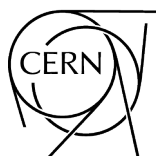


# **Proceedings of the 2022 Asia–Europe–Pacific School of High-Energy Physics**

**Pyeongchang, South Korea, 5–18 October 2022**

Editors: Sascha Stahl, TaeJeong Kim, Pyungwon Ko



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

ISBN 978-92-9083-660-5 (paperback)


ISBN 978-92-9083-659-9 (PDF)

ISSN 2519-804X (Print)

ISSN 2519-805X (Online)

DOI <https://doi.org/10.23730/CYRSP-2024-001>

Copyright © CERN, 2024

 Creative Commons Attribution 4.0

This volume should be cited as:

Proceedings of the 2022 Asia–Europe–Pacific School of High-Energy Physics,  
CERN Yellow Reports: School Proceedings, CERN-2024-001 (CERN, Geneva, 2024)  
<https://doi.org/10.23730/CYRSP-2024-001>.

A contribution in this report should be cited as:

[ChapArticle editor name(s)], in: Proceedings of the 2022 Asia–Europe–Pacific School of High-Energy Physics,  
CERN-2024-001 (CERN, Geneva, 2024), pp. [first page]–[last page],  
<https://doi.org/10.23730/CYRSP-2024-001>. [first page]

Corresponding editor: [Sascha.Stahl@cern.ch](mailto:Sascha.Stahl@cern.ch).

Accepted in May 2024, by the [CERN Reports Editorial Board](#) (contact [Carlos.Lourenco@cern.ch](mailto:Carlos.Lourenco@cern.ch)).

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

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

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

# **Proceedings of the 2022 Asia–Europe–Pacific School of High-Energy Physics**

*Editors: Sascha Stahl, TaeJeong Kim, Pyungwon Ko*

## **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, hadron spectroscopy and practical statistics for particle physics.

## **Keywords**

Neutrino physics, QCD, Hadron spectroscopy, Statistical techniques, Lecture notes



---

## Contents

<b>Preface</b>	
<i>Martijn Mulders</i>	<b>1</b>
<b>Photograph of participants</b>	<b>3</b>
<b>Photographs of school</b>	<b>3</b>
<b>Lecture summaries</b>	<b>5</b>
<b>Statistical techniques</b>	
<i>Nicolas Berger</i>	<b>7</b>
<b>Neutrino physics</b>	
<i>Sin Kyu Kang</i>	<b>47</b>
<b>QCD</b>	
<i>Vajravelu Ravindran</i>	<b>85</b>
<b>Hadron spectroscopy</b>	
<i>Alexei Zhemchugov</i>	<b>105</b>
<b>Scientific programme</b>	<b>115</b>
<b>Organizing committees</b>	<b>116</b>
<b>List of lecturers</b>	<b>117</b>
<b>List of discussion leaders</b>	<b>117</b>
<b>List of students</b>	<b>118</b>
<b>List of posters</b>	<b>119</b>



---

## Preface

The fifth event in the series of the Asia–Europe–Pacific School of High-Energy Physics took place in Pyeongchang, South Korea, from 5 to 18 October 2022. The event was organized in collaboration between a team around Prof. Tae Jeong Kim from Hanyang University, Seoul, and members of the International Organization Committee from CERN and KEK. The school had been delayed by two years due to the COVID-19 pandemic and for many students the school was one of the first opportunities during their studies to travel abroad and meet fellow students.

The staff and students were housed in comfortable accommodation at the Alpensia Resort in Pyeongchang, which also provided excellent conference facilities in a neighbouring building. The students shared accommodation, mixing nationalities to foster cultural exchange between participants from different countries. A total of 96 students of 29 different nationalities attended the school. About 50% of the students were from Asia–Pacific countries, most of the others coming from Europe with a few students from Africa or the US, with 39% female students overall. The majority of the participants were working towards a PhD, while the others were advanced Masters students and a few young postdocs.

A total of 33 lectures were complemented by daily discussion sessions led by six discussion leaders. The teachers (lecturers and discussion leaders) came from institutes all over the world: Austria, France, Germany, India, Japan, Pakistan, Russia, Taiwan, South Korea, Spain, The Netherlands and the USA. Each lecture was allocated 90 minutes including time for questions and discussion. The discussion sessions were also of 90 minutes duration. Professor Fabiola Gianotti, Director-General of CERN, participated in a Q&A Session with the students via video link, and Professor Takaaki Kajita, winner of the 2015 Nobel Prize in Physics, delivered a special lecture about Super Kamiokande by video. 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 about 50% of the 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 six 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 the South Korean culture, including visits to cultural sites in and around Woljeongsa Temple, and excursion to a visitor centre offering views of the demilitarized zone between North and South Korea, and free time in the beach town Gangneung. During the excursions students, lecturers and organizers had the opportunity to appreciate excellent South Korean food.

Our special thanks go to the excellent local-organization team in Korea under the co-chairmanship of Tae Jeong Kim, Hanyang University, and Pyungwon Ko, Korea Institute for Advanced Study, which were supported by Hanna Park and JeongEun Yoon, for all their work and assistance in preparing the School, on both scientific and practical matters, and for their presence throughout the event.

It was a pleasure to welcome Shohei Nishida (KEK) and Pyungwon Ko (KIAS) as members of

---

the International Organizing Committee. Their interest and support for the teaching of fundamental science, and, more generally, international collaboration and cultural exchange, were greatly appreciated. Feedback from the participants after the school was extremely positive in terms of the appreciation of the scientific programme, the quality of the teaching, the beautiful location and practical organization, and the aspects of cultural exchange and building working relationships between promising young scientists from different countries.

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 for her untiring efforts on administration for the School. We would also like to thank the members of the International Committees.

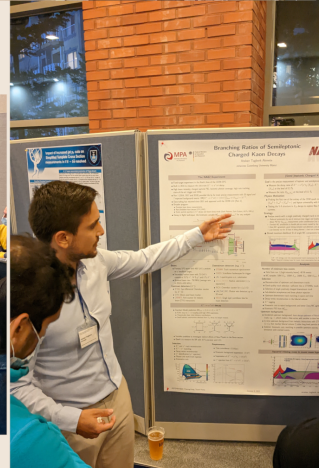
Sponsorship from numerous bodies in many countries covered the cost of travel and/or local expenses of their staff and students who attended the School. In addition, general sponsorship is gratefully acknowledged from: the Center for Extreme Nuclear Matters (CENuM), Hanyang University, Institute for Basic Science (IBS), the Korea-Alice (KoALICE) and Korea-CMS (KCMS) collaborations, the Korea Institute For Advanced Study (KIAS), and the University of Seoul; CNRS/IN2P3, France; CERN; DESY, Germany; and KEK, Japan.

Martijn Mulders<sup>a</sup>  
(Chair of the International Organizing Committee)

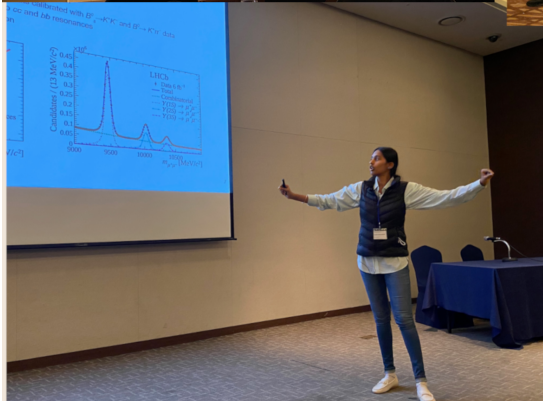
<sup>a</sup>CERN







**AEPSHEP** 2022  
05-18 OCTOBER 2022,  
Pyeongchang, SOUTH KOREA



---

# Lecture summaries

## Statistical techniques

This course covers the main statistical methods used in high-energy physics, focusing in particular on the techniques currently used in LHC experiments. The proceedings cover the following: first, the methods used to describe an experimental setup in probabilistic terms (i.e. to write down a statistical model describing the measurement) are discussed; second, the usage of such a model to produce the usual statistical results in high-energy physics is presented; lastly, as examples, the discovery significances for new signals, confidence intervals for model parameters, and upper limits on signal yields are discussed. The lectures will focus on the use of frequentist techniques.

## Neutrino physics

In this lecture, I start with presenting the history of the neutrino from its invention to what we have discovered about its properties till now. I explain how we can observe neutrinos produced both naturally and artificially. Naturally produced neutrinos come to the Earth from the Sun, supernovae, collisions of cosmic rays with nuclei in the atmosphere, natural radioactivity, etc. On the other hand, those produced in accelerators and nuclear reactors are the examples of artificial neutrinos. I also illustrate what neutrino oscillations are and how such phenomena could be observed from various experiments to detect neutrinos produced in the aforementioned ways. Thanks to the discovery of neutrino oscillations, we are forced to modify the Standard Model, so as to accommodate the masses of neutrinos and lepton flavor mixing, which are essential to make neutrino flavor change. In fact, neutrinos can come in three different flavors, electron, muon and tau, and can change from one flavor to another. The origin of the tiny neutrino masses is still unknown, although we now know a few nice mechanisms capable of generating them. The generation of neutrino masses signifies physics beyond the Standard Model and can, therefore, be related to some of the unresolved fundamental issues, such as the origin of flavors, the unification of forces, the matter-antimatter asymmetry, etc. Some physicists believe that CP violation in neutrinos may be a missing piece in the understanding of the origin of the matter-antimatter asymmetry. I pedagogically explain how we can probe CP violation through neutrino oscillation experiments.

## QCD

The structure of strong interaction dynamics, namely Quantum Chromodynamics, is discussed within the framework of perturbation theory. After a brief introduction to the historical developments, we will discuss in detail the role of perturbative QCD to understand the physics at various high energy colliders involving hadrons. We will discuss how certain large threshold logarithms that show up at every order in perturbation theory can be resummed to all orders. We will also discuss some of the recent advances in performing higher order perturbative corrections.

---

## Hadron spectroscopy

If you query the arXiv electronic database of preprints and look at the topics of published articles, you can see that roughly one in five deals with hadron spectroscopy. During its development, this area of particle physics has experienced several rises and falls. Today, it is a rapidly developing branch of science, which comprises a significant part of the research program of almost every accelerator experiment. It is especially valid for  $B$ -,  $c\tau$ -factories and, certainly, the LHC, where exciting results have been obtained. The purpose of this lecture is to try giving an overview of the current state of hadron spectroscopy through the eyes of an experimentalist.

# Statistical techniques

Nicolas Berger

LAPP, Annecy, France

---

This course covers the main statistical methods used in high-energy physics, focusing in particular on the techniques currently used in LHC experiments. The proceedings cover the following: first, the methods used to describe an experimental setup in probabilistic terms (i.e. to write down a statistical model describing the measurement) are discussed; second, the usage of such a model to produce the usual statistical results in high-energy physics is presented; lastly, as examples, the discovery significances for new signals, confidence intervals for model parameters, and upper limits on signal yields are discussed. The lectures will focus on the use of frequentist techniques.

---

## 1 Introduction

In high-energy physics, as in other fields, experimental processes involve an irreducible random component. For instance, when counting events originating from collider experiments, one can see that the arrival times of these events are randomly distributed. Similarly, measurements of continuous variables are affected by experimental resolution effects that can never be completely removed from the measurement process. This randomness has two underlying sources:

- Experimental noise originating either from the surrounding environment or from imperfections in the measurement apparatus. Reducing its impact is a crucial part of experimental physics, but this cannot be completely achieved.
- Quantum randomness that is inherent to the quantum nature of high-energy physics processes.

The impact of these effects, that manifest themselves for instance as the width of a resonance peak, cannot be accounted for in a deterministic manner: they are described using random processes which account for statistical fluctuations in the description of the measurement.

These lectures will first cover the methods used to describe an experimental setup in probabilistic terms, i.e. to write down a *statistical model* describing the measurement. Secondly, they will present how to use this model to produce the usual statistical results in high-energy physics: discovery significances for new signals, confidence intervals for model parameters, and upper limits on signal yields. The lectures will focus on the use of *frequentist* techniques. Alternative methods based on Bayesian techniques can be found, e.g., in Ref. [1].

## 2 Statistical modeling

We start by presenting the techniques used to build the statistical model of the measurement. This model consists of two components:

---

This article should be cited as: Statistical techniques, Nicolas Berger, DOI: [10.23730/CYRSP-2024-001.7](https://doi.org/10.23730/CYRSP-2024-001.7), in: Proceedings of the 2022 Asia–Europe–Pacific School of High-Energy Physics, CERN Yellow Reports: School Proceedings, CERN-2024-001, DOI: [10.23730/CYRSP-2024-001](https://doi.org/10.23730/CYRSP-2024-001), p. 7.  
© CERN, 2024. Published by CERN under the [Creative Commons Attribution 4.0 license](https://creativecommons.org/licenses/by/4.0/).

- the probability distribution function (PDF) of the measurement, which describes the random process that is assumed to produce the experimental data;
- the *observed data*, i.e. the dataset that was obtained when the measurement was performed.

The PDF of the measurement is the key component of the model. It can generally be written as  $P(n; \alpha)$ , where  $n$  is the set of measured quantities, denoted as the *observables* of the measurement (or *random variables* in mathematical parlance), and  $\alpha$  is a set of parameters that are needed to write down the model. The parameters  $\alpha$  include, for instance, theory quantities such as the value of Standard Model (SM) constants, and experimental quantities such as resolutions, systematic uncertainties and background levels. These parameters are usually separated into two classes:

- *Parameters of interest* (POIs), which are the parameters that the experiment is designed to measure. These are often, but not always, theory quantities. They will be denoted as  $\mu$  in the rest of these notes.
- *Nuisance parameters* (NPs), which are parameters that need to be included in the model to fully describe the measurement process, but are not of interest per se. A typical example would be parameters describing properties of background processes. They will be denoted as  $\theta$  in the rest of these notes.

The rest of these lectures will be focused on how to obtain information on the POIs  $\mu$  based on the knowledge of  $P(n; \mu, \theta)$  and the observed data  $n_{\text{obs}}$ . Building  $P(n; \mu, \theta)$  is the key step in this process, and this will be the focus of the rest of this section.

## 2.1 Building blocks

### 2.1.1 Counting events

In many cases, experiments consist in counting events that pass a given selection. This is particularly true in high-energy physics, where a selection (often a series of “cuts”) is applied to a set of input events produced in a particle collider or another source process.

In general, each step of the selection can be described using a binomial process. The PDFs for each of these, together with the PDF of the source process, can then be used to describe the full measurement. However, in many cases a much simpler description can be used, based on the Poisson approximation. This applies when:

- the number of input events  $N$  to the binomial process is large ( $N \gg 1$ );
- the probability  $p$  to pass the selection is small ( $p \ll 1$ ).

In this case, each binomial process can be approximated by a Poisson distribution

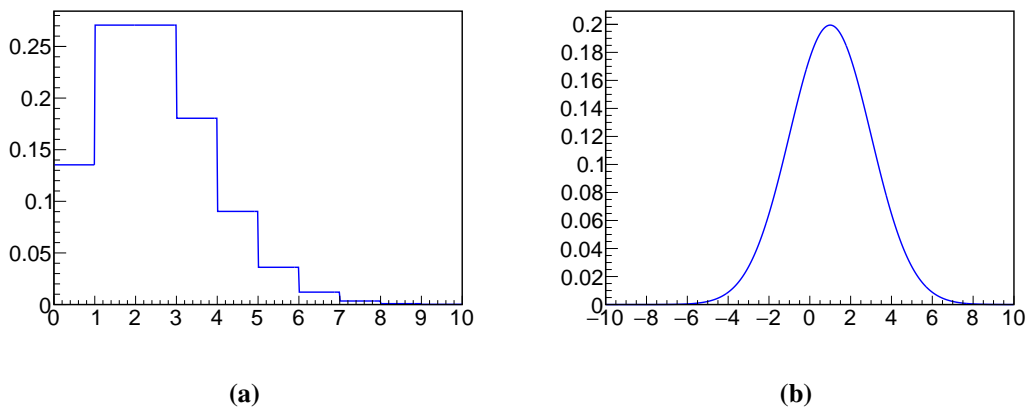
$$P(n; \lambda) = e^{-\lambda} \frac{\lambda^n}{n!} \quad (1)$$

where  $n$  is the measured event count and  $\lambda = Np$  is the expected yield. Furthermore, it can be shown that two successive Poisson processes can be described as a single Poisson process, and that the same

also applies to the combination of a binomial and a Poisson process. If the Poisson approximation holds, the entire counting process can therefore usually be conveniently described as a single Poisson PDF.

Fortunately, this approximation is often valid in high-energy physics. In the specific example of LHC experiments, the production rate of events in  $pp$  collisions is of order  $10^9$  Hz, of which  $10^3$  Hz are recorded by the experiments, and among which interesting signal events typically make up (much) less than 1 Hz. The Poisson approximation of very large input event rates and a very small selection probability is therefore well verified in this case.

Finally, let us recall that the mean and the variance of  $P(n; \lambda)$  are both equal to  $\lambda$ . Its root-mean-square (RMS), the square root of its variance, is therefore  $\sqrt{\lambda}$ . An illustration of the Poisson distribution for  $\lambda = 1$  is shown in Fig. 1a.



**Fig. 1:** (a) Poisson distribution for an expected yield of 1 and (b) Gaussian distribution for a mean of 1 and a width of 2.

## 2.1.2 The Gaussian distribution and the central-limit theorem

### 2.1.2.1 The Gaussian distribution

The Gaussian distribution is a PDF for a single continuous observable  $x$ , defined as

$$G(x; x_0; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x_0}{\sigma}\right)^2} \quad (2)$$

It takes the shape of a symmetric peak with central value  $x_0$  and a width characterized by  $\sigma$ . Its mean is given by  $x_0$ , and its RMS by  $\sigma$ . The shape of the distribution is shown in Fig. 1b.

The Gaussian distribution will play a critical role in much of the rest of these notes. An important feature to keep in mind is the values of its *quantiles*, i.e. the fraction of outcomes that fall within a given interval of the distribution. One can define the *pull*  $z = (x - x_0)/\sigma$  of an observable  $x$  taken from the distribution  $G(x; x_0; \sigma)$ : this quantifies the separation between  $x$  and the mean  $x_0$  of the distribution, in units of the width  $\sigma$ . Simple algebra shows that  $z$  follows a *normal* distribution  $G(z; 0, 1)$ , i.e. a Gaussian with mean 0 and width 1, independently of the parameters of the original Gaussian for  $x$ . This allows us to define quantiles for any Gaussian distributions in terms of  $z$ , and these are shown in Table 1. Key

**Table 1:** Selected quantiles of the Gaussian distribution  $G(x; x_0, \sigma)$ , in terms of the pull  $z = (x - x_0)/\sigma$ .

$Z$	Two-sided		One-sided	
	$p( \frac{x-x_0}{\sigma}  \leq Z)$	$p( \frac{x-x_0}{\sigma}  \geq Z)$	$p(\frac{x-x_0}{\sigma} \leq Z)$	$p(\frac{x-x_0}{\sigma} \geq Z)$
1	0.683	0.317	0.841	0.159
2	0.954	0.046	0.977	0.023
3	0.997	0.0027	0.999	0.0013
5	$\sim 1$	$5.7 \times 10^{-7}$	$\sim 1$	$2.9 \times 10^{-7}$

takeaways include the fact that observations fall within the  $[-1\sigma, +1\sigma]$  interval around the mean about 68.3% of the time, and within  $[-2\sigma, +2\sigma]$  about 95.5% of the time. Gaussians also have "thin tails" so that only 0.3% of outcomes fall beyond the  $\pm 3\sigma$  interval, and only about  $6 \times 10^{-7}$  of the time beyond the  $\pm 5\sigma$  interval. These numbers will all be useful later in these notes. They can all be expressed in terms of the cumulative distribution function (CDF) of the normal distribution,

$$\Phi(z) = \int_{-\infty}^z G(z; 0, 1) dz. \tag{3}$$

For instance the 68.3% quantile corresponding to the  $\pm 1\sigma$  can be obtained as  $\Phi(+1) - \Phi(-1)$ .

### 2.1.2.2 The central-limit theorem

Gaussian distributions occur frequently in experimental settings, in particular to describe resolution effects and uncertainties. The main reason for their ubiquity is a property of the mean of a large number of identical measurements. Let's consider a random process with an observable  $x$ , described by a PDF  $P(x; \alpha)$ , with mean  $\langle x \rangle$  and RMS  $\sigma_x$ . Say that we repeat this process a large number of times  $N$ , and compute the average of  $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$  of the observations  $x_i$  in each case. Then the *central-limit theorem* states that for large  $N$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \sim G\left(\bar{x}, \langle x \rangle, \frac{\sigma_x}{\sqrt{N}}\right). \tag{4}$$

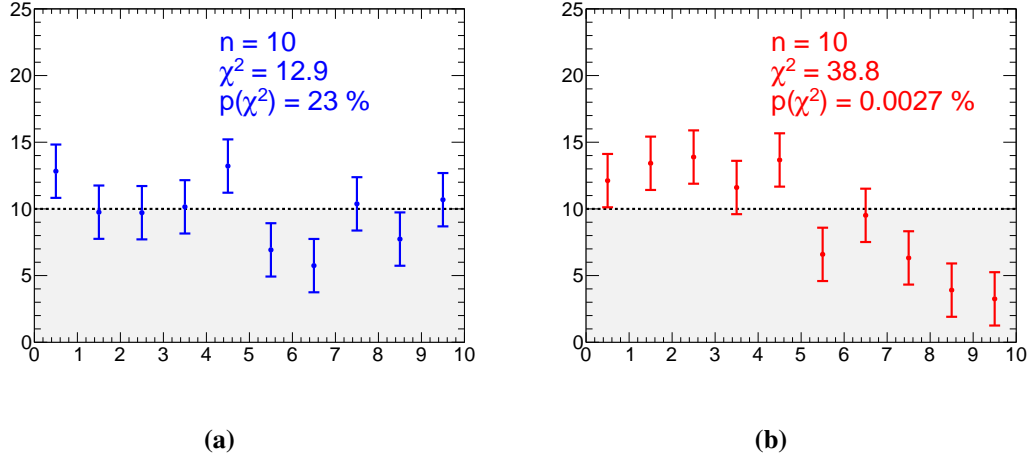
In other words, if we average enough measurements together, then the average will be distributed as a Gaussian no matter what the distribution was for individual measurements. The only residual feature of this distribution is its mean, which is carried over as the mean of the Gaussian, and its RMS, which together with a factor  $1/\sqrt{N}$  gives the width of the Gaussian. The factor  $1/\sqrt{N}$  encodes the fact that our knowledge of the mean improves with more measurements, as one would naively expect.

The central-limit theorem is very often applicable in high-energy physics, as long as the measurements involve a sufficiently large number of events. In particular, Poisson distributions tend towards a Gaussian limit for a sufficiently large expected event yield, so that  $P(n; \lambda) \approx G(n; \lambda, \sqrt{\lambda})$ . As we will see later, this Gaussian regime is reached for relatively small yields, typically  $O(5-10)$ , which motivates the use of Gaussian approximations in a wide range of experimental settings.



### 2.1.3 The $\chi^2$ distribution

Suppose that we produce a histogram of data events, by categorizing events into  $N_{\text{bins}}$  independent bins and counting the number of events  $n_i$  that fall in each bin  $i$ . We also define an *expected* event count  $\mu_i$  in each bin, for example using Monte Carlo simulation. Illustrative examples with a flat expectation are shown in Figs. 2a and 2b. In this situation, it is often useful to quantify the agreement between the



**Fig. 2:** Measurement histograms in a simple 10-bin measurement in which the expected yields in each bin are identical. The measurements are shown as points with error bars, and the expectation by shaded regions. The histogram in (a) was randomly generated from the expectation, while the one in (b) was produced from a different expectation with decreasing yields at high bin numbers. The  $\chi^2$  and  $\chi^2$  probability values for each case are overlaid on the figures, and show that the  $\chi^2$  indicates good agreement in (a) and poor agreement in (b).

observed event counts and the prediction. If one assumes that the measurement in each bin is represented by an independent Gaussian distribution with width  $\sigma_i$ , then the discrepancy between observed and expected counts in each bin can be expressed by the pull  $z_i = (n_i - \mu_i)/\sigma_i$ . To quantify the overall agreement over the entire distribution, one then defines the  $\chi^2$  of the observed with respect to the expected yields as

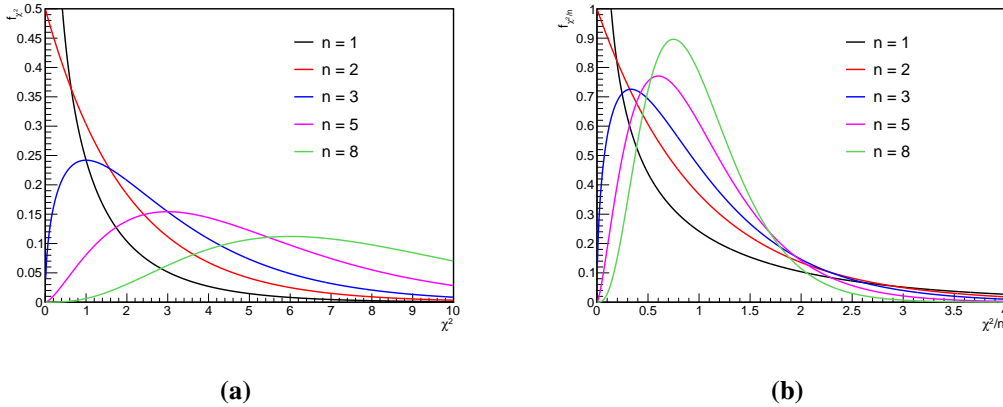
$$\chi^2 = \sum_{i=1}^{N_{\text{bins}}} \left( \frac{n_i - \mu_i}{\sigma_i} \right)^2. \quad (5)$$

This is a positive quantity, and its value is exactly 0 in the case where the observed yields exactly match the expectations. Conversely, large values of the  $\chi^2$  indicate a disagreement between the two.

The observed yields are, however, affected by statistical fluctuations, which lead to small but non-zero values of the  $\chi^2$ . In fact, one can expect on the order of a  $1\sigma$  deviation in each bin, i.e.  $z_i \sim 1$ , which leads to  $\chi^2 \sim N_{\text{bins}}$  overall.

This can be quantified more precisely by introducing the distribution  $f_{\chi^2(N_{\text{bins}})}$  of the  $\chi^2$ , under the hypothesis where the observed yields are produced from the expectation. These distributions are shown in Fig. 3a and show the expected behavior of a peak near  $N_{\text{bins}}$  and a decreasing tail of probabilities to reach high values. These functions are implemented numerically in frameworks such as ROOT and scipy, and allow us to compute the  $\chi^2$  probability as the tail integral of the relevant distribution above the

measured  $\chi^2$  value. Large values of the  $\chi^2$  probability indicate good agreement (since the data is likely for the given expectation), while small values indicate disagreement (since producing this data from the expectation is unlikely). One can also use a rule of thumb based on the *reduced*  $\chi^2$ , defined as  $\chi^2/N_{\text{bins}}$ : as shown in Fig. 3b, the distribution of the reduced  $\chi^2$  is roughly independent from  $N_{\text{bins}}$ , with values of about 1 being fairly typical, and values of 2 being increasingly unlikely. One can therefore gauge agreement by computing the reduced  $\chi^2$  and comparing to a given threshold. For instance a threshold of 1.5 corresponds to a probability of 10–20%, depending on  $N_{\text{bins}}$ . Of course, a more quantitative assessment can be performed by computing the exact  $\chi^2$  probability from the relevant distribution.



**Fig. 3:** Distributions of (a)  $\chi_n^2$  and (b)  $\chi_n^2/n$  for selected values of the number of degrees of freedom  $n$ .

## 2.2 Describing data

Having introduced a few of the basic PDF building blocks, we now turn to how to use this knowledge to model data. The first step is defining the observables, i.e. the measured quantities. These often consist of one or more real numbers that allow us to distinguish signal from background: for instance an invariant mass, or more complex quantities such as the output of a neural network trained to identify the signal process. In other cases, the measured quantities can be one or more event yields, for events passing suitable selection cuts. Some usual modeling choices are described in the following sections.

### 2.2.1 Single-bin counting

The simplest type of measurement is the case where one counts the number of events passing a selection. The observable is then that single number of events  $n$ . As discussed in Section 2.1.1, the counting process can usually be described using a Poisson distribution that is parameterized in terms of an expected event yield ( $\lambda$  in the notation above), which usually receives contributions from both signal and background processes. Assuming we have only one signal process with yield  $S$  and one background process with yield  $B$ , one can write the PDF as

$$p(n; S, B) = e^{-(S+B)} \frac{(S+B)^n}{n!}. \quad (6)$$

Note that, in the formula above,  $n$  is the observable, which is associated with random fluctuations, while  $S$  and  $B$  are model parameters, which have a fixed value (which can be either known or unknown).  $S$  is typically a parameter of interest (POI), while  $B$  is usually a nuisance parameter (NP). The single observable  $n$  cannot be used to determine both  $S$  and  $B$ , so one needs further assumptions to make this a valid measurement. Typically one assumes that  $B$  can be fixed to a predefined value, possibly up to systematic uncertainties (see Section 5.2 for details on how to do this). Recall that for large  $S + B$ , the Poisson distribution can be well-approximated by a Gaussian distribution with mean  $S + B$  and width  $\sqrt{S + B}$ , so one can also use a Gaussian description in this case.

### 2.2.2 Multiple-bin counting

One can go one step further and define a measurement with multiple counting bins. This can occur in two common situations: first, these bins can correspond to several *signal regions* sensitive to different features of the targeted signal; for instance different final states of the same process. Secondly, one can use a set of contiguous bins to describe the distribution of a continuous observable: one just slices the range of the observable into discrete bins to get a discrete approximation to the distribution.

In both scenarios, the bins should be non-overlapping, i.e. a selected event should be assigned to exactly one bin. The bins are then statistically independent, so that the total PDF for the measurement is the product of the measurements in each bin. Assuming as before that the per-bin measurements can each be described by a Poisson distribution, the total PDF can be written as

$$p(\{n_i\}; S, B) = \prod_{k=1}^{N_{\text{bins}}} e^{-(Sf_{S,i} + Bf_{B,i})} \frac{(Sf_{S,i} + Bf_{B,i})^{n_i}}{n_i!}. \quad (7)$$

The observed event yields are denoted as  $n_i$ , for  $i$  running from 1 to the number of bins,  $N_{\text{bins}}$ . As before, this assumes a single signal process and a single background process, with overall expected yields (summed over all bins) respectively  $S$  and  $B$ . The expected yields in each bins are described using the bin fraction  $f_{S,i}$  and  $f_{B,i}$  (with  $\sum_i f_{S,i} = \sum_i f_{B,i} = 1$ ). One could also have expressed the PDF in terms of per-bin yields, but often one is interested in the overall signal yield, so that the form above is more directly useful. Note that in the case where the bins span a continuous distribution, the fractions  $f_{S,i}$  and  $f_{B,i}$  provide a discretized description of the distribution of the observable for signal and background.

Multiple-bin measurements offer a compromise between the simpler single-bin measurements described above and the unbinned measurements that will be covered below, and are therefore very commonly used in high-energy physics experiments. Compared to the single-bin case, they typically provide more sensitive measurements thanks to the extra available information. This information can also allow us to measure the NPs of the model: for instance, with two or more bins one can in principle measure both  $S$  and  $B$ , so that external assumptions on  $B$  are not required. This *data-driven* approach can be built into the design of the measurement, for instance by adding “control region” bins that are specifically designed to constrain the backgrounds. We will come back to this when discussing nuisance parameters in Section 5.

### 2.2.3 Unbinned description

In the case of continuous observables, one can also describe the measurement using a continuous PDF. This is in principle the most sensitive approach, since it avoids the information loss that inevitably occurs when performing a discretization into bins (although this loss can be kept quite small by choosing sufficiently fine bins).

Specializing for simplicity to the case of one signal and one background component, and one observable  $x$ , we need to specify how the events of each type are distributed. This is provided by the signal PDF  $f_S(x)$  and the background PDF  $f_B(x)$ , each describing the distribution in  $x$  of a single event of the respective type.

One then defines the total PDF

$$f_{S+B}(x) = \frac{S}{S+B} f_S(x) + \frac{B}{S+B} f_B(x) \quad (8)$$

which describes the expected single-event distribution in  $x$  for the case of a mixture of signal and background events with total yields  $S$  and  $B$ , respectively.

Since we typically consider datasets consisting of several (and often many!) events, one more ingredient is needed: one needs to describe the random distribution of the total number of events  $n$  in the dataset, which can vary from experiment to experiment. In keeping with the arguments made above, this can be described using the Poisson distribution  $\text{Pois}(n; S+B)$ . Finally, one can put it all together, making use of the fact that events can usually be considered uncorrelated (since e.g. what happens in one collision at the LHC is independent of what may or may not have happened in the previous collisions). The total PDF for the dataset  $\{x_i\}_{1 \leq i \leq n}$  can then be written as

$$p(\{x_i\}; S, B) = e^{-(S+B)} \frac{(S+B)^n}{n!} \prod_{i=1}^n f_{S+B}(x) \quad (9)$$

$$= e^{-(S+B)} \frac{(S+B)^n}{n!} \prod_{i=1}^n \left[ \frac{S}{S+B} f_S(x) + \frac{B}{S+B} f_B(x) \right] \quad (10)$$

$$= \frac{e^{-(S+B)}}{n!} \prod_{i=1}^n [S f_S(x) + B f_B(x)]. \quad (11)$$

This *unbinned* PDF provides more information than a binned description, but is often more complex to implement. In particular, describing  $f_S(x)$  and  $f_B(x)$  can be technically difficult, relying for instance on sampling the distributions using large samples of simulated signal and background events. Computing  $p(\{x_i\}; S, B)$  is also more computationally demanding than a binned approximation, since the product runs over the number of events rather than the number of bins, and the latter is typically much smaller.

In realistic high-energy physics cases, one almost always uses one of the three descriptions above to model the data. The multi-bin description is probably the one used most often, since it often provides a good compromise between the simplicity of the one-bin counting case and the complexity of the unbinned description. Several frameworks have been developed to implement multi-bin cases, for instance the `HistFactory` package [5] available within the `ROOT` [6] framework and the `pyhf` [7, 8] tool. However, the single-bin case is used in some cases, such as measuring total cross-sections [9]. The unbinned

description is often useful in situations where the shape of the signal and background are simple to parameterize, for instance for smooth backgrounds. A well-known example is the study of the  $H \rightarrow \gamma\gamma$  decay by the ATLAS and CMS collaborations [10, 11].

### 3 Introduction to statistical results: the simple Gaussian case

In the previous section we have learned how to build a statistical model for a given experimental setup, using one of the different options described. The next step is to use this model to obtain information about the parameters of interest, for instance on the event yield  $S$ , in the examples given above. The good news is that building the model was the hard part; obtaining these *statistical results* on the POIs will just involve some mathematics.

Before moving to the general methods of obtaining these statistical results, this section will introduce basic concepts in the context of a simple case: the single-bin counting experiment. For simplicity, we assume that the measurement is Gaussian, and that only a single signal process (with yield  $S$ ) and a single background process (with yield  $B$ ) are present. The measurement PDF is

$$p(n; S, B) = G(n; S + B, \sqrt{S + B}). \quad (12)$$

The goal is to determine whether the signal is present or not, and in what amounts, by measuring the parameter of interest  $S$ . As noted before, we need to assume that  $B$  is known a priori in this simple example. We can assume, for instance,  $B = 100$ , which by the central limit theorem (see Section 2.1.2.2) is large enough to give a measurement that is well within the Gaussian regime. Assume now that we measure  $n = 120$ , as is illustrated in Fig. 4a. What can we conclude about  $S$ ?

#### 3.1 Estimating $S$

Very naively, we can compute  $S$  as

$$\hat{S} = n - B, \quad (13)$$

since  $B$  is known exactly. Note the hat on  $S$ : this will be used in the following to refer to *estimators* for parameters, i.e. quantities we use to give information on its true value  $S$ . Whereas  $S$  is fixed (but unknown), the estimator  $\hat{S}$  is a function of the data: a different experimental result would lead to a different  $\hat{S}$ .

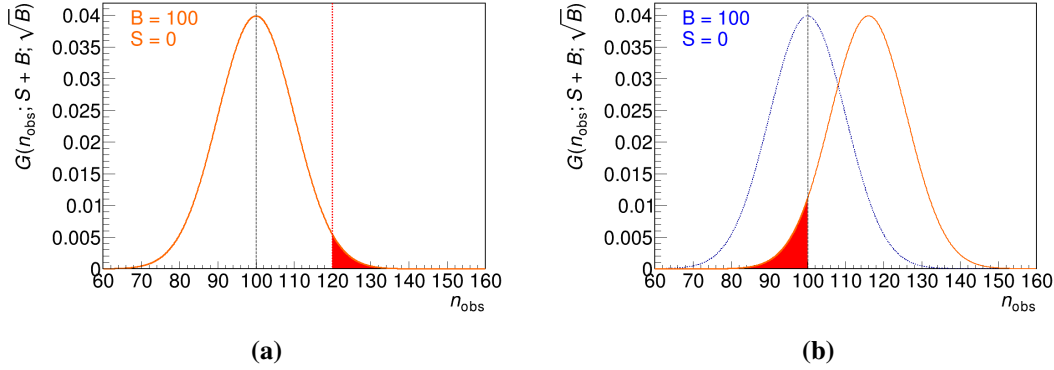
Here, obviously, we have  $\hat{S} = 20$ , which is in some way an exciting result: we have observed a positive  $\hat{S}$ , which seems to indicate that a signal is actually present.

#### 3.2 Significance and $p$ -value

Before getting too excited about this, we need to remember that  $\hat{S}$  is only an estimator, whose values reflect the fluctuations that can occur in the measured  $n$ . It could well be that  $S = 0$  (note that this is  $S$  without the hat, since here we refer to the true value); i.e. there is in fact no signal and the positive  $\hat{S}$  could come from an upward fluctuation in the background, compared to the  $B = 100$  expectation.

How likely is this? Suppose we are indeed in the background-only case,  $S + B = B = 100$ . Then the width of the Gaussian distribution of  $n$  is given by  $\sqrt{S + B} = \sqrt{B} = 10$ , and this gives the typical

size of the fluctuations of  $n$  around its mean value of 100. This is very relevant to our decision as to whether a true signal is indeed present: the observed value of  $\hat{S}$  is twice the typical size of fluctuations, which seems to indicate an outcome that is at least somewhat unusual.



**Fig. 4:** Distribution of the  $G(n; S + B, \sqrt{S + B})$  Gaussian PDF, where  $S$  and  $B$  are respectively the signal and background yields and  $n$  the observable. (a) Discovery scenario where the  $S = 0$  case is presented (orange line). The  $p$ -value for the case  $n_{\text{obs}} = 120$  is the area of the red shaded region. (b) Upper limit scenario, where  $n_{\text{obs}} = 100$ . The distribution for the  $S_{95} + B$  case is shown (red line), where  $S_{95}$  corresponds to the 95% CL upper limit on  $S$  (here  $S \approx 116$ ). The area of the red shaded region is the corresponding  $p$ -value, which is 5% by definition.

More generally, one can define the Gaussian *significance* as

$$z = \frac{\hat{S}}{\sqrt{B}} = \frac{n - B}{\sqrt{B}} \quad (14)$$

i.e. the ratio of the observed  $\hat{S}$  to the size of its statistical fluctuations. This has an intuitive meaning:  $|z| < 1$  corresponds to values of  $\hat{S}$  well within the statistical noise, while large values of  $|z|$  indicate that the observed  $\hat{S}$  likely cannot be explained by statistical fluctuations alone.

We can be a bit more precise by recalling the Gaussian quantiles shown in Table 1. From these, we can conclude that  $90 \leq n \leq 110$  corresponds to the  $\pm 1\sigma$  interval around the mean, and therefore should occur about 68.3% of the time. Similarly, we should have  $80 \leq n \leq 120$  about 95.5% of the time. In other words, observing  $|n - 100| \geq 20$  (i.e.  $|\hat{S}| \geq 20$ ) should occur about 4.5% of the time.

This probability is called the  $p$ -value with respect to the  $S = 0$  hypothesis. Generally, it is defined as the probability to get a result *at least as extreme* as the one that was observed, under the hypothesis one wishes to test (the *null hypothesis*). It is illustrated for the case  $\hat{S} > 0$  as the shaded region in Fig. 4a. This  $p$ -value provides a very quantitative way to decide whether signal is present or not: here it indicates that while  $\hat{S} = 20$  is not a typical outcome in the  $S = 0$  case, it is also not particularly rare, occurring about once every 20 attempts.

Table 1 gives the corresponding numbers for a few other values of  $n$ , both in terms of the significance and  $p$ -value  $p_0$  for the  $S = 0$  hypothesis. In the Gaussian case, the two are closely related through the Gaussian quantiles, since  $p_0$  corresponds to the tail probabilities of the normal distribution beyond  $\pm \hat{S}/\sqrt{B}$ . In terms of the CDF  $\Phi$  of the normal distribution introduced earlier in these lectures, we have

therefore

$$p_0^{2\text{-sided}} = 1 - \left[ \Phi\left(\frac{\hat{S}}{\sqrt{B}}\right) - \Phi\left(-\frac{\hat{S}}{\sqrt{B}}\right) \right] = 2\Phi\left(-\frac{\hat{S}}{\sqrt{B}}\right). \quad (15)$$

The  $p$ -value is denoted as *two-sided* for reasons that will be explained in the next section.

### 3.3 One-sided and two-sided tests

So far we have treated positive and negative values of  $\hat{S}$  on the same footing: i.e. we have defined  $p$ -values that apply both to  $n$  fluctuating above  $B$  (i.e. positive  $\hat{S}$ ) and below  $B$  (negative  $\hat{S}$ ). In high-energy physics, one can often assume that signal will give a positive contribution to the expected event yields (although negative signal yields can occur in some cases, e.g. due to interference effects).

If one knows a priori that  $S > 0$ , then one can restrict the considerations above to only the positive half of the Gaussian; i.e. consider that only  $\hat{S} > 0$  is a bona fide signal, while  $\hat{S} < 0$  is just another manifestation of the background-only hypothesis.

In this case, we consider only the upper tail of the Gaussian in the  $p$ -value is calculation, which now reads

$$p_0^{1\text{-sided}} = 1 - \Phi\left(\frac{\hat{S}}{\sqrt{B}}\right) = \Phi\left(-\frac{\hat{S}}{\sqrt{B}}\right). \quad (16)$$

This  $p$ -value is now denoted as *one-sided*, by opposition to the expression above, since we consider only one side of the Gaussian. This one-sided definition of the  $p$ -value corresponds to the shaded region in Fig. 4a. Compared to the two-sided case, one sees a simple factor-of-2 difference. Note also that the significance is defined in the same way as before. One- and two-sided  $p$ -values for specific significance levels are listed in Table 1.

We will use the one-sided definition of discovery  $p$ -values in the rest of the lectures, unless indicated otherwise.

### 3.4 Significance thresholds

In this Gaussian example, we can now determine how likely a given value of  $\hat{S}$  is to occur in the background-only hypothesis: either in terms of the  $p$ -value (smaller values indicating lower likelihood to occur) or significance (higher values indicate lower likelihood to occur).

In principle, this can be used to decide if one has observed a real signal or not, but there is some arbitrariness on what threshold is used for this purpose. In high-energy physics, one usually defines two thresholds:

- $3\sigma$  threshold ( $z \geq 3$ ), corresponding to *evidence* for new phenomena;
- $5\sigma$  threshold ( $z \geq 5$ ), corresponding to the *observation* (or discovery) of new phenomena.

In each case, one can also define the threshold in terms of the corresponding  $p$ -value: about 0.3% for evidence, and  $3 \times 10^{-7}$  for discovery. These thresholds are quite demanding: discovery corresponds to phenomena that only have about a chance in 3 million to occur in the background-only case. There are several reasons for these high thresholds [4]. The main one is the *look-elsewhere effect*: searches often target a range of signal configurations, for instance by looking for bumps over a range of mass

values. The probability for a fluctuation to occur *anywhere* in a spectrum can be much higher than at one given location, since mass positions separated by an interval larger than the experimental resolution can be considered largely uncorrelated. For this reason, the *global significance* accounting for these possibilities is lower than the *local significance* computed as described here, and fake “discoveries” due to fluctuations are more likely than one could naively estimate. One therefore needs to set a relatively high threshold for the local significance to avoid this. In any case, one should keep in mind that there always remains a chance (however small) that the observed signal is actually due to a very unlikely fluctuation.

Coming back to our example, we can conclude that while  $\hat{S} = 20$  is an intriguing result, it does not meet the criterion for evidence (which would require  $\hat{S} \geq 30$ ), nor the one for discovery ( $\hat{S} \geq 50$ ).

### 3.5 Confidence intervals

So far we have discussed the significance of a measured signal, with the aim of establishing a discovery. Another important class of results is *confidence intervals*, where we add an uncertainty band around the best-fit value of a parameter. This usually takes the form  $\mu = \hat{\mu}_{-\epsilon_-}^{+\epsilon_+}$ , where  $\hat{\mu}$  is the best-fit value of the measurement of  $\mu$  and  $\epsilon_{\pm}$  are the positive and negative uncertainties. This statement is made for a particular *confidence level* (CL). For a single parameter this is often set at the “ $1\sigma$ ” level, i.e. the 68.3% CL that corresponds to the  $1\sigma$  interquartile of a Gaussian distribution. The confidence interval is then defined as

$$p(\hat{\mu} - \epsilon_- \leq \mu \leq \hat{\mu} + \epsilon_+) = 68.3\%. \quad (17)$$

This states that there is a 68.3% chance that the true value  $\mu$  is contained in the confidence interval obtained in the measurement. A very important point is that the probability statement is about *the interval*, and not the true value  $\mu$ : recall that  $\mu$  is a fixed (unknown) value, with no associated probability distribution. What changes randomly from experiment to experiment is the data, and therefore the interval that we compute. Another way to state Eq. (17) is therefore that if we repeat our measurement many times, then the confidence intervals that we computed from each set of observed data will contain the true  $\mu$  68.3% of the time.

Consider a simple Gaussian case where we measure a parameter  $\mu$  using the observable  $m$ . The measurement PDF is  $G(m; \mu, \sigma)$ , and the Gaussian width  $\sigma$  is a known fixed value. Suppose that we observe  $m = m_{\text{obs}}$ , what is the  $1\sigma$  confidence interval on  $\mu$ ?

One knows from Gaussian quantiles that

$$p(\mu - \sigma \leq m_{\text{obs}} \leq \mu + \sigma) = 68.3\%. \quad (18)$$

This can be rewritten as

$$p(|\mu - m_{\text{obs}}| \leq \sigma) = 68.3\%, \quad (19)$$

which one can re-expand in the other direction as

$$p(m_{\text{obs}} - \sigma \leq \mu \leq m_{\text{obs}} + \sigma) = 68.3\%. \quad (20)$$



This is exactly the statement we were looking for: from  $m_{\text{obs}}$  we have computed the interval  $m_{\text{obs}} - \sigma \leq \mu \leq m_{\text{obs}} + \sigma$ , which covers  $\mu$  68.3% of the time. In the usual notation, we can write it as  $\mu = m_{\text{obs}} \pm \sigma$  at 68.3% CL.

### 3.6 Upper limits on a signal yield

The last class of results covered in these lectures is upper limits on signal yields. This is usually reported in the case where a search for new phenomena finds no evidence of its targeted signal, so that reporting a significance is not particularly useful. It allows us to set constraints on physics models that predict such signals, by stating that the true signal cannot be very large since we have not seen evidence of it. These upper limits are in fact one-sided confidence intervals on the true signal yield, with no lower bound. By convention, they are usually reported with a confidence level of 95%.

We can obtain such an upper limit by modifying slightly the example described in the previous section. First, we perform a small computation to determine the point at which the cumulative integral of a normal distribution reaches 5%. Using  $\Phi^{-1}$ , the inverse function of the Gaussian CDF, we find that  $\Phi^{-1}(0.05) \approx -1.64$ , which means that the integral from a point located about  $1.64\sigma$  below the Gaussian mean corresponds to 5% of the total integral. We can write this statement as

$$p(m_{\text{obs}} \geq \mu - 1.64\sigma) = 95\% \quad (21)$$

which can be flipped into

$$p(\mu \leq m_{\text{obs}} + 1.64\sigma) = 95\%. \quad (22)$$

This corresponds to the desired upper limit, i.e.  $\mu \leq m_{\text{obs}} + 1.64\sigma$  at 95% CL. In other words, if we set an upper limit on the signal yield  $\mu$  at a value of  $m_{\text{obs}}$  plus 1.64 times the uncertainty  $\sigma$ , then we know that the true value  $\mu$  will be below this upper limit 95% of the time on average. This is illustrated graphically in Fig. 4b: the Gaussian distribution for the  $S_{95} + B$  scenario, where  $S_{95}$  is the 95% CL upper limit, is shown in red. The shaded region on the left side of the curve amounts to 5% of the outcomes in this scenario, and this shows graphically that this  $S_{95}$  does correspond to a 95% CL upper limit as advertised.

## 4 Computing statistical results

In the previous section, we introduced the main classes of statistical results: parameter estimation (i.e. computing  $\hat{S}$ ); discovery significances and p-value; confidence intervals; and upper limits on signal yields. We also showed that in the simple one-bin Gaussian case, these quantities can be computed rather intuitively. However, we have seen in Section 2 that measurements are often described using much more complex statistical models, for instance with multiple bins and non-Gaussian behavior. The objective of this section is to present a general framework for computing these results, in principle applicable to models of arbitrary complexity. Of course, we will also check that it does give the same results as obtained above for the simple one-bin Gaussian case! The first two sub-sections below will present the general computation framework, while the rest of this section will focus on how to apply this to computing significances, confidence intervals and upper limits.

## 4.1 Maximum-likelihood estimation

### 4.1.1 Likelihood

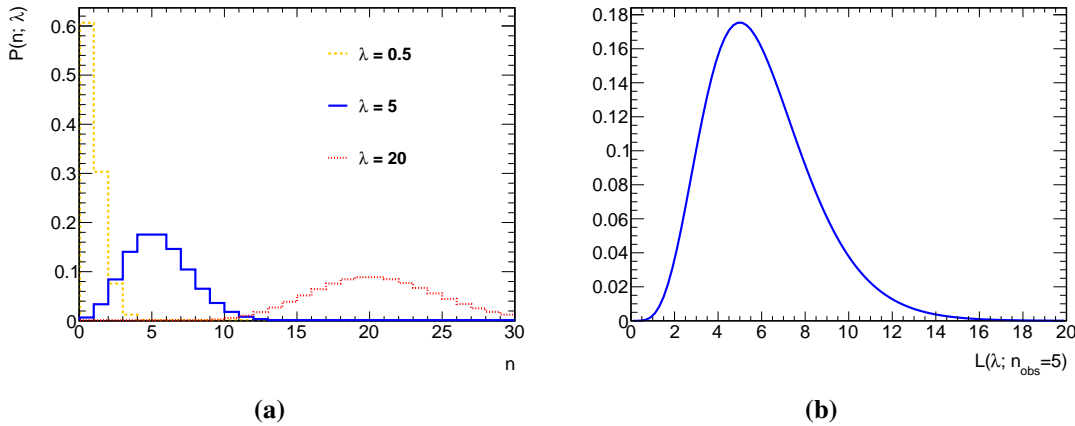
The statistical models described in Section 2 consist in two quantities: the PDF  $P(n; \alpha)$  for the measurement, where  $n$  represents the observables and  $\alpha$  the parameters; and the observed data  $n_{\text{obs}}$ . From these inputs, we would like to obtain an estimator  $\hat{\alpha}$  of the true parameter value  $\alpha$ .

