

Carlos Merino *Editor*

Lectures on Particle Physics, Astrophysics and Cosmology

Proceedings of the Third IDPASC School,
Santiago de Compostela, Spain,
January 21–February 2, 2013

Springer Proceedings in Physics

Volume 161

More information about this series at <http://www.springer.com/series/361>

Carlos Merino
Editor

Lectures on Particle Physics, Astrophysics and Cosmology

Proceedings of the Third IDPASC School,
Santiago de Compostela, Spain,
January 21–February 2, 2013



Springer

Editor

Carlos Merino

Department of Particle Physics

University of Santiago de Compostela (USC)

Santiago de Compostela

Spain

ISSN 0930-8989
Springer Proceedings in Physics
ISBN 978-3-319-12237-3
DOI 10.1007/978-3-319-12238-0

ISSN 1867-4941 (electronic)
ISBN 978-3-319-12238-0 (eBook)

Library of Congress Control Number: 2014957320

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

*Dad, did you know space
shuttles always travel by night?*

*Belén Merino, 3-year-old
Ferrol, Galiza, February 2009*

Participants



Participants at the Third IDPASC School, before the Seminario Maior San Martiño Pinario in Santiago de Compostela, Galiza, Spain, January 22–February 2, 2013 (picture by A.V. Yushkov)

Preface

This volume contains the courses presented at the Third IDPASC School (<http://igfae.usc.es/idpasc/>), held at San Martiño Pinario, Hospedería and Seminario Maior (<http://www.sanmartinpinario.eu>), in the city of Santiago de Compostela, Galiza, Spain, from January 21 to February 2, 2013. This school is the annual common school of the International Doctorate Network in Particle Physics, Astrophysics, and Cosmology, IDPASC (www.idpasc.lip.pt).

IDPASC is an interdisciplinary network formed by more than 20 institutions (universities and research laboratories) from Italy, Slovenia, France, Portugal, Spain, Switzerland, and other non-European countries, as Brazil, and whose international coordinator is Prof. Mário Pimenta of the Instituto Superior Técnico (IST) and Laboratório de Instrumentação e Física Experimental de Partículas (LIP) in Lisbon (Portugal), that has been created with the aim of forming new high-level experts in the fields of Particle Physics, Astrophysics, and Cosmology.

The two former editions of the IDPASC School were held in Sesimbra (Portugal), in December 2010, and in Udine (Italy) in January–February of 2012, respectively, while the Fourth IDPASC School was organized in Braga (Portugal), from January 20 to 28, 2014.

The purpose of the school series is to present before doctorate students from different universities and laboratories in Europe and abroad a large scope of the latest results and current state-of-the-art in the fields of Particle Physics, Astrophysics, and Cosmology, and to further introduce them to the questions now open by the physics potential and experimental challenges connected with the current and future experiments, and, in particular, with the newly available energy ranges.

Following these guidelines, the program of this third edition of the IDPASC School has been planned both by the Academic Council and by the International Committee of the network, whose members ensure every year its adequate and timely formulation, matching the continuous evolution in these fields. The program covers a balanced range of the current activity in these fields worldwide, with courses given by internationally acknowledged physicists, on Basic Features of Hadronic Processes, Quantum Chromodynamics, Physics and Technology of ALICE, LHCb Physics—Parity Violation, The Higgs System in and beyond the

Standard Model, Higgs Searches at the LHC, Theory and Experiment in Cosmic Rays, Numerical Methods and Data Analysis in Particle Physics, Theoretical Cosmology, and AdS/CFT Correspondence. Most of these courses have been complemented by practical and discussion sessions.

The spirit of this series of schools is to favor, not only the chance of young promising physicists to learn the fundamentals, ideas, and methods at the base of hot topics in physics from internationally recognized experts who have contributed to define and conform the topic itself, but also to promote fruitful and informal discussions between young and experience physicists, and among experimentalists and theorists, by giving all of them the opportunity of being together for most of the time during 10 days, while completing a charged but not exhausting program in which many young promising physicists have the first chance to present and discuss their results with other young doctorates and with internationally recognized experts in both experimental and theoretical particle fields.

Apart from the academic and scientific part of the program, the fact of organizing the school in the building of the centuries-old but still acting seminario of San Martiño Pinario has helped in achieving the goal of giving the participants a real approach to the history and culture of the city of Santiago de Compostela, and, more generally, to Galiza as a country in Europe molded by being for more than eight centuries the final point and the goal of the pilgrims way that has given rise and formed the European spirit, through the exchange of knowledge and ideas from all corners in Europe.

The conference dinner at the restaurant of the Hotel-Monumento San Francisco, located at the heart of the old-town in Santiago de Compostela, in which the famous Galician seafood, fish, and meats, cooked in the traditional way but with successful innovative touches, were met by the excellent Galician Albariño white wine and by the reputed red Rioja wine from Northern Spain, was a representation of Galiza, old and new, which combines both tradition and ambition for the future. The banquet ended up by the spelling of the conjure during the burning of queimada (firewater, sugar, and lemon peel), reminiscent of the preromanic times in Galiza, and aimed at chasing away bad spirits and meigas (witches). Let us hope that the drinking degustation of that spelled queimada will prevent Galiza and the whole Europe from giving the needed efforts to support and encourage research and knowledge up.

I want to thank all local organizers for their successful efforts in providing the participants with a comforting atmosphere both for work (with the smooth continuation of the scheduled talks and program), and for getting a true insight into the Galician soul. In particular, the help of María del Carmen Fernández has been fundamental in putting together the financial, informatic, and logistical aspects of the conference. The remarkable work and friendly demeanor by the whole staff of Hospedería San Martiño Pinario has also to be noted, in particular the professional savoir-faire of Adriana Hermida and María Luisa González.

We acknowledge the advice and support of Profs. Mario Pimenta, IST—LIP (Portugal), and Barbara de Lotto, Università di Udine (Italy) to the organization of this school, and we also thank Sandra Dias, LIP (Portugal) and Sandra Muzzin,

Universitá di Udine (Italy) for their collaboration before, during, and after the school.

Finally, we have to thank all participants, both lecturers and students, since they have been the final responsible for the nice environment created during the school and for the high level of the lectures and discussions that have formed the program of this school. We are especially indebted to all professors for their hard and disciplined work in preparing clear and understandable lectures on very vast and technical topics, and in producing the corresponding written contributions to these proceedings by friendly and timely attending to all requests by the editorial committee.

This conference has been organized with the financial support of the European Union through the Lifelong Learning Programme of the Laboratório de Instrumentação e Física Experimental de Partículas (LIP), Lisbon (Portugal), of the Instituto Superior Técnico (IST), Universidade de Lisboa (Portugal), of the Universitá degli Studi di Udine (Italy), and of the Universidade de Santiago de Compostela (USC), Galiza (Spain).

Carlos Merino

Committees and Sponsors, Logos

IDPASC Coordinator

Mário Pimenta, IST—LIP (Portugal)

IDPASC Academic Council

Rolf Heuer, CERN (Switzerland)

Luigi Rolandi, CERN (Switzerland)

Ronald Shellard, CBPF (Brazil)

Frederico Ferrini, EGO (Italy)

Valeria Ferrari, EGO (Italy)

Alberto Ruiz Jimeno, IFCA, SCIC-University of Cantabria (Spain)

Javier Fernández Menéndez, University of Oviedo (Spain)

António Onofre, Universidade do Minho/LIP/MAP-Fis (Portugal)

Miguel Sousa Costa, Universidade do Porto/FCUP/MAP-Fis (Portugal)

Paulo Seara de Sá, Universidade do Algarve (Portugal)

Robertus Potting, Universidade do Algarve/FCT (Portugal)

Rui F. Marques, Universidade de Coimbra/FCTUC/LIP (Portugal)

Orlando Oliveira, Universidade de Coimbra (Portugal)

Nicola Giglietto, INFN, Bari (Italy)

Ilídio Lopes, Universidade de Évora/Departamento de Física (Portugal)

David Berry, Universidade de Évora/Departamento de Física (Portugal)

Alessandro Petrolini, University of Genova (Italy)

Antonio Bueno, University of Granada (Spain)

Manuel Pérez-Victoria, University of Granada (Spain)

Amélia Maio, Universidade de Lisboa/CFNUL/LIP (Portugal)

José Pedro Mimoso, Universidade de Lisboa/CAAUL (Portugal)

Samo Stanic, University of Nova Gorica (Slovenia)
Martin O'Loughlin, University of Nova Gorica (Slovenia)
Jorge Crispim Romão, Universidade Técnica de Lisboa/IST (Portugal)
Mário Pimenta, University Técnica of Lisboa/IST/LIP (Portugal)
Giampiero Naletto, University of Padova (Italy)
Cesare Barbieri, University of Padova (Italy)
Andrea Vitturi, University of Padova (Italy)
Giovanni Busetto, University of Padova (Italy)
Bertrand Laforgue, University of Pierre et Marie Curie-Paris VI (France)
Etienne Parizot, University of Paris Diderot-Paris VII (France)
Jose Ocariz, University of Paris Diderot-Paris VII (France)
Mário João Monteiro, Universidade do Porto (Portugal)
Orfeu Bertolami, Universidade do Porto (Portugal)
Carlos Merino, University of Santiago de Compostela/IGFAE (Spain)
Máximo Pló, University of Santiago de Compostela/IGFAE (Spain)
Luccia di Ciaccio, University of Savoie (France)
Pablo del Amo Sánchez, University of Savoie (France)
Roberto Battiston, University of Trento (Italy)
Alessandro de Angelis, University of Udine/INFN (Italy)
Massimo Persic, University of Udine (Italy)
Antonio Ferrer, University of Valencia/IFIC (Spain)
Francisco Botella, University of Valencia/IFIC (Spain)

Scientific Advisory Committee

Alan Watson, University of Leeds (UK)
Alfredo Barbosa Henriques, Universidade Técnica de Lisboa/IST (Portugal)
Alfredo Soldati, CISM Udine (Italy)
Ana Mourão, Universidade Técnica de Lisboa/IST (Portugal)
Annalisa Celotti, SISSA Trieste (Italy)
Antonio Masiero, University of Padova (Italy)
António Amorim Barbosa, Universidade de Lisboa (Portugal)
Ariella Cattai, CERN (Switzerland)
Augusto Barroso, Universidade de Lisboa/CFTC (Portugal)
Barbara De Lotto, University of Udine (Italy)
Carlos Alberto Ruivo Herdeiro, Universidade de Aveiro (Portugal)
Carlos Pajares, University of Santiago de Compostela/IGFAE (Spain)
Constança Providênci, Universidade de Coimbra (Portugal)
Danilo Zavrtanik, University of Nova Gorica (Slovenia)
Delphine Hardin, University of Paris VI (France)
Enrique Zas, University of Santiago de Compostela/IGFAE (Spain)
Enrique Fernández, IFAE (Spain)

Federico Ferrini, EGO (Italy)
Fernando Cornet, University of Granada (Spain)
Francesco Fidecaro, University of Pisa (Italy)
Francis Halzen, University of Wisconsin-Madison (USA)
Francisco del Águila, University of Granada (Spain)
Gaspar Barreira, LIP (Portugal)
George Smoot, University of California (USA)
Gustavo Castelo Branco, Universidade Técnica de Lisboa/IST (Portugal)
João Manuel Borregana Lopes dos Santos, Universidade do Porto (Portugal)
João Varela, University Técnica of Lisboa/IST/LIP (Portugal)
Joe Silk, University of Oxford (UK)
Jorge Dias de Deus, Universidade Técnica de Lisboa/IST (Portugal)
José Bernabéu, IFIC, University of Valencia-CSIC (Spain)
José Carmelo, Universidade do Minho (Portugal)
José Mariano Gago, Universidade Técnica de Lisboa/IST (Portugal)
José Sande Lemos, Universidade Técnica de Lisboa/IST (Portugal)
José F. Valle, IFIC, University of Valencia-CSIC (Spain)
Juan A. Fuster, IFIC, University of Valencia-CSIC (Spain)
Juan José Hernández, IFIC, University of Valencia-CSIC (Spain)
Luis Álvarez-Gaumé, CERN (Switzerland)
Luís António Ferreira Martins Dias Carlos, Universidade de Aveiro (Portugal)
Manuel Fiolhais, Universidade de Coimbra/FCTUC (Portugal)
Masahiro Teshima, Max-Planck Institute (Germany)
Paolo Spinelli, University of Bari (Italy)
Paulo Crawford, Universidade de Lisboa (Portugal)
Paulo Fonte, Instituto Politecnico de Coimbra/LIP (Portugal)
Pier Simone Marrocchesi, Gruppo Collegato INFN di Siena (Italy)
Riccardo Paoletti, Gruppo Collegato INFN di Siena (Italy)
Sergio Bertolucci, CERN (Switzerland)
Yannick Giraud-Heraud, University of Paris VII (France)
Yuli M. Shabelski, Petersburg Institute of Nuclear Research (Russia)

International Advisory Committee

Néstor Armesto, Universidade de Santiago de Compostela (Spain)
Fernando Barreiro, Universidad Autónoma de Madrid (Spain)
Jochen Bartels, Universität Hamburg (Germany)
Irinel Caprini, University of Bucharest (Romania)
Jiri Chyla, University of Prague (Czech Republic)
Robin Devenish, University of Oxford (UK)
Jorge Dias de Deus, Instituto Superior Técnico de Lisboa (Portugal)
Roberto Fiore, Università di Cosenza (Italy)

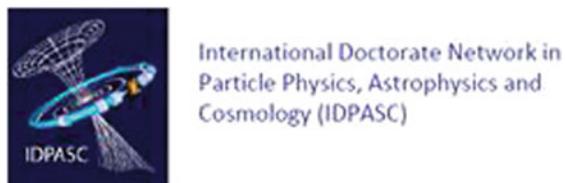
Konstantin Goulianatos, Rockefeller University of New York (USA)
Edmond Iancu, Institut de Physique Théorique, CEA-Saclay (France)
Valery Khoze, University of Durham (UK)
Christos Ktorides, University of Athens (Greece)
Peter V. Landshoff, University of Cambridge (UK)
Uri Maor, University of Tel Aviv (Israel)
Alan D. Martin, University of Durham (UK)
Pierre van Mechelen, University of Antwerpen (Belgium)
Carlos Merino, Universidade de Santiago de Compostela (Spain)
Alfred H. Mueller, Columbia University (USA)
Risto Orava, University of Helsinki (Finland)
Alessandro Papa, Universitá de Cosenza (Italy)
Gonzalo Parente, Universidade de Santiago de Compostela (Spain)
Robi Peschanski, Institut de Physique Théorique, CEA-Saclay (France)
Máximo Pló, Universidade de Santiago de Compostela (Spain)
Albert de Roeck, CERN (Switzerland)
Christophe Royon, IRFU-Service de Physique des Particules, CEA-Saclay (France)
Eddi de Wolf, University of Antwerpen (Belgium)

Local Organizing Committee

Carlos Merino (Chair), Universidade de Santiago de Compostela
Néstor Armesto, Universidade de Santiago de Compostela
José A. Hernando, Universidade de Santiago de Compostela
Carlos Pajares, Universidade de Santiago de Compostela
Gonzalo Parente, Universidade de Santiago de Compostela
Máximo Pló, Universidade de Santiago de Compostela
Juan J. Saborido, Universidade de Santiago de Compostela
Cibrán Santamarina, Universidade de Santiago de Compostela

Sponsors

Intensive Learning Programme, Education and Culture DG, European Union
Università degli Studi di Udine, Italy
Laboratório de Instrumentação e Física Experimental de Partículas, Lisboa, Portugal
Instituto Superior Técnico, Universidade de Lisboa, Portugal
Universidade de Santiago de Compostela, Galiza, Spain



International Doctorate Network in
Particle Physics, Astrophysics and
Cosmology (IDPASC)





Contents

Basic Features of High-Energy Hadronic Interactions	1
Carlos Merino and Yuli M. Shabelski	
Quantum Chromodynamics	49
Néstor Armesto and Carlos Pajares	
Selected Topics from the Experiments with High-Energy Heavy-Ion Collisions: Hot QCD in the Laboratory	99
Mateusz Płoskoń	
Flavour Physics and CP Violation in the Standard Model and Beyond	145
Gustavo Castelo-Branco and David Emmanuel-Costa	
The Higgs System in and Beyond the Standard Model	189
Maria J. Herrero	
Higgs Searches at the LHC	255
Cibrán Santamarina Ríos	
Messengers of the High Energy Universe	283
Alessandro De Angelis, Mário Pimenta and Ruben Conceição	
Numerical Methods and Data Analysis Techniques in Particle Physics	337
Mike Williams	

Theoretical Cosmology	359
Antonio L. Maroto	
Introduction to the AdS/CFT Correspondence	411
Alfonso V. Ramallo	
Welcome Speech by the Rector	475
Queimada	477
List of Participants	479



Seminario Maior San Martiño Pinario in Santiago de Compostela, Galiza, Spain. At its hospedería it was held the Third IDPASC School, January 22–February 2, 2013 (picture by C. Merino)

Basic Features of High-Energy Hadronic Interactions

Carlos Merino and Yuli M. Shabelski

Abstract We discuss some fundamental ideas lying at the base of many modern physical theories, as well as their applications.

1 Introduction

1.1 Theoretical Situation

One unified theory describing the subatomic world is not available, and only some of its components are known (see the Bronstein cube in Fig. 1. The axis correspond to mass M , velocity v , and Plank constant h , which accounts for the inverse distance $1/R$.

Out of all the theories shown into the Bronstein Cube, Newton mechanics, Special Relativity, Quantum Mechanics, and Quantum Field Theory (QED, QCD, ...) form what we know as the Standard Model.

Any established physical theory is based on a core of experimental facts, and these facts are of the largest importance when trying to formulate a new theory enlarging the scope of the first one. In this sense, Quantum Gravity cannot be considered as a real tested physical theory on the same ground as those forming the Standard Model, since there is not yet any experimental information that can be connected to its predictions and results.

In the present lectures, we will consider mainly the Quantum Field Theory, and since by now it is not a complete physical theory, we will rather focus on a

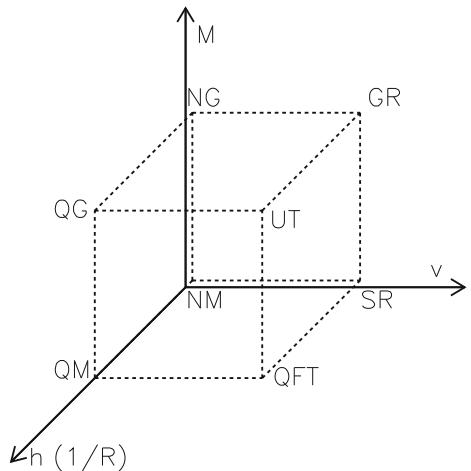
C. Merino (✉)

Departamento de Física de Partículas, Universidade de Santiago de Compostela and Instituto Galego de Física de Altas Enerxías (IGFAE), 15782 Santiago de Compostela, Spain
e-mail: carlos.merino@ucs.es

Yu.M. Shabelski

Petersburg Nuclear Physics Institute, NRC Kurchatov Institute, Gatchina,
188350 St. Petersburg, Russia
e-mail: shabelsk@thd.pnpi.spb.ru

Fig. 1 The *Bronstein Cube*
NM Newton mechanics (+);
SR Special relativity (+); *QM*
 Quantum mechanics (+);
QFT Quantum field theory
 (+/-); *NG* Newton gravity
 (+); *GM* General relativity
 (+); *QG* Quantum gravity
 (-); *UT* Unified theory of all
 interactions (-)



phenomenological point of view by presenting a large amount of experimental information in support of the theory.

1.2 Subnuclear Particles and Their Interactions

In natural units, length and time can be both measured as inverse mass/energy:

$$1 \text{ fm} \simeq 5 \text{ GeV}^{-1}, 1 \text{ GeV}^{-2} \simeq 0.39 \text{ mb}, 1 \text{ GeV}^{-1} \simeq 0.67 \times 10^{-24} \text{ s}$$

The subnuclear particles can be classified attending to the following division:

Hadrons \rightarrow Mesons, bosons ($q\bar{q}$)

Baryons, fermions (qqq)

The number of well-known hadrons is of several hundreds

And theoretically their quantity can be infinitely large.

Hadrons have an internal quark-gluon structure.

Leptons $\rightarrow e^\pm, \mu^\pm, \tau^\pm$

ν_e, ν_μ, ν_τ

Gauge bosons $\rightarrow \gamma$ (photon)

Z^0, W^\pm

g (graviton): not discovered yet.

Higgs boson $\rightarrow H$: discovered in 2012.

Four types of interaction exist: Strong, Electromagnetic, Weak, and Gravitation.
 All of them act via the exchange of one specific gauge boson.

The electromagnetic and gravitation interactions are well-known in the everyday world, and in both cases the strength of the interaction decreases with the distance as $1/R^2$. Such behaviour is connected to the fact that the gauge boson mediating in the interaction (photon and graviton, respectively) a zero mass one.

Everything, including photons, participate of the gravitation interaction, and all particles with non-zero electrical charge participate of the electromagnetical interaction.

On the other hand, the strong and weak interactions only exist in the subatomic world, because their strength exponentially decrease with the distance.

The weak interaction is produced via the exchange of W^\pm ($m_W = 80.4 \text{ GeV}$) or Z bosons ($m_Z = 91.2 \text{ GeV}$). Thus, the typical range of weak interaction is of about 10^{-2} GeV^{-1} , i.e. 0.002 fm . As a rule, we deal with weak interactions when looking at the decay processes. All leptons and hadrons participate of the weak interactions.

At very high energies, the electromagnetic and weak interactions are unified in the framework of the Electroweak Theory.

The strong interaction can be considered from two different points of view: as the interaction between hadrons via meson exchange, or as the interaction between quarks and gluons via gluon exchange.

1.3 One-Pion Exchange

The meson exchange leads to the coupling of protons and neutrons in the atomic nuclei. As the distance between nucleons in the nuclei is of the order of $1.2\text{--}1.5 \text{ fm}$, one can obtain (Yukawa 1935) the mass of the lightest meson obeying this coupling, which turns out to be the pion ($m_\pi = 0.14 \text{ GeV}$).

The meson exchange processes (mainly pion exchange) play an important role in the production of secondaries at not very high energies in so-called soft processes, i.e. in the processes with relatively small transverse momenta. For example, the diagrams presented in Fig. 2 describe the process of one-pion production in pp collision via Δ -resonance excitation.

Due to the isospin invariance, all vertices in the two diagrams of Fig. 2 are the same, except for the corresponding Clebsch-Gordan coefficients.

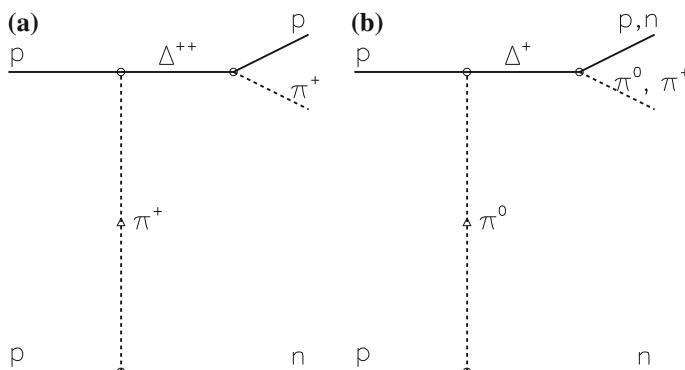


Fig. 2 Pion production in pp collisions via $\Delta(1232)$ resonance excitation

1.4 The Quark-Gluon Structure of Hadrons

At the beginning of 1960s, the number of discovered hadrons became so large (more than 200), that many physicists started to think that they could actually be composite particles, formed by a relatively small number of more fundamental constituents. Following this idea, they tried to obtain some structured classification of the hadrons.

Gell-Mann and Zweig independently proposed in 1964 that hadrons followed the SU(3) symmetry, where nucleons (protons and neutrons) and hyperons (Λ , Σ^+ , Σ^0 , Σ^- , Xi^0 , Xi^-) correspond to the octet representation of SU(3) group symmetry, light mesons (π , K , η and η') correspond to the nonet representation, etc. The lowest representation of the SU(3) group is the triplet one, and it was suggested that this fundamental representation corresponded to some hypothetical particles carrying fractional electric charges, that were called quarks (q).

The most important property of quarks is that they cannot exist as free particles (quark confinement).

The quarks have a special quantum number, called colour, which plays the role of the strong charge (coupling constant). There exist 3 colours, every quark having one of them, and antiquarks having anticolours. The $q\bar{q}$ system with the same colour-anticolour is then colourless (white) and it corresponds to a meson.

The combination of all three colours is also white, these qqq states corresponding to the baryons.

The quark-quark (or quark-antiquark) interaction proceeds via the exchange of a vector particle, the gluon g , which is bi-coloured. As gluons have colours, they also interact with each other, at difference of what happens with photons.

All interaction among quarks and gluons are described by a specific field theory, called Quantum ChromoDynamics (QCD).

Four vertices of quark-gluon interactions exist (see Fig. 3).

The strength of the quark-gluon interactions is determined by the strong coupling constant, α_s .

Contrary to other theories (e.g., QED), α_s increases with the distance between two coloured particles, and it becomes infinitively large at distances of ~ 1 fm. This explains the phenomenon of quark confinement.

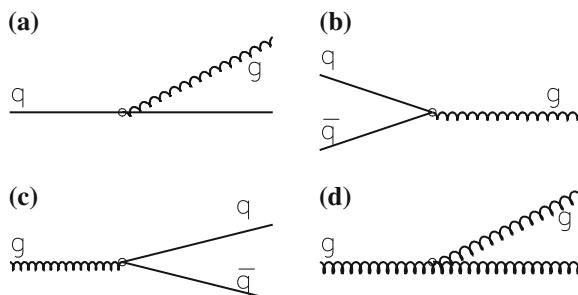


Fig. 3 QCD vertices: **a** gluon emission by quark; **b** quark-antiquark annihilation into gluon; **c** quark-antiquark pair production by gluon; **d** gluon emission by gluon

Table 1 Quantum numbers of quarks

Flavour	Charge	Mass	Other numbers
u	+2/3	1.5–4 MeV	Isospin = 1/2
d	-1/3	4–8 MeV	Isospin = 1/2
s	-1/3	80–130 MeV	Strangeness = -1
c	+2/3	1.15–1.35 GeV	Charm = +1
b	-1/3	4.1–4.9 GeV	Bottom = -1
t	+2/3	170–185 GeV	Top = +1

Quantum ChromoDynamics does not allow to calculate anything when α_s becomes of the order of unity, so the microscopical theory of confinement does not exist yet.

At present, we know 6 types of quarks, namely u , d , s , c , b , and t , with names up, down, strange, charm, beauty, and top. The types of quarks are called flavours. The electrical charges of u , c , and t quarks are equal to +2/3, and that of d , s , and b quarks is equal to -1/3. As quarks cannot be observed as free particles, their masses are not well-defined. However, the first three quarks are considered as light ones, whereas c , b , and t quarks are considered as heavy ones. All quarks are fermions with spin 1/2. The quantum numbers of quarks are presented in Table 1.

When considering hadrons as bound systems of quarks, two approaches can be used. In the case of hard interactions with large transfer momenta, hadrons are seen as consisting of valence quarks, with masses presented in Table 1, as well as of gluons and of quark-antiquark pairs which are produced radiatively by gluons. The quarks from these pairs are called sea-quarks. When considering hadron systematics and soft hadron interactions, gluons and sea pairs can be included in the content of an effective object called constituent quark. So in this cases, we can consider mesons and baryons as $\bar{q}q$ and qqq systems, respectively.

QCD describes the hard processes that occur in hadron-hadron collisions, on the level of hard interactions of partons (quarks and gluons). As examples of such a processes we can think of the production of high- p_T hadron jets, heavy particles (charmed and beauty hadrons, W and Z bosons), high mass Drell-Yan ($\mu^+\mu^-$) pairs, high- p_T direct photons, etc.

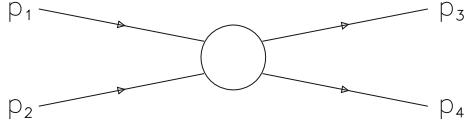
The cross sections of hard processes significantly depend on the quark and gluon structure functions.

2 S-Matrix and Its Unitarity

2.1 Mandelstam Variables

We will consider the particle interactions at high energies, and there the calculations become more simple if one uses 4-dimensional vectors and relativistic invariants. The energy E and the 3-momentum \mathbf{p} of a particle of mass m form a 4-vector $p = (E, \mathbf{p})$,

Fig. 4 Collision of two hadrons with momenta p_1 and p_2 going to a two-particle final state



whose square $p^2 \equiv E^2 - |\mathbf{p}|^2 = \mathbf{m}^2$ is a relativistic invariant. The corresponding velocity of the particle is $\beta = v/c = |\mathbf{p}|/\mathbf{E}$.

The simplest case one can think of is the $2 \rightarrow 2$ -particle interaction, that can be elastic ($pp \rightarrow pp$), or inelastic (e.g., $\pi^- p \rightarrow K^0 \Lambda$) (see Fig. 4).

Let us consider the process

$$1 + 2 \rightarrow 3 + 4 , \quad (1)$$

and let p_1, p_2, p_3, p_4 , and m_1, m_2, m_3, m_4 be the 4-dimensional momenta and the masses of the four hadrons involved, respectively. So we have

$$\begin{aligned} p_1 + p_2 &= p_3 + p_4 \\ p_1^2 &= m_1^2, \quad p_2^2 = m_2^2, \quad p_3^2 = m_3^2, \quad p_4^2 = m_4^2 \\ (p_1 + p_2 - p_3)^2 &= p_4^2 = m_4^2 . \end{aligned} \quad (2)$$

Let us introduce the Lorentz-invariant Mandelstam Variables, which are defined by

$$\begin{aligned} s &= (p_1 + p_2)^2 = (p_3 + p_4)^2 \\ t &= (p_1 - p_3)^2 = (p_2 - p_4)^2 \\ u &= (p_1 - p_4)^2 = (p_2 - p_3)^2 , \end{aligned} \quad (3)$$

and that satisfy

$$s + t + u = m_1^2 + m_2^2 + m_3^2 + m_4^2 , \quad (4)$$

so only two out of the three Mandelstam variables are independent.

The variable s has a physical sense of the energy. In laboratory frame:

$$s = p_1^2 + p_2^2 + 2p_1 \cdot p_2 = m_1^2 + m_2^2 + 2m_2 E_{1lab.}, \quad (5)$$

since $\mathbf{p}_2 = 0$, and $p_1 \cdot p_2 = m_2 E_{1lab.}$

In the c.m. frame, by definition $\mathbf{p}_2 = -\mathbf{p}_1$, so $p_1 \cdot p_2 = E_1 E_2 + |\mathbf{p}|^2$, and

$$s = (E_{1cm} + E_{2cm})^2 . \quad (6)$$

The variables t and u in (3) correspond to the transfer momenta.

In the important case of elastic scattering, when particles in the final state are the same as those in the initial state, we have in c.m. frame:

$$t = -2|\mathbf{p}|^2(1 - \cos\theta) = -4|\mathbf{p}|^2 \sin^2(\theta/2) . \quad (7)$$

2.2 S-Matrix and Transition Amplitude

Here we will consider the formalism of the so-called S-matrix to see that by using the very general properties, such as unitarity, analyticity, conservation laws, etc., one can obtain very important information about the interactions of elementary particles.

Let us consider a sample of physical states and the possible transitions of state Φ_i into state Φ_f , that are written as the modulo squared of the corresponding matrix element of the S-matrix:

$$S_{fi} = \langle \Phi_f^+(t = +\infty) | \hat{S}(+\infty, -\infty) | \Phi_i(t = -\infty) \rangle . \quad (8)$$

The S-matrix elements themselves are evaluated between asymptotic states at times $t = \pm\infty$ (the initial state a long time before the interaction starts and the final state a long time after it finished).

If the case of no-interaction, the state of the system cannot change, what corresponds to the unit S-matrix $I = \delta_{fi}$. Generally

$$S_{fi} = \delta_{fi} + iT_{fi} . \quad (9)$$

The transitions between the states i and f are possible if $T_{fi} \neq 0$, and the matrix elements T_{fi} determine the transition amplitude $A(i \rightarrow f)$, a Lorentz-invariant scalar that is written as

$$T_{fi} = (2\pi)^4 \delta^{(4)}(\sum p_i - \sum p_f) A(i \rightarrow f) , \quad (10)$$

where the Lorentz invariant amplitude $A(i \rightarrow f)$ is called the scattering amplitude, and it depends only on Lorentz invariant scalars.

The scattering amplitudes are analytical functions of their variables, and so, $A(p_1, p_2, p_3, p_4)$ has physical meaning at arbitrary values of p_1, p_2, p_3, p_4 . The important point here is that one can see the incoming particle with negative value of 4-momentum (in the sense that if $p_1 + p_2 = p_3 + p_4, p_3 + p_4 - p_1 - p_2 = 0$) as one outgoing antiparticle with positive 4-momentum, and vice versa.

Let us consider, for example, the elastic proton-neutron scattering. The incident and final state protons have momenta p_1 and p_3 , whereas the incident and final state neutrons have momenta p_2 and p_4 , and the scattering amplitude for the process $1 + 2 \rightarrow 3 + 4$ is $A_{1+2 \rightarrow 3+4}(p_1, p_2, p_3, p_4)$. The crossing assumption means that

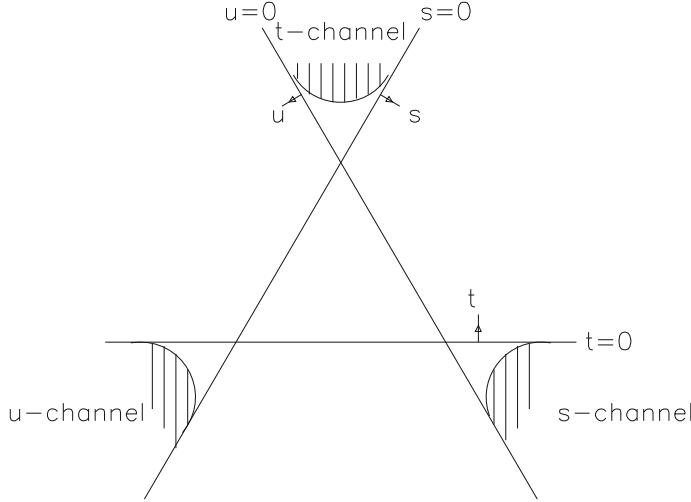


Fig. 5 The Mandelstam plane for the elastic scattering $1 + 2 \rightarrow 3 + 4$

the same analytical function should describe the processes $1 + \bar{3} \rightarrow \bar{2} + 4$ and $1 + \bar{4} \rightarrow \bar{2} + 3$. Thus, we have:

$$\begin{aligned} \text{s-channel, } p + n &\rightarrow p + n, s > 0, t < 0, u < 0 \\ \text{t-channel, } p + \bar{n} &\rightarrow n + \bar{n}, t > 0, s < 0, u < 0 \\ \text{u-channel, } p + \bar{n} &\rightarrow p + \bar{n}, u > 0, t < 0, s < 0 \end{aligned}$$

The values of the variables s, t, u in these three cases are not overlapping. This can be seen in the so-called Mandelstam Plane with the coordinates triangle shown in Fig. 5.

2.3 Probabilities and Cross Sections [1]

The probability of transition from the initial state i into the final state f ($i \neq f$) during a unit of time t is equal to

$$\tilde{w}_{i \rightarrow f} = \frac{1}{t} |T_{fi}|^2 = \frac{1}{t} (2\pi)^8 [\delta(\sum p_i - \sum p_f)]^2 |A(i \rightarrow f)|^2 , \quad (11)$$

where the square of the δ -function appears. The δ -function is an operator defined by

$$\int f(x) \delta(x - a) dx = f(a) . \quad (12)$$

The δ -function in (11) can be represented as

$$\delta^4(\sum p_i - \sum p_f) = \frac{1}{(2\pi)^4} \int d^4x \cdot e^{(\sum p_i - \sum p_f)x} . \quad (13)$$

To account for the fact that different final states are possible, the probability of transition $\tilde{w}_{i \rightarrow f}$ should be multiplied by the number of final states

$$d\tau_n = (2\pi)^4 \delta(\sum p_i - \sum p_f) \prod_{k=1}^n \frac{d^3 p_k}{(2\pi)^3 2E_k} , \quad (14)$$

where n is the number of particles in the final state, p_k is the 3-dimensional momentum of k -th particle, and the last factor is the n -particle phase space, that it can be written in a Lorentz-invariant form as

$$d\tau_n = \frac{d^3 p_k}{(2\pi)^3 2E_k} = \frac{d^4 p_k}{(2\pi)^3} \delta(p_k^2 - m_k^2) . \quad (15)$$

For the decay of a particle with mass m into several (n) particles, in the rest frame, $E_{i=1} = m$,

$$dw_{i \rightarrow f} = \frac{1}{2m} |A(i \rightarrow f)|^2 d\tau_n . \quad (16)$$

The most important physical magnitude is the cross section, that shows the effective range of the particle interaction. The cross section can be obtained by dividing the probability dw by the flow density j of the incident particles

$$d\sigma_{i \rightarrow f} = dw_{i \rightarrow f}/j , \quad (17)$$

where

$$j = \frac{I}{VE_a E_b} , \quad (18)$$

that results in

$$d\sigma = \frac{|A(i \rightarrow f)|^2}{4I} d\tau_n , \quad (19)$$

where

$$I = \sqrt{(p_a p_b)^2 - m_a^2 m_b^2} . \quad (20)$$

In the case of elastic scattering and azimuthal symmetry $d\Omega = 2\pi d\cos(\theta)$, $|p_0| = |p|$, $dt = 2|p|^2 d\cos\theta$, and we obtain

$$\frac{d\sigma}{dt} = \frac{1}{64\pi^2|p|^2 s} |A_{el}|^2 = \frac{1}{64\pi^2|p_{a(lab)}|^2 m_b^2} |A_{el}|^2 . \quad (21)$$

Sometimes it is more suitable to define the scattering amplitude in a different way:

$$f_{i \rightarrow f} = \frac{A_{i \rightarrow f}}{8\pi\sqrt{s}}, \quad \text{or} \quad M_{i \rightarrow f} = \frac{A_{i \rightarrow f}}{16\pi p_0 \sqrt{s}} . \quad (22)$$

The integral of the differential cross section provides the total cross section of the considered process.

2.4 Unitarity and the Optical Theorem

The very important property of the S-matrix is its unitarity:

$$S^+ \cdot S = 1 , \quad (S^+ \cdot S)_{fi} = \sum_n S_{nf}^* \cdot S_{ni} = \delta_{fi} , \quad (23)$$

where the sum is performed over all channels n and we integrate over all variables of the intermediate states.

One can write

$$(S^+ \cdot S)_{fi} = \sum_n (\delta_{fn} - i \cdot T_{fn}^+) (\delta_{ni} + i \cdot T_{ni}) = \delta_{fi} , \quad (24)$$

i.e.

$$T_{fi} - T_{if}^* = \sum_n i \cdot T_{nf}^* \cdot T_{ni} . \quad (25)$$

For the amplitudes A , we obtain:

$$\begin{aligned} A(i \rightarrow f) - A^*(i \rightarrow f) \\ = i \cdot (2\pi)^4 \cdot \sum_n \delta(\sum p_n - \sum p_f) A(i \rightarrow n) \cdot A^*(n \rightarrow f) , \end{aligned} \quad (26)$$

that is, the unitarity condition for the scattering amplitude.

In the case of elastic scattering $i = f$, we have

$$2 \cdot Im A_{el} = (2\pi)^4 \sum_n \delta(\sum p_n - \sum p_f) |A(i \rightarrow n)|^2 , \quad (27)$$

where the imaginary part of elastic scattering amplitude $Im A(i \rightarrow i)$ is expressed via the sum of all possible intermediate states n . Thus:

$$\begin{aligned} ImA_{el}(s, t) &= \frac{1}{2} \sum_n \int |A(i \rightarrow n)|^2 d\tau_n \\ &= \frac{1}{2} \sum_n \Delta A^{(n)}(s, t) = \frac{1}{2} \Delta A(s, t), \end{aligned} \quad (28)$$

that is non-zero only at $\sqrt{s} \geq \sqrt{s_n}$, where s_n is the threshold energy for the corresponding channel.

New inelastic channels will appear when the energy increases, leading to the presence of new contributions to the absorptive part $\Delta A^{(n)}$ of the elastic scattering amplitude.

From the unitarity condition we obtain the relation

$$ImA_{i \rightarrow i}(q = 0) = 2 \cdot p_0 \cdot \sqrt{s} \cdot \sum \sigma_n = 2 \cdot p_0 \cdot \sqrt{s} \cdot \sigma^{tot}, \quad (29)$$

that is known as the optical theorem.

2.5 One Example: $K_L^0 \rightarrow \mu^+ \mu^-$ Decay

Let us consider the probability for the rare decay $K_L^0 \rightarrow \mu^+ \mu^-$ to occur. This was considered as a problem in early 70s, and it took some time before this decay was discovered.

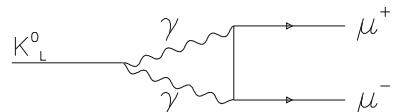
From the theoretical point of view one can consistently argue that the main contribution to this decay comes from the diagram shown in Fig. 6, i.e. via $\gamma\gamma$ intermediate state.

The unitarity condition determines the imaginary part of the corresponding amplitude by

$$ImA(K_L^0 \rightarrow \mu^+ \mu^-) = \int A^*(\mu^+ \mu^- \rightarrow \gamma\gamma) \cdot A(K_L^0 \rightarrow \gamma\gamma) d\tau_{\gamma\gamma}. \quad (30)$$

The two amplitudes in the right-hand side are both known, the first one from QED, and the second one from the known decay $K_L^0 \rightarrow \gamma\gamma$, what allows us to calculate the low boundary for the $K_L^0 \rightarrow \mu^+ \mu^-$ decay probability.

Fig. 6 $K_L^0 \rightarrow \mu^+ \mu^-$ decay via $\gamma\gamma$ intermediate state



2.6 Froissart Theorem

In 1961 Froissart obtained the bound for the increase of the total cross section with energy:

$$\sigma^{tot}(s) \leq C \cdot \ln^2 s, s \rightarrow \infty. \quad (31)$$

This boundary has its origin in the Yukawa potential for strong interactions, where the probability of interaction at large distance r , is $P(r) = P_0 \cdot e^{-b \cdot r}$. Then, the corresponding total inelastic cross section can not be larger than the geometrical cross section, πb^2 , and the corrections to this estimation can not be stronger than logarithmic.

The detailed proof of the Froissart theorem is based on the partial waves (i.e. orbital momentum) expansion, and on dispersion relations (see next Section), and it can be followed in the original paper [2].

3 Analyticity and Dispersion Relations

3.1 Singularities of the Scattering Amplitude

All physical amplitudes are assumed to be the analytical functions of their variables.

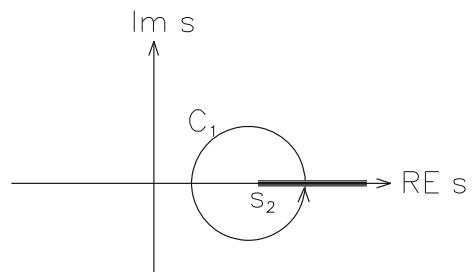
At energies below the elastic threshold, i.e. at $s < (m_1 + m_2)^2$, the elastic scattering amplitude $1 + 2 \rightarrow 1 + 2$ is usually purely real.

At $s > (m_1 + m_2)^2$, the values $A(s + i\varepsilon)$ and $A(s - i\varepsilon)$ become different when $\varepsilon \rightarrow 0$, what reflects the fact that a cut exists in the real axis of the complex s -plane, and that this cut starts at the point $s = s_2 = (m_1 + m_2)^2$ and it goes to infinity, as it is shown in Fig. 7.

The physical amplitude is determined on the up border of the cut, i.e. $A(s + i\varepsilon)$, and so this part of plane is called the physical sheet.

At higher energy, new branching points s_3 and cuts appear. The discontinuity of the elastic scattering amplitude comes from the difference between its values on the

Fig. 7 Elastic scattering amplitude $1 + 2 \rightarrow 1 + 2$, together with its first two-particle cut in the complex s -plane



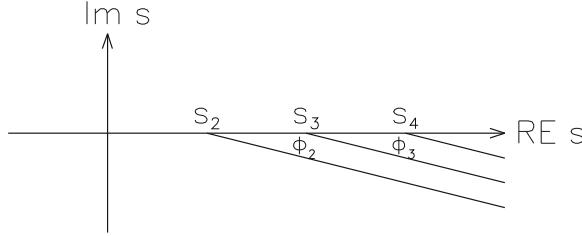


Fig. 8 Elastic scattering amplitude $1 + 2 \rightarrow 1 + 2$ with its two-particle, three-particle and four-particle branching point that correspondent to particle production in the intermediate state together with the corresponding cuts in the complex s -plane. Really all angles $\phi_i \rightarrow 0$

up and on the down borders of the cut:

$$\Delta A(s) = A(s + i\varepsilon) - A(s - i\varepsilon), \quad (32)$$

that coincides with the absorptive part of the amplitude.

To determine the scattering amplitude $A(s)$, all its cuts in the complex s -plane, from $s = s_n$ to infinity, should be taken into account, as it is shown in Fig. 8.

For a fixed value of the variable $t = t_0$, the variable u , that corresponds to the u -channel reaction $1 + \bar{4} \rightarrow \bar{2} + 3$, depends of the s variable as $u = \sum m_i^2 - s - t$. This point is also the branching point of the amplitude $A(s)$, and the corresponding cut exists in the real axis of s in the negative direction, until $s = -\infty$.

Usually, the region where the scattering amplitude is real and $A^*(s) = A(s^*)$ exists. The physical amplitude is then determined upper the right-hand-side cut and below the left-hand-side cut, i.e.:

$$\begin{aligned} A(s) &= A(s + i\varepsilon), \\ A(u) &= A(u + i\varepsilon) = A(s - i\varepsilon). \end{aligned} \quad (33)$$

The discontinuity on the right-hand cut being

$$\Delta A_1(s, t_0) = \sum_n \Delta A^{(n)}(s, t_0) = \text{Im} A_{1+2 \rightarrow 1+2}(s + i\varepsilon, t_0), \quad (34)$$

and the discontinuity on the left-hand cut

$$\Delta A_2(u, t_0) = \sum_n \Delta A^{(n)}(u, t_0) = \text{Im} A_{1+\bar{2} \rightarrow 1+\bar{2}}(s - i\varepsilon, t_0). \quad (35)$$

3.2 Poles of the Scattering Amplitude and the Dispersion Relation

The scattering amplitude at the point $s + i\delta$ can be written as the Cauchy integral over an arbitrary contour around this point on the physical sheet:

$$A(s + i\delta, t_0) = \frac{1}{2\pi i} \cdot \oint \frac{A(s')}{s' - (s + i\delta)} ds' . \quad (36)$$

If the only singularities of the scattering amplitude on the physical sheet are the cuts and their branching points, and s -and u -channel cuts do not overlap, one can deform the contour as it is shown in Fig. 9.

If the amplitude $A(s)$ decreases at $s \rightarrow \infty$ faster than $1/(\ln s)^2$, the first integral over the contour C_R is negligible at $R \rightarrow \infty$, and we obtain

$$\begin{aligned} A(s, t = t_0) &= \frac{1}{\pi} \int_{s_2}^{\infty} \frac{\Delta A_1(s', t_0)}{s' - s - i\delta} \cdot ds' \\ &+ \frac{1}{\pi} \int_{u_2}^{\infty} \frac{\Delta A_2(u', t_0)}{u' - u + i\delta} \cdot du' \quad \delta \rightarrow 0 . \end{aligned} \quad (37)$$

Often, bound states with $m^2 < s_2$, or $m^2 < u_2$, exist in the system 12 (or in the system $1\bar{2}$). The simplest example is a pion in the case of $N\bar{N}$ scattering.

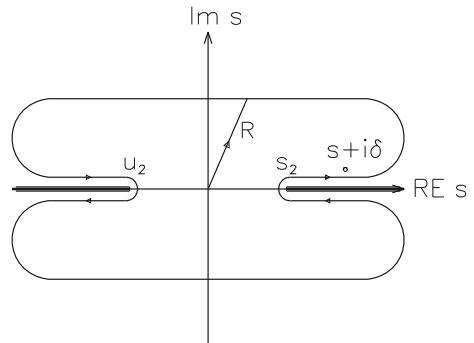
For only one pole we obtain

$$\Delta A_1(s, t_0) = \pi |g_{12}^i|^2 \cdot \delta(s - s_i) , \quad (38)$$

with $s_i = m_i^2$.

Taking into account both the poles in s -channel and in u -channel, we can write the final expression as follows:

Fig. 9 Contour integral which determines the elastic scattering amplitude at the point $s + i\delta$



$$\begin{aligned}
A(s, t_0) = & \sum_i \frac{a_i}{(s_i - s)} + \sum_j \frac{a_j}{(u_j - u)} \\
& + \frac{1}{\pi} \int_{s_2}^{\infty} \frac{\Delta A_1(s', t_0)}{s' - s - i\delta} \cdot ds' + \frac{1}{\pi} \int_{u_2}^{\infty} \frac{\Delta A_2(u', t_0)}{u' - u + i\delta} \cdot du' . \quad (39)
\end{aligned}$$

This expression is known as the dispersion relation for the elastic scattering amplitude at fixed $t = t_0$.

Here, it is assumed that only poles and cuts exist on the physical sheet, and all other singularities of the elastic scattering amplitude are placed on unphysical sheets.

3.3 Real Part of the Elastic Scattering Amplitude

In the operator language

$$\frac{1}{s' - s \mp i\varepsilon} = \frac{1}{s' - s} \pm i \cdot \pi \cdot \delta(s' - s) , \quad (40)$$

so we can write the real part of the amplitude as

$$\begin{aligned}
ReA(s, t_0) = & \sum_i \frac{a_i}{(s_i - s)} + \sum_j \frac{a_j}{(u_j - u)} \\
& + P \cdot \frac{1}{\pi} \cdot \int_{s_2}^{\infty} \frac{\Delta A_1(s', t_0)}{s' - s} ds' + P \cdot \frac{1}{\pi} \cdot \int_{u_2}^{\infty} \frac{\Delta A_2(u', t_0)}{u' - u} du' . \quad (41)
\end{aligned}$$

If the elastic scattering amplitude $A(s, t_0)$ decreases with s slower than $1/\ln^2 s$, then one can write the Couchet integral (3.5) for the quantity

$$A^{(1)}(s) = \frac{A(s) - A(s_0)}{s - s_0} , \quad (42)$$

where s_0 is an arbitrary point, and we obtain the dispersion relation with one subtraction.

The real part of the elastic scattering amplitude can be experimentally measured via the interference between strong and electromagnetic (Coulomb) scattering at very small $|t|$. The discontinuities $\Delta A_1(s, t)$ and $\Delta A_2(s, t)$ at $t = 0$ can be expressed via the total 12 and $1\bar{2}$ cross sections. Thus, in the laboratory frame

$$\begin{aligned}
\Delta A_1(s) &= 2 \cdot p_1 \cdot m_2 \cdot \sigma_{12}^{tot} , \\
\Delta A_2(s) &= 2 \cdot p_1 \cdot m_2 \cdot \sigma_{1\bar{2}}^{tot} , \quad (43)
\end{aligned}$$

and $s' - s = 2 \cdot m_2 \cdot (E'_1 - E_2)$, $u' - u = 2 \cdot m_2 \cdot (E'_1 + E_1)$, and for the real part of elastic 12 amplitude we have

$$\begin{aligned} ReA_{12}^{el} &= \sum_i \frac{a_i}{2 \cdot m_2 \cdot (E_i - E)} + \sum_j \frac{a_j}{2 \cdot m_2 \cdot (E_j - E)} \\ &+ \frac{2 \cdot m_2}{\pi} \int_{m_1}^{\infty} \left[\frac{\sigma_{12}^{tot}}{E' - E} + \frac{\sigma_{12}^{tot}}{E' + E} \right] \cdot \sqrt{E'^2 - m_1^2} \cdot dE' , \end{aligned} \quad (44)$$

where E_i and E_j are the energies in the laboratory frame corresponding to the bound states production, that is

$$E_i = \frac{m_i^2 - m_1^2 - m_2^2}{2m_2} . \quad (45)$$

The dispersion relation can be checked up by comparison to experimental data. Its violation implies the violation of, as a minimum, one of the following fundamental principles of elementary particle physics: conservation of probabilities, micro-causality, Lorentz-invariance, and structure of the S -matrix singularities.

The real part of the elastic scattering amplitude depends on the integrals over total cross sections. This means that the precise measurement of the real part of the elastic scattering amplitude gives us information on the behaviour of the total cross sections at energies where direct measurements of them are not possible.

3.4 Dispersion Relations for πp and pp Scattering

Usually, one considers the ratio of the real to imaginary parts of the scattering amplitude

$$\rho_{hN} = \frac{ReA(s)}{ImA(s)} . \quad (46)$$

The comparison of the experimental data on the values of $\rho_{\pi^- p}$ to the theoretical calculations is presented in Fig. 10.

For proton-proton elastic scattering, the theoretical situation is more complicated. In s -channel there exist only cuts, the first one starting from the point $s = 4m_p^2$, but in u -channel ($p\bar{p}$ scattering) one has the one-pion pole at $u = m_\pi^2$, then the cuts corresponding to the virtual production of 2π , 3π , etc., are present, and, only after that, the elastic scattering cut due to $p\bar{p} \rightarrow p\bar{p}$ elastic scattering appears at larger u . Thus, in this case the pole and the unphysical discontinuity are in the interval $4m_\pi^2 < u < 4m_p^2$, that is experimentally unknown.

Fortunately, the contribution of this discontinuity to ρ_{pp} decreases at high energies as $1/s$, and it can be accounted for by the subtraction term at not very low energy.

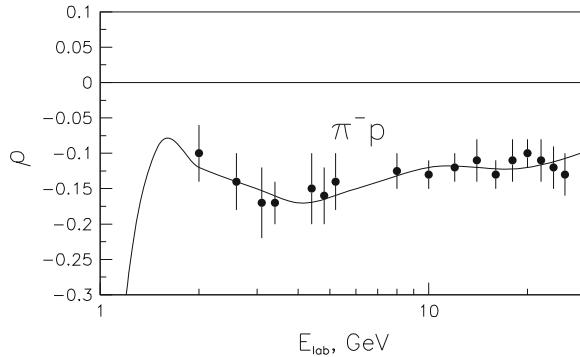


Fig. 10 Experimental data on $\frac{ReA(s)}{ImA(s)}$ for $\pi^- p$ elastic scattering (points) and their comparison to the dispersion relation calculations (curve)

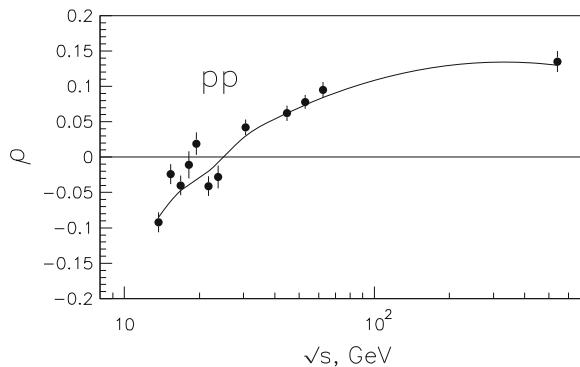


Fig. 11 Experimental data on $\frac{ReA(s)}{ImA(s)}$ for pp elastic scattering (points) and their comparison to the dispersion relation calculations (curve)

So, this unphysical discontinuity can create problems for the theoretical calculations only at comparatively low energies.

The comparison of the experimental data on the values of ρ_{pp} scattering to the theoretical calculations is presented in Fig. 11.

4 General Features of Inelastic Hadron Collisions

4.1 Hadron Interactions in Different Energy Regions

The hadron interaction processes can be classified (of course, rather subjectively) into different energy regions.

At low energies, the elastic scattering dominates, sometimes via s -channel resonances, i.e. via pole diagrams.

At intermediate energies, the structure of elastic hN interactions is rather complicated. Sometimes it is considered as a meson exchange in t -channel. The quasielastic processes come to play an important role when in the processes $1 + 2 \rightarrow 3 + 4$, one or both final particles are resonances. Also a number of s -channel resonances exist, and the processes of real multiparticle production start appearing.

At high energies, $\sqrt{s} = 5 - 50$ GeV, the structure of the interactions becomes simpler. The Regge-pole exchange phenomenology, which we will consider later, provides an useful tool for the description of many quantitative features of high energy hadronic collisions at small transferred momentum, while at high transferred momentum different QCD approaches can be used.

The total cross sections of hN interactions, shown in Fig. 12, are practically constant.

By looking carefully one can appreciate that first the total cross sections slightly decrease with energy, and later start increasing. The elastic cross sections repeat this behaviour.

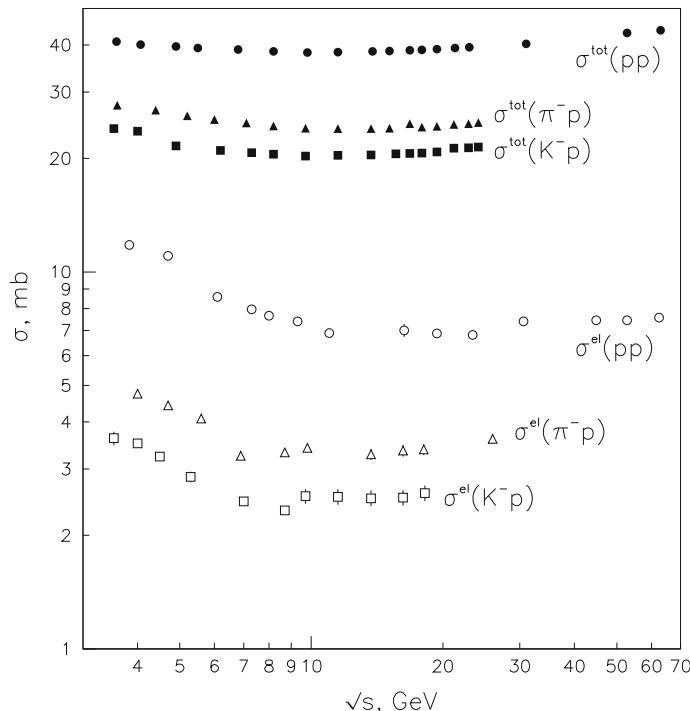


Fig. 12 Total and elastic cross sections of pp , π^-p , and K^-p collisions

$$\frac{\sigma_{pN}^{el}}{\sigma_{pN}^{tot}} \simeq \frac{1}{6}, \quad \frac{\sigma_{\pi N}^{el}}{\sigma_{\pi N}^{tot}} \simeq \frac{1}{8}, \quad \frac{\sigma_{KN}^{el}}{\sigma_{KN}^{tot}} \simeq \frac{1}{10}. \quad (47)$$

At $p_{lab} > 30 \text{ GeV}/c$, more than 80 % of the total cross section is associated with processes in which new particles are produced. As the beam energy increases, the average number of the produced secondaries per inelastic event, $\langle n \rangle$, also increases, first as $\sim \ln s$, and then even faster at higher energies.

All produced secondaries can be divided into two groups, fragmentation particles and centrally in rapidity produced particles. The first ones carry out a significant part of the energy of the incident hadrons, and their quantum numbers correlate with the quantum numbers of the incident hadrons, e.g. in the case of a proton beam the fastest secondaries usually are proton and neutron. The reason for this is that rather often the fragmentation secondaries contain the valence quarks of the incident particles. The centrally produced particles have rather small longitudinal momenta in c.m. frame, are produced in equal particle and antiparticle portions, and they contain only sea quarks produced in the collision.

The main part of produced secondaries are pions, a large fraction of them appearing as the result of meson and baryon resonances decay. The number of produced K -mesons is significantly smaller than the number of pions, and the number of antibaryons is even smaller. Thus, the ratios of particles produced in pp collisions at $\sqrt{s} = 200 \text{ GeV}$ in the central region are the following:

$$K^-/\pi^- \simeq 1/10, \quad \bar{p}/\pi^- \simeq 1/20. \quad (48)$$

The total cross sections of antiparticles with protons are larger than those of particles due to the presence of s -channel poles contributions in the cases of antiparticle interactions, i.e. there is a pion pole in $\bar{p}p$ scattering and a nucleon pole in π^-p scattering, while there are no s -channel poles in the cases of pp , and only small number of poles in π^+p scattering.

At very high energies, up to now (it was written before appearing LHC data) only several experimental points with rather large error bars obtained in colliders exist (see Fig. 13). However, they show an evident rather fast increase of the total cross section with energy. The multiplicities of secondary particles also increase and, at the same time, the probability of hard processes such as gluon jet production becomes significant. The dependence of cross sections with the initial energies is governed by Froissart theorem.

4.2 Neutral Kaon Regeneration [3]

The processes that take place in the system of neutral kaons present a very nice example of quantum effects in high energy physics.

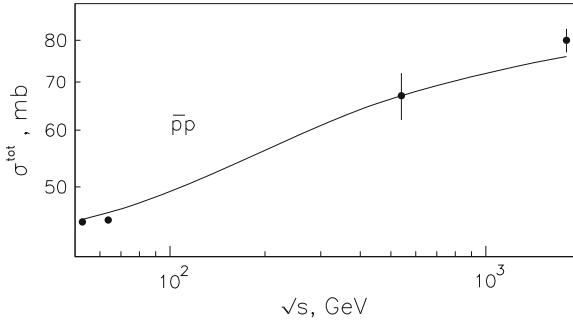


Fig. 13 Total cross sections of $\bar{p}p$ (points) collisions at very high energies

From the point of view of the strong interaction two states exist, K^0 and \bar{K}^0 , which have strangeness equal to +1 and -1, respectively. Strangeness is conserved in strong interactions, so K^0 and \bar{K}^0 are different particles.

Free kaons decay via weak interaction, where strangeness is not conserved, but the CP -parity is approximately conserved (its violation occurs with very small probability). The systems $K^0 = d\bar{s}$ and $\bar{K}^0 = \bar{d}s$ have not well defined values of CP -parity. On the other hand, when looking to the weak decay, two other systems have to be considered, K_S^0 , with $CP = +1$ and which decays into $\pi\pi$ with mean life time 0.9×10^{-10} s, and K_L^0 with $CP = -1$ and which decays into 3π .

Thus from the point of view of the strong interaction (hadrons), the relevant systems are

$$K^0 = \frac{(K_S + K_L)}{\sqrt{2}}, \quad \bar{K}^0 = \frac{(K_S - K_L)}{\sqrt{2}}, \quad (49)$$

while concerning the weak decay, one has to consider the systems

$$K_S = \frac{(K^0 + \bar{K}^0)}{\sqrt{2}}, \quad K_L = \frac{(K^0 - \bar{K}^0)}{\sqrt{2}}. \quad (50)$$

Now, let us consider two interesting cases.

First, the case of the low energy reactions $\pi^- p \rightarrow K^0 \Lambda$ or $pp \rightarrow K^0 \Lambda p$, when only K^0 can be produced below the threshold of $K^0 \bar{K}^0$ pair production, but when K_S and K_L , that, in accordance with (48), contain K^0 , will decay with their own mean life times, which values have been presented above. Then, at some distance, the number of K_L will be many times larger than the number of K_S , and, following with (49), \bar{K}^0 will appear.

In the second case, let us consider the kaon beam at a significantly large distance from the production point, so that all K_S have already decayed, and an only pure K_L beam is still active. Let this beam interact with another extended target T, as it is shown in Fig. 14.

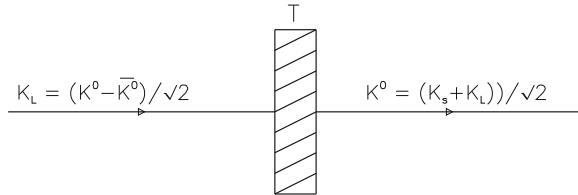


Fig. 14 Regeneration of K -mesons

In strong interactions we should consider separately the interactions of K^0 and \bar{K}^0 . The absorption cross section of \bar{K}^0 on any target is larger than the corresponding K^0 absorption cross section, so the main part of \bar{K}^0 will be absorbed during the collision, and after the target a pure K^0 beam will remain, equally consisting of K_S and of K_L , in agreement with (48). The process of appearance of K_S after a target is called K -meson regeneration.

4.3 Multiperipheral Collisions

The experimental fact is that in inelastic hadron-hadron interactions at high energy the produced particles have, as a rule, small transverse momenta (p_T of the order of their masses). At the same time, the longitudinal momenta (p_L) can be as large as possible with the only constraint of energy conservation.

This situation can be shown in so-called Peyrou plot, where the particle population is presented on the $p_L - p_T$ plane (see Fig. 15).

The main bulk of secondary particles are placed below the horizontal line which schematically separates the regions of low and high p_T , i.e. in the regions of diffraction dissociation (fragmentation regions), and in the central region, and only a few of them can be found in the region of hard (high p_T) collision processes.

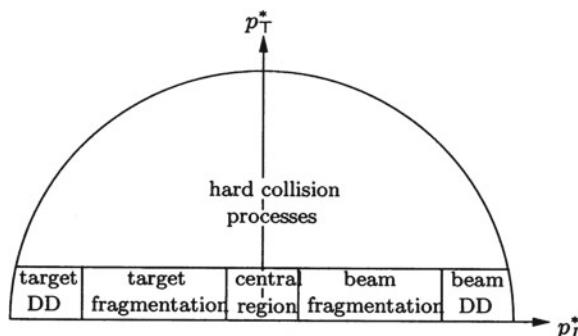


Fig. 15 Peyrou plot defining the various regions for a high energy collision (starred variables refer to the c.m. frame)

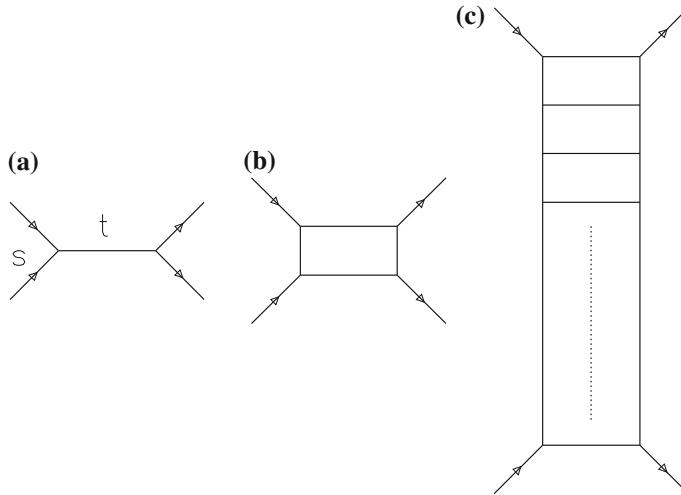


Fig. 16 Sequence of t -channel ladder Feynman diagrams: **a** the single particle exchange (Born approximation); **b** the box diagram; **c** n -rung ladder

The main properties of the soft multiparticle production at high energies are characteristic features of the multiperipheral collisions picture. At high energies, the so-called ladder diagrams dominate (see Fig. 16).

The ladder is called multiperipheral if all the produced secondaries can be enumerated according to the decreasing order of their longitudinal momenta, and if all the transverse momenta are small enough:

$$p_{1L} \geq p_{2L} \geq \cdots \geq p_{nL}, \quad p_{iT} \leq m_i. \quad (51)$$

In the multiperipheral processes all the squared momenta transferred along the ladder, t_i , and all the squared adjacent invariant masses, $s_{i,i+1}$, are limited as $s \rightarrow \infty$.

4.4 Formation Time

The very important theoretical point is the existence of a formation time for secondary production in hadronic collisions.

In QED, the accelerated electron with energy E_e , velocity v_e , and mass m_e , emits the photons with wavelength λ from a distance [4]:

$$\lambda_f \simeq \frac{\lambda}{1 - v_e} \simeq \frac{\lambda}{2} \left(\frac{E_e}{m_e} \right)^2. \quad (52)$$

A similar example was demonstrated for strong interactions on a quantum mechanical level [5, 6]. Let us consider the case of deuteron-nucleus interaction with impact parameter b , such that

$$R_A < b < R_A + R_d . \quad (53)$$

If the transfer momenta are comparatively small, both the deuteron ground state and its disintegration can occur. After the interaction, we should expand the wave function of the final state d^* as the superposition of deuteron and of free proton and neutron:

$$\Psi_{d^*} = C_0 \Psi_d + \sum_i C_i \Psi_{pn}^{(i)} , \quad |C_0|^2 + \sum_i |C_i|^2 = 1 . \quad (54)$$

It is possible to write such a superposition only at rather large time/distance, when nucleons in d^* state will interact. This needs of a time of the order of R_d/v . Thus, the state d^* can be discriminated as being a deuteron or as being a free $p + n$ state only at a distance larger than

$$l_f = \frac{E_d}{m_d} \cdot \frac{R_d}{v} . \quad (55)$$

This length is the formation length of the considered process. The value of R_d/v is of the order of 10 fm, so the formation length for relativistic deuterons with a large Lorenz-factor, E_d/m_d , can be many times larger than the nuclear radii. At the distances smaller than l_f we can not say anything about the nature of the state d^* .

4.5 Secondary Production in QCD

The simplest example of the real QCD mechanism of secondary production is high energy e^+e^- -annihilation.

In the first step of this process, the virtual photon produces a quark-antiquark pair, $q_1\bar{q}_1$, at small distances with large relative momentum (see Fig. 17a). When the distance between these two quarks becomes of about 1 fermi (the confinement range), the tension of the colour field becomes so large that a new quark-antiquark pair, $q_2\bar{q}_2$, can be produced (Fig. 17b).

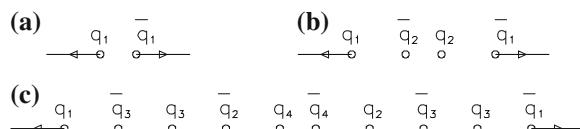


Fig. 17 The multiple quark-antiquark pair production in e^+e^- -annihilation that results in multiple production of secondary hadrons

The probability that these four quarks will turn into two mesons (say, pions) is small. More probably, several new quark-antiquark pairs, $q_3\bar{q}_3$, $q_4\bar{q}_4$ and $q_5\bar{q}_5$, will be produced, what corresponds to the multiple production of secondaries, as it is shown in Fig. 15c, where five secondary mesons will be produced after hadronization of five quark-antiquark pairs. We will call such a configuration of quarks a coloured string.

In the case of hadron-hadron inelastic interaction, the most probable is the production of a gluon (octet) string, $8 \times \bar{8}$.

The most important consideration is that due to the confinement mechanism of secondary production, one needs of about 1 fm in longitudinal distance for the production of every quark-antiquark pair (i.e. for each secondary hadron). So, at high energies the space distance of secondary production can be of about 10 fm, or even more, what is larger than the nucleus radius. This means that in the case of secondary production on a nuclear target, even without accounting for the formation time effect, some part of the secondaries will be produced outside the nucleus.

5 Regge-Gribov Theory [7–9]

5.1 Regge Trajectories

Now we will consider the scattering theory in which the scattering amplitudes are analytically continued in the complex angular-momentum j -plane.

By analytically continuing the partial-wave amplitudes in angular momentum, one can represent the scattering amplitude as a sum of pole and cut contributions in the complex j plane. These singularities are related to the asymptotic behaviour of the scattering amplitude. The summation of ladder diagrams at high energies leads to the scattering amplitudes that contain the so-called Regge trajectories.

$$A(s, t) = \sum_i A_i \sim \sum_i g_i^2 \cdot s^{\alpha_i(t)}, \quad (56)$$

where the Regge trajectories $\alpha_R(t)$ are usually written as

$$\alpha_R(t) = \alpha_R(0) + \alpha'_R t, \quad (57)$$

where $\alpha_R(0)$ (intercept) and α'_R (slope) are some numbers.

The integer values of $\text{Re } \alpha_R(t)$ with $t > 0$ (non-physical region) correspond to the exchange in t -channel of a real particle with mass $M^2 = t$. The Regge trajectory is usually denoted by the lowest mass hadron lying on it. The only exception to this rule is the Pomeron, which is the vacuum Regge trajectory with the highest intercept.

Regge trajectories have a special quantum number, called signature. The t -channel exchange of a Reggeon with positive signature ($\Theta = +1$) corresponds to the exchange of particles having even spin, e.g. f and a_2 , whereas the exchange of a

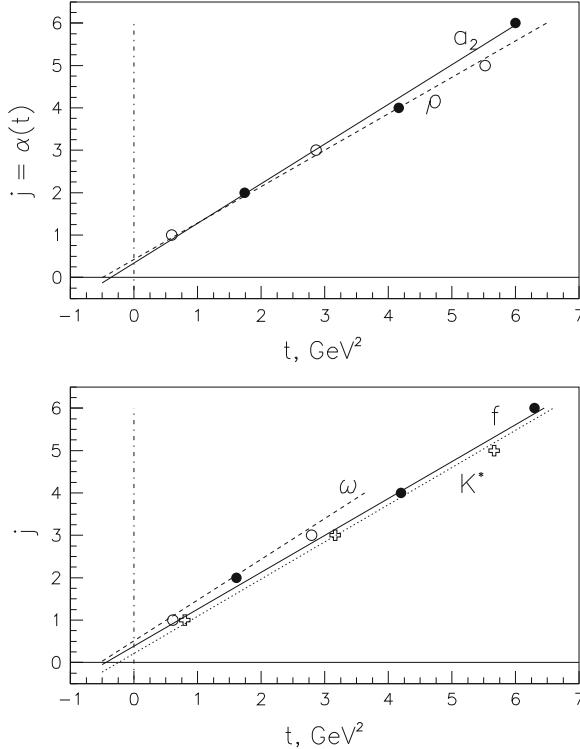


Fig. 18 The Chew-Frautschi plot of $\text{Re}\alpha(t)$ versus t for the well established mesons with relatively large intercept. *Upper panel* particles lying on the a_2 ($\Theta = +1$, solid line and closed points), and on the ρ ($\Theta = -1$, dashed line and open points) trajectories. *Lower panel* particles lying on the f ($\Theta = +1$, solid line and closed points), on the ω ($\Theta = -1$, dashed line and open points), and on the K^* ($\Theta = -1$, dashed line and open crosses) trajectories

Reggeon with negative signature ($\Theta = -1$) corresponds to the exchange of odd spin particles, e.g. ρ and ω particles (see Fig. 18).

At high energies, the most important exchange is the Pomeron exchange. Unfortunately, until now the particles lying on the Pomeron trajectory are unknown.

The experimental data show that all meson trajectories are approximately linear, and that they have approximately equal slopes $\alpha' \approx 0.8\text{--}1 \text{ GeV}^{-2}$, while the Pomeron slope, α'_P , is significantly smaller. The mass of the lightest particle that could lie on the Pomeron trajectory may be estimated with the help of α'_P . Its quantum numbers should be $I^G = 0^+$, $J^P = 2^+$, and its mass should be of about 2.7 GeV. Such a heavy hadron may probably be associated with a glueball state.

Baryon trajectories also exist, but their contributions have been studied in much less detail.

All well established hadrons lie on Reggeon trajectories at $t > 0$.

5.2 Regge Pole Exchange

One Regge pole exchange corresponds to the set of ladder diagrams where we sum over all possible numbers of rungs (see Fig. 19);

The contribution of a single t -channel Regge pole R to the amplitude of the process in Fig. 19 is given by:

$$A^R(s, t) = \gamma^R(t) \cdot \eta(\alpha_R(t)) \cdot \left(\frac{s}{s_0} \right)^{\alpha_R(t)}. \quad (58)$$

Here

$$\gamma^R(t) = \gamma_0^R \cdot e^{r_R^2 \cdot t}, \quad (59)$$

is the coupling of the Regge-pole, that has to be taken from experimental data, and

$$\eta(\alpha_R(t)) = \frac{1 + \Theta_R \cdot e^{-i\pi\alpha_R(t)}}{\sin \pi\alpha_R(t)}, \quad (60)$$

is the signature factor, with $\Theta_R = \pm 1$ being the signature, i.e. the quantum number accounting for the type of the Reggeon. Thus,

$$\eta(\Theta) = \begin{cases} i - \tan^{-1}(\frac{\pi\alpha_R}{2}), & \Theta = +1 \\ i + \tan(\frac{\pi\alpha_R}{2}), & \Theta = -1 \end{cases}. \quad (61)$$

Let us consider the most essential properties of the Reggeon exchange amplitude. First, the only energy dependence of $A^R(s, t)$ comes just from the factor $(s/s_0)^{\alpha_R(t)}$. Second, $A^R(s, t)$ is even (odd) under the crossing transformation $s \leftrightarrow u$ when the signature $\Theta_R = +1(-1)$. Third, $\gamma_R(t)$ can be factorized as follows (factorization property):

Fig. 19 Ladder diagram corresponding to the exchange of one Regge pole

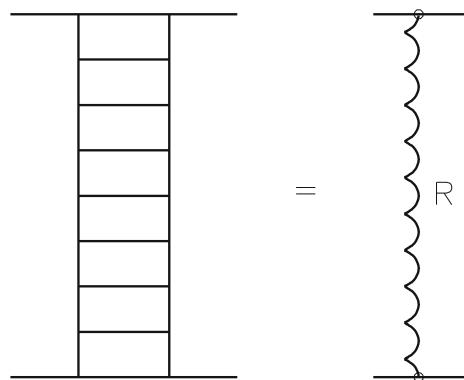


Table 2 Regge poles which determine the elastic hp collisions at high energies

Process	Regge singularities
$pp \rightarrow pp$	$A_P + A_f - A_\omega - A_\rho + A_{a_2}$
$\bar{p}p \rightarrow \bar{p}p$	$A_P + A_f + A_\omega + A_\rho + A_{a_2}$
$\pi^\pm p \rightarrow \pi^\pm p$	$A_P + A_f \pm A_\rho$
$K^\pm p \rightarrow K^\pm p$	$A_P + A_f \pm A_\omega \pm A_\rho + A_{a_2}$

$$\gamma_R(t) = g_R^{13}(t) \cdot g_R^{24}(t) . \quad (62)$$

The contribution of different Reggeon exchanges to the elastic scattering hp amplitudes are presented in Table 2.

The contributions of some trajectories cancel in linear combinations of the total cross sections, and the contribution of a single Reggeon can be isolated:

$$\Delta\sigma(\pi^\pm p) = \sigma_{tot}(\pi^+ p) - \sigma_{tot}(\pi^- p) = 2\gamma_\rho \left(\frac{s}{s_0} \right)^{\alpha_\rho(0)-1} . \quad (63)$$

Thus, the values of the parameters of the Reggeon intercepts $\alpha_R(0)$ and of the couplings γ_R for non-vacuum trajectories (ρ , ω , a_2 , etc.) can be obtained by fitting experimental data on cross section differences.

The differences $\sigma_{\bar{p}p}^{tot} - \sigma_{pp}^{tot}$, $\sigma_{\pi^- p}^{tot} - \sigma_{\pi^+ p}^{tot}$, and $\sigma_{K^- p}^{tot} - \sigma_{K^+ p}^{tot}$, experimentally measured at $p_{lab} > 35 \text{ GeV/c}$, i.e. $\sqrt{s} > 8 \text{ GeV}$, are presented in Fig. 20.

One can see in Fig. 20 that the differences of the presented cross sections decrease with the energy as powers of s . This means that in the high energy region, the Pomeranchuk theorem stating that the total cross sections of particles and antiparticles are equal at asymptotically high energies, is approximately fulfilled.

The differential cross section of the elastic hN scattering can be parametrized in the form

$$\frac{d\sigma}{dt} = \frac{\sigma_{tot}^2}{16\pi} (1 + \rho^2) e^{b(s)t} , \quad (64)$$

where ρ is the ratio of the real and imaginary parts of the amplitude, and the slope can be written as

$$b(s) = a_0 + 2\alpha'_{eff} \ln s/s_0 , \quad (65)$$

so the slope parameter logarithmically increases with the energy.

5.3 Diffractive Dissociation in hN Collisions

At high energies, a new class of processes, the diffractive dissociation of one or both colliding hadrons, appears (see Fig. 21).

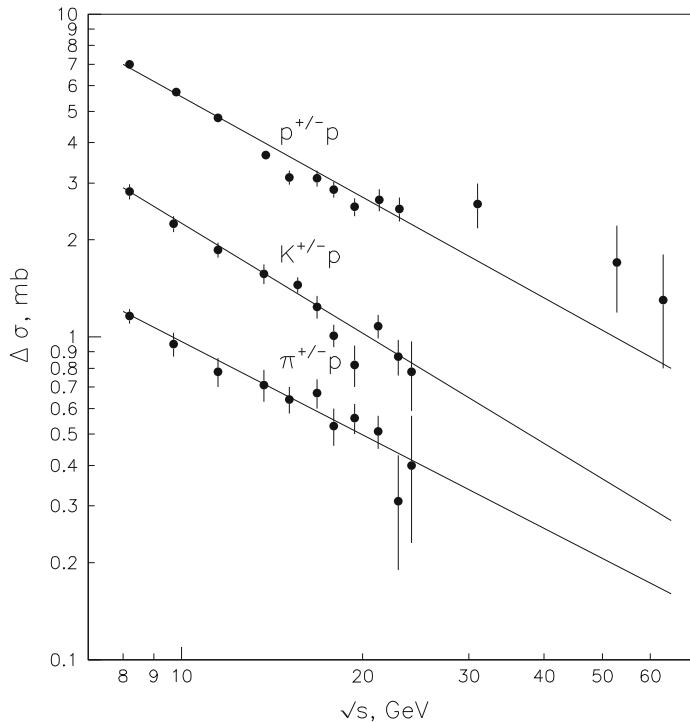


Fig. 20 The differences of the total cross sections of $\bar{p}p - pp$, $\pi^- p - \pi^+ p$, and $K^- p - K^+ p$ collisions

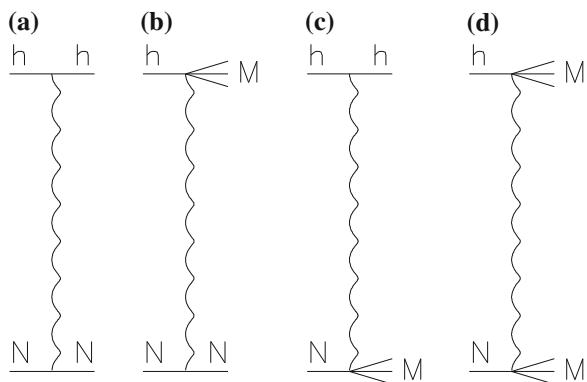


Fig. 21 Regge-pole diagrams for **a** elastic hN scattering, **b** and **c** single diffractive dissociation of one colliding particle, and **d** double diffractive dissociation

In the single (Fig. 21b, c) diffractive processes one colliding particle produces a diffractive jet while the other colliding particle is conserved, while in a double (Fig. 21d) diffractive dissociation both colliding particles produce jets. The integrated cross sections of such a processes are not large, but at high energies they generate new effects in inelastic collisions.

The contribution of the Pomeron exchange to the diffractive dissociation cross section is large, so these processes can exist even at asymptotically high energies.

The cross sections of single diffractive production, σ_D , at high energies can be estimated from the existing experimental data. For the process of Fig. 21b the ratios $R_D = \sigma_D/\sigma^{el}$ were found to be $R_D^{(p)} = 0.16 \pm 0.04$, $R_D^{(\pi)} = 0.37 \pm 0.12$, and $R_D^{(K)} = 0.42 \pm 0.2$ [10, 11], respectively.

The structure of the diffractive jet with mass M is such that at small M it is saturated by resonances.

One important feature of the single diffraction dissociation processes is the existence of a minimal momentum transferred of the recoil particle, say a target nucleon in Fig. 19b. To calculate this, let us find the t value at zero transverse momentum ($p_{3T} = 0$) for the case $m_2 = m_4$ in the center-of-mass system,

$$t = (p_1 - p_3)^2 = (E_1 - E_3)^2 - (p_{1L} - p_{3L})^2 . \quad (66)$$

Now, we use:

$$E_1 = \frac{s + m_1^2 - m_2^2}{2\sqrt{s}} , \quad E_3 = \frac{s + m_3^2 - m_4^2}{2\sqrt{s}} , \quad (67)$$

and, for pure longitudinal momenta,

$$p_{iL} = \sqrt{E_i^2 - m_i^2} = E_i \sqrt{1 - m_i^2/E_i^2} . \quad (68)$$

Let us expand these equations up to the second correction term, $\sqrt{1-x} = 1 - x/2 - x^2/8$. Assuming $p_{3T} = 0$ as corresponding to the minimal value of t , we obtain

$$t_{min} = -m_2^2 \frac{(m_3^2 - m_1^2)^2}{s^2} , \quad M = m_3 , \quad (69)$$

from where, clearly $t_{min} = 0$ for the case of elastic scattering, $m_3 = m_1$.

The minimal longitudinal momentum p_{4L} of the recoil nucleon appears due to the transfer of longitudinal momentum q_z , $q_z^2 = -t_{min}$. At high energies and for $m_i^2 \ll M^2 \ll s$, $s \simeq 2m_2 \cdot E_{lab}$ in the laboratory system, and we obtain

$$q_z = p_{4L} = \frac{M^2 - m_1^2}{2E_{lab}} . \quad (70)$$

5.4 Reggeon Cuts, Planar and Non-planar Diagrams

In Regge-Gribov theory more complicate diagrams exist, first of all, Regge cuts that are the absorptive corrections to the Regge pole exchange. They arise from the exchange of two or more Reggeons.

However, the corresponding diagram in Fig. 22b, considered as a Feynman-diagram, gives zero contribution to the scattering amplitude. The reason for this comes from the space–time picture of strong interactions at high energy, where the growth of characteristic longitudinal distances is quite evident, what reflects a very general feature of soft hadronic interactions. Actually, in soft hadronic interactions each hadron is considered as surrounded by vacuum fluctuations, corresponding to the emission and absorption of virtual particles. Thus, a fast hadron with the momentum $k \gg m$ (with m a mass scale, e.g. $m \sim m_\rho$), can be imagined as a cloud of particles (partons) with different momenta. Then, the parton fluctuation can gather into a hadron, so that its Fock wave function approaches its asymptotic form. Obviously, this requires a time of the order of k/m^2 , k/m being the Lorentz factor and $1/m$ the characteristic time of the interaction. The second interaction with a target nucleon needs again a time of the order of k/m^2 , what seems impossible because at that time a fast particle will be already at a large distance from the target.

In the language of Reggeon diagrams, the simplest graph for the elastic scattering amplitude is the one-Pomeron exchange shown in Fig. 23a, which corresponds to a multiperipheral ladder of hadrons in the intermediate state, (see Fig. 23b). The crosses in the graph mark the hadrons on the mass shell.

The elastic cross section shown in Fig. 23c is one of the possible cuts of the diagram in Fig. 23d, which account for a process highly improbable at high energies, since it corresponds to a repeated interaction with a point-like target. Thus, the amplitude in Fig. 23d should actually vanish.

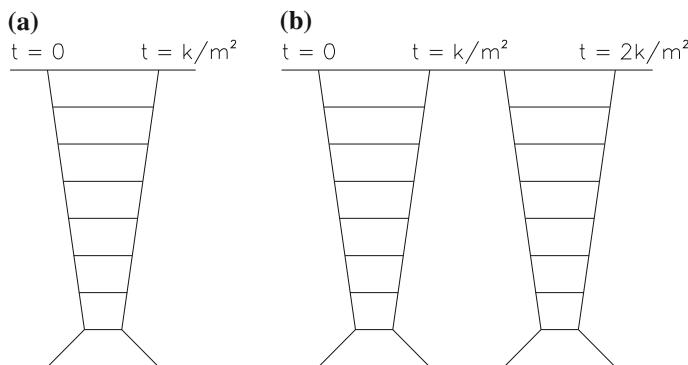


Fig. 22 Diagrams corresponding in the parton model to **a** the interaction of a fast hadron with one nucleon of the target, and **b** two successive interactions with two nucleons of the target

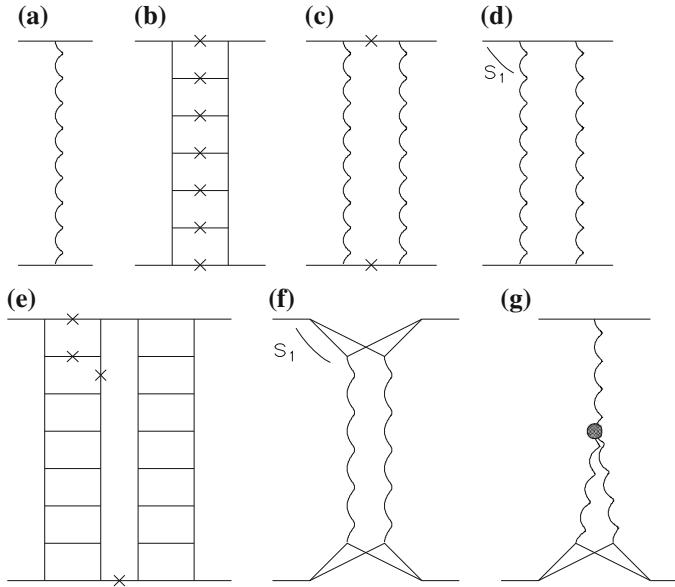


Fig. 23 **a** Pomeron exchange, and **b** its corresponding intermediate states. **c–e** Planar diagrams for the two-Pomeron cut, and **f–g** non-planar diagrams for the two-Pomeron cut

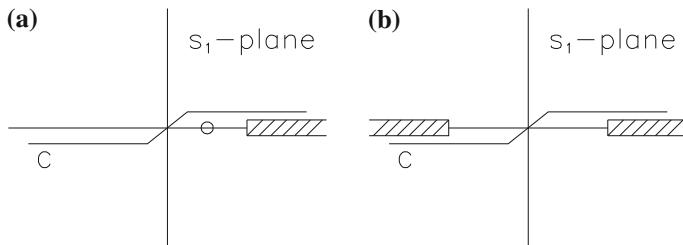


Fig. 24 The integration contour C in the s_1 -plane and the singularities of the diagrams in (a) Fig. 23d and in (b) Fig. 23f

A large imaginary part in the diagram of Fig. 23c is cancelled by the cuts shown in Fig. 23e. Non-zero corrections to the elastic amplitude come from the diagrams with the so-called Mandelstam crosses in Fig. 23f, where both Pomerons simultaneously interact.

The analytical properties of the diagrams in Fig. 23d, f are shown in Fig. 24. The amplitude in Fig. 23d only has a right discontinuity and a pole, both on the same side of the contour C , as shown in Fig. 24a. A positive contribution of the pole is cancelled by the negative contribution of the discontinuity. Contrary to this, the amplitude in Fig. 23f has both left and right discontinuities (see Fig. 24b), giving a non-zero contribution.

5.5 Reggeon Cuts Contribution in the Eikonal Approximation

Let us consider the case when the exchanges by one and two Reggeons are possible, as it is shown in Fig. 25a, b, c.

The eikonal approximation basically means to the diagram of Fig. 25d instead that of Fig. 25c. Since the particle with crosses in Fig. 25d are on mass-shell, we account for only one pole in Fig. 24a, by assuming that the negative contribution of a cut in Fig. 24a and a positive contribution of a cut in Fig. 24b cancel each other.

Then, after rather simple calculations we obtain:

$$\begin{aligned} A^{el}(s, q^2) &= A^{(1)}(s, q^2) + A^{(2)}(s, q^2), \\ A^{(1)}(s, q^2) &= A_1(q^2) + A_2(q^2), \\ A^{(2)}(s, q^2) &= \frac{i}{8\pi^2 s} \int d^2 q_1 A_1(q_1) \cdot A_2(q - q_1). \end{aligned} \quad (71)$$

This last integral is proportional to

$$\int d^2 q_1 e^{-a_1 q_1^2} \cdot e^{-a_2 (q - q_1)^2} = \frac{\pi}{a_1 + a_2} \cdot \exp \left[-\frac{a_1 \cdot a_2}{a_1 + a_2} q^2 \right]. \quad (72)$$

Finally, the amplitude $A^{el}(s, q^2)$ reads as

$$\begin{aligned} A^{el}(s, q^2) &= \gamma_0^{R_1} \eta_{R_1} \left(\frac{s}{s_0} \right)^{\alpha_{R_1}(0)} \cdot e^{-a_1 q^2} + \gamma_0^{R_2} \eta_{R_2} \left(\frac{s}{s_0} \right)^{\alpha_{R_2}(0)} \cdot e^{-a_2 q^2} \\ &+ \frac{i}{8\pi} \frac{\gamma_0^{R_1} \gamma_0^{R_2}}{(a_1 + a_2)} \eta_{R_1} \eta_{R_2} \left(\frac{s}{s_0} \right)^{\alpha_{R_1}(0) + \alpha_{R_2}(0) - 1} \cdot \exp \left[-\frac{a_1 \cdot a_2}{a_1 + a_2} q^2 \right]. \end{aligned} \quad (73)$$

In the case, important from the practical point of view, of a Pomeron-Pomeron cut, $\eta_P \simeq i$, and the double exchange contribution has a negative sign.

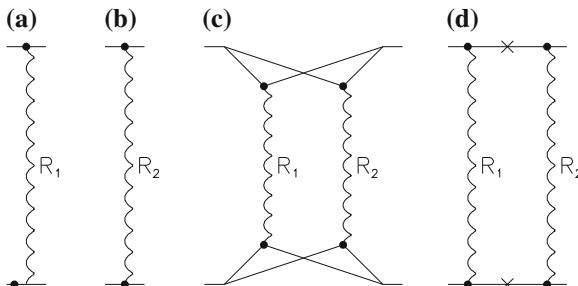


Fig. 25 Elastic scattering amplitudes with exchange of (a) and (b) one and (c) two Reggeons, and (d) the eikonal approximation of the diagrams (c)

The Pomeron trajectory,

$$\alpha_P(t) = 1 + \Delta + \alpha'_P t, \quad \Delta > 0, \quad (74)$$

leads to the one-Pomeron contribution to σ_{hN}^{tot} , equal to

$$\sigma_P = \gamma_0^P \cdot e^{\Delta \xi}, \quad \xi = \ln s/s_0, \quad (75)$$

where γ_0^P is the Pomeron coupling, and σ_P rises with energy as s^Δ . To obey the s -channel unitarity, and the Froissart bound in particular, this contribution should be screened by the multipomeron discontinuities. Then, an eikonal approximation yields

$$\sigma_{hN}^{tot} = \sigma_P \cdot f(z/2), \quad (76)$$

where

$$\begin{aligned} f(z) &= \sum_{k=1}^{\infty} \frac{1}{k \cdot k!} (-z)^{k-1} = \frac{1}{z} \int_0^z \frac{dx}{x} (1 - e^{-x}), \\ z &= \frac{\gamma}{4\pi\alpha_P} \cdot e^{\Delta \xi}, \quad a_P = R^2 + \alpha'_P \cdot \xi. \end{aligned} \quad (77)$$

Here, R_P^2 is the radius of the Pomeron. At asymptotically high energies ($z \gg 1$), we finally obtain

$$\sigma_{hN}^{tot} = 8\pi\alpha'_P \cdot \Delta \cdot \xi^2, \quad (78)$$

that complies with the Froissart limit.

5.6 Inclusive Reactions

In inelastic high energy collisions many secondary particles can be produced. The process $a + b \rightarrow h_1 + h_2 + \dots + h_n$, where all the secondaries are registered, is called an exclusive reaction, though more usually, only one secondary particle h is registered, and such processes

$$a + b \rightarrow h + X, \quad (79)$$

are called inclusive reactions, X denoting all secondaries but h .

The inclusive cross section is defined as

$$F(s, p_L, p_T) = 2E \frac{d^3\sigma_h}{d^3p} . \quad (80)$$

The most important variable is Feynman- x_F :

$$x_F = p^h/p_{max}^h , \quad (81)$$

where p^h is the momentum of secondary h , and p_{max}^h is the kinematically maximum momentum of p^h , that at high energies is $p_{max}^h \simeq p_a$.

To consider the production of secondaries in the central region it turns out to be very convenient to use the variable

$$y = \frac{1}{2} \ln \frac{E + p_L}{E - p_L} = \ln \frac{E + p_L}{m_T} \simeq \ln \frac{2x_F E_a}{m_T} , \quad (82)$$

called rapidity. The rapidity transforms additively under a Lorentz boost with a velocity $\beta = v/c$ along the beam axis, that is:

$$y' = y - \frac{1}{2} \ln \frac{1 + \beta}{1 - \beta} . \quad (83)$$

In some high energy experiments only the angular distributions of the secondaries are measured, and here the so-called pseudorapidity is used:

$$\eta = -\ln \tan(\theta/2) \simeq \ln \frac{2x_F E_a}{p_T} , \quad (84)$$

η being $\eta \simeq y$ for $y \geq 2 - 3$. The following exact equality holds:

$$\frac{d\sigma(ab \rightarrow hX)}{dy} = x_E \cdot \frac{d\sigma(ab \rightarrow hX)}{dx_F} , \quad (85)$$

where $x_E = E_h/E_a$, E_a being the energy of the incident hadron a . Also several sum rules can be established:

$$\begin{aligned} & \int dx_F \frac{x_E}{\sigma_{ab}^{in}} \cdot \frac{d\sigma(ab \rightarrow hX)}{dx_F} = \langle K_h \rangle , \\ & \sum_h \int dx_F x_E \cdot \frac{d\sigma(ab \rightarrow hX)}{dx_F} = \sigma_{ab}^{in} , \end{aligned} \quad (86)$$

if the elastic $a + b \rightarrow a + b$ channel is not included, and, finally,

$$\int \frac{dy}{\sigma_{ab}^{in}} \cdot \frac{d\sigma_h}{dy} = \langle n_h \rangle , \quad (87)$$

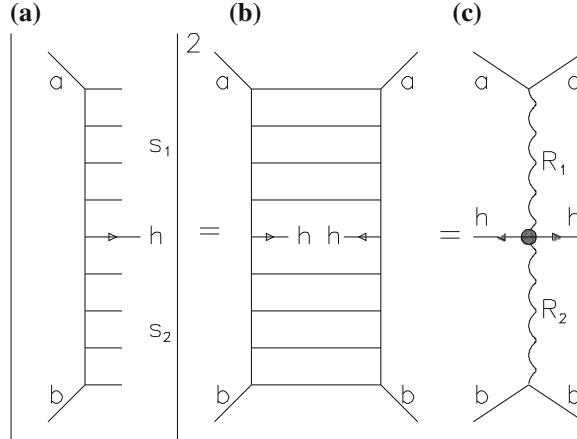


Fig. 26 **a** and **b** Squared modulo of the amplitude of inclusive production in the central region, and **c** the Regge-pole diagram corresponding to this process

where K_h is the part of the initial energy carried away by the secondary hadron h , $\sum \langle K_h \rangle = 1$, and $\langle n_h \rangle$ is the averaged multiplicity of h .

At high energies, the multiperipheral production processes give the main contribution to the inelastic cross section that leads to the flat rapidity distribution in the central region. At small x , the x -dependence of $F(x, p_T)$ is negligible, the behavior of $F(x, p_T)$ in this region being determined by the multiperipheral diagram Fig. 26.

The inclusive production in the central region is determined as [12, 13]:

$$F(x, p_T) = \frac{1}{\pi^2 s} g_{R_1}^{aa}(0) \cdot g_{R_2}^{bb}(0) \cdot g_{R_1 R_2}^{hh}(p_T) \left(\frac{s_1}{s_0} \right)^{\alpha_{R_1}(0)} \cdot \left(\frac{s_2}{s_0} \right)^{\alpha_{R_2}(0)}, \quad (88)$$

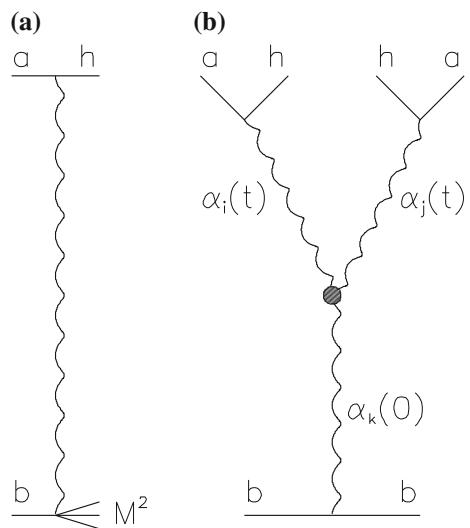
$$\begin{aligned} s_1 &= (p_a + p_h)^2 = m_T \sqrt{s} \cdot e^{-y^*}, \\ s_2 &= (p_b + p_h)^2 = m_T \sqrt{s} \cdot e^{y^*}. \end{aligned} \quad (89)$$

In agreement with the AGK cutting rules, the diagram with a one-Pomeron exchange is the only one contributing to the multiplicity in the central region. Contributions of Regge cuts cancel each other.

The x limit, close to unity, corresponds to the production of a diffractive beam with squared mass $M^2 \ll s$ (see Fig. 27a). Now, $x = 1 - M^2/s$, and so in the case of $M^2 \gg m^2$ the squared modulo of the graph in Fig. 27a leads to the triple-Reggeon limit shown in Fig. 27b.

The triple-Pomeron term rapidly increases the inclusive spectrum at $x \rightarrow 1$.

Fig. 27 **a** Secondary h production at $m_N^2 \ll M^2 \ll s$, and **b** the squared modulo of this amplitude, that corresponds to the triple-Reggeon diagram



6 Hadron-Nuclear Interactions

6.1 Nuclear Shapes

The atomic nucleus is usually seen as a sphere which contains nucleons: protons and neutrons. Light nuclei approximately contain equal numbers of protons and neutrons, whereas the number of neutrons in heavy nuclei is about 1.5 times larger than the number of protons, mainly due to the Coulomb repulsion contribution to the total binding energy.

The binding energy of the nucleons in the nuclei (except in the lightest ones) is of about 8 MeV per nucleon, i.e. two orders of magnitude smaller than the nucleon masses. Thus, the Fermi motion of the nucleon is non-relativistic, and the high-momentum tail of the nucleon motion is relevant just in very few cases.

Nuclear matter is not uniformly distributed, even in spherical nuclei, its radial distribution being a real function of the radial variable r , the distance from the center of the nucleus.

The effective nuclear radius, R_A , can be described as:

$$R_A = R_0 A^{1/3}, \quad R_0 = 1.1 - 1.3 \text{ fm} \quad (90)$$

The elastic high energy eA scattering can be used for the investigation of the nuclear structure. So, if q is the transferred momentum from the electron to the nucleus via one-photon exchange, R_Z is the average electric radius of the nucleus, and $q \cdot R_Z \gg 1$, i.e. the photon can interact with only one proton from the nucleus, the matrix element of such a scattering can be written as

$$M \propto \int \psi_Z(p) \cdot \psi_Z^*(p+q) d^3 p , \quad (91)$$

where $\psi_Z(p)$ and $\psi_Z^*(p+q)$ are the wave functions of the scattered proton, before and after the scattering, respectively.

Then, by taking into account that

$$\psi_Z(p) = \int \psi_Z(r) \cdot e^{ipr} d^3 r , \quad \psi_Z^*(p+q) = \int \psi_Z^*(r) \cdot e^{-i(p+q)r} d^3 r , \quad (92)$$

one gets

$$M \sim \int |\psi_Z(r)|^2 \cdot e^{-iqr} d^3 q . \quad (93)$$

The expression for a charged one-particle nuclear form factor reads as

$$G_Z(q) = \int \rho_Z(r) \cdot e^{-iqr} d^3 r , \quad G_Z(0) = 1 , \quad (94)$$

where the charged nuclear density distribution is $\rho_Z(r) = |\psi_Z(r)|^2$.

From measurements of the elastic eA scattering at different q , we can obtain information about the charged nuclear density distribution $\rho_Z(r)$.

The first detailed information on charged nuclear density distributions came from the experiments by Hofstädter [14] on high energy eA scattering. Different models for charged density distributions were used for the analysis of these experimental data, including the point-like model, $\rho(r) = \delta(r)$, the uniform model, the Gaussian model, the exponential model, etc. From the beginning, it was clear that the point-like model could be excluded, the nuclei being best characterized by form factors.

A reasonable fit of the nuclear form factors can be obtained by using the Fermi (Woods-Saxon [15]) distribution:

$$\rho_A(r) = \frac{\rho_1}{1 + e^{(r-c)/a}} , \quad c \gg a . \quad (95)$$

Here, ρ_1 is the normalization constant, c is a parameter measuring the nuclear size, and a is related to the diffuseness of the surface.

The Woods-Saxon distribution is shown in more detail in Fig. 28. The parameter c shows the value of r at which $\rho(r)$ has decreased by a factor 1/2, when compared to $\rho(r=0)$, $\rho(r=c) = 0.5 \cdot \rho(r=0)$. The value of a determines the distance $a_1 = 4a \ln 3 \sim 4.4a$, at which $\rho(r)$ has decreased from $0.9 \cdot \rho(r=0)$ to $0.1 \cdot \rho(r=0)$.

One also defines

$$\langle R_A^2 \rangle = 4\pi \int_0^\infty \rho(r) \cdot r^4 dr , \quad (96)$$

or its square root.

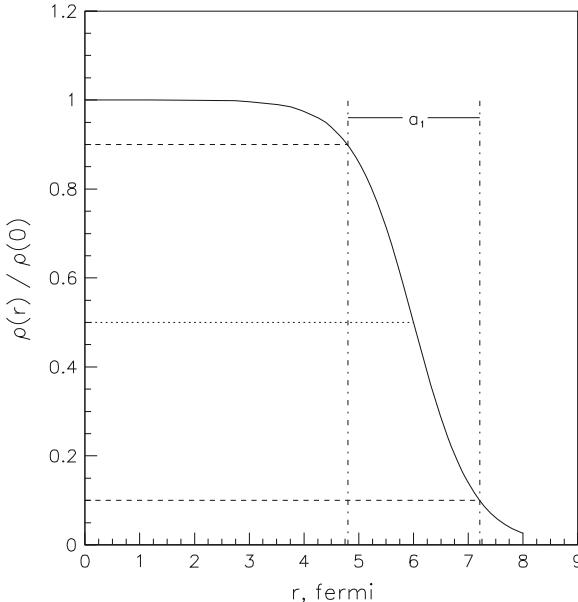


Fig. 28 Woods-Saxon distribution in (6.6), with $c = 6$ fm, and $a_1 = 4a \ln 3$, with $a = 0.55$ fm

In the case of light nuclei (${}^9\text{Be}$, ${}^{12}\text{C}$, ${}^{16}\text{O}$), a Gaussian parametrization can also be used:

$$\begin{aligned} \rho_A(r) &= (\alpha^2/\pi)^{3/2} \cdot e^{-\alpha^2 r^2}, \\ 1/\alpha^2 &= \frac{2}{3} \langle R_A^2 \rangle. \end{aligned} \quad (97)$$

6.2 Main Assumptions in Glauber Theory [16]

The case of hadron-nucleus interactions is more complicated than that of eA scattering, due to the possibility of multiple interactions. A good formalism to describe the hadron-nucleus interaction is the Glauber theory, that has its origin in quantum mechanics.

The basis of the Glauber theory is the eikonal approximation for the scattering of a fast particle. Let a fast particle with momentum k and kinetic energy T in laboratory frame, scatter on the nucleus A , that we will initially treat as a collection of potential wells of size a and depth V_0 . Provided the conditions

$$ka \gg 1, \quad T/V_0 \gg 1 \quad (98)$$

are satisfied, the characteristic scattering angles are small, and thus the phase shift $\chi_A(b)$ is given by the integral of the total nuclear potential:

$$\chi_A(b) = -\frac{m}{k} \int_{-\infty}^{\infty} dz V_A(b, z) , \quad (99)$$

b being the impact parameter (2-dimensional vector). The constraints in (98) prevent the incident particle from interacting more than once with a given target nucleon, the target nucleons having no time for interacting with each other during the scattering process. The next assumption is that the nuclear potential V_A is built up from spatially separated potentials corresponding to each nucleon in the nucleus. Thus, the phase shift on the nucleus, $\chi_A(b)$, is equal to the sum of the phase shifts $\chi_N(b_i)$ from each hadron-nucleon scattering:

$$\chi_A(b) = \sum_{i=1}^A \chi_N(b_i) . \quad (100)$$

The phases $\chi_A(b)$ and $\chi_N(b_i)$ are directly related to the elastic scattering amplitudes on the nucleus, F_{hA}^{el} , and on the isolated nucleon, F_{hN}^{el} , respectively:

$$\begin{aligned} F^{el}(q) &= \frac{ik}{2\pi} \int d^2 b \Gamma(b) \cdot e^{iqb} , \\ \Gamma_N(b_j) &= 1 - e^{i\chi_N(b_j)} , \\ \Gamma_A(b; r_1, \dots, r_A) &= 1 - e^{i\chi_A(b; r_1, \dots, r_A)} = 1 - e^{i \sum_{j=1}^A \chi_j(b - b_j)} , \end{aligned} \quad (101)$$

where r_1, \dots, r_A are the positions of the nucleons, and b_j are their transverse coordinates. Only $A - 1$ of the nucleon coordinates are independent, since $\sum_{j=1}^A r_j = 0$, what leads to the so-called center-of-mass motion correction.

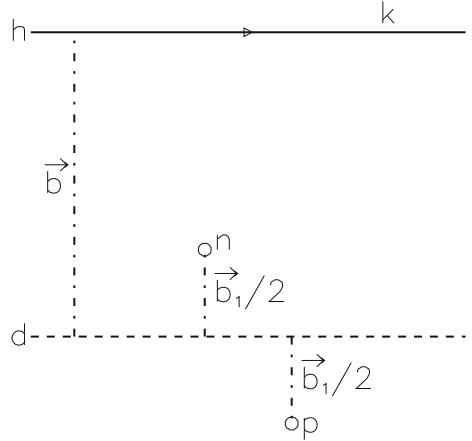
The amplitude F^{el} is defined here in the laboratory system. The optical theorem for F^{el} reads:

$$Im F^{el}(s, 0) = \frac{k}{4\pi} \sigma^{tot} . \quad (102)$$

6.3 Elastic Scattering on a Nuclear Target

Let us consider the interaction with a deuteron target (see Fig. 29), and let r_p and r_n be the coordinates of the proton and the neutron, respectively, with $\mathbf{r} = \mathbf{r}_p - \mathbf{r}_n$, and with $b_1/2$ and $-b_1/2$ their transverse coordinates. The elastic scattering hd amplitude can be written in Glauber theory as:

Fig. 29 Elastic scattering of a fast hadron with momentum k on a deuteron target (vectors are all two-dimensional)



$$F_{hd}^{el}(q) = \frac{ik}{2\pi} \int d^2b e^{iqb} \int \psi_f^*(r) \times \left\{ 1 - e^{[i\chi_p(b+b_1/2) + i\chi_n(b-b_1/2)]} \right\} \cdot \psi_i(r) d^3r . \quad (103)$$

By substituting Γ_N by F_{hN}^{el} ($\equiv F_p$ or F_n), one gets

$$\begin{aligned} F_{hd}^{el} &= \langle f | e^{iqb_1/2} \cdot F_p(q) + e^{-iqb_1/2} \cdot F_n(q) \\ &\quad + \frac{i}{2\pi k} \int d^2q' e^{ib_1 q'} \cdot F_p(q' + q/2) \cdot F_n(q' - q/2) | i \rangle \\ &= G_d(q/2) \cdot F_p(q) + G_d(q/2) \cdot F_n(q) \\ &\quad + \frac{i}{2\pi k} \int G_d(q') \cdot F_p(q/2 - q') \cdot F_n(q/2 + q') d^2q' , \end{aligned} \quad (104)$$

where

$$G_d(q) = \int e^{iqr} \cdot |\psi(r)|^2 d^3r \quad (105)$$

is the deuteron form factor.

From the optical theorem it follows that

$$\sigma_{hd}^{tot} = \sigma_{hp}^{tot} + \sigma_{hn}^{tot} - \Delta_2 , \quad (106)$$

with

$$\Delta_2 = \frac{2}{k^2} \int S_d(q) \cdot \text{Im}F_p(q) \cdot \text{Im}F_n(q) \cdot [1 - \rho_p(q) \cdot \rho_n(q)] d^2q . \quad (107)$$

Here, Δ_2 is the Glauber correction for double scattering, and

$$\rho_N(q) = \frac{ReF_N(q)}{ImF_N(q)} . \quad (108)$$

Let us now consider the elastic hA scattering, for $A \gg 1$:

$$F_{hA}^{el}(q) = \frac{ik}{2\pi} \int d^2 b e^{iqb} \langle A | \Gamma_A(b; r_1, \dots, r_A) | A \rangle . \quad (109)$$

By using the expressions for Γ_A and Γ_N in (101), and the basic assumption of (100), and by carrying out a final transition from the phase back to the amplitudes, we obtain

$$F_{hA}^{el}(q) = \frac{ik}{2\pi} \int d^2 b e^{iqb} \int d^3 r_1 \dots d^3 r_A \rho(r_1, \dots, r_A) \times \prod_{i=1}^A \left[1 - \frac{1}{2\pi ik} \int d^2 q_1 e^{-iq_1(b-b_i)} \cdot F_{hN}(q_1) \right] , \quad (110)$$

where b_i is the transverse coordinate of the i -th nucleon, and $\rho(r_1, \dots, r_A)$ is the probability distribution to find the nucleons in positions r_i , that can be reduced to the product of one-particle densities, $\rho(r_i)$:

$$\rho(r_1, \dots, r_A) = \prod_{i=1}^A \rho(r_i) , \int d^3 r_i \rho(r_i) = 1 . \quad (111)$$

After integrating over the positions of the nucleons we get

$$F_{hA}^{el}(q) = \frac{ik}{2\pi} \int d^2 b e^{iqb} \left[1 - \left(1 - \frac{1}{2\pi ik} \int d^2 q_1 e^{-iq_1 b} \cdot F_{hN}(q_1) \cdot G(q_1) \right)^A \right] , \quad (112)$$

with

$$G(q_1) = \int d^3 r_i \rho(r_i) \cdot e^{iq_1 r_i} , \quad (113)$$

is the one-particle nucleus form factor.

For purely imaginary F_{hN} at $q \simeq 0$, $ReF_{hN}/ImF_{hN} \ll 1$, and for $A \gg 1$, the optical theorem provides further simplification:

$$F_{hA}^{el}(q) = \frac{ik}{2\pi} \int d^2 b e^{iqb} \left[1 - e^{-1/2\sigma T(b)} \right] , \quad \sigma = \sigma_{hN}^{tot} . \quad (114)$$

In particular,

$$\sigma_{hA}^{tot} = 2 \int d^2 b \cdot \left[1 - e^{-(1/2)\sigma T(b)} \right], \quad (115)$$

$$\sigma_{hA}^{el} = \int d^2 b \cdot \left[1 - e^{-(1/2)\sigma T(b)} \right]^2 \quad (116)$$

and

$$\sigma_{hA}^{inel} = \int d^2 b \cdot \left[1 - e^{-\sigma T(b)} \right]. \quad (117)$$

6.4 Experimental Proton and Neutron Distributions [17]

Contrary to the case of eA scattering, hadron-nucleus collisions give information about the distribution not of the nuclear charge, but of nuclear matter. By comparing the eA and hA experimental data, one can obtain separated distributions of protons and neutrons in nuclei.

A lot of experimental data exist on proton-nucleus elastic scattering. Some examples of the available experimental data and of the corresponding theoretical descriptions are shown in Fig. 30.

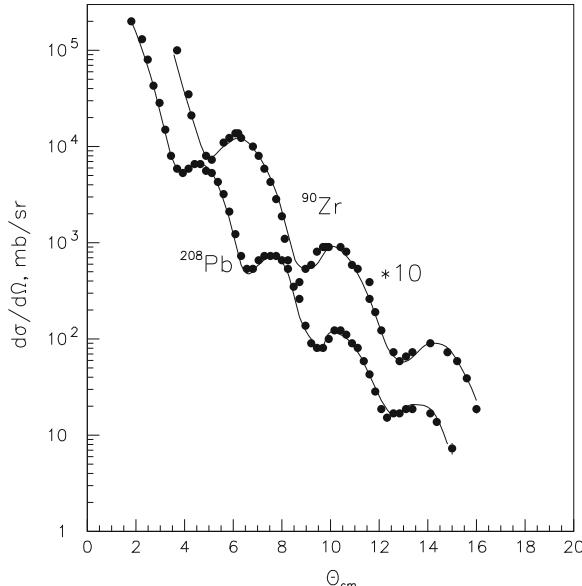


Fig. 30 Differential cross sections for 1 GeV proton scattering on ^{90}Zr and ^{208}Pb nuclei together with the corresponding description by Glauber theory

Table 3 Parameters of the proton and neutron folded densities, with r_p and r_n the r.m.s. radii (in fm) of the proton and neutron distributions, respectively

Nucleus	r_p	r_n	$r_n - r_p$
^{32}S	3.24	3.24	0.00
^{40}Ca	3.48	3.50	0.02
^{48}Ca	3.47	3.66	0.19
^{48}Ti	3.58	3.68	0.10
^{58}Ni	3.76	3.72	-0.04
^{64}Ni	3.84	3.83	-0.01
^{90}Zr	4.26	4.37	0.11
^{208}Pb	5.50	5.50	0.00

The distributions of protons (i.e. electrical charge distributions) are taken from electron scattering. The neutron (i.e. strong interacting but electrically neutral particles) are found by the common fitting of both theoretical eA and pA cross sections to the corresponding experimental data.

The main conclusion of this analysis is that the difference in the radii of proton and neutron distributions is very small. In Table 3 we present some numerical values.

In Table 3 one sees that the difference in the r.m.s. radii $r_n - r_p$ is rather small, with the only exceptions of the cases ^{40}Ca and ^{48}Ca , due to the presence of eight additional neutrons in ^{48}Ca , and in the case of ^{90}Zr .

7 Possible High Energy Physics Applications

7.1 Earth Survey and Research by Neutrino Beams [18]

Low energy neutrinos have a very small interaction cross section, so they can cross the Earth without interacting. This cross section linearly increases with the initial energy, E_ν . In the laboratory system:

$$\sigma(\nu N)/E_\nu \sim 10^{-11} \text{ mb/GeV} , \quad (118)$$

and for the energies of the order of hundreds GeV it gives a significant probability of interactions on a penetration length inside the Earth of several thousands kilometers. In particular, muon neutrinos can be produced by an accelerator (block A in Fig. 31), via the decay of secondary charged pions $\pi \rightarrow \mu + \nu_\mu$ produced by collisions on any target.

When heavy metal (copper, silver, gold, lead, etc.) are placed on the neutrino beam trajectory, one can appreciate an increase of muon production from the charged current process $\nu_\mu + N \rightarrow \mu^- + X$, since the average densities of such a heavy metals are significantly larger than the average density of the Earth's crust. Also, when the neutrino beam would encounter an underground oil-field a specific acoustic signal

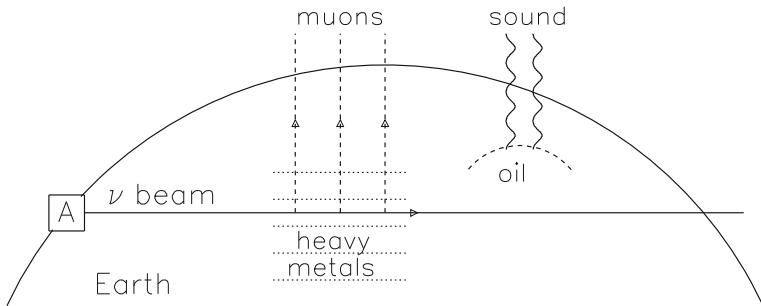


Fig. 31 General scheme for the Earth geological survey by neutrino beams

could be registered. Actually, by using a neutrino beam with a trajectory going near the Earth's center, one could investigate the general structure of the Earth, including its kernel.

7.2 Muon Catalyzed Fusion [19]

The nuclear fusion reaction of heavy hydrogen isotopes, deuteron d and tritium t,

$$d + t \rightarrow {}^4\text{He} + n + 17.6 \text{ MeV}, \quad (119)$$

usually needs of high temperatures due to the Coulomb repulsion between d and t. This repulsion starts being important when the distance between d and t becomes smaller than the range of the corresponding atoms, D and T. The atomic radii are proportional to the inverse value of the electron mass, m_e , so it is clear that if one would change one electron by one negative muon, that is 200 times more heavy ($m_\mu/m_e \sim 200$), the role of repulsion would significantly decrease. Estimations confirmed by the experimental data show that the reaction

$$d + t_\mu \rightarrow {}^4\text{He} + n + \mu^- + 17.6 \text{ MeV}, \quad (120)$$

can occur at normal temperatures and densities. A very important point to be noted is that the final muon in the reaction of (120) will stay free with a probability higher than 99 % (this state overlaps very weakly with the wave functions of ${}^4\text{He}_\mu$ atom in the momentum representation).

Afterwards, this muon can induce a second fusion reaction, and so on so forth, so it can be considered as the trigger of the process of (120), the muon mean life time being many times larger than the time for this processes to occur.

The neutrons obtained in the process of (120) can be directly used as a high intensity monoenergetical neutron source, say, for material science, or for fuel production through the reaction $n + {}^{238}\text{U} \rightarrow {}^{239}\text{Pu} + X$.

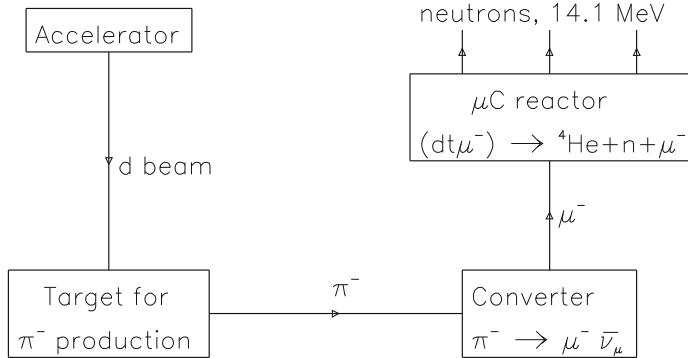


Fig. 32 General scheme for the high intensity production of 14.1 MeV neutrons via muon catalyzed fusion

The general scheme for the production of 14.1 MeV neutrons via muon catalyzed fusion presented in Fig. 32, goes like this. First, one uses, say, a deuteron beam from an accelerator, with kinetic energy of about 1 GeV per nucleon, for the production of π^- on the target. Deuteron beam is preferable because in reactions with neutrons negative pions are mainly produced. The kinetic energy of about 1 GeV per nucleon is close to the optimal value. Then, the produced negative pions decay into muons in the converter, where the special configuration of magnetic fields optimizes the losses of muons.

After that, the μ^- turn to a μ -catalyze reactor with a mixture of deuterium and tritium, where the 14.1 MeV neutrons are produced in the reaction of (120). These neutrons can be used in experiments, or to produce ^{239}Pu from ^{238}U , for the ulterior use of this plutonium in standard nuclear plants for electrical energy production.

7.3 Proton Therapy

Quite often, the standard surgical methods can not be used to treat serious medical problems, as aneurisma, swelling (cancer tumors and others), etc. In the particular

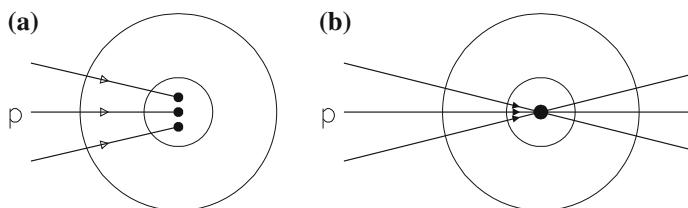


Fig. 33 General scheme of the use of high energy particle beams in medicine. A part of the patient's body is shown by *large circle*, and the affected region is shown by *small circle*

case of brain tumors, high energy particle beams, usually protons, can be used as it is schematically shown in Fig. 33.

Two possibilities exist. Either, the energy is taken in such a way that the beam will be stopped inside the affected region (see Fig. 33a), or in such a way that the beam will penetrate the whole body using different trajectories, but always with the crossing point inside the affected region (see Fig. 33b). Both methods show some advantages and some problems, and both are commonly used with very successful results.

8 Conclusion

Due to the limited extension of this text, we have only presented several topics concerning high energy hadron scattering. We have not considered other topics related to hadron physics as QCD, Standard Model, estatic hadron properties, heavy ion physics, etc., which will be included into a separate publication.

Acknowledgments We are grateful to N.I. Novikova for technical help. This work was supported by Ministerio de Economía y Competitividad of Spain (FPA2011–22776), the Spanish Consolider-Ingenio 2010 Programme CPAN (CSD2007-00042), by Xunta de Galicia, Spain (2011/PC043), and, partially, by grant RSGSS-3628.2008.2.

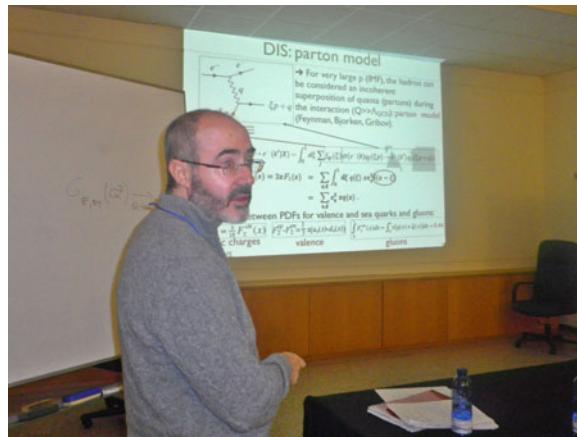
References

1. V.B. Berestetskii, E.M. Lifshitz, L.P. Pitaevskii, Quantum Electrodynamics. in *Landau and Lifshitz Course of Theoretical Physics*, 2nd edn, vol. 4 (Butterworth-Heinemann, London, 1982) (translated from Russian by J.B. Sykes and J.S. Bell). Reprinted in 1989, 1994, 1997, 1999, 2002
2. M. Froissart, Phys. Rev. **123**, 1053 (1961)
3. L.B. Okun, *Weak Interactions of Elementary Particles*. Monographs in Natural Philosophy (Pergamon Press, Oxford, 1965)
4. I.M. Frank, Izv. Acad. Nauk USSR, Ser. Fiz. **6**, 10 (1942)
5. E.L. Feinberg, ZhETF **50**, 202 (1966)
6. E.L. Feinberg, Sov. Phys. JETP **23**, 132 (1966)
7. P.D.B. Collins, E.J. Squires, *Regge Poles in Particle Physics*. Springer Tracts in Modern Physics, vol. 45 (Springer, Berlin, 1968)
8. P.D.B. Collins, *An Introduction to Regge Theory and High Energy Physics* (Cambridge University Press, Cambridge, 1977)
9. A.B. Kaidalov, *Regge Poles in QCD*. [arXiv:hep-ph/0103011](https://arxiv.org/abs/hep-ph/0103011)
10. A.B. Kaidalov, Yad. Fiz. **13**, 401 (1971)
11. A.B. Kaidalov, Nucl. Phys. **13**, 226 (1971)
12. V.A. Abramovsky, O.V. Kancheli, I.D. Mandzhavidze. Yad. Fiz. **13**, 1102 (1971)
13. V.A. Abramovsky, O.V. Kancheli, I.D. Mandzhavidze, Sov. J. Nucl. Phys. **13**, 630 (1971)
14. R. Hofstädter, Rev. Mod. Phys. **28**, 214 (1956)
15. R.D. Woods, D.S. Saxon, Phys. Rev. **95**, 577 (1954)
16. R.J. Glauber, In *Lectures in Theoretical Physics*, vol. 1, ed. by W.E. Brittin et al. (Interscience, New York, 1959), p. 315

17. G.D. Alkhazov, S.I. Belostotsky, A.A. Vorobyov, Phys. Rep. **42**, 89 (1978)
18. A. De Rujula, S.L. Glashow, R.R. Wilson, G. Charpak, Rept. **99**, 3341 (1983)
19. YuV Petrov, Nature **285**, 466 (1980)

General Bibliography

20. W.E. Burcham, M. Jobes, *Nuclear and Particle Physics* (Longman Scientific and Technical, Harlow, 1995)
21. V.V. Anisovich, M.N. Kobrinsky, J. Nyri, Yu.M. Shabelski, *Quark Model and High Energy Collisions* (World Scientific, Singapore, 2004)
22. S. Eidelman et al. Review of particle physics. Phys. Lett. **B592** (2004)
23. J.C. Polkinghorne, *Models of High Energy Processes*. Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, 1980)
24. J.H. Weiss, Regge theory and high energy hadron-nucleus scattering. Acta Phys Pol. **B7**, 851 (1976) and CERN-TH-2197
25. Efectos Sombra Anómalos en las Colisiones Hadrón-Núcleo y Núcleo-Núcleo, Ph.D. Thesis (in Spanish), by A. Vázquez Ramallo (advisor Prof. Carlos Pajares Vales), Universidade de Santiago de Compostela-USC, Spain (1982)
26. El Modelo Dual de Partones y la Búsqueda del Plasma de Quarks y Gluones, Ph.D. Thesis (in Spanish), by C. Merino (advisors Profs. Alphonse Capella and Carlos Pajares Vales), Universidade de Santiago de Compostela-USC (1991)



Professor Néstor Armesto during one of the lectures of his course (in collaboration with Professor Carlos Pajares) Quantum Chromodynamics at the Third IDPASC School (picture by C. Merino)

Quantum Chromodynamics

Néstor Armesto and Carlos Pajares

Abstract In this chapter we present some selected topics in the quantum field theory of the strong interaction, Quantum Chromodynamics. After an introduction on the gauge symmetry in Sect. 1, we briefly present the Renormalisation Group Equation in Sect. 2. Afterwards we turn to asymptotic freedom in Sect. 3 and confinement in Sect. 4. Then, after briefly reviewing the QCD colour factors in Sect. 5, we describe the symmetries of the theory in Sect. 6: isospin, chiral and scale symmetries, and the vacuum structure is reviewed with an introduction about instantons. In Sect. 7 the phase diagram is presented. Finally, in Sect. 8 some phenomenological applications of perturbative Quantum Chromodynamics are reviewed: e^+e^- annihilation, deep inelastic scattering, factorisation, radiation and jets.

1 Introduction

Quantum Chromodynamics (QCD) is the theory of the strong force describing the interactions of quarks forming hadrons, mesons if the interaction is between a quark and an antiquark and baryons if the interaction is among three quarks.

The interaction of quarks is via the exchange of gluons that carry a new quantum number, the colour. The colour was first introduced by Greenberg in order to restore the Pauli principle. The quarks have spin 1/2, three quarks cannot have the same quantum numbers. The baryons Δ^{++} , Δ^- and Ω^- are formed by three quarks uuu , ddd and respectively. Therefore, each quark must have an additional quantum number, the colour, arranged in such a way that each quark has one of three possible values, surviving the Pauli principle. Lately, the colour was introduced dynamically by Han and Nambu, in a non-abelian gauge theory where the gauge bosons are the gluons. The theory has a similar structure to Quantum Electrodynamics (QED), but

N. Armesto (✉) · C. Pajares

Departamento de Física de Partículas and IGFAE, Universidade de Santiago de Compostela,
15782 Santiago de Compostela, Galicia-Spain
e-mail: nestor.armesto@usc.es

C. Pajares

e-mail: pajares@fpaxp1.usc.es

now the group, $SU(3)$, is non abelian and hence the gluons are self-interacting. This fact is very important, implying asymptotic freedom [1, 2] at high energies and strong interactions at low energies. These interactions imply confinement, meaning that only colour-singlet states can propagate at large distances. The colour-singlet stable states are mesons ($q\bar{q}$) and baryons (qqq). There are excellent textbooks about QCD where you can find more details [3–9].

There are many similarities between QED and QCD. QED is a gauge theory, whose gauge group is $U(1)$, and the structure of the lagrangian is derived from this gauge symmetry. In fact, we start from the lagrangian of non-interacting matter

$$\mathcal{L}_0 = \bar{\psi}(i\cancel{d} - m)\psi, \quad (1)$$

with $\cancel{d} = a_\mu \gamma^\mu$ and γ^μ are the Dirac matrixes satisfying

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}, \quad (2)$$

with $g^{\mu\nu}$ the metric tensor. The lagrangian (1) is invariant under the global gauge transformation

$$\begin{aligned} \psi &\longrightarrow \psi' = \exp[iq\alpha]\psi, \\ \bar{\psi} &\longrightarrow \bar{\psi}' = \exp[-iq\alpha]\bar{\psi}. \end{aligned} \quad (3)$$

The QED lagrangian is derived asking how we must modify (1) to make it also invariant under local changes of gauge i.e. when α in (3) is a function $\alpha(x)$. Now, we have

$$\partial_\mu(\exp[iq\alpha(x)]\psi) = \exp[iq\alpha(x)](\partial_\mu\psi(x) + iq\partial_\mu\alpha(x)\psi(x)). \quad (4)$$

There is an extra term which breaks the invariance of the lagrangian (1).

In order to make (1) invariant, one may consider the addition of new terms to compensate the extra term. This can be done by introducing the gauge field A_μ , which transforms under the local change of gauge like

$$A_\mu \longrightarrow A'_\mu = A_\mu - \partial_\mu\alpha(x) = A_\mu + \frac{i}{q}\partial_\mu(\exp[iq\alpha(x)])\exp[-iq\alpha(x)] \quad (5)$$

and by replacing the derivative ∂_μ by the covariant derivative

$$D_\mu = \partial_\mu + iqA_\mu. \quad (6)$$

As we have introduced the gauge field, we must add to the lagrangian the corresponding kinetic term

$$\mathcal{L}_A = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (7)$$

where the field stress tensor is defined by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (8)$$

and the QED lagrangian is therefore given by

$$\mathcal{L}_{QED} = \bar{\psi}(iD - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (9)$$

which is now gauge invariant. Note that there is not any term $\frac{1}{2}m_\gamma^2 A^\mu A_\mu$ corresponding to the mass of the gauge field because it is not gauge invariant. The new lagrangian contains the free lagrangians and the interaction lagrangian:

$$\mathcal{L}_{QED} = \mathcal{L}_0 + \mathcal{L}_A - j^\mu A_\mu, \quad j^\mu = q\bar{\psi}\gamma^\mu\psi. \quad (10)$$

Therefore, the gauge symmetry determines the interaction. This interaction is illustrated in Fig. 1, where an electron (ψ) emits a photon (A_μ) and goes along its path.

QCD can be derived in the same way as QED [10]. We start from the lagrangian density for non-interacting quarks:

$$\mathcal{L}_q = \sum_q \bar{\psi}_q^j i\cancel{d}\psi_q^k - \sum_q m_q \bar{\psi}_q^j \psi_q^j, \quad (11)$$

where the indices $j, k = 1, 2, 3$ stand for the colour and $q = d, u, s, c, b, t$ denotes the flavour. The gauge group is $SU(N_c)$ with $N_c = 3$. Each element U of $SU(N_c)$ infinitesimally close to the identity can be written as

$$U = I + iG, \quad (12)$$

where G are hermitian traceless matrixes. One can choose a set of $N_c^2 - 1$ matrixes such that any G can be written as

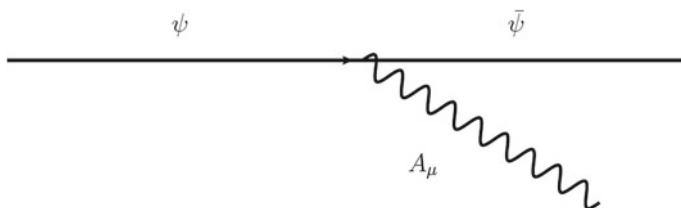


Fig. 1 Diagram of the interaction in QED

$$G = \sum_{a=1}^{N_c^2 - 1} \alpha_a t_a, \quad (13)$$

with α_a some coefficients. It can be shown that $[t_a, t_b]$ are antihermitian and traceless and, hence, they must be a linear combination of the basis t_c ,

$$[t_a, t_b] = i f_{abc} t_c, \quad (14)$$

where f_{abc} are real constant, antisymmetric in all their indices. Equation (14) defines the algebra of the group. The $(N_c^2 - 1) \times (N_c^2 - 1)$ matrixes T_a obeying

$$(T_a)_{bc} = -i f_{abc}, \quad (15)$$

$$[T_a, T_b] = i f_{abc} T_c, \quad (16)$$

define the adjoint representation of the group, while the t_a define the fundamental representation of the group.

