

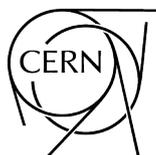
# **Proceedings of the 2018 Asia–Europe–Pacific School of High-Energy Physics**

Quy Nhon, Vietnam, 12–25 September 2018

Editors:

M. Mulders

J. Trân Thanh Vân



ISBN 978-92-9083-571-4 (paperback)  
ISBN 978-92-9083-572-1 (PDF)  
ISSN 2519-8041 (Print)  
ISSN 2519-805X (Online)  
DOI <http://doi.org/10.23730/CYRSP-2020-005>

Copyright © CERN, 2020

 Creative Commons Attribution 4.0

This report should be cited as:

Proceedings of the 2018 Asia–Europe–Pacific School of High-Energy Physics, Quy Nhon, Vietnam, 12–25 September 2018, edited by M. Mulders and J. Trân Thanh Vân, CERN-2020-005 (CERN, Geneva, 2020), <http://doi.org/10.23730/CYRSP-2020-005>

A contribution in this report should be cited as:

[Author name(s)], in Proceedings of the 2018 Asia–Europe–Pacific School of High-Energy Physics, Quy Nhon, Vietnam, 12–25 September 2018, edited by M. Mulders and J. Trân Thanh Vân, CERN-2020-005 (CERN, Geneva, 2020), pp. [first page]–[last page], <http://doi.org/10.23730/CYRSP-2020-005>. [first page]

Corresponding editor: [Martijn.Mulders@cern.ch](mailto:Martijn.Mulders@cern.ch)

Accepted by the [CERN Reports Editorial Board](#) (contact [Carlos.Lourenco@cern.ch](mailto:Carlos.Lourenco@cern.ch)) in September 2020.

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

Indexed in the [CERN Document Server](#), [INSPIRE](#) and in [Scopus](#).

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

## **Abstract**

The Asia–Europe–Pacific School of High-Energy Physics is intended to give young physicists an introduction to the theoretical aspects of recent advances in elementary particle physics. These proceedings contain lecture notes on the theory of quantum chromodynamics, neutrino physics, flavour physics and CP violation, heavy-ion physics, and practical statistics for particle physics.



## Preface

The fourth event in the series of the Asia–Europe–Pacific School of High-Energy Physics took place in Quy Nhon, Vietnam, from 12 to 25 September 2018. A strong team from the Rencontres du Vietnam and the International Center for Interdisciplinary Science and Education (ICISE) took care of the local organization, while CERN and KEK collaborated to provide administrative support in preparation for the School.

The staff and students were housed in comfortable accommodation at the Seagull Hotel in Quy Nhon, whilst excellent conference facilities were provided by the ICISE a short bus ride outside of the city. The students shared accommodation, mixing nationalities to foster cultural exchange between participants from different countries.

A total of 90 students of 28 different nationalities attended the school. About 60% of the students were from Asia–Pacific countries, most of the others coming from Europe, with 42% female students. The majority of the participants were working towards a PhD, while most of the others were advanced Masters students; the School was also open to postdocs.

A total of 32 lectures were complemented by daily discussion sessions led by five discussion leaders. The teachers (lecturers and discussion leaders) came from many different countries: China, France, Germany, Hong Kong, Israel, Japan, Pakistan, Russia, Spain, Switzerland, the UK and the USA.

The programme required the active participation of the students. In addition to the discussion sessions that addressed questions from the lecture courses, there was an evening session in which many students presented posters about their own research work to their colleagues and the teaching staff.

Collaborative student projects in which the students of each discussion group worked together on an in-depth study of a published experimental data analysis were an important activity. This required interacting, outside of the formal teaching sessions, with colleagues from different countries and different cultures. A student representative of each of the five groups presented a short summary of the conclusions of the group's work in a special evening session.

In addition to the academic side of the School, the participants had the occasion to experience many aspects of Vietnamese culture, including visits to cultural sites in and around Quy Nhon, including the Thien Hung Pagoda as well as the Da Dia Reef. They also had ample opportunity to appreciate excellent Vietnamese food.

Our thanks go to the fantastic local-organization team based at the ICISE under the chairmanship of Jean Tran Thanh Van, President of the Rencontres du Vietnam, for all their work and assistance in preparing the School, on both scientific and practical matters, and for their presence throughout the event. Very great thanks are due to the lecturers and discussion leaders for their active participation in the School and for making the scientific programme so stimulating. The students, who in turn manifested their good spirits during two intense weeks, undoubtedly appreciated listening to and discussing with the teaching staff of world renown.

We would like to express our special appreciation to Professor Fabiola Gianotti, Director General of CERN, for participating in a Q&A Session with the students via video link.

We are very grateful to Kate Ross from CERN and to Yoshino Hayes from KEK for their untiring efforts on administration for the School. We would also like to thank the members of the International Committees.

It was a pleasure to welcome high-level guests during the School, including Mr Trần Châu, vice President of the Popular Committee of the province of Binh Dinh in charge of Science and Technology, and Mr. Đỗ Ngọc My, President of the University of Quy Nhon. Their interest and support for the teaching of fundamental science, and, more generally, international collaboration and cultural exchange, were greatly appreciated.

Sponsorship from numerous bodies in many countries covered the cost of travel and/or local expenses of their staff and students who attended the School. In addition, general sponsorship is gratefully acknowledged from: the Rencontres du Vietnam and the Ministry of Science and Technology, Vietnam; CEA/Irfu and CNRS/IN2P3, France; CERN; DESY, Germany; and KEK, Japan.

Nick Ellis  
(Chair of the International Organizing Committee)

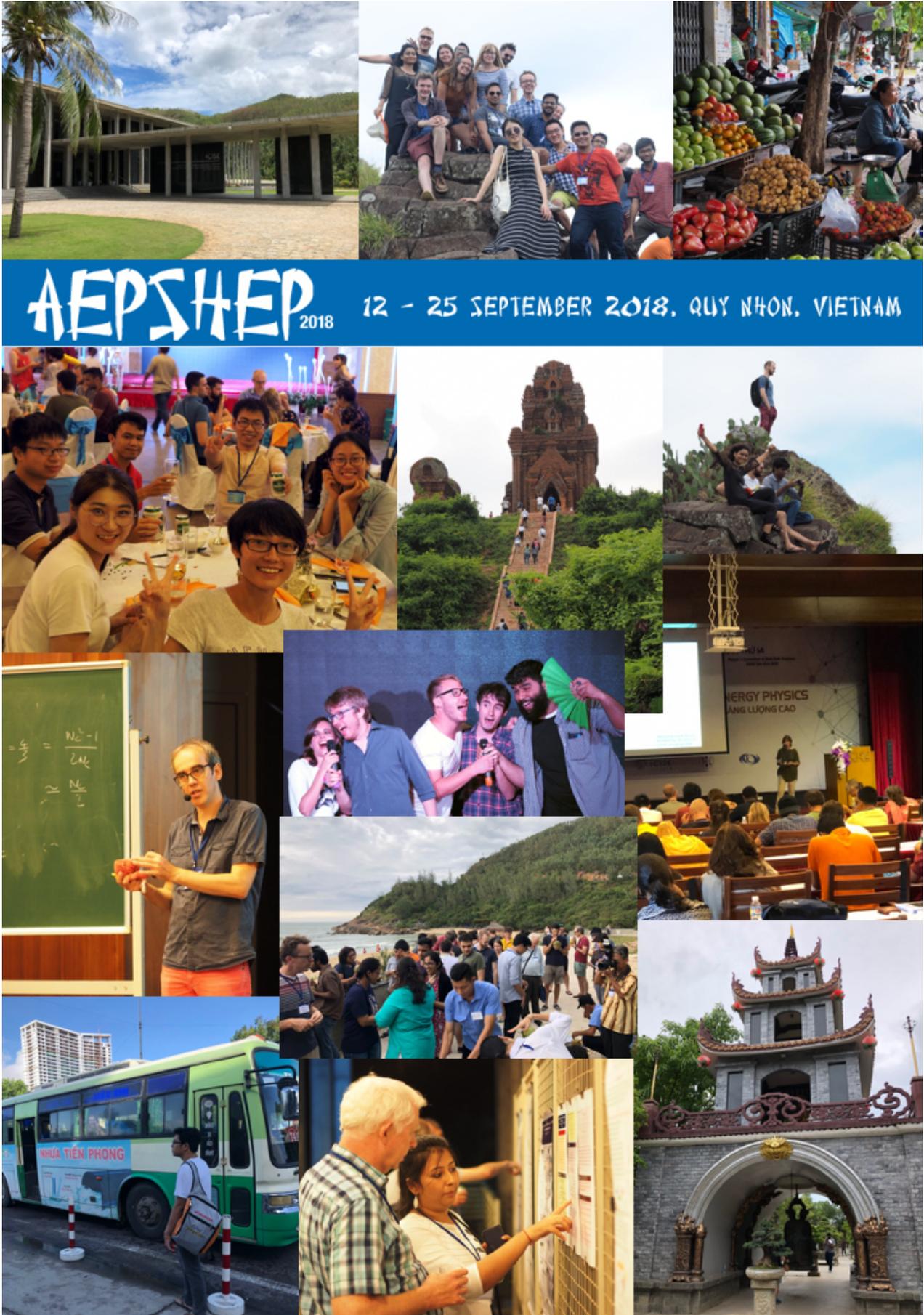




## People in the photograph

1 Thi Hai Lien Tran	39 Diana Seitova	77 Seungcheol Lee
2 Lily Ho	40 Mitsuaki Nozaki	78 Jaydip Singh
3 Saliha Bashir	41 Xinxin Ma	79 Emery Nibigira
4 Huda Siddiqui	42 Andrzej Novak	80 Evgeny Kozyrev
5 Lata Panwar	43 Tigran Mkrtchyan	81 Fariha Vardag
6 Shreya Roy	44 unused number	82 Ondrej Penc
7 Meenakshi Sharma	45 Kevin Floeh	83 Peter Krachkov
8 Duc Ninh Le	46 Dingyu Shao	84 Kosuke Takeda
9 Wadut Shaikh	47 Dajeong Jeon	85 Faiza Khalid
10 Uttiya Sarkar	48 Andrej Arbuzov	86 Yi Zheng
11 Martijn Mulders	49 Thi Thanh Binh Nguyen	87 Debashis Sahoo
12 Daisy Kalra	50 Siew Yan Hoh	88 Naoki Yamaguchi
13 Patrick Bolton	51 Loc TRAN QUANG	89 Marta Babicz
14 Kaustav Chakraborty	52 Kunlin Han	90 Suxian Li
15 Michael Adersberger	53 Hicham Atmani	91 Kate Ross
16 Lucas Altenkamper	54 Hao Zhang	92 Aoife Bharucha
17 Sonia Parmar	55 Dennis Arogancia	93 Iqra Sohail
18 Tran Van Ngoc	56 Shibasis Roy	94 Riccardo Salvatico
19 Châu Tran	57 Minh Truong Nguyen	95 Shan Gu
20 Roger Barlow	58 Quang Hoc Truong	96 Maria Brigida Brunetti
21 Eshwen Bhal	59 Muhammad Mubasher	97 Niharika Rout
22 Jahid Hossain	60 Jhuma Ghosh	98 Ying-Rui Hou
23 Van Tran Thanh	61 Feng-Kun Guo	99 Emmy Gabriel
24 Bharati Naik	62 Vitaliy Popov	100 Maria Onaas
25 Tomas Nosek	63 Grzegorz Żarnecki	101 Jun Koga
26 Eli Rye	64 Dmitrii Neverov	102 Priya Maji
27 Michaela Mlynarikova	65 Kim Smith	103 Xiaoqing Yuan
28 Lukas Sohl	66 Andrew Cohen	104 Xiao Zhao
29 Valentina Zhukova	67 William Barbe	105 Jennifer Zonneveld
30 Rizwan Khalid	68 Thanh Tien Dat Nguyen	106 Yifan Chen
31 Nick Ellis	69 Shyam Balaji	107 Marco Boretto
32 Andrea Cardini	70 Katja Mankinen	108 Hong Van Nguyen
33 Daniel Briglan	71 Dmitrii Pereima	109 Maryvonne Joguet
34 Simen Hellesund	72 Pallabi Das	110 Aimie Fong
35 Luis Pascual Dominguez	73 Arunika Sahu	111 Johanne Le
36 James Kendrick	74 Nizar Septian	112 Thanh Son Tran
37 Suman Chatterjee	75 Dinh Kha Vo	
38 Manoj Kumar Singh	76 Dinh Hung Nguyen	

# PHOTOGRAPHS (MONTAGE)







# Contents

Preface	
<i>N. Ellis</i> .....	v
Photograph of participants .....	vii
Photographs (montage) .....	x
Elements of QCD for hadron colliders	
<i>G.P. Salam</i> .....	1
A $\nu$ hope	
<i>G. Barenboim</i> .....	57
Flavour physics and CP violation	
<i>Y. Nir</i> .....	79
Heavy-ion physics	
<i>M. Nguyen</i> .....	129
Practical statistics for particle physics	
<i>R.J. Barlow</i> .....	149
Scientific Programme .....	199
Organizing Committees .....	200
List of Lecturers .....	200
List of Discussion Leaders .....	201
List of Students .....	202
List of Posters .....	203



## Elements of QCD for hadron colliders

*G. P. Salam*

CERN, Department of Theoretical Physics, Geneva, Switzerland

University of Oxford, Oxford, United Kingdom

### Abstract

The aim of these lectures is to provide students with an introduction to some of the core concepts and methods of QCD that are relevant in an LHC context.

### Keywords

Lectures; QCD; collider physics; parton distribution functions; DGLAP evolution; fixed order; parton showers.

## 1 Introduction

Quantum Chromodynamics, QCD, is the theory of quarks, gluons and their interactions. It is central to all modern colliders. And, for the most part, it is what we are made of.

QCD bears a number of similarities to Quantum Electrodynamics (QED). Just as electrons carry the QED charge, i.e., electric charge, quarks carry the QCD charge, known as colour charge. Whereas there is only one kind of electric charge, colour charge comes in three varieties, sometimes labelled red, green and blue. Anti-quarks have corresponding anti-colour. The gluons in QCD are a bit like the photons of QED. But while photons are electrically neutral, gluons are not colour neutral. They can be thought of as carrying both colour charge and anti-colour charge. There are eight possible different combinations of (anti)colour for gluons. Another difference between QCD and QED lies in its coupling  $\alpha_s$ . In QCD it is only moderately small, it tends to zero at high momentum scales (asymptotic freedom, QED does the opposite), it blows up at small scales, and in between its evolution with scale is quite fast: at the LHC its value will range from  $\alpha_s = 0.08$  at a scale of 5 TeV, to  $\alpha_s \sim 1$  at a scale of 0.5 GeV. These differences between QCD and QED contribute to making QCD a much richer theory.

In these lectures I will attempt to give you a feel for how QCD works at high momentum scales, and for the variety of techniques used by theorists in order to handle QCD at today's high-energy colliders. The hope is that these basics will come in useful for day-to-day work with the QCD facets of hadron collider physics. In the fifty or so pages of these lectures, it will be impossible to give full treatment of any of the topics we will encounter. For that the reader is referred to any of the classic textbooks about QCD at colliders [1–3].

### 1.1 The Lagrangian and colour

Let us start with a brief reminder of the components of the QCD Lagrangian. This section will be rather dense, but we will return to some of the points in more detail later. As already mentioned, quarks come in three colours. So rather than representing them with a single spinor  $\psi$ , we will need the spinor to carry also a colour index  $a$ , which runs from 1 to 3,

$$\psi_a = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix}. \quad (1)$$

The quark part of the Lagrangian (for a single flavour) can be written

$$\mathcal{L}_q = \bar{\psi}_a (i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t_{ab}^C A_\mu^C - m) \psi_b, \quad (2)$$

where the  $\gamma^\mu$  are the usual Dirac matrices; the  $\mathcal{A}_\mu^C$  are gluon fields, with a Lorentz index  $\mu$  and a colour index  $C$  that goes from 1 . . . 8. Quarks are in the fundamental representation of the SU(3) (colour) group, while gluons are in the adjoint representation. Each of the eight gluon fields acts on the quark colour through one of the ‘generator’ matrices of the SU(3) group, the  $t_{ab}^C$  factor in Eq. (2). One convention for writing the matrices is  $t^A \equiv \frac{1}{2}\lambda^A$  with

$$\lambda^1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda^4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$\lambda^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \lambda^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \lambda^7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \lambda^8 = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & \frac{-2}{\sqrt{3}} \end{pmatrix}.$$

By looking at the first of these, together with the  $t_{ab}^C \mathcal{A}_\mu^C \psi_b$  term of  $\mathcal{L}_Q$ , one can immediately get a feel for what gluons do: a gluon with (adjoint) colour index  $C = 1$  acts on quarks through the matrix  $t^1 = \frac{1}{2}\lambda^1$ . That matrix takes green quarks ( $b = 2$ ) and turns them into red quarks ( $a = 1$ ), and vice versa. In other words, when a gluon interacts with a quark it *repaints* the colour of the quark, taking away one colour and replacing it with another. The likelihood with which this happens is governed by the strong coupling constant  $g_s$ . Note that the repainting analogy is less evident for some of the other colour matrices, but it still remains essentially correct.

The second part of the QCD Lagrangian is purely gluonic

$$\mathcal{L}_G = -\frac{1}{4}F_A^{\mu\nu}F^{A\mu\nu} \quad (3)$$

where the gluon field tensor  $F_{\mu\nu}^A$  is given by

$$F_{\mu\nu}^A = \partial_\mu \mathcal{A}_\nu^A - \partial_\nu \mathcal{A}_\mu^A - g_s f_{ABC} \mathcal{A}_\mu^B \mathcal{A}_\nu^C \quad [t^A, t^B] = i f_{ABC} t^C, \quad (4)$$

where the  $f_{ABC}$  are the structure constants of SU(3) (defined through the commutators of the  $t^A$  matrices). Note the major difference with QED here, namely the presence of a term  $g_s f_{ABC} \mathcal{A}_\mu^B \mathcal{A}_\nu^C$  with two gluon fields. The presence of such a term is one of the major differences with QED, and, as we will discuss in more detail below, it will be responsible for the fact that gluons interact directly with gluons. For now, note simply that it has to be there in order for the theory to be gauge invariant under local SU(3) transformations:

$$\psi_a \rightarrow e^{i\theta^C(x)t_{ab}^C} \psi_b \quad (5)$$

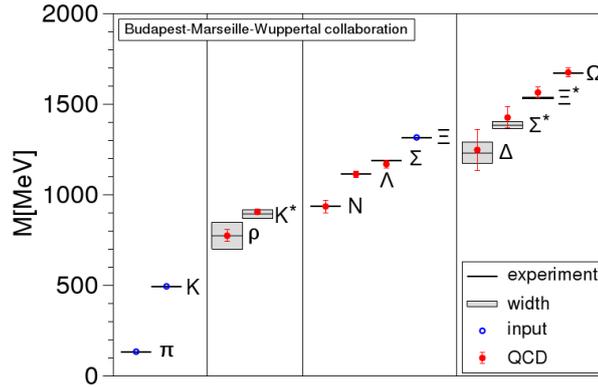
$$\mathcal{A}^C t^C \rightarrow e^{i\theta^D(x)t^D} \left( \mathcal{A}^C t^C - \frac{1}{g_s} \partial_\mu \theta^C(x) t^C \right) e^{-i\theta^E(x)t^E} \quad (6)$$

where, in the second line, we have dropped the explicit subscript  $ab$  indices, and the  $\theta^C(x)$  are eight arbitrary real functions of the space-time position  $x$ .

## 1.2 ‘Solving QCD’

There are two main first-principles approaches to solving QCD: lattice QCD and perturbative QCD.<sup>1</sup>

<sup>1</sup>In addition, effective-theory methods provide ways of looking at QCD that make it easier to solve, given certain ‘inputs’ that generally come from lattice or perturbative QCD (and sometimes also from experimental measurements). These lectures won’t discuss effective theory methods, but for more details you may consult the lectures at this school by Martin Beneke. Another set of methods that has seen much development in recent years makes use of the ‘AdS/CFT’ correspondence [4–6], relating QCD-like models at strong coupling to gravitational models at weak coupling (e.g., [7,8]).



**Fig. 1:** The measured spectrum of hadron masses, compared to a lattice calculation [9]. The open blue circles are the hadron masses that have been used to fix the three parameters of the calculation: the value of the QCD coupling, the average of the up and down quark masses (taken equal) and the strange-quark mass. All other points are results of the calculation.

### 1.2.1 Lattice QCD

The most complete approach is lattice QCD. It involves discretizing space-time, and considering the values of the quark and gluon fields at all the vertices/edges of the resulting 4-dimensional lattice (with imaginary time). Through a suitable Monte Carlo sampling over all possible field configurations, one essentially determines the relative likelihood of different field configurations, and this provides a solution to QCD. This method is particularly suited to the calculation of static quantities in QCD such as the hadron mass spectrum. The results of such a lattice calculation are illustrated in Fig. 1, showing very good agreement.

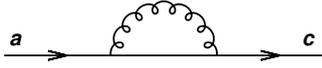
Lattice methods have been successfully used in a range of contexts, for example, in recent years, in helping extract fundamental quantities such as the CKM matrix (and limits on new physics) from the vast array of experimental results on hadron decays and oscillations at flavour factories. Unfortunately lattice calculations aren't suitable in all contexts. Let us imagine, briefly, what would be required in order to carry out lattice calculations for LHC physics: since the centre-of-mass energy is (will be) 14 TeV, we need a lattice spacing of order  $1/(14 \text{ TeV}) \sim 10^{-5} \text{ fm}$  to resolve everything that happens. Non-perturbative dynamics for quarks/hadrons near rest takes place on a timescale  $t \sim \frac{1}{0.5 \text{ GeV}} \sim 0.4 \text{ fm}/c$ . But hadrons at LHC have a boost factor of up to  $10^4$ , so the extent of the lattice should be about 4000 fm. That tells us that if we are to resolve high-momentum transfer interactions and at the same time follow the evolution of quark and gluon fields up to the point where they form hadrons, we would need about  $4 \times 10^8$  lattice units in each direction, of  $\sim 3 \times 10^{34}$  nodes. Not to mention the problem with high particle multiplicities (current lattice calculations seldom involve more than two or three particles) and all the issues that relate to the use of imaginary time in lattice calculations. Of course, that's not to say that it might not be possible, one day, to find clever tricks that would enable lattice calculations to deal with high-energy reactions. However, with today's methods, any lattice calculation of the properties of LHC proton-proton scattering seems highly unlikely. For this reason, we will not give any further discussion of lattice QCD here, but instead refer the curious reader to textbooks and reviews for more details [10–13].

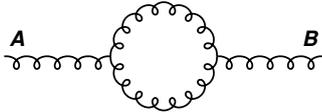
### 1.2.2 Perturbative QCD

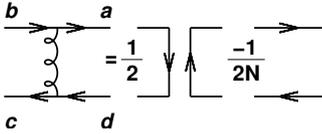
Perturbative QCD relies on the idea of an order-by-order expansion in a small coupling  $\alpha_s = \frac{g_s^2}{4\pi} \ll 1$ . Some given observable  $f$  can then be predicted as

$$f = f_1\alpha_s + f_2\alpha_s^2 + f_3\alpha_s^3 + \dots, \quad (7)$$



$$\sum_A t_{ab}^A t_{bc}^A = C_F \delta_{ac}, \quad C_F = \frac{N_C^2 - 1}{2N_C} = \frac{4}{3} \quad \text{(8b)}$$


$$\sum_{C,D} f^{ACD} f^{BCD} = C_A \delta^{AB}, \quad C_A = N_C = 3 \quad \text{(8c)}$$


$$t_{ab}^A t_{cd}^A = \frac{1}{2} \delta_{bc} \delta_{ad} - \frac{1}{2N_C} \delta_{ab} \delta_{cd} \quad \text{(Fierz)} \quad \text{(8d)}$$


where  $N \equiv N_C = 3$  is the number of colours in QCD and it is useful to express the results for general numbers of colours (because it is sometimes useful to consider how results depend on  $N_C$ , especially in the limit  $N_C \rightarrow \infty$ ). Each mathematical combination of colour factors has a diagrammatic interpretation. Equation (8a) corresponds to a gluon splitting into  $q\bar{q}$  which then join back into a gluon; or, the sum over colours in the squared amplitude for  $g \rightarrow q\bar{q}$ . Equation (8b) corresponds to the square of gluon emission from a quark. Equation (8c) arises as the square of gluon emission from a gluon. One sees that there is almost a factor of 2 between Eqs. (8b) and (8c) (modulo corrections terms  $\sim 1/N_C$ ), which is the mathematical counterpart of our statement above that gluons emit twice as strongly as quarks. Finally the approximate colour-flow interpretation that we had in Fig. 3 (left) can be stated exactly in terms of the Fierz identity, Eq. (8d).

### 1.2.3 The running coupling

Most higher-order QCD calculations are carried out with dimensional regularization (the use of  $4 - \epsilon$  dimensions) in order to handle the ultraviolet divergences that appear in loop diagrams. In the process of going from 4 to  $4 - \epsilon$  dimensions, one needs to introduce an arbitrary ‘renormalization’ scale, generally called  $\mu$ , in order to keep consistent dimensions (units) for all quantities.<sup>2</sup> The value of the QCD coupling,  $\alpha_s = \frac{g_s^2}{4\pi}$ , depends on the scale  $\mu$  at which it is evaluated. That dependence can be expressed in terms of a renormalization group equation

$$\frac{d\alpha_s(\mu^2)}{d \ln \mu^2} = \beta(\alpha_s(\mu^2)), \quad \beta(\alpha_s) = -\alpha_s^2 (b_0 + b_1 \alpha_s + b_2 \alpha_s^2 + \dots), \quad \text{(9)}$$

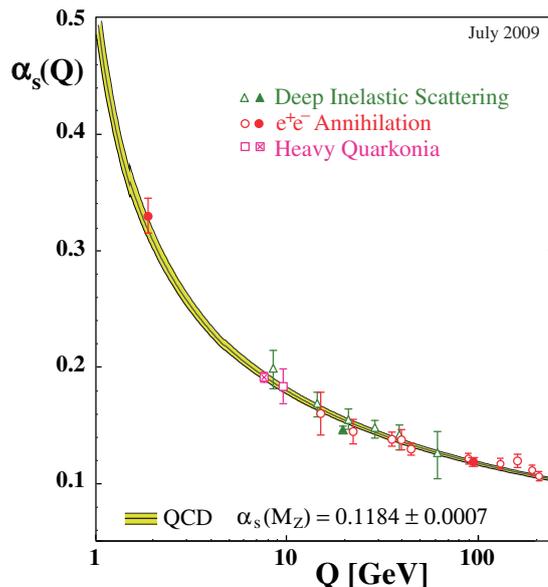
where

$$b_0 = \frac{11C_A - 2n_f}{12\pi}, \quad b_1 = \frac{17C_A^2 - 5C_A n_f - 3C_F n_f}{24\pi^2} = \frac{153 - 19n_f}{24\pi^2}, \quad \text{(10)}$$

with  $n_f$  being the number of ‘light’ quark flavours, those whose mass is lower than  $\mu$ . The negative sign in Eq. (9) is the origin of asymptotic freedom, the fact that the coupling becomes weaker at high momentum scales, i.e., the theory almost becomes a free theory, in which quarks and gluons don’t interact. Conversely at low momentum scales the coupling grows strong, causing quarks and gluons to be tightly bound into hadrons. The importance of the discovery of these features was recognized in the 2004 Nobel prize to Gross, Politzer and Wilczek. Why does the QCD  $\beta$ -function have the opposite sign of that in QED? The fact that the vector particles (gluons) of the theory carry colour charge is central to the result. However, while there have been various attempts to give simple but accurate explanations for the negative sign [15, 16], in practice they all end up being quite involved.<sup>3</sup> So, for the purpose of these lectures, let us just accept the results.

<sup>2</sup>The renormalization procedure itself, i.e., the removal of the  $1/\epsilon$  divergences, is usually carried out in the modified minimal subtraction ( $\overline{\text{MS}}$ ) scheme (see, e.g., Section 11.4 of Ref. [14]), by far the most widespread scheme in QCD.

<sup>3</sup>You might still want to check the sign for yourself: if so, pick up a copy of Peskin and Schroeder [14], arrange to have an afternoon free of interruptions, and work through the derivation.



**Fig. 4:** The QCD coupling as measured in physics processes at different scales  $Q$ , together with the band obtained by running the world average for  $\alpha_s$  within its uncertainties. Figure taken from Ref. [17].

If we ignore all terms on the right of Eq. (9) other than  $b_0$ , and also ignore the subtlety that the number of ‘light’ flavours  $n_f$  depends on  $\mu$ , then there is a simple solution for  $\alpha_s(\mu^2)$ :

$$\alpha_s(\mu^2) = \frac{\alpha_s(\mu_0^2)}{1 + b_0 \alpha_s(\mu_0^2) \ln \frac{\mu^2}{\mu_0^2}} = \frac{1}{b_0 \ln \frac{\mu^2}{\Lambda^2}}, \quad (11)$$

where one can either express the result in terms of the value of the coupling at a reference scale  $\mu_0$ , or in terms of a non-perturbative constant  $\Lambda$  (also called  $\Lambda_{QCD}$ ), the scale at which the coupling diverges. Only for scales  $\mu \gg \Lambda$ , corresponding to  $\alpha_s(\mu^2) \ll 1$ , is perturbation theory valid. Note that  $\Lambda$ , since it is essentially a non-perturbative quantity, is not too well defined: for a given  $\alpha_s(\mu_0)$ , its value depends on whether we used just  $b_0$  in Eq. (9) or also  $b_1$ , etc. However, its order of magnitude, 200 MeV, is physically meaningful insofar as it is closely connected with the scale of hadron masses.

One question that often arises is how  $\mu$ , the renormalization scale, should relate to the physical scale of the process. We will discuss this in detail later (Section 4.1), but for now the following simple statement is good enough: the strength of the QCD interaction for a process involving a momentum transfer  $Q$  is given by  $\alpha_s(\mu)$  with  $\mu \sim Q$ . One can measure the strength of that interaction in a range of processes, at various scales, and Fig. 4 [17] shows a compilation of such measurements, together with the running of an average over many measurements,  $\alpha_s(M_Z) = 0.1184 \pm 0.0007$ , illustrating the good consistency of the measurements with the expected running.

#### 1.2.4 QCD predictions and colliders

Colliders like the Tevatron and the LHC are mainly geared to investigating phenomena involving high-momentum transfers (more precisely large transverse-momenta), say in the range 50 GeV to 5 TeV. There, the QCD coupling is certainly small and we would hope to be able to apply perturbation theory. Yet, the initial state involves protons, at whose mass scale,  $m_p \simeq 0.94$  GeV, the coupling is certainly not weak. And the final states of collider events consist of lots of hadrons. Those aren’t perturbative either. And there are lots of them — tens to hundreds. Even if we wanted to try, somehow, to treat them perturbatively, we would be faced with calculations to some very high order in  $\alpha_s$ , at least as high as the particle multiplicity, which is far beyond what we can calculate exactly: depending on how you count,

at hadron colliders, the best available complete calculation (i.e., all diagrams at a given order), doesn't go beyond  $\alpha_s^2$  or  $\alpha_s^3$ . Certain subsets of diagrams (e.g., those without loops) can be calculated up to  $\alpha_s^{10}$  roughly.

So we are faced with a problem. Exact lattice methods can't deal with the high momentum scales that matter, exact perturbative methods can't deal with low momentum scales that inevitably enter the problem, nor the high multiplicities that events have in practice. Yet, it turns out that we are reasonably successful in making predictions for collider events. These lectures will try to give you an understanding of the methods and approximations that are used.

## 2 Considering $e^+e^- \rightarrow$ hadrons

One simple context in which QCD has been extensively studied over the past 30 years is that of  $e^+e^-$  annihilation to hadrons. This process has the theoretical advantage that only the final state involves QCD. Additionally, huge quantities of data have been collected at quite a number of colliders, including millions of events at the  $Z$  mass at LEP and SLC. We therefore start our investigation of the properties of QCD by considering this process.

### 2.1 Soft and collinear limits

There is one QCD approximation that we will repeatedly make use of, and that is the soft and collinear approximation. 'Soft' implies that an emitted gluon has very little energy compared to the parton (quark or gluon) that emitted it. 'Collinear' means that it is emitted very close in angle to another parton in the event. By considering gluons that are soft and/or collinear one can drastically simplify certain QCD calculations, while still retaining much of the physics.

The soft and collinear approximation is sufficiently important that it's worth, at least once, carrying out a calculation with it, and we'll do that in the context of the emission of a gluon from  $e^+e^- \rightarrow q\bar{q}$  events. Though there are quite a few equations in the page that follows, the manipulations are all quite simple! We're interested in the hadronic side of the  $e^+e^- \rightarrow q\bar{q}$  amplitude, so let's first write the QED matrix element for a virtual photon  $\gamma^* \rightarrow q\bar{q}$  (we can always put back the  $e^+e^- \rightarrow \gamma^*$  and the photon propagator parts later if we need to — which we won't):

$$\mathcal{M}_{q\bar{q}} = \bar{u}_a(p_1) i e_q \gamma_\mu \delta_{ab} v_b(p_2) \quad \begin{array}{c} \nearrow p_1 \\ \text{---} ie \gamma_\mu \text{---} \\ \searrow p_2 \end{array},$$

where the diagram illustrates the momentum labelling. Here  $\bar{u}(p_1)$  and  $v(p_2)$  are the spinors for the outgoing quark and anti-quark (taken massless),  $e_q$  is the quark's electric charge and the  $\gamma_\mu$  are the Dirac matrices. In what follows we shall drop the  $a, b$  quark colour indices for compactness and reintroduce them only at the end.

The corresponding amplitude including the emission of a gluon with momentum  $k$  and polarization vector  $\epsilon$  is

$$\begin{aligned} \mathcal{M}_{q\bar{q}g} &= \begin{array}{c} \nearrow p_1 \\ \text{---} ie \gamma_\mu \text{---} \\ \text{---} \epsilon \text{---} k, \epsilon \\ \searrow p_2 \end{array} + \begin{array}{c} \nearrow p_1 \\ \text{---} ie \gamma_\mu \text{---} \\ \text{---} \epsilon \text{---} k, \epsilon \\ \searrow p_2 \end{array} \\ &= -\bar{u}(p_1) i g_s \not{\epsilon} t^A \frac{i(\not{p}_1 + \not{k})}{(p_1 + k)^2} i e_q \gamma_\mu v(p_2) + \bar{u}(p_1) i e_q \gamma_\mu \frac{i(\not{p}_2 + \not{k})}{(p_2 + k)^2} i g_s \not{\epsilon} t^A v(p_2), \end{aligned} \quad (12)$$

with one term for emission from the quark and the other for emission from the anti-quark and use of the notation  $\not{p} = p_\mu \gamma_\mu$ . Let's concentrate on the first term, collecting the factors of  $i$ , and using the

anti-commutation relation of the  $\gamma$ -matrices,  $\mathbf{A}\mathbf{B} = 2A.B - \mathbf{B}\mathbf{A}$ , to write

$$i\bar{u}(p_1)g_s\epsilon t^A \frac{(\not{p}_1 + \not{k})}{(p_1 + k)^2} e_q \gamma_\mu v(p_2) = ig_s \bar{u}(p_1) \frac{[2\epsilon.(p_1 + k) - (\not{p}_1 + \not{k})\not{\epsilon}]}{(p_1 + k)^2} e_q \gamma_\mu t^A v(p_2), \quad (13a)$$

$$\simeq ig_s \frac{p_1.\epsilon}{p_1.k} \bar{u}(p_1) e_q \gamma_\mu t^A v(p_2), \quad (13b)$$

where to obtain the second line we have made use of the fact that  $\bar{u}(p_1)\not{p}_1 = 0$ ,  $p_1^2 = k^2 = 0$ , and taken the soft approximation  $k_\mu \ll p_\mu$ , which allows us to neglect the terms in the numerator that are proportional to  $k$  rather than  $p$ . The answer including both terms in Eq. (12) is

$$\mathcal{M}_{q\bar{q}g} \simeq \bar{u}(p_1) i e_q \gamma_\mu t^A v(p_2) \cdot g_s \left( \frac{p_1.\epsilon}{p_1.k} - \frac{p_2.\epsilon}{p_2.k} \right), \quad (14)$$

where the first factor has the Lorentz structure of the  $\mathcal{M}_{q\bar{q}}$  amplitude, i.e., apart from the colour matrix  $t^A$ ,  $\mathcal{M}_{q\bar{q}}$  is simply proportional to the  $\mathcal{M}_{q\bar{q}}$  result. We actually need the squared amplitude, summed over polarizations and colour states,

$$\begin{aligned} |\mathcal{M}_{q\bar{q}g}|^2 &\simeq \sum_{A,a,b,\text{pol}} \left| \bar{u}_a(p_1) i e_q \gamma_\mu t^A v_b(p_2) g_s \left( \frac{p_1.\epsilon}{p_1.k} - \frac{p_2.\epsilon}{p_2.k} \right) \right|^2 \\ &= -|\mathcal{M}_{q\bar{q}}|^2 C_F g_s^2 \left( \frac{p_1}{p_1.k} - \frac{p_2}{p_2.k} \right)^2 = |\mathcal{M}_{q\bar{q}}|^2 C_F g_s^2 \frac{2p_1.p_2}{(p_1.k)(p_2.k)}. \end{aligned} \quad (15)$$

We have now explicitly written the quark colour indices  $a, b$  again. To obtain the second line we have made use of the result that  $\sum_{A,a,b} t_{ab}^A t_{ba}^A = C_F N_C$  [cf. Eq. (8b)], whereas for  $|\mathcal{M}_{q\bar{q}}|^2$  we have  $\sum_{A,a,b} \delta_{ab} t_{ba}^A = N_C$ . To carry out the sum over gluon polarizations we have exploited the fact that  $\sum_{\text{pol}} \epsilon_\mu(k) \epsilon_\nu^*(k) = -g_{\mu\nu}$ , plus terms proportional to  $k_\mu$  and  $k_\nu$  that disappear when dotted with the amplitude and its complex conjugate.

One main point of the result here is that in the soft limit, the  $|\mathcal{M}_{q\bar{q}g}|^2$  squared matrix element *factorizes* into two terms: the  $|\mathcal{M}_{q\bar{q}}|^2$  matrix element and a piece with a rather simple dependence on the gluon momentum.

The next ingredient that we need is the the phase space for the  $q\bar{q}g$  system,  $d\Phi_{q\bar{q}g}$ . In the soft approximation, we can write this

$$d\Phi_{q\bar{q}g} \simeq d\Phi_{q\bar{q}} \frac{d^3\vec{k}}{2E(2\pi)^3}, \quad (16)$$

where  $E \equiv E_k$  is the energy of the gluon  $k$ . We see that the phase space also factorizes. Thus we can write the full differential cross section for  $q\bar{q}$  production plus soft gluon emission as the  $q\bar{q}$  production matrix element and phase space,  $|\mathcal{M}_{q\bar{q}}|^2 d\Phi_{q\bar{q}}$ , multiplied by a soft gluon emission probability  $d\mathcal{S}$ ,

$$|\mathcal{M}_{q\bar{q}g}|^2 d\Phi_{q\bar{q}g} \simeq |\mathcal{M}_{q\bar{q}}|^2 d\Phi_{q\bar{q}} d\mathcal{S}, \quad (17)$$

with

$$d\mathcal{S} = E dE d\cos\theta \frac{d\phi}{2\pi} \cdot \frac{2\alpha_s C_F}{\pi} \frac{2p_1.p_2}{(2p_1.k)(2p_2.k)}, \quad (18)$$

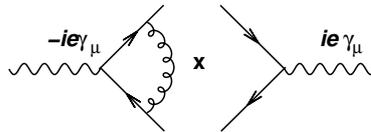
where we have used  $d^3k = E^2 dE d\cos\theta d\phi$ , expressing the result in terms of the polar ( $\theta$ ) and azimuthal ( $\phi$ ) angles of the gluon with respect to the quark (which itself is back-to-back with the antiquark, since we work in the centre-of-mass frame and there is negligible recoil from the soft gluon). With a little more algebra, we get our final result for the probability of soft gluon emission from the  $q\bar{q}$  system

$$d\mathcal{S} = \frac{2\alpha_s C_F}{\pi} \frac{dE}{E} \frac{d\theta}{\sin\theta} \frac{d\phi}{2\pi}. \quad (19)$$

This result has two types of non-integrable divergence: one, called the soft (or infrared) divergence when  $E \rightarrow 0$  and the other, a collinear divergence, when  $\theta \rightarrow 0$  (or  $\pi$ ), i.e., the gluon becomes collinear with the quark (or antiquark) direction. Though derived here in the specific context of  $e^+e^- \rightarrow q\bar{q}$  production, these soft and collinear divergences are a very general property of QCD and appear whenever a gluon is emitted from a quark, regardless of the process.

## 2.2 The total cross section

If we want to calculate the  $\mathcal{O}(\alpha_s)$  corrections to the total cross section, the diagrams included in Eq. (12) are not sufficient. We also need to include a one-loop correction (‘virtual’), specifically, the interference between one-loop  $\gamma^* \rightarrow q\bar{q}$  diagrams and the tree-level  $\gamma^* \rightarrow q\bar{q}$  amplitude, for example a contribution such as



which has the same perturbative order (number of  $g_s$  factors) as the square of Eq. (12).

The total cross section for the production of hadrons must be finite. The integral over the gluon-emission correction has two non-integrable, logarithmic divergences. These divergences must therefore somehow be cancelled by corresponding divergences in the virtual term. This is the requirement of *unitarity*, which is basically the statement that probability of anything happening must add up to 1. The most straightforward way of doing the full calculation for the total cross section is to use dimensional regularization in the phase space integral for the real emission diagram and for the integration over the loop momentum in the virtual diagram. However, in order just to visualize what is happening one can also write

$$\sigma_{tot} = \sigma_{q\bar{q}} \left( 1 + \frac{2\alpha_s C_F}{\pi} \int \frac{dE}{E} \int \frac{d\theta}{\sin\theta} R\left(\frac{E}{Q}, \theta\right) - \frac{2\alpha_s C_F}{\pi} \int \frac{dE}{E} \int \frac{d\theta}{\sin\theta} V\left(\frac{E}{Q}, \theta\right) \right), \quad (20)$$

where the first term, 1, is the ‘Born’ term, i.e., the production of just  $q\bar{q}$ , the second term is the real emission term, and the third term is the loop correction. Since we need to integrate gluon emission beyond the soft and collinear region, we have introduced a function  $R(E/Q, \theta)$ , which parametrizes the deviation of the matrix-element from its soft limit when  $E \sim Q$ , with  $Q$  the centre-of-mass energy of the process.  $R$  has the property  $\lim_{E \rightarrow 0} R(E/Q, \theta) = 1$ . We have written the virtual term in a similar form, using  $V(E/Q, \theta)$  to parametrize its structure (we cheat a bit, since the loop momentum integral includes regions of phase space where the gluon is offshell; this won’t matter though for us here). The statement that real and virtual divergences cancel means that  $V$  should be identical to  $R$  in the soft or collinear limits

$$\lim_{E \rightarrow 0} (R - V) = 0, \quad \lim_{\theta \rightarrow 0, \pi} (R - V) = 0. \quad (21)$$

Thus the corrections to the total cross section come from the region of hard ( $E \sim Q$ ), large-angle gluons (for which perturbation theory is valid). There’s a good reason for this: soft and collinear emission takes place on a time-scale  $\sim 1/(E\theta^2)$  that is long compared to that,  $\sim 1/Q$ , for the production of the  $q\bar{q}$  pair from the virtual photon. Anything that happens long after the production of the  $q\bar{q}$  pair cannot change the fact that there will be a QCD final state (though it can change the properties of that final state), and so it does not affect the total cross section. Similarly, whatever dynamics is involved in effecting the transition between partons and hadrons is also expected to occur on a long time-scale ( $\sim 1/\Lambda$ ) and so should not modify the total cross section. This is important because it allows us to neglect the issue that we cannot directly compute the properties of hadron production.

The fact that the corrections to the total cross section are dominated by a region of hard gluon emission is reflected in a reasonable behaviour for the perturbative series

$$\sigma_{tot} = \sigma_{q\bar{q}} \left( 1 + 1.045 \frac{\alpha_s(Q)}{\pi} + 0.94 \left( \frac{\alpha_s(Q)}{\pi} \right)^2 - 15 \left( \frac{\alpha_s(Q)}{\pi} \right)^3 + \dots \right), \quad (22)$$

where we have expressed the result in terms of  $\alpha_s$  evaluated at a renormalization  $\mu = Q$  and the coefficients correspond to  $Q = M_Z$  (for all known terms in the series, including electroweak effects, see Refs. [18–23] as well as references therein).

### 2.3 The final state

As a first step towards asking questions about the final state, our next exercise is to attempt to determine the mean number of gluons that are emitted from a quark with energy  $\sim Q$ . If the emission probability is small ( $\propto \alpha_s$ ) then to first order in the coupling the mean number of emitted gluons is equal to the probability of emitting one gluon

$$\langle N_g \rangle \simeq \frac{2\alpha_s C_F}{\pi} \int^Q \frac{dE}{E} \int^{\pi/2} \Theta(E\theta > Q_0). \quad (23)$$

The integral diverges for  $E \rightarrow 0$  and  $\theta \rightarrow 0$ , however, we can reasonably argue that the divergent structure only holds as long as perturbation theory is valid. This motivates us to cut the divergences off at a scale  $Q_0 \sim \Lambda$ , because below that scale the language of quarks and gluons loses its meaning. That immediately tells us that we should have  $E \gtrsim Q_0$ , but it's not so immediately clear how the  $\theta$  integral will be cut off. It turns out, for reasons to do with invariance of the small-angle emission pattern as one boosts the quark in the (longitudinal) direction of its motion, that the correct variable to cut on is *transverse momentum*,  $k_t \sim E\theta$ . We therefore find, to first order in the coupling,

$$\langle N_g \rangle \simeq \frac{\alpha_s C_F}{\pi} \ln^2 \frac{Q}{Q_0} + \mathcal{O}(\alpha_s \ln Q), \quad (24)$$

where we have explicitly kept track only of the term with the largest number of logarithms. If we take  $Q_0 = \Lambda$ , how big is this result? We have to decide on the scale for  $\alpha_s$ . Being hopelessly optimistic, i.e., taking  $\alpha_s = \alpha_s(Q) = (2b \ln Q/\Lambda)^{-1}$  gives us

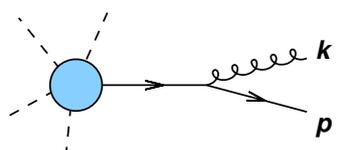
$$\langle N_g \rangle \simeq \frac{C_F}{2b\pi} \ln \frac{Q}{\Lambda} \simeq \frac{C_F}{4b^2\pi\alpha_s}, \quad (25)$$

which, numerically, corresponds to  $\langle N_g \rangle \simeq 2$ . This is neither small numerically, nor parametrically ( $\sim 1/\alpha_s$ ). Does this render perturbation completely useless for anything other than total cross sections?

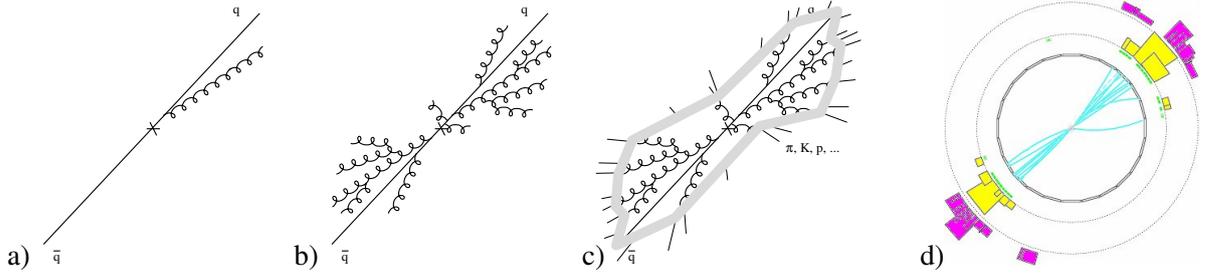
We can follow two possible avenues to help answer this question. One approach is to calculate the next order, and see what structure it has. Alternatively we can ask whether there are final-state observables that have a better-behaved perturbative series than ‘the mean number of gluons’.

#### 2.3.1 Gluon (and hadron) multiplicity

Once one gluon has been emitted, it can itself emit further gluons. To understand what the gluon multiplicity might look like to higher orders, it's useful to write down the general pattern of emission of a soft gluon both from a quark and from a gluon, which is essentially independent of the process that produced the ‘emitter’:



$$\simeq \frac{2\alpha_s C_F}{\pi} \frac{dE}{E} \frac{d\theta}{\theta}, \quad (26a)$$



**Fig. 5:** Emission pattern from a  $q\bar{q}$  event, with first a single gluon (a), then multiple emissions of gluons both from the  $q\bar{q}$  pair and from the previously emitted gluons (b), followed by some process, ‘hadronization’, that causes hadrons to be produced from the gluons, giving an event (c), that structurally resembles a real event (d) ( $e^+e^- \rightarrow Z \rightarrow \text{hadrons}$  at LEP in the OPAL detector)

$$\simeq \frac{2\alpha_s C_A}{\pi} \frac{dE}{E} \frac{d\theta}{\theta}. \quad (26b)$$

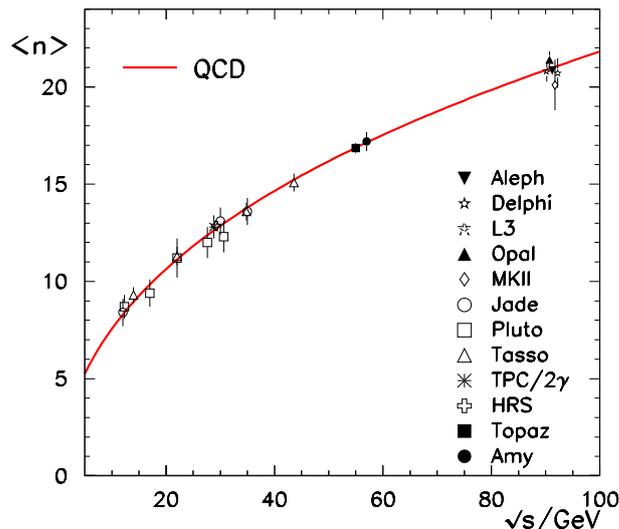
These expressions are valid when the emitted gluon is much lower in energy than the emitter,  $k \ll p$ , and when the emission angle  $\theta$  is much smaller than the angle between the emitter and any other parton in the event (this is known as the condition of angular ordering [24]). The structure of emission of a soft gluon is almost identical from a quark and from a gluon, except for the substitution of the  $C_F = 4/3$  colour factor in the quark case with the  $C_A = 3$  colour factor in the gluon case.

Since quarks and gluons emit in similar ways, every gluon that is emitted from the quark can itself emit further gluons, and so forth. Most of the emissions will either be in almost the same direction as the original quark (due to the collinear divergence) and/or be soft. This is represented in Figs. 5(a) and (b) (for simplicity we’ve not shown gluons splitting to  $q\bar{q}$  pairs, which also occurs, with just a collinear divergence). This still only gives a description of events in terms of quarks and gluons, whereas real events consist of hadrons. Though hadronization, the transition from quarks and gluons to hadrons is not something that we know how to calculate from first principles, one idea that has had some success is Local Parton Hadron Duality (LPHD) (see, e.g., Ref. [25]). It states that after accounting for all gluon and quark production down to scales  $\sim \Lambda$ , the transition from partons to hadrons is essentially local in phase space. Thus the hadron directions and momenta will be closely related to the partons’, and the hadron multiplicity will reflect the parton multiplicity too. This is illustrated in Fig. 5(c), comparing it also to the picture of a real event, Fig. 5(d). The latter illustrates how the hadrons do tend to have the same collimated angular distribution as is predicted for gluons, with the small number of exceptions having low energy (i.e., soft) as can be seen from the larger curvature in the experiment’s magnetic field.

This comparison with a single event is suggestive that our picture of gluon emission and hadronization might be reasonable. A more quantitative test can be obtained by calculating the number of emitted gluons. This requires the extension of Eqs. (23)–(25) to multiple gluon emission. The full calculation doesn’t fit into the space available for these lectures (see instead textbook discussions in Refs. [1, 25]), but the basic idea is that there are terms  $(\alpha_s \ln^2 Q/Q_0)^n$  for all orders  $n$  and that one can calculate their coefficients analytically. The structure of the result is

$$\langle N_g \rangle \sim \frac{C_F}{C_A} \sum_{n=1}^{\infty} \frac{1}{(n!)^2} \left( \frac{C_A}{2\pi b^2 \alpha_s} \right)^n \sim \frac{C_F}{C_A} \exp \left( \sqrt{\frac{2C_A}{\pi b^2 \alpha_s(Q)}} \right), \quad (27)$$

where we’ve neglected to write the prefactor in front of the exponential, and we’ve also not given the subleading terms [26].



**Fig. 6:** Multiplicity of charged hadrons in  $e^+e^- \rightarrow \text{hadrons}$  events, comparing the experimental data at a range of centre-of-mass energies  $Q$ , with the QCD prediction using a fitted normalisation and non-perturbative scale  $\Lambda$ . Figure adapted from Ref. [1].

How is Eq. (27) to be related to the hadron multiplicity? The simplest assumption is that each final parton gives some (unknown) fixed number of hadrons which must be fitted to data. Equation (27) then predicts not the total hadron multiplicity but its energy dependence. That prediction is illustrated in Fig. 6 and shows remarkable agreement with data over a range of energies, providing strong evidence that the picture outlined above is a fair reflection of ‘reality’.

The above approach can be extended to calculate other properties of events such as the energy spectrum of hadrons, the fluctuations in the number of hadrons, and even correlations between hadrons, generally with reasonable success. However, as one probes more detailed features of events, the analytical calculations become significantly more complicated and one also becomes increasingly sensitive to the oversimplicity of the LPHD concept. Having said that, the same ideas that we are using, i.e., the importance of multiple soft and collinear splitting together with a transition from partons to hadrons, are, in numerical form, at the base of widely used Monte Carlo parton-shower event generators like PYTHIA, HERWIG and SHERPA. We will discuss them in more detail in Section 4.2.

### 2.3.2 Infrared safe observables

It is heartening that the above soft-collinear discussion gave such a good description of the data. However, it did involve the application of perturbation theory to kinematic regions where its validity is questionable, the need to calculate dominant contributions at all orders in  $\alpha_s$ , and the introduction of a free parameter to ‘fudge’ the fact that we don’t understand the non-perturbative physics. A natural question is therefore whether one can formulate final-state observables for which these problems are not present.

The answer is that one can. For an observable to be calculated based on just one or two orders of perturbation theory it should be infrared and collinear (IRC) safe. In the words of Ref. [1]:

For an observable’s distribution to be calculable in [fixed-order] perturbation theory, the observable should be infra-red safe, i.e. insensitive to the emission of soft or collinear gluons. In particular if  $\vec{p}_i$  is any momentum occurring in its definition, it must be invariant under the branching

$$\vec{p}_i \rightarrow \vec{p}_j + \vec{p}_k$$

whenever  $\vec{p}_j$  and  $\vec{p}_k$  are parallel [collinear] or one of them is small [infrared].

For example, the multiplicity of gluons is not IRC safe, because it is modified by soft and collinear splitting. The energy of the hardest particle in an event is not IRC safe, because it is modified by collinear splitting. However, the total energy flowing into a given cone is IRC safe, because soft emissions don't modify the energy flow, and collinear emissions don't modify its direction.

This last example comes from Sterman and Weinberg [27], who defined an  $e^+e^-$  event as having 2 'jets' if at least a fraction  $(1 - \epsilon)$  of the event's energy is contained in two cones of half-angle  $\delta$ . If we take  $\delta$  to be  $30^\circ$  and  $\epsilon = 0.1$ , then Fig. 5(d) is an example of such a 2-jet event. We can adapt our expression for the total cross section, Eq. (22), to give us the 2-jet cross section as follows

$$\sigma_{2\text{-jet}} = \sigma_{q\bar{q}} \left( 1 + \frac{2\alpha_s C_F}{\pi} \int \frac{dE}{E} \int \frac{d\theta}{\sin\theta} \left[ R\left(\frac{E}{Q}, \theta\right) \left( 1 - \Theta\left(\frac{E}{Q} - \epsilon\right) \Theta(\theta - \delta) \right) - V\left(\frac{E}{Q}, \theta\right) \right] \right). \quad (28)$$

For small  $E$  or  $\theta$  this is just like the total cross section, with full cancellation of divergences between real and virtual terms [cf. Eq. (21)]. For large  $E$  and large  $\theta$  a *finite* piece of real-emission cross section is cut out by the factor  $(1 - \Theta(\frac{E}{Q} - \epsilon)\Theta(\theta - \delta))$  and it corresponds to scales with  $E \sim Q$  and large angles, for which perturbation theory is valid. This then gives

$$\sigma_{2\text{-jet}} = \sigma_{q\bar{q}}(1 - c_1\alpha_s + c_2\alpha_s^2 + \dots), \quad (29)$$

where  $c_1, c_2$ , etc. are all of order 1 (as long as  $\epsilon$  and  $\delta$  were not taken too small). Similarly one could define a 3-jet cross section by requiring that it not be a 2-jet event and that all but a fraction  $\epsilon$  of the energy be contained in 3 cones of half angle  $\delta$ . This would give a cross section of the form

$$\sigma_{3\text{-jet}} = \sigma_{q\bar{q}}(c'_1\alpha_s + c'_2\alpha_s^2 + \dots), \quad (30)$$

where, again, the coefficients are all  $\mathcal{O}(1)$ . So whereas the cross section for getting an extra gluon is divergent, the cross section for an extra jet is finite and small,  $\mathcal{O}(\alpha_s)$ . One difficulty with the extension of the Sterman–Weinberg definition to multiple jets is to know how to place the cones. Since jet-finding is a well-developed subject in its own right, we will return to in detail in Section 5.

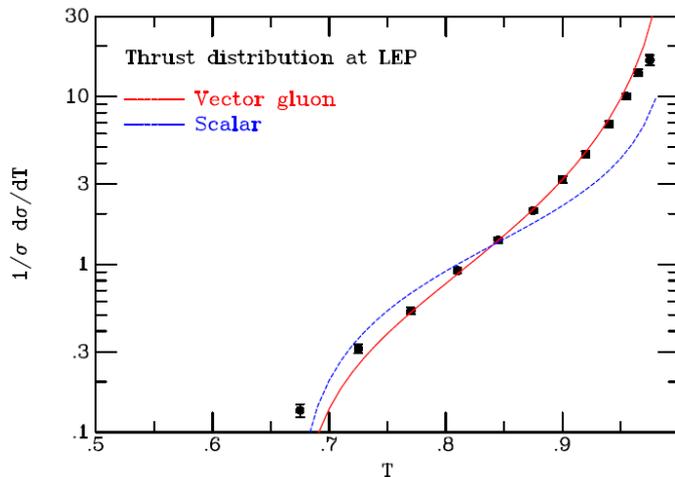
The Sterman–Weinberg jet cross section gives a discrete classification of events: an event either has two jets, or more. An alternative class of infrared and collinear safe observables is that of event shapes, which give a continuous classification of events. The most widely studied example is the thrust,  $T$ ,

$$T = \max_{\vec{n}_T} \frac{\sum_i |\vec{p}_i \cdot \vec{n}_T|}{\sum_i |\vec{p}_i|}, \quad (31)$$

where the sum runs over all particles, and one chooses the thrust axis  $\vec{n}_T$  (a 3-dimensional unit vector) so as to maximize the projection in the numerator. For a perfectly collimated 2-jet event, the thrust axis aligns with the jet axes and the thrust is 1. For a 'Mercedes' type event with three identical collimated jets, the thrust axis will be aligned with any one of the three jets and the thrust will be  $2/3$ . Intermediate events will have intermediate values of the thrust.

One application of the thrust variable is given in Fig. 7. It shows data for the thrust distribution from LEP, compared to  $\mathcal{O}(\alpha_s)$  calculations of the thrust distribution in QCD and in a variant of QCD in which the gluon is a scalar particle rather than a vector particle. The scalar gluon case does not have a divergence for soft emission (the collinear divergence is still there), with the result is that the distribution diverges less strongly in the 2-jet limit than for vector gluons. The data clearly prefer the vector-gluon case, though they do also show the need for higher-order corrections at thrust values close to  $2/3$  and 1.

More generally, event shapes like the thrust have seen broad use in measurements of the QCD coupling, tuning of Monte Carlo event generators (see Section 4.2), studies of the hadronization process, and also as an event-topology discriminant in searches for decays of particles beyond the Standard Model.



**Fig. 7:** Measured thrust distribution at LEP compared to leading order predictions based on QCD (vector gluon, solid red line) and a modified version of QCD in which the gluon is a scalar (spin 0, dashed blue line) rather than a vector (spin-1) particle. Figure taken from CERN academic training lectures by B. R. Webber.

## 2.4 Summary

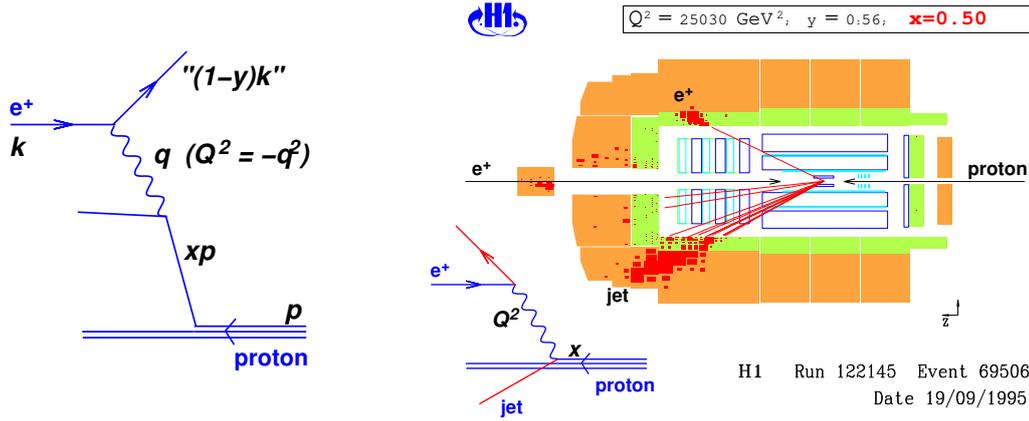
The  $e^+e^- \rightarrow$  hadrons process has allowed us to examine many of the basic ideas of perturbative QCD: soft and collinear divergences, the question of which observables are perturbatively calculable or not (related to infrared and collinear safety) and even what happens if one takes perturbation theory seriously outside its strict domain of applicability (one acquires a rough understanding of the collimated, high-multiplicity structure of real events).

## 3 Parton distribution functions

Having considered processes that involve hadrons in the final state, let us now examine what happens when they are present in the initial state. The importance of understanding initial-state hadrons is obvious at the LHC. Within the ‘parton model’, we write, for example, the hadron collider cross section to produce a  $Z$  and a Higgs boson as

$$\sigma = \int dx_1 f_{q/p}(x_1) \int dx_2 f_{\bar{q}/p}(x_2) \hat{\sigma}_{q\bar{q} \rightarrow ZH}(x_1 x_2 s),$$
(32)

where  $s = (p_1 + p_2)^2$  is the squared  $pp$  centre-of-mass energy,  $f_{q/p}(x_1)$  is the number density of quarks of type  $q$  carrying a fraction  $x_1$  of the momentum of the first proton, and similarly with  $f_{\bar{q}/p}(x_2)$  for the other proton. The  $f_{q/p}(x)$  functions are known as ‘parton distribution functions’ (PDFs). They multiply the ‘hard’ (here, electroweak) cross section,  $\hat{\sigma}_{q\bar{q} \rightarrow ZH}(x_1 x_2 s)$  for the process  $q\bar{q} \rightarrow ZH$ , a function of the squared partonic ( $q\bar{q}$ ) centre-of-mass energy,  $\hat{s} = x_1 x_2 s$ . After integrating over  $x_1$  and  $x_2$  (and summing over quark species), one obtains the total cross section for  $pp \rightarrow ZH$ . The above form seems quite intuitive, but still leaves a number of questions: for example, how do we determine the momentum distributions of quarks and gluons inside the proton? How does the ‘factorization’ into PDFs and a hard part stand up to QCD corrections? Understanding these questions is crucial if, one day, we are to take measured cross sections for  $ZH$  and interpret them, for example, in terms of the coupling of the Higgs to the  $Z$ . And they’re just as crucial for practically any other physics analysis we might want to carry out at the LHC.



**Fig. 8:** Left: kinematic variables of DIS; right: illustration of an event as it appeared in practice in the H1 detector at HERA.

The parton distribution functions are properties of the (non-perturbative) proton. A natural first question is whether we can calculate the PDFs based on lattice QCD. In principle, yes, and there is ongoing work in this direction (see, e.g., Ref. [28]), however, currently lattice QCD has not reached an accuracy in these calculations that is competitive with the combination of experimental measurements and perturbative QCD analyses that we discuss below.

### 3.1 Deep Inelastic Scattering

The process where we have learnt the most about PDFs is Deep Inelastic Scattering (DIS), i.e., lepton–proton scattering in which the photon that is exchanged between lepton and proton has a large virtuality. The kinematics of the “quark-parton-model” DIS process is represented in Fig. 8 (left) and an event from the H1 detector at HERA is shown on the right. Kinematic variables that are usually defined are

$$Q^2 = -q^2, \quad x = \frac{Q^2}{2p \cdot q}, \quad y = \frac{p \cdot q}{p \cdot k}, \quad (33)$$

where  $Q^2$  is the photon virtuality,  $x$  is the longitudinal momentum fraction of the struck quark in the proton, and  $y$  is the momentum fraction lost by the electron (in the proton rest frame).

To zeroth order in  $\alpha_s$  (the ‘quark parton model’), the DIS cross section can be written as

$$\frac{d^2\sigma^{em}}{dx dQ^2} = \frac{4\pi\alpha^2}{xQ^4} \left( \frac{1 + (1-y)^2}{2} F_2^{em} + \mathcal{O}(\alpha_s) \right), \quad (34)$$

written in terms of the  $F_2$  structure function, which, to zeroth order in  $\alpha_s$  is given by

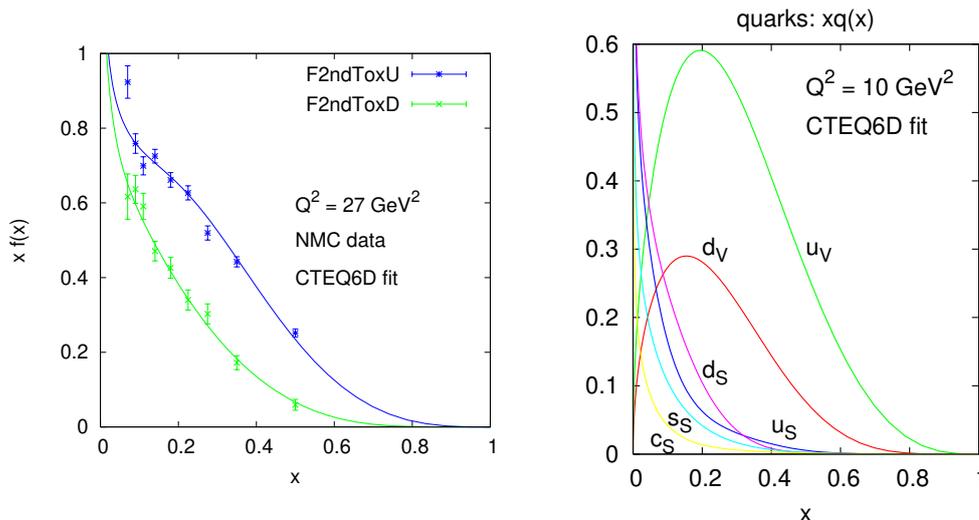
$$F_2^{em} = x \sum_{i=q,\bar{q}} e_i^2 f_{i/p}(x) + \mathcal{O}(\alpha_s), \quad (35)$$

(the “ $em$ ” superscript is a reminder that we’re only considering the electromagnetic part here).

Given the sum over all flavours in Eq. (35), disentangling the information about individual flavours might seem like quite a daunting task. We can attempt to see where the information comes from by starting off with the assumption that the proton consists just of up and down quarks, in which case

$$F_2^{\text{proton}} = x(e_u^2 u_p(x) + e_d^2 d_p(x)) = x \left( \frac{4}{9} u_p(x) + \frac{1}{9} d_p(x) \right), \quad (36)$$

where we have introduced the shorthand  $f_{u/p}(x) = u_p(x)$ , etc. (later we will drop the “ $p$ ” subscript altogether). In Eq. (36) we now have a linear combination of just two functions. The next step is to use



**Fig. 9:** Left: linear combinations of NMC  $F_2$  data [29] for protons and deuterons so as to obtain  $xu(x)$  and  $xd(x)$ , assuming only  $u$  and  $d$  quarks inside the proton, together with the expectations from the same linear combination based on the CTEQ6D PDF parametrizations [30]. Right: results for different valence and sea quark distributions from CTEQ6D at  $Q^2 = 10 \text{ GeV}^2$ .

isospin symmetry, the fact that the neutron is essentially just a proton with  $u \leftrightarrow d$ , i.e.,  $u_n(x) \simeq d_p(x)$  (ignoring small electromagnetic effects), so that

$$\frac{1}{x} F_2^{\text{neutron}} = \frac{4}{9} u_n(x) + \frac{1}{9} u_n(x) \simeq \frac{4}{9} d_p(x) + \frac{1}{9} u_p(x). \quad (37)$$

Appropriate linear combination of  $F_2^{\text{proton}}$  and  $F_2^{\text{neutron}}$  (in practice one uses deuterons, or nuclei as a source of neutrons) therefore provides separate information on  $u_p(x)$  and  $d_p(x)$ .

The results of this exercise are shown in Fig. 9 (left). As expected we see more up quarks than down quarks. But there's also a problem: let's try to extract the total number of up quarks,  $U = \int_0^1 dx u(x)$ . We see that the data increase towards small  $x$ , and a parametrization (CTEQ6D [30]) of these and other data even seems to diverge as  $x \rightarrow 0$ ,  $xu(x) \sim xd(x) \sim x^{-0.25}$ .<sup>4</sup> Given that the plot is supposed to be for  $xu(x)$  and that we need to integrate  $u(x)$  to get the total number of up quarks in the proton, it looks like we'll obtain an infinite number of up quarks, which is hardly consistent with expectations from the picture of a proton as being a  $uud$  state.

One thing we've 'neglected' is that there can also be anti-up and anti-down quarks in the proton, because the proton wavefunction can fluctuate, creating  $u\bar{u}$  and  $d\bar{d}$  pairs, 'sea quarks', and so give rise to  $\bar{u}(x)$  and  $\bar{d}(x)$  distributions. Therefore instead of Eq. (36), we should have written

$$F_2^{\text{proton}} = \frac{4}{9} (xu_p(x) + x\bar{u}_p(x)) + \frac{1}{9} (d_p(x) + \bar{d}_p(x)), \quad (38)$$

since quarks and antiquarks have identical squared charges. So what we called " $xu(x)$ " in Fig. 9 (left) was actually  $xu(x) + x\bar{u}(x)$  (with some admixture of strange and charm quarks too). The infinite number of quarks and antiquarks can then just be interpreted as saying that fluctuations within the proton create infinite numbers of  $q\bar{q}$  pairs, mostly carrying a small fraction of the proton's momentum.

Returning to the statement that the proton has 2 up and 1 down quark, what we mean is that the

<sup>4</sup>The particular parametrization shown here is not very widely used nowadays, but is particularly convenient for comparisons to DIS data (it corresponds to a definition of PDFs known as the "DIS scheme," for which higher-order corrections to many DIS cross sections are particularly simple). The behaviour at small  $x$  is common to essentially all parametrizations.

net number of up minus anti-up quarks is 2,

$$\int_0^1 dx(u(x) - \bar{u}(x)) = 2, \quad \int_0^1 dx(d(x) - \bar{d}(x)) = 1, \quad (39)$$

where  $u(x) - \bar{u}(x)$  is also called the valence quark distribution  $u_V(x)$ . How can we measure the difference between quarks and antiquarks? The answer is through charged-current processes (e.g., neutrino scattering), since a  $W^+$  interacts only with  $d$  and  $\bar{u}$ , but not with  $\bar{d}$  or  $u$ .

We could imagine forming linear combinations of data sets with proton and nuclear targets, with photon and  $W^\pm$  exchange, etc., in order to obtain the different quark-flavour PDFs. In practice it is simpler to introduce parametrizations for each flavour, deduce predictions for DIS, neutrino-scattering and other cross sections, and then fit the parameters so as to obtain agreement with the experimental data. This is known as a ‘global fit’. We will return to such fits below, but for now let’s just look at the results for different quark distributions, shown in Fig. 9 (right).

We see that the valence quark distributions are mainly situated at moderate  $x$  values. For  $x \rightarrow 1$  they fall off as a moderate power of  $(1-x)$ , roughly as  $(1-x)^3$ . This fall-off comes about because to have one  $u$  quark carrying most of the proton momentum implies that the other  $u$  and the  $d$  quark must both carry only a small fraction, and the phase space for that to occur vanishes as  $x \rightarrow 1$  (the proper treatment for this is through ‘quark-counting rules’). For  $x \rightarrow 0$ ,  $xq_V(x) \sim x^{0.5}$ , which is related to something called Regge physics.

The sea-quark distributions ( $u_S(x) \equiv u(x) + \bar{u}(x) - u_V(x) \equiv 2\bar{u}(x)$ , etc.) are concentrated at small  $x$  values. They fall off much more steeply at large  $x$ : for a  $\bar{u}$  quark to have a large momentum fraction, 3  $u$ -quarks and one  $d$ -quark must have correspondingly small- $x$ , so the phase space is even more restricted than in the case of valence quarks. For  $x \rightarrow 0$ , the exact behaviour is not entirely simple, but as a rough rule of thumb  $xq_S(x) \sim x^{-\omega}$  with  $\omega \sim 0.2-0.4$ .

Of course we’re still missing the PDFs for one kind of parton: gluons. It’s a well known fact that if we evaluate the fraction of the proton’s momentum carried by all the quarks,

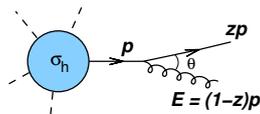
$$\sum_q \int_0^1 dx xq(x), \quad (40)$$

the result is about 0.5. It’s fair to suspect that the gluon is responsible for the other half, but how are we to establish its shape given that it’s not directly probed by photons or  $W^\pm$ ? To answer that question we must go beyond the ‘naive’ leading-order partonic picture of the proton’s quarks interacting with probes, and bring in QCD splitting.

## 3.2 Initial-state parton splitting, DGLAP evolution

### 3.2.1 Final and initial-state divergences

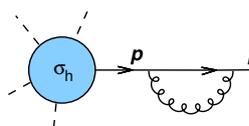
In Eq. (26a) we wrote the universal form for the final-state ‘splitting’ of a quark into a quark and a soft gluon. Let’s rewrite it with different kinematic variables, considering a hard process  $h$  with cross section  $\sigma_h$ , and examining the cross section for  $h$  with an extra gluon in the final state,  $\sigma_{h+g}$ . We have



$$\sigma_{h+g} \simeq \sigma_h \frac{\alpha_s C_F}{\pi} \frac{dz}{1-z} \frac{dk_t^2}{k_t^2}, \quad (41)$$

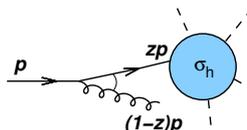
where  $E$  in Eq. (26a) corresponds to  $E = (1-z)p$  and we’ve introduced  $k_t = E \sin \theta \simeq E\theta$ . If we avoid distinguishing a collinear  $q+g$  pair from a plain quark (measurements with IRC safe observables) then, as we argued before, the divergent part of the gluon emission contribution always cancels with a

related virtual correction



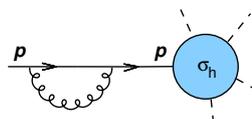
$$\sigma_{h+V} \simeq -\sigma_h \frac{\alpha_s C_F}{\pi} \frac{dz}{1-z} \frac{dk_t^2}{k_t^2}. \quad (42)$$

Now let us examine what happens for initial-state splitting, where the hard process occurs *after* the splitting and the momentum entering the hard process is modified  $p \rightarrow zp$ :



$$\sigma_{g+h}(p) \simeq \sigma_h(zp) \frac{\alpha_s C_F}{\pi} \frac{dz}{1-z} \frac{dk_t^2}{k_t^2}, \quad (43)$$

where we have made explicit the hard process's dependence on the incoming momentum, and we assume that  $\sigma_h$  involves momentum transfers  $\sim Q \gg k_t$ , so that we can ignore the extra transverse momentum entering  $\sigma_h$ . For virtual terms, the momentum entering the process is unchanged, so we have



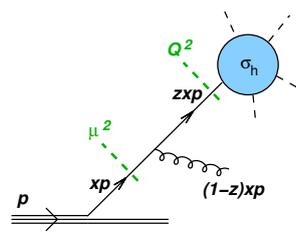
$$\sigma_{g+h}(p) \simeq -\sigma_h(p) \frac{\alpha_s C_F}{\pi} \frac{dz}{1-z} \frac{dk_t^2}{k_t^2}, \quad (44)$$

The total cross section then gets contributions with two different hard cross sections:

$$\sigma_{g+h} + \sigma_{V+h} \simeq \frac{\alpha_s C_F}{\pi} \underbrace{\int_0^{Q^2} \frac{dk_t^2}{k_t^2}}_{\text{infinite}} \underbrace{\int_0^1 \frac{dz}{1-z} [\sigma_h(zp) - \sigma_h(p)]}_{\text{finite}}. \quad (45)$$

Note the limits on the integrals, in particular the  $Q^2$  upper limit on the transverse-momentum integration: the approximations we're using are valid as long as the transverse momentum emitted in the initial state is much smaller than the momentum transfers  $Q$  that are present in the hard process. Of the two integrations in Eq. (45), the one over  $z$  is finite, because in the region of the soft divergence,  $z \rightarrow 1$ , the difference of hard cross sections,  $[\sigma_h(zp) - \sigma_h(p)]$ , tends to zero. In contrast, the  $k_t$  integral diverges in the collinear limit: the cross section with an incoming parton (and virtual corrections) appears not to be collinear safe. This is a general feature of processes with incoming partons: so how are we then to carry out calculations with initial-state hadrons?

In Section 2.3.1, when trying to make sense of final-state divergences, we introduced a (non-perturbative) cutoff. Let's do something similar here, with a cutoff,  $\mu_F$ , called a factorization scale (which will usually be taken at perturbative scales). The main idea in using this cutoff is that any emissions that occur with  $k_t \lesssim \mu_F$  are absorbed ('factorized') into the PDF itself. Thus the PDFs become a function of  $\mu_F$ . We can write the lowest order term for the cross section,  $\sigma_0$ , and the correction with one initial-state emission or virtual loop,  $\sigma_1$ , as follows



$$\sigma_0 = \int dx \sigma_h(xp) q(x, \mu_F^2), \quad (46a)$$

$$\sigma_1 \simeq \frac{\alpha_s C_F}{\pi} \underbrace{\int_{\mu_F^2}^{Q^2} \frac{dk_t^2}{k_t^2}}_{\text{finite (large?)}} \underbrace{\int \frac{dx dz}{1-z} [\sigma_h(zxp) - \sigma_h(xp)]}_{\text{finite}} q(x, \mu_F^2), \quad (46b)$$

where we have now included also the integral over the longitudinal momentum fraction  $x$  of the parton extracted from the proton. The emissions and virtual corrections with  $k_t \lesssim \mu_F$  are now implicitly included inside the  $q(x, \mu_F^2)$  PDF factor that appears in the  $\sigma_0$  contribution, and only those with  $k_t \gtrsim \mu_F$

arise explicitly in the  $\mathcal{O}(\alpha_s)$  term. This term (whose real-emission part is represented in the diagram to the left) is now finite, albeit potentially large if  $\mu_F \ll Q$ .

This situation of having a non-integrable divergence that somehow needs to be regularized and absorbed with a scale choice into some ‘constant’ of the theory (here the PDFs), is reminiscent of renormalization for the coupling constant. The differences are that here we are faced with a divergence in the (collinear) infrared rather than one in the ultraviolet. And that unlike the coupling, the PDFs are not fundamental parameters of the theory, but rather quantities that we could calculate if only we had sufficiently sophisticated theoretical ‘technology’. Nevertheless, as for the coupling, the freedom in choosing the scale that enters the regularization, called the factorization scale, implies the presence of a renormalization group equation for the PDFs, the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equation.

### 3.2.2 The DGLAP equation

To see what form the DGLAP equation takes, let us fix the longitudinal momentum of the quark entering the hard process to be  $xp$  (whereas above we’d fixed the momentum for the quark extracted from the proton). Next we examine the effect on the PDFs of integrating over a small region of transverse momentum  $\mu_F^2 < k_t^2 < (1 + \epsilon)\mu_F^2$ ,

$$\begin{aligned} \frac{dq(x, \mu_F^2)}{d \ln \mu_F^2} &= \frac{1}{\epsilon} \left( \text{Diagram 1} + \text{Diagram 2} \right) \\ &= \frac{\alpha_s}{2\pi} \int_x^1 dz p_{qq}(z) \frac{q(x/z, \mu_F^2)}{z} - \frac{\alpha_s}{2\pi} \int_0^1 dz p_{qq}(z) q(x, \mu_F^2) \end{aligned} \quad (47)$$

where  $p_{qq}(z)$  is the real part of the ‘splitting kernel’ for a quark splitting to a quark plus a gluon,

$$p_{qq}(z) = C_F \frac{1+z^2}{1-z}. \quad (48)$$

Until now, we had concentrated on the soft limit, which was equivalent to approximating  $p_{qq}(z) \simeq \frac{2C_F}{1-z}$ . What Eq. (47) tells us is that as we increase the factorization scale, we get extra partons with longitudinal momentum fraction  $x$  that come from the branching of partons in the proton at lower factorization scales but larger momentum fractions  $x/z$  ( $x < z < 1$ ). There are also loop contributions (second term on the RHS) to the parton density at a fixed  $x$  value, which are negative contributions to the evolution. The way to think about these is that when a parton with momentum fraction  $x$  branches to partons with lower momentum fractions, the original parton is lost and the loop diagram accounts for that.

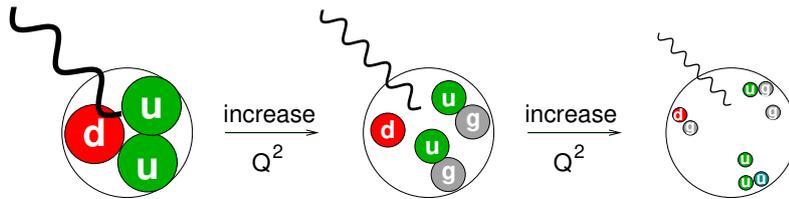
It’s a bit awkward to write the real and virtual parts separately in Eq. (47), especially if one wants to explicitly see the cancellation of the divergences for  $z \rightarrow 1$ . It’s therefore standard to use the more compact notation

$$\frac{dq(x, \mu_F^2)}{d \ln \mu_F^2} = \frac{\alpha_s}{2\pi} \underbrace{\int_x^1 dz P_{qq}(z) \frac{q(x/z, \mu_F^2)}{z}}_{P_{qq} \otimes q}, \quad P_{qq} = C_F \left( \frac{1+z^2}{1-z} \right)_+, \quad (49)$$

where the subscript plus, known as the ‘plus’ prescription, is to be understood as follows:

$$\int_x^1 dz [g(z)]_+ f(z) = \int_x^1 dz g(z) f(z) - \int_0^1 dz g(z) f(1) \quad (50)$$

$$= \int_x^1 dz g(z) (f(z) - f(1)) - \int_0^x dz g(z) f(1) \quad (51)$$



**Fig. 10:** An illustration how with ever shorter wavelength photon probes, one resolves more and more structure inside the proton.

so that the factor  $(f(z) - f(1))$ , which goes to zero at  $z = 1$ , kills the divergence due the singular behaviour of  $g(z)$  for  $z \rightarrow 1$ .

Equation (49) involves just quarks, but the proton contains both quarks and gluons, so the full DGLAP equations are actually coupled evolution equations. Schematically, for just a single quark flavour, they read

$$\frac{d}{d \ln \mu_F^2} \begin{pmatrix} q \\ g \end{pmatrix} = \frac{\alpha_s(\mu_F^2)}{2\pi} \begin{pmatrix} P_{q \leftarrow q} & P_{q \leftarrow g} \\ P_{g \leftarrow q} & P_{g \leftarrow g} \end{pmatrix} \otimes \begin{pmatrix} q \\ g \end{pmatrix} \quad (52)$$

and more generally they span all quark flavours and anti-flavours. In labelling the different flavour entries, we've included arrows (usually not shown), e.g.  $q \leftarrow g$ , so as to emphasize that we have evolution from the right-hand parton type to the left-hand parton type. The splitting functions other than  $P_{qq}$  are given by

$$P_{qg}(z) = T_R [z^2 + (1-z)^2], \quad P_{gq}(z) = C_F \left[ \frac{1 + (1-z)^2}{z} \right], \quad (53a)$$

$$P_{gg}(z) = 2C_A \left[ \frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z) \right] + \delta(1-z) \frac{(11C_A - 4n_f T_R)}{6}. \quad (53b)$$

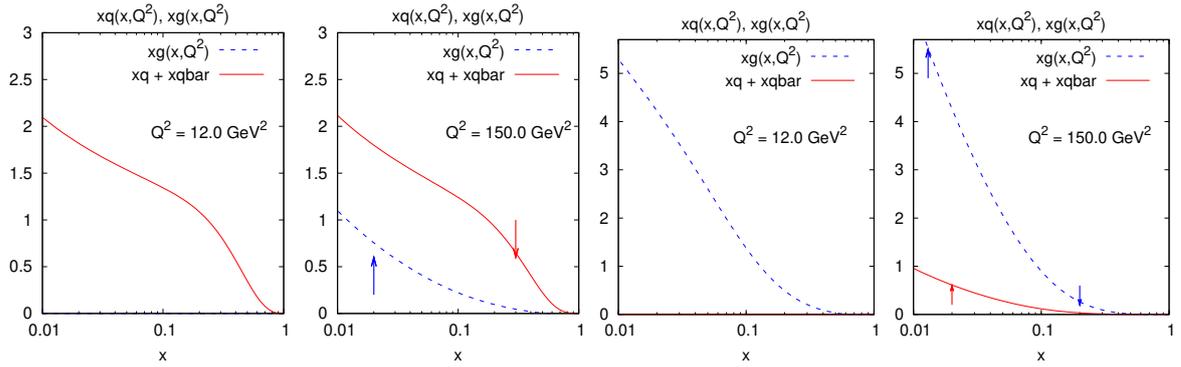
Additionally,  $P_{\bar{q}g} = P_{qg}$  and, to this first order in the coupling,  $P_{qq'}$  and  $P_{q\bar{q}}$  are both zero.

Several features of the splitting functions are worth noting:  $P_{qg}$  and  $P_{gg}$  are both symmetric in  $z \leftrightarrow 1-z$  (except for the virtual part).  $P_{qq}$  and  $P_{gg}$  diverge for  $z \rightarrow 1$ , which corresponds to soft-gluon emission. And  $P_{gg}$  and  $P_{gq}$  both diverge for  $z \rightarrow 0$  (corresponding to a soft gluon entering the 'hard' process). This last point implies that PDFs  $q(x)$  and  $g(x)$  must grow at least as fast as  $1/x$  for  $x \rightarrow 0$ : even if such a divergence is absent in the non-perturbative 'initial conditions' for the quark and gluon distributions at low scales  $\mu_F$ , DGLAP evolution inevitably introduces it into the result for  $q(x, \mu_F^2)$  and  $g(x, \mu_F^2)$  at higher scales.

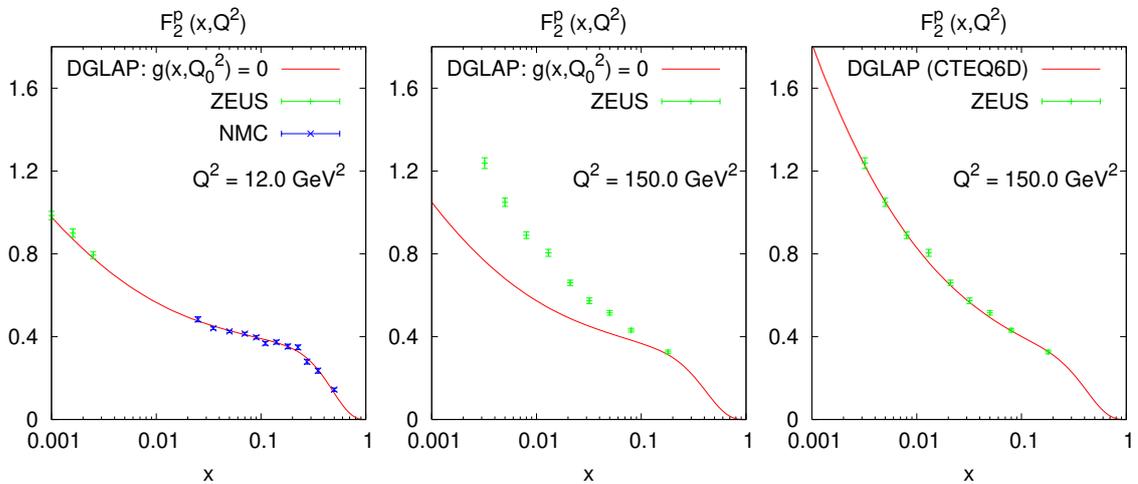
### 3.2.3 Results of DGLAP evolution

Pictorially, the effect of DGLAP evolution is illustrated in Fig. 10. A more quantitative view is given in Fig. 11, which shows the effect of DGLAP evolution with an initial condition that is pure quark (two left plots) or pure gluon (two right plots). In both cases one sees that evolution generates some amount of the missing parton type; one also sees how it depletes the parton distributions at large  $x$ , and increases them at small  $x$  (especially in the case of the gluon). The attentive reader may have observed that the figure labels the scale of the PDFs as  $Q^2$  rather than  $\mu_F^2$ : this is because it is standard to take  $\mu_F^2 = Q^2$  so as to minimize the size of the  $\mathcal{O}(\alpha_s)$  term of Eq. (46) which arises, roughly, from the integral over transverse momenta from  $\mu_F^2$  to  $Q^2$ . I.e., one usually chooses to factorize essentially *all* initial-state radiation into the PDFs and so into the LO cross section.

Since, as we've seen in Fig. 11, the presence of a gluon distribution helps drive quark evolution, we can use the experimentally observed pattern of quark evolution to help constrain the gluon. The left-hand plot of Fig. 12 shows data from ZEUS [31] and NMC [29] on  $F_2(x, Q^2)$  at some low but still perturbative



**Fig. 11:** An illustration of the impact of DGLAP evolution. From left to right: (a) initial condition consisting just of quarks and anti-quarks at  $\mu_F^2 \equiv Q^2 = 12 \text{ GeV}^2$ ; (b) the result of evolution to  $Q^2 = 150 \text{ GeV}^2$ ; (c) a purely gluonic initial condition at  $Q^2 = 12 \text{ GeV}^2$ ; and (d) the result of its evolution to  $Q^2 = 150 \text{ GeV}^2$ .

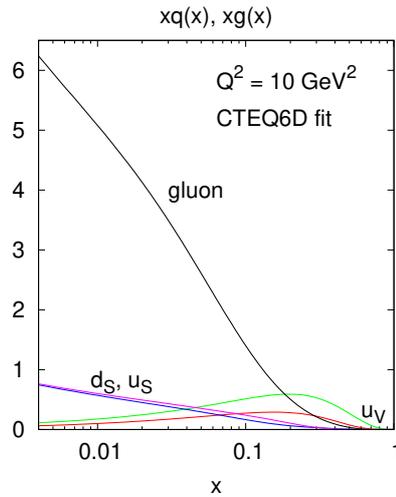


**Fig. 12:** ZEUS and NMC data together with an initial condition that gives a good fit at low  $Q^2$  (left-most plot). If one evolves that initial condition assuming a gluon distribution that is zero at low  $Q^2$ , then agreement with high-scale data is poor (central plot); whereas with a significant low-scale gluon component (taken from the CTEQ6D parametrization), agreement becomes good at high scales (right-most plot).

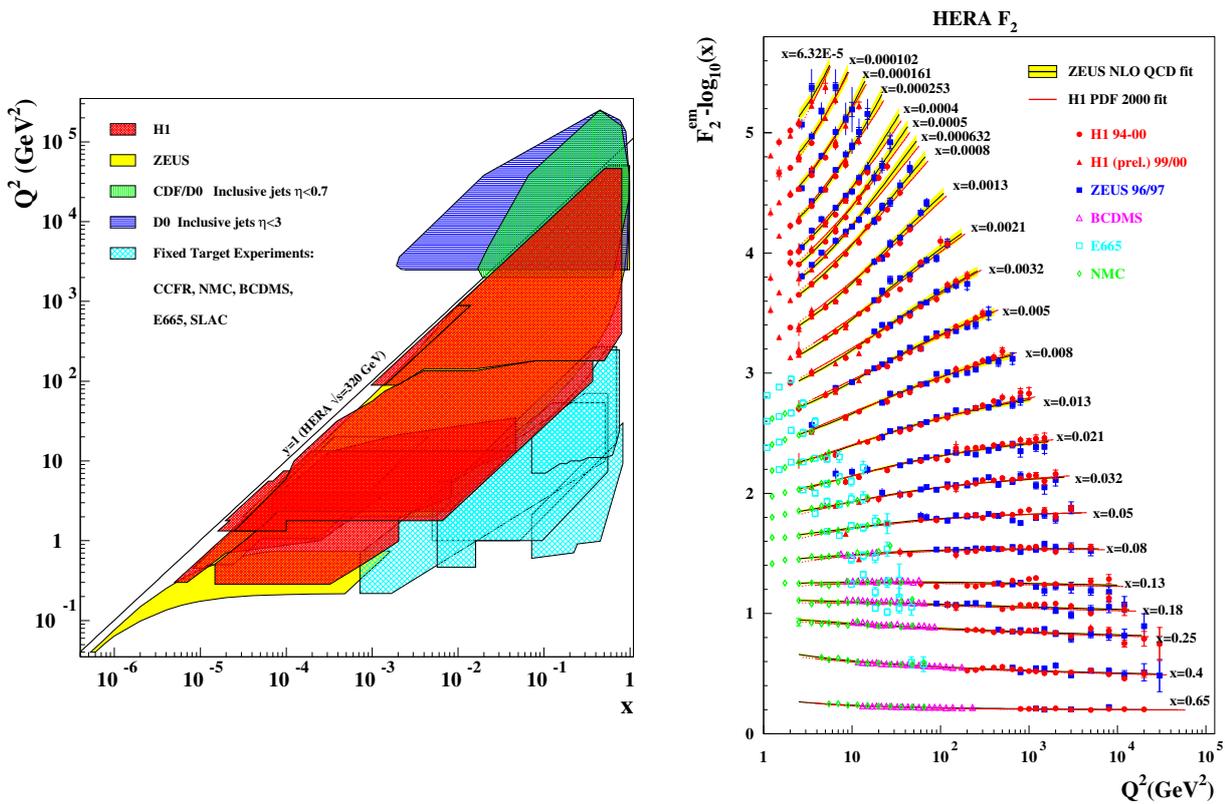
scale  $Q^2 = Q_0^2 \equiv 12 \text{ GeV}^2$ . The data are compared to the expectations based on the CTEQ6D PDFs' quark content at that scale, illustrating the good agreement. Since these are data for  $F_2$ , they have no direct sensitivity to the gluon distribution. The middle plot shows data for  $150 \text{ GeV}^2$ , together with the results of DGLAP evolution from  $Q_0^2 = 12 \text{ GeV}^2$ , assuming that the gluon distribution was zero at  $Q_0^2$ . There's a clear discrepancy. In the right-hand plot, the comparison is made with evolution whose initial condition at  $Q_0^2$  contained a quite large gluon component (exactly that in the CTEQ6D distributions), causing the quark distribution at small  $x$  values to increase faster with  $Q^2$  than would otherwise be the case, bringing the evolution into agreement with the data.

### 3.3 Global fits

It's interesting to ask just how much of a gluon distribution is needed in order to get the agreement shown in Fig. 12. The answer is given in Fig. 13 and one sees that the gluon distribution is *enormous*, especially at small values of  $x$ . It is fair to ask whether we can trust a result such as Fig. 13, so in this section we will examine some of ingredients and issues that are relevant to the 'global fits' that inform our knowledge of PDFs.