We start by defining the *likelihood function* of  $\alpha$  as

$$L(\alpha) = P(n_{\text{obs}}; \alpha). \quad (23)$$

This is in a sense purely formal: the likelihood function is the same as the PDF but seen as a function only of the parameters (here  $\alpha$ ), and for the observables set to their observed values (here  $n = n_{\text{obs}}$ ). It is however an extremely useful quantity that will be used throughout the rest of these lectures.

The likelihood  $L(\alpha_0)$  can be understood as the probability to obtain the data that was observed, if the parameters have the value  $\alpha = \alpha_0$ . As illustrated in Figs. 5a and 5b, this allows us to assign a probability to the parameter values: some parameter values are *likely* in the sense that in this scenario would give rise to  $n_{\text{obs}}$  with a high probability; and other values are *unlikely* in the sense that  $n_{\text{obs}}$  would have a small probability of occurring in this case.



**Fig. 5:** (a) Poisson distributions  $\text{Pois}(n, \lambda)$  for expected yields  $\lambda = 0.5$  (orange),  $\lambda = 5$  (blue) and  $\lambda = 20$  (red), highlighting that the probability to obtain  $n = 5$  (i.e. the likelihood  $L(\lambda; n = 5)$ ) is highest for  $\lambda = 5$ . (b) Graph of  $L(\lambda; n = 5)$  as a function of  $\lambda$ .

### 4.1.2 Maximum-likelihood estimator

This suggests a general method for estimating  $\alpha$ : simply pick the value that gives the highest possible  $L(\alpha)$ . More formally,

$$\hat{\alpha} = \arg \max L(\alpha). \quad (24)$$

This defines the *maximum-likelihood estimator* (MLE) for  $\alpha$ , which we denote again with a hat.<sup>1</sup> Intuitively, we can already guess that the MLE will have good properties: by definition the  $n_{\text{obs}}$  that occur

<sup>1</sup>There is no ambiguity with the previous usage of the hat notation since, as we will see, those instances were, in fact, MLEs.

often are the ones with high  $P(n_{\text{obs}}; \alpha_{\text{true}})$  for the true values  $\alpha_{\text{true}}$  of the parameters. The likelihood  $L(\alpha_{\text{true}})$  will therefore be high in general and therefore  $\hat{\alpha}$  should generally come out quite close to  $\alpha_{\text{true}}$ . The cases where this works less well are those where the observed data is atypical due to a large statistical fluctuation, which translates into low values of  $P(n_{\text{obs}}; \alpha_{\text{true}})$ . However, these cases are rare by definition, so that the MLE  $\hat{\alpha}$  remains a good guess on average.

More formally, the MLE has good statistical properties for very general classes of likelihoods. One can show, in particular, that in the limit of sufficiently large event samples, the MLE is *efficient*, in the sense that its uncertainty is as small as it can get (i.e. matches the limit given by the Cramér-Rao bound); and it is also *unbiased*, i.e. its average over many trials tends toward the true parameter values. More details on the properties of MLEs can be found, for instance, in Ref. [3].

Given these good properties, we will use MLEs to estimate parameter values throughout the rest of these lectures. Before moving to the next topic, we provide examples of MLEs in two simple cases.

#### 4.1.3 Application to the one-bin Gaussian example

Going back to the one-bin Gaussian example of Section 2.1.2.1, the likelihood is defined in this case as

$$L(S) = P(n_{\text{obs}}; S, B) = G(n_{\text{obs}}; S + B, \sqrt{S + B}). \quad (25)$$

Since the Gaussian has a maximum at  $S + B = n_{\text{obs}}$ , one concludes that the MLE corresponds to the value  $\hat{S}$  such that

$$\hat{S} = n_{\text{obs}} - B, \quad (26)$$

which matches the naive estimation in this case. In other words, the general framework of the MLE provides the same numerical answer as obtained in Section 3.1.

#### 4.1.4 Application to multi-bin Gaussian measurements

We now consider the case of a measurement in  $N_{\text{bins}}$  independent bins. Each bin  $i$  consists in a Gaussian measurement with an observed value  $n_i$ , a width  $\sigma_i$  and an expected value given by  $\nu_i(\alpha)$  as a function of the model parameters  $\alpha$ . The total PDF is

$$P(\{n_i\}; \alpha) = \prod_{i=1}^{N_{\text{bins}}} G(n_i; \nu_i(\alpha), \sigma_i). \quad (27)$$

It is often useful to define the *negative twice log-likelihood* (N2LL) as

$$\lambda(\alpha) = -2 \log L(\alpha). \quad (28)$$

Since  $-2 \log$  is a monotonically decreasing function, the MLE can be equivalently obtained by minimizing  $\lambda(\alpha)$ .

This is a useful procedure in particular for Gaussian PDFs such as the one considered here, since

we have<sup>2</sup>

$$\lambda(\alpha) = -2 \log L(\alpha) = -2 \log P(\{n_i\}; \alpha) \quad (29)$$

$$= \sum_{i=1}^{N_{\text{bins}}} -2 \log G(n_i; \nu_i(\alpha), \sigma_i) \quad (30)$$

$$= \sum_{i=1}^{N_{\text{bins}}} \left( \frac{n_i - \nu_i(\alpha)}{\sigma_i} \right)^2. \quad (31)$$

The quantity on the last line is known as the  $\chi^2$  of the  $n_i$  with respect of the prediction  $\nu_i(\alpha)$ , i.e. the sum of the squares of the corresponding pulls, as defined in Section 2.1.2.1. It is often used in so-called  $\chi^2$  fits, in which one adjusts the model parameters to get the smallest  $\chi^2$  with respect to the data.

We have seen that the MLE  $\hat{\alpha}$  is the value that minimizes  $\lambda(\alpha)$ : it is therefore also the value that minimizes the  $\chi^2$ , and thus corresponds to the  $\chi^2$  best-fit value in this Gaussian situation. This illustrates the notion that the MLE matches, in general, the best-fit values of the model parameters to the data.

For example, in the ROOT software, which is widely used in high-energy physics, fitting a histogram to a model prediction is by default done using a  $\chi^2$ . For non-Gaussian cases, one can also use the *likelihood fit* option, which performs a maximum-likelihood estimation based on a model where each bin is described by a Poisson PDF. In ROOT, as in other similar software, fitting a model to data therefore exactly corresponds to performing a MLE.

## 4.2 Testing hypotheses

Now that we have a well-defined method for estimating parameter values, we turn again to the problem of determining whether or not an observed signal yield is in fact significant. This implies computing significances and  $p$ -values as in Section 2.1.2.1, but now for arbitrary statistical models.

### 4.2.1 Tests and errors

What we want to do is in fact to *test a hypothesis*, defined as a set of values for the model parameters. The hypothesis under test is usually referred to as the *null hypothesis*. In the case of discovery, we want to test the null hypothesis  $H_0$  defined by  $S = 0$ , where  $S$  is the signal yield. This hypothesis can in fact be true (i.e. the signal  $S$  does not actually exist) or false (there is actually a non-zero  $S$ ).

Testing the hypothesis  $H_0$  means using the data to come to a decision as to whether it is true or false. There can be, therefore, four possible outcomes. In two of them, the conclusion is correct:

- $H_0$  is false (i.e.,  $S$  exists) and from the data we decided that  $H_0$  was likely false. This is a very positive outcome, where there was a signal to be found and it was successfully detected by the experiment. If the signal is large enough, we have made a discovery.
- $H_0$  is true (i.e., no  $S$ ), and from the data we decided that  $H_0$  was likely true. This is not a very exciting outcome since nothing was found, but we arrived to the correct conclusion that no signal is present.

<sup>2</sup>Note that we have dropped the prefactor in the Gaussian, which would give an additive constant term in  $\lambda(\alpha)$ : since we ultimately wish to minimize  $\lambda(\alpha)$ , this term is irrelevant.

There are however two more outcomes, which correspond to *errors*, in the sense that the test reached the wrong conclusion:

- $H_0$  is true (no  $S$ ) and from the data we decided that  $H_0$  was likely false ( $S$  exists). This is a very embarrassing outcome, where the experimental result is the “discovery” of a signal that does not actually exist. This can create some short-lived excitement but inevitably gets falsified when eventually other experiments fail to reproduce the spurious discovery. This error is called a Type-I error and the probability for it to occur if  $H_0$  is true is called the  $p$ -value<sup>3</sup>. Since Type-I errors are often quite embarrassing for the experimenter, it is important to ensure that their rate (the  $p$ -value) is small.
- $H_0$  is false ( $S$  exists) and from the data we decided that  $H_0$  was likely true. This is another incorrect conclusion, where there was signal to be found but the experiment missed it. This is called a Type-II error and it is again best avoided.

Given these possible outcomes, our goal is to design an optimal test that will lead to minimal rates for both Type-I and Type-II errors. In practice, this is usually done by defining a discriminant, i.e. a function of the observables which has a different distribution in the cases when  $H_0$  is true or false. Ideally, it also captures all or most of the information present in the data to separate these two cases. This discriminant is called the *test statistic*. The result of the test is then based on the value of the test statistic, as illustrated on the left panels of Fig. 6: for instance if a true  $H_0$  corresponds to larger values of a test statistic  $q$  and vice-versa, one would declare that  $H_0$  is likely true if  $q > Q_{\text{thresh}}$  for some threshold value  $Q_{\text{thresh}}$ .

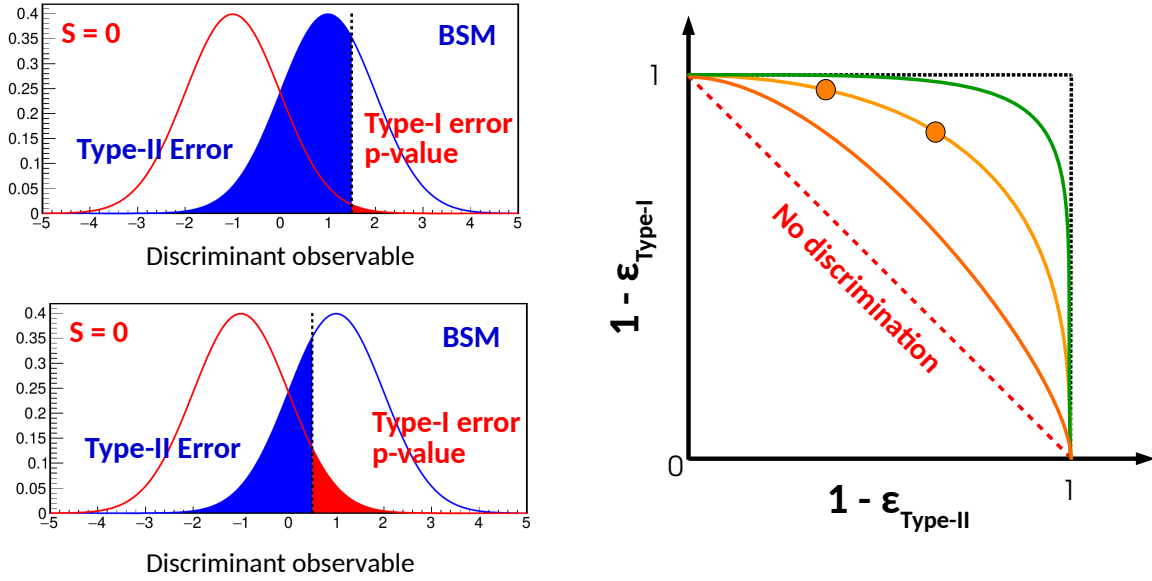
The choice of  $Q_{\text{thresh}}$  determines the Type-I and Type-II error rates. As illustrated on the left panels of Fig. 6, raising  $Q_{\text{thresh}}$  will tighten the test, making it less likely to find a signal whether it is there or not: the Type-I rate will therefore decrease but the Type-II error rate will increase. Lowering  $Q_{\text{thresh}}$  will give a looser test and the opposite behavior. The relation between the two is given by the *ROC curve* shown on the right side of Fig. 6: by changing the threshold one moves along the curve, but one cannot reach arbitrarily small values for both error rates. An optimal test relies on an optimal test statistic, i.e. a discriminant that achieves the best possible separation between the two cases. But this optimal test still cannot yield arbitrarily small error rates of both kinds since lowering one rate raises the other. In fact these rates are bounded from below by the information present in the data and an optimal test is one where this information is captured by the discriminant.

#### 4.2.2 The Neyman–Pearson lemma

At face value, finding such an optimal discriminant is a difficult problem: the test statistic should somehow capture all the information present in the measurement, spanning all the measurement regions and accounting for the distributions of the signals and backgrounds in each case. Fortunately, such a discriminant is provided quite simply in many cases by the *Neyman–Pearson lemma*. This states that when

---

<sup>3</sup>We will see shortly that this coincides with the definition we gave earlier in the Gaussian example.



**Fig. 6:** Representation of the performance of a hypothesis test. The test is defined in terms of the hypotheses  $S = 0$  (i.e. no signal present) and BSM (presence of a signal of physics beyond the Standard Model). The distributions of a discriminant observable under each hypothesis are shown in the two plots on the left. The shaded areas correspond to the Type-I error or  $p$ -value  $\epsilon_{\text{Type-I}}$  (red area) and the Type-II error  $\epsilon_{\text{Type-II}}$  (blue area). In the top and bottom plots, these areas are shown for two different values of the threshold which defines the test (i.e. the BSM hypothesis is chosen for values of the discriminant above the threshold, and the  $S = 0$  hypothesis for values below). The plot on the right shows the ROC curve of the test (yellow line), i.e. the values of  $1 - \epsilon_{\text{Type-I}}$  as a function of  $1 - \epsilon_{\text{Type-II}}$  as the test threshold varies. The situations shown in the plots on the left correspond to the leftmost (top plot) and rightmost (bottom plot) markers on the curve. The orange and green lines correspond to hypothetical situations obtained with, respectively, a less powerful and more powerful discriminant than the one shown here. The dotted red line corresponds to the limiting case, where the discriminant has no sensitivity to the hypotheses.

choosing between two hypotheses  $H_0$  and  $H_1$ , the optimal discriminant is in fact the likelihood ratio

$$\frac{L(\alpha_{H_0})}{L(\alpha_{H_1})} \quad (32)$$

where  $\alpha_{H_0}$  and  $\alpha_{H_1}$  are the parameter values that define  $H_0$  and  $H_1$ , respectively. Note that one does not test  $H_0$  in an absolute sense, but only with respect to an *alternate hypothesis*  $H_1$ .

Just like for the MLE, the Neyman–Pearson lemma can be understood intuitively as following the data: if the data was in fact generated for  $\alpha = \alpha_{H_0}$ , then by definition there is a high probability that  $L(\alpha_{H_0})$  is high and, therefore, that the likelihood ratio takes large values. Conversely, if  $\alpha = \alpha_{H_1}$  then it is  $L(\alpha_{H_1})$  that will take large values and the likelihood ratio will be small. In both cases data fluctuations can lead to the opposite behavior, but these cases occur by definition with low probability. We will not provide a formal proof of the Neyman–Pearson lemma here, but hopefully these arguments make it clear that the likelihood ratio has the right properties for an optimal discriminant. The proof and more details can be found, e.g., in Ref. [2].

The likelihood ratio is optimal in the sense that if we choose a given rate of Type-I error (for instance by adjusting the threshold for the test), then the rate of Type-II errors will be the smallest possible, given the information present in the measurement. In the rest of these lectures we will therefore mostly ignore Type-II error rates: we will instead focus on the Type-I rate (the  $p$ -value), and trust that the Type-II rate is as small as can be thanks to the optimality guaranteed by the Neyman–Pearson lemma.

### 4.3 Discovery testing

#### 4.3.1 The likelihood ratio test statistic

Having established the general framework for hypothesis testing, we can now go back to more practical matters and apply it to the case of discovery testing already covered in Section 3.2 for the special case of a one-bin Gaussian measurement. Here we make a much more general assumption that the measurement is described by a PDF  $p(n; S)$ , in terms of the observables  $n$  and a signal yield parameter  $S$ , and that we have observed  $n = n_{\text{obs}}$ . We define as usual the likelihood as  $L(S) = p(n_{\text{obs}}, S)$ .

Since we want to test for the presence of signal, we define our null hypothesis  $H_0$  to be the case  $S = 0$ . To use the Neyman–Pearson lemma, we need to also define an alternate hypothesis  $H_1$  that will be tested against  $H_0$ . Here we take  $H_1$  to correspond to  $S > 0$ , using a one-sided definition that assumes positive signals.<sup>4</sup>

To compute the numerator of the likelihood ratio, we can simply use  $L(S = 0)$ . For the denominator, we need to choose the value of  $S$  that will represent the  $S > 0$  hypothesis: an obvious choice is to select  $\hat{S}$ , in the case where  $\hat{S} > 0$ . If  $\hat{S} < 0$ , then in keeping with our one-sided assumption we take this to be identical with  $\hat{S} = 0$  (no evidence of signal). Since this is the same as the numerator, the likelihood ratio is simply 1 in this case.

We add one final ingredient: as mentioned in Section 2.1.3, it is often practical to consider  $-2 \log L$  instead of just  $L$ , in particular in the often-seen cases where  $L$  is approximately Gaussian. We therefore define our discriminant as

$$q_0 = \begin{cases} -2 \log \frac{L(S = 0)}{L(\hat{S})} & \text{if } \hat{S} > 0 \\ 0 & \text{if } \hat{S} \leq 0. \end{cases} \quad (33)$$

One can see immediately that  $q_0 \geq 0$ : since  $\hat{S}$  is the MLE, by definition  $L(\hat{S}) > L(S = 0)$ , so that the likelihood ratio in Eq. (33) is negative and  $q_0$  is positive. Furthermore,  $q_0 = 0$  indicates the absence of signal: in this case  $L(S = 0)/L(\hat{S}) = 1$ , so that the best-fit likelihood  $L(\hat{S})$  is identical to that of the background-only hypothesis. Conversely, large values of  $q_0$  indicate the presence of signal: large  $q_0$  means small  $L(S = 0)/L(\hat{S})$ , which in turns means  $L(\hat{S}) \gg L(S = 0)$ : this indicates a strong preference of the data for a  $\hat{S}$  away from 0, and therefore that a signal seems to be present. Setting  $q_0 = 0$  for  $\hat{S} \leq 0$  identifies this case with the absence of signal  $S = 0$ , as mentioned above.

<sup>4</sup>This hypothesis is *composite*, in the sense that it encompasses a range of values of  $S$ . The proof of the Neyman–Pearson lemma applies only *simple* hypotheses corresponding to a single point in parameter space, so that the likelihood ratio is not guaranteed to be optimal in this case [2]. However, in practice this is seen to remain close to being true in many cases.

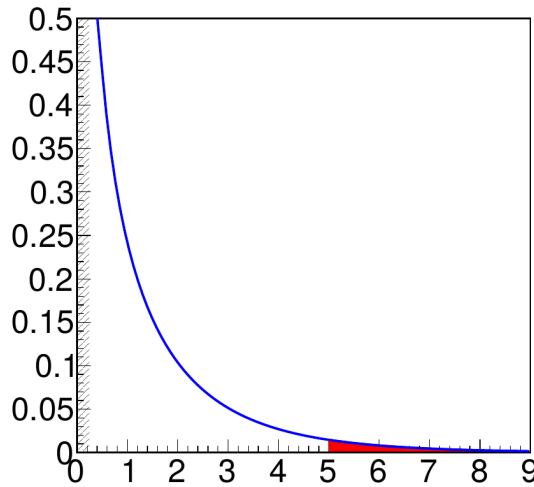
### 4.3.2 The discovery $p$ -value

The test statistic  $q_0$  discriminates between signal and background, but like in any test we can sometimes come to the wrong conclusion based on the observed data. For the discovery case, the main issue is the case of a spurious discovery, when in fact  $S = 0$  (i.e.,  $H_0$  is true), but a large value of  $q_0$  leads to the incorrect conclusion that signal is in fact present. Looking back at the definitions in Section 4.2.1, we see that this corresponds to a Type-I error, and the probability for it to occur under the  $S = 0$  hypothesis is the  $p$ -value.

Graphically, the  $p$ -value can be seen as the tail integral of the PDF of  $q_0$  under the  $S = 0$  hypothesis, as illustrated in Fig. 7. Under  $S = 0$ , the value  $q_0^{\text{obs}}$  observed in data will be usually close to 0, but will occasionally reach higher values if signal-like fluctuations are present. The probability to observe a false discovery at the level of  $q_0^{\text{obs}}$  or higher is given by the tail integral

$$p_0 = \int_{q_0^{\text{obs}}}^{\infty} f(q_0; S = 0) dq_0 \quad (34)$$

where  $f(q_0; S = 0)$  is the distribution of  $q_0$  under the  $S = 0$  hypothesis. This provides the general definition of the  $p$ -value  $p_0$ .



**Fig. 7:** PDF  $f(q_0; S = 0)$  of the test statistic  $q_0$  defined by Eq.( 33) under the  $S = 0$  hypothesis, in the asymptotic approximation. Values of  $q_0$  are on the  $x$ -axis, while the  $y$ -axis gives the values of the half- $\chi^2$  distribution  $\frac{1}{2}f_{\chi^2(n=1)}$ . The hatched region near  $q_0 = 0$  represents the delta function  $\delta(q_0)$  corresponding to the case  $\hat{S} \leq 0$ . The  $p$ -value for the case of  $q_0^{\text{obs}} = 5$  is the area of the red shaded region.

### 4.3.3 The distribution of $q_0$

We are missing a critical ingredient to compute the  $p$ -value (and therefore the significance), namely the distribution  $f(q_0; S = 0)$ . In general, this is a difficult problem since  $q_0$  derives from the usually complex expression for  $L$ . However, one can show that if the measurement is Gaussian, then  $f(q_0; S = 0)$  can be simply expressed in terms of the  $\chi^2$  distribution  $f_{\chi^2(n_f)}$  introduced in Section 2.1.3, with a number of degrees of freedom  $n_f$  equal to the number of parameters used to define  $H_0$  (i.e., one parameter,  $S$ , in the



case we consider here). This result is called *Wilks' theorem* [13]. In a two-sided situation, the distribution of the test statistic would be exactly  $f_{\chi^2(n_f)}$ . In the one-sided situation shown here, the distribution is a “half- $\chi^2$ ”, with the  $\hat{S} > 0$  half of Eq. (33) described by a  $f_{\chi^2(n_f)}$  while the  $\hat{S} \leq 0$  half is represented by a delta function at 0. The distribution is illustrated in Fig. 7.

It is an *asymptotic approximation*, in the sense that it becomes valid in the limit of large event yields, since by the central-limit theorem PDFs generally tend to the Gaussian limit in this case. However, this does not mean that this only applies to the Gaussian case: the key point is that this Gaussian assumption only applies to one part of the computation, namely the distribution  $f(q_0; S = 0)$ . The computation of  $q_0^{\text{obs}}$  itself is performed using the exact form of  $L$ , and therefore accounts for non-Gaussian behavior. For this reason, the asymptotic approximation remains valid over a surprisingly wide range of situations. We will see in Section 4.3.6 that this limit is in fact often already valid for small yields, of order 5 to 10 events.

In the cases where the measurements are so non-Gaussian that Wilks' theorem does not provide a sufficient approximation (e.g. for very small expected event yields), other methods are required to obtain  $f(q_0; S = 0)$ . One solution is to sample  $f(q_0; S = 0)$  using pseudo-experiments: in this case the PDF  $p(n; S = 0)$  is used to generate random datasets, for which the computation of  $q_0$  is performed in the same way as for real data. The distribution of the resulting  $q_0$  values provides an approximation to  $f(q_0; S = 0)$ , which improves as more pseudo-experiments are generated. This procedure can however be quite CPU-intensive, especially to determine the tail of  $f(q_0; S = 0)$  when computing small  $p$ -values.

#### 4.3.4 The $p$ -value and significance under the asymptotic approximation

If we assume that  $q_0^{\text{obs}}$  follows its asymptotic half- $\chi^2$  distribution, then one can compute the  $p$ -value  $p_0$  of a positive signal  $\hat{S}$  with respect to the  $S = 0$  hypothesis as [12]

$$p_0 = \frac{1}{2} \left[ 1 - F_{\chi^2(n_f)}(q_0^{\text{obs}}) \right] \quad (35)$$

and its significance  $z$  as

$$z = \Phi^{-1}(1 - p_0). \quad (36)$$

In Eq. (35),  $F_{\chi^2(n_f)}$  is the cumulative distribution function  $F_{\chi^2(n_f)}(q_0) = \int_0^{q_0} f_{\chi^2(n_f)}(q) dq$ , which is directly related to the tail integral of  $f_{\chi^2(n_f)}$ . The factor  $1/2$  is due to the half- $\chi^2$  nature of the distribution discussed above and ultimately comes from the one-sided nature of the test.

These formulas take a simpler form in the case of a single parameter of interest,  $n_f = 1$ : a  $\chi^2$  observable for a single degree of freedom is by definition the square of a normal observable, so that one has  $F_{\chi^2(1)}(q_0) = \Phi(\sqrt{q_0})$ . Therefore, for  $n_f = 1$ ,

$$p_0 = 1 - \Phi\left(\sqrt{q_0^{\text{obs}}}\right) \quad (37)$$

and

$$z = \sqrt{q_0}. \quad (38)$$

The asymptotic expression for  $z$  in terms of  $q_0$  is, therefore, particularly simple in this case: one simply

needs to take the square root of  $q_0$  to obtain  $z$ .

#### 4.3.5 The one-bin Gaussian example

We consider the case of a one-bin Gaussian measurement with fixed background  $B$ , where a measured event count  $n$  is used to obtain the signal yield  $S$ . The measurement PDF is  $p(n; S) = G(n; S + B, \sqrt{S + B})$  and we assume that we measured  $n = n_{\text{obs}}$ .

In Section 4.1.3, we have already computed  $\hat{S}$ , the  $p$ -value  $p_0$  of the  $S = 0$  hypothesis and the significance  $\hat{S}$  of the signal using elementary methods. We now check that the general methods described in this section give the same results.

For simplicity, we will work with the N2LL

$$\lambda(S) = -2 \log L(S) = \frac{(n - (S + B))^2}{S + B} \quad (39)$$

The MLE  $\hat{S}$  is obtained by finding the minimum of  $\lambda(S)$ , i.e. by solving  $\partial\lambda(\hat{S})/\partial S = 0$ . A simple computation yields

$$\hat{S} = n_{\text{obs}} - B \quad (40)$$

as expected. We can now compute  $q_0$ , which is simply expressed in terms of  $\lambda$  as

$$q_0^{\text{obs}} = \lambda(S = 0) - \lambda(\hat{S}) \quad (41)$$

for  $\hat{S} > 0$ . Plugging in the expressions for  $\lambda$  and  $\hat{S}$ , one obtains

$$q_0^{\text{obs}} = \frac{(n_{\text{obs}} - B)^2}{B} = \left( \frac{\hat{S}}{\sqrt{B}} \right)^2, \quad (42)$$

again assuming  $\hat{S} > 0$ . Using Eqs. (37) and (38), one recovers the expressions

$$p_0 = 1 - \Phi \left( \frac{\hat{S}}{\sqrt{B}} \right) \quad (43)$$

$$z = \frac{\hat{S}}{\sqrt{B}} \quad (44)$$

that were already obtained in Section 3.2. Reassuringly, the general framework therefore yields in this simple situation the same results as those obtained using more pedestrian methods.

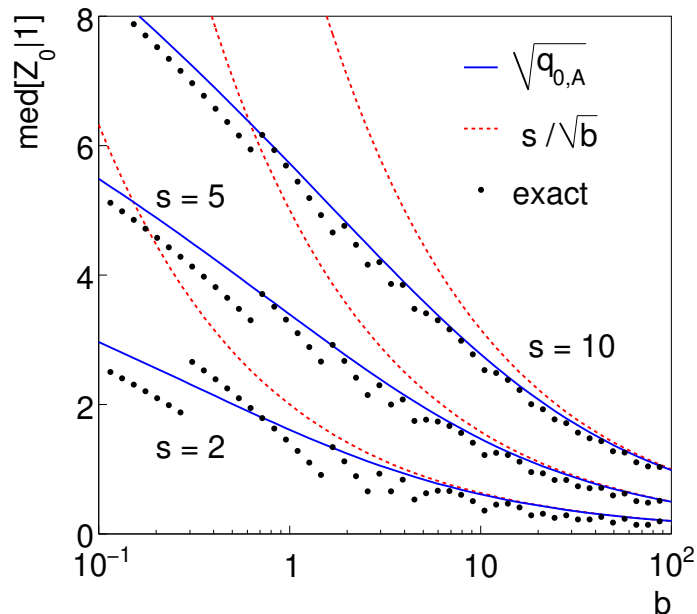
#### 4.3.6 Asymptotic significance for a Poisson measurement

We can apply the same treatment as in the previous section to a measurement described by a Poisson measurement,  $p(n; S) = \text{Pois}(n; S + B)$ . Repeating the same computation in this case, we obtain

$$z = \sqrt{2 \left[ (\hat{S} + B) \log \left( 1 + \frac{\hat{S}}{B} \right) - \hat{S} \right]}. \quad (45)$$

Note that this is obtained using the asymptotic formula  $z = \sqrt{q_0}$ , which assumes Gaussian behavior for this particular step of the computation, but  $q_0$  itself is computed using the Poisson expression. This procedure illustrates the principle behind the asymptotic approximation: the exact (potentially non-Gaussian) PDF of the measurement is used to compute  $q_0$ , but the Gaussian approximation is used to convert  $q_0$  into a significance or a  $p$ -value.

In this particular example, we can check the validity of the asymptotic approximation. Figure 8 shows a comparison of the significance computed using Eq. (45), the fully Gaussian version of Eq. (44), and the exact value (obtained using pseudo-experiments, as discussed in Section 4.3.3), for different values of  $S$  and  $B$ . The results show that the asymptotic Eq. (45) provides a much closer approximation to the exact result than the fully Gaussian form of Eq. (44), and that the approximation remains excellent even for small yields of 5 events or so.



**Fig. 8:** Median significance for a counting experiment with varying numbers of signal and background events. Results are computed using Eq. (44) (dotted red) and Eq. (45) (solid blue), and compared to the exact results computed from pseudo-experiments (black dots). Figure taken from Ref. [12].

## 4.4 Confidence intervals on a model parameter

### 4.4.1 Definition

As already mentioned in Section 3.5, an important class of physics results is *confidence intervals* set on a model parameter  $\mu$ . For a single parameter, they are usually written as in the form  $\mu = \hat{\mu}_{-\epsilon_-}^{+\epsilon_+}$ , where the best-fit value  $\hat{\mu}$  is the central value of the interval and  $\epsilon_{\pm}$  are the positive and negative errors.

Before moving to computations, it is useful to first clarify what we mean exactly by these intervals. First, intervals are accompanied by a probability value called a *confidence level* (CL). For a single parameter this is often set at the "1 $\sigma$ " level, i.e. a confidence level of 68.3% that corresponds to the 1 $\sigma$  inter-quantile of a Gaussian distribution.

We write for instance  $\mu = \hat{\mu}_{-\epsilon_-}^{+\epsilon_+}$  at 68.3% CL, which is the statement that

$$p(\hat{\mu} - \epsilon_- \leq \mu \leq \hat{\mu} + \epsilon_+) = 68.3\%, \quad (46)$$

i.e. that there is a 68.3% chance that the true value  $\mu$  is contained in the stated interval. It is worth noting that the probability statement is about the interval and not  $\mu$  itself: recall that  $\mu$  is a fixed (unknown) value, with no associated probability distribution. What changes from experiment to experiment is the data and therefore the interval that we compute. Another way to state Eq. (46) is, therefore, that the confidence interval that we built for a given observed data will cover the true value  $\mu$  in 68.3% of cases, if we perform the same experiment many times.

#### 4.4.2 The likelihood ratio for intervals

Several methods can be used to compute confidence intervals. A popular one is the *Neyman construction* which is elegant and works very well for small numbers of parameters. It is however difficult to use for larger parameter counts, so we will focus on a different method based on similar principles as those used for discovery testing, namely likelihood ratios.

The basic idea is that defining a confidence interval amounts to finding a range of parameter values that are compatible with the observed data. This in turn can be expressed as a hypothesis test: we define  $H_0(\mu_0)$  as the hypothesis that  $\mu = \mu_0$  and test this against the alternate hypothesis  $\mu \neq \mu_0$ , for an arbitrary value  $\mu_0$ . The values  $\mu_0$  for which  $H_0(\mu_0)$  is likely true will be part of the confidence interval, and vice versa.

The test is naturally two-sided: values of  $\mu$  away from  $\mu_0$  can be either above it or below ( $\mu$  is not necessarily an event yield and can in principle take arbitrary positive or negative values). As before, we will perform the test using the likelihood ratio test statistic. The alternate hypothesis  $\mu \neq \mu_0$  corresponds to a range of values and we need to decide which representative value to use to compute the corresponding likelihood. As before, we choose the best-fit value  $\hat{\mu}$  for this purpose. With the usual  $-2 \log$  modification, the test statistic is then

$$t(\mu_0) = -2 \log \frac{L(\mu = \mu_0)}{L(\hat{\mu})}. \quad (47)$$

Its values are always positive, with a minimum at  $\mu_0 = \hat{\mu}$ . Small values indicate good agreement with the  $\mu = \mu_0$  hypothesis. This agreement is maximal at  $\mu_0 = \hat{\mu}$  and typically gets worse as  $\mu_0$  moves away from  $\hat{\mu}$ . It therefore seems sensible to define the confidence interval with CL  $c$  as the range of values  $\mu_0$  such that  $t(\mu_0) \leq T(c)$ , where  $T(c)$  is a suitable threshold that rises with the confidence level  $c$ .

#### 4.4.3 Asymptotic approximation

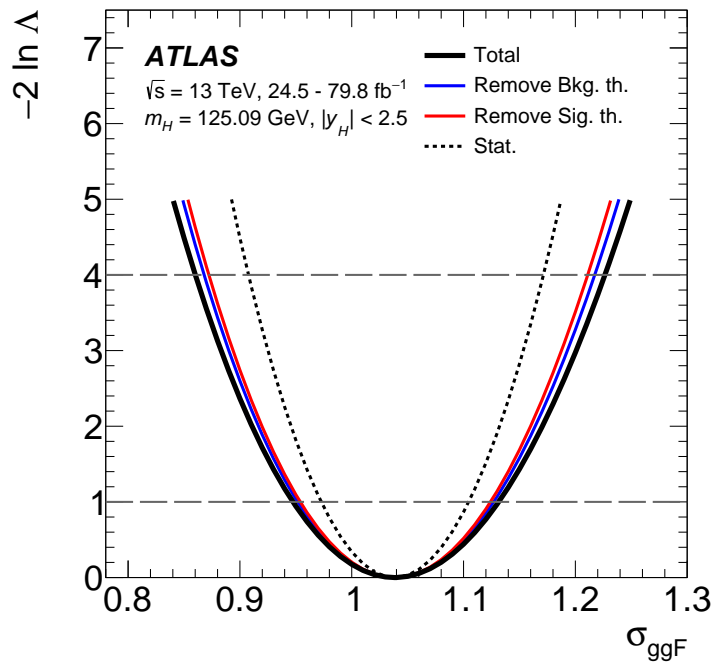
To define  $T(c)$ , we go back to the one-bin Gaussian case discussed in Section 3.5. The likelihood for  $\mu$  is  $L(\mu) = G(m_{\text{obs}}; \mu, \sigma)$  and a short computation shows that, in this case,

$$t(\mu_0) = \left( \frac{\mu_0 - \hat{\mu}}{\sigma} \right)^2. \quad (48)$$

Therefore  $t(\mu_0)$  follows a parabolic shape with a minimum at  $\mu_0 = \hat{\mu}$  and the condition  $t(\mu_0) \leq T(c)$  leads to the confidence interval  $\hat{\mu} \pm \sqrt{T(c)}\sigma$ .

We know from the computation of Section 3.5 that, in this simple Gaussian case, the  $1\sigma$  intervals should be  $\hat{\mu} \pm \sigma$ , and this suggests to use a threshold of  $T(68.3\%) = 1$  in this case: the  $1\sigma$  confidence interval is therefore defined as the range of  $\mu_0$  for which  $t(\mu_0) \leq 1$ . Similarly, a  $2\sigma$  interval would be defined by  $t(\mu_0) \leq 4$ , and so on.

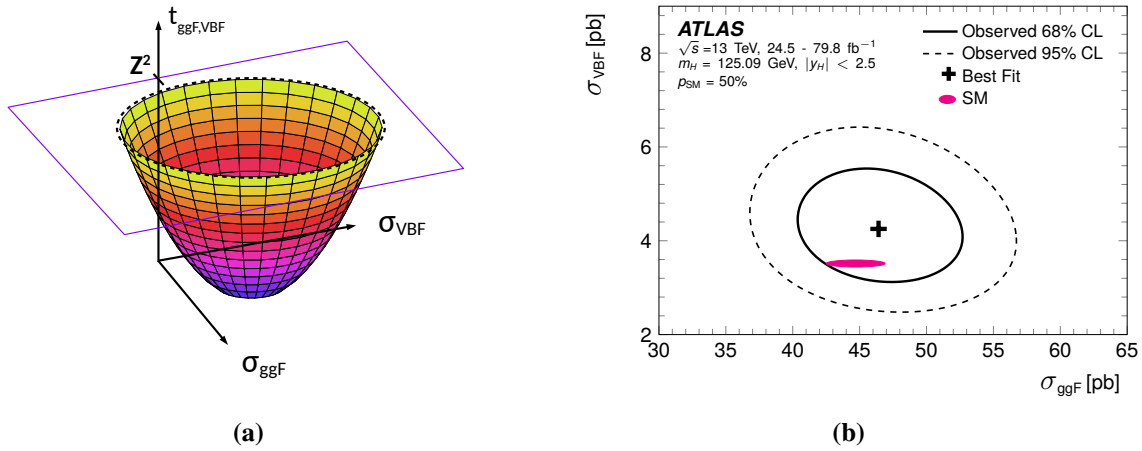
This is a suitable generalization of the results of Section 3.5, which matches the simple computation in the Gaussian case but is applicable to arbitrary forms of  $L$ . These *likelihood intervals* are another example of an asymptotic approximation, in the sense that the computation is exact only in the Gaussian limit. However, this again only applies to the distribution of  $t(\mu_0)$ , since  $t(\mu_0)$  itself is computed from the exact form of  $L$  including non-Gaussian effects. For this reason, the computation remains valid for a wide range of non-Gaussian situations. In practice, this means likelihood scans that are not quite parabolic, as they would be in the Gaussian case, but for which one can still compute confidence intervals by computing the intersections of the scan with the appropriate threshold  $T(c)$ . An example of the application of this method to a real-life example is shown in Fig. 9.



**Fig. 9:** Likelihood scans for the measurement of a Higgs boson production cross-section, taken from Ref. [14]. The scan corresponds to scenarios in which various combinations of measurements uncertainties are considered. The intersections of the scan with horizontal dotted lines at  $y$ -values of 1 and 4 define the endpoints of, respectively, the  $1\sigma$  and  $2\sigma$  confidence intervals on the parameter.

The method can also be extended for larger numbers of parameters, as illustrated in Figs. 10a and 10b for the case of a confidence contour in two dimensions. Since the relevant asymptotic distribution is now a  $\chi^2$  with two degrees of freedom, the thresholds  $T(c)$  differ from the case of a single parameter:

for instance  $1\sigma$  contours correspond to a threshold of about 2.30.



**Fig. 10:** (a) Illustration of the method used to obtain two-dimensional likelihood contours for two model parameters denoted as  $\sigma_{\text{ggF}}$  and  $\sigma_{\text{VBF}}$ . The N2LL  $t(\sigma_{\text{ggF}}, \sigma_{\text{VBF}})$  defines a surface that has a paraboloid shape for Gaussian likelihoods. Likelihood contours are obtained by intersecting this shape with a plane  $t(\sigma_{\text{ggF}}, \sigma_{\text{VBF}}) = Z$  at the appropriate level  $Z$ . For  $1\sigma$  intervals, this level is about 2.30. (b) Example of a real application of this method, taken from Ref. [14].

## 4.5 Upper limits on a signal yield

Finally, we come back to the question of setting *upper limits*, as already introduced in Section 3.6. These are in fact just one-sided confidence intervals set on a quantity such as a signal yield  $S$ , which is known to be positive. They are typically set in the case where the observed signal is small: in this case, rather than reporting a small significance, or a signal yield with a large uncertainty, it is often more useful to frame the result as the exclusion of some large signal hypotheses. Our goal is therefore to be able to state that we can exclude  $S < S_{1-\alpha}$  at a confidence level  $1 - \alpha$ , i.e. that if in fact  $S > S_{1-\alpha}$  then the probability to have observed a signal as small as the one that we did obtain is no more than  $\alpha$ . In high-energy physics, these limits are often set at 95% CL, i.e.  $\alpha = 5\%$ .

### 4.5.1 Hypothesis test

Upper limits are computed in a similar way to confidence intervals, using an hypothesis test. Suppose that the parameter of interest is a signal yield  $S$  and that we are considering an upper limit  $S < S_0$ : then obviously our null hypothesis will be  $S = S_0$ . What is the alternate hypothesis that we should exclude against? For an *upper* limit, this is the case  $S < S_0$  where the true signal is below the limit, since this is the case where the limit would be invalid. If on the other hand  $S > S_0$ , then this does not invalidate our limit, and we can consider this case as part of the good outcomes, together with our  $S = S_0$  null hypothesis.

Therefore the test is naturally one-sided, as for discovery, and we define our test-statistic as<sup>5</sup>

$$q(S_0) = \begin{cases} -2 \log \frac{L(S = S_0)}{L(\hat{S})} & \text{if } \hat{S} < S_0 \\ 0 & \text{if } \hat{S} \geq S_0. \end{cases} \quad (49)$$

As usual, we use the value  $\hat{S}$  as the representative value for the alternate hypothesis  $S < S_0$ . We do this only in the case  $\hat{S} < S_0$ , due to the one-sidedness discussed above; for  $\hat{S} \geq S_0$  we set the test-statistic to 0, the same value as for  $S = S_0$ . This one-sided definition mirrors quite closely the situation of discovery testing in Section 4.3 (compare with Eq. (33)).

#### 4.5.2 Computing $p$ -values and upper limits

We can see from Eq. (49) that values of  $q(S_0)$  that are close to 0 indicate that  $\hat{S}$  is close to  $S_0$  (or above it, in which case  $q(S_0) = 0$  by construction). Conversely, large values of  $q(S_0)$  point to  $\hat{S} \ll S_0$ , i.e. that the observed result is too small to be compatible with  $S = S_0$  or above. As usual, one can quantify this agreement using a  $p$ -value. Assuming as before the asymptotic approximation of a near-Gaussian measurement, the  $p$ -value for an observed test value  $q(S_0) = q^{\text{obs}}(S_0)$  of the test statistic is

$$p(S_0) = 1 - \Phi \left( \sqrt{q^{\text{obs}}(S_0)} \right), \quad (50)$$

following the same steps as for Eq. (37).

There is however a last twist in the case of upper limits: what the  $p$ -value provides is the level of exclusion for a given  $S_0$ , which directly translates into the confidence level for the limit. For instance, if  $p(S_0) = 9\%$  then the  $p$ -value for  $S < S_0$  is 9%, which we can reformulate as the fact that  $S_0$  defines a 91% CL upper limit on  $S$ . However, typically what we want is not this, but instead the value of  $S_0$  that corresponds to a predefined CL, usually 95%. This means that to get the  $(1 - \alpha)$  CL upper limit, one generally needs to find the right value  $S_0$ , by solving the equation  $p(S_0) = \alpha$  for  $S_0$ . In simple examples this can be done in closed form, as we will see below, but generally one needs a numerical procedure that iteratively searches for the solution  $S_0$ .

#### 4.5.3 The one-bin Gaussian example

As usual, we now apply the general method to the case of our simple one-bin Gaussian measurement with fixed background  $B$ . Recall that this is defined by the PDF  $p(n; S) = G(n; S + B, \sigma)$ <sup>6</sup> and the observed yields  $n = n_{\text{obs}}$ .

As before, we have  $\hat{S} = n_{\text{obs}} - B$ . Obviously the upper limits that we will set are above  $\hat{S}$  (we cannot exclude the value that is preferred by the data!), so we consider  $S_0$  hypotheses above  $\hat{S}$ : we are

<sup>5</sup>Alternative definitions, such as the  $\tilde{q}_\mu$  of Ref. [12], can also be used.

<sup>6</sup>Note that we are ‘‘cheating’’ a bit here by using a constant Gaussian width  $\sigma$ , since in principle we should use  $\sigma = \sqrt{S + B}$ , which depends on  $S$ . This is a reasonable assumption in the case where  $S \ll B$ , so that  $\sigma \approx \sqrt{B}$ , and we adopt it here since removing the dependence on  $S$  simplifies the computation.

on the “good” side of the one-sided test defined in Eq. (49). We then have

$$q^{\text{obs}}(S_0) = -2 \log \frac{L(S_0)}{L(\hat{S})} = \left( \frac{S_0 - \hat{S}}{\sigma_S} \right)^2, \quad (51)$$

with the same calculation as the one that led to Eq. (48). Assuming that the asymptotic approximation applies, we have

$$p(S_0) = 1 - \Phi \left( \sqrt{q^{\text{obs}}(S_0)} \right) = 1 - \Phi \left( \frac{S_0 - \hat{S}}{\sigma} \right). \quad (52)$$

Note that we can remove the square root without ambiguity, since we know that  $S_0 > \hat{S}$ . To set the 95% CL upper limit  $S_{95}$ , we therefore need to solve

$$p(S_{95}) = 1 - \Phi \left( \frac{S_{95} - \hat{S}}{\sigma} \right) = 5\%, \quad (53)$$

which gives

$$S_{95} = \hat{S} + \Phi^{-1}(0.95)\sigma \approx \hat{S} + 1.64\sigma. \quad (54)$$

Recall that  $\Phi$  and  $\Phi^{-1}$  are implemented in e.g. `ROOT` and `scipy`, and we can use either to find that  $\Phi^{-1}(0.95) \approx 1.64$ . The computed limit has the expected properties: it rises and decreases with  $\hat{S}$ , so that observing a smaller signal leads to setting a lower upper limit and vice versa; and the upper limit is always above the best-fit signal, by an amount that is proportional to the uncertainty  $\sigma$  in the measurement. The only slightly non-trivial ingredient is the factor 1.64, which corresponds to the desired 95% CL.

#### 4.5.4 $CL_s$ upper limits

We close the discussion of upper limits by briefly discussing the  $CL_s$  modification to upper limits on signal yields, since this procedure is widely used in high-energy physics.

The motivation behind this extra wrinkle can be seen from Eq. (54), in the one-bin Gaussian case: suppose that the true signal value is  $S = 0$ , i.e. that we are looking for a signal that does not actually exist (although we are not aware of this fact!) and that  $\hat{S} < 0$  due to a statistical fluctuation in the background. We see that if  $\hat{S}$  is negative enough, the limit itself will go negative. For a 95% CL limit, this will occur if we are unlucky enough that  $\hat{S} < -1.64\sigma$ .

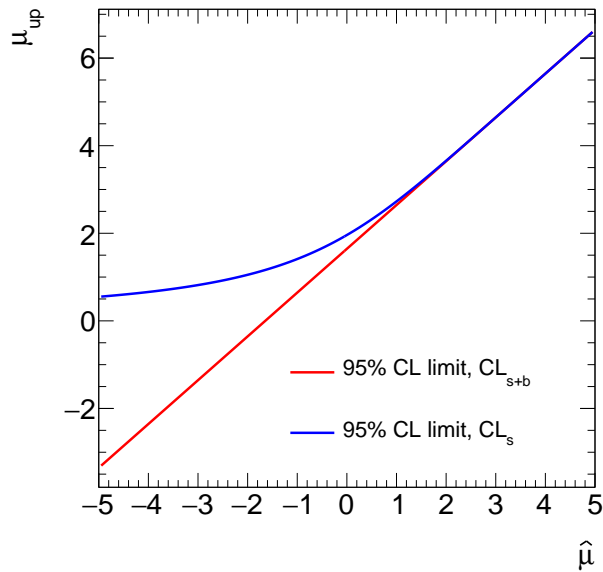
This is in fact completely normal: we know that when setting a 95% CL upper limit, that limit will by definition be wrong in 5% of the cases: this means that if  $S = 0$ , then in 5% of cases we will in fact set a negative limit  $S_{95} < 0$  that wrongly excludes the true value. While this is a basic property of statistical results, it is also somewhat counter-intuitive. Furthermore, if we assume that we know a priori that  $S > 0$ , then we also know that the cases where  $S_{95} < 0$  fall within the 5% of times where the limit fails. This motivates “fixing” the upper limit computation to avoid these cases.

The  $CL_s$  fix consists in modifying the definition of the  $p$ -value: instead of basing the test on  $p(S_0)$  as defined in Eq. (50) we use, instead,

$$p_{CL_s}(S_0) = \frac{p(S_0)}{p_0}, \quad (55)$$



where  $p_0$  is the  $p$ -value for the  $S = 0$  hypothesis. Without going into the technical details of why this particular modification is used, one can check that it has the intended effect: if  $\hat{S}$  is strongly negative, then this excludes  $S = 0$ , which means that the  $p$ -value  $p_0$  is small. Then  $p_{\text{CL}_s}(S_0) \gg p(S_0)$  and this larger  $p$ -value leads to a weaker limit, which almost always avoids spuriously excluding  $S = 0$ . However, if  $\hat{S}$  is compatible with 0 (or positive), then  $p_0 \approx 1$  (a large  $p$ -value indicating no exclusion of  $S = 0$ ) and therefore  $p_{\text{CL}_s}(S_0) \approx p(S_0)$ : in this case the result is unchanged compared to before. This behavior is illustrated in Fig. 11, where we see that the  $\text{CL}_s$  limit coincides with the usual frequentist  $\text{CL}_{s+b}$  (defined by Eq. (50)) for large  $\hat{S}$ ; and for  $\hat{S} < 0$  the  $\text{CL}_s$  case deviates so as to avoid negative limits. While this  $\text{CL}_s$  technique avoids “unphysical” negative limits, the price to pay for this is loss of



**Fig. 11:** Value of the 95% CL upper limit on the mean  $\mu$  of a Gaussian PDF with width 1, as a function of its best-fit value  $\hat{\mu}$ . The  $\text{CL}_{s+b}$  limit computed from Eq. (54) (red line) is shown alongside the  $\text{CL}_s$  limit computed from Eq. (56) (blue line).

coverage: this is still advertised as a 95% CL limit, but for  $\hat{S} \approx 0$  and below it corresponds in fact to a higher CL, and is therefore over-conservative.