Now, the transformation of the gauge fields reads

$$\begin{aligned} \psi_q(x) &\longrightarrow \psi'_q(x) = \exp [i g_s \alpha_a(x) T_a] \psi_q(x), \\ \bar{\psi}_q(x) &\longrightarrow \bar{\psi}'_q(x) = \exp [-i g_s \alpha_a(x) T_a] \bar{\psi}_q(x), \end{aligned} \quad (17)$$

where g_s is the coupling constant of the strong interaction.

In order to make the lagrangian (11) invariant, we introduce $N_c^2 - 1$ gauge fields (gluons) G_μ^a , transforming as

$$G_\mu^a \longrightarrow G_\mu^{a'} = G_\mu^a - \partial_\mu \alpha_a(x) - g_s f_{abc} \alpha_b(x) G_\mu^c, \quad (18)$$

and replace the derivative ∂_μ by the covariant derivative D_μ , with

$$D_\mu = \partial_\mu + i g_s T_a G_\mu^a. \quad (19)$$

Now, comparing with the QED (5), we have introduced an additional term required to cancel the additional term appearing when the transformation (17) is introduced in the free lagrangian (11), due to the non-abelian composition law of T_a .

We have to introduce again a kinetic term for the gauge fields

$$\mathcal{L}_G = -\frac{1}{4} F_a^{\mu\nu} F_{\mu\nu}^a, \quad (20)$$

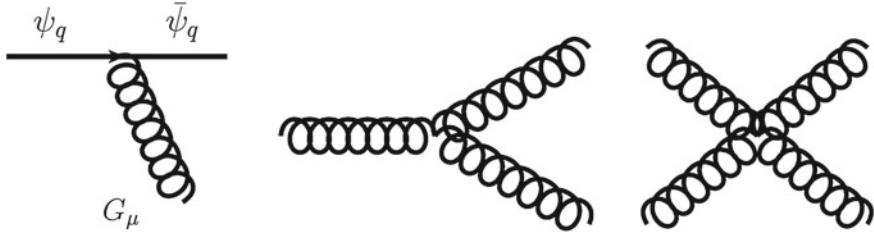


Fig. 2 Diagram of the interactions in QCD

where

$$F_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f_{abc} G_\mu^b G_\nu^c. \quad (21)$$

Note that there is an extra term in (21) comparing to QED, because otherwise (20) would not be gauge invariant.

We observe that in addition to the term $-j_a^\mu G_\mu^a$ with

$$j_a^\mu = g_s \bar{\psi}_q \gamma^\mu \psi_q T_a, \quad (22)$$

corresponding to the emission of the gluon $T_a G_\mu^a$ by a quark ψ_q as represented in Fig. 2, similar to the emission of a photon by a charged particle in the case of QED, we have, according to (20) and (21) the emission of a gluon by a gluon, and the coupling of two gluons to two gluons, Fig. 2. These terms have important consequences. They will give rise to asymptotic freedom at high energies and strong interactions at low energies contrary to what happens in QED. From now on we will use for the coupling constant $\alpha_s = g_s^2/(4\pi)$ instead of g_s .

2 Renormalisation Group Equation

The renormalisation group equation exploits the requirement that a physical quantity Γ does not depend on the particular scale at which it is renormalised. The approach was first introduced by Stückelberg, Peterman, Gell-Mann and Low in the 50s, and later by Callan and Symanzik. Let us denote by $\Gamma_B(s, \alpha_0, \Lambda^2)$ the bare quantity, where α_0 is the bare coupling and Λ is the ultraviolet cut-off to regularise the unrenormalised quantity. The renormalised quantity Γ is defined at scale μ^2 , and also the coupling constant $\alpha(\mu)$ and the normalisation scale factor Z such that

$$\Gamma(s, \alpha(\mu), \mu^2) = Z(\mu^2) \Gamma_B(s, \alpha_0, \Lambda^2). \quad (23)$$

Γ_B has to be independent of the choice of scale μ^2 , therefore

$$\frac{d\Gamma_B}{d\mu^2} = 0, \quad (24)$$

which can be written as

$$\mu^2 \frac{\partial \Gamma}{\partial \mu^2} + \mu^2 \frac{\partial \alpha}{\partial \mu^2} \frac{\partial \Gamma}{\partial \alpha} - \mu^2 \frac{\Gamma}{Z} \frac{\partial Z}{\partial \mu^2} = 0. \quad (25)$$

Defining

$$\beta(\alpha) = \mu^2 \frac{\partial \alpha}{\partial \mu^2} \quad (26)$$

and

$$\gamma(\alpha) = -\mu^2 \frac{\Gamma}{Z} \frac{\partial Z}{\partial \mu^2}, \quad (27)$$

(25) becomes

$$\left[\mu^2 \frac{\partial}{\partial \mu^2} + \beta(\alpha) \frac{\partial}{\partial \alpha} + \gamma(\alpha) \right] \Gamma(s, \alpha(\mu), \mu^2) = 0, \quad (28)$$

which is known as the renormalisation group equation that expresses how you must compensate a change in the scale μ^2 , changing the coupling constant and the overall scale factor Z .

3 Charge Screening: Asymptotic Freedom

From (26), we can write

$$\ln \left(\frac{\mu^2}{\mu_0^2} \right) = \int_{\alpha(\mu_0^2)}^{\alpha(\mu^2)} \frac{d\alpha}{\beta(\alpha)}. \quad (29)$$

In QCD, $\beta(\alpha)$ can be calculated perturbatively with the result

$$\beta(\alpha_s) = -b_0 \alpha_s^2 - b_1 \alpha_s^3 + \mathcal{O}(\alpha_s^4), \quad (30)$$

with

$$b_0 = \frac{11N_c - 2N_f}{12\pi},$$

$$b_1 = \frac{17N_c^2 - [5N_c + 3(N_c^2 - 1)/(2N_c)]N_f}{24\pi^2}, \quad (31)$$

with N_f the number of active flavours. Retaining the first term of (30) in (29), we obtain

$$\alpha_s(\mu^2) = \frac{\alpha_s(\mu_0^2)}{1 + \alpha_s(\mu_0^2)b_0 \ln(\mu^2/\mu_0^2)}. \quad (32)$$

We observe that for large μ^2 or small distances $r^2 \sim 1/\mu^2$, we obtain asymptotic freedom, $\alpha_s(\mu^2) \rightarrow 0$. This behaviour is opposite to the one in QED, where

$$\alpha_{em}(\mu^2) = \frac{\alpha_{em}(\mu_0^2)}{1 - \frac{\alpha_{em}(\mu_0^2)}{3\pi} \ln(\mu^2/\mu_0^2)}. \quad (33)$$

Here, for $r^2 < r_0^2$ ($\mu^2 > \mu_0^2$), $\alpha_{em}(\mu^2) > \alpha_{em}(\mu_0^2)$, and the effective charge grows at smaller distances. This behaviour is nothing but the charge screening effect in QED. In fact, if we consider the electric field produced by two charges located in a medium, the particles in the medium are oriented in the direction of the electric field, producing a reduction of the effective value of the two charges (Fig. 3). Close to each charge, the effective charge is the original charge (bare) minus the induced charge, therefore at smaller distances the screening is smaller and the value of the effective charge becomes closer to the value of the original charge (Fig. 4).

Microscopically, the bare charge is given by the left diagram in Fig. 5. In addition to this diagram, at lowest order in α_{em} we have another diagram (Fig. 5 right) giving rise to (33). The e^+e^- pairs in Fig. 5 (right) lead to charge screening.

In QCD, in addition to similar diagrams to those in QED, we have another diagram, see Fig. 6c, that changes the sign of the denominator of (33) if $11N_c - 2N_f > 0$. The evolution of the coupling constant α_s according to (32) has been checked in deep inelastic scattering, e^+e^- annihilation and heavy flavour experiments as it is shown in Fig. 7.

Although the lagrangian of massless QCD has no scale, the renormalisation introduces a dimensional scale parameter. In fact, from (29) we have

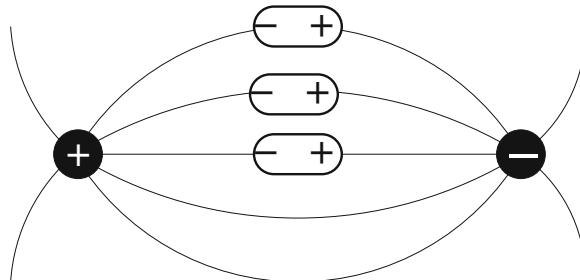


Fig. 3 Polarisation of a medium in QED

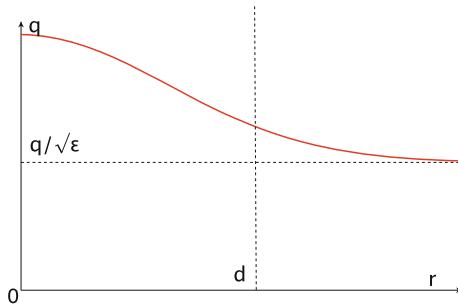


Fig. 4 Effective charge in a polarisable QED medium

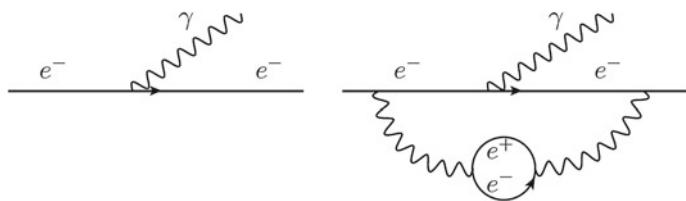


Fig. 5 Diagrams contributing to charge renormalisation in QED

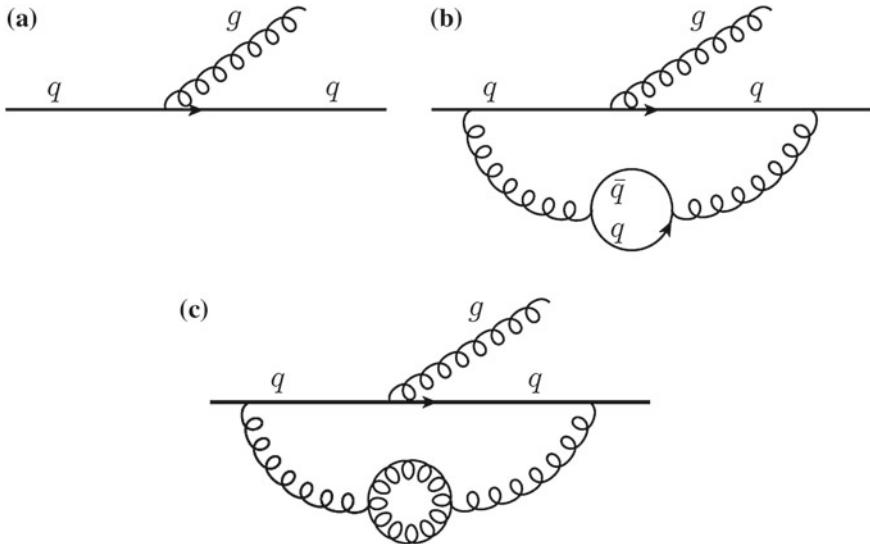


Fig. 6 Diagrams contributing to charge renormalisation in QCD

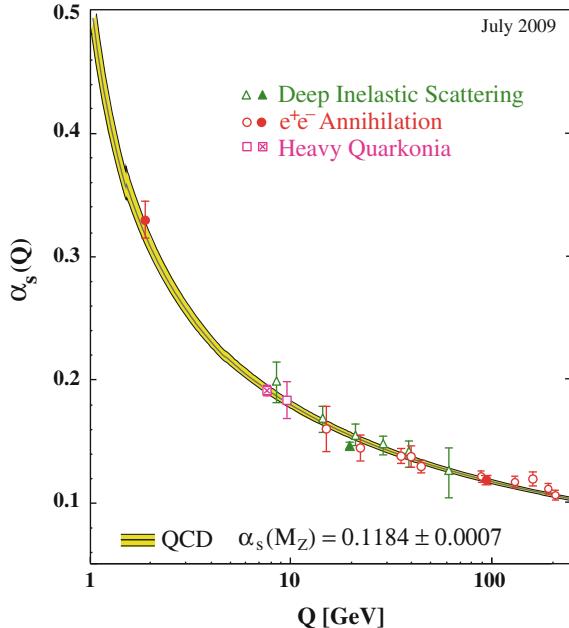


Fig. 7 Compilation of results on the evolution of the strong coupling constant α_s with momentum $Q \propto 1/r$, taken from [11]

$$\Lambda^2 = \mu^2 \exp \left(- \int \frac{dx}{\beta(x)} \right). \quad (34)$$

Equation (32) is derived from (34) taking the leading order expression for $\beta(x)$. It could be said that Λ is the scale at which the coupling constant becomes infinite. However, this is not self-consistent because we use the first term for $\beta(x)$ in a region where the coupling is large. More generally, we can say that Λ is the scale where QCD becomes non-perturbative.

4 Confinement

In QCD, the particles associated to the fields, quarks and gluons, are not asymptotic free states opposite to most of the field theories. Quarks and gluons are confined inside hadrons. This behaviour in QCD is not understood analytically, see [12, 13], and only doing some approximation it can be obtained. There are lattice numerical simulations that allow to reproduce some of the consequences of confinement, the masses of the different hadrons [14]. Notice that more of 95 % of hadron masses is due to the strong interaction and not to the quark masses that are quite small.

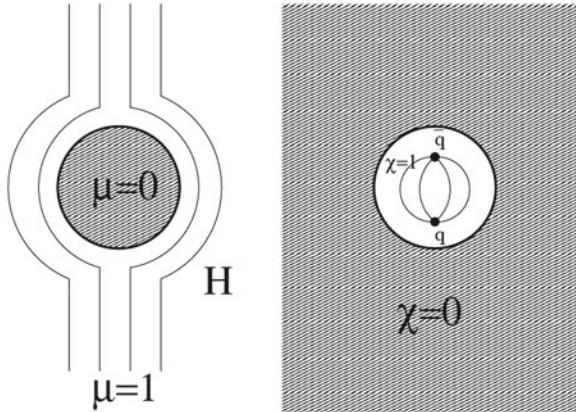


Fig. 8 Illustration of the picture of the dual superconductor

In analogy with QED, confinement can be seen in the following way: the QCD vacuum can be regarded as a gluon and quark–antiquark condensate, being a perfect dielectric (colour dielectric constant $\chi = 0$) similarly to the electron pairs in a superconductor in QED. In this case, the superconductor is a perfect diamagnet (magnetic susceptibility $\mu = 0$). The role of the magnetic field in QED is played by the chromoelectric field in QCD, the QED superconductor is the QCD vacuum and the QED vacuum ($\mu = 1$) is the inside of the hadron in QCD ($\chi = 1$). The role of the inside and outside in QED and QCD are exchanged. Due to this, this picture is known as the dual superconductor picture. In the same way that the magnetic field is expelled from the superconductor, the chromoelectric field is expelled from the QCD vacuum that pushes it towards the inside of the hadron, implying colour confinement as illustrated in Fig. 8.

5 QCD Colour Factors: Jet Structure

In the computation of cross sections, the involved states have to be summed or averaged giving rise to colour factors. In order to compute them, we use the properties of $SU(3)$ or $SU(N_c)$. The relevant properties of the fundamental representation of QCD, t^a used for quarks, and of the adjoint matrix representation of QCD, T^a used for gluons, are (14)–(16). The t^a are normalised such that

$$\text{tr}(t^a t^b) = T_F \delta^{ab} = \frac{1}{2} \delta^{ab}, \quad (35)$$

and the colour matrixes satisfy the following relations:



Fig. 9 Diagrams of the quark–gluon, gluon–gluon and quark–antiquark loops (from *left* to *right*) to which the colour factors C_F , C_A and T_F are respectively associated

$$\sum_a (t^a)_{ik} (t^a)_{kj} = C_F \delta_{ij}, \quad C_F = \frac{N_c^2 - 1}{2N_c},$$

$$\text{tr}(T^a T^b) = \sum_{c,d} f^{acd} f^{bcd} = C_A \delta^{ab}, \quad C_A = N_c, \quad (36)$$

where C_F and C_A are quadratic Casimir invariants. For $SU(3)$, $C_F = 4/3$ and $C_A = 3$. C_F , C_A and T_F are the colour factors associated to quark–gluon, gluon–gluon and quark–antiquark loops respectively as it is illustrated in Fig. 9.

In this way, the annihilation $e^+ e^- \rightarrow \gamma^*, Z \rightarrow q\bar{q}$ is proportional to

$$\sum_{\text{colours}} \delta_{ij} \delta_{ji}^* = \text{tr}(\delta) = N_c. \quad (37)$$

In the Drell-Yan process, we must average over initial quark and antiquark colours, therefore

$$\sigma(q\bar{q} \rightarrow \gamma^*, Z \rightarrow l^+l^-) \propto 1/N_c. \quad (38)$$

In the same way, the deep inelastic scattering $lq \rightarrow \gamma^*, Z \rightarrow lq$ cross section should be proportional to 1 because the sum over colour of the quarks in the final state is compensated by the average over the initial colour quark.

Finally, for the process $\gamma^*, Z \rightarrow q\bar{q}g$ that gives rise to three jets, we have

$$\sum_{\text{colours}} |\mathcal{M}|^2 \propto \delta_{ij} t_{jk}^a (t_{lk}^a \delta_{il})^* = \text{tr}(t^a t^a) = \frac{1}{2} \text{tr}(\delta) = 4. \quad (39)$$

Notice that the multi particle production structure in $e^+e^- \rightarrow \text{hadrons}$ is in jets a consequence of asymptotic freedom. In fact, the process is via the intermediate γ or Z that produces a quark–antiquark pair. This pair evolves as a colour singlet object, emitting gluons that in turn give rise to gluons or new quark–antiquark pairs, recombining at the end of the cascade to produce the observed hadrons in the final state. On the other hand, the probability that a quark emits a gluon at large transverse momentum (Fig. 10) is smaller than the probability of emitting a gluon of small transverse momentum due to asymptotic freedom that tells us that $\alpha_s(|p - p'|^2)$ decreases for large $|p - p'|^2$. Therefore, the largest probability is for the emission of collimated gluons along the directions of the quark and the antiquark, forming two collimated jets. The existence of jets was proved by Sterman and Weinberg [15].

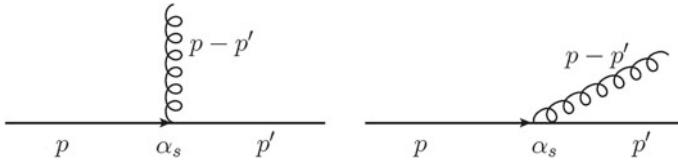


Fig. 10 Diagrams for gluon emission

6 Symmetries

6.1 Isospin Symmetry

Let us consider QCD for u and d quarks. In the limit of equal masses, the QCD lagrangian is invariant under the transformations

$$\begin{aligned} u &\longrightarrow \alpha u + \beta d, \\ d &\longrightarrow \gamma u + \delta d, \end{aligned} \quad V = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \quad \det V = 1, \quad (40)$$

i.e. $V \in SU(2)$.

u and d are isospin doublets (the rest of the quarks s, c, b, t , are isospin singlets). Actually, $m_u < m_d$ and isospin is only an approximate symmetry. We can write the mass term in the lagrangian as

$$m_u \bar{u}u + m_d \bar{d}d = \frac{1}{2}(m_u + m_d)(\bar{u}u + \bar{d}d) + \frac{1}{2}(m_d - m_u)(\bar{d}d - \bar{u}u). \quad (41)$$

The first term is also isospin invariant but the second term breaks the symmetry. The mass difference $m_d - m_u$ plays the role of a parameter that breaks the symmetry. The fact that hadrons are in almost degenerate multiplets implies that the second term is a small perturbation and therefore $m_d - m_u$ must be small.

6.2 Chiral Symmetry

Let us consider the QCD lagrangian for massless u and d quarks. Then, the lagrangian is not only symmetric under the isospin symmetry transformation but also under the chiral transformation [16]

$$\begin{pmatrix} u_R \\ d_R \end{pmatrix} \longrightarrow V_R \begin{pmatrix} u_R \\ d_R \end{pmatrix}, \quad \begin{pmatrix} u_L \\ d_L \end{pmatrix} \longrightarrow V_L \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad V_R, V_L \in SU(2), \quad (42)$$

where u_R, d_R (u_L, d_L) denote the right (left) helicity states of quarks u and d .

This mapping corresponds to the direct product of two isospin groups: $SU(2)_L \times SU(2)_R$. The generators of this group are the three isospin generator \mathbf{I} and the three chiral isospin generators \mathbf{I}_5 . The symmetry implies that the components of $\mathbf{I}(I^+, I^-, I^3)$ and $\mathbf{I}_5(I_5^+, I_5^-, I_5^3)$ are conserved. For instance, for I^+ and I_5^+

$$\begin{aligned} I^+ &= \int d\mathbf{x} \bar{u}\gamma^0 d = \int d\mathbf{x} u^+ d, \\ I_5^+ &= \int d\mathbf{x} \bar{u}\gamma^0 \gamma_5 d = \int d\mathbf{x} u^+ \gamma_5 d, \end{aligned} \quad (43)$$

we have

$$\begin{aligned} \partial_\mu(\bar{u}\gamma^\mu d) &= i(m_u - m_d)\bar{u}d, \\ \partial_\mu(\bar{u}\gamma^\mu \gamma_5 d) &= i(m_u + m_d)\bar{u}\gamma_5 d, \end{aligned} \quad (44)$$

which vanish for $m_u = m_d = 0$ (if $m_u = m_d \neq 0$ i.e. isospin symmetry, only the first equation of (44) vanishes).

Chiral symmetry must be spontaneously broken as pointed out by Nambu [17]. In fact, we observe that in the case of isospin symmetry: (a) The energy levels are degenerate forming multiplets i.e. hadrons are grouped into multiplets of equal mass; (b) The operators \mathbf{I} connect hadrons within the same multiplet—for instance, a neutron is transformed into a proton by $I^+, I^+|n\rangle = |p\rangle$; (c) The ground state $|\Omega\rangle$ is an isospin singlet, $\mathbf{I}|\Omega\rangle = 0$.

If chiral symmetry takes place in the same way, the states should be grouped in degenerate multiplets of $SU(2)_L \times SU(2)_R$. As \mathbf{I}_5 has negative parity, for each isospin multiplet it should exist a multiplet with the same mass and opposite parity. This is not seen in nature. For instance, the partners of the multiplet $|n\rangle$, $|p\rangle$ are $\mathbf{I}_5|n\rangle$, $\mathbf{I}_5|p\rangle$ that have opposite parity to neutron and proton. No such states exist.

If the vacuum is invariant under the action of any element of a given group G , we have

$$S|\Omega\rangle = e^{i\varepsilon Q}|\Omega\rangle \simeq (1 + i\varepsilon Q)|\Omega\rangle = |\Omega\rangle, \quad (45)$$

therefore

$$Q|\Omega\rangle = 0. \quad (46)$$

In our case

$$\mathbf{I}|\Omega\rangle = 0 \quad (47)$$

but

$$\mathbf{I}_5|\Omega\rangle \neq 0. \quad (48)$$

The vacuum breaks the symmetry (i.e. the symmetry is spontaneously broken). As

$$[\mathcal{L}_{QCD}, \mathbf{I}_5] = 0, \quad (49)$$

the three states $\mathbf{I}_5|\Omega\rangle$ have the same energy as $|\Omega\rangle$. On the other hand, as $|\Omega\rangle$ has zero momentum and \mathbf{I}_5 carries no momentum, $\mathbf{I}_5|\Omega\rangle$ must have zero momentum. Therefore, there will be three massless particles. indeed, on general grounds the Goldstone theorem [18] tells us that a spontaneously broken symmetry gives rise to a number of massless particles equal to the number of generators of the symmetry that become broken. The quantum numbers of $\mathbf{I}_5|\Omega\rangle$ are spin $\mathbf{S} = 0$, negative parity $P = -1$ and isospin $\mathbf{I} = 1$ that are just the quantum numbers of the isospin triplet π^+, π^0, π^- . Therefore, it is natural to identify

$$I_5^+|\Omega\rangle = |\pi^+\rangle, \quad I_5^3|\Omega\rangle = |\pi^0\rangle, \quad I_5^-|\Omega\rangle = |\pi^-\rangle \quad (50)$$

as the three Goldstone bosons.

Note that $I_5^+|n\rangle$ is not a singlet state partner of the neutron $|n\rangle$ but the composed state formed by a positive pion and a neutron $|\pi^+|n\rangle$ is.

Actually $m_u, m_d \neq 0$ and therefore the chiral symmetry of the lagrangian is explicitly broken due to the mass term. We can split the hamiltonian in an invariant part \mathcal{H}_0 and the mass term,

$$\begin{aligned} \mathcal{H} &= \mathcal{H}_0 + \mathcal{H}_{sb}, \quad \mathcal{H}_{sb} = \int d\mathbf{x} (m_u \bar{u} u + m_d \bar{d} d), \\ [\mathcal{H}_0, \mathbf{I}] &= 0 = [\mathcal{H}_0, \mathbf{I}_5]. \end{aligned} \quad (51)$$

Note that the quarks s, c, b, t are singlets and their mass do not break the symmetry. They are included in \mathcal{H}_0 .

On the other hand, as

$$\langle \pi^+ | \bar{u} \gamma^\mu \gamma^5 d | 0 \rangle = -ip^\mu \sqrt{2} F_\pi e^{ip \cdot x} \quad (52)$$

and

$$\langle \pi^+ | \bar{u} \gamma^5 d | 0 \rangle = i \sqrt{2} G_\pi e^{ip \cdot x}, \quad (53)$$

where F_π is the axial coupling fixed by the decay width $\pi^+ \rightarrow \mu^+ \nu$, $\Gamma \propto F_\pi^2$, and G_π the pseudo scalar coupling. As the second of (44), known as partially conserved axial current (PCAC), relates the matrix elements of (52) and (53), we obtain

$$m_\pi^2 = (m_u + m_d) \frac{G_\pi}{F_\pi}. \quad (54)$$

The (54) shows that for massless u and d quarks (chiral symmetry), $m_\pi = 0$. We observe that the pion mass does not depend linearly on the quark masses but quadratically.

6.3 Scale Symmetry and the Trace Anomaly

Under the mapping

$$x \longrightarrow \lambda x, \quad \psi_q \longrightarrow \lambda^{3/2} \psi_q, \quad G_\mu^a \longrightarrow \lambda G_\mu^a, \quad (55)$$

the QCD lagrangian is invariant for massless quarks. This symmetry is broken by quantum corrections introduced in the renormalisation process that always includes a scale. Therefore, we have a symmetry at the classical level that is broken by quantum corrections—an anomaly.

The current associated to the transformation is $s^\mu = T^{\mu\nu}x_\nu$, where $T^{\mu\nu}$ is the energy-momentum tensor. Its divergence satisfies

$$\partial_\mu s^\mu = T_\mu^\mu = \frac{\beta(g_s)}{2g_s} F_{\mu\nu}^a F^{\mu\nu a} + [1 + \gamma(g_s)] m_q \bar{\psi}_q \psi_q, \quad (56)$$

where β and γ are the functions defined in (26) and (27).

T_μ^μ is the trace of the energy-momentum tensor, playing a very important role at finite temperatures, because

$$T_\mu^\mu = \varepsilon - 3p \quad (57)$$

with ε and p the energy and pressure density respectively. The equation of state for any free gas is $\varepsilon = 3p$ and T_μ^μ vanishes. Lattice QCD shows that T_μ^μ presents a maximum at a critical temperature T_c , marking the phase transition from a confined to a deconfined phase.

6.4 Vacuum Structure and Instantons

6.4.1 The Vacuum in Quantum Mechanics: Instantons

Let us consider a periodic potential (Fig. 11). Classically, there are infinite degenerate ground states, each state corresponding to the particle sitting at rest at the positions $x_n = na$, $n = 0, \pm 1, \dots$. Quantum mechanically, if the barrier between each well is large enough, the classical ground states will correspond to Gaussian functions centered at x_n , having each an energy $\hbar\omega/2$.

Any linear superposition

$$\psi(x) = \sum_n c_n \psi(x - x_n) \quad (58)$$

is again a ground state with energy $\hbar\omega/2$. The periodic symmetry implies that if $\psi(x)$ is a solution, $\psi(x + a)$ must be a solution. Hence, the solutions must have a

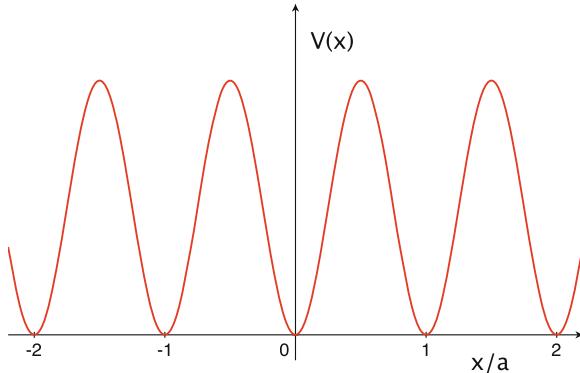


Fig. 11 Scheme of a potential with multiple degenerate vacuums

shape $e^{i\theta}\psi(x)$ and therefore $c_n = e^{in\theta}$. In this way, we write the equation

$$|\theta\rangle = \sum_n e^{in\theta} |n\rangle. \quad (59)$$

The quantum tunnelling between the states $|n\rangle$ breaks the degeneracy of the state $|\theta\rangle$ and the one corresponding to the minimal energy will be the ground state.

A very simple example of this is the wells represented in Fig. 12. The height of the central barrier is V_0/g . Doing $g \rightarrow 0$, there is an infinite barrier. In this case the ground state is doubly degenerate with energy E_0 . For $g \neq 0$, there is tunnelling and the degeneracy is broken. The corresponding states $|\Theta\rangle$ now are

$$\psi_{\pm} = \frac{1}{\sqrt{2}} [\psi(x - b - l) \pm \psi(x + b + l)] \quad (60)$$

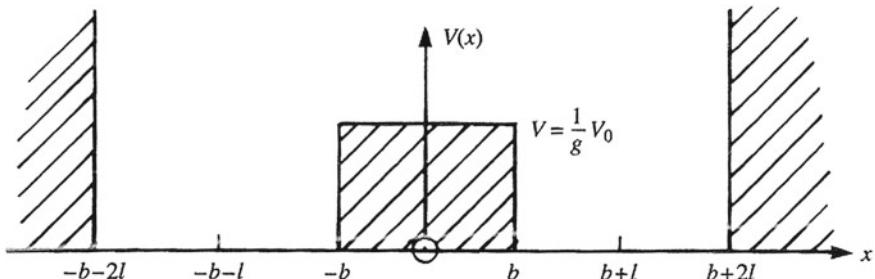


Fig. 12 Two symmetric wells separated by a barrier. (Taken from the book by Leader and Predazzi [8])

corresponding to the energies E_{\pm} with

$$\Delta E = E_- - E_+ \simeq \frac{4\hbar E_0 \sqrt{g}}{l \sqrt{2mV_0}} \exp \left[-\frac{2b}{\hbar} \sqrt{\frac{2mV_0}{g}} \right]. \quad (61)$$

Note that ΔE is not analytic at $g = 0$ and therefore cannot be obtained perturbatively.

In the path integral formalism the transition amplitudes are evaluated by the expression $\exp[iS(x)/\hbar]$ summing over all possible paths $x(t)$ contributing to the classical action $S(x(t))$, adding the contributions from paths obtained by small perturbations. However, in the case of transition amplitudes through barriers there is no classical path. In order to compute it, times t_1, t_2 are continued to imaginary values $-i\tau_1, -i\tau_2$, so that the exponent becomes

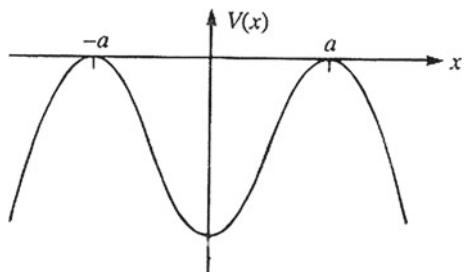
$$\begin{aligned} \frac{i}{\hbar} S(x)|_{t_1=-i\tau_1}^{t_2=-i\tau_2} &= \frac{i}{\hbar} \int_{-i\tau_1}^{-i\tau_2} dt \left[\frac{1}{2m} \left(\frac{dx}{dt} \right)^2 - V(x) \right] \\ &= \frac{1}{\hbar} \int_{\tau_1}^{\tau_2} d\tau \left[-\frac{1}{2m} \left(\frac{dx}{d\tau} \right)^2 - V(x) \right] = -\frac{1}{\hbar} S_E(x)|_{\tau_1}^{\tau_2}, \end{aligned} \quad (62)$$

where S_E , called the Euclidean action, is

$$S_E(x)|_{\tau_1}^{\tau_2} = \int_{\tau_1}^{\tau_2} d\tau \left[\frac{1}{2m} \left(\frac{dx}{d\tau} \right)^2 + V(x) \right], \quad (63)$$

which is the action for a potential $-V(x)$ such as shown in Fig. 13. What was a barrier has become a well, so that classical paths will now exist connecting x_1 at τ_1 with x_2 at τ_2 . ΔE can be obtained computing the transition amplitude $\langle x_1 = a, t = T | x_2 = a, t = 0 \rangle$ in the limit $T \rightarrow \infty$ that will be dominated by the classical paths in the “mirror” problem. These classical paths are peculiar. In fact, it takes an infinitely long time to reach a because the potential is flat at $-a$ and a , hence the velocity at these points is close to zero. Thus, the particles at rest for an infinite time and for a

Fig. 13 Potential for the Euclidean problem. (Taken from the book by Leader and Predazzi [8])



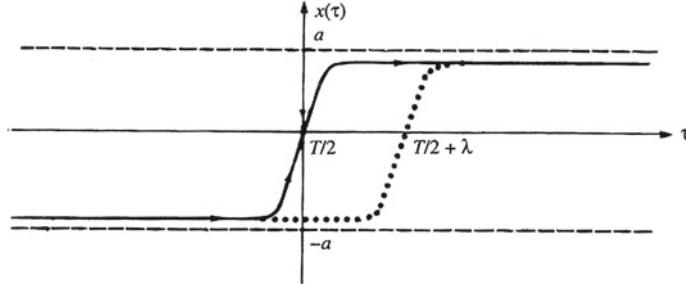


Fig. 14 Trajectories of the instanton in Euclidean time: minimum (solid) and nearby trajectories (dotted). (Taken from the book by Leader and Predazzi [8])

short time it moves fast. The form of a classical path is shown in Fig. 14. A particle moving in this way is called instanton because its kinetic energy is non-zero only for a very short time. The Euclidean action is a minimum for an instant path and there will be infinite paths as the dotted line shown in Fig. 14, that differ from the one of the instant on only over a finite region of time. The sum of all will give an important contribution.

6.4.2 The QCD Vacuum

Let us consider the gauge matrix $G_\mu(x) = T^a G_\mu^a(x)$ and take $G_\mu^a(x) = 0$. The spatial components of the gauge matrix transform as

$$\mathbf{G} \rightarrow U \mathbf{G} U^{-1} + \frac{i}{g} (\nabla U) U^{-1}, \quad (64)$$

with

$$U = \exp(-iT^a \mathbf{G}^a). \quad (65)$$

The field energy is zero when $\mathbf{G} = 0$ and therefore it will be zero when \mathbf{G} is a pure gauge,

$$\mathbf{G} = \frac{i}{g} (\nabla U) U^{-1}, \quad (66)$$

because they are each other related to the gauge transformation (64).

In order to see the degeneracy of the ground state we are going to use $SU(2)$ instead of $SU(3)$ for simplicity. For $SU(2)$, the generators are proportional to the Pauli matrixes, so that the gauge transformations (65) become

$$U = \exp(-i\boldsymbol{\sigma} \cdot \boldsymbol{\theta}(x)/2). \quad (67)$$

In general, they do not have an angle independent limit as $r \rightarrow \infty$, hence we can write the more restrictive form

$$U = \exp(-i\sigma \cdot \mathbf{r} f(r)/r) = I \cos f(r) + i \frac{\sigma \cdot \mathbf{r}}{r} \sin f(r) \quad (68)$$

that will be independent of the angles for $r \rightarrow \infty$ provided $\sin f(r) = 0$ as $r \rightarrow \infty$ or

$$\lim_{r \rightarrow \infty} f(r) = n\pi \quad (69)$$

which implies $\lim_{r \rightarrow \infty} U(\mathbf{r}) = \pm I$.

We also require that $f(0) = 0$ for $U(\mathbf{r})$ to be a single valued functions at $r = 0$.

The gauge transformations (68) can be classified into subsets, labelled by an integer n ,

$$U_n(\mathbf{r}) = \exp(-i\sigma \cdot \mathbf{r} f_n(r)/r), \quad (70)$$

with $f_n(\infty) = n\pi$. A given gauge transformation of one class cannot be continuously deformed into another gauge transformation of another class.

The gauge transformation class (70) satisfies

$$n = \frac{1}{24\pi^2} \int d\mathbf{r} \varepsilon_{ijk} \text{tr}\{(U \partial_i U^{-1})(U \partial_j U^{-1})(U \partial_k U^{-1})\}. \quad (71)$$

The integer n is called the winding number or the topological charge that labels the gauge transformation classes, called homotopy classes.

In summary, the classical theory has degenerate vacuum states (with zero energy) labelled by the winding numbers n , corresponding to pure gauge configurations

$$\mathbf{G}_n = \frac{i}{g} (\nabla U_n) U_n^{-1}. \quad (72)$$

In the semiclassical approximation to quantum QCD, the vacuum will be

$$|\theta\rangle = \sum e^{in\theta} |\mathbf{G}_n\rangle, \quad (73)$$

where the states $|\mathbf{G}_n\rangle$ are eigenstates of the matrix gauge operators \mathbf{G}_n that have the values corresponding to the ground state of winding number n .

Similarly to the one-dimensional period potential, there is tunnelling between the states $|\mathbf{G}_n\rangle$. This tunnelling is generated through the existence of instanton configurations of the gauge matrix $\mathbf{G}_n(\tau, \mathbf{r})$ with n integer that correspond to a finite Euclidean action, being solutions of the classical field equations in the Euclidean space having essentially zero Euclidean field energy and evolving from $\mathbf{G}_n(\mathbf{r})$ at $\tau \rightarrow -\infty$ to $\mathbf{G}_{n+n'}(\mathbf{r})$ at $\tau \rightarrow +\infty$.

The Euclidean action is given by

$$S_E = \frac{1}{2} \int d\tau d\mathbf{r} \operatorname{tr}(F_{\alpha\beta}^{aE} F_{\alpha\beta}^{aE}) = \frac{1}{2} \int d\tau d\mathbf{r} \operatorname{tr}(F_{\mu\nu}^a F^{\mu\nu a}), \quad (74)$$

where α, β are Euclidean indices (x, y, z, it) and $F_{ij}^E = -F^{ij}$, $F_{4j}^E = -iF^{0j}$.

Now

$$\mathbf{G}_\mu \xrightarrow{R \rightarrow \infty} -\frac{i}{g} (\partial_\mu U) U^{-1}, \quad (75)$$

with $R = \sqrt{\tau^2 + r^2}$.

From a similar formula to (71) it is obtained

$$n = -\frac{g^2}{16\pi^2} \int d\tau d\mathbf{r} \operatorname{tr}(F_{\alpha\beta}^{aE} F_{\alpha\beta}^{aE}) = -\frac{g^2}{16\pi^2} \int d\tau d\mathbf{r} \operatorname{tr}(F_{\mu\nu}^a F^{\mu\nu a}). \quad (76)$$

Defining the current

$$K^\mu \equiv 2\varepsilon^{\mu\nu\rho\sigma} \operatorname{tr} \left(\mathbf{G}_\nu^a \partial_\rho \mathbf{G}_\sigma^a - \frac{2}{3} ig f_{abc} \mathbf{G}_\nu^a \mathbf{G}_\rho^b \mathbf{G}_\sigma^c \right), \quad (77)$$

it is observed that

$$\partial_\mu K^\mu = \operatorname{tr}(\tilde{F}_{\mu\nu} \tilde{F}^{\mu\nu}), \quad (78)$$

where the dual tensor $\tilde{F}_{\mu\nu}$ is defined by

$$\tilde{F}_{\mu\nu} = \frac{1}{2} \varepsilon_{\mu\nu\gamma\delta} F_{\gamma\delta}. \quad (79)$$

Introducing (78) in (76) and using the Gauss theorem, we get

$$n = -\frac{g^2}{16\pi^2} \int d\tau d\mathbf{r} \partial_\alpha^K K_\alpha^E = -\frac{g^2}{16\pi^2} \int_{r=\infty} dS_\alpha K_\alpha^E \quad (80)$$

($K_4^E = iK^0$, $K_j^E = K^j$), where the last integral is extended to the three dimensional surface $R = \infty$. At this surface, the finite action configurations are pure gauge, then using (75) and (77) we recover (71).

Note that in the evaluation of the probability amplitude to go from the ground state in t_1 to the ground state in t_2 we will have, going to imaginary time,

$$\langle \theta | e^{-\mathcal{H}(\tau_2 - \tau_1)} | \theta \rangle = \sum_{n,m} e^{i(n-m)\theta} \langle \mathbf{G}_m | e^{-\mathcal{H}(\tau_2 - \tau_1)} | \mathbf{G}_n \rangle. \quad (81)$$

Hence, there is need to compute

$$e^{i(n-m)\theta} \exp[-S_E] = \exp \left\{ - \int d\tau d\mathbf{r} \left[\mathcal{L} - i\theta \frac{g^2}{16\pi^2} \text{tr}(F_{\alpha\beta}^{aE} F_{\alpha\beta}^{aE}) \right] \right\}, \quad (82)$$

which is equivalent to work with the new lagrangian density

$$\mathcal{L}_\theta = \mathcal{L} + \theta \frac{g^2}{32\pi^2} F_{\mu\nu}^a \tilde{F}^{a\mu\nu}. \quad (83)$$

This new term violates CP invariance. Experimental data as the electric dipole moment of the neutron implies a very small number for θ , $\theta < 10^{-10}$. Why this small value? This is known as the QCD CP problem. There are several proposals to answer this question, including the existence of a new particle, the axion, whose field is associated to θ [19, 20].

A detailed discussion of all these aspects can be found in the book by Leader and Predazzi in [8].

6.4.3 $U_A(1)$ Symmetry and the Axial Anomaly

The QCD lagrangian is invariant under the $U(1)$ transformations

$$\psi_q \rightarrow e^{i\alpha} \psi_q, \quad \bar{\psi}_q \rightarrow e^{i\alpha\gamma_5} \bar{\psi}_q. \quad (84)$$

If the chiral transformation is spontaneously broken, there will be a massless Goldstone boson with $I = 0$ and $P = -1$. The lightest pseudo scalar with $I = 0$ is the η whose mass is $549 \text{ MeV}/c^2$, much larger than the pion mass. Therefore the symmetry is not spontaneously broken. This is the $U_A(1)$ problem. We discuss a way out [21, 22].

The symmetry under the chiral $U_A(1)$ gives rise to the current

$$j_{\mu 5}^0 = \bar{u} \gamma_\mu \gamma_5 u + \bar{d} \gamma_\mu \gamma_5 d \quad (85)$$

(we are considering two flavours u and d and the $I = 0$ content as the superscript indicates).

This current is conserved classically but not in a quantised theory, where

$$\partial^\mu j_{\mu 5}^0 = N_f \frac{g^2}{8\pi^2} F_{\mu\nu}^a \tilde{F}^{a\mu\nu}. \quad (86)$$

If we define the current

$$\tilde{j}_{\mu 5} = j_{\mu 5}^0 - N_f \frac{g^2}{8\pi^2} K_\mu, \quad (87)$$

it is obviously conserved,

$$\partial^\mu \tilde{j}_{\mu 5} = 0. \quad (88)$$

Let us consider the charges associated to these currents,

$$\begin{aligned} q_5(t) &= \int d\mathbf{r} j_{05}^0(t, \mathbf{r}), \\ Q_5(t) &= \int d\mathbf{r} \tilde{j}_{05}(t, \mathbf{r}). \end{aligned} \quad (89)$$

q_5 is gauge invariant but it is not time independent because $j_{\mu 5}^0$ is not conserved. Q_5 is conserved but it is not invariant under gauge transformations of a given class U_n ,

$$Q_5 \rightarrow U_n Q_5 U_n^{-1} = Q_5 + 2nN_f \quad (90)$$

or $[U_n, Q_5] = 2nN_f U_n$. Now, we have

$$[\mathcal{H}, U_n] = 0, \quad [\mathcal{H}, Q_5] = 0, \quad [U_n, Q_5] \neq 0 \quad (91)$$

and, therefore, we cannot diagonalise simultaneously \mathcal{H} , U_n and Q_5 . The vacuum $|\theta\rangle$ is an eigenstate of \mathcal{H} and U_n but not of Q_5 . Thus, the vacuum is not chiral invariant and $U_A(1)$ is not a symmetry due to the transition between instantons [21],

$$e^{-i\phi Q_5} |\theta\rangle = |\theta + 2N_f\phi\rangle. \quad (92)$$

7 The QCD Phase Diagram

Forty years ago, Lee and Wick [23] pointed out the possibility of studying new Physics by exploring the behaviour of high density matter or high energy density in a volume. In this way, it would be possible the restoration of broken symmetries, to disturb the the vacuum and the creation of abnormal dense matter. Also, it was very early realised [24–29] that asymptotic freedom should imply the existence of very high density matter formed by deconfined quarks and gluons. Later, these deconfined quarks and gluons, if thermalised, were referred to as Quark–Gluon Plasma (QGP). It would be expected a phase transition between normal nuclear matter and deconfined quarks and gluons because of the large difference in the number of degrees of freedom. In fact, the energy density for a gas of free pions is

$$\varepsilon_{\text{HG}} = \frac{\pi^2}{30} 3 T^4 \simeq T^4, \quad (93)$$

whereas for a gas of free quarks and gluons is

$$\varepsilon_{QGP} = \frac{\pi^2}{30} \left[2 \times 8 + \frac{7}{8} \times 2(3) \times 2 \times 2 \times 3 \right] T^4 = \frac{\pi^2}{30} [16 + 21(31.5)] T^4 \quad (94)$$

where we take into account that each of the 8 gluons has 2 helicities, and the 2(3) quark flavours may have 2 helicities, 3 possible colours and an additional factor 2 takes into account antiquarks. In (93) and (94) we neglect the masses, because their inclusion changes only slightly the formulae.

The energy density as a function of the temperature has been studied in lattice QCD with two and three light quarks and in the more realistic situation of two light and one heavier quark. Figure 15 shows that in a very narrow temperature interval there is rapid increase from the low values of hadronic matter to much higher values as it was expected in the transition from confined to deconfined quarks and gluons. It is observed that even for $T > T_c$ the lattice values are much lower than the values given by (94), 12.25 and 15.61 for the cases of 2 and 3 flavours respectively. This result points out that for moderate temperatures the deconfined quarks and gluons go on interacting strongly. It is denoted by strongly coupled QGP (scQGP). The transition seems to be a cross-over (fast change but without discontinuity) although this depends on the number of flavours used in the simulation and on the masses [30].

At deconfinement, the quark–antiquark binding is dissolved so that the conventional hadrons melt. However, the pions would persist beyond this point unless chiral symmetry is restored here. A measure of chiral symmetry breaking is provided by the quark mass term $\langle \bar{\psi} \psi \rangle$: if $\neq 0$, the chiral symmetry of the lagrangian is spontaneously broken. The chiral order parameter (condensate) is defined by

$$K(T) = \langle \bar{\psi} \psi \rangle = \frac{T}{V} \left(\frac{\partial \ln Z(T, m_q)}{\partial m_q} \right)_{m_q=0}. \quad (95)$$

Fig. 15 Energy density as a function of temperature in lattice QCD [31, 32]. From bottom to top the results are for two light flavours, three light flavours and two light and one heavy flavour. Arrows on the right side indicate the corresponding results for an ideal gas according to the Stefan-Boltzmann law

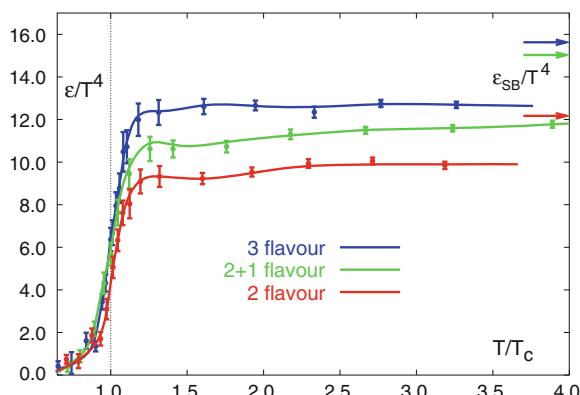
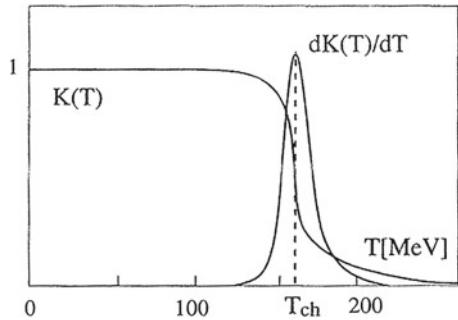


Fig. 16 Temperature dependence of the chiral condensate and of the corresponding susceptibility. (Taken from [30])



The behaviour of $K(T)$ in the case of two light (u,d) and one heavy (s) quark is shown in Fig. 16, with a sharp change at a critical temperature $T_c \simeq 160$ MeV. This temperature coincides with the critical temperature found in the deconfinement transition.

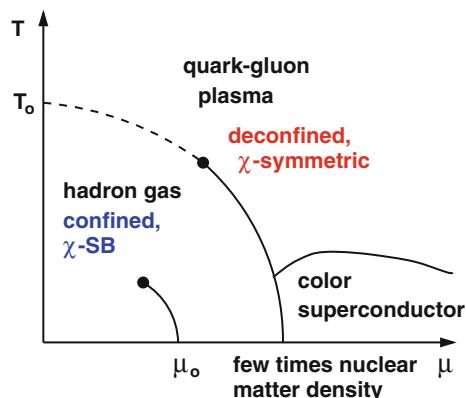
The results of lattice QCD for higher baryon density and low temperature point out to the existence of a first order phase transition. As at high temperature and lower baryon density the transition is just a cross-over, a critical point should exist as shown in Fig. 17.

The energy density reached in central (i.e. at impact parameter $b = 0$) nucleus-nucleus collisions can be evaluated using the Bjorken formula [33]

$$\varepsilon = \frac{\langle p_T \rangle dN/dy}{\tau_0 \pi R_A^2} . \quad (96)$$

Here dN/dy denotes the number of particles produced in central rapidity, $\langle p_T \rangle$ the mean transverse momentum, R_A the nuclear radius and $\tau_0 = 1$ fm/c is the time after the collision when the energy density is evaluated.

Fig. 17 Phase diagram of QCD obtained in lattice for 3 quarks [31]. Solid lines indicate a first order phase transition, dashed ones a cross-over and the dots indicate critical points



At the highest energies studied at the Super Proton Synchrotron (SPS) at CERN and at the Relativistic Heavy-Ion Collider (RHIC) at BNL for PbPb (SPS) and AuAu (RHIC) collisions, the values of the energy density are larger than 2 and 4 GeV/fm³ respectively. At the Large Hadron Collider (LHC) at CERN, for PbPb collisions at 2.76 TeV/nucleon, the value of the energy density is larger than 15 GeV/fm³. In all cases the values are above the critical energy determined in lattice QCD, 0.5 GeV/fm³.

Besides, lattice QCD has studied the behaviour of the pressure density, as shown in Fig. 18, and $(\varepsilon - 3p)/T^4$, shown in Fig. 19, related to the trace anomaly. It can be seen that $(\varepsilon - 3p)/T^4$ above the critical temperature decreases slowly in such a way that it has a sizeable value even for $T \simeq 3 - 4 T_c$. Therefore, even at those temperatures we are far from a state of free quarks and gluons whose equation of state is $\varepsilon = 3p$.

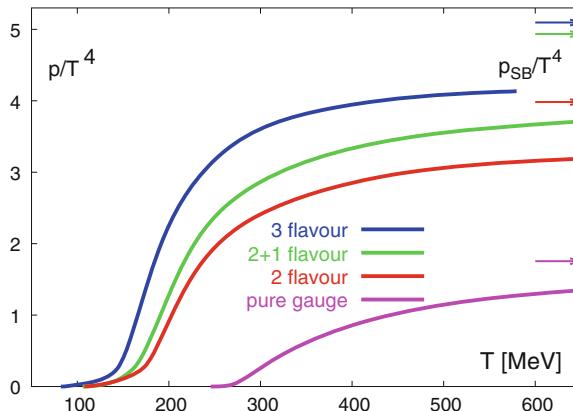


Fig. 18 Pressure versus temperature obtained in lattice QCD [31]. Lines from bottom to top indicate the results for 0, 2 light, 2 light and 1 heavy, and 3 light quarks. Arrows on the right side indicate the corresponding results for an ideal gas according to the Stefan-Boltzmann law

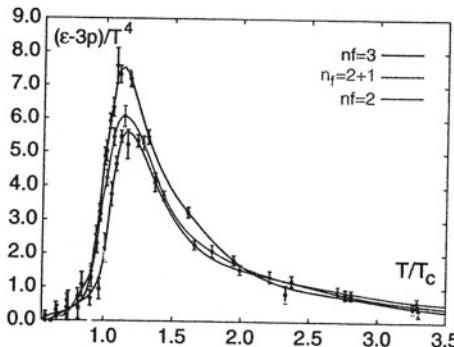


Fig. 19 Interaction measure from lattice QCD, for 2 light, 2 light and 1 heavy, and 3 light quarks. (Taken from [30])

In the last decades there has been much experimental activity concerned with the obtention and study of the high density state of quarks and gluons. First, the Intersecting Storage Rings (ISR) at CERN and the Alternating Gradient Synchrotron (AGS) at BNL explored collisions between light nuclei and at not very high energies. Later, the SPS experiments at 20 GeV/nucleons and RHIC at 62.4 and 200 GeV/nucleon studied extensively a broad range of collisions, centralities and observables. Now the LHC has collided PbPb (also pPb) at 2.76 TeV/nucleon and will reach even higher energies soon. There are many interesting and exciting phenomena seen in these experiments [34–40]: Strong suppression of high p_T particles in nucleus-nucleus relative to the number obtained from nucleon–nucleon collisions scaled by the number of binary collisions; Suppression of back-to-back correlations in jet production (jet quenching); Elliptic flow of particles whose behaviour is consistent with hydrodynamical evaluations assuming that the partonic system created in the collision becomes isotropic in a short time after the collision, and this partonic system has a very low shear viscosity over entropy density value; Sequential melting of bottomonium states according to their binding energy and size, such as expected due to deconfinement and colour screening—related to this, it has been observed an interesting behaviour of the suppression of the J/ψ and the charmonium resonances; Rapidity long range correlations and ridge structure in the two particle correlations. All these phenomena point out to the existence of an initial state with saturated gluons close to a thermalised state with a collective flow similar to a liquid of a very low shear viscosity. Due to the formation of this high density partonic medium, the propagation of particles with large p_T is strongly modified compared to the propagation in vacuum.

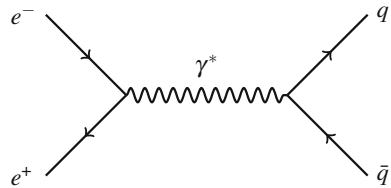
8 Phenomenological Applications of Perturbative QCD

In this Section we address the most important evidences of QCD as the theory of the strong interactions in the perturbative domain. Complete information and details on the subjects treated here can be found in [3, 41].

8.1 e^+e^- Annihilation into Hadrons

Let us consider the process of e^+e^- annihilation into hadrons at the lowest order in perturbation theory (leading order) i.e. at tree level, see Fig. 20. We neglect electro-weak corrections (Z^0 exchange) and consider the centre-of-mass frame and large energies $E_{cm} \gg \Lambda_{QCD}$. As hadronisation happens at scales of order Λ_{QCD} , while the production of the $q\bar{q}$ pair happens at scales E_{cm} , in this situation they can be considered as decoupled and perturbation theory in terms of quarks and gluons is expected to give a good description of the physical process.

Fig. 20 Feynman diagram for $e^+e^- \rightarrow$ hadrons at tree level



The corresponding angular dependence of the cross section for the production of one shower of particles (called jet—see below, that comes from the radiation of a single quark or antiquark) reads

$$\frac{d\sigma}{d\cos\theta} = \frac{\pi\alpha_{em}^2 Q_f^2}{2E_{cm}^2} (1 + \cos^2\theta), \quad (97)$$

with Q_f the electric charge of the quarks with flavour f . The total cross section for production of quarks of a given colour and flavour obtained upon angular integration of (97) reads $\sigma_0 = 4\pi\alpha_{em}^2 Q_f^2 / (3E_{cm}^2)$, and the corresponding ratio

$$R = \frac{\sigma(e^+e^- \rightarrow \text{hadrons})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)} = \frac{\sum_f \sigma(e^+e^- \rightarrow q\bar{q})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)} = N_c \sum_f Q_f^2, \quad (98)$$

where $N_c = 3$ is the number of colours. As commented before, the approximate description that this expression provides of experimental data is one of the evidences for the existence of 3 colours in Nature.

Now we turn to the next-to-leading (radiative) corrections in QCD to this process, that are $\mathcal{O}(\alpha_s)$. They can be classified into those that contain new real partons, called *real corrections*, and those that do not contain them but they participate in loops (*virtual corrections*). The corresponding amplitudes are given by the Feynman diagrams in Figs. 21 and 22 respectively.

The virtual corrections contain two types of divergencies. On the one hand, there are ultraviolet ones that are absorbed through the usual method of renormalisation into the redefinition of the fields and coupling constant. On the other hand, there are

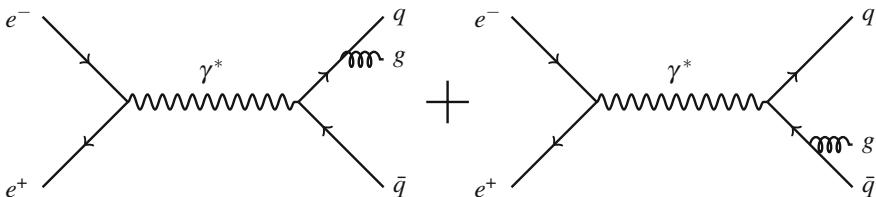


Fig. 21 Feynman diagrams of the real next-to-leading corrections to $e^+e^- \rightarrow$ hadrons

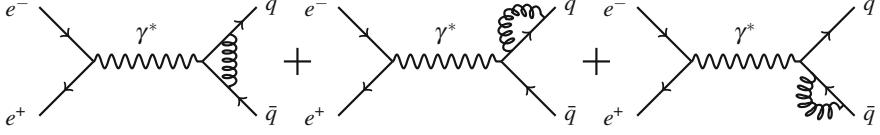


Fig. 22 Feynman diagrams of the virtual next-to-leading corrections to $e^+e^- \rightarrow$ hadrons

infrared divergencies for massless quarks that come through integrals like

$$\begin{aligned} I &= \int \frac{d^4 k}{(2\pi)^4} \frac{1}{k^2 + i\varepsilon} \frac{1}{(p_1 + k)^2 - m^2 + i\varepsilon} \frac{1}{(p_2 - k)^2 - m^2 + i\varepsilon} \\ &\sim \int \frac{d^4 k}{(2\pi)^4} \frac{1}{k^2} \frac{1}{2p_1 \cdot k} \frac{1}{(-2p_2 \cdot k)} \end{aligned} \quad (99)$$

that presents a logarithmic divergency.

The real corrections contain two kinds of divergencies, collinear (or mass, see below) and infrared divergencies, when the emitted parton (gluon) is either collinear to the emitting parton or very soft. They are contained in the gluon emission probability given by the diagram in Fig. 23, that reads in the soft and collinear limit

$$dP \propto \frac{\alpha_s C_R}{\pi} \frac{dx}{x} \frac{dk_\perp^2}{k_\perp^2} \propto \frac{d\omega}{\omega} \frac{d\theta}{\theta}, \quad \theta \simeq \frac{k_\perp}{\omega}, \quad (100)$$

with C_R the quadratic Casimir of the colour representation of the emitting parton.

The singularities can be regularised using dimensional regularisation i.e. working in $d = 4 - 2\varepsilon$ dimensions. The result for the real corrections reads

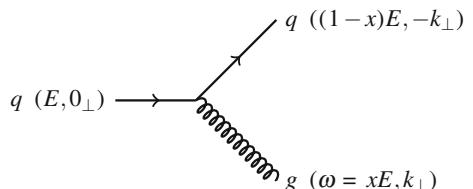
$$\sigma_{real} = \sigma_0 N_c \sum_f Q_f^2 \frac{\alpha_s C_F}{2\pi} H(\varepsilon) \left[\frac{2}{\varepsilon^2} + \frac{3}{\varepsilon} + \frac{19}{2} + \mathcal{O}(\varepsilon) \right], \quad (101)$$

while that for the virtual ones

$$\sigma_{virtual} = \sigma_0 N_c \sum_f Q_f^2 \frac{\alpha_s C_F}{2\pi} H(\varepsilon) \left[-\frac{2}{\varepsilon^2} - \frac{3}{\varepsilon} - 8 + \mathcal{O}(\varepsilon) \right], \quad (102)$$

with $H(\varepsilon) = 1 + \mathcal{O}(\varepsilon)$.

Fig. 23 Feynman diagram giving the gluon emission probability off a quark, with the corresponding kinematical variables



The outcome of this calculation, of general applicability, is that soft divergencies cancel between virtual and real corrections. On the other hand, for sufficient inclusive quantities (e.g. for the total cross section where we sum over initial and final states and thus we cannot distinguish a quark from a quark plus a collinear gluon), collinear divergencies vanish, which is the Kinoshita-Lee-Nauenberg theorem. Thus, this observable is infrared and collinear (IRC) safe and can be computed reliably in perturbation theory. A comparison of data with theoretical calculation can be seen in Fig. 24.

On the other hand, the production of three jets through the emission of an energetic (hard) gluon, in e^+e^- annihilation, see Fig. 25, provides evidence of the existence of the gluon and also of its spin-1 nature. The angular distribution between the most energetic jet in the event and the other two reflects the spin of the emitted gluon, and experimental data clearly favour the spin-1 hypothesis over the spin-0 one. Besides, the existence of non-abelian vertices like ggg produce multi-jet angular distributions that are different from the purely abelian ones $q\bar{q}g$. Again, experimental data demand the existence of non-Abelian vertices and thus provide a test of the non-abelian nature of QCD.

Finally, both the multi-jet distributions and the value of R itself, which receive contributions from radiative corrections, are sensitive to the value and running of α_s and thus allow one of the most precise determinations of this key quantity, see Fig. 7 and [43].

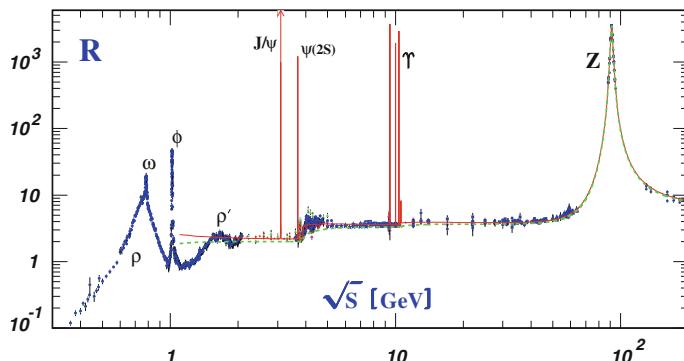


Fig. 24 Compilation of data on R compared to theoretical calculations. (Taken from [42])

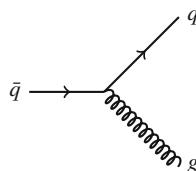


Fig. 25 Illustration of three jet production in e^+e^- annihilation. Each parton will produce a jet

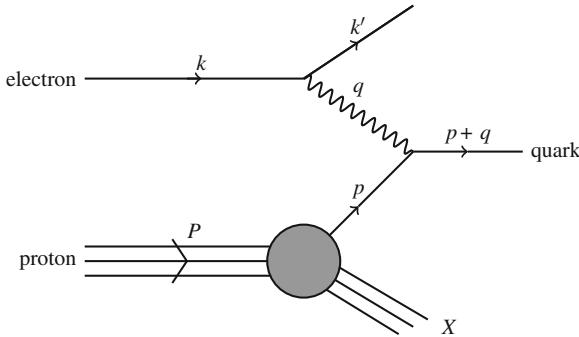


Fig. 26 Feynman diagram for DIS at lowest order in perturbation theory and considering just virtual photon exchange

8.2 Deep Inelastic Scattering

Here we address the analysis of a hadron through a Rutherford-type experiment: the hadron is bombarded by leptons, in a process that is known as Deep Inelastic Scattering (DIS), see Fig. 26. The total cross section is determined by two variables: the scattering angle and the energy, of the outgoing lepton. The variables used to describe this process are

$$\begin{aligned} Q^2 &= -q^2, \\ M^2 &= p^2, \\ v &= p \cdot q = M(E - E'), \\ x &= \frac{Q^2}{2v} = \frac{Q^2}{2M(E - E')}, \\ y &= \frac{p \cdot q}{p \cdot k} = 1 - E'/E, \end{aligned} \quad (103)$$

where the last equalities in the three last lines hold in the reference frame in which the hadron is at rest.

For charged lepton scattering and neglecting electro-weak contributions (i.e. Z^0 or W^\pm exchange), the electromagnetic cross section reads

$$\frac{d^2\sigma}{dxdy} = \frac{8\pi\alpha_{em}^2 ME}{Q^4} \left[\frac{1 + (1-y)^2}{2} 2xF_1 + (1-y)(F_2 - 2xF_1) - \frac{M}{2E} xy F_2 \right], \quad (104)$$

where F_1 and F_2 are the structure functions of the hadron.

The comparison of this expression with the corresponding one for elastic scattering of a lepton on a point-like spin-1/2 particle and on an extended one, leads to the well known conclusions: (i) the experimental fact that, for fixed x , F_1 and F_2 are

roughly independent of Q^2 for large enough Q^2 —*Bjorken scaling*—compared to the $1/Q^4$ behaviour of the proton form factors, leads to the conclusion that the hadron is composed of point-like charged scatterers; and (ii) the experimental fact that $2x F_1 \simeq F_2$ —*Callan-Gross relation*—implies that those scatterers are spin-1/2 particles.

Now, in a frame in which the hadron is moving very fast (the infinite momentum frame IMF), the hadron can be considered as an incoherent superposition of constituents, called partons. This is due to the fact that, for high enough Q , the scale of the interaction between partons, Λ_{QCD} , is much smaller than Q , and constitutes the basis of the *parton model* due to Bjorken, Feynman and Gribov.

In the parton model, the cross section for the DIS process shown in Fig. 26 can be written in the factorised form

$$\begin{aligned} & \sigma[l(k) + \text{hadron}(P) \rightarrow l(k') + p + q + X] \\ &= \int_0^1 d\xi \sum_f f(\xi) \sigma[l(k) + q_f(p = \xi P) \rightarrow l(k') + p + q], \end{aligned} \quad (105)$$

where $f(\xi)$ represents the lower blob in Fig. 26 and can be interpreted as the probability of finding a parton q_f with momentum $p = \xi P$ in the hadron (parton density function PDF), while $\sigma[l(k) + q_f(p = \xi P) \rightarrow l(k') + p + q]$ is the cross section of this parton with the lepton and corresponds to the upper part of the Feynman diagram in the figure.

In the parton model

$$F_2(x) = 2x F_1(x) = \sum_f \int_0^1 d\xi f(\xi) x Q_f^2 \delta(x - \xi) = \sum_f Q_f^2 x f(x), \quad (106)$$

where the sum runs over all (anti)partons in the hadron. In this way, the variable x can be identified with the momentum fraction of the parton in the hadron. Furthermore, if the charged partons are identified with the quarks in the quark model, then

$$F_2^{eN}(x) = \frac{5}{18} F_2^{vN}(x), \quad F_2^{ep}(x) - F_2^{en}(x) = \frac{1}{3} x [u_v(x) - d_v(x)], \quad (107)$$

where N is either a proton p or a neutron n , and u_v and d_v are the PDFs for valence u and d quarks respectively. The experimental verification of these relations supported the identification of the partons in the parton model with the quarks in the quark model and in QCD. Finally, the experimental finding that

$$\int_0^1 dx F_2^{vN}(x) = \int_0^1 dx \sum_{\text{charged } f} x f(x) \simeq 0.44 \quad (108)$$

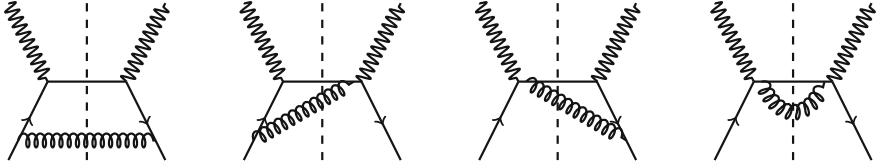


Fig. 27 Squared amplitudes for the QCD corrections to the parton model in the DIS process. The dashed vertical lines correspond to the cut where the amplitude (to the left of the cut) and the complex conjugate amplitude (to the right of the cut) join. All cut particles are on-shell

lead to the conclusion that there is missing momentum in the charged sector of the hadrons, so there exist partons in the hadrons without electric charge, to be identified with the QCD gluons.

Now we turn to the QCD corrections to the parton model, due to the fact that partons: quarks and gluons, radiate. The squared diagrams for the DIS process are given in Fig. 27.

The only diagram that (in light-cone gauge) gives rise to (logarithmic) divergencies is the one on the left in Fig. 27. This logarithmic divergencies lead to an evolution of the parton densities, given by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolutions equations, that read schematically

$$\frac{\partial f(x, Q^2)}{\partial \ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dz}{z} \sum_{f'} P_{f \leftarrow f'} \left(\frac{x}{z} \right) f' (z, Q^2), \quad (109)$$

where the sum runs over all partons f' that may produce parton f with momentum fraction x/z , with probability given by function $P_{f \leftarrow f'} \left(\frac{x}{z} \right)$, known as DGLAP splitting function. It is given by e.g. the diagram in Fig. 23, for the $q \rightarrow g$ splitting function with $x \rightarrow x/z$ and for the $q \rightarrow q$ splitting function with $1 - x \rightarrow x/z$.

While the exact expressions of the evolutions equations and splitting functions (available up to next-to-next-to-leading order) can be found elsewhere, let us discuss their physical origin slightly more in depth. For that, consider both initial state real and virtual corrections to the DIS process, as shown in Figs. 28 and 29.

Fig. 28 Feynman diagram for initial state real corrections to DIS at lowest order in perturbation theory

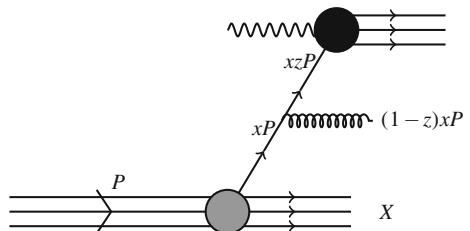
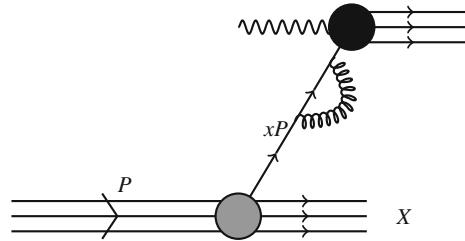


Fig. 29 Feynman diagram for initial state virtual corrections to DIS at lowest order in perturbation theory



Both contributions are IR and collinearly divergent:

$$\sigma_{real} \simeq \sigma_h(xzP) \frac{\alpha_s C_F}{\pi} \frac{dz}{1-z} \frac{dk_\perp^2}{k_\perp^2}, \quad (110)$$

$$\sigma_{virtual} \simeq -\sigma_h(xP) \frac{\alpha_s C_F}{\pi} \frac{dz}{1-z} \frac{dk_\perp^2}{k_\perp^2}, \quad (111)$$

where σ_h denotes the upper black blob representing the γ^* -parton interaction in Figs. 28 and 29. They combine into an IR finite but collinearly divergent cross section,

$$\sigma_{real+virtual} \simeq \frac{\alpha_s C_F}{\pi} \int_0^{Q^2} \frac{dk_\perp^2}{k_\perp^2} \int_0^1 \frac{dz}{1-z} [\sigma_h(xzP) - \sigma_h(xP)], \quad (112)$$

where the soft divergence at $z = 1$ cancels and this cancellation appears as the so-called ‘+’ prescription,

$$\int_0^1 \frac{dz}{(1-z)_+} f(z) \equiv \int_0^1 \frac{dz}{1-z} [f(z) - f(1)]. \quad (113)$$

The collinear divergence in (112) is regulated through a momentum cut-off, called factorisation scale μ_F , and absorbed into the definition of the parton densities:

$$\sigma_{real+virtual} \simeq \frac{\alpha_s C_F}{\pi} \int_{\mu_F^2}^{Q^2} \frac{dk_\perp^2}{k_\perp^2} \int \frac{dx dz}{1-z} [\sigma_h(xzP) - \sigma_h(xP)] q(x, \mu_F^2). \quad (114)$$

This cut-off is arbitrary, so the final physical results must be independent of its value. This gives rise to the DGLAP evolution of the quark densities, which can be seen as the evolution due to radiation between some scale μ_F^2 below the real and virtual emission in Figs. 28 and 29 and some infinitesimal increase $(1+\varepsilon)\mu_F^2$ above.

The result reads

$$\frac{dq(x, \mu_F^2)}{d \ln \mu_F^2} = \int_x^1 \frac{dz}{z} P_{q \leftarrow q}(z) q(x/z, \mu_F^2), \quad P_{q \leftarrow q}(z) = \left(\frac{1+z^2}{1-z} \right)_+. \quad (115)$$

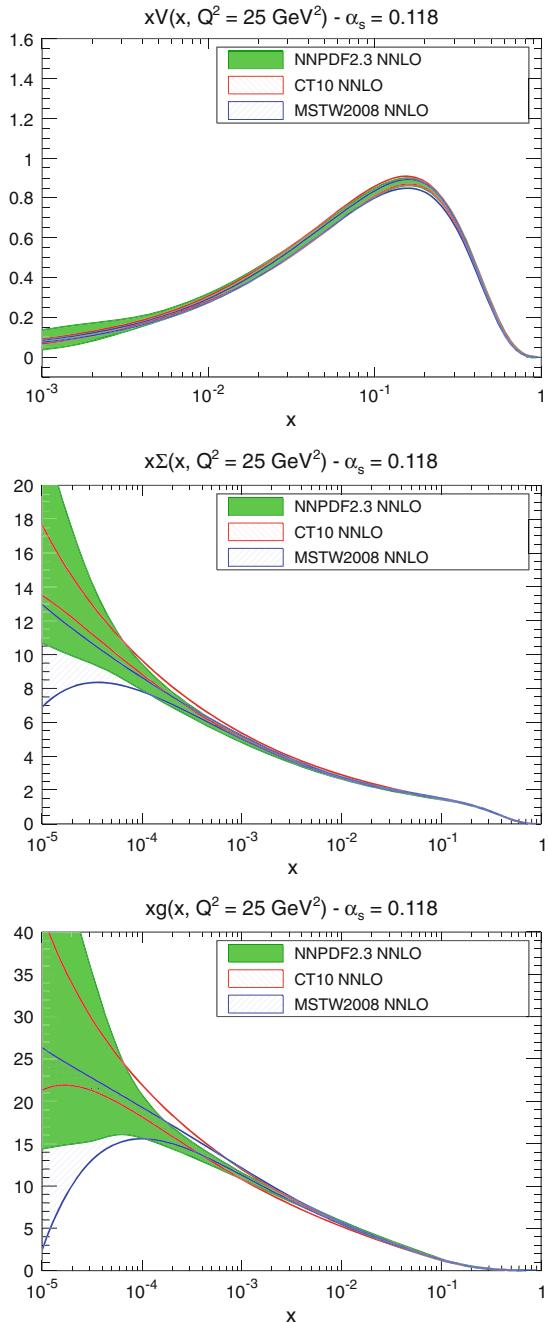
The DGLAP evolution equations are the tools used nowadays to extract the parton densities in proton and nuclei. These quantities are non-perturbative, but their evolution with the factorisation scale, which in DIS is naturally identified with the photon virtuality Q^2 , is controlled by the equations. So the following iterative procedure is used:

1. The parton densities are parametrised at some given low scale Q_0 using some functional form that depends on a series of parameters. Different groups use different functional forms, number of parameters and assumptions in order to reduce the number of parameters, besides momentum and baryon number sum rules. The scale Q_0 has to be low compared with the available range of scales in the experimental data, but larger than Λ_{QCD} in order to apply perturbation theory. A given initial choice of parameters is done.
2. DGLAP evolution is applied to these parton densities.
3. Using the parton densities, now available at all required scales, all cross sections are computed that are believed to be describable in perturbative QCD and for which experimental data exist.
4. Some comparison to experimental data is done, using some criterium as χ^2 , likelihood, ...
5. If such criterium is found to be optimal e.g. a minimum of the χ^2 , then the set of parameters is considered to give the best fit and standard error analysis is used to compute the uncertainties in the extracted parton densities. If no, we go back to 1., varying the initial parameters and the procedure is repeated.

Several groups perform this technique. In Fig. 30, some comparison of the results is shown, taken from [44]. The results show an impressive agreement to data, as shown in Fig. 31, providing one of the most stringent tests of perturbative QCD and an analysis of the proton that shows no evidence of substructure down to lengths of order 10^{-4} fm. For the nuclear case, the same procedure is applied, producing similar results that show clearly that the partons densities inside nuclei cannot be considered as the mere superposition of those in protons and neutrons, see Fig. 32. The only limitation in the nuclear case comes from the scarce data, see e.g. [45].

While this procedure is evidently successful, one comment is in order: in the DGLAP evolution equations the variation of the parton densities is proportional to the parton densities themselves, leading to a strong growth of them towards small x . Such growth strongly suggest that, at some very small x , non-linear phenomena should appear. In fact, there is strong theoretical evidence that it must be so, see e.g. [47, 48] and references therein. There exist realisations of QCD in which these non-linear phenomena are describable in the weak coupling regime and thus by perturbative techniques. Actually, presently available DIS data can be described in

Fig. 30 Comparison of different parton densities for valence quarks (*top*), sea quarks (*middle*) and gluons (*bottom*), extracted using DGLAP analyses. (Taken from [44])



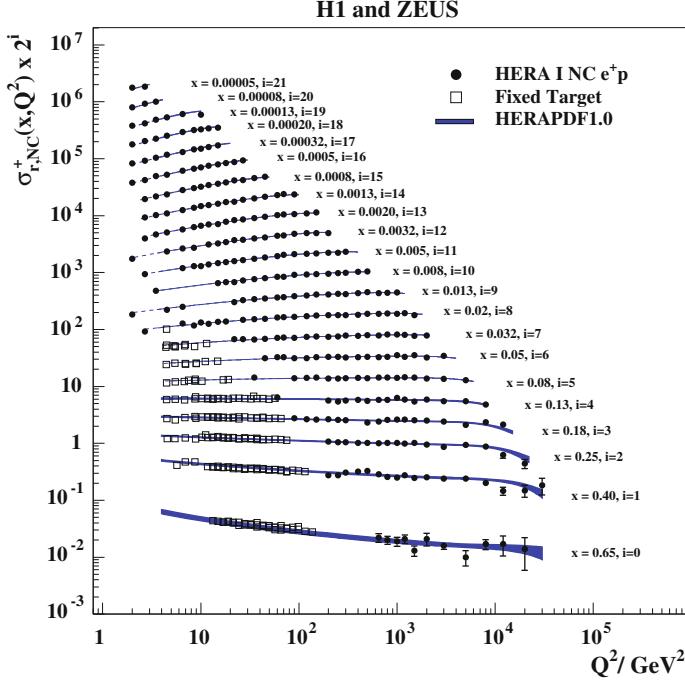


Fig. 31 Comparison of the results of a DGLAP analysis to DIS experimental data. (Taken from [46])

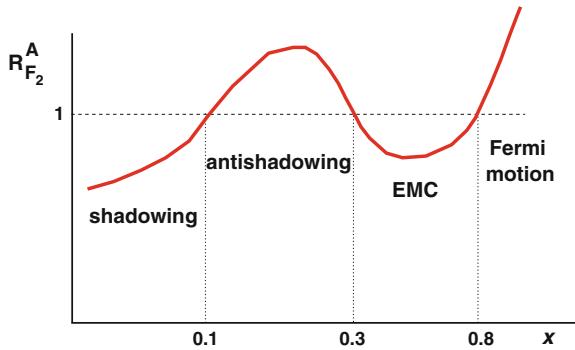


Fig. 32 Schematic plot showing the different regions in the ratio of structure functions and parton densities in nuclei over those in a nucleon. (Taken from [45])

fixed-order perturbation theory (DGLAP evolution), in schemes that resum not only collinear logarithms like DGLAP does but also soft logarithms present in the splitting kernels, see (100), and non-linear approaches. The present experimental debate lies on where in the kinematic $x - Q^2$ plane the existence of effects beyond fixed-order

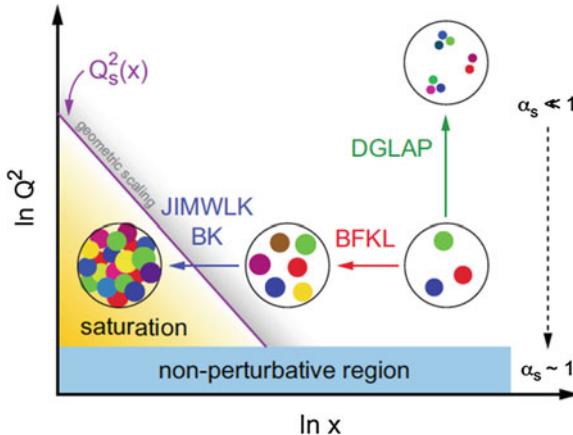


Fig. 33 Schematic plot of hadron structure showing the linear and non-linear (saturation) regions. (Taken from [47])

perturbation theory becomes mandatory. Figure 33 shows a sketch of the hadron structure in this respect.

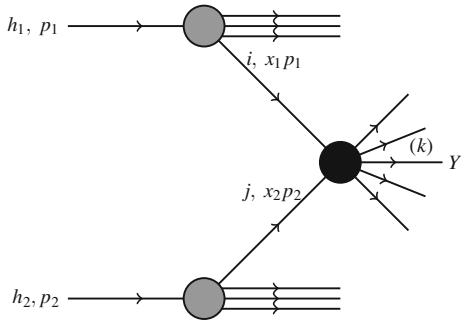
8.3 QCD in Hadronic Collisions: Factorisation

Factorisation is the key tool that allows to perform calculations for physical observables in perturbative QCD. It is illustrated in Fig. 34 for hadronic collisions. Mathematically, the cross section in a hadronic collision to produce observable Y is given by

$$\sigma(h_1(p_1)+h_2(p_2) \rightarrow Y) = \int dx_1 \int dx_2 \sum_{i,j} f_i(x_1) f_j(x_2) \hat{\sigma}(i+j \rightarrow Y) + \mathcal{O}\left(\frac{1}{S^n}\right), \quad (116)$$

where $f_{i(j)}(x_{1(2)})$ (grey blobs in Fig. 34) is the distribution of parton $i(j)$ to take momentum fraction $x_{1(2)}$ from hadron $h_{1(2)}$, $\hat{\sigma}(i+j \rightarrow Y)$ (black blob in Fig. 34) is the cross section to produce observable Y from partons i and j , and observable Y may come directly from $\hat{\sigma}$ (e.g. being a Higgs or electro-weak boson, a photon, . . .), or through a parton k (then containing a jet reconstruction algorithm, a fragmentation function for the projection of the parton onto a hadron, . . .); PDFs f (and fragmentation functions) contain a factorisation scale, and $\hat{\sigma}(i+j \rightarrow Y)$ a renormalisation scale and is computable in perturbation theory; and the corrections are power-suppressed in terms of a given scale S that is related with the observable, being its mass, energy, transverse momentum, . . .

Fig. 34 Illustration of factorisation in hadronic collisions



When the scale S is close to the total available energy in the centre-of-mass, this kind of factorisation is called collinear, and the evolution of PDFs and fragmentation functions is given by the DGLAP equations. The origin of collinear factorisation lies in the separation of scales between the large scale S of the event and the soft scale Λ_{QCD} , that allows a division between short distance pieces (the hard scattering cross section $\hat{\sigma}$ computable order by order in perturbation theory and available at next-to-leading order for all observables of interest, and in many cases at next-to-next-to-leading order), and long distance ones (the PDFs and fragmentation functions). This factorisation has been proved, for inclusive enough observables like 1-particle or 2-jet production and for specific kinematic configurations, in e^+e^- annihilation, in DIS and for the Drell-Yan process. It is assumed to work in hadron-hadron collisions.

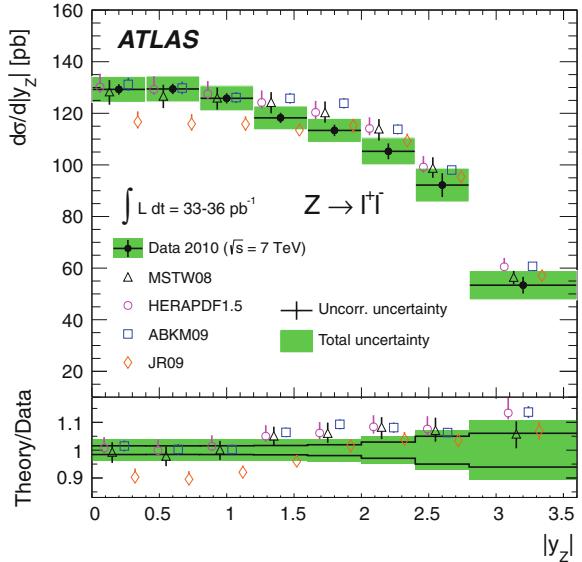
It should be noted that large logarithms may appear for given kinematic configurations that need to be resummed, and that other kinds of factorisation exist like k_T -factorisation for $E_{cm} \gg S \gg \Lambda_{QCD}$.

In Fig. 35 we show one comparison of perturbative calculations with hadron-hadron data at the LHC. It can be seen that the experimental results are sensitive to the differences among the different PDF sets. In fact, this is the main source of uncertainty in the extrapolations for precision measurements and searches for beyond-the-Standard-Model physics in the proposed high-luminosity phase of the LHC [49].

8.4 QCD Radiation

Radiation determines most of the features in perturbative QCD. Let us start by a single emitter, as pictured in Fig. 23. Obviously, an isolated on-shell massless particle cannot radiate. Some virtuality, of order $k_\perp^2/[x(1-x)]$, has to be allowed for radiation to occur. The process of multiple gluon emission proceeds through the iteration of emission given in the soft and collinear limits by the probability in (100). This iteration gives rise to large logarithms in the limit in which subsequent emissions $i = 1, 2, 3, \dots$ are strongly ordered in both longitudinal and transverse momentum i.e.

Fig. 35 Rapidity distribution of Z^0 bosons decaying into l^+l^- measured by the ATLAS Collaboration at the LHC, compared to next-to-next-to-leading order results using different PDFs. (Taken from [50])



$$x_1 \gg x_2 \gg \dots \gg x_n, \quad k_{\perp,1} \ll k_{\perp,2} \ll \dots \ll k_{\perp,n} \quad (117)$$

(for initial state radiation; for final state radiation, the ordering in k_\perp is the opposite, from a high to a low virtuality). These orderings gives rise to powers of two types of logarithms, of x and of k_\perp , and this limit is known as the double-leading-logarithmic limit.

On the other hand, a single massive particle cannot emit collinearly, a phenomenon known as dead cone effect, exemplified in Fig. 36.

Now we consider two emitters: two highly energetic quark and antiquark originated at some given position and with angular separation $\theta_{p\bar{p}}$. In the limit of soft

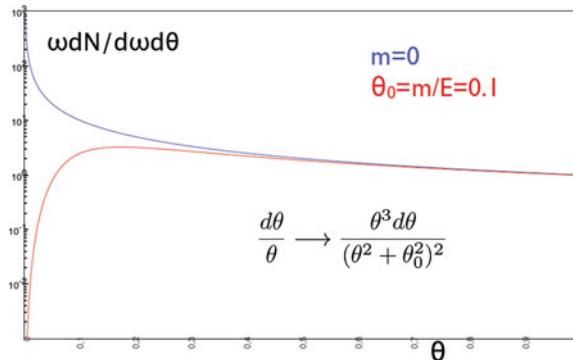
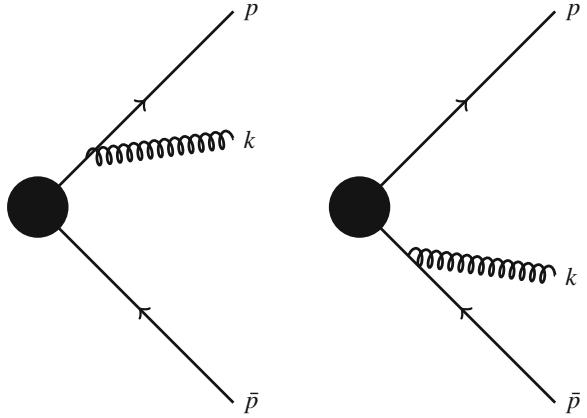


Fig. 36 Illustration of the difference of the gluon radiation spectrum off massless and massive quarks versus the emission angle for fixed gluon energy, in arbitrary units

Fig. 37 Diagrams contributing to gluon emission off a $q\bar{q}$ antenna in the soft limit



emitted gluons, the diagrams that contribute to gluon radiation at the lowest order are given in Fig. 37. The gluon radiation spectrum is given by

$$(2\pi)^2 E \frac{dN}{d^3k} = \alpha_s C_F \frac{p \cdot \bar{p}}{(p \cdot k)(\bar{p} \cdot k)} . \quad (118)$$

It turns out that the spectrum can be separated in parts that contain collinear divergences either from the quark or from the antiquark, $dN = dN_q + dN_{\bar{q}}$. After azimuthal average, the gluon spectrum off the quark (or antiquark) takes the form

$$dN_q \propto \alpha_s \frac{d\omega}{\omega} \frac{d\theta_{pk}}{\theta_{pk}} \Theta(\theta_{p\bar{p}} - \theta_{pk}) , \quad (119)$$

where Θ is the step function. In this way, quantum interferences provide a probabilistic picture of gluon emission, and this gluon emission features angular ordering: gluon emission at angles larger than the opening angle of the $q\bar{q}$ pair are suppressed. The intuitive explanation of this phenomenon is as follows (we use small angle approximations): The decoherence time of the emitted gluon (i.e. the time it takes to decohere the gluon from the $q\bar{q}$ pair) is $t_c \sim \omega/k_\perp^2$. The transverse resolution of the gluon is $D_g \sim 1/k_\perp$. The separation of the $q\bar{q}$ pair at decoherence time is $D_{q\bar{q}} = \theta_{p\bar{p}} t_c = \theta_{p\bar{p}} / (k_\perp \theta_{pk})$. Then the condition that, in order to be radiated, the gluon should resolve the $q\bar{q}$ pair in order to ‘see’ a non-zero colour charge leads to $D_{q\bar{q}} > D_g \Rightarrow \theta_{pk} < \theta_{p\bar{p}}$. While this reasoning is restricted to a colour antenna in the singlet representation, angular ordering holds for arbitrary colour representations: When the antenna is a non-singlet colour state, radiation outside the cone, $\theta_{pk} > \theta_{p\bar{p}}$, happens with the strength given by the total colour charge i.e. the pair acts as a single emitter with the charge of the pair—the charge of the parent parton of the $q\bar{q}$ pair. In this sense, a simple probabilistic picture emerges, which is implemented e.g. in Monte Carlo generators: radiation inside the cone determined by the pair opening angle, $\theta < \theta_{p\bar{p}}$, takes place as independent radiation off each of the partons, while

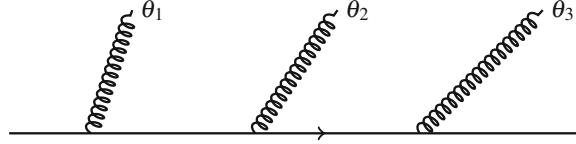


Fig. 38 Multiple gluon emission off a quark with different emission angles, showing angular ordering between subsequent emissions

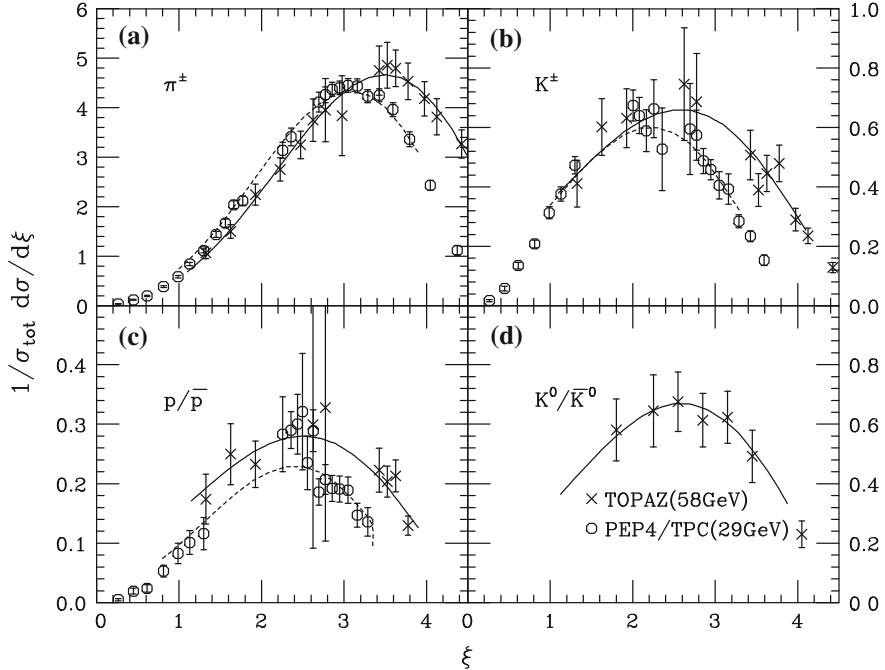


Fig. 39 Comparison of experimental data on particle distribution in $\xi \simeq \ln(E_{jet}/E_{hadron})$. Large ξ corresponds to soft particle production. (Taken from [52])

radiation outside the cone, $\theta > \theta_{p\bar{p}}$, can be reinterpreted as radiation off the parent parton, so that the angular ordering can be effectively introduced.

It should be noted that coherence, through the restriction in the phase space for multiple gluon emission (Fig. 38) that angular ordering provides, implies a strong reduction of particle multiplicities inside jets (particles produced by radiation off a highly energetic parton) in the soft sector compared to successive independent, incoherent emissions. The agreement that angular ordering yields can be seen in Fig. 39. It also provides the basis for a probabilistic interpretation of parton branching that lies at the root of existing implementation of the QCD showering process in Monte Carlo simulators [51], with different ordering variables for the development in the cascade that at high energies are all of them equivalent to emission angle.

It is obvious that, in a Quantum Field Theory like QCD, we would like to compute diagrams with any number of external legs that would give the cross section for production of any number of partons. This is not possible for a large number even a tree level, and constitutes a subject on its own. But the structure of coherent QCD radiation provides a probabilistic picture that allows a sequential treatment of the branching process: iteration of emission kernel at some give value of the ordering variable $t_1 \times$ probability of no resolvable emission (called Sudakov form factor) between t_1 and $t_2 \times$ emission kernel at $t_2 \times \dots$, that should give the dominant contribution to multiplicities of produced partons.

8.5 Jets

In high-energy collisions, the strong interaction tends to produce collimated showers of hadrons, called jets, see [53] and references therein for full information. This is due to several characteristics of QCD: collinear singularities of the emission probabilities, gluon self-interaction, coherence between emitters and, on a more fundamental level, asymptotic freedom and confinement. In principle, in perturbation theory one would like to be able to work with partons but, even at a perturbative level, partons are ill-defined concepts due to the existence of IR and collinear divergences that make it impossible to distinguish between one parton and two collinear ones, or one hard and one very soft parton.

Therefore, jets require a definition and, in order to make it possible a meaningful comparison between experiment and theory, the definition (the counting of the number of jets in an event and their energy) should be:

1. Stable at all order in perturbation theory under the splitting of a parton into two collinear ones or into one very hard and one very soft.
2. Insensitive to the parton-to-hadron transition, so it should be defined in terms of flows of energy i.e. calorimetric measurements.

The problem can be seen in the fact that the definition should be such that the cancellation of singularities between virtual and real corrections, see Fig. 40, happens in order to render a finite cross section.

Jets are defined in the plane determined by the pseudo rapidity and azimuthal angle. Simplifying, there are two types of algorithms for jet reconstruction. One kind of them uses a seed in order to start clustering particles into jets, thus privileging a particle. This kind of algorithms, with many different variations, are IRC unsafe due to an incomplete cancellation between real and virtual corrections, and present problems for the comparison between theory and experiment. They can be made finite at some given order of perturbation theory, but they fail at the next order or when using resummations. This is illustrated in Figs. 41 and 42.

It turns out that seedless or sequential recombination algorithms are IRC safe. Let us briefly define how they work (in the basic implementation). For a list of particles with η_i , $p_{\perp,i}$, ϕ_i , $i = 1, 2, \dots, N$, we define

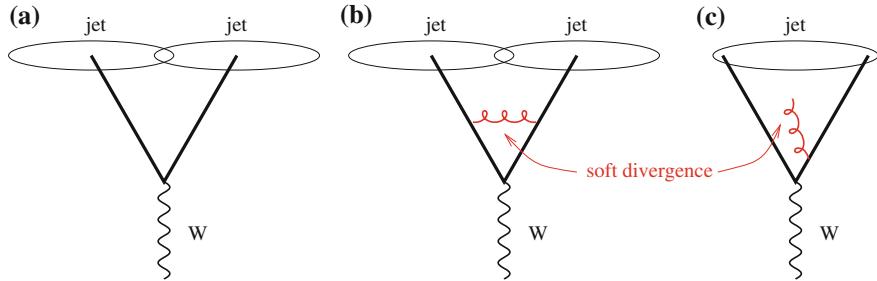


Fig. 40 W decaying into two quark jets (a), the virtual (b) and real (c) next-to-leading corrections whose singularities must cancel in order to make a final finite result. (Taken from [53])

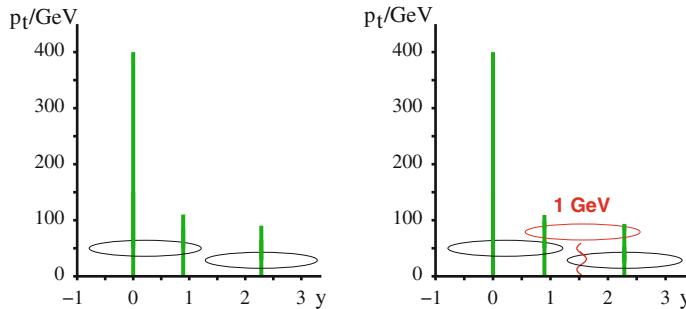


Fig. 41 Exemplification of how the addition of one soft particle (red wavy line) varies the counting of jets between the plot on the *left* and the plot on the *right*, thus showing the problems of algorithms that use a seed. The hardest particle of the event is used as a seed and the value of the parameters inside which particles are merged into jet is smaller than the distance between the first and third particle (from *left* to *right*) in the plot on the *left*. (Taken from [53])

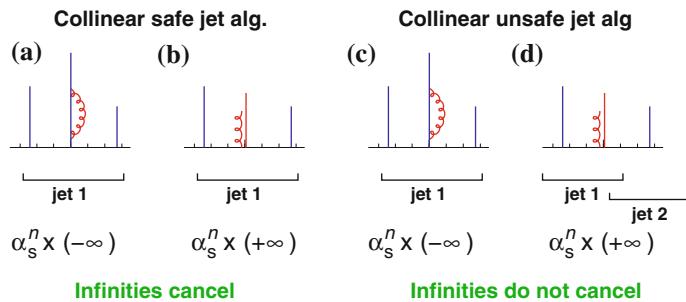


Fig. 42 *Plots on the left* example of how the addition of an collinear gluons does not alter the counting of jets for seedless algorithms. *Plots on the right* example of the contrary in an algorithm that uses a seed, in this case the hardest particle in the event. (Taken from [53])

$$d_{iB} = p_{\perp,i}^{2n}, \quad d_{ij} = \min \left(p_{\perp,i}^{2n}, p_{\perp,j}^{2n} \right) \frac{\Delta R_{ij}^2}{R^2}, \quad \Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2. \quad (120)$$

Now we proceed sequentially:

1. Find the minimum of all $\{d_{iB}, d_{ij}\}$.
2. If it is a d_{iB} , remove particle i , that it is called a jet, from the list.
3. If it is a d_{ij} , merge i and j and redo the list.
4. If the list is not exhausted, go back to 1.

All in all, in this simplest variant any algorithm is defined by the prescription of merging particles (the most widely used nowadays is adding the four-momentum of the particles) and by the value of R . Different values of n produce different algorithms: $n = 1$ gives the so-called k_T algorithm, $n = 0$ Cambridge/Aachen, and $n = -1$ anti- k_T . The latter is the most commonly employed option at the LHC, and gives rise to quite ‘conical’ jets, see Fig. 43 for an illustration. In the past, sequential recombination algorithms were very expensive computationally which prevented their application for hadronic collisions, but new developments have allowed large improvements in computing time, making them practical tools of wide use at the LHC.

One very important aspect is the energy of the parton that is reconstructed when defining a ‘conical’ jet of radius R (note that while the definition of a jet as a cone in the $\eta \times \phi$ space is not, in principle, IRC safe, there exist seedless cone algorithms that are IRC safe and that jets defined through the anti- k_T algorithm have a conical shape). The difference between the energy of the jet and that of the parton receives 3 contributions:

$$\langle p_{\perp,\text{jet}} - p_{\perp,\text{parton}} \rangle = ap_{\perp} \alpha_s \ln R - \frac{b}{R} + cR^2. \quad (121)$$

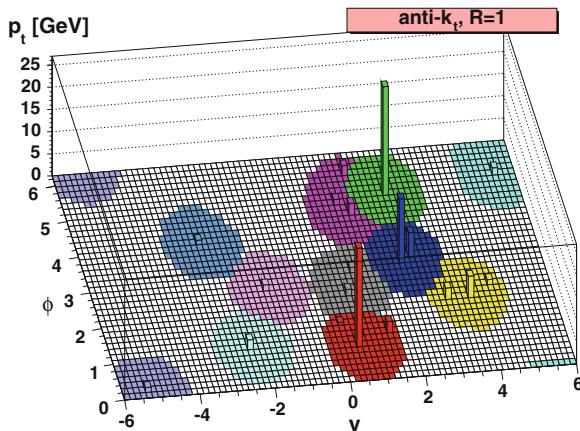


Fig. 43 Illustration of the jets in a pp event and their shapes in the $\eta \times \phi$ plane, defined using the anti- k_T algorithm with $R = 1$. (Taken from [53])

The three contributions to the r.h.s. of this equation are interpreted as follows:

- The first contribution has a radiative origin and corresponds to the energy (p_{\perp}) of the parton that is sent outside the cone by QCD radiation. Thus the presence of the logarithm and of the coupling constant. The constant a is computable using perturbative techniques.
- The second contribution has a non-perturbative origin and accounts for energy of the parton that is sent outside the cone by hadronisation effects. It is usually estimated using Monte Carlo models that consider hadronisation of partons into hadrons.
- The third contribution accounts for the background—the contribution from particles in the collisions that do not come from the parton that originates the jet—that may have three origins: additional soft or hard parton-parton interactions in the collision to the one that originates the parton that produces the jet (the so-called underlying event); additional nucleon-nucleon collisions in the case of nuclear projectiles and targets (called underlying event or heavy-ion background, although it does not require a truly heavy ion); and additional projectile-target collisions that come from the fact that such possibility exists when bunches of particles cross in the interaction regions of the accelerator (so-called pile-up, that maybe as high as 100 in pp collisions at the LHC). Although a strict distinction between particles in the jet and in the background is senseless from both a practical and a theoretical point of view, techniques have been developed to deal with this problem that may be crucial for extracting information off jet measurements. Note that in a head-on PbPb collision at the LHC, the background contribution may be as high as $\mathcal{O}(100 \text{ GeV})$ per unit area in the $\eta \times \phi$ plane.

Jet measurements provide one of the most impressive checks of perturbative QCD, see the agreement between experimental data and theoretical calculations at next-to-leading order in Fig. 44 that extends through several orders of magnitude in the measured cross sections.

Finally, note that jets are not only very interesting objects in QCD, but also most required tools in other topics as Standard Model or Beyond the Standard Model studies. For example, jet reconstruction is required for determining the couplings of the Higgs boson to quarks, couplings that may not only verify their expected dependency with the quark mass in the Standard Model but also be sensitive to new physics. Jets are also required to reconstruct hadronic decays of electro-weak bosons.

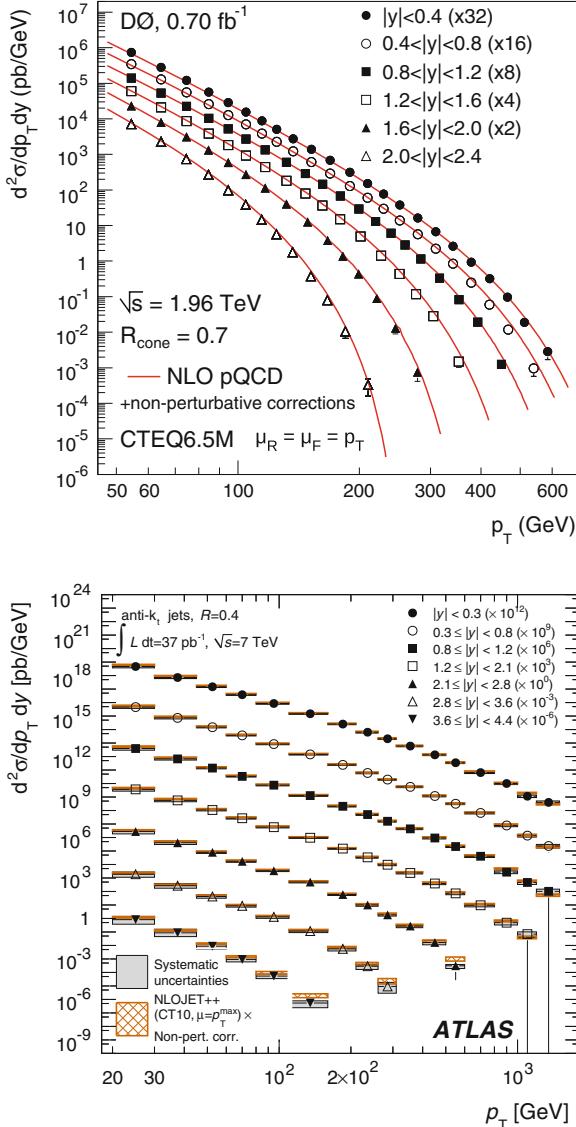


Fig. 44 Comparison of experimental data on single-inclusive jet cross sections at different pseudo-rapidities to perturbative QCD calculations at next-to-leading order. *Top* data from the DO Collaboration at the Tevatron (taken from [54].) *Bottom* data from the ATLAS Collaboration at the LHC (taken from [55])

Acknowledgments We thank Tolga Altinoluk for help in producing Feynman diagrams. This work is supported by the European Research Council grant HotLHC ERC-2011-StG-279579; by Ministerio de Ciencia e Innovación of Spain under projects FPA2009-11951 and FPA2011-22776; by Xunta de Galicia (Consellería de Educación and Consellería de Innovación e Industria—Programa Incite); by the Spanish Consolider-Ingenio 2010 Programme CPAN and by FEDER.

References

1. D.J. Gross, F. Wilczek, Phys. Rev. Lett. **30**, 1343–1346 (1973)
2. H.D. Politzer, Phys. Rev. Lett. **30**, 1346–1349 (1973)
3. F.J. Yndurain, *The Theory of Quark and Gluon Interactions* (Springer, Berlin, 1999)
4. P. Pascual, R. Tarrach, *QCD: Renormalization for the Practitioner* (Springer, New York, 1984)
5. A.V. Smilga, *Lectures on Quantum Chromodynamics* (World Scientific, London, 2001)
6. B.L. Ioffe, V.S. Fadin, L.N. Lipatov, *Quantum Chromodynamics: Perturbative and Nonperturbative Aspects* (Cambridge University Press, Cambridge, 2010)
7. R.K. Ellis, W.J. Stirling, B.R. Webber, *QCD and Collider Physics* (Cambridge University Press, Cambridge, 2003)
8. E. Leader, E. Predazzi, *An Introduction to Gauge Theories and Modern Particle Physics*, vols. 1 and 2 (Cambridge University Press, Cambridge, 1996)
9. Y.L. Dokshitzer, V.A. Khoze, A.H. Mueller, S.I. Troyan, *Basics of Perturbative QCD* (Editions Frontières, Gif-sur-Yvette, 1991)
10. H. Fritzsch, M. Gell-Mann, H. Leutwyler, Phys. Lett. **B47**, 365–368 (1973)
11. S. Bethke, Eur. Phys. J. **C64**, 689–703 (2009)
12. R. Alkofer, J. Greensite, J. Phys. G **34**, S3 (2007). [arXiv:hep-ph/0610365](https://arxiv.org/abs/hep-ph/0610365)
13. J. Greensite, Lect. Notes Phys. **821**, 1 (2011)
14. Z. Fodor, C. Hoelbling, Rev. Mod. Phys. **84**, 449 (2012). [arXiv:1203.4789](https://arxiv.org/abs/1203.4789) [hep-lat]
15. G.F. Sterman, S. Weinberg, Phys. Rev. Lett. **39**, 1436 (1977)
16. A. Dobado, A. Gomez-Nicola, A.L. Maroto, J.R. Pelaez, *Effective Lagrangians for the Standard Model* (Springer, Berlin, 1997)
17. Y. Nambu, Phys. Rev. Lett. **4**, 380 (1960)
18. J. Goldstone, Nuovo Cimento **19**, 154 (1961)
19. H.-Y. Cheng, Phys. Rep. **158**, 1 (1988)
20. R.D. Peccei, H.R. Quinn, Phys. Rev. D **16**, 1791 (1977)
21. G. 't Hooft, Phys. Rev. Lett. **37**, 8 (1976)
22. G. 't Hooft, Phys. Rep. **142**, 357 (1986)
23. T.D. Lee, G.C. Wick, Phys. Rev. D **9**, 2291 (1974)
24. J.C. Collins, M.J. Perry, Phys. Rev. Lett. **34**, 1353 (1975)
25. J. Hofmann, H. Stöcker, W. Scheid, W. Greiner, in *Report of the Workshop on BeV/Nucleon Collisions of Heavy Ions: How and Why*, Bear Mountain, New York, 29 Nov–1 Dec 1974
26. B.A. Freedman, L.D. McLerran, Phys. Rev. D **16**, 1169 (1977)
27. N. Cabibbo, G. Parisi, Phys. Lett. B **59**, 67 (1975)
28. E.V. Shuryak, Phys. Lett. B **78**, 150 (1978)
29. E.V. Shuryak, Sov. J. Nucl. Phys. **28**, 408 (1978)
30. H. Satz, Lect. Notes Phys. **841**, 1 (2012)
31. F. Karsch, Lect. Notes Phys. **583**, 209 (2002) [arXiv:0711.0656](https://arxiv.org/abs/0711.0656) [hep-lat]
32. F. Karsch, J. Phys. Conf. Ser. **46**, 122 (2006) [arXiv:0711.0661](https://arxiv.org/abs/0711.0661) [hep-lat]
33. J.D. Bjorken, Phys. Rev. D **27**, 140 (1983)
34. C.-Y. Wong, *Introduction to High-energy Heavy-ion Collisions* (World Scientific, Singapore, 1994)
35. R. Hwa (ed.), *Quark-Gluon Plasma*, vol. 1 (World Scientific, Singapore, 1990)
36. R. Hwa (ed.), *Quark-Gluon Plasma*, vol. 2 (World Scientific, Singapore, 1995)

37. R. Hwa, X.N. Wang (eds.), *Quark-Gluon Plasma*, vol. 3 (World Scientific, Singapore, 2004)
38. R. Hwa, X.N. Wang (eds.), *Quark-Gluon Plasma*, vol. 4 (World Scientific, Singapore, 2010)
39. C. Pajares, Yu.M. Shabelski, *Relativistic Nuclear Interactions* (URSS, Moscow, 2007)
40. J.-P. Blaizot, Lect. Notes Phys. **583**, 117 (2002). [arXiv:hep-ph/0107131](https://arxiv.org/abs/hep-ph/0107131) [hep-ph]
41. G.P. Salam, CERN Yellow Report CERN-2010-002, 45–100. [arXiv:011.5131](https://arxiv.org/abs/011.5131) [hep-ph]
42. V.V. Ezhela, S.B. Lugovsky, O.V. Zenin, [arXiv:hep-ph/0312114](https://arxiv.org/abs/hep-ph/0312114)
43. J. Beringer et al., [Particle Data Group Collaboration], Phys. Rev. D **86**, 010001 (2012)
44. R.D. Ball, S. Carrazza, L. Del Debbio, S. Forte, J. Gao, N. Hartland, J. Huston, P. Nadolsky et al., JHEP **1304**, 125 (2013). [arXiv:1211.5142](https://arxiv.org/abs/1211.5142) [hep-ph]
45. N. Armesto, J. Phys. G **32**, R367 (2006). [arXiv:hep-ph/0604108](https://arxiv.org/abs/hep-ph/0604108)
46. F.D. Aaron et al. [H1 and ZEUS Collaboration], JHEP **1001**, 109 (2010). [arXiv:0911.0884](https://arxiv.org/abs/0911.0884) [hep-ex]
47. A. Accardi, J.L. Albacete, M. Anselmino, N. Armesto, E.C. Aschenauer, A. Bacchetta, D. Boer, W. Brooks et al., [arXiv:1212.1701](https://arxiv.org/abs/1212.1701) [nucl-ex]
48. Yu.V. Kovchegov, E.M. Levin, *Quantum Chromodynamics at High Energy* (Cambridge University Press, Cambridge, 2012)
49. ATLAS Collaboration, ATL-PHYS-PUB-2012-004
50. G. Aad et al., [ATLAS Collaboration], Phys. Rev. D **85**, 072004 (2012). [arXiv:1109.5141](https://arxiv.org/abs/1109.5141) [hep-ex]
51. A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Hoche, H. Hoeth, F. Krauss, L. Lonnblad et al., Phys. Rep. **504**, 145 (2011). [arXiv:1101.2599](https://arxiv.org/abs/1101.2599) [hep-ph]
52. R. Itoh et al., [TOPAZ Collaboration], Phys. Lett. B **345**, 335 (1995). [arXiv:hep-ex/9412015](https://arxiv.org/abs/hep-ex/9412015)
53. G.P. Salam, Eur. Phys. J. C **67**, 637 (2010). [arXiv:0906.1833](https://arxiv.org/abs/0906.1833) [hep-ph]
54. V.M. Abazov et al., [D0 Collaboration], Phys. Rev. D **85**, 052006 (2012). [arXiv:1110.3771](https://arxiv.org/abs/1110.3771) [hep-ex]
55. G. Aad et al., [ATLAS Collaboration], Phys. Rev. D **86**, 014022 (2012). [arXiv:1112.6297](https://arxiv.org/abs/1112.6297) [hep-ex]



The Rector of the University of Santiago de Compostela, Professor Juan J. Casares Long, and Professor Mateusz Płoskon talk during the reception offered by the Rector to the participants of the Third IDPASC School held at the Pazo de San Xerome (picture by A. Ponte)

Selected Topics from the Experiments with High-Energy Heavy-Ion Collisions: Hot QCD in the Laboratory

Mateusz Płoskoń

Abstract The main aspects of the physics of heavy-ion collisions are presented, as well as the fundamental features of the study of hot QCD in the laboratory. Thus, it is first included a detailed description of the soft probes of quark-gluon plasma (QGP) formation (centrality of a heavy-ion collision, particle production, particle identification, particle spectra, particle yields at freeze-out, freeze-out temperature and thermal models, the radial flow, baryon anomaly, and elliptic flow). Also, the current state-of-the-art of the hard probes of QGP is exposed (jet quenching with hadron observables, particle correlations at high- p_T , jet quenching with fully reconstructed jets, experimental aspects of jet reconstruction in heavy-ion collisions, jets in AA collisions at the LHC, jet fragmentation, energy loss of the heavy-flavor, and thermalization of charm). Finally, the important questions connected to quarkonia ($J/\psi R_{AA}$ at RHIC and the LHC, sequential suppression of quarkonia, $J/\psi v_2$, and cold nuclear matter effects) are discussed.

1 Heavy-Ion Collisions: Hot QCD in the Laboratory

Quantum Chromodynamics (QCD) solved on the lattice predicts a phase transition from normal nuclear matter to the state of deconfined quarks and gluons [1]. The so-called quark-gluon plasma (QGP) is created once the temperature raises above a critical temperature T_c of about 150 MeV at zero baryo-chemical potential, or density of the system is larger than about 0.5 GeV/fm³. Such conditions are often referred to as extreme: in units commonly used in daily life we would have to deal with densities of larger than 10¹⁵ g/cm³ and temperatures beyond 10¹² K. The considerations of the evolution of our Universe point out that such a QGP state could have existed microseconds after the Big Bang [2]. Nowadays, the matter under such extreme conditions can be created and studied in the laboratory by colliding heavy-nuclei at ultra-relativistic energies. Such collisions provide a unique opportunity for studying

M. Płoskoń (✉)

Lawrence Berkeley National Laboratory, Berkeley, CA, USA
e-mail: mploskon@lbl.gov

how the physical properties of the strongly interacting non-abelian partonic matter emerge from the fundamental interactions of QCD.

High-energy heavy ion collisions have been studied experimentally in the last decades at increasing center-of-mass energies at the Brookhaven Alternating Gradient Synchrotron AGS ($\sqrt{s_{NN}} < 5$ GeV), the CERN Super Proton Synchrotron SPS (up to $\sqrt{s_{NN}} = 17.3$ GeV) and the Brookhaven Relativistic Heavy Ion Collider RHIC ($\sqrt{s_{NN}} \leq 200$ GeV), and now also, at the Large Hadron Collider LHC at CERN at $\sqrt{s_{NN}} = 2.76$ TeV, which is more than a factor 10 higher than the maximum collision energy at RHIC.

The experiments at RHIC brought major discovery that the QGP behaves essentially as a strongly coupled liquid and is opaque to high energy partons. The first heavy-ion collisions at the LHC with lead ions took place in 2010. The nuclei were collided at the centre-of-mass energy of 2.76 TeV. In the coming years the expectation is that the beam energy will be even larger (by at least a factor of two). The results from the LHC confirm the findings at RHIC with an improved precision, but also extend the initial measurements into previously unexplored areas.

In this write up we do not attempt to provide a complete set of results and experimental techniques employed in the heavy-ion experiments. We rather discuss a set of selected topics while providing an ample amount of references for further reading. We focus on collisions at RHIC and at the LHC collider energies, which produce matter at the highest energy density in the largest volume and with the longest lifetime attainable in any laboratory experiment. The initial energy density achieved within heavy-ion collisions has been measured to exceed the critical density for the phase transition to quark-gluon plasma (QGP). The analysis of current data combined with theoretical modeling indicates that the ultra-dense system of partons spends the first few fm/c after the collision above the critical energy density, which is long in comparison to typical strong interaction time scales, and achieves a quasi-equilibrated state. However, the lifetime of QGP is far too short to study the system with external probes. Thus, characterization of the properties of the produced system must proceed by studying its decay products. A single heavy-ion collision is a complicated set of causally connected phenomena. In order to infer the properties of QGP we must know the phenomena that are not related to the QGP phase but simply are a consequence of colliding nuclei at high energies. A heavy-ion collision can be decomposed into three distinct stages arranged in time (τ):

- **Initial stage of the collision.** The initial stage comprises of the initial state radiation from the colliding nuclei and initial nucleon-nucleon inelastic collisions, including hard and soft parton-parton scatterings resulting in a large energy density within the overlap region. It is commonly agreed that the initial scatterings take place in the first tenths of fm/c ($\tau \sim 0.3$ fm/c). Note, the actual passage time of the nuclei is $\tau_{pass} = 2R/\gamma_{cm}c$. Very often in order to interpret the experimental data it is important to know the impact parameter of the collision and the physical phenomena that control the initial energy density buildup. Such phenomena may involve saturation effects expected for low-x partons in nucleons and these effects are predicted much stronger for a nucleus as compared to single nucleons. More-

over, the initial random positions of the nucleons participating in the collision set the initial geometrical eccentricity of the colliding system. The understanding of the initial stage is critical for proper modeling and interpretation of experimental results obtained with heavy-ion collisions.

- **Quark Gluon Plasma.** Owing to the many nucleon-nucleon collisions the energy density within the system can be sufficiently high such that the nucleons cease to exist as bound states of partons. The liberated quarks and gluons can interact within a small but finite volume leading to the creation of a small droplet of a thermalized quark-gluon plasma. Within model calculations needed to interpret the experimental data it is usually assumed that the plasma reaches a point of thermal equilibrium at about $\tau \sim 1 \text{ fm}/c$. The main objects of the heavy-ion research are the physical properties of the QGP, such as temperature, density, shear viscosity/entropy density ratio. In this sense, the studies of QGP resemble the scientific program commonly known from the solid-state physics.
- **Hadron gas and freeze-out.** The expanding QGP cools down and freezes into color neutral hadrons at the so-called chemical freeze-out. After the relative abundances of particles are set elastic collisions between hadrons may still occur. Hadrons interact until the so-called kinetic freeze-out, the point at which the particles cease to interact ($\tau > 10 \text{ fm}/c$). The understanding of this stage of the collision is likewise critical for inferring the properties of QGP. In general, the interactions of particles in the hadron gas phase can distort the genuine QGP effects.

In the following we will introduce several experimental methods employed in studies of the properties of heavy-ion collisions that are used to infer the properties of the QGP:

- **Particle production and their p_T spectra.** The density of the produced particles contains information about the state of the hadron gas at freeze-out. Moreover, the relative abundance of the types of particles and their momentum distributions are affected by thermal properties of the created system, by the phenomena of the collective flow, and the stopping power within the collisions. It is worthwhile to note that the measured distributions in AA collisions cannot be trivially extrapolated from the existing measurements of pp collisions.
- **Particle correlations.** The strongly interacting system results in correlations of particles in the final state. These correlations are sensitive to many properties of the system, such as the size of the emitting source (via quantum correlations of produced particles), the shear viscosity/entropy ratio (via the azimuthal momentum anisotropy of the produced particles), and at high- p_T , the jet-like correlations are sensitive to the high energy parton-medium interactions.
- **Photons and di-lepton mass spectrum.** The measurements of low-momentum photon spectra and di-lepton mass spectrum at low and intermediate masses is expected to be sensitive predominantly to the thermal radiation of the QGP via the process $q\bar{q} \rightarrow \gamma^* \rightarrow e^+e^-$. Experiments at SPS and RHIC have measured the excess in the di-lepton spectra consistent with the predictions of the enhanced thermal emission from the QGP phase. The measurements of the photon spectra

allow to infer the temperature of the medium. The inverse slope parameter T of an exponential fit to data at RHIC gives $T \approx 220 \text{ MeV}$ and now at the LHC $T \approx 300 \text{ MeV}$ [3].

- **Jets.** Jets of particles originated by highly virtual partons are used as auto-generated probes of the medium. The parton-medium interactions and subsequent modifications of the jet structure (*jet quenching*) are sensitive to the density and temperature of the medium. The main goal here is to extract the so-called jet quenching coefficient \hat{q} that is characteristic to a given medium. The coefficient is proportional to the density of the medium.
- **Heavy-quarks.** Heavy-quarks are one of the most promising tool to study the mass and flavor dependence of jet quenching, but also provide a stringent constraints on the transport properties of the medium such as the diffusion coefficient.
- **Quarkonia.** Production of bound $q\bar{q}$ states of heavy-quarks is sensitive to the temperature of the medium. The measurements of the various quarkonia states are an important source of information for modeling of the QGP.

We stress that we will not discuss all of those points explicitly. This write up should serve as an introduction to few selected ideas about experimental aspects of heavy-ion collisions and should not be regarded as a complete review.

2 Soft Probes of QGP

Several basic measurements are used to characterize the particle production in heavy-ion collisions. Among these, traditionally the first measurements that are performed is the number and the pseudo-rapidity density of the produced particles (or density of the transverse energy). These measurements are usually followed by the transverse momentum (p_T) spectrum of the produced particles and moments of that distribution.

2.1 Centrality of a Heavy-Ion Collision

Before we review these results and their sensitivity to the properties of the created matter we ought to briefly discuss the so-called centrality of a heavy-ion collision. In fact most of the heavy-ion measurements depend on this quantity. Let us introduce the concept of a heavy-ion collision centrality in terms of the number of produced particles and/or the produced energy density. Note, in a single heavy-ion collision many nucleons may interact inelastically from both of the ions. An experimental control over the number of the so-called participating nucleons (N_{part}) and the number of binary nucleon-nucleon collisions (N_{coll}) can be provided by associating the particle production with the collision geometry—the impact parameter (b). The b is a distance between the centers of the two colliding nuclei perpendicular to the direction of their relative motion. Figure 1 shows schematically two nuclei just before the

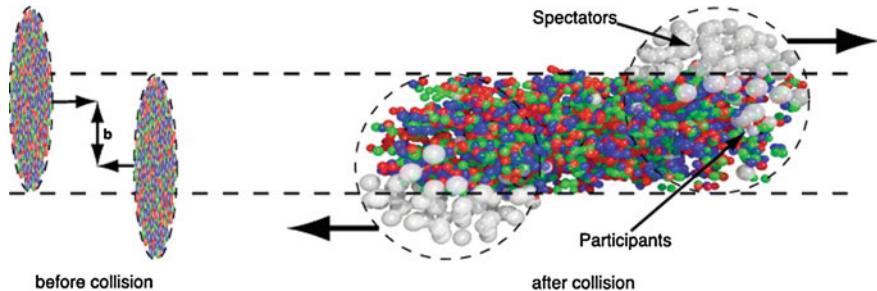


Fig. 1 *Left* Two heavy ions before the collision with the impact parameter b . *Right* The spectators remain unaffected while in the participant zone, particle production takes place—UrQMD calculation [4]

collision with a finite impact parameter and the configuration of nucleons after the collision.

It has been found that the number of produced particles within the collision is inverse proportional and monotonously changes with the impact parameter. More particles are produced for small impact parameters as compared to the number of particles produced for collisions with large b . The collisions with small impact parameters are called *central* while the collisions with large impact parameters are called *peripheral*. The collision centrality is an important estimator of the event selection used to study the magnitude of the nuclear and/or QGP effects. It has been experimentally verified that within the central collisions the energy density thus the QGP related effects are stronger than in peripheral collisions. It is worthwhile to note that almost all measurements reported within heavy-ion collisions will be affected by the precision of the centrality determination.

Experimentally the collision centrality is deduced by measuring the amount of energy produced per collision in a given region of phase space (a rapidity window for example). The precision in the determination of the centrality can be improved by correlating the energy flux registered at different rapidities. Figure 2 shows distribution of the total energy in the Forward Calorimeter (FCal) [5] of the ATLAS experiment [6] divided into bins of event activity or “centrality”. The percentiles correspond to the fraction of the total hadronic cross-section. The figure in the lower panel demonstrates the tight correlation of energy measured at mid-rapidity to energy measured at forward rapidities.

In practice experiments can obtain the number of collisions and the number of participants for a centrality bin utilizing the Glauber [7] modeling of the collision (via Monte Carlo simulation for example) and the fits to experimental data. The fitting assumes a known shape of the nuclei, the inelastic nucleon-nucleon cross-section (often the pp cross-section is known), and that nucleons follow straight trajectories within the interaction. The measurable quantities (particle multiplicities as a function of the percentiles of the cross-section for example) are fitted within the model such that the number of ancestor nucleons is described with the parameter α such that:

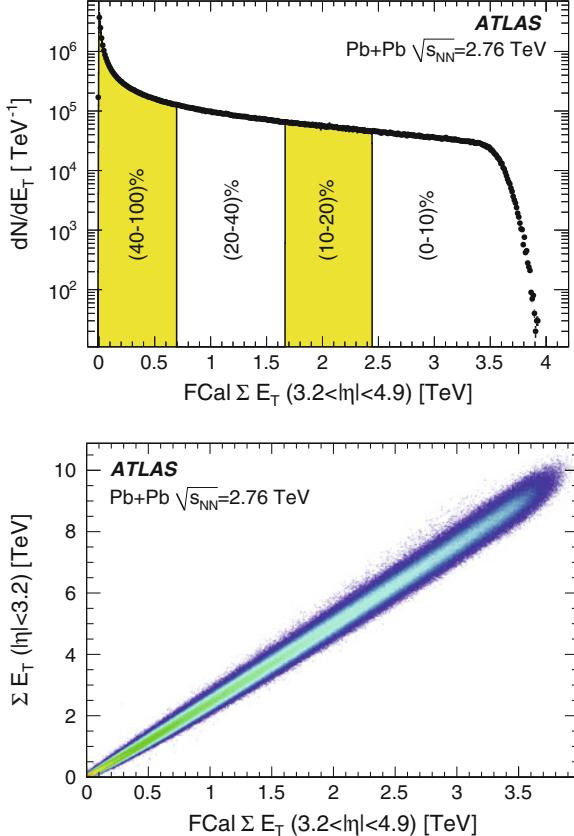


Fig. 2 (top) Distribution of uncorrected ΣE_T in the Forward Calorimeter (FCal). Bins in event activity or “centrality” are indicated by the *alternating bands* (see text for details) and labeled according to increasing fraction of lead-lead total cross section starting from the largest measured ΣE_T . (bottom) Correlation of uncorrected ΣE_T in $|\eta| < 3.2$ with that measured in the FCal ($3.2 < |\eta| < 4.9$). Figure from [5]

$$N_{\text{ancestors}} = \alpha N_{\text{part}} + (1 - \alpha) N_{\text{col}}. \quad (1)$$

For more details please study [8] and references therein.

2.2 Particle Production

The interpretation of the experimental findings strongly relies on theoretical understanding and modeling of heavy-ion collisions. We must insist that any model providing sophisticated interpretations is capable of reproducing the basic observables.

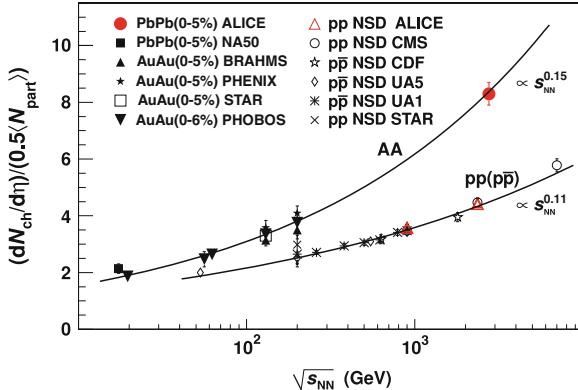


Fig. 3 Charged particle pseudo-rapidity density per participant pair for central nucleus-nucleus and non-single diffractive pp ($p\bar{p}$) collisions, as a function of $\sqrt{s_{NN}}$. The solid lines $\propto s_{NN}^{0.15}$ and $\propto s_{NN}^{0.11}$ are superimposed on the heavy-ion and pp ($p\bar{p}$) data, respectively. Figure from [9]

One of those basic observables is the pseudo-rapidity density of the produced particles. Figure 3 summarizes the experimental data on charged particle production in a narrow interval around the central rapidity ($|\eta| \approx 0$) as a function of the collision energies in pp and most central heavy-ion collisions. Within a single heavy-ion collision a single nucleon may interact more than once and, as expected, a heavy-ion collision produces on average larger particle density per participating nucleon than a pp collision. The increase of the energy of collisions for heavy-ions results in a much steeper rise of the multiplicity of produced particles per number of participating nucleons than in the case of proton–proton collisions. An order of magnitude increase in the collision energy results in roughly doubling the energy density. At the top energy of the LHC the density of produced particles per participating nucleon pair in the most central collisions reaches a factor two larger than in equivalent pp collision. Note that the number of participating nucleons in a central Pb–Pb heavy-ion collision (0–5 %) is about 380, while for pp collisions it is 2.

Figure 4 shows the pseudo-rapidity density of produced particles as a function of pseudo-rapidity. The distribution has two maxima related to the fragmentation regions of the two ions followed by a steep fall towards large rapidity. It illustrates a rapid change of the dense region with the total interaction cross-section indicated by several curves corresponding to percentiles of the hard cross-section (the centrality of the collision). The 0–5 % are the most central collisions (small impact parameter) and 20–30 % are called semi-central. This distribution is a critical benchmark for the models of heavy-ion collisions. In a way one may consider such measurements as calibration tools for the theoretical considerations.

The proton–nucleus collisions are an extremely useful setting for understanding the effects observed in the nucleus–nucleus collisions. In general, in the pA collisions the effects related to QGP are not expected. One of the simplest, but at the same time, a very sensitive observable to the multiplicity of produced particles in pp and

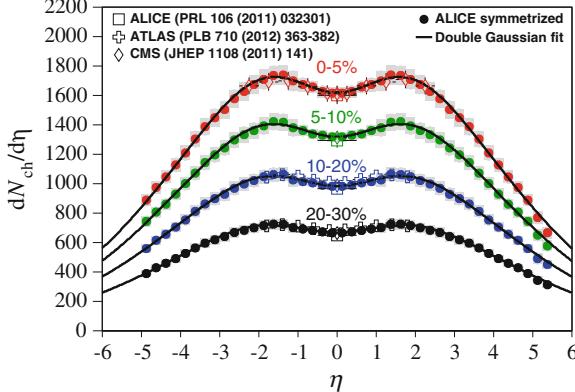


Fig. 4 Comparison of pseudo-rapidity density of produced particles in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV as measured by the LHC experiments. The density was reconstructed separately for several collision centralities. Figure from [10]

p–Pb collisions is the first moment of their transverse momentum distribution—the average p_T . Figure 5 presents a comparison of the average p_T measured in pp , p–Pb, and Pb–Pb collisions as a function of the number of charged particles measured in a pseudo rapidity window of $|\eta| < 0.3$. In pp collisions it is found that the most common and well established event generator, PYTHIA [11], is able to almost exactly reproduce the data points only when the so-called *color reconnection* (CR) mechanism is employed. We must note that the CR mechanism can be regarded as a final-state effect as taking place late in the collision process (later than the initial hard scatterings). It is important to note that at present this is an intriguing observation that resembles collective effects known from heavy-ion collisions and related to the presence of QGP. On the other hand, a strong role that CR plays is visible only within collisions that produce significantly more particles than a “standard” (minimum-bias) pp collision. The events with multiplicity (N_{ch}) of less than 15 exhaust already about 90 % of the hadronic cross-section. Within the p–Pb data a visible change in the trend of the distribution is seen. The p–Pb curve departs from the pp data at N_{ch} larger than about 15 and the increase of the mean p_T is tamed at higher multiplicities. This feature, is similar to the one seen in Pb–Pb collisions where at high multiplicities (central events) a creation of a strongly coupled system (with large correlations) is observed. The comparison of the mean p_T in small systems (pp , and p–Pb) and Pb–Pb collisions was one of the first hints that understanding of how collective effects arise within the small collision systems may become crucial for extraction of the QGP properties in heavy-ion collisions.