**Fig. 13:** The distributions of different parton species in the CTEQ6D parametrization at a scale  $Q^2 = 10 \text{ GeV}^2$

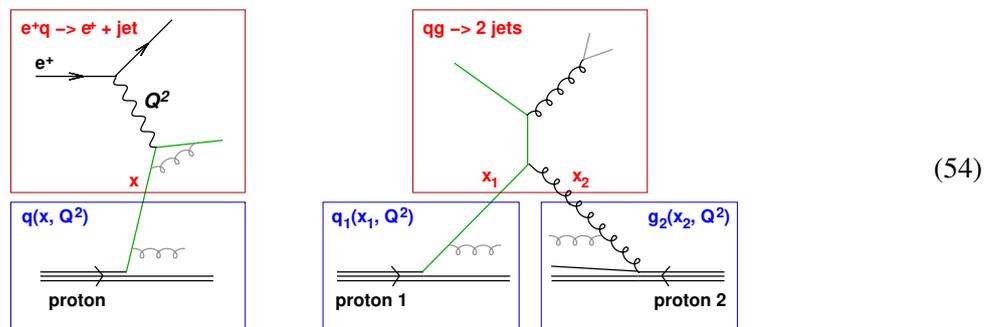


**Fig. 14:** Left: an illustration of the kinematic regions and data sets typically used in PDF fits (based on a fit by ZEUS). Right: experimental results for  $F_2$  as a function of  $Q^2$  for many different  $x$  values, compared to the results of a global fit by the ZEUS collaboration.

Figure 14 (left) illustrates the kinematical regions in the  $x$  and  $Q^2$  plane covered by the experimental data sets typically used in global fits. Everything below the diagonal line corresponds to DIS data, and the right-hand plot shows the comparison between a fit (by ZEUS) and the bulk of the DIS data, illustrating the excellent consistency between fit and data. Agreement with such a broad data set is already a non-trivial achievement. Figure 14 (left) also shows shaded regions that span the diagonal line. These correspond to hadron-collider jets data, which provide valuable direct information on the gluon distribution in global fits, as we will discuss below. The other topics that we will address here relate to the accuracy of our determinations of PDFs.

### 3.3.1 Factorization and $p\bar{p}$ jet production

Perhaps the most convincing cross-check of PDF extractions comes from Tevatron jet data (there are also important jet data from HERA). The process of factorizing initial-state radiation into the PDF at a given scale is equally valid in DIS and  $p\bar{p}$  (or  $pp$ ) collisions, as illustrated in the following pictorial representation of the two cases:



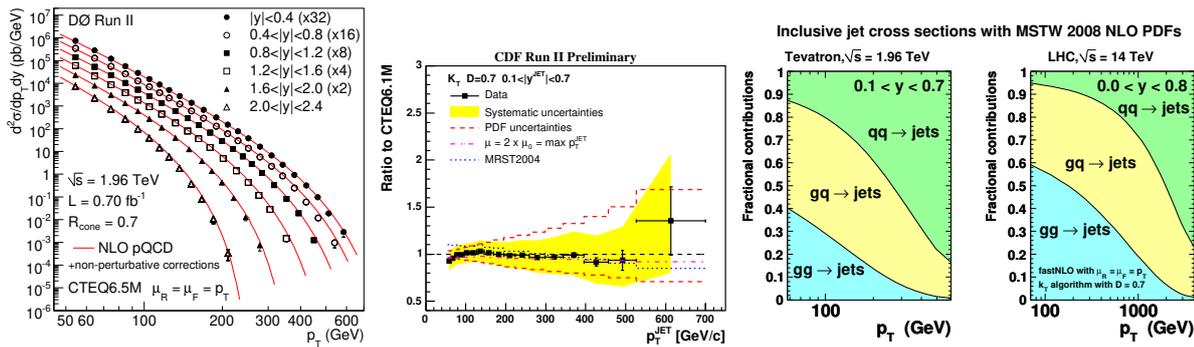
Given factorization and a determination of PDFs in DIS, one can simply take the expression Eq. (32) for a generic cross section in hadron–hadron collisions, and rewrite it with explicit factorization scales:

$$\sigma_{pp \rightarrow ZH} = \int dx_1 f_{q/p}(x_1, \mu_F^2) \int dx_2 f_{\bar{q}/p}(x_2, \mu_F^2) \hat{\sigma}_{q\bar{q} \rightarrow ZH}(x_1 p_1, x_2 p_2, \mu_F^2). \quad (55)$$

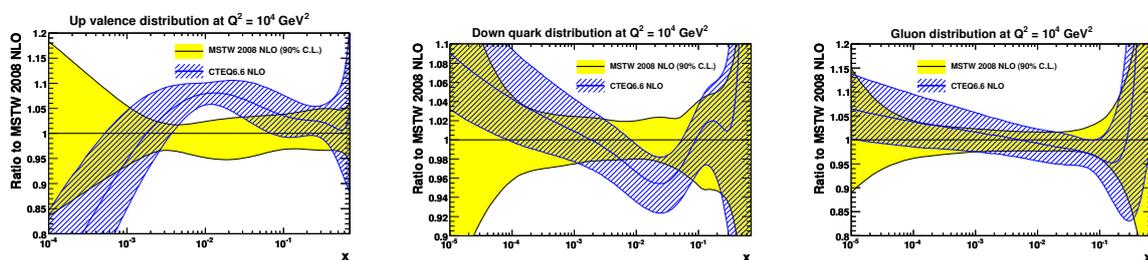
Such a formula, with  $\sigma_{q\bar{q} \rightarrow ZH}$  replaced by  $\sigma_{q\bar{q} \rightarrow q\bar{q}}$  (summing also over processes with gluons, etc.) can be used to obtain predictions for the differential inclusive jet spectrum at the Tevatron. Figure 15 shows comparisons of data from DØ and from CDF with predictions from CTEQ6.5 [32] and MRST2004 [33] PDF parametrizations, illustrating excellent agreement. The two right-hand plots show how different incoming partonic scattering channels contribute to the cross section, highlighting the significant contribution from gluons.

It is perhaps misleading to use the word ‘prediction’ about Fig. 15: most advanced fully global PDF fits actually make use of data such as that in Fig. 15 as part of the fit. Still, it is a powerful consistency check that it is possible to obtain agreement both with the jet data, which is sensitive directly to quark and to gluon distributions, as well as with DIS data, which is directly sensitive to the quarks and indirectly to the gluons, through the scaling violations.

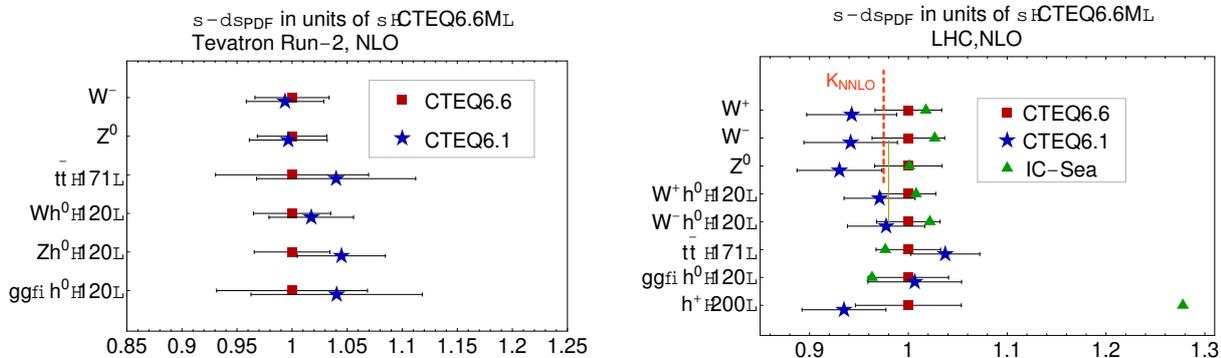
One technical comment is due concerning factorization. While our discussion has been limited to leading order, many of the figures that we have shown also involve higher orders (to which we will return). When using PDFs with predictions beyond LO, it is necessary to specify the ‘factorization scheme’, i.e., the specific procedure by which one separates emissions below and above  $\mu_F$ . The figures in Sections 3.1 and 3.2 made use of the ‘DIS’ scheme (hence CTEQ6D), defined such that  $F_2$  is given by Eq. (35), free of any  $\mathcal{O}(\alpha_s)$  (or higher) corrections. While that scheme has the benefit of pedagogical simplicity, in real calculations (and all plots in this section) it is usually more convenient to use the ‘ $\overline{\text{MS}}$ ’ factorization scheme, based on dimensional regularization.



**Fig. 15:** Left: DØ inclusive jets data [34] compared to predictions with CTEQ6.5M PDFs. Middle: the ratio of CDF data from Ref. [35] to predictions with MRST2004 PDFs. Right: the relative contributions of different scattering channels to the Tevatron and LHC jet spectra, as a function of jet  $p_T$ , taken from Ref. [36].



**Fig. 16:** Uncertainties on recent PDFs from the MSTW [37] and CTEQ groups [38] at a scale of  $Q = 100$  GeV (figure taken from Ref. [37]).



**Fig. 17:** Impact of PDF uncertainties on predictions for standard cross sections at the Tevatron and LHC from Ref. [38].

### 3.3.2 Uncertainties

An important part of the activity on global fits in recent years has been geared to the estimation of the uncertainties on PDFs. Figure 16 shows the uncertainties on recent PDF sets from the two most established PDF fitting groups, MSTW and CTEQ, illustrating uncertainties that are in the couple of per cent to ten per cent range. Figure 17 illustrates the impact of the uncertainties on predictions for a range of cross sections at the Tevatron and the LHC.

Estimating PDF uncertainties is something of an art: for example, one must parametrize the PDFs at some input scale and there is freedom in how flexible a parametrization one uses: too rigid (or with too many theorist's assumptions) and the global fit may not have the flexibility to describe the data or may

appear to have inappropriately small uncertainties in regions where there are no data; with too flexible a parametrization the fits may develop artefacts that adapt to statistical fluctuations of the data. Other issues include reconciling barely compatible data sets and deciding what values of  $\chi^2$  variations are reasonable to estimate errors given the incompatible data sets. In addition to MSTW and CTEQ (and several other groups), a recent entrant to PDF fitting is the NNPDF Collaboration, which is developing procedures that attempt to minimize theoretical bias in questions such as the parametrizations. First global fit results including  $pp$  data have been given in Ref. [39], though for fully accurate treatment of the HERA data, one should await their inclusion of heavy-quark effects. Their results so far tend to be similar to those of MSTW and CTEQ, except in the regions of small- $x$  near the edge of the available HERA phase space and for strange quark distributions, where they find somewhat larger uncertainties.

### 3.3.3 PDFs for LHC and the accuracy of DGLAP evolution

Figure 18 (left) illustrates the kinematic region in the  $x$  and  $Q^2$  plane that is covered by the LHC (with  $\sqrt{s} = 14$  TeV), compared to that for HERA and fixed-target experiments. The LHC region is labelled with a grid corresponding to mass ( $M$ , and one takes  $Q = M$ ) and rapidity ( $y$ ) of the object that is being produced. These are related to the incoming momentum fractions  $x_1$  and  $x_2$  and the  $pp$  squared centre-of-mass energy through

$$M = \sqrt{x_1 x_2 s}, \quad y = \frac{1}{2} \ln \frac{x_1}{x_2}. \quad (56)$$

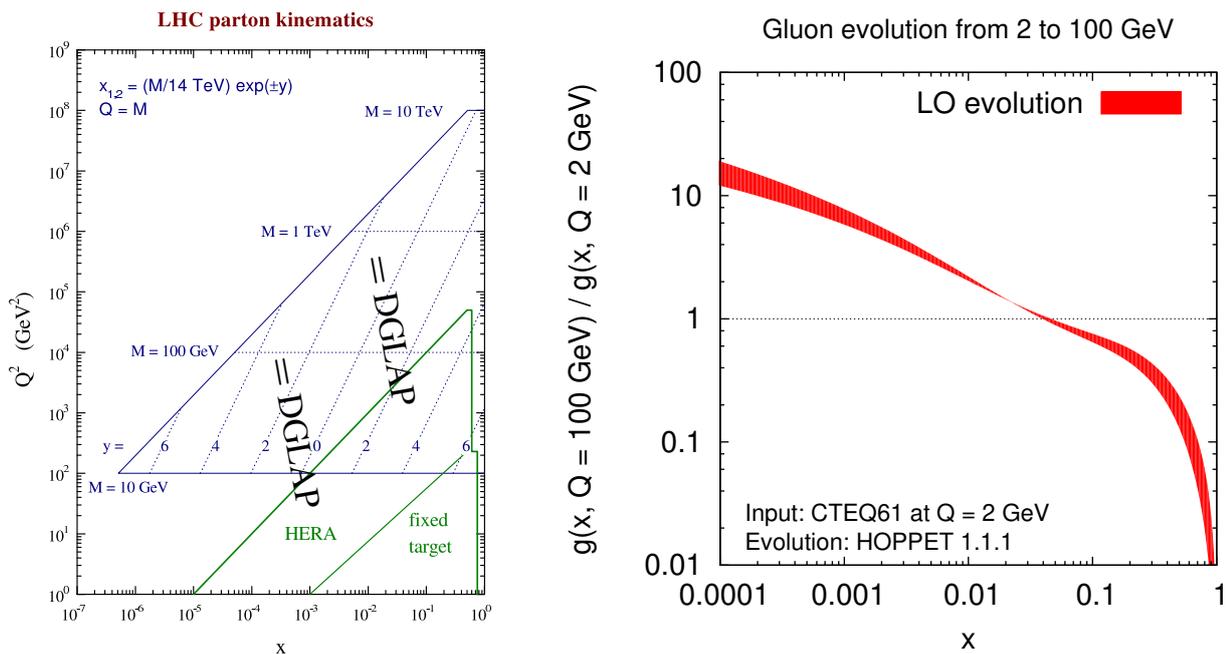
An object produced at a rapidity  $y$  involves  $x_1 = \frac{M}{\sqrt{s}} e^{+y}$  and  $x_2 = \frac{M}{\sqrt{s}} e^{-y}$ . One feature that's immediately visible from the plot is that much of the LHC kinematic plane covers regions where PDFs have not been measured directly. We will therefore rely heavily on DGLAP evolution for our predictions. The right-hand plot in Fig. 18 illustrates just how much, in that it gives the factor by which the gluon distribution evolves in going from  $Q = 2$  GeV to  $Q = 100$  GeV. Depending on the region in  $x$ , this can be a large factor,  $\mathcal{O}(10)$ . When compared to the experimental uncertainties on PDFs as shown in Figs. 16 and 17, we clearly have to ask how well we know the evolution.

In Section 4.1 we will discuss in detail how uncertainties are estimated in theoretical predictions. For now, essentially, there is freedom in Eq. (52) to choose a renormalization scale  $\mu_R^2$  for  $\alpha_s$  that differs from  $\mu_F^2$ . A conventional way of estimating the uncertainties is to choose  $\mu_R^2 = (x_\mu \mu_F)^2$ , varying  $x_\mu$  in the range  $\frac{1}{2} < x_\mu < 2$  (actually one often just takes three values,  $x_\mu = \frac{1}{2}, 1, 2$ ). Starting from a fixed input at 2 GeV and evolving with different choices for  $x_\mu$  gives the width of the band shown in Fig. 18 (right). That width is much larger than the uncertainties that we see in Figs. 16 and 17.

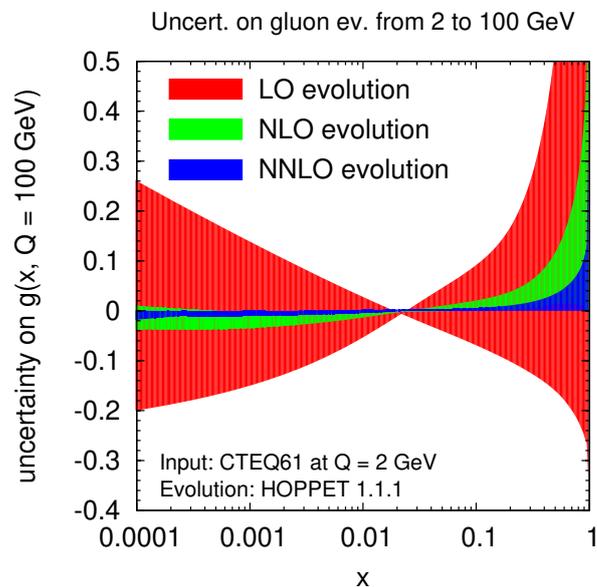
Fortunately we are not limited to leading-order (LO) DGLAP evolution, i.e., just the  $\mathcal{O}(\alpha_s)$  term in Eq. (52). The order  $\alpha_s^2$  (next-to-leading order — NLO) corrections to the DGLAP equation were calculated around 1980 [41, 42] and in 2004 the calculation of the NNLO corrections was completed [43, 44]. To give an idea of the complexity of that calculation, rather than taking a couple of lines as in Eq. (53) at LO, the NNLO results take  $\mathcal{O}(10)$  pages to write down!

The impact of including higher orders on the evolution uncertainties is illustrated in Fig. 19. The same CTEQ61 input distribution is evolved from 2 GeV to 100 GeV with LO, NLO and NNLO splitting kernels, using three scale choices,  $x_\mu = \frac{1}{2}, 1, 2$ . The figure then shows the results for the gluon distribution at scale 100 GeV, normalized to gluon distribution obtained with NNLO evolution and  $x_\mu = 1$ . One sees how the uncertainty comes down substantially as higher orders are included: 30% in some regions at LO, 5% over most of the  $x$  range at NLO and 2% at NNLO. In the NNLO case, the uncertainty is usually smaller than the experimental uncertainties for the PDFs that were shown in Fig. 16.<sup>5</sup>

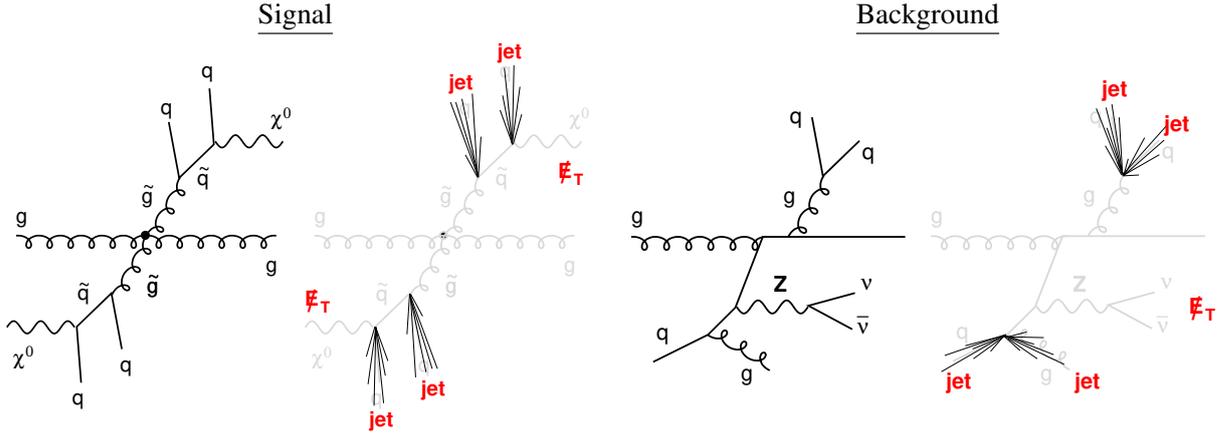
<sup>5</sup>Figure 19 is to be interpreted with some caution. We have taken a fit carried out with  $x_\mu = 1$ , and then evolved it to high scales with  $x_\mu \neq 1$ . However, for example, the gluon distribution is partially determined from the evolution of  $F_2$ , so if the fit itself were carried out with  $x_\mu \neq 1$ , the fit result would probably change, introducing additional dependence of the  $Q = 100$  results on the choice of  $x_\mu$ . This could conceivably cancel some of the dependence seen in Fig. 19.



**Fig. 18:** Left: the kinematic regions covered by HERA, fixed-target scattering experiments, and by the LHC (adapted from the corresponding plot by Stirling). Right: the factor by which the gluon distribution evolves in going from a scale of 2 GeV to 100 GeV using CTEQ61 distributions as a fixed input at the low scale, and carrying out LO DGLAP evolution with HOPPET [40] with  $x_\mu = \frac{1}{2}, 1, 2$  (see text for further details).



**Fig. 19:** Uncertainties on the evolution of the gluon distribution from a fixed (CTEQ61) input at scale 2 GeV up to 100 GeV, for LO, NLO and NNLO evolution. The bands correspond to the envelope of the results with three scale choices  $x_\mu = \frac{1}{2}, 1, 2$  and what is plotted is the ratio to the result at scale 100 GeV, as obtained with NNLO evolution using  $x_\mu = 1$ .



**Fig. 20:** Left: production of a gluino ( $\tilde{g}$ ) pair and the subsequent decay of each gluino through a squark ( $\tilde{q}$ ) to quarks and a neutralino ( $\chi^0$ ); the experimental signature involves four jets and missing transverse energy ( $E_T$ ) from the unobserved neutralino. Right: a background that mimics this signature, with missing energy coming from the production of a  $Z$ -boson that decays to neutrinos.

### 3.4 Summary

Here are some of the points to retain from this section. Firstly, the proton really is what we expect it to be, i.e., a  $uud$  state; however, fluctuations of the proton introduce many extra  $q\bar{q}$  pairs (‘sea’), as well as a substantial amount of ‘glue’, carrying 50% of the proton’s momentum. The sea and gluon distributions diverge at small momentum fractions  $x$ .

Determination of the proton’s PDFs involves fitting data from a range of experiments, with direct sensitivity to quarks (e.g., DIS), indirect sensitivity to the gluon (DIS  $Q^2$  evolution), and direct sensitivity to quarks and gluons (jet data).

One of the keys to being able to measure consistent PDFs from different experiments, thinking about them in perturbative QCD and then applying them to predict results at new experiments is ‘factorization’: initial-state radiation, though collinear divergent, is process-independent; the divergent part can be absorbed into the definition of the PDFs, and then a universal set of PDFs, evolved between different scales with the DGLAP equations, can be used for any process.

Finally, the accuracy with which we know PDFs is quite remarkable: both from the experimental side and the theoretical side, in the best cases we know the PDFs to within a few per cent. This will be important in interpreting future signals of new particles, for example in Higgs-boson production at the LHC when we want to deduce its electroweak couplings given a measurement of its cross section.

If you need to use PDFs yourself, the best place to get them is from the LHAPDF library [45].

## 4 Predictive methods for LHC

In this section we will look at some of the different classes of technique can be used to make QCD predictions at LHC. Among the topics that we’ll touch on are leading order (LO), next-to-leading order (NLO) and next-to-next-leading order (NNLO) calculations, parton-shower Monte Carlos, and then methods to combine the two types of calculation.

Many of the examples that we’ll use will involve  $Z$  (and sometimes  $W$ ) production at hadron colliders. One reason is that  $Z$  and  $W$  bosons decay to leptons and neutrinos (missing energy), both of which are easily-taggable handles that are characteristic of signals in many new-physics scenarios. An illustration is given in Fig. 20, which depicts supersymmetric production of a gluino pair and subsequent decay to four jets and missing transverse energy from the unobserved neutralinos. Because of the complexity of the decays and the fact that the missing energy is the sum of that from two neutralinos, it can be

difficult to extract clear kinematic structures (such as an invariant mass peak) that make a signal emerge unambiguously over the background. In such cases the contribution from the signal may just be to give a cross section that is larger than background expectations over a broad kinematic range. But that will only be a ‘signal’ if we understand what the backgrounds are.

The extent to which we will want to (or have to) rely on QCD predictions of backgrounds in deciding whether there are signals of new physics at the LHC is a subject that deserves in-depth consideration (for a nice discussion of it, see Ref. [46]). But QCD predictions will come into play in many other ways too. Monte Carlo parton shower programs, which simulate the full hadronic final state, are crucial in evaluating detector acceptances and response. And knowing QCD predictions (both for backgrounds and possible signals) is crucial in the design of methods to search for new physics, as well as for extracting meaning from the data (about couplings, spins, etc.) when, it is to be hoped, we finally see signals of something new.

## 4.1 Fixed-order predictions

Fixed-order predictions, which involve the first couple of terms in the QCD perturbative expansion for a given cross section, are conceptually quite simple: it is easy to state which contributions are included, and as one includes further orders in the expansion one can reasonably hope to see systematic improvement in the accuracy of one’s predictions.

We’ll first look at a couple of examples of fixed-order predictions, in order to develop a feel for how the perturbative expansion behaves, and how one estimates its accuracy. We will then examine more generally what theoretical inputs are needed for predictions for a given process, and what practical forms the predictive tools take.

### 4.1.1 Example 1: the cross section for $e^+e^- \rightarrow \text{hadrons}$ and its scale dependence

In Eq. (22), we wrote the total cross section for  $e^+e^- \rightarrow \text{hadrons}$  as a perturbative series expansion in  $\alpha_s$  that multiplied the Born cross section  $e^+e^- \rightarrow q\bar{q}$ . The expansion was formulated in terms of the coupling evaluated at a renormalization scale  $\mu_R$  equal to the centre-of-mass energy  $Q$ , i.e.,  $\alpha_s(\mu_R = Q)$ . That choice is, however, arbitrary: for example, the most energetic gluon that could be produced in  $e^+e^- \rightarrow q\bar{q}g$  would be one with  $E = Q/2$ , so maybe we should be choosing  $\mu_R = Q/2$ . And in loop diagrams, one integrates over gluon energies that go beyond  $Q$ , so maybe  $\mu_R = 2Q$  would be just as reasonable.

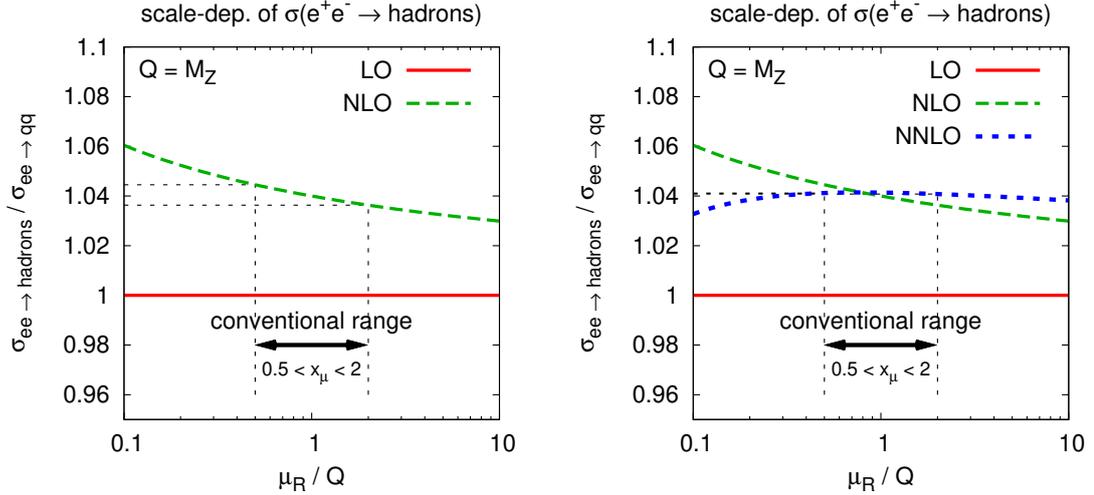
Because of this arbitrariness, a convention has emerged whereby one calculates a ‘central value’ for the prediction by setting the renormalization scale equal to the main physical scale for the process (e.g., the centre-of-mass energy at an  $e^+e^-$  collider; for hadron-collider processes the choice may be less obvious). The uncertainty is then estimated by varying the scale by a factor of two in either direction from the central value, i.e., taking  $\frac{Q}{2} < \mu_R < 2Q$ . This is illustrated in Fig. 21 (left), which plots

$$\sigma^{\text{NLO}} = \sigma_{q\bar{q}}(1 + c_1\alpha_s(\mu_R)), \quad (57)$$

as a function of  $\mu_R$ , showing how the  $\mu_R$ -dependence translates into an uncertainty; note that  $c_1$  can be read from Eq. (22). Given an expansion of the running coupling (i.e., of the middle result of Eq. (11),  $\alpha_s(\mu_R) = \alpha_s(Q) - 2b_0\alpha_s^2(Q) \ln \frac{\mu_R}{Q} + \mathcal{O}(\alpha_s^3)$ ), we can rewrite Eq. (57) as

$$\sigma^{\text{NLO}}(\mu_R) = \sigma_{q\bar{q}} \left( 1 + c_1\alpha_s(Q) - 2c_1b_0\alpha_s^2(Q) \ln \frac{\mu_R}{Q} + \mathcal{O}(\alpha_s^3) \right). \quad (58)$$

This tells us that as we vary the renormalization scale for a prediction up to  $\mathcal{O}(\alpha_s)$  (NLO), we effectively introduce  $\mathcal{O}(\alpha_s^2)$  (NNLO) pieces into the calculation: by generating some fake set of NNLO terms, we are probing the uncertainty of the cross section associated with the missing full NNLO correction.



**Fig. 21:** Renormalization-scale dependence of the NLO (left) and NNLO (right) predictions for the  $e^+e^- \rightarrow$  hadrons total cross section, together with an indication of the conventional choice of scale-variation range

If we calculate the actual NNLO cross section for general  $\mu_R$ , it will have a form

$$\sigma^{\text{NNLO}}(\mu_R) = \sigma_{q\bar{q}} \left( 1 + c_1 \alpha_s(\mu_R) + c_2(\mu_R) \alpha_s^2(\mu_R) \right). \quad (59)$$

Observe that the  $c_2$  coefficient now depends on  $\mu_R$ . This is necessary because the second-order coefficient must cancel the  $\mathcal{O}(\alpha_s^2)$  ambiguity due to the scale choice in Eq. (58). This constrains how  $c_2(\mu_R)$  depends on  $\mu_R$ :

$$c_2(\mu_R) = c_2(Q) + 2c_1 b_0 \alpha_s^2(Q) \ln \frac{\mu_R}{Q}, \quad (60)$$

where  $c_2(Q)$  can again be read from Eq. (22). If we now express  $\sigma^{\text{NNLO}}(\mu_R)$  in terms of  $\alpha_s(Q)$ , we will find that the residual dependence on  $\mu_R$  appears entirely at  $\mathcal{O}(\alpha_s^3)$ , i.e., one order further than in Eq. (58). This is reflected in the right-hand plot of Fig. 21, which illustrates how the impact of the scale variation at NNLO is significantly reduced, since we are now probing the impact of missing  $\alpha_s^3$  terms, rather than  $\alpha_s^2$  terms.

If we had an arbitrarily large number of terms in the  $\alpha_s$  expansion, the scale dependence would disappear exactly. The fact it doesn't in the presence of a fixed number of terms may initially seem like a drawback, but in some respects it's a blessing in disguise because it provides a useful handle on the uncertainties. This is why scale variation has become a standard procedure. It's worth bearing in mind that it isn't a failsafe mechanism: a trivial example comes from the LO curve in Fig. 21. It doesn't have any scale variation because they don't depend on  $\alpha_s$ , yet it differs significantly from the higher-order results.

#### 4.1.2 Example 2: $pp \rightarrow Z$

At LO the  $pp \rightarrow Z$  cross section involves a single underlying hard partonic process, namely  $q\bar{q} \rightarrow Z$ , which is purely electroweak. To go from the  $q\bar{q} \rightarrow Z$  squared matrix element to the  $pp \rightarrow Z$  result, one must integrate over the quark distributions

$$\sigma_{pp \rightarrow Z}^{\text{LO}} = \sum_i \int dx_1 dx_2 f_{q_i}(x_1, \mu_F^2) f_{\bar{q}_i}(x_2, \mu_F^2) \hat{\sigma}_{0, q_i \bar{q}_i \rightarrow Z}(x_1 p_1, x_2 p_2), \quad (61)$$

for which one must choose a factorization scale  $\mu_F$ . A natural choice for this scale is  $\mu_F = M_Z$ , but as with the renormalization scale it is conventional to vary it by a factor of two either side of the central choice in order to obtain a measure of the uncertainties in the prediction.

Adding NLO and NNLO terms, the structure becomes

$$\sigma_{pp \rightarrow Z+X}^{\text{NNLO}} = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \left[ \hat{\sigma}_{0,ij \rightarrow Z}(x_1, x_2) + \alpha_s(\mu_R) \hat{\sigma}_{1,ij \rightarrow Z+X}(x_1, x_2, \mu_F) + \alpha_s^2(\mu_R) \hat{\sigma}_{2,ij \rightarrow Z+X}(x_1, x_2, \mu_F, \mu_R) \right]. \quad (62)$$

We now have a sum over the flavours  $i$  and  $j$  of the initial partons, because starting from NLO there are contributions from (say) gluon-quark scattering [cf. Fig. 22 (left)]. The cross section is written as being for  $Z+X$ , where the  $X$  means that we allow anything (e.g., quarks, gluons) to be produced in addition to the  $Z$ -boson. At  $\mathcal{O}(\alpha_s)$  the  $\mu_F$  dependence of the  $\sigma_1$  coefficient partially cancels the dependence present at  $\mathcal{O}(\alpha_s^0)$  coming from the  $\mu_F$  dependence of the PDFs. That dependence is further cancelled at  $\mathcal{O}(\alpha_s^2)$ , as is part of the  $\mu_R$  dependence that is introduced in the  $\mathcal{O}(\alpha_s(\mu_R))$  term. The plot on the right of Fig. 22 shows the  $Z$ -boson cross section as a function of its rapidity [47]. The bands indicate the uncertainty due to scale variation (taking  $\frac{1}{2}M_Z < \mu_R = \mu_F < 2M_Z$ )<sup>6</sup> and show how this uncertainty undergoes important reductions going from LO to NLO to NNLO.

One of the interesting features that comes out of Fig. 22 is that the LO prediction is only good to within a factor of 1.5 to 2, despite the fact that  $\alpha_s(M_Z) \simeq 0.118$  would imply 10% accuracy. This is because the  $\mathcal{O}(\alpha_s)$  corrections come with large coefficients. This is not uncommon in hadron-collider cross sections. Furthermore the LO uncertainty band seems not to provide a faithful measure of the true uncertainty. Other aspects of the perturbative expansion do seem to behave as one would expect: the size of the uncertainty band decreases significantly going from LO to NLO (10–20%) to NNLO (a few per cent). And the actual shift in the central value in going from NLO to NNLO is substantially smaller than that from NLO to LO.

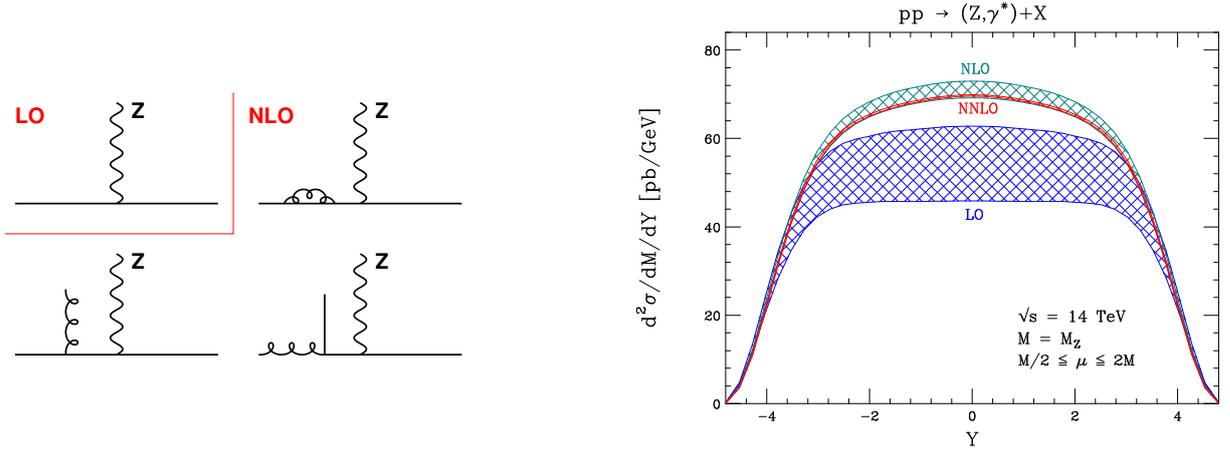
Are these characteristics representative of the ‘typical’ situation for collider observables? We only have predictions up to NNLO in a handful of cases (see below) and in those it is. In cases where we just have NLO predictions, the features of large ‘K-factors’ (NLO/LO enhancements) with a reduced NLO uncertainty band are not uncommon, suggesting that beyond NLO corrections should be small. Exceptions are known to arise in two types of case: those where new enhanced partonic scattering channels open up at NLO (or beyond); and that involve two disparate physical scales. For example, if you ask for the  $Z$ -boson to have a transverse momentum  $p_t$  that is much smaller than  $M_Z$ , then each power of  $\alpha_s$  in the expansion of the cross section will be accompanied by up to two powers of  $\ln^2 M_Z/p_t$ , leading to large coefficients at all orders in the perturbative expansion. These are due to incomplete cancellation between real and virtual (loop) divergences: loop corrections do not affect the  $Z$ -boson  $p_t$  and so are fully integrated over, whereas real emissions do affect the  $Z$   $p_t$  and so are only partially integrated over.

#### 4.1.3 Predictions for more complex processes

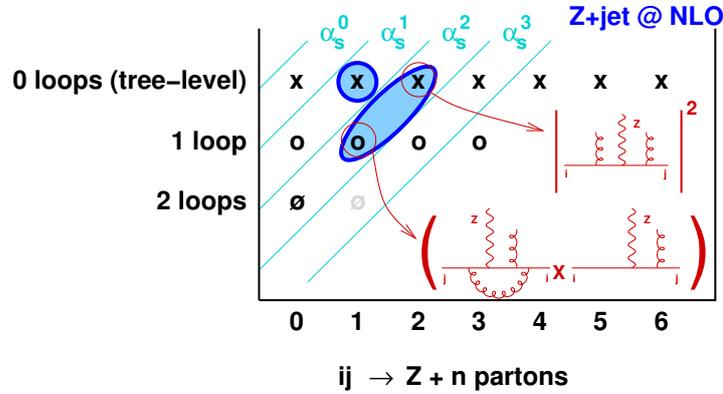
As an example of a more complex process, consider the production of a  $Z$ -boson plus a jet. The leading order cross section requires the calculation of the  $\mathcal{O}(\alpha_s)$  squared diagrams for  $q\bar{q} \rightarrow Zg$ ,  $qg \rightarrow Zq$  and  $\bar{q}g \rightarrow Z\bar{q}$ . The NLO cross section additionally requires all  $\mathcal{O}(\alpha_s^2)$  contributions with a  $Z$  boson and at least one jet, as illustrated in Fig. 23, i.e., the squared tree-level diagram for  $ij \rightarrow Z + 2$  partons and the interference of the 1-loop and tree-level diagrams for  $ij \rightarrow Z + 1$  parton.

More generally, Fig. 23 allows you to read off the contributions that you will need for an  $N^p$ LO calculation of  $ij \rightarrow Z + n$  partons: just take all entries in Fig. 23 with at least  $n$  partons, up to order  $\alpha_s^{n+p}$ . Entries in black are known and have already been used to obtain predictions for LHC and Tevatron. The one entry in grey, the 2-loop  $Z + 1$  parton contribution, is known but has yet to be used in any hadron-collider prediction, for reasons that we will discuss below. Entries that are absent (e.g.,  $Z + 2$  partons at two loops) have so far proven to be too complicated to calculate.

<sup>6</sup>The variation of  $\mu_R$  and  $\mu_F$  simultaneously, though common, is not the only possible procedure. An attractive alternative is to vary both independently around a central scale, with the additional requirement that  $\frac{1}{2} < \mu_R/\mu_F < 2$  [48].



**Fig. 22:** Left: classes of diagram that appear for  $pp \rightarrow Z$  at LO and at NLO. Right: cross section at the LHC for the  $Z$ -boson, differential in rapidity, at LO, NLO and NNLO, as taken from Ref. [47].



**Fig. 23:** Illustration of the contributions that are known for  $ij \rightarrow Z + n$  partons, where  $i$  and  $j$  are arbitrary incoming partons, according to the number of outgoing partons, the number of loops and the number of powers of the coupling. An ‘x’ represents a squared tree-level diagram, an ‘o’ represents the interference of a 1-loop diagram with a tree-level diagram, and a ‘ $\phi$ ’ represents the interference of a two-loop diagram with a tree-level diagram or the square of a 1-loop diagram. Entries in black are known and used; entries in grey are known but have not been used. The entries in the shaded ellipses are those that are relevant for the NLO calculation of the cross section for the production of a  $Z$ -boson with a jet.

The classes of contributions calculated for  $ij \rightarrow Z + n$  partons provide a representative view of the situation for other processes as well, with tree-level diagrams calculated up to quite high final-state multiplicities,  $\sim 10$ , 1-loop diagrams having been used for processes with up to 3 or sometimes 4 final-state particles, and 2-loop diagrams available and used only for  $2 \rightarrow 1$  type processes, essentially  $pp \rightarrow W$ ,  $pp \rightarrow Z/\gamma^*$  and  $pp \rightarrow H$ .

It’s natural to ask in what form these various calculations are available. For certain very simple quantities, for example, the total cross section for  $t\bar{t}$  production, or for  $W$ ,  $Z$  or Higgs production, the result of the perturbative calculation can be written as in Eq. (62),

$$\sigma_{pp \rightarrow A+X}^{\text{N}^{\text{pLO}}} = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \times \sum_{m=0}^p \alpha_s^{n+m}(\mu_R) \hat{\sigma}_{m,ij \rightarrow A+X}(x_1 x_2 s, \mu_R, \mu_F), \quad (63)$$

where the  $\sigma_{m,ij \rightarrow A+X}(x_1 x_2 s, \mu_R, \mu_F)$  are functions whose analytical expressions can be found in the relevant papers (Refs. [49, 50] for  $W$  and  $Z$ , and Refs. [50–53] for Higgs-boson production). To obtain

**Table 1:** The number of Feynman diagrams for tree level  $gg \rightarrow N$  gluon scattering [59]

$N$	2	3	4	5	6	7	8
No. diags	4	25	220	2485	34300	$5 \times 10^5$	$10^7$

a prediction, one just has to type them into a computer program and then integrate over  $x_1$  and  $x_2$ .

In most cases, however, one wants to calculate a cross section that incorporates experimental cuts, such as lepton acceptances, transverse momentum cuts on jets, etc. In these cases the type of tool that should be used depends on the order to which you want the answer.

#### 4.1.3.1 LO predictions

As long as one is dealing with infrared safe observables, then for a LO prediction one need only include tree-level diagrams, in kinematic regions in which their contributions are finite. The simplest approach therefore is to carry out Monte Carlo integration over phase-space points, have a subroutine that determines whether a given phase-space point passes the cuts, and if it does calculate the squared matrix elements and PDF factors for each possible partonic subprocesses.

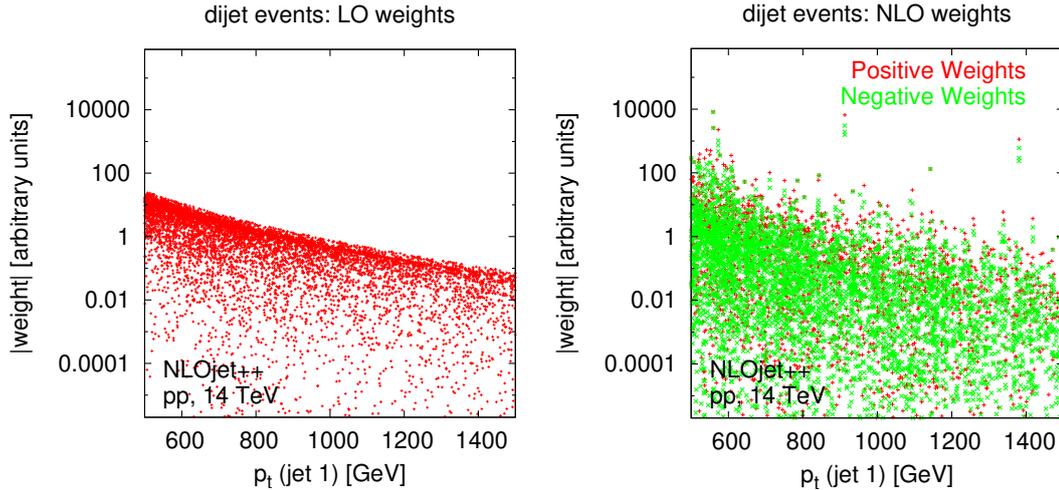
Quite a number of tools enable you to do this: ALPGEN [54], COMIX/SHERPA [55], COMPHEP [56], HELAC/PHEGAS [57] and MADGRAPH [58]. They allow you to calculate cross sections for a broad range of  $2 \rightarrow n$  scattering processes with  $n$  up to 6–8 (or in some cases even beyond). Some of these (COMPHEP, MADGRAPH) use formulae obtained from direct evaluations of Feynman diagrams. This gives them significant flexibility in terms of the range or processes they support (e.g., with easy inclusion of many new physics models), though they suffer at large  $n$  because the number of diagrams to evaluate grows rapidly as  $n$  increases (cf. Table 1). Others (ALPGEN, HELAC/PHEGAS and COMIX/SHERPA) use methods designed to be particularly efficient at high multiplicities, such as Berends–Giele recursion [60], which builds up amplitudes for complex processes by recursively reusing simpler ones (a nice technical review of the technique is given in Ref. [61]).

#### 4.1.3.2 NLO predictions

When, in Sections 2.2 and 2.3.2, we looked at the cancellation between divergences in real and loop diagrams, we wrote the loop diagram with an explicit integral over phase space so as to be able to match the divergences between real and loop diagrams and cancel them easily.

A subtlety that we ignored is that in practical evaluations of loop diagrams, the integral over loop momenta is carried out in  $4 - \epsilon$  dimensions rather than 4 dimensions, in order to regularize the divergences that appear and obtain an answer whose finite part is known independently of any related tree-level diagrams. On the other-hand, experimental cuts are defined in four dimensions, so the real tree-level diagrams must be integrated in four dimensions, which implies divergent results if the real diagrams are taken alone.

This mismatch between the ways loop and tree-level diagrams are handled is one of the main difficulties in carrying out calculations that include experimental cuts beyond LO. For the calculation of a process with  $n$  partons at LO, the standard technique nowadays to deal with the problem is to introduce a  $n+1$ -parton *counterterm* where the  $n+1^{\text{th}}$  parton is always soft and collinear so that it doesn't affect an IR safe observable. It is subtracted from the  $n+1$ -parton real diagram in four dimensions and designed so as to cancel all of its soft and collinear divergences. It is also designed such that the kinematics of its  $n+1^{\text{th}}$  parton can be integrated analytically in  $4 - \epsilon$  dimensions so that the result can be easily added to the loop diagram and cancel *its* divergences. Since the counterterm is subtracted once and added once, its net impact is null, and can just be thought of as a way of reshuffling divergences. This is known as a ‘subtraction’ procedure and the variant most widely used in current NLO computer codes is ‘dipole’



**Fig. 24:** Left: weights associated with tree-level  $LO\ 2 \rightarrow 2$  scattering events in the calculation of the dijet cross section, shown as a function of the transverse momentum of the harder jet. Right: weights at NLO from real  $2 \rightarrow 3$  events, subtraction counterterm events and loop events (after addition of the integrated counterterm), again as a function of  $p_t$  of the hardest jet. Results obtained with NLOjet++ [65].

subtraction [62]; other methods that have seen numerous applications include FKS [63] and antenna [64] subtraction.

An illustration of how subtraction works in practice is given in Fig. 24 for dijet production. In the left-hand plot we see weights for the LO process as a function of the jet  $p_t$ . Though there is some dispersion in the weights, there is a clear upper bound as a function of  $p_t$ , reflecting the finiteness of the matrix-elements. The right-hand plot shows weights at NLO. Here there is no clear upper bound to the weights; however we do see that unusually large (nearly divergent) weights come in clusters: usually one positive (red, '+') weight, accompanied by one or more negative (green, 'x') weights with identical jet transverse momenta, so that each event in the cluster contributes to the same bin of the cross section and their weights sum to a finite result.

Technically, one main consideration has so far limited the range of processes for which NLO results exist: the availability of the loop amplitude. Until recently loop amplitudes were usually calculated semi-manually for each process. The complexity of the calculations increased significantly with the number of outgoing legs, limiting available results to those with at most three outgoing partons. Many NLO results for  $2 \rightarrow 2$  and  $2 \rightarrow 3$  processes are incorporated into programs such as NLOJET++ for jet production [65], MCFM for processes with heavy quarks and/or heavy electroweak bosons [66], VBFNLO for vector-boson fusion processes [67], and the PHOX family [68] for processes with photons in the final state.

In the past couple of years, techniques have come to fruition that offer the prospect of automated calculation of arbitrary loop diagrams. Though full automation is not here yet, a number of  $2 \rightarrow 4$  processes have now been calculated at NLO thanks to these advances, including  $pp \rightarrow t\bar{t}b\bar{b}$  [69, 70],  $pp \rightarrow t\bar{t}jj$  [71] (where  $j$  represents a jet) and  $pp \rightarrow W + 3j$  [72, 73] and  $pp \rightarrow Z + 3j$  [74].<sup>7</sup>

It should be said that NLO calculations are very computing intensive: for some observables it is not unusual to have to devote several years of CPU time in order to get adequate numerical convergence of the Monte Carlo integration.

<sup>7</sup>Since the original version of this writeup, first results for a  $2 \rightarrow 5$  process have also appeared [75].

### 4.1.3.3 NNLO predictions

NNLO predictions suffer from the same problem of cancelling divergences between real and virtual corrections that is present at NLO, with the complication that instead of having one soft and one collinear divergence, there are now two of each, greatly complicating the task of figuring out counterterms to allow experimental cuts to be implemented in four dimensions.

As a result, the only general subtraction type approaches that exist currently are for processes without incoming hadrons, notably  $e^+e^- \rightarrow 3j$  [76, 77]. For hadron collider processes it is only  $2 \rightarrow 1$  processes that are available, specifically vector-boson (FEWZ [78] and DYNNLO [79]) and Higgs-boson (FEHiP [80] and HNNLO [81]) production, using methods that are not so easily generalizable to more complicated processes.

## 4.2 Monte Carlo parton-shower programs

The programs we've discussed so far, known as 'Matrix Element Monte Carlos' provide a powerful combination of accuracy and flexibility as long as you want to calculate IR and collinear safe observables (jets,  $W$ 's,  $Z$ 's, but not pions, kaons, etc.), don't mind dealing with wildly fluctuating positive and negative weights, and don't need to study regions of phase space that involve disparate physical scales.

All these defects are essentially related to the presence of soft and collinear divergences. Yet we know that real life does not diverge. So it is natural to wonder whether we can reinterpret the divergences of perturbation theory physically. It turns out that the right kind of question to ask is "what is the probability of *not* radiating a gluon above some (transverse momentum) scale  $k_t$ ". Starting from a  $q\bar{q}$  system, using the results of Section 2, we know that to  $\mathcal{O}(\alpha_s)$  the answer in the soft and collinear limit goes as

$$P(\text{no emission above } k_t) \simeq 1 - \frac{2\alpha_s C_F}{\pi} \int^Q \frac{dE}{E} \int^{\pi/2} \frac{d\theta}{\theta} \Theta(E\theta - k_t). \quad (64)$$

It so happens that in the soft and collinear limit, this result is easy to work out not just at first order, but at all orders, giving simply the exponential of the first order result

$$P(\text{no emission above } k_t) \equiv \Delta(k_t, Q) \simeq \exp \left[ -\frac{2\alpha_s C_F}{\pi} \int^Q \frac{dE}{E} \int^{\pi/2} \frac{d\theta}{\theta} \Theta(E\theta - k_t) \right]. \quad (65)$$

Whereas Eq. (64) had a 'bare' infinity if one took  $k_t \rightarrow 0$ , Eq. (65) is simply bounded to be between 0 and 1.

The quantity  $\Delta(k_t, Q)$  is known as a Sudakov form factor. We've been very approximate in the way we've calculated it, neglecting for example the running of the coupling ( $\alpha_s$  should be placed inside the integral and evaluated at the scale  $E\theta$ ) and the treatment of hard collinear radiation (the  $dE/E$  integral should be replaced with the full collinear splitting function), but these are just technical details. The importance of the Sudakov form factor is that it allows us to easily calculate the distribution in transverse momentum  $k_{t1}$  of the gluon with largest transverse momentum in an event:

$$\frac{dP}{dk_{t1}} = \frac{d}{dk_{t1}} \Delta(k_{t1}, Q). \quad (66)$$

This distribution is easy to generate by Monte Carlo methods: take a random number  $r$  from a distribution that's uniform in the range  $0 < r < 1$  and find the  $k_{t1}$  that solves  $\Delta(k_{t1}, Q) = r$ . Given  $k_{t1}$  we also need to generate the energy for the gluon, but that's trivial. If we started from a  $q\bar{q}$  system (with some randomly generated orientation), then this gives us a  $q\bar{q}g$  system. As a next step one can work out the Sudakov form factor in the soft/collinear limit for there to be no emission from the  $q\bar{q}g$  system as a whole above some scale  $k_{t2}$  ( $< k_{t1}$ ) and use this to generate a second gluon. The procedure is then repeated over and over again until you find that the next gluon you would generate is below some non-perturbative cutoff scale  $Q_0$ , at which point you stop. This gives you one 'parton shower' event.

a)											c)															
---INITIAL STATE---											---PARTON SHOWERS---															
IHEP	ID	IDPDG	IST	M01	M02	DA1	DA2	P-X	P-Y	P-Z	ENERGY	MASS	IHEP	ID	IDPDG	IST	M01	M02	DA1	DA2	P-X	P-Y	P-Z	ENERGY	MASS	
1	P	2212	101	0	0	0	0	0,00	0,00	7000,0	7000,0	0,94	9	UQRK	94	141	4	6	11	16	2,64	-9,83	592,2	590,2	-49,07	
2	P	2212	102	0	0	0	0	0,00	0,00	-7000,0	7000,0	0,94	10	CONE	0	100	4	5	0	0	-0,27	0,96	0,1	1,0	0,00	
3	CHF		0	103	1	2	0	0,00	0,00	0,0	14000,0	14000,0		11	GLUON	21	2	9	12	32	33	-1,02	3,59	5,6	6,7	0,75
													12	GLUON	21	2	9	13	34	35	0,25	1,46	3,6	4,0	0,75	
													13	GLUON	21	2	9	14	36	37	-0,87	1,62	4,7	5,1	0,75	
													14	GLUON	21	2	9	15	38	39	-0,81	4,17	3611,7	3611,7	0,75	
													15	GLUON	21	2	9	16	40	41	-0,19	-1,01	1727,7	1727,7	0,75	
													16	UD	2101	2	9	25	42	41	0,00	0,00	1054,6	1054,6	0,52	
													17	GLUON	94	142	5	6	19	21	-2,23	0,44	-233,5	232,8	-18,36	
													18	CONE	0	100	5	8	0	0	0,77	0,64	0,2	1,0	0,00	
													19	GLUON	21	2	17	20	43	44	1,60	0,58	-2,1	2,8	0,75	
													20	UD	2101	2	17	21	45	44	0,00	0,00	-2687,6	2687,6	0,52	
													21	UQRK	2	2	17	32	45	45	0,53	-1,02	-4078,3	4076,9	0,32	
													22	Z0/GAMA*	23	195	7	22	251	252	-257,66	-219,68	324,8	477,5	88,56	
													23	UQRK	94	144	8	6	25	31	258,06	210,29	33,9	345,5	86,10	
													24	CONE	0	100	8	5	0	0	0,21	0,17	-1,0	1,0	0,00	
													25	UQRK	2	2	23	26	47	42	26,82	24,33	23,7	43,3	0,32	
													26	GLUON	21	2	23	27	48	49	8,50	8,18	6,0	13,3	0,75	
													27	GLUON	21	2	23	28	50	51	73,27	61,24	12,0	96,2	0,75	
													28	GLUON	21	2	23	29	52	53	73,66	58,54	-6,3	94,3	0,75	
													29	GLUON	21	2	23	30	54	55	67,58	52,13	-7,3	85,7	0,75	
													30	GLUON	21	2	23	31	56	57	6,98	4,60	2,3	8,7	0,75	
													31	GLUON	21	2	23	43	58	59	1,24	1,26	3,6	4,1	0,75	

**Fig. 25:** Sequence of steps in the generation of a  $pp \rightarrow Z + j$  event in HERWIG: (a) specification of the colliding beams and their energy, (b) generation of the kinematics and partonic flavour of the hard subprocess,  $ug \rightarrow Zu$ , and (c) generation of the initial- and final-state parton showers.

This is essentially the procedure that’s present in the shower of PYTHIA 8 [82] and the  $p_t$  ordered option of PYTHIA 6.4 [83], as well as ARIADNE [84] and SHERPA 1.2 (the SHERPA reference [85] describes an earlier version). It is also possible to choose other ordering variables: the original PYTHIA shower [83, 86] is based on virtuality ordering (plus an angular veto). This is still the most widely used shower which works well on a range of data (though there are theoretical issues in some formulations of virtuality ordering). In the HERWIG family of programs [87, 88] it is angular ordering that is used.

The above shower descriptions hold for final-state branching. With initial-state hadrons, one also needs to be careful with the treatment of the PDFs, since the collinear splitting that is accounted for in the parton shower is also connected with the way the PDF is built up at the scale of the hard scattering.

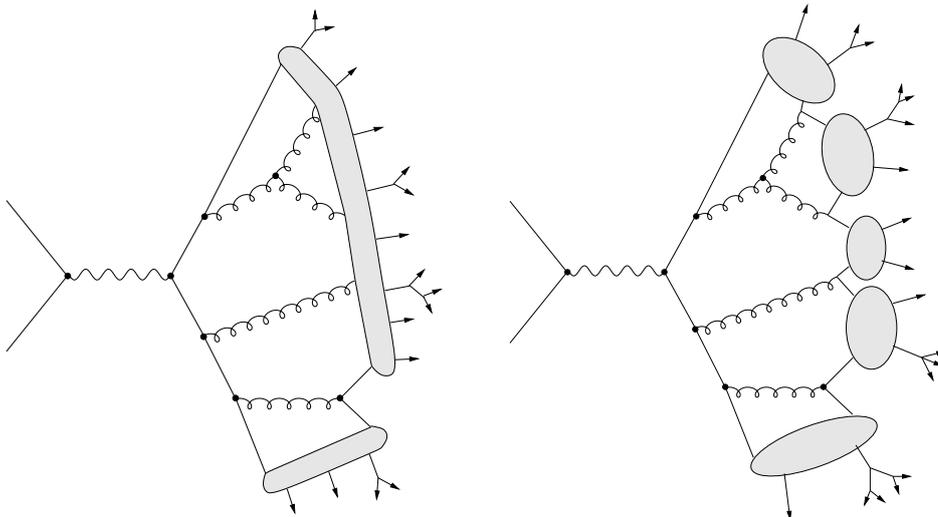
The sequence of steps for the generation of a parton-shower event in  $pp$  collisions is illustrated in Fig. 25.

Real events consist not of partons but of hadrons. Since we have no idea how to calculate the transition between partons and hadrons, Monte Carlo event generators resort to ‘hadronization’ models. One widely-used model involves stretching a colour ‘string’ across quarks and gluons, and breaking it up into hadrons [89, 90]. For a discussion of the implementation of this ‘Lund’ model in the MC program PYTHIA, with further improvements and extensions, Ref. [86] and references therein provide many details. Another model breaks each gluon into a  $q\bar{q}$  pair and then groups quarks and anti-quarks into colourless ‘clusters’, which then give the hadrons. This cluster type hadronization is implemented in the HERWIG event generator [87, 88, 91] and recent versions of SHERPA. Both approaches are illustrated in Fig. 26

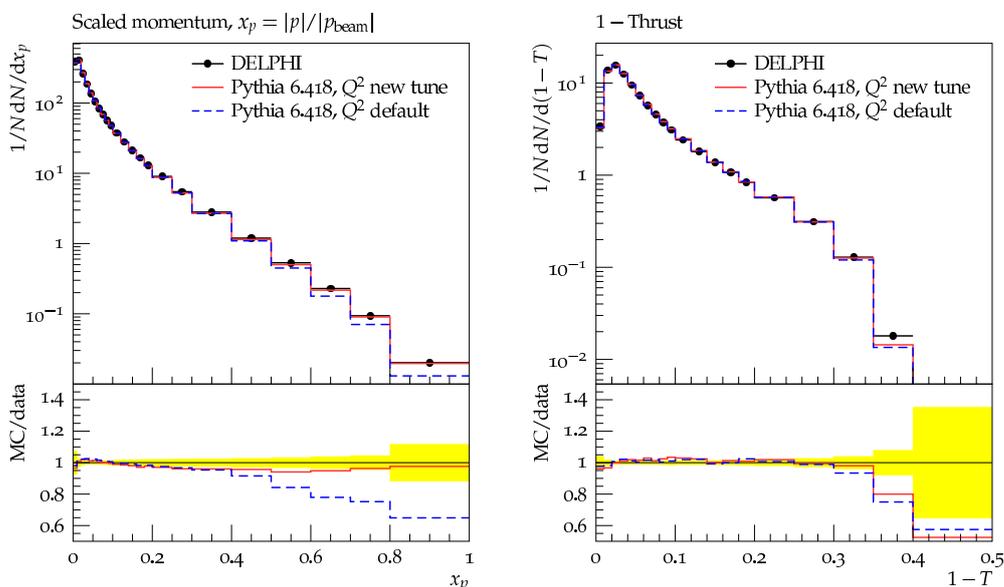
Hadronization models involve a number of ‘non-perturbative’ parameters. The parton-shower itself involves the non-perturbative cutoff  $Q_0$ . These different parameters are usually tuned to data from the LEP experiments. The quality of the description of the data that results is illustrated in Fig. 27.

A final point in our brief description of Monte Carlo event generators concerns the ‘underlying event’ in  $pp$  and  $\gamma p$  collisions. In addition to the hard interaction that is generated by the Monte Carlo simulation, it is also necessary to account for the interactions between the incoming proton (or photon) remnants. This is usually modelled through multiple extra  $2 \rightarrow 2$  scattering, occurring at a scale of a few GeV, and known as multiple parton interactions. This modelling of the underlying event is crucial in order to give an accurate reproduction of the (quite noisy) energy flow that accompanies hard scatterings in hadron-collider events.

Our description here of Monte Carlo event generators has been fairly brief. For a more complete discussion, a good starting point is the lectures notes by Sjöstrand [93] from the 2006 School.



**Fig. 26:** Illustration of string (left) and cluster (right) fragmentation, taken from Ref. [1]



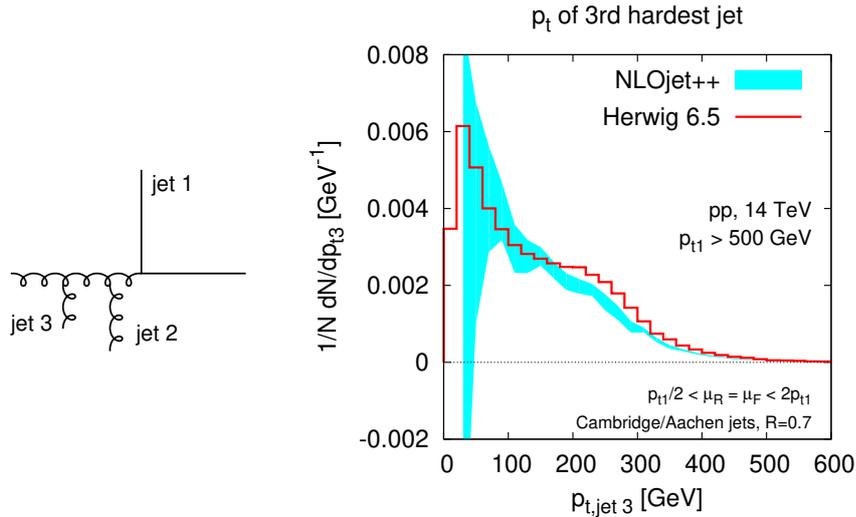
**Fig. 27:** A comparison of DELPHI  $e^+e^-$  data for the particle (scaled) momentum distribution (left) and the thrust event shape distribution (right) with two tunes of the PYTHIA event generator [92].

### 4.3 Comparing fixed-order and parton-shower programs

Parton-shower Monte Carlo programs do a good job of describing most of the features of common events, including the hadron-level detail that is essential for the correct simulation of detector effects on event reconstruction. Another nice feature of theirs is that events have equal weight, just as with real data.

A drawback of parton-shower Monte Carlos is that, because they rely on the soft and collinear approximation, they do not necessarily generate the correct pattern of hard large-angle radiation. This can be important, e.g., if you're simulating backgrounds to new-physics processes, for which often the rare, hard multi-jet configurations are of most interest. In contrast, fixed-order programs do predict these configurations correctly.

The purpose of this section is to give two examples of comparisons between parton-shower predictions and fixed-order predictions, in order to help illustrate their relative strengths.



**Fig. 28:** The predicted  $p_t$  distribution for the jet with the third largest transverse momentum,  $p_{t3}$ , in 14 TeV pp events where the hardest jet has transverse momentum  $p_{t1} > 500$  GeV. The solid histogram is the result from HERWIG 6.5 [87]. The band corresponds to the ratio of the NLO 3-jet cross section, differential in  $p_{t3}$  (with the  $p_{t1}$  cut), to the NLO cross section for events to pass the  $p_{t1}$  cut. The width of the band corresponds to the scale uncertainty and the results have been obtained with NLOJET++ [65].

#### 4.3.1 Jet production

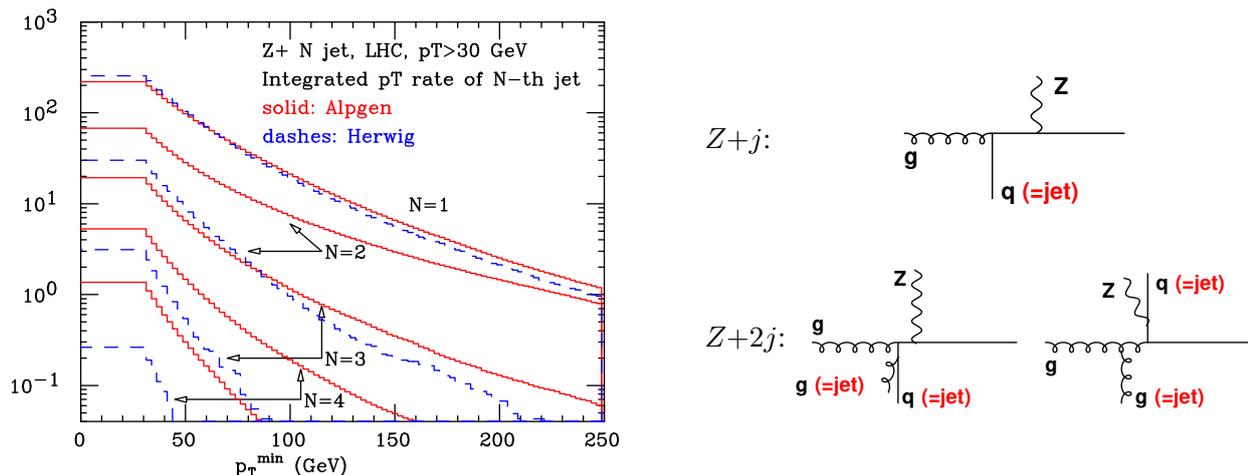
In plain jet production, parton shower Monte Carlo start from a hard event that consists of parton–parton scattering diagrams like  $qq \rightarrow qq$ ,  $qg \rightarrow qg$ , etc., and then rely on the showering to generate extra radiation. While the showering is only correct in the soft and/or collinear limit, it does sometimes generate extra hard radiation. In Fig. 28 we can see distribution in transverse momentum of the 3rd hardest jet, in HERWIG events in which we have imposed a cut on the hardest jet,  $p_{t1} > 500$  GeV. That is compared to the NLO (NLOJET++) prediction for the same distribution, together with its uncertainty band from scale variation.

In much of the range, one observes relatively good agreement between the two distributions: the 20–30% differences that are visible around  $p_{t3} \equiv p_{t,jet 3} \sim 250$  GeV are no larger than the uncertainties that would be obtained from a LO calculation, despite the fact that in this region HERWIG does not even include the exact LO  $2 \rightarrow 3$  matrix element. Of course, it is hard to be sure whether the good agreement is meaningful generally, or instead just a coincidence specific to our particular choice of observable — and the only way to be sure is, for each new observable, to also generate the NLO prediction.

The NLO prediction is not without its limitations though: at low  $p_{t3}$ , the uncertainty band on the NLO prediction blows up. This is a manifestation of large higher-order corrections, which compromise the convergence of the perturbative series. They arise because we have a large ratio of scales between  $p_{t1} (\gtrsim 500 \text{ GeV})$  and  $p_{t3}$  (a few tens of GeV). Such large scale ratios translate into NLO corrections whose size relative to the LO contribution go as  $\alpha_s \ln^2 p_{t1}/p_{t3} \sim 1$  and  $\alpha_s \ln p_{t1}/p_{t3}$ .

#### 4.3.2 Vector-boson plus production

The picture seen above of good agreement between parton shower Monte Carlo and fixed-order predictions does not always hold. Events with vector bosons are among those that parton shower programs have the greatest difficulty reproducing. This is illustrated in Fig. 29 (left), which shows the ‘integrated  $E_T$  spectrum for the  $N^{\text{th}}$  jet’ in events with a  $Z$ -boson, i.e., the cross section for the  $N^{\text{th}}$  jet to have a transverse energy above  $E_T$ . Results are given both from HERWIG and from ALPGEN, which provides an exact LO (tree-level) prediction for each jet’s spectrum.



**Fig. 29:** Left: the cross section for the  $N^{\text{th}}$  jet to have a transverse energy above a given  $E_T$ , in (14 TeV) LHC events with a  $Z$ -boson, as calculated with HERWIG and a tree-level (LO) prediction from ALPGEN. Figure taken from Ref. [46]. Right: kinematic configurations contributing to  $Z$ +jet and  $Z$  + 2jet events.

The distribution for the first jet is fine: this is by construction, since HERWIG (like PYTHIA) includes the full matrix element for  $Z$ +parton production. What is shocking is the result for the second (and higher) jets, for which the  $E_T$  spectra are in complete disagreement between HERWIG and ALPGEN.

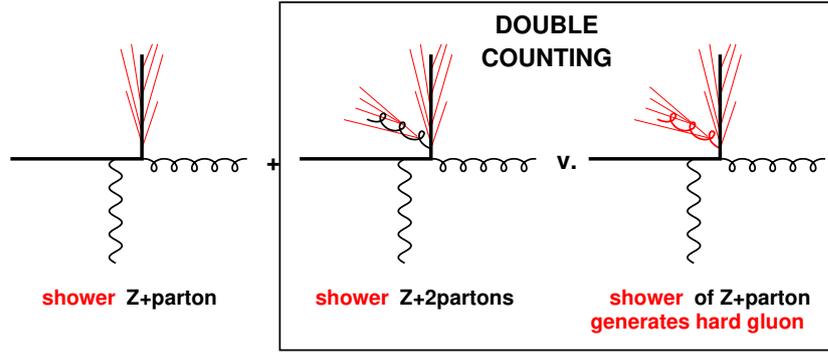
At first sight it is mysterious how HERWIG could be doing such a good job for pure jet production, Fig. 28, yet such a poor job when there's also a  $Z$ -boson in the event, Fig. 29. The diagrams in the right-hand part of Fig. 29 help explain what's going on. HERWIG generates hard configurations like those in the upper line, labelled  $Z$ +j. Events with two jets are generated in HERWIG by emission of a gluon off (say) the high- $p_t$  quark. However, there are also events (bottom right) which look like a dijet event with a  $Z$ -boson radiated off a quark line. Since at high  $p_t$  the  $Z$ -boson's mass becomes irrelevant in its production, such diagrams acquire soft and collinear enhancements (just like gluon radiation). However, today's parton-shower Monte Carlos only include QCD showering, not electroweak showering and therefore they are never in a position to start from a dijet event and radiate a  $Z$ -boson from it. Therefore they miss a very large part of the cross section.

This example helps illustrate a general feature of the use of Monte Carlo: if you are to trust the results, it is crucial that you know what you have asked the Monte Carlo to generate and whether the observable you are interested in is truly likely to be dominated by what the Monte Carlo can generate.

#### 4.4 Combining fixed-order and parton-shower methods

In the above subsections we saw various strengths and weaknesses of different predictive techniques: NLO codes give predictions with well controlled normalizations, for a reasonable range of processes, as long as one isn't faced with observables that involve disparate scales. Tree-level (LO) predictions can be generated up to quite high multiplicities for a broad range of processes, though without good control of the normalization (i.e., often no better than a factor of two). And parton shower Monte Carlos provide reliable behaviour in soft-collinear regions, giving a fully exclusive final state, though they have normalizations which at best are no better than LO normalizations and sometimes they do dramatically badly in reproducing multi-jet structure.

It is natural to ask whether one can develop tools that combine (or merge) the advantages of all three. This is an active research topic, and here we will just outline the ideas behind two well-established merging tasks: combination of different multiplicity LO matrix elements with parton showers; and combination of NLO and parton shower predictions.



**Fig. 30:** Illustration of the double-counting issues that can arise if one attempts to shower  $Z$ +parton and  $Z$ +2-parton events

#### 4.4.1 Matrix elements with parton showers (MEPS)

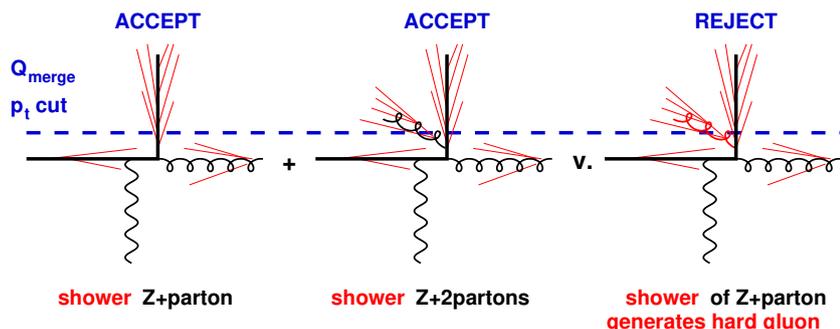
Suppose you ask for  $Z$ +jet production as the initial hard process in PYTHIA or HERWIG. As we saw above, these programs contain the correct matrix element (ME) for  $Z$ +parton production, but do a very bad job of estimating  $Z$ +2 jet production.

One naive solution to this problem would be to generate  $Z$ +2-parton events with ALPGEN, MADGRAPH, or some other preferred LO ME tool and then ask HERWIG or PYTHIA to shower those configurations. However, if one showers both  $Z$ +parton and  $Z$ +2-parton events, then one is faced with a double-counting issue, as illustrated in Fig. 30. In some events (left) the showering off the  $Z$ +parton configuration just leads to soft and collinear emissions. Similarly off the  $Z$ +2-parton events (middle). However, sometimes (right) the showering off the  $Z$ +parton configuration leads to the production of a relatively hard, large-angle gluon that is in the same phase-space region that is already covered in the  $Z$ +2-parton event sample.

Two main methods exist to avoid this double counting: CKKW matching [94] and MLM matching [95]. The latter, named after its inventor, M. L. Mangano, is the one we will describe here (it is the simpler of the two). Let's examine the basics of how it proceeds:

- Introduce a (dimensionful) transverse momentum cutoff scale  $Q_{ME}$  and a (dimensionless) angular cutoff scale  $R_{ME}$  for matrix element generation.
- Generate tree-level events for  $Z$ +1-parton,  $Z$ +2-partons, ... up to  $Z$ + $N$ -partons, where all partons must have  $p_t > Q_{ME}$  and be separated from other partons by an angle greater than  $R_{ME}$  (we will discuss the definition of this 'angle' later in Section 5). The numbers of events that one generates in the different samples are in proportion to their cross sections with these cuts.
- For each tree-level event from these samples, say one with  $n$  partons, shower it with your favourite parton-shower program.
- Apply a jet algorithm to the showered event (choose the algorithm's angular reach  $R$  to be  $\gtrsim R_{ME}$ ) and identify all jets with  $p_t > Q_{\text{merge}}$ , where the merging scale  $Q_{\text{merge}}$  is to be taken  $\gtrsim Q_{ME}$ .
- If each jet corresponds to one of the partons (i.e., is nearby in angle) and there are no extra jets above scale  $Q_{\text{merge}}$ , then accept the event. (For the sample with  $n = N$ , the condition is that there should be no extra jets with  $p_t > p_{tN}$ .)
- Otherwise, reject the event.

The action of the MLM procedure on the events of Fig. 30 is illustrated in Fig. 31, showing which events would be accepted and which ones rejected. One immediately sees how the double-counting issue disappears: in the rightmost event, the showering of a  $Z$ +parton event leads to an extra jet above  $Q_{\text{merge}}$ ; since this event now has more jets above  $Q_{\text{merge}}$  than it had partons, it is rejected. In contrast the middle



**Fig. 31:** Illustration of the application of the MLM matching procedure to the events of Fig. 30, with the  $Q_{\text{merge}}$   $p_t$  cutoff represented by the dashed line.

event, which also has two jets above  $Q_{\text{merge}}$ , was generated from a  $Z+2$ -parton event and is accepted. So is the leftmost event with only one jet, starting from a  $Z+1$ -parton event.

By providing a remedy for the double-counting issue, ensuring that the hard jets always come just from the matrix element, the MLM procedure also ensures that hard jets above  $Q_{\text{merge}}$  have distributions given by the tree-level matrix-element calculations.

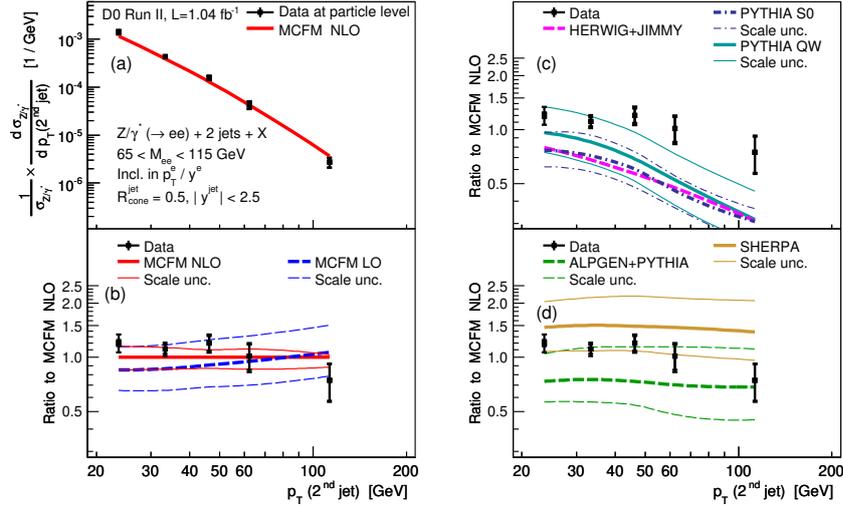
The rejection of extra jets also plays another important role: when there are big differences in scales between the jets and  $Q_{\text{merge}}$  (or between different jets), the Monte Carlo showering would want to ‘fill up’ that phase space with emissions. However, whenever it does so, the event gets rejected by the matching procedure. As long as the Monte Carlo is carrying out a reasonable showering of these multi-parton events,<sup>8</sup> then the procedure is equivalent to the introduction of a Sudakov form factor encoding the probability of not emitting radiation.

The above ‘MLM’ merging of matrix-elements and parton showers (MEPS) is the main procedure available with ALPGEN, for use with both HERWIG and PYTHIA. It is also provided (in a variant form) within MADGRAPH. The SHERPA Monte Carlo also has its own matrix-element generator(s) and provides ‘CKKW’ MEPS matching [94], which instead of the veto steps of MLM matching, uses an analytical calculation of the Sudakov form factors. These and other matrix-element/parton-shower merging schemes are discussed in detail in Ref. [95]. They all share the feature of a matching scale to separate the region under the ‘responsibility’ of matrix elements and that delegated to the parton shower. In all cases physics predictions well above the matching scale should be independent of the scale. Additionally, distributions at scales around the matching scale should be reasonably smooth, as long as the matching scale has been chosen in a region where the parton shower can be expected to give a reasonable description of emission (for caveats, see Ref. [96]; for a method that avoids the need for a matching scale, see Ref. [97]).

MEPS predictions (as well as other predictive methods) are compared to experimental results for  $Z+2$ -jet production in Fig. 32 (bottom right). The MEPS results show good agreement for the shape of the observable in question, the  $p_t$  distribution of the second jet, and they are much more successful than plain parton-shower predictions. Since their normalizations are based on LO calculations, they do, however, suffer from substantial scale dependence, which is much larger than the scale uncertainty one would obtain at NLO (bottom left).

As a result of their considerable success in describing the shapes of experimental distributions, MEPS predictions have become one of the main tools for a broad range of Tevatron and LHC analyses, especially those involving complex final states.

<sup>8</sup>This can depend on subtleties of how the Monte Carlo showers multi-parton events and the communication of information on colour flows between the fixed-order program and the Monte Carlo.



**Fig. 32:** Cross section for  $Z+2$ -jet events at the Tevatron, differential in the transverse momentum of the second jet, as measured by DØ [98]. Comparisons are shown to LO and NLO predictions (from MCFM), parton-shower predictions (PYTHIA, HERWIG) and merged matrix-element + parton-shower predictions (ALPGEN+PYTHIA, SHERPA).

#### 4.4.2 Parton showers and NLO accuracy

We’ve seen how to obtain a parton shower structure with LO accuracy for a range of different jet multiplicities. However, given the large uncertainty on the normalizations of the predictions, there would be considerable advantages to obtaining NLO accuracy. One might think that since we’ve had NLO predictions and parton shower predictions for a couple of decades, such a task should not be too hard. The main difficulty comes from the fact that NLO predictions involve divergent event weights (cf. Fig. 24), which aren’t even positive definite. Two approaches are in use to get around this problem, the MC@NLO [99] and POWHEG [100] methods.

The idea behind the MC@NLO approach is to ‘expand’ the Monte Carlo parton shower to first order in  $\alpha_s$ . I.e., the Monte Carlo’s parton showers already contain some (partially wrong) estimate of the true NLO corrections and the aim is to figure out what that estimate is. This requires a deep understanding of the Monte Carlo program. As a next step, one calculates the difference between the true NLO contributions and the partial ones included in the Monte Carlo. One of the keys to the MC@NLO method is that as long as the Monte Carlo gives the correct pattern of soft and collinear splitting (which it is supposed to), then the differences between true NLO and Monte Carlo partial NLO should be *finite*. Then one can generate partonic configurations with phase-space distributions proportional to those finite differences and shower them.

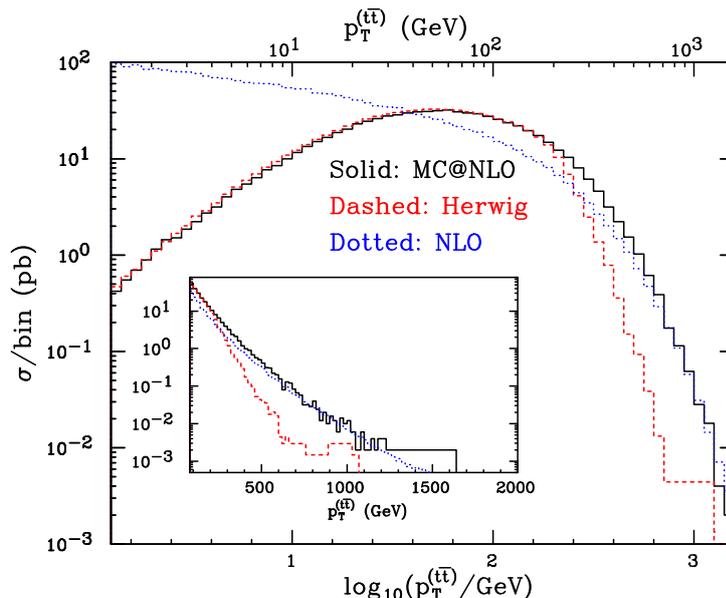
Symbolically, if we imagine a problem with one phase-space variable, say energy  $E$ , then we can write the ‘expansion’ of the Monte Carlo cross section as

$$\sigma^{MC} = 1 \times \delta(E) + \alpha_s \sigma_{1R}^{MC}(E) + \alpha_s \sigma_{1V}^{MC} \delta(E) + \mathcal{O}(\alpha_s^2) \quad (67)$$

where  $\sigma_{1R}^{MC}(E)$  is the coefficient of  $\alpha_s$  for real emission of a gluon with energy  $E$  in the Monte Carlo and  $\sigma_{1V}^{MC}$  is the (divergent) coefficient of the virtual corrections. The MC@NLO prediction is then given by

$$\text{MC@NLO} = \text{MC} \times \left( 1 + \alpha_s (\sigma_{1V} - \sigma_{1V}^{MC}) + \alpha_s \int dE (\sigma_{1R}(E) - \sigma_{1R}^{MC}(E)) \right), \quad (68)$$

where  $\sigma_{1R}(E)$  and  $\sigma_{1V}$  are the true NLO real and virtual coefficients. Each term in (small) brackets in Eq. (68) should separately be finite and corresponds to a given topology of event to be showered: a LO



**Fig. 33:** The transverse-momentum distribution of  $t\bar{t}$  pairs in the MC@NLO approach, compared to the plain HERWIG result (rescaled by the  $\sigma_{\text{NLO}}/\sigma_{\text{LO}}$   $K$ -factor) and to the NLO calculation. Shown for a 14 TeV LHC. Figure taken from Ref. [105].

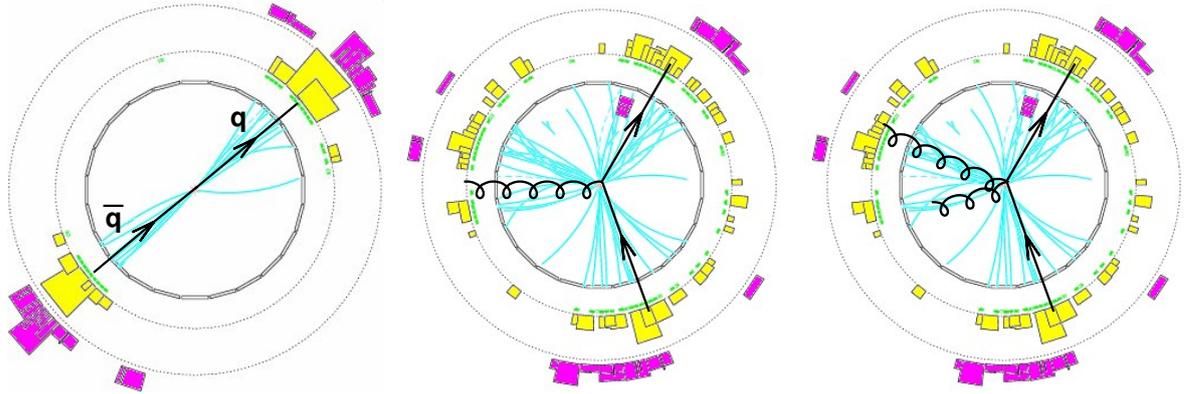
topology for the “1” and  $(\sigma_{1V} - \sigma_{1V}^{MC})$  terms, and a NLO real topology for the  $(\sigma_{1R}(E) - \sigma_{1R}^{MC}(E))$  term. (A more complete and clearer ‘toy’ explanation is given in the original MC@NLO paper [99], which makes for very instructive reading.)

The MC@NLO approach has the practical characteristic that all event weights are  $\pm 1$ . Quite a range of processes are available in the MC@NLO approach for HERWIG, including the production of a Higgs boson, single and double vector-bosons, a heavy-quark pair, various single-top processes, and  $H + W$  and  $H + Z$ . The characteristic of these processes is that they are nearly all free of light jets at LO (except for one of the single-top processes), because this simplifies the set of divergences that need to be dealt with. Very recently one PYTHIA [101] and a couple of HERWIG++ processes (Ref. [102] and references therein) were also interfaced within the MC@NLO approach.<sup>9</sup>

An alternative to MC@NLO is POWHEG [100]. It aims to avoid the (small fraction of) negative weights that are present in MC@NLO and also seeks to be less tightly interconnected with a specific Monte Carlo program. The principle of the POWHEG approach is to write a simplified Monte Carlo that generates just one emission beyond LO. This single-emission Monte Carlo is designed in such a way as to be sufficient to give the correct NLO result. It essentially works by introducing a Sudakov form factor such as Eq. (65) in which the contents of the square bracket are replaced by the integral over the exact real radiation probability above  $k_t$  (plus a constant term that accounts for the finite part of the 1-loop correction). Then emissions with transverse momenta below the scale of the first emission are left to the default Monte Carlo program, for example HERWIG or PYTHIA (implementing a transverse momentum veto to ensure that nothing is generated above the scale of the first, POWHEG, emission). The range of processes available within POWHEG is gradually catching up with that for MC@NLO, and it is hoped that this will be helped by the availability of systematic tools to help with the development of new processes [104].

An illustration of a result with MC@NLO is given in Fig. 33 for the transverse-momentum distribution of a  $t\bar{t}$  pair. It illustrates how MC@NLO reproduces the HERWIG shape in the low- $p_t$  region

<sup>9</sup>In the time since the original version of these lectures were written up, many more processes have been implemented for HERWIG++ [103].



**Fig. 34:** Left: an  $e^+e^-$  event that can be interpreted as having a 2-jet,  $q\bar{q}$ -like structure; middle: an event that can be interpreted as having a 3-jet,  $q\bar{q}g$ , structure; right: the same event reinterpreted as having a 4-jet structure,  $q\bar{q}gg$ .

(with NLO normalization), the NLO distribution at high  $p_t$ , and that neither HERWIG nor plain NLO are able to properly describe both regions.

#### 4.5 Summary

In this section, we have seen quite a range of different predictive methods for QCD at hadron colliders. Two big classes of predictive methods exist: partonic fixed-order calculations, which have well controlled accuracy, but wildly fluctuating positive and negative event weights; and Monte Carlo parton shower tools, which give a much more complete description of events (and uniform event weights).

Development is active on both sets of tools individually. On one hand we've mentioned the challenge of having a broader range of processes known at NLO and a handful at NNLO. And though we have not really touched on it, there is also a very active programme to develop parton shower Monte Carols in C++ as replacements for venerable but ageing Fortran codes like PYTHIA 6.4 and HERWIG 6.5.

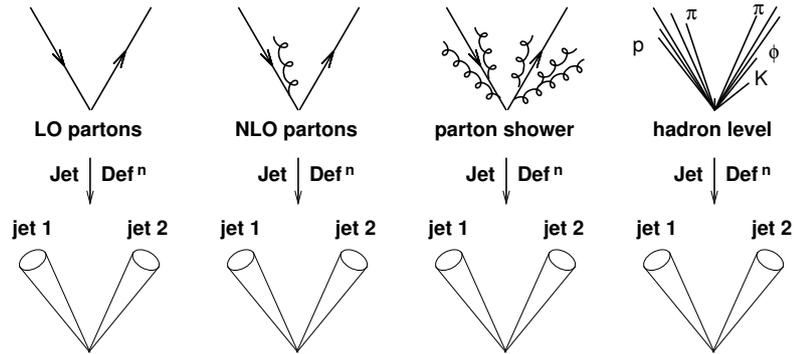
In addition there are methods with the advantages of both fixed-order and parton-shower programs. It is widespread nowadays to merge different LO 'tree-level' predictions (e.g.,  $Z$ +parton,  $Z$ +2-partons,  $Z$ +3-partons, etc.) together with parton showers. And for simple processes, those with at most one light parton in the final state, it is possible to combine NLO accuracy with the full parton-shower description. Ultimately the hope is to be able to combine  $Z$ +jet,  $Z$ +2-jets,  $Z$ +3-jets all at NLO accuracy also merging them with parton showers, so as to obtain accurate descriptions for the whole range of processes that are relevant at the LHC, both as backgrounds and as signals of new particles and new physics.

## 5 Jets

The concept of a jet has already arisen in various contexts, so in this final section we will examine jet-related ideas in more detail.

Consider the three events of Fig. 34. In the left-hand one, one interpretation is that we're seeing an  $e^+e^- \rightarrow q\bar{q}$  event, in which there has been soft and collinear showering followed by a transition to hadrons. This is a classic picture of a '2-jet' event. The middle event is more complex: energy flow is not limited to two cones. One interpretation of the event is that a  $q\bar{q}$  pair has emitted a hard gluon  $g$ , and all three have undergone soft and collinear showering. However, the same event can also be interpreted (right) as a  $q\bar{q}gg$  event, with further soft and collinear showering. Deciding between these two interpretations means choosing just how hard and separated in angle an emission has to be in order for it to be considered a separate jet (cf. the angular and energy parameters,  $\delta$  and  $\epsilon$ , in our discussion of the 2-jet cross section in Section 2.3.2).

In making this choice, it would be highly painful to visually inspect each of the  $\mathcal{O}(10^9)$  events



**Fig. 35:** The application of a jet definition to a variety of events that differ just through soft/collinear branching (and hadronization), should give identical jets in all cases.

written to disk every year at the LHC. Instead one uses a set of rules, a ‘jet definition’, by which a computer can take a list of particle momenta for an event (be they quark and gluons, or hadrons, or even calorimeter deposits), and return a list of jets. If one modifies an event just through soft and collinear emission, then the set of jets should not change, i.e., the result of applying the jet definition should be insensitive to the most common effects of showering and hadronization, as illustrated in Fig. 35.

Jets are central to collider physics: both theory and experimental results are often presented in terms of jet cross sections, and thus jets provide the meeting point between the two. As we saw in Section 4.4.1, jets are also used to assemble together different kinds of theory predictions. And jets are an input to almost all physics analyses: to new physics searches (since new particles may decay to quarks or gluons, giving jets), in Higgs searches, top physics, Monte Carlo validation, fits of PDFs, etc.

## 5.1 Jet definitions

The construction of a jet involves different considerations:

- Which particles are grouped together into a common jet? The set of rules that one follows for deciding this is usually known as a jet algorithm, and it comes with parameters that govern its exact behaviour. A common parameter is  $R$  which determines the angular reach of the jet algorithm.
- How do you combine the momenta of particles inside a jet? One needs to specify a ‘recombination scheme’. The most common is to simply add the 4-vectors of the particles (the ‘E-scheme’). This gives jets that are massive (so jets cannot be thought of as a direct stand-in for partons, which are massless).

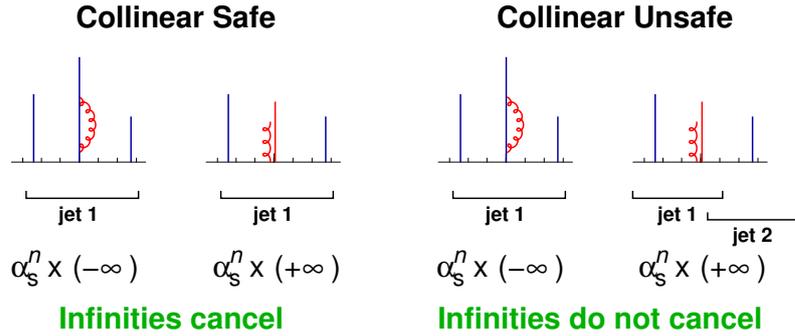
Taken together, the algorithm, its parameters and the recombination scheme specify a ‘jet definition’.

Two broad classes of jet definition are in common use: cone algorithms, which take a top-down approach, and sequential recombination algorithms, based on a bottom-up approach. Below we’ll give a brief discussion of each kind of algorithm, referring the reader to Ref. [106] for a more complete description of all the variants that exist.

## 5.2 Cone algorithms

There are many types of cone algorithm, but all rely on the idea that soft and collinear branching doesn’t modify the basic direction of energy flow.

