

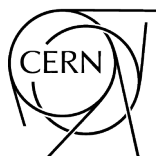
# **Proceedings of the 2019 CERN–Latin-American School of High-Energy Physics**

Villa General Belgrano, Argentina, 13–26 March 2019

Editors:


C. Duhr

M. Mulders



ISBN 978-92-9083-604-9 (Print)  
ISBN 978-92-9083-605-6 (Online)  
ISSN 2519-8041 (Print)  
ISSN 2519-805X (Online)  
DOI <https://doi.org/10.23730/CYRSP-2021-002>

Copyright © CERN, 2021

 Creative Commons Attribution 4.0

This report should be cited as:

Proceedings of the 2019 CERN–Latin-American School of High-Energy Physics, Villa General Belgrano, Argentina, 13–26 March 2019, edited by C. Duhr and M. Mulders, CERN-2021-002 (CERN, Geneva, 2021), <https://doi.org/10.23730/CYRSP-2021-002>

A contribution in this report should be cited as:

[Author name(s)], in Proceedings of the 2019 CERN–Latin-American School of High-Energy Physics, Villa General Belgrano, Argentina, 13–26 March 2019, edited by C. Duhr and M. Mulders, CERN-2021-002 (CERN, Geneva, 2021), pp. [first page]–[last page], <https://doi.org/10.23730/CYRSP-2021-002>. [first page]

Corresponding editors: [Claude.Duhr@cern.ch](mailto:Claude.Duhr@cern.ch) and [Martijn.Mulders@cern.ch](mailto:Martijn.Mulders@cern.ch)

Accepted by the [CERN Reports Editorial Board](#) (contact [Carlos.Lourenco@cern.ch](mailto:Carlos.Lourenco@cern.ch)) in May 2021.

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

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

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

## **Abstract**

The CERN–Latin-American School of High-Energy Physics is intended to give young physicists an introduction to the theoretical aspects of recent advances in elementary particle physics. These proceedings contain lecture notes on quantum field theory, flavour dynamics and CP violation, heavy-ion physics, LHC highlights and prospects, and experimental facilities in Latin America.



## Preface

The tenth Event in the series of CERN–Latin-American Schools of High-Energy Physics took place from 13 to 26 March 2019 in Villa General Belgrano, Argentina. It was organized by CERN with the support of Argentinian colleagues from Comisión Nacional de Energía Atómica, Instituto Balseiro, Universidad de Buenos Aires, Universidad Nacional de Córdoba, Universidad Nacional de La Plata, Universidad Nacional de Mar del Plata, and Universidad de la República, Uruguay.

The School received financial support from: CBPF, Brazil, CERN, CIEMAT, Spain, CONICET, Argentina and SMF, Mexico. Financial and in-kind contributions were also received from the following institutes and universities: Agencia Nacional de Promoción Científica y Tecnológica, Comisión Nacional de Energía Atómica, Universidad Nacional de Córdoba and Universidad Nacional de La Plata.

Our sincere thanks go to all of the sponsors for making it possible to organize the School with a large number of young participants from Latin-American countries, many of whom would otherwise not have been able to attend. We would particularly like to thank the Universidad Nacional de Córdoba for their important in-kind contributions and practical help.

The School was hosted in the Hotel Howard Johnson in Villa General Belgrano. We are indebted to the hotel and its friendly staff for their help in making the Event such a success. In particular, we would like to mention Juan Manuel Laurini and Lucas Emmanuel Medina who helped us greatly in preparing the School, as well as during the Event itself. The excellent meals served in the Restaurante los Personajes at the hotel were highly appreciated, and special thanks are due to the chef, Aida Ieno Pose. Thanks also to Brunnen beer and Bodega Alfredo Roca for supplying drinks during the poster session and the final night dinner.

Tere Dova from UNLP acted as local director for the School, assisted by members of the local organising committee. We are extremely grateful to Tere and her colleagues for their excellent work in organizing the School and for creating such a wonderful atmosphere for the participants.

Seventy-five students of 17 different nationalities attended the School, including 21 from Argentina. Following the tradition of the School the students shared twin rooms mixing nationalities, and in particular the Europeans mixed with Latin Americans.

The 13 lecturers came from Europe, Latin America and the USA. The lectures, which were given in English, were complemented by daily discussion sessions led by five physicists coming from Latin America. The lectures and the discussion sessions were all held using the conference facilities of the hotel. The students displayed their own research work in the form of posters in a special evening session during the first week. The posters were left on display until the end of the School. The students from each discussion group also performed a project, studying in detail the analysis of a published paper from an LHC experiment. A representative of each group gave a brief summary talk during a special evening session during the second week of the School.

Our thanks are due to the lecturers and discussion leaders for their active participation in the School and for making the scientific programme so stimulating. The students who in turn manifested their good spirits during two intense weeks undoubtedly appreciated their personal contributions in answering questions and explaining points of theory. During the school two public lecturers were organised at the University of Cordoba, given by Tere Dova and Juan Martín Maldacena.

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

The participants will certainly remember the two interesting excursions: an afternoon visit to the nearby town of la Cumbrecita, followed by dinner at a local restaurant; and a full-day excursion to the city of Córdoba. They also greatly appreciated evenings spent together in the hotel, especially the farewell party on the last night. The success of the School was to a large extent due to the students themselves. Their poster session and group projects were very well prepared and highly appreciated, and throughout the School they participated actively during the lectures, in the discussion sessions, and in the different activities and excursions.

Nick Ellis  
(On behalf of the Organizing Committee)







## People in the photograph

1	Nick Ellis	46	Elisabetta Spadaro Norella
2	Gustavo Monti	47	Dario Alberto Ramirez Zaldivar
3	Iskya Garcia	48	Paulina Knees
4	Jorge Luis Valdes Albuernes	49	Omar Suarez Navarro
5	Tere Dova	50	Mariana Toscani
6	Ana Martina Botti	51	Tulio Caviquioli
7	Roger J. Hernandez Pinto	52	Lewis Wilkins
8	Juan Carlos Vasquez	53	Julian Schmoeckel
9	Marcela Marin	54	Emily Takeva
10	Gabriel de la Fuente Rosales	55	Fernando Augusto Assunção Neto
11	Andres Borquez	56	Antonio Pich
12	Ana Maria De Sousa Slivar	57	Érison Rocha
13	Juan Carlos Carrasco Martinez	58	Kate Ross
14	Gonzalo Orellana	59	Jefferson Gonçalves
15	Mykola Khandoga	60	Rosa María Sandá Seoane
16	Nicolas Neill	61	Agustín Carulli
17	Julián Calle	62	Alice Morris
18	Rafaela Wiklich Sobrinho	63	Edyvania Emily Pereira Martins
19	Rogério Rosenfeld	64	Daniel Ernani Martins Neto
20	Dener De Souza Lemos	65	Santiago Paredes Saenz
21	Andres Daniel Perez	66	Flavia Gesualdi
22	Antonio Radic	67	Iván Pedron
23	Martijn Mulders	68	Moises Leon
24	Edward Thorpe	69	Rafael Nunez
25	Ignacio Borsa Sanjuan	70	Manuel Szewc
26	Marina Scornavacche	71	Federico Lamagna
27	Adrian Alvarez Fernandez	72	Alison Tully
28	Yufeng Wang	73	Fernanda Gonçalves Abrantes
29	Breno Orzari	74	Gabriel Augusto das Neves
30	Joel Jones-Perez	75	Gabriel Damián Zapata
31	Rolando Calderón Ardila	76	Alan Hernández
32	Carlos Cocha	77	Carlos Díaz
33	Gabriel Ayala	78	Natalia Tapia
34	Paul Asmuss	79	Matthew Heath
35	Brenda Quezadas Vivian	80	Cristobal Morales
36	Mario Aldair Perez de Leon	81	Diego Domingues Lopes
37	Manuel Der	82	Daniel de Florian
38	Máximo Coppola	83	Alejandro De Yta
39	Nicolas Bernal	84	Indira Vergara Quispe
40	Maria Florencia Daneri	85	Isabel Astrid Goos
41	Emma Gonzalez Hernandez	86	Miguel Alves Gallo Pereira
42	Daniel Aguilar	87	Marco Ayala Torres
43	Téssio de Melo	88	Malena Tejeda-Yeomans
44	Felipe Figueroa	89	Dwayne Spiteri
45	Birgit Stapf		

# Photographs (montage)



**10<sup>th</sup> CERN LATIN-AMERICAN SCHOOL OF HIGH-ENERGY PHYSICS**

**Villa General Belgrano, Córdoba, 13–26 March 2019**







## Contents

Preface	
<i>N. Ellis</i> .....	v
Photograph of participants .....	vii
Photographs (montage) .....	x
Introductory lectures on quantum field theory	
<i>L. Alvarez-Gaumé and M. A. Vázquez-Mozo</i> .....	1
Flavour dynamics and violations of CP symmetry	
<i>A. Pich</i> .....	95
Heavy-ion physics: freedom to do hot, dense, exciting QCD	
<i>M. E. Tejeda-Yeomans</i> .....	137
LHC highlights and prospects	
<i>C. E. Gerber</i> .....	161
Experimental facilities in Latin America	
<i>C. O. Dib</i> .....	193
Scientific programme .....	221
Organizing committees .....	222
Local organizing committee .....	222
List of lecturers .....	222
List of discussion leaders .....	223
List of students .....	224
List of posters .....	225



# Introductory Lectures on Quantum Field Theory

*L. Álvarez-Gaumé<sup>a</sup> and M. A. Vázquez-Mozo<sup>b</sup>*

<sup>a</sup> CERN, Geneva, Switzerland

<sup>b</sup> Universidad de Salamanca, Salamanca, Spain

## Abstract

In these lectures we present a few topics in quantum field theory in detail. Some of them are conceptual and some more practical. They have been selected because they appear frequently in current applications to particle physics and string theory.

## 1 Introduction

These notes summarize lectures presented at the 2005 CERN-CLAF School in Malargüe (Argentina), the 2009 CERN-CLAF School in Medellín (Colombia), the 2011 CERN-CLAF School in Natal (Brazil), the 2012 Asia-Europe-Pacific School of High Energy Physics in Fukuoka (Japan), and the CERN–Latin-American Schools of High-Energy Physics in Arequipa (Peru) in 2013, and in Villa General Belgrano (Argentina) in 2019. The audience in all occasions was composed to a large extent by students in experimental High Energy Physics with an important minority of theorists. These lectures were intended to provide a review of those parts of the subject to be used later by other lecturers. Although a cursory acquaintance with the subject of quantum field theory is helpful, the only requirement to follow the lectures is a working knowledge of Quantum Mechanics and Special Relativity.

The guiding principle in choosing the topics presented (apart to serve as introductions to later courses) was to present some basic aspects of the theory that present conceptual subtleties. Those topics one often is uncomfortable with after a first introduction to the subject. Among them we have selected:

- The need to introduce quantum fields, with the great complexity this implies.
- Quantization of gauge theories and the rôle of topology in quantum phenomena. We have included a brief study of the Aharonov-Bohm effect and Dirac's explanation of the quantization of the electric charge in terms of magnetic monopoles.
- Quantum aspects of global and gauge symmetries and their breaking.
- Anomalies.
- The physical idea behind the process of renormalization of quantum field theories.
- Some more specialized topics, like the creation of particle by classical fields and the very basics of supersymmetry.

These notes have been written following closely the original presentation, with numerous clarifications. Sometimes the treatment given to some subjects has been extended, in particular the discussion of the Casimir effect and particle creation by classical backgrounds. Since no group theory was assumed, we have included an Appendix with a review of the basics concepts.

By lack of space and purpose, few proofs have been included. Instead, very often we illustrate a concept or property by describing a physical situation where it arises. A very much expanded version of these lectures, following the same philosophy but including many other topics, has appeared in book form in [1]. For full details and proofs we refer the reader to the many textbooks in the subject, and in particular in the ones provided in the bibliography [2–11]. Specially modern presentations, very much in the spirit of these lectures, can be found in references [5, 6, 10, 11]. We should nevertheless warn the reader that we have been a bit cavalier about references. Our aim has been to provide mostly a (not exhaustive) list of reference for further reading. We apologize to those authors who feel misrepresented.

### A note about notation

Before starting it is convenient to review the notation used. Through these notes we will be using the metric  $\eta_{\mu\nu} = \text{diag}(1, -1, -1, -1)$ . Derivatives with respect to the four-vector  $x^\mu = (ct, \vec{x})$  will be denoted by the shorthand

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = \left( \frac{1}{c} \frac{\partial}{\partial t}, \vec{\nabla} \right). \quad (1)$$

As usual space-time indices will be labelled by Greek letters ( $\mu, \nu, \dots = 0, 1, 2, 3$ ) while Latin indices will be used for spatial directions ( $i, j, \dots = 1, 2, 3$ ). In many expressions we will use the notation  $\sigma^\mu = (\mathbf{1}, \sigma^i)$  where  $\sigma^i$  are the Pauli matrices

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2)$$

Sometimes we use of the Feynman's slash notation  $\not{x} = \gamma^\mu a_\mu$ . Finally, unless stated otherwise, we work in natural units  $\hbar = c = 1$ .

## 2 Why do we need quantum field theory after all?

In spite of the impressive success of Quantum Mechanics in describing atomic physics, it was immediately clear after its formulation that its relativistic extension was not free of difficulties. These problems were clear already to Schrödinger, whose first guess for a wave equation of a free relativistic particle was the Klein-Gordon equation

$$\left( \frac{\partial^2}{\partial t^2} - \nabla^2 + m^2 \right) \psi(t, \vec{x}) = 0. \quad (3)$$

This equation follows directly from the relativistic “mass-shell” identity  $E^2 = \vec{p}^2 + m^2$  using the correspondence principle

$$\begin{aligned} E &\rightarrow i \frac{\partial}{\partial t}, \\ \vec{p} &\rightarrow -i \vec{\nabla}. \end{aligned} \quad (4)$$

Plane wave solutions to the wave equation (3) are readily obtained

$$\psi(t, \vec{x}) = e^{-ip_\mu x^\mu} = e^{-iEt + i\vec{p}\cdot\vec{x}} \quad \text{with} \quad E = \pm\omega_p \equiv \pm\sqrt{\vec{p}^2 + m^2}. \quad (5)$$

In order to have a complete basis of functions, one must include plane wave with both  $E > 0$  and  $E < 0$ . This implies that given the conserved current

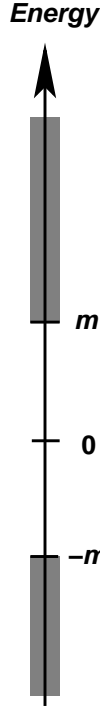
$$j_\mu = \frac{i}{2} \left( \psi^* \partial_\mu \psi - \partial_\mu \psi^* \psi \right), \quad (6)$$

its time-component is  $j^0 = E$  and therefore does not define a positive-definite probability density.

A complete, properly normalized, continuous basis of solutions of the Klein-Gordon equation (3) labelled by the momentum  $\vec{p}$  can be defined as

$$\begin{aligned} f_p(t, \vec{x}) &= \frac{1}{(2\pi)^{\frac{3}{2}} \sqrt{2\omega_p}} e^{-i\omega_p t + i\vec{p}\cdot\vec{x}}, \\ f_{-p}(t, \vec{x}) &= \frac{1}{(2\pi)^{\frac{3}{2}} \sqrt{2\omega_p}} e^{i\omega_p t - i\vec{p}\cdot\vec{x}}. \end{aligned} \quad (7)$$





**Fig. 1:** Spectrum of the Klein-Gordon wave equation

Given the inner product

$$\langle \psi_1 | \psi_2 \rangle = i \int d^3x \left( \psi_1^* \partial_0 \psi_2 - \partial_0 \psi_1^* \psi_2 \right)$$

the states (7) form an orthonormal basis

$$\langle f_p | f_{p'} \rangle = \delta(\vec{p} - \vec{p}'),$$

$$\langle f_{-p} | f_{-p'} \rangle = -\delta(\vec{p} - \vec{p}'), \quad (8)$$

$$\langle f_p | f_{-p'} \rangle = 0. \quad (9)$$

The wave functions  $f_p(t, x)$  describes states with momentum  $\vec{p}$  and energy given by  $\omega_p = \sqrt{\vec{p}^2 + m^2}$ . On the other hand, the states  $|f_{-p}\rangle$  not only have a negative scalar product but they actually correspond to negative energy states

$$i\partial_0 f_{-p}(t, \vec{x}) = -\sqrt{\vec{p}^2 + m^2} f_{-p}(t, \vec{x}). \quad (10)$$

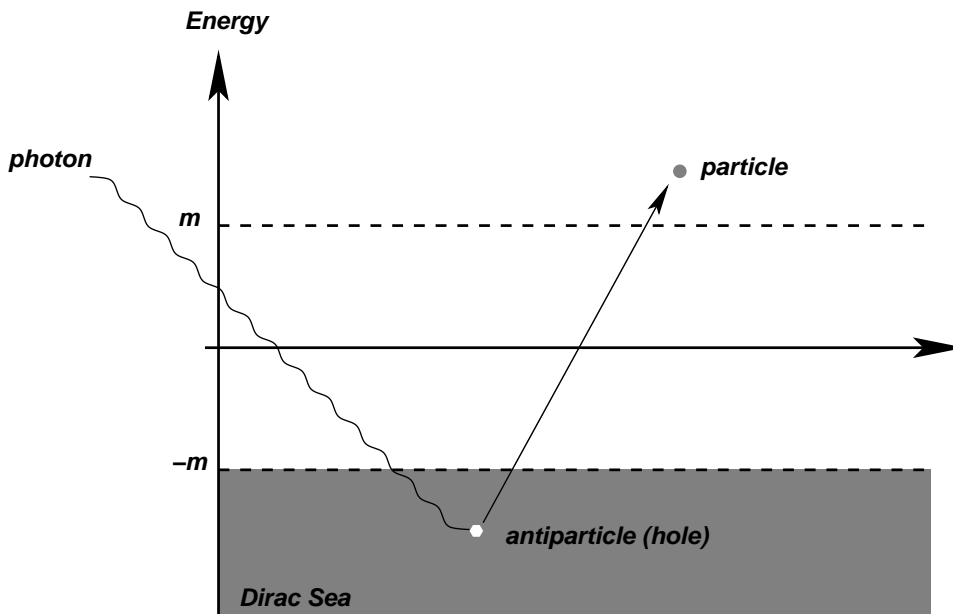
Therefore the energy spectrum of the theory satisfies  $|E| > m$  and is unbounded from below (see Fig. 1). Although in a case of a free theory the absence of a ground state is not necessarily a fatal problem, once the theory is coupled to the electromagnetic field this is the source of all kinds of disasters, since nothing can prevent the decay of any state by emission of electromagnetic radiation.

The problem of the instability of the “first-quantized” relativistic wave equation can be heuristically tackled in the case of spin- $\frac{1}{2}$  particles, described by the Dirac equation

$$\left( -i\beta \frac{\partial}{\partial t} + \vec{\alpha} \cdot \vec{\nabla} - m \right) \psi(t, \vec{x}) = 0, \quad (11)$$

where  $\vec{\alpha}$  and  $\beta$  are  $4 \times 4$  matrices

$$\alpha^i = \begin{pmatrix} 0 & i\sigma^i \\ -i\sigma^i & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}, \quad (12)$$



**Fig. 2:** Creation of a particle-antiparticle pair in the Dirac sea picture

with  $\sigma^i$  the Pauli matrices, and the wave function  $\psi(t, \vec{x})$  has four components. The wave equation (11) can be thought of as a kind of “square root” of the Klein-Gordon equation (3), since the latter can be obtained as

$$\left(-i\beta\frac{\partial}{\partial t} + \vec{\alpha} \cdot \vec{\nabla} - m\right)^\dagger \left(-i\beta\frac{\partial}{\partial t} + \vec{\alpha} \cdot \vec{\nabla} - m\right) \psi(t, \vec{x}) = \left(\frac{\partial^2}{\partial t^2} - \nabla^2 + m^2\right) \psi(t, \vec{x}). \quad (13)$$

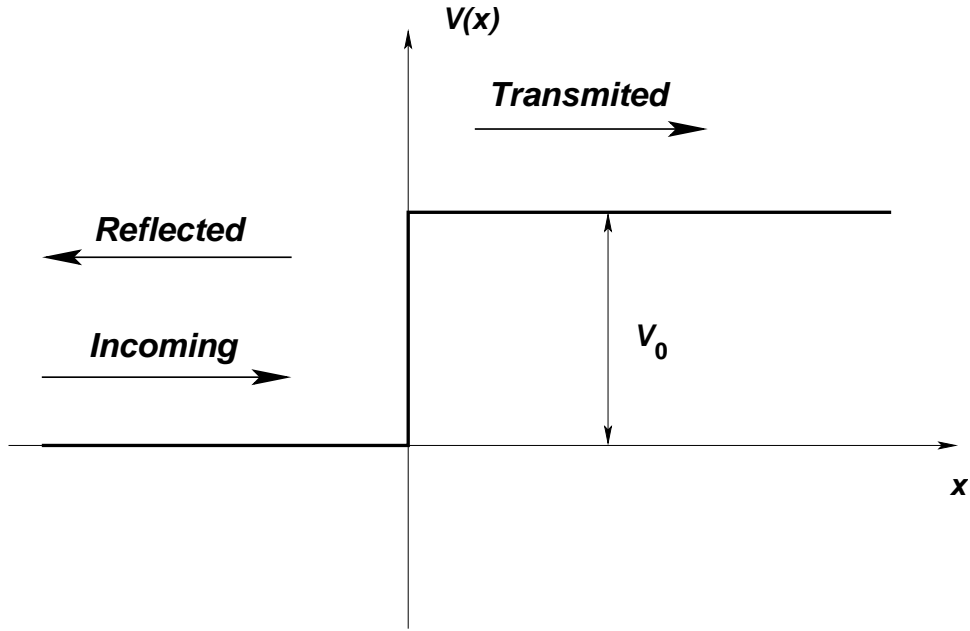
An analysis of Eq. (11) along the lines of the one presented above for the Klein-Gordon equation leads again to the existence of negative energy states and a spectrum unbounded from below as in Fig. 1. Dirac, however, solved the instability problem by pointing out that now the particles are fermions and therefore they are subject to Pauli’s exclusion principle. Hence, each state in the spectrum can be occupied by at most one particle, so the states with  $E = m$  can be made stable if we assume that *all* the negative energy states are filled.

If Dirac’s idea restores the stability of the spectrum by introducing a stable vacuum where all negative energy states are occupied, the so-called Dirac sea, it also leads directly to the conclusion that a single-particle interpretation of the Dirac equation is not possible. Indeed, a photon with enough energy ( $E > 2m$ ) can excite one of the electrons filling the negative energy states, leaving behind a “hole” in the Dirac sea (see Fig. 2). This hole behaves as a particle with equal mass and opposite charge that is interpreted as a positron, so there is no escape to the conclusion that interactions will produce pairs particle-antiparticle out of the vacuum.

In spite of the success of the heuristic interpretation of negative energy states in the Dirac equation this is not the end of the story. In 1929 Oskar Klein stumbled into an apparent paradox when trying to describe the scattering of a relativistic electron by a square potential using Dirac’s wave equation [12] (for pedagogical reviews see [13, 14]). In order to capture the essence of the problem without entering into unnecessary complication we will study Klein’s paradox in the context of the Klein-Gordon equation.

Let us consider a square potential with height  $V_0 > 0$  of the type showed in Fig. 3. A solution to the wave equation in regions I and II is given by

$$\begin{aligned} \psi_I(t, x) &= e^{-iEt+ip_1x} + Re^{-iEt-ip_1x}, \\ \psi_{II}(t, x) &= Te^{-iEt+p_2x}, \end{aligned} \quad (14)$$



**Fig. 3:** Illustration of the Klein paradox.

where the mass-shell condition implies that

$$p_1 = \sqrt{E^2 - m^2}, \quad p_2 = \sqrt{(E - V_0)^2 - m^2}. \quad (15)$$

The constants  $R$  and  $T$  are computed by matching the two solutions across the boundary  $x = 0$ . The conditions  $\psi_I(t, 0) = \psi_{II}(t, 0)$  and  $\partial_x \psi_I(t, 0) = \partial_x \psi_{II}(t, 0)$  imply that

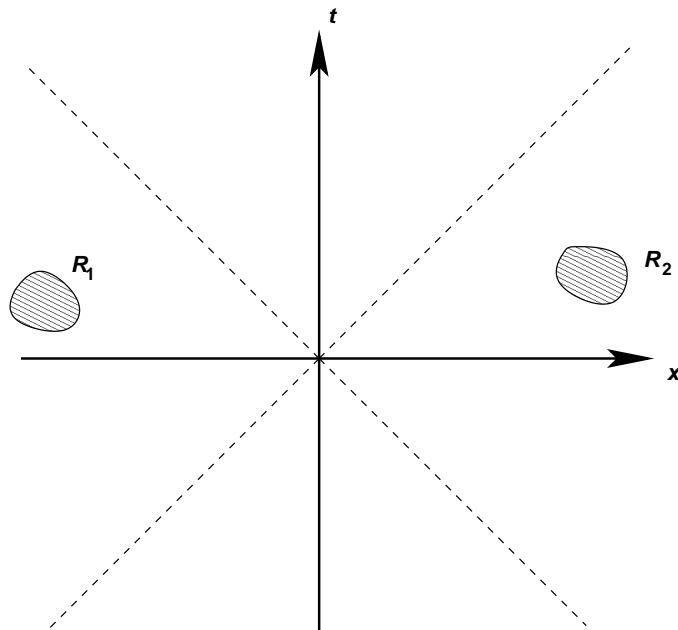
$$T = \frac{2p_1}{p_1 + p_2}, \quad R = \frac{p_1 - p_2}{p_1 + p_2}. \quad (16)$$

At first sight one would expect a behavior similar to the one encountered in the nonrelativistic case. If the kinetic energy is bigger than  $V_0$  both a transmitted and reflected wave are expected, whereas when the kinetic energy is smaller than  $V_0$  one only expect to find a reflected wave, the transmitted wave being exponentially damped within a distance of a Compton wavelength inside the barrier.

Indeed this is what happens if  $E - m > V_0$ . In this case both  $p_1$  and  $p_2$  are real and we have a partly reflected, and a partly transmitted wave. In the same way, if  $V_0 - 2m < E - m < V_0$  then  $p_2$  is imaginary and there is total reflection.

However, in the case when  $V_0 > 2m$  and the energy is in the range  $0 < E - m < V_0 - 2m$  a completely different situation arises. In this case one finds that both  $p_1$  and  $p_2$  are real and therefore the incoming wave function is partially reflected and partially transmitted across the barrier. This is a shocking result, since it implies that there is a nonvanishing probability of finding the particle at any point across the barrier with negative kinetic energy ( $E - m - V_0 < 0$ )! This weird result is known as Klein's paradox.

As with the negative energy states, the Klein paradox results from our insistence in giving a single-particle interpretation to the relativistic wave function. Actually, a multiparticle analysis of the paradox [13] shows that what happens when  $0 < E - m < V_0 - 2m$  is that the reflection of the incoming particle by the barrier is accompanied by the creation of pairs particle-antiparticle out of the energy of the barrier (notice that for this to happen it is required that  $V_0 > 2m$ , the threshold for the creation of a particle-antiparticle pair).



**Fig. 4:** Two regions  $R_1, R_2$  that are causally disconnected.

Actually, this particle creation can be understood by noticing that the sudden potential step in Fig. 3 localizes the incoming particle with mass  $m$  in distances smaller than its Compton wavelength  $\lambda = \frac{1}{m}$ . This can be seen by replacing the square potential by another one where the potential varies smoothly from 0 to  $V_0 > 2m$  in distances scales larger than  $1/m$ . This case was worked out by Sauter shortly after Klein pointed out the paradox [15]. He considered a situation where the regions with  $V = 0$  and  $V = V_0$  are connected by a region of length  $d$  with a linear potential  $V(x) = \frac{V_0 x}{d}$ . When  $d > \frac{1}{m}$  he found that the transmission coefficient is exponentially small<sup>1</sup>.

The creation of particles is impossible to avoid whenever one tries to locate a particle of mass  $m$  within its Compton wavelength. Indeed, from Heisenberg uncertainty relation we find that if  $\Delta x \sim \frac{1}{m}$ , the fluctuations in the momentum will be of order  $\Delta p \sim m$  and fluctuations in the energy of order

$$\Delta E \sim m \quad (17)$$

can be expected. Therefore, in a relativistic theory, the fluctuations of the energy are enough to allow the creation of particles out of the vacuum. In the case of a spin- $\frac{1}{2}$  particle, the Dirac sea picture shows clearly how, when the energy fluctuations are of order  $m$ , electrons from the Dirac sea can be excited to positive energy states, thus creating electron-positron pairs.

It is possible to see how the multiparticle interpretation is forced upon us by relativistic invariance. In non-relativistic Quantum Mechanics observables are represented by self-adjoint operator that in the Heisenberg picture depend on time. Therefore measurements are localized in time but are global in space. The situation is radically different in the relativistic case. Because no signal can propagate faster than the speed of light, measurements have to be localized both in time and space. Causality demands then that two measurements carried out in causally-disconnected regions of space-time cannot interfere with each other. In mathematical terms this means that if  $\mathcal{O}_{R_1}$  and  $\mathcal{O}_{R_2}$  are the observables associated with two measurements localized in two causally-disconnected regions  $R_1, R_2$  (see Fig. 4), they satisfy

$$[\mathcal{O}_{R_1}, \mathcal{O}_{R_2}] = 0, \quad \text{if } (x_1 - x_2)^2 < 0, \text{ for all } x_1 \in R_1, x_2 \in R_2. \quad (18)$$

<sup>1</sup>In section (9.1) we will see how, in the case of the Dirac field, this exponential behavior can be associated with the creation of electron-positron pairs due to a constant electric field (Schwinger effect).

Hence, in a relativistic theory, the basic operators in the Heisenberg picture must depend on the space-time position  $x^\mu$ . Unlike the case in non-relativistic quantum mechanics, here the position  $\vec{x}$  is *not* an observable, but just a label, similarly to the case of time in ordinary quantum mechanics. Causality is then imposed microscopically by requiring

$$[\mathcal{O}(x), \mathcal{O}(y)] = 0, \quad \text{if } (x - y)^2 < 0. \quad (19)$$

A smeared operator  $\mathcal{O}_R$  over a space-time region  $R$  can then be defined as

$$\mathcal{O}_R = \int d^4x \mathcal{O}(x) f_R(x) \quad (20)$$

where  $f_R(x)$  is the characteristic function associated with  $R$ ,

$$f_R(x) = \begin{cases} 1 & x \in R \\ 0 & x \notin R \end{cases}. \quad (21)$$

Eq. (18) follows now from the microcausality condition (19).

Therefore, relativistic invariance forces the introduction of quantum fields. It is only when we insist in keeping a single-particle interpretation that we crash against causality violations. To illustrate the point, let us consider a single particle wave function  $\psi(t, \vec{x})$  that initially is localized in the position  $\vec{x} = 0$

$$\psi(0, \vec{x}) = \delta(\vec{x}). \quad (22)$$

Evolving this wave function using the Hamiltonian  $H = \sqrt{-\nabla^2 + m^2}$  we find that the wave function can be written as

$$\psi(t, \vec{x}) = e^{-it\sqrt{-\nabla^2 + m^2}} \delta(\vec{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x} - it\sqrt{k^2 + m^2}}. \quad (23)$$

Integrating over the angular variables, the wave function can be recast in the form

$$\psi(t, \vec{x}) = \frac{1}{2\pi^2|\vec{x}|} \int_{-\infty}^{\infty} k dk e^{ik|\vec{x}|} e^{-it\sqrt{k^2 + m^2}}. \quad (24)$$

The resulting integral can be evaluated using the complex integration contour  $C$  shown in Fig. 5. The result is that, for any  $t > 0$ , one finds that  $\psi(t, \vec{x}) \neq 0$  for any  $\vec{x}$ . If we insist in interpreting the wave function  $\psi(t, \vec{x})$  as the probability density of finding the particle at the location  $\vec{x}$  in the time  $t$  we find that the probability leaks out of the light cone, thus violating causality.

### 3 From classical to quantum fields

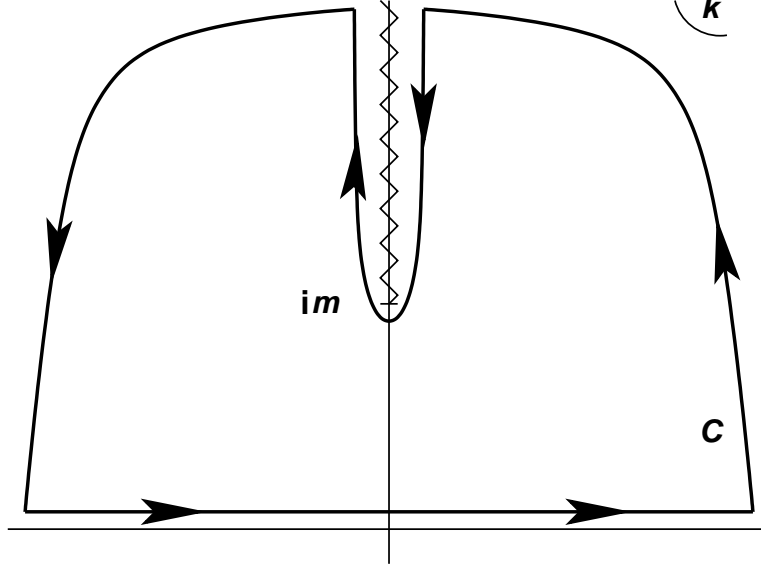
We have learned how the consistency of quantum mechanics with special relativity forces us to abandon the single-particle interpretation of the wave function. Instead we have to consider quantum fields whose elementary excitations are associated with particle states, as we will see below.

In any scattering experiment, the only information available to us is the set of quantum number associated with the set of free particles in the initial and final states. Ignoring for the moment other quantum numbers like spin and flavor, one-particle states are labelled by the three-momentum  $\vec{p}$  and span the single-particle Hilbert space  $\mathcal{H}_1$

$$|\vec{p}\rangle \in \mathcal{H}_1, \quad \langle \vec{p} | \vec{p}' \rangle = \delta(\vec{p} - \vec{p}'). \quad (25)$$

The states  $\{|\vec{p}\rangle\}$  form a basis of  $\mathcal{H}_1$  and therefore satisfy the closure relation

$$\int d^3p |\vec{p}\rangle \langle \vec{p}| = \mathbf{1} \quad (26)$$



**Fig. 5:** Complex contour  $C$  for the computation of the integral in Eq. (24).

The group of spatial rotations acts unitarily on the states  $|\vec{p}\rangle$ . This means that for every rotation  $R \in \text{SO}(3)$  there is a unitary operator  $\mathcal{U}(R)$  such that

$$\mathcal{U}(R)|\vec{p}\rangle = |R\vec{p}\rangle \quad (27)$$

where  $R\vec{p}$  represents the action of the rotation on the vector  $\vec{k}$ ,  $(R\vec{p})^i = R^i_j k^j$ . Using a spectral decomposition, the momentum operator  $\hat{P}^i$  can be written as

$$\hat{P}^i = \int d^3p |\vec{p}\rangle p^i \langle \vec{p}| \quad (28)$$

With the help of Eq. (27) it is straightforward to check that the momentum operator transforms as a vector under rotations:

$$\mathcal{U}(R)^{-1} \hat{P}^i \mathcal{U}(R) = \int d^3p |R^{-1}\vec{p}\rangle p^i \langle R^{-1}\vec{p}| = R^i_j \hat{P}^j, \quad (29)$$

where we have used that the integration measure is invariant under  $\text{SO}(3)$ .

Since, as we argued above, we are forced to deal with multiparticle states, it is convenient to introduce creation-annihilation operators associated with a single-particle state of momentum  $\vec{p}$

$$[a(\vec{p}), a^\dagger(\vec{p}')] = \delta(\vec{p} - \vec{p}'), \quad [a(\vec{p}), a(\vec{p}')] = [a^\dagger(\vec{p}), a^\dagger(\vec{p}')] = 0, \quad (30)$$

such that the state  $|\vec{p}\rangle$  is created out of the Fock space vacuum  $|0\rangle$  (normalized such that  $\langle 0|0\rangle = 1$ ) by the action of a creation operator  $a^\dagger(\vec{p})$

$$|\vec{p}\rangle = a^\dagger(\vec{p})|0\rangle, \quad a(\vec{p})|0\rangle = 0 \quad \forall \vec{p}. \quad (31)$$

Covariance under spatial rotations is all we need if we are interested in a nonrelativistic theory. However in a relativistic quantum field theory we must preserve more than  $\text{SO}(3)$ , actually we need the expressions to be covariant under the full Poincaré group  $\text{ISO}(1, 3)$  consisting in spatial rotations, boosts and space-time translations. Therefore, in order to build the Fock space of the theory we need two key ingredients: first an invariant normalization for the states, since we want a normalized state in

one reference frame to be normalized in any other inertial frame. And secondly a relativistic invariant integration measure in momentum space, so the spectral decomposition of operators is covariant under the full Poincaré group.

Let us begin with the invariant measure. Given an invariant function  $f(p)$  of the four-momentum  $p^\mu$  of a particle of mass  $m$  with positive energy  $p^0 > 0$ , there is an integration measure which is invariant under proper Lorentz transformations<sup>2</sup>

$$\int \frac{d^4 p}{(2\pi)^4} (2\pi) \delta(p^2 - m^2) \theta(p^0) f(p), \quad (32)$$

where  $\theta(x)$  represent the Heaviside step function. The integration over  $p^0$  can be easily done using the  $\delta$ -function identity

$$\delta[f(x)] = \sum_{x_i = \text{zeros of } f} \frac{1}{|f'(x_i)|} \delta(x - x_i), \quad (33)$$

which in our case implies that

$$\delta(p^2 - m^2) = \frac{1}{2p^0} \delta\left(p^0 - \sqrt{\vec{p}^2 + m^2}\right) + \frac{1}{2p^0} \delta\left(p^0 + \sqrt{\vec{p}^2 + m^2}\right). \quad (34)$$

The second term in the previous expression correspond to states with negative energy and therefore does not contribute to the integral. We can write then

$$\int \frac{d^4 p}{(2\pi)^4} (2\pi) \delta(p^2 - m^2) \theta(p^0) f(p) = \int \frac{d^3 p}{(2\pi)^3} \frac{1}{2\sqrt{\vec{p}^2 + m^2}} f\left(\sqrt{\vec{p}^2 + m^2}, \vec{p}\right). \quad (35)$$

Hence, the relativistic invariant measure is given by

$$\int \frac{d^3 p}{(2\pi)^3} \frac{1}{2\omega_p} \quad \text{with} \quad \omega_p \equiv \sqrt{\vec{p}^2 + m^2}. \quad (36)$$

Once we have an invariant measure the next step is to find an invariant normalization for the states. We work with a basis  $\{|p\rangle\}$  of eigenstates of the four-momentum operator  $\hat{P}^\mu$

$$\hat{P}^0 |p\rangle = \omega_p |p\rangle, \quad \hat{P}^i |p\rangle = p^i |p\rangle. \quad (37)$$

Since the states  $|p\rangle$  are eigenstates of the three-momentum operator we can express them in terms of the non-relativistic states  $|\vec{p}\rangle$  that we introduced in Eq. (25)

$$|p\rangle = N(\vec{p}) |\vec{p}\rangle \quad (38)$$

with  $N(\vec{p})$  a normalization to be determined now. The states  $\{|p\rangle\}$  form a complete basis, so they should satisfy the Lorentz invariant closure relation

$$\int \frac{d^4 p}{(2\pi)^4} (2\pi) \delta(p^2 - m^2) \theta(p^0) |p\rangle \langle p| = \mathbf{1} \quad (39)$$

At the same time, this closure relation can be expressed, using Eq. (38), in terms of the nonrelativistic basis of states  $\{|\vec{p}\rangle\}$  as

$$\int \frac{d^4 p}{(2\pi)^4} (2\pi) \delta(p^2 - m^2) \theta(p^0) |p\rangle \langle p| = \int \frac{d^3 p}{(2\pi)^3} \frac{1}{2\omega_p} |N(p)|^2 |\vec{p}\rangle \langle \vec{p}|. \quad (40)$$

<sup>2</sup>The factors of  $2\pi$  are introduced for later convenience.

Using now Eq. (28) for the nonrelativistic states, expression (39) follows provided

$$|N(\vec{p})|^2 = (2\pi)^3 (2\omega_p). \quad (41)$$

Taking the overall phase in Eq. (38) so that  $N(p)$  is real, we define the Lorentz invariant states  $|p\rangle$  as

$$|p\rangle = (2\pi)^{\frac{3}{2}} \sqrt{2\omega_p} |\vec{p}\rangle, \quad (42)$$

and given the normalization of  $|\vec{p}\rangle$  we find the normalization of the relativistic states to be

$$\langle p|p'\rangle = (2\pi)^3 (2\omega_p) \delta(\vec{p} - \vec{p}'). \quad (43)$$

Although not obvious at first sight, the previous normalization is Lorentz invariant. Although it is not difficult to show this in general, here we consider the simpler case of 1+1 dimensions where the two components  $(p^0, p^1)$  of the on-shell momentum can be parametrized in terms of a single hyperbolic angle  $\lambda$  as

$$p^0 = m \cosh \lambda, \quad p^1 = m \sinh \lambda. \quad (44)$$

Now, the combination  $2\omega_p \delta(p^1 - p'^1)$  can be written as

$$2\omega_p \delta(p^1 - p'^1) = 2m \cosh \lambda \delta(m \sinh \lambda - m \sinh \lambda') = 2\delta(\lambda - \lambda'), \quad (45)$$

where we have made use of the property (33) of the  $\delta$ -function. Lorentz transformations in 1 + 1 dimensions are labelled by a parameter  $\xi \in \mathbb{R}$  and act on the momentum by shifting the hyperbolic angle  $\lambda \rightarrow \lambda + \xi$ . However, Eq. (45) is invariant under a common shift of  $\lambda$  and  $\lambda'$ , so the whole expression is obviously invariant under Lorentz transformations.

To summarize what we did so far, we have succeed in constructing a Lorentz covariant basis of states for the one-particle Hilbert space  $\mathcal{H}_1$ . The generators of the Poincaré group act on the states  $|p\rangle$  of the basis as

$$\widehat{P}^\mu |p\rangle = p^\mu |p\rangle, \quad \mathcal{U}(\Lambda)|p\rangle = |\Lambda^\mu{}_\nu p^\nu\rangle \equiv |\Lambda p\rangle \quad \text{with} \quad \Lambda \in \text{SO}(1, 3). \quad (46)$$

This is compatible with the Lorentz invariance of the normalization that we have checked above

$$\langle p|p'\rangle = \langle p|\mathcal{U}(\Lambda)^{-1}\mathcal{U}(\Lambda)|p'\rangle = \langle \Lambda p|\Lambda p'\rangle. \quad (47)$$

On  $\mathcal{H}_1$  the operator  $\widehat{P}^\mu$  admits the following spectral representation

$$\widehat{P}^\mu = \int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} |p\rangle p^\mu \langle p|. \quad (48)$$

Using (47) and the fact that the measure is invariant under Lorentz transformation, one can easily show that  $\widehat{P}^\mu$  transform covariantly under  $\text{SO}(1, 3)$

$$\mathcal{U}(\Lambda)^{-1} \widehat{P}^\mu \mathcal{U}(\Lambda) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} |\Lambda^{-1}p\rangle p^\mu \langle \Lambda^{-1}p| = \Lambda^\mu{}_\nu \widehat{P}^\nu. \quad (49)$$

A set of covariant creation-annihilation operators can be constructed now in terms of the operators  $a(\vec{p})$ ,  $a^\dagger(\vec{p})$  introduced above

$$\alpha(\vec{p}) \equiv (2\pi)^{\frac{3}{2}} \sqrt{2\omega_p} a(\vec{p}), \quad \alpha^\dagger(\vec{p}) \equiv (2\pi)^{\frac{3}{2}} \sqrt{2\omega_p} a^\dagger(\vec{p}) \quad (50)$$

with the Lorentz invariant commutation relations

$$[\alpha(\vec{p}), \alpha^\dagger(\vec{p}')] = (2\pi)^3 (2\omega_p) \delta(\vec{p} - \vec{p}'),$$



$$[\alpha(\vec{p}), \alpha(\vec{p}')] = [\alpha^\dagger(\vec{p}), \alpha^\dagger(\vec{p}')] = 0. \quad (51)$$

Particle states are created by acting with any number of creation operators  $\alpha(\vec{p})$  on the Poincaré invariant vacuum state  $|0\rangle$  satisfying

$$\langle 0|0\rangle = 1, \quad \widehat{P}^\mu|0\rangle = 0, \quad \mathcal{U}(\Lambda)|0\rangle = |0\rangle, \quad \forall \Lambda \in \text{SO}(1, 3). \quad (52)$$

A general one-particle state  $|f\rangle \in \mathcal{H}_1$  can be then written as

$$|f\rangle = \int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} f(\vec{p}) \alpha^\dagger(\vec{p})|0\rangle, \quad (53)$$

while a  $n$ -particle state  $|f\rangle \in \mathcal{H}_1^{\otimes n}$  can be expressed as

$$|f\rangle = \int \prod_{i=1}^n \frac{d^3p_i}{(2\pi)^3} \frac{1}{2\omega_{p_i}} f(\vec{p}_1, \dots, \vec{p}_n) \alpha^\dagger(\vec{p}_1) \dots \alpha^\dagger(\vec{p}_n)|0\rangle. \quad (54)$$

That this states are Lorentz invariant can be checked by noticing that from the definition of the creation-annihilation operators follows the transformation

$$\mathcal{U}(\Lambda)\alpha(\vec{p})\mathcal{U}(\Lambda)^\dagger = \alpha(\Lambda\vec{p}) \quad (55)$$

and the corresponding one for creation operators.

As we have argued above, the very fact that measurements have to be localized implies the necessity of introducing quantum fields. Here we will consider the simplest case of a scalar quantum field  $\phi(x)$  satisfying the following properties:

- **Hermiticity.**

$$\phi^\dagger(x) = \phi(x). \quad (56)$$

- **Microcausality.** Since measurements cannot interfere with each other when performed in causally disconnected points of space-time, the commutator of two fields have to vanish outside the relative lighth-cone

$$[\phi(x), \phi(y)] = 0, \quad (x - y)^2 < 0. \quad (57)$$

- **Translation invariance.**

$$e^{i\widehat{P}\cdot a} \phi(x) e^{-i\widehat{P}\cdot a} = \phi(x - a). \quad (58)$$

- **Lorentz invariance.**

$$\mathcal{U}(\Lambda)^\dagger \phi(x) \mathcal{U}(\Lambda) = \phi(\Lambda^{-1}x). \quad (59)$$

- **Linearity.** To simplify matters we will also assume that  $\phi(x)$  is linear in the creation-annihilation operators  $\alpha(\vec{p}), \alpha^\dagger(\vec{p})$

$$\phi(x) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} \left[ f(\vec{p}, x) \alpha(\vec{p}) + g(\vec{p}, x) \alpha^\dagger(\vec{p}) \right]. \quad (60)$$

Since  $\phi(x)$  should be hermitian we are forced to take  $f(\vec{p}, x)^* = g(\vec{p}, x)$ . Moreover,  $\phi(x)$  satisfies the equations of motion of a free scalar field,  $(\partial_\mu \partial^\mu + m^2)\phi(x) = 0$ , only if  $f(\vec{p}, x)$  is a complete basis of solutions of the Klein-Gordon equation. These considerations leads to the expansion

$$\phi(x) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} \left[ e^{-i\omega_p t + i\vec{p}\cdot\vec{x}} \alpha(\vec{p}) + e^{i\omega_p t - i\vec{p}\cdot\vec{x}} \alpha^\dagger(\vec{p}) \right]. \quad (61)$$

Given the expansion of the scalar field in terms of the creation-annihilation operators it can be checked that  $\phi(x)$  and  $\partial_t\phi(x)$  satisfy the equal-time canonical commutation relations

$$[\phi(t, \vec{x}), \partial_t\phi(t, \vec{y})] = i\delta(\vec{x} - \vec{y}) \quad (62)$$

The general commutator  $[\phi(x), \phi(y)]$  can be also computed to be

$$[\phi(x), \phi(x')] = i\Delta(x - x'). \quad (63)$$

The function  $\Delta(x - y)$  is given by

$$\begin{aligned} i\Delta(x - y) &= -\text{Im} \int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} e^{-i\omega_p(t-t') + i\vec{p}\cdot(\vec{x}-\vec{x}')} \\ &= \int \frac{d^4p}{(2\pi)^4} (2\pi)\delta(p^2 - m^2)\varepsilon(p^0)e^{-ip\cdot(x-x')}, \end{aligned} \quad (64)$$

where  $\varepsilon(x)$  is defined as

$$\varepsilon(x) \equiv \theta(x) - \theta(-x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}. \quad (65)$$

Using the last expression in Eq. (64) it is easy to show that  $i\Delta(x - x')$  vanishes when  $x$  and  $x'$  are space-like separated. Indeed, if  $(x - x')^2 < 0$  there is always a reference frame in which both events are simultaneous, and since  $i\Delta(x - x')$  is Lorentz invariant we can compute it in this reference frame. In this case  $t = t'$  and the exponential in the second line of (64) does not depend on  $p^0$ . Therefore, the integration over  $k^0$  gives

$$\begin{aligned} \int_{-\infty}^{\infty} dp^0 \varepsilon(p^0)\delta(p^2 - m^2) &= \int_{-\infty}^{\infty} dp^0 \left[ \frac{1}{2\omega_p} \varepsilon(p^0)\delta(p^0 - \omega_p) + \frac{1}{2\omega_p} \varepsilon(p^0)\delta(p^0 + \omega_p) \right] \\ &= \frac{1}{2\omega_p} - \frac{1}{2\omega_p} = 0. \end{aligned} \quad (66)$$

So we have concluded that  $i\Delta(x - x') = 0$  if  $(x - x')^2 < 0$ , as required by microcausality. Notice that the situation is completely different when  $(x - x')^2 \geq 0$ , since in this case the exponential depends on  $p^0$  and the integration over this component of the momentum does not vanish.

### 3.1 Canonical quantization

So far we have contented ourselves with requiring a number of properties to the quantum scalar field: existence of asymptotic states, locality, microcausality and relativistic invariance. With these only ingredients we have managed to go quite far. The previous can also be obtained using canonical quantization. One starts with a classical free scalar field theory in Hamiltonian formalism and obtains the quantum theory by replacing Poisson brackets by commutators. Since this quantization procedure is based on the use of the canonical formalism, which gives time a privileged rôle, it is important to check at the end of the calculation that the resulting quantum theory is Lorentz invariant. In the following we will briefly overview the canonical quantization of the Klein-Gordon scalar field.

The starting point is the action functional  $S[\phi(x)]$  which, in the case of a free real scalar field of mass  $m$  is given by

$$S[\phi(x)] \equiv \int d^4x \mathcal{L}(\phi, \partial_\mu\phi) = \frac{1}{2} \int d^4x (\partial_\mu\phi\partial^\mu\phi - m^2\phi^2). \quad (67)$$

The equations of motion are obtained, as usual, from the Euler-Lagrange equations

$$\partial_\mu \left[ \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right] - \frac{\partial \mathcal{L}}{\partial \phi} = 0 \quad \Longrightarrow \quad (\partial_\mu \partial^\mu + m^2)\phi = 0. \quad (68)$$

The momentum canonically conjugated to the field  $\phi(x)$  is given by

$$\pi(x) \equiv \frac{\partial \mathcal{L}}{\partial(\partial_0 \phi)} = \frac{\partial \phi}{\partial t}. \quad (69)$$

In the Hamiltonian formalism the physical system is described not in terms of the generalized coordinates and their time derivatives but in terms of the generalized coordinates and their canonically conjugated momenta. This is achieved by a Legendre transformation after which the dynamics of the system is determined by the Hamiltonian function

$$H \equiv \int d^3x \left( \pi \frac{\partial \phi}{\partial t} - \mathcal{L} \right) = \frac{1}{2} \int d^3x \left[ \pi^2 + (\vec{\nabla} \phi)^2 + m^2 \right]. \quad (70)$$

The equations of motion can be written in terms of the Poisson brackets. Given two functional  $A[\phi, \pi]$ ,  $B[\phi, \pi]$  of the canonical variables

$$A[\phi, \pi] = \int d^3x \mathcal{A}(\phi, \pi), \quad B[\phi, \pi] = \int d^3x \mathcal{B}(\phi, \pi). \quad (71)$$

Their Poisson bracket is defined by

$$\{A, B\} \equiv \int d^3x \left[ \frac{\delta A}{\delta \phi} \frac{\delta B}{\delta \pi} - \frac{\delta A}{\delta \pi} \frac{\delta B}{\delta \phi} \right], \quad (72)$$

where  $\frac{\delta}{\delta \phi}$  denotes the functional derivative defined as

$$\frac{\delta A}{\delta \phi} \equiv \frac{\partial \mathcal{A}}{\partial \phi} - \partial_\mu \left[ \frac{\partial \mathcal{A}}{\partial(\partial_\mu \phi)} \right] \quad (73)$$

Then, the canonically conjugated fields satisfy the following equal time Poisson brackets

$$\begin{aligned} \{\phi(t, \vec{x}), \phi(t, \vec{x}')\} &= \{\pi(t, \vec{x}), \pi(t, \vec{x}')\} = 0, \\ \{\phi(t, \vec{x}), \pi(t, \vec{x}')\} &= \delta(\vec{x} - \vec{x}'). \end{aligned} \quad (74)$$

Canonical quantization proceeds now by replacing classical fields with operators and Poisson brackets with commutators according to the rule

$$i\{\cdot, \cdot\} \longrightarrow [\cdot, \cdot]. \quad (75)$$

In the case of the scalar field, a general solution of the field equations (68) can be obtained by working with the Fourier transform

$$(\partial_\mu \partial^\mu + m^2)\phi(x) = 0 \quad \Longrightarrow \quad (-p^2 + m^2)\tilde{\phi}(p) = 0, \quad (76)$$

whose general solution can be written as<sup>3</sup>

$$\phi(x) = \int \frac{d^4p}{(2\pi)^4} (2\pi) \delta(p^2 - m^2) \theta(p^0) [\alpha(p) e^{-ip \cdot x} + \alpha(p)^* e^{ip \cdot x}]$$

<sup>3</sup>In momentum space, the general solution to this equation is  $\tilde{\phi}(p) = f(p) \delta(p^2 - m^2)$ , with  $f(p)$  a completely general function of  $p^\mu$ . The solution in position space is obtained by inverse Fourier transform.

$$= \int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} \left[ \alpha(\vec{p}) e^{-i\omega_p t + \vec{p} \cdot \vec{x}} + \alpha(\vec{p})^* e^{i\omega_p t - \vec{p} \cdot \vec{x}} \right] \quad (77)$$

and we have required  $\phi(x)$  to be real. The conjugate momentum is

$$\pi(x) = -\frac{i}{2} \int \frac{d^3p}{(2\pi)^3} \left[ \alpha(\vec{p}) e^{-i\omega_p t + \vec{p} \cdot \vec{x}} + \alpha(\vec{p})^* e^{i\omega_p t - \vec{p} \cdot \vec{x}} \right]. \quad (78)$$

Now  $\phi(x)$  and  $\pi(x)$  are promoted to operators by replacing the functions  $\alpha(\vec{p})$ ,  $\alpha(\vec{p})^*$  by the corresponding operators

$$\alpha(\vec{p}) \longrightarrow \hat{\alpha}(\vec{p}), \quad \alpha(\vec{p})^* \longrightarrow \hat{\alpha}^\dagger(\vec{p}). \quad (79)$$

Moreover, demanding  $[\phi(t, \vec{x}), \pi(t, \vec{x}')] = i\delta(\vec{x} - \vec{x}')$  forces the operators  $\hat{\alpha}(\vec{p})$ ,  $\hat{\alpha}(\vec{p})^\dagger$  to have the commutation relations found in Eq. (51). Therefore they are identified as a set of creation-annihilation operators creating states with well-defined momentum  $\vec{p}$  out of the vacuum  $|0\rangle$ . In the canonical quantization formalism the concept of particle appears as a result of the quantization of a classical field.

Knowing the expressions of  $\hat{\phi}$  and  $\hat{\pi}$  in terms of the creation-annihilation operators we can proceed to evaluate the Hamiltonian operator. After a simple calculation one arrives to the expression

$$\hat{H} = \int d^3p \left[ \omega_p \hat{\alpha}^\dagger(\vec{p}) \hat{\alpha}(\vec{p}) + \frac{1}{2} \omega_p \delta(\vec{0}) \right]. \quad (80)$$

The first term has a simple physical interpretation since  $\hat{\alpha}^\dagger(\vec{p}) \hat{\alpha}(\vec{p})$  is the number operator of particles with momentum  $\vec{p}$ . The second divergent term can be eliminated if we defined the normal-ordered Hamiltonian  $:\hat{H}:$  with the vacuum energy subtracted

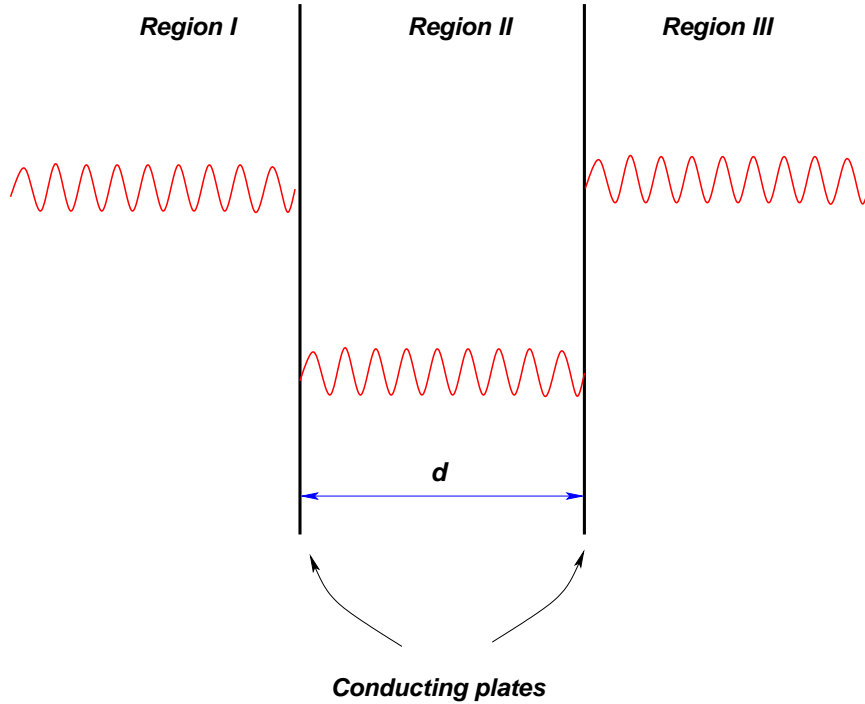
$$:\hat{H}: \equiv \hat{H} - \langle 0 | \hat{H} | 0 \rangle = \int d^3p \omega_p \hat{\alpha}^\dagger(\vec{p}) \hat{\alpha}(\vec{p}) \quad (81)$$

It is interesting to try to make sense of the divergent term in Eq. (80). This term has two sources of divergence. One is associated with the delta function evaluated at zero coming from the fact that we are working in an infinite volume. It can be regularized for large but finite volume by replacing  $\delta(\vec{0}) \sim V$ . Hence, it is of infrared origin. The second one comes from the integration of  $\omega_p$  at large values of the momentum and it is then an ultraviolet divergence. The infrared divergence can be regularized by considering the scalar field to be living in a box of finite volume  $V$ . In this case the vacuum energy is

$$E_{\text{vac}} \equiv \langle 0 | \hat{H} | 0 \rangle = \sum_{\vec{p}} \frac{1}{2} \omega_p. \quad (82)$$

Written in this way the interpretation of the vacuum energy is straightforward. A free scalar quantum field can be seen as an infinite collection of harmonic oscillators per unit volume, each one labelled by  $\vec{p}$ . Even if those oscillators are not excited, they contribute to the vacuum energy with their zero-point energy, given by  $\frac{1}{2} \omega_p$ . This vacuum contribution to the energy adds up to infinity even if we work at finite volume, since even then there are modes with arbitrary high momentum contributing to the sum,  $p_i = \frac{n_i \pi}{L_i}$ , with  $L_i$  the sides of the box of volume  $V$  and  $n_i$  an integer. Hence, this divergence is of ultraviolet origin.

Our discussion leads us to the conclusion that the vacuum in quantum field theory is radically different from the classical idea of the vacuum as “empty space”. Indeed, we have seen that a quantum field can be regarded as a set of an infinite number of harmonic oscillators and that the ground state of the system is obtained when *all* oscillators are in their respective ground states. This being so, we know from elementary quantum mechanics that a harmonic oscillator in its ground state is not “at rest”, but



**Fig. 6:** Illustration of the Casimir effect. In regions I and II the spectrum of modes of the momentum  $p_{\perp}$  is continuous, while in the space between the plates (region II) it is quantized in units of  $\frac{\pi}{d}$ .

fluctuate with an energy given by its zero-point energy. When translated to quantum field theory, this means that the vacuum can be picture as a medium where virtual particles are continuously created and annihilated. As we will see, this nontrivial character of the vacuum has physical consequences ranging from the Casimir effect (see below) to the screening or antiscreening of charges in gauge theories (see Section 8.2).

### 3.2 The Casimir effect

The presence of a vacuum energy is not characteristic of the scalar field. It is also present in other cases, in particular in quantum electrodynamics. Although one might be tempted to discarding this infinite contribution to the energy of the vacuum as unphysical, it has observable consequences. In 1948 Hendrik Casimir pointed out [16] that although a formally divergent vacuum energy would not be observable, any variation in this energy would be (see [17] for comprehensive reviews).

To show this he devised the following experiment. Consider a couple of infinite, perfectly conducting plates placed parallel to each other at a distance  $d$  (see Fig. 6). Because the conducting plates fix the boundary condition of the vacuum modes of the electromagnetic field these are discrete in between the plates (region II), while outside there is a continuous spectrum of modes (regions I and III). In order to calculate the force between the plates we can take the vacuum energy of the electromagnetic field as given by the contribution of two scalar fields corresponding to the two polarizations of the photon. Therefore we can use the formulas derived above.

A naive calculation of the vacuum energy in this system gives a divergent result. This infinity can be removed, however, by subtracting the vacuum energy corresponding to the situation where the plates are removed

$$E(d)_{\text{reg}} = E(d)_{\text{vac}} - E(\infty)_{\text{vac}} \quad (83)$$

This subtraction cancels the contribution of the modes outside the plates. Because of the boundary

conditions imposed by the plates the momentum of the modes perpendicular to the plates are quantized according to  $p_{\perp} = \frac{n\pi}{d}$ , with  $n$  a non-negative integer. If we consider that the size of the plates is much larger than their separation  $d$  we can take the momenta parallel to the plates  $\vec{p}_{\parallel}$  as continuous. For  $n > 0$  we have two polarizations for each vacuum mode of the electromagnetic field, each contributing like  $\frac{1}{2}\sqrt{\vec{p}_{\parallel}^2 + p_{\perp}^2}$  to the vacuum energy. On the other hand, when  $p_{\perp} = 0$  the corresponding modes of the field are effectively (2+1)-dimensional and therefore there is only one polarization. Keeping this in mind, we can write

$$\begin{aligned} E(d)_{\text{reg}} &= S \int \frac{d^2 p_{\parallel}}{(2\pi)^2} \frac{1}{2} |\vec{p}_{\parallel}| + 2S \int \frac{d^2 p_{\parallel}}{(2\pi)^2} \sum_{n=1}^{\infty} \frac{1}{2} \sqrt{\vec{p}_{\parallel}^2 + \left(\frac{n\pi}{d}\right)^2} \\ &\quad - 2Sd \int \frac{d^3 p}{(2\pi)^3} \frac{1}{2} |\vec{p}| \end{aligned} \quad (84)$$

where  $S$  is the area of the plates. The factors of 2 take into account the two propagating degrees of freedom of the electromagnetic field, as discussed above. In order to ensure the convergence of integrals and infinite sums we can introduce an exponential damping factor<sup>4</sup>

$$\begin{aligned} E(d)_{\text{reg}} &= \frac{1}{2} S \int \frac{d^2 p_{\perp}}{(2\pi)^2} e^{-\frac{1}{\Lambda} |\vec{p}_{\perp}|} |\vec{p}_{\perp}| + S \sum_{n=1}^{\infty} \int \frac{d^2 p_{\parallel}}{(2\pi)^2} e^{-\frac{1}{\Lambda} \sqrt{\vec{p}_{\parallel}^2 + \left(\frac{n\pi}{d}\right)^2}} \sqrt{\vec{p}_{\parallel}^2 + \left(\frac{n\pi}{d}\right)^2} \\ &\quad - Sd \int_{-\infty}^{\infty} \frac{dp_{\perp}}{2\pi} \int \frac{d^2 p_{\parallel}}{(2\pi)^2} e^{-\frac{1}{\Lambda} \sqrt{\vec{p}_{\parallel}^2 + p_{\perp}^2}} \sqrt{\vec{p}_{\parallel}^2 + p_{\perp}^2} \end{aligned} \quad (85)$$

where  $\Lambda$  is an ultraviolet cutoff. It is now straightforward to see that if we define the function

$$F(x) = \frac{1}{2\pi} \int_0^{\infty} y dy e^{-\frac{1}{\Lambda} \sqrt{y^2 + \left(\frac{x\pi}{d}\right)^2}} \sqrt{y^2 + \left(\frac{x\pi}{d}\right)^2} = \frac{1}{4\pi} \int_{\left(\frac{x\pi}{d}\right)^2}^{\infty} dz e^{-\frac{\sqrt{z}}{\Lambda}} \sqrt{z} \quad (86)$$

the regularized vacuum energy can be written as

$$E(d)_{\text{reg}} = S \left[ \frac{1}{2} F(0) + \sum_{n=1}^{\infty} F(n) - \int_0^{\infty} dx F(x) \right] \quad (87)$$

This expression can be evaluated using the Euler-MacLaurin formula [19]

$$\begin{aligned} \sum_{n=1}^{\infty} F(n) - \int_0^{\infty} dx F(x) &= -\frac{1}{2} [F(0) + F(\infty)] + \frac{1}{12} [F'(\infty) - F'(0)] \\ &\quad - \frac{1}{720} [F'''(\infty) - F'''(0)] + \dots \end{aligned} \quad (88)$$

Since for our function  $F(\infty) = F'(\infty) = F'''(\infty) = 0$  and  $F'(0) = 0$ , the value of  $E(d)_{\text{reg}}$  is determined by  $F'''(0)$ . Computing this term and removing the ultraviolet cutoff,  $\Lambda \rightarrow \infty$  we find the result

$$E(d)_{\text{reg}} = \frac{S}{720} F'''(0) = -\frac{\pi^2 S}{720 d^3}. \quad (89)$$

Then, the force per unit area between the plates is given by

$$P_{\text{Casimir}} = -\frac{\pi^2}{240} \frac{1}{d^4}. \quad (90)$$

The minus sign shows that the force between the plates is attractive. This is the so-called Casimir effect. It was experimentally measured in 1958 by Sparnaay [18] and since then the Casimir effect has been checked with better and better precision in a variety of situations [17].

<sup>4</sup>Actually, one could introduce any cutoff function  $f(p_{\perp}^2 + p_{\parallel}^2)$  going to zero fast enough as  $p_{\perp}, p_{\parallel} \rightarrow \infty$ . The result is independent of the particular function used in the calculation.

## 4 Theories and Lagrangians

Up to this point we have used a scalar field to illustrate our discussion of the quantization procedure. However, nature is richer than that and it is necessary to consider other fields with more complicated behavior under Lorentz transformations. Before considering other fields we pause and study the properties of the Lorentz group.

### 4.1 Representations of the Lorentz group

In four dimensions the Lorentz group has six generators. Three of them correspond to the generators of the group of rotations in three dimensions  $SO(3)$ . In terms of the generators  $J_i$  of the group a finite rotation of angle  $\varphi$  with respect to an axis determined by a unitary vector  $\vec{e}$  can be written as

$$R(\vec{e}, \varphi) = e^{-i\varphi \vec{e} \cdot \vec{J}}, \quad \vec{J} = \begin{pmatrix} J_1 \\ J_2 \\ J_3 \end{pmatrix}. \quad (91)$$

The other three generators of the Lorentz group are associated with boosts  $M_i$  along the three spatial directions. A boost with rapidity  $\lambda$  along a direction  $\vec{u}$  is given by

$$B(\vec{u}, \lambda) = e^{-i\lambda \vec{u} \cdot \vec{M}}, \quad \vec{M} = \begin{pmatrix} M_1 \\ M_2 \\ M_3 \end{pmatrix}. \quad (92)$$

These six generators satisfy the algebra

$$\begin{aligned} [J_i, J_j] &= i\epsilon_{ijk} J_k, \\ [J_i, M_k] &= i\epsilon_{ijk} M_k, \\ [M_i, M_j] &= -i\epsilon_{ijk} J_k, \end{aligned} \quad (93)$$

The first line corresponds to the commutation relations of  $SO(3)$ , while the second one implies that the generators of the boosts transform like a vector under rotations.

At first sight, to find representations of the algebra (93) might seem difficult. The problem is greatly simplified if we consider the following combination of the generators

$$J_k^\pm = \frac{1}{2}(J_k \pm iM_k). \quad (94)$$

Using (93) it is easy to prove that the new generators  $J_k^\pm$  satisfy the algebra

$$\begin{aligned} [J_i^\pm, J_j^\pm] &= i\epsilon_{ijk} J_k^\pm, \\ [J_i^+, J_j^-] &= 0. \end{aligned} \quad (95)$$

Then the Lorentz algebra (93) is actually equivalent to two copies of the algebra of  $SU(2) \approx SO(3)$ . Therefore the irreducible representations of the Lorentz group can be obtained from the well-known representations of  $SU(2)$ . Since the latter ones are labelled by the spin  $\mathbf{s} = k + \frac{1}{2}, k$  (with  $k \in \mathbb{N}$ ), any representation of the Lorentz algebra can be identified by specifying  $(\mathbf{s}_+, \mathbf{s}_-)$ , the spins of the representations of the two copies of  $SU(2)$  that made up the algebra (93).

To get familiar with this way of labelling the representations of the Lorentz group we study some particular examples. Let us start with the simplest one  $(\mathbf{s}_+, \mathbf{s}_-) = (\mathbf{0}, \mathbf{0})$ . This state is a singlet under  $J_i^\pm$  and therefore also under rotations and boosts. Therefore we have a scalar.

The next interesting cases are  $(\frac{1}{2}, \mathbf{0})$  and  $(\mathbf{0}, \frac{1}{2})$ . They correspond respectively to a right-handed and a left-handed Weyl spinor. Their properties will be studied in more detail below. In the case of

Representation	Type of field
$(\mathbf{0}, \mathbf{0})$	Scalar
$(\frac{1}{2}, \mathbf{0})$	Right-handed spinor
$(\mathbf{0}, \frac{1}{2})$	Left-handed spinor
$(\frac{1}{2}, \frac{1}{2})$	Vector
$(\mathbf{1}, \mathbf{0})$	Selfdual antisymmetric 2-tensor
$(\mathbf{0}, \mathbf{1})$	Anti-selfdual antisymmetric 2-tensor

**Table 1:** Representations of the Lorentz group

$(\frac{1}{2}, \frac{1}{2})$ , since from Eq. (94) we see that  $J_i = J_i^+ + J_i^-$  the rules of addition of angular momentum tell us that there are two states, one of them transforming as a vector and another one as a scalar under three-dimensional rotations. Actually, a more detailed analysis shows that the singlet state corresponds to the time component of a vector and the states combine to form a vector under the Lorentz group.

There are also more “exotic” representations. For example we can consider the  $(\mathbf{1}, \mathbf{0})$  and  $(\mathbf{0}, \mathbf{1})$  representations corresponding respectively to a selfdual and an anti-selfdual rank-two antisymmetric tensor. In Table 1 we summarize the previous discussion.

To conclude our discussion of the representations of the Lorentz group we notice that under a parity transformation the generators of  $SO(1,3)$  transform as

$$P : J_i \longrightarrow J_i, \quad P : M_i \longrightarrow -M_i \quad (96)$$

this means that  $P : J_i^\pm \longrightarrow J_i^\mp$  and therefore a representation  $(\mathbf{s}_1, \mathbf{s}_2)$  is transformed into  $(\mathbf{s}_2, \mathbf{s}_1)$ . This means that, for example, a vector  $(\frac{1}{2}, \frac{1}{2})$  is invariant under parity, whereas a left-handed Weyl spinor  $(\frac{1}{2}, \mathbf{0})$  transforms into a right-handed one  $(\mathbf{0}, \frac{1}{2})$  and vice versa.

## 4.2 Spinors

**Weyl spinors.** Let us go back to the two spinor representations of the Lorentz group, namely  $(\frac{1}{2}, \mathbf{0})$  and  $(\mathbf{0}, \frac{1}{2})$ . These representations can be explicitly constructed using the Pauli matrices as

$$\begin{aligned} J_i^+ &= \frac{1}{2}\sigma^i, & J_i^- &= 0 & \text{for } (\frac{1}{2}, \mathbf{0}), \\ J_i^+ &= 0, & J_i^- &= \frac{1}{2}\sigma^i & \text{for } (\mathbf{0}, \frac{1}{2}). \end{aligned} \quad (97)$$

We denote by  $u_\pm$  a complex two-component object that transforms in the representation  $\mathbf{s}_\pm = \frac{1}{2}$  of  $J_\pm^i$ . If we define  $\sigma_\pm^\mu = (\mathbf{1}, \pm\sigma^i)$  we can construct the following vector quantities

$$u_+^\dagger \sigma_+^\mu u_+, \quad u_-^\dagger \sigma_-^\mu u_-. \quad (98)$$

Notice that since  $(J_i^\pm)^\dagger = J_i^\mp$  the hermitian conjugated fields  $u_\pm^\dagger$  are in the  $(\mathbf{0}, \frac{1}{2})$  and  $(\frac{1}{2}, \mathbf{0})$  respectively.

To construct a free Lagrangian for the fields  $u_\pm$  we have to look for quadratic combinations of the fields that are Lorentz scalars. If we also demand invariance under global phase rotations

$$u_\pm \longrightarrow e^{i\theta} u_\pm \quad (99)$$



we are left with just one possibility up to a sign

$$\mathcal{L}_{\text{Weyl}}^{\pm} = iu_{\pm}^{\dagger} \left( \partial_t \pm \vec{\sigma} \cdot \vec{\nabla} \right) u_{\pm} = iu_{\pm}^{\dagger} \sigma_{\pm}^{\mu} \partial_{\mu} u_{\pm}. \quad (100)$$

This is the Weyl Lagrangian. In order to grasp the physical meaning of the spinors  $u_{\pm}$  we write the equations of motion

$$\left( \partial_0 \pm \vec{\sigma} \cdot \vec{\nabla} \right) u_{\pm} = 0. \quad (101)$$

Multiplying this equation on the left by  $\left( \partial_0 \mp \vec{\sigma} \cdot \vec{\nabla} \right)$  and applying the algebraic properties of the Pauli matrices we conclude that  $u_{\pm}$  satisfies the massless Klein-Gordon equation

$$\partial_{\mu} \partial^{\mu} u_{\pm} = 0, \quad (102)$$

whose solutions are:

$$u_{\pm}(x) = u_{\pm}(k) e^{-ik \cdot x}, \quad \text{with } k^0 = |\vec{k}|. \quad (103)$$

Plugging these solutions back into the equations of motion (101) we find

$$\left( |\vec{k}| \mp \vec{k} \cdot \vec{\sigma} \right) u_{\pm} = 0, \quad (104)$$

which implies

$$\begin{aligned} u_{+} : & \quad \frac{\vec{\sigma} \cdot \vec{k}}{|\vec{k}|} = 1, \\ u_{-} : & \quad \frac{\vec{\sigma} \cdot \vec{k}}{|\vec{k}|} = -1. \end{aligned} \quad (105)$$

Since the spin operator is defined as  $\vec{s} = \frac{1}{2} \vec{\sigma}$ , the previous expressions give the chirality of the states with wave function  $u_{\pm}$ , i.e. the projection of spin along the momentum of the particle. Therefore we conclude that  $u_{+}$  is a Weyl spinor of positive helicity  $\lambda = \frac{1}{2}$ , while  $u_{-}$  has negative helicity  $\lambda = -\frac{1}{2}$ . This agrees with our assertion that the representation  $(\frac{1}{2}, \mathbf{0})$  corresponds to a right-handed Weyl fermion (positive chirality) whereas  $(\mathbf{0}, \frac{1}{2})$  is a left-handed Weyl fermion (negative chirality). For example, in the standard model neutrinos are left-handed Weyl spinors and therefore transform in the representation  $(\mathbf{0}, \frac{1}{2})$  of the Lorentz group.

Nevertheless, it is possible that we were too restrictive in constructing the Weyl Lagrangian (100). There we constructed the invariants from the vector bilinears (98) corresponding to the product representations

$$\left( \frac{1}{2}, \frac{1}{2} \right) = \left( \frac{1}{2}, \mathbf{0} \right) \otimes \left( \mathbf{0}, \frac{1}{2} \right) \quad \text{and} \quad \left( \frac{1}{2}, \frac{1}{2} \right) = \left( \mathbf{0}, \frac{1}{2} \right) \otimes \left( \frac{1}{2}, \mathbf{0} \right). \quad (106)$$

In particular our insistence in demanding the Lagrangian to be invariant under the global symmetry  $u_{\pm} \rightarrow e^{i\theta} u_{\pm}$  rules out the scalar term that appears in the product representations

$$\left( \frac{1}{2}, \mathbf{0} \right) \otimes \left( \frac{1}{2}, \mathbf{0} \right) = \left( \mathbf{1}, \mathbf{0} \right) \oplus \left( \mathbf{0}, \mathbf{0} \right), \quad \left( \mathbf{0}, \frac{1}{2} \right) \otimes \left( \mathbf{0}, \frac{1}{2} \right) = \left( \mathbf{0}, \mathbf{1} \right) \oplus \left( \mathbf{0}, \mathbf{0} \right). \quad (107)$$

The singlet representations corresponds to the antisymmetric combinations

$$\epsilon_{ab} u_{\pm}^a u_{\pm}^b, \quad (108)$$

where  $\epsilon_{ab}$  is the antisymmetric symbol  $\epsilon_{12} = -\epsilon_{21} = 1$ .

At first sight it might seem that the term (108) vanishes identically because of the antisymmetry of the  $\epsilon$ -symbol. However we should keep in mind that the spin-statistic theorem (more on this later) demands that fields with half-integer spin have to satisfy the Fermi-Dirac statistics and therefore satisfy anticommutation relations, whereas fields of integer spin follow the statistic of Bose-Einstein and, as a consequence, quantization replaces Poisson brackets by commutators. This implies that the components of the Weyl fermions  $u_{\pm}$  are anticommuting Grassmann fields

$$u_{\pm}^a u_{\pm}^b + u_{\pm}^b u_{\pm}^a = 0. \quad (109)$$

It is important to realize that, strictly speaking, fermions (i.e., objects that satisfy the Fermi-Dirac statistics) do not exist classically. The reason is that they satisfy the Pauli exclusion principle and therefore each quantum state can be occupied, at most, by one fermion. Therefore the naïve definition of the classical limit as a limit of large occupation numbers cannot be applied. Fermion field do not really make sense classically.

Since the combination (108) does not vanish and we can construct a new Lagrangian

$$\mathcal{L}_{\text{Weyl}}^{\pm} = i u_{\pm}^{\dagger} \sigma_{\pm}^{\mu} \partial_{\mu} u_{\pm} - \frac{m}{2} \epsilon_{ab} u_{\pm}^a u_{\pm}^b + \text{h.c.} \quad (110)$$

This mass term, called of Majorana type, is allowed if we do not worry about breaking the global U(1) symmetry  $u_{\pm} \rightarrow e^{i\theta} u_{\pm}$ . This is not the case, for example, of charged chiral fermions, since the Majorana mass violates the conservation of electric charge or any other gauge U(1) charge. In the standard model, however, there is no such a problem if we introduce Majorana masses for right-handed neutrinos, since they are singlet under all standard model gauge groups. Such a term will break, however, the global U(1) lepton number charge because the operator  $\epsilon_{ab} \nu_R^a \nu_R^b$  changes the lepton number by two units

**Dirac spinors.** We have seen that parity interchanges the representations  $(\frac{1}{2}, \mathbf{0})$  and  $(\mathbf{0}, \frac{1}{2})$ , i.e. it changes right-handed with left-handed fermions

$$P : u_{\pm} \longrightarrow u_{\mp}. \quad (111)$$

An obvious way to build a parity invariant theory is to introduce a pair of Weyl fermions  $u_{+}$  and  $u_{-}$ . Actually, these two fields can be combined in a single four-component spinor

$$\psi = \begin{pmatrix} u_{+} \\ u_{-} \end{pmatrix} \quad (112)$$

transforming in the reducible representation  $(\frac{1}{2}, \mathbf{0}) \oplus (\mathbf{0}, \frac{1}{2})$ .

Since now we have both  $u_{+}$  and  $u_{-}$  simultaneously at our disposal the equations of motion for  $u_{\pm}$ ,  $i\sigma_{\pm}^{\mu} \partial_{\mu} u_{\pm} = 0$  can be modified, while keeping them linear, to

$$\left. \begin{array}{l} i\sigma_{+}^{\mu} \partial_{\mu} u_{+} = m u_{-} \\ i\sigma_{-}^{\mu} \partial_{\mu} u_{-} = m u_{+} \end{array} \right\} \implies i \begin{pmatrix} \sigma_{+}^{\mu} & 0 \\ 0 & \sigma_{-}^{\mu} \end{pmatrix} \partial_{\mu} \psi = m \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \psi. \quad (113)$$

These equations of motion can be derived from the Lagrangian density

$$\mathcal{L}_{\text{Dirac}} = i \psi^{\dagger} \begin{pmatrix} \sigma_{+}^{\mu} & 0 \\ 0 & \sigma_{-}^{\mu} \end{pmatrix} \partial_{\mu} \psi - m \psi^{\dagger} \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \psi. \quad (114)$$

To simplify the notation it is useful to define the Dirac  $\gamma$ -matrices as

$$\gamma^{\mu} = \begin{pmatrix} 0 & \sigma_{-}^{\mu} \\ \sigma_{+}^{\mu} & 0 \end{pmatrix} \quad (115)$$

and the Dirac conjugate spinor  $\bar{\psi}$

$$\bar{\psi} \equiv \psi^\dagger \gamma^0 = \psi^\dagger \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}. \quad (116)$$

Now the Lagrangian (114) can be written in the more compact form

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi. \quad (117)$$

The associated equations of motion give the Dirac equation (11) with the identifications

$$\gamma^0 = \beta, \quad \gamma^i = i\alpha^i. \quad (118)$$

In addition, the  $\gamma$ -matrices defined in (115) satisfy the Clifford algebra

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}. \quad (119)$$

In  $D$  dimensions this algebra admits representations of dimension  $2^{\lfloor \frac{D}{2} \rfloor}$ . When  $D$  is even the Dirac fermions  $\psi$  transform in a reducible representation of the Lorentz group. In the case of interest,  $D = 4$  this is easy to prove by defining the matrix

$$\gamma^5 = -i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix}. \quad (120)$$

We see that  $\gamma^5$  anticommutes with all other  $\gamma$ -matrices. This implies that

$$[\gamma^5, \sigma^{\mu\nu}] = 0, \quad \text{with} \quad \sigma^{\mu\nu} = -\frac{i}{4}[\gamma^\mu, \gamma^\nu]. \quad (121)$$

Because of Schur's lemma (see Appendix) this implies that the representation of the Lorentz group provided by  $\sigma^{\mu\nu}$  is reducible into subspaces spanned by the eigenvectors of  $\gamma^5$  with the same eigenvalue. If we define the projectors  $P_\pm = \frac{1}{2}(1 \pm \gamma^5)$  these subspaces correspond to

$$P_+\psi = \begin{pmatrix} u_+ \\ 0 \end{pmatrix}, \quad P_-\psi = \begin{pmatrix} 0 \\ u_- \end{pmatrix}, \quad (122)$$

which are precisely the Weyl spinors introduced before.

Our next task is to quantize the Dirac Lagrangian. This will be done along the lines used for the Klein-Gordon field, starting with a general solution to the Dirac equation and introducing the corresponding set of creation-annihilation operators. Therefore we start by looking for a complete basis of solutions to the Dirac equation. In the case of the scalar field the elements of the basis were labelled by their four-momentum  $k^\mu$ . Now, however, we have more degrees of freedom since we are dealing with a spinor which means that we have to add extra labels. Looking back at Eq. (105) we can define the helicity operator for a Dirac spinor as

$$\lambda = \frac{1}{2} \vec{\sigma} \cdot \frac{\vec{k}}{|\vec{k}|} \begin{pmatrix} \mathbf{1} & 0 \\ 0 & \mathbf{1} \end{pmatrix}. \quad (123)$$

Hence, each element of the basis of functions is labelled by its four-momentum  $k^\mu$  and the corresponding eigenvalue  $s$  of the helicity operator. For positive energy solutions we then propose the ansatz

$$u(k, s)e^{-ik \cdot x}, \quad s = \pm \frac{1}{2}, \quad (124)$$

where  $u_\alpha(k, s)$  ( $\alpha = 1, \dots, 4$ ) is a four-component spinor. Substituting in the Dirac equation we obtain

$$(\not{k} - m)u(k, s) = 0. \quad (125)$$

In the same way, for negative energy solutions we have

$$v(k, s)e^{ik \cdot x}, \quad s = \pm \frac{1}{2}, \quad (126)$$

where  $v(k, s)$  has to satisfy

$$(\not{k} + m)v(k, s) = 0. \quad (127)$$

Multiplying Eqs. (125) and (127) on the left respectively by  $(\not{k} \mp m)$  we find that the momentum is on the mass shell,  $k^2 = m^2$ . Because of this, the wave function for both positive- and negative-energy solutions can be labeled as well using the three-momentum  $\vec{k}$  of the particle,  $u(\vec{k}, s)$ ,  $v(\vec{k}, s)$ .

A detailed analysis shows that the functions  $u(\vec{k}, s)$ ,  $v(\vec{k}, s)$  satisfy the properties

$$\begin{aligned} \bar{u}(\vec{k}, s)u(\vec{k}, s) &= 2m, & \bar{v}(\vec{k}, s)v(\vec{k}, s) &= -2m, \\ \bar{u}(\vec{k}, s)\gamma^\mu u(\vec{k}, s) &= 2k^\mu, & \bar{v}(\vec{k}, s)\gamma^\mu v(\vec{k}, s) &= 2k^\mu, \\ \sum_{s=\pm\frac{1}{2}} u_\alpha(\vec{k}, s)\bar{u}_\beta(\vec{k}, s) &= (\not{k} + m)_{\alpha\beta}, & \sum_{s=\pm\frac{1}{2}} v_\alpha(\vec{k}, s)\bar{v}_\beta(\vec{k}, s) &= (\not{k} - m)_{\alpha\beta}, \end{aligned} \quad (128)$$

with  $k^0 = \omega_k = \sqrt{\vec{k}^2 + m^2}$ . Then, a general solution to the Dirac equation including creation and annihilation operators can be written as:

$$\hat{\psi}(t, \vec{x}) = \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega_k} \sum_{s=\pm\frac{1}{2}} \left[ u(\vec{k}, s)\hat{b}(\vec{k}, s)e^{-i\omega_k t + i\vec{k} \cdot \vec{x}} + v(\vec{k}, s)\hat{d}^\dagger(\vec{k}, s)e^{i\omega_k t - i\vec{k} \cdot \vec{x}} \right]. \quad (129)$$

The operators  $\hat{b}^\dagger(\vec{k}, s)$ ,  $\hat{b}(\vec{k})$  respectively create and annihilate a spin- $\frac{1}{2}$  particle (for example, an electron) out of the vacuum with momentum  $\vec{k}$  and helicity  $s$ . Because we are dealing with half-integer spin fields, the spin-statistics theorem forces canonical anticommutation relations for  $\hat{\psi}$  which means that the creation-annihilation operators satisfy the algebra<sup>5</sup>

$$\begin{aligned} \{b(\vec{k}, s), b^\dagger(\vec{k}', s')\} &= \delta(\vec{k} - \vec{k}')\delta_{ss'}, \\ \{b(\vec{k}, s), b(\vec{k}', s')\} &= \{b^\dagger(\vec{k}, s), b^\dagger(\vec{k}', s')\} = 0. \end{aligned} \quad (130)$$

In the case of  $d(\vec{k}, s)$ ,  $d^\dagger(\vec{k}, s)$  we have a set of creation-annihilation operators for the corresponding antiparticles (for example positrons). This is clear if we notice that  $d^\dagger(\vec{k}, s)$  can be seen as the annihilation operator of a negative energy state of the Dirac equation with wave function  $v_\alpha(\vec{k}, s)$ . As we saw, in the Dirac sea picture this corresponds to the creation of an antiparticle out of the vacuum (see Fig. 2). The creation-annihilation operators for antiparticles also satisfy the fermionic algebra

$$\begin{aligned} \{d(\vec{k}, s), d^\dagger(\vec{k}', s')\} &= \delta(\vec{k} - \vec{k}')\delta_{ss'}, \\ \{d(\vec{k}, s), d(\vec{k}', s')\} &= \{d^\dagger(\vec{k}, s), d^\dagger(\vec{k}', s')\} = 0. \end{aligned} \quad (131)$$

All other anticommutators between  $b(\vec{k}, s)$ ,  $b^\dagger(\vec{k}, s)$  and  $d(\vec{k}, s)$ ,  $d^\dagger(\vec{k}, s)$  vanish.

The Hamiltonian operator for the Dirac field is

$$\hat{H} = \frac{1}{2} \sum_{s=\pm\frac{1}{2}} \int \frac{d^3k}{(2\pi)^3} \left[ b^\dagger(\vec{k}, s)b(\vec{k}, s) - d(\vec{k}, s)d^\dagger(\vec{k}, s) \right]. \quad (132)$$

At this point we realize again of the necessity of quantizing the theory using anticommutators instead of commutators. Had we use canonical commutation relations, the second term inside the integral in

<sup>5</sup>To simplify notation, and since there is no risk of confusion, we drop from now on the hat to indicate operators.

(132) would give the number operator  $d^\dagger(\vec{k}, s)d(\vec{k}, s)$  with a minus sign in front. As a consequence the Hamiltonian would be unbounded from below and we would be facing again the instability of the theory already noticed in the context of relativistic quantum mechanics. However, because of the *anticommutation* relations (131), the Hamiltonian (132) takes the form

$$\hat{H} = \sum_{s=\pm\frac{1}{2}} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega_k} \left[ \omega_k b^\dagger(\vec{k}, s)b(\vec{k}, s) + \omega_k d^\dagger(\vec{k}, s)d(\vec{k}, s) \right] - 2 \int d^3k \omega_k \delta(\vec{0}). \quad (133)$$

As with the scalar field, we find a divergent vacuum energy contribution due to the zero-point energy of the infinite number of harmonic oscillators. Unlike the Klein-Gordon field, the vacuum energy is negative. In section 9.2 we will see that in certain type of theories called supersymmetric, where the number of bosonic and fermionic degrees of freedom is the same, there is a cancellation of the vacuum energy. The divergent contribution can be removed by the normal order prescription

$$:\hat{H}: = \sum_{s=\pm\frac{1}{2}} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega_k} \left[ \omega_k b^\dagger(\vec{k}, s)b(\vec{k}, s) + \omega_k d^\dagger(\vec{k}, s)d(\vec{k}, s) \right]. \quad (134)$$

Finally, let us mention that using the Dirac equation it is easy to prove that there is a conserved four-current given by

$$j^\mu = \bar{\psi}\gamma^\mu\psi, \quad \partial_\mu j^\mu = 0. \quad (135)$$

As we will explain further in sec. 6 this current is associated to the invariance of the Dirac Lagrangian under the global phase shift  $\psi \rightarrow e^{i\theta}\psi$ . In electrodynamics the associated conserved charge

$$Q = e \int d^3x j^0 \quad (136)$$

is identified with the electric charge.

### 4.3 Gauge fields

In classical electrodynamics the basic quantities are the electric and magnetic fields  $\vec{E}$ ,  $\vec{B}$ . These can be expressed in terms of the scalar and vector potential  $(\varphi, \vec{A})$

$$\begin{aligned} \vec{E} &= -\vec{\nabla}\varphi - \frac{\partial\vec{A}}{\partial t}, \\ \vec{B} &= \vec{\nabla} \times \vec{A}. \end{aligned} \quad (137)$$

From these equations it follows that there is an ambiguity in the definition of the potentials given by the gauge transformations

$$\varphi(t, \vec{x}) \rightarrow \varphi(t, \vec{x}) + \frac{\partial}{\partial t}\epsilon(t, \vec{x}), \quad \vec{A}(t, \vec{x}) \rightarrow \vec{A}(t, \vec{x}) - \vec{\nabla}\epsilon(t, \vec{x}). \quad (138)$$

Classically  $(\varphi, \vec{A})$  are seen as only a convenient way to solve the Maxwell equations, but without physical relevance.

The equations of electrodynamics can be recast in a manifestly Lorentz invariant form using the four-vector gauge potential  $A^\mu = (\varphi, \vec{A})$  and the antisymmetric rank-two tensor:  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ . Maxwell's equations become

$$\begin{aligned} \partial_\mu F^{\mu\nu} &= j^\nu, \\ \epsilon^{\mu\nu\sigma\eta} \partial_\nu F_{\sigma\eta} &= 0, \end{aligned} \quad (139)$$

where the four-current  $j^\mu = (\rho, \vec{j})$  contains the charge density and the electric current. The field strength tensor  $F_{\mu\nu}$  and the Maxwell equations are invariant under gauge transformations (138), which in covariant form read

$$A_\mu \longrightarrow A_\mu + \partial_\mu \epsilon. \quad (140)$$

Finally, the equations of motion of charged particles are given, in covariant form, by

$$m \frac{dw^\mu}{d\tau} = e F^{\mu\nu} u_\nu, \quad (141)$$

where  $e$  is the charge of the particle and  $u^\mu(\tau)$  its four-velocity as a function of the proper time.

The physical rôle of the vector potential becomes manifest only in Quantum Mechanics. Using the prescription of minimal substitution  $\vec{p} \rightarrow \vec{p} - e\vec{A}$ , the Schrödinger equation describing a particle with charge  $e$  moving in an electromagnetic field is

$$i\partial_t \Psi = \left[ -\frac{1}{2m} \left( \vec{\nabla} - ie\vec{A} \right)^2 + e\varphi \right] \Psi. \quad (142)$$

Because of the explicit dependence on the electromagnetic potentials  $\varphi$  and  $\vec{A}$ , this equation seems to change under the gauge transformations (138). This is physically acceptable only if the ambiguity does not affect the probability density given by  $|\Psi(t, \vec{x})|^2$ . Therefore, a gauge transformation of the electromagnetic potential should amount to a change in the (unobservable) phase of the wave function. This is indeed what happens: the Schrödinger equation (142) is invariant under the gauge transformations (138) provided the phase of the wave function is transformed at the same time according to

$$\Psi(t, \vec{x}) \longrightarrow e^{-ie\epsilon(t, \vec{x})} \Psi(t, \vec{x}). \quad (143)$$

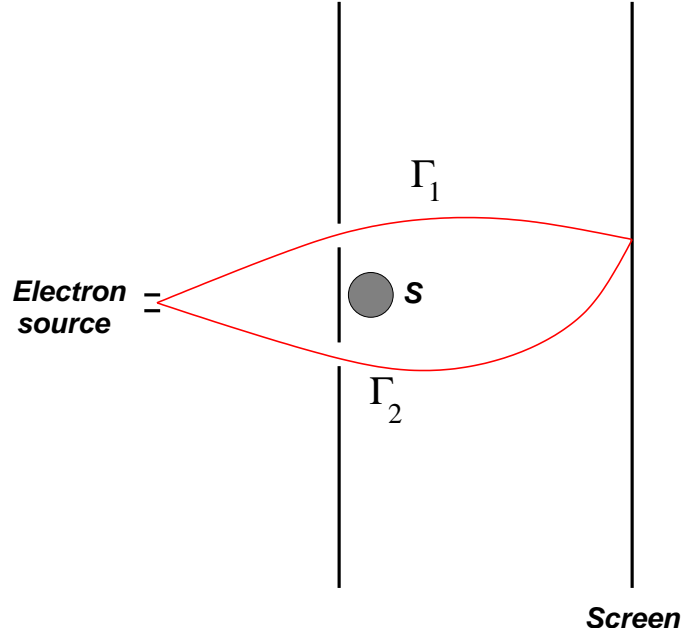
**Aharonov-Bohm effect.** This interplay between gauge transformations and the phase of the wave function give rise to surprising phenomena. The first evidence of the rôle played by the electromagnetic potentials at the quantum level was pointed out by Yakir Aharonov and David Bohm [20]. Let us consider a double slit experiment as shown in Fig. 7, where we have placed a shielded solenoid just behind the first screen. Although the magnetic field is confined to the interior of the solenoid, the vector potential is nonvanishing also outside. Of course the value of  $\vec{A}$  outside the solenoid is a pure gauge, i.e.  $\vec{\nabla} \times \vec{A} = \vec{0}$ , however because the region outside the solenoid is not simply connected the vector potential cannot be gauged to zero everywhere. If we denote by  $\Psi_1^{(0)}$  and  $\Psi_2^{(0)}$  the wave functions for each of the two electron beams in the absence of the solenoid, the total wave function once the magnetic field is switched on can be written as

$$\begin{aligned} \Psi &= e^{ie \int_{\Gamma_1} \vec{A} \cdot d\vec{x}} \Psi_1^{(0)} + e^{ie \int_{\Gamma_2} \vec{A} \cdot d\vec{x}} \Psi_2^{(0)} \\ &= e^{ie \int_{\Gamma_1} \vec{A} \cdot d\vec{x}} \left[ \Psi_1^{(0)} + e^{ie \oint_{\Gamma} \vec{A} \cdot d\vec{x}} \Psi_2^{(0)} \right], \end{aligned} \quad (144)$$

where  $\Gamma_1$  and  $\Gamma_2$  are two curves surrounding the solenoid from different sides, and  $\Gamma$  is any closed loop surrounding it. Therefore the relative phase between the two beams gets an extra term depending on the value of the vector potential outside the solenoid as

$$U = \exp \left[ ie \oint_{\Gamma} \vec{A} \cdot d\vec{x} \right]. \quad (145)$$

Because of the change in the relative phase of the electron wave functions, the presence of the vector potential becomes observable even if the electrons do not feel the magnetic field. If we perform the double-slit experiment when the magnetic field inside the solenoid is switched off we will observe the



**Fig. 7:** Illustration of an interference experiment to show the Aharonov-Bohm effect.  $S$  represent the solenoid in whose interior the magnetic field is confined.

usual interference pattern on the second screen. However if now the magnetic field is switched on, because of the phase (144), a change in the interference pattern will appear. This is the Aharonov-Bohm effect.

The first question that comes up is what happens with gauge invariance. Since we said that  $\vec{A}$  can be changed by a gauge transformation it seems that the resulting interference patterns might depend on the gauge used. Actually, the phase  $U$  in (145) is independent of the gauge although, unlike other gauge-invariant quantities like  $\vec{E}$  and  $\vec{B}$ , is nonlocal. Notice that, since  $\vec{\nabla} \times \vec{A} = \vec{0}$  outside the solenoid, the value of  $U$  does not change under continuous deformations of the closed curve  $\Gamma$ , so long as it does not cross the solenoid.

**The Dirac monopole.** It is very easy to check that the vacuum Maxwell equations remain invariant under the transformation

$$\vec{E} - i\vec{B} \longrightarrow e^{i\theta}(\vec{E} - i\vec{B}), \quad \theta \in [0, 2\pi] \quad (146)$$

which, in particular, for  $\theta = \frac{\pi}{2}$  interchanges the electric and the magnetic fields:  $\vec{E} \rightarrow \vec{B}$ ,  $\vec{B} \rightarrow -\vec{E}$ . This duality symmetry is however broken in the presence of electric sources. Nevertheless the Maxwell equations can be “completed” by introducing sources for the magnetic field  $(\rho_m, \vec{j}_m)$  in such a way that the duality (146) is restored when supplemented by the transformation

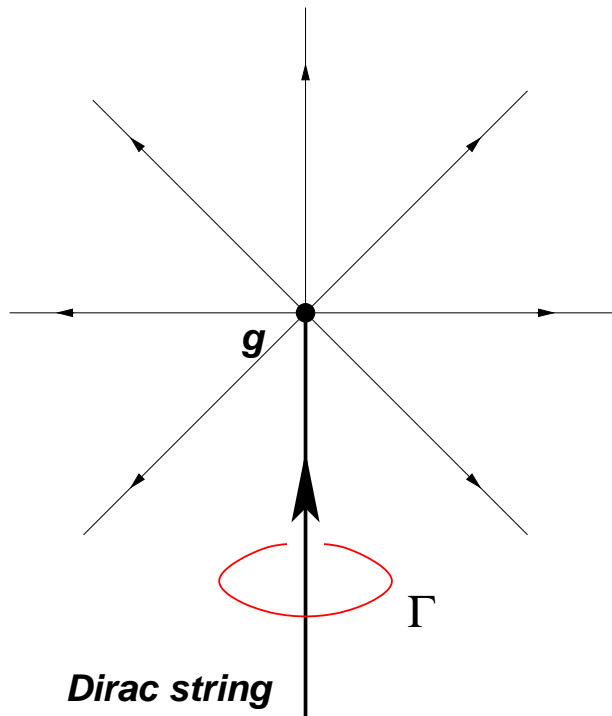
$$\rho - i\rho_m \longrightarrow e^{i\theta}(\rho - i\rho_m), \quad \vec{j} - i\vec{j}_m \longrightarrow e^{i\theta}(\vec{j} - i\vec{j}_m). \quad (147)$$

Again for  $\theta = \pi/2$  the electric and magnetic sources get interchanged.

In 1931 Dirac [21] studied the possibility of finding solutions of the completed Maxwell equation with a magnetic monopoles of charge  $g$ , i.e. solutions to

$$\vec{\nabla} \cdot \vec{B} = g \delta(\vec{x}). \quad (148)$$

Away from the position of the monopole  $\vec{\nabla} \cdot \vec{B} = 0$  and the magnetic field can be still derived locally from a vector potential  $\vec{A}$  according to  $\vec{B} = \vec{\nabla} \times \vec{A}$ . However, the vector potential cannot be regular



**Fig. 8:** The Dirac monopole.

everywhere since otherwise Gauss law would imply that the magnetic flux threading a closed surface around the monopole should vanish, contradicting (148).

We look now for solutions to Eq. (148). Working in spherical coordinates we find

$$B_r = \frac{g}{|\vec{x}|^2}, \quad B_\varphi = B_\theta = 0. \quad (149)$$

Away from the position of the monopole ( $\vec{x} \neq \vec{0}$ ) the magnetic field can be derived from the vector potential

$$A_\varphi = \frac{g}{|\vec{x}|} \tan \frac{\theta}{2}, \quad A_r = A_\theta = 0. \quad (150)$$

As expected we find that this vector potential is actually singular around the half-line  $\theta = \pi$  (see Fig. 8). This singular line starting at the position of the monopole is called the Dirac string and its position changes with a change of gauge but cannot be eliminated by any gauge transformation. Physically we can see it as an infinitely thin solenoid confining a magnetic flux entering into the magnetic monopole from infinity that equals the outgoing magnetic flux from the monopole.

Since the position of the Dirac string depends on the gauge chosen it seems that the presence of monopoles introduces an ambiguity. This would be rather strange, since Maxwell equations are gauge invariant also in the presence of magnetic sources. The solution to this apparent riddle lies in the fact that the Dirac string does not pose any consistency problem as far as it does not produce any physical effect, i.e. if its presence turns out to be undetectable. From our discussion of the Aharonov-Bohm effect we know that the wave function of charged particles pick up a phase (145) when surrounding a region where magnetic flux is confined (for example the solenoid in the Aharonov-Bohm experiment). As explained above, the Dirac string associated with the monopole can be seen as a infinitely thin solenoid. Therefore the Dirac string will be unobservable if the phase picked up by the wave function of a charged particle is equal to one. A simple calculation shows that this happens if

$$e^{ie g} = 1 \quad \implies \quad e g = 2\pi n \text{ with } n \in \mathbb{Z}. \quad (151)$$



Interestingly, this discussion leads to the conclusion that the presence of a single magnetic monopoles somewhere in the Universe implies for consistency the quantization of the electric charge in units of  $\frac{2\pi}{g}$ , where  $g$  the magnetic charge of the monopole.

**Quantization of the electromagnetic field.** We now proceed to the quantization of the electromagnetic field in the absence of sources  $\rho = 0$ ,  $\vec{j} = \vec{0}$ . In this case the Maxwell equations (139) can be derived from the Lagrangian density

$$\mathcal{L}_{\text{Maxwell}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} = \frac{1}{2}\left(\vec{E}^2 - \vec{B}^2\right). \quad (152)$$

Although in general the procedure to quantize the Maxwell Lagrangian is not very different from the one used for the Klein-Gordon or the Dirac field, here we need to deal with a new ingredient: gauge invariance. Unlike the cases studied so far, here the photon field  $A_\mu$  is not unambiguously defined because the action and the equations of motion are insensitive to the gauge transformations  $A_\mu \rightarrow A_\mu + \partial_\mu \varepsilon$ . A first consequence of this symmetry is that the theory has less physical degrees of freedom than one would expect from the fact that we are dealing with a vector field.

The way to tackle the problem of gauge invariance is to fix the freedom in choosing the electromagnetic potential before quantization. This can be done in several ways, for example by imposing the Lorentz gauge fixing condition

$$\partial_\mu A^\mu = 0. \quad (153)$$

Notice that this condition does not fix completely the gauge freedom since Eq. (153) is left invariant by gauge transformations satisfying  $\partial_\mu \partial^\mu \varepsilon = 0$ . One of the advantages, however, of the Lorentz gauge is that it is covariant and therefore does not pose any danger to the Lorentz invariance of the quantum theory. Besides, applying it to the Maxwell equation  $\partial_\mu F^{\mu\nu} = 0$  one finds

$$0 = \partial_\mu \partial^\mu A^\nu - \partial_\nu (\partial_\mu A^\mu) = \partial_\mu \partial^\mu A^\nu, \quad (154)$$

which means that since  $A_\mu$  satisfies the massless Klein-Gordon equation the photon, the quantum of the electromagnetic field, has zero mass.

Once gauge invariance is fixed  $A_\mu$  is expanded in a complete basis of solutions to (154) and the canonical commutation relations are imposed

$$\hat{A}_\mu(t, \vec{x}) = \sum_{\lambda=\pm 1} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2|\vec{k}|} \left[ \epsilon_\mu(\vec{k}, \lambda) \hat{a}(\vec{k}, \lambda) e^{-i|\vec{k}|t + i\vec{k}\cdot\vec{x}} + \epsilon_\mu(\vec{k}, \lambda)^* \hat{a}^\dagger(\vec{k}, \lambda) e^{i|\vec{k}|t - i\vec{k}\cdot\vec{x}} \right] \quad (155)$$

where  $\lambda = \pm 1$  represent the helicity of the photon, and  $\epsilon_\mu(\vec{k}, \lambda)$  are solutions to the equations of motion with well defined momentum and helicity. Because of (153) the polarization vectors have to be orthogonal to  $k_\mu$

$$k^\mu \epsilon_\mu(\vec{k}, \lambda) = k^\mu \epsilon_\mu(\vec{k}, \lambda)^* = 0. \quad (156)$$

The canonical commutation relations imply that

$$\begin{aligned} [\hat{a}(\vec{k}, \lambda), \hat{a}^\dagger(\vec{k}', \lambda')] &= (2\pi)^3 (2|\vec{k}|) \delta(\vec{k} - \vec{k}') \delta_{\lambda\lambda'} \\ [\hat{a}(\vec{k}, \lambda), \hat{a}(\vec{k}', \lambda')] &= [\hat{a}^\dagger(\vec{k}, \lambda), \hat{a}^\dagger(\vec{k}', \lambda')] = 0. \end{aligned} \quad (157)$$

Therefore  $\hat{a}(\vec{k}, \lambda)$ ,  $\hat{a}^\dagger(\vec{k}, \lambda)$  form a set of creation-annihilation operators for photons with momentum  $\vec{k}$  and helicity  $\lambda$ .

Behind the simple construction presented above there are a number of subtleties related with gauge invariance. In particular the gauge freedom seem to introduce states in the Hilbert space with negative

probability. A careful analysis shows that when gauge invariance is properly handled these spurious states decouple from physical states and can be eliminated. The details can be found in standard textbooks [1]-[11].

**Coupling gauge fields to matter.** Once we know how to quantize the electromagnetic field we consider theories containing electrically charged particles, for example electrons. To couple the Dirac Lagrangian to electromagnetism we use as guiding principle what we learned about the Schrödinger equation for a charged particle. There we saw that the gauge ambiguity of the electromagnetic potential is compensated with a  $U(1)$  phase shift in the wave function. In the case of the Dirac equation we know that the Lagrangian is invariant under  $\psi \rightarrow e^{ie\varepsilon}\psi$ , with  $\varepsilon$  a constant. However this invariance is broken as soon as one identifies  $\varepsilon$  with the gauge transformation parameter of the electromagnetic field which depends on the position.

Looking at the Dirac Lagrangian (117) it is easy to see that in order to promote the global  $U(1)$  symmetry into a local one,  $\psi \rightarrow e^{-ie\varepsilon(x)}\psi$ , it suffices to replace the ordinary derivative  $\partial_\mu$  by a covariant one  $D_\mu$  satisfying

$$D_\mu \left[ e^{-ie\varepsilon(x)}\psi \right] = e^{-ie\varepsilon(x)}D_\mu\psi. \quad (158)$$

This covariant derivative can be constructed in terms of the gauge potential  $A_\mu$  as

$$D_\mu = \partial_\mu + ieA_\mu. \quad (159)$$

The Lagrangian of a spin- $\frac{1}{2}$  field coupled to electromagnetism is written as

$$\mathcal{L}_{\text{QED}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\not{D} - m)\psi, \quad (160)$$

invariant under the gauge transformations

$$\psi \longrightarrow e^{-ie\varepsilon(x)}\psi, \quad A_\mu \longrightarrow A_\mu + \partial_\mu\varepsilon(x). \quad (161)$$

Unlike the theories we have seen so far, the Lagrangian (160) describe an interacting theory. By plugging (159) into the Lagrangian we find that the interaction between fermions and photons to be

$$\mathcal{L}_{\text{QED}}^{(\text{int})} = -eA_\mu\bar{\psi}\gamma^\mu\psi. \quad (162)$$

As advertised above, in the Dirac theory the electric current four-vector is given by  $j^\mu = e\bar{\psi}\gamma^\mu\psi$ .

The quantization of interacting field theories poses new problems that we did not meet in the case of the free theories. In particular in most cases it is not possible to solve the theory exactly. When this happens the physical observables have to be computed in perturbation theory in powers of the coupling constant. An added problem appears when computing quantum corrections to the classical result, since in that case the computation of observables are plagued with infinities that should be taken care of. We will go back to this problem in section 8.

**Nonabelian gauge theories.** Quantum electrodynamics (QED) is the simplest example of a gauge theory coupled to matter based in the abelian gauge symmetry of local  $U(1)$  phase rotations. However, it is possible also to construct gauge theories based on nonabelian groups. Actually, our knowledge of the strong and weak interactions is based on the use of such nonabelian generalizations of QED.

Let us consider a gauge group  $G$  with generators  $T^a$ ,  $a = 1, \dots, \dim G$  satisfying the Lie algebra<sup>6</sup>

$$[T^a, T^b] = if^{abc}T^c. \quad (163)$$

---

<sup>6</sup>Some basics facts about Lie groups have been summarized in Appendix A.