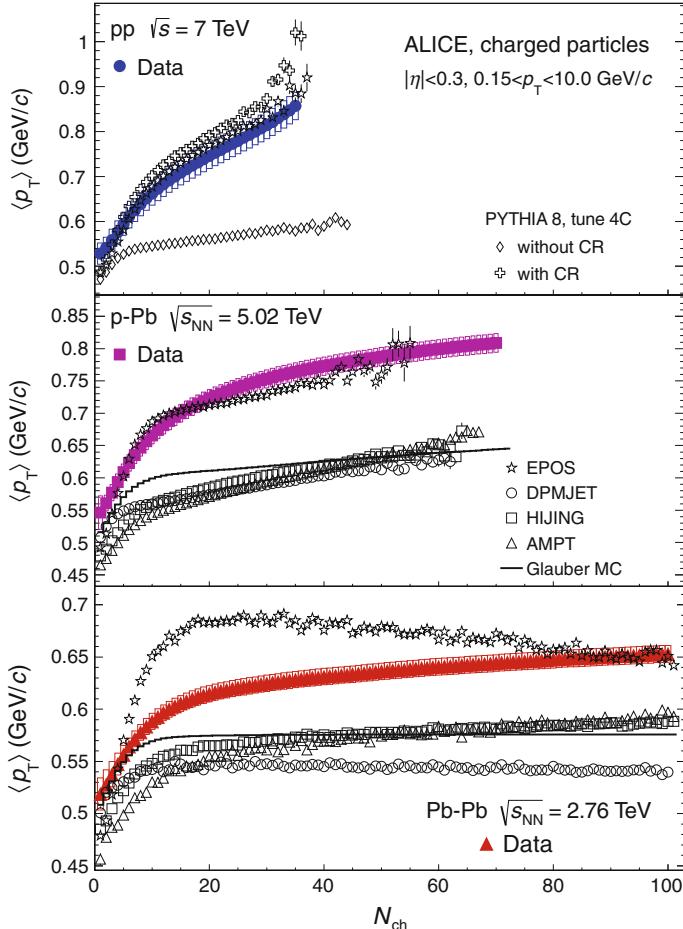


Fig. 5 Average transverse momentum $\langle p_T \rangle$ as a function of charged-particle multiplicity N_{ch} measured in pp (upper panel), $p\text{-Pb}$ (middle panel), and Pb-Pb (lower panel) collisions in comparison to model calculations. The data are compared to calculations with the DPMJET, HIJING, AMPT, and EPOS Monte Carlo event generators. For pp collisions, calculations with PYTHIA 8 [11] with tune 4C are shown with and without the color reconnection (CR) mechanism. The lines show calculations in a Glauber Monte Carlo approach. Figure from [12]

2.3 Particle Identification

Observables based on particle identification play one of the central roles in the experimental studies of heavy-ion collisions. The extraction of majority of signals of QGP depends on observables measured differentially with the type or the mass of the particle. The experiments employ various detection technologies to identify particles via their characteristic signals of their interactions with the detector material.

These include detection of the Cherenkov radiation of an energetic particle traversing through a dielectric material of the detector, measurements of the energy loss within a gas or sensitive solid state volume of the detector, as well as measurements of the time-of-flight of particles.

Figure 6 presents examples of the performance of several detection technologies employed in the ALICE detector [14]. Depending on the momentum of particles the combination of the information from the detectors can be employed to enhance the precision of the identification. The most powerful detector employed for particle identification within ALICE is the Time Projection Chamber which allows to record the energy loss dE/dx of a charged particle traversing the volume of the detector filled with gas. Particles are identified by measuring their characteristic ionization energy loss as defined by the Bethe-Bloch formula. For more details see [15]. The detection of photons, electrons is typically provided by the electromagnetic calorimeters (see [16] for example). For the detection of hadrons ATLAS and CMS detectors employ hadronic calorimeters [17] located outwards of the electromagnetic calorimetry. To detect muons the experiments employ a sequence of absorbers and tracking detectors. The absorbers serve as a passive material in which hadrons are stopped while the energetic muons can propagate with small interaction probability. The setup of the muon tracking detectors in conjunction with the inner tracking systems located in the proximity of the interaction point provide the measurement of the muon trajectory. Another type of commonly used detectors in identification of electrons and muons are the so-called transition radiation detectors, which make use of the electromagnetic radiation emitted when a charged particle passes through inhomogeneous material, such as a boundary between two different materials. More details on currently active detectors can be found in [14] (ALICE detector), [6] (ATLAS), [18] (CMS detector), [19] (LHCb detector), [20] (PHENIX detector), [21] (STAR detector).

A number of particles can be identified via their visible decay products (e.g. charged particles). For instance the neutral kaons with short lifetime (*K-short*) can be identified via the analysis of correlated pairs of charged pions. Similar technique can be applied for Λ baryons and many other particles including open charm mesons. Targeted analyses can of course be employed for not only two, but also three or more body decays. Figure 7 shows examples of the invariant mass distribution of pairs of pions (left), and pion and kaon (right). A clear maxima at the expected kaon and lambda particle mass are visible.

2.4 Particle Spectra

The measurements of spectra of identified hadrons play an important role in establishing the dependence of the physical properties of the created system in heavy-ion collisions. In addition they are sensitive to the dynamics of the system (the expansion) and its temperature at the phase transition to the hadronic phase. Conventionally several p_T -regions of the spectra are highlighted:

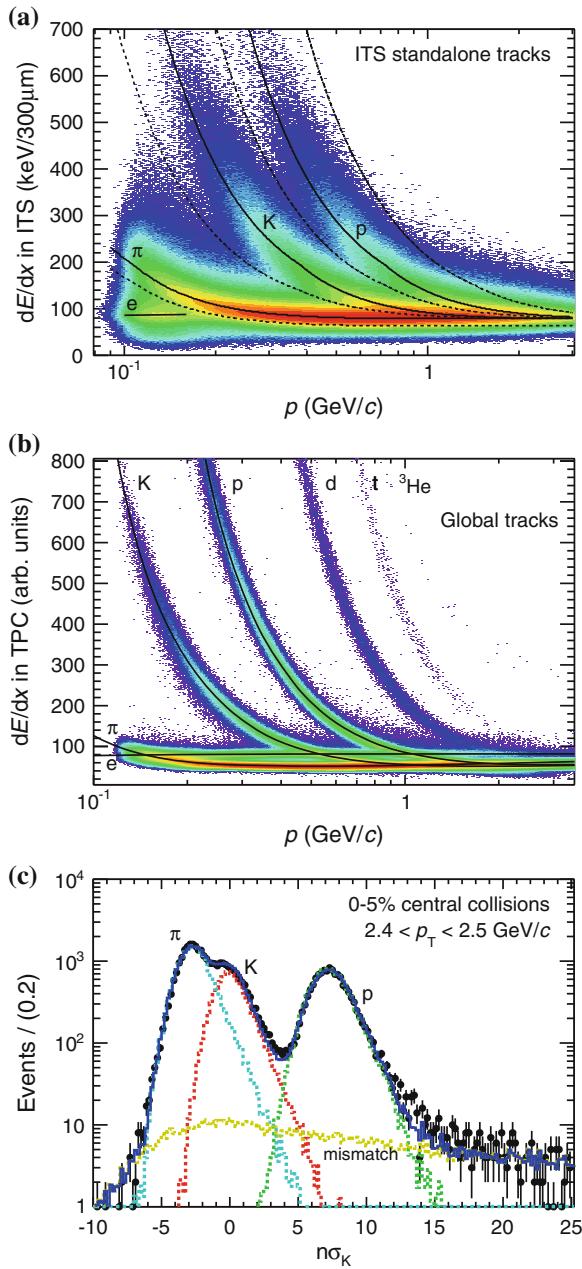


Fig. 6 (color online) Performance of the PID detectors: **a** dE/dx distribution measured in the ITS, the *continuous curves* represent the Bethe-Bloch parametrization, the *dashed curves* the asymmetric bands used in the PID procedure; **b** dE/dx measured in the TPC with global tracks (see text for the definition of global tracks), the *continuous curves* represent the Bethe-Bloch parametrization; **c** fit of the TOF time distribution with the expected contributions for negative tracks and for the kaon mass hypothesis, in the bin $2.4 < p_T < 2.5 \text{ GeV}/c$. Figure from [13]

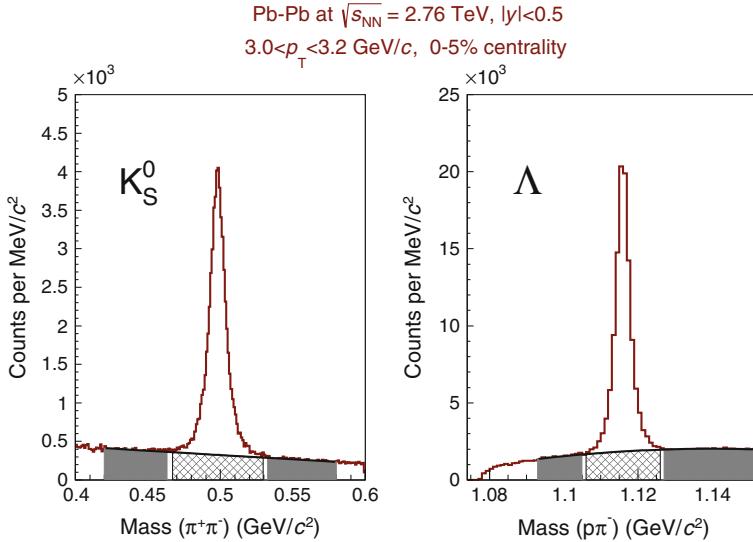


Fig. 7 Examples of invariant mass distributions for K_S^0 and Λ . The *filled areas* to the sides of the peaks were used to fit the background in order to estimate the background level under the peaks, indicated as the *light shaded areas*. Figure from [22]

- low- p_T region ($p_T < \sim 3 \text{ GeV}/c$): The bulk of the particles are produced within this momentum range. From the relative abundances of particles the thermal properties of the system at the kinetic and chemical freeze-out are extracted. The measured yields (typically down to about $100 \text{ MeV}/c$) are extrapolated to zero p_T and used to constrain the statistical hadronization models. More on this topic in Sect. 2.6. Moreover, this region of the spectrum is sensitive to the collective expansion of the system. The particles flow within a common velocity field—the so-called *radial flow*. More on that below in Sects. 2.7 and 2.9.
- intermediate- p_T region ($\sim 3 < p_T < \sim 6 \text{ GeV}/c$): This region is a result of an interplay between the collective expansion of the system and particle production associated to hard scatterings. Several models postulate the so-called coalescence and/or parton recombination phenomena that aim to explain the shape of the spectra and modifications of meson and baryon production as compared to pp collisions. More on that in Sect. 2.8.
- high- p_T region ($p_T > \sim 3 \text{ GeV}/c$): This region is dominated by particle production occurring in jets. Jets are spray of particles resulting from fragmentation of a high-energy gluon or quark. More on that topic in Sect. 3.

Figure 8 shows the measured p_T spectra of charged pions and protons. The measurements are performed in several centrality bins of Pb–Pb collisions. These are just examples of the identified particle spectra—the experiments are able to reconstruct particle yields for a number of particle species through neutral pions, kaons, lambda baryons to much heavier Ξ and Ω baryons, as well as containing charm quark D and

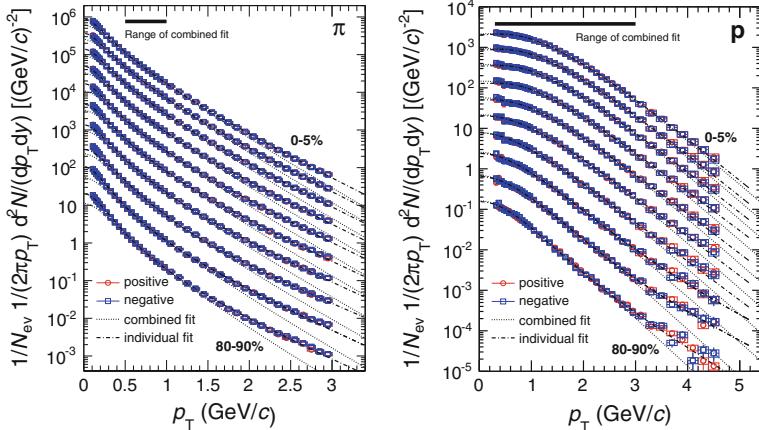


Fig. 8 The p_T -differential spectra of identified charged pions (left) and protons (right) in bins of centrality in Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV by ALICE. Figure from [13]

J/ψ mesons. The future, high statistics runs at the LHC will enable the reconstruction of mesons and baryons containing bottom quarks. Some of those measurements are already now being performed in pp and p–Pb collisions at forward rapidities by the LHCb experiment. Moreover, STAR and ALICE experiments are capable of reconstructing light nuclei produced from the coalescing nucleons.

2.5 Particle Yields at Freeze-Out

To extract the freeze-out parameters from the measured spectra we follow the prescriptions used in [13]. We are interested in the total (p_T -integrated) yields of the produced particles. The particle spectra are fitted individually with a blast-wave function [23]:

$$\frac{1}{p_T} \frac{dN}{dp_T} \propto \int_0^R r dr m_T I_0 \left(\frac{p_T \sinh \rho}{T_{kin}} \right) K_1 \left(\frac{m_T \cosh \rho}{T_{kin}} \right), \quad (2)$$

where the velocity profile ρ is described by

$$\rho = \tanh^{-1} \beta_T = \tanh^{-1} \left(\left(\frac{r}{R} \right)^n \beta_s \right). \quad (3)$$

The $m_T = \sqrt{p_T^2 + m^2}$ is the transverse mass, I_0 and K_1 the modified Bessel functions, r is the radial distance in the transverse plane, R is the radius of the fireball, β_T is the transverse expansion velocity and β_s is the transverse expansion velocity at the surface. The free parameters in the fit are the freeze-out temperature T_{kin} , the average transverse velocity $\langle \beta_T \rangle$ and the exponent of the velocity profile n . The function describes very well the spectrum of all particles within the measured p_T range. However, from fits to a single particle species no physics interpretation of those parameters can be extracted. The correct way to extract the freeze-out condition is to perform a combined fit to different particle species.

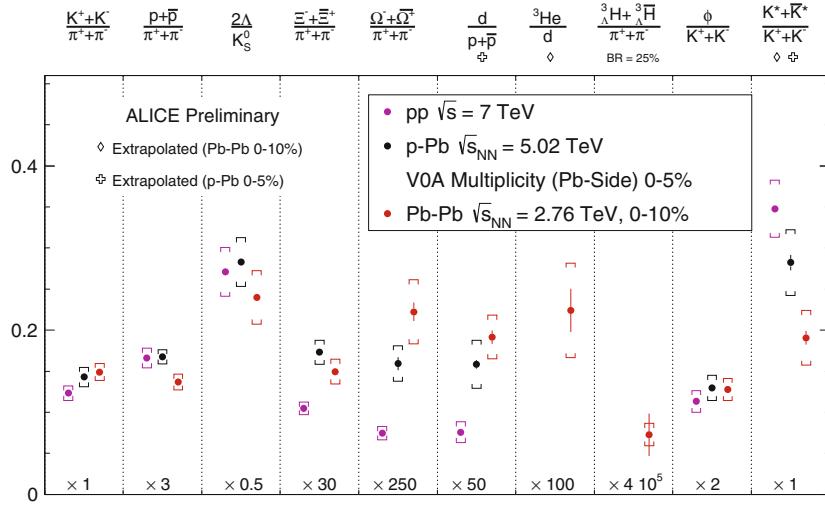
2.6 Freeze-Out Temperature and Thermal Models

Thermal properties of the system can be extracted by the abundances (total yields) and the ratios of abundances of hadrons. Figure 9 shows a recent compilation by ALICE Collaboration of particle ratios in pp , p–Pb, and Pb–Pb collisions. Several features of that diagram are worthwhile to note. First, one observes the relative enhancement of the ratio of strange particle production to charged pions in heavy-ion collisions as compared to pp . This is seen in kaon/pion ratio but also in a strong enhancement of the omega baryon production. The enhancement of strangeness was predicted as one of the signatures of the QGP. Within the hot plasma the production of $s\bar{s}$ strange quark pairs is energetically favorable as compared to strange hadron production in a hadron gas phase. Secondly, we note a clear decrease in K^*/K^\pm ratio. This is an indication of substantial re-scattering processes taking place in the hadronic phase in late stages of heavy-ion collisions. The re-scattering lowers the probability for the resonant kaons to survive until the kinetic freeze-out. Moreover, the diagram clearly shows an enhanced production of deuterons as compared to protons in Pb–Pb collisions.

The so-called *thermal* or *statistical* models are formulated to describe the state of the system at chemical freeze-out. Using the measured abundances of produced hadrons and assuming a local thermal equilibrium a model considering the grand canonical statistical ensemble is often used to extract the freeze-out temperature of the system. The grand canonical ensemble is defined by the chemical potential μ_B , temperature T and the volume V of the system. In particular the number of hadrons produced in a collision can be written as:

$$n_i = g_i V \int \frac{d^3 p}{(2\pi)^3} \left[\exp \left(\frac{E_i(p) - \mu_i}{T} \right) - \varepsilon \right]^{-1}, \quad (4)$$

where the chemical potential μ_i for strongly interacting system is a linear function of the baryon chemical potential μ_B , the strange chemical potential μ_S , and the isospin chemical potential μ_I . At the LHC the chemical potential is close to zero (at high collision energies as many anti-particles as particles are produced).



ALICE-PREL-74423

Fig. 9 Ratios of particle yields in pp , p - Pb , and Pb - Pb collisions. Figure from [24]

The volume of the system is fixed by consideration of the Pb - Pb collision and it is mostly driven by the most abundant species (pions). Finally the temperature is predominantly constrained by the particle ratios with large mass differences (any baryon/pion ratio for example). Figure 10 shows an exemplary comparison of results obtained with models to the experimentally obtained particle abundances. Three models are shown [25–27]. All of the models do remarkably well to describe the particle yields, including the 3He nuclei. While differing in their details of implementation they all extract the same temperature of about 156 MeV. The THERMUS [26] is a software package for statistical-thermal analysis. The GSI model is a calculation of the production of hadrons in nuclear collisions within the framework of the thermal (or statistical hadronization) model [25]. And the SHARE calculations are based on the chemical non-equilibrium statistical hadronization model (SHM) [27]. One can see that, at present, one of the tensions between models and the data needing further understanding is the yield of protons. There are several potential explanations for such a tension: additional late stage baryon-antibaryon annihilation (specifically $p\bar{p}$), sequential freeze-out of different quark flavors, non-equilibrium freeze-out conditions, or incomplete knowledge of the baryon resonance spectrum. For a recent and an in-depth discussion of the thermal models and the data see [24] and references therein.

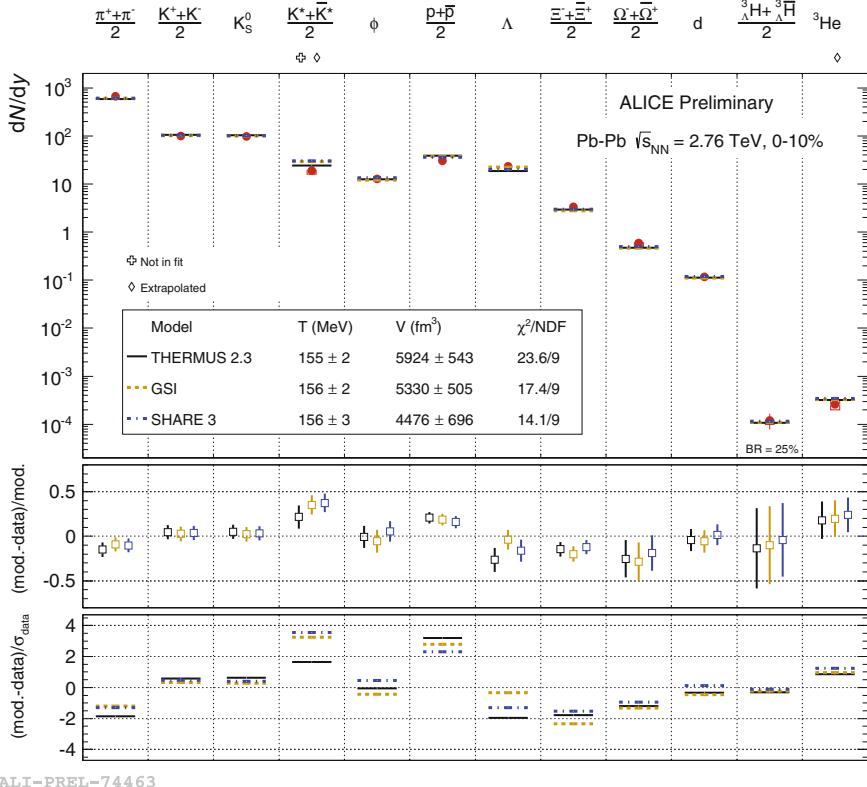


Fig. 10 Abundances of identified particles in most central heavy-ion collisions compared to thermal models GSI [25], THERMUS [26], and SHARE [27]. Figure from [24]

2.7 Radial Flow

As noted in the previous sections the p_T -differential spectra encode the information about the details of the expanding system. In particular, one can extract information about the thermalized system and its radial expansion. To extract features characteristic to the expansion let us come back to Fig. 8 and Eq. 2. We can make two distinct observations:

- The shape of proton spectra evolves with centrality. The distribution for the most central bins are significantly *flatter* at low- p_T as compared to the peripheral events.
- The shape of pion spectra is significantly steeper than of protons within the same centrality bin. We note the vastly different masses of the two particles.

By a simultaneous fit defined by Eq. 2 to all the particle spectra one can quantify the kinetic freeze-out temperature T_{kin} and the transverse expansion velocity $\langle \beta_T \rangle$. The fits are shown within Fig. 8. Figure 11 shows these parameters extracted from ALICE

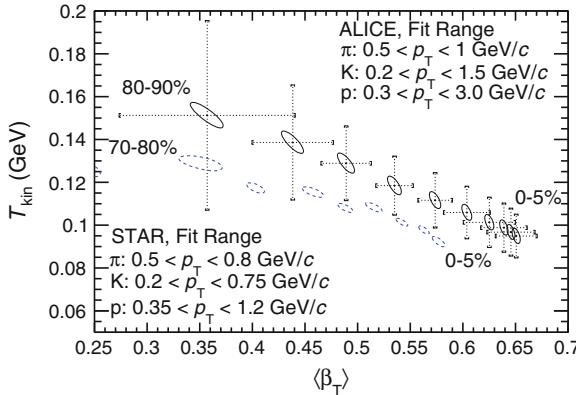


Fig. 11 (color online) Results of blast-wave fits at the LHC [13], compared to similar fits at RHIC energies [28]; the uncertainty contours include the effect of the bin-by-bin systematic uncertainties, the *dashed error bars* represents the full systematic uncertainty, the STAR contours include only statistical uncertainties. Figure from [13]

measurements [13] at the LHC and STAR at RHIC [28]. The T_{kin} is lower for more central events as compared to peripheral events, suggesting a longer lived fireball. This is strongly correlated with the evolution of the average expansion velocity—the $\langle \beta_T \rangle$ grows for large volumes of the system (more central events). Moreover, the almost an order of magnitude increase in the $\sqrt{s_{\text{NN}}}$ results in larger T_{kin} for all centralities at the LHC as well as it energizes the system to reach larger expansion velocities. On the other hand, the similar trends of the blast wave parameters may also suggest a common physical properties of the system and its similar evolution despite the vastly different collision energies. Finally, let us note that for particles expanding within a common velocity field the mass of each particle must have an impact on their p_T -spectra. This is exactly the effect that we see in Fig. 8. Spectra of protons for events with large velocity are deformed (flatter) at low- p_T (where the flow effects are the largest) for events with large $\langle \beta_T \rangle$. Consequently, the effect is stronger for particles with large mass (protons are *pushed out* to higher p_T as compared to pions). Finally, we must note that the T_{kin} is much smaller than the T extracted from the chemical abundances of particles. This is consistent with the picture of the collision evolution where the chemical freeze-out is followed by the kinetic freeze-out.

2.8 Baryon Anomaly

For historical reasons let's consider one of the most puzzling observations in heavy-ion collisions: *the baryon anomaly*. It was first observed at RHIC and coined as a anomalous baryon production, and then later confirmed and explored in detail at the LHC. The anomaly surfaced with the p_T -differential baryon/meson ratio. Figure 12

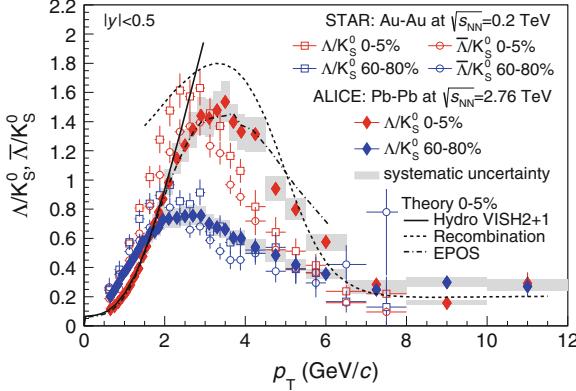


Fig. 12 The Λ/K_s^0 ratio as a function of the hadron p_T . Measurements obtained by the ALICE at the LHC and STAR at RHIC. Strong maximum (above unity) for most 0–5 % central events is observed while for the peripheral collisions the ratio is closer to the values found in pp collisions. Several models are employed to explain the LHC data. The Hydro VISH2+1 model and EPOS involve a hydrodynamical evolution of the system. A model incorporating the parton-parton recombination in hadron creation is also shown. See [22] for details

shows the measured p_T -spectrum of Λ baryons divided by the p_T -spectrum of neutral kaons. The ratio shows a clear maximum at the intermediate p_T . The ratio reaches values larger than unity and it is more than a factor of two larger than in pp collisions. Similar to the considerations above, the ratio can be divided into three regions: (i) the low- p_T where the collective effects dominate; (ii) the intermediate- p_T where the maximum of the ratio is reached and a gradual fall is observed; and (iii) the high- p_T region dominated by the jet fragmentation. As more measurements are being performed it becomes clear at the LHC that the so-called anomaly is a clear consequence of the interplay between radial flow and the jet fragmentation and it is much less puzzling nowadays. In heavy-ion collisions the spectra of the more massive particles (as seen in the previous sections) are deformed according to their mass as compared to pp collisions. Figure 12 in addition to data from LHC and RHIC shows the results of theoretical calculations including hydrodynamic considerations, parton recombination model, and Monte Carlo generator combining all the phenomena relevant for the low- and high- p_T particle production. For more details see [22] and references there in. One interesting feature that can be seen by in Fig. 12 is the apparent shift of the maximum of the ratio from RHIC to LHC for the 0–5 % most central events: The ratio at the LHC reaches maximum at larger p_T than at RHIC. This is in very good agreement with the observation we've made using the blast-wave fits that the most central events at the LHC have a larger average transverse velocity $\langle \beta_T \rangle$ as compared to RHIC.

2.9 Elliptic Flow

The correlations of particles produced in heavy-ion collisions are a powerful tool to assess the properties of QGP. One of the most interesting observations is the final-state momentum anisotropy of the produced particles in non-central collisions (with the impact parameter $b > 0$). The measurements of these anisotropies allow to address fundamental questions on the dynamic properties of the hot and dense QCD matter. An example of a non-central collision is shown in Fig. 14. The two colliding nuclei create an asymmetric fireball. A correlation between the measured azimuthal momentum distribution of particles emitted from the decaying fireball and the initial spatial asymmetry can arise only from multiple interactions between the constituents of the created matter. In other words the momentum anisotropy carries information about how the matter flows. The flow is directly related to the equation of state and thermodynamic transport properties of the created matter. The initial asymmetry is quantified in the transverse plane by the so-called eccentricity

$$\varepsilon = \frac{\langle y^2 \rangle - \langle x^2 \rangle}{\langle y^2 \rangle + \langle x^2 \rangle}, \quad (5)$$

where the x and y are the coordinates in the transverse plane of the collision. The interactions within the fireball may lead to the final energy-momentum anisotropy defined as:

$$\varepsilon_T = \frac{\langle T_{xx} - T_{yy} \rangle}{\langle T_{xx} + T_{yy} \rangle}, \quad (6)$$

where T_{xx} and T_{yy} are the diagonal transverse components of the energy momentum tensor. An example of a model calculation demonstrating how the initial almond shaped fireball while cooling down expands in time is shown in Fig. 13. The energy density asymmetries at late times can be seen by comparing the contours in the orthogonal x and y directions.

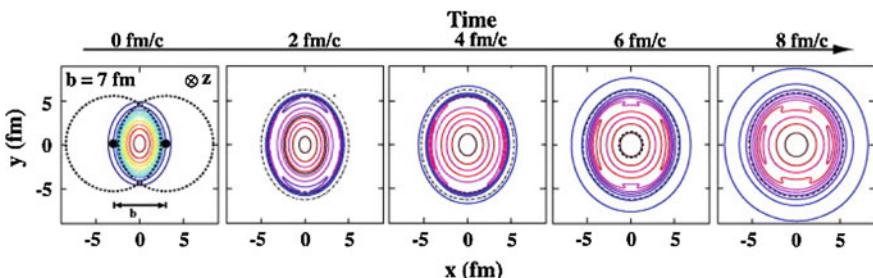


Fig. 13 Result of theoretical calculation [29]. Time evolution of the initial transverse energy density profile in coordinate space for a non-central heavy-ion collision. The z -axis is along the colliding beams and the x -axis is defined by the impact parameter

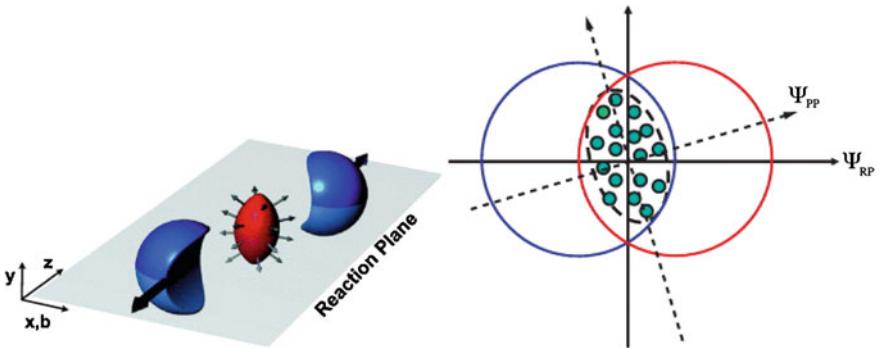


Fig. 14 *Left* Sketch of two nuclei after the collision with the impact parameter $b > 0$ creating the hot fireball in the center region. The reaction plane is defined by the z -coordinate and the vector of the impact parameter b —the x - z plane. The spatial anisotropy with respect to the reaction plane translates into a momentum anisotropy of the produced particles (anisotropic flow). *Right* View of the collision in the transverse plane with the reaction plane angle Ψ_{RP} and the so-called participant plane Ψ_{PP} defined by a particular configuration of nucleons participating (participants) in the collision

Figure 14 introduces the concept of the reaction plane. The reaction plane is defined by the beam axis and the impact parameter b . In the experiment the angle of the reaction plane is not known. What is measurable though is the so-called event plane or participants plane (actually its azimuthal angle). On event-by-event basis the Ψ_{PP} can be inclined at a finite angle with respect to the reaction plane. Thus, the measurements are sensitive to the fluctuations of the initial geometry of the collision. As illustrated by the right panel of Fig. 14 the participants are randomly distributed in the overlap region defining the Ψ_{PP} that fluctuates on an event-by-event basis. Such randomness of the initial distribution of the nucleons has strong consequences of measuring the parameters of the anisotropy.

To characterize the various patterns of the particle distributions one uses a Fourier expansion of the invariant cross-section:

$$E \frac{d^3 N}{d^3 \mathbf{p}} = \frac{1}{2\pi} \frac{d^2 N}{p_T dp_T dy} \left[1 + 2 \sum_{n=1}^{\infty} v_n \cos n(\varphi - \Psi_{RP}) \right], \quad (7)$$

where E and \mathbf{p} are the energy and momentum of the particle, φ is the azimuthal angle, and y is the rapidity. For symmetry reasons the sine terms are zero, while the Fourier coefficients are given by: $v_n(p_T, y) = \langle \cos n(\varphi - \Phi_{RP}) \rangle$. Although the energy-momentum eccentricity ε_T from Eq. 6 is not directly observable one may experimentally study the evolution of the energy density by measuring the v_n coefficients. To relate the ε_T to v_n coefficients it is useful to note that the elliptic

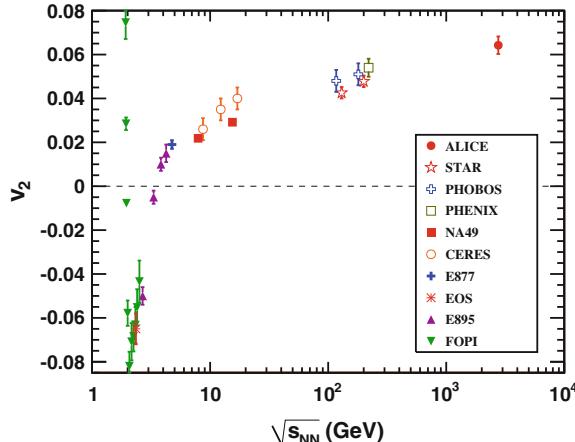


Fig. 15 (color online) Integrated elliptic flow at 2.76 TeV in Pb–Pb 20–30 % centrality class compared with results from lower energies taken at similar centralities [31, 32]. Figure from [33]

flow coefficient v_2 can be written as $v_2 = \frac{\langle p_x^2 \rangle - \langle p_y^2 \rangle}{\langle p_x^2 \rangle + \langle p_y^2 \rangle}$, where p_x and p_y denote the transverse components of the particle momentum.

Finally, let us remind that macroscopic descriptions of a given system are applicable once the mean free path of a particle is much smaller than the size of the system. This is one of the conditions satisfied by QGP allowing to employ relativistic hydrodynamics and describe the evolution of the system up to hadronization. For more details on the application of hydrodynamics to heavy-ion collisions see [30].

Figure 15 shows a compilation of integrated v_2 measurements in heavy-ion collisions as a function of the collision energy. The figure prepared by the ALICE Collaboration [34] demonstrates a continued increase of the magnitude of elliptic flow from few GeV, through RHIC energy to the LHC. It is most interesting to see the transition of v_2 from the negative values (out-of-plane flow) at lower energies to a large positive (in-plane flow) values. The negative v_2 at lower energies is reminiscent of strong nucleon potentials driving the produced particles to escape the dense fireball perpendicular to the reaction plane where the pressure is lower than in-plane. On the other hand, the positive values at the lowest beam energies result from the in-plane rotational motion of the nucleons participating in the collision which predominantly happens in-plane. Again the positive v_2 at energies $\sqrt{s_{NN}} > 4$ GeV indicate an onset of strong interactions within the fireball that eventually at high energies lead to creation of a strongly coupled de-confined medium.

Figure 16 presents the Fourier coefficients measured at the LHC in semi-central (30–40 %) collisions [34]. The v_2 is particularly sensitive to the internal friction or viscosity of the fluid, or more precisely, η/s , the ratio of the shear viscosity (η) to entropy density (s) of the system. For a good fluid such as water, the η/s ratio is small. A “thick” liquid, such as honey, has large values of η/s . Comparison of the

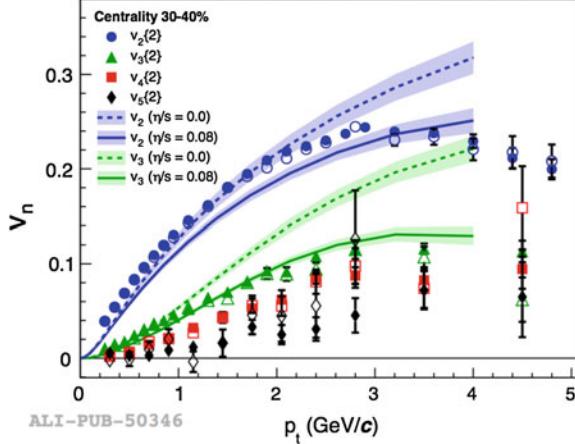


Fig. 16 (color online) v_2, v_3, v_4, v_5 as a function of transverse momentum and for three event centralities. The full, open symbols are for $\Delta\eta > 0.2$ and $\Delta\eta > 1.0$, respectively. The data in the 30–40 % most central collisions compared to hydrodynamic model calculations. Figure from [34]

elliptic flow measured in heavy-ion collisions with theoretical models suggests that the hot matter created in the collision flows like a fluid with little friction, with η/s close to the theoretical limit for a perfect fluid $\eta/s = \hbar/4\pi k_B$, where \hbar is Planck's constant and k_B is the Boltzmann constant. The limiting value for η/s has been derived using string theory methods with the anti-de Sitter/Conformal Field Theory conjecture [35].

Further constraints to the theoretical calculations can be achieved by measuring the higher n Fourier coefficients v_n for $n > 2$. The odd coefficients arise from the event-by-event fluctuating orientations of the nucleons. In some events the overlapping collision zone may be shaped similar to a triangle. The v_3 is associated to the so-called *triangular flow*. Further symmetry planes introduced by the geometrical fluctuations may result in large number of v_n coefficients to be non-zero. The measurements of these flow coefficients provide a direct information about how the initial state asymmetries are transported into the final state momentum anisotropies. Moreover, the larger the number of coefficients is measured by the experiments the more sensitive the details can be extracted from modeling of the liquid QGP state. A powerful tool to study the full spectrum of the v_n comes from two- and many-particle correlations. The description of these techniques is beyond this introduction; however, to give the reader an example the measured correlation function with two-particle azimuthal correlations are shown in Fig. 17. The correlation function can be decomposed into a series of $V_{n\Delta}$ coefficients that can be directly related to the flow coefficients v_n [36].

Furthermore, the hydrodynamical behavior of the system can be assessed with identified particles. Hydrodynamics requires that the $v_2(p_T)$ shows a particular mass ordering. An illustration of such effect is given in Fig. 18. Similar to the effect seen

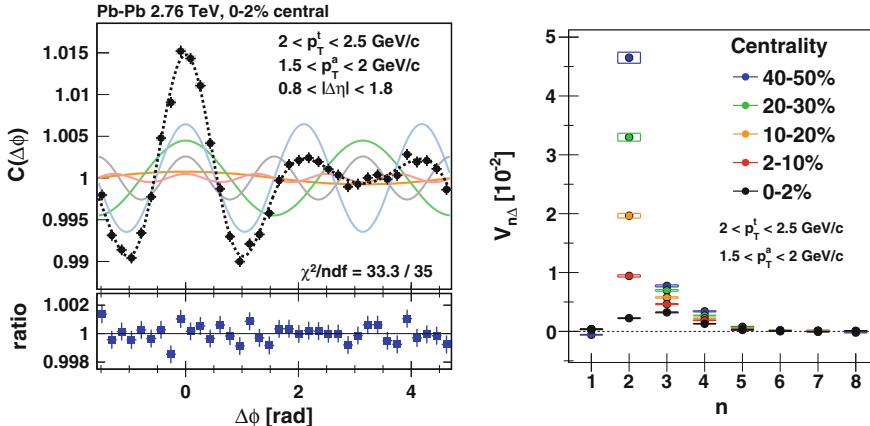


Fig. 17 *Left* Two-particle correlation function and its ratio to the sum of the harmonic functions. The particles are correlated in azimuth into pairs of the so-called *trigger* and *associated* particle for various transverse momentum bins. *Right* The $V_{n\Delta}$ coefficients (up to $n = 8$) extracted from the fits of the correlation function shown on the *left*. Figure from [36]

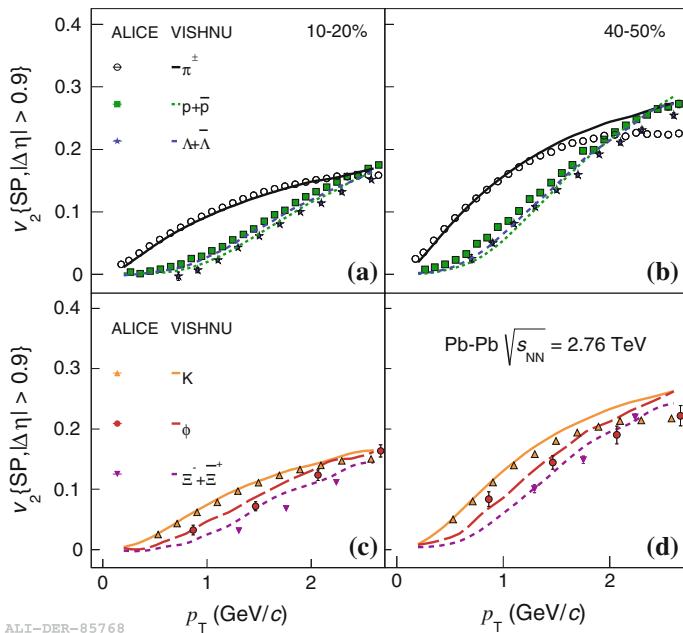


Fig. 18 Elliptic flow measured in two centrality intervals (*left* 10–20 % and *right* 40–50 %) of Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76 \text{ TeV}$. The v_2 is shown for charged pions, protons, Λ (*upper panel*), and kaons, ϕ , Ξ baryons (*lower panel*). Figure from [37]. The theoretical calculation-VISHNU-based on [38]

in the p_T -spectra of identified hadrons, one can clearly observe the apparent push of more massive particles towards the larger p_T . The detailed measurements for most of the hadrons allow for quantitative comparisons to theory. An example of such comparison is given in [37] (see Fig. 18). Studying it in detail we can see that there is room for improvements. Even the state of the art calculation does not reproduce all the measurements in full (the agreement with theory depends on the particle species). Nevertheless, a tremendous progress has been made in the recent years. Both in terms of theory and experiment. On theoretical side, the calculations nowadays couple the initial conditions to the hydrodynamical phase and are followed by calculation of the hadronic re-scattering in the late stages of the collisions [38]. On the experimental side, more particles are measured and with much better precision. One of the most interesting recent result is the measurement of the ϕ -meson. The ϕ has a mass similar to a proton and it was expected that it will provide insight into the understanding of the flow of hadrons—in particular, to what degree the number of constituent quarks plays a role in establishing the flow of mesons and baryons. The new result highlighted the importance of the mass of the particle and disfavored mechanisms that would depend on the particle type.

3 Hard Probes

In this section we will review the experimental signatures of QGP obtained with via measurements of so-called *hard probes* with the created medium. The idea here is quite simple: similar to X-ray scanning of an unknown material we will use highly energetic (hard) partons that interact with the medium. Just as to measure the properties of any form of matter, including ordinary solids and liquids but also a system of quarks and gluons at high temperature, one can study the passage of particles through a given volume of that substance. In QED, this process is described by the Bethe-Bloch formula for the mean rate of energy loss, dE/dx . For electrically charged particles dE/dx depends on several properties of matter, including its atomic excitation and ionization properties and its polarizability, which can be characterized by its plasma energy. By measuring the energy loss of identified and well-calibrated particles one can characterize (unknown) properties of matter. Figure 19 shows the stopping power for a muon traversing through copper. Let's note that in the high-energy limit the energy loss is dominated by the radiative process—a similar situation as in QCD. As we noted in the opening sections, within heavy-ion collisions the usage of external particles to probe the very shortly lived medium is not possible. Instead, one uses the hard scatterings occurring within the heavy-ion collision.

At RHIC and even more so at the LHC energies, partonic interactions with large momentum transfer are abundant, and the hadronic remnants of high-energy scattered partons (*jets*) become experimentally accessible. The highly energetic partons are a class of auto-generated probes whose initial distributions can be calibrated independently. This can be done by comparing the experimental measurements in proton–proton collisions to the perturbative QCD (pQCD) calculations. The partons

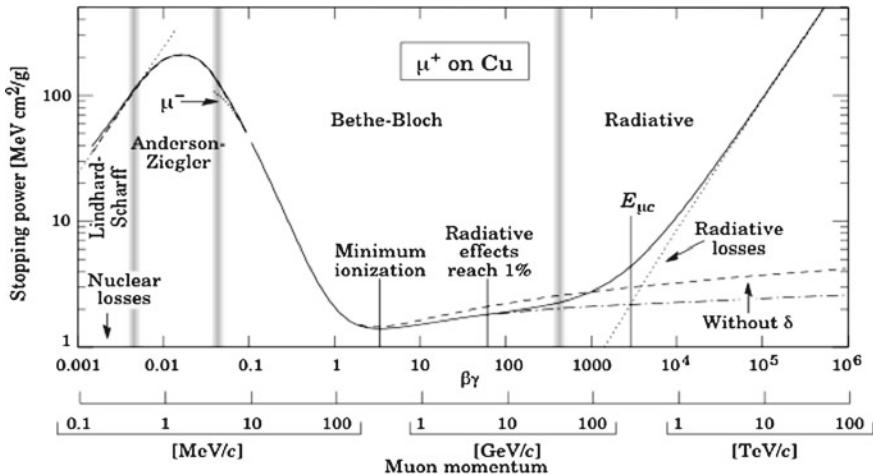


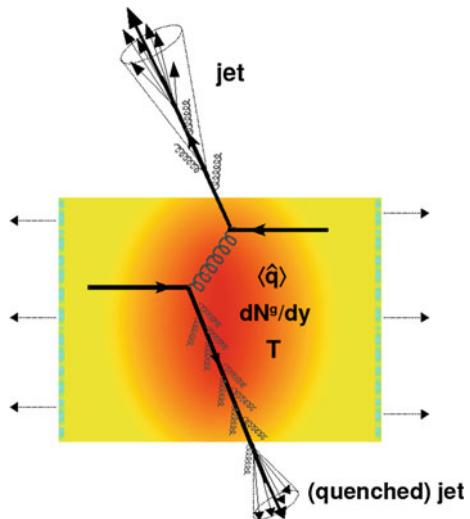
Fig. 19 Stopping power, $-dE/dl$, for positive muons in copper as a function of $\beta\gamma = p/Mc$ (or momentum p). The solid curve indicates the total stopping power [39]

produced in these (high- Q^2) processes in heavy-ion collisions may be viewed as well-identified and calibrated projectiles. These partons while propagating through the dense matter produced in the collision may interact with the medium and loose a portion of their energy (*medium-induced energy loss*). The task is to characterize the medium-modification of the parton propagation and to relate it to fundamental properties of the produced matter. A parameter that is used to characterize the parton-medium interaction is the so-called jet transport coefficient $\hat{q} \propto \mu^2/\lambda$, where λ is the momentum transfer between the medium and the parton, and λ is the mean free path within the medium. We note that \hat{q} is directly proportional to the density of the medium (through the dependence on the mean free path). The energy loss of a parton within a medium of a given density and temperature can be related to \hat{q} such that $dE/dx \propto \alpha_s \hat{q} L^2$, where L is the path length that a parton traversed through the medium.

3.1 Jet Quenching with Hadron Observables

Historically, full jet reconstruction in the presence of large backgrounds in heavy-ion collisions was thought to be impossible. This view predominated until recent developments in background subtraction techniques. Therefore, initially the experiments at RHIC utilized the single particle measurements, which approximated jets via leading hadron observables. STAR and PHENIX at RHIC have measured a strong depletion of high- p_T hadrons created in central heavy-ion collisions as compared to the expected yields derived from proton–proton measurements (Fig. 20).

Fig. 20 Jet quenching in a head-on nucleus–nucleus collision. Two quarks suffer a hard scattering: one goes out directly to the vacuum, radiates a few gluons and hadronises, the other goes through the dense plasma created (characterised by transport coefficient \hat{q} , gluon density dN_g/dy and temperature T), suffers energy loss due to medium-induced gluonstrahlung and finally fragments outside into a (quenched) jet. Figure from [40]



Assuming the scaling of hard processes with the number of independent binary nucleon–nucleon collisions one can define the nuclear modification factor R_{AA} as a ratio of hadron yields measured in heavy-ion collisions to expected yields obtained by superposition of independent nucleon-nucleon inelastic collisions:

$$R_{AA}(p_T) = \frac{dN_h^{AA} p_T / dp_T}{\langle N_{bin} \rangle dN_h^{pp} (p_T) / dp_T} \quad (8)$$

The measured R_{AA} (see Fig. 21) indicates a large deficit (of about factor 5) of high- p_T hadrons in Au–Au collisions as compared to p–p. At the LHC the R_{AA} grows to up to 0.5 at 100 GeV/c. Such loss of hadrons and no sign of deficit for the color neutral objects (direct photons, W and Z bosons recently measured at the LHC— $R_{AA} \sim 1$ [42–44]), as well as no sign of nuclear effects at high- p_T in p–Pb collisions at the LHC [45] supports the argument of substantial interactions of high-energy partons within a hot and dense colored medium created in heavy-ion collisions.

3.2 Particle Correlations at High- p_T

Another powerful way of studying jet quenching are particle correlations. The measurement of the azimuthal angular correlations of hadrons originating from back-to-back jets revealed a strong depletion of the yield of the hadrons associated to the recoiling jet (often called away-side jet). The direction of the near-side jet is approximated with a high- p_T hadron, the so-called trigger particle. The associated particles within the same event with $p_T < p_{T,\text{trigger}}$ are correlated in azimuth to the trigger

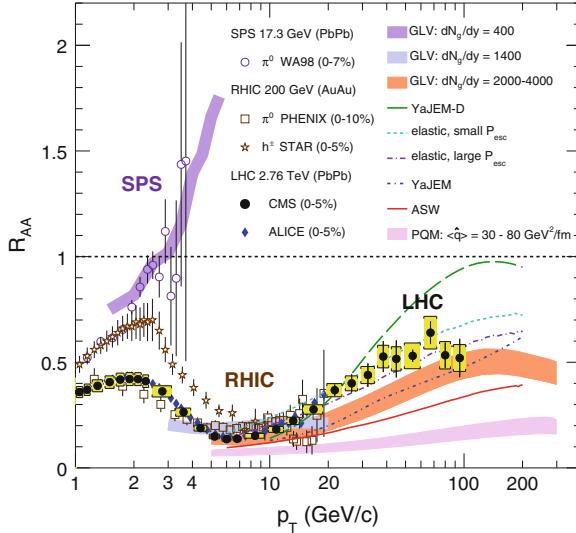


Fig. 21 R_{AA} for charged hadrons measured at the SPS, RHIC and the LHC. The experimental results are compared to theoretical calculations (see [41] and references therein for details)

particle. To quantify the effect the conditional yield in nucleus-nucleus collision is scaled by the same measurement in pp collisions, such that

$$I_{AA} = \frac{1/N_{\text{triggers}} dN_{\text{pairs}}^{\text{AA}} / dp_T}{1/N_{\text{triggers}} dN_{\text{pairs}}^{pp} / dp_T}. \quad (9)$$

Similar to R_{AA} , the I_{AA} is unity when no nuclear effects are present.

The measurements of R_{AA} together with the I_{AA} delivered first constraints to the modeling of the QGP properties. In particular the measurements at RHIC has shown that the initially generated medium is hot, highly opaque to energetic partons and very dense (~ 30 - 50 times cold nuclear matter). Figure 23 summarizes the current status of characterizing the properties of the medium using the jet transport parameter \hat{q}/T^3 . The value of the parameter has been extracted by fitting the experimental data of single hadron R_{AA} and di-hadron I_{AA} at RHIC and the LHC. The values were obtained for a number of jet quenching models. In addition, the figure illustrates the evolution of the parameter with temperature (the ranges for RHIC and the LHC are indicated). For more details see [47] (Fig. 22).

However, measurements relying on a single high momentum hadron suffer from well-known biases. These biases arise from the combined effects of the sharply falling inclusive jet spectrum and the jet fragmentation distribution. The result is that the leading hadrons that dominate the high-energy hadronic spectrum originate from the partons that have lost very little of their initial energy, i.e. not quenched jets. Consequently, the comparison of hadron-triggered data with theoretical predictions

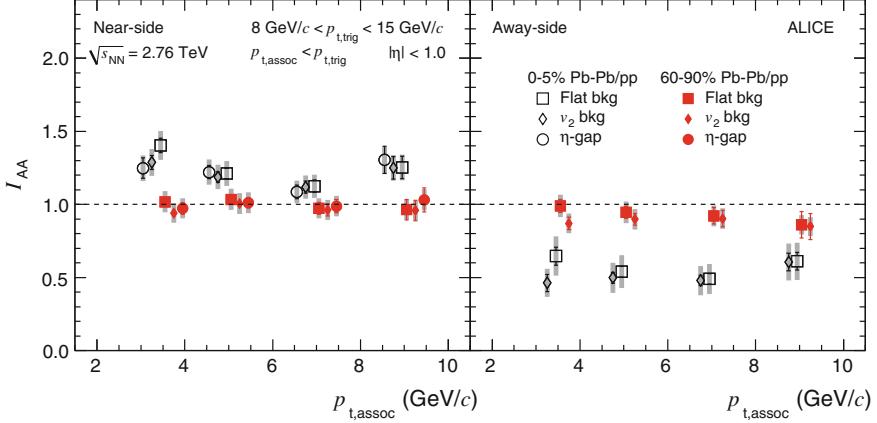


Fig. 22 Ratio of conditional yields (I_{AA}) in \AA and pp of hadrons correlated to the same-jet (*Near-side*) and recoil-jet (*Away-side*). For the most central events 0–5% the I_{AA} is suppressed for the recoil jet that interacted with the created medium. Figure from [46]

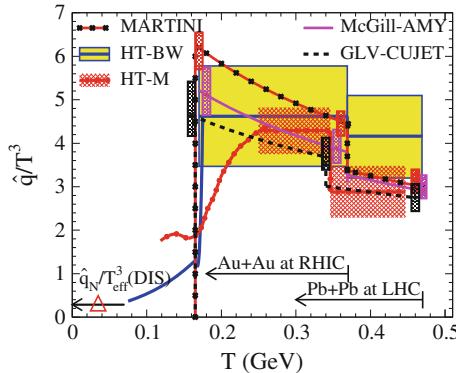


Fig. 23 (Color online) The assumed temperature dependence of the scaled jet transport parameter \hat{q}/T^3 in different jet quenching models for an initial quark jet with energy $E = 10 \text{ GeV}$. Values of \hat{q} at the center of the most central A+A collisions at an initial time $\tau_0 = 0.6 \text{ fm}/c$ in HT-BW and HT-M models are extracted from fitting to experimental data on hadron suppression factor R_{AA} at both RHIC and LHC. In GLV-CUJET, MARTINI and McGill-AMY model, it is calculated within the corresponding model with parameters constrained by experimental data at RHIC and LHC. Errors from the fits are indicated by *filled boxes* at three separate temperatures at RHIC and LHC, respectively. The *arrows* indicate the range of temperatures at the center of the most central A+A collisions. The *triangle* indicates the value of $\hat{q}_N/T_{\text{eff}}^3$ in cold nuclei from DIS experiments. Figure from [47]

does not give the desired access to the full dynamics of the parton energy loss mechanism. One method to overcome the biases and address quantitatively jet quenching is to study fully reconstructed jets. Jets reconstructed in terms of energy flow reduce uncertainties due to hadronization. Since jet quenching is a partonic process, jet

reconstruction are expected to allow for complete exploration of the dynamics of the process enabling more detailed comparison with theoretical descriptions beyond leading hadron observables.

3.3 Jet Quenching with Fully Reconstructed Jets

3.4 Experimental Aspects of Jet Reconstruction in Heavy-Ion Collisions

Jets are an attractive tool to study properties of QCD within heavy-ion collisions. The production of jets in electron-positron collisions, electron-nucleon deep inelastic scatterings, and hadron-hadron collisions has been experimentally studied for a number of decades now. Their production is well understood within the perturbative QCD (pQCD). The agreement between theory calculations and measurements of inclusive jet production cross-sections in hadron-induced reactions over a number of orders of magnitude in center-of-mass energy provides excellent grounding for use of jets as calibrated probes of the hot medium created in heavy-ion collisions (see for example [48]). Before we dive into the jet quenching measurements with fully reconstructed jets we briefly review several important aspects.

Jet definition. Although in general a jet can be defined as a collimated spray of particles originating from a highly virtual parton we need a stricter operational definition that can be applied both in theory and the experiment. The jets are defined by specifying few key parameters: the algorithm, the recombination scheme (the prescription of how the particle/4-vectors are combined), and the resolution parameter R . In the case of cone algorithms it is simply the radius $R = \sqrt{\Delta\eta + \Delta\varphi}$, where $\Delta\eta$ is the pseudo-rapidity distance between the jet axis and a constituent of the jet (a particle belonging to the jet) and $\Delta\varphi$ is the azimuthal separation between the jet centroid and the constituent. Moreover, the jet reconstruction algorithms are to be collinear safe and infrared safe. Clearly a strict relation between a jet and a parton is not unambiguous—jet is not a parton but rather a parton and its radiation constrained by an operational definition. In particular, a single parton shower can be reconstructed into multiple jets depending on the jet definition. For example, for some jet definitions and small resolution parameters the jet finding algorithm run on the very same event (set of particles produced within a collision) may report more than one object as compared to the same algorithm run with a large value of R . Therefore, it is crucial that when comparing to theory the same jet definitions are used. For an overview of the modern jet finder suite FASTJET please see [49] and references there in.

Jet reconstruction in heavy-ion collisions. The jet resolution parameters R used within heavy-ion collisions are smaller than in proton–proton collisions and range between 0.2 and 0.6. The reason for limiting the R is the large background composed of soft particles present in heavy-ion collisions. The larger the area of the jet

the larger the magnitude of the background. Moreover, the background is not uniformly distributed within a single event and fluctuates from region to region. These fluctuations impact the reconstructed jet energy. Authors of the FASTJET package [49] have originally provided a formulae describing the problem:

$$p_T^{\text{rec}} = p_T^{\text{raw}} - \rho A_{\text{jet}} \pm \sigma \sqrt{A_{\text{jet}}}, \quad (10)$$

where p_T^{rec} is the jet p_T corrected for the background effects, the p_T^{raw} denotes the momentum of the object returned by the jet finder (a jet candidate), the ρA_{jet} is the amount of p_T originated from background where ρ is the density of the background (can be obtained by various methods) and A_{jet} is the area of a jet, and finally, the term $\sigma \sqrt{A_{\text{jet}}}$ parametrizes the background fluctuations, where σ is the width of the fluctuations. Approaches using this formula (with some modifications) are nowadays broadly used in the analysis of jets heavy-ion collisions. In practice the reconstructed jet spectra is distorted by the $\sigma \sqrt{A_{\text{jet}}}$ term and may contain a significant number of so-called “false jets”. A false jet (an object reported by the jet finder) is a local upward fluctuations above the average background that is not correlated to any hard scattering but still fulfilling the condition $p_T^{\text{raw}} - \rho A_{\text{jet}} > 0$. The experiments must correct for the contribution of the false jets to the reconstructed jet yield. This can be achieved by either suppressing the false jets by selecting jets with a high- p_T particle that certainly originated from a hard scattering or by subtracting the estimated false jet yield on a statistical basis. Finally, the measured spectrum is a result of convolution of the true jet spectrum with the smearing function $f(p_T, \sigma \sqrt{A_{\text{jet}}})$. To account for the smearing effect the reconstructed spectrum must be de-convoluted. Note, that the strength of the background depends on the selection of the p_T of particles used as the input to the jet finders, and so is the magnitude of the fluctuations. The higher the p_T cuts applied the lower the background and smaller σ . However, with applying the p_T cuts comes the danger of suppressing the physical signal of jet quenching—the soft fragments originated by medium-induced gluon radiation. The hard p_T cuts may remove beyond recovery the information about the soft components of the quenched jet—the very signal that experiment is to measure. To quantify the statistical (region-to-region) fluctuations of the background the transverse momentum of charged particles in the randomly positioned cone (RC) is summed and the difference $\delta p_T^{\text{ch}} = \sum^{\text{RC}} p_{T,i} - \rho \pi R^2$ is calculated. In the left panel of Fig. 24 the distribution of the δp_T measured with charged particles by ALICE is shown for 0–10 % of the most central Pb–Pb collisions. The distribution shows a tail towards large positive values related to the presence of the hard scatterings and depends on R . The right panel of Fig. 24 shows the dependence of the width of the δp_T distribution on the event centrality for $R = 0.2$ and $R = 0.3$. Smaller fluctuations are found for smaller radii with stronger centrality dependence for larger R .

Hadronization corrections in heavy-ion collisions. In the so-called vacuum case (e^+e^- or pp collisions) experiments often correct the jet observables to the so-called *generator-level* or *particle-level* jets. These results can be directly compared to the output of the calculation and/or MC generators. While in heavy-ion collisions one does correct the measurements to the generator level the exact corrections for effects

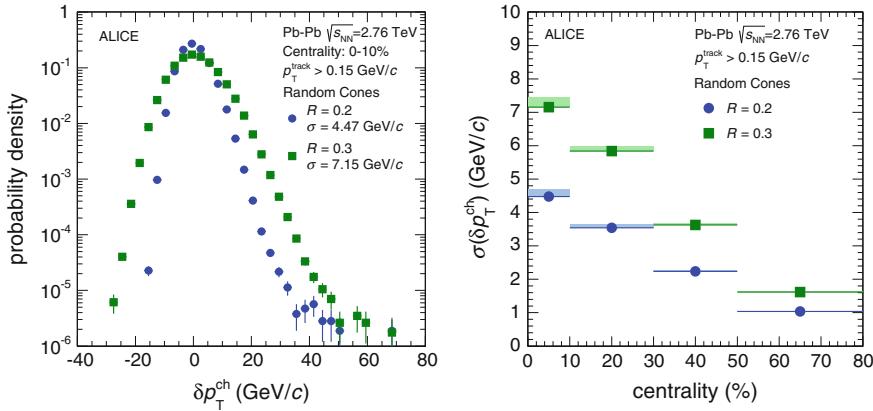


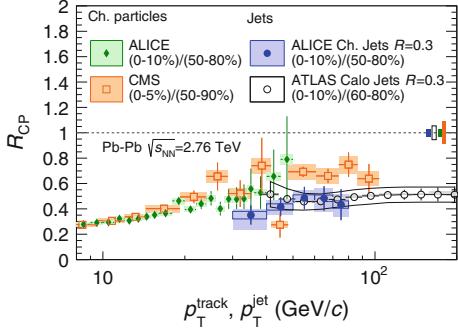
Fig. 24 *Left* δp_T^{ch} distribution for jets with resolution parameter $R = 0.2$ and $R = 0.3$ measured with random cones in central collisions. *Right* Width of the background fluctuation δp_T^{ch} distribution as a function of centrality for cone radii $R = 0.2$ and $R = 0.3$. The shaded uncertainty bands indicate the difference between the width of the δp_T^{ch} distribution from random cones and high p_T probe embedding. Figure from [50]

related strictly to the *hadronization* of the jet are experimentally difficult and to some extent conceptually undefined. While this should be taken into account when comparing to theoretical calculations at present the uncertainty related to hadronization is much smaller than uncertainties related to the background corrections.

3.5 Jets in AA Collisions at the LHC

We have seen the first signal of jet quenching via the suppression of leading hadrons. With fully reconstructed jets one can ask a crucial question: where does the energy go? Note that if the medium-induced radiation does not de-cohere from the jet and remains within the jet cone the total energy (or p_T) of the jet should remain unmodified resulting in R_{AA} consistent with unity. The observations at the LHC point; however, towards a different scenario. Figure 25 compares the R_{CP} for single hadrons measures by ALICE and CMS and with the R_{CP} for jets reconstructed with charged particles (ALICE) and jets measured by CMS. The R_{CP} is simply a ratio of yields from central (C) to the yields recorded in peripheral (P) events scaled by the relative number of participating nucleons such that it becomes unity at high- p_T in the case no nuclear effects are present. At the LHC a strong suppression is found for not only hadrons but also jets. The R_{CP} shows values of about 0.5 and does not depend on the jet energy for $p_T > 40 \text{ GeV}/c$. This is a novel information about the jet quenching process. The modification of the jet p_T spectrum resulting in the $R_{\text{CP}} < 1$ states that in case of heavy-ion collisions a fraction of the jet energy is radiated outside the jet cone.

Fig. 25 Comparison to jet R_{CP} measured by ATLAS [51] and to charged particle suppression by ALICE [52] and CMS [41]. The percentiles of centrality used to build the R_{CP} ratio are indicated within the figure. Figure from [50]



On the other hand, the flatness of the R_{CP} as compared to the evolution of the R_{AA} as a function of p_{T} for charged hadrons is also a valuable information; however, the inclusive distributions depend on the steepness of the production spectra and for further conclusions we must turn to more differential observables.

The first measurement of jet quenching with fully reconstructed jets in heavy-ion collisions focused on the balance between the highest transverse energy pair of jets in events where the two jets have an azimuthal angle separation, $\Delta\varphi = |\varphi_1 - \varphi_2| > \pi/2$. The ATLAS Collaboration has analyzed Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV and extracted the dijet asymmetry A_j defined as:

$$A_j = \frac{E_{\text{T}1} - E_{\text{T}2}}{E_{\text{T}1} + E_{\text{T}2}}, \quad (11)$$

where the first jet is required to have a transverse energy $E_{\text{T}1} > 100$ GeV, and the second jet is the highest transverse energy jet in the opposite hemisphere with $E_{\text{T}2} > 25$ GeV. The jets were reconstructed using the anti- k_{T} algorithm with $R = 0.4$. The results shown in Fig. 26 have two distinct features:

- The asymmetry in peripheral events is consistent with the pp data while for central events it shows strong modifications: a depletion for small A_j and a maximum at about 0.4. This is a clear indication that the recoiling jet loses a significant fraction of its energy.
- The acoplanarity of the dijet system (demonstrated by the $\Delta\varphi$ distribution) remains unchanged with respect to pp collisions. This is observed even for the most central collisions where the quenching effects are expected to be the strongest.

These two observations provide a wealth of information about the nature of jet quenching beyond the single hadron and jet R_{AA} . The modification of the A_j confirmed the conclusion that the medium induced radiation is transported outside of the jet cone and it is consistent with the $R_{\text{AA}} < 1$ (or $R_{\text{CP}} < 1$). On the other hand, the lack of medium induced acoplanarity provides strong constraints to the theoretical considerations of the quenching process: although jets loose significant amount of energy, their original direction remains unmodified.

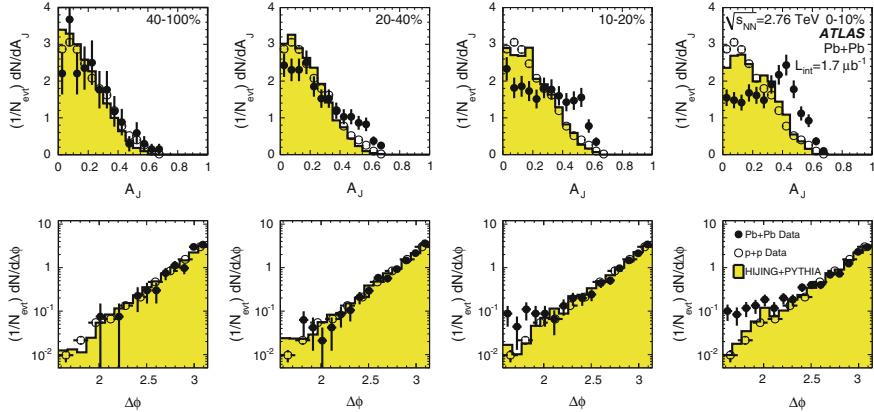


Fig. 26 (top) Dijet asymmetry distributions for data (points) and unquenched HIJING with superimposed PYTHIA dijets (solid yellow histograms), as a function of collision centrality (left to right from peripheral to central events). Proton–proton data from $\sqrt{s} = 7$ TeV, analyzed with the same jet selection, is shown as open circles. (bottom) Distribution of $\Delta\phi$, the azimuthal angle between the two jets, for data and HIJING+PYTHIA, also as a function of centrality. Figure from [5]

3.6 Jet Fragmentation and Where Does the Energy Go?

Jet quenching is a modification of the jet fragmentation. We have seen that the radiation is not captured within the radius R of the reconstruction algorithms. The question is, how is the internal jet structure modified? A fragmentation function $D_i^h(z, Q)$ of a parton is defined as a probability that a hadron of type h carries longitudinal momentum fraction z of the momentum \mathbf{p}_i of a parton type i such that $z = (\mathbf{p}_i \cdot \mathbf{p}_h)/|\mathbf{p}_i|^2$, where Q is the scale of the hard scattering process that produced the parton. The value of D is not calculable perturbatively; however, the variation of D with Q can be predicted for large Q . Experimentally in hadronic collisions one does not measure D directly; however, a related quantity associated to a jet can be measured. Experiments can report a per jet quantity $F(z, p_{T,\text{jet}}) = 1/N_{\text{jet}} dN_h/dz$, where N_h is the number of particles in the jet, and $z = (\mathbf{p}_{\text{jet}} \cdot \mathbf{p}_h)/|\mathbf{p}_{\text{jet}}|^2$. $F(z, p_{T,\text{jet}})$ is a sum over $D_i^h(z, Q)$ weighted by the rate at which parton species (i) is produced from the hard scattering. The hard scattering scale Q is of the same order of magnitude as $p_{T,\text{jet}}$. The jet fragmentation functions for charged particles have been measured extensively in pp collisions at the LHC (see [53] for example).

To enhance the sensitivity to the low-momentum constituents of a jet the heavy ion experiments often use $F(\xi)$ which is related to $F(z)$ as $\xi = \ln(1/z)$. Figure 27 shows the jet fragmentation function using the ξ variable reconstructed by the CMS Collaboration for leading and sub-leading jets in Pb–Pb collisions and the pp reference in several regions of A_J [54]. The leading jet in this case was $p_{T,1} > 100$ GeV/ c and the sub-leading $p_{T,2} > 40$ GeV/ c . The differences between the pp (vacuum) fragmentation and the Pb–Pb case are quite small—clearly seen in the lower panel

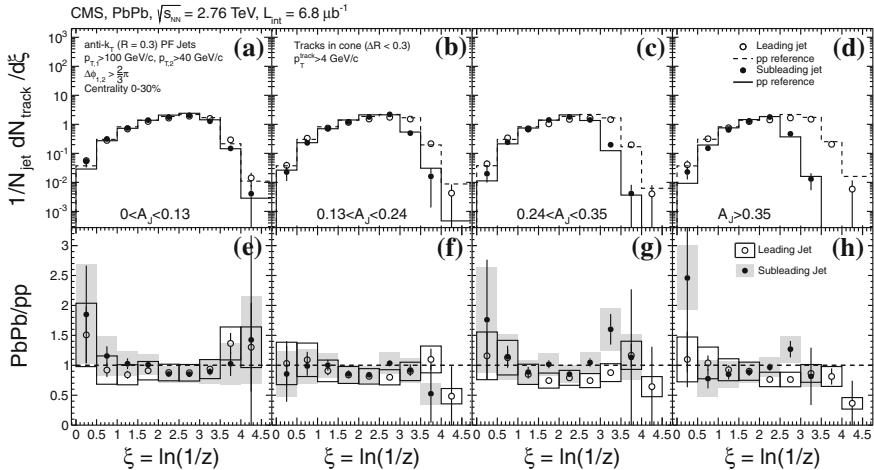


Fig. 27 **a–d** Fragmentation functions for the leading (open circles) and subleading (solid points) jets in four regions of A_J in central PbPb collisions compared to the pp reference. **e–h** Ratio of each fragmentation function to its pp-based reference. Error bars shown are statistical. The systematic uncertainty is represented by hollow boxes (leading jet) or gray boxes (subleading jet). Figure from [54]

where the ratio of Pb–Pb and *pp* measurements is shown. The question remains then: where does the radiated energy go? Recent measurements of jet shapes [55] and particle distributions with respect to the di-jet axis revealed that the energy lost to the medium seems to be “recovered” at much larger angles than R . Moreover, the radiated energy is transformed into an enhanced multiplicity of soft particles $p_T < 2 \text{ GeV}/c$ [56]. Additional information on the fate of the radiated energy can be extracted from the photon-jet correlation measurements. In this process the photon serves as an excellent reference for the jet energy and accoplanarity of the system (for more details see [57]). The dijet and γ -jet measurements at the LHC indicate that the energy loss is constant with the jet p_T and the average energy lost by the jet is of the order of 10–20 GeV. Much of the theoretical work is presently focussed on incorporating the jet measurements into the theoretical frameworks.

3.7 Energy Loss of the Heavy-Flavor

Heavy-quarks are produced early in heavy-ion collisions and their abundances do not change through the evolution of the collision system. They are produced in hard collisions (because of their large mass) and their production can be calculated within the pQCD framework. This makes them excellent auto-generated probes of the medium. However, the central point of the considerations of heavy-quark energy loss is that the gluon radiation from a highly energetic parton traversing the quark-

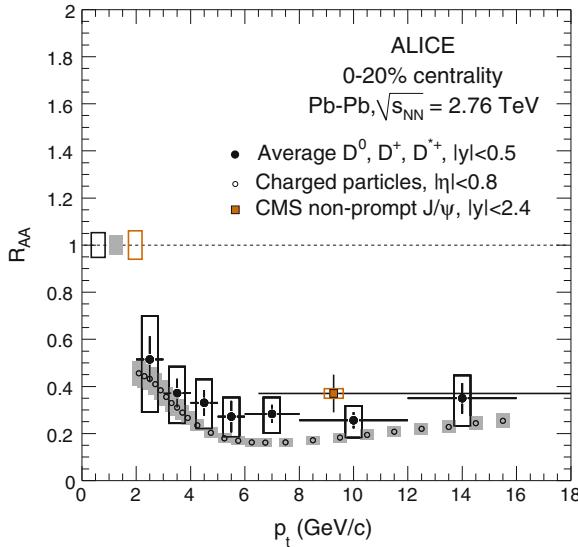


Fig. 28 Average R_{AA} of D mesons in the 0–20 % centrality class compared to the nuclear modification factors of charged particles [52] and non-prompt J/ψ from B decays [58] in the same centrality class. The charged particle R_{AA} is shown only for $2 < p_T < 16$ GeV/c . The three normalization uncertainties shown in the *right-hand panel* are almost fully correlated. Figure from [59]

gluon plasma shows a characteristic hierarchical dependence on the color charge and mass of the parton projectile: $\Delta E_{\text{gluon}} > \Delta E_{\text{light quark}} > \Delta E_{\text{heavy quark}}$. Here, the first inequality follows from the larger color charge of partons in the adjoint representation. The second inequality is due to the dead cone effect, which suppresses radiation of massive particles in the vacuum and in the medium. Moreover, theoretical calculation show that the energy loss of a quark is inverse proportional to its mass ($\Delta E \sim 1/M$) and provides an excellent handle on the diffusion coefficient of the QGP.

Figure 28 shows the nuclear modification factor measured for charged particles, D -mesons but also the non-prompt J/ψ . This is one of the first comparisons of the energy loss of light and heavy-quarks. The non-prompt J/ψ serves as proxy for the B -meson (hence an estimates of the b -quark energy loss). The R_{AA} for the charmed meson seems to be somewhat higher than for all charged particles (predominately originating from the gluon fragmentation); however, the current uncertainties do not allow to formulate firm conclusions. Similar observation can be made for the non-prompt J/ψ . It is somewhat less suppressed than the charmed meson; however, in this case also one must handle the data with appropriate attention to the systematic uncertainties. Moreover, the R_{AA} strongly depends on the production spectra of the particles. In addition the p_T of a hadron is a very poor approximation for the parton p_T . Hence, the J/ψ originates from much higher p_T than the present kinematic reach for the D -meson and the comparison is even more complicated. Nevertheless,

this is an excellent example where the high-rates for rare probes at the LHC allow to significantly extend the current knowledge. In fact, the detail measurements of heavy-flavor are one of the most important milestones for the future runs of the accelerator.

3.8 Thermalization of Charm

Further increase of precision is also needed to establish a precision measurement of the elliptic flow of charm. For instance in the case the charm quark thermalizes within the hot QGP similarly to the light flavor one should observe a positive v_2 . The first measurements at the LHC of the elliptic flow exhibit such signal [61]. The left panel of Fig. 29 presents the invariant mass of D -meson decay into charged kaon and a pion ($D^0 \rightarrow K^-\pi^+$) for two orthogonal selection of azimuthal angle with respect to the event plane. The v_2 was extracted as

$$v_2\{EP\} = \frac{1}{\chi} \frac{\pi}{4} \frac{N_{\text{in-plane}} - N_{\text{out-of-plane}}}{N_{\text{in-plane}} + N_{\text{out-of-plane}}}, \quad (12)$$

where χ is a correction factor for the event plane resolution, and N denotes the yield of measured particles in- or out-of-plane. The $v_2(p_T)$ for semi-central Pb–Pb collisions is shown in the right panel of Fig. 29. While at present the uncertainties are large one can clearly see that the elliptic flow for D -mesons can be as large as for

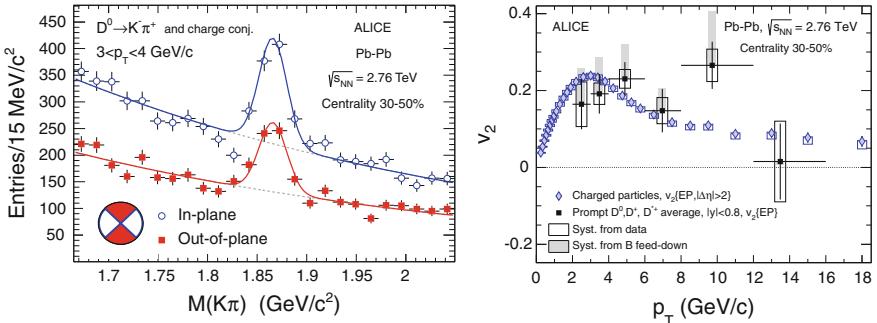


Fig. 29 *Left* Invariant mass distributions for D^0 candidates and their charge conjugates with $3 < p_T < 4 \text{ GeV}/c$ for 9.5×10^6 Pb–Pb collisions in the 30–50% centrality class. The distributions are shown separately for the in-plane (open symbols) and out-of-plane (closed symbols) intervals of azimuthal angle. The curves show the fit functions as described in the text. *Right* Average of D^0 , D^+ and D^{*+} v_2 as a function of p_T , compared to charged-particle v_2 [60] measured with the event plane (EP) method. The symbols are positioned horizontally at the average p_T of the three D meson species. Figure from [61]

the light flavor hadrons. This would support the picture where even the heavy quarks through multiple interactions with the medium constituents may be driven into local equilibrium with the hot bulk of partons.

4 Quarkonia

Quarkonia (bound states of heavy quark and anti-quark) play historically an important role in establishing the properties of QGP; however, the complete description of their production in heavy-ion collisions remains a challenge to theory. Charmonia states for instance are an attractive probe of the strongly coupled system of quarks and gluons as their sizes can be smaller than light hadrons (down to a few tenths of a fm) and having large binding energies (>500 MeV). Early on it was suggested [62] that the quarkonia can be used as a *thermometer* of the QGP [63]. Various states would dissociate within the plasma depending on the screening potential of the QGP. The so-called sequential suppression was predicted: on average the heavy-ion collisions would produce fewer quarkonia states as compared to the extrapolation from the proton–proton collisions, and the states strongly bound would melt at higher temperatures than the less bound states. The color screening would dissociate charmonium in QGP such that first the higher excited states (2S), (1P), then the ground state (1S) would break up (see Fig. 30). The first predictions pointed that the J/ψ (1S) would not survive temperatures of about 2 times T_c , the critical temperature needed for creating the QGP. At present according to lattice QCD calculation this number is estimated between 1.2 and 1.9 T_c depending on the details of the lattice calculations. On the other hand, a compensating scenario may take place. Namely, first we note that all primary charmonia dissociated at high collision energy contribute to the abundance of charm quarks. Secondly, charm and anti-charm quarks can equilibrate with the rest of the plasma and a substantial amount of the $c\bar{c}$ pairs may survive until hadronisation when by statistical combination charmonia states in addition to the open charm hadrons can be created. Such scenario is presented in Fig. 30 in terms of production probability of J/ψ as a function of the energy density. Finally, to fully quantify the effect of hot QGP on the charmonia states it is required to understand the so-called cold nuclear effects. These can be of two types. One, related to the modifications of the parton distribution functions within nuclei (the so-called shadowing and anti-shadowing) that may have a direct impact on the observed number of produced quarkonia. The second effect originates from the final probability for energy loss of quarkonia within the nuclei. This final-state energy loss in the cold nuclear matter may also directly influence the observed p_T spectrum of quarkonia states. Both of these effects are present in the case of AA collisions. An experimental access to these effects can be provided by the collisions of protons and nuclei.

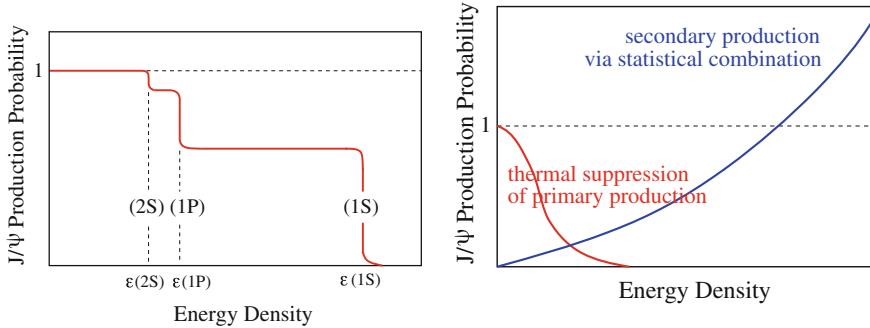


Fig. 30 *Left* A schematic representation of sequential suppression of charmonia states as a function of energy density. The probability is normalized to the total number of charmonia pairs. *Right* A qualitative conjecture of dissociation and re-combination of J/ψ as a function of energy density. The probability is normalized per initial J/ψ particle

4.1 J/ψ R_{AA} at RHIC and the LHC

The experiments at the SPS have delivered the first clear signals for J/ψ suppression in most central Pb–Pb collisions at the top SPS energies ($\sqrt{s_{NN}} = 17.3$ GeV) [64]. Here, however, we focus on the comparison of the results obtained at RHIC and the LHC. The left panel of Fig. 31 shows the comparison of the measurements of $J/\psi R_{AA}$ measured at RHIC and the LHC as a function of $\langle N_{\text{part}} \rangle$. The comparison as a function of $\langle N_{\text{part}} \rangle$ is particularly interesting as it puts the centrality dependence of the R_{AA} on a similar footing at the two energies. The small $\langle N_{\text{part}} \rangle$ corresponds to the peripheral collisions (small volume of the system) while the largest $\langle N_{\text{part}} \rangle$ values correspond to the most central AA collisions (the largest overlap of the colliding ions). The measurements at the lower energy by the PHENIX Collaboration show a decreasing trend with centrality. Greater number of J/ψ are recorded for peripheral collisions as compared to the central collisions where the production seems to be suppressed up to a factor of 5. This observation is consistent with an expectation of J/ψ dissociating in the hot QGP. A quantitatively different picture is observed at the LHC. The R_{AA} measured by the ALICE Collaboration in peripheral collision matches the suppression observed at RHIC; however, for the more central collisions the R_{AA} sets at about 0.6 and remains constant up to the highest $\langle N_{\text{part}} \rangle$. Again, this observation is consistent with the dissociation of J/ψ within the hot medium; however, it also reveals an additional phenomenon that must be at play. Additional information can be obtained from the p_T dependence of the R_{AA} in the most central collisions shown in the right panel of Fig. 31. The R_{AA} measured at RHIC is flat at about 0.2 while the same measurement at the LHC shows a pronounced p_T dependence. The R_{AA} measured by ALICE shows a steep decrease from about 0.8 at the lowest momenta of J/ψ to values of about 0.3 at the highest p_T . This observation demonstrates that on average at the LHC a larger fraction of J/ψ emerges from the heavy-ion collision as compared to RHIC. Note, the R_{AA} is a relative measure of suppression normalized to

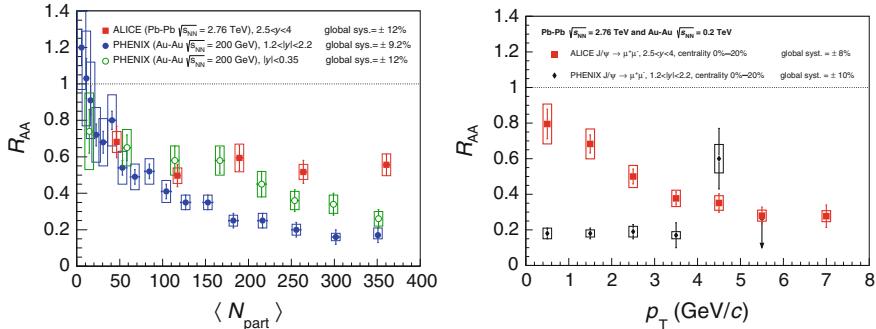


Fig. 31 *Left* Nuclear modification factor as a function of the number of participating nucleons ($\langle N_{part} \rangle$). Data from PHENIX at RHIC and ALICE at the LHC. The clear suppression trend from peripheral to central collisions observed at RHIC (PHENIX) is changed at the LHC (ALICE). The R_{AA} at the LHC shows no dependence on N_{part} for the semi-central and most central collisions—a result consistent with the higher probability for re-generation of J/ψ at the LHC energies. Figure from [65]. *Right* Comparison of the p_T dependence of the R_{AA} at RHIC and the LHC. The trend of the R_{AA} at the LHC is consistent with the expectation of re-generation of J/ψ from the charm rich QGP at the LHC. Figure from [66]

the measurements at the same \sqrt{s} in pp collisions. In other words, we observe that a larger number of J/ψ per nucleon-nucleon collision is emitted in the heavy-ion collisions at the LHC than at RHIC. Moreover, it is the low- p_T J/ψ that seem to drive the R_{AA} at the LHC upwards. Is this consistent with the “hotter” medium at the LHC? One of the possible explanations comes from the re-combination mechanism of the thermalized $c\bar{c}$ pairs mentioned earlier. With the much larger production cross-sections for charm quarks at the LHC as compared to RHIC the medium created at $\sqrt{s_{NN}} = 2.76$ TeV is much more charm-rich and allows for additional J/ψ to be formed in the late stages of the collision. To further substantiate the dissociation and re-combination mechanisms at play more precise measurements (down to lowest momenta) of the total charm created in heavy-ion collisions are needed. This could be achieved, for example, via future measurements of the open charm hadrons down to approximately zero p_T .

4.2 Sequential Suppression of Quarkonia

One of the most striking experimental evidence for the sequential suppression of the quarkonia states is shown in Fig. 32. The invariant-mass distribution of di-muons was reconstructed in pp and $Pb-Pb$ collisions by the CMS collaboration. The plots focus on the mass region of the $b\bar{b}$ states. It is evident that the higher states of 2S and 3S are no longer resolvable in the $Pb-Pb$ case. A very different case than in pp collisions where the three peaks of the Y states are visible.

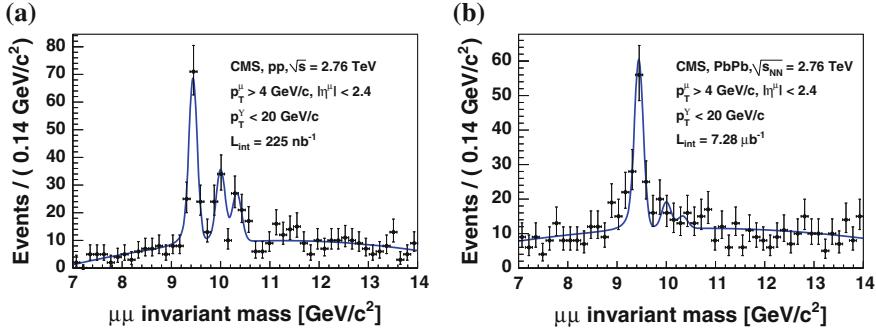


Fig. 32 Dimuon invariant-mass distributions from the pp (a) and $PbPb$ (b) data at $\sqrt{s_{NN}} = 2.76$ TeV. The same reconstruction algorithm and analysis criteria are applied to both data sets, including a transverse momentum requirement on single muons of $p_T^\mu > 4$ GeV/c. The solid lines show the result of the fit described in the text. Figure from [67]

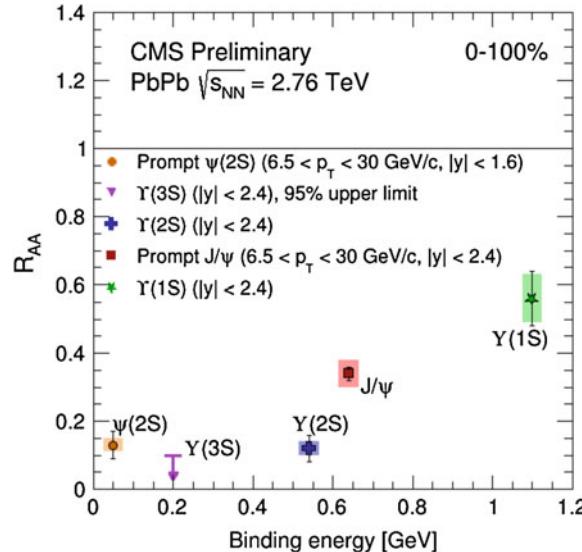


Fig. 33 Nuclear modification factor demonstrating the suppression of quarkonia states as a function of their binding energy. Figure based on [68]

The measurements of the modification factors for J/ψ and Y states can be ordered with respect to their binding energy. Figure 33 shows clearly the suppression pattern that is consistent with the expectation for the sequential melting. The states with small binding energy ($\psi(2S)$, $Y(3S)$ and $Y(2S)$) are strongly suppressed. While the J/ψ and $Y(1S)$ show much larger R_{AA} .

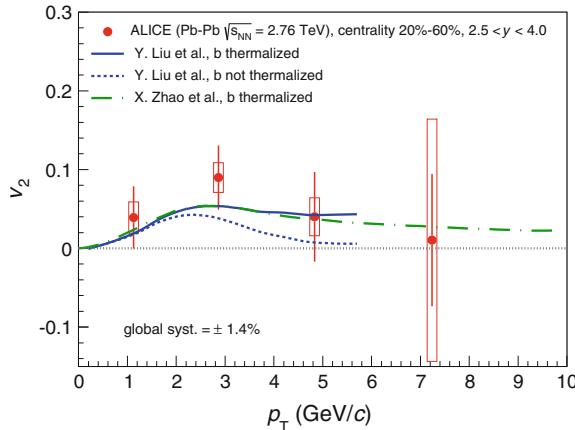


Fig. 34 (color online) Inclusive J/ψ $v_2(p_T)$ for non-central (20–60 %) Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. The used p_T ranges are: 0–2, 2–4, 4–6 and 6–10 GeV/c . Calculations from two transport models [69, 70] are also shown (see text for details). Figure from [71]

4.3 J/ψ v_2

In the previous sections we have discussed that charm quarks interact with the medium and may be affected by the thermal parton bath of QGP. Moreover, as stated earlier we may have hints that a fraction of J/ψ is constructed from the thermalized charm quarks at the hadronization stage. One of the most striking result supporting this scenario is the observation of the elliptic flow of the produced charmonia. The elliptic flow is quantified by the magnitude of the modulation of the azimuthal distribution of particles with respect to the reaction plane. Figure 34 shows the measurement of finite v_2 of J/ψ mesons by the ALICE Collaboration. The measurement shows a clear indication for positive v_2 in the intermediate p_T . This observation is consistent with a scenario where a substantial number of J/ψ mesons is created from quarks that flow together with the medium. While the precision of the measurements will improve with the future LHC runs, already now the v_2 combined with the R_{AA} provides stringent tests for modeling J/ψ within the hot, strongly interacting medium.

4.4 Cold Nuclear Matter Effects

To investigate *cold nuclear matter* (CNM) effects relevant for quarkonium production, LHC experiments studied proton–lead interactions at $\sqrt{s} = 2.76$ TeV. The quarkonia is measured in the proton going direction and the nuclei going direction. Apart from the initial conditions affecting the production cross-section one expects

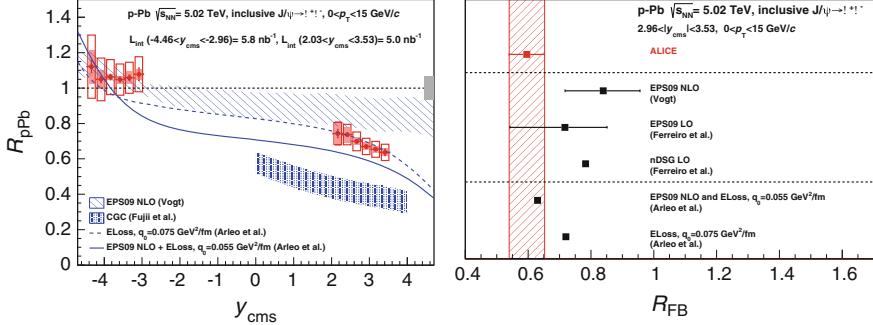


Fig. 35 *Left* The R_{AA} for inclusive J/ψ production at $\sqrt{s_{NN}} = 5.02 \text{ TeV}$, in bins of rapidity. Results from various models are also shown. The theoretical uncertainties for the EPS09 NLO calculation are due to the uncertainty on the shadowing parameterization and to the mass and scale uncertainties on the cross section calculation. For the CGC model, the band is related to the choice of the parton saturation scale and of the charm quark mass. Finally, the q_0 value in the energy loss model represents the value of the transport coefficient in the target nucleons for Bjorken $x_{Bj} = 10^{-2}$ gluons. *Right* The forward to backward ratio R_{FB} of the nuclear modification factors for inclusive J/ψ production, compared to theoretical models. The statistical and systematic uncertainties for the experimental value are added in quadrature. Figures are taken from [72]. See the [72] reference for details on the used models

several final state effects. The J/ψ traversing the nuclei may loose energy and/or dissociate due to interactions with the matter of the nuclei. In summary, the quarkonia production can be affected by the effects related to parton shadowing, gluon saturation, but also by the energy loss or final state break-up. Note, that all these effects with various magnitudes are present in nucleus-nucleus collisions. Thus, the measurements in p - Pb collisions serve as an important reference for nucleus-nucleus collisions.

The nuclear modification factor R_{pPb} for J/ψ [72] as a function of the quarkonium rapidity y_{cms} is shown in Fig. 35. The results are compared with theory predictions, based on a pure nuclear shadowing scenario, as well as partonic energy loss, either in addition to shadowing or as the only nuclear effect. Within the uncertainties, the model based on nuclear shadowing is in fair agreement with the data. Models including a contribution from parton energy loss are likewise in fair agreement with the data, while the prediction based on Color Glass Condensate model underestimate the results. These measurements constitute crucial quantification of the cold nuclear effects and provide additional confirmation that the suppression observed in Pb - Pb collisions is a consequence of the QGP formation. In fact, with some limitations one may consider that the contribution to the cold nuclear matter effects for the R_{AA} measured in Pb - Pb collisions may be constructed from the measurement of R_{pPb} , such that $R_{AA}^{CNM} = R_{pPb}^2 = R_{pA} \times R_{Ap}$. The distinction between R_{Ap} and R_{pA} comes from the consideration of measurements in a rapidity interval for the two collision configuration—either p - A or A - p indicating the proton (or nuclei) going directions. To further quantify the nuclear effects at play one extracts the so-

called forward-backward ratio R_{FB} of the nuclear modification factors. The R_{FB} is attractive since it is not affected (unlike R_{AA}) by the production cross-section of J/ψ in pp collisions, nor by the values of the overlap function T_{AA} used to form the R_{pPb} (scaling by the number of binary collisions). In the right panel of Fig. 35 the R_{FB} is compared to various models (for details see [72]). The agreement between data and the model including both shadowing and coherent energy loss is very good, while pure shadowing scenarios seem to overestimate R_{FB} . We must note, however, that that the models should seek agreement with the measured R_{FB} and the R_{AA} as the forward-backward ratio does not constrain the absolute values of the nuclear effects. For recent results on the ψ (2S) in p–Pb collisions see [73].

References

1. A. Bazavov et al., HotQCD Collaboration, Chiral and deconfinement aspects of the QCD transition. *Phys. Rev. D* **85**, 054503 (2012)
2. D.J. Schwarz, The first second of the universe. *Ann. Phys.* **12**, 220–270 (2003). [arXiv:astro-ph/0303574](https://arxiv.org/abs/astro-ph/0303574)
3. R. Rapp, H. van Hees, M. He, Properties of thermal photons at RHIC and LHC. *Nucl. Phys. A* (2014) (in press)
4. H. Petersen, M. Bleicher, S.A. Bass, H. Stocker, UrQMD v2.3: changes and comparisons. [arXiv:0805.0567](https://arxiv.org/abs/0805.0567)
5. G. Aad et al., ATLAS Collaboration, Observation of a centrality-dependent dijet asymmetry in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS detector at the LHC. *Phys. Rev. Lett.* **105**, 252303 (2010). [arXiv:1011.6182](https://arxiv.org/abs/1011.6182)
6. G. Aad et al., ATLAS Collaboration, The ATLAS experiment at the CERN Large Hadron Collider. *JINST* **3**, S08003 (2008)
7. M.L. Miller, K. Reygers, S.J. Sanders, P. Steinberg, Glauber modeling in high energy nuclear collisions. *Ann. Rev. Nucl. Part. Sci.* **57**, 205–243 (2007). [arXiv:nucl-ex/0701025](https://arxiv.org/abs/nucl-ex/0701025)
8. B. Abelev et al., ALICE Collaboration, Centrality determination of Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV with ALICE. *Phys. Rev.* **C88**(4), 044909 (2013). [arXiv:1301.4361](https://arxiv.org/abs/1301.4361)
9. K. Aamodt et al., ALICE Collaboration, Charged-particle multiplicity density at mid-rapidity in central Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. Lett.* **105**, 252301 (2010). [arXiv:1011.3916](https://arxiv.org/abs/1011.3916)
10. E. Abbas et al., ALICE Collaboration, Centrality dependence of the pseudorapidity density distribution for charged particles in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Lett.* **B726**, 610–622 (2013). [arXiv:1304.0347](https://arxiv.org/abs/1304.0347)
11. T. Sjöstrand, S. Mrenna, P.Z. Skands, A brief introduction to PYTHIA 8.1, *Comput.i Phys. Commun.* **178**, 852–867 (2008). [arXiv:0710.3820](https://arxiv.org/abs/0710.3820)
12. B.B. Abelev et al., ALICE Collaboration, Multiplicity dependence of the average transverse momentum in pp , p–Pb, and Pb–Pb collisions at the LHC. *Phys. Lett.* **B727**, 371–380 (2013). [arXiv:1307.1094](https://arxiv.org/abs/1307.1094)
13. B. Abelev et al., ALICE Collaboration, Centrality dependence of π , K, p production in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev.* **C88**(4), 044910 (2013). [arXiv:1303.0737](https://arxiv.org/abs/1303.0737)
14. K. Aamodt et al., The ALICE Collaboration, The ALICE experiment at the CERN LHC. *J. Instrum.* **3**(08), S08002 (2008)
15. B.B. Abelev et al., ALICE Collaboration, Performance of the ALICE experiment at the CERN LHC. [arXiv:1402.4476](https://arxiv.org/abs/1402.4476)
16. F. Ronchetti, The ALICE Collaboration, The ALICE electromagnetic calorimeter project. *J. Phys.: Conf. Ser.* **160**(1), 012012 (2009)

17. J. Proudfoot, Hadron Calorimetry at the LHC, SLAC, <http://www.slac.stanford.edu/econf/C060717/papers/L012.PDF>. Accessed July 2006
18. S. Chatrchyan et al., CMS Collaboration, The CMS experiment at the CERN LHC. *JINST* **3**, S08004 (2008)
19. J. Alves, A. Augusto et al., LHCb Collaboration, The LHCb detector at the LHC. *JINST* **3**, S08005 (2008)
20. K. Adcox et al., PHENIX Collaboration, PHENIX detector overview. *Nucl. Instrum. Meth.* **A499**, 469–479 (2003)
21. K. Ackermann et al., STAR Collaboration, STAR detector overview. *Nucl. Instrum. Meth.* **A499**, 624–632 (2003)
22. B.B. Abelev et al., ALICE Collaboration, K_S^0 and Λ production in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. Lett.* **111**(22), 222301 (2013). [arXiv:1307.5530](https://arxiv.org/abs/1307.5530)
23. E. Schnedermann, J. Sollfrank, U. Heinz, Thermal phenomenology of hadrons from 200A GeV S+S collisions. *Phys. Rev. C* **48**, 2462–2475 (1993)
24. M. Floris, Hadron yields and the phase diagram of strongly interacting matter. [arXiv:1408.6403](https://arxiv.org/abs/1408.6403)
25. A. Andronic, P. Braun-Munzinger, K. Redlich, J. Stachel, The thermal model on the verge of the ultimate test: particle production in Pb-Pb collisions at the LHC. *J. Phys.* **G38**, 124081 (2011). [arXiv:1106.6321](https://arxiv.org/abs/1106.6321)
26. J. Cleymans, I. Kraus, H. Oeschler, K. Redlich, S. Wheaton, Statistical model predictions for particle ratios at $\sqrt{s_{NN}} = 5.5$ TeV. *Phys. Rev.* **C74**, 034903 (2006). [arXiv:hep-ph/0604237](https://arxiv.org/abs/hep-ph/0604237)
27. M. Petrán, J. Letessier, V. Petráček, J. Rafelski, Hadron production and quark-gluon plasma hadronization in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev.* **C88**(3), 034907 (2013). [arXiv:1303.2098](https://arxiv.org/abs/1303.2098)
28. J. Adams et al., Experimental and theoretical challenges in the search for the quark-gluon plasma: the STAR collaboration’s critical assessment of the evidence from RHIC collisions. *Nucl. Phys. A* **757**(1–2), 102–183 (2005). (First Three Years of Operation of RHIC)
29. P. Kolb, U. Heinz, *Quark-Gluon Plasma 3*, ed by R.C. Hwa, X.N. Wang (World Scientific, Singapore, 2004), pp. 634–714
30. J.-Y. Ollitrault, Relativistic hydrodynamics for heavy-ion collisions. *Eur. J. Phys.* **29**, 275–302 (2008). [arXiv:0708.2433](https://arxiv.org/abs/0708.2433)
31. S.A. Voloshin, A.M. Poskanzer, R. Snellings, Collective phenomena in non-central nuclear collisions. [arXiv:0809.2949](https://arxiv.org/abs/0809.2949)
32. A. Andronic et al., FOPI Collaboration, Excitation function of elliptic flow in Au+Au collisions and the nuclear matter equation of state. *Phys. Lett.* **B612**, 173–180 (2005). [arXiv:nucl-ex/0411024](https://arxiv.org/abs/nucl-ex/0411024)
33. K. Aamodt et al., ALICE Collaboration, Elliptic flow of charged particles in Pb-Pb collisions at 2.76 TeV. *Phys. Rev. Lett.* **105**, 252302 (2010). [arXiv:1011.3914](https://arxiv.org/abs/1011.3914)
34. K. Aamodt et al., ALICE Collaboration, Higher harmonic anisotropic flow measurements of charged particles in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. Lett.* **107**, 032301 (2011). [arXiv:1105.3865](https://arxiv.org/abs/1105.3865)
35. P. Kovtun, D.T. Son, A.O. Starinets, Viscosity in strongly interacting quantum field theories from black hole physics. *Phys. Rev. Lett.* **94**, 111601 (2005). [arXiv:hep-th/0405231](https://arxiv.org/abs/hep-th/0405231)
36. K. Aamodt et al., ALICE Collaboration, Harmonic decomposition of two-particle angular correlations in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Lett.* **B708**, 249–264 (2012). [arXiv:1109.2501](https://arxiv.org/abs/1109.2501)
37. B.B. Abelev et al., ALICE Collaboration, Elliptic flow of identified hadrons in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. [arXiv:1405.4632](https://arxiv.org/abs/1405.4632)
38. H. Song, S. Bass, U.W. Heinz, Spectra and elliptic flow for identified hadrons in 2.76 A TeV Pb+Pb collisions. *Phys. Rev.* **C89**, 034919 (2014). [arXiv:1311.0157](https://arxiv.org/abs/1311.0157)
39. J. Beringer et al., Review of particle physics. *Phys. Rev.* **D86**, 010001 (2012)
40. D. d’Enterria, Jet quenching. [arXiv:0902.2011](https://arxiv.org/abs/0902.2011)
41. S. Chatrchyan et al., CMS Collaboration, Study of high- p_T charged particle suppression in PbPb compared to pp collisions at $\sqrt{s_{NN}} = 2.76$ TeV, *Eur. Phys. J.* **C72**, 1945 (2012). [arXiv:1202.2554](https://arxiv.org/abs/1202.2554)

42. S. Chatrchyan et al., CMS Collaboration, Measurement of isolated photon production in pp and PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Lett. **B710**, 256–277 (2012). [arXiv:1201.3093](#)
43. S. Chatrchyan et al., CMS Collaboration, Study of W boson production in PbPb and pp collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Lett. **B715**, 66–87 (2012). [arXiv:1205.6334](#)
44. S. Chatrchyan et al., CMS Collaboration, Study of Z boson production in PbPb collisions at nucleon-nucleon centre of mass energy = 2.76 TeV. Phys. Rev. Lett. **106**, 212301 (2011). [arXiv:1102.5435](#)
45. B. Abelev et al., ALICE Collaboration, Transverse momentum distribution and nuclear modification factor of charged particles in p -Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. Phys. Rev. Lett. **110**, 082302 (2013). [arXiv:1210.4520](#)
46. K. Aamodt et al., ALICE Collaboration, Particle-yield modification in jet-like azimuthal dihadron correlations in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Rev. Lett. **108**, 092301 (2012). [arXiv:1110.0121](#)
47. K.M. Burke, A. Buzzatti, N. Chang, C. Gale, M. Gyulassy et al., Extracting jet transport coefficient from jet quenching at RHIC and LHC. Phys. Rev. **C90**, 014909 (2014). [arXiv:1312.5003](#)
48. D. Britzger et al., fast NLO, 2014
49. M. Cacciari, G.P. Salam, G. Soyez, FastJet user manual. Eur. Phys. J. **C72**, 1896 (2012). [arXiv:1111.6097](#)
50. B. Abelev et al., ALICE Collaboration, Measurement of charged jet suppression in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. JHEP **1403**, 013 (2014). [arXiv:1311.0633](#)
51. G. Aad et al., ATLAS Collaboration, Measurement of the jet radius and transverse momentum dependence of inclusive jet suppression in lead-lead collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS detector. Phys. Lett. **B719**, 220–241 (2013). [arXiv:1208.1967](#)
52. B. Abelev et al., ALICE Collaboration, Centrality dependence of charged particle production at large transverse momentum in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Lett. **B720**, 52–62 (2013). [arXiv:1208.2711](#)
53. G. Aad et al., ATLAS Collaboration, Measurement of the jet fragmentation function and transverse profile in proton-proton collisions at a center-of-mass energy of 7 TeV with the ATLAS detector. Eur. Phys. J. **C71**, 1795 (2011). [arXiv:1109.5816](#)
54. S. Chatrchyan et al., CMS Collaboration, Measurement of jet fragmentation into charged particles in pp and PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. JHEP **1210**, 087 (2012). [arXiv:1205.5872](#)
55. S. Chatrchyan et al., CMS Collaboration, Modification of jet shapes in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Lett. **B730**, 243–263 (2014). [arXiv:1310.0878](#)
56. CMS Collaboration, Measurement of transverse momentum flow relative to the dijet system in PbPb and pp collisions at $\sqrt{s_{NN}} = 2.76$ TeV. CMS-PAS-HIN-14-010 (2014)
57. S. Chatrchyan et al., CMS Collaboration, Studies of jet quenching using isolated-photon+jet correlations in PbPb and pp collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Lett. **B718**, 773–794 (2013). [arXiv:1205.0206](#)
58. S. Chatrchyan et al., CMS Collaboration, Suppression of non-prompt J/ψ , prompt J/ψ , and $\Upsilon(1S)$ in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. JHEP **1205**, 063 (2012). [arXiv:1201.5069](#)
59. B. Abelev et al., ALICE Collaboration, Suppression of high transverse momentum D mesons in central Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. JHEP **1209**, 112 (2012). [arXiv:1203.2160](#)
60. B. Abelev et al., ALICE Collaboration, Anisotropic flow of charged hadrons, pions and (anti-)protons measured at high transverse momentum in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Lett. **B719**, 18–28 (2013). [arXiv:1205.5761](#)
61. B. Abelev et al., ALICE Collaboration, D meson elliptic flow in non-central Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Rev. Lett. **111**, 102301 (2013). [arXiv:1305.2707](#)
62. T. Matsui, H. Satz, J/ψ suppression by quark-gluon plasma formation. Phys. Lett. B **178**, 416–422 (1986)
63. Á. Mócsy, P. Petreczky, M. Strickland, Quarkonia in the quark gluon plasma. Int. J. Mod. Phys. A **28**(11), 1340012 (2013). <http://www.worldscientific.com/doi/pdf/10.1142/S0217751X13400125>
64. R. Granier de Cassagnac, Quarkonium suppression from SPS to RHIC (and from p+A to A+A). Nucl. Phys. **A783**, 293–300 (2007). [arXiv:nucl-ex/0610003](#)

65. B. Abelev et al., ALICE Collaboration, J/ψ suppression at forward rapidity in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Rev. Lett. **109**, 072301 (2012). [arXiv:1202.1383](#)
66. B. B. Abelev et al., ALICE Collaboration, Centrality, rapidity and transverse momentum dependence of J/Ψ suppression in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Lett. **743**, 314–327 (2014). [arXiv:1311.0214](#)
67. S. Chatrchyan et al., CMS Collaboration, Indications of suppression of excited Υ states in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Rev. Lett. **107**, 052302 (2011). [arXiv:1105.4894](#)
68. S. Chatrchyan et al., CMS Collaboration, Observation of sequential Upsilon suppression in PbPb collisions. Phys. Rev. Lett. **109**, 222301 (2012). [arXiv:1208.2826](#)
69. A. Emerick, X. Zhao, R. Rapp, Bottomonia in the quark-gluon plasma and their production at RHIC and LHC. Eur. Phys. J. **A48**, 72 (2012). [arXiv:1111.6537](#)
70. Y. Liu, N. Xu, P. Zhuang, J/ψ elliptic flow in relativistic heavy ion collisions. Nucl. Phys. **A834**, 317C–319C (2010). [arXiv:0910.0959](#)
71. E. Abbas et al., ALICE Collaboration, J/ψ Elliptic Flow in Pb-Pb Collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Rev. Lett. **111**, 162301 (2013). [arXiv:1303.5880](#)
72. B.B. Abelev et al., ALICE Collaboration, J/ψ production and nuclear effects in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. JHEP **1402**, 073 (2014). [arXiv:1308.6726](#)
73. B.B. Abelev et al., ALICE Collaboration, Suppression of $\psi(2S)$ production in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. [arXiv:1405.3796](#)

Flavour Physics and CP Violation in the Standard Model and Beyond

Gustavo Castelo-Branco and David Emmanuel-Costa

Abstract We present the invited lectures given at the Third IDPASC School which took place in Santiago de Compostela in January 2013. The students attending the school had very different backgrounds, some of them were doing their PhD in experimental particle physics, others in theory. As a result, and in order to make the lectures useful for most of the students, we focused on basic topics of broad interest, avoiding the more technical aspects of Flavour Physics and CP Violation. We make a brief review of the Standard Model, paying special attention to the generation of fermion masses and mixing, as well as to CP violation. We describe some of the simplest extensions of the SM, emphasising novel flavour aspects which arise in their framework.

1 Review of the Standard Model

The Standard Model (SM) of unification of the electroweak and strong interactions [1–4] is based on the gauge group

$$G_{\text{SM}} \equiv \mathbf{SU}(3)_C \times \mathbf{SU}(2)_L \times \mathbf{U}(1)_Y, \quad (1)$$

which has 12 generators. To each one of these generators corresponds a gauge field. The introduction of these gauge fields is essential in order to achieve invariance under local gauge transformations of G_{SM} . This is entirely analogous to what one encounters in electromagnetic interactions, where the photon is the gauge field associated to the $\mathbf{U}(1)_{\text{e.m.}}$, introduced in order to guarantee local gauge invariance. We shall

G. Castelo-Branco (✉)
CFTP, Instituto Superior Técnico, Universidade de Lisboa,
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
e-mail: gbranco@tecnico.ulisboa.pt

D. Emmanuel-Costa
IST-ID and CFTP, Instituto Superior Técnico, Universidade de Lisboa,
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
e-mail: david.costa@tecnico.ulisboa.pt

Table 1 The SM fermionic content

$q_{iL} \equiv \begin{pmatrix} u_i \\ d_i \end{pmatrix}_L$	(3, 2, 1/6)
u_{iR}	(3, 1, 2/3)
d_{iR}	(3, 1, -1/3)
$\ell_{iL} \equiv \begin{pmatrix} v_i \\ e_i^- \end{pmatrix}_L$	(1, 2, -1/2)
e_{iR}^-	(1, 1, -1)

For a given SM representation R one has $(n_3, n_2, y) \equiv (\dim_{\text{SU}(3)}(R), \dim_{\text{SU}(2)}(R), Y(R))$. The index $i = 1, 2, 3$ is the generation index

denote the gauge fields in the following way:

$$\text{SU}(3)_C \longrightarrow G_\mu^k, \quad k = 1, \dots, 8; \quad (2)$$

$$\text{SU}(2)_L \longrightarrow W_\mu^j, \quad j = 1, \dots, 3; \quad (3)$$

$$\text{SU}(2)_L \longrightarrow B_\mu. \quad (4)$$

The electroweak interactions are linear combination of the following gauge bosons:

$$W_\mu^a, B_\mu \longrightarrow W_\mu^+, W_\mu^-, Z_\mu, A_\mu, \quad (5)$$

where A_μ is the photon field, mediator of electromagnetic interactions while the massive bosons W_μ^+ and Z_μ mediate, respectively, the charged and neutral weak currents. Since $\text{U}(1)_{\text{e.m.}}$ is a good symmetry of nature, the photon field should remain massless.

The SM describes all observed fermionic particles, which have definite gauge transformations properties and are replicated in three generations. All the SM fermionic fields carry weak hypercharge Y , defined as

$$Y \equiv Q - T_3, \quad (6)$$

where Q is the electric charge operator and T_3 is the diagonal generator of $\text{SU}(2)_L$. Since experiments only provided evidence for left-handed charged currents, the right-handed components of fermion fields are put in $\text{SU}(2)_L$ -singlets. Only the quarks carry colour, i.e. they are triplets of $\text{SU}(3)_C$, while the leptons carry no colour. We summarise in Table 1 all fermionic content characterised by their transformation properties under the gauge group $\text{SU}(3)_C \times \text{SU}(2)_L \times \text{U}(1)_Y$. It is worth noting that within this matter content the SM is free from anomalies, since $\text{SU}(3)_C$ is non-chiral, all representations of $\times \text{SU}(2)_L$ are real, the $\text{SU}(3)^2 Y$, $\text{SU}(2)^2 Y$ and Y^3 cancel between the quarks and leptons.

Gauge interactions are determined by the covariant derivative which is dictated by the transformation properties of the various fields, under the gauge group. In general one has

$$D_\mu = \partial_\mu - ig_s L^k G_\mu^k - ig T^j W_\mu^j - ig' y B_\mu, \quad (7)$$

where T^j are the three $\mathbf{SU}(2)$ -generators,

$$T^j = \begin{cases} 0, & \text{singlet} \\ \frac{\tau_j}{2}, & \text{fundamental} \end{cases}, \quad (8)$$

while L^k are the eight $\mathbf{SU}(3)$ -generators,

$$L^k = \begin{cases} 0, & \text{singlet} \\ \frac{\lambda_k}{2}, & \text{fundamental} \end{cases}. \quad (9)$$

The matrices τ_j and λ_k are the usual Pauli and Gell-Mann matrices, respectively. For the fermions presented in Table 1 the covariant derivatives read as

$$D_\mu q_L = \left(\partial_\mu - i \frac{g_s}{2} \lambda_k G_\mu^k - i \frac{g}{2} \tau_j W_\mu^j - i \frac{g'}{6} B_\mu \right) q_L, \quad (10)$$

$$D_\mu u_R = \left(\partial_\mu - i \frac{g_s}{2} \lambda_k G_\mu^k - i \frac{2g'}{3} B_\mu \right) u_R, \quad (11)$$

$$D_\mu d_R = \left(\partial_\mu - i \frac{g_s}{2} \lambda_k G_\mu^k + i \frac{g'}{3} B_\mu \right) d_R, \quad (12)$$

$$D_\mu \ell_L = \left(\partial_\mu - i \frac{g}{2} \tau_j W_\mu^j + i \frac{g'}{2} B_\mu \right) q_L, \quad (13)$$

$$D_\mu e_L^- = (\partial_\mu + ig B_\mu) e_R^-. \quad (14)$$

An important feature of the SM is the fact that right-handed neutrinos,

$$\nu_R \sim (1, 1, 0), \quad (15)$$

are not introduced. As a result, neutrinos are strictly massless in the SM, in contradiction with present experimental evidence. We shall come back to this question in the sequel.

In order to account for the massive gauge bosons W_μ^\pm and Z_μ without destroying renormalisability, the gauge symmetry must be spontaneously broken. The simplest scheme to break spontaneously the electroweak gauge symmetry into electromagnetism, involves the introduction of a complex doublet Higgs scalar field ϕ

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \sim (1, 2, 1/2), \quad (16)$$

which leads to the breaking:

$$\mathbf{SU}(3)_C \times \mathbf{SU}(2)_L \times \mathbf{U}(1)_Y \longrightarrow \mathbf{SU}(3)_C \times \mathbf{U}(1)_{\text{e.m.}}. \quad (17)$$

The most general gauge invariant, renormalisable scalar potential is:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (18)$$

Taking $\lambda > 0$ the potential is bounded from below and two minima do exist. For $\mu^2 > 0$ one has $\langle 0 | \phi | 0 \rangle = 0$ while for $\mu^2 < 0$ one has instead

$$\langle 0 | \phi | 0 \rangle = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} v \end{pmatrix}. \quad (19)$$

In Fig. 1 it is drawn the Higgs potential around the two minima. Indeed, the case $\lambda > 0$ and $\mu^2 < 0$ implies the spontaneous breaking of the electroweak gauge as indicated in 17. One can check that the U(1) remains unbroken. The electric charge operator reads as

$$Q = T_3 + Y, \quad (20)$$

and for the Higgs doublet one gets

$$Q = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} + \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \quad (21)$$

Therefore one verifies that the vacuum given in (19) is invariant under the charge operator Q , since

$$Q \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} v \end{pmatrix} = 0, \quad (22)$$

and one gets

$$e^{i \alpha Q} \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} v \end{pmatrix} = \left[1 + i \alpha Q + \dots \right] \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} v \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} v \end{pmatrix}. \quad (23)$$

Electric charge is automatically conserved in the SM. This is no longer true in extensions of the SM with two Higgs doublets, including the case of supersymmetric extensions of the SM. In the general two Higgs doublet model (2HDM) without loss of generality, one has:

$$\langle 0 | \phi_1 | 0 \rangle = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} v_1 \end{pmatrix}, \quad \langle 0 | \phi_2 | 0 \rangle = \begin{pmatrix} \xi \\ \frac{1}{\sqrt{2}} v_2 e^{i \theta} \end{pmatrix}, \quad (24)$$

with ξ real. In order to preserve charge conservation in the 2HDM, one has to choose a region of the parameter space where the minimum is at $\xi = 0$.

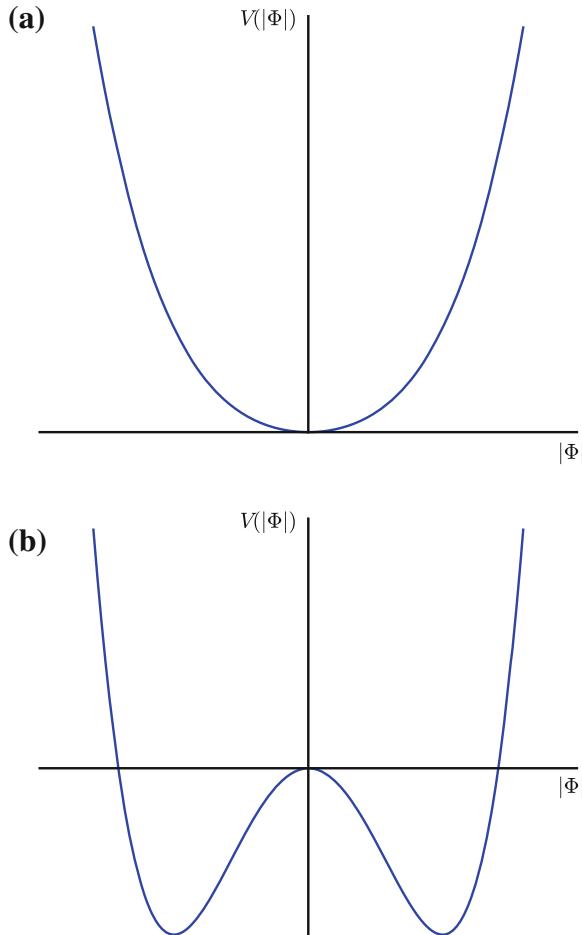


Fig. 1 **a** $\lambda > 0, \mu^2 > 0$, **b** $\lambda > 0, \mu^2 < 0$

The SM does not provide an explanation for the charges of elementary fermions. The values of the hypercharge Y are chosen in such a way that the correct electric charges are obtained. As an example, one can determine Y_{q_L} , by using the (6) and the knowledge of Q_u and Q_d . Thus,

$$Y_{u_L} = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}, \quad (25)$$

$$Y_{d_L} = -\frac{1}{3} + \frac{1}{2} = \frac{1}{6}, \quad (26)$$

and therefore $Y_{q_L} = 1/6$. It is rather intriguing the fact that the requirement of cancellation of the gauge anomaly in the SM together with the fact that the electromagnetic

interactions are non-chiral is sufficient to fully determine all the hypercharges of the fundamental fermions up to an overall factor. In particular one gets relations among quark and lepton charges, leading to:

$$Q_p = -Q_e. \quad (27)$$

Although the hypercharge quantisation may arise from the anomaly-free condition, this is certainly not a satisfactory explanation in the SM. The solution to this fundamental question is elegantly answered in the framework of Grand-Unification, e.g. $SU(5)$, where the quantisation of electric charges is related to some new phenomena like the magnetic monopoles predicted in the theory that can be tested in future experiments.

In order to describe the spontaneous breaking of the electroweak symmetry in the SM, one starts by introducing a convenient parametrisation of the Higgs doublet ϕ as

$$\phi = \begin{pmatrix} G^+ \\ \frac{1}{\sqrt{2}}(v + H + i G_0) \end{pmatrix}, \quad (28)$$

where G^+ is a charged complex scalar field, H is a real scalar field and G_0 is a real pseudo-scalar. The scalar fields G^\pm and G_0 are massless states, the so-called Nambu-Goldstone bosons. Through the Brout-Englert-Higgs mechanism, the charged bosons G^\pm are absorbed as longitudinal components of the W_μ^\pm which acquire a mass:

$$M_W = \frac{g v}{2}, \quad (29)$$

while the neutral boson G_0 becomes the longitudinal component of the gauge boson Z_μ , which is a linear combination of the bosons B_μ and W_μ^3 ,

$$Z_\mu = \cos \theta_W W_\mu^3 - \sin \theta_W B_\mu, \quad (30)$$

where θ_W is simply given by

$$\tan \theta_W \equiv \frac{g'}{g}. \quad (31)$$

The Z_μ boson acquires then a mass given by

$$M_Z = \sqrt{g^2 + g'^2} \frac{v}{2} = \frac{M_W}{\cos \theta_W}. \quad (32)$$

The bosonic state orthogonal bosonic state to Z_μ :

$$A_\mu = \cos \theta_W B_\mu + \sin \theta_W W_\mu^3, \quad (33)$$

remains massless and is identified with the photon. The electron coupling to the photon is directly determined from the weak couplings g and g' as

$$\frac{1}{e^2} = \frac{1}{g^2} + \frac{1}{g'^2}, \quad (34)$$

or

$$e = \frac{g g'}{\sqrt{g^2 + g'^2}} = g \sin \theta_W = g' \cos \theta_W. \quad (35)$$

2 Fermion Masses and Mixings

In the SM, one cannot write directly a mass term for any of the fundamental fermions because they would violate the gauge symmetry, since left-handed and right-handed chiralities do transform differently. The SM fermions acquire mass through Yukawa couplings, once the SM group is spontaneously broken. Therefore, in the SM the Higgs mechanism that is responsible for the breaking of the SM group, also generates fermion masses.

2.1 Quark and Charged Lepton Masses

The Yukawa interactions are the most general terms in the Lagrangian allowed by the SM gauge group that involve fermions and the Higgs doublet. The Yukawa couplings can be written as:

$$-\mathcal{L}_Y = (Y_u)_{ij} \bar{q}_i L \tilde{\phi} u_i R + (Y_d)_{ij} \bar{q}_i L \phi d_i R + (Y_\ell)_{ij} \bar{\ell}_i L \phi e_i R + \text{H.c.}, \quad (36)$$

where $\tilde{\phi} \equiv i\tau_2 \phi^\dagger$. The Yukawa matrices Y_u , Y_d and Y_ℓ are arbitrary complex matrices in flavour space. The first two terms in (36) will generate the up- and down-type quark masses while the third term will give rise to the charged lepton masses. Making use of the Higgs doublet parametrisation given in (28) one can decompose the Lagrangian given in (36) as

$$\begin{aligned} -\mathcal{L}_Y = & \frac{v}{\sqrt{2}} (Y_u)_{ij} \bar{u}_i L u_i R + \frac{v}{\sqrt{2}} (Y_d)_{ij} \bar{d}_i L d_i R + \frac{v}{\sqrt{2}} (Y_\ell)_{ij} \bar{e}_i L e_i R \quad (37) \\ & + \frac{(Y_u)_{ij}}{\sqrt{2}} \bar{u}_i L u_i R H + \frac{(Y_d)_{ij}}{\sqrt{2}} \bar{d}_i L d_i R H + \frac{(Y_\ell)_{ij}}{\sqrt{2}} \bar{e}_i L e_i R H \\ & - \frac{i(Y_u)_{ij}}{\sqrt{2}} \bar{u}_i L u_i R G^0 + \frac{i(Y_d)_{ij}}{\sqrt{2}} \bar{d}_i L d_i R G^0 + \frac{i(Y_\ell)_{ij}}{\sqrt{2}} \bar{e}_i L e_i R G^0 \\ & - (Y_u)_{ij} \bar{d}_i L u_i R G^- + (Y_d)_{ij} \bar{u}_i L d_i R, G^+ + (Y_\ell)_{ij} \bar{v}_i L e_i R G^+ + \text{H.c..} \end{aligned}$$

Once a gauge transformation is performed in order to absorbed the Nambu-Goldstone bosons G^\pm and G^0 , the Lagrangian in (37) becomes

$$\begin{aligned} -\mathcal{L}_Y = & (m_u)_{ij} \bar{u}_{iL} u_{iR} + (m_d)_{ij} \bar{d}_{iL} d_{iR} + (m_\ell)_{ij} \bar{e}_{iL} e_{iR} \\ & + \frac{(Y_u)_{ij}}{\sqrt{2}} \bar{u}_{iL} u_{iR} H + \frac{(Y_d)_{ij}}{\sqrt{2}} \bar{d}_{iL} d_{iR} H + \frac{(Y_\ell)_{ij}}{\sqrt{2}} \bar{e}_{iL} e_{iR} H + \text{H.c.}, \end{aligned} \quad (38)$$

where the quark mass matrices m_u , m_d and the charged lepton mass matrix m_ℓ are simply defined by

$$m_u \equiv \frac{v}{\sqrt{2}} Y_u, \quad m_d \equiv \frac{v}{\sqrt{2}} Y_d, \quad m_\ell \equiv \frac{v}{\sqrt{2}} Y_\ell. \quad (39)$$

Gauge invariance does not constrain the flavour structure of Yukawa couplings and therefore m_u , m_d and m_ℓ are arbitrary complex matrices.

Let us now focus on the mass terms,

$$-\mathcal{L}_m = (m_u)_{ij} \bar{u}_{iL}^0 u_{iR}^0 + (m_d)_{ij} \bar{d}_{iL}^0 d_{iR}^0 + (m_\ell)_{ij} \bar{e}_{iL}^0 e_{iR}^0. \quad (40)$$

A super-script 0 on the fermion fields was used that these fields are the original ones, in the weak basis. The matrices $m_{u,d,e}$ can be diagonalised by the following bi-unitary transformations:

$$u_{iL}^0 = U_L^u u_{iL}; \quad u_{iR}^0 = U_R^u u_{iR}, \quad (41a)$$

$$d_{iL}^0 = U_L^d d_{iL}; \quad d_{iR}^0 = U_R^d d_{iR}, \quad (41b)$$

$$e_{iL}^0 = U_L^e e_{iL}; \quad e_{iR}^0 = U_R^e e_{iR}, \quad (41c)$$

where $U_{R,L}^{u,d,e}$ are a set of unitary matrix such as

$$m_u \longrightarrow U_L^{u\dagger} m_u U_R^u = \text{diag}(m_u, m_c, m_t), \quad (42a)$$

$$m_d \longrightarrow U_L^{d\dagger} m_d U_R^d = \text{diag}(m_d, m_s, m_b), \quad (42b)$$

$$m_\ell \longrightarrow U_L^{e\dagger} m_\ell U_R^e = \text{diag}(m_\ell, m_\mu, m_\tau). \quad (42c)$$

The fields $u_{L,R}$, $d_{L,R}$, $e_{L,R}$ are thus the mass eigenstates. The bi-unitary transformations given in (41a)–(41c) affect the interactions between left-handed particles and the W_μ^\pm bosons—the *charged currents*—which are written in a weak basis as:

$$-\mathcal{L}_{CC} = \frac{g}{\sqrt{2}} \left[\bar{u}_{iL}^0 \gamma^\mu d_{iL}^0 + \bar{v}_{iL}^0 \gamma^\mu e_{iL}^0 \right] W_\mu^+ + \text{H.c..} \quad (43)$$

In the mass eigenstate basis the charged currents become:

$$-\mathcal{L}_{\text{CC}} = \frac{g}{\sqrt{2}} \left[\bar{u}_L \gamma^\mu U_L^{u\dagger} U_L^d d_L + \bar{v}_L^0 \gamma^\mu U_L^e e_L \right] W_\mu^+ + \text{H.c..} \quad (44)$$

The product of unitary matrices in (44) defines the well known Cabibbo-Kobayashi-Maskawa matrix V as

$$V \equiv U_L^{u\dagger} U_L^d. \quad (45)$$

In the SM the unitary matrix U_L^e is physically meaningless. Note that since neutrinos are massless in the SM, one can always redefine neutrino fields as

$$\nu_L^0 \longrightarrow \nu_L = U_L^e \nu_L, \quad (46)$$

and therefore the charged current term $\bar{v}_L^0 \gamma^\mu U_L^e e_L$ in (44) becomes $\bar{v}_L \gamma^\mu e_L$. We then conclude that in the SM there is no leptonic mixing and therefore no neutrino oscillations.

We can show that the electromagnetic and neutral currents are not affected by the transformations given in (41a)–(41c). The electromagnetic $J_{\text{e.m.}}$ given in the weak basis,

$$\begin{aligned} J_{\text{e.m.}}^\mu &= \frac{2}{3} \left[\bar{u}_L^0 \gamma^\mu u_L^0 + \bar{u}_R^0 \gamma^\mu u_R^0 \right] \\ &\quad - \frac{1}{3} \left[\bar{d}_L^0 \gamma^\mu d_L^0 + \bar{d}_R^0 \gamma^\mu d_R^0 \right] \\ &\quad - \left[\bar{e}_L^0 \gamma^\mu e_L^0 + \bar{e}_R^0 \gamma^\mu e_R^0 \right], \end{aligned} \quad (47)$$

do not change in the mass eigenstate, since $J_{\text{e.m.}}^\mu$ transforms as

$$\begin{aligned} J'_{\text{e.m.}}^\mu &= \frac{2}{3} \left[\bar{u}_L \gamma^\mu U_L^{u\dagger} U_L^u u_L + \bar{u}_R \gamma^\mu U_R^{u\dagger} U_R^u u_R \right] \\ &\quad - \frac{1}{3} \left[\bar{d}_L \gamma^\mu U_L^{d\dagger} U_L^d d_L + \bar{d}_R \gamma^\mu U_R^{d\dagger} U_R^d d_R \right] \\ &\quad - \left[\bar{e}_L \gamma^\mu U_L^{e\dagger} U_L^e e_L + \bar{e}_R \gamma^\mu U_R^{e\dagger} U_R^e e_R \right], \end{aligned} \quad (48)$$

and we get the same formal expression as in (47). In a similar way we demonstrate that the neutral currents Lagrangian,

$$\begin{aligned} \mathcal{L}_{\text{NC}} &= \frac{g}{\cos \theta_W} \left[\bar{u}_L^0 \gamma^\mu u_L^0 - \bar{d}_L^0 \gamma^\mu d_L^0 + \bar{v}_L^0 \gamma^\mu v_L^0 - \bar{e}_L \gamma^\mu e_L^0 \right. \\ &\quad \left. - 2 \sin^2 \theta_W J_{\text{e.m.}}^\mu \right] Z_\mu, \end{aligned} \quad (49)$$

are also invariant under the transformations given in (41a)–(41c).

$$\mathcal{L}'_{\text{NC}} = \frac{g}{\cos \theta_W} \left[\bar{u}_L \gamma^\mu u_L - \bar{d}_L \gamma^\mu d_L + \bar{v}_L \gamma^\mu v_L - \bar{e}_L \gamma^\mu e_L - 2 \sin^2 \theta_W J_{\text{e.m.}}^\mu \right] Z_\mu . \quad (50)$$

Flavour changing neutral currents (FCNC) are naturally absent at three-level in the SM, due to the GIM mechanism. Indeed “charm” was invented in order to achieve this cancellation of FCNC.

Exercise 1. Suppose that “charm” did not exist, so that one would have

$$\begin{pmatrix} u_1^0 \\ d_1^0 \end{pmatrix}_L , \quad d_{2L} , \quad u_{1R} , \quad d_{1R} , \quad d_{1R} . \quad (51)$$

Show that in this model FCNC automatically arise.

Historical note: Prior to the appearance of renormalisable gauge interactions, physicists considered the possibility that weak neutral currents could exist. However there was a *strong prejudice* against neutral currents due to the stringent experimental limits on the strength of FCNC.

Example 1 The decay $K_L^0 \rightarrow \mu^+ \mu^-$ has a branching ratio *extremely suppressed*, with respect to the decay $K_L^0 \rightarrow \pi^+ e^- \bar{\nu}_e$. If FCNC existed they would have branching ratios of the same order of magnitude which are shown in Fig. 2.

From (50) we see that neutral current interactions violate parity, since both couplings involving $\bar{\psi} \gamma_\mu \psi$ and $\bar{\psi} \gamma_\mu \gamma_5 \psi$ are present.