Applying the  $\text{CL}_s$  computation to the simple Gaussian example of Section 4.5.3, one finds

$$S_{95}^{\text{CL}_s} = \hat{S} + \Phi^{-1} \left[ 1 - 0.05 \Phi \left( \frac{\hat{S}}{\sigma} \right) \right] \sigma, \quad (56)$$

and one can check that one recovers the result of Eq. (54) in the case of  $\hat{S} \gg \sigma$ , while for  $|\hat{S}| \ll \sigma$  one has

$$S_{95}^{\text{CL}_s} \approx \hat{S} + 1.96\sigma \quad (|\hat{S}| \ll \sigma). \quad (57)$$

This relation is quite useful since the scenario  $|\hat{S}| \ll \sigma$ , where no significant signal is found, is particularly relevant to setting upper limits.

We conclude by stating without proof another very useful result: suppose that we perform a single-bin counting experiment and that we observe  $n_{\text{obs}} = 0$ . Then the exact value of the 95%  $\text{CL}_s$  upper limit

is

$$S_{95}^{\text{CL}_s}(n_{\text{obs}} = 0) = \log(20) \approx 3. \quad (58)$$

This is a remarkable result for two reasons: first, it is independent of the background level  $B$  or, equivalently, of the uncertainty  $\sigma$  of the measurement; and secondly, it is exact in the sense that it does not rely on the asymptotic approximation: it is, in fact, based solely on properties of the Poisson distribution. For this reason, the result cannot be obtained as a limiting case of Eq. (57), but it is relatively easy to derive by going back to the Poisson definition of the  $p$ -values entering Eq. (55).

#### 4.6 Expected results

So far we have covered the computation of so-called *observed* results, i.e. those obtained from a particular observed dataset. It is also often useful to compute *expected* results, i.e. the median expected outcome under a particular hypothesis.

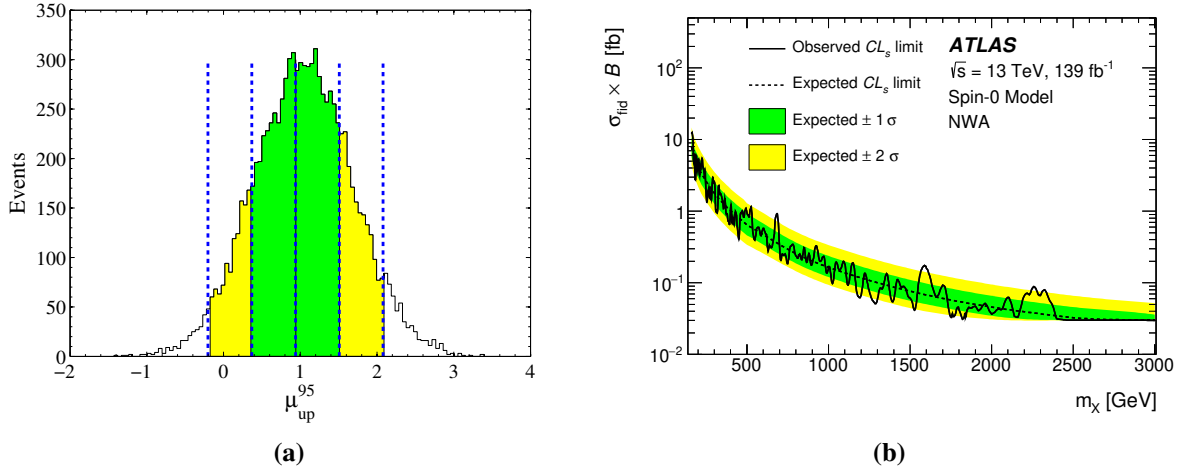
A common use-case for this is to choose between two analysis options: if the choice is done using the observed results, then one may end up picking an option that seems more sensitive due to a lucky fluctuation in the data. While this may be beneficial for this particular dataset, it may not remain so when more data is collected, and such a choice would also systematically overestimate the analysis sensitivity. This is related to the concept of *blind analysis*, where analysis choices are made only based on expected outcomes, without looking at the observed data, in order to avoid biases towards a particular result (e.g. the result found by previous measurements). Another typical use-case for expected results is the projection of analysis sensitivity to as-yet hypothetical situations, for instance to estimate the expected performance at future experimental facilities.

Expected results are computed under a given hypothesis; for instance, the Standard Model expectations. There are two main techniques for this: pseudo-experiments (also often called “toy datasets”) and Asimov datasets. For pseudo-experiments, one uses the measurement PDF to generate random data, i.e. datasets which have not been actually observed in the experimental apparatus, but are randomly generated using the PDF. Recall that the PDF is exactly the tool needed to do so, since it provides the probabilities for different outcomes. Defining the generation hypothesis simply corresponds to setting the PDF parameters to the appropriate values. Technically, tools for random generation are provided by the usual statistics toolkits (e.g. RooFit, ROOT, or pyhf). Statistical results are then computed from each pseudo-dataset, exactly in the same way as for real observed data. The expected result is then reported as the median of these results, as shown in Fig. 12a. One can also compute, e.g.,  $1\sigma$  and  $2\sigma$  bands around the median using the corresponding quantiles of the distribution. These bands are useful to test the agreement of the observed result with the expected result. They are often shown in particular for limits, as in Fig. 12b.

The other method of computing expected results is the so-called *Asimov dataset* technique: in this case, one constructs a single dataset that corresponds exactly to the desired scenario, the Asimov dataset.<sup>7</sup> The expected results are then obtained by simply performing the computation on this dataset. The Asimov dataset is formally defined as a dataset for which the best-fit values of all model parameters are exactly equal to their hypothesis values. So, if the desired scenario is  $\mu = \mu_0$ , then an Asimov dataset

<sup>7</sup>The name originates from a short story by Isaac Asimov, *Franchise*, featuring a form of government based on a similar premise.

should verify  $\hat{\mu} = \mu_0$ . For a counting experiment, one can construct such a dataset by simply setting the observed yield in all bins to their expectations. For unbinned cases there is no similar technique, but one can get a suitable approximation by building a binned dataset with sufficiently fine bins, where the bin yields again match the expectation from the model.



**Fig. 12:** (a) Illustration of the computation of an expected result from an ensemble of pseudo-experiments. The pseudo-experiment results ( $\mu_{95}$ ) are shown as a histogram (black line). The expected result is computed as the median of the histogram (central blue line), and the  $1\sigma$  and  $2\sigma$  bands (green and yellow areas) as the corresponding quantiles around the median. (b) Example computation of expected and observed upper limits taken from Ref. [16]. Limits on the production cross-section of a resonance are shown as a function of its mass  $m_X$ . Each position in  $m_X$  corresponds to a separate result, for which both observed (solid black line) and expected (dotted black lines) upper limits are shown. The green and yellow areas show respectively the  $1\sigma$  and  $2\sigma$  bands around the expected result, as in panel (a). The computation is performed using the Asimov dataset technique at lower values of  $m_X$ , where the measurement is quasi-Gaussian due to large event yields. At higher  $m_X$  the smaller yields invalidate the Gaussian approximation, and the results are instead obtained using the pseudo-experiments technique.

The Asimov dataset technique has the advantage that the expected result can be obtained from a single computation, whereas the pseudo-experiments technique typically requires processing tens to hundreds of datasets. It does not however provide the bands around the results, and these need to be computed using asymptotic formulas [12]. The Asimov dataset technique is therefore usually the preferred choice in Gaussian settings, while pseudo-experiments are required in non-Gaussian cases, where the asymptotic approximation does not apply.

## 5 Profiling and systematic uncertainties

A careful reader may have noticed that the “general” methods presented in Section 4 did not include the treatment of the nuisance parameters of the model, discussed in Section 2. The central concept of systematic uncertainties was also not yet introduced. We will show in this section that both are in fact closely related, and that their treatment can be included as a simple modification of the methods described in Section 4.

## 5.1 Profile likelihood method

### 5.1.1 Definition of the profile likelihood ratio

In Section 4.2, we have considered hypothesis tests in the case of a PDF  $p(n; \mu)$  with a single parameter of interest. We have seen that, according to the Neyman–Pearson lemma, we can use the likelihood ratio  $L(\mu_0)/L(\mu_1)$  to make a decision between two hypotheses  $\mu = \mu_0$  and  $\mu = \mu_1$ . We have extended this to test  $\mu = \mu_0$  against  $\mu \neq \mu_0$  by using  $L(\mu_0)/L(\hat{\mu})$ , where the best-fit value  $\hat{\mu}$  is used as a stand-in for the  $\mu \neq \mu_0$  case.

What if now we have  $P(n; \mu, \theta)$ , with nuisance parameters  $\theta$  also present? In principle, the Neyman–Pearson lemma applies in the same way to this case as well, and we can test hypotheses defined by values of both  $\mu$  and  $\theta$ . However, we are not interested in the values of  $\theta$  (by definition, these are not parameters of interest!); the hypotheses we want to test are only defined by values of  $\mu$ . So, to use the Neyman–Pearson lemma, we need to “fill in” some  $\theta$  values to fully specify the hypotheses.

Following the principles already laid out earlier, the obvious values to use are the ones provided by the data, i.e. the best-fit values:

- For the null hypothesis defined by  $\mu = \hat{\mu}$ , we can just add  $\hat{\theta}$  to the definition so that this becomes the  $(\mu = \hat{\mu}, \theta = \hat{\theta})$  hypothesis.
- For the alternate  $\mu = \mu_0$  hypothesis, we need to account for the fact that we restrict ourselves to a particular value of  $\mu$ , so that for consistency the best-fit value of  $\theta$  should also be computed under this restriction. We therefore introduce the *conditional best-fit value*  $\hat{\theta}(\mu_0)$ , which is the best-fit value of  $\theta$  under the condition  $\mu = \mu_0$ . The alternate hypothesis is then defined in full by  $(\mu = \mu_0, \theta = \hat{\theta}(\mu_0))$ .

One can see immediately that  $\hat{\theta}(\hat{\mu}) = \hat{\theta}$ , but for other values of  $\mu$  the conditional best-fit value may not necessarily match the overall best-fit value  $\hat{\theta}$ . This conditional best-fit value is also called the *profiled* value of  $\theta$  as a function of  $\mu_0$ . Putting it all together, this gives a new definition of the likelihood ratio, which with the usual  $-2 \log$  operation reads

$$t(\mu_0) = -2 \log \frac{L(\mu_0, \hat{\theta}(\mu_0))}{L(\hat{\mu}, \hat{\theta})}. \quad (59)$$

This is the *profile likelihood ratio*, and corresponds to a generalization of Eq. (47) in the presence of nuisance parameters  $\theta$ .

### 5.1.2 Wilks’ theorem for the profile likelihood ratio

One can see immediately that, thanks to the use of best-fit values,  $t(\mu_0)$  remains a function of  $\mu_0$  only, without reference to the  $\theta$ . The  $\theta$  are of course always there in the background, but their impact is baked into  $t(\mu_0)$  through the best-fit values and not explicitly apparent.

Furthermore, there is a truly amazing result on the asymptotic distribution of  $t(\mu_0)$ : *in the  $\mu = \mu_0$  hypothesis,  $t(\mu_0)$  follows a  $\chi^2$  distribution with a number of degrees of freedom equal to the number of parameters of interest.* This result is known as *Wilks’ theorem* [13].

This shouldn't come as too much of a surprise since the same result was already presented in Section 4.3.3 for the case of a simple likelihood ratio without nuisance parameters, defined by Eq. (47). The full version of Wilks' theorem that is stated above generalizes this to the case where nuisance parameters are present, and are profiled as shown in Eq. (59). The fact that it remains true also in this case is somewhat miraculous (it relies on a subtle interplay between the best-fit values of  $\mu$  and  $\theta$  in the Gaussian case), but the upshot is that things do not change very much when nuisance parameters are also included. With the new definition of  $t(\mu_0)$  from Eq. (59) (and the related test statistics of Eqs. (33) and (49)), all the techniques and formulas presented in Section 4 remain applicable as long as the asymptotic approximation is valid.

So, for example, one can still compute the discovery significance as  $z = \sqrt{q_0}$ , following Eq. (36), provided that the definition of  $q_0$  in Eq. (33) is updated to include the conditional (under  $S = 0$ ) and unconditional best-fit values of the nuisance parameters, similarly to Eq. (59). Confidence intervals and upper limits can also still be computed as described in Section 4, with the profiled values of the nuisance parameters included in the definition of the test statistics.

### 5.1.3 Application to a simple Gaussian example

To illustrate the use of the profile likelihood, we consider a measurement where the signal yield  $S$  and background yield  $B$  are both free parameters. The goal is to demonstrate how to deal with  $B$  using profiling, in order to measure  $S$ . Since we need to measure two parameters, we need at least two measurement bins. We therefore now include two independent Gaussian measurements: in one bin, we measure  $S + B$  with uncertainty  $\sigma$ , using an event count  $n$ ; in the other we assume that only background is present so that we measure  $B$  only with uncertainty  $\epsilon$  using an event count  $m$ . This is in fact a fairly standard experimental setup, where the measurement mainly occurs in a signal region (SR) where both signal and background is present, and the background is obtained through a separate control region (CR) which is sensitive to background only. The full measurement PDF is

$$p(n, m; S, B) = G(n; S + B, \sigma) G(m; B, \epsilon). \quad (60)$$

Assuming that we observe  $n_{\text{obs}}$  and  $m_{\text{obs}}$ , we define as usual the likelihood  $L(S, B) = P(n_{\text{obs}}, m_{\text{obs}}; S, B)$  and the N2LL  $\lambda(S, B) = -2 \log L(S, B)$ . We have

$$\lambda(S, B) = \left( \frac{S + B - n_{\text{obs}}}{\sigma} \right)^2 + \left( \frac{B - m_{\text{obs}}}{\epsilon} \right)^2. \quad (61)$$

The best-fit values of  $S$  and  $B$  are obtained from minimizing  $\lambda$ , and we have

$$\hat{S} = n - m \quad (62)$$

$$\hat{B} = m \quad (63)$$

$$\hat{B}(S) = m + \frac{\epsilon^2}{\sigma^2 + \epsilon^2} (\hat{S} - S) \quad (64)$$

As expected, the best-fit values of  $S$  and  $B$  are the ones that best match the data. The profile value  $\hat{B}(S)$  also has the expected properties: for  $S \neq \hat{S}$ , one can see that  $\hat{B}(S)$  deviates from  $\hat{B} = m$  in a way that partially compensates for the deviation of  $S$  from  $\hat{S}$ : if  $S > \hat{S}$  then  $\hat{B}(S) < m$  and vice versa, which in both cases tends to soften the discrepancy between the prediction and the data. Plugging these values into Eq. (59), we then obtain

$$t(S_0) = \left( \frac{S_0 - (n - m)}{\sigma^2 + \epsilon^2} \right)^2. \quad (65)$$

We can then obtain a confidence interval on  $S$  from the intersections  $t(S) = 1$  as described in Section 4.4.3. We get

$$S = (n - m) \pm \sqrt{\sigma^2 + \epsilon^2} \quad \text{at 68.3\% CL} \quad (66)$$

with an uncertainty of  $\sqrt{\sigma^2 + \epsilon^2}$  that is the sum in quadrature of the uncertainties coming from the SR ( $\sigma$ ) and from the CR ( $\epsilon$ ). This illustrates that although  $t(S)$  remains a function of  $S$  only, the profiling accounts for the impact of the nuisance parameters behind the scenes, and the uncertainty from the measurement of  $B$  in the CR was correctly propagated to the estimation of  $S$ .

## 5.2 Systematic uncertainties

### 5.2.1 Statistical and systematic uncertainties

We finally come to one of the central issues of statistical analysis in high-energy physics: systematic uncertainties. First, what are they? Recall that the measurement PDFs that we have been working with are a way to describe uncertainties about the data, as discussed in Section 2. For instance, we use Poisson distributions to encode the fact that the number of events observed in a counting experiments fluctuates, if we repeat the experiment several times. These uncertainties, which are provided by the measurement PDFs, are called *statistical uncertainties*. They are the uncertainties that we have been dealing with up to now.

There is however another class of uncertainties: uncertainties in the form of the PDF itself. For instance, in the counting example studied in Section 4.1.3 we have assumed that the background yield  $B$  is known exactly and this is a critical input to the analysis; e.g. to extract the signal as  $\hat{S} = n_{\text{obs}} - B$ . However, in a real-life situation  $B$  is never known exactly: there is an uncertainty on its value. This uncertainty isn't captured by the PDF itself (Eq. (25) in this example) since it is an uncertainty on the very form of the PDF. These uncertainties on the definition of the PDF are known as *systematic uncertainties*.

There is an alternate definition of statistical uncertainties based on their behavior as the measurement dataset increases: we have seen that, according to the central-limit theorem, the combined precision of  $N$  measurements scales as  $1/\sqrt{N}$  (see Eq. (4)). By the same argument, statistical uncertainties scale as the inverse square root of the size of the dataset, as more data makes the measurement more precise. Systematic uncertainties on the other hand usually remain constant even as the dataset size increases (unless one makes clever use of the new data to improve the measurement!): they represent a fixed bias between the actual measurement process and its imperfect statistical model, which more data does not help to reduce. This can be illustrated by coming back again to the simple Gaussian example of Section 4.1.3, where the signal yield is obtained as  $\hat{S} = n_{\text{obs}} - B$ : the statistical uncertainty comes from  $n_{\text{obs}}$  and a larger dataset will lead to increased relative precision on this term. However, if  $B$  is off from

its true value, then this will lead to a systematic bias on  $\hat{S}$  that more data cannot help to reduce. This bias then needs to be covered by a separate systematic uncertainty.

### 5.2.2 Systematic uncertainties as nuisance parameters

Since systematic uncertainties affect the measurement PDFs themselves, they lie outside the scope of the techniques we have presented in Section 4. We therefore need to find a way to expand our description of the PDFs to also account for these effects. The simplest way to do this is to add more free parameters into the description. For instance, the background yield  $B$  in the example above can be promoted from a fixed value to a floating parameter that can be adjusted from the data itself.

Sometimes it is possible to just do this: we arrange to obtain  $B$  from a *data-driven* estimate, as in the example shown in Section 5.1.3, and remove the source of systematic bias. However, this is not always possible: for instance in the one-bin Gaussian example of Section 4.1.3, we estimate the signal from a single bin yield using  $\hat{S} = n_{\text{obs}} - B$ , and this assumes that we know  $B$  a priori. If both  $S$  and  $B$  are free parameters, then we cannot estimate both their values from just the knowledge of the single yield  $n_{\text{obs}}$ .

The way out is to assume that we have some external knowledge of  $B$ , coming from outside the current measurement. In general this is what happens in realistic situations:  $B$  is not completely unknown but can be estimated from previous experiments, MC simulation, or a combination of the two. We will generally frame this knowledge as the results of *auxiliary measurements* that are independent of the measurement that we are describing: either using a separate dataset or a completely different apparatus.

This is a sensible approach for the background yield  $B$ , but can be adapted to less obvious cases such as theory predictions. Of course, the output of a theory computation can hardly be viewed as the result of a measurement (theory errors do not represent fluctuations in the result of the computation!). However, one can still represent the knowledge on the corresponding theory parameter using, e.g., a Gaussian distribution with a width corresponding to the theory uncertainty.

The general framework is then as follows: suppose that we have a measurement  $P_{\text{main}}(n; \mu)$ , where as usual  $n$  represents the observables and  $\mu$  the measurement parameters. To describe systematic uncertainties in  $P_{\text{main}}$ , we augment  $\mu$  with the parameters  $\{\theta_i^{\text{synt}}\}_{1 \leq i \leq N_{\text{synt}}}$ , which are additional nuisance parameters describing the systematic uncertainties. We assume that we have external knowledge on each  $\theta_i^{\text{synt}}$ , encoded in the PDF  $P_i(\theta_i^{\text{obs}}; \theta_i^{\text{synt}})$ . This PDF represents an auxiliary measurement with observable  $\theta_i^{\text{obs}}$  that provides information on  $\theta_i^{\text{synt}}$ . Since the auxiliary measurements are assumed to be independent from the main measurement, we can combine all of them together with  $P_{\text{main}}$  by taking the product

$$P(n, \{\theta_i^{\text{obs}}\}; \mu, \{\theta_i^{\text{synt}}\}) = P_{\text{main}}(n; \mu, \{\theta_i^{\text{synt}}\}) \prod_i P_i(\theta_i^{\text{obs}}; \theta_i^{\text{synt}}). \quad (67)$$

Remember that in  $P_{\text{main}}(n; \mu, \{\theta_i^{\text{synt}}\})$  alone, we typically would not have enough measurement information to constrain all the  $\theta_i^{\text{synt}}$ . But this is now possible in  $P(n, \{\theta_i^{\text{obs}}\}; \mu, \{\theta_i^{\text{synt}}\})$ , thanks to the extra information coming from the observables  $\theta_i^{\text{obs}}$ . We can therefore now treat this PDF using the profile likelihood techniques described in Section 5.1 to compute the results in the presence of systematic un-

certainties.

In practice, the  $P_i$  will often be represented as simple Gaussians, with a central value corresponding to the nominal value of the nuisance parameter and a width corresponding to the value of its uncertainty. This can be considered as a simplified description of the auxiliary measurement, in the cases where one actually exists, or as a mathematical tool to convey the uncertainty in other cases (e.g. for theory uncertainties). However, it is also possible in principle to provide  $P_i$  as the full PDF of an auxiliary measurement, obtained e.g. as described in Section 2.

### 5.2.3 The simple one-bin Gaussian example

We illustrate the treatment of systematics by returning one last time to our one-bin Gaussian counting example,  $P_{\text{main}}(n; S, B) = G(n; S + B, \sigma)$ . Instead of assuming that  $B$  is known exactly, we will now assume that there is some uncertainty in this value, so that we have  $B = B_{\text{nom}} \pm \sigma_B$ . We represent this systematic uncertainty as a Gaussian auxiliary measurement with PDF  $P_B(B_{\text{nom}}; B) = G(B_{\text{nom}}; B, \sigma_B)$ . Including this extra information, the full measurement PDF is now

$$P(n, B_{\text{nom}}; S, B) = P_{\text{main}}(n; S, B) P_B(B_{\text{nom}}; B) = G(n; S + B, \sigma) G(B_{\text{nom}}; B, \sigma_B). \quad (68)$$

We can now obtain a confidence interval on  $S$  by profiling  $B$  and defining the profile likelihood  $t(S)$  as in Eq. (59). The profiling of  $B$  will then account for the impact of its uncertainty on the measurement of  $S$ , as we saw already in Section 5.1.3. In fact one can check that the computations in this example are formally identical to those in Section 5.1.3: by a simple change of notation, we can obtain immediately

$$S = (n - B_{\text{nom}}) \pm \sqrt{\sigma^2 + \sigma_B^2} \quad \text{at 68.3\% CL.} \quad (69)$$

So, the systematic uncertainty  $\sigma_B$  on the background level gets added in quadrature to the statistical uncertainty  $\sigma$  to form the total uncertainty on  $S$ , as one would have naively expected. The implementation of the systematic uncertainty as a nuisance parameter and its treatment using profiling therefore fully accounts for its effect on the measurement.

The similarity of the computation here and the one in Section 5.1.3 is not completely accidental: as mentioned above, a systematic uncertainty on a model parameter can be seen as information coming from an auxiliary experiment that is sensitive to this parameter. The control region (CR) measurement in Section 5.1.3 can be seen as such an auxiliary measurement: if one considers only the signal region (SR) measurement as *the* measurement, then the CR is an auxiliary measurement and  $B$  is associated with a systematic uncertainty, as in this section. If, however, one considers the measurement as encompassing both the SR and the CR, then both  $S$  and  $B$  are measured simultaneously from data, in a measurement without systematic uncertainties. In this second case, the statistical uncertainty on  $B$  still propagates to the uncertainty on  $S$  in the same way as for a systematic uncertainty, so the two cases are formally equivalent.

Note, however, that in the case of a systematic uncertainty, an increase in the dataset would a priori apply only to the SR and not to the auxiliary measurement, while for a combined measurement, both SR and CR datasets would be expected to increase. This difference reflects the different scaling behaviors



of statistical and systematic uncertainties with luminosity that were described in Section 5.2.1.

### 5.3 Profiling: caveats and pitfalls

As described in Section 5.2.2, there are two types of nuisance parameters: parameters associated with an auxiliary observable, as in Section 5.2.3, which represent systematic uncertainties, and data-driven parameters which are determined fully from the data without additional external information, as in Section 5.1.3. Profiling provides a general way to deal with nuisance parameters of both types. Thanks to the Neyman–Pearson lemma and Wilks’ theorem, the resulting profile likelihood ratio tests statistics are guaranteed (with some caveats) to be optimal, i.e. they make use of all the information in the data to provide statistical results with maximal sensitivity.

There can be some interplay between auxiliary measurements and data-driven constraints: for instance, in a complex measurement with a large number of bins the data can provide constraints on the systematic nuisance parameters. If the constraint from the data is stronger than the one provided by the auxiliary measurement, then the data itself provides a better estimate of the parameter than what was provided externally: the magnitude of the systematic uncertainty is therefore reduced, compared to the value that was given as input in the model.

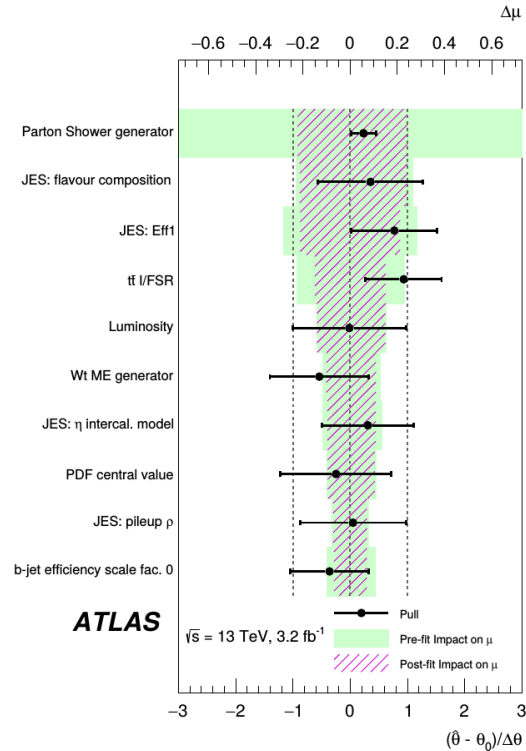
This property of profiling is particularly useful in LHC experiments, where large datasets allow one to set strong constraints that can help to reduce systematic uncertainties. Profiling therefore provides a powerful tool to improve the measurement precision. However, it must also be used with caution, since it relies on the assumption that the systematic parameters provide a complete description of the uncertainties.

To illustrate where this can fail, suppose that a measurement is sensitive to the energy calibration of an experimental observable, say the jet energy, and that the associated systematic uncertainty is described using a single nuisance parameter. Assume further that large amounts of data are available at low jet energies, but that the measurement is performed at higher energies. If profiling is applied, the low-energy data can provide a strong constraint on the parameter, which then translates into a reduced systematic uncertainty that also applies in the high-energy region. In terms of physics modeling this is often wrong: the calibration of high-energy objects is often decorrelated from the low-energy region, so that one should have separate uncertainties described by different parameters. The reduction of the uncertainty in the high-energy region is therefore likely invalid. The issue is not related directly to the profiling itself, but rather to the description of the uncertainties using model parameters. While profiling is a powerful tool, it requires a careful treatment of this point, to avoid spurious reductions in systematic uncertainty.

A realistic example of a case where a systematic uncertainty is heavily reduced is shown in Fig. 13. While such cases can correspond to legitimate uses of the data to improve on the knowledge of systematic effects, they should be checked carefully to ensure this improvement is justified.

## 6 Conclusion

Statistical methods are an essential part of high-energy measurements. Modern tools implemented within the ROOT toolkit or the python ecosystem allow one to describe complex measurements using binned or unbinned PDFs, as well as the associated systematic uncertainties. Frequentist techniques based



**Fig. 13:** Pull and impact plot taken from Ref. [15]. The rows correspond to nuisance parameters describing the leading systematic uncertainties in the analysis. The black bars and dot show the normalized best-fit values and uncertainties (pulls) of the parameters. Uncertainties smaller than 1 indicate that the parameter is constrained by the data. In this case the effective impact of the systematic uncertainty (red hashes) is reduced compared to its input value (green bands).

on the use of profile likelihood ratios can then be used to obtain statistical results, such as discovery significances, confidence intervals for model parameters, and upper limits on signal yields. These results can be obtained using arbitrarily complex likelihoods (limited only by computing power) and make generally optimal use of the information present in the data.

A set of jupyter notebooks providing examples and exercises based on the contents of these lectures can be found at <https://github.com/fastprof-hep/stats-tutorial/tree/main/AEPSHEP2022>. Further reading on these techniques can be found in standard textbooks on statistics, such as Refs. [1, 17, 18].

## References

- [1] G. Cowan, *Statistical data analysis* (Oxford Univ. Press, Oxford, 1998).
- [2] S. Brandt, *Data analysis: Statistical and computational methods for scientists and engineers*, 4th ed. (Springer, Dordrecht, 2014), doi:10.1007/978-3-319-03762-2.
- [3] R.J. Barlow, Practical statistics for particle physics, *CERN Yellow Rep. School Proc.* **5** (2020) 149–197, doi:10.23730/CYRSP-2020-005.149.

- [4] L. Lyons, *Statistical issues in searches for new physics*, Proc. 2nd Conf. on Large Hadron Collider Physics Conference (LHCP 2014), New York, USA, 2–7 June, 2014, [arXiv:1409.1903 [hep-ex]], [doi:10.48550/arXiv.1409.1903](https://doi.org/10.48550/arXiv.1409.1903).
- [5] K. Cranmer *et al.*, HistFactory: A tool for creating statistical models for use with RooFit and RooStats, CERN-OPEN-2012-016 (CERN, Geneva, 2012), [doi:10.17181/CERN-OPEN-2012-016](https://doi.org/10.17181/CERN-OPEN-2012-016).
- [6] R. Brun and F. Rademakers, ROOT – an object-oriented data analysis framework, *Nucl. Instrum. Meth. A* **389** (1996) 81–86, [doi:10.1016/S0168-9002\(97\)00048-X](https://doi.org/10.1016/S0168-9002(97)00048-X).
- [7] L. Heinrich *et al.*, pyhf: pure-python implementation of histfactory statistical models, *J. Open Source Softw.* **6** (2021), no. 58, 2823, [doi:10.21105/joss.02823](https://doi.org/10.21105/joss.02823).
- [8] L. Heinrich, M. Feickert, and G. Stark, *pyhf: v0.7.0* (Scikit-HEP Project, GitHub, 24 Sep. 2022), <https://github.com/scikit-hep/pyhf/releases/tag/v0.7.0>.
- [9] ATLAS Collaboration, Measurement of  $W^\pm$  and Z-boson production cross sections in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector, *Phys. Lett. B* **759** (2016) 601–621, [doi:10.1016/j.physletb.2016.06.023](https://doi.org/10.1016/j.physletb.2016.06.023).
- [10] ATLAS Collaboration, Measurement of the properties of Higgs boson production at  $\sqrt{s} = 13$  TeV in the  $H \rightarrow \gamma\gamma$  channel using  $139 \text{ fb}^{-1}$  of pp collision data with the ATLAS experiment, *JHEP* **07** (2023) 088, [doi:10.1007/JHEP07\(2023\)088](https://doi.org/10.1007/JHEP07(2023)088).
- [11] CMS Collaboration, Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at  $\sqrt{s} = 13$  TeV, *JHEP* **11** (2018) 185, [doi:10.1007/JHEP11\(2018\)185](https://doi.org/10.1007/JHEP11(2018)185).
- [12] G. Cowan *et al.*, Asymptotic formulae for likelihood-based tests of new physics, *Eur. Phys. J. C* **71** (2011) 1554, [doi:10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0).
- [13] S. Wilks, The large-sample distribution of the likelihood ratio for testing composite hypotheses, *Annals Math. Statist.* **9** (1938) 60–62, [doi:10.1214/aoms/1177732360](https://doi.org/10.1214/aoms/1177732360).
- [14] ATLAS Collaboration, Combined measurements of Higgs boson production and decay using up to  $80 \text{ fb}^{-1}$  of proton-proton collision data at  $\sqrt{s} = 13$  TeV collected with the ATLAS experiment, *Phys. Rev. D* **101** (2020) 012002, [doi:10.1103/PhysRevD.101.012002](https://doi.org/10.1103/PhysRevD.101.012002).
- [15] ATLAS Collaboration, Measurement of the cross-section for producing a W boson in association with a single top quark in pp collisions at  $\sqrt{s} = 13$  TeV with ATLAS, *JHEP* **01** (2018) 063, [doi:10.1007/JHEP01\(2018\)063](https://doi.org/10.1007/JHEP01(2018)063).
- [16] ATLAS Collaboration, Search for resonances decaying into photon pairs in  $139 \text{ fb}^{-1}$  of pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector *Phys. Lett. B* **822** (2021) 136651, [doi:10.1016/j.physletb.2021.136651](https://doi.org/10.1016/j.physletb.2021.136651).
- [17] F. James, *Statistical methods in experimental physics*, 2nd ed. (World Scientific, Singapore, 2006), [doi:10.1142/6096](https://doi.org/10.1142/6096).
- [18] R. Barlow, *A guide to the use of statistical methods in the physical sciences* (Wiley, Chichester, 1989).



# Neutrino physics

*Sin Kyu Kang*

Seoul National University of Science and Technology, Seoul, Korea

---

In this lecture, I start with presenting the history of the neutrino from its invention to what we have discovered about its properties till now. I explain how we can observe neutrinos produced both naturally and artificially. Naturally produced neutrinos come to the Earth from the Sun, supernovae, collisions of cosmic rays with nuclei in the atmosphere, natural radioactivity, etc. On the other hand, those produced in accelerators and nuclear reactors are the examples of artificial neutrinos. I also illustrate what neutrino oscillations are and how such phenomena could be observed from various experiments to detect neutrinos produced in the aforementioned ways. Thanks to the discovery of neutrino oscillations, we are forced to modify the Standard Model, so as to accommodate the masses of neutrinos and lepton flavor mixing, which are essential to make neutrino flavor change. In fact, neutrinos can come in three different flavors, electron, muon and tau, and can change from one flavor to another. The origin of the tiny neutrino masses is still unknown, although we now know a few nice mechanisms capable of generating them. The generation of neutrino masses signifies physics beyond the Standard Model and can, therefore, be related to some of the unresolved fundamental issues, such as the origin of flavors, the unification of forces, the matter-antimatter asymmetry, etc. Some physicists believe that CP violation in neutrinos may be a missing piece in the understanding of the origin of the matter-antimatter asymmetry. I pedagogically explain how we can probe CP violation through neutrino oscillation experiments.

---

## 1 Introduction

1998 is the historic year in which neutrinos aroused great interest among not only physicists but also the public. The Super-Kamiokande (SK) Collaboration announced the first evidence of neutrino oscillations in that year [1]. This was the first experimental observation supporting the theory that the neutrino has non-zero mass, a possibility that theorists had speculated about for years. The discovery of neutrino oscillations was expected to bring fundamental changes to our knowledge of physics and astronomy. Though, many more discoveries about neutrinos have yet to be made.

### 1.1 Advent of neutrinos

Neutrinos were postulated by Pauli in 1930 to resolve the puzzle of the electron energy spectrum observed in beta decays, which show a continuous distribution instead of all electrons having the same energy. It seemed to contradict the principle of the conservation of energy. The puzzle can be solved if another unseen particle is emitted along with the electron in beta decay. Pauli originally called this particle the “neutron”, but it was renamed the “neutrino” by Fermi in 1933. After the discovery of the neutron by Chadwick in 1932 [2], Fermi developed the theory of beta decay by proposing that four fermions directly interact with one another at one vertex [3]. By this interaction, the neutron decays directly to a proton, an electron and the proposed neutrino (what we now know to be an electron-antineutrino). The theory developed by Fermi, which proved to be successful, was the precursor to the theory of the weak

---

This article should be cited as: Neutrino physics, Sin Kyu Kang, DOI: [10.23730/CYRSP-2024-001.47](https://doi.org/10.23730/CYRSP-2024-001.47), in: Proceedings of the 2022 Asia–Europe–Pacific School of High-Energy Physics,

CERN Yellow Reports: School Proceedings, CERN-2024-001, DOI: [10.23730/CYRSP-2024-001](https://doi.org/10.23730/CYRSP-2024-001), p. 47.

© CERN, 2024. Published by CERN under the [Creative Commons Attribution 4.0 license](https://creativecommons.org/licenses/by/4.0/).

interaction. Fermi first submitted his “tentative” theory of beta decay to the journal *Nature*, which rejected it “because it contained speculations too remote from reality to be of interest to the reader” [4]. Later, *Nature* admitted the rejection to be one of the great editorial blunders in its history.

## 1.2 Discovery of neutrinos

Neutrinos were first discovered by Reines and Cowan in 1956, who conducted the experiment to observe neutrinos through inverse beta decay ( $\bar{\nu}_e + p \rightarrow n + e^+$ ) by using neutrinos from the Savannah River nuclear reactor [5]. In 1962, Lederman, Schwartz and Steinberger discovered a second neutrino by proving that the “muon” appeared to be accompanied by a neutrino that should be different from the neutrino appearing in beta decays [6]. The third generation charged lepton, called the “tau” lepton, has been discovered by Perl in 1975 [7]. As soon as the discovery of the tau lepton was announced, particle physicists speculated from experience that it would have a neutrino partner, the tau neutrino. It remained elusive until July 2000, when the DONUT experiment from Fermilab announced its discovery [8]. These three species of neutrinos are named electron-neutrino ( $\nu_e$ ), muon-neutrino ( $\nu_\mu$ ) and tau-neutrino ( $\nu_\tau$ ), associated to the charged leptons electron, muon and tau, respectively. Thanks to those experiments, the Standard Model (SM) of particle physics has been established to have six leptons consisting of three families, similarly to the quark sector.

## 1.3 Neutrinos in the Standard Model

In the SM, neutrinos are the only massless particles. There are exactly three neutrinos, one for each of the three charged leptons, as explained before and as confirmed by the observation of  $e^+ + e^- \rightarrow Z^0 \rightarrow f\bar{f}$  at LEP experiments, with  $N_\nu = \frac{\Gamma_{inv}}{\Gamma_{\nu\bar{\nu}}} = 2.984 \pm 0.008$  [10]. All neutrinos are left-handed and all antineutrinos are right-handed in the SM. Since there are no right-handed (left-handed) neutrinos (antineutrinos), neutrinos have no mass. In the SM, neutrinos interact with matter only through weak interactions, which occur in two types, charged-current (CC) interactions and neutral-current (NC) interactions. The CC interactions occur through the exchange of a  $W^\pm$ , where the neutrino converts into the corresponding charged lepton (e.g. inverse beta decay,  $\bar{\nu}_e + p \rightarrow n + e^+$ ). The NC interactions occur through the exchange of a  $Z^0$ , where the neutrino remains a neutrino, but transfers energy and momentum to whatever it interacted with.

## 2 Detection of neutrinos

### 2.1 Neutrino sources

Neutrinos are produced in the stars, the sky, nuclear reactors, human bodies, and even food like bananas. When we consider scientific research, there are several interesting sources that can help physicists study neutrinos: the sun, the atmosphere, reactors, accelerators, earth, the Big Bang, Supernovae, Extragalactic sources, etc. Huge numbers of neutrinos (about  $10^{20}$  per second) are emitted in nuclear reactors, and also artificially produced in man-made accelerators delivering intense neutrino beams. But the main source of neutrinos is the Universe itself. The relic neutrinos from the Big Bang have been wandering for more than 13.6 billion years, with a density of 330 per  $\text{cm}^3$  everywhere. Starting with the fusion of two protons, nuclear reactions in the core of the Sun produce about  $2 \times 10^{38}$   $\nu_e$  per second, which means

65 billions of neutrinos per second per  $\text{cm}^2$  on Earth. Supernova explosions emit about  $10^{58}$  neutrinos in a few seconds and the central engines of active galactic nuclei produce them abundantly.

## 2.2 How to detect neutrinos

Neutrinos rarely interact with ordinary matter, so they are very hard to detect. We can only detect the presence of a neutrino in our laboratory when it interacts through CC or NC interactions. Since a neutrino turns into its partner lepton via a CC interaction, the detection of a charged lepton is considered as a sign of a neutrino event. But, if the neutrino does not have sufficient energy to create its partner lepton, CC interaction is effectively unavailable to it. The neutrino enters and then leaves the detector after having transferred some of its energy and momentum to a target particle via NC interactions. The by-product created by the target particle hit by the neutrino is considered as a sign of a neutrino event. All three neutrinos can participate in NC interactions, regardless of the neutrino energy. In principle, CC interactions are easier to work with, because electrons and muons have characteristic signatures in particle detectors and are thus fairly easy to identify. They also have the advantage that they “flavor-tag” the neutrino. Various different detector technologies have been used in neutrino experiments over the years, depending on the requirements of the particular study. Among them, the following three are the most popular ones widely used.

- **Radiochemical experiments:** The lowest energy thresholds are provided by radiochemical experiments, in which the neutrino is captured by an atom which then (through inverse beta decay, a CC) converts into another element. The classic example is the chlorine solar neutrino experiment. Even lower thresholds were achieved by using gallium as the target: the reaction  ${}^{71}\text{Ga} + \nu \rightarrow {}^{71}\text{Ge} + e^-$  has a threshold of only 0.233 MeV, and is even sensitive to  $pp$  neutrinos from the Sun. The produced isotope is unstable, and will decay back to the original element: neutrinos are counted by extracting the product and observing these decays. Examples of radiochemical experiments are Homestake (Ray Davis; chlorine); SAGE (gallium); GALLEX/GNO (gallium).
- **Liquid scintillator (LS) experiments:** LS has an impressive pedigree as neutrino detectors, since the neutrino was originally discovered using a LS detector. They are primarily sensitive to  $\bar{\nu}_e$ 's, which initiate inverse beta decay of a proton. Being organic compounds, LS is rich in hydrogen nuclei which act as targets for this reaction. The positron promptly annihilates, producing two gamma rays; the neutron is captured on a nucleus after a short time (a few microseconds to a few hundred microseconds), producing another gamma-ray signal. This coincidence of a prompt signal and a delayed signal allows the experiment to reject background effectively. LS detectors have good time and energy resolution, but do not preserve directional information. Examples of LS experiments are Borexino, KamLAND, MiniBooNE and SNO+.
- **Water Cherenkov experiments:** A particle travelling through a transparent medium at faster than the speed of light in that medium emits a kind of “light boom” – a coherent cone of blue light known as Cherenkov radiation. The particle is travelling down the axis of the cone, so if the cone can be reconstructed the direction of the particle can be measured. In a water Cherenkov detector, the Cherenkov radiation is detected, usually by photomultiplier tubes, and the cone of emission reconstructed. The axis of the cone gives the direction of the particle, and the light yield gives the

particle energy. Only charged particles with  $\beta > 1/n$  can be detected, which gives a threshold total energy of about 0.8 MeV for electrons, 160 MeV for muons and 1.4 GeV for protons and neutrons. Neutrinos are detected in the detectors when they interact by  $W$  exchange, converting into muon or electron for  $\nu_\mu$  or  $\nu_e$  respectively, or when they elastically scatter off electrons (when the recoil electron can be detected). Examples of densely instrumented water Cherenkov experiments are SK, far detector for the K2K and T2K experiments, and IMB.

### 3 Neutrino oscillation

#### 3.1 History of neutrino oscillation

The idea of neutrino oscillation was first put forward in 1957 by Bruno Pontecorvo, who proposed that neutrino–antineutrino transitions may occur in analogy with neutral kaon mixing [11]. Although such a matter–antimatter oscillation had not been observed, this idea formed the conceptual foundation for the quantitative theory of neutrino flavor oscillation, which was first developed by Maki, Nakagawa, and Sakata (MNS) in 1962 [12] and further elaborated by Pontecorvo in 1967 [13], who developed the modern theory of neutrino oscillation in vacuum where the new ingredient is the mixing of different families of neutrinos introduced by MNS. One year later the solar neutrino deficit was first observed in Homestake in 1968 [14], and that was followed by the paper by Gribov and Pontecorvo published in 1969 [15]. Atmospheric neutrino experiments, IMB [17] and Kamiokande-II [18], found an anomaly in the ratio of the flux of muon to electron neutrinos. The SK reported the first evidence of the atmospheric neutrino oscillations in 1998 [1], and SNO experiments provided clear evidence of solar neutrino oscillations in 2001 [16]. Thanks to the discovery of neutrino oscillations, Kajita (SK) and McDonald (SNO) received the Nobel prize in 2015.

#### 3.2 Neutrino mixing

To understand what neutrino oscillation is and how it occurs, we need to discriminate two kinds of neutrino quantum eigenstates: one is the flavor eigenstate and the other is the mass eigenstate. Flavor eigenstates, denoted by  $(\nu_e, \nu_\mu, \nu_\tau)$ , are the quantum states produced or detected via weak interactions together with charged leptons with the same flavor  $(e, \mu, \tau)$ . Mass eigenstates, denoted by  $(\nu_1, \nu_2, \nu_3)$ , are the states of definite masses that are created by the interactions with Higgs boson or other mechanisms. The mismatch between flavor states and mass states of neutrinos gives rise to neutrino mixing. Then, a specific flavor state of a neutrino is given by a superposition of three mass eigenstates having definite masses, written as

$$\nu_l = \sum_{i=1}^N U_{li} \nu_i \quad , \quad (1)$$

where  $l = e, \mu, \tau$  and  $i = 1, 2, 3$ , and  $U_{li}$  denotes a  $3 \times 3$  unitary matrix, the so-called Pontecorvo–Maki–Nakagawa–Sakita mixing matrix. The neutrino mixing elements mean how much each flavor can contribute to the composition of each mass states, and how much each flavor is accompanied by a certain mass eigenstate in the weak interaction. The  $3 \times 3$  unitary mixing matrix is generally represented by 3 mixing angles and 6 phases. However, not all phases are physical observables. Let us see how many



phases are physical by assuming neutrinos are Dirac particles. Under global phase transformations of the leptons, neutrinos ( $\nu_k$ ) and charged leptons ( $l_\alpha$ ) are transformed as  $\nu_k \rightarrow e^{i\varphi_k} \nu_k$  ( $k = 1, 2, 3$ ) and  $l_\alpha \rightarrow e^{i\varphi_\alpha} l_\alpha$  ( $\alpha = e, \mu, \tau$ ). Then, the terms of the Lagrangian for the CC interactions in the mass basis,  $\mathcal{L}_{cc} = \frac{g}{\sqrt{2}} \bar{\nu}_k U^\dagger \gamma^\mu l_\alpha W_\mu^+$ , become

$$\sum_{k=1}^3 \sum_{\alpha=e,\mu,\tau} \bar{\nu}_{kL} e^{-i\varphi_k} U_{\alpha k}^* e^{i\varphi_\alpha} \gamma^\mu l_{\alpha L} \Rightarrow e^{-i(\varphi_1 - \varphi_e)} \sum_{k=1}^3 \sum_{\alpha=e,\mu,\tau} \bar{\nu}_{kL} e^{-i(\varphi_k - \varphi_1)} U_{\alpha k}^* e^{i(\varphi_\alpha - \varphi_e)} \gamma^\mu l_{\alpha L} . \quad (2)$$

In the right-handed side of Eq. (2), 5 phases can be eliminated by redefining lepton fields. Finally, we see that the  $3 \times 3$  unitary mixing matrix can be expressed in terms of 3 mixing angles and 1 phase. In the standard parameterization, the lepton mixing matrix can be expressed as

$$\begin{aligned} U &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13} e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13} e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} c_{12} c_{13} & s_{12} c_{13} & s_{13} e^{-i\delta_{\text{CP}}} \\ -s_{12} c_{23} - c_{12} s_{23} s_{13} e^{i\delta_{\text{CP}}} & c_{12} c_{23} - s_{12} s_{23} s_{13} e^{i\delta_{\text{CP}}} & s_{23} c_{13} \\ s_{12} s_{23} - c_{12} c_{23} s_{13} e^{i\delta_{\text{CP}}} & -c_{12} s_{23} - s_{12} c_{23} s_{13} e^{i\delta_{\text{CP}}} & c_{23} c_{13} \end{pmatrix} , \end{aligned} \quad (3)$$

where  $c_{ij}$  and  $s_{ij}$  denote  $\cos \theta_{ij}$  and  $\sin \theta_{ij}$ , respectively, and  $\delta_{\text{CP}}$  is a Dirac CP violating phase. If neutrinos are Majorana particles, then there exist two extra CP violating phases. As will be shown later, the mixing angles  $\theta_{23}$ ,  $\theta_{12}$  and  $\theta_{13}$  are associated with oscillations of atmospheric neutrinos, solar neutrinos and reactor neutrinos, respectively.

### 3.3 Neutrino oscillations in vacuum

Neutrino oscillations are quantum mechanical phenomena where flavor and mass eigenstates mismatch. Then, flavor eigenstates of neutrinos are superpositions of mass eigenstates, written as  $|\nu_\alpha\rangle = \sum_k U_{\alpha k} |\nu_k\rangle$  with  $\alpha = (e, \mu, \tau)$ ,  $k = (1, 2, 3)$ . The time evolution of flavor eigenstates is given by

$$\begin{aligned} |\nu_\alpha(t, x)\rangle &= \sum_k U_{\alpha k} e^{-i(E_k t + p_k x)} |\nu_k\rangle = \sum_\beta \left( \sum_k U_{\alpha k} e^{-i(E_k t + p_k x)} U_{\beta k}^* \right) |\nu_\beta\rangle \\ &= \sum_\beta \mathcal{A}_{\nu_\alpha \rightarrow \nu_\beta}(t, x) |\nu_\beta\rangle , \end{aligned} \quad (4)$$

where  $\mathcal{A}_{\nu_\alpha \rightarrow \nu_\beta}(t, x)$  denotes the amplitude of the neutrino flavor transition and  $E_k, p_k$  are neutrino energy and momentum. From Eq. (4), one can easily obtain the probability of the neutrino flavor transition,

$$P_{\nu_\alpha \rightarrow \nu_\beta}(t, x) = |\mathcal{A}_{\nu_\alpha \rightarrow \nu_\beta}(t, x)|^2 = \left| \sum_k U_{\alpha k} e^{-i(E_k t + p_k x)} U_{\beta k}^* \right|^2 . \quad (5)$$

In natural units,  $t = x$ , we can make an approximation as follows:

$$E_k t - p_k x \simeq (E_k - p_k)L = \frac{E_k^2 - p_k^2}{E_k + p_k}L = \frac{m_k^2}{E_k + p_k}L \simeq \frac{m_k^2}{2E}L, \quad (6)$$

where  $L$  is the distance of neutrino propagation. Then, the probability is given by

$$P_{\nu_\alpha \rightarrow \nu_\beta}(t, x) = \left| \sum_k U_{\alpha k} e^{-im_k^2 L/2E} U_{\beta k}^* \right|^2 = \sum_{k,j} U_{\alpha k} U_{\beta k}^* U_{\alpha j} U_{\beta j}^* \text{Exp} \left( -i \frac{\Delta m_{kj}^2 L}{2E} \right), \quad (7)$$

where  $\Delta m_{kj}^2 \equiv m_k^2 - m_j^2$ .