A gauge field taking values on the Lie algebra of  $\mathcal{G}$  can be introduced  $A_\mu \equiv A_\mu^a T^a$  which transforms under a gauge transformations as

$$A_\mu \longrightarrow -\frac{1}{ig} U \partial_\mu U^{-1} + U A_\mu U^{-1}, \quad U = e^{i\chi^a(x) T^a}, \quad (164)$$

where  $g$  is the coupling constant. The associated field strength is defined as

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c. \quad (165)$$

Notice that this definition of the  $F_{\mu\nu}^a$  reduces to the one used in QED in the abelian case when  $f^{abc} = 0$ . In general, however, unlike the case of QED the field strength is not gauge invariant. In terms of  $F_{\mu\nu} = F_{\mu\nu}^a T^a$  it transforms as

$$F_{\mu\nu} \longrightarrow U F_{\mu\nu} U^{-1}. \quad (166)$$

The coupling of matter to a nonabelian gauge field is done by introducing again a covariant derivative. For a field in a representation of  $\mathcal{G}$

$$\Phi \longrightarrow U \Phi \quad (167)$$

the covariant derivative is given by

$$D_\mu \Phi = \partial_\mu \Phi - ig A_\mu^a T^a \Phi. \quad (168)$$

With the help of this we can write a generic Lagrangian for a nonabelian gauge field coupled to scalars  $\phi$  and spinors  $\psi$  as

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}^a F^{\mu\nu a} + i \bar{\psi} \not{D} \psi + \overline{D_\mu \phi} D^\mu \phi - \bar{\psi} [M_1(\phi) + i\gamma_5 M_2(\phi)] \psi - V(\phi). \quad (169)$$

In order to keep the theory renormalizable we have to restrict  $M_1(\phi)$  and  $M_2(\phi)$  to be at most linear in  $\phi$  whereas  $V(\phi)$  have to be at most of quartic order. The Lagrangian of the standard model is of the form (169).

#### 4.4 Understanding gauge symmetry

In classical mechanics the use of the Hamiltonian formalism starts with the replacement of generalized velocities by momenta

$$p_i \equiv \frac{\partial L}{\partial \dot{q}_i} \quad \Longrightarrow \quad \dot{q}_i = \dot{q}_i(q, p). \quad (170)$$

Most of the times there is no problem in inverting the relations  $p_i = p_i(q, \dot{q})$ . However in some systems these relations might not be invertible and result in a number of constraints of the type

$$f_a(q, p) = 0, \quad a = 1, \dots, N_1. \quad (171)$$

These systems are called degenerate or constrained [23, 24].

The presence of constraints of the type (171) makes the formulation of the Hamiltonian formalism more involved. The first problem is related to the ambiguity in defining the Hamiltonian, since the addition of any linear combination of the constraints do not modify its value. Secondly, one has to make sure that the constraints are consistent with the time evolution in the system. In the language of Poisson brackets this means that further constraints have to be imposed in the form

$$\{f_a, H\} \approx 0. \quad (172)$$

Following [23] we use the symbol  $\approx$  to indicate a “weak” equality that holds when the constraints  $f_a(q, p) = 0$  are satisfied. Notice however that since the computation of the Poisson brackets involves derivatives, the constraints can be used only after the bracket is computed. In principle the conditions (172) can give rise to a new set of constraints  $g_b(q, p) = 0$ ,  $b = 1, \dots, N_2$ . Again these constraints have to be consistent with time evolution and we have to repeat the procedure. Eventually this finishes when a set of constraints is found that do not require any further constraint to be preserved by the time evolution<sup>7</sup>.

Once we find all the constraints of a degenerate system we consider the so-called first class constraints  $\phi_a(q, p) = 0$ ,  $a = 1, \dots, M$ , which are those whose Poisson bracket vanishes weakly

$$\{\phi_a, \phi_b\} = c_{abc}\phi_c \approx 0. \quad (173)$$

The constraints that do not satisfy this condition, called second class constraints, can be eliminated by modifying the Poisson bracket [23]. Then the total Hamiltonian of the theory is defined by

$$H_T = p_i q_i - L + \sum_{a=1}^M \lambda(t) \phi_a. \quad (174)$$

What has all this to do with gauge invariance? The interesting answer is that for a singular system the first class constraints  $\phi_a$  generate gauge transformations. Indeed, because  $\{\phi_a, \phi_b\} \approx 0 \approx \{\phi_a, H\}$  the transformations

$$\begin{aligned} q_i &\longrightarrow q_i + \sum_a^M \varepsilon_a(t) \{q_i, \phi_a\}, \\ p_i &\longrightarrow p_i + \sum_a^M \varepsilon_a(t) \{p_i, \phi_a\} \end{aligned} \quad (175)$$

leave invariant the state of the system. This ambiguity in the description of the system in terms of the generalized coordinates and momenta can be traced back to the equations of motion in Lagrangian language. Writing them in the form

$$\frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j} \ddot{q}_j = - \frac{\partial^2 L}{\partial \dot{q}_i \partial q_j} \dot{q}_j + \frac{\partial L}{\partial q_i}, \quad (176)$$

we find that order to determine the accelerations in terms of the positions and velocities the matrix  $\frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j}$  has to be invertible. However, the existence of constraints (171) precisely implies that the determinant of this matrix vanishes and therefore the time evolution is not uniquely determined in terms of the initial conditions.

Let us apply this to Maxwell electrodynamics described by the Lagrangian

$$L = -\frac{1}{4} \int d^3x F_{\mu\nu} F^{\mu\nu}. \quad (177)$$

The generalized momentum conjugate to  $A_\mu$  is given by

$$\pi^\mu = \frac{\delta L}{\delta(\partial_0 A_\mu)} = F^{0\mu}. \quad (178)$$

In particular for the time component we find the constraint  $\pi^0 = 0$ . The Hamiltonian is given by

$$H = \int d^3x [\pi^\mu \partial_0 A_\mu - \mathcal{L}] = \int d^3x \left[ \frac{1}{2} (\vec{E}^2 + \vec{B}^2) + \pi^0 \partial_0 A_0 + A_0 \vec{\nabla} \cdot \vec{E} \right]. \quad (179)$$

<sup>7</sup>In principle it is also possible that the procedure finishes because some kind of inconsistent identity is found. In this case the system itself is inconsistent as it is the case with the Lagrangian  $L(q, \dot{q}) = q$ .

Requiring the consistency of the constraint  $\pi^0 = 0$  we find a second constraint

$$\{\pi^0, H\} \approx \partial_0 \pi^0 + \vec{\nabla} \cdot \vec{E} = 0. \quad (180)$$

Together with the first constraint  $\pi^0 = 0$  this one implies Gauss' law  $\vec{\nabla} \cdot \vec{E} = 0$ . These two constraints have vanishing Poisson bracket and therefore they are first class. Therefore the total Hamiltonian is given by

$$H_T = H + \int d^3x \left[ \lambda_1(x) \pi^0 + \lambda_2(x) \vec{\nabla} \cdot \vec{E} \right], \quad (181)$$

where we have absorbed  $A_0$  in the definition of the arbitrary functions  $\lambda_1(x)$  and  $\lambda_2(x)$ . Actually, we can fix part of the ambiguity taking  $\lambda_1 = 0$ . Notice that, because  $A_0$  has been included in the multipliers, fixing  $\lambda_1$  amounts to fixing the value of  $A_0$  and therefore it is equivalent to taking a temporal gauge. In this case the Hamiltonian is

$$H_T = \int d^3x \left[ \frac{1}{2} (\vec{E}^2 + \vec{B}^2) + \varepsilon(x) \vec{\nabla} \cdot \vec{E} \right] \quad (182)$$

and we are left just with Gauss' law as the only constraint. Using the canonical commutation relations

$$\{A_i(t, \vec{x}), E_j(t, \vec{x}')\} = \delta_{ij} \delta(\vec{x} - \vec{x}') \quad (183)$$

we find that the remaining gauge transformations are generated by Gauss' law

$$\delta A_i = \{A_i, \int d^3x' \varepsilon \vec{\nabla} \cdot \vec{E}\} = \partial_i \varepsilon, \quad (184)$$

while leaving  $A_0$  invariant, so for consistency with the general gauge transformations the function  $\varepsilon(x)$  should be independent of time. Notice that the constraint  $\vec{\nabla} \cdot \vec{E} = 0$  can be implemented by demanding  $\vec{\nabla} \cdot \vec{A} = 0$  which reduces the three degrees of freedom of  $\vec{A}$  to the two physical degrees of freedom of the photon.

So much for the classical analysis. In the quantum theory the constraint  $\vec{\nabla} \cdot \vec{E} = 0$  has to be imposed on the physical states  $|\text{phys}\rangle$ . This is done by defining the following unitary operator on the Hilbert space

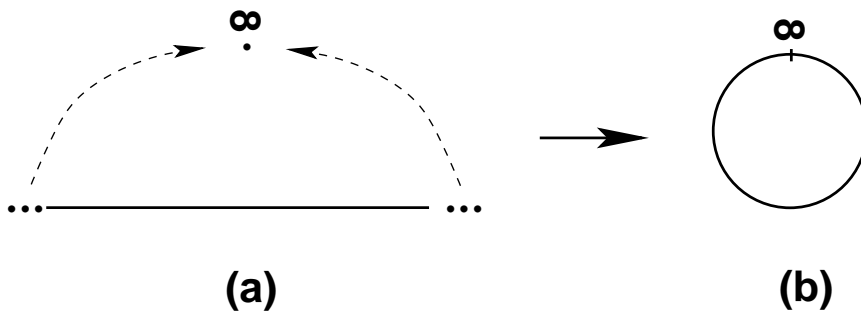
$$\mathcal{U}(\varepsilon) \equiv \exp \left( i \int d^3x \varepsilon(\vec{x}) \vec{\nabla} \cdot \vec{E} \right). \quad (185)$$

By definition, physical states should not change when a gauge transformation is performed. This is implemented by requiring that the operator  $\mathcal{U}(\varepsilon)$  acts trivially on a physical state

$$\mathcal{U}(\varepsilon) |\text{phys}\rangle = |\text{phys}\rangle \quad \Longrightarrow \quad (\vec{\nabla} \cdot \vec{E}) |\text{phys}\rangle = 0. \quad (186)$$

In the presence of charge density  $\rho$ , the condition that physical states are annihilated by Gauss' law changes to  $(\vec{\nabla} \cdot \vec{E} - \rho) |\text{phys}\rangle = 0$ .

The role of gauge transformations in the quantum theory is very illuminating in understanding the real rôle of gauge invariance [25]. As we have learned, the existence of a gauge symmetry in a theory reflects a degree of redundancy in the description of physical states in terms of the degrees of freedom appearing in the Lagrangian. In Classical Mechanics, for example, the state of a system is usually determined by the value of the canonical coordinates  $(q_i, p_i)$ . We know, however, that this is not the case for constrained Hamiltonian systems where the transformations generated by the first class constraints change the value of  $q_i$  and  $p_i$  without changing the physical state. In the case of Maxwell theory for every physical configuration determined by the gauge invariant quantities  $\vec{E}, \vec{B}$  there is an infinite number of possible values of the vector potential that are related by gauge transformations  $\delta A_\mu = \partial_\mu \varepsilon$ .



**Fig. 9:** Compactification of the real line (a) into the circumference  $S^1$  (b) by adding the point at infinity.

In the quantum theory this means that the Hilbert space of physical states is defined as the result of identifying all states related by the operator  $\mathcal{U}(\varepsilon)$  with any gauge function  $\varepsilon(x)$  into a single physical state  $|\text{phys}\rangle$ . In other words, each physical state corresponds to a whole orbit of states that are transformed among themselves by gauge transformations.

This explains the necessity of gauge fixing. In order to avoid the redundancy in the states a further condition can be given that selects one single state on each orbit. In the case of Maxwell electrodynamics the conditions  $A_0 = 0$ ,  $\vec{\nabla} \cdot \vec{A} = 0$  selects a value of the gauge potential among all possible ones giving the same value for the electric and magnetic fields.

Since states have to be identified by gauge transformations the topology of the gauge group plays an important physical rôle. To illustrate the point let us first deal with a toy model of a  $U(1)$  gauge theory in 1+1 dimensions. Later we will be more general. In the Hamiltonian formalism gauge transformations  $g(\vec{x})$  are functions defined on  $\mathbb{R}$  with values on the gauge group  $U(1)$

$$g : \mathbb{R} \longrightarrow U(1). \quad (187)$$

We assume that  $g(x)$  is regular at infinity. In this case we can add to the real line  $\mathbb{R}$  the point at infinity to compactify it into the circumference  $S^1$  (see Fig. 9). Once this is done  $g(x)$  are functions defined on  $S^1$  with values on  $U(1) = S^1$  that can be parametrized as

$$g : S^1 \longrightarrow U(1), \quad g(x) = e^{i\alpha(x)}, \quad (188)$$

with  $x \in [0, 2\pi]$ .

Because  $S^1$  does have a nontrivial topology,  $g(x)$  can be divided into topological sectors. These sectors are labelled by an integer number  $n \in \mathbb{Z}$  and are defined by

$$\alpha(2\pi) = \alpha(0) + 2\pi n. \quad (189)$$

Geometrically  $n$  gives the number of times that the spatial  $S^1$  winds around the  $S^1$  defining the gauge group  $U(1)$ . This winding number can be written in a more sophisticated way as

$$\oint_{S^1} g(x)^{-1} dg(x) = 2\pi n, \quad (190)$$

where the integral is along the spatial  $S^1$ .

In  $\mathbb{R}^3$  a similar situation happens with the gauge group<sup>8</sup>  $SU(2)$ . If we demand  $g(\vec{x}) \in SU(2)$  to be regular at infinity  $|\vec{x}| \rightarrow \infty$  we can compactify  $\mathbb{R}^3$  into a three-dimensional sphere  $S^3$ , exactly as we did in 1+1 dimensions. On the other hand, the function  $g(\vec{x})$  can be written as

$$g(\vec{x}) = a^0(x)\mathbf{1} + \vec{a}(x) \cdot \vec{\sigma} \quad (191)$$

<sup>8</sup>Although we present for simplicity only the case of  $SU(2)$ , similar arguments apply to any simple group.

and the conditions  $g(x)^\dagger g(x) = \mathbf{1}$ ,  $\det g = 1$  implies that  $(a^0)^2 + \vec{a}^2 = 1$ . Therefore  $SU(2)$  is a three-dimensional sphere and  $g(x)$  defines a function

$$g : S^3 \longrightarrow S^3. \quad (192)$$

As it was the case in 1+1 dimensions here the gauge transformations  $g(x)$  are also divided into topological sectors labelled this time by the winding number

$$n = \frac{1}{24\pi^2} \int_{S^3} d^3x \epsilon_{ijk} \text{Tr} [(g^{-1}\partial_i g) (g^{-1}\partial_j g) (g^{-1}\partial_k g)] \in \mathbb{Z}. \quad (193)$$

In the two cases analyzed we find that due to the nontrivial topology of the gauge group manifold the gauge transformations are divided into different sectors labelled by an integer  $n$ . Gauge transformations with different values of  $n$  cannot be smoothly deformed into each other. The sector with  $n = 0$  corresponds to those gauge transformations that can be connected with the identity.

Now we can be a bit more formal. Let us consider a gauge theory in 3+1 dimensions with gauge group  $G$  and let us denote by  $\mathcal{G}$  the set of all gauge transformations  $\mathcal{G} = \{g : S^3 \rightarrow G\}$ . At the same time we define  $\mathcal{G}_0$  as the set of transformations in  $\mathcal{G}$  that can be smoothly deformed into the identity. Our theory will have topological sectors if

$$\mathcal{G}/\mathcal{G}_0 \neq \mathbf{1}. \quad (194)$$

In the case of the electromagnetism we have seen that Gauss' law annihilates physical states. For a nonabelian theory the analysis is similar and leads to the condition

$$\mathcal{U}(g_0)|\text{phys}\rangle \equiv \exp \left[ i \int d^3x \chi^a(\vec{x}) \vec{\nabla} \cdot \vec{E}^a \right] |\text{phys}\rangle = |\text{phys}\rangle, \quad (195)$$

where  $g_0(\vec{x}) = e^{i\chi^a(\vec{x})T^a}$  is in the connected component of the identity  $\mathcal{G}_0$ . The important point to realize here is that only the elements of  $\mathcal{G}_0$  can be written as exponentials of the infinitesimal generators. Since these generators annihilate the physical states this implies that  $\mathcal{U}(g_0)|\text{phys}\rangle = |\text{phys}\rangle$  only when  $g_0 \in \mathcal{G}_0$ .

What happens then with the other topological sectors? If  $g \in \mathcal{G}/\mathcal{G}_0$  there is still a unitary operator  $\mathcal{U}(g)$  that realizes gauge transformations on the Hilbert space of the theory. However since  $g$  is not in the connected component of the identity, it cannot be written as the exponential of Gauss' law. Still gauge invariance is preserved if  $\mathcal{U}(g)$  only changes the overall global phase of the physical states. For example, if  $g_1$  is a gauge transformation with winding number  $n = 1$

$$\mathcal{U}(g_1)|\text{phys}\rangle = e^{i\theta}|\text{phys}\rangle. \quad (196)$$

It is easy to convince oneself that all transformations with winding number  $n = 1$  have the same value of  $\theta$  modulo  $2\pi$ . This can be shown by noticing that if  $g(\vec{x})$  has winding number  $n = 1$  then  $g(\vec{x})^{-1}$  has opposite winding number  $n = -1$ . Since the winding number is additive, given two transformations  $g_1, g_2$  with winding number 1,  $g_1^{-1}g_2$  has winding number  $n = 0$ . This implies that

$$|\text{phys}\rangle = \mathcal{U}(g_1^{-1}g_2)|\text{phys}\rangle = \mathcal{U}(g_1)^\dagger \mathcal{U}(g_2)|\text{phys}\rangle = e^{i(\theta_2 - \theta_1)}|\text{phys}\rangle \quad (197)$$

and we conclude that  $\theta_1 = \theta_2 \pmod{2\pi}$ . Once we know this it is straightforward to conclude that a gauge transformation  $g_n(\vec{x})$  with winding number  $n$  has the following action on physical states

$$\mathcal{U}(g_n)|\text{phys}\rangle = e^{in\theta}|\text{phys}\rangle, \quad n \in \mathbb{Z}. \quad (198)$$

To find a physical interpretation of this result we are going to look for similar things in other physical situations. One of them is borrowed from condensed matter physics and refers to the quantum

states of electrons in the periodic potential produced by the ion lattice in a solid. For simplicity we discuss the one-dimensional case where the minima of the potential are separated by a distance  $a$ . When the barrier between consecutive degenerate vacua is high enough we can neglect tunneling between different vacua and consider the ground state  $|na\rangle$  of the potential near the minimum located at  $x = na$  ( $n \in \mathbb{Z}$ ) as possible vacua of the theory. This vacuum state is, however, not invariant under lattice translations

$$e^{ia\hat{P}}|na\rangle = |(n+1)a\rangle. \quad (199)$$

However, it is possible to define a new vacuum state

$$|k\rangle = \sum_{n \in \mathbb{Z}} e^{-ikna} |na\rangle, \quad (200)$$

which under  $e^{ia\hat{P}}$  transforms by a global phase

$$e^{ia\hat{P}}|k\rangle = \sum_{n \in \mathbb{Z}} e^{-ikna} |(n+1)a\rangle = e^{ika} |k\rangle. \quad (201)$$

This ground state is labelled by the momentum  $k$  and corresponds to the Bloch wave function.

This looks very much the same as what we found for nonabelian gauge theories. The vacuum state labelled by  $\theta$  plays a rôle similar to the Bloch wave function for the periodic potential with the identification of  $\theta$  with the momentum  $k$ . To make this analogy more precise let us write the Hamiltonian for nonabelian gauge theories

$$H = \frac{1}{2} \int d^3x \left( \vec{\pi}_a \cdot \vec{\pi}_a + \vec{B}_a \cdot \vec{B}_a \right) = \frac{1}{2} \int d^3x \left( \vec{E}_a \cdot \vec{E}_a + \vec{B}_a \cdot \vec{B}_a \right), \quad (202)$$

where we have used the expression of the canonical momenta  $\pi_a^i$  and we assume that the Gauss' law constraint is satisfied. Looking at this Hamiltonian we can interpret the first term within the brackets as the kinetic energy  $T = \frac{1}{2} \vec{\pi}_a \cdot \vec{\pi}_a$  and the second term as the potential energy  $V = \frac{1}{2} \vec{B}_a \cdot \vec{B}_a$ . Since  $V \geq 0$  we can identify the vacua of the theory as those  $\vec{A}$  for which  $V = 0$ , modulo gauge transformations. This happens wherever  $\vec{A}$  is a pure gauge. However, since we know that the gauge transformations are labelled by the winding number we can have an infinite number of vacua which cannot be continuously connected with one another using trivial gauge transformations. Taking a representative gauge transformation  $g_n(\vec{x})$  in the sector with winding number  $n$ , these vacua will be associated with the gauge potentials

$$\vec{A} = -\frac{1}{ig} g_n(\vec{x}) \vec{\nabla} g_n(\vec{x})^{-1}, \quad (203)$$

modulo topologically trivial gauge transformations. Therefore the theory is characterized by an infinite number of vacua  $|n\rangle$  labelled by the winding number. These vacua are not gauge invariant. Indeed, a gauge transformation with  $n = 1$  will change the winding number of the vacua in one unit

$$\mathcal{U}(g_1)|n\rangle = |n+1\rangle. \quad (204)$$

Nevertheless a gauge invariant vacuum can be defined as

$$|\theta\rangle = \sum_{n \in \mathbb{Z}} e^{-in\theta} |n\rangle, \quad \text{with } \theta \in \mathbb{R} \quad (205)$$

satisfying

$$\mathcal{U}(g_1)|\theta\rangle = e^{i\theta} |\theta\rangle. \quad (206)$$



We have concluded that the nontrivial topology of the gauge group have very important physical consequences for the quantum theory. In particular it implies an ambiguity in the definition of the vacuum. Actually, this can also be seen in a Lagrangian analysis. In constructing the Lagrangian for the nonabelian version of Maxwell theory we only consider the term  $F_{\mu\nu}^a F^{\mu\nu a}$ . However this is not the only Lorentz and gauge invariant term that contains just two derivatives. We can write the more general Lagrangian

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^a F^{\mu\nu a} - \frac{\theta g^2}{32\pi^2}F_{\mu\nu}^a \tilde{F}^{\mu\nu a}, \quad (207)$$

where  $\tilde{F}_{\mu\nu}^a$  is the dual of the field strength defined by

$$\tilde{F}_{\mu\nu}^a = \frac{1}{2}\epsilon_{\mu\nu\sigma\lambda}F^{\sigma\lambda}. \quad (208)$$

The extra term in (207), proportional to  $\vec{E}^a \cdot \vec{B}^a$ , is actually a total derivative and does not change the equations of motion or the quantum perturbation theory. Nevertheless it has several important physical consequences. One of them is that it violates both parity  $P$  and the combination of charge conjugation and parity  $CP$ . This means that since strong interactions are described by a nonabelian gauge theory with group  $SU(3)$  there is an extra source of  $CP$  violation which puts a strong bound on the value of  $\theta$ . One of the consequences of a term like (207) in the QCD Lagrangian is a nonvanishing electric dipole moment for the neutron [26]. The fact that this is not observed impose a very strong bound on the value of the  $\theta$ -parameter

$$|\theta| < 10^{-9} \quad (209)$$

From a theoretical point of view it is still to be fully understood why  $\theta$  either vanishes or has a very small value.

Finally, the  $\theta$ -vacuum structure of gauge theories that we found in the Hamiltonian formalism can be also obtained using path integral techniques from the Lagrangian (207). The second term in Eq. (207) gives then a contribution that depends on the winding number of the corresponding gauge configuration.

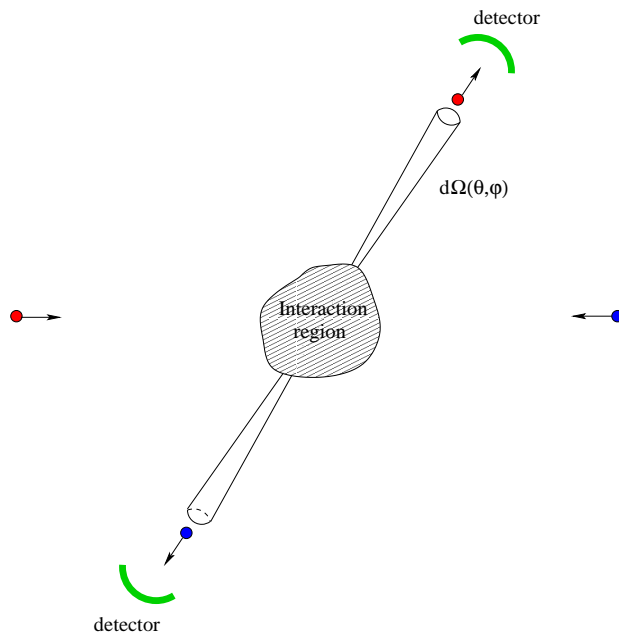
## 5 Towards computational rules: Feynman diagrams

As the basic tool to describe the physics of elementary particles, the final aim of quantum field theory is the calculation of observables. Most of the information we have about the physics of subatomic particles comes from scattering experiments. Typically, these experiments consist of arranging two or more particles to collide with a certain energy and to setup an array of detectors, sufficiently far away from the region where the collision takes place, that register the outgoing products of the collision and their momenta (together with other relevant quantum numbers).

Next we discuss how these cross sections can be computed from quantum mechanical amplitudes and how these amplitudes themselves can be evaluated in perturbative quantum field theory. We keep our discussion rather heuristic and avoid technical details that can be found in standard texts [2]- [11]. The techniques described will be illustrated with the calculation of the cross section for Compton scattering at low energies.

### 5.1 Cross sections and S-matrix amplitudes

In order to fix ideas let us consider the simplest case of a collision experiment where two particles collide to produce again two particles in the final state. The aim of such an experiments is a direct measurement of the number of particles per unit time  $\frac{dN}{dt}(\theta, \varphi)$  registered by the detector flying within a solid angle  $d\Omega$  in the direction specified by the polar angles  $\theta, \varphi$  (see Fig. 10). On general grounds we know that



**Fig. 10:** Schematic setup of a two-to-two-particles single scattering event in the center of mass reference frame.

this quantity has to be proportional to the flux of incoming particles<sup>9</sup>,  $f_{\text{in}}$ . The proportionality constant defines the differential cross section

$$\frac{dN}{dt}(\theta, \varphi) = f_{\text{in}} \frac{d\sigma}{d\Omega}(\theta, \varphi). \quad (210)$$

In natural units  $f_{\text{in}}$  has dimensions of  $(\text{length})^{-3}$ , and then the differential cross section has dimensions of  $(\text{length})^2$ . It depends, apart from the direction  $(\theta, \varphi)$ , on the parameters of the collision (energy, impact parameter, etc.) as well as on the masses and spins of the incoming particles.

Differential cross sections measure the angular distribution of the products of the collision. It is also physically interesting to quantify how effective the interaction between the particles is to produce a nontrivial dispersion. This is measured by the total cross section, which is obtained by integrating the differential cross section over all directions

$$\sigma = \int_{-1}^1 d(\cos \theta) \int_0^{2\pi} d\varphi \frac{d\sigma}{d\Omega}(\theta, \varphi). \quad (211)$$

To get some physical intuition of the meaning of the total cross section we can think of the classical scattering of a point particle off a sphere of radius  $R$ . The particle undergoes a collision only when the impact parameter is smaller than the radius of the sphere and a calculation of the total cross section yields  $\sigma = \pi R^2$ . This is precisely the cross area that the sphere presents to incoming particles.

In Quantum Mechanics in general and in quantum field theory in particular the starting point for the calculation of cross sections is the probability amplitude for the corresponding process. In a scattering experiment one prepares a system with a given number of particles with definite momenta  $\vec{p}_1, \dots, \vec{p}_n$ . In the Heisenberg picture this is described by a time independent state labelled by the incoming momenta of the particles (to keep things simple we consider spinless particles) that we denote by

$$|\vec{p}_1, \dots, \vec{p}_n; \text{in}\rangle. \quad (212)$$

<sup>9</sup>This is defined as the number of particles that enter the interaction region per unit time and per unit area perpendicular to the direction of the beam.

On the other hand, as a result of the scattering experiment a number  $k$  of particles with momenta  $\vec{p}'_1, \dots, \vec{p}'_k$  are detected. Thus, the system is now in the “out” Heisenberg picture state

$$|\vec{p}'_1, \dots, \vec{p}'_k; \text{out}\rangle \quad (213)$$

labelled by the momenta of the particles detected at late times. The probability amplitude of detecting  $k$  particles in the final state with momenta  $\vec{p}'_1, \dots, \vec{p}'_k$  in the collision of  $n$  particles with initial momenta  $\vec{p}_1, \dots, \vec{p}_n$  defines the  $S$ -matrix amplitude

$$S(\text{in} \rightarrow \text{out}) = \langle \vec{p}'_1, \dots, \vec{p}'_k; \text{out} | \vec{p}_1, \dots, \vec{p}_n; \text{in} \rangle. \quad (214)$$

It is very important to keep in mind that both the (212) and (213) are time-independent states in the Hilbert space of a very complicated interacting theory. However, since both at early and late times the incoming and outgoing particles are well apart from each other, the “in” and “out” states can be thought as two states  $|\vec{p}_1, \dots, \vec{p}_n\rangle$  and  $|\vec{p}'_1, \dots, \vec{p}'_k\rangle$  of the Fock space of the corresponding free theory in which the coupling constants are zero. Then, the overlaps (214) can be written in terms of the matrix elements of an  $S$ -matrix operator  $\widehat{S}$  acting on the free Fock space

$$\langle \vec{p}'_1, \dots, \vec{p}'_k; \text{out} | \vec{p}_1, \dots, \vec{p}_n; \text{in} \rangle = \langle \vec{p}'_1, \dots, \vec{p}'_k | \widehat{S} | \vec{p}_1, \dots, \vec{p}_n \rangle. \quad (215)$$

The operator  $\widehat{S}$  is unitary,  $\widehat{S}^\dagger = \widehat{S}^{-1}$ , and its matrix elements are analytic in the external momenta.

In any scattering experiment there is the possibility that the particles do not interact at all and the system is left in the same initial state. Then it is useful to write the  $S$ -matrix operator as

$$\widehat{S} = \mathbf{1} + i\widehat{T}, \quad (216)$$

where  $\mathbf{1}$  represents the identity operator. In this way, all nontrivial interactions are encoded in the matrix elements of the  $T$ -operator  $\langle \vec{p}'_1, \dots, \vec{p}'_k | i\widehat{T} | \vec{p}_1, \dots, \vec{p}_n \rangle$ . Since momentum has to be conserved, a global delta function can be factored out from these matrix elements to define the invariant scattering amplitude  $i\mathcal{M}$

$$\langle \vec{p}'_1, \dots, \vec{p}'_k | i\widehat{T} | \vec{p}_1, \dots, \vec{p}_n \rangle = (2\pi)^4 \delta^{(4)} \left( \sum_{\text{initial}} p_i - \sum_{\text{final}} p'_f \right) i\mathcal{M}(\vec{p}_1, \dots, \vec{p}_n; \vec{p}'_1, \dots, \vec{p}'_k) \quad (217)$$

Total and differential cross sections can be now computed from the invariant amplitudes. Here we consider the most common situation in which two particles with momenta  $\vec{p}_1$  and  $\vec{p}_2$  collide to produce a number of particles in the final state with momenta  $\vec{p}'_i$ . In this case the total cross section is given by

$$\sigma = \frac{1}{(2\omega_{p_1})(2\omega_{p_2})|\vec{v}_{12}|} \int \left[ \prod_{\text{final states}} \frac{d^3 p'_i}{(2\pi)^3} \frac{1}{2\omega_{p'_i}} \right] |\mathcal{M}_{i \rightarrow f}|^2 (2\pi)^4 \delta^{(4)} \left( p_1 + p_2 - \sum_{\text{final states}} p'_i \right), \quad (218)$$

where  $\vec{v}_{12}$  is the relative velocity of the two scattering particles. The corresponding differential cross section can be computed by dropping the integration over the directions of the final momenta. We will use this expression later in Section 5.3 to evaluate the cross section of Compton scattering.

We seen how particle cross sections are determined by the invariant amplitude for the corresponding process, i.e.  $S$ -matrix amplitudes. In general, in quantum field theory it is not possible to compute exactly these amplitudes. However, in many physical situations it can be argued that interactions are weak enough to allow for a perturbative evaluation. In what follows we will describe how  $S$ -matrix elements can be computed in perturbation theory using Feynman diagrams and rules. These are very convenient bookkeeping techniques allowing both to keep track of all contributions to a process at a given order in perturbation theory, and computing the different contributions.

## 5.2 Feynman rules

The basic quantities to be computed in quantum field theory are vacuum expectation values of products of the operators of the theory. Particularly useful are time-ordered Green functions,

$$\langle \Omega | T \left[ \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \right] | \Omega \rangle, \quad (219)$$

where  $|\Omega\rangle$  is the the ground state of the theory and the time ordered product is defined

$$T \left[ \mathcal{O}_i(x) \mathcal{O}_j(y) \right] = \theta(x^0 - y^0) \mathcal{O}_i(x) \mathcal{O}_j(y) + \theta(y^0 - x^0) \mathcal{O}_j(y) \mathcal{O}_i(x). \quad (220)$$

The generalization to products with more than two operators is straightforward: operators are always multiplied in time order, those evaluated at earlier times always to the right. The interest of these kind of correlation functions lies in the fact that they can be related to  $S$ -matrix amplitudes through the so-called reduction formula. To keep our discussion as simple as possible we will not derived it or even write it down in full detail. Its form for different theories can be found in any textbook. Here it suffices to say that the reduction formula simply states that any  $S$ -matrix amplitude can be written in terms of the Fourier transform of a time-ordered correlation function. Morally speaking

$$\begin{aligned} & \langle \vec{p}'_1, \dots, \vec{p}'_m; \text{out} | \vec{p}_1, \dots, \vec{p}_n; \text{in} \rangle \\ & \quad \Downarrow \\ & \int d^4x_1 \dots \int d^4y_n \langle \Omega | T \left[ \phi(x_1)^\dagger \dots \phi(x_m)^\dagger \phi(y_1) \dots \phi(y_n) \right] | \Omega \rangle e^{ip'_1 \cdot x_1} \dots e^{-ip_n \cdot y_n}, \end{aligned} \quad (221)$$

where  $\phi(x)$  is the field whose elementary excitations are the particles involved in the scattering.

The reduction formula reduces the problem of computing  $S$ -matrix amplitudes to that of evaluating time-ordered correlation functions of field operators. These quantities are easy to compute exactly in the free theory. For an interacting theory the situation is more complicated, however. Using path integrals, the vacuum expectation value of the time-ordered product of a number of operators can be expressed as

$$\langle \Omega | T \left[ \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \right] | \Omega \rangle = \frac{\int \mathcal{D}\phi \mathcal{D}\phi^\dagger \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) e^{iS[\phi, \phi^\dagger]}}{\int \mathcal{D}\phi \mathcal{D}\phi^\dagger e^{iS[\phi, \phi^\dagger]}}. \quad (222)$$

For an theory with interactions, neither the path integral in the numerator or in the denominator is Gaussian and they cannot be calculated exactly. However, Eq. (222) is still very useful. The action  $S[\phi, \phi^\dagger]$  can be split into the free (quadratic) piece and the interaction part

$$S[\phi, \phi^\dagger] = S_0[\phi, \phi^\dagger] + S_{\text{int}}[\phi, \phi^\dagger]. \quad (223)$$

All dependence in the coupling constants of the theory comes from the second piece. Expanding now  $\exp[iS_{\text{int}}]$  in power series of the coupling constant we find that each term in the series expansion of both the numerator and the denominator has the structure

$$\int \mathcal{D}\phi \mathcal{D}\phi^\dagger \left[ \dots \right] e^{iS_0[\phi, \phi^\dagger]}, \quad (224)$$

where “ $\dots$ ” denotes certain monomial of fields. The important point is that now the integration measure only involves the free action, and the path integral in (224) is Gaussian and therefore can be computed exactly. The same conclusion can be reached using the operator formalism. In this case the correlation function (219) can be expressed in terms of correlation functions of operators in the interaction picture. The advantage of using this picture is that the fields satisfy the free equations of motion and therefore

can be expanded in creation-annihilation operators. The correlations functions are then easily computed using Wick's theorem.

Putting together all the previous ingredients we can calculate  $S$ -matrix amplitudes in a perturbative series in the coupling constants of the field theory. This can be done using Feynman diagrams and rules, a very economical way to compute each term in the perturbative expansion of the  $S$ -matrix amplitude for a given process. We will not detail the the construction of Feynman rules but just present them heuristically.

For the sake of concreteness we focus on the case of QED first. Going back to Eq. (160) we expand the covariant derivative to write the action

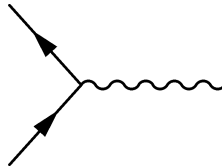
$$S_{\text{QED}} = \int d^4x \left[ -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi}(i\cancel{\partial} - m)\psi + e\bar{\psi}\gamma^\mu\psi A_\mu \right]. \quad (225)$$

The action contains two types of particles, photons and fermions, that we represent by straight and wavy lines respectively



The arrow in the fermion line does not represent the direction of the momentum but the flux of (negative) charge. This distinguishes particles from antiparticles: if the fermion propagates from left to right (i.e. in the direction of the charge flux) it represents a particle, whereas when it does from right to left it corresponds to an antiparticle. Photons are not charged and therefore wavy lines do not have orientation.

Next we turn to the interaction part of the action containing a photon field, a spinor and its conjugate. In a Feynman diagram this corresponds to the vertex



Now, in order to compute an  $S$ -matrix amplitude to a given order in the coupling constant  $e$  for a process with certain number of incoming and outgoing asymptotic states one only has to draw all possible diagrams with as many vertices as the order in perturbation theory, and the corresponding number and type of external legs. It is very important to keep in mind that in joining the fermion lines among the different building blocks of the diagram one has to respect their orientation. This reflects the conservation of the electric charge. In addition one should only consider diagrams that are topologically non-equivalent, i.e. that they cannot be smoothly deformed into one another keeping the external legs fixed<sup>10</sup>.

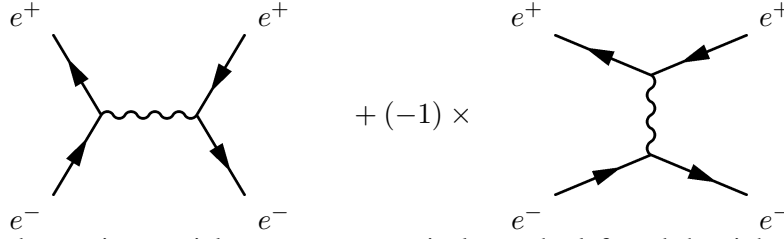
To show in a practical way how Feynman diagrams are drawn, we consider Bhabha scattering, i.e. the elastic dispersion of an electron and a positron:

$$e^+ + e^- \longrightarrow e^+ + e^-.$$

Our problem is to compute the  $S$ -matrix amplitude to the leading order in the electric charge. Because the QED vertex contains a photon line and our process does not have photons either in the initial or the

<sup>10</sup>From the point of view of the operator formalism, the requirement of considering only diagrams that are topologically nonequivalent comes from the fact that each diagram represents a certain Wick contraction in the correlation function of interaction-picture operators.

final states we find that drawing a Feynman diagram requires at least two vertices. In fact, the leading contribution is of order  $e^2$  and comes from the following two diagrams, each containing two vertices:



Incoming and outgoing particles appear respectively on the left and the right of this diagram. Notice how the identification of electrons and positrons is done comparing the direction of the charge flux with the direction of propagation. For electrons the flux of charges goes in the direction of propagation, whereas for positrons the two directions are opposite. These are the only two diagrams that can be drawn at this order in perturbation theory. It is important to include a relative minus sign between the two contributions. To understand the origin of this sign we have to remember that in the operator formalism Feynman diagrams are just a way to encode a particular Wick contraction of field operators in the interaction picture. The factor of  $-1$  reflects the relative sign in Wick contractions represented by the two diagrams, due to the fermionic character of the Dirac field.

We have learned how to draw Feynman diagrams in QED. Now one needs to compute the contribution of each one to the corresponding amplitude using the so-called Feynman rules. The idea is simple: given a diagram, each of its building blocks (vertices as well as external and internal lines) has an associated contribution that allows the calculation of the corresponding diagram. In the case of QED in the Feynman gauge, we have the following correspondence for vertices and internal propagators:

$$\begin{aligned}
 \alpha \longrightarrow \beta &\implies \left( \frac{i}{\not{p} - m + i\varepsilon} \right)_{\beta\alpha} \\
 \mu \text{ wavy } \nu &\implies \frac{-i\eta_{\mu\nu}}{p^2 + i\varepsilon} \\
 \begin{array}{c} \beta \\ \nearrow \\ \alpha \end{array} \text{ vertex } \mu &\implies -ie\gamma_{\beta\alpha}^{\mu} (2\pi)^4 \delta^{(4)}(p_1 + p_2 + p_3).
 \end{aligned}$$

A change in the gauge would reflect in an extra piece in the photon propagator. The delta function implementing conservation of momenta is written using the convention that all momenta are entering the vertex. In addition, one has to perform an integration over all momenta running in internal lines with the measure

$$\int \frac{d^d p}{(2\pi)^4}, \tag{226}$$

and introduce a factor of  $-1$  for each fermion loop in the diagram<sup>11</sup>.

<sup>11</sup>The contribution of each diagram comes also multiplied by a degeneracy factor that takes into account in how many ways a given Wick contraction can be done. In QED, however, these factors are equal to 1 for many diagrams.

In fact, some of the integrations over internal momenta can actually be done using the delta function at the vertices, leaving just a global delta function implementing the total momentum conservation in the diagram [cf. Eq. (217)]. It is even possible that all integrations can be eliminated in this way. This is the case when we have tree level diagrams, i.e. those without closed loops. In the case of diagrams with loops there will be as many remaining integrations as the number of independent loops in the diagram.

The need to perform integrations over internal momenta in loop diagrams has important consequences in Quantum Field Theory. The reason is that in many cases the resulting integrals are ill-defined, i.e. are divergent either at small or large values of the loop momenta. In the first case one speaks of *infrared divergences* and usually they cancel once all contributions to a given process are added together. More profound, however, are the divergences appearing at large internal momenta. These *ultraviolet divergences* cannot be cancelled and have to be dealt through the renormalization procedure. We will discuss this problem in some detail in Section 8.

Were we computing time-ordered (amputated) correlation function of operators, this would be all. However, in the case of  $S$ -matrix amplitudes this is not the whole story. In addition to the previous rules here one needs to attach contributions also to the external legs in the diagram. These are the wave functions of the corresponding asymptotic states containing information about the spin and momenta of the incoming and outgoing particles. In the case of QED these contributions are:

Incoming fermion:  $\alpha \longrightarrow \text{circle with diagonal lines} \implies u_\alpha(\vec{p}, s)$

Incoming antifermion:  $\alpha \longleftarrow \text{circle with diagonal lines} \implies \bar{v}_\alpha(\vec{p}, s)$

Outgoing fermion:  $\text{circle with diagonal lines} \longrightarrow \alpha \implies \bar{u}_\alpha(\vec{p}, s)$

Outgoing antifermion:  $\text{circle with diagonal lines} \longleftarrow \alpha \implies v_\alpha(p, s)$

Incoming photon:  $\mu \text{ wavy line} \text{ circle with diagonal lines} \implies \epsilon_\mu(\vec{k}, \lambda)$

Outgoing photon:  $\text{circle with diagonal lines} \text{ wavy line} \mu \implies \epsilon_\mu(\vec{k}, \lambda)^*$

Here we have assumed that the momenta for incoming (resp. outgoing) particles are entering (resp. leaving) the diagram. It is important also to keep in mind that in the computation of  $S$ -matrix amplitudes all external states are on-shell. In Section 5.3 we illustrate the use of the Feynman rules for QED with the case of the Compton scattering.

The application of Feynman diagrams to carry out computations in perturbation theory is extremely convenient. It provides a very useful bookkeeping technique to account for all contributions to a process at a given order in the coupling constant. This does not mean that the calculation of Feynman diagrams is an easy task. The number of diagrams contributing to the process grows very fast with the order in perturbation theory and the integrals that appear in calculating loop diagrams also get very complicated. This means that, generically, the calculation of Feynman diagrams beyond the first few orders very often requires the use of computers.

Above we have illustrated the Feynman rules with the case of QED. Similar rules can be computed for other interacting quantum field theories with scalar, vector or spinor fields. In the case of the nonabelian gauge theories introduced in Section 4.3 we have:

$$\begin{aligned}
 \alpha, i \longrightarrow \beta, j &\implies \left( \frac{i}{\not{p} - m + i\varepsilon} \right)_{\beta\alpha} \delta_{ij} \\
 \mu, a \text{ (wavy)} \nu, b &\implies \frac{-i\eta_{\mu\nu}}{p^2 + i\varepsilon} \delta^{ab} \\
 \begin{array}{l} \beta, j \\ \alpha, i \end{array} \text{ (fermion lines)} \text{ meeting a wavy line } \mu, a &\implies -ig\gamma_{\beta\alpha}^{\mu} t_{ij}^a \\
 \begin{array}{l} \sigma, c \\ \nu, b \end{array} \text{ (wavy lines)} \text{ meeting a wavy line } \mu, a &\implies g f^{abc} \left[ \eta^{\mu\nu} (p_1^{\sigma} - p_2^{\sigma}) + \text{permutations} \right] \\
 \begin{array}{l} \sigma, c \\ \mu, a \end{array} \text{ (wavy lines)} \text{ meeting a wavy line } \lambda, d \\
 \begin{array}{l} \nu, b \\ \nu, b \end{array} \text{ (wavy lines)} \text{ meeting a wavy line } \nu, b &\implies -ig^2 \left[ f^{abe} f^{cde} \left( \eta^{\mu\sigma} \eta^{\nu\lambda} - \eta^{\mu\lambda} \eta^{\nu\sigma} \right) + \text{permutations} \right]
 \end{aligned}$$

It is not our aim here to give a full and detailed description of the Feynman rules for nonabelian gauge theories. It suffices to point out that, unlike the case of QED, here the gauge fields can interact



among themselves. Indeed, the three and four gauge field vertices are a consequence of the cubic and quartic terms in the action

$$S = -\frac{1}{4} \int d^4x F_{\mu\nu}^a F^{\mu\nu a}, \quad (227)$$

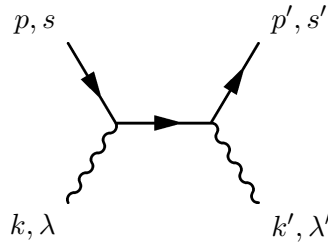
where the nonabelian gauge field strength  $F_{\mu\nu}^a$  is given in Eq. (165). The self-interaction of the non-abelian gauge fields has crucial dynamical consequences and its at the very heart of its success in describing the physics of elementary particles.

### 5.3 An example: Compton scattering

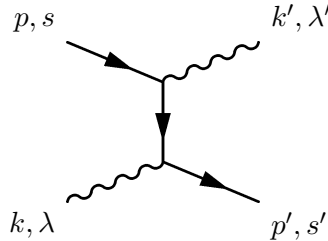
To illustrate the use of Feynman diagrams and Feynman rules we compute the cross section for the dispersion of photons by free electrons, the so-called Compton scattering:

$$\gamma(k, \lambda) + e^-(p, s) \longrightarrow \gamma(k', \lambda') + e^-(p', s').$$

In brackets we have indicated the momenta for the different particles, as well as the polarizations and spins of the incoming and outgoing photon and electrons respectively. The first step is to identify all the diagrams contributing to the process at leading order. Taking into account that the vertex of QED contains two fermion and one photon leg, it is straightforward to realize that any diagram contributing to the process at hand must contain at least two vertices. Hence the leading contribution is of order  $e^2$ . A first diagram we can draw is:



This is, however, not the only possibility. Indeed, there is a second possible diagram:



It is important to stress that these two diagrams are topologically nonequivalent, since deforming one into the other would require changing the label of the external legs. Therefore the leading  $\mathcal{O}(e^2)$  amplitude has to be computed adding the contributions from both of them.

Using the Feynman rules of QED we find

$$\begin{aligned} \text{Diagram 1} + \text{Diagram 2} &= (ie)^2 \bar{u}(\vec{p}', s') \not{\epsilon}'(\vec{k}', \lambda')^* \frac{\not{p} + \not{k} + m_e}{(p+k)^2 - m_e^2} \not{\epsilon}(\vec{k}, \lambda) u(\vec{p}, s) \\ &+ (ie)^2 \bar{u}(\vec{p}', s') \not{\epsilon}(\vec{k}, \lambda) \frac{\not{p} - \not{k}' + m_e}{(p-k')^2 - m_e^2} \not{\epsilon}'(\vec{k}', \lambda')^* u(\vec{p}, s). \end{aligned} \quad (228)$$

Because the leading order contributions only involve tree-level diagrams, there is no integration over internal momenta and therefore we are left with a purely algebraic expression for the amplitude. To get

an explicit expression we begin by simplifying the numerators. The following simple identity turns out to be very useful for this task

$$\not{a}\not{b} = -\not{b}\not{a} + 2(a \cdot b)\mathbf{1}. \quad (229)$$

Indeed, looking at the first term in Eq. (228) we have

$$\begin{aligned} (\not{p} + \not{k} + m_e)\not{\epsilon}(\vec{k}, \lambda)u(\vec{p}, s) &= -\not{\epsilon}(\vec{k}, \lambda)(\not{p} - m_e)u(\vec{p}, s) + \not{k}\not{\epsilon}(\vec{k}, \lambda)u(\vec{p}, s) \\ &+ 2p \cdot \epsilon(\vec{k}, \lambda)u(\vec{p}, s), \end{aligned} \quad (230)$$

where we have applied the identity (229) on the first term inside the parenthesis. The first term on the right-hand side of this equation vanishes identically because of Eq. (125). The expression can be further simplified if we restrict our attention to the Compton scattering at low energy when electrons are nonrelativistic. This means that all spatial momenta are much smaller than the electron mass

$$|\vec{p}|, |\vec{k}|, |\vec{p}'|, |\vec{k}'| \ll m_e. \quad (231)$$

In this approximation we have that  $p^\mu, p'^\mu \approx (m_e, \vec{0})$  and therefore

$$p \cdot \epsilon(\vec{k}, \lambda) = 0. \quad (232)$$

This follows from the absence of temporal photon polarization. Then we conclude that at low energies

$$(\not{p} + \not{k} + m_e)\not{\epsilon}(\vec{k}, \lambda)u(\vec{p}, s) = \not{k}\not{\epsilon}(\vec{k}, \lambda)u(\vec{p}, s) \quad (233)$$

and similarly for the second term in Eq. (228)

$$(\not{p} - \not{k}' + m_e)\not{\epsilon}'(\vec{k}', \lambda')^*u(\vec{p}, s) = -\not{k}'\not{\epsilon}'(\vec{k}', \lambda')^*u(\vec{p}, s). \quad (234)$$

Next, we turn to the denominators in Eq. (228). As it was explained in Section 5.2, in computing scattering amplitudes incoming and outgoing particles should have on-shell momenta,

$$p^2 = m_e^2 = p'^2 \quad \text{and} \quad k^2 = 0 = k'^2. \quad (235)$$

Then, the two denominator in Eq. (228) simplify respectively to

$$(p + k)^2 - m_e^2 = p^2 + k^2 + 2p \cdot k - m_e^2 = 2p \cdot k = 2\omega_p|\vec{k}| - 2\vec{p} \cdot \vec{k} \quad (236)$$

and

$$(p - k')^2 - m_e^2 = p^2 + k'^2 + 2p \cdot k' - m_e^2 = -2p \cdot k' = -2\omega_p|\vec{k}'| + 2\vec{p} \cdot \vec{k}'. \quad (237)$$

Working again in the low energy approximation (231) these two expressions simplify to

$$(p + k)^2 - m_e^2 \approx 2m_e|\vec{k}|, \quad (p - k')^2 - m_e^2 \approx -2m_e|\vec{k}'|. \quad (238)$$

Putting together all these expressions we find that at low energies

$$\begin{aligned} &\text{[Diagram 1]} + \text{[Diagram 2]} \\ &\approx \frac{(ie)^2}{2m_e} \bar{u}(\vec{p}', s') \left[ \not{\epsilon}'(\vec{k}', \lambda')^* \frac{\not{k}}{|\vec{k}|} \epsilon(\vec{k}, \lambda) + \epsilon(\vec{k}, \lambda) \frac{\not{k}'}{|\vec{k}'|} \not{\epsilon}'(\vec{k}', \lambda')^* \right] u(\vec{p}, s). \end{aligned} \quad (239)$$

Using now again the identity (229) a number of times as well as the transversality condition of the polarization vectors (156) we end up with a handier equation

$$\begin{aligned}
 \text{Diagram 1} + \text{Diagram 2} &\approx \frac{e^2}{m_e} \left[ \epsilon(\vec{k}, \lambda) \cdot \epsilon'(\vec{k}', \lambda')^* \right] \bar{u}(\vec{p}', s') \frac{\not{k}}{|\vec{k}|} u(\vec{p}, s) \\
 &+ \frac{e^2}{2m_e} \bar{u}(\vec{p}', s') \not{\epsilon}(\vec{k}, \lambda) \not{\epsilon}'(\vec{k}', \lambda')^* \left( \frac{\not{k}}{|\vec{k}|} - \frac{\not{k}'}{|\vec{k}'|} \right) u(\vec{p}, s). \quad (240)
 \end{aligned}$$

With a little bit of effort we can show that the second term on the right-hand side vanishes. First we notice that in the low energy limit  $|\vec{k}| \approx |\vec{k}'|$ . If in addition we make use the conservation of momentum  $k - k' = p' - p$  and the identity (125)

$$\begin{aligned}
 \bar{u}(\vec{p}', s') \not{\epsilon}(\vec{k}, \lambda) \not{\epsilon}'(\vec{k}', \lambda')^* \left( \frac{\not{k}}{|\vec{k}|} - \frac{\not{k}'}{|\vec{k}'|} \right) u(\vec{p}, s) \\
 \approx \frac{1}{|\vec{k}|} \bar{u}(\vec{p}', s') \not{\epsilon}(\vec{k}, \lambda) \not{\epsilon}'(\vec{k}', \lambda')^* (\not{p}' - m_e) u(\vec{p}, s). \quad (241)
 \end{aligned}$$

Next we use the identity (229) to take the term  $(\not{p}' - m_e)$  to the right. Taking into account that in the low energy limit the electron four-momenta are orthogonal to the photon polarization vectors [see Eq. (232)] we conclude that

$$\begin{aligned}
 \bar{u}(\vec{p}', s') \not{\epsilon}(\vec{k}, \lambda) \not{\epsilon}'(\vec{k}', \lambda')^* (\not{p}' - m_e) u(\vec{p}, s) \\
 = \bar{u}(\vec{p}', s') (\not{p}' - m_e) \not{\epsilon}(\vec{k}, \lambda) \not{\epsilon}'(\vec{k}', \lambda')^* u(\vec{p}, s) = 0 \quad (242)
 \end{aligned}$$

where the last identity follows from the equation satisfied by the conjugate positive-energy spinor,  $\bar{u}(\vec{p}', s') (\not{p}' - m_e) = 0$ .

After all these lengthy manipulations we have finally arrived at the expression of the invariant amplitude for the Compton scattering at low energies

$$i\mathcal{M} = \frac{e^2}{m_e} \left[ \epsilon(\vec{k}, \lambda) \cdot \epsilon'(\vec{k}', \lambda')^* \right] \bar{u}(\vec{p}', s') \frac{\not{k}}{|\vec{k}|} u(\vec{p}, s). \quad (243)$$

The calculation of the cross section involves computing the modulus squared of this quantity. For many physical applications, however, one is interested in the dispersion of photons with a given polarization by electrons that are not polarized, i.e. whose spins are randomly distributed. In addition in many situations either we are not interested, or there is no way to measure the final polarization of the outgoing electron. This is for example the situation in cosmology, where we do not have any information about the polarization of the free electrons in the primordial plasma before or after the scattering with photons (although we have ways to measure the polarization of the scattered photons).

To describe this physical situations we have to average over initial electron polarization (since we do not know them) and sum over all possible final electron polarization (because our detector is blind to this quantum number),

$$\overline{|i\mathcal{M}|^2} = \frac{1}{2} \left( \frac{e^2}{m_e |\vec{k}|} \right)^2 \left| \epsilon(\vec{k}, \lambda) \cdot \epsilon'(\vec{k}', \lambda')^* \right|^2 \sum_{s=\pm\frac{1}{2}} \sum_{s'=\pm\frac{1}{2}} \left| \bar{u}(\vec{p}', s') \not{k} u(\vec{p}, s) \right|^2. \quad (244)$$

The factor of  $\frac{1}{2}$  comes from averaging over the two possible polarizations of the incoming electrons. The sums in this expression can be calculated without much difficulty. Expanding the absolute value explicitly

$$\sum_{s=\pm\frac{1}{2}} \sum_{s'=\pm\frac{1}{2}} \left| \bar{u}(\vec{p}', s') \not{k} u(\vec{p}, s) \right|^2 = \sum_{s=\pm\frac{1}{2}} \sum_{s'=\pm\frac{1}{2}} \left[ u(\vec{p}, s)^\dagger \not{k}^\dagger \bar{u}(\vec{p}', s')^\dagger \right] \left[ \bar{u}(\vec{p}', s') \not{k} u(\vec{p}, s) \right], \quad (245)$$

using that  $\gamma^{\mu\dagger} = \gamma^0 \gamma^\mu \gamma^0$  and after some manipulation one finds that

$$\begin{aligned} \sum_{s=\pm\frac{1}{2}} \sum_{s'=\pm\frac{1}{2}} \left| \bar{u}(\vec{p}', s') \not{k} u(\vec{p}, s) \right|^2 &= \left[ \sum_{s=\pm\frac{1}{2}} u_\alpha(\vec{p}, s) \bar{u}_\beta(\vec{p}, s) \right] (\not{k})_{\beta\sigma} \left[ \sum_{s'=\pm\frac{1}{2}} u_\sigma(\vec{p}', s') \bar{u}_\rho(\vec{p}', s') \right] (\not{k})_{\rho\alpha} \\ &= \text{Tr} \left[ (\not{p} + m_e) \not{k} (\not{p}' + m_e) \not{k} \right], \end{aligned} \quad (246)$$

where the final expression has been computed using the completeness relations in Eq. (128). The final evaluation of the trace can be done using the standard Dirac matrices identities. Here we compute it applying again the relation (229) to commute  $\not{p}'$  and  $\not{k}$ . Using that  $k^2 = 0$  and that we are working in the low energy limit we have<sup>12</sup>

$$\text{Tr} \left[ (\not{p} + m_e) \not{k} (\not{p}' + m_e) \not{k} \right] = 2(p \cdot k)(p' \cdot k) \text{Tr} \mathbf{1} \approx 8m_e^2 |\vec{k}|^2. \quad (247)$$

This gives the following value for the invariant amplitude

$$|\overline{i\mathcal{M}}|^2 = 4e^4 \left| \epsilon(\vec{k}, \lambda) \cdot \epsilon'(\vec{k}', \lambda')^* \right|^2 \quad (248)$$

Plugging  $|\overline{i\mathcal{M}}|^2$  into the formula for the differential cross section we get

$$\frac{d\sigma}{d\Omega} = \frac{1}{64\pi^2 m_e^2} |\overline{i\mathcal{M}}|^2 = \left( \frac{e^2}{4\pi m_e} \right)^2 \left| \epsilon(\vec{k}, \lambda) \cdot \epsilon'(\vec{k}', \lambda')^* \right|^2. \quad (249)$$

The prefactor of the last equation is precisely the square of the so-called classical electron radius  $r_{\text{cl}}$ . In fact, the previous differential cross section can be rewritten as

$$\frac{d\sigma}{d\Omega} = \frac{3}{8\pi} \sigma_T \left| \epsilon(\vec{k}, \lambda) \cdot \epsilon'(\vec{k}', \lambda')^* \right|^2, \quad (250)$$

where  $\sigma_T$  is the total Thomson cross section

$$\sigma_T = \frac{e^4}{6\pi m_e^2} = \frac{8\pi}{3} r_{\text{cl}}^2. \quad (251)$$

The result (250) is relevant in many areas of Physics, but its importance is paramount in the study of the cosmological microwave background (CMB). Just before recombination the universe is filled by a plasma of electrons interacting with photons via Compton scattering, with temperatures of the order of 1 keV. Electrons are then nonrelativistic ( $m_e \sim 0.5$  MeV) and the approximations leading to Eq. (250) are fully valid. Because we do not know the polarization state of the photons before being scattered by electrons we have to consider the cross section averaged over incoming photon polarizations. From Eq. (250) we see that this is proportional to

$$\frac{1}{2} \sum_{\lambda=1,2} \left| \epsilon(\vec{k}, \lambda) \cdot \epsilon'(\vec{k}', \lambda')^* \right|^2 = \left[ \frac{1}{2} \sum_{\lambda=1,2} \epsilon_i(\vec{k}, \lambda) \epsilon_j(\vec{k}, \lambda)^* \right] \epsilon_j(\vec{k}', \lambda') \epsilon_i(\vec{k}', \lambda')^*. \quad (252)$$

The sum inside the brackets can be computed using the normalization of the polarization vectors,  $|\vec{\epsilon}(\vec{k}, \lambda)|^2 = 1$ , and the transversality condition  $\vec{k} \cdot \vec{\epsilon}(\vec{k}, \lambda) = 0$

$$\frac{1}{2} \sum_{\lambda=1,2} \left| \epsilon(\vec{k}, \lambda) \cdot \epsilon'(\vec{k}', \lambda')^* \right|^2 = \frac{1}{2} \left( \delta_{ij} - \frac{k_i k_j}{|\vec{k}|^2} \right) \epsilon_j(\vec{k}', \lambda') \epsilon_i(\vec{k}', \lambda')^*$$

<sup>12</sup>We use also the fact that the trace of the product of an odd number of Dirac matrices is always zero.

$$= \frac{1}{2} \left[ 1 - |\vec{\ell} \cdot \vec{\epsilon}'(\vec{k}', \lambda')|^2 \right], \quad (253)$$

where  $\vec{\ell} = \frac{\vec{k}}{|\vec{k}|}$  is the unit vector in the direction of the incoming photon.

From the last equation we conclude that Thomson scattering suppresses all polarizations parallel to the direction of the incoming photon  $\vec{\ell}$ , whereas the differential cross section reaches the maximum in the plane normal to  $\vec{\ell}$ . If photons would collide with the electrons in the plasma with the same intensity from all directions, the result would be an unpolarized CMB radiation. The fact that polarization is actually measured in the CMB carries crucial information about the physics of the plasma before recombination and, as a consequence, about the very early universe (see for example [22] for a throughout discussion).

## 6 Symmetries

### 6.1 Noether's theorem

In Classical Mechanics and Classical Field Theory there is a basic result that relates symmetries and conserved charges. This is called Noether's theorem and states that for each continuous symmetry of the system there is conserved current. In its simplest version in Classical Mechanics it can be easily proved. Let us consider a Lagrangian  $L(q_i, \dot{q}_i)$  which is invariant under a transformation  $q_i(t) \rightarrow q'_i(t, \epsilon)$  labelled by a parameter  $\epsilon$ . This means that  $L(q', \dot{q}') = L(q, \dot{q})$  without using the equations of motion<sup>13</sup>. If  $\epsilon \ll 1$  we can consider an infinitesimal variation of the coordinates  $\delta_\epsilon q_i(t)$  and the invariance of the Lagrangian implies

$$0 = \delta_\epsilon L(q_i, \dot{q}_i) = \frac{\partial L}{\partial q_i} \delta_\epsilon q_i + \frac{\partial L}{\partial \dot{q}_i} \delta_\epsilon \dot{q}_i = \left[ \frac{\partial L}{\partial q_i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} \right] \delta_\epsilon q_i + \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_i} \delta_\epsilon q_i \right). \quad (254)$$

When  $\delta_\epsilon q_i$  is applied on a solution to the equations of motion the term inside the square brackets vanishes and we conclude that there is a conserved quantity

$$\dot{Q} = 0 \quad \text{with} \quad Q \equiv \frac{\partial L}{\partial \dot{q}_i} \delta_\epsilon q_i. \quad (255)$$

Notice that in this derivation it is crucial that the symmetry depends on a continuous parameter since otherwise the infinitesimal variation of the Lagrangian in Eq. (254) does not make sense.

In Classical Field Theory a similar result holds. Let us consider for simplicity a theory of a single field  $\phi(x)$ . We say that the variations  $\delta_\epsilon \phi$  depending on a continuous parameter  $\epsilon$  are a symmetry of the theory if, without using the equations of motion, the Lagrangian density changes by

$$\delta_\epsilon \mathcal{L} = \partial_\mu K^\mu. \quad (256)$$

If this happens then the action remains invariant and so do the equations of motion. Working out now the variation of  $\mathcal{L}$  under  $\delta_\epsilon \phi$  we find

$$\partial_\mu K^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \partial_\mu \delta_\epsilon \phi + \frac{\partial \mathcal{L}}{\partial \phi} \delta_\epsilon \phi = \partial_\mu \left( \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \delta_\epsilon \phi \right) + \left[ \frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left( \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) \right] \delta_\epsilon \phi. \quad (257)$$

If  $\phi(x)$  is a solution to the equations of motion the last terms disappears, and we find that there is a conserved current

$$\partial_\mu J^\mu = 0 \quad \text{with} \quad J^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \delta_\epsilon \phi - K^\mu. \quad (258)$$

---

<sup>13</sup>The following result can be also derived a more general situations where the Lagrangian changes by a total time derivative.

Actually a conserved current implies the existence of a charge

$$Q \equiv \int d^3x J^0(t, \vec{x}) \quad (259)$$

which is conserved

$$\frac{dQ}{dt} = \int d^3x \partial_0 J^0(t, \vec{x}) = - \int d^3x \partial_i J^i(t, \vec{x}) = 0, \quad (260)$$

provided the fields vanish at infinity fast enough. Moreover, the conserved charge  $Q$  is a Lorentz scalar. After canonical quantization the charge  $Q$  defined by Eq. (259) is promoted to an operator that generates the symmetry on the fields

$$\delta\phi = i[\phi, Q]. \quad (261)$$

As an example we can consider a scalar field  $\phi(x)$  which under a coordinate transformation  $x \rightarrow x'$  changes as  $\phi'(x') = \phi(x)$ . In particular performing a space-time translation  $x^{\mu'} = x^\mu + a^\mu$  we have

$$\phi'(x) - \phi(x) = -a^\mu \partial_\mu \phi + \mathcal{O}(a^2) \quad \Longrightarrow \quad \delta\phi = -a^\mu \partial_\mu \phi. \quad (262)$$

Since the Lagrangian density is also a scalar quantity, it transforms under translations as

$$\delta\mathcal{L} = -a^\mu \partial_\mu \mathcal{L}. \quad (263)$$

Therefore the corresponding conserved charge is

$$J^\mu = -\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} a^\nu \partial_\nu \phi + a^\mu \mathcal{L} \equiv -a_\nu T^{\mu\nu}, \quad (264)$$

where we introduced the energy-momentum tensor

$$T^{\mu\nu} = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \partial^\nu \phi - \eta^{\mu\nu} \mathcal{L}. \quad (265)$$

We find that associated with the invariance of the theory with respect to space-time translations there are four conserved currents defined by  $T^{\mu\nu}$  with  $\nu = 0, \dots, 3$ , each one associated with the translation along a space-time direction. These four currents form a rank-two tensor under Lorentz transformations satisfying

$$\partial_\mu T^{\mu\nu} = 0. \quad (266)$$

The associated conserved charges are given by

$$P^\nu = \int d^3x T^{0\nu} \quad (267)$$

and correspond to the total energy-momentum content of the field configuration. Therefore the energy density of the field is given by  $T^{00}$  while  $T^{0i}$  is the momentum density. In the quantum theory the  $P^\mu$  are the generators of space-time translations.

Another example of a symmetry related with a physically relevant conserved charge is the global phase invariance of the Dirac Lagrangian (117),  $\psi \rightarrow e^{i\theta}\psi$ . For small  $\theta$  this corresponds to variations  $\delta_\theta\psi = i\theta\psi$ ,  $\delta_\theta\bar{\psi} = -i\theta\bar{\psi}$  which by Noether's theorem result in the conserved charge

$$j^\mu = \bar{\psi}\gamma^\mu\psi, \quad \partial_\mu j^\mu = 0. \quad (268)$$

Thus implying the existence of a conserved charge

$$Q = \int d^3x \bar{\psi} \gamma^0 \psi = \int d^3x \psi^\dagger \psi. \quad (269)$$

In physics there are several instances of global U(1) symmetries that act as phase shifts on spinors. This is the case, for example, of the baryon and lepton number conservation in the standard model. A more familiar case is the U(1) local symmetry associated with electromagnetism. Notice that although in this case we are dealing with a local symmetry,  $\theta \rightarrow e\alpha(x)$ , the invariance of the Lagrangian holds in particular for global transformations and therefore there is a conserved current  $j^\mu = e\bar{\psi}\gamma^\mu\psi$ . In Eq. (162) we saw that the spinor is coupled to the photon field precisely through this current. Its time component is the electric charge density  $\rho$ , while the spatial components are the current density vector  $\vec{j}$ .

This analysis can be carried over also to nonabelian unitary global symmetries acting as

$$\psi_i \longrightarrow U_{ij} \psi_j, \quad U^\dagger U = \mathbf{1} \quad (270)$$

and leaving invariant the Dirac Lagrangian when we have several fermions. If we write the matrix  $U$  in terms of the hermitian group generators  $T^a$  as

$$U = \exp(i\alpha_a T^a), \quad (T^a)^\dagger = T^a, \quad (271)$$

we find the conserved current

$$j^{\mu a} = \bar{\psi}_i T_{ij}^a \gamma^\mu \psi_j, \quad \partial_\mu j^\mu = 0. \quad (272)$$

This is the case, for example of the approximate flavor symmetries in hadron physics. The simplest example is the isospin symmetry that mixes the quarks  $u$  and  $d$

$$\begin{pmatrix} u \\ d \end{pmatrix} \longrightarrow M \begin{pmatrix} u \\ d \end{pmatrix}, \quad M \in \text{SU}(2). \quad (273)$$

Since the proton is a bound state of two quarks  $u$  and one quark  $d$  while the neutron is made out of one quark  $u$  and two quarks  $d$ , this isospin symmetry reduces at low energies to the well known isospin transformations of nuclear physics that mixes protons and neutrons.

## 6.2 Symmetries in the quantum theory

We have seen that in canonical quantization the conserved charges  $Q^a$  associated to symmetries by Noether's theorem are operators implementing the symmetry at the quantum level. Since the charges are conserved they must commute with the Hamiltonian

$$[Q^a, H] = 0. \quad (274)$$

There are several possibilities in the quantum mechanical realization of a symmetry:

**Wigner-Weyl realization.** In this case the ground state of the theory  $|0\rangle$  is invariant under the symmetry. Since the symmetry is generated by  $Q^a$  this means that

$$\mathcal{U}(\alpha)|0\rangle \equiv e^{i\alpha_a Q^a}|0\rangle = |0\rangle \implies Q^a|0\rangle = 0. \quad (275)$$

At the same time the fields of the theory have to transform according to some irreducible representation of the group generated by the  $Q^a$ . From Eq. (261) it is easy to prove that

$$\mathcal{U}(\alpha)\phi_i\mathcal{U}(\alpha)^{-1} = U_{ij}(\alpha)\phi_j, \quad (276)$$

where  $U_{ij}(\alpha)$  is an element of the representation in which the field  $\phi_i$  transforms. If we consider now the quantum state associated with the operator  $\phi_i$

$$|i\rangle = \phi_i|0\rangle \quad (277)$$

we find that because of the invariance of the vacuum (275) the states  $|i\rangle$  transform in the same representation as  $\phi_i$

$$\mathcal{U}(\alpha)|i\rangle = \mathcal{U}(\alpha)\phi_i\mathcal{U}(\alpha)^{-1}\mathcal{U}(\alpha)|0\rangle = U_{ij}(\alpha)\phi_j|0\rangle = U_{ij}(\alpha)|j\rangle. \quad (278)$$

Therefore the spectrum of the theory is classified in multiplets of the symmetry group. In addition, since  $[H, \mathcal{U}(\alpha)] = 0$  all states in the same multiplet have the same energy. If we consider one-particle states, then going to the rest frame we conclude that all states in the same multiplet have exactly the same mass.

**Nambu-Goldstone realization.** In our previous discussion the result that the spectrum of the theory is classified according to multiplets of the symmetry group depended crucially on the invariance of the ground state. However this condition is not mandatory and one can relax it to consider theories where the vacuum state is not left invariant by the symmetry

$$e^{i\alpha_a Q^a}|0\rangle \neq |0\rangle \quad \implies \quad Q^a|0\rangle \neq 0. \quad (279)$$

In this case it is also said that the symmetry is spontaneously broken by the vacuum.

To illustrate the consequences of (279) we consider the example of a number scalar fields  $\varphi^i$  ( $i = 1, \dots, N$ ) whose dynamics is governed by the Lagrangian

$$\mathcal{L} = \frac{1}{2}\partial_\mu\varphi^i\partial^\mu\varphi^i - V(\varphi), \quad (280)$$

where we assume that  $V(\phi)$  is bounded from below. This theory is globally invariant under the transformations

$$\delta\varphi^i = \epsilon^a (T^a)^i_j \varphi^j, \quad (281)$$

with  $T^a$ ,  $a = 1, \dots, \frac{1}{2}N(N-1)$  the generators of the group  $\text{SO}(N)$ .

To analyze the structure of vacua of the theory we construct the Hamiltonian

$$H = \int d^3x \left[ \frac{1}{2}\pi^i\pi^i + \frac{1}{2}\vec{\nabla}\varphi^i \cdot \vec{\nabla}\varphi^i + V(\varphi) \right] \quad (282)$$

and look for the minimum of

$$\mathcal{V}(\varphi) = \int d^3x \left[ \frac{1}{2}\vec{\nabla}\varphi^i \cdot \vec{\nabla}\varphi^i + V(\varphi) \right]. \quad (283)$$

Since we are interested in finding constant field configurations,  $\vec{\nabla}\varphi = \vec{0}$  to preserve translational invariance, the vacua of the potential  $\mathcal{V}(\varphi)$  coincides with the vacua of  $V(\varphi)$ . Therefore the minima of the potential correspond to the vacuum expectation values<sup>14</sup>

$$\langle\varphi^i\rangle : \quad V(\langle\varphi^i\rangle) = 0, \quad \left. \frac{\partial V}{\partial\varphi^i} \right|_{\varphi^i=\langle\varphi^i\rangle} = 0. \quad (284)$$

We divide the generators  $T^a$  of  $\text{SO}(N)$  into two groups: Those denoted by  $H^\alpha$  ( $\alpha = 1, \dots, h$ ) that satisfy

$$(H^\alpha)^i_j \langle\varphi^j\rangle = 0. \quad (285)$$

<sup>14</sup>For simplicity we consider that the minima of  $V(\phi)$  occur at zero potential.



This means that the vacuum configuration  $\langle \varphi^i \rangle$  is left invariant by the transformation generated by  $H^\alpha$ . For this reason we call them *unbroken generators*. Notice that the commutator of two unbroken generators also annihilates the vacuum expectation value,  $[H^\alpha, H^\beta]_{ij} \langle \varphi^j \rangle = 0$ . Therefore the generators  $\{H^\alpha\}$  form a subalgebra of the algebra of the generators of  $SO(N)$ . The subgroup of the symmetry group generated by them is realized à la Wigner-Weyl.

The remaining generators  $K^A$ , with  $A = 1, \dots, \frac{1}{2}N(N-1) - h$ , by definition do not preserve the vacuum expectation value of the field

$$(K^A)^i_j \langle \varphi^j \rangle \neq 0. \quad (286)$$

These will be called the *broken generators*. Next we prove a very important result concerning the broken generators known as the Goldstone theorem: for each generator broken by the vacuum expectation value there is a massless excitation.

The mass matrix of the excitations around the vacuum  $\langle \varphi^i \rangle$  is determined by the quadratic part of the potential. Since we assumed that  $V(\langle \varphi \rangle) = 0$  and we are expanding around a minimum, the first term in the expansion of the potential  $V(\varphi)$  around the vacuum expectation values is given by

$$V(\varphi) = \frac{\partial^2 V}{\partial \varphi^i \partial \varphi^j} \Big|_{\varphi = \langle \varphi \rangle} (\varphi^i - \langle \varphi^i \rangle)(\varphi^j - \langle \varphi^j \rangle) + \mathcal{O}[(\varphi - \langle \varphi \rangle)^3] \quad (287)$$

and the mass matrix is:

$$M_{ij}^2 \equiv \frac{\partial^2 V}{\partial \varphi^i \partial \varphi^j} \Big|_{\varphi = \langle \varphi \rangle}. \quad (288)$$

In order to avoid a cumbersome notation we do not show explicitly the dependence of the mass matrix on the vacuum expectation values  $\langle \varphi^i \rangle$ .

To extract some information about the possible zero modes of the mass matrix, we write down the conditions that follow from the invariance of the potential under  $\delta \varphi^i = \epsilon^a (T^a)^i_j \varphi^j$ . At first order in  $\epsilon^a$

$$\delta V(\varphi) = \epsilon^a \frac{\partial V}{\partial \varphi^i} (T^a)^i_j \varphi^j = 0. \quad (289)$$

Differentiating this expression with respect to  $\varphi^k$  we arrive at

$$\frac{\partial^2 V}{\partial \varphi^i \partial \varphi^k} (T^a)^i_j \varphi^j + \frac{\partial V}{\partial \varphi^i} (T^a)^i_k = 0. \quad (290)$$

Now we evaluate this expression in the vacuum  $\varphi^i = \langle \varphi^i \rangle$ . Then the derivative in the second term cancels while the second derivative in the first one gives the mass matrix. Hence we find

$$M_{ik}^2 (T^a)^i_j \langle \varphi^j \rangle = 0. \quad (291)$$

Now we can write this expression for both broken and unbroken generators. For the unbroken ones, since  $(H^\alpha)^i_j \langle \varphi^j \rangle = 0$ , we find a trivial identity  $0 = 0$ . On the other hand for the broken generators we have

$$M_{ik}^2 (K^A)^i_j \langle \varphi^j \rangle = 0. \quad (292)$$

Since  $(K^A)^i_j \langle \varphi^j \rangle \neq 0$  this equation implies that the mass matrix has as many zero modes as broken generators. Therefore we have proven Goldstone's theorem: associated with each broken symmetry there is a massless mode in the theory. Here we have presented a classical proof of the theorem. In the quantum theory the proof follows the same lines as the one presented here but one has to consider the effective action containing the effects of the quantum corrections to the classical Lagrangian.

As an example to illustrate this theorem, we consider a  $SO(3)$  invariant scalar field theory with a “mexican hat” potential

$$V(\vec{\varphi}) = \frac{\lambda}{4} (\vec{\varphi}^2 - a^2)^2. \quad (293)$$

The vacua of the theory correspond to the configurations satisfying  $\langle \vec{\varphi} \rangle^2 = a^2$ . In field space this equation describes a two-dimensional sphere and each solution is just a point in that sphere. Geometrically it is easy to visualize that a given vacuum field configuration, i.e. a point in the sphere, is preserved by  $SO(2)$  rotations around the axis of the sphere that passes through that point. Hence the vacuum expectation value of the scalar field breaks the symmetry according to

$$\langle \vec{\varphi} \rangle : \quad SO(3) \longrightarrow SO(2). \quad (294)$$

Since  $SO(3)$  has three generators and  $SO(2)$  only one we see that two generators are broken and therefore there are two massless Goldstone bosons. Physically this massless modes can be thought of as corresponding to excitations along the surface of the sphere  $\langle \vec{\varphi} \rangle^2 = a^2$ .

Once a minimum of the potential has been chosen we can proceed to quantize the excitations around it. Since the vacuum only leaves invariant a  $SO(2)$  subgroup of the original  $SO(3)$  symmetry group it seems that the fact that we are expanding around a particular vacuum expectation value of the scalar field has resulted in a lost of symmetry. This is however not the case. The full quantum theory is symmetric under the whole symmetry group  $SO(3)$ . This is reflected in the fact that the physical properties of the theory do not depend on the particular point of the sphere  $\langle \vec{\varphi} \rangle^2 = a^2$  that we have chosen. Different vacua are related by the full  $SO(3)$  symmetry and therefore should give the same physics.

It is very important to realize that given a theory with a vacuum determined by  $\langle \vec{\varphi} \rangle$  all other possible vacua of the theory are inaccessible in the infinite volume limit. This means that two vacuum states  $|0_1\rangle, |0_2\rangle$  corresponding to different vacuum expectation values of the scalar field are orthogonal  $\langle 0_1|0_2\rangle = 0$  and cannot be connected by any local observable  $\Phi(x)$ ,  $\langle 0_1|\Phi(x)|0_2\rangle = 0$ . Heuristically this can be understood by noticing that in the infinite volume limit switching from one vacuum into another one requires changing the vacuum expectation value of the field everywhere in space at the same time, something that cannot be done by any local operator. Notice that this is radically different to our expectations based on the Quantum Mechanics of a system with a finite number of degrees of freedom.

In High Energy Physics the typical example of a Goldstone boson is the pion, associated with the spontaneous breaking of the global chiral isospin  $SU(2)_L \times SU(2)_R$  symmetry. This symmetry acts independently in the left- and right-handed spinors as

$$\begin{pmatrix} u_{L,R} \\ d_{L,R} \end{pmatrix} \longrightarrow M_{L,R} \begin{pmatrix} u_{L,R} \\ d_{L,R} \end{pmatrix}, \quad M_{L,R} \in SU(2)_{L,R} \quad (295)$$

Presumably since the quarks are confined at low energies this symmetry is spontaneously broken down to the diagonal  $SU(2)$  acting in the same way on the left- and right-handed components of the spinors. Associated with this symmetry breaking there is a Goldstone mode which is identified as the pion. Notice, nevertheless, that the  $SU(2)_L \times SU(2)_R$  would be an exact global symmetry of the QCD Lagrangian only in the limit when the masses of the quarks are zero  $m_u, m_d \rightarrow 0$ . Since these quarks have nonzero masses the chiral symmetry is only approximate and as a consequence the corresponding Goldstone boson is not massless. That is why pions have masses, although they are the lightest particle among the hadrons.

Symmetry breaking appears also in many places in condensed matter. For example, when a solid crystallizes from a liquid the translational invariance that is present in the liquid phase is broken to a discrete group of translations that represent the crystal lattice. This symmetry breaking has Goldstone

bosons associated which are identified with phonons which are the quantum excitation modes of the vibrational degrees of freedom of the lattice.

**The Higgs mechanism.** Gauge symmetry seems to prevent a vector field from having a mass. This is obvious once we realize that a term in the Lagrangian like  $m^2 A_\mu A^\mu$  is incompatible with gauge invariance.

However certain physical situations seem to require massive vector fields. This happened for example during the 1960s in the study of weak interactions. The Glashow model gave a common description of both electromagnetic and weak interactions based on a gauge theory with group  $SU(2) \times U(1)$  but, in order to reproduce Fermi's four-fermion theory of the  $\beta$ -decay it was necessary that two of the vector fields involved would be massive. Also in condensed matter physics massive vector fields are required to describe certain systems, most notably in superconductivity.

The way out to this situation is found in the concept of spontaneous symmetry breaking discussed previously. The consistency of the quantum theory requires gauge invariance, but this invariance can be realized à la Nambu-Goldstone. When this is the case the full gauge symmetry is not explicitly present in the effective action constructed around the particular vacuum chosen by the theory. This makes possible the existence of mass terms for gauge fields without jeopardizing the consistency of the full theory, which is still invariant under the whole gauge group.

To illustrate the Higgs mechanism we study the simplest example, the Abelian Higgs model: a  $U(1)$  gauge field coupled to a self-interacting charged complex scalar field  $\Phi$  with Lagrangian

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \overline{D}_\mu \Phi D^\mu \Phi - \frac{\lambda}{4} (\overline{\Phi} \Phi - \mu^2)^2, \quad (296)$$

where the covariant derivative is given by Eq. (159). This theory is invariant under the gauge transformations

$$\Phi \rightarrow e^{i\alpha(x)} \Phi, \quad A_\mu \rightarrow A_\mu + \partial_\mu \alpha(x). \quad (297)$$

The minimum of the potential is defined by the equation  $|\Phi| = \mu$ . We have a continuum of different vacua labelled by the phase of the scalar field. None of these vacua, however, is invariant under the gauge symmetry

$$\langle \Phi \rangle = \mu e^{i\vartheta_0} \rightarrow \mu e^{i\vartheta_0 + i\alpha(x)} \quad (298)$$

and therefore the symmetry is spontaneously broken. Let us study now the theory around one of these vacua, for example  $\langle \Phi \rangle = \mu$ , by writing the field  $\Phi$  in terms of the excitations around this particular vacuum

$$\Phi(x) = \left[ \mu + \frac{1}{\sqrt{2}} \sigma(x) \right] e^{i\vartheta(x)}. \quad (299)$$

Independently of whether we are expanding around a particular vacuum for the scalar field we should keep in mind that the whole Lagrangian is still gauge invariant under (297). This means that performing a gauge transformation with parameter  $\alpha(x) = -\vartheta(x)$  we can get rid of the phase in Eq. (299). Substituting then  $\Phi(x) = \mu + \frac{1}{\sqrt{2}} \sigma(x)$  in the Lagrangian we find

$$\begin{aligned} \mathcal{L} &= -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + e^2 \mu^2 A_\mu A^\mu + \frac{1}{2} \partial_\mu \sigma \partial^\mu \sigma - \frac{1}{2} \lambda \mu^2 \sigma^2 \\ &\quad - \lambda \mu \sigma^3 - \frac{\lambda}{4} \sigma^4 + e^2 \mu A_\mu A^\mu \sigma + e^2 A_\mu A^\mu \sigma^2. \end{aligned} \quad (300)$$

What are the excitation of the theory around the vacuum  $\langle \Phi \rangle = \mu$ ? First we find a massive real scalar field  $\sigma(x)$ . The important point however is that the vector field  $A_\mu$  now has a mass given by

$$m_\gamma^2 = 2e^2 \mu^2. \quad (301)$$

The remarkable thing about this way of giving a mass to the photon is that at no point we have given up gauge invariance. The symmetry is only hidden. Therefore in quantizing the theory we can still enjoy all the advantages of having a gauge theory but at the same time we have managed to generate a mass for the gauge field.

It is surprising, however, that in the Lagrangian (300) we did not find any massless mode. Since the vacuum chosen by the scalar field breaks the  $U(1)$  generator of  $U(1)$  we would have expected one massless particle from Goldstone's theorem. To understand the fate of the missing Goldstone boson we have to revisit the calculation leading to Eq. (300). Were we dealing with a global  $U(1)$  theory, the Goldstone boson would correspond to excitation of the scalar field along the valley of the potential and the phase  $\vartheta(x)$  would be the massless Goldstone boson. However we have to keep in mind that in computing the Lagrangian we managed to get rid of  $\vartheta(x)$  by shifting it into  $A_\mu$  using a gauge transformation. Actually by identifying the gauge parameter with the Goldstone excitation we have completely fixed the gauge and the Lagrangian (300) does not have any gauge symmetry left.

A massive vector field has three polarizations: two transverse ones  $\vec{k} \cdot \vec{\epsilon}(\vec{k}, \pm 1) = 0$  plus a longitudinal one  $\vec{\epsilon}_L(\vec{k}) \sim \vec{k}$ . In gauging away the massless Goldstone boson  $\vartheta(x)$  we have transformed it into the longitudinal polarization of the massive vector field. In the literature this is usually expressed saying that the Goldstone mode is “eaten up” by the longitudinal component of the gauge field. It is important to realize that in spite of the fact that the Lagrangian (300) looks pretty different from the one we started with we have not lost any degrees of freedom. We started with the two polarizations of the photon plus the two degrees of freedom associated with the real and imaginary components of the complex scalar field. After symmetry breaking we end up with the three polarizations of the massive vector field and the degree of freedom of the real scalar field  $\sigma(x)$ .

We can also understand the Higgs mechanism in the light of our discussion of gauge symmetry in section 4.4. In the Higgs mechanism the invariance of the theory under infinitesimal gauge transformations is not explicitly broken, and this implies that Gauss' law is satisfied quantum mechanically,  $\vec{\nabla} \cdot \vec{E}_a|_{\text{phys}} = 0$ . The theory remains invariant under gauge transformations in the connected component of the identity  $\mathcal{G}_0$ , the ones generated by Gauss' law. This does not pose any restriction on the possible breaking of the invariance of the theory with respect to transformations that cannot be continuously deformed to the identity. Hence in the Higgs mechanism the invariance under gauge transformation that are not in the connected component of the identity,  $\mathcal{G}/\mathcal{G}_0$ , can be broken. Let us try to put it in more precise terms. As we learned in section 4.4, in the Hamiltonian formulation of the theory finite energy gauge field configurations tend to a pure gauge at spatial infinity

$$\vec{A}_\mu(\vec{x}) \longrightarrow -\frac{1}{ig}g(\vec{x})\vec{\nabla}g(\vec{x})^{-1}, \quad |\vec{x}| \rightarrow \infty \quad (302)$$

The set transformations  $g_0(\vec{x}) \in \mathcal{G}_0$  that tend to the identity at infinity are the ones generated by Gauss' law. However, one can also consider in general gauge transformations  $g(\vec{x})$  which, as  $|\vec{x}| \rightarrow \infty$ , approach any other element  $g \in G$ . The quotient  $\mathcal{G}_\infty \equiv \mathcal{G}/\mathcal{G}_0$  gives a copy of the gauge group at infinity. There is no reason, however, why this group should not be broken, and in general it is if the gauge symmetry is spontaneously broken. Notice that this is not a threat to the consistency of the theory. Properties like the decoupling of unphysical states are guaranteed by the fact that Gauss' law is satisfied quantum mechanically and are not affected by the breaking of  $\mathcal{G}_\infty$ .

In condensed matter physics the symmetry breaking described by the nonrelativistic version of the Abelian Higgs model can be used to characterize the onset of a superconducting phase in the BCS theory, where the complex scalar field  $\Phi$  is associated with the Cooper pairs. In this case the parameter  $\mu^2$  depends on the temperature. Above the critical temperature  $T_c$ ,  $\mu^2(T) > 0$  and there is only a symmetric vacuum  $\langle \Phi \rangle = 0$ . When, on the other hand,  $T < T_c$  then  $\mu^2(T) < 0$  and symmetry breaking takes place. The onset of a nonzero mass of the photon (301) below the critical temperature explains the Meissner effect: the magnetic fields cannot penetrate inside superconductors beyond a distance of the order  $\frac{1}{m_\gamma}$ .

The Abelian Higgs model discussed here can be regarded as a toy model of the Brout-Englert-Higgs mechanism responsible for giving mass to the  $W^\pm$  and  $Z^0$  gauge bosons in the standard model. Giving mass to these three bosons requires the introduction of a two-component complex scalar field transforming as a doublet under  $SU(2)$ . Three of its four degrees of freedom are incorporated as the longitudinal components of the three massive gauge fields, whereas the fourth one remains as a scalar propagating degree of freedom. Its elementary excitations are spin zero neutral particles known as Higgs bosons.

The Higgs boson couples to the massive gauge fields, as well as to quarks and leptons. Moreover, its coupling to the fermions is proportional to the fermion masses and therefore very weak for light fermions. This, together with the fact that Higgs production processes have large standard model backgrounds, complicates its experimental detection. After decades of searches in various experiments, a Higgs boson candidate was finally detected at the ATLAS and CMS collaborations at the Large Hadron Collider (LHC) in 2012 with a mass of approximately 125 GeV. At the time of writing, all evidences point to the fact that this new particle is indeed the so much coveted standard model Higgs.

## 7 Anomalies

So far we did not worry too much about how classical symmetries of a theory are carried over to the quantum theory. We have implicitly assumed that classical symmetries are preserved in the process of quantization, so they are also realized in the quantum theory.

This, however, does not have to be necessarily the case. Quantizing an interacting field theory is a very involved process that requires regularization and renormalization and sometimes, it does not matter how hard we try, there is no way for a classical symmetry to survive quantization. When this happens one says that the theory has an *anomaly* (for reviews see [28]). It is important to avoid here the misconception that anomalies appear due to a bad choice of the way a theory is regularized in the process of quantization. When we talk about anomalies we mean a classical symmetry that *cannot* be realized in the quantum theory, no matter how smart we are in choosing the regularization procedure.

In the following we analyze some examples of anomalies associated with global and local symmetries of the classical theory. In Section 8 we will encounter yet another example of an anomaly, this time associated with the breaking of classical scale invariance in the quantum theory.

### 7.1 Axial anomaly

Probably the best known examples of anomalies appear when we consider axial symmetries. If we consider a theory of two Weyl spinors  $u_\pm$

$$\mathcal{L} = i\bar{\psi}\not{\partial}\psi = iu_+^\dagger\sigma_+^\mu\partial_\mu u_+ + iu_-^\dagger\sigma_-^\mu\partial_\mu u_- \quad \text{with} \quad \psi = \begin{pmatrix} u_+ \\ u_- \end{pmatrix} \quad (303)$$

the Lagrangian is invariant under two types of global  $U(1)$  transformations. In the first one both helicities transform with the same phase, this is a *vector* transformation:

$$U(1)_V : u_\pm \longrightarrow e^{i\alpha}u_\pm, \quad (304)$$

whereas in the second one, the axial  $U(1)$ , the signs of the phases are different for the two chiralities

$$U(1)_A : u_\pm \longrightarrow e^{\pm i\alpha}u_\pm. \quad (305)$$

Using Noether's theorem, there are two conserved currents, a vector current

$$J_V^\mu = \bar{\psi}\gamma^\mu\psi = u_+^\dagger\sigma_+^\mu u_+ + u_-^\dagger\sigma_-^\mu u_- \quad \implies \quad \partial_\mu J_V^\mu = 0 \quad (306)$$

and an axial vector current

$$J_A^\mu = \bar{\psi} \gamma^\mu \gamma_5 \psi = u_+^\dagger \sigma_+^\mu u_+ - u_-^\dagger \sigma_-^\mu u_- \implies \partial_\mu J_A^\mu = 0. \quad (307)$$

The theory described by the Lagrangian (303) can be coupled to the electromagnetic field. The resulting classical theory is still invariant under the vector and axial U(1) symmetries (304) and (305). Surprisingly, upon quantization it turns out that the conservation of the axial current (307) is spoiled by quantum effects

$$\partial_\mu J_A^\mu \sim \hbar \vec{E} \cdot \vec{B}. \quad (308)$$

To understand more clearly how this result comes about we study first a simple model in two dimensions that captures the relevant physics involved in the four-dimensional case [29]. We work in Minkowski space in two dimensions with coordinates  $(x^0, x^1) \equiv (t, x)$  and where the spatial direction is compactified to a circle  $S^1$ . In this setup we consider a fermion coupled to the electromagnetic field. Notice that since we are living in two dimensions the field strength  $F_{\mu\nu}$  only has one independent component that corresponds to the electric field along the spatial direction,  $F^{01} \equiv \mathcal{E}$  (in two dimensions there are no magnetic fields!).

To write the Lagrangian for the spinor field we need to find a representation of the algebra of  $\gamma$ -matrices

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} \quad \text{with} \quad \eta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (309)$$

In two dimensions the dimension of the representation of the  $\gamma$ -matrices is  $2^{\lfloor \frac{2}{2} \rfloor} = 2$ . Here take

$$\gamma^0 \equiv \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^1 \equiv i\sigma^2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (310)$$

This is a chiral representation since the matrix  $\gamma_5$  is diagonal<sup>15</sup>

$$\gamma_5 \equiv -\gamma^0 \gamma^1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (311)$$

Writing the two-component spinor  $\psi$  as

$$\psi = \begin{pmatrix} u_+ \\ u_- \end{pmatrix} \quad (312)$$

and defining as usual the projectors  $P_\pm = \frac{1}{2}(\mathbf{1} \pm \gamma_5)$  we find that the components  $u_\pm$  of  $\psi$  are respectively a right- and left-handed Weyl spinor in two dimensions.

Once we have a representation of the  $\gamma$ -matrices we can write the Dirac equation. Expressing it in terms of the components  $u_\pm$  of the Dirac spinor we find

$$(\partial_0 - \partial_1)u_+ = 0, \quad (\partial_0 + \partial_1)u_- = 0. \quad (313)$$

The general solution to these equations can be immediately written as

$$u_+ = u_+(x^0 + x^1), \quad u_- = u_-(x^0 - x^1). \quad (314)$$

Hence  $u_\pm$  are two wave packets moving along the spatial dimension respectively to the left ( $u_+$ ) and to the right ( $u_-$ ). Notice that according to our convention the left-moving  $u_+$  is a right-handed spinor (positive helicity) whereas the right-moving  $u_-$  is a left-handed spinor (negative helicity).

<sup>15</sup>In any even number of dimensions  $\gamma_5$  is defined to satisfy the conditions  $\gamma_5^2 = \mathbf{1}$  and  $\{\gamma_5, \gamma^\mu\} = 0$ .

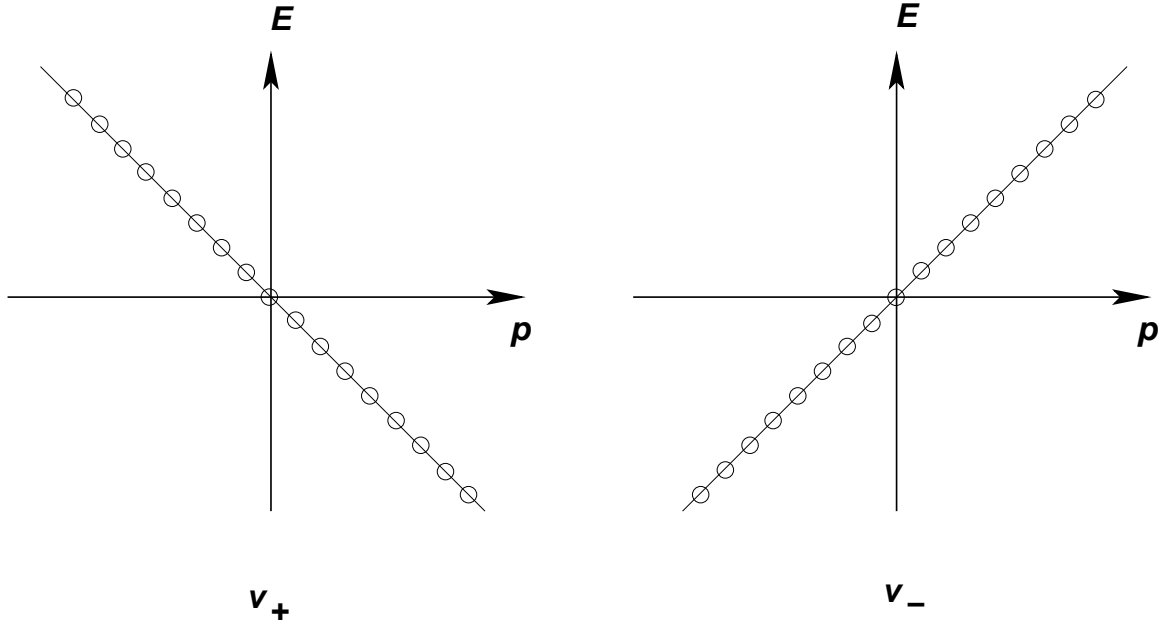


Fig. 11: Spectrum of the massless two-dimensional Dirac field.

If we want to interpret (313) as the wave equation for two-dimensional Weyl spinors we have the following wave functions for free particles with well defined momentum  $p^\mu = (E, p)$ .

$$u_{\pm}^{(E)}(x^0 \pm x^1) = \frac{1}{\sqrt{L}} e^{-iE(x^0 \pm x^1)} \quad \text{with} \quad p = \mp E. \quad (315)$$

As it is always the case with the Dirac equation we have both positive and negative energy solutions. For  $u_+$ , since  $E = -p$ , we see that the solutions with positive energy are those with negative momentum  $p < 0$ , whereas the negative energy solutions are plane waves with  $p > 0$ . For the left-handed spinor  $u_-$  the situation is reversed. Besides, since the spatial direction is compact with length  $L$  the momentum  $p$  is quantized according to

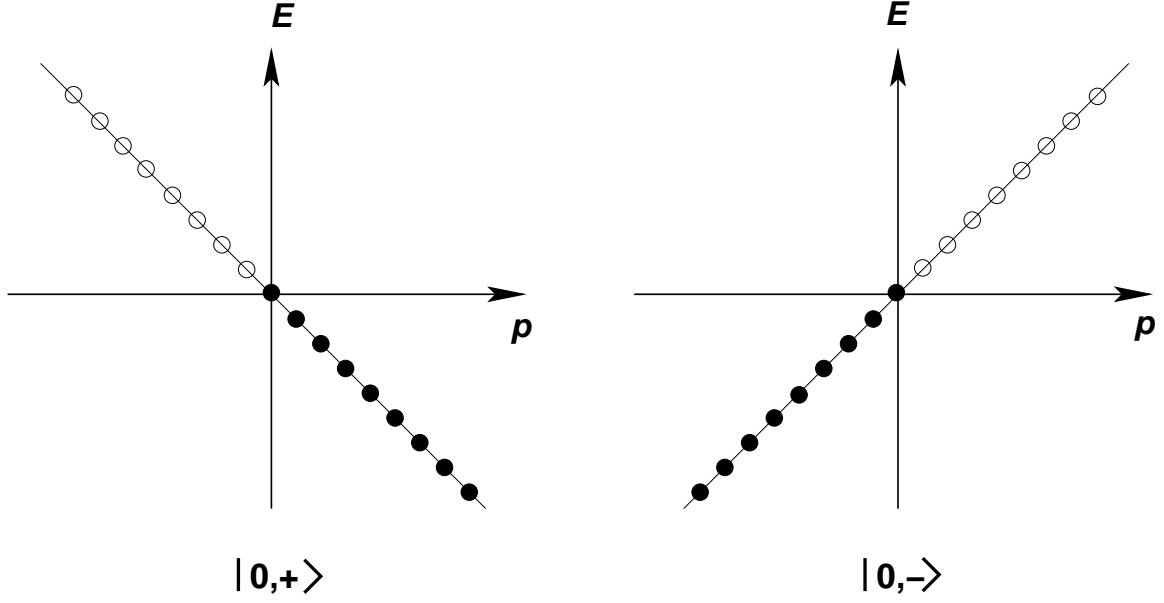
$$p = \frac{2\pi n}{L}, \quad n \in \mathbb{Z}. \quad (316)$$

The spectrum of the theory is represented in Fig. 11.

Once we have the spectrum of the theory the next step is to obtain the vacuum. As with the Dirac equation in four dimensions we fill all the states with  $E \leq 0$  (Fig. 12). Exciting of a particle in the Dirac sea produces a positive energy fermion plus a hole that is interpreted as an antiparticle. This gives us the clue on how to quantize the theory. In the expansion of the operator  $u_{\pm}$  in terms of the modes (315) we associate positive energy states with annihilation operators whereas the states with negative energy are associated with creation operators for the corresponding antiparticle

$$u_{\pm}(x) = \sum_{E>0} \left[ a_{\pm}(E) v_{\pm}^{(E)}(x) + b_{\pm}^{\dagger}(E) v_{\pm}^{(E)}(x)^* \right]. \quad (317)$$

The operator  $a_{\pm}(E)$  acting on the vacuum  $|0, \pm\rangle$  annihilates a particle with positive energy  $E$  and momentum  $\mp E$ . In the same way  $b_{\pm}^{\dagger}(E)$  creates out of the vacuum an antiparticle with positive energy  $E$  and spatial momentum  $\mp E$ . In the Dirac sea picture the operator  $b_{\pm}(E)^{\dagger}$  is originally an annihilation operator for a state of the sea with negative energy  $-E$ . As in the four-dimensional case the problem of the negative energy states is solved by interpreting annihilation operators for negative energy states as



**Fig. 12:** Vacuum of the theory.

creation operators for the corresponding antiparticle with positive energy (and vice versa). The operators appearing in the expansion of  $u_{\pm}$  in Eq. (317) satisfy the usual algebra

$$\{a_{\lambda}(E), a_{\lambda'}^{\dagger}(E')\} = \{b_{\lambda}(E), b_{\lambda'}^{\dagger}(E')\} = \delta_{E,E'}\delta_{\lambda\lambda'}, \quad (318)$$

where we have introduced the label  $\lambda, \lambda' = \pm$ . Also,  $a_{\lambda}(E), a_{\lambda'}^{\dagger}(E)$  anticommute with  $b_{\lambda'}(E'), b_{\lambda'}^{\dagger}(E')$ .

The Lagrangian of the theory

$$\mathcal{L} = iu_{+}^{\dagger}(\partial_0 + \partial_1)u_{+} + iu_{-}^{\dagger}(\partial_0 - \partial_1)u_{-} \quad (319)$$

is invariant under both  $U(1)_V$ , Eq. (304), and  $U(1)_A$ , Eq. (305). The associated Noether currents are in this case

$$J_V^{\mu} = \begin{pmatrix} u_{+}^{\dagger}u_{+} + u_{-}^{\dagger}u_{-} \\ -u_{+}^{\dagger}u_{+} + u_{-}^{\dagger}u_{-} \end{pmatrix}, \quad J_A^{\mu} = \begin{pmatrix} u_{+}^{\dagger}u_{+} - u_{-}^{\dagger}u_{-} \\ -u_{+}^{\dagger}u_{+} - u_{-}^{\dagger}u_{-} \end{pmatrix}. \quad (320)$$

The associated conserved charges are given, for the vector current by

$$Q_V = \int_0^L dx^1 \left( u_{+}^{\dagger}u_{+} + u_{-}^{\dagger}u_{-} \right) \quad (321)$$

and for the axial current

$$Q_A = \int_0^L dx^1 \left( u_{+}^{\dagger}u_{+} - u_{-}^{\dagger}u_{-} \right). \quad (322)$$

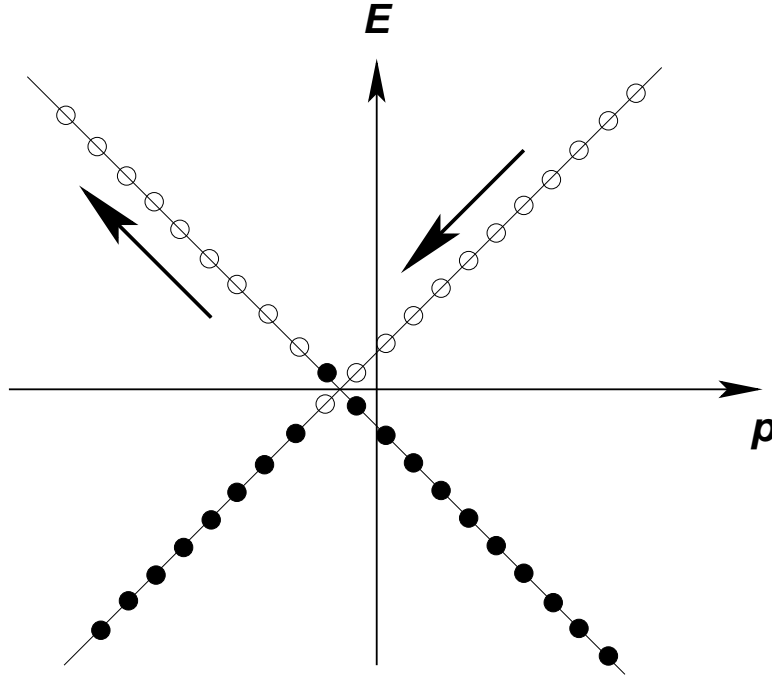
Using the orthonormality relations for the modes  $v_{\pm}^{(E)}(x)$

$$\int_0^L dx^1 v_{\pm}^{(E)}(x) v_{\pm}^{(E')}(x) = \delta_{E,E'} \quad (323)$$

we find for the conserved charges:

$$Q_V = \sum_{E>0} \left[ a_{+}^{\dagger}(E)a_{+}(E) - b_{+}^{\dagger}(E)b_{+}(E) + a_{-}^{\dagger}(E)a_{-}(E) - b_{-}^{\dagger}(E)b_{-}(E) \right],$$





**Fig. 13:** Effect of the electric field.

$$Q_A = \sum_{E>0} \left[ a_+^\dagger(E)a_+(E) - b_+^\dagger(E)b_+(E) - a_-^\dagger(E)a_-(E) + b_-^\dagger(E)b_-(E) \right]. \quad (324)$$

We see that  $Q_V$  counts the net number (particles minus antiparticles) of positive helicity states plus the net number of states with negative helicity. The axial charge, on the other hand, counts the net number of positive helicity states minus the number of negative helicity ones. In the case of the vector current we have subtracted a formally divergent vacuum contribution to the charge (the “charge of the Dirac sea”).

In the free theory there is of course no problem with the conservation of either  $Q_V$  or  $Q_A$ , since the occupation numbers do not change. What we want to study is the effect of coupling the theory to electric field  $\mathcal{E}$ . We work in the gauge  $A_0 = 0$ . Instead of solving the problem exactly we are going to simulate the electric field by adiabatically varying in a long time  $\tau_0$  the vector potential  $A_1$  from zero value to  $-\mathcal{E}\tau_0$ . From our discussion in section 4.3 we know that the effect of the electromagnetic coupling in the theory is a shift in the momentum according to

$$p \longrightarrow p - eA_1, \quad (325)$$

where  $e$  is the charge of the fermions. Since we assumed that the vector potential varies adiabatically, we can assume it to be approximately constant at each time.

Then, we have to understand what is the effect of (325) on the vacuum depicted in Fig. (12). What we find is that the two branches move as shown in Fig. (13) resulting in some of the negative energy states of the  $v_+$  branch acquiring positive energy while the same number of the empty positive energy states of the other branch  $v_-$  will become empty negative energy states. Physically this means that the external electric field  $\mathcal{E}$  creates a number of particle-antiparticle pairs out of the vacuum. Denoting by  $N \sim e\mathcal{E}$  the number of such pairs created by the electric field per unit time, the final values of the charges  $Q_V$  and  $Q_A$  are

$$\begin{aligned} Q_A(\tau_0) &= (N - 0) + (0 - N) = 0, \\ Q_V(\tau_0) &= (N - 0) - (0 - N) = 2N. \end{aligned} \quad (326)$$

Therefore we conclude that the coupling to the electric field produces a violation in the conservation of the axial charge per unit time given by  $\Delta Q_A \sim e\mathcal{E}$ . This implies that

$$\partial_\mu J_A^\mu \sim e\hbar\mathcal{E}, \quad (327)$$

where we have restored  $\hbar$  to make clear that the violation in the conservation of the axial current is a quantum effect. At the same time  $\Delta Q_V = 0$  guarantees that the vector current remains conserved also quantum mechanically,  $\partial_\mu J_V^\mu = 0$ .

We have just studied a two-dimensional example of the Adler-Bell-Jackiw axial anomaly [30]. The heuristic analysis presented here can be made more precise by computing the quantity

$$C^{\mu\nu} = \langle 0|T [J_A^\mu(x)J_V^\nu(0)] |0\rangle = \begin{array}{c} \text{---} \\ J_A^\mu \end{array} \circlearrowleft \begin{array}{c} \text{---} \\ \gamma \end{array} \quad (328)$$

The anomaly is given then by  $\partial_\mu C^{\mu\nu}$ . A careful calculation yields the numerical prefactor missing in Eq. (327) leading to the result

$$\partial_\mu J_A^\mu = \frac{e\hbar}{2\pi} \varepsilon^{\nu\sigma} F_{\nu\sigma}, \quad (329)$$

with  $\varepsilon^{01} = -\varepsilon^{10} = 1$ .