One of the simpler cone algorithms (we’ll call it IC-PR, for iterative cone with progressive removal of particles) is that used by the CMS experiment during much of their preparation for LHC running. One first sorts all particles according to their transverse momentum, and identifies the one with largest transverse momentum. This is referred to as a seed particle,  $s$ . One draws a cone of radius  $R$  around the



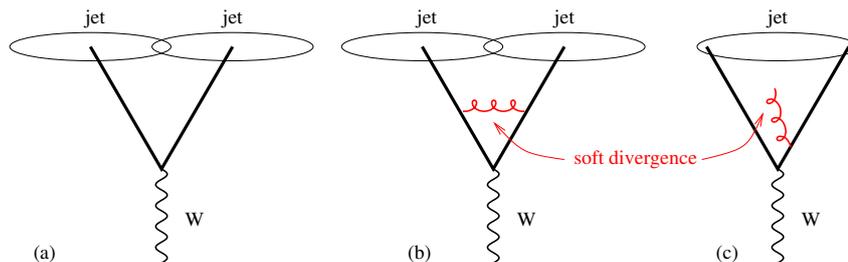
**Fig. 36:** Sample events illustrating the result of applying collinear safe and unsafe jet algorithms. The height of a given line corresponds to the particle’s transverse momentum, its horizontal position to its rapidity ( $\phi = 0$  for all particles here). Left: expectations for the behaviour of a collinear-safe jet algorithm, where the jets found by the algorithm should be independent on the collinear splitting of the hardest particles. Right: in a collinear unsafe algorithm such as the IC-PR type CMS cone, the splitting of the central hard particle causes the leftmost particle to become the hardest in the event, leading to a two-jet rather than a one-jet event.

seed — in hadron-collider variables this means identifying all particles with  $\Delta R_{is}^2 = (y_i - y_s)^2 + (\phi_i - \phi_s)^2 < R^2$ , where  $y_i = \frac{1}{2} \ln \frac{E_i + p_{zi}}{E_i - p_{zi}}$  is the rapidity of particle  $i$ ,  $\phi_i$  its azimuth, and  $y_s$  and  $\phi_s$  the rapidity and azimuth of the seed. One then identifies the direction of the sum of the momenta of those particles. If it doesn’t coincide with the seed direction then one uses that sum as a new seed direction, and iterates until the sum of the cone contents coincides with the previous seed. This is referred to as finding a stable cone. The particles inside that stable cone make a jet, and they’re removed from the list of particles in the event. The procedure is then repeated on the particles that remain, removing particles each time one finds a stable cone ( $\rightarrow$  jet), until no particles remain and one has the complete set of jets. Of these jets one retains only those above some transverse-momentum threshold  $p_{t,\min}$ .

There is one major drawback to the above procedure: the use of the particles’  $p_t$ ’s to decide which one to take as the first seed. This is problematic, because particle  $p_t$ ’s are not collinear safe quantities. As illustrated in the two right-hand events of Fig. 36, in an IC-PR algorithm, if the hardest particle undergoes a collinear splitting then this can result in another particle in the event becoming the ‘new’ hardest particle, giving a different set of final jets as compared to events without the splitting. Thus in the example of Fig. 36 there is a divergent (real, positive) contribution to the 2-jet cross section and a separate divergent (1-loop virtual, negative) contribution to the 1-jet cross section. In contrast, for a collinear-safe algorithm (two leftmost events), the collinear-splitting of the hardest particle does not change the set of final jets. Then the real and virtual divergences both contribute to the 1-jet cross section and so cancel each other.

Collinear unsafety means that certain cross sections cannot be calculated at NLO (or sometimes NNLO) — one will only obtain nonsensical infinite answers. Furthermore, even if one is working at some low perturbative order which is not divergent (e.g., LO), the fact that higher orders diverge means that the convergence of the whole perturbative series becomes questionable, compromising the usefulness even of the low orders.

Over the past two decades there has been significant discussion of such problems. There are many other variants of cone algorithm, and nearly all suffer from problems either of collinear safety, or infrared safety. One class that has been widely used at the Tevatron avoids the ordering of initial seeds, and instead obtains stable cones using all possible particles as seeds: stable cones are not immediately converted into jets, but instead, once one has the list of the stable cones found by iterating from all possible seeds one then uses a ‘split–merge’ procedure to decide how particles that appear in multiple stable cones should be unambiguously assigned to a single jet.



**Fig. 37:** Configurations illustrating IR unsafety of iterative cone algorithms with a split–merge procedure, in events with a  $W$  and two hard partons. The addition of a soft gluon converts the event from having two jets to just one jet. In contrast to Fig. 36, here the explicit angular structure is shown (rather than  $p_t$  as a function of rapidity).

This procedure avoids the collinear-safety issue (the order of particles’  $p_t$ ’s no longer matters), however, it turns out that it instead introduces an infrared-safety issue: adding an extra soft particle creates a new seed, which can lead to an extra stable cone being found, which feeds through the split–merge procedure, altering the final set of (hard) jets. This tends to happen when two hard particles are separated by a distance  $\Delta R$  that satisfies  $R < \Delta R < 2R$  (so that a cone centred on either fails to capture the other and each hard particle leads to its own jet) and one adds a soft particle in between the two (so that a cone centred on the soft particle includes both hard ones, which then end up in a single jet), as illustrated in Fig. 37. A partial fix for the problem was given in Ref. [107] (adopted by the Tevatron in Ref. [108]), which involves adding extra seeds at the midpoints between all pairs of stable cones and uses those as new starting points for finding additional stable cones before starting the split–merge step. A full fix involves a non-seed-based approach to exhaustively finding all possible stable cones in an event, in an algorithm known as the Seedless Infrared Safe Cone (SIS Cone) [109].

### 5.3 Sequential-recombination algorithms

Sequential-recombination jet algorithms take a bottom-up approach to constructing jets, as if they were inverting the sequence of splittings of the parton shower. Of course that sequence doesn’t really exist unambiguously, since gluon emission is a quantum-mechanical process involving coherent emission from all colour sources in an event. However, for collinear emissions the picture that there is a single ‘emitter’ is not a poor approximation.

#### 5.3.1 The $e^+e^- k_t$ algorithm

The most widely used sequential recombination algorithm to date is the  $k_t$  algorithm, originally formulated for  $e^+e^-$  events [110]. Recall, from Eq. (26), that the soft and collinear limit of the gluon-emission probability for  $a \rightarrow ij$  is

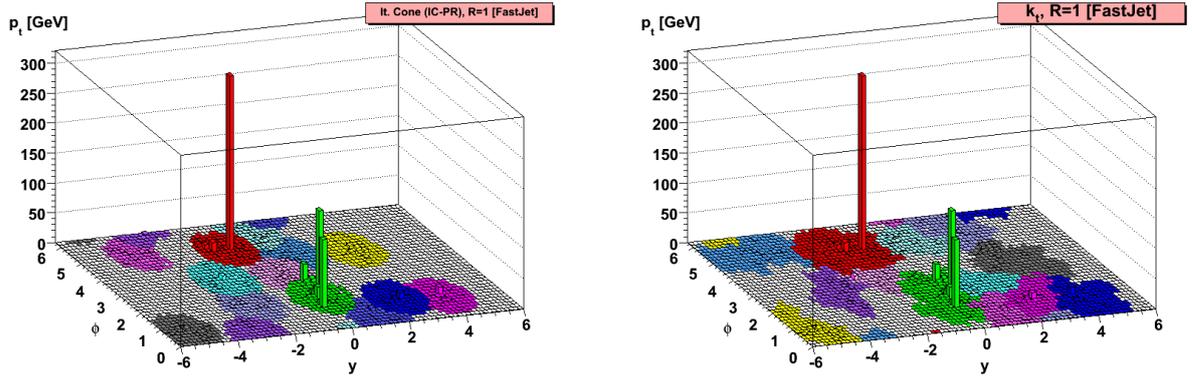
$$dS \simeq \frac{2\alpha_s C_i}{\pi} \frac{dE_i}{\min(E_i, E_j)} \frac{d\theta_{ij}}{\theta_{ij}}, \quad (69)$$

where  $C_i$  is  $C_A$  ( $C_F$ ) if  $a$  is a gluon (quark), and where we’ve written  $\min(E_i, E_j)$  in the denominator to avoid specifying which of  $i$  and  $j$  is the soft particle.

The essential idea of the  $k_t$  algorithm is to define a ‘distance measure’  $y_{ij}$  between every pair of particles  $i, j$ ,

$$y_{ij} = \frac{2 \min(E_i^2, E_j^2) (1 - \cos \theta)}{Q^2}. \quad (70)$$

In the collinear limit,  $y_{ij}$  reduces to  $\min(E_i^2, E_j^2) \theta_{ij}^2$ , which is the relative transverse momentum between particles  $i$  and  $j$  (hence the name  $k_t$  algorithm), normalized to the total visible (or sometimes centre-of-mass) energy  $Q$ . Apart from the normalization, this is just what appears in the denominator of the



**Fig. 38:** Regions of the  $y$ - $\phi$  plane covered by jets in an illustrative (simulated) hadron-collider event with an IC-PR type cone algorithm (left) and the inclusive longitudinally-invariant  $k_t$  algorithm (right). The jet finding was carried out with the FASTJET package [113, 114], with the addition of very soft ghost particles to trace the extent of the jets [115].

splitting probability, Eq. (69), so that pairs of particles that would arise from a splitting with a strong divergence are considered very close together.

The algorithm works by identifying the pair of particles that has the smallest  $y_{ij}$  and recombining them into a single particle (also called a ‘pseudojet’). It then recalculates all  $y_{ij}$  distances taking into account the new particle, and again recombines the pair that’s closest. The procedure is repeated until all  $y_{ij}$  distances are above some threshold  $y_{\text{cut}}$ , at which point the pseudojets that remain become the event’s jets.

### 5.3.2 The $k_t$ algorithm for hadron collisions

For hadron collisions and deep-inelastic scattering, the version of the  $k_t$  algorithm that is most commonly used reads as follows [111, 112]. For every pair of particles define a (dimensionful) inter-particle distance  $d_{ij}$ ,

$$d_{ij} = \min(p_{ti}^2, p_{tj}^2) \frac{\Delta R_{ij}^2}{R^2}, \quad (71)$$

where  $R$  is a parameter whose role is similar to that of  $R$  in cone algorithms. Also define a beam distance for every particle,

$$d_{iB} = p_{ti}^2. \quad (72)$$

The algorithm proceeds by searching for the smallest of the  $d_{ij}$  and the  $d_{iB}$ . If it is a  $d_{ij}$  then particles  $i$  and  $j$  are recombined into a single new particle. If it is a  $d_{iB}$  then  $i$  is removed from the list of particles, and called a jet. This is repeated until no particles remain.

Note that the distance in Eq. (71) just reduces to that of Eq. (70) in the collinear limit (modulo  $Q^2$  normalization). So one is still dealing with the relative transverse momentum between pairs of particles. As with cone algorithms, in this ‘inclusive longitudinally invariant  $k_t$  algorithm,’ arbitrarily soft particles can form jets. It is therefore standard to place a  $p_{t,\text{min}}$  cutoff on the jets one uses for ‘hard’ physics.

One can verify that  $R$  in the  $k_t$  algorithm plays a similar role to  $R$  in cone algorithms, using the following observations: if two particles  $i$  and  $j$  are within  $R$  of each other, i.e.,  $\Delta R_{ij} < R$ , then  $d_{ij} < d_{iB}, d_{jB}$  and so  $i$  and  $j$  will prefer to recombine rather than forming separate jets. If a particle  $i$  is separated by more than  $R$  from all other particles in the event then it will have  $d_{iB} < d_{ij}$  for all  $j$  and so it will form a jet on its own.

Despite this similarity to the behaviour of cone algorithms for pairs of particles, the  $k_t$  algorithm gives jets that ‘look’ somewhat different. Figure 38 illustrates what happens when one clusters a simulated hadron-collider event with an IC-PR type cone algorithm and with the  $k_t$  algorithm. In both cases the figure shows (in a given colour) the calorimeter cells that are included in each jet. For the IC-PR algorithm the hardest jets all appear circular, as expected given the use of cones in the definition of the algorithm (in cone algorithms with split–merge steps, the jets are often not circular, because of the less trivial mapping from stable cones to jets). For the  $k_t$  algorithm, the jets have quite irregular (or jagged) edges, because many of the soft particles cluster together early in the recombination sequence (owing to their small  $p_t$  and hence  $d_{ij}$ ) in patterns that are determined by the random distributions of those particles in  $p_t$  and rapidity and azimuth.

The irregularity of the jet boundaries has often been held against the  $k_t$  algorithm. One reason is that it makes it harder to calculate acceptance corrections: for example, if you know that some part of a detector is misbehaving, it is not obvious how far a  $k_t$  jet must be from that part of the detector in order not to be affected by it. Another reason relates to the linearity of the dependence of the jet momentum on soft-particle momenta: in the IC-PR algorithm, the hard core of the jet essentially determines the jet boundary, and the algorithm depends linearly on the momenta of any soft particles within the boundary, and is independent of particles outside it. In the  $k_t$  algorithm, varying the momentum of one soft particle in the vicinity of the jet core can affect whether it and other soft particles get clustered with that core or not. This can complicate energy calibrations for the jet algorithm, though techniques exist to correct for this to some extent (jet-area-based subtraction [116], which leaves just a residual term known as back-reaction [115]).

A feature of the  $k_t$  algorithm that is attractive is that it not only gives you jets, but also assigns a clustering sequence to the particles within the jet. One can therefore undo the clustering and look inside the jet. This has been exploited in a range of QCD studies (e.g., Ref. [117]), and also in discussions of searches of hadronic decays of boosted massive particles such as  $W$ ,  $H$ , or  $Z$  bosons, top quarks, or new particles (early examples include Refs. [118, 119]; for more recent examples, see the reviews in Refs. [106, 120]). Jet substructure studies are also often carried out with the Cambridge/Aachen (C/A) algorithm [121, 122], which essentially replaces  $p_{ti} \rightarrow 1, p_{tj} \rightarrow 1$  in Eqs. (71,72) but is otherwise like the  $k_t$  algorithm.

### 5.3.3 The anti- $k_t$ algorithm

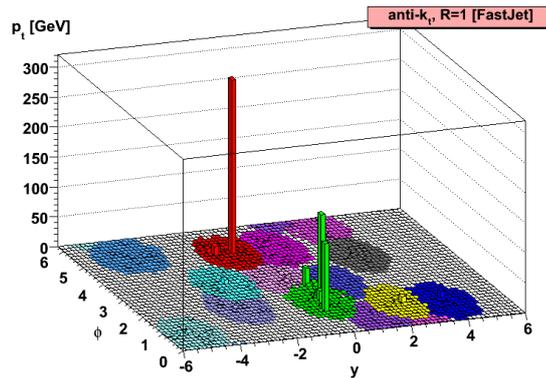
It turns out that it is possible to design a sequential-recombination algorithm with many of the nice properties of cone algorithms via a simple modification of the  $k_t$  algorithm’s distance measures [123]:

$$d_{ij} = \frac{1}{\max(p_{ti}^2, p_{tj}^2)} \frac{\Delta R_{ij}^2}{R^2}, \quad (73a)$$

$$d_{iB} = \frac{1}{p_{ti}^2}. \quad (73b)$$

The correspondence with the divergences of Eq. (69) is partially lost: objects that are close in angle still prefer to cluster early, but that clustering tends to occur with a hard particle (rather than necessarily involving soft particles). This means that jets ‘grow’ in concentric circles out from a hard core, until they reach a radius  $R$ , giving circular jets just as with the IC-PR cone, as shown in Fig. 39. However, unlike the IC-PR cone, this ‘anti- $k_t$ ’ algorithm is collinear (and infrared) safe, meaning that it is safe to use with fixed-order QCD predictions. This, combined with the fact that it has been implemented efficiently in the FASTJET jet-finding code [113, 114], has led to it being adopted as the default jet algorithm by both the ATLAS and CMS collaborations.

Note that the anti- $k_t$  algorithm does not provide useful information on jet substructure: if a jet contains two hard cores, then the  $k_t$  (or C/A) algorithms first reconstruct those hard cores and merge the



**Fig. 39:** As in Fig. 38, but shown for the anti- $k_t$  algorithm

resulting two subjects. The anti- $k_t$  algorithm will often first cluster the harder of the two cores and then gradually agglomerate the contents of the second hard core.

#### 5.4 Using jets

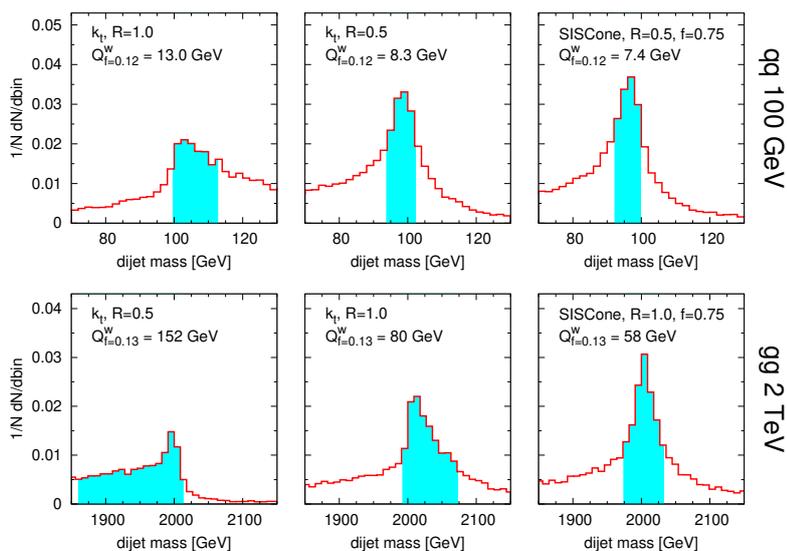
Because of the intricacies of calibrating jets, past experiments have tended to concentrate their efforts on just one or two jet definitions, which are used across the board for QCD studies, top-quark studies, Higgs and new physics searches. Typically the choices have been for cone-type algorithms with  $R$  values in the range 0.4–0.7.

At the LHC there are prospects of experimental methods (for example topoclusters at ATLAS and particle flow at ATLAS) that make it easier to use a broad range of jet definitions. We’ve already mentioned, above, the use of jet substructure in new physics searches. These methods will definitely benefit from experiments’ flexibility in choosing different jet definitions (for example, many of the proposed searches use quite a large jet radius,  $R \sim 1$ –1.5). More generally, when using jets, it is important to establish what you are trying to get out of the jet algorithm, what it is that you’re trying to optimize.

Different studies will want to optimize different things. For example in QCD studies, such as the inclusive jet-spectrum measurement that goes into PDF fits, one criterion might be to choose a jet definition that minimizes the non-perturbative corrections that need to be applied when comparing perturbative predictions with data.

In searches for new physics, for example when looking for a resonance in the dijet invariant mass spectrum, the criterion might be the following: given a narrow resonance, what jet definition will lead to the narrowest peak in the dijet mass spectrum? It turns out that the answer depends significantly on properties of the resonance you’re trying to reconstruct. Figure 40 illustrates how SIS Cone with  $R = 0.5$  does well for reconstructing a 100 GeV resonance that decays to  $q\bar{q}$ . While for a 2 TeV resonance decaying to  $gg$ , you’re better off with an algorithm (such as SIS Cone, or others supplemented with the trick of filtering and/or variants [126, 127]) with a substantially larger radius,  $R \sim 1$  [124, 125, 128, 129]. And if you find a resonance, you might then want to know which jet definition will allow you to measure its mass most accurately, which may not be the same choice that allowed you to find the resonance most easily.

Understanding analytically which jet choices work well is the subject of ongoing theoretical work [115, 130, 131], which involves figuring out how perturbative radiation, hadronization, the underlying event and pileup all come together to affect jet momenta.



**Fig. 40:** Illustration of the impact of different jet definition on the quality of reconstruction of dijet invariant mass peaks for a 100 GeV resonance decaying to  $q\bar{q}$  (top) and a 2 TeV resonance decaying to  $gg$  (bottom). The  $Q_{f=0.12}^W$  indicates the width of the region that contains 12% of the resonance events (corresponding to about 25% of the resonance events that pass basic selection cuts). Smaller numbers imply better reconstruction. Figure taken from Ref. [124]. Many further plots available from Ref. [125].

## 6 Conclusions

In these lectures we have covered material that ranges from the QCD Lagrangian through to  $e^+e^-$  scattering, soft and collinear divergences, PDFs, fixed-order calculations, parton showers and jets. It is my hope that the material here will have given you enough background information to at least understand the basics of these different concepts.

The aspects of QCD that we have examined will play a role in nearly all analyses at the LHC. Whether because Monte Carlo parton-shower programs are used for estimating detector effects on almost every measurement; or because in searches for new physics one might look for an excess with respect to QCD predictions of backgrounds; or perhaps because an understanding of QCD will make it possible to more clearly pull out a signal of new physics that involves QCD final states.

If you want to find out more about the topics discussed here (or those left out due to time constraints, notably heavy quarks), good places to look include the textbooks [1–3] mentioned at the beginning of these lectures, the references given throughout, and transparencies from the various summer schools dedicated to QCD, for example, the CTEQ (and MCNET) schools [132, 133].

## References

- [1] R. K. Ellis, W. J. Stirling and B. R. Webber, *QCD and collider physics*, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **8** (1996) 1, doi:10.1017/CBO9780511628788.
- [2] G. Dissertori, I. G. Knowles and M. Schmelling, *Quantum Chromodynamics: High-Energy Experiments and Theory*, (Clarendon, Oxford, 2002), doi:10.1093/acprof:oso/9780199566419.001.0001.
- [3] R. Brock *et al.* [CTEQ Collaboration], *Rev. Mod. Phys.* **67** (1995) 157, doi:10.1103/RevModPhys.67.157 ; see also <http://www.phys.psu.edu/~cteq/handbook/v1.1/handbook.pdf>.
- [4] J. M. Maldacena, *Adv. Theor. Math. Phys.* **2** (1998) 231, doi:10.1023/A:1026654312961,

- Int. J. Theor. Phys.* **38** (1999) 1113, doi:10.4310/ATMP.1998.v2.n2.a1, arXiv:hep-th/9711200.
- [5] S. S. Gubser, I. R. Klebanov and A. M. Polyakov, *Phys. Lett.* **B428** (1998) 105, doi:10.1016/S0370-2693(98)00377-3, arXiv:hep-th/9802109.
- [6] E. Witten, *Adv. Theor. Math. Phys.* **2** (1998) 253, doi:10.4310/ATMP.1998.v2.n2.a2, arXiv:hep-th/9802150.
- [7] J. Erlich, E. Katz, D. T. Son and M. A. Stephanov, *Phys. Rev. Lett.* **95** (2005) 261602, doi:10.1103/PhysRevLett.95.261602, arXiv:hep-ph/0501128.
- [8] L. Da Rold and A. Pomarol, *Nucl. Phys.* **B721** (2005) 79, doi:10.1016/j.nuclphysb.2005.05.009, arXiv:hep-ph/0501218.
- [9] S. Durr *et al.*, *Science* **322** (2008) 1224, doi: 10.1126/science.1163233, arXiv:0906.3599 [hep-lat].
- [10] J. Smit, *Introduction to quantum fields on a lattice: A robust mate*, *Cambridge Lect. Notes Phys.* **15** (2002) 1, doi:10.1017/CBO9780511583971.
- [11] T. DeGrand and C. E. Detar, *Lattice Methods for Quantum Chromodynamics*, (World Scientific, Singapore, 2006), doi:10.1142/6065.
- [12] Perspectives in Lattice QCD, Proceedings of the Workshop, Nara International Seminar House, Nara, Japan, 2005, Ed. Y. Kuramashi, (World Scientific, Singapore, 2007), doi:10.1142/6216.
- [13] T. Onogi, arXiv:0906.2344 [hep-ph].
- [14] M. E. Peskin and D. V. Schroeder, *An Introduction To Quantum Field Theory* (Addison-Wesley, Reading, MA, 1995).
- [15] V. N. Gribov, Lectures at the 12th Winter School of the Leningrad Nuclear Physics Institute, SLAC-TRANS-176 (1977), <https://www.slac.stanford.edu/pubs/slactrans/trans01/slac-trans-0176.pdf>; S. D. Drell, in a Festschrift for Maurice Goldhaber, *Trans. NY Acad. Sci., Series II* **40** (1980) 76, doi:10.1111/j.2164-0947.1980.tb03001.x.
- [16] N. K. Nielsen, *Am. J. Phys.* **49** (1981) 1171, doi:10.1119/1.12565; R. J. Hughes, *Phys. Lett.* **B97** (1980) 246, doi:10.1016/0370-2693(80)90593-6; *idem* *Nucl. Phys.* **B186** (1981) 376, doi:10.1016/0550-3213(81)90076-6; K. Johnson, in *Asymptotic Realms of Physics*, Eds. A. Guth, K. Huang and R. L. Jaffe (MIT Press, Cambridge, MA, 1983).
- [17] S. Bethke, *Eur. Phys. J.* **C64** (2009) 689, doi:10.1140/epjc/s10052-009-1173-1, arXiv:0908.1135 [hep-ph].
- [18] S. G. Gorishnii, A. L. Kataev and S. A. Larin, *Phys. Lett.* **B259** (1991) 144, doi:10.1016/0370-2693(91)90149-K.
- [19] L. R. Surguladze and M. A. Samuel, *Phys. Rev. Lett.* **66**, (1991) 560, doi:10.1103/PhysRevLett.66.560, [Erratum-*ibid.* **66**, 2416 (1991)].
- [20] T. Hebbeker, M. Martinez, G. Passarino and G. Quast, *Phys. Lett.* **B331** (1994) 165, doi:10.1016/0370-2693(94)90958-X.
- [21] K. G. Chetyrkin, J. H. Kuhn and A. Kwiatkowski, *Phys. Rep.* **277** (1996) 189, doi:10.1016/S0370-1573(96)00012-9.
- [22] P. A. Baikov, K. G. Chetyrkin and J. H. Kuhn, *Phys. Rev. Lett.* **101** (2008) 012002, doi:10.1103/PhysRevLett.101.012002, arXiv:0801.1821 [hep-ph].
- [23] P. A. Baikov, K. G. Chetyrkin and J. H. Kuhn, *Nucl. Phys. B Proc. Suppl.* **189** (2009), 49, doi:10.1016/j.nuclphysbps.2009.03.010, arXiv:0906.2987 [hep-ph].
- [24] V. S. Fadin, *Yad. Fiz.* **37** (1983) 408; B. I. Ermolaev and V. S. Fadin, *JETP Lett.* **33** (1981) 269 [*Pisma Zh. Eksp. Teor. Fiz.* **33** (1981) 285], [http://www.jetpletters.ac.ru/ps/1505/article\\_23008.shtml](http://www.jetpletters.ac.ru/ps/1505/article_23008.shtml); A. H. Mueller, *Phys. Lett.* **B104**, (1981) 161, doi:10.1016/0370-2693(81)90581-5;

- Y. L. Dokshitzer, V. S. Fadin and V. A. Khoze, *Z. Phys.* **C15** (1982) 325, doi:10.1007/BF01614423;
- A. Bassetto, M. Ciafaloni and G. Marchesini, *Phys. Rept.* **100** (1983) 201, doi:10.1016/0370-1573(83)90083-2.
- [25] Y. L. Dokshitzer, V. A. Khoze, A. H. Mueller and S. I. Troyan, *Basics of perturbative QCD*, (Editions Frontières, Gif-sur-Yvette, 1991), <http://www.lpthe.jussieu.fr/~yuri/BPQCD/cover.html>.
- [26] A. H. Mueller, *Nucl. Phys.* **B213** (1983) 85, doi:10.1016/0550-3213(83)90176-1.
- [27] G. Sterman and S. Weinberg, *Phys. Rev. Lett.* **39** (1977) 1436, doi:10.1103/PhysRevLett.39.1436.
- [28] J. M. Zanotti, *PoS LC2008* (2008) 051, doi:10.22323/1.061.0051.
- [29] M. Arneodo *et al.* [New Muon Collaboration], *Nucl. Phys.* **B483** (1997) 3, doi:10.1016/S0550-3213(96)00538-X, arXiv:hep-ph/9610231.
- [30] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky and W. K. Tung, *JHEP* **07** (2002) 012, doi:10.1088/1126-6708/2002/07/012, arXiv:hep-ph/0201195.
- [31] S. Chekanov *et al.* [ZEUS Collaboration], *Eur. Phys. J.* **C21** (2001) 443, doi:10.1007/s100520100749 arXiv:hep-ex/0105090.
- [32] W. K. Tung, H. L. Lai, A. Belyaev, J. Pumplin, D. Stump and C. P. Yuan, *JHEP* **02** (2007) 053, doi:10.1088/1126-6708/2007/02/053, arXiv:hep-ph/0611254.
- [33] A. D. Martin, R. G. Roberts, W. J. Stirling and R. S. Thorne, *Phys. Lett.* **B604** (2004) 61, doi:10.1016/j.physletb.2004.10.040, arXiv:hep-ph/0410230.
- [34] V. M. Abazov *et al.* [D0 Collaboration], *Phys. Rev. Lett.* **101** (2008) 062001, doi:10.1103/PhysRevLett.101.062001, arXiv:0802.2400 [hep-ex].
- [35] A. Abulencia *et al.* [CDF - Run II Collaboration], *Phys. Rev.* **D75** (2007) 092006, doi:10.1103/PhysRevD.75.092006, [Erratum-*ibid.* **D75** (2007) 119901], arXiv:hep-ex/0701051.
- [36] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, *Eur. Phys. J.* **C64** (2009) 653, doi:10.1140/epjc/s10052-009-1164-2, arXiv:0905.3531 [hep-ph].
- [37] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, *Eur. Phys. J.* **C63** (2009) 189, doi:10.1140/epjc/s10052-009-1072-5, arXiv:0901.0002 [hep-ph].
- [38] P. M. Nadolsky *et al.*, *Phys. Rev.* **D78** (2008) 013004, doi:10.1103/PhysRevD.78.013004, arXiv:0802.0007 [hep-ph].
- [39] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo and M. Ubiali, *Nucl. Phys.* **B838** (2010) 136, doi:10.1016/j.nuclphysb.2010.05.008, arXiv:1002.4407 [hep-ph].
- [40] G. P. Salam and J. Rojo, *Comput. Phys. Commun.* **180** (2009) 120, doi:10.1016/j.cpc.2008.08.010, arXiv:0804.3755 [hep-ph].
- [41] W. Furmanski and R. Petronzio, *Phys. Lett.* **B97** (1980) 437, doi:10.1016/0370-2693(80)90636-X.
- [42] G. Curci, W. Furmanski and R. Petronzio, *Nucl. Phys.* **B175** (1980) 27, doi:10.1016/0550-3213(80)90003-6.
- [43] A. Vogt, S. Moch and J. A. M. Vermaseren, *Nucl. Phys.* **B691** (2004) 129, doi:10.1016/j.nuclphysb.2004.04.024, arXiv:hep-ph/0404111.
- [44] S. Moch, J. A. M. Vermaseren and A. Vogt, *Nucl. Phys.* **B688** (2004) 101, doi:10.1016/j.nuclphysb.2004.03.030, arXiv:hep-ph/0403192.
- [45] LHAPDF, *LHAPDF Documentation*, <http://projects.hepforge.org/lhapdf/>.
- [46] M. L. Mangano, *Eur. Phys. J.* **C59** (2009) 373, doi:10.1140/epjc/s10052-008-0757-5, arXiv:0809.1567 [hep-ph].
- [47] C. Anastasiou, L. J. Dixon, K. Melnikov and F. Petriello, *Phys. Rev.* **D69** (2004) 094008, doi:10.1103/PhysRevD.69.094008, arXiv:hep-ph/0312266.

- [48] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason and G. Ridolfi, *JHEP* **04** (2004) 068, doi:10.1088/1126-6708/2004/04/068, arXiv:hep-ph/0303085.
- [49] R. Hamberg, W. L. van Neerven and T. Matsuura, *Nucl. Phys.* **B359** (1991) 343, doi:10.1016/0550-3213(91)90064-5, [Erratum-*ibid.* **B644** (2002) 403].
- [50] R. V. Harlander and W. B. Kilgore, *Phys. Rev. Lett.* **88** (2002) 201801, doi:10.1103/PhysRevLett.88.201801, arXiv:hep-ph/0201206.
- [51] C. Anastasiou and K. Melnikov, *Nucl. Phys.* **B646** (2002) 220, doi:10.1016/S0550-3213(02)00837-4, arXiv:hep-ph/0207004.
- [52] V. Ravindran, J. Smith and W. L. van Neerven, *Nucl. Phys.* **B665** (2003) 325, doi:10.1016/S0550-3213(03)00457-7, arXiv:hep-ph/0302135.
- [53] R. V. Harlander and K. J. Ozeren, *JHEP* **11** (2009) 088, doi:10.1088/1126-6708/2009/11/088, arXiv:0909.3420 [hep-ph].
- [54] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa, *JHEP* **07** (2003) 001, doi:10.1088/1126-6708/2003/07/001, arXiv:hep-ph/0206293; *ALPGEN V2.14*, <http://cern.ch/mlm/alpgen/>.
- [55] T. Gleisberg and S. Hoche, *JHEP* **12** (2008) 039, doi:10.1088/1126-6708/2008/12/039, arXiv:0808.3674 [hep-ph].
- [56] E. Boos *et al.* [CompHEP Collaboration], *Nucl. Instrum. Meth.* **A534** (2004) 250, doi:10.1016/j.nima.2004.07.096, arXiv:hep-ph/0403113; *CompHEP homepage*, <http://comphep.sinp.msu.ru/>.
- [57] A. Cafarella, C. G. Papadopoulos and M. Worek, *Comput. Phys. Commun.* **180** (2009) 1941, doi:10.1016/j.cpc.2009.04.023, arXiv:0710.2427 [hep-ph]; *HELAC-NLO and associated tools*, <http://cern.ch/helac-phegas/>.
- [58] J. Alwall *et al.*, *JHEP* **09** (2007) 028, doi:10.1088/1126-6708/2007/09/028, arXiv:0706.2334 [hep-ph]; *The MadGraph5\_aMC at NLO homepage*, <http://madgraph.phys.ucl.ac.be/>.
- [59] E. W. N. Glover, Precision phenomenology and collider physics, talk given at Computer Algebra and Particle Physics Workshop, DESY Zeuthen, April 2005, <http://www-zeuthen.desy.de/theory/capp2005/>.
- [60] F. A. Berends and W. T. Giele, *Nucl. Phys.* **B306** (1988) 759, doi:10.1016/0550-3213(88)90442-7.
- [61] L. J. Dixon, arXiv:hep-ph/9601359.
- [62] S. Catani and M. H. Seymour, *Nucl. Phys.* **B485** (1997) 291, doi:10.1016/S0550-3213(96)00589-5, [Erratum-*ibid.* **B510** (1998) 503], arXiv:hep-ph/9605323.
- [63] S. Frixione, Z. Kunszt and A. Signer, *Nucl. Phys.* **B467** (1996) 399, doi:10.1016/0550-3213(96)00110-1, arXiv:hep-ph/9512328.
- [64] D. A. Kosower, *Phys. Rev.* **D57** (1998) 5410, doi:10.1103/PhysRevD.57.5410, arXiv:hep-ph/9710213; J. M. Campbell, M. A. Cullen and E. W. N. Glover, *Eur. Phys. J.* **C9** (1999) 245, doi:10.1007/s100529900034, arXiv:hep-ph/9809429; D. A. Kosower, *Phys. Rev.* **D71** (2005) 045016, doi:10.1103/PhysRevD.71.045016, arXiv:hep-ph/0311272.
- [65] Z. Nagy, *Phys. Rev.* **D68** (2003) 094002, doi:10.1103/PhysRevD.68.094002, arXiv:hep-ph/0307268; *NLOJet++*, <https://www.desy.de/~znagy/Site/NLOJet++.html>.
- [66] J. M. Campbell and R. K. Ellis, *Phys. Rev.* **D62** (2000) 114012, doi:10.1103/PhysRevD.62.114012, arXiv:hep-ph/0006304; *MCFM - Monte Carlo for FeMtobarn processes*, <http://mcfm.fnal.gov/>.

- [67] K. Arnold *et al.*, *Comput. Phys. Commun.* **180** (2009) 1661, doi:10.1016/j.cpc.2009.03.006, arXiv:0811.4559 [hep-ph]. *VBFNLO - A parton level Monte Carlo for processes with electroweak bosons*, <https://www.itp.kit.edu/vbfnlo/wiki/doku.php>.
- [68] T. Binoth, J. P. Guillet, E. Pilon and M. Werlen, *Eur. Phys. J.* **C16** (2000) 311, doi:10.1007/s100520050024, arXiv:hep-ph/9911340; *The PHOX Family*, [https://lapth.cnrs.fr/PHOX\\_FAMILY/main.html](https://lapth.cnrs.fr/PHOX_FAMILY/main.html).
- [69] A. Bredenstein, A. Denner, S. Dittmaier and S. Pozzorini, *Phys. Rev. Lett.* **103** (2009) 012002, doi:10.1103/PhysRevLett.103.012002, arXiv:0905.0110 [hep-ph].
- [70] G. Bevilacqua, M. Czakon, C. G. Papadopoulos, R. Pittau and M. Worek, *JHEP* **09** (2009) 109, doi:10.1088/1126-6708/2009/09/109, arXiv:0907.4723 [hep-ph].
- [71] G. Bevilacqua, M. Czakon, C. G. Papadopoulos and M. Worek, *Phys. Rev. Lett.* **104** (2010) 162002, doi:10.1103/PhysRevLett.104.162002, arXiv:1002.4009 [hep-ph].
- [72] C. F. Berger *et al.*, *Phys. Rev.* **D80** (2009) 074036, doi:10.1103/PhysRevD.80.074036, arXiv:0907.1984 [hep-ph].
- [73] R. Keith Ellis, K. Melnikov and G. Zanderighi, *Phys. Rev.* **D80** (2009) 094002, doi:10.1103/PhysRevD.80.094002, arXiv:0906.1445 [hep-ph].
- [74] C. F. Berger *et al.*, *Phys. Rev.* **D82** (2010) 074002, doi:10.1103/PhysRevD.82.074002, arXiv:1004.1659 [hep-ph].
- [75] C. F. Berger *et al.*, *Phys. Rev. Lett.* **106** (2011) 092001, doi:10.1103/PhysRevLett.106.092001, arXiv:1009.2338 [hep-ph].
- [76] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover and G. Heinrich, *JHEP* **12** (2007) 094, doi:10.1088/1126-6708/2007/12/094, arXiv:0711.4711 [hep-ph].
- [77] S. Weinzierl, *Phys. Rev. Lett.* **101** (2008) 162001, doi:10.1103/PhysRevLett.101.162001, arXiv:0807.3241 [hep-ph]; *JHEP* **06** (2009) 041, doi:10.1088/1126-6708/2009/06/041, arXiv:0904.1077 [hep-ph].
- [78] K. Melnikov and F. Petriello, *Phys. Rev.* **D74** (2006) 114017, doi:10.1103/PhysRevD.74.114017, arXiv:hep-ph/0609070.
- [79] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Phys. Rev. Lett.* **103** (2009) 082001, doi:10.1103/PhysRevLett.103.082001, arXiv:0903.2120 [hep-ph].
- [80] C. Anastasiou, K. Melnikov and F. Petriello, *Nucl. Phys.* **B724** (2005) 197, doi:10.1016/j.nuclphysb.2005.06.03, arXiv:hep-ph/0501130.
- [81] S. Catani and M. Grazzini, *Phys. Rev. Lett.* **98** (2007) 222002, doi:10.1103/PhysRevLett.98.222002, arXiv:hep-ph/0703012.
- [82] T. Sjostrand, S. Mrenna and P. Skands, *Comput. Phys. Commun.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820 [hep-ph]; *Present program version: PYTHIA 8.2*, <http://home.thep.lu.se/~torbjorn/Pythia.html>.
- [83] T. Sjostrand, S. Mrenna and P. Skands, *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175; *PYTHIA 6.4*, <https://pythiasix.hepforge.org>.
- [84] L. Lonnblad, *Comput. Phys. Commun.* **71** (1992) 15, doi:10.1016/0010-4655(92)90068-A.
- [85] T. Gleisberg, S. Hoche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert and J. Winter, *JHEP* **02** (2009) 007, doi:10.1088/1126-6708/2009/02/007, arXiv:0811.4622 [hep-ph]; *Sherpa Homepage*, <https://sherpa-team.gitlab.io/>.
- [86] T. Sjostrand, P. Eden, C. Friberg, L. Lonnblad, G. Miu, S. Mrenna and E. Norrbin, *Comput. Phys. Commun.* **135** (2001) 238, doi:10.1016/S0010-4655(00)00236-8, arXiv:hep-ph/0010017.
- [87] G. Corcella *et al.*, *JHEP* **01** (2001) 010, doi:10.1088/1126-6708/2001/01/010, arXiv:hep-ph/0011363; *The Herwig Event Generator*, <https://herwig.hepforge.org/>.

- [88] M. Bahr *et al.*, *Eur. Phys. J.* **C58** (2008) 639, doi:[10.1140/epjc/s10052-008-0798-9](https://doi.org/10.1140/epjc/s10052-008-0798-9), arXiv:[0803.0883](https://arxiv.org/abs/0803.0883) [hep-ph]; *The Herwig Event Generator*, <http://projects.hepforge.org/herwig/>.
- [89] B. Andersson, G. Gustafson, G. Ingelman and T. Sjostrand, *Phys. Rep.* **97** (1983) 31, doi:[10.1016/0370-1573\(83\)90080-7](https://doi.org/10.1016/0370-1573(83)90080-7).
- [90] T. Sjostrand, *Nucl. Phys.* **B248** (1984) 469, doi:[10.1016/0550-3213\(84\)90607-2](https://doi.org/10.1016/0550-3213(84)90607-2).
- [91] B. R. Webber, *Nucl. Phys.* **B238** (1984) 492, doi:[10.1016/0550-3213\(84\)90333-X](https://doi.org/10.1016/0550-3213(84)90333-X).
- [92] A. Buckley, H. Hoeth, H. Lacker, H. Schulz and J. E. von Seggern, *Eur. Phys. J.* **C65** (2010) 331, doi:[10.1140/epjc/s10052-009-1196-7](https://doi.org/10.1140/epjc/s10052-009-1196-7), arXiv:[0907.2973](https://arxiv.org/abs/0907.2973) [hep-ph].
- [93] T. Sjostrand, Monte Carlo Generators, 2006 European School of High-Energy Physics, Aronsborg, Sweden, Ed. R. Fleischer, CERN-2007-005 (CERN, Geneva, Switzerland, 2007), p. 51, doi:[10.5170/CERN-2007-005.51](https://doi.org/10.5170/CERN-2007-005.51), arXiv:[hep-ph/0611247](https://arxiv.org/abs/hep-ph/0611247).
- [94] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, *JHEP* **11** (2001) 063, doi:[10.1088/1126-6708/2001/11/063](https://doi.org/10.1088/1126-6708/2001/11/063), arXiv:[hep-ph/0109231](https://arxiv.org/abs/hep-ph/0109231).
- [95] J. Alwall *et al.*, *Eur. Phys. J.* **C53** (2008) 473, doi:[10.1140/epjc/s10052-007-0490-5](https://doi.org/10.1140/epjc/s10052-007-0490-5), arXiv:[0706.2569](https://arxiv.org/abs/0706.2569) [hep-ph].
- [96] N. Lavesson and L. Lonnblad, *JHEP* **04** (2008) 085, doi:[10.1088/1126-6708/2008/04/085](https://doi.org/10.1088/1126-6708/2008/04/085), arXiv:[0712.2966](https://arxiv.org/abs/0712.2966) [hep-ph].
- [97] W. T. Giele, D. A. Kosower and P. Z. Skands, *Phys. Rev.* **D78** (2008) 014026, doi:[10.1103/PhysRevD.78.014026](https://doi.org/10.1103/PhysRevD.78.014026), arXiv:[0707.3652](https://arxiv.org/abs/0707.3652) [hep-ph].
- [98] V. M. Abazov *et al.* [D0 Collaboration], *Phys. Lett.* **B678** (2009) 45, doi:[10.1016/j.physletb.2009.05.058](https://doi.org/10.1016/j.physletb.2009.05.058), arXiv:[0903.1748](https://arxiv.org/abs/0903.1748) [hep-ex].
- [99] S. Frixione and B. R. Webber, *JHEP* **06** (2002) 029, doi:[10.1088/1126-6708/2002/06/029](https://doi.org/10.1088/1126-6708/2002/06/029), arXiv:[hep-ph/0204244](https://arxiv.org/abs/hep-ph/0204244).
- [100] P. Nason, *JHEP* **11** (2004) 040, doi:[10.1088/1126-6708/2004/11/040](https://doi.org/10.1088/1126-6708/2004/11/040), arXiv:[hep-ph/0409146](https://arxiv.org/abs/hep-ph/0409146).
- [101] P. Torrielli and S. Frixione, *JHEP* **04** (2010) 110, doi:[10.1007/JHEP04\(2010\)110](https://doi.org/10.1007/JHEP04(2010)110), arXiv:[1002.4293](https://arxiv.org/abs/1002.4293) [hep-ph].
- [102] O. Latunde-Dada, *JHEP* **05** (2009) 112, doi:[10.1088/1126-6708/2009/05/112](https://doi.org/10.1088/1126-6708/2009/05/112), arXiv:[0903.4135](https://arxiv.org/abs/0903.4135) [hep-ph].
- [103] S. Frixione, F. Stoeckli, P. Torrielli and B. R. Webber, *JHEP* **01** (2011) 053, doi:[10.1007/JHEP01\(2011\)053](https://doi.org/10.1007/JHEP01(2011)053), arXiv:[1010.0568](https://arxiv.org/abs/1010.0568) [hep-ph].
- [104] S. Alioli, P. Nason, C. Oleari and E. Re, *JHEP* **06** (2010) 043, doi:[10.1007/JHEP06\(2010\)043](https://doi.org/10.1007/JHEP06(2010)043), arXiv:[1002.2581](https://arxiv.org/abs/1002.2581) [hep-ph].
- [105] S. Frixione, P. Nason and B. R. Webber, *JHEP* **08** (2003) 007, doi:[10.1088/1126-6708/2003/08/007](https://doi.org/10.1088/1126-6708/2003/08/007), arXiv:[hep-ph/0305252](https://arxiv.org/abs/hep-ph/0305252).
- [106] G. P. Salam, *Eur. Phys. J.* **C67** (2010) 637, doi:[10.1140/epjc/s10052-010-1314-6](https://doi.org/10.1140/epjc/s10052-010-1314-6), arXiv:[0906.1833](https://arxiv.org/abs/0906.1833) [hep-ph].
- [107] S. D. Ellis, private communication to the OPAL Collaboration; D. E. Soper and H.-C. Yang, private communication to the OPAL Collaboration; L. A. del Pozo, PhD thesis, University of Cambridge, RALT-002, 1993; R. Akers *et al.* [OPAL Collaboration], *Z. Phys.* **C63**, (1994) 197, doi:[10.1007/BF01411011](https://doi.org/10.1007/BF01411011).
- [108] G. C. Blazey *et al.*, arXiv:[hep-ex/0005012](https://arxiv.org/abs/hep-ex/0005012).
- [109] G. P. Salam and G. Soyez, *JHEP* **05** (2007) 086, doi:[10.1088/1126-6708/2007/05/086](https://doi.org/10.1088/1126-6708/2007/05/086), arXiv:[0704.0292](https://arxiv.org/abs/0704.0292) [hep-ph].
- [110] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock and B. R. Webber, *Phys. Lett.* **B269**, (1991) 432, doi:[10.1016/0370-2693\(91\)90196-W](https://doi.org/10.1016/0370-2693(91)90196-W).

- [111] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, *Nucl. Phys.* **B406** (1993) 187, doi:10.1016/0550-3213(93)90166-M.
- [112] S. D. Ellis and D. E. Soper, *Phys. Rev.* **D48**, (1993) 3160, doi:10.1103/PhysRevD.48.3160, arXiv:hep-ph/9305266.
- [113] M. Cacciari and G. P. Salam, *Phys. Lett.* **B641** (2006) 57, doi:10.1016/j.physletb.2006.08.037, arXiv:hep-ph/0512210.
- [114] M. Cacciari, G. P. Salam and G. Soyez, *FastJet*, <http://fastjet.fr/>.
- [115] M. Cacciari, G. P. Salam and G. Soyez, *JHEP* **04** (2008) 005, doi:10.1088/1126-6708/2008/04/005, arXiv:0802.1188 [hep-ph].
- [116] M. Cacciari and G. P. Salam, *Phys. Lett.* **B659** (2008) 119, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378 [hep-ph].
- [117] V. M. Abazov *et al.* [D0 Collaboration], *Phys. Rev.* **D65**, 052008 (2002), doi:10.1103/PhysRevD.65.052008, arXiv:hep-ex/0108054.
- [118] M. H. Seymour, *Z. Phys.* **C62** (1994) 127, doi:10.1007/BF01559532.
- [119] J. M. Butterworth, B. E. Cox and J. R. Forshaw, *Phys. Rev.* **D65** (2002) 096014, doi:10.1103/PhysRevD.65.096014, arXiv:hep-ph/0201098.
- [120] J. M. Butterworth *et al.*, arXiv:1003.1643 [hep-ph].
- [121] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, *JHEP* **08** (1997) 001, doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323.
- [122] M. Wobisch and T. Wengler, arXiv:hep-ph/9907280;  
M. Wobisch, Diploma thesis, RWTH Aachen University, DESY-THESIS-2000-049, <https://inspirehep.net/literature/538471>.
- [123] M. Cacciari, G. P. Salam and G. Soyez, *JHEP* **0804** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189 [hep-ph].
- [124] M. Cacciari, J. Rojo, G. P. Salam and G. Soyez, *JHEP* **0812** (2008) 032, doi:10.1088/1126-6708/2008/12/032, arXiv:0810.1304 [hep-ph].
- [125] M. Cacciari, J. Rojo, G. P. Salam and G. Soyez, *Testing jet definitions: qq and gg cases*, <http://quality.fastjet.fr>.
- [126] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Phys. Rev. Lett.* **100** (2008) 242001, doi:10.1103/PhysRevLett.100.242001, arXiv:0802.2470 [hep-ph].
- [127] D. Krohn, J. Thaler and L. T. Wang, *JHEP* **02** (2010), 084, doi:10.1007/JHEP02(2010)084, arXiv:0912.1342 [hep-ph].
- [128] C. Buttar *et al.*, arXiv:0803.0678 [hep-ph].
- [129] D. Krohn, J. Thaler and L. T. Wang, *JHEP* **06** (2009), 059, doi:10.1088/1126-6708/2009/06/059, arXiv:0903.0392 [hep-ph].
- [130] M. Dasgupta, L. Magnea and G. P. Salam, *JHEP* **02** (2008) 055, doi:10.1088/1126-6708/2008/02/055 arXiv:0712.3014 [hep-ph].
- [131] M. Rubin, *JHEP* **05** (2010), 005, doi:10.1007/JHEP05(2010)005 arXiv:1002.4557 [hep-ph].
- [132] 2008 CTEQ - MCnet Summer School on QCD Phenomenology and Monte Carlo Event Generators, Debrecen, Hungary, 2008, <https://www.physics.smu.edu/scalise/cteq/schools/summer08/>.
- [133] CTEQ Summer School on QCD Analysis and Phenomenology, Madison, WI, USA, 2009, <https://www.physics.smu.edu/scalise/cteq/schools/summer09/>.

## A $\nu$ hope

*G. Barenboim*

Department of Theoretical Physics and IFIC, University of Valencia, Valencia, Spain

### **Abstract**

The Standard Model has been successful beyond expectations in predicting the results of almost all the experimental tests done so far. In it, neutrinos are massless. Nonetheless, in recent years we have collected solid proofs indicating little but non zero masses for the neutrinos (when contrasted with those of the charged leptons). These masses permit neutrinos to change their flavour and oscillate, indeed a unique treat. In these lectures, I discuss the properties and the amazing potential of neutrinos in and beyond the Standard Model. I review also the pieces of evidence that do not fit with the vanilla three neutrino picture and the prospects to discover the physics that hides beyond the Standard Model (if any) using neutrinos.

### **Keywords**

Lectures; neutrinos; neutrino oscillations; mass; flavor; Majorana particles.

## **1 Introduction**

Last decade witnessed a brutal transformation in neutrino physics. It has been proven beyond any doubt that neutrinos have non-zero masses, implying that leptons mix. This fact was demonstrated by the experimental evidence that neutrinos can change from one state, or “flavour”, to another. All the information we have gathered about neutrinos is relatively new. Less than thirty years old. Neutrino physics as a solid science is in its teenage years and therefore as any adolescent, in a wild and very exciting (and excited) state. However, before jumping into the latest “news” about neutrinos, let’s understand how and why neutrinos were conceived.

The ’20s saw the death of numerous sacred cows, and physics was no exception. One of physics most holly principles, energy conservation, apparently showed up not to hold inside the subatomic world. For some radioactive nuclei, it appeared that a non-negligible fraction of its energy simply vanished, leaving no trace of its presence.

In 1920, in a (by now famous) letter to a meeting [1], Pauli quasi apologetically wrote, “Dear radioactive Ladies and Gentlemen, . . . as a desperate remedy to save the principle of energy conservation in beta decay, . . . I propose the idea of a neutral particle of spin half”. Pauli hypothesised that the missing energy was taken off by another particle, whose properties were such that made it invisible and impossible to detect: it had no electric charge, no mass and only very rarely interacted with matter. With these properties, the neutrino was introduced as one of the few inhabitants of the particle zoo. Before long, Fermi postulated the four-Fermi Hamiltonian in order to describe beta decay utilising the neutrino, electron, neutron and proton. Another field was born: weak interactions took the stage to never leave it. Closing the loop, twenty something years after Pauli’s letter, Cowan and Reines got the experimental signature of anti-neutrinos emitted by a nuclear power plant.

As more particles who participated in weak interactions were found in the years following neutrino discovery, weak interactions got credibility as an authentic new force of nature and the neutrino got to be a key element of it.

Further experimental tests spanning the following years demonstrated that there was not one but three sort, or “flavours” of neutrinos (tagged according to the charged lepton they were produces in association with: electron neutrinos ( $\nu_e$ ), muon neutrinos ( $\nu_\mu$ ) and tau neutrinos ( $\nu_\tau$ )) and that, to the extent we could test, had no mass (and no charge) whatsoever.

The neutrino adventure could have easily finished there, but new tests using neutrinos coming from the sun have shown that the neutrino saga was just beginning . . .

In the canonical Standard Model, neutrinos are completely massless and as a consequence are flavour eigenstates:

$$\begin{aligned} W^+ &\longrightarrow e^+ + \nu_e \quad ; \quad Z \longrightarrow \nu_e + \bar{\nu}_e \quad , \\ W^+ &\longrightarrow \mu^+ + \nu_\mu \quad ; \quad Z \longrightarrow \nu_\mu + \bar{\nu}_\mu \quad , \\ W^+ &\longrightarrow \tau^+ + \nu_\tau \quad ; \quad Z \longrightarrow \nu_\tau + \bar{\nu}_\tau \quad . \end{aligned} \quad (1)$$

Precisely because they are massless, they travel at the speed of light. But the masslessness not only defines the speed at which they propagate, it fixes its flavour as they travel as well. It is evident then, that as flavour is concerned, zero mass neutrinos are not an attractive object to study, specially when contrasted with quarks.

However, if neutrinos were massive, and these masses were not degenerate (degenerate masses flavour-wise is identical to the zero mass case) would mean that neutrino mass eigenstates exist  $\nu_i, i = 1, 2, \dots$ , each with a mass  $m_i$ . The impact of leptonic mixing becomes apparent by looking at the leptonic decays,  $W^+ \longrightarrow \nu_i + \bar{\ell}_\alpha$  of the charged vector boson  $W$ , where  $\alpha = e, \mu$ , or  $\tau$ , and  $\ell_e$  refers to the electron,  $\ell_\mu$  the muon, or  $\ell_\tau$  the tau.

We call particle  $\ell_\alpha$  as the charged lepton of flavour  $\alpha$ . Mixing basically implies that when the charged boson  $W^+$  decays to a given kind of charged lepton  $\bar{\ell}_\alpha$ , the neutrino that goes along is not generally the same mass eigenstate  $\nu_i$ . Any of the different  $\nu_i$  can appear.

The amplitude for the decay of a vector boson  $W^+$  to a particular mix  $\bar{\ell}_\alpha + \nu_i$  is given by  $U_{\alpha i}^*$ . The neutrino that is emitted in this decay alongside the given charged lepton  $\bar{\ell}_\alpha$  is then

$$|\nu_\alpha\rangle = \sum_i U_{\alpha i}^* |\nu_i\rangle . \quad (2)$$

This specific mixture of mass eigenstates yields the neutrino of flavour  $\alpha$ .

The different  $U_{\alpha i}$  can be gathered in a unitary matrix (in the same way they were collected in the CKM matrix in the quark sector) that receives the name of the leptonic mixing matrix or  $U_{PNMS}$  [2]. The unitarity of  $U$  ensures that each time a neutrino of flavour  $\alpha$  through its interaction produces a charged lepton, the produced charged lepton will always be  $\ell_\alpha$ , the charged lepton of flavour  $\alpha$ . That is, a  $\nu_e$  produces exclusively an  $e$ , a  $\nu_\mu$  exclusively a  $\mu$ , and in a similar way  $\nu_\tau$  a  $\tau$ .

The expression (2), portraying each neutrino of a given flavour as a linear combination of the three mass eigenstates, can be easily inverted to depict every mass eigenstate  $\nu_i$  as an analogous linear combination of the three flavours:

$$|\nu_i\rangle = \sum_\alpha U_{\alpha i} |\nu_\alpha\rangle . \quad (3)$$

The amount of  $\alpha$ -flavour (or the  $\alpha$ -fraction) of  $\nu_i$  is obviously  $|U_{\alpha i}|^2$ . When a  $\nu_i$  interacts and creates a charged lepton, this  $\alpha$ -content (or fraction) expresses the probability that the created charged lepton be of flavour  $\alpha$ .

## 2 Neutrino oscillations basics

The phenomenon of neutrino flavour transitions or in short oscillation, can be understood in the following form. A neutrino is created or emitted by a source along with a charged lepton  $\bar{\ell}_\alpha$  of flavour  $\alpha$ . In this way, at the emission point, the neutrino does have a definite flavour. It is a  $\nu_\alpha$ . After that point, the neutrino propagates, i.e., covers some distance  $L$  until it is absorbed.

At this point, when it has already reached the detector, the neutrino (sometimes) interacts and these interactions create another charged lepton  $\ell_\beta$  of flavour  $\beta$ , which we can detect. In this way, at the

target, we can know that the neutrino is again a neutrino of definite flavour, a  $\nu_\beta$ . Of course not always both flavours are identical, sometimes  $\beta \neq \alpha$  (for instance, if  $\ell_\alpha$  is a  $\mu$  however  $\ell_\beta$  is a  $\tau$ ), then, all along his journey from the source to the identification point, the neutrino has changed or mutated from a  $\nu_\alpha$  into a  $\nu_\beta$ .

This transition from one flavour to the other,  $\nu_\alpha \longrightarrow \nu_\beta$ , is only one more realization of the widely known quantum-mechanical effect present in a variety of two state systems and not a particular property of neutrinos.

Since, as shown clearly by Eq. (2), a  $\nu_\alpha$  is truly a coherent superposition of the three mass eigenstates  $\nu_i$ , the neutrino that travels since it is produced until it is detected, can be any of the three  $\nu_i$ 's. Because of that, we should include the contributions of each of the  $\nu_i$  in a coherent way. As a consequence, the transition amplitude,  $\text{Amp}(\nu_\alpha \longrightarrow \nu_\beta)$  receives a contribution of each  $\nu_i$  and turns out to be the product of three pieces. The first factor is the amplitude for the neutrino created at the generation point along with a charged lepton  $\bar{\ell}_\alpha$  to be, particularly, a  $\nu_i$  and is given by  $U_{\alpha i}^*$ .

The second component of our product is the amplitude for the  $\nu_i$  made by the source to cover the distance up to the detector. We will name this element  $\text{Prop}(\nu_i)$  for the time being and will postpone the calculation of its value until later. The last (third) piece is the amplitude for the charged lepton born out of the interaction of the neutrino  $\nu_i$  with the target to be, particularly, a  $\ell_\beta$ .

Being the Hamiltonian that describes the interactions between neutrinos, charged leptons and charged bosons  $W$  hermitian (otherwise probabilities wouldn't be conserved), it follows that if  $\text{Amp}(W \longrightarrow \bar{\ell}_\alpha \nu_i) = U_{\alpha i}^*$ , then  $\text{Amp}(\nu_i \longrightarrow \ell_\beta W) = U_{\beta i}$ . In this way, the third and last component of the product the  $\nu_i$  contribution is given by  $U_{\beta i}$ , and

$$\text{Amp}(\nu_\alpha \longrightarrow \nu_\beta) = \sum_i U_{\alpha i}^* \text{Prop}(\nu_i) U_{\beta i} . \quad (4)$$

It still remains to be established the value of  $\text{Prop}(\nu_i)$ . To determine it, we'd better study the  $\nu_i$  in its rest frame. We will label the time in that system  $\tau_i$ . If  $\nu_i$  does have a rest mass  $m_i$ , then in this frame of reference its state vector satisfies the Schrödinger equation

$$i \frac{\partial}{\partial \tau_i} |\nu_i(\tau_i)\rangle = m_i |\nu_i(\tau_i)\rangle , \quad (5)$$

whose solution is given clearly by

$$|\nu_i(\tau_i)\rangle = e^{-im_i\tau_i} |\nu_i(0)\rangle . \quad (6)$$

Then, the amplitude for a given mass eigenstate  $\nu_i$  to travel freely during a time  $\tau_i$ , is simply the amplitude  $\langle \nu_i(0) | \nu_i(\tau_i) \rangle$  for observing the initial state  $\nu_i$ ,  $|\nu_i(0)\rangle$  after some time as the evolved state  $|\nu_i(\tau_i)\rangle$ , *i.e.*  $\exp[-im_i\tau_i]$ . Thus  $\text{Prop}(\nu_i)$  is only this amplitude where we have used that the time taken by  $\nu_i$  to cover the distance from the source to the detector is just  $\tau_i$ , the proper time.

Nevertheless, if we want  $\text{Prop}(\nu_i)$  to be of any use to us, we must write it first in terms of variables we can measure, this means to express it, in variables in the lab frame. The natural choice is obviously the distance,  $L$ , that the neutrino covers between the source and the detector as seen in the lab frame, and the time,  $t$ , that slips away during the journey, again in the lab frame. The distance  $L$  is set by the experimentalists through the selection of the place of settlement of the source and that of the detector and is unique to each experimental setting. Likewise, the value of  $t$  is selected by the experimentalists through their election for the time at which the neutrino is made and that when it dies (or gets detected). Therefore,  $L$  and  $t$  are determined by the experimental set up, and are the same for all the  $\nu_i$  in the beam. The different  $\nu_i$  do travel through an identical distance  $L$ , in an identical time  $t$ .

We still have two additional lab frame variables to determine, the energy  $E_i$  and three momentum  $p_i$  of the neutrino mass eigenstate  $\nu_i$ . By using the Lorentz invariance of the four component internal product (scalar product), we can obtain the expression for the  $m_i\tau_i$  appearing in the  $\nu_i$  propagator

$\text{Prop}(\nu_i)$  in terms of the (easy to measure) lab frame variables we have been looking for, which is given by

$$m_i \tau_i = E_i t - p_i L . \quad (7)$$

At this point however one may argue that, in real life, neutrino sources are basically constant in time, and that the time  $t$  that slips away since the neutrino is produced till it dies in the detector is actually not measured. This argument is absolutely right. In reality, an experiment smears over the time  $t$  used by the neutrino to complete its route. However, let's consider that two constituents of the neutrino beam, the first one with energy  $E_1$  and the second one with energy  $E_2$  (both measured in the lab frame), add up coherently to the neutrino signal produced in the detector. Now, let us call  $t$  the time used by the neutrino to cover the distance separating the production and detection points. Then by the time the constituent whose energy is  $E_j$  ( $j = 1, 2$ ) arrives to the detector, it has raised a phase factor  $\exp[-iE_j t]$ . Therefore, we will have an interference between the  $E_1$  and  $E_2$  beam participants that will include a phase factor  $\exp[-i(E_1 - E_2)t]$ . When smeared over the non-observed travel time  $t$ , this factor goes away, *except when*  $E_2 = E_1$ . Therefore, only those constituents of the neutrino beam that share the same energy contribute coherently to the neutrino oscillation signal [3,4]. Specifically, only the different mass eigenstates constituents of the beams that have the same energy weight in. The rest gets averaged out.

Courtesy to its dispersion relation, a mass eigenstate  $\nu_i$ , with mass  $m_i$ , and energy  $E$ , has a three momentum  $p_i$  whose absolute value is given by

$$p_i = \sqrt{E^2 - m_i^2} \cong E - \frac{m_i^2}{2E} . \quad (8)$$

Where, we have utilised that as the masses of the neutrinos are miserably small,  $m_i^2 \ll E^2$  for a typical energy  $E$  attainable at any experiment (the lowest energy neutrinos have are MeV energies while masses are sub-eV). From Eqs. (7) and (8), it is easy to see that at a given energy  $E$  the phase  $m_i \tau_i$  appearing in  $\text{Prop}(\nu_i)$  takes the value

$$m_i \tau_i \cong E(t - L) + \frac{m_i^2}{2E} L . \quad (9)$$

As the phase  $E(t - L)$  appears in all the interfering terms it will eventually disappear when calculating the transition amplitude. After all it is a common phase factor (its absolute value is one). Thus, we can get rid of it already now and use

$$\text{Prop}(\nu_i) = \exp\left[-im_i^2 \frac{L}{2E}\right] . \quad (10)$$

Plugging this into Eq. (4), we can obtain that the amplitude for a neutrino born as a  $\nu_\alpha$  to be detected as a  $\nu_\beta$  after covering a distance  $L$  with energy  $E$  yields

$$\text{Amp}(\nu_\alpha \longrightarrow \nu_\beta) = \sum_i U_{\alpha i}^* e^{-im_i^2 \frac{L}{2E}} U_{\beta i} . \quad (11)$$

The expression above is valid for an arbitrary number of neutrino flavours and an identical number of mass eigenstates, as far as they travel through vacuum. The probability  $P(\nu_\alpha \longrightarrow \nu_\beta)$  for  $\nu_\alpha \longrightarrow \nu_\beta$  can be found by squaring it, giving

$$P(\nu_\alpha \longrightarrow \nu_\beta) = |\text{Amp}(\nu_\alpha \longrightarrow \nu_\beta)|^2 = \delta_{\alpha\beta} - 4 \sum_{i>j} \Re(U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*) \sin^2 \left( \Delta m_{ij}^2 \frac{L}{4E} \right) + 2 \sum_{i>j} \Im(U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*) \sin \left( \Delta m_{ij}^2 \frac{L}{2E} \right) , \quad (12)$$

with  $\Delta m_{ij}^2 \equiv m_i^2 - m_j^2$ . In order to get Eq. (12) we have used that the mixing matrix  $U$  is unitary.

The oscillation probability  $P(\nu_\alpha \rightarrow \nu_\beta)$  we have just obtained corresponds to that of a *neutrino*, and not to an *antineutrino*, as we have used that the oscillating neutrino was produced along with a charged *antilepton*  $\bar{\ell}$ , and gives birth to a charged *lepton*  $\ell$  once it reaches the detector. The corresponding probability  $P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta)$  for an antineutrino oscillation can be obtained from  $P(\nu_\alpha \rightarrow \nu_\beta)$  taking advantage of the fact that the two transitions  $\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta$  and  $\nu_\beta \rightarrow \nu_\alpha$  are CPT conjugated processes. Thus, assuming that neutrino interactions respect CPT [5],

$$P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) = P(\nu_\beta \rightarrow \nu_\alpha) . \quad (13)$$

Then it is clear that if the mixing matrix  $U$  is complex,  $P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta)$  and  $P(\nu_\alpha \rightarrow \nu_\beta)$  will not be identical, in general. As  $\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta$  and  $\nu_\alpha \rightarrow \nu_\beta$  are CP conjugated processes,  $P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) \neq P(\nu_\alpha \rightarrow \nu_\beta)$  would provide evidence of CP violation in neutrino oscillations (if Nature has chosen its mixing parameters so that the mixing matrix is indeed complex). Until now, CP violation has been observed only in the quark sector, so its measurement in neutrino physics would be quite exciting.

So far, we have been working in natural units. A fact that becomes transparent by looking at the dispersion relation Eq. (9). If we restore now the  $\hbar$ 's and  $c$  factors (we have happily set to one) into the oscillation probability we find that

$$\sin^2 \left( \Delta m_{ij}^2 \frac{L}{4E} \right) \rightarrow \sin^2 \left( \Delta m_{ij}^2 c^4 \frac{L}{4\hbar c E} \right) . \quad (14)$$

Having done that, it is easy and instructive to explore the semi-classical limit,  $\hbar \rightarrow 0$ . In this limit the oscillation length goes to zero (the oscillation phase goes to infinity) and the oscillations are averaged to 1/2. The interference pattern is lost. A similar situation appears if we let the mass difference  $\Delta m^2$  become large. This is exactly what happens in the quark sector (and the reason why we never study quark oscillations despite knowing that mass eigenstates do not coincide with flavour eigenstates).

In terms of real life units (which are not "natural" units), the oscillation phase is given by

$$\Delta m_{ij}^2 \frac{L}{4E} = 1.27 \Delta m_{ij}^2 (\text{eV}^2) \frac{L (\text{km})}{E (\text{GeV})} . \quad (15)$$

then, since  $\sin^2[1.27 \Delta m_{ij}^2 (\text{eV}^2) L (\text{km}) / E (\text{GeV})]$  can be experimentally observed (*ie.* not smeared out) only if its argument is in a ballpark around one, an experimental set-up with a baseline  $L$  (km) and an energy  $E$  (GeV) is sensitive to neutrino mass squared differences  $\Delta m_{ij}^2 (\text{eV}^2)$  of order  $\sim [L (\text{km}) / E (\text{GeV})]^{-1}$ . For example, an experiment with a baseline of  $L \sim 10^4$  km, roughly the size of Earth's diameter, and  $E \sim 1$  GeV would explore mass differences  $\Delta m_{ij}^2$  down to  $\sim 10^{-4}$  eV<sup>2</sup>. This fact makes it clear that neutrino long-baseline experiments can test even miserably small neutrino mass differences. It does so by exploiting the quantum mechanical interference between amplitudes whose relative phases are given precisely by these super tiny neutrino mass differences, which can be transformed into sizeable effects by choosing  $L/E$  appropriately.

But let's keep analysing the oscillation probability and see whether we can learn more about neutrino oscillations by studying its expression.

It is clear from  $P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta)$  that if neutrinos have zero mass, in such a way that all  $\Delta m_{ij}^2 = 0$ , then,  $P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) = \delta_{\alpha\beta}$ . Therefore, the experimental observation that neutrinos can morph from one flavour to a different one indicates that neutrinos are not only massive but also that their masses are not degenerate. Actually, it was precisely this evidence the one that proved beyond any reasonable doubt that neutrinos are massive.

However, every neutrino oscillation seen so far has involved at some point neutrinos that travel through matter. But the expression we derived is valid only for flavour change in vacuum, and does not take into account any interaction between the neutrinos and the matter traversed between their source and their detector. Thus, the question remains whether it may be that some unknown flavour changing

interactions between neutrinos and matter are indeed responsible of the observed flavour transitions, and not neutrino masses. Regarding this question, a couple of things should be said. First, although it is true that the Standard Model of elementary particle physics contains only massless neutrinos, it provides an amazingly well corroborated description of weak interactions, and therefore of all the ways a neutrino interacts. Such a description does not include flavour change. Second, for some of the processes experimentally observed where neutrinos do change flavour, matter effects are expected to be miserably small, and on those cases the evidence points towards a dependence on  $L$  and  $E$  in the flavour transition probability through the combination  $L/E$ , as anticipated by the oscillation hypothesis. Modulo a constant,  $L/E$  is precisely the proper time that goes by in the rest frame of the neutrino as it covers a distance  $L$  possessing an energy  $E$ . Therefore, these flavour transitions behave as if they were a true progression of the neutrino itself over time, and not a result of an interaction with matter.

Now, let's explore the case where the leptonic mixing were trivial. This would imply that in the charged boson decay  $W^+ \rightarrow \bar{\ell}_\alpha + \nu_i$ , which as we established has an amplitude  $U_{\alpha i}^*$ , the emerging charged antilepton  $\bar{\ell}_\alpha$  of flavour  $\alpha$  comes along always with the *same* neutrino mass eigenstate  $\nu_i$ . That is, if  $U_{\alpha i}^* \neq 0$ , then due to unitarity,  $U_{\alpha j}$  becomes zero for all  $j \neq i$ . Therefore, from Eq. (12) it is clear that,  $P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) = \delta_{\alpha\beta}$ . Thus, the observation that neutrinos morph indicates non trivial a mixing matrix.

Then, we are left with basically two ways to detect neutrino flavour change. The first one is to observe, in a beam of neutrinos which are all created with the same flavour, say  $\alpha$ , some amount of neutrinos of a new flavour  $\beta$  that is different from the flavour  $\alpha$  we started with. This goes under the name of appearance experiments. The second way is to start with a beam of identical  $\nu_\alpha$ s, whose flux is either measured or known, and observe that after travelling some distance this flux is depleted. Such experiments are called disappearance experiments.

As Eq. (12) shows, the transition probability in vacuum does not only depend on  $L/E$  but also oscillates with it. It is because of this fact that neutrino flavour transitions are named “neutrino oscillations”. Now notice also that neutrino transition probabilities do not depend on the individual neutrino masses (or masses squared) but on the squared-mass *differences*. Thus, oscillation experiments can only measure the neutrino mass squared spectrum. Not its absolute scale. Experiments can test the pattern but cannot determine the distance above zero the whole spectra lies.

It is clear that neutrino transitions cannot modify the total flux in a neutrino beam, but simply alter its distribution between the different flavours. Actually, from Eq. (12) and the unitarity of the  $U$  matrix, it is obvious that

$$\sum_{\beta} P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) = 1 \quad , \quad (16)$$

where the sum runs over all flavours  $\beta$ , including the original one  $\alpha$ . Eq. (16) makes it transparent that the probability that a neutrino morphs its flavour, added to the probability that it keeps the flavour it had at birth, is one. Ergo, flavour transitions do not modify the total flux. Nevertheless, some of the flavours  $\beta \neq \alpha$  into which a neutrino can oscillate into may be *sterile* flavours; that is, flavours that do not take part in weak interactions and therefore escape detection. If any of the original (active) neutrino flux turns into sterile, then an experiment able to measure the total *active* neutrino flux—that is, the flux associated to those neutrinos that couple to the weak gauge bosons:  $\nu_e$ ,  $\nu_\mu$ , and  $\nu_\tau$ —will observe it to be not exactly the original one, but smaller than it. In the experiments performed up today, no clear evidence of missing fluxes was found (although there are some hints in this direction).

In the literature, description of neutrino oscillations normally assume that the different mass eigenstates  $\nu_i$  that contribute coherently to a beam share the same *momentum*, rather than the same *energy* as we have argued they must have. While the supposition of equal momentum is technically wrong, it is an inoffensive mistake, since, as can easily be shown, it conveys to the same oscillation probabilities as the ones we have obtained.

A relevant and interesting case of the (not that simple) formula for  $P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta)$  is the case where

only two flavours participate in the oscillation. The only-two-neutrino scenario is a rather rigorous description of a vast number of experiments. In fact only recently (and in few experiments) a more sophisticated (three neutrino description) was needed to fit observations. Lets assume then, that only two mass eigenstates, which we will name  $\nu_1$  and  $\nu_2$ , and two reciprocal flavour states, which we will name  $\nu_\mu$  and  $\nu_\tau$ , are relevant, in such a way that only one squared-mass difference,  $m_2^2 - m_1^2 \equiv \Delta m^2$  arises. Even more, neglecting phase factors that can be proven to have no impact on oscillation probabilities, the mixing matrix  $U$  can be written as

$$\begin{pmatrix} \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} . \quad (17)$$

The unitary mixing matrix  $U$  of Eq. (17) is just a  $2 \times 2$  rotation matrix, and as such, parameterized by a single rotation angle  $\theta$  which is named (in neutrino physics) as the mixing angle. Plugging the  $U$  of Eq. (17) and the unique  $\Delta m^2$  into the general formula of the transition probability  $P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta)$ , we can readily see that, for  $\beta \neq \alpha$ , when only two neutrinos are relevant,

$$P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) = \sin^2 2\theta \sin^2 \left( \frac{\Delta m^2 L}{4E} \right) . \quad (18)$$

Moreover, the survival probability, *ie.* the probability that the neutrino remains with the same flavour its was created with is, as expected, one minus the probability that it changes flavour.

### 3 Neutrino oscillations in a medium

When we create a beam of neutrinos on earth through an accelerator and send it up to thousand kilometres away to meet a detector, the beam does not move through vacuum, but through matter, earth matter. The beam of neutrinos then scatters from the particles it meets along the way. Such a coherent forward scattering can have a large effect on the transition probabilities. We will assume for the time being that neutrino interactions with matter are flavour conserving, as described by the Standard Model, and comment on the possibility of flavour changing interactions later. Then as there are only two types of weak interactions (mediated by charged and neutral currents) there would be accordingly only two possibilities for this coherent forward scattering from matter particles to take place. Charged current mediated weak interactions will occur only if and only if the incoming neutrino is an electron neutrino. As only the  $\nu_e$  can exchange charged boson  $W$  with an Earth electron. Thus neutrino-electron coherent forward scattering via  $W$  exchange opens up an extra source of interaction energy  $V_W$  suffered exclusively by electron neutrinos. Obviously, this additional energy being from weak interactions origin has to be proportional to  $G_F$ , the Fermi coupling constant. In addition, the interaction energy coming from  $\nu_e - e$  scattering grows with the number of targets,  $N_e$ , the number of electrons per unit volume (given by the density of the Earth). Putting everything together it is not difficult to see that

$$V_W = +\sqrt{2} G_F N_e , \quad (19)$$

clearly, this interaction energy affects also antineutrinos (in a opposite way though). It changes sign if we switch the  $\nu_e$  by  $\bar{\nu}_e$ .

The interactions mediated by neutral currents correspond to the case where a neutrino in matter interacts with a matter electron, proton, or neutron by exchanging a neutral  $Z$  boson. According to the Standard Model weak interactions are flavour blind. Every flavour of neutrino enjoys them, and the amplitude for this  $Z$  exchange is always the same. It also teaches us that, at zero momentum transfer, electrons and protons couple to the  $Z$  boson with equal strength. The interaction has though, opposite sign. Therefore, counting on the fact that the matter through which our neutrino moves is electrically neutral (it contains equal number of electrons and protons), the contribution of both, electrons and protons to coherent forward neutrino scattering through  $Z$  exchange will add up to zero. Consequently only

interactions with neutrons will survive so that, the effect of the  $Z$  exchange contribution to the interaction potential energy  $V_Z$  reduces exclusively to that with neutrons and will be proportional to  $N_n$ , the number density of neutrons. It goes without saying that it will be equal to all flavours. This time, we find that

$$V_Z = -\frac{\sqrt{2}}{2} G_F N_n , \quad (20)$$

as was the case before, for  $V_W$ , this contribution will flip sign if we replace the neutrinos by anti-neutrinos.

But if, as we said, the Standard Model interactions do not change neutrino flavour, neutrino flavour transitions or neutrino oscillations point undoubtedly to neutrino mass and mixing even when neutrinos are propagating through matter. Unless non-Standard-Model flavour changing interactions play a role.

Neutrino propagation in matter is easy to understand when analysed through the time dependent Schrödinger equation in the lab frame

$$i \frac{\partial}{\partial t} |\nu(t)\rangle = \mathcal{H} |\nu(t)\rangle . \quad (21)$$

where,  $|\nu(t)\rangle$  is a (three component) neutrino vector state, in which each neutrino flavour corresponds to one component. In the same way, the Hamiltonian  $\mathcal{H}$  is a (three  $\times$  three) matrix in flavour space. To make our lives easy, lets analyse the case where only two neutrino flavours are relevant, say  $\nu_e$  and  $\nu_\mu$ . Then

$$|\nu(t)\rangle = \begin{pmatrix} f_e(t) \\ f_\mu(t) \end{pmatrix} , \quad (22)$$

with  $f_i(t)^2$  the amplitude of the neutrino to be a  $\nu_i$  at time  $t$ . This time the Hamiltonian,  $\mathcal{H}$ , is a  $2 \times 2$  matrix in neutrino flavour space, i.e.,  $\nu_e - \nu_\mu$  space.

It will prove to be clarifying to work out the two flavour case in vacuum first, and add matter effects afterwards. Using Eq. (2) to express  $|\nu_\alpha\rangle$  as a linear combination of mass eigenstates, we can see that the  $\nu_\alpha - \nu_\beta$  matrix element of the Hamiltonian in vacuum,  $\mathcal{H}_{\text{vac}}$ , can be written as

$$\langle \nu_\alpha | \mathcal{H}_{\text{vac}} | \nu_\beta \rangle = \langle \sum_i U_{\alpha i}^* \nu_i | \mathcal{H}_{\text{vac}} | \sum_j U_{\beta j}^* \nu_j \rangle = \sum_j U_{\alpha j} U_{\beta j}^* \sqrt{p^2 + m_j^2} . \quad (23)$$

where we are supposing that the neutrinos belong to a beam where all its mass components (the mass eigenstates) share the same definite momentum  $p$ . As we have already mentioned, despite this supposition being technically wrong, it leads anyway to the right transition amplitude. In the second line of Eq. (23), we have used that the neutrinos  $\nu_j$  with momentum  $p$ , the mass eigenstates, are the asymptotic states of the hamiltonian,  $\mathcal{H}_{\text{vac}}$  for which constitute an orthonormal basis, satisfy

$$\mathcal{H}_{\text{vac}} |\nu_j\rangle = E_j |\nu_j\rangle \quad (24)$$

and have the standard dispersion relation,  $E_j = \sqrt{p^2 + m_j^2}$ .

As we have already mentioned, neutrino oscillations are the archetype quantum interference phenomenon, where only the *relative* phases of the interfering states play a role. Therefore, only the *relative* energies of these states, which set their relative phases, are relevant. As a consequence, if it proves to be convenient (and it will), we can feel free to happily remove from the Hamiltonian  $\mathcal{H}$  any contribution proportional to the identity matrix  $I$ . As we have said, this subtraction will leave unaffected the differences between the eigenvalues of  $\mathcal{H}$ , and therefore will leave unaffected the prediction of  $\mathcal{H}$  for flavour transitions.

It goes without saying that as in this case only two neutrinos are relevant, there are only two mass eigenstates,  $\nu_1$  and  $\nu_2$ , and only one mass splitting  $\Delta m^2 \equiv m_2^2 - m_1^2$ , and therefore there should be,

as before a unitary  $U$  matrix given by Eq. (17) which rotates from one basis to the other. Inserting it into Eq. (23), and assuming that our neutrinos have low masses as compared to their momenta, i.e.,  $(p^2 + m_j^2)^{1/2} \cong p + m_j^2/2p$ , and removing from  $\mathcal{H}_{\text{Vac}}$  a term proportional to the identity matrix (a removal we know is going to be harmless), we get

$$\mathcal{H}_{\text{Vac}} = \frac{\Delta m^2}{4E} \begin{pmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix}. \quad (25)$$

To write this expression, the highly relativistic approximation, which says that  $p \cong E$  is used. Where  $E$  is the average energy of the neutrino mass eigenstates in our neutrino beam of ultra high momentum  $p$ .

It is not difficult to corroborate that the Hamiltonian  $\mathcal{H}_{\text{Vac}}$  of Eq. (25) for the two neutrino scenario would give an identical oscillation probability, Eq. (18), as the one we have already obtained in a different way. An easy way to do it is to analyse the transition probability for the process  $\nu_e \longrightarrow \nu_\mu$ . From Eq. (17) it is clear that in terms of the mixing angle, the electron and muon neutrino states composition is

$$|\nu_e\rangle = |\nu_1\rangle \cos \theta + |\nu_2\rangle \sin \theta, \quad |\nu_\mu\rangle = -|\nu_1\rangle \sin \theta + |\nu_2\rangle \cos \theta. \quad (26)$$

In the same way, we can also write the eigenvalues of the vacuum hamiltonian  $\mathcal{H}_{\text{Vac}}$ , Eq.25, in terms of the mass squared differences as

$$\lambda_1 = -\frac{\Delta m^2}{4E}, \quad \lambda_2 = +\frac{\Delta m^2}{4E}. \quad (27)$$

The mass eigenbasis of this Hamiltonian,  $|\nu_1\rangle$  and  $|\nu_2\rangle$ , can also be written in terms of flavour eigenbasis  $|\nu_e\rangle$  and  $|\nu_\mu\rangle$  by means of Eqs. (26). Therefore, the Schrödinger equation of Eq. (21), with the identification of  $\mathcal{H}$  in this case with  $\mathcal{H}_{\text{Vac}}$  tells us that if at time  $t = 0$  we begin from a  $|\nu_e\rangle$ , then once some time  $t$  elapses this  $|\nu_e\rangle$  will progress into the state given by

$$|\nu(t)\rangle = |\nu_1\rangle e^{+i\frac{\Delta m^2}{4E}t} \cos \theta + |\nu_2\rangle e^{-i\frac{\Delta m^2}{4E}t} \sin \theta. \quad (28)$$

Thus, the probability  $P(\nu_e \longrightarrow \nu_\mu)$  that this evolved neutrino be detected as a different flavour  $\nu_\mu$ , from Eqs. (26) and (28), is given by,

$$\begin{aligned} P(\nu_e \longrightarrow \nu_\mu) &= |\langle \nu_\mu | \nu(t) \rangle|^2 = |\sin \theta \cos \theta (-e^{i\frac{\Delta m^2}{4E}t} + e^{-i\frac{\Delta m^2}{4E}t})|^2 \\ &= \sin^2 2\theta \sin^2 \left( \Delta m^2 \frac{L}{4E} \right). \end{aligned} \quad (29)$$

Where we have substituted the time  $t$  travelled by our highly relativistic state by the distance  $L$  it has covered. The flavour transition or oscillation probability of Eq. (29), as expected, is exactly the same we have found before, Eq. (18).

We can now move on to analyse neutrino propagation in matter. In this case, the  $2 \times 2$  Hamiltonian representing the propagation in vacuum  $\mathcal{H}_{\text{Vac}}$  receives the two additional contributions we have discussed before, and becomes  $\mathcal{H}_M$ , which is given by

$$\mathcal{H}_M = \mathcal{H}_{\text{Vac}} + V_W \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + V_Z \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (30)$$

In the new Hamiltonian, the first additional contribution corresponds to the interaction potential due to the charged bosons exchange, Eq. (19). As this interaction is suffered only by  $\nu_e$ , this contribution is different from zero only in the  $\mathcal{H}_M(1,1)$  element or the  $\nu_e - \nu_e$  element. The second additional contribution, the last term of Eq. (30) comes from the  $Z$  boson exchange, Eq. (20). Since this interaction is flavour blind,

it affects every neutrino flavour in the same way, its contribution to  $\mathcal{H}_M$  is proportional to the identity matrix, and can be safely neglected. Thus

$$\mathcal{H}_M = \mathcal{H}_{\text{Vac}} + \frac{V_W}{2} + \frac{V_W}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (31)$$

where (for reasons that are going to become clear later) we have divided the  $W$ -exchange contribution into two pieces, one proportional to the identity (that we will disregard in the next step) and, a piece that it is not proportional to the identity, that we will keep. Disregarding the first piece as promised, we have from Eqs. (25) and (31)

$$\mathcal{H}_M = \frac{\Delta m^2}{4E} \begin{pmatrix} -(\cos 2\theta - A) & \sin 2\theta \\ \sin 2\theta & (\cos 2\theta - A) \end{pmatrix}, \quad (32)$$

where we have defined

$$A \equiv \frac{V_W/2}{\Delta m^2/4E} = \frac{2\sqrt{2}G_F N_e E}{\Delta m^2}. \quad (33)$$

Clearly,  $A$  parameterizes the relative size of the matter effects as compared to the vacuum contribution given by the neutrino squared-mass splitting and signals the situations when they become important.

Now, if we introduce (a physically meaningful) short-hand notation

$$\Delta m_M^2 \equiv \Delta m^2 \sqrt{\sin^2 2\theta + (\cos 2\theta - A)^2} \quad (34)$$

and

$$\sin^2 2\theta^M \equiv \frac{\sin^2 2\theta}{\sin^2 2\theta + (\cos 2\theta - A)^2}, \quad (35)$$

then the Hamiltonian in a medium  $\mathcal{H}_M$  turns out to be

$$\mathcal{H}_M = \frac{\Delta m_M^2}{4E} \begin{pmatrix} -\cos 2\theta^M & \sin 2\theta^M \\ \sin 2\theta^M & \cos 2\theta^M \end{pmatrix}. \quad (36)$$

and can be diagonalised by inspection, i.e., as a result of our choice, the Hamiltonian in a medium,  $\mathcal{H}_M$ , becomes formally indistinguishable to the vacuum one,  $\mathcal{H}_{\text{Vac}}$ , Eq. (25). The difference being that in this case what used to be the vacuum parameters  $\Delta m^2$  and  $\theta$  are presently given by the matter ones,  $\Delta m_M^2$  and  $\theta^M$ , respectively.

Obviously, the mass eigenstates and eigenvalues (which determine the mass differences and mixing angle) of  $\mathcal{H}_M$  are not identical to the ones in vacuum. The eigenstates in matter, *ie.* the files of the unitary matrix that rotates from the flavour basis to the mass basis, are different from the vacuum eigenvalues that form the vacuum mixing matrix, and therefore  $\theta_M$  is not  $\theta$ . But, the matter Hamiltonian  $\mathcal{H}_M$  does indeed contain all about the propagation of neutrinos in matter, in the same way  $\mathcal{H}_{\text{Vac}}$  contains all about the propagation in vacuum.

According to Eq. (36),  $\mathcal{H}_M$  has the same functional dependence on the matter parameters  $\Delta m_M^2$  and  $\theta^M$  as the vacuum Hamiltonian  $\mathcal{H}_{\text{Vac}}$ , Eq. (25), on the vacuum ones,  $\Delta m^2$  and  $\theta$ . Therefore,  $\Delta m_M^2$  can be identified with an effective mass squared difference in matter, and accordingly  $\theta^M$  can be unidentified with an effective mixing angle in matter.

In a typical experimental set-up where the neutrino beam is generated by an accelerator and sent away to a detector that is several hundred, or even thousand kilometres away, it traverses through earth matter, but only superficially, it does not get deep into the earth. Then, during this voyage the matter density encountered by such a beam can be taken to be approximately constant<sup>1</sup>. But if the density of

<sup>1</sup>This approximation is clearly not valid for neutrinos that cross the Earth

the earth's matter is constant, the same happens with the electron density  $N_e$ , and the  $A$  parameter in which it is incorporated. And it is also true about the Hamiltonian  $\mathcal{H}_M$ . They all become approximately constant, and therefore quite identical to the vacuum Hamiltonian  $\mathcal{H}_{\text{vac}}$ , except for the particular values of their parameters. By comparing Eqs. (36) and (25), we can immediately conclude that exactly in the same way  $\mathcal{H}_{\text{vac}}$  gives rise to vacuum oscillations with probability  $P(\nu_e \rightarrow \nu_\mu)$  of Eq. (29),  $\mathcal{H}_M$  must give rise to matter oscillations with probability

$$P_M(\nu_e \rightarrow \nu_\mu) = \sin^2 2\theta^M \sin^2 \left( \Delta m_M^2 \frac{L}{4E} \right) . \quad (37)$$

Namely, the transition and survival probabilities in matter are the same as those in vacuum, except that the vacuum parameters  $\Delta m^2$  and  $\theta$  are now replaced by their matter counterparts,  $\Delta m_M^2$  and  $\theta^M$ .

In theory, judging simply by its potential, matter effects can have very drastic repercussions in the oscillation probabilities. The exact impact (if any) can be estimated only after the details of the experimental set-up of the experiment in question are given. As a rule of thumb, to guess the importance of matter effects, we should keep in mind that for neutrinos propagating through the earth's mantle (not deeper than 200 km below the surface) and if the kinematic phase associated to the solar mass difference is still negligible,

$$A \cong \frac{E}{13 \text{ GeV}} \quad (38)$$

so that only for beam energies of several GeV matter effects do matter.

And how much do they matter? They matter a lot! From Eq. (35) for the matter mixing angle,  $\theta^M$ , we can appreciate that even when the vacuum mixing angle  $\theta$  is incredible small, say,  $\sin^2 2\theta = 10^{-4}$ , if we get to have  $A \cong \cos 2\theta$ , i.e., for energies of a few tens of GeV, then  $\sin^2 2\theta^M$  can be brutally enhanced as compared to its vacuum value and can even reach maximal mixing, i.e.  $\sin^2 2\theta^M = 1$ . This wild enhancement of a small mixing angle in vacuum up to a sizeable (even maximal) one in matter is the ‘‘resonant’’ enhancement, the largest possible version of the Mikheyev-Smirnov-Wolfenstein effect [6–9]. In the beginning of solar neutrino experiments, people entertained the idea that this brutal enhancement was actually taking place while neutrinos crossed the sun. Nonetheless, as we will see soon the mixing angle associated with solar neutrinos is quite sizeable ( $\sim 34^\circ$ ) already in vacuum [10]. Then, although matter effects on the sun are important and they do enhance the solar mixing angle, unfortunately they are not as drastic as we once dreamt. Nevertheless, for long-baselines they will play (they are already playing!) a key role in the determination of the ordering of the neutrino spectrum.

## 4 Evidence for neutrino oscillations

### 4.1 Atmospheric and accelerator neutrinos

Almost twenty years have elapsed since we were presented solid and convincing evidence of neutrino masses and mixings, and since then, the evidence has only grown. SuperKamiokande (SK) was the first experiment to present compelling evidence of  $\nu_\mu$  disappearance in their atmospheric neutrino fluxes, see Ref. [11]. In Fig. 1 the zenith angle (the angle subtended with the horizontal) dependence of the multi-GeV  $\nu_\mu$  sample is shown together with the disappearance as a function of  $L/E$ . These data fit amazingly well the naive two component neutrino hypothesis with

$$\Delta m_{\text{atm}}^2 = 2 - 3 \times 10^{-3} \text{eV}^2 \quad \text{and} \quad \sin^2 \theta_{\text{atm}} = 0.50 \pm 0.13 . \quad (39)$$

Roughly speaking SK corresponds to an  $L/E$  for oscillations of 500 km/GeV and almost maximal mixing (the mass eigenstates are nearly even admixtures of muon and tau neutrinos). No signal of an involvement of the third flavour,  $\nu_e$  is found so the assumption is that atmospheric neutrino disappearance is basically  $\nu_\mu \rightarrow \nu_\tau$ . Notice however, that the first NOvA results seem to point toward a mixing angle which is not maxima.

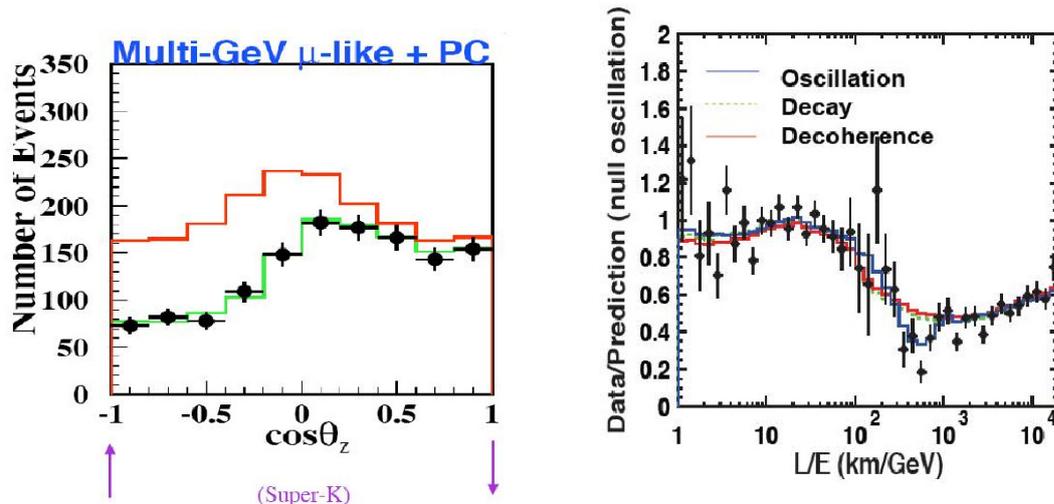


Fig. 1: Superkamiokande's evidence for neutrino oscillations both in the zenith angle and  $L/E$  plots

After atmospheric neutrino oscillations were established, a new series of neutrino experiments were built, sending (man-made) beams of  $\nu_\mu$  neutrinos to detectors located at large distances: the K2K (T2K) experiment [12, 13], sends neutrinos from the KEK accelerator complex to the old SK mine, with a baseline of 120 (235) km while the MINOS (NOvA) experiment [14, 15], sends its beam from Fermilab, near Chicago, to the Soudan mine (Ash river) in Minnesota, a baseline of 735 (810) km. All these experiments have seen evidence for  $\nu_\mu$  disappearance consistent with the one found by SK. Their results are summarised in Fig. 2.