As a result of the GIM mechanism there are no tree-level contributions to $K^0 - \overline{K^0}$, $B^0 - \overline{B^0}$, $B_S - \overline{B_S}$ and $D^0 - \overline{D^0}$ mixings. However in the SM there are higher order contributions to these processes which are calculable. The contributions from the diagrams given in Fig. 3 led to the correct estimate to the charm quark mass [5] and the size of $B_d - \overline{B_d}$ mixing provided the first indirect evidence of a large top mass.

Exercise 2. Consider a simple extension of the SM which consists of the addition of an isosinglet quark D ,

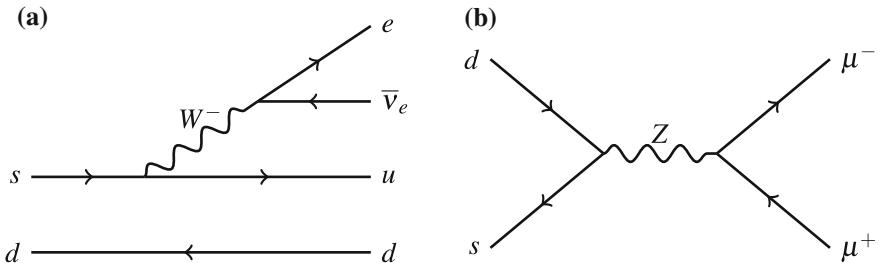
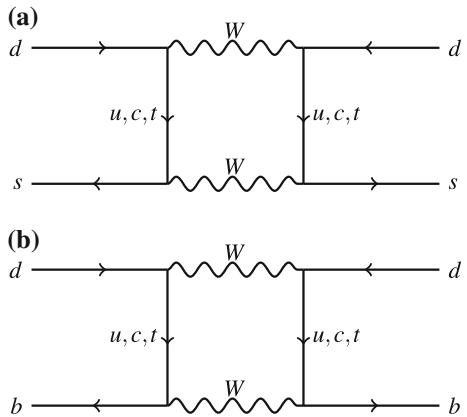


Fig. 2 **a** $K_L^0 \rightarrow \pi^+ e^- \bar{\nu}_e$, **b** does not exist at tree-level in SM

Fig. 3 a $K^0 - \overline{K^0}$ mixing, **b** $B_d^0 - \overline{B_d^0}$ mixing



$$D_L, D_R \sim (3, 1, -1/3). \quad (52)$$

- (a) Write down the most general quark mass terms which are obtained in the framework of this model.
- (b) Derive the structure of the charged currents.
- (c) Derive the structure of neutral currents, showing that there are FCNC in this model.
- (d) Show that although non-vanishing at tree level, FCNC are naturally suppressed in this model, provided the isosinglet quark D is much heavier than the standard quarks.

Neutral currents have played a crucial rôle in the construction of the SM and its experimental tests and the discovery of Neutral weak currents was the first great success of the SM. As it was here described, the important feature of FCNC is that they are forbidden at tree-level, both in the SM and in most of its extensions. At loop level FCNC are generated and have played a crucial rôle in testing the SM and in putting bounds on New Physics beyond the SM through the study of process like: $K^0 - \overline{K^0}$, $B^0 - \overline{B^0}$, $B_S - \overline{B_S}$ and $D^0 - \overline{D^0}$; rare kaon decays; rare b-meson decays; CP violation. In this framework, SM contributes to these processes at loop level and therefore New Physics has a chance to give significant contributions. On the other hand, the need to suppress FCNC has lead to two dogmas:

no Z-mediated FCNC at tree level and no FCNC in the scalar sector, at tree level.

Glashow and Weinberg [6] and Paschos [7] derived necessary and sufficient conditions for having diagonal neutral currents, namely:

- (i) All quarks of fixed charge and helicity must transform according to the same irreducible representation of $SU(2)$ and correspond to the same eigenvalue of T_3 .
- (ii) All quarks should receive their contributions to the quark mass matrix from a single neutral scalar VEV.

**Can one violate the above two dogmas in reasonable extensions of the SM?
The answer is yes!**

“Reasonable” means that FCNC should be naturally suppressed without fine-tuning. In the gauge sector, the *dogma* can be violated through the introduction of a $Q = 1/3$ and/or $Q = 2/3$ vector-like quark [8–14], since in this model one has naturally small violation of 3×3 unitarity of the CKM matrix V which in turn leads to Z -mediated FCNC at tree level, which are naturally suppressed.

In the Higgs sector, the dogma can be violated and yet having FCNC automatically suppressed by small CKM matrix elements [8].

2.2 Fundamental Properties of the CKM Matrix

We have introduced in (45) the CKM matrix V , which characterises the flavour changing charged currents in the quark sector:

$$\mathcal{L}_{\text{CC}} = (\bar{u} \bar{c} \bar{t})_L \gamma^\mu \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L W_\mu^+ + \text{H.c.}, \quad (53)$$

The CKM matrix is complex, but some of its phases have no physical meaning. This is due to the fact that one has the freedom to rephase the mass eigenstate quark fields u_α, d_k :

$$u_\alpha = e^{i\varphi_\alpha} u'_\alpha, \quad d_k = e^{i\varphi_k} d'_k. \quad (54)$$

Under this rephasing one has:

$$V'_{\alpha k} = e^{i(\varphi_k - \varphi_\alpha)} V_{\alpha k}. \quad (55)$$

It is clear from (55) that the individual phases of V_{ij} have no Physical meaning. It is useful to look for rephasing invariant quantities, which do not change under this rephasing. The simplest examples are moduli $|V_{\alpha k}|$ and quartets $Q_{\alpha i \beta j}$, defined as

$$Q_{\alpha i \beta j} \equiv V_{\alpha i} V_{\beta j} V_{\alpha j}^* V_{\beta i}^*, \quad (56)$$

with $\alpha \neq \beta$ and $i \neq j$. Invariants of higher order may in general be written as functions of the quartets and the moduli.

Exercise 3. Show that:

$$V_{\alpha i} V_{\beta j} V_{\gamma k} V_{\alpha j}^* V_{\beta k}^* V_{\gamma i}^* = \frac{Q_{\alpha i \beta j} Q_{\beta i \alpha j}}{|V_{\beta i}|^2}. \quad (57)$$

The quartets are easily constructed through the following scheme,

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \quad (58)$$

where the two quartets,

$$V_{us} V_{cb} V_{ub}^* V_{cs}^* = Q_{uscb}, \quad V_{cd} V_{ts} V_{td}^* V_{cs}^* = Q_{cdts}, \quad (59)$$

are illustrated. The diagonal dotted line refers to the product of the corresponding CKM elements.

2.3 Neutrino Masses

In the SM, neutrinos are exactly massless. No Dirac mass terms can be written since right-handed neutrino fields are not introduced in the SM. On the other hand, Majorana mass terms are not generated in higher orders, due to exact $(B - L)$ conservation in the SM. As a result of having massless neutrinos, neither leptonic mixing nor leptonic CP violation can be generated in the SM. Indeed, any mixing arising from the diagonalisation of the charged-lepton masses can be rotated away by a redefinition of the neutrino fields.

In the view of above, one concludes that the discovery of leptonic mixing and non-vanishing neutrino masses, rules out the SM, as it was proposed. However a simple extension of the SM, sometimes denoted ν SM, can easily accommodate leptonic mixing and provide an explanation for the smallness of neutrino masses, through the seesaw mechanism [15–19]. The nature of neutrinos (i.e. Majorana or Dirac) is still an important open question. Both in the case of Majorana [15–19] or Dirac neutrinos [20] one has to have a mechanism to understand the smallness of neutrino masses.

2.4 The Flavour Sector of the SM

Let us now discuss the flavour sector of the SM. The gauge invariance does not constrain the flavour structure of the Yukawa matrices Y_u , Y_d and using (39) one obtains two arbitrary mass matrices m_u and m_d . The two quark mass matrices are

arbitrary complex matrices which need not to be Hermitian [21]. The two matrices m_u, m_d contain $(18 + 18)$ parameters, but most of them are not physical. Due to the fermion family replication the gauge interaction part of \mathcal{L}_{SM} has a very large flavour symmetry. One can make Weak-basis transformations which change m_u, m_d but do not change the physical content of m_u, m_d . One has then a large redundancy in m_u, m_d . By making a WB transformation such as:

$$u_L^0 = W_L u_L^{0'}; \quad u_R^0 = W_R^u u_R^{0'}, \quad (60\text{a})$$

$$d_L^0 = W_L d_L^{0'}; \quad d_R^0 = W_R^d d_R^{0'}, \quad (60\text{b})$$

the gauge currents remain flavour diagonal but m_u, m_d change as follows:

$$m_u \longrightarrow m_u' = W_L^\dagger m_u W_R^u, \quad (61\text{a})$$

$$m_d \longrightarrow m_d' = W_L^\dagger m_d W_R^d, \quad (61\text{b})$$

but the physical content does not change! Therefore, without loss of generality, one can make a WB transformation so that m_u is diagonal, i.e. $m_u \text{diag}(m_u, m_c, m_t)$ and m_d is Hermitian

$$m_d = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{12}^* & m_{22} & m_{23} \\ m_{13}^* & m_{23}^* & m_{33} \end{pmatrix}. \quad (62)$$

In this basis, the only rephasing invariant phase is

$$\varphi \equiv \arg(m_{12} m_{23} m_{13}^*), \quad (63)$$

and there are ten independent parameters: 3 up-quark masses m_u, m_c, m_t , 6 moduli down-type matrix elements $|m_{dij}|$ and one rephasing invariant phase φ . There is to a difficulty in following a bottom-up approach in the search for a solution to the **Flavour Puzzle**: even if there is a Flavour Symmetry behind the spectrum or fermion masses and mixings, in what Weak-Basis will the symmetry be transparent? For example Texture Zeroes are Weak-Basis dependent [22].

2.5 CP Violation

In order to study the CP properties of a Lagrangian, it is convenient to separate the Lagrangian in two parts:

$$\mathcal{L} = \mathcal{L}_{(\text{CP})} + \mathcal{L}', \quad (64)$$

where $\mathcal{L}_{(CP)}$ denotes the part of the Lagrangian which one knows that conserves CP. At this stage it is important to recall that a pure gauge Lagrangian is necessarily CP invariant [23]. One should allow for *the most general CP transformations* allowed by \mathcal{L}_{CP} . Typically, \mathcal{L}_{CP} leaves a large freedom of choice in the definition of CP transformations. CP is violated if and only if there is no possible choice of CP transformation which leaves the Lagrangian invariant [24]. CP can be investigated in the fermion mass eigenstate or in a weak basis. We shall consider both cases. Let us study the CP properties of the SM, after spontaneous gauge symmetry breaking, and after diagonalisation of the quark mass matrices, i.e.,

$$m_u = \text{diag}(m_u, m_c, m_t), \quad (65a)$$

$$m_d = \text{diag}(m_d, m_s, m_b), \quad (65b)$$

which are non-degenerate. In the mass eigenstate basis, the most general CP transformation is:

$$\begin{aligned} (\text{CP}) W^{+\mu}(t, \mathbf{r}) (\text{CP})^{-1} &= -e^{i\zeta_W} W^{-\mu}(t, -\mathbf{r}), \\ (\text{CP}) W^{-\mu}(t, \mathbf{r}) (\text{CP})^{-1} &= -e^{-i\zeta_W} W^{+\mu}(t, -\mathbf{r}), \\ (\text{CP}) u_\alpha(t, \mathbf{r}) (\text{CP})^{-1} &= e^{i\zeta_\alpha} \gamma^0 C \bar{u}_\alpha^\top(t, -\mathbf{r}), \\ (\text{CP}) d_k(t, \mathbf{r}) (\text{CP})^{-1} &= e^{i\zeta_k} \gamma^0 C \bar{u}_k^\top(t, -\mathbf{r}), \end{aligned} \quad (66)$$

where the conjugation matrix C obeys to the relation $\gamma_\mu C = -C \gamma_\mu^T$. Invariance of charged current weak interactions under CP constrains $V_{\alpha k}$ to satisfy the following condition:

$$V_{\alpha k}^* = e^{i(\zeta_W + \zeta_k - \zeta_\alpha)} V_{\alpha k}. \quad (67)$$

If one considers a single element of the CKM matrix V , the previous condition can always be satisfied by using the freedom to choose $\zeta_W, \zeta_k, \zeta_\alpha$. However, it can be readily shown that the condition constrains all quartets and all rephasing invariant functions of V to real. Therefore there is CP violation in the SM if and only if any of the rephasing invariant functions of the CKM matrix V is not real.

In the SM with n_g generations, the CKM matrix V is a $n_g \times n_g$ unitary matrix and it can be then parametrised by n_g^2 independent parameters. Through rephasing of quark fields, one can remove $2n_g - 1$ phases. Thus, the total number of parameters, denoted N , is given by

$$N = n_g^2 - 2n_g - 1 = (n_g - 1)^2, \quad (68)$$

which shows that for three generations ($n_g = 3$) one is left with 4 real parameters. If one takes into account that a unitary matrix is described by $n_g(n_g - 1)/2$ “angles”, one can further count the total number of physical phases N_{ph} as:

$$N_{ph} = N - \frac{1}{2}n_g(n_g - 1) = \frac{1}{2}(n_g - 1)(n_g - 2). \quad (69)$$

We conclude that for 2 generations ($n_g = 2$), there are no physical phases left and therefore CP is conserved. In the case of three generations ($n_g = 3$), one has only one CP violating phase. There is another way of confirming this. For two generations, there is only one rephasing invariant quartet Q_{udcs} , defined as

$$Q_{udcs} \equiv V_{ud} V_{cs} V_{us}^* V_{cd}^*. \quad (70)$$

However using the orthogonality relation:

$$V_{ud} V_{cd}^* + V_{us} V_{cs}^* = 0, \quad (71)$$

and multiplying by $V_{us}^* V_{cs}$, one obtains:

$$Q_{udcs} = -|V_{us}|^2 |V_{cs}|^2, \quad (72)$$

which shows that Q_{udcs} is real.

Considering now the case of three generations, we see that orthogonality of the first two rows of V leads to

$$V_{ud} V_{cd}^* + V_{us} V_{cs}^* + V_{ub} V_{cb}^* = 0. \quad (73)$$

Multiplying by $V_{us}^* V_{cs}$ and taking imaginary parts one obtains:

$$iQ_{udcs} = -iQ_{ubcs}. \quad (74)$$

In an analogous way, one can show that for $n_g = 3$ the imaginary parts of all quartets are equal, up to a sign. In the SM with three generations $|iQ|$ gives the strength of CP violation. If we consider the orthogonality between the first and third columns of V :

$$V_{ud} V_{ub}^* + V_{cd} V_{cb}^* + V_{td} V_{tb}^* = 0. \quad (75)$$

This equation may be interpreted as a “triangle” as represented in Fig. 4. One verifies easily that under rephasing, the triangle rotates. Therefore the orientation of the

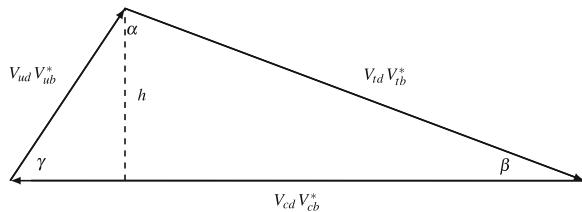


Fig. 4 Unitarity triangle

triangle has no physical meaning. Obviously, the internal angles of the triangles are rephasing invariant, namely

$$\alpha \equiv \arg [-V_{td} V_{ub} V_{ud}^* V_{tb}^*] = \arg(-Q_{ubtd}), \quad (76a)$$

$$\beta \equiv \arg [-V_{cd} V_{tb} V_{cb}^* V_{td}^*] = \arg(-Q_{tbc}), \quad (76b)$$

$$\gamma \equiv \arg [-V_{ud} V_{cb} V_{ub}^* V_{cd}^*] = \arg(-Q_{cbud}), \quad (76c)$$

and one gets the following relation

$$\alpha + \beta + \gamma = \arg(-1) = \pi \pmod{\pi}. \quad (77)$$

This is true “by definition”, and therefore it is not a test of unitarity!!

The quantity iQ has a simple geometrical interpretation. It is twice the area of the unitarity triangles, as sketched in Fig. 4. The area of the triangles, A , is given by

$$A = |V_{cd} V_{cb}^*| \frac{h}{2}, \quad (78)$$

where the height of triangle, h , is given by

$$h = |V_{ud} V_{ub}^*| \sin \gamma, \quad (79)$$

with γ defined in (76c). One then obtains

$$A = \frac{1}{2} |iQ_{udcb}|. \quad (80)$$

Since all $|iQ|$ are equal then all triangles have the same area.

Experimentally we know that:

$$|V_{\text{CKM}}| \simeq \begin{pmatrix} 1 & \lambda & \lambda^3 \\ \lambda & 1 & \lambda^2 \\ \lambda^3 & \lambda^2 & 1 \end{pmatrix}, \quad (81)$$

with $\lambda \approx 0.22$.

The six unitarity triangles are given by

$$\begin{aligned} V_{ud}^{\lambda} V_{us}^* + V_{cd}^{\lambda} V_{cs}^* + V_{td}^{\lambda^5} V_{ts}^* &= 0, & T_{ds} \\ V_{ud}^{\lambda^3} V_{ub}^* + V_{cd}^{\lambda^3} V_{cb}^* + V_{td}^{\lambda^3} V_{tb}^* &= 0, & T_{db} \\ V_{us}^{\lambda^4} V_{ub}^* + V_{cs}^{\lambda^2} V_{cb}^* + V_{ts}^{\lambda^2} V_{tb}^* &= 0, & T_{sb} \\ V_{ud}^{\lambda} V_{cd}^* + V_{cs}^{\lambda} V_{cs}^* + V_{ub}^{\lambda^5} V_{cb}^* &= 0, & T_{uc} \end{aligned} \quad (82)$$

$$\begin{aligned} V_{ud}^{\lambda^3} V_{td}^* + V_{us}^{\lambda^3} V_{ts}^* + V_{ub}^{\lambda^3} V_{td}^* &= 0, \quad T_{ut} \\ V_{cd}^{\lambda^4} V_{td}^* + V_{cs}^{\lambda^2} V_{ts}^* + V_{cb}^{\lambda^2} V_{tb}^* &= 0, \quad T_{ct} \end{aligned}$$

Let us now comment on the strength of CP violation in the SM, which

$$|iQ| = \left| \begin{array}{cccc} \lambda^0 & \lambda^3 & \lambda & \lambda^2 \\ V_{ud} & V_{ub} & V_{cd} & V_{cb} \end{array} \right| \sin \gamma. \quad (83)$$

In order to account for CP violation in the kaon sector, $\sin \gamma$ should be of order 1. So $|iQ| \approx \lambda^6$.

The strength of CP violation (measured by iQ) is small in the SM, due to the smallness of some CKM moduli $|V_{ij}|$, like $|V_{ub}|, |V_{cb}|$. What would be the maximal possible value of iQ ? The maximal value is obtained for the following mixing matrix with universal moduli as

$$V = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^* \\ 1 & \omega^* & \omega \end{pmatrix}, \quad (84)$$

with $\omega \equiv \exp(i2\pi/3)$, yielding

$$iQ = \frac{1}{6\sqrt{3}} \approx 0.096. \quad (85)$$

A convenient parametrisation of the CKM matrix is the so-called Standard Parametrisation, which is defined by the product of three rotations, namely:

$$\begin{aligned} & V(\theta_{12}, \theta_{13}, \theta_{23}, \delta_{13}) \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13} e^{-i\delta_{13}} \\ 0 & 1 & 0 \\ -s_{13} e^{i\delta_{13}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} c_{12} c_{13} & s_{12} c_{13} & s_{13} e^{-i\delta_{13}} \\ -s_{12} c_{23} - c_{12} s_{23} s_{13} e^{i\delta_{13}} & c_{12} c_{23} - s_{12} s_{23} s_{13} e^{i\delta_{13}} & s_{23} c_{13} \\ s_{12} s_{23} - c_{12} c_{23} s_{13} e^{i\delta_{13}} & -c_{12} s_{23} - s_{12} c_{23} s_{13} e^{i\delta_{13}} & c_{23} c_{13} \end{pmatrix}, \end{aligned} \quad (86)$$

where $c_{ij} \equiv \cos \theta_{ij}$ and $s_{ij} \equiv \sin \theta_{ij}$. One of the advantages of the Standard Parametrisation is that the s_{ij} are simply related to directly measured quantities:

$$s_{13} = |V_{ub}|, \quad s_{12} = \frac{|V_{us}|}{\sqrt{1 - |V_{ub}|^2}}, \quad s_{23} = \frac{|V_{cb}|}{\sqrt{1 - |V_{ub}|^2}}. \quad (87)$$

Once s_{ij} are fixed, all data has to be fit by a single parameter: δ_{13} .

2.6 Invariant Approach to CP Violation

In this section we review the invariant approach to CP violation [25]. As previously indicated we write \mathcal{L} as in (64). In order to analyse whether the whole Lagrangian violates CP, one has to check whether the CP transformation under which $\mathcal{L}_{(CP)}$ is invariant implies non-trivial restrictions, i.e. restrictions which may not be satisfied by \mathcal{L}' defined in (64). In the case of the SM, the most general CP transformations which leave $\mathcal{L}_{(CP)}$ invariant are:

$$\begin{aligned} (\text{CP}) u_L^0 (\text{CP})^{-1} &= e^{i\xi_W} K_L \gamma^0 C \overline{u_L^0}^\top, \\ (\text{CP}) d_L^0 (\text{CP})^{-1} &= K_L \gamma^0 C \overline{d_L^0}^\top, \\ (\text{CP}) u_R^0 (\text{CP})^{-1} &= K_R^u \gamma^0 C \overline{u_R^0}^\top, \\ (\text{CP}) d_R^0 (\text{CP})^{-1} &= K_R^d \gamma^0 C \overline{d_R^0}^\top, \end{aligned} \quad (88)$$

where K_L , K_L^u and K_R^d are unitary matrices acting in flavour space. It can be shown that in order for the Yukawa interactions (or equivalently m_u , m_d) to be CP invariant, the following relations have to be satisfied:

$$\begin{aligned} K_L^\dagger m_u K_R^u &= m_u^*, \\ K_L^\dagger m_d K_R^d &= m_d^*. \end{aligned} \quad (89)$$

The existence of the matrices K_L , K_L^u and K_R^d is a necessary and sufficient condition for CP invariance in the SM.

Exercise 4. Prove the above result.

It is rather convenient to define the Hermitian matrices H_u , H_d as:

$$H_u \equiv m_u m_u^\dagger, \quad H_d \equiv m_d m_d^\dagger. \quad (90)$$

Thus, from (89) one derives:

$$\begin{aligned} K_L^\dagger H_u K_L &= H_u^* = H_u^\top, \\ K_L^\dagger H_d K_L &= H_d^* = H_d^\top, \end{aligned} \quad (91)$$

and therefore one has

$$K_L^\dagger [H_u, H_d] K_L = [H_u^\top, H_d^\top] = -[H_u, H_d]^\top. \quad (92)$$

Taking the trace of an odd r power of the above equation we find

$$\text{Tr}[H_u, H_d]^r = 0. \quad (r \text{ odd}) \quad (93)$$

Therefore, one concludes that in the SM CP invariance implies $\text{Tr}[H_u, H_d]^r = 0$. For the case of $r = 1$ this relation is trivially satisfied, since the trace of a commutator is automatically zero. The minimum non-trivial case is for $r = 3$. Note that (93) is a necessary condition for CP invariance for any number of generations. For two generations the invariant automatically vanishes. In the case of three generations one obtains:

$$\begin{aligned} \text{Tr}[H_u, H_d]^3 &= 6i (m_t^2 - m_c^2)(m_t^2 - m_u^2)(m_c^2 - m_u^2) \\ &\quad (m_b^2 - m_s^2)(m_b^2 - m_d^2)(m_s^2 - m_d^2) iQ. \end{aligned} \quad (94)$$

It can be shown [25] that the vanishing of $\text{Tr}[H_u, H_d]^3$ is a necessary and sufficient condition for CP invariance in the SM with three generations. For more than three generations the vanishing of this invariant continuous being a necessary condition for CP invariance, but no longer is a sufficient condition. In the case of three generations one has¹:

$$\det[H_u, H_d] = \frac{1}{3} \text{tr}[H_u, H_d]^3, \quad (95)$$

and thus the vanishing of the above determinant can be used [26] as a necessary and sufficient condition for CP invariance.

Exercise 5. Prove the above result given by (94). Hint: choose to work in a weak basis where either H_u or H_d is diagonal.

3 Physics Beyond the Standard Model

3.1 Neutrino Masses

It was already mentioned that in the SM, neutrinos are strictly massless. There are no Dirac mass terms since no fermionic right-handed field ν_R is not introduced. There are no Majorana mass terms at tree level, since no scalar $SU(2)$ -triplet is introduced. Also no Majorana mass is generated at higher orders, due to exact $B - L$ conservation. Note that the Majorana mass term:

$$\nu_L^\top C \nu_L, \quad (96)$$

violates $B - L$ by 2 units. The same applies to $SU(5)$ GUT, where $B - L$ is also an accidental symmetry.

It can be readily seen that in the SM, there is no leptonic mixing. The leptonic charged currents is given by

$$\bar{\nu}_{Li}^0 \gamma^\mu \ell_{Li}^0 W_\mu, \quad (97)$$

¹ For any 3×3 traceless matrix A : $\text{Tr}A^3 = 3 \det(A)$.

where ν_{Li}^0 and ℓ_{Li}^0 are weak eigenstates. After diagonalisation of the charged lepton mass matrix the leptonic charged currents become

$$\bar{\nu}_L \gamma^\mu U \ell_L W_\mu , \quad (98)$$

where ℓ_L is now in the mass eigenstate. But the unitary matrix U can be eliminated through a redefinition of $n u_L^0$ so that the charged currents become flavour diagonal:

$$\bar{\nu}_L \gamma^\mu \ell_L W_\mu . \quad (99)$$

Observation of neutrino oscillations provides clear evidence for New Physics beyond the SM. The Minimal extension of the SM which allows for non-vanishing neutrino masses introduces right-handed neutrino fields, ν_{Ri} , a “strange” missing feature of the SM. Once the right-handed neutrino fields are introduced, Dirac masses for neutrinos are generated. Yukawa interactions can then be written as

$$(Y_\nu)_{ij} \bar{\ell}_i \tilde{\phi} \nu_{Ri} + \text{H.c.} , \quad (100)$$

where the Yukawa mass matrix Y_ν is arbitrarily complex. After the spontaneous breaking of the electroweak gauge symmetry Dirac neutrino masses are then generated as

$$m_D = (Y_\nu)_{ij} \frac{v}{\sqrt{2}} . \quad (101)$$

If one writes the most general Lagrangian consistent with gauge invariance and renormalisability, one has to include the mass term:

$$(M_R)_{ij} \nu_{Ri}^\top C \nu_{Rj} , \quad (102)$$

where M_R is a symmetric complex matrix. One may have $M_R \gg v$, since the mass term is gauge invariant. This leads to the seesaw mechanism, with:

$$(m_\nu)_{\text{light}} \approx \frac{v^2}{M_R} , \quad (103a)$$

$$(M)_{\text{heavy}} \approx M_R . \quad (103b)$$

Note that this minimal extension of the SM, sometimes denoted ν SM, is actually “simpler” and more “natural” than the SM, providing a simple and plausible explanation for the smallness of neutrino masses.

For the moment let us consider the low energy limit of the ν SM. Let us consider the neutrino masses and mixing at low energies, i.e the mass term Lagrangian

$$\mathcal{L}_{\text{mass}} = -\bar{\ell}_L m_\ell \ell_R - \frac{1}{2} \nu_L^\top C m_\nu \nu_L + \text{H.c.} , \quad (104)$$

and the charged current Lagrangian

$$\mathcal{L}_W = \frac{g}{\sqrt{2}} \bar{\ell}_L \gamma^\mu v_L W_\mu + \text{H.c.}, \quad (105)$$

where m_ℓ , an arbitrary complex matrix, and m_ν , a symmetric complex matrix, encode all information about lepton masses and mixing. There is a great redundancy in m_ℓ, m_ν , since not all their parameters are physical. This redundancy stems from the freedom to make Weak-Basis transformations:

$$v_L = W_L v_L, \quad \ell_L = W_L \ell_L, \quad \ell_R = W_R \ell_R, \quad (106)$$

where W_L, W_R are unitary matrices. The matrices m_ℓ, m_ν transform then as:

$$m'_\ell = W_L^\dagger m_\ell W_R, \quad m'_\nu = W_L^\top m_\nu W_L. \quad (107)$$

One can use the freedom to make WB transformations to go to a basis where

$$m_\ell = d_\ell \quad (108)$$

is diagonal and real. In this basis, one can still make a rephasing:

$$\ell''_{L,R} = K_L \ell_{L,R}, \quad v''_L = K_L v_L, \quad (109)$$

with $K_L = \text{diag}(e^{i\varphi_1}, e^{i\varphi_2}, e^{i\varphi_3})$. Under this rephasing d_ℓ remains invariant, but m_ν transforms as:

$$(m''_\nu)_{ij} = e^{i(\varphi_i + \varphi_j)} (m'_\nu)_{ij}. \quad (110)$$

One can eliminate n phases from m_ν . The number of physical phases in m_ν is:

$$N_{ph} = \frac{1}{2} n(n+1) - n = \frac{1}{2} n(n-1), \quad (111)$$

where n is the number of right-handed neutrino fields. For $n = 3$ one has $N_{ph} = 3$. So altogether one has in m_ν : 6 real moduli $|(m''_\nu)_{ij}|$ and three phases ($N_{ph} = 3$). The individual phases of $(m''_\nu)_{ij}$ have no physical meaning because they are not rephasing invariant. But one can construct polynomials of $(m''_\nu)_{ij}$ which are rephasing invariant. Examples of rephasing invariant polynomials:

$$\begin{aligned} P_1 &\equiv (m_\nu^*)_{11}(m_\nu^*)_{22}(m_\nu^*)_{12}^2, \\ P_2 &\equiv (m_\nu^*)_{11}(m_\nu^*)_{33}(m_\nu^*)_{13}^2, \\ P_3 &\equiv (m_\nu^*)_{33}(m_\nu^*)_{12}(m_\nu^*)_{13}(m_\nu^*)_{23}. \end{aligned} \quad (112)$$

Let us now discuss the generation of the leptonic mixing in the charged current. The charged leptonic mass matrix is diagonalised as

$$U_{eL}^\dagger m_\ell U_{eR} = \text{diag}(m_e, m_\mu, m_\tau) \quad (113)$$

through the unitary matrices U_{eL} , U_{eR} , while the effective neutrino mass matrix is diagonalised as

$$U_v^\top m_v U_v = \text{diag}(m_1, m_2, m_3), \quad (114)$$

through the unitary matrix U_v . Under these unitary transformations the charged currents are

$$\mathcal{L}_W = \frac{g}{\sqrt{2}} \bar{\ell}_L \gamma^\mu U v_L W_\mu + \text{H.c..} \quad (115)$$

The unitary matrix U , which measure the mixing in the leptonic sector, given by

$$U \equiv U_\ell^\dagger U_v, \quad (116)$$

is the so-called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) [27–29] matrix. In this basis, there is still freedom to rephase the charged lepton fields

$$\ell_j \longrightarrow \ell'_j = \exp(i \phi_j) \ell_j, \quad (117)$$

with arbitrary phases ϕ_j . Due to the Majorana nature of neutrinos, the rephasing:

$$v_{Lk} \longrightarrow v'_{Lk} = \exp(i \psi_j) v_{Lk}, \quad (118)$$

with arbitrary phases ψ_k , is not allowed, since it would not leave the Majorana mass terms

$$v_{Lk}^\top C m_k v_{Lk}, \quad (119)$$

invariant. In the mass eigenstate basis, the charred currents are

$$\mathcal{L}_W = \frac{g}{\sqrt{2}} (\bar{e} \bar{\mu} \bar{\tau})_L \gamma^\mu \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}_L W_\mu + \text{H.c..} \quad (120)$$

For the moment we do not introduce the constraints of 3×3 unitarity. Note that in the context of type-one seesaw the PMNS matrix is not unitary.

3.1.1 Rephasing Invariant Quantities

Let us recall the situation in the quark sector where the charged current are given in the mass eigenstate basis as

$$\mathcal{L}_W = \frac{g}{\sqrt{2}} (\bar{u} \bar{c} \bar{t})_L \gamma^\mu \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L W_\mu + \text{H.c..} \quad (121)$$

The CKM matrix V , has 9 moduli and the 4 rephasing invariant phases defined as

$$\beta \equiv \arg(-V_{cd} V_{tb} V_{cb}^* V_{td}^*) , \quad (122a)$$

$$\gamma \equiv \arg(-V_{ud} V_{cb} V_{ub}^* V_{cd}^*) , \quad (122b)$$

$$\chi = \beta_s \equiv \arg(-V_{cb} V_{ts} V_{cs}^* V_{tb}^*) , \quad (122c)$$

$$\chi' \equiv \arg(-V_{us} V_{cd} V_{ud}^* V_{cs}^*) . \quad (122d)$$

A novel feature in the leptonic sector with Majorana neutrinos, is the presence of rephasing invariant bilinear:

$$\arg(U_{\ell\alpha} U_{\ell\beta}^*) , \quad (123)$$

where there is no summation of repeated indices. These are the so-called **Majorana-type phase**. There are six independent Majorana-type phase. This is true even when unitarity is not imposed on the PMNS matrix U . It applies to a general framework with an arbitrary number of right-handed neutrinos. A possible choice for the six independent Majorana-type phases is [30]:

$$\beta_1 \equiv \arg(U_{e1}, U_{e2}^*) , \quad \gamma_1 \equiv \arg(U_{e1}, U_{e3}^*) , \quad (124a)$$

$$\beta_2 \equiv \arg(U_{\mu 1}, U_{\mu 2}^*) , \quad \gamma_2 \equiv \arg(U_{\mu 1}, U_{\mu 3}^*) , \quad (124b)$$

$$\beta_3 \equiv \arg(U_{\tau 1}, U_{\tau 2}^*) , \quad \gamma_3 \equiv \arg(U_{\tau 1}, U_{\tau 3}^*) . \quad (124c)$$

One can choose the following four independent Dirac-type invariant phases:

$$\sigma_{e\mu}^{12} \equiv \arg(U_{e1} U_{\mu 2} U_{e2}^* U_{\mu 1}^*) = \beta_1 - \beta_2 , \quad (125a)$$

$$\sigma_{e\tau}^{12} \equiv \arg(U_{e1} U_{\tau 2} U_{e2}^* U_{\tau 1}^*) = \beta_1 - \beta_3 , \quad (125b)$$

$$\sigma_{e\mu}^{13} \equiv \arg(U_{e1} U_{\mu 3} U_{e3}^* U_{\mu 1}^*) = \gamma_1 - \gamma_2 , \quad (125c)$$

$$\sigma_{e\tau}^{13} \equiv \arg(U_{e1} U_{\tau 3} U_{e3}^* U_{\tau 1}^*) = \gamma_1 - \gamma_3 . \quad (125d)$$

If one assumes 3×3 unitarity of the PMNS matrix, the full leptonic mixing matrix can be reconstructed [22, 30, 31] from the six independent Majorana phases given

in (124a)–(124c). Normalisation of rows and columns plays an important rôle. It prevents the “blowing up” of unitarity triangles. For three generations and assuming 3×3 unitarity of the PMNS matrix can be parametrised by:

$$U = V K, \quad (126)$$

where V is parametrised through the Standard Parametrisation:

$$V(\theta_{12}, \theta_{13}, \theta_{23}, \delta_{13}) = \begin{pmatrix} c_{12} c_{13} & s_{12} c_{13} & s_{13} e^{-i\delta_{13}} \\ -s_{12} c_{23} - c_{12} s_{23} s_{13} e^{i\delta_{13}} & c_{12} c_{23} - s_{12} s_{23} s_{13} e^{i\delta_{13}} & s_{23} c_{13} \\ s_{12} s_{23} - c_{12} c_{23} s_{13} e^{i\delta_{13}} & -c_{12} s_{23} - s_{12} c_{23} s_{13} e^{i\delta_{13}} & c_{23} c_{13} \end{pmatrix}, \quad (127)$$

and the phase matrix K given by $K = \text{diag}(1, e^{i\alpha_1/2}, e^{i\alpha_2/2})$. One can eliminate the phase δ from the first row by writing:

$$U = V K', \quad (128)$$

where

$$K' = \text{diag}(1, 1, e^{i\delta}) K, \quad (129)$$

convenient for the analysis of the neutrinoless double-beta decay ($0\nu\beta\beta$). The unitarity triangles in the leptonic sector, within the hypothesis of unitarity of the PMNS matrix, can be split into two category: Dirac unitarity triangles

$$\begin{aligned} U_{e1} U_{\mu 1}^* + U_{e2} U_{\mu 2}^* + U_{e3} U_{\mu 3}^* &= 0, & T_{e\mu} \\ U_{e1} U_{\tau 1}^* + U_{e2} U_{\tau 2}^* + U_{e3} U_{\tau 3}^* &= 0, & T_{e\tau} \\ U_{\mu 1} U_{\tau 1}^* + U_{\mu 2} U_{\tau 2}^* + U_{\mu 3} U_{\tau 3}^* &= 0, & T_{\mu\tau} \end{aligned} \quad (130)$$

Majorana unitarity triangles

$$\begin{aligned} U_{e1} U_{e2}^* + U_{\mu 1} U_{\mu 2}^* + U_{\tau 1} U_{\tau 2}^* &= 0, & T_{12} \\ U_{e1} U_{e3}^* + U_{\mu 1} U_{\mu 3}^* + U_{\tau 1} U_{\tau 3}^* &= 0, & T_{13} \\ U_{e2} U_{e3}^* + U_{\mu 2} U_{\mu 3}^* + U_{\tau 2} U_{\tau 3}^* &= 0, & T_{23} \end{aligned} \quad (131)$$

Let us analyse an example of a Majorana triangle, which is shown in Fig. 5. This triangle is a typical triangle that depends only on Majorana phases. The Majorana phases give the directions of the sides of the Majorana unitary triangles. The arrows in Fig. 5 have no meaning! They can be reversed if one makes, for example the rephasing:

$$\nu_3 \longrightarrow \nu'_3 = -\nu_3. \quad (132)$$

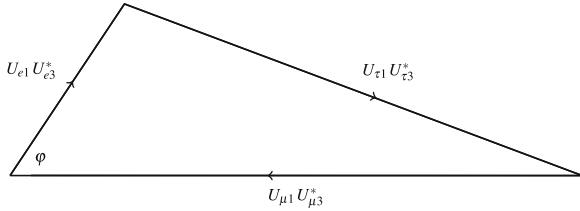


Fig. 5 Unitarity triangle

The angle φ is a Dirac-type phase, since:

$$\varphi = \arg(-U_{e1} U_{e3}^* U_{\tau 1} U_{\tau 3}) = \pi - (\gamma_3 - \gamma_1). \quad (133)$$

Majorana triangles provide necessary and sufficient conditions for having CP invariance with Majorana neutrinos, namely:

- Vanishing of their common area:

$$A = \frac{1}{2} |\mathrm{i}Q|; \quad (134)$$

- Orientation of all the “collapsed” triangles along the real axis or the imaginary axis. If one of these triangles T_{ik} is parallel to the imaginary axis, that means that the neutrinos i, k have opposite CP parities.

The $0\nu\beta\beta$, which is sensitive to Majorana-type phases, is proportional to the quantity m_{ee} given by

$$\begin{aligned} m_{ee}^2 &= m_1^2 |U_{e1}|^4 + m_2^2 |U_{e2}|^4 + m_3^2 |U_{e3}|^4 \\ &\quad + 2m_1 m_2 |U_{e1}|^2 |U_{e1}|^2 \cos 2\beta_1 + 2m_1 m_3 |U_{e1}|^2 |U_{e3}|^2 \cos 2\gamma_1 \\ &\quad + 2m_2 m_3 |U_{e2}|^2 |U_{e3}|^2 \cos 2(\beta_1 - \gamma_1). \end{aligned} \quad (135)$$

Note that the angle $(\beta_1 - \gamma_1)$ is the argument of $(U_{e1}^* U_{e2} U_{e1} U_{e3}^*)$ which is not a rephasing invariant Dirac-type quartet. If one adopts the parametrisation given by (128) with the definition given in (127), one then has:

$$m_{ee} = \left| c_{13}^2 \left(m_1 c_{12}^2 + m_2 e^{-i\alpha_1} s_{12}^2 \right) + m_3 e^{-i\alpha_2} s_{13}^2 \right|. \quad (136)$$

This is the reason why this parametrisation is useful for the analysis of $0\nu\beta\beta$.

One may ask whether the effective neutrino mass matrix can be determined from experiment. We have seen that in the WB where the charged lepton mass matrix is diagonal, real one has:

$$m_\ell = \text{diag}(m_e, m_\mu, m_\tau), \quad m_v = \begin{pmatrix} m_{ee} & m_{e\mu} & m_{e\tau} \\ m_{e\mu} & m_{\mu\mu} & m_{\mu\tau} \\ m_{e\tau} & m_{\mu\tau} & m_{\tau\tau} \end{pmatrix}. \quad (137)$$

One can use rephasing freedom to make m_{ee} , $m_{\mu\mu}$, $m_{\tau\tau}$ real, but $m_{e\mu}$, $m_{e\tau}$, $m_{\mu\tau}$ complex. Altogether we end up with 6 real parameters and three phases, a total of 9 real parameters. Let us then compare with the physical quantities in m_ν , that can be obtained through feasible experiments: two mass squared differences, Δm_{12}^2 , Δm_{23}^2 , three mixing angles θ_{12} , θ_{23} , θ_{13} , the imaginary of the quartet iQ from CP violation in ν oscillations and $0\nu\beta\beta$ from the measurement of m_{ee} . This is a total of 7 measurable quantities (Glashow's counting). We arrive at the dreadful conclusion ($7 < 9$) that no presently conceivable set of feasible experiments can fully determine the effective neutrino matrix [32]. Some of the ways out:

- Postulate “texture zeroes” in m_ν [32]: various sets of zeroes allowed by experiment.
- Postulate $\det(m_\nu) = 0$. This leads to 7 parameters in m_ν [33].

Let us now derive CP-odd Weak-basis invariants in the leptonic sector with Majorana neutrinos. The relevant part of the Lagrangian is:

$$\mathcal{L}_{\text{mass}} = -\bar{\ell}_L m_\ell \ell_R - \frac{1}{2} v_L^\top C m_\nu v_L + \frac{g}{\sqrt{2}} \bar{\ell}_L \gamma_\mu v_L W^\mu + \text{H.c.} \quad (138)$$

The CP transformation properties of the various fields are dictated by the part of the Lagrangian which conserves CP, namely the gauge interactions. One has to keep in mind the fact that the gauge sector of the SM does not distinguish the various families of fermions. The most general CP transformations which leave $\mathcal{L}_{\text{gauge}}$ invariant is:

$$\text{CP } \ell_L (\text{CP})^\dagger = W_L \gamma^0 C \bar{\ell}_L^\top, \quad (139a)$$

$$\text{CP } v_L (\text{CP})^\dagger = W_L \gamma^0 C \bar{v}_L^\top, \quad (139b)$$

$$\text{CP } \ell_R (\text{CP})^\dagger = W_R \gamma^0 C \bar{\ell}_R^\top, \quad (139c)$$

where W_L , W_R are unitarity matrices acting in generation space. The Lagrangian of the leptonic sector conserves CP if and only if the leptonic mass matrices m_ν , m_ℓ satisfy

$$W_L^\top m_\nu W_L = -m_\nu^*, \quad W_L^\dagger m_\ell W_R = m_\ell^*, \quad (140)$$

or

$$W_L^\dagger \tilde{h}_\nu W_L = \tilde{h}_\nu^*, \quad W_L^\dagger h_e W_R = h_e^*, \quad (141)$$

where $\tilde{h}_\nu \equiv h_\nu^*$, $h_\nu \equiv m_\nu m_\nu^*$ and $h_e \equiv m_e m_e^\dagger$. One then gets

$$W_L^\dagger [\tilde{h}_\nu, h_\ell] W_L = [\tilde{h}_\nu^*, h_\ell^*] = [\tilde{h}_\nu^\top, h_\ell^\top] = -[\tilde{h}_\nu, h_\ell]^\top. \quad (142)$$

Making the cube of both sides, one gets

$$W_L^\dagger \left[\tilde{h}_\nu, h_\ell \right]^3 W_L = - \left(\left[\tilde{h}_\nu, h_\ell \right]^3 \right)^\top. \quad (143)$$

Therefore, CP invariance implies

$$\text{Tr} \left[\tilde{h}_\nu, h_\ell \right]^3 = 0, \quad (144)$$

which is valid for an arbitrary number of generations. This relation, first derived by [25], can be written for three generations in terms of measurable quantities as

$$\text{Tr} \left[\tilde{h}_\nu, h_\ell \right]^3 = -6i (m_\mu^2 - m_e^2) (m_\tau^2 - m_\mu^2) (m_\tau^2 - m_e^2) \Delta m_{21}^2 \Delta m_{31}^2 \Delta m_{32}^2 iQ, \quad (145)$$

where

$$iQ = \frac{1}{8} \sin 2\theta_{12} \sin 2\theta_{13} \sin 2\theta_{23} \sin \delta. \quad (146)$$

This invariant is sensitive to Dirac type CP violation. For three generations, the vanishing of this invariant is the necessary and sufficient for the absence of Dirac-type CP violation. Using the previous method, one can derive invariants sensitive to Majorana-type CP violation [34]. The following invariant is sensitive to Majorana-type CP violation:

$$I_{\text{Majorana}}^{\text{CP}} = i\text{Tr} \left(m_\ell m_\ell^\dagger m_\nu^* m_\nu m_\nu^* m_\ell^\top m_\ell^* m_\nu \right). \quad (147)$$

The simplest way to check that $I_{\text{Majorana}}^{\text{CP}}$ is sensitive to Majorana-type CP violation is by evaluating it in the case of two generations of Majorana neutrinos

$$I_{\text{Majorana}}^{\text{CP}} = \frac{1}{4} m_1 m_2 \Delta m_{21}^2 (m_\nu^2 - m_e^2) \sin^2 \theta \sin 2\gamma, \quad (148)$$

where the PMNS matrix U is parametrised as

$$U = \begin{pmatrix} \cos \theta & -\sin \theta e^{i\gamma} \\ \sin \theta e^{-i\gamma} & \cos \theta \end{pmatrix}. \quad (149)$$

The invariant $I_{\text{Majorana}}^{\text{CP}}$ vanishes, for example, if $\gamma = \pi/2$, since this corresponds to having CP invariance, with the two neutrinos with opposite CP parities.

3.2 Violating 3×3 CKM Unitarity

Suppose that one drops the requirement of 3×3 unitarity. How many parameters are there in the 3×3 CKM matrix? By taking into account the elements of the CKM matrix V :

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} & \cdots \\ V_{cd} & V_{cs} & V_{cb} & \cdots \\ V_{td} & V_{ts} & V_{tb} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (150)$$

One counts 9 moduli plus 4 ($9 - 5$) rephasing invariant phases for a total of 13 parameters. A convenient choice for 4 independent rephasing invariant phases is:

$$\beta \equiv \arg(-V_{cd} V_{tb} V_{cb}^* V_{td}^*) , \quad (151a)$$

$$\gamma \equiv \arg(-V_{ud} V_{cb} V_{ub}^* V_{cd}^*) , \quad (151b)$$

$$\chi = \beta_s \equiv \arg(-V_{cb} V_{ts} V_{cs}^* V_{tb}^*) , \quad (151c)$$

$$\chi' \equiv \arg(-V_{us} V_{cd} V_{ud}^* V_{cs}^*) , \quad (151d)$$

The SM with three generations predicts a series of exact relations among the 13 measurable (in principle) quantities. Again we should emphasise that the relation

$$\alpha + \beta + \gamma = \pi , \quad (152)$$

is not a test of unitarity. It is true, by definition!

$$\alpha \equiv \arg(-V_{td} V_{ub} V_{ud}^* V_{tb}^*) , \quad (153a)$$

$$\beta \equiv \arg(-V_{cd} V_{tb} V_{cb}^* V_{id}^*) , \quad (153b)$$

$$\gamma \equiv \arg(-V_{ud} V_{cb} V_{ub}^* V_{cd}^*) . \quad (153c)$$

In the derivation of the unitary relations, it is useful to adopt a convenient phase convention [35]. Without loss of generality one can choose:

$$\arg(V) = \begin{pmatrix} 0 & \chi' & -\gamma \\ \pi & 0 & 0 \\ -\beta \pi + \chi & 0 & 0 \end{pmatrix} . \quad (154)$$

We have used the 5 rephasing degrees of freedom to fix 5 of the nine phases. We are left with 4 phases.

In the case of the SM where the CKM matrix is strictly unitary, one has exact relations predicted by the SM, such as [30, 31]:

$$|V_{ub}| = \frac{|V_{cd}||V_{cb}|}{|V_{ub}|} \frac{\sin \beta}{\sin(\beta + \gamma)}, \quad (155a)$$

$$\sin \chi = \frac{|V_{td}|}{|V_{ts}|} \frac{|V_{cd}|}{|V_{cs}|} \sin \beta, \quad (155b)$$

$$\frac{|V_{ub}|}{|V_{tb}|} = \frac{\sin \beta}{\sin \gamma} \frac{|V_{tb}|}{|V_{ud}|}, \quad (155c)$$

$$\sin \chi = \frac{|V_{us}||V_{ub}|}{|V_{cs}||V_{cb}|} \sin(-\chi + \chi' + \gamma). \quad (155d)$$

Violation of any of these exact relations signals the presence of New Physics which may involve deviations of 3×3 unitarity or not. The presence of New Physics contributions to $B_d - \bar{B}_d$ and $B_s - \bar{B}_s$ mixings affects the extraction of $|V_{td}|$, $|V_{ts}|$ from the data, even in the framework of New Physics which respects 3×3 unitarity. An example of that is the Supersymmetric extension of the SM. In many of the extensions of the SM, the dominant effect of New Physics arises from new contributions to $B_d - \bar{B}_d$ and $B_s - \bar{B}_s$ mixings, which is convenient to parametrise as:

$$M_{12}^q = (M_{12}^q)^{\text{SM}} r_q^2 e^{2i\theta_q}, \quad (156)$$

with $q = d, s$. Thus the mass difference ΔM_{B_d} is now given by

$$\Delta M_{B_d} = r_d^2 (\Delta M_{B_d})^{\text{SM}}, \quad (157)$$

that affects the extraction of $|V_{td}|$ from experiment. On the other hand, the mass difference ΔM_{B_s} is

$$\Delta M_{B_s} = r_s^2 (\Delta M_{B_s})^{\text{SM}}, \quad (158)$$

that affects the extraction of $|V_{ts}|$.

From the CP asymmetries $S_{J/\psi K_s}$ and $S_{\rho^+ \rho^-}$ given by

$$S_{J/\psi K_s} = \sin(2\beta + 2\theta_d) \equiv \sin 2\bar{\beta}, \quad (159a)$$

$$S_{\rho^+ \rho^-} = \sin(2\alpha - 2\theta_s) \equiv \sin 2\bar{\alpha}. \quad (159b)$$

How to detect the presence of New Physics? The answer is: one can use the exact relations predicted by the SM. The extraction θ_d from

$$|V_{ub}| = \frac{|V_{cd}||V_{cb}|}{|V_{ub}|} \frac{\sin \beta}{\sin(\beta + \gamma)}, \quad (160)$$

which then leads to

$$\tan \theta_d = \frac{R_u \sin(\gamma + \bar{\beta}) - \sin \bar{\beta}}{\cos \bar{\beta} - R_u \cos(\gamma + \bar{\beta})}, \quad (161)$$

where

$$R_u = \frac{|V_{ud}| |V_{ub}|}{|V_{cd}| |V_{cb}|}. \quad (162)$$

While to extract θ_s one must use the exact relation

$$\sin \chi = \frac{|V_{us}| |V_{ub}|}{|V_{cs}| |V_{cb}|} \sin(\gamma - \chi + \chi'), \quad (163)$$

which then leads to

$$\tan \theta_s = \frac{\sin \bar{\chi} - C \sin(\gamma - \bar{\chi})}{C \cos(\gamma - \bar{\chi}) - \cos \bar{\chi}}, \quad (164)$$

where

$$C = \frac{|V_{us}| |V_{ub}|}{|V_{cs}| |V_{cb}|}. \quad (165)$$

To an excellent approximation one has [36]:

$$\sin \chi = \frac{|V_{us}|^2}{|V_{ud}|^2} \frac{\sin \beta \sin \gamma}{\sin \gamma + \beta}, \quad (166)$$

or [35]:

$$\sin \chi = \frac{|V_{td}|}{|V_{ts}|} \frac{|V_{cd}|}{|V_{cs}|} \sin \beta. \quad (167)$$

If either (γ, χ) or $(\frac{\Delta M_{B_d}}{\Delta M_{B_s}}, \chi)$ are measured with some precision, one has novel stringent tests of the SM, where contribution of New Physics can be significant. At this point, the following point should be emphasised. There is clear evidence for a complex CKM matrix even if one allows for the presence of New Physics [35]. This is essentially due to the evidence for a non-vanishing γ , which is not contaminated by the presence of New Physics!

Since we are considering experimental tests of 3×3 unitarity of the CKM matrix, one should ask the following questions:

- Can one have self-consistent extensions of the SM, where deviations of 3×3 unitarity of the CKM matrix may occur?
- Can these deviations be naturally small?

The answer to both questions is positive! In the next section we describe an extension of the SM with the addition of a vector-like quarks.

3.3 SM with the Addition of a Isosinglet Down-Type Quark

We shall consider in this section extensions of the SM with vector-like isosinglet quarks of $Q = -1/3$ and $Q = 2/3$. One question one may raise is whether the addition of vector-like quarks to SM can bring important features to solve flavour issues presented in the SM. We point out several reasons to consider vector-like quarks:

- (i) they provide a self-consistent framework with naturally small violations of 3×3 unitarity of the CKM matrix.
- (ii) Lead to naturally small Flavour Changing Neutral Currents (FCNC) mediated by Z_μ .
- (iii) Provide the simplest framework to have spontaneous CP violation [37, 38], with a vacuum Phase generating a non-trivial CKM phase. An important requirement is that there is experimental evidence of a complex CKM matrix even if one allows for the presence of New Physics.
- (iv) Provide New Physics contributions to $B_d - \bar{B}_d$ mixing and $B_s - \bar{B}_s$ mixing.
- (v) Provide a simple solution to the *Strong CP problem*, which does not require Axions.
- (vi) May contribute to the understanding of the observed pattern of fermion masses and mixing.
- (vii) Provide a framework where there is a common origin of all CP violations [39]:
 - (i) CP violation in the Quark Sector;
 - (ii) CP violation in the Lepton Sector detectable through neutrino oscillations $U_{e3} \neq 0$ and “relatively large”. This is a great feature!
 - (iii) CP violation need to generate the Baryon Asymmetry of the Universe through Leptogenesis.

There is nothing strange in having deviations of 3×3 unitarity. The PMNS matrix in the leptonic sector in the context of type-one seesaw (ν SM) is not 3×3 unitarity.

For simplicity let us study the Minimal Model where one adds a vector-like quark field D^0 into SM. This down-type quark particle D^0 has the property that both chiral fields D_L^0 and D_R^0 are $SU(2)_L$ singlets with electric charge $Q = -1/3$ (one could also have introduced a isosinglet of the up-type instead with electric charge $Q = 2/3$). To complete the fermionic content of this Minimal Model we introduce 3 right-handed neutrinos ν_{Rj}^0 . The Higgs sector is just extended with a neutral complex singlet field S .

Since we want to have *Spontaneous CP Violation*, we impose CP invariance at the Lagrangian level, i.e. all couplings are taken real. We add a Z_4 symmetry, under which the SM fields transform as [37–39]:

$$\ell^0 \rightarrow i \ell^0, \quad e_{Rj}^0 \rightarrow i e_{Rj}^0, \quad \nu_{Rj}^0 \rightarrow i \nu_{Rj}^0, \quad (168)$$

and the remaining SM fermions transform trivially. The new particles transform as

$$D^0 \rightarrow -D^0, \quad S \rightarrow -S. \quad (169)$$

The discrete symmetry Z_4 is crucial to obtain a solution of the Strong CP problem and Leptogenesis. The scalar potential contains various terms which do not have phase dependence but there are terms with phase dependence, which are given by

$$V_{\text{phase}}(\phi, S) = \left[\mu^2 + \lambda_1 S^* S + \lambda_2 \phi^\dagger \phi \right] (S^2 + S^{*2}) + \lambda_3 (S^4 + S^{*4}). \quad (170)$$

There is a range of the parameters of the Higgs potential, where the minimum is at:

$$\langle \phi \rangle = \frac{v}{\sqrt{2}}, \quad \langle S \rangle = \frac{V}{\sqrt{2}} e^{i\theta}. \quad (171)$$

The most general $SU(2)_C \times SU(2)_L \times U(1) \times Z_4$ invariant Yukawa couplings

$$-\mathcal{L} = (\bar{u}^0 \bar{d}^0)_{Li} \left[g_{ij} \phi d_{Rj}^0 + h_{ij} \phi u_{Rj}^0 \right] + \overline{M} \overline{D}_L^0 D_R^0 + \text{H.c..} \quad (172)$$

This implies that the quark mass matrix for down-type quarks has the following form:

$$\bar{\mathbf{d}}_L \mathcal{M} \mathbf{d}_R = \begin{pmatrix} \bar{d}_{1L}^0 & \bar{d}_{2L}^0 & \bar{d}_{3L}^0 & \overline{D}_L^0 \end{pmatrix} \begin{pmatrix} m_d & 0 \\ M_1 & M_2 & M_3 & \overline{M} \end{pmatrix} \begin{pmatrix} d_{1R}^0 \\ d_{2R}^0 \\ d_{3R}^0 \\ D_R^0 \end{pmatrix}, \quad (173)$$

where

$$M_j = f_j V e^{i\theta} + f'_j V e^{i\theta}. \quad (174)$$

The zero $3 \times$ column in the down-type quark matrix in (173) is due to the Z_4 symmetry. The down quarks masses are then obtained through the diagonalisation:

$$\mathcal{U}_L^\dagger \mathcal{M} \mathcal{M}^\dagger \mathcal{U}_L = \text{diag}(m_d^2, m_s^2, m_b^2, m_D^2). \quad (175)$$

Defining the block-entries of the unitary matrix \mathcal{U} as

$$\mathcal{U}_L = \begin{pmatrix} K & R \\ S & T \end{pmatrix}, \quad (176)$$

one can easily derive approximative the effective down quark mass matrix $m_{\text{eff}} m_{\text{eft}}^\dagger$ as:

$$K^{-1} m_{\text{eff}} m_{\text{eft}}^\dagger K = \text{diag}(m_d^2, m_s^2, m_b^2), \quad (177)$$

where $m_{\text{eff}} m_{\text{eft}}^\dagger$ is given by

$$m_{\text{eff}} m_{\text{eft}}^\dagger = m_d m_d^\dagger - \frac{m_d M^\dagger M m_d^\dagger}{M M^\dagger + \bar{M}^2}. \quad (178)$$

It is worth to point out that $m_{\text{eff}} m_{\text{eff}}^\dagger$ is complex since the combination $M^\dagger M$ is complex because of (174). A remarkable feature of the Model is that the phase θ arising from $\langle S \rangle$, generates a non-trivial CKM phase, provided $|M_j|$ and \bar{M} are of the same order of magnitude, which it is natural. We shall now see that within this model that deviations of 3×3 unitarity and Flavour-Changing Neutral Currents are naturally small. Let us write down the charged and neutral currents in the model:

$$\mathcal{L}_W = -\frac{g}{\sqrt{2}} (\bar{u} \bar{c} \bar{t})_L \gamma^\mu \begin{pmatrix} d \\ s \\ b \\ D \end{pmatrix}_L W_\mu^+, \quad (179)$$

and

$$\mathcal{L}_Z = -\frac{g}{2 \cos \theta_W} \left\{ (\bar{u} \bar{c} \bar{t})_L \gamma^\mu \begin{pmatrix} u \\ c \\ t \end{pmatrix}_L + (\bar{d} \bar{s} \bar{b} \bar{D}) \begin{pmatrix} K^\dagger K & K^\dagger R \\ R^\dagger K & R^\dagger R \end{pmatrix} \gamma^\mu \begin{pmatrix} d \\ s \\ b \\ D \end{pmatrix}_L \right. \\ \left. - \sin^2 \theta_W J_{\text{em}}^\mu \right\} Z_\mu, \quad (180)$$

Let us now quantify the deviations of 3×3 unitarity. Since the full 6×6 matrix \mathcal{U}_L is unitary, the following relations are verified:

$$K^\dagger K + S^\dagger S = \mathbf{1}, \quad R^\dagger R + T^\dagger T = \mathbf{1}, \quad R^\dagger K + T^\dagger S = \mathbf{0}, \quad (181)$$

and one derives

$$S \approx -\frac{M m_d^\dagger K}{\bar{M}^2} \rightarrow \mathcal{O}\left(\frac{m}{M}\right), \quad (182)$$

Table 2 Deviations between experiment and theory

	Prediction	Measurement	Pull
$\sin 2\beta$	0.81 ± 0.05	0.680 ± 0.023	2.4
γ	$68^\circ \pm 3^\circ$	$76^\circ \pm 11^\circ$	<1
α	$88^\circ \pm 4^\circ$	$91^\circ \pm 6^\circ$	<1
$ V_{cb} \times 10^3$	42.3 ± 0.9	41.0 ± 1.0	<1
$ V_{ub} \times 10^3$	3.62 ± 0.14	3.82 ± 0.56	<1
$\epsilon_K \times 10^3$	1.96 ± 0.20	2.23 ± 0.01	1.4
$BR(B \rightarrow \tau \nu) \times 10^4$	0.82 ± 0.08	1.67 ± 0.30	-2.7

From UTfit collaboration, Cecilia Tarantino talk at ICHEP2012 conference

and therefore, making use of the relations given in (181), one concludes that the deviations of 3×3 unitarity,

$$K^\dagger K = \mathbf{1} - \mathcal{O}\left(\frac{m^2}{M^2}\right) \quad (183)$$

are naturally small. Note that there is nothing strange about violations of 3×3 unitarity. The PMNS matrix is not unitary in the framework of seesaw mechanism, type-one.

Can extensions of the SM with vector-like quarks “solve” some of the tensions between SM and experiment? The answer is yes! In the framework of an extension of the SM, with one $Q = 2/3$ vector like quark, it has been shown that the tensions can be solved and various correlations are predicted [40]. But, the important point is for experiment/theory to confirm that deviations are really there. In Table 2 we present deviations between experiment and theory presented from UTfit Collaboration in the ICHEP2012 Conference.

3.3.1 Leptonic Sector

We recall that the leptonic fields transform under Z_4 as:

$$\ell^0 \rightarrow i \ell^0, \quad e_{Rj}^0 \rightarrow i e_{Rj}^0, \quad \nu_{Rj}^0 \rightarrow i \nu_{Rj}^0. \quad (184)$$

This constrains the Yukawa Lagrangian terms as:

$$-\mathcal{L}_\ell = \bar{\ell}_L^0 G_\ell \phi e_R^0 + \bar{\ell}_L^0 G_\nu \phi \nu_R^0 + \frac{1}{2} \nu_R^{0\top} C (f_\nu S + f'_\nu S^*) \nu_R^0 + \text{H.c.}, \quad (185)$$

which after spontaneous breaking become as

$$-\mathcal{L}_\ell = \bar{e}_L^0 G_\ell \phi e_R^0 + \frac{1}{2} n_L^\top C \mathcal{M}^* n_L + \text{H.c.}, \quad (186)$$

with

$$n_L \equiv \begin{pmatrix} v_L^0 \\ (v_R)^C \end{pmatrix}. \quad (187)$$

The 6×6 matrix is then given by

$$\mathcal{M}_\nu = \begin{pmatrix} 0 & m \\ m^\top & M_\nu \end{pmatrix}, \quad (188)$$

where

$$m_\ell = \frac{\nu}{\sqrt{2}} G_\ell, \quad m = \frac{\nu}{\sqrt{2}} G_\nu, \quad (189)$$

and M_ν is given by

$$M_\nu = \frac{V}{\sqrt{2}} (f_\nu^+ \cos \theta + i f_\nu^- \sin \theta), \quad (190)$$

where $f_\nu^\pm \equiv f_\nu \pm f_\nu$.

In the weak-basis where m_ℓ is diagonal, real, the light neutrino masses and low energy leptonic mixing are obtained from

$$K^\dagger \left(m M_\nu^{-1} m^\top \right) K = \text{diag}(m_1, m_2, m_3) \quad (191)$$

where m is real, but since M_ν is a generic complex matrix, the effective light neutrino mass m_ν is also a generic complex symmetric matrix. Thus, the matrix K has three complexes phases: one Dirac-type and two Majorana-type.

3.3.2 Conclusions

Vector-like quarks provide a very interesting scenario for New Physics. They are a crucial ingredient in the simplest realistic model of spontaneous CP violation where a complex CKM matrix is generated from a vacuum phase. They provide a consistent framework where there are naturally small deviations of 3×3 unitarity in the CKM matrix, leading to naturally small Z-FCNC. They provide a simple solution to the *Strong CP problem*, without the need of introducing *axions*.

The Standard Model and its CKM mechanism for mixing and CP violation is in good agreement with experiment. This is a remarkable fact in view of the *large amount of data*: $|V_{us}|, |V_{ub}|, |V_{cb}|, \gamma$ completely fix the CKM matrix. Then with no free parameters, one has to accommodate a large number of measurable quantities like ε_K , $B_d - \bar{B}_d$ mixing, $B_s - \bar{B}_s$, β, β_s , rare B -decays, rare kaon decays, etc. Unfortunately there are *hadronic uncertainties*.

There is room for New Physics which could be detected in LHCb and future super-B factories. The spectrum of Fermion Masses and the Pattern of quark and lepton mixing remains one of the Fundamental Questions in Particle Physics. It is very likely that detectable New Physics be involved in the solution of the Flavour Puzzle. The observation of neutrino oscillation is a strong indication to search for an extension of the SM that can account for neutrino masses.

3.4 Baryon Asymmetry of the Universe

We now address the question how to generate the Baryon Asymmetry of the Universe (BAU). The ingredients to dynamically generate BAU from an initial state with zero Baryon Asymmetry were formulate in 1967 by Sakharov [41]:

- (i) Baryon number violation;
- (ii) C and CP violation;
- (iii) Departure from thermal equilibrium.

All these ingredients exist in the SM, but it has been established that in the SM, one cannot generate the observed BAU:

$$\eta_B \equiv \frac{n_B - n_{\bar{B}}}{n_\gamma} = (6.20 \pm 0.15) \times 10^{-10}, \quad (192)$$

where n_B , $n_{\bar{B}}$, n_γ correspond to number densities of baryons, anti-baryons and photons at present time, respectively. There are mainly two reasons why the SM cannot generate sufficient BAU:

- (i) CP violation in the SM is too small. Indeed one has:

$$\frac{\text{Tr}[H_u, H_d]^3}{T_{\text{EW}}^{12}} \simeq 10^{-20}. \quad (193)$$

- (ii) Successful Baryogenesis requires a strongly first order phase transition which would require a light Higgs mass:

$$m_H \leq 70 \text{ GeV}. \quad (194)$$

One concludes that an explanation of the observed BAU requires New Physics beyond the SM. Leptogenesis, suggested by Fukugita and Yanagida is one the simplest and most attractive mechanism to generate BAU. In the framework of leptogenesis, BAU is generated through out of equilibrium decays of right-handed neutrinos that create a lepton asymmetry which is in turn converted into a baryon asymmetry by $(B + L)$ violating but $(B - L)$ conserving sphaleron interactions.

The ν SM, i.e. the extension of the SM consisting of adding 3 right-handed neutrinos has all the ingredients to have Leptogenesis. For recent reviews, see [42, 43].

$$\begin{aligned} -\mathcal{L}_m &= \bar{v}_L^0 m_D v_R^0 + \frac{1}{2} v_R^{0\top} C M_R v_R^0 + \bar{e}_L^0 m_\ell e_R^0 + \text{H.c.} \\ &= \frac{1}{2} n_L^\top C \mathcal{M}^* n_L + \bar{e}_L^0 m_\ell e_R^0 + \text{H.c.}, \end{aligned} \quad (195)$$

with

$$n_L = \begin{pmatrix} v_L^0 \\ (v_R^0)^c \end{pmatrix}. \quad (196)$$

The full neutrino mass matrix is a 6×6 matrix:

$$\mathcal{M} = \begin{pmatrix} 0 & m \\ m^\top & M_R \end{pmatrix}, \quad (197)$$

diagonalised by

$$V^\top \mathcal{M}^* V = \begin{pmatrix} d & 0 \\ 0 & D \end{pmatrix}, \quad (198)$$

with

$$d \equiv \text{diag}(m1, m2, m3) \quad \text{and} \quad D \equiv \text{diag}(M1, M2, M3). \quad (199)$$

The unitary matrix V can be described by

$$V = \begin{pmatrix} K & G \\ S & T \end{pmatrix}. \quad (200)$$

One can show that:

$$S^\dagger \approx K^\dagger m M_R^{-1}, \quad \text{and} \quad G^\dagger \approx K^\dagger m T^* D^{-1} \approx m D^{-1}, \quad (201)$$

so that one gets the usual seesaw formula

$$-K^\dagger m \frac{1}{M} m^\top K^* = d. \quad (202)$$

The leptonic charged current interactions are:

$$-\frac{g}{\sqrt{2}} (\bar{\ell}_{iL} \gamma_\mu K_{ij} v_{jL} + \bar{\ell}_{iL} \gamma_\mu G_{ij} N_{jL}) W^\mu + \text{H.c..} \quad (203)$$

Let us count the parameters in the leptonic sector. Without loss of generality, one can choose a Weak basis where the charged lepton mass matrix is diagonal, real and also the right-handed neutrino mass matrix is diagonal, real. In this basis, the Yukawa

coupling matrix Y_D entering in the Dirac neutrino mass matrix is an arbitrary complex matrix. Moreover, 3 of 9 phases in Y_D can be eliminated by rephasing. So altogether one has 3 charged lepton masses, 3 diagonal entries in M_R , 9 real parameters in Y_D plus 6 phases in Y_D . This gives a total of 21 parameters.

The *Lepton-Asymmetry* generated through CP violating decays of the Heavy neutrinos,

$$N \longrightarrow \ell + H, \quad (204)$$

within unflavoured Leptogenesis approximation, which do not include flavour effects, is given by

$$A^j = \frac{\sum_i \Gamma_i^j - \bar{\Gamma}_i^j}{\sum_i \Gamma_i^j + \bar{\Gamma}_i^j} \propto \sum_{k \neq j} C_K i \left[(m_D^\dagger m_D)_{jk}^2 \right] \quad (205)$$

where

$$\Gamma_i^j \equiv \Gamma(N_j \longrightarrow \ell_i + H), \quad \bar{\Gamma}_i^j \equiv \Gamma(N_j \longrightarrow \bar{\ell}_i + H). \quad (206)$$

Taking into account the Casas and Ibarra parametrisation [44]:

$$m_D = i U_\nu \sqrt{d} R \sqrt{D}, \quad (207)$$

with R a complex orthogonal matrix, one sees that the combination $m_D^\dagger m_D$ is given by:

$$m_D^\dagger m_D = -\sqrt{D} R^\dagger d R \sqrt{D}, \quad (208)$$

and thus one concludes that leptogenesis is independent of U_ν .

In general, it is not possible to establish a connection between CP asymmetries needed for leptogenesis and CP violation detectable in Neutrino Oscillations. *One may have leptogenesis even if U_ν is real* [45]. The connection may be established with further theoretical assumptions [46–48].

Can one have a WB invariant which is sensitive to the CP violating phases entering in *Unflavoured Leptogenesis*? It is indeed possible. The WB invariant sensitive to the CP violating phases entering in *Unflavoured Leptogenesis* is given by [49]:

$$\begin{aligned} I \equiv i \text{Tr} [h H M_R^* h^* M_R] &= M_1 M_2 (M_2^2 - M_1^2) i(h_{12}^2) \\ &+ M_1 M_3 (M_3^2 - M_1^2) i(h_{13}^2) + M_2 M_3 (M_3^2 - M_2^2) i(h_{23}^2), \end{aligned} \quad (209)$$

where $h \equiv m_D^\dagger m_D$ and $H \equiv M_R^\dagger M_R$.

4 Conclusions

Neutrino Oscillations provide clear evidence for Physics beyond the SM and the discovery of $U_{e3} \neq 0$ opens up the exciting possibility of detecting leptonic Dirac-type CP violation through neutrino oscillations.

Leptogenesis is an attractive framework to generate BAU which can occur in the framework of ν . SM It is difficult to test experimentally, but one should try to find a framework where this is possible.

It is urgent to conceive *feasible experiments* which can measure physical quantities in m_ν beyond the seven quantities mentioned by Glashow et al. Difficult but what looks impossible today, may be possible tomorrow! It would be very nice if some years from now, we have a workshop with a title like:

The leptonic unitarity triangle fit

Acknowledgments We would like to thank the organisers of the Third IDPASC School, specially Carlos Merino, for the very warm hospitality extended to us and for the very nice atmosphere in the School. This work was supported by *Fundação para a Ciência e a Tecnologia* (FCT, Portugal) through the projects CERN/FP/123580/2011, PTDC/FISNUC/ 0548/2012 and CFTP-FCT Unit 777 (PEst-OE/FIS/UI0777/2013) which are partially funded through POCTI (FEDER). The work of D.E.C. was also supported by Associação do Instituto Superior Técnico para a Investigação e Desenvolvimento (IST-ID).

References

1. S. Glashow, Partial symmetries of weak interactions. *Nucl. Phys.* **22**, 579–588 (1961)
2. S. Weinberg, A model of leptons. *Phys. Rev. Lett.* **19**, 1264–1266 (1967)
3. A. Salam, Weak and electromagnetic interactions. *Conf. Proc. C* **680519**, 367–377 (1968)
4. S. Glashow, J. Iliopoulos, L. Maiani, Weak interactions with Lepton-Hadron symmetry. *Phys. Rev. D* **2**, 1285–1292 (1970)
5. M. Gaillard, B.W. Lee, Rare decay modes of the K-mesons in Gauge theories. *Phys. Rev. D* **10**, 897 (1974)
6. S.L. Glashow, S. Weinberg, Natural conservation laws for neutral currents. *Phys. Rev. D* **15**, 1958 (1977)
7. E. Paschos, Diagonal neutral currents. *Phys. Rev. D* **15**, 1966 (1977)
8. G. Branco, L. Lavoura, On the addition of vector like quarks to the standard model. *Nucl. Phys. B* **278**, 738 (1986)
9. F. del Aguila, M. Chase, J. Cortes, Vector like fermion contributions to epsilon-prime. *Nucl. Phys. B* **271**, 61 (1986)
10. G. Branco, T. Morozumi, P. Parada, M. Rebelo, CP asymmetries in B0 decays in the presence of flavor changing neutral currents. *Phys. Rev. D* **48**, 1167–1175 (1993)
11. G. Branco, P. Parada, M. Rebelo, D0—anti-D0 mixing in the presence of isosinglet quarks. *Phys. Rev. D* **52**, 4217–4222 (1995). [arxiv:hep-ph/9501347](https://arxiv.org/abs/hep-ph/9501347)
12. T. Morozumi, T. Satou, M. Rebelo, M. Tanimoto, The top quark mass and flavor mixing in a seesaw model of quark masses. *Phys. Lett. B* **410**, 233–240 (1997). [arxiv:hep-ph/9703249](https://arxiv.org/abs/hep-ph/9703249)
13. G. Barenboim, F. Botella, G. Branco, O. Vives, How sensitive to FCNC can B0 CP asymmetries be? *Phys. Lett. B* **422**, 277–286 (1998). [arxiv:hep-ph/9709369](https://arxiv.org/abs/hep-ph/9709369)

14. G. Barenboim, F. Botella, Delta F=2 effective Lagrangian in theories with vector-like fermions. *Phys. Lett. B* **433**, 385–395 (1998). [arxiv:hep-ph/9708209](https://arxiv.org/abs/hep-ph/9708209)
15. P. Minkowski, mu → e gamma at a rate of one out of 1-billion muon decays? *Phys. Lett. B* **67**, 421 (1977)
16. T. Yanagida, in *Horizontal Gauge Symmetry and Masses of Neutrinos*. Proceedings of the Workshop on the Baryon Number of the Universe and Unified Theories, Tsukuba, Japan, 13–14 Feb 1979
17. M. Gell-Mann, P. Ramond, R. Slansky, *Complex Spinors and Unified Theories*, eds. by P. van Nieuwenhuizen, D.Z. Freedman. Supergravity (North Holland Publishing Co, Amsterdam, 1979), p. 315
18. S.L. Glashow, in *Quarks and Leptons, Quarks and Leptons*, eds. by M. Lévy et al. Cargèse Lectures (Plenum, New York, 1980), p. 687
19. R.N. Mohapatra, G. Senjanovic, Neutrino mass and spontaneous parity nonconservation. *Phys. Rev. Lett.* **44**, 912 (1980)
20. G. Branco, G. Senjanovic, The question of neutrino mass. *Phys. Rev. D* **18**, 1621 (1978)
21. G. Branco, J. Silva-Marcos, NonHermitian Yukawa couplings? *Phys. Lett. B* **331**, 390–394 (1994)
22. G. Branco, D. Emmanuel-Costa, R. Gonzalez Felipe, Texture zeros and weak basis transformations. *Phys. Lett. B* **477**, 147–155 (2000). [arxiv:hep-ph/99111418](https://arxiv.org/abs/hep-ph/99111418)
23. W. Grimus, M. Rebelo, Automorphisms in gauge theories and the definition of CP and P. *Phys. Rep.* **281**, 239–308 (1997). [arxiv:hep-ph/9506272](https://arxiv.org/abs/hep-ph/9506272)
24. G.C. Branco, L. Lavoura, J.P. Silva, CP violation. *Int. Ser. Monogr. Phys.* **103**, 1–536 (1999)
25. J. Bernabéu, G. Branco, M. Gronau, CP restrictions on quark mass matrices. *Phys. Lett. B* **169**, 243–247 (1986)
26. C. Jarlskog, Commutator of the quark mass matrices in the standard electroweak model and a measure of maximal CP violation. *Phys. Rev. Lett.* **55**, 1039 (1985)
27. B. Pontecorvo, Mesonium and antimesonium. *Sov. Phys. JETP* **6**, 429 (1957)
28. B. Pontecorvo, Inverse beta processes and nonconservation of lepton charge. *Sov. Phys. JETP* **7**, 172–173 (1958)
29. Z. Maki, M. Nakagawa, S. Sakata, Remarks on the unified model of elementary particles. *Prog. Theor. Phys.* **28**, 870–880 (1962)
30. G.C. Branco, M. Rebelo, Building the full PMNS matrix from six independent majorana-type phases. *Phys. Rev. D* **79**, 013001 (2009). [arXiv:0809.2799](https://arxiv.org/abs/0809.2799)
31. F. Botella, G. Branco, M. Nebot, M. Rebelo, Unitarity triangles and the search for new physics. *Nucl. Phys. B* **651**, 174–190 (2003). [arxiv:hep-ph/0206133](https://arxiv.org/abs/hep-ph/0206133)
32. P.H. Frampton, S.L. Glashow, D. Marfatia, Zeroes of the neutrino mass matrix. *Phys. Lett. B* **536**, 79–82 (2002). [arxiv:hep-ph/0201008](https://arxiv.org/abs/hep-ph/0201008)
33. G. Branco, R. Gonzalez Felipe, F. Joaquim, T. Yanagida, Removing ambiguities in the neutrino mass matrix. *Phys. Lett. B* **562**, 265–272 (2003). [arxiv:hep-ph/0212341](https://arxiv.org/abs/hep-ph/0212341)
34. G. Branco, L. Lavoura, M. Rebelo, Majorana neutrinos and CP violation in the leptonic sector. *Phys. Lett. B* **180**, 264 (1986)
35. F. Botella, G. Branco, M. Nebot, M. Rebelo, New physics and evidence for a complex CKM. *Nucl. Phys. B* **725**, 155–172 (2005). [arxiv:hep-ph/0502133](https://arxiv.org/abs/hep-ph/0502133)
36. J.P. Silva, L. Wolfenstein, Detecting new physics from CP violating phase measurements in B decays. *Phys. Rev. D* **55**, 5331–5333 (1997). [arxiv:hep-ph/9610208](https://arxiv.org/abs/hep-ph/9610208)
37. L. Bento, G.C. Branco, Generation of a K-M phase from spontaneous CP breaking at a high-energy scale. *Phys. Lett. B* **245**, 599–604 (1990)
38. L. Bento, G.C. Branco, P.A. Parada, A minimal model with natural suppression of strong CP violation. *Phys. Lett. B* **267**, 95–99 (1991)
39. G. Branco, P. Parada, M. Rebelo, A common origin for all CP violations. [arxiv:hep-ph/0307119](https://arxiv.org/abs/hep-ph/0307119)
40. F. Botella, G. Branco, M. Nebot, The hunt for new physics in the flavour sector with up vector-like quarks. *JHEP* **1212**, 040 (2012). [arXiv:1207.4440](https://arxiv.org/abs/1207.4440)
41. A. Sakharov, Violation of CP invariance, c asymmetry, and Baryon asymmetry of the universe. *Pisma Zh. Eksp. Teor. Fiz.* **5**, 32–35 (1967)

42. S. Davidson, E. Nardi, Y. Nir, Leptogenesis. *Phys. Rep.* **466**, 105–177 (2008). [arXiv:0802.2962](https://arxiv.org/abs/0802.2962)
43. G. Branco, R.G. Felipe, F. Joaquim, Leptonic CP violation. *Rev. Mod. Phys.* **84**, 515–565 (2012). [arXiv:1111.5332](https://arxiv.org/abs/1111.5332)
44. J. Casas, A. Ibarra, Oscillating neutrinos and muon $\rightarrow e, \gamma$. *Nucl. Phys. B* **618**, 171–204 (2001). [arxiv:hep-ph/0103065](https://arxiv.org/abs/hep-ph/0103065)
45. M. Rebelo, Leptogenesis without CP violation at low-energies. *Phys. Rev. D* **67**, 013008 (2003). [arxiv:hep-ph/0207236](https://arxiv.org/abs/hep-ph/0207236)
46. G. Branco, R. Gonzalez Felipe, F. Joaquim, I. Masina, M. Rebelo et al., Minimal scenarios for leptogenesis and CP violation. *Phys. Rev. D* **67**, 073025 (2003). [arxiv:hep-ph/0211001](https://arxiv.org/abs/hep-ph/0211001)
47. P. Frampton, S. Glashow, T. Yanagida, Cosmological sign of neutrino CP violation. *Phys. Lett. B* **548**, 119–121 (2002). [arxiv:hep-ph/0208157](https://arxiv.org/abs/hep-ph/0208157)
48. G. Branco, R. Gonzalez Felipe, F. Joaquim, M. Rebelo, Leptogenesis, CP violation and neutrino data: what can we learn? *Nucl. Phys. B* **640**, 202–232 (2002). [arxiv:hep-ph/0202030](https://arxiv.org/abs/hep-ph/0202030)
49. G.C. Branco, T. Morozumi, B. Nobre, M. Rebelo, A bridge between CP violation at low-energies and leptogenesis. *Nucl. Phys. B* **617**, 475–492 (2001). [arxiv:hep-ph/0107164](https://arxiv.org/abs/hep-ph/0107164)



Professor María José Herrero during one of the lectures of her course The Higgs System in and beyond the Standard Model at the Third IDPASC School (picture by C. Merino)

The Higgs System in and Beyond the Standard Model

Maria J. Herrero

Abstract After the discovery of the Higgs boson particle on the 4th of July of 2012 at the Large Hadron Collider, sited at the European CERN laboratory, we are entering in a fascinating period for Particle Physics where both theorists and experimentalists are devoted to fully understand the features of this new particle and the possible consequences for High Energy Physics of the Higgs system both within and beyond the Standard Model of fundamental particle interactions. This paper is a summary of the lectures given at the third IDPASC school (Santiago de Compostela, Feb. 2013, Spain) addressed to PhD students, and contains a short introduction to the main basic aspects of the Higgs boson particle in and beyond the Standard Model.

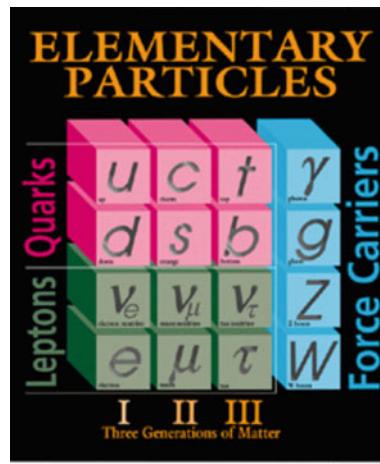
1 Introduction

The Standard Model (SM) [1–5] describes with unprecedent precision (0.1 %) the properties of all known elementary particles, Leptons and Quarks, and their fundamental interactions, electromagnetic, strong and weak. Gravity is not included in the SM. The complete gauge symmetry group of the SM is $SU(3)_C \times SU(2)_L \times U(1)_Y$, with $SU(3)_C$ being the symmetry group of the strong interactions and $SU(2)_L \times U(1)_Y$ the symmetry group of the electroweak interactions.

The different elementary particles described by the SM are collected in the following figure and include: (1) the three fermion families with the three charged leptons, the electron e the muon μ and the tau τ , the three neutral leptons, i.e. the neutrinos ν_e , ν_μ and ν_τ , the three up-type quarks u (up), c (charm) and t (top), the three down-type quarks d (down), s (strange) and b (bottom); (2) the force carriers: the photon γ , mediator of the electromagnetic interactions, the eight gluons g_a ($a = 1, \dots, 8$), mediators of the strong interactions, and the three weak bosons, mediators of the weak interactions, the neutral Z boson and the two charged W^\pm bosons.

M.J. Herrero()

Departamento de Física Teórica and Instituto de Física Teórica, IFT-UAM/CSIC,
Universidad Autónoma de Madrid, Cantoblanco, Madrid, Spain
e-mail: Maria.Herrero@uam.es



The particle content of the Standard Model

All the above particles have been experimentally seen and their properties have been measured in many cases with very high precision [6]. The elementary particles that are the constituents of matter are of two types, leptons and quarks, and they are fermions with spin equal to 1/2. With respect to the mediators of the three interactions within the SM, i.e. the gauge particles, they are bosons with spin equal to 1, and their properties have also been tested in the experiments. One of the most important properties of these gauge bosons is their mass. The carriers of electromagnetic (photon) and strong interactions (gluons) are massless gauge bosons. But the carriers of weak interactions, W^\pm and Z , are massive. The present measurements give: $M_W^{\text{exp}} = 80.385 \pm 0.015$ GeV and $M_Z^{\text{exp}} = 91.1876 \pm 0.0021$ GeV.

The fact that these weak bosons, W^\pm and Z , have non zero masses leads to a problem in Gauge Quantum Field Theory which can be read as how to reconcile gauge invariance and massive gauge bosons. The SM, as any other Gauge Quantum Field Theory is built under the construction principle of gauge invariance where the exchanged field quanta with spin one defines the gauge particle and it must be massless in order to preserve this gauge invariance. Therefore, the observed weak boson masses must be explained in a different way. Within the SM, this way is the Higgs Mechanism [7–11] that will be described in these lectures and that leads to the prediction of a new massive scalar particle the Higgs boson particle. The present consensus in the High Energy Physics Community points towards the interpretation that the recently discovered scalar particle at the LHC [12, 13] with a mass between 125 and 126 GeV is indeed this Higgs boson, predicted in the SM. The most recent measurements by the ATLAS and CMS collaborations set $M_H^{\text{ATLAS}} = 125.5 \pm 0.6$ GeV [14] and $M_H^{\text{CMS}} = 125.7 \pm 0.4$ GeV [15], respectively. They also show that the most probable J^P quantum numbers for this discovered particle are 0^+ , and conclude that the measured Higgs couplings to the other SM particles are in agreement so far with the values predicted in the SM. But, although all these first LHC data are really

encouraging, there is still a long way to fully check the SM Higgs boson hypothesis. These commented properties above and many other features of this new scalar particle will be measured in the future with much higher precision than at present, and we hopefully will be able to disentangle finally which particle is really this one, the SM Higgs boson or something else. This is really a fascinating period in the History of Particle Physics.

This paper is organized in two main blocks, corresponding to the two given lectures:

Lecture 1: The Higgs boson in the Standard Model

The building of the Electroweak Theory

Electroweak Symmetry Breaking

The spectra of the SM and the particle masses

SM Higgs boson couplings

SM Higgs boson decays and production at LHC

Other interesting properties of the SM Higgs system

Lecture 2: Some avenues beyond Standard Model Higgs

Motivations for looking beyond the Standard Model

The hierarchy problem of the SM Higgs sector

Two main avenues to solve the hierarchy problem

Supersymmetry

Compositeness

Electroweak Chiral Lagrangians

2 The Building of the Electroweak Theory

In the following we shortly remind the basics of the gauge principle, using QED as an illustrative example, and then apply it to the Electroweak Theory.

The gauge principle:

In order to get a Lagrangian that is invariant under local (gauge) transformations, massless gauge fields A_μ must be introduced with specific interactions with matter. The concrete prescription is provided by the covariant derivative. Number of gauge bosons = Number of symmetries = Number of generators of the symmetry group.

In practice, one follows three steps: (1) Start with the Lagrangian for propagating fermion fields without interactions, i.e., for free fields. (2) Replace the usual derivative by the covariant derivative. (3) Add the proper invariant kinetic terms for the gauge fields, such that they can propagate.

QED as an example:

Let use Ψ to describe the field of a fermion with electric charge Q (in units of e , the electron charge) and mass m . The associated free Lagrangian is:

$$\mathcal{L}_{\text{free}} = \bar{\Psi}(i\cancel{\partial} - m)\Psi$$

where, $\not{d} \equiv \partial_\mu \gamma^\mu$, γ^μ = Dirac matrices.

The corresponding equation of motion for Ψ is the Dirac equation:

$$(i\not{d} - m)\Psi = 0$$

Then we replace the normal derivative by the covariant derivative that includes the gauge field, here denoted by A_μ which defines the photon particle (γ in the figure),

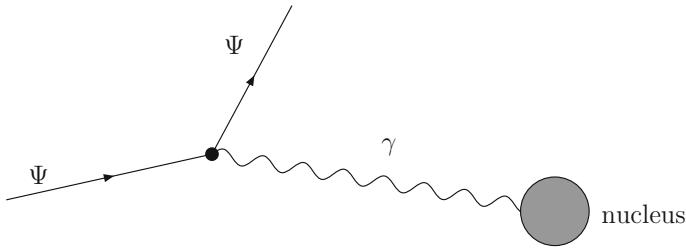
$$\partial_\mu \Psi \rightarrow D_\mu \Psi \equiv (\partial_\mu - ieQA_\mu)\Psi ;$$

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

With this covariant derivative we then build the QED Lagrangian:

$$\Rightarrow \mathcal{L}_{\text{QED}} = \bar{\Psi}(i\not{d} - m)\Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field stress tensor.



One can check easily that \mathcal{L}_{QED} is invariant under $U(1)$ Gauge transformations, with one single generator give by Q :

$$\Psi \rightarrow e^{ieQ\theta(x)}\Psi ; A_\mu \rightarrow A_\mu - \frac{1}{e}\partial_\mu\theta(x)$$

Notice that a mass term for the photon of the type $m^2 A^\mu A_\mu$ is not $U(1)$ gauge invariant, therefore in QED the gauge invariance principle implies that the photon is massless, which is in total agreement with data. However, this is not the case of the W and Z electroweak gauge bosons and the immediate questions arise: why are they massive? How do they get their masses?...

The gauge invariance in the Electroweak Theory:

The Electroweak Theory (EW) refers to the part of the SM that describes together the electromagnetic and weak interactions within the same framework of a Gauge Quantum Field Theory based on the gauge principle invariance of the electroweak interactions.

The gauge symmetry group of the Electroweak Theory is $SU(2)_L \times U(1)_Y$, with 4 generators. $SU(2)_L$ is the weak isospin group which is non abelian, and has 3 generators $T_{1,2,3} = \sigma_{1,2,3}/2$, with $\sigma_{1,2,3}$ being the Pauli matrices. $U(1)_Y$ is the weak hypercharge group which is abelian and has 1 generator $Y/2$. The electromagnetic

group appears as a subgroup of the electroweak group, $U(1)_{\text{em}} \subset SU(2)_L \times U(1)_Y$; and the corresponding generator is a combination of the third component of the weak isospin and the weak hypercharge, $Q = T_3 + Y/2$.

The elementary particles of the SM, Quarks and Leptons, transform as:

(1) Under $SU(2)_L$: $\Psi_L \rightarrow e^{i\frac{\tilde{\phi}}{2}\tilde{\theta}(x)}\Psi_L$, doublets ; $\Psi_R \rightarrow \Psi_R$, singlets

(2) Under $U(1)_Y$: $\Psi \rightarrow e^{i\frac{Y}{2}\beta(x)}\Psi$.

Where $\Psi_L = (1 - \gamma_5)/2$ and $\Psi_R = (1 + \gamma_5)/2$ refer to the two possible chiral projections, for left and right handed chiralities of the fermion Ψ , respectively.

The corresponding quantum numbers for the first generation of quarks and leptons are collected in the tables.

Lepton	T	T_3	Q	Y
ν_L	$\frac{1}{2}$	$\frac{1}{2}$	0	-1
e_L	$\frac{1}{2}$	$-\frac{1}{2}$	-1	-1
e_R	0	0	-1	-2

Quark	T	T_3	Q	Y
u_L	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$
d_L	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$\frac{1}{3}$
u_R	0	0	$\frac{2}{3}$	$\frac{4}{3}$
d_R	0	0	$-\frac{1}{3}$	$-\frac{2}{3}$

The particle content and the Lagrangian of the Electroweak Theory

The particle content of the SM is summarized schematically in the following:

Matter particles
1st family: $\begin{pmatrix} v_e \\ e^- \end{pmatrix}_L, e_R^-, \begin{pmatrix} u \\ d \end{pmatrix}_L, u_R, d_R$
2nd family: $\begin{pmatrix} v_\mu \\ \mu^- \end{pmatrix}_L, \mu_R^-, \begin{pmatrix} c \\ s \end{pmatrix}_L, c_R, s_R$
3rd family: $\begin{pmatrix} v_\tau \\ \tau^- \end{pmatrix}_L, \tau_R^-, \begin{pmatrix} t \\ b \end{pmatrix}_L, t_R, b_R$

Gauge particles
$SU(2)_L$: 3 generators T_i , 3 gauge bosons W_i^μ
$U(1)_Y$: 1 generator $\frac{Y}{2}$, 1 gauge boson B^μ
$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g e^{ijk} W_\mu^j W_\nu^k$
$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$
Physical EW bosons
$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp i W_\mu^2)$
$Z_\mu = \cos \theta_W W_\mu^3 - \sin \theta_W B_\mu$
$A_\mu = \sin \theta_W W_\mu^3 + \cos \theta_W B_\mu$

The electroweak interactions are introduced via the gauge principle, as in the previous example, by the replacement in the free Lagrangian of the normal derivative by the corresponding covariant derivative:

$$\partial_\mu \Psi \rightarrow D_\mu \Psi = (\partial_\mu - ig \vec{T} \cdot \vec{W}_\mu - ig' \frac{Y}{2} B_\mu) \Psi$$

where g is the $SU(2)_L$ gauge coupling and g' is the $U(1)_Y$ gauge coupling. The relation between the electromagnetic coupling e and these two couplings g and g' is a consequence of $U(1)_{\text{em}}$ being a subgroup of $SU(2)_L \times U(1)_Y$:

$$g = \frac{e}{\sin \theta_W} ; \quad g' = \frac{e}{\cos \theta_W}$$

where θ_W is the weak angle that defines the physical neutral gauge bosons Z_μ and A_μ in terms of the EW interaction eigenstates W_μ^3 and B_μ .

The Lagrangian of the Electroweak Theory is then given by:

$$\mathcal{L}_{\text{EW}} = \sum_{\Psi} i \bar{\Psi} \gamma^\mu D_\mu \Psi - \frac{1}{4} W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}$$

where the sum runs over all the fermions of the SM: quarks and leptons.

This Lagrangian \mathcal{L}_{EW} is invariant under $SU(2)_L \times U(1)_Y$ gauge transformations. However, notice that it doesn't contain any mass term for any of the SM fields. It can be easily checked that a mass term like $m \bar{\Psi} \Psi = m(\bar{\Psi}_L \Psi_R + \bar{\Psi}_R \Psi_L)$ and a mass term like $M_W^2 W_\mu W^\mu$ are not $SU(2)_L$ invariant. Therefore \mathcal{L}_{EW} does not describe properly yet the masses for the fermions nor the weak gauge bosons and a new piece in the SM Lagrangian must be introduced to generate the particle masses which is directly related with some sort of breaking of the Electroweak symmetry. Then the full SM Lagrangian will be finally built from the two terms:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{EWSB}}$$

where $\mathcal{L}_{\text{EWSB}}$ refers to the Lagrangian for the Electroweak Symmetry Breaking that will be described next.

3 Electroweak Symmetry Breaking

In this section we shortly summarize the basics of the Electroweak Symmetry Breaking in the SM and set the steps to follow for the building of $\mathcal{L}_{\text{EWSB}}$.

The most relevant aspects of the Electroweak Symmetry Breaking can be organized in three main points:

1. The Phenomenon of Spontaneous Symmetry Breaking
2. Spontaneous Symmetry Breaking: the Goldstone Theorem
3. Electroweak Symmetry Breaking: the Higgs Mechanism

These are important to understand separately and will be commented in the following.

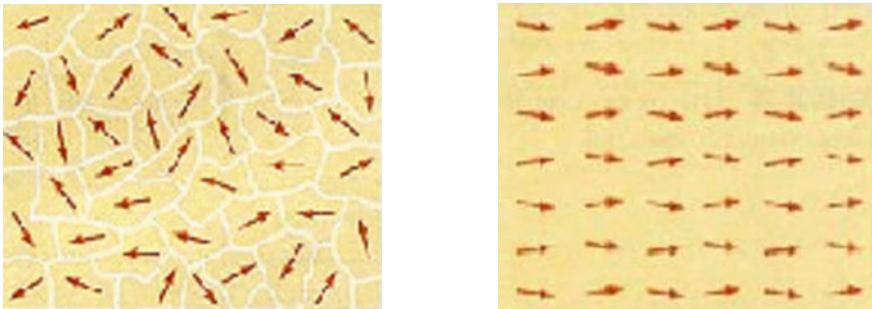
The Phenomenon of Spontaneous Symmetry Breaking

A simple definition:

A physical system has a symmetry that is spontaneously broken if the interactions governing the dynamics of the system possess such a symmetry but the ground state of this system does not.

A simple example:

Let us consider an infinitely extended ferromagnet at temperature T close to the Curie temperature T_C . The system is described by an infinite set of elementary spins and their interactions (given by the Lagrangian) are rotational invariant. The ground state of this system presents two different situations depending on the value of T being above or below the Curie temperature. These two situations are schematically described below:



With regard the mathematical description of this behavior of the ground state in the extended ferromagnet example, there is a very simple theoretical framework that describes successfully this phenomenon of spontaneous symmetry breaking, the Theory of Ginzburg-Landau.

The Theory of Ginzburg-Landau (1950)

In this theory, for T near T_C , the free energy density $u(\vec{M})$ for small \vec{M} is given by:

$$u(\vec{M}) = (\partial_i \vec{M})(\partial_i \vec{M}) + V(\vec{M}) ; \quad i = 1, 2, 3$$

where the potential is:

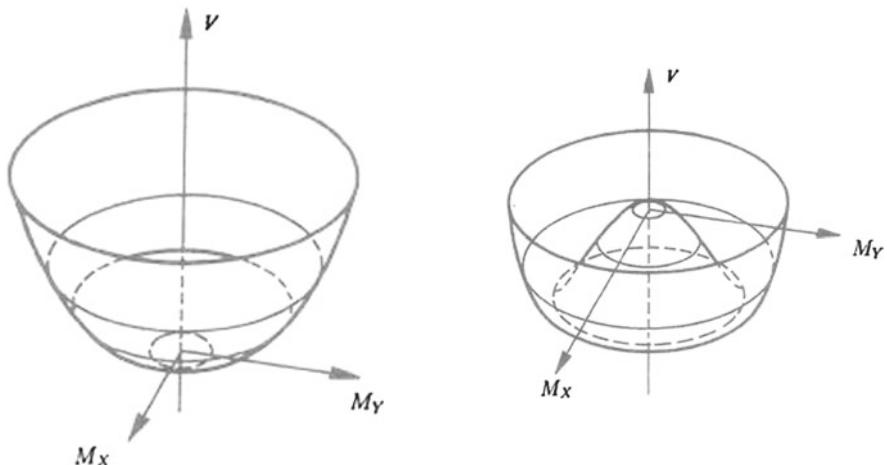
$$V(\vec{M}) = \alpha_1(T - T_C)(\vec{M} \cdot \vec{M}) + \alpha_2(\vec{M} \cdot \vec{M})^2 ; \quad \alpha_1, \alpha_2 > 0$$

Notice that in the drawings to simplify we have chosen a two dimensional (instead of three) magnetization vector $\vec{M} = (M_X, M_Y)$.

The magnetization of the ground state is obtained from the condition of extremum:

$$\frac{\delta V(\vec{M})}{\delta M_i} = 0 \Rightarrow \vec{M} \cdot [\alpha_1(T - T_C) + 2\alpha_2(\vec{M} \cdot \vec{M})] = 0$$

This leads to two solutions for \vec{M} , depending on the value of T which correspond respectively to the previous described situations I and II. These two qualitative different solutions describe the so-called symmetric and non symmetric phases of the system.



Goldstone Theorem (Nambu, Goldstone, 1960–1962)

The Goldstone Theorem applies to Quantum Field Theories (QFT) with Spontaneous Symmetry Breaking (SSB).

SSB stated in simple words:

In QFT, a system is said to have a symmetry that is spontaneously broken if the Lagrangian describing the dynamics of the system is invariant under this symmetry transformation, but the vacuum of the theory is not. The vacuum $|0\rangle$ is the state where the Hamiltonian expectation value $\langle 0|H|0\rangle$ is minimum.

Goldstone Theorem stated in simple words:

If a QFT has a global symmetry of the Lagrangian which is not a symmetry of the vacuum \Rightarrow there must exist one massless boson, scalar or pseudoscalar, associated to each generator which does not annihilate the vacuum and having its same quantum numbers. These modes are referred to as Nambu-Goldstone bosons or simply as Goldstone bosons (GBs).

Notice that:

$$U|0\rangle = |0\rangle \text{ with } U = \exp(i\epsilon^a Q^a) \Rightarrow Q^a|0\rangle = 0 \quad \forall a$$

and:

$$U|0\rangle \neq |0\rangle \text{ with } U = \exp(i\epsilon^a Q^a) \Rightarrow \exists Q^a / Q^a|0\rangle \neq 0$$

QCD as an example

One illustrative example of the phenomenon of SSB and the consequences of the Goldstone Theorem is provided by the well known case of QCD with two flavors, u, d , where there is a global symmetry, the chiral symmetry, that is known to be spontaneously broken. We comment briefly on this next.

Let us start with the QCD Lagrangian, given in terms of quarks, $q_i \equiv q (i = 1, 2, 3)$, and gluons, $g_\mu^a \equiv A_\mu^a (a = 1, \dots, 8)$, by:

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{2} \text{Tr} G^{\mu\nu} G_{\mu\nu} + \sum_{u,d} (i\bar{q}\gamma^\mu D_\mu q - m_q \bar{q}q)$$

where,

$$\begin{aligned} G_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu - ig_s [A_\mu, A_\nu] \\ D_\mu q &= (\partial_\mu - ig_s A_\mu) q \\ A_\mu &= \sum_{a=1}^8 \frac{1}{2} A_\mu^a \lambda_a \end{aligned}$$

The generators of the $SU(3)_C$ color group are the eight $\lambda_a/2$ matrices, with λ_a being the well known 3×3 Gell-Mann matrices. g_s is the strong coupling constant, and m_q is the mass of the quark q .

In addition to the $SU(3)_C$ gauge symmetry of QCD, that is the responsible for the strong interactions among quarks and gluons, \mathcal{L}_{QCD} has an extra global symmetry for the case of massless quarks, $m_{u,d} = 0$:

$$SU(2)_L \times SU(2)_R \equiv \text{Chiral Symmetry}$$

defined by:

$$\Psi_L \rightarrow \Psi'_L = U_L \Psi_L = \exp(i\alpha_L^a Q_L^a) \Psi_L ; \quad Q_L^{1,2,3} \text{ generators of } SU(2)_L$$

$$\Psi_R \rightarrow \Psi'_R = U_R \Psi_R = \exp(i\alpha_R^a Q_R^a) \Psi_R ; \quad Q_R^{1,2,3} \text{ generators of } SU(2)_R$$

where,

$$\Psi = \begin{pmatrix} u \\ d \end{pmatrix} ; \quad \Psi_L = \frac{1}{2}(1 - \gamma_5)\Psi ; \quad \Psi_R = \frac{1}{2}(1 + \gamma_5)\Psi$$

When the $m_{u,d} \neq 0$ terms are included into \mathcal{L}_{QCD} then the chiral symmetry is explicitly broken, but not much since these quark masses are small. Then, the chiral symmetry is not an exact global symmetry but it is a very good approximate symmetry of QCD.

On the other hand, it happens that this chiral symmetry is not a symmetry of the QCD vacuum, therefore it must be a spontaneously broken symmetry. Indeed, this chiral symmetry is spontaneously broken down to the isospin symmetry, given by the subgroup $SU(2)_V$ of the chiral group, $SU(2)_L \times SU(2)_R$:

$$\begin{aligned} SU(2)_L \times SU(2)_R &= SU(2)_V \times SU(2)_A \rightarrow SU(2)_V ; \\ SU(2)_V &= SU(2)_{R+L} ; \quad SU(2)_A = SU(2)_{R-L} \end{aligned}$$

The SSB phenomenon occurs here because \mathcal{L}_{QCD} is invariant under $SU(2)_L \times SU(2)_R$ but the QCD vacuum is NOT fully $SU(2)_L \times SU(2)_R$ invariant. It is only invariant under the subgroup $SU(2)_V \subset SU(2)_L \times SU(2)_R$. Schematically we write

this SSB as:

$$SU(2)_L \times SU(2)_R \rightarrow SU(2)_V$$

But, how do we know from experiment that the QCD vacuum is not $SU(2)_L \times SU(2)_R$ symmetric? The demonstration of this fact goes by starting with the ‘negative’ hypothesis, i.e assuming a QCD symmetric vacuum, and ending in an acceptable conclusion.

Let us assume that $|0\rangle$ is chiral invariant \Rightarrow

$$U_L|0\rangle = |0\rangle ; U_R|0\rangle = |0\rangle \Rightarrow Q_L^a|0\rangle = 0 ; Q_R^a|0\rangle = 0$$

Let $|\Psi\rangle$ be an eigenstate of the Hamiltonian and parity operator such that:

$$H|\Psi\rangle = E|\Psi\rangle ; P|\Psi\rangle = |\Psi\rangle$$

Then, from the two previous assumptions, one finds a new eigenstate $|\Psi'\rangle$ of the Hamiltonian with the same eigenvalue as $|\Psi\rangle$ but with opposite parity:

$$\exists|\Psi'\rangle = \frac{1}{\sqrt{2}}(Q_R^a - Q_L^a)|\Psi\rangle / H|\Psi'\rangle = E|\Psi'\rangle ; P|\Psi'\rangle = -|\Psi'\rangle$$

But, it turns out that *there are not such parity doublets in the hadronic spectrum* \Rightarrow $SU(2)_A$ is NOT a symmetry of the vacuum, or equivalently, $Q_A^a|0\rangle \neq 0(a = 1, 2, 3)$. \Rightarrow *chiral symmetry must be spontaneously broken to the reduced symmetry of the vacuum, $SU(2)_V$.*

Now, according to Goldstone Theorem, and as a consequence of the previous breaking, there must exist one massless Goldstone boson, scalar or pseudoscalar, associated to each generator which does not annihilate the vacuum and having its same quantum numbers. More specifically, the spontaneous breaking of the chiral symmetry in QCD, implies the existence of three massless Goldstone bosons:

$$SU(2)_L \times SU(2)_R \rightarrow SU(2)_V ; \text{ with, } Q_A^a|0\rangle \neq 0(a = 1, 2, 3)$$

$$\Rightarrow \exists 3 \text{ massless GBs, pseudoscalars, } \pi^a(x) \quad a = 1, 2, 3.$$

A very important feature for the phenomenology of low energy QCD is that these three GBs are identified with the physical pions. More specifically, their combinations: $\pi^+ = (\pi^1 - i\pi^2)/\sqrt{2}$, $\pi^- = (\pi^1 + i\pi^2)/\sqrt{2}$ and $\pi^0 = \pi^3$.

Since, in Nature, $m_\pi \neq 0 \Rightarrow$ chiral symmetry is explicitly broken, and the pions are pseudo-GB. But the important outcome is that the hierarchy $m_\pi << m_{\text{hadrons}}$ is explained.

The dynamics of pion interactions is well described by the so-called Chiral Lagrangian of QCD and the associated Effective Quantum Field Theory called Chiral Perturbation Theory (ChPT). We will come back to the subject of Chiral Lagrangians

in the next lecture where we will comment on some applications of these type of effective Lagrangians for beyond the Standard Model Physics.

The Higgs Mechanism:

The Goldstone Theorem is for theories with spontaneously broken global symmetries but does not hold for gauge theories. When a spontaneous symmetry breaking takes place in a gauge theory, the so-called Higgs Mechanism operates. As will be seen in the following, the Higgs Mechanism when applied to the case of the SM leads to the prediction of a new scalar particle, the so-called Higgs boson particle, whose experimental discovery by the collaborations ATLAS and CMS at the Large Hadron Collider (LHC), placed at the laboratory CERN, close to Geneva, was announced in an international open web-conference on the 4th of July 2012. Recently, on the 8th of October 2013, Peter Higgs and Francois Englert have received the Physics Nobel Prize 2013 for the proposal of this mass generation mechanism and for the prediction of the Higgs boson particle.

In the historical development of the guiding ideas that ended up with the final Higgs Mechanism there were indeed many authors involved, including: [7, 8, 10, 11]. Many of these contributions were inspired in previous works within Solid State Physics, including those by [16]. See also the works by [17] where the generation of mass for gauge fields was already mentioned. See also the BCS Theory of Superconductivity, the existence of Cooper pairs and the absence of massless GBs in presence of electromagnetic interactions which can be found in the works by [18].

How to generate mass for gauge bosons in gauge theories (in simple words):

When a spontaneous symmetry breaking takes place in a gauge theory the would-be Goldstone bosons associated to the global symmetry breaking do not manifest explicitly in the physical spectrum but instead they ‘combine’ with the massless gauge bosons and as result, once the spectrum of the theory is built up on the non-symmetric vacuum, there appear massive vector bosons. The number of vector bosons that acquire a mass is precisely equal to the number of these would-be-Goldstone bosons, which in turn are equal to the number of symmetries that the vacuum has lost.

Before going to the SM case, we illustrate first the Higgs Mechanism with one very simple example.

An illustrative example: U(1) gauge symmetry breaking:

Consider the simplest case of a gauge theory based in a $U(1)$ gauge symmetry, with one complex scalar $\Phi = \frac{1}{\sqrt{2}}(\Phi_1 + i\Phi_2)$, one gauge boson A_μ , and a potential of Ginzburg-Landau type. The Lagrangian for this $U(1)$ gauge theory is:

$$\mathcal{L} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - V(\Phi), \text{ with } D_\mu \Phi = (\partial_\mu - igA_\mu)\Phi ; F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \text{ and } V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 ; \lambda > 0.$$

\mathcal{L} is invariant under $U(1)$ gauge transformations given by:

$$\Phi \rightarrow e^{-i\alpha(x)}\Phi ; D_\mu \Phi \rightarrow e^{-i\alpha(x)}D_\mu \Phi ; e^{-i\alpha(x)} \subset U(1)$$

$$A_\mu \rightarrow A_\mu - \frac{1}{g}\partial_\mu \alpha(x)$$

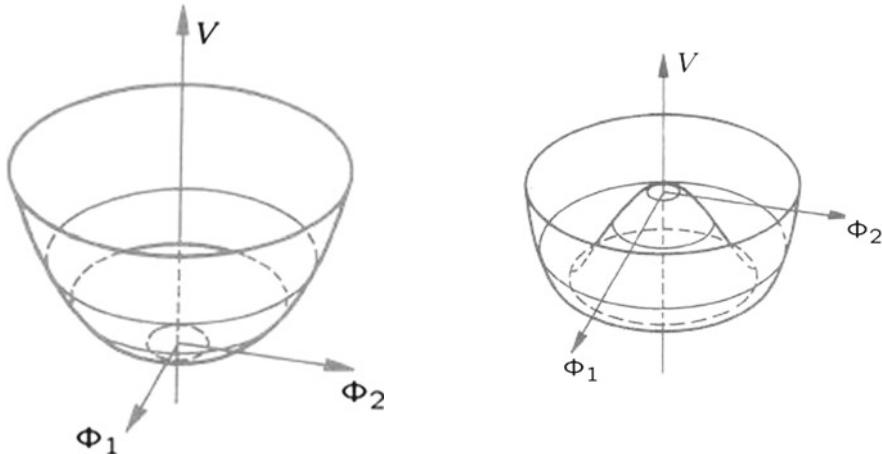
It is interesting to compare $V(\Phi)$ with the previous ferromagnet case:

$$V(\vec{M}) = \alpha_1(T - T_C)(\vec{M} \cdot \vec{M}) + \alpha_2(\vec{M} \cdot \vec{M})^2 ; \alpha_1, \alpha_2 > 0$$

All said previously applies now with the replacements: $(M_X, M_Y) \rightarrow \frac{1}{\sqrt{2}}(\Phi_1 + i\Phi_2)$

$$\alpha_1(T - T_C) \rightarrow \mu^2; \alpha_2 \rightarrow \lambda; \tilde{M}_{\text{ground state}} \rightarrow <0|\Phi|0> \equiv <\Phi>$$

Then, in this case one similarly finds two different situations, but now corresponding to having either $\mu^2 > 0$ or $\mu^2 < 0$. These two situations describe the symmetric phase and the non-symmetric phase of the $U(1)$ gauge theory, as summarized schematically in the following:



At this point, there are two important features that are worth to emphasize:

- (1) *the choice of a particular vacuum (complex phase) is what generates the spontaneous breaking of $U(1)$,*
- (2) *the building of the spectra on top of this non-invariant vacuum (minimum) is what generates the gauge boson mass.*

The first point is clear, since once a particular complex phase has been chosen to describe the vacuum, this is not any more invariant under a $U(1)$ transformation which precisely rotates this phase and would change the starting vacuum into another one with a different phase. Let us see now the second point in more detail.

Let us choose first a particular vacuum configuration, for instance, let us take a real one:

$$|\langle \Phi \rangle| = \sqrt{\frac{-\mu^2}{2\lambda}} \neq 0; \arg \langle \Phi \rangle = 0 \Rightarrow \langle \Phi_1 \rangle = \sqrt{\frac{-\mu^2}{\lambda}} = v, \langle \Phi_2 \rangle = 0$$

Then, we change coordinates to new fields (\equiv shifting the origin):

$$\Phi'_1 \equiv \Phi_1 - v; \Phi'_2 \equiv \Phi_2 \text{ such that } \langle \Phi'_1 \rangle = 0; \langle \Phi'_2 \rangle = 0$$

Next, write everything in terms of these new $\Phi'_{1,2}$ fields:

$$(D_\mu \Phi)^\dagger (D_\mu \Phi) = \left((\partial_\mu + igA_\mu) \frac{1}{\sqrt{2}} (\Phi_1 - i\Phi_2) \right) \left((\partial_\mu - igA_\mu) \frac{1}{\sqrt{2}} (\Phi_1 + i\Phi_2) \right) = \dots$$

$$\frac{1}{2} (\partial_\mu \Phi'_1 + gA_\mu \Phi'_2)^2 + \frac{1}{2} (\partial_\mu \Phi'_2 - gA_\mu \Phi'_1)^2 - gvA^\mu (\partial_\mu \Phi'_2 + gA_\mu \Phi'_1) + \frac{1}{2} g^2 v^2 A_\mu A^\mu$$

And we see that a mass term for A_μ has appeared, i.e. the last term above. But it is not the physical basis yet since there is a (nonphysical) mixing term $\sim gvA^\mu \partial_\mu \Phi'_2$

which ‘combines’ the gauge boson and the scalar fields. In order to find the physical states this mixing term has to be removed. It is convenient then to first choose some proper coordinates: for instance, let us take ‘polar’ coordinates to describe ‘small oscillations’ around vacuum configuration:

$$\Phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x))e^{i\frac{\xi(x)}{v}}$$

Second, let us choose the proper gauge, i.e., make a gauge transformation to the unitary gauge (by fixing the gauge parameter to $\alpha(x) = \frac{\xi(x)}{v}$) where the unwanted mixing terms do not appear:

$$\begin{aligned}\Phi(x) &\rightarrow e^{-i\frac{\xi(x)}{v}}\Phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x)) \\ A_\mu(x) &\rightarrow A_\mu(x) - \frac{1}{gv}\partial_\mu\xi(x) \equiv B_\mu(x)\end{aligned}$$

Finally, we write the Lagrangian in terms of the new fields B_μ and η :

$$\begin{aligned}\mathcal{L} = \frac{1}{2}(\partial_\mu\eta)^2 + \mu^2\eta^2 - \frac{1}{4}(\partial_\mu B_\nu - \partial_\nu B_\mu)^2 + \frac{1}{2}(gv)^2B_\mu B^\mu \\ + \frac{1}{2}g^2B_\mu B^\mu\eta(2v + \eta) - \lambda v\eta^3 - \frac{1}{4}\lambda\eta^4\end{aligned}$$

And we see clearly that these new fields, which are now physical, describe a massive gauge boson particle B_μ with spin 1 and mass $M_{B_\mu} = gv$, and a massive scalar particle η with spin 0 and mass $m_\eta = \sqrt{2}|\mu|$. Notice also that the would-be-Goldstone boson in this example is the ξ field and it has disappeared from the spectrum. There is one symmetry of the Lagrangian that is not preserved by the vacuum and as a consequence there is one gauge boson getting mass. There is also one remaining scalar particle in the physical spectrum, the η particle that is the Higgs particle of this example.

The ‘nice’ properties of the Higgs Mechanism:

Here we collect some of the general properties of the Higgs Mechanism:

- The gauge symmetry of the interactions (i.e. of \mathcal{L}) is preserved
- The renormalizability of the massless gauge theories is preserved
- The total number of polarization degrees is preserved
For instance, in the previous $U(1)$ case:
Before SSB: total polarization degrees = 4 = (2 of A_μ) + (2 of Φ)
After SSB: total polarization degrees = 4 = (3 of B_μ) + (1 of η)
- The nonphysical fields (i.e. the would-be-GBs) have disappeared from the spectrum. In the previous $U(1)$ case: $\xi(x)$
- The number of gauge bosons getting a mass = number of would-be-GBs = number of symmetries of \mathcal{L} that are not symmetries of the vacuum. In the previous $U(1)$ case, this number is 1.

- The would-be-GBs combine with the massless gauge bosons to give them a mass. This ‘combine’ in the U(1) example occurs indeed due to the mixing term $\sim gvA^\mu\partial_\mu\Phi'_2$.
- Comment: The Higgs mechanism does not necessarily imply the existence of a Higgs particle. It appears JUST when required by the polarization degrees preservation property.

The Higgs Mechanism applied to the Standard Model:

We want to generate masses for 3 gauge fields: Z , W^+ and W^- , but we want to keep the photon γ massless.

Strategy: Introduce (ad hoc) a new scalar field, Φ , and a potential of Ginzburg-Landau type, $V(\Phi)$ that make the job. Then one requires the following properties to this Φ :

- ⇒ It must provide the 3 needed polarization degrees to play the role of the would-be-GBs.
- ⇒ It must have non-zero $SU(2)_L \times U(1)_Y$ quantum numbers, such that the vacuum is not invariant under the complete symmetry, but just invariant under the subgroup $U(1)_{\text{em}}$.
- ⇒ The field component in Φ acquiring a vev must be electrically neutral to preserve $U(1)_{\text{em}}$.

Within the SM these Φ and $V(\Phi)$ are chosen to be the simplest ones fulfilling all the above requirements:

The SM introduces one complex scalar SU(2) doublet: $\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$,

with particular $SU(2)_L \times U(1)_Y$ quantum numbers given by:

$$T(\Phi) = \frac{1}{2}, \quad Y(\Phi) = 1,$$

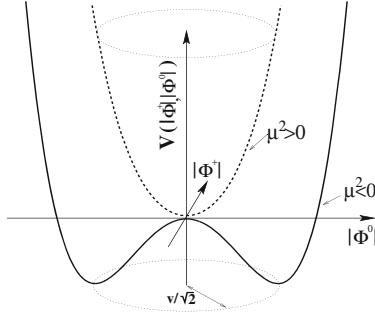
and with a potential defined as:

$$V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2, \quad \lambda > 0,$$

that is a copy of the Ginzburg-Landau one with the replacements: $\alpha_1(T - T_C) \rightarrow \mu^2$; $\alpha_2 \rightarrow \lambda$ and $M \rightarrow \Phi$.

The two phases in the SM case, are reached, as in the previous example, by setting the sign of μ^2 , either to $\mu^2 > 0$ if we want to place the SM vacuum in the symmetric phase, or to $\mu^2 < 0$ if we want the SM vacuum to be in the non-symmetric phase. These two phases are schematically described below:

$$\begin{aligned}\mu^2 > 0 : & SU(2)_L \times U(1)_Y \\ \mu^2 < 0 : & SU(2)_L \times U(1)_Y \rightarrow U(1)_{\text{em}}\end{aligned}$$



Summarizing the outcome of this spontaneous EW symmetry breaking: for $\mu^2 < 0$ the SM vacuum is not $SU(2)_L \times U(1)_Y$ invariant but it is just $U(1)_{\text{em}}$ invariant.

This SSB is usually represented by the sequence: full EW symmetry \rightarrow vacuum symmetry, namely, $SU(2)_L \times U(1)_Y \rightarrow U(1)_{\text{em}}$.

4 The Spectra of the SM and the Particle Masses

In order to get the proper gauge boson and fermion masses by means of the Higgs Mechanism the EW symmetry breaking Lagrangian $\mathcal{L}_{\text{EWSB}}$ has to be properly defined. Within the SM, $\mathcal{L}_{\text{EWSB}}$ is built by including the previous $V(\Phi)$, the proper covariant derivatives of Φ and the Yukawa interactions of Φ with fermions. Specifically, one defines:

$$\mathcal{L}_{\text{EWSB}} = \mathcal{L}_{\text{SBS}} + \mathcal{L}_{\text{YW}},$$

where the Symmetry Breaking Sector Lagrangian (SBS) is given by:

$$\mathcal{L}_{\text{SBS}} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi),$$

with

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2,$$

and

$$D_\mu \Phi = (\partial_\mu - \frac{1}{2} i g \vec{\tau} \cdot \vec{W}_\mu - \frac{1}{2} i g' B_\mu) \Phi,$$

The Yukawa Lagrangian (YW) is given in terms of the λ_q Yukawa couplings by:

$$\mathcal{L}_{\text{YW}} = \lambda_e \bar{l}_L \Phi e_R + \lambda_u \bar{q} \tilde{\Phi} u_R + \lambda_d \bar{q}_L \Phi d_R + h.c. + \text{2nd and 3rd families},$$

with

$$\begin{aligned} l_L &= \begin{pmatrix} v_L \\ e_L \end{pmatrix} ; \quad q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \\ \Phi &= \begin{pmatrix} \phi^+ \\ \phi_0 \end{pmatrix} ; \quad \tilde{\Phi} = i\tau_2 \Phi^* = \begin{pmatrix} \phi_0^* \\ -\phi^- \end{pmatrix} \end{aligned}$$

It is a simple exercise to check that the $\mathcal{L}_{\text{EWSB}}$ above is gauge $SU(2)_L \times U(1)_Y$ invariant.

Once $\mathcal{L}_{\text{EWSB}}$ is defined, then one follows the following steps:

(1) Fix a particular non-symmetric vacuum. For instance:

$$\langle 0 | \Phi | 0 \rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} ; \quad \arg \Phi = 0$$

(2) Perform ‘small oscillations’ around this vacuum:

$$\Phi(x) = \exp \left(i \frac{\vec{\xi}(x) \vec{\tau}}{v} \right) \begin{pmatrix} 0 \\ \frac{v+H(x)}{\sqrt{2}} \end{pmatrix}$$

where $\vec{\xi}(x) = (\xi_1(x), \xi_2(x), \xi_3(x))$ and $H(x)$ are ‘small’ fields.

(3) To eliminate the nonphysical (would-be-GBs) fields $\vec{\xi}$ make the following gauge transformation (i.e. go to the unitary gauge):

$$\begin{aligned} \Phi' &= U(\xi) \Phi = \begin{pmatrix} 0 \\ \frac{v+H}{\sqrt{2}} \end{pmatrix} ; \quad U(\xi) = \exp \left(-i \frac{\vec{\xi} \cdot \vec{\tau}}{v} \right) \\ l'_L &= U(\xi) l_L ; \quad e'_R = e_R ; \quad q'_L = U(\xi) q_L ; \quad u'_R = u_R ; \quad d'_R = d_R \\ \left(\frac{\vec{\tau} \cdot \vec{W}'_\mu}{2} \right) &= U(\xi) \left(\frac{\vec{\tau} \cdot \vec{W}_\mu}{2} \right) U^{-1}(\xi) - \frac{i}{g} (\partial_\mu U(\xi)) U^{-1}(\xi) ; \quad B'_\mu = B_\mu \end{aligned}$$

(4) Rotate the weak eigenstates to the mass eigenstates:

$$\begin{aligned} W_\mu^\pm &= \frac{W_\mu'^1 \mp i W_\mu'^2}{\sqrt{2}} ; \\ Z_\mu &= \cos \theta_W \ W_\mu'^3 - \sin \theta_W \ B'_\mu ; \\ A_\mu &= \sin \theta_W \ W_\mu'^3 + \cos \theta_W \ B'_\mu ; \end{aligned}$$

where the weak angle θ_W , defining the physical electroweak neutral gauge bosons, gives also the relations between the electromagnetic and the weak couplings:

$$g = \frac{e}{\sin \theta_W} ; \quad g' = \frac{e}{\cos \theta_W}$$

(5) Read the (tree level) particle masses from the proper terms in $\mathcal{L}_{\text{EWSB}}$:

$$(D_\mu \Phi')^\dagger (D^\mu \Phi') = \left(\frac{g^2 v^2}{4} \right) W_\mu^+ W^{\mu -} + \frac{1}{2} \left(\frac{(g^2 + g'^2)v^2}{4} \right) Z_\mu Z^\mu + \dots$$

$$V(\Phi') = \mu^2 H^2 + \dots$$

$$\mathcal{L}_{\text{YW}} = - \left(\lambda_e \frac{v}{\sqrt{2}} \right) \bar{e}'_L e'_R - \left(\lambda_u \frac{v}{\sqrt{2}} \right) \bar{u}'_L u'_R - \left(\lambda_d \frac{v}{\sqrt{2}} \right) \bar{d}'_L d'_R + h.c. + \dots$$

And, from these expressions above, one finally gets the tree level particle masses:

$$M_W = \frac{gv}{2}; M_Z = \frac{\sqrt{g^2 + g'^2}v}{2}; M_H = \sqrt{2}|\mu|;$$

$$m_e = \lambda_e \frac{v}{\sqrt{2}}; m_u = \lambda_u \frac{v}{\sqrt{2}}; m_d = \lambda_d \frac{v}{\sqrt{2}}; \dots$$

And, by construction, the photon and the neutrinos are got massless within the SM.

The first immediate conclusion, after the building of the SM spectra on top of the non-symmetric vacuum, is that one finds three massive weak gauge bosons, W^+ , W^- and Z and one physical scalar massive boson with positive parity, the H particle. This 0^+ particle is named the Higgs boson particle of the SM.

Notice also that, as expected, the number of bosonic degrees of freedom is preserved:

$$\begin{aligned} \text{before SSB} &= 12 \text{ (4 } \times \text{ 2 gauge + 4 scalar)}; \\ \text{after SSB} &= 12 \text{ (3 } \times \text{ 3 gauge + 1 } \times \text{ 2 gauge + 1 scalar).} \end{aligned}$$

The second conclusions from the above expressions is that all the SM particle masses, as predicted from the Higgs Mechanism, are given in terms of the parameter v with energy dimension, and whose relation with the input parameters μ and λ in the potential is given by:

$$v = \sqrt{\frac{-\mu^2}{\lambda}}, \text{ with } \mu^2 < 0, \text{ and } \lambda > 0.$$

Notice that both μ and λ are unknown parameters of the model. Therefore, the predicted tree level Higgs boson mass above, $M_H = \sqrt{2}|\mu|$ and the Higgs self interactions given by λ are unknown quantities within the SM. In contrast, it is worth recalling that the value of the parameter v , i.e. the vacuum expectation value of the Φ field, was known from the experiments long time ago, indeed before the discovery of W^\pm and Z . It was obtained from physical observables, well known from experiment. For instance, for the muon decay width $\Gamma(\mu^- \rightarrow \nu_\mu \bar{\nu}_e e^-)$ it was known the prediction from the V-A Theory (Feynman, Gell-Mann 1958) in terms of the Fermi constant G_F and the muon mass m_μ given by:

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

which provides a rather good prediction for the muon life time:

$$\tau_\mu = 2.2 \times 10^{-6} \text{ s}, \text{ for } G_F = 1.167 \times 10^{-5} \text{ GeV}^{-2}, \text{ and } m_\mu = 0.10566 \text{ GeV}.$$

On the other hand, within the SM, the muon decay proceeds via an intermediate virtual W exchange: Therefore, by matching the above Γ to the prediction in the SM one gets:

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2} = \frac{1}{2v^2} \Rightarrow v = 246 \text{ GeV}$$

Finally, by using this v , the experimental value for $\sin^2 \theta_W \simeq 0.23$ from e.g. DIS data and $g = e/\sin \theta_W$, with e set by the fine structure constant, $\alpha = e^2/(4\pi)$, one gets the tree level mass values:

$$\Rightarrow M_W^{\text{tree}} \simeq 78 \text{ GeV}, M_Z^{\text{tree}} \simeq 89 \text{ GeV}.$$

Thus, these values were known much before the weak bosons W^\pm and Z were discovered at CERN in 1983 and, indeed, they were pretty close to the experimental measured values!!.

At present there are much more precise predictions of M_W and M_Z , beyond tree level, i.e., with radiative corrections included, and there are also much more precise measurements of these masses, and the agreement between theory and experiment is excellent. These predictions of the gauge boson masses are probably one of the greatest successes of the SM.

In contrast to the gauge boson masses, the values of the fermion masses were not predicted from the Yukawa couplings, since these later were not extracted previously from other physical observables. In fact, in the fermion sector the predictions were rather the other way around: namely, the Yukawa couplings were extracted from the experimental measurements of the fermion masses. For instance, for $m_e \simeq 0.5 \times 10^{-3}$ GeV one gets $\lambda_e \simeq 3 \times 10^{-6}$, for $m_t \simeq 173$ GeV one gets $\lambda_t \simeq 1$, and similarly for the other fermions. One of the greatest mysteries nowadays in particle physics, not explained by the SM, is to understand the origin of such a widely spread fermion masses, or equivalently, the origin of the large hierarchy among the various fermion Yukawa couplings, which vary from extremely small values for leptons, more specifically those of the first generation and obviously all the tiny neutrino Yukawa couplings, to quite sizable values in the quark sector, specially that of the top quark being close to one!!!. Whatever explains this must be beyond SM physics.

5 SM Higgs Boson Couplings

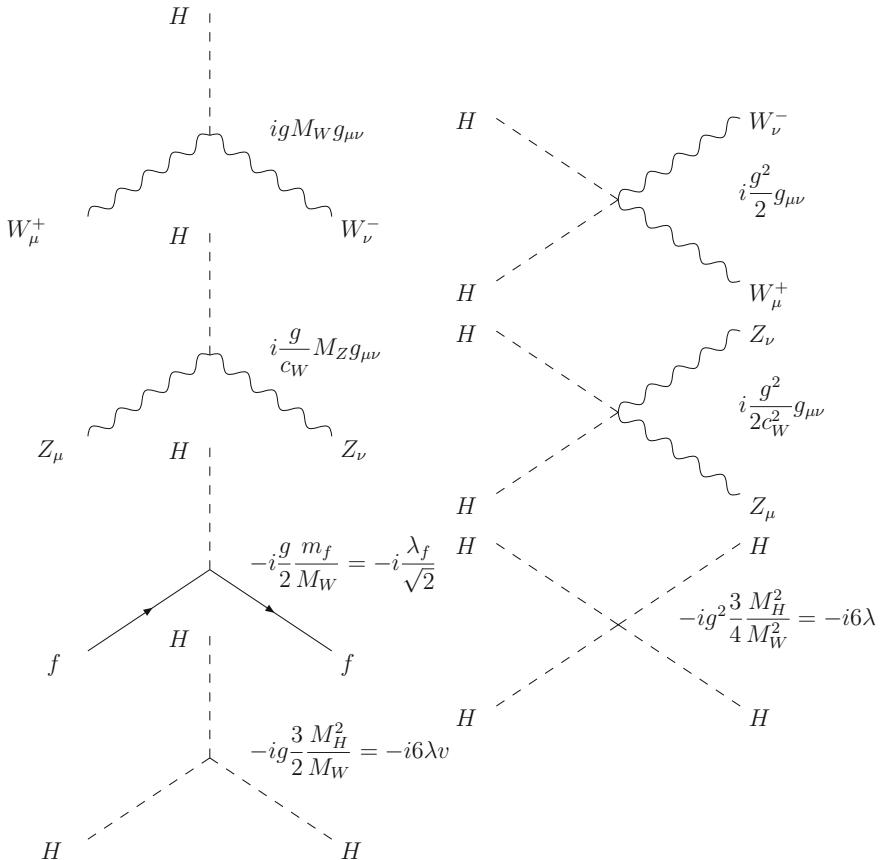
In order to get the SM Higgs boson couplings to gauge bosons and to fermions one has to work out the interaction Lagrangian terms from the previous \mathcal{L}_{SBS} and \mathcal{L}_{YW} and express them in terms of the physical basis. Instead of writing the final

interaction Lagrangian, we prefer here to express the Higgs interaction terms by the corresponding Feynman rules and the corresponding Higgs boson couplings. These are collected in the following drawings.

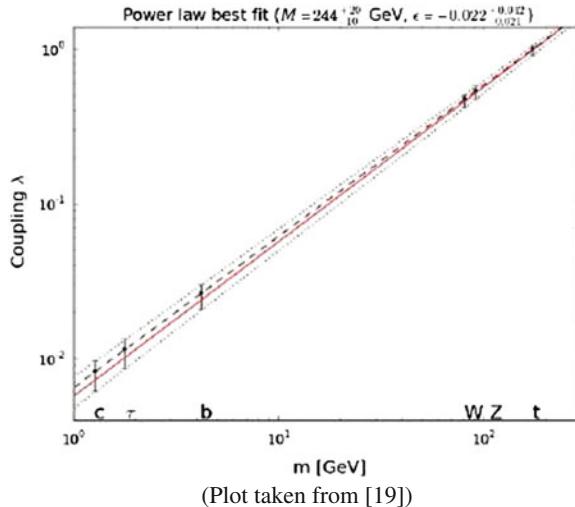
The most remarkable feature of the Higgs couplings is that they grow with the mass of the particle that is coupled to. Thus, the Higgs coupling to fermions f are larger for larger m_f , the Higgs couplings to the gauge bosons W and Z go respectively with M_W and M_Z , and the Higgs self-couplings, both the triple and the quartic, are more intense for heavier M_H . A very simple exercise is to use the present experimentally measured value at LHC of $M_H \simeq 125$ GeV and using the SM tree level relation. Thus, by assuming $M_H \simeq 125$ GeV and using the SM tree level relation:

$$\lambda = \frac{g^2 M_H^2}{8 M_W^2},$$

one gets, $\lambda \simeq 0.12$, which is indeed a small coupling if we compare it with either the electromagnetic coupling, $e \simeq 0.3$ and the weak coupling, $g \simeq 0.63$.