The existence of an anomaly in the axial symmetry that we have illustrated in two dimensions is present in all even dimensional of space-times. In particular in four dimensions the axial anomaly it is given by

$$\partial_\mu J_A^\mu = -\frac{e^2}{16\pi^2} \varepsilon^{\mu\nu\sigma\lambda} F_{\mu\nu} F_{\sigma\lambda}. \quad (330)$$

This result has very important consequences in the physics of strong interactions as we will see in what follows

## 7.2 Chiral symmetry in QCD

Our knowledge of the physics of strong interactions is based on the theory of Quantum Chromodynamics (QCD) [32]. This is a nonabelian gauge theory with gauge group  $SU(N_c)$  coupled to a number  $N_f$  of quarks. These are spin- $\frac{1}{2}$  particles  $Q^{if}$  labelled by two quantum numbers: color  $i = 1, \dots, N_c$  and flavor  $f = 1, \dots, N_f$ . The interaction between them is mediated by the  $N_c^2 - 1$  gauge bosons, the gluons  $A_\mu^a$ ,  $a = 1, \dots, N_c^2 - 1$ . In the real world  $N_c = 3$  and the number of flavors is six, corresponding to the number of different quarks: up ( $u$ ), down ( $d$ ), charm ( $c$ ), strange ( $s$ ), top ( $t$ ) and bottom ( $b$ ).

For the time being we are going to study a general theory of QCD with  $N_c$  colors and  $N_f$  flavors. Also, for reasons that will be clear later we are going to work in the limit of vanishing quark masses,  $m_f \rightarrow 0$ . In this cases the Lagrangian is given by

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} F_{\mu\nu}^a F^{a\mu\nu} + \sum_{f=1}^{N_f} \left[ i\bar{Q}_L^f \not{D} Q_L^f + i\bar{Q}_R^f \not{D} Q_R^f \right], \quad (331)$$

where the subscripts  $L$  and  $R$  indicate respectively left and right-handed spinors,  $Q_{L,R}^f \equiv P_\pm Q^f$ , and the field strength  $F_{\mu\nu}^a$  and the covariant derivative  $D_\mu$  are respectively defined in Eqs. (165) and (168). Apart

from the gauge symmetry, this Lagrangian is also invariant under a global  $U(N_f)_L \times U(N_f)_R$  acting on the flavor indices and defined by

$$U(N_f)_L : \begin{cases} Q_L^f \rightarrow \sum_{f'} (U_L)_{ff'} Q_L^{f'} \\ Q_R^f \rightarrow Q_R^f \end{cases} \quad U(N_f)_R : \begin{cases} Q_L^f \rightarrow Q_L^f \\ Q_R^f \rightarrow \sum_{f'} (U_R)_{ff'} Q_R^{f'} \end{cases} \quad (332)$$

with  $U_L, U_R \in U(N_f)$ . Actually, since  $U(N) = U(1) \times SU(N)$  this global symmetry group can be written as  $SU(N_f)_L \times SU(N_f)_R \times U(1)_L \times U(1)_R$ . The abelian subgroup  $U(1)_L \times U(1)_R$  can be now decomposed into their vector  $U(1)_B$  and axial  $U(1)_A$  subgroups defined by the transformations

$$U(1)_B : \begin{cases} Q_L^f \rightarrow e^{i\alpha} Q_L^f \\ Q_R^f \rightarrow e^{i\alpha} Q_R^f \end{cases} \quad U(1)_A : \begin{cases} Q_L^f \rightarrow e^{i\alpha} Q_L^f \\ Q_R^f \rightarrow e^{-i\alpha} Q_R^f \end{cases} \quad (333)$$

According to Noether's theorem, associated with these two abelian symmetries we have two conserved currents:

$$J_V^\mu = \sum_{f=1}^{N_f} \bar{Q}^f \gamma^\mu Q^f, \quad J_A^\mu = \sum_{f=1}^{N_f} \bar{Q}^f \gamma^\mu \gamma_5 Q^f. \quad (334)$$

The conserved charge associated with vector charge  $J_V^\mu$  is actually the baryon number defined as the number of quarks minus number of antiquarks.

The nonabelian part of the global symmetry group  $SU(N_f)_L \times SU(N_f)_R$  can also be decomposed into its vector and axial subgroups,  $SU(N_f)_V \times SU(N_f)_A$ , defined by the following transformations of the quarks fields

$$SU(N_f)_V : \begin{cases} Q_L^f \rightarrow \sum_{f'} (U_L)_{ff'} Q_L^{f'} \\ Q_R^f \rightarrow \sum_{f'} (U_L)_{ff'} Q_R^{f'} \end{cases} \quad SU(N_f)_A : \begin{cases} Q_L^f \rightarrow \sum_{f'} (U_L)_{ff'} Q_L^{f'} \\ Q_R^f \rightarrow \sum_{f'} (U_R^{-1})_{ff'} Q_R^{f'} \end{cases} \quad (335)$$

Again, the application of Noether's theorem shows the existence of the following nonabelian conserved charges

$$J_V^{I\mu} \equiv \sum_{f,f'=1}^{N_f} \bar{Q}^f \gamma^\mu (T^I)_{ff'} Q^{f'}, \quad J_A^{I\mu} \equiv \sum_{f,f'=1}^{N_f} \bar{Q}^f \gamma^\mu \gamma_5 (T^I)_{ff'} Q^{f'}. \quad (336)$$

To summarize, we have shown that the initial chiral symmetry of the QCD Lagrangian (331) can be decomposed into its chiral and vector subgroups according to

$$U(N_f)_L \times U(N_f)_R = SU(N_f)_V \times SU(N_f)_A \times U(1)_B \times U(1)_A. \quad (337)$$

The question to address now is which part of the classical global symmetry is preserved by the quantum theory.

As argued in section 7.1, the conservation of the axial currents  $J_A^\mu$  and  $J_A^{a\mu}$  can in principle be spoiled due to the presence of an anomaly. In the case of the abelian axial current  $J_A^\mu$  the relevant quantity is the correlation function

$$C^{\mu\nu\sigma} \equiv \langle 0 | T \left[ J_A^\mu(x) j_{\text{gauge}}^{a\nu}(x') j_{\text{gauge}}^{b\sigma}(0) \right] | 0 \rangle = \sum_{f=1}^{N_f} \left[ \begin{array}{c} \text{Diagram} \end{array} \right]_{\text{symmetric}} \quad (338)$$

Here  $j_{\text{gauge}}^{a\mu}$  is the nonabelian conserved current coupling to the gluon field

$$j_{\text{gauge}}^{a\mu} \equiv \sum_{f=1}^{N_f} \bar{Q}^f \gamma^\mu \tau^a Q^f, \quad (339)$$

where, to avoid confusion with the generators of the global symmetry we have denoted by  $\tau^a$  the generators of the gauge group  $\text{SU}(N_c)$ . The anomaly can be read now from  $\partial_\mu C^{\mu\nu\sigma}$ . If we impose Bose symmetry with respect to the interchange of the two outgoing gluons and gauge invariance of the whole expression,  $\partial_\nu C^{\mu\nu\sigma} = 0 = \partial_\sigma C^{\mu\nu\sigma}$ , we find that the axial abelian global current has an anomaly given by<sup>16</sup>

$$\partial_\mu J_A^\mu = -\frac{g^2 N_f}{32\pi^2} \varepsilon^{\mu\nu\sigma\lambda} F_{\mu\nu}^a F_{\sigma\lambda}^a. \quad (340)$$

In the case of the nonabelian axial global symmetry  $\text{SU}(N_f)_A$  the calculation of the anomaly is made as above. The result, however, is quite different since in this case we conclude that the nonabelian axial current  $J_A^{a\mu}$  is not anomalous. This can be easily seen by noticing that associated with the axial current vertex we have a generator  $T^I$  of  $\text{SU}(N_f)$ , whereas for the two gluon vertices we have the generators  $\tau^a$  of the gauge group  $\text{SU}(N_c)$ . Therefore, the triangle diagram is proportional to the group-theoretic factor

$$\left[ \begin{array}{c} \text{Diagram: Triangle with } J_A^{I\mu} \text{ on the left, } Q^f \text{ on the top and bottom, and } g \text{ on the right.} \\ \text{symmetric} \end{array} \right] \sim \text{tr } T^I \text{tr } \{ \tau^a, \tau^b \} = 0 \quad (341)$$

which vanishes because the generators of  $\text{SU}(N_f)$  are traceless.

From here we would conclude that the nonabelian axial symmetry  $\text{SU}(N_f)_A$  is nonanomalous. However this is not the whole story since quarks are charged particles that also couple to photons. Hence there is a second potential source of an anomaly coming from the one-loop triangle diagram coupling  $J_A^{I\mu}$  to two photons

$$\langle 0 | T \left[ J_A^{I\mu}(x) j_{\text{em}}^\nu(x') j_{\text{em}}^\sigma(0) \right] | 0 \rangle = \sum_{f=1}^{N_f} \left[ \begin{array}{c} \text{Diagram: Triangle with } J_A^{I\mu} \text{ on the left, } Q^f \text{ on the top and bottom, and } \gamma \text{ on the right.} \\ \text{symmetric} \end{array} \right] \quad (342)$$

where  $j_{\text{em}}^\mu$  is the electromagnetic current

$$j_{\text{em}}^\mu = \sum_{f=1}^{N_f} q_f \bar{Q}^f \gamma^\mu Q^f, \quad (343)$$

with  $q_f$  the electric charge of the  $f$ -th quark flavor. A calculation of the diagram in (342) shows the existence of an Adler-Bell-Jackiw anomaly given by

$$\partial_\mu J_A^{I\mu} = -\frac{N_c}{16\pi^2} \left[ \sum_{f=1}^{N_f} (T^I)_{ff} q_f^2 \right] \varepsilon^{\mu\nu\sigma\lambda} F_{\mu\nu} F_{\sigma\lambda}, \quad (344)$$

<sup>16</sup>The normalization of the generators  $T^I$  of the global  $\text{SU}(N_f)$  is given by  $\text{tr}(T^I T^J) = \frac{1}{2} \delta^{IJ}$ .

where  $F_{\mu\nu}$  is the field strength of the electromagnetic field coupling to the quarks. The only chance for the anomaly to cancel is that the factor between brackets in this equation be identically zero.

Before proceeding let us summarize the results found so far. Because of the presence of anomalies the axial part of the global chiral symmetry,  $SU(N_f)_A$  and  $U(1)_A$  are not realized quantum mechanically in general. We found that  $U(1)_A$  is always affected by an anomaly. However, because the right-hand side of the anomaly equation (340) is a total derivative, the anomalous character of  $J_A^\mu$  does not explain the absence of  $U(1)_A$  multiplets in the hadron spectrum, since a new current can be constructed which is conserved. In addition, the nonexistence of candidates for a Goldstone boson associated with the right quantum numbers indicates that  $U(1)_A$  is not spontaneously broken either, so it has to be explicitly broken somehow. This is the so-called  $U(1)$ -problem which was solved by 't Hooft [33], who showed how the contribution of quantum transitions between vacua with topologically nontrivial gauge field configurations (instantons) results in an explicit breaking of this symmetry.

Due to the dynamics of the  $SU(N_c)$  gauge theory the axial nonabelian symmetry is spontaneously broken due to the presence at low energies of a vacuum expectation value for the fermion bilinear  $\bar{Q}^f Q^f$

$$\langle 0 | \bar{Q}^f Q^f | 0 \rangle \neq 0 \quad (\text{No summation in } f!). \quad (345)$$

This nonvanishing vacuum expectation value for the quark bilinear actually breaks chiral invariance spontaneously to the vector subgroup  $SU(N_f)_V$ , so the only subgroup of the original global symmetry that is realized by the full theory at low energy is

$$U(N_f)_L \times U(N_f)_R \longrightarrow SU(N_f)_V \times U(1)_B. \quad (346)$$

Associated with this breaking a Goldstone boson should appear with the quantum numbers of the broken nonabelian current. For example, in the case of QCD the Goldstone bosons associated with the spontaneously symmetry breaking induced by the vacuum expectation values  $\langle \bar{u}u \rangle$ ,  $\langle \bar{d}d \rangle$  and  $\langle (\bar{u}d - \bar{d}u) \rangle$  have been identified as the pions  $\pi^0$ ,  $\pi^\pm$ . These bosons are not exactly massless because of the nonvanishing mass of the  $u$  and  $d$  quarks. Since the global chiral symmetry is already slightly broken by mass terms in the Lagrangian, the associated Goldstone bosons also have masses although they are very light compared to the masses of other hadrons.

In order to have a better physical understanding of the role of anomalies in the physics of strong interactions we particularize now our analysis of the case of real QCD. Since the  $u$  and  $d$  quarks are much lighter than the other four flavors, QCD at low energies can be well described by including only these two flavors and ignoring heavier quarks. In this approximation, from our previous discussion we know that the low energy global symmetry of the theory is  $SU(2)_V \times U(1)_B$ , where now the vector group  $SU(2)_V$  is the well-known isospin symmetry. The axial  $U(1)_A$  current is anomalous due to Eq. (340) with  $N_f = 2$ . In the case of the nonabelian axial symmetry  $SU(2)_A$ , taking into account that  $q_u = \frac{2}{3}e$  and  $q_d = -\frac{1}{3}e$  and that the three generators of  $SU(2)$  can be written in terms of the Pauli matrices as  $T^K = \frac{1}{2}\sigma^K$  we find

$$\sum_{f=u,d} (T^1)_{ff} q_f^2 = \sum_{f=u,d} (T^2)_{ff} q_f^2 = 0, \quad \sum_{f=u,d} (T^3)_{ff} q_f^2 = \frac{e^2}{6}. \quad (347)$$

Therefore  $J_A^{3\mu}$  is anomalous.

Physically, the anomaly in the axial current  $J_A^{3\mu}$  has an important consequence. In the quark model, the wave function of the neutral pion  $\pi^0$  is given in terms of those for the  $u$  and  $d$  quark by

$$|\pi^0\rangle = \frac{1}{\sqrt{2}} (|\bar{u}\rangle|u\rangle - |\bar{d}\rangle|d\rangle). \quad (348)$$

The isospin quantum numbers of  $|\pi^0\rangle$  are those of the generator  $T^3$ . Actually the analogy goes further since  $\partial_\mu J_A^{3\mu}$  is the operator creating a pion  $\pi^0$  out of the vacuum

$$|\pi^0\rangle \sim \partial_\mu J_A^{3\mu} |0\rangle. \quad (349)$$

This leads to the physical interpretation of the triangle diagram (342) with  $J_A^{3\mu}$  as the one loop contribution to the decay of a neutral pion into two photons

$$\pi^0 \longrightarrow 2\gamma. \quad (350)$$

This is an interesting piece of physics. In 1967 Sutherland and Veltman [34] presented a calculation, using current algebra techniques, according to which the decay of the pion into two photons should be suppressed. This however contradicted the experimental evidence that showed the existence of such a decay. The way out to this paradox, as pointed out in [30], is the axial anomaly. What happens is that the current algebra analysis overlooks the ambiguities associated with the regularization of divergences in quantum field theory. A QED evaluation of the triangle diagram leads to a divergent integral that has to be regularized somehow. It is in this process that the Adler-Bell-Jackiw axial anomaly appears resulting in a nonvanishing value for the  $\pi^0 \rightarrow 2\gamma$  amplitude<sup>17</sup>.

The existence of anomalies associated with global currents does not necessarily mean difficulties for the theory. On the contrary, as we saw in the case of the axial anomaly it is its existence what allows for a solution of the Sutherland-Veltman paradox and an explanation of the electromagnetic decay of the pion. The situation, however, is very different if we deal with local symmetries. A quantum mechanical violation of gauge symmetry leads to all kinds of problems, from lack of renormalizability to nondecoupling of negative norm states. This is because the presence of an anomaly in the theory implies that the Gauss' law constraint  $\vec{\nabla} \cdot \vec{E}_a = \rho_a$  cannot be consistently implemented in the quantum theory. As a consequence states that classically are eliminated by the gauge symmetry become propagating fields in the quantum theory, thus spoiling the consistency of the theory.

Anomalies in a gauge symmetry can be expected only in chiral theories where left and right-handed fermions transform in different representations of the gauge group. Physically, the most interesting example of such theories is the electroweak sector of the standard model where, for example, left handed fermions transform as doublets under SU(2) whereas right-handed fermions are singlets. On the other hand, QCD is free of gauge anomalies since both left- and right-handed quarks transform in the fundamental representation of SU(3).

We consider the Lagrangian

$$\mathcal{L} = -\frac{1}{4} F^{a\mu\nu} F_{\mu\nu}^a + i \sum_{i=1}^{N_+} \bar{\psi}_+^i \mathcal{D}^{(+)} \psi_+^i + i \sum_{j=1}^{N_-} \bar{\psi}_-^j \mathcal{D}^{(-)} \psi_-^j, \quad (351)$$

where the chiral fermions  $\psi_\pm^i$  transform according to the representations  $\tau_{i,\pm}^a$  of the gauge group  $G$  ( $a = 1, \dots, \dim G$ ). The covariant derivatives  $D_\mu^{(\pm)}$  are then defined by

$$D_\mu^{(\pm)} \psi_\pm^i = \partial_\mu \psi_\pm^i + ig A_\mu^K \tau_{i,\pm}^K \psi_\pm^i. \quad (352)$$

As for global symmetries, anomalies in the gauge symmetry appear in the triangle diagram with one axial and two vector gauge current vertices

$$\langle 0 | T \left[ j_A^{a\mu}(x) j_V^{b\nu}(x') j_V^{c\sigma}(0) \right] | 0 \rangle = \left[ \text{triangle diagram} \right]_{\text{symmetric}} \quad (353)$$

<sup>17</sup>An early computation of the triangle diagram for the electromagnetic decay of the pion was made by Steinberger in [31].

where gauge vector and axial currents  $j_V^{a\mu}$ ,  $j_A^{a\mu}$  are given by

$$\begin{aligned} j_V^{a\mu} &= \sum_{i=1}^{N_+} \bar{\psi}_+^i \tau_+^a \gamma^\mu \psi_+^i + \sum_{j=1}^{N_-} \bar{\psi}_-^j \tau_-^a \gamma^\mu \psi_-^j, \\ j_A^{a\mu} &= \sum_{i=1}^{N_+} \bar{\psi}_+^i \tau_+^a \gamma^\mu \psi_+^i - \sum_{j=1}^{N_-} \bar{\psi}_-^j \tau_-^a \gamma^\mu \psi_-^j. \end{aligned} \quad (354)$$

Luckily, we do not have to compute the whole diagram in order to find an anomaly cancellation condition, it is enough if we calculate the overall group theoretical factor. In the case of the diagram in Eq. (353) for every fermion species running in the loop this factor is equal to

$$\text{tr} \left[ \tau_{i,\pm}^a \{ \tau_{i,\pm}^b, \tau_{i,\pm}^c \} \right], \quad (355)$$

where the sign  $\pm$  corresponds respectively to the generators of the representation of the gauge group for the left and right-handed fermions. Hence the anomaly cancellation condition reads

$$\sum_{i=1}^{N_+} \text{tr} \left[ \tau_{i,+}^a \{ \tau_{i,+}^b, \tau_{i,+}^c \} \right] - \sum_{j=1}^{N_-} \text{tr} \left[ \tau_{j,-}^a \{ \tau_{j,-}^b, \tau_{j,-}^c \} \right] = 0. \quad (356)$$

Knowing this we can proceed to check the anomaly cancellation in the standard model  $SU(3) \times SU(2) \times U(1)$ . Left handed fermions (both leptons and quarks) transform as doublets with respect to the  $SU(2)$  factor whereas the right-handed components are singlets. The charge with respect to the  $U(1)$  part, the hypercharge  $Y$ , is determined by the Gell-Mann-Nishijima formula

$$Q = T_3 + Y, \quad (357)$$

where  $Q$  is the electric charge of the corresponding particle and  $T_3$  is the eigenvalue with respect to the third generator of the  $SU(2)$  group in the corresponding representation:  $T_3 = \frac{1}{2}\sigma^3$  for the doublets and  $T_3 = 0$  for the singlets. For the first family of quarks ( $u$ ,  $d$ ) and leptons ( $e$ ,  $\nu_e$ ) we have the following field content

$$\begin{aligned} \text{quarks:} & \quad \begin{pmatrix} u^\alpha \\ d^\alpha \end{pmatrix}_{L, \frac{1}{6}} & \quad u_{R, \frac{2}{3}}^\alpha & \quad d_{R, \frac{2}{3}}^\alpha \\ \text{leptons:} & \quad \begin{pmatrix} \nu_e \\ e \end{pmatrix}_{L, -\frac{1}{2}} & \quad e_{R, -1} \end{aligned} \quad (358)$$

where  $\alpha = 1, 2, 3$  labels the color quantum number and the subscript indicates the value of the weak hypercharge  $Y$ . Denoting the representations of  $SU(3) \times SU(2) \times U(1)$  by  $(n_c, n_w)_Y$ , with  $n_c$  and  $n_w$  the representations of  $SU(3)$  and  $SU(2)$  respectively and  $Y$  the hypercharge, the matter content of the standard model consists of a three family replication of the representations:

$$\begin{aligned} \text{left-handed fermions:} & \quad (3, 2)_{\frac{1}{6}}^L & \quad (1, 2)_{-\frac{1}{2}}^L \\ \text{right-handed fermions:} & \quad (3, 1)_{\frac{2}{3}}^R & \quad (3, 1)_{-\frac{1}{3}}^R & \quad (1, 1)_{-1}^R. \end{aligned} \quad (359)$$

In computing the triangle diagram we have 10 possibilities depending on which factor of the gauge group

$SU(3) \times SU(2) \times U(1)$  couples to each vertex:

$$\begin{array}{lll}
 SU(3)^3 & SU(2)^3 & U(1)^3 \\
 SU(3)^2 SU(2) & SU(2)^2 U(1) & \\
 SU(3)^2 U(1) & SU(2) U(1)^2 & \\
 SU(3) SU(2)^2 & & \\
 SU(3) SU(2) U(1) & & \\
 SU(3) U(1)^2 & & 
 \end{array}$$

It is easy to check that some of them do not give rise to anomalies. For example the anomaly for the  $SU(3)^3$  case cancels because left and right-handed quarks transform in the same representation. In the case of  $SU(2)^3$  the cancellation happens term by term because of the Pauli matrices identity  $\sigma^a \sigma^b = \delta^{ab} + i\varepsilon^{abc} \sigma^c$  that leads to

$$\text{tr} \left[ \sigma^a \{ \sigma^b, \sigma^c \} \right] = 2 (\text{tr} \sigma^a) \delta^{bc} = 0. \quad (360)$$

However the hardest anomaly cancellation condition to satisfy is the one with three  $U(1)$ 's. In this case the absence of anomalies within a single family is guaranteed by the nontrivial identity

$$\begin{aligned}
 \sum_{\text{left}} Y_+^3 - \sum_{\text{right}} Y_-^3 &= 3 \times 2 \times \left(\frac{1}{6}\right)^3 + 2 \times \left(-\frac{1}{2}\right)^3 - 3 \times \left(\frac{2}{3}\right)^3 - 3 \times \left(-\frac{1}{3}\right)^3 - (-1)^3 \\
 &= \left(-\frac{3}{4}\right) + \left(\frac{3}{4}\right) = 0.
 \end{aligned} \quad (361)$$

It is remarkable that the anomaly exactly cancels between leptons and quarks. Notice that this result holds even if a right-handed sterile neutrino is added since such a particle is a singlet under the whole standard model gauge group and therefore does not contribute to the triangle diagram. Therefore we see how the matter content of the standard model conspires to yield a consistent quantum field theory.

In all our discussion of anomalies we only considered the computation of one-loop diagrams. It may happen that higher loop orders impose additional conditions. Fortunately this is not so: the Adler-Bardeen theorem [35] guarantees that the axial anomaly only receives contributions from one loop diagrams. Therefore, once anomalies are canceled (if possible) at one loop we know that there will be no new conditions coming from higher-loop diagrams in perturbation theory.

The Adler-Bardeen theorem, however, only applies in perturbation theory. It is nonetheless possible that nonperturbative effects can result in the quantum violation of a gauge symmetry. This is precisely the case pointed out by Witten [36] with respect to the  $SU(2)$  gauge symmetry of the standard model. In this case the problem lies in the nontrivial topology of the gauge group  $SU(2)$ . The invariance of the theory with respect to gauge transformations which are not in the connected component of the identity makes all correlation functions equal to zero. Only when the number of left-handed  $SU(2)$  fermion doublets is even gauge invariance allows for a nontrivial theory. It is again remarkable that the family structure of the standard model makes this anomaly to cancel

$$3 \times \begin{pmatrix} u \\ d \end{pmatrix}_L + 1 \times \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L = 4 \text{ SU(2)-doublets}, \quad (362)$$

where the factor of 3 comes from the number of colors.



## 8 Renormalization

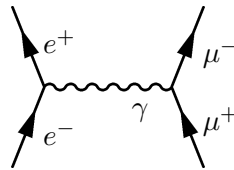
### 8.1 Removing infinities

From its very early stages, quantum field theory was faced with infinities. They emerged in the calculation of most physical quantities, such as the correction to the charge of the electron due to the interactions with the radiation field. The way these divergences were handled in the 1940s, starting with Kramers, was physically very much in the spirit of the Quantum Theory emphasis in observable quantities: since the observed magnitude of physical quantities (such as the charge of the electron) is finite, this number should arise from the addition of a “bare” (unobservable) value and the quantum corrections. The fact that both of these quantities were divergent was not a problem physically, since only its finite sum was an observable quantity. To make things mathematically sound, the handling of infinities requires the introduction of some regularization procedure which cuts the divergent integrals off at some momentum scale  $\Lambda$ . Morally speaking, the physical value of an observable  $\mathcal{O}_{\text{physical}}$  is given by

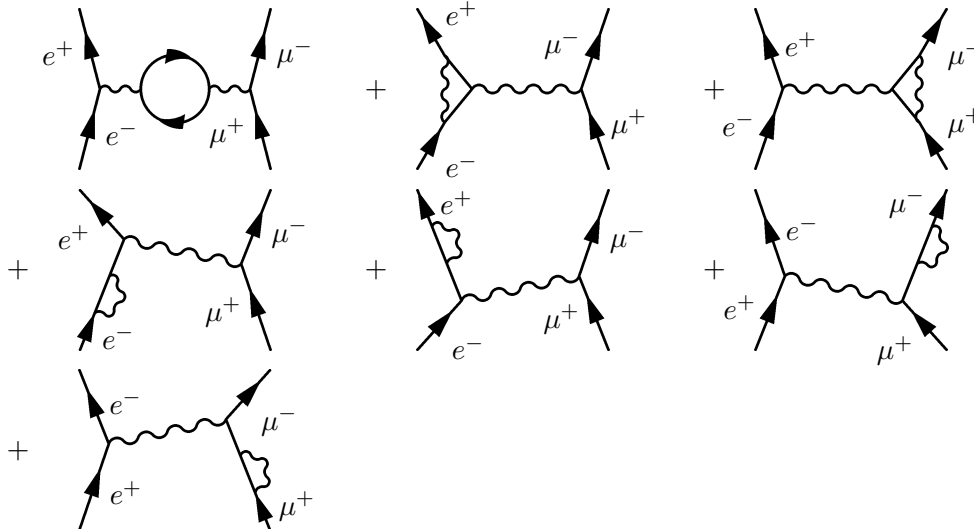
$$\mathcal{O}_{\text{physical}} = \lim_{\Lambda \rightarrow \infty} [\mathcal{O}(\Lambda)_{\text{bare}} + \Delta\mathcal{O}(\Lambda)_{\hbar}], \quad (363)$$

where  $\Delta\mathcal{O}(\Lambda)_{\hbar}$  represents the regularized quantum corrections.

To make this qualitative discussion more precise we compute the corrections to the electric charge in Quantum Electrodynamics. We consider the process of annihilation of an electron-positron pair to create a muon-antimuon pair  $e^-e^+ \rightarrow \mu^+\mu^-$ . To lowest order in the electric charge  $e$  the only diagram contributing is



However, the corrections at order  $e^4$  to this result requires the calculation of seven more diagrams



In order to compute the renormalization of the charge we consider the first diagram which takes into account the first correction to the propagator of the virtual photon interchanged between the pairs

due to vacuum polarization. We begin by evaluating

$$\begin{array}{c} \text{---} \circlearrowleft \text{---} \end{array} = \frac{-i\eta^{\mu\alpha}}{q^2 + i\epsilon} \left[ \begin{array}{c} \alpha \circlearrowleft \beta \end{array} \right] \frac{-i\eta^{\beta\nu}}{q^2 + i\epsilon}, \quad (364)$$

where the diagram between brackets is given by

$$\begin{array}{c} \alpha \circlearrowleft \beta \end{array} \equiv \Pi^{\alpha\beta}(q) = i^2(-ie)^2(-1) \int \frac{d^4k}{(2\pi)^4} \frac{\text{Tr}(\not{k} + m_e)\gamma^\alpha(\not{k} + \not{q} + m_e)\gamma^\beta}{[k^2 - m_e^2 + i\epsilon][(k+q)^2 - m_e^2 + i\epsilon]}. \quad (365)$$

Physically this diagram includes the correction to the propagator due to the polarization of the vacuum, i.e. the creation of virtual electron-positron pairs by the propagating photon. The momentum  $q$  is the total momentum of the electron-positron pair in the intermediate channel.

It is instructive to look at this diagram from the point of view of perturbation theory in nonrelativistic Quantum Mechanics. In each vertex the interaction consists of the annihilation (resp. creation) of a photon and the creation (resp. annihilation) of an electron-positron pair. This can be implemented by the interaction Hamiltonian

$$H_{\text{int}} = e \int d^3x \bar{\psi} \gamma^\mu \psi A_\mu. \quad (366)$$

All fields inside the integral can be expressed in terms of the corresponding creation-annihilation operators for photons, electrons and positrons. In Quantum Mechanics, the change in the wave function at first order in the perturbation  $H_{\text{int}}$  is given by

$$|\gamma, \text{in}\rangle = |\gamma, \text{in}\rangle_0 + \sum_n \frac{\langle n | H_{\text{int}} | \gamma, \text{in}\rangle_0}{E_{\text{in}} - E_n} |n\rangle \quad (367)$$

and similarly for  $|\gamma, \text{out}\rangle$ , where we have denoted symbolically by  $|n\rangle$  all the possible states of the electron-positron pair. Since these states are orthogonal to  $|\gamma, \text{in}\rangle_0$ ,  $|\gamma, \text{out}\rangle_0$ , we find to order  $e^2$

$$\langle \gamma, \text{in} | \gamma', \text{out} \rangle = {}_0\langle \gamma, \text{in} | \gamma', \text{out} \rangle_0 + \sum_n \frac{{}_0\langle \gamma, \text{in} | H_{\text{int}} | n \rangle \langle n | H_{\text{int}} | \gamma', \text{out} \rangle_0}{(E_{\text{in}} - E_n)(E_{\text{out}} - E_n)} + \mathcal{O}(e^4). \quad (368)$$

Hence, we see that the diagram of Eq. (364) really corresponds to the order- $e^2$  correction to the photon propagator  $\langle \gamma, \text{in} | \gamma', \text{out} \rangle$

$$\begin{array}{c} \text{---} \gamma \text{---} \gamma' \text{---} \\ \text{---} \gamma \text{---} \circlearrowleft \text{---} \gamma' \text{---} \end{array} \longrightarrow \begin{array}{c} {}_0\langle \gamma, \text{in} | \gamma', \text{out} \rangle_0 \\ \sum_n \frac{\langle \gamma, \text{in} | H_{\text{int}} | n \rangle \langle n | H_{\text{int}} | \gamma', \text{out} \rangle}{(E_{\text{in}} - E_n)(E_{\text{out}} - E_n)}. \end{array} \quad (369)$$

Once we understood the physical meaning of the Feynman diagram to be computed we proceed to its evaluation. In principle there is no problem in computing the integral in Eq. (364) for nonzero values of the electron mass. However since here we are going to be mostly interested in seeing how the divergence of the integral results in a scale-dependent renormalization of the electric charge, we will set  $m_e = 0$ . This is something safe to do, since in the case of this diagram we are not inducing new infrared divergences in taking the electron as massless. Implementing gauge invariance and using standard techniques in the computation of Feynman diagrams (see references [1]- [11]) the polarization tensor  $\Pi_{\mu\nu}(q)$  defined in Eq. (365) can be written as

$$\Pi_{\mu\nu}(q) = (q^2 \eta_{\mu\nu} - q_\mu q_\nu) \Pi(q^2) \quad (370)$$

with

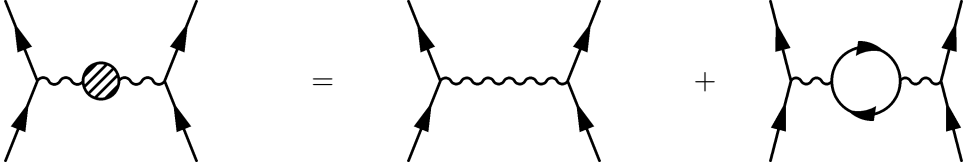
$$\Pi(q) = 8e^2 \int_0^1 dx \int \frac{d^4k}{(2\pi)^4} \frac{x(1-x)}{[k^2 - m^2 + x(1-x)q^2 + i\epsilon]^2} \quad (371)$$

To handle this divergent integral we have to figure out some procedure to render it finite. This can be done in several ways, but here we choose to cut the integrals off at a high energy scale  $\Lambda$ , where new physics might be at work,  $|p| < \Lambda$ . This gives the result

$$\Pi(q^2) \simeq \frac{e^2}{12\pi^2} \log\left(\frac{q^2}{\Lambda^2}\right) + \text{finite terms.} \quad (372)$$

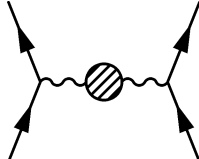
If we would send the cutoff to infinity  $\Lambda \rightarrow \infty$  the divergence blows up and something has to be done about it.

If we want to make sense out of this, we have to go back to the physical question that led us to compute Eq. (364). Our primordial motivation was to compute the corrections to the annihilation of two electrons into two muons. Including the correction to the propagator of the virtual photon we have



$$\begin{aligned} &= \eta_{\alpha\beta} (\bar{v}_e \gamma^\alpha u_e) \frac{e^2}{4\pi q^2} (\bar{v}_\mu \gamma^\beta u_\mu) + \eta_{\alpha\beta} (\bar{v}_e \gamma^\alpha u_e) \frac{e^2}{4\pi q^2} \Pi(q^2) (\bar{v}_\mu \gamma^\beta u_\mu) \\ &= \eta_{\alpha\beta} (\bar{v}_e \gamma^\alpha u_e) \left\{ \frac{e^2}{4\pi q^2} \left[ 1 + \frac{e^2}{12\pi^2} \log\left(\frac{q^2}{\Lambda^2}\right) \right] \right\} (\bar{v}_\mu \gamma^\beta u_\mu). \end{aligned} \quad (373)$$

Now let us imagine that we are performing a  $e^- e^+ \rightarrow \mu^- \mu^+$  with a center of mass energy  $\mu$ . From the previous result we can identify the effective charge of the particles at this energy scale  $e(\mu)$  as



$$= \eta_{\alpha\beta} (\bar{v}_e \gamma^\alpha u_e) \left[ \frac{e(\mu)^2}{4\pi q^2} \right] (\bar{v}_\mu \gamma^\beta u_\mu). \quad (374)$$

This charge,  $e(\mu)$ , is the quantity that is physically measurable in our experiment. Now we can make sense of the formally divergent result (373) by assuming that the charge appearing in the classical Lagrangian of QED is just a “bare” value that depends on the scale  $\Lambda$  at which we cut off the theory,  $e \equiv e(\Lambda)_{\text{bare}}$ . In order to reconcile (373) with the physical results (374) we must assume that the dependence of the bare (unobservable) charge  $e(\Lambda)_{\text{bare}}$  on the cutoff  $\Lambda$  is determined by the identity

$$e(\mu)^2 = e(\Lambda)_{\text{bare}}^2 \left[ 1 + \frac{e(\Lambda)_{\text{bare}}^2}{12\pi^2} \log\left(\frac{\mu^2}{\Lambda^2}\right) \right]. \quad (375)$$

If we still insist in removing the cutoff,  $\Lambda \rightarrow \infty$  we have to send the bare charge to zero  $e(\Lambda)_{\text{bare}} \rightarrow 0$  in such a way that the effective coupling has the finite value given by the experiment at the energy scale  $\mu$ . It is not a problem, however, that the bare charge is small for large values of the cutoff, since the only measurable quantity is the effective charge that remains finite. Therefore all observable quantities should be expressed in perturbation theory as a power series in the physical coupling  $e(\mu)^2$  and not in the unphysical bare coupling  $e(\Lambda)_{\text{bare}}$ .

## 8.2 The beta-function and asymptotic freedom

We can look at the previous discussion, in particular Eq. (375), from a different point of view. In order to remove the ambiguities associated with infinities we have been forced to introduce a dependence of the coupling constant on the energy scale at which a process takes place. From the expression of the physical coupling in terms of the bare charge (375) we can actually eliminate the cutoff  $\Lambda$ , whose value after all should not affect the value of physical quantities. Taking into account that we are working in perturbation theory in  $e(\mu)^2$ , we can express the bare charge  $e(\Lambda)_{\text{bare}}^2$  in terms of  $e(\mu)^2$  as

$$e(\Lambda)^2 = e(\mu)^2 \left[ 1 + \frac{e(\mu)^2}{12\pi^2} \log \left( \frac{\mu^2}{\Lambda^2} \right) \right] + \mathcal{O}[e(\mu)^6]. \quad (376)$$

This expression allow us to eliminate all dependence in the cutoff in the expression of the effective charge at a scale  $\mu$  by replacing  $e(\Lambda)_{\text{bare}}$  in Eq. (375) by the one computed using (376) at a given reference energy scale  $\mu_0$

$$e(\mu)^2 = e(\mu_0)^2 \left[ 1 + \frac{e(\mu_0)^2}{12\pi^2} \log \left( \frac{\mu^2}{\mu_0^2} \right) \right]. \quad (377)$$

From this equation we can compute, at this order in perturbation theory, the effective value of the coupling constant at an energy  $\mu$ , once we know its value at some reference energy scale  $\mu_0$ . In the case of the electron charge we can use as a reference Thompson's scattering at energies of the order of the electron mass  $m_e \simeq 0.5$  MeV, at where the value of the electron charge is given by the well known value

$$e(m_e)^2 \simeq \frac{1}{137}. \quad (378)$$

With this we can compute  $e(\mu)^2$  at any other energy scale applying Eq. (377), for example at the electron mass  $\mu = m_e \simeq 0.5$  MeV. However, in computing the electromagnetic coupling constant at any other scale we must take into account the fact that other charged particles can run in the loop in Eq. (373). Suppose, for example, that we want to calculate the fine structure constant at the mass of the  $Z^0$ -boson  $\mu = M_Z \equiv 92$  GeV. Then we should include in Eq. (377) the effect of other fermionic standard model fields with masses below  $M_Z$ . Doing this, we find<sup>18</sup>

$$e(M_Z)^2 = e(m_e)^2 \left[ 1 + \frac{e(m_e)^2}{12\pi^2} \left( \sum_i q_i^2 \right) \log \left( \frac{M_Z^2}{m_e^2} \right) \right], \quad (379)$$

where  $q_i$  is the charge in units of the electron charge of the  $i$ -th fermionic species running in the loop and we sum over all fermions with masses below the mass of the  $Z^0$  boson. This expression shows how the electromagnetic coupling grows with energy. However, in order to compare with the experimental value of  $e(M_Z)^2$  it is not enough with including the effect of fermionic fields, since also the  $W^\pm$  bosons

<sup>18</sup>In the first version of these notes the argument used to show the growing of the electromagnetic coupling constant could have led to confusion to some readers. To avoid this potential problem we include in the equation for the running coupling  $e(\mu)^2$  the contribution of all fermions with masses below  $M_Z$ . We thank Lubos Motl for bringing this issue to our attention.

can run in the loop ( $M_W < M_Z$ ). Taking this into account, as well as threshold effects, the value of the electron charge at the scale  $M_Z$  is found to be [37]

$$e(M_Z)^2 \simeq \frac{1}{128.9} . \quad (380)$$

This growing of the effective fine structure constant with energy can be understood heuristically by remembering that the effect of the polarization of the vacuum shown in the diagram of Eq. (364) amounts to the creation of a plethora of electron-positron pairs around the location of the charge. These virtual pairs behave as dipoles that, as in a dielectric medium, tend to screen this charge and decreasing its value at long distances (i.e. lower energies).

The variation of the coupling constant with energy is usually encoded in quantum field theory in the *beta function* defined by

$$\beta(g) = \mu \frac{dg}{d\mu} . \quad (381)$$

In the case of QED the beta function can be computed from Eq. (377) with the result

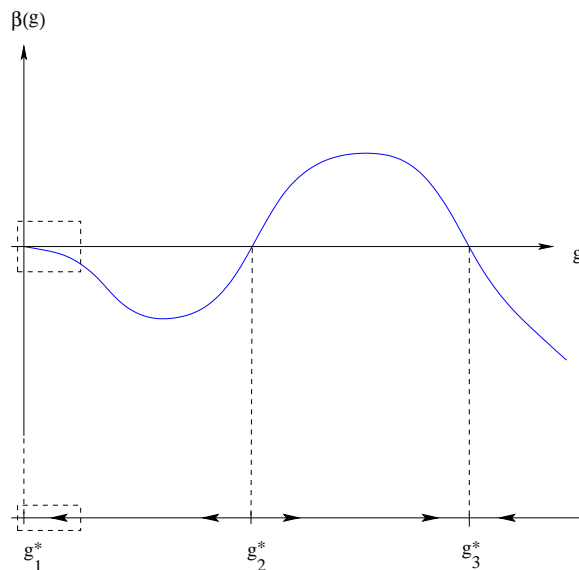
$$\beta(e)_{\text{QED}} = \frac{e^3}{12\pi^2} . \quad (382)$$

The fact that the coefficient of the leading term in the beta-function is positive  $\beta_0 \equiv \frac{1}{6\pi} > 0$  gives us the overall behavior of the coupling as we change the scale. Eq. (382) means that, if we start at an energy where the electric coupling is small enough for our perturbative treatment to be valid, the effective charge grows with the energy scale. This growing of the effective coupling constant with energy means that QED is infrared safe, since the perturbative approximation gives better and better results as we go to lower energies. Actually, because the electron is the lighter electrically charged particle and has a finite nonvanishing mass the running of the fine structure constant stops at the scale  $m_e$  in the well-known value  $\frac{1}{137}$ . Would other charged fermions with masses below  $m_e$  be present in Nature, the effective value of the fine structure constant in the interaction between these particles would run further to lower values at energies below the electron mass.

On the other hand if we increase the energy scale  $e(\mu)^2$  grows until at some scale the coupling is of order one and the perturbative approximation breaks down. In QED this is known as the problem of the Landau pole but in fact it does not pose any serious threat to the reliability of QED perturbation theory: a simple calculation shows that the energy scale at which the theory would become strongly coupled is  $\Lambda_{\text{Landau}} \simeq 10^{277}$  GeV. However, we know that QED does not live that long! At much lower scales we expect electromagnetism to be unified with other interactions, and even if this is not the case we will enter the uncharted territory of quantum gravity at energies of the order of  $10^{19}$  GeV.

So much for QED. The next question that one may ask at this stage is whether it is possible to find quantum field theories with a behavior opposite to that of QED, i.e. such that they become weakly coupled at high energies. This is not a purely academic question. In the late 1960s a series of deep-inelastic scattering experiments carried out at SLAC showed that the quarks behave essentially as free particles inside hadrons. The apparent problem was that no theory was known at that time that would become free at very short distances: the example set by QED seem to be followed by all the theories that were studied. This posed a very serious problem for quantum field theory as a way to describe subnuclear physics, since it seemed that its predictive power was restricted to electrodynamics but failed miserably when applied to describe strong interactions.

Nevertheless, this critical time for quantum field theory turned out to be its finest hour. In 1973 David Gross and Frank Wilczek [38] and David Politzer [39] showed that nonabelian gauge theories can actually display the required behavior. For the QCD Lagrangian in Eq. (331) the beta function is given



**Fig. 14:** Beta function for a hypothetical theory with three fixed points  $g_1^*$ ,  $g_2^*$  and  $g_3^*$ . A perturbative analysis would capture only the regions shown in the boxes.

by<sup>19</sup>

$$\beta(g) = -\frac{g^3}{16\pi^2} \left[ \frac{11}{3}N_c - \frac{2}{3}N_f \right]. \quad (383)$$

In particular, for real QCD ( $N_C = 3$ ,  $N_f = 6$ ) we have that  $\beta(g) = -\frac{7g^3}{16\pi^2} < 0$ . This means that for a theory that is weakly coupled at an energy scale  $\mu_0$  the coupling constant decreases as the energy increases  $\mu \rightarrow \infty$ . This explains the apparent freedom of quarks inside the hadrons: when the quarks are very close together their effective color charge tends to zero. This phenomenon is called *asymptotic freedom*.

Asymptotic free theories display a behavior that is opposite to that found above in QED. At high energies their coupling constant approaches zero whereas at low energies they become strongly coupled (infrared slavery). These features are at the heart of the success of QCD as a theory of strong interactions, since this is exactly the type of behavior found in quarks: they are quasi-free particles inside the hadrons but the interaction potential between them increases at large distances.

Although asymptotic free theories can be handled in the ultraviolet, they become extremely complicated in the infrared. In the case of QCD it is still to be understood (at least analytically) how the theory confines color charges and generates the spectrum of hadrons, as well as the breaking of the chiral symmetry (345).

In general, the ultraviolet and infrared properties of a theory are controlled by the fixed points of the beta function, i.e. those values of the coupling constant  $g$  for which it vanishes

$$\beta(g^*) = 0. \quad (384)$$

Using perturbation theory we have seen that for both QED and QCD one of such fixed points occurs at zero coupling,  $g^* = 0$ . However, our analysis also showed that the two theories present radically different behavior at high and low energies. From the point of view of the beta function, the difference lies in the energy regime at which the coupling constant approaches its critical value. This is in fact governed by the sign of the beta function around the critical coupling.

<sup>19</sup>The expression of the beta function of QCD was also known to 't Hooft [40]. There are even earlier computations in the Russian literature [41].

We have seen above that when the beta function is negative close to the fixed point (the case of QCD) the coupling tends to its critical value,  $g^* = 0$ , as the energy is increased. This means that the critical point is *ultraviolet stable*, i.e. it is an attractor as we evolve towards higher energies. If, on the contrary, the beta function is positive (as it happens in QED) the coupling constant approaches the critical value as the energy decreases. This is the case of an *infrared stable* fixed point.

This analysis that we have motivated with the examples of QED and QCD is completely general and can be carried out for any quantum field theory. In Fig. 14 we have represented the beta function for a hypothetical theory with three fixed points located at couplings  $g_1^*$ ,  $g_2^*$  and  $g_3^*$ . The arrows in the line below the plot represent the evolution of the coupling constant as the energy increases. From the analysis presented above we see that  $g_1^* = 0$  and  $g_3^*$  are ultraviolet stable fixed points, while the fixed point  $g_2^*$  is infrared stable.

In order to understand the high and low energy behavior of a quantum field theory it is then crucial to know the structure of the beta functions associated with its couplings. This can be a very difficult task, since perturbation theory only allows the study of the theory around “trivial” fixed points, i.e. those that occur at zero coupling like the case of  $g_1^*$  in Fig. 14. On the other hand, any “nontrivial” fixed point occurring in a theory (like  $g_2^*$  and  $g_3^*$ ) cannot be captured in perturbation theory and requires a full nonperturbative analysis.

The moral to be learned from our discussion above is that dealing with the ultraviolet divergences in a quantum field theory has the consequence, among others, of introducing an energy dependence in the measured value of the coupling constants of the theory (for example the electric charge in QED). This happens even in the case of renormalizable theories without mass terms. These theories are scale invariant at the classical level because the action does not contain any dimensionful parameter. In this case the running of the coupling constants can be seen as resulting from a quantum breaking of classical scale invariance: different energy scales in the theory are distinguished by different values of the coupling constants. Remembering what we learned in Section 7, we conclude that classical scale invariance is an anomalous symmetry. One heuristic way to see how the conformal anomaly comes about is to notice that the regularization of an otherwise scale invariant field theory requires the introduction of an energy scale (e.g. a cutoff). This breaking of scale invariance cannot be restored after renormalization.

Nevertheless, scale invariance is not lost forever in the quantum theory. It is recovered at the fixed points of the beta function where, by definition, the coupling does not run. To understand how this happens we go back to a scale invariant classical field theory whose field  $\phi(x)$  transform under coordinate rescalings as

$$x^\mu \longrightarrow \lambda x^\mu, \quad \phi(x) \longrightarrow \lambda^{-\Delta} \phi(\lambda^{-1}x), \quad (385)$$

where  $\Delta$  is called the canonical scaling dimension of the field. An example of such a theory is a massless  $\phi^4$  theory in four dimensions

$$\mathcal{L} = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{g}{4!} \phi^4, \quad (386)$$

where the scalar field has canonical scaling dimension  $\Delta = 1$ . The Lagrangian density transforms as

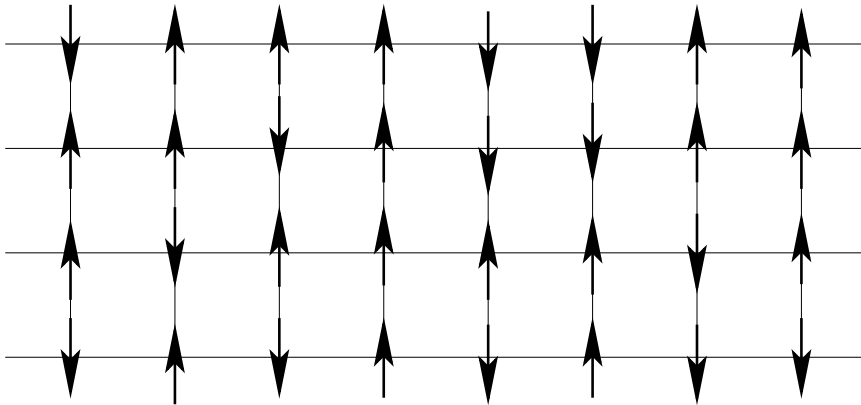
$$\mathcal{L} \longrightarrow \lambda^{-4} \mathcal{L}[\phi] \quad (387)$$

and the classical action remains invariant<sup>20</sup>.

If scale invariance is preserved under quantization, the Green’s functions transform as

$$\langle \Omega | T[\phi'(x_1) \dots \phi'(x_n)] | \Omega \rangle = \lambda^{n\Delta} \langle \Omega | T[\phi(\lambda^{-1}x_1) \dots \phi(\lambda^{-1}x_n)] | \Omega \rangle. \quad (388)$$

<sup>20</sup>In a  $D$ -dimensional theory the canonical scaling dimensions of the fields coincide with its engineering dimension:  $\Delta = \frac{D-2}{2}$  for bosonic fields and  $\Delta = \frac{D-1}{2}$  for fermionic ones. For a Lagrangian with no dimensionful parameters classical scale invariance follows then from dimensional analysis.



**Fig. 15:** Systems of spins in a two-dimensional square lattice.

This is precisely what happens in a free theory. In an interacting theory the running of the coupling constant destroys classical scale invariance at the quantum level. Despite of this, at the fixed points of the beta function the Green's functions transform again according to (388) where  $\Delta$  is replaced by

$$\Delta_{\text{anom}} = \Delta + \gamma^*. \quad (389)$$

The canonical scaling dimension of the fields are corrected by  $\gamma^*$ , which is called the anomalous dimension. They carry the dynamical information about the high-energy behavior of the theory.

### 8.3 The renormalization group

In spite of its successes, the renormalization procedure presented above can be seen as some kind of prescription or recipe to get rid of the divergences in an ordered way. This discomfort about renormalization was expressed in occasions by comparing it with “sweeping the infinities under the rug”. However thanks to Ken Wilson to a large extent [42] the process of renormalization is now understood in a very profound way as a procedure to incorporate the effects of physics at high energies by modifying the value of the parameters that appear in the Lagrangian.

**Statistical mechanics.** Wilson's ideas are both simple and profound and consist in thinking about quantum field theory as the analog of a thermodynamical description of a statistical system. To be more precise, let us consider an Ising spin system in a two-dimensional square lattice as the one depicted in Fig 15. In terms of the spin variables  $s_i = \pm \frac{1}{2}$ , where  $i$  labels the lattice site, the Hamiltonian of the system is given by

$$H = -J \sum_{\langle i,j \rangle} s_i s_j, \quad (390)$$

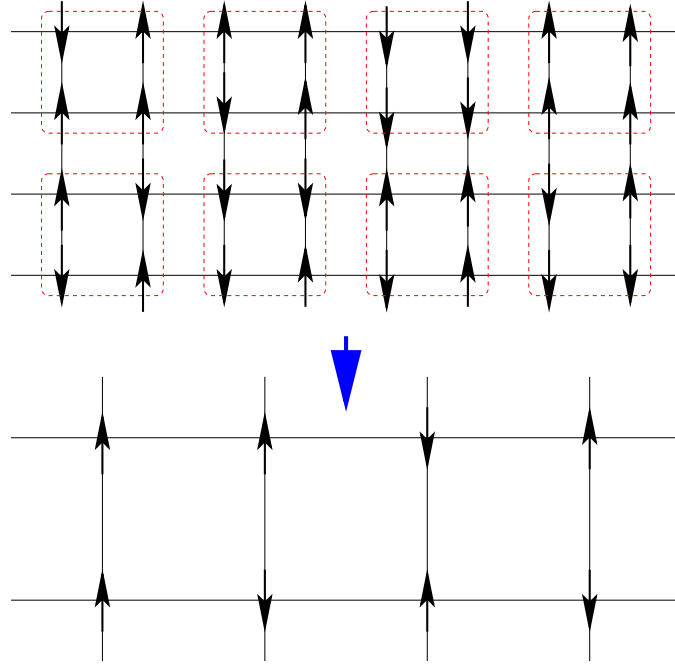
where  $\langle i,j \rangle$  indicates that the sum extends over nearest neighbors and  $J$  is the coupling constant between neighboring spins (here we consider that there is no external magnetic field). The starting point to study the statistical mechanics of this system is the partition function defined as

$$\mathcal{Z} = \sum_{\{s_i\}} e^{-\beta H}, \quad (391)$$

where the sum is over all possible configurations of the spins and  $\beta = \frac{1}{T}$  is the inverse temperature. For  $J > 0$  the Ising model presents spontaneous magnetization below a critical temperature  $T_c$ , in any dimension higher than one. Away from this temperature correlations between spins decay exponentially at large distances

$$\langle s_i s_j \rangle \sim e^{-\frac{|x_{ij}|}{\xi}}, \quad (392)$$





**Fig. 16:** Decimation of the spin lattice. Each block in the upper lattice is replaced by an effective spin computed according to the rule (394). Notice also that the size of the lattice spacing is doubled in the process.

with  $|x_{ij}|$  the distance between the spins located in the  $i$ -th and  $j$ -th sites of the lattice. This expression serves as a definition of the correlation length  $\xi$  which sets the characteristic length scale at which spins can influence each other by their interaction through their nearest neighbors.

Suppose now that we are interested in a macroscopic description of this spin system. We can capture the relevant physics by integrating out somehow the physics at short scales. A way in which this can be done was proposed by Leo Kadanoff [43] and consists in dividing our spin system in spin-blocks like the ones showed in Fig 16. Now we can construct another spin system where each spin-block of the original lattice is replaced by an effective spin calculated according to some rule from the spins contained in each block  $B_a$

$$\{s_i : i \in B_a\} \longrightarrow s_a^{(1)}. \quad (393)$$

For example we can define the effective spin associated with the block  $B_a$  by taking the majority rule with an additional prescription in case of a draw

$$s_a^{(1)} = \frac{1}{2} \text{sgn} \left( \sum_{i \in B_a} s_i \right), \quad (394)$$

where we have used the sign function,  $\text{sign}(x) \equiv \frac{x}{|x|}$ , with the additional definition  $\text{sign}(0) = 1$ . This procedure is called decimation and leads to a new spin system with a doubled lattice space.

The idea now is to rewrite the partition function (391) only in terms of the new effective spins  $s_a^{(1)}$ . Then we start by splitting the sum over spin configurations into two nested sums, one over the spin blocks and a second one over the spins within each block

$$\mathcal{Z} = \sum_{\{\vec{s}\}} e^{-\beta H[\vec{s}]} = \sum_{\{\vec{s}^{(1)}\}} \sum_{\{\vec{s} \in B_a\}} \delta \left[ s_a^{(1)} - \text{sign} \left( \sum_{i \in B_a} s_i \right) \right] e^{-\beta H[\vec{s}]} \quad (395)$$

The interesting point now is that the sum over spins inside each block can be written as the exponential of a new effective Hamiltonian depending only on the effective spins,  $H^{(1)}[s_a^{(1)}]$

$$\sum_{\{s \in B_a\}} \delta \left[ s_a^{(1)} - \text{sign} \left( \sum_{i \in B_a} s_i \right) \right] e^{-\beta H[s_i]} = e^{-\beta H^{(1)}[s_a^{(1)}]}. \quad (396)$$

The new Hamiltonian is of course more complicated

$$H^{(1)} = -J^{(1)} \sum_{\langle i,j \rangle} s_i^{(1)} s_j^{(1)} + \dots \quad (397)$$

where the dots stand for other interaction terms between the effective block spins. This new terms appear because in the process of integrating out short distance physics we induce interactions between the new effective degrees of freedom. For example the interaction between the spin block variables  $s_i^{(1)}$  will in general not be restricted to nearest neighbors in the new lattice. The important point is that we have managed to rewrite the partition function solely in terms of this new (renormalized) spin variables  $s^{(1)}$  interacting through a new Hamiltonian  $H^{(1)}$

$$\mathcal{Z} = \sum_{\{s^{(1)}\}} e^{-\beta H^{(1)}[s_a^{(1)}]}. \quad (398)$$

Let us now think about the space of all possible Hamiltonians for our statistical system including all kinds of possible couplings between the individual spins compatible with the symmetries of the system. If denote by  $\mathcal{R}$  the decimation operation, our previous analysis shows that  $\mathcal{R}$  defines a map in this space of Hamiltonians

$$\mathcal{R} : H \rightarrow H^{(1)}. \quad (399)$$

At the same time the operation  $\mathcal{R}$  replaces a lattice with spacing  $a$  by another one with double spacing  $2a$ . As a consequence the correlation length in the new lattice measured in units of the lattice spacing is divided by two,  $\mathcal{R} : \xi \rightarrow \frac{\xi}{2}$ .

Now we can iterate the operation  $\mathcal{R}$  an indefinite number of times. Eventually we might reach a Hamiltonian  $H_*$  that is not further modified by the operation  $\mathcal{R}$

$$H \xrightarrow{\mathcal{R}} H^{(1)} \xrightarrow{\mathcal{R}} H^{(2)} \xrightarrow{\mathcal{R}} \dots \xrightarrow{\mathcal{R}} H_*. \quad (400)$$

The fixed point Hamiltonian  $H_*$  is *scale invariant* because it does not change as  $\mathcal{R}$  is performed. Notice that because of this invariance the correlation length of the system at the fixed point do not change under  $\mathcal{R}$ . This fact is compatible with the transformation  $\xi \rightarrow \frac{\xi}{2}$  only if  $\xi = 0$  or  $\xi = \infty$ . Here we will focus in the case of nontrivial fixed points with infinite correlation length.

The space of Hamiltonians can be parametrized by specifying the values of the coupling constants associated with all possible interaction terms between individual spins of the lattice. If we denote by  $\mathcal{O}_a[s_i]$  these (possibly infinite) interaction terms, the most general Hamiltonian for the spin system under study can be written as

$$H[s_i] = \sum_{a=1}^{\infty} \lambda_a \mathcal{O}_a[s_i], \quad (401)$$

where  $\lambda_a \in \mathbb{R}$  are the coupling constants for the corresponding operators. These constants can be thought of as coordinates in the space of all Hamiltonians. Therefore the operation  $\mathcal{R}$  defines a transformation in the set of coupling constants

$$\mathcal{R} : \lambda_a \longrightarrow \lambda_a^{(1)}. \quad (402)$$

For example, in our case we started with a Hamiltonian in which only one of the coupling constants is different from zero (say  $\lambda_1 = -J$ ). As a result of the decimation  $\lambda_1 \equiv -J \rightarrow -J^{(1)}$  while some of the originally vanishing coupling constants will take a nonzero value. Of course, for the fixed point Hamiltonian the coupling constants do not change under the scale transformation  $\mathcal{R}$ .

Physically the transformation  $\mathcal{R}$  integrates out short distance physics. The consequence for physics at long distances is that we have to replace our Hamiltonian by a new one with different values for the coupling constants. That is, our ignorance of the details of the physics going on at short distances result in a *renormalization* of the coupling constants of the Hamiltonian that describes the long range physical processes. It is important to stress that although  $\mathcal{R}$  is sometimes called a renormalization group transformation in fact this is a misnomer. Transformations between Hamiltonians defined by  $\mathcal{R}$  do not form a group: since these transformations proceed by integrating out degrees of freedom at short scales they cannot be inverted.

In statistical mechanics fixed points under renormalization group transformations with  $\xi = \infty$  are associated with phase transitions. From our previous discussion we can conclude that the space of Hamiltonians is divided in regions corresponding to the basins of attraction of the different fixed points. We can ask ourselves now about the stability of those fixed points. Suppose we have a statistical system described by a fixed-point Hamiltonian  $H_*$  and we perturb it by changing the coupling constant associated with an interaction term  $\mathcal{O}$ . This is equivalent to replace  $H_*$  by the perturbed Hamiltonian

$$H = H_* + \delta\lambda \mathcal{O}, \quad (403)$$

where  $\delta\lambda$  is the perturbation of the coupling constant corresponding to  $\mathcal{O}$  (we can also consider perturbations in more than one coupling constant). At the same time thinking of the  $\lambda_a$ 's as coordinates in the space of all Hamiltonians this corresponds to moving slightly away from the position of the fixed point.

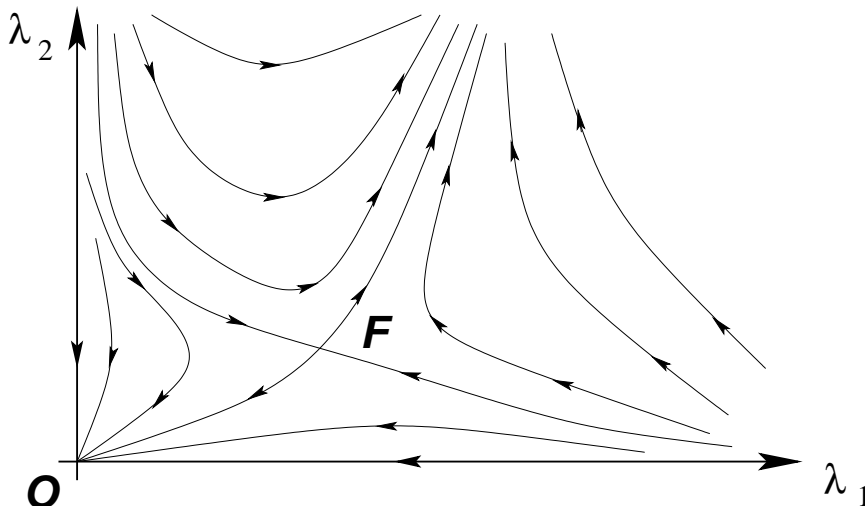
The question to decide now is in which direction the renormalization group flow will take the perturbed system. Working at first order in  $\delta\lambda$  there are three possibilities:

- The renormalization group flow takes the system back to the fixed point. In this case the corresponding interaction  $\mathcal{O}$  is called *irrelevant*.
- $\mathcal{R}$  takes the system away from the fixed point. If this is what happens the interaction is called *relevant*.
- It is possible that the perturbation actually does not take the system away from the fixed point at first order in  $\delta\lambda$ . In this case the interaction is said to be *marginal* and it is necessary to go to higher orders in  $\delta\lambda$  in order to decide whether the system moves to or away the fixed point, or whether we have a family of fixed points.

Therefore we can picture the action of the renormalization group transformation as a flow in the space of coupling constants. In Fig. 17 we have depicted an example of such a flow in the case of a system with two coupling constants  $\lambda_1$  and  $\lambda_2$ . In this example we find two fixed points, one at the origin  $O$  and another at  $F$  for a finite value of the couplings. The arrows indicate the direction in which the renormalization group flow acts. The free theory at  $\lambda_1 = \lambda_2 = 0$  is a stable fix point since any perturbation  $\delta\lambda_1, \delta\lambda_2 > 0$  makes the theory flow back to the free theory at long distances. On the other hand, the fixed point  $F$  is stable with respect to certain type of perturbations (along the line with incoming arrows) whereas for any other perturbations the system flows either to the free theory at the origin or to a theory with infinite values for the couplings.

**Quantum field theory.** Let us see now how these ideas of the renormalization group apply to Field Theory. Let us begin with a quantum field theory defined by the Lagrangian

$$\mathcal{L}[\phi_a] = \mathcal{L}_0[\phi_a] + \sum_i g_i \mathcal{O}_i[\phi_a], \quad (404)$$



**Fig. 17:** Example of a renormalization group flow.

where  $\mathcal{L}_0[\phi_a]$  is the kinetic part of the Lagrangian and  $g_i$  are the coupling constants associated with the operators  $\mathcal{O}_i[\phi_a]$ . In order to make sense of the quantum theory we introduce a cutoff in momenta  $\Lambda$ . In principle we include all operators  $\mathcal{O}_i$  compatible with the symmetries of the theory.

In section 8.2 we saw how in the cases of QED and QCD, the value of the coupling constant changed with the scale from its value at the scale  $\Lambda$ . We can understand now this behavior along the lines of the analysis presented above for the Ising model. If we would like to compute the effective dynamics of the theory at an energy scale  $\mu < \Lambda$  we only have to integrate out all physical models with energies between the cutoff  $\Lambda$  and the scale of interest  $\mu$ . This is analogous to what we did in the Ising model by replacing the original spins by the block spins. In the case of field theory the effective action  $S[\phi_a, \mu]$  at scale  $\mu$  can be written in the language of functional integration as

$$e^{iS[\phi'_a, \mu]} = \int_{\mu < p < \Lambda} \prod_a \mathcal{D}\phi_a e^{iS[\phi_a, \Lambda]}. \quad (405)$$

Here  $S[\phi_a, \Lambda]$  is the action at the cutoff scale

$$S[\phi_a, \Lambda] = \int d^4x \left\{ \mathcal{L}_0[\phi_a] + \sum_i g_i(\Lambda) \mathcal{O}_i[\phi_a] \right\} \quad (406)$$

and the functional integral in Eq. (405) is carried out only over the field modes with momenta in the range  $\mu < p < \Lambda$ . The action resulting from integrating out the physics at the intermediate scales between  $\Lambda$  and  $\mu$  depends not on the original field variable  $\phi_a$  but on some renormalized field  $\phi'_a$ . At the same time the couplings  $g_i(\mu)$  differ from their values at the cutoff scale  $g_i(\Lambda)$ . This is analogous to what we learned in the Ising model: by integrating out short distance physics we ended up with a new Hamiltonian depending on renormalized effective spin variables and with renormalized values for the coupling constants. Therefore the resulting effective action at scale  $\mu$  can be written as

$$S[\phi'_a, \mu] = \int d^4x \left\{ \mathcal{L}_0[\phi'_a] + \sum_i g_i(\mu) \mathcal{O}_i[\phi'_a] \right\}. \quad (407)$$

This Wilsonian interpretation of renormalization sheds light to what in section 8.1 might have looked just a smart way to get rid of the infinities. The running of the coupling constant with the energy scale can be understood now as a way of incorporating into an effective action at scale  $\mu$  the effects of field excitations at higher energies  $E > \mu$ .

As in statistical mechanics there are also quantum field theories that are fixed points of the renormalization group flow, i.e. whose coupling constants do not change with the scale. We have encountered them already in Section 8.2 when studying the properties of the beta function. The most trivial example of such theories are massless free quantum field theories, but there are also examples of four-dimensional interacting quantum field theories which are scale invariant. Again we can ask the question of what happens when a scale invariant theory is perturbed with some operator. In general the perturbed theory is not scale invariant anymore but we may wonder whether the perturbed theory flows at low energies towards or away the theory at the fixed point.

In quantum field theory this can be decided by looking at the canonical dimension  $d[\mathcal{O}]$  of the operator  $\mathcal{O}[\phi_a]$  used to perturb the theory at the fixed point. In four dimensions the three possibilities are defined by:

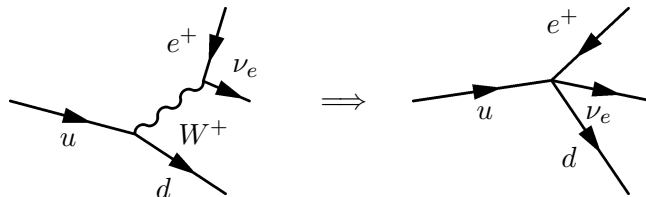
- $d[\mathcal{O}] > 4$ : irrelevant perturbation. The running of the coupling constants takes the theory back to the fixed point.
- $d[\mathcal{O}] < 4$ : relevant perturbation. At low energies the theory flows away from the scale-invariant theory.
- $d[\mathcal{O}] = 4$ : marginal deformation. The direction of the flow cannot be decided only on dimensional grounds.

As an example, let us consider first a massless fermion theory perturbed by a four-fermion interaction term

$$\mathcal{L} = i\bar{\psi}\not{\partial}\psi - \frac{1}{M^2}(\bar{\psi}\psi)^2. \tag{408}$$

This is indeed a perturbation by an irrelevant operator, since in four-dimensions  $[\psi] = \frac{3}{2}$ . Interactions generated by the extra term are suppressed at low energies since typically their effects are weighted by the dimensionless factor  $\frac{E^2}{M^2}$ , where  $E$  is the energy scale of the process. This means that as we try to capture the relevant physics at lower and lower energies the effect of the perturbation is weaker and weaker rendering in the infrared limit  $E \rightarrow 0$  again a free theory. Hence, the irrelevant perturbation in (408) makes the theory flow back to the fixed point.

On the other hand relevant operators dominate the physics at low energies. This is the case, for example, of a mass term. As we lower the energy the mass becomes more important and once the energy goes below the mass of the field its dynamics is completely dominated by the mass term. This is, for example, how Fermi's theory of weak interactions emerges from the standard model at energies below the mass of the  $W^\pm$  boson



At energies below  $M_W = 80.4$  GeV the dynamics of the  $W^+$  boson is dominated by its mass term and therefore becomes nonpropagating, giving rise to the effective four-fermion Fermi theory.

To summarize our discussion so far, we found that while relevant operators dominate the dynamics in the infrared, taking the theory away from the fixed point, irrelevant perturbations become suppressed in the same limit. Finally we consider the effect of marginal operators. As an example we take the interaction term in massless QED,  $\mathcal{O} = \bar{\psi}\gamma^\mu\psi A_\mu$ . Taking into account that in  $d = 4$  the dimension of the electromagnetic potential is  $[A_\mu] = 1$  the operator  $\mathcal{O}$  is a marginal perturbation. In order to decide whether the fixed point theory

$$\mathcal{L}_0 = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\psi \tag{409}$$

is restored at low energies or not we need to study the perturbed theory in more detail. This we have done in section 8.1 where we learned that the effective coupling in QED decreases at low energies. Then we conclude that the perturbed theory flows towards the fixed point in the infrared.

As an example of a marginal operator with the opposite behavior we can write the Lagrangian for a  $SU(N_c)$  gauge theory,  $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^a F^{a\mu\nu}$ , as

$$\begin{aligned} \mathcal{L} &= -\frac{1}{4}(\partial_\mu A_\nu^a - \partial_\nu A_\mu^a)(\partial^\mu A^{a\nu} - \partial^\nu A^{a\mu}) - 4gf^{abc}A_\mu^a A_\nu^b \partial^\mu A^{c\nu} \\ &+ g^2 f^{abc} f^{ade} A_\mu^b A_\nu^c A^{d\mu} A^{e\nu} \equiv \mathcal{L}_0 + \mathcal{O}_g, \end{aligned} \quad (410)$$

i.e. a marginal perturbation of the free theory described by  $\mathcal{L}_0$ , which is obviously a fixed point under renormalization group transformations. Unlike the case of QED we know that the full theory is asymptotically free, so the coupling constant grows at low energies. This implies that the operator  $\mathcal{O}_g$  becomes more and more important in the infrared and therefore the theory flows away the fixed point in this limit.

It is very important to notice here that in the Wilsonian view the cutoff is not necessarily regarded as just some artifact to remove infinities but actually has a physical origin. For example in the case of Fermi's theory of  $\beta$ -decay there is a natural cutoff  $\Lambda = M_W$  at which the theory has to be replaced by the standard model. In the case of the standard model itself the cutoff can be taken at Planck scale  $\Lambda \simeq 10^{19}$  GeV or the Grand Unification scale  $\Lambda \simeq 10^{16}$  GeV, where new degrees of freedom are expected to become relevant. The cutoff serves the purpose of cloaking the range of energies at which new physics has to be taken into account.

Provided that in the Wilsonian approach the quantum theory is always defined with a physical cutoff, there is no fundamental difference between renormalizable and nonrenormalizable theories. Actually, a renormalizable field theory, like the standard model, can generate nonrenormalizable operators at low energies such as the effective four-fermion interaction of Fermi's theory. They are not sources of any trouble if we are interested in the physics at scales much below the cutoff,  $E \ll \Lambda$ , since their contribution to the amplitudes will be suppressed by powers of  $\frac{E}{\Lambda}$ .

## 9 Special topics

### 9.1 Creation of particles by classical fields

**Particle creation by a classical source.** In a free quantum field theory the total number of particles contained in a given state of the field is a conserved quantity. For example, in the case of the quantum scalar field studied in section 3 we have that the number operator commutes with the Hamiltonian

$$\hat{n} \equiv \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega_k} \alpha^\dagger(\vec{k})\alpha(\vec{k}), \quad [\hat{H}, \hat{n}] = 0. \quad (411)$$

This means that any states with a well-defined number of particle excitations will preserve this number at all times. The situation, however, changes as soon as interactions are introduced, since in this case particles can be created and/or destroyed as a result of the dynamics.

Another case in which the number of particles might change is if the quantum theory is coupled to a classical source. The archetypical example of such a situation is the Schwinger effect, in which a classical strong electric field produces the creation of electron-positron pairs out of the vacuum. However, before plunging into this more involved situation we can illustrate the relevant physics involved in the creation of particles by classical sources with the help of the simplest example: a free scalar field theory coupled to a classical external source  $J(x)$ . The action for such a theory can be written as

$$S = \int d^4x \left[ \frac{1}{2} \partial_\mu \phi(x) \partial^\mu \phi(x) - \frac{m^2}{2} \phi(x)^2 + J(x) \phi(x) \right], \quad (412)$$

where  $J(x)$  is a real function of the coordinates. Its identification with a classical source is obvious once we calculate the equations of motion

$$(\nabla^2 + m^2) \phi(x) = J(x). \quad (413)$$

Our plan is to quantize this theory but, unlike the case analyzed in section 3, now the presence of the source  $J(x)$  makes the situation a bit more involved. The general solution to the equations of motion can be written in terms of the retarded Green function for the Klein-Gordon equation as

$$\phi(x) = \phi_0(x) + i \int d^4x' G_R(x - x') J(x'), \quad (414)$$

where  $\phi_0(x)$  is a general solution to the homogeneous equation and

$$\begin{aligned} G_R(t, \vec{x}) &= \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2 + i\epsilon \operatorname{sign}(k^0)} e^{-ik \cdot x} \\ &= i \theta(t) \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega_k} \left( e^{-i\omega_k t + \vec{k} \cdot \vec{x}} - e^{i\omega_k t - i\vec{k} \cdot \vec{x}} \right), \end{aligned} \quad (415)$$

with  $\theta(x)$  the Heaviside step function. The integration contour to evaluate the integral over  $p^0$  surrounds the poles at  $p^0 = \pm\omega_k$  from above. Since  $G_R(t, \vec{x}) = 0$  for  $t < 0$ , the function  $\phi_0(x)$  corresponds to the solution of the field equation at  $t \rightarrow -\infty$ , before the interaction with the external source<sup>21</sup>

To make the argument simpler we assume that  $J(x)$  is switched on at  $t = 0$ , and only last for a time  $\tau$ , that is

$$J(t, \vec{x}) = 0 \quad \text{if } t < 0 \text{ or } t > \tau. \quad (416)$$

We are interested in a solution of (413) for times after the external source has been switched off,  $t > \tau$ . In this case the expression (415) can be written in terms of the Fourier modes  $\tilde{J}(\omega, \vec{k})$  of the source as

$$\phi(t, \vec{x}) = \phi_0(x) + i \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega_k} \left[ \tilde{J}(\omega_k, \vec{k}) e^{-i\omega_k t + i\vec{k} \cdot \vec{x}} - \tilde{J}(\omega_k, \vec{k})^* e^{i\omega_k t - i\vec{k} \cdot \vec{x}} \right]. \quad (417)$$

On the other hand, the general solution  $\phi_0(x)$  has been already computed in Eq. (77). Combining this result with Eq. (417) we find the following expression for the late time general solution to the Klein-Gordon equation in the presence of the source

$$\begin{aligned} \phi(t, x) &= \int \frac{d^3k}{(2\pi)^3} \frac{1}{\sqrt{2\omega_k}} \left\{ \left[ \alpha(\vec{k}) + \frac{i}{\sqrt{2\omega_k}} \tilde{J}(\omega_k, \vec{k}) \right] e^{-i\omega_k t + i\vec{k} \cdot \vec{x}} \right. \\ &\quad \left. + \left[ \alpha^*(\vec{k}) - \frac{i}{\sqrt{2\omega_k}} \tilde{J}(\omega_k, \vec{k})^* \right] e^{i\omega_k t - i\vec{k} \cdot \vec{x}} \right\}. \end{aligned} \quad (418)$$

We should not forget that this is a solution valid for times  $t > \tau$ , i.e. once the external source has been disconnected. On the other hand, for  $t < 0$  we find from Eqs. (414) and (415) that the general solution is given by Eq. (77).

Now we can proceed to quantize the theory. The conjugate momentum  $\pi(x) = \partial_0 \phi(x)$  can be computed from Eqs. (77) and (418). Imposing the canonical equal time commutation relations (74) we find that  $\alpha(\vec{k})$ ,  $\alpha^\dagger(\vec{k})$  satisfy the creation-annihilation algebra (51). From our previous calculation we find that for  $t > \tau$  the expansion of the operator  $\phi(x)$  in terms of the creation-annihilation operators  $\alpha(\vec{k})$ ,  $\alpha^\dagger(\vec{k})$  can be obtained from the one for  $t < 0$  by the replacement

$$\alpha(\vec{k}) \longrightarrow \beta(\vec{k}) \equiv \alpha(\vec{k}) + \frac{i}{\sqrt{2\omega_k}} \tilde{J}(\omega_k, \vec{k}),$$

<sup>21</sup>We could have taken instead the advanced propagator  $G_A(x)$  in which case  $\phi_0(x)$  would correspond to the solution to the equation at large times, after the interaction with  $J(x)$ .

$$\alpha^\dagger(\vec{k}) \longrightarrow \beta^\dagger(\vec{k}) \equiv \alpha^\dagger(\vec{k}) - \frac{i}{\sqrt{2\omega_k}} \tilde{J}(\omega_k, \vec{k})^*. \quad (419)$$

Actually, since  $\tilde{J}(\omega_k, \vec{k})$  is a c-number, the operators  $\beta(\vec{k}), \beta^\dagger(\vec{k})$  satisfy the same algebra as  $\alpha(\vec{k}), \alpha^\dagger(\vec{k})$  and therefore can be interpreted as well as a set of creation-annihilation operators. This means that we can define two vacuum states,  $|0_-\rangle, |0_+\rangle$  associated with both sets of operators

$$\left. \begin{aligned} \alpha(\vec{k})|0_-\rangle &= 0 \\ \beta(\vec{k})|0_+\rangle &= 0 \end{aligned} \right\} \quad \forall \vec{k}. \quad (420)$$

For an observer at  $t < 0$ ,  $\alpha(\vec{k})$  and  $\alpha^\dagger(\vec{k})$  are the natural set of creation-annihilation operators in terms of which to expand the field operator  $\phi(x)$ . After the usual zero-point energy subtraction the Hamiltonian is given by

$$\hat{H}^{(-)} = \frac{1}{2} \int \frac{d^3k}{(2\pi)^3} \alpha^\dagger(\vec{k}) \alpha(\vec{k}) \quad (421)$$

and the ground state of the spectrum for this observer is the vacuum  $|0_-\rangle$ . At the same time, a second observer at  $t > \tau$  will also see a free scalar quantum field (the source has been switched off at  $t = \tau$ ) and consequently will expand  $\phi$  in terms of the second set of creation-annihilation operators  $\beta(\vec{k}), \beta^\dagger(\vec{k})$ . In terms of this operators the Hamiltonian is written as

$$\hat{H}^{(+)} = \frac{1}{2} \int \frac{d^3k}{(2\pi)^3} \beta^\dagger(\vec{k}) \beta(\vec{k}). \quad (422)$$

Then for this late-time observer the ground state of the Hamiltonian is the second vacuum state  $|0_+\rangle$ .

In our analysis we have been working in the Heisenberg picture, where states are time-independent and the time dependence comes in the operators. Therefore the states of the theory are globally defined. Suppose now that the system is in the “in” ground state  $|0_-\rangle$ . An observer at  $t < 0$  will find that there are no particles

$$\hat{n}^{(-)}|0_-\rangle = 0. \quad (423)$$

However the late-time observer will find that the state  $|0_-\rangle$  contains an average number of particles given by

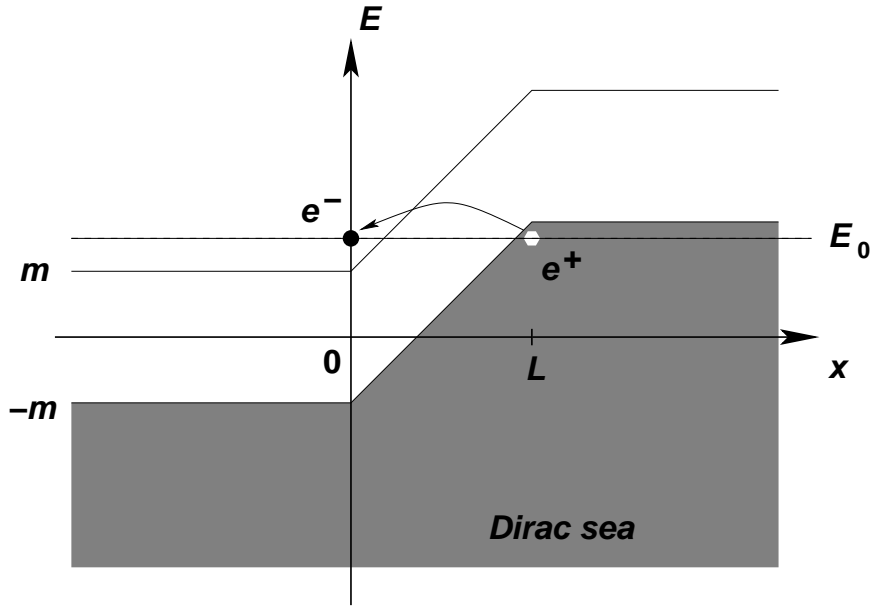
$$\langle 0_- | \hat{n}^{(+)} | 0_- \rangle = \int \frac{d^3k}{(2\pi)^3} \frac{1}{2\omega_k} \left| \tilde{J}(\omega_k, \vec{k}) \right|^2. \quad (424)$$

Moreover,  $|0_-\rangle$  is no longer the ground state for the “out” observer. On the contrary, this state have a vacuum expectation value for  $\hat{H}^{(+)}$

$$\langle 0_- | \hat{H}^{(+)} | 0_- \rangle = \frac{1}{2} \int \frac{d^3k}{(2\pi)^3} \left| \tilde{J}(\omega_k, \vec{k}) \right|^2. \quad (425)$$

The key to understand what is going on here lies in the fact that the external source breaks the invariance of the theory under space-time translations. In the particular case we have studied here where  $J(x)$  has support over a finite time interval  $0 < t < \tau$ , this implies that the vacuum is not invariant under time translations, so observers at different times will make different choices of vacuum that will not necessarily agree with each other. This is clear in our example. An observer in  $t < \tau$  will choose the vacuum to be the lowest energy state of her Hamiltonian,  $|0_-\rangle$ . On the other hand, the second observer at late times  $t > \tau$  will naturally choose  $|0_+\rangle$  as the vacuum. However, for this second observer, the





**Fig. 18:** Pair creation by a electric field in the Dirac sea picture.

state  $|0_-\rangle$  is not the vacuum of his Hamiltonian, but actually an excited state that is a superposition of states with well-defined number of particles. In this sense it can be said that the external source has the effect of creating particles out of the “in” vacuum. Besides, this breaking of time translation invariance produces a violation in the energy conservation as we see from Eq. (425). Particles are actually created from the energy pumped into the system by the external source.

**The Schwinger effect.** A classical example of creation of particles by a external field was pointed out by Schwinger [44] and consists of the creation of electron-positron pairs by a strong electric field. In order to illustrate this effect we are going to follow a heuristic argument based on the Dirac sea picture and the WKB approximation.

In the absence of an electric field the vacuum state of a spin- $\frac{1}{2}$  field is constructed by filling all the negative energy states as depicted in Fig. 2. Let us now connect a constant electric field  $\vec{\mathcal{E}} = \mathcal{E}\vec{u}_x$  in the range  $0 < x < L$  created by an electrostatic potential

$$V(\vec{r}) = \begin{cases} 0 & x < 0 \\ -\mathcal{E}x & 0 < x < L \\ -\mathcal{E}L & x > L \end{cases} \quad (426)$$

After the field has been switched on, the Dirac sea looks like in Fig. 18. In particular we find that if  $e\mathcal{E}L > 2m$  there are negative energy states at  $x > L$  with the same energy as the positive energy states in the region  $x < 0$ . Therefore it is possible for an electron filling a negative energy state with energy close to  $-2m$  to tunnel through the forbidden region into a positive energy state. The interpretation of such a process is the production of an electron-positron pair out of the electric field.

We can compute the rate at which such pairs are produced by using the WKB approximation. Focusing for simplicity on an electron on top of the Fermi surface near  $x = L$  with energy  $E_0$ , the transmission coefficient in this approximation is given by<sup>22</sup>

$$T_{\text{WKB}} = \exp \left[ -2 \int_{\frac{1}{e\mathcal{E}}(E_0 - \sqrt{m^2 + \vec{p}_T^2})}^{\frac{1}{e\mathcal{E}}(E_0 + \sqrt{m^2 + \vec{p}_T^2})} dx \sqrt{m^2 - [E_0 - e\mathcal{E}(x - x_0)]^2 + \vec{p}_T^2} \right]$$

<sup>22</sup>Notice that the electron satisfy the relativistic dispersion relation  $E = \sqrt{\vec{p}^2 + m^2} + V$  and therefore  $-p_x^2 = m^2 - (E - V)^2 + \vec{p}_T^2$ . The integration limits are set by those values of  $x$  at which  $p_x = 0$ .

$$= \exp \left[ -\frac{\pi}{e\mathcal{E}} (\vec{p}_T^2 + m^2) \right], \quad (427)$$

where  $p_T^2 \equiv p_y^2 + p_z^2$ . This gives the transition probability per unit time and per unit cross section  $dydz$  for an electron in the Dirac sea with transverse momentum  $\vec{p}_T$  and energy  $E_0$ . To get the total probability per unit time and per unit volume we have to integrate over all possible values of  $\vec{p}_T$  and  $E_0$ . Actually, in the case of the energy, because of the relation between  $E_0$  and the coordinate  $x$  at which the particle penetrates into the barrier we can write  $\frac{dE_0}{2\pi} = \frac{e\mathcal{E}}{2\pi} dx$  and the total probability per unit time and per unit volume for the creation of a pair is given by

$$W = 2 \left( \frac{e\mathcal{E}}{2\pi} \right) \int \frac{d^2 p_T}{(2\pi)^2} e^{-\frac{\pi}{e\mathcal{E}} (\vec{p}_T^2 + m^2)} = \frac{e^2 \mathcal{E}^2}{4\pi^3} e^{-\frac{\pi m^2}{e\mathcal{E}}}, \quad (428)$$

where the factor of 2 accounts for the two polarizations of the electron.

Then production of electron-positron pairs is exponentially suppressed and it is only sizeable for strong electric fields. To estimate its order of magnitude it is useful to restore the powers of  $c$  and  $\hbar$  in (428)

$$W = \frac{e^2 \mathcal{E}^2}{4\pi^3 c \hbar^2} e^{-\frac{\pi m^2 c^3}{\hbar e \mathcal{E}}} \quad (429)$$

The exponential suppression of the pair production disappears when the electric field reaches the critical value  $\mathcal{E}_{\text{crit}}$  at which the exponent is of order one

$$\mathcal{E}_{\text{crit}} = \frac{m^2 c^3}{\hbar e} \simeq 1.3 \times 10^{16} \text{ V cm}^{-1}. \quad (430)$$

This is indeed a very strong field which is extremely difficult to produce. A similar effect, however, takes place also in a time-varying electric field [45] and there is the hope that pair production could be observed in the presence of the alternating electric field produced by a laser.

The heuristic derivation that we followed here can be made more precise in QED. There the decay of the vacuum into electron-positron pairs can be computed from the imaginary part of the effective action  $\Gamma[A_\mu]$  in the presence of a classical gauge potential  $A_\mu$

$$\begin{aligned} i\Gamma[A_\mu] &\equiv \text{Diagram 1} + \text{Diagram 2} + \text{Diagram 3} + \dots \\ &= \log \det \left[ 1 - ieA \frac{1}{i\cancel{\partial} - m} \right]. \end{aligned} \quad (431)$$

This determinant can be computed using the standard heat kernel techniques. The probability of pair production is proportional to the imaginary part of  $i\Gamma[A_\mu]$  and gives

$$W = \frac{e^2 \mathcal{E}^2}{4\pi^3} \sum_{n=1}^{\infty} \frac{1}{n^2} e^{-n \frac{\pi m^2}{e\mathcal{E}}}. \quad (432)$$

Our simple argument based on tunneling in the Dirac sea gave only the leading term of Schwinger's result (432). The remaining terms can be also captured in the WKB approximation by taking into account the probability of production of several pairs, i.e. the tunneling of more than one electron through the barrier.

Here we have illustrated the creation of particles by semiclassical sources in quantum field theory using simple examples. Nevertheless, what we learned has important applications to the study of quantum fields in curved backgrounds. In quantum field theory in Minkowski space-time the vacuum state

is invariant under the Poincaré group and this, together with the covariance of the theory under Lorentz transformations, implies that all inertial observers agree on the number of particles contained in a quantum state. The breaking of such invariance, as happened in the case of coupling to a time-varying source analyzed above, implies that it is not possible anymore to define a state which would be recognized as the vacuum by all observers.

This is precisely the situation when fields are quantized on curved backgrounds. In particular, if the background is time-dependent (as it happens in a cosmological setup or for a collapsing star) different observers will identify different vacuum states. As a consequence what one observer call the vacuum will be full of particles for a different observer. This is precisely what is behind the phenomenon of Hawking radiation [46]. The emission of particles by a physical black hole formed from gravitational collapse of a star is the consequence of the fact that the vacuum state in the asymptotic past contain particles for an observer in the asymptotic future. As a consequence, a detector located far away from the black hole detects a stream of thermal radiation with temperature

$$T_{\text{Hawking}} = \frac{\hbar c^3}{8\pi G_N k M} \quad (433)$$

where  $M$  is the mass of the black hole,  $G_N$  is Newton's constant and  $k$  is Boltzmann's constant. There are several ways in which this results can be obtained. A more heuristic way is perhaps to think of this particle creation as resulting from quantum tunneling of particles across the potential barrier posed by gravity [47].

## 9.2 Supersymmetry

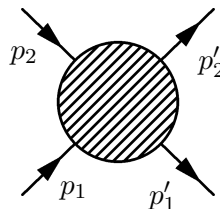
One of the things that we have learned in our journey around the landscape of quantum field theory is that our knowledge of the fundamental interactions in Nature is based on the idea of symmetry, and in particular gauge symmetry. The Lagrangian of the standard model can be written just including all possible renormalizable terms (i.e. with canonical dimension smaller or equal to 4) compatible with the gauge symmetry  $SU(3) \times SU(2) \times U(1)$  and Poincaré invariance. All attempts to go beyond start with the question of how to extend the symmetries of the standard model.

As explained in Section 5.1, in a quantum field theoretical description of the interaction of elementary particles the basic observable quantity to compute is the scattering or  $S$ -matrix giving the probability amplitude for the scattering of a number of incoming particles with a certain momentum into some final products

$$\mathcal{A}(\text{in} \longrightarrow \text{out}) = \langle \vec{p}'_1, \dots; \text{out} | \vec{p}_1, \dots; \text{in} \rangle. \quad (434)$$

An explicit symmetry of the theory has to be necessarily a symmetry of the  $S$ -matrix. Hence it is fair to ask what is the largest symmetry of the  $S$ -matrix.

Let us ask this question in the simple case of the scattering of two particles with four-momenta  $p_1$  and  $p_2$  in the  $t$ -channel



We will make the usual assumptions regarding positivity of the energy and analyticity. Invariance of the theory under the Poincaré group implies that the amplitude can only depend on the scattering angle  $\vartheta$  through

$$t = (p'_1 - p_1)^2 = 2(m_1^2 - p_1 \cdot p'_1) = 2(m_1^2 - E_1 E'_1 + |\vec{p}_1| |\vec{p}'_1| \cos \vartheta). \quad (435)$$

If there would be any extra bosonic symmetry of the theory it would restrict the scattering angle to a set of discrete values. In this case the  $S$ -matrix cannot be analytic since it would vanish everywhere except for the discrete values selected by the extra symmetry.

Actually, the only way to extend the symmetry of the theory without renouncing to the analyticity of the scattering amplitudes is to introduce “fermionic” symmetries, i.e. symmetries whose generators are anticommuting objects [48]. This means that in addition to the generators of the Poincaré group<sup>23</sup>  $P^\mu$ ,  $M^{\mu\nu}$  and the ones for the internal gauge symmetries  $G$ , we can introduce a number of fermionic generators  $Q_a^I$ ,  $\bar{Q}_{\dot{a}I}$  ( $I = 1, \dots, \mathcal{N}$ ), where  $\bar{Q}_{\dot{a}I} = (Q_a^I)^\dagger$ . The most general algebra that these generators satisfy is the  $\mathcal{N}$ -extended supersymmetry algebra [49]

$$\begin{aligned} \{Q_a^I, \bar{Q}_{\dot{b}J}\} &= 2\sigma_{ab}^\mu P_\mu \delta^I_J, \\ \{Q_a^I, Q_b^J\} &= 2\varepsilon_{ab} \mathcal{Z}^{IJ}, \end{aligned} \quad (436)$$

$$\{\bar{Q}_{\dot{a}I}, \bar{Q}_{\dot{b}J}\} = 2\varepsilon_{\dot{a}\dot{b}} \bar{\mathcal{Z}}^{IJ}, \quad (437)$$

where  $\mathcal{Z}^{IJ} \in \mathbb{C}$  commute with any other generator and satisfies  $\mathcal{Z}^{IJ} = -\mathcal{Z}^{JI}$ . Besides we have the commutators that determine the Poincaré transformations of the fermionic generators  $Q_a^I, \bar{Q}_{\dot{a}J}$

$$\begin{aligned} [Q_a^I, P^\mu] &= [\bar{Q}_{\dot{a}I}, P^\mu] = 0, \\ [Q_a^I, M^{\mu\nu}] &= \frac{1}{2}(\sigma^{\mu\nu})_a{}^b Q_b^I, \\ [\bar{Q}_{\dot{a}I}, M^{\mu\nu}] &= -\frac{1}{2}(\bar{\sigma}^{\mu\nu})_{\dot{a}}{}^{\dot{b}} \bar{Q}_{\dot{b}I}, \end{aligned} \quad (438)$$

where  $\sigma^{0i} = -i\sigma^i$ ,  $\sigma^{ij} = \varepsilon^{ijk}\sigma^k$  and  $\bar{\sigma}^{\mu\nu} = (\sigma^{\mu\nu})^\dagger$ . These identities simply mean that  $Q_a^I, \bar{Q}_{\dot{a}J}$  transform respectively in the  $(\frac{1}{2}, \mathbf{0})$  and  $(\mathbf{0}, \frac{1}{2})$  representations of the Lorentz group.

We know that the presence of a global symmetry in a theory implies that the spectrum can be classified in multiplets with respect to that symmetry. In the case of supersymmetry start with the case  $\mathcal{N} = 1$  in which there is a single pair of supercharges  $Q_a, \bar{Q}_{\dot{a}}$  satisfying the algebra

$$\{Q_a, \bar{Q}_{\dot{b}}\} = 2\sigma_{ab}^\mu P_\mu, \quad \{Q_a, Q_b\} = \{\bar{Q}_{\dot{a}}, \bar{Q}_{\dot{b}}\} = 0. \quad (439)$$

Notice that in the  $\mathcal{N} = 1$  case there is no possibility of having central charges.

We study now the representations of the supersymmetry algebra (439), starting with the massless case. Given a state  $|k\rangle$  satisfying  $k^2 = 0$ , we can always find a reference frame where the four-vector  $k^\mu$  takes the form  $k^\mu = (E, 0, 0, E)$ . Since the theory is Lorentz covariant we can obtain the representation of the supersymmetry algebra in this frame where the expressions are simpler. In particular, the right-hand side of the first anticommutator in Eq. (439) is given by

$$2\sigma_{ab}^\mu P_\mu = 2(P^0 - \sigma^3 P^3) = \begin{pmatrix} 0 & 0 \\ 0 & 4E \end{pmatrix}. \quad (440)$$

Therefore the algebra of supercharges in the massless case reduces to

$$\begin{aligned} \{Q_1, Q_1^\dagger\} &= \{Q_1, Q_2^\dagger\} = 0, \\ \{Q_2, Q_2^\dagger\} &= 4E. \end{aligned} \quad (441)$$

The commutator  $\{Q_1, Q_1^\dagger\} = 0$  implies that the action of  $Q_1$  on any state gives a zero-norm state of the Hilbert space  $\|Q_1|\Psi\rangle\| = 0$ . If we want the theory to preserve unitarity we must eliminate these null

<sup>23</sup>The generators  $M^{\mu\nu}$  are related with the ones for boost and rotations introduced in section 4.1 by  $J^i \equiv M^{0i}$ ,  $M^i = \frac{1}{2}\varepsilon^{ijk}M^{jk}$ . In this section we also use the “dotted spinor” notation, in which spinors in the  $(\frac{1}{2}, \mathbf{0})$  and  $(\mathbf{0}, \frac{1}{2})$  representations of the Lorentz group are indicated respectively by undotted ( $a, b, \dots$ ) and dotted ( $\dot{a}, \dot{b}, \dots$ ) indices.

states from the spectrum. This is equivalent to setting  $Q_1 \equiv 0$ . On the other hand, in terms of the second generator  $Q_2$  we can define the operators

$$a = \frac{1}{2\sqrt{E}}Q_2, \quad a^\dagger = \frac{1}{2\sqrt{E}}Q_2^\dagger, \quad (442)$$

which satisfy the algebra of a pair of fermionic creation-annihilation operators,  $\{a, a^\dagger\} = 1$ ,  $a^2 = (a^\dagger)^2 = 0$ . Starting with a vacuum state  $a|\lambda\rangle = 0$  with helicity  $\lambda$  we can build the massless multiplet

$$|\lambda\rangle, \quad |\lambda + \frac{1}{2}\rangle \equiv a^\dagger|\lambda\rangle. \quad (443)$$

Here we consider two important cases:

- Scalar multiplet: we take the vacuum state to have zero helicity  $|0^+\rangle$  so the multiplet consists of a scalar and a helicity- $\frac{1}{2}$  state

$$|0^+\rangle, \quad |\frac{1}{2}\rangle \equiv a^\dagger|0^+\rangle. \quad (444)$$

However, this multiplet is not invariant under the CPT transformation which reverses the sign of the helicity of the states. In order to have a CPT-invariant theory we have to add to this multiplet its CPT-conjugate which can be obtained from a vacuum state with helicity  $\lambda = -\frac{1}{2}$

$$|0^-\rangle, \quad |-\frac{1}{2}\rangle. \quad (445)$$

Putting them together we can combine the two zero helicity states with the two fermionic ones into the degrees of freedom of a complex scalar field and a Weyl (or Majorana) spinor.

- Vector multiplet: now we take the vacuum state to have helicity  $\lambda = \frac{1}{2}$ , so the multiplet contains also a massless state with helicity  $\lambda = 1$

$$|\frac{1}{2}\rangle, \quad |1\rangle \equiv a^\dagger|\frac{1}{2}\rangle. \quad (446)$$

As with the scalar multiplet we add the CPT conjugated obtained from a vacuum state with helicity  $\lambda = -1$

$$|-\frac{1}{2}\rangle, \quad |-1\rangle, \quad (447)$$

which together with (446) give the propagating states of a gauge field and a spin- $\frac{1}{2}$  gaugino.

In both cases we see the trademark of supersymmetric theories: the number of bosonic and fermionic states within a multiplet are the same.

In the case of extended supersymmetry we have to repeat the previous analysis for each supersymmetry charge. At the end, we have  $\mathcal{N}$  sets of fermionic creation-annihilation operators  $\{a^I, a_I^\dagger\} = \delta^I_J$ ,  $(a_I)^2 = (a_I^\dagger)^2 = 0$ . Let us work out the case of  $\mathcal{N} = 8$  supersymmetry. Since for several reasons we do not want to have states with helicity larger than 2, we start with a vacuum state  $|-2\rangle$  of helicity  $\lambda = -2$ . The rest of the states of the supermultiplet are obtained by applying the eight different creation operators  $a_I^\dagger$  to the vacuum:

$$\begin{aligned} \lambda = 2 : & \quad a_1^\dagger \dots a_8^\dagger |-2\rangle & \binom{8}{8} = 1 \text{ state,} \\ \lambda = \frac{3}{2} : & \quad a_{I_1}^\dagger \dots a_{I_7}^\dagger |-2\rangle & \binom{8}{7} = 8 \text{ states,} \\ \lambda = 1 : & \quad a_{I_1}^\dagger \dots a_{I_6}^\dagger |-2\rangle & \binom{8}{6} = 28 \text{ states,} \end{aligned}$$

$$\begin{aligned}
 \lambda = \frac{1}{2} : & \quad a_{I_1}^\dagger \dots a_{I_5}^\dagger | - 2 \rangle & \quad \binom{8}{5} = 56 \text{ states,} \\
 \lambda = 0 : & \quad a_{I_1}^\dagger \dots a_{I_4}^\dagger | - 2 \rangle & \quad \binom{8}{4} = 70 \text{ states,} \\
 \lambda = -\frac{1}{2} : & \quad a_{I_1}^\dagger a_{I_2}^\dagger a_{I_3}^\dagger | - 2 \rangle & \quad \binom{8}{3} = 56 \text{ states,} \\
 \lambda = -1 : & \quad a_{I_1}^\dagger a_{I_2}^\dagger | - 2 \rangle & \quad \binom{8}{2} = 28 \text{ states,} \\
 \lambda = -\frac{3}{2} : & \quad a_{I_1}^\dagger | - 2 \rangle & \quad \binom{8}{1} = 8 \text{ states,} \\
 \lambda = -2 : & \quad | - 2 \rangle & \quad 1 \text{ state.}
 \end{aligned} \tag{448}$$

Putting together the states with opposite helicity we find that the theory contains:

- 1 spin-2 field  $g_{\mu\nu}$  (a graviton),
- 8 spin- $\frac{3}{2}$  gravitino fields  $\psi_\mu^I$ ,
- 28 gauge fields  $A_\mu^{[IJ]}$ ,
- 56 spin- $\frac{1}{2}$  fermions  $\psi^{[IJK]}$ ,
- 70 scalars  $\phi^{[IJKL]}$ ,

where by  $[IJ\dots]$  we have denoted that the indices are antisymmetrized. We see that, unlike the massless multiplets of  $\mathcal{N} = 1$  supersymmetry studied above, this multiplet is CPT invariant by itself. As in the case of the massless  $\mathcal{N} = 1$  multiplet, here we also find as many bosonic as fermionic states:

$$\begin{aligned}
 \text{bosons:} & \quad 1 + 28 + 70 + 28 + 1 = 128 \quad \text{states,} \\
 \text{fermions:} & \quad 8 + 56 + 56 + 8 = 128 \quad \text{states.}
 \end{aligned}$$

Now we study briefly the case of massive representations  $|k\rangle$ ,  $k^2 = M^2$ . Things become simpler if we work in the rest frame where  $P^0 = M$  and the spatial components of the momentum vanish. Then, the supersymmetry algebra becomes:

$$\{Q_a^I, \bar{Q}_{bJ}\} = 2M\delta_{ab}\delta^I_J. \tag{449}$$

We proceed now in a similar way to the massless case by defining the operators

$$a_a^I \equiv \frac{1}{\sqrt{2M}} Q_a^I, \quad a_{\dot{a}I}^\dagger \equiv \frac{1}{\sqrt{2M}} \bar{Q}_{\dot{a}I}. \tag{450}$$

The multiplets are found by choosing a vacuum state with a definite spin. For example, for  $\mathcal{N} = 1$  and taking a spin-0 vacuum  $|0\rangle$  we find three states in the multiplet transforming irreducibly with respect to the Lorentz group:

$$|0\rangle, \quad a_{\dot{a}}^\dagger |0\rangle, \quad \varepsilon^{\dot{a}\dot{b}} a_{\dot{a}}^\dagger a_{\dot{b}}^\dagger |0\rangle, \tag{451}$$

which, once transformed back from the rest frame, correspond to the physical states of two spin-0 bosons and one spin- $\frac{1}{2}$  fermion. For  $\mathcal{N}$ -extended supersymmetry the corresponding multiplets can be worked out in a similar way.

The equality between bosonic and fermionic degrees of freedom is at the root of many of the interesting properties of supersymmetric theories. For example, in section 4 we computed the divergent vacuum energy contributions for each real bosonic or fermionic propagating degree of freedom is<sup>24</sup>

$$E_{\text{vac}} = \pm \frac{1}{2} \delta(\vec{0}) \int d^3p \omega_p, \tag{452}$$

<sup>24</sup>For a boson, this can be read off Eq. (80). In the case of fermions, the result of Eq. (134) gives the vacuum energy contribution of the four real propagating degrees of freedom of a Dirac spinor.

where the sign  $\pm$  corresponds respectively to bosons and fermions. Hence, for a supersymmetric theory the vacuum energy contribution exactly cancels between bosons and fermions. This boson-fermion degeneracy is also responsible for supersymmetric quantum field theories being less divergent than non-supersymmetric ones.

### Appendix: A crash course in Group Theory

In this Appendix we summarize some basic facts about Group Theory. Given a group  $G$  a representation of  $G$  is a correspondence between the elements of  $G$  and the set of linear operators acting on a vector space  $V$ , such that for each element of the group  $g \in G$  there is a linear operator  $D(g)$

$$D(g) : V \longrightarrow V \tag{453}$$

satisfying the group operations

$$D(g_1)D(g_2) = D(g_1g_2), \quad D(g_1^{-1}) = D(g_1)^{-1}, \quad g_1, g_2 \in \mathcal{G}. \tag{454}$$

The representation  $D(g)$  is irreducible if and only if the only operators  $A : V \rightarrow V$  commuting with all the elements of the representation  $D(g)$  are the ones proportional to the identity

$$[D(g), A] = 0, \quad \forall g \iff A = \lambda \mathbf{1}, \quad \lambda \in \mathbb{C} \tag{455}$$

More intuitively, we can say that a representation is irreducible if there is no proper subspace  $U \subset V$  (i.e.  $U \neq V$  and  $U \neq \emptyset$ ) such that  $D(g)U \subset U$  for every element  $g \in G$ .

Here we are specially interested in Lie groups whose elements are labelled by a number of continuous parameters. In mathematical terms this means that a Lie group is a manifold  $\mathcal{M}$  together with an operation  $\mathcal{M} \times \mathcal{M} \rightarrow \mathcal{M}$  that we will call multiplication that satisfies the associativity property  $g_1 \cdot (g_2 \cdot g_3) = (g_1 \cdot g_2) \cdot g_3$  together with the existence of unity  $g\mathbf{1} = \mathbf{1}g = g$ , for every  $g \in \mathcal{M}$  and inverse  $gg^{-1} = g^{-1}g = \mathbf{1}$ .

The simplest example of a Lie group is  $SO(2)$ , the group of rotations in the plane. Each element  $R(\theta)$  is labelled by the rotation angle  $\theta$ , with the multiplication acting as  $R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2)$ . Because the angle  $\theta$  is defined only modulo  $2\pi$ , the manifold of  $SO(2)$  is a circumference  $S^1$ .

One of the interesting properties of Lie groups is that in a neighborhood of the identity element they can be expressed in terms of a set of generators  $T^a$  ( $a = 1, \dots, \dim G$ ) as

$$D(g) = \exp(-i\alpha_a T^a) \equiv \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \alpha_{a_1} \dots \alpha_{a_n} T^{a_1} \dots T^{a_n}, \tag{456}$$

where  $\alpha_a \in \mathbb{C}$  are a set of coordinates of  $\mathcal{M}$  in a neighborhood of  $\mathbf{1}$ . Because of the general Baker-Campbell-Hausdorff formula, the multiplication of two group elements is encoded in the value of the commutator of two generators, that in general has the form

$$[T^a, T^b] = if^{abc}T^c, \tag{457}$$

where  $f^{abc} \in \mathbb{C}$  are called the structure constants. The set of generators with the commutator operation form the Lie algebra associated with the Lie group. Hence, given a representation of the Lie algebra of generators we can construct a representation of the group by exponentiation (at least locally near the identity).

We illustrate these concept with some particular examples. For  $SU(2)$  each group element is labelled by three real number  $\alpha_i$ ,  $i = 1, 2, 3$ . We have two basic representations: one is the fundamental representation (or spin  $\frac{1}{2}$ ) defined by

$$D_{\frac{1}{2}}(\alpha_i) = e^{-\frac{i}{2}\alpha_i\sigma^i}, \tag{458}$$

with  $\sigma^i$  the Pauli matrices. The second one is the adjoint (or spin 1) representation which can be written as

$$D_1(\alpha_i) = e^{-i\alpha_i J^i}, \quad (459)$$

where

$$J^1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \quad J^2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad J^3 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (460)$$

Actually,  $J^i$  ( $i = 1, 2, 3$ ) generate rotations around the  $x$ ,  $y$  and  $z$  axis respectively. Representations of spin  $j \in \mathbb{N} + \frac{1}{2}$  can be also constructed with dimension

$$\dim D_j(g) = 2j + 1. \quad (461)$$

As a second example we consider  $SU(3)$ . This group has two basic three-dimensional representations denoted by  $\mathbf{3}$  and  $\bar{\mathbf{3}}$  which in QCD are associated with the transformation of quarks and antiquarks under the color gauge symmetry  $SU(3)$ . The elements of these representations can be written as

$$D_{\mathbf{3}}(\alpha^a) = e^{\frac{i}{2}\alpha^a \lambda_a}, \quad D_{\bar{\mathbf{3}}}(\alpha^a) = e^{-\frac{i}{2}\alpha^a \lambda_a^T} \quad (a = 1, \dots, 8), \quad (462)$$

where  $\lambda_a$  are the eight hermitian Gell-Mann matrices

$$\begin{aligned} \lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ \lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, & \lambda_5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, & \lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \\ \lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, & \lambda_8 &= \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & -\frac{2}{\sqrt{3}} \end{pmatrix}. \end{aligned} \quad (463)$$

Hence the generators of the representations  $\mathbf{3}$  and  $\bar{\mathbf{3}}$  are given by

$$T^a(\mathbf{3}) = \frac{1}{2}\lambda_a, \quad T^a(\bar{\mathbf{3}}) = -\frac{1}{2}\lambda_a^T. \quad (464)$$

Irreducible representations can be classified in three groups: real, complex and pseudoreal.