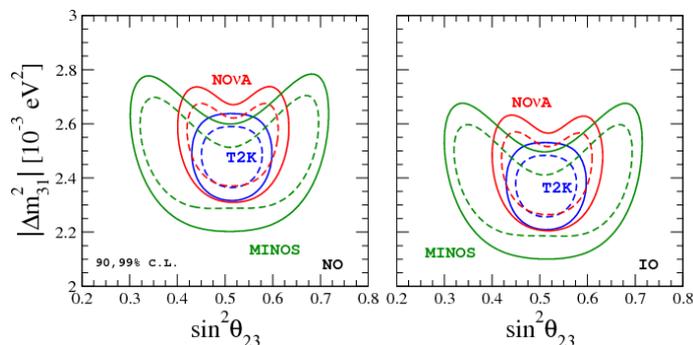


Fig. 2: Allowed regions in the  $\Delta m^2_{\text{atm}}$  vs  $\sin^2 \theta_{\text{atm}}$  plane for MINOS and NOVA data as well as for T2K data and two of the SK analyses. Results from <https://globalfit.astroparticles.es/>.

## 4.2 Reactor and solar neutrinos

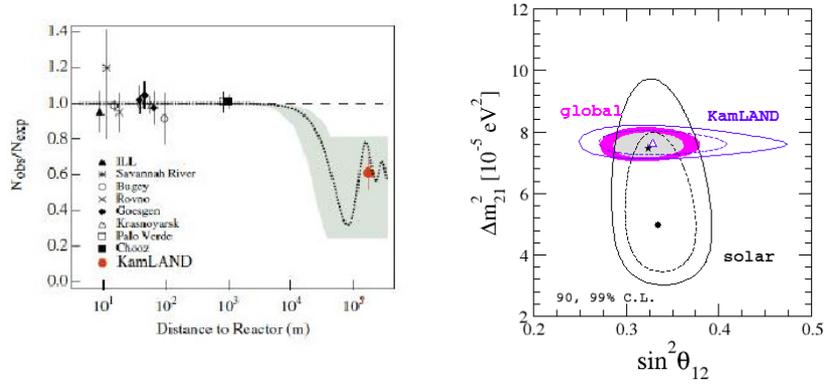
The KamLAND reactor experiment, an antineutrino disappearance experiment, receiving neutrinos from sixteen different reactors, at distances ranging from hundred to thousand kilometres, with an average baseline of 180 km and neutrinos of a few eV [16, 17], has seen evidence of neutrino oscillations. Such evidence was collected not only at a different  $L/E$  than the atmospheric and accelerator experiments but also consists on oscillations involving electron neutrinos,  $\nu_e$ , the ones which were not involved before. These oscillations have also been seen for neutrinos coming from the sun (the sun produces only electron neutrinos). However, in order to compare the two experiments we should assume that neutrinos (solar)

and antineutrinos (reactor) behave in the same way, *ie.* assume CPT conservation. The best fit values in the two neutrino scenario for the KamLAND experiment are

$$\Delta m_{\odot}^2 = 7.55 \pm 0.2 \times 10^{-5} \text{eV}^2 \quad \text{and} \quad \sin^2 \theta_{\odot} = 0.32 \pm 0.03 . \quad (40)$$

In this case, the  $L/E$  involved is 15 km/MeV which is more than an order of magnitude larger than the atmospheric scale and the mixing angle, although large, is clearly not maximal.

Figure 3 shows the disappearance probability for the  $\bar{\nu}_e$  for KamLAND as well as several older reactor experiments with shorter baselines <sup>2</sup>. The second panel depicts the flavour content of the  $^8\text{B}$  solar neutrino flux (with GeV energies) measured by SNO [18] and SK [19]. The reactor outcome can be explained in terms of two flavour oscillations in vacuum, given that the fit to the disappearance probability, is appropriately averaged over  $E$  and  $L$ .



**Fig. 3:** Disappearance of  $\bar{\nu}_e$  observed in reactor experiments as a function of distance from the reactor (favoured region for all solar and reactor experiments). Results from <https://globalfit.astroparticles.es/>.

The analysis of neutrinos originating from the sun is marginally more complex than the one we did before because it should incorporate the matter effects that the neutrinos endure since they are born (at the centre of the sun) until they abandon it, which are imperative at least for the  $^8\text{B}$  Boron neutrinos. The pp and  $^7\text{Be}$  neutrinos are less energetic and therefore are not significantly altered by the presence of matter and leave the sun as though it were ethereal.  $^8\text{B}$  Boron neutrinos on the other hand, leave the sun unequivocally influenced by the presence of matter and this is evidenced by the fact that they leave the sun as  $\nu_2$ , the second mass eigenstate and therefore do not experience oscillations. This distinction among neutrinos coming from different reaction chains is, as mentioned, due mainly to their disparities at birth. While pp ( $^7\text{Be}$ ) neutrinos are created with an average energy of 0.2 MeV (0.9 MeV),  $^8\text{B}$  are born with 10 MeV and as we have seen the impact of matter effects grows with the energy of the neutrino.

However, we ought to emphasise that we do not really see solar neutrino oscillations. To trace the oscillation pattern, to be able to test its distinctive shape, we need a kinematic phase of order one otherwise the oscillations either do not develop or get averaged to 1/2. In the case of neutrinos coming from the sun the kinematic phase is

$$\Delta_{\odot} = \frac{\Delta m_{\odot}^2 L}{4E} = 10^{7 \pm 1} . \quad (41)$$

Consequently, solar neutrinos behave as "effectively incoherent" mass eigenstates once they leave the sun, and remain so once they reach the earth. Consequently the  $\nu_e$  disappearance or survival probability

<sup>2</sup>Shorter baseline reactor neutrino experiments, which has seen no evidence of flux depletion suffer the so-called reactor neutrino anomaly, which may point toward the existence of light sterile states

is given by

$$\langle P_{ee} \rangle = f_1 \cos^2 \theta_\odot + f_2 \sin^2 \theta_\odot \quad (42)$$

where  $f_1$  is the  $\nu_1$  content or fraction of  $\nu_\mu$  and  $f_2$  is the  $\nu_2$  content of  $\nu_\mu$  and therefore both fractions satisfy  $f_1 + f_2 = 1$ .

Nevertheless, as we have already mentioned, solar neutrinos originating from the pp and  ${}^7\text{Be}$  chains are not affected by the solar matter and oscillate as in vacuum and thus, in their case  $f_1 \approx \cos^2 \theta_\odot = 0.69$  and  $f_2 \approx \sin^2 \theta_\odot = 0.31$ . In the  ${}^8\text{B}$  a neutrino case, however, the impact of solar matter is sizeable and the corresponding fractions are substantially altered. In a two neutrino scenario, the day-time CC/NC measured by SNO, which is roughly identical to the day-time average  $\nu_e$  survival probability,  $\langle P_{ee} \rangle$ , reads

$$\left. \frac{CC}{NC} \right|_{\text{day}} = \langle P_{ee} \rangle = f_1 \cos^2 \theta_\odot + f_2 \sin^2 \theta_\odot , \quad (43)$$

where  $f_1$  and  $f_2 = 1 - f_1$  are the  $\nu_1$  and  $\nu_2$  contents of the muon neutrino, respectively, averaged over the  ${}^8\text{B}$  neutrino energy spectrum appropriately weighted with the charged current current cross section. Therefore, the  $\nu_1$  fraction (or how much  $f_2$  differs from 100%) is given by

$$f_1 = \frac{\left( \left. \frac{CC}{NC} \right|_{\text{day}} - \sin^2 \theta_\odot \right)}{\cos 2\theta_\odot} = \frac{(0.347 - 0.311)}{0.378} \approx 10\% \quad (44)$$

where the central values of the last SNO analysis [18] were used. As there are strong correlations between the uncertainties of the CC/NC ratio and  $\sin^2 \theta_\odot$  it is not obvious how to estimate the uncertainty on  $f_1$  from their analysis. Note, that if the fraction of  $\nu_2$  were 100%, then  $\left. \frac{CC}{NC} \right|_{\text{day}} = \sin^2 \theta_\odot$ .

Utilising the analytic analysis of the Mikheyev-Smirnov-Wolfenstein (MSW) effect, gave in Ref. [20], one can obtain the mass eigenstate fractions in a medium, which are given by

$$f_2 = 1 - f_1 = \langle \sin^2 \theta_\odot^M + P_x \cos 2\theta_\odot^M \rangle_{s\text{B}} , \quad (45)$$

with  $\theta_\odot^M$  being the mixing angle as given at the  $\nu_e$  production point and  $P_x$  is the probability of the neutrino to hop from one mass eigenstate to the second one during the Mikheyev-Smirnov resonance crossing. The average  $\langle \dots \rangle_{s\text{B}}$  is over the electron density of the  ${}^8\text{B}$   $\nu_e$  production region in the centre of the Sun as given by the Solar Standard Model and the energy spectrum of  ${}^8\text{B}$  neutrinos has been appropriately weighted with SNO's charged current cross section. All in all, the  ${}^8\text{B}$  energy weighted average content of  $\nu_2$ 's measured by SNO is

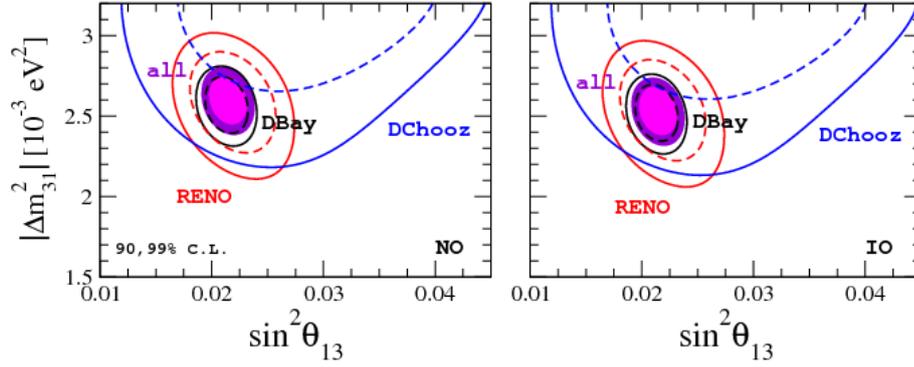
$$f_2 = 91 \pm 2\% \text{ at the } 95\% \text{ C.L.} . \quad (46)$$

Therefore, it is obvious that the  ${}^8\text{B}$  solar neutrinos are the purest mass eigenstate neutrino beam known so far and SK super famous picture of the sun taken (from underground) with neutrinos is made with approximately 90% of  $\nu_2$ , *ie.* almost a pure beam of mass eigenstates.

Last but not least, six years ago, a newly built reactor neutrino experiment, the Daya Bay experiment, located in China, announced the measurement of the third mixing angle [22], the only one which was still missing and found it to be

$$\sin^2(2\theta_{12}) = 0.092 \pm 0.017 . \quad (47)$$

Following this announcement, several experiments confirmed the finding and during the last years the last mixing angle to be measured became the best (most precisely) measured one. The fact that this angle, although smaller than the other two, is still sizeable opens the door to a new generation of neutrino experiments aiming to answer the open questions in the field.

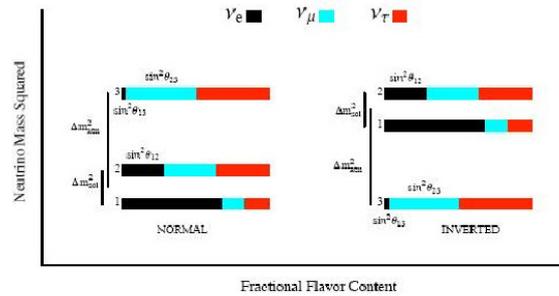


**Fig. 4:** Reactor mixing angle. Results from <https://globalfit.astroparticles.es/>

## 5 $\nu$ Standard Model

Now that we have comprehended the physics behind neutrinos oscillations and have leaned the experimental evidence about the parameters driving these oscillations, we can move ahead and construct the Neutrino Standard Model: it comprises three light ( $m_i < 1$  eV) neutrinos, *ie.* it involves just two mass differences.

So far we have not seen any solid experimental indication (or need) for additional neutrinos<sup>3</sup>. As we have measured long time ago the invisible width of the  $Z$  boson and found it to be 3, within errors, if additional neutrinos are going to be incorporated into the model, they cannot couple to the  $Z$  boson, *ie.* they cannot enjoy weak interactions, so we call them sterile. However, as sterile neutrinos have not been seen (although they may have been hinted), and are not needed to explain any solid experimental evidence, our Neutrino Standard Model will contain just the three active flavours:  $e$ ,  $\mu$  and  $\tau$ .



**Fig. 5:** Flavour content of the three neutrino mass eigenstates (not including the dependence on the cosine of the CP violating phase  $\delta$ ). If CPT is conserved, the flavour content must be the same for neutrinos and anti-neutrinos. Notice that oscillation experiments cannot tell us how far above zero the entire spectrum lies.

The unitary mixing matrix which rotates from the flavour to the mass basis, called the PMNS matrix, comprises three mixing angles (the so called solar mixing angle:  $\theta_{12}$ , the atmospheric mixing angle  $\theta_{23}$ , and the last to be measured, the reactor mixing angle  $\theta_{13}$ ), one Dirac phase ( $\delta$ ) and potentially two Majorana phases ( $\alpha$ ,  $\beta$ ) and is given by

$$|\nu_\alpha\rangle = U_{\alpha i} |\nu_i\rangle$$

<sup>3</sup>Although it must be noted that there are several not significant hint pointing in this direction

$$U_{\alpha i} = \begin{pmatrix} 1 & & \\ & c_{23} & s_{23} \\ & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & & s_{13}e^{-i\delta} \\ & 1 & \\ -s_{13}e^{i\delta} & & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} \\ -s_{12} & c_{12} \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ & e^{i\alpha} & \\ & & e^{i\beta} \end{pmatrix}$$

where  $s_{ij} = \sin \theta_{ij}$  and  $c_{ij} = \cos \theta_{ij}$ . Courtesy of the hierarchy in mass differences (and to a less extent to the smallness of the reactor mixing angle) we are permitted to recognise the (23) label in the three neutrino scenario as the atmospheric  $\Delta m_{\text{atm}}^2$  we obtained in the two neutrino scenario, in a similar fashion the (12) label can be assimilated to the solar  $\Delta m_{\odot}^2$ . The (13) sector drives the  $\nu_e$  flavour oscillations at the atmospheric scale, and the depletion in reactor neutrino fluxes, see Ref. [22]. According to the experiments done so far, the three sigma ranges for the neutrino mixing angles are

$$0.273 < \sin^2 \theta_{12} < 0.379 \quad ; \quad 0.445 < \sin^2 \theta_{23} < 0.599 \quad ; \quad 0.0196 < \sin^2 \theta_{13} < 0.0241$$

while the corresponding ones for the mass splittings are

$$2.41 \times 10^{-3} \text{eV}^2 < |\Delta m_{32}^2| < 2.60 \times 10^{-3} \text{eV}^2 \quad \text{and} \quad 7.05 \times 10^{-5} \text{eV}^2 < \Delta m_{21}^2 < 8.14 \times 10^{-5} \text{eV}^2.$$

These mixing angles and mass splittings are summarised in Fig. 5.

As oscillation experiments only explore the two mass differences, two orderings are possible, as shown in Fig. 5. They are called normal and inverted hierarchy and roughly identify whether the mass eigenstate with the smaller electron neutrino content is the lightest or the heaviest.

The absolute mass scale of the neutrinos, or the mass of the lightest neutrino is not known yet, but cosmological bounds already say that the heaviest one must be lighter than about 0.3 eV.

As transition or survival probabilities depend on the combination  $U_{\alpha i}^* U_{\beta i}$  no trace of the Majorana phases could appear on oscillation phenomena, however they will have observable effects in those processes where the Majorana character of the neutrino is essential for the process to happen, like neutrino-less double beta decay.

## 6 Neutrino mass and character

### 6.1 Absolute neutrino mass

The absolute mass scale of the neutrino, *ie.* the mass of the lightest/heaviest neutrino, cannot be obtained from oscillation experiments, however this does not mean we have no access to it. Direct experiments like tritium beta decay, or neutrinoless double beta decay and indirect ones, like cosmological observations, have the potential to give us the information on the absolute scale of neutrino mass, we so desperately need. The KATRIN tritium beta decay experiment [23] has sensitivity down to 200 meV for the "mass" of  $\nu_e$  defined as

$$m_{\nu_e} = |U_{e1}|^2 m_1 + |U_{e2}|^2 m_2 + |U_{e3}|^2 m_3 \quad . \quad (48)$$

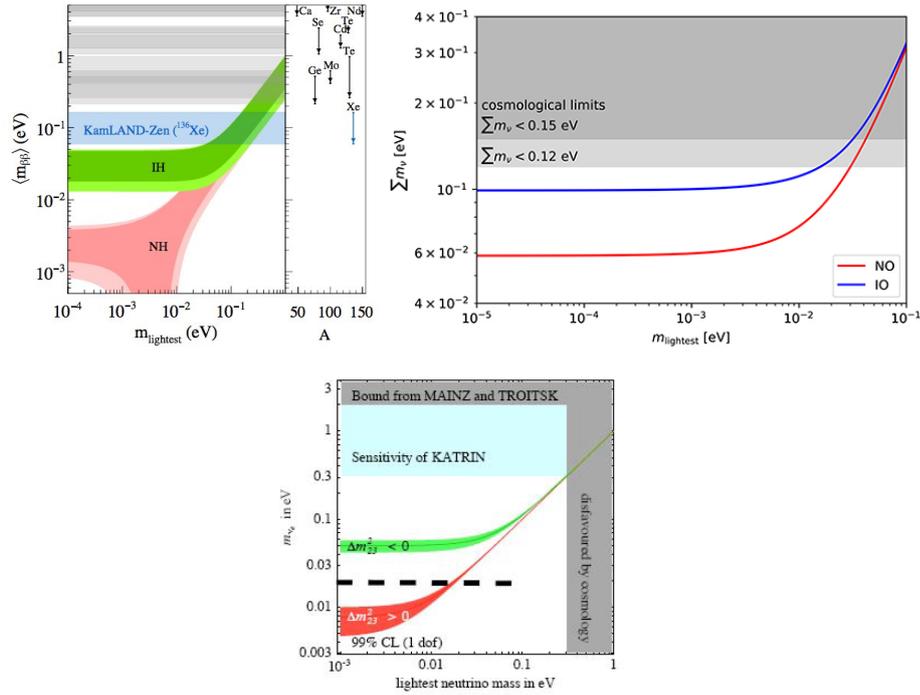
Neutrino-less double beta decay experiments, see Ref. [24] for a review, do not measure the absolute mass of the neutrino directly but a particular combination of neutrino masses and mixings,

$$m_{\beta\beta} = \left| \sum m_i U_{ei}^2 \right| = \left| m_1 c_{13}^2 c_{12}^2 + m_2 c_{13}^2 s_{12}^2 e^{2i\alpha} + m_3 s_{13}^2 e^{2i\beta} \right| \quad , \quad (49)$$

where it is understood that neutrinos are taken to be Majorana particles, *ie.* truly neutral particles (having all their quantum numbers to be zero). The new generation of experiments seeks to reach below 10 meV for  $m_{\beta\beta}$  in double beta decay.

Cosmological probes (CMB and Large Scale Structure experiments) measure the sum of the neutrino masses

$$m_{\text{cosmo}} = \sum_i m_i \quad . \quad (50)$$



**Fig. 6:** The effective mass measured in double  $\beta$  decay, in cosmology and in Tritium  $\beta$  decay versus the mass of the lightest neutrino. Below the dashed lines, only the normal hierarchy is allowed. Notice that while double  $\beta$  decay experiments bound the neutrino mass only in the Majorana case, Planck bounds apply for either case.

and may have a say on the mass ordering (direct or inverted spectrum) as well as test other neutrino properties like neutrino asymmetries [25]. If  $\sum m_i \approx 10$  eV, the energy balance of the universe saturates the bound coming from its critical density. The current limit [26], is a few % of this number,  $\sim 0.3$  eV. These bounds are model dependent but they do all give numbers of the same order of magnitude. However, given the systematic uncertainties characteristic of cosmology, a solid limit of less than 200 meV seems way too aggressive. Figure 6 shows the allowed parameter space for the neutrino masses (as a function of the absolute scale) for both the normal and inverted hierarchy.

## 6.2 Majorana vs Dirac

A fermion mass is nothing but a coupling between a left handed state and a right handed one. Thus, if we examine a massive fermion at rest, then one can regard this state as a linear combination of two massless particles, one right handed and one left handed. If the particle we are examining is electrically charged, like an electron or a muon, both particles, the left handed as well as the right handed must have the same charge (we want the mass term to be electrically neutral). This is a Dirac mass term. However, for a neutral particle, like the neutrino, a new possibility opens up, the left handed particle can be coupled to the right handed anti-particle, (a term which would have a net charge, if the fields are not absolutely and totally neutral) this is a Majorana mass term.

Thus a truly and absolutely neutral particle (who will inevitably be its own antiparticle) does have two ways of getting a mass term, a la Dirac or a la Majorana, and if there are no reasons to forbid one of them, will have them both.

In the case of a neutrino, the left chiral field couples to  $SU(2) \times U(1)$  implying that a Majorana mass term is forbidden by gauge symmetry. However, the right chiral field carries no quantum numbers,

is totally and absolutely neutral. Then, the Majorana mass term is unprotected by any symmetry and it is expected to be very large, of the order of the largest scale in the theory. On the other hand, Dirac mass terms are expected to be of the order of the electroweak scale times a Yukawa coupling, giving a mass of the order of magnitude of the charged lepton or quark masses. Putting all the pieces together, the mass matrix for the neutrinos results as in Fig. 7.

$$\begin{array}{ccc}
 \nu_L \text{ to } \bar{\nu}_R & & \nu_L \text{ to } \nu_R \\
 & \searrow & \swarrow \\
 & 0 & m_D \\
 & \swarrow & \searrow \\
 \bar{\nu}_R \text{ to } \bar{\nu}_L & & \nu_R \text{ to } \bar{\nu}_L
 \end{array}
 \left( \begin{array}{cc}
 0 & m_D \\
 m_D & M
 \end{array} \right)$$

**Fig. 7:** The neutrino mass matrix with the various right to left couplings,  $m_D$  is the Dirac mass terms while 0 and  $M$  are Majorana masses for the charged and uncharged (under  $SU(2) \times U(1)$ ) chiral components.

To get the mass eigenstates we need to diagonalise the neutrino mass matrix. By doing so, one is left with two Majorana neutrinos, one super-heavy Majorana neutrino with mass  $\simeq M$  and one super-light Majorana neutrino with mass  $m_D^2/M$ , *ie.* one mass goes up while the other sinks, this is what we call the seesaw mechanism [27–29]<sup>4</sup>. The light neutrino(s) is(are) the one(s) observed in current experiments (its mass differences) while the heavy neutrino(s) are not accessible to current experiments and could be responsible for explaining the baryon asymmetry of the universe through the generation of a lepton asymmetry at very high energy scales since its decays can in principle be CP violating (they depend on the two Majorana phases on the PNMS matrix which are invisible for oscillations). The super heavy Majorana neutrinos being their masses so large can play a role at very high energies and can be related to inflation [30].

If neutrinos are Majorana particles lepton number is no longer a good quantum number and a plethora of new processes forbidden by lepton number conservation can take place, it is not only neutrino-less double beta decay. For example, a muon neutrino can produce a positively charged muon. However, this process and any processes of this kind, would be suppressed by  $(m_\nu/E)^2$  which is tiny,  $10^{-20}$ , and therefore, although they are technically allowed, are experimentally unobservable. To most stringent limit nowadays comes from KamLAND-zen [31], and constraints the half-life of neutrino-less double beta decay to be  $T_{1/2}^{0\nu} > 1.07 \times 10^{26}$  years at 90% C.L. Forthcoming experiments such as GERDA-PhaseII, Majorana, SuperNEMO, CUORE, and nEXO will improve this sensitivity by one order of magnitude.

Recently low energy seesaw models [32] have experienced a revival and are actively being explored [33]. In such models the heavy states, of only few tens of TeV can be searched for at the LHC. The heavy right handed states in these models will be produced at LHC either through Yukawa couplings or through gauge coupling to right handed gauge bosons. Some models contain also additional scalar that can be looked for.

## 7 Conclusions

The experimental observations of neutrino oscillations, meaning that neutrinos have mass and mix, have answered questions that have been present since the establishment of the Standard Model. As those veils have disappeared, new questions open up and challenge our understanding:

<sup>4</sup>Depending on the envisioned high energy theory, the simplest see saw mechanism can be categorised into three different classes or types (as they are called) depending on their scalar content.

- What is the true nature of the neutrinos? Are they Majorana or Dirac particles?
- Is there any new scale associated to neutrinos masses? Is this new scale accessible using particle colliders?
- Is the spectrum normal or inverted? Is the lightest neutrino the one with the least electron content on it, or is it the heaviest one?
- Is CP violated (is  $\sin \delta \neq 0$ )? If so, is this phase related at any rate with the baryon asymmetry of the Universe? What about the other two phases? Which is the absolute mass scale of the neutrinos?
- Are there new interactions? Are neutrinos related to the open questions in cosmology like dark matter and/or dark energy? Do (presumably heavy) neutrinos play a role in inflation?
- Can neutrinos violate CPT [34]? What about Lorentz invariance? If we ever measure a different spectrum for neutrinos and antineutrinos (after matter effects are properly taken into account), how can we distinguish whether it is due to a true (genuine) CPT violation or to a non-standard neutrino interaction?
- Are these intriguing signals in short baseline reactor neutrino experiments (the missing fluxes) a real effect? Do they imply the existence of sterile neutrinos?

We would like to answer these questions, so we plan to do new experiments. These experiments will for sure bring some answers and clearly open new and pressing questions. Only one thing is clear—our journey into the neutrino world has just begun.

## Acknowledgements

I would like to thank the students and the organisers of the Asia-Europe-Pacific School of High-Energy Physics for giving me the opportunity to present these lectures in such a wonderful atmosphere. I enjoyed each day of the school enormously. I also thank Pablo Martinez for his comments about this manuscript. Support from the MEC and FEDER (EC) Grants SEV-2014-0398, FIS2015-72245-EXP, and FPA2014-54459 and the Generalitat Valenciana under grant PROMETEOII/2017/033 is acknowledged. Also partial support from the European Union FP7 ITN INVISIBLES MSCA PITN-GA-2011-289442 and InvisiblesPlus (RISE) H2020-MSCA-RISE-2015-690575.

## References

- [1] *Symmetry* **4** (2007) ibc, <https://www.symmetrismagazine.org/article/march-2007/neutrino-invention>
- [2] Z. Maki, M. Nakagawa and S. Sakata, *Prog. Theor. Phys.* **28**, (1962) 870, doi:10.1143/PTP.28.870.
- [3] H. J. Lipkin, *Phys. Lett.* **B579** (2004) 355, doi:10.1016/j.physletb.2003.11.013, arXiv:hep-ph/0304187.
- [4] L. Stodolsky, *Phys. Rev.* **D58** (1998) 036006, doi:10.1103/PhysRevD.58.036006, arXiv:hep-ph/9802387.
- [5] An analysis of CPT violation in the neutrino sector can be found in: G. Barenboim and J. D. Lykken, *Phys. Lett.* **B554** (2003) 73, doi:10.1016/S0370-2693(02)03262-8, arXiv:hep-ph/0210411; G. Barenboim, J. F. Beacom, L. Borissoff and B. Kayser, *Phys. Lett.* **B537** (2002) 227, doi:10.1016/S0370-2693(02)01947-0, arXiv:hep-ph/0203261; The world best bound in CPT conservation is found in G. Barenboim, C. A. Ternes and M. Tortola, *Phys. Lett.* **B780** (2018) 631, doi:10.1016/j.physletb.2018.03.060, arXiv:1712.01714 [hep-ph].
- [6] L. Wolfenstein, *Phys. Rev.* **D17** (1978) 2369, doi:10.1103/PhysRevD.17.2369.
- [7] S. P. Mikheev and A. Y. Smirnov, *Sov. J. Nucl. Phys.* **42** (1985) 913, [*Yad. Fiz.* **42** (1985) 1441].
- [8] S. P. Mikheev and A. Y. Smirnov, *Sov. Phys. JETP* **64** (1986) 4, [*Zh. Eksp. Teor. Fiz.* **91** (1986) 7], arXiv:0706.0454 [hep-ph].

- [9] S. P. Mikheev and A. Y. Smirnov, *Nuovo Cim.* **C9**, (1986) 17, doi:[10.1007/BF02508049](https://doi.org/10.1007/BF02508049).
- [10] N. Tolich [SNO Collaboration], *J. Phys. Conf. Ser.* **375** (2012) 042049, doi:[10.1088/1742-6596/375/1/042049](https://doi.org/10.1088/1742-6596/375/1/042049).
- [11] J. Hosaka *et al.* [Super-Kamiokande Collaboration], *Phys. Rev.* **D73**, (2006) 112001, doi:[10.1103/PhysRevD.73.112001](https://doi.org/10.1103/PhysRevD.73.112001), arXiv:[hep-ex/0508053](https://arxiv.org/abs/hep-ex/0508053);  
J. P. Cravens *et al.* [Super-Kamiokande Collaboration], *Phys. Rev.* **D78** (2008) 032002, doi:[10.1103/PhysRevD.78.032002](https://doi.org/10.1103/PhysRevD.78.032002), arXiv:[0803.4312](https://arxiv.org/abs/0803.4312) [hep-ex];  
K. Abe *et al.* [Super-Kamiokande Collaboration], *Phys. Rev.* **D83**, (2011) 052010 doi:[10.1103/PhysRevD.83.052010](https://doi.org/10.1103/PhysRevD.83.052010), arXiv:[1010.0118](https://arxiv.org/abs/1010.0118) [hep-ex].
- [12] M. H. Ahn *et al.* [K2K Collaboration], *Phys. Rev.* **D74** (2006) 072003, doi:[10.1103/PhysRevD.74.072003](https://doi.org/10.1103/PhysRevD.74.072003), arXiv:[hep-ex/0606032](https://arxiv.org/abs/hep-ex/0606032);  
C. Mariani [K2K Collaboration], *AIP Conf. Proc.* **981** (2008) 247, doi:[10.1063/1.2898949](https://doi.org/10.1063/1.2898949).
- [13] K. Abe *et al.* [T2K Collaboration], *Phys. Rev. Lett.* **118**, (2017) 151801, doi:[10.1103/PhysRevLett.118.151801](https://doi.org/10.1103/PhysRevLett.118.151801), arXiv:[1701.00432](https://arxiv.org/abs/1701.00432) [hep-ex].
- [14] A. B. Sousa [MINOS and MINOS+ Collaborations], *AIP Conf. Proc.* **1666** (2015) 110004, doi:[10.1063/1.4915576](https://doi.org/10.1063/1.4915576), arXiv:[1502.07715](https://arxiv.org/abs/1502.07715) [hep-ex];  
L. H. Whitehead [MINOS Collaboration], *Nucl. Phys.* **B908** (2016) 130, doi:[10.1016/j.nuclphysb.2016.03.004](https://doi.org/10.1016/j.nuclphysb.2016.03.004), arXiv:[1601.05233](https://arxiv.org/abs/1601.05233) [hep-ex].
- [15] P. Adamson *et al.* [NOvA Collaboration], *Phys. Rev. Lett.* **118**, (2017) 151802, doi:[10.1103/PhysRevLett.118.151802](https://doi.org/10.1103/PhysRevLett.118.151802), arXiv:[1701.05891](https://arxiv.org/abs/1701.05891) [hep-ex];  
P. Adamson *et al.* [NOvA Collaboration], *Phys. Rev. Lett.* **118**, (2017) 231801, doi:[10.1103/PhysRevLett.118.231801](https://doi.org/10.1103/PhysRevLett.118.231801), arXiv:[1703.03328](https://arxiv.org/abs/1703.03328) [hep-ex].
- [16] I. Shimizu [KamLAND Collaboration], *Nucl. Phys. Proc. Suppl.* **188** (2009) 84, doi:[10.1016/j.nuclphysbps.2009.02.020](https://doi.org/10.1016/j.nuclphysbps.2009.02.020).
- [17] M. P. Decowski [KamLAND Collaboration], *Nucl. Phys.* **B908** (2016) 52, doi:[10.1016/j.nuclphysb.2016.04.014](https://doi.org/10.1016/j.nuclphysb.2016.04.014).
- [18] B. Aharmim *et al.* [SNO Collaboration], *Phys. Rev.* **C88** (2013) 025501, doi:[10.1103/PhysRevC.88.025501](https://doi.org/10.1103/PhysRevC.88.025501), arXiv:[1109.0763](https://arxiv.org/abs/1109.0763) [nucl-ex].
- [19] A. Hime, *Nucl. Phys. Proc. Suppl.* **221** (2011) 110, doi:[10.1016/j.nuclphysbps.2011.03.104](https://doi.org/10.1016/j.nuclphysbps.2011.03.104).
- [20] S. J. Parke and T. P. Walker, *Phys. Rev. Lett.* **57** (1986) 2322, doi:[10.1103/PhysRevLett.57.2322](https://doi.org/10.1103/PhysRevLett.57.2322), [Erratum: *ibid.*, **57** (1986) 3124].
- [21] F. P. An *et al.* [Daya Bay Collaboration], *Phys. Rev.* **D95** (2017) 072006, doi:[10.1103/PhysRevD.95.072006](https://doi.org/10.1103/PhysRevD.95.072006), arXiv:[1610.04802](https://arxiv.org/abs/1610.04802) [hep-ex];  
S. B. Kim [RENO Collaboration], *Nucl. Phys.* **B908** (2016) 94, doi:[10.1016/j.nuclphysb.2016.02.036](https://doi.org/10.1016/j.nuclphysb.2016.02.036).
- [22] F. P. An *et al.* [Daya Bay Collaboration], *Phys. Rev.* **D95** (2017) 072006, doi:[10.1103/PhysRevD.95.072006](https://doi.org/10.1103/PhysRevD.95.072006), arXiv:[1610.04802](https://arxiv.org/abs/1610.04802) [hep-ex];  
J. H. Choi *et al.* [RENO Collaboration], *Phys. Rev. Lett.* **116** (2016) 211801, doi:[10.1103/PhysRevLett.116.211801](https://doi.org/10.1103/PhysRevLett.116.211801), arXiv:[1511.05849](https://arxiv.org/abs/1511.05849) [hep-ex];  
Y. Abe *et al.* (Double Chooz), *JHEP* **10**, 086 (2014), doi:[10.1007/JHEP10\(2014\)086](https://doi.org/10.1007/JHEP10(2014)086), [Erratum: *JHEP* **02** (2015) 074], arXiv:[1406.7763](https://arxiv.org/abs/1406.7763) [hep-ex].
- [23] S. Mertens [KATRIN Collaboration], *Phys. Procedia* **61** (2015) 267, doi:[10.1016/j.phpro.2014.12.043](https://doi.org/10.1016/j.phpro.2014.12.043).
- [24] S. R. Elliott and P. Vogel, *Ann. Rev. Nucl. Part. Sci.* **52** (2002) 115, doi:[10.1146/annurev.nucl.52.050102.090641](https://doi.org/10.1146/annurev.nucl.52.050102.090641), arXiv:[hep-ph/0202264](https://arxiv.org/abs/hep-ph/0202264).
- [25] G. Barenboim, W. H. Kinney and W. I. Park, *Phys. Rev.* **D95** (2017) 043506, doi:[10.1103/PhysRevD.95.043506](https://doi.org/10.1103/PhysRevD.95.043506), arXiv:[1609.01584](https://arxiv.org/abs/1609.01584) [hep-ph];

- G. Barenboim, W. H. Kinney and W. I. Park, *Eur. Phys. J.* **C77** (2017) 590, [doi:10.1140/epjc/s10052-017-5147-4](https://doi.org/10.1140/epjc/s10052-017-5147-4), [arXiv:1609.03200](https://arxiv.org/abs/1609.03200) [astro-ph.CO].
- [26] M. Lattanzi [Planck Collaboration], *J. Phys. Conf. Ser.* **718** (2016) 032008, [doi:10.1088/1742-6596/718/3/032008](https://doi.org/10.1088/1742-6596/718/3/032008).
- [27] M. Gell-Mann, P. Ramond and R. Slansky, Complex spinors and unified theories, Supergravity Workshop, Stony Brook, NY, United States, 1979, Eds. by P. van Nieuwenhuizen, D. Freedman, (North-Holland, 1979), p.315, [arXiv:1306.4669](https://arxiv.org/abs/1306.4669) [hep-th].
- [28] R. N. Mohapatra and G. Senjanovic, *Phys. Rev. Lett.* **44** (1980) 912, [doi:10.1103/PhysRevLett.44.912](https://doi.org/10.1103/PhysRevLett.44.912).
- [29] M. Fukugita and T. Yanagida, *Phys. Lett.* **B174** (1986) 45, [doi:10.1016/0370-2693\(86\)91126-3](https://doi.org/10.1016/0370-2693(86)91126-3).
- [30] G. Barenboim, *JHEP* **03** (2009) 102, [doi:10.1088/1126-6708/2009/03/102](https://doi.org/10.1088/1126-6708/2009/03/102), [arXiv:0811.2998](https://arxiv.org/abs/0811.2998) [hep-ph];  
G. Barenboim, *Phys. Rev.* **D82** (2010) 093014, [doi:10.1103/PhysRevD.82.093014](https://doi.org/10.1103/PhysRevD.82.093014), [arXiv:1009.2504](https://arxiv.org/abs/1009.2504) [hep-ph].
- [31] Y. Gando [KamLAND-Zen Collaboration], *Nucl. Part. Phys. Proc.* **273-275** (2016) 1842, [doi:10.1016/j.nuclphysbps.2015.09.297](https://doi.org/10.1016/j.nuclphysbps.2015.09.297).
- [32] F. Borzumati and Y. Nomura, *Phys. Rev.* **D64** (2001) 053005, [doi:10.1103/PhysRevD.64.053005](https://doi.org/10.1103/PhysRevD.64.053005), [arXiv:hep-ph/0007018](https://arxiv.org/abs/hep-ph/0007018).
- [33] C. G. Cely, A. Ibarra, E. Molinaro and S. T. Petcov, *Phys. Lett.* **B718** (2013) 957, [doi:10.1016/j.physletb.2012.11.026](https://doi.org/10.1016/j.physletb.2012.11.026), [arXiv:1208.3654](https://arxiv.org/abs/1208.3654) [hep-ph].
- [34] G. Barenboim, L. Borisso, J. D. Lykken and A. Y. Smirnov, *JHEP* **10** (2002) 001, [doi:10.1088/1126-6708/2002/10/001](https://doi.org/10.1088/1126-6708/2002/10/001), [arXiv:hep-ph/0108199](https://arxiv.org/abs/hep-ph/0108199);  
G. Barenboim and J. D. Lykken, *Phys. Rev.* **D80** (2009) 113008, [doi:10.1103/PhysRevD.80.113008](https://doi.org/10.1103/PhysRevD.80.113008), [arXiv:0908.2993](https://arxiv.org/abs/0908.2993) [hep-ph].



# Flavour physics and CP violation

*Y. Nir*

Department of Particle Physics and Astrophysics, Weizmann Institute of Science, Rehovot, Israel

## Abstract

We explain the reasons for the interest in flavor physics. We describe flavor physics and the related CP violation within the Standard Model, and explain how the B-factories proved that the CKM (KM) mechanism dominates the flavor changing (CP violating) processes that have been observed in meson decays. We explain the implications of flavor physics for new physics, with emphasis on the “new physics flavor puzzle”, and present the idea of minimal flavor violation as a possible solution. We explain the “standard model flavor puzzle”, and present the Froggatt-Nielsen mechanism as a possible solution. We show that measurements of the Higgs boson decays may provide new opportunities for making progress on the various flavor puzzles. We briefly discuss two sets of measurements and their possible theoretical implications:  $\text{BR}(h \rightarrow \tau\mu)$  and  $R(D^{(*)})$ .

## Keywords

Lectures; flavors; CP violation; CKM matrix; flavor changing neutral current; lepton flavor universality

## 1 Introduction

### 1.1 What is flavor?

The term “**flavors**” is used, in the jargon of particle physics, to describe several copies of the same gauge representation, namely several fields that are assigned the same quantum charges. Within the Standard Model, when thinking of its unbroken  $SU(3)_C \times U(1)_{\text{EM}}$  gauge group, there are four different types of particles, each coming in three flavors:

- Up-type quarks in the  $(3)_{+2/3}$  representation:  $u, c, t$ ;
- Down-type quarks in the  $(3)_{-1/3}$  representation:  $d, s, b$ ;
- Charged leptons in the  $(1)_{-1}$  representation:  $e, \mu, \tau$ ;
- Neutrinos in the  $(1)_0$  representation:  $\nu_1, \nu_2, \nu_3$ .

The term “**flavor physics**” refers to interactions that distinguish between flavors. By definition, gauge interactions, namely interactions that are related to unbroken symmetries and mediated therefore by massless gauge bosons, do not distinguish among the flavors and do not constitute part of flavor physics. Within the Standard Model, flavor-physics refers to the weak and Yukawa interactions.

The term “**flavor parameters**” refers to parameters that carry flavor indices. Within the Standard Model, these are the nine masses of the charged fermions and the four “mixing parameters” (three angles and one phase) that describe the interactions of the charged weak-force carriers ( $W^\pm$ ) with quark-antiquark pairs. If one augments the Standard Model with Majorana mass terms for the neutrinos, one should add to the list three neutrino masses and six mixing parameters (three angles and three phases) for the  $W^\pm$  interactions with lepton-antilepton pairs.

The term “**flavor universal**” refers to interactions with couplings (or to parameters) that are proportional to the unit matrix in flavor space. Thus, the strong and electromagnetic interactions are flavor-

universal.<sup>1</sup> An alternative term for “flavor-universal” is “**flavor-blind**”.

The term “**flavor diagonal**” refers to interactions with couplings (or to parameters) that are diagonal, but not necessarily universal, in the flavor space. Within the Standard Model, the Yukawa interactions of the Higgs particle are flavor diagonal in the mass basis.

The term “**flavor changing**” refers to processes where the initial and final flavor-numbers (that is, the number of particles of a certain flavor minus the number of anti-particles of the same flavor) are different. In “flavor changing charged current” (FCCC) processes, both up-type and down-type flavors, and/or both charged lepton and neutrino flavors are involved. Examples are (i) muon decay via  $\mu \rightarrow e\bar{\nu}_e\nu_\mu$ , (ii)  $K^- \rightarrow \mu^-\bar{\nu}_\mu$  (which corresponds, at the quark level, to  $s\bar{u} \rightarrow \mu^-\bar{\nu}_\mu$ ), and (iii)  $B \rightarrow \psi K$  ( $b \rightarrow c\bar{c}s$ ). Within the Standard Model, these processes are mediated by the  $W$ -bosons and occur at tree level. In “**flavor changing neutral current**” (FCNC) processes, either up-type or down-type flavors but not both, and/or either charged lepton or neutrino flavors but not both, are involved. Example are (i) muon decay via  $\mu \rightarrow e\gamma$ , (ii)  $K_L \rightarrow \mu^+\mu^-$  (which corresponds, at the quark level, to  $s\bar{d} \rightarrow \mu^+\mu^-$ ), and (iii)  $B \rightarrow \phi K$  ( $b \rightarrow s\bar{s}s$ ). Within the Standard Model, these processes do not occur at tree level, and are often highly suppressed.

Another useful term is “**flavor violation**”. We will explain it later in these lectures.

## 1.2 Why is flavor physics interesting?

Flavor physics is interesting, on one hand, as a tool for discovery and, on the other hand, because of intrinsic puzzling features:

- Flavor physics can discover new physics or probe it before it is directly observed in experiments. Here are some examples from the past:
  - The smallness of  $\frac{\Gamma(K_L \rightarrow \mu^+\mu^-)}{\Gamma(K^+ \rightarrow \mu^+\nu)}$  led to predicting a fourth (the charm) quark;
  - The size of  $\Delta m_K$  led to a successful prediction of the charm mass;
  - The size of  $\Delta m_B$  led to a successful prediction of the top mass;
  - The measurement of  $\varepsilon_K$  led to predicting the third generation;
  - The measurement of neutrino flavor transitions led to the discovery of neutrino masses.
- CP violation is closely related to flavor physics. Within the Standard Model, there is a single CP violating parameter, the Kobayashi-Maskawa phase  $\delta_{\text{KM}}$  [2]. Baryogenesis tells us, however, that there must exist new sources of CP violation. Measurements of CP violation in flavor changing processes might provide evidence for such sources.
- The fine-tuning problem of the Higgs mass, and the puzzle of the dark matter imply that there exists new physics at, or below, the TeV scale. If such new physics had a generic flavor structure, it would contribute to flavor changing neutral current (FCNC) processes orders of magnitude above the observed rates. The question of why this does not happen constitutes the *new physics flavor puzzle*.
- Most of the charged fermion flavor parameters are small and hierarchical. The Standard Model does not provide any explanation of these features. This is the *Standard Model flavor puzzle*. The puzzle became even deeper after neutrino masses and mixings were measured because, so far, neither smallness nor hierarchy in these parameters have been established.

## 2 The Standard Model

A model of elementary particles and their interactions is defined by the following ingredients: (i) The symmetries of the Lagrangian and the pattern of spontaneous symmetry breaking (SSB); (ii) The repre-

<sup>1</sup>In the interaction basis, the weak interactions are also flavor-universal, and one can identify the source of all flavor physics in the Yukawa interactions among the gauge-interaction eigenstates.

sentations of fermions and scalars. The Standard Model (SM) is defined as follows:

- The symmetry is a local

$$G_{\text{SM}} = SU(3)_C \times SU(2)_L \times U(1)_Y \quad . \quad (1)$$

- It is spontaneously broken by the VEV of a single Higgs scalar,

$$\phi(1, 2)_{+1/2}, \quad (\langle \phi^0 \rangle = v/\sqrt{2}) \quad , \quad (2)$$

$$G_{\text{SM}} \rightarrow SU(3)_C \times U(1)_{\text{EM}} \quad (Q_{\text{EM}} = T_3 + Y) \quad . \quad (3)$$

- There are three fermion generations, each consisting of five representations of  $G_{\text{SM}}$ :

$$Q_{Li}(3, 2)_{+1/6}, \quad U_{Ri}(3, 1)_{+2/3}, \quad D_{Ri}(3, 1)_{-1/3}, \quad L_{Li}(1, 2)_{-1/2}, \quad E_{Ri}(1, 1)_{-1} \quad . \quad (4)$$

## 2.1 The Lagrangian

The most general renormalizable Lagrangian with scalar and fermion fields can be decomposed into

$$\mathcal{L} = \mathcal{L}_{\text{kin}} + \mathcal{L}_{\psi} + \mathcal{L}_{\text{Yuk}} + \mathcal{L}_{\phi} \quad . \quad (5)$$

Here  $\mathcal{L}_{\text{kin}}$  describes free propagation in spacetime, as well as gauge interactions,  $\mathcal{L}_{\psi}$  gives fermion mass terms,  $\mathcal{L}_{\text{Yuk}}$  describes the Yukawa interactions, and  $\mathcal{L}_{\phi}$  gives the scalar potential. We now find the specific form of the Lagrangian made of the fermion fields  $Q_{Li}$ ,  $U_{Ri}$ ,  $D_{Ri}$ ,  $L_{Li}$  and  $E_{Ri}$  (4), and the scalar field (2), subject to the gauge symmetry (1) and leading to the SSB of Eq. (3).

### 2.1.1 $\mathcal{L}_{\text{kin}}$

The local symmetry requires the following gauge boson degrees of freedom:

$$G_a^\mu(8, 1)_0, \quad W_a^\mu(1, 3)_0, \quad B^\mu(1, 1)_0 \quad . \quad (6)$$

The corresponding field strengths are given by

$$\begin{aligned} G_a^{\mu\nu} &= \partial^\mu G_a^\nu - \partial^\nu G_a^\mu - g_s f_{abc} G_b^\mu G_c^\nu \quad , \\ W_a^{\mu\nu} &= \partial^\mu W_a^\nu - \partial^\nu W_a^\mu - g \epsilon_{abc} W_b^\mu W_c^\nu \quad , \\ B^{\mu\nu} &= \partial^\mu B^\nu - \partial^\nu B^\mu \quad . \end{aligned} \quad (7)$$

The covariant derivative is

$$D^\mu = \partial^\mu + ig_s G_a^\mu L_a + ig W_b^\mu T_b + ig' B^\mu Y \quad , \quad (8)$$

where the  $L_a$ 's are  $SU(3)_C$  generators (the  $3 \times 3$  Gell-Mann matrices  $\frac{1}{2}\lambda_a$  for triplets, 0 for singlets), the  $T_b$ 's are  $SU(2)_L$  generators (the  $2 \times 2$  Pauli matrices  $\frac{1}{2}\tau_b$  for doublets, 0 for singlets), and the  $Y$ 's are the  $U(1)_Y$  charges. Explicitly, the covariant derivatives acting on the various scalar and fermion fields are given by

$$\begin{aligned} D^\mu \phi &= \left( \partial^\mu + \frac{i}{2} g W_b^\mu \tau_b + \frac{i}{2} g' B^\mu \right) \phi \quad , \\ D^\mu Q_{Li} &= \left( \partial^\mu + \frac{i}{2} g_s G_a^\mu \lambda_a + \frac{i}{2} g W_b^\mu \tau_b + \frac{i}{6} g' B^\mu \right) Q_{Li} \quad , \\ D^\mu U_{Ri} &= \left( \partial^\mu + \frac{i}{2} g_s G_a^\mu \lambda_a + \frac{2i}{3} g' B^\mu \right) U_{Ri} \quad , \end{aligned}$$

$$\begin{aligned}
D^\mu D_{Ri} &= \left( \partial^\mu + \frac{i}{2} g_s G_a^\mu \lambda_a - \frac{i}{3} g' B^\mu \right) D_{Ri} \quad , \\
D^\mu L_{Li} &= \left( \partial^\mu + \frac{i}{2} g W_b^\mu \tau_b - \frac{i}{2} g' B^\mu \right) L_{Li} \quad , \\
D^\mu E_{Ri} &= (\partial^\mu - i g' B^\mu) E_{Ri} \quad .
\end{aligned} \tag{9}$$

$\mathcal{L}_{\text{kin}}$  is given by

$$\begin{aligned}
\mathcal{L}_{\text{kin}}^{\text{SM}} &= -\frac{1}{4} G_a^{\mu\nu} G_{a\mu\nu} - \frac{1}{4} W_b^{\mu\nu} W_{b\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} \\
&\quad - i \overline{Q_{Li}} \not{D} Q_{Li} - i \overline{U_{Ri}} \not{D} U_{Ri} - i \overline{D_{Ri}} \not{D} D_{Ri} - i \overline{L_{Li}} \not{D} L_{Li} - i \overline{E_{Ri}} \not{D} E_{Ri} \\
&\quad - (D^\mu \phi)^\dagger (D_\mu \phi) \quad .
\end{aligned} \tag{10}$$

This part of the interaction Lagrangian is flavor-universal. In addition, it conserves CP.

### 2.1.2 $\mathcal{L}_\psi$

There are no mass terms for the fermions in the SM. We cannot write Dirac mass terms for the fermions because they are assigned to chiral representations of the gauge symmetry. We cannot write Majorana mass terms for the fermions because they all have  $Y \neq 0$ . Thus,

$$\mathcal{L}_\psi^{\text{SM}} = 0 \quad . \tag{11}$$

### 2.1.3 $\mathcal{L}_{\text{Yuk}}$

The Yukawa part of the Lagrangian is given by

$$\mathcal{L}_Y^{\text{SM}} = Y_{ij}^d \overline{Q_{Li}} \phi D_{Rj} + Y_{ij}^u \overline{Q_{Li}} \tilde{\phi} U_{Rj} + Y_{ij}^e \overline{L_{Li}} \phi E_{Rj} \text{h.c.} \quad , \tag{12}$$

where  $\tilde{\phi} = i\tau_2 \phi^\dagger$ , and the  $Y^f$  are general  $3 \times 3$  matrices of dimensionless couplings. This part of the Lagrangian is, in general, flavor-dependent (that is,  $Y^f \not\propto \mathbf{1}$ ) and CP violating.

Without loss of generality, we can use a bi-unitary transformation,

$$Y^e \rightarrow \hat{Y}_e = U_{eL} Y^e U_{eR}^\dagger \quad , \tag{13}$$

to change the basis to one where  $Y^e$  is diagonal and real:

$$\hat{Y}_e = \text{diag}(y_e, y_\mu, y_\tau) \quad . \tag{14}$$

In the basis defined in Eq. (14), we denote the components of the lepton  $SU(2)$ -doublets, and the three lepton  $SU(2)$ -singlets, as follows:

$$\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \quad \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \quad \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix}; \quad e_R, \quad \mu_R, \quad \tau_R, \tag{15}$$

where  $e, \mu, \tau$  are ordered by the size of  $y_{e,\mu,\tau}$  (from smallest to largest).

Similarly, without loss of generality, we can use a bi-unitary transformation,

$$Y^u \rightarrow \hat{Y}_u = V_{uL} Y^u V_{uR}^\dagger \quad , \tag{16}$$

to change the basis to one where  $\hat{Y}_u$  is diagonal and real:

$$\hat{Y}_u = \text{diag}(y_u, y_c, y_t) \quad . \tag{17}$$

In the basis defined in Eq. (17), we denote the components of the quark  $SU(2)$ -doublets, and the quark up  $SU(2)$ -singlets, as follows:

$$\begin{pmatrix} u_L \\ d_{uL} \end{pmatrix}, \quad \begin{pmatrix} c_L \\ d_{cL} \end{pmatrix}, \quad \begin{pmatrix} t_L \\ d_{tL} \end{pmatrix}; \quad u_R, \quad c_R, \quad t_R, \quad (18)$$

where  $u, c, t$  are ordered by the size of  $y_{u,c,t}$  (from smallest to largest).

We can use yet another bi-unitary transformation,

$$Y^d \rightarrow \hat{Y}_d = V_{dL} Y^d V_{dR}^\dagger, \quad (19)$$

to change the basis to one where  $\hat{Y}^d$  is diagonal and real:

$$\hat{Y}^d = \text{diag}(y_d, y_s, y_b). \quad (20)$$

In the basis defined in Eq. (20), we denote the components of the quark  $SU(2)$ -doublets, and the quark down  $SU(2)$ -singlets, as follows:

$$\begin{pmatrix} u_{dL} \\ d_L \end{pmatrix}, \quad \begin{pmatrix} u_{sL} \\ s_L \end{pmatrix}, \quad \begin{pmatrix} u_{bL} \\ b_L \end{pmatrix}; \quad d_R, \quad s_R, \quad b_R, \quad (21)$$

where  $d, s, b$  are ordered by the size of  $y_{d,s,b}$  (from smallest to largest).

Note that if  $V_{uL} \neq V_{dL}$ , as is the general case, then the interaction basis defined by (17) is different from the interaction basis defined by (20). In the former,  $Y^d$  can be written as a unitary matrix times a diagonal one,

$$Y^u = \hat{Y}^u, \quad Y^d = V \hat{Y}^d. \quad (22)$$

In the latter,  $Y^u$  can be written as a unitary matrix times a diagonal one,

$$Y^d = \hat{Y}^d, \quad Y^u = V^\dagger \hat{Y}^u. \quad (23)$$

In either case, the matrix  $V$  is given by

$$V = V_{uL} V_{dL}^\dagger, \quad (24)$$

where  $V_{uL}$  and  $V_{dL}$  are defined in Eqs. (16) and (19), respectively. Note that  $V_{uL}$ ,  $V_{uR}$ ,  $V_{dL}$  and  $V_{dR}$  depend on the basis from which we start the diagonalization. The combination  $V = V_{uL} V_{dL}^\dagger$ , however, does not. This is a hint that  $V$  is physical. Indeed, below we see that it plays a crucial role in the charged current interactions.

#### 2.1.4 $\mathcal{L}_\phi$

The scalar potential is given by

$$\mathcal{L}_\phi^{\text{SM}} = -\mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2. \quad (25)$$

Choosing  $\mu^2 < 0$  and  $\lambda > 0$  leads to the required spontaneous symmetry breaking. This part of the Lagrangian is also CP conserving.

## 2.2 The spectrum

The spectrum of the standard model is presented in Table 1.

All masses are proportional to the VEV of the scalar field,  $v$ . For the three massive gauge bosons, and for the fermions, this is expected: In the absence of spontaneous symmetry breaking, the former would be protected by the gauge symmetry and the latter by their chiral nature. For the Higgs boson, the situation is different, as a mass-squared term does not violate any symmetry.

**Table 1:** The SM particles

particle	spin	color	$Q_{\text{EM}}$	mass [ $v$ ]
$W^\pm$	1	(1)	$\pm 1$	$\frac{1}{2}g$
$Z^0$	1	(1)	0	$\frac{1}{2}\sqrt{g^2 + g'^2}$
$A^0$	1	(1)	0	0
$g$	1	(8)	0	0
$h$	0	(1)	0	$\sqrt{2\lambda}$
$e, \mu, \tau$	1/2	(1)	-1	$y_{e,\mu,\tau}/\sqrt{2}$
$\nu_e, \nu_\mu, \nu_\tau$	1/2	(1)	0	0
$u, c, t$	1/2	(3)	+2/3	$y_{u,c,t}/\sqrt{2}$
$d, s, b$	1/2	(3)	-1/3	$y_{d,s,b}/\sqrt{2}$

**Table 2:** The SM fermion interactions

interaction	fermions	force carrier	coupling	flavor
Electromagnetic	$u, d, \ell$	$A^0$	$eQ$	universal
Strong	$u, d$	$g$	$g_s$	universal
NC weak	all	$Z^0$	$\frac{e(T_3 - s_W^2 Q)}{s_W c_W}$	universal
CC weak	$\bar{u}d/\bar{\ell}\nu$	$W^\pm$	$gV/g$	non-universal/universal
Yukawa	$u, d, \ell$	$h$	$y_q$	diagonal

For the charged fermions, the spontaneous symmetry breaking allows their masses because they are in vector-like representations of the  $SU(3)_C \times U(1)_{\text{EM}}$  group: The LH and RH charged lepton fields,  $e, \mu$  and  $\tau$ , are in the  $(1)_{-1}$  representation; The LH and RH up-type quark fields,  $u, c$  and  $t$ , are in the  $(3)_{+2/3}$  representation; The LH and RH down-type quark fields,  $d, s$  and  $b$ , are in the  $(3)_{-1/3}$  representation. On the other hand, the neutrinos remain massless in spite of the fact that they are in the  $(1)_0$  representation of  $SU(3)_C \times U(1)_{\text{EM}}$ , which allows for Majorana masses. Such masses require a VEV carried by a scalar field in the  $(1, 3)_{+1}$  representation of the  $SU(3)_C \times SU(2)_L \times U(1)_Y$  symmetry, but there is no such field in the SM.

The experimental values of the charged fermion masses are [1]<sup>2</sup>

$$\begin{aligned}
m_e &= 0.510998946(3) \text{ MeV} , \quad m_\mu = 105.6583745(24) \text{ MeV} , \quad m_\tau = 1776.86(12) \text{ MeV} , \\
m_u &= 2.2_{-0.4}^{+0.5} \text{ MeV} , \quad m_c = 1.275_{-0.035}^{+0.025} \text{ GeV} , \quad m_t = 173.1 \pm 0.9 \text{ GeV} , \\
m_d &= 4.7_{-0.3}^{+0.5} \text{ MeV} , \quad m_s = 95_{-3}^{+9} \text{ MeV} , \quad m_b = 4.18_{-0.03}^{+0.04} \text{ GeV} .
\end{aligned} \tag{26}$$

### 2.3 The interactions

Within the SM, the fermions have five types of interactions. These interactions are summarized in Table 2. In the next few subsections, we explain the entries of this table.

<sup>2</sup>See [1] for detailed explanations of the quoted quark masses. For  $q = u, d, s, c, b$ ,  $m_q$  are the running quark masses in the  $\overline{\text{MS}}$  scheme, with  $m_{u,d,s} = m_{u,d,s}(\mu = 2 \text{ GeV})$  and  $m_{c,b} = m_{c,b}(\mu = m_{c,b})$ .

### 2.3.1 EM and strong interactions

By construction, a local  $SU(3)_C \times U(1)_{EM}$  symmetry survives the SSB. The SM has thus the photon and gluon massless gauge fields. All charged fermions interact with the photon:

$$\mathcal{L}_{QED,\psi} = -\frac{2e}{3}\bar{u}_i \not{A} u_i + \frac{e}{3}\bar{d}_i \not{A} d_i + e\bar{\ell}_i \not{A} \ell_i \quad , \quad (27)$$

where  $u_{1,2,3} = u, c, t$ ,  $d_{1,2,3} = d, s, b$  and  $\ell_{1,2,3} = e, \mu, \tau$ . We emphasize the following points:

1. The photon couplings are *vector-like* and *parity conserving*.
2. *Diagonality*: The photon couples to  $e^+e^-$ ,  $\mu^+\mu^-$  and  $\tau^+\tau^-$ , but not to  $e^\pm\mu^\mp$ ,  $e^\pm\tau^\mp$  or  $\mu^\pm\tau^\mp$  pairs, and similarly in the up and down sectors.
3. *Universality*: The couplings of the photon to different generations are universal.

All colored fermions (namely, quarks) interact with the gluon:

$$\mathcal{L}_{QCD,\psi} = -\frac{g_s}{2}\bar{q}\lambda_a \not{G}_a q \quad , \quad (28)$$

where  $q = u, c, t, d, s, b$ . We emphasize the following points:

1. The gluon couplings are *vector-like* and *parity conserving*.
2. *Diagonality*: The gluon couples to  $\bar{t}t$ ,  $\bar{c}c$ , etc., but not to  $\bar{t}c$  or any other flavor changing pair.
3. *Universality*: The couplings of the gluon to different quark generations are universal.

The universality of the photon and gluon couplings are a result of the  $SU(3)_C \times U(1)_{EM}$  gauge invariance, and thus hold in any model, and not just within the SM.

### 2.3.2 Z-mediated weak interactions

All SM fermions couple to the Z-boson:

$$\begin{aligned} \mathcal{L}_{Z,\psi} = & \frac{e}{s_W c_W} \left[ -\left(\frac{1}{2} - s_W^2\right) \bar{e}_{Li} \not{Z} e_{Li} + s_W^2 \bar{e}_{Ri} \not{Z} e_{Ri} + \frac{1}{2} \bar{\nu}_{L\alpha} \not{Z} \nu_{L\alpha} \right. \\ & \left. + \left(\frac{1}{2} - \frac{2}{3}s_W^2\right) \bar{u}_{Li} \not{Z} u_{Li} - \frac{2}{3}s_W^2 \bar{u}_{Ri} \not{Z} u_{Ri} - \left(\frac{1}{2} - \frac{1}{3}s_W^2\right) \bar{d}_{Li} \not{Z} d_{Li} + \frac{1}{3}s_W^2 \bar{d}_{Ri} \not{Z} d_{Ri} \right] \quad . \end{aligned} \quad (29)$$

where  $\nu_\alpha = \nu_e, \nu_\mu, \nu_\tau$ . We emphasize the following points:

1. The Z-boson couplings are *chiral* and *parity violating*.
2. *Diagonality*: The Z-boson couples diagonally and, as a result of this, there are no Z-mediated flavor changing neutral current (FCNC) processes.
3. *Universality*: The couplings of the Z-boson to different fermion generations are universal.

The universality is a result of a special feature of the SM: all fermions of given chirality and given charge come from the same  $SU(2)_L \times U(1)_Y$  representation.

As an example to experimental tests of diagonality and universality, we can take the leptonic sector. The branching ratios of the Z-boson into charged lepton pairs [1],

$$\begin{aligned} \text{BR}(Z \rightarrow e^+e^-) &= (3.363 \pm 0.004)\% \quad , \\ \text{BR}(Z \rightarrow \mu^+\mu^-) &= (3.366 \pm 0.007)\% \quad , \\ \text{BR}(Z \rightarrow \tau^+\tau^-) &= (3.367 \pm 0.008)\% \quad . \end{aligned} \quad (30)$$

beautifully confirms universality:

$$\begin{aligned}\Gamma(\mu^+\mu^-)/\Gamma(e^+e^-) &= 1.0009 \pm 0.0028 \quad , \\ \Gamma(\tau^+\tau^-)/\Gamma(e^+e^-) &= 1.0019 \pm 0.0032 \quad .\end{aligned}$$

Diagonality is also tested by the following experimental searches:

$$\begin{aligned}\text{BR}(Z \rightarrow e^+\mu^-) &< 7.5 \times 10^{-7} \quad , \\ \text{BR}(Z \rightarrow e^+\tau^-) &< 9.8 \times 10^{-6} \quad , \\ \text{BR}(Z \rightarrow \mu^+\tau^-) &< 1.2 \times 10^{-5} \quad .\end{aligned}\tag{31}$$

### 2.3.3 $W$ -mediated weak interactions

We now study the couplings of the charged vector bosons,  $W^\pm$ , to fermion pairs. For the lepton mass eigenstates, things are simple, because there exists an interaction basis that is also a mass basis. Thus,

$$\mathcal{L}_{W,\ell} = -\frac{g}{\sqrt{2}} (\bar{\nu}_{eL} W^+ e_L^- + \bar{\nu}_{\mu L} W^+ \mu_L^- + \bar{\nu}_{\tau L} W^+ \tau_L^- + \text{h.c.}) \quad .\tag{32}$$

Eq. (32) reveals some important features of the model:

1. Only left-handed particles take part in charged-current interactions. Consequently, parity is violated.
2. *Diagonality*: the charged current interactions couple each charged lepton to a single neutrino, and each neutrino to a single charged lepton. Note that a global  $SU(2)$  symmetry would allow off-diagonal couplings; It is the local symmetry that leads to diagonality.
3. *Universality*: the couplings of the  $W$ -boson to  $\tau\bar{\nu}_\tau$ , to  $\mu\bar{\nu}_\mu$  and to  $e\bar{\nu}_e$  are equal. Again, a global symmetry would have allowed an independent coupling to each lepton pair.

All of these predictions have been experimentally tested. As an example of how well universality works, consider the decay rates of the  $W$ -bosons to the three lepton pairs [1]:

$$\begin{aligned}\text{BR}(W^+ \rightarrow e^+\nu_e) &= (10.71 \pm 0.16) \times 10^{-2} \quad , \\ \text{BR}(W^+ \rightarrow \mu^+\nu_\mu) &= (10.63 \pm 0.15) \times 10^{-2} \quad , \\ \text{BR}(W^+ \rightarrow \tau^+\nu_\tau) &= (11.38 \pm 0.21) \times 10^{-2} \quad .\end{aligned}\tag{33}$$

You must be impressed by the nice agreement!

As concerns quarks, things are more complicated, since there is no interaction basis that is also a mass basis. In the interaction basis where the down quarks are mass eigenstates (21), the  $W$  interactions have the following form:

$$\mathcal{L}_{W,q} = -\frac{g}{\sqrt{2}} (\bar{u}_{dL} W^+ d_L + \bar{u}_{sL} W^+ s_L + \bar{u}_{bL} W^+ b_L + \text{h.c.}) \quad .\tag{34}$$

The Yukawa matrices in this basis have the form (23), and in particular, for the up sector, we have

$$\mathcal{L}_{\text{Yuk}}^u = (\bar{u}_{dL} \bar{u}_{sL} \bar{u}_{bL}) V^\dagger \hat{Y}^u \begin{pmatrix} u_R \\ c_R \\ t_R \end{pmatrix} \quad ,\tag{35}$$

which tells us straightforwardly how to transform to the mass basis:

$$\begin{pmatrix} u_L \\ c_L \\ t_L \end{pmatrix} = V \begin{pmatrix} u_{dL} \\ u_{sL} \\ u_{bL} \end{pmatrix} \quad .\tag{36}$$

Using Eq. (36), we obtain the form of the  $W$  interactions (34) in the mass basis:

$$-\frac{g}{\sqrt{2}} (\bar{u}_L \bar{c}_L \bar{t}_L) V W^+ \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} + \text{h.c.} \quad . \quad (37)$$

You can easily convince yourself that we would have obtained the same form starting from any arbitrary interaction basis. We remind you that

$$V = V_{uL} V_{dL}^\dagger \quad (38)$$

is basis independent. The matrix  $V$  is called the CKM matrix [2, 3].

Similarly to the leptons, only left-handed quarks take part in charged-current interactions and, consequently, parity is violated by these interactions. But then there is an important difference:

1. The  $W$  couplings to the quark mass eigenstates are neither universal nor diagonal. The universality of gauge interactions is hidden in the unitarity of the matrix  $V$ .

Omitting common factors (particularly, a factor of  $g^2/4$ ) and phase space factors, we obtain the following predictions for the  $W$  decays:

$$\begin{aligned} \Gamma(W^+ \rightarrow \ell^+ \nu_\ell) &\propto 1 \quad , \\ \Gamma(W^+ \rightarrow u_i \bar{d}_j) &\propto 3|V_{ij}|^2 \quad (i = 1, 2; j = 1, 2, 3) \quad . \end{aligned} \quad (39)$$

The top quark is not included because it is heavier than the  $W$  boson. Taking this fact into account, and the CKM unitarity relations

$$|V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2 = |V_{cd}|^2 + |V_{cs}|^2 + |V_{cb}|^2 = 1 \quad , \quad (40)$$

we obtain

$$\Gamma(W \rightarrow \text{hadrons})/\Gamma(W \rightarrow \text{leptons}) \approx 2 \quad . \quad (41)$$

Experimentally

$$\text{BR}((W \rightarrow \text{leptons})) = (32.40 \pm 0.27)\% \quad \text{BR}((W \rightarrow \text{hadrons})) = (67.60 \pm 0.27)\% \quad , \quad (42)$$

which leads to

$$\Gamma(W \rightarrow \text{hadrons})/\Gamma(W \rightarrow \text{leptons}) = 2.09 \pm 0.1 \quad , \quad (43)$$

which, taking into account radiative corrections, is in beautiful agreement with the SM prediction. The (hidden) universality within the quark sector is tested by the prediction

$$\Gamma(W \rightarrow uX) = \Gamma(W \rightarrow cX) = \frac{1}{2}\Gamma(W \rightarrow \text{hadrons}) \quad . \quad (44)$$

Experimentally,

$$\Gamma(W \rightarrow cX)/\Gamma(W \rightarrow \text{hadrons}) = 0.49 \pm 0.04 \quad . \quad (45)$$

### 2.3.4 Yukawa interactions

The Yukawa interactions are given by

$$\begin{aligned} \mathcal{L}_{\text{Yuk}} = & - \frac{h}{v} (m_e \bar{e}_L e_R + m_\mu \bar{\mu}_L \mu_R + m_\tau \bar{\tau}_L \tau_R \\ & + m_u \bar{u}_L u_R + m_c \bar{c}_L c_R + m_t \bar{t}_L t_R + m_d \bar{d}_L d_R + m_s \bar{s}_L s_R + m_b \bar{b}_L b_R + \text{h.c.}) \quad . \end{aligned}$$

**Table 3:** Higgs decays: The SM predictions for the branching ratios, and the experimental  $\mu$  values

Mode	$\text{BR}_{\text{SM}}$	$\mu_{\text{experiment}}$	Comments
$b\bar{b}$	0.58	$0.98 \pm 0.20$	
$WW^*$	0.21	$0.99 \pm 0.15$	3-body
$gg$	0.09		loop
$\tau^+\tau^-$	0.06	$1.09 \pm 0.23$	
$ZZ^*$	0.03	$1.17 \pm 0.23$	3-body
$c\bar{c}$	0.03		
$\gamma\gamma$	0.002	$1.14 \pm 0.14$	loop

To see that the Higgs boson couples diagonally to the quark mass eigenstates, let us start from an arbitrary interaction basis:

$$\begin{aligned}
h\overline{D}_L Y^d D_R &= h\overline{D}_L (V_{dL}^\dagger V_{dL}) Y^d (V_{dR}^\dagger V_{dR}) D_R \\
&= h(\overline{D}_L V_{dL}^\dagger) (V_{dL} Y^d V_{dR}^\dagger) (V_{dR} D_R) \\
&= h(\overline{d}_L \overline{s}_L \overline{b}_L) \hat{Y}^d (d_R s_R b_R)^T .
\end{aligned} \tag{46}$$

We conclude that the Higgs couplings to the fermion mass eigenstates have the following features:

1. *Diagonality.*
2. *Non-universality.*
3. *Proportionality* to the fermion masses: the heavier the fermion, the stronger the coupling. The factor of proportionality is  $m_\psi/v$ .

Thus, the Higgs boson decay is dominated by the heaviest particle which can be pair-produced in the decay. For  $m_h \sim 125$  GeV, this is the bottom quark. Indeed, the SM predicts the following branching ratios quoted in Table 3 for the leading decay modes. The following comments are in order with regard to the predicted branching ratios:

1. From the seven branching ratios, three ( $b, \tau, c$ ) stand for two-body tree-level decays. Thus, at tree level, the respective branching ratios obey  $\text{BR}_{b\bar{b}} : \text{BR}_{\tau^+\tau^-} : \text{BR}_{c\bar{c}} = 3m_b^2 : m_\tau^2 : 3m_c^2$ . QCD radiative corrections somewhat suppress the two modes with the quark final states ( $b, c$ ) compared to one with the lepton final state ( $\tau$ ).
2. The  $WW^*$  and  $ZZ^*$  modes stand for the three-body tree-level decays, where one of the vector bosons is on-shell and the other off-shell.
3. The Higgs boson does not have a tree-level coupling to gluons since it carries no color (and the gluons have no mass). The decay into final gluons proceeds via loop diagrams. The dominant contribution comes from the top-quark loop.
4. Similarly, the Higgs decays into final two photons via loop diagrams with small ( $\text{BR}_{\gamma\gamma} \sim 0.002$ ), but observable, rate. The dominant contributions come from the  $W$  and the top-quark loops which interfere destructively.

Experimentally, the decays into final  $ZZ^*$ ,  $WW^*$ ,  $\gamma\gamma$ ,  $b\bar{b}$  and  $\tau^+\tau^-$  have been established.

## 2.4 Global symmetries

The SM has an accidental global symmetry:

$$G_{\text{global}}^{\text{SM}}(Y^{u,d,e} \neq 0) = U(1)_B \times U(1)_e \times U(1)_\mu \times U(1)_\tau . \tag{47}$$

This symmetry leads to various testable predictions. Here are a few examples:

- The proton must not decay, *e.g.*  $p \rightarrow e^+ \pi$  is forbidden.
- FCNC decays of charged leptons must not occur, *e.g.*  $\mu \rightarrow e \gamma$  is forbidden.
- Neutrinos are massless,  $m_\nu = 0$ .

The last prediction is, however, violated in Nature. Neutrino flavor transitions are observed, implying that at least two of the neutrino masses are different from zero.

Accidental symmetries are broken by higher-dimensional (non-renormalizable) terms. Two examples are the following:

- At dimension five,  $\frac{z_{ij}^\nu}{\Lambda} L_i L_j \phi \phi$  terms break  $U(1)_e \times U(1)_\mu \times U(1)_\tau$ .
- At dimension six,  $\frac{y_{ijkl}}{\Lambda^2} Q_i Q_j Q_k L_l$  terms break  $U(1)_B$ .

Thus, given that  $m_\nu \neq 0$ , we learn that the SM is, at best, a good low energy effective field theory.

In the absence of the Yukawa matrices,  $\mathcal{L}_{\text{Yuk}} = 0$ , the SM has a large  $U(3)^5$  global symmetry:

$$G_{\text{global}}^{\text{SM}}(Y^{u,d,e} = 0) = SU(3)_q^3 \times SU(3)_\ell^2 \times U(1)^5 \quad , \quad (48)$$

where

$$\begin{aligned} SU(3)_q^3 &= SU(3)_Q \times SU(3)_U \times SU(3)_D \quad , \\ SU(3)_\ell^2 &= SU(3)_L \times SU(3)_E \quad , \\ U(1)^5 &= U(1)_B \times U(1)_L \times U(1)_Y \times U(1)_{\text{PQ}} \times U(1)_{E_R} \quad . \end{aligned} \quad (49)$$

Out of the five  $U(1)$  charges, three can be identified with baryon number ( $B$ ), lepton number ( $L$ ) and hypercharge ( $Y$ ), which are respected by the Yukawa interactions. The two remaining  $U(1)$  groups can be identified with the PQ symmetry whereby the Higgs and  $D_R, E_R$  fields have opposite charges, and with a global rotation of  $E_R$  only.

The point that is important for our purposes is that  $\mathcal{L}_{\text{kin}}$  respects the non-Abelian flavor symmetry  $SU(3)_q^3 \times SU(3)_\ell^2$ , under which

$$Q_L \rightarrow V_Q Q_L \quad , \quad U_R \rightarrow V_U U_R \quad , \quad D_R \rightarrow V_D D_R \quad , \quad L_L \rightarrow V_L L_L \quad , \quad E_R \rightarrow V_E E_R \quad , \quad (50)$$

where the  $V_i$  are unitary matrices. The Yukawa interactions (12) break the global symmetry into the subgroup of Eq. (47). (Of course, the gauged  $U(1)_Y$  also remains a good symmetry.) Thus, the transformations of Eq. (50) are not a symmetry of  $\mathcal{L}_{\text{SM}}$ . Instead, they correspond to a change of the interaction basis. These observations also offer an alternative way of defining flavor physics: it refers to interactions that break the  $SU(3)^5$  symmetry (50). Thus, the term “**flavor violation**” is often used to describe processes or parameters that break the symmetry.

One can think of the quark Yukawa couplings as spurions that break the global  $SU(3)_q^3$  symmetry (but are neutral under  $U(1)_B$ ),

$$Y^u \sim (3, \bar{3}, 1)_{SU(3)_q^3} \quad , \quad Y^d \sim (3, 1, \bar{3})_{SU(3)_q^3} \quad , \quad (51)$$

and of the lepton Yukawa couplings as spurions that break the global  $SU(3)_\ell^2$  symmetry (but are neutral under  $U(1)_e \times U(1)_\mu \times U(1)_\tau$ ),

$$Y^e \sim (3, \bar{3})_{SU(3)_\ell^2} \quad . \quad (52)$$

The spurion formalism is convenient for several purposes: parameter counting (see below), identification of flavor suppression factors (see Section 7), and the idea of minimal flavor violation (see Section 7.2).

## 2.5 Counting parameters

How many independent parameters are there in  $\mathcal{L}_{\text{Yuk}}^q$ ? The two Yukawa matrices,  $Y^u$  and  $Y^d$ , are  $3 \times 3$  and complex. Consequently, there are 18 real and 18 imaginary parameters in these matrices. Not all of them are, however, physical. The pattern of  $G_{\text{global}}$  breaking means that there is freedom to remove 9 real and 17 imaginary parameters (the number of parameters in three  $3 \times 3$  unitary matrices minus the phase related to  $U(1)_B$ ). For example, we can use the unitary transformations  $Q_L \rightarrow V_Q Q_L$ ,  $U_R \rightarrow V_U U_R$  and  $D_R \rightarrow V_D D_R$ , to lead to the following interaction basis:

$$Y^d = \lambda_d, \quad Y^u = V^\dagger \lambda_u \quad , \quad (53)$$

where  $\lambda_{d,u}$  are diagonal,

$$\lambda_d = \text{diag}(y_d, y_s, y_b) \quad , \quad \lambda_u = \text{diag}(y_u, y_c, y_t) \quad , \quad (54)$$

while  $V$  is a unitary matrix that depends on three real angles and one complex phase. We conclude that there are 10 quark flavor parameters: 9 real ones and a single phase. In the mass basis, we identify the nine real parameters as six quark masses and three mixing angles, while the single phase is  $\delta_{\text{KM}}$ .

How many independent parameters are there in  $\mathcal{L}_{\text{Yuk}}^\ell$ ? The Yukawa matrix  $Y^e$  is  $3 \times 3$  and complex. Consequently, there are 9 real and 9 imaginary parameters in this matrix. There is, however, freedom to remove 6 real and 9 imaginary parameters (the number of parameters in two  $3 \times 3$  unitary matrices minus the phases related to  $U(1)^3$ ). For example, we can use the unitary transformations  $L_L \rightarrow V_L L_L$  and  $E_R \rightarrow V_E E_R$ , to lead to the following interaction basis:

$$Y^e = \lambda_e = \text{diag}(y_e, y_\mu, y_\tau) \quad . \quad (55)$$

We conclude that there are 3 real lepton flavor parameters. In the mass basis, we identify these parameters as the three charged lepton masses. We must, however, modify the model when we take into account the evidence for neutrino masses.

## 3 The CKM matrix

Among the SM interactions, the  $W$ -mediated interactions are the only ones that are not diagonal. Consequently, all flavor changing processes depend on the CKM parameters. The fact that there are only four independent CKM parameters, while the number of measured flavor changing processes is much larger, allows for stringent tests of the CKM mechanism for flavor changing processes.

### 3.1 Parametrization of the CKM matrix

The CKM matrix  $V$  is a  $3 \times 3$  unitary matrix. Its form, however, is not unique:

(i) There is freedom in defining  $V$  in that we can permute between the various generations. This freedom is fixed by ordering the up quarks and the down quarks by their masses, *i.e.*  $(u_1, u_2, u_3) \rightarrow (u, c, t)$  and  $(d_1, d_2, d_3) \rightarrow (d, s, b)$ . The elements of  $V$  are written as follows:

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \quad . \quad (56)$$

(ii) There is further freedom in the phase structure of  $V$ . This means that the number of physical parameters in  $V$  is smaller than the number of parameters in a general unitary  $3 \times 3$  matrix which is nine (three real angles and six phases). Let us define  $P_q$  ( $q = u, d$ ) to be diagonal unitary (phase) matrices. Then, if instead of using  $V_{qL}$  and  $V_{qR}$  for the rotations (16) and (19) to the mass basis we use  $\tilde{V}_{qL}$  and

$\tilde{V}_{qR}$ , defined by  $\tilde{V}_{qL} = P_q V_{qL}$  and  $\tilde{V}_{qR} = P_q V_{qR}$ , we still maintain a legitimate mass basis since  $M_q^{\text{diag}}$  remains unchanged by such transformations. However,  $V$  does change:

$$V \rightarrow P_u V P_d^* \quad . \quad (57)$$

This freedom is fixed by demanding that  $V$  has the minimal number of phases. In the three generation case  $V$  has a single phase. (There are five phase differences between the elements of  $P_u$  and  $P_d$  and, therefore, five of the six phases in the CKM matrix can be removed.) This is the Kobayashi-Maskawa phase  $\delta_{\text{KM}}$  which is the single source of CP violation in the quark sector of the Standard Model [2].

The fact that  $V$  is unitary and depends on only four independent physical parameters can be made manifest by choosing a specific parametrization. The standard choice is [78]

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \quad , \quad (58)$$

where  $c_{ij} \equiv \cos \theta_{ij}$  and  $s_{ij} \equiv \sin \theta_{ij}$ . The  $\theta_{ij}$ 's are the three real mixing parameters while  $\delta$  is the Kobayashi-Maskawa phase. The experimental central values of the four parameters are given by

$$s_{12} = 0.225, \quad s_{23} = 0.042, \quad s_{13} = 0.0037, \quad \delta = 74^\circ \quad . \quad (59)$$

Since  $s_{13} \ll s_{23} \ll s_{12} \ll 1$ , it is convenient to choose an approximate expression where this hierarchy is manifest. This is the Wolfenstein parametrization, where the four mixing parameters are  $(\lambda, A, \rho, \eta)$  with  $\lambda = |V_{us}| \approx 0.23$  playing the role of an expansion parameter and  $\eta$  representing the CP violating phase [79, 80]:

$$V = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 - \frac{1}{8}\lambda^4 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda + \frac{1}{2}A^2\lambda^5[1 - 2(\rho + i\eta)] & 1 - \frac{1}{2}\lambda^2 - \frac{1}{8}\lambda^4(1 + 4A^2) & A\lambda^2 \\ A\lambda^3[1 - (1 - \frac{1}{2}\lambda^2)(\rho + i\eta)] & -A\lambda^2 + \frac{1}{2}A\lambda^4[1 - 2(\rho + i\eta)] & 1 - \frac{1}{2}A^2\lambda^4 \end{pmatrix} \quad . \quad (60)$$

The experimental ranges for the four parameters are given by

$$\begin{aligned} \lambda &= 0.2251 \pm 0.0005 \quad , \\ A &= 0.81 \pm 0.03 \quad , \\ \rho &= +0.160 \pm 0.007 \quad , \\ \eta &= +0.350 \pm 0.006 \quad . \end{aligned} \quad (61)$$

### 3.2 Unitarity triangles

A very useful concept is that of the *unitarity triangles*. The unitarity of the CKM matrix leads to various relations among the matrix elements, *e.g.*

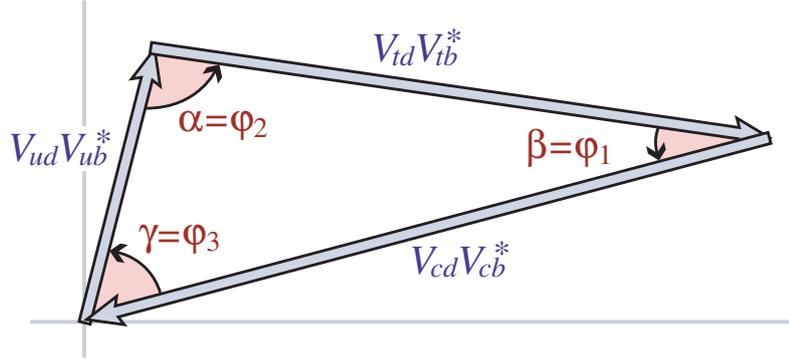
$$V_{ud}V_{us}^* + V_{cd}V_{cs}^* + V_{td}V_{ts}^* = 0 \quad , \quad (62)$$

$$V_{us}V_{ub}^* + V_{cs}V_{cb}^* + V_{ts}V_{tb}^* = 0 \quad , \quad (63)$$

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \quad . \quad (64)$$

Each of these three relations requires the sum of three complex quantities to vanish and so can be geometrically represented in the complex plane as a triangle. These are “the unitarity triangles”, though the term “unitarity triangle” is usually reserved for the relation (64) only. The unitarity triangle related to Eq. (64) is depicted in Fig. 1.

The rescaled unitarity triangle is derived from (64) by (a) choosing a phase convention such that  $(V_{cd}V_{cb}^*)$  is real, and (b) dividing the lengths of all sides by  $|V_{cd}V_{cb}^*|$ . Step (a) aligns one side of the triangle



**Fig. 1:** Graphical representation of the unitarity constraint  $V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0$  as a triangle in the complex plane.

with the real axis, and step (b) makes the length of this side 1. The form of the triangle is unchanged. Two vertices of the rescaled unitarity triangle are thus fixed at  $(0,0)$  and  $(1,0)$ . The coordinates of the remaining vertex correspond to the Wolfenstein parameters  $(\rho, \eta)$ . The area of the rescaled unitarity triangle is  $|\eta|/2$ .

Depicting the rescaled unitarity triangle in the  $(\rho, \eta)$  plane, the lengths of the two complex sides are

$$R_u \equiv \left| \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \right| = \sqrt{\rho^2 + \eta^2} \quad , \quad R_t \equiv \left| \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} \right| = \sqrt{(1-\rho)^2 + \eta^2} \quad . \quad (65)$$

The three angles of the unitarity triangle are defined as follows [81, 82]:

$$\alpha \equiv \arg \left[ -\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*} \right] \quad , \quad \beta \equiv \arg \left[ -\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*} \right] \quad , \quad \gamma \equiv \arg \left[ -\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \right] \quad . \quad (66)$$

They are physical quantities and can be independently measured by CP asymmetries in  $B$  decays. It is also useful to define the two small angles of the unitarity triangles (63,62):

$$\beta_s \equiv \arg \left[ -\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*} \right] \quad , \quad \beta_K \equiv \arg \left[ -\frac{V_{cs}V_{cd}^*}{V_{us}V_{ud}^*} \right] \quad . \quad (67)$$

### 3.3 The CKM matrix from tree level processes

The absolute values of seven entries, and in addition one phase, of the CKM matrix are extracted from tree level processes, see Table 4.

These eight measurements already over-constrain the four Wolfenstein parameters, but the CKM mechanism passes this test successfully. The ranges that are consistent with all tree level measurements are the following:

$$\lambda = 0.2245 \pm 0.0005 \quad , \quad A = 0.84 \pm 0.02 \quad , \quad \rho = 0.14 \pm 0.04 \quad , \quad \eta = 0.37 \pm 0.03 \quad . \quad (68)$$

The  $\lambda$  and  $A$  parameters are very well determined. The main effort in CKM measurements is thus aimed at further improving our knowledge of  $\rho$  and  $\eta$ . The present status of our knowledge is best seen in a plot of the various constraints and the final allowed region in the  $\rho - \eta$  plane. This is shown in Fig. 3. The present status of our knowledge of the absolute values of the various entries in the CKM matrix can be summarized as follows:

$$|V| = \begin{pmatrix} 0.97417 \pm 0.00021 & 0.2248 \pm 0.0006 & (4.1 \pm 0.4) \times 10^{-3} \\ 0.2249 \pm 0.0005 & 0.9735 \pm 0.0001 & (4.05 \pm 0.15) \times 10^{-2} \\ (8.7 \pm 0.3) \times 10^{-3} & (4.03 \pm 0.13) \times 10^{-2} & 0.99915 \pm 0.00005 \end{pmatrix} . \quad (69)$$

**Table 4:** FCCC processes and CKM entries

Process	CKM
$u \rightarrow d\ell^+\nu$	$ V_{ud}  = 0.97417 \pm 0.00021$
$s \rightarrow u\ell^-\bar{\nu}$	$ V_{us}  = 0.2248 \pm 0.0006$
$c \rightarrow d\ell^+\nu$ or $\nu_\mu + d \rightarrow c + \mu^-$	$ V_{cd}  = 0.220 \pm 0.005$
$c \rightarrow s\ell^+\nu$ or $c\bar{s} \rightarrow \ell^+\nu$	$ V_{cs}  = 0.995 \pm 0.016$
$b \rightarrow c\ell^-\bar{\nu}$	$ V_{cb}  = 0.0405 \pm 0.0015$
$b \rightarrow u\ell^-\bar{\nu}$	$ V_{ub}  = 0.0041 \pm 0.0004$
$pp \rightarrow tX$	$ V_{tb}  = 1.01 \pm 0.03$
$b \rightarrow sc\bar{u}$ and $b \rightarrow su\bar{c}$	$\gamma = 73 \pm 5^o$

**Table 5:** Measurements related to neutral meson mixing

Sector	CP-conserving	CP-violating
$sd$	$\Delta m_K/m_K = 7.0 \times 10^{-15}$	$\epsilon_K = 2.3 \times 10^{-3}$
$cu$	$\Delta m_D/m_D = 8.7 \times 10^{-15}$	$A_\Gamma/y_{\text{CP}} \lesssim 0.2$
$bd$	$\Delta m_B/m_B = 6.3 \times 10^{-14}$	$S_{\psi K} = +0.70 \pm 0.02$
$bs$	$\Delta m_{B_s}/m_{B_s} = 2.1 \times 10^{-12}$	$S_{\psi\phi} = -0.04 \pm 0.06$

## 4 Flavor changing neutral current (FCNC) processes

A very useful class of FCNC is that of neutral meson mixing. Nature provides us with four pairs of neutral mesons:  $K^0 - \bar{K}^0$ ,  $B^0 - \bar{B}^0$ ,  $B_s^0 - \bar{B}_s^0$ , and  $D^0 - \bar{D}^0$ . Mixing in this context refers to a transition such as  $K^0 \rightarrow \bar{K}^0$  ( $\bar{s}d \rightarrow \bar{d}s$ ).<sup>3</sup> The experimental results for CP conserving and CP violating observables related to neutral meson mixing (mass splittings and CP asymmetries in tree level decays, respectively) are given in Table 5.

### 4.1 The SM suppression factors

Our aim in this section is to explain the suppression factors that affect FCNC within the SM.

(a) **Loop suppression.** The  $W$ -boson cannot mediate FCNC processes at tree level, since it couples to up-down pairs, or to neutrino-charged lepton pairs. Obviously, only neutral bosons can mediate FCNC at tree level. The SM has four neutral bosons: the gluon, the photon, the  $Z$ -boson and the Higgs-boson. As concerns the massless gauge bosons, the gluon and the photon, their couplings are flavor-universal and, in particular, flavor-diagonal. This is guaranteed by gauge invariance. The universality of the kinetic terms in the canonical basis requires universality of the gauge couplings related to the unbroken symmetries. Hence neither the gluon nor the photon can mediate flavor changing processes at tree level. The situation concerning the  $Z$ -boson and the Higgs-boson is more complicated. In fact, the diagonality of their tree-level couplings is a consequence of special features of the SM, and can be violated with new physics.

The  $Z$ -boson, similarly to the  $W$ -boson, does not correspond to an unbroken gauge symmetry (as manifest in the fact that it is massive). Hence, there is no fundamental symmetry principle that forbids

<sup>3</sup>These transitions involve four-quark operators. When calculating the matrix elements of these operators between meson-antimeson states, approximate symmetries of QCD are of no help. Instead, one uses lattice calculations to relate, for example, the  $B^0 \rightarrow \bar{B}^0$  transition to the corresponding quark process,  $\bar{b}d \rightarrow \bar{d}b$ .

flavor changing couplings. Yet, as mentioned in Section 2.3.2, in the SM this does not happen. The key point is the following. For each sector of mass eigenstates, characterized by spin,  $SU(3)_C$  representation and  $U(1)_{EM}$  charge, there are two possibilities:

1. All mass eigenstates in this sector originate from interaction eigenstates in the same  $SU(2)_L \times U(1)_Y$  representation.
2. The mass eigenstates in this sector mix interaction eigenstates of different  $SU(2)_L \times U(1)_Y$  representations (but, of course, with the same  $T_3 + Y$ ).

Let us examine the  $Z$  couplings in the interaction basis in the subspace of all states that mix within a given sector of mass eigenstates:

1. In the first class, the  $Z$  couplings in this subspace are universal, namely they are proportional to the unit matrix (times  $T_3 - Q \sin^2 \theta_W$  of the relevant interaction eigenstates). The rotation to the mass basis maintains the universality:  $V_{fM} \times \mathbf{1} \times V_{fM}^\dagger = \mathbf{1}$  ( $f = u, d, e; M = L, R$ ).
2. In the second class, the  $Z$  couplings are only “block-universal”. In each sub-block  $i$  of  $m_i$  interaction eigenstates that have the same  $(T_3)_i$ , they are proportional to the  $m_i \times m_i$  unit matrix, but the overall factor of  $(T_3)_i - Q \sin^2 \theta_W$  is different between the sub-blocks. In this case, the rotation to the mass basis,  $V_{fM} \times \text{diag}\{[(T_3)_1 - Qs_W^2]\mathbf{1}_{m_1}, [(T_3)_2 - Qs_W^2]\mathbf{1}_{m_2}, \dots\} \times V_{fM}^\dagger$ , does not maintain the universality, nor even the diagonality.

The special feature of the SM fermions is that they belong to the first class: All fermion mass eigenstates in a given  $SU(3)_C \times U(1)_{EM}$  representation come from the same  $SU(3)_C \times SU(2)_L \times U(1)_Y$  representation.<sup>4</sup> For example, all the left-handed up quark mass eigenstates, which are in the  $(3)_{+2/3}$  representation, come from interaction eigenstates in the  $(3, 2)_{+1/6}$  representation. This is the reason that the SM predicts universal  $Z$  couplings to fermions. If, for example, Nature had left-handed quarks in the  $(3, 1)_{+2/3}$  representation, then the  $Z$  couplings in the left-handed up sector would be non-universal and the  $Z$  could mediate FCNC.

The Yukawa couplings of the Higgs boson are not universal. In fact, in the interaction basis, they are given by completely general  $3 \times 3$  matrices. Yet, as explained in Section 2.3.4, in the fermion mass basis they are diagonal. The reason is that the fermion mass matrix is proportional to the corresponding Yukawa matrix. Consequently, the mass matrix and the Yukawa matrix are simultaneously diagonalized. The special features of the SM in this regard are the following:

1. All the SM fermions are chiral, and therefore there are no bare mass terms.
2. The scalar sector has a single Higgs doublet.

In contrast, either of the following possible extensions would lead to flavor changing Higgs couplings:

1. There are quarks or leptons in vector-like representations, and thus there are bare mass terms.
2. There is more than one  $SU(2)_L$ -doublet scalar.

We conclude that within the SM, all FCNC processes are loop suppressed. However, in extensions of the SM, FCNC can appear at the tree level, mediated by the  $Z$  boson or by the Higgs boson or by new massive bosons.