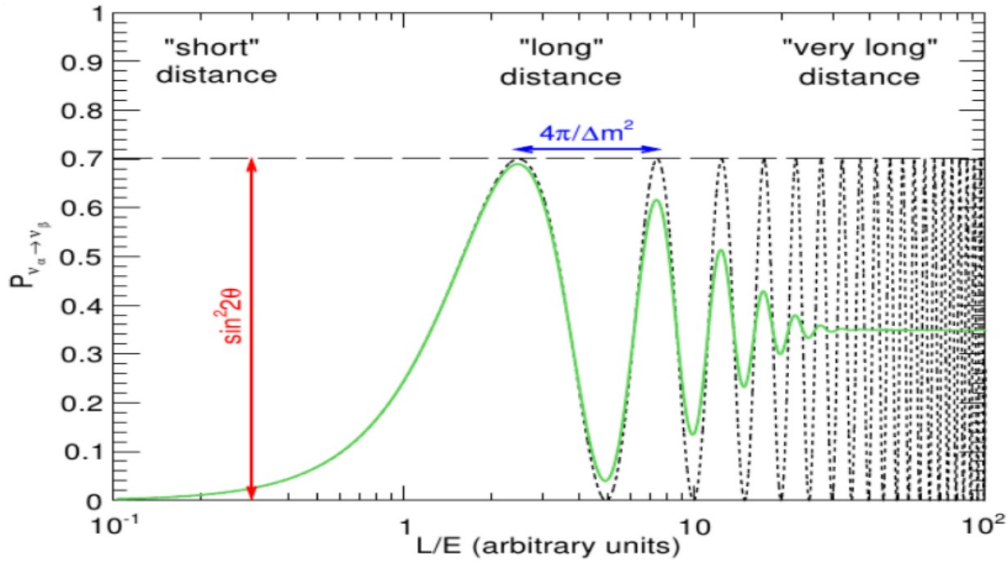
As an example, let us consider the two-flavor oscillation with  $\nu_e$  and  $\nu_\mu$ , for which the states produced and detected are supposedly  $|\nu_\mu\rangle = -\sin\theta|\nu_1\rangle + \cos\theta|\nu_2\rangle$  and  $|\nu_e\rangle = \cos\theta|\nu_1\rangle + \sin\theta|\nu_2\rangle$ , respectively. For this case, the transition (appearance) probability is given by

$$P_{\nu_\mu \rightarrow \nu_e}(L, E) = 2 \sin^2 \theta \cos^2 \theta \left( 1 - \cos \left( \frac{\Delta m_{21}^2 L}{2E} \right) \right) = \sin^2 2\theta \sin^2 \left( \frac{\Delta m_{21}^2 L}{4E} \right). \quad (8)$$

In the above expression, there are two fundamental parameters,  $\theta$  and  $\Delta m_{21}^2$ , which are determined from neutrino oscillation experiments for given  $L$  and  $E$ . The neutrino oscillation length is defined by  $L^{\text{osc}} \equiv \frac{4\pi E}{\Delta m^2}$ . The so-called survival (disappearance) probability is given by

$$P_{\nu_\mu \rightarrow \nu_\mu}(L, E) = 1 - P_{\nu_\mu \rightarrow \nu_e}(L, E) = 1 - \sin^2 2\theta \sin^2 \left( \frac{\Delta m_{21}^2 L}{4E} \right). \quad (9)$$

Figure 1 shows how the transition probability evolves along with  $L/E$ . The maximum height of the



**Fig. 1:** Plot of the transition probability  $P_{\nu_\alpha \rightarrow \nu_\beta}(L, E)$  in terms of  $L/E$  (figure taken from Ref. [19]).

curve corresponds to  $\sin^2 2\theta$ , and the horizontal distance between the first and second peak is  $4\pi/\Delta m^2$ . At very long distances, it is averaged out to be half of the maximum height.

Extending to the three-flavor paradigm, the probability of neutrino oscillation is given explicitly by

$$\begin{aligned}
 P_{\nu_\alpha \rightarrow \nu_\beta} = & \underbrace{\delta_{\alpha\beta} - 4 \sum_{i < j}^3 \operatorname{Re}(U_{\alpha i} U_{\beta i} U_{\alpha j}^* U_{\beta j}^*) \sin^2 \left( \frac{\Delta m_{ji}^2 L}{2E} \right)}_{\text{CP conserving part: } = P_{\nu_\alpha \rightarrow \nu_\beta}^{\text{CPC}}} \\
 & + \underbrace{2 \sum_{i < j}^3 \operatorname{Im}(U_{\alpha i} U_{\beta i} U_{\alpha j}^* U_{\beta j}^*) \sin \left( \frac{\Delta m_{ji}^2 L}{2E} \right)}_{\text{CP violating part: } = P_{\nu_\alpha \rightarrow \nu_\beta}^{\text{CPV}}}, \quad (10)
 \end{aligned}$$

where the first two terms correspond to the CP conserving part ( $\equiv P_{\nu_\alpha \rightarrow \nu_\beta}^{\text{CPC}}$ ), whereas the last one to the CP violating part ( $\equiv P_{\nu_\alpha \rightarrow \nu_\beta}^{\text{CPV}} = 8J \sum_{\gamma} \epsilon_{\alpha\beta\gamma} \sin \frac{\Delta m_{21}^2 L}{4E} \sin \frac{\Delta m_{31}^2 L}{4E} \sin \frac{\Delta m_{32}^2 L}{4E}$ ). In the limit that  $\Delta m_{21}^2 = \Delta m_{\text{sol}}^2 \ll |\Delta m_{\text{atm}}^2| = |\Delta m_{31}^2| \simeq |\Delta m_{32}^2|$ , the survival and transition probabilities are approximately given by

$$\begin{aligned}
 P_{\nu_\alpha \rightarrow \nu_\alpha} & \simeq 1 - 4|U_{\alpha 1}|^2 |U_{\alpha 2}|^2 \sin^2 \left( \frac{\Delta m_{21}^2 L}{4E} \right) - 4(1 - |U_{\alpha 3}|^2) |U_{\alpha 3}|^2 \sin^2 \left( \frac{\Delta m_{31}^2 L}{4E} \right), \\
 P_{\nu_\alpha \rightarrow \nu_\beta} & \simeq -4(U_{\alpha 1} U_{\beta 1} U_{\alpha 2}^* U_{\beta 2}^*) \sin^2 \left( \frac{\Delta m_{21}^2 L}{4E} \right) + 4|U_{\alpha 3}|^2 |U_{\beta 3}|^2 \sin^2 \left( \frac{\Delta m_{31}^2 L}{4E} \right). \quad (11)
 \end{aligned}$$

In addition, since the mixing angle  $\theta_{13}$  is small compared with the two others,  $|U_{e3}|^2 \ll |U_{e1}|^2, |U_{e2}|^2$ , and  $U_{e1} \simeq \cos \theta_{12}, U_{e2} \simeq \sin \theta_{12}$ . Adopting those approximations, the electron neutrino survival probability is simply given by

$$P_{\nu_e \rightarrow \nu_e} \simeq 1 - \sin^2 2\theta_{12} \sin^2 \left( \frac{\Delta m_{21}^2 L}{4E} \right). \quad (12)$$

This result shows that the effect of solar neutrino oscillations is decoupled from that of atmospheric neutrino oscillations. So, the formula (12) is good at probing solar neutrinos. On the other hand, in the case that we can ignore the oscillating terms involving  $\Delta m_{21}^2$ , the oscillation probabilities are given by

$$\begin{aligned}
 P_{\nu_\alpha \rightarrow \nu_\alpha} & \simeq 1 - 4(1 - |U_{\alpha 3}|^2) |U_{\alpha 3}|^2 \sin^2 \left( \frac{\Delta m_{31}^2 L}{4E} \right), \\
 P_{\nu_\alpha \rightarrow \nu_\beta} & \simeq 4|U_{\alpha 3}|^2 |U_{\beta 3}|^2 \sin^2 \left( \frac{\Delta m_{31}^2 L}{4E} \right). \quad (13)
 \end{aligned}$$

These expressions are relevant to the atmospheric and short baseline reactor neutrino experiments.

### 3.4 Neutrino oscillations in matter

Wolfenstein for the first time studied a matter effect that may convert the flavor of a neutrino into another one [20]. When neutrinos pass through matter, they experience forward scattering, mostly from electrons they encounter along the way, which represents a contribution to the Hamiltonian with a potential proportional to the density of electrons in matter. Similarly to optics, the net effect of this coherent elastic scattering is the appearance of a phase difference, a refractive index, or equivalently, a neutrino effective mass. As a result, the oscillation probability can be rather different from that in vacuum. For the case of

the two-flavor scenario, the  $2 \times 2$  Hamiltonian in matter is given by

$$\mathcal{H}_M = \mathcal{H}_{\text{vac}} + \begin{pmatrix} V_e & 0 \\ 0 & V_\mu \end{pmatrix} = \begin{pmatrix} V & 0 \\ 0 & 0 \end{pmatrix} + \lambda'' I, \quad (14)$$

where  $V_{\alpha(=e,\mu)}$  denotes the potential energy associated with  $\nu_\alpha$  and the term proportional to the unit matrix  $I$  is irrelevant for flavor evolution.  $V$  is defined by  $V_e - V_\mu = \sqrt{2}G_F N_e$ , where  $G_F$  and  $N_e$  are the Fermi constant and electron number density in matter, respectively.  $\mathcal{H}_{\text{vac}}$  is given by

$$\mathcal{H}_{\text{vac}} = \frac{\Delta m^2}{4E} \begin{pmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix}, \quad (15)$$

where we have used  $p \simeq E$ , with  $E$  being the average energy of the neutrinos. If  $N_e$  is constant, diagonalizing  $\mathcal{H}_M$  leads us to modifications of the mixing angle and of the mass-squared difference, denoted by  $\theta_M$  and  $\Delta m_{M}^2$ , respectively. The explicit forms of  $\theta_M$  and  $\Delta m_{M}^2$  are given as

$$\begin{aligned} \tan 2\theta_M &= \frac{\tan 2\theta}{1 - \frac{A_{CC}}{\Delta m^2 \cos 2\theta}}, \\ \Delta m_{M}^2 &= \sqrt{(\Delta m^2 \cos 2\theta - A_{CC})^2 + (\Delta m^2 \sin 2\theta)^2}, \end{aligned} \quad (16)$$

where  $A_{CC} = 2\sqrt{2}G_F N_e E$  and  $(\theta, \Delta m^2)$  are the parameters in vacuum. Such a modification of the parameters in matter leads to a shift of the mass eigenstates, which are related to flavor eigenstates as follows:

$$\begin{aligned} |\nu_e\rangle &= \cos \theta_M |\nu_{1m}\rangle + \sin \theta_M |\nu_{2m}\rangle, \\ |\nu_\mu\rangle &= -\sin \theta_M |\nu_{1m}\rangle + \cos \theta_M |\nu_{2m}\rangle, \end{aligned} \quad (17)$$

where  $|\nu_{im}\rangle$  denotes the mass eigenstate in matter. Mikheyev and Smirnov found that when  $A_{CC} = \Delta m^2 \cos 2\theta$ , resonance occurs and neutrino mixing becomes maximal, with  $\theta_M = \pi/4$  [21]. In a medium with constant density, there is no transition between  $\nu_{1m}$  and  $\nu_{2m}$ , which are the eigenstates of propagation, so oscillation probability is simply given by  $P_{\nu_e \rightarrow \nu_\mu} \simeq \sin^2 2\theta_M \sin^2 \left( \frac{\Delta m_{M}^2 L}{4E} \right)$ , which is similar to the probability in vacuum.

In case that matter density varies with time, it is hard to solve the time-dependent Schrödinger equations for neutrinos analytically. In this case,  $\nu_{1m}$  and  $\nu_{2m}$  are not propagation eigenstates any longer and transition between them occurs. Let us suppose that  $\nu_e$  is produced in matter and detected later in vacuum. Then, the flavor eigenstates in the production and detection are given by

$$\begin{aligned} \text{production :} \quad & |\nu_e\rangle = \cos \theta_M |\nu_{1m}\rangle + \sin \theta_M |\nu_{2m}\rangle, \\ \text{detection :} \quad & |\nu_e(x)\rangle = \cos \theta |\nu_1(x)\rangle + \sin \theta |\nu_2(x)\rangle. \end{aligned} \quad (18)$$

Neglecting the interference term, the average survival (appearance) probability is given by [22]

$$\bar{P}_{\nu_e \rightarrow \nu_\mu}(x) = \frac{1}{2} + \left( \frac{1}{2} - P_c \right) \cos 2\theta_M \cos 2\theta, \quad (19)$$

where  $P_c$  represents the transition probability between  $\nu_{1m}$  and  $\nu_{2m}$ . An interesting limit exists, which is called ‘‘adiabatic limit’’, where the time evolution of the neutrino state is sufficiently slow, and then each state evolves independently and transitions between  $\nu_{1m}$  and  $\nu_{2m}$  can be neglected. In the adiabatic limit,  $P_c = 0$  and then the probability  $\bar{P}_{\nu_e \rightarrow \nu_\mu}$  becomes

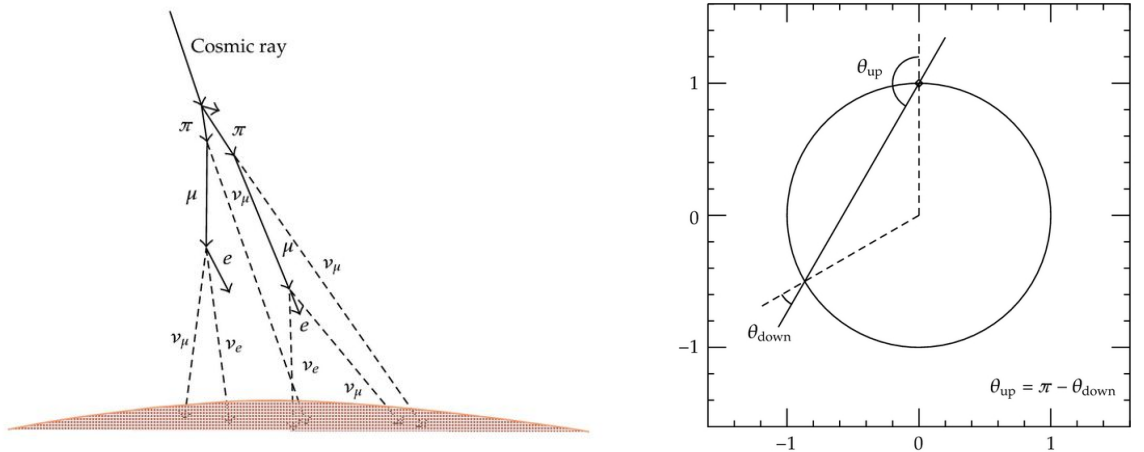
$$\bar{P}_{\nu_e \rightarrow \nu_\mu}(x) = \cos^2 \theta \cos^2 \theta_M + \sin^2 \theta \sin^2 \theta_M. \quad (20)$$

There are two interesting limits. In the limit that  $\Delta m^2/2E \ll \sqrt{2}G_F N_e$ ,  $\theta_M$  goes to  $\pi/2$  and then the probability becomes  $\bar{P}_{\nu_e \rightarrow \nu_\mu} \simeq \sin^2 \theta$ . In the limit that  $\Delta m^2/2E \gg \sqrt{2}G_F N_e$ ,  $\theta_M$  goes to the vacuum angle and then the probability becomes  $\bar{P}_{\nu_e \rightarrow \nu_\mu} \simeq 1 - \frac{1}{2} \sin^2 2\theta$ . These results are useful to interpret the flux deficits of solar neutrinos observed at various experiments, as will be discussed later.

## 4 Neutrino experiments

### 4.1 Atmospheric neutrino experiments

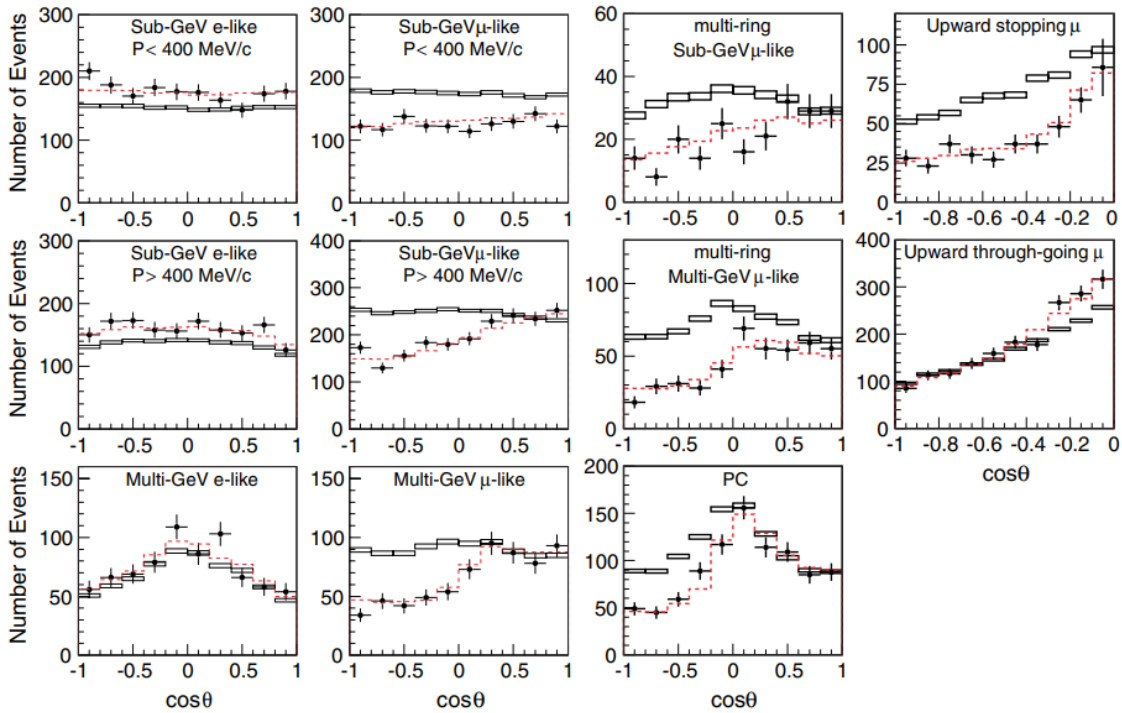
Neutrino oscillation was discovered for the first time through the studies of atmospheric neutrinos, which are produced by cosmic-ray interactions with nuclei in the atmosphere. Electron-neutrinos and muon-neutrinos are produced mainly by the decay chain of charged pions to muons and to electrons. The event ratio of  $\nu_\mu$  to  $\nu_e$ ,  $R_{\mu/e} = (N_{\nu_\mu} + N_{\bar{\nu}_\mu})/(N_{\nu_e} + N_{\bar{\nu}_e})$ , is expected to be nearly 2 below about 1 GeV, based on the calculations of neutrino fluxes produced from interactions in the atmosphere, as shown in Ref. [23]. Above this energy, the ratio increases due to the increasing probability of muons reaching the ground before their decay. So, whether the ratio remains 2 in the detection of atmospheric neutrinos or not is a good indicator for neutrino oscillation. In early-stage experiments such as Soudan 2 [24], IMB [17] and Kamiokande [18], a deficit of  $R_{\mu/e}$  was observed, but it was unclear whether this was due to neutrino oscillations or not. Another important hint toward the understanding of the atmospheric neutrino flux deficit was given in Ref. [25]. The atmospheric neutrinos enter the spherical Earth at a point



**Fig. 2:** Schematic figure of the production of atmospheric neutrinos (left) and a neutrino trajectory that enters a spherical Earth with a zenith angle  $\theta_{\text{down}}$  and exits with  $\theta_{\text{up}}$  (right). These figures are taken from Ref. [26].

with a zenith angle  $\theta_{\text{down}}$  and should exit the Earth at a point with  $\theta_{\text{up}}$ . Obviously,  $\theta_{\text{down}}$  and  $\theta_{\text{up}}$  are related by  $\theta_{\text{down}} = \pi - \theta_{\text{up}}$ , as can be seen in Fig. 2. Since cosmic rays enter into the atmosphere with approximately equal rate in every position in the Earth, the numbers of downward-going and upward-going neutrinos must be the same. Thus the flux is expected to be up-down symmetric. The Kamiokande data [18] showed that the deficit of  $\mu$ -like events depended on the zenith angle, but with relatively poor event statistics.

In 1996, a much larger detector, SK, started taking data. The SK experiment used a 50 kt water Cherenkov detector, and has obtained data with substantially improved statistics in 1998. The events observed in SK are classified by 4 types. Events where vertex positions are located inside the fiducial volume of the detector and all visible secondary particles stop in the detector are called ‘‘fully-contained’’ (FC) events. The ‘‘partially-contained’’ (PC) events are  $\nu_{\mu}$  events with multi-GeV neutrino energies, producing energetic muons which do not stop in the detector. High-energy  $\nu_{\mu}$  interactions in the rock below the detector produce high-energy muons, which enter into the detector. Some of them stop in the detector and are called ‘‘upward stopping muons’’, others penetrate through the detector and are called ‘‘upward through-going muons’’. Figure 3 shows several plots of zenith angle dependence for the results

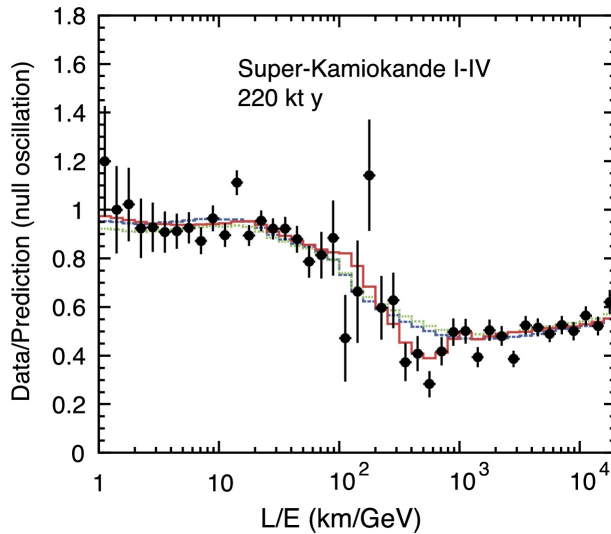


**Fig. 3:** Zenith angle distributions of  $\mu$ - and  $e$ -like events at sub-GeV and multi-GeV scales from the SK 92 kt-year data for various data samples [26].  $\cos\theta = 1(-1)$  corresponds to down(up)-going. The solid-line histograms show the prediction without neutrino oscillations. The dashed-line histograms show the prediction with oscillation ( $\nu_{\mu} \rightarrow \nu_{\tau}$ ) for  $\Delta m_{32}^2 = 2.1 \times 10^{-3} \text{ eV}^2$  and  $\sin^2 2\theta_{23} = 1.0$ .

obtained by SK [26]. From the results, we see that the deficit of upward-going  $\mu$ -like events depends on the zenith angle in the multi-GeV energy range, the  $\nu_{\mu}$  to  $\nu_e$  event ratio is smaller than what was expected, the ratio of upward-going stopping/through-going muons is smaller than what was expected,

and the zenith angle distribution for the upward through-going muons is distorted. These results represent an evidence of the neutrino oscillation for which muon-neutrinos convert to other flavors of neutrinos through their flight inside the Earth.

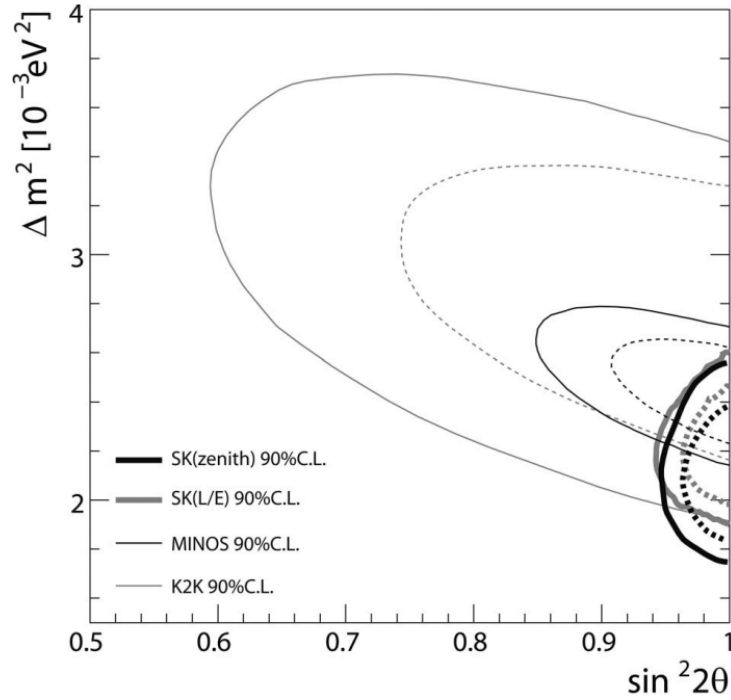
According to the neutrino oscillation formulae, the neutrino survival probability should follow the sinusoidal function. The  $\nu_\mu$  survival probability should have a minimum at a certain  $L/E$  value, come back to unity after traveling twice the distance, and continue oscillating. Using high  $L/E$  resolution events only, SK found that the  $\nu_\mu$  survival probability shows a dip at a position corresponding to the first minimum of the survival probability. Figure 4 shows the updated plot based on the 220 kt-yrs data of



**Fig. 4:** The ratio of data to MC events without oscillation as a function of the reconstructed  $L/E$ , together with the best-fit 3-flavor expectation for neutrino oscillation and two alternative hypotheses with similar shape. The dashed (blue) and dotted (green) lines show the best-fit expectations for neutrino decay and neutrino decoherence, respectively. Figure taken from Ref. [27].

SK-I through SK-IV [27]. This was the first evidence that the  $\nu_\mu$  survival probability is represented by a sinusoidal function as predicted by neutrino oscillations. In Fig. 4, the expected  $\nu_\mu$  survival probabilities by neutrino oscillations as well as those from alternative models, which were able to explain the zenith angle distributions, are shown with the detector  $L/E$  resolution taken into account. It is clear that the alternative models cannot reproduce the dip seen near  $L/E = 500$  km/GeV. Thanks to this result, alternative hypotheses for the atmospheric neutrino flux deficit, such as neutrino decay (blue dashed) and neutrino decoherence (green dotted), are ruled out.

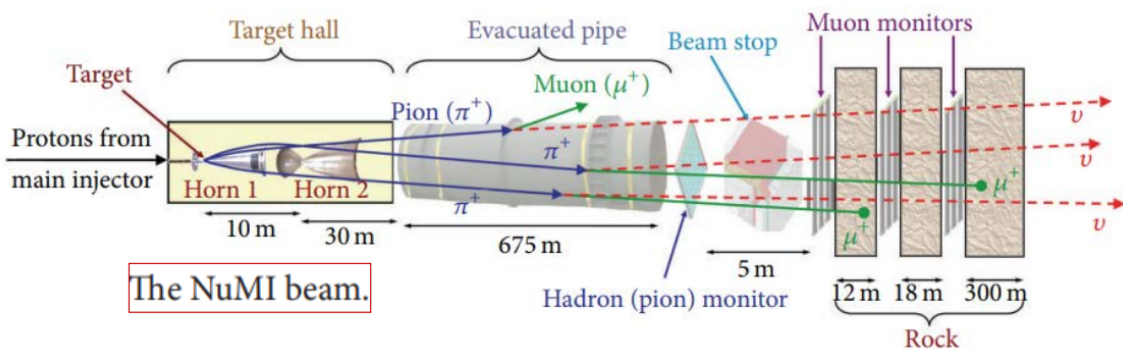
Figure 5 shows the allowed regions of neutrino oscillation parameters ( $\Delta m^2$ ,  $\sin^2 2\theta$ ) at 68% (dashed curves) and 90% (solid curves) C.L. from several experiments [27]. The thick-black and thick-gray curves represent the allowed regions based, respectively, on the zenith angle analysis and the  $L/E$  analyses in SK. Results from K2K (thin-gray) and MINOS (thin-black) experiments are plotted. The mixing angle is consistent with the maximum mixing ( $\sin^2 2\theta = 1.0$ ). These parameters are much more accurately measured compared with those in 1998.



**Fig. 5:** Allowed regions of ( $\Delta m^2$  and  $\sin^2 2\theta$ ) at 68 (dashed lines) and 90% (solid lines) C.L. from various experiments (figure taken from Ref. [27]). Thick-black and thick-gray lines show the allowed regions based on the zenith-angle analysis and  $L/E$  analyses in SK, respectively. Also shown are the allowed regions from K2K (thin-gray lines) and MINOS (thin-black lines) experiments.

## 4.2 Accelerator based neutrino experiments

One of the ways physicists can study neutrinos effectively is by making intense neutrino beams using proton accelerators. The neutrinos produced in accelerators are typically muon neutrinos, and the machine can be tuned to create either neutrinos or antineutrinos. Figure 6 represents an overview of the neutrino beam production from NuMI (Neutrinos at Main Injector), which is used at Fermilab to create an intense beam of neutrinos aimed toward detectors in experiments such as MINOS, MINER $\nu$ A, NO $\nu$ A etc. Accelerator neutrinos are used to study neutrino interactions and neutrino oscillations taking advan-

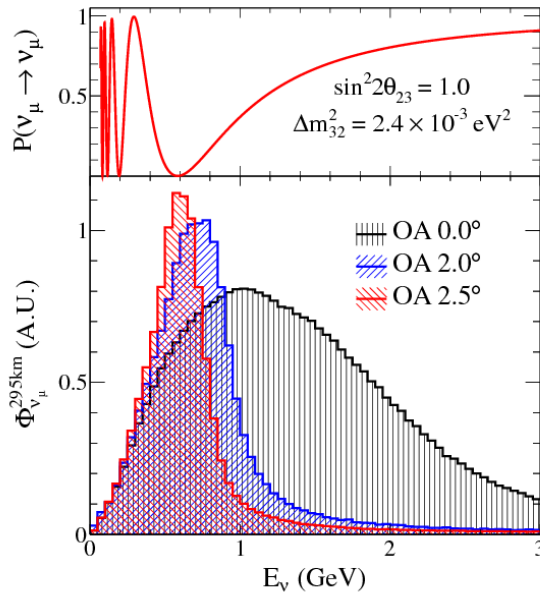


**Fig. 6:** Overview of neutrino beam production from NuMI [28].



tage of high-intensity neutrino beams, as well as a possibility of controlling and understanding their type and kinematic properties to a much larger extent than for neutrinos from other sources. Neutrino beams can be used for two different kinds of experiments, distinguished by how far away the detectors are from where the neutrinos are made: short baseline (SBL) and long baseline (LBL) experiments. In the case of SBL experiments using accelerator neutrinos, the detector sits close to the neutrino source, so that the beam is very concentrated when it reaches the detector. The experiments are good for characterizing the beam and learning about the neutrinos before they oscillate, and are also a good place to hunt for sterile neutrinos and see how neutrinos interact with other particles. In the case of LBL experiments, they focus on the oscillations while traveling a long distance through the Earth. Neutrinos have many opportunities to interact with matter and have sufficient distance to change flavors. They are a good place to figure out mass ordering (MO) and CP violation in the neutrino sector.

The off-axis neutrino beam results in a narrow band energy distribution of the produced neutrinos, due to the correlation between the off-axis angle and neutrino energy [29]. The accelerator neutrino beam is primarily a wide beam that has no clear boundaries, because the neutrinos in it do not move in parallel, but have a certain angular distribution. The further away from the axis of the beam, the smaller the number of neutrinos, and the distribution of energy also changes. The energy spectrum becomes narrower and its maximum moves to lower energy. The off-axis angle can be optimized to maximize the neutrino oscillation probability or to select an energy range in which the desired type of neutrino interaction is dominant.



**Fig. 7:**  $\nu_\mu$  survival probability at 295 km (up) and  $\nu_\mu$  flux (down) versus  $E_\nu$  at T2K. Back, blue and red regions correspond to the flux for axis angles of  $0^\circ$ ,  $2.0^\circ$ ,  $2.5^\circ$ , respectively [32].

The first experiment with an off-axis neutrino beam was the T2K experiment, a LBL experiment located in Japan [30]. The main components of T2K include a neutrino beam line, muon monitors, a near detector complex ND280 located at 280 m from the proton interaction target, and the far detector of SK (295 km from the neutrino source), at a 2.5 degree off-axis angle from the beam. The physics goals

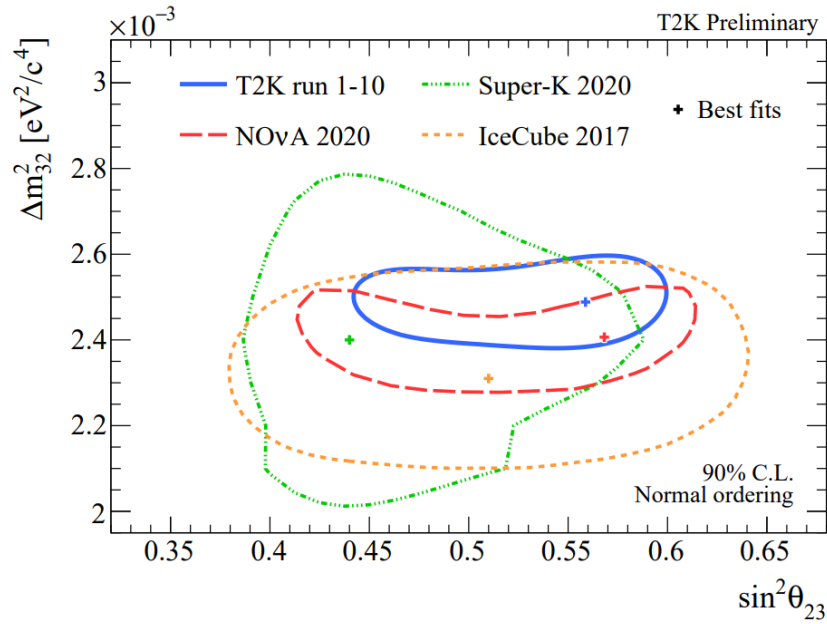
of T2K are (1) to improve the measurement of  $\Delta m_{32}^2$  and  $\sin^2 \theta_{23}$  using the  $\nu_\mu$  disappearance analysis, and (2) to measure for the first time the  $\theta_{13}$  oscillation parameter through the  $\nu_e$  appearance analysis. As illustrated in the lower panel of Fig. 7, the peak energy of the neutrino beam varies with different off-axis angles. For T2K, the off-axis angle is at  $2.5^\circ$ , so that the neutrino beam at SK has a peak energy at about 0.6 GeV, close to the expected first oscillation maximum [32]. In 2011, T2K first announced the observation of  $\nu_e$  appearance events in a  $\nu_\mu$  beam [31]. It also provided the world's best measurement of the oscillation parameter  $\theta_{23}$  and the first hint of CP violation in neutrino oscillations.

The MINOS experiment is designed to study the phenomena of neutrino oscillations by making precision neutrino oscillation measurements using the neutrino beam produced by NuMI [33]. Neutrinos are observed in two detectors, one very close to where the beam is produced (near detector), and another much larger detector 735 km away from northern Minnesota (far detector). MINOS measures the difference in neutrino beam composition and energy distribution in both detectors for the purpose of precision measurements of  $\Delta m_{23}^2$  and  $\theta_{23}$ . In addition, MINOS looks for the appearance of  $\nu_e$  in the far detector, and will either measure or set a limit on the oscillation probability of  $\nu_\mu$  into  $\nu_e$ . By observing the disappearance of  $\nu_\mu$ , MINOS has made the world's most precise measurement of the larger neutrino mass splitting and has measured  $\theta_{23}$  [34]. Using a dedicated antineutrino beam, MINOS has also made the first direct precision measurements of the corresponding antineutrino parameters [34]. A search for  $\nu_e$  and  $\bar{\nu}_e$  appearance has enabled a measurement of the mixing angle  $\theta_{13}$ . MINOS has performed the first search for  $\bar{\nu}_e$  appearance in a  $\nu_\mu$  beam and the first search for  $\nu_e$  and  $\bar{\nu}_e$  appearance with significant matter effects. MINOS will continue as MINOS+ [35] using an upgraded beam with higher energy and intensity, allowing precision tests of the three-flavour neutrino oscillation picture, in particular a very sensitive search for the existence of sterile neutrinos.

The NO $\nu$ A experiment [36] is designed to mainly observe the oscillation of  $\nu_\mu$  to  $\nu_e$  by using the NuMI beam and consists of two detectors, one in Fermilab and the other in northern Minnesota, allowing neutrinos to travel more than 810 km. It is also capable of measuring  $\delta_{\text{CP}}$  through the comparison between the  $\nu_\mu \rightarrow \nu_e$  and  $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$  oscillation channels. Recently, NO $\nu$ A [37] provided a less precise measurement of  $\delta_{\text{CP}}$ , which is in slight tension with the T2K result (as will be shown later). Figure 8 shows the allowed regions of oscillation parameters ( $\Delta m_{32}^2, \sin^2 \theta_{23}$ ) at 90% C.L. We note that the contours overlap. NO $\nu$ A, T2K and IceCube prefer the upper octant of  $\theta_{23}$ , while SK prefers the lower one.

### 4.3 Solar neutrino experiments

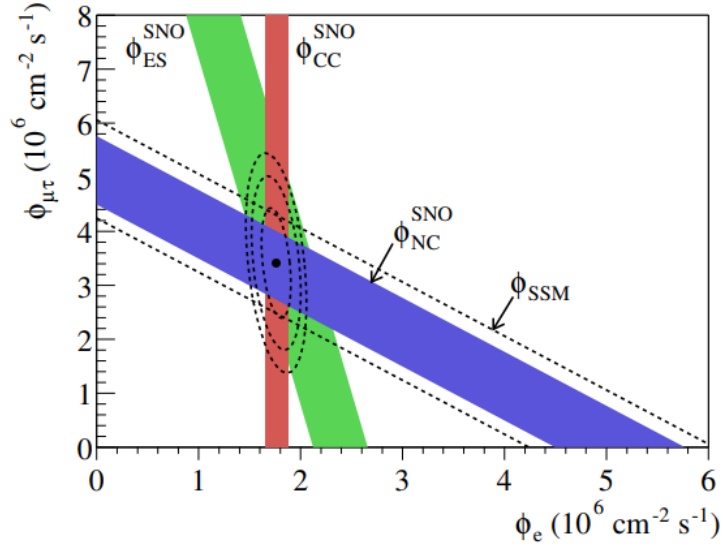
Solar neutrinos are produced by nuclear fusion in the Sun's core and are the most common type of neutrinos passing through any source observed on Earth. The vast majority of neutrinos are produced through the  $pp$  chain, a process in which four protons are combined to produce two protons, two neutrons, two positrons, and two electron neutrinos, but their energy is so low ( $< 0.425$  MeV) [13] that they are very difficult to detect [39]. The electron capture of  ${}^7\text{Be}$  produces neutrinos at either roughly 0.862 MeV ( $\sim 90\%$ ) or 0.384 MeV ( $\sim 10\%$ ). A rare side branch of the  $pp$  chain produces the  ${}^8\text{B}$  neutrinos with a maximum energy of roughly 15 MeV, and these are the easiest neutrinos to detect [39]. A very rare interaction in the  $pp$  chain produces the “ $hep$ ” neutrinos, the highest energy neutrinos (up to 18 MeV) predicted to be produced by the Sun [39]. Neutrinos are also produced by the CNO cycle, but that process



**Fig. 8:** Allowed regions of  $(\Delta m_{32}^2, \sin^2 \theta_{23})$  at 90% C.L. from various experiments (figure taken from Ref. [38]).

is considerably less important in the Sun than in other stars [39].

The timeline of solar neutrinos and their discovery dates back to the 1960s, beginning with the two astrophysicists John N. Bahcall and Raymond Davis Jr. The experiment, known as the Homestake experiment, aimed to count the solar neutrinos arriving on Earth. Using the standard solar model (SSM), Bahcall was able to calculate the number of neutrinos arriving on Earth from the Sun [40]. At the same time, Davis had proposed an idea to detect solar neutrinos by using a radioactive chemical process:  $\nu_e + {}^{37}\text{Cl} \rightarrow {}^{37}\text{Ar} + e^-$  [41]. By conducting the experiment deep underground, they were able to avoid cosmic ray interactions which could affect the process and results. As a surprise, the experimental value of observed solar neutrinos was less than 20% of the theoretical prediction calculated by Bahcall [14]. It was unknown at the time whether there were errors in the experiment or calculations, or whether Bahcall and Davis did not explain all variables, but this discrepancy gave birth to what became known as the solar neutrino problem. Later, the deficits of the solar neutrino flux were observed in some other experiments such as GALLEX [42], SAGE [43] and SK [44], with numbers ranging from one half to two thirds. One year after the discovery of neutrino oscillation at SK, the Sudbury Neutrino Observatory (SNO) started collecting data [45]. That experiment aimed at observing the  ${}^8\text{B}$  solar neutrinos with around 10 MeV energy, and was designed to employ a large quantity of heavy water as the detection medium, which could make it possible to observe both the electron neutrinos produced in the core of the Sun and all flavors of neutrinos. The experiment was able to observe two separate reactions on deuteron, a CC reaction that was sensitive only to  $\nu_e$  and a NC that was equally sensitive to all flavors. Also, SNO could observe neutrinos of all flavors via the elastic scattering (ES) of electrons by neutrinos. The neutrino flux in each



**Fig. 9:** Flux of  ${}^8\text{B}$  solar neutrinos from SNO, measured via CC, NC and ES interactions. The axes represent the inferred fluxes of  $\nu_\mu + \nu_\tau$  and  $\nu_e$ . The sensitivity to NC and ES interactions give the slopes of the bands. The solar neutrino flux predicted by the SSM [49] is indicated as  $\phi_{\text{SSM}}$  (dashed line). The intercepts of these bands with the axes represent the  $\pm 1\sigma$  errors. The dashed ellipses represent the best estimates of  $\phi(\nu_e)$  and  $\phi(\nu_{\mu\tau})$  at  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  C.L.

reaction is parameterized in terms of the flux of each flavor as follows [46]:

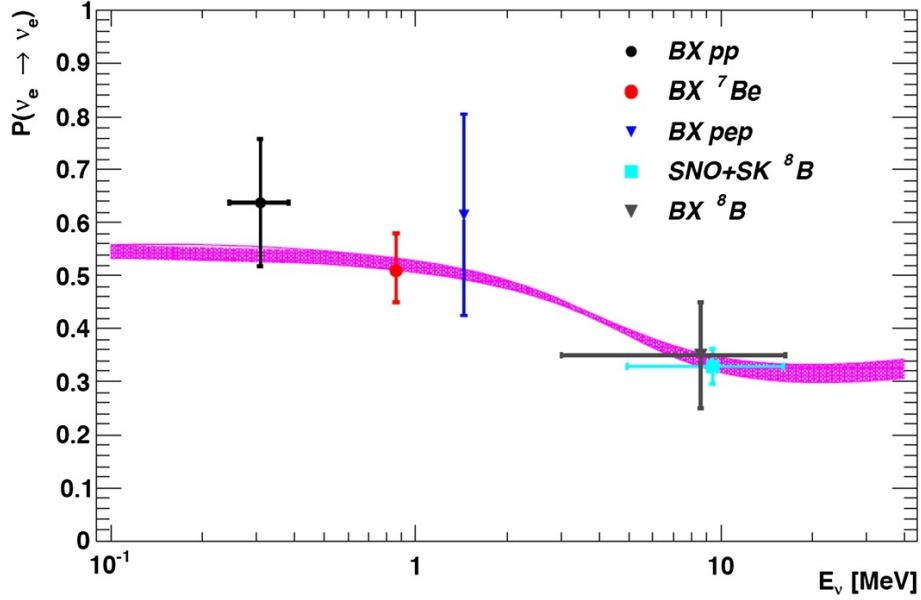
$$\begin{aligned}\phi_{CC} &= \phi(\nu_e), \\ \phi_{ES} &= \phi(\nu_e) + 0.1559 \phi(\nu_{\mu\tau}), \\ \phi_{NC} &= \phi(\nu_e) + \phi(\nu_{\mu,\tau}),\end{aligned}\tag{21}$$

where  $\phi(\nu_{\mu\tau}) = \phi(\nu_\mu) + \phi(\nu_\tau)$  and the factor of 0.1559 is the ratio of the ES cross sections for  $\nu_{\mu\tau}$  and  $\nu_e$  above  $T_{\text{eff}} = 5.0$  MeV. Making this change of variables and fitting directly for the flavor content, the null hypothesis test of no-oscillation is reduced to a test of the condition  $\phi(\nu_{\mu\tau}) = 0$ . With the measurements of the systematic uncertainties on both acceptance and detector response, the results of the flux for the constrained fit are given in units of  $10^6 \text{ cm}^{-2}\text{s}^{-1}$  by [47]

$$\begin{aligned}\phi_{CC} &= 1.76^{+0.06}_{-0.05}(\text{stat})^{+0.09}_{-0.09}(\text{syst}), \\ \phi_{ES} &= 2.39^{+0.24}_{-0.23}(\text{stat})^{+0.12}_{-0.12}(\text{syst}), \\ \phi_{NC} &= 5.09^{+0.44}_{-0.43}(\text{stat})^{+0.46}_{-0.43}(\text{syst}).\end{aligned}\tag{22}$$

The physical interpretation of the “flux” for each interaction type is that it is the equivalent flux of  ${}^8\text{B}$   $\nu_e$ ’s produced from an undistorted energy spectrum that would yield the same number of events inside the signal region from that interaction as was seen in the data. It turns out that about 1/3 of solar  $\nu_e$  survived while 2/3 are transformed into the combined  $\nu_\mu$  and  $\nu_\tau$ . The inequality of the fluxes determined from the CC, ES, and NC reactions strongly supported the evidence for the existence of non- $\nu_e$  components to the  ${}^8\text{B}$  solar neutrinos. Figure 9 shows the constraints on the  $\nu_e$  flux  $\phi(\nu_e)$  versus the combined  $\nu_\mu$  and

$\nu_\tau$  fluxes  $\phi(\nu_{\mu\tau})$ , derived from the CC, ES, and NC rates. Both measurements of the total active fluxes  $\phi_{NC}$ , as well as the sum of  $\phi(\nu_e) + \phi(\nu_{\mu\tau})$ , were in good agreement with SSM predictions, which could confirm the validity of the SSM predictions on the solar neutrino fluxes.



**Fig. 10:** Measurements of the  $\nu_e$  survival probability, obtained by Borexino, as well as SNO and SK (figure taken from Ref. [51]). The shaded regions represent the MSW-LMA predictions.

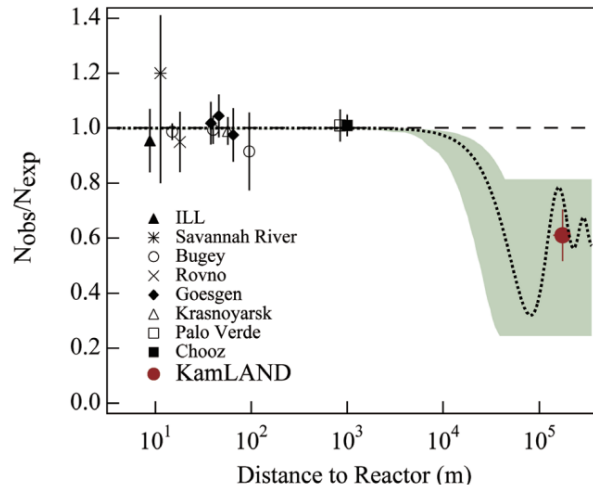
The phenomenon of solar neutrino transition is complicated by the matter effect, as neutrinos propagate outwards from their production point inside the Sun. As discussed before, the so-called “MSW” effect, proposed by Mikheyev, Smirnov and Wolfenstein, enhances oscillation in an energy-dependent fashion. As a result, neutrinos produced in different fusion reactions are affected to a different extent, due to their different energies. While the MSW effect is negligible at sub-MeV energies, where vacuum oscillation dominates, it becomes significant above about 5 MeV. The former case is relevant for the results from the gallium experiments such as GALLEX and SAGE, which were sensitive to low energy  $pp$  neutrinos, resulting in  $\bar{P}_{\nu_e \rightarrow \nu_\mu} \simeq 1 - \frac{1}{2} \sin^2 2\theta \simeq 0.6$ . The latter condition is assumed to explain the results from SK and SNO, which mostly detected  ${}^8\text{B}$  neutrinos, resulting in  $\bar{P}_{\nu_e \rightarrow \nu_\mu} \simeq \sin^2 \theta \simeq 0.32$ . A transition is predicted in between these two regimes, where the survival probability falls from the vacuum-averaged value to the additionally-suppressed matter oscillation value. It is worthwhile to note that in both the vacuum- and matter-dominated regions the survival probability is determined by the value of the mixing angle,  $\theta_{12}$ , and not by the details of the interaction of neutrinos with matter.

In 2011, the Borexino experiment [50] first confirmed the energy dependence of the oscillation probability as well as the transition between two regimes for solar neutrinos, as presented in Fig. 10, which is taken from Ref. [51].