- Real representations: a representation is said to be real if there is a *symmetric matrix*  $S$  which acts as intertwiner between the generators and their complex conjugates

$$\bar{T}^a = -S T^a S^{-1}, \quad S^T = S. \quad (465)$$

This is for example the case of the adjoint representation of  $SU(2)$  generated by the matrices (460)

- Pseudoreal representations: are the ones for which an *antisymmetric matrix*  $S$  exists with the property

$$\bar{T}^a = -S T^a S^{-1}, \quad S^T = -S. \quad (466)$$

As an example we can mention the spin- $\frac{1}{2}$  representation of  $SU(2)$  generated by  $\frac{1}{2}\sigma^i$ .



- Complex representations: finally, a representation is complex if the generators and their complex conjugate are not related by a similarity transformation. This is for instance the case of the two three-dimensional representations  $\mathbf{3}$  and  $\bar{\mathbf{3}}$  of  $SU(3)$ .

There are a number of invariants that can be constructed associated with an irreducible representation  $R$  of a Lie group  $G$  and that can be used to label such a representation. If  $T_R^a$  are the generators in a certain representation  $R$  of the Lie algebra, it is easy to see that the matrix  $\sum_{a=1}^{\dim G} T_R^a T_R^a$  commutes with every generator  $T_R^a$ . Therefore, because of Schur's lemma, it has to be proportional to the identity<sup>25</sup>. This defines the Casimir invariant  $C_2(R)$  as

$$\sum_{a=1}^{\dim G} T_R^a T_R^a = C_2(R) \mathbf{1}. \quad (467)$$

A second invariant  $T_2(R)$  associated with a representation  $R$  can also be defined by the identity

$$\text{Tr } T_R^a T_R^b = T_2(R) \delta^{ab}. \quad (468)$$

Actually, taking the trace in Eq. (467) and combining the result with (468) we find that both invariants are related by the identity

$$C_2(R) \dim R = T_2(R) \dim G, \quad (469)$$

with  $\dim R$  the dimension of the representation  $R$ .

These two invariants appear frequently in quantum field theory calculations with nonabelian gauge fields. For example  $T_2(R)$  comes about as the coefficient of the one-loop calculation of the beta-function for a Yang-Mills theory with gauge group  $G$ . In the case of  $SU(N)$ , for the fundamental representation, we find the values

$$C_2(\text{fund}) = \frac{N^2 - 1}{2N}, \quad T_2(\text{fund}) = \frac{1}{2}, \quad (470)$$

whereas for the adjoint representation the results are

$$C_2(\text{adj}) = N, \quad T_2(\text{adj}) = N. \quad (471)$$

A third invariant  $A(R)$  is specially important in the calculation of anomalies. As discussed in section (7), the chiral anomaly in gauge theories is proportional to the group-theoretical factor  $\text{Tr} [T_R^a \{T_R^b, T_R^c\}]$ . This leads us to define  $A(R)$  as

$$\text{Tr} [T_R^a \{T_R^b, T_R^c\}] = A(R) d^{abc}, \quad (472)$$

where  $d^{abc}$  is symmetric in its three indices and does not depend on the representation. Therefore, the cancellation of anomalies in a gauge theory with fermions transformed in the representation  $R$  of the gauge group is guaranteed if the corresponding invariant  $A(R)$  vanishes.

It is not difficult to prove that  $A(R) = 0$  if the representation  $R$  is either real or pseudoreal. Indeed, if this is the case, then there is a matrix  $S$  (symmetric or antisymmetric) that intertwines the generators  $T_R^a$  and their complex conjugates  $\bar{T}_R^a = -S T_R^a S^{-1}$ . Then, using the hermiticity of the generators we can write

$$\text{Tr} [T_R^a \{T_R^b, T_R^c\}] = \text{Tr} [T_R^a \{T_R^b, T_R^c\}]^T = \text{Tr} [\bar{T}_R^a \{\bar{T}_R^b, \bar{T}_R^c\}]. \quad (473)$$

<sup>25</sup>Schur's lemma states that if there is a matrix  $A$  that commutes with all elements of an irreducible representation of a Lie algebra, then  $A = \lambda \mathbf{1}$ , for some  $\lambda \in \mathbb{C}$ .

Now, using (465) or (466) we have

$$\mathrm{Tr} \left[ \overline{T}_R^a \{ \overline{T}_R^b, \overline{T}_R^c \} \right] = -\mathrm{Tr} \left[ ST_R^a S^{-1} \{ ST_R^b S^{-1}, ST_R^c S^{-1} \} \right] = -\mathrm{Tr} \left[ T_R^a \{ T_R^b, T_R^c \} \right], \quad (474)$$

which proves that  $\mathrm{Tr} \left[ T_R^a \{ T_R^b, T_R^c \} \right]$  and therefore  $A(R) = 0$  whenever the representation is real or pseudoreal. Since the gauge anomaly in four dimensions is proportional to  $A(R)$  this means that anomalies appear only when the fermions transform in a complex representation of the gauge group.

## References

- [1] L. Álvarez-Gaumé and M. A. Vázquez-Mozo, *An Invitation to Quantum Field Theory*, Springer 2011.
- [2] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields*, McGraw-Hill 1965.
- [3] C. Itzykson and J.-B. Zuber, *Quantum Field Theory*, McGraw-Hill 1980.
- [4] P. Ramond, *Field Theory: A Modern Primer*, Addison-Wesley 1990.
- [5] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Addison Wesley 1995.
- [6] S. Weinberg, *The Quantum Theory of Fields*, Vols. 1-3, Cambridge 1995
- [7] P. Deligne et al. (editors), *Quantum Fields and Strings: a Course for Mathematicians*, American Mathematical Society 1999.
- [8] A. Zee, *Quantum Field Theory in a Nutshell*, Princeton 2003.
- [9] B. S. DeWitt, *The Global Approach to Quantum Field Theory*, Vols. 1 & 2, Oxford 2003.
- [10] V. P. Nair, *Quantum Field Theory. A Modern Perspective*, Springer 2005.
- [11] T. Banks, *Modern Quantum Field Theory*, Cambridge 2008.
- [12] O. Klein, *Die Reflexion von Elektronen an einem Potentialsprung nach der Relativischen Dynamik von Dirac*, Z. Phys. **53** (1929) 157.
- [13] B. R. Holstein, *Klein's paradox*, Am. J. Phys. **66** (1998) 507.
- [14] N. Dombey and A. Calogeracos, *Seventy years of the Klein paradox*, Phys. Rept. **315** (1999) 41.  
N. Dombey and A. Calogeracos, *History and Physics of the Klein Paradox*, Contemp. Phys. **40** (1999) 313 (quant-ph/9905076).
- [15] F. Sauter, *Zum Kleinschen Paradoxon*, Z. Phys. **73** (1932) 547.
- [16] H. B. G. Casimir, *On the attraction between two perfectly conducting plates*, Proc. Kon. Ned. Akad. Wet. **60** (1948) 793.
- [17] G. Plunien, B. Müller and W. Greiner, *The Casimir Effect*, Phys. Rept. **134** (1986) 87.  
K. A. Milton, *The Casimir Effect: Physical Manifestation of Zero-Point Energy*, (hep-th/9901011).  
K. A. Milton, *The Casimir effect: recent controversies and progress*, J. Phys. **A37** (2004) R209 (hep-th/0406024).  
S. K. Lamoreaux, *The Casimir force: background, experiments, and applications*, Rep. Prog. Phys. **68** (2005) 201.
- [18] M. J. Sparnaay, *Measurement of attractive forces between flat plates*, Physica **24** (1958) 751.
- [19] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover 1972.
- [20] Y. Aharonov and D. Bohm, *Significance of the electromagnetic potentials in the quantum theory*, Phys. Rev. **115** (1955) 485.
- [21] P. A. M. Dirac, *Quantised Singularities in the Electromagnetic Field*, Proc. Roy. Soc. **133** (1931) 60.
- [22] S. Dodelson, *Modern Cosmology*, Academic Press 2003.
- [23] P. A. M. Dirac, *Lectures on Quantum Mechanics*, Dover 2001.

- [24] M. Henneaux and C. Teitelboim, *Quantization of Gauge Systems*, Princeton 1992.
- [25] R. Jackiw, *Quantum meaning of classical field theory*, Rev. Mod. Phys. **49** (1977) 681  
 R. Jackiw, *Introduction to the Yang-Mills quantum theory*, Rev. Mod. Phys. **52** (1980) 661.
- [26] P. Ramond, *Journeys Beyond the Standard Model*, Perseus Books 1999.  
 R. N. Mohapatra, *Unification and Supersymmetry. The Frontiers of Quark-Lepton Physics*, Springer 2003.
- [27] C. P. Burgess, *Goldstone and pseudogoldstone bosons in nuclear, particle and condensed matter physics*, Phys. Rept. **330** (2000) 193 (hep-th/9808176).
- [28] L. Álvarez-Gaumé, *An introduction to anomalies*, in: “Fundamental problems of gauge field theory”, eds. G. Velo and A. S. Wightman, Plenum Press 1986.  
 R. A. Bertlmann, *Anomalies in Quantum Field Theory*, Oxford 1996.  
 K. Fujikawa and H. Suzuki, *Path Integrals and Quantum Anomalies*, Oxford 2004.  
 J. A. Harvey, *TASI lectures on anomalies*, hep-th/0509097.  
 L. Álvarez-Gaumé and M. A. Vázquez-Mozo, *Introduction to Anomalies*, Springer (to appear).
- [29] R. Jackiw, *Topological investigations of quantized gauge theories*, in: “Current Algebra and Anomalies”, eds. S. B. Treiman, R. Jackiw, B. Zumino and E. Witten, Princeton 1985.
- [30] S. Adler, *Axial-Vector Vertex in Spinor Electrodynamics*, Phys. Rev. **177** (1969) 2426.  
 J. S. Bell and R. Jackiw, *A PCAC puzzle:  $\pi^0 \rightarrow 2\gamma$  in the sigma model*, Nuovo Cimento **A60** (1969) 47.
- [31] J. Steinberger, *On the Use of Subtraction Fields and the Lifetimes of Some Types of Meson Decay*, Phys. Rev. **76** (1949) 1180.
- [32] F. J. Ynduráin, *The Theory of Quark and Gluon Interactions*, Springer 1999.
- [33] G. 't Hooft, *How the instantons solve the U(1) problem*, Phys. Rept. **142** (1986) 357.
- [34] D. G. Sutherland, *Current Algebra and Some Nonstrong Mesonic Decays*, Nucl. Phys. **B2** (1967) 433.  
 M. J. G. Veltman, *Theoretical aspects of high-energy neutrino interactions*, Proc. R. Soc. **A301** (1967) 107.
- [35] S. L. Adler and W. A. Bardeen, *Absence of higher order corrections in the anomalous axial vector divergence equation*, Phys. Rev. **182** (1969) 1517.
- [36] E. Witten, *An SU(2) anomaly*, Phys. Lett. **B117** (1982) 324.
- [37] S. Eidelman et al. *Review of Particle Physics*, Phys. Lett. **B592** (2004) 1 (<http://pdg.lbl.gov>).
- [38] D. J. Gross and F. Wilczek, *Ultraviolet behavior of nonabelian gauge theories*, Phys. Rev. Lett. **30** (1973) 1343.
- [39] H. D. Politzer, *Reliable perturbative results for strong interactions?*, Phys. Rev. Lett. **30** (1973) 1346.
- [40] G. 't Hooft, remarks at the *Colloquium on Renormalization of Yang-Mills fields and applications to particle physics*, Marseille 1972.
- [41] I. B. Khriplovich, *Green's functions in theories with a non-abelian gauge group*, Yad. Fiz. **10** (1969) 409 [Sov. J. Nucl. Phys. **10** (1970) 235].  
 M. V. Terentiev and V. S. Vanyashin, *The vacuum polarization of a charged vector field*, Zh. Eksp. Teor. Fiz. **48** (1965) 565 [Sov. Phys. JETP **21** (1965) 375].
- [42] K. G. Wilson, *Renormalization group and critical phenomena 1. Renormalization group and the Kadanoff scaling picture*, Phys. Rev. **B4** (1971) 3174.  
 K. G. Wilson, *Renormalization group and critical phenomena 2. Phase space cell analysis of critical behavior*, Phys. Rev. **B4** (1971) 3184  
 K. G. Wilson, *The renormalization group and critical phenomena*, Rev. Mod. Phys. **55** (1983) 583.
- [43] L. P. Kadanoff, *Scaling Laws for Ising Models Near  $T_c$* , Physics **2** (1966) 263.

- [44] J. Schwinger, *On Gauge Invariance and Vacuum Polarization*, Phys. Rev. **82** (1951) 664.
- [45] E. Brezin and C. Itzykson, *Pair Production in Vacuum by an Alternating Field*, Phys. Rev. **D2** (1970) 1191.
- [46] S. W. Hawking, *Particle Creation by Black Holes*, Commun. Math. Phys. **43** (1975) 199.
- [47] M. K. Parikh and F. Wilczek, *Hawking Radiation as Tunneling*, Phys. Rev. Lett. **85** (2000) 5042 (hep-th/9907001)
- [48] Yu. A. Golfand and E. P. Likhtman, *Extension of the Algebra of Poincaré group generators and violations of P-invariance*, JETP Lett. **13** (1971) 323.  
D. V. Volkov and V. P. Akulov, *Is the Neutrino a Goldstone Particle*, Phys. Lett. **B46** (1973) 109.  
J. Wess and B. Zumino, *A Lagrangian Model Invariant under Supergauge Transformations*, Phys. Lett. **B49** (1974) 52.
- [49] R. Haag, J. Łopuszański and M. Sohnius, *All possible generators of supersymmetries of the S-matrix*, Nucl. Phys. **B88** (1975) 257.

# Flavour Dynamics and Violations of the CP Symmetry

A. Pich

IFIC, University of Valencia – CSIC, E-46071 Valencia, Spain

## Abstract

An overview of flavour physics and  $\mathcal{CP}$ -violating phenomena is presented. The Standard Model quark-mixing mechanism is discussed in detail and its many successful experimental tests are summarized. Flavour-changing transitions put very stringent constraints on new-physics scenarios beyond the Standard Model framework. Special attention is given to the empirical evidences of  $\mathcal{CP}$  violation and their important role in our understanding of flavour dynamics. The current status of the so-called flavour anomalies is also reviewed.

## Keywords

Flavour physics; quark mixing;  $\mathcal{CP}$  violation; electroweak interactions.

## 1 Fermion families

We have learnt experimentally that there are six different quark flavours; three of them,  $u$ ,  $c$ ,  $t$ , with electric charge  $Q = +\frac{2}{3}$  (up-type), and the other three,  $d$ ,  $s$ ,  $b$ , with  $Q = -\frac{1}{3}$  (down-type). There are also three different charged leptons,  $e$ ,  $\mu$ ,  $\tau$ , with  $Q = -1$  and their corresponding neutrinos,  $\nu_e$ ,  $\nu_\mu$ ,  $\nu_\tau$ , with  $Q = 0$ . We can include all these particles into the  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$  Standard Model (SM) framework [1–3], by organizing them into three families of quarks and leptons:

$$\begin{bmatrix} \nu_e & u \\ e^- & d' \end{bmatrix}, \quad \begin{bmatrix} \nu_\mu & c \\ \mu^- & s' \end{bmatrix}, \quad \begin{bmatrix} \nu_\tau & t \\ \tau^- & b' \end{bmatrix}, \quad (1)$$

where (each quark appears in three different colours)

$$\begin{bmatrix} \nu_i & u_i \\ \ell_i^- & d'_i \end{bmatrix} \equiv \begin{pmatrix} \nu_i \\ \ell_i^- \end{pmatrix}_L, \quad \begin{pmatrix} u_i \\ d'_i \end{pmatrix}_L, \quad \ell_{iR}^-, \quad u_{iR}, \quad d'_{iR}, \quad (2)$$

plus the corresponding antiparticles. Thus, the left-handed fields are  $SU(2)_L$  doublets, while their right-handed partners transform as  $SU(2)_L$  singlets. The three fermionic families appear to have identical properties (gauge interactions); they differ only by their mass and their flavour quantum numbers.

The fermionic couplings of the photon and the  $Z$  boson are flavour conserving, i.e., the neutral gauge bosons couple to a fermion and its corresponding antifermion. In contrast, the  $W^\pm$  bosons couple any up-type quark with all down-type quarks because the weak doublet partner of  $u_i$  turns out to be a quantum superposition of down-type mass eigenstates:  $d'_i = \sum_j \mathbf{V}_{ij}^- d_j$ . This flavour mixing generates a rich variety of observable phenomena, including  $\mathcal{CP}$ -violation effects, which can be described in a very successful way within the SM [4, 5].

In spite of its enormous phenomenological success, The SM does not provide any real understanding of flavour. We do not know yet why fermions are replicated in three (and only three) nearly identical copies. Why the pattern of masses and mixings is what it is? Are the masses the only difference among the three families? What is the origin of the SM flavour structure? Which dynamics is responsible for the observed  $\mathcal{CP}$  violation? The fermionic flavour is the main source of arbitrary free parameters in the SM: 9 fermion masses, 3 mixing angles and 1 complex phase, for massless neutrinos. Another 7 (9) additional parameters arise with non-zero Dirac (Majorana) neutrino masses: 3 masses, 3 mixing angles and 1 (3) phases. The problem of fermion mass generation is deeply related with the mechanism responsible for the electroweak Spontaneous Symmetry Breaking (SSB). Thus, the origin of these parameters lies in the



**Fig. 1:** Flavour-changing transitions through the charged-current couplings of the  $W^\pm$  bosons.

most obscure part of the SM Lagrangian: the scalar sector. Clearly, the dynamics of flavour appears to be “terra incognita” which deserves a careful investigation.

The following sections contain a short overview of the quark flavour sector and its present phenomenological status. The most relevant experimental tests are briefly described. A more pedagogic introduction to the SM can be found in Ref. [4].

## 2 Flavour structure of the Standard Model

In the SM flavour-changing transitions occur only in the charged-current sector (Fig. 1):

$$\mathcal{L}_{\text{CC}} = -\frac{g}{2\sqrt{2}} \left\{ W_\mu^\dagger \left[ \sum_{ij} \bar{u}_i \gamma^\mu (1 - \gamma_5) \mathbf{V}_{ij} d_j + \sum_\ell \bar{\nu}_\ell \gamma^\mu (1 - \gamma_5) \ell \right] + \text{h.c.} \right\}. \quad (3)$$

The so-called Cabibbo–Kobayashi–Maskawa (CKM) matrix  $\mathbf{V}$  [6, 7] is generated by the same Yukawa couplings giving rise to the quark masses. Before SSB, there is no mixing among the different quarks, i.e.,  $\mathbf{V} = \mathbf{I}$ . In order to understand the origin of the matrix  $\mathbf{V}$ , let us consider the general case of  $N_G$  generations of fermions, and denote  $\nu'_j, \ell'_j, u'_j, d'_j$  the members of the weak family  $j$  ( $j = 1, \dots, N_G$ ), with definite transformation properties under the gauge group. Owing to the fermion replication, a large variety of fermion-scalar couplings are allowed by the gauge symmetry. The most general Yukawa Lagrangian has the form

$$\begin{aligned} \mathcal{L}_Y = & - \sum_{jk} \left\{ (\bar{u}'_j, \bar{d}'_j)_L \left[ c_{jk}^{(d)} \begin{pmatrix} \phi^{(+)} \\ \phi^{(0)} \end{pmatrix} d'_{kR} + c_{jk}^{(u)} \begin{pmatrix} \phi^{(0)*} \\ -\phi^{(-)} \end{pmatrix} u'_{kR} \right] \right. \\ & \left. + (\bar{\nu}'_j, \bar{\ell}'_j)_L c_{jk}^{(\ell)} \begin{pmatrix} \phi^{(+)} \\ \phi^{(0)} \end{pmatrix} \ell'_{kR} \right\} + \text{h.c.}, \end{aligned} \quad (4)$$

where  $\phi^T(x) \equiv (\phi^{(+)}, \phi^{(0)})$  is the SM scalar doublet and  $c_{jk}^{(d)}$ ,  $c_{jk}^{(u)}$  and  $c_{jk}^{(\ell)}$  are arbitrary coupling constants. The second term involves the charge-conjugate scalar field  $\phi^c(x) = i\sigma_2 \phi^*(x)$ .

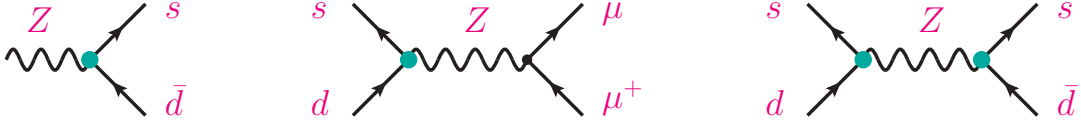
In the unitary gauge  $\phi^T(x) \equiv \frac{1}{\sqrt{2}} (0, v + H(x))$ , where  $v$  is the electroweak vacuum expectation value and  $H(x)$  the Higgs field. The Yukawa Lagrangian can then be written as

$$\mathcal{L}_Y = - \left( 1 + \frac{H}{v} \right) \{ \bar{\mathbf{d}}'_L \mathbf{M}'_d \mathbf{d}'_R + \bar{\mathbf{u}}'_L \mathbf{M}'_u \mathbf{u}'_R + \bar{\mathbf{\ell}}'_L \mathbf{M}'_\ell \mathbf{\ell}'_R + \text{h.c.} \}. \quad (5)$$

Here,  $\mathbf{d}'$ ,  $\mathbf{u}'$  and  $\mathbf{\ell}'$  denote vectors in the  $N_G$ -dimensional flavour space, with components  $d'_j$ ,  $u'_j$  and  $\ell'_j$ , respectively, and the corresponding mass matrices are given by

$$(\mathbf{M}'_d)_{ij} \equiv c_{ij}^{(d)} \frac{v}{\sqrt{2}}, \quad (\mathbf{M}'_u)_{ij} \equiv c_{ij}^{(u)} \frac{v}{\sqrt{2}}, \quad (\mathbf{M}'_\ell)_{ij} \equiv c_{ij}^{(\ell)} \frac{v}{\sqrt{2}}. \quad (6)$$

The diagonalization of these mass matrices determines the mass eigenstates  $d_j$ ,  $u_j$  and  $\ell_j$ , which are linear combinations of the corresponding weak eigenstates  $d'_j$ ,  $u'_j$  and  $\ell'_j$ , respectively.



**Fig. 2:** Tree-level FCNC couplings (green solid vertices) are absent in the SM. Therefore, very suppressed (experimentally) transitions such as  $K^0 \rightarrow \mu^+ \mu^-$  or  $K^0 - \bar{K}^0$  mixing cannot occur through tree-level exchange.

The matrix  $\mathbf{M}'_d$  can be decomposed as<sup>1</sup>  $\mathbf{M}'_d = \mathbf{H}_d \mathbf{U}_d = \mathbf{S}_d^\dagger \mathcal{M}_d \mathbf{S}_d \mathbf{U}_d$ , where  $\mathbf{H}_d \equiv \sqrt{\mathbf{M}'_d \mathbf{M}'_d{}^\dagger}$  is an Hermitian positive-definite matrix, while  $\mathbf{U}_d$  is unitary.  $\mathbf{H}_d$  can be diagonalized by a unitary matrix  $\mathbf{S}_d$ ; the resulting matrix  $\mathcal{M}_d$  is diagonal, Hermitian and positive definite. Similarly, one has  $\mathbf{M}'_u = \mathbf{H}_u \mathbf{U}_u = \mathbf{S}_u^\dagger \mathcal{M}_u \mathbf{S}_u \mathbf{U}_u$  and  $\mathbf{M}'_\ell = \mathbf{H}_\ell \mathbf{U}_\ell = \mathbf{S}_\ell^\dagger \mathcal{M}_\ell \mathbf{S}_\ell \mathbf{U}_\ell$ . In terms of the diagonal mass matrices

$$\mathcal{M}_d = \text{diag}(m_d, m_s, m_b, \dots), \quad \mathcal{M}_u = \text{diag}(m_u, m_c, m_t, \dots), \quad \mathcal{M}_\ell = \text{diag}(m_e, m_\mu, m_\tau, \dots), \quad (7)$$

the Yukawa Lagrangian takes the simpler form

$$\mathcal{L}_Y = - \left( 1 + \frac{H}{v} \right) \{ \bar{\mathbf{d}} \mathcal{M}_d \mathbf{d} + \bar{\mathbf{u}} \mathcal{M}_u \mathbf{u} + \bar{\ell} \mathcal{M}_\ell \ell \}, \quad (8)$$

where the mass eigenstates are defined by

$$\begin{aligned} \mathbf{d}_L &\equiv \mathbf{S}_d \mathbf{d}'_L, & \mathbf{u}_L &\equiv \mathbf{S}_u \mathbf{u}'_L, & \ell_L &\equiv \mathbf{S}_\ell \ell'_L, \\ \mathbf{d}_R &\equiv \mathbf{S}_d \mathbf{U}_d \mathbf{d}'_R, & \mathbf{u}_R &\equiv \mathbf{S}_u \mathbf{U}_u \mathbf{u}'_R, & \ell_R &\equiv \mathbf{S}_\ell \mathbf{U}_\ell \ell'_R. \end{aligned} \quad (9)$$

Note, that the Higgs couplings are flavour-conserving and proportional to the corresponding fermion masses.

Since,  $\bar{\mathbf{f}}'_L \mathbf{f}'_L = \bar{\mathbf{f}}_L \mathbf{f}_L$  and  $\bar{\mathbf{f}}'_R \mathbf{f}'_R = \bar{\mathbf{f}}_R \mathbf{f}_R$  ( $f = d, u, \ell$ ), the form of the neutral-current part of the  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$  Lagrangian does not change when expressed in terms of mass eigenstates. Therefore, there are no flavour-changing neutral currents (FCNCs) in the SM. This is a consequence of treating all equal-charge fermions on the same footing (GIM mechanism [8]), and guarantees that weak transitions such as  $B_{s,d}^0 \rightarrow \ell^+ \ell^-$ ,  $K^0 \rightarrow \mu^+ \mu^-$  or  $K^0 - \bar{K}^0$  mixing (Fig. 2), which are known experimentally to be very suppressed, cannot happen at tree level. The absence of FCNCs is crucial for the phenomenological success of the SM. However,  $\bar{\mathbf{u}}'_L \mathbf{d}'_L = \bar{\mathbf{u}}_L \mathbf{S}_u \mathbf{S}_d^\dagger \mathbf{d}_L \equiv \bar{\mathbf{u}}_L \mathbf{V} \mathbf{d}_L$ . In general,  $\mathbf{S}_u \neq \mathbf{S}_d$ ; thus, if one writes the weak eigenstates in terms of mass eigenstates, a  $N_G \times N_G$  unitary mixing matrix  $\mathbf{V}$  appears in the quark charged-current sector as indicated in Eq. (3).

If neutrinos are assumed to be massless, we can always redefine the neutrino flavours, in such a way as to eliminate the mixing in the lepton sector:  $\bar{\nu}'_L \ell'_L = \bar{\nu}'_L \mathbf{S}_\ell^\dagger \ell_L \equiv \bar{\nu}_L \ell_L$ . Thus, we have lepton-flavour conservation in the minimal SM without right-handed neutrinos. If sterile  $\nu_R$  fields are included in the model, one has an additional Yukawa term in Eq. (4), giving rise to a neutrino mass matrix  $(\mathbf{M}'_\nu)_{ij} \equiv c_{ij}^{(\nu)} v / \sqrt{2}$ . Thus, the model can accommodate non-zero neutrino masses and lepton-flavour violation through a lepton mixing matrix  $\mathbf{V}_L$  analogous to the one present in the quark sector. Note, however, that the total lepton number  $L \equiv L_e + L_\mu + L_\tau$  is still conserved. We know experimentally that neutrino masses are tiny and, as shown in Table 1, there are strong bounds on lepton-flavour violating decays. However, we do have a clear evidence of neutrino oscillation phenomena [9]. Moreover, since right-handed neutrinos are singlets under  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ , the SM gauge symmetry group

<sup>1</sup>The condition  $\det \mathbf{M}'_f \neq 0$  ( $f = d, u, \ell$ ) guarantees that the decomposition  $\mathbf{M}'_f = \mathbf{H}_f \mathbf{U}_f$  is unique:  $\mathbf{U}_f \equiv \mathbf{H}_f^{-1} \mathbf{M}'_f$ . The matrices  $\mathbf{S}_f$  are completely determined (up to phases) only if all diagonal elements of  $\mathbf{M}'_f$  are different. If there is some degeneracy, the arbitrariness of  $\mathbf{S}_f$  reflects the freedom to define the physical fields. When  $\det \mathbf{M}'_f = 0$ , the matrices  $\mathbf{U}_f$  and  $\mathbf{S}_f$  are not uniquely determined, unless their unitarity is explicitly imposed.

**Table 1:** Best published limits on lepton-flavour-violating transitions [9].

$\text{Br}(\mu^- \rightarrow X^-) \cdot 10^{12}$ (90% CL)							
$e^- \gamma$	0.42	$e^- 2\gamma$	72	$e^- e^- e^+$	1.0		
$\Gamma(\mu^- + N \rightarrow e^- + N)/\Gamma(\mu^- + N \rightarrow \text{capture}) \cdot 10^{12}$ (90% CL)							
Au	0.7	Ti	4.3	Pb	46		
$\text{Br}(\tau^- \rightarrow X^-) \cdot 10^8$ (90% CL)							
$e^- \gamma$	3.3	$e^- e^+ e^-$	2.7	$e^- \mu^+ \mu^-$	2.7	$e^- e^- \mu^+$	1.5
$\mu^- \gamma$	4.4	$\mu^- e^+ e^-$	1.8	$\mu^- \mu^+ \mu^-$	2.1	$\mu^- \mu^- e^+$	1.7
$e^- \pi^0$	8.0	$\mu^- \pi^0$	11	$e^- \phi$	3.1	$\mu^- \phi$	8.4
$e^- \eta$	9.2	$e^- \eta'$	16	$e^- \rho^0$	1.8	$e^- \omega$	4.8
$\mu^- \eta$	6.5	$\mu^- \eta'$	13	$\mu^- \rho^0$	1.2	$\mu^- \omega$	4.7
$e^- K_S$	2.6	$e^- K^{*0}$	3.2	$e^- \bar{K}^{*0}$	3.4	$e^- K^+ \pi^-$	3.1
$\mu^- K_S$	2.3	$\mu^- K^{*0}$	5.9	$\mu^- \bar{K}^{*0}$	7.0	$\mu^- K^+ \pi^-$	4.5
$e^- K_S K_S$	7.1	$e^- K^+ K^-$	3.4	$e^- \pi^+ \pi^-$	2.3	$e^- \pi^+ K^-$	3.7
$\mu^- K_S K_S$	8.0	$\mu^- K^+ K^-$	4.4	$\mu^- \pi^+ \pi^-$	2.1	$\mu^- \pi^+ K^-$	8.6
$e^- f_0(980) \rightarrow e^- \pi^+ \pi^-$			3.2	$\mu^- f_0(980) \rightarrow \mu^- \pi^+ \pi^-$			3.4
$\text{Br}(Z \rightarrow X^0) \cdot 10^6$ (95% CL)							
$e^\pm \mu^\mp$	0.75	$e^\pm \tau^\mp$	9.8	$\mu^\pm \tau^\mp$	12		
$\text{Br}(H \rightarrow X^0) \cdot 10^3$ (95% CL)							
$e^\pm \mu^\mp$	0.061	$e^\pm \tau^\mp$	4.7	$\mu^\pm \tau^\mp$	2.5		
$\text{Br}(\pi^0 \rightarrow X^0) \cdot 10^9$ (90% CL)			$\text{Br}(K^+ \rightarrow X^+) \cdot 10^{11}$ (90% CL)				
$\mu^+ e^-$	0.38	$\mu^- e^+$	3.4	$\pi^+ \mu^+ e^-$	1.3	$\pi^+ \mu^- e^+$	52
$\text{Br}(K_L^0 \rightarrow X^0) \cdot 10^{11}$ (90% CL)							
$e^\pm \mu^\mp$	0.47	$e^\pm e^\pm \mu^\mp \mu^\mp$	4.12	$\pi^0 \mu^\pm e^\mp$	7.6	$\pi^0 \pi^0 \mu^\pm e^\mp$	17
$\text{Br}(B_{(s)}^0 \rightarrow X^0) \cdot 10^9$ (90% CL)			$\text{Br}(B^+ \rightarrow X^+) \cdot 10^9$ (90% CL)				
$B^0 \rightarrow e^\pm \mu^\mp$	1.0	$B_s^0 \rightarrow e^\pm \mu^\mp$	5.4	$K^+ e^- \mu^+$	6.4	$K^+ e^+ \mu^-$	7.0

allows for a right-handed Majorana neutrino mass term, violating lepton number by two units. Non-zero neutrino masses clearly imply interesting new phenomena [4].

The fermion masses and the quark mixing matrix  $\mathbf{V}$  are all generated by the Yukawa couplings in Eq. (4). However, the complex coefficients  $c_{ij}^{(f)}$  are not determined by the gauge symmetry; therefore, we have a large number of arbitrary parameters. A general  $N_G \times N_G$  unitary matrix is characterized by  $N_G^2$  real parameters:  $N_G(N_G - 1)/2$  moduli and  $N_G(N_G + 1)/2$  phases. In the case of  $\mathbf{V}$ , many of these parameters are irrelevant because we can always choose arbitrary quark phases. Under the phase redefinitions  $u_i \rightarrow e^{i\phi_i} u_i$  and  $d_j \rightarrow e^{i\theta_j} d_j$ , the mixing matrix changes as  $\mathbf{V}_{ij} \rightarrow \mathbf{V}_{ij} e^{i(\theta_j - \phi_i)}$ ; thus,  $2N_G - 1$  phases are unobservable. The number of physical free parameters in the quark-mixing matrix



then gets reduced to  $(N_G - 1)^2$ :  $N_G(N_G - 1)/2$  moduli and  $(N_G - 1)(N_G - 2)/2$  phases.

In the simpler case of two generations,  $\mathbf{V}$  is determined by a single parameter. One then recovers the Cabibbo rotation matrix [6]

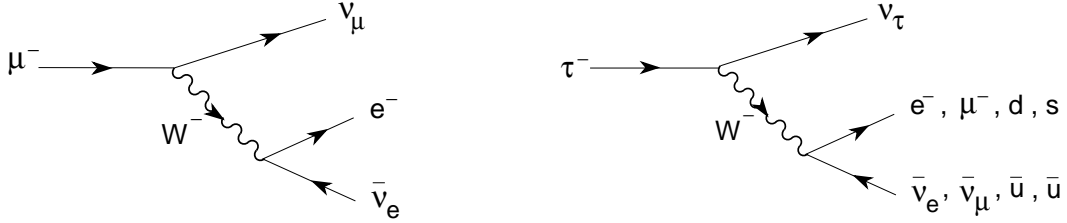
$$\mathbf{V} = \begin{pmatrix} \cos \theta_C & \sin \theta_C \\ -\sin \theta_C & \cos \theta_C \end{pmatrix}. \quad (10)$$

With  $N_G = 3$ , the CKM matrix is described by three angles and one phase. Different (but equivalent) representations can be found in the literature. The Particle data Group [9] advocates the use of the following one as the ‘standard’ CKM parametrization:

$$\begin{aligned} \mathbf{V} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \cdot \begin{bmatrix} c_{13} & 0 & s_{13} e^{-i\delta_{13}} \\ 0 & 1 & 0 \\ -s_{13} e^{i\delta_{13}} & 0 & c_{13} \end{bmatrix} \cdot \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} c_{12} c_{13} & s_{12} c_{13} & s_{13} e^{-i\delta_{13}} \\ -s_{12} c_{23} - c_{12} s_{23} s_{13} e^{i\delta_{13}} & c_{12} c_{23} - s_{12} s_{23} s_{13} e^{i\delta_{13}} & s_{23} c_{13} \\ s_{12} s_{23} - c_{12} c_{23} s_{13} e^{i\delta_{13}} & -c_{12} s_{23} - s_{12} c_{23} s_{13} e^{i\delta_{13}} & c_{23} c_{13} \end{bmatrix}. \quad (11) \end{aligned}$$

Here  $c_{ij} \equiv \cos \theta_{ij}$  and  $s_{ij} \equiv \sin \theta_{ij}$ , with  $i$  and  $j$  being generation labels ( $i, j = 1, 2, 3$ ). The real angles  $\theta_{12}$ ,  $\theta_{23}$  and  $\theta_{13}$  can all be made to lie in the first quadrant, by an appropriate redefinition of quark field phases; then,  $c_{ij} \geq 0$ ,  $s_{ij} \geq 0$  and  $0 \leq \delta_{13} \leq 2\pi$ . Notice that  $\delta_{13}$  is the only complex phase in the SM Lagrangian. Therefore, it is the only possible source of  $\mathcal{CP}$ -violation phenomena. In fact, it was for this reason that the third generation was assumed to exist [7], before the discovery of the  $b$  and the  $\tau$ . With two generations, the SM could not explain the observed  $\mathcal{CP}$  violation in the  $K$  system.

### 3 Lepton decays



**Fig. 3:** Tree-level Feynman diagrams for  $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$  and  $\tau^- \rightarrow \nu_\tau X^-$  ( $X^- = e^- \bar{\nu}_e, \mu^- \bar{\nu}_\mu, d\bar{u}, s\bar{u}$ ).

The simplest flavour-changing process is the leptonic decay of the muon, which proceeds through the  $W$ -exchange diagram shown in Fig. 3. The momentum transfer carried by the intermediate  $W$  is very small compared to  $M_W$ . Therefore, the vector-boson propagator reduces to a contact interaction,

$$\frac{-g_{\mu\nu} + q_\mu q_\nu / M_W^2}{q^2 - M_W^2} \xrightarrow{q^2 \ll M_W^2} \frac{g_{\mu\nu}}{M_W^2}. \quad (12)$$

The decay can then be described through an effective local four-fermion Hamiltonian,

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} [\bar{e} \gamma^\alpha (1 - \gamma_5) \nu_e] [\bar{\nu}_\mu \gamma_\alpha (1 - \gamma_5) \mu], \quad (13)$$

where

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2} = \frac{1}{2v^2} \quad (14)$$

**Table 2:** Experimental determinations of the ratios  $g_\ell/g_{\ell'}$  [9, 17].

	$\Gamma_{\tau \rightarrow \mu}/\Gamma_{\tau \rightarrow e}$	$\Gamma_{\pi \rightarrow \mu}/\Gamma_{\pi \rightarrow e}$	$\Gamma_{K \rightarrow \mu}/\Gamma_{K \rightarrow e}$	$\Gamma_{K \rightarrow \pi \mu}/\Gamma_{K \rightarrow \pi e}$	$\Gamma_{W \rightarrow \mu}/\Gamma_{W \rightarrow e}$
$ g_\mu/g_e $	1.0017 (16)	1.0010 (9)	0.9978 (18)	1.0010 (25)	0.993 (7)
	$\Gamma_{\tau \rightarrow e}/\Gamma_{\mu \rightarrow e}$	$\Gamma_{\tau \rightarrow \pi}/\Gamma_{\pi \rightarrow \mu}$	$\Gamma_{\tau \rightarrow K}/\Gamma_{K \rightarrow \mu}$	$\Gamma_{W \rightarrow \tau}/\Gamma_{W \rightarrow \mu}$	
$ g_\tau/g_\mu $	1.0011 (14)	0.9965 (26)	0.9860 (73)	1.034 (13)	
	$\Gamma_{\tau \rightarrow \mu}/\Gamma_{\mu \rightarrow e}$	$\Gamma_{W \rightarrow \tau}/\Gamma_{W \rightarrow e}$			
$ g_\tau/g_e $	1.0028 (14)	1.021 (12)			

is called the Fermi coupling constant.  $G_F$  is fixed by the total decay width,

$$\frac{1}{\tau_\mu} = \Gamma[\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu (\gamma)] = \frac{G_F^2 m_\mu^5}{192\pi^3} (1 + \delta_{\text{RC}}) f\left(\frac{m_e^2}{m_\mu^2}\right), \quad (15)$$

where  $f(x) = 1 - 8x + 8x^3 - x^4 - 12x^2 \ln x$ , and  $\delta_{\text{RC}} \approx \frac{\alpha}{2\pi}(\frac{25}{4} - \pi^2)$  takes into account higher-order QED corrections, which are known to  $\mathcal{O}(\alpha^2)$  [10–12]. The tiny neutrino masses can be safely neglected. The measured lifetime [13],  $\tau_\mu = (2.196\,981\,1 \pm 0.000\,002\,2) \cdot 10^{-6}$  s, implies the value

$$G_F = (1.166\,378\,7 \pm 0.000\,000\,6) \cdot 10^{-5} \text{ GeV}^{-2} \approx \frac{1}{(293 \text{ GeV})^2}. \quad (16)$$

The decays of the  $\tau$  lepton proceed through the same  $W$ -exchange mechanism. The only difference is that several final states are kinematically allowed:  $\tau^- \rightarrow \nu_\tau e^- \bar{\nu}_e$ ,  $\tau^- \rightarrow \nu_\tau \mu^- \bar{\nu}_\mu$ ,  $\tau^- \rightarrow \nu_\tau d \bar{u}$  and  $\tau^- \rightarrow \nu_\tau s \bar{u}$ . Owing to the universality of the  $W$  couplings in  $\mathcal{L}_{\text{CC}}$ , all these decay modes have equal amplitudes (if final fermion masses and QCD interactions are neglected), except for an additional  $N_C |\mathbf{V}_{ui}|^2$  factor ( $i = d, s$ ) in the semileptonic channels, where  $N_C = 3$  is the number of quark colours. Making trivial kinematical changes in Eq. (15), one easily gets the lowest-order prediction for the total  $\tau$  decay width:

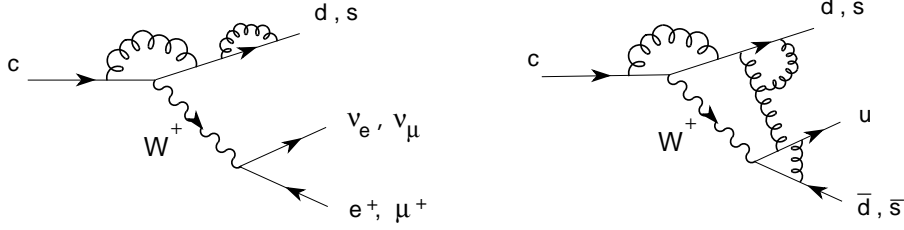
$$\frac{1}{\tau_\tau} \equiv \Gamma(\tau) \approx \Gamma(\mu) \left(\frac{m_\tau}{m_\mu}\right)^5 \left\{2 + N_C \left(|\mathbf{V}_{ud}|^2 + |\mathbf{V}_{us}|^2\right)\right\} \approx \frac{5}{\tau_\mu} \left(\frac{m_\tau}{m_\mu}\right)^5, \quad (17)$$

where we have used the CKM unitarity relation  $|\mathbf{V}_{ud}|^2 + |\mathbf{V}_{us}|^2 = 1 - |\mathbf{V}_{ub}|^2 \approx 1$  (we will see later that this is an excellent approximation). From the measured muon lifetime, one has then  $\tau_\tau \approx 3.3 \cdot 10^{-13}$  s, to be compared with the experimental value  $\tau_\tau^{\text{exp}} = (2.903 \pm 0.005) \cdot 10^{-13}$  s [9]. The numerical difference is due to the effect of QCD corrections, which enhance the hadronic  $\tau$  decay width by about 20%. The size of these corrections has been accurately predicted in terms of the strong coupling [14], allowing us to extract from  $\tau$  decays one of the most precise determinations of  $\alpha_s$  [15, 16].

In the SM all lepton doublets have identical couplings to the  $W$  boson. Comparing the measured decay widths of leptonic or semileptonic decays which only differ in the lepton flavour, one can test experimentally that the  $W$  interaction is indeed the same, i.e., that  $g_e = g_\mu = g_\tau \equiv g$ . As shown in Table 2, the present data verify the universality of the leptonic charged-current couplings to the 0.2% level.

#### 4 Quark mixing

In order to measure the CKM matrix elements  $\mathbf{V}_{ij}$ , one needs to study hadronic weak decays of the type  $H \rightarrow H' \ell^- \bar{\nu}_\ell$  or  $H' \rightarrow H \ell^+ \nu_\ell$  that are associated with the corresponding quark transitions



**Fig. 4:**  $\mathbf{V}_{ij}$  are measured in semileptonic decays (left), where a single quark current is present. Hadronic decays (right) involve two different quark currents and are more affected by QCD effects (gluons can couple everywhere).

$d_j \rightarrow u_i \ell^- \bar{\nu}_\ell$  and  $u_i \rightarrow d_j \ell^+ \nu_\ell$  (Fig. 4). Since quarks are confined within hadrons, the decay amplitude

$$T[H \rightarrow H' \ell^- \bar{\nu}_\ell] = \frac{G_F}{\sqrt{2}} \mathbf{V}_{ij} \langle H' | \bar{u}_i \gamma^\mu (1 - \gamma_5) d_j | H \rangle [\bar{\ell} \gamma_\mu (1 - \gamma_5) \nu_\ell] \quad (18)$$

always involves an hadronic matrix element of the weak left current. The evaluation of this matrix element is a non-perturbative QCD problem, which introduces unavoidable theoretical uncertainties.

One usually looks for a semileptonic transition where the matrix element can be fixed at some specific kinematic point by a symmetry principle. This has the virtue of reducing the theoretical uncertainties to the level of symmetry-breaking corrections and kinematic extrapolations. The standard example is a  $0^- \rightarrow 0^-$  decay such as  $K \rightarrow \pi \ell \nu_\ell$ ,  $D \rightarrow K \ell \nu_\ell$  or  $B \rightarrow D \ell \nu_\ell$ , where, owing to parity (the vector and axial-vector currents have  $J^P = 1^-$  and  $1^+$ , respectively), only the vector current contributes. The most general Lorentz decomposition of the hadronic matrix element contains two terms:

$$\langle P'(k') | \bar{u}_i \gamma^\mu d_j | P(k) \rangle = C_{PP'} \{ (k + k')^\mu f_+(t) + (k - k')^\mu f_-(t) \}. \quad (19)$$

Here,  $C_{PP'}$  is a Clebsch–Gordan factor relating  $P \rightarrow P'$  transitions that only differ by the meson electromagnetic charges, and  $t = (k - k')^2 \equiv q^2$  is the momentum transfer. The unknown strong dynamics is fully contained in the form factors  $f_\pm(t)$ .

In the limit of equal quark masses,  $m_{u_i} = m_{d_j}$ , the divergence of the vector current is zero. Thus  $q_\mu [\bar{u}_i \gamma^\mu d_j] = 0$ , which implies  $f_-(t) = 0$ . Moreover, as shown in the appendix,  $f_+(0) = 1$  to all orders in the strong coupling because the associated flavour charge is a conserved quantity.<sup>2</sup> Therefore, one only needs to estimate the corrections induced by the quark mass differences.

Since  $q_\mu [\bar{\ell} \gamma^\mu (1 - \gamma_5) \nu_\ell] \sim m_\ell$ , the contribution of  $f_-(t)$  is kinematically suppressed in the electron and muon decay modes. The decay width can then be written as ( $\ell = e, \mu$ )

$$\Gamma(P \rightarrow P' \ell \nu) = \frac{G_F^2 M_P^5}{192\pi^3} |\mathbf{V}_{ij}|^2 C_{PP'}^2 |f_+(0)|^2 \mathcal{I} (1 + \delta_{\text{RC}}), \quad (20)$$

where  $\delta_{\text{RC}}$  is an electroweak radiative correction factor and  $\mathcal{I}$  denotes a phase-space integral, which in the limit  $m_\ell = 0$  takes the form

$$\mathcal{I} \approx \int_0^{(M_P - M_{P'})^2} \frac{dt}{M_P^8} \lambda^{3/2}(t, M_P^2, M_{P'}^2) \left| \frac{f_+(t)}{f_+(0)} \right|^2. \quad (21)$$

The usual procedure to determine  $|\mathbf{V}_{ij}|$  involves three steps:

1. Measure the shape of the  $t$  distribution. This fixes  $|f_+(t)/f_+(0)|$  and therefore determines  $\mathcal{I}$ .

<sup>2</sup>This is completely analogous to the electromagnetic charge conservation in QED. The conservation of the electromagnetic current implies that the proton electromagnetic form factor does not get any QED or QCD correction at  $q^2 = 0$  and, therefore,  $Q_p = 2Q_u + Q_d = |Q_e|$ . An explicit proof can be found in Ref. [18].

2. Measure the total decay width  $\Gamma$ . Since  $G_F$  is already known from  $\mu$  decay, one gets then an experimental value for the product  $|f_+(0) \mathbf{V}_{ij}|$ , provided the radiative correction  $\delta_{\text{RC}}$  is known to the needed accuracy.
3. Get a theoretical prediction for  $f_+(0)$ .

It is important to realize that theoretical input is always needed. Thus, the accuracy of the  $|\mathbf{V}_{ij}|$  determination is limited by our ability to calculate the relevant hadronic parameters and radiative corrections.

#### 4.1 Determination of $|\mathbf{V}_{ud}|$ and $|\mathbf{V}_{us}|$

The conservation of the vector QCD currents in the massless quark limit allows for precise determinations of the light-quark mixings. The most accurate measurement of  $\mathbf{V}_{ud}$  is done with superallowed nuclear  $\beta$  decays of the Fermi type ( $0^+ \rightarrow 0^+$ ), where the nuclear matrix element  $\langle N' | \bar{u} \gamma^\mu d | N \rangle$  can be fixed by vector-current conservation. The CKM factor is obtained through the relation [19]

$$|\mathbf{V}_{ud}|^2 = \frac{\pi^3 \ln 2}{ft G_F^2 m_e^5 (1 + \delta_{\text{RC}})} = \frac{(2984.430 \pm 0.003) \text{ s}}{ft (1 + \delta_{\text{RC}})} = \frac{(2984.430 \pm 0.003) \text{ s}}{\mathcal{F}t (1 + \Delta_R^V)}, \quad (22)$$

where  $ft$  denotes the product of a phase-space statistical decay-rate factor and the measured half-life of the transition. In order to determine  $|\mathbf{V}_{ud}|$ , one needs to perform a careful analysis of radiative corrections, including electroweak contributions, nuclear-structure corrections and isospin-violating nuclear effects. The nucleus-dependent corrections, which are reabsorbed into an effective nucleus-independent  $\mathcal{F}t$ -value, have a crucial role in bringing the results from different nuclei into good agreement. The weighted average of the fourteen most precise determinations yields  $\mathcal{F}t = (3072.07 \pm 0.63) \text{ s}$  [19, 20]. The remaining universal correction  $\Delta_R^V$  is sizeable and its previously accepted value [21] has been questioned by recent re-evaluations [22, 23]. Taking  $\Delta_R^V = 0.02426 \pm 0.00032$  [24], one gets

$$|\mathbf{V}_{ud}| = 0.97389 \pm 0.00018. \quad (23)$$

An independent determination of  $|\mathbf{V}_{ud}|$  can be obtained from neutron decay,  $n \rightarrow p e^- \bar{\nu}_e$ . The axial current also contributes in this case; therefore, one needs to use the experimental value of the axial-current matrix element at  $q^2 = 0$ ,  $\langle p | \bar{u} \gamma^\mu \gamma_5 d | n \rangle = G_A \bar{p} \gamma^\mu n$ , which can be extracted from the distribution of the neutron decay products. Using the current world averages,  $g_A \equiv G_A/G_V = -1.2732 \pm 0.0023$  and  $\tau_n = (879.4 \pm 0.6) \text{ s}$  [9], and the estimated radiative corrections [24], one gets

$$|\mathbf{V}_{ud}| = \left\{ \frac{(4906.4 \pm 1.7) \text{ s}}{\tau_n (1 + 3g_A^2)} \right\}^{1/2} = 0.9755 \pm 0.0015, \quad (24)$$

which is  $1.1 \sigma$  larger than (23) but less precise. The uncertainty on the input value of  $g_A$  has been inflated because the most recent and accurate measurements of  $g_A$  disagree with the older experiments. Using instead the post-2002 average  $g_A = -1.2762 \pm 0.0005$  [24], results in  $|\mathbf{V}_{ud}| = 0.9736 \pm 0.0005$ ; three times more precise and in better agreement with (23).

The pion  $\beta$  decay  $\pi^+ \rightarrow \pi^0 e^+ \nu_e$  offers a cleaner way to measure  $|\mathbf{V}_{ud}|$ . It is a pure vector transition, with very small theoretical uncertainties. At  $q^2 = 0$ , the hadronic matrix element does not receive isospin-breaking contributions of first order in  $m_d - m_u$ , i.e.,  $f_+(0) = 1 + \mathcal{O}[(m_d - m_u)^2]$  [25]. The small available phase space makes it possible to theoretically control the form factor with high accuracy over the entire kinematical domain [26]; unfortunately, it also implies a very suppressed branching fraction of  $\mathcal{O}(10^{-8})$ . From the currently measured value [27], one gets  $|\mathbf{V}_{ud}| = 0.9749 \pm 0.0026$  [9]. A tenfold improvement of the experimental accuracy would be needed to get a determination competitive with (23).

The standard determination of  $|\mathbf{V}_{us}|$  takes advantage of the theoretically well-understood decay amplitudes in  $K \rightarrow \pi \ell \nu_\ell$ . The high accuracy achieved in high-statistics experiments [9], supplemented

with theoretical calculations of electromagnetic and isospin corrections [28, 29], allows us to extract the product  $|\mathbf{V}_{us} f_+(0)| = 0.2165 \pm 0.0004$  [30, 31], with  $f_+(0) = 1 + \mathcal{O}[(m_s - m_u)^2]$  the vector form factor of the  $K^0 \rightarrow \pi^- \ell^+ \nu_\ell$  decay [25, 32]. The exact value of  $f_+(0)$  has been thoroughly investigated since the first precise estimate by Leutwyler and Roos,  $f_+(0) = 0.961 \pm 0.008$  [33]. The most recent and precise lattice determinations exhibit a clear shift to higher values [34, 35], in agreement with the analytical chiral perturbation theory predictions at two loops [36–38]. Taking the current lattice average (with  $2 + 1 + 1$  active fermions),  $f_+(0) = 0.9706 \pm 0.0027$  [39], one obtains

$$|\mathbf{V}_{us}| = 0.2231 \pm 0.0007. \quad (25)$$

The ratio of radiative inclusive decay rates  $\Gamma[K \rightarrow \mu\nu(\gamma)]/\Gamma[\pi \rightarrow \mu\nu(\gamma)]$  provides also information on  $\mathbf{V}_{us}$  [30, 40]. With a careful treatment of electromagnetic and isospin-violating corrections, one extracts  $|\mathbf{V}_{us}/\mathbf{V}_{ud}| |f_K/f_\pi| = 0.2760 \pm 0.0004$  [31, 41, 42]. Taking for the ratio of meson decay constants the lattice average  $f_K/f_\pi = 1.1932 \pm 0.0019$  [39], one finally gets

$$\frac{|\mathbf{V}_{us}|}{|\mathbf{V}_{ud}|} = 0.2313 \pm 0.0005. \quad (26)$$

With the value of  $|\mathbf{V}_{ud}|$  in Eq. (23), this implies  $|\mathbf{V}_{us}| = 0.2253 \pm 0.0005$  that is  $2.6\sigma$  larger than Eq. (25).

Hyperon decays are also sensitive to  $\mathbf{V}_{us}$  [43]. Unfortunately, in weak baryon decays the theoretical control on  $SU(3)$ -breaking corrections is not as good as for the meson case. A conservative estimate of these effects leads to the result  $|\mathbf{V}_{us}| = 0.226 \pm 0.005$  [44].

The accuracy of all previous determinations is limited by theoretical uncertainties. The ratio of the inclusive  $\Delta S = 0$  ( $\tau^- \rightarrow \nu_\tau \bar{u}d$ ) and  $|\Delta S| = 1$  ( $\tau^- \rightarrow \nu_\tau \bar{u}s$ ) tau decay widths provides a very clean observable to directly measure  $|\mathbf{V}_{us}|$  [17, 45] because  $SU(3)$ -breaking corrections are suppressed by two powers of the  $\tau$  mass. The present  $\tau$  decay data imply  $|\mathbf{V}_{us}| = 0.2195 \pm 0.0019$  [46], which is  $1.8\sigma$  lower than Eq. (25), and  $3.0\sigma$  lower than the value extracted from Eqs. (26) and (23).

## 4.2 Determination of $|\mathbf{V}_{cb}|$ and $|\mathbf{V}_{ub}|$

In the limit of very heavy quark masses, QCD has additional flavour and spin symmetries [47–50] that can be used to make precise determinations of  $|\mathbf{V}_{cb}|$ , either from exclusive semileptonic decays such as  $B \rightarrow D\ell\bar{\nu}_\ell$  and  $B \rightarrow D^*\ell\bar{\nu}_\ell$  [51, 52] or from the inclusive analysis of  $b \rightarrow c\ell\bar{\nu}_\ell$  transitions. In the rest frame of a heavy-light meson  $\bar{Q}q$ , with  $M_Q \gg (m_q, \Lambda_{\text{QCD}})$ , the heavy quark  $Q$  is practically at rest and acts as a static source of gluons ( $\lambda_Q \sim 1/M_Q \ll R_{\text{had}} \sim 1/\Lambda_{\text{QCD}}$ ). At  $M_Q \rightarrow \infty$ , the interaction becomes then independent of the heavy-quark mass and spin. Moreover, assuming that the charm quark is heavy enough, the  $b \rightarrow c\ell\bar{\nu}_\ell$  transition within the meson does not modify the interaction with the light quark at zero recoil, i.e., when the meson velocity remains unchanged ( $v_D = v_B$ ).

Taking the limit  $m_b > m_c \rightarrow \infty$ , all form factors characterizing the decays  $B \rightarrow D\ell\bar{\nu}_\ell$  and  $B \rightarrow D^*\ell\bar{\nu}_\ell$  reduce to a single function [47], which depends on the product of the four-velocities of the two mesons  $w \equiv v_B \cdot v_{D^{(*)}} = (M_B^2 + M_{D^{(*)}}^2 - q^2)/(2M_B M_{D^{(*)}})$ . Heavy quark symmetry determines the normalization of the rate at  $w = 1$ , the maximum momentum transfer to the leptons, because the corresponding vector current is conserved in the limit of equal  $B$  and  $D^{(*)}$  velocities. The  $B \rightarrow D^*$  mode has the additional advantage that corrections to the infinite-mass limit are of second order in  $1/m_b - 1/m_c$  at zero recoil ( $w = 1$ ) [52].

The exclusive determination of  $|\mathbf{V}_{cb}|$  is obtained from an extrapolation of the measured spectrum to  $w = 1$ . Using the CLN parametrization of the relevant form factors [53], which is based on heavy-quark symmetry and includes  $1/M_Q$  corrections, the *Heavy Flavor Averaging group* (HFLAV) [46] quotes the experimental value  $\eta_{\text{EW}} \mathcal{F}(1) |\mathbf{V}_{cb}| = (35.27 \pm 0.38) \cdot 10^{-3}$  from  $B \rightarrow D^*\ell\bar{\nu}_\ell$  data, while the measured

$B \rightarrow D\ell\bar{\nu}_\ell$  distribution results in  $\eta_{\text{EW}} \mathcal{G}(1) |\mathbf{V}_{cb}| = (42.00 \pm 1.00) \cdot 10^{-3}$ , where  $\mathcal{F}(1)$  and  $\mathcal{G}(1)$  are the corresponding form factors at  $w = 1$  and  $\eta_{\text{EW}}$  accounts for small electroweak corrections. Lattice simulations are used to estimate the deviations from unity of the two form factors at zero recoil. Using  $\eta_{\text{EW}} \mathcal{F}(1) = 0.910 \pm 0.013$  [39] and  $\eta_{\text{EW}} \mathcal{G}(1) = 1.061 \pm 0.010$  [54], one gets [46]

$$|\mathbf{V}_{cb}| = \begin{cases} (38.76 \pm 0.42_{\text{exp}} \pm 0.55_{\text{th}}) \cdot 10^{-3} & (B \rightarrow D^* \ell \bar{\nu}_\ell) \\ (39.58 \pm 0.94_{\text{exp}} \pm 0.37_{\text{th}}) \cdot 10^{-3} & (B \rightarrow D \ell \bar{\nu}_\ell) \end{cases} = (39.02 \pm 0.57) \cdot 10^{-3}. \quad (27)$$

It has been pointed out recently that the CLN parametrization is only valid within 2% and this uncertainty has not been properly taken into account in the experimental extrapolations [55–58]. Using instead the more general BGL parametrization [59], combined with lattice and light-cone sum rules information, the analysis of the most recent  $B \rightarrow D^* \ell \bar{\nu}_\ell$  Belle data [60, 61] gives [62]

$$|\mathbf{V}_{cb}| = (39.6^{+1.1}_{-1.0}) \cdot 10^{-3}, \quad (28)$$

while a similar analysis of BaBar [63] and Belle [64]  $B \rightarrow D \ell \bar{\nu}_\ell$  data obtains [55]

$$|\mathbf{V}_{cb}| = (40.49 \pm 0.97) \cdot 10^{-3}. \quad (29)$$

These numbers are significantly higher than the corresponding HFLAV results in Eq. (27) and indicate the presence of underestimated uncertainties.

The inclusive determination of  $|\mathbf{V}_{cb}|$  uses the Operator Product Expansion [65, 66] to express the total  $b \rightarrow c \ell \bar{\nu}_\ell$  rate and moments of the differential energy and invariant-mass spectra in a double expansion in powers of  $\alpha_s$  and  $1/m_b$ , which includes terms of  $\mathcal{O}(\alpha_s^2)$  and up to  $\mathcal{O}(1/m_b^5)$  [12, 67–77]. The non-perturbative matrix elements of the corresponding local operators are obtained from a global fit to experimental moments of inclusive lepton energy and hadronic invariant mass distributions. The most recent analyses find [78, 79]

$$|\mathbf{V}_{cb}| = (42.00 \pm 0.64) \cdot 10^{-3}. \quad (30)$$

This value, which we will adopt in the following, agrees within errors with the exclusive  $B \rightarrow D$  determination in Eq. (29) and it is only  $1.9\sigma$  away from the  $B \rightarrow D^*$  value in Eq. (28).

The presence of a light quark makes more difficult to control the theoretical uncertainties in the analogous determinations of  $|\mathbf{V}_{ub}|$ . Exclusive  $B \rightarrow \pi \ell \nu_\ell$  decays involve a non-perturbative form factor  $f_+(t)$  which is estimated through light-cone sum rules [80–83] and lattice simulations [84, 85]. The inclusive measurement requires the use of stringent experimental cuts to suppress the  $b \rightarrow X_c \ell \nu_\ell$  background that has fifty times larger rates. This induces sizeable errors in the theoretical predictions [86–94], which become sensitive to non-perturbative shape functions and depend much more strongly on  $m_b$ . The HFLAV group quotes the values [46]

$$|\mathbf{V}_{ub}| = \begin{cases} (3.67 \pm 0.09_{\text{exp}} \pm 0.12_{\text{th}}) \cdot 10^{-3} & (B \rightarrow \pi \ell \bar{\nu}_\ell) \\ (4.32 \pm 0.12_{\text{exp}} \pm 0.13_{\text{th}}) \cdot 10^{-3} & (B \rightarrow X_u \ell \bar{\nu}_\ell) \end{cases} = (3.95 \pm 0.32) \cdot 10^{-3}. \quad (31)$$

Since the exclusive and inclusive determinations of  $|\mathbf{V}_{ub}|$  disagree, we have averaged both values scaling the error by  $\sqrt{\chi^2/\text{dof}} = 2.8$ .

LHCb has extracted  $|\mathbf{V}_{ub}|/|\mathbf{V}_{cb}|$  from the measured ratio of high- $q^2$  events between the  $\Lambda_b$  decay modes into  $p\mu\nu$  ( $q^2 > 15 \text{ GeV}^2$ ) and  $\Lambda_c\mu\nu$  ( $q^2 > 7 \text{ GeV}^2$ ) [46, 95]:

$$\frac{|\mathbf{V}_{ub}|}{|\mathbf{V}_{cb}|} = 0.079 \pm 0.004_{\text{exp}} \pm 0.004_{\text{FF}}, \quad (32)$$

where the second error is due to the limited knowledge of the relevant form factors. This ratio is compatible with the values of  $|\mathbf{V}_{cb}|$  and  $|\mathbf{V}_{ub}|$  in Eqs. (30) and (31), at the  $1.6\sigma$  level.

$|\mathbf{V}_{ub}|$  can be also extracted from the  $B^- \rightarrow \tau^- \bar{\nu}_\tau$  decay width, taking the B-meson decay constant  $f_B$  from lattice calculations [39]. Unfortunately, the current tension between the BaBar [96] and Belle [97] measurements does not allow for a very precise determination. The particle data group quotes  $|\mathbf{V}_{ub}| = (4.01 \pm 0.37) \cdot 10^{-3}$  [9], which agrees with either the exclusive or inclusive values in Eq. (31).

### 4.3 Determination of the charm and top CKM elements

The analytic control of theoretical uncertainties is more difficult in semileptonic charm decays, because the symmetry arguments associated with the light and heavy quark limits get corrected by sizeable symmetry-breaking effects. The magnitude of  $|\mathbf{V}_{cd}|$  can be extracted from  $D \rightarrow \pi \ell \nu_\ell$  and  $D \rightarrow \ell \nu_\ell$  decays, while  $|\mathbf{V}_{cs}|$  is obtained from  $D \rightarrow K \ell \nu_\ell$  and  $D_s \rightarrow \ell \nu_\ell$ , using the lattice determinations of the relevant form factor normalizations and decay constants [39]. The HFLAV group quotes the averages [46]

$$|\mathbf{V}_{cd}| = 0.2204 \pm 0.0040, \quad |\mathbf{V}_{cs}| = 0.969 \pm 0.010. \quad (33)$$

The difference of the ratio of double-muon to single-muon production by neutrino and antineutrino beams is proportional to the charm cross section off valence  $d$  quarks and, therefore, to  $|\mathbf{V}_{cd}|$  times the average semileptonic branching ratio of charm mesons. This allows for an independent determination of  $|\mathbf{V}_{cd}|$ . Averaging data from several experiments, the PDG quotes [9]

$$|\mathbf{V}_{cd}| = 0.230 \pm 0.011, \quad (34)$$

which agrees with (33) but has a larger uncertainty. The analogous determination of  $|\mathbf{V}_{cs}|$  from  $\nu s \rightarrow cX$  suffers from the uncertainty of the  $s$ -quark sea content.

The top quark has only been seen decaying into bottom. From the ratio of branching fractions  $\text{Br}(t \rightarrow Wb)/\text{Br}(t \rightarrow Wq)$ , CMS has extracted [98]

$$\frac{|\mathbf{V}_{tb}|}{\sqrt{\sum_q |\mathbf{V}_{tq}|^2}} > 0.975 \quad (95\% \text{CL}), \quad (35)$$

where  $q = b, s, d$ . A more direct determination of  $|\mathbf{V}_{tb}|$  can be obtained from the single top-quark production cross section, measured at the LHC and the Tevatron. The PDG quotes the world average [9]

$$|\mathbf{V}_{tb}| = 1.019 \pm 0.025. \quad (36)$$

### 4.4 Structure of the CKM matrix

Using the previous determinations of CKM elements, we can check the unitarity of the quark mixing matrix. The most precise test involves the elements of the first row:

$$|\mathbf{V}_{ud}|^2 + |\mathbf{V}_{us}|^2 + |\mathbf{V}_{ub}|^2 = 0.99825 \pm 0.00047, \quad (37)$$

where we have taken as reference values the determinations in Eqs. (23), (25) and (31). Radiative corrections play a crucial role at the quoted level of uncertainty, while the  $|\mathbf{V}_{ub}|^2$  contribution is negligible. This relation exhibits a  $3.7\sigma$  violation of unitarity, at the per-mill level, which calls for an independent re-evaluation of the very precise  $|\mathbf{V}_{ud}|$  value in Eq. (23) and improvements on the  $|\mathbf{V}_{us}|$  determination.

With the  $|\mathbf{V}_{cq}|^2$  values in Eqs. (30) and (33) we can also test the unitarity relation in the second row,

$$|\mathbf{V}_{cd}|^2 + |\mathbf{V}_{cs}|^2 + |\mathbf{V}_{cb}|^2 = 0.989 \pm 0.019, \quad (38)$$

and, adding the information on  $|\mathbf{V}_{tb}|$  in Eq. (36), the relation involving the third column,

$$|\mathbf{V}_{ub}|^2 + |\mathbf{V}_{cb}|^2 + |\mathbf{V}_{tb}|^2 = 1.040 \pm 0.051. \quad (39)$$

The ratio of the total hadronic decay width of the  $W$  to the leptonic one provides the sum [99, 100]

$$\sum_{j=d,s,b} \left( |\mathbf{V}_{uj}|^2 + |\mathbf{V}_{cj}|^2 \right) = 2.002 \pm 0.027, \quad (40)$$

which involves the first and second rows of the CKM matrix. Although much less precise than Eq. (37), these three results test unitarity at the 2%, 5% and 1.4% level, respectively.

From Eq. (40) one can also obtain an independent estimate of  $|\mathbf{V}_{cs}|$ , using the experimental knowledge on the other CKM matrix elements, i.e.,  $|\mathbf{V}_{ud}|^2 + |\mathbf{V}_{us}|^2 + |\mathbf{V}_{ub}|^2 + |\mathbf{V}_{cd}|^2 + |\mathbf{V}_{cb}|^2 = 1.0486 \pm 0.0018$ . This gives

$$|\mathbf{V}_{cs}| = 0.976 \pm 0.014, \quad (41)$$

which agrees with the slightly more accurate direct determination in Eq. (33).

The measured entries of the CKM matrix show a hierarchical pattern, with the diagonal elements being very close to one, the ones connecting the first two generations having a size

$$\lambda \approx |\mathbf{V}_{us}| = 0.2231 \pm 0.0007, \quad (42)$$

the mixing between the second and third families being of order  $\lambda^2$ , and the mixing between the first and third quark generations having a much smaller size of about  $\lambda^3$ . It is then quite practical to use the approximate parametrization [101]:

$$\mathbf{V} = \begin{bmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{bmatrix} + \mathcal{O}(\lambda^4), \quad (43)$$

where

$$A \approx \frac{|\mathbf{V}_{cb}|}{\lambda^2} = 0.844 \pm 0.014, \quad \sqrt{\rho^2 + \eta^2} \approx \left| \frac{\mathbf{V}_{ub}}{\lambda \mathbf{V}_{cb}} \right| = 0.422 \pm 0.035. \quad (44)$$

Defining to all orders in  $\lambda$  [102]  $s_{12} \equiv \lambda$ ,  $s_{23} \equiv A\lambda^2$  and  $s_{13} e^{-i\delta_{13}} \equiv A\lambda^3(\rho - i\eta)$ , Eq. (43) just corresponds to a Taylor expansion of Eq. (11) in powers of  $\lambda$ .

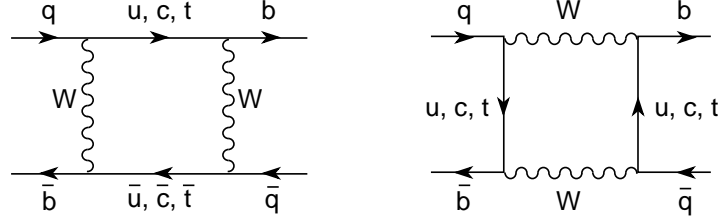
## 5 Meson-antimeson mixing

Additional information on the CKM parameters can be obtained from FCNC transitions, occurring at the one-loop level. An important example is provided by the mixing between the  $B_d^0$  meson and its antiparticle. This process occurs through the box diagrams shown in Fig. 5, where two  $W$  bosons are exchanged between a pair of quark lines. The mixing amplitude is proportional to

$$\langle \bar{B}_d^0 | \mathcal{H}_{\Delta B=2} | B_d^0 \rangle \sim \sum_{ij} \mathbf{V}_{id} \mathbf{V}_{ib}^* \mathbf{V}_{jd} \mathbf{V}_{jb}^* S(r_i, r_j) \sim \mathbf{V}_{td}^2 S(r_t, r_t), \quad (45)$$

where  $S(r_i, r_j)$  is a loop function [103] which depends on  $r_i \equiv m_i^2/M_W^2$ , with  $m_i$  the masses of the up-type quarks running along the internal fermionic lines. Owing to the unitarity of the CKM matrix, the mixing vanishes for equal (up-type) quark masses (GIM mechanism [8]); thus the flavour-changing transition is governed by the mass splittings between the  $u$ ,  $c$  and  $t$  quarks. Since the different CKM factors have all a similar size,  $\mathbf{V}_{ud} \mathbf{V}_{ub}^* \sim \mathbf{V}_{cd} \mathbf{V}_{cb}^* \sim \mathbf{V}_{td} \mathbf{V}_{tb}^* \sim A\lambda^3$ , the final amplitude is completely dominated by the top contribution. This transition can then be used to perform an indirect determination of  $\mathbf{V}_{td}$ .





**Fig. 5:** Box diagrams contributing to  $B_q^0 - \bar{B}_q^0$  mixing ( $q = d, s$ ).

Notice that this determination has a qualitatively different character than the ones obtained before from tree-level weak decays. Now, we are going to test the structure of the electroweak theory at the quantum level. This flavour-changing transition could then be sensitive to contributions from new physics at higher energy scales. Moreover, the mixing amplitude crucially depends on the unitarity of the CKM matrix. Without the GIM mechanism embodied in the CKM mixing structure, the calculation of the analogous  $K^0 \rightarrow \bar{K}^0$  transition (replace the  $b$  by a strange quark  $s$  in the box diagrams) would have failed to explain the observed  $K^0 - \bar{K}^0$  mixing by several orders of magnitude [104].

### 5.1 Mixing formalism

Since weak interactions can transform a  $P^0$  state ( $P = K, D, B$ ) into its antiparticle  $\bar{P}^0$ , these flavour eigenstates are not mass eigenstates and do not follow an exponential decay law. Let us consider an arbitrary mixture of the two flavour states,

$$|\psi(t)\rangle = a(t) |P^0\rangle + b(t) |\bar{P}^0\rangle \equiv \begin{pmatrix} a(t) \\ b(t) \end{pmatrix}, \quad (46)$$

with the time evolution (in the meson rest frame)

$$i \frac{d}{dt} |\psi(t)\rangle = \mathcal{M} |\psi(t)\rangle. \quad (47)$$

Assuming  $\mathcal{CPT}$  symmetry to hold, the  $2 \times 2$  mixing matrix can be written as

$$\mathcal{M} = \begin{pmatrix} M & M_{12} \\ M_{12}^* & M \end{pmatrix} - \frac{i}{2} \begin{pmatrix} \Gamma & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma \end{pmatrix}. \quad (48)$$

The diagonal elements  $M$  and  $\Gamma$  are real parameters, which would correspond to the mass and width of the neutral mesons in the absence of mixing. The off-diagonal entries contain the  $\Delta F = 2$  transition amplitude ( $F = S, C, B$ ):

$$M_{12} - \frac{i}{2} \Gamma_{12} = \frac{1}{2M} \left\{ \langle P^0 | \mathcal{H}_{\Delta F=2}(0) | \bar{P}^0 \rangle - \frac{1}{2} \int d^4x \langle P^0 | T(\mathcal{H}_{\Delta F=1}(x) \mathcal{H}_{\Delta F=1}(0)) | \bar{P}^0 \rangle \right\}. \quad (49)$$

In addition to the short-distance  $\Delta F = 2$  Hamiltonian generated by the box diagrams, the mixing amplitude receives non-local contributions involving two  $\Delta F = 1$  transitions, which contain both *dispersive* and *absorptive* components, contributing to  $M_{12}$  and  $\Gamma_{12}$ , respectively. The *absorptive* contribution  $\Gamma_{12}$  arises from on-shell intermediate states:

$$\Gamma_{12} = \frac{1}{2M} (2\pi)^4 \sum_n \delta^{(4)}(p_{P^0} - p_n) \langle P^0 | \mathcal{H}_{\Delta F=1}(0) | n \rangle \langle n | \mathcal{H}_{\Delta F=1}(0) | \bar{P}^0 \rangle. \quad (50)$$

The sum extends over all possible intermediate states  $|n\rangle$  into which the  $|\bar{P}^0\rangle$  and  $|P^0\rangle$  can both decay:  $\bar{P}^0 \rightarrow f \rightarrow P^0$ . In the SM, the  $\Delta F = 1$  Hamiltonian is generated through a single  $W^\pm$  emission,

as shown in Fig. 4 for charm decay. If  $\mathcal{CP}$  were an exact symmetry,  $M_{12}$  and  $\Gamma_{12}$  would also be real parameters.

The physical eigenstates of  $\mathcal{M}$  are

$$|P_{\mp}\rangle = \frac{1}{\sqrt{|p|^2 + |q|^2}} \left[ p |P^0\rangle \mp q |\bar{P}^0\rangle \right], \quad (51)$$

with

$$\frac{q}{p} \equiv \frac{1 - \bar{\varepsilon}}{1 + \bar{\varepsilon}} = \left( \frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}} \right)^{1/2}. \quad (52)$$

The corresponding eigenvalues are

$$M_{P_{\mp}} - \frac{i}{2}\Gamma_{P_{\mp}} = \left( M \mp \frac{1}{2}\Delta M \right) - \frac{i}{2} \left( \Gamma \mp \frac{1}{2}\Delta\Gamma \right), \quad (53)$$

where<sup>3</sup>  $\Delta M \equiv M_{P_+} - M_{P_-}$  and  $\Delta\Gamma \equiv \Gamma_{P_+} - \Gamma_{P_-}$  satisfy

$$(\Delta M)^2 - \frac{1}{4}(\Delta\Gamma)^2 = 4|M_{12}|^2 - |\Gamma_{12}|^2 \quad (54)$$

and

$$\Delta M \Delta\Gamma = 4 \operatorname{Re}(M_{12}\Gamma_{12}^*) = 4|M_{12}||\Gamma_{12}| \cos\phi, \quad (55)$$

with  $\phi \equiv \arg(-M_{12}/\Gamma_{12})$ .

If  $M_{12}$  and  $\Gamma_{12}$  were real then  $q/p = 1$  and the mass eigenstates  $|P_{\mp}\rangle$  would correspond to the  $\mathcal{CP}$ -even and  $\mathcal{CP}$ -odd states (we use the phase convention<sup>4</sup>  $\mathcal{CP}|P^0\rangle = -|\bar{P}^0\rangle$ )

$$|P_{1,2}\rangle \equiv \frac{1}{\sqrt{2}} \left( |P^0\rangle \mp |\bar{P}^0\rangle \right), \quad \mathcal{CP}|P_{1,2}\rangle = \pm |P_{1,2}\rangle. \quad (56)$$

The two mass eigenstates are no longer orthogonal when  $\mathcal{CP}$  is violated:

$$\langle P_- | P_+ \rangle = \frac{|p|^2 - |q|^2}{|p|^2 + |q|^2} = \frac{2 \operatorname{Re}(\bar{\varepsilon})}{(1 + |\bar{\varepsilon}|^2)}. \quad (57)$$

The time evolution of a state which was originally produced as a  $P^0$  or a  $\bar{P}^0$  is given by

$$\begin{pmatrix} |P^0(t)\rangle \\ |\bar{P}^0(t)\rangle \end{pmatrix} = \begin{pmatrix} g_1(t) & \frac{q}{p}g_2(t) \\ \frac{p}{q}g_2(t) & g_1(t) \end{pmatrix} \begin{pmatrix} |P^0\rangle \\ |\bar{P}^0\rangle \end{pmatrix}, \quad (58)$$

where

$$\begin{pmatrix} g_1(t) \\ g_2(t) \end{pmatrix} = e^{-iMt} e^{-\Gamma t/2} \begin{pmatrix} \cos[(x - iy)\Gamma t/2] \\ -i \sin[(x - iy)\Gamma t/2] \end{pmatrix}, \quad (59)$$

with

$$x \equiv \frac{\Delta M}{\Gamma}, \quad y \equiv \frac{\Delta\Gamma}{2\Gamma}. \quad (60)$$

<sup>3</sup>Be aware of the different sign conventions in the literature. Quite often,  $\Delta M$  and  $\Delta\Gamma$  are defined to be positive.

<sup>4</sup>Since flavour is conserved by strong interactions, there is some freedom in defining the phases of flavour eigenstates. One could use  $|P_{\zeta}^0\rangle \equiv e^{-i\zeta}|P^0\rangle$  and  $|\bar{P}_{\zeta}^0\rangle \equiv e^{i\zeta}|\bar{P}^0\rangle$ , which satisfy  $\mathcal{CP}|P_{\zeta}^0\rangle = -e^{-2i\zeta}|\bar{P}_{\zeta}^0\rangle$ . Both basis are trivially related:  $M_{12}^{\zeta} = e^{2i\zeta}M_{12}$ ,  $\Gamma_{12}^{\zeta} = e^{2i\zeta}\Gamma_{12}$  and  $(q/p)_{\zeta} = e^{-2i\zeta}(q/p)$ . Thus,  $q/p \neq 1$  does not necessarily imply  $\mathcal{CP}$  violation.  $\mathcal{CP}$  is violated if  $|q/p| \neq 1$ ; i.e.,  $\operatorname{Re}(\bar{\varepsilon}) \neq 0$  and  $\langle P_- | P_+ \rangle \neq 0$ . Note that  $\langle P_- | P_+ \rangle_{\zeta} = \langle P_- | P_+ \rangle$ . Another phase-convention-independent quantity is  $(q/p)(\bar{A}_f/A_f)$ , where  $A_f \equiv A(P^0 \rightarrow f)$  and  $\bar{A}_f \equiv -A(\bar{P}^0 \rightarrow f)$ , for any final state  $f$ .

## 5.2 Experimental measurements

The main difference between the  $K^0-\bar{K}^0$  and  $B^0-\bar{B}^0$  systems stems from the different kinematics involved. The light kaon mass only allows the hadronic decay modes  $K^0 \rightarrow 2\pi$  and  $K^0 \rightarrow 3\pi$ . Since  $\mathcal{CP}|\pi\pi\rangle = +|\pi\pi\rangle$ , for both  $\pi^0\pi^0$  and  $\pi^+\pi^-$  final states, the  $\mathcal{CP}$ -even kaon state decays into  $2\pi$  whereas the  $\mathcal{CP}$ -odd one decays into the phase-space-suppressed  $3\pi$  mode. Therefore, there is a large lifetime difference and we have a short-lived  $|K_S\rangle \equiv |K_-\rangle \approx |K_1\rangle + \bar{\varepsilon}_K|K_2\rangle$  and a long-lived  $|K_L\rangle \equiv |K_+\rangle \approx |K_2\rangle + \bar{\varepsilon}_K|K_1\rangle$  kaon, with  $\Gamma_{K_L} \ll \Gamma_{K_S} \approx 2\Gamma_{K^0}$ . One finds experimentally that  $\Delta\Gamma_{K^0} \approx -\Gamma_{K_S} \approx -2\Delta M_{K^0}$  [9]:

$$\Delta M_{K^0} = (0.5293 \pm 0.0009) \cdot 10^{10} \text{ s}^{-1}, \quad \Delta\Gamma_{K^0} = -(1.1149 \pm 0.0005) \cdot 10^{10} \text{ s}^{-1}. \quad (61)$$

Thus, the two  $K^0-\bar{K}^0$  oscillations parameters are sizeable and of similar magnitudes:  $x_{K^0} \approx -y_{K^0} \approx 1$ .

In the  $B$  system, there are many open decay channels and a large part of them are common to both mass eigenstates. Therefore, the  $|B_{\mp}\rangle$  states have a similar lifetime; i.e.,  $|\Delta\Gamma_{B^0}| \ll \Gamma_{B^0}$ . Moreover, whereas the  $B^0-\bar{B}^0$  transition is dominated by the top box diagram, the decay amplitudes get obviously their main contribution from the  $b \rightarrow c$  process. Thus,  $|\Delta\Gamma_{B^0}/\Delta M_{B^0}| \sim m_b^2/m_t^2 \ll 1$ . To experimentally measure the mixing transition requires the identification of the  $B$ -meson flavour at both its production and decay time. This can be done through flavour-specific decays such as  $B^0 \rightarrow X\ell^+\nu_\ell$  and  $\bar{B}^0 \rightarrow X\ell^-\bar{\nu}_\ell$ , where the lepton charge labels the initial  $B$  meson. In general, mixing is measured by studying pairs of  $B$  mesons so that one  $B$  can be used to *tag* the initial flavour of the other meson. For instance, in  $e^+e^-$  machines one can look into the pair production process  $e^+e^- \rightarrow B^0\bar{B}^0 \rightarrow (X\ell\nu_\ell)(Y\ell\nu_\ell)$ . In the absence of mixing, the final leptons should have opposite charges; the amount of like-sign leptons is then a clear signature of meson mixing.

Evidence for a large  $B_d^0-\bar{B}_d^0$  mixing was first reported in 1987 by ARGUS [105]. This provided the first indication that the top quark was very heavy. Since then, many experiments have analysed the mixing probability. The present world-average values are [9, 46]:

$$\Delta M_{B_d^0} = (0.5065 \pm 0.0019) \cdot 10^{12} \text{ s}^{-1}, \quad x_{B_d^0} = 0.769 \pm 0.004, \quad (62)$$

while  $y_{B_d^0} = 0.001 \pm 0.005$  confirms the expected suppression of  $\Delta\Gamma_{B_d^0}$ .

The first direct evidence of  $B_s^0-\bar{B}_s^0$  oscillations was obtained by CDF [106]. The large measured mass difference reflects the CKM hierarchy  $|\mathbf{V}_{ts}|^2 \gg |\mathbf{V}_{td}|^2$ , implying very fast oscillations [9, 46]:

$$\begin{aligned} \Delta M_{B_s^0} &= (17.757 \pm 0.021) \cdot 10^{12} \text{ s}^{-1}, & x_{B_s^0} &= 26.81 \pm 0.08, \\ \Delta\Gamma_{B_s^0} &= -(0.090 \pm 0.005) \cdot 10^{12} \text{ s}^{-1}, & y_{B_s^0} &= -0.068 \pm 0.004. \end{aligned} \quad (63)$$

Evidence of mixing has been also obtained in the  $D^0-\bar{D}^0$  system. The present world averages [46],

$$x_{D^0} = -\left(0.39^{+0.11}_{-0.12}\right) \cdot 10^{-2}, \quad y_{D^0} = -\left(0.65^{+0.06}_{-0.07}\right) \cdot 10^{-2}, \quad (64)$$

confirm the SM expectation of a very slow oscillation, compared with the decay rate. Since the short-distance mixing amplitude originates in box diagrams with down-type quarks in the internal lines, it is very suppressed by the relevant combination of CKM factors and quark masses.

## 5.3 Mixing constraints on the CKM matrix

Long-distance contributions arising from intermediate hadronic states completely dominate the  $D^0-\bar{D}^0$  mixing amplitude and are very sizeable for  $\Delta M_{K^0}$ , making difficult to extract useful information on the CKM matrix. The situation is much better for  $B^0$  mesons, owing to the dominance of the short-distance

top contribution which is known to next-to-leading order (NLO) in the strong coupling [107, 108]. The main uncertainty stems from the hadronic matrix element of the  $\Delta B = 2$  four-quark operator

$$\langle \bar{B}_d^0 | (\bar{b}\gamma^\mu(1-\gamma_5)d) (\bar{b}\gamma_\mu(1-\gamma_5)d) | B_d^0 \rangle \equiv \frac{8}{3} M_{B^0}^2 \xi_B^2, \quad (65)$$

which is characterized through the non-perturbative parameter  $\xi_B(\mu) \equiv f_B \sqrt{B_B(\mu)}$  [109]. The current  $(2+1)$  lattice averages [39] are  $\hat{\xi}_{B_d} = (225 \pm 9)$  MeV,  $\hat{\xi}_{B_s} = (274 \pm 8)$  MeV and  $\hat{\xi}_{B_s}/\hat{\xi}_{B_d} = 1.206 \pm 0.017$ , where  $\hat{\xi}_B \approx \alpha_s(\mu)^{-3/23} \xi_B(\mu)$  is the corresponding renormalization-group-invariant quantity. Using these values, the measured mass differences in Eq. (62) and Eq. (63) imply

$$|\mathbf{V}_{tb}^* \mathbf{V}_{td}| = 0.0080 \pm 0.0003, \quad |\mathbf{V}_{tb}^* \mathbf{V}_{ts}| = 0.0388 \pm 0.0012, \quad \frac{|\mathbf{V}_{td}|}{|\mathbf{V}_{ts}|} = 0.205 \pm 0.003. \quad (66)$$

The last number takes advantage of the smaller uncertainty in the ratio  $\hat{\xi}_{B_s}/\hat{\xi}_{B_d}$ . Since  $|\mathbf{V}_{tb}| \approx 1$ , the mixing of  $B_{d,s}^0$  mesons provides indirect determinations of  $|\mathbf{V}_{td}|$  and  $|\mathbf{V}_{ts}|$ . The resulting value of  $|\mathbf{V}_{ts}|$  is in agreement with Eq. (30), satisfying the unitarity constraint  $|\mathbf{V}_{ts}| \approx |\mathbf{V}_{cb}|$ . In terms of the  $(\rho, \eta)$  parametrization of Eq. (43), one obtains

$$\sqrt{(1-\rho)^2 + \eta^2} = \begin{cases} \left| \frac{\mathbf{V}_{td}}{\lambda \mathbf{V}_{cb}} \right| = 0.86 \pm 0.04 \\ \left| \frac{\mathbf{V}_{td}}{\lambda \mathbf{V}_{ts}} \right| = 0.920 \pm 0.013 \end{cases}. \quad (67)$$

## 6 $\mathcal{CP}$ violation

While parity ( $\mathcal{P}$ ) and charge conjugation ( $\mathcal{C}$ ) are violated by the weak interactions in a maximal way, the product of the two discrete transformations is still a good symmetry of the gauge interactions (left-handed fermions  $\leftrightarrow$  right-handed antifermions). In fact,  $\mathcal{CP}$  appears to be a symmetry of nearly all observed phenomena. However, a slight violation of the  $\mathcal{CP}$  symmetry at the level of 0.2% is observed in the neutral kaon system and more sizeable signals of  $\mathcal{CP}$  violation have been established at the  $B$  factories. Moreover, the huge matter–antimatter asymmetry present in our Universe is a clear manifestation of  $\mathcal{CP}$  violation and its important role in the primordial baryogenesis.

The  $\mathcal{CPT}$  theorem guarantees that the product of the three discrete transformations is an exact symmetry of any local and Lorentz-invariant quantum field theory, preserving micro-causality. A violation of  $\mathcal{CP}$  implies then a corresponding violation of time reversal ( $\mathcal{T}$ ). Since  $\mathcal{T}$  is an antiunitary transformation, this requires the presence of relative complex phases between different interfering amplitudes.

The electroweak SM Lagrangian only contains a single complex phase  $\delta_{13}(\eta)$ . This is the sole possible source of  $\mathcal{CP}$  violation and, therefore, the SM predictions for  $\mathcal{CP}$ -violating phenomena are quite constrained. The CKM mechanism requires several necessary conditions in order to generate an observable  $\mathcal{CP}$ -violation effect. With only two fermion generations, the quark mixing matrix cannot give rise to  $\mathcal{CP}$  violation; therefore, for  $\mathcal{CP}$  violation to occur in a particular process, all three generations are required to play an active role. In the kaon system, for instance,  $\mathcal{CP}$  violation can only appear at the one-loop level, where the top quark is present. In addition, all CKM matrix elements must be non-zero and the quarks of a given charge must be non-degenerate in mass. If any of these conditions were not satisfied, the CKM phase could be rotated away by a redefinition of the quark fields.  $\mathcal{CP}$ -violation effects are then necessarily proportional to the product of all CKM angles, and should vanish in the limit where any two (equal-charge) quark masses are taken to be equal. All these necessary conditions can be summarized as a single requirement on the original quark mass matrices  $\mathbf{M}'_u$  and  $\mathbf{M}'_d$  [110]:

$$\mathcal{CP} \text{ violation} \quad \iff \quad \text{Im} \left\{ \det \left[ \mathbf{M}'_u \mathbf{M}'_u{}^\dagger, \mathbf{M}'_d \mathbf{M}'_d{}^\dagger \right] \right\} \neq 0. \quad (68)$$

Without performing any detailed calculation, one can make the following general statements on the implications of the CKM mechanism of  $\mathcal{CP}$  violation:

- Owing to unitarity, for any choice of  $i, j, k, l$  (between 1 and 3),

$$\text{Im} [\mathbf{V}_{ij} \mathbf{V}_{ik}^* \mathbf{V}_{lk} \mathbf{V}_{lj}^*] = \mathcal{J} \sum_{m,n=1}^3 \epsilon_{ilm} \epsilon_{jkn}, \quad (69)$$

$$\mathcal{J} = c_{12} c_{23} c_{13}^2 s_{12} s_{23} s_{13} \sin \delta_{13} \approx A^2 \lambda^6 \eta < 10^{-4}. \quad (70)$$

Any  $\mathcal{CP}$ -violation observable involves the product  $\mathcal{J}$  [110]. Thus, violations of the  $\mathcal{CP}$  symmetry are necessarily small.

- In order to have sizeable  $\mathcal{CP}$ -violating asymmetries  $\mathcal{A} \equiv (\Gamma - \bar{\Gamma})/(\Gamma + \bar{\Gamma})$ , one should look for very suppressed decays, where the decay widths already involve small CKM matrix elements.
- In the SM,  $\mathcal{CP}$  violation is a low-energy phenomenon, in the sense that any effect should disappear when the quark mass difference  $m_c - m_u$  becomes negligible.
- $B$  decays are the optimal place for  $\mathcal{CP}$ -violation signals to show up. They involve small CKM matrix elements and are the lowest-mass processes where the three quark generations play a direct (tree-level) role.

The SM mechanism of  $\mathcal{CP}$  violation is based on the unitarity of the CKM matrix. Testing the constraints implied by unitarity is then a way to test the source of  $\mathcal{CP}$  violation. The unitarity tests in Eqs. (37), (38), (39) and (40) involve only the moduli of the CKM parameters, while  $\mathcal{CP}$  violation has to do with their phases. More interesting are the off-diagonal unitarity conditions:

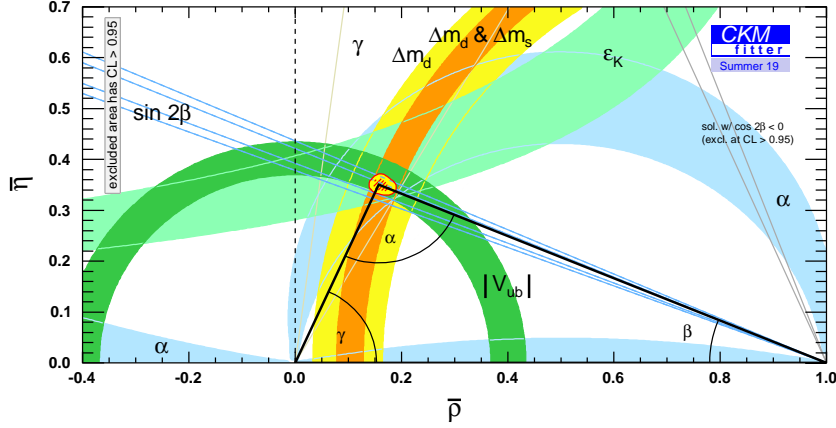
$$\begin{aligned} \mathbf{V}_{ud}^* \mathbf{V}_{us} + \mathbf{V}_{cd}^* \mathbf{V}_{cs} + \mathbf{V}_{td}^* \mathbf{V}_{ts} &= 0, \\ \mathbf{V}_{us}^* \mathbf{V}_{ub} + \mathbf{V}_{cs}^* \mathbf{V}_{cb} + \mathbf{V}_{ts}^* \mathbf{V}_{tb} &= 0, \\ \mathbf{V}_{ub}^* \mathbf{V}_{ud} + \mathbf{V}_{cb}^* \mathbf{V}_{cd} + \mathbf{V}_{tb}^* \mathbf{V}_{td} &= 0. \end{aligned} \quad (71)$$

These relations can be visualized by triangles in a complex plane which, owing to Eq. (69), have the same area  $|\mathcal{J}|/2$ . In the absence of  $\mathcal{CP}$  violation, these triangles would degenerate into segments along the real axis.

In the first two triangles, one side is much shorter than the other two (the Cabibbo suppression factors of the three sides are  $\lambda$ ,  $\lambda$  and  $\lambda^5$  in the first triangle, and  $\lambda^4$ ,  $\lambda^2$  and  $\lambda^2$  in the second one). This is why  $\mathcal{CP}$  effects are so small for  $K$  mesons (first triangle), and why certain asymmetries in  $B_s^0$  decays are predicted to be tiny (second triangle). The third triangle looks more interesting, since the three sides have a similar size of about  $\lambda^3$ . They are small, which means that the relevant  $b$ -decay branching ratios are small, but once enough  $B_d^0$  mesons have been produced, the  $\mathcal{CP}$ -violation asymmetries are sizeable. The present experimental constraints on this triangle are shown in Fig. 6, where it has been scaled by dividing its sides by  $\mathbf{V}_{cb}^* \mathbf{V}_{cd}$ . This aligns one side of the triangle along the real axis and makes its length equal to 1; the coordinates of the 3 vertices are then  $(0, 0)$ ,  $(1, 0)$  and  $(\bar{\rho}, \bar{\eta}) \approx (1 - \lambda^2/2)(\rho, \eta)$ .

We have already determined the sides of the unitarity triangle in Eqs. (44) and (67), through two  $\mathcal{CP}$ -conserving observables:  $|\mathbf{V}_{ub}/\mathbf{V}_{cb}|$  and  $B_{d,s}^0$  mixing. This gives the circular rings shown in Fig. 6, centered at the vertices  $(0, 0)$  and  $(1, 0)$ . Their overlap at  $\eta \neq 0$  establishes that  $\mathcal{CP}$  is violated (assuming unitarity). More direct constraints on the parameter  $\eta$  can be obtained from  $\mathcal{CP}$ -violating observables, which provide sensitivity to the angles of the unitarity triangle ( $\alpha + \beta + \gamma = \pi$ ):

$$\alpha \equiv \arg \left[ -\frac{\mathbf{V}_{td} \mathbf{V}_{tb}^*}{\mathbf{V}_{ud} \mathbf{V}_{ub}^*} \right], \quad \beta \equiv \arg \left[ -\frac{\mathbf{V}_{cd} \mathbf{V}_{cb}^*}{\mathbf{V}_{td} \mathbf{V}_{tb}^*} \right], \quad \gamma \equiv \arg \left[ -\frac{\mathbf{V}_{ud} \mathbf{V}_{ub}^*}{\mathbf{V}_{cd} \mathbf{V}_{cb}^*} \right]. \quad (72)$$



**Fig. 6:** Experimental constraints on the SM unitarity triangle [111]. (Copyright CKMfitter group, reused with permission.)

### 6.1 Indirect and direct $\mathcal{CP}$ violation in the kaon system

Any observable  $\mathcal{CP}$ -violation effect is generated by the interference between different amplitudes contributing to the same physical transition. This interference can occur either through meson-antimeson mixing or via final-state interactions, or by a combination of both effects.

The flavour-specific decays  $K^0 \rightarrow \pi^- \ell^+ \nu_\ell$  and  $\bar{K}^0 \rightarrow \pi^+ \ell^- \bar{\nu}_\ell$  provide a way to measure the departure of the  $K^0$ - $\bar{K}^0$  mixing parameter  $|p/q|$  from unity. In the SM, the decay amplitudes satisfy  $|A(\bar{K}^0 \rightarrow \pi^+ \ell^- \bar{\nu}_\ell)| = |A(K^0 \rightarrow \pi^- \ell^+ \nu_\ell)|$ ; therefore ( $\ell = e, \mu$ ),

$$A_L \equiv \frac{\Gamma(K_L \rightarrow \pi^- \ell^+ \nu_\ell) - \Gamma(K_L \rightarrow \pi^+ \ell^- \bar{\nu}_\ell)}{\Gamma(K_L \rightarrow \pi^- \ell^+ \nu_\ell) + \Gamma(K_L \rightarrow \pi^+ \ell^- \bar{\nu}_\ell)} = \frac{|p|^2 - |q|^2}{|p|^2 + |q|^2} = \frac{2 \operatorname{Re}(\bar{\varepsilon}_K)}{(1 + |\bar{\varepsilon}_K|^2)}. \quad (73)$$

The experimental measurement [9],  $A_L = (3.32 \pm 0.06) \cdot 10^{-3}$ , implies

$$\operatorname{Re}(\bar{\varepsilon}_K) = (1.66 \pm 0.03) \cdot 10^{-3}, \quad (74)$$

which establishes the presence of *indirect*  $\mathcal{CP}$  violation generated by the mixing amplitude.

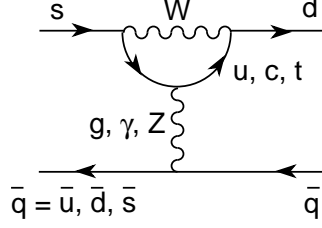
If the flavour of the decaying meson  $P$  is known, any observed difference between the decay rate  $\Gamma(P \rightarrow f)$  and its  $\mathcal{CP}$  conjugate  $\Gamma(\bar{P} \rightarrow \bar{f})$  would indicate that  $\mathcal{CP}$  is directly violated in the decay amplitude. One could study, for instance,  $\mathcal{CP}$  asymmetries in decays such as  $K^\pm \rightarrow \pi^\pm \pi^0$  where the pion charges identify the kaon flavour; however, no positive signals have been found in charged kaon decays. Since at least two interfering contributions are needed, let us write the decay amplitudes as

$$A[P \rightarrow f] = M_1 e^{i\phi_1} e^{i\delta_1} + M_2 e^{i\phi_2} e^{i\delta_2}, \quad A[\bar{P} \rightarrow \bar{f}] = M_1 e^{-i\phi_1} e^{i\delta_1} + M_2 e^{-i\phi_2} e^{i\delta_2}, \quad (75)$$

where  $\phi_i$  denote weak phases,  $\delta_i$  strong final-state interaction phases and  $M_i$  the moduli of the matrix elements. Notice that the weak phase gets reversed under  $\mathcal{CP}$ , while the strong one remains of course invariant. The rate asymmetry is given by

$$\mathcal{A}_{P \rightarrow f}^{\mathcal{CP}} \equiv \frac{\Gamma[P \rightarrow f] - \Gamma[\bar{P} \rightarrow \bar{f}]}{\Gamma[P \rightarrow f] + \Gamma[\bar{P} \rightarrow \bar{f}]} = \frac{-2M_1 M_2 \sin(\phi_1 - \phi_2) \sin(\delta_1 - \delta_2)}{|M_1|^2 + |M_2|^2 + 2M_1 M_2 \cos(\phi_1 - \phi_2) \cos(\delta_1 - \delta_2)}. \quad (76)$$

Thus, to generate a direct  $\mathcal{CP}$  asymmetry one needs: 1) at least two interfering amplitudes, which should be of comparable size in order to get a sizeable asymmetry; 2) two different weak phases [ $\sin(\phi_1 - \phi_2) \neq 0$ ], and 3) two different strong phases [ $\sin(\delta_1 - \delta_2) \neq 0$ ].


**Fig. 7:**  $\Delta S = 1$  penguin diagrams.

Direct  $\mathcal{CP}$  violation has been searched for in decays of neutral kaons, where  $K^0-\bar{K}^0$  mixing is also involved. Thus, both direct and indirect  $\mathcal{CP}$  violation need to be taken into account simultaneously. A  $\mathcal{CP}$ -violation signal is provided by the ratios:

$$\eta_{+-} \equiv \frac{A(K_L \rightarrow \pi^+ \pi^-)}{A(K_S \rightarrow \pi^+ \pi^-)} = \varepsilon_K + \varepsilon'_K, \quad \eta_{00} \equiv \frac{A(K_L \rightarrow \pi^0 \pi^0)}{A(K_S \rightarrow \pi^0 \pi^0)} = \varepsilon_K - 2\varepsilon'_K. \quad (77)$$

The dominant effect from  $\mathcal{CP}$  violation in  $K^0-\bar{K}^0$  mixing is contained in  $\varepsilon_K$ , while  $\varepsilon'_K$  accounts for direct  $\mathcal{CP}$  violation in the decay amplitudes [42]:

$$\varepsilon_K = \bar{\varepsilon}_K + i\xi_0, \quad \varepsilon'_K = \frac{i}{\sqrt{2}} \omega (\xi_2 - \xi_0), \quad \omega \equiv \frac{\text{Re}(A_2)}{\text{Re}(A_0)} e^{i(\delta_2 - \delta_0)}, \quad \xi_I \equiv \frac{\text{Im}(A_I)}{\text{Re}(A_I)}. \quad (78)$$

$A_I$  are the transition amplitudes into two pions with isospin  $I = 0, 2$  (these are the only two values allowed by Bose symmetry for the final  $2\pi$  state) and  $\delta_I$  their corresponding strong phase shifts. Although  $\varepsilon'_K$  is strongly suppressed by the small ratio  $|\omega| \approx 1/22$ , a non-zero value has been established through very accurate measurements, demonstrating the existence of direct  $\mathcal{CP}$  violation in K decays [112–115]:

$$\text{Re}(\varepsilon'_K/\varepsilon_K) = \frac{1}{3} \left( 1 - \left| \frac{\eta_{00}}{\eta_{+-}} \right| \right) = (16.6 \pm 2.3) \cdot 10^{-4}. \quad (79)$$

In the SM the necessary weak phases are generated through the gluonic and electroweak penguin diagrams shown in Fig. 7, involving virtual up-type quarks of the three generations in the loop. These short-distance contributions are known to NLO in the strong coupling [116, 117]. However, the theoretical prediction involves a delicate balance between the two isospin amplitudes and is sensitive to long-distance and isospin-violating effects. Using chiral perturbation theory techniques, one finds  $\text{Re}(\varepsilon'_K/\varepsilon_K) = (14 \pm 5) \cdot 10^{-4}$  [118–122], in agreement with Eq. (79) but with a large uncertainty.<sup>5</sup>

Since  $\text{Re}(\varepsilon'_K/\varepsilon_K) \ll 1$ , the ratios  $\eta_{+-}$  and  $\eta_{00}$  provide a measurement of  $\varepsilon_K = |\varepsilon_K| e^{i\phi_\varepsilon}$  [9]:

$$|\varepsilon_K| = \frac{1}{3} (2|\eta_{+-}| + |\eta_{00}|) = (2.228 \pm 0.011) \cdot 10^{-3}, \quad \phi_\varepsilon = (43.52 \pm 0.05)^\circ, \quad (80)$$

in perfect agreement with the semileptonic asymmetry  $A_L$ . In the SM  $\varepsilon_K$  receives short-distance contributions from box diagrams involving virtual top and charm quarks, which are proportional to

$$\varepsilon_K \propto \sum_{i,j=c,t} \eta_{ij} \text{Im}[\mathbf{V}_{id} \mathbf{V}_{is}^* \mathbf{V}_{jd} \mathbf{V}_{js}^*] S(r_i, r_j) \propto A^2 \lambda^6 \bar{\eta} \left\{ \eta_{tt} A^2 \lambda^4 (1 - \bar{\rho}) + P_c \right\}. \quad (81)$$

The first term shows the CKM dependence of the dominant top contribution,  $P_c$  accounts for the charm corrections [124] and the short-distance QCD corrections  $\eta_{ij}$  are known to NLO [107, 108, 125]. The measured value of  $|\varepsilon_K|$  determines an hyperbolic constraint in the  $(\bar{\rho}, \bar{\eta})$  plane, shown in Fig. 6, taking into account the theoretical uncertainty in the hadronic matrix element of the  $\Delta S = 2$  operator [39].

<sup>5</sup>A very recent lattice calculation gives  $\text{Re}(\varepsilon'_K/\varepsilon_K) = (22 \pm 8) \cdot 10^{-4}$  [123], with an even larger error. However, this result does not include yet important isospin-breaking corrections that are known to be negative [122].

## 6.2 $\mathcal{CP}$ asymmetries in B decays

The semileptonic decays  $B_q^0 \rightarrow X^- \ell^+ \nu_\ell$  and  $\bar{B}_q^0 \rightarrow X^+ \ell^- \bar{\nu}_\ell$  ( $q = d, s$ ) provide the most direct way to measure the amount of  $\mathcal{CP}$  violation in the  $B^0$ - $\bar{B}^0$  mixing matrix, through

$$\begin{aligned} a_{\text{sl}}^q &\equiv \frac{\Gamma(\bar{B}_q^0 \rightarrow X^- \ell^+ \nu_\ell) - \Gamma(B_q^0 \rightarrow X^+ \ell^- \bar{\nu}_\ell)}{\Gamma(\bar{B}_q^0 \rightarrow X^- \ell^+ \nu_\ell) + \Gamma(B_q^0 \rightarrow X^+ \ell^- \bar{\nu}_\ell)} = \frac{|p|^4 - |q|^4}{|p|^4 + |q|^4} \approx 4 \operatorname{Re}(\bar{\varepsilon}_{B_q^0}) \\ &\approx \frac{|\Gamma_{12}|}{|M_{12}|} \sin \phi_q \approx \frac{|\Delta\Gamma_{B_q^0}|}{|\Delta M_{B_q^0}|} \tan \phi_q. \end{aligned} \quad (82)$$

This asymmetry is expected to be tiny because  $|\Gamma_{12}/M_{12}| \sim m_b^2/m_t^2 \ll 1$ . Moreover, there is an additional GIM suppression in the relative mixing phase  $\phi_q \equiv \arg(-M_{12}/\Gamma_{12}) \sim (m_c^2 - m_u^2)/m_b^2$ , implying a value of  $|q/p|$  very close to 1. Therefore,  $a_{\text{sl}}^q$  could be very sensitive to new sources of  $\mathcal{CP}$  violation beyond the SM, contributing to  $\phi_q$ . The present measurements give [9, 46]

$$\operatorname{Re}(\bar{\varepsilon}_{B_d^0}) = (-0.5 \pm 0.4) \cdot 10^{-3}, \quad \operatorname{Re}(\bar{\varepsilon}_{B_s^0}) = (-0.15 \pm 0.70) \cdot 10^{-3}. \quad (83)$$

The large  $B^0$ - $\bar{B}^0$  mixing provides a different way to generate the required  $\mathcal{CP}$ -violating interference. There are quite a few nonleptonic final states which are reachable both from a  $B^0$  and a  $\bar{B}^0$ . For these flavour non-specific decays the  $B^0$  (or  $\bar{B}^0$ ) can decay directly to the given final state  $f$ , or do it after the meson has been changed to its antiparticle via the mixing process; i.e., there are two different amplitudes,  $A(B^0 \rightarrow f)$  and  $A(B^0 \rightarrow \bar{B}^0 \rightarrow f)$ , corresponding to two possible decay paths.  $\mathcal{CP}$ -violating effects can then result from the interference of these two contributions.