Regarding the comparison of the SM predictions for the Higgs boson couplings to fermions and bosons with the experimental data from LHC, there seems to be a good agreement up to now, although the statistical significance of this agreement is still not very high. In the next plot we see that when comparing the value of the measured Higgs couplings to a particle P , λ_P , versus the corresponding mass of the particle P that the Higgs is coupled to, m_P , one also finds a good agreement (black dashed line is best fit to data and the dotted black lines are 68 % CL ranges) with the SM prediction (red solid line). This is clearly signaling that the Higgs particle that has been discovered at LHC has couplings to the fundamental fermions and gauge bosons that are proportional to their mass, as in the SM.



6 SM Higgs Boson Decays and Production at LHC

(1) Higgs decay into fermions

The Higgs boson coupling to fermion f can be written in terms of the Fermi constant G_F as:

$$g_{f\bar{f}H} = \left[\sqrt{2} G_F \right]^{1/2} m_f$$

The Higgs boson partial decay width to fermions can then be expressed as:

$$\Gamma(H \rightarrow f\bar{f}) = K \frac{G_F M_H}{4\sqrt{2}\pi} m_f^2 (M_H^2) \left(1 - 4 \frac{m_f^2}{M_H^2} \right)^{3/2}$$

where $K = N_c$ = number of colors if f is a quark, or $K = 1$ if f is a charged lepton.

Notice that the previous expression has the same functional form as the tree level partial decay width, except that the dominant radiative corrections are included into the value of the running fermion mass. Thus, for instance, the bulk of QCD corrections for decays to quarks are mapped into

$$m_q^2(\text{pole}) \rightarrow m_q^2(M_H^2)$$

A simple numerical estimate of the previous partial decay width shows that the dominant decay process is: $H \rightarrow b\bar{b}$.

(2) Decay to heavy gauge bosons ($V = W, Z$)

The relevant Higgs boson coupling here is the coupling to a couple of gauge bosons VV , with $V = W$ or $V = Z$, that can be written as:

$$g_{VVH} = 2 \left[\sqrt{2} G_F \right]^{1/2} M_V^2$$

The on-shell decay width ($M_H > 2M_V$) at the tree level can be easily computed and gives:

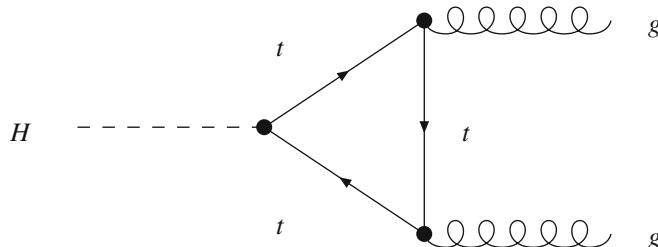
$$\Gamma(H \rightarrow VV) = \delta_V \frac{G_F M_H^3}{16 \sqrt{2} \pi} \left(1 - 4 \frac{M_V^2}{M_H^2} + 12 \frac{M_V^4}{M_H^4} \right) \left(1 - 4 \frac{M_V^2}{M_H^2} \right)^{1/2}$$

with $\delta_{W,Z} = 2, 1$

(3) Higgs boson decay to massless gauge bosons ($gg, \gamma\gamma$)

The Higgs decays to two gluons and two two photons do not exit at the tree level, and the first non vanishing contributions within the SM appear at the one loop level. These are very relevant decays precisely because of this fact, and any deviation from these SM predictions coming from new physics will be noticed most probably in these kind of decay channels.

The Higgs decay width into two gluons can be estimated with the dominant one-loop diagram where the top quark is propagated through the triangle, as in the figure below:

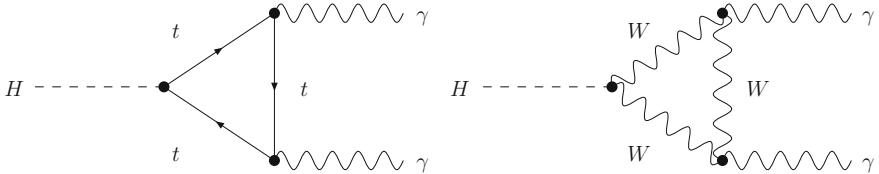


Besides, to account for the huge higher order QCD radiative corrections one has to correct the one-loop result by using the running strong coupling constant and some extra terms summarized into the one with C below.

$$\Gamma(H \rightarrow gg) = \frac{G_F \alpha_s^2(M_H^2) M_H^3}{36 \sqrt{2} \pi^3} \left[1 + C \frac{\alpha_s(\mu)}{\pi} \right]$$

$$C = \frac{215}{12} - \frac{23}{6} \log\left(\frac{\mu^2}{M_H^2}\right) + \mathcal{O}(\alpha_s)$$

In the case of Higgs boson decays to a couple of photons, there are two relevant one-loop diagrams, the triangular one with top quarks and the triangular one with W gauge bosons propagating through the triangle, as given below:



In this case, the computation of these two diagrams is sufficient to approximately describe this decay. The resulting partial decay width can be written as:

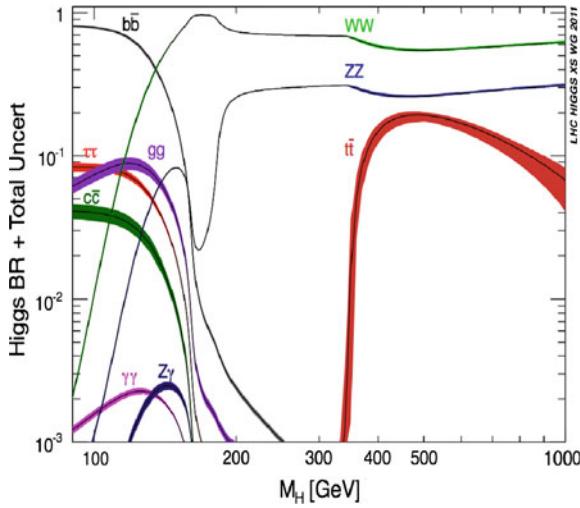
$$\Gamma(H \rightarrow \gamma\gamma) = \frac{G_F \alpha^2 M_H^3}{128 \sqrt{2} \pi^3} \left| \frac{4}{3} e_t^2 - 7 \right|^2$$

where the first term is from the top quark loop and the second one from the W boson loop, and e_t is the top quark electric charge.

Summary of the branching ratios for the SM Higgs boson decays

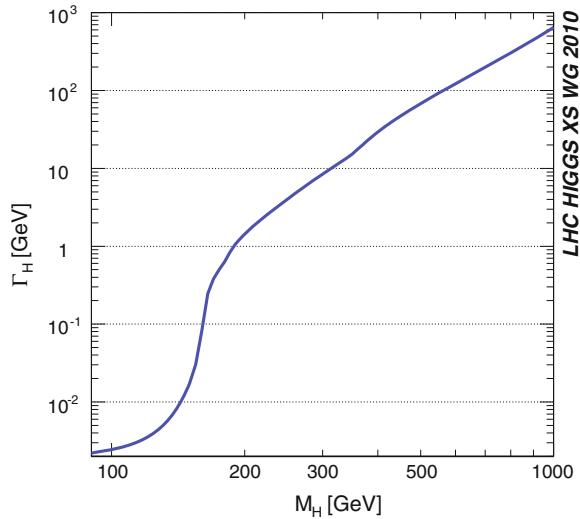
The SM predictions for the branching ratios for all the Higgs boson decays as a function of the Higgs mass are collected in the plot below. The total uncertainties in these predictions are also included.

At low values of M_H the decay to $b\bar{b}$ dominates and for high M_H the decays to weak gauge bosons dominate. In the region of most interest, namely with M_H close to 126 GeV, most of the channels enter into play and, in particular, the $H \rightarrow \gamma\gamma$ channel, even though it has a small branching ratio of about 2×10^{-3} , is indeed one of the most relevant channels at LHC due to the experimental feasibility to detect photons. The other most relevant channels proceed via ZZ and these are the decays to four fermions: $H \rightarrow ZZ \rightarrow f_1 \bar{f}_1 f_2 \bar{f}_2$. The cases where these fermions are either a muon or an electron, i.e., $H \rightarrow \mu^+ \mu^- \mu^+ \mu^-$, $H \rightarrow e^+ e^- e^+ e^-$, and $H \rightarrow \mu^+ \mu^- e^+ e^-$ (the so-called golden-channels) have received much attention in the recent years since these leptons are well measured at LHC. But overall, noways all the channels are being studied at LHC and ATLAS and CMS provide indeed measurements for most of these branching ratios and the couplings involved.



The total width of the SM Higgs boson

The prediction of total Higgs width as a function of the Higgs mass within the SM is shown in the plot below.



The total width grows with M_H , and reaches large values at large M_H . In fact, Γ_H gets comparable with the mass itself in the extreme case of M_H close to 1,000 GeV. However, for M_H close to 126 GeV the total width is very narrow, below 10^{-2} GeV, and therefore it is very difficult to measure. At present there is not an experimental measurement of the total Higgs width.

Higgs boson production at the LHC

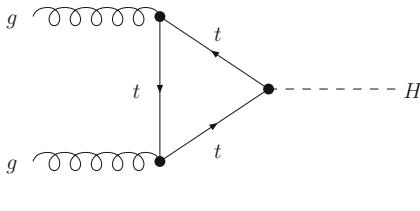
The most important SM Higgs production channels at the LHC are:

- (1) gluon fusion: $gg \rightarrow H$

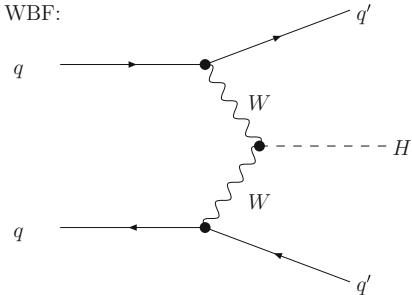
- (2) weak boson fusion (WBF): $q\bar{q} \rightarrow q'\bar{q}'H$
- (3) W boson associated production: $q\bar{q}' \rightarrow WH$
- (4) Z boson associated production: $q\bar{q} \rightarrow ZH$
- (5) top quark associated production: $gg, q\bar{q} \rightarrow t\bar{t}H$.

The relevant Feynman diagrams for the two first channels are shown below:

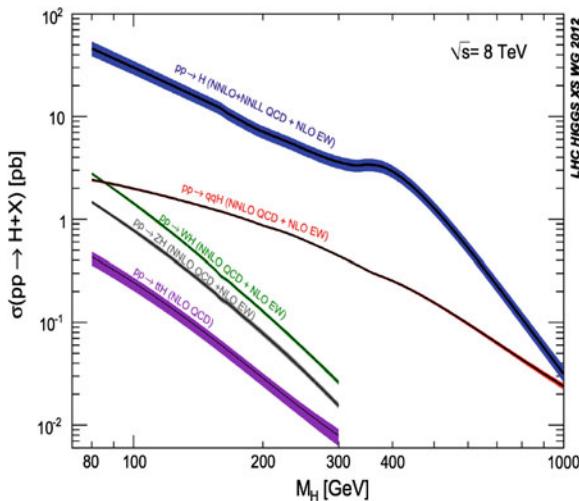
Gluon-Fusion:



WBF:



The predictions within the SM for the Higgs boson cross sections at LHC with $\sqrt{s} = 8$ TeV in the various channels as a function of the Higgs boson mass are collected in the figure below:



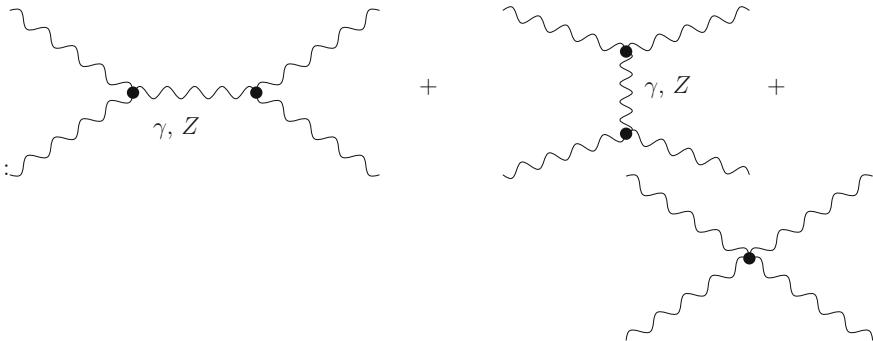
7 Other Interesting Properties of the SM Higgs System

There are several properties of the SM Higgs system that are worth to keep in mind, since they may give us some clue in the future studies of the fundamental physics underlying the EW symmetry breaking sector. We select here three: (1) the role of the Higgs boson in the scattering of longitudinal weak gauge bosons, (2) the Equivalence

Theorem, and (3) the theoretical limits on the value on the Higgs mass. We will comment shortly on these next.

Higgs boson role in scattering of longitudinal W and Z bosons:

It is interesting to recall that when one computes the scattering amplitude for longitudinal gauge bosons, W_L and/or Z_L without including the diagrams with a Higgs boson, one gets a resulting amplitude that does not preserve unitarity at high energies. For instance, computing the three diagrams below for $W_L W_L \rightarrow W_L W_L$ scattering:

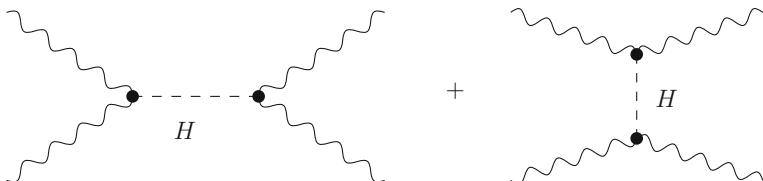


one gets, for large E , the following behavior:

$$T_V = -g^2 \frac{E^2}{M_W^2} + \dots$$

which implies a *violation of unitarity* at high energies.

However, if one considers, in addition, the contribution of a scalar particle H with couplings proportional to the mass, given by the two extra diagrams below:



whose behavior at high energies is given by:

$$T_S = g_{WWH}^2 \frac{E^2}{M_W^4} + \dots$$

Then the final result for the total amplitude is:

$$T_{\text{tot}} = T_V + T_S = \frac{E^2}{M_W^4} \left(g_{WWH}^2 - g^2 M_W^2 \right) + \dots$$

and the potential terms with bad high-energy behavior happens to cancel for

$$g_{WWH} = g M_W$$

which is exactly the value of the SM Higgs coupling to the W gauge bosons. In conclusion, the SM Higgs particle repairs the bad high energy behavior of the longitudinal weak gauge bosons and the resulting total amplitudes are unitary at all energies. However, it is also convenient to keep in mind that this is not a unique solution and the SM Higgs system could be replaced by something else (new Higgs particles, new gauge bosons, new resonances, etc) which could effectively play this same role.

Comparing the WW scattering with the would-be-GB scattering

An interesting result is provided by the so-called Equivalence Theorem that relates the scattering of massive gauge bosons and the scattering of GBs.

Equivalence Theorem [20, 21]:

The scattering amplitudes of longitudinal gauge bosons V_L ($V = W^\pm, Z$), at high energies, $\sqrt{s} \gg M_V$, are equivalent to the scattering amplitudes of their corresponding would-be Goldstone bosons w

$$|T(V_L^1 V_L^2 \dots V_L^N \rightarrow V_L^1 V_L^2 \dots V_L^{N'})| \approx |T(w_1 w_2 \dots w_N \rightarrow w_1 w_2 \dots w_{N'})|$$

For instance, instead of using the unitary gauge one can use the more general Feynman rules of R_ξ gauges and get the following relations: (1)

$$T(W_L^+ W_L^- \rightarrow W_L^+ W_L^-) = T(w^+ w^- \rightarrow w^+ w^-) + O\left(\frac{M^2}{s}\right), \text{ for } \sqrt{s} \gg M_W, M_Z$$

and, (2) for $M_H \gg M_{W,Z}$:

$$\begin{aligned} \Gamma(H \rightarrow W_L^+ W_L^-) &= \Gamma(H \rightarrow w^+ w^-) + O\left(\frac{M_W}{M_H}\right) \\ \Gamma(H \rightarrow Z_L Z_L) &= \Gamma(H \rightarrow zz) + O\left(\frac{M_Z}{M_H}\right) \end{aligned}$$

These results above and some others provided by the Equivalence Theorem, apart of being of practical use, may give us also some clue in the future to further understand the fundamental dynamics underlying the EW symmetry breaking sector.

Theoretical limits on the Higgs mass

Next we present the three most popular Higgs mass limits from theory: (I) The upper Higgs mass limit from unitarity, (II) The upper Higgs mass limit from triviality, and (III) The Lower Higgs mass limit from vacuum stability.

I: Upper Higgs mass bound from unitarity

Let us study here the behavior of the scattering amplitude of longitudinal gauge bosons with respect to the value of M_H . The complete tree level result is given by:

$$\begin{aligned} T(W_L^+ W_L^- \rightarrow W_L^+ W_L^-) = & -\frac{1}{v^2} \left\{ -s - t + \frac{s^2}{s - M_H^2} + \frac{t^2}{t - M_H^2} + 2M_Z^2 \right. \\ & \left. + \frac{2M_Z^2 s}{t - M_Z^2} + \frac{2t}{s} (M_Z^2 - 4M_W^2) - \frac{8s_W^2 M_W^2 M_Z^2 s}{t(t - M_Z^2)} \right\} \end{aligned}$$

Next we decompose T in partial waves a_J defined by:

$$T(s, \cos \theta) = 16\pi \sum_{J=0}^{\infty} (2J+1) a_J(s) P_J(\cos \theta), \quad P_J = \text{Legendre polynomials}$$

One can then compute the cross-section in terms of these partial waves:

$$\sigma_{\text{tot}} \simeq \sigma_{\text{el}} = \frac{16\pi}{s} \sum_{J=0}^{\infty} (2J+1) |a_J(s)|^2$$

On the other hand if we require σ to fulfill the Optical Theorem (this OT is a consequence of unitarity $T^\dagger T = TT^\dagger = 1$):

$$\sigma_{\text{tot}}(1+2 \rightarrow \text{anything}) = \frac{1}{s} \text{Im } T(s, \cos \theta = 1)$$

when applied to σ_{el} , the OT can then be written in terms of the partial waves as:

$$|a_J(s)|^2 = \text{Im } a_J(s); \quad \forall J \Rightarrow |a_J|^2 \leq 1; \quad 0 \leq \text{Im } a_J \leq 1; \quad |\text{Re } a_J| \leq \frac{1}{2}; \quad \forall J$$

For instance, when applied to the lowest partial wave, defined by:

$$a_0(W_L^+ W_L^- \rightarrow W_L^+ W_L^-) = \frac{1}{32\pi} \int_{-1}^1 T(s, \cos \theta) d(\cos \theta)$$

one finds an expression for $|\text{Re } a_0|$ valid in the high energy limit, $\sqrt{s} \gg M_H, M_W$:

$$|a_0| \xrightarrow{s \gg M_H^2, M_V^2} \frac{M_H^2}{8\pi v^2}$$

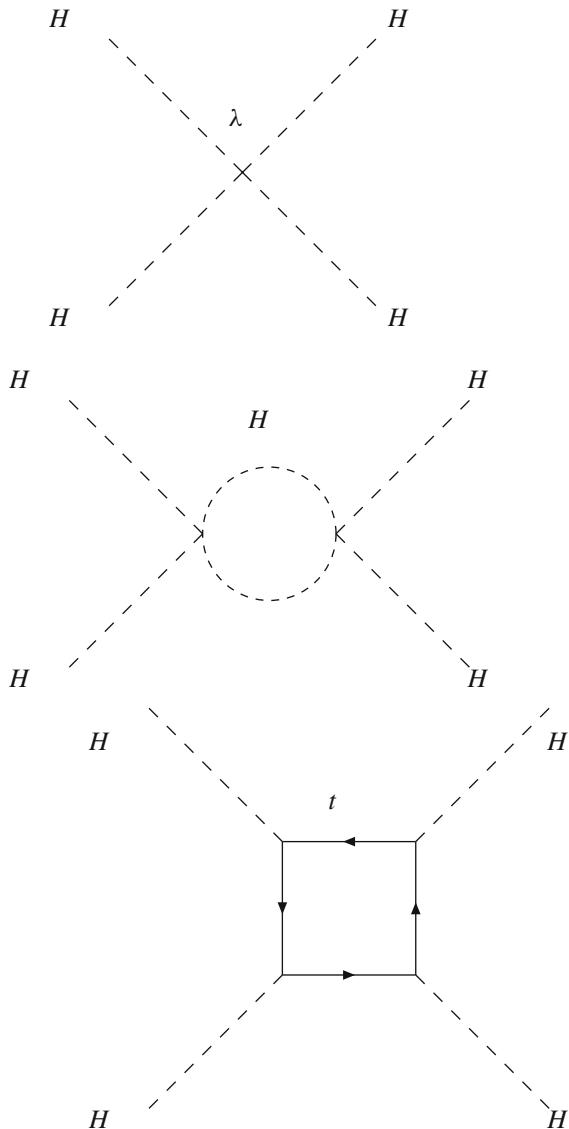
Therefore, the corresponding unitary bound for a_0 leads to an upper bound for M_H , which in this particular case is:

$$|\text{Re } a_0| \leq \frac{1}{2} \Rightarrow M_H < 860 \text{ GeV}$$

This unitary bound can be improved if higher order corrections beyond tree level are included, and also by considering other possible channels. But the size of the final upper bound remains close to this.

Upper Higgs mass bound from triviality

Let us first consider the running of the Higgs self-coupling at the one-loop level, whose dominant contributions are given by the three diagrams below:



The relevant renormalization group equation (RGE) for the self-coupling λ is:

$$\frac{d\lambda}{dt} = \frac{3}{16\pi^2} \left[4\lambda^2 + 2\lambda g_t^2 - g_t^4 + \frac{1}{16} (2g_2^4 + (g_2^2 + g_1^2)^2) \right], \quad t = \log \left(\frac{Q^2}{v^2} \right)$$

Notice that we use here a different notation than before: the top Yukawa coupling is g_t , and the SM gauge couplings are g_1 , g_2 and g_3 respectively.

The so-called ‘Triviality Problem’ arises when λ is large and it is related to the existence of a pole, named the Landau pole, in the solution to the previous RGE. For large λ , one can neglect the contributions from g_t , g_1 , g_2 and g_3 in the RGE and keep just the dominant contribution from λ , leading to a simple solution for the running coupling constant $\lambda(Q)$ in terms of the bare coupling constant λ_0 :

$$\begin{aligned} \frac{d\lambda}{dt} &= \frac{3}{4\pi^2} [\lambda^2] \\ \Rightarrow \lambda(Q) &= \frac{\lambda_0}{1 - \frac{3\lambda_0}{2\pi^2} \log \left(\frac{Q}{\Lambda} \right)} ; \quad \lambda_0 \equiv \lambda(\Lambda) \end{aligned}$$

where the presence of the Landau pole is manifest. Now, by taking the $\Lambda \rightarrow \infty$ limit, while fixing λ_0 to a finite value, one finds that the effective coupling $\lambda(Q) \rightarrow 0$ and in consequence the theory is trivial, i.e. non-interacting. The only way out from this is to assume the existence of a finite physical cut-off Λ_{phys} such that $\lambda(Q) \neq 0$ all the way up to this cut-off. Then, by defining the renormalized Higgs mass in terms of $\lambda(v)$ as:

$$M_H^2 = 2\lambda(v)v^2 \quad \text{with} \quad \lambda(v) = \frac{\lambda_0}{1 - \frac{3}{2\pi^2}\lambda_0 \log \left(\frac{v}{\Lambda_{\text{phys}}} \right)},$$

one finds that for decreasing (increasing) $\Lambda_{\text{phys}} \Rightarrow M_H$ increases (decreases) and indeed they may cross. This crossing point where $M_H(\Lambda_{\text{phys}}) \simeq \Lambda_{\text{phys}}$ is what gives the upper bound to M_H . Clearly, this is a cut-off dependent bound.

Lower Higgs mass bound from vacuum stability

The problem of vacuum instability arises for small or negative λ . It can be understood either from the behavior of the effective potential or from the behaviour of the solution to the RGE for λ .

In few words, the behavior of the effective potential is as follows. The minimum of the effective potential (including loop corrections) changes with $\lambda(Q)$ and, a too small or negative $\lambda(Q)$ may change the true vacuum: from wanted stable vacuum $V(v) < V(0)$ to the unwanted unstable vacuum with $V(v) > V(0) \Rightarrow$ in which case the electroweak symmetry breaking does not take place. Indeed the situation can be even worse, since it can lead to an effective potential that is not even bounded from below!! In summary, by requiring vacuum stability, namely by imposing $V(v) < V(0)$, one then gets a lower bound on $\lambda(v)$ and in consequence also on M_H . This lower bound on M_H is also cut-off dependent.

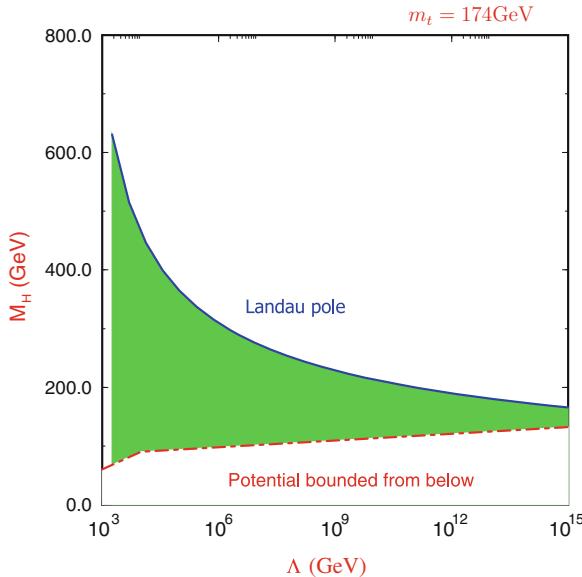
One can alternatively solve the one loop RGE in the small λ regime by, for instance, keeping just the dominant terms:

$$\begin{aligned} \frac{d\lambda}{dt} &= \frac{3}{16\pi^2} \left[-g_t^4 + \frac{1}{16} (2g_2^4 + (g_2^2 + g_1^2)^2) \right] \\ \Rightarrow \quad \lambda(Q^2) &= \lambda(v^2) + \frac{3}{16\pi^2} \left[-g_t^4 + \frac{1}{16} (2g_2^4 + (g_2^2 + g_1^2)^2) \right] \log \left(\frac{Q^2}{v^2} \right) \end{aligned}$$

and then require a positive $\lambda(\Lambda)$. Thus, one gets a lower limit on M_H that again depends on Λ :

$$\lambda(\Lambda) > 0 \Rightarrow M_H^2 > \frac{3v^2}{8\pi^2} \left[g_t^4 - \frac{1}{16} (2g_2^4 + (g_2^2 + g_1^2)^2) \right] \log \left(\frac{\Lambda^2}{v^2} \right)$$

Both M_H limits, upper and lower, combined



In the figure above the two limits on M_H previously commented are plotted together as a function of the cut-off Λ , which is interpreted to be the scale up to which the SM is valid. The upper blue line is the upper limit from avoiding the Landau pole/Triviality and the lower dot-dashed red line is the lower limit from avoiding the vacuum instability and by requiring the potential to be bounded from below. The shaded green region in between the two lines is the allowed area for the SM value of M_H . For instance, if one requires the SM to be valid up to the scale of the Grand Unification Theory, then the Higgs mass should be within the following approximate interval:

$$\text{For } \Lambda = M_{\text{GUT}} \Rightarrow 130 \text{ GeV} \lesssim M_H \lesssim 180 \text{ GeV}$$

and, as can be seen in the above figure, the allowed interval gets narrower for larger Λ .

Recent computations of the stability lower bound include a NNLO analysis of the Higgs potential and realistic error estimates.

The condition for absolute stability up to the Planck scale gives the following bound (see for instance, [22]):

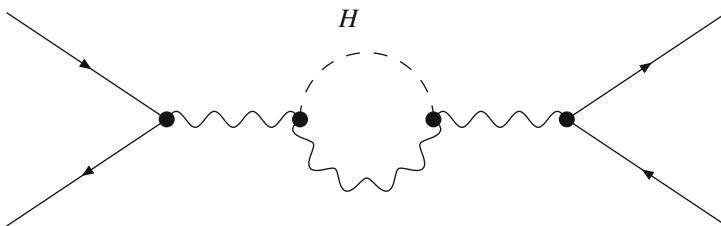
$$M_H(\text{GeV}) > 129.4 + 1.4 \left(\frac{m_t(\text{GeV}) - 173.1}{0.7} \right) - 0.5 \left(\frac{\alpha_s(M_Z) - 0.1184}{0.0007} \right) \pm 1.0_{\text{th}}$$

$$\Rightarrow M_H > 129.4 \pm 1.8 \text{ GeV}$$

From this lower bound one then may conclude that vacuum stability of the SM up to the Planck scale is excluded at 2σ (98 %CL) for $M_H < 126$ GeV !!! This is a quite remarkable result, given the present experimental measurement which is precisely pretty close to 126 GeV.

Higgs mass limits from radiative corrections

Another interesting Higgs mass limits can be extracted from the contributions of the Higgs particle, via radiative corrections, to the electro-weak precision observables (EWPO). For instance, the Higgs particle can propagate into the loops that contribute to the $e^+e^- \rightarrow \mu^+\mu^-$ scattering, and correct the tree level prediction for the observables associated to this process by an amount whose size depend, among other parameters, on the value of M_H . One example of a one-loop diagram where the Higgs enters in a relevant way is the one shown in the figure below, where the Higgs and the Z bosons propagate inside the loop correcting the intermediate Z boson propagator. A comparison between the prediction for a EWPO from the SM at a given order in perturbation theory and for a given M_H value with the experimental measurement for this EWPO allows to set an allowed interval on M_H (or equivalently a preferred by data M_H window) and also set exclusion limits on M_H .



There are many examples of EWPO where the Higgs particle contributes. For illustration, we choose here one of the most studied observables in the literature: The prediction for M_W in terms of M_Z , α , G_F and Δr :

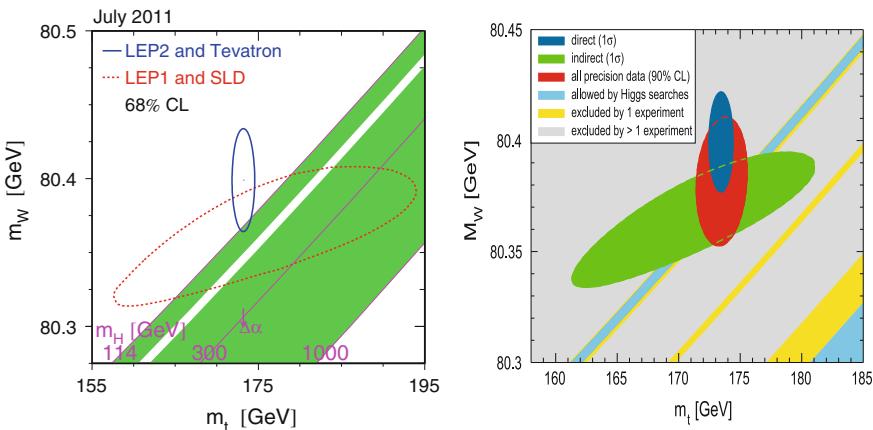
$$M_W^2 \left(1 - \frac{M_W^2}{M_Z^2} \right) = \frac{\pi \alpha}{\sqrt{2} G_F} \left(\frac{1}{1 - \Delta r} \right).$$

The parameter Δr collects all the loop corrections and summarizes the deviations from the tree level relation: $M_W^2 \left(1 - \frac{M_W^2}{M_Z^2}\right) = \frac{\pi \alpha}{\sqrt{2} G_F}$. This Δr can be evaluated, for instance, from μ decay and from this one then gets M_W . The one-loop result for M_W in the SM is well known (see, A. Sirlin '80 and W. Marciano, A. Sirlin '80) and contains three relevant contributions given schematically by:

$$\begin{aligned} \Delta r_{1\text{-loop}} = & \Delta\alpha - \frac{c_W^2}{s_W^2} \Delta\rho + \Delta r_{\text{rem}}(M_H) \\ & \sim \log \frac{M_Z}{m_t} \quad \sim m_t^2 \quad \sim \log(M_H/M_W) \\ & \sim 6\% \quad \sim 3.3\% \quad \sim 1\% \end{aligned}$$

the contribution from the top quark is larger than the Higgs contribution, since the dependence with the top mass is quadratic. In contrast, the dependence on the Higgs mass is logarithmic, and indeed this dependence is quite general for many EWPO.

This interesting exercise of comparing the SM prediction for M_W as a function of the top mass and the Higgs mass with the data was done very often in the past, and more intensely in the era of LEP, SLD and TeVatron. These were done before the starting of the LHC and there were already some indications that the data seemed to prefer a light Higgs boson. We include below, on the left, one of these plots produced by the LEP Electroweak Working Group (LEPEWWG) in 2011. The oblique lines are the predictions for specific values of M_H (GeV), 114, 300, 1000 ... and we can see clearly that the lines at the lowest values, i.e. those close to 114 GeV, fit better to the experimental measurement than the lines for heavier Higgs boson.



For comparison, we also include, on the right, the same kind of plot but corresponding to the situation after the LHC had started taking data. In this plot, the red area is allowed by all precision data at 90 %CL and the light blue oblique bands are the SM prediction for M_W as a function of m_t , with M_H allowed by Higgs searches at LHC, before the Higgs discovery: (a) Central band: $115.5 \text{ GeV} < M_H < 127 \text{ GeV}$,

(b) band at lower-right corner: $M_H > 600$ GeV. Again we see that a light Higgs mass in this 115.5 GeV $< M_H < 127$ GeV window was preferred by all data before the Higgs discovery. It is certainly a good lesson to learn for the future searches of new physics beyond the SM, since the radiative corrections from this new physics could contribute to the precision observables in a relevant way and give us some clue on the scale where to look for this new physics.

The ρ parameter and the custodial symmetry

Another interesting parameter that measures the relevance of radiative corrections in the EW theory, and that is very sensitive to new physics beyond the SM is the so-called ρ parameter. This parameter is defined as the ratio of neutral to charged current amplitudes at low energies:

$$\rho \equiv \frac{T_{NC}(q^2 \ll M_Z^2)}{T_{CC}(q^2 \ll M_W^2)}$$

From ν -scattering experiments and others there is a good measurement: $\rho_{\text{exp}} = 1.0008^{+0.0020}_{-0.0011}$ (PDG 2012)

The SM prediction at tree level is:

$$\rho_{\text{tree}}^{\text{SM}} = \frac{M_W^2 \text{ tree}}{M_Z^2 \text{ tree} \cos^2 \theta_W^{\text{tree}}} = 1$$

At one loop and keeping just the so-called ‘oblique’ corrections,

$$\rho = \frac{\rho_{\text{tree}}}{1 - \Delta\rho} ; \quad \Delta\rho = \frac{\Sigma_Z^R(0)}{M_Z^2} - \frac{\Sigma_W^R(0)}{M_W^2} \quad \text{related to } T \text{ parameter}$$

For instance, the leading top and Higgs loop contributions give:

$$\begin{aligned} (\Delta\rho)_t &= \frac{g^2}{64\pi^2} N_C \frac{m_t^2}{M_W^2} + \dots \\ (\Delta\rho)_H &= -\frac{g^2}{64\pi^2} 3 \tan^2 \theta_W \log \frac{M_H^2}{M_W^2} + \dots \end{aligned}$$

The ρ parameter being close to one is due to the so-called custodial symmetry: a global symmetry of the SM Higgs sector in absence of gauge interactions.

In order to illustrate clearer this custodial symmetry it is convenient to use an alternative way of writing the (ungauged) Lagrangian of the SBS:

$$\begin{aligned} \mathcal{L}_{\text{SBS}} &= \frac{1}{4} \text{Tr} \left[(\partial_\mu M)^\dagger (\partial^\mu M) \right] - V(M) ; \\ V(M) &= \frac{1}{4} \lambda \left[\frac{1}{2} \text{Tr}(M^\dagger M) + \frac{\mu^2}{\lambda} \right]^2 \end{aligned}$$

where M is a 2×2 matrix containing the four real scalar fields of the Φ doublet:

$$\begin{aligned} M &\equiv \sqrt{2}(\tilde{\Phi}\Phi) = \sqrt{2} \begin{pmatrix} \phi_0^* & \phi^+ \\ -\phi^- & \phi_0 \end{pmatrix}; \\ \Phi &= \begin{pmatrix} \phi^+ \\ \phi_0 \end{pmatrix}; \\ \tilde{\Phi} &= i\tau_2\Phi^* = \begin{pmatrix} \phi_0^* \\ -\phi^- \end{pmatrix} \end{aligned}$$

It is immediate to check that \mathcal{L}_{SBS} is invariant under the global transformations:

$$M \rightarrow g_L M g_R^+; \quad g_L \subset SU(2)_L; \quad g_R \subset SU(2)_R$$

This global symmetry $SU(2)_L \times SU(2)_R$ is called chiral EW symmetry (for analogy with QCD) and it is spontaneously broken down to the diagonal subgroup $SU(2)_{L+R} \equiv SU(2)_{\text{custodial}}$.

The pattern of global symmetry breaking is:

$$SU(2)_L \times SU(2)_R \rightarrow SU(2)_{\text{custodial}}$$

Once $SU(2)_L \times U(1)_Y$ is gauged, the chiral symmetry (and the custodial) is explicitly broken. This custodial symmetry has many interesting implications for phenomenology, and it could provide some clue in the future searches of new physics beyond the Standard Model.

8 Motivations for Looking Beyond the Standard Model

In spite of the extraordinary success of the SM describing all the Particle Physics phenomena known so far in Nature, there is the general believe that the SM cannot be the ultimate theory of Fundamental Physics. There are several aspects where the SM does not provide a satisfactory answer. Here we list very briefly some of the issues that require going beyond the Standard Model:

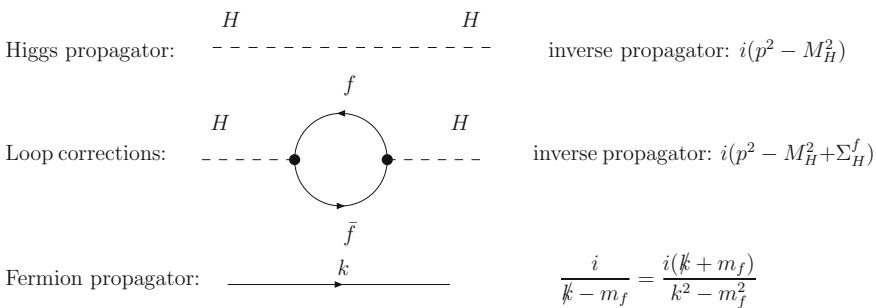
- The SM does not contain gravity. A fundamental theory including all the four known interactions: electromagnetic, weak, strong and gravitational, must go beyond SM.
- The SM does not provide gauge coupling unification. A fundamental theory that unifies all known gauge interactions must go beyond SM.
- The SM does not include neutrino masses nor intergenerational mixings for neutrinos. An explanation of the origin of non-vanishing neutrino masses and neutrino oscillations that are found in experiments require going beyond SM.

- The SM has no proper candidate for Dark Matter. The observations demonstrate that Dark Matter exists in Nature. Explaining the origin of Dark Matter requires going beyond SM.
- The Higgs sector of the SM suffers of the so-called hierarchy problem. This will be shortly described below. Solving this problem requires going beyond SM.

9 The Hierarchy Problem of the SM Higgs Sector

This problem can be expressed as the instability of the value of the Higgs boson mass when radiative corrections are included in presence of a physical cut-off that is placed at energies far above the electroweak scale. It should be emphasized that it appears exclusively when the SM is considered as a low energy effective theory that is valid up to this given cut-off. But this assumption seems to be the most probable one, given the previous list of unsolved issues within the SM, one is lead to think that the SM is not a fundamental theory but rather a successfull effective theory.

Let us illustrate with a bit more detail how this hierarchy problem appears when computing the one-loop radiative corrections to the Higgs mass in presence of an ultraviolet cut-off Λ . One starts with the Higgs propagator at the tree level, describing the free propagation, and then one adds the one-loop diagrams correcting this propagator, as the one shown below with the self-energy correction Σ_H given by the fermionic contribution Σ_H^f , and this later being easily computable from the corresponding Feynman integral over the internal fermion momentum k . Schematically:



Integrating over all possible loop momenta k , considering N_f degrees for fermion f with mass m_f and Yukawa coupling λ_f , and keeping just the dominant terms for large cut-off Λ , gives:

$$\Sigma_H^f \sim N_f \lambda_f^2 \int d^4k \left(\frac{1}{k^2 - m_f^2} + \frac{2m_f^2}{(k^2 - m_f^2)^2} \right) (-1)$$

$$\text{for } \Lambda \rightarrow \infty : \quad \Sigma_H^f \sim N_f \lambda_f^2 \left(\underbrace{\int \frac{d^4 k}{k^2}}_{\sim \Lambda^2} + \underbrace{2m_f^2 \int \frac{d^4 k}{k^4}}_{\sim \ln \Lambda} \right) (-1)$$

And from this, one finally gets the mass corrections as a function of Λ :

$$\delta M_H^2 = N_f \frac{\lambda_f^2}{16\pi^2} \left(-2\Lambda^2 + 6m_f^2 \log \frac{\Lambda}{m_f} + \dots \right).$$

There are two dominant contributions at high values of the cut-off: the largest one that grows quadratically as Λ^2 and the other one that grows logarithmically as $\log \Lambda$. Thus, for instance, if one takes the cut-off at the Planck energy scale: $\Lambda = M_{\text{Pl}}$:

$$\delta M_H^2 \sim M_{\text{Pl}}^2 \Rightarrow \delta M_H^2 \approx 10^{30} M_H^2,$$

i.e., one finds an unacceptable huge correction that is 30 orders of magnitude larger than the starting tree level squared Higgs mass (for $M_H \lesssim 1$ TeV).

Another popular example is the case of Grand Unified Theories (GUT), where the physical cut-off is at $\Lambda = M_{\text{GUT}} \sim 10^{16}$ GeV and one also gets huge corrections given by $\delta M_H^2 \approx M_{\text{GUT}}^2$.

In summary, the hierarchy problem is the instability of the small Higgs mass to large corrections in a context where the SM is a low energy remnant of a more fundamental theory with a large mass scale in addition to the weak scale. Furthermore, this instability occurs because:

- there is no additional symmetry for $M_H = 0$,
- and in consequence, there is no protection against large corrections.

10 Two Main Avenues to Solve the Hierarchy Problem

There are two qualitatively different proposals to solve the hierarchy problem of the SM. Generically: One avenue assumes new symmetries and the Higgs boson is an elementary particle; The other avenue assumes new interactions and the Higgs boson is a composite particle. At present there is not yet any experimental evidence in favor of none of these two possibilities, therefore the issue of the elementarity/compositeness of the observed Higgs particle is still an open question. The generic features of these two main avenues are summarized schematically below:

Elementary higgs	Composite higgs
There should exist an extra symmetry (at least) and new particles with couplings dictated by this symmetry such that the most problematic quadratic sensitivity to the high scale cancels	At some scale the Higgs dissolves and the theory of constituents is at work
The typical example is <i>Supersymmetry</i> where the sparticle partner cancels the quadratic divergence generated by the particle	Similar to QCD where the pions dissolve into quarks
The soft SUSY breaking scale acts as a cutoff of divergences	The compositeness scale acts as a cutoff of quadratic divergences
The Higgs boson is weakly interacting	The typical example is <i>Technicolor</i> Theories where the Higgs boson is strongly interacting and the Higgs mass is at TeV scale
The Higgs self-coupling is related to the EW gauge coupling	Modern theories of compositeness involve an <i>extra dimension</i> through the AdS/CFT correspondence. The Higgs mass value and the size of its couplings are very model dependent
The Higgs boson mass is close to the EW scale	The smallness of m_H , close to the EW scale, can be explained if H is a Pseudo-Goldstone boson
Typically a bunch of new elementary particles appear in the spectrum	Typically, new composite resonances appear in the spectrum
Including several Higgs particles, besides the SM Higgs-like boson	

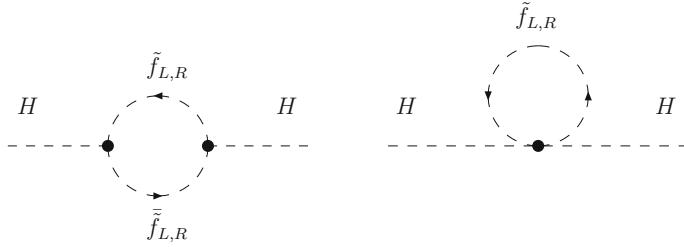
11 Supersymmetry

The existence of one (or more) new symmetry relating fermions and bosons is the most popular proposal to solve the hierarchy problem of the SM Higgs sector. This new symmetry is called supersymmetry (SUSY) and generically acts as:

$$\begin{aligned} Q|\text{boson}\rangle &= |\text{fermion}\rangle \\ Q|\text{fermion}\rangle &= |\text{boson}\rangle \end{aligned}$$

which, effectively, produces an enlargement with respect to the SM spectrum: the SM particles have SUSY partners that share their same quantum numbers but differ by one half unit in their spin. Thus, the SM left-handed and right-handed fermions, $f_{L,R}$, have their corresponding SUSY partners, named sfermions, $\tilde{f}_{L,R}$, that are scalar particles, the SM gauge bosons have their SUSY partners named gauginos that are fermions, and the Higgs bosons have their SUSY partners named higgssinos that are also fermions.

SUSY then solves the hierarchy problem by the additional contributions from sfermions. When computing the one-loop radiative corrections to the Higgs boson propagator in SUSY theories one has to add new contributions from the scalar fermion partners that are given by the two graphs below. As previously done, one then focus on the behavior of these new contributions at large values of the cut-off Λ :



$$\Sigma_H^{\tilde{f}} \sim \text{first diagram } (\sim \log \Lambda) + N_{\tilde{f}} \lambda_{\tilde{f}} \int d^4k \left(\frac{1}{k^2 - m_{\tilde{f}_L}^2} + \frac{1}{k^2 - m_{\tilde{f}_R}^2} \right)$$

$$\text{for } \Lambda \rightarrow \infty: \delta M_H^2 = 2N_{\tilde{f}} \frac{\lambda_{\tilde{f}}}{16\pi^2} \left(\Lambda^2 - 2m_{\tilde{f}}^2 \log \frac{\Lambda}{m_{\tilde{f}}} \right) + \dots$$

where, for simplicity, $m_{\tilde{f}_L} = m_{\tilde{f}_R} = m_{\tilde{f}}$ is assumed. $N_{\tilde{f}}$ is the number of sfermion modes and $\lambda_{\tilde{f}}$ is the sfermions coupling to Higgs bosons involved in the second diagram.

From the previous result, it is clear that when adding these sfermion corrections to the previous fermionic corrections one finds that the quadratic contributions, $\mathcal{O}(\Lambda^2)$, cancel in the total Higgs boson mass squared if the following equations are satisfied:

$$\begin{aligned} N_{\tilde{f}_L} &= N_{\tilde{f}_R} = N_{\tilde{f}} = N_f \\ \lambda_{\tilde{f}} &= \lambda_f^2 \end{aligned}$$

and these are precisely the conditions imposed by SUSY, namely, the identity in the number of bosonic and fermionic degrees of freedom, and the specific relations between their couplings. Notice also that the total dominant corrections, including $\mathcal{O}(\Lambda^2)$ and $\mathcal{O}(\log \Lambda)$, vanish if furthermore

$$m_{\tilde{f}} = m_f$$

i.e. for exact degeneracy between fermions and sfermions.

Generically, one may then characterize the SUSY breaking by the mass splitting between fermions and their sfermion partners. Namely, if $m_{\tilde{f}}^2 = m_f^2 + \Delta^2$ and $\lambda_{\tilde{f}} = \lambda_f^2$ then one gets a total correction given by:

$$\Sigma_H^{f+\tilde{f}} \sim N_f \lambda_f^2 \Delta^2 + \dots$$

and this correction stays acceptably small if the mass splitting Δ is small, say not much heavier than the weak scale.

One then concludes that the stability of the Higgs boson mass corrections is realized if the mass scale of the SUSY partners is not very far above the weak scale,

or simply:

$$M_{\text{SUSY}} \lesssim \mathcal{O}(1 \text{ TeV})$$

Therefore, SUSY at the TeV scale provides an attractive solution to the hierarchy problem. But this implies finding SUSY at these scales, which has not happened so far in the experiments. Setting M_{SUSY} above these values, say M_{SUSY} larger than a few TeV, leads to the so-called Split SUSY models, but all these suffer of some kind of instabilities due to reintroduction of large Higgs mass corrections.

11.1 SUSY-Breaking Models

As we have seen, exact SUSY requires mass degeneracy between particles and sparticles: $m_f = m_{\tilde{f}}, \dots$, etc. However in a realistic model SUSY must be broken somehow, since the SUSY partners with such masses have not been found in Nature.

On the other hand, it is known that satisfactory models of SUSY breaking must proceed via spontaneous SUSY breaking at some high energy scale. Specific SUSY-breaking schemes (see below) in general yield an effective Lagrangian at low energies that is supersymmetric except for some explicit *soft* SUSY-breaking terms. These soft SUSY-breaking terms have the interesting property of not altering, via the radiative corrections, the dimensionless couplings of the theory. Therefore, they preserve the nice property of the unbroken SUSY case with no quadratic divergences, as required by the SUSY solution to the hierarchy problem. In fact this is true in all orders of perturbation theory and the specific types of soft SUSY-breaking terms are well known and classified in the literature. For the case of minimal SUSY particle content and two Higgs doublets, $H_d (= H_1)$ and $H_u (= H_2)$, the soft SUSY-breaking Lagrangian can be written as,

$$\begin{aligned} \mathcal{L}_{\text{soft}} = & -\frac{1}{2} \left(M_1 \tilde{B} \tilde{B} + M_2 \tilde{W} \tilde{W} + M_3 \tilde{g} \tilde{g} \right) + \text{h.c.} \\ & - (m_{H_u}^2 + |\mu|^2) H_u^+ H_u - (m_{H_d}^2 + |\mu|^2) H_d^+ H_d - (B H_u H_d + \text{h.c.}) \\ & - (\tilde{u}_R A_u \tilde{Q} H_u - \tilde{d}_R A_d \tilde{Q} H_d - \tilde{e}_R A_e \tilde{L} H_d) + \text{h.c.} \\ & - \tilde{Q}^+ m_{\tilde{Q}}^2 \tilde{Q} - \tilde{L}^+ m_{\tilde{L}}^2 \tilde{L} - \tilde{u}_R m_{\tilde{u}}^2 \tilde{u}_R^* - \tilde{d}_R m_{\tilde{d}}^2 \tilde{d}_R^* - \tilde{e}_R m_{\tilde{e}}^2 \tilde{e}_R^*. \end{aligned}$$

This is the most general parameterization of SUSY-breaking terms that keeps relations between dimensionless couplings unchanged; hence not generating quadratic divergences. It includes mass terms for the gauginos, M_i ($i = 1, 2, 3$), for the scalars m_S^2 , a Higgs bilinear B term and also trilinear couplings between the Higgs bosons and the sfermions A_f . Notice that $m_{\tilde{f}}^2$ and A_f are 3×3 matrices in family space, therefore, in general, they introduce many new parameters. Most of the SUSY-breaking

models assume that the soft SUSY-breaking mass scales involved are not far above the TeV scale, $M_{\text{SUSY}} \lesssim 1$ TeV, such that they all avoid the hierarchy problem.

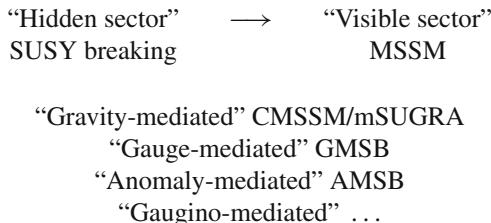
Generically, these SUSY-breaking Models can be classified in two big groups: Unconstrained Models and Constrained Models, according to the following general features:

Unconstrained models (MSSM,...)

- These are agnostic about how SUSY breaking is achieved and no particular SUSY breaking mechanism is assumed.
- They implement instead a general parameterization of all possible soft SUSY-breaking terms.
- The relations between dimensionless couplings are unchanged and, therefore, no quadratic divergences are re-introduced by the SUSY breaking.
- The simplest and most popular of these models is the Minimal Supersymmetric Standard Model (MSSM). It is called minimal because it is based in the minimal number of Supersymmetries and in the minimal particle content. In spite of being the simplest case, the MSSM still has plenty of parameters. In the most general case there are 105 new parameters, including couplings, masses, mixing angles and phases.

Constrained models (mSUGRA, ...):

- In these models there are specific assumptions on the scenario that achieves spontaneous SUSY breaking.
- Therefore, they provide specific predictions for the soft SUSY-breaking terms in terms of a smaller set of parameters.
- An experimental determination of the SUSY parameters would imply setting the patterns of SUSY breaking.
- All constrained models are special versions of the MSSM.
- There are different kinds of Constrained models, mainly according to the origin of the SUSY breaking and the way it is transmitted from the so-called “Hidden sector” to the “Visible sector”. For illustration, we include below, in an schematic way, some examples:



For instance, in two of the most popular ones CMSSM (Constrained MSSM) and mSUGRA (Minimal Supergravity), the mediating interactions are gravitational. In contrast in the GMSB (Gauge Mediated SUSY Breaking) the mediating interactions are gauge interactions, etc.

Table 1 MSSM Spectrum

	SUSY particles			
Extended standard	$SU(3)_C \times SU(2)_L \times U(1)_Y$		Mass eigenstates	
Model spectrum	Interaction eigenstates			
	Notation	Name	Notation	Name
$q = u, d, s, c, b, t$	\tilde{q}_L, \tilde{q}_R	squarks	\tilde{q}_1, \tilde{q}_2	squarks
$l = e, \mu, \tau$	\tilde{l}_L, \tilde{l}_R	sleptons	\tilde{l}_1, \tilde{l}_2	sleptons
$\nu = \nu_e, \nu_\mu, \nu_\tau$	$\tilde{\nu}$	sneutrino	$\tilde{\nu}$	sneutrino
g	\tilde{g}	gluino	\tilde{g}	gluino
W^\pm	\tilde{W}^\pm	wino		
$H_1^+ \supset H^+$	\tilde{H}_1^+	higgsino	$\tilde{\chi}_i^\pm \ (i=1,2)$	charginos
$H_2^- \supset H^-$	\tilde{H}_2^-	higgsino		
γ	$\tilde{\gamma}$	photino		
Z	\tilde{Z}	zino		
$H_1^o \supset h^0, H^0, A^0$	\tilde{H}_1^o	higgsino	$\tilde{\chi}_j^o \ (j=1, \dots, 4)$	neutralinos
$H_2^o \supset h^0, H^0, A^0$	\tilde{H}_2^o	higgsino		
W^3	\tilde{W}^3	wino		
B	\tilde{B}	bino		

Since all models are specific versions of the MSSM, we will focus next in the general features of the MSSM spectrum and in particular of the MSSM Higgs sector.

11.2 MSSM Spectrum

We summarize in Table 1 the particle content within the MSSM. Besides the SM particles, there are their corresponding SUSY partners and the extra Higgs boson particles that correspond to the enlarged Higgs sector of the MSSM with two Higgs doublets. The SM interaction eigenstates are also specified in the table. The interaction eigenstates that have the same quantum numbers mix and give rise to the physical mass eigenstates which are also specified in the table. The MSSM physical states include the squarks, sleptons, sneutrinos, gluinos, charginos, neutralinos and the physical Higgs bosons, these particles and all these particles are being searched for at the present experiments.

11.3 Enlarged Higgs Sector of the MSSM Versus SM

In the MSSM two Higgs doublets H_d ($=H_1$) and H_u ($=H_2$) are needed to give masses to down- and up-type fermions. This is in contrast to the SM where with just one

Higgs doublet both u and d type masses can be generated via the Higgs Mechanism. For instance, for the quarks of the first generation:

$$\mathcal{L}_{\text{SM}} = \underbrace{m_d \bar{Q}_L \Phi d_R}_{\text{d-quark mass}} + \underbrace{m_u \bar{Q}_L \tilde{\Phi} u_R}_{\text{u-quark mass}}$$

$$Q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L, \quad \tilde{\Phi} = i\sigma_2 \Phi^*, \quad \Phi \rightarrow \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad \tilde{\Phi} \rightarrow \begin{pmatrix} v \\ 0 \end{pmatrix}$$

There are two main reasons for two Higgs doublets in the MSSM and not just one: On one hand, Supersymmetry implies that the Superpotential must be an holomorphic function of the chiral superfields, i.e. it depends only on φ_i , not on φ_i^* . Therefore a term like $\bar{Q}_L \Phi^*$ is not allowed.

On the other hand, two Higgs doublets are also needed in SUSY for cancellation of anomalies. The fermionic partners of these Higgs scalars would otherwise contribute to non-vanishing anomalies.

The specific components of these two Higgs doublets are given by:

$$H_1 = \begin{pmatrix} H_1^1 \\ H_1^2 \end{pmatrix} = \begin{pmatrix} v_1 + (\phi_1 + i\chi_1)/\sqrt{2} \\ \phi_1^- \end{pmatrix}$$

$$H_2 = \begin{pmatrix} H_2^1 \\ H_2^2 \end{pmatrix} = \begin{pmatrix} \phi_2^+ \\ v_2 + (\phi_2 + i\chi_2)/\sqrt{2} \end{pmatrix}$$

where v_1 and v_2 are the respective vacuum expectation values of the two neutral scalar bosons, and their ratio defines a very relevant MSSM parameter: $\tan \beta = v_2/v_1$. Notice that these two complex doublets imply the introduction of eight degrees of freedom, but only three of them are really needed to provide masses to the three weak bosons. Therefore, the implementation of the Higgs Mechanism in this case will lead to five physical scalars remaining in the spectrum.

The Higgs potential of the MSSM is given in terms of these two Higgs doublets by:

$$V = m_1^2 H_1 \bar{H}_1 + m_2^2 H_2 \bar{H}_2 - m_{12}^2 (\epsilon_{ab} H_1^a H_2^b + \text{h.c.})$$

$$+ \frac{g'^2 + g^2}{8} (H_1 \bar{H}_1 - H_2 \bar{H}_2)^2 + \frac{g^2}{2} |H_1 \bar{H}_2|^2$$

One remarkable feature of this potential is that the Higgs self-couplings are given in terms of the EW gauge couplings, in contrast to the SM potential where the self-couplings were given by the unknown λ .

After the spontaneous EW symmetry breaking, one finds the announced five physical Higgs particles, h^0, H^0, A^0, H^\pm , which appear as ‘oscillations’ around the asymmetric vacuum state.

The would-be Goldstone bosons disappear in the unitary gauge and one gets the wanted gauge boson masses:

$$M_W^2 = \frac{1}{2}g'^2(v_1^2 + v_2^2), \quad M_Z^2 = \frac{1}{2}(g^2 + g'^2)(v_1^2 + v_2^2), \quad M_\gamma = 0$$

All the predictions in the MSSM Higgs sector can then be expressed in terms of two input parameters, $\tan \beta$ and the mass of the CP-odd Higgs boson M_A :

$$\tan \beta = \frac{v_2}{v_1}, \quad M_A^2 = -m_{12}^2(\tan \beta + \cot \beta)$$

The masses of the MSSM Higgs bosons

The tree-level prediction for the Higgs boson masses in the MSSM are:

$$m_{H,h}^2 = \frac{1}{2} \left[M_A^2 + M_Z^2 \pm \sqrt{(M_A^2 + M_Z^2)^2 - 4M_Z^2 M_A^2 \cos^2 2\beta} \right],$$

and

$$m_{H^\pm}^2 = M_A^2 + M_W^2.$$

This implies an upper mass bound for the lightest Higgs mass m_h :

$$\Rightarrow m_h^2 \leq M_Z^2 \cos^2 2\beta \quad (\text{the equality holds for } M_A \gg M_Z).$$

Therefore, at tree level: $m_h < M_Z$, what is clearly in contradiction with the experiments that have not found such a light Higgs particle.

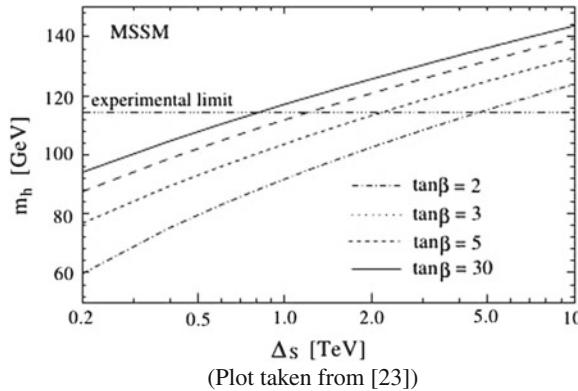
However, when radiative corrections are included, the Higgs masses get shifted respect to the tree level values and, in particular, the lightest Higgs corrected mass m_h (we use here the same notation for the tree and corrected masses, for shortness) can be considerably enhanced in some regions of the MSSM parameter space. In fact there are large corrections from the Yukawa couplings of the third generation quarks, being the corrections from the top quark the largest ones. For instance, the dominant 1-loop corrections from the top-stop sector can be written (for $M_A \gtrsim 150$ GeV) approximately as:

$$m_h^2 \simeq M_Z^2 \cos^2 2\beta + \frac{3g^2}{8\pi^2} \frac{m_t^4}{M_W^2} \log\left(\frac{m_t^2}{m_b^2}\right),$$

which clearly indicates the large size of the correction, since it goes with a large factor m_t^4 in front, and we also see the logarithmic growing of this correction with the relevant SUSY scale, here the stop mass.

The approximate behavior of the corrected m_h as a function of the relevant SUSY mass for several values of $\tan \beta$ is illustrated in the plot below. In this plot we see a very relevant increase of m_h with $\Delta_S = m_{\tilde{t}}$ and with $\tan \beta$ that lead to Higgs mass values that are compatible with data.

The present status of the predictions for m_h in the MSSM is that a complete one-loop result and ‘almost complete’ two-loop result are available. In particular, one can find some regions of the MSSM parameter space where the predicted m_h is compatible with the Higgs mass value that has been recently measured at LHC. Generically, one can conclude that for soft SUSY masses at or slightly below a few TeV the lightest Higgs boson of the MSSM is indeed light, with m_h of order $\sim \mathcal{O}(100 \text{ GeV})$, and it could perfectly fit the LHC measured Higgs mass value.



The couplings of the MSSM Higgs bosons:

The MSSM (neutral) Higgs boson couplings to weak gauge bosons and fermions are predicted to be at tree level as follows:

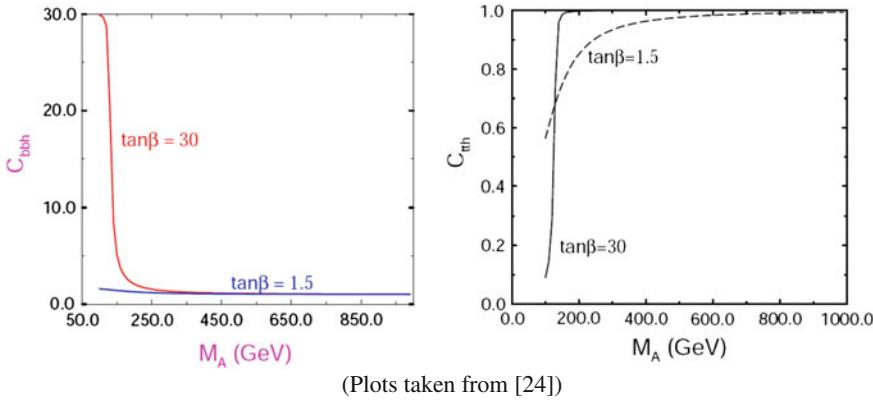
$$\begin{aligned}
g_{hVV} &= \sin(\beta - \alpha) g_{HVV}^{\text{SM}}, \quad V = W^\pm, Z \\
g_{HVV} &= \cos(\beta - \alpha) g_{HVV}^{\text{SM}} \\
g_{hAZ} &= \cos(\beta - \alpha) \frac{g'}{2 \cos \theta_W} \\
g_{hb\bar{b}}, g_{h\tau^+\tau^-} &= -\frac{\sin \alpha}{\cos \beta} g_{Hb\bar{b}, H\tau^+\tau^-}^{\text{SM}} \\
g_{Hb\bar{b}}, g_{H\tau^+\tau^-} &= \frac{\cos \alpha}{\cos \beta} g_{Hb\bar{b}, H\tau^+\tau^-}^{\text{SM}} \\
g_{ht\bar{t}} &= \frac{\cos \alpha}{\sin \beta} g_{Ht\bar{t}}^{\text{SM}} \\
g_{Ht\bar{t}} &= \frac{\sin \alpha}{\sin \beta} g_{Ht\bar{t}}^{\text{SM}} \\
g_{Ab\bar{b}}, g_{A\tau^+\tau^-} &= \gamma_5 \tan \beta g_{Hb\bar{b}, H\tau^+\tau^-}^{\text{SM}} \\
g_{At\bar{t}} &= \gamma_5 \cot \beta g_{Ht\bar{t}}^{\text{SM}}
\end{aligned}$$

where α is the mixing angle in the CP-even Higgs sector that is given at tree level by:

$$\alpha = \arctan \left[\frac{-(m_A^2 + M_Z^2) \sin \beta \cos \beta}{M_Z^2 \cos^2 \beta + m_A^2 \sin^2 \beta - m_h^2} \right], \quad -\frac{\pi}{2} < \alpha < 0,$$

and the g^{SM} couplings are the corresponding Higgs boson couplings of the SM.

The first obvious conclusion from the formulas above is that the lightest MSSM Higgs boson couplings to weak gauge bosons are always smaller than the corresponding SM Higgs boson couplings, $g_{hVV} \leq g_{HVV}^{\text{SM}}$, $V = W^\pm, Z$, whereas, the couplings to b quarks (t quarks) and to τ leptons $g_{hb\bar{b}}, g_{h\tau^+\tau^-}$ ($g_{h t\bar{t}}$) can get a significant enhancement (suppression) with respect to the corresponding SM coupling at large $\tan \beta$. This enhancement (suppression) is illustrated in the plot below, where the ratio $C_{qqh} = g_{hq\bar{q}}/g_{Hq\bar{q}}^{\text{SM}}$ is shown as a function of M_A for two values of $\tan \beta$.



(Plots taken from [24])

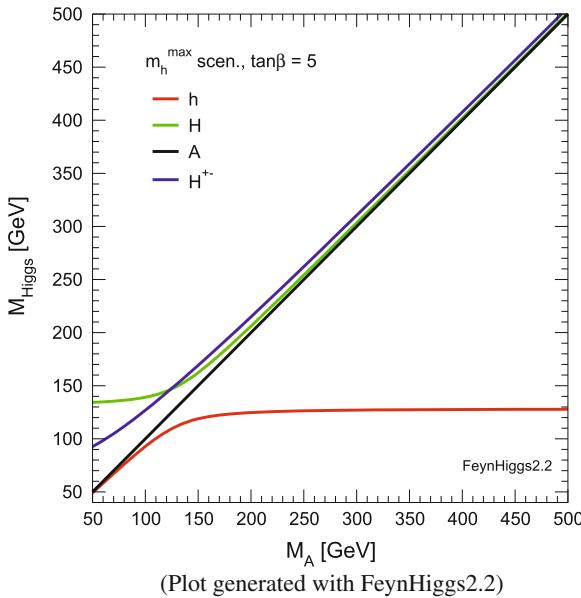
As the Higgs boson masses, all the couplings above also get corrected by loops when going beyond the tree level and in some regions of the MSSM parameter space these radiative corrections can be sizable.

The decoupling limit

The so-called decoupling limit corresponds to taking large values of the input mass M_A , $M_A \gg M_Z$, and it has the interesting peculiarity of getting the MSSM Higgs sector converging to the SM Higgs sector. More specifically, the lightest MSSM Higgs boson behaves as the SM Higgs boson, whereas the other MSSM Higgs bosons get heavy, close to M_A , and they decouple from the low energy physics. In particular, one can check that for $M_A \gg M_Z$ the following limits are obtained:

$$\begin{aligned} m_h^{\text{tree}} &\rightarrow M_Z |\cos 2\beta| \\ -\frac{\sin \alpha}{\cos \beta} &\rightarrow 1, \frac{\cos \alpha}{\sin \beta} \rightarrow 1, \sin(\beta - \alpha) \rightarrow 1 \Rightarrow g_{hVV} \rightarrow g_{HVV}^{\text{SM}}, g_{hf\bar{f}} \rightarrow g_{Hf\bar{f}}^{\text{SM}} \\ M_A \approx M_H &\approx M_{H^\pm} \end{aligned}$$

In fact, as can be seen in the plot below, this decoupling limit is already effective at $M_A \gtrsim 150\text{ GeV}$:



And one can conclude that the SM is the resulting low energy effective theory of the MSSM once the heavy Higgs bosons (and the heavy SUSY partners) are decoupled.

11.4 Another Interesting Properties of SUSY

Among the most interesting properties of SUSY theories that are specially relevant for phenomenology, there are the following: (1) Coupling constant unification, (2) Radiative Electroweak Symmetry Breaking and (3) R-parity.

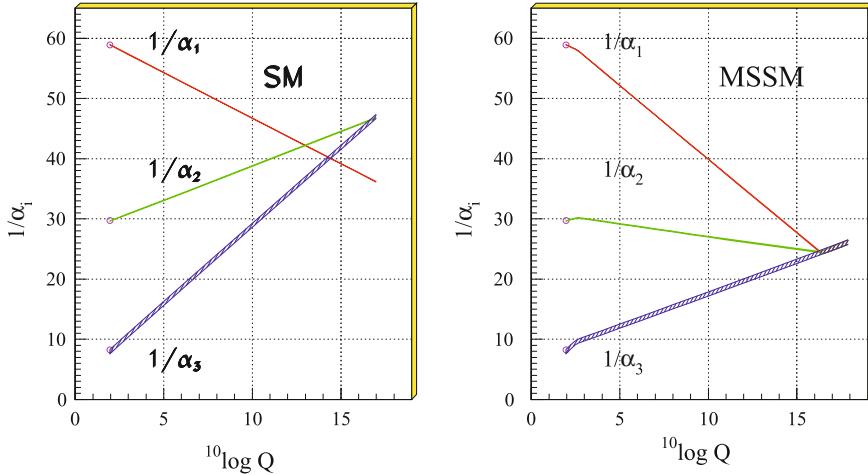
Coupling constant unification

Let us consider the running of the three coupling constants α_1 , α_2 and α_3 of the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge theory. This running is dictated by the Renormalization Group Equations (RGE's) that connect parameters at different energy scales and by the particle content of the theory. In particular, if we use the RGE's to evolve the gauge coupling constants from the electroweak (EW) scale to the Gran Unification Theory (GUT) scale,

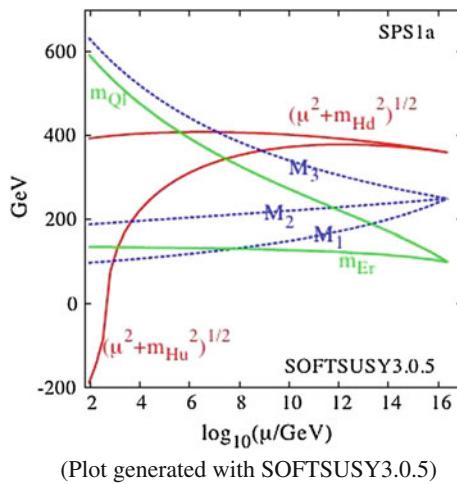
$$\alpha_i(Q_{\text{EW}}) \rightarrow \alpha_i(Q_{\text{GUT}})$$

one gets different answers depending if one uses just the SM or the SUSY enlarged models. For instance, comparing the SM with the MSSM, one gets that the three coupling constants do not meet in the SM case, whereas they do unify in the MSSM at $Q = Q_{\text{GUT}} \simeq 10^{16}$ GeV. This unification is illustrated in the next plot where the inverse of the coupling constants are shown as functions of $\log Q$, with Q expressed in GeV. Again, this is under the assumption that the soft SUSY breaking masses are not far above the TeV scale, $M_{\text{SUSY}} \lesssim 1$ TeV.

Unification of the Coupling Constants
in the SM and the minimal MSSM



Electroweak Radiative Symmetry Breaking



(Plot generated with SOFTSUSY3.0.5)

In SUSY theories, one can get the Electroweak Symmetry Breaking by means of radiative corrections that produce the needed negative squared scalar mass, as required by the Higgs Mechanism. This can be clearly illustrated by means of the running of the relevant parameters from the high energy scale down to the EW scale as provided by the RGE's. For instance, if one starts with a set of universal conditions at the GUT scale for the soft masses and run them down to the EW scale, one finds out that one scalar mass parameter gets negative, and it is precisely the needed one to produce the wanted $SU(2)_L \times U(1)_Y \rightarrow U(1)_{\text{em}}$ breaking. This is illustrated in the included plot where the $(\mu^2 + m_{H_u}^2)^{1/2}$ mass runs into negative values. And it turns out that it only works properly if again $M_{\text{SUSY}} \lesssim 1$ TeV.

R-parity

The MSSM and other SUSY models have an extra symmetry, called the “R-parity” that implies the conservation of a new multiplicative quantum number defined for each particle as,

$$P_R = (-1)^{3(B-L)+2s},$$

where B , L and s are the baryon number, the lepton number and the spin of the particle respectively. It happens that all SM-particles and Higgs bosons have even R-parity, $P_R = +1$, whereas, all superpartners have odd R-parity, $P_R = -1$.

This symmetry has very important consequences for phenomenology. First, the SUSY particles appear only in pairs. For instance, the production of neutralinos in electron-positron collisions occur in pairs, like, e.g.: $e^+e^- \rightarrow \tilde{\chi}_1^+\tilde{\chi}_1^-$.

It also has very important consequences for Dark Matter Physics, since it provides a natural particle candidate for explaining the Dark Matter: the lightest SUSY particle (LSP) that, due R-parity, is stable. More specifically, the lightest neutralino is usually the most popular candidate for (Cold) Dark Matter in the MSSM and other SUSY models.

Another interesting and very relevant consequence of R-parity for collider phenomenology is that, since the LSP is neutral and uncolored, it leaves no traces in collider detectors and, therefore, the typical SUSY signatures are events with unbalanced energy, i.e. with apparent “missing energy”.

Another proposals as: the gravitino, the axino and others have also been considered as Dark Matter candidates in the literature.

12 Compositeness

Another different avenue is to think of the Higgs boson as a composite particle instead of a fundamental one which was the avenue in the previous SUSY models. The hypothesis of a composite Higgs is perfectly compatible with present data, therefore why not to consider it. Another important qualitative difference with respect to the previous SUSY case is that typically the interactions leading to a composite state must be strong. So the composite hypothesis for the Higgs system leads generically

to new strong interactions beyond the Standard Model, in contrast to the SUSY hypothesis that does not introduce generically new interactions but instead leads to new fundamental particles.

Once one assumes a composite Higgs, then one has to make a particular assumption on what is the strong dynamics that is responsible for the production of such a composite state if one is interested in studying further phenomenological consequences of its existence in Nature. This implies to set a specific underlying Quantum Field Theory describing these new strong interactions beyond the Standard Model and, therefore, the predictions will be obviously model dependent. There are many of such strongly interacting models and we will not go through all of them. However, one can learn some interesting general features by just looking for singularities with some well known physical examples where composite particles appear as a result of strong dynamics. First we know that in a theory with strong dynamics it is common to generate a populated spectra with a dense tower of resonances built from the more fundamental objects that interact strongly. The mass of these resonances are in general related to the typical scale where the strong dynamics condensates, Λ_{strong} . Apart from resonances there are another interesting composite systems that appear in strong interacting theories having a global symmetry that is spontaneously broken. According to the Goldstone Theorem, in these cases there may appear in the spectrum new scalar or pseudoscalar composite states, the Goldstone Bosons, that should be massless if the global symmetry of the Lagrangian was exact, or they may get a small mass in the case that this is not exact but an approximate global symmetry. The small breaking in this later case is usually accounted for by some small explicit breaking masses in the Lagrangian, given generically by a new scale, Λ_{break} that may be or not related with Λ_{strong} . Thus, under the generic hypothesis of the Higgs boson being a composite/dynamical state of a strongly interacting gauge theory, still there are two generic possibilities. Either it emerges as a resonance, and its mass is related to the condensation scale, Λ_{strong} , or it is a Pseudo-Goldstone boson and its mass emerges as a consequence of the ‘small’ breaking of the global symmetry, therefore it is related with Λ_{break} . In the following we will look into particular examples that are illustrative of what could happen in case that the Higgs boson is a composite particle. The most popular example is to think that the Higgs system of the SM relies in a strongly theory that is a copy of QCD, but operating at higher energies.

12.1 Electroweak Chiral Symmetry and Composite Higgs

In order to understand the possible similarities of the Higgs system with QCD, it is more convenient to use the alternative parametrization of the SM Higgs field that were commented in a previous section of these lectures.

Let us again consider here this alternative way of writing the (ungauged) Lagrangian for the Symmetry Breaking Sector of the SM, given by:

$$\begin{aligned}\mathcal{L}_{\text{SBS}} &= \frac{1}{4} \text{Tr} \left[(\partial_\mu M)^\dagger (\partial^\mu M) \right] - V(M) ; \\ V(M) &= \frac{1}{4} \lambda \left[\frac{1}{2} \text{Tr}(M^\dagger M) + \frac{\mu^2}{\lambda} \right]^2\end{aligned}$$

where M is a 2×2 matrix containing the four real scalar fields of the doublet Φ :

$$\begin{aligned}M &\equiv \sqrt{2}(\tilde{\Phi}\Phi) = \sqrt{2} \begin{pmatrix} \phi_0^* & \phi^\dagger \\ -\phi^- & \phi_0 \end{pmatrix} ; \\ \Phi &= \begin{pmatrix} \phi^+ \\ \phi_0 \end{pmatrix} ; \quad \tilde{\Phi} = i\tau_2\Phi^* = \begin{pmatrix} \phi_0^* \\ -\phi^- \end{pmatrix}\end{aligned}$$

Written in this way it is easier to see the existence of an extra global symmetry of this SBS Lagrangian. More specifically, \mathcal{L}_{SBS} is invariant under the global and independent transformations g_L and g_R acting on M :

$$M \rightarrow g_L M g_R^+ ; \quad g_L \subset SU(2)_L ; \quad g_R \subset SU(2)_R$$

This global symmetry $SU(2)_L \times SU(2)_R$ is called *Electroweak Chiral Symmetry* due to the obvious analogy with the well known Chiral Symmetry of QCD. One can also check that this Electroweak Chiral Symmetry is spontaneously broken down to the diagonal subgroup, usually called the custodial symmetry group, $SU(2)_{L+R} \equiv SU(2)_{\text{custodial}}$.

The pattern of the global symmetry breaking in the SM Higgs system is therefore:

$$SU(2)_L \times SU(2)_R \rightarrow SU(2)_{\text{custodial}}$$

This breaking, according to the Goldstone Theorem, would give rise to the appearance in the spectrum of three massless Goldstone Bosons, which in this case are pseudo scalars (i.e. with negative parity) since it is the $SU(2)_{L-R}$ symmetry the one that is not a symmetry of the vacuum. Once the gauge interactions are incorporated as usual, by the gauge principle associated to the $SU(2)_L \times U(1)_Y$ gauge symmetry, then these three GBs disappear from the spectrum and the three corresponding gauge bosons, W^\pm and Z , get the proper masses. On the other hand, within the full SM, there are symmetry breaking terms in the Lagrangian that indeed break the two global symmetries $SU(2)_L \times SU(2)_R$ and $SU(2)_{\text{custodial}}$, and this is why the associated GBs of the EW Chiral symmetry breaking are not strictly massless and they are usually referred to as pseudo-GBs with a ‘small’ associated mass.

Some immediate possibilities arise for a composite Higgs from the above way of thinking: It could be that the Higgs is a scalar resonance emerging from some new strong interactions among new fermions, as it happens in QCD where many resonances (like σ , ρ , etc.) emerge from the strong interactions among quarks. Or it could happen that the Higgs particle emerges as a pseudo-Goldstone boson associated to a spontaneous symmetry breaking of a larger symmetry group, con-

taining the previous 'minimal breaking pattern' given by $SU(2)_L \times SU(2)_R \rightarrow SU(2)_{\text{custodial}}$.

12.2 Chiral Symmetry and the Chiral Lagrangian of QCD

We have seen in the previous lecture that in QCD with two massless quarks, u and d , there is an extra global symmetry, the Chiral Symmetry of QCD, that is spontaneously broken down to the isospin group:

$$SU(2)_L \times SU(2)_R \rightarrow SU(2)_V,$$

by the non-vanishing quark condensate, $\langle 0 | \bar{q}q | 0 \rangle \neq 0$, and this breaking explains the smallness of the pion masses as compared to the typical resonance masses of QCD. The three pions, π^\pm and π^0 are the three pseudo-GBs associated to this breaking whereas the other hadrons emerge in the QCD spectrum as resonances made up by quarks and antiquarks. In this case, the masses of the pions are associated to the scale of the Chiral Symmetry breaking in QCD, $m_\pi \sim \Lambda_{\text{break}}^{\text{QCD}}$ whereas the masses of the other hadrons are related to the scale of the QCD strong interactions, $\Lambda_{\text{strong}}^{\text{QCD}}$ (usually called Λ_{QCD} in short). If these two scales of QCD are or are not related is still under debate and we will not go further into this point here.

Regarding the dynamics of pions in QCD to low energies, it is well described by an effective Lagrangian, the so-called Chiral Lagrangian of QCD, that is invariant under the Chiral Symmetry. The Effective Quantum Field Theory that is built from this Chiral Symmetry is called Chiral Perturbation Theory (ChPT).

The Chiral Lagrangian of QCD is usually written in terms of a non-linear representation of the GBs:

$$U(x) = \exp \left(\frac{i}{f_\pi} \pi_a(x) \sigma^a \right) \quad \text{with } \sigma^a (a = 1, 2, 3) = \text{Pauli matrices},$$

where f_π is the pion decay constant that is measured, for instance, from the $\pi^+ \rightarrow \mu^+ \nu_\mu$ decay:

$$\langle 0 | J^{+\mu} | \pi^-(p) \rangle = \frac{if_\pi}{\sqrt{2}} p^\mu, \quad f_\pi = 94 \text{ MeV}$$

Under a chiral transformation the $U(x)$ transforms linearly (but π transform non-linearly):

$$U(x) \rightarrow g_L U(x) g_R^+ \quad \text{with } g_L \in SU(2)_L, \quad g_R \in SU(2)_R$$

The most general chiral invariant Lagrangian is a sum of an infinite number of terms with increasing number of derivatives in the $U(x)$ and the $U^+(x)$ fields and with

an infinite number of arbitrary parameters. This provides a systematic expansion in powers of momenta and also in powers of m_π if the explicit Chiral symmetry breaking terms are included into the Lagrangian.

Thus in ChPT to lowest order, $\mathcal{O}(p^2)$, and by neglecting the explicit chiral symmetry breaking terms, there is just one term in the Chiral Lagrangian, given by:

$$\mathcal{L}_0 = \frac{f_\pi^2}{4} \text{Tr}(\partial_\mu U \partial^\mu U^+),$$

and from this one gets the well known expressions for the pion-pion scattering amplitudes, which were referred to previously as the Low Energy Theorems (LET's) of QCD:

$$T(\pi^+ \pi^- \rightarrow \pi^+ \pi^-) = -\frac{u}{f_\pi^2}, \quad T(\pi^+ \pi^- \rightarrow \pi^0 \pi^0) = \frac{s}{f_\pi^2}.$$

Going to next to leading order in ChPT (we are still ignoring the explicit chiral symmetry breaking terms to make this presentation here as simpler as possible) amounts to consider the next order Lagrangian \mathcal{L}_1 with all the possible chiral invariant terms with four derivatives, i.e. $\mathcal{O}(p^4)$ terms, with the corresponding chiral parameters in front, usually called L_i parameters in QCD. Some of these terms are, for instance:

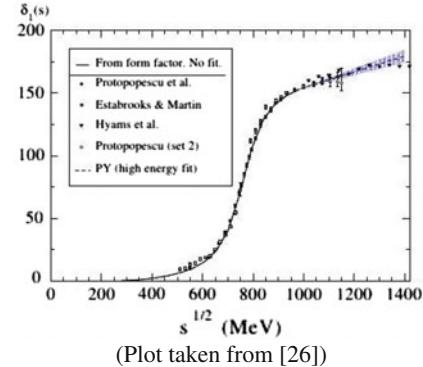
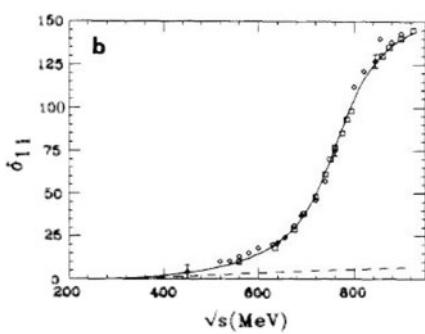
$$\mathcal{L}_1 = L_1 \left[\text{Tr} \left((\partial_\mu U) U^\dagger (\partial^\mu U) U^\dagger \right) \right]^2 + L_2 \left[\text{Tr} \left((\partial_\mu U) U^\dagger (\partial_\nu U) U^\dagger \right) \right]^2 + \dots$$

In practical terms, and in order to perform a one-loop computation of a given observable with the Chiral Lagrangian (CL) of QCD,

$$\mathcal{L}_{\text{CL}} = \mathcal{L}_0 + \mathcal{L}_1 + \dots,$$

by means of Feynman diagrams, one has to compute all the contributing tree-level diagrams from \mathcal{L}_0 and from \mathcal{L}_1 , add the one-loop contributing diagrams generated with the Feynman rules of \mathcal{L}_0 , and finally perform the renormalization of all the entering chiral parameters L_i . All together provide well defined predictions for low energy observables like, for instance, the pion-pion scattering amplitudes, which are given in terms of a finite set of renormalized chiral parameters, usually called L_i^r . These predictions are usually compared with data, and from this comparison one gets the preferred by data values for these L_i^r . This procedure can be done to higher orders in ChPT, i.e. to $\mathcal{O}(p^6)$ with two-loop contributions included, etc, and it can also be generalized to include all the proper explicit chiral symmetry breaking terms into the Lagrangian. The typical dimensionless parameter of the low energy expansion provided by ChPT, once loop contributions are incorporated, is given by $p/(4\pi f_\pi)$. Therefore, one expects a good convergence of this expansion for energies well below $4\pi f_\pi \sim 1,200$ MeV. ChPT has been checked to work pretty well in comparison with data for many years.

In addition, by going to higher orders in ChPT, i.e. $\mathcal{O}(p^4)$ and above, and using either unitarization methods or dispersion relations, the emerging resonances can also be implemented. These are seen as resonant peaks in $\pi\pi \rightarrow \pi\pi$ scattering. For instance, the ρ vector meson appears clearly in the phase shift δ_{IJ} for $I = J = 1$. The next figures illustrate $\delta_{11}(\sqrt{s})$ as a function of the energy \sqrt{s} and we see that $\delta_{11} \simeq 90^\circ$ when $\sqrt{s} = 775$ MeV, signaling clearly the emergent ρ resonance with mass at $m_\rho = 775$ MeV.



12.3 From QCD to Technicolor Theories

The most popular examples of strongly interacting theories that can provide the needed framework for a composite Higgs boson are the Technicolor Models. Generically, these models assume a new $SU(N_{TC})$ gauge symmetry describing the dynamics of new strong interactions in analogy to the usual $SU(3)_C$ gauge interactions of QCD. Also following the guide of QCD, Technicolor assumes the existence of new fermionic constituents of matter, and new gauge bosons that are the intermediate bosons of the new strong interactions among these new constituents. By analogy to QCD these are named respectively as

- New constituents: Techniquarks q_{TC}
- New gauge bosons: Technigluons g_{TC}

And the number of Technicolors is given by N_{TC} .

Regarding the global symmetries, the simplest Technicolor Models also assume that the Lagrangian has the global chiral symmetry of the Electroweak Theory, i.e. the Electroweak Chiral Symmetry, and that it is broken spontaneously to the custodial symmetry group by the techniquark condensate:

$$\langle 0 | \bar{q}_{TC} q_{TC} | 0 \rangle \neq 0 \Rightarrow SU(2)_L \times SU(2)_R \rightarrow SU(2)_{L+R}.$$

Then, following again the guide of QCD, the 3 resulting Goldstone bosons are identified with the 3 Technipions: π_{TC}^\pm and π_{TC}^0 . When the subgroup $SU(2)_L \times U(1)_Y$ is gauged: the three GBs, π_{TC}^\pm and π_{TC}^0 , disappear from the spectrum and they are replaced by the longitudinal gauge bosons, W_L^\pm, Z_L . The EW bosons then get the proper mass by means of the Higgs Mechanism, but in this case without the appearance in the spectrum of an elementary Higgs boson. Notice that preserving the degrees of freedom here does not require the introduction of any extra scalars other than the needed three GBs. Notice also that, as in the SM, the gauging of the EW symmetry breaks explicitly the global EW Chiral symmetry and therefore the would-be GBs are not massless but they acquire masses (M_W and M_Z) that are typically smaller than the masses of the other resonances or composite particles in these Technicolor Models that typically appear at or above 1 TeV.

Regarding the dynamics of these technipions to low energies, one also follows the guide of QCD and uses effective Lagrangians that are based in the Electroweak Chiral Symmetry but applied to the technipions case. Similarly, the coupling of the technipions to the weak current (in analogy to f_π) is given by:

$$\langle 0 | J_L^{+\mu} | \pi_{TC}^-(p) \rangle = \frac{iF_\pi^{TC}}{\sqrt{2}} p^\mu \text{ with } F_\pi^{TC} = v = 246 \text{ GeV}$$

The spectrum of $SU(N_{TC})$ appear as a replica of the QCD spectrum but with all the masses shifted upwards. Thus, there are Technipions (π_{TC}^\pm, π_{TC}^0), Technirhos ($\rho_{TC}^\pm, \rho_{TC}^0$), Technimegas, Technietas, etc..

By using large N techniques one can re-scale QCD quantities to the Technicolor ones. For instance, the ratio between the technimeson mass and the meson mass can be estimated as:

$$\frac{m_{T\text{meson}}}{m_{\text{meson}}} \sim \frac{F_\pi^{TC}}{f_\pi} \cdot \sqrt{\frac{N_C}{N_{TC}}} \text{ with } \frac{F_\pi^{TC}}{f_\pi} = \frac{246 \text{ GeV}}{0.094 \text{ GeV}} \sim 2,700$$

Thus, the first expected resonance is the technirho with a mass and a total width given by:

$$\begin{aligned} m_{\rho_{TC}} &= \frac{F_\pi^{TC}}{f_\pi} \sqrt{\frac{N_C}{N_{TC}}} m_\rho, \\ \Gamma_{\rho_{TC}} &= \frac{N_C}{N_{TC}} \frac{m_{\rho_{TC}}}{m_\rho} \Gamma_\rho. \end{aligned}$$

For example, taking $N_C = 3, N_{TC} = 4, m_\rho = 760 \text{ MeV}$, and $\Gamma_\rho = 151 \text{ MeV}$ lead to: $m_{\rho_{TC}} = 1.8 \text{ TeV}$ and $\Gamma_{\rho_{TC}} = 260 \text{ GeV}$.

In Technicolor Theories the Higgs particle does not appear as an elementary particle but as a composite scalar resonance. Therefore, as any other resonance of Technicolor, the mass of the Higgs boson resonance should be at the $O(1 \text{ TeV})$ energy

scale. Correspondingly, the effective cut-off of Technicolor Theory where the new physics sets in is at:

$$\Lambda_{TC}^{\text{eff}} \sim O(1 \text{ TeV})$$

and therefore there is not hierarchy problem in Technicolor Theories. However, this avenue to solve the hierarchy problem having the simplest implementation of a composite Higgs boson as an emerging resonance in Technicolor Models at $O(1 \text{ TeV})$ is not anymore acceptable, if one identifies this boson with the recently discovered Higgs particle that has a relative low mass at $m_H = 125.6 \text{ GeV}$. There could be however, different implementations of the leading ideas of Technicolor Models that could lead to more compatible with data predictions, even though we do not have yet a quite satisfactory Theory of Technicolor.

Following the sensibilities with QCD, the resonances of Technicolor would then appear in $V_L V_L$ scattering ($V = W, Z$), as the ρ of QCD appears in $\pi\pi$ scattering. In order to describe these scattering processes and other interesting observables that could be measured it is common to use the effective Lagrangians technique. In particular, Technicolor and other Strongly Interacting theories of EWSB can be described generically with effective Electroweak Chiral Lagrangians and with the equivalent to ChPT in this other context.

Present bounds on Technicolor

Although the general motivation for Techicolor Theories is very appealing, it turns out that these theories are very much constrained by past and present data.

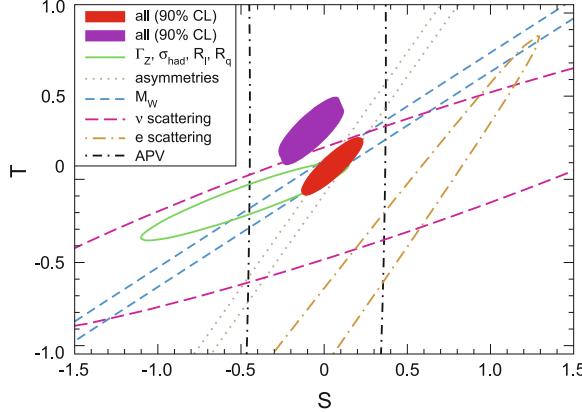
On one hand, Technicolor models when connecting quarks with techniquarks tend to produce too much Flavor Changing Neutral Currents (FCNC). The absence of FCNC in data sets very strong constraints on these models. However, these bounds are very model dependent.

On the other hand, the Electroweak Precision Observables (EWPO) also set very restrictive bounds in these and other models. In particular, the so-called oblique parameters S and T (Peskin, Takeuchi, 1990) that measure possible deviations from the SM predictions due to new physics in self-energies Π_{XY} of EW gauge bosons are the most constraining ones. These are defined by:

$$\hat{\alpha}(M_Z)T \equiv \frac{\Pi_{WW}^{\text{new}}(0)}{M_W^2} - \frac{\Pi_{ZZ}^{\text{new}}(0)}{M_Z^2},$$

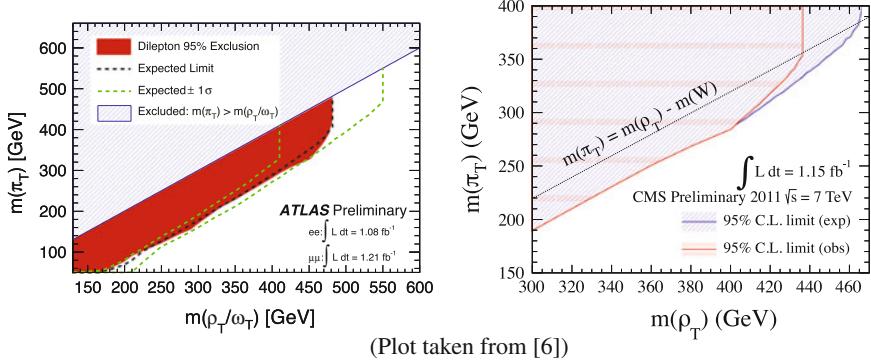
$$\frac{\hat{\alpha}(M_Z)}{4\hat{s}_Z^2\hat{c}_Z^2}S \equiv \frac{\Pi_{ZZ}^{\text{new}}(M_Z^2) - \Pi_{ZZ}^{\text{new}}(0)}{M_Z^2}.$$

Indeed, the present experimental bounds on S and T exclude already many Technicolor models, as can be seen in the next plot.



Particularly, the models that are based on simple scaling from QCD, are already excluded by these data, since each technicolor and each technidoblet contributes to S and one has: $S_{TC} \propto N_{TC} N_D$, for N_{TC} technicolors and N_D technidoblets. This leads to $S_{TC} \sim 0.45$ for $N_{TC} = 4$ and $N_D = 1$. If this is compared with (1σ , 39.35 %): $S_{\text{exp}} = 0.04 \pm 0.09$, we conclude that this simple model is indeed many sigmas away from data!.

Finally, one can also get very restrictive bounds from present colliders data. As we can see in the next plots, both experiments at LHC, ATLAS and CMS, exclude light ρ_{TC} masses from direct searches and from its couplings to standard fermions. Concretely, from the last PDG2012 one gets the exclusion region: $m_{\rho_{TC}} < 260 - 480$ GeV (depending on channels).



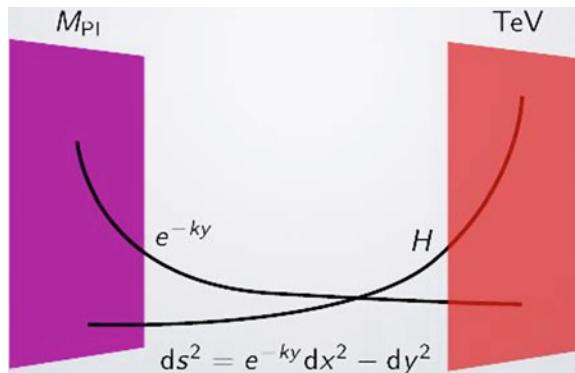
12.4 Composite Higgs in Extra Dimensions

Here we comment shortly on the main ideas underlying some proposals for a composite Higgs boson within the context of theoretical models with extra dimensions. Usually the number of total dimensions is chosen to be five.

To get a light Higgs boson $\mathcal{O}(100 \text{ GeV})$ in theories with extra dimensions, one interesting possibility is to assume that the Higgs boson field is the scalar component of a new gauge field living in five dimensions. Another interesting feature of these models in extra dimensions is that the mass of the Higgs is then protected by gauge symmetry (this happens in Gauge-Higgs Unification Models). More specifically, the Higgs mass is zero at the tree level and a non-zero mass value is generated radiatively at one-loop, in a similar way as in the well known Coleman-Weinberg Model.

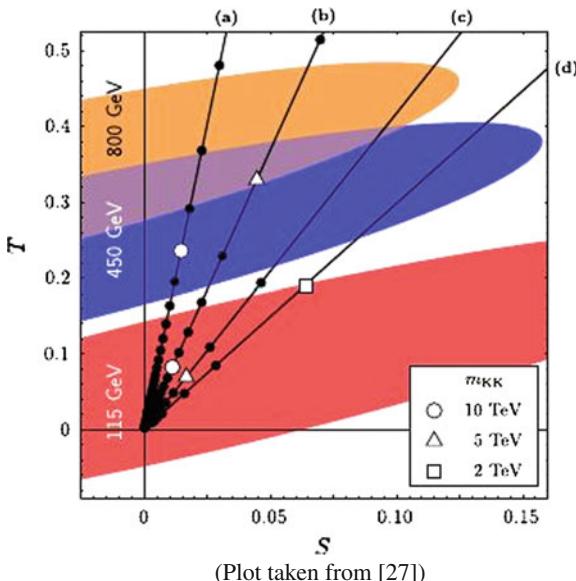
The connection of a light Higgs boson with the physics of strongly interacting theories comes by means of the famous $\text{AdS}_5/\text{CFT}_4$ correspondence that relates weakly coupled theories of gravity (Anti-de Sitter, AdS) in 5 dimensions (5D) with strongly coupled Conformal Field Theories (CFT) in 4 dimensions (4D). More specifically, the breaking of the bulk (5D) gauge group by boundary conditions on the InfraRed (IR) brane is described in the CFT (4D) as the Spontaneous Symmetry Breaking of a global symmetry G into a subgroup H , and this breaking $G \rightarrow H$ occurs by the strong dynamics at the TeV scale. The Higgs boson in 4D is then identified with one of the associated GBs of this Spontaneous Symmetry Breaking (similar to Little Higgs Models).

When the 5th dimension y is compactified and the geometry is warped (Randall Sundrum (RS) Models) the small ratio between the infrared brane at the TeV scale and the Ultraviolet brane at the Planck energy scale, $1 \text{ TeV}/M_{\text{Pl}}$, is explained in terms of the exponential suppression produced by the ‘warp’ factor e^{-ky} , with k being the AdS_5 curvature $\sim \mathcal{O}(M_{\text{Pl}})$.



The main problem of all these models with extra-dimensions is that they are strongly constrained by the EWPO data. It turns out that the typical excitation modes in these models, the so-called Kaluza-Klein (KK) modes that appear due to the

compact extra 5th dimension, contribute dangerously to S and/or T parameters and therefore very stringent lower mass bounds are found leading to very heavy KK masses. The specific mass bounds are model dependent. In the plot shown next: $m_{KK} > \mathcal{O}(10\text{--}5 \text{ TeV})$ for models (a) and (b) with RS metric, and $m_{KK} > \mathcal{O}(5\text{--}2 \text{ TeV})$ for models (c) and (d) with RS-deformed metric. Indeed, these later models are the only ones that allow for a light Higgs having a compatible with data mass value. Generically, the most important restrictions of extra dimension models come from the breaking of custodial symmetry that is very common in these models. Therefore, usually the most realistic models include an additional symmetry in 5D leading to the needed custodial symmetry protection in 4D.



13 Electroweak Chiral Lagrangians

Electroweak Chiral Lagrangians are based on the existence of the electroweak chiral symmetry, the accidental symmetry of the symmetry breaking sector of the SM that has been introduced in these lectures before. The building of these effective field theories rely on specific non-linear implementations of the global symmetry breaking pattern of the electroweak chiral symmetry down to the custodial symmetry:

$$SU(2)_L \times SU(2)_R \rightarrow SU(2)_{\text{custodial}}.$$

And non-linear here means that the three associated GBs of these breaking transform non-linearly under the global $SU(2)_L \times SU(2)_R$ transformations. Once the

subgroup $SU(2)_L \times U(1)_Y$ is gauged in order to include the EW gauge interactions by means of the gauge principle, then both the electroweak chiral symmetry and the custodial symmetry are explicitly broken by the hypercharge gauge interactions with coupling g' . In addition, when including fermions into the EW Theory, also the mass differences between the two fermion components of the $SU(2)_L$ doublets break explicitly the custodial symmetry.

There are many models that one can build with Electroweak Chiral Lagrangians. Here we do not review them, but just for illustration of some of their generic properties, we simply classify them into two main qualitative different classes of models:

- Electroweak Chiral Lagrangians without a light Higgs (ECL)
- Electroweak Chiral Lagrangians with a light Higgs (ECLh)

Obviously, the first ones are not so much interesting at present, however they were proposed first in the literature (before the discovery of the Higgs boson) and they provided already many of the most relevant features of these Non-linear Effective Theories. Indeed, some of these features are common with the other class of models, therefore, it is worth to remind here some of their most relevant properties.

13.1 Electroweak Chiral Lagrangians Without a Light Higgs (ECL)

The simplest Electroweak Chiral Lagrangian (ECL) is similar to the lowest order Chiral Lagrangian for QCD, but with the proper gauging for $SU(2)_L \times U(1)_Y$:

$$\mathcal{L}_{\text{ECL}}^0 = \frac{v^2}{4} \text{Tr} \left[D_\mu U^\dagger D^\mu U \right] + \mathcal{L}_{\text{YM}}$$

where the first term is usually called the gauged non-linear sigma model Lagrangian and the second one, \mathcal{L}_{YM} , is the usual Yang Mills Lagrangian for the electroweak gauge fields. The unitary matrix containing the three GBs of the EWSB, w^+, w^-, w^0 , and its covariant derivative are usually written as:

$$U \equiv \exp \left(i \frac{\vec{\tau} \cdot \vec{w}}{v} \right), \quad v = 246 \text{ GeV}, \quad \vec{w} = (w^1, w^2, w^3)$$

$$D_\mu U \equiv \partial_\mu U + i \frac{g}{2} \vec{W}_\mu \cdot \vec{\tau} U - i \frac{g'}{2} U B_\mu \tau^3$$

where $\tau_i, i = 1, 2, 3$ are the 2×2 Pauli matrices.

When going to next to leading order, as in the case of QCD, one adds to the previous Lagrangian, $\mathcal{L}_{\text{ECL}}^0$, all possible $SU(2)_L \times U(1)_Y$ gauge invariant terms with higher dimension (i.e. with dimension 4 in this case). These new terms must also be invariant under the global $SU(2)_L \times SU(2)_R$ symmetry (and the custodial symmetry) in the same way as in the SM case, that is if the g' and the mass differences

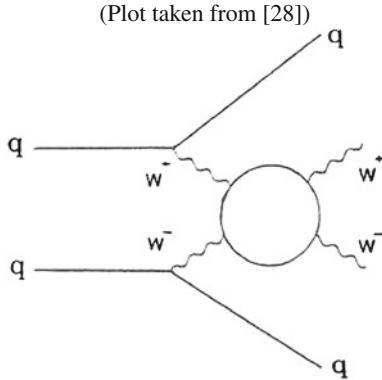
between the two components of the fermion $SU(2)_L$ doublets are set to zero. For instance, in the next to leading order Lagrangian, $\mathcal{L}_{\text{ECL}}^1$, one has the following two terms, among others, that are of $\mathcal{O}(p^4)$ in momentum space:

$$\mathcal{L}_{\text{ECL}}^1 = a_4 \left[\text{Tr} \left((D_\mu U) U^\dagger (D_\nu U) U^\dagger \right) \right]^2 + a_5 \left[\text{Tr} \left((D_\mu U) U^\dagger (D^\mu U) U^\dagger \right) \right]^2 + \dots$$

Notice, that these particular a_4 and a_5 would correspond to L_2 and L_1 respectively of the Chiral Lagrangian for low energy QCD.

Then, for a specific computation of an observable to next to leading order, one includes the tree level contributions from $\mathcal{L}_{\text{ECL}}^0$ and $\mathcal{L}_{\text{ECL}}^1$ and adds the one loop contributions using the Feynman Rules of the lowest order Lagrangian $\mathcal{L}_{\text{ECL}}^0$. Then, finally one defines properly the renormalized ECL parameters, a_i^r , such that all the predictions at the one-loop level are finite and well defined, following a similar effective field theory procedure as in the ChPT of QCD. Previous to the discovery of the Higgs particle at LHC, there were indeed many works in the literature studying phenomenological implications of these kind of Electroweak Chiral Lagrangians and some comparisons with data were also done. By using these effective field theory techniques it is possible, for instance, to make well defined predictions to one-loop level for the scattering amplitudes of longitudinal electroweak gauge bosons in terms of these renormalized a_i^r parameters and then compare the expected events from these processes at high energy colliders, both hadronic ones like LHC etc, and also future e^+e^- colliders like ILC etc for various specific scenarios corresponding to different settings for the numerical values of these a_i^r parameters. The final goal would be obviously to determine which is the underlying fundamental dynamics in the SBS that explains these particular values.

Besides, and also following similar techniques as in the QCD case, one can incorporate the description of resonances with the ECL. It turns out that the two particular previous ECL parameters, a_4 and a_5 , have important consequences for the phenomenology of $V_L V_L$ scattering, in particular for the appearance of a resonant behavior in the scattering amplitudes like, for instance, $W_L^+ W_L^- \rightarrow W_L^+ W_L^-$, $W_L^+ W_L^- \rightarrow Z_L Z_L$, and $W_L^+ Z_L \rightarrow W_L^+ Z_L$. In the next plot we include some examples of possible resonances that could show up in the LHC when the initial quarks radiate these longitudinal electroweak gauge bosons and they re-scatter with amplitudes as computed to one-loop level from the previous ECL with some particular values of a_4 and a_5 and once the proper dispersion relations (or unitarization requirements) have been implemented. One can see in these plots that several possibilities could arise, and several kinds of resonances, scalar, vector, etc, could be seen at LHC with masses around $\mathcal{O}(1 \text{ TeV})$.



13.2 Electroweak Chiral Lagrangians with a Light Higgs (ECLh)

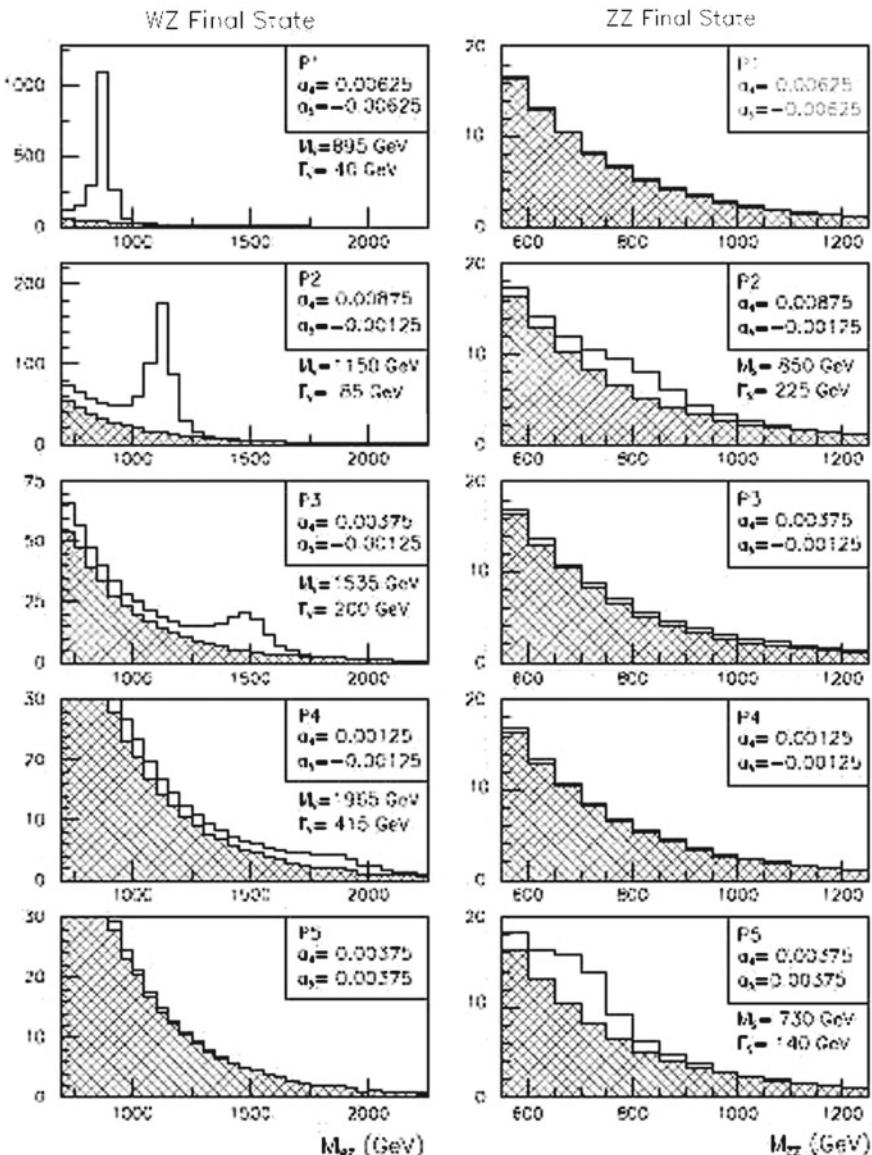
After the discovery of the Higgs particle at the LHC, new Electroweak Chiral Lagrangians have been proposed that include a dynamical Higgs-like particle into the formulation of the EFT. Following a similar procedure as explained in the previous case, one can build non-linear effective Lagrangians that are $SU(2)_L \times U(1)_L$ gauge invariant and that have, in addition to the three GBs, transforming non-linearly under the Electroweak Chiral Symmetry, a new dynamical field $h(x)$ that is a singlet under this symmetry. This field then appears explicitly in the effective Lagrangian interaction terms by means of factor functions $(1 + k_1(h(x)/v)^1 + k_2(h(x)/v)^2 + \dots)$ containing a series of powers of the dimensionless field $(h(x)/v)$.

If one assumes again that the custodial symmetry breaking is as in the SM, and that no large CP violation nor Flavor Violation occur beyond SM, then the simplest Electroweak Chiral Lagrangian with a light Higgs, keeping for instance just up to $\mathcal{O}(h(x)/v)$ terms in these series expansions, and including also fermions, is given by (we use here the first generation quarks notation):

$$\begin{aligned} \mathcal{L}_{\text{ECLh}} = & \frac{v^2}{4} \text{Tr} \left[D_\mu U^\dagger D^\mu U \right] \left(1 + 2a \frac{h}{v} \right) \\ & - \frac{v}{\sqrt{2}} (\bar{u}_L \bar{d}_L) U \left(1 + c^{u,d} \frac{h}{v} \right) (y^u u_R \ y^d d_R)^T + h.c. \\ & + \mathcal{L}_{\text{YM}} + \frac{1}{2} (\partial_\mu h) (\partial^\mu h) \end{aligned}$$

where, \mathcal{L}_{YM} is again the Yang Mills Lagrangian including all the kinetic terms for the $SU(2)_L \times U(1)_Y$ gauge fields, and the coefficients a and $c^{u,d}$, are the relevant parameters of the Electroweak Chiral Lagrangian at this leading order of the non-linear effective field theory. These ECLh parameters are usually treated as phenomeno-

logical parameters, but the final goal with this approach would be to have a list of predictions for these parameters from different underlying theories. Then from the comparison with data, that will set experimentally preferred values for these ECLh parameters one expects to be able to conclude on the preferred fundamental theory underlying the Higgs System. As in the previous ECL case, one can go beyond leading order and include higher dimensional terms also in the ECLh.

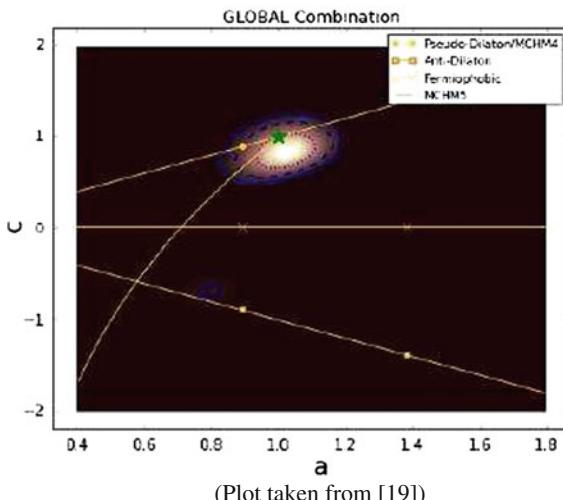


There are recent studies with this Lagrangian $\mathcal{L}_{\text{ECLh}}$ and also comparisons of their predictions at the tree level with present data at LHC and they are already providing some constraints on the values of the c and a parameters that are compatible with data. For instance, a global fit analysis, combining several channels, assuming universality in the fermion parameters, $c_f = c$, and also assuming that no non-SM particles contribute to potential anomalous effective vertices like $(h/v)G_{a\mu\nu}G^{a\mu\nu}$ or $(h/v)F_{\mu\nu}F^{\mu\nu}$ leads to constraints on the (a, c) plane as those shown in the next plot [19]. Here, the more likely regions of this parameter space have lighter shading, and the 68, 95 and 99 % CL contours are indicated by dotted, dashed and solid lines respectively. The yellow lines are the predictions in various models.

We see clearly in this plot, that the space left by data in the (a, c) plane for potential deviations from the SM values, $a_{\text{SM}} = c_{\text{SM}} = 1$, is being reduced considerably with the increasing statistics at LHC, and this kind of analysis indeed sets already constraints on the ECLh parameters.

As in the previous ECL case, one can also include here higher dimensional operators, and in addition there are also terms with higher powers of $(h(x)/v)$. All these have been considered in the literature, but the total number of invariant terms increases considerably, as well as the number of parameters involved in the ECLh and the phenomenological analysis of these models with so many parameters gets very involved.

Besides, as in any of the previously commented non-linear Chiral Effective Field theories , when doing predictions with the ECLh beyond leading order, one has to include not only the the tree level contributions from the previous Lagrangian, but also the one-loop generated contributions from the lowest order terms, and then finally one has to renormalize all the ECLh parameters to provide finite predictions for the observables at one-loop.



This interesting program of building consistent and well defined effective quantum field theories for the Electroweak Theory with the help of Chiral Lagrangians is very

active nowadays and seems promising. In fact there are already some predictions in the literature showing that, not only in the ECL but also in the ECLh case, there could appear new resonances in the scattering of longitudinal W and Z gauge bosons close to the TeV scale, and these could be seen at LHC in the next stage.

Acknowledgments I wish to thank Carlos Merino for inviting me to give these lectures at the 2013 IDPASC school at Santiago de Compostela, for the very efficient organization and for contributing to create the kind atmosphere with the students. I also acknowledge Sven Heinemeyer and Mariano Quiros for interesting discussions and for lending me their slides of previous Lectures on related subjects to some considered here. This work was partially supported by the European Union FP7 ITN INVISIBLES (Marie Curie Actions, PITN-GA-2011-289442), by the CICYT through the project FPA2012-31880 and by the CM (Comunidad Autonoma de Madrid) through the project HEPHACOS S2009/ESP-1473. The work is also supported in part by the Spanish Consolider-Ingenio 2010 Programme CPAN (CSD2007-00042). The author also acknowledges the support of the Spanish MINECO's "Centro de Excelencia Severo Ochoa" Programme under grant SEV-2012-0249.

References

1. S.L. Glashow, Nucl. Phys. **22**, 579 (1961)
2. M. Gell-Mann, Phys. Lett. **8**, 214 (1964)
3. G. Zweig, CERN-TH-412
4. S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967)
5. A. Salam, Conf. Proc. C **680519**, 367 (1968)
6. J. Beringer et al., Particle data group, Phys. Rev. **D86**, 010001 (2012)
7. P.W. Higgs, Phys. Lett. **12**, 132 (1964)
8. P.W. Higgs, Phys. Rev. Lett. **13**, 508 (1964)
9. P.W. Higgs, Phys. Rev. **145**, 1156 (1966)
10. F. Englert, R. Brout, Phys. Rev. Lett. **13**, 321 (1964)
11. G.S. Guralnik, C.R. Hagen, T.W.B. Kibble, Phys. Rev. Lett. **13**, 585 (1964)
12. G. Aad et al., ATLAS collaboration, Phys. Lett. B **716**, 1 (2012). [arXiv:1207.7214](https://arxiv.org/abs/1207.7214) [hep-ex]
13. S. Chatrchyan et al., CMS collaboration, Phys. Lett. B **716**, 30 (2012). [arXiv:1207.7235](https://arxiv.org/abs/1207.7235) [hep-ex]
14. G. Aad et al., ATLAS collaboration, Phys. Lett. B **726**, 88 (2013). [arXiv:1307.1427](https://arxiv.org/abs/1307.1427) [hep-ex]
15. S. Chatrchyan et al., CMS collaboration, JHEP **1306**, 081 (2013). [arXiv:1303.4571](https://arxiv.org/abs/1303.4571) [hep-ex]
16. P. Anderson, Phys. Rev. **130**, 439 (1963)
17. J. Schwinger, Phys. Rev. **125**, 397 (1962)
18. Y. Nambu, Phys. Rev. Lett. **4**, 380 (1960)
19. J. Ellis, T. You, JHEP **1306**, 103 (2013). [arXiv:1303.3879](https://arxiv.org/abs/1303.3879) [hep-ph]
20. J.M. Cornwall, D.N. Levin, G. Tiktopoulos, Phys. Rev. D **10**, 1145 (1974)
21. B. Lee, C. Quigg, H. Thacker, Phys. Rev. D **16**, 1519 (1977)
22. G. Degrassi, S. Di Vita, J. Elias-Miro, J.R. Espinosa, G.F. Giudice, G. Isidori, A. Strumia, JHEP **1208**, 098 (2012)
23. J.D. Wells, *Lectures on Higgs Boson Physics in the Standard Model and Beyond*. [arXiv:0909.4541](https://arxiv.org/abs/0909.4541) [hep-ph]
24. S. Dawson, *The MSSM and Why It Works*. [arXiv:hep-ph/9712464](https://arxiv.org/abs/hep-ph/9712464)
25. A. Dobado, M.J. Herrero, T.N. Truong, Phys. Lett. B **235**, 134 (1990)
26. J.R. Pelaez, F.J. Yndurain, Phys. Rev. D **71**, 074016 (2005). [arXiv:hep-ph/0411334](https://arxiv.org/abs/hep-ph/0411334)
27. J.A. Cabrer, G. von Gersdorff, M. Quiros, Phys. Rev. D **84**, 035024 (2011). [arXiv:1104.3149](https://arxiv.org/abs/1104.3149) [hep-ph]
28. A. Dobado, M.J. Herrero, J.R. Pelaez, E. Ruiz, Morales. Phys. Rev. D **62**, 055011 (2000). [arXiv:hep-ph/9912224](https://arxiv.org/abs/hep-ph/9912224)



Doctoral Student Carlos Naya during his presentation at the Third IDPASC School (picture by C. Merino)

Higgs Searches at the LHC

Cibrán Santamarina Ríos

Abstract An overview of the experimental status of the Standard Model Higgs boson studies at the Large Hadron Collider general purpose detectors, ATLAS and CMS, is presented. This ensemble of analyses has lead to the conclusion that a Standard Model Higgs boson has been found at a mass around $125.5 \text{ GeV}/c^2$. The evidence for a new resonance was announced by CERN in December 12th 2011 and the discovery confirmed in July 4th 2012. In March 14th 2013, CERN declared that the new resonance was a Higgs boson since the measurement of some of its properties, particularly its spin, production and decay branching fractions, shows no disagreement with the theoretical predictions. This concluded with the award of the 2013 Physics Nobel prize to François Englert and Peter Higgs *for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider* announced on October 9th 2013.

1 Introduction

The Standard Model originates in the unification of the electromagnetic and weak interactions proposed by Sheldon Glashow in 1961 [1]. This model, introduces a $SU(2)_L \times U(1)_Y$ symmetric lagrangian, including four Yang-Mills fields, one isospin triplet \mathbf{W} and one singlet B and is governed by the lagrangian

$$\begin{aligned} \mathcal{L} = & -\frac{1}{2}W_{\mu\nu}W^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \bar{\psi}_L\gamma^\mu(i\partial_\mu - g'\frac{Y}{2}B_\mu - g\frac{1}{2}\tau\mathbf{W}_\mu)\psi_L \\ & + \bar{\psi}_R\gamma^\mu(i\partial_\mu - g'\frac{Y}{2}B_\mu)\psi_R, \end{aligned} \quad (1)$$

C. Santamarina Ríos (✉)

Universidade de Santiago de Compostela, Santiago de Compostela, Galiza, Spain
e-mail: cibran.santamarina@usc.es

that describes the different behaviour of left-handed and right-handed fermions. The model was incomplete since it did not resolve the short range of the weak interaction and the absence of four massless exchange bosons. Moreover, it posed problems with unitarity and renormalization.

In 1964 Robert Brout and François Englert [2] and independently Peter Higgs [3] proposed a mechanism for spontaneous symmetry breakup as a possible way to obtain massive Goldstone bosons. This idea was incorporated to Glashow's model by Stephen Weinberg [4] and Abdus Salam [5] adding a term to the lagrangian of (1) given by

$$\mathcal{L}^H = |(i\partial_\mu - g' \frac{Y}{2} B_\mu - g \frac{1}{2} \tau \mathbf{W}_\mu) \Phi^\dagger \Phi|^2 - V(\Phi) \quad (2)$$

where Φ is a weak isospin scalar doublet and $V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2$ is a fourth order potential with multiple minima ($\lambda > 0, \mu^2 < 0$).

The selection of a particular minimum breaks-up the original $SU(2)_L \times U(1)_Y$ symmetry that, after rearranging the interaction fields reveals three massive interaction bosons (W^+ , W^- and Z^0) and only a massless photon, consequence of a conserved $U(1)_{EM}$ symmetry. The accuracy of this picture was demonstrated with the discovery of the weak vector bosons in 1983 by the UA1 and UA2 experiments at CERN [6–9].

The symmetry break-up and the choice of a particular vacuum state of the Higgs doublet produce the absorption of one of the fields into the masses of the weak interaction bosons and the appearance of a single massive scalar field. Additionally, by adding ad-hoc Yukawa terms coupling the Higgs doublet to the fermion fields, the spontaneous symmetry breaking produces mass terms for the fermions. The presence of the Higgs scalar field also restores unitarity at tree level and is necessary for the renormalization of the theory as t'Hooft and Veltman showed [10]. Either a Higgs or something else is needed to explain the nature of the W and the Z bosons and this is at the origin of the Higgs searching program of different experiments in the last decades culminated in the LHC.

Historically Higgs searches only started after the discovery of the W and Z bosons in 1983. But in reality a serious attempt was only possible after the Large Electron-Positron collider (LEP) at CERN was in operation in 1989. The first publication on Higgs searches was published by the ALEPH collaboration [11], looking for a Higgs produced through the Higgs-strahlung mechanism, that will be discussed later, after the Z boson was produced in e^+e^- collisions. The search tried to find decays into a fermion pair and excluded a Higgs boson in the mass range from $32 \text{ MeV}/c^2$ to $15 \text{ GeV}/c^2$ at 95 % confidence level. The first stage of LEP finished in 1995 and the combined analysis made by the four experiments using the beams resulted in an exclusion up to a mass of $65.1 \text{ GeV}/c^2$ [12].

The collider was upgraded in 1995 to enable production of a pair of W bosons, each having a mass of $80 \text{ GeV}/c^2$. LEP collider energy eventually topped at 209 GeV at the end in 2000. This enlarged the Higgs production cross sections and put LEP

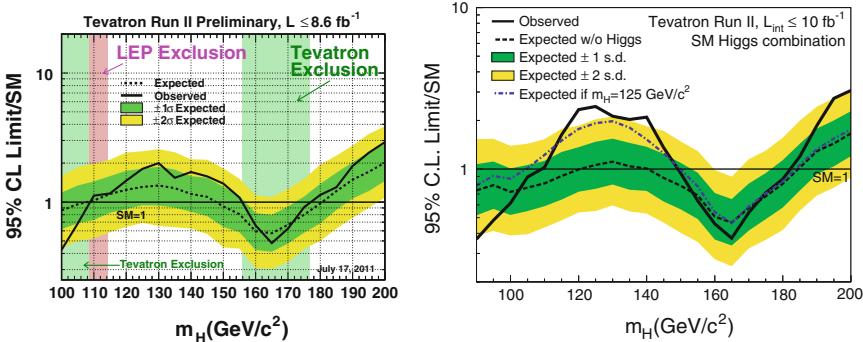


Fig. 1 *Left* Tevatron limits on the SM Higgs in summer 2011 [15]. *Right* The most recent results of the Tevatron experiments [14]

at the limit of discovering a Higgs boson. The final combined LEP result excluded masses up to $114.4 \text{ GeV}/c^2$ at 95 % confidence level [13].

The effort was continued at the Tevatron by the D0 and CDF collaborations. The SM Higgs searches at this collider, placed in Fermilab, near Chicago, are very similar to the ones performed at the LHC since this is a proton-antiproton collider with center-of-mass energy around 1 TeV (hence its name). The Tevatron was close to find the Higgs and actually excluded mass ranges above the resonance found at CERN discussed in the next sections. The Tevatron collaborations (even after the shutdown of the collider in 2011) continue to produce analysis results, the latest published in [14] where they even observe a significant excess of events in the mass range between 115 and $140 \text{ GeV}/c^2$. Even in summer 2011, when ATLAS and CMS made public their first mass limits in the Higgs searches, the Tevatron collaborations had the most stringent results [15]. The historical results of 2011 and the most up to date limits can be seen in Fig. 1.

The Large Hadron Collider, located at the European Laboratory for Particle Physics (CERN), near Geneva, is the highest energy particle accelerator (actually it is a collider, thus two accelerators) ever made by the humankind. It is hosted in a 27 km ring and collides two proton beams of 4 TeV each (3.5 in 2011). The beams are being upgraded to run at more than 13 TeV from early 2015. One of the LHC main goals is to search for the Standard Model Higgs boson and for signatures of new physics beyond this well established model.

In the four interaction points of the Large Hadron Collider four major experiments, LHCb, ALICE, ATLAS and CMS are installed. Two of them, ATLAS and CMS, count with General Purpose Detectors, characterized by almost full angular coverage, excellent particle identification, track reconstruction, and energy determination. The proton bunches of 10^{11} protons interact every 50 ns, the instantaneous top luminosity corresponding to $7.7 \times 10^{33} \text{ cm}^{-2} \text{s}^{-1}$.

This document presents a review of the Standard Model Higgs searches performed at the LHC. Particular emphasis is made in the analysis strategies rather than in the final results, since these evolve rapidly and the quoted numbers will be soon outdated.

2 Higgs Production Modes at the LHC

The inelastic proton-proton collisions at the LHC produce copious amount of particles. Also, the high beam energies enable the creation of very massive and small production cross section particles. One of them could be the Standard Model Higgs boson. Intuitively, the mechanism to produce SM Higgs bosons would be the direct coupling to fermion-antifermion pairs. The coupling constant of the Higgs boson to the fermions is given by

$$g_{Hf} = \sqrt{2} \frac{m_f}{v} \sim 4 \times 10^{-6}$$

where m_f is the mass of the corresponding fermion (in this case a light quark was considered) and v is the vacuum expectation of the Higgs field. However, this direct coupling is much smaller than the coupling of fermions to weak bosons or gluons to heavy quarks (top) and W bosons. The SM Higgs is therefore produced at the LHC mainly via other four main mechanisms: gluon-gluon fusion (ggF), vector (or weak) boson fusion (VBF), higgstrahlung or production associated to a W or Z boson (VH) and production with $t\bar{t}$ pairs ($t\bar{t}H$) that involve vector bosons or top quark couplings to the Higgs. The diagrams for these mechanisms are shown in Fig. 2. The calculation of such production cross sections is beyond the scope of this document but a very complete discussion and references can be found in the document produced by the LHC Higgs Cross Section Working Group [16–18]. As an illustration the results for a $125.5 \text{ GeV}/c^2$ SM Higgs are shown in Table 1 and the Higgs production cross section for a center of mass energy of the proton-proton system $\sqrt{s} = 8 \text{ TeV}$ in Fig. 3.

Since Standard Model Higgs decays almost instantly, an invariant mass reconstruction from its decay products is required. The Standard Model predicts the branching fractions to different decay channels that are dependent on the Higgs mass. Indeed, the signal strength (corresponding to the product of the cross section times the branching fraction) of every mode is one of the main observables to test the Standard Model predictions. The SM Higgs preferred decay modes are given by the decays into a fermion anti-fermion pair and into a pair of Vector Bosons (WW or ZZ). At leading order the Higgs into leptons branching fraction is given by [19].

$$\Gamma(H \rightarrow l^+ l^-) = \frac{G_F M_l^2}{4\sqrt{2}\pi} M_H \beta_l^3, \quad \beta_l = \sqrt{1 - 4M_l^2/M_H^2} \quad (3)$$

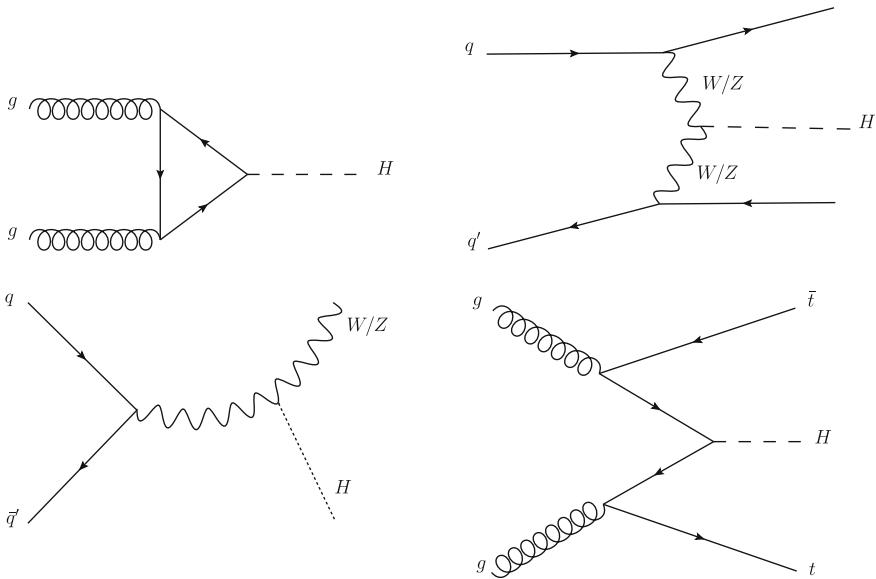


Fig. 2 Diagrams for the main four Higgs production mechanisms at the LHC

Table 1 Standard Model Higgs production cross sections in the four main modes for a Higgs mass of 125.5 GeV/c²

Production mechanism	σ (pb) at $\sqrt{s} = 7$ TeV	σ (pb) at $\sqrt{s} = 8$ TeV
Gluon-gluon fusion	$15.01^{+1.56}_{-1.58}$	19.12 ± 1.99
Vector boson fusion	$1.219^{+0.307}_{-0.259}$	$1.573^{+0.041}_{-0.044}$
WH Higgs-strahlung	0.5703 ± 0.0159	0.6951 ± 0.0181
ZH Higgs-strahlung	0.3309 ± 0.0129	0.4102 ± 0.0132
$t\bar{t}$ Associated production	$0.08528^{+0.00767}_{-0.01069}$	$0.1277^{+0.0114}_{-0.0157}$

The results are the most precise calculations to date given by the dedicated working group [16]

where G_F is the Fermi constant, M_l is the lepton mass and M_H is the Higgs boson mass. At the same time, the decay into a quark-antiquark pair is given by a very similar relation,

$$\Gamma(H \rightarrow q\bar{q}) = 3 \frac{G_F M_q^2}{4\sqrt{2}\pi} M_H \beta_q^3 \left(1 + \frac{4}{3} \frac{\alpha_s}{\pi} \Delta_H^{QCD} \right) \quad (4)$$

where two main factors arise, first the scaling by 3 due to the color, and a significant QCD correction. Nevertheless, the main conclusion from (3) and (4) is that, considering only fermions, the M_f^2 factor dominates the branching fraction and therefore the Higgs boson predominantly decays to the heaviest available fermion. For Higgs

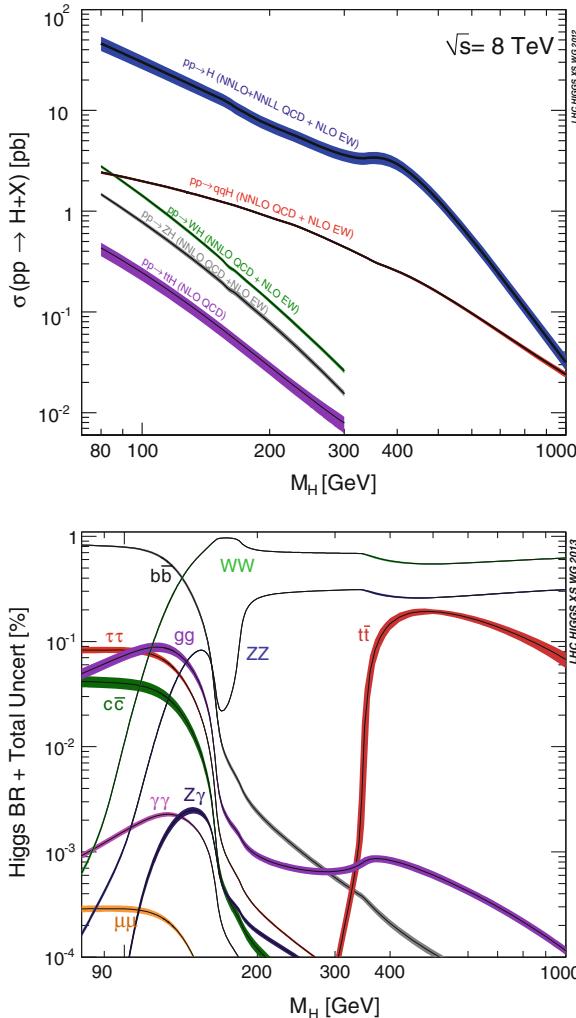


Fig. 3 *Top* Higgs boson production cross section as a function of its mass. *Bottom* Branching fraction of the main decay modes of the Standard Model Higgs boson as calculated in [16]

masses well below the top mass this happens to be the b-quark, followed by the τ lepton and the c-quark.