(b) **CKM suppression.** Obviously, all flavor changing processes are proportional to off-diagonal entries in the CKM matrix. A quick look at the absolute values of the off-diagonal entries of the CKM matrix (69) reveals that they are small. A rough estimate of the CKM suppression can be acquired by

---

<sup>4</sup>This is not true for the SM bosons. The vector boson mass eigenstates in the  $(1)_0$  representation come from interaction eigenstates in the  $(1, 3)_0$  and  $(1, 1)_0$  representations ( $W_3$  and  $B$ , respectively).

counting powers of  $\lambda$  in the Wolfenstein parametrization (60):  $|V_{us}|$  and  $|V_{cd}|$  are suppressed by  $\lambda$ ,  $|V_{cb}|$  and  $|V_{ts}|$  by  $\lambda^2$ ,  $|V_{ub}|$  and  $|V_{td}|$  by  $\lambda^3$ .

For example, the amplitude for  $b \rightarrow s\gamma$  decay comes from penguin diagrams, dominated by the intermediate top quark, and suppressed by  $|V_{tb}V_{ts}| \sim \lambda^2$ . As another example, the  $B^0 - \bar{B}^0$  mixing amplitude comes from box diagrams, dominated by intermediate top quarks, and suppressed by  $|V_{tb}V_{td}|^2 \sim \lambda^6$ .

**(c) GIM suppression.** If all quarks in a given sector were degenerate, then there would be no flavor changing  $W$ -couplings. A consequence of this fact is that FCNC in the down (up) sector are proportional to mass-squared differences between the quarks of the up (down) sector. For FCNC processes that involve only quarks of the first two generations, this leads to a strong suppression factor related to the light quark masses, and known as Glashow-Iliopoulos-Maiani (GIM) suppression.

Let us take as an example  $\Delta m_K$ , the mass splitting between the two neutral  $K$ -mesons. We have  $\Delta m_K = 2|M_{K\bar{K}}|$ , where  $M_{K\bar{K}}$  corresponds to the  $\bar{K}^0 \rightarrow K^0$  transition and comes from box diagrams. The top contribution is CKM-suppressed compared to the contributions from intermediate up and charm, so we consider only the latter:

$$M_{K\bar{K}} \simeq \sum_{i,j=u,c} \frac{G_F^2 m_W^2}{16\pi^2} \langle K^0 | (\bar{d}_L \gamma^\mu s_L)^2 | \bar{K}^0 \rangle (V_{is} V_{id}^* V_{js} V_{jd}^*) \times F(x_i, x_j) \quad , \quad (70)$$

where  $x_i = m_i^2/m_W^2$ . If we had  $m_u = m_c$ , the amplitude would be proportional to  $(V_{us}V_{ud}^* + V_{cs}V_{cd}^*)^2$ , which vanishes in the two generation limit. We conclude that  $\Delta m_K \propto (m_c^2 - m_u^2)/m_W^2$ , which is the GIM suppression factor.

For the  $B^0 - \bar{B}^0$  and  $B_s - \bar{B}_s$  mixing amplitudes, the top-mediated contribution is not CKM suppressed compared to the lighter generations. The mass ratio  $m_t^2/m_W^2$  enhances, rather than suppresses, the top contribution. Consequently, the  $M_{B\bar{B}}$  amplitude is dominated by the top contribution:

$$M_{B\bar{B}} \simeq \frac{G_F^2 m_W^2}{16\pi^2} \langle B^0 | (\bar{d}_L \gamma^\mu b_L)^2 | \bar{B}^0 \rangle (V_{tb} V_{td}^*)^2 \times F(x_t, x_t) \quad . \quad (71)$$

## 4.2 SM1.5: FCNC at tree level

Consider a model with the SM gauge group and pattern of SSB, but with only three quark flavors:  $u, d, s$ . Such a situation cannot fit into a model with all left-handed quarks in doublets of  $SU(2)_L$ . How can we incorporate the interactions of the strange quark in this picture? The solution that we now describe is wrong. Yet, it is of historical significance and, moreover, helps us to understand some of the unique properties of the SM described above. In particular, it leads to FCNC at tree level. We define the three flavor Standard Model (SM1.5) as follows (we ignore the lepton sector):

- The symmetry is a local

$$G_{\text{SM}} = SU(3)_C \times SU(2)_L \times U(1)_Y \quad . \quad (72)$$

- It is spontaneously broken by the VEV of a single Higgs scalar,

$$\phi(1, 2)_{+1/2} \quad , \quad (\langle \phi^0 \rangle = v/\sqrt{2}) \quad , \quad (73)$$

$$G_{\text{SM}} \rightarrow SU(3)_C \times U(1)_{\text{EM}} \quad (Q_{\text{EM}} = T_3 + Y) \quad . \quad (74)$$

- The colored fermion representations are the following:

$$Q_L(3, 2)_{+1/6} \quad , \quad D_L(3, 1)_{-1/3} \quad , \quad U_R(3, 1)_{+2/3} \quad , \quad D_{Ri}(3, 1)_{-1/3} \quad (i = 1, 2) \quad . \quad (75)$$

We point out two important ingredients that are different from the SM:

1. There are quarks in a vector-like representation ( $D_L + D_R$ );
2. Not all  $(3)_{-1/3}$  quarks come from the same type of  $SU(2)_L \times U(1)_Y$  representations.

We first note that  $D_L$  does not couple to the  $W$ -bosons:

$$\mathcal{L}_W = \frac{g}{2} \overline{Q_L} \not{W}_b \tau_b Q_L \quad . \quad (76)$$

The Yukawa interactions are given by

$$\mathcal{L}_{\text{Yuk}} = -y_u \overline{Q_L} \tilde{\phi} U_R - Y_i^d \overline{Q_L} \phi D_{Ri} + \text{h.c.} \quad . \quad (77)$$

Unlike the SM, we now have bare mass terms for fermions:

$$\mathcal{L}_q = -m_{di} \overline{D_L} D_{Ri} + \text{h.c.} \quad . \quad (78)$$

Given that there is a single up generation, the interaction basis is also the up mass basis. Explicitly, we identify the up-component of  $Q_L$  with  $u_L$  (and denote the down component of the doublet as  $d_{uL}$ ), and  $U_R$  with  $u_R$ . With the SSB, we have the following mass terms:

$$-\mathcal{L}_{\text{mass}} = (\overline{d_{uL}} \ \overline{D_L}) \begin{pmatrix} Y_{d1} \frac{v}{\sqrt{2}} & Y_{d2} \frac{v}{\sqrt{2}} \\ m_{d1} & m_{d2} \end{pmatrix} \begin{pmatrix} D_{R1} \\ D_{R2} \end{pmatrix} + y_u \frac{v}{\sqrt{2}} \overline{u_L} u_R + \text{h.c.} \quad . \quad (79)$$

We now rotate to the down mass basis:

$$V_{dL} \begin{pmatrix} Y_{d1} \frac{v}{\sqrt{2}} & Y_{d2} \frac{v}{\sqrt{2}} \\ m_{d1} & m_{d2} \end{pmatrix} V_{dR}^\dagger = \begin{pmatrix} m_d & \\ & m_s \end{pmatrix} \quad . \quad (80)$$

The resulting mixing matrix for the charged current interactions is a  $1 \times 2$  matrix:

$$-\mathcal{L}_{W,q} = \frac{g}{\sqrt{2}} \overline{u_L} \not{W}^+ (\cos \theta_C \ \sin \theta_C) \begin{pmatrix} d_L \\ s_L \end{pmatrix} + \text{h.c.} \quad , \quad (81)$$

where  $\theta_C$  is the rotation angle of  $V_{dL}$ . The neutral current interactions in the left-handed down sector are neither universal nor diagonal:

$$\begin{aligned} \mathcal{L}_{Z,q} &= \frac{g}{c_W} \left[ \left( \frac{1}{2} - \frac{2}{3} s_W^2 \right) \overline{u_L} \not{Z} u_L - \frac{2}{3} s_W^2 \overline{u_R} \not{Z} u_R + \frac{1}{3} s_W^2 (\overline{d_L} \not{Z} d_L + \overline{s_L} \not{Z} s_L + \overline{d_R} \not{Z} d_R + \overline{s_R} \not{Z} s_R) \right] \\ &\quad - \frac{g}{2c_W} (\overline{d_L} \ \overline{s_L}) \not{Z} \begin{pmatrix} \cos^2 \theta_C & \cos \theta_C \sin \theta_C \\ \cos \theta_C \sin \theta_C & \sin^2 \theta_C \end{pmatrix} \begin{pmatrix} d_L \\ s_L \end{pmatrix} \quad . \quad (82) \end{aligned}$$

The Higgs interactions in the down sector are neither proportional to the mass matrix nor diagonal:

$$\mathcal{L}_{\text{Yuk}}^q = y_u h \overline{u_L} u_R + h (\overline{d_L} \ \overline{s_L}) \left[ V_{dL} \begin{pmatrix} Y_{d1} & Y_{d2} \\ 0 & 0 \end{pmatrix} V_{dR}^\dagger \right] \begin{pmatrix} d_R \\ s_R \end{pmatrix} + \text{h.c.} \quad . \quad (83)$$

Thus, in this model, both the  $Z$ -boson and the  $h$ -boson mediate FCNC at tree level. For example,  $K_L \rightarrow \mu^+ \mu^-$  and  $K^0 - \overline{K}^0$  mixing get  $Z$ - and  $h$ -mediated tree-level contributions.

### 4.3 2HDM: FCNC at tree level

Consider a model with two Higgs doublets. The symmetry structure, the pattern of spontaneous symmetry breaking, and the fermion content are the same as in the SM. However, the scalar content is extended:

– The scalar representations are

$$\phi_i(1, 2)_{+1/2}, \quad i = 1, 2 \quad . \quad (84)$$

We are particularly interested in the modification of the Yukawa terms:

$$\mathcal{L}_{\text{Yuk}} = (Y_k^u)_{ij} \overline{Q_{Li}} U_{Rj} \phi_k + (Y_k^d)_{ij} \overline{Q_{Li}} D_{Rj} \phi_k + (Y_k^e)_{ij} \overline{L_{Li}} E_{Rj} \phi_k + \text{h.c.} \quad . \quad (85)$$

Without loss of generality, we can work in a basis (commonly called "the Higgs basis")  $(\phi_A, \phi_M)$ , where one the Higgs doublets carries the VEV,  $\langle \phi_M \rangle = v$ , while the other has zero VEV,  $\langle \phi_A \rangle = 0$ . In this basis,  $Y_M^f$  is known and related to the fermions masses in the same way as the Yukawa matrices of the SM:

$$Y_M^f = \sqrt{2} M_f / v \quad . \quad (86)$$

The entries Yukawa matrices  $Y_A^f$  are, however, free parameters and, in general, unrelated to the fermion masses. The rotation angle from the Higgs basis to the basis of neutral CP-even Higgs states,  $(\phi_h, \phi_H)$ , is denoted by  $(\alpha - \beta)$ . The Yukawa matrix of the light Higgs field  $h$  is given by

$$Y_h^f = c_{\alpha-\beta} Y_A^f - s_{\alpha-\beta} Y_M^f \quad . \quad (87)$$

Given the arbitrary structure of  $Y_A^f$ , the Higgs boson can have couplings that are neither proportional to the mass matrix nor diagonal.

It is interesting to note, however, that not all multi Higgs doublet models lead to flavor changing Higgs couplings. If all the fermions of a given sector couple to one and the same doublet, then the Higgs couplings in that sector would still be diagonal. For example, in a model with two Higgs doublets,  $\phi_1$  and  $\phi_2$ , and Yukawa terms of the form

$$\mathcal{L}_{\text{Yuk}} = Y_{ij}^u \overline{Q_{Li}} U_{Rj} \phi_2 + Y_{ij}^d \overline{Q_{Li}} D_{Rj} \phi_1 + Y_{ij}^e \overline{L_{Li}} E_{Rj} \phi_1 + \text{h.c.} \quad , \quad (88)$$

the Higgs couplings are flavor diagonal:

$$Y_h^u = (c_\alpha / s_\beta) Y_M^u, \quad Y_h^d = -(s_\alpha / c_\beta) Y_M^d, \quad Y_h^e = -(s_\alpha / c_\beta) Y_M^e \quad , \quad (89)$$

where  $\beta$  [ $\alpha$ ] is the rotation angle from the  $(\phi_1, \phi_2)$  basis to the  $(\phi_A, \phi_M)$  [ $(\phi_h, \phi_H)$ ] basis. In the physics jargon, we say that such models have *natural flavor conservation* (NFC).

## 5 CP violation

There are two main reasons for the interest in CP violation:

- CP asymmetries provide some of the theoretically cleanest probes of flavor physics. The reason for that is that CP is a good symmetry of the strong interactions. Consequently, for some hadronic decays, QCD-related uncertainties cancel out in the CP asymmetries.
- There is a cosmological puzzle related to CP violation. The baryon asymmetry of the Universe is a CP violating observable, and it is many orders of magnitude larger than the SM prediction. Hence, there must exist new sources of CP violation beyond the single phase of the CKM matrix.

In this section we explain why CP violation is related to complex parameters of the Lagrangian. Based on this fact, we prove that CP violation in a two generation SM is impossible, while CP violation in a three generation SM requires a long list of conditions on its flavor parameters in order to occur.

## 5.1 CP violation and complex couplings

The CP transformation combines charge conjugation C with parity P. Under C, particles and antiparticles are interchanged by conjugating all internal quantum numbers, *e.g.*,  $Q \rightarrow -Q$ . Under P, the handedness of space is reversed,  $\vec{x} \rightarrow -\vec{x}$ . Thus, for example, a left-handed electron  $e_L^-$  is transformed under CP into a right-handed positron  $e_R^+$ .

At the Lagrangian level, CP is a good symmetry if there is a basis where all couplings are real. Let us provide a simple explanation of this statement. Consider fields  $\Phi_i$ . We can define the CP transformation of the fields as

$$\Phi_i \rightarrow \Phi_i^\dagger \quad . \quad (90)$$

Take, for example, terms in the Lagrangian that consist of three fields. (These could be Yukawa terms, if two of the  $\Phi_i$ 's are fermions and one is a scalar, or terms in the scalar potential, if all three are scalars, *etc.*) The hermiticity of the Lagrangian dictates that the following two terms should be included:

$$Y_{ijk} \Phi_i \Phi_j \Phi_k + Y_{ijk}^* \Phi_i^\dagger \Phi_j^\dagger \Phi_k^\dagger \quad . \quad (91)$$

Under the CP transformation, the field content of the two terms is exchanged, but the couplings remain the same. Thus, CP is a good symmetry if  $Y_{ijk} = Y_{ijk}^*$ , *i.e.*, the coupling is real.

In practice, things are more subtle, since one can define the CP transformation as  $\Phi_i \rightarrow e^{i\theta_i} \Phi_i^\dagger$ , with  $\theta_i$  a convention dependent phase. Then, there can be complex couplings, yet CP would be a good symmetry. Therefore, the correct statement is that CP is violated if, using all freedom to redefine the phases of the fields, one cannot find any basis where all couplings are real.

Let us examine the situation in the mass basis of the SM. The couplings of the gluons, the photon and the  $Z$ -boson are all real, as are the two parameters of the scalar potential. As concerns the fermion mass terms (or, equivalently, the Yukawa couplings) and the weak gauge interactions, the relevant CP transformation laws are

$$\bar{\psi}_i \psi_j \rightarrow \bar{\psi}_j \psi_i \quad , \quad \bar{\psi}_i \gamma^\mu W_\mu^+ (1 - \gamma_5) \psi_j \rightarrow \bar{\psi}_j \gamma^\mu W_\mu^- (1 - \gamma_5) \psi_i \quad . \quad (92)$$

Thus the mass terms and CC weak interaction terms are CP invariant if all the masses and couplings are real. We can always choose the masses to be real. Then, let us focus on the couplings of  $W^\pm$  to quarks:

$$- \frac{g}{\sqrt{2}} (V_{ij} \bar{u}_i \gamma^\mu W_\mu^+ (1 - \gamma_5) d_j + V_{ij}^* \bar{d}_j \gamma^\mu W_\mu^- (1 - \gamma_5) u_i) \quad . \quad (93)$$

The CP operation exchanges the two terms, except that  $V_{ij}$  and  $V_{ij}^*$  are not interchanged. Thus CP would be a good symmetry of the SM only if there were a mass basis and choice of phase convention where all masses and entries of the CKM matrix are real.

## 5.2 SM2: CP conserving

Consider a two generation Standard Model, SM2. This model is similar to the one defined in Section 2, which in this section will be referred to as SM3, except that there are two, rather than three fermion generations. Many features of SM2 are similar to SM3, but there is one important difference: CP is a good symmetry of SM2, but not of SM3. To see how this difference comes about, let us examine the accidental symmetries of SM2. We follow here the line of analysis of SM3 in Section 2.5.

If we set the Yukawa couplings to zero,  $\mathcal{L}_{\text{Yuk}}^{\text{SM2}} = 0$ , SM2 gains an accidental global symmetry:

$$G_{\text{SM2}}^{\text{global}} (Y^{u,d,e} = 0) = U(2)_Q \times U(2)_U \times U(2)_D \times U(2)_L \times U(2)_E \quad , \quad (94)$$

where the two generations of each gauge representation are a doublet of the corresponding  $U(2)$ . The Yukawa couplings break this symmetry into the subgroup

$$G_{\text{SM2}}^{\text{global}} = U(1)_B \times U(1)_e \times U(1)_\mu \quad . \quad (95)$$

A-priori, the Yukawa terms depend on three  $2 \times 2$  complex matrices, namely  $12_R + 12_I$  parameters. The global symmetry breaking,  $[U(2)]^5 \rightarrow [U(1)]^3$ , implies that we can remove  $5 \times (1_R + 3_I) - 3_I = 5_R + 12_I$  parameters. Thus the number of physical flavor parameters is 7 real parameters and no imaginary parameter. The real parameters can be identified as two charged lepton masses, four quark masses, and the single real mixing angle,  $\sin \theta_c = |V_{us}|$ .

The important conclusion for our purposes is that all imaginary couplings can be removed from SM2, and CP is an accidental symmetry of the model.

### 5.3 SM3: Not necessarily CP violating

A-priori, CP is not necessarily violated in SM3. If two quarks of the same charge had equal masses, one mixing angle and the phase could be removed from  $V$ . This can be written as a condition on the quark mass differences. CP violation requires

$$(m_t^2 - m_c^2)(m_t^2 - m_u^2)(m_c^2 - m_u^2)(m_b^2 - m_s^2)(m_b^2 - m_d^2)(m_s^2 - m_d^2) \neq 0 \quad . \quad (96)$$

Likewise, if the value of any of the three mixing angles were 0 or  $\pi/2$ , then the phase can be removed. Finally, CP would not be violated if the value of the single phase were 0 or  $\pi$ . These last eight conditions are elegantly incorporated into one, parametrization-independent condition. To find this condition, note that the unitarity of the CKM matrix,  $VV^\dagger = 1$ , requires that for any choice of  $i, j, k, l = 1, 2, 3$ ,

$$\mathcal{I}m[V_{ij}V_{kl}V_{il}^*V_{kj}^*] = J \sum_{m,n=1}^3 \epsilon_{ikm}\epsilon_{jln} \quad . \quad (97)$$

Then the conditions on the mixing parameters are summarized by

$$J \neq 0 \quad . \quad (98)$$

The quantity  $J$  is of much interest in the study of CP violation from the CKM matrix. The maximum value that  $J$  could assume in principle is  $1/(6\sqrt{3}) \approx 0.1$ , but it is found to be  $\sim 4 \times 10^{-5}$ .

The fourteen conditions incorporated in Eqs. (96) and (98) can all be written as a single requirement on the quark mass matrices in the interaction basis:

$$X_{CP} \equiv \mathcal{I}m \left\{ \det \left[ M_d M_d^\dagger, M_u M_u^\dagger \right] \right\} \neq 0 \Leftrightarrow \text{CP violation} \quad . \quad (99)$$

This is a convention independent condition.

### 5.4 $S_{\psi K_S}$

As an example of a CP violating observable, we take the CP asymmetry in  $B \rightarrow \psi K_S$  decays, which plays a major role in testing the KM mechanism. Before we explain the test itself, we should understand why the theoretical interpretation of the asymmetry is exceptionally clean, and what are the theoretical parameters on which it depends.

The CP asymmetry in neutral  $B$  meson decays into final CP eigenstates  $f_{CP}$  is defined as follows:

$$\mathcal{A}_{f_{CP}}(t) \equiv \frac{d\Gamma/dt[\overline{B}_{\text{phys}}^0(t) \rightarrow f_{CP}] - d\Gamma/dt[B_{\text{phys}}^0(t) \rightarrow f_{CP}]}{d\Gamma/dt[\overline{B}_{\text{phys}}^0(t) \rightarrow f_{CP}] + d\Gamma/dt[B_{\text{phys}}^0(t) \rightarrow f_{CP}]} \quad . \quad (100)$$

A detailed evaluation of this asymmetry is given in Appendix A. It leads to the following form:

$$\mathcal{A}_{f_{CP}}(t) = S_{f_{CP}} \sin(\Delta m_B t) - C_{f_{CP}} \cos(\Delta m_B t) \quad ,$$

$$S_{f_{CP}} \equiv \frac{2\mathcal{I}m(\lambda_{f_{CP}})}{1 + |\lambda_{f_{CP}}|^2} \quad , \quad C_{f_{CP}} \equiv \frac{1 - |\lambda_{f_{CP}}|^2}{1 + |\lambda_{f_{CP}}|^2} \quad , \quad (101)$$

where

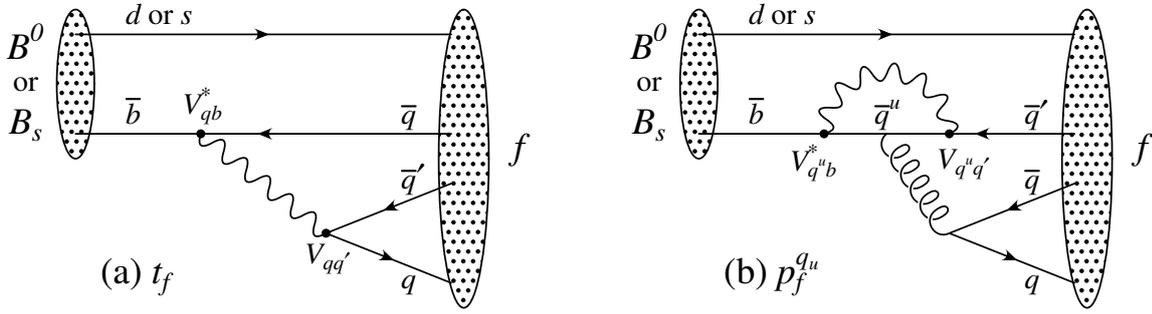
$$\lambda_{f_{CP}} = e^{-i\phi_B} (\bar{A}_{f_{CP}} / A_{f_{CP}}) \quad . \quad (102)$$

Here  $\phi_B$  refers to the phase of  $M_{B\bar{B}}$  [see Eq. (A.23)]. Within the Standard Model, the corresponding phase factor is given by

$$e^{-i\phi_B} = (V_{tb}^* V_{td}) / (V_{tb} V_{td}^*) \quad . \quad (103)$$

The decay amplitudes  $A_f$  and  $\bar{A}_f$  are defined in Eq. (A.1).

**Fig. 2:** Feynman diagrams for (a) tree and (b) penguin amplitudes contributing to  $B^0 \rightarrow f$  or  $B_s \rightarrow f$  via a  $\bar{b} \rightarrow \bar{q}q q'$  quark-level process.



The  $B^0 \rightarrow J/\psi K^0$  decay [4,5] proceeds via the quark transition  $\bar{b} \rightarrow \bar{c}c\bar{s}$ . There are contributions from both tree ( $t$ ) and penguin ( $p^{q_u}$ , where  $q_u = u, c, t$  is the quark in the loop) diagrams (see Fig. 2) which carry different weak phases:

$$A_f = (V_{cb}^* V_{cs}) t_f + \sum_{q_u=u,c,t} (V_{q_u b}^* V_{q_u s}) p_f^{q_u} \quad . \quad (104)$$

(The distinction between tree and penguin contributions is a heuristic one, the separation by the operator that enters is more precise. For a detailed discussion of the more complete operator product approach, which also includes higher order QCD corrections, see, for example, Ref. [6].) Using CKM unitarity, these decay amplitudes can always be written in terms of just two CKM combinations:

$$A_{\psi K} = (V_{cb}^* V_{cs}) T_{\psi K} + (V_{ub}^* V_{us}) P_{\psi K}^u \quad , \quad (105)$$

where  $T_{\psi K} = t_{\psi K} + p_{\psi K}^c - p_{\psi K}^t$  and  $P_{\psi K}^u = p_{\psi K}^u - p_{\psi K}^t$ . A subtlety arises in this decay that is related to the fact that  $B^0 \rightarrow J/\psi K^0$  and  $\bar{B}^0 \rightarrow J/\psi \bar{K}^0$ . A common final state, e.g.  $J/\psi K_S$ , can be reached via  $K^0 - \bar{K}^0$  mixing. Consequently, the phase factor corresponding to neutral  $K$  mixing,  $e^{-i\phi_K} = (V_{cd}^* V_{cs}) / (V_{cd} V_{cs}^*)$ , plays a role:

$$\frac{\bar{A}_{\psi K_S}}{A_{\psi K_S}} = - \frac{(V_{cb} V_{cs}^*) T_{\psi K} + (V_{ub} V_{us}^*) P_{\psi K}^u}{(V_{cb}^* V_{cs}) T_{\psi K} + (V_{ub}^* V_{us}) P_{\psi K}^u} \times \frac{V_{cd}^* V_{cs}}{V_{cd} V_{cs}^*} \quad . \quad (106)$$

The crucial point is that, for  $B \rightarrow J/\psi K_S$  and other  $\bar{b} \rightarrow \bar{c}c\bar{s}$  processes, we can neglect the  $P^u$  contribution to  $A_{\psi K}$ , in the SM, to an approximation that is better than one percent:

$$|P_{\psi K}^u / T_{\psi K}| \times |V_{ub} / V_{cb}| \times |V_{us} / V_{cs}| \sim (\text{loop factor}) \times 0.1 \times 0.23 \lesssim 0.005 \quad . \quad (107)$$

Thus, to an accuracy better than one percent,

$$\lambda_{\psi K_S} = \left( \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*} \right) \left( \frac{V_{cb} V_{cd}^*}{V_{cb}^* V_{cd}} \right) = -e^{-2i\beta} \quad , \quad (108)$$

where  $\beta$  is defined in Eq. (66), and consequently

$$S_{\psi K_S} = \sin 2\beta, \quad C_{\psi K_S} = 0 \quad . \quad (109)$$

(Below the percent level, several effects modify this equation [7–10].)

**Exercise 3:** Show that, if the  $B \rightarrow \pi\pi$  decays were dominated by tree diagrams, then  $S_{\pi\pi} = \sin 2\alpha$ .

**Exercise 4:** Estimate the accuracy of the predictions  $S_{\phi K_S} = \sin 2\beta$  and  $C_{\phi K_S} = 0$ .

## 6 Testing CKM

Measurements of rates, mixing, and CP asymmetries in  $B$  decays in the two B factories, BaBar and Belle, and in the two Tevatron detectors, CDF and D0, signified a new era in our understanding of flavor physics and CP violation. The progress has been both qualitative and quantitative. Various basic questions concerning CP and flavor violation have received, for the first time, answers based on experimental information. These questions include, for example,

- Is the Kobayashi-Maskawa mechanism at work (namely, is  $\delta_{KM} \neq 0$ )?
- Does the KM phase dominate the observed CP violation?
- Does the CKM mechanism dominate FCNC?

As a first step, one may assume the SM and test the overall consistency of the various measurements. However, the richness of data from the B factories allow us to go a step further and answer these questions model independently, namely allowing new physics to contribute to the relevant processes. We here explain the way in which this analysis proceeds.

### 6.1 Is the CKM assumption self-consistent?

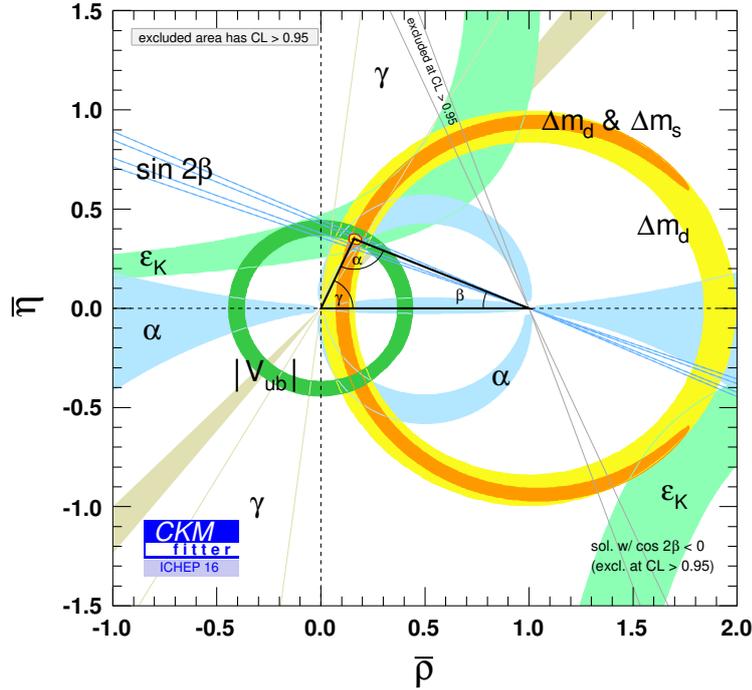
The three generation standard model has room for CP violation, through the KM phase in the quark mixing matrix. Yet, one would like to make sure that indeed CP is violated by the SM interactions, namely that  $\sin \delta_{KM} \neq 0$ . If we establish that this is the case, we would further like to know whether the SM contributions to CP violating observables are dominant. More quantitatively, we would like to put an upper bound on the ratio between the new physics and the SM contributions.

As a first step, one can assume that flavor changing processes are fully described by the SM, and check the consistency of the various measurements with this assumption. There are four relevant mixing parameters, which can be taken to be the Wolfenstein parameters  $\lambda$ ,  $A$ ,  $\rho$  and  $\eta$  defined in Eq. (60). The values of  $\lambda$  and  $A$  are known rather accurately [1] from, respectively,  $K \rightarrow \pi\ell\nu$  and  $b \rightarrow c\ell\nu$  decays:

$$\lambda = 0.2251 \pm 0.0005 \quad , \quad A = 0.81 \pm 0.03 \quad . \quad (110)$$

Then, one can express all the relevant observables as a function of the two remaining parameters,  $\rho$  and  $\eta$ , and check whether there is a range in the  $\rho - \eta$  plane that is consistent with all measurements. The list of observables includes the following:

- The rates of inclusive and exclusive charmless semileptonic  $B$  decays depend on  $|V_{ub}|^2 \propto \rho^2 + \eta^2$
- The CP asymmetry in  $B \rightarrow \psi K_S$ ,  $S_{\psi K_S} = \sin 2\beta = \frac{2\eta(1-\rho)}{(1-\rho)^2 + \eta^2}$
- The rates of various  $B \rightarrow DK$  decays depend on the phase  $\gamma$ , where  $e^{i\gamma} = \frac{\rho + i\eta}{\sqrt{\rho^2 + \eta^2}}$



**Fig. 3:** Allowed region in the  $\rho, \eta$  plane. Superimposed are the individual constraints from charmless semileptonic  $B$  decays ( $|V_{ub}|$ ), mass differences in the  $B^0$  ( $\Delta m_d$ ) and  $B_s$  ( $\Delta m_s$ ) neutral meson systems, and CP violation in  $K \rightarrow \pi\pi$  ( $\epsilon_K$ ),  $B \rightarrow \psi K$  ( $\sin 2\beta$ ),  $B \rightarrow \pi\pi, \rho\pi, \rho\rho$  ( $\alpha$ ), and  $B \rightarrow DK$  ( $\gamma$ ). Taken from [12].

- The rates of various  $B \rightarrow \pi\pi, \rho\pi, \rho\rho$  decays depend on the phase  $\alpha = \pi - \beta - \gamma$
- The ratio between the mass splittings in the neutral  $B$  and  $B_s$  systems is sensitive to  $|V_{td}/V_{ts}|^2 = \lambda^2[(1 - \rho)^2 + \eta^2]$
- The CP violation in  $K \rightarrow \pi\pi$  decays,  $\epsilon_K$ , depends in a complicated way on  $\rho$  and  $\eta$ .

The resulting constraints are shown in Fig. 3.

The consistency of the various constraints is impressive. In particular, the following ranges for  $\rho$  and  $\eta$  can account for all the measurements [1]:

$$\rho = +0.160 \pm 0.007 \quad , \quad \eta = +0.350 \pm 0.006 \quad . \quad (111)$$

One can make then the following statements [13]:

**Very likely, flavor changing processes are dominated by the Cabibbo-Kobayashi-Maskawa mechanism and, in particular, CP violation in flavor changing processes is dominated by the Kobayashi-Maskawa phase.**

In the following subsections, we explain how we can remove the phrase “very likely” from this statement, and how we can quantify the KM-dominance.

## 6.2 $S_{\psi K_S}$

As an example of how to use FCNC in probing new physics, we take  $S_{\psi K_S}$ . When we consider extensions of the SM, we still do not expect any significant new contribution to the tree level decay,  $b \rightarrow c\bar{c}s$ , beyond the SM  $W$ -mediated diagram. Thus, the expression  $\bar{A}_{\psi K_S}/A_{\psi K_S} = (V_{cb}V_{cd}^*)/(V_{cb}^*V_{cd})$  remains

valid, though the approximation of neglecting sub-dominant phases can be somewhat less accurate than Eq. (107). On the other hand, since  $B^0 - \bar{B}^0$  mixing is an FCNC process,  $M_{B\bar{B}}$  can in principle get large and even dominant contributions from new physics. We can parameterize the modification to the SM in terms of a complex parameter  $\Delta_d$ :

$$M_{B\bar{B}} = \Delta_d M_{B\bar{B}}^{\text{SM}}(\rho, \eta) \quad . \quad (112)$$

This leads to the following generalization of Eq. (109):

$$S_{\psi_{K_S}} = \sin[2\arctan(\eta/(1-\rho)) + \arg(\Delta_d)], \quad C_{\psi_{K_S}} = 0 \quad . \quad (113)$$

The experimental measurements give the following ranges [11]:

$$S_{\psi_{K_S}} = +0.70 \pm 0.02, \quad C_{\psi_{K_S}} = -0.005 \pm 0.015 \quad . \quad (114)$$

### 6.3 Is the KM mechanism at work?

In proving that the KM mechanism is at work, we assume that charged-current tree-level processes are dominated by the  $W$ -mediated SM diagrams (see, for example, [14]). This is a very plausible assumption. It is difficult to construct a model where new physics competes with the SM in flavor changing charged current processes, and does not violate the constraints from flavor changing neutral current processes. Thus we can use all tree level processes and fit them to  $\rho$  and  $\eta$ , as we did before. The list of such processes includes the following:

1. Charmless semileptonic  $B$ -decays,  $b \rightarrow u\ell\nu$ , measure  $R_u$  [see Eq. (65)].
2.  $B \rightarrow DK$  decays, which go through the quark transitions  $b \rightarrow c\bar{u}s$  and  $b \rightarrow u\bar{c}s$ , measure the angle  $\gamma$  [see Eq. (66)].
3.  $B \rightarrow \rho\rho$  decays (and, similarly,  $B \rightarrow \pi\pi$  and  $B \rightarrow \rho\pi$  decays) go through the quark transition  $b \rightarrow u\bar{u}d$ . With an isospin analysis, one can determine the relative phase between the tree decay amplitude and the mixing amplitude. By incorporating the measurement of  $S_{\psi_{K_S}}$ , one can subtract the phase from the mixing amplitude, finally providing a measurement of the angle  $\gamma$  [see Eq. (66)].

In addition, we can use loop processes, but then we must allow for new physics contributions, in addition to the  $(\rho, \eta)$ -dependent SM contributions. Of course, if each such measurement adds a separate mode-dependent parameter, then we do not gain anything by using this information. However, there is a number of observables where the only relevant loop process is  $B^0 - \bar{B}^0$  mixing. The list includes  $S_{\psi_{K_S}}$ ,  $\Delta m_B$  and the CP asymmetry in semileptonic  $B$  decays:

$$\begin{aligned} S_{\psi_{K_S}} &= \sin[2\arctan(\eta/(1-\rho)) + \arg(\Delta_d)] \quad , \\ \Delta m_B &= 2|M_{B\bar{B}}^{\text{SM}}(\rho, \eta)| \times |\Delta_d| \quad , \\ \mathcal{A}_{\text{SL}} &= -\mathcal{R}e\left(\frac{\Gamma_{B\bar{B}}}{M_{B\bar{B}}}\right)^{\text{SM}} \frac{\sin[\arg(\Delta_d)]}{|\Delta_d|} + \mathcal{I}m\left(\frac{\Gamma_{B\bar{B}}}{M_{B\bar{B}}}\right)^{\text{SM}} \frac{\cos[\arg(\Delta_d)]}{|\Delta_d|} \quad . \end{aligned} \quad (115)$$

As explained above, such processes involve two new parameters [see Eq. (112)]. Since there are three relevant observables, we can further tighten the constraints in the  $(\rho, \eta)$ -plane. Similarly, one can use measurements related to  $B_s - \bar{B}_s$  mixing. One gains three new observables at the cost of two new parameters (see, for example, [15]).

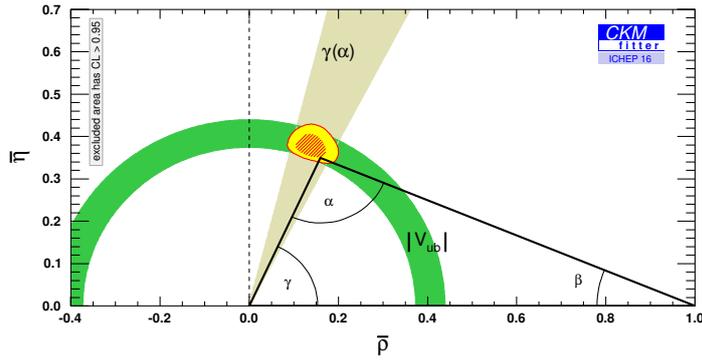
The results of such fit, projected on the  $\rho - \eta$  plane, can be seen in Fig. 4. It gives [12]

$$\eta = 0.38 \pm 0.02 \quad . \quad (116)$$

It is clear that  $\eta \neq 0$  is well established:

**The Kobayashi-Maskawa mechanism of CP violation is at work.**

The consistency of the experimental results (114) with the SM predictions (109) means that the KM mechanism of CP violation dominates the observed CP violation. In the next subsection, we make this statement more quantitative.

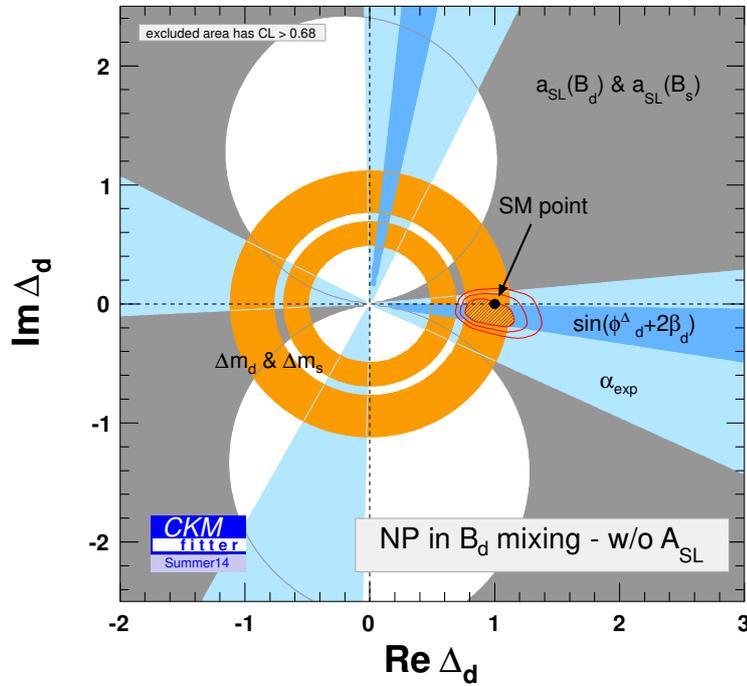


**Fig. 4:** The allowed region in the  $\rho - \eta$  plane, assuming that tree diagrams are dominated by the Standard Model [12].

#### 6.4 How much can new physics contribute to $B^0 - \bar{B}^0$ mixing?

All that we need to do in order to establish whether the SM dominates the observed CP violation, and to put an upper bound on the new physics contribution to  $B^0 - \bar{B}^0$  mixing, is to project the results of the fit performed in the previous subsection on the  $\text{Re}(\Delta_d) - \text{Im}(\Delta_d)$  plane. If we find that  $|\text{Im}(\Delta_d)| \ll 1$ , then the SM dominance in the observed CP violation will be established. The constraints are shown in Fig. 5.

**Fig. 5:** Constraints in the  $\text{Re}(\Delta_d) - \text{Im}(\Delta_d)$  plane, assuming that NP contributions to tree level processes are negligible [12].



We obtain:

$$\text{Re}(\Delta_d) = +0.94^{+0.18}_{-0.15} ,$$

$$\mathcal{I}m(\Delta_d) = -0.11_{-0.05}^{+0.11} . \quad (117)$$

This can be translated into the following approximate (one sigma) upper bounds:

$$\begin{aligned} |M_{B\bar{B}}^{\text{NP}}/M_{B\bar{B}}^{\text{SM}}| &\lesssim 0.2 , \\ \mathcal{I}m(M_{B\bar{B}}^{\text{NP}}/M_{B\bar{B}}^{\text{SM}}) &\lesssim 0.1 . \end{aligned} \quad (118)$$

We can make the following two statements:

1. **A new physics contribution to  $B^0 - \bar{B}^0$  mixing amplitude that carries a phase that is significantly different from the KM phase is constrained to lie below the 10% level.**
2. **A new physics contribution to the  $B^0 - \bar{B}^0$  mixing amplitude which is aligned with the KM phase is constrained to lie below the 20% level.**

Analogous upper bounds can be obtained for new physics contributions to the  $K^0 - \bar{K}^0$ ,  $B_s^0 - \bar{B}_s^0$ , and  $D^0 - \bar{D}^0$  mixing amplitudes.

## 7 The new physics flavor puzzle

### 7.1 A model independent discussion

It is clear that the Standard Model is not a complete theory of Nature:

1. It does not include gravity, and therefore it cannot be valid at energy scales above  $m_{\text{Planck}} \sim 10^{19}$  GeV;
2. It does not allow for neutrino masses, and therefore it cannot be valid at energy scales above  $m_{\text{seesaw}} \sim 10^{15}$  GeV;
3. The fine-tuning problem of the Higgs mass and the puzzle of the dark matter suggest that the scale where the SM is replaced with a more fundamental theory is actually much lower,  $m_{\text{top-partners}}$ ,  $m_{\text{wimp}} \lesssim$  a few TeV.

Given that the SM is only an effective low energy theory, non-renormalizable terms must be added to  $\mathcal{L}_{\text{SM}}$ . These are terms of dimension higher than four in the fields which, therefore, have couplings that are inversely proportional to the scale of new physics  $\Lambda_{\text{NP}}$ .

The lowest dimension non-renormalizable terms are dimension-five:

$$-\mathcal{L}_{\text{Seesaw}}^{\text{dim-5}} = \frac{Z_{ij}^\nu}{\Lambda_{\text{NP}}} L_{Li} L_{Lj} \phi \phi + \text{h.c.} . \quad (119)$$

These are the seesaw terms, leading to neutrino masses.

**Exercise 5:** *How does the global symmetry breaking pattern (47) change when (119) is taken into account?*

**Exercise 6:** *What is the number of physical lepton flavor parameters in this case? Identify these parameters in the mass basis.*

As concerns quark flavor physics, consider, for example, the following dimension-six set of operators:

$$\mathcal{L}_{\Delta F=2}^{\text{dim-6}} = \sum_{i \neq j} \frac{z_{ij}}{\Lambda^2} (\bar{Q}_{Li} \gamma_\mu Q_{Lj})^2 , \quad (120)$$

where the  $z_{ij}$  are dimensionless couplings. These terms contribute to the mass splittings between the corresponding two neutral mesons. For example, the term  $\mathcal{L}_{\Delta B=2} \propto (\bar{d}_L \gamma_\mu b_L)^2$  contributes to  $\Delta m_B$ , the mass difference between the two neutral  $B$ -mesons. We use

$$M_{B\bar{B}}^{\text{NP}} = \frac{1}{6} \frac{z_{db}}{\Lambda^2} m_B f_B^2 B_B . \quad (121)$$

**Table 6:** Lower bounds on the scale of new physics  $\Lambda$ , in units of TeV, for  $|z_{ij}| = 1$ , and upper bounds on  $z_{ij}$ , assuming  $\Lambda = 1$  TeV.

Operator	$\Lambda$ [TeV] CPC	$\Lambda$ [TeV] CPV	$ z_{ij} $	$\mathcal{I}m(z_{ij})$	Observables
$(\bar{s}_L \gamma^\mu d_L)^2$	$9.8 \times 10^2$	$1.6 \times 10^4$	$9.0 \times 10^{-7}$	$3.4 \times 10^{-9}$	$\Delta m_K; \epsilon_K$
$(\bar{s}_R d_L)(\bar{s}_L d_R)$	$1.8 \times 10^4$	$3.2 \times 10^5$	$6.9 \times 10^{-9}$	$2.6 \times 10^{-11}$	$\Delta m_K; \epsilon_K$
$(\bar{c}_L \gamma^\mu u_L)^2$	$1.2 \times 10^3$	$2.9 \times 10^3$	$5.6 \times 10^{-7}$	$1.0 \times 10^{-7}$	$\Delta m_D; A_\Gamma$
$(\bar{c}_R u_L)(\bar{c}_L u_R)$	$6.2 \times 10^3$	$1.5 \times 10^4$	$5.7 \times 10^{-8}$	$1.1 \times 10^{-8}$	$\Delta m_D; A_\Gamma$
$(\bar{b}_L \gamma^\mu d_L)^2$	$6.6 \times 10^2$	$9.3 \times 10^2$	$2.3 \times 10^{-6}$	$1.1 \times 10^{-6}$	$\Delta m_B; S_{\psi K}$
$(\bar{b}_R d_L)(\bar{b}_L d_R)$	$2.5 \times 10^3$	$3.6 \times 10^3$	$3.9 \times 10^{-7}$	$1.9 \times 10^{-7}$	$\Delta m_B; S_{\psi K}$
$(\bar{b}_L \gamma^\mu s_L)^2$	$1.4 \times 10^2$	$2.5 \times 10^2$	$5.0 \times 10^{-5}$	$1.7 \times 10^{-5}$	$\Delta m_{B_s}; S_{\psi\phi}$
$(\bar{b}_R s_L)(\bar{b}_L s_R)$	$4.8 \times 10^2$	$8.3 \times 10^2$	$8.8 \times 10^{-6}$	$2.9 \times 10^{-6}$	$\Delta m_{B_s}; S_{\psi\phi}$

Analogous expressions hold for the other neutral mesons. Taking into account the bounds of Eq. (118), we obtain

$$\frac{|z_{db}|}{\Lambda^2} < \frac{2.3 \times 10^{-6}}{\text{TeV}^2}, \quad \frac{\mathcal{I}m(z_{db})}{\Lambda^2} < \frac{1.1 \times 10^{-6}}{\text{TeV}^2}. \quad (122)$$

A more detailed list of the bounds derived from the  $\Delta F = 2$  observables in Table 5 is given in Table 6. The bounds refer to two representative sets of dimension-six operators: (i) left-left operators, that are also present in the SM, and (ii) operators with different chirality, where the bounds are strongest because of larger hadronic matrix elements.

The first lesson that we draw from these bounds on  $\Lambda$  is that new physics can contribute to FCNC at a level comparable to the SM contributions even if it takes place at a scale that is six orders of magnitude above the electroweak scale. A second lesson is that if the new physics has a generic flavor structure, that is  $z_{ij} = \mathcal{O}(1)$ , then its scale must be above  $10^4 - 10^5$  TeV (or, if the leading contributions involve electroweak loops, above  $10^3 - 10^4$  TeV). *If indeed  $\Lambda \gg TeV$ , it means that we have misinterpreted the hints from the fine-tuning problem and the dark matter puzzle.*

A different lesson can be drawn from the bounds on  $z_{ij}$ . *It could be that the scale of new physics is of order TeV, but its flavor structure is far from generic.* Specifically, if new particles at the TeV scale couple to the SM fermions, then there are two ways in which their contributions to FCNC processes, such as neutral meson mixing, can be suppressed: degeneracy and alignment. Either of these principles, or a combination of both, signifies non-generic structure.

One can use the language of effective operators also for the SM, integrating out all particles significantly heavier than the neutral mesons (that is, the top, the Higgs and the weak gauge bosons). Thus, the scale is  $\Lambda_{\text{SM}} \sim m_W$ . Since the leading contributions to neutral meson mixings come from box diagrams, the  $z_{ij}$  coefficients are suppressed by  $\alpha_2^2$ . To identify the relevant flavor suppression factor, one can employ the spurion formalism. For example, the flavor transition that is relevant to  $B^0 - \bar{B}^0$  mixing involves  $\bar{d}_L b_L$  which transforms as  $(8, 1, 1)_{SU(3)_q}$ . The leading contribution must then be proportional to  $(Y^u Y^{u\dagger})_{13} \propto y_t^2 V_{tb} V_{td}^*$ . Indeed, an explicit calculation (using VIA for the matrix element and neglecting QCD corrections) gives<sup>5</sup>

$$\frac{2M_{B\bar{B}}}{m_B} \approx -\frac{\alpha_2^2}{12} \frac{f_B^2}{m_W^2} S_0(x_t) (V_{tb} V_{td}^*)^2, \quad (123)$$

where  $x_i = m_i^2/m_W^2$  and

$$S_0(x) = \frac{x}{(1-x)^2} \left[ 1 - \frac{11x}{4} + \frac{x^2}{4} - \frac{3x^2 \ln x}{2(1-x)} \right]. \quad (124)$$

<sup>5</sup>A detailed derivation can be found in Appendix B of [16].

Similar spurion analyses, or explicit calculations, allow us to extract the weak and flavor suppression factors that apply in the SM:

$$\begin{aligned}
 \mathcal{I}m(z_{sd}^{\text{SM}}) &\sim \alpha_2^2 y_t^2 |V_{td}V_{ts}|^2 \sim 1 \times 10^{-10} \quad , \\
 z_{sd}^{\text{SM}} &\sim \alpha_2^2 y_c^2 |V_{cd}V_{cs}|^2 \sim 5 \times 10^{-9} \quad , \\
 \mathcal{I}m(z_{cu}^{\text{SM}}) &\sim \alpha_2^2 y_b^2 |V_{ub}V_{cb}|^2 \sim 2 \times 10^{-14} \quad , \\
 z_{bd}^{\text{SM}} &\sim \alpha_2^2 y_t^2 |V_{td}V_{tb}|^2 \sim 7 \times 10^{-8} \quad , \\
 z_{bs}^{\text{SM}} &\sim \alpha_2^2 y_t^2 |V_{ts}V_{tb}|^2 \sim 2 \times 10^{-6} \quad .
 \end{aligned} \tag{125}$$

(We did not include  $z_{cu}^{\text{SM}}$  in the list because it requires a more detailed consideration. The naively leading short distance contribution is  $\propto \alpha_2^2 (y_s^4/y_c^2) |V_{cs}V_{us}|^2 \sim 5 \times 10^{-13}$ . However, higher dimension terms can replace a  $y_s^2$  factor with  $(\Lambda/m_D)^2$  [17]. Moreover, long distance contributions are expected to dominate. In particular, peculiar phase space effects [18, 19] have been identified which are expected to enhance  $\Delta m_D$  to within an order of magnitude of its measured value. The CP violating part, on the other hand, is dominated by short distance physics.)

It is clear then that contributions from new physics at  $\Lambda_{\text{NP}} \sim 1$  TeV should be suppressed by factors that are comparable or smaller than the SM ones. Why does that happen? This is the new physics flavor puzzle.

The fact that the flavor structure of new physics at the TeV scale must be non-generic means that flavor measurements are a good probe of the new physics. Perhaps the best-studied example is that of supersymmetry. Here, the spectrum of the superpartners and the structure of their couplings to the SM fermions will allow us to probe the mechanism of dynamical supersymmetry breaking.

## 7.2 Minimal flavor violation (MFV)

Models of gauge mediated supersymmetry breaking (GMSB) provide a concrete example of a large class of models that obey a simple principle called *minimal flavor violation* (MFV) [28]. This principle guarantees that low energy flavor changing processes deviate only very little from the SM predictions. The basic idea can be described as follows. The gauge interactions of the SM are universal in flavor space. The only breaking of this flavor universality comes from the three Yukawa matrices,  $Y^u$ ,  $Y^d$  and  $Y^e$ . If this remains true in the presence of the new physics, namely  $Y^u$ ,  $Y^d$  and  $Y^e$  are the only flavor non-universal parameters, then the model belongs to the MFV class.

Let us now formulate this principle in a more formal way, using the language of spurions that we presented in Section 2.4. The Standard Model with vanishing Yukawa couplings has a large global symmetry (48,49). In this section we concentrate only on the quarks. The non-Abelian part of the flavor symmetry for the quarks is  $SU(3)_q^3$  of Eq. (49) with the three generations of quark fields transforming as follows:

$$Q_L(3, 1, 1) \quad , \quad U_R(1, 3, 1) \quad , \quad D_R(1, 1, 3) \quad . \tag{126}$$

The Yukawa interactions,

$$\mathcal{L}_{\text{Yuk}}^q = \overline{Q}_L Y^d D_R H + \overline{Q}_L Y^u U_R H_c \quad , \tag{127}$$

( $H_c = i\tau_2 H^*$ ) break this symmetry. The Yukawa couplings can thus be thought of as spurions with the following transformation properties under  $SU(3)_q^3$  [see Eq. (51)]:

$$Y^u \sim (3, \bar{3}, 1) \quad , \quad Y^d \sim (3, 1, \bar{3}) \quad . \tag{128}$$

When we say ‘‘spurions’’, we mean that we pretend that the Yukawa matrices are fields which transform under the flavor symmetry, and then require that all the Lagrangian terms, constructed from the SM fields,  $Y^d$  and  $Y^u$ , must be (formally) invariant under the flavor group  $SU(3)_q^3$ . Of course, in reality,  $\mathcal{L}_{\text{Yuk}}^q$  breaks  $SU(3)_q^3$  precisely because  $Y^{d,u}$  are *not* fields and do not transform under the symmetry.

**Table 7:** The MFV values and the experimental bounds on the coefficients of  $\Delta F = 1$  operators

Operator	$z_{ij} \propto$	CKM+GIM	$ z_{ij}  < (\Lambda/\text{TeV})^2 \times$
$(\bar{s}_L \gamma^\mu d_L)^2$	$y_t^4 (V_{ts} V_{td}^*)^2$	$10^{-7}$	$9.0 \times 10^{-7}$
$(\bar{s}_R d_L)(\bar{s}_L d_R)$	$y_t^4 y_s y_d (V_{ts} V_{td}^*)^2$	$10^{-14}$	$6.9 \times 10^{-9}$
$(\bar{c}_L \gamma^\mu u_L)^2$	$y_b^4 (V_{cb} V_{ub}^*)^2$	$10^{-14}$	$5.6 \times 10^{-7}$
$(\bar{c}_R u_L)(\bar{c}_L u_R)$	$y_b^4 y_c y_u (V_{cb} V_{ub}^*)^2$	$10^{-20}$	$5.7 \times 10^{-8}$
$(\bar{b}_L \gamma^\mu d_L)^2$	$y_t^4 (V_{tb} V_{td}^*)^2$	$10^{-4}$	$2.3 \times 10^{-6}$
$(\bar{b}_R d_L)(\bar{b}_L d_R)$	$y_t^4 y_b y_d (V_{tb} V_{td}^*)^2$	$10^{-9}$	$3.9 \times 10^{-7}$
$(\bar{b}_L \gamma^\mu s_L)^2$	$y_t^4 (V_{tb} V_{ts}^*)^2$	$10^{-3}$	$5.0 \times 10^{-5}$
$(\bar{b}_R s_L)(\bar{b}_L s_R)$	$y_t^4 y_b y_s (V_{tb} V_{ts}^*)^2$	$10^{-6}$	$8.8 \times 10^{-6}$

The idea of minimal flavor violation is relevant to extensions of the SM, and can be applied in two ways:

1. If we consider the SM as a low energy effective theory, then all higher-dimension operators, constructed from SM-fields and  $Y$ -spurions, are formally invariant under  $G_{\text{global}}$ .
2. If we consider a full high-energy theory that extends the SM, then all operators, constructed from SM and the new fields, and from  $Y$ -spurions, are formally invariant under  $G_{\text{global}}$ .

That MFV allows new physics at the TeV scale is demonstrated in Table 7.

**Exercise 10:** Use the spurion formalism to argue that, in MFV models, the  $K_L \rightarrow \pi^0 \nu \bar{\nu}$  decay amplitude is proportional to  $y_t^2 V_{td} V_{ts}^*$ .

Examples of MFV models include models of supersymmetry with gauge-mediation or with anomaly-mediation of its breaking.

## 8 The Standard Model flavor puzzle

The SM has thirteen flavor parameters: six quark Yukawa couplings, four CKM parameters (three angles and a phase), and three charged lepton Yukawa couplings. (One can use fermions masses instead of the fermion Yukawa couplings,  $y_f = \sqrt{2}m_f/v$ .) The orders of magnitudes of these thirteen dimensionless parameters are as follows:

$$\begin{aligned}
y_t &\sim 1, & y_c &\sim 10^{-2}, & y_u &\sim 10^{-5}, \\
y_b &\sim 10^{-2}, & y_s &\sim 10^{-3}, & y_d &\sim 10^{-4}, \\
y_\tau &\sim 10^{-2}, & y_\mu &\sim 10^{-3}, & y_e &\sim 10^{-6}, \\
|V_{us}| &\sim 0.2, & |V_{cb}| &\sim 0.04, & |V_{ub}| &\sim 0.004, & \delta_{\text{KM}} &\sim 1 \quad .
\end{aligned} \tag{129}$$

Only two of these parameters are clearly of  $\mathcal{O}(1)$ , the top-Yukawa and the KM phase. The other flavor parameters exhibit smallness and hierarchy. Their values span six orders of magnitude. It may be that this set of numerical values are just accidental. More likely, the smallness and the hierarchy have a reason. The question of why there is smallness and hierarchy in the SM flavor parameters constitutes “The Standard Model flavor puzzle.”

The motivation to think that there is indeed a structure in the flavor parameters is strengthened by considering the values of the four SM parameters that are not flavor parameters, namely the three gauge couplings and the Higgs self-coupling:

$$g_s \sim 1 \quad , \quad g \sim 0.6 \quad , \quad e \sim 0.3 \quad , \quad \lambda \sim 0.12 \quad . \tag{130}$$

This set of values does seem to be a random distribution of order-one numbers, as one would naively expect.

A few examples of mechanisms that were proposed to explain the observed structure of the flavor parameters are the following:

- An approximate Abelian symmetry (“The Froggatt-Nielsen mechanism” [29]);
- An approximate non-Abelian symmetry (see *e.g.* [30]);
- Conformal dynamics (“The Nelson-Strassler mechanism” [31]);
- Location in an extra dimension [32];
- Loop corrections (see *e.g.* [33]).

We take as an example the Froggatt-Nielsen mechanism.

### 8.1 The Froggatt-Nielsen (FN) mechanism

Small numbers and hierarchies are often explained by approximate symmetries. For example, the small mass splitting between the charged and neutral pions finds an explanation in the approximate isospin (global  $SU(2)$ ) symmetry of the strong interactions.

Approximate symmetries lead to selection rules which account for the size of deviations from the symmetry limit. Spurion analysis is particularly convenient to derive such selection rules. The Froggatt-Nielsen mechanism postulates a  $U(1)_H$  symmetry, that is broken by a small spurion  $\epsilon_H$ . Without loss of generality, we assign  $\epsilon_H$  a  $U(1)_H$  charge of  $H(\epsilon_H) = -1$ . Each SM field is assigned a  $U(1)_H$  charge. In general, different fermion generations are assigned different charges, hence the term ‘horizontal symmetry.’ The rule is that each term in the Lagrangian, made of SM fields and the spurion, should be formally invariant under  $U(1)_H$ .

The approximate  $U(1)_H$  symmetry thus leads to the following selection rules:

$$\begin{aligned} Y_{ij}^u &= \epsilon_H^{|H(\bar{Q}_i)+H(U_j)+H(\phi_u)|} , \\ Y_{ij}^d &= \epsilon_H^{|H(\bar{Q}_i)+H(D_j)+H(\phi_d)|} , \\ Y_{ij}^e &= \epsilon_H^{|H(\bar{L}_i)+H(E_j)-H(\phi_d)|} . \end{aligned} \quad (131)$$

The consequent parametric suppression of the physical parameters is then

$$\begin{aligned} y_f &\propto \epsilon_H^{H(\bar{f}_L)+H(f_R)+H(\phi)} , \\ |V_{ij}| &\propto \epsilon_H^{H(Q_{Li})-H(Q_{Lj})} . \end{aligned} \quad (132)$$

As a concrete example, we take the following set of charges:

$$\begin{aligned} H(\bar{Q}_i) &= H(U_i) = H(E_i) = (2, 1, 0) , \\ H(\bar{L}_i) &= H(D_i) = (0, 0, 0) , \\ H(\phi_u) &= H(\phi_d) = 0 . \end{aligned} \quad (133)$$

It leads to the following parametric suppressions of the Yukawa couplings:

$$Y^u \sim \begin{pmatrix} \epsilon^4 & \epsilon^3 & \epsilon^2 \\ \epsilon^3 & \epsilon^2 & \epsilon \\ \epsilon^2 & \epsilon & 1 \end{pmatrix} , \quad Y^d \sim (Y^e)^T \sim \begin{pmatrix} \epsilon^2 & \epsilon^2 & \epsilon^2 \\ \epsilon & \epsilon & \epsilon \\ 1 & 1 & 1 \end{pmatrix} . \quad (134)$$

We emphasize that for each entry we give the parametric suppression (that is the power of  $\epsilon$ ), but each entry has an unknown (complex) coefficient of order one, and there are no relations between the order one coefficients of different entries.

The structure of the Yukawa matrices dictates the parametric suppression of the physical observables:

$$\begin{aligned}
y_t &\sim 1, & y_c &\sim \epsilon^2, & y_u &\sim \epsilon^4, \\
y_b &\sim 1, & y_s &\sim \epsilon, & y_d &\sim \epsilon^2, \\
y_\tau &\sim 1, & y_\mu &\sim \epsilon, & y_e &\sim \epsilon^2, \\
|V_{us}| &\sim \epsilon, & |V_{cb}| &\sim \epsilon, & |V_{ub}| &\sim \epsilon^2, & \delta_{\text{KM}} &\sim 1.
\end{aligned} \tag{135}$$

For  $\epsilon \sim 0.05$ , the parametric suppressions are roughly consistent with the observed hierarchy. In particular, this set of charges predicts that the down and charged lepton mass hierarchies are similar, while the up hierarchy is the square of the down hierarchy. These features are roughly realized in Nature.

**Exercise 13:** *Derive the parametric suppression and approximate numerical values of  $Y^u$ , its eigenvalues, and the three angles of  $V_L^u$ , for  $H(Q_i) = 4, 2, 0$ ,  $H(U_i) = 3, 2, 0$  and  $\epsilon_H = 0.2$*

Could we explain any set of observed values with such an approximate symmetry? If we could, then the FN mechanism cannot be really tested. The answer however is negative. Consider, for example, the quark sector. Naively, we have 11  $U(1)_H$  charges that we are free to choose. However, the  $U(1)_Y \times U(1)_B \times U(1)_{\text{PQ}}$  symmetry implies that there are only 8 independent choices that affect the structure of the Yukawa couplings. On the other hand, there are 9 physical parameters. Thus, there should be a single relation between the physical parameters that is independent of the choice of charges. Assuming that the sum of charges in the exponents of Eq. (131) is of the same sign for all 18 combinations, the relation is

$$|V_{ub}| \sim |V_{us}V_{cb}|, \tag{136}$$

which is fulfilled to within a factor of 2. There are also interesting inequalities (here  $i < j$ ):

$$|V_{ij}| \gtrsim m(U_i)/m(U_j), \quad m(D_i)/m(D_j). \tag{137}$$

All six inequalities are fulfilled. Finally, if we order the up and the down masses from light to heavy, then the CKM matrix is predicted to be  $\sim \mathbf{1}$ , namely the diagonal entries are not parametrically suppressed. This structure is also consistent with the observed CKM structure.

## 8.2 The flavor of neutrinos

Five neutrino flavor parameters have been measured in recent years (see *e.g.* [34]): two mass-squared differences,

$$\Delta m_{21}^2 = (7.5 \pm 0.2) \times 10^{-5} \text{ eV}^2, \quad |\Delta m_{32}^2| = (2.5 \pm 0.1) \times 10^{-3} \text{ eV}^2, \tag{138}$$

and the three mixing angles,

$$|U_{e2}| = 0.55 \pm 0.01, \quad |U_{\mu 3}| = 0.67 \pm 0.03, \quad |U_{e3}| = 0.148 \pm 0.003. \tag{139}$$

These parameters constitute a significant addition to the thirteen SM flavor parameters and provide, in principle, tests of various ideas to explain the SM flavor puzzle.

The numerical values of the parameters show various surprising features:

- $|U_{\mu 3}| > \text{any } |V_{ij}|$  ;
- $|U_{e2}| > \text{any } |V_{ij}|$  ;
- $|U_{e3}|$  is not particularly small ( $|U_{e3}| \not\ll |U_{e2}U_{\mu 3}|$ ) ;
- $m_2/m_3 \gtrsim 1/6 > \text{any } m_i/m_j$  for charged fermions.

These features can be summarized by the statement that, in contrast to the charged fermions, neither smallness nor hierarchy have been observed so far in the neutrino related parameters.

One way of interpretation of the neutrino data comes under the name of neutrino mass anarchy [35–37]. It postulates that the neutrino mass matrix has no special structure, namely all entries are of the same order of magnitude. Normalized to an effective neutrino mass scale,  $v^2/\Lambda_{\text{seesaw}}$ , the various entries are random numbers of order one. Note that anarchy means neither hierarchy nor degeneracy.

If true, the contrast between neutrino mass anarchy and quark and charged lepton mass hierarchy may be a deep hint for a difference between the flavor physics of Majorana and Dirac fermions. The source of both anarchy and hierarchy might, however, be explained by a much more mundane mechanism. In particular, neutrino mass anarchy could be a result of a FN mechanism, where the three left-handed lepton doublets carry the same FN charge. In that case, the FN mechanism predict parametric suppression of neither neutrino mass ratios nor leptonic mixing angles, which is quite consistent with (138) and (139). Indeed, the viable FN model presented in Section 8.1 belongs to this class.

Another possible interpretation of the neutrino data is to take  $m_2/m_3 \sim |U_{e3}| \sim 0.15$  to be small, and require that they are parametrically suppressed (while the other two mixing angles are order one). Such a situation is impossible to accommodate in a large class of FN models [38].

The same data, and in particular the proximity of  $(|U_{\mu 3}|, |U_{\tau 3}|)$  to  $(1/\sqrt{2}, 1/\sqrt{2})$ , and the proximity of  $|U_{e2}|$  to  $1/\sqrt{3} \simeq 0.58$ , led to a very different interpretation. This interpretation, termed ‘tribimaximal mixing’ (TBM), postulates that the leptonic mixing matrix is parametrically close to the following special form [39]:

$$|U|_{\text{TBM}} = \begin{pmatrix} \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{pmatrix} . \quad (140)$$

Such a form is suggestive of discrete non-Abelian symmetries, and indeed numerous models based on an  $A_4$  symmetry have been proposed [40, 41]. A significant feature of of TBM is that the third mixing angle should be close to  $|U_{e3}| = 0$ . Until 2012, there have been only upper bounds on  $|U_{e3}|$ , consistent with the models in the literature. In recent years, however, a value of  $|U_{e3}|$  close to the previous upper bound has been established [42], see Eq. (139). Such a large value (and the consequent significant deviation of  $|U_{\mu 3}|$  from maximal bimixing) puts in serious doubt the TBM idea. Indeed, it is difficult in this framework, if not impossible, to account for  $\Delta m_{12}^2/\Delta m_{23}^2 \sim |U_{e3}|^2$  without fine-tuning [43].

## 9 Higgs physics: the new flavor arena

The SM relates the Yukawa couplings to the corresponding mass matrices:

$$Y^f = \sqrt{2}M_f/v . \quad (141)$$

This simple equation implies three features:

1. *Proportionality*:  $y_i \equiv Y_{ii}^f \propto m_i$  ;
2. *Factor of proportionality*:  $y_i/m_i = \sqrt{2}/v$  ;
3. *Diagonality*:  $Y_{ij}^f = 0$  for  $i \neq j$  .

In extensions of the SM, each of these three features might be violated. Thus, testing these features might provide a window to new physics and to allow progress in understanding the flavor puzzles.

A Higgs-like boson  $h$  has been discovered by the ATLAS and CMS experiments at the LHC [44, 45]. The experiments normalize their results to the SM rates:

$$\mu_f \equiv \frac{\sigma(pp \rightarrow h)\text{BR}(h \rightarrow f)}{[\sigma(pp \rightarrow h)\text{BR}(h \rightarrow f)]^{\text{SM}}} . \quad (142)$$

The measurements give [46–50]:

$$\begin{aligned}
\mu_{\gamma\gamma} &= 1.14 \pm 0.14 \quad , \\
\mu_{ZZ^*} &= 1.17 \pm 0.23 \quad , \\
\mu_{WW^*} &= 0.99 \pm 0.15 \quad , \\
\mu_{b\bar{b}} &= 0.98 \pm 0.20 \quad , \\
\mu_{\tau\tau} &= 1.09 \pm 0.23 \quad , \\
\mu_{t\bar{t}h} &= 1.29 \pm 0.18 \quad .
\end{aligned} \tag{143}$$

(Here  $\mu_{t\bar{t}h} \equiv \sigma(pp \rightarrow t\bar{t}h)/\sigma(pp \rightarrow t\bar{t}h)^{\text{SM}}$ .) In addition, there are upper bounds on decays into the first two generation charged leptons [51–53]:

$$\begin{aligned}
\mu_{\mu\mu} &< 2.8 \quad , \\
\mu_{ee} &< 4 \times 10^5 \quad .
\end{aligned} \tag{144}$$

As concerns quark flavor changing Higgs couplings, these have been searched for in  $t \rightarrow qh$  decays ( $q = c, u$ ) [54, 55]:

$$\begin{aligned}
\text{BR}(t \rightarrow ch) &< 2.2 \times 10^{-3} \quad , \\
\text{BR}(t \rightarrow uh) &< 2.4 \times 10^{-3} \quad ,
\end{aligned} \tag{145}$$

and for lepton flavor violating (LFV) decays [56–58]:

$$\begin{aligned}
\text{BR}(h \rightarrow \tau\mu) &< 2.5 \times 10^{-3} \quad , \\
\text{BR}(h \rightarrow \tau e) &< 6.1 \times 10^{-3} \quad , \\
\text{BR}(h \rightarrow \mu e) &< 3.4 \times 10^{-4} \quad .
\end{aligned} \tag{146}$$

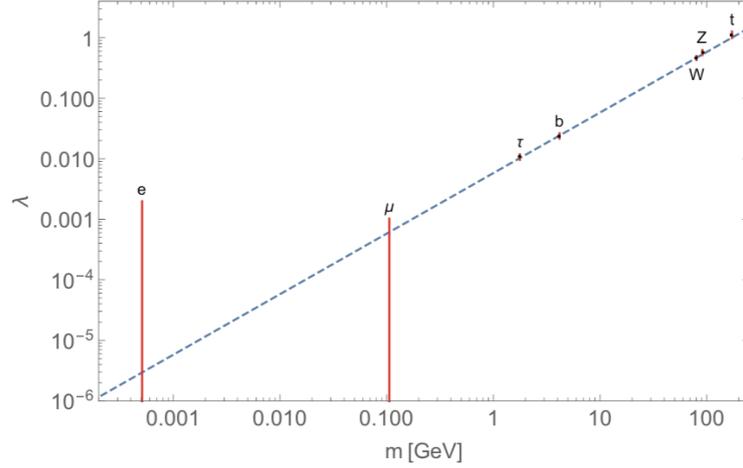
The measurements quoted in Eqs. (143) and (144) can be presented in the  $y_i - m_i$  plane. We do so in Fig. 6. The first two features quoted above are already being tested. The upper bounds on flavor violating decays quoted in Eqs. (145) and (146) test the third feature. We can make the following statements:

- $y_e, y_\mu < y_\tau$ . This goes in the direction of proportionality.
- The third generation Yukawa couplings,  $y_t, y_b, y_\tau$ , obey  $y_3/m_3 \approx \sqrt{2}/v$ . This is in agreement with the predicted factor of proportionality.
- There are strong upper bounds on violation of diagonality:  $Y_{tq}/Y_{yy} \lesssim 0.1$  and  $Y_{\tau\ell}/Y_{\tau\tau} \lesssim 0.1$ .
- The era of Higgs flavor physics has begun.

Beyond the search for new physics via Higgs decays, it is interesting to ask whether the measurements of the Higgs couplings to quarks and leptons can shed light on the standard model and/or new physics flavor puzzles. If eventually the values of  $y_b$  and/or  $y_\tau$  deviate from their SM values, the most likely explanation of such deviations will be that there are more than one Higgs doublets, and that the doublet(s) that couple to the down and charged lepton sectors are not the same as the one that couples to the up sector. A more significant test of our understanding of flavor physics, which might provide a window into new flavor physics, will come further in the future, when  $\mu_{\mu^+\mu^-}$  is measured. The ratio

$$X_{\mu^+\mu^-} \equiv \frac{\text{BR}(h \rightarrow \mu^+\mu^-)}{\text{BR}(h \rightarrow \tau^+\tau^-)} \quad , \tag{147}$$

is predicted within the SM with impressive theoretical cleanliness. To leading order, it is given by  $X_{\mu^+\mu^-} = m_\mu^2/m_\tau^2$ , and the corrections of order  $\alpha_W$  and of order  $m_\mu^2/m_\tau^2$  to this leading result are



**Fig. 6:** The allowed ranges for the Higgs couplings. The SM prediction is presented by the dashed line (Avital Dery, private communication).

known. It is an interesting question to understand what can be learned from a test of this relation [59]. In fact, as mentioned above, the bound (144) already shows that  $X_{\mu^+\mu^-} < 1$ , namely that second generation Yukawa couplings are smaller than third generation ones. It is also interesting to test diagonality via the search for the SM-forbidden decay modes,  $h \rightarrow \mu^\pm \tau^\mp$  [60–63]. A measurement of, or an upper bound on

$$X_{\mu\tau} \equiv \frac{\text{BR}(h \rightarrow \mu^+\tau^-) + \text{BR}(h \rightarrow \mu^-\tau^+)}{\text{BR}(h \rightarrow \tau^+\tau^-)} \quad , \quad (148)$$

would provide additional information relevant to flavor physics. We demonstrate below the potential power of Higgs flavor physics to lead to progress in our understanding of the flavor puzzles by focussing on the measurements of  $\mu_{\tau^+\tau^-}$ ,  $X_{\mu^+\mu^-}$  and  $X_{\mu\tau}$  [59].

Let us take as an example how we can use the set of these three measurements if there is a single light Higgs boson. A violation of the SM relation  $Y_{ij}^{\text{SM}} = \frac{\sqrt{2}m_i}{v} \delta_{ij}$ , is a consequence of nonrenormalizable terms. The leading ones are the  $d = 6$  terms. In the interaction basis, we have

$$\begin{aligned} \mathcal{L}_Y^{d=4} &= -\lambda_{ij} \bar{f}_L^i f_R^j \phi + \text{h.c.} \quad , \\ \mathcal{L}_Y^{d=6} &= -\frac{\lambda'_{ij}}{\Lambda^2} \bar{f}_L^i f_R^j \phi (\phi^\dagger \phi) + \text{h.c.} \quad , \end{aligned} \quad (149)$$

where expanding around the vacuum we have  $\phi = (v + h)/\sqrt{2}$ . Defining  $V_{L,R}$  via

$$\sqrt{2}m = V_L \left( \lambda + \frac{v^2}{2\Lambda^2} \lambda' \right) V_R^\dagger v \quad , \quad (150)$$

where  $m = \text{diag}(m_e, m_\mu, m_\tau)$ , and defining  $\hat{\lambda}$  via

$$\hat{\lambda} = V_L \lambda' V_R^\dagger \quad , \quad (151)$$

we obtain

$$Y_{ij} = \frac{\sqrt{2}m_i}{v} \delta_{ij} + \frac{v^2}{\Lambda^2} \hat{\lambda}_{ij} \quad . \quad (152)$$

To proceed, one has to make assumptions about the structure of  $\hat{\lambda}$ . In what follows, we consider first the assumption of minimal flavor violation (MFV) and then a Froggatt-Nielsen (FN) symmetry.

#### Exercise 14:

Find the predictions of models with Natural Flavor Conservation (NFC) for  $\mu_{\tau^+\tau^-}$ ,  $X_{\mu^+\mu^-}$  and  $X_{\tau\mu}$ .

## 9.1 MFV

MFV requires that the leptonic part of the Lagrangian is invariant under an  $SU(3)_L \times SU(3)_E$  global symmetry, with the left-handed lepton doublets transforming as  $(3, 1)$ , the right-handed charged lepton singlets transforming as  $(1, 3)$  and the charged lepton Yukawa matrix  $Y$  is a spurion transforming as  $(3, \bar{3})$ .

Specifically, MFV means that, in Eq. (149),

$$\lambda' = a\lambda + b\lambda\lambda^\dagger\lambda + \mathcal{O}(\lambda^5) \quad , \quad (153)$$

where  $a$  and  $b$  are numbers. Note that, if  $V_L$  and  $V_R$  are the diagonalizing matrices for  $\lambda$ ,  $V_L\lambda V_R^\dagger = \lambda^{\text{diag}}$ , then they are also the diagonalizing matrices for  $\lambda\lambda^\dagger\lambda$ ,  $V_L\lambda\lambda^\dagger\lambda V_R^\dagger = (\lambda^{\text{diag}})^3$ . Then, Eqs. (150), (151) and (152) become

$$\begin{aligned} \frac{\sqrt{2}m}{v} &= \left(1 + \frac{av^2}{2\Lambda^2}\right) \lambda^{\text{diag}} + \frac{bv^2}{2\Lambda^2} (\lambda^{\text{diag}})^3 \quad , \\ \hat{\lambda} &= a\lambda^{\text{diag}} + b(\lambda^{\text{diag}})^3 = a\frac{\sqrt{2}m}{v} + \frac{2\sqrt{2}bm^3}{v^3} \quad , \\ Y_{ij} &= \frac{\sqrt{2}m_i}{v} \delta_{ij} \left[1 + \frac{av^2}{\Lambda^2} + \frac{2bm_i^2}{\Lambda^2}\right] \quad , \end{aligned} \quad (154)$$

where, in the expressions for  $\hat{\lambda}$  and  $Y$ , we included only the leading universal and leading non-universal corrections to the SM relations.

We learn the following points about the Higgs-related lepton flavor parameters in this class of models:

1.  $h$  has no flavor off-diagonal couplings:

$$Y_{\mu\tau} \quad , \quad Y_{\tau\mu} = 0 \quad . \quad (155)$$

2. The values of the diagonal couplings deviate from their SM values. The deviation is small, of order  $v^2/\Lambda^2$ :

$$y_\tau \approx \left(1 + \frac{av^2}{\Lambda^2}\right) \frac{\sqrt{2}m_\tau}{v} \quad . \quad (156)$$

3. The ratio between the Yukawa couplings to different charged lepton flavors deviates from its SM value. The deviation is, however, very small, of order  $m_\tau^2/\Lambda^2$ :

$$\frac{y_\mu}{y_\tau} = \frac{m_\mu}{m_\tau} \left(1 - \frac{2b(m_\tau^2 - m_\mu^2)}{\Lambda^2}\right) \quad . \quad (157)$$

The predictions of the SM with MFV non-renormalizable terms are then the following:

$$\begin{aligned} \mu_{\tau^+\tau^-} &= 1 + 2av^2/\Lambda^2 \quad , \\ X_{\mu^+\mu^-} &= (m_\mu/m_\tau)^2(1 - 4bm_\tau^2/\Lambda^2) \quad , \\ X_{\tau\mu} &= 0 \quad . \end{aligned} \quad (158)$$

Thus, MFV will be excluded if experiments observe the  $h \rightarrow \mu\tau$  decay. On the other hand, MFV allows for a universal deviation of  $\mathcal{O}(v^2/\Lambda^2)$  of the flavor-diagonal dilepton rates, and a smaller non-universal deviation of  $\mathcal{O}(m_\tau^2/\Lambda^2)$ .

## 9.2 FN

An attractive explanation of the smallness and hierarchy in the Yukawa couplings is provided by the Froggatt-Nielsen (FN) mechanism [29]. In this framework, a  $U(1)_H$  symmetry, under which different generations carry different charges, is broken by a small parameter  $\epsilon_H$ . Without loss of generality,  $\epsilon_H$  is taken to be a spurion of charge  $-1$ . Then, various entries in the Yukawa mass matrices are suppressed by different powers of  $\epsilon_H$ , leading to smallness and hierarchy.

Specifically for the leptonic Yukawa matrix, taking  $h$  to be neutral under  $U(1)_H$ ,  $H(h) = 0$ , we have

$$\lambda_{ij} \propto \epsilon_H^{H(E_j)-H(L_i)} . \quad (159)$$

We emphasize that the FN mechanism dictates only the parametric suppression. Each entry has an arbitrary order one coefficient. The resulting parametric suppression of the masses and leptonic mixing angles is given by [64]

$$m_{\ell_i}/v \sim \epsilon_H^{H(E_i)-H(L_i)} , \quad |U_{ij}| \sim \epsilon_H^{H(L_j)-H(L_i)} . \quad (160)$$

Since  $H(\phi^\dagger\phi) = 0$ , the entries of the matrix  $\lambda'$  have the same parametric suppression as the corresponding entries in  $\lambda$  [65], though the order one coefficients are different:

$$\lambda'_{ij} = \mathcal{O}(1) \times \lambda_{ij} . \quad (161)$$

This structure allows us to estimate the entries of  $\hat{\lambda}_{ij}$  in terms of physical observables:

$$\begin{aligned} \hat{\lambda}_{33} &\sim m_\tau/v , \\ \hat{\lambda}_{22} &\sim m_\mu/v , \\ \hat{\lambda}_{23} &\sim |U_{23}|(m_\tau/v) , \\ \hat{\lambda}_{32} &\sim (m_\mu/v)/|U_{23}| . \end{aligned} \quad (162)$$

We learn the following points about the Higgs-related lepton flavor parameters in this class of models:

1.  $h$  has flavor off-diagonal couplings:

$$\begin{aligned} Y_{\mu\tau} &= \mathcal{O}\left(\frac{|U_{23}|vm_\tau}{\Lambda^2}\right) , \\ Y_{\tau\mu} &= \mathcal{O}\left(\frac{vm_\mu}{|U_{23}|\Lambda^2}\right) . \end{aligned} \quad (163)$$

2. The values of the diagonal couplings deviate from their SM values:

$$y_\tau \approx \frac{\sqrt{2}m_\tau}{v} \left[ 1 + \mathcal{O}\left(\frac{v^2}{\Lambda^2}\right) \right] . \quad (164)$$

3. The ratio between the Yukawa couplings to different charged lepton flavors deviates from its SM value:

$$\frac{y_\mu}{y_\tau} = \frac{m_\mu}{m_\tau} \left[ 1 + \mathcal{O}\left(\frac{v^2}{\Lambda^2}\right) \right] . \quad (165)$$

The predictions of the SM with FN-suppressed non-renormalizable terms are then the following:

$$\begin{aligned} \mu_{\tau^+\tau^-} &= 1 + \mathcal{O}(v^2/\Lambda^2) , \\ X_{\mu^+\mu^-} &= (m_\mu/m_\tau)^2(1 + \mathcal{O}(v^2/\Lambda^2)) , \end{aligned}$$

$$X_{\tau\mu} = \mathcal{O}(v^4/\Lambda^4) \quad . \quad (166)$$

Thus, FN will be excluded if experiments observe deviations from the SM of the same size in both flavor-diagonal and flavor-changing  $h$  decays. On the other hand, FN allows non-universal deviations of  $\mathcal{O}(v^2/\Lambda^2)$  in the flavor-diagonal dilepton rates, and a smaller deviation of  $\mathcal{O}(v^4/\Lambda^4)$  in the off-diagonal rate.

## 10 Anomalies in $B$ -meson decays?

In this section we discuss two sets of recent measurements of flavor changing processes that arouse much interest:  $B \rightarrow K^{(*)}\mu^+\mu^-$  and  $B \rightarrow D^{(*)}\tau\nu$ . Both classes of decays test lepton flavor universality (LFU).

### 10.1 $B \rightarrow K^{(*)}\mu\mu$

Within the SM, lepton flavor universality (LFU) is respected by the weak interactions. Consequently, LFU is predicted to hold – up to (calculable) phase-space effects – in processes where the Yukawa interactions are negligible. Hints of violation of LFU have, however, been observed by the LHCb experiment in  $B \rightarrow K^{(*)}\ell^+\ell^-$  decays. While LFU implies that the ratios

$$R_{K^{(*)},[a,b]} = \frac{\int_a^b dq^2 [d\Gamma(B \rightarrow K^{(*)}\mu^+\mu^-)/dq^2]}{\int_a^b dq^2 [d\Gamma(B \rightarrow K^{(*)}e^+e^-)/dq^2]} \quad (167)$$

( $q^2$  is the invariant dilepton mass-squared) should be very close to unity,

$$\begin{aligned} R_{K,[1,6]\text{GeV}^2}^{\text{SM}} &= 1.00 \pm 0.01 \quad , \\ R_{K^*,[1.1,6.0]\text{GeV}^2}^{\text{SM}} &= 1.00 \pm 0.01 \quad , \\ R_{K^*,[0.045,1.1]\text{GeV}^2}^{\text{SM}} &= 0.91 \pm 0.03 \quad , \end{aligned} \quad (168)$$

the measurements give [67, 68]

$$\begin{aligned} R_{K,[1,6]\text{GeV}^2} &= 0.745_{-0.074}^{+0.090} \pm 0.036 \quad , \\ R_{K^*,[1.1,6.0]\text{GeV}^2} &= 0.69_{-0.07}^{+0.11} \pm 0.05 \quad , \\ R_{K^*,[0.045,1.1]\text{GeV}^2} &= 0.66_{-0.07}^{+0.11} \pm 0.03 \quad , \end{aligned} \quad (169)$$

which stand in a  $2.2 - 2.6\sigma$  discrepancy with the SM predictions.

Given various additional measurements, it is plausible that, if indeed the discrepancy is due to new physics, it is related to modification of the muon mode, rather than to the electron mode. In this case, there should be destructive interference between the SM and the NP contributions. We assume that the new physics that affects the  $b \rightarrow s\mu\mu$  transition takes place at an energy scale larger than the electroweak breaking scale, in which case it can be represented by higher dimensional operators. There are two dimension-six operators that interfere with the SM contribution:

$$\mathcal{L}_{\text{eff}} = \frac{G_F\alpha}{\sqrt{2}\pi} V_{tb}V_{ts}^* [C_{LL}(\bar{s}\gamma^\mu P_L b)(\bar{\mu}\gamma^\mu P_L \mu) + C_{RL}(\bar{s}\gamma^\mu P_R b)(\bar{\mu}\gamma^\mu P_L \mu)] + \text{h.c.} \quad . \quad (170)$$

To leading order in  $C_{AB}^{\text{NP}}$ , we have

$$\begin{aligned} R_{K,[1,6]\text{GeV}^2} &= 1 + 2\mathcal{R}e\left(\frac{C_{LL}^{\text{NP}} + C_{RL}^{\text{NP}}}{C_{LL}^{\text{SM}}}\right) \quad , \\ R_{K^*,[0.045,1.1]\text{GeV}^2} &\approx 1 + 2\mathcal{R}e\left(\frac{C_{LL}^{\text{NP}} - C_{RL}^{\text{NP}}}{C_{LL}^{\text{SM}}}\right) \quad , \end{aligned} \quad (171)$$

where we approximated the polarization fraction  $p$  by  $p = 1$ . Given that both  $R_K$  and  $R_{K^*}$  are smaller than the SM value,  $C_{LL}^{\text{NP}}/C_{LL}^{\text{SM}} \sim -0.15$  is singled out as the prime candidate to explain the anomalies.

New physics that generates  $C_{LL}$  can do so at tree level or at the loop level. We focus on the first class. The mediator can be a scalar or a vector. We focus on scalar leptoquarks. There are four scalar leptoquark representations that couple to a down-type quark and a charged lepton:

- $(3, 1)_{-4/3}$  couples to  $\bar{D}\bar{E}$  and generates  $C_{RR}$ .
- $(3, 2)_{+1/6}$  couples to  $\bar{D}L$  and generates  $C_{RL}$ .
- $(3, 2)_{+7/6}$  couples to  $\bar{Q}E$  and generates  $C_{LR}$ .
- $(3, 3)_{-1/3}$  couples to  $\bar{Q}\bar{L}$  and generates  $C_{LL}$ .

Thus, only  $T(3, 3)_{-1/3}$  can account for both  $R_K$  and  $R_{K^*}$ . To do so, its mass and couplings have to obey

$$\frac{\mathcal{R}e(Y_{\mu s}^T Y_{\mu b}^{T*})}{m_T^2} \sim -\frac{0.004}{\text{TeV}^2} . \quad (172)$$

For  $\mathcal{R}e(Y_{\mu s}^T Y_{\mu b}^{T*}) \lesssim 1$ , we have  $m_T \lesssim 15$  TeV.

The same couplings as those modifying  $R_{K^{(*)}}$  modify also other processes. In particular, when the only operator generated by the NP is  $C_{LL}$ , we find the interesting predictions

$$\frac{\text{BR}(B_s \rightarrow \mu^+ \mu^-)}{\text{BR}(B_s \rightarrow \mu^+ \mu^-)_{\text{SM}}} = \frac{\text{BR}(B_s \rightarrow \phi \mu^+ \mu^-)}{\text{BR}(B_s \rightarrow \phi \mu^+ \mu^-)_{\text{SM}}} = R_K = R_{K^*} . \quad (173)$$

The  $T(3, 3)_{-1/3}$  contributes also to  $B_s - \bar{B}_s$  mixing via box diagrams. The resulting constraint is

$$\frac{|Y_{\mu s}^T Y_{\mu b}^{T*}|^2}{m_T^2} \lesssim -\frac{0.06}{\text{TeV}^2} . \quad (174)$$

Let us now consider subjecting this model to the MFV principle [69]. To have a  $T\bar{Q}\bar{L}$  coupling,  $T$  can transform under  $SU(3)_L \times SU(3)_E$  as either  $(\bar{3}, 1)$  or  $(1, \bar{3})$ . In the first case,

$$\frac{Y_{\tau s}^T Y_{\tau b}^{T*}}{Y_{\mu s}^T Y_{\mu b}^{T*}} = \frac{y_\tau^4}{y_\mu^4} , \quad (175)$$

and the experimental upper bound on  $B \rightarrow K\tau\tau$  is violated. In the latter case,

$$\frac{Y_{\tau s}^T Y_{\tau b}^{T*}}{Y_{\mu s}^T Y_{\mu b}^{T*}} = \frac{y_\tau^2}{y_\mu^2} . \quad (176)$$

The requirement from  $R_K$  and the constraint from  $B_s - \bar{B}_s$  mixing can be simultaneously satisfied only for  $m_T \lesssim 0.5$  TeV, which is excluded by LHC direct searches. We conclude that if the  $R_{K^{(*)}}$  anomaly is generated by a  $T(3, 3)_{-1/3}$  leptoquark, then MFV will be excluded [69].

## 10.2 $B \rightarrow D^{(*)}\tau\nu$

One can use the following ratios to test lepton flavor universality (LFU):

$$R(D^{(*)}) \equiv \frac{\Gamma(B \rightarrow D^{(*)}\tau\nu)}{\Gamma(B \rightarrow D^{(*)}\ell\nu)} , \quad (177)$$

where  $\ell = e, \mu$ . Babar [70, 71], Belle [72, 73] and LHCb [74] have measured  $R(D^{(*)})$ . The HFAG average of these measurements gives [75]

$$R(D^{(*)}) = 0.306 \pm 0.015 ,$$

$$R(D) = 0.407 \pm 0.046 \quad . \quad (178)$$

The SM values (averaged over several calculations) are given by

$$\begin{aligned} R(D^*)_{\text{SM}} &= 0.258 \pm 0.005 \quad , \\ R(D)_{\text{SM}} &= 0.299 \pm 0.003 \quad . \end{aligned} \quad (179)$$

Thus

$$\begin{aligned} R(D^*)/R(D^*)_{\text{SM}} &= 1.19 \pm 0.06 \quad , \\ R(D)/R(D)_{\text{SM}} &= 1.36 \pm 0.15 \quad . \end{aligned} \quad (180)$$

The  $p$ -value is  $1.57 \times 10^{-4}$ . In this section we entertain the idea that a deviation from the SM will indeed be established. We follow mainly the analyses of Refs. [76, 77].

We assume that the new physics that affects the  $b \rightarrow c\tau\nu$  transition takes place at an energy scale larger than the electroweak breaking scale, in which case it can be represented by higher dimensional operators. The most general  $d = 6$  terms are [76]

$$\begin{aligned} \mathcal{L}_{\text{eff}} &= c_{\bar{Q}QLL}^{ijkl} (\bar{Q}_i \gamma_\mu \sigma^a Q_j) (\bar{L}_k \gamma^\mu \sigma_a L_l) + c_{\bar{Q}uLe}^{ijkl} (\bar{Q}_i U_j) i\sigma^2 (\bar{L}_k E_l) \\ &+ c_{\bar{d}QLe}^{ijkl} (\bar{d}_i Q_j) (\bar{L}_k E_l) + c_{\bar{d}QLe'}^{ijkl} (\bar{d}_i \sigma_{\mu\nu} Q_j) (\bar{L}_k \sigma^{\mu\nu} E_l) + \text{h.c.} \quad . \end{aligned} \quad (181)$$

It has been shown that a non-zero  $c_{\bar{Q}uLe}$  alone cannot accommodate both  $R(D^{(*)})$  and be consistent with the measurements of the corresponding decay spectra.

In what follows, we write the Wilson coefficients in the down and charged lepton mass bases, *e.g.*  $Q_3 = (V_{ib}u_{iL}, b_L)^T$  and  $L_3 = (\nu_\tau, \tau_L)^T$ .

Given that the deviation from the SM is large, of order 30%, and that within the SM the semileptonic decay  $b \rightarrow c\tau\nu$  is a  $W$ -mediated tree-level decay, it is likely that the new physics contribution is also a tree-level one. The mediator can be a Lorentz scalar or vector, and a color singlet or triplet, for example:

- A scalar  $H'(1, 2)_{+1/2}$  ;
- A vector  $W'_\mu(1, 3)_0$  ;
- A vector  $U_\mu(3, 1)_{+2/3}$  ;
- A scalar  $\Delta(3, 2)_{+7/6}$  .

We discuss the first three.

The scalar  $H'$  with Yukawa couplings

$$\mathcal{L}_{H'} = -Y_b \bar{Q}_3 H' b_R - Y_c \bar{Q}_3 \tilde{H}' c_R - Y_\tau \bar{L}_3 H' \tau_R + \text{h.c.} \quad (182)$$

generates

$$c_{\bar{d}QLe}^{3333} = Y_b Y_\tau^* / M_{H'}^2, \quad c_{\bar{Q}uLe}^{3233} = Y_c Y_\tau / M_{H'}^2 \quad . \quad (183)$$

It can account for (180) with  $c_{\bar{d}QLe}^{3333} = (50 \pm 14) \text{TeV}^{-2}$  and  $c_{\bar{Q}uLe}^{3233} = (-1.6 \pm 0.5) \text{TeV}^{-2}$ .

The vector  $W'_\mu$  with couplings

$$\mathcal{L}_{W'} = W'_\mu \left( \lambda_{ij}^q \bar{Q}_i \gamma^\mu \sigma^a Q_j + \lambda_{ij}^\ell \bar{L}_i \gamma^\mu \sigma^a L_j \right) \quad (184)$$

generates

$$c_{\bar{Q}QLL}^{3333} = -\lambda_{33}^q \lambda_{33}^\ell / M_{W'}^2 \quad . \quad (185)$$

It can account for (180) (note a  $V_{cb}$  suppression factor) with  $c_{QQLL}^{3333} = (-2.1 \pm 0.5)\text{TeV}^{-2}$ . At the same time, this operator leads to  $b\bar{b} \rightarrow W'^0 \rightarrow \tau^+\tau^-$  at unacceptably large rate.