The time-dependent decay probabilities for the decay of a neutral  $B$  meson created at the time  $t_0 = 0$  as a pure  $B^0$  ( $\bar{B}^0$ ) into the final state  $f$  ( $\bar{f} \equiv \mathcal{CP} f$ ) are:

$$\begin{aligned} \Gamma[B^0(t) \rightarrow f] &\propto \frac{1}{2} e^{-\Gamma_{B^0} t} \left( |A_f|^2 + |\bar{A}_f|^2 \right) \left\{ 1 + C_f \cos(\Delta M_{B^0} t) - S_f \sin(\Delta M_{B^0} t) \right\}, \\ \Gamma[\bar{B}^0(t) \rightarrow \bar{f}] &\propto \frac{1}{2} e^{-\Gamma_{B^0} t} \left( |\bar{A}_{\bar{f}}|^2 + |A_{\bar{f}}|^2 \right) \left\{ 1 - C_{\bar{f}} \cos(\Delta M_{B^0} t) + S_{\bar{f}} \sin(\Delta M_{B^0} t) \right\}, \end{aligned} \quad (84)$$

where the tiny  $\Delta\Gamma_{B^0}$  corrections have been neglected and we have introduced the notation

$$\begin{aligned} A_f &\equiv A[B^0 \rightarrow f], & \bar{A}_f &\equiv -A[\bar{B}^0 \rightarrow f], & \bar{\rho}_f &\equiv \bar{A}_f/A_f, \\ A_{\bar{f}} &\equiv A[B^0 \rightarrow \bar{f}], & \bar{A}_{\bar{f}} &\equiv -A[\bar{B}^0 \rightarrow \bar{f}], & \rho_{\bar{f}} &\equiv A_{\bar{f}}/\bar{A}_{\bar{f}}, \\ C_f &\equiv \frac{1 - |\bar{\rho}_f|^2}{1 + |\bar{\rho}_f|^2}, & S_f &\equiv \frac{2 \operatorname{Im}\left(\frac{q}{p} \bar{\rho}_f\right)}{1 + |\bar{\rho}_f|^2}, & C_{\bar{f}} &\equiv -\frac{1 - |\rho_{\bar{f}}|^2}{1 + |\rho_{\bar{f}}|^2}, & S_{\bar{f}} &\equiv \frac{-2 \operatorname{Im}\left(\frac{p}{q} \rho_{\bar{f}}\right)}{1 + |\rho_{\bar{f}}|^2}. \end{aligned} \quad (85)$$

$\mathcal{CP}$  invariance demands the probabilities of  $\mathcal{CP}$ -conjugate processes to be identical. Thus,  $\mathcal{CP}$  conservation requires  $A_f = \bar{A}_{\bar{f}}$ ,  $A_{\bar{f}} = \bar{A}_f$ ,  $\bar{\rho}_f = \rho_{\bar{f}}$  and  $\operatorname{Im}\left(\frac{q}{p} \bar{\rho}_f\right) = \operatorname{Im}\left(\frac{p}{q} \rho_{\bar{f}}\right)$ , i.e.,  $C_f = -C_{\bar{f}}$  and  $S_f = -S_{\bar{f}}$ . Violation of any of the first three equalities would be a signal of direct  $\mathcal{CP}$  violation. The fourth equality tests  $\mathcal{CP}$  violation generated by the interference of the direct decay  $B^0 \rightarrow f$  and the mixing-induced decay  $B^0 \rightarrow \bar{B}^0 \rightarrow f$ .

For  $B^0$  mesons

$$\frac{q}{p} \Big|_{B_q^0} \approx \sqrt{\frac{M_{12}^*}{M_{12}}} \approx \frac{\mathbf{V}_{tb}^* \mathbf{V}_{tq}}{\mathbf{V}_{tb} \mathbf{V}_{tq}^*} \equiv e^{-2i\phi_q^M}, \quad (86)$$



where  $\phi_d^M = \beta + \mathcal{O}(\lambda^4)$  and  $\phi_s^M = -\beta_s + \mathcal{O}(\lambda^6)$ . The angle  $\beta$  is defined in Eq. (72), while  $\beta_s \equiv \arg[-(\mathbf{V}_{ts}\mathbf{V}_{tb}^*)/(\mathbf{V}_{cs}\mathbf{V}_{cb}^*)] = \lambda^2\eta + \mathcal{O}(\lambda^4)$  is the equivalent angle in the  $B_s^0$  unitarity triangle, which is predicted to be tiny. Therefore, the mixing ratio  $q/p$  is given by a known weak phase.

An obvious example of final states  $f$  which can be reached both from the  $B^0$  and the  $\bar{B}^0$  are  $\mathcal{CP}$  eigenstates; i.e., states such that  $\bar{f} = \zeta_f f$  ( $\zeta_f = \pm 1$ ). In this case,  $A_{\bar{f}} = \zeta_f A_f$ ,  $\bar{A}_{\bar{f}} = \zeta_f \bar{A}_f$ ,  $\rho_{\bar{f}} = 1/\bar{\rho}_f$ ,  $C_{\bar{f}} = C_f$  and  $S_{\bar{f}} = S_f$ . A non-zero value of  $C_f$  or  $S_f$  signals then  $\mathcal{CP}$  violation. The ratios  $\bar{\rho}_f$  and  $\rho_{\bar{f}}$  depend in general on the underlying strong dynamics. However, for  $\mathcal{CP}$  self-conjugate final states, all dependence on the strong interaction disappears if only one weak amplitude contributes to the  $B^0 \rightarrow f$  and  $\bar{B}^0 \rightarrow f$  transitions [126, 127]. In this case, we can write the decay amplitude as  $A_f = M e^{i\phi^D} e^{i\delta_s}$ , with  $M = M^*$  and  $\phi^D$  and  $\delta_s$  weak and strong phases. The ratios  $\bar{\rho}_f$  and  $\rho_{\bar{f}}$  are then given by

$$\rho_{\bar{f}} = \bar{\rho}_f^* = \zeta_f e^{2i\phi^D}. \quad (87)$$

The modulus  $M$  and the unwanted strong phase cancel out completely from these two ratios;  $\rho_{\bar{f}}$  and  $\bar{\rho}_f$  simplify to a single weak phase, associated with the underlying weak quark transition. Since  $|\rho_{\bar{f}}| = |\bar{\rho}_f| = 1$ , the time-dependent decay probabilities become much simpler. In particular,  $C_f = 0$  and there is no longer any dependence on  $\cos(\Delta M_{B^0} t)$ . Moreover, the coefficients of the sinusoidal terms are then fully known in terms of CKM mixing angles only:  $S_f = S_{\bar{f}} = -\zeta_f \sin[2(\phi_q^M + \phi^D)] \equiv -\zeta_f \sin(2\Phi)$ . In this ideal case, the time-dependent  $\mathcal{CP}$ -violating decay asymmetry

$$A_{B^0 \rightarrow \bar{f}}^{\mathcal{CP}} \equiv \frac{\Gamma[\bar{B}^0(t) \rightarrow \bar{f}] - \Gamma[B^0(t) \rightarrow f]}{\Gamma[\bar{B}^0(t) \rightarrow \bar{f}] + \Gamma[B^0(t) \rightarrow f]} = -\zeta_f \sin(2\Phi) \sin(\Delta M_{B^0} t) \quad (88)$$

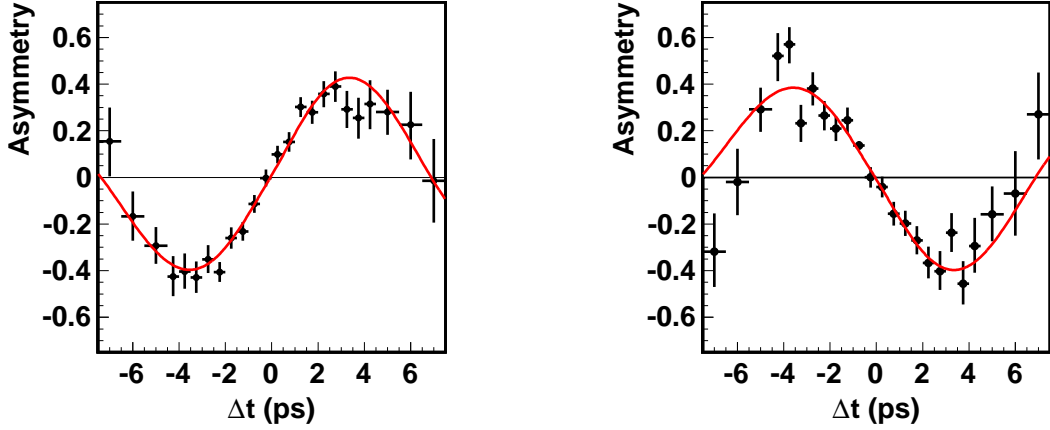
provides a direct and clean measurement of the CKM parameters [128].

When several decay amplitudes with different phases contribute,  $|\bar{\rho}_f| \neq 1$  and the interference term will depend both on CKM parameters and on the strong dynamics embodied in  $\bar{\rho}_f$ . The leading contributions to  $\bar{b} \rightarrow \bar{q}' q' \bar{q}$  are either the tree-level  $W$  exchange or penguin topologies generated by gluon ( $\gamma$ ,  $Z$ ) exchange. Although of higher order in the strong (electroweak) coupling, penguin amplitudes are logarithmically enhanced by the virtual  $W$  loop and are potentially competitive. Table 3 contains the CKM factors associated with the two topologies for different  $B$  decay modes into  $\mathcal{CP}$  eigenstates.

**Table 3:** CKM factors and relevant angle  $\Phi$  for some  $B$  decays into  $\mathcal{CP}$  eigenstates.

Decay	Tree-level CKM	Penguin CKM	Exclusive channels	$\Phi$
$\bar{b} \rightarrow \bar{c} c \bar{s}$	$A\lambda^2$	$-A\lambda^2$	$B_d^0 \rightarrow J/\psi K_S, J/\psi K_L$ $B_s^0 \rightarrow D_s^+ D_s^-, J/\psi \phi$	$\beta$ $-\beta_s$
$\bar{b} \rightarrow \bar{s} s \bar{s}$		$-A\lambda^2$	$B_d^0 \rightarrow K_S \phi, K_L \phi$ $B_s^0 \rightarrow \phi \phi$	$\beta$ $-\beta_s$
$\bar{b} \rightarrow \bar{d} d \bar{s}$		$-A\lambda^2$	$B_s^0 \rightarrow K_S K_S, K_L K_L$	$-\beta_s$
$\bar{b} \rightarrow \bar{c} c \bar{d}$	$-A\lambda^3$	$A\lambda^3(1 - \rho - i\eta)$	$B_d^0 \rightarrow D^+ D^-, J/\psi \pi^0$ $B_s^0 \rightarrow J/\psi K_S, J/\psi K_L$	$\approx \beta$ $\approx -\beta_s$
$\bar{b} \rightarrow \bar{u} u \bar{d}$	$A\lambda^3(\rho + i\eta)$	$A\lambda^3(1 - \rho - i\eta)$	$B_d^0 \rightarrow \pi^+ \pi^-, \rho^0 \pi^0, \omega \pi^0$ $B_s^0 \rightarrow \rho^0 K_{S,L}, \omega K_{S,L}, \pi^0 K_{S,L}$	$\approx \beta + \gamma$ $\neq \gamma - \beta_s$
$\bar{b} \rightarrow \bar{s} s \bar{d}$		$A\lambda^3(1 - \rho - i\eta)$	$B_d^0 \rightarrow K_S K_S, K_L K_L, \phi \pi^0$ $B_s^0 \rightarrow K_S \phi, K_L \phi$	$0$ $-\beta - \beta_s$

The gold-plated decay mode is  $B_d^0 \rightarrow J/\psi K_S$ . In addition of having a clean experimental signature, the two topologies have the same (zero) weak phase. The  $\mathcal{CP}$  asymmetry provides then a clean



**Fig. 8:** Time-dependent asymmetries for  $\mathcal{CP}$ -odd ( $B_d^0 \rightarrow J/\psi K_S$ ,  $B_d^0 \rightarrow \psi' K_S$ ,  $B_d^0 \rightarrow \chi_{c1} K_S$ ;  $\zeta_f = -1$ ; left) and  $\mathcal{CP}$ -even ( $B_d^0 \rightarrow J/\psi K_L$ ;  $\zeta_f = +1$ ; right) final states, measured by Belle [129].

measurement of the mixing angle  $\beta$ , without strong-interaction uncertainties. Fig. 8 shows the Belle measurement [129] of time-dependent  $\bar{b} \rightarrow c\bar{c}\bar{s}$  asymmetries for  $\mathcal{CP}$ -odd ( $B_d^0 \rightarrow J/\psi K_S$ ,  $B_d^0 \rightarrow \psi' K_S$ ,  $B_d^0 \rightarrow \chi_{c1} K_S$ ) and  $\mathcal{CP}$ -even ( $B_d^0 \rightarrow J/\psi K_L$ ) final states. A very nice oscillation is manifest, with opposite signs for the two different choices of  $\zeta_f = \pm 1$ . Including the information obtained from other  $\bar{b} \rightarrow c\bar{c}\bar{s}$  decays, one gets the world average [46]:

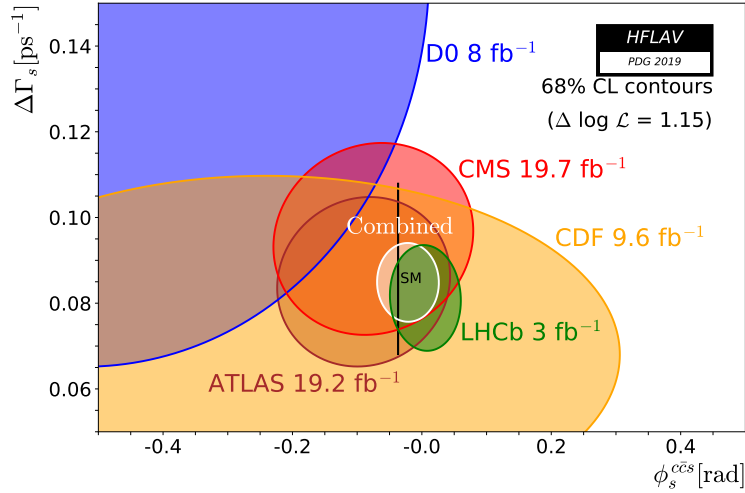
$$\sin(2\beta) = 0.699 \pm 0.017. \quad (89)$$

Fitting an additional  $\cos(\Delta M_{B^0} t)$  term in the measured asymmetries results in  $C_f = -0.005 \pm 0.015$  [46], confirming the expected null result. An independent measurement of  $\sin 2\beta$  can be obtained from  $\bar{b} \rightarrow s\bar{s}\bar{s}$  and  $\bar{b} \rightarrow d\bar{d}\bar{s}$  decays, which only receive penguin contributions and, therefore, could be more sensitive to new-physics corrections in the loop diagram. These modes give  $\sin(2\beta) = 0.648 \pm 0.038$  [46], in good agreement with (89).

Eq. (89) determines the angle  $\beta$  up to a four-fold ambiguity:  $\beta$ ,  $\frac{\pi}{2} - \beta$ ,  $\pi + \beta$  and  $\frac{3\pi}{2} - \beta$ . The solution  $\beta = (22.2 \pm 0.7)^\circ$  is in remarkable agreement with the other phenomenological constraints on the unitarity triangle in Fig. 6. The ambiguity has been resolved through a time-dependent analysis of the Dalitz plot distribution in  $B_d^0 \rightarrow D^{(*)} h^0$  decays ( $h^0 = \pi^0, \eta, \omega$ ), showing that  $\cos(2\beta) = 0.91 \pm 0.25$  and  $\beta = (22.5 \pm 4.6)^\circ$  [130]. This proves that  $\cos(2\beta)$  is positive with a  $3.7\sigma$  significance, while the multifold solution  $\frac{\pi}{2} - \beta = (67.8 \pm 0.7)^\circ$  is excluded at the  $7.3\sigma$  level.

A determination of  $\beta + \gamma = \pi - \alpha$  can be obtained from  $\bar{b} \rightarrow \bar{u}u\bar{d}$  decays, such as  $B_d^0 \rightarrow \pi\pi$  or  $B_d^0 \rightarrow \rho\rho$ . However, the penguin contamination that carries a different weak phase can be sizeable. The time-dependent asymmetry in  $B_d^0 \rightarrow \pi^+\pi^-$  shows indeed a non-zero value for the  $\cos(\Delta M_{B^0} t)$  term,  $C_f = -0.32 \pm 0.04$  [46], providing evidence of direct  $\mathcal{CP}$  violation and indicating the presence of an additional decay amplitude; therefore,  $S_f = -0.63 \pm 0.04 \neq \sin 2\alpha$ . One could still extract useful information on  $\alpha$  (up to 16 mirror solutions), using the isospin relations among the  $B_d^0 \rightarrow \pi^+\pi^-$ ,  $B_d^0 \rightarrow \pi^0\pi^0$  and  $B^+ \rightarrow \pi^+\pi^0$  amplitudes and their  $\mathcal{CP}$  conjugates [131]; however, only a loose constraint is obtained, given the limited experimental precision on  $B_d^0 \rightarrow \pi^0\pi^0$ . Much stronger constraints are obtained from  $B_d^0 \rightarrow \rho^+\rho^-, \rho^0\rho^0$  because one can use the additional polarization information of two vectors in the final state to resolve the different contributions and, moreover, the small branching fraction  $\text{Br}(B_d^0 \rightarrow \rho^0\rho^0) = (9.6 \pm 1.5) \cdot 10^{-7}$  [9] implies a very small penguin contribution. Additional information can be obtained from  $B_d^0, \bar{B}_d^0 \rightarrow \rho^\pm\pi^\mp, a_1^\pm\pi^\mp$ , although the final states are not  $\mathcal{CP}$  eigenstates. Combining all pieces of information results in [46]

$$\alpha = (84.9_{-4.5}^{+5.1})^\circ. \quad (90)$$



**Fig. 9:** 68% CL regions in  $\Delta\Gamma_{B_s^0}$  and  $\phi_s^{ccs}$ , extracted from  $\bar{b} \rightarrow \bar{c}c\bar{s}$   $\mathcal{CP}$  asymmetries of  $B_s^0$  mesons [46]. The vertical black line shows the SM prediction [135–137].

The angle  $\gamma$  cannot be determined in  $\bar{b} \rightarrow u\bar{u}\bar{d}$  decays such as  $B_s^0 \rightarrow \rho^0 K_S$  because the colour factors in the hadronic matrix element enhance the penguin amplitude with respect to the tree-level contribution. Instead,  $\gamma$  can be measured through the tree-level decays  $B \rightarrow D^{(*)}K^{(*)}$  ( $\bar{b} \rightarrow \bar{u}c\bar{s}$ ) and  $B \rightarrow \bar{D}^{(*)}K^{(*)}$  ( $\bar{b} \rightarrow \bar{c}u\bar{s}$ ), using final states accessible in both  $D^{(*)0}$  and  $\bar{D}^{(*)0}$  decays and playing with the interference of both amplitudes [132–134]. The sensitivity can be optimized with Dalitz-plot analyses of the  $D^0, \bar{D}^0$  decay products. The extensive experimental studies performed so far result in [46]

$$\gamma = (71.1^{+4.6}_{-5.3})^\circ. \quad (91)$$

Another ambiguous solution with  $\gamma \leftrightarrow \gamma + \pi$  also exists.

Mixing-induced  $\mathcal{CP}$  violation has been also searched for in the decays of  $B_s^0$  and  $\bar{B}_s^0$  mesons into  $J/\psi\phi, \psi(2S)\phi, J/\psi K^+K^-, J/\psi\pi^+\pi^-$  and  $D_s^+D_s^-$ . From the corresponding time-dependent  $\mathcal{CP}$  asymmetries,<sup>6</sup> one extracts correlated constraints on  $\Delta\Gamma_s$  and the weak phase  $\phi_s^{ccs} \equiv 2\Phi_s^{ccs} \approx 2\phi_s^M \approx -2\beta_s$  in Eq. (88), which are shown in Fig. 9. They lead to [46]

$$\beta_s = (0.60 \pm 0.89)^\circ, \quad (92)$$

in good agreement with the SM prediction  $\beta_s \approx \eta\lambda^2 \approx 1^\circ$ .

### 6.3 Global fit of the unitarity triangle

The CKM parameters can be more accurately determined through a global fit to all available measurements, imposing the unitarity constraints and taking properly into account the theoretical uncertainties. The global fit shown in Fig. 6 uses frequentist statistics and gives [111]

$$\lambda = 0.22484^{+0.00025}_{-0.00006}, \quad A = 0.824^{+0.006}_{-0.015}, \quad \bar{\rho} = 0.157^{+0.010}_{-0.006}, \quad \bar{\eta} = 0.350^{+0.008}_{-0.007}. \quad (93)$$

This implies  $\mathcal{J} = (3.06^{+0.07}_{-0.08}) \cdot 10^{-5}$ ,  $\alpha = (91.7^{+1.7}_{-1.1})^\circ$ ,  $\beta = (22.6^{+0.5}_{-0.4})^\circ$  and  $\gamma = (65.8^{+0.9}_{-1.3})^\circ$ . Similar results are obtained by the UTfit group [138], using instead a Bayesian approach and a slightly different treatment of theoretical uncertainties.

<sup>6</sup>The  $\Delta\Gamma_{B_s^0}$  corrections to Eq. (84) must be taken into account at the current level of precision.

## 6.4 Direct $\mathcal{CP}$ violation in $B$ decays

The big data samples accumulated at the  $B$  factories and the collider experiments have established the presence of direct  $\mathcal{CP}$  violation in several decays of  $B$  mesons. The most significant signals are [9, 46]

$$\begin{aligned}
\mathcal{A}_{\bar{B}_d^0 \rightarrow K^- \pi^+}^{\mathcal{CP}} &= -0.083 \pm 0.004, & \mathcal{A}_{\bar{B}_d^0 \rightarrow \bar{K}^{*0} \eta}^{\mathcal{CP}} &= 0.19 \pm 0.05, & \mathcal{A}_{\bar{B}_d^0 \rightarrow K^{*-} \pi^+}^{\mathcal{CP}} &= -0.27 \pm 0.04, \\
C_{B_d^0 \rightarrow \pi^+ \pi^-} &= -0.32 \pm 0.04, & \gamma_{B \rightarrow D^{(*)} K^{(*)}} &= (71.1_{-5.3}^{+4.6})^\circ, & \mathcal{A}_{\bar{B}_s^0 \rightarrow K^+ \pi^-}^{\mathcal{CP}} &= 0.221 \pm 0.015, \\
\mathcal{A}_{B^- \rightarrow K^- \rho^0}^{\mathcal{CP}} &= 0.37 \pm 0.10, & \mathcal{A}_{B^- \rightarrow K^- \eta}^{\mathcal{CP}} &= -0.37 \pm 0.08, & \mathcal{A}_{B^- \rightarrow \pi^- \pi^+ \pi^-}^{\mathcal{CP}} &= 0.057 \pm 0.013, \\
\mathcal{A}_{B^- \rightarrow K^- K^+ K^-}^{\mathcal{CP}} &= -0.033 \pm 0.008, & \mathcal{A}_{B^- \rightarrow K^- K^+ \pi^-}^{\mathcal{CP}} &= -0.122 \pm 0.021, \\
\mathcal{A}_{B^- \rightarrow K^- f_2(1270)}^{\mathcal{CP}} &= -0.68_{-0.17}^{+0.19}, & \mathcal{A}_{B^- \rightarrow \pi^- f_2(1270)}^{\mathcal{CP}} &= 0.40 \pm 0.06, & & (94)
\end{aligned}$$

Another prominent observation of direct  $\mathcal{CP}$  violation has been done recently in charm decays [139]:

$$\mathcal{A}_{D^0 \rightarrow K^- K^+}^{\text{direct } \mathcal{CP}} - \mathcal{A}_{D^0 \rightarrow \pi^- \pi^+}^{\text{direct } \mathcal{CP}} = (-15.7 \pm 2.9) \cdot 10^{-4}, \quad (95)$$

where the small contribution from  $D^0$ - $\bar{D}^0$  mixing has been subtracted, using the measured difference of reconstructed mean decay times of the two modes. Unfortunately, owing to the unavoidable presence of strong phases, a real theoretical understanding of the corresponding SM predictions is still lacking. Progress in this direction is needed to perform meaningful tests of the CKM mechanism of  $\mathcal{CP}$  violation and pin down any possible effects from new physics beyond the SM framework.

## 7 Rare decays

Complementary and very valuable information can be obtained from rare decays that in the SM are strongly suppressed by the GIM mechanism. These processes are then sensitive to new-physics contributions with a different flavour structure. Well-known examples are the  $\bar{K}^0 \rightarrow \mu^+ \mu^-$  decay modes [9, 140],

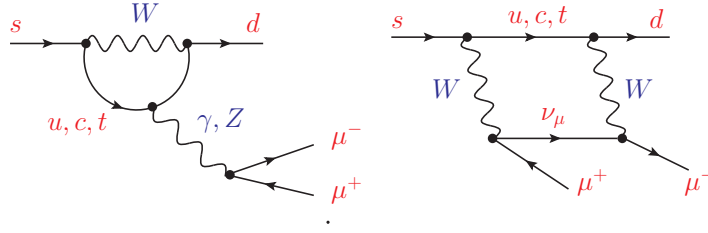
$$\text{Br}(K_L \rightarrow \mu^+ \mu^-) = (6.84 \pm 0.11) \cdot 10^{-9}, \quad \text{Br}(K_S \rightarrow \mu^+ \mu^-) < 2.1 \cdot 10^{-10} \quad (90\% \text{ CL}), \quad (96)$$

which tightly constrain any hypothetical flavour-changing ( $s \rightarrow d$ ) tree-level coupling of the  $Z$  boson. In the SM, these decays receive short-distance contributions from the penguin and box diagrams displayed in Fig. 10. Owing to the unitarity of the CKM matrix, the resulting amplitude vanishes for equal up-type quark masses, which entails a heavy suppression:

$$\mathcal{M} \propto \sum_{i=u,c,t} \mathbf{V}_{is} \mathbf{V}_{id}^* F(m_i^2/M_W^2) = \mathbf{V}_{cs} \mathbf{V}_{cd}^* \tilde{F}(m_c^2/M_W^2) + \mathbf{V}_{ts} \mathbf{V}_{td}^* \tilde{F}(m_t^2/M_W^2), \quad (97)$$

where  $F(x)$  is a loop function and  $\tilde{F}(x) \equiv F(x) - F(0)$ . In addition to the loop factor  $g^4/(16\pi^2)$ , the charm contribution carries a mass suppression  $\lambda m_c^2/M_W^2$ , while the top term is proportional to  $\lambda^5 A^2(1 - \rho + i\eta) m_t^2/M_W^2$ . The large top mass compensates the strong Cabibbo suppression so that the top contribution is finally larger than the charm one. However, the total short-distance contribution to the  $K_L$  decay,  $\text{Br}(K_L \rightarrow \mu^+ \mu^-)_{\text{sd}} = (0.79 \pm 0.12) \cdot 10^{-9}$  [124, 141], is nearly one order of magnitude below the experimental measurement (96), while  $\text{Br}(K_S \rightarrow \mu^+ \mu^-)_{\text{sd}} \approx 1.7 \cdot 10^{-13}$  [42].

The decays  $K_L \rightarrow \mu^+ \mu^-$  and  $K_S \rightarrow \mu^+ \mu^-$  are actually dominated by long-distance contributions, through  $\bar{K}_L \rightarrow \pi^0, \eta, \eta' \rightarrow \gamma\gamma \rightarrow \mu^+ \mu^-$  and  $K_S \rightarrow \pi^+ \pi^- \rightarrow \gamma\gamma \rightarrow \mu^+ \mu^-$ . The absorptive component from two on-shell intermediate photons nearly saturates the measured  $\text{Br}(K_L \rightarrow \mu^+ \mu^-)$  [142], while the  $K_S$  decay rate is predicted to be  $\text{Br}(K_S \rightarrow \mu^+ \mu^-) = 5.1 \cdot 10^{-12}$  [42, 143]. These decays can be rigorously analysed with chiral perturbation theory techniques [42], but the strong suppression of their short-distance contributions does not make possible to extract useful information on the



**Fig. 10:** Short-distance penguin (left) and box (right) SM contributions to  $\bar{K}^0 \rightarrow \mu^+ \mu^-$ .

CKM parameters. Nevertheless, it is possible to predict the longitudinal polarization  $P_L$  of either muon in the  $K_L$  decay, a  $\mathcal{CP}$ -violating observable which in the SM is dominated by indirect  $\mathcal{CP}$  violation from  $K^0 - \bar{K}^0$  mixing:  $|P_L| = (2.6 \pm 0.4) \cdot 10^{-3}$  [143].

Other interesting kaon decay modes such as  $K^0 \rightarrow \gamma\gamma$ ,  $K \rightarrow \pi\gamma\gamma$  and  $K \rightarrow \pi\ell^+\ell^-$  are also governed by long-distance physics [42, 144]. Of particular interest is the decay  $K_L \rightarrow \pi^0 e^+ e^-$ , since it is sensitive to new sources of  $\mathcal{CP}$  violation. At lowest order in  $\alpha$  the decay proceeds through  $K_2^0 \rightarrow \pi^0 \gamma^*$  that violates  $\mathcal{CP}$ , while the  $\mathcal{CP}$ -conserving contribution through  $K_L^0 \rightarrow \pi^0 \gamma^* \gamma^*$  is suppressed by an additional power of  $\alpha$  and it is found to be below  $10^{-12}$  [42]. The  $K_L \rightarrow \pi^0 e^+ e^-$  transition is then dominated by the  $\mathcal{O}(\alpha)$   $\mathcal{CP}$ -violating contributions [144], both from  $K^0 - \bar{K}^0$  mixing and direct  $\mathcal{CP}$  violation. The estimated rate  $\text{Br}(K_L \rightarrow \pi^0 e^+ e^-) = (3.1 \pm 0.9) \cdot 10^{-11}$  [42, 145, 146] is only a factor ten smaller than the present (90% CL) upper bound of  $2.8 \cdot 10^{-10}$  [147] and should be reachable in the near future.

The decays  $K^\pm \rightarrow \pi^\pm \nu \bar{\nu}$  and  $K_L \rightarrow \pi^0 \nu \bar{\nu}$  provide a more direct access to CKM information because long-distance effects play a negligible role. The decay amplitudes are dominated by  $Z$ -penguin and  $W$ -box loop diagrams of the type shown in Fig. 10 (replacing the final muons by neutrinos), and are proportional to the hadronic  $\bar{K}\pi$  matrix element of the  $\Delta S = 1$  vector current, which (assuming isospin symmetry) is extracted from  $K \rightarrow \pi \ell \nu_\ell$  decays. The small long-distance and isospin-violating corrections can be estimated with chiral perturbation theory. The neutral decay is  $\mathcal{CP}$  violating and proceeds almost entirely through direct  $\mathcal{CP}$  violation (via interference with mixing). Taking the CKM inputs from other observables, the predicted SM rates are [148–150]:

$$\text{Br}(K^+ \rightarrow \pi^+ \nu \bar{\nu})_{\text{th}} = (8.5 \pm 0.6) \cdot 10^{-11}, \quad \text{Br}(K_L \rightarrow \pi^0 \nu \bar{\nu})_{\text{th}} = (2.9 \pm 0.3) \cdot 10^{-11}. \quad (98)$$

The uncertainties are largely parametrical, due to CKM input, the charm and top masses and  $\alpha_s(M_Z)$ . On the experimental side, the current upper bounds on the charged [151] and neutral [152] modes are

$$\text{Br}(K^+ \rightarrow \pi^+ \nu \bar{\nu}) < 1.85 \cdot 10^{-10}, \quad \text{Br}(K_L \rightarrow \pi^0 \nu \bar{\nu}) < 3.0 \cdot 10^{-9} \quad (90\% \text{ CL}). \quad (99)$$

The ongoing CERN NA62 experiment aims to reach  $\mathcal{O}(100)$   $K^+ \rightarrow \pi^+ \nu \bar{\nu}$  events (assuming SM rates), while increased sensitivities on the  $K_L \rightarrow \pi^0 \nu \bar{\nu}$  mode are expected to be achieved by the KOTO experiment at J-PARC. These experiments will start to seriously probe the new-physics potential of these decays.

The inclusive decay  $\bar{B} \rightarrow X_s \gamma$  provides another powerful test of the SM flavour structure at the quantum loop level. It proceeds through a  $b \rightarrow s\gamma$  penguin diagram with an on-shell photon. The present experimental world average,  $\text{Br}(\bar{B} \rightarrow X_s \gamma)_{E_\gamma \geq 1.6 \text{ GeV}} = (3.32 \pm 0.15) \cdot 10^{-4}$  [46], agrees very well with the SM theoretical prediction,  $\text{Br}(\bar{B} \rightarrow X_s \gamma)_{E_\gamma \geq 1.6 \text{ GeV}}^{\text{th}} = (3.40 \pm 0.17) \cdot 10^{-4}$  [153–155], obtained at the next-to-next-to-leading order.

The LHC experiments have recently reached the SM sensitivity for the  $B_d^0$  and  $B_s^0$  decays into

$\mu^+\mu^-$  pairs [156–158]. The current world averages [9],<sup>7</sup>

$$\overline{\text{Br}}(B_d^0 \rightarrow \mu^+\mu^-) = (1.4_{-1.4}^{+1.6}) \cdot 10^{-10}, \quad \overline{\text{Br}}(B_s^0 \rightarrow \mu^+\mu^-) = (3.0 \pm 0.4) \cdot 10^{-9}, \quad (100)$$

agree with the SM predictions  $\overline{\text{Br}}(B_d^0 \rightarrow \mu^+\mu^-) = (1.06 \pm 0.09) \cdot 10^{-10}$  and  $\overline{\text{Br}}(B_s^0 \rightarrow \mu^+\mu^-) = (3.65 \pm 0.23) \cdot 10^{-9}$  [160]. Other interesting FCNC decays with  $B$  mesons are  $\bar{B} \rightarrow K^{(*)}l^+l^-$  and  $\bar{B} \rightarrow K^{(*)}\nu\bar{\nu}$  [86].

## 8 Flavour constraints on new physics

The CKM matrix provides a very successful description of flavour phenomena, as it is clearly exhibited in the unitarity test of Fig. 6, showing how very different observables converge into a single choice of CKM parameters. This is a quite robust and impressive result. One can perform separate tests with different subsets of measurements, according to their  $\mathcal{CP}$ -conserving or  $\mathcal{CP}$ -violating nature, or splitting them into tree-level and loop-induced processes. In all cases, one finally gets a closed triangle and similar values for the fitted CKM entries [111, 138]. However, the SM mechanism of flavour mixing and  $\mathcal{CP}$  violation is conceptually quite unsatisfactory because it does not provide any dynamical understanding of the numerical values of fermion masses, and mixings. We completely ignore the reasons why the fermion spectrum contains such a hierarchy of different masses, spanning many orders of magnitude, and which fundamental dynamics is behind the existence of three flavour generations and their observed mixing structure.

The phenomenological success of the SM puts severe constraints on possible scenarios of new physics. The absence of any clear signals of new phenomena in the LHC searches is pushing the hypothetical new-physics scale at higher energies, above the TeV. The low-energy implications of new dynamics beyond the SM can then be analysed in terms of an effective Lagrangian, containing only the known SM fields:

$$\mathcal{L}_{\text{eff}} = L_{\text{SM}} + \sum_{d>4} \sum_k \frac{c_k^{(d)}}{\Lambda_{\text{NP}}^{d-4}} O_k^{(d)}. \quad (101)$$

The effective Lagrangian is organised as an expansion in terms of dimension- $d$  operators  $O_k^{(d)}$ , invariant under the SM gauge group, suppressed by the corresponding powers of the new-physics scale  $\Lambda_{\text{NP}}$ . The dimensionless couplings  $c_k^{(d)}$  encode information on the underlying dynamics. The lowest-order term in this dimensional expansion is the SM Lagrangian that contains all allowed operators of dimension 4.

At low energies, the terms with lower dimensions dominate the physical transition amplitudes. There is only one operator with  $d = 5$  (up to Hermitian conjugation and flavour assignments), which violates lepton number by two units and is then related with the possible existence of neutrino Majorana masses [161]. Taking  $m_\nu \gtrsim 0.05$  eV, one estimates a very large lepton-number-violating scale  $\Lambda_{\text{NP}}/c^{(5)} \lesssim 10^{15}$  GeV [17]. Assuming lepton-number conservation, the first signals of new phenomena should be associated with  $d = 6$  operators.

One can easily analyse the possible impact of  $\Delta F = 2$  ( $F = S, C, B$ ) four-quark operators ( $d = 6$ ), such as the SM left-left operator in Eq. (65) that induces a  $\Delta B = 2$  transition. Since the SM box diagram provides an excellent description of the data, hypothetical new-physics contributions can only be tolerated within the current uncertainties, which puts stringent upper bounds on the corresponding Wilson coefficients  $\tilde{c}_k \equiv c_k^{(6)}/\Lambda_{\text{NP}}^2$ . For instance,  $\Delta M_{B_d^0}$  and  $S_{J/\psi K_S}$  imply that the real ( $\mathcal{CP}$ -conserving) and imaginary ( $\mathcal{CP}$ -violating) parts of  $\tilde{c}_k$  must be below  $10^{-6}$   $\text{TeV}^{-2}$  for a  $(\bar{b}_L \gamma^\mu d_L)^2$  operator, and nearly one order of magnitude smaller ( $10^{-7}$ ) for  $(\bar{b}_R d_L)(\bar{b}_L d_R)$  [162]. Stronger bounds are obtained in the kaon system from  $\Delta M_{K^0}$  and  $\varepsilon_K$ . For the  $(\bar{s}_L \gamma^\mu d_L)^2$  operator the real (imaginary) coefficient must be below  $10^{-6}$  ( $3 \cdot 10^{-9}$ ), while the corresponding bounds for  $(\bar{s}_R d_L)(\bar{s}_L d_R)$  are  $7 \cdot 10^{-9}$  ( $3 \cdot 10^{-11}$ ), in the

<sup>7</sup> $\overline{\text{Br}}$  is the time-integrated branching ratio, which for  $B_s^0$  is slightly affected by the sizeable value of  $\Delta\Gamma_s$  [159].

same  $\text{TeV}^{-2}$  units [162]. Taking the coefficients  $c_k^{(6)} \sim \mathcal{O}(1)$ , this would imply  $\Lambda_{\text{NP}} > 3 \cdot 10^3 \text{ TeV}$  for  $B_d^0$  and  $\Lambda_{\text{NP}} > 3 \cdot 10^5 \text{ TeV}$  for  $K^0$ . Therefore, two relevant messages emerge from the data:

1. A generic flavour structure with coefficients  $c_k^{(6)} \sim \mathcal{O}(1)$  is completely ruled out at the TeV scale.
2. New flavour-changing physics at  $\Lambda_{\text{NP}} \sim 1 \text{ TeV}$  could only be possible if the corresponding Wilson coefficients  $c_k^{(6)}$  inherit the strong SM suppressions generated by the GIM mechanism.

The last requirement can be satisfied by assuming that the up and down Yukawa matrices are the only sources of quark-flavour symmetry breaking (minimal flavour violation) [163–165]. In the absence of Yukawa interactions, the SM Lagrangian has a  $\mathcal{G} \equiv U(3)_{L_L} \otimes U(3)_{Q_L} \otimes U(3)_{\ell_R} \otimes U(3)_{u_R} \otimes U(3)_{d_R}$  global flavour symmetry, because one can rotate arbitrarily in the 3-generation space each one of the five SM fermion components in Eq. (2). The Yukawa matrices are the only explicit breakings of this large symmetry group. Assuming that the new physics does not introduce any additional breakings of the flavour symmetry  $\mathcal{G}$  (beyond insertions of Yukawa matrices), one can easily comply with the flavour bounds. Otherwise, flavour data provide very strong constraints on models with additional sources of flavour symmetry breaking and probe physics at energy scales not directly accessible at accelerators.

The subtle SM cancellations suppressing FCNC transitions could be easily destroyed in the presence of new physics contributions. To better appreciate the non-generic nature of the flavour structure, let us analyse the simplest extension of the SM scalar sector with a second Higgs doublet, which increases the number of quark Yukawas:

$$\mathcal{L}_Y = - \sum_{a=1}^2 \left\{ \bar{Q}'_L \mathcal{Y}_d^{(a)'} \phi_a d'_R + \bar{Q}'_L \mathcal{Y}_u^{(a)'} \phi_a^c u'_R + \bar{L}'_L \mathcal{Y}_\ell^{(a)'} \phi_a \ell'_R \right\} + \text{h.c.}, \quad (102)$$

where  $\phi_a^T = (\phi_a^{(+)}, \phi_a^{(0)})$  are the two scalar doublets,  $\phi_a^c$  their  $\mathcal{C}$ -conjugate fields, and  $Q'_L$  and  $L'_L$  the left-handed quark and lepton doublets, respectively. All fermion fields are written as three-dimensional flavour vectors and  $\mathcal{Y}_f^{(a)'}$  are  $3 \times 3$  complex matrices in flavour space. With an appropriate scalar potential, the neutral components of the scalar doublets acquire vacuum expectation values  $\langle 0 | \phi_a^{(0)} | 0 \rangle = v_a e^{i\theta_a}$ . It is convenient to make a  $U(2)$  transformation in the space of scalar fields,  $(\phi_1, \phi_2) \rightarrow (\Phi_1, \Phi_2)$ , so that only the first doublet has a non-zero vacuum expectation value  $v = (v_1^2 + v_2^2)^{1/2}$ .  $\Phi_1$  plays then the same role as the SM Higgs doublet, while  $\Phi_2$  does not participate in the electroweak symmetry breaking. In this scalar basis the Yukawa interactions become more transparent:

$$\sum_{a=1}^2 \mathcal{Y}_{d,\ell}^{(a)'} \phi_a = \sum_{a=1}^2 Y_{d,\ell}^{(a)'} \Phi_a, \quad \sum_{a=1}^2 \mathcal{Y}_u^{(a)'} \phi_a^c = \sum_{a=1}^2 Y_u^{(a)'} \Phi_a^c. \quad (103)$$

The fermion masses originate from the  $\Phi_1$  couplings, because  $\Phi_1$  is the only field acquiring a vacuum expectation value:

$$M'_f = Y_f^{(1)'} \frac{v}{\sqrt{2}}. \quad (104)$$

The diagonalization of these fermion mass matrices proceeds in exactly the same way as in the SM, and defines the fermion mass eigenstates  $d_i, u_i, \ell_i$ , with diagonal mass matrices  $M_f$ , as described in Section 2. However, in general, one cannot diagonalize simultaneously all Yukawa matrices, i.e., in the fermion mass-eigenstate basis the matrices  $Y_f^{(2)}$  remain non-diagonal, giving rise to dangerous flavour-changing transitions mediated by neutral scalars. The appearance of FCNC interactions represents a major phenomenological shortcoming, given the very strong experimental bounds on this type of phenomena.

To avoid this disaster, one needs to implement ad-hoc dynamical restrictions to guarantee the suppression of FCNC couplings at the required level. Unless the Yukawa couplings are very small or

the scalar bosons very heavy, a specific flavour structure is required by the data. The unwanted non-diagonal neutral couplings can be eliminated requiring the alignment in flavour space of the Yukawa matrices [166]:

$$Y_{d,\ell}^{(2)} = \varsigma_{d,\ell} Y_{d,\ell}^{(1)} = \frac{\sqrt{2}}{v} \varsigma_{d,\ell} M_{d,\ell}, \quad Y_u^{(2)} = \varsigma_u^* Y_u^{(1)} = \frac{\sqrt{2}}{v} \varsigma_u^* M_u, \quad (105)$$

with  $\varsigma_f$  arbitrary complex proportionality parameters.<sup>8</sup>

Flavour alignment constitutes a very simple implementation of minimal flavour violation. It results in a very specific dynamical structure, with all fermion-scalar interactions being proportional to the corresponding fermion masses. The Yukawas are fully characterized by the three complex alignment parameters  $\varsigma_f$ , which introduce new sources of  $\mathcal{CP}$  violation. The aligned two-Higgs doublet model Lagrangian satisfies the flavour constraints [168–178], and leads to a rich collider phenomenology with five physical scalar bosons [179–186]:  $h, H, A$  and  $H^\pm$ .

## 9 Flavour anomalies

The experimental data exhibit a few deviations from the SM predictions [187]. For instance, Table 2 shows a  $2.6\sigma$  violation of lepton universality in  $|g_\tau/g_\mu|$  at the 1% level, from  $W \rightarrow \ell\nu$  decays, that is difficult to reconcile with the precise 0.15% limits extracted from virtual  $W^\pm$  transitions, shown also in the same table. In fact, it has been demonstrated that it is not possible to accommodate this deviation from universality with an effective Lagrangian and, therefore, such a signal could only be explained by the introduction of new light degrees of freedom that so far remain undetected [188]. Thus, the most plausible explanation is a small problem (statistical fluctuation or underestimated systematics) in the LEP-2 measurements that will remain unresolved until more precise high-statistics  $W \rightarrow \ell\nu$  data samples become available.

Some years ago BaBar reported a non-zero  $\mathcal{CP}$  asymmetry in  $\tau^\pm \rightarrow \pi^\pm K_S \nu$  decays at the level of  $3.6 \cdot 10^{-3}$  [189], the same size than the SM expectation from  $K^0-\bar{K}^0$  mixing [190, 191] but with the opposite sign, which represents a  $2.8\sigma$  anomaly. So far, Belle has only reached a null result with a smaller  $10^{-2}$  sensitivity and, therefore, has not been able to either confirm or refute the asymmetry. Nevertheless, on very general grounds, it has been shown that the BaBar signal is incompatible with other sets of data ( $K^0$  and  $D^0$  mixing, neutron electric dipole moment) [192].

Another small flavour anomaly was triggered by the unexpected large value of  $\text{Br}(B^- \rightarrow \tau^- \bar{\nu})$ , found in 2006 by Belle [193] and later confirmed by BaBar [96], which implied values of  $|\mathbf{V}_{ub}|$  higher than the ones measured in semileptonic decays or extracted from global CKM fits. While the BaBar results remain unchanged, the reanalysis of the full Belle data sample resulted in a sizeable  $\sim 40\%$  reduction of the measured central value [97], eliminating the discrepancy with the SM but leaving a pending disagreement with the BaBar results.

In the last few years a series of anomalies have emerged in  $b \rightarrow c\tau\nu$  and  $b \rightarrow s\mu^+\mu^-$  transitions. Given the difficulty of the experimental analyses, the results should be taken with some caution and further studies with larger data sets are still necessary. Nevertheless, these anomalies exhibit a quite consistent pattern that makes them intriguing.

### 9.1 $b \rightarrow c\tau\nu$ decays

In 2012 the BaBar collaboration [194] observed an excess in  $B \rightarrow D^{(*)}\tau\nu_\tau$  decays with respect to the SM predictions [195], indicating a violation of lepton-flavour universality at the 30% level. The

<sup>8</sup>Actually, since one only needs that  $Y_f^{(1)'}$  and  $Y_f^{(2)'}$  can be simultaneously diagonalized, in full generality the factors  $\varsigma_f$  could be 3-dimensional diagonal matrices in the fermion mass-eigenstate basis (generalized alignment) [167]. The fashionable models of types I, II, X and Y, usually considered in the literature, are particular cases of the flavour-aligned Lagrangian with all  $\varsigma_f$  parameters real and fixed in terms of  $\tan\beta = v_2/v_1$  [166].



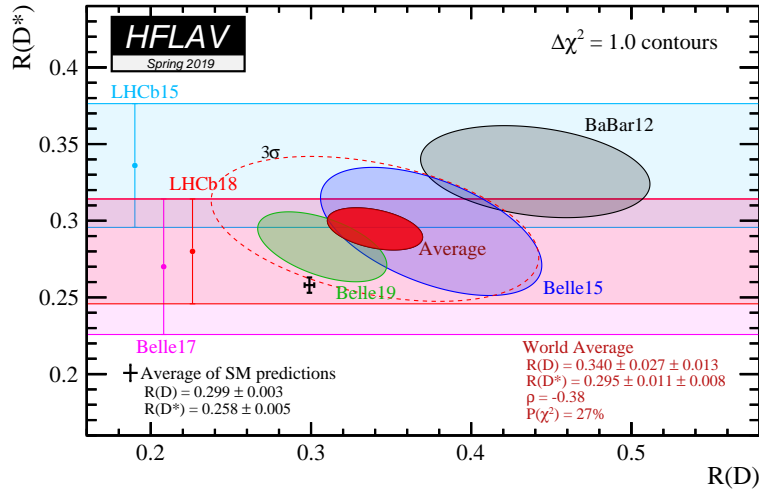
measured observables are the ratios

$$R(D) \equiv \frac{\text{Br}(B \rightarrow D\tau\nu)}{\text{Br}(B \rightarrow D\ell\nu)}, \quad R(D^*) \equiv \frac{\text{Br}(B \rightarrow D^*\tau\nu)}{\text{Br}(B \rightarrow D^*\ell\nu)}, \quad (106)$$

with  $\ell = e, \mu$ , where many sources of experimental and theoretical errors cancel. The effect has been later confirmed by LHCb [196] ( $D^*$  mode only) and Belle [197] (Fig. 11). Although the results of the last two experiments are slightly closer to the SM expected values,  $R(D) = 0.302 \pm 0.004$  and  $R(D^*) = 0.258^{+0.006}_{-0.005}$  [55, 58, 198–202], the resulting world averages [46]

$$R(D) = 0.340 \pm 0.027 \pm 0.013, \quad R(D^*) = 0.295 \pm 0.011 \pm 0.008, \quad (107)$$

deviate at the  $3.2\sigma$  level (considering their correlation of  $-0.38$ ) from the SM predictions, which is a very large effect for a tree-level SM transition.



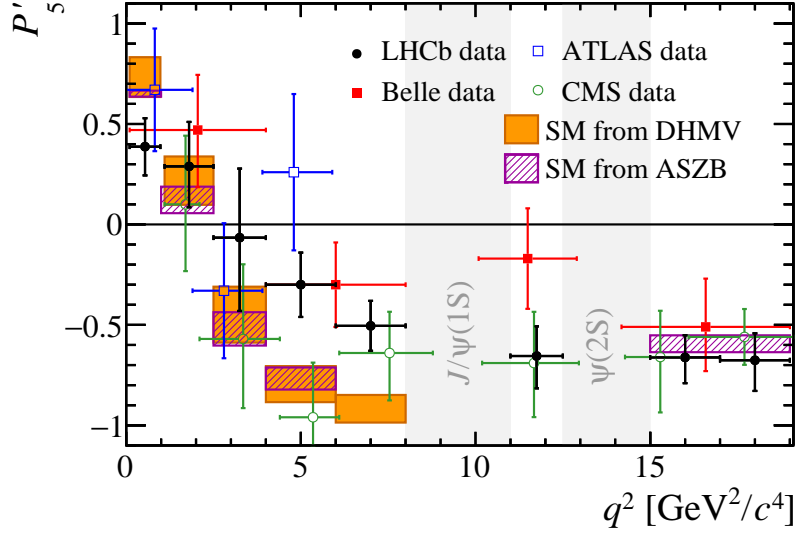
**Fig. 11:** Measurements of  $R(D)$  and  $R(D^*)$  and their average compared with the SM predictions. Filled contours correspond to  $\Delta\chi^2 = 1$  (68% CL for the bands and 39% CL for the ellipses), while the dashed ellipse displays the  $3\sigma$  region [46].

Different new-physics explanations of the anomaly have been put forward: new charged vector or scalar bosons, leptoquarks, right-handed neutrinos, etc.<sup>9</sup> The normalized  $q^2$  distributions measured by BaBar [194] and Belle [197] do not favour large deviations from the SM [203]. One must also take into account that the needed enhancement of the  $b \rightarrow c\tau\nu$  transition is constrained by the cross-channel  $b\bar{c} \rightarrow \tau\nu$ . A conservative (more stringent) upper bound  $\text{Br}(B_c \rightarrow \tau\nu) < 30\%$  (10%) can be extracted from the  $B_c$  lifetime [203, 204] (LEP data [205]). A global fit to the data, using a generic low-energy effective Hamiltonian with four-fermion effective operators [201, 202], finds several viable possibilities. However, while the fitted results clearly indicate that new-physics contributions are needed (much lower  $\chi^2$  than in the SM), they don't show any strong preference for a particular Wilson coefficient. The simplest solution would be some new-physics contribution that only manifests in the Wilson coefficient of the SM operator  $(\bar{c}_L\gamma^\mu b_L)(\bar{\tau}_L\gamma_\mu\nu_L)$ . This would imply a universal enhancement of all  $b \rightarrow c\tau\nu$  transitions, in agreement with the recent LHCb observation of the  $B_c \rightarrow J/\psi\tau\nu$  decay [206],

$$R(J/\psi) \equiv \frac{\text{Br}(B_c \rightarrow J/\psi\tau\nu)}{\text{Br}(B_c \rightarrow J/\psi\ell\nu)} = 0.71 \pm 0.17 \pm 0.18, \quad (108)$$

which is  $2\sigma$  above the SM expected value  $R(J/\psi) \sim 0.25\text{--}0.28$  [207–210]. Writing the four-fermion left-left operator in terms of  $SU(2)_L \otimes U(1)_Y$  invariant operators at the electroweak scale, and imposing that the experimental bounds on  $\text{Br}(b \rightarrow s\nu\bar{\nu})$  are satisfied, this possibility would imply rather large rates in  $b \rightarrow s\tau^+\tau^-$  transitions [211–213], but still safely below the current upper limits [214].

<sup>9</sup>A long, but not exhaustive, list of relevant references is given in Refs. [201, 202].



**Fig. 12:** Comparison between the SM predictions for  $P'_5$  [226–228] and the experimental measurements [229].

## 9.2 $b \rightarrow s\ell^+\ell^-$ decays

The rates of several  $b \rightarrow s\mu^+\mu^-$  transitions have been found at LHCb to be consistently lower than their SM predictions:  $B^+ \rightarrow K^+\mu^+\mu^-$  [215, 216],  $B^+ \rightarrow K^{*+}\mu^+\mu^-$  [215],  $B_d^0 \rightarrow K^0\mu^+\mu^-$  [215],  $B_d^0 \rightarrow K^{*0}\mu^+\mu^-$  [217],  $B_s^0 \rightarrow \phi\mu^+\mu^-$  [218] and  $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$  [219]. The angular and invariant-mass distributions of the final products in  $B \rightarrow K^*\mu^+\mu^-$  have been also studied by ATLAS [220], BaBar [221], Belle [222], CDF [223], CMS [224] and LHCb [217]. The rich variety of angular dependences in the four-body  $K\pi\mu^+\mu^-$  final state allows one to disentangle different sources of dynamical contributions. Particular attention has been put in the so-called *optimised observables*  $P'_i(q^2)$ , where  $q^2$  is the dilepton invariant-mass squared, which are specific combinations of angular observables that are free from form-factor uncertainties in the heavy quark-mass limit [225]. A sizeable discrepancy with the SM prediction [226–228], shown in Fig. 12, has been identified in two adjacent bins of the  $P'_5$  distribution, just below the  $J/\Psi$  peak. Belle has also included  $K^*e^+e^-$  final states in the analysis, but the results for this electron mode are compatible with the SM expectations [222].

The SM predictions for the previous observables suffer from hadronic uncertainties that are not easy to quantify. However, LHCb has also reported sizeable violations of lepton flavour universality, at the 2.1-2.6  $\sigma$  level, through the ratios [230]

$$R_{K^{*0}} \equiv \frac{\text{Br}(B_d^0 \rightarrow K^{*0}\mu^+\mu^-)}{\text{Br}(B_d^0 \rightarrow K^{*0}e^+e^-)} = \begin{cases} 0.66_{-0.07}^{+0.11} \pm 0.03, & q^2 \in [0.045, 1.1] \text{ GeV}^2, \\ 0.69_{-0.07}^{+0.11} \pm 0.05, & q^2 \in [1.1, 6.0] \text{ GeV}^2, \end{cases} \quad (109)$$

and [231]

$$R_K \equiv \frac{\text{Br}(B^+ \rightarrow K^+\mu^+\mu^-)}{\text{Br}(B^+ \rightarrow K^+e^+e^-)} \Bigg|_{q^2 \in [1.1, 6.0] \text{ GeV}^2} = 0.846_{-0.054-0.014}^{+0.060+0.016}. \quad (110)$$

These observables constitute a much cleaner probe of new physics because most theoretical uncertainties cancel out [232–235]. In the SM, the only difference between the muon and electron channels is the lepton mass. The SM theoretical predictions,  $R_K = 1.00 \pm 0.01_{\text{QED}}$ ,  $R_{K^{*0}}[0.045, 1.1] = 0.906 \pm 0.028_{\text{th}}$  and  $R_{K^{*0}}[1.1, 6] = 1.00 \pm 0.01_{\text{QED}}$  [236], deviate from these experimental measurements by 2.4  $\sigma$ , 2.1  $\sigma$  and 2.6  $\sigma$ , respectively. Owing to their larger uncertainties, the recent Belle measurements of  $R_{K^*}$  [237] and  $R_K$  [238] are compatible with the SM as well as with LHCb.

Global fits to the  $b \rightarrow s\ell^+\ell^-$  data with an effective low-energy Lagrangian

$$\mathcal{L}_{\text{eff}} = \frac{G_F}{\sqrt{2}} \mathbf{V}_{tb} \mathbf{V}_{ts}^* \frac{\alpha}{\pi} \sum_{i,\ell} C_{i,\ell} O_i^\ell \quad (111)$$

show a clear preference for new-physics contributions to the operators  $O_9^\ell = (\bar{s}_L \gamma_\mu b_L)(\bar{\ell} \gamma^\mu \ell)$  and  $O_{10}^\ell = (\bar{s}_L \gamma_\mu b_L)(\bar{\ell} \gamma^\mu \gamma_5 \ell)$ , with  $\ell = \mu$  [239–245]. Although the different analyses tend to favour slightly different solutions, two main common scenarios stand out: either  $\delta C_{9,\mu}^{\text{NP}} \approx -0.98$  or  $\delta C_{9,\mu}^{\text{NP}} = -\delta C_{10,\mu}^{\text{NP}} \approx -0.46$ . Both constitute large shifts (−24% and −11%, respectively) from the SM values:  $C_{9,\mu}^{\text{SM}}(\mu_b) \approx 4.1$  and  $C_{10,\mu}^{\text{SM}}(\mu_b) \approx -4.3$ , at  $\mu_b = 4.8$  GeV. The first possibility is slightly preferred by the global analysis of all data, while the left-handed new-physics solution accommodates better the lepton-flavour-universality-violating observables [240].

The left-handed scenario is theoretically appealing because it can be easily generated through  $SU(2)_L \otimes U(1)_Y$ -invariant effective operators at the electroweak scale that, moreover, could also provide an explanation to the  $b \rightarrow c\tau\nu$  anomaly. This possibility emerges naturally from the so-called  $U_1$  vector leptoquark model [246], and can be tested experimentally, since it implies a  $b \rightarrow s\tau^+\tau^-$  rate three orders of magnitude larger than the SM expectation [213]. For a recent review of theoretical models with a quite complete list of references, see Ref. [247].

## 10 Discussion

The flavour structure of the SM is one of the main pending questions in our understanding of weak interactions. Although we do not know the reasons of the observed family replication, we have learnt experimentally that the number of SM generations is just three (and no more). Therefore, we must study as precisely as possible the few existing flavours, to get some hints on the dynamics responsible for their observed structure.

In the SM all flavour dynamics originate in the fermion mass matrices, which determine the measurable masses and mixings. Thus, flavour is related through the Yukawa interactions with the scalar sector, the part of the SM Lagrangian that is more open to theoretical speculations. At present, we totally ignore the underlying dynamics responsible for the vastly different scales exhibited by the fermion spectrum and the particular values of the measured mixings. The SM Yukawa matrices are just a bunch of arbitrary parameters to be fitted to data, which is conceptually unsatisfactory.

The SM incorporates a mechanism to generate  $\mathcal{CP}$  violation, through the single phase naturally occurring in the CKM matrix. This mechanism, deeply rooted into the unitarity structure of  $\mathbf{V}$ , implies very specific requirements for  $\mathcal{CP}$  violation to show up. The CKM matrix has been thoroughly investigated in dedicated experiments and a large number of  $\mathcal{CP}$ -violating processes have been studied in detail. The flavour data seem to fit into the SM framework, confirming that the fermion mass matrices are the dominant source of flavour-mixing phenomena. However, a fundamental explanation of the flavour dynamics is still lacking.

At present, a few flavour anomalies have been identified in  $b \rightarrow c\tau\nu$  and  $b \rightarrow s\mu^+\mu^-$  transitions. Whether they truly represent the first signals of new phenomena, or just result from statistical fluctuations and/or underestimated systematics remains to be understood. New experimental input from LHC and Belle-II should soon clarify the situation. Very valuable information on the flavour dynamics is also expected from BESS-III and from several kaon (NA62, KOTO) and muon (MEG-II, Mu2e, Mu3e, COMET) experiments, complementing the high-energy searches for new phenomena at LHC. Unexpected surprises may well be discovered, probably giving hints of new physics at higher scales and offering clues to the problems of fermion mass generation, quark mixing and family replication.

## Acknowledgements

I want to thank the organizers for the charming atmosphere of this school and all the students for their many interesting questions and comments. This work has been supported in part by the Spanish Government and ERDF funds from the EU Commission [Grants FPA2017-84445-P and FPA2014-53631-C2-1-P], by Generalitat Valenciana [Grant Prometeo/2017/053] and by the Spanish Centro de Excelencia Severo Ochoa Programme [Grant SEV-2014-0398].

## Appendices

### A Conservation of the vector current

In the limit where all quark masses are equal, the QCD Lagrangian remains invariant under global  $SU(N_f)$  transformations of the quark fields in flavour space, where  $N_f$  is the number of (equal-mass) quark flavours. This guarantees the conservation of the corresponding Noether currents  $V_{ij}^\mu = \bar{u}_i \gamma^\mu d_j$ . In fact, using the QCD equations of motion, one immediately finds that

$$\partial_\mu V_{ij}^\mu \equiv \partial_\mu (\bar{u}_i \gamma^\mu d_j) = i (m_{u_i} - m_{d_j}) \bar{u}_i d_j, \quad (\text{A.1})$$

which vanishes when  $m_{u_i} = m_{d_j}$ . In momentum space, this reads  $q_\mu V_{ij}^\mu = \mathcal{O}(m_{u_i} - m_{d_j})$ , with  $q_\mu$  the corresponding momentum transfer.

Let us consider a  $0^-(k) \rightarrow 0^-(k')$  weak transition mediated by the vector current  $V_{ij}^\mu$ . The relevant hadronic matrix element is given in Eq. (19) and contains two form factors  $f_\pm(q^2)$ . The conservation of the vector current implies that  $f_-(q^2)$  is identically zero when  $m_{u_i} = m_{d_j}$ . Therefore,

$$\langle P'_i(k') | V_{ij}^\mu(x) | P_j(k) \rangle = e^{iq \cdot x} C_{PP'} (k + k')^\mu f_+(q^2). \quad (\text{A.2})$$

We have made use of translation invariance to write  $V_{ij}^\mu(x) = e^{iP \cdot x} V_{ij}^\mu(0) e^{-iP \cdot x}$ , with  $P^\mu$  the four-momentum operator. This determines the dependence of the matrix element on the space-time coordinate, with  $q^\mu = (k' - k)^\mu$ .

The conserved Noether charges

$$\mathcal{N}_{ij} = \int d^3x V_{ij}^0(x) = \int d^3x u_i^\dagger(x) d_j(x), \quad (\text{A.3})$$

annihilate one quark  $d_j$  and create instead one  $u_i$  (or replace  $\bar{u}_i$  by  $\bar{d}_j$ ), transforming the meson  $P$  into  $P'$  (up to a trivial Clebsch-Gordan factor  $C_{PP'}$  that, for light-quarks, takes the value  $1/\sqrt{2}$  when  $P'$  is a  $\pi^0$  and is 1 otherwise). Thus,

$$\langle P'(k') | \mathcal{N}_{ij} | P(k) \rangle = C_{PP'} \langle P'(k') | P(k) \rangle = C_{PP'} (2\pi)^3 2k^0 \delta^{(3)}(\vec{k}' - \vec{k}). \quad (\text{A.4})$$

On the other side, inserting in this matrix element the explicit expression for  $\mathcal{N}_{ij}$  in (A.3) and making use of (A.2),

$$\langle P'(k') | \mathcal{N}_{ij} | P(k) \rangle = \int d^3x \langle P'(k') | V_{ij}^0(x) | P(k) \rangle = C_{PP'} (2\pi)^3 2k^0 \delta^{(3)}(\vec{q}) f_+(0). \quad (\text{A.5})$$

Comparing Eqs. (A.4) and (A.5), one finally obtains the result

$$f_+(0) = 1. \quad (\text{A.6})$$

Therefore, the flavour symmetry  $SU(N_f)$  determines the normalization of the form factor at  $q^2 = 0$ , when  $m_{u_i} = m_{d_j}$ . It is possible to prove that the deviations from 1 are of second order in the symmetry-breaking quark mass difference, i.e.,  $f_+(0) = 1 + \mathcal{O}[(m_{u_i} - m_{d_j})^2]$  [25].

## References

- [1] S.L. Glashow, *Nucl. Phys.* **22** (1961) 579, doi:[10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [2] S. Weinberg, *Phys. Rev. Lett.* **19** (1967) 1264, doi:[10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [3] A. Salam, in *Elementary particle theory*, ed. N. Svartholm (Almquist and Wiksells, Stockholm, 1969), p. 367, repr. in *Selected papers of Abdus Salam*, eds. A. Ali *et al.* (World Scientific, Singapore, 1994) p. 244, doi:[10.1142/9789812795915\\_0034](https://doi.org/10.1142/9789812795915_0034).
- [4] A. Pich, *The Standard Model of electroweak interactions*, in Proc. European School of High-Energy Physics, 2004 Barcelona and 2006 Aronsborg, ed. R. Fleischer, CERN-2006-003, p. 1, doi:[10.5170/CERN-2006-003.1](https://doi.org/10.5170/CERN-2006-003.1); CERN-2007-005, p. 1, doi:[10.5170/CERN-2007-005.1](https://doi.org/10.5170/CERN-2007-005.1).
- [5] A. Pich, *Flavour physics and CP violation*, in Proc. 6th CERN – Latin-American School of High-Energy Physics, CLASHEP 2011, Natal, Brazil, March 33 - April 5, 2011, CERN-2013-003, p. 119, ed. C. Grojean, M. Mulders and M. Spiropulu, doi:[10.5170/CERN-2013-003.119](https://doi.org/10.5170/CERN-2013-003.119).
- [6] N. Cabibbo, *Phys. Rev. Lett.* **10** (1963) 531, doi:[10.1103/PhysRevLett.10.531](https://doi.org/10.1103/PhysRevLett.10.531).
- [7] M. Kobayashi and T. Maskawa, *Prog. Theor. Phys.* **42** (1973) 652, doi:[10.1143/PTP.49.652](https://doi.org/10.1143/PTP.49.652).
- [8] S.L. Glashow, J. Iliopoulos and L. Maiani, *Phys. Rev.* **D2** (1970) 1285, doi:[10.1103/PhysRevD.2.1285](https://doi.org/10.1103/PhysRevD.2.1285).
- [9] M. Tanabashi *et al.* [Particle Data Group], *Phys. Rev. D* **98** (2018) 030001, doi:[10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001); and 2019 update at <https://pdg.lbl.gov/2019/>.
- [10] W.J. Marciano and A. Sirlin, *Phys. Rev. Lett.* **61** (1988) 1815, doi:[10.1103/PhysRevLett.61.1815](https://doi.org/10.1103/PhysRevLett.61.1815).
- [11] T. van Ritbergen and R.G. Stuart, *Phys. Rev. Lett.* **82** (1999) 488, doi:[10.1103/PhysRevLett.82.488](https://doi.org/10.1103/PhysRevLett.82.488), arXiv:[hep-ph/9808283](https://arxiv.org/abs/hep-ph/9808283); *Nucl. Phys.* **B564** (2000) 343, doi:[10.1016/S0550-3213\(99\)00572-6](https://doi.org/10.1016/S0550-3213(99)00572-6), arXiv:[hep-ph/9904240](https://arxiv.org/abs/hep-ph/9904240).
- [12] A. Pak and A. Czarnecki, *Phys. Rev. Lett.* **100** (2008) 241807, doi:[10.1103/PhysRevLett.100.241807](https://doi.org/10.1103/PhysRevLett.100.241807), arXiv:[0803.0960](https://arxiv.org/abs/0803.0960) [hep-ph].
- [13] MuLan Collaboration, *Phys. Rev.* **D87** (2013) 052003, doi:[10.1103/PhysRevD.87.052003](https://doi.org/10.1103/PhysRevD.87.052003), arXiv:[1211.0960](https://arxiv.org/abs/1211.0960) [hep-ex].
- [14] E. Braaten, S. Narison and A. Pich, *Nucl. Phys.* **B373** (1992) 581, doi:[10.1016/0550-3213\(92\)90267-F](https://doi.org/10.1016/0550-3213(92)90267-F).
- [15] A. Pich and A. Rodríguez-Sánchez, *Phys. Rev.* **D94** (2016) 034027, doi:[10.1103/PhysRevD.94.034027](https://doi.org/10.1103/PhysRevD.94.034027), arXiv:[1605.06830](https://arxiv.org/abs/1605.06830) [hep-ph].
- [16] A. Pich, EPJ Web Conf. **137** (2017) 01016, doi:[10.1051/epjconf/201713701016](https://doi.org/10.1051/epjconf/201713701016).
- [17] A. Pich, *Prog. Part. Nucl. Phys.* **75** (2014) 41, doi:[10.1016/j.pnpnp.2013.11.002](https://doi.org/10.1016/j.pnpnp.2013.11.002), arXiv:[1310.7922](https://arxiv.org/abs/1310.7922) [hep-ph].
- [18] A. Pich, *Flavourdynamics*, arXiv:[hep-ph/9601202](https://arxiv.org/abs/hep-ph/9601202).
- [19] J. C. Hardy and I. S. Towner, *Phys. Rev.* **C91** (2015) 025501, doi:[10.1103/PhysRevC.91.025501](https://doi.org/10.1103/PhysRevC.91.025501), arXiv:[1411.5987](https://arxiv.org/abs/1411.5987) [nucl-ex].
- [20] J. Hardy and I. Towner, *Nuclear Beta Decays and CKM Unitarity*, arXiv:[1807.01146](https://arxiv.org/abs/1807.01146) [nucl-ex].
- [21] W. J. Marciano and A. Sirlin, *Phys. Rev. Lett.* **96** (2006) 032002, doi:[10.1103/PhysRevLett.96.032002](https://doi.org/10.1103/PhysRevLett.96.032002), arXiv:[hep-ph/0510099](https://arxiv.org/abs/hep-ph/0510099) [hep-ph].
- [22] C. Seng *et al.*, *Phys. Rev. Lett.* **121** (2018) 241804, doi:[10.1103/PhysRevLett.121.241804](https://doi.org/10.1103/PhysRevLett.121.241804), arXiv:[1807.10197](https://arxiv.org/abs/1807.10197) [hep-ph].
- [23] C. Seng *et al.*, *Phys. Rev.* **D101** (2020) 111301, doi:[10.1103/PhysRevD.101.111301](https://doi.org/10.1103/PhysRevD.101.111301).
- [24] A. Czarnecki, W. J. Marciano and A. Sirlin, *Phys. Rev.* **D100** (2019) 073008, doi:[10.1103/PhysRevD.100.073008](https://doi.org/10.1103/PhysRevD.100.073008).
- [25] M. Ademollo and R. Gatto, *Phys. Rev. Lett.* **13** (1964) 264, doi:[10.1103/PhysRevLett.13.264](https://doi.org/10.1103/PhysRevLett.13.264).

- [26] V. Cirigliano *et al.*, *Eur. Phys. J.* **C27** (2003) 255, doi:10.1140/epjc/s2002-01093-2, arXiv:hep-ph/0209226.
- [27] D. Poganic *et al.*, *Phys. Rev. Lett.* **93** (2004) 181803, doi:10.1103/PhysRevLett.93.181803, arXiv:hep-ex/0312030.
- [28] V. Cirigliano, M. Giannotti and H. Neufeld, *JHEP* **11** (2008) 006, doi:10.1088/1126-6708/2008/11/006, arXiv:0807.4507 [hep-ph].
- [29] A. Kastner and H. Neufeld, *Eur. Phys. J.* **C57** (2008) 541, doi:10.1140/epjc/s10052-008-0703-6, arXiv:0805.2222 [hep-ph].
- [30] M. Antonelli *et al.* [FlaviaNet Working Group on Kaon Decays], *Eur. Phys. J.* **C69** (2010) 399, doi:10.1140/epjc/s10052-010-1406-3, arXiv:1005.2323 [hep-ph].
- [31] M. Moulson, *Experimental determination of  $V_{us}$  from kaon decays*, arXiv:1411.5252 [hep-ex].
- [32] R.E. Behrends and A. Sirlin, *Phys. Rev. Lett.* **4** (1960) 186, doi:10.1103/PhysRevLett.4.186.
- [33] H. Leutwyler and M. Roos, *Z. Phys.* **C25** (1984) 91, doi:10.1007/BF01571961.
- [34] A. Bazavov *et al.*, *Phys. Rev. Lett.* **112** (2014) 112001, doi:10.1103/PhysRevLett.112.112001, arXiv:1312.1228 [hep-ph].
- [35] N. Carrasco *et al.*, *Phys. Rev.* **D93** (2016) 114512, doi:10.1103/PhysRevD.93.114512, arXiv:1602.04113 [hep-lat].
- [36] J. Bijnens and P. Talavera, *Nucl. Phys.* **B669** (2003) 341, doi:10.1016/S0550-3213(03)00581-9, arXiv:hep-ph/0303103.
- [37] M. Jamin, J.A. Oller and A. Pich, *JHEP* **02** (2004) 047, doi:10.1088/1126-6708/2004/02/047, arXiv:hep-ph/0401080.
- [38] V. Cirigliano *et al.*, *JHEP* **04** (2005) 006, doi:10.1088/1126-6708/2005/04/006, arXiv:hep-ph/0503108.
- [39] S. Aoki *et al.* [Flavour Lattice Averaging Group], *Eur. Phys. J.* **C80** (2020) 113, doi:10.1140/epjc/s10052-019-7354-7.
- [40] W.J. Marciano, *Phys. Rev. Lett.* **93** (2004) 231803, doi:10.1103/PhysRevLett.93.231803, arXiv:hep-ph/0402299.
- [41] V. Cirigliano and H. Neufeld, *Phys. Lett.* **B700** (2011) 7, doi:10.1016/j.physletb.2011.04.038, arXiv:1102.0563 [hep-ph].
- [42] V. Cirigliano *et al.*, *Rev. Mod. Phys.* **84** (2012) 399, doi:10.1103/RevModPhys.84.399, arXiv:1107.6001 [hep-ph].
- [43] N. Cabibbo, E.C. Swallow and R. Winston, *Annu. Rev. Nucl. Part. Sci.* **53** (2003) 39, doi:10.1146/annurev.nucl.53.013103.155258, arXiv:hep-ph/0307298; *Phys. Rev. Lett.* **92** (2004) 251803, doi:10.1103/PhysRevLett.92.251803, arXiv:hep-ph/0307214.
- [44] V. Mateu and A. Pich, *JHEP* **10** (2005) 041, doi:10.1088/1126-6708/2005/10/041, arXiv:hep-ph/0509045.
- [45] E. Gámiz *et al.*, *Phys. Rev. Lett.* **94** (2005), 011803, doi:10.1103/PhysRevLett.94.011803, arXiv:hep-ph/0408044; *JHEP* **01** (2003) 060, doi:10.1088/1126-6708/2003/01/060, arXiv:hep-ph/0212230.
- [46] Y. S. Amhis *et al.* [HFLAV Collaboration], *Averages of  $b$ -hadron,  $c$ -hadron, and  $\tau$ -lepton properties as of 2018* arXiv:1909.12524 [hep-ex]; <https://hflav.web.cern.ch/>, last accessed 11 Feb. 2021.
- [47] N. Isgur and M. Wise, *Phys. Lett.* **B232** (1989) 113, doi:10.1016/0370-2693(89)90566-2; **B237** (1990) 527, doi:10.1016/0370-2693(90)91219-2.
- [48] B. Grinstein, *Nucl. Phys.* **B339** (1990) 253, doi:10.1016/0550-3213(90)90349-I.
- [49] E. Eichten and B. Hill, *Phys. Lett.* **B234** (1990) 511, doi:10.1016/0370-2693(90)92049-O.

- [50] H. Georgi, *Phys. Lett.* **B240** (1990) 447, doi:[10.1016/0370-2693\(90\)91128-X](https://doi.org/10.1016/0370-2693(90)91128-X).
- [51] M. Neubert, *Phys. Lett.* **B264** (1991) 455, doi:[10.1016/0370-2693\(91\)90377-3](https://doi.org/10.1016/0370-2693(91)90377-3).
- [52] M. Luke, *Phys. Lett.* **B252** (1990) 447, doi:[10.1016/0370-2693\(90\)90568-Q](https://doi.org/10.1016/0370-2693(90)90568-Q).
- [53] I. Caprini, L. Lellouch and M. Neubert, *Nucl. Phys.* **B530** (1998) 153, doi:[10.1016/S0550-3213\(98\)00350-2](https://doi.org/10.1016/S0550-3213(98)00350-2), arXiv:[hep-ph/9712417](https://arxiv.org/abs/hep-ph/9712417).
- [54] J. A. Bailey *et al.* [MILC Collaboration], *Phys. Rev.* **D92** (2015) 034506, doi:[10.1103/PhysRevD.92.034506](https://doi.org/10.1103/PhysRevD.92.034506), arXiv:[1503.07237](https://arxiv.org/abs/1503.07237) [hep-lat].
- [55] D. Bigi and P. Gambino, *Phys. Rev.* **D94** (2016) 094008, doi:[10.1103/PhysRevD.94.094008](https://doi.org/10.1103/PhysRevD.94.094008), arXiv:[1606.08030](https://arxiv.org/abs/1606.08030) [hep-ph].
- [56] D. Bigi, P. Gambino and S. Schacht, *Phys. Lett.* **B769** (2017) 441, doi:[10.1016/j.physletb.2017.04.022](https://doi.org/10.1016/j.physletb.2017.04.022); *JHEP* **11** (2017) 061, doi:[10.1007/JHEP11\(2017\)061](https://doi.org/10.1007/JHEP11(2017)061).
- [57] B. Grinstein and A. Kobach, *Phys. Lett.* **B771** (2017) 359, doi:[10.1016/j.physletb.2017.05.078](https://doi.org/10.1016/j.physletb.2017.05.078).
- [58] F.U. Bernlochner *et al.*, *Phys. Rev.* **D95** (2017)115008, doi:[10.1103/PhysRevD.95.115008](https://doi.org/10.1103/PhysRevD.95.115008), [Erratum **D97** (2018) 059902, [10.1103/PhysRevD.97.059902](https://doi.org/10.1103/PhysRevD.97.059902)], arXiv: [1703.05330](https://arxiv.org/abs/1703.05330) [hep-ph]; **D96** (2017) 091503, doi:[10.1103/PhysRevD.96.091503](https://doi.org/10.1103/PhysRevD.96.091503), arXiv: [1708.07134](https://arxiv.org/abs/1708.07134) [hep-ph].
- [59] C.G. Boyd, B. Grinstein and R.F. Lebed, *Phys. Rev.* **D56** (1997) 6895, doi:[10.1103/PhysRevD.56.6895](https://doi.org/10.1103/PhysRevD.56.6895), arXiv:[hep-ph/9705252](https://arxiv.org/abs/hep-ph/9705252).
- [60] Belle Collaboration, arXiv:[1702.01521](https://arxiv.org/abs/1702.01521) [hep-ex].
- [61] E. Waheed *et al.* [Belle], *Phys. Rev.* **D100** (2019) 052007, doi:[10.1103/PhysRevD.100.052007](https://doi.org/10.1103/PhysRevD.100.052007).
- [62] P. Gambino, M. Jung and S. Schacht, *Phys. Lett.* **B795** (2019) 386, doi:[10.1016/j.physletb.2019.06.039](https://doi.org/10.1016/j.physletb.2019.06.039).
- [63] BaBar Collaboration, *Phys. Rev. Lett.* **104** (2010) 011802, doi:[10.1103/PhysRevLett.104.011802](https://doi.org/10.1103/PhysRevLett.104.011802), arXiv:[0904.4063](https://arxiv.org/abs/0904.4063) [hep-ex].
- [64] Belle Collaboration, *Phys. Rev.* **D93** (2016) 032006, doi:[10.1103/PhysRevD.93.032006](https://doi.org/10.1103/PhysRevD.93.032006), arXiv:[1510.03657](https://arxiv.org/abs/1510.03657) [hep-ex].
- [65] I.I.Y. Bigi *et al.*, *Phys. Rev. Lett.* **71** (1993) 496, doi:[10.1103/PhysRevLett.71.496](https://doi.org/10.1103/PhysRevLett.71.496), arXiv:[hep-ph/9304225](https://arxiv.org/abs/hep-ph/9304225); *Phys. Lett.* **B323** (1994) 408, doi:[10.1016/0370-2693\(94\)91239-4](https://doi.org/10.1016/0370-2693(94)91239-4), arXiv:[hep-ph/9311339](https://arxiv.org/abs/hep-ph/9311339).
- [66] A.V. Manohar and M.B. Wise, *Phys. Rev.* **D49** (1994) 1310, doi:[10.1103/PhysRevD.49.1310](https://doi.org/10.1103/PhysRevD.49.1310), arXiv:[hep-ph/9308246](https://arxiv.org/abs/hep-ph/9308246).
- [67] M. Gremm and A. Kapustin, *Phys. Rev.* **D55** (1997) 6924, doi:[10.1103/PhysRevD.55.6924](https://doi.org/10.1103/PhysRevD.55.6924), arXiv:[hep-ph/9603448](https://arxiv.org/abs/hep-ph/9603448).
- [68] D. Benson *et al.*, **B665** (2003) 367, doi:[10.1016/S0550-3213\(03\)00452-8](https://doi.org/10.1016/S0550-3213(03)00452-8), arXiv:[hep-ph/0302262](https://arxiv.org/abs/hep-ph/0302262).
- [69] C.W. Bauer *et al.*, *Phys. Rev.* **D70** (2004) 094017, doi:[10.1103/PhysRevD.70.094017](https://doi.org/10.1103/PhysRevD.70.094017), arXiv:[hep-ph/0408002](https://arxiv.org/abs/hep-ph/0408002).
- [70] P. Gambino and N. Uraltsev, *Eur. Phys. J.* **C34** (2004) 181, doi:[10.1140/epjc/s2004-01671-2](https://doi.org/10.1140/epjc/s2004-01671-2), arXiv:[hep-ph/0401063](https://arxiv.org/abs/hep-ph/0401063).
- [71] D. Benson, I.I. Bigi and N. Uraltsev, *Nucl. Phys.* **B710** (2005) 371, doi:[10.1016/j.nuclphysb.2004.12.035](https://doi.org/10.1016/j.nuclphysb.2004.12.035), arXiv:[hep-ph/0410080](https://arxiv.org/abs/hep-ph/0410080).
- [72] O. Buchmuller and H. Flacher, *Phys. Rev.* **D73** (2006) 073008, doi:[10.1103/PhysRevD.73.073008](https://doi.org/10.1103/PhysRevD.73.073008), arXiv:[hep-ph/0507253](https://arxiv.org/abs/hep-ph/0507253).
- [73] T. Mannel, S. Turczyk and N. Uraltsev, *JHEP* **11** (2010) 109, doi:[10.1007/JHEP11\(2010\)109](https://doi.org/10.1007/JHEP11(2010)109), arXiv:[1009.4622](https://arxiv.org/abs/1009.4622) [hep-ph].
- [74] P. Gambino, *JHEP* **09** (2011) 055, doi:[10.1007/JHEP09\(2011\)055](https://doi.org/10.1007/JHEP09(2011)055), arXiv:[1107.3100](https://arxiv.org/abs/1107.3100) [hep-ph].

- [75] A. Alberti *et al.*, *Nucl. Phys.* **B870** (2013) 16, doi:[10.1016/j.nuclphysb.2013.01.005](https://doi.org/10.1016/j.nuclphysb.2013.01.005), arXiv:[1212.5082](https://arxiv.org/abs/1212.5082) [hep-ph].
- [76] A. Alberti, P. Gambino and S. Nandi, *JHEP* **01** (2014) 147, doi:[10.1007/JHEP01\(2014\)147](https://doi.org/10.1007/JHEP01(2014)147).
- [77] T. Mannel, A.A. Pivovarov and D. Rosenthal, *Phys. Rev.* **D92** (2015) 054025, doi:[10.1103/PhysRevD.92.054025](https://doi.org/10.1103/PhysRevD.92.054025), arXiv:[1506.08167](https://arxiv.org/abs/1506.08167) [hep-ph].
- [78] A. Alberti *et al.*, *Phys. Rev. Lett.* **114** (2015) 061802, doi:[10.1103/PhysRevLett.114.061802](https://doi.org/10.1103/PhysRevLett.114.061802), arXiv:[1411.6560](https://arxiv.org/abs/1411.6560) [hep-ph].
- [79] P. Gambino, K. J. Healey and S. Turczyk, *Phys. Lett.* **B763** (2016) 60, doi:[10.1016/j.physletb.2016.10.023](https://doi.org/10.1016/j.physletb.2016.10.023).
- [80] P. Ball and R. Zwicky, *Phys. Rev.* **D71** (2005) 014015, doi:[10.1103/PhysRevD.71.014015](https://doi.org/10.1103/PhysRevD.71.014015), arXiv:[hep-ph/0406232](https://arxiv.org/abs/hep-ph/0406232).
- [81] G. Duplancic *et al.*, *JHEP* **04** (2008) 014, doi:[10.1088/1126-6708/2008/04/014](https://doi.org/10.1088/1126-6708/2008/04/014), arXiv:[0801.1796](https://arxiv.org/abs/0801.1796) [hep-ph].
- [82] A. Khodjamirian *et al.*, *Phys. Rev.* **D83** (2011) 094031, doi:[10.1103/PhysRevD.83.094031](https://doi.org/10.1103/PhysRevD.83.094031), arXiv:[1103.2655](https://arxiv.org/abs/1103.2655) [hep-ph].
- [83] A. Bharucha, *JHEP* **05** (2012) 092, doi:[10.1007/JHEP05\(2012\)092](https://doi.org/10.1007/JHEP05(2012)092), arXiv:[1203.1359](https://arxiv.org/abs/1203.1359) [hep-ph].
- [84] J.A. Bailey *et al.* [Fermilab Lattice and MILC Collaborations], *Phys. Rev.* **D92** (2015) 014024, doi:[10.1103/PhysRevD.92.014024](https://doi.org/10.1103/PhysRevD.92.014024), arXiv:[1503.07839](https://arxiv.org/abs/1503.07839) [hep-lat].
- [85] J.M. Flynn *et al.*, *Phys. Rev.* **D91** (2015) 074510, doi:[10.1103/PhysRevD.91.074510](https://doi.org/10.1103/PhysRevD.91.074510), arXiv:[1501.05373](https://arxiv.org/abs/1501.05373) [hep-lat].
- [86] M. Antonelli *et al.*, *Phys. Rept.* **494** (2010) 197, doi:[10.1016/j.physrep.2010.05.003](https://doi.org/10.1016/j.physrep.2010.05.003), arXiv:[0907.5386](https://arxiv.org/abs/0907.5386) [hep-ph].
- [87] B.O. Lange *et al.*, *Phys. Rev.* **D72** (2005) 073006, doi:[10.1103/PhysRevD.72.073006](https://doi.org/10.1103/PhysRevD.72.073006), arXiv:[hep-ph/0504071](https://arxiv.org/abs/hep-ph/0504071); *Nucl. Phys.* **B699** (2004) 335, doi:[10.1016/j.nuclphysb.2004.07.041](https://doi.org/10.1016/j.nuclphysb.2004.07.041), arXiv:[hep-ph/0402094](https://arxiv.org/abs/hep-ph/0402094).
- [88] J.R. Andersen and E. Gardi, *JHEP* **01** (2006) 097, doi:[10.1088/1126-6708/2006/01/097](https://doi.org/10.1088/1126-6708/2006/01/097), arXiv:[hep-ph/0509360](https://arxiv.org/abs/hep-ph/0509360).
- [89] P. Gambino *et al.*, *JHEP* **10** (2007) 058, doi:[10.1088/1126-6708/2007/10/058](https://doi.org/10.1088/1126-6708/2007/10/058), arXiv:[0707.2493](https://arxiv.org/abs/0707.2493) [hep-ph].
- [90] U. Aglietti *et al.*, *Eur. Phys. J.* **C59** (2009) 831, doi:[10.1140/epjc/s10052-008-0817-x](https://doi.org/10.1140/epjc/s10052-008-0817-x), arXiv:[0711.0860](https://arxiv.org/abs/0711.0860) [hep-ph].
- [91] C. Greub, M. Neubert and B. D. Pecjak, *Eur. Phys. J.* **C65** (2010) 501, doi:[10.1140/epjc/s10052-009-1210-0](https://doi.org/10.1140/epjc/s10052-009-1210-0), arXiv:[0909.1609](https://arxiv.org/abs/0909.1609) [hep-ph].
- [92] G. Paz, *JHEP* **06** (2009) 083, doi:[10.1088/1126-6708/2009/06/083](https://doi.org/10.1088/1126-6708/2009/06/083), arXiv:[0903.3377](https://arxiv.org/abs/0903.3377) [hep-ph].
- [93] Z. Ligeti *et al.*, *Phys. Rev.* **D82** (2010) 033003, doi:[10.1103/PhysRevD.82.033003](https://doi.org/10.1103/PhysRevD.82.033003), arXiv:[1003.1351](https://arxiv.org/abs/1003.1351) [hep-ph]; **D78** (2008) 114014, doi:[10.1103/PhysRevD.78.114014](https://doi.org/10.1103/PhysRevD.78.114014), arXiv:[0807.1926](https://arxiv.org/abs/0807.1926) [hep-ph]; *Status of SIMBA*, arXiv:[1101.3310](https://arxiv.org/abs/1101.3310) [hep-ph].
- [94] P. Gambino and J.F. Kamenik, *Nucl. Phys.* **B840** (2010) 424, doi:[10.1016/j.nuclphysb.2010.07.019](https://doi.org/10.1016/j.nuclphysb.2010.07.019), arXiv:[1004.0114](https://arxiv.org/abs/1004.0114) [hep-ph].
- [95] LHCb Collaboration, *Nature Phys.* **11** (2015) 743, doi:[10.1038/nphys3415](https://doi.org/10.1038/nphys3415).
- [96] BaBar Collaboration, *Phys. Rev.* **D77** (2008) 011107, doi:[10.1103/PhysRevD.77.011107](https://doi.org/10.1103/PhysRevD.77.011107), arXiv:[0708.2260](https://arxiv.org/abs/0708.2260) [hep-ex]; **D81** (2010) 051101, doi:[10.1103/PhysRevD.81.051101](https://doi.org/10.1103/PhysRevD.81.051101), arXiv:[0912.2453](https://arxiv.org/abs/0912.2453) [hep-ex]; **D88** (2013) 031102, doi:[10.1103/PhysRevD.88.031102](https://doi.org/10.1103/PhysRevD.88.031102), arXiv:[1207.0698](https://arxiv.org/abs/1207.0698) [hep-ex].
- [97] Belle Collaboration, *Phys. Rev. Lett.* **110** (2013) 131801, doi:[10.1103/PhysRevLett.110.131801](https://doi.org/10.1103/PhysRevLett.110.131801), arXiv:[1208.4678](https://arxiv.org/abs/1208.4678) [hep-ex]; *Phys. Rev.* **D92** (2015) 051102, doi:[10.1103/PhysRevD.92.051102](https://doi.org/10.1103/PhysRevD.92.051102), arXiv:[1503.05613](https://arxiv.org/abs/1503.05613) [hep-ex].



- [98] CMS Collaboration, *Phys. Lett.* **B736** (2014) 33, doi:10.1016/j.physletb.2014.06.076.
- [99] The ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the Tevatron Electroweak Working Group and the SLD Electroweak and Heavy Flavour Groups, *Precision electroweak measurements and constraints on the Standard Model*, arXiv:1012.2367 [hep-ex]; *The LEP Electroweak Working Group*, <http://lepewwg.web.cern.ch/LEPEWWG/>.
- [100] The ALEPH, DELPHI, L3, OPAL and SLD Collaborations, the LEP Electroweak Working Group and the SLD Electroweak and Heavy Flavour Groups, *Phys. Rept.* **427** (2006) 257, doi:10.1016/j.physrep.2005.12.006, arXiv:hep-ex/0509008.
- [101] L. Wolfenstein, *Phys. Rev. Lett.* **51** (1983) 1945, doi:10.1103/PhysRevLett.51.1945.
- [102] A.J. Buras, M.E. Lautenbacher and G. Ostermaier, *Phys. Rev.* **D50** (1994) 3433, doi:10.1103/PhysRevD.50.3433, arXiv:hep-ph/9403384.
- [103] T. Inami and C.S. Lim, *Progr. Theor. Phys.* **65** (1981) 297, doi:10.1143/PTP.65.297, [Erratum-ibid. 1772, doi:10.1143/PTP.65.1772].
- [104] M.K. Gaillard and B.W. Lee, *Phys. Rev.* **D10** (1974) 897, doi:10.1103/PhysRevD.10.897.
- [105] ARGUS Collaboration, *Phys. Lett.* **B192** (1987) 245, doi:10.1016/0370-2693(87)91177-4.
- [106] CDF Collaboration, *Phys. Rev. Lett.* **97** (2006) 242003, doi:10.1103/PhysRevLett.97.242003.
- [107] A.J. Buras, M. Jamin and P.H. Weisz, *Nucl. Phys.* **B347** (1990) 491, doi:10.1016/0550-3213(90)90373-L.
- [108] S. Herrlich and U. Nierste, *Nucl. Phys.* **B419** (1994) 292; **B476** (1996) 27, doi:10.1016/0550-3213(96)00324-0, arXiv:hep-ph/9604330.
- [109] A. Pich and J. Prades, *Phys. Lett.* **B346** (1995) 342, doi:10.1016/0370-2693(95)00019-H, arXiv:hep-ph/9407226.
- [110] C. Jarlskog, *Phys. Rev. Lett.* **55** (1985) 1039, doi:10.1103/PhysRevLett.55.1039; *Z. Phys.* **C29** (1985) 491, doi:10.1007/BF01565198.
- [111] CKMfitter Group, *Eur. Phys. J.* **C41** (2005) 1, doi:10.1140/epjc/s2005-02169-1, arXiv:hep-ph/0406184; 2019 update at <http://ckmfitter.in2p3.fr/>.
- [112] NA48 Collaboration, *Phys. Lett.* **B544** (2002) 97, doi:10.1016/S0370-2693(02)02476-0, arXiv:hep-ex/0208009; **B465** (1999) 335, doi:10.1016/S0370-2693(99)01030-8, arXiv:hep-ex/9909022; *Eur. Phys. J.* **C22** (2001) 231, doi:10.1007/s100520100822, arXiv:hep-ex/0110019.
- [113] KTeV Collaboration, *Phys. Rev.* **D83** (2011) 092001, doi:10.1103/PhysRevD.83.092001, arXiv:1011.0127 [hep-ex]; **D70** (2004) 079904, doi:10.1103/PhysRevD.67.012005, arXiv:hep-ex/0208007; *Phys. Rev. Lett.* **83** (1999) 22, doi:10.1103/PhysRevLett.83.22, arXiv:hep-ex/9905060.
- [114] NA31 Collaboration, *Phys. Lett.* **B317** (1993) 233, doi:10.1016/0370-2693(93)91599-I; **B206** (1988) 169, doi:10.1016/0370-2693(88)91282-8.
- [115] E731 Collaboration, *Phys. Rev. Lett.* **70** (1993) 1203, doi:10.1103/PhysRevLett.70.1203.
- [116] A.J. Buras, M. Jamin and M.E. Lautenbacher, *Nucl. Phys.* **B408** (1993) 209, doi:10.1016/0550-3213(93)90535-W, arXiv:hep-ph/9303284; *Phys. Lett.* **B389** (1996) 749, doi:10.1016/S0370-2693(96)80019-0, arXiv:hep-ph/9608365.
- [117] M. Ciuchini *et al.*, *Phys. Lett.* **B301** (1993) 263, doi:10.1016/0370-2693(93)90699-I, arXiv:hep-ph/9212203; *Z. Phys.* **C68** (1995) 239, doi:10.1007/BF01566672, arXiv:hep-ph/9501265.
- [118] E. Pallante and A. Pich, *Phys. Rev. Lett.* **84** (2000) 2568, doi:10.1103/PhysRevLett.84.2568, arXiv:hep-ph/9911233; *Nucl. Phys.* **B592** (2001) 294, doi:10.1016/S0550-3213(00)00601-5, arXiv:hep-ph/0007208.

- [119] E. Pallante, A. Pich and I. Scimemi, *Nucl. Phys.* **B617** (2001) 441, doi:10.1016/S0550-3213(01)00418-7, arXiv:hep-ph/0105011.
- [120] V. Cirigliano *et al.*, *Eur. Phys. J.* **C33** (2004) 369, doi:10.1140/epjc/s2003-01579-3, arXiv:hep-ph/0310351.
- [121] H. Gisbert and A. Pich, *Rept. Prog. Phys.* **81** (2018) 076201, doi:10.1088/1361-6633/aac18e, arXiv:1712.06147 [hep-ph].
- [122] V. Cirigliano *et al.*, *JHEP* **02** (2020) 032, doi:10.1007/JHEP02(2020)032.
- [123] R. Abbott *et al.* [RBC and UKQCD Collaborations], *Phys. Rev.* **D102** (2020) 054509, doi:10.1103/PhysRevD.102.054509.
- [124] G. Buchalla, A.J. Buras and M.E. Lautenbacher, *Rev. Mod. Phys.* **68** (1996) 1125, doi:10.1103/RevModPhys.68.1125, arXiv:hep-ph/9512380.
- [125] J. Brod and M. Gorbahn, *Phys. Rev.* **D82** (2010) 094026, doi:10.1103/PhysRevD.82.094026, arXiv:1007.0684 [hep-ph].
- [126] A.B. Carter and A.I. Sanda, *Phys. Rev. Lett.* **45** (1980) 952, doi:10.1103/PhysRevLett.45.952; *Phys. Rev.* **D23** (1981) 1567, doi:10.1103/PhysRevD.23.1567.
- [127] I.I. Bigi and A.I. Sanda, *Nucl. Phys.* **B193** (1981) 85, doi:10.1016/0550-3213(81)90519-8.
- [128] P. Krawczyk *et al.*, *Nucl. Phys.* **B307** (1988) 19, doi:10.1016/0550-3213(88)90520-2.
- [129] Belle Collaboration, *Phys. Rev. Lett.* **108** (2012) 171802, doi:10.1103/PhysRevLett.108.171802, arXiv:1201.4643 [hep-ex].
- [130] BaBar and Belle Collaborations, *Phys. Rev. Lett.* **121** (2018) 261801, doi:10.1103/PhysRevLett.121.261801; *Phys. Rev.* **D98** (2018) 112012, doi:10.1103/PhysRevD.98.112012.
- [131] M. Gronau and D. London, *Phys. Rev. Lett.* **65** (1990) 3381, doi:10.1103/PhysRevLett.65.3381.
- [132] M. Gronau and D. London, *Phys. Lett.* **B253** (1991) 483, doi:10.1016/0370-2693(91)91756-L.
- [133] M. Gronau and D. Wyler, *Phys. Lett.* **B265** (1991) 172, doi:10.1016/0370-2693(91)90034-N.
- [134] D. Atwood *et al.*, *Phys. Rev. Lett.* **78** (1997) 3257, doi:10.1103/PhysRevLett.78.3257, arXiv:hep-ph/9612433; *Phys. Rev.* **D63** (2001) 036005, doi:10.1103/PhysRevD.63.036005, arXiv:hep-ph/0008090.
- [135] A. Lenz and U. Nierste, *JHEP* **06** (2007) 072, doi:10.1088/1126-6708/2007/06/072, arXiv:hep-ph/0612167.
- [136] M. Artuso, G. Borissov and A. Lenz, *Rev. Mod. Phys.* **88** (2016) 045002, doi:10.1103/RevModPhys.88.045002, arXiv:1511.09466 [hep-ph].
- [137] T. Jubb *et al.*, *Nucl. Phys.* **B915** (2017) 431, doi:10.1016/j.nuclphysb.2016.12.020.
- [138] UTfit Collaboration, *JHEP* **07** (2005) 028, doi:10.1088/1126-6708/2005/07/028, arXiv:hep-ph/0501199; 2018 update at <http://www.utfit.org/UTfit/>.
- [139] LHCb Collaboration, *Phys. Rev. Lett.* **122** (2019) 211803, doi:10.1103/PhysRevLett.122.211803.
- [140] LHCb Collaboration, *Phys. Rev. Lett.* **125** (2020) 231801, doi:10.1103/PhysRevLett.125.231801.
- [141] M. Gorbahn and U. Haisch, *Phys. Rev. Lett.* **97** (2006) 122002, doi:10.1103/PhysRevLett.97.122002, arXiv:hep-ph/0605203.
- [142] D. Gomez Dumm and A. Pich, *Phys. Rev. Lett.* **80** (1998) 4633, doi:10.1103/PhysRevLett.80.4633, arXiv:hep-ph/9801298.
- [143] G. Ecker and A. Pich, *Nucl. Phys.* **B366** (1991) 189, doi:10.1016/0550-3213(91)90056-4.
- [144] G. Ecker, A. Pich and E. de Rafael, *Nucl. Phys.* **B303** (1988) 665, doi:10.1016/0550-3213(88)90425-7; **B291** (1987) 692, doi:10.1016/0550-3213(87)90491-3; *Phys. Lett.* **B189** (1987) 363, doi:10.1016/0370-2693(87)91448-1.

- [145] A.J. Buras *et al.*, *Nucl. Phys.* **B423** (1994) 349, doi:10.1016/0550-3213(94)90138-4, arXiv:hep-ph/9402347.
- [146] G. Buchalla, G. D’Ambrosio and G. Isidori, *Nucl. Phys.* **B672** (2003) 387, doi:10.1016/j.nuclphysb.2003.09.010, arXiv:hep-ph/0308008.
- [147] KTeV Collaboration, *Phys. Rev. Lett.* **93** (2004) 021805, doi:10.1103/PhysRevLett.93.021805, arXiv:hep-ex/0309072.
- [148] A.J. Buras *et al.*, *Phys. Rev. Lett.* **95** (2005) 261805, doi:10.1103/PhysRevLett.95.261805, arXiv:hep-ph/0508165; *JHEP* **11** (2006) 002, doi:10.1088/1126-6708/2006/11/002, [Erratum: **11** (2012) 167, doi:10.1007/JHEP11(2012)167], arXiv:hep-ph/0603079.
- [149] J. Brod, M. Gorbahn and E. Stamou, *Phys. Rev.* **D83** (2011) 034030, doi:10.1103/PhysRevD.83.034030, arXiv:1009.0947 [hep-ph].
- [150] M. Gorbahn, talk at Kaon 2019, Perugia, Italy, 10 September 2019, slides from Indico.
- [151] NA62 Collaboration, G. Ruggiero, *J. Phys. Conf. Ser.* **1526** (2020) 012003, doi:10.1088/1742-6596/1526/1/012003.
- [152] KOTO Collaboration, *Phys. Rev. Lett.* **122** (2019) 021802, doi:10.1103/PhysRevLett.122.021802.
- [153] M. Misiak *et al.*, *Phys. Rev. Lett.* **114** (2015) 221801, doi:10.1103/PhysRevLett.114.221801.
- [154] M. Czakon *et al.*, *JHEP* **04** (2015) 168, doi:10.1007/JHEP04(2015)168.
- [155] M. Misiak, A. Rehman and M. Steinhauser, *JHEP* **06** (2020), 175, doi:10.1007/JHEP06(2020)175.
- [156] CMS and LHCb Collaborations, *Nature* **522** (2015) 68, doi:10.1038/nature14474.
- [157] LHCb Collaboration, *Phys. Rev. Lett.* **118** (2017) 191801, doi:10.1103/PhysRevLett.118.191801.
- [158] ATLAS Collaboration, *JHEP* **04** (2019) 098, doi:10.1007/JHEP04(2019)098.
- [159] K. De Bruyn *et al.*, *Phys. Rev. Lett.* **109** (2012) 041801, doi:10.1103/PhysRevLett.109.041801, arXiv:1204.1737 [hep-ph].
- [160] C. Bobeth *et al.*, *Phys. Rev. Lett.* **112** (2014) 101801, doi:10.1103/PhysRevLett.112.101801, arXiv:1311.0903 [hep-ph].
- [161] S. Weinberg, *Phys. Rev. Lett.* **43** (1979) 1566, doi:10.1103/PhysRevLett.43.1566.
- [162] G. Isidori, Y. Nir and G. Perez, *Ann. Rev. Nucl. Part. Sci.* **60** (2010) 355, doi:10.1146/annurev.nucl.012809.104534, arXiv:1002.0900 [hep-ph].
- [163] L.J. Hall and L. Randall, *Phys. Rev. Lett.* **65** (1990) 2939, doi:10.1103/PhysRevLett.65.2939.
- [164] R.S. Chivukula and H. Georgi, *Phys. Lett.* **B188** (1987) 99, doi:10.1016/0370-2693(87)90713-1.
- [165] G. D’Ambrosio *et al.*, *Nucl. Phys.* **B645** (2002) 155, doi:10.1016/S0550-3213(02)00836-2, arXiv:hep-ph/0207036.
- [166] A. Pich and P. Tuzón, *Phys. Rev.* **D80** (2009) 091702, doi:10.1103/PhysRevD.80.091702, arXiv:0908.1554 [hep-ph].
- [167] A. Peñuelas and A. Pich, *JHEP* **12** (2017), 084, doi:10.1007/JHEP12(2017)084.
- [168] M. Jung, A. Pich and P. Tuzón, *JHEP* **11** (2010) 003, doi:10.1007/JHEP11(2010)003; *Phys. Rev.* **D83** (2011) 074011, doi:10.1103/PhysRevD.83.074011, arXiv:1011.5154 [hep-ph].
- [169] M. Jung, X. Q. Li and A. Pich, *JHEP* **10** (2012) 063, doi:10.1007/JHEP10(2012)063, arXiv:1208.1251 [hep-ph].
- [170] M. Jung and A. Pich, *JHEP* **04** (2014) 076, doi:10.1007/JHEP04(2014)076.
- [171] X. Q. Li, J. Lu and A. Pich, *JHEP* **06** (2014) 022, doi:10.1007/JHEP06(2014)022.
- [172] V. Ilisie, *JHEP* **04** (2015) 077, doi:10.1007/JHEP04(2015)077.
- [173] A. Cherchiglia *et al.*, *JHEP* **01** (2017) 007, doi:10.1007/JHEP01(2017)007.
- [174] Q. Chang, P.F. Li and X.Q. Li, *Eur. Phys. J.* **C75** (2015) 594, doi:10.1140/epjc/s10052-015-3813-y.

- [175] Q. Y. Hu, X. Q. Li and Y. D. Yang, *Eur. Phys. J.* **C77** (2017) 190, doi:10.1140/epjc/s10052-017-4748-2; **C77** (2017) 228, doi:10.1140/epjc/s10052-017-4794-9.
- [176] N. Cho *et al.*, *Adv. High Energy Phys.* **2017** (2017) 2863647, doi:10.1155/2017/2863647.
- [177] T. Han, S.K. Kang and J. Sayre, *JHEP* **02** (2016) 097, doi:10.1007/JHEP02(2016)097.
- [178] T. Enomoto and R. Watanabe, *JHEP* **05** (2016) 002, doi:10.1007/JHEP05(2016)002.
- [179] A. Celis, V. Ilisie and A. Pich, *JHEP* **07** (2013) 053, doi:10.1007/JHEP07(2013)053, arXiv:1302.4022 [hep-ph]; **12** (2013) 095, doi:10.1007/JHEP12(2013)095, arXiv:1310.7941 [hep-ph].
- [180] G. Abbas *et al.*, *JHEP* **06** (2015) 005, doi:10.1007/JHEP06(2015)005.
- [181] V. Ilisie and A. Pich, *JHEP* **09** (2014) 089, doi:10.1007/JHEP09(2014)089.
- [182] W. Altmannshofer, S. Gori and G.D. Kribs, *Phys. Rev.* **D86** (2012) 115009, doi:10.1103/PhysRevD.86.115009, arXiv:1210.2465 [hep-ph].
- [183] Y. Bai *et al.*, *Phys. Rev.* **D87** (2013) 115013, doi:10.1103/PhysRevD.87.115013, arXiv:1210.4922 [hep-ph].
- [184] L. Duarte, G.A. González-Sprinberg and J. Vidal, *JHEP* **11** (2013) 114, doi:10.1007/JHEP11(2013)114, arXiv:1308.3652 [hep-ph].
- [185] C. Ayala *et al.*, *JHEP* **03** (2017) 128, doi:10.1007/JHEP03(2017)128.
- [186] L. Wang and X.F. Han, *JHEP* **04** (2014) 128, doi:10.1007/JHEP04(2014)128.
- [187] A. Pich, *PoS LHCP2019* (2019) 078, doi:10.22323/1.350.0078.
- [188] A. Filipuzzi, J. Portolés and M. González-Alonso, *Phys. Rev.* **D85** (2012) 116010, doi:10.1103/PhysRevD.85.116010, arXiv:1203.2092 [hep-ph].
- [189] BaBar Collaboration, *Phys. Rev.* **D85** (2012) 031102, doi:10.1103/PhysRevD.85.031102, [Erratum-ibid. 099904, doi:10.1103/PhysRevD.85.099904], arXiv:1109.1527 [hep-ex].
- [190] I.I. Bigi and A.I. Sanda, *Phys. Lett.* **B625** (2005) 47, doi:10.1016/j.physletb.2005.08.033, arXiv:hep-ph/0506037.
- [191] Y. Grossman and Y. Nir, *JHEP* **04** (2012) 002, doi:10.1007/JHEP04(2012)002, arXiv:1110.3790 [hep-ph].
- [192] V. Cirigliano, A. Crivellin and M. Hoferichter, *Phys. Rev. Lett.* **120** (2018) 141803, doi:10.1103/PhysRevLett.120.141803.
- [193] Belle Collaboration, *Phys. Rev. Lett.* **97** (2006) 251802, doi:10.1103/PhysRevLett.97.251802, arXiv:hep-ex/0604018; *Phys. Rev.* **D82** (2010) 071101, doi:10.1103/PhysRevD.82.071101, arXiv:1006.4201 [hep-ex].
- [194] BaBar Collaboration, *Phys. Rev. Lett.* **109** (2012) 101802, doi:10.1103/PhysRevLett.109.101802, arXiv:1205.5442 [hep-ex]; *Phys. Rev.* **D88** (2013) 072012, doi:10.1103/PhysRevD.88.072012, arXiv:1303.0571 [hep-ex].
- [195] S. Fajfer, J.F. Kamenik and I. Nisandzic, *Phys. Rev.* **D85** (2012) 094025, doi:10.1103/PhysRevD.85.094025, arXiv:1203.2654 [hep-ph].
- [196] LHCb Collaboration, *Phys. Rev. Lett.* **115** (2015) 111803, doi:10.1103/PhysRevLett.115.111803, [Erratum-ibid. 159901, doi:10.1103/PhysRevLett.115.159901]; **120** (2018) 171802, doi:10.1103/PhysRevLett.120.171802; *Phys. Rev.* **D97** (2018) 072013, doi:10.1103/PhysRevD.97.072013.
- [197] Belle Collaboration, *Phys. Rev.* **D92** (2015) 072014, doi:10.1103/PhysRevD.92.072014, arXiv:1507.03233 [hep-ex]; **D94** (2016) 072007, doi:10.1103/PhysRevD.94.072007, arXiv:1607.07923 [hep-ex]; **D97** (2018) 012004, doi:10.1103/PhysRevD.97.012004; *Phys. Rev. Lett.* **118** (2017) 211801, doi:10.1103/PhysRevLett.118.211801, arXiv:1612.00529 [hep-ex]; *Measurement of  $\mathcal{R}(D)$  and  $\mathcal{R}(D^*)$  with a semileptonic tagging method*, arXiv:1904.08794 [hep-ex].