## 4.4 Reactor based neutrino experiments

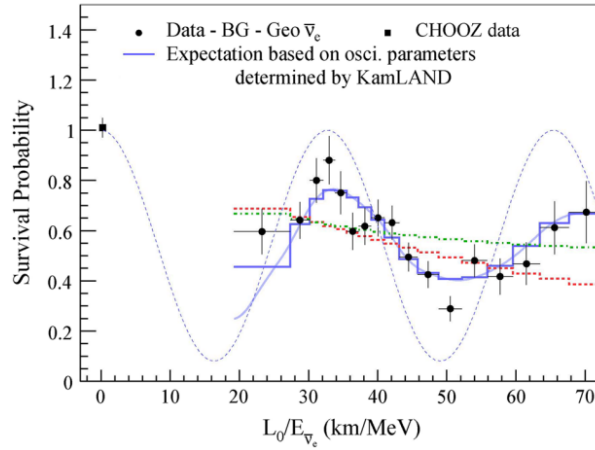
### 4.4.1 KamLAND experiment

This was the first experiment aiming to find evidence for neutrino oscillation using a terrestrial source of  $\bar{\nu}_e$  produced from nuclear reactors in Japan. KamLAND detected hundreds of  $\bar{\nu}_e$  through the inverse beta decay  $\bar{\nu}_e + p \rightarrow n + e^+$  with a  $\bar{\nu}_e$  energy threshold of 1.8 MeV, achieving an enormous improvement over previous attempts from any other detectors. The 180 km baseline, together with the emitted  $\bar{\nu}_e$  spectrum, made KamLAND sensitive to neutrino oscillation with large mixing angle (LMA) as a solution to the solar neutrino problem. Figure 11 shows the ratio of the measured to the expected flux for KamLAND, as well as for previous reactor experiments, as a function of the average distance from the source. The solid red circle corresponds to the KamLAND result obtained at a flux weighted average distance of 180 km. The shaded region indicates the range of flux predictions corresponding to the 95% C.L. LMA region from a global analysis of the solar neutrino data. Earlier measurements have seen no trace of anomaly, whereas the first data from KamLAND [52] give a lower ratio, exactly as expected by the LMA solution to the solar neutrino problem. The dotted curve, drawn with  $\sin^2 2\theta = 0.833$  and  $\Delta m^2 = 5.5 \times 10^{-5} \text{ eV}^2$ , is representative of a best-fit LMA prediction. The  $L_0/E$  distribution plotted in Fig. 12 shows the oscillatory behavior of the KamLAND data [53]. The solid (blue), dash (red) and dash-dot (green) histograms are the expectations from the best-fit oscillations, best-fit decay and best-fit decoherence, taking into account the individual time-dependent flux variations of all reactors and detector effects. It turns out that the data from KamLAND follow the oscillatory shape of reactor  $\bar{\nu}_e$ 's arising from the neutrino oscillation.

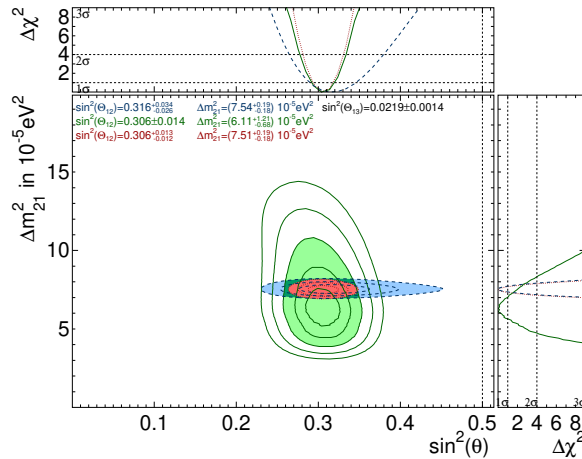


**Fig. 11:** The ratio of measured to expected  $\bar{\nu}_e$  flux for reactor experiments (figure taken from Ref. [52]).

**Compatibility of KamLAND with solar neutrino experiments:** Figure 13 presents the allowed region of parameter space  $(\Delta m_{21}^2, \sin^2 \theta_{12})$  from solar neutrino experiments (green region) and from KamLAND (blue region). The two results are compatible at the  $1.1\sigma$  level, and the tension between solar global result and KamLAND data is significantly reduced compared with the old data from solar neutrinos. Combining both results, we obtain the region in red.



**Fig. 12:** Ratio of the background- and geoneutrino-subtracted  $\bar{\nu}_e$  spectrum to the expectation for no-oscillation, as a function of  $L_0(= 180\text{km})/E$  [53]. The solid (blue), dashed (red) and dot-dashed (green) histograms are the expectations from the best-fit oscillations, best-fit decay and best-fit decoherence.



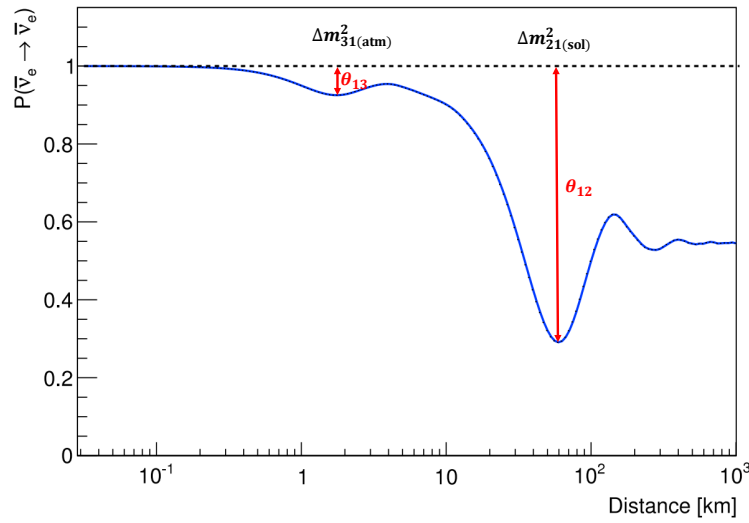
**Fig. 13:** Allowed regions of parameter space ( $\Delta m_{21}^2, \sin^2 \theta_{12}$ ) from solar neutrino experiments (green contours and green region) and from KamLAND (blue region) (figure taken from Ref. [54]). The red region is the combined result. The filled regions give the  $3\sigma$  C.L. results, the other contours shown are at the  $1$  and  $2\sigma$  C.L.

#### 4.4.2 Measuring $\theta_{13}$

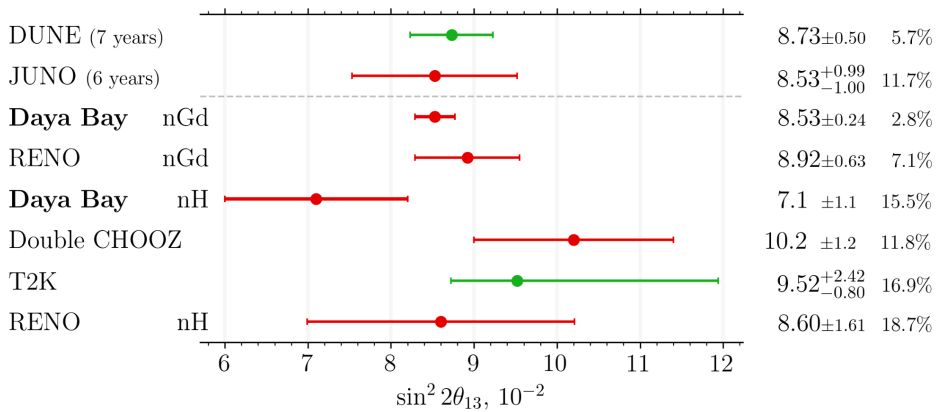
The mixing angle  $\theta_{13}$  is a key parameter to understand the underlying structure of neutrino mixing as well as to explore whether CP is violated in the lepton sector. There are lots of different ways to learn about  $\theta_{13}$ . Two of the most popular involve particle accelerators and nuclear reactors. The best measurements of  $\theta_{13}$  come from nuclear reactor experiments such as Double Chooz [55], RENO [56] and Daya Bay [57]. Detectors located near nuclear reactors provide such wonderful readings of  $\theta_{13}$

because reactors produce an extremely pure fountain of  $\bar{\nu}_e$ 's, and  $\theta_{13}$  is closely tied to how  $\nu_e$ 's mix. Reactor experiments look for the disappearance of  $\bar{\nu}_e$  in the flux from the operating fission reactors. This provides an intense source of neutrinos, in the energy range of a few MeV. The signal channel is the inverse beta decay reaction on protons. The coincidence signal from the prompt positron and the delayed neutron capture allows the unique identification of  $\bar{\nu}_e$  events. For neutrinos in this energy range we can ignore Earth's matter effect and the survival probability is quite simply given by

$$P_{\bar{\nu}_e \rightarrow \bar{\nu}_e} \simeq 1 - \sin^2 2\theta_{13} \sin^2 \frac{\Delta m_{31}^2 L}{4E} - \cos^4 \theta_{13} \sin^2 2\theta_{12} \sin^2 \frac{\Delta m_{21}^2 L}{4E}. \quad (23)$$



**Fig. 14:**  $P_{\bar{\nu}_e \rightarrow \bar{\nu}_e}$  in terms of  $L$  for short baseline ( $L = 1\text{--}2$  km) and medium baseline ( $L \sim 60$  km) reactor neutrinos.



**Fig. 15:** Comparison of the recent published results for  $\theta_{13}$  from Daya Bay, RENO, Double Chooz, and T2K, as well as the expected sensitivities of future experiments released by Daya Bay.

Figure 14 shows how  $P_{\bar{\nu}_e \rightarrow \bar{\nu}_e}$  evolves with  $L$  for short baseline and medium baseline neutrinos. On a 1–2 km baseline, the third (solar scale) term is around 1% of the second (atmospheric scale) term, so



it contributes very little uncertainty to the probability. As a result the survival probability gives direct access to  $\sin^2 2\theta_{13}$ . Unlike the LBL experiments, the measurement of  $\theta_{13}$  through the reactor experiment is theoretically clean but, by the same token, they cannot determine the MO or find CP violation. On the other hand, the third term is dominant over the second one for about 60 km baseline, such as for the JUNO experiment [58]. Based on clear deficits of  $\bar{\nu}_e$  fluxes compared with the no-oscillation predictions as well as the energy dependence of the oscillation probability, Daya Bay, RENO and Double Chooz have measured the values of  $\theta_{13}$  in 2012 [59–61], which turned out to be surprisingly large. Figure 15 presents the comparison of the recent published results for  $\theta_{13}$  from Daya Bay, RENO, Double Chooz and T2K, as well as the expected sensitivities of future experiments released by Daya Bay.

#### 4.5 Global fit to neutrino data

Three-flavor oscillation parameters from the global fit to data as of November 2022 are presented in Fig. 16 [62]. The results shown in the upper (lower) section are obtained without (with) the inclusion of the tabulated  $\chi^2$  data on atmospheric neutrinos provided by SK (SK-atm). The numbers in the 1st and 2nd column are obtained by assuming the neutrino mass spectrum to be normal ordering (NO) and inverted ordering (IO), respectively. The minimization with respect to IO provides the same results as NO, except for the  $3\sigma$  range of  $\Delta m_{3l}^2$  in the analysis without SK-atm. Note that  $\Delta m_{3l}^2 = \Delta m_{31}^2 > 0$  for NO and  $\Delta m_{3l}^2 = \Delta m_{32}^2 < 0$  for IO. Figure 17 shows the results of the  $\chi^2$  analysis for the 3-flavor oscillation [62]. The red (blue) curves are for NO (IO). The solid (dashed) lines are obtained without (with) the inclusion of SK-atm  $\chi^2$  data. For atmospheric  $\Delta m^2$ , we use  $\Delta m_{31}^2$  for NO and  $\Delta m_{32}^2$  for IO. From the results, we see that  $\theta_{13}, \theta_{12}, \Delta m_{21}^2, |\Delta m_{31}^2|$  are well-measured, whereas the MO, the octant of  $\theta_{23}$  and CP phase have yet to be determined.

### 5 Neutrino mass

Neutrinos were presumed to be massless. Despite decades of experimental efforts, the only evidence we have that neutrino masses are not zero comes from neutrino oscillation experiments. The masses of neutrinos are still a mystery. It is a fundamental and important question, “how much mass do neutrinos have?”. Measuring their masses would help point toward new physics beyond the SM. What we know from the neutrino oscillation experiments are the two square-mass differences  $\Delta m_{21}^2 \sim 7.5 \times 10^{-5} \text{ eV}^2$  and  $|\Delta m_{31}^2| \sim 2.5 \times 10^{-3} \text{ eV}^2$ . A recent combination of Planck data with Type Ia supernova luminosity distances, Baryon Acoustic Oscillation (BAO), and determinations of the growth rate parameter set the most constraining bound to date,  $\sum_i m_i < 0.12 \text{ eV}$  at 95% C.L. [63], which is based on the assumption of the  $\Lambda$  (cosmological constant) cold dark matter model of the Big Bang theory. Although such constraint from cosmic surveys is indirect, it is safe to say that all three neutrino masses are smaller than 1 eV.

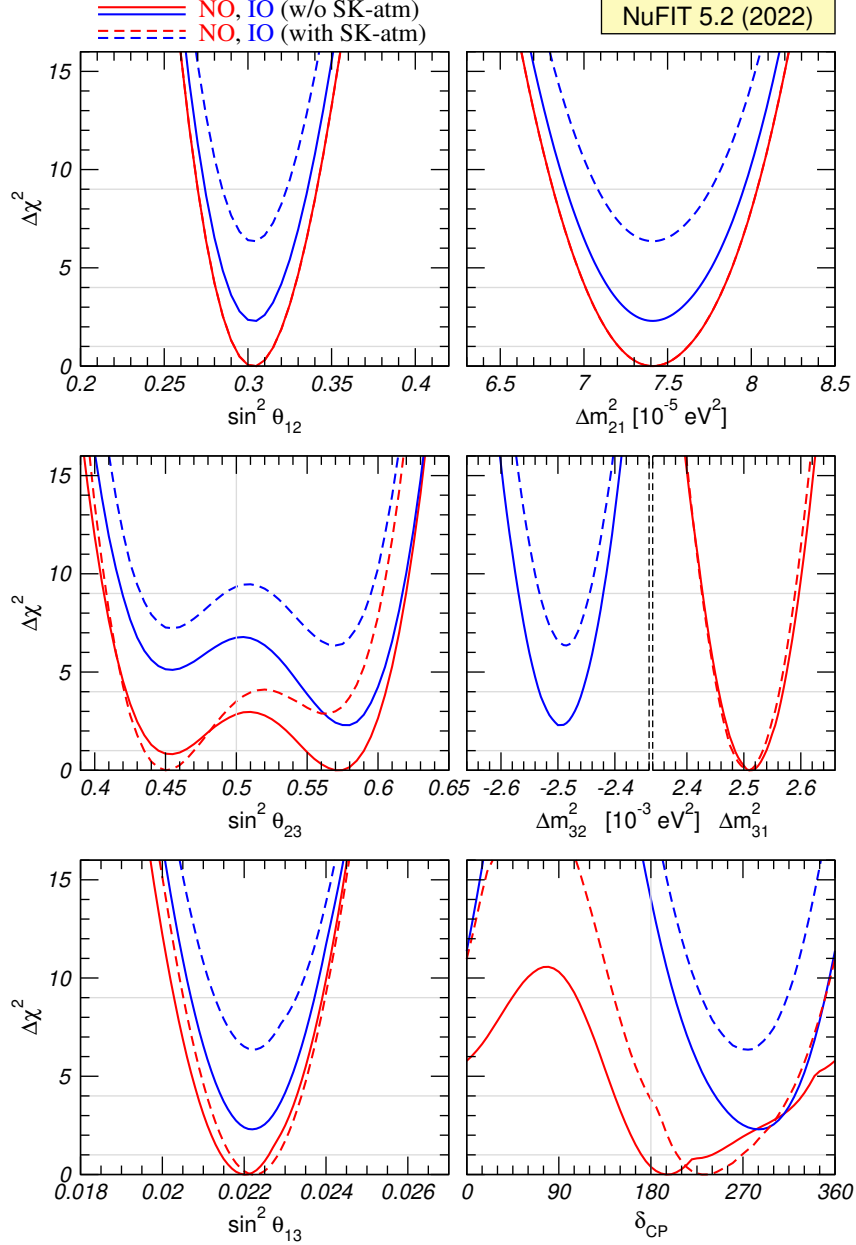
#### 5.1 Neutrino mass ordering

Since the sign of the atmospheric mass splitting ( $|\Delta m_{31}^2| \sim 2 \times 10^{-3} \text{ eV}^2$ ) remains unknown, there are two options for the neutrino MO: *normal* and *inverted*. On the other hand, since  $\Delta m_{21}^2$  turned out to be positive from the solar neutrino experiments considering the matter effect,  $\nu_2$  is always assumed to be heavier than  $\nu_1$ . In the NO (IO),  $\nu_3$  is the heaviest (lightest) neutrino. Figure 18 shows the pictorial

NuFIT 5.2 (2022)

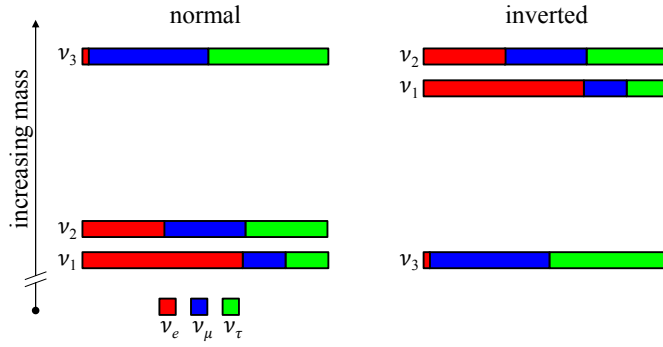
	Normal Ordering (best fit)		Inverted Ordering ( $\Delta\chi^2 = 2.3$ )		
	bfp $\pm 1\sigma$	$3\sigma$ range	bfp $\pm 1\sigma$	$3\sigma$ range	
without SK atmospheric data	$\sin^2 \theta_{12}$	$0.303^{+0.012}_{-0.011}$	$0.270 \rightarrow 0.341$	$0.303^{+0.012}_{-0.011}$	$0.270 \rightarrow 0.341$
	$\theta_{12}/^\circ$	$33.41^{+0.75}_{-0.72}$	$31.31 \rightarrow 35.74$	$33.41^{+0.75}_{-0.72}$	$31.31 \rightarrow 35.74$
	$\sin^2 \theta_{23}$	$0.572^{+0.018}_{-0.023}$	$0.406 \rightarrow 0.620$	$0.578^{+0.016}_{-0.021}$	$0.412 \rightarrow 0.623$
	$\theta_{23}/^\circ$	$49.1^{+1.0}_{-1.3}$	$39.6 \rightarrow 51.9$	$49.5^{+0.9}_{-1.2}$	$39.9 \rightarrow 52.1$
	$\sin^2 \theta_{13}$	$0.02203^{+0.00056}_{-0.00059}$	$0.02029 \rightarrow 0.02391$	$0.02219^{+0.00060}_{-0.00057}$	$0.02047 \rightarrow 0.02396$
	$\theta_{13}/^\circ$	$8.54^{+0.11}_{-0.12}$	$8.19 \rightarrow 8.89$	$8.57^{+0.12}_{-0.11}$	$8.23 \rightarrow 8.90$
	$\delta_{CP}/^\circ$	$197^{+42}_{-25}$	$108 \rightarrow 404$	$286^{+27}_{-32}$	$192 \rightarrow 360$
	$\frac{\Delta m_{21}^2}{10^{-5} \text{ eV}^2}$	$7.41^{+0.21}_{-0.20}$	$6.82 \rightarrow 8.03$	$7.41^{+0.21}_{-0.20}$	$6.82 \rightarrow 8.03$
	$\frac{\Delta m_{3\ell}^2}{10^{-3} \text{ eV}^2}$	$+2.511^{+0.028}_{-0.027}$	$+2.428 \rightarrow +2.597$	$-2.498^{+0.032}_{-0.025}$	$-2.581 \rightarrow -2.408$
	with SK atmospheric data	$\sin^2 \theta_{12}$	$0.303^{+0.012}_{-0.012}$	$0.270 \rightarrow 0.341$	$0.303^{+0.012}_{-0.011}$
$\theta_{12}/^\circ$		$33.41^{+0.75}_{-0.72}$	$31.31 \rightarrow 35.74$	$33.41^{+0.75}_{-0.72}$	$31.31 \rightarrow 35.74$
$\sin^2 \theta_{23}$		$0.451^{+0.019}_{-0.016}$	$0.408 \rightarrow 0.603$	$0.569^{+0.016}_{-0.021}$	$0.412 \rightarrow 0.613$
$\theta_{23}/^\circ$		$42.2^{+1.1}_{-0.9}$	$39.7 \rightarrow 51.0$	$49.0^{+1.0}_{-1.2}$	$39.9 \rightarrow 51.5$
$\sin^2 \theta_{13}$		$0.02225^{+0.00056}_{-0.00059}$	$0.02052 \rightarrow 0.02398$	$0.02223^{+0.00058}_{-0.00058}$	$0.02048 \rightarrow 0.02416$
$\theta_{13}/^\circ$		$8.58^{+0.11}_{-0.11}$	$8.23 \rightarrow 8.91$	$8.57^{+0.11}_{-0.11}$	$8.23 \rightarrow 8.94$
$\delta_{CP}/^\circ$		$232^{+36}_{-26}$	$144 \rightarrow 350$	$276^{+22}_{-29}$	$194 \rightarrow 344$
$\frac{\Delta m_{21}^2}{10^{-5} \text{ eV}^2}$		$7.41^{+0.21}_{-0.20}$	$6.82 \rightarrow 8.03$	$7.41^{+0.21}_{-0.20}$	$6.82 \rightarrow 8.03$
$\frac{\Delta m_{3\ell}^2}{10^{-3} \text{ eV}^2}$		$+2.507^{+0.026}_{-0.027}$	$+2.427 \rightarrow +2.590$	$-2.486^{+0.025}_{-0.028}$	$-2.570 \rightarrow -2.406$

**Fig. 16:** Three-flavor oscillation parameters from the fit to global data as of November 2022. The numbers in the 1st (2nd) column are obtained assuming NO (IO), i.e., they are relative to the respective local minimum.



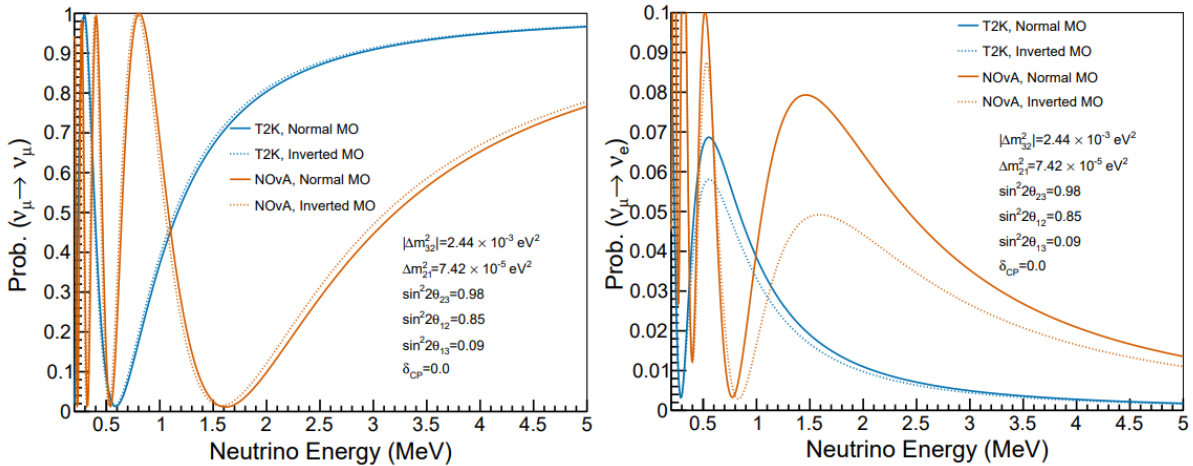
**Fig. 17:** Resulting  $\chi^2$  distributions for the global  $3\nu$  oscillation analysis of Fig. 16. The red (blue) curves are for NO (IO). The solid (dashed) lines are obtained without (with) the inclusion of SK-atm  $\chi^2$  data.

representations of the two MO possibilities [65]. While recent T2K [66], NO $\nu$ A [67] and SK [68] experiments report individually that their data favor mildly NO, a study [69] shows no favor in that indication with the combined data from the experiments. This ambiguity in determining the neutrino MO is worthy of further investigation. Measurements of  $|\Delta m_{31}^2|$  are carried on in three main chain channels: (i)  $\nu_\mu \rightarrow \nu_\mu$  (or  $\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$ ) with accelerator-based LBL (A-LBL) neutrino experiments and atmospheric neutrino experiments; (ii)  $\nu_\mu \rightarrow \nu_e$  (or  $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ ) with A-LBL neutrino experiment; and (iii)  $\bar{\nu}_e \rightarrow \bar{\nu}_e$  with the reactor-based SBL(R-SBL) neutrino experiments. The probabilities for the two cases, NO and IO, as function of the neutrino energies for the first two channels are shown in Fig. 19.



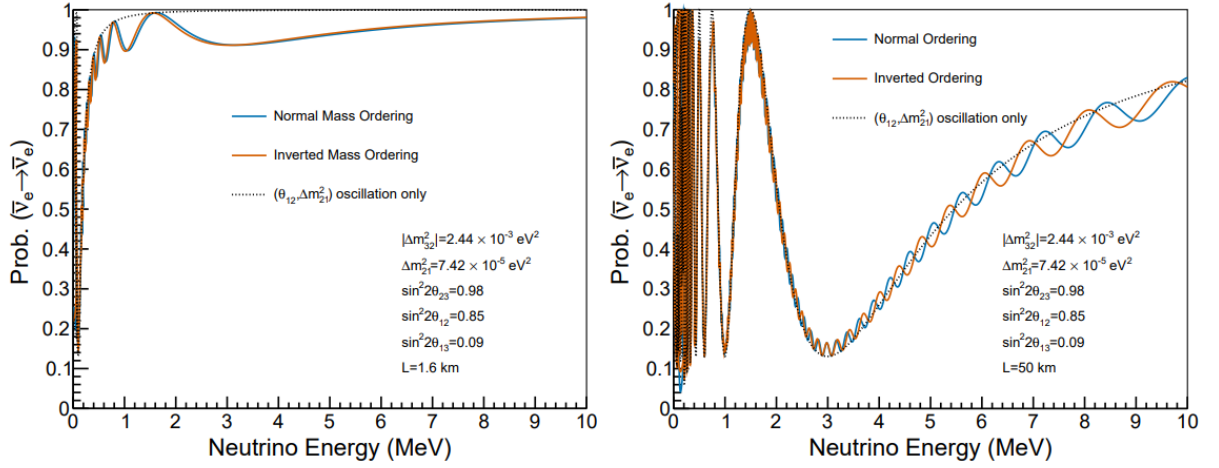
**Fig. 18:** Pictorial representations of two possible neutrino mass orderings (figure taken from Ref. [64]).

It is deduced from Fig. 19 that the sensitivity to the neutrino MO is marginal in the  $\nu_\mu \rightarrow \nu_\mu$  (or  $\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$ ) disappearance channels. The effect of the neutrino MO, on the other hand, is much stronger in the  $\nu_\mu \rightarrow \nu_e$  (or  $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ ) appearance channels. The relatively large modification of the oscillation probabilities in this channel is due to the coherent scattering of electron (anti-) neutrinos on the electrons present in the matter—the MSW effect. However, one must consider the fact that the  $\nu_\mu \rightarrow \nu_e$  (or  $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ ) appearance probability is just a few percentage, limiting the statistics of the collected data sample. Moreover, extracting the neutrino MO effect from the appearance probabilities is non-trivial since the sign of  $\Delta m_{31}^2$  is tangled severely with  $\delta_{CP}$  and the mixing angle  $\theta_{23}$  (it also depends on the mixing angle  $\theta_{13}$ , which we know with 3% precision), which have been measured with relatively large uncertainty. In addition, it is important to note that the modifications of the  $\nu_\mu \rightarrow \nu_e$  and  $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$  appearance probabilities due to matter effects are not the same.



**Fig. 19:** Neutrino oscillation probabilities for  $\nu_\mu$  (or  $\bar{\nu}_\mu$ ) disappearance (left) and  $\nu_e$  appearance (right) for the T2K and NO $\nu$ A experiments with two possible MO hypotheses (figure taken from Ref. [70]).

Due to the mutual dependence of the considered parameters in the neutrino oscillation probabilities, determining the neutrino MO will apparently enhance the sensitivity of the CP-violation search and vice



**Fig. 20:** The  $\bar{\nu}_e$  survival probabilities in the reactor-based neutrino experiments with short (left) and medium (right) experimental baselines (figure taken from Ref. [70]).

versa. As a result, the program to elucidate the neutrino MO and the search for CP violation in the A-LBL neutrino experiments are inextricably linked. So far, in the case of the R-SBL neutrino experiments, we have investigated with detectors placed relatively close to the reactor core, a few hundred meters to a few kilometers from the neutrino source. As shown on the left side of Fig. 20, the sensitivity to the neutrino MO is marginal. However, JUNO [58], with a medium-baseline of 50 km, can improve the sensitivity to the neutrino MO thanks to the interference between two oscillation terms [71], which are driven by  $\Delta m_{21}^2$  and  $\Delta m_{31}^2$ , respectively.

For JUNO, the most challenging thing to achieve is an excellent resolution of the reconstructed neutrino energy to unravel the neutrino MO effect from the detector response effect. The recent progress on the JUNO calibration [72] demonstrates that this unprecedented achievement in energy resolution is viable. In addition, it is important to note that, unlike when measuring the neutrino MO with an A-LBL experiment, the sensitivity to the neutrino MO in the reactor-based medium baseline (R-MBL) neutrino experiments is independent of the value of  $\delta_{\text{CP}}$ . Thus, resolving the neutrino MO with A-LBL and R-MBL neutrino experiments is complementary [73].

## 5.2 Dirac or Majorana masses ?

### 5.2.1 Dirac masses

For a Dirac field given by  $\psi = \begin{pmatrix} \varphi \\ \chi \end{pmatrix}$ , we can separate left-handed and right-handed fields by using projection operators,  $P_{L(R)}$ , as follows:

$$\begin{aligned} \psi_L &= P_L \psi = \frac{1}{2}(1 - \gamma^5)\psi = \begin{pmatrix} \varphi \\ 0 \end{pmatrix}, \\ \psi_R &= P_R \psi = \frac{1}{2}(1 + \gamma^5)\psi = \begin{pmatrix} 0 \\ \chi \end{pmatrix}. \end{aligned} \quad (24)$$

Similarly,  $\bar{\psi}_L = \bar{\psi}P_R, \bar{\psi}_R = \bar{\psi}P_L$ . Then a Dirac field is composed of two fields  $\psi_L$  and  $\psi_R$  as  $\psi = \psi_L + \psi_R$ . The fermion mass term is written as  $\bar{\psi}\psi = \bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R$ , and the coefficients of the terms correspond to the mass of the fermion  $\psi$ . This result shows that a fermion mass can be thought of as a  $L \leftrightarrow R$  transition. For an electron (or positron),  $e_L$  (or  $\bar{e}_R \equiv (e^c)_R$ : a short bar (-) on top stands for an antiparticle) is a component of an iso-doublet with  $I_3 = -(+)\frac{1}{2}$ , whereas  $e_R(\bar{e}_L)$  is an iso-singlet. Then, a Dirac mass of electrons (positrons) is constructed by combining both  $e_L(\bar{e}_R)$  and  $e_R(\bar{e}_L)$ . But the combination of  $e_L(e_R)$  and  $\bar{e}_R(\bar{e}_L)$  is not allowed due to violation of conservation of electric charge. In order to make neutrinos massive, a new degree of freedom with an iso-singlet is required. A very simple possibility is to postulate the existence of new Weyl fermions,  $N_R$ , with no SM quantum numbers. They do not affect SM gauge anomaly cancellations and are only modestly constrained by experiments. SM gauge singlet fermions couple to the SM only via neutrino – Higgs boson Yukawa interactions;

$$\mathcal{L}_{\nu Yuk} = y_\nu \bar{L} H N_R + \text{h.c.} , \quad (25)$$

where  $y_\nu$  is the Yukawa coupling,  $L$  is a lepton doublet and  $H$  is a SM Higgs. After electroweak symmetry is broken, the left-handed neutrino  $\nu$  in  $L$  and a singlet  $N_R$  combine into a massive neutrino with mass  $m_D = y_\nu v$ , where  $v$  is the vacuum expectation value of  $H$ , which is called a neutrino Dirac mass. This type of neutrino mass is similar to that of other charged leptons. In order to make the neutrino Dirac mass tiny, of order 0.1 eV, the Yukawa coupling  $y_\nu$  must be around  $10^{-12}$ . The size of  $y_\nu$  is very small compared with other Yukawa couplings in the SM. Such a small Yukawa coupling may serve as a hint of new physics beyond the SM. In general,  $m_D$  is a  $N \times N$  complex matrix when we consider  $N$  generation of leptons. Then,  $m_D$  is diagonalized by a bi-unitary transformation like  $m_D = U M^D V^\dagger$ , where  $U$  corresponds to the rotation matrix of left-handed neutrinos and  $V$  to that of right-handed neutrinos.

### 5.2.2 Majorana masses

In 1937, Majorana formulated a new theory of neutrinos, whereby the neutrino and the antineutrino are indistinguishable, and suggested the antineutrino-induced  $\beta$ -decay as an experimental verification of this hypothesis [74]. To understand the properties of Majorana neutrinos, let us consider some basic relations. The charge conjugation of the field  $\psi$  is defined by

$$\psi^c \equiv \bar{\psi} = C \bar{\psi}^T , \quad C = i\gamma^2 \gamma^0 , \quad (26)$$

where  $C$  is the charge conjugation operator and  $\gamma^i$  is the gamma matrix. Then, the following relations hold:

$$(\psi_R)^c = \bar{\psi}_L , \quad (\psi_L)^c = \bar{\psi}_R . \quad (27)$$

Let  $\psi_L = \begin{pmatrix} \varphi \\ 0 \end{pmatrix}$ , then its charge conjugate becomes  $(\psi_L)^c = C \gamma^0 \psi_L^* = \begin{pmatrix} 0 \\ -i\sigma^2 \varphi^* \end{pmatrix}$ . Combining both gives a fermion satisfying the relation  $\psi = \bar{\psi}$ , which is the condition for a Majorana neutrino. We need

only  $\varphi$  to describe a Majorana neutrino, as follows:

$$\psi_M = \psi_L + (\psi_L)^c = \begin{pmatrix} \varphi \\ -i\sigma^2\varphi^* \end{pmatrix} = \bar{\psi}_M. \quad (28)$$

Note that, in the same representation for a Majorana neutrino, a Dirac fermion can be written as  $\psi_D = \begin{pmatrix} \varphi \\ -i\sigma^2\chi^* \end{pmatrix}$ , ( $\varphi \neq \chi$ ). Then, if  $\varphi = \chi$ , it becomes a Majorana fermion. Using Eq. (27), a Majorana fermion can be written as

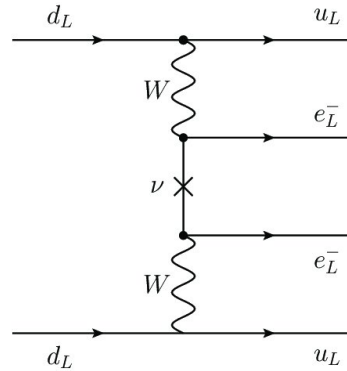
$$\psi_M = \begin{cases} \psi_L + (\psi_L)^c = \psi_L + \bar{\psi}_R \\ \psi_R + (\psi_R)^c = \psi_R + \bar{\psi}_L \end{cases}. \quad (29)$$

Then, the Majorana mass term becomes

$$\mathcal{L} = \bar{\psi}_M M \psi_M + \text{h.c.} = \bar{\psi}_L M (\psi_L)^c + \bar{\psi}_L M \psi_L + \text{h.c.} \quad (30)$$

Note that the Majorana mass matrix is symmetric, and can therefore be diagonalized by an unitary mixing matrix such as  $U^\dagger M U^* = M_{\text{diag}}$ . The Majorana mass terms are not invariant under  $\psi \rightarrow e^{i\alpha}\psi$ . So, lepton number is not conserved in the Majorana mass term.

If the nature of the neutrinos is Majorana, then they can be emitted and absorbed in the same process without showing up in the corresponding final state [75]. The neutrinoless double beta decay ( $0\nu\beta\beta$ ),  $(A, Z) \rightarrow (A, Z+2) + 2e^-$ , is a commonly proposed and experimentally pursued theoretical radioactive decay process that would prove a Majorana nature of the neutrinos, as shown in Fig. 21. It would also indicate the first ever signal of non-conservation of total lepton number. The amplitude for the decay rate depends on the effective neutrino mass, defined by  $\langle m_{\beta\beta} \rangle = \sum_i U_{ei}^2 m_i$ . Experimentally of interest and thus measured is the sum of the kinetic energies of the two emitted electrons. It should equal the total released kinetic energy of the respective nucleus for neutrinoless double beta emission. To search for neutrinoless double beta decay, there are currently a number of experiments underway and several future experiments were proposed for increased sensitivity.



**Fig. 21:** Feynman diagram for neutrinoless double beta decay due to exchange of light Majorana neutrino  $\nu$  (figure taken from Ref. [76]).

### 5.2.3 Seesaw mechanism

Once  $N_R$  is introduced to make neutrinos massive, as far as the SM is concerned, there is no symmetry to prohibit the mass term,  $M\bar{N}_R\bar{N}_L$  where  $M$  is the Majorana mass. Then, it is possible to consider the Majorana neutrino  $N = N_R + \bar{N}_L$ . Putting possible mass terms for neutrinos together, the mass terms can be written in the matrix form as follows:

$$(\overline{\nu_L}, \bar{N}_L) \begin{bmatrix} 0 & m_D \\ m_D^T & M \end{bmatrix} \begin{pmatrix} \bar{\nu}_R \\ N_R \end{pmatrix}. \quad (31)$$

Note that the mass term combining  $\nu_L$  and  $\bar{\nu}_R$  implies  $I_3 = 1, Y = -2$  and thus is forbidden by weak isospin. But if a new scalar triplet with  $I_3 = 1, Y = 2$  is introduced, the mass term can be generated. Assuming  $M \gg m_D$ , diagonalization of the mass matrix in (31) leads to two mass eigenvalues of which the lighter one is given by  $m_\nu = -m_D M^{-1} m_D^T$ . As  $M$  gets larger,  $m_\nu$  gets smaller. Its behavior looks like a seesaw, so we call it a seesaw mechanism [77–83] that successfully makes  $m_\nu$  tiny when  $M$  is large enough. There is no guide to determine the scale of  $m_D$  as well as of  $M$ . Taking  $m_D \sim 100$  GeV, we need  $M \sim 10^{15}$  GeV to achieve  $m_\nu \sim \sqrt{\Delta m_{atm}^2} \sim 0.05$  eV. Alternatively, there are so-called type-II [84–87] and type-III [88] seesaw models in which a new scalar triplet and a new fermion triplet are introduced, respectively. There are also several models that realize tiny neutrino masses via quantum effects [89,90]. The seesaw mechanisms, as well as the radiative generations of neutrino masses, provide possible answers to why light neutrino masses are so tiny. However, there is no experimental hint for them yet.

### 5.3 Determination of neutrino mass

As mentioned, we cannot determine the absolute values of neutrino masses through neutrino oscillation experiments. Then, how can we determine them experimentally? One possible way is to use decay kinematics. The simplest case is the 2-body at-rest-decay kinematics of  $\pi \rightarrow \mu\nu_\mu$ , for which one can easily obtain the relation

$$m_\nu^2 = m_\pi^2 + m_\mu^2 - \sqrt{4m_\pi^2(|\vec{p}_\mu|^2 + m_\mu^2)}. \quad (32)$$

Using the relation, we can extract the value of  $m_\nu$ . However, it is hard to do it with this method, mainly because of uncertainties in measuring  $m_\pi, m_\mu$  and  $|\vec{p}_\mu|$ . Now, the most plausible method is to extract information on the scale of neutrino mass from the endpoint spectrum of the  ${}^3\text{H} \rightarrow {}^3\text{He}^+ + e^- + \bar{\nu}_e$ , beta decay, with energy threshold  $E_0 = 18.6$  keV. The idea that the neutrino mass can be deduced in that way was already recognized by Fermi [3] in 1934, when he formulated the theory of beta decay. This method is called a ‘‘direct measurement’’ since it is model independent and does only rely on energy and momentum conservation. The number of events of the tritium beta decay depends on

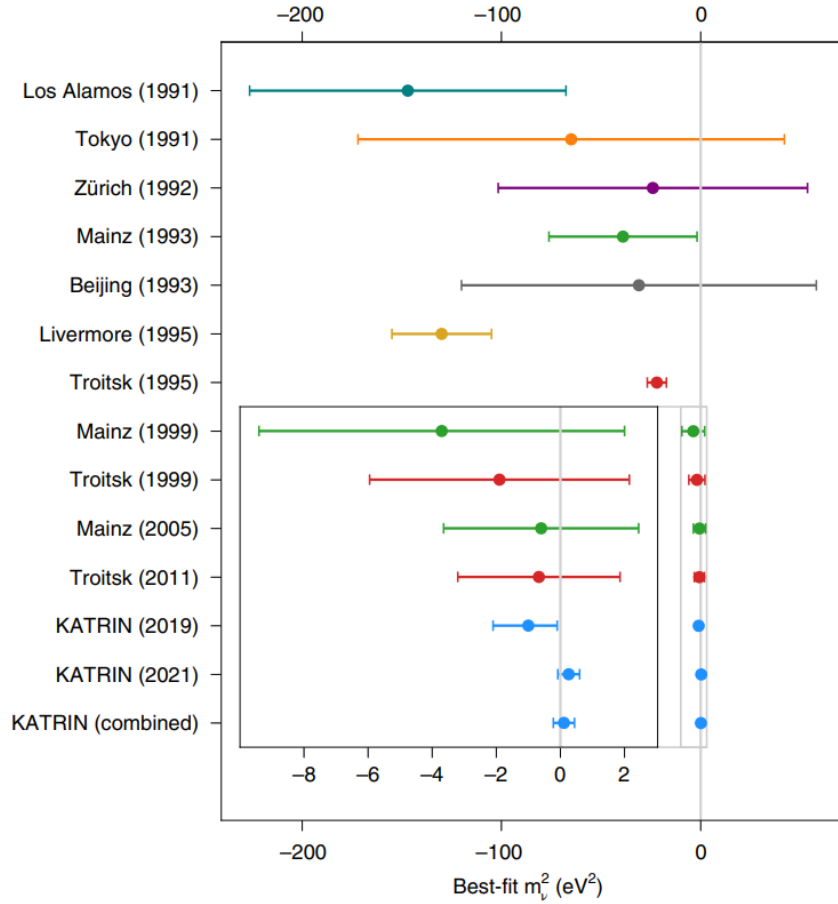
$$m^2(\nu_e) = \sum |U_{ei}|^2 m^2(\nu_i). \quad (33)$$

The endpoint spectrum of the beta decay would be shifted along with the value of  $m^2(\nu_e)$ . But, to observe modification of the endpoint spectrum, we need an eV-scale  $E$  resolution, very high luminosity,



very low background and an accurate theoretical prediction of the integral spectrum. There are a few experiments to determine the absolute scale of neutrino mass.

The KATRIN experiment [91] addresses these challenges by combining a high-activity molecular tritium source with a high-resolution spectrometer of the magnetic adiabatic collimation and electrostatic (MAC-E)-filter type16. Recently, KATRIN reported a new upper limit on the neutrino mass of 0.8 eV, by using the Lokhov–Tkachov or Feldman–Cousins technique, which is the first time that a direct neutrino mass experiment has entered the sub-eV mass range [92]. Figure 22 presents the evolution of the best fit  $m^2(\nu_e)$  results from various neutrino-mass measurements to date. Compared with other measurements, the result from the KATRIN experiment has narrowed the statistical and systematic uncertainties.



**Fig. 22:** Evolution of the  $m^2(\nu_e)$  best-fit values and total uncertainties from neutrino mass experiments [92].

## 6 CP violation in neutrino oscillations

Charge conjugation is the transformation associated with the exchange between a particle and an antiparticle, whereas parity changes a left-handed state into a right-handed state, and vice-versa. Under CP transformation, left-handed and right-handed fermion fields become, respectively,

$$(\psi_L)^{\text{CP}} = i\sigma^2\psi_L^*, \quad (\psi_R)^{\text{CP}} = -i\sigma^2\psi_R^*. \quad (34)$$

An important observable concerned with CP symmetry is the so-called Jarlskog invariant defined by  $\text{Im}[U_{\alpha k}^* U_{\beta k} U_{\alpha j} U_{\beta j}^*] = \pm J = \pm \hat{J} \sin \delta_{\text{CP}}$ , with  $\hat{J} = c_{12}s_{12}c_{23}s_{23}c_{13}^2s_{13}$  for the lepton sector. Under CP transformation, the mixing element  $U$  and  $J$  go to  $U^*$  and  $-J$ , respectively. As can be seen from Eq. (10), the oscillation probability is composed of a CP conserving part (second term) and a CP violating part (third term). Since the CP violating part vanishes for  $\alpha = \beta$ , CP violation can be explored through the appearance channels,  $\nu_\alpha \rightarrow \nu_\beta$ . Under CP transformation, the transition probability  $P_{\nu_\alpha \rightarrow \nu_\beta} (= P_{\nu_\alpha \rightarrow \nu_\beta}^{\text{CPC}} + P_{\nu_\alpha \rightarrow \nu_\beta}^{\text{CPV}})$  goes to  $P_{\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta} (= P_{\nu_\alpha \rightarrow \nu_\beta}^{\text{CPC}} - P_{\nu_\alpha \rightarrow \nu_\beta}^{\text{CPV}})$ . CP violation shows up as a difference between  $P_{\nu_\alpha \rightarrow \nu_\beta}$  and  $P_{\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta}$ , which is called the CP asymmetry,  $A_{\alpha\beta}^{\text{CP}}$ , defined by

$$A_{\alpha\beta}^{\text{CP}} \equiv \frac{P_{\nu_\alpha \rightarrow \nu_\beta} - P_{\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta}}{P_{\nu_\alpha \rightarrow \nu_\beta} + P_{\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta}} = \frac{P_{\nu_\alpha \rightarrow \nu_\beta}^{\text{CPV}}}{P_{\nu_\alpha \rightarrow \nu_\beta}^{\text{CPC}}} \quad (\alpha \neq \beta). \quad (35)$$

Since the detection of  $\nu_e$  and  $\nu_\mu$  is far easier than that of  $\nu_\tau$ , the golden channel for CP asymmetry is  $A_{\mu e}^{\text{CP}}$ . To leading order in  $\Delta m_{21}^2 L / (2E) \equiv \Delta_{21}$ , the CP asymmetry  $A_{\mu e}^{\text{CP}}$  approximately becomes

$$A_{\mu e}^{\text{CP}} \simeq \frac{4 \sin \Delta_{21} \hat{J} \sin \delta_{\text{CP}}}{\sin^2 \theta_{23} \sin^2 2\theta_{13}} \simeq \frac{c_{23} \sin 2\theta_{12}}{s_{12} s_{13}} \left( \frac{\Delta m_{21}^2}{\Delta m_{31}^2} \right) \frac{\Delta m_{31}^2 L}{4E} \sim 0.26 \left( \frac{\Delta m_{31}^2 L}{4E} \right). \quad (36)$$

We see from Eq. (36) that the asymmetry grows linearly with  $L$ , but for fixed detector size and neutrino energy  $E$ , the flux of neutrinos decreases as  $\sim 1/L^2$ . The first oscillation maximum occurs at

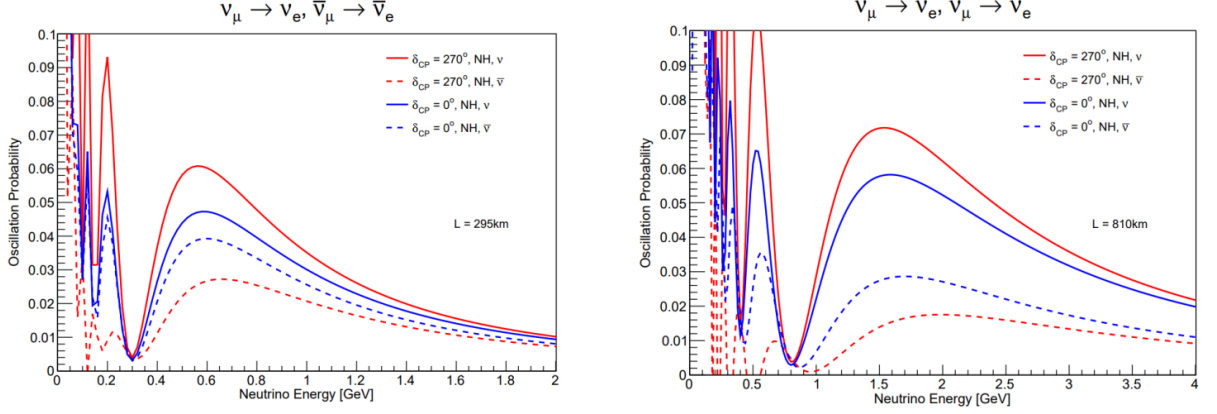
$$L_0 = \frac{2\pi E}{\Delta m_{31}^2} \simeq 495 \left( \frac{E}{\text{GeV}} \right) \left( \frac{2.5 \times 10^{-3}}{\Delta m_{31}^2} \right) \text{ km}. \quad (37)$$

For an example, the detector of the T2K experiment is located at 295 km away from the accelerator, so as to observe the first oscillation maximum for neutrinos with  $E \sim 0.6$  GeV. The observation of CP asymmetry is achievable at LBL experiments. Indeed, the measurement of CP violation can become more complicated because of the fact that the oscillation probabilities for neutrinos and anti-neutrinos are in general different in matter even if  $\delta_{\text{CP}} = 0$ . The approximated form of the transition probability in matter is given by Ref. [93]

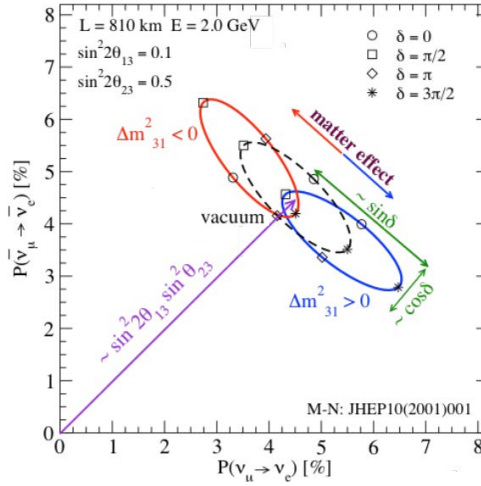
$$\begin{aligned} P_{\nu_\alpha \rightarrow \nu_\beta} &\sim \sin^2 \theta_{23} \sin^2 2\theta_{13} \sin^2 \frac{\Delta_{31}}{2} \left( 1 - \frac{8a}{\Delta m_{31}^2} \cos 2\theta_{13} \right) \\ &+ c_{13}^2 (c_{23}^2 \sin^2 2\theta_{12} + 4s_{13}^2 s_{23}^2 s_{12}^4 - 2s_{13} s_{12}^2 \sin 2\theta_{12} \sin 2\theta_{23} \cos \delta) \sin^2 \frac{\Delta_{21}}{2} \\ &+ c_{13}^2 (s_{13}^2 \sin 2\theta_{12} \sin \theta_{23} \cos \delta_{\text{CP}} - 4s_{23}^2 s_{12}^2 s_{13}^2) \sin^2 \frac{\Delta_{31}}{2} \sin \frac{\Delta_{21}}{2} \\ &+ 8\hat{J} \sin \frac{\Delta_{31}}{2} \sin \frac{\Delta_{21}}{2} \sin \frac{\Delta_{32}}{2} \sin \delta_{\text{CP}} \\ &+ 2 \cos 2\theta_{13} \sin^2 2\theta_{13} s_{23}^2 \left( \frac{aL}{4E} \right) \sin \frac{\Delta_{31}}{2} \cos \frac{\Delta_{32}}{2}, \end{aligned} \quad (38)$$

where  $a[\text{eV}^2] = 2\sqrt{2}G_F N_e E = 7.6 \times 10^{-3} \rho[\text{g}/\text{cm}^3] E[\text{GeV}]$  stands for the matter effect.  $G_F$ ,  $N_e$  and  $\rho$  represent the Fermi constant, number density of electron and matter density of Earth, respectively. Under CP transformation,  $a$  goes to  $-a$ , which mimics CP violation.