The vector leptoquark  $U_\mu$  with couplings

$$\mathcal{L}_U = g_U U_\mu \bar{Q}_3 \gamma^\mu L_3 + \text{h.c.} \quad (186)$$

generates

$$c_{QQLL}^{3333} = g_U^2 / (2M_U^2) \quad . \quad (187)$$

It can account for (180) (note a  $V_{cb}$  suppression factor) with  $c_{QQLL}^{3333} = (-2.1 \pm 0.5)\text{TeV}^{-2}$ .

The presence of a quark doublet  $Q_i$  and a lepton doublet  $L_j$  in all operators implies that, in parallel to the charged current operator of interest, there are unavoidably also neutral current four fermion operators of the type  $\bar{u}ul^-\ell^+$  or  $\bar{d}d\ell^-\ell^+$ . Let us take as an example the case of  $W'_\mu{}^a$  with only  $\lambda_{33}^q \neq 0$  among the quark couplings. Even in this case, a potentially large contribution to  $\Delta C = 2$  transition is generated [77]:

$$\mathcal{L}_{\Delta C=2}^{W'} = -\frac{(\lambda_{33}^q V_{ub} V_{cb}^*)^2}{2M_{W'}^2} (\bar{u}_L \gamma_\mu c_L)^2 + \text{h.c.} \quad . \quad (188)$$

The experimental bounds on CP violation in  $D^0 - \bar{D}^0$  mixing require [77]

$$\frac{\lambda_{33}^{q2}}{M_{W'}^2} \leq 10 \text{ TeV}^{-2} \quad , \quad (189)$$

to be compared with the  $R(D^{(*)})$  constraint [76, 77],

$$\frac{\lambda_{33}^q \lambda_{33}^\ell}{M_{W'}^2} \sim 2 \text{ TeV}^{-2} \quad . \quad (190)$$

Thus, the model is viable for  $\lambda_{33}^q / \lambda_{33}^\ell < 5$ .

## 11 Conclusions

- (i) Measurements of CP violating  $B$ -meson decays have established that the Kobayashi-Maskawa mechanism is the dominant source of the observed CP violation.
- (ii) Measurements of flavor changing  $B$ -meson decays have established the Cabibbo-Kobayashi-Maskawa mechanism is a major player in flavor violation.
- (iii) The consistency of all these measurements with the CKM predictions sharpens the new physics flavor puzzle: If there is new physics at, or below, the TeV scale, then its flavor structure must be highly non-generic.
- (iv) Measurements of neutrino flavor parameters have not only not clarified the standard model flavor puzzle, but actually deepened it. Whether they imply an anarchical structure, or a tribimaximal mixing, it seems that the neutrino flavor structure is very different from that of quarks.
- (v) If the LHC experiments discover new particles that couple to the Standard Model fermions, then, in principle, they will be able to measure new flavor parameters. Consequently, the new physics flavor puzzle is likely to be understood.
- (vi) If the flavor structure of such new particles is affected by the same physics that sets the flavor structure of the Yukawa couplings, then the LHC experiments (and future flavor factories) may be able to shed light also on the standard model flavor puzzle.
- (vii) The recently discovered Higgs-like boson provides an opportunity to make progress in our understanding of the flavor puzzle(s).

(viii) Extensions of the SM where new particles couple to quark- and/or lepton-pairs are constrained by flavor.

The huge progress in flavor physics in recent years has provided answers to many questions. At the same time, new questions arise. The LHC era is likely to provide more answers and more questions.

## Acknowledgments

I thank my students – Avital Dery, Daniel Aloni, and Shahaf Aharony – for many useful discussions. YN is supported by grants from the I-CORE program of the Planning and Budgeting Committee of the Israel Science Foundation (grant number 1937/12), from the Israel Science Foundation (grant number 394/16) and from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel (grant number 2014230).

## Appendices

### A CPV in $B$ decays to final CP eigenstates

We define decay amplitudes of  $B$  (which could be charged or neutral) and its CP conjugate  $\bar{B}$  to a multi-particle final state  $f$  and its CP conjugate  $\bar{f}$  as

$$A_f = \langle f | \mathcal{H} | B \rangle \quad , \quad \bar{A}_f = \langle f | \mathcal{H} | \bar{B} \rangle \quad , \quad A_{\bar{f}} = \langle \bar{f} | \mathcal{H} | B \rangle \quad , \quad \bar{A}_{\bar{f}} = \langle \bar{f} | \mathcal{H} | \bar{B} \rangle \quad , \quad (\text{A.1})$$

where  $\mathcal{H}$  is the Hamiltonian governing weak interactions. The action of CP on these states introduces phases  $\xi_B$  and  $\xi_f$  according to

$$\begin{aligned} CP |B\rangle &= e^{+i\xi_B} |\bar{B}\rangle \quad , \quad CP |f\rangle = e^{+i\xi_f} |\bar{f}\rangle \quad , \\ CP |\bar{B}\rangle &= e^{-i\xi_B} |B\rangle \quad , \quad CP |\bar{f}\rangle = e^{-i\xi_f} |f\rangle \quad , \end{aligned} \quad (\text{A.2})$$

so that  $(CP)^2 = 1$ . The phases  $\xi_B$  and  $\xi_f$  are arbitrary and unphysical because of the flavor symmetry of the strong interaction. If CP is conserved by the dynamics,  $[CP, \mathcal{H}] = 0$ , then  $A_f$  and  $\bar{A}_{\bar{f}}$  have the same magnitude and an arbitrary unphysical relative phase

$$\bar{A}_{\bar{f}} = e^{i(\xi_f - \xi_B)} A_f \quad . \quad (\text{A.3})$$

A state that is initially a superposition of  $B^0$  and  $\bar{B}^0$ , say

$$|\psi(0)\rangle = a(0)|B^0\rangle + b(0)|\bar{B}^0\rangle \quad , \quad (\text{A.4})$$

will evolve in time acquiring components that describe all possible decay final states  $\{f_1, f_2, \dots\}$ , that is,

$$|\psi(t)\rangle = a(t)|B^0\rangle + b(t)|\bar{B}^0\rangle + c_1(t)|f_1\rangle + c_2(t)|f_2\rangle + \dots \quad . \quad (\text{A.5})$$

If we are interested in computing only the values of  $a(t)$  and  $b(t)$  (and not the values of all  $c_i(t)$ ), and if the times  $t$  in which we are interested are much larger than the typical strong interaction scale, then we can use a much simplified formalism [83]. The simplified time evolution is determined by a  $2 \times 2$  effective Hamiltonian  $\mathcal{H}$  that is not Hermitian, since otherwise the mesons would only oscillate and not decay. Any complex matrix, such as  $\mathcal{H}$ , can be written in terms of Hermitian matrices  $M$  and  $\Gamma$  as

$$\mathcal{H} = M - \frac{i}{2} \Gamma \quad . \quad (\text{A.6})$$

$M$  and  $\Gamma$  are associated with  $(B^0, \bar{B}^0) \leftrightarrow (B^0, \bar{B}^0)$  transitions via off-shell (dispersive) and on-shell (absorptive) intermediate states, respectively. Diagonal elements of  $M$  and  $\Gamma$  are associated with the

flavor-conserving transitions  $B^0 \rightarrow B^0$  and  $\bar{B}^0 \rightarrow \bar{B}^0$  while off-diagonal elements are associated with flavor-changing transitions  $B^0 \leftrightarrow \bar{B}^0$ .

The eigenvectors of  $\mathcal{H}$  have well defined masses and decay widths. We introduce complex parameters  $p$  and  $q$  to specify the components of the strong interaction eigenstates,  $B^0$  and  $\bar{B}^0$ , in the light ( $B_L$ ) and heavy ( $B_H$ ) mass eigenstates:

$$|B_{L,H}\rangle = p|B^0\rangle \pm q|\bar{B}^0\rangle \quad (\text{A.7})$$

with the normalization  $|p|^2 + |q|^2 = 1$ . The special form of Eq. (A.7) is related to the fact that CPT imposes  $M_{11} = M_{22}$  and  $\Gamma_{11} = \Gamma_{22}$ . Solving the eigenvalue problem gives

$$\left(\frac{q}{p}\right)^2 = \frac{M_{12}^* - (i/2)\Gamma_{12}^*}{M_{12} - (i/2)\Gamma_{12}} \quad (\text{A.8})$$

If either CP or T is a symmetry of  $\mathcal{H}$ , then  $M_{12}$  and  $\Gamma_{12}$  are relatively real, leading to

$$\left(\frac{q}{p}\right)^2 = e^{2i\xi_B} \quad \Rightarrow \quad \left|\frac{q}{p}\right| = 1 \quad , \quad (\text{A.9})$$

where  $\xi_B$  is the arbitrary unphysical phase introduced in Eq. (A.2).

The real and imaginary parts of the eigenvalues of  $\mathcal{H}$  corresponding to  $|B_{L,H}\rangle$  represent their masses and decay-widths, respectively. The mass difference  $\Delta m_B$  and the width difference  $\Delta\Gamma_B$  are defined as follows:

$$\Delta m_B \equiv M_H - M_L \quad , \quad \Delta\Gamma_B \equiv \Gamma_H - \Gamma_L \quad (\text{A.10})$$

Note that here  $\Delta m_B$  is positive by definition, while the sign of  $\Delta\Gamma_B$  is to be experimentally determined. The average mass and width are given by

$$m_B \equiv \frac{M_H + M_L}{2} \quad , \quad \Gamma_B \equiv \frac{\Gamma_H + \Gamma_L}{2} \quad (\text{A.11})$$

It is useful to define dimensionless ratios  $x$  and  $y$ :

$$x \equiv \frac{\Delta m_B}{\Gamma_B} \quad , \quad y \equiv \frac{\Delta\Gamma_B}{2\Gamma_B} \quad (\text{A.12})$$

Solving the eigenvalue equation gives

$$(\Delta m_B)^2 - \frac{1}{4}(\Delta\Gamma_B)^2 = (4|M_{12}|^2 - |\Gamma_{12}|^2) \quad , \quad \Delta m_B \Delta\Gamma_B = 4\mathcal{R}e(M_{12}\Gamma_{12}^*) \quad (\text{A.13})$$

All CP-violating observables in  $B$  and  $\bar{B}$  decays to final states  $f$  and  $\bar{f}$  can be expressed in terms of phase-convention-independent combinations of  $A_f$ ,  $\bar{A}_f$ ,  $A_{\bar{f}}$  and  $\bar{A}_{\bar{f}}$ , together with, for neutral-meson decays only,  $q/p$ . CP violation in charged-meson decays depends only on the combination  $|\bar{A}_{\bar{f}}/A_f|$ , while CP violation in neutral-meson decays is complicated by  $B^0 \leftrightarrow \bar{B}^0$  oscillations and depends, additionally, on  $|q/p|$  and on  $\lambda_f \equiv (q/p)(\bar{A}_f/A_f)$ .

For neutral  $D$ ,  $B$ , and  $B_s$  mesons,  $\Delta\Gamma/\Gamma \ll 1$  and so both mass eigenstates must be considered in their evolution. We denote the state of an initially pure  $|B^0\rangle$  or  $|\bar{B}^0\rangle$  after an elapsed proper time  $t$  as  $|B_{\text{phys}}^0(t)\rangle$  or  $|\bar{B}_{\text{phys}}^0(t)\rangle$ , respectively. Using the effective Hamiltonian approximation, we obtain

$$\begin{aligned} |B_{\text{phys}}^0(t)\rangle &= g_+(t)|B^0\rangle - \frac{q}{p}g_-(t)|\bar{B}^0\rangle \quad , \\ |\bar{B}_{\text{phys}}^0(t)\rangle &= g_+(t)|\bar{B}^0\rangle - \frac{p}{q}g_-(t)|B^0\rangle \quad , \end{aligned} \quad (\text{A.14})$$

where

$$g_{\pm}(t) \equiv \frac{1}{2} \left( e^{-im_H t - \frac{1}{2}\Gamma_H t} \pm e^{-im_L t - \frac{1}{2}\Gamma_L t} \right) . \quad (\text{A.15})$$

One obtains the following time-dependent decay rates:

$$\begin{aligned} \frac{d\Gamma[B^0_{\text{phys}}(t) \rightarrow f]/dt}{e^{-\Gamma t} \mathcal{N}_f} &= (|A_f|^2 + |(q/p)\bar{A}_f|^2) \cosh(y\Gamma t) + (|A_f|^2 - |(q/p)\bar{A}_f|^2) \cos(x\Gamma t) \\ &+ 2 \operatorname{Re}((q/p)A_f^* \bar{A}_f) \sinh(y\Gamma t) - 2 \operatorname{Im}((q/p)A_f^* \bar{A}_f) \sin(x\Gamma t) \quad , (\text{A.16}) \end{aligned}$$

$$\begin{aligned} \frac{d\Gamma[\bar{B}^0_{\text{phys}}(t) \rightarrow f]/dt}{e^{-\Gamma t} \mathcal{N}_f} &= (|(p/q)A_f|^2 + |\bar{A}_f|^2) \cosh(y\Gamma t) - (|(p/q)A_f|^2 - |\bar{A}_f|^2) \cos(x\Gamma t) \\ &+ 2 \operatorname{Re}((p/q)A_f \bar{A}_f^*) \sinh(y\Gamma t) - 2 \operatorname{Im}((p/q)A_f \bar{A}_f^*) \sin(x\Gamma t) \quad , (\text{A.17}) \end{aligned}$$

where  $\mathcal{N}_f$  is a common normalization factor. Decay rates to the CP-conjugate final state  $\bar{f}$  are obtained analogously, with  $\mathcal{N}_f = \mathcal{N}_{\bar{f}}$  and the substitutions  $A_f \rightarrow A_{\bar{f}}$  and  $\bar{A}_f \rightarrow \bar{A}_{\bar{f}}$  in Eqs. (A.16, A.17). Terms proportional to  $|A_f|^2$  or  $|\bar{A}_f|^2$  are associated with decays that occur without any net  $B \leftrightarrow \bar{B}$  oscillation, while terms proportional to  $|(q/p)\bar{A}_f|^2$  or  $|(p/q)A_f|^2$  are associated with decays following a net oscillation. The  $\sinh(y\Gamma t)$  and  $\sin(x\Gamma t)$  terms of Eqs. (A.16, A.17) are associated with the interference between these two cases. Note that, in multi-body decays, amplitudes are functions of phase-space variables. Interference may be present in some regions but not others, and is strongly influenced by resonant substructure.

One possible manifestation of CP-violating effects in meson decays [84] is in the interference between a decay without mixing,  $B^0 \rightarrow f$ , and a decay with mixing,  $B^0 \rightarrow \bar{B}^0 \rightarrow f$  (such an effect occurs only in decays to final states that are common to  $B^0$  and  $\bar{B}^0$ , including all CP eigenstates). It is defined by

$$\operatorname{Im}(\lambda_f) \neq 0 \quad , \quad (\text{A.18})$$

with

$$\lambda_f \equiv \frac{q \bar{A}_f}{p A_f} . \quad (\text{A.19})$$

This form of CP violation can be observed, for example, using the asymmetry of neutral meson decays into final CP eigenstates  $f_{CP}$

$$\mathcal{A}_{f_{CP}}(t) \equiv \frac{d\Gamma/dt[\bar{B}^0_{\text{phys}}(t) \rightarrow f_{CP}] - d\Gamma/dt[B^0_{\text{phys}}(t) \rightarrow f_{CP}]}{d\Gamma/dt[\bar{B}^0_{\text{phys}}(t) \rightarrow f_{CP}] + d\Gamma/dt[B^0_{\text{phys}}(t) \rightarrow f_{CP}]} . \quad (\text{A.20})$$

For  $\Delta\Gamma = 0$  and  $|q/p| = 1$  (which is a good approximation for  $B$  mesons),  $\mathcal{A}_{f_{CP}}$  has a particularly simple form [85–87]:

$$\begin{aligned} \mathcal{A}_f(t) &= S_f \sin(\Delta m t) - C_f \cos(\Delta m t) \quad , \\ S_f &\equiv \frac{2 \operatorname{Im}(\lambda_f)}{1 + |\lambda_f|^2} \quad , \quad C_f \equiv \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2} \quad , \end{aligned} \quad (\text{A.21})$$

Consider the  $B \rightarrow f$  decay amplitude  $A_f$ , and the CP conjugate process,  $\bar{B} \rightarrow \bar{f}$ , with decay amplitude  $\bar{A}_{\bar{f}}$ . There are two types of phases that may appear in these decay amplitudes. Complex parameters in any Lagrangian term that contributes to the amplitude will appear in complex conjugate form in the CP-conjugate amplitude. Thus their phases appear in  $A_f$  and  $\bar{A}_{\bar{f}}$  with opposite signs. In the Standard Model, these phases occur only in the couplings of the  $W^{\pm}$  bosons and hence are often called “weak phases”. The weak phase of any single term is convention dependent. However, the difference between the weak phases in two different terms in  $A_f$  is convention independent. A second type of phase

can appear in scattering or decay amplitudes even when the Lagrangian is real. Their origin is the possible contribution from intermediate on-shell states in the decay process. Since these phases are generated by CP-invariant interactions, they are the same in  $A_f$  and  $\bar{A}_f$ . Usually the dominant rescattering is due to strong interactions and hence the designation ‘‘strong phases’’ for the phase shifts so induced. Again, only the relative strong phases between different terms in the amplitude are physically meaningful.

The ‘weak’ and ‘strong’ phases discussed here appear in addition to the ‘spurious’ CP-transformation phases of Eq. (A.3). Those spurious phases are due to an arbitrary choice of phase convention, and do not originate from any dynamics or induce any CP violation. For simplicity, we set them to zero from here on.

It is useful to write each contribution  $a_i$  to  $A_f$  in three parts: its magnitude  $|a_i|$ , its weak phase  $\phi_i$ , and its strong phase  $\delta_i$ . If, for example, there are two such contributions,  $A_f = a_1 + a_2$ , we have

$$\begin{aligned} A_f &= |a_1|e^{i(\delta_1+\phi_1)} + |a_2|e^{i(\delta_2+\phi_2)} \quad , \\ \bar{A}_f &= |a_1|e^{i(\delta_1-\phi_1)} + |a_2|e^{i(\delta_2-\phi_2)} \quad . \end{aligned} \quad (\text{A.22})$$

Similarly, for neutral meson decays, it is useful to write

$$M_{12} = |M_{12}|e^{i\phi_M} \quad , \quad \Gamma_{12} = |\Gamma_{12}|e^{i\phi_\Gamma} \quad . \quad (\text{A.23})$$

Each of the phases appearing in Eqs. (A.22,A.23) is convention dependent, but combinations such as  $\delta_1 - \delta_2$ ,  $\phi_1 - \phi_2$ ,  $\phi_M - \phi_\Gamma$  and  $\phi_M + \phi_1 - \bar{\phi}_1$  (where  $\bar{\phi}_1$  is a weak phase contributing to  $\bar{A}_f$ ) are physical.

In the approximations that only a single weak phase contributes to decay,  $A_f = |a_f|e^{i(\delta_f+\phi_f)}$ , and that  $|\Gamma_{12}/M_{12}| = 0$ , we obtain  $|\lambda_f| = 1$  and the CP asymmetries in decays to a final CP eigenstate  $f$  [Eq. (A.20)] with eigenvalue  $\eta_f = \pm 1$  are given by

$$A_{fCP}(t) = \mathcal{I}m(\lambda_f) \sin(\Delta mt) \quad \text{with} \quad \mathcal{I}m(\lambda_f) = \eta_f \sin(\phi_M + 2\phi_f) \quad . \quad (\text{A.24})$$

Note that the phase so measured is purely a weak phase, and no hadronic parameters are involved in the extraction of its value from  $\mathcal{I}m(\lambda_f)$ .

## References

- [1] C. Patrignani, *Chin. Phys.* **C40** (2016) 100001, [doi:10.1088/1674-1137/40/10/100001](https://doi.org/10.1088/1674-1137/40/10/100001).
- [2] M. Kobayashi and T. Maskawa, *Prog. Theor. Phys.* **49** (1973) 652, [doi:10.1143/PTP.49.652](https://doi.org/10.1143/PTP.49.652).
- [3] N. Cabibbo, *Phys. Rev. Lett.* **10** (1963) 531, [doi:10.1103/PhysRevLett.10.531](https://doi.org/10.1103/PhysRevLett.10.531).
- [4] A. B. Carter and A. I. Sanda, *Phys. Rev. Lett.* **45** (1980) 952, [doi:10.1103/PhysRevLett.45.952](https://doi.org/10.1103/PhysRevLett.45.952); *Phys. Rev.* **D23** (1981) 1567, [doi:10.1103/PhysRevD.23.1567](https://doi.org/10.1103/PhysRevD.23.1567).
- [5] I. I. Y. Bigi and A. I. Sanda, *Nucl. Phys.* **B193** (1981) 85, [doi:10.1016/0550-3213\(81\)90519-8](https://doi.org/10.1016/0550-3213(81)90519-8).
- [6] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, *Rev. Mod. Phys.* **68** (1996) 1125, [doi:10.1103/RevModPhys.68.1125](https://doi.org/10.1103/RevModPhys.68.1125), [arXiv:hep-ph/9512380](https://arxiv.org/abs/hep-ph/9512380).
- [7] Y. Grossman, A. L. Kagan and Z. Ligeti, *Phys. Lett.* **B538** (2002) 327, [doi:10.1016/S0370-2693\(02\)02027-0](https://doi.org/10.1016/S0370-2693(02)02027-0), [arXiv:hep-ph/0204212](https://arxiv.org/abs/hep-ph/0204212).
- [8] H. Boos, T. Mannel and J. Reuter, *Phys. Rev.* **D70** (2004) 036006, [doi:10.1103/PhysRevD.70.036006](https://doi.org/10.1103/PhysRevD.70.036006), [arXiv:hep-ph/0403085](https://arxiv.org/abs/hep-ph/0403085).
- [9] H. n. Li and S. Mishima, *JHEP* **03** (2007) 009, [doi:10.1088/1126-6708/2007/03/009](https://doi.org/10.1088/1126-6708/2007/03/009), [arXiv:hep-ph/0610120](https://arxiv.org/abs/hep-ph/0610120).
- [10] M. Gronau and J. L. Rosner, *Phys. Lett.* **B672** (2009) 349, [doi:10.1016/j.physletb.2009.01.049](https://doi.org/10.1016/j.physletb.2009.01.049), [arXiv:0812.4796](https://arxiv.org/abs/0812.4796) [hep-ph].
- [11] Y. Amhis *et al.* [Heavy Flavor Averaging Group Collaboration], [arXiv:1207.1158](https://arxiv.org/abs/1207.1158) [hep-ex] and online update at <http://www.slac.stanford.edu/xorg/hfag>.

- [12] CKMfitter Group (J. Charles et al.), *Eur. Phys. J.* **C41** (2005) 1, doi:[10.1140/epjc/s2005-02169-1](https://doi.org/10.1140/epjc/s2005-02169-1), arXiv:[hep-ph/0406184](https://arxiv.org/abs/hep-ph/0406184), updated results and plots available at: <http://ckmfitter.in2p3.fr>.
- [13] Y. Nir, *Nucl. Phys. Proc. Suppl.* **117** (2003) 111, doi:[10.1016/S0920-5632\(03\)01414-2](https://doi.org/10.1016/S0920-5632(03)01414-2), arXiv:[hep-ph/0208080](https://arxiv.org/abs/hep-ph/0208080).
- [14] Y. Grossman, Y. Nir and M. P. Worah, *Phys. Lett.* **B407** (1997) 307, doi:[10.1016/S0370-2693\(97\)00675-8](https://doi.org/10.1016/S0370-2693(97)00675-8), arXiv:[hep-ph/9704287](https://arxiv.org/abs/hep-ph/9704287).
- [15] Y. Grossman, Y. Nir and G. Raz, *Phys. Rev. Lett.* **97** (2006) 151801, doi:[10.1103/PhysRevLett.97.151801](https://doi.org/10.1103/PhysRevLett.97.151801), arXiv:[hep-ph/0605028](https://arxiv.org/abs/hep-ph/0605028).
- [16] G. C. Branco, L. Lavoura and J. P. Silva, *CP violation*, (Clarendon Press, Oxford, 1999).
- [17] I. I. Y. Bigi and N. G. Uraltsev, *Nucl. Phys. B* **592**, 92 (2001) [arXiv:[hep-ph/0005089](https://arxiv.org/abs/hep-ph/0005089)].
- [18] A. F. Falk, Y. Grossman, Z. Ligeti and A. A. Petrov, *Phys. Rev.* **D65** (2002) 054034, doi:[10.1103/PhysRevD.65.054034](https://doi.org/10.1103/PhysRevD.65.054034), arXiv:[hep-ph/0110317](https://arxiv.org/abs/hep-ph/0110317).
- [19] A. F. Falk, Y. Grossman, Z. Ligeti, Y. Nir and A. A. Petrov, *Phys. Rev.* **D69** (2004) 114021, doi:[10.1103/PhysRevD.69.114021](https://doi.org/10.1103/PhysRevD.69.114021), arXiv:[hep-ph/0402204](https://arxiv.org/abs/hep-ph/0402204).
- [20] G. Raz, *Phys. Rev.* **D66** (2002) 037701, doi:[10.1103/PhysRevD.66.037701](https://doi.org/10.1103/PhysRevD.66.037701), arXiv:[hep-ph/0205310](https://arxiv.org/abs/hep-ph/0205310).
- [21] G. Isidori, Y. Nir and G. Perez, *Ann. Rev. Nucl. Part. Sci.* **60** (2010) 355, doi:[10.1146/annurev.nucl.012809.104534](https://doi.org/10.1146/annurev.nucl.012809.104534), arXiv:[1002.0900](https://arxiv.org/abs/1002.0900) [hep-ph].
- [22] N. Arkani-Hamed and S. Dimopoulos, *JHEP* **06** (2005) 073, doi:[10.1088/1126-6708/2005/06/073](https://doi.org/10.1088/1126-6708/2005/06/073), arXiv:[hep-th/0405159](https://arxiv.org/abs/hep-th/0405159).
- [23] A. G. Cohen, D. B. Kaplan and A. E. Nelson, *Phys. Lett.* **B388** (1996) 588, doi:[10.1016/S0370-2693\(96\)01183-5](https://doi.org/10.1016/S0370-2693(96)01183-5), arXiv:[hep-ph/9607394](https://arxiv.org/abs/hep-ph/9607394).
- [24] M. Ciuchini, E. Franco, D. Guadagnoli, V. Lubicz, M. Pierini, V. Porretti and L. Silvestrini, *Phys. Lett.* **B655** (2007) 162, doi:[10.1016/j.physletb.2007.08.055](https://doi.org/10.1016/j.physletb.2007.08.055), arXiv:[hep-ph/0703204](https://arxiv.org/abs/hep-ph/0703204).
- [25] Y. Nir, *JHEP* **05** (2007) 102, doi:[10.1088/1126-6708/2007/05/102](https://doi.org/10.1088/1126-6708/2007/05/102), arXiv:[hep-ph/0703235](https://arxiv.org/abs/hep-ph/0703235).
- [26] O. Gedalia, J. F. Kamenik, Z. Ligeti and G. Perez, *Phys. Lett.* **B714** (2012) 55, doi:[10.1016/j.physletb.2012.06.050](https://doi.org/10.1016/j.physletb.2012.06.050), arXiv:[1202.5038](https://arxiv.org/abs/1202.5038) [hep-ph].
- [27] K. Blum, Y. Grossman, Y. Nir and G. Perez, *Phys. Rev. Lett.* **102** (2009) 211802, doi:[10.1103/PhysRevLett.102.211802](https://doi.org/10.1103/PhysRevLett.102.211802), arXiv:[0903.2118](https://arxiv.org/abs/0903.2118) [hep-ph].
- [28] G. D'Ambrosio, G. F. Giudice, G. Isidori and A. Strumia, *Nucl. Phys.* **B645** (2002) 155, doi:[10.1016/S0550-3213\(02\)00836-2](https://doi.org/10.1016/S0550-3213(02)00836-2), arXiv:[hep-ph/0207036](https://arxiv.org/abs/hep-ph/0207036).
- [29] C. D. Froggatt and H. B. Nielsen, *Nucl. Phys.* **B147** (1979) 277, doi:[10.1016/0550-3213\(79\)90316-X](https://doi.org/10.1016/0550-3213(79)90316-X).
- [30] M. Dine, R. G. Leigh and A. Kagan, *Phys. Rev.* **D48** (1993) 4269, doi:[10.1103/PhysRevD.48.4269](https://doi.org/10.1103/PhysRevD.48.4269), arXiv:[hep-ph/9304299](https://arxiv.org/abs/hep-ph/9304299).
- [31] A. E. Nelson and M. J. Strassler, *JHEP* **09** (2000) 030, doi:[10.1088/1126-6708/2000/09/030](https://doi.org/10.1088/1126-6708/2000/09/030), arXiv:[hep-ph/0006251](https://arxiv.org/abs/hep-ph/0006251).
- [32] N. Arkani-Hamed and M. Schmaltz, *Phys. Rev.* **D61** (2000) 033005, doi:[10.1103/PhysRevD.61.033005](https://doi.org/10.1103/PhysRevD.61.033005), arXiv:[hep-ph/9903417](https://arxiv.org/abs/hep-ph/9903417).
- [33] W. Altmannshofer, C. Frugiuele and R. Harnik, *JHEP* **12** (2014) 180, doi:[10.1007/JHEP12\(2014\)180](https://doi.org/10.1007/JHEP12(2014)180), arXiv:[1409.2522](https://arxiv.org/abs/1409.2522) [hep-ph].
- [34] M. C. Gonzalez-Garcia, M. Maltoni, J. Salvado and T. Schwetz, *JHEP* **12** (2012) 123, doi:[10.1007/JHEP12\(2012\)123](https://doi.org/10.1007/JHEP12(2012)123), arXiv:[1209.3023](https://arxiv.org/abs/1209.3023) [hep-ph];  
M. C. Gonzalez-Garcia, M. Maltoni and T. Schwetz, *Nucl. Phys.* **B908** (2016) 199, doi:[10.1016/j.nuclphysb.2016.02.033](https://doi.org/10.1016/j.nuclphysb.2016.02.033), arXiv:[1512.06856](https://arxiv.org/abs/1512.06856) [hep-ph].
- [35] L. J. Hall, H. Murayama and N. Weiner, *Phys. Rev. Lett.* **84** (2000) 2572, doi:[10.1103/PhysRevLett.84.2572](https://doi.org/10.1103/PhysRevLett.84.2572), arXiv:[hep-ph/9911341](https://arxiv.org/abs/hep-ph/9911341).

- [36] N. Haba and H. Murayama, *Phys. Rev.* **D63** (2001) 053010, doi:[10.1103/PhysRevD.63.053010](https://doi.org/10.1103/PhysRevD.63.053010), arXiv:[hep-ph/0009174](https://arxiv.org/abs/hep-ph/0009174).
- [37] A. de Gouvea and H. Murayama, *Phys. Lett.* **B573** (2003) 94, doi:[10.1016/j.physletb.2003.08.045](https://doi.org/10.1016/j.physletb.2003.08.045), arXiv:[hep-ph/0301050](https://arxiv.org/abs/hep-ph/0301050); *Phys. Lett.* **B747** (2015) 479, doi:[10.1016/j.physletb.2015.06.028](https://doi.org/10.1016/j.physletb.2015.06.028), arXiv:[1204.1249](https://arxiv.org/abs/1204.1249) [hep-ph].
- [38] S. Amitai, arXiv:[1211.6252](https://arxiv.org/abs/1211.6252) [hep-ph].
- [39] P. F. Harrison, D. H. Perkins and W. G. Scott, *Phys. Lett.* **B530** (2002) 167, doi:[10.1016/S0370-2693\(02\)01336-9](https://doi.org/10.1016/S0370-2693(02)01336-9), arXiv:[hep-ph/0202074](https://arxiv.org/abs/hep-ph/0202074).
- [40] E. Ma and G. Rajasekaran, *Phys. Rev.* **D64** (2001) 113012, doi:[10.1103/PhysRevD.64.113012](https://doi.org/10.1103/PhysRevD.64.113012), arXiv:[hep-ph/0106291](https://arxiv.org/abs/hep-ph/0106291).
- [41] G. Altarelli and F. Feruglio, *Rev. Mod. Phys.* **82** (2010) 2701, doi:[10.1103/RevModPhys.82.2701](https://doi.org/10.1103/RevModPhys.82.2701), arXiv:[1002.0211](https://arxiv.org/abs/1002.0211) [hep-ph]; *Nucl. Phys.* **B741** (2006) 215, doi:[10.1016/j.nuclphysb.2006.02.015](https://doi.org/10.1016/j.nuclphysb.2006.02.015), arXiv:[hep-ph/0512103](https://arxiv.org/abs/hep-ph/0512103).
- [42] F. P. An *et al.* [DAYA-BAY Collaboration], *Phys. Rev. Lett.* **108** (2012) 171803, doi:[10.1103/PhysRevLett.108.171803](https://doi.org/10.1103/PhysRevLett.108.171803), arXiv:[1203.1669](https://arxiv.org/abs/1203.1669) [hep-ex].
- [43] S. Amitai, arXiv:[1212.5165](https://arxiv.org/abs/1212.5165) [hep-ph].
- [44] G. Aad *et al.* [ATLAS Collaboration], *Phys. Lett.* **B716** (2012) 1, doi:[10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020), arXiv:[1207.7214](https://arxiv.org/abs/1207.7214) [hep-ex].
- [45] S. Chatrchyan *et al.* [CMS Collaboration], *Phys. Lett.* **B716** (2012) 30, doi:[10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021), arXiv:[1207.7235](https://arxiv.org/abs/1207.7235) [hep-ex].
- [46] G. Aad *et al.* [ATLAS and CMS Collaborations], *JHEP* **08** (2016) 045, doi:[10.1007/JHEP08\(2016\)045](https://doi.org/10.1007/JHEP08(2016)045), arXiv:[1606.02266](https://arxiv.org/abs/1606.02266) [hep-ex].
- [47] A. M. Sirunyan *et al.* [CMS Collaboration], *Phys. Rev. Lett.* **120** (2018) 231801, doi:[10.1103/PhysRevLett.120.231801](https://doi.org/10.1103/PhysRevLett.120.231801), arXiv:[1804.02610](https://arxiv.org/abs/1804.02610) [hep-ex].
- [48] M. Aaboud *et al.* [ATLAS Collaboration], *Phys. Lett.* **B784** (2018) 173, doi:[10.1016/j.physletb.2018.07.035](https://doi.org/10.1016/j.physletb.2018.07.035), arXiv:[1806.00425](https://arxiv.org/abs/1806.00425) [hep-ex].
- [49] M. Aaboud *et al.* [ATLAS Collaboration], *JHEP* **12** (2017) 024, doi:[10.1007/JHEP12\(2017\)024](https://doi.org/10.1007/JHEP12(2017)024), arXiv:[1708.03299](https://arxiv.org/abs/1708.03299) [hep-ex].
- [50] A. M. Sirunyan *et al.* [CMS Collaboration], *Phys. Lett.* **B780** (2018) 501, doi:[10.1016/j.physletb.2018.02.050](https://doi.org/10.1016/j.physletb.2018.02.050), arXiv:[1709.07497](https://arxiv.org/abs/1709.07497) [hep-ex].
- [51] V. Khachatryan *et al.* [CMS Collaboration], *Phys. Lett.* **B744** (2015) 184, doi:[10.1016/j.physletb.2015.03.048](https://doi.org/10.1016/j.physletb.2015.03.048), arXiv:[1410.6679](https://arxiv.org/abs/1410.6679) [hep-ex].
- [52] M. Aaboud *et al.* [ATLAS Collaboration], *Phys. Rev. Lett.* **119** (2017) 051802, doi:[10.1103/PhysRevLett.119.051802](https://doi.org/10.1103/PhysRevLett.119.051802), arXiv:[1705.04582](https://arxiv.org/abs/1705.04582) [hep-ex].
- [53] A. M. Sirunyan *et al.* [CMS Collaboration], *Phys. Rev. Lett.* **122** (2019) 021801, doi:[10.1103/PhysRevLett.122.021801](https://doi.org/10.1103/PhysRevLett.122.021801), arXiv:[1807.06325](https://arxiv.org/abs/1807.06325) [hep-ex].
- [54] M. Aaboud *et al.* [ATLAS Collaboration], *JHEP* **10** (2017) 129, doi:[10.1007/JHEP10\(2017\)129](https://doi.org/10.1007/JHEP10(2017)129), arXiv:[1707.01404](https://arxiv.org/abs/1707.01404) [hep-ex].
- [55] V. Khachatryan *et al.* [CMS Collaboration], *JHEP* **02** (2017) 079, doi:[10.1007/JHEP02\(2017\)079](https://doi.org/10.1007/JHEP02(2017)079), arXiv:[1610.04857](https://arxiv.org/abs/1610.04857) [hep-ex].
- [56] G. Aad *et al.* [ATLAS Collaboration], *Eur. Phys. J.* **C77** (2017) 70, doi:[10.1140/epjc/s10052-017-4624-0](https://doi.org/10.1140/epjc/s10052-017-4624-0), arXiv:[1604.07730](https://arxiv.org/abs/1604.07730) [hep-ex].
- [57] A. M. Sirunyan *et al.* [CMS Collaboration], *JHEP* **06** (2018) 001, doi:[10.1007/JHEP06\(2018\)001](https://doi.org/10.1007/JHEP06(2018)001), arXiv:[1712.07173](https://arxiv.org/abs/1712.07173) [hep-ex].
- [58] V. Khachatryan *et al.* [CMS Collaboration], *Phys. Lett.* **B763** (2016) 472, doi:[10.1016/j.physletb.2016.09.062](https://doi.org/10.1016/j.physletb.2016.09.062), arXiv:[1607.03561](https://arxiv.org/abs/1607.03561) [hep-ex].

- [59] A. Dery, A. Efrati, Y. Hochberg and Y. Nir, *JHEP* **05** (2013) 039, [doi:10.1007/JHEP05\(2013\)039](https://doi.org/10.1007/JHEP05(2013)039), [arXiv:1302.3229](https://arxiv.org/abs/1302.3229) [hep-ph].
- [60] G. Blankenburg, J. Ellis and G. Isidori, *Phys. Lett.* **B712** (2012) 386, [doi:10.1016/j.physletb.2012.05.007](https://doi.org/10.1016/j.physletb.2012.05.007), [arXiv:1202.5704](https://arxiv.org/abs/1202.5704) [hep-ph].
- [61] R. Harnik, J. Kopp and J. Zupan, *JHEP* **03** (2013) 026, [doi:10.1007/JHEP03\(2013\)026](https://doi.org/10.1007/JHEP03(2013)026), [arXiv:1209.1397](https://arxiv.org/abs/1209.1397) [hep-ph].
- [62] S. Davidson and P. Verdier, *Phys. Rev.* **D86** (2012) 111701, [doi:10.1103/PhysRevD.86.111701](https://doi.org/10.1103/PhysRevD.86.111701), [arXiv:1211.1248](https://arxiv.org/abs/1211.1248) [hep-ph].
- [63] A. Arhrib, Y. Cheng and O. C. W. Kong, *Phys. Rev.* **D87** (2013) 015025, [doi:10.1103/PhysRevD.87.015025](https://doi.org/10.1103/PhysRevD.87.015025), [arXiv:1210.8241](https://arxiv.org/abs/1210.8241) [hep-ph].
- [64] Y. Grossman and Y. Nir, *Nucl. Phys.* **B448** (1995) 30, [doi:10.1016/0550-3213\(95\)00203-5](https://doi.org/10.1016/0550-3213(95)00203-5), [arXiv:hep-ph/9502418](https://arxiv.org/abs/hep-ph/9502418).
- [65] M. Leurer, Y. Nir and N. Seiberg, *Nucl. Phys.* **B420** (1994) 468, [doi:10.1016/0550-3213\(94\)90074-4](https://doi.org/10.1016/0550-3213(94)90074-4), [arXiv:hep-ph/9310320](https://arxiv.org/abs/hep-ph/9310320).
- [66] D. Aloni, Y. Nir and E. Stamou, *JHEP* **04** (2016) 162, [doi:10.1007/JHEP04\(2016\)162](https://doi.org/10.1007/JHEP04(2016)162), [arXiv:1511.00979](https://arxiv.org/abs/1511.00979) [hep-ph].
- [67] R. Aaij *et al.* [LHCb Collaboration], *Phys. Rev. Lett.* **113** (2014) 151601, [doi:10.1103/PhysRevLett.113.151601](https://doi.org/10.1103/PhysRevLett.113.151601), [arXiv:1406.6482](https://arxiv.org/abs/1406.6482) [hep-ex].
- [68] R. Aaij *et al.* [LHCb Collaboration], *JHEP* **08** (2017) 055, [doi:10.1007/JHEP08\(2017\)055](https://doi.org/10.1007/JHEP08(2017)055), [arXiv:1705.05802](https://arxiv.org/abs/1705.05802) [hep-ex].
- [69] D. Aloni, A. Dery, C. Frugiuele and Y. Nir, *JHEP* **11** (2017) 109, [doi:10.1007/JHEP11\(2017\)109](https://doi.org/10.1007/JHEP11(2017)109), [arXiv:1708.06161](https://arxiv.org/abs/1708.06161) [hep-ph].
- [70] J. P. Lees *et al.* [BaBar Collaboration], *Phys. Rev. Lett.* **109** (2012) 101802, [doi:10.1103/PhysRevLett.109.101802](https://doi.org/10.1103/PhysRevLett.109.101802), [arXiv:1205.5442](https://arxiv.org/abs/1205.5442) [hep-ex].
- [71] J. P. Lees *et al.* [BaBar Collaboration], *Phys. Rev.* **D88** (2013) 072012, [doi:10.1103/PhysRevD.88.072012](https://doi.org/10.1103/PhysRevD.88.072012), [arXiv:1303.0571](https://arxiv.org/abs/1303.0571) [hep-ex].
- [72] M. Huschle *et al.* [Belle Collaboration], *Phys. Rev.* **D92** (2015) 072014, [doi:10.1103/PhysRevD.92.072014](https://doi.org/10.1103/PhysRevD.92.072014), [arXiv:1507.03233](https://arxiv.org/abs/1507.03233) [hep-ex].
- [73] A. Abdesselam *et al.* [Belle Collaboration], [arXiv:1603.06711](https://arxiv.org/abs/1603.06711) [hep-ex].
- [74] R. Aaij *et al.* [LHCb Collaboration], *Phys. Rev. Lett.* **115** (2015) 111803, [doi:10.1103/PhysRevLett.115.111803](https://doi.org/10.1103/PhysRevLett.115.111803), Erratum: [*Phys. Rev. Lett.* **115** (2015)] 159901], [arXiv:1506.08614](https://arxiv.org/abs/1506.08614) [hep-ex].
- [75] Y. Amhis *et al.* [Heavy Flavor Averaging Group (HFAG) Collaboration], [arXiv:1412.7515](https://arxiv.org/abs/1412.7515) [hep-ex].
- [76] D. A. Faroughy, A. Greljo and J. F. Kamenik, *Phys. Lett.* **B764** (2017) 126, [doi:10.1016/j.physletb.2016.11.011](https://doi.org/10.1016/j.physletb.2016.11.011) [arXiv:1609.07138](https://arxiv.org/abs/1609.07138) [hep-ph].
- [77] A. Greljo, G. Isidori and D. Marzocca, *JHEP* **07** (2015) 142, [doi:10.1007/JHEP07\(2015\)142](https://doi.org/10.1007/JHEP07(2015)142), [arXiv:1506.01705](https://arxiv.org/abs/1506.01705) [hep-ph].
- [78] L. Chau and W. Keung, *Phys. Rev. Lett.* **53** (1984) 1802, [doi:10.1103/PhysRevLett.53.1802](https://doi.org/10.1103/PhysRevLett.53.1802).
- [79] L. Wolfenstein, *Phys. Rev. Lett.* **51** (1983) 1945, [doi:10.1103/PhysRevLett.51.1945](https://doi.org/10.1103/PhysRevLett.51.1945).
- [80] A. J. Buras, M. E. Lautenbacher, and G. Ostermaier, *Phys. Rev.* **D50** (1994) 3433, [doi:10.1103/PhysRevD.50.3433](https://doi.org/10.1103/PhysRevD.50.3433), [arXiv:hep-ph/9403384](https://arxiv.org/abs/hep-ph/9403384).
- [81] C. Dib, I. Dunietz, F. J. Gilman and Y. Nir, *Phys. Rev.* **D41** (1990) 1522, [doi:10.1103/PhysRevD.41.1522](https://doi.org/10.1103/PhysRevD.41.1522).
- [82] J. L. Rosner, A. I. Sanda and M. P. Schmidt, *Conf. Proc.* **C871111** (1987) 165, presented at Workshop on High Sensitivity Beauty Physics, Batavia, IL, Nov 11-14, 1987, <https://inspirehep.net/literature/258107>.

- [83] V. Weisskopf and E. P. Wigner, *Z. Phys.* **63** (1930) 54, [doi:10.1007/BF01336768](https://doi.org/10.1007/BF01336768);  
*Z. Phys.* **65** (1930) 18, [doi:10.1007/BF01397406](https://doi.org/10.1007/BF01397406);  
P. K. Kabir, *The CP Puzzle: Strange Decays of the Neutral Kaon*, (Academic Press (1968),  
Appendix A.
- [84] Y. Nir, *Conf. Proc. C9207131* (1992) 81, lectures given at 20th Annual SLAC Summer Institute on  
Particle Physics (Stanford, CA, 1992), <https://inspirehep.net/literature/356250>.
- [85] I. Dunietz and J. L. Rosner, *Phys. Rev.* **D34** (1986) 1404, [doi:10.1103/PhysRevD.34.1404](https://doi.org/10.1103/PhysRevD.34.1404).
- [86] Ya. I. Azimov, N. G. Uraltsev, and V. A. Khoze, *Sov. J. Nucl. Phys.* **45** (1987) 878, [*Yad. Fiz.* **45**  
(1987) 1412].
- [87] I. I. Bigi and A. I. Sanda, *Nucl. Phys.* **B281** (1987) 41, [doi:10.1016/0550-3213\(87\)90246-X](https://doi.org/10.1016/0550-3213(87)90246-X).
- [88] Y. Nir and G. Raz, *Phys. Rev.* **D66** (2002) 035007, [doi:10.1103/PhysRevD.66.035007](https://doi.org/10.1103/PhysRevD.66.035007),  
[arXiv:hep-ph/0206064](https://arxiv.org/abs/hep-ph/0206064).



## Heavy-ion physics

*M. Nguyen*

Laboratoire Leprince-Ringuet (CNRS), Palaiseau, France

### Abstract

These proceedings cover two lectures on heavy-ion physics given at the 2018 edition of the Asia-Europe-Pacific School of High-Energy Physics. After a brief introduction of basic concepts that are relevant for heavy-ion collisions, I discuss several interesting phenomena that have been observed, and mention currently still open questions.

### Keywords

Lectures; heavy ion collisions; quark-gluon plasma; jet quenching; strangeness enhancement; collective flow.

## 1 Introduction

These lectures are an introduction to the physics of ultra-relativistic heavy-ion collisions and the quark-gluon plasma (QGP). In the first lecture, I start with a brief review of some salient features of quantum chromodynamics (QCD), which is the underlying theory that describes the dynamics of the QGP, the calculational difficulties of that theory notwithstanding. I then discuss properties of the QGP as seen through the lens of thermodynamics, which turns out to be an appropriate language with which to describe heavy-ion collisions, due to the typically large number of individual particles in such collisions. I then trace out the space-time evolution of a heavy-ion collision focusing first on collision centrality, a key collision property employed in nearly all heavy-ion measurements. Then I look at a simple picture attributed to J.D. Bjorken, who estimated the energy density of the QGP from mid-rapidity particle yields. I then turn to hadro-chemistry, first looking at the statistical hadroproduction model, which connects the abundance of the various particle species to the properties of the QCD matter prior to hadronization. Then I focus on the particular role of strangeness in this type of model, and what it tells us about chemical equilibrium.

In the second lecture I focus on the various phenomena that we use to infer the properties of the QGP. There are many, and I have not attempted a complete survey. Rather I focused on three types of measurements, which have been of particular interest of late, namely those related to collective flow, jet quenching and quarkonia dissociation. Finally, I close the second set of lectures with a discussion of “small systems”, mainly referring to proton-proton and proton-ion collisions, which surprisingly exhibit certain effects previously only expected in ion-ion collisions.

## 2 Generalities

Let us begin by defining the term “heavy-ion collision”. Ions are of course nuclei that have been stripped of their electrons, which is necessary to accelerate them to large energies. The Large Hadron Collider typically collides the lightest possible ion, the hydrogen ion, also known as the proton. By heavy ions, we mean nuclei with  $O(100)$  nucleons. At the Relativistic Heavy-Ion Collider (RHIC) at Brookhaven, the most commonly collided ion is gold (Au), although quite a few other species, the heaviest being the oblong uranium (U) ion, have also been collided. The Au isotope which is collided (the only stable one) consists of 197 nucleons, 79 protons and 118 neutrons. At the Large Hadron Collider the most frequently employed ion is Pb(208), with 82 protons and 126 neutrons. Due to the difficulty of isolating this particular isotope with high purity, the Pb source is perhaps surprisingly *more* expensive than Au. The center of mass energy of the highest energy ion collisions at the LHC is about 5 TeV per nucleon pair. That corresponds to roughly a PeV of total energy! Not all of this energy is however available for particle production. This energy will be estimated in Bjorken’s picture in Section 5.

The primary aim of colliding heavy ions is to observe matter in the deconfined phase, that of the quark-gluon plasma. Such a phase transition was predicted not long after the discovery of asymptotic freedom, first at large density [1], for example at the core of compact astrophysical objects, and then at large temperature [2], as in the first few microseconds of the universe. Although finite density calculations are complicated by the so-called sign problem, the temperature of the phase transition at vanishing net baryon density is amenable to calculation in lattice QCD. State-of-the-art calculations give a critical temperature of around 160 MeV [3,4], which corresponds to about  $10^{12}$  K! As it turns out, this temperature coincides well with an estimate for a “limiting temperature” of hadronic matter due to Hagedorn, that pre-dates the advent of QCD [5]. Hagedorn’s estimate was based on the exponential growth in the number of bound states with increasing energy, and is intimately related to the statistical model of hadroproduction that will be discussed in Section 6. We will see in Section 5 that the energy delivered by heavy-ion collisions is sufficient to reach the critical temperature.

### 3 Brief reminders of QCD

The Lagrangian of QCD is similar in its form to that of QED, but with different constants related to the SU(3) structure, which are defined by the commutation relations of the generators of the group. The non-trivial part of the theory, which limits the applicability of perturbative calculations to hard processes, is the self-coupling of the gauge field, in other words the interaction between gluons. The particle content of the theory is given by the gluons, which come in 8 different linearly independent color combinations, and by the quarks, which come in 3 colors, but also in 6 different flavors, which have very different masses. Most of the matter in the nucleon has nothing to do with the bare quark mass, but is rather generated dynamically by virtual quarks and gluons that buzz around the valence quarks. The bare quark masses imparted by the Higgs mechanisms are basically free parameters as far as QCD is concerned. These masses will turn out to be important for the discussion of the QGP. The light quarks, up and down, have masses on the order of a few MeV, that are small compared to the temperature of the QGP, such that the energy can easily be converted to creation of these particles. The strange quark mass of around 100 MeV is of the same order as the transition temperature. The heavy quarks, charm and bottom, are interesting precisely because they cannot be created easily by the plasma. In Section 10 I will discuss how bound states of these quarks are modified by the QGP. Top quarks are so heavy that they decay before the QGP is formed, so I will not discuss them in these lectures.

The reason that the gluon self-interaction is important has to do with the effect it has on the QCD coupling constant. In QED virtual particles screen the (electric) charge, such that the coupling grows weaker with distance. In QCD the self-interaction of gluons causes the opposite effect, such that the coupling increases with distance. The small coupling at short distance scales is what allows us to probe the quarks inside the nucleons with deep inelastic scattering experiments. This is also what allows us to solve QCD using a perturbative expansion. But this linearly increasing potential, rather than one which goes to zero in QED, is what keeps quarks confined inside hadrons. It is also what makes the low energy behavior of QCD difficult to calculate analytically. In this regime, one has to resort to lattice QCD calculations, which we’ll come back to in a moment.

### 4 QCD thermodynamics

First we can get some guidance on what to expect from QCD matter via thermodynamics. The idea of a QGP emerges directly from the concept of asymptotic freedom. In the limit of large temperature, the coupling constant approaches zero, so one ends up with a weakly interacting gas of quarks and gluons. That is not to say that the QGP we create in the laboratory satisfies this condition. As we’ll see, this is quite far from what actually happens. Rather the QGP just above the confinement temperature is strongly coupled. However, it’s still instructive to consider this limit, where we can derive the equation of state directly from kinetic theory.

The equation of state is usually expressed as the relationship between the pressure and the temperature. Equivalently we may express the E.O.S. in terms of the energy density and temperature. From kinetic theory, this is simply given by the density of states multiplied by their energy:

$$\epsilon(T) = \int \frac{d^3p}{(2\pi)^3} \sum_i \frac{E_i}{e^{E_i/T} \pm 1} \quad . \quad (1)$$

Recall that the plus-minus takes into account quantum effects. It would be zero for a classical system, i.e., Maxwell-Boltzmann statistics. The plus corresponds to fermions, i.e., Fermi-Dirac statistics, and the minus to bosons, i.e., Bose-Einstein statistics. The sum is over the different species of particles. Performing the integration (by parts), one obtains

$$\epsilon(T) = \frac{\pi}{30} N T^4 \quad . \quad (2)$$

If one performs the integration, one finds that the energy density is proportional to  $T^4$ , the familiar dependence of blackbody radiation.  $N$  is the number of different states. For matter below the critical temperature, one may approximate matter to be a pion gas, as pions are the lightest and therefore most abundant particle. The pions form an isospin triplet, so  $N=3$ . For the QGP, one obtains the following result, which is identical save for a slightly different numerical pre-factor for fermions:

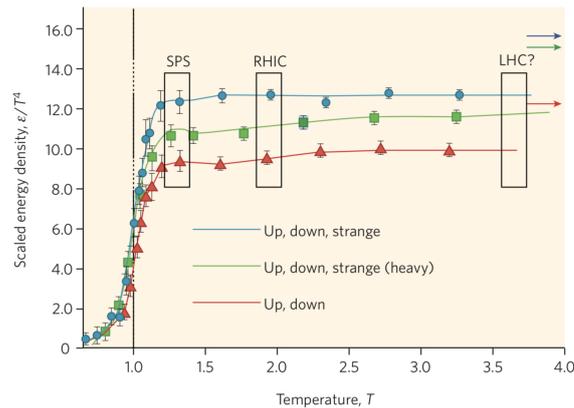
$$\epsilon(T) = \frac{\pi}{30} \left( N_B + \frac{7}{8} N_F \right) T^4 \quad . \quad (3)$$

The number of degrees of freedom is much larger for the case of the QGP, however. For the bosons one has 8 different linearly independent color combinations. As a massless spin-1 boson, the gluon also has two allowable spin states (compare to 3 spin states for the massive pion). As fermions, the quarks each contribute two spin states. Another factor of two comes from the anti-quarks, as well as a factor of 3 for the three colors of the SU(3) group. Finally we have flavor. Here we can consider that 3 colors are available to be created, up, down and strange. In practice, strange production is somewhat suppressed by its mass, which we've neglected, but it doesn't change the result dramatically. In total, instead of 3, we find 47.5 for the number of degrees of freedom, which is about a factor 15 jump in the energy density at  $T_c$ .

Now we can compare what we get from kinetic theory for an ideal gas to what we get from the lattice QCD. The equation of state from the lattice is shown in Fig. 1. Indeed one finds a large jump in the ratio of  $\epsilon/T^4$ , indicative of a jump in the number of degrees of freedom. Inserting a critical temperature of 160 MeV into equation 3, one obtains an energy density of around 1 GeV/fm<sup>3</sup><sup>1</sup>.

Although the height of the plateau above  $T_c$  in Fig. 1 depends a bit on how exactly one treats the strange quark, one observes a jump all the same. It's reasonable to ask, however, whether the ideal gas E.O.S. is actually a good description of the QGP we observe in the lab. Note that although this jump is a large factor, it does not quite saturate the Stefan-Boltzmann limit, indicated by arrows on the figure. To get some intuition, it's useful to look at N=4 Super Yang Mills, which is a 10 dimensional theory that has nothing to do with QCD. But this theory is a dual to a conformal field theory, which is a particular kind of quantum field theory that is scale invariant. Such a theory is very much the opposite of an ideal gas, as it has infinitely strong coupling. So even though this is not QCD, one may reasonably expect that the details of the theory aren't important in the strong coupling limit. And so it turns out that the scaled energy density is rather close to what one gets in such a theory. So this is by no means a proof that the QGP is strongly coupled, however it gives a hint that this departure from the weak-coupling limit could be significant. We will see evidence of the strongly coupled behavior of the plasma in Section 8.

<sup>1</sup>Keep in mind the handy conversion factor for natural units:  $\hbar c = 0.197$  GeV fm



**Fig. 1:** A lattice QCD calculation of the equation of state of matter in terms of the scaled energy density, as a function of temperature (normalized by the critical temperature for a phase transition) [6, 7]. The lattice data demonstrate a precipitous jump near  $T_c$ , as expected for a phase transition from hadronic to partonic degrees of freedom.

## 5 Space-time evolution

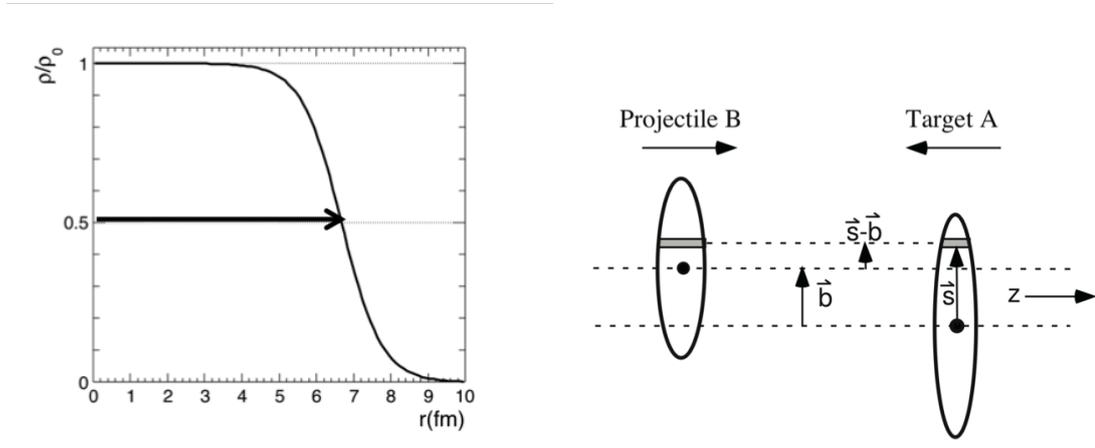
So far we've been discussing the quark-gluon plasma as if it's a static object, when in fact it's a very transient state. Before the collision, if we are standing in the lab frame, the two nuclei appear as flattened pancakes. These pancakes pass through one another, and the particles start to scatter. The time taken for the scattering to occur can be roughly estimated from the uncertainty principle, as one over the momentum of the particles. This implies that by a time of 1 fm, the bulk of particles have been created and can rescatter with each other. We'll see later from measurements of collectivity that we do in fact have evidence that thermalization is achieved over such a short timescale. Then the QGP expands and cools until about 10 fm. At this point the temperature falls below the critical one and hadronization occurs. This is referred to as chemical freeze-out as the abundances of different particle species are fixed, barring long-lived decays. There's a short period of additional rescattering before the particles stop interacting and free stream to our detectors. So what we end up detecting is a mess of thousands of final state hadrons. What I'll try to show you is how we can learn about the QGP, which lasts shorter than 10 fm, which is something like  $3 \times 10^{-23}$  seconds!

In addition to the final state particles, we also need to consider the role of the initial state of the colliding nuclei. In particular, it's important to consider the special role that geometry plays in heavy-ion collisions. When collisions have a relatively small impact parameter, the overlap region is going to be relatively large. This is referred to as a central event. In more glancing or "peripheral" collisions, one expects nuclear effects to be smaller, eventually looking very much like proton collisions in the limit of two nucleon scattering. Nucleons that scatter are referred to as participants, while others are spectators. Generally in a collider experiment one cannot directly measure how many nucleons participate in the collision, event-by-event. So if the energy density of the matter depends crucially on impact parameter, i.e., collision "centrality", how can we access this in our experiments?

The approach used is to make use of a so-called Glauber model, named after Nobel prize winner Roy Glauber, who passed away since these lectures were given. The model is based on work he did in the late 1950's [8], although this description follows a modern review [9]. The basic assumptions are: 1) the probability of scattering is independent of the scatterings which came before, and 2) the struck nucleons continue along straight line trajectories. The starting point is to take the nuclear charge density, which is well-known from all sorts of low energy scattering experiments. This is typically parameterized by a Woods-Saxon distribution, which for spherical nuclei is expressed as

$$\rho(r) = \frac{\rho_0}{1 + \exp[(r - R)/a]} \quad , \quad (4)$$

where  $\rho_0$  is the density at the core. However, we're only interested in how the density falls off as a function of the radius ( $r$ ), such that we can look at the ratio  $\rho/\rho_0$ . As shown for Pb(208) in Fig. 2 (left), this density is essentially flat until you get close to the surface, where it falls off exponentially. The parameter  $R$  corresponds to the nuclear radius, while the surface thickness  $a$  describes how rapidly the nuclear density falls off. For the example of Pb(208),  $R = 6.7$  fm and  $a$  is about 0.5 fm. The nuclear thickness of nucleus  $A$  is defined by the following integral over the nuclear density,  $T_A(\vec{s}) = \int \rho_A(\vec{s}, z) dz$ , as shown in Fig 2 (right).  $T$  is normalized such that the integral over the entire nucleus gives back the number of nucleons:  $\int T_A(\vec{s}) d^2s = A$ . The thickness function is then the product of the two individual nuclear thicknesses integrated over the impact parameter of the collision:  $T_{AB}(b) = \int T_A(\vec{s}) \cdot T_B(\vec{s} - \vec{b}) d^2s$ .



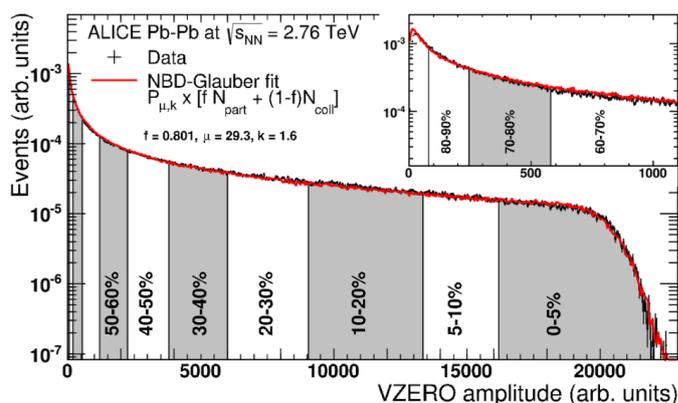
**Fig. 2:** Left: Woods-Saxon profile for Pb(208). Right: Diagram of nuclear scattering in the Glauber model [9].

There are two numbers that are key for describing heavy-ion collisions. One is the mean number of inelastic collisions for a given impact parameter. Elastic collisions don't contribute much to the mid-rapidity energy density, so they're not taken into account. The total number of binary inelastic collisions in a heavy-ion collision is simply the nuclear thickness function times the nucleon-nucleon cross section:  $N_{\text{coll}} = T_{AB}(b) \cdot \sigma_{\text{inel}}^{\text{nn}}$ . The other important number is the number of participating nucleons, sometimes called the number of wounded nucleons. This is the number of nucleons that are struck at least once. Given a nucleon in nucleus A, the probability for it to scatter with a nucleon from B is  $p_{\text{int}} = T_B(\vec{s} - \vec{b}) \cdot \sigma_{\text{inel}}^{\text{nn}}/B$ . The probability for it not to interact with any nucleon in B is then just  $(1 - p_{\text{int}})^B$ . The probability that at least one nucleon interacts is  $1 - (1 - p_{\text{int}})$ . To get the total number of participants in A, we integrate this probability over all nucleons in A, taking into account the thickness:  $N_{\text{part}}^A = \int T_A(\vec{s}) \cdot \left(1 - [1 - T_B(\vec{s} - \vec{b}) \cdot \sigma_{\text{inel}}^{\text{nn}}/B]^B\right) d^2s$ . The total number of participants is then the sum over the two nuclei:  $N_{\text{part}}(b) = N_{\text{part}}^A(b) + N_{\text{part}}^B(b)$ .

The Glauber model gives us a connection between the geometry, i.e., the impact parameter, on one hand, and which nucleons are struck on the other. Still missing is a connection to what we measure in our detectors. Centrality is defined based on some observable, usually multiplicity in some detector, that has a monotonic dependence on the Glauber quantities, most importantly on  $N_{\text{part}}$ . To make the connection, one has to model the multiplicity distribution for a  $nn$  scattering. We know from  $pp$  and  $p\bar{p}$  experiments that the multiplicity distribution is well described by a negative binomial distribution. One can randomly sample this distribution  $N_{\text{part}}$  times to get the total multiplicity. One then divides the total cross section in equal bins or percentiles. The top 5% highest multiplicity events become centrality class

0-5% for example. One then quotes Glauber quantities, typically  $\langle N_{\text{part}} \rangle$ , for a given centrality class, which has the advantage of being independent of one's detector, to good approximation.

Although centrality determination varies by experiment, it's instructive to look at an example. Figure 3 shows how ALICE determines their centrality from the multiplicity measured in a forward detector called the VZERO [10]. Using forward detectors is rather typical, as one wants to avoid correlations between the centrality determination and mid-rapidity measurements. Their Glauber calculation is shown in red, which is actually a fit to the data. They assume that particle production scales like a linear sum of a component proportional to  $N_{\text{part}}$  and one proportional to  $N_{\text{coll}}$ . The  $N_{\text{part}}$  dominates, as expected if the multiplicity is dominated by particle production. To model the multiplicity this linear combination of  $N_{\text{part}}$  and  $N_{\text{coll}}$  is convoluted with a negative binomial distribution where the mean and variance are also free parameters in the fit. The fit turns out to give an excellent description of the data, as shown in the figure, giving confidence that the extracted Glauber parameters are meaningful.



**Fig. 3:** Centrality determination from the VZERO amplitude in ALICE [10]. The red line is a fit to a linear combination of  $N_{\text{part}}$  and  $N_{\text{coll}}$ , which is smeared by a negative binomial distribution, as described in the text.

We've been discussing multiplicity as a way to characterize the initial state, but multiplicity can also be used to estimate how much energy is deposited in a heavy-ion collision. Recall that although the total collision energy is on the order of a PeV, a lot of this energy will just continue with the remnants of the nuclei down the beampipe. On the other hand, we're interested to know how much of this energy is made available for particle production. The multiplicities involved are quite large, around 1.9k particles per unit  $\eta$  at mid-rapidity in the 5% most central collisions at top LHC energy [11]. The distribution in  $\eta$  has a characteristic shape with a plateau region at mid-rapidity. This shape can be understood in a simple space-time picture, which is commonly used to estimate the energy density.

That picture is due to J.D. Bjorken [12], more famous for his work on deep inelastic scattering. He considered the nuclei to be infinitely thin pancakes moving at the speed of light. The beam remnants continue moving forward, but mid-rapidity is dominated by produced particles. He considered the produced matter to expand homogeneously in the longitudinal direction, until some formation time  $t_0$  in the particle rest frame. At that point the particles will materialize and free stream towards your detector. The velocity of a particle in this system is then simply  $z/t$ , or  $z\gamma/\tau$ . Considering an infinitely thin slice at  $z = 0$  and time  $= \tau_0$ , the energy density is then just the area of the cylinder ( $A$ ), so the size of the nucleus basically, times  $dE/dz$  (evaluated at  $z = 0$ ):

$$\epsilon = \frac{E}{V} = \frac{1}{A} \frac{dE}{dz} \Big|_{z=0} = \frac{1}{A} \frac{dy}{dz} \Big|_{z=0} \frac{dE}{dy} \Big|_{y=0} = \frac{1}{A \cdot \tau_0} \frac{dE}{dy} \Big|_{y=0} . \quad (5)$$

Here we have used the chain rule to get the energy density in terms of rapidity, as well as using a useful relation for rapidity:  $\sinh y = \beta/\gamma = z/\tau$ . This equation now relates the actual energy density in the

produced matter to a quantity we can observe in our detector. Inserting some typical numbers,  $A$  is about 7 fm for the Pb nucleus.  $\tau_0$  is usually assumed to be on the order of 1 fm based on the uncertainty principle, although we'll see from flow measurements in Section 8 that such a short formation time is justified. The energy we measure at mid-rapidity can be up to 2 GeV per unit  $\eta$  in central collisions at the LHC. That gives an energy density of around 13 GeV/fm<sup>3</sup> which is well in excess of the energy density of 1 GeV/fm<sup>3</sup> predicated for a phase transition from the lattice.

## 6 Statistical hadroproduction

In addition to looking at the overall number of particles produced, we can look at what types of particles are produced. What actually determines this? For elementary collisions, we don't rely on a first principles theory. Instead, hadronization is described by fragmentation functions which are universal, but non-perturbative. There are various phenomenological models to describe hadronization such as the string model in the Pythia generator [13] and the cluster model in the Herwig generator [14]. In heavy ions we may however choose to think in terms of an ensemble, where the language of statistical mechanics may be suitable. In fact that was what Fermi had tried to do back in the 50's before QCD [15].

We postulate that the QGP turns into an equilibrated gas of hadrons. Equilibrium is assumed to be established over a large volume, such that particle numbers are conserved only over the entire volume, not locally. In the language of statistical mechanics this is a grand canonical ensemble. So, for example, the number density of each state ( $n_i$ ) will be given by a derivative of the partition function as follows,

$$n_i = N/V = -\frac{T}{V} \frac{\partial \ln Z_i}{\partial \mu} = \frac{g_i}{2\pi^2} \int_0^\infty \frac{p^2 dp}{e^{(E_i - \mu_i)/T} \pm 1} \quad (6)$$

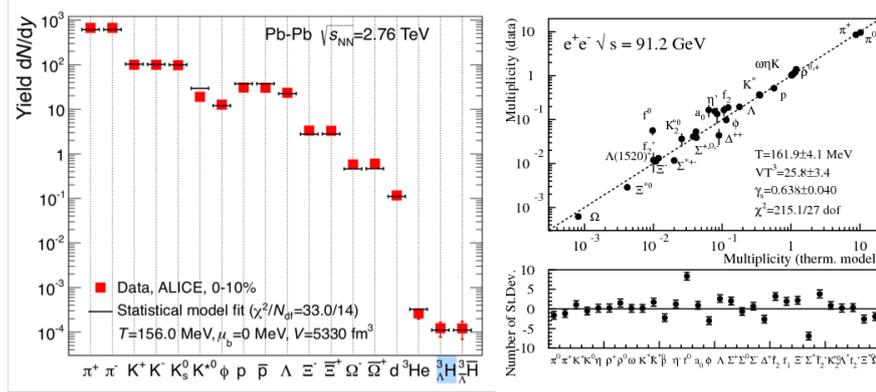
The abundance of each state depends on its mass and the temperature. Each conserved quantum number gives you a chemical potential ( $\mu$ ) such that you can conserve charge, baryon number, etc. The key point is that, if this holds, you can determine the temperature via a fit to the abundance of different states. This temperature corresponds not to the critical temperature, but to the temperature at chemical freezeout, when the identity of particles is fixed.

Figure 4 (left) shows an example of such a fit in central PbPb collisions. The ratio of data to the statistical hadronization model is plotted for a large variety of hadron species. The free parameters are the freezeout temperature, the baryon chemical potential and the volume. The model describes the data remarkably well, given the small number of free parameters, which are the following. The baryon chemical potential extracted from the fit is consistent with zero, which is what we expect at such large energy. The volume is 5000 fm<sup>3</sup>, which corresponds to a radius of about 10 fm. The extracted temperature of freezeout is 156 MeV, so below the critical temperature, but only slightly. This means the QGP hadronizes and the particle abundances are fixed almost immediately afterwards.

The fact that the fit works so well supports a picture in which we have chemical equilibrium over a large volume. This implies that the system is thermalized. What's surprising then is that such a fit actually works pretty well in elementary collisions, as well. This is shown in Fig. 4 (right), where the multiplicity of various hadron species in data is plotted against the results from a statistical model. While an excellent fit is obtained, crucially, one has to use a canonical ensemble, rather than a grand canonical one, such that quantum numbers are conserved only locally. One also needs to add an additional chemical potential for strange particles, which we'll discuss in the next section. These differences notwithstanding, there is quite a bit of debate as to why such a description works in elementary collisions, where thermalization is not expected to be reached [18].

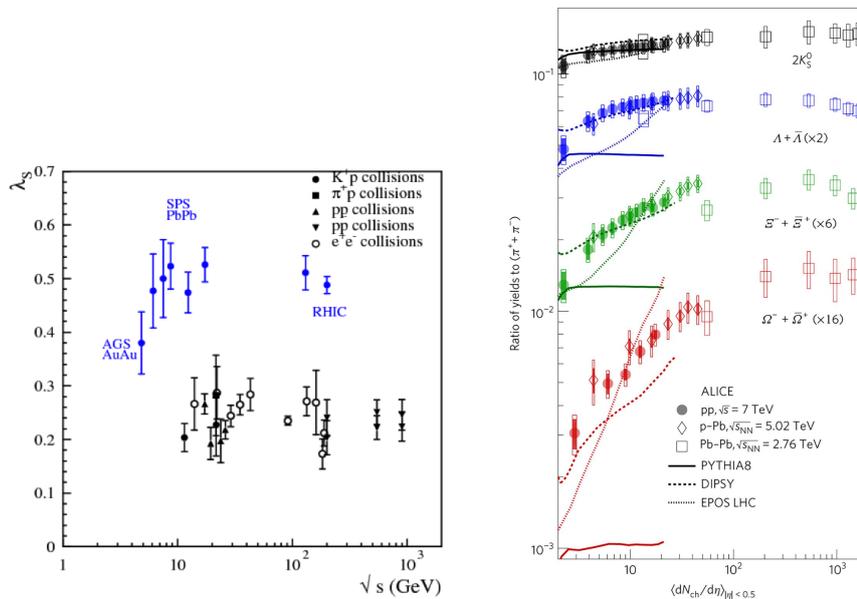
## 7 Strangeness enhancement

The strange quark is indeed rather peculiar with respect to the QGP. For the light quarks, the hadron masses are almost entirely generated dynamically by the confinement mechanism, whereas for the heavy



**Fig. 4:** Left: Data-to-model ratio for the yields of various species of hadrons in central PbPb collisions [16]. Right: Multiplicity of various hadron species in data, compared to a statistical model in e<sup>+</sup>e<sup>-</sup> collisions. [17]

ones, the mass comes from the Higgs mechanism. The strange quark kind of splits the difference, in that bare mass isn't entirely negligible. It is however smaller than the temperature of the QGP, such that we expect thermal production, according to the statistical model. However, if we only have local conservation then the created strange quark pairs have a high probability to annihilate shortly after their creation. On the other hand if we have a situation which may be described as grand canonical ensemble, where we have a large volume of equilibrated matter, the strange quarks can wander off and hadronize with other quarks. That's indeed exactly what we see in the data. If we look at the ratio of strangeness production compared to light quark production, for example by looking at the ratio of kaons to pions, we see that there's an enhancement of a factor 2, which reaches a plateau at large collision energy. This enhancement factor is even larger for particles that contain more than one strange quark, up to about a factor of 15 for the triply strange Ω baryon. This supports a picture of the QGP as an equilibrated state of matter.



**Fig. 5:** Left: The ratio of strange quarks to light quarks vs. collision energy for various collision species [19]. Right: The ratio of yields of various strange hadrons to charged pions vs. charged particle multiplicity for pp, pPb and PbPb collisions [20].

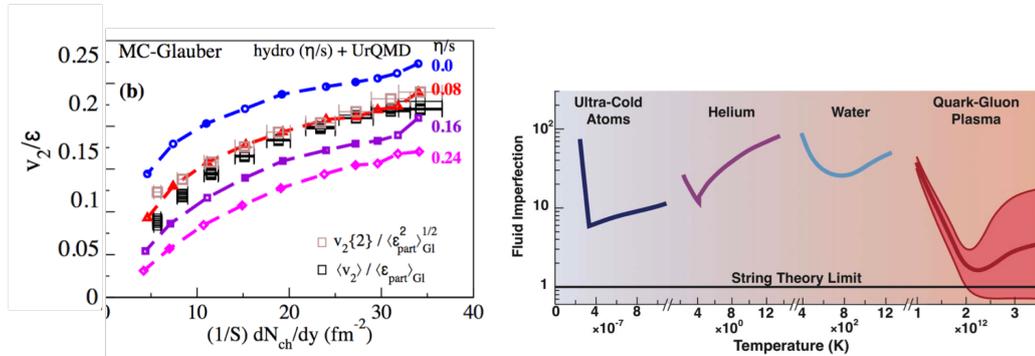
## 8 Collective flow

We now turn to the 2<sup>nd</sup> part of the lectures, which is more focused on observables that measure properties of the QGP, starting with collectivity. In non-central collisions the overlap zone of a heavy-ion collisions has an ellipsoidal shape. Given rapid thermalization, this leads to larger pressure along the short axis compared to the long one. This *spatial asymmetry* hence gives rise to a *momentum space anisotropy*. This effect is typically quantified via a Fourier decomposition of the particle yield vs. azimuth as

$$E \frac{d^3\sigma}{d\mathbf{p}^3} = \frac{1}{2\pi} \frac{d^2\sigma}{p_T dp_T dy} \left( 1 + 2 \sum_n v_n \cos[n(\phi - \psi_R)] \right), \quad (7)$$

where  $v_n$  are the flow coefficients, and  $\psi_R$  is the reaction plane angle, defined by the short axis of the ellipsoid. Given the ellipsoidal overlap shape, the 2<sup>nd</sup> harmonic  $v_2$  is expected to dominate, the so-called elliptic flow term. Naively odd terms are not expected to contribute, due to the symmetry of the overlap shape.

Collective flow is modeled using relativistic hydrodynamics. Corrections to ideal hydrodynamics are encoded in the viscous correction terms. The shear viscosity term is particularly relevant as it serves to damp the elliptic flow. Hydrodynamics calculations including viscous terms [21] are compared to the data from STAR [22] in Fig. 6 (left). The elliptic flow term (scaled by the eccentricity of the overlap zone) is plotted as a function of the multiplicity density and compared to hydro calculations for various value of the dimensionless ratio shear viscosity divided by entropy density ( $\eta/s$ ). The result comes remarkably close to the value of  $1/4\pi$ , calculated for conformal field theory, via the AdS/CFT correspondence. Although such a theory does not necessarily represent QCD, as an infinitely strongly coupled theory, it has been conjectured to give a universal lower bound for  $\eta/s$ . A comparison of the fluid imperfection index,  $\eta/s$  scaled by the value of  $1/4\pi$ , for several low viscosity materials in very different temperature regimes is shown in Fig. 6 (right). Albeit with a sizable systematic uncertainty, the QGP shows the lowest  $\eta/s$  of any known material, leading it to be dubbed the “perfect fluid”.

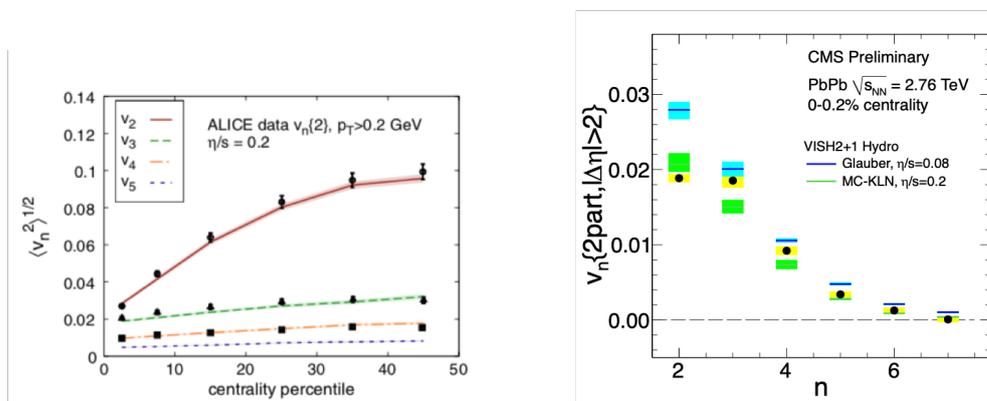


**Fig. 6:** Left: A comparison of STAR  $v_2$  data [22] to hydrodynamic calculations with different values of  $\eta/s$  [21]. Right: The fluid imperfection index  $4\pi\eta/s$  as a function of temperature for various low viscosity material in the vicinity of their respective phase transitions.

Whereas a large  $v_2$  coefficient was expected, higher orders terms were not expected to be large. In particular the odd terms were expected to be vanishing owing to the symmetry of the overlap zone. This turned out not to be the case, as shown in Fig. 7, where the various  $v_n$  coefficients are plotted vs. centrality [23]. The  $v_3$  and  $v_4$  terms were not only found to be non-zero, but have a markedly different centrality dependence than  $v_2$ , which increases as collisions become less central. These higher order terms were eventually understood to be due to event-by-event fluctuations of the overlap shape [24], which are not accounted for in the analytic Glauber approach described in Section 5.

These higher order terms turn out to be quite useful, which we can understand by analogy with the

cosmic microwave background. Variation in the temperature of the CMB reflects the density fluctuations of the early universe before the inflationary period [25]. In much the same way, the azimuthal anisotropy in long range correlations in central heavy ions collisions reflect fluctuations of the initial state of dense QCD matter before its hydrodynamic expansion. Figure 7 (right) [26] shows the  $v_n$  coefficients for a selection of the 0.2% most central PbPb events, where all azimuthal flow should be dominated by initial state fluctuations, as the initial overlap zone is symmetric. The data are compared to hydrodynamic calculations, but with two different initial states: the standard Glauber ones, and MC-KLN, a model that additionally takes into account color field fluctuation at the sub-nucleon scale [27]. While such hydrodynamic calculations provide an accurate description of flow measurements overall, there is quite some sensitivity to how the initial state is modeled, which is demonstrated by the difference between these two calculations. That in turn effects the estimation of the viscosity, and in fact our knowledge of the initial state is at present the limiting factor in how well we can extract this quantity.



**Fig. 7:** Left: Azimuthal flow harmonics of different orders vs. centrality for PbPb collisions at 2.76 TeV from ALICE [23]. Right: Flow coefficients in ultra-central PbPb collisions from CMS [26], compared to hydrodynamic calculations with two different models of initial state fluctuations, as described in the text.

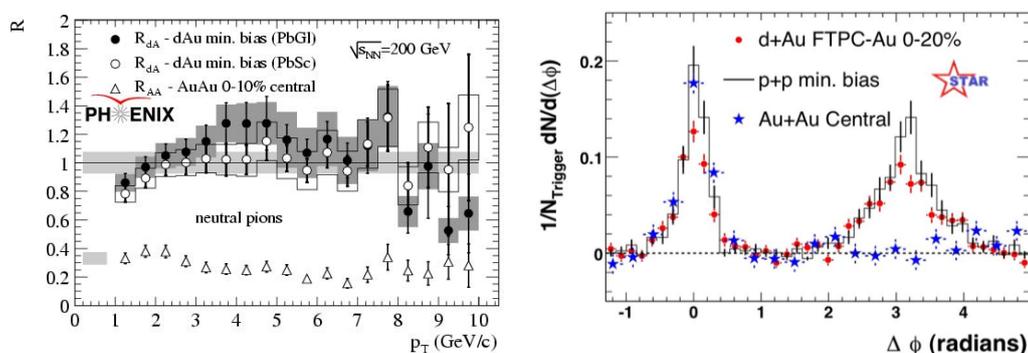
## 9 Jet quenching

The observables we have discussed so far, such as azimuthal flow and strangeness production, are related to bulk particle production. We now turn to hard probes, i.e., processes with large momentum exchange that occur before the QGP is formed. Hard scattering, whether in elementary collisions or in heavy ions, leads to production of jets, a collimated spray of particles resulting from the hadronization of an energetic parton. In heavy ions, however, one expects these jets to be “quenched”, an idea first proposed in the early 1980’s by Bjorken [28]. In much that same way as a charged particle loses energy as it traverses normal matter, a particle with color charge, i.e., a parton, should lose energy as it passes through the colored medium of the QGP. If one can describe the energy loss process theoretically, then one can in turn extract properties of the medium itself, such as the color charge density.

Jet quenching was first observed at RHIC. Although full reconstruction of jets was not yet possible in the busy heavy-ion environment, the phenomenon was clearly visible as a suppression in the yield of high  $p_T$  hadrons [29]. Nuclear effects on hard probes are quantified by the nuclear modification factor  $R_{AA}$ , which ratio of the yield observed in AA collisions compared to the expectation from pp collisions, taking into account the nuclear thickness from the Glauber model discussed in Section 5. For example, as a function of  $p_T$

$$R_{AA} \equiv \frac{dN^{AA}/dp_T}{\langle T_{AA} \rangle d\sigma/dp_T} = \frac{dN^{AA}/dp_T}{\langle N_{coll} \rangle dN/dp_T} , \quad (8)$$

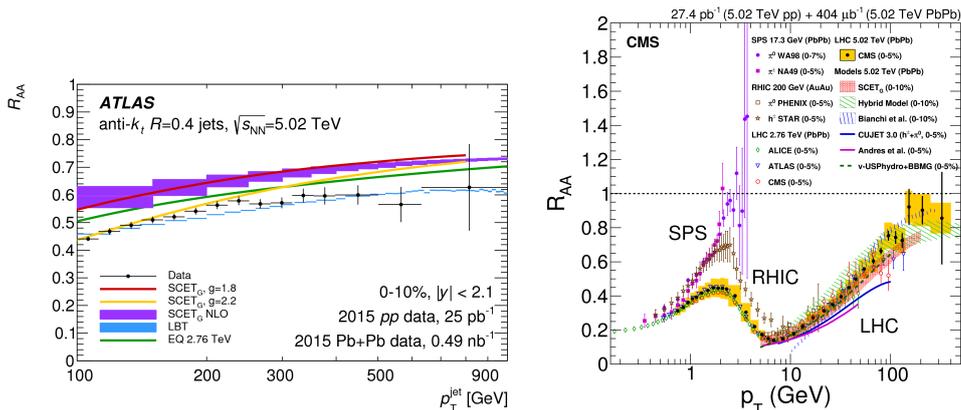
where, in the first expression, the denominator contains the inelastic cross section for pp. In the 2<sup>nd</sup> equivalent form of  $R_{AA}$  one sees explicitly that hard processes are expected to scale with  $N_{coll}$ , such that an  $R_{AA}$  of unity would indicate the absence of nuclear effects. Figure 8 (left) shows  $R_{AA}$  for the  $\pi^0$  meson in central AuAu collisions, which shows a suppression of nearly a factor of 5. This suppression is nearly independent of  $p_T$ , which, in the absence of any modification to the fragmentation pattern of jets, would indicate that partons lose a constant fraction of their  $p_T$ . No such suppression is observed in deuteron-gold (dAu) collisions, as would arise if the presence of a nucleus in the initial state was responsible for the effect. In addition to single particle yields, one can look at back-to-back correlations from high  $p_T$  particles from dijet correlations, as shown in Fig. 8 (right). A high  $p_T$  “trigger” charged hadron is selected in the range  $4 < p_T < 6$  GeV. The yield of hadron “partners” with  $p_T$  lower than the trigger particle, but still above a hard cutoff of 2 GeV, is then studied as function of azimuthal angle between the two particles. One observes a nearly complete disappearance of back-to-back correlations in central AuAu collisions, that once again is not observed in dAu. The standard interpretation of this disappearance is that the selection of a high  $p_T$  trigger particle preferentially selects jets that come from the surface of the overlap zone, and hence suffer little quenching. As a result, the recoiling jet traverses a large medium path-length, on average, and hence suffers a large quenching.



**Fig. 8:** First observations of jet quenching at RHIC. Left: Nuclear modification factor of  $\pi^0$  in AuAu and dAu collisions at 200 GeV from PHENIX [30]. Right: Disappearance of back-to-back high  $p_T$  dihadron azimuthal correlations in AuAu collisions compared to pp and dAu collisions with STAR [31].

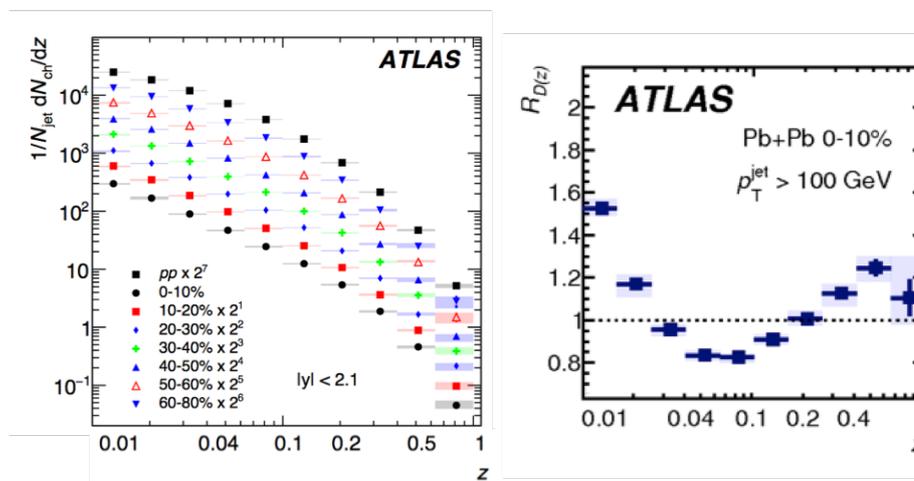
Due to the larger cross section for hard processes, as well as the availability of hermetic calorimeters, full reconstruction of jets in heavy ions became possible for the first time at the LHC. Figure 9 (left) shows jet  $R_{AA}$  measured over a very wide range of  $p_T$  out to nearly 1 TeV. A similar approximate flatness of  $R_{AA}$  is seen, as was observed for hadrons at RHIC. However, if one looks at the charged hadrons [32], as shown in Fig. 9 (right), one sees that with the much large range of  $p_T$  that can be measured at the LHC, that the trend is not nearly as flat as was observed at RHIC. In fact, the  $R_{AA}$  of hadrons rises rather steeply nearly reaching unity above 100 GeV, albeit with non-negligible uncertainties. How can the measurements of jets and charged hadrons be reconciled? For them to give a different qualitative behavior requires that there be interplay between quenching and the fragmentation pattern of jets.

An advantage of full jet reconstruction is that their fragmentation pattern can then be measured rather directly. Figure 10 (left) shows the jet fragmentation function, defined as the distribution of the ratio ( $z$ ) of charged hadron  $p_T$  to the associated jet  $p_T$ , for pp and various centrality selections of PbPb collisions [34]. One can study the modification of the fragmentation pattern by forming  $R_{D(z)}$ , the ratio of the PbPb and pp measurements, as shown in Fig. 10 (right). The fragmentation pattern shows a characteristic modification. The enhancement at low  $z$  is generally understood to be the response of the medium to the deposited energy. The depletion and rise at intermediate and high  $z$  are still debated. One possibility is that gluons are preferentially suppressed compared to quarks, due to their larger color factor. Because of the softer fragmentation function of gluons, the resulting sample of jets, which initially



**Fig. 9:** Left: Jet  $R_{AA}$  vs.  $p_T$  in central PbPb collisions from ATLAS [33]. Right: Charged hadron  $R_{AA}$  vs.  $p_T$  in central PbPb collisions from CMS [32].

contains a roughly even mix of quarks and gluons, is biased towards quark jets. Another possibility is that there is a direct dependence of quenching of the fragmentation pattern, via the parton shower. Partons that shower into a larger number of secondary partons may effectively interact with the medium more, compared to lower multiplicity showers. Understanding these details of parton-medium interactions is at the forefront of current investigations, which involve, for example, measurements of jet substructure [35].



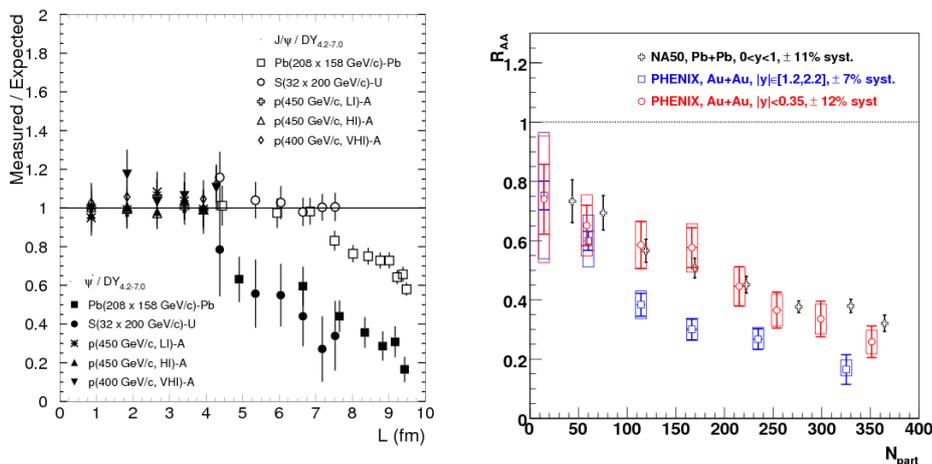
**Fig. 10:** Left: Jet fragmentation functions measured in pp and in different centrality selections of PbPb collisions [34]. Right: Ratio of jet fragmentation function in central PbPb to pp collisions [34].

## 10 Quarkonium dissociation

In addition to jets, hard processes can also produce massive states such as quarkonia, bound states of a quark and its anti-quark. The most abundant of these is the  $J/\psi$  meson, a bound state of  $c$  and  $\bar{c}$ . Since the charm mass is above the threshold for thermal production in the QGP, one expects it to be produced only via hard scattering. Just as electric charges are screened by a (QED) plasma, Matsui and Satz predicted that the attractive QCD potential between the quarks could be screened in the QGP [36]. Moreover, different quarkonium states have quite different binding energies, which implies that they should dissociate at different temperatures. For this reason, quarkonia were proposed as a thermometer

of the QGP.

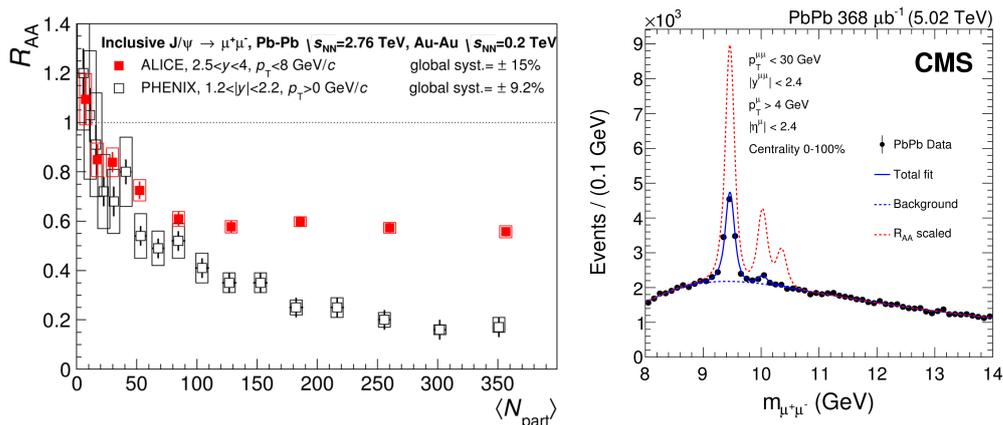
In line with this expectation, the  $J/\psi$  meson was found to be suppressed in fixed target PbPb collisions at the CERN SPS [37]. As shown in Fig. 11 (left), the data are plotted vs. the pathlength  $L$ , another Glauber derived quantity which scales with collision centrality. As a baseline, the  $J/\psi$  yield is normalized by that of Drell-Yan, which is not expected to be modified by nuclear effects. The  $J/\psi$  shows a rather abrupt suppression in central PbPb events, which is not seen in more peripheral events, nor in lighter collision systems. This was interpreted as one of the earliest observed signatures of QGP formation. The more loosely bound excited state,  $\psi(2S)$ , was also measured. It shows a suppression starting from even less central events. Naively, one would interpret this as the  $\psi(2S)$  dissociating at a lower temperature, a phenomenon often referred to as “sequential melting”. As it turns out the more modern interpretation of  $\psi(2S)$  data is more nuanced, as it is so fragile that other final state effects likely come into play.