- [198] D. Bigi, P. Gambino and S. Schacht, *JHEP* **11** (2017) 061, doi:[10.1007/JHEP11\(2017\)061](https://doi.org/10.1007/JHEP11(2017)061).
- [199] S. Jaiswal, S. Nandi and S.K. Patra, *JHEP* **12** (2017) 060, doi:[10.1007/JHEP12\(2017\)060](https://doi.org/10.1007/JHEP12(2017)060).
- [200] M. Jung and D.M. Straub, *JHEP* **01** (2019) 009, doi:[10.1007/JHEP01\(2019\)009](https://doi.org/10.1007/JHEP01(2019)009).
- [201] C. Murgui *et al.*, *JHEP* **09** (2019) 103, doi:[10.1007/JHEP09\(2019\)103](https://doi.org/10.1007/JHEP09(2019)103).
- [202] R. Mandal *et al.*, *JHEP* **08** (2020) 022, doi:[10.1007/JHEP08\(2020\)022](https://doi.org/10.1007/JHEP08(2020)022).
- [203] A. Celis *et al.*, *Phys. Lett.* **B771** (2017) 168, doi:[10.1016/j.physletb.2017.05.037](https://doi.org/10.1016/j.physletb.2017.05.037); *JHEP* **01** (2013) 054, doi:[10.1007/JHEP01\(2013\)054](https://doi.org/10.1007/JHEP01(2013)054), arXiv:[1210.8443](https://arxiv.org/abs/1210.8443) [hep-ph].
- [204] R. Alonso, B. Grinstein and J. Martin Camalich, *Phys. Rev. Lett.* **118** (2017) 081802, doi:[10.1103/PhysRevLett.118.081802](https://doi.org/10.1103/PhysRevLett.118.081802).
- [205] A. Akeroyd and C. Chen, *Phys. Rev.* **D96** (2017) 075011, doi:[10.1103/PhysRevD.96.075011](https://doi.org/10.1103/PhysRevD.96.075011), arXiv:[1708.04072](https://arxiv.org/abs/1708.04072) [hep-ph].
- [206] LHCb Collaboration, *Phys. Rev. Lett.* **120** (2018) 121801, doi:[10.1103/PhysRevLett.120.121801](https://doi.org/10.1103/PhysRevLett.120.121801).
- [207] A.Y. Anisimov *et al.*, *Phys. Lett.* **B452** (1999) 129, doi:[10.1016/S0370-2693\(99\)00273-7](https://doi.org/10.1016/S0370-2693(99)00273-7), arXiv:[hep-ph/9812514](https://arxiv.org/abs/hep-ph/9812514).
- [208] V.V. Kiselev, arXiv:[hep-ph/0211021](https://arxiv.org/abs/hep-ph/0211021).
- [209] M.A. Ivanov, J.G. Korner and P. Santorelli, *Phys. Rev.* **D73** (2006) 054024, doi:[10.1103/PhysRevD.73.054024](https://doi.org/10.1103/PhysRevD.73.054024), arXiv:[hep-ph/0602050](https://arxiv.org/abs/hep-ph/0602050).
- [210] E. Hernandez, J. Nieves and J.M. Verde-Velasco, *Phys. Rev.* **D74** (2006) 074008, doi:[10.1103/PhysRevD.74.074008](https://doi.org/10.1103/PhysRevD.74.074008), arXiv:[hep-ph/0607150](https://arxiv.org/abs/hep-ph/0607150).
- [211] R. Alonso, B. Grinstein and J. Martin Camalich, *JHEP* **10** (2015) 184, doi:[10.1007/JHEP10\(2015\)184](https://doi.org/10.1007/JHEP10(2015)184).
- [212] A. Crivellin, D. Müller and T. Ota, *JHEP* **09** (2017) 040, doi:[10.1007/JHEP09\(2017\)040](https://doi.org/10.1007/JHEP09(2017)040).
- [213] B. Capdevila *et al.*, *Phys. Rev. Lett.* **120** (2018) 181802, doi:[10.1103/PhysRevLett.120.181802](https://doi.org/10.1103/PhysRevLett.120.181802).
- [214] C. Bobeth and U. Haisch, *Acta Phys. Polon.* **B44** (2013) 127, doi:[10.5506/APhysPolB.44.127](https://doi.org/10.5506/APhysPolB.44.127).
- [215] LHCb Collaboration, *JHEP* **06** (2014) 133, doi:[10.1007/JHEP06\(2014\)133](https://doi.org/10.1007/JHEP06(2014)133).
- [216] LHCb Collaboration, *JHEP* **02** (2013) 105, doi:[10.1007/JHEP02\(2013\)105](https://doi.org/10.1007/JHEP02(2013)105).
- [217] LHCb Collaboration, *JHEP* **11** (2016) 047, doi:[10.1007/JHEP11\(2016\)047](https://doi.org/10.1007/JHEP11(2016)047), [Erratum **04** (2017) 142, doi:[10.1007/JHEP04\(2017\)142](https://doi.org/10.1007/JHEP04(2017)142)]; **02** (2016) 104, doi:[10.1007/JHEP02\(2016\)104](https://doi.org/10.1007/JHEP02(2016)104); **08** (2013) 131, doi:[10.1007/JHEP08\(2013\)131](https://doi.org/10.1007/JHEP08(2013)131); *Phys. Rev. Lett.* **111** (2013) 191801, doi:[10.1103/PhysRevLett.111.191801](https://doi.org/10.1103/PhysRevLett.111.191801).
- [218] LHCb Collaboration, *JHEP* **09** (2015) 179, doi:[10.1007/JHEP09\(2015\)179](https://doi.org/10.1007/JHEP09(2015)179); **07** (2013) 084, doi:[10.1007/JHEP07\(2013\)084](https://doi.org/10.1007/JHEP07(2013)084).
- [219] LHCb Collaboration, *JHEP* **06** (2015) 115, doi:[10.1007/JHEP06\(2015\)115](https://doi.org/10.1007/JHEP06(2015)115), [Erratum **09** (2018) 145, doi:[10.1007/JHEP09\(2018\)145](https://doi.org/10.1007/JHEP09(2018)145)]; *Phys. Lett.* **B725** (2013) 25, doi:[10.1016/j.physletb.2013.06.060](https://doi.org/10.1016/j.physletb.2013.06.060).
- [220] ATLAS Collaboration, *JHEP* **10** (2018) 047, doi:[10.1007/JHEP10\(2018\)047](https://doi.org/10.1007/JHEP10(2018)047).
- [221] BaBar Collaboration, *Phys. Rev.* **D93** (2016) 052015, doi:[10.1103/PhysRevD.93.052015](https://doi.org/10.1103/PhysRevD.93.052015), arXiv:[1508.07960](https://arxiv.org/abs/1508.07960) [hep-ex].
- [222] Belle Collaboration, *Phys. Rev. Lett.* **118** (2017) 111801, doi:[10.1103/PhysRevLett.118.111801](https://doi.org/10.1103/PhysRevLett.118.111801), arXiv:[1612.05014](https://arxiv.org/abs/1612.05014) [hep-ex]; **103** (2009) 171801, doi:[10.1103/PhysRevLett.103.171801](https://doi.org/10.1103/PhysRevLett.103.171801), arXiv:[0904.0770](https://arxiv.org/abs/0904.0770) [hep-ex]; *Angular analysis of  $B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-$* , arXiv:[1604.04042](https://arxiv.org/abs/1604.04042) [hep-ex].
- [223] CDF Collaboration, *Phys. Rev. Lett.* **108** (2012) 081807, doi:[10.1103/PhysRevLett.108.081807](https://doi.org/10.1103/PhysRevLett.108.081807), arXiv:[1108.0695](https://arxiv.org/abs/1108.0695) [hep-ex].
- [224] CMS Collaboration, *Phys. Lett.* **B781** (2018) 517, doi:[10.1016/j.physletb.2018.04.030](https://doi.org/10.1016/j.physletb.2018.04.030); *Phys. Lett.* **B753** (2016) 424, doi:[10.1016/j.physletb.2015.12.020](https://doi.org/10.1016/j.physletb.2015.12.020).

- [225] S. Descotes-Genon *et al.*, *JHEP* **01** (2013) 048, doi:10.1007/JHEP01(2013)048.
- [226] S. Descotes-Genon *et al.*, *JHEP* **12** (2014) 125, doi:10.1007/JHEP12(2014)125.
- [227] W. Altmannshofer and D.M. Straub, *Eur. Phys. J.* **C75** (2015) 382, doi:10.1140/epjc/s10052-015-3602-7.
- [228] A. Bharucha, D.M. Straub and R. Zwicky, *JHEP* **08** (2016) 098, doi:10.1007/JHEP08(2016)098.
- [229] F. Dettori [LHCb Collaboration], *Search for new physics in  $b \rightarrow q\ell\ell$  decays*, arXiv:1805.05073 [hep-ex].
- [230] LHCb Collaboration, *JHEP* **08** (2017) 055, doi:10.1007/JHEP08(2017)055.
- [231] LHCb Collaboration, *Phys. Rev. Lett.* **122** (2019) 191801, doi:10.1103/PhysRevLett.122.191801.
- [232] C. Bobeth *et al.*, *JHEP* **01** (2012) 107, doi:10.1007/JHEP01(2012)107, arXiv:1111.2558 [hep-ph].
- [233] G. Hiller and F. Kruger, *Phys. Rev.* **D69** (2004) 074020, doi:10.1103/PhysRevD.69.074020, arXiv:hep-ph/0310219.
- [234] C. Bobeth, G. Hiller and G. Piranishvili, *JHEP* **12** (2007) 040, doi:10.1088/1126-6708/2007/12/040, arXiv:0709.4174 [hep-ph].
- [235] HPQCD Collaboration, *Phys. Rev. Lett.* **111** (2013) 162002, 10.1103/PhysRevLett.111.162002, [Erratum **112** (2014) 149902, doi:10.1103/PhysRevLett.112.149902].
- [236] M. Bordone, G. Isidori and A. Pattori, *Eur. Phys. J.* **C76** (2016) 440, doi:10.1140/epjc/s10052-016-4274-7.
- [237] Belle Collaboration, *Test of lepton flavor universality in  $B \rightarrow K^*\ell^+\ell^-$  decays at Belle*, arXiv:1904.02440 [hep-ex].
- [238] Belle Collaboration, *Test of lepton flavor universality in  $B \rightarrow K\ell^+\ell^-$  decays*, arXiv:1908.01848 [hep-ex].
- [239] J. Aebischer *et al.*, *Eur. Phys. J.* **C80** (2020) 252, doi:10.1140/epjc/s10052-020-7817-x.
- [240] M. Algueró *et al.*, *Eur. Phys. J.* **C79** (2019) 714, doi:10.1140/epjc/s10052-019-7216-3.
- [241] M. Ciuchini *et al.*, *Eur. Phys. J.* **C79** (2019) 719, doi:10.1140/epjc/s10052-019-7210-9.
- [242] A. Datta, J. Kumar and D. London, *Phys. Lett.* **B797** (2019) 134858, doi:10.1016/j.physletb.2019.134858.
- [243] K. Kowalska, D. Kumar and E.M. Sessolo, *Eur. Phys. J.* **C79** (2019) 840, doi:10.1140/epjc/s10052-019-7330-2.
- [244] A. Arbey *et al.*, *Phys. Rev.* **D100** (2019) 015045, doi:10.1103/PhysRevD.100.015045.
- [245] A.K. Alok *et al.*, *JHEP* **06** (2019) 089, doi:10.1007/JHEP06(2019)089.
- [246] A. Angelescu *et al.*, *JHEP* **10** (2018) 183, doi:10.1007/JHEP10(2018)183.
- [247] S. Bifani *et al.*, *J. Phys.* **G46** (2019) 023001, doi:10.1088/1361-6471/aaf5de, arXiv:1809.06229 [hep-ex].

## Heavy-ion physics: freedom to do hot, dense, exciting QCD

M. E. Tejeda-Yeomans

Faculty of Science, CUICBAS, University of Colima, Mexico

### Abstract

In these two lectures I review the basics of heavy-ion collisions at relativistic energies and the physics we can do with them. I aim to cover the basics on the kinematics and observables in heavy-ion collider experiments, the basics on the phenomenology of the nuclear matter phase diagram, some of the model building and simulations currently used in the heavy-ion physics community and a selected list of amazing phenomenological discoveries and predictions.

### Keywords

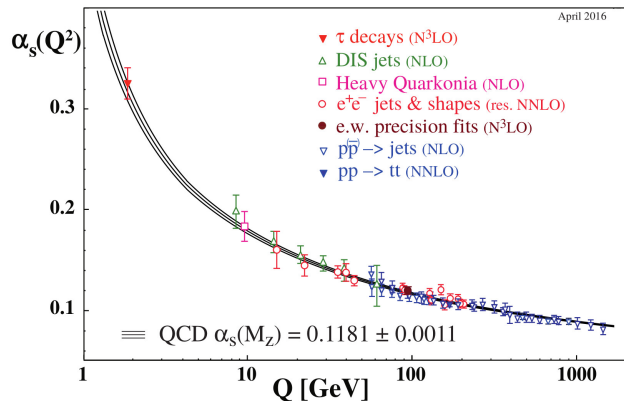
QCD; heavy-ion physics; quark-gluon plasma; QCD phase diagrams; temperature; baryon density.

### 1 The not-so-simple questions that remain

The recent discoveries in particle physics are built upon the platform created by a solid quantum field theory of the strong interactions: quantum chromodynamics (QCD). Without it, we could not "see" through the haze of multiple particle production at the LHC and study different decay channels of the Higgs boson. Eventhough QCD is robust enough to encompass perturbative and non-perturbative approaches that allow us to study a variety of phenomena in hadron colliders, we are always tempted to put it to the test in the high temperature and/or density regimes. For example, if we already know how to study proton-proton collisions ( $pp$  collisions) at LHC energies, starting with parton level cross sections using QCD matrix elements that involve quarks and gluons, that later become hadron level cross sections and can be connected to shape observables or jet observables, how much of this knowledge can be used to simulate jet formation in nucleus-nucleus collisions ( $A - A$  collisions, where  $A$  can be any nucleus)? Do we assume that hadronization and jet formation in the collision of large systems happens just as it does in small systems? It turns out that these questions are just the tip of the iceberg: the physics of heavy-ion collisions at relativistic energies are a portal to new phenomena of the strong interactions. They are the means to explore regimes of QCD under *extreme* conditions: what happens to strongly interacting matter when it is in a thermal bath and/or when it is highly compressed? Does this mean that we expect ordinary nuclear matter to phase transition into new states of matter? How do we know theoretically and experimentally what are the signals we should look for when these phase transitions occur? Suddenly, heavy-ion collision experiments represent our best tool to learn about the questions that remain: the rich dynamical QCD phenomena that emerges under these conditions, an arena to further extend the vast knowledge we already have on hadron and QCD physics.

Now, if you are already an expert on using QCD to study  $pp$  collisions, then you might be surprised to learn that from there, it is an easy road to learn more about  $A - A$  collisions. This is the road I plan to take in these couple of lectures, to discuss about the phases of nuclear matter, about QCD now under extreme conditions, about quantum field theory techniques for larger systems and about code used in  $pp$  collisions to be able to simulate  $A - A$  collisions. In these two lectures I will aim to give you the basics of heavy-ion collisions at relativistic energies and the physics we can do with them. These lectures will rely heavily on previous enlightening talks and short courses given at different venues (see for example Ref. [1]) and on various articles and textbooks (see Refs. [2–8]).

The structure of the lectures is as follows: in Section 2 we review the essential ideas that helped establish QCD as the theory of strong interactions and use them to have an initial discussion about the



**Fig. 1:** Measurements of  $\alpha_s$  as a function of the energy scale  $Q$ , as reported in the quantum chromodynamics review of the Particle Data Group [9]. From each set of experiments, there is a value of  $\alpha_s$  that has been extracted using different levels of QCD perturbation theory, which are indicated in brackets. Re-used with permission from PDG.

nuclear matter phase diagram, then in Section 3 we review the basic ideas used to study relativistic heavy-ion collisions and the evolution of the system after the collisions, in Section 4 we discuss the quark-gluon plasma and its basic properties, in Section 5 we introduce the tools required to study the phase diagram of nuclear matter and the observables with which phase transitions and critical behavior can be probed, finally in Section 6 we highlight observables that help characterize the QGP formation and evolution. We provide final remarks and future prospects in Section 7.

## 2 QCD as usual—said no one ever

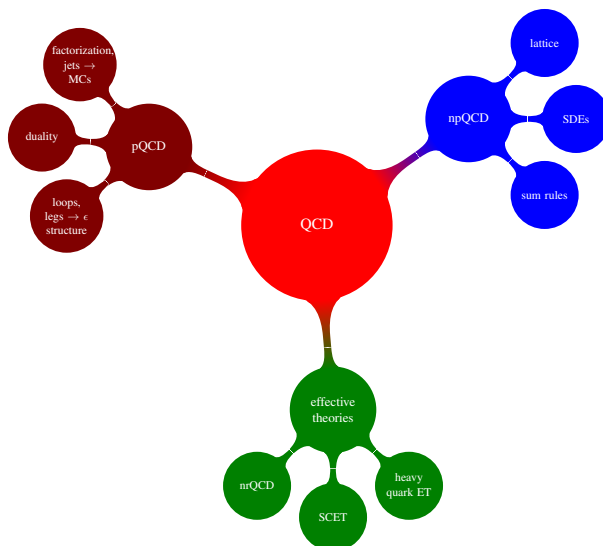
### 2.1 The basics

It is now carved in stone (which is to say, it has been established by numerous experiments) that nature favours an  $SU(3)_c$  gauge theory as the mathematical framework with huge phenomenological consequences that shape our current understanding of the strong interactions at a fundamental level. As is shown in Fig. 1, QCD, the theory of strong interactions that Nature favours, is everywhere. The emblematic characteristics of this theory make it unique and powerful:  $\alpha_s$  running with the energy scale of a wide range of experiments and  $SU(3)_c$  well established as the underlying symmetry<sup>1</sup>. This supports the formulation of a microscopic theory in terms of elementary fields (quarks and gluons), whose interactions obey the principles of a relativistic quantum field theory with non-abelian gauge invariance.

Since the early development of QCD, people started to amass tools that would allow to study phenomena at the heart of every experiment conceived, with the highest precision possible. As shown in Fig. 2, nowadays the calculation of observables using perturbative and non-perturbative QCD-inspired tools or QCD-inspired effective theories and their connections, makes it possible to do Standard Model precise calculations and to look for new physics (see for example Ref. [12]). QCD in heavy-ion physics has flourished (via many effective theories and extensions to QFT at finite temperature) that provide new tools to describe the initial conditions of ions moving at relativistic velocities, the intermediate stages where the QGP has formed and the later stages where the high temperature and density of gluons and quarks, become hadrons and jets (see for example Ref. [13]).

<sup>1</sup>For more interesting discussions about the experimental tests of QCD and about how to extract  $\alpha_s$  beyond perturbative QCD, see examples in Refs. [10, 11].





**Fig. 2:** QCD is everywhere. Non-perturbative QCD (npQCD) has a vast reach and here we highlight how Schwinger-Dyson equations (SDEs) are solved for the quark and gluon propagator with dressed versions of the QCD vertices to get hadron masses. Lattice QCD can solve numerically the QCD equations of motion and produce incredible results for hadron spectra and thermodynamic properties of nuclear matter. Perturbative QCD (pQCD) was the historical way to establish QCD as a gauge theory of the strong interactions and provided the tools to predict and describe *jets*, which are now known as the standard signature of hard processes. Throughout the development of hadron and electroweak physics, effective theories have paved the way for modern QCD phenomenology.

## 2.2 The running and confinement

Under *ordinary* circumstances, quarks are confined within hadrons. The color potential between quarks inside hadrons at long distances is of order 1 fm and is linear. Separation of quarks requires an infinite amount of energy, so this potential makes hadrons combine into zero net color charge hadrons. From the perturbative regime, confinement can be seen as a direct consequence of the gluon self-interaction. Given that the strength of the strong interaction is characterized by the coupling constant  $\alpha_s = \frac{g_s^2}{4\pi}$ , we can expose its running (made evident by experiment), by demanding that all relevant observables  $\mathcal{O}(Q^2; \alpha_s)$  constructed with QCD should be renormalization scale independent. This is encoded in the renormalization group equation as

$$\left[ \mu^2 \frac{\partial}{\partial \mu^2} + \beta(\alpha_s) \right] \mathcal{O} \left( \alpha_s, \frac{Q^2}{\mu^2} \right) = 0 \quad ,$$

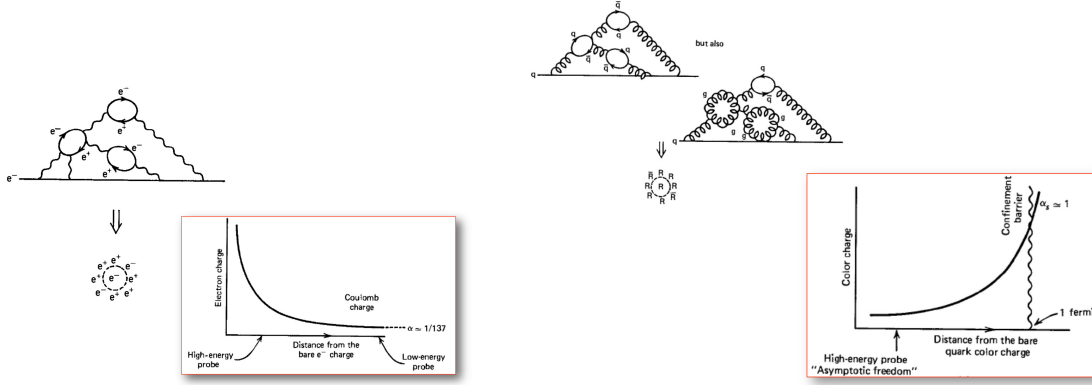
where the *beta function* is defined as

$$\beta(\alpha_s) \equiv \frac{d\alpha_s}{d \log \frac{Q^2}{\mu^2}} \quad . \quad (1)$$

The beta function can be obtained perturbatively to study the running of alpha in the high energy regime (small values of the coupling). At lowest order,

$$\beta(\alpha_s) = -b_0 \alpha_s^2 \quad \text{with} \quad b_0 = \frac{1}{2\pi} \left( \frac{11}{6} N_c - \frac{1}{3} N_f \right), \quad (2)$$

where  $N_c = 3$  and  $N_f$  are the number of color and flavour degrees of freedom of the theory, respectively. We can now solve Eq. 1 and 2, assuming  $b_0 > 0$  (which is true provided  $N_f < 16$ ) and get the famous



**Fig. 3:** On the left we can see the screening effect that quark loops provide for the electric charge, which corresponds to the notion of an effective charge  $e(r)$  that becomes smaller with larger distances. This is encoded in the QED beta function  $\beta(r) = -\frac{de(r)}{d\log(r)}$ , which is positive. On the right we see that the color charge gets screening from the quark loops but antiscreening from the gluon loops, which corresponds to the two contributions with opposite signs to  $b_0$  in the QCD beta function. The QCD beta function is negative and so the effective strong coupling becomes small at short distances [14]. Copyright John Wiley and Sons, re-used with permission.

running of  $\alpha_s$ :

$$\alpha_s(\mu^2) = \frac{1}{b_0 \log(\mu^2/\Lambda^2)} \quad . \quad (3)$$

The parameter  $\Lambda$  describes the boundary condition of the first order differential equation defining the running of  $\alpha_s$ , and corresponds to the scale at which the coupling becomes infinity. Since  $\alpha_s$  is not an observable, it can contain all the terms that are  $\mu$  dependent, in order to achieve a  $\mu$ -independent observable  $\mathcal{O}$  that has a power-series representation in terms of  $\alpha_s$ . As shown in Fig. 3, this behaviour shows the difference between the screening and anti-screening effects of the color charge as compared to the electric charge. The observables that can be described where perturbation theory can be applied are usually those where the transferred momentum satisfies  $Q^2 \gtrsim 1 \text{ GeV}^2$  which marks the separation between the perturbative and non-perturbative regimes.

With the aid of Eqs. 1 and 2, we can define a transferred momentum value  $\Lambda_{\text{QCD}}$  small enough such that the coupling blows up

$$1 + b_0 \alpha_s(\mu^2) \log(\Lambda_{\text{QCD}}^2/\mu^2) = 0 \quad \Rightarrow \quad \Lambda_{\text{QCD}}^2 = \mu^2 e^{-\frac{1}{b_0 \alpha_s(\mu^2)}} \quad , \quad (4)$$

where  $\Lambda_{\text{QCD}}$  is a renormalization scheme dependent quantity. In the  $\overline{\text{MS}}$  scheme and for three active flavors, its value is of order  $\Lambda_{\text{QCD}} \sim 200 - 300 \text{ MeV}$ . This has a huge impact on the phenomenological applications of pQCD, since all dimensionful QCD results with small transferred momentum, scale with  $\Lambda_{\text{QCD}}$ . In this context, we are tempted to say that the existence of this scale is a key ingredient for the emergence of baryon masses and thus of the mass of the visible universe, and hastily conclude that this is what explains hadron confinement. Nevertheless, the emergence of confinement is much more than this, so we must pursue an all encompassing strategy to understand confinement and the breaking of chiral symmetry in QCD, which is also connected to the emergence of hadron masses.

### 2.3 Chiral symmetry in QCD

The QCD lagrangian,

$$\mathcal{L}_{\text{QCD}} = \sum_{\alpha=1}^{N_c^2-1} \sum_{a,b=1}^{N_c} \sum_{i=1}^{N_f} \bar{\psi}_i^a \left( i\gamma^\mu (\partial_\mu \delta^{ab} + i g_s A_\mu^{ab}) - m_i \delta^{ab} \right) \psi_i^b - \frac{1}{4} G_{\mu\nu}^\alpha G_{\mu\nu}^\alpha \quad , \quad (5)$$

is that of a non-abelian gauge theory with a local symmetry group  $SU(N_c)$  with  $N_c = 3$  where the fundamental fields are matter fields  $\psi_i^a$  (quarks) with masses  $m_i$  and massless gauge fields  $A_\mu^\alpha$  (gluons). In this lagrangian the strong coupling  $\alpha_s = g_s^2/4\pi$  enters both in the quark-gluon interaction term  $\bar{\psi}^a \gamma^\mu A_\mu^{ab} \psi^b$  (with  $A_\mu^{ab} = A_\mu^\sigma (\tau_\sigma)^{ab}$ ) and in the gluon self-interaction term through the field strength tensor, defined as  $G_{\mu\nu}^\alpha = \partial_\mu A_\nu^\alpha - \partial_\nu A_\mu^\alpha + g_s f^{\alpha\beta\sigma} A_\mu^\beta A_\nu^\sigma$ . The  $N_f$  quark fields belong to the fundamental representation of the color group ( $N_c$ -dimensional), antiquark fields to the complex conjugate of the fundamental representation (also  $N_c$ -dimensional) and gluon fields to the adjoint representation ( $N_c^2 - 1$ -dimensional), where  $a, b$  run from 1 to  $N_c$  and  $\alpha, \beta, \sigma$  run from 1 to  $N_c^2 - 1$ .

If we imagine a universe in which all the quarks have the same mass  $m$ , then the quark sector of the QCD lagrangian in Eq. 5 is (we focus on the flavour indices and we omit colour indices for the purpose of the discussion)

$$\mathcal{L}_q = \sum_{i=1}^{N_f} \bar{\psi}_i (i\gamma^\mu (\partial_\mu + i g_s A_\mu) - m) \psi_i \quad .$$

This lagrangian is invariant under continuous global *vector flavour* transformations of  $SU(N_f)$  since  $\psi_i \rightarrow \psi'_i = e^{-i\alpha^B (T^B)_i^j} \psi_j$ . In fact using Noether's theorem, we find  $N_f^2 - 1$  conserved currents associated with this symmetry since  $\partial^\mu j_\mu^A = 0$  for

$$j_\mu^A(x) = \bar{\psi}_i(x) \gamma_\mu (T^A)^j_i \psi_j^i(x) \quad , \quad (6)$$

where  $A, B, C = 1, \dots, N_f^2 - 1$ . The charges (or generators) are obtained from the space integration of the density current  $j_0^A$

$$Q^A = \int d^3x j_0^A(x) \quad ,$$

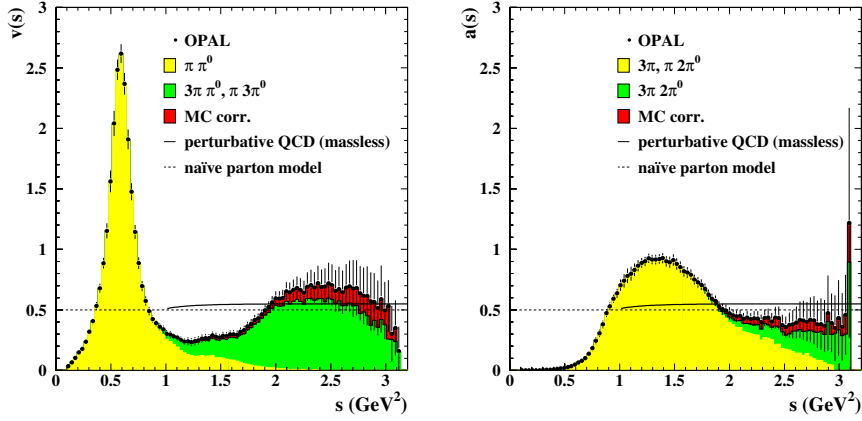
where  $Q^A$  satisfy the  $SU(N_f)$  algebra  $[Q^A, Q^B] = i f^{ABC} Q^C$ . Furthermore, current conservation implies that these charges are constant in time, in the sense that they commute with the Hamiltonian  $[H, Q^A] = 0$  of the theory. The transformation properties of the fields translate into transformation properties of the states (e.g.  $Q^A |\mathbf{p}, i\rangle = (T^A)^j_i |\mathbf{p}, j\rangle$ , suppressing spin indices) and if the vacuum state is also invariant under these transformations, all one-particle states of the fundamental representation multiplet, have equal masses  $m$ . In quantum mechanics, a symmetry of the hamiltonian of the theory reflects in its energy spectrum, most of the time. In this case we have a flavour symmetry transformation that reflects on a degeneracy on the mass spectrum of the theory. This realization of a symmetry is called *Wigner-Weyl mode*. As we will see in what follows, at energies 1.5 – 3 GeV there is data showing parity doubling in the hadron spectrum which could be explained through this flavour symmetry in a universe we imagined with quarks that have the same mass.

If we modify the universe slightly to now consider quark flavors with different masses

$$\mathcal{L}_q = \sum_{i=1}^{N_f} \bar{\psi}_i (i\gamma^\mu (\partial_\mu + i g_s A_\mu) - m_i) \psi_i \quad , \quad (7)$$

the mass term now spoils  $SU(N_f)$  invariance, therefore currents are not conserved ( $\partial^\mu j_\mu^A \neq 0$ ) and instead of Eq. 6, we find

$$\partial^\mu j_\mu^A = -i \sum_{i,j=1}^{N_f} (m_i - m_j) \bar{\psi}_i (T^A)^j_i \quad .$$



**Fig. 4:** OPAL Collaboration measurement of the  $\rho$  (vector meson,  $I^G/J^P = 1^+/1^-$ ) and  $a_1$  (axial-vector meson,  $I^G/J^P = 1^-/1^+$ ) mass spectral distributions. The  $\rho$  spectrum peaks at  $\sim 0.6 \text{ GeV}^2$  while the  $a_1$  spectrum peaks at  $\sim 1.2 \text{ GeV}^2$  [16]. Copyright CERN for the benefit of the OPAL Collaboration.

We could still get back the flavour symmetry in an approximate manner: we could divide quarks into two groups *light quarks*  $u, d, s$  and *heavy quarks*  $c, b, t$ . The mass difference between each group is large ( $\gtrsim 1 \text{ GeV}$ ) but, an approximate flavour symmetry could still be realised for light quarks:  $SU(2)$  for  $u, d$  or  $SU(3)$  for  $u, d, s$ , whereby  $\partial^\mu j_\mu^A \sim 0$ .

Quarks can also be transformed by means of unitary transformations that include the  $\gamma_5$  matrix which are called *axial flavour* transformations. In infinitesimal form these transformations are  $\delta\psi_i = -i\delta\alpha^A (T^A)_i^j \gamma_5 \psi_j$  and the current associated with this symmetry is the axial current

$$j_{5\mu}^A(x) = \bar{\psi}_i(x) \gamma_\mu \gamma_5 (T^A)_i^j \psi_j(x) . \quad (8)$$

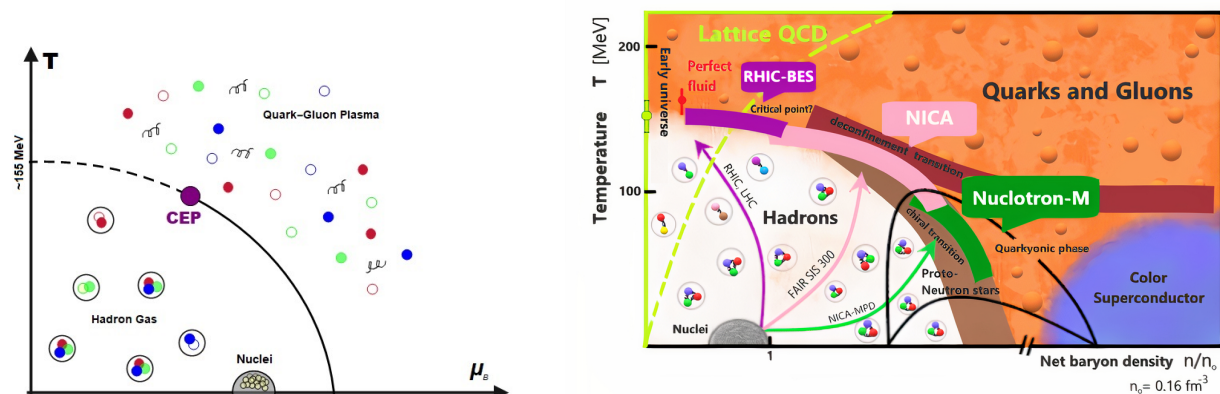
Again, if we consider the universe where the quark masses are equal, then under these transformations the invariance of the lagrangian  $\mathcal{L}_q$  in Eq. 7 is spoiled by the mass as  $\delta\mathcal{L}_q = 2i m \delta\alpha^A \bar{\psi}^i (T^A)_i^j \gamma_5 \psi_j$ . So, contrary to the invariance of the lagrangian under vector flavour transformations which is attained with quarks of the same mass, invariance under axial flavor transformations is spoiled even in a universe with quarks of the same mass. Now, of course, in a universe where quarks are massless, the quark lagrangian is invariant under both the vector flavour and axial flavour transformations, which together form the *chiral transformations*. In this case, both the vector and axial currents are conserved:  $\partial^\mu j_\mu^A = 0, \partial^\mu j_{5\mu}^A = 0$  with the corresponding charges

$$Q^A = \int d^3x j_0^A(x), \quad Q_5^A = \int d^3x j_{50}^A(x) ,$$

where the  $Q^A$  and  $Q_5^A$  satisfy the commutation relations  $[Q^A, Q^B] = if^{ABC} Q^C$ ,  $[Q^A, Q_5^B] = if^{ABC} Q_5^C$  and  $[Q_5^A, Q_5^B] = if^{ABC} Q^C$ .

In this context, the axial charges do not form an algebra but, if we define left-handed and right-handed charges as  $Q_L^A = \frac{1}{2}(Q^A - Q_5^A)$  and  $Q_R^A = \frac{1}{2}(Q^A + Q_5^A)$ , the newly defined charges decouple and operate separately  $[Q_L^A, Q_R^B] = 0$ , generating each an  $SU(N_f)$  group as  $[Q_L^A, Q_L^B] = if^{ABC} Q_L^C$  and  $[Q_R^A, Q_R^B] = if^{ABC} Q_R^C$ . The chiral group is then decomposed into the direct product of two  $SU(N_f)$  groups:  $SU(N_f)_L \otimes SU(N_f)_R$ .

It turns out that in our universe, chiral symmetry is not exact, quark masses break it explicitly. In practise, in the light quark sector chiral symmetry is approximate, so the breaking can be treated



**Fig. 5:** On the left, the phase diagram for QCD matter shows the hadron gas phase and the quark-gluon plasma phase in the temperature ( $T$ ) and baryon chemical potential ( $\mu_B$ ) plane [24]. On the right, the phase diagram for QCD matter shows the hadron gas phase, the quark-gluon plasma phase and the quarkyonic phase in the temperature ( $T$ ) and baryon density ( $n$ ) normalized to the cold nuclei baryon density ( $n_0$ ) plane [25].

as a perturbation. How can we *see* that this symmetry is broken? and, what is the signature of this approximate symmetry? Well, suppose the symmetry is realized in the Wigner-Weyl mode. In the massless quark limit a chiral transformation acting on a massive state changes the parity of the state  $Q_5^A |M, s, \mathbf{p}, +, i\rangle = (T^A)^j_i |M, s, \mathbf{p}, -, j\rangle$ . This means that we should find parity partners for the massive states. Even in the massive quark limit, when we consider the light quark sector, the degeneracy within parity doublets is lifted, but the masses should remain close to each other. We do not observe parity doublets (copies of, say the neutron and the proton) in the hadron spectrum, so under these conditions, the Wigner-Weyl mode realization of chiral symmetry does not happen. We could then consider another mechanism, the spontaneous breaking of chiral symmetry, also referred to as *Nambu-Goldstone mode*. In this case the generators do not annihilate the vacuum ( $Q_5^A |0\rangle \neq 0$ ) and in fact new pseudoscalar states are created:  $Q_5^A |0\rangle = |A, -\rangle$ , with  $A = 1, \dots, N_f^2 - 1$ .

In the massless limit, the charges commute with the hamiltonian therefore these states are  $N_f^2 - 1$  pseudoscalar massless particles, Nambu-Goldstone bosons. If  $N_f = 3$ , this would correspond to eight pseudoscalar bosons with small masses. Indeed, the observed pseudoscalar mesons made up of only  $u, d, s$  quarks form an octet with  $J^P = 0^-$ :  $\pi^0, \eta, \pi^+, \pi^-, K^0, K^+, K^-, \bar{K}^0$ , which are the lightest mesons. Beyond the massless limit, an important measurement involves comparing the experimental mass spectral distributions of parity partner mesons. The main ingredients for these mass spectral distributions are the vector and axial correlators  $\text{Im}\Pi_V(q), \text{Im}\Pi_A(q)$ . Chiral symmetry implies that these correlators should be identical to all orders in perturbation theory and they were measured at LEP with  $\tau$  decays into even and odd number of pions [15, 16]. In particular, the seminal measurement of the  $\rho$  (vector meson,  $I^G/J^P = 1^+/1^-$ ) and  $a_1$  (axial-vector meson,  $I^G/J^P = 1^-/1^+$ ) mass spectral distributions are shown in Fig. 4. The  $\rho$  spectrum peaks at  $\sim 0.6 \text{ GeV}^2$  while the  $a_1$  spectrum peaks at  $\sim 1.2 \text{ GeV}^2$ . This large difference cannot be due to finite current quark masses since, as we already discussed, this produces small differences compared with the masses of the mesons themselves. In this case the differences are of the order of the mass of the  $\rho$ , so the mechanism that could help explain this, is the spontaneous breaking of chiral symmetry in vacuum<sup>2</sup>.

## 2.4 The many faces/phases of nuclear matter

In heavy-ion collisions at relativistic energies, we reach the conditions under which confinement is no more and we obtain a medium of thermally equilibrated hadronic matter where now the quarks and gluons can move freely. In order to describe these conditions, we identify two important parameters: the temperature  $T$  and the baryon number density  $n_B$  (or its conjugate variable, the baryon chemical potential  $\mu_B$ ). Now, as we discussed before, the intrinsic scale in QCD is  $\Lambda_{\text{QCD}} \sim 200 \text{ MeV}$ , so we expect a partons-to-hadrons phase transition around  $T \simeq \Lambda_{\text{QCD}} \sim 200 \text{ MeV}$  and  $n_B \sim \Lambda_{\text{QCD}}^3 \sim 1 \text{ fm}^{-3}$ . Also, the coupling constant runs towards smaller values with increasing energy scale so we can anticipate that both confinement dissolves and this may be somewhat connected to the fact that chiral symmetry breaks. We could argue that QCD matter must undergo phase transitions at high baryon and/or energy densities. The anticipation and excitement of the possibility to create and probe these *novel* (back then!) phases of matter was proposed [18] just a decade after the birth of the hadron Quark Model [19] and two years after QCD was established as a non-Abelian quantum field theory with asymptotic freedom [20]<sup>3</sup>.

In Fig. 5, the panel on the left is a conjectured phase diagram for QCD matter that shows the hadron gas phase and the quark-gluon plasma phase separated by a phase transition line that ends in the *critical end point* (CEP). At  $\mu_B \sim 0$  the transition is more of a continuous crossover, which as mentioned before, must happen around  $\Lambda_{\text{QCD}}$ . Recently, lattice calculations and effective model theoretical estimations have provided constraints and have allowed for a better picture of where the CEP might be located and for a better determination of the phase transitions and this has been matched by the identification of observables relevant for criticality studies [22]. This has been accompanied by several new heavy-ion collision experiments that will be able to scan this phase diagram, beyond what current experiments have achieved<sup>4</sup>. This is schematically shown in the panel of the right in Fig. 5 where now three phases are separated by a chiral phase transition, a deconfinement transition and a pseudocritical crossover line. The experimental scans using heavy-ion collisions are shown for the CERN-LHC, RHIC-BES, FAIR-SIS and JINR-NICA programs and where the CEP location can be studied using overlapping scan programs. The QGP created at RHIC and LHC at high energies, contains almost equal amounts of matter and antimatter, so the corresponding experimental scan, stays close to the vertical axis: low  $\mu_b$  or  $n_B$ , high  $T$ . The planned collisions at FAIR and NICA, will allow to create quark-gluon plasma with an excess of matter over antimatter, and so the corresponding experimental scans will be able to explore the bulk of the phase diagram.

In the next sections, we will go deeper into some of these features of the QCD phase diagram. For now, let us turn to heavy-ion physics and collision experiments and discuss the basics of their well-established usefulness to scan this phase diagram.

## 3 Relativistic heavy ion collisions

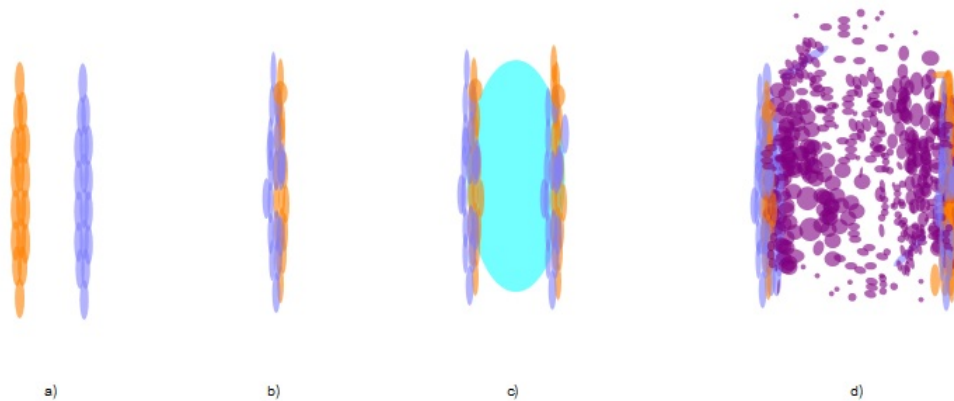
### 3.1 What happens when two ultra relativistic nuclei collide?

In preparation for a collision, the ions are accelerated to relativistic velocities so much so, that the Lorentz contraction factor  $\gamma$  is of the order of 100 at RHIC and of the order of 2500 at LHC. If we choose the laboratory frame as the centre-of-mass (CM) frame in a collider, this means that each incident nucleus is a Lorentz contracted disc of the order of  $14/\gamma \text{ fm}$  for Au and Pb nuclei (14 fm being the typical size of these large nuclei). For example, heavy-ion collisions at BNL and CERN have total collision energy ( $\sqrt{s_{NN}}$ ) per nucleon-nucleon pair in the CM frame of 200 GeV at RHIC and 2.76 TeV at LHC. Fig. 6 shows four stages of the heavy-ion collision and its evolution: (a) first the Lorentz contracted heavy-

<sup>2</sup>For further details see for example Ref. [17]

<sup>3</sup>For more historical insight into the early developments of the QCD phase diagram and the early discussions on the quark-gluon plasma, see for example Ref. [21].

<sup>4</sup>If you are interested on the thermodynamics and statistical mechanics tools to study QCD and hadrons at high temperature, you can start here [23, 32] and go from there.



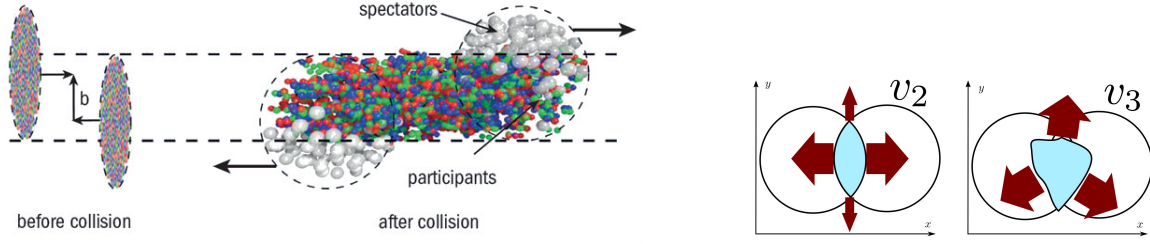
**Fig. 6:** Four stages of a heavy-ion collision and its evolution. (a) Shows the two Lorentz contracted heavy-ions moving towards each other in the CM frame, then (b) shows the overlapping nuclei at the collision stage. (c) Just after the collision the heavy-ions travel across each other and generate a volume of high temperature and energy density and finally in (d), the system expands and cools down, eventually fragmenting into hadrons that travel to the detector [26].

ions come into the collision stage, (b) then they travel across each other and generate (c) a volume of high temperature and energy density and finally the system expands and cools down, (d) eventually fragmenting into hadrons that travel to the detector. The stage highlighted in panel (c), is just after the collision where the conditions are optimal for creating QGP.

The incoming ions are very thin discs of interacting transverse color fields and have on average more quarks than antiquarks with color charges, so that they become sources of colored gluons. When the nuclei pass each other, long color fields fill the space between the receding two Lorentz contracted ions, which makes them lose energy and then they gradually decay into  $q - \bar{q}$  pairs and gluons. Now this initial stage is actually dynamic and non-uniform per event, the initial state energy and momentum distributions fluctuate, so modeling the initial state is the first non-trivial task to be approached using high-density/temperature QCD techniques. Furthermore, we know that most of the incident partons lose energy but only a small—albeit important—fraction of them, will undergo hard interactions. These high- $p_T$  hard-interacting partons will be the seed to produce in-medium jets.

On average the energy density around the midpoint between the two Lorentz contracted ions is far greater than the energy density in a hadron (for example at the LHC with  $\sqrt{s_{NN}} = 2.76$  TeV,  $\langle \epsilon \rangle \sim 12 \text{ GeV/fm}^3 \sim 20 \epsilon_{\text{hadron}}$ ). On the other hand, lattice QCD tells us that QCD matter in thermal equilibrium at  $T \sim 300$  MeV has  $\epsilon \sim 12 T^4 = 12.7 \text{ GeV/fm}^3$ . This means that the medium formed after heavy-ion collisions cannot be described just as a collection of distinct individual hadrons, rather it has to be made out of a high density of quarks and gluons, given the sheer amount of energy density available in a small volume after the collision. Moreover, before the collision the entropy of the two incident nuclei is practically null whereas after the collision the final state can contain as many as  $10^4$  particles, so this entropy is produced quickly, in the initial moments after the collision.

Furthermore, as can be seen on the left panel in Fig. 7, the ions may collide head-on or may only partially overlap at the collision stage, so in the overlap region there are conditions that facilitate QGP formation. This *centrality* of the collision gives also a characteristic initial shape in the transverse plane to the formed medium, that tends to be more lenticular as the collisions are more head-on or central, as shown in the right hand side of Fig. 7. In fact, the deviations from circular symmetry—or lumpiness—in the initial shape of the QGP and fluctuations of the incident nuclei, lead to pressure anisotropies in the



**Fig. 7:** On the left, a schematic representation of the collision *centrality*. The ions may collide head-on or may only partially overlap at the collision stage, so in the overlap region there are conditions that facilitate QGP formation. On the right, the collision centrality promotes a characteristic initial shape in the transverse plane to the formed medium, that tends to be more lenticular as the collisions are more head-on or central [27, 28].

hydrodynamic fluid stage of the plasma. This in turn, leads to anisotropies in the expansion velocity of the products of the hadronized plasma, so that these finally produced hadrons display an azimuthal momentum distribution with the footprint of the initial symmetry or lumpiness. These features can be characterized with the extraction of the *flow coefficients*  $v_n$  of the azimuthal particle distribution as

$$\frac{1}{N} \frac{dN}{d\phi} = \frac{1}{2\pi} \left[ 1 + \sum_n 2v_n \cos(n(\phi - \Psi)) \right], \quad (9)$$

where  $\phi$  is the azimuthal angle and  $\Psi$  is the reaction plane, which is determined by the impact parameter and beam direction in a Glauber model, or is determined event-by-event, after the fact, by the weighted projection of the all the  $p_T$  tracks.

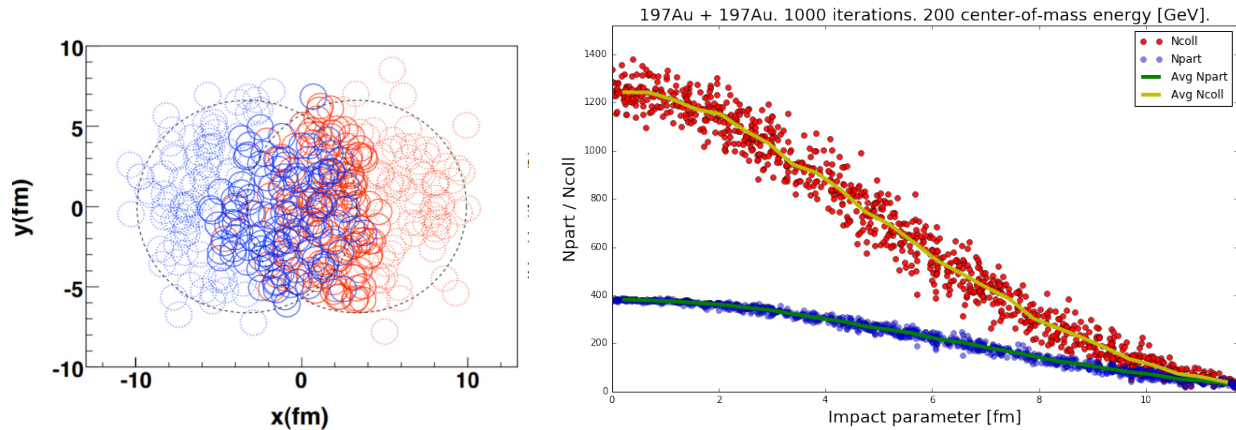
The extreme limit of an off-centre collision, is the case where the nuclei miss each other completely. Still these Lorentz-contracted discs contain charges moving at relativistic speeds, so the electromagnetic fields from each ion do interact. These *ultraperipheral* collisions give rise to  $\gamma + \gamma$  and  $\gamma + A$  interactions, which dominate the total nucleus-nucleus cross section.

The initial state of the collision can also be characterized using models that allow to correlate the centrality of the collision with the number of *participant* nucleons that collide with at least another nucleon, the number of *spectator* nucleons that do not collide and keep on moving along the beam direction and the number of nucleon pairs that collide or *binary collisions*, assuming transparency of the collision. At early times, simulations show that both spectators and participants create intense but short-lived magnetic fields in the collision zone. The effects of these primordial magnetic fields in the evolution of the collision and the available observables to measure these effects, are a subject of current interest in the community.

### 3.2 How do we access the different stages of a heavy-ion collision?

Even though we can use models to simulate the initial condition of a heavy-ion collision as we described before, we only have two quantities under our direct control: which two species of nuclei we accelerate in preparation for the collision and with what amount of energy we make them collide. So with the help of many experiments we have to observe the outcome of the collisions, event by event, and infer the transverse distance between the centres of the colliding nuclei (*impact parameter*  $b$ ) and the location and motion of partons in nucleons and of nucleons in the colliding nuclei. We can select events with a narrow distribution of impact parameters or *centrality bin* only after the experiment or simulation is done, and we can match this result with the impact parameter distributions expected from the initial condition models. Also, we need nuclear and particle physics studies to support these initial conditions models, e.g. the nuclei can be reasonably well approximated by a collection of nucleons, distributed on average according to three dimensional distributions that provide the usual charge distributions for each type of





**Fig. 8:** On the left, modelling the colliding nuclei as made of transparent spheres where the dashed circles represent spectator nucleons that travel down the beam pipe and the solid circles represent the wounded or participant nucleons which collide with at least one other nucleon. On the right, the number of binary collisions and the number of participants are plotted for 1000 events in Au–Au collisions at 200 GeV with respect to the impact parameter, using MC Glauber [29].

ion. Moreover, the average quark and gluon content of the nucleons in nuclei can be given in terms of parton distribution functions (nPDF) which can be sometimes approximated by those of free nucleons (PDF). In fact, the measured hadronic total inelastic cross section can be used to model the nucleons in nuclei as hard spheres with a radius that depends on the collision energy, so that the initial condition of the heavy-ion collision and the subsequent stages can be simulated with well known phenomenological inputs such as  $\sigma_{\text{inel}}^{pp}(\sqrt{s})$  for  $pp$  collisions.

For example, we can model colliding nuclei as made of  $A$  transparent spheres of radius  $\sqrt{\sigma_{\text{inel}}^{pp}/4\pi}$ ,  $A_L$  and  $A_R$  being the number of nucleons inside the left- and right-moving nuclei. Then, as shown in Fig. 8, in every collision, there are  $N_{\text{spec}}$  spectator nucleons (dashed circles) that travel down the beam pipe,  $N_{\text{part}}$  wounded or participant nucleons which collide with at least one other nucleon (solid circles), so that  $N_{\text{spec}} + N_{\text{part}} = A_L + A_R$  and  $N_{\text{coll}}$  is the number of encounters between nucleons in  $A_L$  and in  $A_R$ . So a toy event with a “nucleus” of 8 nucleons lined up in a row that collides head-on ( $b = 0$ ) with a “nucleus” of 4 nucleons lined up in a row would have  $N_{\text{part}} = 12$ ,  $N_{\text{coll}} = 32$ ,  $N_{\text{spec}} = 0$ . But in real central heavy-ion collisions, the nucleons at the centre of nucleus will hit about 12 nucleons on average, but less if it is located at the edge of the collision. So we expect that in general  $N_{\text{coll}} \gg N_{\text{part}}$ , with a more marked difference for the most central collisions. Indeed, this can be seen in the right panel of Fig. 8 where the values of  $N_{\text{coll}}$  and  $N_{\text{part}}$  are plotted for 1000 events in Au–Au collisions at 200 GeV with respect to the impact parameter. These results were obtained using the so called *Glauber model* that is one of the simplest tools that can be used to simulate the initial conditions of heavy-ion collisions, upon which more sophisticated approaches can be constructed where now the Glauber modelling serves as a benchmark.

### 3.3 How do we study the quark-gluon plasma in heavy-ion collisions?

The QGP is made of quarks and gluons that are strongly coupled. They form a collective medium that expands and flows as a relativistic hydrodynamic fluid, with low viscosity to entropy density ratio  $\eta/s \gtrsim 1/4\pi$ , during the initial stages after the collision  $\Delta t \sim 1 \text{ fm}/c$  (in the fluid rest frame). Just after the Lorentz contracted heavy-ions collide, droplets of QGP fluid form and flow hydrodynamically: the initial high pressure generated by the collision, drives fluid motion, expansion, and subsequent cooling, until the energy density of the droplet is of the order or smaller than that of a hadron. This is when the system hadronizes and a mist of hadrons continues to expand, allowing for hadrons to scatter off each

other a few times. Finally these hadrons lose sufficient energy to stop scattering and just stream away freely until they are collected by the detector.

In order to study the quark-gluon plasma in heavy-ion collisions we need adequate experimental probes for the different stages of the formation and evolution of the QGP and a model that incorporates the theoretical ideas relevant for different time and energy scales. Some of the most prominent experimental probes that have provided key information about the different stages of the collision are hadronic probes, such as the relative abundances of pions and kaons, electromagnetic probes such as dilepton and photon production, jet quenching and  $J/\Psi$  resonance production and decay [2,30]. In the last decade, the nuclear physics community has converged towards an initial consensus on a model with marked epochs in the time evolution of heavy-ion collisions. This model not only helps with the construction of useful physical observables, but it also guides the design of physics plans for current and future experiments. The main features of this model of the stages in the evolution of heavy-ion collisions are as follows [31]:

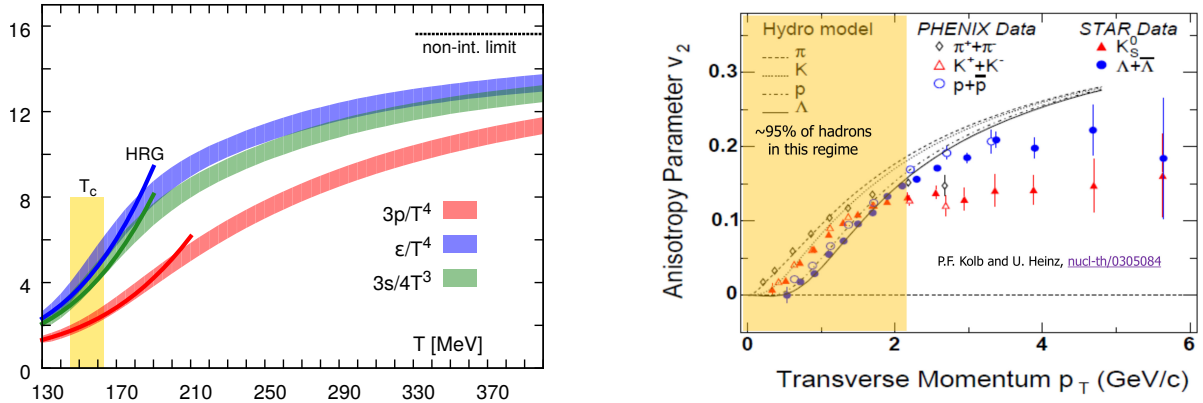
- I. The Lorentz-contracted nuclei collide with short transversal time ( $\ll 1\text{fm}/c$ ). The energy deposited into the medium mainly through gluon field interactions, creates an inhomogeneous initial condition in the transverse plane.
- II. At this stage matter is out-of-equilibrium and needs time to equilibrate, so the expansion occurs at almost the speed of light in both the longitudinal direction and radially in the transverse plane.
- III. Now matter is nearly equilibrated and behaves as a fluid (collective modes). We can use viscous relativistic hydrodynamics with an equation of state from lattice QCD to describe the QGP. The small deviations from equilibrium happen because of shear/bulk viscous medium.
- IV. Now, this fluid cools down to about the cross-over temperature of  $T \sim 170\text{ MeV}$  and breaks up into hadrons. This process of hadronization of fluid modes can be modelled using e.g., Cooper–Frye hadronization.
- V. Finally, the hadrons scatter inelastically until they reach a *chemical freeze-out*, where no more decays or secondary production is possible and continue to scatter elastically until *kinetic freeze-out*, where their momentum distribution is set. This then produces hadrons in a final-state with momenta as measured experimentally.

This model of particle production in heavy-ion collisions with an intermediate epoch during which a hydrodynamic fluid forms and expands, is quite different from the ones used for particle production in elementary hadron collisions in which only a few new particles are created. In fact, the question of whether we can create droplets of QGP in  $pp$  collisions has caused a big excitement and the possibility to have collective effects in relativistic collisions of small systems is now being pursued with both theoretical and experimental approaches<sup>5</sup>.

#### 4 The quark-gluon plasma: extreme QCD

One way to start exploring the basic properties of the QGP is to observe the hadron mass spectra. At a first glance, we can see that there is a “QCD mass gap”: the pion mass is separated from the rest of the hadron masses. So we could imagine a hadronic phase of nuclear matter (pion gas) where the relevant number of degrees of freedom are 3 ( $\pi^-$ ,  $\pi^0$ ,  $\pi^+$ ) and the rest of the hadrons are basically “excited hadron states”. Then the corresponding quark-gluon phase (quark-gluon gas) would be one where the relevant degrees of freedom are 37 for two light-quark flavours ( $2_{\text{spin}} \times 8_{\text{colour}} = 16$  for gluons,  $(7/8) \times 2_{\text{spin}} \times 2_{\text{flavour}} \times 2_{q\bar{q}} \times 3_{\text{colour}} = 21$  for quarks). So the quark-gluon phase has more than ten times the number of degrees of freedom than the hadron phase. If we use basic thermodynamics, we can do back-of-the-envelope estimates of the pressure of a pion gas  $p_\pi$  and compare it with that of a bag

<sup>5</sup>For a recent review on small system collectivity effects in relativistic nuclear and hadron collisions see Ref. [31].



**Fig. 9:** On the left, thermodynamic properties calculated with Lattice QCD [33], pressure, energy and entropy densities. On the right,  $v_2$  for most hadrons at RHIC which uses a hydrodynamics description of the QGP [34, 35].

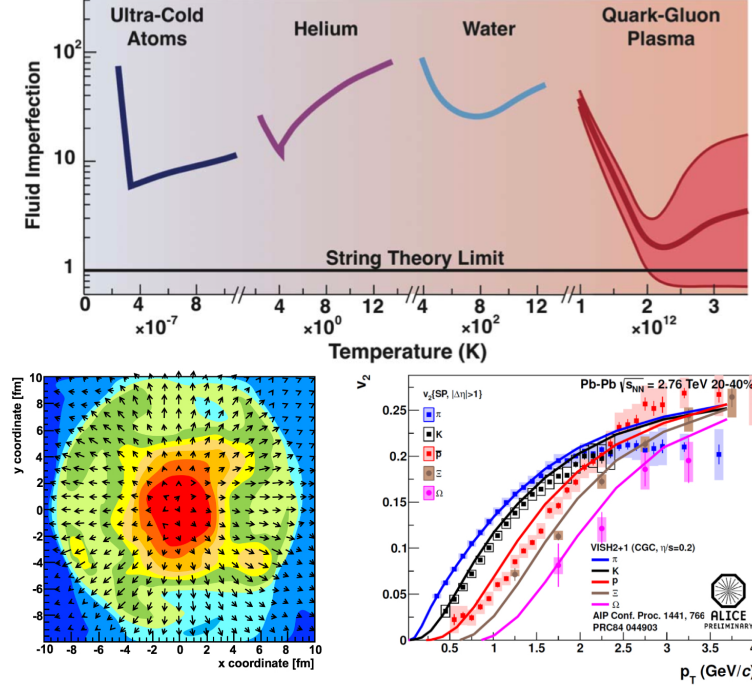
of quarks and gluons  $p_{\text{QGP}}$ . We will find out that

$$p_{\pi}(T) = 3 \frac{\pi^2}{90} T^4 \ll p_{\text{QGP}}(T) = 37 \frac{\pi^2}{90} T^4 - B \quad ,$$

so even for a bag constant  $B^{1/4} \sim 200$  MeV, Nature prefers the system with higher pressure for increasing temperatures ( $T > 0.68B^{1/4}$ ). So, under these simple arguments a transition from the hadron phase to the deconfined phase is expected [32].

Lattice QCD has been able to provide the results not only for the pressure  $p$ , but also the energy  $\varepsilon$  and entropy  $s$  densities, for the quark-gluon phase. The left panel on Fig. 9 shows these three thermodynamic quantities using solid curves in a temperature range that starts below the critical temperature and goes higher than the temperatures achieved at the LHC. We can see, for example, that the pressure curve (red) is actually  $3p/T^4$  and the values at both ends of this curve, compare reasonably well with the ones in our back-of-the-envelope estimate, which are for the hadron phase  $3\pi^2/30$  and  $37\pi^2/30$  for the quark-gluon phase. Also in this figure, it is clear that the values of these thermodynamic properties increase dramatically for temperatures  $T \gtrsim 140$  MeV and a smooth crossover is apparent at  $T_c = (154 \pm 9)$  MeV. This figure also has the Hadron Resonance Gas (HRG) model curves at low temperatures, which predicts an exponential increase of states. Finally, the non-interacting ideal gas limit is shown at high temperature which would take an energy density of  $\varepsilon/T^4 = 95\pi^2/60$  to have an ideal non-interacting gas of quarks and gluons. Here we could be tempted to study QGP dynamics with Lattice QCD results above the cross over  $T_c$ , but this conclusion would be based on a *static* picture, since as we just discussed, thermodynamically this system approaches that of ideal non-interacting gas of quarks and gluons. Now, beyond the static picture, dynamically things look different: the QGP phase is a system more like a viscous hydrodynamic fluid made of strongly interacting quarks and gluons with large cross sections. For example, most of the hadrons produced in heavy-ion collisions at RHIC are best described with a hydrodynamical model of the QGP that expands and generates anisotropies in the transverse momentum distribution, as is shown on the right in Fig. 9 for the case of  $v_2$  (second coefficient in r.h.s. of Eq. 9), the so called *elliptic flow*.

One important and seminal work for the applicability of viscous hydrodynamics to the description of the QGP, was reported by Danielewicz and Gyulassy back in 1985 [36]. They came up with a nice way to have bounds for the size of the viscosity in these systems. Since the QGP phase is a system made of strongly interacting quarks and gluons with large cross sections, then we can estimate the viscosity  $\eta$  as the transverse momentum in-medium diffusion coefficient, in terms of the number density  $n$ , average



**Fig. 10:** In the upper panel, comparison of the ratio  $4\pi k \eta/\hbar s$  in natural units [38]. On the lower panels, a viscous hydrodynamics calculation for a collision between heavy ions, as a one-time snapshot at  $t = 5 \text{ fm}/c$  [2,31], where the colour scheme indicates the temperature and the arrows indicate the fluid velocity.

momentum  $\langle p \rangle$ , mean-free path  $\lambda$  and cross section  $\sigma$  of the medium constituents

$$\eta \sim \frac{1}{3} n \langle p \rangle \lambda \quad \text{or} \quad \eta \sim \frac{\langle p \rangle}{3\sigma} \quad , \quad (10)$$

where on the right hand side, we have taken the mean-free path to be  $\lambda = (n\sigma)^{-1}$ , so that we can start to see the immediate implications on the small size of the viscosity for large cross sections and low average momentum. If we consider the QGP regime achieved with large chemical potential and/or large temperature, plus the fact that quarks and gluons are almost free under these conditions, then naively we expect small particle correlations, so that the mean-free path is small compared with the system size.

In order to give a better estimate of the viscosity, we can build a quantity that provides a comparison between the viscosity as a *drag force* and the *inertial mass* of the volume element of fluid, say

$$\frac{\text{drag force}}{\text{inertial mass}} \rightarrow \frac{\eta}{\varepsilon + p} = \left(\frac{\eta}{s}\right) \frac{1}{T} \quad ,$$

where the energy and pressure densities can be related thermodynamically to the entropy density  $s$  as  $\varepsilon + p = s T$  at null chemical potential  $\mu = 0$ . Now, we can use the first expression in Eq. 10 to give a lower bound of  $\eta/s$ : for a system made out of massless non interacting quanta we know that  $s = 4n$  and we can estimate  $\langle p \rangle \lambda \gtrsim \hbar$  so finally the lower bound is  $\frac{\eta}{s} \gtrsim \frac{\hbar}{4 \times 3}$ . Later, using string theory, Kovtun and collaborators [37] obtained the bound  $\frac{\eta}{s} \gtrsim \frac{\hbar}{4\pi}$ .

Soon after, the QGP was popularized as the fluid that might have the smallest viscosity or *fluid imperfection*—a measurement of the departure from ideal hydrodynamics—with the ratio  $4\pi k \eta/\hbar s$ . In natural units then the QGP would have near-fluid perfection in the lower bound for  $\eta/s$ , as is shown in the upper panel of Fig. 10 in comparison with ultra-cold atoms, helium and water. So, the extreme conditions that prevail in heavy-ion collisions with extreme expansion rates for the fire ball and the smallness of  $\eta/s$

warrants an extended period of applicability of relativistic viscous hydrodynamics (vHydro) between the pre-equilibrium and the final decoupling epochs. This has resulted in an impressive success in the use of vHydro to describe and predict central observables that help characterize the QGP. For example, the lower panel of Fig. 10 shows the elliptic flow coefficient  $v_2$  for identified hadrons in Pb–Pb collisions at the LHC, compared to hydrodynamic calculations for  $\eta/s = 0.2$ . The strategy used to obtain the curves shown in this figure, is the numerical solution of coupled relativistic hydrodynamic equations, together with a hadronization and flow analysis as was described in the previous section. There are several very successful public and private codes that solve 2 + 1 and 3 + 1 vHydro that can be coupled to codes that generate heavy-ion initial conditions and to codes that help in the later stages of the evolution (freeze-out and hadronization). Even a simple linearized version of vHydro equations

$$\partial_\mu \delta T^{\mu\nu} = J^\nu; \quad \partial_\mu T_0^{\mu\nu} = 0 \quad , \quad (11)$$

where the medium total energy-momentum  $T^{\mu\nu} = T_0^{\mu\nu} + \delta T^{\mu\nu}$  with  $T_0^{\mu\nu}$  the energy-momentum for underlying medium in equilibrium and  $\delta T^{\mu\nu}$  a *small* perturbation ( $J^\nu$  being the source of disturbance), can be solved analytically for certain source models and has provided physical insight into the effects of hydro-modes in experimental observables [40].

From the point of view of a microscopic theory, the viscosity is a transport coefficient that quantifies the inefficiency of dissipation in the medium, which would translate into a persistent anisotropy in the energy-momentum tensor  $T_{\mu\nu}$ . Then in the perturbative picture, the QGP would be made up of quasiparticle excitations with momenta the size of the dominant scale, the temperature of the system. So the anisotropy in the energy-momentum tensor, arises from the momentum distributions of these quasiparticles. Then, the scattering and splitting of these constituents will drive the system into its equilibrium value. So in this perturbative approach the challenge is to determine the form of the source of anisotropies by the calculation of the collision operator that enables this system relaxation. At fixed order in perturbation theory, this implies solving a Boltzmann equation using finite temperature techniques

$$(\partial_t + \hat{\mathbf{p}} \cdot \nabla_{\mathbf{x}})f(\mathbf{x}, \mathbf{p}, t) = -C^{2 \rightarrow 2}[f] - C^{1 \rightarrow 2}[f] \quad , \quad (12)$$

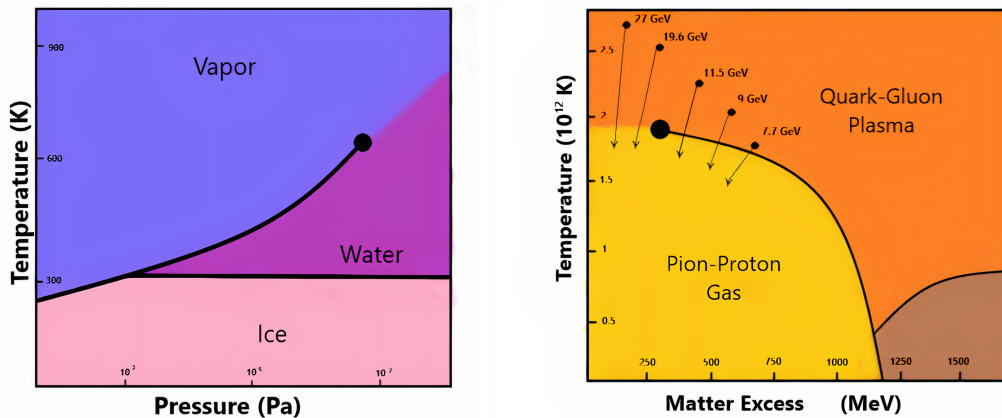
where  $C_s^{2 \rightarrow 2}$  is a scattering operator,  $C_s^{1 \rightarrow 2}$  is a splitting operator and  $f(\mathbf{x}, \mathbf{p}, t) = dN/d^3\mathbf{x}d^3\mathbf{p}$  is the phase space distribution for either quarks, antiquarks or gluons. The solution of this equation then, is used to obtain the transport coefficients, via the vHydro equations [41]. So a full hydrodynamical study, will contain the non-linearized version of Eq. 11 coupled to a Boltzmann equation similar to Eq. 12<sup>6</sup>.

## 5 QCD phase diagram and criticality signatures

### 5.1 Phase transition, tradition vs. praxis

In general, *phase transitions* refer to the transformations of a substance -from one matter state to another- as a result of variations of external conditions such as pressure, temperature, etc. Typically, when the substance goes through a phase transition, there are certain quantities that often change in a discontinuous manner. Perhaps the most familiar phase transitions of a substance, are those of water: in our daily lives we use water in the liquid, solid and gas phases and under certain conditions of, say, pressure and temperature, we know which phase to expect for water. This information is summarized in phase diagrams in Fig. 11, where the solid black lines are the values of temperature and pressure under which the phase transitions occur. Notice there is a *triple point* -three phases coexist- where the phase transition lines intersect and there is a black circle for the *critical end-point* (CEP) where the phase transition line ends. For water, the CEP is known experimentally but, for the QGP, the CEP is predicted theoretically but not known experimentally. The experimental discovery of this CEP is part of current and future heavy-ion collision experiments and the theoretical connection between regions of criticality on a phase diagram and observables at the experiment, are a subject of much interest.

<sup>6</sup>For further details on the applicability of relativistic viscous hydrodynamics and for a complete description of the implementation of vHydro beyond the linearized regime, see for example Ref. [39].



**Fig. 11:** On the left, the phase diagram of water which has continuous lines representing the phase transitions and a critical end-point. On the right, the phase diagram of nuclear matter also shows the phase transitions with continuous lines and a critical end-point, where the cross-over region begins. [42]

Since phase transitions are the result of interactions of a large number of particles—so size of the system and number of particles are relevant to the discussion- in technical terms, they occur when the free energy—the energetic balance between changes in internal energy and changes in temperature and entropy—is non-analytic (one of its derivatives diverges) for some values of the relevant thermodynamical variables of the system. So, on the phase transition lines the free energies in both phases coincide. For certain systems such as nuclear matter, it is possible to change the state of the substance, without crossing a phase transition line. In this case, the thermodynamic conditions define a *cross-over*, rather than a phase transition. On the right panel of Fig. 11, there is a cross-over region for nuclear matter that transforms between the hadron phase and the QGP phase, for low baryon chemical potential and high temperature.

Again, in general, phase transitions are characterized as either discontinuous or continuous. The former are a result of a discontinuous change in entropy at a fixed temperature (the change in entropy corresponds to  $latent\ heat\ L = T\Delta S$ ) and typical examples are solid-liquid and liquid-gas transitions at temperatures below the critical temperature. The latter involve a continuous change in entropy, which means there is no latent heat in the process, such as liquid-gas transitions at temperatures above the critical temperature. In practise, sometimes we do phase transition classification à la Ehrenfest: the order of a transition is the order of the lowest derivative of the free energy, which shows a discontinuity. For example in boiling water, a first order phase transition is that of a discontinuity in the density, the derivative of free energy with respect to chemical potential, and would be labeled a discontinuous phase transition under the previous criteria. In a second order phase transition for the magnetization of a ferromagnet, the derivative of the free energy with respect to the external field is continuous, but the *susceptibility* (the derivative of the magnetization with respect to the external field) is discontinuous and this would be labeled a continuous phase transition with the previous criteria. Finally, the modern classification of phase transitions have a more subtle description of them: the first order phase transitions involve latent heat, so the system absorbs or releases heat at a constant temperature, its phases coexist so only some parts of the system have completed the transition; the second order phase transitions are simply continuous transitions and then, the systems' susceptibilities diverge and the characteristic correlation lengths become large.

## 5.2 Hadron spectra and Hagedorn temperature

In order to study the impact of the nuclear matter phase transitions in the hadron spectra, we can start with a simple model for the hadron gas phase: an ideal gas of identical neutral scalar particles of mass  $m$ <sup>7</sup>. The grand canonical partition function for these neutral scalar particles, contained in a box volume  $V$  at temperature  $T$ , assuming Boltzmann statistics is  $2\pi^2 \ln \mathcal{Z}(T, V) = VTm^2 K_2(m/T)$ . So for temperatures much greater than the mass of these particles, the energy density of the system grows as  $T^4$ , the particle density grows as  $T^3$  and the average energy per particle grows as  $T$ . The more energy that is available to the system, the higher the temperatures are accessible to it and the more energetic its constituents. Now, if we allow these hadrons to interact and we allow for resonance formation, we can use an improved model of an ideal gas like before, but now with a probability to form certain resonances,

$$\ln \mathcal{Z}(T, V) = \sum_{i=0} \frac{VTm_i^2}{2\pi^2} \rho(m_i) K_2\left(\frac{m_i}{T}\right) \quad , \quad (13)$$

where we sum over all possible resonances ( $i = 0$  being the ground state) with weights relative to the ground state, encoded in  $\rho(m)$ . In the early days of hadron discoveries, R. Hagedorn [45] collected the data on the hadron spectra and concluded the hadron spectra could be well described with this model with  $\rho(m) \propto \exp(m/T_H)$  where the *Hagedorn temperature* is  $T_H \simeq 0.19 \text{ GeV}$ . This hypothesis of the experimental growth with mass of the number of hadronic resonances is at the core of the phenomenology of particle production as a direct consequence of the phase transition between the QGP and the hadron gas. The amount of data we have accumulated since Hagedorn's time, means that we can revise this model, use it as a guiding light in the search of possible missing states and explore the effects this would have for hadron flow observables [46]. Now, if we are in a regime where not only do we have a high temperature  $T$  but also a finite baryon chemical potential  $\mu_B$ , then this exponential growth is in direct competition with a Boltzmann factor and  $T_H$  is now a fixed temperature limit  $T = (1 - \mu_B/m) T_H$ , so in this case, the momentum of the constituents of this hadron gas, does not continue to increase and more species of heavier particles can be produced.

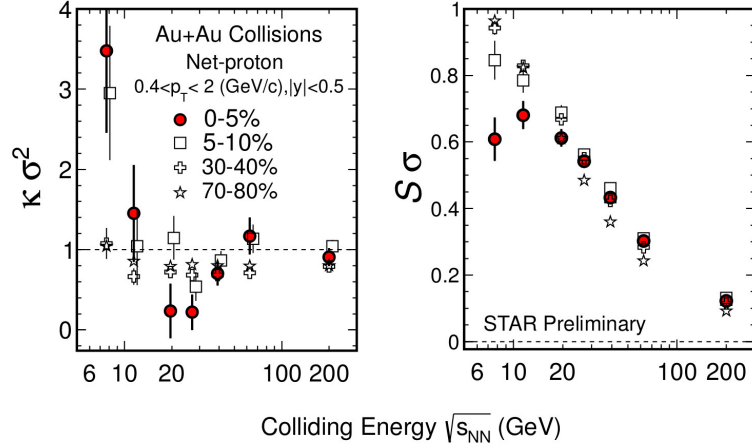
## 5.3 Chiral transition and hadronization

In section 2.3, we laid out the ideas about how we should expect a phase transition between a state with light current quarks and state with heavy constituent quarks a *chiral phase transition*. In other words, we can study the QCD phase diagram at finite  $T$  and  $\mu_B$  from the point of view of chiral symmetry restoration, because quarks turn bare in hot and/or dense energetic nuclear matter. For these studies the typical order parameter to follow in the search of a transition is the *chiral condensate*  $\langle \bar{\psi}\psi \rangle$ , a quark-antiquark bound state with Bose-Einstein statistics. The easiness of this condensate to form in vacuum is characterized by the value  $\langle \bar{\psi}\psi \rangle_0 = -(0.24 \text{ GeV})^3$ , which sets the scale for the critical temperature of chiral restoration. Chiral perturbation theory provides a strategy to calculate the value of this condensate at finite -but small- temperature and chemical potential and the result shows that the condensate melts, so there is a set of values of temperature and chemical potential that indicate a chiral phase transition in the QCD phase diagram.

In order to probe this chiral phase transition in experiments, we should keep tabs of the particle abundances and make sure that we have a statistical model that includes variations of these abundances due to a high particle density in the phase transition region<sup>8</sup>. For central collisions where  $\mu_B \sim 1/\sqrt{s_{NN}}$  and for current collider based experiments where the highest energies are reached—such as the LHC and RHIC—the baryon chemical potential associated to the reaction is the smallest. In these experiments there is then an increase of the *degree of transparency* of the colliding nuclear matter: the energy injected to the interaction zone, produces roughly an equal number of particles and antiparticles. New

<sup>7</sup>For a review on the thermodynamics and hadron spectra of QCD matter, see Ref. [44].

<sup>8</sup>For a review of the statistical-thermal model of particle production in heavy-ion collisions is Ref. [47].



**Fig. 12:** Skewness and kurtosis of net-proton number is reported by the STAR Collaboration [51].

experiments—such as NICA and FAIR—have collisions with lower centre-of-mass energies, a smaller degree of transparency to create an interaction zone that is rich in baryons.

#### 5.4 The critical end-point (CEP)

So far, we have complemented the discussion that we started back in section 2.4, on the theoretical ideas and observed phenomena that motivate a conjectured QCD phase diagram for nuclear matter under extreme conditions of temperature and density, such as the ones in Fig. 5. The main features of the conjectured QCD phase diagram can be summarized as follows: there is a crossover for  $\mu_B = 0$ , there is a first order phase transition that turns into a second order phase transition somewhere in the middle of the phase diagram and there is a CEP somewhere in the middle of the phase diagram, where the crossover becomes a first order phase transition line<sup>9</sup>.

One way to find experimental evidence of the CEP in relativistic heavy-ion collisions, is to look for event-by-event fluctuations of conserved quantities. The possibility to detect *non-Gaussian fluctuations* in conserved charges is one of the tools thought to be sensitive to the early thermal properties of the medium created at heavy-ion collision experiments. In principle, the amount of any conserved charge  $Q$  in a given volume of phase space  $V$  is an integral over the volume of the charge density  $n(x)$ , but when the measurement of the charge is performed event-by-event for a volume in a thermodynamical system, we expect thermal fluctuations of  $Q$ . These fluctuations can be quantified with the help of moments of the charge distribution functions e.g. the variance of  $Q$  given in terms of a *correlation function* as

$$\langle \delta Q^2 \rangle_V = \langle (Q - \langle Q \rangle_V)^2 \rangle_V = \int_V dx_1 dx_2 \langle \delta n(x_1) \delta n(x_2) \rangle \quad .$$

In general for a conserved quantity  $c$ , the *cumulants*  $\langle x^n \rangle_c$  associated to a probability distribution function  $\mathcal{P}(x)$  of a stochastic variable  $x$  are

$$\langle x^n \rangle_c = \left. \frac{d^n}{d\theta^n} \ln G(\theta) \right|_{\theta=0} \quad ,$$

where  $\langle x \rangle_c = \langle x \rangle$ ,  $\langle x^2 \rangle_c = \langle x^2 \rangle - \langle x \rangle^2 = \langle \delta x^2 \rangle$ ,  $\langle x^3 \rangle_c = \langle \delta x^3 \rangle$ ,  $\langle x^4 \rangle_c = \langle \delta x^4 \rangle - 3\langle \delta x^2 \rangle^2$ , etc. For a thermodynamical system the partition function  $\ln \mathcal{Z}$ , is the moment generating function - the so called *cumulant generating function*-  $\ln G$ . Since a conserved quantity will be connected to the volume  $V$  of the

<sup>9</sup>For a review on the location of the CEP in the QCD phase diagram see Ref. [48] and references therein.



system in a grand canonical ensemble framework, cumulants are extensive quantities. Notice also that non-Gaussian fluctuations will be driven by non-vanishing higher order cumulants. Since the asymmetry and sharpness of a distribution function can also be established through the *skewness*  $S$  and *kurtosis*  $\kappa$ , then there is a direct connection between these two quantities and the cumulants: when the stochastic variable  $x$  is normalized to the square root of the variance  $\sigma$ , the skewness and the kurtosis are given as the third and fourth-order cumulants:  $S = \langle \tilde{x}^3 \rangle_c$ ,  $\kappa = \langle \tilde{x}^4 \rangle_c$ . If we can describe the fluctuations of conserved charges in heavy-ion experiments using hadron degrees of freedom, then the cumulants should be consistent with those of the Hadron Resonance Gas model (HRG) and so deviations from this model are used as experimental signals of the emergence of non-hadron or non-thermal physics. Near the CEP, higher order cumulants change sign and are sensitive to the increase of correlation lengths. In particular when passing through the CEP in a beam energy scan in heavy-ion collision experiments, the fluctuations of multiplicities or mean transverse momenta of particles are a manifestation of higher non-Gaussian cumulants with a non-monotonic behavior of the correlation length. For example, the top panels in Fig. 12 show the skewness and kurtosis for net-proton number as reported by the STAR Collaboration for a beam energy scan from 6 GeV to 200 GeV for Au–Au collisions at RHIC. In peripheral and mid-central collisions, the  $\kappa\sigma^2$  values are close to 1 and the  $S\sigma$  show strong monotonic increase when the energy decreases, giving the first hints of criticality.

## 6 Heavy-ion collision physics: season finale and season premiere

The experiments that have established the production of the QGP -RHIC at BNL and LHC at CERN- and are still providing data to characterise it, represent the golden era of heavy-ion collisions. Now, a new era begins with new experiments—NICA at JINR, FAIR at GSI and RHIC at BNL—that are going to provide data to scan regions of the QCD phase diagram that we could not access before. Together with heavy-ion physics, astrophysics and condensed matter will continue to enrich the description of the phases of nuclear matter from different perspectives. Even though the collisions are at lower energies, this poses new technological challenges both for the preparation of the ion-beam and for the design of detectors that can have adequate coverage to capture the products of the QGP evolution. Throughout these lectures, we have shown several examples of the theoretical and experimental efforts that provide the most comprehensive picture, up to today, of the QGP and its evolution. In this section we highlight certain aspects of heavy-ion phenomenology that have resolved some issues in the past, but that also present new challenges in this new experimental era.

### 6.1 $R_{AA}$ , jet quenching and correlations

In order to quantify the effect of the medium created in heavy-ion collisions in the production of particles, we use the *nuclear modification factor*  $R_{AA}$ , which is the ratio of the multiplicity of particles produced in  $A - A$  collisions with respect to the multiplicity in  $pp$  collisions, normalized to the average number of binary collisions

$$R_{AA}(p_T) = \frac{1}{\langle N_{\text{coll}} \rangle} \frac{dN^{AA}(p_T)}{dp_T} \cdot \frac{dN^{pp}(p_T)}{dp_T} .$$

So, naively, we expect this ratio to be closer to one for primordial photons for which the QGP should be transparent, and perhaps smaller than one for hadrons that form at different stages of QGP evolution. In the left panel of Fig. 13 we can see the data for  $R_{AA}$  for photons and hadrons in the most central Au–Au collisions at  $\sqrt{s_{NN}} = 200$  GeV, as reported by the PHENIX collaboration [49]. Indeed hadrons show large suppression—*quenching*—for a wide range of  $p_T$  and photons mostly do not. Looking closely at the plot, we can see that for  $p_T < 5$  GeV there is an excess of photons in Au–Au collisions and there is a slight recovery on the  $p_T > 5$  GeV observed hadron suppression. At the LHC this translates into *jet quenching*, a large -by a factor of 2 for most central collisions- suppression of single-inclusive jets produced in Pb–Pb collisions, when compared to  $pp$  collisions, as shown in the right panel of Fig. 13.

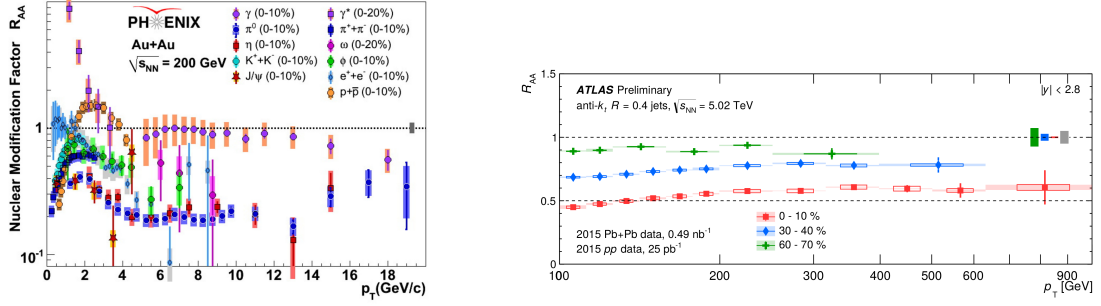


Fig. 13: Hadron species show *quenching* at RHIC [49] and single-inclusive jet quenching at LHC [50].

Jet quenching in heavy-ion collisions is a tool to understand the formation and evolution of in-medium jets, since there is an enhancement of soft activity at the edges of the jet and far away from the jet. So jet physics in heavy-ion collisions represents one of the most important tools to study both the process of equilibration -through the interactions between the jet and the medium- and gives access to the scale dependence of medium properties. The missing  $p_T$  in an event, due to jet quenching, has made it possible to perform in-medium jet-tomography [53] and to inspire new developments for in-medium fragmentation and jet-formation in Monte Carlo simulations for heavy-ion experiments. Jet angular and rapidity distributions, have shown that there are *ridges* or long-range correlations, that support the emergence of collective phenomena. Since these ridges are present at large rapidities, and show enhancement for less central collisions and lower values of  $p_T$ , then this complements the expected and observed behavior for the values of  $v_n$  and further supports the hydrodynamical modelling of the QGP, that we presented back in Sections 4 and 5.

## 6.2 Photons as thermometers and viscometers

Photons are a special probe of the QGP that, unlike the hard-probes, can traverse the medium from which they are emitted without further interaction. The photon spectra provides access to the evolution and temperature of the fireball, since they are emitted at all the evolution stages of a heavy-ion collision. Even more, the mechanisms for photo-production are known from many other experiments, so an excess or suppression of photons produced in  $A - A$  collisions, would be a signal of new mechanisms possible due to the conditions that prevail in heavy-ion collisions.

Photons are thermometers of the QGP in the following sense: the effective temperature of the QGP can be extracted from the photon spectrum in the low  $p_T$  region, since it represents the inverse logarithmic slope  $dN/dp_T \sim \text{Exp} \{-p_T/T_{\text{eff}}\}$ . These *thermal photons* provide an estimate of the effective temperature from different experiments, e.g.  $T_{\text{eff}} = 211 \pm 19 \text{ MeV}$  in 0-20% Au–Au collisions at  $\sqrt{s_{NN}} = 200 \text{ GeV}$  at RHIC and  $T_{\text{eff}} = 304 \pm 51 \text{ MeV}$  in 0–40% Pb–Pb collisions at  $\sqrt{s_{NN}} = 2.76 \text{ TeV}$  at LHC. The photon  $p_T$  spectrum even shows photons at  $T_{\text{eff}} < T_c$  with blue-shift to  $T_{\text{eff}} > T_c$  due to a strong radial hydrodynamic flow for  $T_c = 155 - 170$ . Recently, there was even a *direct photon's flow "puzzle"* whereby photons from hard scatterings become dominant at high- $p_T$  and are well described by NLO pQCD plus photons from thermal emission of the QGP and the hadronic phase, but there was a photon excess at low- $p_T$  with large values of elliptic flow, that could not be accounted for by the regular mechanisms [54]. This prompted several possibilities for new sources of primordial photons, such as the photo-production via intense magnetic fields present at early stages after the heavy-ion collision [55]. Photons are viscometers of the QGP in the following sense: using event-by-event simulations of heavy-ion collisions based on viscous relativistic hydrodynamics, we learn that  $\eta/s$  suppresses the flow coefficients  $v_n$  from the soft photon spectra and the effect is more pronounced for more peripheral collisions [56].

## 7 Summary and open questions

In these lectures we presented the basic experimental and theoretical tools to study the collisions—and evolution of the colliding system—in heavy-ion experiments, where QGP is produced, the “simplest form of complex matter described by the fundamental laws of QCD” [2]. The QGP is probed with a range of experimental set-ups and we can characterize it with several relevant observables. We have a standard model of heavy-ion collisions, that includes the QGP formation and decay, but that also incorporates the formation of hadrons from the primordial fire ball. The QGP is modeled as a relativistic fluid with very small viscosity and this model can describe most of the flow-related observables across a variety of colliding systems and conditions. Finally, the study of phase transitions in nuclear matter using QCD-inspired models, fuels the experimental plans for the discovery of the critical end-point in this diagram and the ultimate understanding of in-medium hadron formation. As was nicely summarized in Ref. [2], the field of heavy-ion physics has spread into new theoretical and experimental areas of development that have provided a framework for the exploration of new *extreme* QCD phenomenology, but that still has many open new and exciting questions to tackle:

- What new insights can we obtain about well established QCD phenomenology such as hadronization and confinement, from the analysis of heavy ion collision data?
- Do we have an understanding of all the stages of a heavy ion collision? In particular, how well do we understand the initial stages of the collision process, up to the creation of QGP?
- Can we create QGP with collisions of small systems such as proton-proton and proton-light ion collisions?
- What are the properties of QGP? What have we learned about the dynamics of different probes as they traverse this medium?
- Which aspects of heavy-ion collisions phenomenology involve weakly coupled dynamics or strongly coupled dynamics?
- What are the regions of the nuclear matter phase diagram we can explore using heavy-ion physics?

I hope that with these lectures, you are motivated to do hot, dense and exciting QCD, and join the efforts of the heavy-ion community in this era where new experiments are helping us explore the phase diagram of nuclear matter.

## References

- [1] These lectures are based on talks given at Quark Matter 2018, <https://indico.cern.ch/event/656452>, by K. Zapp, J. Jia and W. Zajc; talks given at Confinement 2018 by H. Satz, *PoS Confinement2018* (2018), 028, doi:10.22323/1.336.0028 and 087, doi:10.22323/1.336.0087; A. K. Chaudhuri, *A short course on relativistic heavy ion collisions* (IOP, Bristol, 2014), doi:10.1088/978-0-750-31060-4, arXiv:1207.7028; K. Reyers, *Lecture notes on quark-gluon plasma physics* (Heidelberg University, 2017); previous heavy-ion and extreme QCD lectures given at CLASHEP by A. Ayala (2017), doi:10.23730/CYRSP-2018-004.125, E.S. Fraga (2015), doi:10.5170/CERN-2016-005.165 and J. Takahashi (2013), doi:10.5170/CERN-2013-003.273.
- [2] W. Busza, K. Rajagopal and W. van der Schee, *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 339, doi:10.1146/annurev-nucl-101917-020852, arXiv:1802.04801 [hep-ph].
- [3] C.Y. Wong, *Introduction to high-energy heavy-ion collisions* (World Scientific, Singapore, 1994), doi:10.1142/1128.
- [4] L. Csernai, *Introduction to relativistic heavy-ion collisions* (John Wiley, Chichester, 1994).
- [5] E.V. Shuryak, *The QCD vacuum, hadrons and superdense matter*, 2nd ed. (World Scientific, Singapore, 2004), doi:10.1142/5367.

- [6] R. Vogt, *Ultrarelativistic heavy-ion collisions* (Elsevier, Amsterdam, 2007), doi:10.1016/B978-0-444-52196-5.X5000-2.
- [7] W. Florkowski, *Phenomenology of ultra-relativistic heavy-ion collisions* (World Scientific, Singapore, 2010), doi:10.1142/7396.
- [8] Eds. S. Sarkar, H. Satz and B. Sinha, *The physics of the quark-gluon plasma* (Lect. Notes Phys. 785, Springer, Berlin, 2010) doi:10.1007/978-3-642-02286-9.
- [9] M. Tanabashi *et al.* [Particle Data Group], *Phys. Rev.* **D98** (2018) 030001, doi:10.1103/PhysRevD.98.030001.
- [10] S. Kluth, *Rept. Prog. Phys.* **69** (2006) 1771, doi:10.1088/0034-4885/69/6/R04, arXiv:hep-ex/0603011; G. Altarelli, *Collider physics within the Standard Model: A primer*, Ed. J. Wells, (Lect. Notes Phys. 937, Springer, Berlin, 2017), doi:10.1007/978-3-319-51920-3.
- [11] A. Deur, S. Brodsky, G. de Teramond, *Prog.Part.Nucl.Phys.* **90** (2016) 1, doi:10.1016/j.pnpnp.2016.04.003, arXiv:1604.08082 [hep-ph].
- [12] M. Boggia *et al.*, *J.Phys.* **G45** (2018) 065004, doi:10.1088/1361-6471/aab812, arXiv:1711.09875 [hep-ph].
- [13] E. Iancu, QCD in heavy ion collisions, Proc. 2011 European School of High-Energy Physics (ESHEP 2011), Cheile Gradistei, Romania, Sep. 7–20, 2011, Eds. C. Grojean, M. Mulders, CERN-2014-003, p. 197, doi:10.5170/CERN-2014-003.197.
- [14] M. Botje **PP-II QCD lectures** (Nikhef, Watergraafsmeer, 2013); F. Halzen, Alan D. Martin, *Quarks and leptons: An introductory course in modern particle physics* (Wiley, New York, NY, 1984).
- [15] H.J. Behrend *et al.* [CELLO Collaboration], *Z. Phys.* **C46** (1990) 537, doi:10.1007/BF01560254; H. Albrecht *et al.* [ARGUS Collaboration], *Z. Phys.* **C58** (1993) 61, doi:10.1007/BF01554080; R. Akers *et al.* [OPAL Collaboration], *Z. Phys.* **C67** (1995) 45, doi:10.1007/BF01564820; K. Ackerstaff *et al.* [OPAL Collaboration], *Z. Phys.* **C75** (1997) 593, doi:10.1007/s002880050505; R. Barate *et al.* [ALEPH Collaboration], *Eur. Phys. J.* **C4** (1998) 409, doi:10.1007/s100520050217.
- [16] K. Ackerstaff *et al.* [OPAL Collaboration], *Eur. Phys. J.* **C7** (1999) 571, doi:10.1007/s100529901061, arXiv:hep-ex/9808019.
- [17] R.L. Jaffe, D. Pirjol, A. Scardicchio, *Phys.Rep.* **435** (2006) 157, doi:10.1016/j.physrep.2006.09.004, arXiv:hep-ph/0602010.
- [18] J.C. Collins, M.J. Perry, *Phys. Rev. Lett.* **34** (1975) 1353, doi:10.1103/PhysRevLett.34.1353; N. Cabibbo, G. Parisi, *Phys. Lett.* **B59** (1975) 67, doi:10.1016/0370-2693(75)90158-6.
- [19] M. Gell-Mann, *Phys. Lett.* **8** (1964) 214, doi:10.1016/S0031-9163(64)92001-3; G. Zweig, **An  $SU_3$  model for strong interaction symmetry and its breaking**, version 1, CERN-TH-401 (1964).
- [20] H. Fritzsch, M. Gell-Mann, H. Leutwyler, *Phys. Lett.* **B47** (1973) 365, doi:10.1016/0370-2693(73)90625-4; D.J. Gross, F. Wilczek, *Phys. Rev. Lett.* **30** (1973) 1343, doi:10.1103/PhysRevLett.30.1343; H.D. Politzer, *Phys. Rev. Lett.* **30** (1973) 1346, doi:10.1103/PhysRevLett.30.1346.
- [21] R. Pasechnik, M. Šumbera, *Universe* **3** (2017) 7, doi:10.3390/universe3010007.
- [22] A. Bzdak *et al.*, *Phys. Rep.* **853** (2020) 1, doi:10.1016/j.physrep.2020.01.005, arXiv:1906.00936 [nucl-th].
- [23] K. Yagi, T. Hatsuda, Y. Miake, *Quark-gluon plasma: From big bang to little bang* (Cambridge Univ. Press, 2005); Ed. X. Wang, *Quark-gluon plasma 5* (World Scientific, Singapore, 2016), doi:10.1142/9533.

- [24] H. T. Ding, F. Karsch, S. Mukherjee, *Int. J. Mod. Phys.* **E24** (2015) 1530007, doi:10.1142/S0218301315300076, arXiv:1504.05274 [hep-lat].
- [25] NICA physics, Joint Institute for Nuclear Research, Dubna, <https://nica.jinr.ru/physics.php>, last accessed September 11, 2019.
- [26] Adapted from: A. Harmon, *Modeling heavy-ion collisions with Open Science Grid*, <https://sciencenode.org/feature/modeling-heavy-ion-collisions-open-science-grid.php>, last accessed September 11, 2019.
- [27] A. Toia, *Participants and spectators at the heavy-ion fireball*, *CERN Cour.* **53** no. 4 (2013) 31.
- [28] A. Mazeliauskas, *Fluctuations in ultra-relativistic heavy ion collisions*, Ph.D. thesis, Stony Brook University, 2017, arXiv:1702.01202 [nucl-th].
- [29] MCGlauber: An open-source iPython-based Monte Carlo Glauber model of nuclear collisions, <https://github.com/MCGlauber>, last accessed September 11, 2019.
- [30] H. Satz, *J. Phys.* **G32** (2006) R25, doi:10.1088/0954-3899/32/3/R01, arXiv:hep-ph/0512217.
- [31] J.L. Nagle, W.A. Zajc, *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 211, doi:10.1146/annurev-nucl-101916-123209, arXiv:1801.03477 [nucl-ex].
- [32] J. Letessier, J. Rafelski, *Hadrons and quark-gluon plasma* (Cambridge Univ. Press, Cambridge, 2002), doi:10.1017/CBO9780511534997.
- [33] A. Bazavov *et al.* [HotQCD Collaboration], *Phys. Rev.* **D90** (2014) 094503, doi:10.1103/PhysRevD.90.094503, arXiv:1407.6387 [hep-lat].
- [34] Adapted from W.A. Zajc, *The modern view of quark-gluon plasma*, presented at Int. Conf. on Ultra-relativistic Nucleus-Nucleus Collisions (Quark Matter 2018), Venice, Italy, 13–19 May 2018.
- [35] P.F. Kolb, U.W. Heinz, Hydrodynamic description of ultrarelativistic heavy ion collisions, in Eds. R.C.Hwa, X.-N. Wang, *Quark gluon plasma 4* (World Scientific, Singapore, 2010) p. 223, doi:10.1142/7588, nucl-th/0305084.
- [36] P. Danielewicz, M. Gyulassy, *Phys. Rev.* **D31** (1985) 53, doi:10.1103/PhysRevD.31.53.
- [37] P.K. Kovtun, D.T. Son, A.O. Starinets, *Phys. Rev. Lett.* **94** (2005) 111601, doi:10.1103/PhysRevLett.94.111601, arXiv:hep-th/0405231.
- [38] R. Tribble, *Implementing the 2007 Long Range Plan*, report to the Nuclear Science Advisory Committee, 2013, archived 22 October 2020.
- [39] S. Jeon, U. Heinz, *Int. J. Mod. Phys.* **E24** (2015) 1530010, doi:10.1142/S0218301315300106, arXiv:1503.03931 [hep-ph].
- [40] A. Ayala, I. Dominguez, M.E. Tejeda-Yeomans, *Phys. Rev.* **C88** (2013) 025203, doi:10.1103/PhysRevC.88.025203, arXiv:1212.1127 [hep-ph].
- [41] J. Ghiglieri, G.D. Moore, D. Teaney, *JHEP* **03** (2018) 179, doi:10.1007/JHEP03(2018)179; P. B. Arnold, G. D. Moore, L. G. Yaffe, *JHEP* **01** (2003) 030, doi:10.1088/1126-6708/2003/01/030, arXiv:hep-ph/0209353.
- [42] Adapted from *Interview with Krishna Rajagopal*, *ALICE Matters*, 15 April 2013.
- [43] J.-B. Rose *et al.*, *Phys. Rev.* **C97** (2018) 055204, doi:10.1103/PhysRevC.97.055204, arXiv:1709.03826 [nucl-th]; A. Dubla *et al.*, *Nucl. Phys.* **A979** (2018) 251, doi:10.1016/j.nuclphysa.2018.09.046, arXiv:1805.02985 [nucl-th].
- [44] P. Blanchard, S. Fortunato, H. Satz, *Eur.Phys.J.* **C34** (2004) 361, doi:10.1140/epjc/s2004-01673-0, arXiv:hep-ph/0401103.
- [45] R. Hagedorn, *Nuovo Cimento, Suppl.* **3** (1965) 147, CERN-TH-520; Thermodynamics of strong interactions, CERN 71-12 (1971), doi:10.5170/CERN-1971-012; *NATO Sci. Ser.* **B346** (1995) 13, doi:10.1007/978-3-319-17545-4\_17.

- [46] J. Noronha-Hostler, Implications of missing resonances in heavy ions collisions, [arXiv:1612.07765 \[nucl-th\]](#), Proc. Workshop on Excited Hyperons in QCD Thermodynamics at Freeze-Out (YSTAR2016), Thomas Jefferson National Accelerator Facility, 16–17 November 2016, Eds. M. Amaryan *et al.*, [arXiv:1701.07346 \[hep-ph\]](#).
- [47] S.K. Tiwari, C.P. Singh, *Adv. High Energy Phys.* **2013** (2013) 805413, [doi:10.1155/2013/805413](#).
- [48] M.A. Stephanov, *Int. J. Mod. Phys. A* **20** (2005) 4387, [doi:10.1142/S0217751X05027965](#), [arXiv:hep-ph/0402115](#);  
A. Ayala *et al.*, *Eur. Phys. J. A* **56** (2020) 71, [doi:10.1140/epja/s10050-020-00086-z](#), [arXiv:1904.11905 \[hep-ph\]](#).
- [49] M.J. Tannenbaum, Highlights from BNL-RHIC, Proc. Int. School of Subnuclear Physics, Erice, Italy, 23 June–2 July 2012, Ed. A. Zichichi, (World Scientific, Singapore, 2014) p. 347, [doi:10.1142/9789814603904\\_0022](#), [arXiv:1302.1833 \[nucl-ex\]](#).
- [50] M. Aaboud *et al.* [ATLAS], *Phys. Lett.* **B790** (2019) 108, [doi:10.1016/j.physletb.2018.10.076](#).
- [51] X. Luo [STAR Collaboration], *PoS CPOD2014* (2015) 019, [doi:10.22323/1.217.0019](#).
- [52] A. Bazavov *et al.* [HotQCD Collaboration], *Phys. Rev.* **D96** (2017) 074510, [doi:10.1103/PhysRevD.96.074510](#), [arXiv:1708.04897 \[hep-lat\]](#).
- [53] M. Gyulassy *et al.*, Jet quenching and radiative energy loss in dense nuclear matter, in Eds. R.C. Hwa, X.-N. Wang, Quark–gluon plasma 3 (World Scientific, Singapore, 2004) p. 123, [doi:10.1142/9789812795533\\_0003](#), [arXiv:nucl-th/0302077](#);  
G. Y. Qin, X. N. Wang, *et al.*, *Int. J. Mod. Phys. E* **24** (2015) 1530014, [doi:10.1142/S0218301315300143](#), [arXiv:1511.00790 \[hep-ph\]](#);  
I. Vitev *et al.*, *JHEP* **11** (2008) 093, [doi:10.1088/1126-6708/2008/11/093](#), [arXiv:0810.2807 \[hep-ph\]](#);  
I. Vitev *et al.*, *Phys. Rev. Lett.* **104** (2010) 132001, [doi:10.1103/PhysRevLett.104.132001](#), [arXiv:0910.1090 \[hep-ph\]](#).
- [54] A. Adare *et al.* [PHENIX Collaboration], *Phys. Rev. Lett.* **109** (2012) 122302, [doi:10.1103/PhysRevLett.109.122302](#), [arXiv:1105.4126 \[nucl-ex\]](#);  
D. Lohner [ALICE Collaboration], *J. Phys. Conf. Ser.* **446** (2013) 012028, [doi:10.1088/1742-6596/446/1/012028](#);  
J. Adam *et al.* [ALICE Collaboration], *Phys. Lett.* **B754** (2016) 235, [doi:10.1016/j.physletb.2016.01.020](#);  
C. Shen, *Nucl. Phys. A* **956** (2016) 184, [doi:10.1016/j.nuclphysa.2016.02.033](#), [arXiv:1601.02563 \[nucl-th\]](#).
- [55] A. Ayala *et al.*, *Phys. Rev.* **D96** (2017) 014023, [doi:10.1103/PhysRevD.96.014023](#), Erratum-*ibid.* 119901, [doi:10.1103/PhysRevD.96.119901](#), [arXiv:1704.02433 \[hep-ph\]](#);  
A. Ayala *et al.*, *Eur. Phys. J. A* **56** (2020) 53, [doi:10.1140/epja/s10050-020-00060-9](#), [arXiv:1904.02938 \[hep-ph\]](#).
- [56] C. Shen *et al.*, *Nucl. Phys. A* **932** (2014) 184, [doi:10.1016/j.nuclphysa.2014.07.042](#), [arXiv:1403.7558 \[nucl-th\]](#).

## LHC highlights and prospects

*C. E. Gerber*

University of Illinois at Chicago, Chicago, IL, USA

### Abstract

These lectures were presented at the 2019 CERN–Latin-American School of High-Energy Physics. They were centered on the experimental methods used in hadron colliders to advance our understanding in the field of high-energy particle physics. From accelerators, to particle detector technologies, object identification and data analyses techniques, the lectures did not attempt to provide a comprehensive, in-depth technical background, but rather focused on an overview of experimental techniques that enabled our advances in supporting and challenging the predictions of the Standard Model. This document includes a selection of the material presented in the lectures, focusing on how advances in detector technologies and object identification enabled the development of increasingly sophisticated data analysis techniques. This write-up also includes an outlook to the future LHC program and beyond.

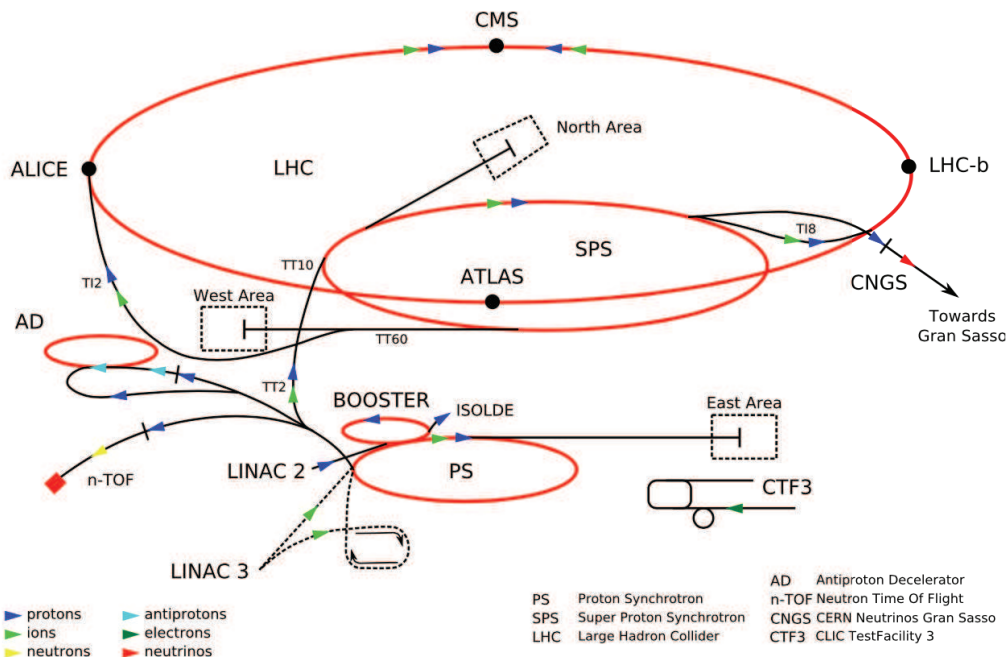
### Keywords

LHC physics; Hadron colliders; Experimental techniques.