Figure 23 shows how  $P_{\nu_\mu \rightarrow \nu_e}$  and  $P_{\bar{\nu}_\mu \rightarrow \bar{\nu}_e}$  differently evolve over  $E$  for fixed  $\delta_{\text{CP}}$  and  $L$ . The left

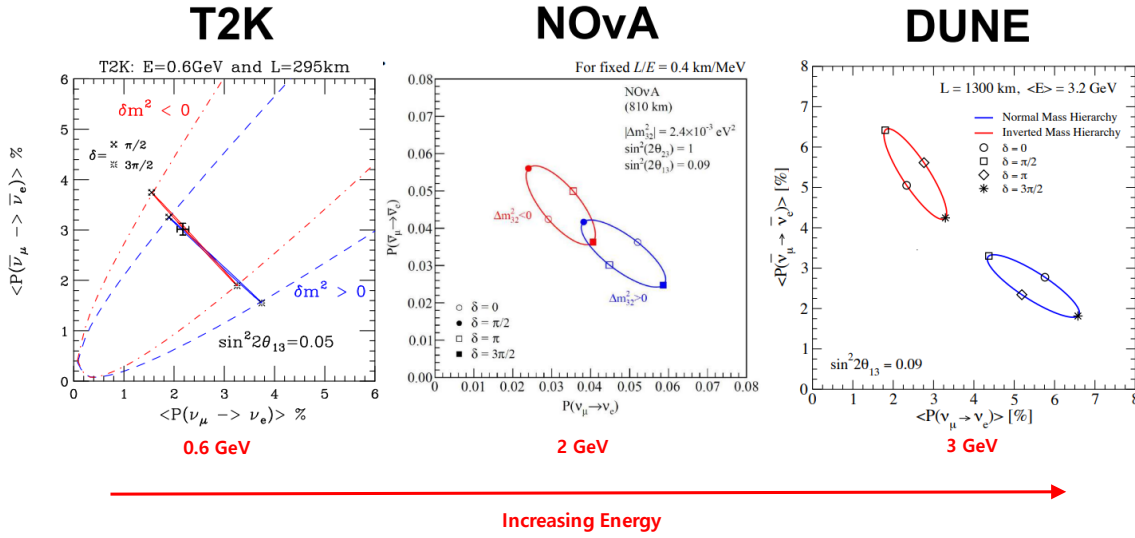


**Fig. 23:** Transition probabilities  $P_{\nu_\mu \rightarrow \nu_e}$  and  $P_{\bar{\nu}_\mu \rightarrow \bar{\nu}_e}$  as functions of neutrino energy for T2K baseline  $L = 295 \text{ km}$  (a) and NOνA baseline  $L = 810 \text{ km}$  (b) [94].



**Fig. 24:** Bi-probability diagram in the space  $P_{\nu_\mu \rightarrow \nu_e}$  vs.  $P_{\bar{\nu}_\mu \rightarrow \bar{\nu}_e}$  for the NOνA experiment.

(right) panel corresponds to the T2K (NOνA) baseline. The solid (dashed) lines correspond to the neutrino (anti-neutrino) oscillation, and the red (blue) curves to  $\delta_{\text{CP}} = 270^\circ (0^\circ)$ . Figure 24 presents a CP trajectory diagram in bi-probability space as a powerful tool for a pictorial representation of the genuine CP and matter effects in neutrino oscillations [95]. If we vary  $\delta_{\text{CP}}$  from 0 to  $2\pi$ , we can draw a closed trajectory, which becomes an ellipse, in the  $P - \bar{P}$  plane. How far the ellipse is away from the origin is proportional to  $\sin^2 \theta_{13}$ . For  $\delta_{\text{CP}} = 0$  or  $\pi$ , there is no difference between the two probabilities. Taking into account matter effects, the ellipse is shifted to two different directions according to the sign of  $\Delta m_{31}^2$ , i.e. matter effects for (+)  $\Delta m_{31}^2$  enhance  $P$ , whereas for (-) they suppress  $P$ . The magnitude of shift is larger as the matter effects, as well as the baseline, are larger. If the distance is short, the two trajectories may overlap. The octant of  $\theta_{23}$  can be distinguished using this diagram. Figure 25 shows

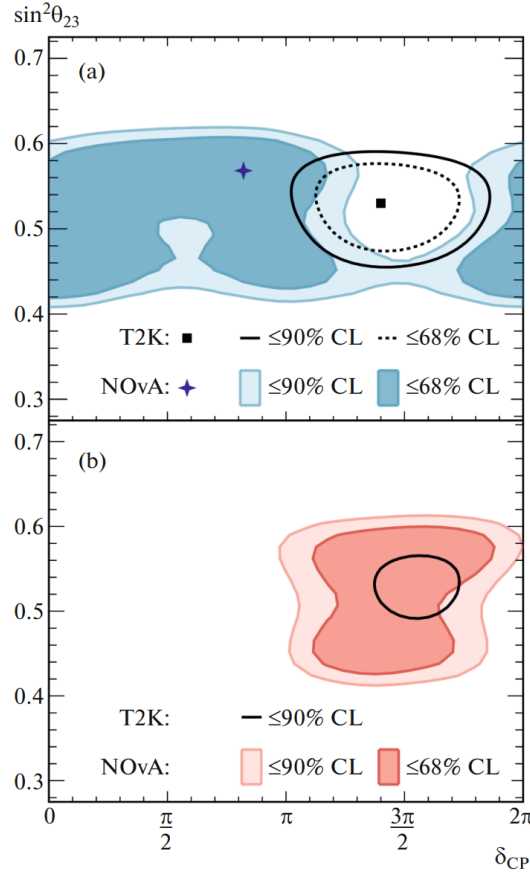


**Fig. 25:** Comparison of bi-probabilities for the T2K, NO $\nu$ A and DUNE experiments (figure taken from Ref. [96]).

how the neutrino MO and CP violation can be disentangled for experiments with different  $L$  and  $E$ . As expected, the chance to disentangle them would be increased as  $E$  and  $L$  get larger. There are some implications on CP violation in the lepton sector coming from the results of NO $\nu$ A [97] and T2K [98] experiments. Figure 26 shows the favored regions of the parameter space  $(\delta_{CP}, \sin^2 \theta_{23})$ . The upper (lower) panel corresponds to NO (IO). The colored regions are results obtained from NO $\nu$ A, whereas the regions surrounded by black contours are from T2K. We see that both results for IO are consistent, whereas there is a tension between them for NO at  $2\sigma$  level. The NO $\nu$ A and T2K experiments will continue to take data till 2026 and 2027, respectively, and then the statistics of present analyses are expected to be double. The future experiments such as DUNE and Hyper-Kamiokande will achieve the sensitivities to determine whether CP is violated in the lepton sector or not.

## 7 Conclusion

In the past few decades, a very important breakthrough in particle physics was made by the discovery of neutrino oscillations, which has shown neutrino properties beyond the SM. In this lecture, a full understanding of the various aspects of the neutrino oscillations was provided. The various experiments studying neutrinos from different sources were discussed, ranging from the pioneering ones to the experiments still in operation and to those in preparation. I have shown how the parameters concerned with neutrino oscillations were investigated and determined from the experiments. The most important milestones and the results of neutrino experiments were presented. The origin of neutrino masses and the principle behind the observed leptonic structure have been addressed. I have presented various attempts to identify the nature of neutrinos, establish the absolute values of neutrino masses, determine their or-



**Fig. 26:** Allowed regions of  $(\delta_{\text{CP}}, \sin^2 \theta_{23})$  from the T2K (black contours) and  $\text{NO}\nu\text{A}$  (colored regions) experiments [97, 98]. The upper (lower) plot corresponds to NO (IO).

dering, and measure the CP violating phase. Despite being challenging, exciting experimental programs are underway and planned for the future and will be able to address the unsolved issues.

### Acknowledgement

I sincerely thank the students and the organizers of the 2022 Asia–Europe–Pacific School of High-Energy Physics for the amazing success of the school. I was supported by the National Research Foundation of Korea (NRF) grants NRF-2019R1A2C1088953 and NRF-2020K1A3A7A09080135.

### References

- [1] Y. Fukuda *et al.* [Super-Kamiokande] *Phys. Rev. Lett.* **81** (1998) 1562–1567, [doi:10.1103/PhysRevLett.81.1562](https://doi.org/10.1103/PhysRevLett.81.1562).
- [2] J. Chadwick, *Proc. Royal Soc. A* **136** (1932) 692–708, [doi:10.1098/rspa.1932.0112](https://doi.org/10.1098/rspa.1932.0112).
- [3] E. Fermi, *Z. Phys.* **88** (1934) 161–177, [doi:10.1007/BF01351864](https://doi.org/10.1007/BF01351864).
- [4] A. Pais, *Inward bound of matter and forces in the physical world* (Oxford Univ. Press, Oxford, 1986) p. 418.
- [5] C.L. Cowan *et al.*, *Sci.* **124** (1956) 103–104, [doi:10.1126/science.124.3212.103](https://doi.org/10.1126/science.124.3212.103).

- [6] G. Danby *et al.*, *Phys. Rev. Lett.* **9** (1962) 36–44, doi:[10.1103/PhysRevLett.9.36](https://doi.org/10.1103/PhysRevLett.9.36).
- [7] M.L. Perl *et al.*, *Phys. Rev. Lett.* **35** (1975) 1489–1492, doi:[10.1103/PhysRevLett.35.1489](https://doi.org/10.1103/PhysRevLett.35.1489).
- [8] J. Jackson *et al.*, Press release: [Physicists find first direct evidence for tau neutrino at Fermilab](#) (Fermilab, Batavia, IL, 20 Jul. 2000); K. Kodama *et al.* [DONUT], *Phys. Lett. B* **504** (2001) 218–224, doi:[10.1016/S0370-2693\(01\)00307-0](https://doi.org/10.1016/S0370-2693(01)00307-0).
- [9] G. Barbiellini *et al.*, Neutrino counting, Proc. Workshop on Z Physics at LEP1, CERN, Switzerland, edited by G. Altarelli, R.H.P. Kleiss and C. Verzegnassi, CERN-89-08, vol. 1 (CERN, Geneva, 1989), pp. 129–170, doi:[10.5170/CERN-1989-008-V-1.129](https://doi.org/10.5170/CERN-1989-008-V-1.129);  
D. Decamp *et al.* [ALEPH], *Phys. Lett. B* **231** (1989) 519–529, doi:[10.1016/0370-2693\(89\)90704-1](https://doi.org/10.1016/0370-2693(89)90704-1);  
P. Aarnio *et al.* [DELPHI], *Phys. Lett. B* **231** (1989) 539–547, doi:[10.1016/0370-2693\(89\)90706-5](https://doi.org/10.1016/0370-2693(89)90706-5);  
B. Adeva *et al.* [L3], *Phys. Lett. B* **231** (1989) 509–518, doi:[10.1016/0370-2693\(89\)90703-X](https://doi.org/10.1016/0370-2693(89)90703-X);  
M.Z. Akrawy *et al.* [OPAL], *Phys. Lett. B* **231** (1989) 530–538, doi:[10.1016/0370-2693\(89\)90705-3](https://doi.org/10.1016/0370-2693(89)90705-3).
- [10] S. Eidelman *et al.*, *Phys. Lett. B* **592** (2004) 1, doi:[10.1016/j.physletb.2004.06.001](https://doi.org/10.1016/j.physletb.2004.06.001);  
<http://pdg.lbl.gov>.
- [11] B. Pontecorvo, *Zh. Eksp. Teor. Fiz.* **33** (1958) 549–551; *Sov. Phys. JETP* **6** (1957) 429–431, <https://inspirehep.net/literature/2884>.
- [12] Z. Maki, M. Nakagawa and S. Sakata, *Prog. Theor. Phys.* **28** (1962) 870–880, doi:[10.1143/PTP.28.870](https://doi.org/10.1143/PTP.28.870).
- [13] B. Pontecorvo, *Zh. Eksp. Teor. Fiz.* **53** (1967) 1717–1725; *Sov. Phys. JETP* **26** (1968) 984–988, <https://inspirehep.net/literature/51319>.
- [14] R. Davis, Jr., D.S. Harmer and K.C. Hoffman, *Phys. Rev. Lett.* **20** (1968) 1205–1209, doi:[10.1103/PhysRevLett.20.1205](https://doi.org/10.1103/PhysRevLett.20.1205).
- [15] V.N. Gribov and B. Pontecorvo, *Phys. Lett. B* **28** (1969) 493–496, doi:[10.1016/0370-2693\(69\)90525-5](https://doi.org/10.1016/0370-2693(69)90525-5).
- [16] Q.R. Ahmad *et al.* [SNO], *Phys. Rev. Lett.* **87** (2001) 071301, doi:[10.1103/PhysRevLett.87.071301](https://doi.org/10.1103/PhysRevLett.87.071301).
- [17] D. Casper *et al.*, *Phys. Rev. Lett.* **66** (1991) 2561–2564, doi:[10.1103/PhysRevLett.66.2561](https://doi.org/10.1103/PhysRevLett.66.2561).
- [18] K.S. Hirata *et al.* [Kamiokande-II], *Phys. Lett. B* **205** (1998) 416–420, doi:[10.1016/0370-2693\(88\)91690-5](https://doi.org/10.1016/0370-2693(88)91690-5).
- [19] M. Hirsch, The quest for neutrino mass, lecture at the 4th Chilean School of High Energy Physics, 13–15 Jan. 2016, [Indico](#).
- [20] L. Wolfenstein, *Phys. Rev. D* **17** (1978) 2369–2374, doi:[10.1103/PhysRevD.17.2369](https://doi.org/10.1103/PhysRevD.17.2369).
- [21] S.P. Mikheyev and A.Y. Smirnov, *Sov. J. Nucl. Phys.* **42** (1985) 913.
- [22] S.J. Parke, *Phys. Rev. Lett.* **57** (1986) 1275–1278, doi:[10.1103/PhysRevLett.57.1275](https://doi.org/10.1103/PhysRevLett.57.1275).
- [23] T.K. Gaisser and M. Honda, *Ann. Rev. Nucl. Part. Sci.* **52** (2002) 153–199, doi:[10.1146/annurev.nucl.52.050102.090645](https://doi.org/10.1146/annurev.nucl.52.050102.090645).
- [24] W.W.M. Allison *et al.*, *Phys. Lett. B* **391** (1997) 491–500, doi:[10.1016/S0370-2693\(96\)01609-7](https://doi.org/10.1016/S0370-2693(96)01609-7).

- [25] Y. Fukuda *et al.* [Kamiokande], *Phys. Lett. B* **335** (1994) 237–245, [doi:10.1016/0370-2693\(94\)91420-6](https://doi.org/10.1016/0370-2693(94)91420-6).
- [26] T. Kajita and P. Lipari, "*Comptes Rendus Physique* **6** (2005) 739–748, [doi:10.1016/j.crhy.2005.07.004](https://doi.org/10.1016/j.crhy.2005.07.004).
- [27] T. Kajita *et al.* [Super-Kamiokande], *Nucl. Phys. B* **908** (2016) 14–29, [doi:10.1016/j.nuclphysb.2016.04.017](https://doi.org/10.1016/j.nuclphysb.2016.04.017).
- [28] P. Adamson *et al.*, *Nucl. Instrum. Meth. A* **806** (2016) 279–306, [doi:10.1016/j.nima.2015.08.063](https://doi.org/10.1016/j.nima.2015.08.063).
- [29] D. Beavis *et al.* [E899], Long Baseline Neutrino Oscillation Experiment: Physics design report, BNL 52459 (Brookhaven National Lab., Upton, NY, 1995), [doi:10.2172/52878](https://doi.org/10.2172/52878).
- [30] Y. Hayato *et al.* [T2K] Letter of intent: Neutrino Oscillation Experiment at JHF (2003), <https://inspirehep.net/literature/640246>.
- [31] K. Abe *et al.* [T2K], *Phys. Rev. Lett.* **107** (2011) 041801, [doi:10.1103/PhysRevLett.107.041801](https://doi.org/10.1103/PhysRevLett.107.041801).
- [32] K. Abe *et al.* [T2K], *Phys. Rev. D* **87** (2013) 012001, [doi:10.1103/PhysRevD.87.012001](https://doi.org/10.1103/PhysRevD.87.012001).
- [33] D. Michael, *Nucl. Phys. B Proc. Suppl.* **118** (2003) 189–196, [doi:10.1016/S0920-5632\(03\)01317-3](https://doi.org/10.1016/S0920-5632(03)01317-3).
- [34] P. Adamson *et al.* [MINOS], *Phys. Rev. Lett.* **110** (2013) 251801, [doi:10.1103/PhysRevLett.110.251801](https://doi.org/10.1103/PhysRevLett.110.251801).
- [35] G. Tzanankos *et al.* [MINOS+], MINOS+: a Proposal to FNAL to run MINOS with the medium energy NuMI beam, FERMILAB-PROPOSAL-1016 (2011), <https://inspirehep.net/literature/944685>
- [36] A. Habig [NOvA], *Nucl. Phys. B Proc. Suppl.* **229–232** (2012) 460, [doi:10.1016/j.nuclphysbps.2012.09.097](https://doi.org/10.1016/j.nuclphysbps.2012.09.097).
- [37] M.A. Acero *et al.* [NOvA], *Phys. Rev. Lett.* **123** (2019) 151803, [doi:10.1103/PhysRevLett.123.151803](https://doi.org/10.1103/PhysRevLett.123.151803).
- [38] L. Berns [T2K], [arXiv:2105.06732 [hep-ex]], [10.48550/arXiv.2105.06732](https://arxiv.org/abs/2105.06732).
- [39] J.N. Bahcall, *Neutrino astrophysics* (Cambridge Univ. Press, 1989).
- [40] J.N. Bahcall, *Phys. Rev. Lett.* **12** (1964) 300–302, [doi:10.1103/PhysRevLett.12.300](https://doi.org/10.1103/PhysRevLett.12.300).
- [41] R. Davis, *Phys. Rev. Lett.* **12** (1964) 303–305, [doi:10.1103/PhysRevLett.12.303](https://doi.org/10.1103/PhysRevLett.12.303).
- [42] W. Hampel *et al.* [GALLEX], *Phys. Lett. B* **447** (1999) 127–133, [doi:10.1016/S0370-2693\(98\)01579-2](https://doi.org/10.1016/S0370-2693(98)01579-2).
- [43] J.N. Abdurashitov *et al.* [SAGE], *Phys. Rev. C* **60** (1999) 055801, [doi:10.1103/PhysRevC.60.055801](https://doi.org/10.1103/PhysRevC.60.055801).
- [44] Y. Fukuda *et al.* [Super-Kamiokande], *Phys. Rev. Lett.* **81** (1998) 1158–1162, [doi:10.1103/PhysRevLett.81.1158](https://doi.org/10.1103/PhysRevLett.81.1158), [Erratum-*ibid.* **81** (1998) 4279, [doi:10.1103/PhysRevLett.81.4279](https://doi.org/10.1103/PhysRevLett.81.4279)].
- [45] J. Boger *et al.* [SNO], *Nucl. Instrum. Meth. A* **449** (2000) 172–207, [doi:10.1016/S0168-9002\(99\)01469-2](https://doi.org/10.1016/S0168-9002(99)01469-2).
- [46] B. Aharmim *et al.* [SNO], *Phys. Rev. C* **75** (2007) 045502, [doi:10.1103/PhysRevC.75.045502](https://doi.org/10.1103/PhysRevC.75.045502).

- [47] Q.R. Ahmad *et al.* [SNO], *Phys. Rev. Lett.* **89** (2002) 011301, doi:[10.1103/PhysRevLett.89.011301](https://doi.org/10.1103/PhysRevLett.89.011301).
- [48] A. Bellerive *et al.* [SNO], *Nucl. Phys. B* **908** (2016) 30–51, doi:[10.1016/j.nuclphysb.2016.04.035](https://doi.org/10.1016/j.nuclphysb.2016.04.035).
- [49] J.N. Bahcall, M.H. Pinsonneault and S. Basu, *Astrophys. J.* **555** (2001) 990–1012, doi:[10.1086/321493](https://doi.org/10.1086/321493).
- [50] G. Bellini *et al.*, *Phys. Rev. Lett.* **107** (2011) 141302, doi:[10.1103/PhysRevLett.107.141302](https://doi.org/10.1103/PhysRevLett.107.141302).
- [51] G. Bellini, *Nucl. Phys. B* **908** (2016) 178–198, doi:[10.1016/j.nuclphysb.2016.04.011](https://doi.org/10.1016/j.nuclphysb.2016.04.011).
- [52] K. Eguchi *et al.* [KamLAND], *Phys. Rev. Lett.* **90** (2003) 021802, doi:[10.1103/PhysRevLett.90.021802](https://doi.org/10.1103/PhysRevLett.90.021802).
- [53] S. Abe *et al.* [KamLAND], *Phys. Rev. Lett.* **100** (2008) 221803, doi:[10.1103/PhysRevLett.100.221803](https://doi.org/10.1103/PhysRevLett.100.221803).
- [54] Y. Nakajima, Recent results and future prospects from Super-Kamiokande, talk given at XXIX Int. Conf. on Neutrino Physics and Astrophysics, 22 Jun. –2 Jul. 2020, doi:[10.5281/zenodo.4134680](https://doi.org/10.5281/zenodo.4134680).
- [55] F. Ardellier *et al.* [Double Chooz], [arXiv:hep-ex/0606025 [hep-ex]], doi:[10.48550/arXiv.hep-ex/0606025](https://doi.org/10.48550/arXiv.hep-ex/0606025).
- [56] J.K. Ahn *et al.* [RENO], [arXiv:1003.1391 [hep-ex]], doi:[10.48550/arXiv.1003.1391](https://doi.org/10.48550/arXiv.1003.1391).
- [57] X. Guo *et al.* [Daya Bay], [arXiv:hep-ex/0701029 [hep-ex]], doi:[10.48550/arXiv.hep-ex/0701029](https://doi.org/10.48550/arXiv.hep-ex/0701029).
- [58] Z. Djurcic *et al.* [JUNO], [arXiv:1508.07166 [physics.ins-det]], doi:[10.48550/arXiv.1508.07166](https://doi.org/10.48550/arXiv.1508.07166).
- [59] F.P. An *et al.* [Daya Bay], *Phys. Rev. Lett.* **108** (2012) 171803, doi:[10.1103/PhysRevLett.108.171803](https://doi.org/10.1103/PhysRevLett.108.171803).
- [60] J.K. Ahn *et al.* [RENO], *Phys. Rev. Lett.* **108** (2012) 191802, doi:[10.1103/PhysRevLett.108.191802](https://doi.org/10.1103/PhysRevLett.108.191802).
- [61] Y. Abe *et al.* [Double Chooz], *Phys. Rev. D* **86** (2012) 052008, doi:[10.1103/PhysRevD.86.052008](https://doi.org/10.1103/PhysRevD.86.052008).
- [62] The results in Figs. 16 and 17 are taken from <http://www.nu-fit.org>.
- [63] E. Di Valentino and A. Melchiorri, *Astrophys. J. Lett.* **931** (2022) L18, doi:[10.3847/2041-8213/ac6ef5](https://doi.org/10.3847/2041-8213/ac6ef5).
- [64] R. B. Patterson, *Ann. Rev. Nucl. Part. Sci.* **65** (2015) 177, doi:[10.1146/annurev-nucl-102014-021916](https://doi.org/10.1146/annurev-nucl-102014-021916).
- [65] P. Fernández, *SciPost Phys. Proc.* **1** (2019) 029, doi:[10.21468/SciPostPhysProc.1.029](https://doi.org/10.21468/SciPostPhysProc.1.029).
- [66] J. Lagoda [T2K], *PoS NuFact2021* (2022) 054, doi:[10.22323/1.402.0054](https://doi.org/10.22323/1.402.0054).
- [67] E. Catano-Mur [NOvA], [arXiv:2206.03542 [hep-ex]], doi:[10.48550/arXiv.2206.03542](https://doi.org/10.48550/arXiv.2206.03542).
- [68] P. Fernandez Menendez [Super-Kamiokande], *PoS ICRC2021* (2021) 008, doi:[10.22323/1.395.0008](https://doi.org/10.22323/1.395.0008).
- [69] K.J. Kelly *et al.*, *Phys. Rev. D* **103** (2021) 013004, doi:[10.1103/PhysRevD.103.013004](https://doi.org/10.1103/PhysRevD.103.013004).
- [70] S. Cao *et al.*, *Symmetry* **14** (2022) 1, doi:[10.3390/sym14010056](https://doi.org/10.3390/sym14010056).
- [71] S.T. Petcov and M. Piai, *Phys. Lett. B* **533** (2002) 94–106, doi:[10.1016/S0370-2693\(02\)01591-5](https://doi.org/10.1016/S0370-2693(02)01591-5).
- [72] A. Abusleme *et al.* [JUNO], *JHEP* **03** (2021) 004, doi:[10.1007/JHEP03\(2021\)004](https://doi.org/10.1007/JHEP03(2021)004).
- [73] S. Cao *et al.*, *Phys. Rev. D* **103** (2021) 112010, doi:[10.1103/PhysRevD.103.112010](https://doi.org/10.1103/PhysRevD.103.112010).



- [74] E. Majorana, *Nuovo Cim.* **14** (1937) 171–184, [doi:10.1007/BF02961314](https://doi.org/10.1007/BF02961314).
- [75] W. Rodejohann, *Int. J. Mod. Phys. E* **20** (2011) 1833–1930, [doi:10.1142/S0218301311020186](https://doi.org/10.1142/S0218301311020186).
- [76] J.J. Gomez-Cadenas *et al.*, *Riv. Nuovo Cim.* **35** (2012) no.2, 29–98, [doi:10.1393/ncr/i2012-10074-9](https://doi.org/10.1393/ncr/i2012-10074-9).
- [77] H. Fritzsch, M. Gell-Mann and P. Minkowski, *Phys. Lett. B* **59** (1975) 256–260, [doi:10.1016/0370-2693\(75\)90040-4](https://doi.org/10.1016/0370-2693(75)90040-4).
- [78] P. Minkowski, *Phys. Lett. B* **67** (1977) 421–428, [doi:10.1016/0370-2693\(77\)90435-X](https://doi.org/10.1016/0370-2693(77)90435-X).
- [79] T. Yanagida, *Prog. Theor. Phys.* **64** (1980) 1103–1105, [doi:10.1143/PTP.64.1103](https://doi.org/10.1143/PTP.64.1103).
- [80] M. Gell-Mann, P. Ramond and R. Slansky, Proc. Supergravity Workshop, Stony Brook, NY, United States, 27–28 Sep. 1979, edited by D.Z. Freedman and P.van Nieuwenhuizen (North-Holland, Amsterdam, 1979), pp. 315–321, <https://inspirehep.net/literature/9686>.
- [81] S.L. Glashow, The future of elementary particle physics, Proc. Cargese Summer Institute: Quarks and leptons, edited by M. Lévy *et al.* (Plenum, New York, NY, 1980), pp. 687–713, [doi:10.1007/978-1-4684-7197-7\\_15](https://doi.org/10.1007/978-1-4684-7197-7_15).
- [82] 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).
- [83] J. Schechter and J.W.F. Valle, *Phys. Rev. D* **22** (1980) 2227, [doi:10.1103/PhysRevD.22.2227](https://doi.org/10.1103/PhysRevD.22.2227).
- [84] M. Magg and C. Wetterich, *Phys. Lett. B* **94** (1980) 61–64, [doi:10.1016/0370-2693\(80\)90825-4](https://doi.org/10.1016/0370-2693(80)90825-4).
- [85] T.P. Cheng and L.F. Li, *Phys. Rev. D* **22** (1980) 2860, [doi:10.1103/PhysRevD.22.2860](https://doi.org/10.1103/PhysRevD.22.2860).
- [86] G. Lazarides, Q. Shafi and C. Wetterich, *Nucl. Phys. B* **181** (1981) 287–300, [doi:10.1016/0550-3213\(81\)90354-0](https://doi.org/10.1016/0550-3213(81)90354-0).
- [87] R.N. Mohapatra and G. Senjanovic, *Phys. Rev. D* **23** (1981) 165, [doi:10.1103/PhysRevD.23.165](https://doi.org/10.1103/PhysRevD.23.165).
- [88] R. Foot *et al.*, *Z. Phys. C* **44** (1989) 441, [doi:10.1007/BF01415558](https://doi.org/10.1007/BF01415558).
- [89] A. Zee, *Phys. Lett. B* **93**, 389 (1980), [doi:10.1016/0370-2693\(80\)90349-4](https://doi.org/10.1016/0370-2693(80)90349-4), [Erratum-*ibid.* **95**, 461 (1980), [doi:10.1016/0370-2693\(80\)90193-8](https://doi.org/10.1016/0370-2693(80)90193-8)].
- [90] Y. Cai *et al.*, *Front. in Phys.* **5** (2017) 63, [doi:10.3389/fphy.2017.00063](https://doi.org/10.3389/fphy.2017.00063).
- [91] A. Osipowicz *et al.* [KATRIN], [arXiv:hep-ex/0109033 [hep-ex]], [doi:10.48550/arXiv.hep-ex/0109033](https://doi.org/10.48550/arXiv.hep-ex/0109033).
- [92] M. Aker *et al.* [KATRIN], *Nature Phys.* **18** (2022) 160, [doi:10.1038/s41567-021-01463-1](https://doi.org/10.1038/s41567-021-01463-1).
- [93] H. Nunokawa, S.J. Parke and J.W.F. Valle, *Prog. Part. Nucl. Phys.* **60** (2008) 338–402, [doi:10.1016/j.pnpnp.2007.10.001](https://doi.org/10.1016/j.pnpnp.2007.10.001).
- [94] T. van Ngoc, C.V. Son and N.T.H. Van, *Commun. Phys.* **28** (2018) 337, [doi:10.15625/0868-3166/28/4/12930](https://doi.org/10.15625/0868-3166/28/4/12930).
- [95] H. Minakata and H. Nunokawa, *JHEP* **10** (2001) 001, [doi:10.1088/1126-6708/2001/10/001](https://doi.org/10.1088/1126-6708/2001/10/001).
- [96] J. Hartnell, Status and prospects of neutrino mass hierarchy and CP-violation measurements, talk given at Symmetries in Subatomic Physics, Aachen, 11 Jun. 2018, [Indico](https://indico.in2p3.fr/event/12000/10000).
- [97] A. Himmel, New oscillation results from the NO $\nu$ A experiment, talk given at XXIX Int. Conf. on Neutrino Physics and Astrophysics, 22 Jun. –2 Jul. 2020, [doi:10.5281/zenodo.4142045](https://doi.org/10.5281/zenodo.4142045).

- [98] P. Dunne, Latest neutrino oscillations from T2K, talk given at XXIX Int. Conf. on Neutrino Physics and Astrophysics, 22 Jun. –2 Jul. 2020, <https://doi.org/10.5281/zenodo.4154355>.

# QCD

*Vajravelu Ravindran*

The Institute of Mathematical Sciences, C.I.T Campus, Chennai, India

---

The structure of strong interaction dynamics, namely Quantum Chromodynamics, is discussed within the framework of perturbation theory. After a brief introduction to the historical developments, we will discuss in detail the role of perturbative QCD to understand the physics at various high energy colliders involving hadrons. We will discuss how certain large threshold logarithms that show up at every order in perturbation theory can be resummed to all orders. We will also discuss some of the recent advances in performing higher order perturbative corrections.

---

## 1 Introduction

The question of what constitutes the visible matter around us has been there for several centuries. There have been different answers at different times. Thanks to several experiments and theoretical studies, we could probe distances that are of the order of less than a few fermi to unravel what constitutes the matter and the dynamics that govern them. The Standard Model (SM) of particle physics is extremely successful in explaining electromagnetic, weak and strong forces within a single framework. However, there are several phenomena that we do not have explanations for within the SM and the efforts to understand them in a single framework are still going on. The Large hadron collider (LHC) at Geneva is designed not only to test the SM to unprecedented accuracy but also probe physics beyond the SM. At LHC, high energetic protons are collided to shed light on the dynamics at the smallest length scale possible.

Among the four forces, the force that binds the nucleons inside the nucleus of every atom is the strongest. In addition, it is blind to electric charges of the nucleons. Since the colliding particles at the LHC are hadrons, the strong interaction plays an important role. Hence, it is important to understand this force at short distances. In this article, we discuss how to apply QCD to test SM at high energies.

In order to set stage, we will give brief introduction to Quark Model that describes the structure of hadrons in terms of its constituents called quarks and anti quarks. Then, we present how observables at hadron colliders can be expressible in terms of these constituents. We exploit QCD factorisation of short and large distance physics at high energies to study these observables. We show that the key property of QCD, namely asymptotic freedom allows us to compute short distance part reliably. Often the fixed order perturbative predictions are affected by large logarithms resulting from soft gluons in the threshold region. We show how resummation of such logarithms to all orders can be done to make sensible predictions. Precision measurements at the LHC demand precise predictions from QCD. The later is difficult to obtain due to presence of large number of Feynman diagrams and the corresponding multi loop and phase space integrals. We present some of the modern technique to deal with these quantities.

---

This article should be cited as: QCD, Vajravelu Ravindran, DOI: [10.23730/CYRSP-2024-001.85](https://doi.org/10.23730/CYRSP-2024-001.85), in: Proceedings of the 2022 Asia–Europe–Pacific School of High-Energy Physics, CERN Yellow Reports: School Proceedings, CERN-2024-001, DOI: [10.23730/CYRSP-2023-001](https://doi.org/10.23730/CYRSP-2023-001), p. 85. © CERN, 2024. Published by CERN under the [Creative Commons Attribution 4.0 license](https://creativecommons.org/licenses/by/4.0/).

## 2 Quark model

In the 1950s [1], large numbers of strongly interacting particles, called hadrons, were discovered. It was a challenging task not only to classify them in a systematic fashion but also to look for constituents that act as building blocks of these composite objects. The hadrons can be classified in two classes based on their spin quantum numbers. Those with half integral spins are called baryons and the ones with integral spins are called mesons. Hadrons such as the proton, neutron,  $\Lambda$  and  $\Sigma$  are baryons, while  $\pi^\pm$ ,  $\pi^0$ ,  $K^\pm$  and  $K^0$  are mesons.

Symmetry transformations play an important role in physics and strong interaction is no exception. One finds that the interchange of protons with neutrons does not affect the strong interaction dynamics. Hence, it was proposed that these two hadrons can form a basis for isospin transformations, in particular the isospin doublets of the continuous group  $SU(2)$  and the strong interaction is said to be invariant under  $SU(2)$  transformations. The symmetry group  $SU(2)$  allows to accommodate other hadrons. For example, charged as well as neutral pions,  $\pi^\pm$  and  $\pi^0$ , form an isotriplet of the same  $SU(2)$ . Similarly the  $\Sigma$  baryons,  $\Sigma^\pm$  and  $\Sigma^0$ , form an isotriplet. Hadrons such as  $\eta$ ,  $\omega$ ,  $\Lambda$ , etc. are simply isosinglets. Among the particles,  $K$  mesons and  $\Lambda$  baryons showed a peculiar behaviour, namely, they were produced in large numbers through the strong interaction, while their life times were longer and their decays proceeded through weak interactions. This led to the introduction of a new quantum number called ‘strangeness’, akin to the electric charge. One finds that the strong interaction preserves strangeness, while the weak interaction violates it. The symmetry transformation was found to be a  $U(1)$  symmetry and the conserved charge is called hypercharge  $Y$ , where  $Y$  is sum of baryon number and strangeness. Remarkably, the electric charge of every hadron that was observed was found to satisfy a relation, called the Gell-Mann–Nishijima relation,  $Q = T_3 + Y/2$ , where  $T_3$  is the generator of  $SU(2)$ . In summary, both isospin  $SU(2)$  and strangeness  $U(1)$  symmetries enormously simplified the classification of baryons in terms of a pair of quantum numbers  $(T_3, Y)$ . Gell-Mann proposed a larger symmetry group  $SU(3)$ , where the symmetry group  $SU(2) \otimes U(1)$  is a subgroup, and arranged the mesons and baryons in a scheme called the eight-fold way. Using the higher-dimensional representations of  $SU(3)$ , a large number of baryons and mesons were classified and, interestingly, this led to the prediction of so far unobserved hadrons that were later discovered. While this approach was successful, the mere proliferation of the number of hadrons posed a serious challenge that was eventually resolved thanks to a proposal made by Gell-Mann and Zweig independently. It states that all known hadrons can be thought of as composite particles made of point-like spin-1/2 fermions, called quarks. The model assumes that there are three types of quarks, called up, down and strange quarks, which form a multiplet that transform in the fundamental representation of  $SU(3)$ . The anti-quarks are in the conjugate representation. According to this model, mesons are made up of a pair of quark and anti-quark, and they come in both singlet and octet representations, because  $8 \times 8 = 1 + 8$ . Similarly, the baryons are made of three types of quarks and they show up in one of the representations of  $3 \times 3 \times 3 = 1 + 8 + 8 + 10$ . While the quark model was enormously successful in explaining most of the hadronic phenomena, the existence of hadronic states having  $J^P = 3/2^+$  posed a serious problem, as it apparently violates the spin-statistics theorem. The reason is that the wave functions of these spin-3/2 hadrons remain symmetric under the interchange of their constituents, as a result of their spins being aligned and the spatial part of the wave function being in the symmetric zero-angular-momentum state. The solution to this problem was the introduction of a new quantum number,

called color, for each of the quark states. An additional  $SU_c(3)$  symmetry group corresponding to the color quantum number was introduced. Each type of quark can come in three different colors and, hence, each of them transforms in the fundamental representation of  $SU_c(3)$ . Before we study the role of this symmetry group in strong interaction physics, we will discuss two developments that played an important role in understanding the structure of the hadrons [2].

### 3 Form factors and structure functions of hadrons

Scattering experiments provide valuable information on both the structure of the particles involved and the dynamics that govern the various interactions. In the following we will discuss two distinct scattering processes involving hadrons.

#### 3.1 Form factors

In elastic electron–proton scattering one can explore the spatial structure of the proton in terms of electric and magnetic form factors. For example, one finds, for  $e^-(k) + P(P) \rightarrow e^-(k') + P(P')$ ,

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{4\omega^2 \sin^4(\theta/2)} \frac{\omega'}{\omega} \{A \cos^2(\theta/2) + B \sin^2(\theta/2)\} \quad , \quad (1)$$

where  $\alpha$  is the fine structure constant,  $\omega$  and  $\omega'$  are the energies initial and final state leptons respectively,  $A = (G_E^2 + \tau G_M^2)/(1 + \tau)$  and  $B = 2\tau G_M^2$ . Here  $G_E = \mathcal{F}_1 - \tau\kappa\mathcal{F}_2$  and  $G_M = \mathcal{F}_1 + \kappa\mathcal{F}_2$ , with  $\tau = Q^2/4M^2$ ,  $Q^2 = -q^2$ . Here  $q = k' - k$  and  $M$  is the mass of the hadron. The angle  $\theta$  is the scattering angle in the rest frame of the hadron.  $\kappa$  defines the anomalous magnetic moment of the proton, related to it by  $\mu_p = (1 + \kappa)e/2M$ . The form factors  $\mathcal{F}_i$  capture the spatial structure of the hadrons and can be extracted from experiments. The best fit to the data leads, for the proton, to  $G_E = [1 + Q^2/(0.71 \text{ GeV}^2)]^{-2}$  up to  $Q^2 = 10 \text{ GeV}^2$ . There exist several fits demonstrating the non-trivial structure of the hadrons through the  $Q^2$  dependence of the form factors.

#### 3.2 Deep inelastic scattering

When the energy transfer in the scattering is large, the scattered electrons lose a lot of energy, leading to deep inelastic scattering events [3]. This leads to breaking up the proton into pieces, which then fragment into a bunch of hadrons in the final state. The inclusive cross section, where the hadronic final states ( $X$ ) are summed over, can be used to study the structure of hadrons at very high energies. The dominant contribution arises from the scattering involving a single photon exchange and the corresponding cross section factorises into leptonic and hadronic pieces. That is,

$$d\sigma = \frac{1}{4k \cdot P} \left[ \frac{4\pi e^4}{Q^4} L_{\mu\nu}(k, q) W^{\mu\nu}(q, P) \right] \frac{d^3k'}{2\omega'(2\pi)^3} \quad , \quad (2)$$

where we have dropped the mass of the lepton. The lepton tensor is given by  $L_{\mu\nu} = 2(k_\mu k'_\nu + k_\nu k'_\mu - Q^2/2g_{\mu\nu})$ . The hadronic tensor is found to be  $W_{\mu\nu} = (8\pi M)^{-1} \sum_{X(P'), s} \langle P(P, s) | j_{\mu,em}(0) | X(P') \rangle \langle X(P') | j_{\nu,em}(0) | P(P, s) \rangle \times (2\pi)^4 \delta^{(4)}(P + q - P')$ . Here  $s$  stands for the spin of the hadron  $P$  and subscript  $em$  indicates the current is electromagnetic. Unlike the leptonic

tensor, the hadronic tensor is not calculable perturbatively, however, using Lorentz covariance and the invariance of the strong interaction under parity and time reversal transformations, it can be decomposed as  $W_{\mu\nu} = (-g_{\mu\nu} + q_\mu q_\nu / q^2)W_1(q, P) + (P_\mu - P \cdot q / q^2 q_\mu)(P_\nu - P \cdot q / q^2 q_\nu)W_2(q, P)$ , where  $W_i(q, P) = W_i(P \cdot q, Q^2, M^2)$ , with  $i = 1, 2$ , are unknown scalar functions, called structure functions. The inclusive cross section is found to be

$$\frac{d\sigma}{dQ^2 d\nu} = \frac{\pi\alpha^2}{4\omega^2 \sin^4(\theta/4)} \frac{1}{\omega\omega'} \left( W_2(\nu, Q^2) \cos^2(\theta/2) + 2W_1(\nu, Q^2) \sin^2(\theta/2) \right) . \quad (3)$$

Since we sum over all the final state hadrons, it is straightforward to show that the structure functions are related to the hadronic matrix elements of the commutator of a pair of electromagnetic currents at different space-time points. Applying the ideas of current algebra, Björken found a remarkable scaling behaviour of the structure functions when  $Q^2$  and  $P \cdot q / M = \nu$  are taken to be infinity while keeping their ratio  $x_{Bj} = Q^2 / (2M\nu)$  fixed. These limits are called Björken limits. The scaling here means that the structure functions in the Björken limit ( $\lim_{Bj}$ ) depend only on the Björken variable  $x_{Bj}$ , irrespectively of  $Q^2$ . In particular,  $W_i(\nu, Q^2)$  satisfy  $\lim_{Bj}[MW_1(\nu, Q^2)] = F_1(x_{Bj})$  and  $\lim_{Bj}[\nu W_2(\nu, Q^2)] = F_2(x_{Bj})$ . The scaling behaviour of the structure functions was confirmed in a series of deep inelastic scattering experiments.

### 3.3 Parton model

Let us try to find out the consequence of the scaling [4]. We find that if we compute the differential cross section  $d\sigma/dQ^2 d\nu$  for the elastic scattering of an electron on a point-like object, then the elastic cross section takes exactly the same form as in Eq. (3), with  $W_1(\mu, Q^2)$  replaced by  $Q^2/(4M^2)\delta(\nu - Q^2/(2M))$  and  $W_2(\nu, Q^2)$  by  $\delta(\nu - Q^2/(2M))$ . This implies that  $F_2(x_{Bj}) = \int_{x_{Bj}}^1 dz F_2(z) \frac{\nu}{z} \delta(\nu - Q^2/(Mz))$ . This can be interpreted as the incoherent sum of elastic scatterings on point-like objects, each one carrying a momentum fraction  $z$  of the parent hadron. Along this line, Feynman and Björken proposed a model to explain the observed Björken scaling of the structure functions in terms of point-like objects, called partons. It is called the parton model. The salient features of the parton model are easy to understand in the infinite-momentum frame of the target hadron. The model assumes that, for an observer at rest, a fast-moving hadron will appear like a collection of weakly interacting point-like particles, called partons. Inspired by the success of the quark model, one assumes that the partons have spin 1/2 and carry fractional charges like quarks. Due to time dilation, the time scale of the interactions among partons will be much longer than that of the scattering against the highly energetic probe: the partons look almost free. Hence, the inelastic scattering can be thought of as an incoherent sum of elastic scatterings on point-like particles, each one weighted by a probability function. Naively, we can write,

$$d\sigma(x_{Bj}, Q^2) = \sum_a \int_{x_{Bj}}^1 dz \hat{f}_{a/P}(z) d\hat{\sigma}_a \left( \frac{x_{Bj}}{z}, Q^2 \right) , \quad (4)$$

where  $\hat{f}_{a/P}(z)$  is the probability of finding a parton of type ‘ $a$ ’ with a momentum fraction  $z$  of the parent hadron (i.e., the parton momentum is  $p = zP$ ) and is called parton distribution function (PDF), while  $d\hat{\sigma}_a$  represents the elastic scattering of an electron on a parton of type  $a$ . Remarkably, the hadronic

cross section factorises into process-independent but target-dependent PDFs and target-independent but parton-dependent cross sections. Note that the partonic cross section depends on the parton-level Björken variable given by  $z_{Bj} = Q^2/2p \cdot q = x_{Bj}/z$ . Using this parton model, one finds  $2x_{Bj}F_1(x_{Bj}) = F_2(x_{Bj}) = \sum_i e_i^2 x_{Bj} f_i(x_{Bj})$ , where  $e_i$  is the electric charge of the parton of type  $i$ . Measurements of these structure functions using charged- as well as neutral-current probes can be used to extract PDFs of different types for a wide range of  $x_{Bj}$ . Note that the PDFs contain the long-distance physics and hence they are not computable within the framework of perturbation theory. The PDFs satisfy various sum rules: for example, the energy momentum conservation leads to the momentum sum rule given by  $\sum_i \int_0^1 dz z f_i(z) = 1$ . Such sum rules can be used to constrain the fits of PDFs from various experiments. The parton model can be used to study other high-energy scattering processes, involving hadrons in both the initial and final states. For example, within the parton model, the cross section for the production of a pair of leptons in proton–(anti)proton scattering experiments can be expressed in terms of two PDFs as

$$\frac{d\sigma}{dQ^2} = \sum_{ab} \int_{\tau}^1 dx_1 \hat{f}_{a/P_1}(x_1) \int_{x_1}^1 dx_2 \hat{f}_{b/P_2}(x_2) d\hat{\sigma}_{ab} \left( \frac{\tau}{x_1 x_2}, Q^2 \right) . \quad (5)$$

Here the PDFs  $f_{c/P_i}(x_i)$ , where  $c = a, b$  are the parton types, are process independent and  $d\hat{\sigma}_{ab}$  are parton-level cross sections. Semi-inclusive DIS processes, where one tags a specific hadron with momentum  $P'$  in the final state, can be described in the parton model as

$$d\sigma(x_{Bj}, z_F, Q^2) = \sum_{a,b} \int_{x_{Bj}}^1 dz_1 \hat{f}_{a/P}(z_1) \int_{z_F}^1 dz_2 \hat{D}_{P'/b}(z_2) d\hat{\sigma}_{ab} \left( \frac{x_{Bj}}{z_1}, \frac{z_F}{z_2}, Q^2 \right) . \quad (6)$$

Here,  $D_{P'/b}(z_2)$  is called the fragmentation function and describes the probability that a parton of type  $b$  fragments into a hadron  $P'$  that carries away a fraction  $z_2$  of the parent parton momentum.  $z_F = 2P' \cdot q / Q^2$  is the scaling variable corresponding to the final-state hadron. It is worth emphasising that both PDFs and fragmentation functions are process independent and, hence, can be used to predict various observables at hadron colliders. In summary, the parton model provides an elegant framework to compute a variety of observables in high-energy hadronic scattering experiments.