**Fig. 11:** Left: Suppression of the  $J/\psi$  and  $\psi(2S)$  in PbPb collisions, and the corresponding lack of suppression in lighter collision systems, by the NA50 experiment [37]. The data are plotted vs. the nuclear pathlength and normalized to the expectation from Drell-Yan events. Right: Nuclear suppression factor for  $J/\psi$  vs. the number of participants for a central and a more forward rapidity selection, as measured by PHENIX [38].

Initial results from RHIC showed a similar suppression factor for the  $J/\psi$ , despite the larger collision energy, as seen in the mid-rapidity results in Fig. 11 (right), this time evaluated in terms of  $R_{AA}$  vs.  $N_{part}$ . These results would be most naturally understood as a saturation of the dissociation effect, such that an increase in temperatures would not lead to further suppression. A measurement at forward rapidity showed a larger suppression factor, however, that was not anticipated (also shown in the same figure). Interestingly, when even more forward rapidity measurements were performed at an order of magnitude larger collision energy (2.76 TeV, as compared to 200 GeV), a much reduced suppression was observed, as shown in Fig. 12 (left). This effect is now understood to correspond to an additional mechanism of  $J/\psi$  production referred to as statistical regeneration [39]. Although charm quarks are relatively rare at the SPS, at RHIC energies they are abundant enough that charm quarks can combine with anti-charm quarks produced in a different nucleon-nucleon scattering. This effect is more pronounced at mid-rapidity, where the charm quark density is larger, explaining the relatively smaller suppression factor, compared to forward rapidity, observed at RHIC. The charm cross section also grows rather rapidly with collision energy, explaining the further reduced suppression observed at the LHC.

While the statistical regeneration picture provides a satisfactory explanation to the  $J/\psi$  puzzle, it certainly complicates a simple interpretation in terms of the afore-mentioned sequential melting picture. In addition to the charmonia states, the large collision energies provided by the LHC enabled the study of the bottomonia states with reasonably high statistics, for the first time. The mass and radii of the



**Fig. 12:** Left:  $R_{AA}$  of  $J/\psi$  at vs.  $N_{part}$  in 2.76 TeV PbPb collisions at the LHC, compared to 200 GeV collisions at RHIC [40]. Right: CMS measurement of the invariant mass spectrum of dimuon pairs showing the peaks of the three lowest lying  $\Upsilon$  states in PbPb, compared to the expectation from pp collisions [41].

state	$J/\psi$	$\psi(2S)$	$\Upsilon(1S)$	$\Upsilon(2S)$	$\Upsilon(3S)$
mass (GeV)	3.10	3.68	9.46	10.0	10.4
radius (fm)	0.25	0.45	0.14	0.28	0.39

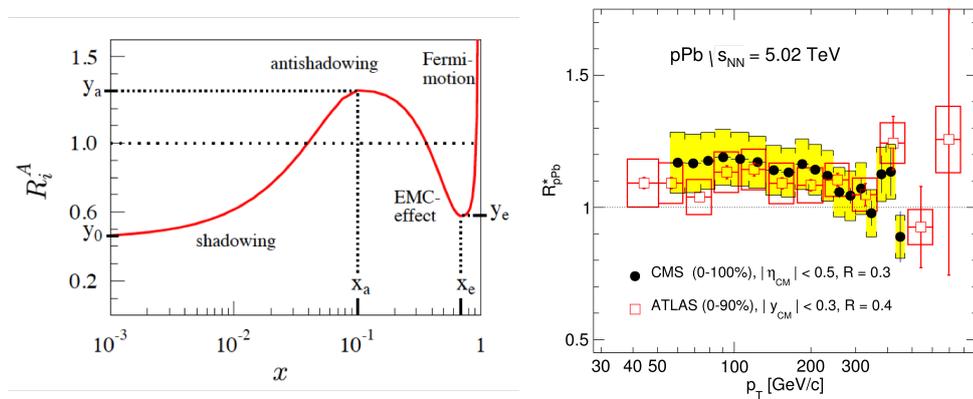
five quarkonia states that are readily measured in heavy ions are listed in Table 10. The  $\Upsilon$  family has three states that are readily measurable, compared to two for charmonia. The ground state  $\Upsilon(1S)$  is the most tightly bound of the 5 states, while the  $\Upsilon(2S)$  and  $\Upsilon(3S)$  states are roughly comparable to the  $J/\psi$  and  $\psi(2S)$ , respectively. In contrast to charmonia though, bottom quark production is sufficiently rare that statistical regeneration effects are expected to be small. Figure 12 (right) shows the dimuon invariant mass spectrum in PbPb collisions in the vicinity of the  $\Upsilon$  resonances [41]. The PbPb data are compared to a line-shape based on pp collisions, i.e., in the absence of nuclear effects. A strong suppression is observed for all 3 states, with  $R$  values smaller than 0.15 for both excited states (the 3S state is only an upper limit). The 1S state is less suppressed, with an  $R_{AA}$  value of around 0.45, as expected from the sequential melting picture. Moreover a sizable fraction of  $\Upsilon(1S)$  come from feed-down of excited states (both S-wave and P-wave), which may account for the majority of its suppression. The  $\Upsilon$  measurements therefore support the scenario of sequential melting of the onium states in the QGP. A quantitative extraction of the medium temperature requires a thorough evaluation of so called cold nuclear matter effects, which will be introduced in the next and final section.

## 11 Small systems

Thus far I have mainly discussed results in ion-ion collisions, using pp collisions as a baseline from which to benchmark nuclear effects. In Section 9 I mentioned that dAu collisions were studied, to ensure that jet quenching did not arise from the presence of a heavy nucleus in the initial state (see Fig. 8). Similarly, in Section 10, we have seen that the anomalous suppression of the charmonium states is not observed in lighter collision systems, such as pA with various types of heavy nuclei. This is not to say, however, that nuclear effects are entirely absent from these lighter systems. In this section, I revisit some of the observables that have been discussed, taking a deeper look at their behavior in pA, as well as in pp collisions.

One class of initial state nuclear effects that has been known since the 1970s is modification of the parton distributions in nuclei. The pattern of these modifications, illustrated in Fig. 13 (left), is quantified by looking at the ratio of nuclear PDF to the free nucleon one, as function of Bjorken  $x$ .

At low  $x$  one observes a depletion of partons, known as shadowing, which essentially corresponds to partons obscuring one another in the nucleus. At intermediate values of  $x$ , one sees a compensating anti-shadowing effect, where the PDF is enhanced in the nucleus. Finally, at large values of  $x$ , one again sees a depletion, referred to as the EMC effect, whose physical explanation remains debated to this day. As for free nucleons, the nuclear PDFs are universal. One may then ask whether data are consistent with various global fits of nuclear PDFs, or whether there are other types of nuclear effects potentially at play. Figure 13 (right) shows the nuclear modification factor in pPb ( $R_{pPb}$ ) for jets, as a function of  $p_T$  from ATLAS [42] and CMS [43]. The results are compatible with a light enhancement from anti-shadowing, which is dominant in the region of Bjorken  $x$  probed by this measurement. By varying the kinematics of dijets, one is able to span a larger range of Bjorken  $x$ , including the shadowing, anti-shadowing and EMC regions. The results are also compatible with global fits to the nuclear parton distributions, such as those in Ref. [44].

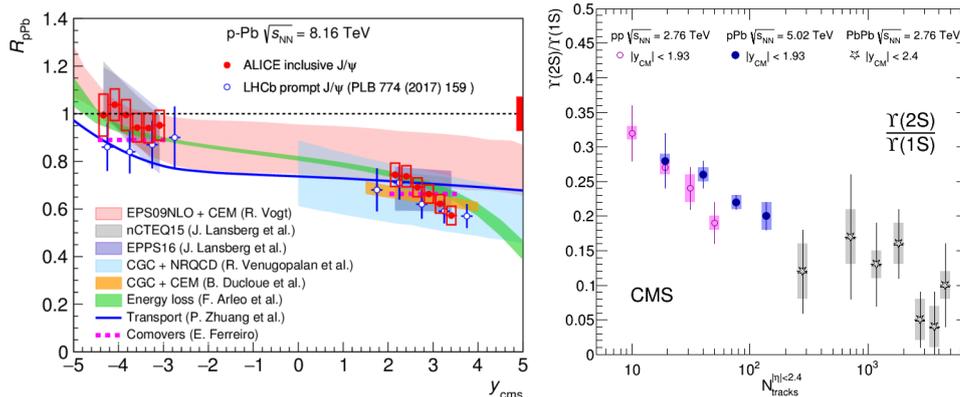


**Fig. 13:** Left: A schematic of nuclear modifications to the parton distributions, as a function of Bjorken  $x$  [45]. Right: Nuclear modification factor of jets in pPb collisions, as a function of  $p_T$  [43].

In the charmonia sector, one is also able to reproduce the nuclear modification factor based on parameterization of nuclear PDF effects. In Fig. 14, the  $R_{pPb}$  of  $J/\psi$  is plotted as function of rapidity [46]. While nuclear effects are modest at backward rapidity, the Pb-going direction, they are rather sizable at forward rapidity, which is the proton-side going direction. These effects are consistent with the level of shadowing found in commonly used global fits (the first three theory calculations on the plots) [44,45,47]. Although the suppression due to shadowing is smaller than that observed in PbPb collisions, clearly it needs to be taken into account in the interpretation of quarkonium dissociation measurements.

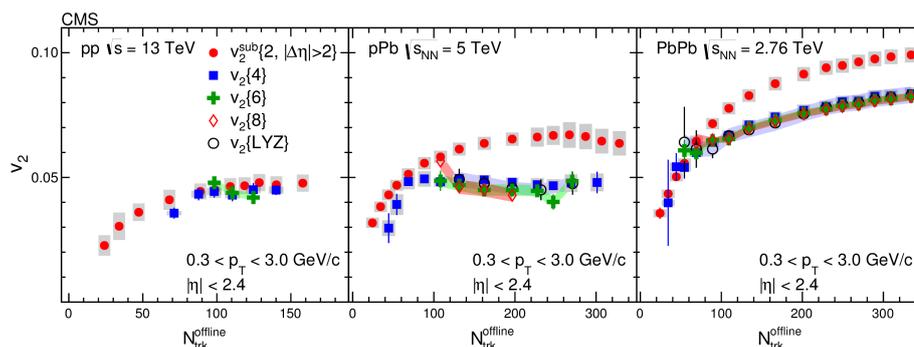
Whereas both the jet and  $J/\psi$  data can be described by nuclear PDF parameterizations, there are effects that cannot be easily described only by these effects. Fig 14 (right) shows the ratio of the yield of the excited state  $\Upsilon(2S)$  to the ground state  $\Upsilon(1S)$ , as a function of event multiplicity, for 3 collision systems: pp, pPb and PbPb [49]. One observes that this ratio varies with multiplicity not only in PbPb collisions, where such an effect is expected based on the corresponding variation of collision centrality, but also in pPb, and even in pp. Although the statistical precision of the data is not sufficient to tell whether the data fall along the same trend, clearly there are multiplicity dependent effects present in all three systems. The interpretation of this data remains a subject of debate.

Such multiplicity dependent effects in small systems have also been observed in the flow sector. The elliptic flow can be measured there, by using a technique known as multi-particle cumulants. Figure 15 shows  $v_2$  as a function of event multiplicity in pp, pPb and PbPb. Aside from two-particle correlation results, which are sensitive to a “non-flow” signal from jet correlations, the results for correlations amongst different numbers of particles are very close to one another. This supports collectivity as the origin of the effect. In PbPb collisions, one observes an increasing  $v_2$  with multiplicity, which



**Fig. 14:** Left: Nuclear modification factor  $R_{pPb}$  of  $J/\psi$ , as a function of rapidity [46], compared to calculations using various parameterizations of nuclear PDF effects, as well as other models. Right: Ratio of the yield of the excited state  $\Upsilon(2S)$  to the ground state  $\Upsilon(1S)$ , as a function of event multiplicity, for 3 collision systems: pp, pPb and PbPb [48].

can be traced back to the centrality dependence of the collisions, as discussed in Section 8. Surprisingly, however, one also observes a non-zero flow appearing in both pp and pPb, which saturates above a multiplicity of about 100 charged particles, and which is of comparable magnitude in the two systems. Just as the higher order flow coefficients (see Fig. 7) reflect event-by-event fluctuations of the overlap zone in PbPb collisions, the non-zero  $v_2$  observed in smaller systems is thought to share a similar origin. The collectivity is thought to arise from fluctuations of color fields on the sub-nucleonic scale. A similar effect is observed in the ratios of strange particle to light hadron yields in Fig. 5 (right). Together these results tend to support the surprising conclusion that there is collectivity and chemical equilibrium over an extended domain, even in light systems. These effects are not captured by standard event generators of proton collisions, and thus could have an impact on measurements thought to be far outside the domain of heavy ion physics, where such ideas have arisen. Detailed studies of the origin of such effects in small systems represents one of the most interesting and active current frontiers in the field.



**Fig. 15:** The elliptic flow coefficient  $v_2$  measured by various methods, as function of event multiplicity in pp (left), pPb (middle) and PbPb (right) collisions [49].

## References

- [1] J. C. Collins and M. J. Perry, “Superdense Matter: Neutrons Or Asymptotically Free Quarks?,” *Phys. Rev. Lett.* **34** (1975) 1353.
- [2] N. Cabibbo and G. Parisi, “Exponential Hadronic Spectrum and Quark Liberation,” *Phys. Lett.* **59B** (1975) 67–69.

- [3] Y. Aoki *et al.*, “The QCD transition temperature: results with physical masses in the continuum limit II.,” *JHEP* **06** (2009) 088, [arXiv:0903.4155 \[hep-lat\]](#).
- [4] A. Bazavov *et al.*, “The chiral and deconfinement aspects of the QCD transition,” *Phys. Rev.* **D85** (2012) 054503, [arXiv:1111.1710 \[hep-lat\]](#).
- [5] R. Hagedorn, “Statistical thermodynamics of strong interactions at high energies,” *Nuovo Cimento, Suppl.* **3** no. CERN-TH-520, (1965) 147–186. <http://cds.cern.ch/record/346206>.
- [6] P. Braun-Munzinger and J. Stachel, “The quest for the quark-gluon plasma,” *Nature* **448** (2007) 302–309.
- [7] R. C. Hwa and X. N. Wang, eds., *Quark-gluon plasma. Vol. 3.* 2004.
- [8] R. J. Glauber, “Cross-sections in deuterium at high-energies,” *Phys. Rev.* **100** (1955) 242–248.
- [9] M. L. Miller, K. Reygers, S. J. Sanders, and P. Steinberg, “Glauber modeling in high energy nuclear collisions,” *Ann. Rev. Nucl. Part. Sci.* **57** (2007) 205–243, [arXiv:nucl-ex/0701025 \[nucl-ex\]](#).
- [10] ALICE Collaboration, B. Abelev *et al.*, “Centrality determination of Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV with ALICE,” *Phys. Rev.* **C88** (2013) 044909, [arXiv:1301.4361 \[nucl-ex\]](#).
- [11] ALICE Collaboration, J. Adam *et al.*, “Centrality dependence of the charged-particle multiplicity density at midrapidity in Pb-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV,” *Phys. Rev. Lett.* **116** (2016) 222302, [arXiv:1512.06104 \[nucl-ex\]](#).
- [12] J. D. Bjorken, “Highly Relativistic Nucleus-Nucleus Collisions: The Central Rapidity Region,” *Phys. Rev.* **D27** (1983) 140–151.
- [13] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, “Parton Fragmentation and String Dynamics,” *Phys. Rept.* **97** (1983) 31–145.
- [14] G. Marchesini and B. R. Webber, “Simulation of QCD Jets Including Soft Gluon Interference,” *Nucl. Phys.* **B238** (1984) 1–29.
- [15] E. Fermi, “High-energy nuclear events,” *Prog. Theor. Phys.* **5** (1950) 570–583.
- [16] A. Andronic, P. Braun-Munzinger, K. Redlich, and J. Stachel, “Decoding the phase structure of QCD via particle production at high energy,” *Nature* **561** (2018) 321–330, [arXiv:1710.09425 \[nucl-th\]](#).
- [17] F. Becattini, P. Castorina, J. Manninen, and H. Satz, “The Thermal Production of Strange and Non-Strange Hadrons in e+ e- Collisions,” *Eur. Phys. J.* **C56** (2008) 493–510, [arXiv:0805.0964 \[hep-ph\]](#).
- [18] R. Stock, “Hadronization revisited: The Dynamics behind hadro-chemical equilibrium,” *PoS CPOD2006* (2006) 040, [arXiv:nucl-th/0703050 \[nucl-th\]](#).
- [19] F. Becattini and J. Manninen, “Strangeness production from SPS to LHC,” *J. Phys.* **G35** (2008) 104013, [arXiv:0805.0098 \[nucl-th\]](#).
- [20] ALICE Collaboration, J. Adam *et al.*, “Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions,” *Nature Phys.* **13** (2017) 535–539, [arXiv:1606.07424 \[nucl-ex\]](#).
- [21] H. Song, S. A. Bass, U. Heinz, T. Hirano, and C. Shen, “200 A GeV Au+Au collisions serve a nearly perfect quark-gluon liquid,” *Phys. Rev. Lett.* **106** (2011) 192301, [arXiv:1011.2783 \[nucl-th\]](#). [Erratum: *Phys. Rev. Lett.* 109,139904(2012)].
- [22] STAR Collaboration, J. Adams *et al.*, “Azimuthal anisotropy in Au+Au collisions at  $\sqrt{s_{NN}} = 200$ -GeV,” *Phys. Rev.* **C72** (2005) 014904, [arXiv:nucl-ex/0409033 \[nucl-ex\]](#).
- [23] ALICE Collaboration, J. Adam *et al.*, “Pseudorapidity dependence of the anisotropic flow of charged particles in Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV,” *Phys. Lett.* **B762** (2016) 376–388, [arXiv:1605.02035 \[nucl-ex\]](#).

- [24] B. Alver and G. Roland, “Collision geometry fluctuations and triangular flow in heavy-ion collisions,” *Phys. Rev.* **C81** (2010) 054905, [arXiv:1003.0194 \[nucl-th\]](#). [Erratum: *Phys. Rev.* **C82**, 039903(2010)].
- [25] **WMAP** Collaboration, G. Hinshaw *et al.*, “Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: temperature analysis,” *Astrophys. J. Suppl.* **170** (2007) 288, [arXiv:astro-ph/0603451 \[astro-ph\]](#).
- [26] **CMS** Collaboration, S. Chatrchyan *et al.*, “Studies of azimuthal dihadron correlations in ultra-central PbPb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV,” *JHEP* **02** (2014) 088, [arXiv:1312.1845 \[nucl-ex\]](#).
- [27] T. Hirano, U. W. Heinz, D. Kharzeev, R. Lacey, and Y. Nara, “Hadronic dissipative effects on elliptic flow in ultrarelativistic heavy-ion collisions,” *Phys. Lett.* **B636** (2006) 299–304, [arXiv:nucl-th/0511046 \[nucl-th\]](#).
- [28] J. D. Bjorken, “Energy Loss of Energetic Partons in Quark - Gluon Plasma: Possible Extinction of High  $p(t)$  Jets in Hadron - Hadron Collisions,” 1982. FERMILAB-PUB-82-059-THY, FERMILAB-PUB-82-059-T.
- [29] **PHENIX** Collaboration, “Suppression of hadrons with large transverse momentum in central Au+Au collisions at  $\sqrt{s_{NN}} = 130$ -GeV,” *Phys. Rev. Lett.* **88** (2002) 022301, [arXiv:nucl-ex/0109003 \[nucl-ex\]](#).
- [30] **PHENIX** Collaboration, S. S. Adler *et al.*, “Absence of suppression in particle production at large transverse momentum in  $\sqrt{s_{NN}} = 200$ -GeV d + Au collisions,” *Phys. Rev. Lett.* **91** (2003) 072303, [arXiv:nucl-ex/0306021 \[nucl-ex\]](#).
- [31] **STAR** Collaboration, “Disappearance of back-to-back high  $p_T$  hadron correlations in central Au+Au collisions at  $\sqrt{s_{NN}} = 200$ -GeV,” *Phys. Rev. Lett.* **90** (2003) 082302, [arXiv:nucl-ex/0210033 \[nucl-ex\]](#).
- [32] **CMS** Collaboration, V. Khachatryan *et al.*, “Charged-particle nuclear modification factors in PbPb and pPb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV,” *JHEP* **04** (2017) 039, [arXiv:1611.01664 \[nucl-ex\]](#).
- [33] **ATLAS** Collaboration, M. Aaboud *et al.*, “Measurement of the nuclear modification factor for inclusive jets in Pb+Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV with the ATLAS detector,” [arXiv:1805.05635 \[nucl-ex\]](#).
- [34] **ATLAS** Collaboration, M. Aaboud *et al.*, “Measurement of jet fragmentation in Pb+Pb and  $pp$  collisions at  $\sqrt{s_{NN}} = 2.76$  TeV with the ATLAS detector at the LHC,” *Eur. Phys. J.* **C77** (2017) 379, [arXiv:1702.00674 \[hep-ex\]](#).
- [35] **CMS** Collaboration, A. M. Sirunyan *et al.*, “Measurement of the Splitting Function in  $pp$  and Pb-Pb Collisions at  $\sqrt{s_{NN}} = 5.02$  TeV,” *Phys. Rev. Lett.* **120** (2018) 142302, [arXiv:1708.09429 \[nucl-ex\]](#).
- [36] T. Matsui and H. Satz, “ $J/\psi$  Suppression by Quark-Gluon Plasma Formation,” *Phys. Lett.* **B178** (1986) 416–422.
- [37] **NA50** Collaboration, B. Alessandro *et al.*, “psi-prime production in Pb-Pb collisions at 158-GeV/nucleon,” *Eur. Phys. J.* **C49** (2007) 559–567, [arXiv:nucl-ex/0612013 \[nucl-ex\]](#).
- [38] **PHENIX** Collaboration, A. Adare *et al.*, “ $J/\psi$  Production vs Centrality, Transverse Momentum, and Rapidity in Au+Au Collisions at  $\sqrt{s_{NN}} = 200$  GeV,” *Phys. Rev. Lett.* **98** (2007) 232301, [arXiv:nucl-ex/0611020 \[nucl-ex\]](#).
- [39] R. L. Thews, M. Schroedter, and J. Rafelski, “Enhanced  $J/\psi$  production in deconfined quark matter,” *Phys. Rev.* **C63** (2001) 054905, [arXiv:hep-ph/0007323 \[hep-ph\]](#).
- [40] **ALICE** Collaboration, J. Adam *et al.*, “Differential studies of inclusive  $J/\psi$  and  $\psi(2S)$  production at forward rapidity in Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV,” *JHEP* **05** (2016) 179, [arXiv:1506.08804 \[nucl-ex\]](#).

- [41] **CMS** Collaboration, A. M. Sirunyan *et al.*, “Measurement of nuclear modification factors of  $\Upsilon(1S)$ ,  $\Upsilon(2S)$ , and  $\Upsilon(3S)$  mesons in PbPb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV,” [arXiv:1805.09215 \[hep-ex\]](#).
- [42] **ATLAS** Collaboration, G. Aad *et al.*, “Centrality and rapidity dependence of inclusive jet production in  $\sqrt{s_{NN}} = 5.02$  TeV proton-lead collisions with the ATLAS detector,” *Phys. Lett. B* **748** (2015) 392–413, [arXiv:1412.4092 \[hep-ex\]](#).
- [43] **CMS** Collaboration, V. Khachatryan *et al.*, “Measurement of inclusive jet production and nuclear modifications in pPb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV,” *Eur. Phys. J. C* **76** (2016) 372, [arXiv:1601.02001 \[nucl-ex\]](#).
- [44] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, “EPPS16: Nuclear parton distributions with LHC data,” *Eur. Phys. J. C* **77** (2017) 163, [arXiv:1612.05741 \[hep-ph\]](#).
- [45] K. J. Eskola, H. Paukkunen, and C. A. Salgado, “EPS09: A New Generation of NLO and LO Nuclear Parton Distribution Functions,” *JHEP* **04** (2009) 065, [arXiv:0902.4154 \[hep-ph\]](#).
- [46] **ALICE** Collaboration, S. Acharya *et al.*, “Inclusive  $J/\psi$  production at forward and backward rapidity in p-Pb collisions at  $\sqrt{s_{NN}} = 8.16$  TeV,” *JHEP* **07** (2018) 160, [arXiv:1805.04381 \[nucl-ex\]](#).
- [47] K. Kovarik *et al.*, “nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTQ framework,” *Phys. Rev. D* **93** (2016) 085037, [arXiv:1509.00792 \[hep-ph\]](#).
- [48] **CMS** Collaboration, S. Chatrchyan *et al.*, “Event activity dependence of  $Y(nS)$  production in  $\sqrt{s_{NN}}=5.02$  TeV pPb and  $\sqrt{s}=2.76$  TeV pp collisions,” *JHEP* **04** (2014) 103, [arXiv:1312.6300 \[nucl-ex\]](#).
- [49] **CMS** Collaboration, V. Khachatryan *et al.*, “Evidence for collectivity in pp collisions at the LHC,” *Phys. Lett. B* **765** (2017) 193–220, [arXiv:1606.06198 \[nucl-ex\]](#).



## Practical statistics for particle physics

R. J. Barlow

The University of Huddersfield, Huddersfield, United Kingdom

### Abstract

This is the write-up of a set of lectures given at the Asia Europe Pacific School of High Energy Physics in Quy Nhon, Vietnam in September 2018, to an audience of PhD students in all branches of particle physics. They cover the different meanings of ‘probability’, particularly frequentist and Bayesian, the binomial, the Poisson and the Gaussian distributions, hypothesis testing, estimation, errors (including asymmetric and systematic errors) and goodness of fit. Several different methods used in setting upper limits are explained, followed by a discussion on why 5 sigma are conventionally required for a ‘discovery’.

### Keywords

Lectures; statistics; particle physics, probability, estimation, confidence limits.

## 1 Introduction

To interpret the results of your particle physics experiment and see what it implies for the relevant theoretical model and parameters, you need to use statistical techniques. These are a part of your experimental toolkit, and to extract the maximum information from your data you need to use the correct and most powerful statistical tools.

Particle physics (like, probably, any field of science) has its own special set of statistical processes and language. Our use is in some ways more complicated (we often fit multi-parameter functions, not just straight lines) and in some ways more simple (we do not have to worry about ethics, or law suits). So the generic textbooks and courses you will meet on ‘Statistics’ are not really appropriate. That’s why HEP schools like this one include lectures on statistics as well as the fundamental real physics, like field theory and physics beyond the Standard Model (BSM).

There are several textbooks [1–6] available which are designed for an audience of particle physicists. You will find these helpful—more helpful than general statistical textbooks. You should find one whose language suits you and keep a copy on your bookshelf—preferably purchased—but at least on long term library loan. You will also find useful conference proceedings [7–9], journal papers (particularly in Nuclear Instruments and Methods) and web material: often your own experiment will have a set of pages devoted to the topic.

## 2 Probability

We begin by looking at the concept of probability. Although this is familiar (we use it all the time, both inside and outside the laboratory), its use is not as obvious as you would think.

### 2.1 What is probability?

A typical exam for Statistics101 (or equivalent) might well contain the question:

---

Q1 Explain what is meant by the *probability*  $P_A$  of an event  $A$  [1]

---

The ‘1’ in square brackets signifies that the answer carries one mark. That’s an indication that just a sentence or two are required, not a long essay.

Asking a group of physicists this question produces answers falling into four different categories

1.  $P_A$  is number obeying certain mathematical rules,
2.  $P_A$  is a property of  $A$  that determines how often  $A$  happens,
3. For  $N$  trials in which  $A$  occurs  $N_A$  times,  $P_A$  is the limit of  $N_A/N$  for large  $N$ , and
4.  $P_A$  is my belief that  $A$  will happen, measurable by seeing what odds I will accept in a bet.

Although all these are generally present, number 3 is the most common, perhaps because it is often explicitly taught as the definition. All are, in some way, correct! We consider each in turn.

## 2.2 Mathematical probability

The Kolmogorov axioms are: For all  $A \subset S$

$$\begin{aligned} P_A &\geq 0 \\ P_S &= 1 \quad . \end{aligned} \tag{1}$$

$$P_{A \cup B} = P_A + P_B \text{ if } A \cap B = \phi \text{ and } A, B \subset S$$

From these simple axioms a complete and complicated structure of theorems can be erected. This is what pure mathematicians do. For example, the 2nd and 3rd axiom show that the probability of not- $A$   $P_{\bar{A}}$ , is  $1 - P_A$ , and then the 1st axiom shows that  $P_A \leq 1$ : probabilities must be less than 1.

But the axioms and the ensuing theorems says nothing about what  $P_A$  actually means. Kolmogorov had frequentist probability in mind, but these axioms apply to any definition: he explicitly avoids tying  $P_A$  down in this way. So although this apparatus enables us to compute numbers, it does not tell us what we can use them for.

## 2.3 Real probability

Also known as Classical probability, this was developed during the 18th–19th centuries by Pascal, Laplace and others to serve the gambling industry.

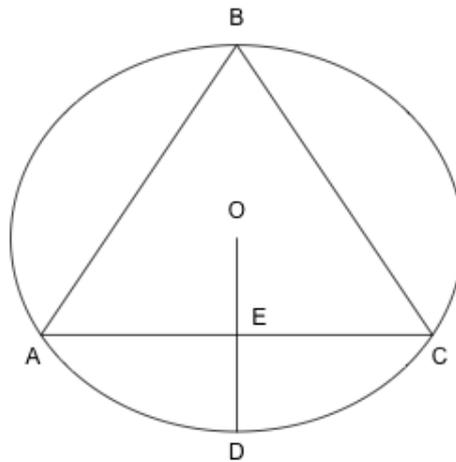
If there are several possible outcomes and there is a symmetry between them so they are all, in a sense, identical, then their individual probabilities must be equal. For example, there are two sides to a coin, so if you toss it there must be a probability  $\frac{1}{2}$  for each face to land uppermost. Likewise there are 52 cards in a pack, so the probability of a particular card being chosen is  $\frac{1}{52}$ . In the same way there are 6 sides to a dice, and 33 slots in a roulette wheel.

This enables you to answer questions like ‘What is the probability of rolling more than 10 with 2 dice?’. There are 3 such combinations (5-6, 6-5 and 6-6) out of the  $6 \times 6 = 36$  total possibilities, so the probability is  $\frac{3}{36} = \frac{1}{12}$ . Compound instances of  $A$  are broken down into smaller instances to which the symmetry argument can be applied. This is satisfactory and clearly applicable—you know that if someone offers you a 10 to 1 bet on this dice throw, you should refuse; in the long run knowledge of the correct probabilities will pay off.

The problem arises that this approach cannot be applied to continuous variables. This is brought out in Bertan’s paradoxes, one of which runs:

*In a circle of radius  $R$  an equilateral triangle is drawn. A chord is drawn at random. What is the probability that the length of the chord is greater than the side of the triangle?*

Considering Fig. 1 one can give three answers:



**Fig. 1:** Bertan's paradox

1. If the chord, without loss of generality, starts at A, then it will be longer than the side if the end point is anywhere between B and C. So the answer is obviously  $\frac{1}{3}$ .
2. If the centre of the chord, without loss of generality, is chosen at random along the line OD, then it will be longer than the side of the triangle if it is in OE rather than ED. E is the midpoint of OD so the answer is obviously  $\frac{1}{2}$ .
3. If the centre of the chord, without loss of generality, is chosen at random within the circle, then it will be longer than the side of the triangle if it lies within the circle of radius  $\frac{R}{2}$ . So the answer is obviously  $\frac{1}{4}$ .

So we have three obvious but contradictory answers. The whole question is built on a false premise: drawing a chord 'at random' is, unlike tossing a coin or throwing a dice, not defined. Another way of seeing this is that a distribution which is uniform in one variable, say  $\theta$ , is not uniform in any non-trivial transformation of that variable, say  $\cos \theta$  or  $\tan \theta$ . Classical probability has therefore to be discarded.

## 2.4 Frequentist probability

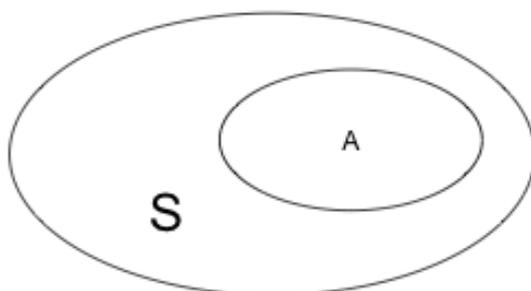
Because of such difficulties, Real Probability was replaced by Frequentist Probability in the early 20th century. This is the usual definition taught in schools and undergraduate classes. A very readable account is given by von Mises [10]:

$$P_A = \lim_{N \rightarrow \infty} \frac{N_A}{N} .$$

$N$  is the total number of events in the ensemble (or collective). It can be visualised as a Venn diagram, as in Fig. 2.

The probability of a coin landing heads up is  $\frac{1}{2}$  because if you toss a coin 1000 times, one side will come down  $\sim 500$  times. That is an empirical definition (Frequentist probability has roots in the Vienna school and logical positivism). Similarly, the lifetime of a muon is  $2.2\mu\text{s}$  because if you take 1000 muons and wait  $2.2\mu\text{s}$ , then  $\sim 368$  (that's a fraction  $e^{-1}$ ) will remain.

With this definition  $P_A$  is not just a property of  $A$  but a joint property of  $A$  and the ensemble. The same coin will have a different probability for showing head depending on whether it is in a purse or



**Fig. 2:** Frequentist probability

in a numismatic collection. This leads to two distinctive properties (or, some would say, problems) for frequentist probability.

Firstly, there may be more than one ensemble. To take an everyday example from von Mises, German life insurance companies pay out on 0.4% of 40 year old male clients. Your friend Hans is 40 today. What is the probability that he will survive to see his 41st birthday? 99.6% is an answer (if he's insured). But he is also a non-smoker and non-drinker—so perhaps the figure is higher (maybe 99.8%)? But if he drives a Harley-Davidson it should be lower (maybe 99.0%)? All these numbers are acceptable. The individual Hans belongs to several different ensembles, and the probability will be different for each of them.

To take an example from physics, suppose your experiment has a Particle Identification (PID) system using Cherenkov, time-of-flight and/or  $\frac{dE}{dx}$  measurements. You want to talk about the probability that a  $K^+$  will be correctly recognised by your PID. You determine this by considering many  $K^+$  mesons and counting the number accepted to get  $P = N_{acc}/N_{tot}$ . But these will depend on the kaon sample you work with. It could be all kaons, or kaons above a certain energy threshold, or that actually enter the detector. The ensemble can be defined in various ways, each giving a valid but different value for the probability.

On the other hand, there may be no Ensemble. To take an everyday example we might want to calculate the probability that it will rain tomorrow. This is impossible. There is only one tomorrow. It will either rain or not rain.  $P_{rain}$  is either 0 or 1, and we won't know which until tomorrow gets here. Von Mises insists that statements like 'It will probably rain tomorrow' are loose and unscientific.

To take an example from physics, consider the probability that there is a supersymmetric particle with mass below 2 TeV. Again, either there is or there isn't.

But, despite von Mises' objections, it does seem sensible, as the pressure falls and the gathering clouds turn grey, to say 'It will probably rain'. So this is a drawback to the frequentist definition. We will return to this and show how frequentists can talk meaningfully and quantitatively about unique events in the discussion of confidence intervals in Section 8.1.

## 2.5 Bayes' theorem

Before presenting Bayesian statistics we need to discuss Bayes' theorem, though we point out that Bayes' theorem applies (and is useful) in any probability model: it goes right back to the Kolmogorov axioms.

First we need to define the conditional probability:  $P(A|B)$ : this is the probability for  $A$ , given that  $B$  is true. For example: if a playing card is drawn at random from a pack of 52, then  $P(\spadesuit A) = \frac{1}{52}$ , but if you are told that the card is black, then  $P(\spadesuit A|Black) = \frac{1}{26}$  (and obviously  $P(\spadesuit A|Red) = 0$ ).

Bayes' theorem is just

$$P(A|B) = \frac{P(B|A)}{P(B)} \times P(A) \quad . \quad (2)$$

The proof is gratifyingly simple: the probability that  $A$  and  $B$  are both true can be written in two ways

$$P(A|B) \times P(B) = P(A \& B) = P(B|A) \times P(A) \quad .$$

Throw away middle term and divide by  $P(B)$  to get the result.

As a first example, we go back to the ace of spades above. A card is drawn at random, and you are told that it is black. Bayes' theorem says

$$P(\spadesuit A | Black) = \frac{P(Black | \spadesuit A)}{P(Black)} P(\spadesuit A) = \frac{1}{2} \times \frac{1}{52} = \frac{1}{26} \quad ;$$

i.e. the original probability of drawing  $\spadesuit A$ ,  $\frac{1}{52}$ , is multiplied by the probability that the ace of spades is black (just 1) and divided by the overall probability of drawing a black card ( $\frac{1}{2}$ ) to give the obvious result.

For a less trivial example, suppose you have a momentum-selected beam which is 90%  $\pi$  and 10%  $K$ . This goes through a Cherenkov counter for which pions exceed the threshold velocity but kaons do not. In principle pions will give a signal, but suppose there is a 5% chance, due to inefficiencies, that they will not. Again in principle kaons always give no Cherenkov signal, but suppose that probability is only 95% due to background noise. What is the probability that a particle identified as a kaon, as it gave no signal, is truly one?

Bayes' theorem runs

$$P(K | no\ signal) = \frac{P(no\ signal | K)}{P(no\ signal)} \times P(K) = \frac{0.95}{0.95 \times 0.1 + 0.05 \times 0.9} \times 0.1 = 0.68 \quad ,$$

showing that the probability is only  $\frac{2}{3}$ . The positive identification is not enough to overwhelm the 9:1  $\pi : K$  ratio. Incidentally this uses the (often handy) expression for the denominator:  $P(B) = P(B|A) \times P(A) + P(B|\bar{A}) \times \bar{P}(A)$ .

## 2.6 Bayesian probability

The Bayesian definition of probability is that  $P_A$  represents your belief in  $A$ . 1 represents certainty, 0 represents total disbelief. Intermediate values can be calibrated by asking whether you would prefer to bet on  $A$ , or on a white ball being drawn from an urn containing a mix of white and black balls.

This avoids the limitations of frequentist probability—coins, dice, kaons, rain tomorrow, existence of supersymmetry (SUSY) can all have probabilities assigned to them.

The drawback is that your value for  $P_A$  may be different from mine, or anyone else's. It is also called subjective probability.

Bayesian probability makes great use of Bayes' theorem, in the form

$$P(Theory|Data) = \frac{P(Data|Theory)}{P(Data)} \times P(Theory) \quad . \quad (3)$$

$P(Theory)$  is called the *prior*: your initial belief in *Theory*.  $P(Data|Theory)$  is the *Likelihood*: the probability of getting *Data* if *Theory* is true.  $P(Theory|Data)$  is the *Posterior*: your belief in *Theory* in the light of a particular *Data* being observed.

So this all works very sensibly. If the data observed is predicted by the theory, your belief in that theory is boosted, though this is moderated by the probability that the data could have arisen anyway. Conversely, if data is observed which is disfavoured by the theory, your belief in that theory is weakened.

The process can be chained. The posterior from a first experiment can be taken as the prior for a second experiment, and so on. When you write out the factors you find that the order doesn't matter.

### 2.6.1 Prior distributions

Often, though, the theory being considered is not totally defined: it may contain a parameter (or several parameters) such as a mass, coupling constant, or decay rate. Generically we will call this  $a$ , with the proviso that it may be multidimensional.

The prior is now not a single number  $P(\text{Theory})$  but a probability distribution  $P_0(a)$ .  $\int_{a_1}^{a_2} P_0(a) da$  is your prior belief that  $a$  lies between  $a_1$  and  $a_2$ .  $\int_{-\infty}^{\infty} P_0(a) da$  is your original  $P(\text{Theory})$ . This is generally taken as 1, which is valid provided the possibility that the theory that is false is matched by some value of  $a$ —for example if the coupling constant for a hypothetical particle is zero, that accommodates any belief that it might not exist. Bayes' theorem then runs:

$$P_1(a; x) \propto L(a; x)P_0(a) \quad . \quad (4)$$

If the range of  $a$  is infinite,  $P_0(a)$  may be vanishingly small (this is called an 'improper prior'). However this is not a problem. Suppose, for example, that all we know about  $a$  is that it is non-negative, and we are genuinely equally open to its having any value. We write  $P_0(a)$  as  $C$ , so  $\int_{a_1}^{a_2} P_0(a) da = C(a_2 - a_1)$ . This probability is vanishingly small: if you were offered the choice of a bet on  $a$  lying within the range  $[a_1, a_2]$  or of drawing a white ball from an urn containing 1 white ball and  $N$  black balls, you would choose the latter, however large  $N$  was. However it is not zero: if the urn contained  $N$  black balls, but no white ball, your betting choice would change. After a measurement you have  $P_1(a; x) = \frac{L(a; x)}{\int L(a; x)C da} C$ , and the factors of  $C$  can be cancelled (which, and this is the point, you could *not* do if  $C$  were exactly zero) giving  $P_1(a; x) = \frac{L(a; x)}{\int L(a; x) da}$  or,  $P_1(a; x) \propto L(a; x)$ , and you can then just normalize  $P_1(a)$  to 1.

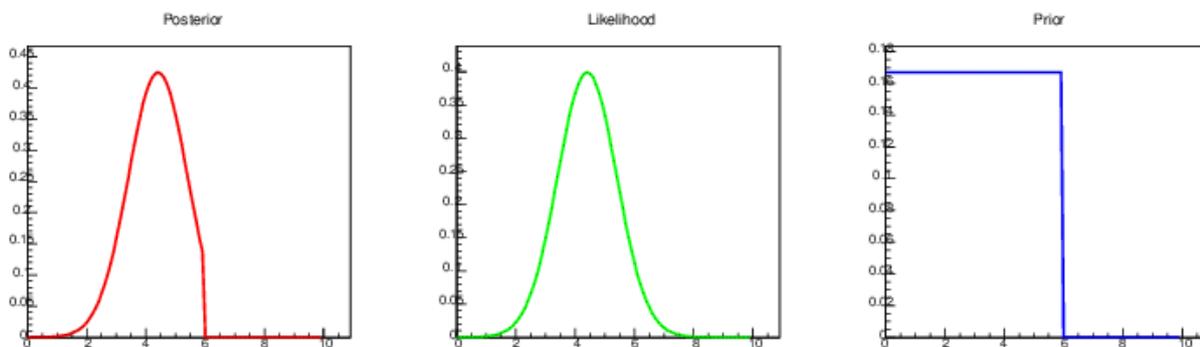


Fig. 3: Bayes at work

Figure 3 shows Eq. 4 at work. Suppose  $a$  is known to lie between 0 and 6, and the prior distribution is taken as flat, as shown in the left hand plot. A measurement of  $a$  gives a result  $4.4 \pm 1.0$ , as shown in the central plot. The product of the two gives (after normalization) the posterior, as shown in the right hand plot.

### 2.6.2 Likelihood

The likelihood—the number  $P(\text{Data}|\text{Theory})$ —is now generalised to the function  $L(a, x)$ , where  $x$  is the observed value of the data. Again,  $x$  may be multidimensional, but in what follows it is not misleading to ignore that.

This can be confusing. For example, anticipating Section 3.2.2, the probability of getting  $x$  counts from a Poisson process with mean  $a$  is

$$P(x, a) = e^{-a} \frac{a^x}{x!} . \quad (5)$$

We also write

$$L(a, x) = e^{-a} \frac{a^x}{x!} . \quad (6)$$

What's the difference? Technically there is none. These are identical joint functions of two variables ( $x$  and  $a$ ) to which we have just happened to have given different names. Pragmatically we regard Eq. 5 as describing the probability of getting various different  $x$  from some fixed  $a$ , whereas Eq. 6 describes the likelihood for various different  $a$  from some given  $x$ . But be careful with the term 'likelihood'. If  $P(x_1, a) > P(x_2, a)$  then  $x_1$  is more probable (whatever you mean by that) than  $x_2$ . If  $L(a_1, x) > L(a_2, x)$  it does not mean that  $a_1$  is more likely (however you define that) than  $a_2$ .

### 2.6.3 Shortcomings of Bayesian probability

The big problem with Bayesian probability is that it is subjective. Your  $P_0(a)$  and my  $P_0(a)$  may be different—so how can we compare results? Science does, after all, take pride in being objective: it handles real facts, not opinions. If you present a Bayesian result from your search for the  $X$  particle this embodies the actual experiment and your irrational prior prejudices. I am interested in your experiment but not in your irrational prior prejudices—I have my own—and it is unhelpful if you combine the two.

Bayesians sometimes ask about the right prior they should use. This is the wrong question. The prior is what you believe, and only you know that.

There is an argument made for taking the prior as uniform. This is sometimes called the 'Principle of ignorance' and justified as being impartial. But this is misleading, even dishonest. If  $P_0(a)$  is taken as constant, favouring no particular value, then it is not constant for  $a^2$  or  $\sqrt{a}$  or  $\ln a$ , which are equally valid parameters.

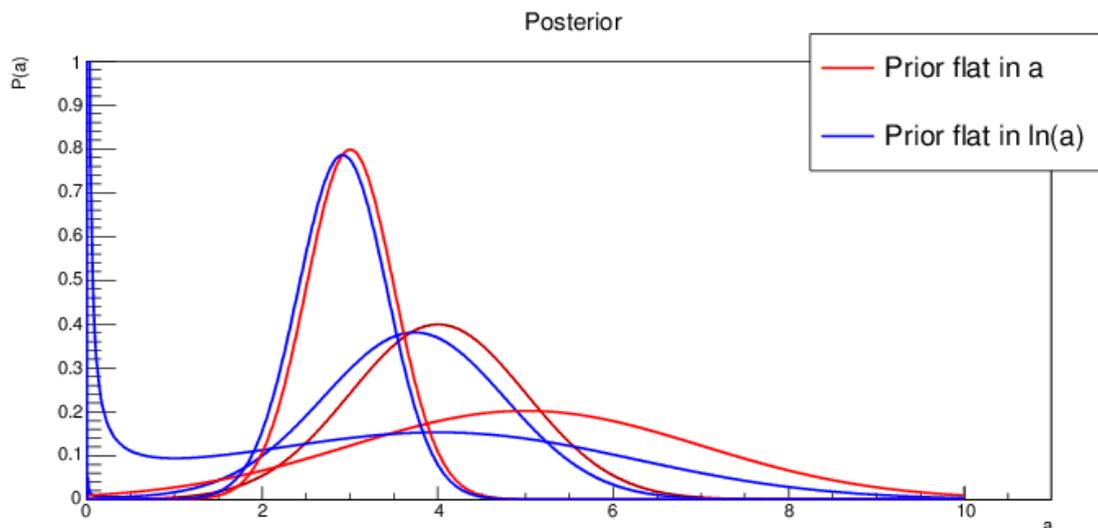
It is true that with lots of data,  $P_1(a)$  decouples from  $P_0(a)$ . The final result depends only on the measurements. But this is not the case with little data—and that's the situation we're usually in—when doing statistics properly matters.

As an example, suppose you make a Gaussian measurement (anticipating slightly Section 3.2.3). You consider a prior flat in  $a$  and a prior flat in  $\ln a$ . This latter is quite sensible—it says you expect a result between 0.1 and 0.2 as being equally likely as a result between 1 and 2, or 10 and 20. The posteriors are shown in Fig. 4. For an 'accurate' result of  $3 \pm 0.5$  the posteriors are very close. For an 'intermediate' result of  $4.0 \pm 1.0$  there is an appreciable difference in the peak value and the shape. For a 'poor' measurement of  $5.0 \pm 2.0$  the posteriors are *very* different.

So you should never just quote results from a single prior. Try several forms of prior and examine the spread of results. If they are pretty much the same you are vindicated. This is called 'robustness under choice of prior' and it is standard practice for statisticians. If they are different then the data are telling you about the limitations of your results.

### 2.6.4 Jeffreys' prior

Jeffreys [11] suggested a technique now known as the Jeffreys' or *objective prior*: that you should choose a prior flat in a transformed variable  $a'$  for which the Fisher information,  $\mathcal{I} = -\left\langle \frac{\partial^2 L(x;a)}{\partial a^2} \right\rangle$  is constant. The Fisher information (which is important in maximum likelihood estimation, as described in Section 5.2) is a measure of how much a measurement tells you about the parameter: a large  $\mathcal{I}$  has a likelihood function with a sharp peak and will tell you (by some measure) a lot about  $a$ ; a small  $\mathcal{I}$  has a featureless likelihood function which will not be useful. Jeffrey's principle is that the prior should not favour or disfavour particular values of the parameter. It is equivalently—and more conveniently—used as taking a prior in the original  $a$  which is proportional to  $\sqrt{\mathcal{I}}$ .



**Fig. 4:** Posteriors for two different priors for the results  $3.0 \pm 0.5$ ,  $4.0 \pm 1.0$  and  $5.0 \pm 2.0$

It has not been universally adopted for various reasons. Some practitioners like to be able to include their own prior belief into the analysis. It also makes the prior dependent on the experiment (in the form of the likelihood function). Thus if ATLAS and CMS searched for the same new  $X$  particle they would use different priors for  $P_0(M_X)$ , which is (to some people) absurd.

So it is not universal—but when you are selecting a bunch of priors to test robustness—the Jefferys’ prior is a strong contender for inclusion.

## 2.7 Summary

So mathematical probability has no meaning, and real probability is discredited. That leaves the Frequentist and Bayesian definitions. Both are very much in use.

They are sometimes presented as rivals, with adherents on either side (‘frequentists versus Bayesians’). This is needless drama. They are both tools that help us understand our results. Both have drawbacks. Sometimes it is clear which is the best tool for a particular job, sometimes it is not and one is free to choose either. It is said—probably accurately—that particle physicists feel happier with frequentist probability as they are used to large ensembles of similar but different events, whereas astrophysicists and cosmologists are more at home with Bayesian probability as they only have one universe to consider.

What is important is not which version you prefer—these are not football teams—but that you know the limitations of each, that you use the best definition when there is a reason to do so, and, above all, that you are aware of which form you are using.

As a possibly heretical afterthought, perhaps classical probability still has a place? Quantum Mechanics, after all, gives probabilities. If  $P_A$  is not ‘real’—either because it depends on an arbitrary ensemble, or because it is a subjective belief—then it looks like there is nothing ‘real’ in the universe.

The state of a coin—or an electron spin—having probability  $\frac{1}{2}$  makes sense. There is a symmetry that dictates it. The lifetime of a muon—i.e. probability per unit time that it will decay—seems to be a well-defined quantity, a property of the muon and independent of any ensemble, or any Bayesian belief.

The probability a muon will produce a signal in your muon detector seems like a ‘real well-defined quantity’, if you specify the 4 momentum and the state of the detector. Of course the inverse probability

‘What is the probability that a muon signal in my detector comes from a real muon, not background’ is not intrinsically defined, So perhaps classical probability has a place in physics—but not in interpreting results. However you should not mention this to a statistician or they will think you’re crazy.

### 3 Probability distributions and their properties

We have to make a simple distinction between two sorts of data: *integer* data and *real-number* data<sup>1</sup>.

The first covers results which are of their nature whole numbers: the numbers of kaons produced in a collision, or the number of entries falling into some bin of a histogram. Generically let’s call such numbers  $r$ . They have probabilities  $P(r)$  which are dimensionless.

The second covers results whose values are real (or floating-point) numbers. There are lots of these: energies, angles, invariant masses . . . Generically let’s call such numbers  $x$ , and they have probability density functions  $P(x)$  which have dimensions of  $[x]^{-1}$ , so  $\int_{x_1}^{x_2} P(x)dx$  or  $P(x) dx$  are probabilities.

You will also sometimes meet the cumulative distribution  $C(x) = \int_{-\infty}^x P(x') dx'$ .

#### 3.1 Expectation values

From  $P(r)$  or  $P(x)$  one can form the expectation value

$$\langle f \rangle = \sum_r f(r)P(r) \quad \text{or} \quad \langle f \rangle = \int f(x)P(x) dx \quad , \quad (7)$$

where the sum or integral is taken as appropriate. Some authors write this as  $E(f)$ , but I personally prefer the angle-bracket notation. You may think it looks too much like quantum mechanics, but in fact it’s quantum mechanics which looks like statistics: an expression like  $\langle \psi | \hat{Q} | \psi \rangle$  is the average value of an operator  $\hat{Q}$  in some state  $\psi$ , where ‘average value’ has exactly the same meaning and significance.

##### 3.1.1 Mean and standard deviation

In particular the *mean*, often written  $\mu$ , is given by

$$\langle r \rangle = \sum_r rP(r) \quad \text{or} \quad \langle x \rangle = \int xP(x) dx \quad .$$

Similarly one can write higher *moments*

$$\mu_k = \langle r^k \rangle = \sum_r r^k P(r) \quad \text{or} \quad \langle x^k \rangle = \int x^k P(x) dx \quad ,$$

and *central moments*

$$\mu'_k = \langle (r - \mu)^k \rangle = \sum_r (r - \mu)^k P(r) \quad \text{or} \quad \langle (x - \mu)^k \rangle = \int (x - \mu)^k P(x) dx \quad .$$

The second central moment is known as the *variance*

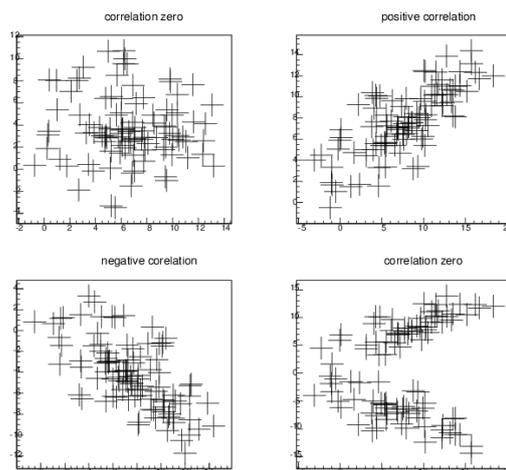
$$\mu'_2 = V = \sum_r (r - \mu)^2 P(r) = \langle r^2 \rangle - \langle r \rangle^2 \quad \text{or} \quad \int (x - \mu)^2 P(x) dx = \langle x^2 \rangle - \langle x \rangle^2$$

It is easy to show that  $\langle (x - \mu)^2 \rangle = \langle x^2 \rangle - \mu^2$ . The *standard deviation* is just the square root of the variance  $\sigma = \sqrt{V}$ .

Statisticians usually use variance, perhaps because formulae come out simpler. Physicists usually use standard deviation, perhaps because it has the same dimensions as the variable being studied, and can be drawn as an error bar on a plot.

You may also meet *skew*, which is  $\gamma = \langle (x - \mu)^3 \rangle / \sigma^3$  and *kurtosis*,  $h = \langle (x - \mu)^4 \rangle / \sigma^4 - 3$ . Definitions vary, so be careful. Skew is a dimensionless measure of the asymmetry of a distribution. Kurtosis is (thanks to that rather arbitrary looking 3 in the definition) zero for a Gaussian distribution (see Section 3.2.3): positive kurtosis indicates a narrow core with a wide tail, negative kurtosis indicates the tails are reduced.

<sup>1</sup>Other branches of science have to include a third, *categorical* data, but we will ignore that.



**Fig. 5:** Examples of two dimensional distributions. The top right has positive covariance (and correlation), the bottom left negative. In the top left the covariance is zero and  $x$  and  $y$  are independent; in the bottom right the covariance is also zero, but they are not independent.

### 3.1.2 Covariance and correlation

If your data are 2-dimensional pairs  $(x, y)$ , then besides forming  $\langle x \rangle$ ,  $\langle y \rangle$ ,  $\sigma_x$  etc., you can also form the *Covariance*

$$\text{Cov}(x, y) = \langle (x - \mu_x)(y - \mu_y) \rangle = \langle xy \rangle - \langle x \rangle \langle y \rangle \quad .$$

Examples are shown in Fig. 5. If there is a tendency for positive fluctuations in  $x$  to be associated with positive fluctuations in  $y$  (and therefore negative with negative) then the product  $(x_i - \bar{x})(y_i - \bar{y})$  tends to be positive and the covariance is greater than 0. A negative covariance, as in the 3rd plot, happens if a positive fluctuation in one variable is associated with a negative fluctuation in the other. If the variables are independent then a positive variation in  $x$  is equally likely to be associated with a positive or a negative variation in  $y$  and the covariance is zero, as in the first plot. However the converse is not always the case, there can be two-dimensional distributions where the covariance is zero, but the two variables are not independent, as is shown in the fourth plot.

Covariance is useful, but it has dimensions. Often one uses the *correlation*, which is just

$$\rho = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} \quad . \quad (8)$$

It is easy to show that  $\rho$  lies between 1 (complete correlation) and -1 (complete anticorrelation).  $\rho = 0$  if  $x$  and  $y$  are independent.

If there are more than two variables—the alphabet runs out so let's call them  $(x_1, x_2, x_3 \dots x_n)$ —then these generalise to the *covariance matrix*

$$\mathbf{V}_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$$

and the *correlation matrix*

$$\rho_{ij} = \frac{\mathbf{V}_{ij}}{\sigma_i \sigma_j} \quad .$$

The diagonal of  $\mathbf{V}$  is  $\sigma_i^2$ . The diagonal of  $\rho$  is 1.

## 3.2 Binomial, Poisson and Gaussian

We now move from considering the general properties of distributions to considering three specific ones. These are the ones you will most commonly meet for the distribution of the original data (as opposed to

quantities constructed from it). Actually the first, the binomial, is not nearly as common as the second, the Poisson; and the third, the Gaussian, is overwhelmingly more common. However it is useful to consider all three as concepts are built up from the simplest to the more sophisticated.

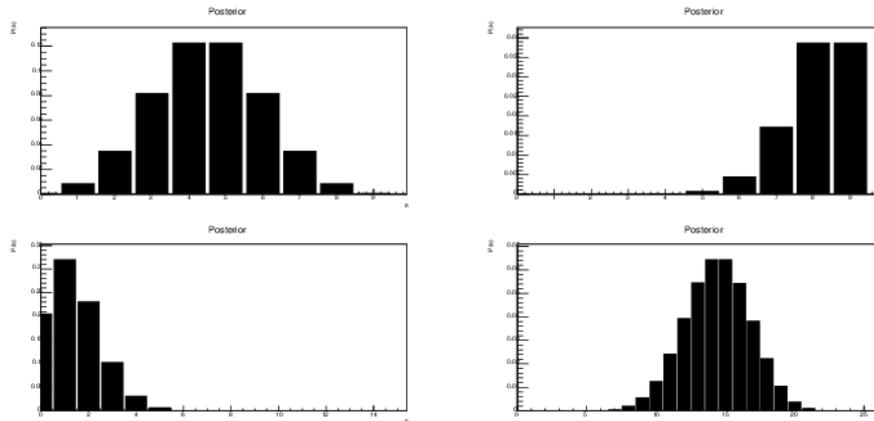
### 3.2.1 The binomial distribution

The binomial distribution is easy to understand as it basically describes the familiar<sup>2</sup> tossing of coins. It describes the number  $r$  of successes in  $N$  trials, each with probability  $p$  of success.  $r$  is discrete so the process is described by a probability distribution

$$P(r; p, N) = \frac{N!}{r!(N-r)!} p^r q^{N-r} \quad , \quad (9)$$

where  $q \equiv 1 - p$ .

Some examples are shown in Fig. 6.



**Fig. 6:** Some examples of the binomial distribution, for (1)  $N = 10, p = 0.6$ , (2)  $N = 10, p = 0.9$ , (3)  $N = 15, p = 0.1$ , and (4)  $N = 25, p = 0.6$ .

The distribution has mean  $\mu = Np$ , variance  $V = Npq$ , and standard deviation  $\sigma = \sqrt{Npq}$ .

### 3.2.2 The Poisson distribution

The Poisson distribution also describes the probability of some discrete number  $r$ , but rather than a fixed number of ‘trials’ it considers a random rate  $\lambda$ :

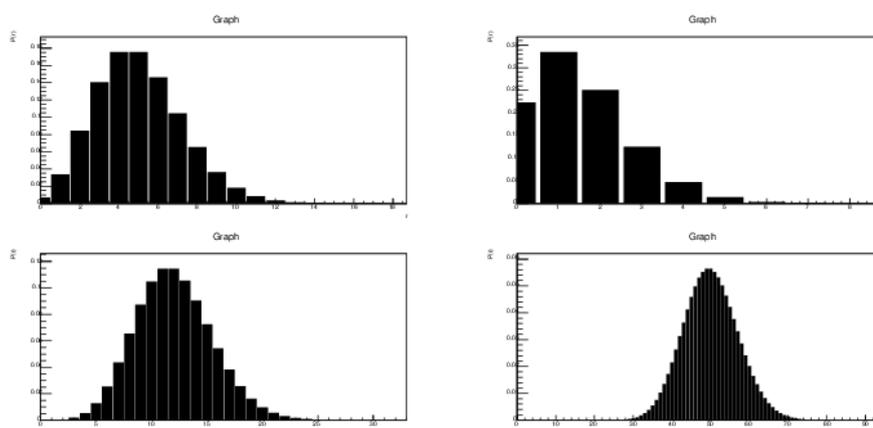
$$P(r; \lambda) = e^{-\lambda} \frac{\lambda^r}{r!} \quad . \quad (10)$$

It is linked to the binomial—the Poisson is the limit of the binomial—as  $N \rightarrow \infty, p \rightarrow 0$  with  $np = \lambda = \text{constant}$ . Figure 7 shows various examples. It has mean  $\mu = \lambda$ , variance  $V = \lambda$ , and standard deviation  $\sigma = \sqrt{\lambda} = \sqrt{\mu}$ .

The clicks of a Geiger counter are the standard illustration of a Poisson process. You will meet it a lot as it applies to event counts—on their own or in histogram bins.

To help you think about the Poisson, here is a simple question (which describes a situation I have seen in practice, more than once, from people who ought to know better).

<sup>2</sup>Except, as it happens, in Vietnam, where coins have been completely replaced by banknotes.



**Fig. 7:** Poisson distributions for (1)  $\lambda = 5$ , (2)  $\lambda = 1.5$ , (3)  $\lambda = 12$  and (4)  $\lambda = 50$

You need to know the efficiency of your PID system for positrons.

You find 1000 data events where 2 tracks have a combined mass of 3.1 GeV ( $J/\psi$ ) and the negative track is identified as an  $e^-$  ('Tag-and-probe' technique).

In 900 events the  $e^+$  is also identified. In 100 events it is not. The efficiency is 90%.

What about the error?

Colleague A says  $\sqrt{900} = 30$  so efficiency is  $90.0 \pm 3.0\%$ ,

colleague B says  $\sqrt{100} = 10$  so efficiency is  $90.0 \pm 1.0\%$ .

Which is right?

Please think about this before turning the page...

Neither—both are wrong. This is binomial not Poisson:  $p = 0.9, N = 1000$ .

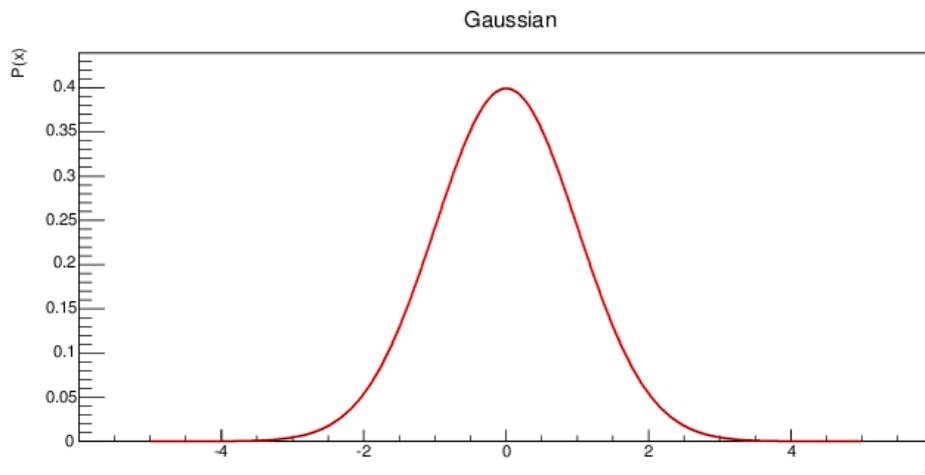
The error is  $\sqrt{Npq} = \sqrt{1000 \times 0.9 \times 0.1}$  (or  $\sqrt{1000 \times 0.1 \times 0.9} = \sqrt{90} = 9.49$  so the efficiency is  $90.0 \pm 0.9 \%$ .

### 3.2.3 The Gaussian distribution

This is by far the most important statistical distribution. The probability density function (PDF) for a variable  $x$  is given by the formula

$$P(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (11)$$

Pictorially this is shown in Fig. 8.



**Fig. 8:** The Gaussian distribution

This is sometimes called the ‘bell curve’, though in fact a real bell does not have flared edges like that. There is (in contrast to the Poisson and binomial) only one Gaussian curve, as  $\mu$  and  $\sigma$  are just location and scale parameters.

The mean is  $\mu$  and the standard deviation is  $\sigma$ . The Skew is zero, as it is symmetric, and the kurtosis is zero by construction.

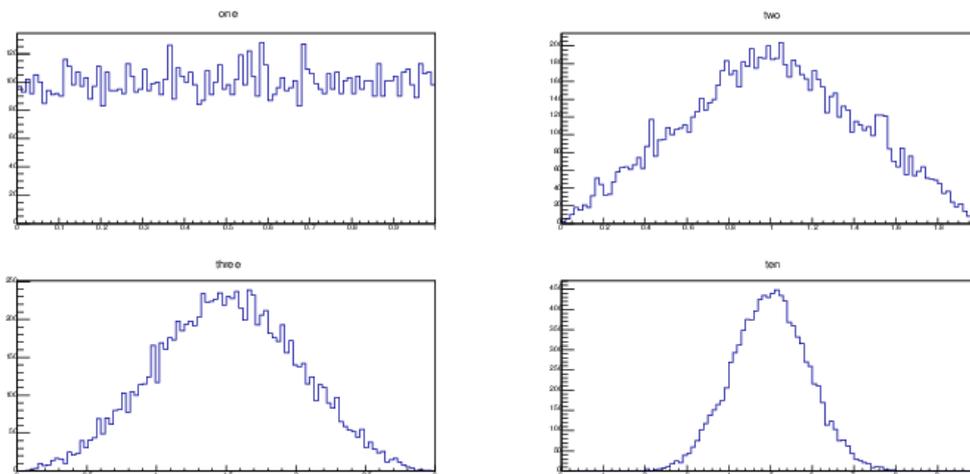
In statistics, and most disciplines, this is known as the *normal distribution*. Only in physics is it known as ‘The Gaussian’—perhaps because the word ‘normal’ already has so many meanings.

The reason for the importance of the Gaussian is the *central limit theorem* (CLT) that states: if the variable  $X$  is the sum of  $N$  variables  $x_1, x_2 \dots x_N$  then:

1. Means add:  $\langle X \rangle = \langle x_1 \rangle + \langle x_2 \rangle + \dots \langle x_N \rangle$ ,
2. Variances add:  $V_X = V_1 + V_2 + \dots V_N$ ,
3. If the variables  $x_i$  are independent and identically distributed (i.i.d.) then  $P(X)$  tends to a Gaussian for large  $N$ .

(1) is obvious, (2) is pretty obvious, and means that standard deviations add in quadrature, and that the standard deviation of an average falls like  $\frac{1}{\sqrt{N}}$ , (3) applies whatever the form of the original  $P(x)$ .

Before proving this, it is helpful to see a demonstration to convince yourself that the implausible assertion in (3) actually does happen. Take a uniform distribution from 0 to 1, as shown in the top left subplot of Fig. 9. It is flat. Add two such numbers and the distribution is triangular, between 0 and 2, as shown in the top right.



**Fig. 9:** Demonstration of the central limit theorem

With 3 numbers, at the bottom left, it gets curved. With 10 numbers, at the bottom right, it looks pretty Gaussian. The proof follows.

*Proof.* First, introduce the characteristic function  $\langle e^{ikx} \rangle = \int e^{ikx} P(x) dx = \tilde{P}(k)$ .

This can usefully be thought of as an expectation value and as a Fourier transform, FT.

Expand the exponential as a series

$\langle e^{ikx} \rangle = \langle 1 + ikx + \frac{(ikx)^2}{2!} + \frac{(ikx)^3}{3!} \dots \rangle = 1 + ik\langle x \rangle + (ik)^2 \frac{\langle x^2 \rangle}{2!} + (ik)^3 \frac{\langle x^3 \rangle}{3!} \dots$ . Take the logarithm and use the expansion  $\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} \dots$ . This gives a power series in  $(ik)$ , where the coefficient  $\frac{\kappa_r}{r!}$  of  $(ik)^r$  is made up of expectation values of  $x$  of total power  $r$

$$\kappa_1 = \langle x \rangle, \kappa_2 = \langle x^2 \rangle - \langle x \rangle^2, \kappa_3 = \langle x^3 \rangle - 3\langle x^2 \rangle \langle x \rangle + 2\langle x \rangle^3 \dots$$

These are called the semi-invariant cumulants of Thièle. Under a change of scale  $\alpha$ ,  $\kappa_r \rightarrow \alpha^r \kappa_r$ . Under a change in location only  $\kappa_1$  changes.

If  $X$  is the sum of i.i.d. random variables,  $x_1 + x_2 + x_3 \dots$ , then  $P(X)$  is the convolution of  $P(x)$  with itself  $N$  times.

The FT of a convolution is the product of the individual FTs,

the logarithm of a product is the sum of the logarithms,

so  $P(X)$  has cumulants  $K_r = N\kappa_r$ .

To make graphs commensurate, you need to scale the  $X$  axis by the standard deviation, which grows like  $\sqrt{N}$ . The cumulants of the scaled graph are  $K'_r = N^{1-r/2} \kappa_r$ .

As  $N \rightarrow \infty$ , these vanish for  $r > 2$ , leaving a quadratic.

If the log is a quadratic, the exponential is a Gaussian. So  $\tilde{P}(X)$  is Gaussian.

And finally, the inverse FT of a Gaussian is also a Gaussian. □

Even if the distributions are not identical, the CLT tends to apply, unless one (or two) dominates. Most ‘errors’ fit this, being compounded of many different sources.

#### 4 Hypothesis testing

‘Hypothesis testing’ is another piece of statistical technical jargon. It just means ‘making choices’—in a logical way—on the basis of statistical information.

- Is some track a pion or a kaon?
- Is this event signal or background?
- Is the detector performance degrading with time?
- Do the data agree with the Standard Model prediction or not?

To establish some terms: you have a *hypothesis* (the track is a pion, the event is signal, the detector is stable, the Standard Model is fine . . .) and an alternative hypothesis (kaon, background, changing, new physics needed . . .) Your hypothesis is usually *simple* i.e. completely specified, but the alternative is often *composite* containing a parameter (for example, the detector decay rate) which may have any non-zero value.

##### 4.1 Type I and type II errors

As an example, let’s use the signal/background decision. Do you accept or reject the event (perhaps in the trigger, perhaps in your offline analysis)? To make things easy we consider the case where both hypotheses are simple, i.e. completely defined.

Suppose you measure some parameter  $x$  which is related to what you are trying to measure. It may well be the output from a neural network or other machine learning (ML) systems. The expected distributions for  $x$  under the hypothesis and the alternative,  $S$  and  $B$  respectively, are shown in Fig. 10.

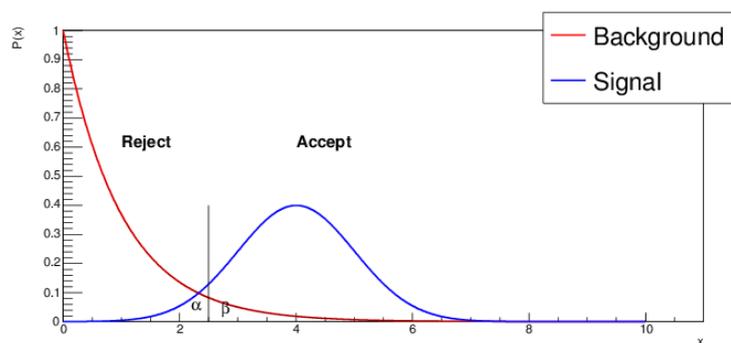


Fig. 10: Hypothesis testing example

You impose a cut as shown—you have to put one somewhere—accepting events above  $x = x_{cut}$  and rejecting those below.

This means losing a fraction  $\alpha$  of signal. This is called a *type I error* and  $\alpha$  is known as the *significance*.

You admit a fraction  $\beta$  of background. This is called a *type II error* and  $1 - \beta$  is the power.

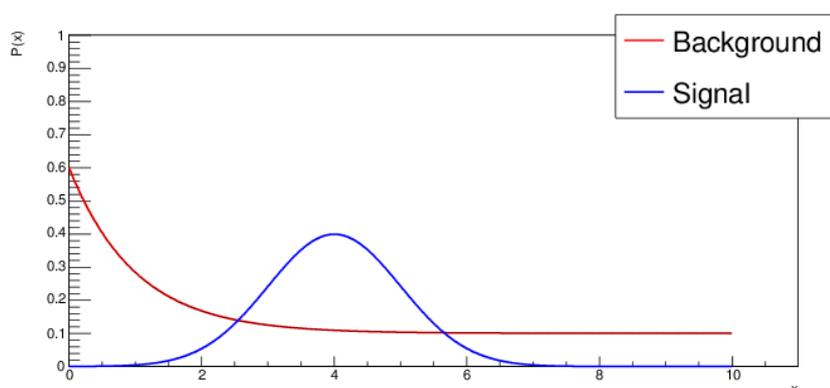
You would like to know the best place to put the cut. This graph cannot tell you! The strategy for the cut depends on three things—hypothesis testing only covers one of them.

The second is the prior signal to noise ratio. These plots are normalized to 1. The red curve is (probably) MUCH bigger. A value of  $\beta$  of, say, 0.01 looks nice and small—only one in a hundred background events get through. But if your background is 10,000 times bigger than your signal (and it often is) you are still swamped.

The third is the cost of making mistakes, which will be different for the two types of error. You have a trade-off between efficiency and purity: what are they worth? In a typical analysis, a type II error is more serious than a type I: losing a signal event is regrettable, but it happens. Including background events in your selected pure sample can give a very misleading result. By contrast, in medical decisions, type I errors are much worse than type II. Telling healthy patients they are sick leads to worry and perhaps further tests, but telling sick patients they are healthy means they don't get the treatment they need.

## 4.2 The Neymann-Pearson lemma

In Fig. 10 the strategy is plain—you choose  $x_{cut}$  and evaluate  $\alpha$  and  $\beta$ . But suppose the  $S$  and  $B$  curves are more complicated, as in Fig. 11? Or that  $x$  is multidimensional?



**Fig. 11:** A more complicated case for hypothesis testing

Neymann and Pearson say: your acceptance region just includes regions of greatest  $\frac{S(x)}{B(x)}$  (the ratio of likelihoods). For a given  $\alpha$ , this gives the smallest  $\beta$  ('Most powerful at a given significance')

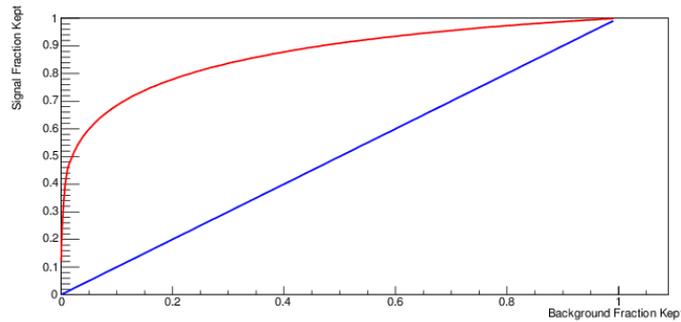
The proof is simple: having done this, if you then move a small region from 'accept' to 'reject' it has to be replaced by an equivalent region, to balance  $\alpha$ , which (by construction) brings more background, increasing  $\beta$ .

However complicated, such a problem reduces to a single monotonic variable  $\frac{S}{B}$ , and you cut on that.

## 4.3 Efficiency, purity, and ROC plots

ROC plots are often used to show the efficacy of different selection variables. You scan over the cut value (in  $x$ , for Fig. 10 or in  $S/B$  for a case like Fig. 11 and plot the fraction of background accepted ( $\beta$ ) against fraction of signal retained ( $1 - \alpha$ ), as shown in Fig. 12.

For a very loose cut all data is accepted, corresponding to a point at the top right. As the cut is tightened both signal and background fractions fall, so the point moves to the left and down, though hopefully the background loss is greater than the signal loss, so it moves more to the left than it does downwards. As the cut is increased the line moves towards the bottom left, the limit of a very tight cut where all data is rejected.



**Fig. 12:** ROC curves

A diagonal line corresponds to no discrimination—the  $S$  and  $B$  curves are identical. The further the actual line bulges away from that diagonal, the better.

Where you should put your cut depends, as pointed out earlier, also on the prior signal/background ratio and the relative costs of errors. The ROC plots do not tell you that, but they can be useful in comparing the performance of different discriminators.

The name ‘ROC’ stands for ‘receiver operating characteristic’, for reasons that are lost in history. Actually it is good to use this meaningless acronym, otherwise they get called ‘efficiency-purity plots’ even though they definitely do not show the purity (they cannot, as that depends on the overall signal/background ratio). Be careful, as the phrases ‘background efficiency’, ‘contamination’, and ‘purity’ are used ambiguously in the literature.

#### 4.4 The null hypothesis

An analysis is often (but not always) investigating whether an effect is present, motivated by the hope that the results will show that it is:

- Eating broccoli makes you smart.
- Facebook advertising increases sales.
- A new drug increases patient survival rates.
- The data show Beyond-the-Standard-Model physics.

To reach such a conclusion you have to use your best efforts to try, and to fail, to prove the opposite: the *Null Hypothesis*  $H_0$ .

- Broccoli lovers have the same or small IQ than broccoli loathers.
- Sales are independent of the Facebook advertising budget.
- The survival rates for the new treatment is the same.
- The Standard Model (functions or Monte-Carlo) describe the data.

If the null hypothesis is not tenable, you’ve proved—or at least, supported—your point.

The reason for calling  $\alpha$  the ‘significance’ is now clear. It is the probability that the null hypothesis will be wrongly rejected, and you’ll claim an effect where there isn’t any.

There is a minefield of difficulties. Correlation is not causation. If broccoli eaters are more intelligent, perhaps that’s because it’s intelligent to eat green vegetables, not that vegetables make you intelligent. One has to consider that if similar experiments are done, self-censorship will influence which results get published. This is further discussed in Section 9.

This account is perhaps unconventional in introducing the null hypothesis at such a late stage. Most treatments bring it in right at the start of the description of hypothesis testing, because they assume that all decisions are of this type.

## 5 Estimation

What statisticians call ‘estimation’, physicists would generally call ‘measurement’.

Suppose you know the probability (density) function  $P(x; a)$  and you take a set of data  $\{x_i\}$ . What is the best value for  $a$ ? (Sometimes one wants to estimate a property (e.g. the mean) rather than a parameter, but this is relatively uncommon, and the methodology is the same.)

$x_i$  may be single values, or pairs, or higher-dimensional. The unknown  $a$  may be a single parameter or several. If it has more than one component, these are sometimes split into ‘parameters of interest’ and ‘nuisance parameters’.

The *estimator* is defined very broadly: an estimator  $\hat{a}(x_1 \dots x_N)$  is a function of the data that gives a value for the parameter  $a$ . There is no ‘correct’ estimator, but some are better than others. A perfect estimator would be:

- Consistent.  $\hat{a}(x_1 \dots x_N) \rightarrow a$  as  $N \rightarrow \infty$ ,
- Unbiased:  $\langle \hat{a} \rangle = a$ ,
- Efficient:  $\langle (\hat{a} - a)^2 \rangle$  is as small as possible,
- Invariant:  $\hat{f}(a) = f(\hat{a})$ .

No estimator is perfect—these 4 goals are incompatible. In particular the second and the fourth; if an estimator  $\hat{a}$  is unbiased for  $a$  then  $\sqrt{\hat{a}}$  is not an unbiased estimator of  $\sqrt{a}$ .

### 5.1 Bias

Suppose we estimate the mean by taking the obvious<sup>3</sup>  $\hat{\mu} = \bar{x}$

$$\langle \hat{\mu} \rangle = \langle \frac{1}{N} \sum x_i \rangle = \frac{1}{N} \sum \mu = \mu.$$

So there is no bias. This expectation value of this estimator of  $\mu$  is just  $\mu$  itself. By contrast suppose we estimate the variance by the apparently obvious  $\hat{V} = \overline{x^2} - \bar{x}^2$ .

$$\text{Then } \langle \hat{V} \rangle = \langle \overline{x^2} \rangle - \langle \bar{x}^2 \rangle.$$

The first term is just  $\langle x^2 \rangle$ . To make sense of the second term, note that  $\langle x \rangle = \langle \bar{x} \rangle$  and add and subtract  $\langle x \rangle^2$  to get

$$\begin{aligned} \langle \hat{V} \rangle &= \langle x^2 \rangle - \langle x \rangle^2 - (\langle \bar{x}^2 \rangle - \langle \bar{x} \rangle^2) \\ \langle \hat{V} \rangle &= V(x) - V(\bar{x}) = V - \frac{V}{N} = \frac{N-1}{N} V. \end{aligned}$$

So the estimator is biased!  $\hat{V}$  will, on average, give too small a value.

This bias, like any known bias, can be corrected for. Using  $\hat{V} = \frac{N}{N-1} (\overline{x^2} - \bar{x}^2)$  corrects the bias.

The familiar estimator for the standard deviation follows:  $\hat{\sigma} = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N-1}}$ .

(Of course this gives a biased estimate of  $\sigma$ . But  $V$  is generally more important in this context.)

<sup>3</sup>Note the difference between  $\langle x \rangle$  which is an average over a PDF and  $\bar{x}$  which denotes the average over a particular sample: both are called ‘the mean  $x$ ’.

## 5.2 Efficiency

Somewhat surprisingly, there is a limit to the efficiency of an estimator: the *minimum variance bound* (MVB), also known as the *Cramér-Rao bound*.

For any unbiased estimator  $\hat{a}(x)$ , the variance is bounded

$$V(\hat{a}) \geq -\frac{1}{\left\langle \frac{d^2 \ln L}{da^2} \right\rangle} = \frac{1}{\left\langle \left( \frac{d \ln L}{da} \right)^2 \right\rangle} . \quad (12)$$

$L$  is the likelihood (as introduced in Section 2.6.2) of a sample of independent measurements, i.e. the probability for the whole data sample for a particular value of  $a$ . It is just the product of the individual probabilities:

$$L(a; x_1, x_2, \dots, x_N) = P(x_1; a)P(x_2; a)\dots P(x_N; a).$$

We will write  $L(a; x_1, x_2, \dots, x_N)$  as  $L(a; x)$  for simplicity.

*Proof.* Proof of the MVB

$$\text{Unitarity requires } \int P(x; a) dx = \int L(a; x) dx = 1$$

Differentiate wrt  $a$ :

$$0 = \int \frac{dL}{da} dx = \int L \frac{d \ln L}{da} dx = \left\langle \frac{d \ln L}{da} \right\rangle \quad (13)$$

$$\text{If } \hat{a} \text{ is unbiased } \langle \hat{a} \rangle = \int \hat{a}(x)P(x; a) dx = \int \hat{a}(x)L(a; x) dx = a$$

$$\text{Differentiate wrt } a: \quad 1 = \int \hat{a}(x) \frac{dL}{da} dx = \int \hat{a}L \frac{d \ln L}{da} dx$$

$$\text{Subtract Eq. 13 multiplied by } a, \text{ and get } \int (\hat{a} - a) \frac{d \ln L}{da} L dx = 1$$

$$\text{Invoke the Schwarz inequality } \int u^2 dx \int v^2 dx \geq \left( \int uv dx \right)^2 \text{ with } u \equiv (\hat{a} - a)\sqrt{L}, v \equiv \frac{d \ln L}{da} \sqrt{L}$$

$$\text{Hence } \int (\hat{a} - a)^2 L dx \int \left( \frac{d \ln L}{da} \right)^2 L dx \geq 1$$

$$\left\langle (\hat{a} - a)^2 \right\rangle \geq 1 / \left\langle \left( \frac{d \ln L}{da} \right)^2 \right\rangle \quad (14)$$

□

Differentiating Eq. 13 again gives

$$\frac{d}{da} \int L \frac{d \ln L}{da} dx = \int \frac{dL}{da} \frac{d \ln L}{da} dx + \int L \frac{d^2 \ln L}{da^2} dx = \left\langle \left( \frac{d \ln L}{da} \right)^2 \right\rangle + \left\langle \frac{d^2 \ln L}{da^2} \right\rangle = 0,$$

$$\text{hence } \left\langle \left( \frac{d \ln L}{da} \right)^2 \right\rangle = - \left\langle \frac{d^2 \ln L}{da^2} \right\rangle.$$

This is the *Fisher information* referred to in Section 2.6.4. Note how it is intrinsically positive.

## 5.3 Maximum likelihood estimation

The *maximum likelihood* (ML) estimator just does what it says:  $a$  is adjusted to maximise the likelihood of the sample (for practical reasons one actually maximises the log likelihood, which is a sum rather than a product).

$$\text{Maximise } \ln L = \sum_i \ln P(x_i; a) \quad , \quad (15)$$

$$\left. \frac{d \ln L}{da} \right|_{\hat{a}} = 0 \quad . \quad (16)$$

The ML estimator is very commonly used. It is not only simple and intuitive, it has lots of nice properties.

- It is consistent.
- It is biased, but bias falls like  $1/N$ .
- It is efficient for the large  $N$ .
- It is invariant—doesn't matter if you reparametrize  $a$ .

A particular maximisation problem may be solved in 3 ways, depending on the complexity

1. Solve Eq. 16 algebraically,
2. Solve Eq. 16 numerically, and
3. Solve Eq. 15 numerically.

## 5.4 Least squares

*Least squares estimation* follows from maximum likelihood estimation. If you have Gaussian measurements of  $y$  taken at various  $x$  values, with measurement error  $\sigma$ , and a prediction  $y = f(x; a)$  then the Gaussian probability

$$P(y; x, a) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-f(x,a))^2/2\sigma^2}$$

gives the log likelihood

$$\ln L = - \sum \frac{(y_i - f(x_i; a))^2}{2\sigma_i^2} + \text{constants.}$$

To maximise  $\ln L$ , you minimise  $\chi^2 = \sum \frac{(y_i - f(x_i; a))^2}{\sigma_i^2}$ , hence the name ‘least squares’.

Differentiating gives the *normal equations*:  $\sum \frac{(y_i - f(x_i; a))}{\sigma_i^2} f'(x_i; a) = 0$ .

If  $f(x; a)$  is linear in  $a$  then these can be solved exactly. Otherwise an iterative method has to be used.

## 5.5 Straight line fits

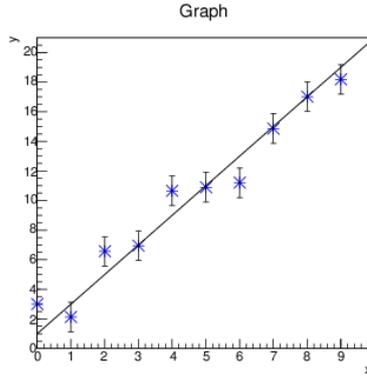
As a particular instance of least squares estimation, suppose the function is  $y = mx + c$ , and assume all  $\sigma_i$  are the same (the extension to the general case is straightforward). The normal equations are then  $\sum (y_i - mx_i - c)x_i = 0$  and  $\sum (y_i - mx_i - c) = 0$ , for which the solution, shown in Fig. 13, is  $m = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}$ ,  $c = \bar{y} - m\bar{x}$ .

Statisticians call this *regression*. Actually there is a subtle difference, as shown in Fig. 14.

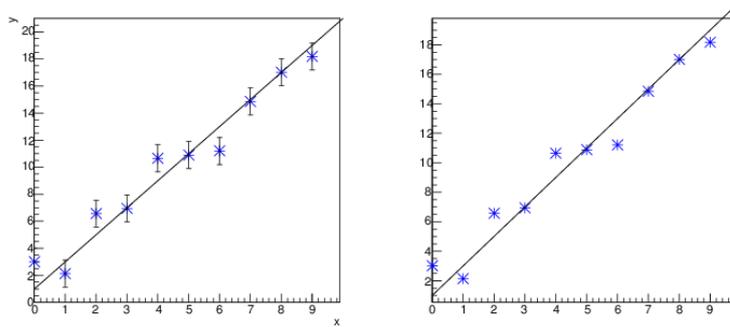
The straight line fit considers well-defined  $x$  values and  $y$  values with measurement errors—if it were not for those errors then presumably the values would line up perfectly, with no scatter. The scatter in regression is not caused by measurement errors, but by the fact that the variables are linked only loosely.

The history of regression started with Galton, who measured the heights of fathers and their (adult) sons. Tall parents tend to have tall children so there is a correlation. Because the height of a son depends not just on his paternal genes but on many factors (maternal genes, diet, childhood illnesses . . .), the points do not line up exactly—and using a high accuracy laser interferometer to do the measurements, rather than a simple ruler, would not change anything.

Galton, incidentally, used this to show that although tall fathers tend to have tall sons, they are not that tall. An outstandingly tall father will have (on average) quite tall children, and only tallish grandchildren. He called this ‘Regression towards mediocrity’, hence the name.



**Fig. 13:** A straight line fit



**Fig. 14:** A straight line fit (left) and linear regression (right)

It is also true that tall sons tend to have tall fathers—but not that tall—and only tallish grandfathers. Regress works in both directions!

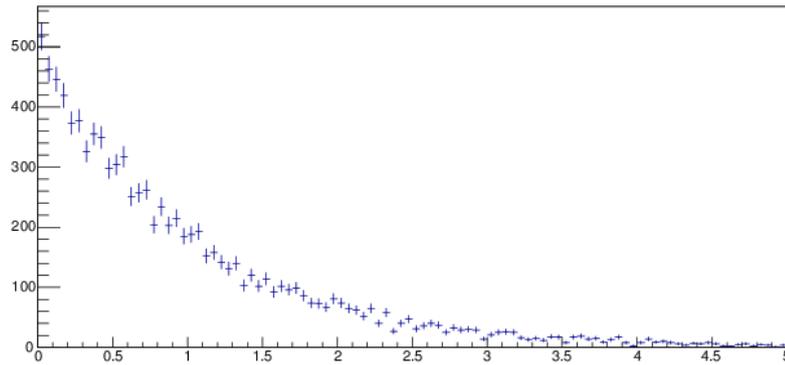
Thus for regression there is always an ambiguity as to whether to plot  $x$  against  $y$  or  $y$  against  $x$ . For a straight line fit as we usually meet them this does not arise: one variable is precisely specified and we call that one  $x$ , and the one with measurement errors is  $y$ .

### 5.6 Fitting histograms

When fitting a histogram the error is given by Poisson statistics for the number of events in each bin.

There are 4 methods of approaching this problem—in order of increasing accuracy and decreasing speed. It is assumed that the bin width  $W$  is narrow, so that  $f(x_i, a) = \int_{x_i}^{x_i+W} P(x, a) dx$  can be approximated by  $f_i(x_i; a) = P(x_i; a) \times W$ .  $W$  is almost always the same for all bins, but the rare cases of variable bin width can easily be included.

1. Minimise  $\chi^2 = \sum_i \frac{(n_i - f_i)^2}{n_i}$ . This is the simplest but clearly breaks if  $n_i = 0$ .
2. Minimise  $\chi^2 = \sum_i \frac{(n_i - f_i)^2}{f_i}$ . Minimising the Pearson  $\chi^2$  (which is valid here) avoids the division-by-zero problem. It assumes that the Poisson distribution can be approximated by a Gaussian.
3. Maximise  $\ln L = \sum \ln(e^{-f_i} f_i^{n_i} / n_i!) \sim \sum n_i \ln f_i - f_i$ . This, known as *binned maximum likelihood*, remedies that assumption.
4. Ignore bins and maximise the total likelihood. Sums run over  $N_{events}$  not  $N_{bins}$ . So if you have large data samples this is much slower. You have to use it for sparse data, but of course in such cases the sample is small and the time penalty is irrelevant.



**Fig. 15:** Fitting a histogram

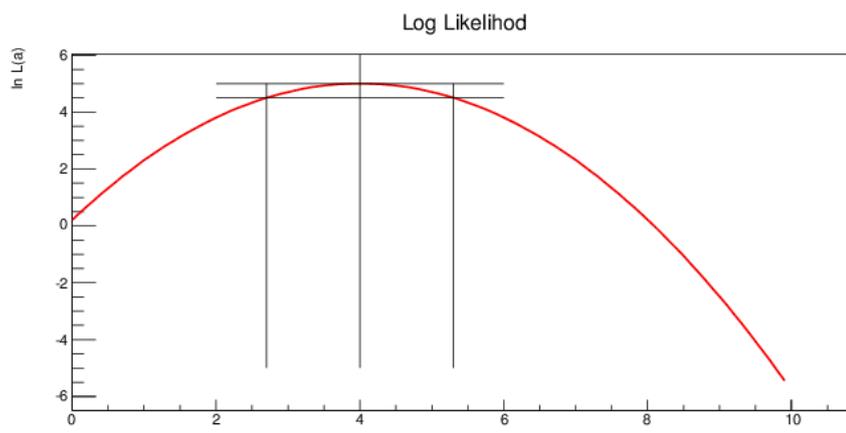
Which method to use is something you have to decide on a case by case basis. If you have bins with zero entries then the first method is ruled out (and removing such bins from the fit introduces bias so this should not be done). Otherwise, in my experience, the improvement in adopting a more complicated method tends to be small.

## 6 Errors

Estimation gives you a value for the parameter(s) that we have called  $a$ . But you also—presumably—want to know something about the uncertainty on that estimate. The maximum likelihood method provides this.

### 6.1 Errors from likelihood

For large  $N$ , the  $\ln L(a, x)$  curve is a parabola, as shown in Fig. 16.



**Fig. 16:** Reading off the error from a Maximum Likelihood fit

At the maximum, a Taylor expansion gives  $\ln L(a) = \ln L(\hat{a}) + \frac{1}{2}(a - \hat{a})^2 \frac{d^2 \ln L}{da^2} \dots$

The maximum likelihood estimator saturates the MVB, so

$$V_{\hat{a}} = -1 / \left\langle \frac{d^2 \ln L}{da^2} \right\rangle \quad \sigma_{\hat{a}} = \sqrt{-\frac{1}{\frac{d^2 \ln L}{da^2}}} \quad . \quad (17)$$

We approximate the expectation value  $\left\langle \frac{d^2 \ln L}{da^2} \right\rangle$  by the actual value in this case  $\frac{d^2 \ln L}{da^2} \Big|_{a=\hat{a}}$  (for a discussion of the introduced inaccuracy, see Ref. [12]).

This can be read off the curve, as also shown in Fig. 16. The maximum gives the estimate. You then draw a line  $\frac{1}{2}$  below that (of course nowadays this is done within the code, not with pencil and ruler, but the visual image is still valid). This line  $\ln L(a) = \ln L(\hat{a}) - \frac{1}{2}$  intersects the likelihood curve at the points  $a = \hat{a} \pm \sigma_{\hat{a}}$ . If you are working with  $\chi^2$ ,  $L \propto e^{-\frac{1}{2}\chi^2}$  so the line is  $\Delta\chi^2 = 1$ .

This gives  $\sigma$ , or 68% errors. You can also take  $\Delta \ln L = -2$  to get 2 sigma or 95% errors, or  $-4.5$  for 3 sigma errors as desired. For large  $N$  these will all be consistent.

## 6.2 Combining errors

Having obtained—by whatever means—errors  $\sigma_x, \sigma_y \dots$  how does one combine them to get errors on derived quantities  $f(x, y \dots), g(x, y, \dots)$ ?