At masses around $115\text{ GeV}/c^2$ the mode $H \rightarrow WW^*$, where W^* indicates a virtual W starts to be significant. The same happens for the $H \rightarrow ZZ^*$ decay at slightly higher masses. The branching fraction for these channels is proportional to m_H^3 and dominates for masses above $135\text{ GeV}/c^2$.

Other than direct decays into fermions and vector bosons, loop-mediated decays into two gluons, two photons or into a $Z\gamma$ pair are also important. The $H \rightarrow \gamma\gamma$

mode, although with a small branching fraction of 2.3×10^{-3} for a Higgs mass of $125.5\text{ GeV}/c^2$, is very important since benefits from a reduced background in the context of proton-proton collisions, as compared to $H \rightarrow b\bar{b}$, $H \rightarrow c\bar{c}$ and $H \rightarrow gg$ modes.

The most precise calculations of the branching fractions, provided by [16], are also shown in Fig. 3. Having that in mind, and before entering into the details, it is worth mentioning some facts. Even before the discovery of a scalar resonance at $\sim 125.5\text{ GeV}/c^2$ three were the main decay modes to tackle the Higgs search: $H \rightarrow \gamma\gamma$ in the low mass range ($m_H \simeq 130\text{ GeV}/c^2$) and the $H \rightarrow WW^*$ and $H \rightarrow ZZ^*$ in higher masses. The reason for this is that these three channels, considering leptonic decays of at least one of the vector bosons in the final state ($Z \rightarrow l^+l^-$ and $W^- \rightarrow l^-\bar{\nu}$), contain either two large transverse momentum photons or a large transverse momentum lepton in the final state and therefore suffer from significant less background than channels such as $H \rightarrow b\bar{b}$ (and $H \rightarrow c\bar{c}$) that only present jets (although heavy flavor jets) in the final state and are much more difficult to discriminate from the overwhelming background of di-jet events produced in the context of the LHC. A different story is the $H \rightarrow \tau\tau$ decay since τ reconstruction is a category on its own.

These channels are the subject of the following sections where an overview of the key features of the subdetectors involved is also presented. For the $H \rightarrow \gamma\gamma$ a description of the electromagnetic calorimeters of the LHC general purpose detectors is included. For the $H \rightarrow ZZ$ a brief discussion of the tracking system is made. Finally, for the $H \rightarrow WW$ the hadronic calorimeters are presented as well as the performance of the missing transverse energy reconstruction, and for the $H \rightarrow b\bar{b}$ the jet reconstruction performance is briefly described. The aim is not to present an exhaustive description but just the main ideas behind the search strategies and analysis results. For additional details the reader will be pointed to the relevant bibliography.

3 The $H \rightarrow \gamma\gamma$ Decay

For low mass SM Higgs, the decay into two photons has the largest sensitivity at the LHC. This is something that was historically known and in documents such as the ATLAS CSC-book [20] this channel was seen as the discovery mode for a Higgs with masses below $130\text{ GeV}/c^2$. There are two reasons for such a suppressed mode to be so relevant. The first one is that it suffers from reduced background since it produces a rare signature composed of isolated high transverse momentum (p_T) photons. The second is that it is feasible to reconstruct the momentum of these photons quite accurately and produce a mass narrow peak. For that an excellent performance of the electromagnetic calorimeters. A big challenge for the performance of the calorimeter is the existence of significant pile-up of events. In 2012 the ATLAS experiment measured a mean of 20.7 inelastic proton-proton interactions per bunch crossing [21]. This means that, whenever a signal event happens, there is additional activity superimposed in the

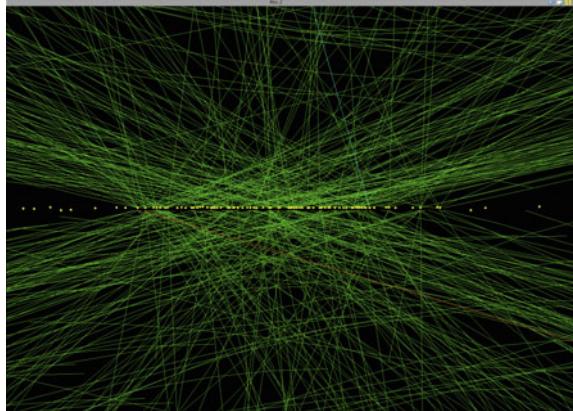


Fig. 4 A CMS event with 78 collisions in the bunch crossing. Taken from [22]. *Image credit Andre Holzner*

calorimeter due to other interactions. An example of an event with large pile-up (up to 78 vertices) is shown in Fig. 4. Of course, many of the secondary tracks do not have enough momentum to get through the spectrometer solenoids but some of them do and reach the calorimeters degrading their reconstruction performance.

The invariant mass of the two photon system dependents on their energy (E_1 and E_2) and on their opening angle (α) as,

$$m_{\gamma\gamma}^2 = 2E_1E_2(1 - \cos \alpha). \quad (5)$$

The determination of the photon energy is limited by the energy resolution of the calorimeter. Ultimately, the limit for the energy resolution of a calorimeter is determined by fluctuations in the development of EM showers and also by instrumental and calibration limits. The basic phenomena in showers are statistical processes, and the intrinsic limiting accuracy, expressed as a fraction of total energy, improves with increasing energy as

$$\left(\frac{\sigma}{E}\right)_{fluct} = \frac{a}{\sqrt{E}}. \quad (6)$$

Over much of the energy range this is the dominating term. The contribution due to instrumental effects is (noise, pedestal) depends on the energy as

$$\left(\frac{\sigma}{E}\right)_{instrm} = \frac{b}{E}. \quad (7)$$

This component may limit the low-energy performance. The third component is due to calibration errors, non-uniformities and non-linearity in photomultipliers, proportional counters, ADC's, etc. This contribution is energy-independent,

$$\left(\frac{\sigma}{E}\right)_{\text{syst}} = c. \quad (8)$$

The three components are added in quadrature¹:

$$\left(\frac{\sigma}{E}\right) = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \quad (9)$$

For the LHC General Purpose Detectors the a , b and c coefficients are $a = 0.1 \sqrt{\text{GeV}/c^2}$, $b = 0.07$ and $c = 0.2 \text{ GeV}/c^2$ for ATLAS and $a = 0.027 \sqrt{\text{GeV}/c^2}$, $b = 0.055$ and $c = 0.16 \text{ GeV}/c^2$ for CMS. Therefore the CMS spectrometer, which is based in the use of PbWO₄ scintillating crystals with fine lateral segmentation, has an advantage in this particular feature due to the design of its EM calorimeter. However, ATLAS, that uses a Liquid Argon sampling calorimeter takes advantage of its excellent lateral and longitudinal segmentation. The first sampling of the calorimeter is particularly finely segmented and this provides additional precision, when combined with the inner detector information, to assign the primary vertex from which the signal candidate originates, allowing a significant improvement of the energy reconstruction.

There are three main background sources for the $H \rightarrow \gamma\gamma$ mode. The first one is the continuous QED diphoton background. This source is irreducible. The second source of background consists of events containing a photon and a parton in the final state, the parton producing a jet (γj events). Finally, events with two jets in the final state (jj events) are copiously produced at the LHC. The last two background categories can be reduced mainly by using shower-shape discriminating variables. The shower produced in the calorimeter by a jet is often wider and less contained than the shower produced by an electron or a photon. The ATLAS EM calorimeter is particularly designed to provide rejection of jets containing a leading π^0 . Since these particles decay into two close photons, the first sampling of the calorimeter has a very fine pseudo-rapidity granularity of $\Delta\eta = 0.0031$. A typical misidentification factor of a jet as a photon is $\sim 2,000$, meaning that one out of 2,000 jets is misreconstructed as a photon.

The analysis of $H \rightarrow \gamma\gamma$ divides the events into mutually exclusive categories with different mass resolutions and signal-to-background ratios. It is worth mentioning that, in 20 % of the cases, a high energy photon undergoes a conversion into an e^+e^- pair. These events are recovered with specific reconstruction tools. If recovery is not possible an electron veto is applied. The events categorization also responds to the different Higgs production mechanism. Gluon-gluon fusion should not contain additional objects (other than the underlying event from the proton fragments), VBF events contain two well separated high transverse momentum jets, Higgs-strahlung events might contain a high transverse momentum lepton or jets from the W or Z decay and events from associated production contain the signature of a $t\bar{t}$ pair (either one or two high transverse momentum leptons and b-jets). If a leptonic

¹ The symbol \oplus indicates: $a \oplus b = \sqrt{a^2 + b^2}$.

Table 2 Number of events in the data (N_D) and expected number of SM Higgs signal events (N_S) for $m_H = 126.5\text{ GeV}$ from the $H \rightarrow \gamma\gamma$ analysis of the ATLAS experiment, for each category in the mass range 100–160 GeV at $\sqrt{s} = 8\text{ TeV}$

\sqrt{s}	8 TeV/c						
Category	N_D	N_S	$gg \rightarrow H$ (%)	VBF (%)	WH (%)	ZH (%)	ttH (%)
Unconv. central, low pTt	10,900	51.8	93.7	4.0	1.4	0.8	0.2
Unconv. central, high pTt	553	7.9	79.3	12.6	4.1	2.5	1.4
Unconv. rest, low pTt	41,236	107.9	93.2	4.0	1.6	1.0	0.1
Unconv. rest, high pTt	2,558	16.0	78.1	13.3	4.7	2.8	1.1
Conv. central, low pTt	7,109	33.1	93.6	4.0	1.3	0.9	0.2
Conv. central, high pTt	363	5.1	78.9	12.6	4.3	2.7	1.5
Conv. rest, low pTt	38,156	97.8	93.2	4.1	1.6	1.0	0.1
Conv. rest, high pTt	2,360	14.4	77.7	13.0	5.2	3.0	1.1
Conv. transition	14,864	40.1	90.7	5.5	2.2	1.3	0.2
Loose high-mass two-jet	276	5.3	45.0	54.1	0.5	0.3	0.1
Tight high-mass two-jet	136	8.1	23.8	76.0	0.1	0.1	0.0
Low-mass two-jet	210	3.3	48.1	3.0	29.7	17.2	1.9
E_{miss}^T significance	49	1.3	4.1	0.5	35.7	47.6	12.1
One-lepton	123	2.9	2.2	0.6	63.2	15.4	18.6
All categories	118,893	395.0	88.0	7.3	2.7	1.5	0.5

The statistical uncertainties in N_S are less than 1 %. The fractions of expected signal events from the $gg \rightarrow H$, VBF, WH, ZH, ttH processes are detailed. Taken from [24]

decay of a W occurs in the event it also presents missing transverse energy (\cancel{E}_T). The existence of additional objects is superimposed to the common requirement of two large transverse momentum photons with large invariant mass. The results of the ATLAS analysis for 2012 data are shown in Table 2. CMS uses a selection based on multi-variate discrimination, also categorizing events [23].

The results of the different categories are combined to obtain a single signal significance. The p-value is the probability that a background-only scenario produces a fluctuation that reproduces the data. The statistical significance is related to the p-value as

$$p = 2(1 - \Phi(n)), \quad (10)$$

where Φ is the cumulative distribution function of the standard normal distribution. With the accumulated data of 2011 and 2012 ATLAS [24] and CMS [23] obtain clear evidence of a resonance decaying into two photons with a significance of 7.4σ and 4.1σ . With the look elsewhere effect these significances reduce to 6.1^2 and 3.2σ . Quoting Lyons [25]: *while the chance of obtaining a 5σ effect in a particular bin is very small, it is to be remembered that histograms contain many bins, they could be*

² With the analyzed statistics the expected significance for a SM Higgs was 2.9σ .

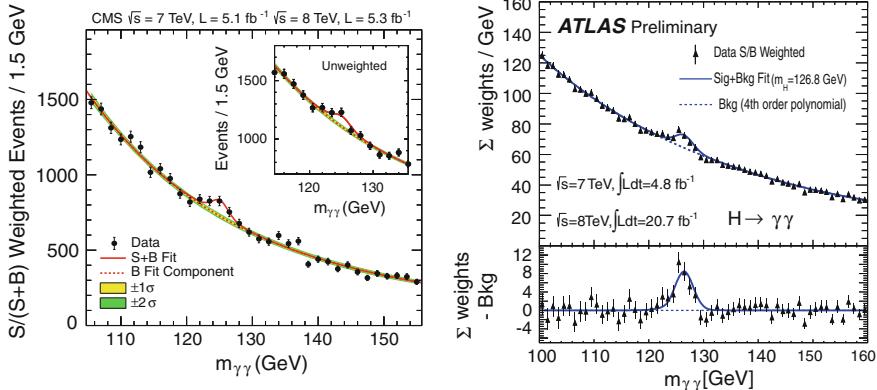


Fig. 5 Invariant mass distributions of the $\gamma\gamma$ spectrum analyzed by the CMS [27] and ATLAS [24] experiments. The CMS events are weighted as $S_i/(S_i + B_i)$ and the ATLAS events as $\log(1 + S_i/B_i)$ with the expected signal and background events of each of the i categories in which the sample is divided

plotted with different selection criteria and different binning, and there are very many other histograms that could be plotted with different selection and different binning. In Higgs boson searches the procedure is to compare the selected events sample to a model consisting of a background plus a signal: $\mu S + D$, where μ is the signal strength and the mass is the only free parameter. In the absence of a signal this parameter is equal to zero. The look elsewhere effect reflects the impact in the significance of the appearance of a signal anywhere in the signal range as a consequence of the freedom of the nuisance parameter. This can be evaluated in different ways, a common one being the use of Monte-Carlo experiments [26]. Figure 5 shows a plot where the different event categories are combined for a convenient display of the data according to their expected signal and background.

As the usual consensus to claim for a new discovery is to have a signal significance above 5σ , this mode is enough to claim for a discovery in ATLAS, which does not necessarily mean it is a SM Higgs boson. ATLAS has determined its mass: $126.8 \pm 0.2(\text{stat}) \pm 0.7(\text{syst}) \text{ GeV}/c^2$. Other than that, CMS establishes an upper limit of $6.9 \text{ GeV}/c^2$ in its natural width at 90 % confidence level, to be compared to the SM prediction (< 5.9 at 95 % CL).

3.1 Other Properties of the New Resonance

The only way to determine the nature of a particle is by measuring its properties. The first property to be determined, already mentioned above, is the resonance mass. This is a crucial piece of information since, if this particle is the SM Higgs boson, this is the last unknown parameter of the most successful theory of elementary particle

physics. Once the mass is obtained many observables can be predicted within the SM. The first one is the signal strength. This is the amount of events expected in the analysis conditions. This number results of mixing properly the integrated luminosity, the production cross sections, the $H \rightarrow \gamma\gamma$ branching fraction and the event selection of the analysis. The signal strength is presented normalized to the SM expectation, and therefore a signal strength of one means that the signal matches the theoretical prediction. For a $126.8 \text{ GeV}/c^2$ mass the signal strength measured by ATLAS is $\mu = 1.65 \pm 0.24(\text{stat})^{+0.25}_{-0.18}(\text{syst})$ times the SM expectation [24]. This is a little too high. The CMS collaboration finds a mass of $125.4 \pm 0.5(\text{stat}) \pm 0.6(\text{syst})$ and a signal strength of $0.78^{+0.28}_{-0.26}$ [23]. It is too early to draw conclusions out of the disagreements between the two experiment results (that are reasonably compatible) and of possible tensions with the SM. However, it is illustrative of the convenience of having two different experiments involved in the same physics program.

The LHC experiments have also measured the signal strengths for the different Higgs production modes (or a combination of them). ATLAS finds [24]: $\mu_{ggF+ttH} = 1.6^{+0.3}_{-0.3}(\text{stat})^{+0.3}_{-0.2}(\text{syst})$, $\mu_{VBF} = 1.7^{+0.8}_{-0.8}(\text{stat})^{+0.5}_{-0.5}(\text{syst})$ and $\mu_{VH} = 1.8^{+1.5}_{-1.3}(\text{stat})^{+0.3}_{-0.3}(\text{syst})$, where the gluon-gluon fusion (ggF) and the associated production (ttH) are added together. CMS finds, in a combined fit, central values of $(\mu_{ggF+ttH}, \mu_{VBF+VH}) = (0.52, 1.48)$ [23].

The Standard Model predicts spin 0 for the Higgs boson and positive parity. The spin of a particle can be determined from the angular distribution of the particles into which it decays. As the observed resonance decays into two photons, that have spin 1 and odd parity, one thing is for sure, that the discovered particle is a boson. The spin information is extracted from the distribution of the absolute value of the cosine of the polar angle θ^* of the photons with respect to the z-axis of the Collins-Soper frame [28]

$$|\cos \theta^*| = \frac{|\sinh(\Delta\eta^{\gamma\gamma})|}{\sqrt{1 + (p_T^{\gamma\gamma}/m_{\gamma\gamma})^2}} \frac{2p_T^{\gamma 1} p_T^{\gamma 2}}{m_{\gamma\gamma}^2}. \quad (11)$$

The distribution of the data, shown in Fig. 6, favors the 0^+ hypothesis over other alternatives [29]. This will be completed with the analysis of other decay modes. The result is very relevant since the SM Higgs boson is the only elementary particle of the model with spin 0. Of course, the elementary (non-composite) nature of the particle should be elucidated before identifying the discovered resonance with the elementary SM Higgs, but this is, nonetheless, remarkable.

ATLAS has also tackled the study of the differential distributions of relevant variables, such as the p_T and rapidity of the $\gamma\gamma$ system and the p_T of the accompanying jet. Within the experimental and theoretical uncertainties, no significant deviation from the SM expectation is observed [30].

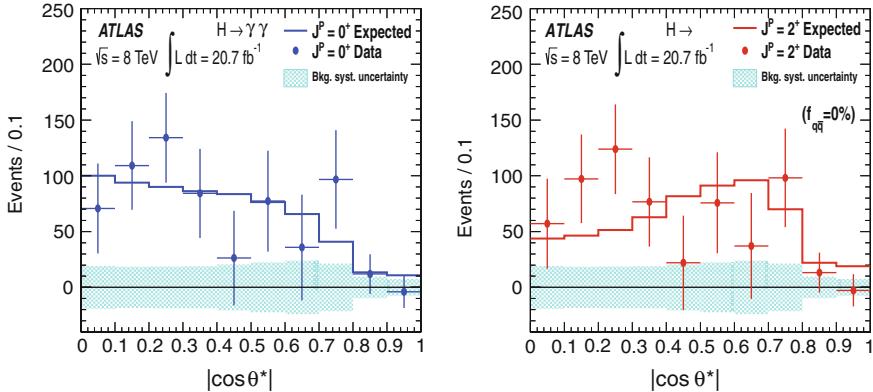


Fig. 6 $\cos \theta^*$ angle distribution and comparison of the spin-parity hypothesis 0^+ and 2^+ with the background subtracted experimental distribution that clearly favor the first possibility. Taken from [29]

4 The $H \rightarrow ZZ^*$ Mode

The SM predicts a significant branching fraction of $(2.76^{+0.12}_{-0.11}) \times 10^{-2}$ [16] for a Higgs with a mass of $125.5 \text{ GeV}/c^2$ decaying into a pair of Z bosons, one of them being out of the mass shell³. The cleaner way to detect this decay is by requiring that both Z bosons decay into two leptons (either muons or electrons). The signature of such events is the dream of the experimental particle physicist since the final state consists of two pairs of leptons, one with invariant mass at the Z resonance, the other with large invariant mass and the four of them combining to the Higgs mass. There are three sources of background for this channel. One of them is irreducible and consists of ZZ^* , $\gamma\gamma$ or $Z^{(*)}\gamma$ diboson production. The second source are $Z+jets$ events, where the jets originate from $b\bar{b}$ or light parton pairs. Finally, $t\bar{t}$ events, where the two W s and the two b jets from the $t\bar{t}$ decay pass the tight selection criteria, also constitute a background. The $Z+jets$ and $t\bar{t}$ background sources can be reduced requiring the impact parameter of each lepton along the beam axis to be within 10 mm of the reconstructed primary vertex⁴. Leptons from $t\bar{t}$ and $Zb\bar{b}$ backgrounds are most likely to originate from decays of charmed or beauty hadrons. These decays are mediated by the weak interaction and have larger lifetimes and often decay after flying a significant distance that can be measured with the tracking detectors. Consequently, these particles have larger transverse impact parameters and further rejection of these backgrounds can be achieved by applying a cut on the impact parameter significance of the tracks associated to the two softest leptons.

³ Could not be otherwise since $2 \times M_Z = 182.4 \text{ GeV}/c^2 > 125.5 \text{ GeV}/c^2$.

⁴ This cut is not applied for muons in the forward region without an inner detector track.

4.1 The Inner Tracking Detectors

At the LHC thousands of particles emerge from the collision point every 50 ns. This creates a large track density in the inner detector. A high momentum and vertex resolution are required by the analysis of physics processes such as the search for the $H \rightarrow ZZ$. These resolution standards are met with arrays of silicon pixel and silicon strip detectors, both in the ATLAS [31] and CMS [32] spectrometers. The ATLAS detector is also equipped with a system of straw tubes, the Transition Radiation Tracker (TRT), that provides additional electron identification by measuring the transition radiation of a particle that crosses the boundary between two different media. To give some numbers, CMS counts with 25,000 silicon strip sensors that cover a surface of 210 m^2 . They are equipped with 75,000 APV chips that provide the readout of 9,600,000 channels. The ATLAS TRT consists of 353,536 4 mm-diameter Kapton straws filled with $\text{Xe}/\text{CO}_2/\text{O}_2$ gas, the Semi-Conductor Tracker is organized in 4 barrel layers and 9 disks per end-cap. This means 4,088 modules with 6.3 M channels. Finally, the pixel detector is organized in 3 barrel layers, 2×3 -layer end-cap disks. This amounts to a total of 1,744 pixel modules and 80M channels.

In the CMS detector the charged particles evolve under the action of a 3.8 T axial magnetic field, this is to be compared with the ATLAS figure of 2 T.⁵ The nominal momentum resolution is typically 0.7 (1.5) % at 1 (100) GeV/c in the central region and the impact parameter resolution for high-momentum tracks is typically $10\text{ }\mu\text{m}$ whereas for ATLAS the momentum resolution is 1.3 (3.8) % and the IP $\sim 11\text{ }\mu\text{m}$ [33].

A remark must be made concerning the electron reconstruction. Whereas, as commented before, photons undergo pair conversions, an almost reverse reaction can also happen and electrons suffer from bremsstrahlung, where a high momentum electron (positron) interacts with the electric field of an atomic nucleus of the detector material radiating a photon. The Bremsstrahlung by electrons was described 70 years ago by Bethe and Heitler [34]. In the bremsstrahlung process the electron mostly retains its direction of propagation while keeping a fraction z of its energy. The energy fraction probability density $f(z)$ depends on the amount of traversed material

$$f(z) = \frac{(-\ln z)^{c-1}}{\Gamma(c)}, \quad c = (x/X_0)/\ln 2 \quad (12)$$

where X_0 is the radiation length of the material. Dedicated methods are used to recover events that suffer bremsstrahlung. In ATLAS they allow a better matching between the silicon and TRT track segments and the recovery of $\sim 12\text{ }%$ of the events.

⁵ ATLAS also counts with a toroid magnet that makes its muon spectrometer unique.

4.2 Results on the $H \rightarrow ZZ^*$ Mode

The analysis of high invariant mass events containing four large transverse momentum leptons, with one of the pairs resonating in the Z boson invariant mass and the other in invariant masses above $12\text{ GeV}/c^2$ gives in CMS [35] an observation with local significance of 6.7 standard deviations above the expected background. Its mass is measured to be $125.8 \pm 0.5(\text{stat}) \pm 0.2(\text{syst})\text{ GeV}/c^2$. The analysis categorizes the events into two classes: events not accompanied by jets and accompanied by two large transverse momentum separated jets. This classification aims at observing Higgs bosons produced by means of gluon-gluon fusion or VH mechanisms in the first case and by the VBF or $t\bar{t}H$ mechanisms in the second. The signal strength relative to the expectation for the standard model Higgs boson at the measured mass is measured to be $\mu_{gg+t\bar{t}H} = 0.91^{+0.30}_{-0.24}$ for events of the first category and $\mu_{VBF+VH} = 1.22^{+0.84}_{-0.57}$ for events of the second category. No other significant standard model Higgs-like excess is found and the range $130\text{--}827\text{ GeV}/c^2$ is excluded with 95 % confidence level. The angular analysis of the events disfavors non Standard Model hypotheses for the spin-parity of the observed resonance, confirming the $H \rightarrow \gamma\gamma$ results.

Similar results are reported by the ATLAS collaboration [36] where a signal significance of 6.6 standard deviations is achieved, and a mass of $124.3^{+0.6}_{-0.5}(\text{stat})^{+0.5}_{-0.3}(\text{syst})\text{ GeV}/c^2$ is measured. The corresponding signal strengths are $\mu_{gg+t\bar{t}H} = 1.8^{+0.8}_{-0.5}$ and $\mu_{VBF+VH} = 1.2^{+3.8}_{-1.4}$. The Higgs-like boson is found to be compatible with the spin-parity SM expectation of 0^+ whereas the 0^- and 1^+ states are excluded at the 97.8 % confidence level.

More limited is the result of the analysis of $H \rightarrow ZZ^*$ events where the resonating Z decays into two τ leptons. The difficulty to reconstruct a τ is quite challenging (as will be shown in Sect. 7). CMS has tackled this project and found no significant signal.

The study of the $H \rightarrow ZZ^*$ is complemented with the search of $H \rightarrow ZZ$ decays where both Z bosons are on shell. These searches can also be made with one of the Z bosons decays into a pair of quarks $Z \rightarrow q\bar{q}$. This gives additional sensitivity to high mass regions (above twice the mass of the Z boson) to discard heavy Higgs bosons since, although everything points to the new resonance to be the SM Higgs, there are models that predict a more complex Higgs sector containing additional particles coupling to vector bosons [18]. The $Z \rightarrow q\bar{q}$ branching fraction is about 20 times that of $Z \rightarrow l^+l^-$ and this makes this channel more sensitive to large mass Higgs-like decays. Both ATLAS and CMS have performed searches in the mass range above $230\text{ GeV}/c^2$ (exploiting the flavor tagging looking for $Z \rightarrow b\bar{b}$ decays) without finding evidence of a boson up to $600\text{ GeV}/c^2$ [37]. CMS has also studied the $H \rightarrow ZZ$ with one of the Z s decaying into neutrinos that helped to exclude SM Higgs regions in the high mass range [38].

A remark must be also made with regard to invisible Higgs decays. These are possible in the SM through the $H \rightarrow ZZ^{(*)}$ ⁶ channel followed by $Z^{(*)} \rightarrow \nu\bar{\nu}$ decays but their yield would be tiny at the LHC. However, alternative models predict a much larger branching fraction into invisible particles that would be a clear signature of physics beyond the SM. The study of invisible decays of a particle is very challenging and is tackled by requiring the HZ production mode. The main backgrounds are $ZZ \rightarrow 2l2\nu$ (almost irreducible) and $WZ \rightarrow vlll$ where the lepton from the W is not detected. CMS finds at 95 % CL an upper limit on the branching fraction of the Higgs boson to invisible particles of 64 % [39] and ATLAS of 60 % [40]. CMS finds a limit of 75 % using the VBF production mode [41].

5 The $H \rightarrow WW^*$ Decay Mode

This is the sister mode of the $H \rightarrow ZZ^*$ decay. The coupling essentially goes with m_V^3 and this favours the ZZ^* mode, but since one of the two bosons is off shell and the W meson is lighter than the Z meson this channel has a larger branching fraction and would be the main decay mode for a Higgs boson with a mass above 135 GeV/c². This is the most sensitive channel for Higgs searches over the 125–180 GeV/c² mass range. For trigger and selection purposes at least one of the W s (and in the main analysis the two) is (are) required to decay leptonically, either as $W \rightarrow \mu\nu_\mu$ or $W \rightarrow e\nu_e$. This makes the analysis of the mode quite challenging. Since the neutrino cannot be reconstructed, only information through the missing transverse energy of the event gives a crude estimation of its transverse momentum. This means that there is no mass peak and the signal is to be searched for in other variable distributions (or in a combination) and there are more important systematic effects, mainly due to the transverse energy scale. The signatures of the signal, if both W s are required to decay leptonically, are: 2 isolated opposite-sign leptons and large transverse missing energy. Additionally the analysis on jet multiplicities (0, 1, 2-jet bins) provides information about the production mode. The main background sources are: WW^* , $t\bar{t}$, $Z + jets$ and $W + jets$ events. The events containing a Z boson can be suppressed by applying a veto on the dilepton mass (except for inefficiencies) and the $t\bar{t}$ by applying a veto to the presence of b-jets in the event. Also, some topological cuts against “irreducible” WW background are applied: cuts on the p_T of the pair of leptons, their invariant mass and their axial angle difference ($\Delta\phi_{ll}$). Also the transverse missing energy in the event and the transverse mass of the lepton pair are used to discriminate the signal from the background (Fig. 7).

The analysis of the $H \rightarrow WW^*$ decay mode relies on the measurement of missing transverse energy \cancel{E}_T . The two LHC proton beams collide with their momenta aligned in the horizontal direction (often referred to as z). Therefore the E/p in the transverse direction is zero before and after the proton-proton collision. If the event contains one (or more) weakly interacting particles carrying a large transverse momentum,

⁶ The (*) indicates that the Z can be either in or off shell.

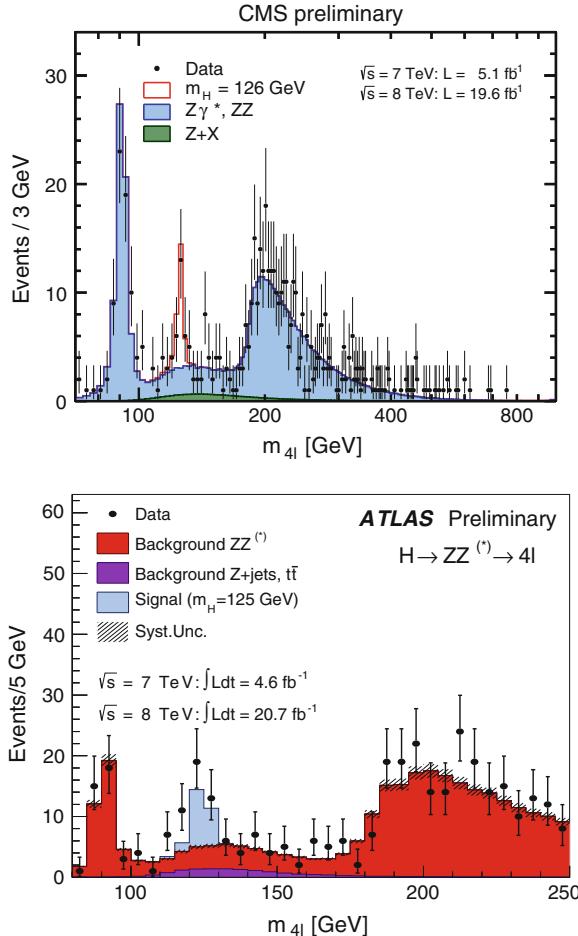


Fig. 7 Four lepton invariant mass distribution showing the Higgs-like signal above the models of expected background for the CMS (*top*) and ATLAS (*bottom*) experiments

this will be observed as an unbalance in the overall transverse momentum in the event. Here event might refer either to a single proton-proton collision or to the piled up collisions in a bunch crossing. In principle the trigger and analysis selects events with objects carrying very large transverse momentum (up to hundreds of GeVs). The piled-up collisions would most likely not contain such objects. Therefore, pile-up only adds up symmetric low transverse energy/momenta activity, also known as noise, in the calorimeter.

The missing transverse energy is calculated as follows. After reconstruction of all the objects in an event: mainly isolated leptons (either electrons or muons),

hadrons, photons and jets, their transverse momentum components p_x^{object} and p_y^{object} are summed up⁷

$$E_x = \sum_{objects} p_x^{object}, \quad E_y = \sum_{objects} p_y^{object}, \quad (13)$$

The transverse missing energy is thus given by

$$\cancel{E}_T = -\sqrt{E_x^2 + E_y^2}. \quad (14)$$

An alternative calculation is obtained replacing the jets by all the calorimeter cells in the object list.

Lately, and particularly in the LHC experiments, the significance of the transverse energy determination, instead of the actual value of this magnitude, is used as a discriminating variable. In the simplified case of gaussian behaviour this significance is given by:

$$\mathcal{S} = \frac{\cancel{E}_T^2}{\sigma_{\cancel{E}_T}^2 (1 - \rho^2)} \quad (15)$$

where $\sigma_{\cancel{E}_T}$ is the variance of the magnitude of \cancel{E}_T and ρ is the correlation coefficient between the variances parallel to and perpendicular to the measured \cancel{E}_T .⁸ An illustration of the discriminating power of this variable can be seen in Fig. 8.

The resolution in jet reconstruction is a key point to achieve a good \cancel{E}_T determination, therefore a good hadronic calorimeter is needed. The physical principle behind hadron calorimeters is the development of hadronic showers. This is determined by the mean free path between inelastic collisions of hadrons in the calorimeter material. The nuclear interaction length, in g cm^{-2} , is given by $\lambda = A/(N_A \sigma_I)$, where A is the atomic mass, N_A is Avogadro's number and σ_I the nuclear interaction cross section.

For CMS and the ATLAS barrel the same technology is used: metal energy-absorbing material interleaved with scintillating tiles that sample energy deposit. The ATLAS Tile calorimeter is 12 m long and 1.6 m thick (corresponding to 7.4 nuclear interaction lengths) the scintillating tiles signal is collected with wavelength shifting fibers and grouped in “channels” read by photomultipliers. They amount to 9,856 channels and 5,184 cells, that provide transverse and longitudinal information (since they are segmented in 3 radial layers) [43]. For the end-cap ATLAS has a copper-liquid argon sampling calorimeter structured in two superimposed wheels. Each wheel has an outer diameter of about 4 m. Their lengths are 0.82 and 0.96 m respectively. Each wheel is composed of 32 identical modules segmented with fine granularity and read out by 3,072 channels [44]. The CMS hadron barrel end-cap calorimeters are sampling calorimeters with 50 mm thick copper absorber plates

⁷ The x and y axis are perpendicular to the proton beams and define an orthogonal system with it.

⁸ This is only an approximation, for a full discussion see [42].

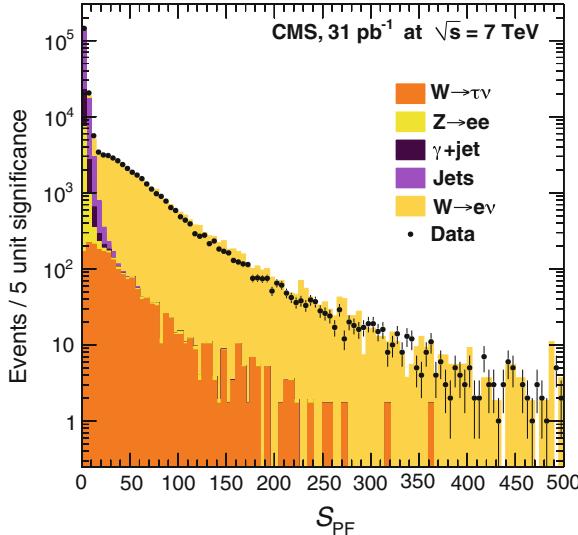


Fig. 8 Distributions for the \cancel{E}_T significance in candidate $W \rightarrow e\nu$ events from CMS data (points) and simulation (histograms). Taken from [42]

interleaved with 4 mm thick scintillator sheets. The barrel calorimeter consists of two half barrels of 4.3 m length each contained within the CMS solenoid. Additional scintillation layers are placed behind the solenoid to contain the energy of high energy showers. The full depth of the barrel system is approximately 11 absorption lengths. The end-cap calorimeter consists of two large structures, situated at each end of the barrel detector with technical characteristics very similar to the end-cap (Fig. 9).

5.1 Results

The latest results of the ATLAS collaboration [45] in the $H \rightarrow WW^*$ decay mode, with $W^{(*)} \rightarrow \mu\nu$ or $W^{(*)} \rightarrow e\nu$, show an excess of events over the expected background for $m_H \lesssim 150 \text{ GeV}/c^2$, with a maximum local significance of 2.8 standard deviations at $m_H = 111 \text{ GeV}/c^2$. At the mass value of $125 \text{ GeV}/c^2$, in the region of the discovered resonance in the $\gamma\gamma$ and four lepton spectra, the significance is 2.6 standard deviations. The best fit signal strength at $m_H = 125 \text{ GeV}/c^2$ is 1.5 ± 0.6 . The results are consistent with the observed resonance being a Higgs boson.

CMS obtains, with slightly more data analyzed in this channel, for a mass of 125 GeV , a significance with respect to the background only hypothesis of 4.0 standard deviations (expected 5.1 assuming the SM prediction). No other significant deviations are observed and the SM Higgs boson is excluded in the mass range

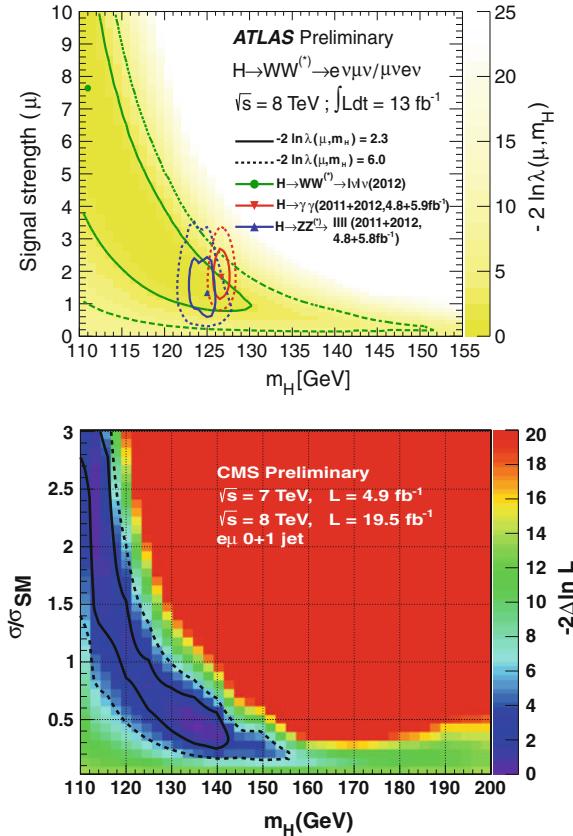


Fig. 9 Higgs mass versus signal strengths likelihood contours in ATLAS (top) and CMS (bottom) for the analysis of the $H \rightarrow WW^*$ decay. Taken from [47] and [46] respectively

128–600 GeV at 95 % confidence level. The angular analysis of the data favors the spin parity of 0^+ over 2^+ [46].

Near and above the threshold given by twice the mass of the W the branching fraction of a SM Higgs decaying through this mode dramatically increases and dominates. This is the reason why this mode is particularly sensitive for high mass Higgs searches. Both ATLAS and CMS have performed such studies excluding, at 95 % confidence level, mass ranges $260 < m_H < 642$ GeV/c 2 [47] and $145 < m_H < 710$ GeV/c 2 [48] respectively. After completing these first studies, the collaborations are tackling now more ambitious goals like searching this decay mode for a Higgs produced through Higgs-strahlung [49].

6 The $H \rightarrow b\bar{b}$ Decay

Back in 1994 this was a channel considered almost impossible to be studied at the LHC. The ATLAS Technical proposal expected signal significances of ~ 0.8 standard deviations for a SM Higgs boson of $120 \text{ GeV}/c^2$ [50] for 10 fb^{-1} integrated luminosity. With this mode Higgs boson searches rely on the Higgs-strahlung production mechanism where the weak boson is required to decay leptonically and a high p_T lepton used to trigger and analyze the event. This is so because at the LHC huge amounts of dijet events are produced and a signal containing only jets in the final state would be overwhelmed by the huge background. In particular, the $b\bar{b}$ pairs cross section is of the order of $\sim 100 \mu\text{b}$, although decreases fast with the invariant mass of the pair. Additionally $t\bar{t}$, single top and resonating background such as $Z \rightarrow b\bar{b}$ pair make this decay mode really challenging. Two ingredients are required to be able to obtain significant results for this decay. One is an excellent jet reconstruction, with a good performance of the jet energy scale. The other is a high efficiency to tag a b-jet while keeping a good rejection factor of light parton and c quark jets.

It is out of the scope of this document to review the jet reconstruction at the ATLAS and CMS experiments. It is, however, worth to mention that the use of the anti- k_T jet clustering algorithm [51] has greatly improved the performance because of the infrared (insensitive to soft radiation) and collinear safeness (insensitive to collinear radiation) of the algorithm. A similar comment applies to the Jet Energy Scale. The reconstructed energy of a jet is just a fraction of the initial parton energy that is aimed at being reconstructed. The scale factor depends on the nature of the parton. Heavy quark jets (b, c) are tagged according to the presence of a displaced vertex in the event but additionally there are tagging algorithms to distinguish light quark jets from gluon jets using information referred to the track multiplicity, the track-jet width and other topological characteristics. Other than that, implementing particle flow techniques, such as substituting calorimeter clusters of low transverse momentum tracks by a matching track⁹ are also employed. Finally, using the maximum of the calorimeter segmentations, thus calibrating every part of the detector independently, also improves the performance. These methods can be checked with data driven dedicated methods such us analyzing Z+jets events.

Despite of its difficulties this channel is very relevant for the understanding of the new resonance nature. This is the largest coupling for the mass where it was discovered and, in principle, the only available decay mode into a quark antiquark pair at the LHC and together with the $H \rightarrow \tau\tau$ and $H \rightarrow \mu\mu$ one of the three decays into fermions that have been tackled. CMS has analyzed data samples [52] corresponding to integrated luminosity of 5.0 fb^{-1} at $\sqrt{s} = 7 \text{ TeV}$ and 19.0 fb^{-1} at $\sqrt{s} = 8 \text{ TeV}$, establishing upper limits that vary from 1.1 to 3.1 times the standard model prediction, at the 95 % confidence level, on the VH production cross section times the $H \rightarrow b\bar{b}$ branching ratio, for a standard model Higgs boson in the mass range $110\text{--}135 \text{ GeV}/c^2$. Corresponding expected limits vary from 0.7 to 1.5. At a

⁹ The tracking and calorimeter resolution curves cross each other. Below a certain momentum the tracking detectors are more accurate at determining the particle energy than the calorimeter.

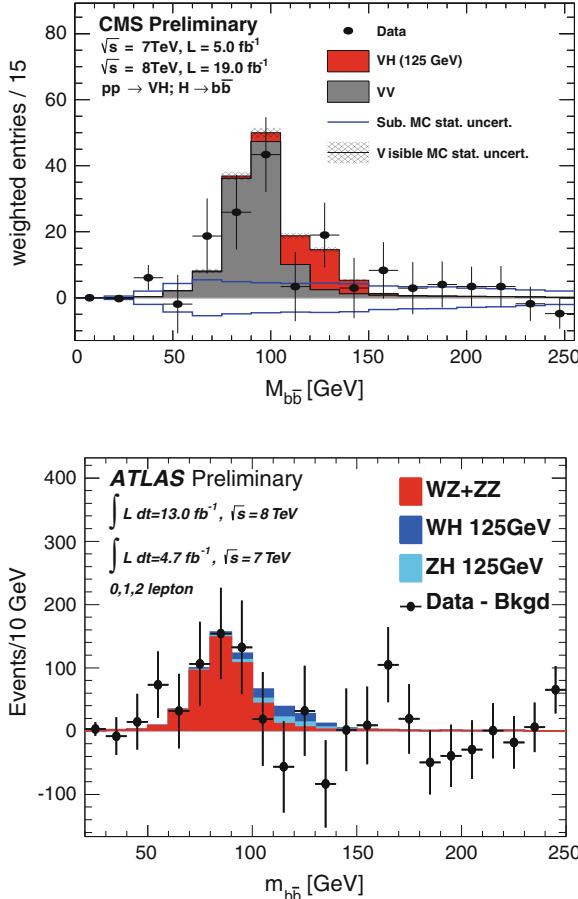


Fig. 10 Background subtracted spectrum $b\bar{b}$ spectra obtained by the CMS (top) and ATLAS (bottom) collaborations. Taken from [52] and [53] respectively

Higgs boson mass of 125 GeV an excess of events is observed above the expected background with a local significance of 2.1 standard deviations, consistent with the expectation for the standard model Higgs boson. The signal strength corresponding relative to that of the standard model Higgs boson, is found to be $\mu = 1.0 \pm 0.5$. With a dataset of 4.7 fb^{-1} at $\sqrt{s} = 7 \text{ TeV}$ and 13.0 fb^{-1} at $\sqrt{s} = 8 \text{ TeV}$ ATLAS [53] establishes similar limits without observing any significant excess. For $m_H = 125 \text{ GeV}/c^2$ the observed (expected) upper limit on the production cross section times branching ratio is 1.8 (1.9) times the SM prediction.

Recently, results on searches of $H \rightarrow b\bar{b}$ produced in association with $t\bar{t}$ or in VBF modes have been published by CMS with little significance [54]. No deviation from the SM is observed (Fig. 10).

7 The $H \rightarrow \tau\tau$ Mode

The $H \rightarrow \tau\tau$ decay should be very sensitive at Higgs masses around $125.5 \text{ GeV}/c^2$. This is because its high branching ratio of 0.062. The ATLAS CSC book [20] established in 2008 that this was a discovery channel in the VBF production mode with 30 fb^{-1} . The analysis of the complete data set of the years 2011 and 2012 took significant longer than other channels with potentially the same discovery power. The reason is that the reconstruction of τ relies on missing transverse energy and, as commented before, this is quite challenging. Indeed the channel suffers from large background sources such as Drell-Yan $\tau\tau$ production, $W + jets$ and QCD dijet production.

The τ decays into hadrons and a neutrino 65 % of the times. The remaining 45 % leptonic $\tau \rightarrow l\nu$ decays occur. The identification of a $H \rightarrow \tau\tau$ signal event often requires at least one of the τ s decaying leptonically. The selection relies on identifying an isolated lepton and significant missing transverse energy. If the second τ also decays leptonically isolation criteria and larger \cancel{E}_T selections are applicable. If the second τ decays hadronically a τ jet is required. The topology of a jet produced after the decay of a τ lepton is very particular. They contain very few charged particles: 1 in 49.5 % of the cases, 3 in 15.2 % and seldom 5 (0.1 %). The lifetime of the τ ($c\tau = 87 \mu\text{m}$) and small mass ($m_\tau = 1.78 \text{ GeV}/c^2$) produce a sizeable decay length that allow for the reconstruction of the decay vertex. Also, the invariant mass of the particles composing the jet can be used for identification purposes. All these variables are used (either separately or combined in a single multi-variate discriminator) to tag a τ jet. If that was not enough, the channel suffers from almost irreducible background from $Z \rightarrow \tau\tau$ decays.

The latest results of the ATLAS and CMS collaborations, updating preliminary works of 2012, finally gave the promising fruits and significances of 4.1σ and 2.93σ respectively. The CMS experiment has analyzed 24.3 fb^{-1} of 7 and 8 TeV center-of-mass LHC data [55]. Five final states were considered: $\mu h_\tau + X$, $e h_\tau + X$, $e\mu + X$, $h_\tau h_\tau + X$, and $\mu\mu + X$. The modes are also combined with the possibility of finding an accompanying W or Z or two well separated jets and, therefore to detect Higgs-strahlung and VBF as the production mode. The detected excess of events is compatible with the SM Higgs of $m_H = 125 \text{ GeV}/c^2$, for which the local significance is 2.85σ . A result that gives a strength coupling compatible to the SM prediction.

On its side ATLAS obtains an spectacular result after exploring all decay modes [56] with 20.3 fb^{-1} integrated luminosity. The analysis categorized the events in two sets: one containing additional jets, targeting VBF produced events, and a second one where boosted Higgs with $p_T > 100 \text{ GeV}$ are required together with a well-measured mass. The result gives a significance of 4.1 standard deviations above the background only hypothesis (3.2 expected). This result significantly updates a preliminary analysis [57] showing the improvement achieved using multivariate methods and a Missing Mass Calculator that achieves a mass resolution of 13–20 % [58].

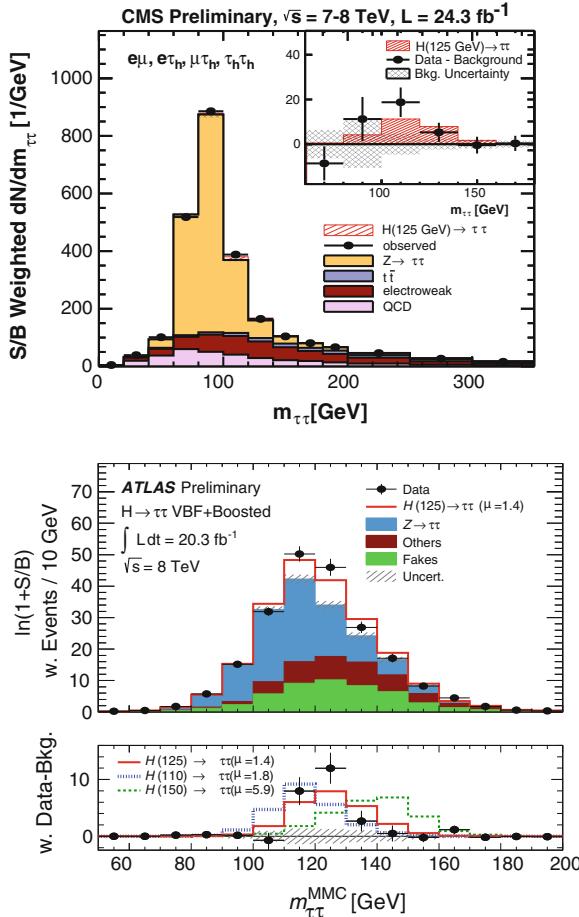


Fig. 11 Mass spectra with background and background subtracted for a SM Higgs in the $H \rightarrow \tau\tau$ search by the CMS (*top*) and ATLAS (*bottom*) collaborations. Taken from [55] and [56] respectively

The results are fully compatible with the SM. This is the first evidence¹⁰ of a decay of the Higgs into fermions that is therefore a key mode to check its properties.

The mass spectra obtained by the ATLAS and CMS collaborations for the $H \rightarrow \tau\tau$ can be seen in Fig. 11.

¹⁰ A sample is considered to provide evidence of a signal if the signal significance is more than three standard deviations.

8 Combined Results

The results combination obtained in the different modes provides the most complete picture of the present status of the analyses. Of course, the combination mainly relies in the channels with more sensitivity.

In particular, the best mass determination by ATLAS uses the data from the $H \rightarrow \gamma\gamma$ and the $H \rightarrow ZZ^*$ decays to produce a combined result of $m_H = 125.5 \pm 0.2(stat)_{-0.6}^{+0.5}(sys)$ MeV/c². It is quite interesting that this average is produced with results from the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^*$ that present some tension since they are apart 2.4 standard deviations.

ATLAS also finds that measurements of the signal strengths for $H \rightarrow \gamma\gamma$, $H \rightarrow ZZ^* \rightarrow 4l$, $H \rightarrow WW^* \rightarrow l\bar{lv}\bar{l}$, $H \rightarrow \tau\tau$ and $H \rightarrow b\bar{b}$ give a combined average value of $1.30 \pm 0.13(stat) \pm 0.14(sys)$ at the mass of 125.5 GeV/c². The more conservative compatibility test gives 11 % probability for this hypothesis. With respect to the signal strengths of the production modes ATLAS reports that their ratio $\mu_{VBF+VH}/\mu_{ggF+nH} = 1.2_{-0.5}^{+0.7}$ is in agreement with the SM expectation. ATLAS states that these and other tests made in the signal strengths and related magnitudes show compatibility with the SM Higgs expectations.

CMS summarizes the combination of results in [59]. As for ATLAS, they find a combined mass from $H \rightarrow \gamma\gamma$ and the $H \rightarrow ZZ^*$ decays to be $125.7 \pm 0.3(stat) \pm 0.3(syst)$ GeV/c². They also find that *the event yields obtained by the different analyses targeting specific decay modes and production mechanisms are consistent with those expected for the SM Higgs boson*, with a combined signal strength ratio with the SM of 0.80 ± 0.14 at the measured mass. The spin of the resonance is also found to favor the SM prediction of 0⁺.

In summary, the state of the art of the data analyses, that will only be resumed after the long break in the LHC, shows the discovery of a 0⁺ spin parity resonance of mass around 125 GeV/c². The couplings of this new particle are, within uncertainties, compatible with an Standard Model Higgs boson in its simplest formulation of a single Higgs doublet.

References

1. S. Glashow, Nucl. Phys. **22**, 579 (1961). doi:[10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2)
2. F. Englert, R. Brout, Phys. Rev. Lett. **13**, 321 (1964). doi:[10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321)
3. P.W. Higgs, Phys. Rev. Lett. **13**, 508 (1964). doi:[10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508)
4. S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967). doi:[10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264)
5. A. Salam, Conf. Proc. **C680519**, 367 (1968)
6. G. Arnison et al., Phys. Lett. **B122**, 103 (1983). doi:[10.1016/0370-2693\(83\)91177-2](https://doi.org/10.1016/0370-2693(83)91177-2)
7. M. Banner et al., Phys. Lett. **B122**, 476 (1983). doi:[10.1016/0370-2693\(83\)91605-2](https://doi.org/10.1016/0370-2693(83)91605-2)
8. G. Arnison et al., Phys. Lett. **B126**, 398 (1983). doi:[10.1016/0370-2693\(83\)90188-0](https://doi.org/10.1016/0370-2693(83)90188-0)
9. P. Bagnaia et al., Phys. Lett. **B129**, 130 (1983). doi:[10.1016/0370-2693\(83\)90744-X](https://doi.org/10.1016/0370-2693(83)90744-X)
10. G. 't Hooft, M. Veltman, Nucl. Phys. **B44**, 189 (1972). doi:[10.1016/0550-3213\(72\)90279-9](https://doi.org/10.1016/0550-3213(72)90279-9)
11. D. Decamp et al., Phys. Lett. B **236(2)**, 233 (1990). [http://dx.doi.org/10.1016/0370-2693\(90\)90834-S](http://dx.doi.org/10.1016/0370-2693(90)90834-S)

12. A. Sopczak, (1995). [arXiv.hep-ph/9504300](https://arxiv.org/abs/hep-ph/9504300)
13. R. Barate et al., Phys. Lett. **B565**, 61 (2003). doi:[10.1016/S0370-2693\(03\)00614-2](https://doi.org/10.1016/S0370-2693(03)00614-2)
14. T. Altonen et al., Phys. Rev. **D88**, 052014 (2013). doi:[10.1103/PhysRevD.88.052014](https://doi.org/10.1103/PhysRevD.88.052014)
15. Tevatron New Phenomena and Higgs Working Group, CDF Collaboration, D0 Collaboration, (FERMILAB-CONF-11-354-E, 2011). [arXiv.1107.5518](https://arxiv.org/abs/1107.5518)
16. LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka (eds.), CERN-2011-002 (CERN, Geneva, 2011)
17. LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka (eds.), CERN-2012-002 (CERN, Geneva, 2012)
18. LHC Higgs Cross Section Working Group, S. Heinemeyer, C. Mariotti, G. Passarino, R. Tanaka (eds.), CERN-2013-004 (CERN, Geneva, 2013)
19. S. Dawson, (1998). [arXiv.hep-ph/9901280](https://arxiv.org/abs/hep-ph/9901280)
20. G. Aad et al., (2009). [arXiv.0901.0512](https://arxiv.org/abs/0901.0512)
21. G. Aad et al., Eur. Phys. J. **C71**, 1630 (2011). doi:[10.1140/epjc/s10052-011-1630-5](https://doi.org/10.1140/epjc/s10052-011-1630-5)
22. K. Stenson, A. Rao, (2012). <http://cms.web.cern.ch/news/reconstructing-magnitude-particle-tracks-within-cms>
23. Updated measurements of the Higgs boson at 125 gev in the two photon decay channel. Technical Report CMS-PAS-HIG-13-001, (CERN, Geneva, 2013)
24. Measurements of the properties of the Higgs-like boson in the two photon decay channel with the atlas detector using 25 fb^{-1} of proton-proton collision data. Technical Report ATLAS-CONF-2013-012 (CERN, Geneva, 2013)
25. L. Lyons, (2008). [ArXiv:0811.1663](https://arxiv.org/abs/0811.1663)
26. E. Gross, O. Vitells, Eur. Phys. J. C **70**, 525 (2010). doi:[10.1140/epjc/s10052-010-1470-8](https://doi.org/10.1140/epjc/s10052-010-1470-8)
27. S. Chatrchyan et al., JHEP **1306**, 081 (2013). doi:[10.1007/JHEP06\(2013\)081](https://doi.org/10.1007/JHEP06(2013)081)
28. J.C. Collins, D.E. Soper, Phys. Rev. **D16**, 2219 (1977). doi:[10.1103/PhysRevD.16.2219](https://doi.org/10.1103/PhysRevD.16.2219)
29. G. Aad et al., (2013). [arXiv:1307.1432](https://arxiv.org/abs/1307.1432)
30. Differential cross sections of the Higgs boson measured in the diphoton decay channel using 8 tev pp collisions. Technical Report ATLAS-CONF-2013-072 (CERN, Geneva, 2013)
31. G. Aad et al., JINST **3**, S08003 (2008). doi:[10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003)
32. S. Chatrchyan et al., JINST **3**, S08004 (2008). doi:[10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004)
33. D. Froidevaux, P. Sphicas, Annu. Rev. Nucl. Part. Sci. (2006). [10.1146/annurev.nucl.54.070103.181209](https://doi.org/10.1146/annurev.nucl.54.070103.181209)
34. H. Bethe, W. Heitler, Proc. Roy. Soc. Lond. **A146**, 83 (1934)
35. Properties of the Higgs-like boson in the decay $h \rightarrow zz$ to $4l$ in pp collisions at $\sqrt{s} = 7$ and 8 tev. Technical Report CMS-PAS-HIG-13-002 (CERN, Geneva, 2013)
36. Measurements of the properties of the Higgs-like boson in the four lepton decay channel with the atlas detector using 25 fb^{-1} of proton-proton collision data. Technical Report ATLAS-CONF-2013-013 (CERN, Geneva, 2013)
37. Search for a standard model like Higgs boson in the decay channel $h \rightarrow zz$ to $l+l^-q\bar{q}$ at cms. Technical Report CMS-PAS-HIG-12-024 (CERN, Geneva, 2013)
38. S. Chatrchyan et al., JHEP **1203**, 040 (2012). doi:[10.1007/JHEP03\(2012\)040](https://doi.org/10.1007/JHEP03(2012)040)
39. Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 gev. Technical Report CMS-PAS-HIG-12-045 (CERN, Geneva, 2012)
40. Combined coupling measurements of the Higgs-like boson with the atlas detector using up to 25 fb^{-1} of proton-proton collision data. Technical Report ATLAS-CONF-2013-034 (CERN, Geneva, 2013)
41. Search for an invisible Higgs boson. Technical Report CMS-PAS-HIG-13-013 (CERN, Geneva, 2013)
42. S. Chatrchyan et al., JINST **6**, P09001 (2011). doi:[10.1088/1748-0221/6/09/P09001](https://doi.org/10.1088/1748-0221/6/09/P09001)
43. G. Aad et al., Eur. Phys. J. **C70**, 1193 (2010). doi:[10.1140/epjc/s10052-010-1508-y](https://doi.org/10.1140/epjc/s10052-010-1508-y)
44. T. Barillari, Nuc. Phys. B Proc. Suppl. **150**, 102 (2006). doi:[10.1016/j.nuclphysbps.2004.10.087](https://doi.org/10.1016/j.nuclphysbps.2004.10.087)
45. Update of the $H \rightarrow WW^{(*)} \rightarrow e\nu\mu\nu$ analysis with 13 fb^{-1} of $\sqrt{s} = 8$ tev data collected with the atlas detector. Technical Report ATLAS-CONF-2012-158 (CERN, Geneva, 2012)

46. Evidence for a particle decaying to $w+w-$ in the fully leptonic final state in a standard model higgs boson search in pp collisions at the lhc. Technical Report CMS-PAS-HIG-13-003 (CERN, Geneva, 2013)
47. Search for a high-mass Higgs boson in the $h \rightarrow ww \rightarrow llvv$ decay channel with the atlas detector using 21 fb^{-1} of proton-proton collision data. Technical Report ATLAS-CONF-2013-067 (CERN, Geneva, 2013)
48. S. Chatrchyan et al., Eur. Phys. J. **C73**, 2469 (2013). doi:[10.1140/epjc/s10052-013-2469-8](https://doi.org/10.1140/epjc/s10052-013-2469-8)
49. Search for associated production of the higgs boson in the $wh \rightarrow www^{(*)} \rightarrow llvv$ and $zh \rightarrow zww^{(*)} \rightarrow llvv$ channels with the atlas detector at the lhc. Technical Report ATLAS-CONF-2013-075 (CERN, Geneva, 2013)
50. W. Armstrong et al., *ATLAS: Technical proposal for a general-purpose p p experiment at the Large Hadron Collider at CERN*. CERN-LHCC-94-43 (CERN, Geneva, 1994)
51. M. Cacciari, G.P. Salam, G. Soyez, JHEP **0804**, 063 (2008). doi:[10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063)
52. Search for the standard model Higgs boson produced in association with w or z bosons, and decaying to bottom quarks for lhcp 2013. Technical Report CMS-PAS-HIG-13-012 (CERN, Geneva, 2013)
53. Search for the standard model Higgs boson in produced in association with a vector boson and decaying to bottom quarks with the atlas detector. Technical Report ATLAS-CONF-2012-161 (CERN, Geneva, 2012)
54. Search for Higgs boson production in association with a top-quark pair and decaying to bottom quarks or tau leptons. Technical Report CMS-PAS-HIG-13-019 (CERN, Geneva, 2013)
55. Search for the Standard-Model Higgs boson decaying to tau pairs in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV. Technical Report CMS-PAS-HIG-13-004 (CERN, Geneva, 2013)
56. Evidence for Higgs Boson Decays to the $\tau^+\tau^-$ Final State with the ATLAS Detector. Technical Report ATLAS-CONF-2013-108 (CERN, Geneva, 2013)
57. Search for the standard model Higgs boson in $h \rightarrow \tau\tau$ decays in proton-proton collisions with the atlas detector. Technical Report ATLAS-CONF-2012-160 (CERN, Geneva, 2012)
58. A. Elagin, P. Murat, A. Pranko, A. Safonov, Nucl. Instrum. Meth. **A654**, 481 (2011). doi:[10.1016/j.nima.2011.07.009](https://doi.org/10.1016/j.nima.2011.07.009)
59. Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV. Technical Report CMS-PAS-HIG-13-005 (CERN, Geneva, 2013)

Messengers of the High Energy Universe

Alessandro De Angelis, Mário Pimenta and Ruben Conceição

Abstract The scales of distances and energies studied in astroparticle physics range from the realm of elementary particles to the realm of cosmology. The importance of this field is increasing steadily, since it allows to perform measurements at energies far higher than those reachable by present and future human-made accelerators, and it gives a handle for understanding the characteristics of emission of the most energetic astrophysical sources.

1 Introduction

For thousands of years Humankind observe the Universe through a narrow optical light window. Nevertheless that was enough to build a solid scientific domain, the Astronomy, which was able to understand our basic situation of a small planet in the neighbourhood of one of the countless galaxies of an expanding Universe. In the last century this window was dramatically enlarged basically to all the electromagnetic spectrum from the infra-red to the very highest energies, the gamma rays. Sources emitting at the highest energies were then found, one by one, and nowadays a rich sky map of gamma ray sources is already available allowing new visions of our Universe.

On the other hand charged cosmic rays were definitely discovered 100 years ago and, since then, the quest of their origin and nature was always present. Their energy spectrum spans through many orders of magnitude from lower than the MeV

A. De Angelis (✉)
Università di Udine, Udine, Italy
e-mail: alessandro.deangelis@uniud.it

A. De Angelis
INFN, Pisa, Italy

M. Pimenta · A. De Angelis
IST, Lisbon, Portugal
e-mail: pimenta@lip.pt

R. Conceição · A. De Angelis · M. Pimenta
LIP, Lisbon, Portugal
e-mail: ruben@lip.pt

to higher than the EeV. However the flux at the highest energies is really scarce (at the EeV, around 1 cosmic ray per km^2 per century) which implies the construction of detectors able to cover several thousand of km^2 .

Finally the search for high energetic astrophysical neutrinos has been, in the last 20 years, in the centre of many efforts. Their extremely weak interaction cross-sections make them the ideal messengers of processes occurring in the core of the astrophysical sources but, at the same time, make their detection extremely difficult. Very recently a strong evidence of non-atmospheric PeV neutrinos was reported opening this new observation channel.

In this review a short summary of the production, propagation and detection of these unique high energy messengers of our Universe is given.

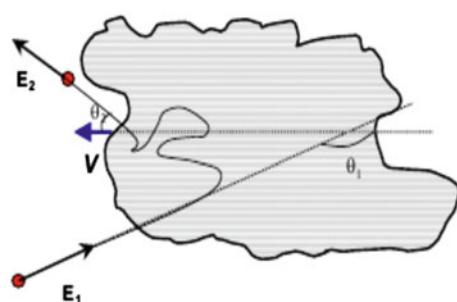
2 Charged Cosmic Rays

2.1 Cosmic Rays Production, Acceleration and Propagation

The production of cosmic rays occurs in several different astrophysical sources. Usually after being ejected the cosmic ray will traverse many regions with strong variable magnetic fields, that induce variable electric fields, contributing in this way for the progressive acceleration of these particles. Fermi proposed, in 1949, an acceleration mechanism [1] in which the cosmic ray would gain energy by colliding stochastically with the molecules of a massive interstellar cloud (see Fig. 1). The cloud system would be basically unaffected by the collision. However, for each cloud that the particle would encounter would provide an average energy gain¹ of

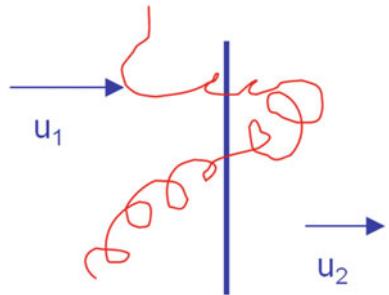
$$\left\langle \frac{\Delta E}{E} \right\rangle \approx \frac{4}{3} \beta^2. \quad (1)$$

Fig. 1 Scattering of a cosmic ray in a interstellar cloud



¹ see Exercise 5.2-2 for the derivation of such result.

Fig. 2 Cosmic ray acceleration in the shock wave rest frame



This mechanism is known as second order Fermi acceleration. Although particle energy increases quadratically with the velocity of the cloud this is not sufficient to explain the cosmic ray energy spectrum as these clouds have usually a small velocity, $\beta \sim 10^{-4}$.

The first order Fermi acceleration is mechanism that occurs when the particle traverses a shock wave (diffuse shock acceleration, DSA, Fig. 2), created for instance by a supernova. In this case the gain is linear,

$$\left\langle \frac{\Delta E}{E} \right\rangle \approx \frac{4}{3}\beta \quad (2)$$

These shock waves can have velocities of $\beta \sim 10^{-2}$ being in this way a natural candidate for the acceleration of high energy cosmic rays (see Exercise 5.2-4).

The Fermi mechanism leads to a power law energy spectrum,

$$\frac{dN}{dE} \approx \left(\frac{E}{E_0} \right)^{-\gamma} \quad (3)$$

with an index $\gamma = 2$. However, when measuring the cosmic ray energy spectrum at Earth one has to take into account that the cosmic rays escape probability from the galaxy is proportional to its energy. This leads to a measurement of a steeper energy spectrum since,

$$\left. \frac{dN}{dE} \right|_{Earth} \propto \left. \frac{dN}{dE} \right|_{sources} \times E^{-\delta} \propto \left(\frac{E}{E_0} \right)^{-\gamma-\delta}. \quad (4)$$

Experimentally the spectral index $\gamma \in [-2.7; -3]$, as it will be shown in Sect. 2.6. Currently, supernova remnants are considered responsible for the acceleration of high energy cosmic ray up to the 10^{17} eV (the *knee* region).

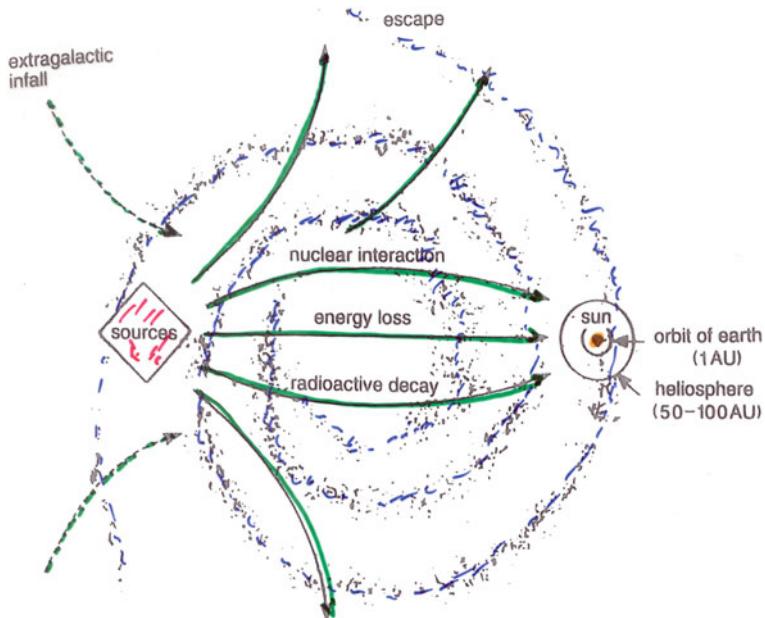


Fig. 3 Galactic cosmic ray propagation

2.2 Galactic Cosmic Rays

The cosmic rays that are produced in the galaxy may have their trajectories changed either by diffusion in random magnetic fields or by convection due to Galactic winds. Their composition can also change due to spallation on the interstellar medium or radioactive decay. They can even gain or lose energy during their propagation. At some point they can either reach the Earth or accumulate enough energy to escape the galactic magnetic field (see Fig. 3). Their lifetime in the galaxy is usually long (10^7 years).

The description of the galactic cosmic ray propagation is done by coupled transport equations that includes all the phenomena described above. These equations allow to determine the abundances of each cosmic ray species. The solutions of these equations are usually achieved using semi-analytical or numerical calculations or even using Monte Carlo propagation codes (where the 3D distribution of the sources can be added). However, in order to provide accurate solutions all physical processes and spatial/energy dependences must be included. This leads to a high number of parameters that cannot be fully constrained by the available data.

Nonetheless, there are simpler models, for instance the *Leaky-Box* models, that can be used to understand the main features of data. This model, in its simplest version, consists of cosmic ray sources uniformly distributed in a volume, the so called *box*. When produced the cosmic rays travel freely inside the box with a given escape probability (see Fig. 4).

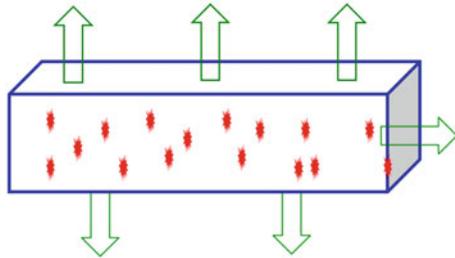


Fig. 4 Sketch of a *Leaky-Box*. The red stars represent the cosmic ray sources while the green arrows show the regions where the particle can escape

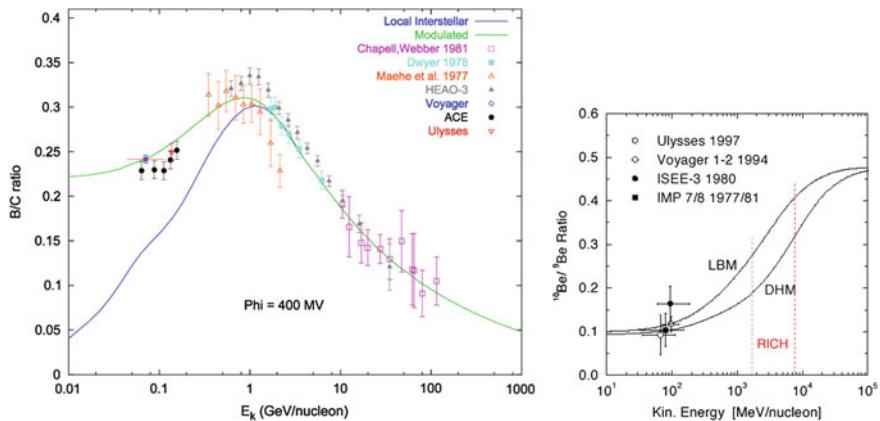


Fig. 5 On the *left* boron over carbon (B/C) (secondary over primary cosmic rays) as a function of the energy and on the *right* (Be^{10}/Be^9) (unstable over stable elements). The curves correspond to different diffusion models: Diffusive Halo Model (DHM), Leaky Box Model (LBM)

These simple models have parameters that can be adjusted using experimental cosmic ray data. One of the most useful measurements is the ratio of secondary elements over the primary elements as a function of the cosmic ray energy, Fig. 5. The former is produced by spallation of heavier elements during the propagation while the latter attains direct information about the production at the sources. Another interesting quantity to constrain these kind of models is the ratios between stable and unstable elements.

It is interesting to note that ratios between primary elements have nearly no dependence on the energy. However, the same is not true for the secondary over primary ratios as shown in Fig. 6. In particular, as energy increases this ratio decreases as a consequence of the increase of the escape probability for higher energy particles.

The *Leaky-Box* models cannot be applied for electrons and positrons as both energy losses and escape probability are significantly higher than for nuclei cosmic rays. The energy losses for electrons and positrons are mainly due to synchrotron

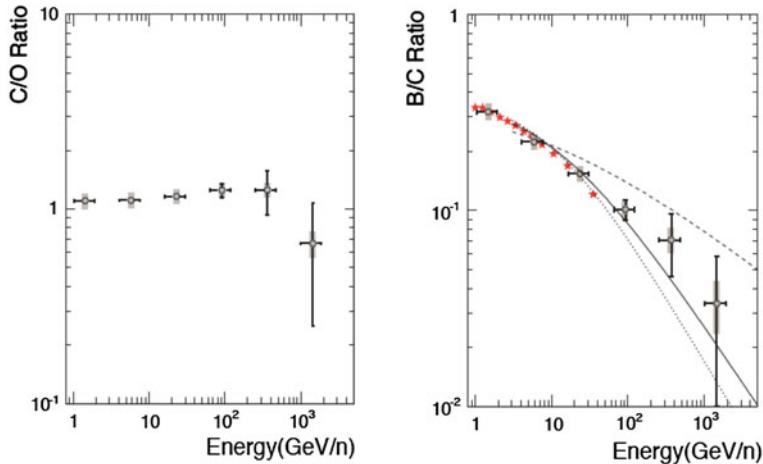


Fig. 6 On the *left* carbon over oxygen (C/O) ratio (primary/primary) as a function of the energy and on the *right* boron over carbon (B/C) (secondary/primary)

radiation and inverse Compton scattering. Their fluxes at Earth have been a subject of intense debate and experimental measurements in the last years [2].

2.3 Extra-Galactic Cosmic Rays

Extra-Galactic cosmic rays travel large distances (tens or hundreds of Mpc) before they can reach us. In fact, it is nowadays believed that the highest energy cosmic rays have an extra-galactic source. However, the Universe is not empty but full of Cosmic Microwave Background (CMB) radiation [3]. These photons, a reminiscence from the Big Bang, have a density of $n_\gamma \sim 410 \gamma \text{ cm}^{-3}$ and an average temperature of $T \sim 2.73 \text{ K}$. Short after its discovery, Greisen, Zatsepin and Kuz'min realised, in 1966, that high energetic protons could interact with the CMB photons through the following inelastic channels (photo-pion production):

$$p \gamma_{CMB} \rightarrow (\Delta^+) \rightarrow p \pi^0 \quad (5)$$

$$p \gamma_{CMB} \rightarrow (\Delta^+) \rightarrow n \pi^+ \quad (6)$$

This would lead to a strong decrease of the proton interaction length, the so called GZK effect [4, 5]. The threshold energy for this effect can be obtained using relativistic kinematics (see Exercise 5.2-1),

$$E_p = \frac{m_\pi^2 + 2m_p m_\pi}{4E_\gamma} \approx 6 \times 10^{19} \text{ eV} \quad (7)$$

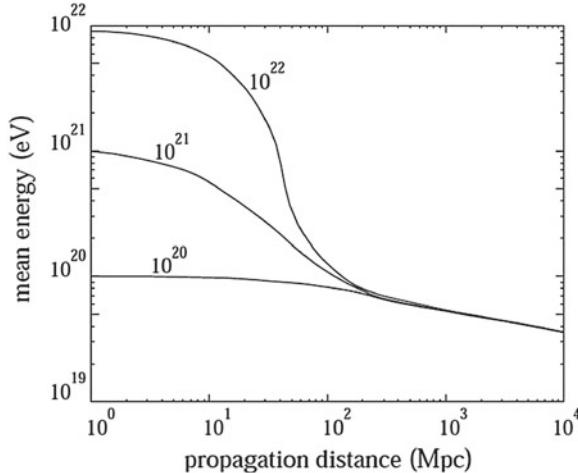


Fig. 7 Simulation results for the proton energy as a function of its propagation distance [6]

and if one takes into account that the photo-pion cross-section, $\sigma_{\gamma p}$, which has values as high as $\sim 500 \mu\text{b}$ just above the threshold, then the mean free path for these high energy protons would be of the order of,

$$\lambda_p \sim \frac{1}{n_\gamma \sigma_{\gamma p}} \sim 10 \text{ Mpc} \quad (8)$$

For each interaction that would occur the proton would lose about 20 % of its initial energy. The accurate assessment of the impact of such effect in the cosmic ray energy spectrum depends not only of the photo-pion cross-section and the CMB energy spectrum but also on the knowledge over the sources: spatial distribution and injection power. The expansion of the Universe has also to be account in these simulation as CMB photons get red-shifted by it. Hence, a proton with a energy above $\sim 10^{20}$ eV would lose most of its initial energy in about 50–100 Mpc until it gets below the GZK energy threshold as shown in Fig. 7.

The decay of the neutrons and pions produced during the photo-pion production should give origin to a flux of high energy photons and neutrinos providing an experimental signature for this process. For the protons there is another possible inelastic channel at around $\sim 2 \times 10^{18}$ eV if the CMB photons convert into $e^- e^+$. Berenzinsky proposed that this could be an explanation for the *ankle* structure observed in the ultra high energy cosmic ray spectrum.

High energy heavier nuclei also can be suppressed by a similar mechanism. In this case they interact with the CMB or Infrared photons leading to their disintegration into lighter nuclei. The photo disintegration process is dominated by the Giant Dipole resonance. The threshold for this process is proportional to the nuclei binding energy per nucleon, and thus naturally stable nuclei are expected to survive longer. The most

stable nuclei is the iron element, *Fe*, and its interaction length is, at the GZK energy, similar to the proton GZK interaction length. The interaction length of lighter nuclei is significantly smaller than for iron. Therefore, during their propagation until Earth they have a high probability of suffering spallation.

2.4 The Detection of Charged Cosmic Ray

The detection of cosmic ray can be done in a direct or indirect way depending if the flux is high enough to perform a direct measurement (typically it is possible to detect it directly up to some tens or hundreds of GeV). For the particles above this energy one must rely on the measurement of the particle showers induced by the interaction of the cosmic rays with the atmosphere, usually designated as Extensive Air Showers (EAS).

The direct detection of cosmic rays is done using balloon or satellite based experiments. These experiments are usually composed by several particle detectors that combined can be used to learn about the nature and energy of the particles that cross it. Among all the experiments that were launch in the last 30 years we choose to enumerate some as examples:

- ACE: Advanced Composition Explorer [7] was launched by NASA in 1997 and is still in operation. It has been producing a large set of measurements on the composition (from H to Ni) of solar and Galactic Cosmic rays covering energies from the 1 KeV/nucleon to 500 MeV/nucleon.
- BESS: Balloon-borne Experiment with Superconducting Spectrometer [8] is a balloon experiment which performed several flights starting in 1993 with the main aim to search for anti-matter, namely anti-helium, and to measure the low energy anti-proton spectrum. The last two flights were performed over Antarctica (BESS-Polar) and had a reasonably long duration (8.5 days in 2004 and 29.5 days in 2007/2008).
- PAMELA: Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics [9] is a satellite launched in June 2006 and has measured during 6 years charged particle and anti-particles out of the Earth atmosphere.
- AMS-02: Alpha Magnetic Spectrometer 02 [10] is a detector installed, in May 2011, on the International Space Station (ISS). The layout of the experiment is similar to the one from PAMELA but with a larger acceptance and a more complete set of sophisticated and higher performing detectors. The experiment total weight is 8,500kg and its estimated cost was over \$1.5 billion. AMS-02 has a broad scientific research objective, among them: anti-matter searches, look for dark matter signatures and study the cosmic ray abundances (particle ratios) over a wide energy range (Fig. 8).

The properties of cosmic rays with energies above $\sim 10^{14}$ eV are inferred through the analysis of Extensive Air Showers. There are essentially three different techniques used to detect these objects, Fig. 9:

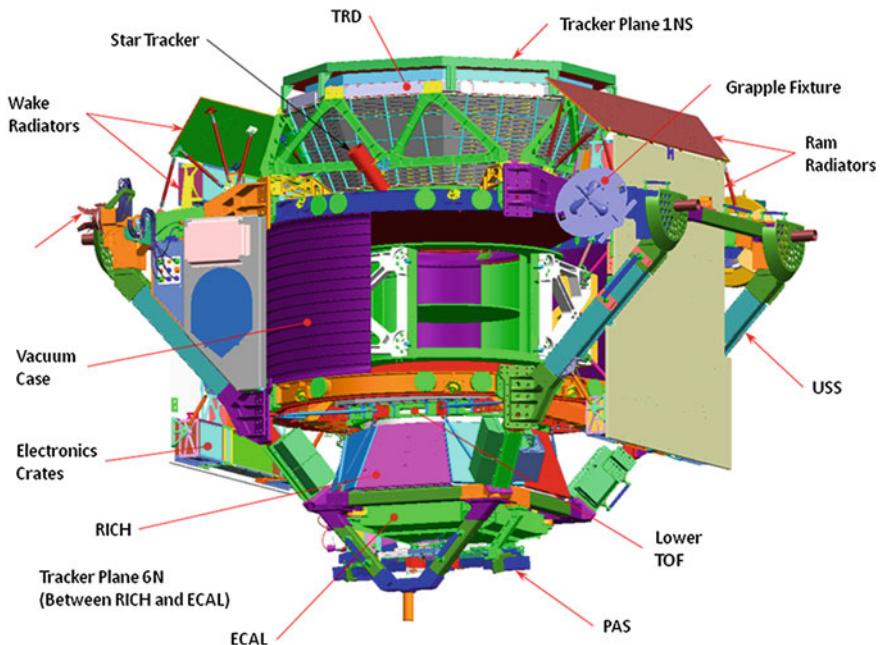


Fig. 8 The AMS-02 detector layout

- The sampling of the charge particles of shower that arrive at ground by an array of surface detectors.
- The measurement of the fluorescence light, in moonless nights, emitted by the excited nitrogen molecules due to passage of the shower low energy electromagnetic component. The emitted photons are isotropic and mainly in the ultra-violet region. This light can be acquired telescopes in the earth or in satellites.
- The shower travels essentially at the velocity of light which means that there is Cherenkov light associated with the development of the shower. This light travels in a narrow cone around the shower axis and can be measured by dedicated telescopes as the Imaging Atmosphere Cherenkov Telescopes (IACTs) (see Exercise 5.3).

Surface detectors measure the time of arrival of individual particles at a given location. There are many types of detectors that can be used to build arrays of surface detectors but the most common are scintillator counters and water Cherenkov tanks. While the core position of the shower can be obtained using the density signal at the stations, the arrival direction is reconstructed using the arrival time information of the shower front at the different surface detectors. The shower front can be, in a first approximation, described by a thin disk propagating at the speed of light (see Fig. 10).

The density of particles at the ground show a fast decrease with the distance to the shower core. The measured lateral distribution function, LDF, is usual fitted

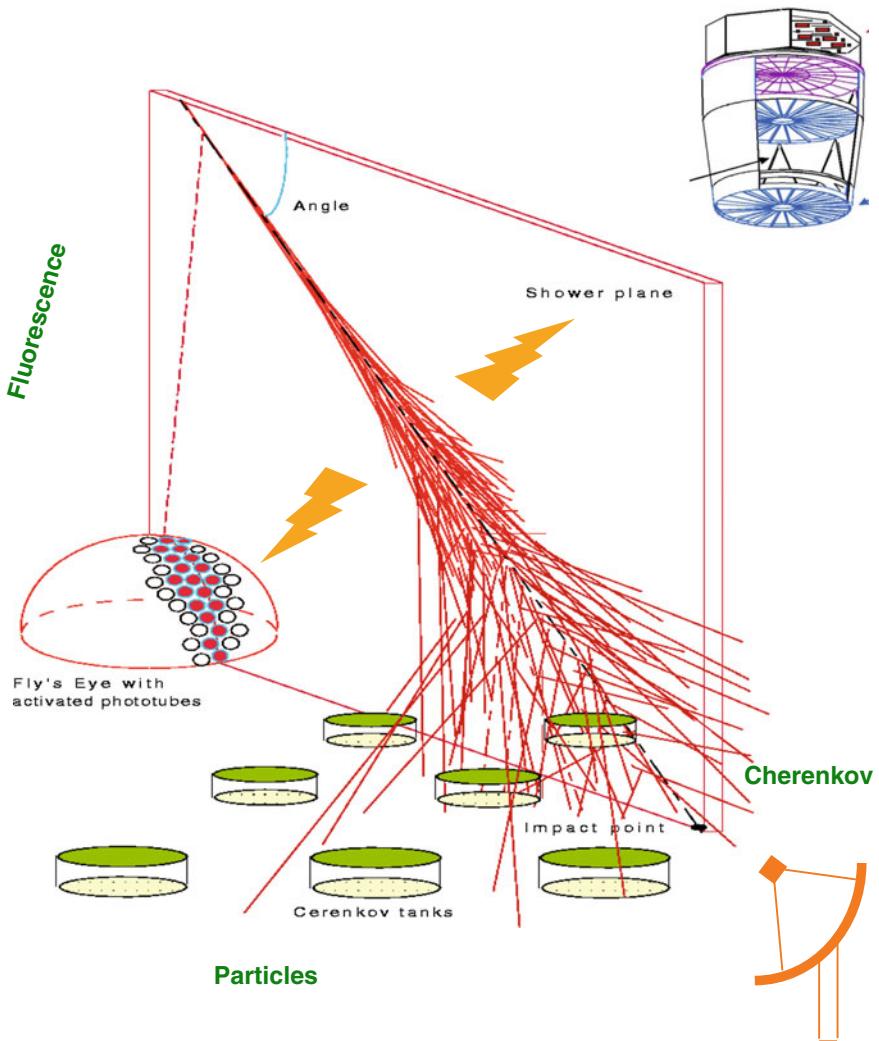


Fig. 9 Scheme for the extended air showers detection technique

with an empirical or phenomenological expression which are generally modifications to the Nishimura-Kamata-Greisen, NKG, formula [11, 12]. The function depends on the radial distance to the shower core and on the shower age, i.e., the shower development level at the time of impact on the soil. These functions are useful to interpolate to regions where there is no detectors, allowing to improve the shower core reconstruction and are commonly used to determine the particle density at a reference distance which is sensitive to the shower energy. The LDF-function as well as the reference distance depends of the array spacing and configuration.

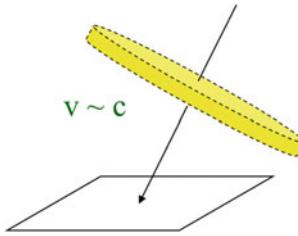


Fig. 10 Scheme of the shower front

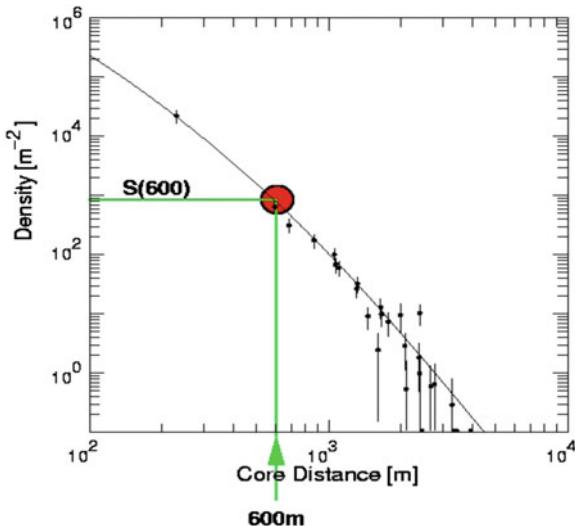


Fig. 11 SD-event measured by the AGASA experiment. The shower energy estimator, $S(600)$, is marker at the plot

For instance, for the AGASA experiment [13] the chosen reference distance was 600 meters from the shower core. This means that the energy estimator for AGASA was $S(600)$, i.e. the signal at 600 m (see Fig. 11).

Fluorescence telescopes can observe the light emitted by the nitrogen molecules due to the development of the shower through the atmosphere. It records the intensity and the arrival time of the light emitted in a specific solid angle region. The shower geometry is reconstructed assuming that the shower is a line in the atmosphere and using the timing information. The longitudinal shower profile is built integrating the signal collected for each given pixel. In Fig. 12 is shown the full image of a single shower event in the focal plane of a FD telescope from the Pierre Auger Observatory [14]. The colour code represents the arrival time with blue being the early part and red latter photons to arrive.

The shower geometry (Fig. 13) reconstruction is performed in two steps: finding the shower detector plane (SDP) and then find the shower inclination within the SDP.

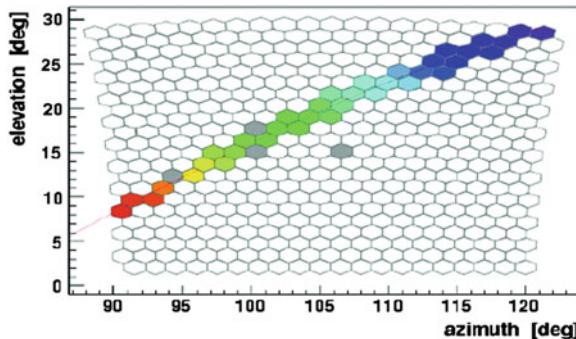


Fig. 12 Display of a shower event in the FD camera of one of the Pierre Auger fluorescence telescopes

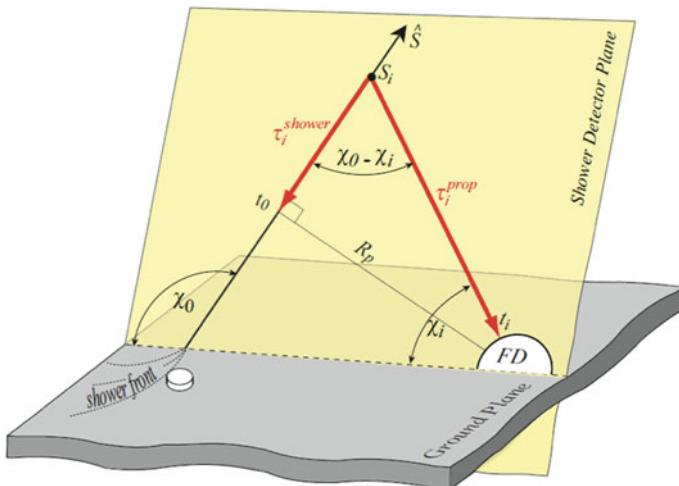


Fig. 13 The shower geometry as seen by a fluorescence telescope

The first step consists in minimising the direction of the SDP normal to the mean directions of the triggered pixels. Afterwards, the shower inclination is obtained using the arrival time of the photons in each pixel while assuming, once again, that the shower develops at the speed of light.

The flux of collected light by the FD telescope has to be corrected along the shower axis for the detector efficiency, the solid angle seen by each detector pixel, the attenuation in the atmosphere and the night sky background. The light contributions from fluorescence light and Cherenkov light (direct or scattered) have also to be assessed. The fluorescence light has an isotropic emission while the Cherenkov light is strongly coupled to the shower geometry. At the end, the longitudinal profile (Fig. 14) can be reconstructed assuming a proportionality between the detected fluorescence light

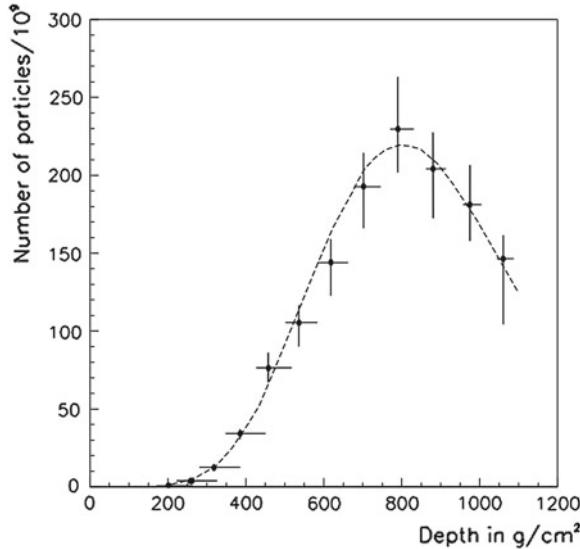


Fig. 14 Shower longitudinal profile of the most energetic event measured by the fly's eye experiment. The shower energy was estimated to be $E = 3 \times 10^{20} \text{ eV}$

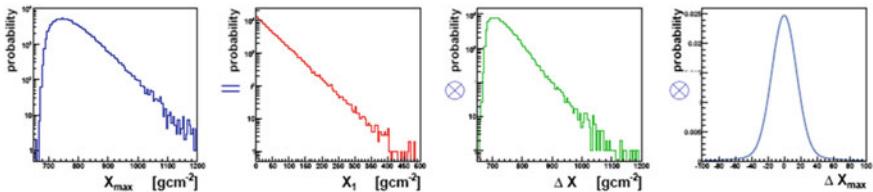


Fig. 15 Construction of the X_{max} distribution

and the number of particles (essentially electromagnetic low energy particles) in the shower. Apart from a small correction related with the energy carried by muons and neutrinos, the *invisible energy*, the integral of the longitudinal shower profile is a good estimator of the shower energy, i.e., the primary particle energy.

As it can be seen in Fig. 14 the longitudinal shower profile as a maximum. The maximum is also a fairly good energy estimator but the depth at which the maximum occurs, X_{max} , has even more interesting properties. This observable, which is usually presented in units of traversed matter (g cm^{-2}), can be expressed as the sum of the first interaction point, X_1 , and the shower development length, ΔX ,

$$X_{max} = X_1 + \Delta X \quad (9)$$

The X_1 distribution is related with the primary cross-section and so is a negative exponential, $\exp(-X_1/\lambda_I)$, where λ_I is the mean interaction length. The X_{max} distribution is simply the convolution of the first interaction point with the ΔX distribution, which has itself a distribution similar to the one from X_{max} (see Fig. 15).

For iron induced showers both X_1 and ΔX have on average smaller values than for proton induced showers. Therefore the X_{max} distribution can be used as an estimator for the cosmic rays mass composition. Moreover, it is easy to see from Fig. 15, that the exponential tail is related essentially with the X_1 exponential. Thus the measurement of the X_{max} tail can be used to infer the primary cross-section, as it will be discussed in Sect. 2.8.

The operation of the Fluorescence Telescope can only be done in low light background conditions and as a consequence it can only operate in moonless nights. A natural consequence of this is a duty cycle of around 15 % as opposed to the surface detectors arrays which operate continuously.

Some experiments, such as the Pierre Auger Observatory or the Telescope Array Experiment [15], have more than one Fluorescence Detectors overviewing a surface detector array. Therefore, occasionally a shower gets seen by two FD telescopes—the *stereo* events. These showers are very important as they can be used to test the knowledge over the detector: energy and geometry resolution. On the other hand, sometimes a shower gets triggered by both the Fluorescence Detector and the Surface Detector array. These events are designated of *hybrid* events and are of the utmost importance for these kind of experiments as they can be used to:

- improve the FD geometry reconstruction;
- calibrate the surface detector energy estimator, which is more model dependent, to the Fluorescence detector energy measurement;

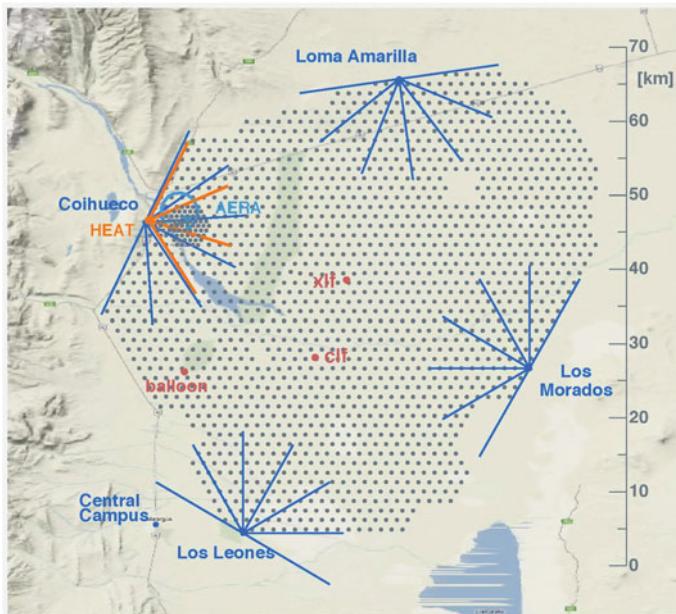


Fig. 16 The Pierre Auger Observatory at Malargüe, Argentina

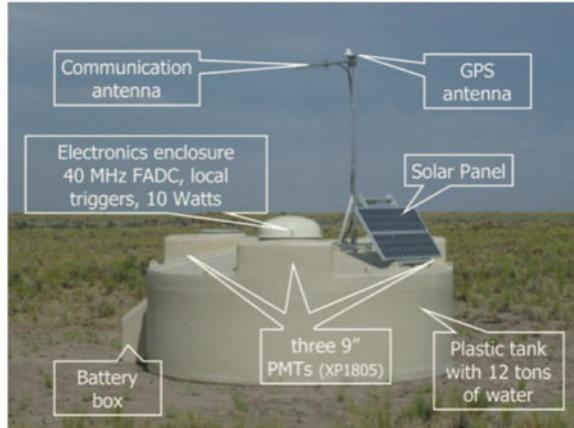


Fig. 17 Layout of a surface detector station of the Pierre Auger Observatory

- create a consistent picture of the shower description, i.e. combine the FD information (sensitive to the longitudinal development of the electromagnetic component) to the SD results (with information about all charged particles at ground including muons).

Nowadays, the biggest cosmic ray detector is the Pierre Auger Observatory [16] located in Malargüe, in the province of Mendoza, Argentina, Fig. 16. It has a surface detector array with more than 1,600 station units covering a total area of around $3,000 \text{ km}^2$. The SD stations are Water Cherenkov Tanks disposed in a triangular grid with 1.5 km of spacing. The array is overviewed by 4 Fluorescence Detectors each with 6 telescopes.

Each water Cherenkov tank covers a circular area of 10 m^2 containing 12 tons of pure water, Fig. 17. The water is involved by a Tyvek material with high reflectivity and there are 3 photomultipliers overviewing the water volume that collect the Cherenkov light emitted by the charge shower particles that cross the tank. Each station is equipped with a GPS unit as well as with a solar panel and a battery making it completely autonomous. The collected data is send to a central data acquisition system through a radio antenna.

The fluorescence detector is a modified Schmidt telescope with a field of view of $30^\circ \times 28.6^\circ$, Fig. 18. The photons enter through a circular diaphragm with a radius of 1.1 m which has a filter to select only ultraviolet photons and a corrector ring to reduce the spherical aberration. The light is collected by a spherical mirror of $3.5 \text{ m} \times 3.5 \text{ m}$ which focus the photons into a camera. The camera is composed by 440 hexagonal photomultipliers.

The JEM-EUSO collaboration [17] has proposed an alternative approach to detect extensive air shower. It consists in installing in the Japanese Experimental Module of the International Space Station a large and sophisticated double sided Fresnel lens telescope. The working principle is similar to the one of the fluorescence detector. The

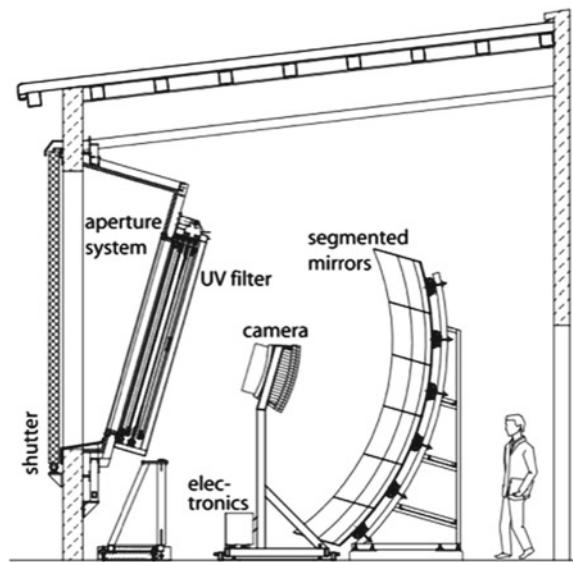


Fig. 18 Layout of a fluorescence detector of the Pierre Auger Observatory

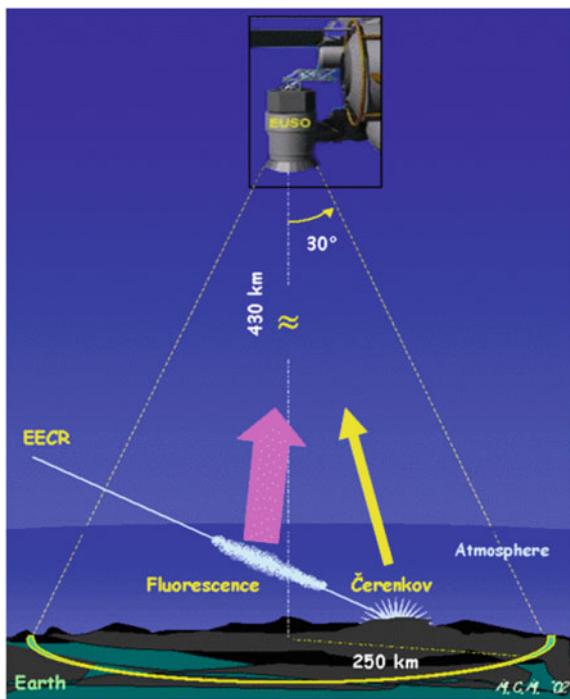


Fig. 19 JEM-EUSO observational principle

telescope would detect the fluorescence light emitted during the shower development as well as the Cherenkov photons created when the shower hits the ground, Fig. 19. The camera is composed by about 5,000 multi-anode photomultipliers (PMTs) with a total number of pixels of 3.2×10^5 . The telescope acceptance would be huge (the Earth surface covered $\sim 1.9 \times 10^5 \text{ km}^2$) due to the large field of view $\pm 30^\circ$ and because the detector is observing the ground at a distance of 400 km. This means that this experiment would be able to collect lots of shower events at the highest energies. However, the price to pay is a high energy threshold ($\sim 3 \times 10^{19} \text{ eV}$) and a poorer resolution on the X_{\max} when compared with the results of the FD at ground.

2.5 Arrival Direction

Due to the galactic and extra galactic magnetic fields, charged cosmic rays have an isotropic arrival direction. However, at the highest energies if the cosmic rays composition is light and if the sources are sufficiently nearby one could expect some degree of anisotropy. In 2007 the Pierre Auger Collaboration reported a positive correlation for the 27 events above 57 EeV and the Véron-Cetty-Véron, VCV, active nuclei galaxies, AGN, catalog. However, the strength of the correlation decreased significantly and is nowadays in a constant plateau of around 3σ above the isotropy hypothesis.

2.6 Energy Spectrum

The cosmic ray energy spectrum spans over nearly 30 orders of magnitude in flux and almost 11 order of magnitude in energy, as one can see in Fig. 20. While at low energies, $\sim 1 \text{ GeV}$ the fluxes are high, i.e. around 1 particle per square meter per second, at the highest energies, $\sim 10^{20} \text{ eV}$, the fluxes are extremely scarce, about 1 particle per square kilometre per century.

The lowest energy region of the cosmic ray energy spectrum, below 1 GeV, is modulated by the solar wind (anti-correlation) and it also depends on the Earth geomagnetic field. Above a few GeV the flux of the cosmic rays is basically ruled by a power law

$$\Phi(E) \propto E^{-\gamma} \quad (10)$$

with a differential spectral index between 2.7 and 3.

Two break points can be clearly observed, the so called *knee* and *ankle* (see Fig. 21). The first one occurs at $\sim 10^{15} \text{ eV}$ and is commonly believed to be associated with the transition from galactic to extra-galactic cosmic rays. The second one can be found at $\sim 5 \times 10^{18} \text{ eV}$ and its nature is still controversial.

At the end of the energy spectrum, $E \sim 5 \times 10^{19} \text{ eV}$, there is a strong suppression of the cosmic ray flux. This suppression, which is nowadays well established by the

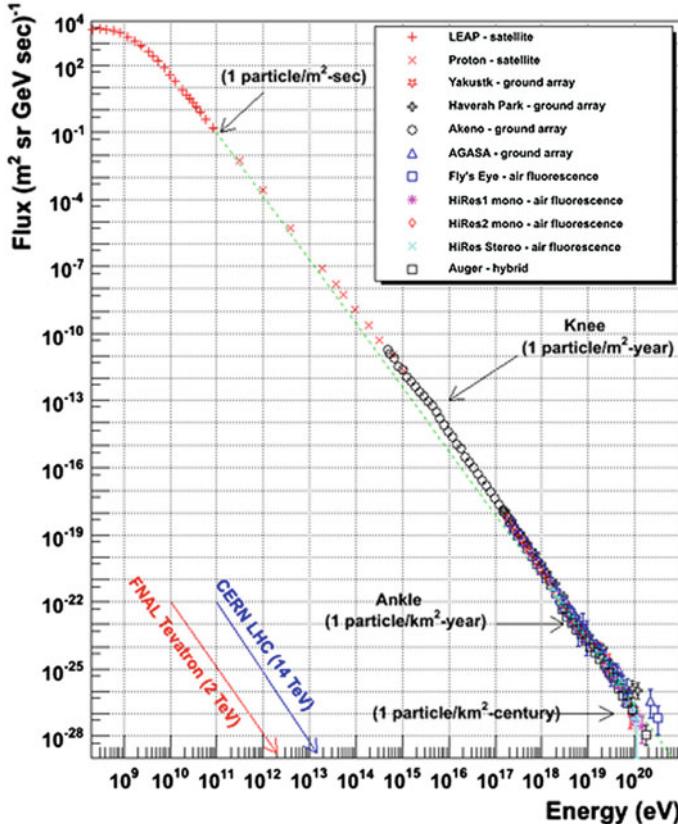


Fig. 20 The cosmic ray energy spectrum

Pierre Auger Observatory [18] and the Telescope Array [15] data, is compatible with the expected GZK suppression, discussed in Sect. 2.3. However, since the sources are still unknown, one cannot discard a scenario of heavy composition associated with the exhaustion of the sources. In other words, without knowing the sources distributions and the injection spectra it is not possible to distinguish between the two scenarios using only the cosmic ray energy spectrum. This ambiguity may be solved by learning about the nature of ultra high energy cosmic rays.

2.7 Mass Composition

If one considers the full cosmic ray energy spectrum then the cosmic ray are mainly protons ($\sim 80\%$) and nuclei.

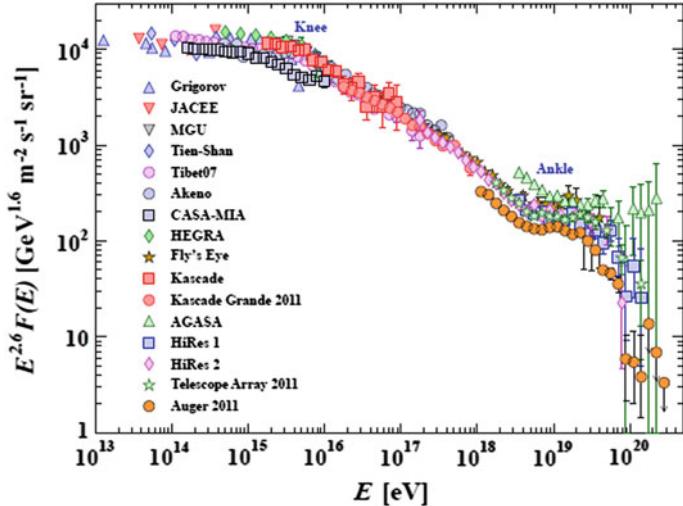


Fig. 21 The cosmic ray energy spectrum multiplied by a power of the energy to enhance the spectrum features

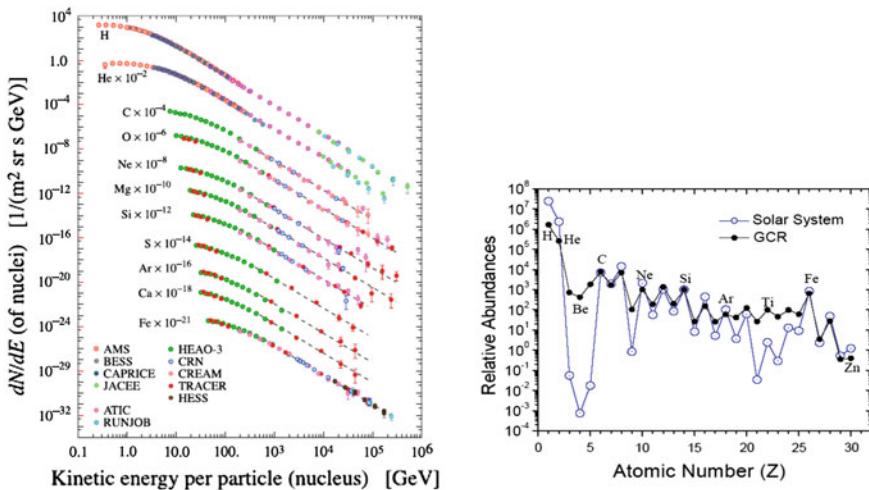


Fig. 22 The cosmic ray energy spectrum and relative abundances of chemical composition

Figure 22 shows the absolute and the relative fluxes of the main hadronic components of cosmic rays arriving at Earth. The relative abundances that arrive at Earth are compared to the relative abundances in the solar system. The features of this plot can be explained within a scenario with primary and secondary production of cosmic rays. The local maximums in the plot are primary cosmic rays, which are produced in stellar like sources, while the local minimums are mainly secondaries

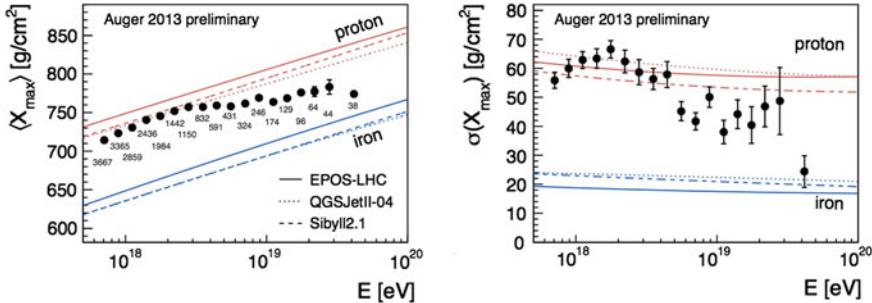


Fig. 23 Evolution of the average X_{\max} (left) and of its fluctuation (right) versus energy. The number of events in each energy bin is given

produced in the interaction of the primary cosmic rays with the interstellar medium, i.e. *spallation*, which justifies the discrepancy found in the *valleys*.

At the highest energy the nature of the cosmic ray has to be inferred through the analysis of EAS observables. One of the most sensitive shower observables to mass composition is the first two momenta of the X_{\max} distribution: the mean, $\langle X_{\max} \rangle$, and its fluctuations, $RMS(X_{\max})$.

The Pierre Auger Observatory has measured these two quantities as a function of the primary energy and the results are displayed in Fig. 23 [19]. The lines represent the predictions of different hadronic interaction models for both proton and iron primaries. A quick inspection of the two plots suggests that the data is dominated by a light mass composition at lower energies becoming increasingly heavier as the energy increases. However, a careful analysis shows that a simple bimodal proton–iron transition scenario is not able to explain simultaneously the two observables. In fact, using the current hadronic interaction models, the most probable mass composition scenario is a complex mixture that at the highest energies is dominated by intermediate state element such as helium or nitrogen. Such scenario is difficult to reconcile with the current cosmic ray acceleration and transport models scenarios. The shower observables are as well sensitive to the hadronic interaction physical description and so one cannot discard the hypothesis that at the highest energies there are surprises in the hadronic sector. The solution for this puzzle might pass through the understanding of the shower development namely its muon content, as it is a direct probe to the hadronic interactions.

2.8 Particle Physics Measurement with EAS

The ultra high energy cosmic ray have energies well above the current human-made accelerators capabilities. Also their interaction with the atmosphere can occur at energies of centre of mass up to $\sqrt{s} \sim 400 \text{ TeV}$, when currently the highest energy

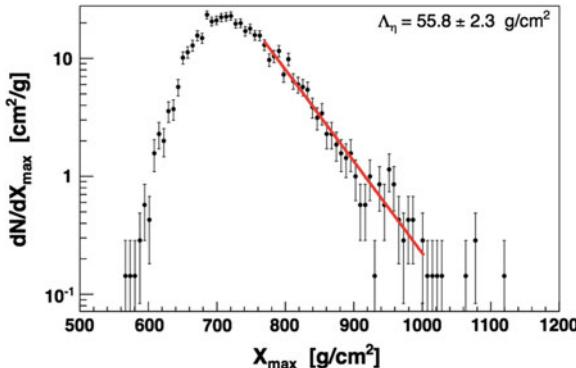


Fig. 24 X_{\max} distribution as measured by the Pierre Auger collaboration in the energy bin $\log(E/\text{eV}) \in [18; 18.5]$

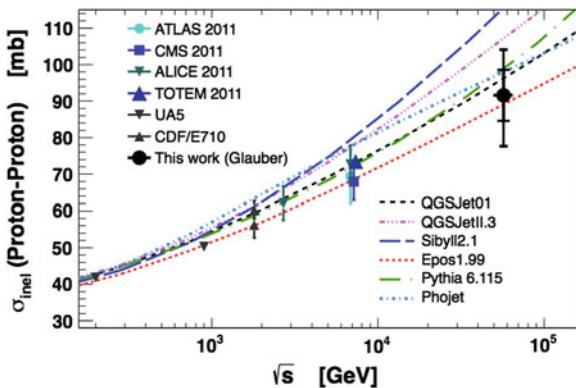


Fig. 25 Proton-proton inelastic cross-section as a function of \sqrt{s}

accelerator, the LHC, collides protons at 8 TeV.² Therefore, although the ultra high energy cosmic ray interaction with the atmosphere does not occur in a controlled laboratory as in the LHC, the resulting shower of such interaction is an object of interest as it might attain information about the hadronic interactions at the highest possible energy.

As seen in the previous sections, the Extensive Air Shower development depends not only on the nature of the cosmic ray that induced it but also on the hadronic interactions. For instance, the maximum of the shower, X_{\max} is quite sensitive to the first interaction point, or in other words to the primary cross-section. Figure 24 shows the X_{\max} distribution measured by the Pierre Auger Collaboration in the energy bin $\log(E/\text{eV}) \in [18; 18.5]$ for the 20 % most deeply penetrating showers. The tail of this distribution has an exponential behaviour as expected. The selection of the

² In the next year it is expected to start colliding protons at 14 TeV.

shower with an higher X_{max} strongly enhances the selection of proton showers as, in average, they penetrate more deeply in the atmosphere than any other nuclei.

The measurement of the slope of the exponential tail can be converted into the proton-air cross-section by means of EAS Monte Carlo simulations. The conversion of proton-air into proton-proton cross-section (both total and inelastic) can be done using a standard Glauber model [20] which combines geometrical arguments with the multiple-scattering probability of partons inside the nuclei.

The Pierre Auger collaboration result in terms of inelastic proton-air cross-section is presented in Fig. 25 as the red point [21]. It is also presented previous accelerator measurements of this quantity including the new LHC data points. Also the predictions of the different phenomenological codes used to describe hadronic interactions at the highest energies are represented in the plot by the lines. The result is in line with the evolution of the previous data points. This supports further that this energy bin should, not only be dominated by a light composition, but also it should have an important fraction of protons.

3 Photons and Neutrinos

As seen previously, the detection of the sources of charged cosmic rays are seriously limited due to the presence of the magnetic fields in the Universe. Studies of the Faraday rotation of the polarisation suggest that the Galactic magnetic fields are of the order of a few μG , while the highly directional structures led to the creation of magnetic fields maps for the Galaxy. However, the extra-galactic magnetic field remains yet unknown. Observations of clusters of galaxies and of the bridges between clusters allowed us to conclude that this space has a non-zero magnetic field. Currently the limits for this field (ExtraGalactic Magnetic Field, EGMF) are estimated to be $10^{15} < B < 10^{-9}$ G. The structure of EGMF is believed to be cellular, i.e. the magnetic field B should have a correlation length $\lambda_B \in [0.1; 1]$ Mpc, which is randomly changing its direction from one region to another while keeping approximately the field strength [22].

The Larmor radius, which gives the deflection of a charged particle of energy E , in a magnetic field of strength B , can be written as,

$$\frac{R_L}{1 \text{ pc}} \approx \frac{1}{Q} \frac{\frac{E}{1 \text{ PeV}}}{\frac{B}{1 \mu\text{G}}} \quad (11)$$

where Q is the charge of the particle.

The galactic centre is at 8 kpc from Earth and, as told previously, the average magnetic field is around $1 \mu\text{G}$. This means that in order to see protons coming from the direction of the galactic centre, its energies should be at minimum, 2×10^{19} eV. However, the flux at these energies is very scarce, and, moreover, the black hole inside

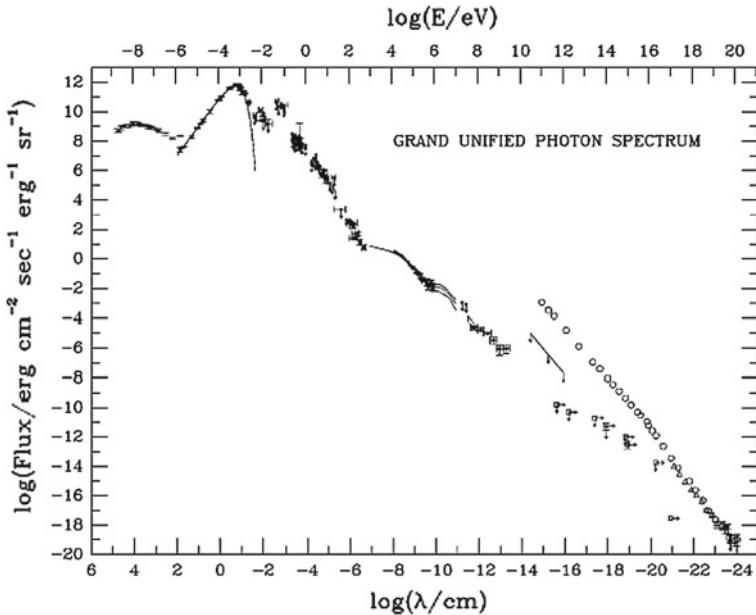


Fig. 26 Energy spectrum of photons experimentally detected

our galaxy, with about 4 million solar masses, is not a good candidate to accelerate particles at these energies.

Hence, it is only natural to use neutral messengers to investigate the emission processes of charged cosmic rays. However, this is not an easy task as the yield of photons at 1 TeV is about one thousand times smaller than the one for protons. Additionally, the yield of neutrinos is expected to be of the same order for these energies.

3.1 Gamma Rays

High energy astrophysical processes have, as a signature, photon radiation over a large range of wavelengths. The available experimental data on the cosmic photon radiation covers about 20 orders of magnitude in energy (see Fig. 26). From the lowest energies to the highest the first visible structure is a bump corresponding to the CMB radiation. As the energy increases, the leading behaviour of the gamma rays can be approximated by a power law with a spectral index of -2 . Although the so called ultra and extremely high energy photons (EeV and PeV, respectively) are expected to exist, cosmic gamma rays have been only detected up to now in the low (MeV), the high (GeV), and the very high (TeV) energy regions. The region above ~ 50 TeV is extrapolated by low energy data with additional experimental upper limits.

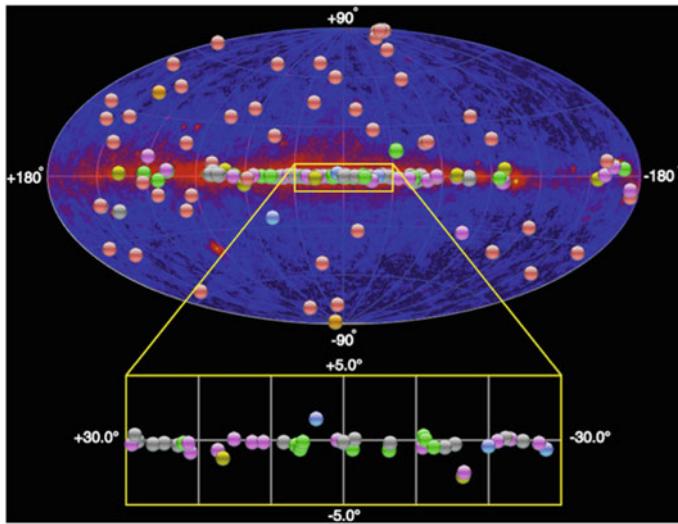


Fig. 27 Sources of VHE emission plotted in galactic coordinates. The *background* represents the HE sources detected by the Fermi satellite. The region near the Galactic centre is blown up

Usually gamma rays are classified according to their energy: high energy (HE) photons are the ones above 30 MeV, i.e., the threshold for the e^+e^- pair production plus some phase space; and very high energy (VHE) photons, above 30 GeV. This classification is somewhat arbitrary, but it corresponds to different detection techniques, as will be seen later.

Gamma rays probe the non-thermal Universe. Up to now, the hottest thermalised processes observed were in the accretion region of a supermassive black hole (SMBH) at a temperature of 10^5 K, corresponding to a few tenth of eV, or in other words, the X-ray region. Apart of the evident astrophysical interest of VHE photons, they can also be used to perform tests of fundamental physics.

Gamma rays in the HE region come mostly from a diffuse background plus a set of localised sources. Nowadays about 2,500 HE emitters have been identified; among these, 200 emit also on the VHE region (Fig. 27). Approximately half of the discovered emitters are in our Galaxy and most of them can be associated to supernova remnants (SNR). The other half has an extra-galactic origin and the present detector resolutions (which is of the order of 0.1°) it is not sufficiently good to draw correlation with a particular object in these galaxies. However, it is believed, that it should be produced in the vicinity of SMBH in the galaxies, as it will be discussed later.

There are several possible production mechanisms of HE photons [23]. They can be produced, for instance, by the interaction of charged particles, such as electrons, protons or ions, accelerated by the shock waves of remnants of gravitational collapses, with nuclear targets, in particular molecular clouds, or with radiation fields. The decay of heavy particles could also produce HE photons.

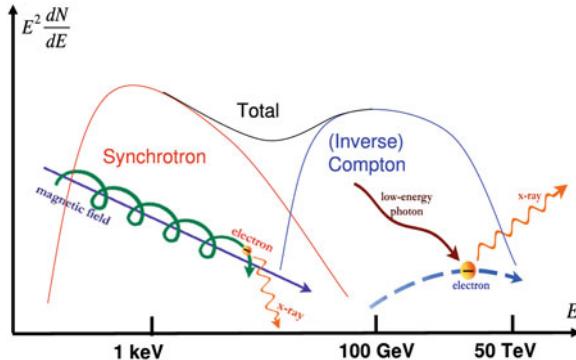


Fig. 28 Differential energy spectrum of photons in the self-synchrotron Compton (SSC) model

The primary production mechanism is the so called self-synchrotron Compton (SSC) mechanism. This is a purely leptonic mechanism in which accelerated relativistic electrons generate synchrotron photons. These electrons can be accelerated in magnetic fields such as the ones present in the accretion disk of an AGN or the surrounding of a supernova remnant. The synchrotron photons have, usually, an energy spectrum that is peaked in the infrared/X-ray region due to the typical strength of these fields. Such photons can, afterwards, interact through Compton scattering with their own parent electron population. As the electrons are ultra relativistic, i.e., with a Lorentz factor of $\gamma \sim 10^{4-5}$, the energy of the upscattered photon can be largely boosted. This phenomenon is known as inverse Compton scattering (IC). The peak of the Compton component can be found at the GeV–TeV energies. The inspection of photon energy spectrum in these regions reveals the two characteristics synchrotron and Compton peaks, which appear on top of the E_γ^{-2} dependency (Fig. 28). A clear signature of the SSC model is a definite correlation between the yields from the synchrotron radiation and from the IC during a flare event.

A complementary model for the VHE gamma ray emissions states that they can be generated in cascades initiated by primary protons/nuclei. The beam of hadrons can be either dumped by a nuclear target (a molecular cloud) or by a sea of photons (hadron-photon production). In this case, the energy of the primary photons should be about 1 or 2 orders of magnitude below the energy of the protons. The main reason resides on the dominant photon production mechanism: π^0 mesons, which are only produced at the end of the hadronic cascade, decays into boosted γ pairs. This means that the study of gamma rays can also be used to give insight on the production of charged cosmic rays. Furthermore, it has been demonstrated, from the morphology of the sources and from the reconstructed of π^0 peak in cosmic photons, that these hadronic components exist. Naturally, the observation of neutrinos in a flare event would be a smoking-gun for the presence of this hadronic component.

Another possibility for the origin of these VHE photons could be the decay of heavier particles—top-down scenarios.

3.2 The Propagation of Cosmic Gamma Rays

After its production, VHE photons propagate through the Universe until they reach the observer. During their propagation, they can interact with extragalactic background photons, producing electron-positron (e^-e^+) pairs. This is the main source of opacity of the Universe to gamma rays and it occurs whenever the corresponding photon mean free path is smaller, or of the order, of the source distance [24].

The air creation $\gamma + \gamma_{background} \rightarrow e^+ + e^-$ is kinematically possible whenever

$$\varepsilon > \varepsilon_{thr}(E, \phi) \equiv \frac{2m_e^2 c^4}{E(1 - \cos \phi)}, \quad (12)$$

where m_e is the electron mass and ϕ the scattering angle. Due to the cosmic expansion, E and ε change in the longitudinal direction with a factor proportional to $(1+z)$.

The corresponding Breit-Wheeler cross section is given by:

$$\sigma_{\gamma\gamma}(E, \varepsilon, \phi) = \frac{2\pi\alpha^2}{3m_e^3} W(\beta) \simeq 1.25 \times 10^{-25} W(\beta) \text{ cm}^2, \quad (13)$$

where

$$W(\beta = v/c) = (1 - \beta^2) \left[2\beta(\beta^2 - 2) + (3 - \beta^4) \ln \left(\frac{1 + \beta}{1 - \beta} \right) \right]. \quad (14)$$

Assuming an isotropic background of photons, the cross-section has its maximum when the energy of the background photons is around

$$\varepsilon(E) \simeq \left(\frac{900 \text{ GeV}}{E} \right) \text{ eV}. \quad (15)$$

Summarising:

- For $10 \text{ GeV} \leq E < 10^5 \text{ GeV}$, the Extragalactic Background Light (*EBL*) represents the main opacity source. Explicitly, when $E \sim 10 \text{ GeV}$, the integral of the cross section $\sigma_{\gamma\gamma}(E, \varepsilon)$, over an isotropic distribution of background sources has its maximum value for far-ultraviolet photons, $\varepsilon \simeq 90 \text{ eV}$. As for $E \sim 10^5 \text{ GeV}$, the cross-section is maximum for soft photons in the far-infrared, $\varepsilon \sim 9 \times 10^{-3} \text{ eV}$;
- For $10^5 \text{ GeV} \leq E < 10^{10} \text{ GeV}$, the interaction with the CMB is the dominant contribution;
- For $E \geq 10^{10} \text{ GeV}$, the radio background plays the leading role.

If one neglects the expansion of the Universe, it is possible to compute the mean free path from (13). It is larger than the Hubble radius for energies smaller than some 10 GeV, but for energies above 100 GeV, it is comparable to the distance of observed sources. The mean free path will be of the order of the Galactic centre when

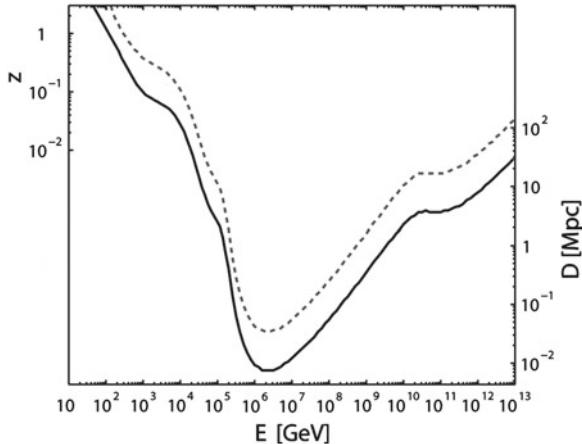


Fig. 29 Curves corresponding to the gamma-ray horizon $\tau(E, z) = 1$ (lower) and to a survival probability of $e^{-\tau(E, z)} = 1\%$ (upper curve)

the Universe becomes mostly opaque, i.e., for photons with an energy around 10^{15} GeV (see Exercise 5.2-1).

At the end, the probability for a photon of energy E to survive when travelling from its source to an observer is mainly ruled by an absorption factor for the radiation flux. This is commonly expressed in the form,

$$e^{-\tau(E, z)} \quad (16)$$

where z is the redshift of the source relative to the observer and $\tau(E, z)$ is called the *optical depth*.

The *horizon* or *attenuation edge* for a photon of energy E is the distance (redshift z) at which $\tau(E, z) = 1$, i.e., the distance where the attenuation reduces the initial population by a factor of $1/e$ (see Fig. 29).

The attenuation of γ -rays might be changed in the presence of other interactions apart from the ones described above. The present data on the absorption of photons, in particular from the observed luminosity of VHE photon sources, are hardly compatible with a pure QED picture as the Universe appears to be more transparent to γ -rays than expected. This has motivated many thorough studies. One of the possible explanations for such result would be that the gamma rays might transform into *sterile* or *quasi-sterile* particles, for instance axions, along their propagation. This speculative explanation would effectively decrease the path length of the photon. Another possible explanation consists in the absorption change through violation of the Lorentz invariance. Such models, currently under strong scrutiny, are very appealing for the scenarios inspired to quantum gravity.

3.3 Detectors for Gamma Rays

When compared to other charged particles of similar energy, the detection of high-energy photons is more difficult due to the faintness of the signal. Furthermore, the absorption rate in the atmosphere, whose thickness is about 28 radiation length at sea level, poses another complication. For these reasons, the detection of primary X/ γ -rays below some 10 GeV is made essentially with satellite-based detectors, although the maximum size of such instruments is constrained by the cost of space technology, resulting into small detectors, with an area around 1 m². However, an indirect detection of these photons is possible when the energy of the primary particle is large enough. In this case, a shower of particles is generated through the interactions with the atmosphere (as in a calorimeter), and the resulting products can be detected by ground-based detectors. To reduce the atmospheric attenuation, it is more convenient to build these detectors at high altitudes. The advantage of this technique lies on the larger area covered by the ground-based detectors when compared to the typical area covered by satellites. For this reason, VHE and UHE photons can only be detected using ground-based detection techniques since the fluxes at these energies are very small and decrease with increasing energy. The separation between HE and VHE gammas can also be understood by its detection techniques: while HE photons are detected using satellites, VHE photons can only be detected from the atmospheric showers that they produce when crossing the atmosphere (Fig. 30).

Therefore, the two detection techniques (satellite and at ground) are complementary: for very low energies (≤ 1 GeV), the atmospheric shower induced by the primary attenuates very rapidly and cannot be detected at ground level. As such, the

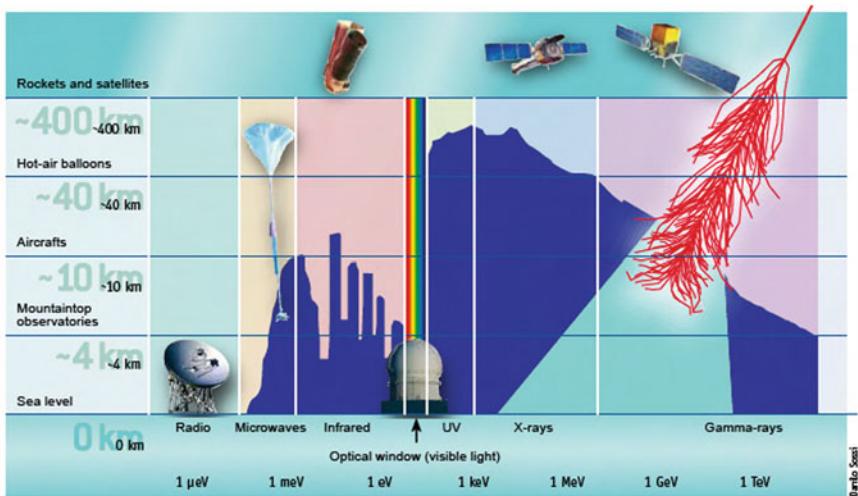


Fig. 30 Transparency of the atmosphere for different photon energies and possible detection techniques

only possible detection method for gamma-rays below this energy concerns the use of satellites. When the energy increases up to the TeV scale, the flux is too small to be detected by the small effective area of a satellite-based telescope (less than one photon/m² every 10 h for the most luminous sources). The solution is to rely on low cost and large sensitivity of ground-based detectors despite the huge amount of background events.

3.3.1 Satellites

The most important characteristics of a satellite, responsible for its capability as a photon detector, are: the energy resolution; the space or angular resolution (also known as point-spread function, or PSF); and the effective area, i.e., the area times the detection efficiency.