## 4 Quantum chromodynamics

While the parton model is enormously successful, one can not ignore the effect of strong interactions among partons. Hence, the next task was to look for a suitable quantum field theory that captures the underlying dynamics of the strong interactions among the partons [5]. Gross, Wilzcek and Politzer independently found the right gauge theory that correctly describes the interaction among the partons. It is called Quantum Chromodynamics. The underlying gauge group turned out to be  $SU_c(3)$ , where the subscript ‘c’ denotes the color quantum number. It contains quark fields,  $\psi_i(x)$ , and anti-quarks,  $\bar{\psi}_i(x)$ , that come in three different colors,  $i = 1, 2, 3$ , and transform in the fundamental representation of  $SU_c(3)$ . Indicating with  $\psi$  the column vector of components  $\psi_i$ , one finds  $\psi(x) \rightarrow \psi'(x) = U(\vec{\alpha}(x))\psi(x)$  where  $U(\vec{\alpha}(x)) = \exp(i\vec{\alpha}(x) \cdot \vec{T})$  is an element of the  $SU_c(3)$  group and  $\vec{\alpha}(x)$  is the space-time-dependent angle and  $\vec{T}$  is the short-hand notation for the eight generators  $T^a$  of  $SU_c(3)$ . Similarly, the anti-quarks transform in the conjugate representation of  $SU(3)$ . The gauge fields in this theory

are called gluons ( $A_\mu^a$ ,  $a = 1, \dots, 8$ ) and transform in the adjoint representation of  $SU_c(3)$ . Denoting  $A_\mu = A_\mu^a T^a$ , one finds that under  $SU_c(3)$   $A_\mu$  transforms as  $A_\mu \rightarrow A'_\mu = U(\vec{\alpha})(A_\mu - i/g_s \partial_\mu)U^\dagger(\vec{\alpha})$ . These transformations are gauge transformations. The gauge-invariant Lagrangian takes the following form:

$$\mathcal{L}_{\text{QCD}} = \bar{\psi}_j (i\mathcal{D}_{jk} - m\delta_{jk}) \psi_k - \frac{1}{4} F_{\mu\nu}^a F^{\mu\nu a} + \text{G.F.} \quad (7)$$

where  $\mathcal{D} = (I - ig_s T^a A_\mu^a)$  and  $F_{\mu\nu} = F_{\mu\nu}^a T^a = \partial_\mu A_\nu - \partial_\nu A_\mu - ig_s [A_\mu, A_\nu]$ . The parameter  $g_s$  in the gauge transformation as well as in the covariant derivative  $D_\mu$  is the strong coupling constant and  $m$  is the mass of the quark field. While the gauge symmetry provides the right framework to understand the interaction between quarks and anti-quarks, in order to remove the two unphysical degrees of freedom associated with gluons a gauge-fixing term (G.F.) is introduced. We choose the Lorenz gauge, i.e.,  $-1/2\xi(\partial_\mu A^{\mu a})^2$ . Within the framework of perturbation theory, this introduces additional scalar fields in the formalism. These scalar fields, unlike the standard scalar field operators, do not commute, but they anti-commute like fermions and hence they are called ghost fields. They are denoted by  $c^a$  (for ghost field operators) and  $\bar{c}^a$  (anti-ghost field operators). Since the ghost fields are introduced to deal with unphysical degrees of freedom of the gluons, they are merely a mathematical construction and hence do not correspond to any real physical particle. In the Lagrangian, the terms that are bilinear in the fields describe their propagation:

$$\mathcal{L}_{\text{K.E.}} = \bar{\psi}_i (i\not{\partial} - m)\psi_i - \frac{1}{2}(\partial_\mu A_\nu^a - \partial_\nu A_\mu^a)^2 - \frac{1}{2\xi}(\partial_\mu A^{\mu a})^2 + (\partial_\mu c^a)(\partial^\mu \bar{c}^a) \quad . \quad (8)$$

The terms that describe the interactions of quark, anti-quark and gluon fields in the Lagrangian are found to contain the following terms:

$$\mathcal{L}_{\text{int}} = g_s A_\mu^a \bar{\psi} \gamma^\mu T^a \psi - g_s f^{abc} (\partial_\mu A_\nu^a) A^{\mu b} A^{\nu c} - g_s^2 f^{eab} f^{ecd} A_\mu^a A_\nu^b A^{\mu c} A^{\nu d} \quad . \quad (9)$$

The first term describes the vectorial interaction of quarks and anti-quarks with gluons, the second and third terms describe self interactions of three and four gluons. These are present due to the non-Abelian nature of the underlying gauge group  $SU_c(3)$ . The G.F. part of the Lagrangian takes the following form in the Lorenz gauge:

$$\mathcal{L}_{\text{G.F.}} = \partial \bar{c}^a \partial^\nu c^a - g_s f^{abc} (\partial_\mu \bar{c}^a) A^{\mu b} c^c \quad . \quad (10)$$

The first term describes the propagation of ghost particles and the second term their interaction with the gauge fields.

#### 4.1 Ultraviolet renormalisation

As mentioned previously, we use perturbation theory to compute various quantities from the Lagrangian and the strong coupling constant is assumed to be small enough to treat it as an expansion parameter. We use Feynman's diagrammatic approach throughout. Feynman rules for propagators of the fields and for the interaction terms are obtained from  $\mathcal{L}_{\text{K.E.}}$  and  $\mathcal{L}_{\text{int}}$ ,  $\mathcal{L}_{\text{G.F.}}$ , respectively. The standard perturbative



techniques along with the Feynman rules,  $n$ -point Green's functions of the fields, on-shell amplitudes etc. can be computed in powers of the coupling constant  $g_s$ . Explicit calculations show that the Green's functions as well as the on-shell amplitudes diverge. These divergences come from loop integrals of virtual diagrams. When the momentum circulating in the loop approaches  $\pm\infty$ , the corresponding loop integral develops divergences. They are called ultra-violet (UV) divergences. Consider, for example, the quark propagator at one-loop level. It is easy to show that in 4 dimensions the Feynman integral that appears in the computation is proportional to a divergent integral of the form  $\int d^4k (k^2(k-p)^2)^{-1}$ , which behaves as  $\log(k^2)$  for large  $k^2$  and diverges when  $k \rightarrow \pm\infty$ . Here  $p$  is the momentum of the external quark propagator.

The presence of UV divergences in these quantities poses a serious problem beyond leading order in perturbation theory. This can be solved using the technique of renormalisation. To apply this technique, we first need to regularise the theory so that the Feynman integrals are rendered finite. While there exist several ways to regularise the theory, dimensional regularisation (DR) turns out to be the most elegant as well as convenient one. In DR, the space-time dimension is extended to  $n$  dimensions, where  $n$  is not only a continuous parameter but also a complex number. If we parametrise  $n$  as  $n = 4 + \varepsilon$ , then the integrals diverge in the limit  $n \rightarrow 4$ , or equivalently  $\varepsilon \rightarrow 0$ . In  $4 + \varepsilon$  dimensions, the mass dimensions of the fields will start depending on the variable  $\varepsilon$ . In addition, the dimensionless coupling constant develops non-zero mass dimension in the regularised theory. We use an arbitrary mass  $\mu$  to define a dimensionless coupling constant in the regularised theory: i.e., the coupling in  $n$  dimensions, denoted by  $g_{s,n}$ , can be written as  $\hat{g}_s/\mu^{\varepsilon/2}$ , where  $\hat{g}_s$  is dimensionless. Here,  $\mu$  is called regularisation scale. In addition, the computations involving contraction of Lorentz indices as well as traces of Dirac matrices are performed in  $n$  dimensions.

Having regularised the theory, the next step is to renormalise the fields, coupling and the mass by performing a set of transformations involving re-scaling, as explained in the following. We first denote quark, anti-quark, gluon and ghost fields in the regularised theory collectively by  $\{\Phi_\alpha\}$ . Then, the renormalised fields  $\{\Phi_{\alpha,R}\}$  are defined by  $\Phi_\alpha = Z_{\Phi_\alpha}^{\frac{1}{2}}(\mu_R^2) \Phi_{\alpha,R}(\mu_R^2)$ . Similarly, the coupling constant is renormalised as  $g_s/\mu^{\frac{\varepsilon}{2}} S_\varepsilon^{\frac{1}{2}} = Z_{a_s}^{\frac{1}{2}}(\mu_R^2) g_{s,R}(\mu_R^2)/\mu_R^{\frac{\varepsilon}{2}}$  and the mass is renormalised by  $Z_m$ . The  $Z$ 's are called renormalisation constants. The constant  $S_\varepsilon = \exp(\varepsilon/2(\gamma_E - \log(4\pi)))$ , with  $\gamma_E$  is Euler–Mascheroni constant. The re-scaling can be interpreted as a factorisation of the unregularised quantity into the renormalisation constant and the renormalised quantity. Hence, renormalisation is multiplicative in nature.

Note that there is an arbitrariness in defining the renormalisation constants. The scale  $\mu_R$  is introduced to quantify the arbitrariness. In other words, the choice of  $\mu_R$  uniquely defines the divergent part of the renormalised term. It is called renormalisation scale. Like the coupling constant, the mass  $m$  can also be expressed in terms  $m_R$  using the mass renormalisation constant.

Finally, one can substitute these relations in the Lagrangian to write it in terms of renormalised fields, coupling and mass and their renormalisation constants. One can easily show that the Lagrangian for  $4 + \varepsilon$  dimensions takes the following form:

$$\mathcal{L}_{\text{QCD}}(\Phi_\alpha, g_{s,n}, m_n, \mu, \varepsilon) = \mathcal{L}_R(\Phi_{\alpha,R}, g_{s,R}, m_R, \mu_R, \varepsilon) + \mathcal{L}_{\text{CT}}(\Phi_{\alpha,R}, g_{s,R}, m_R, \{Z\}, \mu_R) \quad , \quad (11)$$

where the first term on the right hand side is identical to the original Lagrangian except that it contains only the renormalised fields, coupling and mass. The second term depends, in addition, on the renormalisation constants and is called counter term (CT).

Given this Lagrangian in terms of renormalised quantities and renormalisation constants, our next task is to obtain finite  $n$ -point Green's functions and on-shell amplitudes in the limit  $\varepsilon \rightarrow 0$ . This is possible provided that the renormalisation constants can be adjusted in such a way that they contain all the UV divergences at every order in perturbation theory, without any introduction of new renormalisation constants. These divergences show up as poles in  $\varepsilon$  in the limit  $n \rightarrow 4$ . While the renormalisation scale parameterises the scale at which divergences are factored into the renormalisation constant, the arbitrariness associated with the finite part of each constant remains, and can be fixed by defining a suitable scheme, called renormalisation scheme. We choose the  $\overline{MS}$  scheme, in which we combine the Euler constant  $\gamma_E$  and  $S_\varepsilon$  with the poles in the renormalisation constants.

Given any scheme, the renormalisation scale plays an important role in the understanding of the underlying dynamics of the quantum field theory. The unrenormalised fields, coupling and mass or, equivalently, the corresponding Green's functions or on-shell amplitudes do not depend on the renormalisation scale. For example, the  $n$ -point unrenormalised Green's function  $\langle 0|T \prod_{i=1}^n \Phi_i(p_i)|0\rangle$ , where  $T$  is the time-ordering operator, satisfies  $\mu_R^2 d/d\mu_R^2 \langle 0|T \prod_{i=1}^n \Phi_i(p_i)|0\rangle = 0$ , which leads to

$$\mu_R^2 \frac{d}{d\mu_R^2} \ln \langle 0|T \prod_{i=1}^n \Phi_i(p_i)|0\rangle = -\mu_R^2 \frac{d}{d\mu_R^2} \ln Z_\Phi(a_s(\mu_R^2), \varepsilon) = \Gamma_\Phi(a_s(\mu_R^2)) \quad , \quad (12)$$

where  $a_s = g_s^2(\mu_R^2)/16\pi^2$  and  $\Gamma_\Phi(a_s(\mu_R^2))$  is called the anomalous dimension. The above equation is called the renormalisation group equation (RGE). The solution to the RGE provides a relation among the Green's functions at different scales. In addition, using RGE and the fixed-order results for  $\Gamma_\Phi$  and  $a_s(\mu_R^2)$  one can systematically sum up the large logarithms of physical scale(s) to all orders in perturbation theory. In the following we show how the coupling constant can be renormalised in the  $\overline{MS}$  scheme and its consequences. The unrenormalised  $\hat{a}_s = g_s^2/16\pi^2$  is related to renormalised  $a_s(\mu_R^2) = g_s^2(\mu_R^2)/16\pi^2$  through  $\hat{a}_s = Z_{a_s}(a_s(\mu_R^2), \varepsilon)(\mu^2/\mu_R^2)^{\varepsilon/2} S_\varepsilon^{-1} a_s(\mu_R^2)$ . The fact that  $\mu_R^2(d/d\mu_R^2)\hat{a}_s = 0$  leads to

$$\mu_R^2 \frac{d}{d\mu_R^2} a_s(\mu_R^2) = \beta(a_s(\mu_R^2), \varepsilon) = -a_s(\mu_R^2) \mu_R^2 \frac{d}{d\mu_R^2} \log Z_{a_s}(a_s(\mu_R^2), \varepsilon) - \frac{\varepsilon}{2} a_s(\mu_R^2) \quad . \quad (13)$$

With the knowledge of the renormalisation constant  $Z_{a_s}$ , we can determine  $\beta(a_s(\mu_R^2), \varepsilon)$  order by order in  $a_s$ . In the limit  $\varepsilon \rightarrow 0$ , we find that  $\beta_0(a_s) = -\beta_0 a_s^2 - \beta_1 a_s^3 + \dots$ . To leading order, we have  $\mu_R^2(d/d\mu_R^2)a_s(\mu_R^2) = -\beta_0 a_s^2(\mu_R^2) + \mathcal{O}(a_s^3)$ . The solution to the RGE for  $a_s$  is found to be

$$a_s(\mu_f^2) = \frac{a_s(\mu_i^2)}{1 + a_s(\mu_i^2) \beta_0 \log \left( \frac{\mu_f^2}{\mu_i^2} \right)} + \mathcal{O}(a_s^2) \quad , \quad (14)$$

where  $\mu_i$  and  $\mu_f$  are initial and final scales. The constant  $\beta_0$  in QCD is found out to be  $\beta_0 = 11/3 C_A - 2/3 n_f$ , where  $C_A$  is the Casimir of SU(N). For QCD,  $C_A = 3$ .  $n_f$  is the number flavours of quarks. Because  $\beta_0 > 0$ , one sees that the coupling constant falls off as the scale increases. This implies that

at high energies quarks (anti-quarks), gluons and ghost particles will behave like free particles. This phenomenon is called asymptotic freedom. This property allows one to use perturbative techniques to study strong interaction at high energies. At low energies, we find that particles confine to form hadrons through the phenomenon called confinement.

## 4.2 QCD improved parton model

Our next task is to apply perturbative QCD (pQCD) to study high-energy scattering processes involving hadrons. We consider the DIS process discussed already using the parton model. Recall that in the parton model the cross section can be expressed in terms of two structure functions  $F_i$ , with  $i = 1, 2$ , which are related to tree-level scattering cross sections of leptons off the quarks as well as the anti-quarks, convoluted with the PDFs of the respective quark or anti-quark. This is called the leading-order prediction. This result will get modified if the strong interaction dynamics are included. In pQCD, the parton-level cross section  $\hat{\sigma}_a(\varepsilon)$  admits the expansion in powers of the strong coupling constant  $a_s$ :  $\hat{\sigma}_a(\varepsilon) = \sum_{i=0}^{\infty} a_s^i(\mu_R^2) \hat{\sigma}_a^{(i)}(a_s(\mu_R^2), \varepsilon)$ . Here the subscript  $a$  can be  $q$ ,  $\bar{q}$ , or  $g$ . The perturbative corrections resulting from the order- $a_s$  term are called next-to-leading-order contributions; similarly, those from the  $a_s^2$  term are called next-to-next-to-LO (NNLO) corrections, and so on. At  $a_s$  order we encounter scattering processes, such as one-loop corrections to  $e^- + q(p) \rightarrow e^- + qf(p')$ , and real emission processes, namely  $e^- + q(p)(\bar{q}(p)) \rightarrow e^- + q(p')(\bar{q}(p')) + g(k)$  and  $e^- + g(p) \rightarrow q(p_1) + \bar{q}(p_2)$ . These contributions are UV finite. However, they are sensitive to infrared (IR) divergences in 4 space-time dimensions. There are two types of IR divergences: soft and collinear. The soft divergences show up in both real and virtual diagrams when the momentum of the gluon vanishes, i.e.,  $k_i^\mu \rightarrow 0$ . Collinear divergences arise whenever two massless partons become collinear to each other.

We illustrate below the origin of soft and collinear divergences at  $a_s$  level. Let us begin with pure virtual contributions to quark/anti-quark initiated processes. We find that there will be QCD corrections to the quark–anti-quark–photon vertex as well as self-energy corrections to the quark/anti-quark legs. Let us consider, for example, the vertex correction within the dimensionally regularised set up. If we restrict ourselves to the region where the momentum of the gluon approaches zero, the leading contribution results from the integral

$$\int d^n k \frac{1}{k^2(k+p)^2(k+p')^2} \quad , \quad (15)$$

where  $k$  is the loop momentum and  $p$  and  $p'$  are the momenta of the incoming quark and the outgoing quark, respectively. Note that the above integral diverges in 4 dimensions when  $k \rightarrow 0$ . In addition, we observe that, due to the presence of the propagators  $1/(k+p)^2$  and  $1/(k+p')^2$ , we encounter collinear divergences. For example when  $k$  is parallel to  $p$  or  $p'$  and in the centre of mass frame of the quarks, in 4 dimensions, the angular part of the integral, namely  $\int_{-1}^1 d\cos(\theta) 1/(1 \pm \cos(\theta))$ , diverges. We observe that in  $4 + \varepsilon$  dimensions the soft and collinear divergences appear as poles in  $\varepsilon$ . Often, there will be configurations in which both soft and collinear divergences appear together, giving rise to double-pole terms in  $\varepsilon$ .

Like virtual contributions, the real emission processes also develop soft and collinear divergences through the phase space integrals [6]. For example, the one corresponding to the parton-level process

$e^- + q(\bar{q}) \rightarrow q(\bar{q}) + g + g$  contains an integral of the form  $\int d^{n-1}k/(2k^0) \int d^{n-1}p'/(2p^0)(k-p)^{-2}(k+p')^{-2}\delta^{(4)}(q+p-p'-k)$ . This integral diverges when  $k \rightarrow 0$ ; similarly, when  $k$  is parallel to  $p$  or  $p'$  it gives rise to collinear divergences in  $n = 4$  dimensions and develops poles in  $\varepsilon$  when  $\varepsilon \rightarrow 0$ . We also find configurations containing both soft and collinear divergences simultaneously. They are, again, represented by double-pole terms in  $\varepsilon$ .

According to the Kinoshita–Lee–Nauenberg (KLN) theorem, the soft divergences cancel between pure virtual corrections and those with at least one real parton emission. Similarly, the final-state collinear divergences also cancel among themselves when they are summed up. However, those resulting from configurations involving an initial-state parton do not vanish. According to the KLN theorem, if we sum over degenerate states the resulting cross section will be free of collinear divergences. In DIS, as we do not integrate inclusively over the momentum of the initial-state parton, the collinear singularities arising between initial- and final-state partons remain at every perturbative order.

In summary, the higher-order contributions to parton-level subprocesses always develop initial-state collinear divergences due to the presence of massless partons. This is consistent with the KLN theorem and hence not unexpected. In the following we discuss how these initial-state collinear divergences go away when we perform a sum over the momenta and quantum numbers of initial-state partons, as we expect from the KLN theorem. We demonstrate this using the mass factorisation theorem.

The mass factorisation theorem encapsulates the factorisation properties of parton-level cross sections that develop initial-state collinear divergences due to the presence of massless partons. As per the theorem, the collinear divergences factor out from the parton-level process in a way that only depends on the state of the incoming parton before it scatters off the electron state. The result depends on how the initial-state parton undergoes the QCD dynamics to become another parton that eventually scatters with the electron. According to the factorisation theorem, we can express in DIS a generic parton-level subprocess involving a parton type “ $a$ ” as

$$\frac{\hat{\sigma}_a(z, Q^2, \varepsilon)}{z} = \hat{\sigma}^{(0)}(Q^2) \sum_{b=q, \bar{q}, g} \int_z^1 \frac{dz_1}{z_1} \frac{\Delta_b\left(\frac{z}{z_1}, Q^2, \mu_F^2, \varepsilon\right)}{(z/z_1)} \Gamma_{ba}(z_1, \mu_F^2, \varepsilon) \quad . \quad (16)$$

Here,  $\Delta_a$  is called coefficient function, which is finite when  $\varepsilon \rightarrow 0$ , and  $\Gamma_{ba}$  is called the Altarelli–Parisi kernel, which contains the collinear divergences present in  $\hat{\sigma}_a$ . Note that the collinear divergences are factored out from  $\hat{\sigma}_a$  at the scale  $\mu_F^2$ , called factorisation scale.  $\Gamma_{ba}$  in the  $\overline{MS}$  scheme only contains the poles in  $\varepsilon$ . Since the collinear divergences are purely due to the QCD dynamics among the partons, they are process independent. In other words, the kernels  $\Gamma_{ba}(z, \mu_F^2, \varepsilon)$  do not depend on the interaction of the partons with the leptons. Both  $\Delta_a$  and  $\Gamma_{ba}$  are normalised in such a way that they can be expanded in powers of  $\alpha_s(\mu_F^2)$ :  $X = \sum_{i=0}^{\infty} \alpha_s^i(\mu_F^2) X^{(i)}(z, Q^2, \mu_F^2, \varepsilon)$ , where  $X = \Delta_a, \Gamma_{ba}$ , with  $X^{(0)} = \delta(1-z)$ .

In the above expression, we have arranged the integrand in such a way that the integral is a Mellin convolution of two functions, namely  $\Delta_a(x)/x$  and  $f_b(x)$ . The Mellin convolution of “ $n$ ” functions  $f_1(x), f_2(x) \dots f_n(x)$  is defined by

$$f_1(x) \otimes f_2(x) \otimes \dots \otimes f_n(x) = \prod_{i=1}^n \left( \int_0^1 dx_i f_i(x_i) \right) \delta \left( x - \prod_{i=1}^n x_i \right) \quad . \quad (17)$$

Note that the convolution is symmetric under interchange of the functions. In addition, under Mellin transformation the above convolution results in a simple product of Mellin moments of  $f_i(x)$ . If we define the Mellin transformation by  $f_N = \int_0^1 dx x^{N-1} f(x)$ , then

$$\int_0^1 dx x^{N-1} f_1(x) \otimes f_2(x) \otimes \dots \otimes f_n(x) = \prod_{i=1}^n f_{i,N} \quad . \quad (18)$$

Note that the right hand side contains simple products of Mellin moments of the functions  $f_i(x_i)$ .

In the notation of convolutions, the mass factorisation formula reads as

$$\frac{\hat{\sigma}_a(z, Q^2, \varepsilon)}{z} = \hat{\sigma}_0(Q^2, \varepsilon) \sum_{b=q, \bar{q}, g} \frac{\Delta_b(z, Q^2, \mu_F^2, \varepsilon)}{z} \otimes \Gamma_{ba}(z, \mu_F^2, \varepsilon) \quad . \quad (19)$$

Expressing the parton model result for DIS in the convolution form and substituting the mass factorised result, we obtain

$$\frac{1}{x_{Bj}} \sigma(x_{Bj}, Q^2) = \hat{\sigma}_0(Q^2, \varepsilon) \sum_{a,b=q, \bar{q}, g} \hat{f}_a(x_{Bj}) \otimes \Gamma_{ba}(x_{Bj}, \mu_F^2, \varepsilon) \otimes \frac{1}{x_{Bj}} \Delta_b(x_{Bj}, Q^2, \mu_F^2, \varepsilon) \quad . \quad (20)$$

Since the left hand side is finite, one expects that the convolution of  $\hat{f}_a$  and  $\Gamma_{ba}$  should be finite. Since the convolution between  $f_a$  and  $\Gamma_{ba}$  sums up the initial state configurations, the collinear divergences contained in  $\Gamma_{ba}$  cancel against those in  $\hat{f}_a$  in accordance with the KLN theorem. Hence, we can relate their convolution to a finite function,  $f_b(z, \mu_F^2)$ , as

$$f_b(z, \mu_F^2) = \sum_{a=q, \bar{q}, g} \hat{f}_a(z) \otimes \Gamma_{ba}(z, \mu_F^2, \varepsilon) \quad . \quad (21)$$

We call  $f_b$  the mass-factorised parton distribution function, which is defined at the factorisation scale  $\mu_F^2$  and is finite when  $\varepsilon \rightarrow 0$ . In terms of  $f_b$ , the hadronic cross section reads

$$\frac{1}{x_{Bj}} \sigma(x_{Bj}, Q^2) = \hat{\sigma}_0(Q^2) \sum_{a,b=q, \bar{q}, g} f_a(x_{Bj}, \mu_F^2) \otimes \frac{1}{x_{Bj}} \Delta_b(x_{Bj}, Q^2, \mu_F^2) \quad . \quad (22)$$

The fact that  $\hat{f}_a$  is independent of  $\mu_F^2$  leads to the renormalisation group equation in the infrared (collinear) sector of QCD:

$$\mu_F^2 \frac{d}{d\mu_F^2} f_a(z, \mu_F^2) = \frac{1}{2} P_{ab}(z, \mu_F^2) \otimes f_b(z, \mu_F^2) \quad , \quad (23)$$

where  $P_{ab}$  is the matrix element of  $P(z, \mu_F^2) = \mu_F^2 d \log \Gamma(z, \mu_F^2, \varepsilon) / d\mu_F^2$  are finite and are called Altarelli–Parisi (AP) splitting functions. They are computable in perturbative QCD as:  $P = \sum_{i=0}^{\infty} a_s^i(\mu_F^2) P^{(i)}(z)$ . Few comments are in order: while  $f_b$  and  $\Delta_b$  depend on the scale  $\mu_F$ , the convolution of them is independent of  $\mu_F$ , provided that the AP splitting functions  $P$  and the coefficient functions  $\Delta_a$  are known to all orders in  $a_s(\mu_F^2)$ . Since these perturbative results are known only to few orders in  $a_s$ , the predictions will always be sensitive to  $\mu_F$ . However, by varying the scale around the hard scale  $Q^2$ , we can estimate the theoretical uncertainty due to the truncation of the perturbative series.

The other source of theoretical uncertainty is from ultraviolet renormalisation. Note that  $\Delta_b$  is computable in perturbative QCD as a power series in  $a_s(\mu_R^2)$ . While  $\Delta_b$  is  $\mu_R$ -independent, the truncated result will be sensitive to  $\mu_R$ . Like  $\mu_F$ , we can vary  $\mu_R$  to estimate the error resulting from the truncation of the perturbative series of  $\Delta_b$ . In addition to these dependences, the predictions will depend on the hard scale  $Q^2$  even if the series is summed to all orders. Explicit calculations reveal that the  $Q^2$  dependence is through certain logarithms whose coefficients are controlled by the  $\beta$  function as well as by AP splitting functions. Due to the presence of these logarithms, the hadronic cross section will depend on the hard scale  $Q^2$ , leading to a violation of the Björken scaling. The fact that the coefficients of these  $\log(Q^2)$  terms, namely beta and the AP splitting functions, are computable order by order in  $a_s$ , one can predict the exact dependence of the hadronic cross section on  $Q^2$ . Remarkably, precise measurements of DIS cross sections at various  $Q^2$  values confirm the predictions of perturbative QCD.

### 4.3 Threshold resummation

So far, we studied the factorisation of collinear divergences in the partonic cross sections and their universal/process-independent structure in terms of  $\beta$  and AP splitting functions. Our next task is to study the factorisation properties of  $\hat{\sigma}_a$  in the threshold limit [8]. The threshold limit in DIS is defined by the limit when  $z \rightarrow 1$ . We restrict ourselves to quark or anti-quark initiated processes. The threshold-enhanced terms in the mass factorised cross sections  $\Delta_{q(\bar{q})}$  take the following form:

$$\begin{aligned} \Delta_{q(\bar{q})}(z, a_s) = & \Delta_{q(\bar{q})}^{\delta}(a_s) \delta(1-z) + \sum_{j=0}^{\infty} \left[ \Delta_{q(\bar{q}),j}^{\mathcal{D}}(a_s) \mathcal{D}_j(z) \right. \\ & \left. + \Delta_{q(\bar{q}),j}^{\log z}(a_s) \log^j(1-z) \right] + \Delta_{q(\bar{q})}^{\text{bNSV}}(z) \quad . \end{aligned} \quad (24)$$

The upperscript labels  $\delta$  and  $\mathcal{D}$  indicate the terms proportional to  $\delta(1-z)$  and to the distributions  $\mathcal{D}_j$  (defined below). Both terms, which are the leading ones, are called “soft-plus-virtual” (SV). The label  $\log z$  indicates the “next-to-SV” (NSV) terms, proportional to  $\log^j(1-z)$ , while the last term describes the remaining “beyond NSV” (bNSV) processes. The  $\mathcal{D}$  terms contain the “+” distributions  $\mathcal{D}_j(z) = (\log^j(1-z)/(1-z))_+$ , defined by

$$\int_0^1 dz \mathcal{D}_j(z) f(z) = \int_0^1 dz \left( \frac{\log^j(1-z)}{1-z} \right) (f(z) - f(1)) \quad . \quad (25)$$

In the threshold region, these distributions can become dominant. In addition, at a given order “ $n$ ” in  $a_s$ , these distributions will be as big as the inverse of  $1/a_s^n$ , resulting in order-one terms of the form  $a_s \beta_0 \log N$ . These terms can spoil the reliability of the perturbative approach in the threshold region. The solution to this problem is to sum up these order-one terms in a systematic fashion, so that the modified perturbative expansion provides reliable predictions. This was achieved independently by Sterman and by Catani and Trentedue in the Mellin “ $N$ ” space. The result takes the form

$$\log \Delta_{q(\bar{q}),N}^{\text{res}}(a_s) = \log \tilde{g}_0(a_s) + \log(N) g_1(w) + \sum_{i=0}^{\infty} a_s^i g_{i+2}(w) \quad , \quad (26)$$

where  $w = a_s \beta_0 \log(N)$ . In the above equation  $g_0$  is independent of  $N$  and they result from  $\delta(1 - z)$  terms in the threshold limit, while  $g_i(w)$  are from  $\mathcal{D}_j(z)$  terms. The SV limit in the  $N$  space corresponds to taking the large- $N$  limit. In the large- $N$  limit, being  $a_s$  small,  $w$  becomes order-one and, hence, we need to rearrange the perturbative series in such a way that the sum over “ $w$ ” terms is performed to all orders, as in the above equation.

In the following, see Ref. [7], we study the all-order perturbative structure of the coefficient function in terms of  $Q^2$  and  $z$ , by setting up a Sudakov-type differential equation in the kinematic region where  $z$  is closer to the threshold limit  $z = 1$ . We begin with the mass factorisation of  $\hat{\sigma}_q$  and restrict ourselves only to  $\mathcal{D}$  and  $\log z$  terms in  $\hat{\sigma}_q$  and  $\Gamma_{bq}$ , with  $a, b = q, \bar{q}, g$ ; then we find that in  $\Gamma_{bq}$  only  $b = q$  will contribute. Note that  $\Gamma_{gq}$  does not contain any “+” distributions or terms that can lead to SV or NSV terms for  $\Delta_q$  in the limit  $z \rightarrow 1$ . Similar arguments can easily convince one that  $\Delta_g$  does not contain any SV terms. You will recall that the partonic cross section  $\hat{\sigma}_q$  beyond leading order gets contributions from processes of three different sources: pure virtual, pure real emissions and real emission-virtual together. The pure virtual contribution to  $\hat{\sigma}_q$  is found to be  $|\hat{F}_q(\hat{a}_s, \mu^2, Q^2, \varepsilon)|^2$  where  $F_q$  is nothing but the form factor of the vector-boson–quark–anti-quark vertex. We factor out  $|\hat{F}_q|^2$  from  $\hat{\sigma}_q$  and define the “jet function”  $\mathcal{S}_{J,q}$  as the following quotient:

$$\mathcal{S}_{J,q}(\hat{a}_s, \mu^2, q^2, z, \varepsilon) = |\hat{F}_q(\hat{a}_s, \mu^2, Q^2, \varepsilon)|^{-2} \delta(1 - z) \otimes \hat{\sigma}^{\text{SV+NSV}}(q^2, z, \varepsilon) \quad . \quad (27)$$

Note that  $\mathcal{S}_{J,q}$  is computable order by order in  $a_s$  and is also renormalisation group invariant with respect to the scale  $\mu_R$ . A great deal of understanding is provided about the infrared and UV structure of the form factors (FF) by the Sudakov “ $K + G$ ” equation (see below), and about the AP kernels by the AP evolution equation in terms of universal anomalous dimensions. The factorisation of the IR singularity in a form factor implies that  $\hat{F}_q(Q^2) = Z_{\hat{F}_q}(Q^2, \mu_s^2) F_{q,\text{fin}}(Q^2, \mu_s^2)$ , where  $Z_{\hat{F}_q}$  is IR singular, while  $F_{q,\text{fin}}$  is IR finite, and the scale  $\mu_s$  is the IR factorisation scale. The peculiar IR singularity structure of  $Z_{\hat{F}_q}$  implies that the kernel defined by  $K_q = 2d \log Z_{\hat{F}_q} / d \log Q^2$  is independent of  $Q^2$  and contains only IR poles in  $\varepsilon$ , while  $G_q = 2d \log \hat{F}_q / d \log Q^2$  is finite, as well as dependent on  $Q^2$ . This implies that  $\hat{F}_q$  satisfies the  $K + G$  equation, namely  $d \log \hat{F}_q / d \log Q^2 = 1/2(K_q(\mu_s^2, \varepsilon) + G_q(Q^2, \mu_s^2, \varepsilon))$ . The solution to the  $K + G$  equation is given by

$$\hat{F}_q(Q^2, \varepsilon) = \exp \left( \int_0^{Q^2} \frac{d\lambda^2}{\lambda^2} \Gamma_{\hat{F}_q}(\lambda^2, \varepsilon) \right) \quad , \quad (28)$$

where  $\hat{F}_q(Q^2 = 0, \varepsilon) = 1$  and  $\Gamma_{\hat{F}_q} = (K_q + G_q)/2$ . In terms of  $\mathcal{S}_J$ , the mass-factorised cross section reads as

$$\Delta_q(z, Q^2, \varepsilon) = |\hat{F}_q(Q^2, \varepsilon)|^2 \delta(1 - z) \otimes \mathcal{S}_{J,q}(q^2, z, \varepsilon) \otimes \Gamma_{qq}^{-1}(z, \mu_F^2, \varepsilon) \quad . \quad (29)$$

Differentiating the above equation with respect to  $\log Q^2$  and using the  $K + G$  equation for  $\hat{F}_q$ , we obtain a ( $K + G$ )-like equation for  $\mathcal{S}_{J,q}$ :

$$q^2 \frac{d\mathcal{S}_{J,q}}{dq^2} = \Gamma_{\mathcal{S}_{J,q}}(q^2, z) \otimes \mathcal{S}_{J,q}(z, q^2) \quad , \quad (30)$$

where  $\Gamma_{S_{J,q}} = q^2 d\Delta_q/dq^2 \left( \mathcal{C} \log \Delta_q - \log |\hat{F}_q|^2 \delta(1-z) \right)$ . Here  $\mathcal{C}$  means that in the perturbative expansion of  $\log \Delta_q$  in powers of  $a_s$ , the product of  $z$  dependent functions should be understood as convolution of  $z$  dependent functions. We find that  $\Gamma_{S_{J,q}}$  admits a remarkable structure, namely, it can be written as the sum of a  $q^2$ -independent IR-divergent term and a  $q^2$ -dependent IR-finite term. If we identify the former as a “ $K$ -type” term and the latter as a “ $G$ -type” term, then the jet function  $\mathcal{S}_{J,q}$  does satisfy a  $(K + G)$ -type equation. This implies that the jet function  $\mathcal{S}_{J,q}$  can be factorized into an IR-divergent renormalisation constant  $Z_q(z, q^2, \mu_s^2, \varepsilon)$  and a finite quotient  $\mathcal{S}_{J,q,\text{fin}}(z, q^2, \mu_s^2, \varepsilon)$ , where  $\mu_s$  is the factorisation scale. That is,  $\mathcal{S}_{J,q}(z, q^2, \varepsilon) = Z_q(z, q^2, \mu_s^2, \varepsilon) \mathcal{S}_{J,q,\text{fin}}(z, q^2, \mu_s^2, \varepsilon)$ . The solution to the above differential equation takes the following form:

$$\mathcal{S}_{J,q}(q^2, z, \varepsilon) = \mathcal{C} \exp \left( \int_0^{q^2} \frac{d\lambda^2}{\lambda^2} \Gamma_{S_{J,q}}(\lambda^2, z, \varepsilon) \right) . \quad (31)$$

The general structure of the exponent can be derived from the explicit perturbative results for the mass factorisation coefficient function  $\Delta_q$ , the form factor  $\hat{F}_q$  and the AP factorisation kernel  $\Gamma_{qq}$ , the latter being known to the third order in  $a_s$ . In particular, the divergent part of the jet function can be determined by noting that  $\mathcal{S}_{J,q}$  should contain the right singularities to cancel those from the form factor and the AP kernel. You will recall that the singularity structures of the form factor and the AP kernel are controlled by universal anomalous dimensions such as  $A_q, B_q, f_q, C_q, D_q$  and the  $\beta$  function of QCD. We also observe that in dimensional regularisation both the form factor and  $\Delta_q$  show certain structures related to transcendentality at every order in perturbation theory. Using the fact that  $\Delta_q$  is finite and its transcendentality structure, we propose a solution for the jet function  $\mathcal{S}_{J,q}$  to all orders:

$$\log \mathcal{S}_{J,q} = \sum_{i=1}^{\infty} \hat{a}_s^i \left( \frac{q^2(1-z)}{\mu^2} \right)^{i \frac{\varepsilon}{2}} S_\varepsilon^i \left( \frac{i\varepsilon}{2(1-z)} \right) \left( \hat{\Phi}_q^{\text{SV},(i)}(\varepsilon) + \frac{2}{i\varepsilon} (1-z) \hat{\varphi}_q^{(i)}(z, \varepsilon) \right) . \quad (32)$$

In the above equation,  $\hat{\Phi}^{\text{SV}}(\varepsilon)$  encodes all contributions from the pure distributions, while  $\hat{\varphi}_q^{(i)}(z, \varepsilon)$  encodes  $z$  dependent next-to-SV terms. The AP kernels  $\Gamma_{qq}$  satisfy the AP evolution equation and, in the approximation we work with, they are controlled only by the diagonal AP slitting functions  $P_{qq}$ . Hence, the all-order solution takes the simple form:

$$\Gamma_{qq}(\mu_F^2, z, \varepsilon) = \mathcal{C} \exp \left( \frac{1}{2} \int_0^{\mu_F^2} \frac{d\lambda^2}{\lambda^2} P_{qq}(\lambda^2, z, \varepsilon) \right) . \quad (33)$$

The AP splitting function is known to the third order in perturbation theory and the SV distributions and NSV logarithms present in them are controlled by universal cusp and collinear anomalous dimensions. Putting all of them together we obtain

$$\Delta_q(Q^2, z, \varepsilon) = \mathcal{C} \exp \left( \int_{\mu_F^2}^{Q^2(1-z)} \frac{d\lambda^2}{\lambda^2} P'_{qq}(a_s(\lambda^2), z) + Q_q(a_s(Q^2(1-z)), z) \right) , \quad (34)$$



where

$$Q_q = \left( \frac{1}{2(1-z)} \overline{G}_{q,J}^{\text{SV}}(a_s(Q^2(1-z))) \right)_+ + \overline{\varphi}_{f,q}(a_s(Q^2(1-z)), z) + \ln C_{0,q}(a_s(\mu_R^2), Q^2, \mu_R^2, \mu_F^2) \quad (35)$$

In the above equation, we have decomposed  $Q_q$  in terms of pure SV and  $z$  dependent contributions denoted by  $\overline{G}_{q,J}^{\text{SV}}$ , next to SV  $\overline{\varphi}_{f,q}$  and  $z$  independent matching term  $C_{0,q}$ . Expanding the exponent in powers of  $a_s$ , we can obtain both SV and NSV terms. For example, if we know the exponent to order  $a_s$ , the expansion of the exponential will provide leading SV terms  $(\mathcal{D}_3, \mathcal{D}_2), (\mathcal{D}_5, \mathcal{D}_4), \dots, (\mathcal{D}_{2i-1}, \mathcal{D}_{2i-2})$  and leading NSV terms  $\log^3(1-z), \log^5(1-z), \dots, \log^{2i-1}(1-z)$  at  $a_s^2, a_s^3, \dots, a_s^i$ , respectively, for all  $i$ . Similarly, from the knowledge of  $a_s^2$  results, we can provide coefficients of  $(\mathcal{D}_2, \mathcal{D}_1), (\mathcal{D}_4, \mathcal{D}_3), \dots, (\mathcal{D}_{2i-3}, \mathcal{D}_{2i-4})$  and leading NSV terms  $\log^4(1-z), \log^6(1-z), \dots, \log^{2i-2}(1-z)$  at  $a_s^3, a_s^4, \dots, a_s^i$ , respectively, for all  $i$ . This can be generalised for an arbitrary order in  $a_s$  for the exponent. In Mellin space, after reorganising the exponent according to the logarithmic accuracy we obtain

$$\begin{aligned} \log \Delta_q(Q^2, N, \varepsilon) &= \log g_0^q(a_s(\mu_R^2)) + \tilde{g}_1^q(\omega) \log N + \sum_{i=0}^{\infty} a_s^i(\mu_R^2) \tilde{g}_{i+2}^q(\omega) \\ &+ \frac{1}{N} \sum_{i=0}^{\infty} a_s(\mu_R^2) h_i^q(\omega, N) \quad , \end{aligned} \quad (36)$$

where

$$\tilde{g}_i^q(\omega) = g_i^q(\omega) + \frac{1}{N} \overline{g}_i^q(\omega), \quad \omega = a_s(\mu_R^2) \beta_0 \log N \quad (37)$$

and

$$h_0^q(\omega, N) = h_{00}^q(\omega) + h_{01}^q(\omega) \log N, \quad h_i^q(\omega, N) = \sum_{k=0}^{\infty} h_{ik}^q \log N \quad . \quad (38)$$

In the above equation, the result of the Mellin integrals are decomposed into  $z$  dependent SV, NSV terms and  $z$  independent matching terms. Using the above equation, one can predict resummed contributions to leading logarithmic (LL) accuracy, next-to-leading logarithmic accuracy etc. in a systematic fashion for the inclusive cross section at various  $Q^2$  values.

In summary, we find that the perturbative results not only help us to make precise predictions from the theory but also unravel universal structures of the theory. The comparison of the predictions against experimental observations can put the theory on firm footing. In addition, they can put stringent bounds on the parameters of physics beyond the SM. Similarly, understanding the UV and IR structures of the theory can provide ingredients to sum up potentially important contributions from all orders and also shed light on the power corrections. For example, the resummation of threshold and next-to-threshold logarithms was possible due to the universal structure in the perturbative predictions.

#### 4.4 Multi-leg and multi-loop Feynman diagrams

In the following we will discuss various methods of obtaining perturbative results [9]. Let us restrict ourselves to the computation of scattering cross sections at hadron colliders. The task reduces to writing down on-shell scattering amplitudes, squaring their moduli, performing loop integrals and then phase-space integrals, taking into account the constraints from experiments. One begins with setting up Feynman rules and then using the rules to write down the amplitudes. Often, one finds that the number of Feynman diagrams becomes very large as we increase the number of legs or the order of the perturbative expansion. For example, the number of Feynman diagrams is four for the tree-level process  $g + g \rightarrow g + g$ , while for  $g + g \rightarrow 5g$  it becomes 2485, and 10525900 for  $g + g \rightarrow 8g$ . These numbers will increase if we include loop corrections to the tree-level amplitudes. One often resorts to computer programs to generate these diagrams and obtain the corresponding amplitudes in the analytical form. Packages such as FeynArts and QGRAF are found to be very useful for this purpose. Next, we need to compute the square of the modulus of the total amplitude. Note that the amplitudes are made up of Dirac spinors,  $u_i(p, s), v_i(p, s), \bar{u}_i(p, s), \bar{v}_i(p, s)$ , chains of Dirac matrices, polarisation vectors of gauge fields,  $\epsilon_\mu(q, \lambda), \epsilon_\nu^*(q, \lambda)$ . Here,  $p$  is the momentum,  $s$  is the spin of the Dirac particle, while  $q$  is the momentum and  $\lambda$  is the polarisation of the gauge field. In addition, the amplitude will contain chains of Dirac matrices and generators and structure constants of the  $SU(N)$  group. Note that the index  $i$  in the Dirac spinor is due to  $SU(N)$ . The modulus of the sum of the amplitudes involves the computation of a large number of traces of Dirac matrices and also the simplification of the  $SU(N)$  generators and structure constants. Again, one can set up computer codes to perform this task, if we have a small number of amplitudes. For processes with larger numbers of amplitudes, it is desirable to simplify the amplitudes so that the evaluation of their moduli is manageable. One can make the simplification at the amplitude level if the properties of Dirac spinors and the freedom of gauge choice are exploited. The documentation of a powerful technique using helicity amplitudes can be found in the celebrated book “The Ubiquitous photon: Helicity method for QED and QCD” by R. Gastmans and T.T. Wu. The modern versions of this approach provide a set up suitable for faster computer codes. For example, one defines simpler notations for helicity amplitudes, namely  $(1 \pm \gamma_5)u(k_i) \equiv u_\pm(k_i)$ ,  $(1 \pm \gamma_5)v(k_i) \equiv v_\mp(k_i)$  and  $\bar{u}_\pm(k_i) = \bar{v}_\mp(k_i) \equiv \langle i^\pm | \equiv \langle k_i^\pm |$ . Furthermore, we can define  $|i\rangle = |i^+\rangle$  and  $|i] = |i^-\rangle$  and obtain  $\bar{u}_-(k_i)u_+(k_j) = \langle ij \rangle$  and  $\bar{u}_+(k_i)u_-(k_j) = [ij]$ . In addition, the Gordon identity  $[p\gamma^\mu p] = 2p^\mu$ , the Fierz identity  $\langle p\gamma^\mu q|r\gamma_\mu s\rangle = 2\langle ps\rangle[rq]$  and the Schouten identity  $\langle pq\rangle\langle rs\rangle + \langle pr\rangle\langle sq\rangle + \langle ps\rangle\langle qr\rangle = 0$  can be used to simplify the expressions at the amplitude level. For gluon polarisations, one uses  $\epsilon_\mu^+(p, q) = \langle q|\gamma_\mu p\rangle/(\sqrt{2}\langle qp\rangle)$  and  $\epsilon_\mu^-(p, q) = -[q|\gamma_\mu p\rangle/(\sqrt{2}[qp])$  with  $p_\mu \epsilon^\mu = 0$  and  $q$  being any light-like vector.

Like Dirac spinors and gamma matrices, the generators and structure constants of the  $SU(N)$  group in the vertices of the amplitudes pose additional complexity. However, a remarkable simplification is achieved by stripping them off from each amplitude. Using the  $SU(N)$  algebra  $[T^a, T^b] = if^{abc}T^c$ , or, equivalently,  $if^{abc} = 2(T_r(T^a T^b T^c) - T_r(T^b T^a T^c))$ , one can replace all the  $if^{abc}$  terms by the latter identity, to obtain a color-stripped amplitude. One finds a tree-level amplitude involving  $n$  gluons,

$$\mathcal{A}_n^{(0)}(g_1, g_2, \dots, g_n) = g_s^{n-2} \sum_{\alpha \in S_n/Z_n} 2T_r(T^{a_{\sigma(1)}} \dots T^{a_{\sigma(n)}}) A_n^{(0)}(g_{\sigma(1)}, \dots, g_{\sigma(n)}) \quad , \quad (39)$$

where  $A_n^{(0)}$  are called partial amplitudes. They do not contain any  $SU(N)$  factors and, in addition, each of them is gauge invariant. By repeatedly applying the trace identity one can simplify the original Feynman rules to obtain a new set of Feynman rules for the partial amplitudes. The advantage of this approach is the reduction of the number of independent amplitudes. For example, one gets 12925 amplitudes instead of 10525900. Several approaches were developed in order to improve the speed of the computation. For example, using off-shell currents, Berends and Giele constructed recursion relations, which not only give enormous simplifications, but also reduce the computation time significantly. Thanks to these approaches, the computation of tree level amplitudes is now an accessible task. In addition, the results of certain amplitudes show remarkable simplifications. For example, a certain class of amplitudes, called Parke–Taylor amplitudes, describing  $n$  gluons with specific polarisation assignments, reduces to a single term. The mysterious simplification in the Parke–Taylor amplitudes was explained by Britto, Cachazo, Feng and Witten (BCFW) through their recursion relations, which use Cauchy’s residue theorem together with the analytical structure of the  $SU(N)$  gauge theory amplitudes. Further progress was made by working in twistor space. In addition, there are also efforts to apply some of these methods to amplitudes with loops. Finally, it is worth mentioning that, while all these approaches provide enormous simplification, as well as insights in the theory, each one presents disadvantages when dealing with amplitudes having a large number of legs.

Having obtained the amplitudes, the next task is to perform Feynman loop and phase space integrals [10]. As we had seen in the lectures, the Feynman integrals are sensitive to UV and IR divergences. We use dimensional regularisation regulate them and then proceed to compute them. Often we need to deal with large number of multi loop and many-body phase space integrals, each of them is highly complicated to solve. The standard text book methods do not work and hence one resorts to alternate ones. We present two important developments that revolutionized the computation of Feynman diagrams. Note that loop integrals and phase integrals differ as the later contain delta functions from on-shell external legs. Replacing the delta functions by the imaginary part of the corresponding propagator, we can evaluate phase space integrals the way we evaluate loop integrals. We relate each  $\delta(p_i^2 - m_j^2)$  with the imaginary part of  $1/(p_i^2 - m_j^2 + i\epsilon)$  where  $\epsilon$  is infinitesimally small and positive number in the Feynman prescription. This approach is called the method of reverse unitarity. The method of integration by parts (IBP) identities reduces a set of large number of Feynman integrals to fewer integrals, called master integrals (MI). The method of differential equations (DE) solves the MIs in a most efficient way. The results of these master integrals can often be expressed in terms of certain class of special functions namely classical polylogarithms, multiple polylogarithms, Nielsen integrals, generalised polylogarithms or Goncherev polylogarithms or Chen integrated integrals. One finds cases where more complicated integrals such as elliptic integrals. We give brief account on both IBP and DE in the following.

The typical  $L$  loop Feynman integral in  $n$  space time dimensions with  $p_j, j = 1, \dots, n_e$  external momenta takes the following form

$$\int \prod_{i=1}^L d^m l_i \frac{\mathcal{N}(l_i, p_j)}{D_1^{\alpha_1} \dots D_M^{\alpha_M}} \quad (40)$$

where  $D_i$  are propagators involving the momenta  $\{l_i\}$  and  $\{p_j\}$  and masses  $\{m_k\}$ . The number of scalar products here at the most is  $L(n_e + L/2 - 1/2)$ . Beyond one loop, this number is always greater than  $M$

and hence we can introduce auxiliary propagators so that the extra scalar products are expressed in terms of them. This way we can express any loop integral in terms of the expanded set of  $N$  propagators where the exponent of the propagator can be negative. Since the loop integrals are invariant under the shift of the loop momenta namely  $l_i \rightarrow l_i + \sum_k c_k p_k$  for some constants  $c_k$ , we can relate many integrals and reduce number of integrals that we deal with. Similarly, integrals are invariant under Lorentz transformation of external momenta, i.e  $p_j^\mu \rightarrow p_j^\mu + \omega^{\mu\nu} p_{j,\nu}$ , giving

$$\omega^{\mu\nu} \sum_j^{n_e} p_{j,\nu} \frac{\partial}{\partial p_j^\mu} \int \prod_{i=1}^L d^n l_i \frac{1}{D_1^{\alpha_1} \dots D_N^{\alpha_N}} = 0 \quad (41)$$

for arbitrary antisymmetric constant  $\omega^{\mu\nu}$

$$\sum_j^{n_e} \left( p_{j,\nu} \frac{\partial}{\partial p_j^\mu} - p_{j,\mu} \frac{\partial}{\partial p_j^\nu} \right) \int \prod_{i=1}^L d^n l_i \frac{1}{D_1^{\alpha_1} \dots D_N^{\alpha_N}} = 0 \quad . \quad (42)$$

The generalisation of the above two transformations is given by  $l_i \rightarrow q_i = c_{ij} l_j + d_{ij} p_j$ . The invariance under this transformation gives

$$\int \prod_{i=1}^L d^n l_i \frac{\partial}{\partial l_{i,\mu}} \left( \frac{q_t^\mu}{D_1^{\alpha_1} \dots D_N^{\alpha_N}} \right) = 0 \quad . \quad (43)$$

The above equation is called integration by parts identity (IBP). Since the exponents are arbitrary, we can generate infinite of IBP identities, of which most of them are redundant. One can show that a finite set of integrals can solve these identities, we call these integrals the Master Integrals. This process reduces the task of computing too many integrals.