Suppose  $f = Ax + By + C$ , with  $A, B$  and  $C$  constant. Then it is easy to show that

$$\begin{aligned} V_f &= \left\langle (f - \langle f \rangle)^2 \right\rangle \\ &= \left\langle (Ax + By + C - \langle Ax + By + C \rangle)^2 \right\rangle \\ &= A^2 (\langle x^2 \rangle - \langle x \rangle^2) + B^2 (\langle y^2 \rangle - \langle y \rangle^2) + 2AB (\langle xy \rangle - \langle x \rangle \langle y \rangle) \\ &= A^2 V_x + B^2 V_y + 2AB \text{Cov}_{xy} \quad . \end{aligned} \quad (18)$$

If  $f$  is not a simple linear function of  $x$  and  $y$  then one can use a first order Taylor expansion to approximate it about a central value  $f_0(x_0, y_0)$

$$f(x, y) \approx f_0 + \left( \frac{\partial f}{\partial x} \right) (x - x_0) + \left( \frac{\partial f}{\partial y} \right) (y - y_0) \quad (19)$$

and application of Eq. 18 gives

$$V_f = \left( \frac{\partial f}{\partial x} \right)^2 V_x + \left( \frac{\partial f}{\partial y} \right)^2 V_y + 2 \left( \frac{\partial f}{\partial x} \right) \left( \frac{\partial f}{\partial y} \right) \text{Cov}_{xy} \quad (20)$$

writing the more familiar  $\sigma^2$  instead of  $V$  this is equivalent to

$$\sigma_f^2 = \left( \frac{\partial f}{\partial x} \right)^2 \sigma_x^2 + \left( \frac{\partial f}{\partial y} \right)^2 \sigma_y^2 + 2\rho \left( \frac{\partial f}{\partial x} \right) \left( \frac{\partial f}{\partial y} \right) \sigma_x \sigma_y \quad . \quad (21)$$

If  $x$  and  $y$  are independent, which is often but not always the case, this reduces to what is often known as the ‘combination of errors’ formula

$$\sigma_f^2 = \left( \frac{\partial f}{\partial x} \right)^2 \sigma_x^2 + \left( \frac{\partial f}{\partial y} \right)^2 \sigma_y^2 \quad . \quad (22)$$

Extension to more than two variables is trivial: an extra squared term is added for each and an extra covariance term for each of the variables (if any) with which it is correlated.

This can be expressed in language as *errors add in quadrature*. This is a friendly fact, as the result is smaller than you would get from arithmetic addition. If this puzzles you, it may be helpful to think of this as allowing for the possibility that a positive fluctuation in one variable may be cancelled by a negative fluctuation in the other.

There are a couple of special cases we need to consider. If  $f$  is a simple product,  $f = Axy$ , then Eq. 22 gives

$$\sigma_f^2 = (Ay)^2 \sigma_x^2 + (Ax)^2 \sigma_y^2,$$

which, dividing by  $f^2$ , can be written as

$$\left(\frac{\sigma_f}{f}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2. \quad (23)$$

Furthermore this also applies if  $f$  is a simple quotient,  $f = Ax/y$  or  $Ay/x$  or even  $A/(xy)$ .

This is very elegant, but it should not be overemphasised. Equation 23 is not fundamental: it only applies in certain cases (products or quotients). Equation 22 is the fundamental one, and Eq. 23 is just a special case of it.

For example: if you measure the radius of a cylinder as  $r = 123 \pm 2$  mm and the height as  $h = 456 \pm 3$  mm then the volume  $\pi r^2 h$  is  $\pi \times 123^2 \times 456 = 21673295$  mm<sup>3</sup> with error  $\sqrt{(2\pi r h)^2 \times \sigma_r^2 + (\pi r^2)^2 \times \sigma_h^2} = 719101$ , so one could write it as  $v = (216.73 \pm 0.72) \times 10^5$  mm<sup>3</sup>. The surface area  $2\pi r^2 + 2\pi r h$  is  $2\pi \times 123^2 + 2\pi \times 123 \times 456 = 447470$  mm<sup>2</sup> with error  $\sqrt{(4\pi r + 2\pi h)^2 \sigma_r^2 + (2\pi r)^2 \sigma_h^2} = 9121$  mm<sup>2</sup>—so one could write the result as  $a = (447.5 \pm 9.1) \times 10^3$  mm<sup>2</sup>.

A full error analysis has to include the treatment of the covariance terms—if only to show that they can be ignored. Why should the  $x$  and  $y$  in Eq. 20 be correlated? For direct measurements very often (but not always) they will not be. However the interpretation of results is generally a multistage process. From raw numbers of events one computes branching ratios (or cross sections...), from which one computes matrix elements (or particle masses...). Many quantities of interest to theorists are expressed as ratios of experimental numbers. And in this interpretation there is plenty of scope for correlations to creep into the analysis.

For example, an experiment might measure a cross section  $\sigma(pp \rightarrow X)$  from a number of observed events  $N$  in the decay channel  $X \rightarrow \mu^+ \mu^-$ . One would use a formula

$$\sigma = \frac{N}{B\eta\mathcal{L}},$$

where  $\eta$  is the efficiency for detecting and reconstructing an event,  $B$  is the branching ratio for  $X \rightarrow \mu^+ \mu^-$ , and  $\mathcal{L}$  is the integrated luminosity. These will all have errors, and the above prescription can be applied.

However it might also use the  $X \rightarrow e^+ e^-$  channel and then use

$$\sigma' = \frac{N'}{B'\eta'\mathcal{L}}.$$

Now  $\sigma$  and  $\sigma'$  are clearly correlated; even though  $N$  and  $N'$  are independent, the same  $\mathcal{L}$  appears in both. If the estimate of  $\mathcal{L}$  is on the high side, that will push both  $\sigma$  and  $\sigma'$  downwards, and vice versa.

On the other hand, if a second experiment did the same measurement it would have its own  $N$ ,  $\eta$  and  $\mathcal{L}$ , but would be correlated with the first through using the same branching ratio (taken, presumably, from the Particle Data Group).

To calculate correlations between results we need the equivalent of Eq. 18

$$\begin{aligned}\text{Cov}_{fg} &= \langle (f - \langle f \rangle)(g - \langle g \rangle) \rangle \\ &= \left( \frac{\partial f}{\partial x} \right) \left( \frac{\partial g}{\partial x} \right) \sigma_x^2 \quad ,\end{aligned}\quad (24)$$

This can all be combined in the general formula which encapsulates all of the ones above

$$\mathbf{V}_f = \mathbf{G} \mathbf{V}_x \tilde{\mathbf{G}} \quad ,\quad (25)$$

where  $\mathbf{V}_x$  is the covariance matrix of the primary quantities (often, as pointed out earlier, this is diagonal),  $\mathbf{V}_f$  is the covariance matrix of secondary quantities, and

$$G_{ij} = \frac{\partial f_i}{\partial x_j} \quad .\quad (26)$$

The  $\mathbf{G}$  matrix is rectangular but need not be square. There may be more—or fewer—derived quantities than primary quantities. The matrix algebra of  $\mathbf{G}$  and its transpose  $\tilde{\mathbf{G}}$  ensures that the numbers of rows and columns match for Eq. 25.

To show how this works, we go back to our earlier example of a cylinder.  $v$  and  $a$  are correlated: if  $r$  or  $h$  fluctuate upwards (or downwards), that makes both volume and area larger (or smaller). The matrix  $\mathbf{G}$  is

$$\mathbf{G} = \begin{pmatrix} 2\pi r h & \pi r^2 \\ 2\pi(2r + h) & 2\pi r \end{pmatrix} = \begin{pmatrix} 352411 & 47529 \\ 4411 & 773 \end{pmatrix} \quad ,\quad (27)$$

the variance matrix  $V_x$  is

$$\mathbf{V}_x = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix}$$

and Eq. 25 gives

$$\mathbf{V}_f = \begin{pmatrix} 517.1 \times 10^9 & 6.548 \times 10^9 \\ 6.548 \times 10^9 & 83.20 \times 10^6 \end{pmatrix}$$

from which one obtains, as before,  $\sigma_v = 719101$ ,  $\sigma_a = 9121$  but also  $\rho = 0.9983$ .

This can be used to provide a useful example of why correlation matters. Suppose you want to know the volume to surface ratio,  $z = v/a$ , of this cylinder. Division gives  $z = 21673295/447470 = 48.4352$  mm.

If we just use Eq. 22 for the error, this gives  $\sigma_z = 1.89$  mm. Including the correlation term, as in Eq. 21, reduces this to 0.62 mm—three times smaller. It makes a big difference.

We can also check that this is correct, because the ration  $\frac{v}{a}$  can be written as  $\frac{\pi r^2 h}{2\pi r^2 + 2\pi r h}$ , and applying the uncorrelated errors of the original  $r$  and  $h$  to this also gives an error of 0.62 mm.

As a second, hopefully helpful, example we consider a simple straight line fit,  $y = mx + c$ . Assuming that all the  $N$   $y$  values are measured with the same error  $\sigma$ , least squares estimation gives the well known results

$$m = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} \quad c = \frac{\bar{y}\overline{x^2} - \bar{x}\bar{y}\bar{x}}{\overline{x^2} - \bar{x}^2} \quad .\quad (28)$$

For simplicity we write  $D = 1/(\overline{x^2} - \bar{x}^2)$ . The differentials are

$$\frac{\partial m}{\partial y_i} = \frac{D}{N}(x_i - \bar{x}) \quad \frac{\partial c}{\partial y_i} = \frac{D}{N}(\overline{x^2} - x_i \bar{x}) \quad ,$$

from which, remembering that the  $y$  values are uncorrelated,

$$\begin{aligned} V_m &= \sigma^2 \left(\frac{D}{N}\right)^2 \sum (x_i - \bar{x})^2 = \sigma^2 \frac{D}{N} \\ V_c &= \sigma^2 \left(\frac{D}{N}\right)^2 \sum (\overline{x^2} - x_i \bar{x})^2 = \sigma^2 \overline{x^2} \frac{D}{N} \\ \text{Cov}_{mc} &= \sigma^2 \left(\frac{D}{N}\right)^2 \sum (x_i - \bar{x})(\overline{x^2} - x_i \bar{x}) = -\sigma^2 \bar{x} \frac{D}{N} \end{aligned}$$

from which the correlation between  $m$  and  $c$  is just  $\rho = -\bar{x}/\sqrt{\overline{x^2}}$ .

This makes sense. Imagine you're fitting a straight line through a set of points with a range of positive  $x$  values (so  $\bar{x}$  is positive). If the rightmost point happened to be a bit higher, that would push the slope  $m$  up and the intercept  $c$  down. Likewise if the leftmost point happened to be too high that would push the slope down and the intercept up. There is a negative correlation between the two fitted quantities.

Does it matter? Sometimes. Not if you're just interested in the slope—or the constant. But suppose you intend to use them to find the expected value of  $y$  at some extrapolated  $x$ . Equation 21 gives

$$y = mx + c \pm \sqrt{x^2 \sigma_m^2 + \sigma_c^2 + 2x\rho\sigma_m\sigma_c}$$

and if, for a typical case where  $\bar{x}$  is positive so  $\rho$  is negative, you leave out the correlation term you will overestimate your error.

This is an educational example because this correlation can be avoided. Shifting to a co-ordinate system in which  $\bar{x}$  is zero ensures that the quantities are uncorrelated. This is equivalent to rewriting the well-known  $y = mx + c$  formula as  $y = m(x - \bar{x}) + c'$ , where  $m$  is the same as before and  $c' = c + m\bar{x}$ .  $m$  and  $c'$  are now uncorrelated, and error calculations involving them become a lot simpler.

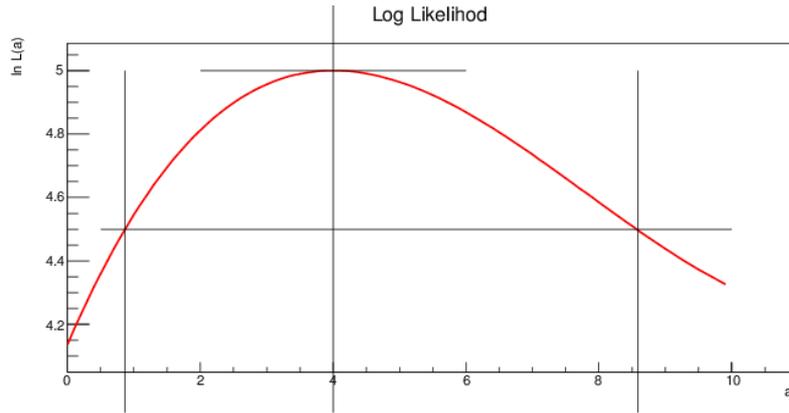
### 6.3 Asymmetric errors

So what happens if you plot the likelihood function and it is not symmetric like Fig. 16 but looks more like Fig. 17? This arises in many cases when numbers are small. For instance, in a simple Poisson count suppose you observe one event.  $P(1; \lambda) = \lambda e^{-\lambda}$  is not symmetric:  $\lambda = 1.5$  is more likely to fluctuate down to 1 than  $\lambda = 0.5$  is to fluctuate up to 1.

You can read off  $\sigma_+$  and  $\sigma_-$  from the two  $\Delta \ln L = -\frac{1}{2}$  crossings, but they are different. The result can then be given as  $a_{-\sigma_-}^{+\sigma_+}$ . What happens after that?

The first advice is to avoid this if possible. If you get  $\hat{a} = 4.56$  with  $\sigma_+ = 1.61$ ,  $\sigma_- = 1.59$  then quote this as  $4.6 \pm 1.6$  rather than  $4.56_{-1.59}^{+1.61}$ . Those extra significant digits have no real meaning. If you can convince yourself that the difference between  $\sigma_+$  and  $\sigma_-$  is small enough to be ignored then you should do so, as the alternative brings in a whole lot of trouble and it's not worth it.

But there will be some cases where the difference is too great to be swept away, so let's consider that case. There are two problems that arise: combination of measurements and combination of errors.

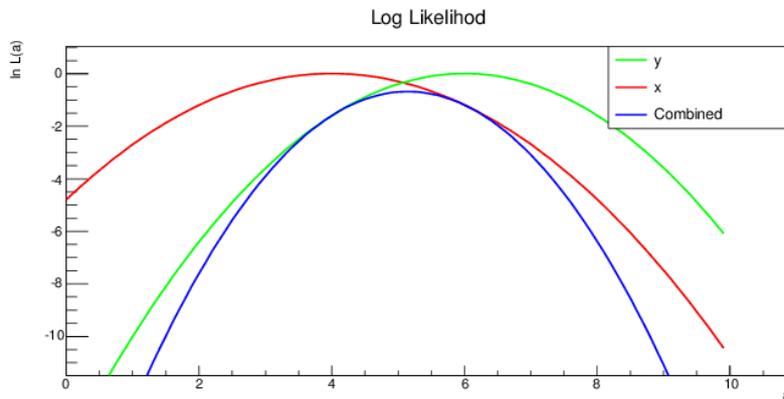


**Fig. 17:** An asymmetric likelihood curve

**6.3.1** *Combination of measurements with asymmetric errors*

Suppose you have two measurements of the same parameter  $a$ :  $\hat{a}_1^{+\sigma_1}$  and  $\hat{a}_2^{+\sigma_2}$  and you want to combine them to give the best estimate and, of course, its error. For symmetric errors the answer is well established to be  $\hat{a} = \frac{\hat{a}_1/\sigma_1^2 + \hat{a}_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}$ .

If you know the likelihood functions, you can do it. The joint likelihood is just the sum. This is shown in Fig. 18 where the red and green curves are measurements of  $a$ . The log likelihood functions just add (blue), from which the peak is found and the  $\Delta \ln L = -\frac{1}{2}$  errors read off.



**Fig. 18:** Combination of two likelihood functions (red and green) to give the total (blue)

But you don't know the full likelihood function: just 3 points (and that it had a maximum at the second). There are, of course, an infinite number of curves that could be drawn, and several models have been tried (cubics, constrained quartic...) on likely instances—see Ref. [13] for details. Some do better than others. The two most plausible are

$$\ln L = -\frac{1}{2} \left( \frac{a - \hat{a}}{\sigma + \sigma'(a - \hat{a})} \right)^2 \quad \text{and} \quad (29)$$

$$\ln L = -\frac{1}{2} \frac{(a - \hat{a})^2}{V + V'(a - \hat{a})} \quad . \quad (30)$$

These are similar to the Gaussian parabola, but the denominator is not constant. It varies with the value of  $a$ , being linear either in the standard deviation or in the variance. Both are pretty good. The first does better with errors on  $\log a$  (which are asymmetric if  $a$  is symmetric: such asymmetric error bars are often seen on plots where the  $y$  axis is logarithmic), the second does better with Poisson measurements.

From the 3 numbers given one readily obtains

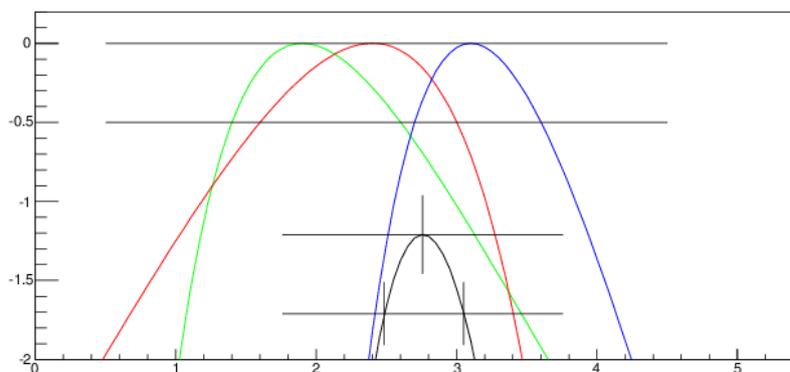
$$\sigma = \frac{2\sigma^+\sigma^-}{\sigma^+ + \sigma^-} \quad \sigma' = \frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-} \quad (31)$$

or, if preferred

$$V = \sigma^+\sigma^- \quad V' = \sigma^+ - \sigma^- \quad . \quad (32)$$

From the total likelihood you then find the maximum of sum, numerically, and the  $\Delta \ln L = -\frac{1}{2}$  points.

Code for doing this is available on GitHub<sup>4</sup> in both R and Root.



**Fig. 19:** Combining three asymmetric measurements

An example is shown in Fig. 19. Combining  $1.9_{-0.5}^{+0.7}$ ,  $2.4_{-0.8}^{+0.6}$  and  $3.1_{-0.4}^{+0.5}$  gives  $2.76_{-0.27}^{+0.29}$ .

### 6.3.2 Combination of errors for asymmetric errors

For symmetric errors, given  $x \pm \sigma_x, y \pm \sigma_y$ , (and  $\rho_{xy} = 0$ ) the error on  $f(x, y)$  is the sum in quadrature:  $\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2$ . What is the equivalent for the error on  $f(x, y)$  when the errors are asymmetric,  $x_{-\sigma_x}^{+\sigma_x}, y_{-\sigma_y}^{+\sigma_y}$ ? Such a problem arises frequently at the end of an analysis when the systematic errors from various sources are all combined.

The standard procedure—which you will see done, though it has not, to my knowledge, been written down anywhere—is to add the positive and negative errors in quadrature separately:  $\sigma_f^{+2} = \sigma_x^{+2} + \sigma_y^{+2}$ ,  $\sigma_f^{-2} = \sigma_x^{-2} + \sigma_y^{-2}$ . This looks plausible, but it is *manifestly wrong* as it breaks the central limit theorem.

To see this, suppose you have to average  $N$  i.i.d. variables each with the same errors which are asymmetric:  $\sigma^+ = 2\sigma^-$ . The standard procedure reduces both  $\sigma^+$  and  $\sigma^-$  by a factor  $1/\sqrt{N}$ , but the skewness remains. The positive error is twice the negative error. This is therefore not Gaussian, and never will be, even as  $N \rightarrow \infty$ .

<sup>4</sup><https://github.com/RogerJBarlow/Asymmetric-Errors>

You can see what’s happening by considering the combination of two of these measurements. They both may fluctuate upwards, or they both may fluctuate downwards, and yes, the upward fluctuation will be, on average, twice as big. But there is a 50% chance of one upward and one downward fluctuation, which is not considered in the standard procedure.

For simplicity we write  $z_i = \frac{\partial f}{\partial x_i}(x_i - x_i^0)$ , the deviation of the parameter from its nominal value, scaled by the differential. The individual likelihoods are again parametrized as Gaussian with a linear dependence of the standard deviation or of the variance, giving

$$\ln L(\vec{z}) = -\frac{1}{2} \sum_i \left( \frac{z_i}{\sigma_i + \sigma'_i z_i} \right)^2 \quad \text{or} \quad -\frac{1}{2} \sum_i \frac{z_i^2}{V_i + V'_i z_i} \quad , \quad (33)$$

where  $\sigma, \sigma', V, V'$  are obtained from Eqs. 31 or 32.

The  $z_i$  are nuisance parameters (as described later) and can be removed by profiling. Let  $u = \sum z_i$  be the total deviation in the quoted  $f$  arising from the individual deviations. We form  $\hat{L}(u)$  as the maximum of  $L(\vec{z})$  subject to the constraint  $\sum_i z_i = u$ . The method of undetermined multipliers readily gives the solution

$$z_i = u \frac{w_i}{\sum_j w_j} \quad , \quad (34)$$

where

$$w_i = \frac{(\sigma_i + \sigma'_i z_i)^3}{2\sigma_i} \quad \text{or} \quad \frac{(V_i + V'_i z_i)^2}{2V_i + V'_i z_i} \quad . \quad (35)$$

The equations are nonlinear, but can be solved iteratively. At  $u = 0$  all the  $z_i$  are zero. Increasing (or decreasing)  $u$  in small steps, Eqs. 34 and 35 are applied successively to give the  $z_i$  and the  $w_i$ : convergence is rapid. The value of  $u$  which maximises the likelihood should in principle be applied as a correction to the quoted result.

Programs to do this are also available on the GitHub site.

As an example, consider a counting experiment with a number of backgrounds, each determined by an ancillary Poisson experiment, and that for simplicity each background was determined by running the apparatus for the same time as the actual experiment. (In practice this is unlikely, but scale factors can easily be added.)

Suppose two backgrounds are measured, one giving four events and the other five. These would be reported, using  $\Delta \ln L = -\frac{1}{2}$  errors, as  $4^{+2.346}_{-1.682}$  and  $5^{+2.581}_{-1.916}$ . The method, using linear  $V$ , gives the combined error on the background count as  $^{+3.333}_{-2.668}$ .

In this simple case we can check the result against the total background count of nine events, which has errors  $^{+3.342}_{-2.676}$ . The agreement is impressive. Further examples of the same total, partitioned differently, are shown in table 1.

Inputs	Linear $\sigma$		Linear $V$	
	$\sigma^-$	$\sigma^+$	$\sigma^-$	$\sigma^+$
4+5	2.653	3.310	2.668	3.333
3+6	2.653	3.310	2.668	3.333
2+7	2.653	3.310	2.668	3.333
2+7	2.653	3.310	2.668	3.333
3+3+3	2.630	3.278	2.659	3.323
1+1+1+1+1+1+1+1	2.500	3.098	2.610	3.270

**Table 1:** Various combinations of Poisson errors. The target value is  $\sigma^- = 2.676, \sigma^+ = 3.342$

## 6.4 Errors in 2 or more dimensions

For 2 (or more) dimensions, one plots the log likelihood and defines regions using contours in  $\Delta \ln L$  (or  $\Delta\chi^2 \equiv -2\Delta \ln L$ ). An example is given in Fig. 20.

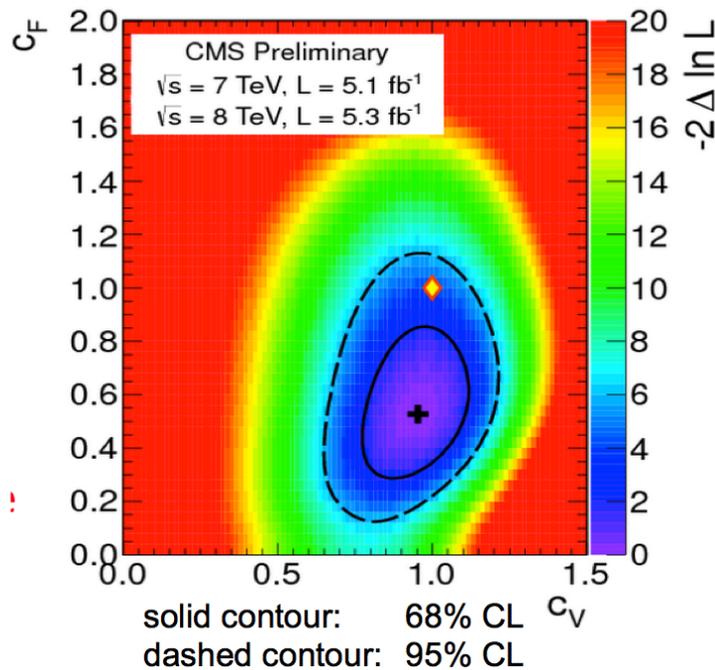


Fig. 20: CMS results on  $C_V$  and  $C_F$ , taken from Ref. [14]

The link between the  $\Delta \ln L$  values and the significance changes. In 1D, there is a 68% probability of a measurement falling within  $1\sigma$ . In 2D, a  $1\sigma$  square would give a probability  $0.68^2 = 47\%$ . If one rounds off the corners and draws a  $1\sigma$  contour at  $\Delta \ln L = -\frac{1}{2}$  this falls to 39%. To retrieve the full 68% one has to draw a contour at  $\Delta \ln L = -1.14$ , or equivalently  $\Delta\chi^2 = 2.27$ . For 95% use  $\Delta\chi^2 = 5.99$  or  $\Delta \ln L = -3.00$ .

The necessary value is obtained from the  $\chi^2$  distribution—described later. It can be found by the R function `qchisq(p,n)` or the Root function `TMath::ChiSquareQuantile(p,n)`, where the desired probability  $p$  and number of degrees of freedom  $n$  are the arguments given.

### 6.4.1 Nuisance parameters

In the example of Fig. 20, both  $C_V$  and  $C_F$  are interesting. But in many cases one is interested only in one (or some) of the quantities and the others are ‘nuisance parameters’ that one would like to remove, reducing the dimensionality of the quoted result. There are two methods of doing this, one (basically) frequentist and one Bayesian.

The frequentist uses the *profile likelihood* technique. Suppose that there are two parameters,  $a_1$  and  $a_2$ , where  $a_2$  is a nuisance parameter, and so one wants to reduce the joint likelihood function  $L(x; a_1, a_2)$  to some function  $\hat{L}(a_1)$ . To do this one scans across the values of  $a_1$  and inserts  $\hat{a}_2(a_1)$ , the value of  $a_2$  which maximises the likelihood for that particular  $a_1$

$$\hat{L}(x, a_1) = L(a_1, \hat{a}_2(a_1)) \quad (36)$$

and the location of the maximum and the  $\Delta \ln L = \frac{1}{2}$  errors are read off as usual.

To see why this works—though this is not a very rigorous motivation—suppose one had a likelihood function as shown in Fig. 21.

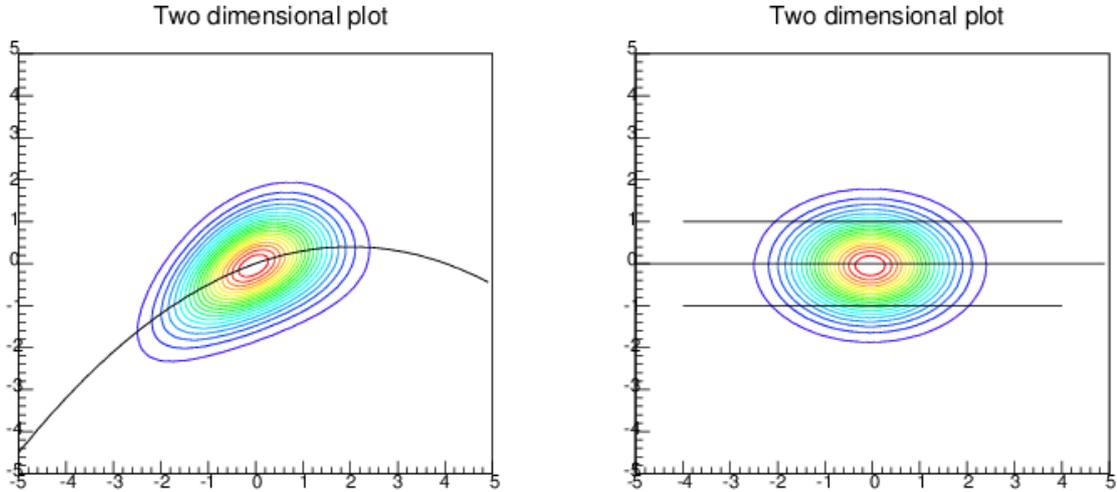


Fig. 21: Justification of the likelihood profile method

The horizontal axis is for the parameter of interest,  $a_1$ , and the vertical for the nuisance parameter  $a_2$ .

Different values of  $a_2$  give different results (central and errors) for  $a_1$ .

If it is possible to transform to  $a'_2(a_1, a_2)$  so that  $L$  factorises, then we can write  $L(a_1, a'_2) = L_1(a_1)L_2(a'_2)$ : this is shown in the plot on the right. We suppose that this is indeed possible. In the case here, and other not-too-complicated cases, it clearly is, although it will not be so in more complicated topologies with multiple peaks.

Then using the transformed graph, whatever the value of  $a'_2$ , one would get the same result for  $a_1$ . Then one can present this result for  $a_1$ , independent of anything about  $a'_2$ .

There is no need to factorise explicitly: the path of central  $a'_2$  value as a function of  $a_1$  (the central of the 3 lines on the right hand plot) is the path of the peak, and that path can be located in the first plot (the transformation only stretches the  $a_2$  axis, it does not change the heights).

The Bayesian method uses the technique called *marginalisation*, which just integrates over  $a_2$ . Frequentists can not do this as they are not allowed to integrate likelihoods over the parameter:  $\int P(x; a) dx$  is fine, but  $\int P(x; a) da$  is off limits. Nevertheless this can be a very helpful alternative to profiling, specially for many nuisance parameters. But if you use it you must be aware that this is strictly Bayesian. Reparametrizing  $a_2$  (or choosing a different prior) will give different results for  $a_1$ . In many cases, where the effect of the nuisance parameter is small, this does not have a big effect on the result.

### 6.5 Systematic errors

This can be a touchy subject. There is a lot of bad practice out there. Muddled thinking and following traditional procedures without understanding. When statistical errors dominated, this didn't matter much. In the days of particle factories and big data samples, it does.

#### 6.5.1 What is a systematic error?

Consider these two quotations, from eminent and widely-read authorities.

R. Bevington defines

‘Systematic error: reproducible inaccuracy introduced by faulty equipment, calibration, or technique.’ [15],

whereas J. Orear writes

‘Systematic effects is a general category which includes effects such as background, scanning efficiency, energy resolution, variation of counter efficiency with beam position, and energy, dead time, etc. The uncertainty in the estimation of such a systematic effect is called a systematic error.’ [16].

Read these carefully and you will see that they are contradictory. They are not talking about the same thing. Furthermore, Orear is RIGHT and Bevington is WRONG—as are a lot of other books and websites.

We teach undergraduates the difference between measurement *errors*, which are part of doing science, and *mistakes*. They are not the same. If you measure a potential of 12.3 V as 12.4 V, with a voltmeter accurate to 0.1V, that is fine. Even if you measure 12.5 V. If you measure it as 124 V, that is a mistake.

In the quotes above, Bevington is describing *systematic mistakes* (the word ‘faulty’ is the key) whereas Orear is describing *systematic uncertainties*—which are ‘errors’ in the way we use the term.

There is a case for saying one should avoid the term ‘systematic error’ and always use ‘uncertainty’ or ‘mistake’. This is probably impossible. But you should always know which you mean.

Restricting ourselves to uncertainties (we will come back to mistakes later) here are some typical examples:

- Track momenta from  $p_i = 0.3B\rho_i$  have statistical errors from  $\rho$  and systematic errors from  $B$ ,
- Calorimeter energies from  $E_i = \alpha D_i + \beta$  have statistical errors from the digitised light signal  $D_i$  and systematic errors from the calibration  $\alpha, \beta$ , and
- Branching ratios from  $Br = \frac{N_D - B}{\eta N_T}$  have statistical errors from  $N_D$  and systematic errors from efficiency  $\eta$ , background  $B$ , total  $N_T$ .

Systematic uncertainties can be either Bayesian or Frequentist. There are clearly frequentist cases where errors have been determined by an *ancillary experiment* (real or simulated), such as magnetic field measurements, calorimeter calibration in a testbeam, and efficiencies from Monte Carlo simulations. (Sometimes the ancillary experiment is also the main experiment—e.g. in estimating background from sidebands.) There are also uncertainties that can only be Bayesian, e.g. when a theorist tells you that their calculation is good to 5% (or whatever) or an experimentalist affirms that the calibration will not have shifted during the run by more than 2% (or whatever).

### 6.5.2 How to handle them: correlations

Working with systematic errors is actually quite straightforward. They obey the same rules as statistical uncertainties.

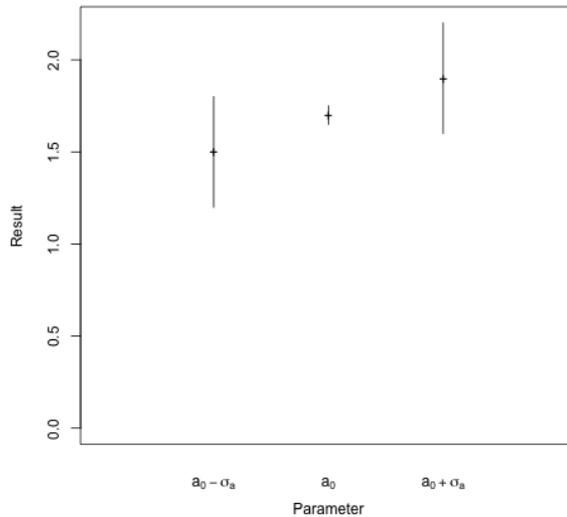
We write  $x = 12.2 \pm 0.3 \pm 0.4$  ‘where the first error is statistical and the second is systematic’, but it would be valid to write  $x = 12.2 \pm 0.5$ . For single measurement the extra information given by the two separate numbers is small. (In this case it just tells you that there is little to be gained by increasing the size of the data sample). For multiple measurements e.g.  $x_a = 12.2 \pm 0.3$ ,  $x_b = 17.1 \pm 0.4$ , *all*  $\pm 0.5$  the extra information is important, as results are correlated. Such cases arise, for example, in cross section measurements with a common luminosity error, or branching ratios with common efficiency.

Such a correlation means that taking more measurements and averaging does not reduce the error. Also there is no way to estimate  $\sigma_{sys}$  from the data—hence no check on the goodness of fit from a  $\chi^2$  test.

**6.5.3 Handling systematic errors in your analysis**

It is useful to consider systematic errors as having three types:

1. Uncertainty in an explicit continuous parameter. For example an uncertainty in efficiency, background and luminosity in determining a branching ratio or cross section. For these the standard combination of errors formula and algebra are usable, just like undergraduate labs.
2. Uncertainty in an implicit continuous parameter. For example: MC tuning parameters ( $\sigma_{p_T}$ , polarisation . . .). These are not amenable to algebra. Instead one calculates the result for different parameter values, typically at  $\pm\sigma$ , and observes the variation in the result, as illustrated in Fig. 22.



**Fig. 22:** Evaluating the effect of an implicit systematic uncertainty

Hopefully the effect is equal but opposite—if not then one can reluctantly quote an asymmetric error. Also your analysis results will have errors due to finite MC statistics. Some people add these in quadrature. This is wrong. The technically correct thing to do is to subtract them in quadrature, but this is not advised.

3. Discrete uncertainties:

These typically occur in model choices. Using a different Monte Carlo for background—or signal—gives you a (slightly) different result. How do you include this uncertainty? The situation depends on the status of the models. Sometimes one is preferred, sometimes they are all equal (more or less).

With 1 preferred model and one other, quote  $R_1 \pm |R_1 - R_2|$ .

With 2 models of equal status, quote  $\frac{R_1+R_2}{2} \pm \left| \frac{R_1-R_2}{\sqrt{2}} \right|$ .

With N models: take  $\bar{R} \pm \sqrt{\frac{N}{N-1}(\bar{R}^2 - \overline{R^2})}$  or similar mean value.

2 extreme models: take  $\frac{R_1+R_2}{2} \pm \frac{|R_1-R_2|}{\sqrt{12}}$ .

These are just ballpark estimates. Do not push them too hard. If the difference is not small, you have a problem—which can be an opportunity to study model differences.

**6.5.4 Checking the analysis**

*“As we know, there are known knowns. There are things we know that we know. There are known*

*unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know."*

Donald H. Rumsfeld

Errors are not mistakes—but mistakes still happen. Statistical tools can help find them. Check your result by repeating the analysis with changes which *should* make no difference:

- Data subsets,
- Magnet up/down,
- Different selection cuts,
- Changing histogram bin size and fit ranges,
- Changing parametrization (including order of polynomial),
- Changing fit technique,
- Looking for impossibilities,
- ...

The more tests the better. You cannot prove the analysis is correct. But the more tests it survives the more likely your colleagues<sup>5</sup> will be to believe the result.

For example: in the paper reporting the first measurement of CP violation in  $B$  mesons the BaBar Collaboration [17] reported

‘... consistency checks, including separation of the decay by decay mode, tagging category and  $B_{tag}$  flavour ... We also fit the samples of non-CP decay modes for  $\sin 2\beta$  with no statistically significant difference found.’

If your analysis passes a test then *tick the box and move on*. Do not add the discrepancy to the systematic error. Many people do—and your supervisor and your review committee may want you to do so. Do not give in.

- It's illogical,
- It penalises diligence, and
- Errors get inflated.

If your analysis fails a test then worry!

- Check the test. Very often this turns out to be faulty.
- Check the analysis. Find mistake, enjoy improvement.
- Worry. Consider whether the effect might be real. (E.g. June's results are different from July's. Temperature effect? If so can (i) compensate and (ii) introduce implicit systematic uncertainty).
- Worry harder. Ask colleagues, look at other experiments.

Only as a last resort, add the term to the systematic error. Remember that this could be a hint of something much bigger and nastier.

---

<sup>5</sup>and eventually even you

### 6.5.5 *Clearing up a possible confusion*

What’s the difference between?

Evaluating implicit systematic errors: vary lots of parameters, see what happens to the result, and include in systematic error.

Checks: vary lots of parameters, see what happens to the result, and don’t include in systematic error.

If you find yourself in such a situation there are actually two ways to tell the difference.

(1) Are you expecting to see an effect? If so, it’s an evaluation, if not, it’s a check.

(2) Do you clearly know how much to vary them by? If so, it’s an evaluation. If not, it’s a check.

These cover even complicated cases such as a trigger energy cut where the energy calibration is uncertain—and it may be simpler to simulate the effect by varying the cut rather than the calibration.

### 6.5.6 *So finally:*

1. Thou shalt never say ‘systematic error’ when thou meanest ‘systematic effect’ or ‘systematic mistake’.
2. Thou shalt know at all times whether what thou performest is a check for a mistake or an evaluation of an uncertainty.
3. Thou shalt not incorporate successful check results into thy total systematic error and make thereby a shield to hide thy dodgy result.
4. Thou shalt not incorporate failed check results unless thou art truly at thy wits’ end.
5. Thou shalt not add uncertainties on uncertainties in quadrature. If they are larger than chickenfeed thou shalt generate more Monte Carlo until they shrink.
6. Thou shalt say what thou doest, and thou shalt be able to justify it out of thine own mouth; not the mouth of thy supervisor, nor thy colleague who did the analysis last time, nor thy local statistics guru, nor thy mate down the pub.

Do these, and thou shalt flourish, and thine analysis likewise.

## 7 Goodness of fit

You have the best fit model to your data—but is it good enough? The upper plot in Fig. 23 shows the best straight line through a set of points which are clearly not well described by a straight line. How can one quantify this?

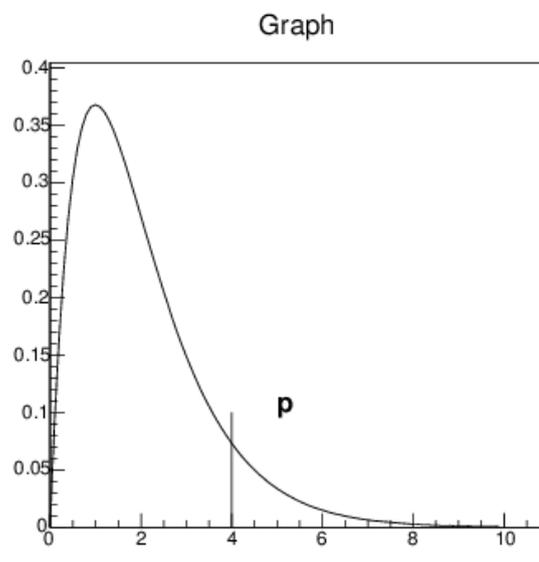
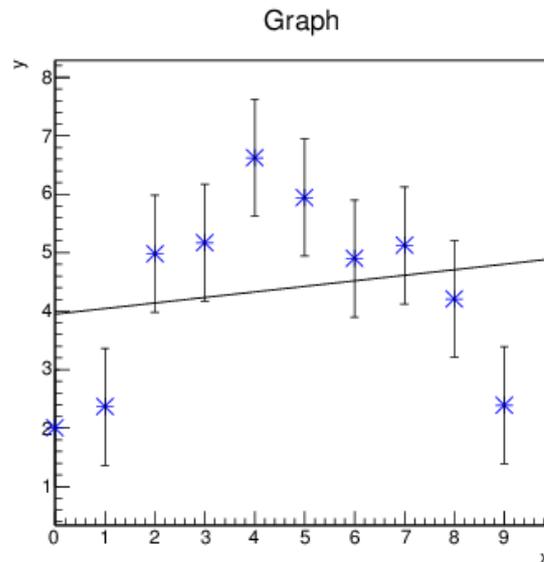
You construct some measure of agreement—call it  $t$ —between the model and the data. Convention:  $t \geq 0$ ,  $t = 0$  is perfect agreement. Worse agreement implies larger  $t$ . The null hypothesis  $H_0$  is that the model did indeed produce this data. You calculate the  $p$ -value: the probability under  $H_0$  of getting a  $t$  this bad, or worse. This is shown schematically in the lower plot. Usually this can be done using known algebra—if not one can use simulation (a so-called ‘Toy Monte Carlo’).

### 7.1 The $\chi^2$ distribution

The overwhelmingly most used such measure of agreement is the quantity  $\chi^2$

$$\chi^2 = \sum_1^N \left( \frac{y_i - f(x_i)}{\sigma_i} \right)^2 . \quad (37)$$

In words: the total of the squared differences between prediction and data, scaled by the expected error. Obviously each term will be about 1, so  $\langle \chi^2 \rangle \approx N$ , and this turns out to be exact.



**Fig. 23:** The best fit to the data may not be good enough

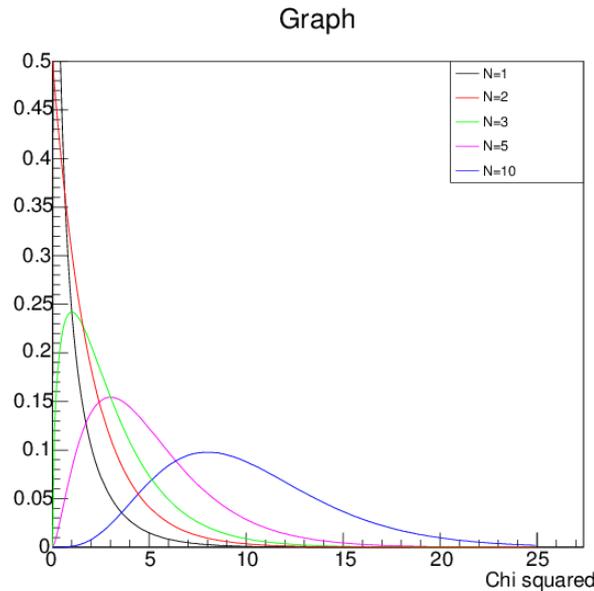
The distribution for  $\chi^2$  is given by

$$P(\chi^2; N) = \frac{1}{2^{N/2}\Gamma(N/2)} \chi^{N-2} e^{-\chi^2/2} \quad (38)$$

shown in Fig. 24, though this is in fact not much used: one is usually interested in the  $p$ -value, the probability (under the null hypothesis) of getting a value of  $\chi^2$  as large as, or larger than, the one observed. This can be found in ROOT with `TMath::Prob(chisquared,ndf)`, and in R from `1-pchisq(chisquared,ndf)`.

Thus for example with  $N = 10$ ,  $\chi^2 = 15$  then  $p = 0.13$ . This is probably OK. But for  $N = 10$ ,  $\chi^2 = 20$  then  $p = 0.03$ , which is probably not OK.

If the model has parameters which have been adjusted to fit the data, this clearly reduces  $\chi^2$ . It is a very useful fact that the result also follows a  $\chi^2$  distribution for  $NDF = N_{data} - N_{parameters}$  where  $NDF$  is called the ‘number of degrees of freedom’.



**Fig. 24:** The  $\chi^2$  distribution for various  $N$

If your  $\chi^2$  is suspiciously big, there are 4 possible reasons:

1. Your model is wrong,
2. Your data are wrong,
3. Your errors are too small, or
4. You are unlucky.

If your  $\chi^2$  is suspiciously small there are 2 possible reasons:

1. Your errors are too big, or
2. You are lucky.

## 7.2 Wilks' theorem

The Likelihood on its own tells you *nothing*. Even if you include all the constant factors normally omitted in maximisation. This may seem counter-intuitive, but it is inescapably true.

There is a theorem due to Wilks which is frequently invoked and appears to link likelihood and  $\chi^2$ , but it does so only in very specific circumstances. Given two nested models, for large  $N$  the improvement in  $\ln L$  is distributed like  $\chi^2$  in  $-2\Delta \ln L$ , with  $NDF$  the number of extra parameters.

So suppose you have some data with many  $(x, y)$  values and two models, Model 1 being linear and Model 2 quadratic. You maximise the likelihood using Model 1 and then using Model 2: the Likelihood increases as more parameters are available ( $NDF = 1$ ). If this increase is significantly more than  $N$  that justifies using Model 2 rather than Model 1. So it may tell you whether or not the extra term in a quadratic gives a meaningful improvement, but not whether the final quadratic (or linear) model is a good one.

Even this has an important exception. it does not apply if Model 2 contains a parameter which is meaningless under Model 1. This is a surprisingly common occurrence. Model 1 may be background, Model 2 background plus a Breit-Wigner with adjustable mass, width and normalization ( $NDF = 3$ ). The mass and the width are meaningless under Model 1 so Wilks' theorem does not apply and the improvement in likelihood cannot be translated into a  $\chi^2$  for testing.

### 7.3 Toy Monte Carlos and likelihood for goodness of fit

Although the likelihood contains no information about the goodness of fit of the model, an obvious way to get such information is to run many simulations of the model, plot the spread of fitted likelihoods and use it to get the  $p$ -value.

This may be obvious, but it is wrong [18]. Consider a test case observing decay times where the model is a simple exponential  $P(t) = \frac{1}{\tau} e^{-t/\tau}$ , with  $\tau$  an adjustable parameter. Then you get the Log Likelihood  $\sum(-t_i/\tau - \ln \tau) = -N(\bar{t}/\tau + \ln \tau)$  and maximum likelihood gives  $\hat{t} = \bar{t} = \frac{1}{N} \sum_i t_i$ , so  $\ln L(\hat{t}; x) = -N(1 + \ln \bar{t})$ . This holds whatever the original sample  $\{t_i\}$  looks like: any distribution with the same  $\bar{t}$  has the same likelihood, after fitting.

## 8 Upper limits

Many analyses are ‘searches for...’ and most of these are unsuccessful. But you have to say something! Not just ‘We looked, but we didn’t see anything’. This is done using the construction of frequentist confidence intervals and/or Bayesian credible intervals.

### 8.1 Frequentist confidence

Going back to the discussion of the basics, for frequentists the probability that it will rain tomorrow is meaningless: there is only one tomorrow, it will either rain or it will not, there is no ensemble. The probability  $N_{\text{rain}}/N_{\text{tomorrows}}$  is either 0 or 1. To talk about  $P_{\text{rain}}$  is "unscientific" [10].

This is unhelpful. But there is a workaround.

Suppose some forecast says it will rain and studies show this forecast is correct 90% of the time. We now have an ensemble of statements, and can say: ‘The statement ‘It will rain tomorrow’ has a 90% probability of being true’. We shorten this to ‘It will rain tomorrow, with 90% confidence’. We state X with confidence  $P$  if X is a member of an ensemble of statements of which at least  $P$  are true.

Note the ‘at least’ which has crept into the definition. There are two reasons for it:

1. Higher confidences embrace lower ones. If X at 95% then X at 90%, and
2. We can cater for composite hypotheses which are not completely defined.

The familiar quoted error is in fact a confidence statement. Consider as an illustration the Higgs mass measurement (current at the time of writing)  $M_H = 125.09 \pm 0.24$  GeV. This does not mean that the probability of the Higgs mass being in the range  $124.85 < M_H < 125.33$  GeV is 68%: the Higgs mass is a single, unique, number which either lies in this interval or it does not. What we are saying is that  $M_H$  has been measured to be 125.09 GeV with a technique that will give a value within 0.24 GeV of the true value 68% of the time. We say:  $124.85 < M_H < 125.33$  GeV with 68% confidence. The statement is either true or false (time will tell), but it belongs to a collection of statements of which (at least) 68% are true.

So we construct *confidence regions* also known as confidence intervals  $[x_-, x_+]$  such that  $\int_{x_-}^{x_+} P(x) dx = CL$ . We have not only a choice of the probability content (68%, 90%, 95%, 99%...) to work with but also of strategy. Common options are:

1. Symmetric:  $\hat{x} - x_- = x_+ - \hat{x}$ ,
2. Shortest: Interval that minimises  $x_+ - x_-$ ,
3. Central:  $\int_{-\infty}^{x_-} P(x) dx = \int_{x_+}^{\infty} P(x) dx = \frac{1}{2}(1 - CL)$ ,
4. Upper Limit:  $x_- = -\infty$ ,  $\int_{x_+}^{\infty} P(x) dx = 1 - CL$ , and
5. Lower Limit:  $x_+ = \infty$ ,  $\int_{-\infty}^{x_-} P(x) dx = 1 - CL$ .

For the Gaussian (or any symmetric PDF) 1-3 are the same.

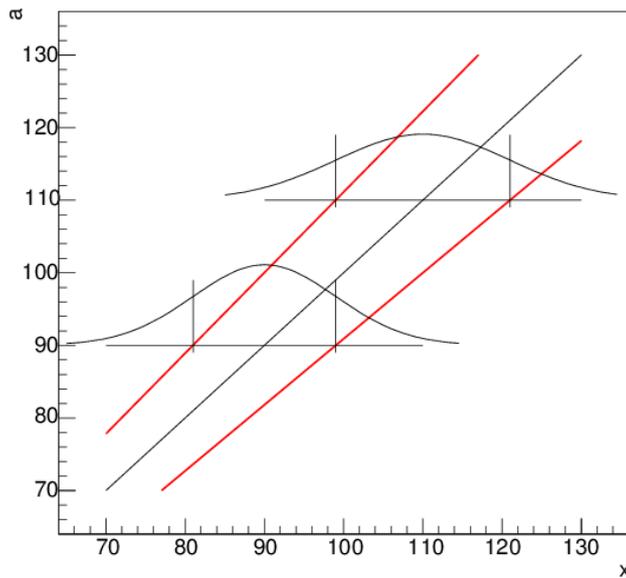
We are particularly concerned with the upper limit: we observe some small value  $x$ . We find a value  $x_+$  such that for values of  $x_+$  or more the probability of getting a result as small as  $x$ , or even less, is  $1 - CL$ , or even less.

### 8.2 Confidence belts

We have shown that a simple Gaussian measurement is basically a statement about confidence regions.  $x = 100 \pm 10$  implies that [90,110] is the 68% central confidence region.

We want to extend this to less simple scenarios. As a first step, we consider a proportional Gaussian. Suppose we measure  $x = 100$  from Gaussian measurement with  $\sigma = 0.1x$  (a 10% measurement—which is realistic). If the true value is 90 the error is  $\sigma = 9$  so  $x = 100$  is more than one standard deviation, whereas if the true value is 110 then  $\sigma = 11$  and it is less than one standard deviation. 90 and 110 are not equidistant from 100.

This is done with a technique called a confidence belt. The key point is that they are constructed horizontally and read vertically, using the following procedure (as shown in Fig. 25). Suppose that  $a$  is the parameter of interest and  $x$  is the measurement.



**Fig. 25:** A confidence belt for a proportional Gaussian

1. For each  $a$ , construct desired confidence interval (here 68% central).
2. The result  $(x, a)$  lies inside the belt (the red lines), with 68% confidence.
3. Measure  $x$ .
4. The result  $(x, a)$  lies inside the belt, with 68% confidence. And now we know  $x$ .
5. Read off the belt limits  $a_+$  and  $a_-$  at that  $x$ : in this case they are 111.1, 90.9. So we can report that  $a$  lies in [90.9,111.1] with 68% confidence.
6. Other choices for the confidence level value and for the strategy are available.

This can be extended to the case of a Poisson distribution, Fig. 26.

The only difference is that the horizontal axis is discrete as the number observed,  $x$ , is integer. In constructing the belt (horizontally) there will not in general be  $x$  values available to give  $\sum_{x_-}^{x_+} = CL$

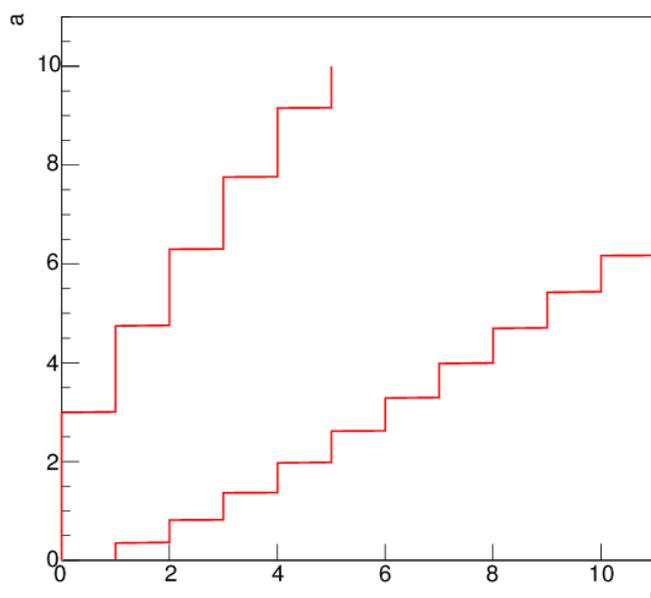


Fig. 26: A confidence belt for a Poisson

and we call, again, on the ‘at least’ in the definition and allow it to be  $\sum_{x_-}^{x_+} \geq CL$ .

Thus for a central 90% confidence we require for each  $a$  the largest integer  $x_{lo}$  and smallest  $x_{hi}$  for which  $\sum_{x=0}^{x_{lo}-1} e^{-a} \frac{a^x}{x!} \leq 0.05$  and  $\sum_{x=x_{hi}+1}^{\infty} e^{-a} \frac{a^x}{x!} \leq 0.05$ . For the second sum it is easier to calculate  $\sum_{x=0}^{x_{hi}} e^{-a} \frac{a^x}{x!} \geq 0.95$ .

Whatever the value of  $a$ , the probability of the result falling in the belt is 90% or more. We proceed as for the Gaussian.

### 8.3 Coverage

This is an appropriate point to introduce *coverage*: the probability, given  $a$ , that the statement ‘ $a_{lo} \leq a \leq a_{hi}$ ’ will be true. Ideally this would be the same as the confidence level, however it may (because of the ‘at least’ clauses) exceed it (‘overcover’); this is allowed though in principle inefficient. It should never be less (‘undercover’).

For example: suppose we have a Poisson process with  $a = 3.5$  and we want a 90% central limit.

There is a probability  $e^{-3.5} = 3\%$  of getting zero events, leading to  $a_+ = 3.0$ , which would be wrong as  $3.0 < 3.5$ .

Continuing in sequence, there is a probability  $3.5e^{-3.5} = 11\%$  of getting one event, leading to  $a_+ = 4.7$ , which would be right.

Right answers continue up to seven events (with probability  $3.5^7 e^{-3.5} / 7! = 4\%$ ): this gives a safely large value for  $a_+$  and  $a_- = 3.3$ , which is right as  $3.3 < 3.5$ , though only just, The next outcome, eight events (probability 2%) gives  $a_- = 4.0$  which is wrong, as are all subsequent results.

Adding up the probabilities for the outcomes 1 thru 7 that give a true answer totals 94%, so there is 4% overcoverage.

Note that coverage is a function of the true value of the parameter on which limits are being placed. Values of  $a$  other than 3.5 will give different coverage numbers—though all are over 90%.

### 8.4 Upper limits

The one-sided upper limit—option 4 in the list above—gives us a way of quantifying the outcome of a null experiment. ‘We saw nothing (or nothing that might not have been background), so we say  $a \leq a_+$  at some confidence level’.

One simple and enlightening example occurs if you see no events, and there is no expected background. Now  $P(0; 2.996) = 0.05$  and  $2.996 \sim 3$ . So if you see zero events, you can say with 95% confidence that the true value is less than 3.0. You can then directly use this to calculate a limit on the branching fraction, cross section, or whatever you’re measuring.

### 8.5 Bayesian ‘credible intervals’

A Bayesian has no problems saying ‘It will probably rain tomorrow’ or ‘The probability that  $124.85 < M_H < 125.33$  GeV is 68%’. The downside, of course, is that another Bayesian can say ‘It will probably not rain tomorrow’ and ‘The probability that  $124.85 < M_H < 125.33$  GeV is 86%’ with equal validity and the two cannot resolve their subjective difference in any objective way.

A Bayesian has a prior belief PDF  $P(a)$  and defines a region  $R$  such that  $\int_R P(a) da = CL$ . There is the same ambiguity regarding choice of content (68%, 90%, 95%...) and strategy (central, symmetric, upper limit...). So Bayesian credible intervals look a lot like frequentist confidence intervals even if their meaning is different.

There are two happy coincidences.

The first is that Bayesian credible intervals on Gaussians, with a flat prior, are the same as Frequentist confidence intervals. If F quotes 68% or 95% or . . . confidence intervals and B quotes 68% or 95% or . . . credible interval, their results will agree.

The second is that although the Frequentist Poisson upper limit is given by  $\sum_{r=0}^{r=r_{data}} e^{-a_{hi}} a_{hi}^r / r!$  whereas the Bayesian Poisson flat prior upper limit is given by  $\int_0^{a_{hi}} e^{-a} a^{r_{data}} / r_{data}! da$ , integration by parts of the Bayesian formula gives a series which is same as the Frequentist limit. A Bayesian will also say : ‘I see zero events—the probability is 95% that the true value is 3.0 or less.’ This is (I think) a coincidence—it does not apply for lower limits. But it does avoid heated discussions as to which value to publish.

### 8.6 Limits in the presence of background

This is where it gets tricky. Typically an experiment may observe  $N_D$  events, with an expected background  $N_B$  and efficiency  $\eta$ , and wants to present results for  $N_S = \frac{N_D - N_B}{\eta}$ . Uncertainties in  $\eta$  and  $N_B$  are handled by profiling or marginalising. The problem is that the *actual number* of background events is not  $N_B$  but Poisson in  $N_B$ .

So in a straightforward case, if you observe twelve events, with expected background 3.4 and  $\eta = 1$  it is obviously sensible to say  $N_S = 8.6$  (though the error is  $\sqrt{12}$  not  $\sqrt{8.6}$ )

But suppose, with the same background, you see four events, three events or zero events. Can you say  $N_S = 0.6?$  or  $-0.4?$  Or  $-3.4???$

We will look at four methods of handling this, considering as an example the observation of three events with expected background 3.40 and wanting to present the 95% CL upper limit on  $N_S$ .

#### 8.6.1 Method 1: Pure frequentist

$N_D - N_B$  is an unbiased estimator of  $N_S$  and its properties are known. Quote the result. Even if it is non-physical.

The argument for doing so is that this is needed for balance: if there is really no signal, approximately half of the experiments will give positive values and half negative. If the negative results are not

published, but the positive ones are, the world average will be spuriously high. For a 95% confidence limit one accepts that 5% of the results can be wrong. This (unlikely) case is clearly one of them. So what?

A counter-argument is that if  $N_D < N_B$ , we *know* that the background has fluctuated downwards. But this cannot be incorporated into the formalism.

Anyway, the upper limit from 3 is 7.75, as  $\sum_0^3 e^{-7.75} 7.75^r / r! = 0.05$ , and the 95% upper limit on  $N_S = 7.75 - 3.40 = 4.35$ .

### 8.6.2 Method 2: Go Bayesian

Assign a uniform prior to  $N_S$ , for  $N_S > 0$ , zero for  $N_S < 0$ . The posterior is then just the likelihood,  $P(N_S | N_D, N_B) = e^{-(N_S + N_B)} \frac{(N_S + N_B)^{N_D}}{N_D!}$ . The required limit is obtained from integrating  $\int_0^{N_{hi}} P(N_S) dN_S = 0.95$  where  $P(N_S) \propto e^{-(N_S + 3.40)} \frac{(N_S + 3.4)^3}{3!}$ ; this is illustrated in Fig. 27 and the value of the limit is 5.21.

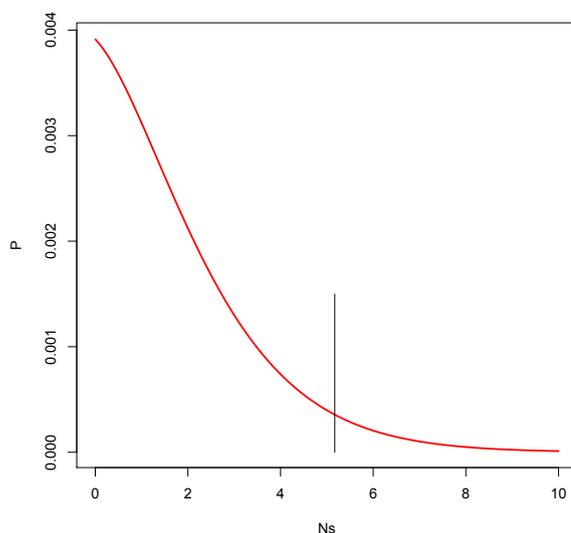


Fig. 27: The Bayesian limit construction

### 8.6.3 Method 3: Feldman-Cousins

This—called ‘the unified approach’ by Feldman and Cousins [19]—takes a step backwards and considers the ambiguity in the use of confidence belts.

In principle, if you decide to work at, say, 90% confidence you may choose to use a 90% central or a 90% upper limit, and in either case the probability of the result lying in the band is at least 90%. This is shown in Fig. 28.

In practice, if you happen to get a low result you would quote an upper limit, but if you get a high result you would quote a central limit. This, which they call ‘flip-flopping’, is illustrated in the plot by a break shown here for  $r = 10$ .

Now the confidence belt is the green one for  $r < 10$  and the red one for  $r \geq 10$ . The probability of lying in the band is no longer 90%! Flip-flopping invalidates the Frequentist construction, leading to undercoverage.

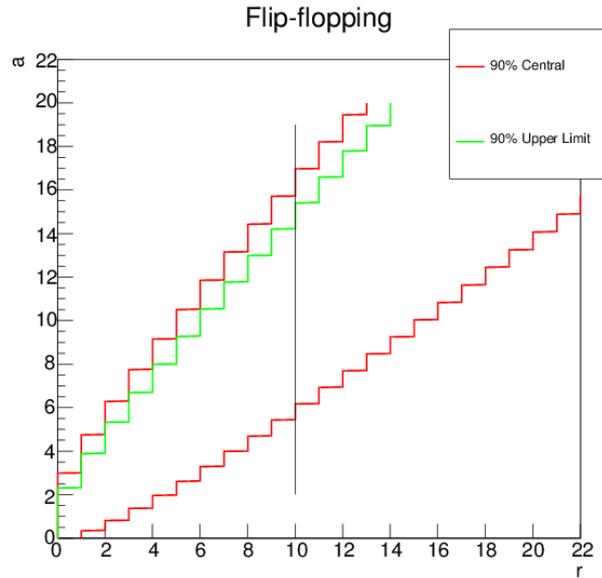


Fig. 28: The flip-flopping problem

They show how to avoid this. You draw the plot slightly differently:  $r \equiv N_D$  is still the horizontal variable, but as the vertical variable you use  $N_S$ . (This means a different plot for any different  $N_B$ , whereas the previous Poisson plot is universal, but this is not a problem.) This is to be filled using  $P(r; N_S) = e^{-(N_S+N_B)} \frac{(N_S+N_B)^r}{r!}$ .

For each  $N_S$  you define a region  $R$  such that  $\sum_{r \in R} P(r; N_S) \geq 90\%$ . You have a choice of strategy that goes beyond ‘central’ or ‘upper limit’: one plausible suggestion would be to rank  $r$  by probability and take them in order until the desired total probability content is achieved (which would, incidentally, give the shortest interval). However this has the drawback that outcomes with  $r < N_B$  will have small probabilities and be excluded for all  $N_S$ , so that, if such a result does occur, one cannot say anything constructive, just ‘This was unlikely’.

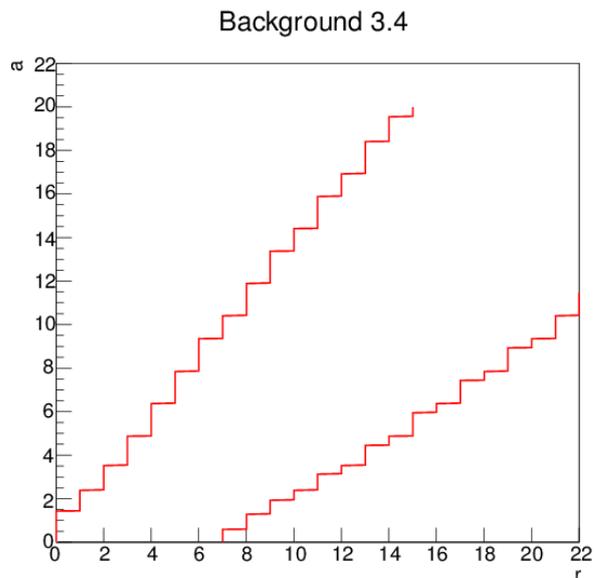
An improved form of this suggestion is that for each  $N_S$ , considering each  $r$  you compare  $P(r; N_S)$  with the largest possible value obtained by varying  $N_S$ . This is easier than it sounds because this highest value is either at  $N_S = r - N_B$  (if  $r \geq N_B$ ) or  $N_S = 0$  (if  $r \leq N_B$ ). Rank on the ratio  $P(r; N_S)/P(r; N_S^{best})$  and again take them in order till their sum gives the desired probability.

This gives a band as shown in Fig. 29, which has  $N_B = 3.4$ . You can see that ‘flip-flopping’ occurs naturally: for small values of  $r$  one just has an upper limit, whereas for larger values, above  $r = 7$ , one obtains a lower limit as well. Yet there is a single band, and the coverage is correct (i.e. it does not undercover). In the case we are considering,  $r = 3$ , just an upper limit is given, at 4.86.

Like other good ideas, this has not found universal favour. Two arguments are raised against the method.

First, that it deprives the physicist of the choice of whether to publish an upper limit or a range. It could be embarrassing if you look for something weird and are ‘forced’ to publish a non-zero result. But this is actually the point, and in such cases one can always explain that the limits should not be taken as implying that the quantity actually is nonzero.

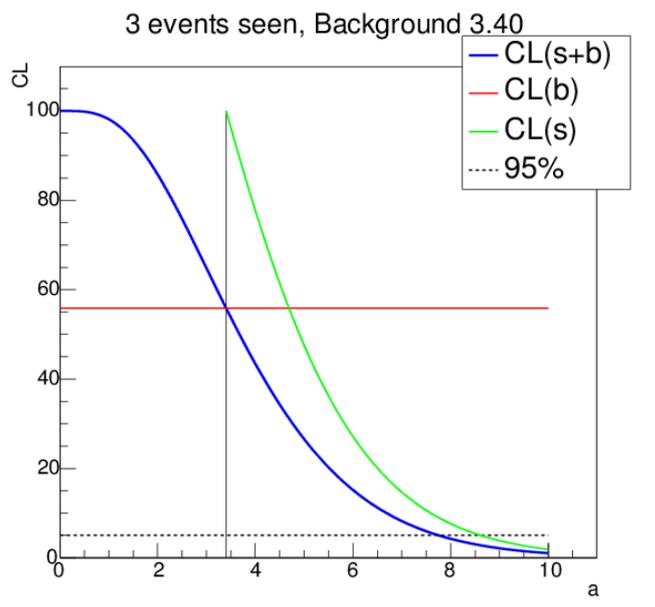
Secondly, if two experiments with different  $N_B$  get the same small  $N_D$ , the one with the higher  $N_B$  will quote a smaller limit on  $N_S$ . The worse experiment gets the better result, which is clearly unfair! But this is not comparing like with like: for a ‘bad’ experiment with large background to get a small number of events is much less likely than it is for a ‘good’ low background experiment.



**Fig. 29:** A Feldman-Cousins confidence band

**8.6.4 Method 4:  $CL_s$**

This is a modification of the standard frequentist approach to include the fact, as mentioned above, that a small observed signal implies a downward fluctuation in background [20]. Although presented here using just numbers of events, the method is usually extended to use the full likelihood of the result, as will be discussed in Section 8.6.6.



**Fig. 30:** The  $CL_s$  construction

Denote the (strict frequentist) probability of getting a result this small (or less) from  $s + b$  events as  $CL_{s+b}$ , and the equivalent probability from pure background as  $CL_b$  (so  $CL_b = CL_{s+b}$  for  $s = 0$ ). Then introduce

$$CL_s = \frac{CL_{s+b}}{CL_b} . \quad (39)$$

Looking at Fig. 30, the  $CL_{s+b}$  curve shows that if  $s + b$  is small then the probability of getting three events or less is high, near 100%. As  $s + b$  increases this probability falls, and at  $s + b = 7.75$  the probability of only getting three events or less is only 5%. This, after subtraction of  $b = 3.4$ , gives the strict frequentist value.

The point  $s + b = 3.4$  corresponds to  $s = 0$ , at which the probability  $CL_b$  is 56%. As  $s$  must be non-negative, one can argue that everything to the left of that is unmeaningful. So one attempts to incorporate this by renormalizing the (blue)  $CL_{s+b}$  curve to have a maximum of 100% in the physically sensible region, dividing it by 0.56 to get the (green)  $CL_s$  curve. This is treated in the same way as the  $CL_{s+b}$  curve, reading off the point at  $s + b = 8.61$  where it falls to 5%. This is a limit on  $s + b$  so we subtract 3.4 to get the limit on  $s$  as 5.21. This is larger than the strict frequentist limit: the method over-covers (which, as we have seen, is allowed if not encouraged) and is, in this respect ‘conservative’<sup>6</sup>. This is the same value as the Bayesian Method 2, as it makes the same assumptions.

$CL_s$  is not frequentist, just ‘frequentist inspired’. In terms of statistics there is perhaps little in its favour. But it has an intuitive appeal, and is widely used.

### 8.6.5 Summary so far

Given three observed events, and an expected background of 3.4 events, what is the 95% upper limit on the ‘true’ number of events? Possible answers are shown in table 2.

Strict Frequentist	4.35
Bayesian (uniform prior)	5.21
Feldman-Cousins	4.86
$CL_s$	5.21

**Table 2:** Upper limits from different methods

Which is ‘right’? Take your pick! All are correct. (Well, not wrong.). The golden rule is to say what you are doing, and if possible give the raw numbers.

### 8.6.6 Extension: not just counting numbers

These examples have used simple counting experiments. But a simple number does not (usually) exploit the full information.

Consider the illustration in Fig. 31. One is searching for (or putting an upper limit on) some broad resonance around 7 GeV. One could count the number of events inside some window (perhaps 6 to 8 GeV?) and subtract the estimated background. This might work with high statistics, as in the left, but would be pretty useless with small numbers, as in the right. It is clearly not optimal just to count an event as ‘in’, whether it is at 7.0 or 7.9, and to treat an event as ‘out’, if it is at 8.1 or 10.1.

It is better to calculate the Likelihood  $\ln L_{s+b} = \sum_i \ln N_s S(x_i) + N_b B(x_i)$  ;  $\ln L_b = \sum_i \ln N_b B(x_i)$ . Then, for example using  $CL_s$ , you can work with  $L_{s+b}/L_b$ , or  $-2 \ln(L_{s+b}/L_b)$ . The confidence/probability quantities can be found from simulations, or sometimes from data.

<sup>6</sup>‘Conservative’ is a misleading word. It is used by people describing their analyses to imply safety and caution, whereas it usually entails cowardice and sloppy thinking.

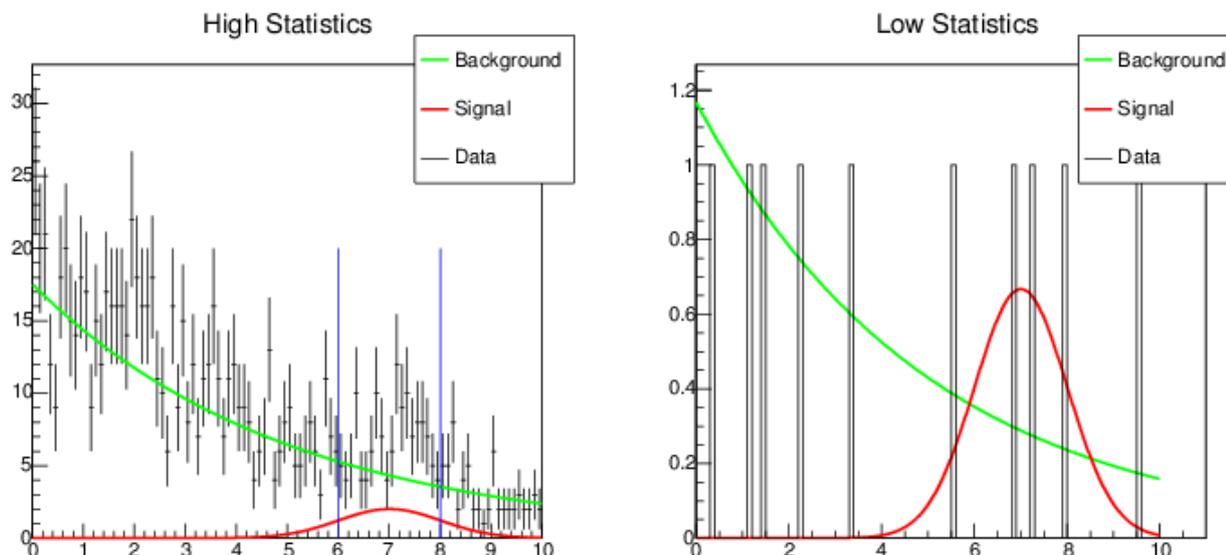


Fig. 31: Just counting numbers may not give the full information

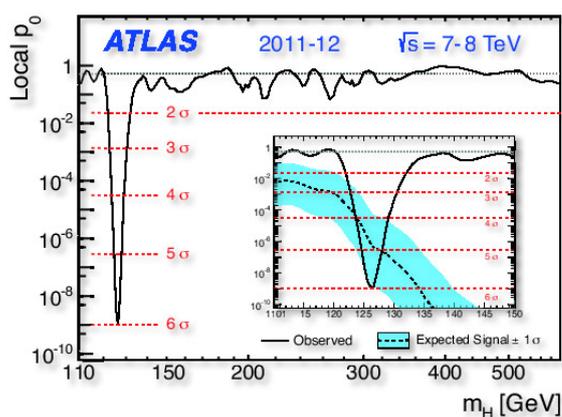


Fig. 32: Significance plot for the Higgs search

### 8.6.7 Extension: From numbers to masses

Limits on numbers of events can readily be translated into limits on branching ratios,  $BR = \frac{N_s}{N_{total}}$ , or limits on cross sections,  $\sigma = \frac{N_s}{\int \mathcal{L} dt}$ .

These may translate to limits on other, theory, parameters.

In the Higgs search (to take an example) the cross section depends on the mass,  $M_H$ —and so does the detection efficiency—which may require changing strategy (hence different backgrounds). This leads to the need to basically repeat the analysis for all (of many)  $M_H$  values. This can be presented in two ways.

The first is shown in Fig. 32, taken from Ref. [21]. For each  $M_H$  (or whatever is being studied) you search for a signal and plot the  $CL_s$  (or whatever limit method you prefer) significance in a *Significance Plot*. Small values indicate that it is unlikely to get a signal this large just from background.

One often also plots the expected (from MC) significance, assuming the signal hypothesis is true. This is a measure of a ‘good experiment’. In this case there is a discovery level drop at  $M_H \approx 125$  GeV, which exceeds the expected significance, though not by much: ATLAS were lucky but not incredibly

lucky.

The second method is—for some reason—known as the green-and-yellow plot. This is basically the same data, but fixing  $CL$  at a chosen value: in Fig. 33 it is 95%. You find the limit on signal strength, at this confidence level, and interpret it as a limit on the cross section  $\sigma/\sigma_{SM}$ . Again, as well as plotting the actual data one also plots the expected (from MC) limit, with variations. If there is no signal, 68% of experiments should give results in the green band, 95% in the yellow band.

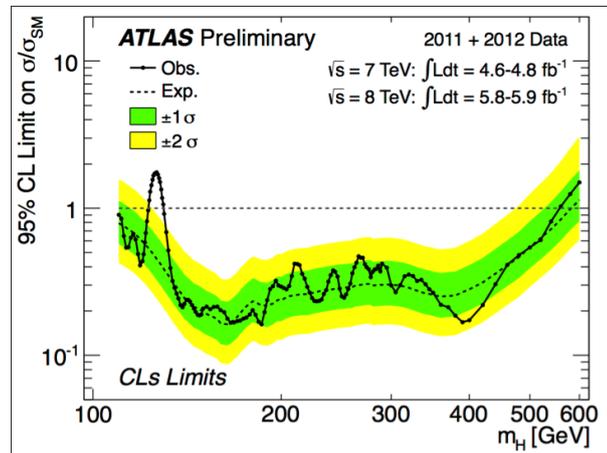


Fig. 33: Green and yellow plot showing the Higgs discovery

So this figure shows the experimental result as a black line. Around 125 GeV the 95% upper limit is more than the Standard Model prediction indicating a discovery. There are peaks between 200 and 300 GeV, but they do not approach the SM value, indicating that they are just fluctuations. The value rises at 600 GeV, but the green (and yellow) bands rise also, showing that the experiment is not sensitive for such high masses: basically it sees nothing but would expect to see nothing.

## 9 Making a discovery

We now turn from setting limits, to say what you did not see, to the more exciting prospect of making a discovery.

Remembering hypothesis testing, in claiming a discovery you have to show that your data can't be explained without it. This is quantified by the  $p$ -value: the probability of getting a result this extreme (or worse) under the null hypothesis/Standard Model. (This is *not* 'The probability that the Standard Model is correct', but it seems impossible for journalists to understand the difference.)

Some journals (particularly in psychology) refuse to publish papers giving  $p$ -values. If you do lots of studies, some will have low  $p$ -values (5% below 0.05 etc.). The danger is that these get published, but the unsuccessful ones are binned.

Is  $p$  like the significance  $\alpha$ ? Yes and no. The formula is the same, but  $\alpha$  is a property of the test, computed before you see the data.  $p$  is a property of the data.

### 9.1 Sigma language

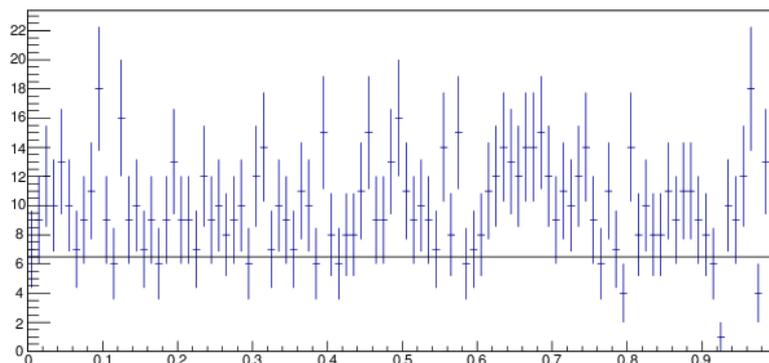
The probability ( $p$ -value) is often translated into Gaussian-like language: the probability of a result more than  $3\sigma$  from the mean is 0.27% so a  $p$ -value of 0.0027 is a '3  $\sigma$  effect' (or 0.0013 depending on whether one takes the 1-tailed or 2-tailed option. Both are used.) In reporting a result with a significance of 'so many  $\sigma$ ' there is no actual  $\sigma$  involved: it is just a translation to give a better feel for the size of the probability.

By convention, 3 sigma,  $p = 0.0013$  is reported as ‘Evidence for’ whereas a full 5 sigma  $p = 0.0000003$  is required for ‘discovery of’.

## 9.2 The look-elsewhere effect

You may think that the requirement for 5  $\sigma$  is excessively cautious. Its justification comes from history—too many 3- and 4- sigma ‘signals’ have gone away when more data was taken.

This is partly explained by the ‘look-elsewhere effect’. How many peaks can you see in the data in Fig. 34?



**Fig. 34:** How many peaks are in this data?

The answer is that there are none. The data is in fact purely random and flat. But the human eye is very good at seeing features.

With 100 bins, a  $p$ -value below 1% is pretty likely. This can be factored in, to some extent, using pseudo-experiments, but this does not allow for the sheer number of plots being produced by hard-working physicists looking for something. Hence the need for caution.

This is not just ancient history. ATLAS and CMS recently observed a signal in the  $\gamma\gamma$  mass around 750 GeV, with a significance of  $3.9\sigma$  (ATLAS) and  $3.4\sigma$  (CMS), which went away when more data was taken.

## 9.3 Blind analysis

It is said<sup>7</sup> that when Michaelangelo was asked how he created his masterpiece sculpture ‘David’ he replied ‘It was easy—all I did was get a block of marble and chip away everything that didn’t look like David’. Such creativity may be good for sculpture, but it’s bad for physics. If you take your data and devise cuts to remove all the events that don’t look like the signal you want to see, then whatever is left at the end will look like that signal.

Many/most analyses are now done ‘blind’. Cuts are devised using Monte Carlo and/or non-signal data. You only ‘open the box’ once the cuts are fixed. Most collaborations have a formal procedure for doing this.

This may seem a tedious imposition, but we have learnt the hard way that it avoids embarrassing mistakes.

<sup>7</sup>This story is certainly not historically accurate, but it’s still a good story (*quoteinvestigator.com*: <https://quoteinvestigator.com/2014/06/22/chip-away/>).

## 10 Conclusions

Statistics is a tool for doing physics. Good physicists understand their tools. Don't just follow without understanding, but read books and conference proceedings, go to seminars, talk to people, experiment with the data, and understand what you are doing. Then you will succeed. And you will have a great time!

## References

- [1] R. J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, (Wiley, Chichester, 1989).
- [2] G. Cowan, *Statistical Data Analysis*, (Oxford Univ. Press, Oxford, 1998).
- [3] I. Narsky and F. C. Porter, *Statistical Analysis Techniques in Particle Physics*, (Wiley, Weinheim, 2014), doi:10.1002/9783527677320.
- [4] O. Behnke *et al* (Eds.) *Data Analysis in High Energy Physics*, (Wiley, Weinheim, 2013), doi:10.1002/9783527653416.
- [5] L. Lyons, *Statistics for Nuclear and Particle Physicists*, Cambridge Univ. Press, Cambridge, 1986).
- [6] G. Bohm and G. Zech, *Introduction to Statistics and Data Analysis for Physicists*, 3rd revised ed. (Verlag Deutsches Elektronen-Synchrotron, Hamburg, 2017), doi:10.3204/PUBDB-2017-08987.
- [7] M. R. Whalley and L. Lyons (Eds) *Advanced Statistical Techniques in Particle Physics*, Durham report IPPP/02/39, 2002, <https://inspirehep.net/literature/601052>.
- [8] L. Lyons, R. Mount and R. Reitmayer (Eds.) *Proceedings of PHYSTAT03*, SLAC-R-703, 2003, <https://www.slac.stanford.edu/econf/C030908/>.
- [9] L. Lyons and M. K. Unel (Eds.) *Proceedings of PHYSTAT05*, (Imperial College Press, London, 2006), doi:10.1142/p446.
- [10] R. von Mises, *Probability, Statistics and Truth*, reprint of the second revised 1957 English edition (Dover, Mineola, NY, 1981).
- [11] H. Jeffreys, *Theory of Probability*, 3rd ed. (Oxford Univ. Press, Oxford, 1961).
- [12] R. J. Barlow, *A note on  $\Delta \ln L = -\frac{1}{2}$  Errors*, arXiv:physics/0403046, 2004.
- [13] R. J. Barlow, *Asymmetric Statistical Errors*, Proceedings of PHYSTAT05, Eds. L. Lyons and M. K. Unel (Imperial College Press, London, 2006), p.56, doi:10.1142/9781860948985\_0013, arXiv:physics/0406120, 2004.
- [14] CMS Collaboration, *CMS 2D Cf-Cv Likelihood Profile*, <https://root.cern.ch/cms-2d-cf-cv-likelihood-profile>, accessed 26 May 2019.
- [15] P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, 3rd ed. (McGraw Hill, New York, NY, 2003).
- [16] J. Orear, *Notes on Statistics for Physicists*, UCRL-8417, 1958, <http://nedwww.ipac.caltech.edu/level5/Sept01/Orear/frames.html>
- [17] B. Aubert *et al.* [BaBar Collaboration], *Phys. Rev. Lett.* **86** (2001) 2515, doi:10.1103/PhysRevLett.86.2515, arXiv:hep-ex/0102030.
- [18] J. G. Heinrich, CDF internal note CDF/MEMO/BOTTOM.CDFR/5639 (Many thanks to Jonas Rademacker for pointing this out); L. Lyons, R. Mount and R. Reitmayer (Eds.) *Proceedings of PHYSTAT03*, SLAC-R-703, 2003, p.52, <https://www.slac.stanford.edu/econf/C030908/>.
- [19] G. J. Feldman and R. D. Cousins, *Phys. Rev.* **D57** (1998) 3873, doi:10.1103/PhysRevD.57.3873, arXiv:physics/9711021.
- [20] A. L. Read, *J. Phys.* **G28** (2002), 2693, doi:10.1088/0954-3899/28/10/313.
- [21] ATLAS Collaboration, *ATLAS Higgs Search Update*, <https://atlas.cern/updates/atlas-news/atlas-higgs-search-update>, accessed 26 May 2019.



# Scientific Programme <sup>1</sup>

## Neutrino Physics

*Gabriela Barenboim (Valencia)*

## Practical Statistics for Particle Physicists

*Roger Barlow (Huddersfield)*

## Field Theory and the EW Standard Model

*Andy Cohen (HKUST and Boston)*

## Physics Beyond Colliders

*Michael Doser (CERN)*

## Higgs Physics and Physics Beyond the Standard Model

*Christophe Grojean (DESY and Humboldt)*

## Hadron Spectroscopy

*Feng-Kun Guo (ITP)*

## Instrumentation and Detectors

*Junji Haba (KEK)*

## Future e+e- Accelerator-Based Experiments

*Sachi Komamiya (Tokyo)*

## Cosmology and Dark Matter

*Hitoshi Murayama (IPMU)*

## Heavy-Ion Physics

*Matthew Nguyen (IN2P3 and LLR)*

## Flavour Physics and CP Violation

*Yossi Nir (Weizmann)*

## QCD

*Gavin Salam (CERN)*

## LHC Results: Highlights and Perspectives

*Nhan Tran (Fermilab)*

## Q & A Session

*Fabiola Gianotti (CERN)*

---

<sup>1</sup>Slides available at <https://indico.cern.ch/event/706886/timetable/?daysPerRow=5&view=nicecompact>

## **Local Organizing Committee**

Jean Trân Thanh Vân (Rencontres du Vietnam)  
Trân Thanh Sơn (ICISE)  
Nguyen Thi Hong Van (IFIRSE)  
Le Duc Ninh (IFIRSE)

## **International Organizing Committee**

Ashfaq Ahmad (NCP)  
Elisabetta Barberio (Melbourne)  
Marc Besancon (CEA/Irfu)  
Subhasis Chattopadhyay (VECC)  
Nick Ellis (CERN, Chair)  
Kazunori Hanagaki (KEK)  
Yee Bob Hsiung (National Taiwan University)  
Kiyotomo Kawagoe (Kyushu)  
Pyungwon Ko (KIAS)  
Martijn Mulders (CERN)  
Sreerup Raychaudhuri (TIFR)  
Lydia Roos (IN2P3/LPNHE)  
Thomas Schoerner-Sadenius (DESY)  
Boris Shwartz (BINP)  
Changzheng Yuan (IHEP)  
Shi-Lin Zhu (Peking University)

## **International Advisory Committee**

Ursula Bassler (IN2P3/CNRS)  
Alexander Bondar (BINP)  
Rohini Godbole (IISc)  
Hafeez Hoorani (NCP)  
Joachim Mnich (DESY)  
Mitsuaki Nozaki (KEK, Chair)  
Martin Sevier (University of Melbourne)  
Xiaoyan Shen (IHEP)  
Dongchul Son (KNU)  
Henry Tsz-king Wong (Academia Sinica)  
Rüdiger Voss (CERN)

## **Lecturers**

Gabriela Barenboim (Valencia)  
Roger Barlow (Huddersfield)  
Andy Cohen (HKUST and Boston)  
Michael Doser (CERN)  
Fabiola Gianotti (CERN)  
Christophe Grojean (DESY and Humboldt)  
Feng-Kun Guo (ITP)

Junji Haba (KEK)  
Sachi Komamiya (Tokyo)  
Hitoshi Murayama (IPMU)  
Matthew Nguyen (IN2P3 and LLR)  
Yossi Nir (Weizmann)  
Gavin Salam (CERN)  
Nhan Tran (Fermilab)

## **Discussion Leaders**

Andrej Arbuzov (JINR)  
Aoife Bharucha (CNRS and CPT)  
Rizwan Khalid (NUST)  
DingYu Shao (CERN)  
Hao Zhang (IHEP)

# Students

Michael ADERSBERGER	Dajeong JEON	Shibasis ROY
Lucas ALTENKAMPER	Daisy KALRA	Shreya ROY
Dennis AROGANCIA	James KENDRICK	Eli Baverfjord RYE
Hicham ATMANI	Faiza KHALID	Debashis SAHOO
Marta BABICZ	Jun KOGA	Arunika SAHU
Shyam BALAJI	Evgeny KOZYREV	Riccardo SALVATICO
William Mickael BARBE	Peter KRACHKOV	Uttiya SARKAR
Saliha BASHIR	Seungcheol LEE	Diana SEITOVA
Eshwen BHAL	Suxian LI	Nizar SEPTIAN
Patrick BOLTON	Xinxin MA	Wadut SHAIKH
Marco BORETTO	Priya MAJI	Meenakshi SHARMA
Daniel Lawrence BRIGLIN	Katja MANKINEN	Huda SIDDIQUI
Maria Brigida BRUNETTI	Nguyen MINH TRUONG	Jaydip SINGH
Zaineb CALCUTTAWALA	Tigran MKRTCHYAN	Manoj Kumar SINGH
Andrea CARDINI	Michaela MLYNARIKOVA	Kim SMITH
Kaustav CHAKRABORTY	Muhammad MUBASHER	Iqra SOHAIL
Suman CHATTERJEE	Bharati NAIK	Lukas SOHL
Yifan CHEN	Dmitrii NEVEROV	Mahdi TAANI
Pallabi DAS	Thanh Tien Dat NGUYEN	Kosuke TAKEDA
Kevin FLOEH	Emery NIBIGIRA	Loc TRAN QUANG
Emmy Pauline Maria GABRIEL	Tomas NOSEK	Tran VAN NGOC
Jhuma GHOSH	Andrzej NOVAK	Fariha VARDAG
Swagata GHOSH	Maria ONAAS	Naoki YAMAGUCHI
Shan GU	Lata PANWAR	Xiaoqing YUAN
Kunlin HAN	Sonia PARMAR	Tamar ZAKAREISHVILI
Simen HELLESUND	Luis PASCUAL DOMINGUEZ	Grzegorz ZARNECKI
Siew Yan HOH	Ondrej PENC	Xiao ZHAO
Jahid HOSSAIN	Dmitrii PEREIMA	Yi ZHENG
Ying-Rui HOU	Vitaliy POPOV	Valentina ZHUKOVA
Lorenza IACOBUZIO	Niharika ROUT	Jennifer ZONNEVELD

## Posters

Poster title	Presenter
Search for Stable Massive Particles	ADERSBERGER, M.
$J/\psi$ -hadron correlations in high-multiplicity proton-proton collisions at 13 TeV at mid-rapidity with ALICE	ALTENKAMPER, L.
Dark Matter Searches at CMS at 13 TeV	BHAL, E.
New physics with the lepton flavor violating decay $\tau \rightarrow 3\mu$	CALCUTTAWALA, Z.
Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in CMS	CARDINI, A.
Partial $\mu$ - $\tau$ Reflection Symmetry and its Consequences	CHAKRABORTY, K.
Radius Scan for Inclusive Jets in CMS Experiment at 13 TeV	CHATTERJEE, S.
Mapping E-field with a UV-laser system in MicroBooNE	CHEN, Y.
Search for the single top production with Higgs in the multilepton final state	DAS, P.
Single top + Higgs at CMS in pp collisions at 13 TeV	FLÖH, K.
The TORCH Detector	GABRIEL, E.
Inclusive $\Psi(2S)$ suppression in p-Pb 8.16 TeV collisions with ALICE	GHOSH, J.
Potential of a singlet scalar enhanced Standard Model	GHOSH, S.
Status of the NOvA neutrino-induced neutral current neutral $\pi^0$ production cross-section measurements	KALRA, D.

Poster title	Presenter
Measurement of the angular distribution of prompt gamma-rays emitted from $^{117}\text{Sn}(n,\gamma)$ reaction for T-violation search	KOGA, J.
Lepton Flavor Non-Universality in Tau Sector	MAJI, P.
DeeMe - A Search for $\mu - e$ Conversion Experiment at J-PARC	MINH TRUONG, N.
Hadron response studies of the ATLAS Tile Hadronic Calorimeter at Test Beams	MKRTCHYAN, T.
Study of the $H \rightarrow \tau\tau$ decay channel with ATLAS	MLYNARIKOVA, M.
Measurement of azimuthal correlations of D mesons with charged particles in pp collisions at 13 TeV with ALICE at the LHC	NAIK, B.
Search for Magnetic Monopoles with the Belle II experiment	NEVEROV, D.
Search for $W' \rightarrow tb$ resonances in the lepton+jets final states with ATLAS detector	NIBIGIRA, E.
Neutrino and Antineutrino Oscillations in NOvA	NOSEK, T.
Performance of Deep Learning Based Tagging Algorithms for Boosted Double Quark Jet Topology	NOVAK, A.
Photon Calorimetric isolation with the ATLAS detector: robustness against pileup	PASCUAL DOMÍNGUEZ, L.
Event-by-event study of charge separation in Pb-Pb collisions at 2.76 TeV using AMPT	PARMAR, S.
Charm and B rediscoveries at Belle II	ROUT, N.
Lepton mass effects and angular observables in $\Lambda_b \rightarrow \Lambda(\rightarrow p\pi)\ell^+\ell^-$	ROY, S. (SHIBASIS)

<b>Poster title</b>	<b>Presenter</b>
Detectors for High Energy Physics Experiments and Cosmic Rays	ROY, S. (SHREYA)
Search for lepton flavor violating tau decays at Belle	SAHOO, D.
Microscopic Black Hole production at LHC with CMS experiment	SEITOVA, D.
Upsilon production in p-Pb collisions at $\sqrt{s_{NN}} = 8.16$ TeV with ALICE at the LHC	SHAIKH, W.
Strangeness production in p-Pb collision system at 8.16 TeV	SHARMA, M.
Constraining Nuclear Effects in the Ar-Target using the High-Resolution STT Near Detector	SINGH, J.
Development of Picosec Micromegas for fast timing in high rate environments	SOHL, L.
Development of software-based data acquisition system for the Level-1 Endcap Muon Trigger in ATLAS Run-3	TAKEDA, K.
$\mu$ -hybrid inflation with low reheat temperature and observable gravity waves	VARDAG, F.
Development of the assemble procedure of silicon pixel detectors for HL-LHC	YAMAGUCHI, N.
Studies of the ATLAS hadronic calorimeter response to muons at Test Beams	ZAKAREISHVILI, T.
Event selection for the cross section measurement of the charged current muon antineutrino single pion production in the T2K near detector	ZARNECKI, G.
Determination of the CP-violating phase $\phi_s$ originating in $B_s \rightarrow J/\psi\Phi$ decays at LHCb	ZONNEVELD, J.