The energy of the primary photons detected by satellite HE gamma telescopes is smaller than for ground-based telescopes. They present some limits in the sensitivity due to the small effective area ($\simeq 1\text{m}^2$) and are quite expensive. The cost is essentially dominated by the launch program and by the demanding instrumental capabilities as the equipment must be sent into space with little or no possibility of intervention to fix possible errors. On the other hand, they are less affected by background events, such as charged cosmic rays, since it is possible to introduce anti-coincidence systems to reject this kind of events. Moreover, the duty cycle of these detectors is larger.

Currently there are two gamma-ray telescopes in orbit: AGILE, an Italian satellite, and Fermi (see Fig. 31), an international collaboration lead by NASA [25]. The former was launched in April 2007 and the latter in 2008. They present very similar structures, only differing on the effective area: while Fermi has an effective area around 1m², AGILE is one order of magnitude smaller. The idea was acquired from their less technological precursor, EGRET, that was operational during almost 10 years (1991–2000). The main detector (for FERMI it is called the Large Area Telescope, LAT) consists of foils of heavy materials to convert the incident photons into e^+e^- pairs, that are later reconstructed by planes of silicon detectors; a calorimeter in the bottom reconstructs the energy. From the kinematical configuration of the charged particles, it is possible to determine the direction of the incident photon. The background events that are produced by incoming charged particles are rejected by a system of anticoincidences (see Fig. 31, right).

Fermi was already able to detect many new classes of gamma ray emitters, among them, more than 2,000 sources with energies larger than 100 MeV.

A summary of the main characteristics of Fermi in the energy region of main interest, $10\text{ MeV} \leq E \leq 30\text{ GeV}$, can be found in Table 1. Due to the high cost of space missions, with present technology, it is difficult to surpass the performance of the Fermi LAT in the detection of GeV gammas by exceeding its size. Nonetheless, there are satellites under discussion that will improve other aspects of Fermi, such as calorimetry.

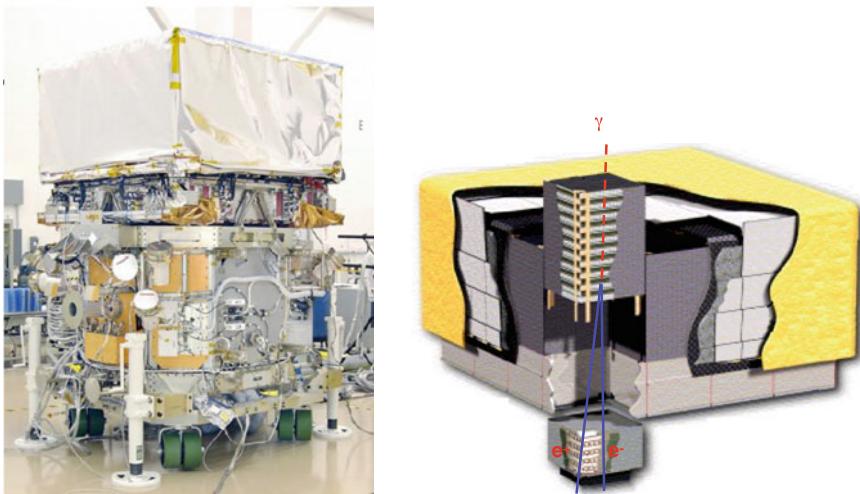


Fig. 31 On the *left*, the Fermi satellite. On the *right*, the layout of the LAT and its principle of operation

3.3.2 Ground-Based Detectors

The indirect detection of VHE gamma-rays can be done through air showers that are originated by the interaction of primary photons and cosmic rays with the atmospheric molecules. From them, secondary particles are originated that can be detected by ground-based telescopes.

Ground-based detectors can be divided into two main categories: the Extensive Air Showers arrays and the Cherenkov telescopes (see Fig. 32).

EAS Detectors

This kind of instruments, among them MILAGRO, Tibet-AS, ARGO and HAWC [26], essentially use the same detection technique as charged cosmic rays detectors: a large array covered with stations that are able to detect the charged secondary particles generated by the photon-induced atmospheric shower.

The field of view (FoV) and the duty cycle are large, but the sensitivity is relatively low. Since the maximum of a shower initiated by a 1 TeV photon occurs, on average, at 8 Km above the sea level, the threshold energy lies, at best, within the 0.5–1 TeV range. The corresponding flux at such energies is quite small and arrays with a total area of the order of 10^4 m^2 are necessary to provide a good statistic. To reduce the background contamination coming from hadron-initiated showers, muon detectors coupled to the EAS detectors can be used, or, alternatively, the shape of the shower can be reconstructed.

Table 1 A comparison of the characteristics of Fermi/LAT, of the IACTs and of the EAS particle detector arrays

Quantity	Fermi	IACTs	EAS
Energy range	20 MeV–200 GeV	100 GeV–50 TeV	400 GeV–100 TeV
Energy res. (%)	5–10	15–20	~50
Duty cycle (%)	80	15	>90
FoV	$4\pi/5$	$5^\circ \times 5^\circ$	$4\pi/6$
PSF (degree)	0.1	0.07	0.5
Sensitivity	1 % Crab (1 GeV)	1 % Crab (0.5 TeV)	0.5 Crab (5 TeV)

Sensitivity computed over 1 year for Fermi and the EAS, and over 50 h for the IACTs

The reconstruction of the primary direction is made by considering the arrival times of the shower particles, as seen previously in Sect. 2.4. The reconstructed primary arrival direction has an angular resolution of about 1 degree, while the resolution in energy is relatively poor. The angular resolution can be estimated studying the deficit of cosmic rays induced by the shadow of the Moon.

Still, current EAS detectors were not able to detect a large number of sources. In Mexico, at 4,100 m of altitude, a new detector with a larger effective area and using water Cherenkov tanks to detect atmospheric air showers is currently under construction: the HAWC detector. This very high-energy gamma-ray observatory consists of 300 steel water Cherenkov units of 7.3 m of diameter and 4.5 m high, making a total area of about 22,000 m² when fully deployed (this phase is scheduled for 2015). Each tank is filled with purified water and equipped with three PMTs with diameter of 20 cm each. By crossing the water, the secondary particles that were originated in the EAS can produce Cherenkov light that will be collected by the PMTs. As opposed to the previous generation of detectors, this method of detection is able to collect not only the charged secondary particles, but also photons that represent a significant fraction of the electromagnetic component of an air shower at ground level (photons will undergo Compton scattering with the water molecules or will convert into e^+e^- pairs, subsequently producing Cherenkov light). Therefore, the HAWC sensitivity to a Crab-like point spectrum is larger by a factor of 15 in comparison to its predecessor MILAGRO. Also, this improvement should enable the detection of Gamma-Ray Burst emissions at high energy.

Cherenkov Telescopes

Most of the ground-based detectors of VHE gamma-rays use Imaging Atmospheric Cherenkov Telescopes (IACTs). This technique has provided most of the VHE photons results and is used by several generations of detectors: WHIPPLE, the first successful cosmic gamma ray detector; HEGRA and CANGAROO, the second generation of such detectors; and currently, H.E.S.S., MAGIC and VERITAS.

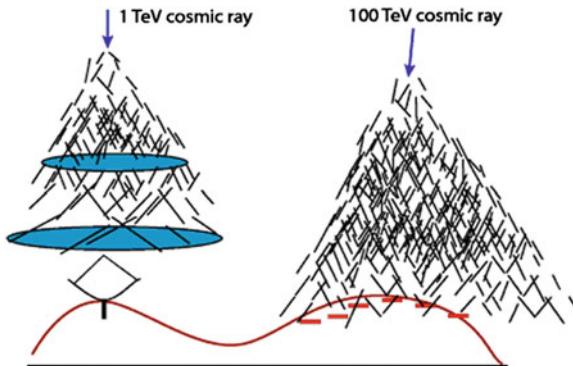


Fig. 32 Sketch of the operation of Cherenkov telescopes (*left*) and of EAS detectors (*right*)

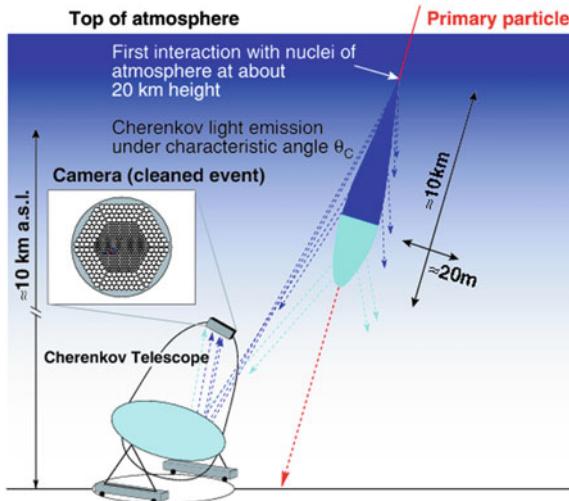


Fig. 33 The observational technique adopted by the imaging atmospheric Cherenkov telescopes (IACTs)

These experiments have small FoV and duty cycle, but a large sensitivity and a small energy threshold. The secondary charged particles originated from the EAS may produce Cherenkov radiation in their propagation through the atmosphere. The resulting light is detected by the telescope that is composed by a large optical surface (see Fig. 33), that reflects the photons onto the focal plane. Photons are collected by a camera of 1 m, corresponding to a FoV of $5^\circ \times 5^\circ$, that consists of several PMTs with a quantum efficiency of about 30 %.

The duration of the shower, that is around 2 up to 3 ns at ground, is maintained by an isochronous (parabolic) reflector.

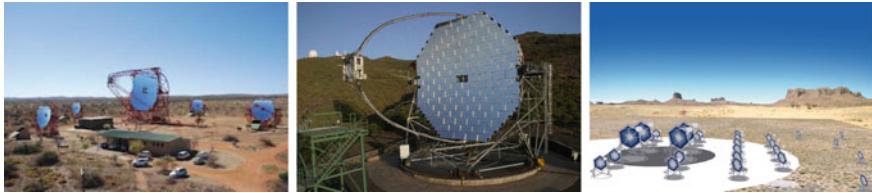


Fig. 34 *Left* The H.E.S.S. telescopes; *middle* one of the MAGIC telescopes; *right* a possible layout for the Cherenkov telescope array (CTA)

In the energy range of interest, GeV–TeV, the background events coming from hadron-initiated showers produce a flux of charged particles that is three orders of magnitude larger than the one originated by the photons. Nonetheless, hadronic showers show larger fluctuations with respect to electromagnetic ones, e.g., the ones initiated by photons. Therefore, it is possible to elaborate a veto based on the shower shape to select gamma-rays induced showers. This background rejection has a larger efficiency for experiments that use more than one Cherenkov telescope. Besides, a larger number of IACTs in the same observatory also result into a better energy and angular resolution.

Three experiments are operational nowadays:

- H.E.S.S. (Fig. 34, left) located in Namibia and working since 2003. This experiment consists of four telescopes with 12 m of diameter, plus a fifth larger telescope with an area of $\simeq 600 \text{ m}^2$, that was inaugurated in 2012 [27].
- MAGIC (Fig. 34, middle) located in Canary Island of La Palma. It consists of two telescopes, that are the largest single-dish Cherenkov telescopes in operation. Each telescope has a diameter of 17 m and 236 m² of reflecting area [28].
- VERITAS in Tucson, Arizona and working since April 2007. It is composed by four telescopes of 12 m of diameter each [29].

In 50 h of observation, these detectors have a sensitivity of about 1 % of the Crab supernova remnant (which is kind of a benchmark for these experiments, since it is the second most visible source of gamma rays presently known, the most luminous among the ones visible both from the northern and the southern hemisphere detectors, and it is rather stable); see Fig. 35.

The next future experiment, the Cherenkov Telescope Array (Fig. 34, right), is expected to increase this sensitivity by one order of magnitude [30]. The CTA collaboration consists of almost all the scientists from current IACTs experiments with 400 newcomers, making a total of 800 people approximately.

The layout to detect low energy photons will involve four telescopes of 23 m class, with a FoV of about 4–5°. For the intermediate energy range, 100 GeV–1 TeV, the array will be covered with 20 telescopes of the 12 m class, with a FoV of 6–8°. The high energy range, above 10 TeV will be surveyed by approximately 40 small telescopes, with a diameter of 4–6 m and a FoV around 10°. The telescopes will be disposed in concentrical circles, where the largest ones will be placed in the center

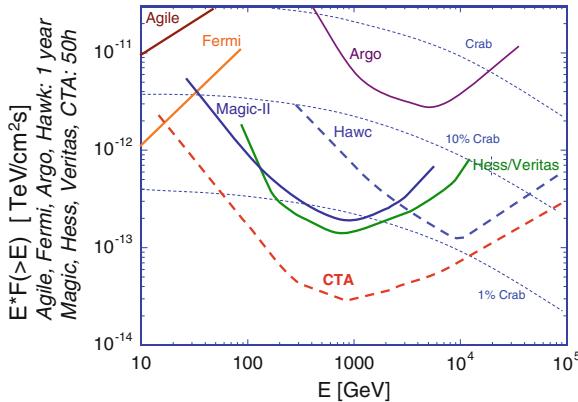


Fig. 35 Sensitivities of some present and future HE gamma detectors, measured as the minimum intensity source detectable at 5σ . The performance for EAS and satellite detector is based on 1 year of data taking; for Cherenkov telescopes it is based on 50 h of data

(Fig. 34, right). In the design scenario, the experiment will be composed by two arrays: a southern hemisphere array, composed by three types of telescopes with different mirror sizes (in this way, it is possible to cover the full range of energies); and a northern hemisphere array that consists of two larger telescopes types.

3.3.3 Summary of the Performance of Gamma Ray Detectors

Table 1 shows a simplified comparison of the characteristics of the Fermi LAT satellite detector, the IACTs, and the EAS detectors (ground-based). In Fig. 35 a comparison of the sensitivity of several gamma-ray experiments as a function of the photon energy is shown.

3.4 Neutrinos

Neutrinos are very important particles to astrophysics [31]. Having a very small cross-section, they can leave their production regions, without interacting. This means that, contrarily to photons, neutrinos carry information about the deep core of their astrophysical sources. Neutrinos are massively produced in nuclear reactions that stars use to generate their energy (the Sun emits about 2×10^{38} neutrinos per second). For instance, each conversion of four protons into a helium produces two neutrinos. Neutrinos can also be produced in the Universe most violent explosions such as in the core-collapse supernovae, the accretion disks encircling supermassive black holes, and have been certainly produced from the Big Bang. Neutrinos are responsible for

most of the cooling of many astrophysical objects including the degenerate helium cores of red giants and young neutron stars.

Furthermore, neutrinos appear also as by-products of cosmic ray collisions in the atmosphere. Hence, these neutrinos can also be used to constrain the properties of the primary cosmic ray spectrum, even more effectively than photons—which can come also from purely electromagnetic reactions.

However, the same property that makes these particles extremely interesting (i.e., their low interaction cross-section) poses significant challenges to their detection. In order to surpass this difficulty, UHE neutrino researchers must build telescopes with the necessary sensitivity to see events. The low detection probability implies the use of large volumes of water or ice, to increase the collected number of events. The detection techniques include optical and coherent radio detection of the secondary particles produced.

Among the experiments in operation, the largest sensitivity detectors are Baikal NT-200 in the Lake Baikal, ANTARES in the Mediterranean sea (in front of Marseille), and the IceCube experiment in the ices of Antarctica.

Being located at the South Pole and with a volume of nearly 1 km^3 , IceCube (Fig. 36) is nowadays the largest neutrino experiment. This telescope surveys the ice using approximately 5,160 optical sensors, deployed on an array of 80 sparse and 6 dense vertical strings, at a depth of 1,450–2,450 m. On top there is a surface air shower array—IceTop—whose data can be combined with IceCube. Most of the background in this experiment comes from muons produced during the Extensive Air Shower development. However, this kind of background can be removed considering events coming from the Earth, i.e., once again take advantage of the neutrino low cross-section. Such cross section grows linearly with energy: hence, for UHE neutrino the Earth is opaque, which means that their detection must be done considering events coming from the horizon or above. Intermediate energy neutrinos, in turn, are more likely to be detected from below.

Up to now, only two astrophysical sources of neutrinos have been identified: the Sun and one collapsing supernova.

The neutrinos from the Sun were firstly observed by Davis and his collaborators in the 1960s. For this they built a 650 ton neutrino detector in the Homestake Gold Mine in South Dakota. Their data allowed to conclude that the flux of neutrinos coming from the Sun was only about one third of what expected from the solar power, thus paving the way to the discovery of neutrino oscillation and consequently of the neutrino nonzero mass.

The neutrinos coming from the collapse of a supernovae can be detected by experiments such as the Super-Kamiokande and the Sudbury Neutrino Observatory (SNO). The first and only observation up to now of a neutrino burst occurred on February 23rd, 1987. The neutrinos, emitted from a supernova in the Large Magellanic Could, some 0.2 Mly from Earth, was observed in the Kamiokande and in the Irvine-Michigan-Brookhaven (IMB) detectors, both primarily intended to study the proton decay. There were approximately 20 events spread over about 10 s. The optical counterpart reached an apparent magnitude of about 3 (about 50 times smaller than Sirius, the brightest star at night), and it could even be observed a few hours after

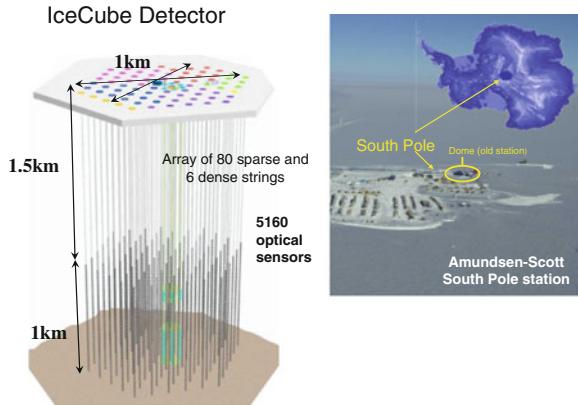


Fig. 36 *Left* IceCube experiment layout. *Right* location of the IceCube experiment

the neutrino burst. Despite the low statistics and resolutions of these first generation detectors, the burst duration and the number of collected events are consistent the standard estimates of the energy release and cooling time of a supernova.

Recently the IceCube Collaboration published an article claiming the observation of 27 high energy neutrinos. The two highest energy events have a reconstructed energy of (1.04 ± 0.16) PeV and (1.14 ± 0.17) PeV. At the PeV energy scale neutrinos are expected to come from astrophysical sources beyond the Solar System. However, all the conducted studies on these events are still inconclusive about the neutrinos production sources. In any case, the detection of these high energy neutrinos opens the door for neutrino astrophysics, suggesting that the 1 km^3 volume of IceCube is close to the right size to explore the high energy neutrinos.

There is another neutrino experiment with a size of the order of the IceCube experiment that is already planned, the Km3NeT. This will be a large underwater neutrino detector with a volume of several cubic kilometres at the bottom of the Mediterranean Sea. Currently the experiment design predicts around 12,000 pressure-resistant glass spheres attached to about 300 detection units. Each unit is composed by a vertical structure with a height of nearly 1 km. Each glass sphere is composed by 31 PMTs and the collected signal will be sent to the shore through a high-bandwidth optical link. This huge experiment has three pioneer smaller experiments that collect experience for the final design. All three pilot project are in the Mediterranean Sea: the ANTARES experiment near Marseille, the NESTOR detector in Greece and the NEMO project situated in the sea of Sicily. ANTARES was completed in 2008 and is currently the largest neutrino telescope operating in the northern hemisphere.

Apart from these astrophysical sources, another important production mechanism of neutrinos is the interaction of cosmic rays with the atmosphere. As seen previously, when a cosmic ray interacts with the atmosphere it produces secondary particles such

as pions, kaons and muons. The atmospheric neutrinos that are produced during the decay of these particles come essentially from the following processes:

- $\pi^\pm \rightarrow \mu^\pm + \nu_\mu (\bar{\nu}_\mu)$
- $K^\pm \rightarrow \mu^\pm + \nu_\mu (\bar{\nu}_\mu)$
- $\mu^\pm \rightarrow e^\pm + \nu_e (\bar{\nu}_e) + \bar{\nu}_\mu (\nu_\mu)$

Hence, the ratio $r \equiv (\nu_e + \bar{\nu}_e) / (\nu_\mu + \bar{\nu}_\mu)$ is expected to be around 1/2 for atmospheric neutrinos.

4 Final Remarks

Astrophysics with high energy cosmic rays has a double purpose: use the energetic cosmic particles reaching Earth as *bullets*,³ or use them to investigate their origin and the characteristics of astrophysical sources. Especially for the second topic the multi messenger approach plays a relevant role.

Related to their use in the field of particle physics, i.e., the use of cosmic rays as ultra-high-energy bullets, charged cosmic rays have no rivals. Their scarce flux led to the construction of huge detectors that allow the recording of particle collisions with atmospheric nuclei at a c.m. energy one order of magnitude above the LHC (~ 100 TeV). The present results in mass composition, fluxes and arrival directions suggest an unexpected astrophysical scenario or possibly dramatic changes in the hadronic interaction multi particle production.

However, the Galactic ($\sim 1\mu\text{G}$) and extragalactic ($\sim 1\text{nG} < B < 10^{-6}\text{ ng}$) magnetic fields, together with the GZK cutoff, do not allow the study of cosmic ray sources with charged cosmic rays, being their gyroradii smaller than the typical distance of possible sources. Recent important results on the origin of cosmic rays come however from high-energy gamma rays.

Among the categories of possible cosmic ray accelerators, several have been studied trying to infer the relation between gammas and charged particles. In the Milky Way in particular, SNRs are since longtime [32] thought to be possible accelerators to energies of the order of 1 PeV and beyond. The particle acceleration in SNRs is accompanied by production of gammas due to interactions of accelerated protons and nuclei with the ambient medium.

In particular, in SNRs with molecular clouds, a possible mechanism involves a source of cosmic rays illuminating clouds at different distances, and generating hadronic showers by pp collisions. This allows to study the generation of cosmic rays by the study of photons coming from π^0 decays in the hadronic showers.

An example of such a mechanism at work could be IC443. Recent results from AGILE and Fermi support the hypothesis. In particular, the centroid of emission by Fermi is significantly displaced from the centroid of emission from MAGIC (which, in turn, is consistent with a later measurement from VERITAS, and with a molecular

³ for instance, as particle beams as it is done in particle physics accelerators.

cloud). The spectral energy distributions also supports a two-component emissions, with a rate of production of primary electrons consistent with the rate of production of protons.

Besides indications from the studies of the morphology (see also the results on the morphology of the SNR W51 by MAGIC, FERmi and AGILE, where the emissions from a central accelerator and from a molecular cloud can be clearly separated), the detection of photons of energies of the order of 100 TeV and above could be a direct indication of production via π^0 decay, since the emission via leptonic mechanisms should be strongly suppressed at those energies ([33] and references therein) where the inverse Compton scattering cross-section enters the Klein-Nishina regime. Recently it has been possible to fit directly a π^0 peak from the SNR W44 [34].

As the energetics of SNRs might explain the production of galactic CR, the energetics of AGN might explain the production of CR up to the highest energies.

Although the spatial resolution of Cherenkov telescopes is not so good to study the morphology of extragalactic emitters, a recent study of a flare from M87 by the main Cherenkov telescopes plus the VLBA radio array (H.E.S.S., MAGIC, VERITAS and VLBA [35]) has shown, based on the VLBA imaging power, that this AGN accelerates particles to very high energies in the immediate vicinity of its central black hole (less than 60 Schwarzschild radii).

The study of the morphology of galactic sources continues, and is telling us more and more, also in the context of multiwavelength analyses; in the future, the planned Cherenkov Telescope Array (CTA) will give the possibility to explore the highest energies, and to contribute, together with high energy CR arrays and possibly with neutrino detectors, to the final solution of the CR problem. Incidentally, Centaurus A is a very-high-energy gamma emitter.

Cosmic gamma rays could perhaps come from new particles interacting or decaying; such an emission is for example foreseen in the self-annihilation of the neutralino. Also in this case the multi wavelength and multi messenger approach is crucial to separate the possible signal from the astrophysical background.

Finally, also high energy neutrinos may be used as probes for the violent processes that are occurring in supernovas and their remnants. With the discovery of several PeV neutrinos, by the IceCube experiment, and with the foreseen giant neutrino observatories, the field of neutrino astrophysics is open and might bring important information about our surrounding Universe.

5 Exercises

5.1 Extensive Air Showers

The shower main characteristic can be obtained using a simple Heitler model [36].

1. In a first step, assume that our shower is initiated by a photon—a particle having only electromagnetic interaction. Let E_0 be the energy of the primary particle and

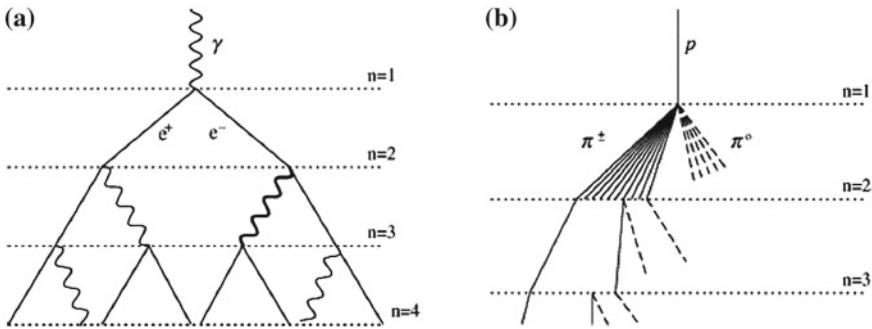


Fig. 37 Schematic views of **a** an electromagnetic cascade and **b** a hadronic shower. In the hadron shower, dashed lines indicate neutral pions which do not re-interact, but quickly decay, yielding electromagnetic subshowers (not shown). Not all pion lines are shown after the $n = 2$ level. Neither diagram is to scale

consider that the electrons, positrons and photons in the cascade always interact after travelling a certain atmospheric depth d , and that the energy is always equally shared between the two particles. With this assumptions, we can schematically represent the cascade as in Fig. 37a.

- Write the analytical expressions for the number of particles and for the energy of each particle at depth X as a function of d , n and E_0 .
 - The multiplication of the cascade stop when the particles reach a critical energy, E_c (when the decay probability surpasses the interaction probability). Using the expressions obtained in the previous question, write as a function of E_0 , E_c and $\lambda = d/\ln(2)$, the expressions, at the shower maximum, for:
 - the average energy of the particles,
 - the number of particles, N_{max} ,
 - the atmospheric depth X_{max} .
- Consider a shower initiated by a proton of energy E_0 . We will describe it with the simple model of Fig. 37b: after each depth d an equal number of pions, n_π , and each of the 3 types is produced: π^0 , π^+ , π^- . Neutral pions decay through $\pi^0 \rightarrow \gamma\gamma$ and their energy is transferred to the electromagnetic cascade. Only the charged pions will feed the hadronic cascade. We consider that the cascade ends when these particles decay as they reach a given decay energy E_{dec} , after n interactions, originating a muon (plus an undetected neutrino).
 - How many generations are needed to have more than 90 % of the primary energy, E_0 in the electromagnetic component?
 - Assuming the validity of the superposition principle, according to which a nucleus of mass number A and energy E_0 behaves like A nucleons of energy E_0/A . Derive expressions for:
 - the depth where this maximum is reached, X_{max} ,

- (ii) the number of particles at the shower maximum,
- (iii) the number of muons produced in the shower, N_μ .

5.2 Acceleration and Propagation

1. The transparency of the Universe to a given particle depends critically on its nature and energy. In fact, whenever it is possible to open an inelastic channel of the interaction between the *travelling* particle and the Cosmic Microwave Background, its mean free path diminishes drastically. Assuming that the only relevant phenomena that rules the mean free path of the *travelling* particle is the CMB (CvB), estimate the order of magnitude energies at which the transparency of the Universe changes significantly, for:

- (a) Photons;
- (b) Protons;
- (c) Neutrinos.

$$\langle E_{\gamma_{CMB}} \rangle \approx 0.24 \text{ meV} ; \langle E_{\nu_{CvB}} \rangle \approx 0.17 \text{ meV}$$

$$m_e = 511 \text{ KeV} ; m_{\Delta^+} \approx 1.2 \text{ GeV} ; m_p \approx 1 \text{ GeV} ; m_\pi \approx 0.14 \text{ GeV} ; m_{Z^0} \approx 91 \text{ GeV}$$

2. Show that the second order Fermi acceleration is proportional to β^2 .
3. Repeat the previous exercises for the first order Fermi acceleration, i.e. consider again a cosmic ray scattering at magnetic irregularities that are now separated by a planar shock moving with velocity V . Show that in this case it is proportional to β .
4. Calculate, for both acceleration mechanism, how many times has the particle to cross the cloud(shock) to gain a factor 10 on its initial energy. Assume $\beta = 10^{-4}$ for the magnetic cloud and $\beta = 10^{-2}$ for the shock wave.
5. Repeat the previous exercise assuming $\beta = 10^{-4}$ for both acceleration mechanisms.

5.3 Imaging Array Cherenkov Telescopes

In the *isothermal* approximation, the depth x of the atmosphere at a height h (i.e., the amount of atmosphere above h) can be approximated as

$$x \simeq X e^{-h/7 \text{ km}},$$

with $X \simeq 1,030 \text{ g/cm}^2$. If a shower is generated by a gamma ray of $E = 1 \text{ TeV}$ penetrating the atmosphere vertically, considering that the radiation length X_0 of air

Table 2 Shower parameters according to Rossi approximation B

	Incident electron	Incident photon
Peak of shower t_{max}	$1.0 \times (\ln y - 1)$	$1.0 \times (\ln y - 0.5)$
Centre of gravity t_{med}	$t_{max} + 1.4$	$t_{max} + 1.7$
Number of e^+ and e^- at peak	$0.3y/\sqrt{\ln y - 0.37}$	$0.3y/\sqrt{\ln y - 0.31}$
Total track length	y	y

$y = E/E_c$; unit of length is the radiation length

is approximately 36.6 g/cm² (440 m) and its critical energy E_c is about 88 MeV (Table 2):

- (a) Calculate the height h_M of the maximum of the shower in the Heitler model and in the Rossi approximation B.

If 2,000 useful Cherenkov photons per radiation length are emitted by charged particles in the visible and near UV,

- (b) compute the total number N_γ of Cherenkov photons generated by the shower
(note: the critical energy is larger than the Cherenkov threshold).

Supposing that the Cherenkov photons are all emitted at the centre of gravity of the shower—that in the Heitler approximation is just the maximum of the shower minus one radiation length,

- (c) compute how many photons per square meter arrive to a detector at a height h_d of 2,000 m, supposing that the average attenuation length of photons in air is 3 km, and that the light pool can be derived by a opening of $\sim 1.3^\circ$ from the shower maximum (1.3° is the Cherenkov angle and 0.5° comes from the intrinsic shower spread). Comment on the size of a Cherenkov telescope, considering an average reflectivity of the mirrors (including absorption in transmission) of 70 %, and a photodetection efficiency (including all the chains of acquisition) of 20 %.

- (d) Redo the calculations for $E = 50$ GeV, and comment.

6 Exercises Solutions

6.1 Extensive Air Showers

1. (a) At the level n ,

$$X = n \times d \quad (17)$$

and the number of produced particles is simply,

$$N = 2^n. \quad (18)$$

As the energy of the particles are distributed equally, the energy of the particles is the primary energy divided by the number of particles in this level, i.e.,

$$E_i = \frac{E_0}{2^n} \quad (19)$$

(b) (i) By construction,

$$E = E_c \quad (20)$$

(ii)

$$N_{max} = \frac{E_0}{E_c} \quad (21)$$

(iii) Using (18),

$$N_{max} = 2^{n_{max}} \Leftrightarrow n_{max} = \frac{\ln(N_{max})}{\ln(2)} \quad (22)$$

Since $d = \lambda \log(2)$ the maximum atmospheric depth can be written as,

$$X_{max} = n_{max} \times d = \frac{\ln(N_{max})}{\ln(2)} d = \lambda \ln \left(\frac{E_0}{E_c} \right) \quad (23)$$

where it was used (22) and (21) to substitute n_{max} and N_{max} , respectively.

2. In this case:

$$N_{tot} = n_\pi^n ; N_{ch} = \left(\frac{2}{3} n_\pi\right)^n ; E_i = \frac{E_0}{n_\pi^n} \quad (24)$$

where N_{ch} is the number of charged particles at the level n and n_π the number of pions produced at each interaction.

(a) At each interaction 1/3 of the energy goes into the electromagnetic channel through the π^0 decay. Therefore the energy that remains for the charged particles is,

$$E_{ch} = \left(\frac{2}{3}\right)^n E_0. \quad (25)$$

Thus the of electromagnetic energy rises as $E_{em} = E_0 - E_{ch}$. Hence,

$$\frac{E_{em}}{E_0} = 1 - \left(\frac{2}{3}\right)^n \quad (26)$$

Taking $E_{em}/E_0 = 0.9$ and inverting (26) to obtain the number of generations (levels), one gets,

$$n = \frac{\log(0.1)}{\log(2/3)} \approx 5.7 \text{ generations.} \quad (27)$$

- (b) (i) Let us start by evaluate X_{max} for protons. In this case from (24) and recalling that the shower development stops when the energy of the particles reaches E_{dec} one obtains,

$$X_{max} = d \times n_{dec}. \quad (28)$$

The maximum number of generations is $E_{max} = E_0/n_\pi^{n_{dec}}$. Inverting this last expression one gets,

$$n_{dec} = \frac{\log(E_0/E_{dec})}{\log(n_\pi)}, \quad (29)$$

which leads, using (28), to,

$$X_{max} = d \frac{\log(E_0/E_{dec})}{\log(n_\pi)} \quad (30)$$

For iron we have 56 nucleons (i.e. the atomic number $A = 26$ protons + 30 neutrons). Using the superposition principle each nucleon carries $E_0/56$ of the primary energy. Substituting in (31),

$$X_{max} = d \frac{\log\left(\frac{E_0}{AE_{dec}}\right)}{\log(n_\pi)} = \frac{d}{\log(n_\pi)} \left(\log\left(\frac{E_0}{E_{dec}}\right) - \log(A) \right) \quad (31)$$

Notice that the X_{max} evolution with energy is the same for proton and iron and the curves are separated by a constant term: $\log(A)$.

- (ii) Again starting with protons we have for the number of particles at the shower maximum,

$$N_{max} = n_\pi^{n_{dec}} = \frac{E_0}{E_{dec}} \quad (32)$$

For iron primaries,

$$N_{max} = A n_\pi^{n_{dec}} \quad (33)$$

Using the superposition principle and the result of (29) is easy to see that,

$$N_{max} = An_\pi^{\frac{\log\left(\frac{E_0}{AE_{dec}}\right)}{\log(n\pi)}} = A \frac{E_0}{AE_{dec}} = \frac{E_0}{E_{dec}} \quad (34)$$

which means that the number of particles at the shower maximum does not depend on the primary mass composition.

- (iii) The number of muons in the shower, for this simplified model, is given by,

$$N_\mu = N_{ch}|_{X=X_{max}} = \left(\frac{2}{3}n_\pi\right)^{n_{dec}} \quad (35)$$

Therefore, for proton primaries, using (29),

$$\begin{aligned} N_\mu &= \left(\frac{2}{3}n_\pi\right)^{\frac{\log(E_0/E_{dec})}{\log(n\pi)}} \\ &= \left[\left(\frac{2}{3}n_\pi\right)^{\log\frac{2}{3}n_\pi(E_0/E_{dec})}\right]^{\frac{\log\left(\frac{2}{3}n_\pi\right)}{\log(n\pi)}} \\ &= \left(\frac{E_0}{E_{dec}}\right)^{\frac{\log\left(\frac{2}{3}n_\pi\right)}{\log(n\pi)}} = \left(\frac{E_0}{E_{dec}}\right)^\beta \end{aligned} \quad (36)$$

where β is a parameter related with the hadronic interactions multi particle production, in particular, the ratio between the hadronic and the electromagnetic component of the interaction.

For iron, using again the superposition principle and the final result of (36) one gets,

$$N_\mu = A \left(\frac{E_0/A}{E_{dec}}\right)^\beta = A^{1-\beta} \left(\frac{E_0}{E_{dec}}\right)^\beta. \quad (37)$$

6.2 Acceleration and Propagation

1. (a) For photons the dominant process is the interaction with the photons of the cosmic microwave background, γ_{CMB} , through pair creation,

$$\gamma + \gamma_{CMB} \rightarrow e^+ + e^- \quad (38)$$

To determine the minimum energy at which this process can occur it is useful to compute the inner product of the quadri-momentum vector. This quantity is a Lorentz invariant and thus we can easily relate quantities in the Laboratory framework with the ones in the centre-of-mass framework.

For convenience, the calculations for the photons (before the interaction) will be considered in the Lab while the products of such interaction will be considered in the centre-of-mass.

Therefore,

$$P_\mu^{LAB} = (E_\gamma + E_b, \mathbf{P}_b + \mathbf{P}_\gamma) \quad (39)$$

$$P_\mu^{CM} = (2m_e, \mathbf{0}) \quad (40)$$

where it was used the nomenclature $E_b \equiv E_{\gamma CMB}$. Using,

$$s = (P_\mu P^\mu)_{LAB} = (P_\mu P^\mu)_{CM} \quad (41)$$

$$P_\mu P^\mu = E^2 - \mathbf{P} \cdot \mathbf{P} \quad (42)$$

one obtains the following equation,

$$4m_e^2 = (E_b + E_\gamma)^2 - (P_b^2 + P_\gamma^2 + 2\mathbf{P}_b \cdot \mathbf{P}_\gamma) \quad (43)$$

Taking into account that for photons $E = P$ and that $\mathbf{P}_1 \cdot \mathbf{P}_2 = |P_1||P_2|\cos(\theta)$, one gets that the energy of the incoming photon is given by,

$$E_\gamma = \frac{2m_e^2}{E_b(1 - \cos\theta)} \quad (44)$$

Notice that we are looking for the minimal energy that allows for this process to happen, so $\cos(\theta) = -1$. Inputting the values given in the problem one gets,

$$E_\gamma = \frac{m_e^2}{E_b} \sim 10^{14} \text{ eV} \quad (45)$$

(b) For protons the dominant inelastic channel is via,

$$\begin{aligned} p + \gamma_{CMB} &\rightarrow (\Delta^+) \rightarrow p + \pi^0 \\ p + \gamma_{CMB} &\rightarrow (\Delta^+) \rightarrow n + \pi^+ \end{aligned} \quad (46)$$

As in the previous problem, we want to find the minimum energy of the proton for which the process is possible. Moreover, we will use the Lorentz invariant s and consider once again that the proton and the gamma are in the Lab framework and the products of the interaction in the centre-of-mass framework. Let us then start by defining our cinematic,

$$P_\mu^{LAB} = (E_p + E_b, \mathbf{P}_p + \mathbf{P}_b) \quad (47)$$

$$P_\mu^{CM} = (m_p + m_\pi, \mathbf{0}) \quad (48)$$

Therefore, using (42) one can write the following equation,

$$(E_b + E_p)^2 - P_b^2 - P_p^2 - 2\mathbf{P}_b \cdot \mathbf{P}_p = (m_p + m_\pi)^2 \quad (49)$$

Solving the above equation in order to E_p and recalling that $P_p = \sqrt{E_p^2 + m_p^2}$ one gets,

$$2E_b E_p - 2E_b \sqrt{E_p^2 + m_p^2} \cos(\theta) = m_\pi^2 + 2m_p m_\pi \quad (50)$$

Since $E_p \gg m_p$ the proton momentum can be approximated by its energy ($P_p \approx E_p$). Therefore,

$$E_p = \frac{m_\pi^2 + 2m_p m_\pi}{4E_b} \sim 6 \times 10^{19} \text{ eV} \quad (51)$$

where the $\cos(\theta)$ was taken to be -1 , i.e. the proton and the photon have opposite directions.

- (c) The Universe becomes opaque to the neutrinos when they the following inelastic interaction channel opens,

$$\nu + \nu_{C\nu B} \rightarrow Z^0 \quad (52)$$

Again considering that the interaction occurs in the Lab framework and noticing that the interaction occurs if there is enough energy to produce a Z^0 at rest, one can write,

$$P_\mu^{LAB} = (E_\nu + E_b, \mathbf{P}_\nu + \mathbf{P}_b) \quad (53)$$

$$P_\mu^{CM} = (m_{Z^0}, \mathbf{0}) \quad (54)$$

and similarly as before one can write the following expression,

$$(E_b + E_\nu)^2 - P_b^2 - P_\nu^2 - 2\mathbf{P}_b \cdot \mathbf{P}_\nu = m_{Z^0}^2 \quad (55)$$

$$E_\nu = \frac{m_{Z^0}^2}{2E_b(1 - \cos \theta)} \quad (56)$$

Therefore, the minimum energy of the neutrino that allows the process in (52) is,

$$E_\nu = \frac{m_{Z^0}^2}{4E_b} \sim 10^{24} \text{ eV} \quad (57)$$

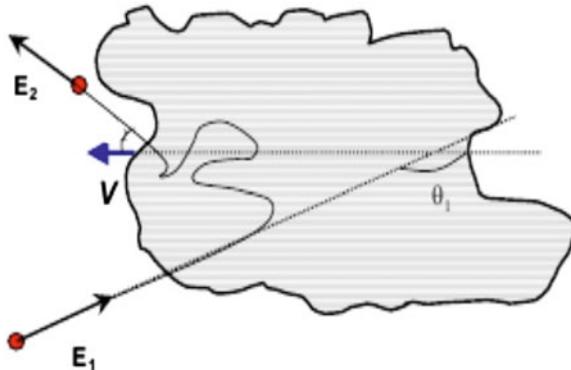


Fig. 38 Illustration of the problem main variables

2. Let us start by considering that a particle of energy E_1 encounters a cloud with velocity V (see Fig. 38). The angle at which the particle enters in cloud is θ_1 . After some scatterings in the cloud the particle should leave the cloud with an energy E_2 and θ_2 , again in the reference frame. To compute the energy of the particle in the reference frame of the cloud, E_1^* , one can simply use the Lorentz transformations,

$$E_1^* = \gamma E_1 (1 - \cos \theta_1) \quad (58)$$

where $\beta = V/c$, $\gamma = (1 - \beta^2)^{-1/2}$ and we have neglected the particle mass with respect to its kinetic energy.

In the cloud reference frame, the internal energy gets conserved and so $E_1^* = E_2^*$. This is the equivalent of a classical collision to a wall. The expression for the energy after the collision in the Laboratory framework is given by,

$$E_2 = \gamma E_2^* (1 + \cos \theta_2^*) \quad (59)$$

Using (58) and (59) and recalling that $E_1^* = E_2^*$ one can write the relative energy gain for the particle,

$$\frac{\Delta E}{E} = \frac{E_2 - E_1}{E_1} = \frac{1 - \beta \cos \theta_1 + \beta \cos \theta_2^* - \beta^2 \cos \theta_1 \cos \theta_2^*}{1 - \beta^2} - 1 \quad (60)$$

Let us now compute the average energy gain. For that we have to make some considerations about the particle average entry/escaping angles. First, it is easy to see that as the particle suffers a great number of collisions while inside the cloud, and so it should lose the memory of the initial direction. Thus,

$$\langle \cos \theta_2^* \rangle = 0, \quad (61)$$

leading to,

$$\left\langle \frac{\Delta E}{E} \right\rangle = \frac{1 - \beta \langle \cos \theta_1 \rangle}{1 - \beta^2} - 1. \quad (62)$$

We have now to compute $\langle \cos \theta_1 \rangle$, i.e. the probability of the particle with velocity v to encounter a cloud moving with a velocity V . Clearly this depends on θ_1 and on the relative velocity, i.e. it is more probable that the particle hits the cloud that is coming against it, than one that it is running away from it. Hence,

$$P \propto (v - V \cos \theta_1) \sim c(1 - \beta \cos \theta_1). \quad (63)$$

Therefore using the above proportionality,

$$\langle \cos \theta_1 \rangle = \frac{\int_{-1}^1 \cos \theta_1 (1 - \beta \cos \theta_1) d \cos \theta_1}{\int_{-1}^1 (1 - \beta \cos \theta_1) d \cos \theta_1} = -\frac{\beta}{3} \quad (64)$$

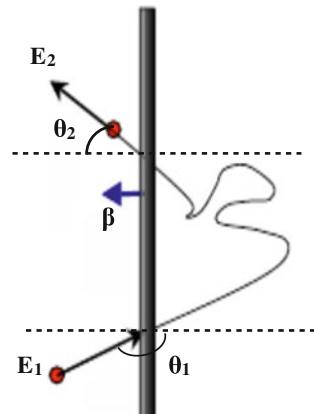
Substituting in (62), one gets that the average energy increases by,

$$\left\langle \frac{\Delta E}{E} \right\rangle \approx \frac{4}{3} \beta^2. \quad (65)$$

3. In this case the particle encounters a diffusive shock wave that is travelling through space with a velocity, V . The particle enters the shock wave carrying an energy E_1 and with an angle θ_1 (upstream). While in the shock gas the particle can undergo several elastic interactions, due to the turbulent magnetic field, and may invert its initial direction crossing once again the shock front (downstream), now with an energy E_2 and an angle θ_2 (see Fig. 39).

Thus, as in the previous problem, the particle will move from a Laboratory framework to a high density region moving at a velocity V . Hence, the relative energy change, derived in the previous problem, is given by,

Fig. 39 Illustration of the problem main variables



$$\frac{\Delta E}{E} = \frac{1 - \beta \cos \theta_1 + \beta \cos \theta_2^* - \beta^2 \cos \theta_1 \cos \theta_2^*}{1 - \beta^2} - 1. \quad (66)$$

However, the entry and escaping angles are now constrained to a specific geometry, as it is shown in Fig. 39,

$$-1 \leq \cos \theta_1 \leq 0 \quad (67)$$

$$0 \leq \cos \theta_2^* \leq 1 \quad (68)$$

On the other hand, the probability of the particle to collide with the shock wave is simply $\cos \theta_1$ (upstream) and $\cos \theta_2^*$ (downstream).

Therefore the mean values for $\cos \theta_1$ and $\cos \theta_2^*$ are, respectively,

$$\langle \cos \theta_1 \rangle = \frac{\int_{-1}^0 \cos^2 \theta_1 d \cos \theta_1}{\int_{-1}^0 \cos \theta_1 d \cos \theta_1} = -\frac{2}{3} \quad (69)$$

$$\langle \cos \theta_2^* \rangle = \frac{\int_0^1 \cos^2 \theta_2^* d \cos \theta_2^*}{\int_0^1 \cos \theta_2^* d \cos \theta_2^*} = \frac{2}{3} \quad (70)$$

Consequently, the energy gain, at each passage, is,

$$\frac{\Delta E}{E} \approx \frac{4}{3} \beta \quad (71)$$

4. At each passage through the cloud or shock wake the particle gains some energy that is proportional to its energy. Therefore for n crosses the energy of the particle relatively to its initial energy, E_0 is given by,

$$E = E_0(1 + \varepsilon)^n \quad (72)$$

where ε is the gain and it is proportional to β for the acceleration in a shock wave (Fermi first order acceleration mechanism) and proportional to β^2 for the *collision* with the magnetic cloud (Fermi second order acceleration mechanism). β is the velocity of the astrophysical object (shock wave or cloud).

Inverting (72) one can obtain the number of times that a particle should cross in order to increase its energy from E_0 to E ,

$$n = \frac{\log \left(\frac{E}{E_0} \right)}{\log(1 + \varepsilon)} \quad (73)$$

In this problem we want to know how many times a particle should cross a cloud or a shock wave to increase its energy by a factor of 10, so $E/E_0 = 10$.

Finally using (73) we have,

$$n(\varepsilon \propto \beta; \beta = 10^{-2}) \approx 2.3 \times 10^2 \text{ cycles} \quad (74)$$

$$n(\varepsilon \propto \beta^2; \beta = 10^{-4}) \approx 2.3 \times 10^8 \text{ cycles} \quad (75)$$

Therefore, in realistic astrophysical conditions the particle needs to cross only 230 times the shock wave to gain a factor of 10 on its energy while it should cross a cloud 2.3×10^8 times to gain the same energy.

5. Here we assume that both the shock wave and the magnetic cloud have the same velocity $\beta = 10^{-4}$.

$$n(\varepsilon \propto \beta; \beta = 10^{-4}) \approx 2.3 \times 10^4 \text{ cycles} \quad (76)$$

$$n(\varepsilon \propto \beta^2; \beta = 10^{-4}) \approx 2.3 \times 10^8 \text{ cycles} \quad (77)$$

Even considering the same velocity for the two astrophysical phenomena the Fermi first order acceleration need 10,000 times less cycles than the Fermi second order mechanism.

6.3 Imaging Array Cherenkov Telescopes

- (a) Let us start by computing the depth of the shower maximum, X_{max} , for the two models and afterwards convert it into an altitude using the atmosphere model provided in the problem,

$$X_{max} \approx X e^{-\frac{h}{7\text{km}}} \quad (78)$$

For the heitler model, as seen in Exercise 6.1, is given by,

$$X_{max} = X_0 \left(1 + \frac{\log(E/E_c)}{\log(2)} \right) \approx 530 \text{ g cm}^{-2} \quad (79)$$

where $E = 1 \text{ TeV}$, $E_c = 88 \text{ MeV}$ and $X_0 = 36.6 \text{ g cm}^{-2}$.

For the Rossi approximation B model,

$$X_{max} = X_0 \left(\log \left(\frac{E}{E_c} \right) - 0.5 \right) \approx 325 \text{ g cm}^{-2} \quad (80)$$

Inverting (78) to obtain the height of X_{max} one gets,

$$h_M = -7 \log \left(\frac{X_{max}}{X} \right) \quad (81)$$

Hence,

- Heitler: $h_M \approx 4.9$ km
- Rossi: $h_M \approx 8.6$ km
- MC: $h_M \approx 9.5$ km

where MC is the value for a full Monte Carlo EAS simulation.

- (b) The total number of Cherenkov photons generated by the shower is,

$$N_{\gamma}^{total} = \left(\frac{E}{E_c} \right) N_{\gamma}^{Ch} \approx 2.27 \times 10^7 \text{ photons} \quad (82)$$

where N_{γ}^{Ch} is the number of Cherenkov photons per radiation length and (E/E_c) the total track length in radiation units.

- (c) In this problem we will use the Rossi approximation B model to evaluate the shower main characteristics for a 1 TeV photon induced shower.

It shall be assumed that all the photons are coming from the centre of gravity of the shower, t_{med} . This quantity can be computed using,

$$t_{med} = t_{max} + 1.7 = \log \left(\frac{E}{E_c} \right) - 0.5 + 1.7 \approx 10.5 \quad (83)$$

or in traversed matter units,

$$X_{med} = t_{med} X_0 \approx 385.7 \text{ g cm}^2 \quad (84)$$

Using the *isothermal* approximation as atmosphere model one obtains for the altitude,

$$h_{med} = -7 \log \left(\frac{X_{med}}{X} \right) \approx 6.88 \text{ km}. \quad (85)$$

Using this altitude as emission point and knowing that the light pool can be derived from an opening angle of $\sim 1.3^\circ$ (see Fig. 40), one gets for r_p ,

$$r_p = \tan(\theta)(h_{med} - h_d) \approx 110.6 \text{ m}. \quad (86)$$

Therefore the number of photons at ground is,

$$n_{\gamma} = \frac{N_{\gamma}|_{ground}}{A} = \frac{N_{\gamma} e^{-(h_{med}-h_d)/3}}{\pi r_p^2} = 116 \text{ photons m}^{-2} \quad (87)$$

where A is the area of the light pool at ground and $N_{\gamma}|_{ground}$ is the number of Cherenkov photons that reach the ground. N_{γ} is the total number of photons produced by the shower, calculated in the previous problem, while the exponential term represents the attenuation of these photons while travelling through the atmosphere.

IACTs telescopes of VERITAS and HESS have areas, A_{det} of $\sim 100 \text{ m}^2$. Thus,

Fig. 40 Illustration of the problem main variables

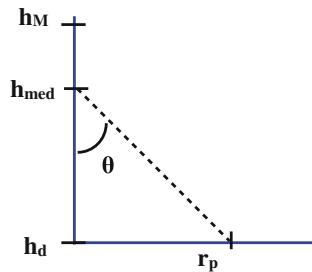


Table 3 Comparison table of the parameters evaluated for the resolution of this exercise

Energy (GeV)	t_{med}	X_{med} (g cm $^{-2}$)	h_{med} (km)	r_p (m)	N_γ (ph)	n_γ (ph m $^{-2}$)	n_γ^{det} (ph)
50	7.5	276.0	9.22	163.7	1.1×10^6	1	17
1,000	10.5	385.7	6.88	110.6	2.2×10^7	116	1,629

the number of detected photons is,

$$n_\gamma^{det} = n_\gamma A_{det} \varepsilon_{ref} \varepsilon_{acq} \approx 1,629 \text{ photons} \quad (88)$$

where ε_{ref} and ε_{act} are the reflective and acquisition efficiencies, respectively.

- (d) This problem is solved in the same way than the previous one taking into account that the primary energy is now of 50 GeV. We summarise the difference between the parameters necessary to compute the number of Cherenkov photons detected by the IACT experiments in Table 3.

For showers induced by photons with $E = 50$ GeV the number of detected Cherenkov photons is extremely low (of the order of the background fluctuations). Therefore these kind of IACT are not suited to measure gamma-ray induced showers bellow 50 GeV.

References

1. E. Fermi, Phys. Rev. **75**, 1169 (1949)
2. A.W. Strong, I.V. Moskalenko, V.S. Ptuskin, Annu. Rev. Nucl. Part. Sci. **57**, 285 (2007). doi:[10.1146/annurev.nucl.57.090506.123011](https://doi.org/10.1146/annurev.nucl.57.090506.123011)
3. A.A. Penzias, R.W. Wilson, Astrophys. J. **142**, 419 (1965). doi:[10.1086/148307](https://doi.org/10.1086/148307)
4. K. Greisen, Phys. Rev. Lett. **16**, 748 (1966). doi:[10.1103/PhysRevLett.16.748](https://doi.org/10.1103/PhysRevLett.16.748)
5. G.T. Zatsepin, V.A. Kuzmin, JETP Lett. **4**, 78 (1966)
6. J.W. Cronin, Nucl. Phys. Proc. Suppl. **138**, 465 (2005). doi:[10.1016/j.nuclphysbps.2004.11.107](https://doi.org/10.1016/j.nuclphysbps.2004.11.107)
7. <http://home.web.cern.ch/about/experiments/ace>
8. <http://asd.gsfc.nasa.gov/bess/>
9. <http://pamela.roma2.infn.it/index.php>
10. <http://www.ams02.org/>
11. K. Greisen, Annu. Rev. Nucl. Sci. **10**, 63 (1960)
12. K. Kamata, J. Nishimura, Prog. Theor. Phys. Suppl. **6**, 93 (1958)

13. <http://www-akeno.icrr.u-tokyo.ac.jp/AGASA/>
14. J.A. Abraham et al., Nucl. Instrum. Methods **A620**, 227 (2010). doi:[10.1016/j.nima.2010.04.023](https://doi.org/10.1016/j.nima.2010.04.023)
15. <http://www.telescopearray.org/>
16. <http://www.auger.org/>
17. <http://jemeuso.riken.jp/en/>
18. J. Abraham et al., Phys. Lett. **B685**, 239 (2010). doi:[10.1016/j.physletb.2010.02.013](https://doi.org/10.1016/j.physletb.2010.02.013)
19. J. Abraham et al., Phys. Rev. Lett. **104**, 091101 (2010). doi:[10.1103/PhysRevLett.104.091101](https://doi.org/10.1103/PhysRevLett.104.091101)
20. R. Glauber, G. Matthiae, Nucl. Phys. **B21**, 135 (1970)
21. P. Abreu et al., Phys. Rev. Lett. **109**, 062002 (2012). doi:[10.1103/PhysRevLett.109.062002](https://doi.org/10.1103/PhysRevLett.109.062002)
22. R. Durrer, A. Neronov, Cosmological Magnetic Fields: Their Generation, Evolution and Observation (2013) [ARXIV:1303.7121](https://arxiv.org/abs/1303.7121)
23. M.S. Longair, Introduction to high energy astrophysics (2001), pp. 3–48
24. A. De Angelis, G. Galanti, M. Roncadelli, Mon. Not. R. Astron. Soc. **432**, 3245 (2013). doi:[10.1093/mnras/stt684](https://doi.org/10.1093/mnras/stt684)
25. <http://fermi.gsfc.nasa.gov>
26. <http://www.hawc-observatory.org>
27. <http://www.mpi-hd.mpg.de/hfm/HESS>
28. <https://magic.mpp.mpg.de>
29. <http://veritas.sao.arizona.edu>
30. <http://www.cta-observatory.org>
31. F. Halzen, S.R. Klein, Rev. Sci. Instrum. **81**, 081101 (2010). doi:[10.1063/1.3480478](https://doi.org/10.1063/1.3480478), <http://icecube.wisc.edu>
32. W. Baade, F. Zwicky, Proc. Natl. Acad. Sci. **20**, 259 (1934)
33. F.A. Aharonian, C.W. Akerlof, Annu. Rev. Nucl. Part. Sci. **47**, 273 (1997). doi:[10.1146/annurev.nucl.47.1.273](https://doi.org/10.1146/annurev.nucl.47.1.273)
34. A. Giuliani et al., Astrophys. J. **742**, L30 (2011). doi:[10.1088/2041-8205/742/2/L30](https://doi.org/10.1088/2041-8205/742/2/L30)
35. V. Acciari et al., Science **325**, 444 (2009). doi:[10.1126/science.1175406](https://doi.org/10.1126/science.1175406)
36. H.J. Bhabha, W. Heitler, Proc. R. Soc. Lond. Ser. A **159**, 432 (1937)

Numerical Methods and Data Analysis Techniques in Particle Physics

Mike Williams

Abstract Experimental physics is now in the era of big data. Modern experiments are collecting petabytes of data and physicists are trying to extract from these vast data sets information about the nature of the universe. Gone are the days when data analysis was done by eye and where physicists relied on the algorithms hardcoded into the human brain. Modern analyses have replaced brain power with CPU power. From the perspective of the physicist, statistical methods are tools; they are a means to an end. As an analogy, consider the car. People use cars to get from point *A* to point *B*. Most people don't care how or why a car works. In fact, many people think that car enthusiasts (often referred to as *gearheads*) are strange. You certainly don't need to be a gearhead to use a car, but it's a good idea to learn how to drive properly. The consequences of not doing so can be catastrophic! The same can be said for physicists when it comes to statistical analysis. The point of these lectures is not to make you a *statshead*. The goal is to teach you how to use statistical methods properly so you don't metaphorically drive off of a cliff in your analysis. My lectures covered the following list of topics: machine learning; resampling; two-sample testing; parametric and non-parametric regression; multivariate goodness-of-fit tests; Bayesian versus Frequentist statistics; and limit setting. I gave three 1-h lectures covering all of these topics. This was not enough time to go into great detail on any topic. Instead, I chose to try and convey the important aspects of each and have provided references so you can learn more on your own if you desire to.

1 Machine Learning

Machine learning involves learning to accurately make predictions by studying data provided for training. Here, the physicist is the teacher and the machine is the student. The teacher provides the data sample on which the student will train. I will only discuss what is known as supervised learning, where the teacher gives the student

M. Williams (✉)

Massachusetts Institute of Technology (MIT), 77 Massachusetts Avenue, Bldg. 26-445,
Cambridge, MA 02139, USA
e-mail: mwill@mit.edu

the answers during training. The most common usage of machine learning in physics is for classification, e.g., identifying an event as signal or background. The basic strategy is for the teacher (physicist) to provide the student (machine) with data samples for which the class (signal or background) is known for each event. The machine will learn how best to classify events from this sample and then apply what it has learned to data where the classification is not known. A few notes before moving on: it is possible to deal with any number of classes of data, but in physics we tend to only use two (signal and background); the methods I will discuss can also be used for regression but I will not discuss this.

The two most common classifiers used in physics are boosted decision trees (BDTs) and artificial neural networks (ANNs). There are plenty of other classifiers available on the market; however, I will only cover BDTs and ANNs due to lack of time. For most real-world applications, what matters is how the training is done and not which classifier is used; i.e., it is the teacher that matters not the student. There are counter examples, of course, but generally the difference in performance between classifiers is small. Despite this fact, many physicists can be adamant about the merits of one type of classifier over all others. An analogy here can be made to smartphones. For the vast majority of people, their smartphone usage is such that either an iPhone or Android would perfectly suit their needs; however, large camps have been formed that are passionate about the superiority of one of these types of phones. In physics, unfortunately, this is also often true of BDT and ANN *enthusiasts*.

1.1 Boosted Decision Trees

I will start by describing a decision tree (DT). Consider a pooled sample of data containing signal and background. The classification of each event is known. Each event is described by an D -dimensional vector of variates \mathbf{x} . These variates could be kinematic quantities, track quality variables, particle identification (PID) information, etc. The goal is to maximize some figure of merit (FOM), e.g., purity or the Gini index.

A DT is a fairly simple object. To build one we start with the full pooled sample. A loop over each variate in \mathbf{x} is performed. For each variate the point at which splitting the data into two subsamples produces the largest FOM is determined. Only one split is actually made at this step: the data is split into two subsamples using the variable x_i that provides the largest FOM of any variable in \mathbf{x} . Next, the same procedure is followed for splitting the two subsamples of the data. This process is repeated repeatedly to build the DT. Figure 1 shows a toy schematic diagram of a DT. In general, one continues splitting subsamples until one of the following conditions is reached: the FOM cannot be improved by splitting the subsample; a pre-defined maximum number of subsamples has been reached; or the subsample has the pre-defined minimum number of events allowed. The final set of subsamples are referred to as the *leaves* of the tree.

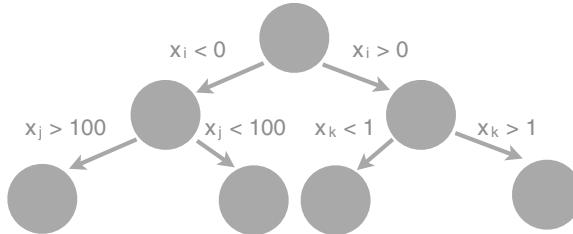


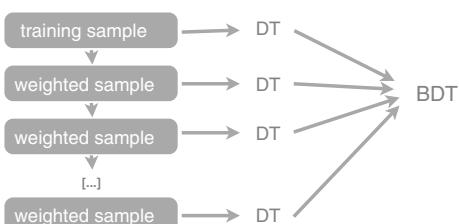
Fig. 1 Toy schematic diagram of a single decision tree

Once the DT is constructed, it can be used to classify events that were not used in the training. The performance of the DT on the training sample itself is not of great interest since the classification of these events is already known. What we really care about is the ability to classify events where we do not know the true class. One can quickly see a problem then with the approach taken in building the DT. Consider, e.g., a subsample with two events with x_i values of 2.1 and 2.2 and take the resolution on x_i to be 0.5; thus, these events are experimentally indistinguishable. One is signal and the other is background so splitting between the two data points to make two new subsamples improves whatever FOM has been chosen for this DT. Clearly, however, making such a split will not help classify events from an independent sample. In fact, it will most likely hurt performance! This effect is known as overtraining. The classifier has learned something specific to the training sample and not about the parent PDF.

One method that is commonly used to avoid overtraining is *boosting* [1]. In this approach, a series of classifiers is trained. The training sample for each member of the series is augmented based on the performance of previous members. Incorrectly classified events are assigned larger weights to boost their importance. The result is that each successive classifier is designed to improve the overall performance of the series in the regions where its predecessors have failed. In the end, the members of the series are combined to produce a single more powerful classifier than any of the individual members. Figure 2 shows a toy schematic diagram of a BDT. See [2] for a pedagogical description of boosting in the context of a physics analysis.

A general approach to avoid overtraining involves first splitting the data where the classifications are known into two subsamples: (1) the training sample and (2) the so-called validation sample. The BDT is trained on sample (1) and then evaluated on sample (2). If the performance is similar on the two samples, where the first is clearly

Fig. 2 Toy schematic diagram of a boosted decision tree



biased as it was used to create the BDT, then the performance on an independent sample is expected to be predictable. If the performance is not similar, it is advisable to evaluate the performance on a third (test) data sample and to take further measures to insure that the final product has predictable performance (otherwise it is not useful).

1.2 Bootstrapping

Let us take a quick detour now and discuss a resampling method known as the *bootstrap*. Bootstrapping is a way of simulating repeated observations using only the data itself (see, e.g., [3]) by creating many *bootstrap copy* data sets. A bootstrap copy data set is made by sampling with replacement from the original data set. If the data has n events, then a bootstrap copy data set is obtained by randomly drawing n events from the data. The copy data set may contain multiple copies of some events and no copies of others. By making many bootstrap copy data sets we can determine estimators for variances.

As an example, consider a group of 10 people whose mean height is μ (assume that none of the people are related). To estimate the uncertainty on μ , we could make 100 bootstrap copy groups from the original and determine the mean height value for each group μ_i . The variance of μ_i can be used to obtain an estimate of the uncertainty on μ . As another example consider data sampled from the PDF $f(x) = 2x$ on the unit interval. Figure 3 shows a data set with 100 events sampled from $f(x)$. A linear

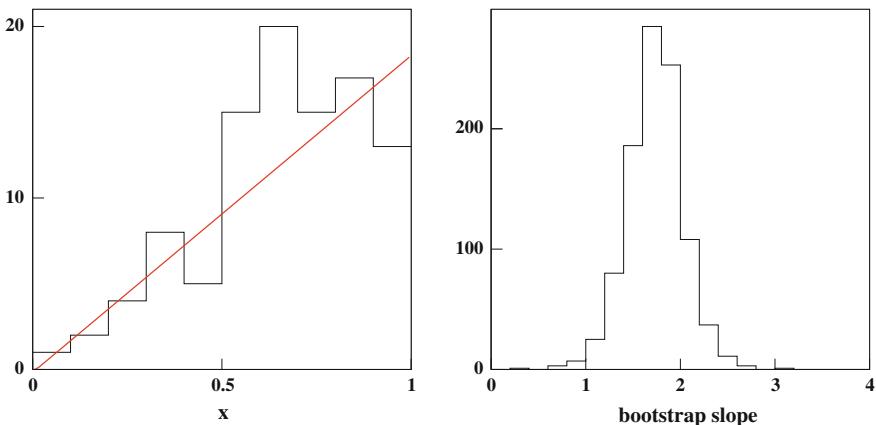


Fig. 3 Toy example of bootstrapping: (*left*) 100 events sampled from the PDF $f(x) = 2x$ on the unit interval and (*right*) slopes obtained by fitting 1,000 bootstrap copy data sets sampled from the original (*left*) one. Notice that the mean value of the bootstrap slope distribution is very close to that obtained by fitting the original data set. The bootstrap does not add information that would improve the estimate of the slope, it merely provides a way of estimating the uncertainty of the slope estimator

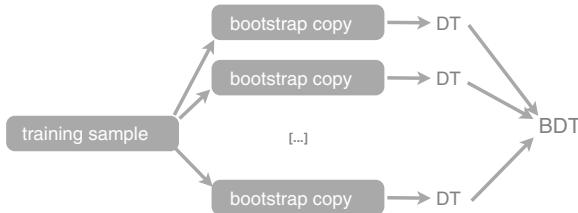


Fig. 4 Toy schematic diagram of a bagged decision tree

regression yields a slope estimator of 1.85 ± 0.26 . The estimate of the uncertainty on the slope obtained using 1,000 bootstrap copy data sets is 0.29; this is agrees well with the fit result. Of course, the bootstrap is much more useful when a simple method for estimating an uncertainty is not available.

We can apply bootstrapping to the BDT [4]. The idea is simple: make n bootstrap copy data sets from the original training data sample and train an independent DT on each. The BDT is then obtained by combining the responses of the set of DTs. Figure 4 shows a schematic diagram of the so-called boosted aggregate (or bagged) decision tree (still referred to as a BDT). The power of the BDT can often times be enhanced further by randomly selecting a subset of the variables to use in each DT [5]. This is known as the *random forest* approach to boosting.¹ The random forest increases classification power by making the DTs less correlated. It also makes training faster since fewer variables are considered at each split.

1.3 Artificial Neural Networks

ANNs send data from input neurons via synapses to a hidden layer of neurons and then to output neurons via more synapses. Figure 5 shows a schematic diagram of an ANN. Learning typically involves forward propagation of weights, followed by determination of classification error and then backwards propagation of errors to update the weights. Of course, one can also boost an ANN. There are many variants of ANNs on the market. I will not discuss them in detail here.

1.4 Evaluating Performance

Performance is often visualized using a receiver operating characteristic (ROC) curve. ROCs were originally developed during WWII for use with radar systems. Figure 6 shows an example ROC curve. ROC curves are useful; however, typically

¹ I often wonder what fraction of their time machine learning experts spend coming up with names for their algorithms.

Fig. 5 Toy schematic diagram of an artificial neural network

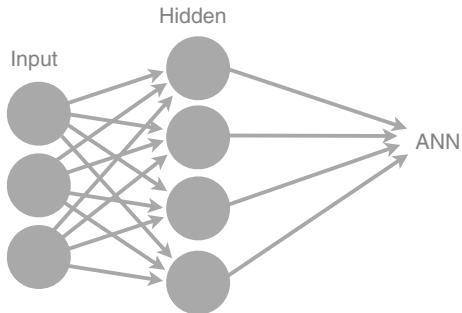


Fig. 6 Toy schematic diagram of a ROC curve



an analyst really is concerned with a specific FOM. In such cases it is better to study the performance with respect to that FOM (not the ROC).

BDTs, ANNs, etc. tend to perform roughly the same in real-world use cases even in situations where one would not expect so. For example, a BDT is naturally able to handle a mixture of discrete and continuous variables. This is the case because splitting data samples works for discrete and continuous variables. The ANN, however, is not well suited for dealing with such a mixture. The ANN constructs a variable transformation from the input to hidden layer. Such a transformation is unstable if some of the variables are discrete. This shortcoming can be overcome by running *preprocessor* algorithms or via other modifications. As long as one uses polished algorithms, such deficiencies should not be apparent to the user. Finally, I note that each of these algorithms has input parameters that can be tuned (chosen) by the user. It is important to study the dependence of the performance of whatever classifier you use on these input parameters.

1.5 Real-World Examples

Machine learning classifiers have now been used in many published analyses in particle physics. BDTs and ANNs are used for signal-background separation, but also for things such as triggering, PID, etc. In this section I quickly discuss three uses and note that the use of such techniques is now commonplace.

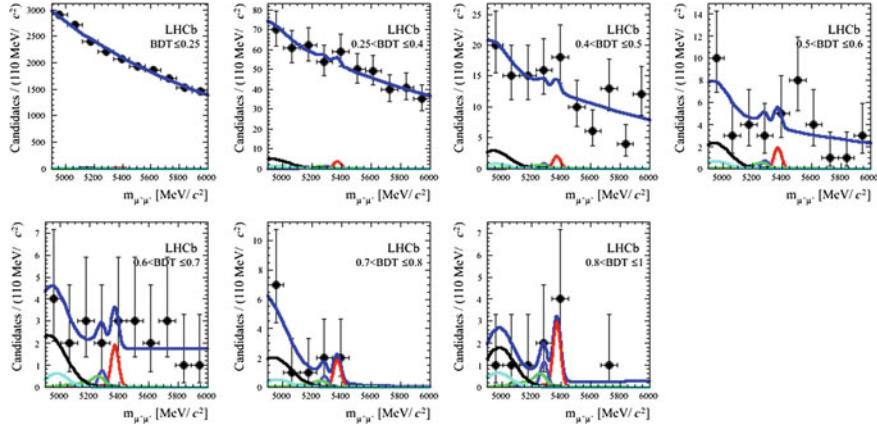


Fig. 7 First evidence for the rare decay $B_s \rightarrow \mu^+ \mu^-$ from LHCb [6]. The BDT was trained on the decay $B \rightarrow \pi^+ \pi^-$ which looks almost identical in the detector to the signal mode. The response was calibrated such that the expected signal distribution was uniform, while the background falls off rapidly with increasing BDT response. This plot shows about half of the data collected by LHCb in 2012. A roughly equal amount of data collected in 2011 was also used in this publication (but not shown in this plot). The red peaking curve near 5,366 is the $B_s \rightarrow \mu^+ \mu^-$ contribution, while the purple curve peaking near 5,280 is $B_d \rightarrow \mu^+ \mu^-$. The black and cyan curves that peak around 5,000 are partially reconstructed b -hadron decays. The green curve that peaks just below 5,280 is $B_d \rightarrow h^+ h^-$, where h denotes a kaon or pion that has been misidentified as a muon. The BDT does an excellent job separating signal from background. In the highest bins of BDT response, the only backgrounds that remain come from closely related decays of b hadrons

1.5.1 Rare Decay Search at LHCb

The standard model of particle physics (SM) predicts that the branching fraction $\mathcal{B}(B_s \rightarrow \mu^+ \mu^-)$ is about 3.5 parts per billion. Prior to the LHC turning on, this process had never been observed and the best upper limit on the branching fraction was about 12 times the SM prediction. The branching fraction of this decay is greatly enhanced (or reduced) by many extensions to the SM, e.g., supersymmetric extensions. This is a very rare process and the LHC is a *dirty* environment. Figure 7 shows the data observed by the LHCb collaboration [6] and published as the first evidence for this rare process.² One can see that the BDT does an excellent job of separating the signal from the background.

1.5.2 Higgs Search at CMS

In 2011, CMS published Higgs to two W boson searches using both cut-based and BDT-based selections [8]. The BDT-based performance was clearly better: the

² This is the data I showed in the lecture. Since then both LHCb and CMS have seen this decay and the combined significance is above threshold to claim observation [7]. Both the LHCb and CMS results are consistent with the SM prediction.

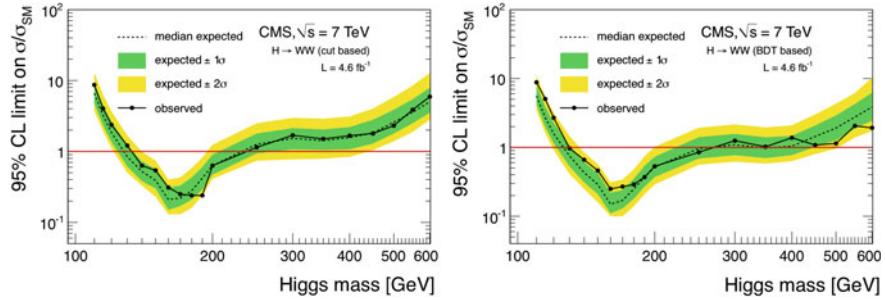


Fig. 8 $H \rightarrow WW$ results published by CMS in 2011 [8] for (left) cuts-based and (right) BDT-based selections. The excluded regions are (132–238) GeV by the cuts and (129–270) GeV by the BDT

cuts-based exclusion region was (132–238) GeV, while the BDT excluded (129–270) GeV. Figure 8 shows the famous *Brazil* plots for each selection.

1.5.3 Triggering

The LHCb collaboration also uses a BDT in its high level trigger (HLT) [9]. There are three major concerns that need to be addressed prior to using a BDT in an HLT:

- If the leaves are small relative to the resolution or data-taking stability, the signal could oscillate in and out of accepted regions. This would result in a lower efficiency and also one that is very difficult to understand. Furthermore, the background could also oscillate resulting in a wildly varying output rate.
- Often times the signal PDFs are not known at the time of training since the experiment itself is supposed to be the first to observe them. A related issue is that many triggers are meant to be *inclusive* and select types of signal rather than a single process.
- All HLT algorithms must be extremely fast as they run while data is being recorded.

There is a single simple solution to all three of these issues and it is detailed in [10]. This BDT algorithm was run at LHCb in 2011 and 2012 and performed well beyond the expected design parameters of the experiment.

1.6 Discussion

Machine learning methods are powerful but should not be treated as black boxes. It is advisable to assume that you have overtrained and then convince yourself, your referees, etc., otherwise. Most software packages provide some form of variable ranking that informs the analyst of how important each variable was in the construction

of the classifier. Ask yourself “Does this make sense?” Be sure to boost and to study the effect on performance of any input parameter values that the algorithm requires.

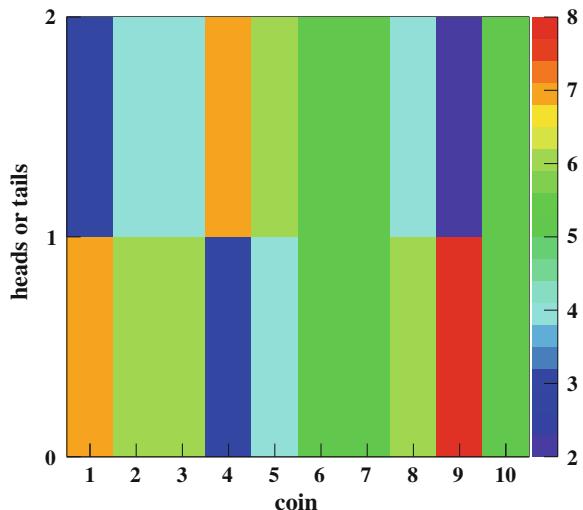
2 Regression and Two-Sample Testing

The second lecture focused on analyzing the data once it has been selected. In this section \mathbf{x} again refers to an arbitrary number of variables, $f(\mathbf{x})$ denotes the true PDF, $f_0(\mathbf{x})$ is a test PDF and T is any test statistic that quantifies (in some way) agreement between the data and f_0 . Here I chose to always use smaller T to mean better agreement but in general this does not have to be the case. This lecture follows closely [11].

2.1 The χ^2 Statistic

Consider a very simple example: I flip 10 coins 10 times each. Is my collection of 10 coins *fair*? Figure 9 shows the results of such a pseudo-experiment. We define $\chi^2 = \sum(o_i - e_i)^2/e_i$, where o_i is the observed number of counts in bin i and e_i is the expected number. For the example shown in Fig. 9, $\chi^2 = 8.4$ and the number of degrees of freedom is 10. Are the coins fair? We need a quantity that tells us how well our data agrees with our hypothesis in a well-defined way.

Fig. 9 Results obtained by flipping 10 coins 10 times each



2.2 *p*-values

If we denote the PDF of T as $g(T)$, which may depend on f_0 , then the *p*-value is defined as follows:

$$p = \int_T^\infty g_{f_0}(T')dT'.$$

The *p*-value is the probability of finding a T -value corresponding to lesser agreement than the observed T -value if $f = f_0$. It is not the probability that $f = f_0$!

If $f = f_0$, then the *p*-value distribution is uniform on $(0, 1)$. One can reject the hypothesis $f = f_0$ at confidence level α if $p < 1 - \alpha$; e.g., the test hypothesis is rejected at 95 % confidence level if $p < 0.05$. If $f = f_0$, $p < 0.001$ should happen in 1/1,000 experiments; rare but it should happen. The case $p > 0.999$ should also happen in 1/1,000 experiments. You should be suspicious of these too. *N.b.*, one of the reasons the χ^2 is so popular is that its g does not depend on f ; however, this is really only true in the limit $n \rightarrow \infty$. Everything you know about χ^2 is never strictly true on your data.

As a simple pedagogical example about *p*-values, assume that I give a physics exam to one million physics students and build g for T defined as the scores on the exam. If I give this exam to a student whose major is unknown, what can I conclude about the student's major from my exam? Figure 10 shows the *p*-value distributions for physics students, math students, art students and physics professors. By construction the *p*-value distribution of physics students is uniform. All values of p are equally likely for physics students. Math students do, on average, slightly worse on the physics exam than the physics students. The art students do poorly, while the physics professors do very well. Notice that a large *p*-value does not mean that the person taking the exam was a physics student. This test is a decent discriminator of

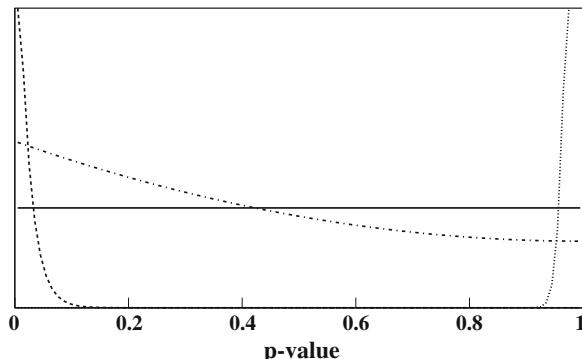


Fig. 10 *p*-value distributions for (solid) physics students, (dash-dotted) math students, (dashed) art students and (dotted) physics professors on a physics exam

physics and art students but not physics and math students. All of these observations hold in general for p -values and goodness-of-fit tests. A large p -value does not mean that your hypothesis is correct.

2.3 Two-Sample Tests in 1-D

Consider the situation where two data sets have been collected, call them A and B , and one simply wants to know if they share the same parent PDF. For example, there may be a (presumed) physical symmetry that relates A to B ; thus, observing that A and B do not share the same PDF would be evidence for new physics. A less exciting example is a systematic study comparing two measurements of the same process. How do we test this? I.e., how do we quantify the level of agreement between the two data sets given that we do not know the PDF?

One option is the χ^2 test discussed above. The normalized two-sample χ^2 statistic is given as follows:

$$\chi^2 = \sum_{i=1}^n \frac{(a_i n_B - b_i n_A)^2}{n_A n_B (a_i + b_i)}, \quad (1)$$

where a_i and b_i are the observed number of events in the i th bin in each of the data sets, n_A and n_B are the total sample sizes and n is the number of bins. The test statistic defined in 1 will approximately follow a χ^2 distribution with $n - 1$ degrees of freedom for any *reasonable* choice of binning scheme; i.e., given that the bins are chosen such that there are not too many low-occupancy bins. This is a situation where many analysts commonly use an improper formula for χ^2 . The above formula is the correct one for this situation. In principle, the χ^2 test works in any number of dimensions; however, in practice it tends to fall victim to the *curse of dimensionality* as the number of bins required in higher dimensions leads to low bin occupancies (the data becomes sparse).

Another well-known test that works in one dimension is the Kolmogorov-Smirnov (KS) test [12, 13]. The KS test defines the test statistic T as the maximum difference between the two empirical cumulative distribution functions (CDFs). In one dimension the KS test works very well. Unfortunately, it does not work in higher dimensions. The reason is that CDFs are integrals and in more than 1-D there is no unambiguous way to define the direction (or order) of integration. I close this subsection by noting that there are many more 1-D two-sample tests on the market: Cramer-von Mises [14], Anderson-Darling [15], etc. Which is the most powerful depends on the application, but all of them are very good tests and one should not hesitate to use them for one-dimensional problems.

2.4 Two-Sample Tests in D Dimensions

Before discussing higher-dimensional two-sample tests I need to introduce the concept of distance in a multivariate space. There are a number of valid ways to define distance. One popular choice of distance metric is the *normalized Euclidean* distance

$$|\mathbf{x}_i - \mathbf{x}_j|^2 = \sum_{v=1}^D \left(\frac{x_i^v - x_j^v}{w_v} \right)^2, \quad (2)$$

where w_v are weights for each variate. A common choice is to use the RMS for each variate. The choice is arbitrary, just like choosing a binning scheme for a histogram. The conclusions drawn from any test that depends on distance should not depend strongly on the choice of metric. *N.b.*, distance-based methods do not work well with mixed input types (discrete and continuous).

2.4.1 Permutation Test

It is often possible to define some statistic T that is a powerful discriminator between the null and alternative hypotheses; however, it is rarely the case that one can analytically (even asymptotically) convert T into a p -value. The permutation test is a numerical way of converting any statistic into a p -value [16, 17]. If $f_A = f_B$, then we can treat A and B as just labels. I.e., for each event recorded, there was an equal chance of it being in the A or B data set. So, any relabeling is just as likely as the measured labels.

example permutation test with very small data sets		
A_0, A_1, A_2	B_0, B_1, B_2	$\rightarrow T$
A_0, A_2, B_0	A_1, B_1, B_2	$\rightarrow T_0$
A_1, B_1, B_2	A_0, A_2, B_1	$\rightarrow T_1$
\vdots	\vdots	\vdots

The permutation test randomly assigns n_A datum the label A and the remaining n_B the label B , then recalculates T . This is done many times and the p -value is the fraction of these permutations for which $T < T_i$ is true. In this way we can obtain the p -value for any test statistic T . The permutation test is a very powerful and useful tool.

2.4.2 Mixed Sample Tests

If two data sets are combined to form a pooled sample, the mixing of the two samples is only optimal if they share the same parent distribution (see Fig. 11). The method described here [18, 19] does not require any knowledge concerning the PDFs of

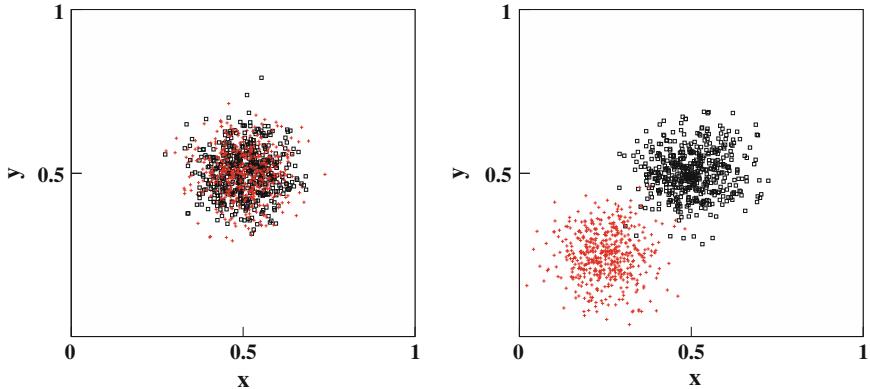


Fig. 11 Example distributions of data randomly sampled from the PDFs $f_A(\mathbf{x})$ (black open squares) and $f_B(\mathbf{x})$ (red crosses) for the cases: (left) $f_A(\mathbf{x}) = f_B(\mathbf{x})$; (right) $f_A(\mathbf{x}) \neq f_B(\mathbf{x})$. The two samples are optimally mixed if $f_A(\mathbf{x}) = f_B(\mathbf{x})$ but not so if $f_A(\mathbf{x}) \neq f_B(\mathbf{x})$. This fact is exploited by GOF tests in the mixed-sample category. Figure taken from [11]

either of the samples. The statistic that will be used to test the hypothesis $f_A = f_B$ is defined as follows:

$$T = \frac{1}{n_k(n_A + n_B)} \sum_{i=1}^{n_A+n_B} \sum_{k=1}^{n_k} I(i, k),$$

where $I(i, k) = 1$ if the i th event and its k th nearest neighbor belong to the same sample and $I(i, k) = 0$ otherwise, and n_k is the number of nearest-neighbor events being considered. The quantity T is then simply the mean fraction of like-sample nearest-neighbor events in the pooled sample of the two data sets. The expectation value of T is larger for the case $f_A \neq f_B$ due to the lack of complete mixing of the two samples that occurs if their parent distributions are not the same. For the extreme example shown in Fig. 11, one can see that the left panel has $T \approx 1/2$ ($n_A = n_B$) while the right panel has $T \approx 1$. A p -value can be obtained from T using an approximate limiting expression (see [11]) or using the permutation test.

2.4.3 Energy Test

The potential energy of a system of charged point particles interacting via potential $\psi(\Delta\mathbf{x})$ ($\Delta\mathbf{x} \equiv |\mathbf{x} - \mathbf{x}'|$) is given by [20, 21]:

$$\begin{aligned} T &= \frac{1}{2} \int \int (f_A(\mathbf{x}) - f_B(\mathbf{x})) (f_A(\mathbf{x}') - f_B(\mathbf{x}')) \psi(\Delta\mathbf{x}) d\mathbf{x} d\mathbf{x}' \\ &= \frac{1}{2} \int \int [f_A(\mathbf{x}) f_A(\mathbf{x}') + f_B(\mathbf{x}) f_B(\mathbf{x}') - 2 f_A(\mathbf{x}) f_B(\mathbf{x}')] \psi(\Delta\mathbf{x}) d\mathbf{x} d\mathbf{x}', \quad (3) \end{aligned}$$

which is minimal when $f_A = f_B$ (e.g., consider $\psi(x) = 1/x$ and recall your E&M courses; a better choice here is $\psi(x) = -\log(x + \varepsilon)$). Given that we want to test this same hypothesis, this quantity seems like it may be useful. There is one big problem: We don't know f_A or f_B ! Luckily, we do not need to know anything about either PDF.

T can be estimated without the need for any knowledge about the forms of f_A and f_B using the data as follows:

$$T \approx \frac{1}{n_A(n_A - 1)} \sum_{i,j>i}^{n_A} \psi(\Delta \mathbf{x}_{ij}) + \frac{1}{n_B(n_B - 1)} \sum_{i,j>i}^{n_B} \psi(\Delta \mathbf{x}_{ij}) - \frac{1}{n_A n_B} \sum_{i,j}^{n_A, n_B} \psi(\Delta \mathbf{x}_{ij}). \quad (4)$$

This is simply the previous equation rewritten using the standard Monte Carlo integration approximation ($\int \phi(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \approx \frac{1}{n} \sum \phi(\mathbf{x}_i)$), along with the fact that $\int f_A(\mathbf{x}) d\mathbf{x} = \int f_B(\mathbf{x}) d\mathbf{x} = 1$. The data is sampled from the PDFs, whatever they are, so we are done! The permutation test can be used to convert T into a p -value.

2.5 Real World Example: CP Violation

As a pedagogical example of applying these ideas to a physics analysis, consider searching for CP violation in a (fictitious) process $X \rightarrow abc$ and its charge conjugate decay [22]. If CP is conserved, the two decays share the same parent PDF. There are many processes in nature where the SM prediction is that no CP violation should be observable. In such situations determining that the data sets are not consistent with sharing the same PDF would be evidence for new physics. Figure 12 shows the Dalitz-plot distributions for $X \rightarrow abc$ and its charge conjugate decay. The model has a small CP -violating phase in one of the resonances (the 1^- resonance in the ac system). Visually the two Dalitz plots look indistinguishable. An ensemble of 100 data sets was generated from these PDFs. The χ^2 test was run on each and found to be *blind* to the presence of CP violation. The energy test, however, was able to detect the discrepancy. Figure 13 provides a visualization tool showing the location of the discrepancy in the two data sets identified by the energy test. This agrees with the known CP violation input into the model PDFs.

2.6 Goodness-of-Fit Tests

Regression involves determining the relationship between a set of dependent variables (parameters) and independent variables (data points). The name originates from one of its early uses in biology involving *regression to the mean*. In physics, regression is (almost) always either:

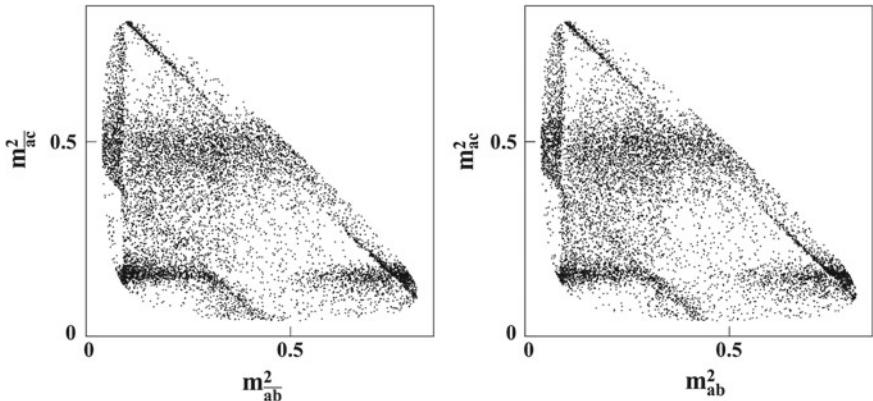
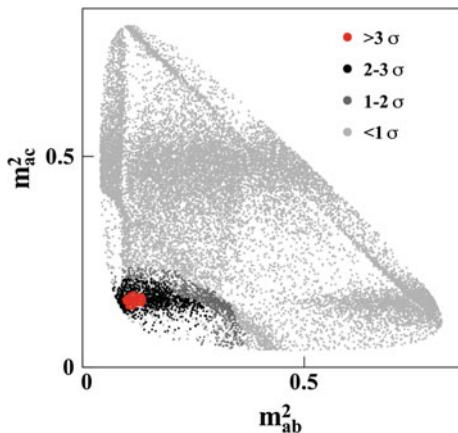


Fig. 12 Dalitz-plot distributions of the toy CP -violating model from [22]

Fig. 13 Dalitz-plot distributions of the toy CP -violating model from [22]. The colored bands show where CP violation has been found by the energy test



- χ^2 minimization: the data are binned (in some number of dimensions) and the quantity $\chi^2 = \sum(o_c - e_c)^2/e_c$, where $o_c (e_c)$ is the number of observed(expected) counts in bin c , is minimized by varying the parameters;
- MLE: the likelihood function $\mathcal{L} = \prod f_0(\mathbf{x}_i)$ is maximized (really $-\log \mathcal{L}$ is minimized) by varying the parameters (sometimes the *extended* likelihood is used if n_{observed} is relevant).

I assume the reader is familiar with these techniques and will not discuss them.

A *goodness-of-fit* (GOF) test describes how well a model describes a data set. GOF tests should produce p -value distributions that are uniform when testing the true PDF. The χ^2 test is the *classic* example; however, one expects that unbinned methods perform better in large D dimensions (especially with sparse data). What other choices are there? All of the two-sample tests discussed above can be used as GOF tests. This simply requires generating a large MC sample from the test PDF

(f_0) which is then used as one of the data samples (the data itself serves as the other sample). Of course, there will be some bias due to the fact that f_0 is determined from the data; however, provided the correlation between the fit statistic (e.g., χ^2 or $\log \mathcal{L}$) and T is small, this should not be an issue. I have studied the performance of the energy test on Dalitz-plot fits where f_0 was determined from an unbinned maximum likelihood fit and found the bias in the p -values to be at the few percent level (compatible with using the asymptotic χ^2 p -values, i.e., the bias can safely be ignored).

I will introduce one GOF test here that was not discussed in the two-sample test section: the K function [23, 24]:

$$K(r) \propto \sum_{i=1}^n \sum_{j \neq i} \frac{I(|\mathbf{x}_i - \mathbf{x}_j| < r)}{v(i, j) f_0(\mathbf{x}_i) f_0(\mathbf{x}_j)}, \quad (5)$$

where $v(i, j)$ is a volume edge-correction factor. Often times a related function $L(r) = K(r)^{1/D}$ is used instead of K . If $f_0 = f$, then $\langle L(r) \rangle = r$. To get a p -value, the test statistic $T = (L(r) - r)_{\max}$ is used (similar to KS test). The basic idea is to compare the local event density (out to radius r) around each event to the expected value. Figure 14 shows an example from [11]. I found that K (in a 2-D Dalitz analysis) is much more powerful than the χ^2 test and comparable to the energy test.

There are other GOF tests on the market, e.g., kernel-based tests; distance to nearest neighbor based tests; etc. There is no uniformly most powerful GOF test. For some analyses it pays to use a specialized test. This is fine but be sure that you know the test works. The p -values should be uniform when $f_0 = f$. If they are not then you have not constructed a GOF test. It is recommended to verify this using toy MC where the true PDF is known.

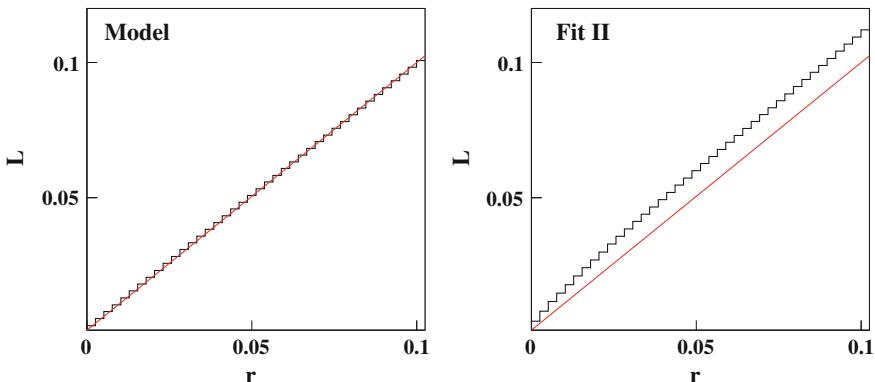


Fig. 14 Example use of the K function from [11]. The *left plot* is when $f = f_0$, while the *right* is for $f \neq f_0$. The red line shows the expected distribution under the hypothesis that $f = f_0$. The difference between the red line and the data is used to define T

2.7 Confidence Intervals and Significance

After obtaining an estimator \hat{x} for some parameter x , one quotes \hat{x}_{-l}^{+u} . If the experiment is repeated many times, 68 % of the $(\hat{x} - l, \hat{x} + u)$ intervals should contain the true value of x . Notice that 68 % is not the probability that the true value of x is in the interval (which is either 0 or 1 for any interval). How is a confidence interval defined? For example, consider the case where we obtain \hat{x} by maximizing the likelihood. In the limit $n \rightarrow \infty$, \mathcal{L} is a parabola and $\Delta \log \mathcal{L} = \pm 1/2$ can be used to get the 68 % confidence interval. For finite n , this result is not strictly true. It may or may not provide a good estimate for the confidence interval. A useful rule of thumb: Asymptotic statistical results tend to converge surprisingly quickly. The places where they fail are when one observes something that is forbidden in the $n \rightarrow \infty$ limit (e.g., less events than the background-only expectation).

The statistical significance is the probability that a result as extreme as the observed data would occur by chance in the absence of *signal*. In physics it is quoted in “units” of σ ; i.e., how many σ away from the mean of a Gaussian one would need to be to obtain the same p -value (one needs to be careful here about whether a one-sided or two-sided test is being performed). For *nested* models, $-2\Delta \log \mathcal{L}$ is asymptotically χ^2 distributed with $n_{\text{dof}} = \Delta n_{\text{parameters}}$ (this is known as Wilkes’ Theorem). There are several important points about this theorem that are often overlooked by analysts:

- the models must be nested (e.g., a signal Gaussian with mean and/or width free can not be nested with a no signal model!);
- converting this to $n\sigma$ using $\sqrt{-2\Delta \log \mathcal{L}}$ is only (possibly) valid if $\Delta n_{\text{par}} = 1$;
- this is asymptotically correct only.

Finally, one must always be aware of the *look elsewhere effect* (a.k.a. trials factor). Recall that p -values are uniformly distributed if the null hypothesis is true; thus, if a large number of tests are performed on a data set for which the null is true, then a small p -value is likely to be obtained in one of the tests. This is often mistakenly interpreted as evidence that the null hypothesis is false. This is rather well demonstrated in an amusing way by this XKCD comic: xkcd.com/882.

2.8 Discussion

It is important to always remember that a p -value is not the probability that the null hypothesis is true. If the null is true, the p -value distribution obtained from an ensemble of data sets should be uniform. The KS test works well in one dimension but it does not work in higher dimensions. The χ^2 test works well in one dimension but its power in higher dimensions tends to be limited by the curse of dimensionality. Mixed-sample tests work in higher dimensions, but my experience with these is that they are only moderately more powerful than the χ^2 test. The energy test is the most

powerful two-sample test that I have ever used. It is costly in terms of CPU power but, in my experience, it is worth the expense. The permutation test can be used to obtain the p -value of any test statistic T when performing a two-sample test.

3 Bayesian Versus Frequentist

It's now time to confront the epic division in statistics: Frequentists versus Bayesians. I will frame this debate rather informally as I did in the actual lecture. The language is rather imprecise but this tends to be more useful for non-experts. Frequentists work with objective probability. They refer to the fraction of data sets from an ensemble that possess some property. Frequentists can only discuss things for which there can be an ensemble, but for these things they are in total agreement (in theory). Bayesians work with subjective probability. Given some prior beliefs, they use observations (data) to update their beliefs. Bayesians can discuss anything; however, they leave room for disagreement (your prior belief may not agree with mine).

Bayes' Theorem states that $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$. Consider this simple example: My friend belongs to a cult that thinks the world will end on Feb. 1, 2013 (the day after this lecture was given). I am skeptical. I think there's a 10% chance that he is right. One of the predicted signs of the coming doom is that it will rain in Santiago de Compostela on Jan. 31. I check historical data and find that the probability of rain on this date is 50% (in the absence of doom). Applying Bayes' Theorem here gives

$$P(\text{doom}|\text{rain}) = \frac{P(\text{rain}|\text{doom})P(\text{doom})}{P(\text{rain})} = \frac{1 \times 0.1}{1 \times 0.1 + 0.5 \times 0.9} = 0.18. \quad (6)$$

I.e., if it rains today I will think there's an 18% that his cult is right. If it does not rain, then I believe there is no chance his cult is right since his cult assigned 100% probability to it raining today.

Now consider a Bayesian and Frequentist locked in a room with no windows.³ They have a neutrino detector that can be used to determine whether or not the sun has exploded. Inside this detector is a random number generator. When asked whether the sun has exploded a random number is generated between 1 and 3.5 million. If the number 66 is obtained, the detector lies about the state of the sun, otherwise it tells the truth. Our pair of physicists locked in the windowless room ask the machine if the sun has exploded and it replies "yes". Our Frequentist notes that $p < 5\sigma$ and rejects the "sun has not exploded" hypothesis, but accepts that there is a one in 3.5M chance that they are wrong. Our Bayesian notes that Bayes' theorem says $P(\text{explode|yes}) = \frac{P(\text{yes}|\text{explode})P(\text{explode})}{P(\text{yes})}$. The Bayesian's prior belief for the probability of the sun exploding during the required time window is ε (some very small value), while the probability of the machine lying is ℓ ; thus, $P(\text{explode|yes}) = (1 - \ell)\varepsilon / [(1 - \ell)\varepsilon + \ell(1 - \varepsilon)] \approx \varepsilon / (\varepsilon + \ell)$. If $\varepsilon \ll \ell$, then $P(\text{explode|yes}) \approx \varepsilon / \ell$;

³ This example is a slightly modified version of xkcd.com/1132.

i.e., the Bayesian's belief that the sun has exploded will be $3.5M$ times bigger than before consulting the untrustworthy machine, but still very very small.

In physics, if $p < 5\sigma$, we claim “discovery”; however, nothing special actually happens at 5σ . In principle, frequentism is objective, but the claims made by physicists based on frequentist results are subjective. In my experience, most physicists claim to be frequentists but actually hold Bayesian views on science. To truly avoid Bayesism completely a scientist must take the following view: *Scientific theories say nothing about truth. Scientific theories are simply useful tools for assigning probabilities to future events.* Most physicists, in my experience, do not take this view of science. Most physicists are really (implicit) Bayesians.

3.1 Limits

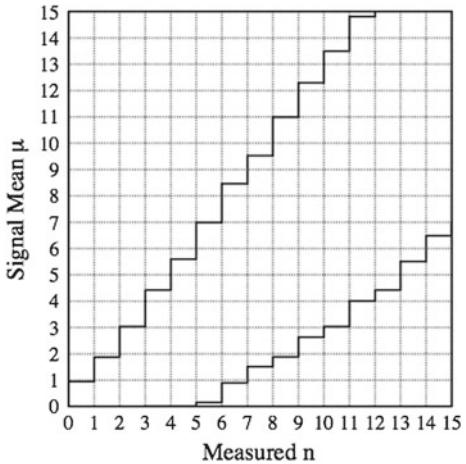
Setting limits is important in physics since, unfortunately, most of the things we look for we don't find (probably they don't exist). Below are the confidence intervals for a Poisson process:

n observed	low: $\sum P(n, \lambda) = 0.05$	high: $\sum P(n, \lambda) = 0.95$
0	—	3.00
1	0.05	4.74
2	0.36	6.30
3	0.82	7.75
:	:	:

Unfortunately, it is rarely (ever?) as easy as reading off this chart to set limits. For example, if the expected number of background events is 3.0 and the observed number of events is $n = 0$, then at the 95 % CL $\lambda = s + b < 3.0$. The upper limit on the signal obtained is $s < 0$. On one had, we know that as many as 5 % of our claims at 95 % CL should be wrong. We could accept that this is one of them and publish the result (to avoid bias). On the other hand, we also know that $s \geq 0$. Can't we use this knowledge somehow?

I will now give a quick overview of the procedure of Feldman and Cousins [25]. First, it is important to realize that prior to looking at the data, one must decide whether to quote a CI or a limit. If this decision is made after viewing the data, it will bias the results. The method of Feldman and Cousins starts by considering each value of s and constructing a confidence *belt* that contains the desired probability. This is done by *ranking* on $P(n|s) = (s + b)^n \exp(-(s + b))/n!$, where $s = n - b$ or zero. For each s , one starts with the highest $P(n|s)$ and adds to the interval until reaching the desired CL. This is done for all s . For any measured n , the CI is read *vertically*, (e.g., for $n = 0$ the interval is $[0, 1.1]$, while for $n = 6$ its $(0.2, 8.5)$) (see Fig. 15). This procedure guarantees correct coverage and automatically returns either a limit or CI.

Fig. 15 Confidence belts from the Feldman-Cousins [25] method for 90 % CL and $b = 3.0$



To calculate a Bayesian limit we need to first determine the *posterior* distribution for the signal

$$p(s|n) \propto \int \int P(n|s, b, \varepsilon) \pi(s) \pi(b) \pi(\varepsilon) db d\varepsilon, \quad (7)$$

where $\pi(x)$ is the prior for x and $P(n|s, b, \varepsilon) = \exp(-(\varepsilon s + b))(\varepsilon s + b)^n / n!$ (ε is the efficiency/normalization). The limit @ 90 % *credibility level* is

$$\int_0^{\mathcal{B}_{90}} p(\mathcal{B}|n_{\text{obs}}) d\mathcal{B} = 0.9 \int_0^{\infty} p(\mathcal{B}|n_{\text{obs}}) d\mathcal{B}. \quad (8)$$

This procedure naturally produces a one or two-sided interval; however, the credibility is based on priors used (assumed). A common choice for the signal prior is a uniform distribution on $[0, \infty)$ as this is meant to represent *total ignorance* about the signal; however, if one works with x^2 instead of x , a different result is obtained (something uniform in x is not uniform in x^2). This means that the *total ignorance* declaration does not survive a change of variables. It is generally considered good practice to try several priors (e.g., uniform, Jeffreys) to see how much the result changes. If the difference large, then the result is strongly reflecting the prior belief about x . The more data that is available the less dependent the result will be on the prior.

3.2 Discussion

Bayesian and Frequentist philosophies are very different; however, with enough data the data *speaks for itself* and (roughly) the same result will be obtained using either

approach. It is very important to choose a strategy prior to looking at the data. There are many other limit-setting options on the market that I did not discuss (e.g., CLs, profile likelihood).

4 Summary

These lectures were a *tour de force* of numerical methods. One parting piece of advice I would like to give is *know the assumptions built into any method that you use*. It is important to recognize the situations that do not satisfy those assumptions.

References

1. L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, *Classification and Regression Trees* (Wadsworth International Group, Belmont, California, 1984)
2. B.P. Roe et al, Boosted decision trees as an alternative to artificial neural networks for particle identification. Nucl. Instr. Meth. **A543**, 577 (2005). [arXiv:physics/0408124](https://arxiv.org/abs/physics/0408124)
3. B. Efron, R.J. Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall, London, 1993)
4. L. Breiman, Bagging predictors. Mach. Learn. **24**, 123 (1996)
5. L. Breiman, Random forests. Mach. Learn. **45**, 5 (2001)
6. R. Aaij et al., [LHCb Collaboration], *First evidence for the decay $B_s \rightarrow \mu^+ \mu^-$.* Phys. Rev. Lett. **110**, 021801 (2013)
7. CMS and LHCb Collaborations, *Combination of results on the rare decays $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ from the CMS and LHCb experiments.* CMS-PAS-BPH-13-007 and LHCb-CONF-2013-012
8. CMS Collaboration, Search for the standard model Higgs boson decaying to $W^+ W^-$ in the fully leptonic final state in pp collisions at $\sqrt{s} = 7$ TeV, Phys. Lett. B **710**, 91113 (2012). <https://twiki.cern.ch/twiki/bin/view/CMSPublic/Hig11024TWiki>
9. R. Aaij et al., [LHCb Trigger Group], The LHCb trigger and its performance. JINST (2013). [arxiv:1211.3055](https://arxiv.org/abs/1211.3055)
10. V. Gligorov, M. Williams, Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree. JINST **8**, P02013 (2013). [arXiv:1210.6861](https://arxiv.org/abs/1210.6861)
11. M. Williams, JINST **5**, P09004 (2010)
12. A. Kolmogorov, G. Ist. Ital. Attuari **4**, 83 (1933)
13. N. Smirnov, Ann. Math. Stat. **19**, 279 (1948)
14. T.W. Anderson, On the distribution of the two-sample Cramervon Mises criterion. Ann. Math. Stat. (Institute of Mathematical Statistics) **33**, 1148 (1962)
15. T.W. Anderson, D.A. Darling, Asymptotic theory of certain “goodness-of-fit” criteria based on stochastic processes. Ann. Math. Stat. **23**, 193 (1952)
16. R.A. Fisher, *The Design of Experiments* (Oliver and Boyd Ltd., London, 1935)
17. P. Good, *Permutation Tests: a Practical Guide to Resampling Methods for Testing Hypotheses* (Springer-Verlag, New York, 1994)
18. M.F. Schilling, Multivariate two-sample tests based on nearest neighbors. J. Am. Stat. Assoc. **81**(395), 799–806 (1986)
19. N. Henze, A multivariate two-sample test based on the number of nearest neighbor type coincidences. Ann. Stat. **16**(2), 772–783 (1988)
20. B. Aslan, G. Zech, New test for the multivariate two-sample problem based on the concept of minimum energy. Stat. Comp. Simul. **75**(2), 109–119 (2004)

21. B. Aslan, G. Zech, Statistical energy as a tool for binning-free, multivariate goodness-of-fit tests, two-sample comparison and unfolding. *Nucl. Instr. Meth.* **A537**, 626–636 (2005)
22. M. Williams, Observing CP violation in many-body decays. *Phys. Rev. D* **84**, 054015 (2011)
23. B.D. Ripley, Modelling spatial patterns. *J. Roy. Stat. Soc. B Met.* **39**(2), 172–212 (1977)
24. A.J. Baddeley, J. Møller, R. Waagepetersen, Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Stat. Neerl.* **54**(3), 329–350 (2000)
25. G. Feldman, R.D. Cousins, A unified approach to the classical statistical analysis of small signals. *Phys. Rev. D* **57**, 3873 (1998)

Theoretical Cosmology

Antonio L. Maroto

Abstract The aim of these lectures is to provide an introduction to some fundamental aspects of cosmology from a theoretical point of view. The selected topics include a brief review of the observational basis of cosmology, from the distribution of matter on large scales to the Hubble expansion and acceleration, together with a description of the main properties of the cosmic microwave background. In this first lecture a summary of the standard Friedmann–Lemaître–Robertson–Walker models including their kinematics and dynamics is also presented, together with a discussion on distance indicators. The second lecture is devoted to thermodynamics, including both equilibrium thermodynamics and the important phenomenon of decoupling in an expanding universe which will allow us to present explicit calculations of the relic abundances of different particle species. The third lecture is a brief introduction to inflation, where we discuss the fundamental problems of standard cosmology and how inflation solves them. We present also basic results on the slow-roll inflation models based on scalar inflaton fields. The last section is an introduction to the theory of cosmological perturbations which focuses on the evolution of perturbations and their generation from quantum fluctuations during inflation.

1 Cosmological Observations and the Friedmann–Lemaître–Robertson–Walker Models

1.1 Cosmological Observations

Cosmology studies the origin, evolution and structure of the universe on very large scales (much larger than the size of a galaxy) [1–4]. A typical galaxy is a gravitationally bound object containing around 10^{11} stars. The mass of a typical star like the sun is $M_{\odot} = 2 \times 10^{30}$ kg and its luminosity $L_{\odot} = 3.84 \times 10^{26}$ J s⁻¹. The mass

A.L. Maroto (✉)
Departamento de Física Teórica, Universidad Complutense de Madrid,
28040 Madrid, Spain
e-mail: maroto@ucm.es

of a galaxy is thus of the order of $M_{gal} \simeq 10^{41}$ kg, and its size (radius) is around $R_{gal} \simeq 10^{20}$ m $\simeq 10$ kpc¹

In systems of this size the only relevant interaction is gravitation.² For this reason, Einstein's General Relativity (GR) plays a fundamental role in Cosmology. However before studying GR in cosmological contexts, we will briefly review what observations tell us about the universe on large scales.

1.1.1 Matter Distribution on Large Scales

Galaxies are not uniformly distributed in space, rather they form structures of different sizes:

- Groups: typically containing 5–100 galaxies, with a size of 200kpc–2Mpc and typical velocity dispersion of $v_{gal} \sim 200$ km/s. The Milky Way is one of the largest galaxies of the Local Group.
- Clusters: contain 100–1,000 galaxies, with a few Mpc size, and velocity dispersions of $v_{gal} \sim 1,000$ km/s. The Virgo cluster is the nearest cluster to the Local Group.
- Superclusters: formed out of ~ 100 groups and clusters, with typical sizes around 25 Mpc. These objects are still evolving non-virialized structures. Our Local Group is contained in the Virgo supercluster.
- Filaments, Voids and Walls: on larger scales, a cell like structure emerges in which filaments of galaxies connect superclusters with huge underdense regions (voids) in between. The typical size of these superstructures could reach ~ 100 Mpc.

Therefore we see that on small scales $\lesssim 100$ Mpc the matter distribution is very inhomogeneous, but on large scales $\gtrsim 100$ Mpc, the distribution is statistically homogeneous as has been revealed by large galaxy surveys such as 2dF or SDSS [5] which have completed large 3D maps of galaxy distribution up to distances which could reach 600 Mpc³. Fig. 1 shows how the amplitude of the density fluctuations decreases with the averaging scale.

1.1.2 Matter Density of the Universe

Galaxy clusters are the largest stable and bound (virialized) objects in the universe. Using virial theorem it is possible to determine the total mass of a cluster by measuring the average velocity dispersion of the galaxies contained in it. Fritz Zwicky (1933) was the first in using this method to estimate the mass of the Coma cluster, finding that total mass to luminosity ratio $M/L \sim 200$ (in units $M_\odot/L_\odot = 1$). Similar

¹ 1 pc = 3.26 light year = 3.086×10^{16} m.

² It is usually assumed that there are no charged structures with this size although the origin of galactic and extragalactic magnetic fields is still unknown.

³ For many years homogeneity on large scales has been one the main fundamental assumptions of Cosmology known as the Cosmological Principle.

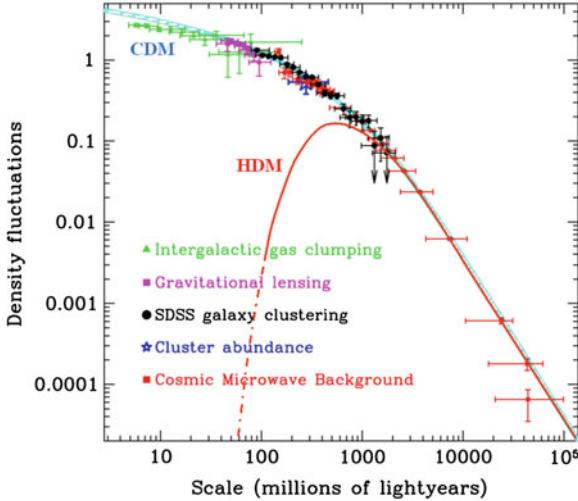


Fig. 1 Scale dependence of density fluctuations. The *blue curve* corresponds to the generic prediction of cold *dark matter* (CDM) models, whereas the *red curve* corresponds to hot *dark matter* (HDM) ones which show a suppression on small scales because of the free-streaming damping

results have been obtained from observation of X-ray emission of clusters, produced by hot intracluster gas at $T \sim 10^8$ K. This emission requires the total mass of the cluster to be much larger than its luminous mass in order to keep the hot gas bound in the cluster, i.e. most of mass in the clusters is in the form of dark matter. Given the large size of clusters, it is possible to extrapolate these results to the whole universe. Thus we can estimate the average matter density ρ_M as:

$$\Omega_M = \frac{\rho_M}{\rho_c} = \frac{\langle \mathcal{L} \rangle M}{\rho_c L} \quad (1)$$

where $\langle \mathcal{L} \rangle = (2.0 \pm 0.2) \times 10^8 h L_\odot \text{Mpc}^{-3}$ is the average luminosity density of the universe. ρ_c is the critical density, defined for a given expansion rate $H_0 = 100 h \text{ km/s/Mpc}$, as $\rho_c = 3H_0^2/(8\pi G)$. As we will see, the critical density corresponds to the matter density that a universe which is expanding at a rate H_0 should have in order for its spatial sections to be flat. Its value reads:

$$\rho_c = 1.88 h^2 \times 10^{-29} \text{ g cm}^{-3}. \quad (2)$$

Using these definitions, it is possible to obtain [2]:

$$\Omega_{M, \text{cluster}} = 0.15 \pm 0.02 \pm 0.04 (\text{sys}). \quad (3)$$

Additional evidence for the existence of dark matter comes from the rotation curves of elliptical galaxies, i.e. the rotation velocity of objects orbiting away from

the galactic center as a function of radius $v_{rot}(r)$. It is found that the velocity grows linearly from the center until it reaches a maximum and remains constant up to the maximum measured radius. However, if the galactic mass is concentrated within a few disk radii r_D , then for $r \gg r_D$, the mass distribution can be considered as pointlike and the orbits should be Keplerian, i.e.:

$$\frac{mv_{rot}^2}{r} = \frac{GM_D m}{r^2} \Rightarrow v_{rot}(r) = \left(\frac{GM_D}{r} \right)^{1/2} \propto r^{-1/2}, \quad r \gg r_D \quad (4)$$

in contradiction with observations. In order to have $v_{rot}(r) = v_m = \text{const.}$ at large radii, it is required that $M(r)$, the mass contained within a sphere of radius r , grows as r , i.e. $M(r) = v_m^2 r / G$, which means that if matter at large radii would be isotropically distributed, its density should follow the profile $\rho(r) = v_m^2 / (4\pi Gr^2)$. More precise dark matter profiles have been obtained by fitting different N-body simulations with observations, yielding:

- Navarro-Frenk-White (NFW) profile:

$$\rho(r) = \frac{\rho_0}{\frac{r}{R_s} \left(1 + \frac{r}{R_s} \right)^2} \quad (5)$$

- Einasto profile:

$$\rho(r) = \rho_0 \exp \left(\frac{-r}{R_s} \right)^\alpha \quad (6)$$

with R_s the dark matter halo radius scale and $\alpha \sim 0.1 - 1$.

1.1.3 The Expansion of the Universe

In 1915, Vesto Slipher discovered that except for the galaxies in the Local Group, all the other spiral galaxies which he was able to observe recede from us, as derived from the measurements of the redshift z in their absorption lines:

$$\frac{\lambda_{obs}}{\lambda_{em}} = 1 + z \quad (7)$$

Later on, in 1927–1929 Lemaître and Hubble proposed a linear relation between recession velocity and distance, in such a way that the more distant the galaxy is, the faster it moves away from us:

$$v \simeq zc \simeq H_0 L \quad (8)$$

where the proportionality constant H_0 is known as the Hubble constant. This relation is satisfied only for $z \ll 1$. The value measured by the Hubble space telescope (2011) is [6]:

$$H_0 = 73.8 \pm 2.4 \text{ km/s/Mpc} \quad (9)$$

As we will show, the Hubble law is interpreted in the context of GR as due to the expansion of the space and therefore H_0 can be understood as the rate of expansion of the universe. As we move backwards in time, the physical distance between galaxies would be smaller and smaller, eventually reaching a point of infinite density and temperature usually referred to as the Big Bang. In the standard cosmological model, the time elapsed from the Big Bang (the age of the universe) is given by [9] (Planck 2013):

$$t_0 = 13.813 \pm 0.058 \text{ Gyr} \quad (10)$$

so that $t_0 \simeq H_0^{-1}$.

Extending the Hubble diagrams to higher redshifts using supernovae type Ia as distance indicators, it has been possible [11, 12] to measure not only the rate of expansion, but also the acceleration rate. The unexpected result shows that contrary to what we would expect for a gravitational attractive interaction, the rate of expansion is accelerating, which as we will see has led to propose the existence of dark energy.

1.1.4 Abundances of Light Elements

It was proposed by G. Gamow in 1946 that if the universe in the past was in a stage of high density and temperature, the conditions could be appropriate for the formation of heavier nuclei from the fusion of protons and neutrons in the process of primordial nucleosynthesis. Elements heavier than ^7Li are produced by nuclear processes in stars, however H, D, ^3He , ^4He and ^7Li have a primordial origin and were produced when the universe was $t_{nuc} \sim 1 - 200$ s old with $T_{nuc} \sim 1 - 0.1$ MeV. The observed relative abundances:

$$\begin{aligned} {}^4\text{He}/H &\sim 0.25, \\ D/H &\sim 3 \times 10^{-5}, \\ {}^3\text{He}/H &\sim 2 \times 10^{-5}, \\ {}^7\text{Li}/H &\sim 2 \times 10^{-10} \end{aligned} \quad (11)$$

can be compared to the theoretically predicted ones which can be computed using the nuclear reaction rates measured in laboratory, the neutron decay rate, the neutrino number density, etc., in terms of a single free parameter which is the ratio between the baryon and photon number densities $\eta = n_B/n_\gamma$. Observations are consistent with theoretical predictions in the range:

$$\eta = (4.6 - 5.9) \times 10^{-10} \quad (12)$$

This extremely small ratio means that there is one baryon (proton or neutron) per 10^9 photons in the universe.

Using the photon density derived from cosmic microwave background (CMB) measurements $n_\gamma = 410 \text{ cm}^{-3}$ and the nucleon mass $m_N = 939 \text{ MeV}/c^2$, it is possible to obtain the baryon energy density:

$$\Omega_B h^2 = \frac{\rho_B h^2}{\rho_c} = \frac{\eta n_\gamma m_N h^2}{\rho_c} = 0.02207 \pm 0.00033 \quad (13)$$

(quoted value corresponds to Planck, 2013 results [9]). This value can be compared to the luminous matter density $\Omega_{lum} \simeq 0.0035 - 0.007$ so that:

$$\Omega_B \gg \Omega_{lum} \quad (14)$$

i.e. most of the baryonic matter is not luminous. It is believed that most of the baryonic dark matter should be in the hot highly ionized halo gas around galaxies.

Comparing with the matter density in clusters, or in the total matter density derived for CMB observations, we get:

$$\Omega_M \gg \Omega_B \gg \Omega_{lum} \quad (15)$$

i.e. most of the matter in the universe is dark and non-baryonic. None of the known particles in the Standard Model (SM) seems to be a viable candidate to explain the nature of this non-baryonic dark matter.

1.1.5 Cosmic Microwave Background

At sufficiently early times, the temperature was high enough for the electrons not to be bound into atoms. Collisions of free electrons with photons were so frequent as to keep photons in thermal equilibrium according to the Planck formula:

$$n_T(\omega)d\omega = \frac{1}{\pi^2} \frac{\omega^2}{e^{(\frac{\omega}{T})} - 1} d\omega \quad (16)$$

As the universe expands, matter becomes less abundant and colder, electrons recombine with nuclei and radiation decouples from matter, but keeping the same thermal spectrum with a temperature which decreases as the universe expands. G. Gamow realized that if around t_{nuc} , the temperature was the required one for nucleosynthesis to take place, then today the radiation temperature should be around $T_0 \sim 5 \text{ K}$. In 1965 A. Penzias and R. Wilson discovered this CMB radiation. In 1992 the FIRAS instrument of the COBE satellite confirmed with high accuracy the thermal spectrum at a temperature:

$$T_0 = 2.725 \pm 0.002 \text{ K} \quad (17)$$

The corresponding energy density is:

$$\rho_\gamma(T_0) = \frac{1}{\pi^2} \int_0^\infty \frac{\omega^3 d\omega}{e^{\omega/T_0} - 1} = \frac{\pi^2}{15} T_0^4 = 4.64 \times 10^{-34} \text{ g cm}^{-3} \quad (18)$$

and accordingly:

$$\Omega_\gamma = \frac{\rho_\gamma}{\rho_c} = 2.47 \times 10^{-5} h^{-2} \quad (19)$$

i.e. today the CMB radiation contribution to the energy density is negligible. The photon number density is given by:

$$n_\gamma(T_0) = \frac{1}{\pi^2} \int_0^\infty \frac{\omega^2 d\omega}{e^{\omega/T_0} - 1} = \frac{2\zeta(3)}{\pi^2} T_0^3 = 410 \text{ cm}^{-3} \quad (20)$$

where $\zeta(3) = 1.202057 \dots$

The COBE DMR instrument showed that the CMB radiation is extremely isotropic. The leading anisotropic contribution is a dipole with an amplitude $\delta T_1 = 3.372 \pm 0.007 \text{ mK}$, i.e.:

$$\frac{\delta T_1}{T_0} = 1.23 \times 10^{-3} \quad (21)$$

which is usually interpreted as due to the motion of the Solar System with respect to the CMB rest frame: $T(\theta) = T_0(1 + \frac{v_\odot}{c} \cos \theta)$, where $v_\odot = v_\odot \mathbf{u}$ is the Solar System velocity with respect to the CMB and $\cos \theta = \mathbf{u} \cdot \mathbf{n}$ with \mathbf{n} a unit vector in the observation direction. This has allowed to determine that the Solar System velocity is $v_\odot = 371 \pm 0.5 \text{ km/s}$ in the direction of the Virgo cluster.

Subtracting the dipole contribution, still some small high multipole anisotropies remain with a quadrupole amplitude given by $\delta T_2 = 30 \pm 5 \mu\text{K}$, i.e.:

$$\frac{\delta T_2}{T_0} = 1.1 \times 10^{-5} \quad (22)$$

These anisotropies play a fundamental role in order to understand the origin of structures in the universe [7] and their measurements by satellites such as WMAP (NASA 2003) [8] or Planck (ESA 2013) [9] has allowed to determine cosmological parameters with unprecedented precision (see Fig. 2).

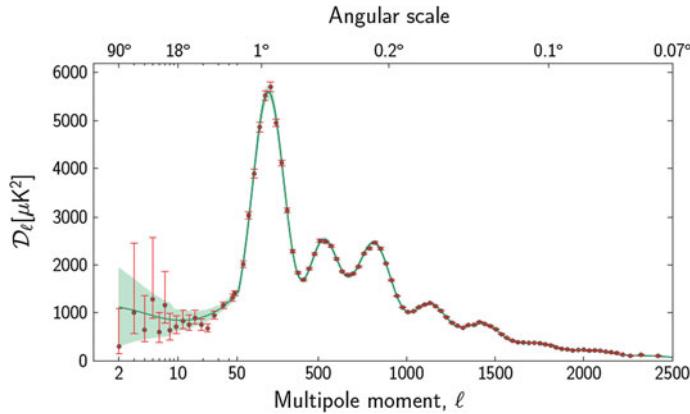


Fig. 2 Temperature power spectrum measured by Planck satellite [10]. The shaded region corresponds to the uncertainty at low multipoles due to cosmic variance. *Full line* corresponds to the standard Λ CDM prediction

1.2 Friedmann–Lemaître–Robertson–Walker Models

1.2.1 Robertson–Walker Metric

Homogeneity and isotropy of matter and CMB radiation suggest that on large scales and for comoving observers (those at rest with the CMB), the geometry of the universe should not depend on the position or the spatial direction. Using the coordinates corresponding to those observers, the most general form of the metric with homogeneous and isotropic spatial sections is the Robertson–Walker (RW) metric, given by:

$$ds^2 = dt^2 - a^2(t) \left(\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right) \quad (23)$$

where $a(t)$ is the scale factor and k determines the curvature of the spatial sections:

$${}^3R = \frac{6k}{a^2} \quad (24)$$

Rewriting the metric as $ds^2 = dt^2 - a^2(t)(h_{ij}dx^i dx^j)$, we see that the scale factor $a(t)$ determines the global scale of cosmological distances. Indeed, if at time t two comoving objects are located at $\{x^i\}$, $\{x^i + dx^i\}$, their coordinate or comoving distance will be given by:

$$dL = \sqrt{h_{ij}dx^i dx^j} \quad (25)$$

However, the physical distance, obtained from the RW line element will be:

$$dD = \sqrt{-g_{ij}dx^i dx^j} = a(t)dL \quad (26)$$

Thus, physical distances are related to comoving ones by:

$$D(t) = a(t)L \quad (27)$$

i.e. the physical distance between two comoving objects (with constant coordinates) grows with the scale factor.

Let us introduce a new radial coordinate χ given by:

$$\chi = \int \frac{dr}{\sqrt{1 - kr^2}} = \begin{cases} \frac{1}{|k|^{1/2}} \arcsin(r|k|^{1/2}), & k > 0 \\ r, & k = 0 \\ \frac{1}{|k|^{1/2}} \operatorname{argsinh}(r|k|^{1/2}), & k < 0 \end{cases} \quad (28)$$

which is nothing but the comoving radial distance. We will also introduce a new time coordinate η (conformal time), defined as:

$$dt = a(\eta)d\eta \quad (29)$$

In these coordinates the RW metric reads:

$$ds^2 = a^2(\eta)(d\eta^2 - d\chi^2 - f^2(\chi)(d\theta^2 + \sin^2 \theta d\phi^2)) \quad (30)$$

where:

$$f(\chi) = \frac{1}{|k|^{1/2}} \operatorname{senn}(\chi|k|^{1/2}) = \begin{cases} \frac{1}{|k|^{1/2}} \sin(\chi|k|^{1/2}), & k > 0 \\ \chi, & k = 0 \\ \frac{1}{|k|^{1/2}} \sinh(\chi|k|^{1/2}), & k < 0 \end{cases} \quad (31)$$

Let us consider a radial signal emitted from $\chi = \chi_S$ at $\eta = \eta_S$ which is received at $\chi = 0$ at $\eta = \eta_0$. A second signal is emitted from the same source at $\chi = \chi_S$ at $\eta = \eta_S + d\eta_S$ and received at the same place $\chi = 0$ at $\eta = \eta_0 + d\eta_0$. The signals follow radial null geodesics (see Fig. 3), $ds^2 = 0$, i.e.:

$$0 = a^2(\eta) \left(d\eta^2 - d\chi^2 \right) \quad (32)$$

so that for the incoming geodesics we have $d\eta = -d\chi$. After integration along the geodesic, we get:

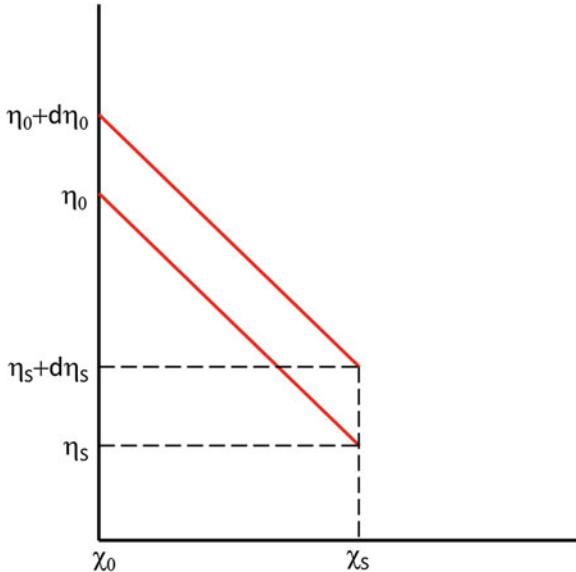


Fig. 3 Radial null geodesics in RW geometry in (χ, η) coordinates

$$\begin{aligned} \eta_0 - \eta_s &= \chi_s \\ (\eta_0 + d\eta_0) - (\eta_s + d\eta_s) &= \chi_s \end{aligned} \quad (33)$$

so that $d\eta_0 = d\eta_s$. Thus in physical time:

$$\frac{dt_0}{a(t_0)} = \frac{dt_s}{a(t_s)} \quad (34)$$

If we take dt_0 as the physical period of the signal as measured by the receiver and dt_s the physical period at emission, we find that the physical frequencies $\nu = 1/dt$, will satisfy:

$$\frac{\nu_s}{\nu_0} = \frac{a(t_0)}{a(t_s)} \quad (35)$$

Thus, in an expanding universe $a(t_0) > a(t_s)$ and therefore $\nu_0 < \nu_s$, the physical frequency decreases with the expansion. Similarly, since the comoving propagation speed is $c = 1$, i.e. along the geodesics $d\chi/d\eta = 1$, the comoving wavelength L is constant. Therefore, the physical wavelength $\lambda = La(t)$, satisfies:

$$\frac{\lambda_0}{a(t_0)} = \frac{\lambda_s}{a(t_s)} \quad (36)$$

and we can define the signal redshift as:

$$z = \frac{\lambda_0 - \lambda_S}{\lambda_S} = \frac{a(t_0)}{a(t_S)} - 1 \quad (37)$$

or

$$1 + z = \frac{a(t_0)}{a(t_S)} \quad (38)$$

Normalizing the scale factor today as $a(t_0) = 1$ we have:

$$1 + z = \frac{1}{a} \quad (39)$$

Therefore $z > 0$ implies an expanding universe whose expansion rate is given by the Hubble parameter:

$$H(t) = \frac{\dot{a}(t)}{a(t)} \quad (40)$$

whose present value corresponds to the Hubble constant $H_0 = H(t_0)$ (notice that in general the expansion rate depends on time).

1.2.2 Hubble Horizon and Particle Horizon

Consider two comoving objects separated by a comoving distance L . As shown before, the corresponding physical distance will grow as $D(t) = a(t)L$, and the recession velocity will be: $v(t) = \dot{D}(t) = \dot{a}(t)L$, which can be rewritten as:

$$v(t) = \frac{\dot{a}(t)}{a(t)}a(t)L = H(t)D(t) \quad (41)$$

This means that objects that at a given time t are separated by a physical distance $D(t) = H^{-1}(t)$ will recede at the speed of light. The sphere of radius $H^{-1}(t)$ is known as the Hubble sphere or Hubble horizon. Thus comoving objects inside(outside) our Hubble sphere recede from us sub(super)luminally. Notice that superluminal recession velocities are not in contradiction with Special Relativity theory as the comoving velocities of these objects are actually zero.

The Hubble sphere, in turn, recedes from us at a velocity:

$$\dot{H}^{-1}(t) = \frac{d}{dt} \left(\frac{a(t)}{\dot{a}(t)} \right) = 1 - \frac{a(t)\ddot{a}(t)}{\dot{a}^2(t)} \quad (42)$$

In such a way that if the universe expansion is decelerated $\ddot{a} < 0$ then $\dot{H}^{-1}(t) > 1$, i.e. the Hubble sphere grows superluminally and conversely, if the expansion is accelerated, it grows subluminally. This means that in an universe with decelerated expansion more and more objects enter the Hubble horizon, whereas in an accelerated universe objects which are initially inside will exit the Hubble sphere at some later time. The same effect happens to the physical wavelengths, which grow with the expansion. In an accelerated universe all the wavelengths will end up being larger than the Hubble radius at sufficiently late times. As we will show this effect will play a fundamental role for the generation of density perturbations during inflation.

On the other hand, if the universe had an origin in time (Big Bang), the maximum physical distance that a signal could have traveled up to a given time t , will be given by:

$$d_H(t) = a(t) \int_0^r \frac{dr}{\sqrt{1 - kr^2}} = a(t) \int_0^t \frac{dt}{a(t)} \quad (43)$$

where we have used the null geodesics equation, $ds^2 = 0$, that for radial signals reads:

$$dt^2 = a^2(t) \frac{dr^2}{1 - kr^2} \quad (44)$$

This distance is known as the particle horizon and is finite. The particle horizon sets the maximum size of a causally connected region at a given time.