## 1 Introduction

The Large Hadron Collider [1], a 26.7 km diameter superconducting proton-proton collider, is the largest particle accelerator ever built. It is operated by the European Organization for Nuclear Research (CERN) in Geneva Switzerland, and is the last stage of a multi-accelerator complex that results in proton-proton collisions at a center of mass energy in the 7 – 14 TeV range. A schematic view of the CERN accelerator complex can be seen in Fig. 1. The first stage is the LINAC 2 linear accelerator, where a proton source extracts 90 keV protons from a Hydrogen bottle and accelerates the beam to 50 MeV over a distance of 33 m, providing a pulse every 1.2 s. Next, the PS Booster, which is the first synchrotron in the accelerator chain, and has 157 m in circumference, increases the proton energy to 1.4 GeV in 1.2 s. Protons are then injected into the PS, the oldest operating synchrotron at CERN with a circumference of 628 m (4 times the size of the PS Booster), and their energy is increased to 26 GeV. Next, protons are injected into the SPS, the first superconducting synchrotron in the chain, with 6.9 km circumference and 30 m underground, which was originally a proton-antiproton collider that led to the discovery of the  $W$  and  $Z$  bosons. The SPS increases the proton energy to 450 GeV and provides beam to the LHC and to fixed target areas. The LHC consists of 1232 main dipoles of 15 m each that deviate the beams around the 27 km circumference, 858 main quadrupoles that keep the beam focused and 6000 corrector magnets to preserve the beam quality. The main magnets use superconducting cables (Cu-cladded Nb-Ti), with 12 000 A providing a nominal field of 8.33 Tesla. The LHC started operations in 2010, delivering  $36 \text{ pb}^{-1}$  of data at  $\sqrt{s} = 7 \text{ TeV}$ , followed by  $5 \text{ fb}^{-1}$  of data in 2011. The center of mass energy was increased to  $\sqrt{s} = 8 \text{ TeV}$  in 2012, when  $20 \text{ pb}^{-1}$  of data were delivered. Finally, in what is known as Run 2,  $150 \text{ pb}^{-1}$  of data at  $\sqrt{s} = 13 \text{ TeV}$  were delivered between 2015 and 2018.

These lectures will focus only on the two multi-purpose experiments, ATLAS [2] and CMS [3], shown schematically in Figs. 2 and 3. Both detectors rely on layered instrumentation technologies to detect the passage of particles and measure their position and their energies with the best possible precision. The innermost layer is typically a silicon-based semiconductor tracker, that is both radiation hard and has low mass, to minimize the multiple scattering from detected tracks. Electrically charged particles leave hits in the tracker layers, which allows trajectories to be reconstructed from consecutive

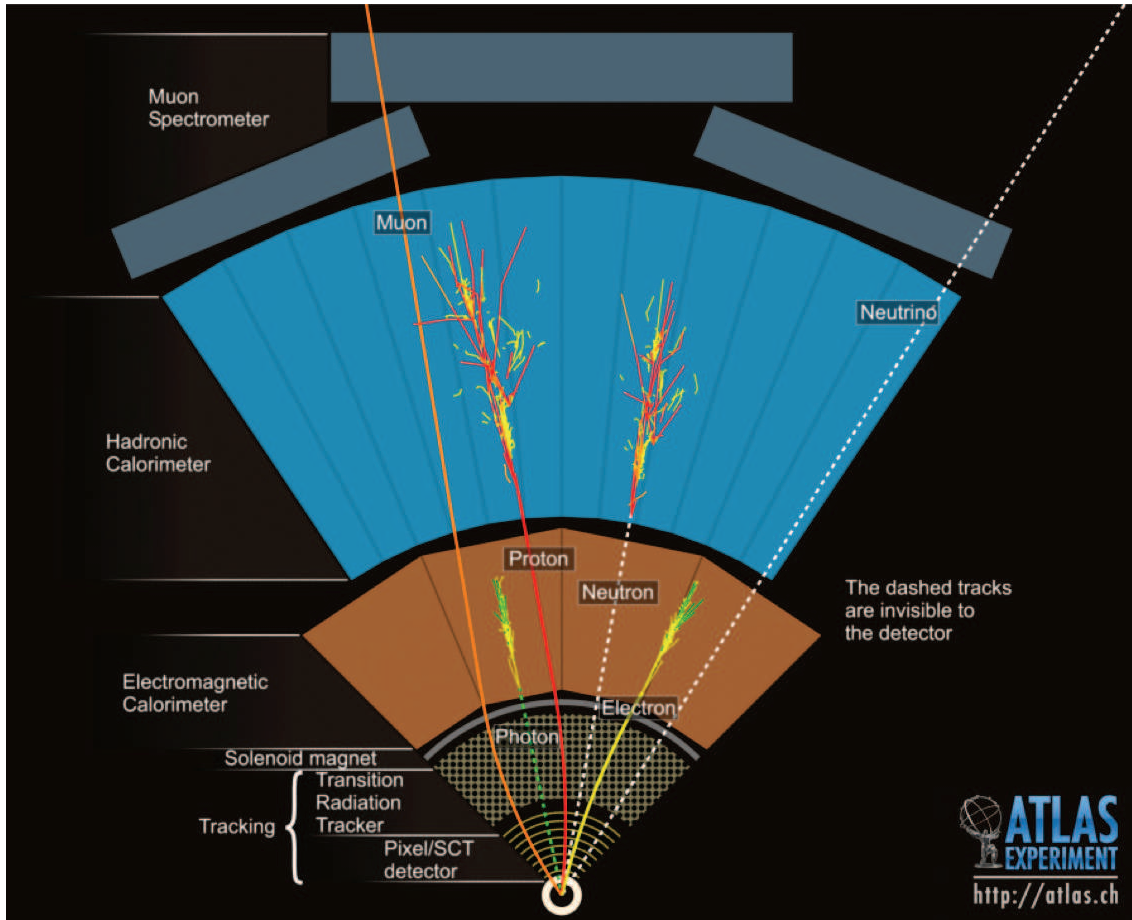


**Fig. 1:** Schematic view of the CERN accelerator complex, showing the different accelerator stages that result in proton-proton collisions at the center of the four detector areas.

measurements as particles traverse the detector volume. CMS is the first hadron collider experiment to use an all-silicon tracker. ATLAS innermost tracker is silicon-based, but the outer part uses a gas and wire transition radiation tracker. Outside of the trackers are the calorimeters, that destructively measure the energy of charged and neutral particles. Calorimeters complement the information obtained from the magnetic spectrometers, as they are able to measure the energy of neutral particles and have an energy resolution which improves with the particle energy, while the spectrometer resolution degrades with particle energy. Finally, the outermost part of the detectors are the muon spectrometers, that are typically gas-based detectors that cover huge volumes and identify the only particles that escape the calorimeters, which are the muons. Both collaborations have superconducting magnets. CMS has a single solenoidal 4 Tesla magnet, cooled at  $-270^{\circ}\text{C}$ , outside the combined tracker and calorimeter volume. The muon chambers are interleaved with a 12-sided, 3-layers iron structure that surrounds the magnet coils and contains the B field. ATLAS has a 2 Tesla central solenoid between the tracker and the calorimeters and two toroids (a central and a forward toroid) as part of the ATLAS muon system. Finally, the trigger systems decide, in real-time, which subset of data is to be readout by the detector and stored for offline analysis, and the data acquisition (DAQ) system collects the data from the different parts of the detector, converts it to a suitable format, and saves it to permanent storage. Both detectors use a multi-level trigger system, with a first trigger level with very short latency, high signal efficiency but modest background rejection, which is typically firmware-based, single-detector or limited combination. This is followed by subsequent trigger level(s) that achieve high background rejection, with typically larger latency, and which are computer-based and use the information from multiple detectors, with the highest level being a speed-up version of the offline reconstruction.

The latest development in offline event reconstruction uses the particle flow algorithm [4] to create a comprehensive list of all final-state particles in the collision. This approach makes use of the best combination of all subdetectors to reconstruct and identify all particles, and also provides robust handles





**Fig. 2:** Cross-sectional view of the ATLAS detector.

against energy deposits originating from the underlying event and from multiple proton-proton interactions in the same bunch crossing. It also opens the door to a new field of studies that uses jet substructure techniques to differentiate between quark and gluon jets based on the differences in their radiation patterns and lifetimes [5]. The technique can also be extended to distinguish jets from hadronic decays of high transverse momentum heavy particles, in particular  $W$  and  $Z$  bosons, top quarks and Higgs bosons. As the largest branching ratios for these particles is into their hadronic decays, these tools open up a large amount of acceptance that was previously inaccessible. Furthermore, the ability to reconstruct subjects from merged decays can be used to separate these heavy objects from the much more copious quantum chromodynamics (QCD) multi-jet production. This is particularly important when studying hadron collisions that give rise to a wide variety of processes with production cross sections that span 12–13 orders of magnitude. While they enable a rich physics program, the interesting processes are overwhelmed by mundane processes that occur at much higher rates. Background discrimination and residual background modeling are crucial ingredients in any physics analysis using hadron collider data. Figure 4 shows the cross sections and events per second produced at hadron colliders as a function of the center of mass energy.

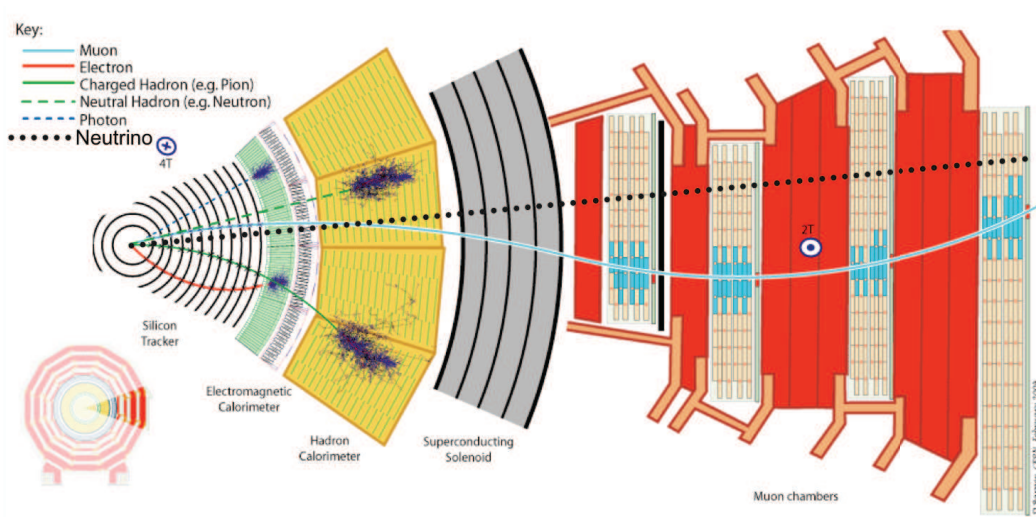


Fig. 3: Cross-sectional view of the CMS detector.

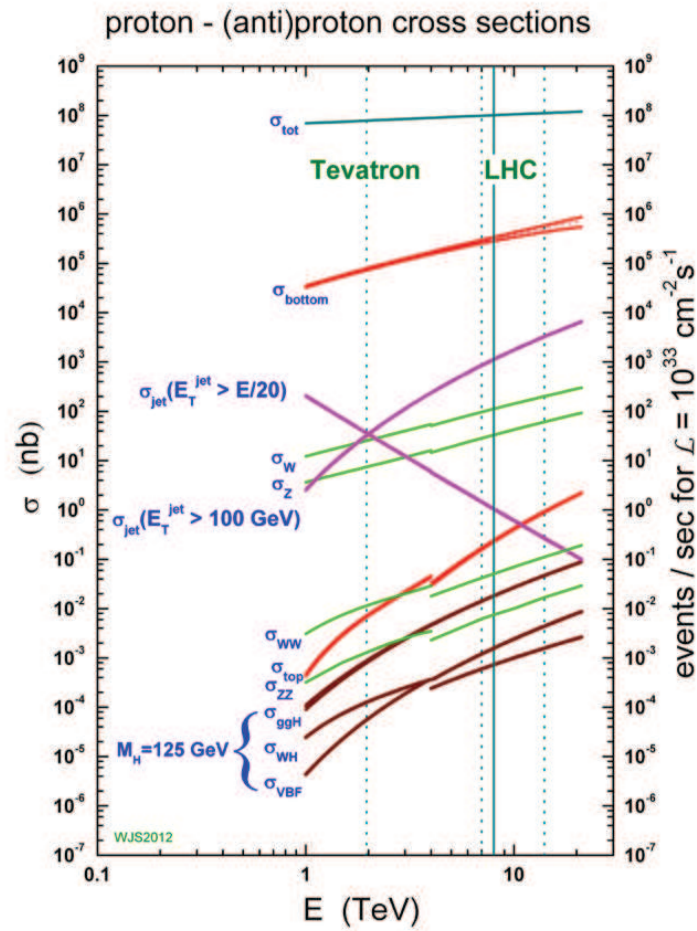


Fig. 4: Production cross sections as a function of center of mass energy in hadron colliders.

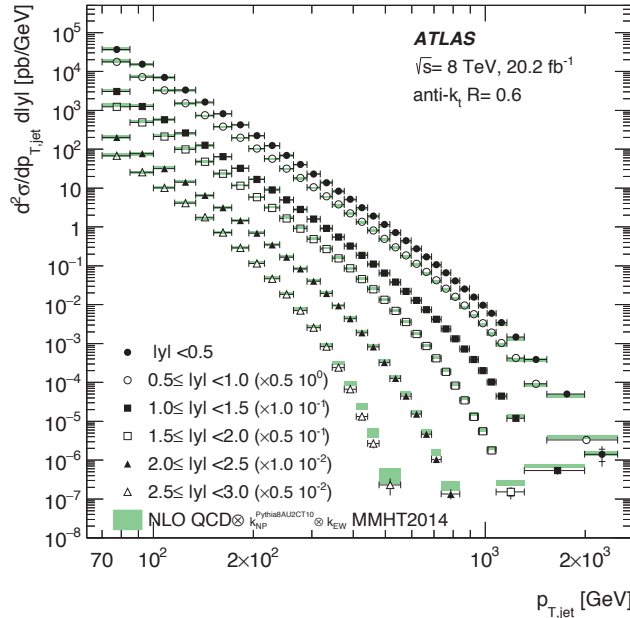
## 2 Jet production

At LHC energies, the dominant process in pp collisions is jet production. Within the framework of QCD, inelastic scattering between two protons can be described as an elastic collision between a single constituent of each proton. These constituents are called partons. After the collision, the outgoing partons manifest themselves as localized streams of particles referred to as “jets”. Theoretical predictions for jet production are given by the folding of the parton scattering cross sections with experimentally determined parton density functions (pdf’s). These predictions are known at next-to-leading order (NLO) in perturbative QCD scattering calculations [6,7] and accurately measured pdf’s [8]. Some of the questions that can be addressed with studies of jet production are testing of NLO QCD, extraction of pdf’s, measuring the value of the strong coupling constant  $\alpha_s$ , and testing of beyond the Standard Model (BSM) theories.

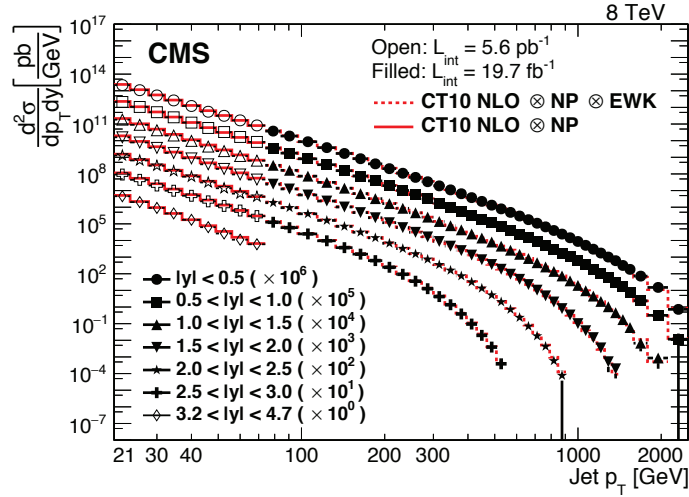
The simplest test that can be performed is the measurement of the production rate of jets as a function of the jet transverse momentum ( $p_T$ ) in different rapidity bins  $y$ , a study in which each jet in the event corresponds to an entry in the histogram. The ATLAS and CMS collaborations measure the double differential inclusive jet cross section in pp collisions at  $\sqrt{s} = 8$  TeV, which can be expressed as:

$$d^2\sigma/(dp_T dy) = (N_{Jet})/(\varepsilon \Delta p_T \Delta |y| \int L dt) ,$$

where  $N_{Jet}$  is the total number of jets observed in a certain jet transverse energy  $E_T$  bin,  $\varepsilon$  is the selection efficiency,  $\Delta p_T$  is the bin width,  $\Delta y$  is the rapidity range considered, and  $\int L dt$  is the integrated luminosity associated with the data set. The results are shown in Fig. 5 for ATLAS [9] and Fig. 6 for CMS [10]. The cross sections for different rapidity bins are multiplied by factors indicated in the legends for better visibility. The dominant systematic uncertainties are the jet energy scale and resolution, and range from 2 – 30%, being largest for low  $p_T$  jets and jets in high rapidity regions. Overall, reasonable good agreement is observed between data and the NLO predictions. The data should provide improved constraints on parton distribution functions.

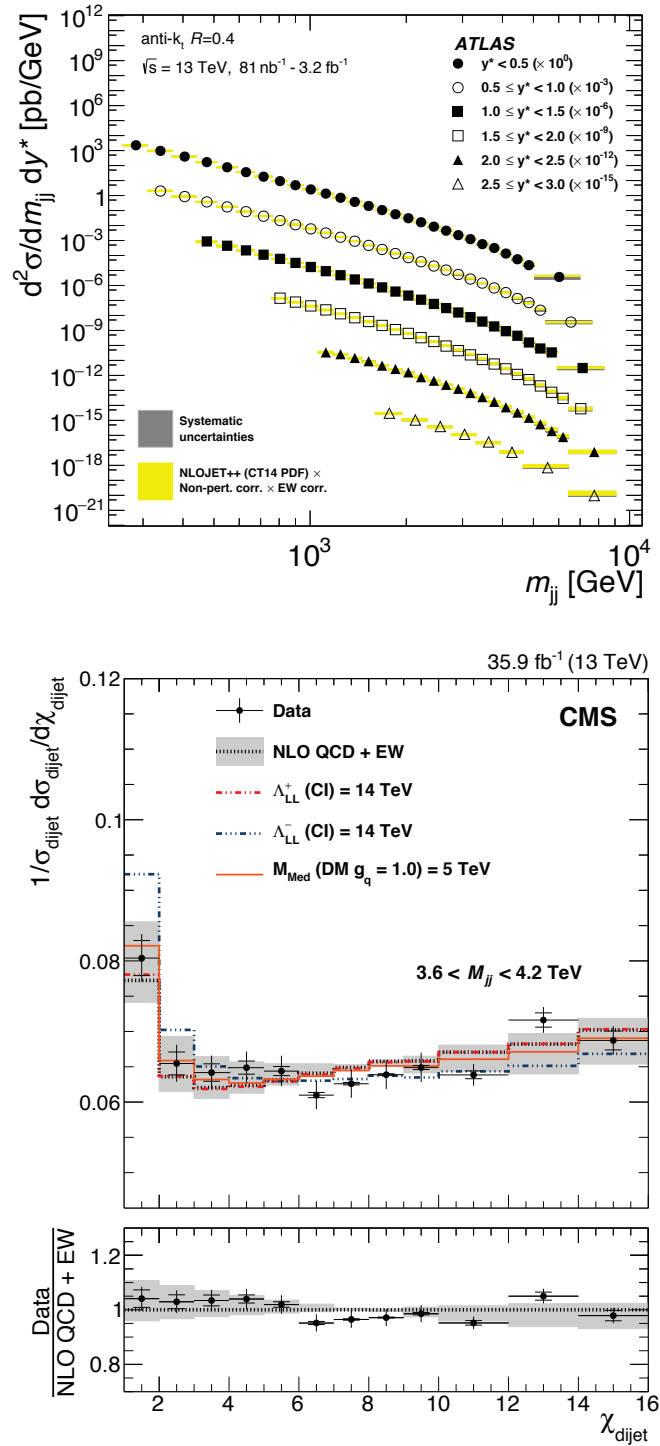


**Fig. 5:** Double-differential inclusive jet cross sections as function of jet  $p_T$  from ATLAS. The cross sections for different rapidity bins are multiplied by factors indicated in the legends for better visibility.



**Fig. 6:** Double-differential inclusive jet cross sections as function of jet  $p_T$  from CMS. The cross sections for different rapidity bins are multiplied by factors indicated in the legends for better visibility.

Both collaborations also study the characteristics of the system given by the two leading jets in an event. Discrepancies with QCD predictions could indicate beyond the Standard Model physics like quark compositeness, excited quarks, quark contact interactions, extra spatial dimensions, quantum black holes, or dark matter. The results are shown in Fig. 7 (top) for ATLAS [11] and Fig. 7 (bottom) for CMS [12]. ATLAS presents its result as the dijet cross section as a function of the dijet invariant mass  $m_{jj}$ . CMS chooses to concentrate on the angular distribution of dijets relative to the beam direction by studying  $\chi_{\text{dijet}} = e^{|y_1 - y_2|}$ , where  $y_1$  and  $y_2$  are the rapidities of the two leading jets. The choice of  $\chi_{\text{dijet}}$  as the sensitive variable is motivated by the fact that BSM processes, that are expected to have isotropic angular distributions, would result in an excess of events over QCD predictions at low values of  $\chi_{\text{dijet}}$ .

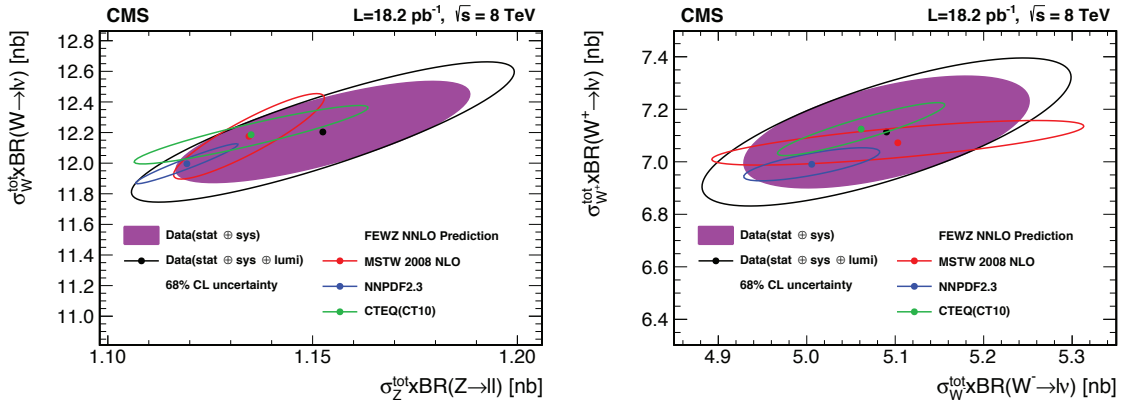


**Fig. 7:** Dijet cross-sections as a function of the dijet invariant mass  $m_{jj}$ . The ATLAS data are compared to NLO pQCD predictions (top). Example of a normalized  $\chi_{dijet}$  distribution where unfolded data from CMS is compared to NLO predictions and various BSM scenarios (bottom).

### 3 Vector boson production

$W$  and  $Z$  bosons, the carriers of the weak force, are directly produced in high-energy pp collisions at the CERN LHC. In addition to probing electroweak physics, the study of the production of  $W$  and  $Z$  bosons provides an avenue to explore QCD.  $W$  and  $Z$  bosons, when produced in association with jets, and in particular with  $b$ -jets, constitute an irreducible background to many processes that decay to  $W$  or  $Z$ , like top, Higgs and BSM production. The leptonic  $Z$  decays provide clean samples with adequate statistics for detector performance measurements and the extraction of data-to-MC correction factors for trigger and lepton identification.

The most precise measurement of the leptonic  $W$  production cross section at  $\sqrt{s} = 8$  TeV was performed by CMS using special low pileup data collected in 2012 [13]. The data sample corresponds to an integrated luminosity of  $18.2 \pm 0.5 \text{ pb}^{-1}$ , and has an average of 4 interactions per bunch crossing, to be compared with an average of 21 interactions for the regular beam conditions during 2012. Events are selected in their decays to high  $p_T$ , isolated electrons or muons. The leading systematic uncertainty arises from the measurement of the lepton reconstruction and identification. The uncertainty on the integrated luminosity cancels when calculating the ratio of cross sections, which are measured with a precision of 2%. Figure 8 shows the measured and predicted total production cross sections times branching ratio to leptons for  $W$  vs.  $Z$  (left) and  $W^+$  vs.  $W^-$  (right). The measurements in the electron and muon channel are in agreement with NNLO theoretical predictions and among the channels, in accordance with the expectation from lepton universality.



**Fig. 8:** Total production cross sections times branching ratio to leptons for  $W$  vs.  $Z$  (left) and  $W^+$  vs.  $W^-$  (right). The ellipses illustrate the 68% CL coverage for total uncertainties (open) and excluding the luminosity uncertainty (filled). The uncertainties in the theoretical predictions correspond to the PDF uncertainty components only and are evaluated for different PDF sets, as indicated in the figure.

### 4 Measurement of the $W$ mass

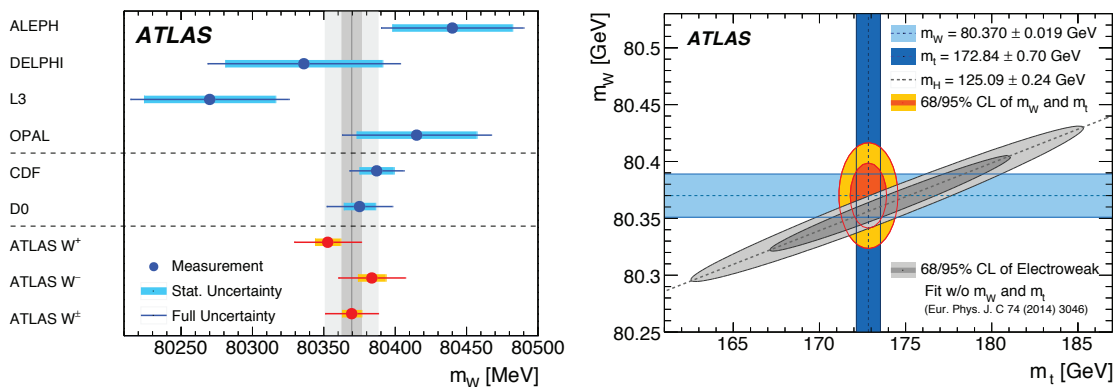
In the Standard Model of the electroweak interactions, the mass of the  $W$  boson can be expressed as

$$m_W^2 \left( 1 - \frac{m_W^2}{m_Z^2} \right) = \frac{\pi \alpha (m_Z^2)}{\sqrt{2} G_F} (1 + \Delta r_W). \quad (1)$$

A measurement of  $m_W$ , together with  $m_Z$ , the Fermi constant ( $G_F$ ), and the electromagnetic coupling constant ( $\alpha$ ), determines the electroweak radiative corrections  $\Delta r_W$  experimentally. The dominant contributions to  $\Delta r_W$  arise from loop diagrams that involve the top quark and the Higgs boson. The correction from the  $t\bar{t}$  loop is substantial because of the large mass difference between the two quarks. It is proportional to  $m_t^2$  for large values of the top quark mass  $m_t$ . If additional particles which couple to the

$W$  boson exist, they would give rise to additional contributions to  $\Delta r_W$ . Therefore, a measurement of  $m_W$  is one of the most stringent experimental tests of SM predictions. Deviations from the predictions may indicate the existence of new physics. Within the SM, measurements of  $m_W$  and the mass of the top quark and the Higgs boson are a crucial test of the overall consistency of the SM, and discrepancies could indicate new physics. The experimental challenge is thus to measure the  $W$  boson mass to sufficient precision, about 0.1%, to be sensitive to BSM effects.

The ATLAS collaboration has used data collected in 2011 in pp collisions at  $\sqrt{s} = 7$  TeV to measure the  $W$  boson mass from template fits to the reconstructed distributions for the charged lepton  $p_T$  and the  $W$  boson transverse mass [14]. The analysis uses  $4.6 \text{ pb}^{-1}$  of data, which results in  $7.8 \times 10^6$   $W \rightarrow \mu\nu$  and  $5.9 \times 10^6$   $W \rightarrow e\nu$  candidates. In addition,  $1.23 \times 10^6$   $Z \rightarrow \mu\mu$  and  $0.58 \times 10^6$   $Z \rightarrow ee$  candidates are used to pin down the lepton energy and  $W$  recoil calibration. The measured mass of  $m_W = 80370 \pm 7(\text{stat.}) \pm 11(\text{exp.syst.}) \pm 14(\text{mod.syst.}) \text{ MeV} = 80370 \pm 19 \text{ MeV}$ , is consistent with previous results and SM expectations. The measurement is limited by the understanding of the detector modeling, the PDFs and theoretical uncertainties. Figure 9 shows the measured values for the  $W$  boson mass and results from global electroweak fits.



**Fig. 9:** Comparison of the measured values for the  $W$  boson mass by various collaborations (left). ATLAS measured values for the mass of the  $W$  boson and the top quark are compared to the results from global electroweak fits based on the measured LHC Higgs Boson mass of  $125.09 \pm 0.24 \text{ GeV}$  (right).

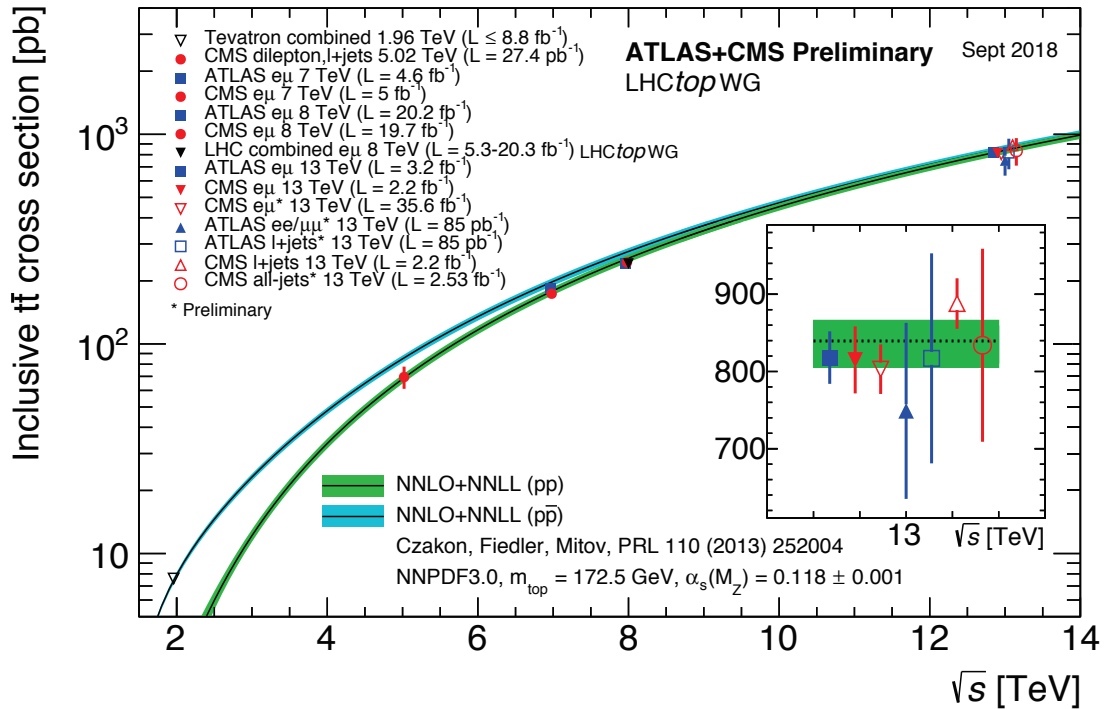
## 5 Top quark production and decay

The top quark is the heaviest known elementary particle and completes the quark sector of the three-generation structure of the Standard Model (SM). It differs from the other quarks not only by its much larger mass, but also by its lifetime which is too short to build hadronic bound states.

The SM predicts that top quarks are created via two independent production mechanisms at hadron colliders. The primary mode, in which a  $t\bar{t}$  pair is produced from a  $g t\bar{t}$  vertex via the strong interaction, was used by the D0 and CDF collaborations to establish the existence of the top quark in 1995 [15, 16]. The second production mode of top quarks at hadron colliders is the electroweak production of a single top quark from a  $Wtb$  vertex. At Tevatron energies, the predicted cross section for single top quark production is about half that of  $t\bar{t}$  pairs but the signal-to-background ratio is much worse; observation of single top quark production has therefore been impeded by its low rate and difficult background environment compared to the top pair production, and only detected by the same collaborations in 2009, fourteen years after the strong production [17, 18].

Within the SM, the top quark decays almost exclusively into a  $W$  boson and a  $b$  quark, resulting in two  $W$  bosons and two  $b$  jets in each  $t\bar{t}$  pair event. The  $W$  boson itself decays into one lepton and

its associated neutrino, or hadronically. The  $t\bar{t}$  pair decay channels have been classified as follows: the dilepton channels where both  $W$  bosons decay leptonically into an electron or a muon, the lepton jets channels where one of the  $W$  bosons decays leptonically and the other hadronically, and the all-jets channel where both  $W$  bosons decay hadronically.  $t\bar{t}$  pair production cross sections have been measured in  $p\bar{p}$  collisions at the Tevatron and in  $pp$  collisions at the LHC in all decay channels except in the dilepton channel where both Tau leptons decay hadronically. Figure 10 summarizes the measurements of the  $t\bar{t}$  production cross-section as a function of the center of mass energy compared to theoretical predictions at NNLO QCD with NNLL resummation [19]. Excellent agreement between measurements and theoretical predictions is observed in all channels. Precise measurements of the  $t\bar{t}$  production cross section represent a test of pQCD at high  $Q^2$ , can be used to constrain PDFs, to determine the top quark mass  $m_t$ , and measure the strong coupling constant  $\alpha_s$ . Precise comparisons of the measured cross sections in different channels and different methods with theoretical expectations are sensitive to new physics. Good understanding of the composition of the samples is crucial to enable the measurement of top quark properties and as input to searches for which the top quarks are the dominant backgrounds. Well understood samples of top quarks can also be used to constrain the energy scale of jets and to measure efficiencies to tag b and top jets.

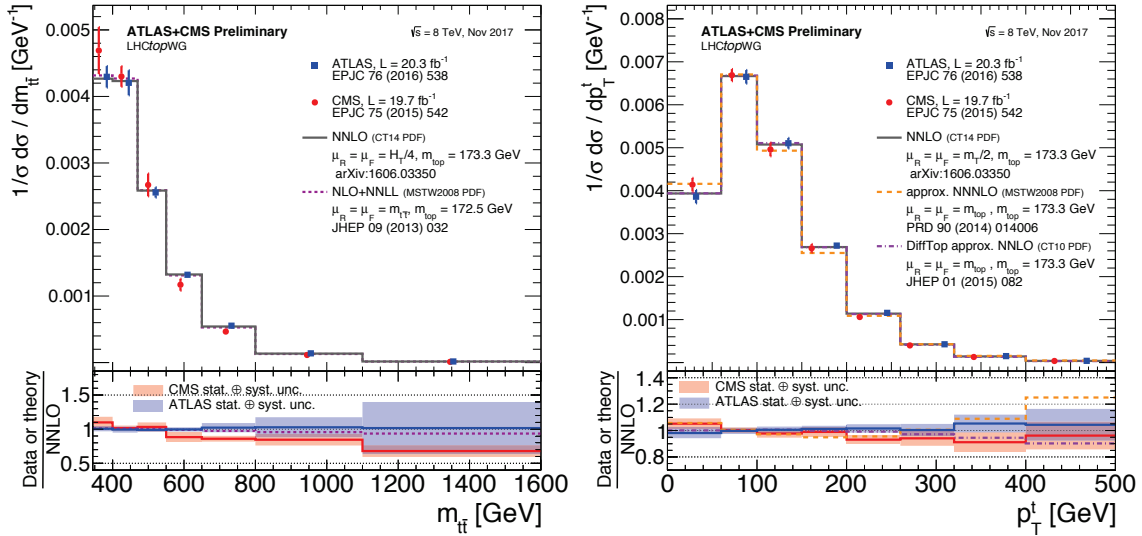


**Fig. 10:** Comparison of the measured values for the  $t\bar{t}$  production cross section as a function of the center of mass energy for  $p\bar{p}$  and  $pp$  collisions with NNLO QCD with NNLL resummation theoretical predictions.

The large number of  $t\bar{t}$  events produced at LHC energies allows for full phase-space normalized differential cross section measurements as those in Fig. 11, in which the ATLAS and CMS measurements are compared to theoretical predictions [19]. Overall good agreement is observed between data and predictions and the measurements can be used to improve the theoretical models.

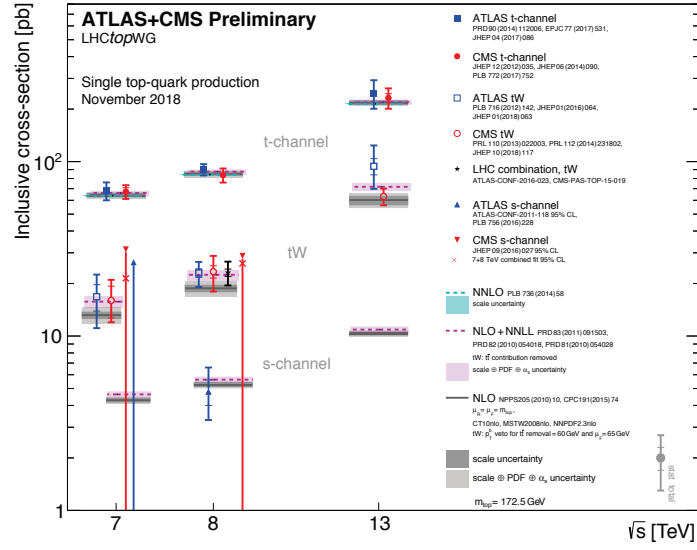
Single top production at hadron colliders provides an opportunity to study the charged-current weak-interaction of the top quark. The SM predicts that the top quark decays almost exclusively to a  $W$  boson and a bottom quark with  $B(t \rightarrow Wb) \approx 1$ . The rate for the process leads to a firm prediction



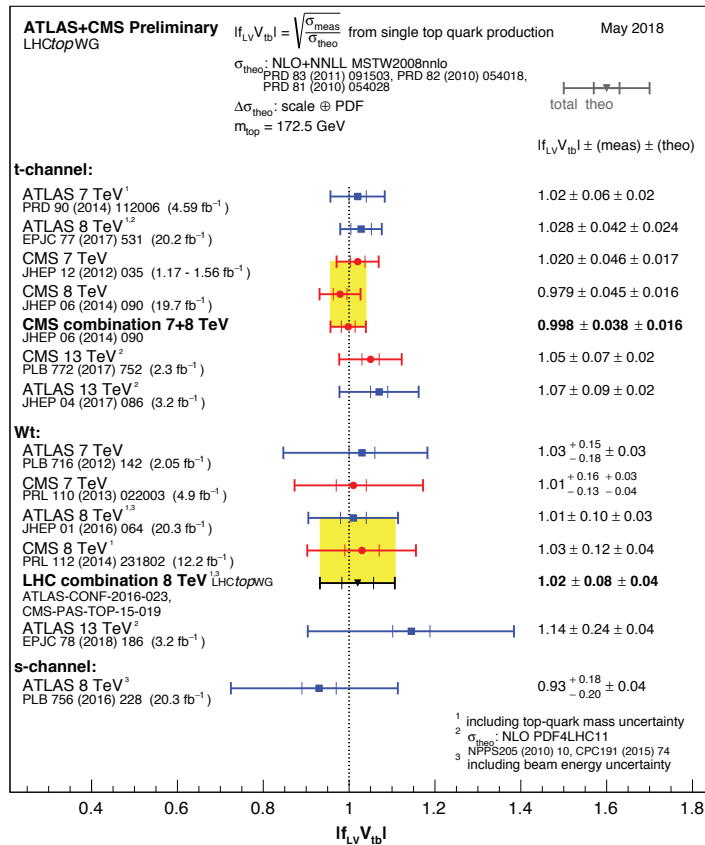


**Fig. 11:** Full phase-space normalized differential  $t\bar{t}$  cross-section as a function of the top quark pair invariant mass (left) and the transverse momentum of the top quark (right). The ATLAS and CMS results are compared to theoretical predictions at NNLO and NLO+NNLL.

for the top quark decay width  $\Gamma_t$ . A direct measurement of  $\Gamma_t$  is of great importance, because the width would be affected by any non-expected decay modes of the top quark, whether they are observed or not. Unfortunately,  $\Gamma_t$  cannot be directly measured in the  $t\bar{t}$  sample at hadron colliders, but its main component can be accessed through single top processes. If there are only three generations, the unitarity constraint of the CKM matrix implies that  $|V_{tb}|$  is very close to unity. But, the presence of a heavy fourth generation quark with a large CKM coupling to the top quark could be consistent with large values of  $B(t \rightarrow Wb)$ , while resulting in an almost entirely unconstrained value for  $|V_{tb}|$ . A direct measurement of  $|V_{tb}|$  can therefore explore the possibility of a fourth generation, and confirm that the top quark is indeed the SU(2) partner of the bottom quark. A measurement of the single top quark production cross section provides the only known way to directly measure  $|V_{tb}|$  at a hadron collider. Figure 12 present the most recent ATLAS and CMS measurements of the single top production cross section in its three channels ( $s$ ,  $t$  and  $Wt$  associated production) and for three center of mass energies, while Fig. 13 summarizes the corresponding extractions of the CKM matrix element  $|V_{tb}|$  [19]. Good overall agreement is observed between measurements and the predictions.



**Fig. 12:** Comparison of measured single top production cross sections as a function of the center of mass energy with various theoretical predictions. Good overall agreement is observed between measurements and the predictions.



**Fig. 13:** Summary of the ATLAS and CMS extractions of the CKM matrix element  $V_{tb}$  from single top quark measurements. Good overall agreement is observed between measurements and the predictions.

## 6 Top quark properties

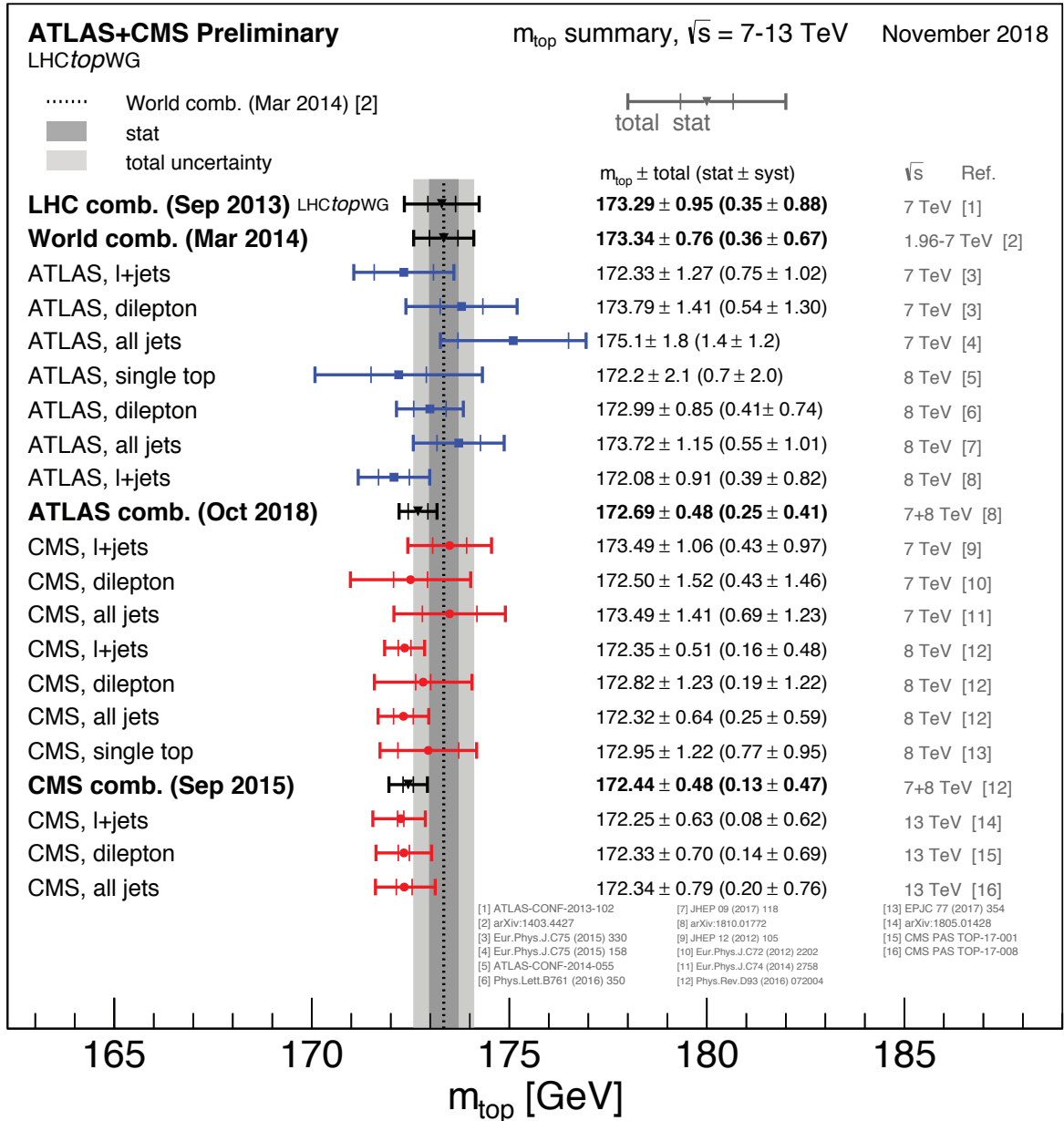
Top quark properties are predicted by the SM and can be modified by BSM effects at the production or decay level. Precision measurements can uncover deviations of SM predictions which could serve as indirect evidence of BSM processes due to particles with masses not currently accessible at the LHC. Analyses typically rely on the  $t\bar{t}$  1+jets channel, due to its large samples, low background and constrained final state. Dedicated sensitive observables are defined for each property, with the LHC benefiting from and expanding the methods developed for those same studies at the Tevatron.

From the many properties of interest for the top quark, few are so fundamental as its mass. The ATLAS and CMS collaboration have published measurements of the top quark mass using candidate events in all  $t\bar{t}$  decay channels. The top quark mass has also been measured from single top events and extracted indirectly from the  $t\bar{t}$  cross section measurement. Comparing precision measurements of the top quark mass, the  $W$  and the Higgs boson masses with SM predictions provides a powerful tool to search for BSM effects. Deviations would be an indication of new physics in the mass loops.

The most precise measurements of the top quark mass  $m_t$  are obtained from the lepton jets  $t\bar{t}$  sample, where the mass is reconstructed from a kinematic fit of the decay products to the  $t\bar{t}$  hypothesis. The mass is determined from a simultaneous fit of  $m_t$  and the main uncertainties arising from the jet energy scale and the b-jet energy scale. The contributions from these uncertainties are statistical in nature, and will benefit from larger data samples. Figure 14 presents a summary of the direct ATLAS and CMS top  $m_t$  measurements [19], compared to the LHC and World averages. The results show good agreement between measurements and with the world average.

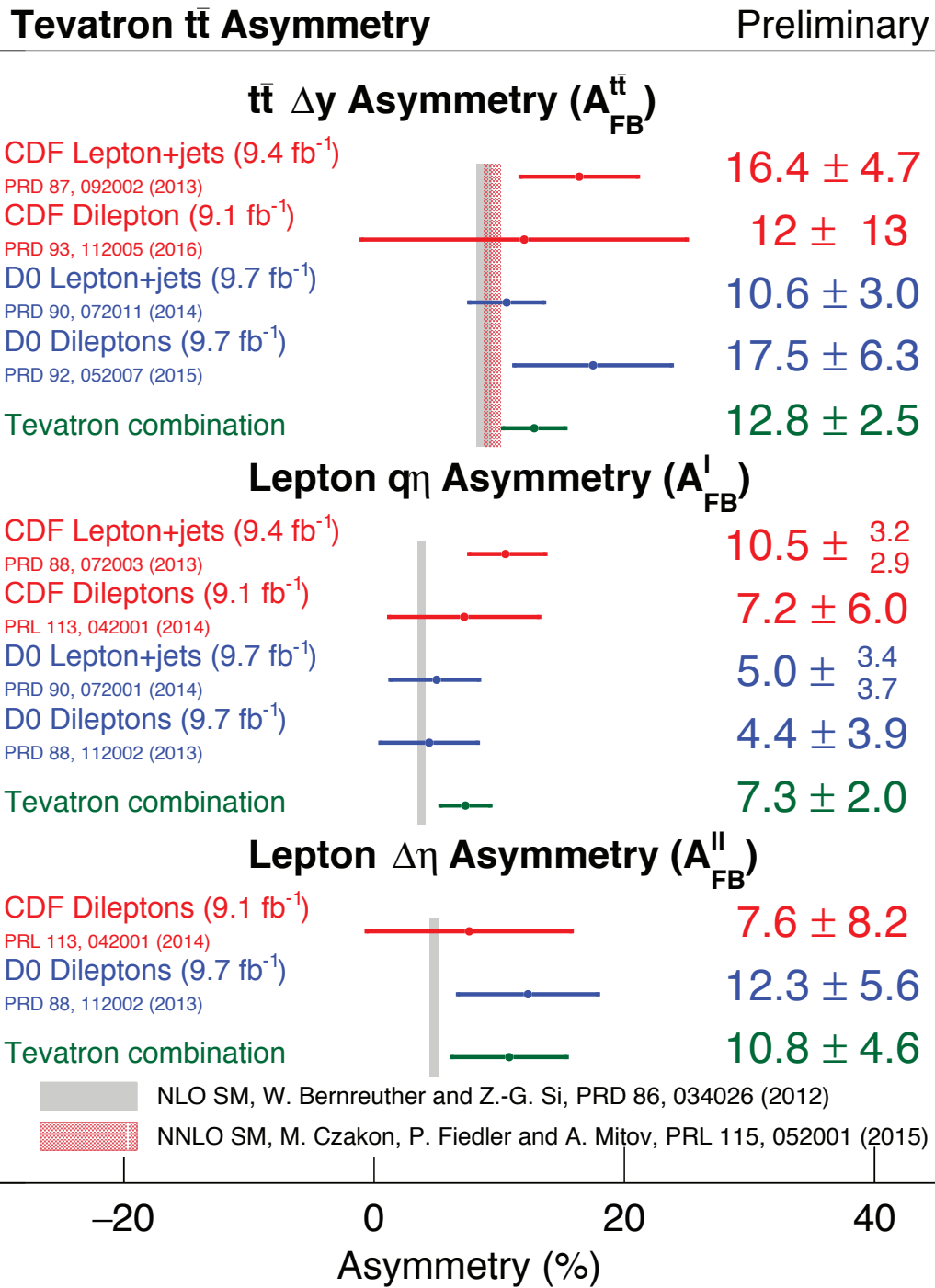
A fundamental difference between the  $t\bar{t}$  production at the Tevatron and at the LHC is that the production is dominated by  $q\bar{q}$  annihilation in the former and by gluon fusion in the latter. At leading order, the SM predicts that the  $t\bar{t}$  production is forward-backward symmetric in  $q\bar{q}$  annihilation. However, higher order SM effects result in a small ( $\approx 6.6\%$ ) positive asymmetry  $A_{FB}$ , such that the top (anti-top) quark is preferentially emitted in the direction of the incoming quark (anti-quark) [20]. BSM production mechanisms that exchange new bosons could enhance  $A_{FB}$ . There is no asymmetry in the gluon fusion  $t\bar{t}$  production that dominates at the LHC, but because quarks carry, on average, larger momentum than antiquarks, the rapidity distribution of top quarks is expected to be broader than that of anti-tops, which results in a  $t\bar{t}$  charge asymmetry  $A_C$  of ( $\approx 1\%$ ). Early Tevatron  $A_{FB}$  measurements [21, 22], based on about half the data that would eventually become available, sparked a huge interest when they showed larger asymmetries than those predicted by the SM at the time [23], especially because the discrepancies grew with larger top quark pair masses and rapidity difference. Measurements using the full Tevatron dataset, and combining the results from the two collaborations, recently became available [24]. Even though all measurements favor somewhat larger positive asymmetries than the predictions [25], none of the observed differences are larger than 2 standard deviations, as can be seen in Fig. 15. The ATLAS and CMS collaborations have combined their inclusive and differential measurements of  $A_C$  at two center of mass energies (7 and 8 TeV), obtaining  $A_C = 0.005 \pm 0.007(\text{stat}) \pm 0.006(\text{syst})$  and  $A_C = 0.0055 \pm 0.0023(\text{stat}) \pm 0.0025(\text{syst})$  at 7 and 8 TeV, respectively, in good agreement with the respective SM predictions [26]. Figure 16 shows the measured combined inclusive  $A_C$  at 8 TeV versus the combined Tevatron  $A_{FB}$  compared with the SM prediction at NNLO+EW NLO [27] and various BSM predictions that could affect the asymmetries. The combined Tevatron/LHC measurements uniquely restrict the phase space of possible BSM phenomena which would produce large asymmetries, including models that predict the existence of heavy  $W'$  bosons, heavy axigluons, scalar isodoublets, color triplet scalars and color sextet scalars [28].

A property characterizing the dynamics of the top-quark decay is the helicity state of the on-shell  $W$  boson. The  $W$  boson can have three possible helicity states, and the fractions of  $W^+$  bosons produced in these states are denoted as  $f_0$  (longitudinal),  $f_-$  (left-handed), and  $f_+$  (right-handed). In the SM, the top quark decays through the  $V - A$  weak charged-current interaction, which strongly suppresses right-handed  $W^+$  bosons or left-handed  $W^-$  bosons. Significant deviations from these expectations would

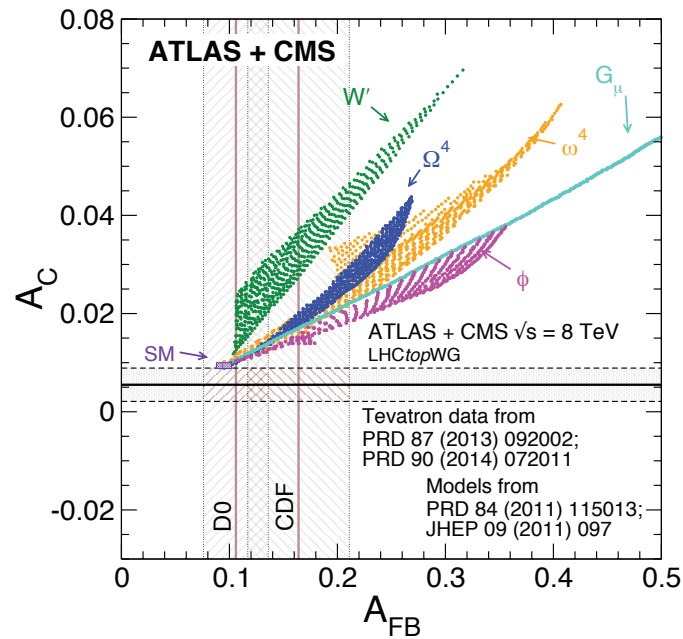


**Fig. 14:** Summary of the ATLAS and CMS direct top quark mass measurements compared to the LHC and Tevatron+LHC combinations.

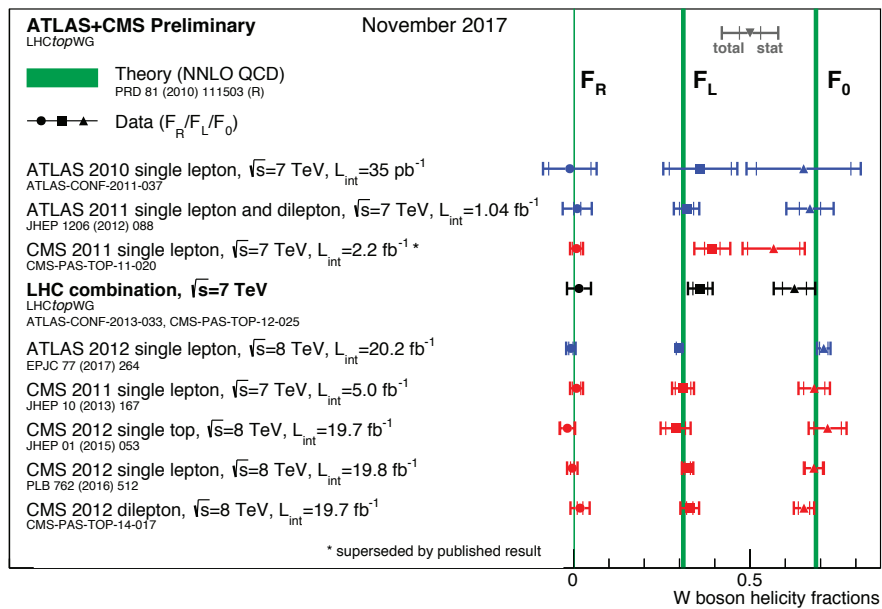
indicate either a departure from the  $V - A$  structure of the  $tWb$  vertex or the presence of a non-SM contribution to the  $t\bar{t}$  candidate sample. Both ATLAS and CMS have measured the helicity fractions to a precision better than 5% [19] and found good agreement with the SM prediction at NNLO [29], as shown in Fig. 17.



**Fig. 15:** Summary of inclusive forward-backward asymmetries in  $t\bar{t}$  events at the Tevatron. Even though all measurements favor somewhat larger positive asymmetries than the predictions [25], none of the observed differences are larger than 2 standard deviations.



**Fig. 16:** Combined inclusive LHC  $A_C$  at 8 TeV versus the combined Tevatron  $A_{FB}$  compared with the SM and BSM predictions. The combined Tevatron/LHC measurements uniquely restrict the phase space of possible BSM phenomena which would produce large asymmetries.



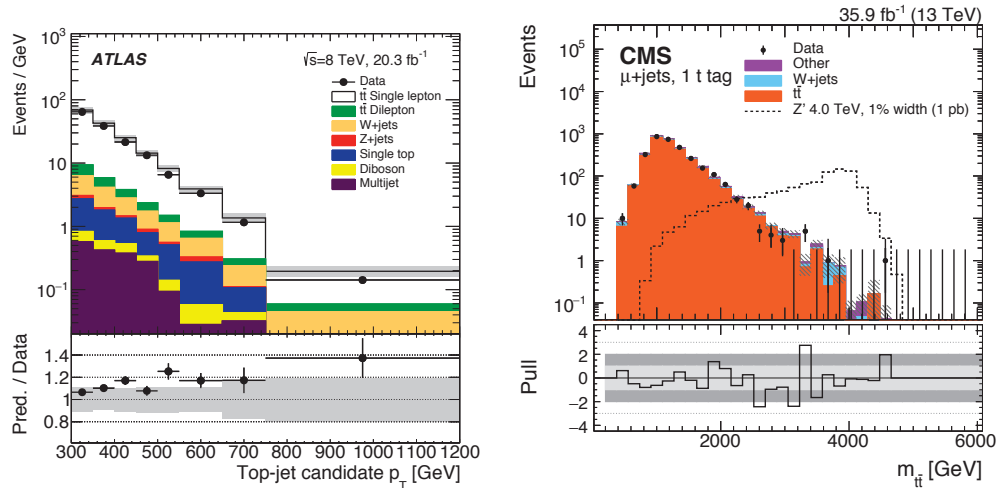
**Fig. 17:** Summary of the  $W$  boson helicity fractions from top quark decays measured in  $t\bar{t}$  events collected at 7 and 8 TeV by the ATLAS and CMS collaboration. The results, which have a precision better than 5%, agree well with SM NNLO predictions.

## 7 Boosted top quark production

At Tevatron energies, the majority of  $t\bar{t}$  events were produced at rest. This has changed at the LHC, where the higher center of mass energies result in top quarks that are often produced with a high Lorentz-boost in momentum, which yields decay products that are partially or fully merged because the angular distance between partons is smaller than the jet clustering distance parameter. As a consequence, the three quarks from the hadronically decaying top quark may be reconstructed as one fat jet, and similarly, for the leptonically decaying top quark, the lepton may appear as non-isolated due to its proximity to the b-quark. Special techniques were developed by both ATLAS and CMS to reconstruct these boosted top quarks, in both the hadronic and the leptonic case. The cornerstone of these techniques relies on the ability to reconstruct a single jet that contains the full energy of the decay. This “fat jet” is then distinguished from ordinary jets through the identification of the jet internal structure. In the case of the hadronic decay of the top quark, the internal structure can identify the b-jet and the individual light jets that result from the  $W$  boson decay. In the case of the leptonic decay of the top quark, the energy deposits from the leptons are identified and excluded from the b-jet reconstruction. This is particularly challenging in the case of the electron, as no isolation requirement can be applied.

The ATLAS collaboration has measured the  $t\bar{t}$  differential production cross section at  $\sqrt{s} = 8$  TeV for lepton jets events with high transverse momentum. The measurement is reported as a function of the transverse momentum of the hadronically decaying top quark for values of  $p_T > 300$  GeV [30]. Figure 18 (left) shows the  $p_T$  of the leading fat jet compared to the SM prediction. The lower panel shows the ratio of the MC prediction to the data, where good agreement is observed within the uncertainties.

The CMS collaboration has searched for boosted anomalous resonant  $t\bar{t}$  production in events with zero, one and two leptons [31], by reconstructing the  $t\bar{t}$  invariant mass distribution and focusing in the area of masses of at least 1 TeV. Figure 18 (right) shows the  $t\bar{t}$  invariant mass for events in the muon+jets channel in the cases in which the hadronic top has been reconstructed as a fat jet. No excess of data over the SM predictions is observed, and limits are set on the production cross section times branching fraction, probing a region of parameter space for certain models of new physics not yet constrained by precision measurements.



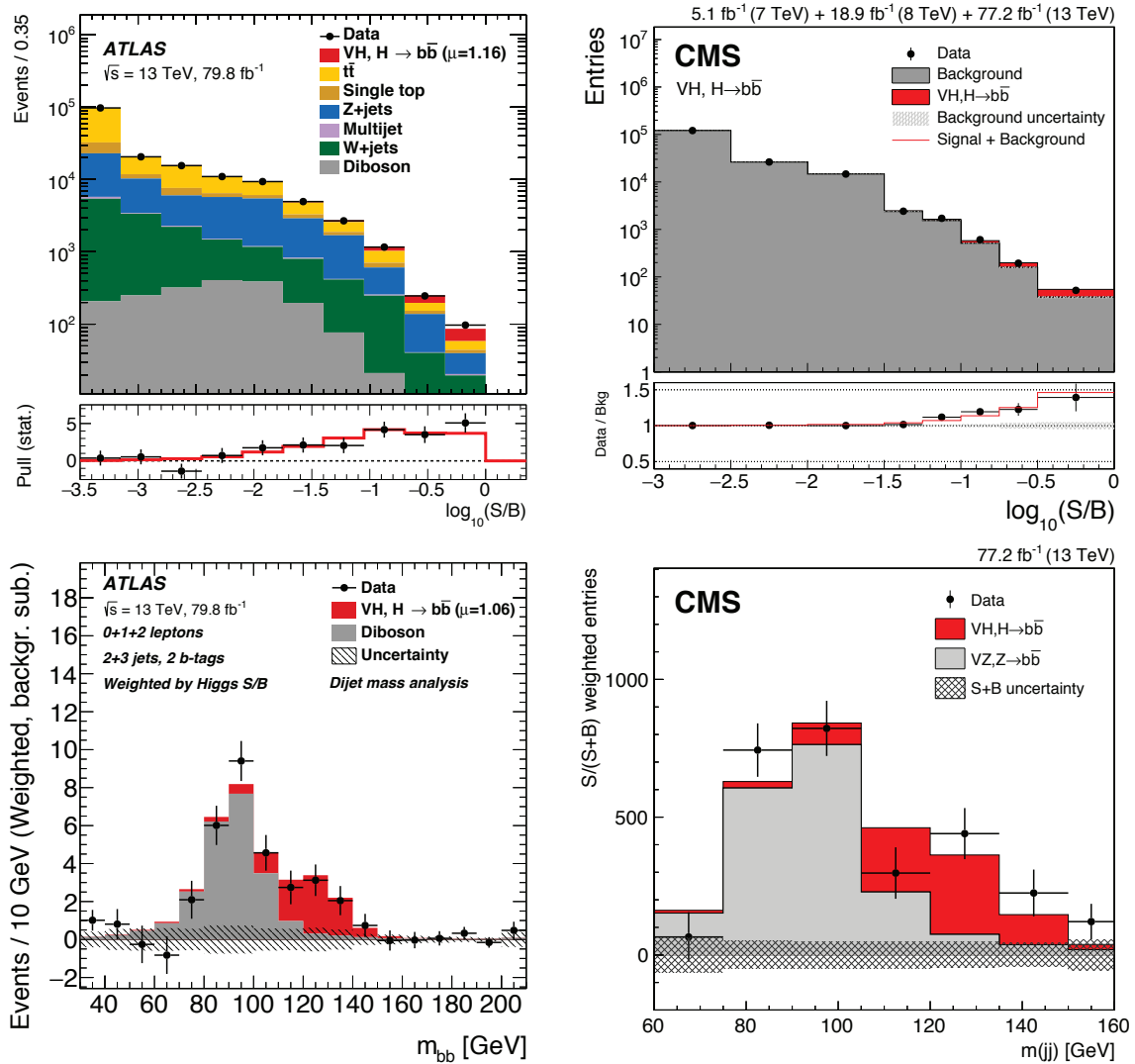
**Fig. 18:**  $t\bar{t}$  differential production cross section in the lepton jets channels as a function of the  $p_T$  of the hadronically decaying top quark for values of  $p_T > 300$  GeV (left). Invariant mass of  $t\bar{t}$  events in the boosted lepton jets channel in the cases in which the hadronic top has been reconstructed as a fat jet (right). The lower panels show the ratio of the MC prediction to the data, where good agreement is observed within the uncertainties.

## 8 Higgs boson studies

The Higgs Boson, that had been predicted by the SM since the seventies, was observed on July 4, 2012 by the ATLAS [32] and CMS [33] collaborations. The analyses used the fully reconstructed Higgs boson decay products and their excellent mass resolution to drive the discovery; however, the channel in which the Higgs boson decays into a  $b\bar{b}$  pair, even though it represents 58% of the branching ratio, was not observed. In the years since the discovery, both collaborations worked hard to extract the  $H \rightarrow b\bar{b}$  signal from the overwhelming multi-jet  $b\bar{b}$  background taking advantage of the associated production, in which the Higgs boson is produced together with a  $W$  or a  $Z$  boson. That process had already been used by the LEP collaborations to exclude Higgs masses below 114.4 GeV [34], and by the Tevatron collaborations to claim evidence of a observation of the Higgs boson with a significance of  $3.3\sigma$  [35]. At the LHC, both collaborations select events with zero, one or two charged leptons (electrons or muons) to include the cases in which the  $Z$  boson decays to two neutrinos, and the leptonic decays of the  $W$  and the  $Z$  boson, respectively. In all cases, the events are required to include two b-jets. Several tools were developed to make the signal accessible. First, the vector boson was required to be highly boosted, which had the effect of suppressing the QCD multi-jet background and also of merging the  $b\bar{b}$  pair so that it was reconstructed as a single fat jet. Jet substructure techniques and a special b-jet tagger based on a deep neural network discriminant were developed to identify the  $b\bar{b}$  jet with high efficiency and a low 0.1% misidentification rate. The resolution of the  $b\bar{b}$  invariant mass was improved by applying a multivariate regression technique. A deep neural network was then used together with 7 signal and 21 background regions to simultaneously extract the background normalization and the signal. Figure 19 shows the output of the multivariate discriminant and the dijet invariant mass distribution. Both collaborations obtained significances larger than  $5\sigma$  and the observation of the  $VH(b\bar{b})$  process [36, 37]. Both collaborations also measured the signal strength, defined as the ratio of the number of observed  $H \rightarrow b\bar{b}$  events over the number predicted by the SM, to be consistent with 1 within an uncertainty of about 20%.

Seven years after the Higgs boson observation, Higgs physics has entered a precision era. The  $\gamma\gamma$ ,  $ZZ$  and  $WW$  decays were firmly established and the Higgs mass has been measured to a 0.15% precision. The Yukawa mechanism has been established in the last two years by the observation of the  $\tau\tau$  and  $b\bar{b}$  decays and the  $t\bar{t}H$  process. And differential cross section measurements are being used to compare the data to state-of-the-art calculations. But there is still a lot to learn, in particular, searching for di-Higgs production is vitally important to start to understand the self-couplings of the Higgs.





**Fig. 19:** Top: Event yields as a function of the logarithm of signal over background for data, background and a 125 GeV Higgs boson for ATLAS (left) and CMS (right). Bottom: The distribution of the invariant mass of the  $b\bar{b}$  pair, with the Higgs boson signal shown in red.

### 9 Searches for beyond the Standard Model processes

Since its inception more than 50 years ago, the SM has been extremely successful in predicting the existence of particles and processes that were later observed in experiments. In particular, the ATLAS and CMS collaborations have developed a broad and rich program, probing processes that span 9 orders of magnitude in production cross section. Some of the rare processes had not been observed before and are being produced at rates comparable with those expected from BSM predictions. Figure 20 shows a summary of SM total production cross section measurements, corrected for leptonic branching fractions, compared to the corresponding theoretical expectations and ratio with respect to best theory. Figure 21 shows a summary of the cross section measurements for SM processes by the CMS collaboration.

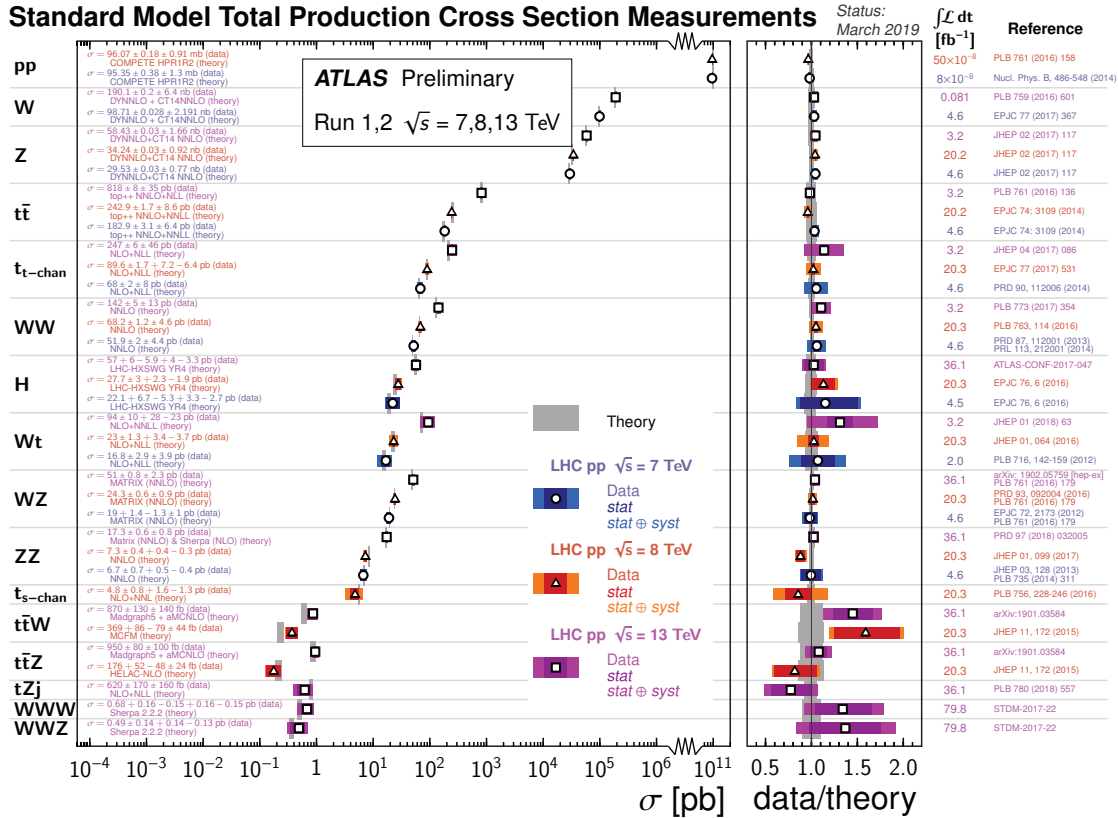
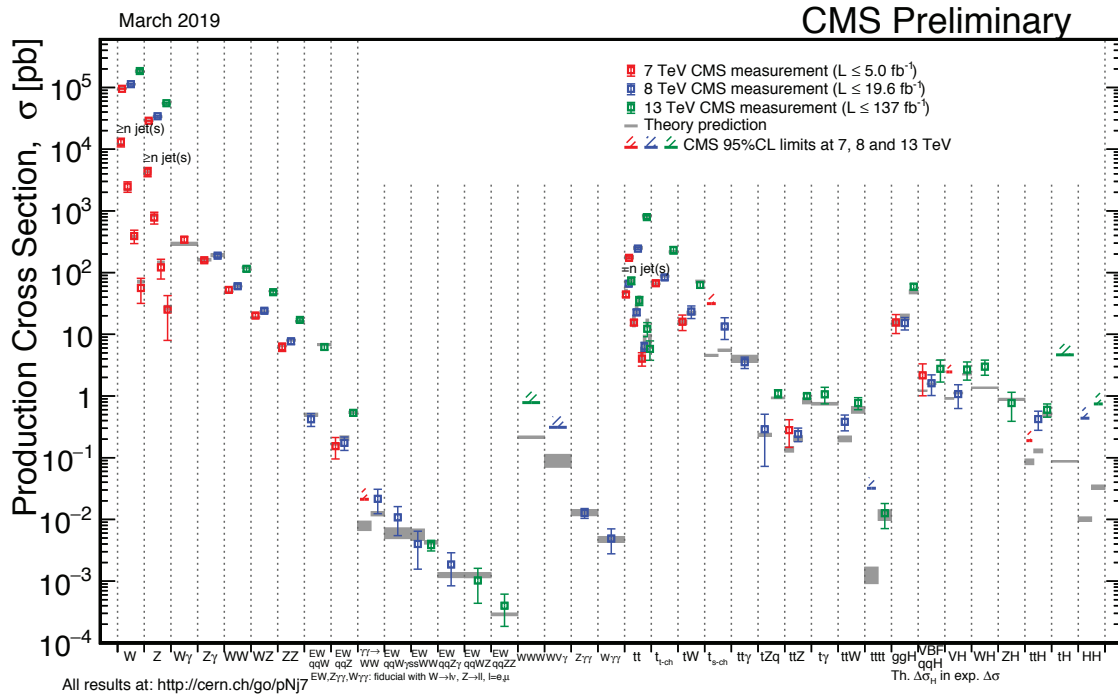


Fig. 20: Summary of measured SM production cross sections reported by the ATLAS collaboration, corrected for leptonic branching fractions, and compared to the corresponding theoretical expectations.

In spite of the SM success, there remain several unanswered questions: why are there exactly three generations of quarks and leptons? Are quarks and leptons actually fundamental, or made up of even more fundamental particles? Why can't the SM predict a particle's mass? How come neutrinos have mass? Why do we observe matter and almost no antimatter if we believe there is a symmetry between the two in the universe? What is this "dark matter" that we can't see that has visible gravitational effects in the cosmos? How does gravity fit into all of this? And the uncomfortable issue of fine tuning. The BSM front runner around 2007 was supersymmetry (SUSY), that hypothesizes that a symmetry exists between fermions and bosons in which each boson has a fermion super-partner with the same mass and quantum numbers and vice-versa. These superpartners contribute with opposite sign to the loop corrections to the Higgs mass providing cancellation of the divergent terms. SUSY is theoretically compelling, providing a solution to the Higgs hierarchy problem, allowing unification of gauge couplings, and even predicting a dark matter particle candidate. For experimentalists SUSY predicted the existence of a large number of

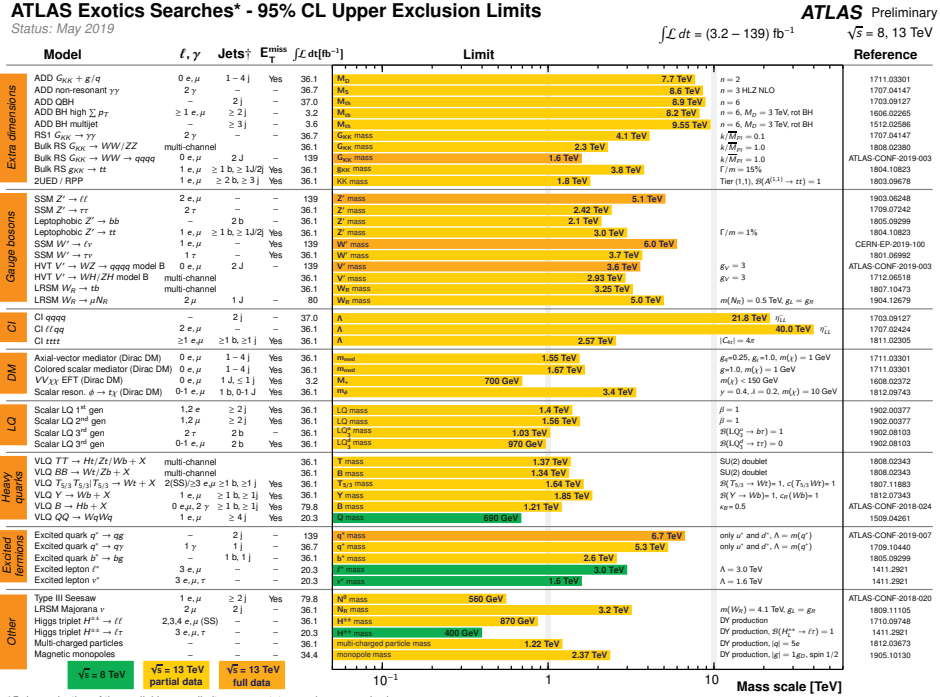


**Fig. 21:** Summary of the measured cross sections for SM processes as reported by the CMS collaboration. Each result is compared to the best available theoretical prediction.

new particles, giving rise to striking experimental signatures ready to be discovered. The LHC years have witnessed a systematic exploration of TeV scale gluinos and squarks, and the LHC collaborations are just starting to gain sensitivity to Higgsinos, which in many models are expected to be the lowest mass SUSY particles with masses around a few hundreds GeV. Both collaborations are also in the process of extending their searches to unexplored regions of parameter space characterized by challenging manifestations of SUSY. In parallel to the continuing exploration of the SUSY parameter space, both collaborations have also vigorously pursued a plethora of non-SUSY BSM ideas, from new gauge bosons, quark compositeness, high-mass resonances, extra dimensions and back holes, both in a model-independent and in a model-directed way, making sure no stone is left un-turned. Examples of a selection of results of such searches can be seen in Figs. 22 and 23 for the ATLAS and CMS collaborations, respectively.

In recent years, many novel techniques have been developed that rely on alternative methods to trigger and reconstruct events. One such example is the search for massive long-lived particles that would lose their kinetic energy and stop while traversing the detector. Such particles would give rise to energy deposits in the calorimeters or the muon systems, but have no associated hits in the tracking detectors (displaced particles). Alternatively, they would appear as tracks with no associated hits in the calorimeters or beyond (disappearing particles). These signatures pose a difficult experimental challenge and depend on modified object reconstruction techniques that do not assume the presence of a prompt vertex and rely on timing information for the energy deposits. Figure 24 shows a diagram of these hypothetical particles and their signatures and where they would decay depending on their lifetime.

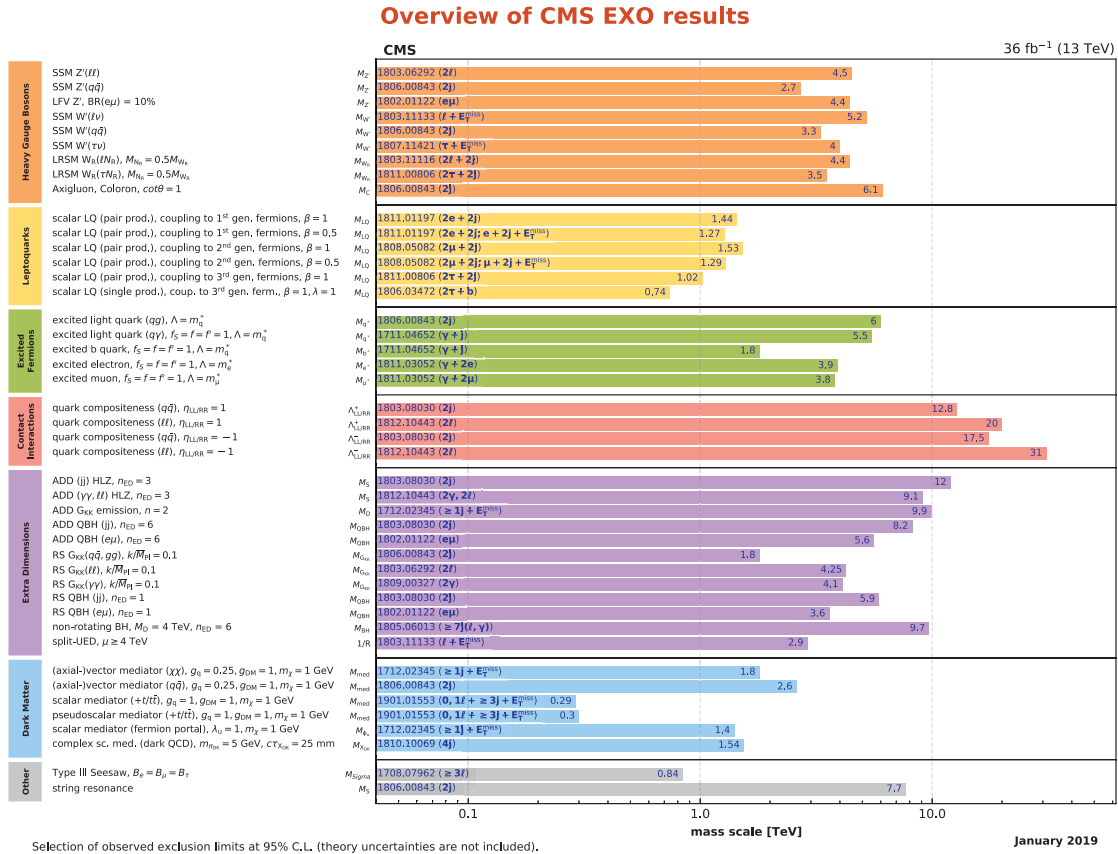
One example of such an analysis is the search for a single neutral long-lived particle  $Z_d$ , decaying hadronically, produced in association with a SM  $Z$  boson which decays leptonically to electrons or muons [38]. Such  $Z_d$  particle would be produced in popular scenarios in hidden or dark sector models with additional  $U(1)_d$  dark gauge symmetry [39,40] and would travel from a few centimeters to hundreds of meters, depending on the model. The ATLAS collaboration selected events with two opposite-sign isolated leptons and a jet that has no associated tracks and a timing inconsistent with out-of-time pileup



**Fig. 22:** A representative selection of available results from the ATLAS collaboration on searches for new phenomena other than SUSY. Green bands indicate 8 TeV data results; yellow (orange) bands indicate 13 TeV data results with partial (full) dataset.

and beam-induced backgrounds. The timing of the jet is obtained from the timing of its constituent calorimeter cells as measured by the tile calorimeter, relative to the expected time-of-flight from the bunch crossing to the cell [41]. Figure 25 shows the reconstruction efficiency for the jet as a function of the transversal decay length and the obtained 95% CL limits as a function of the decay length of  $Z_d$ . As can be seen, the efficiency is high beyond the volume of the tracker and maximum within the hadronic calorimeter. No significant excess of events is observed above the expected background, which allows the ATLAS collaboration to set limits on the production cross section of the  $Z_d$  particle as a function of its mass for decay lengths from a few centimeters to one hundred meters.

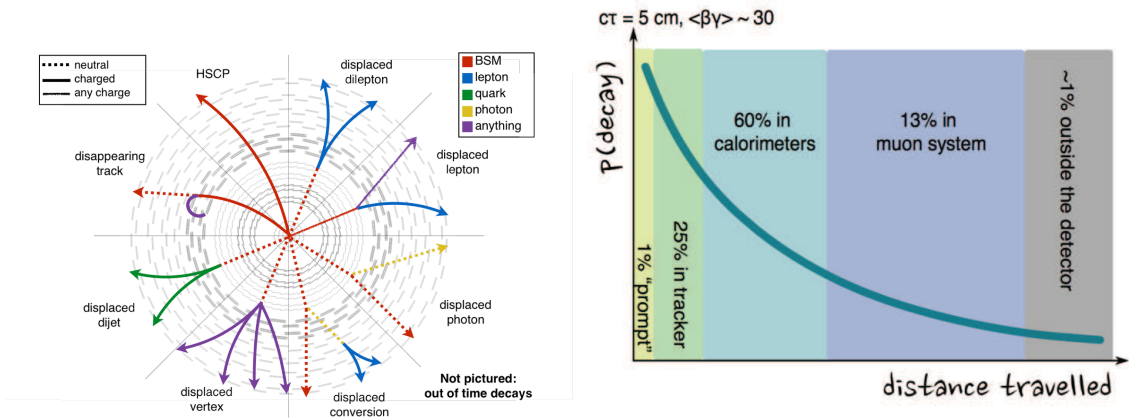
The CMS collaboration also searched for the decay of heavy long-lived particles that come to rest in the detector, and whose decays would be visible during periods of time well separated from the pp collisions [42]. For particles with lifetimes longer than tens of nanoseconds, their decays would be reconstructed as part of a separate event from their production. The search thus focuses on times when there are no proton bunches in the detector. Two cases are considered: a hadronic decay that would be detected as a large energy deposit in the calorimeters in the interval between collisions, and the case in which the particle decays into muons and appears as displaced muon tracks out of time with the collision. Dedicated triggers were deployed that are live only during specific time windows when the detector was quiet. Backgrounds arise from cosmic rays, beam halo, and detector noise, and are estimated from control samples. Figure 25 (left) shows the difference in the time of the muon track between the upper and the lower hemisphere for data and estimated backgrounds. The data agrees well with events expected from cosmic rays. Figure 26 (right) shows an example of 95% CL upper limits obtained from the calorimeter search in the neutralino vs. gluino mass plane, for lifetimes between 10 microseconds and 1000 seconds.



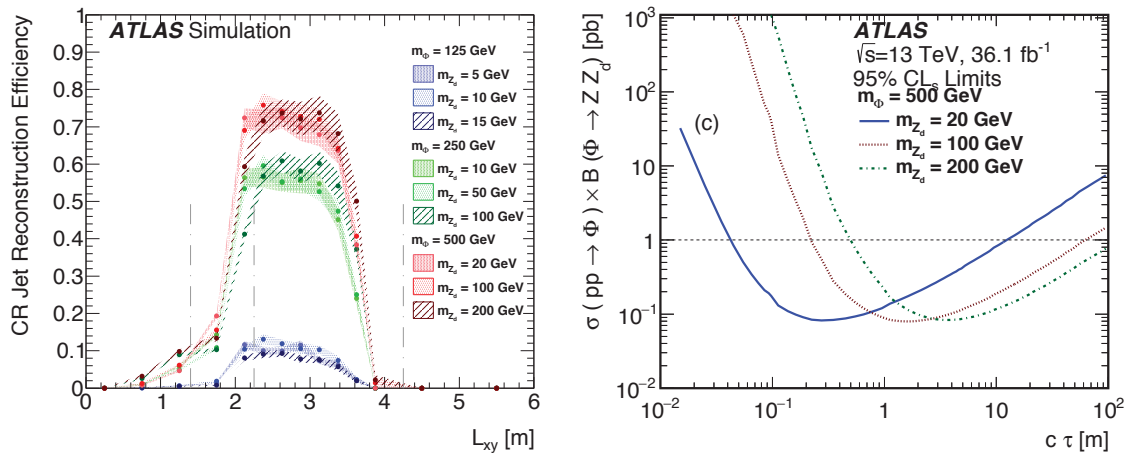
**Fig. 23:** Bar chart representing the mass scale reach of CMS BSM analyses using data collected in 2016 for a selected set of new physics phenomena.

## 10 Dark matter searches

Cosmological observations indicate that 85% of the matter of the universe is dark matter (DM). However, there is no evidence yet for non-gravitational interactions between DM and SM particles. If such interaction existed, hadron colliders would offer a complementary strategy to look for non-gravitational DM interactions via the collision of SM particles at high energies. DM candidates are assumed to be weakly interacting and would leave no signal in the detectors. However, they could be identified by looking at the production of other particles decaying against them, giving rise to spectacular signatures in which jets or heavy particles would be seen recoiling against missing transverse energy. Early searches presented their results using effective field theory operators to describe the DM-SM interaction, allowing for direct comparison with non-collider searches in the contact interaction approximation. For cases in which the mediator of the DM-SM interaction is not very heavy, simplified models need to be used that include the particles and their BSM interactions and are valid at LHC energies [43]. These models are described by a small number of free parameters but make it harder to compare with direct and indirect detection experiments. Nevertheless, a rich phenomenology of DM searches at colliders has been pursued in the X+MET topology, where X includes single jets, photons,  $W$  and  $Z$  bosons, top and bottom quarks (single and pair). In all cases, control regions are used to understand the background contribution to the signal region and to ensure that spurious detector signals do not appear as fake missing transverse energy, mimicking the DM signal. Results are typically presented as exclusion plots in the DM vs. mediator mass plane, and the spin-independent (SI) or spin-dependent (SD) DM-nucleon or DM-proton cross section vs. DM mass. Figure 27 shows examples in which LHC searches from ATLAS [44] (left) and CMS [45] (right)



**Fig. 24:** Diagram of hypothetical long-lived particles and their signatures in the detectors (left). Distance travelled by a long-lived particle depending on its lifetime (right).

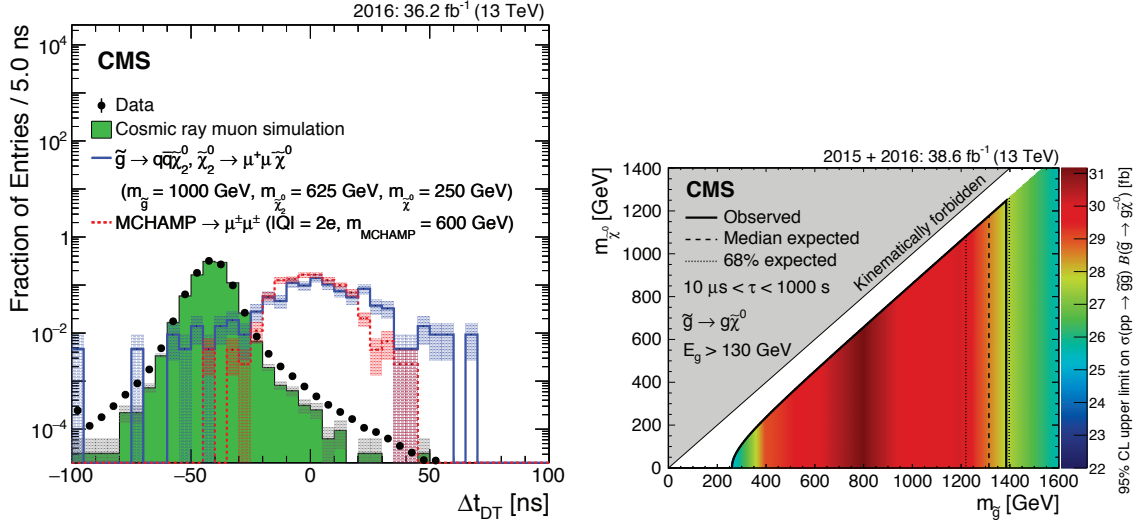


**Fig. 25:** Reconstruction efficiency for the jet as a function of the transversal decay length (left). 95% CL limits as a function of the decay length of  $Z_d$  (right).

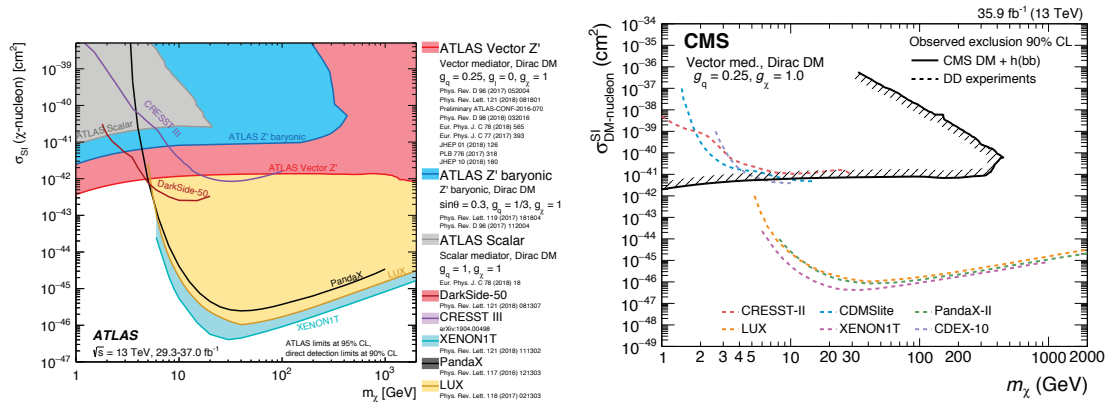
are compared to a selection of direct detection (DD) experiments. For the ATLAS result, the shaded areas are excluded both for the collider and the DD results. For the CMS result, the regions above the curves are excluded for DD experiments. The reinterpretation of the collider results in terms of a nucleon scattering cross section yields a higher sensitivity for lower masses than existing results from DD experiments, under the assumptions imposed by the model.

## 11 The High-Luminosity LHC era

The LHC run plans for the next twenty years is well defined and summarized in Fig. 28. We are currently in the Long Shutdown 2 (LS2) period after having collected about  $150 \text{ fb}^{-1}$  of data during Run 2 and Run 3, currently scheduled for 2021–2023, will see an average of up to 80 simultaneous proton interactions per beam crossing, peak instantaneous luminosities of  $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  and an anticipated increase of the center of mass energy to 14 TeV. The total integrated luminosity expected for Run 2 and 3 is  $300 \text{ fb}^{-1}$ . After an extended long shutdown 3 (LS3), currently scheduled for 2024–2026, the proposed High-Luminosity LHC (HL-LHC) Run 4 would begin in late 2026, with an average number of simulta-



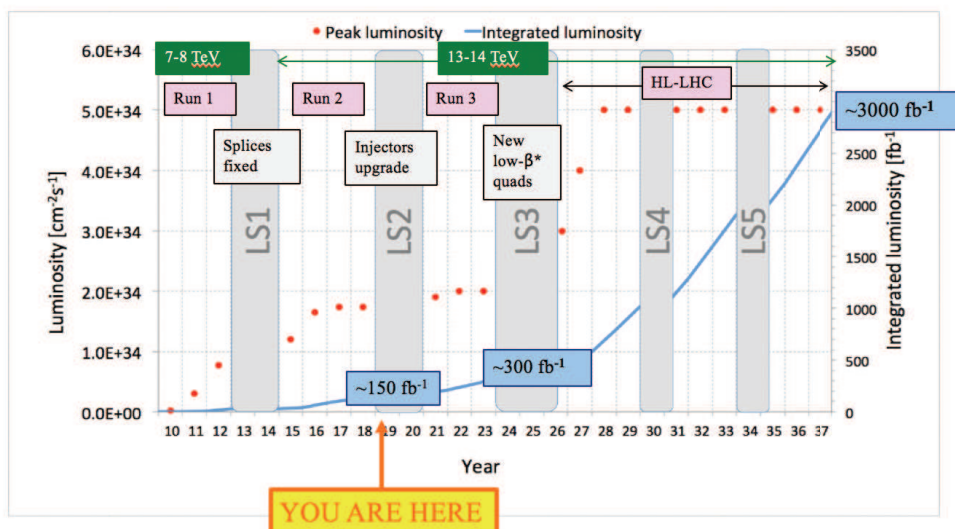
**Fig. 26:** Difference in the time of the muon track between the upper and the lower hemisphere for data and estimated backgrounds (left). Example of 95% CL upper limits obtained from the calorimeter search (right).



**Fig. 27:** Examples of dark matter searches in which the collider results are compared with results from direct detection experiments. Plots included as examples only and taken from [44] (left) and CMS [45] (right).

neous pp collisions per beam crossing of 120, peak instantaneous luminosities of  $5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ , and a total integrated luminosity of  $3000 \text{ fb}^{-1}$ . The ATLAS and CMS collaborations are planning a series of upgrades [46–49] that will ensure the capabilities of the detector are matched to the running conditions expected from the LHC machine, while taking the opportunity to improve the performance and repair any problems uncovered during data-taking periods. The installation of the Phase 1 upgrades will be completed during LS2, and the Phase 2 upgrades are planned to coincide with LS3.

During the HL-LHC era, the large dataset will be delivered at the cost of having as many as 200 concurrent pp interactions every 25 ns and large radiation doses. Both ATLAS and CMS are planning on significant changes to their detectors to maintain their performance in these challenging conditions. In particular, both detectors will add timing capabilities to cope with the increased pileup and preserve the lepton identification (via isolation), the b-tagging effectiveness (via primary vertex reconstruction and combinatorics) and energy measurement of the jets. Both experiments will also replace their entire tracking detectors, with ATLAS joining CMS in having an all-silicon tracker. Both collaborations will



**Fig. 28:** The LHC run plans for the next twenty years, in which we expect to collect up to  $3000 \text{ fb}^{-1}$  of data, most of it at a center of mass energy of 14 TeV.

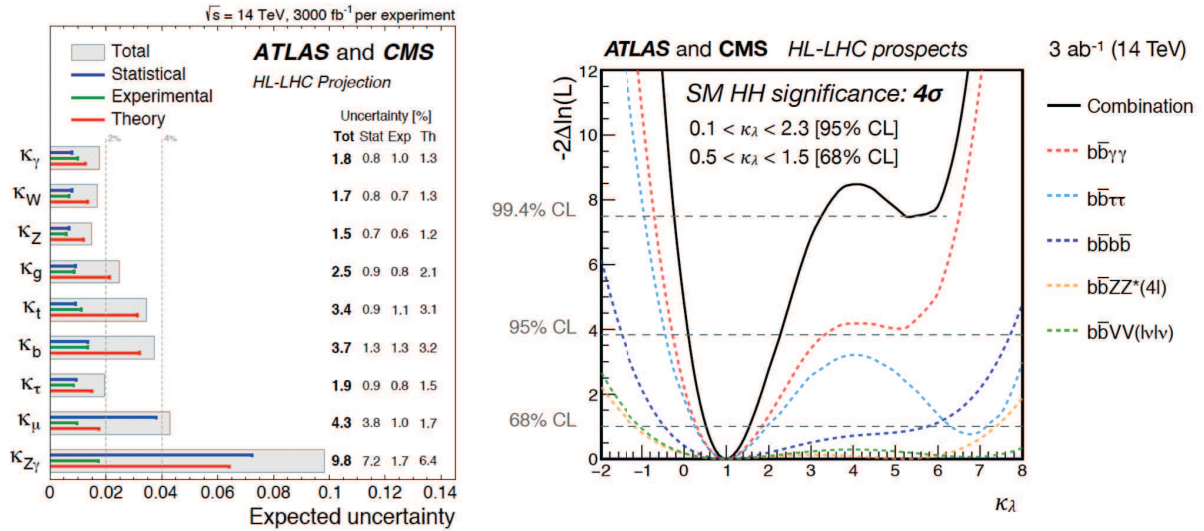
increase the granularity and the coverage of their tracking volumes and reduce the material, which will allow them to preserve the reconstruction efficiency. Trigger capabilities require significant improvements to preserve the trigger thresholds and are implemented by installing higher bandwidth readout systems and adding fast tracking to the first trigger level. In addition, detectors that would be damaged by radiation are also being replaced by higher granularity, radiation-hard options.

Physics Projection studies for the HL-LHC were prepared and submitted to the CERN Council Open Symposium on the Update of the European Strategy for Particle Physics [50–52]. From those studies, it is clear that the HL-LHC has an uncontested leadership in areas of direct searches for new particles, precision measurements of the Higgs boson, measurements of precision electroweak parameters and closure test ( $W$  boson mass, top quark mass, Higgs boson mass), and some topics in rare B decays and other topics in B physics.

One of the main goals of the HL-LHC studies is to measure the Higgs couplings to a precision close to the percent level. Another goal is the measurement of the Higgs boson trilinear self-coupling  $\gamma_{HHH}$  via the study of di-Higgs production. Figure 29, see Ref. [52], shows that the Higgs couplings are expected to be measured with a precision that would be sensitive to new physics, and that the measurements of the Higgs trilinear interaction would provide constraints on the shape of the Higgs potential close to the minimum, verifying the electroweak symmetry breaking mechanism of the Standard Model.

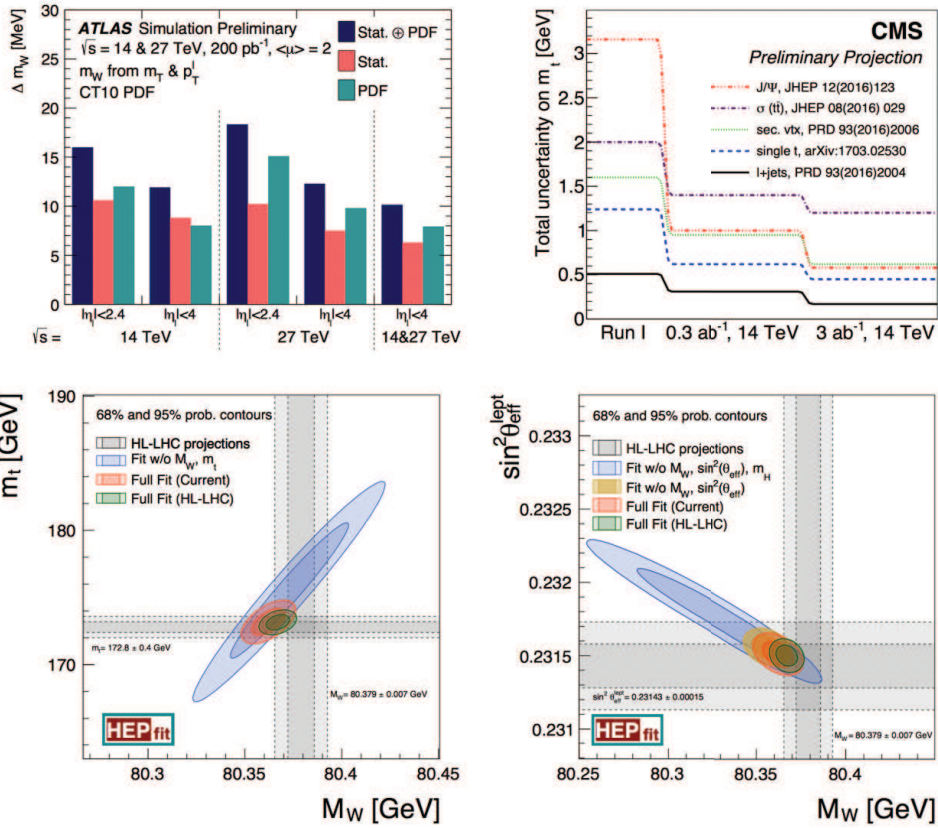
The LHC results have confirmed the predictions of the Standard Model to unprecedented precision, and it is expected that the data collected during the HL-LHC era will extend the sensitivity to possible anomalies that might indicate the presence of new physics. Of particular interest are the studies of precision electroweak measurement and precision top quark physics, which can be combined together in a global fit of electroweak precision observables now that the Higgs Boson mass has been measured. This last input to the global fit of electroweak precision observables (EWPO) can be used to constrain new physics, a key goal of the HL-LHC physics program. Figure 30 shows the projections on the uncertainty of the top quark and the  $W$  boson mass, as well as comparisons between the indirect constraints and current and projected measurements. What is apparent from the comparisons is that if the central values of the measured inputs were to remain unchanged, the expected improvement on their uncertainties would significantly increase the tension between the indirect determinations from the electroweak fit and the corresponding measurements.



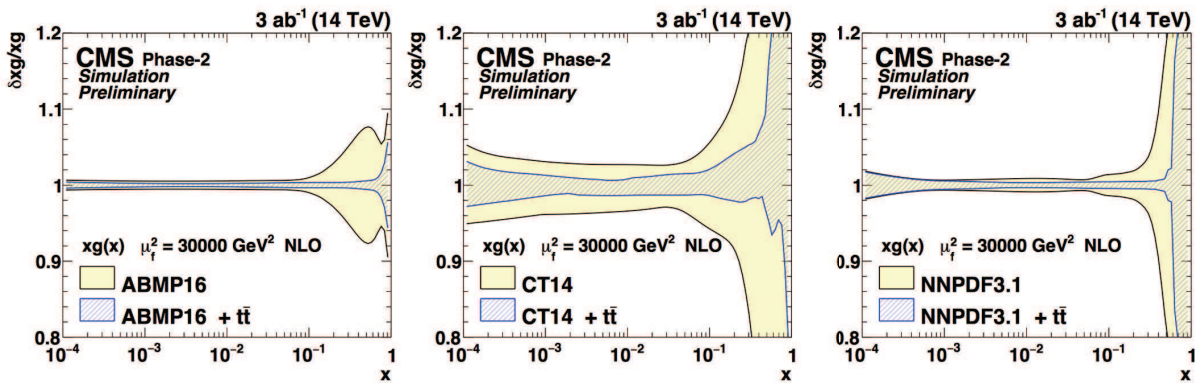


**Fig. 29:** Expected uncertainties on the coupling modifier parameters that are introduced to investigate potential deviations from the Standard Model prediction of the Higgs boson couplings to bosons and fermions (left). Minimum negative log-likelihood as a function of  $\kappa_\lambda = \gamma_{HHH}/\gamma_{HHH}^{SM}$  (right).

The enormous amount of data expected during the HL-LHC era will open the door to precision multi-dimensional differential  $t\bar{t}$  cross section measurements that can be used as input to fits for parton distribution functions. The extended forward coverage that will be available with the upgraded ATLAS and CMS detectors will allow for fine-binned measurements in regions of phase space that were not previously accessible. These combined effects will result in unprecedented reductions on the uncertainties of the gluon parton distribution functions once the  $t\bar{t}$  data is incorporated in the fit, as can be seen in Fig. 31.



**Fig. 30:** Projections on the uncertainties for the measurements of the  $W$  boson mass (top left), top quark mass (top right) and indirect constrains from fits to the electroweak precision observables compared with their measured values in the  $W$  boson mass vs. top quark mass plane (bottom left) and vs. the effective weak mixing angle (bottom right).



**Fig. 31:** Projections on the reduction of the relative gluon parton distribution function uncertainties after incorporating double-differential cross section  $t\bar{t}$  data to the fit.

## 12 Outlook

The LHC recently entered a two-year shutdown. We expect to double the data sample during Run 3, scheduled for 2021, with 10 times more data following in the HL-LHC era, currently scheduled for 2025-2035. The Higgs boson will continue to play a central role in the LHC physics program. The formidable precision era enabled by the HL-LHC will allow us to continue to probe the Standard Model predictions and look for cracks that might indicate the presence of new physics processes, even if their masses are above the LHC reach. Direct searches for beyond the Standard Model phenomena will continue to cover previously unexplored ground, with the collaborations pursuing both model-guided and model-independent searches to make sure all options are covered. Furthermore, new detector capabilities will allow us to search in previously unexplored regions like long-lived particles and very forward processes.

Exciting technical challenges lie ahead. If history is our guide, prior projections will be regularly surpassed by real results once the data is in hand and new techniques are developed, and surprises might be just around the corner. We have only collected 5% of the data we expect from the LHC, and analyzed 1% in most cases. We may not have seen an obvious sign of new physics in the data yet, however, what that implies is that we have to get cleverer and make sure we look in every corner and leave no stone unturned. Fun and exciting times lie ahead of us and there is no better time to join the quest.

## Acknowledgements

I wish to thank the ATLAS and CMS colleagues for their help in the preparation of these lectures. In particular, I would like to acknowledge the following individuals who gave me access to material that was ultimately incorporated into my presentations, some of which also made it to this summary: Jamie Antonelli, Doug Berry, Richard Cavanaugh, Matteo Cremonesi, Albert De Roeck, Sarah Eno, Frank Hartmann, Jim Hirschauer, Markus Klute, Andrey Korytov, Jeremy Mans, Corrinne Mills, Chris Palmer, Justin Pilot, Caterina Vernieri.

## References

- [1] L. Evans and P. Bryant, *JINST* **3** (2008) S08001, doi:[10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).
- [2] ATLAS Collaboration, *JINST* **3** (2008) S08003, doi:[10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [3] CMS Collaboration, *JINST* **3** (2008) S08004, doi:[10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [4] CMS Collaboration, *JINST* **12** (2017) P10003, doi:[10.1088/1748-0221/12/10/P10003](https://doi.org/10.1088/1748-0221/12/10/P10003).
- [5] L. Asquith *et al.*, *Rev. Mod. Phys.* **91** (2019) 045003, doi:[10.1103/RevModPhys.91.045003](https://doi.org/10.1103/RevModPhys.91.045003), [arXiv:1803.06991](https://arxiv.org/abs/1803.06991) [hep-ex].
- [6] Z. Nagy, *Phys. Rev. Lett.* **88** (2002) 122003, doi:[10.1103/PhysRevLett.88.122003](https://doi.org/10.1103/PhysRevLett.88.122003), [arXiv:hep-ph/0110315](https://arxiv.org/abs/hep-ph/0110315) [hep-ph].
- [7] Z. Nagy, *Phys. Rev.* **D68** (2003) 094002, doi:[10.1103/PhysRevD.68.094002](https://doi.org/10.1103/PhysRevD.68.094002), [arXiv:hep-ph/0307268](https://arxiv.org/abs/hep-ph/0307268) [hep-ph].
- [8] H.-L. Lai *et al.*, *Phys. Rev.* **D82** (2010) 074024, doi:[10.1103/PhysRevD.82.074024](https://doi.org/10.1103/PhysRevD.82.074024), [arXiv:1007.2241](https://arxiv.org/abs/1007.2241) [hep-ph].
- [9] ATLAS Collaboration, *JHEP* **09** (2017) 020, doi:[10.1007/JHEP09\(2017\)020](https://doi.org/10.1007/JHEP09(2017)020).
- [10] CMS Collaboration, *JHEP* **03** (2017) 156, doi:[10.1007/JHEP03\(2017\)156](https://doi.org/10.1007/JHEP03(2017)156).
- [11] ATLAS Collaboration, *JHEP* **05** (2018) 195, doi:[10.1007/JHEP05\(2018\)195](https://doi.org/10.1007/JHEP05(2018)195).
- [12] CMS collaboration, *Eur. Phys. J.* **C78** (2018) 789, doi:[10.1140/epjc/s10052-018-6242-x](https://doi.org/10.1140/epjc/s10052-018-6242-x).
- [13] CMS Collaboration, *Phys. Rev. Lett.* **112** (2014) 191802, doi:[10.1103/PhysRevLett.112.191802](https://doi.org/10.1103/PhysRevLett.112.191802).
- [14] ATLAS Collaboration, *Eur. Phys. J.* **C78** (2018) 110, doi:[10.1140/epjc/s10052-017-5475-4](https://doi.org/10.1140/epjc/s10052-017-5475-4).
- [15] CDF Collaboration, *Phys. Rev. Lett.* **74** (1995) 2626, doi:[10.1103/PhysRevLett.74.2626](https://doi.org/10.1103/PhysRevLett.74.2626), [arXiv:hep-ex/9503002](https://arxiv.org/abs/hep-ex/9503002).