Given the set of MIs, our next task to compute each one of them efficiently. The standard approach is to apply Feynman's trick or Schwinger parameterization to the integrands which allows to perform integration over loop momenta. However, the standard approach brings in parametric integrals which are hard to perform beyond one loop integrals. The method of DE is the alternate approach to solve MIs beyond one loop ones. Here, we first make set of Lorentz scalars out of all the external momenta and their masses. Let this set be  $\{s_{ij}\} = \{x_1, \dots, x_m\} = \vec{x}$  where  $m$  is the number of scalars  $s_{ij} = (p_i + p_j)^2$  constructed out of  $p_j$ , including their masses. Then construct a set  $\{\partial/\partial s_{ij}\} = \partial/\partial \vec{x}$  and apply them on all the MIs. Denoting MIs by  $\vec{I}(\{s_{ij}/\varepsilon\}) = I_1(\vec{x}, \varepsilon), \dots, I_{N_m}(\vec{x}, \varepsilon)$ , where  $N_m$  is the number of MIs, we obtain

$$\frac{\partial}{\partial \vec{x}} \vec{I}(\vec{x}, \varepsilon) = \hat{B}(\vec{x}, \varepsilon) \cdot \vec{J}(\vec{x}, \varepsilon) = \hat{A}(\vec{x}, \varepsilon) \cdot \vec{I}(\vec{x}, \varepsilon) \quad . \quad (44)$$

In the above we used IBP identities to express  $\vec{J}$  in terms of  $\vec{I}$  which converts  $\hat{B}$  to  $\hat{A}$ . The coupled differential equations being first order ones are straightforward to solve provided  $\vec{A}$  has fewer entries and the boundary integrals are known. Often, the later are easy to obtain for certain choice of  $\vec{x} = \vec{x}_0$ . There are several ways to solve the system of DEs depending on the structure of  $\hat{A}$ . In fact one can transform  $\vec{I} \rightarrow \vec{I}' = U \cdot \vec{I}$  in such a way that  $\hat{A}' = U \cdot \hat{A} \cdot U^{-1} - U \cdot \partial/\partial \vec{x} \cdot U^{-1}$  takes the simple form. For the case when  $\hat{A}'(\vec{x}, \varepsilon) = \varepsilon \tilde{A}(\vec{x})$ , the solution demonstrates a peculiar all order structure in  $\varepsilon$ . One finds the

solution of  $\vec{I}'$  is a set of iterated integrals with uniform transcendentality:

$$I'(\vec{x}, \varepsilon) = \mathcal{P} \exp \left( \varepsilon \int_{\mathcal{C}} d\vec{x} \cdot \tilde{A}(\vec{x}) \right) \cdot \vec{I}'(\vec{x}_0, \varepsilon) \quad , \quad (45)$$

where  $\mathcal{P}$  is path ordering along the curve  $\mathcal{C}$ . If we assign transcendentality weight  $i$  for  $\varepsilon^{-i}$  and  $\log^i(g(x))$ , the terms in the expansion will have uniform weight.

## 5 Conclusion

We have demonstrated how the theory of strong interaction, namely quantum chromodynamics, plays an important role to understand physics at subatomic level in high energy experiments. We have shown that QCD can be applied in a systematic way using its factorisation properties. The perturbative methods demonstrate that various observables can be computed reliably. In addition we have discussed how large logarithms show up at the threshold region and how resummation of them to all orders can be performed. We have also discussed few modern techniques that are available to perform various computations efficiently.

## Acknowledgements

I sincerely thank the organizers of the 2022 Asia–Europe–Pacific School of High-Energy Physics for their invitation and thank the students for their interactions during the school.

## References

- [1] R. Bowley and E. Copeland, *Quarks in Sixty Symbols* by Brady Haran for the University of Nottingham, [YouTube recording](#); D.J. Griffiths, *Introduction to elementary particles*, 2nd ed. (Wiley, New York, 2008), [doi:10.1002/9783527618460](https://doi.org/10.1002/9783527618460); I.S. Hughes, *Elementary particles*, 3rd ed. (Cambridge University Press, Cambridge, 1991).
- [2] Ta-Pei Cheng and Ling-Fong Li, *Gauge theory of elementary particle physics* (Oxford University Press, Oxford, 1984), <https://library.oapen.org/handle/20.500.12657/59106>.
- [3] J. Blümlein, Deep-inelastic scattering: What do we know?, DESY 23-066, DO-TH 23/06, arXiv:2306.01362 [hep-ph], [doi:10.48550/arXiv.2306.01362](https://doi.org/10.48550/arXiv.2306.01362).
- [4] G.F. Sterman, QCD and jets, Contribution to: TASI 2004, 67–145, hep-ph/0412013, [doi:10.48550/arXiv.hep-ph/0412013](https://doi.org/10.48550/arXiv.hep-ph/0412013).
- [5] T. Muta, *Foundations of quantum chromodynamics: An introduction to perturbative methods in gauge theories*, 3rd ed. (World Scientific, Singapore, 2009), [doi:10.1142/6766](https://doi.org/10.1142/6766).
- [6] R.D. Field, *Applications of perturbative QCD* (Addison-Wesley, Redwood City, CA, 1989).
- [7] V. Ravindran, Higher-order threshold effects to inclusive processes in QCD, *Nucl. Phys. B* **752** (2006) 173–196, [doi:10.1016/j.nuclphysb.2006.06.025](https://doi.org/10.1016/j.nuclphysb.2006.06.025), arXiv:hep-ph/0603041, [doi:10.48550/arXiv.hep-ph/0603041](https://doi.org/10.48550/arXiv.hep-ph/0603041);  
V. Ravindran, On Sudakov and soft resummations in QCD, *Nucl. Phys. B* **746** (2006) 58–76, [doi:10.1016/j.nuclphysb.2006.04.008](https://doi.org/10.1016/j.nuclphysb.2006.04.008), arXiv:hep-ph/0512249, [doi:10.48550/arXiv.hep-ph/0512249](https://doi.org/10.48550/arXiv.hep-ph/0512249).

- [8] S. Catani, Soft gluon resummation: A short review, in *32nd Rencontres de Moriond: QCD and High-Energy Hadronic Interactions*, pp. 331–336, LPTHE-ORSAY 97/46, arXiv:hep-ph/9709503, [10.48550/arXiv.hep-ph/9709503](https://doi.org/10.48550/arXiv.hep-ph/9709503).
- [9] L.J. Dixon, A brief introduction to modern amplitude methods, in *Journeys Through the Precision Frontier: Amplitudes for Colliders* (TASI 2014), pp. 39–97, doi:[10.1142/9789814678766\\_0002](https://doi.org/10.1142/9789814678766_0002).
- [10] S. Weinzierl, *Feynman integrals: A comprehensive treatment for students and researchers* (Springer, Cham, 2023), <https://doi.org/10.1007/978-3-030-99558-4>, arXiv:2201.03593 [hep-th], doi:[10.48550/arXiv.2201.03593](https://doi.org/10.48550/arXiv.2201.03593).

# Hadron spectroscopy

Alexey Zhemchugov

JINR

---

If you query the arXiv electronic database of preprints and look at the topics of published articles, you can see that roughly one in five deals with hadron spectroscopy. During its development, this area of particle physics has experienced several rises and falls. Today, it is a rapidly developing branch of science, which comprises a significant part of the research program of almost every accelerator experiment. It is especially valid for  $B$ -,  $c\tau$ -factories and, certainly, the LHC, where exciting results have been obtained. The purpose of this lecture is to try giving an overview of the current state of hadron spectroscopy through the eyes of an experimentalist.

---

## 1 Introduction: why do we call it spectroscopy?

The term ‘spectroscopy’ was coined by Robert Bunsen and Gustav Kirchhoff in the 1860s, when they built the first *spectroscope*. The idea of the device was to use a prism to split the light into a spectrum and study the resulting picture at the screen. Pretty soon it became clear that the atom of a chemical element is capable of emitting and absorbing light of only a definite frequency (or photon energy, as we say today). In the experiment, this was expressed as the presence in the spectrum of clear emission lines (for heated substances) or absorption lines (for white light going through a cold gas), with a unique line pattern for each element. As a result, the victorious march of atomic spectroscopy began with the discovery of several new chemical elements and the study of the chemical composition of stars. Further, it led to the emergence of the Bohr model of the atom, which made it possible to describe in detail the spectrum of the hydrogen atom and became a keystone of quantum mechanics. At present, atomic spectroscopy remains a powerful and sensitive method of chemical analysis used in almost every laboratory.

After the discovery of radioactivity, the method of spectroscopy was applied to the investigation of nuclear radiation. Naturally, the technique of the experiment has changed a lot. The energy of the radioactive decay made it possible to abandon the use of a gas burner, and more complex instruments began to be used to detect radiation instead of a prism and a screen. The radiation energy was measured either by deflecting charged rays by a magnetic field (so-called *mass spectrometry*) or by measuring the energy released during the absorption of radiation in a scintillator. It is curious that both methods, although with much more sophisticated equipment, are used in the experiment even today. Nuclear spectroscopy made it possible to ‘weigh’ various nuclei and study their energy levels. The problem turned out to be difficult, and the study of the spectra of nuclei and their theoretical description remain on the agenda of nuclear physics till now.

Advances in the study of atomic nuclei led, by the end of the Second World War, to a clear and consistent picture of the structure of matter. The nucleus consists of nucleons (protons and neutrons) bound by the strong nuclear force, carried by the exchange of pions. The weak force is responsible for

---

This article should be cited as: Hadron spectroscopy, Alexey Zhemchugov, DOI: [10.23730/CYRSP-2024-001.105](https://doi.org/10.23730/CYRSP-2024-001.105), in: Proceedings of the 2022 Asia–Europe–Pacific School of High-Energy Physics, CERN Yellow Reports: School Proceedings, CERN-2024-001, DOI: [10.23730/CYRSP-2024-001](https://doi.org/10.23730/CYRSP-2024-001), p. 105.  
© CERN, 2024. Published by CERN under the [Creative Commons Attribution 4.0 license](https://creativecommons.org/licenses/by/4.0/).

the beta decay. Indeed, the remaining problems were the lack of a detailed theoretical description of the nuclear force and the problem of the existence of the muon. The role of the muon in the structure of matter was unclear. The development of technology during the war opened up opportunities for the construction of new, much more powerful accelerators and the start of a series of experiments on them. However, contrary to expectations, instead of solving existing problems, new experiments have brought new mysteries. A large number of new elementary particles was discovered, including a great many short-lived resonances. Among the new particles, analogs of nucleons (antiprotons, antineutrons, hyperons) and interaction carrier particles (kaons,  $\rho$ ,  $\eta$  and  $\omega$  mesons, etc.) were found. It became evident that nuclear matter has a much more complicated structure. Note that the Particle Data Group counts over 250 established particles now, of which more than 50 have been detected in the past two decades.

It is this reason that led to the appearance of hadron spectroscopy. By analogy with atomic and nuclear spectroscopy, hadron spectroscopy aims to study the possible quantum states of strongly interacting particles – hadrons. The purpose of hadron spectroscopy is essentially to find answers to the questions: which hadrons exist? What properties do they possess? What reactions can they undergo? Just like atomic spectroscopy for atoms or nuclear spectroscopy for nuclei, hadron spectroscopy is a key source of experimental data for understanding the structure of hadrons and, ultimately, the strong interaction.

## 2 What are we really measuring?

Since we are talking about an experiment, we should start by discussing what exactly we want and can measure. First of all, we can experimentally establish the very existence of the particle. Then, we can determine its electric charge, mass and lifetime (or particle width). Finally, we can study its possible decays, the relative probabilities of different decay channels, and the angular and energy distributions of the decay products.

From the decay information, one can extract the P-symmetry and C-symmetry of the particle, its total angular momentum  $J$ , as well as the quantum numbers associated with its internal composition: the flavor numbers (strangeness, charm, beauty, isospin), as well as the baryon number (that is, the difference between the number of quarks and antiquarks). Already difficult, these measurements are further complicated by the quantum nature of the experiment: we are measuring a state that can be a superposition of several or many quantum states, and we usually have no way to distinguish them.

Now let us consider the most important experimental techniques.

**The invariant mass spectrum** of a system of particles is perhaps the simplest and most powerful tool for hadron spectroscopy. Indeed, by registering particles, measuring their energies or momenta, and identifying (or assuming) their types, we can easily calculate the mass of a hypothetical parent particle. Having plotted the distribution of this mass, it is easy to detect peaks in the resulting spectrum if the parent particle really exists. At the same time, the spectrum of the invariant mass gives the mass of the particle and, if the resolution of the detector allows it, its width.

**The Dalitz plot**, invented by R. H. Dalitz [1], is a more sophisticated technique used in the analysis of multiparticle decays. The Dalitz plot is constructed as a correlation of the momenta and energies of product particles in the decay. Thus, in the case of a decay into three particles, each decay event is



represented in the Dalitz plot by a point, the position of which is determined by the invariant mass of two pairs of particles in this event. Obviously, the area populated by decay events is confined according to the energy and momentum conservation laws.

Analyzing the population of the Dalitz plot by events, one can establish the presence of intermediate states during the decay of the parent particle to the final state under study, derive their quantum numbers, and quantify the probabilities of individual decay channels (that is, measure the decay matrix elements). It is important to emphasize that all variations in the event density distribution are due to the dynamics, which defines the dependence of the decay matrix elements on the momenta (or energies) of the decay particles, and not by their particular kinematics. This is one of the most important properties of the Dalitz plot.

However, the most complex technique is **the partial-wave analysis, or PWA**. This technique is based on the idea that the interaction of particles is modeled by a coherent sum of resonances. The resulting set of partial waves is fitted to the experimental data, and the energy and angular information are used simultaneously in the fit. As a result of the fit, the individual contributions of partial waves to the final state are determined. This method allows one to deal with broad and overlapping resonances that cannot be separated, for example, in the invariant mass spectrum. An important advantage of partial wave analysis is the proper accounting for the interference between different states with the same quantum numbers. Nevertheless, the PWA is not at all a simple and easy method. Its great technical complexity, heavy computation, and the need for strong theoretical support are the main obstacles to starting to use it. In many cases, you have to work with sets of dozens of waves, which leads to the need to adjust up to hundreds of free parameters in the fit. Therefore, as in any multi-parameter fit, in PWA one has to overcome the instability of the fit, to solve the problem of finding a global minimum, and also deal with the ambiguity due to the inevitable presence of multiple solutions. Interpretation of PWA results is often hampered by rescattering effects, although some methods like using the K-matrix formalism and simultaneous analysis of several complementary reactions may, of course, help. Lastly, experimental conditions, such as background processes or detector resolution, often present a serious problem.

Finally, note that even simple methods should be used with care. As an illustration, we can consider the observation of the  $\tilde{X}(3872)$  state in the COMPASS experiment [2]. The experiment studied the hadron production by a virtual photon in the reaction  $\mu^+ N \rightarrow \mu^+ (J/\psi \pi^+ \pi^-) \pi^\pm N'$ , using COMPASS data collected with incoming muons of 160 GeV/c and 200 GeV/c momentum. In the invariant mass spectrum of the system  $(J/\psi \pi^+ \pi^-)$ , a peak with a mass of  $3860.0 \pm 10.4$  MeV/c<sup>2</sup> and width of less than 51 MeV/c<sup>2</sup> was found near the  $\psi(3686)$  resonance with a statistical significance of 4.1  $\sigma$ . While the mass and width were compatible with the  $X(3872)$  state observed in other experiments before, this turned out to be not enough to draw a conclusion that it was the same particle. The reason was the fact that the shape of the  $\pi^+ \pi^-$  mass distribution from the observed decay into  $J/\psi \pi^+ \pi^-$  had shown disagreement with previous observations for  $X(3872)$ . The observed state was designated as  $\tilde{X}(3872)$  and interpreted as a possible evidence of a new charmonium state. Its nature is still not clear.

### 3 Conventional hadron spectra

The first successful attempt to describe the spectrum of hadrons using a simple model was made by M. Gell-Mann [3] and independently by G. Zweig [4] in 1964. The basis of the model is the assumption about the constituent parts of hadrons, which Gell-Mann called *quarks*. It turned out that three varieties (flavors) of quarks –  $u$  (up),  $d$  (down) and  $s$  (strange) – were enough to describe all hadrons known at that time as their various combinations and even to predict a few yet unknown. Quarks have a fractional electric charge. In the quark model, all baryons are combinations of three quarks, and all mesons are combinations of a quark and an antiquark. The quark model explained the whole variety of particles. From the three constituents you can add up not so many combinations. However, the same combination of quarks can form particles of different masses. These are ‘excited states’, where quarks move around each other in a higher orbit. Over time, the number of quarks that make up hadrons increased to five:  $c$  (charm) and  $b$  (beauty, or bottom) quarks were discovered, along with entire families of charmed and beautiful particles. In the beginning, the quark model was just a convenient way to classify particles. However, after the experimental observation of partons inside the nucleon, and especially after the successful construction of the fundamental theory of the strong interaction – quantum chromodynamics, it is difficult to doubt the reality of quarks, despite their astonishing properties, such as confinement and fractional electric charge.

The mesons are classified in  $J^{PC}$  multiplets. The states with orbital angular momentum equal to zero are the pseudoscalars ( $0^{-+}$ ) and the vectors ( $1^{--}$ ), depending on the quark spins. The orbital excitations with orbital angular momentum equal to 1 are the scalars ( $0^{++}$ ), the axial vectors ( $1^{++}$  and  $1^{+-}$ ), and the tensors ( $2^{++}$ ). The mesons in the multiplet have different quark compositions and are close in mass and in their properties. Thus, light mesons form multiplets of nine states each (nonets):

- $J^{PC} = 0^{-+} : (\pi, K, \eta, \eta')$
- $J^{PC} = 1^{--} : (\rho, K^*, \omega, \phi)$
- $J^{PC} = 1^{+-} : (b_1, K_1, h_1, h'_1)$

and so on.

Introducing the heavier charm and beauty quarks makes this scheme more involved. For instance, including the charm quark transforms the meson nonet into a 16-plet, which includes both mesons with open charm ( $D, D_s$ ) and charmonium states ( $c\bar{c}$ ), where the charm quantum number is hidden. Taking into account the  $b$  quark makes this picture even more complicated, though keeping the general principle untouched.

Mesons consisting of two heavy quarks – charmonium ( $c\bar{c}$ ) and bottomonium ( $b\bar{b}$ ) – deserve special attention. The large mass of quarks makes it possible to use various potential models or effective theories, like HQET or NRQCD, to calculate the spectrum of such a system. It turns out that these calculations are in excellent agreement with experiment, at least up to the production threshold of mesons with an open charm or beauty. This makes it possible to rely on a good understanding of the excited states of quarkonium in order to use them, for example, to verify calculations using lattice QCD, or to search for unconventional hadrons.

The mechanism for baryons is the same as for the mesons. Of course, for baryons, a much larger

number of combinations is obtained. In the ground state, only for light quarks, multiplets of 8 baryons (octet) with  $J^P = 1/2^+$  and 10 baryons (decuplet) with  $J^P = 3/2^+$  are formed. With the addition of heavy  $b$  or  $c$  quarks, the number of combinations increases even more. All baryons of the ground state multiplets are known. Many of their properties, in particular their masses, are in good agreement even with the most basic versions of the quark model. The picture for excited states is much less clear.

The conventional hadron spectra, derived from the quark model, make it possible to describe the properties of hundreds of hadrons and do not contradict the available experimental facts. However, today hadronic spectroscopy is as far as possible from drawing a conclusion about a complete understanding of the structure of hadrons and from limiting itself to a compilation of reference books with higher and higher precision. Fortunately, the existing mysteries in the spectrum of long-known particles and the discoveries made in recent decades do not give grounds for such conclusions and turn hadron spectroscopy into a rapidly developing and fascinating field of research.

#### 4 Exotic hadrons

The assumption that hadrons can be formed not only by combinations of two or three quarks but also by a larger number of them, was made by Gell-Mann in the very first work devoted to the quark model. Subsequently, in the framework of QCD, it was shown that the existence of hadrons including one or more valence gluons is possible. Such particles are called *hybrids*, and the ones consisting only of gluons are called *glueballs*. Naturally, such particles, called exotic ones, have been sought and tried to be discovered experimentally for more than 30 years. The main problem faced in this search by experimentalists concerns not so much the discovery itself of new particles, as the proof of their ‘exotic’ nature.

Indeed, how can we identify exotic hadrons? Unfortunately, there are not many distinct signatures. Essentially, there are only three main ways to go. First, we can search for particles with quantum numbers that are not compatible with the assumption of two- or three-quark composition. For instance, for mesons it could be  $J^{PC} = 0^{--}, 0^{+-}, 1^{-+}, 2^{+-}, 3^{-+}$ , etc. which cannot be constructed having only a quark and an antiquark.

Secondly, we can look for superfluous states. That is, those particles whose existence does not match theoretical predictions, or several particles found instead of one predicted. In the latter case, it will be necessary to find out which of the extra states is a conventional hadron and which is an exotic one. Unfortunately, we are on shaky ground here, as theoretical predictions are still not accurate enough, especially for excited states.

Finally, one can look for unexpected decay patterns. For example, in a decay  $X^+ \rightarrow J/\psi\pi^+$ , we need charm quark and antiquark to form charmonium, as well as at least two more light quarks to get the unit electric charge of a pion. That is, such decay can be interpreted as the decay of a multi-quark state of four or more quarks.

The lack of experimental findings over the decades has been exacerbated several times by reports of false observations. The most striking case was the report of the observation of a narrow pentaquark state  $\Theta^+(1530)$  in 2002. The first indication of a pentaquark in the  $nK^+$  system in the photoproduction on nuclei received more than ten independent confirmations within two years before it was proved that

this was an erroneous result [6]. All this, together with the undoubted successes of QCD in explaining the spectrum of conventional hadrons, gave every reason to doubt the existence of anything other than generally accepted mesons and baryons.

#### 4.1 Light hadrons: hunt for glueballs, search for diquarks

The relatively poor theoretical understanding of hadrons at low energies is not particularly surprising, since it is known that QCD becomes nonperturbative at low energies and does not allow one to calculate the spectrum of light hadrons. The progress of lattice QCD calculations is impressive and encouraging, but the results have not yet reached sufficient precision and reliability either. Combined with the experimental difficulties of studying broad mixing resonances, this explains why there are still quite a few unsolved problems in the spectrum of light hadrons. Let's consider some of them.

Among the light hadrons, the lightest  $f_0(500)$  meson, also known as the  $\sigma$  meson, attracts special attention. The existence and properties of this meson have been controversial for almost six decades. The existence of this meson was suggested in 1955, even before the quark model was formulated, to explain short-range nucleon-nucleon interactions by two-pion exchange. Experimental difficulties of studying the light and broad state in a  $\pi\pi$  system, together with the model-dependent interpretation, resulted in a highly uncertain mass range, from 400 to 1200 MeV, and a similarly large range, from 500 to 1000 MeV, for the width, for a long time. General agreement on the  $\sigma$  meson properties was reached only in the mid-2000s after extensive theoretical and experimental efforts. However, the composition of this resonance in terms of quarks and gluons is still a puzzle. It is well established that it cannot be interpreted as predominantly made of a quark and an antiquark. Whether it is a glueball or a tetraquark, there is a lot of controversies, but no firm conclusion is made yet. More details can be found in an excellent review paper, see Ref. [7]. Similar difficulties, theoretical and experimental, exist when considering the light strange meson,  $K^*(700)$  or  $\kappa$ , showing up as an S-wave in  $\pi K$  scattering. However, unlike the  $f_0(500)$ , they have not yet been finally resolved [8].

Among the light mesons, the search for possible glueballs and other exotic hadrons has been going on for a long time. In particular, among scalar mesons with masses less than 1 GeV, the mesons  $a_0(980)$  and  $f_0(980)$  are known, which have almost identical masses, just below the  $K\bar{K}$  production threshold, but different isospin. These two particles have long been considered as candidates for exotic hadrons – tetraquarks, meson molecules, etc. or a mixture of exotic and ordinary mesons. In the spectrum of light hadrons above 1 GeV, the same can be said about the  $f_0(1370)$ ,  $f_0(1500)$  and  $f_0(1710)$  mesons, at least one of which looks as a surplus state and possibly a glueball. This corresponds, among other things, to predictions by lattice QCD calculations that the mass of the lightest scalar glueball lays the range 1500 to 1800 MeV. Just below the  $p\bar{p}$  production threshold, the BES experiment observed the  $X(1835)$  state [9], later confirmed by BESIII in several different reactions, which is also rather difficult to interpret within the framework of the classical quark model. These are just some examples of the existing difficulties in interpreting the spectrum of light mesons.

As regards the light baryons, there is an opposite problem. There is a great inconsistency between theoretical predictions and the experimentally observed spectrum of baryons [10–12]. The number of experimentally discovered baryons is much smaller than those predicted. Of course, this can be ex-

plained by the lack of experimental data and the difficulty of their interpretation. Indeed, the number of established excited states of light baryons increases with time. An alternative explanation for the problem of missing baryons is that our ideas about the structure of baryons need to be refined. There is an assumption that two valence quarks can combine into a colored correlation, or diquark, which leads to a decrease in possible combinations in the construction of excited baryons. The experimental verification of this assumption is complicated, since diquarks are not a color singlet and, like quarks, cannot exist as free particles.

## 4.2 Heavy hadrons: multiquarks and other exotics

The revolution in the study of exotic hadrons began rather quietly, with the discovery of the  $X(3872)$  state. This state was found in the BELLE experiment and appeared as a clear peak in the spectrum of invariant masses of the system  $(\pi^+\pi^-J/\psi)$  during the study of the decay  $B^+ \rightarrow K^+\pi^+\pi^-J/\psi$ . The mass of the new state was  $M = 3872.0 \pm 0.6(\text{stat}) \pm 0.5(\text{syst})$  MeV and the width was narrow, less than 2.3 MeV at 90% C.L. This state did not fit into the conventional charmonium spectrum and became a vivid example of an extra, unexpected particle. Of course, a great many assumptions have been made about the nature of the  $X(3872)$  since then, but none of them was confirmed or rejected. After 20 years, we know more. The quantum numbers of the  $X(3872)$  have been measured:  $J^{PC} = 1^{++}$  [14]. The mass is suspiciously (within 1 MeV) close to the mass of a  $D^*D$  pair, but the comparable decay rate to  $D^*D$  and  $\gamma\psi(3686)$  tells us that the  $X(3872)$  is unlikely to be a  $D^*D$  bound state. The decay rates of  $X(3872)$  to  $\omega J/\psi$  and  $\rho J/\psi$  are approximately the same, which suggests a large violation of isospin. A charged partner of the  $X(3872)$  is not found. So, despite all these findings, we still do not know the nature of this mysterious state.

After the  $X(3872)$ , the epoch of charged charmonium-like and bottomonium-like states began. These are resonances decaying into charmonium or bottomonium and a charged meson, which hints at their multiquark nature. It was the BELLE experiment at the B-factory that made the first observations of states like these. Charged charmonium-like states,  $Z_c^\pm(4430)$ ,  $Z_c^\pm(4050)$ ,  $Z_c^\pm(4250)$ , and bottomonium-like states,  $Z_b^\pm(10610)$  and  $Z_b^\pm(10650)$ , have been reported [15–17]. However, this evidence was not confirmed by the BaBar experiment in similar conditions and remained contradictory. The first charged charmonium-like state beyond doubt was the  $Z_c^\pm(3900)$  in the BESIII experiment, discovered in 2013 and quickly confirmed by the BELLE and CLEO-c experiments [18]. Since then, the existence of a large number of exotic charmonium-like and bottomonium-like states has been firmly established, mainly in the BELLE, BESIII, and LHCb experiments [19]. Since the new states did not fit into the traditional charmonium and bottomonium spectra, they were named  $X$ ,  $Y$ , or  $Z$ , depending on their electrical charges and quantum numbers, and they were collectively referred to as the  $XYZ$  states.

Pentaquarks have been found in a similar way, in decays to charmonium and baryons. In 2015, at the LHCb experiment, the  $P_c^+(4450)$  and  $P_c^+(4380)$  pentaquarks were discovered in  $\Lambda_b^0 \rightarrow J/\psi p K^-$  decays in the invariant mass spectrum of the  $(J/\psi p)$  system [20]. Later, also at the LHCb experiment, when analyzing samples of increased statistics, another pentaquark,  $P_c^+(4312)$ , was found, and the pentaquark  $P_c^+(4450)$  was reinterpreted as two separate states,  $P_c^+(4440)$  and  $P_c^+(4457)$  [21]. In addition to pentaquarks, more than 10 candidates for other exotic states, including two possible tetraquarks, and about 50 excited states of ordinary mesons and baryons were discovered at the LHC [22].

Thus, the existence of a large number of hadrons that are not mesons nor baryons is now firmly established. They are not ‘exotic’ any more. No doubt that the number of such hadrons will increase in the coming years, both among heavy and light hadrons. Discussions whether these particles are multiquarks, diquark molecules, hybrids or some more complicated structures are still ongoing and it will probably take a long time before the nature of the new states becomes clear. However, we can say right now that these discoveries once again confirmed the validity of the quark model and provided rich material for refining our understanding of hadronic spectra. Ultimately, the discovery of new hadronic states does not simply aim at adding the properties of more particle to the set of particles known to us. The main result is the verification and refinement of the theory that describes the properties of these particles, based on fundamental concepts of nuclear matter, and a deeper understanding of the laws of strong interaction.

## 5 Where do we stand?

Currently, the data accumulated in the B factory experiments (Belle, BaBar), the  $\tau$ -charm experiment (BESIII), and the LHC experiments (LHCb, ATLAS, CMS) are the main source of knowledge about new hadrons. In 2017, the GlueX experiment began, aimed at high-sensitivity searches for pentaquarks and other exotic mesons in photoproduction processes. The Belle II experiment started data taking in 2019 after a major upgrade. There is no doubt that it is the Belle II, BESIII, GlueX, LHCb and other experiments at the LHC that will determine the future of hadron spectroscopy in the next 10–15 years. In addition, some new experiments are expected to join exotic hadron research efforts in the near future: the PANDA experiment in Darmstadt, Germany, the AMBER experiment at CERN, the  $\tau$ -charm Super Factories STCF in Hefei, China, and the SCTF in Sarov, Russia. Later, experiments at the planned electron-ion collider EIC might also contribute a lot to our understanding of baryon structure and hadron spectra.

As a final note, we can add that, throughout the history of particle physics, hadron spectroscopy has repeatedly led to significant changes in our understanding of the structure of matter. It remains a unique tool for gaining knowledge about the intrinsic properties and composition of hadrons today. There is no doubt that the discovery of a large number of exotic, supposedly multiquark states is one of the most exciting events in hadron physics recently, perhaps comparable to the discovery of the  $J/\psi$  meson. To understand the nature of exotic states, and to solve the problem of extra mesons and missing baryons, both more accurate and high-quality experimental data and new theoretical insights are needed. Special hopes are placed here on lattice QCD, which is making impressive progress; but there is still a long way to go before constructing a theory that fully describes hadronic spectra.

## References

- [1] R.H. Dalitz, On the analysis of  $\tau$ -meson data and the nature of the  $\tau$ -meson, *Phil. Mag. Ser. 7* **44** (1953) 1068–1080, [doi:10.1080/14786441008520365](https://doi.org/10.1080/14786441008520365).
- [2] M. Aghasyan *et al.* (COMPASS Collaboration), Search for muoproduction of  $X(3872)$  at COMPASS and indication of a new state  $\tilde{X}(3872)$ , *Phys. Lett B* **783** (2018) 334–340, [doi:10.1016/j.physletb.2018.07.008](https://doi.org/10.1016/j.physletb.2018.07.008).

- [3] M. Gell-Mann, A schematic model of baryons and mesons, *Phys. Lett.* **8** (1964) 214–215, doi:10.1016/S0031-9163(64)92001-3.
- [4] G. Zweig, An  $SU_3$  model for strong interaction symmetry and its breaking, CERN-TH-401 (CERN, Geneva, 1964), doi:10.17181/CERN-TH-401.
- [5] S.L. Olsen, T. Skwarnicki and D. Zieminska, Nonstandard heavy mesons and baryons: Experimental evidence, *Rev. Mod. Phys.* **90** (2018) 015003, doi:10.1103/RevModPhys.90.015003.
- [6] R.A. Schumacher, The rise and fall of pentaquarks in experiments, *AIP Conf. Proc.* **842** (2006) 409–417, doi:10.1063/1.2220285.
- [7] J. Peláez, From controversy to precision on the sigma meson: a review on the status of the non-ordinary  $f_0(500)$  resonance, *Phys. Rept.* **658** (2016) 1–111, doi:10.1016/j.physrep.2016.09.001.
- [8] J. Peláez and A. Rodas, Dispersive  $\pi K \rightarrow \pi K$  and  $\pi\pi \rightarrow K\bar{K}$  amplitudes from scattering data, threshold parameters, and the lightest strange resonance  $\kappa$  or  $K^*(700)$ , *Phys. Rept.* **969** (2022) 1–126, doi:10.1016/j.physrep.2022.03.004.
- [9] J.Z. Bai *et al.* (BES Collaboration), Observation of a near-threshold enhancement in the  $p\bar{p}$  mass spectrum from radiative  $J/\psi \rightarrow \gamma p\bar{p}$  decay, *Phys. Rev. Lett.* **91** (2003) 022001, doi:10.1103/PhysRevLett.91.022001.
- [10] U. Loring, B.C. Metsch and H.R. Petry, The Light baryon spectrum in a relativistic quark model with instanton induced quark forces: The Nonstrange baryon spectrum and ground states, *Eur. Phys. J. A* **10** (2001) 395–446, doi:10.1007/s100500170105.
- [11] R.G. Edwards *et al.*, Excited state baryon spectroscopy from lattice QCD, *Phys. Rev. D* **84** (2011) 074508, doi:10.1103/PhysRevD.84.074508.
- [12] G. Eichmann and C.S. Fischer, Baryon structure and reactions, *Few Body Syst.* **60** (2019) 2, doi:10.1007/s00601-018-1469-5.
- [13] S.-K. Choi *et al.* (Belle Collaboration), Observation of a narrow charmonium-like state in exclusive  $B^\pm \rightarrow K^\pm \pi^+ \pi^- J/\psi$  decays, *Phys. Rev. Lett.* **91** (2003) 262001, doi:10.1103/PhysRevLett.91.262001.
- [14] R. Aaij *et al.* (LHCb Collaboration), Determination of the X(3872) meson quantum numbers, *Phys. Rev. Lett.* **110** (2013) 22001, doi:10.1103/PhysRevLett.110.222001.
- [15] S.-K. Choi *et al.* (Belle Collaboration), Observation of a resonance-like structure in the  $\pi^\pm \psi'$  mass distribution in exclusive  $B \rightarrow K \pi^\pm \psi'$  decays, *Phys. Rev. Lett.* **100** (2008) 142001, doi:10.1103/PhysRevLett.100.142001.
- [16] R. Mizuk *et al.* (Belle Collaboration) Observation of two resonance-like structures in the  $\pi^+ \chi_{c1}$  mass distribution in exclusive  $\bar{B}^0 \rightarrow K^- \pi^+ \chi_{c1}$  decays, *Phys. Rev. D* **78** (2008) 072004, doi:10.1103/PhysRevD.78.072004.
- [17] A. Bondar *et al.* (Belle Collaboration), Observation of two charged bottomonium-like resonances in  $\Upsilon(5S)$  decays, *Phys. Rev. Lett.* **108** 122001, doi:10.1103/PhysRevLett.108.122001.

- [18] M. Ablikim *et al.* (BESIII Collaboration), Observation of a charged charmonium-like structure in  $e^+e^- \rightarrow \pi^+\pi^- J/\psi$  at  $\sqrt{s} = 4.26$  GeV, *Phys. Rev. Lett.* **110** (2013) 252001, [doi:10.1103/PhysRevLett.110.252001](https://doi.org/10.1103/PhysRevLett.110.252001).
- [19] N. Brambilla *et al.*, The XYZ states: Experimental and theoretical status and perspectives, *Phys. Rep.* **873** (2020) 1–154, [doi:10.1016/j.physrep.2020.05.001](https://doi.org/10.1016/j.physrep.2020.05.001).
- [20] R. Aaij *et al.* (LHCb Collaboration), Observation of  $J/\psi p$  resonances consistent with pentaquark states in  $\Lambda_b^0 \rightarrow J/\psi K^- p$  decays, *Phys. Rev. Lett.* **115** (2015) 072001, [doi:10.1103/PhysRevLett.115.072001](https://doi.org/10.1103/PhysRevLett.115.072001).
- [21] R. Aaij *et al.* (LHCb Collaboration), Observation of a narrow pentaquark state,  $P_c(4312)^+$ , and of the two-peak structure of the  $P_c(4450)^+$ , *Phys. Rev. Lett.* **122** (2019) 222001, [doi:10.1103/PhysRevLett.122.222001](https://doi.org/10.1103/PhysRevLett.122.222001).
- [22] T. Gershon *et al.* (LHCb Collaboration), Exotic hadron naming convention, CERN-LHCb-PUB-2022-013 (CERN, Geneva, 2022), [doi:10.17181/CERN.7XZO.HPH7](https://doi.org/10.17181/CERN.7XZO.HPH7).



---

## Scientific programme<sup>8</sup>

### Neutrino physics

*Sin Kyu Kang, Seoul Tech., South Korea*

### Statistical techniques

*Nicolas Berger, LAPP, France*

### Field theory and the EW Standard Model

*Rohini Godbole, IISc, Bangalore, India*

### Super Kamiokande

*Takaaki Kajita, U. Tokyo, Japan*

### Q&A with Director-General of CERN

*Fabiola Gianotti, CERN*

### Higgs and BSM physics

*Ryuichiro Kitano, KEK, Japan*

### Hadron spectroscopy

*Alexei Zhemchugov, JINR*

### Instrumentation

*Ingrid-Maria Gregor, DESY, Germany*

### High-energy physics outlook

*Young-Kee Kim, U. of Chicago, USA*

### Cosmology and dark matter

*Hitoshi Murayama, IPMU, Japan*

### Heavy-ion physics

*Yen-Jie Lee, MIT, USA*

### Flavour physics and CP violation

*Xiao-Gang He, SJTU/NTU, Taiwan*

### QCD

*Vajravelu Ravindran, IMSc, Chennai, Pakistan*

### Prospects at LHC in Run 3 and HL-LHC

*Sarah Demers, Yale, USA*

### Gravitational waves

*Jo van den Brand, Nikhef, The Netherlands*

---

<sup>8</sup>Slides available at <https://indico.cern.ch/event/884244/>.

---

## Organizing committees

### International advisory committee

Alexander Bondar (BINP)  
Amol Dighe (TIFR)  
Kazunori Hanagaki (KEK, Chair)  
Beate Heinemann (DESY)  
Hafeez Hoorani (NCP)  
Martin Seviar (Melbourne)  
Xiaoyan Shen (IHEP)  
Dongchul Son (KNU)  
Patrice Verdier (IN2P3-CNRS)  
Rüdiger Voss (CERN)  
Henry Tsz-King Wong (Academia Sinica)

### International organizing committee

Ashfaq Ahmad (NCP)  
Elisabetta Barberio (Melbourne)  
Marc Besancon (CEA/Irfu)  
Nick Ellis (CERN)  
Yee Hsiung (National Taiwan University)  
Kiyotomo Kawagoe (Kyushu)  
Pyungwon Ko (KIAS)  
Gobinda Majumdar (TIFR)  
Bedanga Mohanty (NISER)  
Martijn Mulders (CERN, Chair)  
Shohei Nishida (KEK)  
Lydia Roos (IN2P3/LPNHE)  
Thomas Schoerner-Sadenius (DESY)  
Boris Shwartz (BINP)  
Changzheng Yuan (IHEP)  
Shilin Zhu (PKU)

### Local organizing committee

Sanghyeon Chang (IBS CTPU)  
Byunggu Cheon (Hanyang University)  
Suyong Choi (Korea University)  
Byungsik Hong (Korea University)  
Donghee Kim (Kyungpook National University)  
Hyunsoo Kim (Sejong University)  
Tae Jeong Kim (Hanyang University, Co-chair)  
Pyungwon Ko (Korea Institute for Advanced Study, Co-chair)  
Youngjoon Kwon (Yosei University)  
Jason Lee (University of Seoul)  
Chang-Seong Moon (Kyungpook National University)  
Dongho Moon (Chonnam National University)  
Inkyu Park (University of Seoul)  
Dongchul Son (Kyungpook National University)  
Minho Son (KAIST)  
Unki Yang (Seoul National University)  
Jaehyeok Yoo (Korea University)  
Intae Yu (Sungkyunkwan University)

---

## List of lecturers

Nicolas Berger, LAPP, France  
Jo van den Brand, Nikhef, The Netherlands  
Sarah Demers, Yale, USA  
Fabiola Gianotti, CERN  
Rohini Godbole, IISc, Bangalore, India  
Ingrid-Maria Gregor, DESY, Germany  
Xiao-Gang He, SJTU/NTU, Taiwan  
Takaaki Kajita, U. Tokyo, Japan  
Sin Kyu Kang, Seoul Tech., South Korea  
Young-Kee Kim, U. of Chicago, USA  
Ryuichiro Kitano, KEK, Japan  
Yen-Jie Lee, MIT, USA  
Hitoshi Murayama, IPMU, Japan  
Vajravelu Ravindran, IMSc, Chennai, Pakistan  
Alexei Zhemchugov, JINR

## List of discussion leaders

Faisal Akram (CHEP, Punjab U., Pakistan)  
Yu Seon Jeong (Chung-Ang U., South Korea)  
Suchita Kulkarni (U. of Graz, Austria)  
John Gargalionis (IFIC, Spain)  
Hua-Sheng Shao (LPTHE, France)  
Takahiro Terada (IBS, Japan)

---

## List of students

Malak AIT TAMLIHAT  
Atakan Tugberk AKMETE  
Shiwen AN  
Stefio Yosse ANDREAN  
Ryan Justin ATKIN  
Irene BACHILLER PEREA  
Sweta BARADIA  
Aashwin BASNET  
Aya BESHR  
Fionn BISHOP  
Francesco CARNEVALI  
Vilius CEPAITIS  
Wai Yuen CHAN  
Ashish DAHAL  
Sudipta DAS  
Nicola DE BIASE  
Lindamulage Malinda Shiram DE SILVA  
Olexiy DVORNIKOV  
José Antonio FERNÁNDEZ PRETEL  
Eszter FRAJNA  
Guangyong FU  
Quentin FÜHRING  
Tommaso FULGHESU  
Gabor GALGOCZI  
Simran Sunil GURDASANI  
Laszlo GYULAI  
Mohammadmehdi HAJIMAGHSOUD  
SAGAR HAZRA  
Mourad HIDAOU  
Benjamin HODKINSON  
Jieun HONG  
Moe ISSHIKI  
Nour JALAL ABDULAMEER  
Pranati JANA  
Dominic JONES  
Amandeep KAUR  
Harjot KAUR  
Niamat Ullah KHAN  
Elham KHAZAIE  
Honey KHINDRI  
Jinyoung KIM  
Jihun KIM  
JINHEUNG KIM  
Jiwoong KIM  
Yongkyu KIM  
Giorgi KISTAURI  
Anežka KLUSTOVÁ  
Byeonghak KO  
Joon-Bin LEE  
SooJin LEE  
Younghoon LEE  
Yunjae LEE  
Xiaowen LI  
Geliang LIU  
Kalpanie Madara LIYANAGE  
Matteo MARCHEGIANI  
Lazar MARKOVIC  
Daniel MARTIN  
Michael MEWS  
Marcos MIRALLES LOPEZ  
Poulami MONDAL  
Muhammad NUMAN ANWAR  
David MUNOZ PEREZ  
Madeeha NAZISH  
Dang Bao Nhi NGUYEN  
Lukas NOVOTNY  
Jayashri PADMANABAN  
Alessandra PALAZZO  
Tong PAN  
Papia PANDA  
Hanseo PARK  
Laura PEREIRA SANCHEZ  
Jonas RÜBENACH  
YeonSu RYOU  
Ritik SAXENA  
Hyonsan SEO  
Upasana SHARMA  
Thanaporn SICHANUGRIST  
Dinesh Kumar SINGHA  
Elizaveta SITNIKOVA  
Youngwan SON  
Juhee SONG  
Rongrong SONG  
Lakshmi Priya SREELATHA PRAMOD  
Henrikas SVIDRAS  
Bongho TAE  
Aravind THACHAYATH SUGUNAN  
Shusaku TSUMURA  
Vismaya V S  
Elena VERNAZZA  
Zebing WANG  
Chu WANG  
Cheng-Han WU  
Sze Chun YIU  
Chaochen YUAN  
Jing ZHAO

---

## List of posters

Poster title	Presenter
Branching Ratios of Semileptonic Charged Kaon Decays	AKMETE, A. T.
Silicon Sensor Qualification for the Outer Tracker of CMS	ANWAR, M. N.
Search for $B_c^+$ decays to two charm mesons	BISHOP, F.
A deep neural network to search for new long-lived particle states decaying to jets	CEPAITIS, V.
Application of quantum computing techniques in particle tracking at LHC	CHAN, W. Y.
Clustering and tracking in dense hadronic environments with the ITk	DE BIASE, N.
SiPM-on-Tile Technology for the Phase II upgrade of the CMS High Granularity Endcap Calorimeter (HGCALE)	DE SILVA, M.
Precision measurement of fiducial and differential cross sections of WW production with the ATLAS Detector	FERNÁNDEZ PRETEL, J. A.
NLO Calculations of $D - \bar{D}$ Azimuthal Correlations with HERWIG	FRAJNA, E.
Search for new physics in the tt+MET final state	GURDASANI, S.
Defining the Underlying-Event Activity in the Presence of Heavy-Flavour Processes in Proton-Proton Collisions at LHC Energies	GYULAI, L.
Measurement of the photon-energy spectrum in inclusive $B \rightarrow X_s \gamma$ decays identified using hadronic decays of the recoil $B$ meson at Belle II	SVIDRAS, H.
METNet: A combined missing transverse momentum working point using a neural network with the ATLAS detector	HODKINSON, B.
FPGA based firmware implementation of MET algorithm for CMS Phase-2 L1 trigger	HONG, J.
Study of $\tau$ lepton pair production in the $e\mu$ final state using ultraperipheral nucleus-nucleus collisions	JANA, P.

---

<b>Poster title</b>	<b>Presenter</b>
Measurement of Higgs boson production in association with a W or Z boson using the full Run 2 dataset at the CMS	KAUR, A.
SUSY Searches via VBF in 1-lepton final states using LHC Run 2 data collected by the CMS detector	KAUR, H.
Search for $H \rightarrow a_1 a_1 \rightarrow \mu\mu bb$ at 13 TeV	KHAZAIE, E.
Dark photon search using $B \rightarrow Kl^+l^-l^+l^-$ decay at Belle	KIM, Y.-K.
Measurement of CP violation in single top t-channel production at 13 TeV	KO, B.
Particle identification for dual-readout calorimeter	LEE, Y.
Search for High-mass Resonances Decaying into Muon Pairs in pp Collisions at $\sqrt{s} = 13$ TeV with the Full Run2 Data Collected by the CMS Detector	LIYANAGE, K.
Hierarchy problem and dimension-6 effective operators	MONDAL, P.
Measurement of pulse arrival time at CMS HCAL by using Cross-Correlation	PADMANABAN, J.
DAQ and DCS development for the ATLAS Inner Pixel Tracker and search for new physics in diboson production	PALAZZO, A.
Extracting the best physics sensitivity from T2HKK: a study on optimal detector volume at Japan and Korea	PANDA, P.
Observing Axion Emission from Supernova with Collider Detectors	SICHANUGRIST, T.
3+1 sterile neutrino study of P2O	SINGHA, D. K.
Search for heavy resonances in four-top-quark final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector	SITNIKOVA, E.
Identification of additional b jets in ttbb production in the lep+jet channel using DNN	SONG, J.
Search for lepton flavor violation in $B \rightarrow K^* \tau^\pm e^\mp$ decays	FULGHESU, T.
Charged Particle Identification with the TOP detector at Belle II	V S, VISMAYA
Irradiation testing of HGCROC3, the front-end readout ASIC for HGCALE	VERNAZZA, E.

---

<b>Poster title</b>	<b>Presenter</b>
TeV gamma-ray analysis of eHWC J1842-035 region with HAWC observatory	SON, Y.
Impact of increased jet $p_T$ cuts on Simplified Template Cross Section measurements in $VH \rightarrow b\bar{b}$ resolved	ATKIN, R. J.
Prospects of triggering $H/Z \rightarrow \phi\gamma$ and $H/Z \rightarrow \rho\gamma$ at $\sqrt{s} = 14\text{TeV}$ at the High Luminosity LHC (HL-LHC)	BARADIA, S.
Study of fully leptonic WZgamma production	KIM J.-W.
Decay-time dependent analysis of $B_s^0 \rightarrow D_s^\mp h^\pm$ decays at the LHCb experiment	FUEHRING, Q.
Search for electroweak production of photon+jets	HAJIMAGHSOUD, M.
Effect of magnetic field variation on the precision measurements at ICAL: Simulation study	HONEY, K.
Simulations and Magnetic field measurements on mini-ICAL	HONEY, K.
The KM3NeT Project	KISTAURI, G.
Measurement of Nuclear Dependence in Inclusive Antineutrino Scattering with MINERvA	KLUSTOVÁ, A.
Itinerary to New Physics with Non-Isolated Muons Inside a b-Jet	LEE, J.-B.
Experience and Results of the 2021 HGCAL Beam Test Analysis	LIU, G.
Search for the leptonic charge asymmetry in ttW in final states with three charged leptons	MIRALLES LÓPEZ, M.
Improved track reconstruction for prompt and long-lived particles in ATLAS for the LHC Run 3	MUNOZ PEREZ, D.
Measuring CP properties of Higgs boson interactions with $\tau$ leptons with the ATLAS detector	TONG, P.
Development of ILC shower clustering algorithm using GNN	TSUMURA, S.
Electron identification and energy calibration in search for $H \rightarrow \gamma * \gamma \rightarrow ee\gamma$ at CMS	WU, C.-H.