- [16] D0 Collaboration, *Phys. Rev., Lett.* **74** (1995) 2632, doi:10.1103/PhysRevLett.74.2632, arXiv:hep-ex/9503003.
- [17] D0 Collaboration, *Phys. Rev. Lett.* **103** (2009) 092001, doi:10.1103/PhysRevLett.103.092001, arXiv: 0903.0850 [hep-ex].
- [18] CDF Collaboration, *Phys. Rev. Lett.* **103** (2009) 092002, doi:10.1103/PhysRevLett.103.092002, arXiv: 0903.0885 [hep-ex].
- [19] LHCTopWG–LHC Top Physics Working Group , <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWG>.
- [20] M. Czakon *et al.*, *Phys. Rev.* **D98** (2018) 014003, doi:10.1103/PhysRevD.98.014003.
- [21] CDF Collaboration, *Phys. Rev.* **D83** (2011) 112003, doi:10.1103/PhysRevD.83.112003, arXiv:1101.0034 [hep-ex].
- [22] D0 Collaboration, *Phys. Rev.* **D84** (2011) 112005, doi:10.1103/PhysRevD.84.112005, arXiv:1107.4995 [hep-ex].
- [23] S. Frixione and B.R. Webber, *JHEP* **06** (2002) 029, doi:10.1088/1126-6708/2002/06/029, arXiv:hep-ph/0204244 [hep-ph]; S. Frixione, P. Nason, and B.R. Webber, *JHEP* **08** (2003) 007, doi:10.1088/1126-6708/2003/08/007, arXiv:hep-ph/0305252.
- [24] CDF and D0 Collaborations, *Phys. Rev. Lett.* **120** (2018) 042001, doi:10.1103/PhysRevLett.120.042001.
- [25] J.A. Aguilar-Saavedra *et al.*, *Rev. Mod. Phys.* **87** (2015) 421, doi:10.1103/RevModPhys.87.421, arXiv:1406.1798 [hep-ph].
- [26] ATLAS and CMS Collaborations, *JHEP* **04** (2018) 033, doi:10.1007/JHEP04(2018)033.
- [27] M. Czakon *et al.*, *JHEP* **10** (2017) 186, doi:10.1007/JHEP10(2017)186; M. Czakon, D. Heymes and A. Mitov, *JHEP* **04** (2017) 071, doi:10.1007/JHEP04(2017)071; M. Czakon *et al.*, *JHEP* **05** (2016) 034, doi:10.1007/JHEP05(2016)034.
- [28] J.A. Aguilar-Saavedra and M. Pérez-Victoria, *Phys. Rev.* **D84** (2011) 115013, doi:10.1103/PhysRevD.84.115013, arXiv:1105.4606 [hep-ph]; *JHEP* **09** (2011) 097, doi:10.1007/JHEP09(2011)097, arXiv:1107.0841 [hep-ph].
- [29] A. Czarnecki, J.G. Körner, and J.H. Piclum, *Phys. Rev.* **D81** (2010) 111503, doi:10.1103/PhysRevD.81.111503, arXiv:1005.2625 [hep-ph].
- [30] ATLAS Collaboration, *Phys. Rev.* **D93** (2016) 032009, doi:10.1103/PhysRevD.93.032009.
- [31] CMS Collaboration, *JHEP* **04** (2019) 031, doi:10.1007/JHEP04(2019)031.
- [32] ATLAS Collaboration, *Phys. Lett.* **B716** (2012) 1, doi:10.1016/j.physletb.2012.08.020.
- [33] CMS Collaboration, *Phys. Lett.* **B716** (2012) 30, doi:10.1016/j.physletb.2012.08.021.
- [34] ALEPH Collaboration *et al.*, *Phys. Lett.* **B565** (2003) 61, doi:10.1016/S0370-2693(03)00614-2.
- [35] CDF and D0 Collaborations, *Phys. Rev. Lett.* **109** (2012) 071804, doi:10.1103/PhysRevLett.109.071804, arXiv:1207.6436 [hep-ex].
- [36] ATLAS Collaboration, *Phys. Lett.* **B786** (2018) 59, doi:10.1016/j.physletb.2018.09.013.
- [37] CMS Collaboration, *Phys. Rev. Lett.* **121** (2018) 121801, doi:10.1103/PhysRevLett.121.121801.
- [38] ATLAS Collaboration, *Phys. Rev. Lett.* **122** (2019) 151801, doi:10.1103/PhysRevLett.122.151801.
- [39] H. Davoudiasl, H.-S. Lee, and W.J. Marciano, *Phys. Rev.* **D85** (2012) 115019, doi:10.1103/PhysRevD.85.115019, arXiv:1203.2947 [hep-ph].
- [40] H. Davoudiasl *et al.*, *Phys. Rev.* **D88** (2013) 015022, doi:10.1103/PhysRevD.88.015022, arXiv:1304.4935 [hep-ph].
- [41] ATLAS Collaboration, *Eur. Phys. J.* **C77** (2017) 490, doi:10.1140/epjc/s10052-017-5004-5.
- [42] CMS Collaboration, *JHEP* **05** (2018) 127, doi:10.1007/JHEP05(2018)127.
- [43] A. Boveia *et al.*, *Phys. Dark Univ.* **27** (2020) 100365, doi:10.1016/j.dark.2019.100365.

- [44] ATLAS Collaboration, *JHEP* **05** (2019) 142, [doi:10.1007/JHEP05\(2019\)142](https://doi.org/10.1007/JHEP05(2019)142).
- [45] CMS Collaboration, *Eur. Phys. J.* **C79** (2019) 280, [doi:10.1140/epjc/s10052-019-6730-7](https://doi.org/10.1140/epjc/s10052-019-6730-7).
- [46] ATLAS Collaboration, ATLAS liquid argon calorimeter Phase-I upgrade : Technical design report, [CERN-LHCC-2013-017](https://cds.cern.ch/record/2711113); Technical design report for the Phase-I upgrade of the ATLAS TDAQ system, [CERN-LHCC-2013-018](https://cds.cern.ch/record/2711113).
- [47] CMS Collaboration, Technical proposal for the upgrade of the CMS detector through 2020, [CERN-LHCC-2011-006](https://cds.cern.ch/record/1411113).
- [48] ATLAS Collaboration, Technical design report for the ATLAS inner tracker strip detector, [CERN-LHCC-2017-005](https://cds.cern.ch/record/2211113); Technical design report for the Phase-II upgrade of the ATLAS muon spectrometer, [CERN-LHCC-2017-017](https://cds.cern.ch/record/2211113); ATLAS liquid argon calorimeter Phase-II upgrade : Technical design report, [CERN-LHCC-2017-018](https://cds.cern.ch/record/2211113); Technical design report for the Phase-II upgrade of the ATLAS tile calorimeter, [CERN-LHCC-2017-019](https://cds.cern.ch/record/2211113); Technical design report for the Phase-II upgrade of the ATLAS TDAQ system, [CERN-LHCC-2017-020](https://cds.cern.ch/record/2211113).
- [49] CMS Collaboration, Technical proposal for the Phase-II upgrade of the CMS detector, [CERN-LHCC-2015-010](https://cds.cern.ch/record/1411113); The Phase-2 upgrade of the CMS tracker, [CERN-LHCC-2017-009](https://cds.cern.ch/record/2211113); The Phase-2 upgrade of the CMS muon detectors, [CERN-LHCC-2017-012](https://cds.cern.ch/record/2211113); The Phase-2 upgrade of the CMS DAQ interim technical design report, [CERN-LHCC-2017-014](https://cds.cern.ch/record/2211113); The Phase-2 upgrade of the CMS endcap calorimeter, [CERN-LHCC-2017-023](https://cds.cern.ch/record/2211113); Technical proposal for a MIP timing detector in the CMS experiment Phase-2 upgrade, [CERN-LHCC-2017-027](https://cds.cern.ch/record/2211113).
- [50] A. Dainese (ed.) *et al.*, The physics potential of HL-LHC, Executive summary of the findings of the Workshop on The Physics of HL-LHC, and Perspectives on HE-LHC, <https://twiki.cern.ch/twiki/pub/LHCPhysics/HLHELHCWorkshop/report.pdf>.
- [51] P. Azzi (ed.) *et al.*, Report from Working Group 1 on the Physics of the HL-LHC and Perspectives at the HE-LHC: Standard Model Physics at the HL-LHC and HE-LHC, CERN-2019-007, p. 1, [doi:10.23731/CYRM-2019-007.1](https://cds.cern.ch/record/2211113).
- [52] M. Cepeda (ed.) *et al.*, Report from Working Group 2 on the Physics of the HL-LHC and Perspectives at the HE-LHC: Higgs Physics at the HL-LHC and HE-LHC, CERN-2019-007, p. 221, [doi:10.23731/CYRM-2019-007.221](https://cds.cern.ch/record/2211113).



## Experimental facilities in Latin America

*C. O. Dib*

*CCTVal and Physics Department, Federico Santa María Technical University, Valparaíso, Chile*

### Abstract

These lecture notes briefly describe the current and planned experimental facilities for high-energy physics research in Latin America. The list is not exhaustive nor the descriptions are complete, but I tried to select some of the most representative facilities and large international experiments at the time. Given that particle physics today is tightly related to cosmology and astrophysics, and South America is where most of the major astronomical observatories in the world are located, I start by listing some of the representative observatories. Then we move on to the main accelerator facilities, which are few and mainly for applications, but the core of particle physics research infrastructure in Latin America is in astroparticle physics, which is where I take most of the time to describe. The name and location of each infrastructure are listed, the scientific goals and some introductory description of the detecting techniques. Specific details are not included, which can be found in the literature.

### Keywords

Latin America; experimental facilities; accelerators; detectors; telescopes; cosmology; particle physics; dark matter; dark energy.

## 1 Introduction: a short historical view of experiments in particle physics

One can historically classify the experiments in particle physics in three types according to a mixed criterion of size, technique: (a) those built on a table top, (b) those that detect cosmic rays, which were originally done at high altitude either on top of mountains or aboard of flying balloons, and (c) those that accelerate subatomic particles to make them collide at high energies. This order is also a progression in time, for the simple reason of the increasing level of difficulty: cosmic ray experiments require detectors, but accelerator experiments require, obviously, to build accelerators in addition to detectors, and this task has been a major technological achievement on its own. Now, the observation of cosmic rays is an interesting subject on its own, and so it has continued developing in parallel to experiments with accelerators and colliders. Moreover, due to the ever increasing size and cost of accelerator facilities, there has been an increasing interest in cosmic ray observation facilities in the last decades, a fact that is particularly noticeable in Latin America.

The discovery of the electron by J.J. Thomson in 1897 can be considered as the beginning of subatomic physics, and the gold foil experiment of H. Geiger and E. Mardsen, led by E. Rutherford (former disciple of J.J. Thomson) on the discovery of the atomic nucleus, as the beginning of experimental particle physics. Indeed, the experiment of bombarding a thin gold foil with alpha particles has the essence of all subsequent particle physics experiments: to observe what comes out of collisions of subatomic particles. These experiments and several others at that time are part of the class of experiments done on a table top. Within this class is also the discovery of the neutron by James Chadwick in 1932, himself a disciple of Rutherford.

Concerning experiments done by observing cosmic rays, one must add that the transition from the table top on a laboratory is not sharp: some cosmic ray (CR) experiments can be done on top of a table as

well! The main point here is to stress the origin of the particles that are detected: they do not come from a piece of material or an apparatus in a laboratory, but from outer space. The discovery of cosmic rays is attributed to Victor Hess who, in a remarkable experiment in 1912, showed that there was a radiation that increased with altitude, by climbing with instruments in a free balloon flight up to 5 300 m. He even ruled out the Sun as the source of the radiation by flying during a solar eclipse, noticing that the measurement did not change during the eclipse itself.

A series of experiments with cosmic rays followed, where several of the subatomic particles we know today were discovered, in particular the positron, the muon, the pions and kaons. Without disregarding the value of all these experiments, we want to mention specifically the work of C. Powell, G. Occhialini and C. Lattes in 1947, where they were able to resolve a pending issue at that time. About twelve years earlier, Hideki Yukawa had predicted a “meson”, a particle that would mediate the force between protons and neutrons, with a mass around one or two hundred MeV. Such a particle would be expected to interact strongly and be absorbed by nuclei. A particle with such a mass was found but in some experiments it was shown not to be absorbed by nuclei. The work of Powell, Occhialini and Lattes, using photographic emulsions in experiments on top of Pic du Midi in the Pyrenees, and then on mount Chacaltaya, Bolivia, at an altitude of 5 240 m, was able to show that there were actually two particles with mass in the range predicted by Yukawa: a strongly interacting one that fitted Yukawa’s prediction, and a slightly lighter and weakly interacting particle that did not. These are the pion and the muon, respectively. We single out this experiment for two reasons: first because it was done in Latin America, and second because one of the key experimentalists was the then young Brazilian physicist Cesare Lattes. The leader of the team, Cecil Powell, received the Nobel Prize for this work in 1950, Giuseppe Occhialini was honored in recent years with the naming of the BeppoSAX Satellite for X-ray Astronomy (“Beppo” was Occhialini’s nickname), and Cesare Lattes’s name is being honored with a proposed ground array observatory of cosmic gamma rays that is mentioned further in this lecture.

Our third class of experiments is comprised by those that use particle accelerators. The accelerated particles are most commonly electrons or protons and their antiparticles, but there are also accelerators for heavier ions. The advantage of using particle accelerators instead of cosmic rays is that the collisions are more controlled: we know precisely what particles collide, with what energies, where they collide and how often they collide. In these controlled conditions, the detectors can be placed right around the interaction point and can be much more specialized in their detecting capabilities. Several types of accelerators have been invented since the 1920’s, which go from electrostatic accelerators such as the Van de Graaff and the Cockcroft-Walton generators to the most modern and large scale accelerators with varying fields: the linear accelerators (linacs) and synchrotrons. Moreover, two types of collisions can be devised: a single beam of accelerated particles hitting a fixed piece of material (fixed target collisions) or two beams colliding against each other. The latter can reach much higher centre-of-momentum (CM) energies. The largest linear accelerator in existence is the 3 km long Stanford Linear Accelerator in the USA, and the largest synchrotron was LEP, an electron positron collider with almost 27 km circumference at CERN, which was dismantled in order to install in the same tunnel the currently operating Large Hadron Collider, a proton (and heavy ion) collider, to date the highest energy accelerator, achieving 6.5 TeV per proton, and thus 13 TeV of CM energy in the collisions. It is in high energy particle accelerators where most of the known subatomic particles and antiparticles have been discovered, including the elementary tau lepton, muon-neutrino, tau-neutrino, the top quark and finally the Higgs boson.

With the success of particle accelerators, cosmic ray experiments became less prominent for many decades. However, with the cost and complexity of building ever larger accelerators, there has been again a growing attention to experiments with cosmic rays. Moreover, with a continuous improvement in detector, telescope and satellite technology, major breakthroughs in astrophysics and cosmology have been achieved in the last decades, which have turned again our attention to the heavens. Examples of this are the discovery of gamma ray bursts (GRB) in far away galaxies, which are the most powerful electromagnetic events known in the Cosmos (in a few seconds a GRB releases the energy that our Sun will release



in its entire 10-billion year lifetime), the measurement of the cosmic microwave background (CMB) spectrum by a succession of experiments starting with the COBE satellite in 1992, and the discovery of the accelerated expansion of the Universe.

Here we will present some of the current and planned experimental facilities in Latin America, within the context of particle physics. Our list tries to be representative, but not exhaustive. We will start by briefly mentioning the few existing or planned accelerator facilities in Latin America, although they are mainly oriented to applications than to basic research in particle physics. As we will see, most of the Particle research facilities in Latin America are astroparticle observatories. Moreover, since most of the optical and radio astronomical observatories in the world today (and increasingly so in the near future) are in South America, We will include a short description of some of the most representative astronomical observatories operating, or planned for the near future. Then we will move on to describe the main astroparticle physics facilities and experiments in Latin America starting from those that are operating today and continuing with those that are planned, concluding with the initiative where I am most involved, which is the proposal of an international underground laboratory inside a road tunnel under the Andes mountains between Chile and Argentina. This laboratory will be one of the deepest in the world, the first in South America and, unless another initiative catches up quickly, the first of its kind in the Southern Hemisphere.

## 2 Accelerator facilities in Latin America

Here we will mention two existing facilities, which are Tandem accelerators for ions, and one facility under planning or construction, which is a major Synchrotron Radiation source.

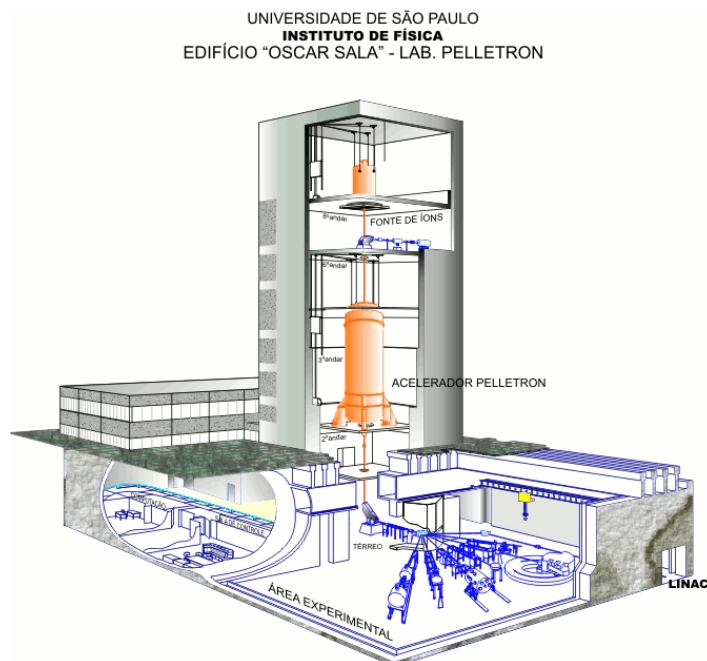
### 2.1 The USP Pelletron

The USP Pelletron is a tandem accelerator existing at the University of São Paulo (USP) since 1972 [1]. The general layout of the facility is shown in Fig. 1(top).

A tandem accelerator is an electrostatic accelerator for ions, derived from simpler types such as Van de Graaf accelerators. In a tandem accelerator there are two stages of acceleration, aligned one after the other. A high positive voltage (let us call it, without much imagination,  $V$ ) is located in the mid point along the acceleration trajectory. Negative ions (with charge  $-1$  in units of the fundamental charge  $e$ ) are produced externally in an ion source, where their mass is selected. These ions are injected into the accelerator and are attracted by the positive terminal at the midpoint, where they reach a kinetic energy  $eV$ . There, the ions cross a thin sheet of material (e.g. carbon), where some or all the electrons of the ion are removed by the collisions, resulting in positive ions with charge  $Ne$  ( $N$  being the number of electrons removed—the ions' atomic number  $Z$  or less). In the second stage, these positive ions are repelled by the central positive terminal, being again accelerated to the end of the machine, gaining an additional kinetic energy  $NeV$ . In total the ions emerge from the accelerator with a kinetic energy  $(N + 1)eV$ .

The name “pelletron” has to do with the method for the charge transport to the positive terminal, which is basically a current of electrically charged metal pellets bound to each other by plastic insulator material.

The USP pelletron has a 8 Megavolt positive terminal, so ions can reach several tens of MeV of energy. The ions are used for different research and application purposes, such as nuclear physics, crystal structure and properties, material analysis, the development of high energy instrumentation, among many other applications.



**Fig. 1:** Top: Drawing of the USP Pelletron (credit: Physics Department, University of São Paulo, Brazil). Bottom: TANDAR at Centro Atómico Constituyentes, Buenos Aires, Argentina. Bottom left: view of the acceleration tower and the  $SF_6$  tanks. Bottom right: insert view of the accelerating column.

## 2.2 TANDAR

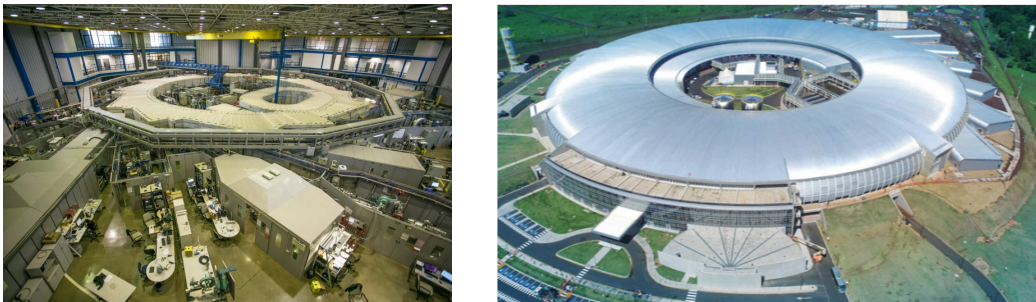
TANDAR is another tandem accelerator, operating since 1985, located at the *Centro Atómico Constituyentes*, Buenos Aires, Argentina, a national scientific laboratory that belongs to CNEA (the Argentinian National Atomic Energy Commission) [2]. The accelerator, similar to the USP pelletron, is a 34.84 m long vertical column inside 73 m tall tower, with a central terminal at 20 MV, as shown in Fig. 1 (bottom). The high voltage terminal is also charged using the pelletron technique. The ion beam runs inside a vacuum tube at  $10^{-8}$  mbar, inside a cylindrical tank with  $SF_6$  (sulphur hexafluoride) at 10 atm to provide dielectric insulation and avoid discharges.

TANDAR is used for advanced education in science and nuclear technology in the country, basic research in nuclear physics and condensed matter physics, applications to material science, biology, environmental studies and many other applications.

### 2.3 LNLS

LNLS (Brazilian Synchrotron Light Laboratory) is a centre that hosts two synchrotron light sources: UVX, operating since 1997 [3], and SIRIUS, a new and much larger synchrotron radiation apparatus currently under construction and planned to start operations in 2020 [4], both shown in Fig. 2.

UVX, a second generation synchrotron radiation source, is a 1.37 GeV electron storage ring of 29.7 m average circumference that provides synchrotron radiation. The injection system comprises a 120 MeV linac and a 500 MeV booster synchrotron. The storage ring provides 17 synchrotron radiation beamlines (experimental stations), where experiments in microscopic analysis techniques using infrared, ultraviolet and X-ray radiation are performed. Most of the beams are in the X-ray range (1 to 30 keV), two beams in the soft X-ray range (100 to 1500 eV), one in the UV range (3 to 330 eV) and one in the IR range ( $70$  to  $300\text{ mm}^{-1}$ ).



**Fig. 2:** Synchrotron light source facilities at LNLS, Campinas, Brazil; left: UVX; right: SIRIUS. Credit: LNLS.

SIRIUS is one of the few 4th generation synchrotron radiation sources in the world. Still under construction, it is planned to start operations in 2020. It is a 3 GeV electron synchrotron, 165 m in diameter. It is a sizable improvement over UVX, not only with twice the energy in the electron beam, but also with an emittance (electron beam divergence) about 360 times smaller, resulting in a much brighter radiation beam. Due to the higher energy and brightness it will allow the study of dense materials at depths up to a few centimetres. Also due to the extremely focused radiation beam, it will be a great improvement over experiments in nano- and biotechnology.

### 3 A brief account on astronomy in Latin America

While astronomy is a different field than high-energy physics with facilities that are different in many respects, we have two reasons to include here a brief description of the astronomical observatories in Latin America.

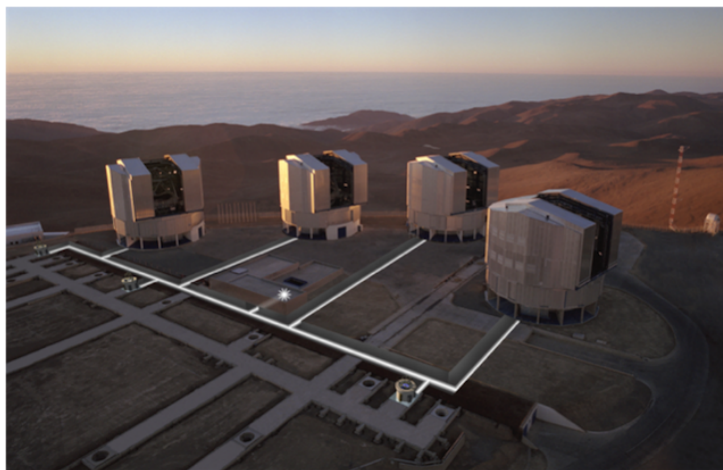
First, as the two fields advance they have more and more points in common, not only in technology but also in research interests. It is clear that the observation of the Cosmos today implies not only detection of light, be it IR, visible, UV or even X-ray, but also very high energy gamma rays, cosmic rays and neutrinos; even the unraveling of the mystery of dark matter is a subject of interest to both fields.

Second, a large proportion of the main astronomical observatories of the world are located in Latin America. In particular, due to the optimal conditions for astronomical observation provided by the desert skies in northern Chile, more than 70% of the light-catching surface of telescopes in the world will be in the Chilean deserts in the near future.

We cannot do justice here to the whole history or set of astronomical observatories in Latin America, so we will limit ourselves to describe just a few representative current facilities and some of those planned or in construction to start operating in the near future.

### 3.1 VLT — the Very Large Telescope

VLT is operated by ESO (European Southern Observatory), and is located in Cerro Paranal, northern Chile, at 2 635 m.a.s.l., coordinates  $24^{\circ}37'38''\text{S}$ ,  $70^{\circ}24'15''\text{W}$ . It is about 100 km south of the city of Antofagasta [5]. It is a set of four large 8.2 m diameter telescopes (called Antu, Kueyen, Melipal and Yepun—words in Mapuche language for astronomical objects), and four movable auxiliary telescopes 1.8 m diameter. See Fig. 3 for a photograph showing the general layout, with a schematic overlay illustrating the interferometry light path.



**Fig. 3:** The Very Large Telescope (VLT), showing the four large telescopes. Three of the four (much smaller) auxiliary telescopes are superimposed on the picture. White lines show the paths of the light beams for interferometry. The asterisk shows the site of the interferometric lab, part underground. Credit: ESO.

VLT is the most modern and the most productive ground-based optical telescope facility in the world today. It operates at visible and IR light. Each telescope can operate independently reaching an angular resolution of 0.05 arc-sec, or all four combined reaching 0.02 arc-sec. It uses interferometry and adaptive optics to overcome diffraction effects in the atmosphere. VLT has been the first telescope to get an image of an exoplanet, has tracked stars around the supermassive black hole (SMBH) at the centre of our galaxy, and has seen the afterglow of the furthest known gamma ray burst (GRB) to date.

### 3.2 ALMA — Atacama Large Millimeter Array

ALMA is the world's largest observatory in the millimetre wavelength (range 9.6 mm to 0.3 mm). It is located on the Chajnantor plateau, in the Atacama desert in Chile, at an altitude of 5 059 m.a.s.l. [6] The very dry desert site at high altitude is necessary to avoid the high absorption of the microwave by water in the atmosphere. Its coordinates are  $23^{\circ}01'09''\text{S}$ ,  $67^{\circ}45'12''\text{W}$ .

ALMA is an interferometric array of 66 movable radio telescopes or *antennas*. The main array (Fig. 4, left) has 50 antennas of 12 m diameter arranged in specific layouts at distances from 150 m up to 16 km, simulating a giant telescope. While each antenna gives an angular resolution of about 20 arcseconds, the giant array working with interferometry as a single device gives an angular resolution higher than the Hubble Space Telescope. Four additional 12 m antennas and twelve 7 m antennas form the Atacama Compact Array (ACA). The different antenna configurations allow the study of both the general structure of astronomic sources as well as its minute details. The 100 ton antennas can be moved by a special transporter that places them on concrete pads with millimetre precision (see Fig. 4, right).

ALMA, the most expensive ground-based observatory to date, is a partnership of ESO (Europe), NSF (USA) and NINS (Japan), in collaboration with the Republic of Chile, and funded by ESO, NSF,



**Fig. 4:** The Atacama Large Millimeter Array. Left: view of the array. Right: telescope being transported. Credit: ALMA.

NRC (Canada), NSC (Taiwan) and KASI (South Korea). ALMA can study star-forming regions seeing through interstellar dust, chemical compounds of the stellar medium, disks and structures around stars, distant galaxies, dynamics of black holes, and many other astronomical phenomena not accessible to any other observatory.

### 3.3 DSA 3— Deep Space Antenna 3

The Deep Space Antenna 3 (Malargüe Station) shown in Fig. 5 (left), is part of the European Space Tracking Network (ESTRACK) for the European Space Agency (ESA). It is a 35 m diameter radio antenna, located 40 km south of Malargüe, Argentina, at an altitude of 1 550 m.a.s.l. Its coordinates are  $35^{\circ}46'34''\text{S}$ ,  $69^{\circ}23'54''\text{W}$ .

DSA3 is a 35 m diameter parabolic radio antenna, working in the range of 8 GHz to 32 GHz. It has two sister stations for very long baseline interference, one in Spain and the other in Australia. DSA3 provides support for communication and tracking with several deep space probes and other scientific probes such as XMM-Newton Rosetta, Herschel, Solar Orbiter, Gaia, Mars Express and Planck. It is also used in radio astronomy to study gamma ray sources, radio galaxies, AGNs, nebula chemistry, and other astronomical radio sources.



**Fig. 5:** Left: DSA3 in Malargüe, Argentina. Right: LMT Alfonso Serrano, Sierra Negra mountain, Mexico.

### 3.4 LMT — Large Millimeter Telescope

Its full name in Spanish is Gran Telescopio Milimétrico Alfonso Serrano [7] (see Fig. 5, right). It is the world's largest single-aperture telescope in its frequency range, built for observing millimetre

wavelengths from 0.85 to 4 mm. It is located at 4 600 m.a.s.l. on top of the Sierra Negra mountain near Puebla, Mexico. The location is  $18^{\circ}59'06''\text{N}$ ,  $97^{\circ}18'53''\text{W}$ .

LMT is a bent Cassegrain optical system with a 50 m-diameter reflecting primary surface (M1) formed by 180 segments, distributed in five concentric rings, a 2.6 m diameter reflecting secondary surface (M2), and a reflecting tertiary surface (M3) almost flat, elliptical with a 1.6m major axis.

Millimetre wavelength allows observation through the interstellar dust and of relatively cold objects that emit mainly at millimetre wavelengths. Among the objects of interest are: comets, planets, protoplanetary discs, evolved stars, star-forming regions and galaxies, molecular clouds, active galactic nuclei (AGNs), high-redshift galaxies, clusters of galaxies and the cosmic microwave background.

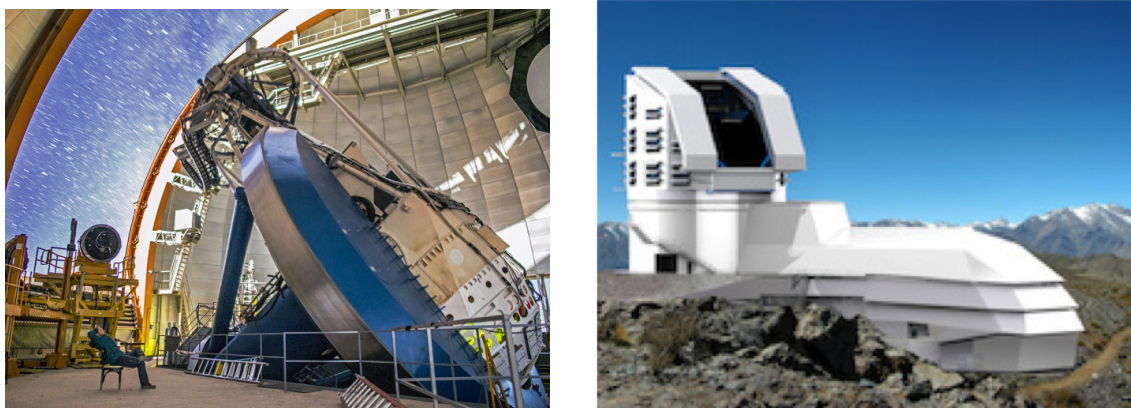
LMT is also part of the Event Horizon array of telescopes around the globe, dedicated to obtain images of black holes. ALMA is also part of the Event Horizon.

### 3.5 DES — Dark Energy Survey

The Dark Energy Survey (DES) [8] is an international collaboration dedicated to map several hundreds of millions of galaxies and the patterns of the cosmos, in order to reveal the character of the so called dark energy that seems cause the accelerated expansion of the Universe. It aims at measuring with high precision the history of cosmic expansion. The survey mapped 5 000 sq. degrees of the southern skies, through five optical filters to obtain additional spectral information of each galaxy. It also focused on smaller patches of the sky to look for thousands of supernovae and transient sources.

The DES Collaboration built a highly sensitive 570 megapixel digital camera, called DECam, that was installed on the Victor Blanco 4 metre telescope at Cerro Tololo Inter-American Observatory (CTIO) [9], 84 km southeast of the city of La Serena, Chile. It is located on top of cerro Tololo, a hill at 2 200 m.a.s.l., with coordinates  $30^{\circ}10'11''\text{S}$ ,  $70^{\circ}48'23''\text{W}$ , as shown in Fig. 6 (left).

DES took data for over six years, from 2013 to 2019.



**Fig. 6:** Left: Dark Energy Survey: Victor Blanco telescope at Cerro Tololo, with the camera DECam installed (credit: DES). Right: Artist rendition of the LSST telescope and building for operations and maintenance at Cerro Pachón, Chile (credit: LSST).

### 3.6 LSST — Large Synoptic Survey Telescope

The Large Synoptic Survey Telescope [10], Fig. 6 (right), is currently under construction at Cerro Pachón, at an altitude of 2 662 m.a.s.l. and 20 km away from Cerro Tololo and part of the CTIO Observatory [9]. The site is at  $30^{\circ}14'40.7''\text{S}$ ,  $70^{\circ}44'58''\text{W}$ .

LSST has a wide field 8.4 metre telescope formed by a three-mirror and three-lens optical assembly. A single exposure has a field of view of 3.5 degrees, which is about 40 times the solid angle of the moon. Its camera of 3.2 gigapixels, is the largest digital camera ever built, 3 m  $\times$  1.65 m, weighing 2.8 ton. LSST also includes a data management system, processing and storing 20 terabytes of data per night, whose ultimate deliverable is the fully reduced data.

Aiming at transient phenomena in the cosmos, LSST will map the entire visible sky with extremely high resolution every 3 nights, storing 45 terabytes of data, building the deepest and widest image of the Universe. This survey will be done over a period of 10 years.

LSST is designed to study primarily four science issues: explore the transient sky in the optical range, understand the nature of dark matter and dark energy, study the formation of the Milky Way, and study the remote solar system, including survey of possible dangerous asteroids.

### 3.7 LLAMA — Large Latin American Millimeter Array

LLAMA [11] is a joint proposal of Argentina and Brazil to install a 12 m single radio telescope, similar to those of ALMA, in Alto Chorrillos, northern Argentina, coordinates 24°11'31"S, 66°28'29"W, at an altitude of 4 820 m.a.s.l.

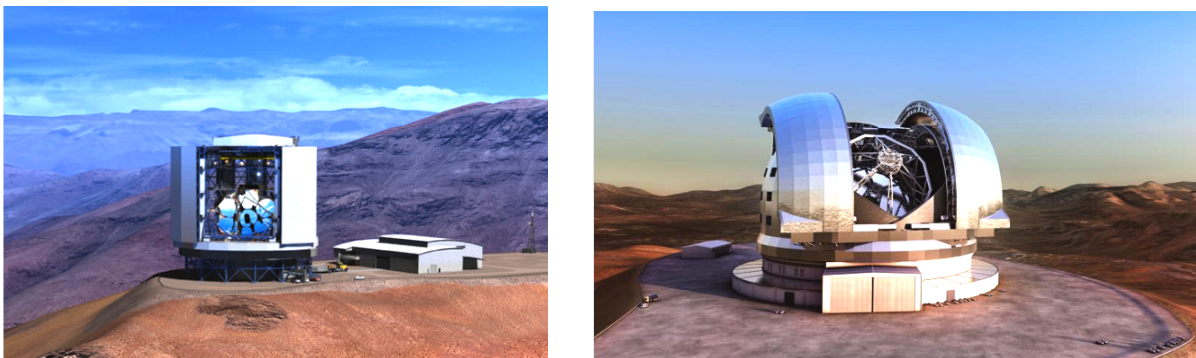
LLAMA will detect in the 35 to 1 000 GHz frequency band, with an angular resolution of 8 arcsec at 900 GHz to 3 arcmin at 35 GHz. The telescope is also planned to be one of the first in a series of antennas of a very long baseline interferometry (VLBI) network in Latin America. In this sense, working with ALMA will increase 10 times the angular resolution of the latter, to about 1 millisecond of arc.

LLAMA aims at studying the solar atmosphere, exo planetary systems, molecules and other features of the intergalactic medium, and the distortion of space around black holes.

### 3.8 GMT — Giant Magellan Telescope

The Giant Magellan Telescope [12] is a planned optical telescope composed of seven 8.4 m mirrors that form a 24.5 m diameter array, with a total of 368 m<sup>2</sup> of light-catching surface. GMT is planned to see first light in 2029, and is rendered in Fig. 7 (left).

GMT will be located at Las Campanas Observatory [13], coordinates 29°0'52.6"S, 70°41'33.4"W, at an altitude of 2 550 m.a.s.l. GMT will use adaptive optics, reaching a resolving power 10 times higher than the Hubble Space Telescope. Operating in the visible to near IR spectrum (320 nm to 25 000 nm), it will study extra-solar planets, search for their chemical composition and possible evidence of life, galaxy formation and evolution, evidence of dark matter and dark energy, and the evolution of the Universe.



**Fig. 7:** Renderings of the future giant telescopes in northern Chile. Left: the Giant Magellan Telescope, GMT at Las Campanas Observatory. Right: the european Extremely Large Telescope, ELT, at Cerro Armazones.

### 3.9 ELT — Extremely Large Telescope

The Extremely Large Telescope (ELT) [14] of the European Southern Observatory (ESO) is under construction at Cerro Armazones, in the northern desert of Chile, 140 km south of the city of Antofagasta, and close to Cerro Paranal where VLT is located. The site is at an altitude of 3 046 m.a.s.l., with coordinates  $24^{\circ}35'36''\text{S}$ ,  $70^{\circ}11'50''\text{W}$ . A rendering of this future telescope is shown in Fig. 7 (right).

ELT is a reflecting telescope with a 39.3 m segmented primary mirror and a 4.2 m secondary mirror. It will have adaptive optics with 8 laser guide star units to correct for atmospheric distortion. It will have the largest light gathering area ever built, 256 times the Hubble Space Telescope and should provide images 16 times sharper than the Hubble. It is expected to see first light in 2025. ELT will provide detailed studies of extra solar planets, detect water and organic molecules in extra solar protoplanetary disks, see the first galaxies of the Universe, study their formation and evolution, study supermassive black holes, study the nature of dark matter and dark energy and the evolution of the Universe.

## 4 Current astroparticle physics facilities in Latin America

After this brief description of some of the Astronomy facilities in the continent, we now describe some of the main astroparticle physics facilities operating in Latin America: the cosmic ray observatories Pierre Auger and LAGO, and HAWC, the high energy gamma ray observatory. Our descriptions will be general, focusing on the infrastructure and science goals. More details on cosmic ray or gamma ray physics can be found in the literature [15].

### 4.1 The Pierre Auger Observatory

The Pierre Auger Observatory [16] is currently the largest cosmic ray detector array in the Southern Hemisphere and the world, aiming at the detection of the highest energy cosmic rays. It combines water Cherenkov detectors and fluorescence telescopes, and grew from previous ground arrays in the north, such as CASA-MIA [17] and HiRes [18].

Auger is located near the town of Malargüe, Argentina, approximate coordinates  $35^{\circ}28'00''\text{S}$ ,  $69^{\circ}18'41''\text{W}$ , at altitudes 1 330 to 1 620 m.a.s.l., covering an area of about  $3\,000\text{ km}^2$  with a surface array of 1 600 water Cherenkov detector (WCD) separated by about 1.5 km from one another, and 27 fluorescence telescopes concentrated in four sites. An overview is shown in Fig. 8.

Each water Cherenkov detector (Fig. 9, left) is a plastic tank with  $12\text{ m}^3$  of ultra pure water inside a Tyvek light diffusing bag, with three 9 inch photomultiplier tubes (PMTs) on the upper part of the tank looking down. Each tank is a self-supported unit with solar panel and batteries, with GPS time, 40 MHz analog to digital converter and radio communication with the control centre. The PMTs detect the Cherenkov light emitted in the water when a relativistic charged particle (muons or electrons) from the secondary cosmic shower enters the tank.

Each fluorescence telescope (Fig. 9, right) is a  $14\text{ m}^2$  parabolic mirror with a camera at its focal plane formed by 440 PMTs. Each set of fluorescence telescopes cover a field of view (FoV) of  $30^{\circ}\times 180^{\circ}$ . The telescopes catch the fluorescence light emitted by the nitrogen gas in the atmosphere which is excited by the pass of the secondary shower of particles created by an incoming cosmic ray. However this light is very dim (about 1 visible photon per metre per electron or positron in the shower), so that the fluorescence telescopes have a low duty cycle of about 10% only (they operate only during moonless nights).

#### *Some words on cosmic rays*

Cosmic rays are actually relativistic charged and massive particles, mainly protons and heavier nuclei, that fly around the galaxy and the whole Universe, that constantly hit the atmosphere. For energies below



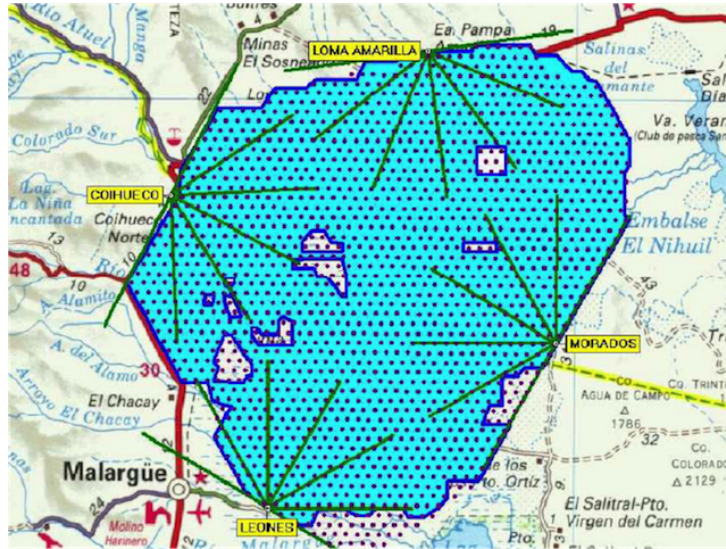


Fig. 8: Map of Auger, showing the extension of the surface array and the fluorescence telescope sites.

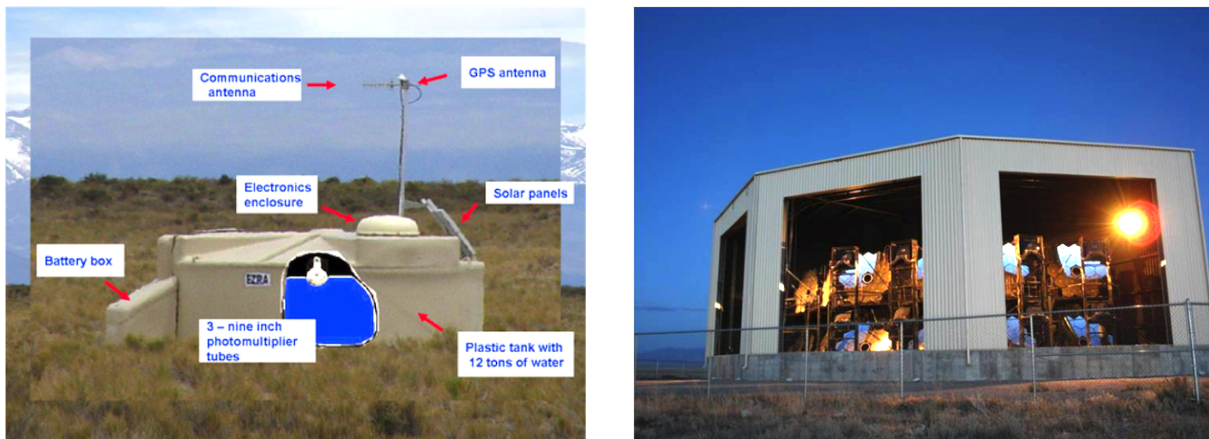
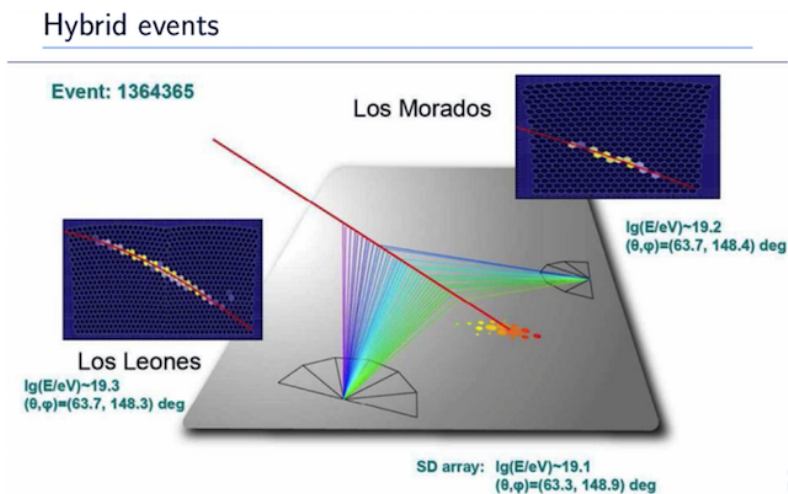


Fig. 9: Left: Auger water Cherenkov detector. Right: Auger fluorescence telescope site.

10 GeV they are mostly of solar origin, and are easily deflected by the solar magnetic field. They have the highest flux, but they are absorbed in the upper atmosphere, so they can only be detected by satellites or in high altitude balloons. For higher energies the flux decreases sharply (about a factor  $1/1000$  per decade of energy). For energies above 100 TeV, the flux is so low that it is difficult to detect them with the small area of detectors in balloons or satellites. However, when they enter the atmosphere they produce *extensive air showers* (EAS) of secondary particles. These secondary particles are more hadrons, mainly pions. Charged pions then decay in muons, electrons and neutrinos, while neutral pions decay quickly into photons. Electrons, positrons and photons keep splitting by bremsstrahlung and pair production until the energy per particle goes below some threshold (near 80 MeV): this is the *electromagnetic* component of the shower. The muons on the other hand, are the remains of the hadronic component of the shower. This extensive shower can even reach the surface of the earth. The higher the energy of the primary particle, the larger the multiplicity of secondary particles, the lower in the atmosphere the shower reaches before it starts getting absorbed, and the wider the spread of the shower at ground level. The point along the shower where it reaches maximal multiplicity is usually called (without much creativity)  $X_{MAX}$ . The multiplicity, spread and  $X_{MAX}$  are the important quantities to measure in order to estimate the energy and composition of the primary ray.

One last issue we should mention on cosmic rays is the so called GZK cutoff [19] of the spectrum at ultra high energies. UHECR cannot be trapped by the galactic magnetic field, so they must be of extragalactic origin. However, for energies above  $10^{20}$  eV, Greisen Zatsepin and Kuzmin conjectured that cosmic rays would degrade their energy by producing pions due to collisions with the photons of the cosmic microwave background at the  $\Delta$  resonance, and consequently cosmic rays of higher energies cannot originate from sources beyond 150 megaparsecs. Since there are not many possible sources in such a small neighbourhood (!), the spectrum should show a cutoff at those energies.



**Fig. 10:** A hybrid event detected by Auger: two telescope sites reconstruct the shower path in the atmosphere and the WCD's detect the hit on the ground. The timing on different WCD's also reconstruct the incoming direction of the CR. (credit: X. Bertou).

After that digression on cosmic rays, let us go back to Auger. Auger aims at the highest energies (the ultra high energy cosmic rays or UHECR), above  $10^{18}$  eV, where the flux is very low (the flux of the highest UHECR are about 1 CR per  $\text{km}^2$  per century). That is why the WCD tanks cover such a large area of  $3\,000\text{ km}^2$ . On the other hand, the shower spread on the ground of the UHECR is large, so the WCD can be quite far apart (near 1.5 km). For this same reason, Auger is insensitive to  $E_{CR} \ll 10^{18}$  eV. With this in mind, we can understand the design of Auger: the WCD tanks will detect the spread and shape of the shower on the ground, while the fluorescence telescopes can get a picture of the development of the shower as it goes down the atmosphere, and have a better estimate of  $X_{MAX}$ . In events with surface water tanks hits only, or fluorescence detectors only, it is possible to obtain an estimate of the energy and composition of the primary, but having hybrid events with signals at the water tanks as well as more than one fluorescence telescope, the resolution is much improved. An example of such a hybrid event is shown in Fig. 10.

Auger has obtained many novel scientific results, such as: UHECR are nuclei (H to Fe), no gammas, no neutrinos; at the highest energies, they seem to be composed more of heavier elements; there may not be a GZK cutoff; there seem to be no excess of UHECR from the galactic centre.

An upgrade of Auger is planned, where scintillator plates will be placed on top of each WCD and a muon detector underground for better separation of electromagnetic vs. muon components.

## 4.2 LAGO — Latin American Giant Observatory

LAGO, the Latin American Giant Observatory [20] (formerly called Large Aperture Gamma ray Observatory) is a set of surface WCDs distributed all along the continent, in order to have an ultra long baseline array of cosmic ray detectors. The tanks are located one per site, each one at different altitude along the

continent (from sea level to about 5 000 m.a.s.l.). The sites so far are Mexico, Guatemala, Venezuela, Colombia, Brazil, Ecuador, Bolivia, Perú, Argentina and recently Chile.

The design of the water Cherenkov tanks aims at being simple, inexpensive and reliable. Commercial tanks with a detection area from  $1.5 \text{ m}^2$  to  $10 \text{ m}^2$  are preferable, with an inner coating of Tyvek for light diffusion, and they must be filled with purified water. The PMT and the electronic design are available for new sites to join in.

The LAGO network is designed to measure the temporal evolution of cosmic radiation at ground level. Its basic research focuses on three areas: high energy phenomena, space weather and atmospheric radiation at ground level. As such, it studies the high energy component of GRB at the very high altitude sites. It also monitors cosmic radiation at the continental scale, observing the effect of solar activity on cosmic rays. These science goals are complemented by two academic goals, which are to train students in astroparticle and high-energy physics techniques and to support the development of astroparticle physics in Latin America [21].

### 4.3 HAWC — High Altitude Water Cherenkov

HAWC, the High Altitude Water Cherenkov gamma ray observatory [22] is an *air shower array*, that is, an array of water tanks at high altitude, designed to detect the shower of secondary particles created in the atmosphere, in this case of interest for high-energy gamma rays. HAWC is located on the side of Sierra Negra mountain, near Puebla, Mexico (not far from the Alfonso Serrano Telescope, see Section 3.4), at high altitude (4 100 m.a.s.l.). Its coordinates are  $18^\circ 59' 41'' \text{N}$ ,  $97^\circ 18' 28'' \text{W}$ .

HAWC aims at the detection of *very high energy* (VHE) gamma rays (GR), from 1 TeV to 100 TeV. These are the highest GR expected to exist.

The array is composed of 300 steel tanks filled with ultrapure water, each tank 7.3 m in diameter and 4.5 m tall, with four PMT's at the bottom of each tank looking up. Unlike the Auger water tanks where the PMT's detect from the top of the tank the diffuse Cherenkov light, here they detect the direct Cherenkov light from the bottom. The array covers a ground surface of about  $20\,000 \text{ m}^2$ , see Fig. 11.



**Fig. 11:** Views of the HAWC array of water tanks. Left: ground view (J. Goodman, 2016). Right: satellite view; at the centre is the Counting House with the electronics and computing; on the right is the HAWC Utility Building with the water filtration plant (Google Earth, 2016).

HAWC is a design improved over MILAGRO [23] (a humorous name for a serious scientific experiment), a previous detector of GR showers at Los Alamos, New Mexico, USA. MILAGRO was a  $80 \text{ m} \times 60 \text{ m}$ , 8 m deep pond at an altitude of 2 530 m.a.s.l. HAWC's higher altitude and water tank isolation makes it 15 times more sensitive to VHEGR than MILAGRO. Nevertheless, MILAGRO proved to be a successful experiment by detecting the first TeV GR from the galactic plane, obtaining a map of

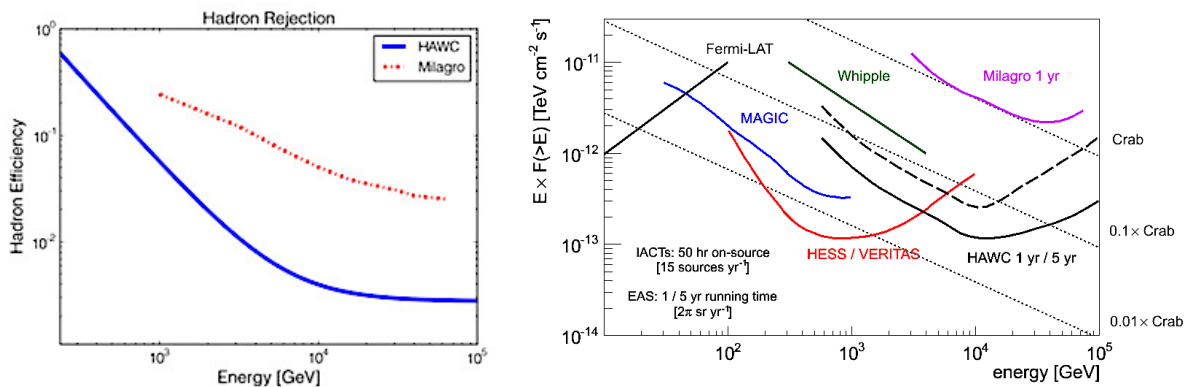
diffuse galactic TeV GR, discovering a TeV GR emission from Cygnus, as well as TeV GR from the Crab Nebula and AGN's.

### Some words on VHE gamma rays

VHE gamma rays, which are photons with energies above 10 GeV up to 100 TeV, produce extensive air showers in the atmosphere as cosmic rays do, but only (or mainly) electromagnetic (electrons, positrons and photons), while cosmic ray showers also have the hadronic component that produces many muons. The electromagnetic showers tend to be more regularly spread around the shower axis, while CR showers, with their hadronic components, are more irregularly scattered and tend to have several lumps away from the axis. Also GR showers do not reach so deep down in the atmosphere, so ground arrays have to be set at quite high altitudes. Another important distinction is that the flux of VHE cosmic rays is about  $10^3$  times more intense than that of gamma rays, so a ground array aimed at observing GR must have a good technique to separate them from the large background of cosmic rays.

HAWC discriminates the GR from the CR background by the pattern of the shower on the ground. The collaboration developed a very clear and user friendly website, which includes a game to illustrate this method [24]. The CR rejection is better the higher the energy of the shower, because the GR and CR patterns become more and more distinct (see Fig. 12, left). Figure 12 (right) show the sensitivity of HAWC compared to other GR detectors, including MILAGRO. It is clear that HAWC performs better at the highest energies. This is due in part to its large area, high altitude and the large duty cycle and large field of view (FoV) of a surface array in comparison to air Cherenkov telescopes (see next sections); however the latter have a better energy and angular resolution.

HAWC is able to study VHE gammas up to 100 TeV from galactic sources (point sources), detect unrequent events such as extreme galactic accelerators thanks to its large duty cycle and FoV, study extended TeV sources (most sources are extended), diffuse emission from the galactic plane (form interactions of CR with gas and dust), extragalactic CR sources such as AGN's. It can also do multimessenger and multiwavelength searches in collaboration with other observatories (can do early detection of most transient sources due to its large duty cycle and FoV to warn other observatories with better resolution).



**Fig. 12:** Left: Cosmic ray rejection power of HAWC and MILAGRO. Right: HAWC sensitivity, compared to other observatories based on air Cherenkov telescopes, and the Fermi satellite.

## 5 Future astroparticle physics facilities in Latin America

Finally, we move on to the future of the astroparticle research infrastructure in Latin America. From this growing field of research we will describe the CTA gamma ray observatory, the cosmic ray detector ALPACA, the proposed LATTES gamma ray detector, SWGO (previously called SGO) which is an

upgrade of HAWC in the south, and ANDES, the proposed first deep underground laboratory in the Southern Hemisphere.

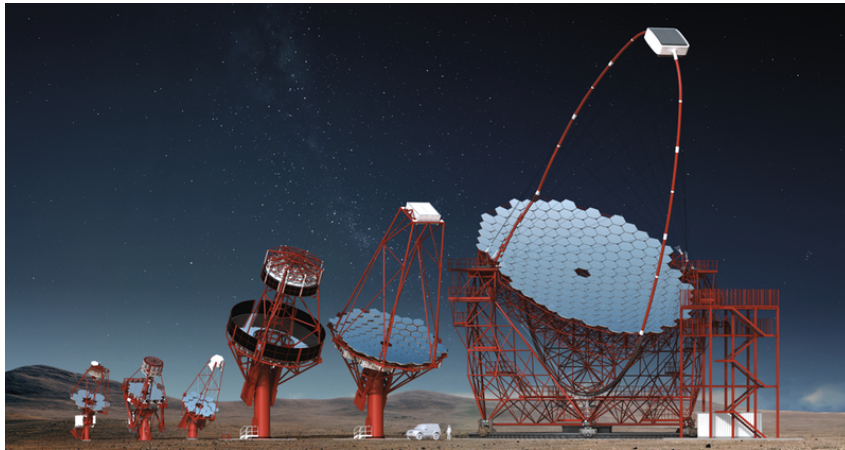
### 5.1 CTA — Cherenkov Telescope Array

CTA, the Cherenkov Telescope Array [25] will be by far the largest VHE gamma ray observatory ever built. It will have one North site in La Palma island (Islas Canarias, Spain, altitude about 2 200 m.a.s.l. and coordinates  $28^{\circ}45'42''\text{N}$ ,  $17^{\circ}53'30''\text{W}$ ) and one South site near Cerro Paranal, in the Atacama Desert, Chile, close to the VLT optical observatory. The south site is at an altitude of 2 100 m.a.s.l. and at the coordinates  $24^{\circ}41'0.34''\text{S}$ ,  $70^{\circ}18'58.84''\text{W}$ .

CTA consists of two large arrays of imaging air Cherenkov telescopes (IACT), a design and technique evolved from previous experiences with the smaller arrays MAGIC (La Palma), HESS (Namibia) and VERITAS (Arizona, USA). An IACT is essentially a spherical or parabolic mirror with a camera on its focal plane, that detects the Cherenkov light (mainly visible and UV) emitted by the relativistic charged particles of a shower as they travel through the atmosphere at speeds greater than light in the medium.

VHE gamma ray showers are produced at high altitudes, of the order of  $10^4$  m.a.s.l. For example a 1 TeV gamma-ray produces a shower that reaches maximum multiplicity at  $\sim 8$  km.a.s.l. (see Fig. 19 below in Section 5.3), in a cone opening of  $\sim 1^{\circ}$ . At 2 000 m.a.s.l. the light pool (ground illuminated area) reaches a radius of  $\sim 120$  m, with a photon density  $\sim 150$  photon/m<sup>2</sup>. For lower energy gamma rays the light pool is of the same size, but much thinner (e.g.  $\sim 10$  photon/m<sup>2</sup> for a 100 GeV GR). However, while GR of lower energy produce a dimmer signal, they arrive more frequently.

As a consequence of these GR features, in order to observe lower energy GR's one needs larger size telescopes (more light-catching are for dimmer signals), but there is no need to cover large ground areas (large GR flux—many GR's per unit area per unit time). On the other hand, to detect higher energy GR's the telescopes can be smaller (intense signals) but one needs many of them to cover a lot of ground (low GR flux).

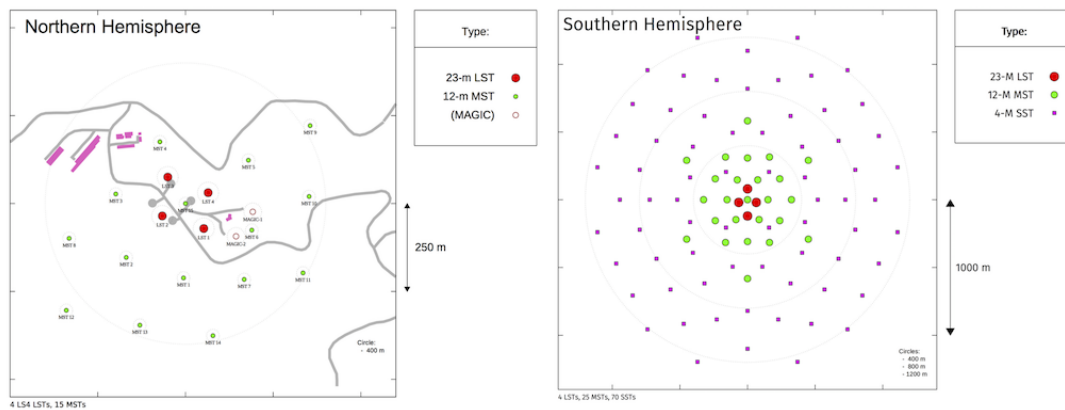


**Fig. 13:** CTA telescopes: three different SST prototype designs, the two MST prototype designs and the LST design (credit: G. Pérez Díaz).

Each IACT has a field of view of a few degrees only. Therefore, to see the shower it has to be looking more or less in the right direction and the shower must be pointing more or less towards the telescope to be inside the light pool. The camera then registers a very short and dim signal that lasts about 10 nanoseconds. Because of that, IACT's are only operated on moonless nights, and therefore they have a duty cycle of less than about 18% or 1,500 h/year (without considering the time loss due to bad

weather).

CTA will have IACT of three different sizes (see Fig. 13): small size telescopes (SST) 4 m diameter to cover the high energy range 1 TeV to 300 TeV, medium size (MST) 12 m diameter to cover the core energy range 150 GeV to 5 TeV, and large size (LST) 23 m diameter to cover the lower energy range 20 to 150 GeV, although they are sensitive up to TeV.

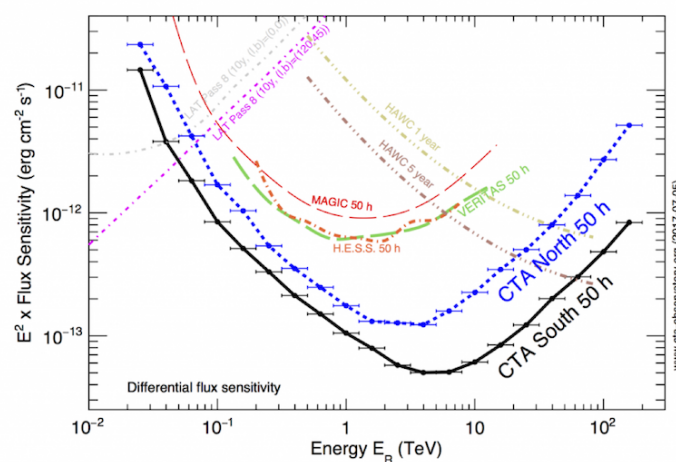


**Fig. 14:** CTA layouts in the North Site (left) and South Site (right).

Figure 14 shows the array layouts on both sites. The North Site in La Palma will be a smaller array, with 4 LST and 15 MST, covering 20 GeV to 20 TeV. The South Site in Chile will have the larger array, as one of the main goals is the galactic centre, which is visible only from the south. The array will consist in 4 LST, 25 MST and 70 SST.

The CTA Observatory headquarters is in Bologna, Italy, while the Science Data Management Centre in Berlin-Zeuthen, Germany.

CTA is expected to cover higher energies and improve on the sensitivity by an order of magnitude with respect to all previous IACT observatories. Figure 15 shows the differential flux sensitivity for CTA and several other observatories. Notice that at the high end of 100 TeV, HAWC gives a comparable sensitivity but for a much longer observation time. Nevertheless, one must take into account that CTA has a much lower duty cycle, 18% or less.

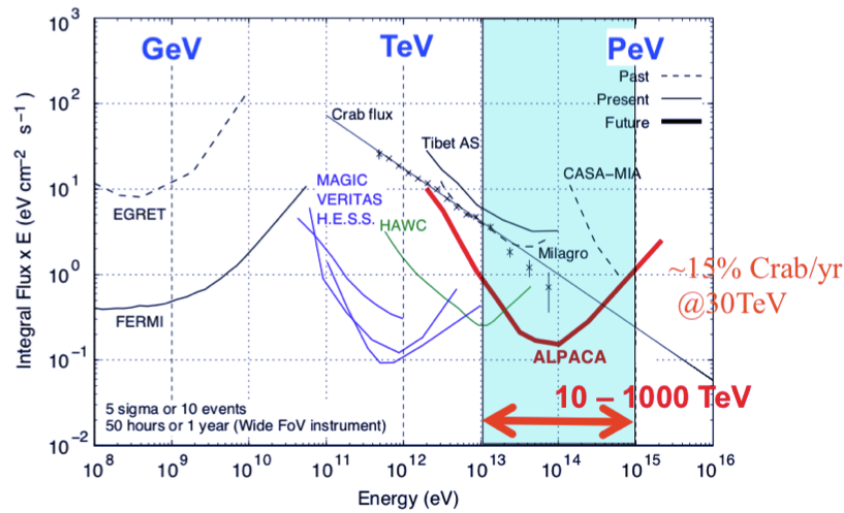


**Fig. 15:** Differential flux as a function of reconstructed GR energy for CTA and other observatories.

The science goals of CTA can be classified in three major groups, that cover astroparticle physics, cosmology and fundamental physics: a) the origin and role of relativistic cosmic particles (where are high energy particles accelerated, what are the CR acceleration mechanisms, what is the role of CR on star formation and galaxy evolution); b) the exploration of extreme environments (what processes occur near neutron stars and black holes, what are the features of relativistic jets, winds and explosions, how are radiation and magnetic fields in cosmic voids); and c) the exploration of frontiers in physics (what is dark matter, are there quantum gravity effects on photon propagation, are there axion-like particles).

## 5.2 ALPACA — Andes Large area Particle detector for Cosmic ray physics and Astronomy

ALPACA [26], a collaboration between Japan and Bolivia, is a proposal to build an air shower array at very high altitude (4740 m.a.s.l.) on the Estuquería plateau near mount Chacaltaya, Bolivia (this site is close to the historical Cosmic Ray Observatory at Mt. Chacaltaya, mentioned in the Introduction). The approximate site coordinates of ALPACA are  $16^{\circ}23'S$ ,  $68^{\circ}08'W$ .



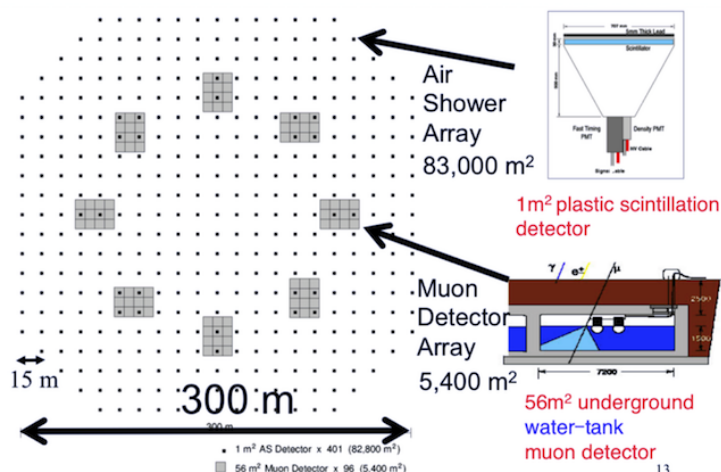
**Fig. 16:** The ALPACA sensitivity to gamma ray point sources.

ALPACA aims at the detection of ultra high energy cosmic rays, or UHECR ( $> 10^{18}$  eV) and very high-energy gamma rays, or VHEGR from 10 to 1000 TeV (indicated in Fig. 16). Figure 17 displays the layout of the ALPACA array, which consists of 401 plastic scintillator detectors  $1 \text{ m}^2$  each, covering a ground surface of  $82000 \text{ m}^2$ , together with 96 water Cherenkov detector (WCD) tanks underground,  $56 \text{ m}^2$  each, in some specific points under the surface array, covering a total area of  $5400 \text{ m}^2$ . The underground WCD's serve to detect the muons in the shower, thus providing a discrimination between cosmic and gamma rays.

The scientific goals of ALPACA can be summarized as the study of: the 10 to 1000 TeV gamma ray sources; galactic and nearby extragalactic sources of UHECR; the CR composition around the “knee” of the spectrum ( $\sim 10^{15}$  eV) in order to understand the CR acceleration mechanisms; the solar coronal magnetic field by detecting the Sun's shadow to CR; the flux from young SNR's (supernova remnants); the cosmic ray anisotropy for energies above TeV in the southern sky; very extended GR sources.

## 5.3 LATTES — Large Array Telescope for Tracking Energy Sources

LATTES, with a clever acronym that honors the famous Brazilian physicist Cesare Lattes, is a proposed surface detector array at very high altitude for gamma ray showers, dedicated to fill the gap in sensitivity

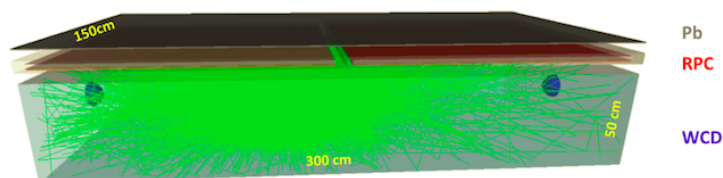


**Fig. 17:** A schematic drawing of the ALPACA air shower array, showing the surface plastic scintillator detectors and the underground water tank muon detectors.

between satellite and VHE ground arrays, i.e. from below 100 GeV up to 100 TeV [27]. It is a collaboration from Brazil, Portugal and Italy. LATTES is proposed to be built on the Chajnantor plateau in northern Chile, near the ALMA astronomical observatory, at an altitude of 5 200 m.a.s.l. and an approximate location  $23^{\circ}01'S$ ,  $67^{\circ}45'W$ .

The LATTES design consists of a dense surface array covering an area of 20 000 m<sup>2</sup>. Each unit is a hybrid detector composed of two resistive plate chambers (RPC)  $1.5 \times 1.5$  m<sup>2</sup> each, covered by a thin lead slab (5.6 mm), all on top of a water Cherenkov detector (WCD) of dimensions  $3 \times 1.5 \times 0.5$  m<sup>3</sup> with two PMT's on the smallest vertical faces, as shown in Fig. 18.

An upgrade to “Full LATTES” is also envisioned by adding an additional array of sparse detectors around the core array of 20 000 m<sup>2</sup>, covering a total area of 100 000 m<sup>2</sup>.



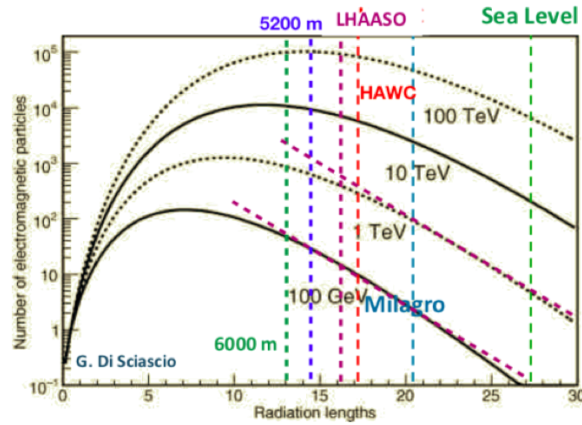
**Fig. 18:** LATTES detector unit, showing the lead plate on top, the two resistive plate chambers and the water Cherenkov detector at the bottom with the two photomultiplier tubes on the left and right walls (credit: B. Tomé, LATTES coll.).

The sensitivity to lower GR is achieved by being able to trigger on the shower secondary photons which are 5 to 7 times more numerous than secondary charged particles. The thin lead slab converts the secondary photons which have stronger correlation with the shower axis, while absorbs the lower energy electrons which have poorer correlation. In turn, the RPC's give good timing and geometry resolution, and the WCD gives a good calorimetric measurement of the shower.

The very high altitude location is necessary for a relatively low energy electromagnetic shower reach the ground. Figure 19 shows how the e.m. shower develops and then it is absorbed as it travels down the atmosphere. For lower energy gamma rays it is clear that a high altitude is essential.

In summary, LATTES will be a surface array designed to explore gamma rays in the 100 GeV region, covering the gap between satellite and other ground based observatories. Due to its wide field





**Fig. 19:** Particle multiplicity in an electromagnetic shower as a function of atmospheric depth, for various primary GR energies. Vertical lines correspond to the altitudes of several GR ground array experiments (credit: G. di Sciascio).

of view and large duty cycle, it will be able to trigger observation of variable sources for other observatories like CTA, survey the southern sky and in particular the galactic centre, and detect galactic and extragalactic transient phenomena.

#### 5.4 SGSO — now SWGO

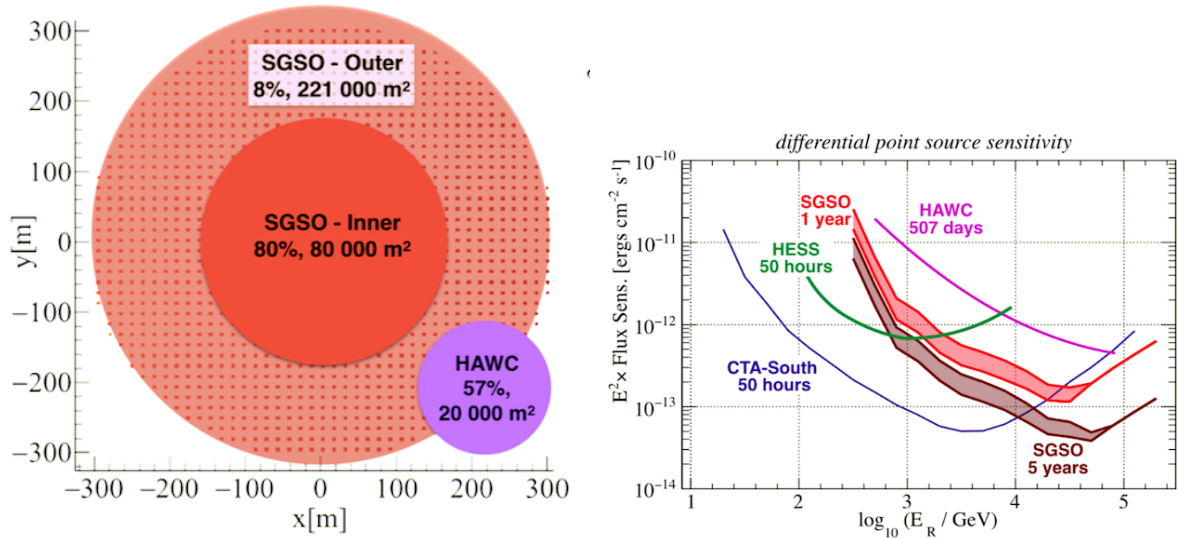
SGSO, the Southern Gamma-Ray Survey Observatory [28, 29], is a proposed large water Cherenkov surface array for VHE gamma rays in South America. SGSO grows from the experience of HAWC and other observatories, with the intention to observe the southern skies. One of the main motivations for having a ground air shower array is the complementarity with air Cherenkov observatories like CTA, by providing a much larger duty cycle and field of view than the latter. This is a major advantage for finding variable sources and transient phenomena. A south site for a ground array is particularly important if one of the goals is the study of the galactic centre, which has a ground view only from the South.

As all surface observatories for gamma ray showers, it needs to be at high altitudes. The site is not defined yet, but possible candidate sites are: Cerro Vecar, Argentina (altitude 4 800 m.a.s.l., latitude  $24^{\circ}\text{S}$ ), which is near the LLAMA site; Chajnantor plateau, Chile (5 300 m.a.s.l.,  $23^{\circ}\text{S}$ ), near ALMA; and Lake Sibinacocha, Cuzco Region, Peru (4 870 m.a.s.l.,  $13^{\circ}\text{S}$ ).

The optimization of the design of SGSO, supported by simulations, is still in progress. A “strawman detector” is considered, with larger size than HAWC, as well as higher altitude and larger *fill factor* (the fraction of the full area covered with detectors). The strawman design, shown in Fig. 20, considers a core array of  $80\,000\text{ m}^2$  with a fill factor of 80%, surrounded by a sparse outer array of  $221\,000\text{ m}^2$  with a fill factor of 8%. These figures can be compared to HAWC, which covers  $20\,000\text{ m}^2$  with a fill factor of 57%.

SGSO main science goals are *i*) unveil galactic and extragalactic CR accelerators (galactic centre, galactic plane, star-forming regions, pulsars, Fermi bubbles); *ii*) monitor VHE GR transients (AGN’s, Blazars, GRB’s, G Waves, HE neutrinos); *iii*) probe new particle physics (dark matter, axion-like particles, Lorentz violation); and *iv*) characterize the CR flux (CR spectrum and composition, flux anisotropy, space weather, heliosphere physics).

Please notice that from the time I gave this lecture to the time I wrote these proceedings, the name of this observatory and collaboration changed from SGSO to SWGO, the *Southern Wide-Field Gamma-Ray Observatory* [29, 30], the name it should be referred to from now on. Thus have we observed another



**Fig. 20:** SGSO strawman design; left: layout with instrumented areas and fill factors; right: differential point source sensitivity vs. reconstructed gamma-ray energy, compared to other GR observatories (from A. Albert *et al.* [28]).

type of transient phenomenon.

## 5.5 ANDES — Agua Negra Deep Experiment Site

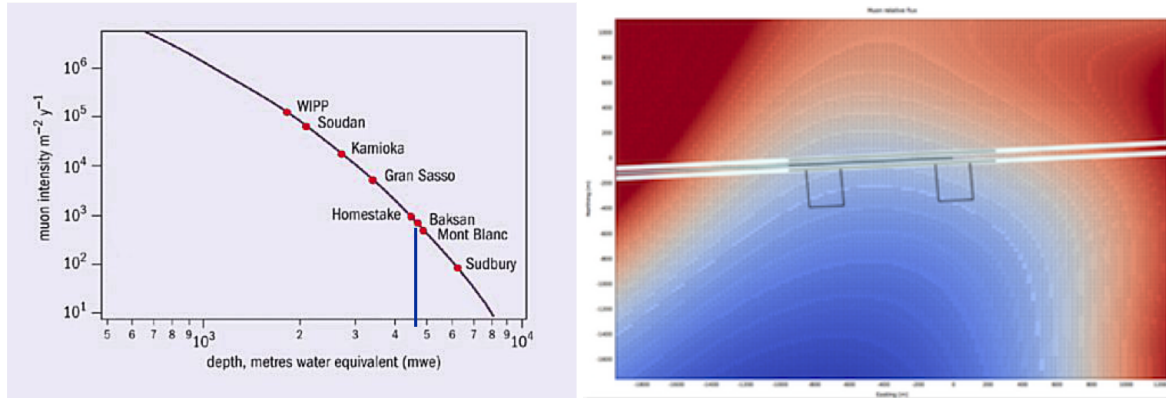
ANDES (Agua Negra Deep Experiment Site) [31] is a proposal between Argentina, Chile, Brazil and Mexico to build a world class deep underground international laboratory under the Andes mountains, at the border between Argentina and Chile. The proposal takes advantage of the future construction of a 14 km long road tunnel, *Tunel Agua Negra* that will cross the Andes between the Region of Coquimbo (Chile) and the Province of San Juan. The nearest major cities are about 250 km away from the tunnel: La Serena in Chile and San Juan in Argentina. La Serena is also an active centre for astronomy in the region, close to the observatories Cerro Tololo, Las Campanas and La Silla. The underground site will be at an altitude about 3 900 m.a.s.l., with coordinates  $33^{\circ}11'34''\text{S}$ ,  $69^{\circ}49'25''\text{W}$ .

ANDES will be the first underground laboratory of its kind in the Southern Hemisphere and among the deepest in the world, with 1750 m of rock overburden in all directions, providing excellent shielding against background radiation from cosmic ray muons, as illustrated in Fig. 21.

ANDES should be managed by an international Consortium formed by country members and possible Associate members. The line of organization of ANDES should follow that of SESAME, the international Synchrotron Laboratory in the Middle East [32]. Membership will be open to countries around the world. The laboratory is designed to host a large experiment in neutrino physics, and other experiments in astroparticle physics for e.g. dark matter searches, as well as nuclear astrophysics, biology, and geophysics.

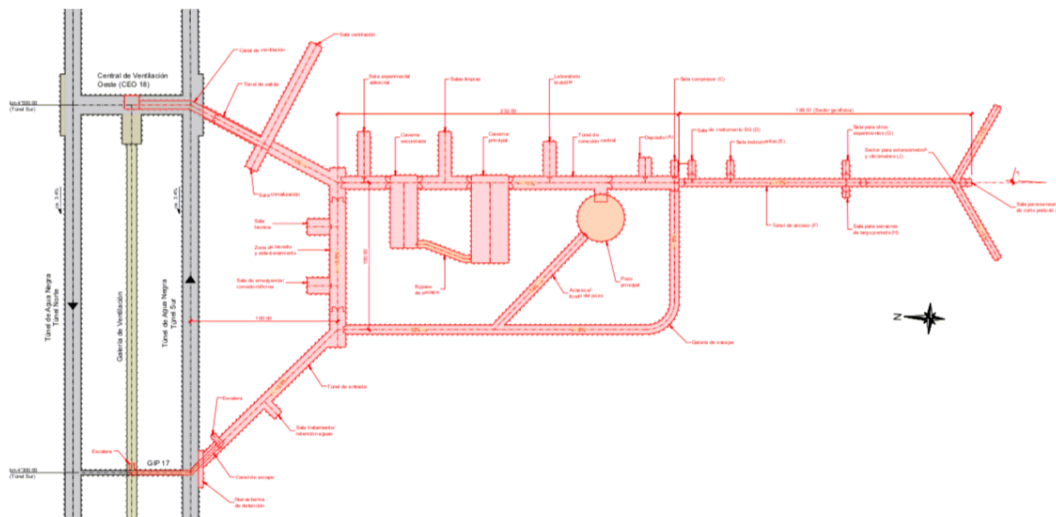
The site of ANDES is special in many ways. Besides being in the Southern Hemisphere, it is in a region with much tectonic activity, and with very little background from nuclear power plants. The quality of the rock at the underground site is not yet known, although samples from 600 m depth were obtained, showing a proportion of andesite and rhyolite with normal level of radioactivity.

The underground site will lie at the deepest point along the tunnel. The tunnel is actually a pair of two-lane 14 km long tunnels separated by about 100 m. The laboratory will be on the south side of the eastbound tunnel, so the access will be about 4 km from the Chilean entrance. The underground



**Fig. 21:** Left: muon flux attenuation for different laboratories; ANDES is shown by the vertical line. Right: muon multidirectional relative flux; blue denotes more attenuation; the left box corresponds to the border between the countries and the right box the position of the ventilation plant; the actual site must be determined by the quality of the rock (credit: X. Bertou)

site will have an entrance and exit for vehicles, and the following caverns: a) one large cavern for large experiments, 42 m long  $\times$  21 m wide  $\times$  23 m high, with a 40 ton curved bridge crane; b) a secondary cavern for several smaller experiments and offices, 30 m long  $\times$  16 m wide  $\times$  14 m high, also with a 40 ton bridge crane; c) main pit for large neutrino experiments, 30 m diameter  $\times$  38 m high, with access from the top and from the bottom. d) insulated room 30 m long  $\times$  10 m wide and 5 m high, designed for nuclear astrophysics experiments. e) Biology room 16 m long  $\times$  9 m wide  $\times$  3.5 m high. f) Two clean rooms 10 m  $\times$  10 m  $\times$  8 m, ISO Class 7. g) Tunnel for Geoscience installations, 200 m long  $\times$  3 m wide  $\times$  3.5 m high. f) Emergency cavern with office space, dining room, meeting room, emergency first aid and data processing and communications centre, 10 m long  $\times$  10 m wide  $\times$  8 m high. h) Technical room, for transformers, compressors and generally noisy installations, and welding workshop. The general layout is shown in Fig. 22.



**Fig. 22:** ANDES underground site layout, showing the different caverns and main pit, with the access galleries connecting to the eastbound tunnel (credit: Lombardi S.A.)

## 6 Final words

I presented a brief description of some of the major current and future facilities in Latin America, knowing that I could not possibly do justice to all of them, and possibly I may have missed some important details, but my goal was to convey a flavour of the opportunities that exist and what can be expected in the near future. Facilities in Latin America have clearly grown in the last decades, and this tendency seems to continue for the next decades. This is clearly so for Astronomy due to the clean skies of the South American desert; in fact, more than half of the optical telescope surface in the world will be in northern Chile in the next decades. Moreover, as astronomy and particle physics have growing common interests as research fields, world level astroparticle physics facilities are also planning to find sites in the continent. Finally, the high-energy physics community in Latin America shows a steady growth for a variety of reasons we may argue, but a clear one is the appearance of large experiments that require worldwide collaborations.

## Acknowledgements

I want to thank the organizers for the wonderful hospitality, as well as the lecturers, discussion leaders and students for the warm and stimulating atmosphere of this School. This work was supported in part by Fondecyt, Chile grant 1170171 and Conicyt, Chile PIA/Basal FB0821.

## References

- [1] Acelerador Pelletron, Laboratório Aberto de Física Nuclear – LAFN, <http://portal.if.usp.br/fnc/pt-br/acelerador-pelletron> (in Portuguese); J.R.B. Oliveira, *J. Phys. Conf. Ser.* **1291** (2019) 012003, doi:10.1088/1742-6596/1291/1/012003.
- [2] Gerencia Investigación y Aplicaciones – Laboratorio Tandara, <http://www.tandar.cnea.gov.ar> (in Spanish); N. Fazzini *et al.*, *Nucl. Instrum. Meth.* **A268** 330, doi:10.1016/0168-9002(88)90529-3.
- [3] UVX Synchrotron Light Source – Brazilian Synchrotron Light Laboratory, <https://www.lnls.cnpem.br/uvx-en/>; A. F. Craievich, *Radiat. Phys. Chem.* **167** (2020) 108253, 10.1016/j.radphyschem.2019.04.003.
- [4] Sirius – Accelerating the Future of Brazilian Science, <https://www.lnls.cnpem.br/sirius-en/>.
- [5] The Very Large Telescope array (VLT) – European Southern Observatory, <https://www.eso.org/public/teles-instr/paranal-observatory/vlt/>; *The VLT White Book*, reprint ed. (ESO, Garching, 1998), [http://www.eso.org/sci/libraries/historicaldocuments/VLT\\_White\\_Book/VLT\\_Whitebook\\_A1b.pdf](http://www.eso.org/sci/libraries/historicaldocuments/VLT_White_Book/VLT_Whitebook_A1b.pdf).
- [6] The Atacama Large Millimeter/submillimeter Array (ALMA), <https://www.almaobservatory.org>.
- [7] The Large Millimeter Telescope Alfonso Serrano, [https://en.wikipedia.org/wiki/Large\\_Millimeter\\_Telescope](https://en.wikipedia.org/wiki/Large_Millimeter_Telescope); D.H. Hughes *et al.*, *Proc. SPIE* **11445** (2020) 1144522, doi:10.1117/12.2561893.
- [8] The Dark Energy Survey (DES), <https://www.darkenergysurvey.org>; H.T. Diehl *et al.*, *The Dark Energy Survey and Operations: Year 6 – The Finale*, FERMILAB-TM-2720-AE, doi:10.2172/1596042.
- [9] Cerro Tololo Inter-American Observatory, <http://www.ctio.noao.edu/noao/>.
- [10] Vera C. Rubin Observatory, <https://www.lsst.org>; S.J. Thomas *et al.*, *Proc. SPIE* **11445** (2020) 114450I, doi:10.1117/12.2561581.

- [11] The Large Latin American Millimeter Array,  
<http://tux.iar.unlp.edu.ar/llama-web/english.html>;  
 G.E. Romero, *Sci. Rev.—from the end of the world* **1** (4) (2020) 34,  
<http://scirevfew.net/index.php/sciencereviews/article/view/10>.
- [12] The Giant Magellan Telescope, <https://www.gmto.org>;  
 J. Fanson *et al.*, *Proc. SPIE* **10700** (2018) 1070012, doi:10.1117/12.2313340.
- [13] The Las Campanas Observatory, <http://www.lco.cl>;  
 F. Di Mille *et al.*, *Proc. SPIE* **10704** (2018) 107041S, doi:10.1117/12.2312825.
- [14] The Extremely Large Telescope (ELT) – European Southern Observatory,  
<https://www.eso.org/public/teles-instr/elt/>;  
 R. Tamai, *Proc. SPIE* **11445** (2020) 114451E, doi:10.1117/12.2562828.
- [15] A. de Angelis and M. Pimenta, *Introduction to particle and astroparticle physics*, 2nd ed. (Springer, Cham, 2018), doi:10.1007/978-3-319-78181-5; C. Grupen, *Astroparticle physics*, 2nd ed. (Springer, Cham, 2020), doi:10.1007/978-3-030-27339-2; G. Sigl, *Astroparticle physics: Theory and phenomenology* (Atlantis Press, Paris, 2017), doi:10.2991/978-94-6239-243-4.
- [16] Pierre Auger Observatory, <https://www.auger.org>;  
 The Pierre Auger Collaboration, *Sci. Rev.—from the end of the world* **1** (4) (2020) 8,  
<http://scirevfew.net/index.php/sciencereviews/article/view/31>.
- [17] R. Ong, *Ultra High Energy Cosmic Ray Research with CASA-MIA*, Pres. at Cronin Fest: Celebration of James Cronin’s 75th Birthday, Chicago, IL, USA, 8–9 Sep. 2006,  
<http://www.astro.ucla.edu/~rene/talks/Cronin-Fest-Ong-Writeup.pdf>.
- [18] The High Resolution Fly’s Eye (HiRes) – University of Utah,  
<http://www.cosmic-ray.org>.
- [19] K. Greisen, *Phys. Rev. Lett* **16** (1966) 748, doi:10.1103/PhysRevLett.16.748;  
 G.T. Zatsepin and V.A. Kuz’min, *JETP Lett.* **4** (1966) 78,  
[http://jetpletters.ru/ps/1624/article\\_24846.shtml](http://jetpletters.ru/ps/1624/article_24846.shtml).
- [20] The Latin American Giant Observatory (LAGO), <http://lagoproject.net>;  
 L. Villaseñor, *Forum Int. Phys. Newsl.* **Spring** (2012) 19,  
<https://engage.aps.org/fip/resources/newsletters/spring-2012>.
- [21] I. Sidelnik *et al.*, *Nucl. Instrum. Meth.* **A876** (2017) 173, doi:10.1016/j.nima.2017.02.069,  
[arXiv:1703.05337](https://arxiv.org/abs/1703.05337) [astro-ph.IM].
- [22] HAWC: the High-Altitude Water Cherenkov Gamma-Ray Observatory,  
<https://www.hawc-observatory.org>;  
 K. Malone, *EPJ Web Conf.* **235** (2020) 07001, doi:10.1051/epjconf/202023507001.
- [23] Experimental particle physics at NYU: MILAGRO,  
<https://physics.nyu.edu/experimentalparticle/milagro.html>;  
 J.A. Goodman, *Nucl. Phys. B, Proc. Suppl.*, **151** (2006) 101,  
 doi:10.1016/j.nuclphysbps.2005.07.018.
- [24] HAWC: the High-Altitude Water Cherenkov Gamma-Ray Observatory: The gamma/hadron separation game, <https://www.hawc-observatory.org/observatory/ghsep.php>.
- [25] Cherenkov Telescope Array Observatory, <https://www.cta-observatory.org>;  
 F. Ferrini and W. Wild, *The Messenger* **180** (2020) 3, doi:10.18727/0722-6691/5194.
- [26] The ALPACA experiment, <https://alpaca-experiment.org>;  
 C. Calle, *J. Phys.: Conf. Ser.* **1468** (2020) 012091, doi:10.1088/1742-6596/1468/1/012091.
- [27] P. Assis *et al.*, *PoS ICRC2017* (2018), 784, doi:10.22323/1.301.0784.
- [28] A. Albert *et al.*, *Science case for a Wide Field-of-View Very-High-Energy Gamma-Ray Observatory in the Southern Hemisphere*, [arXiv:1902.08429](https://arxiv.org/abs/1902.08429) [astro-ph.HE].

- [29] The Southern Wide-field Gamma-ray Observatory (SWGGO), <https://www.swgo.org>;  
U. Barres de Almeida, *Astron. Nachr.* **342** (2021) 431, doi:10.1002/asna.202113946.
- [30] P. Abreu *et al.*, *The Southern Wide-Field Gamma-Ray Observatory (SWGGO): A next-generation ground-based survey instrument for VHE gamma-ray astronomy*, arXiv:1907.07737 [astro-ph.IM].
- [31] ANDES: Towards a unique deep underground laboratory in the Southern Hemisphere,  
<https://www.andeslab.org/>;  
X. Bertou, *PoS EDSU2018* (2018), 027, doi:10.22323/1.335.0027.
- [32] SESAME: Synchrotron-light for experimental science and applications in the Middle East,  
<https://www.sesame.org.jo>;  
H. Schopper, *Riv. Nuovo Cim.* **40** (2017) 199, doi:10.1393/ncr/i2017-10134-8.

**Appendix: Exercises for the Lecture on experimental facilities in Latin America**

Notice: the problems are not self-contained; you may need to look up some formulae somewhere else. I listed some suggested references at the end.

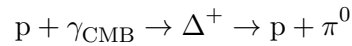
1. The radiation length of electrons in air is  $X_r \simeq 37 \text{ g/cm}^2$ . Considering that the atmospheric pressure at sea level is  $p_0 \simeq 1 \times 10^5 \text{ N/m}^2$ , show that the vertical thickness of the atmosphere is about 30 radiation lengths.
2. (a) Determine the Cherenkov radiation emission angle  $\theta_C$  in the atmosphere at sea level for a charged particle moving with  $\beta \sim 1$ . The refractive index of the atmosphere at sea level is  $n_{\text{at}} \simeq 1.00029$ .  
 (b) Same question but when the medium is liquid water.  
 (c) Determine the threshold (minimum) energy for an electron to emit Cherenkov light in air.  
 (d) Same question but for a muon.
3. Considering that the average galactic magnetic field is a few microgauss, show that cosmic rays (protons) with energies below  $10^{18} \text{ eV}$  should be of galactic origin while those with energies  $10^{19} \text{ eV}$  or above should be extragalactic.
4. What components of galactic cosmic rays can have a larger energy, protons or heavier nuclei?
5. Consider observatories of cosmic rays and gammas based on IACT (imaging atmospheric Cherenkov telescopes) and SDA (surface particle detector arrays). Explain the meaning of the following concepts:
  - (a) Field of view
  - (b) Energy threshold
  - (c) Sensitivity
  - (d) Duty cycle
  - (e) Effective area of detection of a ground array. Explain why it depends on the energy of the CR or gamma ray.
6. (a) Why are IACT's (Cherenkov Telescopes) in general more sensitive to lower energies than surface arrays?  
 (b) Why IACT's have a much shorter duty cycle than surface arrays?  
 (c) Which one of the detector techniques, IACT's or surface arrays, have a larger field of view? Why is that?  
 (d) How can a surface array measure the energy of the primary CR or gamma ray? How is that done with IACT's?
7. In an extensive air shower (EAS), a primary proton produces pions as it collides with the atmospheric nuclei. As it develops, the shower contains many muons and also many electromagnetic cascades as well. Explain why it contains muons and additional e.m. cascades.
8. In an EAS, a muon is produced at an altitude of 10 km. If the muon is able to reach sea level before decaying, determine the minimum energy it should have. Will this muon produce Cherenkov light in the atmosphere?
9. From the graph of the CR spectrum, in the region  $10 \text{ GeV} < E < 1 \text{ PeV}$ , the spectral flux (particles per unit area, time, solid angle and energy) can be fitted to a function:

$$j(E) = A \left( \frac{\text{eV}}{E} \right)^\alpha, \quad \text{with } \alpha \simeq 2.7$$

- (a) By fitting the graph (choosing an adequate point), find the value of  $A$  in  $[1/(\text{m}^2 \cdot \text{s} \cdot \text{sr} \cdot \text{eV})]$ . (sr = steradian)
- (b) Consider a satellite with a flat detector face of  $1 \text{ m}^2$ . Determine the rate of primary cosmic rays with energies above 1 TeV that hit the detector (do not forget that the incidence is from the outside only, and do the solid angle integration correctly).
- (c) What about energies above 100 TeV? Express the result in  $[\text{particles}/(\text{m}^2 \cdot \text{day})]$ .
10. Consider an electromagnetic shower caused by a vertical cosmic gamma ray of 1 TeV. Assume that the shower begins at 20 km.a.s.l. ( $x = 0$ ) and as it develops, all particles in the shower (electrons, positrons and photons) at a given height have the same energy.
- (a) Determine the number of particles in the shower as a function of depth  $x$ . You will need the radiation length  $X_r$  from (1).
- (b) The shower reaches its maximum when the energy of the particles reaches the critical value  $E_C \simeq 80 \text{ MeV}$ . Determine  $x_{\text{MAX}}$ .
- (c)  $E_C$  occurs when the energy loss in electrons and positrons start to be dominated by ionization instead of bremsstrahlung. Find in the literature the expression for the energy loss by ionization and show that  $E_C$  for electrons is indeed around 80 MeV.
11. Consider the Sun, a star  $1.5 \times 10^{11} \text{ m}$  away from Earth. The luminous solar radiation intensity on Earth is about  $1.5 \text{ kW}/\text{m}^2$ .
- (a) Determine the total power emitted by the Sun.
- (b) The net reaction where that power is produced is  $p + p \rightarrow {}^4_2\text{He} + 2e^+ + 2\nu$ , releasing about 26 MeV per reaction. Most of this energy is eventually emitted as photons. However the neutrinos escape, carrying in average about 0.4 MeV each (these are the so-called “pp neutrinos”). There are other reactions in the Sun that produce neutrinos, but these are the most abundant. Determine the flux of neutrinos on Earth, in  $1/\text{cm}^2 \text{ s}$ .
- (c) The solar neutrinos that are detectable by Super KamiokaNDE (SuperK) are not pp neutrinos. They have energies above 5 MeV and they come from the decay of the isotope  ${}^8_5\text{B}$  (boron-8), and they constitute a very small fraction of the total flux that arrive to Earth, about  $5 \times 10^6 \text{ } 1/\text{cm}^2 \text{ s}$ . Try to estimate roughly the number of Solar neutrinos detected by SuperK per day. (You need to look up the size of SuperK and the  $\nu$ -e cross section at those energies).
12. Let us try to recreate a Cherenkov emission from the e.m. shower produced by a 1 TeV primary photon. Assume that the shower is vertical and initiates at 12 km.a.s.l. Also assume for simplicity that the radiation length is a fixed value of 400 m, independent of altitude. An e.m. shower is quite collimated, with a spread not more than  $0.5^\circ$ , while the Cherenkov angle in air is about  $1.5^\circ$ . So clearly the size of the light pool on the ground will be determined mainly by the Cherenkov spread. Now consider that, along the shower, each electron or positron as it travels, emits about 10 Cherenkov photons per metre travelled. Consider this emission only up to  $X_{\text{MAX}}$ , which is where the shower has maximal multiplicity (particles reach the critical energy  $E_C \simeq 80 \text{ MeV}$ ).
- (a) Determine the diameter of the Cherenkov light pool on the ground, which is at an altitude of 2000 m.a.s.l.
- (b) Determine the number of Cherenkov photons in the light pool (here we are disregarding absorption in the lower atmosphere, which is not quite realistic).
- (c) Assuming that the Cherenkov photons are uniformly spread inside the light pool, determine how many photons will be caught by a 12 m diameter IACT that lies completely inside the light pool.
13. How can one learn about the acceleration regions of cosmic rays by measuring gamma rays?
14. Find out about the mechanism called *Fermi Acceleration*.



15. Why is it important to measure the VHE gamma rays that come from the galactic centre?
16. How can one test Lorentz invariance by measuring cosmic gamma rays?
17. Why gamma rays are better to identify the sources of cosmic rays than the cosmic rays themselves?
18. What are “Fermi bubbles”?
19. Concerning the search for the particles that constitute the dark matter of the Universe, explain what is called “direct searches” and “indirect searches”.
20. The GZK cutoff is the upper limit for the energy of CR that can reach the Earth from very far extragalactic distances, due to the energy loss of a CR particle (e.g. a proton) when it collides with a CMB photon and produces a pion at the resonance of the  $\Delta$  baryon, e.g.:



Determine the energy of the GZK cutoff.

### Some references and reading

1. A. de Angelis and M. Pimenta, *Introduction to particle and astroparticle physics*, 2nd ed. (Springer, Cham, 2018), doi:10.1007/978-3-319-78181-5.
2. P.A. Zyla *et al.* [Particle Data Group], *Review of Particle Physics*, *PTEP* **2020** (2020) 083C01, doi:10.1093/ptep/ptaa104.
3. HAWC website: <https://www.hawc-observatory.org>.
4. A. Albert *et al.*, *Science case for a Wide Field-of-View Very-High-Energy Gamma-Ray Observatory in the Southern Hemisphere*, arXiv:1902.08429 [astro-ph.HE].
5. CTA website: <https://www.cta-observatory.org>.
6. B.S. Acharya *et al.* [CTA Consortium], *Science with the Cherenkov Telescope Array* (World Scientific, Singapore, 2019), doi:10.1142/10986.



## Scientific programme <sup>1</sup>

Field theory and the electro-weak Standard Model

*Luis Álvarez-Gaumé, Stony Brook, USA / CERN*

Higgs physics

*Marcela Carena, U. of Chicago/Fermilab, USA*

QCD

*Daniel De Florian, ICAS/UNSAM, Argentina*

Facilities in Latin America

*Claudio Dib, U. Técnica Federico Santa Maria, Chile*

Highlights of LHC Run 2 and future prospects

*Cecilia E. Gerber, U. of Illinois at Chicago, USA*

Special lecture on gravitational waves

*Gabriela González, Louisiana State U., USA*

Practical statistics

*Eilam Gross, Weizmann Institute, Israel*

Evening lecture: the music of physics

*Eilam Gross, Weizmann Institute, Israel*

Neutrino physics

*Pilar Hernández, IFIC, U. Valencia-CSIC, Spain*

Cosmological collider physics

*Juan Martín Maldacena, IAS, Princeton, USA*

Flavour physics and CP violation

*Antonio Pich, IFIC, U. Valencia-CSIC, Spain*

Cosmology

*Rogério Rosenfeld, IFT-UNESP/ICTP-SAIFR, Brazil*

Heavy-ion physics

*María Elena Tejeda-Yeomans, U. de Colima, Mexico*

Physics beyond the Standard Model

*Carlos Wagner, U. of Chicago/Argonne National Laboratory, USA*

Q & A session

*Fabiola Gianotti (CERN)*

---

<sup>1</sup>Slides available at <https://indico.cern.ch/event/763013/timetable/?daysPerRow=5&view=nicecompact>

## Organizing committee

M. Aguilar (CIEMAT, Spain)  
L. Álvarez-Gaumé (Stony Brook, USA and CERN)  
E. Carrera (USFQ, Ecuador)  
M. Cerrada (CIEMAT, Spain)  
C. Dib (UTFSM, Chile)  
M.T. Dova (UNLP, Argentina)  
C. Duhr (CERN)  
J. Ellis (King's College London, UK and CERN)  
N. Ellis (Schools Director, CERN (Chair))  
A. Gago-Medina (PUCP, Peru)  
M. Gandelman (UFRJ, Brazil)  
P. Garcia (CIEMAT, Spain)  
M. Losada (UAN, Colombia)  
M. Mulders (Schools Deputy-Director, CERN)  
K. Ross (Schools Administrator, CERN)  
A. Zepeda (Cinvestav, Mexico)

## Local organizing committee

L. Da Rold (Instituto Balseiro, CONICET)  
M.T. Dova (Universidad Nacional de La Plata, IFLP (UNLP- CONICET))  
G. González-Sprinberg (Universidad de la República, Uruguay)  
G. Otero y Garzón (University of Buenos Aires, CONICET)  
O.A. Sampayo (Universidad Nacional de Mar del Plata, IFIMAR (CONICET))  
N.N. Scoccola (Comisión Nacional de Energía Atómica, CONICET)  
A. Szyrkman Universidad Nacional de La Plata, IFLP (UNLP- CONICET)  
F. Tamarit (FAMAF, Universidad Nacional de Córdoba, IFEG (UNC-CONICET))  
H. Wahlberg (Universidad Nacional de La Plata, IFLP (UNLP- CONICET))

## Lecturers

L. Álvarez-Gaumé (Stony Brook, USA and CERN)  
M. Carena (University of Chicago and Fermilab, USA)  
D. De Florian (ICAS/UNSAM, Argentina)  
C. Dib (Universidad Técnica Federico Santa María, Chile)  
C.E. Gerber (University of Illinois at Chicago, USA)  
G. González (Louisiana State University, USA)  
E. Gross (Weizmann Institute, Israel)  
P. Hernández (IFIC, University of Valencia - CSIC, Spain)  
J.M. Maldacena (IAS, Princeton, USA)  
A. Pich (IFIC, University of Valencia - CSIC, Spain)  
R. Rosenfeld (IFT-UNESP, ICTP-SAIFR, Brazil)  
M.E. Tejeda Yeomans (Universidad de Colima, Mexico)  
C. Wagner (University of Chicago and Argonne National Laboratory, USA)

## **Discussion leaders**

N. Bernal (Universidad Antonio Narino, Colombia)

R. Hernández Pinto (U. Autónoma de Sinaloa, Mexico)

J. Jones Pérez (PUCP, Peru)

N. Neill (Universidad Técnica Federico Santa María, Chile)

J.C. Vasquez Carmona (Universidad Técnica Federico Santa María, Chile)

## Students

Daniel AGUILAR  
Adrian ALVAREZ FERNANDEZ  
Miguel ALVES GALLO PEREIRA  
Paul ASMUSS  
Fernando Augusto ASSUNÇÃO NETO  
Gabriel AUGUSTO DAS NEVES  
Gabriel Artemio AYALA SANCHEZ  
Marco AYALA TORRES  
Téssio B. DE MELO  
Andrés BÓRQUEZ CÁRCAMO  
Ignacio BORSA SANJUAN  
Ana Martina BOTTI  
Rolando CALDERÓN ARDILA  
Julián CALLE  
Juan Carlos CARRASCO MARTINEZ  
Agustín CARULLI  
Tulio CAVIQUIOLI CARDOSO  
Carlos COCHA  
Máximo COPPOLA  
Maria Florencia DANERI  
Gabriel DE LA FUENTE ROSALES  
Ana Maria DE SOUSA SLIVAR  
Dener DE SOUZA LEMOS  
Alejandro DE YTA HERNÁNDEZ  
Manuel DER  
María Roberta DEVESA  
Carlos Eduardo DÍAZ JARAMILLO  
Diego DOMINGUES LOPES  
Daniel ERNANI MARTINS NETO  
Felipe FIGUEROA VILAR DEL VALLE  
Flavia GESUALDI  
Jefferson GON CALVES  
Fernanda GON CALVES ABRANTES  
Emma GONZALEZ HERNÁNDEZ  
Isabel Astrid GOOS  
Matthew Peter HEATH  
Alan Ignacio HERNÁNDEZ JUÁREZ  
Mykola KHANDOGA  
Paulina KNEES  
Federico LAMAGNA  
Moises LEON  
Bibiana Marcela MARÍN OCHOA  
Edyvania MARTINS  
Cristobal MORALES  
Alice MORRIS  
Rafael David NUNEZ PALACIO  
Gonzalo Enrique ORELLANA  
Breno ORZARI  
Santiago Rafael PAREDES SAENZ  
Iván PEDRON  
Andres Daniel PEREZ  
Mario Aldair PEREZ DE LEON  
Brenda QUEZADAS VIVIAN  
Antonio RADIC  
Dario RAMIREZ  
Érison ROCHA  
Rosa María SANDÁ SEOANE  
Julian SCHMOECKEL  
Marina Andrea SCORNAVACCHE  
Elisabetta SPADARO NORELLA  
Dwayne SPITERI  
Birgit Sylvia STAPF  
Omar Giancarlo SUAREZ NAVARRO  
Manuel SZEWC  
Emily TAKEVA  
Natalia TAPIA  
Edward THORPE  
Mariana TOSCANI  
Alison Maria TULLY  
Jorge Luis VALDES ALBUERNES  
Indira Dajhana VERGARA QUISPE  
Yufeng WANG  
Rafaela WIKLICH SOBRINHO  
Lewis WILKINS  
Gabriel Damián ZAPATA

# Posters

Poster title	Presenter
Search for additional Higgs bosons produced in association with b quarks and decaying into two b quarks	ASMUSS, P.
Tests of trigger detectors based on plastic scintillators coupled to silicon photomultipliers	AYALA TORRES, M.
Neutrino masses in a two Higgs doublet model with a U(1) gauge symmetry	B. DE MELO, T.
Parton distributions and fragmentation functions at a future electron-ion collider	BORSA, I.
AMIGA underground muon detector at AugerPrime	BOTTI, A.M.
A first study on improving the spatial resolution of plastic scintillator based muon Telescopes	CALDERÓN ARDILA, R.
LINAC commissioning at SIRIUS: First data and beam dynamics	COCHA, C.
Charged pion masses under strong magnetic fields in the NJL model	COPPOLA, M.
Precision measurements in high energy physics at the ATLAS experiment	DANERI, M.F.
Study of mono-Z production at the LHC as signal of dark matter	DE SOUSA SLIVAR, A.M.
Mixed QCD $\oplus$ QED corrections to Drell Yan production	DER, M.
The fermion mass hierarchy from a non-universal Abelian extension to the Standard Model	DIAZ, C.E.

<b>Poster title</b>	<b>Presenter</b>
Muon deficit in air shower simulations from AGASA data	GESUALDI, F.
Unfolding method for momentum of atmospheric muons	GONZÁLEZ, E.
Modeling of extensive air showers generated by high-energy cosmic rays	GOOS, I.
General conditions to CP violation in neutral gauge couplings $ZZV^*$ ( $V = Z, Z'$ )	HERNÁNDEZ-JUÁREZ, A.I.
Data-driven shower shapes corrections	KHANDOGA, M.
Neutrino physics in supersymmetry	KNEES, P.
Discrimination between $tq$ and $ttH$ Higgs production modes in the $H \rightarrow \gamma\gamma$ decay channel	LÉON M.
Describing high-energy collisions with hydrodynamics	LEMOS, D.S.
Hadronic showers with high-electromagnetic production	MARTINS, E.E.
Search for exotic long-lived particles using displaced jets in the ATLAS calorimeters	MORRIS, A.
Angular asymmetries in the decay $\tau \rightarrow K\pi\nu_\tau$	NEVES, G.A.
CP sensitivity Dalitz plot analysis of the decay $B^0 \rightarrow K_s^0 K^+ K^-$	NUÑEZ, R.D.



Poster title	Presenter
Strategy for a search for supersymmetry in the ATLAS detector	ORELLANA, G.
Search for the dark Higgs in the CMS Experiment	ORZARI, B.
Variable-radius sub-jets in a boosted hh $\rightarrow$ 4b search in ATLAS	PAREDES SAENZ, S.
Supersymmetric dark matter candidates	PEREZ, A.D.
A private SUSY 4HDM with FCNC in the up-sector	PEREZ DE LEON, M.A.
Weak dipole moments of the tau lepton	QUEZADAS VIVIAN, B.
Study of Drell-Yan Z-boson production in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV	RAMIREZ, D; DE LA FUENTE, G; ALBUERNES, J.L.
Phenomenological approaches of diffractive exclusive production of quarkonia in collisions proton-nucleus at LHC	ROCHA, E.
Universality with AMIGA underground muon detector	SCORNAVACCHE, M.
Observation of $B_s^0 \rightarrow J/\psi p\bar{p}$ decays and precision measurements of the $B_s^0$ masses	SPADARO NORELLA, E.
ATLAS VH(bb) analysis - 2 lepton (2L) channel	SPITERI, D
Majorana and pseudo-Dirac neutrinos at the ILC	SUAREZ NAVARRO, O.

<b>Poster title</b>	<b>Presenter</b>
One photon pair production from foil bremsstrahlung at the LUXE strong field experiment	TAPIA, N.
Doubly charmed $B$ decays with LHCb	TULLY, A.
Photons search using dedicated triggers in the Pierre Auger Observatory	VERGARA QUISPE, I.
Search for resonances in the 65 to 110 GeV diphoton invariant mass range using pp collisions collected with the ATLAS detector	WANG, Y.
Study of the phenomenology of the hadronic interactions in the reconstruction of extensive air showers	WIKLICH SOBRINHO, R.
Final taus and initial state polarization signatures from effective interactions of Majorana neutrinos at future $e^+e^-$ colliders	ZAPATA, G.