1.2.3 Dynamics of FLRW Models

The evolution of the scale factor $a(t)$ and the curvature parameter k can be obtained by solving the Einstein equations for a RW metric:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi GT_{\mu\nu} \quad (45)$$

where $T_{\mu\nu}$ is the energy-momentum tensor representing the matter and energy content of the universe, which is the sum of the contributions of the different components: baryons, dark matter, radiation, dark energy, etc.

$$T_{\mu\nu} = \sum_{\alpha} T_{\mu\nu}^{(\alpha)} \quad (46)$$

In general each of these components $T_{\mu\nu}^{(\alpha)}$ can be well described as a perfect fluid: $T_{\mu\nu}^{(\alpha)} = (\rho^{(\alpha)}(t) + p^{(\alpha)}(t))u_{\mu}u_{\nu} - p^{(\alpha)}(t)g_{\mu\nu}$, where $\rho^{(\alpha)}(t)$ and $p^{(\alpha)}(t)$ are the energy density and pressure. On the other hand, for comoving observers in

coordinates (t, r, θ, ϕ) , the four velocities read:

$$u^\mu = (1, 0, 0, 0) \quad (47)$$

Thus, for each component, we will have:

$$T_{\nu}^{(\alpha)\mu} = \begin{pmatrix} \rho^{(\alpha)}(t) & & & \\ & -p^{(\alpha)}(t) & & \\ & & -p^{(\alpha)}(t) & \\ & & & -p^{(\alpha)}(t) \end{pmatrix} \quad (48)$$

On the other hand, the equation of state for each component will read:

- Non-relativistic matter (baryons and cold dark matter): $p = 0$
- Radiation (photons and relativistic particles): $p = \frac{1}{3}\rho$
- Cosmological constant: $p = -\rho$ (in general it could be possible to consider other dark energy models in which the equation of state $p = \omega\rho$ satisfies $\omega < -1/3$ in order to have accelerated expansion).

Assuming for simplicity that each component $T_{\nu}^{(\alpha)\mu}$ is independently conserved i.e.:

$$T_{;\nu}^{(\alpha)\mu\nu} = 0 \quad (49)$$

and taking the $\mu = 0$ component, we get:

$$\dot{\rho} + 3H(\rho + p) = 0 \quad (50)$$

which can be written as:

$$\frac{d}{da}(\rho a^3) = -3pa^2 \quad (51)$$

For a general equation of state: $p = \omega\rho$ with ω constant, we get:

$$\rho \propto a^{-3(1+\omega)} \quad (52)$$

Therefore:

- Non-relativistic matter: $\rho \propto a^{-3}$
- Radiation: $\rho \propto a^{-4}$
- Cosmological constant: $\rho \propto const.$

The non-vanishing components of the Einstein equations read:

$$\frac{\dot{a}^2}{a^2} + \frac{k}{a^2} = \frac{8\pi G}{3} \sum_{\alpha} \rho^{(\alpha)} \quad (53)$$

which is the Friedmann equation and:

$$2\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{k}{a^2} = -8\pi G \sum_{\alpha} p^{(\alpha)} \quad (54)$$

Subtracting (53) from (54), we get the acceleration equation:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \sum_{\alpha} (\rho^{(\alpha)} + 3p^{(\alpha)}) \quad (55)$$

This equation implies that for ordinary matter or radiation $\rho + 3p > 0$ (strong energy condition) and therefore the expansion would be decelerated $\ddot{a} < 0$. In order to have accelerated expansion as suggested by observations, a dominant component with negative pressure is required:

$$p < -\frac{1}{3}\rho \quad (56)$$

which is usually referred to as dark energy.

Evaluating the Friedmann equation at the present time, we get:

$$H_0^2 = \frac{8\pi G}{3}\rho_M^0 + \frac{8\pi G}{3}\rho_R^0 + \frac{8\pi G}{3}\rho_A^0 - \frac{k}{a_0^2} \quad (57)$$

so that:

$$1 = \underbrace{\frac{8\pi G}{3H_0^2}\rho_{M0}}_{\Omega_M} + \underbrace{\frac{8\pi G}{3H_0^2}\rho_{R0}}_{\Omega_R} + \underbrace{\frac{8\pi G}{3H_0^2}\rho_A}_{\Omega_A} - \underbrace{\frac{k}{a_0^2 H_0^2}}_{\Omega_K} \quad (58)$$

and we obtain the so called cosmic sum rule:

$$1 = \Omega_M + \Omega_R + \Omega_A + \Omega_K \quad (59)$$

Notice that the density parameters can be written as $\Omega_{\alpha} = \rho_{\alpha}/\rho_c$, where $\rho_c = 3H_0^2/(8\pi G)$ is the critical density. Accordingly, if the total energy density $\rho_{tot} = \rho_M + \rho_R + \rho_A$ is larger than the critical density then $\Omega_K < 0$ which implies positive curvature $k > 0$.

The values of the density parameters measured by Planck+WMAP 2013 [9] are:

$$\begin{aligned} \Omega_M &= \frac{8\pi G}{3H_0^2}\rho_{M0} = 0.314 \pm 0.020 \\ \Omega_R &= \frac{8\pi G}{3H_0^2}\rho_{R0} = 4.17 \times 10^{-5}h^{-2} \end{aligned}$$

$$\begin{aligned}\Omega_\Lambda &= \frac{8\pi G}{3H_0^2} \rho_\Lambda = 0.686 \pm 0.020 \\ \Omega_K &= -0.037 \pm 0.046 \quad (95\%, \text{ C.L.})\end{aligned}\quad (60)$$

where three neutrino families have been included in Ω_R .

Notice that Ω_K is compatible with 0 and therefore in the standard ΛCDM cosmology spatial sections are nearly flat. As a matter of fact, combining data from Planck with WMAP polarization and BAO it is possible to obtain even stronger limits on the curvature parameter [9]:

$$\Omega_K = 0.0000^{+0.0066}_{-0.0067} \quad (95\%, \text{ C.L.}) \quad (61)$$

On the other hand, the energy density in the form of a cosmological constant can be written as:

$$\rho_\Lambda \simeq (2 \times 10^{-3} \text{ eV})^4 \quad (62)$$

In Fig. 4 the different scaling of the various components are shown. At early times radiation dominates, but at a certain point, matter density equals radiation density:

$$\rho_M(z_{eq}) = \rho_R(z_{eq}) \quad (63)$$

$z_{eq} = 3,200 \pm 200$ which corresponds to $a_{eq} = 3.1 \times 10^{-4}$. After that the universe is matter dominated until the time when the cosmological constant starts dominating:

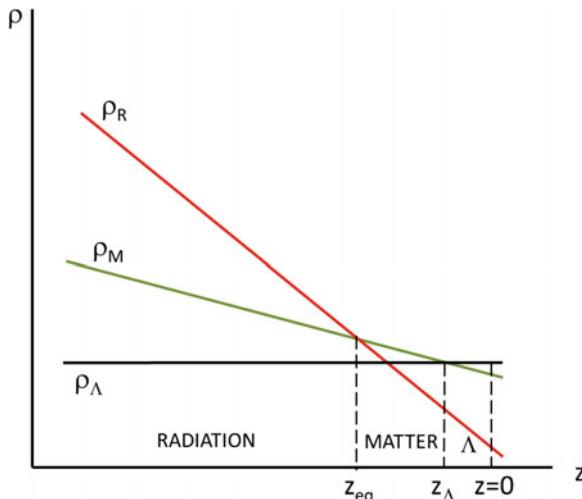


Fig. 4 Diagram showing the scaling of the different energy density components (logarithmic scales). At early times radiation is the dominant component, whereas at late times, the cosmological constant dominates

$$\rho_M(z_\Lambda) = \rho_\Lambda(z_\Lambda) \quad (64)$$

which corresponds to $z_\Lambda \simeq 0.30$ in Λ CDM. We observe that whereas throughout most of the expansion history the cosmological constant density has been negligible, its present value is however similar to the matter density. This has been referred to as the coincidence problem of the cosmological constant.

1.2.4 Explicit Solutions

Notice that giving the set of parameters $(H_0, \Omega_M, \Omega_R, \Omega_\Lambda)$ together with the cosmic sum rule, a cosmological model is completely determined (at the background level). It is possible to obtain explicit solutions for the scale factor as a function of time in some cases.

- Models without cosmological constant $\Omega_\Lambda = 0$,
 (a) Flat models with $\Omega_K = 0$:

$$a(t) = \begin{cases} \left(\frac{3}{2}H_0t\right)^{2/3} & \text{matter domination } (\Omega_M = 1) \\ (2H_0t)^{1/2} & \text{radiation domination } (\Omega_R = 1) \end{cases} \quad (65)$$

or, in conformal time:

$$a(\eta) = \begin{cases} \frac{H_0^2\eta^2}{4} & \text{matter domination } (\Omega_M = 1) \\ H_0\eta & \text{radiation domination } (\Omega_R = 1) \end{cases} \quad (66)$$

(b) In general, for $\Omega_K \neq 0$ the sign of k determines the fate of the universe:

$$\begin{aligned} k > 0 & \quad \text{recollapse} \\ k = 0 & \quad \text{critical} \\ k < 0 & \quad \text{eternal expansion} \end{aligned} \quad (67)$$

- Models with cosmological constant $\Omega_\Lambda \neq 0$.

In the general case, the Friedmann equation is analogous to the energy conservation equation of a particle moving in one dimensional space. Introducing the dimensionless time $\tau = H_0 t$ and taking $a_0 = 1$, we get from (51):

$$\underbrace{\dot{a}^2}_T - \underbrace{\frac{\Omega_M}{a}}_V - \Omega_\Lambda a^2 = \underbrace{\Omega_K}_E \quad (68)$$

where the particle energy would be given by Ω_K and the potential is:

$$V(a) = -\frac{\Omega_M}{a} - \Omega_\Lambda a^2 \quad (69)$$

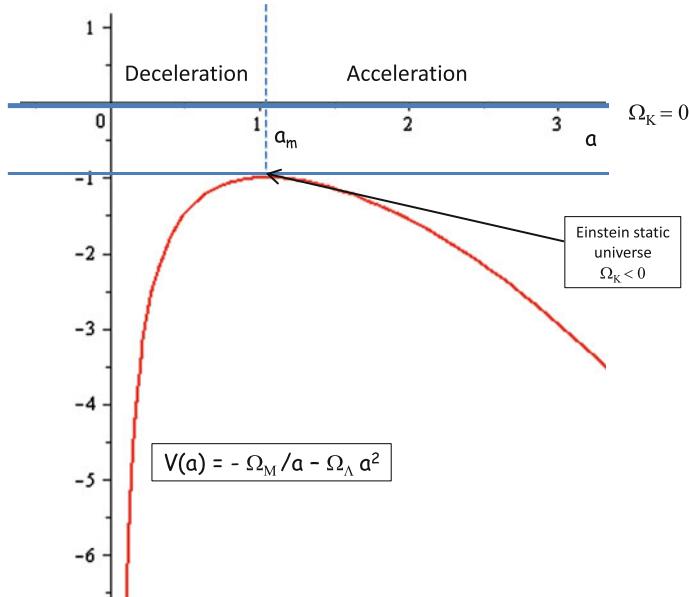


Fig. 5 Effective potential $V(a)$ determining the evolution of the scale factor in the case $\Omega_\Lambda > 0$

For $\Omega_\Lambda < 0$ the potential is a monotonically growing function, in such a way that all solution expand from the initial singularity up to a maximum value of the scale factor and then recollapse.

For $\Omega_\Lambda > 0$ we have $V < 0 \forall a$ and the potential has a maximum at:

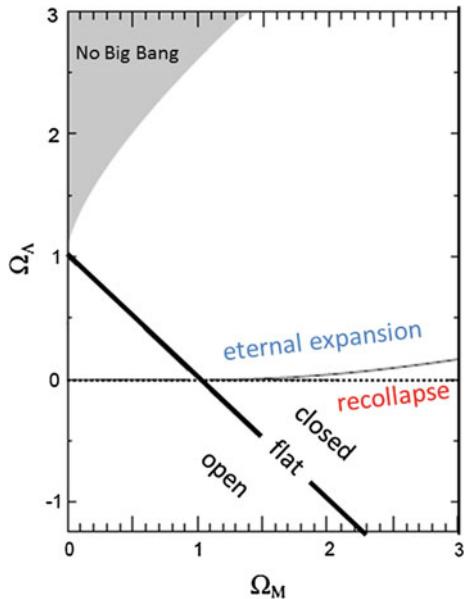
$$a_m = \left(\frac{\Omega_M}{2\Omega_\Lambda} \right)^{1/3} \quad (70)$$

see Fig. 5. So that for a given energy Ω_K , we have the following possible solutions depending on the value of a_r which is the turning point for which $\Omega_K = V(a_r)$:

$$\begin{aligned} \Omega_K &> V(a_m) && \text{eternal expansion} \\ \Omega_K &< V(a_m), \ a_r > 1 && \text{recollapse} \\ \Omega_K &< V(a_m), \ a_r < 1 && \text{no Big-Bang} \\ \Omega_K &= V(a_m) && \text{Einstein static universe} \end{aligned} \quad (71)$$

Notice that unlike the models without cosmological constant, now the sign of k does not determine the future evolution (see Fig. 6).

Fig. 6 Regions in the $\Omega_M - \Omega_A$ plane corresponding to eternal expansion, recollapse and absence of big bang



1.3 Distance Indicators

1.3.1 Luminosity Distance

In flat space-time, the apparent luminosity l of an object with known absolute luminosity L (standard candle) is given by:

$$l = \frac{L}{4\pi d^2} \quad (72)$$

where d is the distance to the object. Thus, by measuring the apparent luminosity it is possible to determine d . However, in an expanding universe physical distances grow with time so that the distance derived from this expression does not correspond to either the distance at emission or the distance at reception. This distance is known as luminosity distance.

Let us thus assume that an object with absolute luminosity L is located at a radial coordinate $r = r_1$ and starts to emit radiation at time t_1 during a time interval dt_1 . This radiation is received at $r = 0$ at time t_0 during a time interval dt_0 . Therefore, the apparent luminosity at reception will be:

$$l = \frac{L dt_1}{dt_0} \frac{a(t_1)}{a(t_0)} \frac{1}{4\pi r_1^2 a^2(t_0)} \quad (73)$$

where the factor $\frac{a(t_1)}{a(t_0)}$ accounts for the energy loss because of the frequency redshift and $r_1 a(t_0)$ is the physical radius of the sphere passing through $r = 0$ and centered at $r = r_1$ at time t_0 . Using again (34):

$$\frac{dt_1}{dt_0} = \frac{a(t_1)}{a(t_0)} \quad (74)$$

we get:

$$l = \frac{L}{4\pi r_1^2 a^2(t_0)} \left(\frac{a(t_1)}{a(t_0)} \right)^2 = \frac{L}{4\pi r_1^2 (1+z)^2} \quad (75)$$

where as usual we have taken $a(t_0) = 1$. Thus, we define the luminosity distance d_L as:

$$l = \frac{L}{4\pi d_L^2} \quad (76)$$

so that for a source with radial coordinate r , we have:

$$d_L(z) = r(1+z) \quad (77)$$

In order to obtain $r(z)$ we can integrate along the null photon geodesics $ds^2 = 0$, so that:

$$\begin{aligned} \int_0^r \frac{dr}{\sqrt{1 - kr^2}} &= \int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_{a_1}^{a_0} \frac{da}{a\dot{a}} = \int_{a_1}^{a_0} \frac{da}{a^2 H(a)} \\ &= - \int_z^0 \frac{dz}{H(z)} = \int_0^z \frac{dz}{H(z)} \end{aligned} \quad (78)$$

Using the Friedmann equation, we can write:

$$\frac{H^2}{H_0^2} = \Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_\Lambda + \Omega_K(1+z)^2 \quad (79)$$

so that we get:

$$\frac{1}{|k|^{1/2}} \operatorname{senn}^{-1}(r|k|^{1/2}) = \int_0^z \frac{dz}{H_0 E(z)} \quad (80)$$

where $\text{senn}(x)$ was defined in (31) and:

$$E(z) = \sqrt{\Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_A + \Omega_K(1+z)^2} \quad (81)$$

Finally, taking into account that $|k|^{1/2} = \sqrt{|\Omega_K|}H_0$, we get:

$$r = H_0^{-1} \frac{1}{\sqrt{|\Omega_K|}} \text{senn} \left(\sqrt{|\Omega_K|} \int_0^z \frac{dz}{E(z)} \right) \quad (82)$$

so that from (78) we can write:

$$d_L(z) = (1+z)H_0^{-1} \frac{1}{\sqrt{|\Omega_K|}} \text{senn} \left(\sqrt{|\Omega_K|} \int_0^z \frac{dz}{E(z)} \right) \quad (83)$$

The luminosity distance is related to the distance modulus μ (the difference between the apparent m and the absolute M magnitudes of the object) by this expression:

$$\mu = m - M = 5 \log \left(\frac{d_L}{\text{Mpc}} \right) + 25 \quad (84)$$

The curve $\mu(z)$ known as Hubble diagram is thus a prediction of a given cosmological model which can be compared with observations. By extending the Hubble diagrams to high redshifts it would be possible to discriminate between different cosmologies (see Fig. 7). This possibility became real by means of the use of type Ia supernovae as standard candles [11, 12].

Type Ia SN explosions occur when a white dwarf star present in a binary system accretes enough material from the companion star in such a way that its mass reaches the Chandrasekhar limit of $1.44 M_\odot$. A white dwarf is a stellar remnant which is formed when a star has exhausted all its nuclear fuel, so that the gravitational attraction is only compensated by the electron degeneracy pressure. When the mass reaches the Chandrasekhar limit, the degeneracy pressure is unable to support the gravitational pull, the star becomes unstable and an explosion takes place with absolute luminosity comparable to that of the host galaxy. The exploding stars always have masses close to the Chandrasekhar limit and therefore there is small variability in the absolute magnitude, this fact makes SN Ia good distance indicators. There are however small variations in the luminosity curves which are correlated with the explosion duration. The maximum luminosity is reached in around three weeks and then declines in some months period. The longer explosions are also more luminous. Calibrating this relation with nearby SNIa it has been possible to determine distances of several thousands Mpc with a precision around 10 %. The SNIa rate is quite low (around 1 SNIa explosion per 500 years per galaxy), so that constructing high-redshift Hubble diagrams requires the observation of large portions of the sky for several

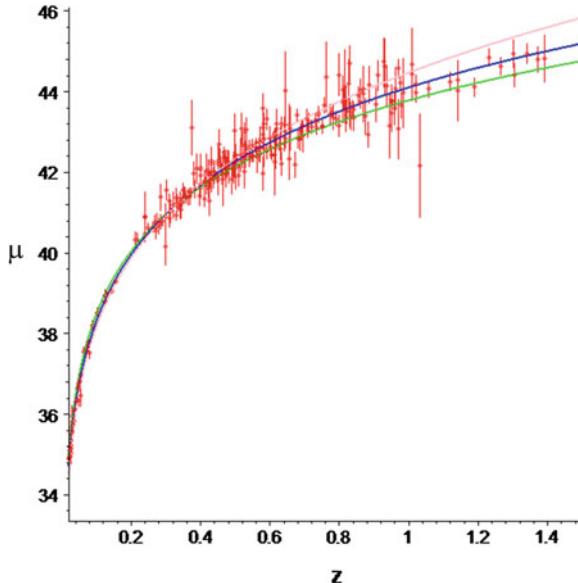


Fig. 7 High-redshift Hubble diagram $\mu(z)$. Data points correspond to the Union data set [13]. Green (lower) curve corresponds to an Einstein-de Sitter universe with $\Omega_M = 1$ and decelerated expansion. The blue (middle) curve corresponds to the Standard Λ CDM cosmology with $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ (which provides the best fit) and the pink (upper) curve corresponds to the de Sitter universe with $\Omega_\Lambda = 1$

years. This method has allowed to determine [11, 12] the acceleration rate of the universe, see Fig. 7.

1.3.2 Angular Diameter Distance

Another way of measuring distances in cosmology is using the so called angular diameter distance d_A . Thus, in flat space-time the angle subtended by an object of diameter D which is located at a distance d is given by:

$$\theta = \frac{D}{d} \quad (85)$$

where we have assumed that D is transverse to the line of sight and $d \gg D$. If we know D (standard ruler), measuring θ we can obtain the distance to the object d . In an expanding universe it is again necessary to interpret physically this way of measuring.

Let us thus assume that we know the physical size D at a given time t_1 of an object located at $r = r_1$ which emits radiation at time t_1 which is received at $r = 0$ at time t_0 . The comoving size of the object will be $\lambda = D/a(t_1)$ which is constant in time.

Therefore in the comoving frame, the angle subtended by the object will be:

$$\theta = \frac{D}{r_1 a(t_1)} \quad (86)$$

and the distance obtained from the expression:

$$\theta = \frac{D}{d_A} \quad (87)$$

is called the angular diameter distance. It is given by:

$$d_A = r_1 a(t_1) = r_1 a(t_0) \frac{a(t_1)}{a(t_0)} = r_1 \frac{a(t_0)}{1+z} \quad (88)$$

so that, taking $a(t_0) = 1$, we get:

$$d_A = \frac{r}{1+z} = \frac{d_L(z)}{(1+z)^2} \quad (89)$$

Galaxies tend to group at a comoving distance of around 153 Mpc, i.e. there is an excess in the number of galaxy pairs separated by this distance. This scale, known as the baryon acoustic oscillations (BAO) peak scale, has been used recently as standard ruler [14–17].

2 Thermodynamics in an Expanding Universe

In order to determine the abundances of the different particle species throughout the evolution of the universe, it is necessary to know the corresponding distribution functions. If the reaction rates of a given species is fast enough compared to the rate of expansion:

$$\Gamma > H \quad (\text{local thermodynamic equilibrium}) \quad (90)$$

then the species will be in local thermodynamic equilibrium. For much of its history, the universe has been in equilibrium, however departures from equilibrium have played a fundamental role in cosmology. Indeed, if the universe had been always in equilibrium then at present it would only contain a uniform background of cosmic radiation with a negligible amount of baryons and no structure. Thus the phenomenon of decoupling in an expanding universe in which a particle species abandons thermodynamic equilibrium, i.e.:

$$\Gamma \lesssim H \quad (\text{decoupling}) \quad (91)$$

plays a fundamental role in cosmology and it is the basis of big-bang nucleosynthesis, matter-radiation decoupling, the existence of a cosmic neutrino background or the relic density of dark matter.

2.1 Equilibrium Thermodynamics

Let us briefly review the main concepts of equilibrium thermodynamics in an expanding universe before moving to the decoupling phenomenon. For rapidly interacting species, the phase-space distribution function, which provides the occupation number of each state, reads:

$$f(\mathbf{k}) = \frac{1}{e^{\frac{E-\mu}{T}} \pm 1} \quad (92)$$

where the + sign corresponds to fermions (Fermi-Dirac) and the - sign to bosons (Bose-Einstein). \mathbf{k} corresponds to the physical momentum of the particles (which scales with the expansion) and

$$E^2 = |\mathbf{k}|^2 + m^2 \quad (93)$$

is the energy of the particle. μ is the chemical potential which is introduced whenever conserved charges are present.

From the occupation number and taking into account that for a particle species with g internal degrees of freedom, the density of states in phase space is $\frac{g}{(2\pi)^3}$, it is possible to calculate the different thermodynamic quantities. Thus the particle number density reads:

$$n = \frac{g}{(2\pi)^3} \int f(\mathbf{k}) d^3k = \frac{g}{2\pi^2} \int_m^\infty \frac{(E^2 - m^2)^{1/2}}{e^{(E-\mu)/T} \pm 1} E dE \quad (94)$$

The energy density is:

$$\rho = \frac{g}{(2\pi)^3} \int Ef(\mathbf{k}) d^3k = \frac{g}{2\pi^2} \int_m^\infty \frac{(E^2 - m^2)^{1/2}}{e^{(E-\mu)/T} \pm 1} E^2 dE \quad (95)$$

and finally, taking into account that the contribution per particle to the pressure is $\mathbf{k}^2/(3E)$, we get:

$$p = \frac{g}{(2\pi)^3} \int \frac{k^2}{3E} f(\mathbf{k}) d^3k = \frac{g}{6\pi^2} \int_m^\infty \frac{(E^2 - m^2)^{3/2}}{e^{(E-\mu)/T} \pm 1} dE \quad (96)$$

The entropy density can be obtained from the second law of thermodynamics in an expanding universe:

$$TdS = dE + pdV = d(\rho a^3) + pda^3 = d((\rho + p)a^3) - a^3 dp \quad (97)$$

using:

$$\frac{dp}{dT} = \frac{\rho + p}{T} \quad (98)$$

which can be obtained from (96) assuming $\mu/T \ll 1$, we get

$$TdS = d((\rho + p)a^3) - (\rho + p)a^3 \frac{dT}{T} = Td \left(\frac{\rho + p}{T} a^3 \right) \quad (99)$$

So that, the entropy density is defined as:

$$s = \frac{(\rho + p)}{T} \quad (100)$$

and the total entropy is $S = a^3 s$. It is possible to show using the energy-momentum tensor conservation (51), that the total entropy is indeed conserved:

$$\begin{aligned} \frac{d}{dt}(\rho a^3) &= -p \frac{da^3}{dt} \\ \frac{d}{dT}(\rho a^3) &= -p \frac{da^3}{dT} \\ \frac{d}{dT}((\rho + p)a^3) &= a^3 \frac{dp}{dT} = a^3 \frac{\rho + p}{T} \\ T \frac{d}{dT} \left(\frac{(\rho + p)a^3}{T} \right) &= T \frac{ds}{dT} = 0 \end{aligned} \quad (101)$$

i.e. the universe expansion is adiabatic.

2.1.1 Relativistic Particles ($T \gg m, \mu$)

It is useful to get the limits of the previous expressions in the case of ultrarelativistic particles. Thus:

$$\rho = \begin{cases} \frac{\pi^2}{30} g T^4, & \text{bosons} \\ \frac{7}{8} \frac{\pi^2}{30} g T^4, & \text{fermions} \end{cases} \quad (102)$$

$$n = \begin{cases} \frac{\zeta(3)}{\pi^2} g T^3, & \text{bosons} \\ \frac{3}{4} \frac{\zeta(3)}{\pi^2} g T^3, & \text{fermions} \end{cases} \quad (103)$$

where $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$ with $\zeta(3) \simeq 1.202$. Finally, as expected:

$$p = \frac{1}{3} \rho. \quad (104)$$

2.1.2 Non-relativistic Particles ($T \ll m$)

$$n = g \left(\frac{mT}{2\pi} \right)^{3/2} e^{-\frac{m-\mu}{T}} \quad (105)$$

$$\rho = mn \quad (106)$$

$$p = nT \ll \rho \quad (107)$$

Notice that for a given T and g we always find that:

$$n_R \gg n_{NR} \quad (108)$$

$$\rho_R \gg \rho_{NR} \quad (109)$$

i.e. due to the exponential suppression of the equilibrium abundances of non-relativistic particles, equilibrium particle density and energy are always dominated by ultrarelativistic particles, i.e. for non-relativistic particles to give non-negligible contributions they must be decoupled.

Since for species in thermal equilibrium, the relativistic particles contribute the most, we can write in a good approximation:

$$\rho = \frac{\pi^2}{30} g_*(T) T^4 \quad (110)$$

$$p = \frac{\pi^2}{90} g_*(T) T^4 \quad (111)$$

$$s = \frac{2\pi^2}{45} g_{*S}(T) T^3 \quad (112)$$

with:

$$g_*(T) = \sum_{bosons} g_i \left(\frac{T_i}{T} \right)^4 + \frac{7}{8} \sum_{fermions} g_i \left(\frac{T_i}{T} \right)^4 \quad (113)$$

$$g_{*S}(T) = \sum_{bosons} g_i \left(\frac{T_i}{T} \right)^3 + \frac{7}{8} \sum_{fermions} g_i \left(\frac{T_i}{T} \right)^3 \quad (114)$$

are the effective number of degrees of freedom contributing to the energy density and entropy respectively, and T_i is the temperature of species i assuming that it decoupled when it was still relativistic. For species coupled to the photons, they all share the same temperature $T_i = T$.

Since the total entropy S is conserved, we have:

$$S = a^3 s = \frac{2\pi^2}{45} g_{*S}(T) T^3 a^3 = const. \quad (115)$$

and therefore:

$$T \propto a^{-1} g_{*S}^{-1/3}(T) \quad (116)$$

i.e. for constant g_{*S} the temperature scales as $T \propto 1/a$.

However, when a particle species in equilibrium with radiation becomes non-relativistic, g_{*S} decreases, but since the total entropy is conserved, then the radiation temperature T should increase, i.e. the entropy is transferred to the relativistic particles which are heated up with respect to the decoupled species.

2.2 Decoupling in an Expanding Universe

The interaction rate of a particle species can be written as:

$$\Gamma = n_{eq} \langle \sigma v \rangle \quad (117)$$

where $\langle \rangle$ denotes the thermal average at a given temperature, σ is the interaction cross section and v is the particle velocity. We will denote n_{eq} the equilibrium abundance to distinguish it from the actual abundance n .

Let us consider for simplicity a stable particle Ψ which annihilates by pairs into $X\bar{X}$ which are also stable, i.e.:



assuming also for simplicity that X, \bar{X} have equilibrium distributions. In this case, it can be shown that the number density of particles Ψ, n will be given by the Boltzmann equation:

$$\frac{dn}{dt} = -3Hn - \langle \sigma_A v \rangle \left(n^2 - n_{eq}^2 \right) \quad (119)$$

where the first term on the r.h.s takes into account the dilution due to the expansion and the second one is the collision term which takes into account the change in the number density due to the creation and annihilation of $\psi\bar{\psi}$ pairs. We see from the Boltzmann equation that:

- when $\Gamma(T) \gg H(T)$ the collision term dominates over the dilution one, and the species remain in thermal equilibrium with $n \simeq n_{eq}$
- when $\Gamma(T) \ll H(T)$ the second term is negligible and $n \propto a^{-3}$, the species is decoupled. On the other hand, we have shown before that the entropy density always scales as $s \propto a^{-3}$, and accordingly, after decoupling:

$$\frac{n}{s} = \text{const.} \quad (120)$$

i.e. after decoupling the abundance remains frozen with respect to the entropy density. The condition:

$$\Gamma(T_{dec}) = H(T_{dec}) \quad (121)$$

sets the decoupling temperature T_{dec} .

We will study the decoupling phenomenon in the two limiting cases of interest: species which are ultrarelativistic at the decoupling time (hot relics) and non-relativistic species at decoupling (cold relics).

2.2.1 Hot Relics $m \ll T_{dec}$

For ultrarelativistic particles, the equilibrium abundance is given by:

$$n_{eq} = g_{eff} \frac{\zeta(3)}{\pi^2} T^3 \quad (122)$$

with $g_{eff} = g$ (bosons) and $3g/4$ (fermions).

On the other hand the entropy density reads:

$$s = \frac{2\pi^2}{45} g_{*S}(T) T^3 \quad (123)$$

and therefore, using the fact that the relative abundance remains frozen after decoupling, we have that the present relative abundance (at $T = T_0$) will be given by:

$$\left. \frac{n}{s} \right|_{T_0} = \left. \frac{n}{s} \right|_{T_{dec}} = \frac{45\zeta(3)g_{eff}}{2\pi^4 g_{*S}(T_{dec})} \quad (124)$$

and accordingly, the present abundance will be:

$$n_0 = 0.278 \frac{g_{eff} s_0}{g_{*S}(T_{dec})} \quad (125)$$

Thus in order to determine the relic density, we need to calculate:

- T_{dec}
- $g_{*S}(T_{dec})$
- s_0 (entropy density today)

T_{dec} can be obtained from the decoupling condition $\Gamma(T_{dec}) = H(T_{dec})$. Assuming that decoupling took place in the radiation dominated era we will have:

$$H^2 = \frac{8\pi G}{3} \rho_R = \frac{8\pi G}{3} \frac{\pi^2}{30} g_*(T) T^4 \quad (126)$$

and therefore:

$$\Gamma(T) = 1.66 \sqrt{g_* G} T^2 \quad (127)$$

2.2.2 Example: Relic Neutrino Density

Let us consider the case of SM neutrinos which annihilate in processes:

$$\nu \bar{\nu} \leftrightarrow e^+ e^- \quad (128)$$

The corresponding annihilation cross-section can be estimated to be of order:

$$\sigma_A \simeq G_F s \quad (129)$$

where $G_F = (300 \text{ GeV})^{-2}$ is the Fermi constant and s is the collision center of mass energy. The thermal average can be estimated to be of order:

$$\langle \sigma_A v \rangle \simeq G_F T^2 \quad (130)$$

and therefore

$$\Gamma = n_{eq} \langle \sigma_A v \rangle \simeq G_F T^5 \quad (131)$$

that using (127) implies:

$$T_{dec} \simeq 1 \text{ MeV} \quad (132)$$

Thus, we see that indeed neutrino decoupling takes place in the radiation dominated era. When $T = T_{dec}$ the relativistic degrees of freedom include: e^+ , e^- , ν , $\bar{\nu}$ and γ , therefore:

$$g_{*S}(T_{dec}) = \underbrace{2}_{\gamma} + \frac{7}{8} \underbrace{4}_{e^+ e^-} + \frac{7}{8} \underbrace{2 \times 3}_{\nu \bar{\nu}} = 10.75 \quad (133)$$

where we have used the fact that neutrinos only have two degrees of freedom per family coupled to the rest of SM particles. Finally, in order to calculate the entropy density s_0 , we have:

$$s_0 = \frac{2\pi^2}{45} g_{*S}(T_0) T_0^3 \quad (134)$$

Thus, we need to obtain $g_{*S}(T_0)$. However here we find a subtlety. In principle, the only relativistic degrees of freedom today would be photons (assuming that neutrinos have small masses but larger than T_0). However, as commented before, neutrinos decoupled when they were still ultrarelativistic and its distribution function got frozen as that of a relativistic species, thus contributing in a non-negligible amount to the energy and entropy densities. However, today the temperature of the neutrino background T_ν^0 does not necessarily agree with that of the photon background T_0 . In order to calculate T_ν^0 we use the fact that the temperature of a decoupled relativistic species scales as $T_\nu \propto a^{-1}$ whereas as shown before the photon temperature scales as $T \propto a^{-1} g_{*S\gamma}(T)^{-1/3}$ where $g_{*S\gamma}$ denotes the number of relativistic degrees of freedom which remain coupled to the photons. Since both temperatures coincide at decoupling, we have:

$$\frac{T^3}{T_\nu^3} = \frac{g_{*S\gamma}(T_{dec})}{g_{*S\gamma}(T)} \quad (135)$$

When $T \simeq m_e \simeq 0.5$ MeV, $e^+ e^-$ become non-relativistic and therefore $g_{*S\gamma}$ decreases in such a way that:

$$g_{*S\gamma}(T_{dec}) = \underbrace{2}_{\gamma} + \frac{7}{8} \cdot \underbrace{4}_{e^+ e^-} = \frac{11}{2}$$

$$g_{*S\gamma}(T_0) = 2 \quad (136)$$

Therefore, today:

$$\frac{T_0}{T_\nu^0} = \left(\frac{g_{*S\gamma}(T_{dec})}{g_{*S\gamma}(T_0)} \right)^{1/3} = \left(\frac{11}{4} \right)^{1/3} \quad (137)$$

i.e. the present neutrino temperature is:

$$T_\nu^0 = 1.96 \text{ K} \quad (138)$$

Thus we get:

$$g_{*S}(T_0) = 2 + \frac{7}{8} \cdot 2 \cdot 3 \left(\frac{T_\nu^0}{T_0} \right)^3 = 2 + \frac{7}{8} \cdot 2 \cdot 3 \left(\frac{4}{11} \right) = 3.91 \quad (139)$$

and therefore from (134) the present entropy density reads:

$$s_0 = 2,970 \text{ cm}^{-3} \quad (140)$$

Finally, taking into account that for each neutrino family $g_{eff} = 2 \cdot 3/4 = 3/2$ (we have included a 2 factor to take into account the degrees of freedom corresponding to neutrino-antineutrino), we get the present number density of each neutrino species as:

$$n_0 = 2,970 \cdot 0.278 \frac{3/2}{10.75} = 115 \text{ cm}^{-3} \quad (141)$$

If neutrinos are non-relativistic today, then the corresponding energy density would be for each family:

$$\rho_\nu = n_0 m_\nu \quad (142)$$

and therefore:

$$\Omega_\nu h^2 = \frac{\rho_\nu h^2}{\rho_c} = \frac{n_0 m_\nu h^2}{\frac{3H_0^2}{8\pi G}} = \frac{m_\nu}{91.5 \text{ eV}} \quad (143)$$

Using $\Omega_M h^2 = 0.14$ we can thus obtain an upper bound to the sum of all the neutrino masses:

$$\sum_\nu \Omega_\nu h^2 = \frac{\sum_\nu m_\nu}{91.5 \text{ eV}} \leq 0.14 \quad (144)$$

and therefore:

$$\sum_\nu m_\nu \leq 13 \text{ eV} \quad (145)$$

However, as shown in Fig. 1 models containing only hot dark matter will be in contradiction with observations of structures on small scales. This is due to the fact that hot dark matter particles propagating close to the speed of light could escape

from overdense regions, erasing the density fluctuations on scales smaller than the free-streaming scale λ_{FS} , which is the maximum distance that a particle can travel from the initial time until matter-radiation equality [1].

The present limits on neutrino masses obtained by Planck from CMB anisotropies read [9]:

$$\sum_{\nu} m_{\nu} \leq 0.23 \text{ eV } 95\% \text{ C.L.} \quad (146)$$

The obtained upper limits can be compared with the laboratory limits for the three kinds of neutrinos:

$$\begin{aligned} m_{\nu_e} &< 3 \text{ eV} \\ m_{\nu_{\mu}} &< 0.19 \text{ MeV} \\ m_{\nu_{\tau}} &< 18.2 \text{ MeV} \end{aligned} \quad (147)$$

2.2.3 Cold Relics $m \gg T_{dec}$

In this case, particles are non-relativistic at decoupling and the particle density before decoupling reads:

$$n(T) = n_{eq}(T) = g \left(\frac{mT}{2\pi} \right)^{3/2} e^{-\frac{m}{T}} \quad (148)$$

Therefore n_{eq}/s decreases exponentially in this case. When the species decouples $n/s = \text{const}$. The solution of the Boltzmann equation shows that the larger the annihilation cross-section σ_A , the later decoupling takes place. On the other hand, the later decoupling takes place, the smaller the relic abundance. Thus, in order to have a large relic abundance, we require weakly interacting massive particles (WIMP) (see Fig. 8).

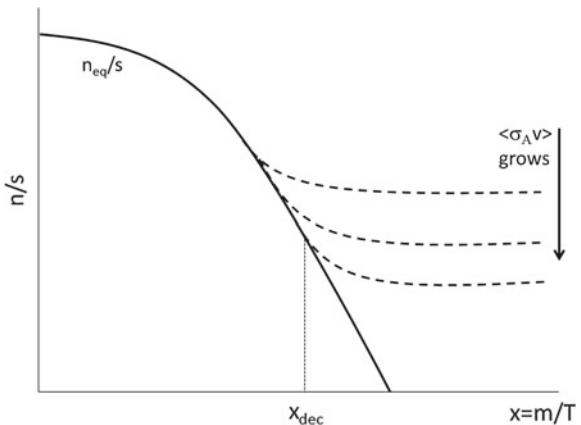
Now, since n_{eq} decreases exponentially, in order to determine the relic abundance, the decoupling temperature should be precisely calculated, so that a numerical solution of the Boltzmann equation is required. Thus, defining $x = m/T$ it is possible to find an implicit equation for the decoupling temperature [1]:

$$x_{dec} = \ln \left(\frac{0.038c(c+2)M_p m \langle \sigma_A v \rangle}{g_*^{1/2} x_{dec}^{1/2}} \right) \quad (149)$$

where the fit to the numerical solution requires $c \simeq 0.5$. This equation can be solved iteratively, so that generically, we have:

$$x_{dec} = \frac{m}{T_{dec}} \sim 20 \quad (150)$$

Fig. 8 Evolution of n/s as a function of $x = m/T \cdot x_{dec}$ corresponds to the decoupling temperature below which the relative abundance gets frozen. The different dotted lines correspond to increasing values of the annihilation cross-section



On the other hand, the solution of the Boltzmann equation shows:

$$\Omega_{CDM} h^2 = 8.8 \cdot 10^{-11} \text{ GeV}^{-2} \frac{x_{dec}}{\langle \sigma_A v \rangle} \quad (151)$$

Thus, in order to obtain the correct abundance of cold dark matter $\Omega_{CDM} h^2 = 0.14$, the typical value of the annihilation cross section should correspond to that of electroweak interactions (WIMP miracle). However, none of the SM particles are suitable CDM candidates as the potential WIMP candidates are unstable. This is one of the most pressing arguments for the existence of new particles beyond the SM.

3 Inflation

The standard cosmological model provides a good description of the universe evolution from approximately one second after the Big Bang until present, some 13,800 Myr after. Before the formation of the light elements in the primordial nucleosynthesis process, we do not have observational means to test the model. However, from a theoretical point of view we know that the standard evolution must fail at some point in the early history of the universe.

This claim is based on the existence of several problems of the standard cosmology, namely, the flatness and the horizon problems. A possible solution to these problems (and the only viable to date) is a phase of inflation in the early universe [18]. Although there are some evidence that supports the idea of inflation, at the moment it is still a paradigm which can be implemented by many different models which, in some cases, present data are already able to falsify.

3.1 The Flatness Problem

Measurements of the total energy content of the universe suggest that $|\Omega_K| < 0.08$ at 95 % C.L. (Planck+WMAP [9]). However, from the definition of $\Omega_K(t)$, we can determine its time evolution as follows:

$$\Omega_K(t) = -\frac{k}{(aH)^2} \quad (152)$$

Thus, we know that $a(t) \propto t^{1/2}$ in the radiation era and $a(t) \propto t^{2/3}$ in the matter era, so that in both cases $H(t) \propto 1/t$. This means that:

$$\Omega_K(t) \propto \begin{cases} t, & \text{radiation era} \\ t^{2/3}, & \text{matter era} \end{cases} \quad (153)$$

i.e. $\Omega_K(t)$ grows with time. Thus, in order for the present value to be < 0.08 , the value at the nucleosynthesis era $t_{nuc} \sim 1$ s should be:

$$|\Omega_K(t_{nuc})| < 0.08 \left(\frac{t_{nuc}}{t_{eq}} \right) \left(\frac{t_{eq}}{t_0} \right)^{2/3} \simeq 0.08 \left(\frac{1 \text{ s}}{10^{10} \text{ s}} \right) \left(\frac{10^{10} \text{ s}}{10^{17} \text{ s}} \right)^{2/3} \simeq 10^{-16} \quad (154)$$

i.e.:

$$\Omega_{tot}(t_{nuc}) = 1 - \Omega_K(t_{nuc}) = 1 \pm 10^{-16} \quad (155)$$

Thus, the spatial sections should have been extremely flat at the nucleosynthesis era otherwise the present curvature would be too large to agree with observations. From a dynamical point of view, we see that the solutions with $\Omega_K \simeq 0$ are unstable, a very small perturbation would make the universe either positively or negatively curved, in other words, the universe we observed is very “unnatural” within the standard model evolution and it requires an extreme fine tuning of its initial conditions.

3.2 Horizon Problem

The cosmic microwave background radiation is extremely isotropic over the whole sky. CMB photons reaching us today were last scattered at decoupling time $t_{dec} \sim 380,000$ years after the Big Bang, which corresponds to a $z_{dec} \simeq 1100$. This redshift determines the last scattering surface (LSS), a sphere with a comoving radius of:

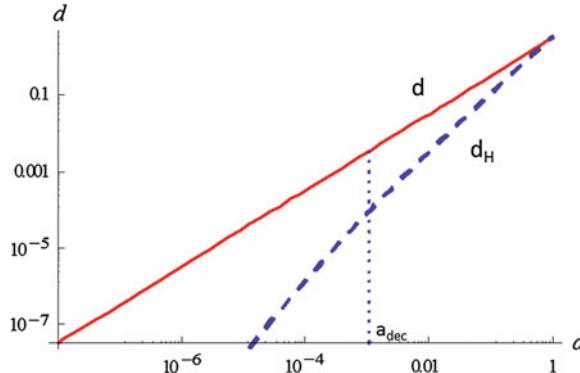


Fig. 9 Evolution of the physical radius of the LSS d and particle horizon d_H in H_0^{-1} units as a function of scale factor a (logarithmic scale) when no inflation period took place in the past. The particle horizon grows faster than the physical radius of the LSS, but it is still several orders of magnitude smaller at decoupling time a_{dec}

$$R = \int_0^{z_{dec}} \frac{dz}{H_0 E(z)} \simeq 3.2 H_0^{-1} \quad (156)$$

where we have used (78). The corresponding physical distance at decoupling being $d(t_{dec}) = R a_{dec}$. We can compare this distance with the size of the causally connected region at decoupling, i.e. the particle horizon, which can be obtained from:

$$d_H(t) = a(t) \int_{t_0}^t \frac{dt}{a(t)} = \frac{1}{1+z} \int_z^\infty \frac{dz}{H_0 E(z)} \quad (157)$$

i.e. today $d_H(z=0) \sim 3.3 H_0^{-1}$, so that $d(t_0) \simeq d_H(t_0)$. Assuming a power law behaviour for the scale factor $a(t) \propto t^\alpha$, it is easy to see that:

$$d_H(t) \propto \begin{cases} a^2, & \text{radiation era} \\ a^{2/3}, & \text{matter era} \end{cases} \quad (158)$$

whereas $d \propto a$, i.e. at decoupling $d(t_{dec}) \gg d_H(t_{dec})$ (see Fig. 9). The physical distance between two opposite points on the LSS at decoupling was much larger than the physical size of the causally connected region at that time. As a matter of fact, points on the LSS separated in the sky by more than $\sim 1^\circ$ were never in causal contact. Then, why is the CMB temperature approximately the same in any direction?

3.3 Inflation

A simple way of solving these problems is to modify the cosmological evolution in the very early times, replacing the decelerated expansion by a short period of accelerated expansion with $\ddot{a} > 0$ known as inflation [19, 20]. If this phase lasted for a sufficiently long time, then inflation solves the mentioned problems. A typical inflationary evolution would correspond to a (quasi-) exponential growth $a(t) \propto e^{H_I t}$, with $H_I = \text{const.}$, being the Hubble parameter during inflation.

3.3.1 Inflation Solves the Flatness Problem

During the exponential phase $(aH)^2 \propto e^{2H_I t}$, so that $\Omega_K(t_f)/\Omega_K(t_i) = e^{-2N}$, where $N = H_I(t_f - t_i)$, with t_f, t_i the final and initial times of inflation. N is the number of e-folds of inflation which determines the growth produced and is used as a measure of the duration of the inflationary phase. Typical inflationary models are able to generate $N \gtrsim 50$, so that the value of the curvature parameter can be typically reduced at least in a factor 10^{-44} at the end of inflation, thus solving the flatness problem.

3.3.2 Inflation Solves the Horizon Problem

Let us consider a simple example with three phases in which a short period of inflation takes place between two decelerated eras:

$$a(t) \propto \begin{cases} t^\alpha, & t_0 < t < t_i \\ e^{H_I t}, & t_i < t < t_f \\ t^\alpha, & t > t_f \end{cases} \quad (159)$$

with $\alpha < 1$. The particle horizon in this case grows as follows:

$$d_H(t) \propto \begin{cases} a^{1/\alpha}, & t_0 < t < t_i \\ a, & t_i < t < t_c \\ a^{1/\alpha}, & t > t_c \end{cases} \quad (160)$$

with $t_c > t_f$. As shown in Fig. 10, if the physical size $d(t_i) < H_I^{-1}$ at the beginning of inflation, then the particle horizon is always larger than $d(t)$ after inflation. Therefore we can write the condition for inflation to solve the horizon problem as:

$$H_0^{-1} a(t_i) < H_I^{-1} \quad (161)$$

This condition translates into a limit on the minimum number of e-folds of inflation required to solve the problem. Using the Friedmann equation during inflation:

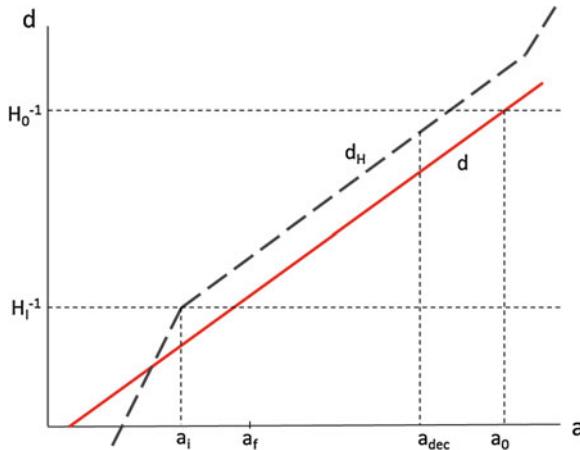


Fig. 10 Evolution of the physical radius of the LSS d and particle horizon d_H (logarithmic scales) as a function of scale factor a when an inflation period took place in the past (a_i and a_f denote the initial and final times of the inflation period). The particle horizon grows faster than the physical radius of the LSS, and can be several orders of magnitude above d at matter-radiation decoupling a_{dec} . The present time is denoted by a_0

$$H_I^2 = \frac{8\pi G}{3} \rho_I \quad (162)$$

with ρ_I the (almost) constant energy density driving inflation, we can write [18]:

$$N \gtrsim 53 + \frac{1}{3} \log \left(\frac{T_{RH}}{10^{10} \text{ GeV}} \right) + \frac{2}{3} \log \left(\frac{\rho_I^{1/4}}{10^{14} \text{ GeV}} \right) \quad (163)$$

where it is usually assumed that after inflation there is a period of reheating with a maximum temperature T_{RH} in which the energy density driving inflation ρ_I is transferred to radiation. In this way a transition from the quasi-de Sitter inflationary phase to the radiation dominated era takes place. We see that for typical values of T_{RH} and of the energy scale of inflation $\rho_I^{1/4}$, the required number of e-folds is around 50.

3.4 Inflation Models

As commented before, inflation is a paradigm without a concrete implementation. There are many models of inflation proposed to date, but the most popular and also the simplest ones are based on the existence of a scalar field (inflaton field) whose almost constant potential energy density is responsible for the accelerated expansion.

3.4.1 Slow-Roll Inflation

Let us consider a scalar field in an expanding universe, whose action is given by:

$$S = \int d^4x \sqrt{g} \left(\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right) \quad (164)$$

Varying with respect to the ϕ field we obtain the corresponding equations of motion:

$$\square\phi + V'(\phi) = 0 \quad (165)$$

Assuming that $\phi = \phi(t)$ is a homogeneous field and the background metric is RW, we obtain:

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0 \quad (166)$$

Thus we see that the scalar field behaves as an oscillator with a friction term which depends on the Hubble parameter.

On the other hand, varying with respect to the metric tensor, we get the corresponding energy-momentum tensor:

$$T_{\mu\nu} = -g_{\mu\nu} \left(\frac{1}{2} g^{\rho\sigma} \partial_\rho \phi \partial_\sigma \phi - V(\phi) \right) + \partial_\mu \phi \partial_\nu \phi \quad (167)$$

so that the non-vanishing components are:

$$\begin{aligned} \rho &= T_0^0 = \frac{1}{2} \dot{\phi}^2 + V(\phi) \\ p &= -T_i^i = \frac{1}{2} \dot{\phi}^2 - V(\phi) \end{aligned} \quad (168)$$

Thus, assuming that during inflation, the total energy-momentum tensor is dominated by the scalar field, we get from the Einstein equations:

$$\begin{aligned} H^2 &= \frac{8\pi G}{3} \left(\frac{1}{2} \dot{\phi}^2 + V(\phi) \right) \\ \dot{H} &= -4\pi G(\rho + p) = -4\pi G\dot{\phi}^2 \end{aligned} \quad (169)$$

In order to get an almost exponential phase of inflation with almost constant H , we need $p \simeq -\rho$, which according to (168), requires $\dot{\phi}^2 \ll V$ during a sufficiently long period. This can be achieved if the friction terms in (166) dominates (overdamped oscillator), this is the so called slow-roll regime. Thus the conditions for slow-roll inflation can be summarized as follows:

- $\ddot{\phi} \ll \{V', 3H\dot{\phi}\}$ which implies:

$$\dot{\phi} \simeq -\frac{V'}{3H} \quad (170)$$

- $\dot{\phi}^2 \ll V$, so that using (169) we can define the first slow-roll parameter as:

$$\varepsilon = -\frac{\dot{H}}{H^2} = \frac{4\pi G\dot{\phi}^2}{\frac{8\pi G}{3}V} = \frac{1}{16\pi G} \left(\frac{V'}{V}\right)^2 \ll 1 \quad (171)$$

which is nothing but the condition of almost constant expansion rate.

- Imposing that $\dot{\phi}$ is also slowly evolving, we get:

$$\delta = -\frac{\ddot{\phi}}{\dot{\phi}H} = \frac{V''}{8\pi GV} - \varepsilon \quad (172)$$

so that we define the second slow-roll parameter:

$$\eta = \delta + \varepsilon = \frac{V''}{8\pi GV} \ll 1 \quad (173)$$

i.e. we have two conditions on the potential $V(\phi)$ in order to get slow-roll inflation.

3.4.2 Example: $V(\phi) = \frac{1}{2}m^2\phi^2$

A mass term provides the simplest example of a slow-roll inflation model. The corresponding slow-roll parameters read:

$$\varepsilon = \eta = \frac{1}{4\pi G\phi^2} \ll 1 \quad (174)$$

which implies:

$$\phi \gg \frac{M_p}{2\sqrt{\pi}} \quad (175)$$

In this kind of models (large field models), the field has to be displaced from the origin in order for the slow-roll conditions to be satisfied. This is in fact the original chaotic inflation model [20], in which it is assumed that in the early universe there were domains in which ϕ was sufficiently homogeneous so that $|\nabla\phi| \ll |\dot{\phi}|$ with $\phi \sim M_p$. One of the domains in which these conditions are satisfied will inflate giving rise to our present universe, but other domains in which similar conditions are satisfied will also inflate giving rise to other exponentially large regions which could be causally disconnected from our patch (multiverse picture).

Apart from the large-field models, there are also other kind of proposals with symmetry breaking like potentials in which the slow-roll conditions are satisfied for small field values. For a review of particles physics models of inflation see [21].

It is possible to calculate the number of e-folds of inflation generated in these models during the slow-roll phase, thus:

$$\begin{aligned} N &= \ln \frac{a_f}{a_i} = \int_{a_i}^{a_f} \frac{da}{a} = \int_{t_i}^{t_f} \frac{\dot{a}}{a} dt = \int_{\phi_i}^{\phi_f} H(\phi) \frac{d\phi}{\dot{\phi}} \simeq -3 \int_{\phi_i}^{\phi_f} \frac{H^2}{V'} d\phi \\ &= -8\pi G \int_{\phi_i}^{\phi_f} \frac{V}{V'} d\phi \end{aligned} \quad (176)$$

where we have made use of the slow-roll conditions and ϕ_f satisfies the condition $\max\{\varepsilon(\phi_f), \eta(\phi_f)\} = 1$.

For the previous example $V(\phi) = \frac{1}{2}m^2\phi^2$, we get:

$$N = -8\pi G \int_{\phi_i}^{\phi_f} \frac{\phi}{2} d\phi = -\frac{2\pi}{M_p^2} (\phi_f^2 - \phi_i^2) = -\frac{2\pi}{M_p^2} \left(\frac{M_p^2}{4\pi} - \phi_i^2 \right) \quad (177)$$

If we set $N \sim 50$ as required to solve the flatness and horizon problems, we find the minimum initial value of the inflation field required:

$$\phi_{50} \simeq 2.8M_p \quad (178)$$

Notice that although this value is larger than M_p the corresponding energy density during inflation can be smaller than the Planck density M_p^4 and still a classical description of gravity would be correct during inflation.

4 Cosmological Perturbation Theory

Although the universe is statistically homogeneous on very large scales, the origin of structures on smaller scales is not explained by the standard cosmological model. We will show in this section that inflation, not only solves the horizon and flatness problems, but also provides a mechanism for the generation of fluctuations which grew by gravitational attraction giving rise to the large scale structures and CMB anisotropies that we observe today.

4.1 Perturbations of the Robertson–Walker Metric

In order to describe the first stages in the evolution of perturbations, we will consider small deviations around the RW metric and work in the linear regime [22]. When the density contrast approaches unity, the approximation will break down and different methods, such as N-body simulations will be required in order to study the evolution. Thus, let us consider the most general form of the perturbed RW metric:

$$ds^2 = g_{\mu\nu}^{(0)} dx^\mu dx^\nu + \delta g_{\mu\nu} dx^\mu dx^\nu \quad (179)$$

where for simplicity we will consider the background metric to be a flat RW metric in conformal time:

$$g_{\mu\nu}^{(0)} dx^\mu dx^\nu = a^2(\eta) (d\eta^2 - d\mathbf{x}^2) \quad (180)$$

The metric perturbation $\delta g_{\mu\nu}$ can be decomposed in three pieces according to their properties under rotations on the spatial sections. Thus we have:

$$\delta g_{\mu\nu} = \delta g_{\mu\nu}^{(S)} + \delta g_{\mu\nu}^{(V)} + \delta g_{\mu\nu}^{(T)} \quad (181)$$

where $\delta g_{\mu\nu}^{(S)}$ corresponds to the scalar perturbations which are responsible for density perturbations. $\delta g_{\mu\nu}^{(V)}$ are vector perturbations associated to vorticity perturbations in the fluid. This kind of perturbation decays in the absence of sources of vorticity and we will not consider them in the following. Finally, $\delta g_{\mu\nu}^{(T)}$ corresponds to gravity waves (tensor perturbations). In the linear regime, the evolution of each type of perturbations is independent. In the following we will restrict ourselves to the case of scalar perturbations.

The most general form of the scalarly perturbed flat RW metric can be written as:

$$ds^2 = a^2(\eta) \left((1 + 2\phi) d\eta^2 - 2B_{,i} dx^i d\eta - ((1 - 2\psi)\delta_{ij} + 2E_{,ij}) dx^i dx^j \right) \quad (182)$$

where comma denotes the ordinary derivative. Here $\phi(\eta, \mathbf{x})$, $B(\eta, \mathbf{x})$, $\psi(\eta, \mathbf{x})$ and $E(\eta, \mathbf{x})$ are four arbitrary scalar functions describing the scalar perturbation of the metric. However very much as in the case of gauge fields, not all these functions are physical. Thus, in the case of U(1) gauge fields it is always possible to perform a gauge transformation:

$$A_\mu \rightarrow \hat{A}_\mu = A_\mu + \partial_\mu \Lambda \quad (183)$$

such that A_μ and \hat{A}_μ represent the same physical electric and magnetic fields.

In the gravitational case, the infinitesimal coordinate transformations:

$$x^\mu \rightarrow \tilde{x}^\mu = x^\mu + \xi^\mu(\eta, \mathbf{x}) \quad (184)$$

imply:

$$\tilde{g}_{\mu\nu}(x) = g_{\mu\nu}(x) + \mathcal{L}_\xi g_{\mu\nu}(x) = g_{\mu\nu}(x) - \xi_{\mu;\nu} - \xi_{\nu;\mu} \quad (185)$$

where \mathcal{L}_ξ denotes the Lie derivative. Substituting for the perturbed metric, and assuming that ξ is also first order in perturbations, we get:

$$\tilde{g}_{\mu\nu}(x) = g_{\mu\nu}^{(0)}(x) + \delta g_{\mu\nu}(x) + \mathcal{L}_\xi g_{\mu\nu}^{(0)}(x) + \dots \quad (186)$$

so that we can obtain the transformation rule of the metric perturbation under general coordinate transformation as:

$$\delta\tilde{g}_{\mu\nu}(x) = \delta g_{\mu\nu}(x) + \mathcal{L}_\xi g_{\mu\nu}^{(0)}(x) \quad (187)$$

which is analogous to the gauge transformation showed before. We apply this rule to the perturbed metric (182) so that we can obtain the transformation rules of the scalar potentials:

$$\begin{aligned} \tilde{\phi} &= \phi - \mathcal{H}\xi^0 - \xi^{0'} \\ \tilde{\psi} &= \psi + \mathcal{H}\xi^0 \\ \tilde{B} &= B + \xi^0 - \xi' \\ \tilde{E} &= E - \xi \end{aligned} \quad (188)$$

where $\mathcal{H} = a'/a$, a prime denotes derivative with respect to the conformal time and we have written $\xi^\mu = (\xi^0, \xi^i)$. Notice that since we are considering only scalar perturbations, we will only take into account the scalar part of the transformation parameter, i.e. $\xi^i = \partial^i \xi$, where indices are raised and lowered using the δ_{ij} metric on the spatial sections. Thus, we see that the four scalar potential are not gauge invariant, but transform with two arbitrary functions ξ^0 and ξ . However, as in the electromagnetic case, it is possible to define two independent combinations which are indeed gauge invariant:

$$\begin{aligned} \Phi &= \phi + \frac{1}{a}((B - E')a)' \\ \Psi &= \psi - \mathcal{H}(B - E') \end{aligned} \quad (189)$$

which are known as Bardeen potentials [23]. Similarly to the choice of gauge in electromagnetism, which allows us to fix one of the components of A_μ using the freedom in the function Λ , it is possible in this case to fix two of the scalar potentials using the two arbitrary functions ξ^0 and ξ . Thus, in the following we will work in the so called longitudinal gauge in which $B = E = 0$. In this gauge, the Bardeen potentials are simply $\Phi = \phi$ and $\Psi = \psi$, so that we can write the perturbed metric as:

$$ds^2 = a^2(\eta) \left((1 + 2\Phi)d\eta^2 - (1 - 2\Psi)\delta_{ij}dx^i dx^j \right) \quad (190)$$

4.2 Evolution of Cosmological Perturbations

In order to determine the evolution of the metric perturbations we will solve Einstein equations for the perturbed metric:

$$G^\mu_\nu = 8\pi G T^\mu_\nu \quad (191)$$

Expanding to first order in perturbations we can write:

$$\begin{aligned} G^\mu_\nu &= G^\mu_\nu^{(0)} + \delta G^\mu_\nu \\ T^\mu_\nu &= T^\mu_\nu^{(0)} + \delta T^\mu_\nu \end{aligned} \quad (192)$$

So that, the zero order equations give rise to the (53) and (54) already studied, whereas to first order we get:

$$\delta G^\mu_\nu = 8\pi G \delta T^\mu_\nu \quad (193)$$

Notice that the perturbed Einstein tensor can be computed straightforwardly from (190). The non-vanishing components being:

$$\begin{aligned} \delta G_0^0 &= \frac{2}{a^2} (\nabla^2 \Psi - 3\mathcal{H}(\Psi' + \mathcal{H}\Phi)), \\ \delta G_j^i &= \frac{1}{a^2} ([-2\Psi'' - 2(\mathcal{H}^2 + 2\mathcal{H}')\Phi - 2\mathcal{H}\Phi' - 4\mathcal{H}\Psi' \\ &\quad - \nabla^2(\Phi - \Psi)]\delta_i^j - \partial_i \partial^j(\Psi - \Phi)), \\ \delta G_0^i &= -\frac{2}{a^2} \partial^i(\Psi' + \mathcal{H}\Phi) \end{aligned} \quad (194)$$

Thus we only need to determine the perturbed energy-momentum tensor. We will concentrate in those cases in which the energy content can be described by a perfect fluid.

4.2.1 Perturbations in a Perfect Fluid

Let us consider the general form of the energy-momentum tensor of a perfect fluid:

$$T^\mu_\nu = (\rho + p)u^\mu u_\nu - p\delta^\mu_\nu \quad (195)$$

The homogeneous (unperturbed) part will be written as:

$$T_v^\mu{}^{(0)} = (\rho^{(0)} + p^{(0)}) u^\mu{}^{(0)} u_v{}^{(0)} - p^{(0)} \delta_v^\mu \quad (196)$$

where in conformal time:

$$\begin{aligned} \rho^{(0)} &= \rho^{(0)}(\eta) \\ p^{(0)} &= p^{(0)}(\eta) \\ u^\mu{}^{(0)} &= (a^{-1}, 0, 0, 0) \end{aligned} \quad (197)$$

To first order we find:

$$\begin{aligned} \delta T_0^0 &= \delta\rho \\ \delta T_j^i &= -\delta p \delta_j^i \\ \delta T_0^i &= (\rho^{(0)} + p^{(0)}) v^i \end{aligned} \quad (198)$$

where $\delta\rho$ and δp are the energy density and pressure perturbations respectively and v is the velocity perturbation potential such that $v^i = \partial^i v$. Substituting in the first order Einstein equations (193), we get from the $i \neq j$ equation:

$$\Phi = \Psi \quad (199)$$

i.e. there is only one independent scalar perturbation. Moreover, if we assume that the perturbations in the fluid are adiabatic, i.e. there are no entropy perturbations, it is possible to set:

$$\delta p = c_s^2 \delta\rho \quad (200)$$

where c_s is the speed of sound of perturbations which for a single component fluid is nothing but $c_s^2 = \omega$, with $p^{(0)} = \omega\rho^{(0)}$. Using the combination of the perturbed equations $c_s^2(00) - (ii)$ and (200) it is straightforward to obtain the evolution equation for the scalar perturbation:

$$\Phi'' + 3\mathcal{H}(1 + c_s^2)\Phi' - c_s^2 \nabla^2 \Phi = 0 \quad (201)$$

After obtaining Φ from this equation, it is possible to derive $\delta\rho$ from the (00) Einstein equation and v from the $(0i)$, so that all the perturbed quantities can be determined.

Since the coefficients in (201) only depend on time it is useful to Fourier transform the spatial dependence:

$$\Phi(\eta, \mathbf{x}) = \frac{1}{(2\pi)^{3/2}} \int d^3k e^{i\mathbf{k}\mathbf{x}} \Phi_k(\eta) \quad (202)$$

so that (201) becomes:

$$\Phi_k'' + 3\mathcal{H}(1 + c_s^2)\Phi_k' + c_s^2 k^2 \Phi_k = 0 \quad (203)$$

Notice that due to the isotropy of the background geometry the Fourier modes $\Phi_k(\eta)$ only depend on $k = |\mathbf{k}|$.

Equation (203) exhibits two different regimes in which it is possible to obtain explicit solutions, namely:

$$\begin{aligned} k \gg \mathcal{H}, (k\eta \gg 1) & \text{ sub-Hubble modes} \\ k \ll \mathcal{H}, (k\eta \ll 1) & \text{ super-Hubble modes} \end{aligned} \quad (204)$$

Notice that sub(super)-Hubble modes correspond to those modes whose physical wavelength is smaller(larger) than the Hubble radius, i.e. $2\pi a/k \ll H^{-1}$ for sub-Hubble modes and $2\pi a/k \gg H^{-1}$ for super-Hubble ones. This means that for sub-Hubble modes the curvature of space-time plays a negligible role in the evolution, and space-time can be considered as effectively Minkowskian.

4.2.2 Solutions

- super-Hubble modes

$$\begin{aligned} \Phi_k &= \text{const.} \\ \frac{\delta\rho_k}{\rho^{(0)}} &= -2\Phi_k = \text{const.} \end{aligned} \quad (205)$$

- sub-Hubble modes

$$\Phi_k \propto \begin{cases} \text{const. matter era } (c_s^2 = 0) \\ \frac{1}{a^2} e^{ikc_s\eta} \text{ radiation era } (c_s^2 = 1/3) \end{cases} \quad (206)$$

$$\frac{\delta\rho_k}{\rho^{(0)}} \propto \begin{cases} a \text{ matter era } (c_s^2 = 0) \\ e^{ikc_s\eta} \text{ radiation era } (c_s^2 = 1/3) \end{cases} \quad (207)$$

Thus, we see that in the matter dominated era, the evolution of the Φ_k potential is the same for sub- or super-Hubble modes. On the other hand, the density contrast can only start growing in the matter era once the mode becomes sub-Hubble. In the radiation era, density perturbations only oscillate but do not grow and this behaviour is also followed by baryons which are strongly coupled to radiation until matter-radiation decoupling ($a_{dec} \simeq 10^{-3}$). After decoupling, the baryonic density contrast

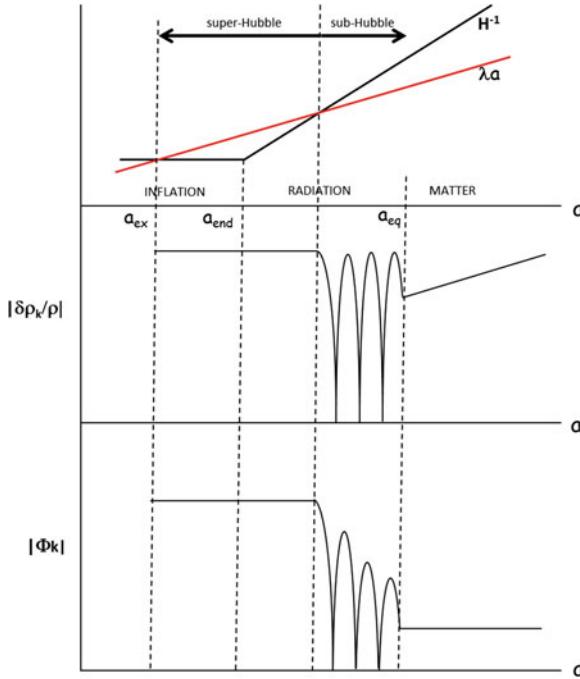


Fig. 11 *Upper panel* shows the evolution of the physical size of a given comoving scale $\lambda = 2\pi/k$ which exits the Hubble horizon H^{-1} during inflation reentering in the radiation dominated era. *Middle panel* shows the evolution of the density contrast which remains constant while the scale is super-Hubble, oscillates when the scale reenters the horizon in the radiation era and starts to grow in the matter era. *Lower panel* shows the evolution of the scalar perturbation potential which is also constant in the super-Hubble regime and oscillates in the radiation era becoming constant again in the matter era

could start growing as $\delta\rho/\rho \propto a$. This means that since today $a_0 = 1$, the maximum growth in the density contrast of baryons from decoupling till present would be a factor $a_0/a_{dec} \sim 10^3$. However, CMB anisotropies shows that at decoupling time (Fig. 11):

$$\left. \frac{\delta\rho}{\rho^{(0)}} \right|_{dec} \simeq \frac{\delta T}{T_0} \simeq 10^{-5} \quad (208)$$

This means that in an universe containing only baryons and radiation, the maximum density contrast today would be of order:

$$\left. \frac{\delta\rho}{\rho^{(0)}} \right|_0 \simeq 10^3 \left. \frac{\delta\rho}{\rho^{(0)}} \right|_{dec} \simeq 10^{-2} \quad (209)$$

i.e. structures would not have reached the non-linear regime and therefore no galaxies or clusters could have been formed. A possible solution to this problem is also provided by the existence of dark matter, which could have been decoupled from radiation before baryons did and therefore the cold dark matter density contrast could have started to grow earlier. After decoupling, baryons would fall in the potential wells created by dark matter overdensities.

4.3 The Origin of Perturbations

So far we have only considered the evolution of perturbations in the different epochs, in this section we will focus on their generation mechanism and how to determine their primordial amplitude.

Inflation provides a mechanism for the generation of perturbations from quantum fluctuations of the fields Φ and $\delta\phi$ (the fluctuation of the inflaton field). Let us then consider the full action of the theory, including the Einstein-Hilbert term for the gravitational field and the inflaton term:

$$S_{tot} = S_{grav} + S_{inflation} = \int d^4x \sqrt{g} \left(-\frac{R}{16\pi G} + \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right) \quad (210)$$

substituting:

$$\begin{aligned} g_{\mu\nu} &= g_{\mu\nu}^{(0)} + \delta g_{\mu\nu} \\ \phi &= \phi_0(\eta) + \delta\phi(\eta, \mathbf{x}) \end{aligned} \quad (211)$$

it can be seen that the action for perturbations depends on a single field:

$$u = a\delta\phi + z\Phi \quad (212)$$

which is called the Mukhanov–Sasaki variable and where:

$$z = a \frac{\phi'_0}{\mathcal{H}} \quad (213)$$

On super-Hubble scales, it is easy to show that:

$$\frac{u_k}{z} = \frac{\Phi_k}{\varepsilon} = R_k \simeq const. \quad (214)$$

with ε the slow-roll parameter defined in (171) and R_k is the perturbation of the curvature of the spatial hypersurfaces.

In terms of the u field, the quadratic part of the total action reads:

$$\delta S_{tot}^{(2)} = \int d^4x \left(\frac{1}{2} \eta^{\mu\nu} \partial_\mu u \partial_\nu u + \frac{1}{2} \frac{z''}{z} u^2 \right) \quad (215)$$

which is nothing but the action for a scalar field in Minkowski space-time but with a time-dependent mass. We can proceed with the canonical quantization of this theory. The Fourier modes of the field satisfy:

$$u_k'' + k^2 u_k - \frac{z''}{z} u_k = 0 \quad (216)$$

Thus we can expand in creation and annihilation operators as:

$$u(\eta, \mathbf{x}) = \sum_{\mathbf{k}} \left(a_{\mathbf{k}} u_{\mathbf{k}}(\eta) e^{i\mathbf{k}\mathbf{x}} + a_{\mathbf{k}}^\dagger u_{\mathbf{k}}^*(\eta) e^{-i\mathbf{k}\mathbf{x}} \right) \quad (217)$$

with:

$$[a_{\mathbf{k}}, a_{\mathbf{k}'}^\dagger] = \delta_{\mathbf{k}\mathbf{k}'} \quad (218)$$

The $u_k(\eta)$ modes appearing in this expression are not uniquely defined and depend on the boundary conditions imposed on the field. However, there is a way to define them provided we impose that in the short-wavelength limit $k^2 \gg z''/z$ the modes behave as the standard Minkowski modes, i.e.:

$$u_k(\eta) \xrightarrow{k^2 \rightarrow \infty} \frac{e^{-ik\eta}}{\sqrt{2kV}} \quad (219)$$

with V the volume of the box which will be taken to infinity. This particular choice of modes is equivalent to a definition of vacuum state:

$$a_{\mathbf{k}}|0\rangle = 0, \quad \forall \mathbf{k} \quad (220)$$

which is usually referred to as the Bunch-Davies vacuum.

Imposing this asymptotic behaviour and using the slow-roll evolution of the inflation field ϕ_0 during inflation, it is possible to determine z and obtain the exact solution to (216):

$$u_k(\eta) = \frac{\sqrt{\pi}}{2\sqrt{V}} e^{i(v+\frac{1}{2})\frac{\pi}{2}} (-\eta)^{1/2} H_v^{(1)}(-k\eta) \quad (221)$$

where $H_v^{(1)}$ denote the Hankel functions of the first kind with:

$$v = \frac{3}{2} + 2\varepsilon - \delta \quad (222)$$

with ε and δ the slow-roll parameters defined before. Taking the super-Hubble limit $|k\eta| \ll 1$ we get:

$$|u_k| = \frac{1}{\sqrt{2kV}} \left(\frac{k}{aH} \right)^{1/2-\nu} \quad (223)$$

We can finally compute the amplitude of the quantum fluctuations generated by this mechanism from the field dispersion. Notice that in the vacuum state, using (217) we have $\langle 0|u|0\rangle = 0$, but:

$$\begin{aligned} \langle 0|u^2|0\rangle &= \sum_{\mathbf{k}} \sum_{\mathbf{k}'} u_{\mathbf{k}}^* u_{\mathbf{k}'} \langle 0|a_{\mathbf{k}'} a_{\mathbf{k}}^\dagger|0\rangle = \sum_{\mathbf{k}} |u_{\mathbf{k}}|^2 \xrightarrow{\nu \rightarrow \infty} \int \frac{d^3 k}{(2\pi)^3} |u_{\mathbf{k}}|^2 \\ &= 4\pi \int \frac{dk}{(2\pi)^3} k^2 |u_{\mathbf{k}}|^2 = \int \frac{dk}{k} P(k) \end{aligned} \quad (224)$$

where we have used the commutation relations in (218) and the fact that u_k only depends on $k = |\mathbf{k}|$ so that we can perform the integration in the angular variables in a trivial way. The power spectrum $P(k)$ is thus defined as:

$$P(k) = \frac{k^3}{2\pi^2} |u_{\mathbf{k}}|^2 \quad (225)$$

and provides the amplitude of the fluctuation at a given scale k . It is usual to give the result in terms of the curvature perturbation $R_k = u_k/z$, so that:

$$P_R(k) = \frac{k^3}{2\pi^2} \frac{|u_{\mathbf{k}}|^2}{z^2} = \frac{k^2}{4\pi^2 z^2} \left(\frac{k}{aH} \right)^{1-2\nu} \quad (226)$$

and using the expression for z it can be rewritten as:

$$P_R(k) = A_S \left(\frac{k}{aH} \right)^{n_s - 1} \quad (227)$$

where A_S is the amplitude of the scalar perturbations:

$$A_S = \frac{4\pi G}{\varepsilon} \left(\frac{H}{2\pi} \right)^2 \Big|_{k_0=aH} = \frac{128\pi}{3} \frac{V^3}{M_P^6 V'^2} \Big|_{\phi_{50}} \quad (228)$$

where the amplitude is evaluated when the scale $k_0 = 0.05 \text{ Mpc}^{-1}$ (Planck pivot scale) crossed the Hubble horizon during inflation, which as shown before corresponds roughly to 50 e-folds before the end of inflation. On the other hand n_s is the scalar spectral index which measures the dependence of the fluctuations on the scale and is given by:

$$n_s = 1 - 4\varepsilon + 2\delta \quad (229)$$

which as we see must be very close to $n_s = 1$ (Harrison-Zeldovich spectrum) in slow-roll inflation.

As shown in (214), the curvature perturbation is related to the gravitational potential Φ which, in turn, is related to the density contrast. This implies that photons propagating on the perturbed background acquire perturbations in their energies: on one hand because of the intrinsic density fluctuations on the LSS and secondly because of the gravitational Doppler effect produced when a photon moves across a metric perturbations. This is the basis of the Sachs-Wolfe effect [7] which allows to relate the observed temperature anisotropies of the CMB to the primordial curvature perturbations spectrum. Thus, from Planck observations [9], it is possible to obtain:

$$\begin{aligned} A_S &= (2.23 \pm 0.16) \times 10^{-9} \\ n_s &= 0.9616 \pm 0.0094 \end{aligned} \quad (230)$$

Notice that as expected the spectral index is closed to $n_s = 1$. However Planck data excludes exact scale invariance at over 4σ .

A similar analysis can be performed for tensor perturbations. It is possible to define a tensor power spectrum as:

$$P_T(k) = A_T \left(\frac{k}{aH} \right)^{n_T} \quad (231)$$

where the amplitude A_T and the tensor spectral index n_T are given by:

$$\begin{aligned} A_T &= \frac{128V}{3M_p^4} \Big|_{\phi_{50}} \\ n_T &= -2\varepsilon \end{aligned} \quad (232)$$

In single-field models of inflation, the tensor to scalar ratio, defined as:

$$r = \frac{A_T}{A_S} \quad (233)$$

satisfies the so called consistency condition:

$$r = -8n_T \quad (234)$$

Non-observation of gravity waves by Planck satellite has allowed to set an upper limit on r [24]:

$$r < 0.12, \text{ } 95\% \text{ } C.L. \quad (235)$$

at the pivot scale $k_* = 0.002 \text{ Mpc}^{-1}$.

As an example, we can now calculate what are the predictions of the simplest inflationary models for the power spectrum parameters. Let us then consider an inflation potential $V(\phi) = \frac{1}{2}m^2\phi^2$. As shown in Sect. 3.4.2, the values of the slow parameters for this particular model, given in (174), imply $\delta = 0$. On the other hand, for the value of $\phi_{50} = 2.8M_p$, obtained in (178), we get:

$$\varepsilon = 0.01 \quad (236)$$

so that:

$$n_s = 1 - 4\varepsilon = 0.96 \quad (237)$$

in very good agreement with observations. On the other hand, using the expression for the amplitude, we can fix the mass of the inflation field. In this particular case:

$$A_S = \frac{128\pi}{3} \frac{V^3}{M_p^6 V'^2} \Big|_{\phi_{50}} = \frac{128\pi}{24} \frac{m^2 \phi_{50}^4}{M_p^6} \quad (238)$$

so that using the Planck results (230), we get for the inflation mass:

$$m = 1.8 \times 10^{13} \text{ GeV} \quad (239)$$

and the corresponding scale of inflation can be estimated as:

$$V^{1/4} = \left(\frac{1}{2} m^2 \phi_{50}^2 \right)^{1/4} \simeq 2 \times 10^{16} \text{ GeV} \quad (240)$$

The high precision of Planck measurements has started to shrink the space for the allowed forms of the inflation potential [24].

Acknowledgments I would like to thank the organizers of the Third IDPASC school and in particular Carlos Merino for their kind hospitality and the excellent work in the preparation of the school. This work was supported by the Spanish MINECO projects numbers FIS2011-23000 and MULTIDARK CSD2009-00064 (Consolider-Ingenio 2010 Programme).

References

1. E.W. Kolb, M.S. Turner, *The Early Universe* (Addison-Wesley, New York, 1990)
2. S. Weinberg, *Cosmology* (Oxford University Press, Oxford, 2008)
3. S. Dodelson, *Modern Cosmology* (Academic Press, Amsterdam, 2003)
4. V.F. Mukhanov, *Physical Foundations of Cosmology* (Cambridge University Press, Cambridge, 2005)
5. M. Tegmark et al., SDSS collaboration. Phys. Rev. D **74**, 123507 (2006). [[astro-ph/0608632](#)]

6. A.G. Riess, L. Macri, S. Casertano, H. Lampeitl, H.C. Ferguson, A.V. Filippenko, S.W. Jha, W. Li et al., *Astrophys. J.* **730**, 119 (2011) [Erratum-*ibid.* **732**, 129 (2011)] [arXiv:1103.2976](https://arxiv.org/abs/1103.2976) [astro-ph.CO]
7. R. Durrer, *The Cosmic Microwave Background* (Cambridge University Press, Cambridge, 2008)
8. G. Hinshaw et al., [WMAP Collaboration], [arXiv:1212.5226](https://arxiv.org/abs/1212.5226) [astro-ph.CO]
9. P.A.R. Ade et al., [Planck Collaboration], [arXiv:1303.5076](https://arxiv.org/abs/1303.5076) [astro-ph.CO]
10. P.A.R. Ade et al., [Planck Collaboration], [arXiv:1303.5075](https://arxiv.org/abs/1303.5075) [astro-ph.CO]
11. S. Perlmutter et al., [Supernova Cosmology Project Collaboration], *Astrophys. J.* **517**, 565 (1999) [astro-ph/9812133]
12. A.G. Riess et al., [Supernova Search Team Collaboration]. *Astron. J.* **116**, 1009 (1998) [astro-ph/9805201]
13. M. Kowalski et al., [Supernova Cosmology Project Collaboration]. *Astrophys. J.* **686**, 749 (2008). [arXiv:0804.4142](https://arxiv.org/abs/0804.4142) [astro-ph]
14. D.J. Eisenstein et al., [SDSS Collaboration], *Astrophys. J.* **633**, 560 (2005) [astro-ph/0501171]
15. S. Cole et al., [2dFGRS Collaboration], *Mon. Not. Roy. Astron. Soc.* **362**, 505 (2005) [astro-ph/0501174]
16. F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders et al., *Mon. Not. Roy. Astron. Soc.* **416**, 3017 (2011). [arXiv:1106.3366](https://arxiv.org/abs/1106.3366) [astro-ph.CO]
17. C. Blake, E. Kazin, F. Beutler, T. Davis, D. Parkinson, S. Brough, M. Colless, C. Contreras et al., *Mon. Not. Roy. Astron. Soc.* **418**, 1707 (2011). [arXiv:1108.2635](https://arxiv.org/abs/1108.2635) [astro-ph.CO]
18. A.R. Liddle, D.H. Lyth, *Cosmological Inflation and Large-Scale Structure* (Cambridge University Press, Cambridge, 2000)
19. A. Guth, *Phys. Rev. D* **23**, 347 (1981)
20. A. Linde, *Phys. Lett. B* **108**, 389 (1982)
21. D.H. Lyth, A. Riotto, *Phys. Rept.* **314**, 1 (1999). [hep-ph/9807278]
22. V.F. Mukhanov, H.A. Feldman, R.H. Brandenberger, *Phys. Rept.* **215**, 203 (1992)
23. J.M. Bardeen, *Phys. Rev. D* **22**, 1882 (1980)
24. P.A.R. Ade et al. Planck collaboration]. [arXiv:1303.5082](https://arxiv.org/abs/1303.5082) [astro-ph.CO]

Introduction to the AdS/CFT Correspondence

Alfonso V. Ramallo

Abstract This is a pedagogical introduction to the AdS/CFT correspondence. Starting with the conceptual basis of the holographic dualities, the subject is developed emphasizing some concrete topics, which are discussed in detail. A very brief introduction to string theory is provided, containing the minimal ingredients to understand the origin of the AdS/CFT duality. Other topics covered are the holographic calculation of correlation functions, quark–antiquark potentials and transport coefficients.

1 Introduction and Motivation

The AdS/CFT correspondence is a duality relating quantum field theory (QFT) and gravity. More precisely, the correspondence relates the quantum physics of strongly correlated many-body systems to the classical dynamics of gravity in one higher dimension. This duality is also referred to as the holographic duality or the gauge/gravity correspondence. In its original formulation [1–3], the correspondence related a four-dimensional Conformal Field Theory (CFT) to the geometry of an anti-de Sitter (AdS) space in five dimensions.

In the study of collective phenomena in condensed matter physics it is quite common that when a system is strongly coupled it reorganizes itself in such a way that new weakly coupled degrees of freedom emerge dynamically and the system can be better described in terms of fields representing the emergent excitations. The holographic duality is a new example of this paradigm. The new (and surprising!) feature is that the emergent fields live in a space with one extra dimension and that the dual theory is a gravity theory. As we will argue below, the extra dimension is related to the energy scale of the QFT. The holographic description is a geometrization of the quantum dynamics of the systems with a large number of degrees of freedom,

A.V. Ramallo (✉)

Departamento de Física de Partículas, Universidade de Santiago de Compostela and Instituto Galego de Física de Altas Enerxías (IGFAE), E-15782 Santiago de Compostela, Spain
e-mail: alfonso@fpaxp1.usc.es

which makes manifest that there are deep connections between quantum mechanics and gravity.

The gauge/gravity duality was discovered in the context of string theory, where it is quite natural to realize (gauge) field theories on hypersurfaces embedded in a higher dimensional space, in a theory containing gravity. However, the study of the correspondence has been extended to include very different domains, such as the analysis of the strong coupling dynamics of QCD and the electroweak theories, the physics of black holes and quantum gravity, relativistic hydrodynamics or different applications in condensed matter physics (holographic superconductors, quantum phase transitions, cold atoms, . . .). In these lectures we will concentrate on some particular topics, trying to present in clear terms the basic conceptual ideas, as well as the more practical calculational aspects of the subject. For reviews on the different aspects of the duality see [4–10].

We will start by motivating the duality from the Kadanoff-Wilson renormalization group approach to the analysis of lattice systems. Let us consider a non-gravitational system in a lattice with lattice spacing a and hamiltonian given by:

$$H = \sum_{x,i} J_i(x, a) \mathcal{O}^i(x), \quad (1)$$

where x denotes the different lattice sites and i labels the different operators \mathcal{O}^i . The $J_i(x, a)$ are the coupling constants (or sources) of the operators at the point x of the lattice. Notice that we have included a second argument in J^i , to make clear they correspond to a lattice spacing a . In the renormalization group approach we coarse grain the lattice by increasing the lattice spacing and by replacing multiple sites by a single site with the average value of the lattice variables. In this process the hamiltonian retains its form (1) but different operators are weighed differently. Accordingly, the couplings $J_i(x, a)$ change in each step. Suppose that we double the lattice spacing in each step. Then, we would have a succession of couplings of the type:

$$J_i(x, a) \rightarrow J_i(x, 2a) \rightarrow J_i(x, 4a) \rightarrow \dots \quad (2)$$

Therefore, the couplings acquire in this process a dependence on the scale (the lattice spacing) and we can write them as $J_i(x, u)$, where $u = (a, 2a, 4a, \dots)$ is the length scale at which we probe the system. The evolution of the couplings with the scale is determined by flow equations of the form:

$$u \frac{\partial}{\partial u} J_i(x, u) = \beta_i(J_j(x, u), u), \quad (3)$$

where β_i is the so-called β -function of the i th coupling constant. At weak coupling the β_i 's can be determined in perturbation theory. At strong coupling the AdS/CFT proposal is to consider u as an extra dimension. In this picture the succession of lattices at different values of u are considered as layers of a new higher-dimensional

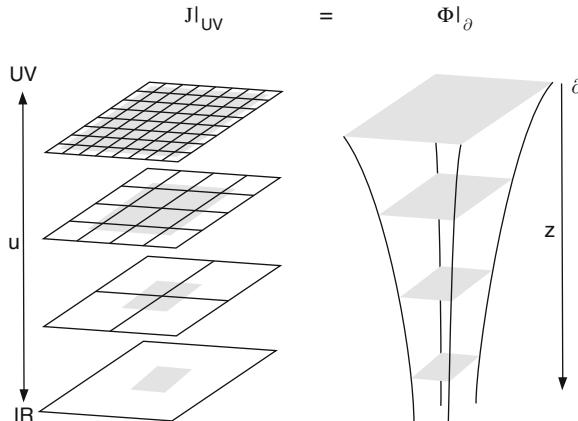


Fig. 1 On the *left* we illustrate the Kadanoff-Wilson renormalization of a lattice system. In the AdS/CFT correspondence the lattices at different scale are considered as the layers of the higher dimensional space represented on the *right* of the figure

space. Moreover, the sources $J_i(x, u)$ are regarded as fields in a space with one extra dimension and, accordingly we will simply write:

$$J_i(x, u) = \phi_i(x, u). \quad (4)$$

The dynamics of the sources ϕ_i 's will be governed by some action. Actually, in the AdS/CFT duality the dynamics of the ϕ_i 's is determined by some gravity theory (i.e. by some metric). Therefore, one can consider the holographic duality as a geometrization of the quantum dynamics encoded by the renormalization group. The microscopic couplings of the field theory in the UV can be identified with the values of the bulk fields at the boundary of the extra-dimensional space. Thus, one can say that the field theory lives on the boundary of the higher-dimensional space (see Fig. 1).

The sources ϕ_i of the dual gravity theory must have the same tensor structure of the corresponding dual operator \mathcal{O}^i of field theory, in such a way that the product $\phi_i \mathcal{O}^i$ is a scalar. Therefore, a scalar field will be dual to a scalar operator, a vector field A_μ will be dual to a current J^μ , whereas a spin-two field $g_{\mu\nu}$ will be dual to a symmetric second-order tensor $T_{\mu\nu}$ which can be naturally identified with the energy-momentum tensor $T_{\mu\nu}$ of the field theory.

The holographic duality raises several conceptual issues which should be addressed in order to fully understand it. The simplest of these issues is the matching of the degrees of freedom on both sides of the correspondence. Let us consider a QFT in a d -dimensional spacetime (i.e. in $d - 1$ spatial dimensions plus time). The number of degrees of freedom of a system is measured by the entropy. On the QFT side the entropy is an extensive quantity. Therefore, if R_{d-1} is $(d - 1)$ -dimensional spatial region, at constant time, its entropy should be proportional to its volume in $d - 1$ dimensions:

$$S_{QFT} \propto \text{Vol}(R_{d-1}) . \quad (5)$$

On the gravity side the theory lives in a $(d + 1)$ -dimensional spacetime. How such a higher dimensional theory can contain the same information as its lower dimensional dual? The crucial point to answer this question is the fact that the entropy in quantum gravity is subextensive. Indeed, in a gravitational theory the entropy in a volume is bounded by the entropy of a black hole that fits inside the volume and, according to the so-called holographic principle, the entropy is proportional to the surface of the black hole horizon (and not to the volume enclosed by the horizon). More concretely, the black hole entropy is given by the Bekenstein-Hawking formula:

$$S_{BH} = \frac{1}{4G_N} A_H , \quad (6)$$

where A_H is the area of the event horizon and G_N is the Newton constant. In order to apply (6) for our purposes, let R_d be a spatial region in the $(d + 1)$ -dimension spacetime where the gravity theory lives and let us assume that R_d is bounded by a $(d - 1)$ -dimensional manifold R_{d-1} ($R_{d-1} = \partial R_d$). Then, according to (6), the gravitational entropy associated to R_d scales as:

$$S_{GR}(R_d) \propto \text{Area}(\partial R_d) \propto \text{Vol}(R_{d-1}) , \quad (7)$$

which agrees with the QFT behavior (5). In this lectures we will present several refinements of this argument and we will establish a more precise matching of the degrees of freedom of the two dual theories.

2 The Anti-de Sitter Space

In general, finding the geometry associated to a given QFT is a very difficult problem. However, if the theory is at a fixed point of the renormalization group flow (with a vanishing β -function) it has conformal invariance (is a CFT) and one can easily find such a metric. Indeed, let us consider a QFT in d spacetime dimensions. The most general metric in $(d + 1)$ -dimensions with Poincaré invariance in d -dimensions is:

$$ds^2 = \Omega^2(z) (-dt^2 + d\mathbf{x}^2 + dz^2) , \quad (8)$$

where z is the coordinate of the extra dimension, $\mathbf{x} = (x^1, \dots, x^{d-1})$ and $\Omega(z)$ is a function to be determined. If z represents a length scale and the theory is conformal invariant, then ds^2 must be invariant under the transformation:

$$(t, \mathbf{x}) \rightarrow \lambda(t, \mathbf{x}) , \quad z \rightarrow \lambda z . \quad (9)$$

Then, by imposing the invariance of the metric (8) under the transformation (9), we obtain that the function $\Omega(z)$ must transform as:

$$\Omega(z) \rightarrow \lambda^{-1} \Omega(z), \quad (10)$$

which fixes $\Omega(z)$ to be:

$$\Omega(z) = \frac{L}{z}, \quad (11)$$

where L is a constant. Thus, by plugging the function (11) into the metric (8) we arrive at the following form of ds^2 :

$$ds^2 = \frac{L^2}{z^2} (-dt^2 + d\mathbf{x}^2 + dz^2), \quad (12)$$

which is the line element of the *AdS* space in $(d+1)$ -dimensions, that we will denote by AdS_{d+1} . The constant L is a global factor in (12), which we will refer to as the anti-de Sitter radius. The (conformal) boundary of the *AdS* space is located at $z = 0$. Notice that the metric (12) is singular at $z = 0$. This means that we will have to introduce a regularization procedure in order to define quantities in the *AdS* boundary.

The *AdS* metric (12) is a solution of the equation of motion of a gravity action of the type:

$$I = \frac{1}{16\pi G_N} \int d^{d+1}x \sqrt{-g} [-2\Lambda + R + c_2 R^2 + c_3 R^3 + \dots], \quad (13)$$

with G_N is the Newton constant, the c_i are constants, $g = \det(g_{\mu\nu})$, $R = g^{\mu\nu} R_{\mu\nu}$ is the scalar curvature and Λ is a cosmological constant. In particular, if $c_2 = c_3 = \dots = 0$ the action (13) becomes the Einstein-Hilbert (EH) action of general relativity with a cosmological constant. In this case the equations of motion are just the Einstein equations:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = -\Lambda g_{\mu\nu}. \quad (14)$$

Taking the trace on both sides of (14), we get that the scalar curvature is given by:

$$R = g^{\mu\nu} R_{\mu\nu} = 2 \frac{d+1}{d-1} \Lambda. \quad (15)$$

Inserting back this result in the Einstein equation (14) we get that the Ricci tensor and the metric are proportional:

$$R_{\mu\nu} = \frac{2}{d-1} \Lambda g_{\mu\nu}. \quad (16)$$

Therefore the solution of (14) defines an Einstein space. Moreover, the Ricci tensor for the metric (12) can be computed directly from its definition. We get:

$$R_{\mu\nu} = -\frac{d}{L^2} g_{\mu\nu}. \quad (17)$$

By comparing (16) and (17) we get that the AdS_{d+1} space solves the equations of motion (14) of EH gravity with a cosmological constant equal to:

$$\Lambda = -\frac{d(d-1)}{2L^2}, \quad (18)$$

which is negative. It follows from (15) and (18) that the scalar curvature for the AdS_{d+1} space with radius L is given by:

$$R = -\frac{d(d+1)}{L^2}. \quad (19)$$

In these lectures we will be mostly interested in gauge theories in $3 + 1$ dimensions, which corresponds to taking $d = 4$ in our formulas. The dual geometry found above for this case is AdS_5 . This was precisely the system studied in [1] by Maldacena, who conjectured that the dual QFT is super Yang-Mills theory with four supersymmetries ($\mathcal{N} = 4$ SYM).

2.1 Counting the Degrees of Freedom in AdS

After having identified the AdS space as the gravity dual of a field theory with conformal invariance, we can refine the argument of Sect. 1 to match the number of degrees of freedom of both sides of the duality.

Let us consider first the QFT side. To regulate the theory we put both a UV and IR regulator. We place the system in a spatial box of size R (which serves as an IR cutoff) and we introduce a lattice spacing ε that acts as a UV regulator. In d spacetime dimensions the system has $R^{d-1}/\varepsilon^{d-1}$ cells. Let c_{QFT} be the number of degrees of freedom per lattice site, which we will refer to as the central charge. Then, the total number of degrees of freedom of the QFT is:

$$N_{dof}^{QFT} = \left(\frac{R}{\varepsilon}\right)^{d-1} c_{QFT}. \quad (20)$$

The central charge is one of the main quantities that characterize a CFT. If the CFT is a $SU(N)$ gauge field theory, such as the theory with four supersymmetries which will be described below, the fields are $N \times N$ matrices in the adjoint representation which, for large N , contain N^2 independent components. Thus, in these $SU(N)$ CFT's the central charge scales as $c_{SU(N)} \sim N^2$.

Let us now compute the number of degrees of freedom of the AdS_{d+1} solution. According to the holographic principle and to the Bekenstein-Hawking formula (6), the number of degrees of freedom contained in a certain region is equal to the maximum entropy, given by

$$N_{dof}^{AdS} = \frac{A_\partial}{4G_N}, \quad (21)$$

with A_∂ being the area of the region at boundary $z \rightarrow 0$ of AdS_{d+1} . Let us evaluate A_∂ by integrating the volume element corresponding to the metric (12) at a slice $z = \varepsilon \rightarrow 0$:

$$A_\partial = \int_{\mathbb{R}^{d-1}, z=\varepsilon} d^{d-1}x \sqrt{g} = \left(\frac{L}{\varepsilon}\right)^{d-1} \int_{\mathbb{R}^{d-1}} d^{d-1}x. \quad (22)$$

The integral on the right-hand-side of (22) is the the volume of \mathbb{R}^{d-1} , which is infinite. As we did on the QFT side, we regulate it by putting the system in a box of size R :

$$\int_{\mathbb{R}^{d-1}} d^{d-1}x = R^{d-1}. \quad (23)$$

Thus, the area of the A_∂ is given by:

$$A_\partial = \left(\frac{RL}{\varepsilon}\right)^{d-1}. \quad (24)$$

Let us next introduce the Planck length l_P and the Planck mass M_P for a gravity theory in $d + 1$ dimensions as:

$$G_N = (l_P)^{d-1} = \frac{1}{(M_P)^{d-1}}. \quad (25)$$

Then, the number of degrees of freedom of the AdS_{d+1} space is:

$$N_{dof}^{AdS} = \frac{1}{4} \left(\frac{R}{\varepsilon}\right)^{d-1} \left(\frac{L}{l_P}\right)^{d-1}. \quad (26)$$

By comparing N_{dof}^{QFT} and N_{dof}^{AdS} we conclude that they scale in the same way with the IR and UV cutoffs R and ε and we can identify:

$$\frac{1}{4} \left(\frac{L}{l_P}\right)^{d-1} = c_{QFT}. \quad (27)$$

This gives the matching condition between gravity and QFT that we were looking for. Notice that a theory is (semi)classical when the coefficient multiplying its action is large. In this case the path integral is dominated by a saddle point. The action of our gravity theory in the AdS_{d+1} space of radius L contains a factor L^{d-1}/G_N . Thus, taking into account the definition of the Planck length in (25), we conclude that the classical gravity theory is reliable if:

$$\text{classical gravity in AdS} \rightarrow \left(\frac{L}{l_P}\right)^{d-1} \gg 1, \quad (28)$$

which happens when the AdS radius is large in Planck units. Since the scalar curvature goes like $1/L^2$, the curvature is small in Planck units. Thus, a QFT has a classical gravity dual when c_{QFT} is large, or equivalently if there is a large number of degrees of freedom per unit volume or a large number of species (which corresponds to large N for $SU(N)$ gauge theories).

3 String Theory Basics

As mentioned in the introduction, the AdS/CFT correspondence was originally discovered in the context of string theory. Although it can be formulated without making reference to its stringy origin, it is however very convenient to know its connection to the theory of relativistic strings in order to acquire a deep understanding of the duality. We will start reviewing qualitatively this connection in this section (see [11–16], which contain systematic expositions of string theory with different levels of technical details).

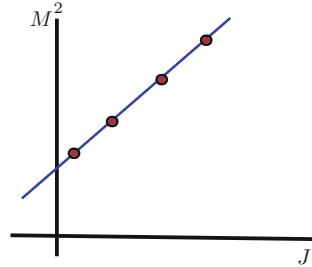
Historically, string theory was introduced in the sixties as an attempt to describe the hadronic resonances of high spin observed in the experiments. Experimentally, the mass square of these particles is linearly related to its spin J :

$$M^2 \sim J. \quad (29)$$

It is then said that the hadrons are distributed along Regge trajectories (see Fig. 2). String theory was introduced to reproduce this behavior. Actually, it is not difficult to verify qualitatively that the rotational degree of freedom of the relativistic string gives rise to Regge trajectories as in (29). Indeed, let us suppose that we have an open string with length L and tension T which is rotating around its center of mass. The mass of this object would be $M \sim TL$, whereas its angular momentum J would be $J \sim PL$, with P being its linear momentum. In a relativistic theory $P \sim M$, which implies that $J \sim PL \sim ML \sim T^{-1} M^2$ or, equivalently, $M^2 \sim TJ$. Thus, we reproduce the Regge behavior (29) with the slope being proportional to the string tension T .

The basic object of string theory is an object extended along some characteristic distance l_s . Therefore, the theory is non-local. It becomes local in the point-like limit

Fig. 2 In a Regge trajectory the mass square M^2 of the particles grows linearly with their spin J



in which the size $l_s \rightarrow 0$. The rotational degree of freedom of the string gives rise to Regge trajectories similar to those observed experimentally. In modern language one can regard a meson as a quark–antiquark pair joined by a string. The energy of such a configuration grows linearly with the length, and this constitutes a model of confinement.

The classical description of a relativistic string is directly inspired from that of a point particle in the special theory of relativity. Indeed, let us consider a relativistic point particle of mass m moving in a flat spacetime with Minkowski metric $\eta_{\mu\nu}$. As it moves the particle describes a curve in spacetime (the so-called worldline), which can be represented by a function of the type:

$$x^\mu = x^\mu(\tau), \quad (30)$$

where x^μ is the coordinate in the space in which the point particle is moving (the target space) and τ parameterizes the path of the particle (the worldline coordinate). The action of the particle is proportional to the integral of the line element along the trajectory in spacetime, with the coefficient being given by the mass m of the particle:

$$S = -m \int ds = -m \int_{\tau_0}^{\tau_1} d\tau \sqrt{-\eta_{\mu\nu} \dot{x}^\mu \dot{x}^\nu}. \quad (31)$$

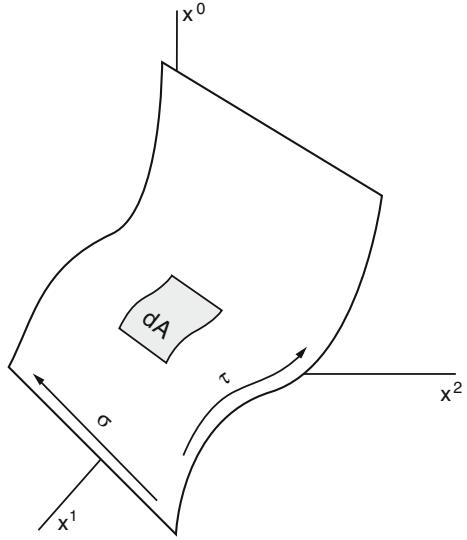
Let us now consider a relativistic string. A one-dimensional object moving in spacetime describes a surface (the so-called worldsheet). Let dA be the area element of the worldsheet (see Fig. 3). Then, the analogue of the action (31) for a string is the so-called Nambu-Goto action:

$$S_{NG} = -T \int dA, \quad (32)$$

where T is the tension of the string, given by:

$$T = \frac{1}{2\pi\alpha'}, \quad (33)$$

Fig. 3 The string sweeps in spacetime a two-dimensional worldsheet



and α' is the Regge slope (α' has units of $(length)^2$ or $(mass)^{-2}$). The string length and mass are defined as:

$$l_s = \sqrt{\alpha'} = \frac{1}{M_s}. \quad (34)$$

Then, the string tension is related to l_s and M_s as:

$$T = \frac{1}{2\pi l_s^2} = \frac{M_s^2}{2\pi}. \quad (35)$$

Let us write more explicitly the Nambu-Goto action (32). With this purpose we will take two coordinates ξ^α ($\alpha = 0, 1$) to parameterize the worldsheet Σ ($(\xi^0, \xi^1) = (\tau, \sigma)$). Let us assume that the string moves in a target space \mathcal{M} with metric $G_{\mu\nu}$. The embedding of Σ in the spacetime \mathcal{M} is characterized by a map $\Sigma \rightarrow \mathcal{M}$ with $\xi^\alpha \rightarrow X^\mu(\xi^\alpha)$. The induced metric on Σ is:

$$\hat{G}_{\alpha\beta} \equiv G_{\mu\nu} \partial_\alpha X^\mu \partial_\beta X^\nu. \quad (36)$$

Thus, the Nambu-Goto action of the string is:

$$S_{NG} = -T \int \sqrt{-\det \hat{G}_{\alpha\beta}} d^2\xi. \quad (37)$$

The action (37) depends non-linearly on the embedding functions $X^\mu(\tau, \sigma)$. The classical equations of motion derived from (37) are partial differential equations

which, remarkably, can be solved for a flat target spacetime with $G_{\mu\nu} = \eta_{\mu\nu}$, both for Neumann and Dirichlet boundary conditions. Actually, the functions $X^\mu(\tau, \sigma)$ can be represented in a Fourier expansion as an infinite superposition of oscillation modes, much as in the string of a violin.

The quantization of the string can be carried out by using the standard methods in quantum physics. The simplest way is just by canonical quantization, i.e. by considering that the X^μ are operators and by imposing canonical commutation relations between coordinates and momenta. As a result one finds that the different oscillation modes can be interpreted as particles and that the spectrum of the string contains an infinite tower of particles with growing masses and spins that are organized in Regge trajectories, with $1/l_s$ being the mass gap.

The close scrutiny of the consistency conditions of the quantum string reveals many surprises (quantizing the string is like opening Pandora's box!). First of all, the mass spectrum contains tachyons (particles with $m^2 < 0$), which is a signal of instability. In order to avoid this problem one must consider a string which has also fermionic coordinates and require that the system is supersymmetric (i.e. that there is a symmetry between bosonic and fermionic degrees of freedom). This generalization of the bosonic string (37) is the so-called superstring. Another consistency requirement imposed by the quantization is that the number D of dimensions of the space in which the string is moving is fixed. For a superstring $D = 10$. This does not mean that the extra dimensions have the same meaning as the ordinary ones of the four-dimensional Minkowski spacetime. Actually, the extra dimensions should be regarded as defining a configuration space (as the phase space in classical mechanics does). We will see below that this is precisely the interpretation that they have in the context of the AdS/CFT correspondence.

Another important piece of information about the nature of string theory is obtained from the analysis of the spectrum of massless particles. Massive particles have a mass which a multiple of $1/l_s$ and, therefore, they become unobservable in the low-energy limit $l_s \rightarrow 0$. Then, after eliminating the tachyons by using supersymmetry, the massless particles are the low-lying excitations of the spectrum. We must distinguish the case of open and closed strings. The spectrum of open strings contains massless particles of spin one with the couplings needed to have gauge symmetry. These particles can be naturally identified with gauge bosons (photons, gluons, ...). The big surprise comes when looking at the spectrum of closed strings, since it contains a particle of spin two and zero mass which can only be interpreted as the graviton (the quantum of gravity). Besides, quantum consistency of the propagation of the string in a curved space implies Einstein equations in ten dimension plus corrections:

$$R_{\mu\nu} + \dots = 0. \quad (38)$$

Then, one is led to conclude that string theory is not a theory of hadrons but a theory of quantum gravity!! Thus, the string length l_s should be of the order of the Planck length l_P and not of the order of the hadronic scale ~ 1 fm. Elementary strings without internal structure and zero thickness were born for the wrong purpose. Moreover, in

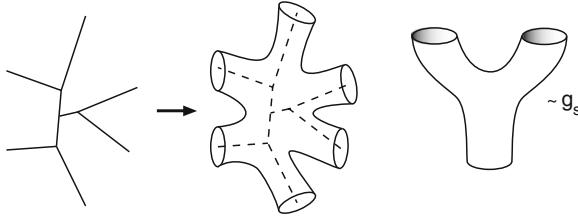


Fig. 4 A vertex in field theory corresponds to a two-dimensional surface with boundaries (*left*). The triple vertex for three closed strings is represented on the *right* as a pants surface

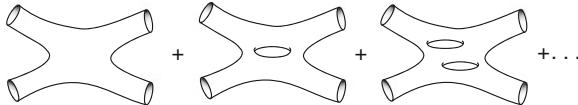


Fig. 5 Perturbative expansion of the amplitude corresponding to four closed strings

addition to the graviton, the massless spectrum contains antisymmetric tensor fields of the form $A_{\mu_1 \dots \mu_{p+1}}$, which will become very relevant in formulating the AdS/CFT correspondence.

As strings propagate through the target space \mathcal{M} , they can suffer interactions by splitting in two or more parts or by joining with other strings. The worldsheet for one of these interactions is just a two-dimensional surface with holes and boundaries which can be thought as a vertex in QFT perturbation theory in which the lines have been thickened, as shown in Fig. 4. In closed string theory the basic interaction is a string splitting into two (or the inverse process of two closed strings merging into one). The corresponding worldsheet has the form of pants, which can be obtained by thickening a triple vertex in QFT. A loop in string perturbation theory can be obtained with two triple vertices, which produces a worldsheet which is a Riemann surface with a hole. Higher order terms in perturbation theory correspond to having more holes on the worldsheet (see Fig. 5). Therefore, the perturbative series in string theory is a topological expansion!. In the topology of two-dimensional surfaces, the number h of holes (or handles) of the surface is called the genus of the surface. In string perturbation theory, h is just the number of string loops.

Let us suppose that we weigh every triple vertex with a string coupling constant coupling g_s . It is clear from the discussion above that string perturbation theory is an expansion of the type:

$$\mathcal{A} = \sum_{h=0}^{\infty} g_s^{2h-2} F_h(\alpha') , \quad (39)$$

where \mathcal{A} is some amplitude. In the next section we will identify a topological expansion in gauge theories which has the same structure as the series (39).

4 Large N Expansion in Gauge Theories

Let us consider $U(N)$ Yang-Mills theory with lagrangian

$$\mathcal{L} = -\frac{1}{g^2} \text{Tr}[F_{\mu\nu} F^{\mu\nu}], \quad (40)$$

where the non-abelian gauge field strength is given by:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]. \quad (41)$$

In the equations above A_μ is an $N \times N$ matrix whose elements can be written as $A_{\mu,b}^a$, where a, b are indices that run from 1 to N . Let us rewrite \mathcal{L} as:

$$\mathcal{L} = -\frac{N}{\lambda} \text{Tr}[F_{\mu\nu} F^{\mu\nu}], \quad (42)$$

where $\lambda = g^2 N$ is the so-called 't Hooft coupling. The 't Hooft expansion [17] corresponds to keeping λ fixed and to perform an expansion of the amplitudes in powers of N . It turns out that the different powers of N correspond to the different topologies. One can prove this by adopting a double line notation for the gauge propagator. Then, one can verify that the color lines form the perimeter of an oriented polygon (a face). Polygons join at a common edges in such a way that every vacuum graph is associated to a triangulated two-dimensional surface formed by sewing all polygons along the edges.

It is not difficult to find the powers of N and λ appearing in a given diagram D with no external lines. The contributions of the gauge propagator and the vertices are displayed in Fig. 6. Moreover, every index loop contributes with a power of N . Suppose that E is the number of propagators (edges) of D connecting two vertices, V is the number of vertices and F is the number of index loops (faces). Then, one can show that:

$$D \sim \left(\frac{\lambda}{N}\right)^E \left(\frac{N}{\lambda}\right)^V N^F = N^{F-E+V} \lambda^{E-V}. \quad (43)$$

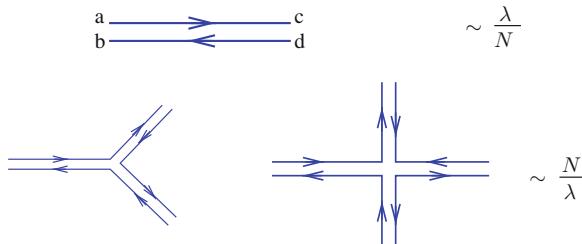


Fig. 6 Gauge propagator and vertices in the *double line* notation of Yang-Mills $U(N)$ theories

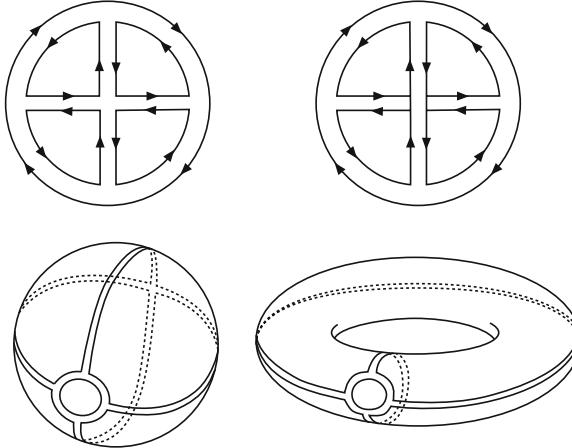


Fig. 7 The planar diagram on the *left* can be drawn on a sphere, whereas the non-planar diagram on the *right* must be drawn on a torus

Let us denote by χ the combination of F , E and V that appears in the exponent of N in (43):

$$F - E + V = \chi , \quad (44)$$

which is nothing but the Euler characteristic of the triangulation, which only depends on the topology of the surface. If the diagram triangulates a surface with h handles (an h -torus), χ is given by:

$$\chi = 2 - 2h . \quad (45)$$

Thus the diagrams are weighed with a power of N determined by the number of handles of the surface that they triangulate (see Fig. 7). The planar diagrams are those that can be drawn on a piece of paper without self-crossing. They correspond to $h = 0$ and their contribution is of the type $N^{2\lambda^n}$, for some power n which depends on the diagram considered. Clearly, as the dependence on N goes like N^{2-2h} the diagrams with $h = 0$ are the dominant ones in the large N expansion. For this reason the large N limit is also called the planar limit of the gauge theory (see [18] for more details and examples).

The effective action Z can be obtained by computing the sum over the connected vacuum-to-vacuum diagrams. It has the structure:

$$\log Z = \sum_{h=0}^{\infty} N^{2-2h} \sum_{l=0}^{\infty} c_{l,h} \lambda^l = \sum_{h=0}^{\infty} N^{2-2h} f_h(\lambda) , \quad (46)$$

where $f_h(\lambda)$ is the sum of Feynman diagrams that can be drawn in a surface of genus h . This clearly suggest a connection with the perturbative expansion of string theory written in (39) if one identifies the string coupling constant g_s as:

$$g_s \sim \frac{1}{N}. \quad (47)$$

Heuristically one can say that gauge theory diagrams triangulate the worldsheet of an effective string. The AdS/CFT correspondence is a concrete realization of this connection in the limit $(N, \lambda) \rightarrow \infty$ (i.e. for planar theories in the strongly coupled regime). This dual stringy description of gauge theory is nothing but a quantum version of the familiar description of electromagnetism in terms of the string-like lines of force.

5 D-Branes

Besides the perturbative structure reviewed above, string theories have a non-perturbative sector, which plays a crucial role in connecting them with gauge theories. The relevant objects in this non-perturbative sector are the solitons, which are extended objects. The most important solitons for the AdS/CFT correspondence are the Dp-branes, that are objects extended in $p + 1$ directions (p spatial + time). The Dp-branes can be defined as hypersurfaces where strings end (see Fig. 8). They can be obtained by quantizing the string with fixed ends along hyperplanes (Dirichlet boundary conditions). In addition, they can also be understood as objects charged under the antisymmetric tensor fields $A_{\mu_1 \dots \mu_{p+1}}$ of string theory, which naturally couple to the Dp-brane worldvolume as:

$$A_{\mu_1 \dots \mu_{p+1}} \rightarrow \int_{\mathcal{M}_{p+1}} A_{\mu_1 \dots \mu_{p+1}} dx^{\mu_1} \dots dx^{\mu_{p+1}}, \quad (48)$$

with \mathcal{M}_{p+1} being the worldvolume of the Dp-brane. The Dp-branes are dynamical objects that can move and get excited. Schematically, their action takes the form:

$$S_{Dp} = -T_{Dp} \int d^{p+1}x [\dots], \quad (49)$$

with T_{Dp} being the tension of the Dp-brane, which in terms of the string coupling constant g_s and the string length l_s is given by:

$$T_{Dp} = \frac{1}{(2\pi)^p g_s l_s^{p+1}}. \quad (50)$$

Notice from (50) that $T_{Dp} \sim g_s^{-1}$, which confirms that the Dp-branes are non-perturbative objects in string theory. The dependence of T_{Dp} on l_s in (50) is fixed by dimensional analysis.

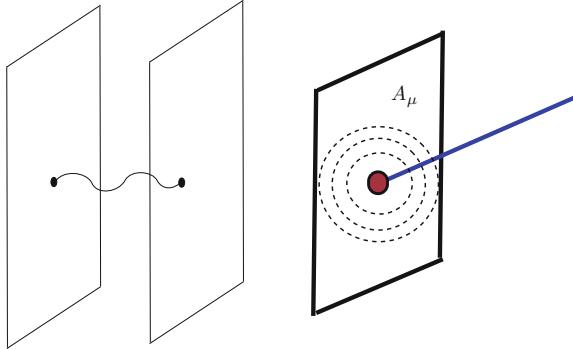


Fig. 8 The D-branes are hyperplanes where open strings can end. The endpoints of the strings source gauge fields A_μ on the D-brane worldvolume

The D p -branes can have two classes of excitations. The first one correspond to rigid motions and deformations of their shape. These degrees of freedom can be parameterized by the $9 - p$ coordinates ϕ^i ($i = 1, \dots, 9 - p$) transverse to the $(p + 1)$ -dimensional worldvolume in the ten-dimensional target space. The ϕ^i 's are just scalar fields on the D p -brane worldvolume. Besides, the D p -branes can have internal excitations. To get a clue on how to represent these excitations, let us recall that the endpoint of the string is a charge. When there is a charge, a gauge field is sourced, as illustrated in Fig. 8. Then, it follows that a D p -brane has an abelian gauge field A_μ ($\mu = 0, \dots, p$) living in its worldvolume. The action of the D p -brane which takes into account these two types of excitations is the so-called Dirac-Born-Infeld action, which can be written as:

$$S_{DBI} = -T_{Dp} \int d^{p+1}x \sqrt{-\det(g_{\mu\nu} + 2\pi l_s^2 F_{\mu\nu})}, \quad (51)$$

where $g_{\mu\nu}$ is the induced metric on the worldvolume and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field strength of the gauge field A_μ . When the worldvolume gauge field is not excited ($F_{\mu\nu} = 0$) the DBI action (51) becomes the natural generalization of the Nambu-Goto action (37) for an object extended along p spatial directions.

Let us consider a D p -brane in flat space. The induced metric in this case takes the form:

$$g_{\mu\nu} = \eta_{\mu\nu} + (2\pi l_s^2)^2 \partial_\mu \phi^i \partial_\nu \phi^i, \quad (52)$$

where the ϕ^i are the coordinates that parameterize the embedding of the brane. Let us now expand the square root of (51) in powers of $F_{\mu\nu}$ and $\partial_\mu \phi$. The quadratic terms in this expansion can be written as:

$$S_{DBI}^{(2)} = -\frac{1}{g_{YM}^2} \left(\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} \partial_\mu \phi^i \partial^\mu \phi^i + \dots \right), \quad (53)$$

which is just the ordinary action of a gauge field and $9 - p$ scalar fields. In (53) g_{YM} is the Yang-Mills coupling, which, in terms of l_s and g_s , is given by:

$$g_{YM}^2 = 2(2\pi)^{p-2} l_s^{p-3} g_s. \quad (54)$$

In addition to A_μ and ϕ , in superstring theory the branes have fermionic excitations, which can be represented in terms of fermionic fields.

We have just discovered one of the most important features of the Dp-branes: they contain a gauge theory living in their worldvolume!. In the case of a single Dp-brane the gauge group is $U(1)$, as in (53). However, if we consider a stack of coincident parallel Dp-branes the gauge group gets promoted to $U(N)$. Indeed, one can verify in this case that the fields A_μ and ϕ^i are matrices transforming in the adjoint representation of $U(N)$. The non-diagonal elements of these fields correspond to excitations connecting different branes, whereas the diagonal ones are excitations on a single brane. Actually, the $U(1)$ component of the $U(N)$ gauge theory can be decoupled from the $SU(N)$ fields. Therefore, we conclude that a stack of N Dp-branes realizes a $SU(N)$ gauge theory in $p + 1$ dimensions.

The D-branes have brought a completely new perspective on gauge theories which gives rise to what is called brane engineering. The geometric realization of the gauge symmetry that they provide allows to perform a series of transformations which lead to discover new unexpected properties of gauge theories. For example, one can move the branes, place them in different spaces, etc. In the gauge theory, these transformations lead to dualities, reductions of the amount of supersymmetry, changes of the field content, etc. In general we have a novel geometric insight on gauge dynamics. The AdS/CFT correspondence is an important outcome of this new realization of the gauge symmetry.

The particular case of the D3-branes is specially relevant in what follows. In this case we have a $3 + 1$ worldvolume and $10 - 4 = 6$ scalar fields. The corresponding four-dimensional $SU(N)$ gauge theory can be identified with super Yang-Mills theory with four supersymmetries ($\mathcal{N} = 4$ SYM). This theory is an exact CFT at the quantum level and will be the basic example of the AdS/CFT correspondence. Notice from (54) that the Yang-Mills coupling is dimensionless in this case and is related to the string coupling constant g_s as:

$$g_{YM}^2 = 4\pi g_s, \quad (\text{D3-branes}). \quad (55)$$

6 D-Branes and Gravity

String theory is a gravity theory in which all types of matter distort the spacetime. This distortion is determined by solving Einstein equations which follow from the action:

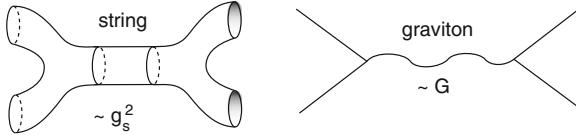


Fig. 9 In string theory the exchange of a graviton is obtained as the low-energy limit of the exchange of a closed string

$$S = \frac{1}{16\pi G} \int d^{10}x \sqrt{-g} R + \dots \quad (56)$$

The ten-dimensional Newton constant in (56) is related to string parameters as:

$$16\pi G = (2\pi)^7 g_s^2 l_s^8. \quad (57)$$

The dependence of G on l_s in (57) follows from dimensional analysis. The dependence on g_s follows by comparing two-to-two scattering amplitudes in string theory and in supergravity (see Fig. 9).

The Dp-branes (and other extended objects) are solutions of the Einstein equations. Let us consider these solutions at the linearized level of weak gravity (i.e. when we are far from the object). The linearized metric for a point-like object in a D -dimensional spacetime is:

$$ds^2 \approx -(1 + 2\varphi) dt^2 + \left(1 - \frac{2}{D-3}\varphi\right) (dx_1^2 + \dots + dx_{D-1}^2), \quad (58)$$

where the function φ parameterizes the deviation of the metric from the flat Minkowski metric in D dimensions. The specific form (58) can be obtained by solving the linearized Einstein equations and generalizes the standard result in general relativity in four dimensions. As in this latter particular case, one can identify the function φ with the Newtonian gravitational potential, as seen by comparing the motion along a geodesic and Newton's law. Thus:

$$\varphi \sim \frac{GM}{r^{D-3}}, \quad (59)$$

where M is the mass of the particle and $r = \sqrt{x_1^2 + \dots + x_{D-1}^2}$ is the radial coordinate of the space. The function φ is a solution of the Poisson equation in $D - 1$ dimensions. Notice that the power $D - 3$ is equal to the number d_T of dimensions transverse to the object (i.e. $d_T = D - 1$) minus 2:

$$D - 3 = d_T - 2. \quad (60)$$

For an extended object (along p spatial directions) the corresponding linearized metric takes the form:

$$ds^2 \approx (1 + 2\varphi) (-dt^2 + dx_1^2 + \cdots + dx_p^2) + \left(1 - \frac{2(p+1)}{D-p-3}\varphi\right) (dx_{p+1}^2 + \cdots + dx_{D-1}^2). \quad (61)$$

Since now the number of transverse directions is $d_T = D - 1 - p$, the function φ must be:

$$\varphi \sim \frac{GM}{r^{D-p-3}}. \quad (62)$$

The D3-brane solution is a solution of 10d supergravity corresponding to a stack of N coincident D3-branes. The (exact) metric takes the form:

$$ds^2 = H^{-\frac{1}{2}} \left(-dt^2 + dx_1^2 + dx_2^2 + dx_3^2 \right) + H^{\frac{1}{2}} (dr^2 + r^2 d\Omega_5^2), \quad (63)$$

with $d\Omega_5^2$ is the line element of a unit \mathbb{S}^5 . Notice that $dr^2 + r^2 d\Omega_5^2$ is the flat metric of \mathbb{R}^6 . The function $H(r)$ is the so-called warp factor:

$$H = 1 + \frac{L^4}{r^4}, \quad (64)$$

with the constant L being given by:

$$L^4 = 4\pi g_s N l_s^4. \quad (65)$$

At linearized level (when $r \rightarrow \infty$) the solution (63) corresponds to the general solution in (61) with $D = 10$, $p = 3$ and the function φ given by:

$$\varphi = -\frac{1}{4} \frac{L^4}{r^4}. \quad (66)$$

Comparing this expression of φ with the one written in (62) for $D = 10$ and $p = 3$ we conclude that:

$$GM \sim L^4. \quad (67)$$

Let us find this last relation from a different reasoning. As $G \sim g_s^2 l_s^8$ (see 57) and since M should be the mass of a stack of N D3-branes with tension $T_{D3} \sim 1/(g_s l_s^4)$, we should have:

$$GM \sim g_s^2 l_s^8 \frac{N}{g_s l_s^4} \sim N g_s l_s^4, \quad (68)$$

which, apart from a numerical factor is precisely the value of L^4 written in (65).

The geometry of this solution is asymptotically the Minkowski spacetime in 10d with a throat of infinite size (see Fig. 1). In the throat we can take $r \ll 1$ and neglect the 1 in the function H . This defines the so-called near-horizon limit. Actually, taking

$$H \approx \frac{L^4}{r^4}, \quad (69)$$

the metric (63) becomes:

$$ds^2 = \frac{r^2}{L^2} (-dt^2 + dx_1^2 + dx_2^2 + dx_3^2) + \frac{L^2}{r^2} dr^2 + L^2 d\Omega_5^2. \quad (70)$$

To identify this near-horizon metric, let us change variable as:

$$r = \frac{L}{z}. \quad (71)$$

Then, the line element (70) can be rewritten as:

$$ds^2 = \frac{L^2}{z^2} (-dt^2 + dx_1^2 + dx_2^2 + dx_3^2 + dz^2) + L^2 d\Omega_5^2, \quad (72)$$

which is precisely the metric of the product space $AdS_5 \times S^5$, with L being the radius of both factors.

7 The AdS/CFT for $\mathcal{N}=4$ $SU(N)$ SYM

We are now in a position to formulate the AdS/CFT correspondence, as originally proposed by Maldacena in [1]. We have seen two different descriptions of the D3-brane: as a gauge theory and as a gravity solution. Maldacena proposed that the two descriptions are dual to each other. More generally, it was conjectured in [1] that $\mathcal{N}=4$ SYM theory with gauge group $SU(N)$ is equivalent to string theory in $AdS_5 \times S^5$.

In order to make the duality more precise, it is quite convenient to find the relation of the two parameters on the two sides of the correspondence. First of all, by combining (65) and (55) we can write the ratio between the AdS radius L and the string length l_s in terms of gauge theory quantities:

$$\left(\frac{L}{l_s}\right)^4 = N g_{YM}^2. \quad (73)$$

Notice that the right-hand side of (73) is just the 't Hooft coupling $\lambda = N g_{YM}^2$. Therefore, we can rewrite (73) as:

$$\frac{l_s^2}{L^2} = \frac{1}{\sqrt{\lambda}}. \quad (74)$$

Moreover, by combining (57) and (55) we can write the ten-dimensional Newton constant G , and the corresponding Planck length l_P , in terms of g_{YM} , namely:

$$G = l_P^8 = \frac{\pi^4}{2} g_{YM}^4 l_s^8. \quad (75)$$

Then, the relation between the Planck length and the AdS radius is given by

$$\left(\frac{l_P}{L}\right)^8 = \frac{\pi^4}{2N^2}. \quad (76)$$

The relations (74) and (76) are essential to delimitate the domain of validity of the dual description in terms of classical gravity. As discussed above, we must require that $l_P/L \ll 1$ if we want to avoid having quantum gravity corrections (this is equivalent to the condition of having small curvature). From (76) is clear that one should require that $N \gg 1$. Notice this conclusion agrees with our general discussion at the end of Sect. 2. Moreover, in order to avoid stringy corrections due to the full tower of massive states of the string, we must require that $l_s/L \ll 1$ which, in view of (74), means that the 't Hooft coupling must be large. Therefore, we conclude that the gravitational description of $\mathcal{N} = 4$ SYM is reliable if:

$$N \gg 1, \quad \lambda \gg 1. \quad (77)$$

Then, the planar (large N), strongly coupled (large λ) SYM theory can be described as classical gravity.

7.1 Reduction on the \mathbb{S}^5

Let us now discuss further the role of the five-sphere in the AdS dual of $\mathcal{N} = 4$ SYM. We begin by noticing that any field on $AdS_5 \times \mathbb{S}^5$ can be reduced to a tower of fields on AdS_5 by expanding it in terms of the harmonics on \mathbb{S}^5 :

$$\phi(x, \Omega) = \sum_l \phi_l(x) Y_l(\Omega), \quad (78)$$

with x being coordinates of AdS_5 , Ω are coordinates of \mathbb{S}^5 and $Y_l(\Omega)$ are spherical harmonics on \mathbb{S}^5 . The gravity action, after reducing on \mathbb{S}^5 , becomes:

$$S = \frac{1}{16\pi G_5} \int d^5x \left[\mathcal{L}_{grav} + \mathcal{L}_{matter} \right], \quad (79)$$

where G_5 is the five-dimensional Newton constant, and the gravitational part of the \mathcal{L} is:

$$\mathcal{L}_{grav} = \sqrt{-g} \left[R + \frac{12}{L^2} \right], \quad (80)$$

which corresponds to a negative cosmological constant $\Lambda = -6/L^2$. The five-dimensional Newton constant can be related to the 10d constant by considering the reduction of the Einstein-Hilbert term:

$$\frac{1}{16\pi G} \int d^5x d^5\Omega \sqrt{-g_{10}} R_{10} \rightarrow \frac{L^5 \Omega_5}{16\pi G} \int d^5x \sqrt{-g_5} R_5, \quad (81)$$

where $\Omega_5 = \pi^3$ is the volume of a unit \mathbb{S}^5 . Then it follows that G_5 and G can be related as:

$$G_5 = \frac{G}{L^5 \Omega_5} = \frac{G}{\pi^3 L^5}. \quad (82)$$

Using the value of the ten-dimensional Newton constant G written in (57), we get that G_5 is given by:

$$G_5 = \frac{\pi}{2N^2} L^3. \quad (83)$$

From this formula and (27) we can compute the central charge of the SYM theory:

$$c_{SYM} = \frac{1}{4} \frac{L^3}{G_5} = \frac{N^2}{2\pi}. \quad (84)$$

Then, the large N limit corresponds to having a large central charge, in agreement with our previous discussions on the validity domain of the gravitational description.

Let us now consider the reduction (78) of a massless scalar field in $AdS_5 \times \mathbb{S}^5$. The Klein-Gordon equation in ten dimensions is:

$$\nabla^2 \phi = 0. \quad (85)$$

Since the metric factorizes into a AdS_5 and \mathbb{S}^5 part, the D'Alembertian is additive

$$\nabla^2 = \nabla_{AdS_5}^2 + \nabla_{\mathbb{S}^5}^2, \quad (86)$$

where $\nabla_{\mathbb{S}^5}^2$ is nothing but the quadratic Casimir operator in $SO(6)$. The eigenvalues of $\nabla_{\mathbb{S}^5}^2$ acting on the spherical harmonics are:

$$\nabla_{\mathbb{S}^5}^2 Y_l(\Omega) = -\frac{C_l^{(5)}}{L^2} Y_l(\Omega), \quad (87)$$

where the $C_l^{(5)}$ are given by:

$$C_l^{(5)} = l(l+4), \quad l = 0, 1, 2, \dots . \quad (88)$$

Thus, the reduced AdS_5 fields ϕ_l satisfy the massive Klein-Gordon equation

$$\nabla_{AdS_5}^2 \phi_l = m_l^2 \phi_l, \quad m_l^2 = \frac{l(l+4)}{L^2}. \quad (89)$$

Therefore, we have a tower of massive fields ϕ_l , with a particular set of masses, which originate, after dimensional Kaluza-Klein (KK) reduction on the five-sphere, from a single massless scalar field ϕ in ten-dimensions. These fields should have a field theory dual in the $\mathcal{N} = 4$ theory and the mass spectrum (89) should have a counterpart on the field theory side. We will see that this is indeed the case in the next section.

8 A Scalar Field in AdS

We argued in Sect. 1 that fields in AdS correspond to sources of operators on the field theory side and that we can learn about these dual operators by analyzing the dynamics of the sources in the curved space. In this section we study the simplest case of a scalar field in AdS . Accordingly, let us consider the AdS_{d+1} space in euclidean signature with metric:

$$ds^2 = \frac{L^2}{z^2} [dz^2 + \delta_{\mu\nu} dx^\mu dx^\nu]. \quad (90)$$

The action of a scalar field ϕ in the AdS_{d+1} space is:

$$S = -\frac{1}{2} \int d^{d+1}x \sqrt{g} \left[g^{MN} \partial_M \phi \partial_N \phi + m^2 \phi^2 \right]. \quad (91)$$

The equation of motion derived from the action (91) is:

$$\frac{1}{\sqrt{g}} \partial_M \left(\sqrt{g} g^{MN} \partial_N \phi \right) - m^2 \phi = 0. \quad (92)$$

More explicitly, after using the metric (90), this equation becomes :

$$z^{d+1} \partial_z (z^{1-d} \partial_z \phi) + z^2 \delta^{\mu\nu} \partial_\mu \partial_\nu \phi - m^2 L^2 \phi = 0. \quad (93)$$

Let us perform the Fourier transform of ϕ in the x^μ coordinates:

$$\phi(z, x^\mu) = \int \frac{d^d k}{(2\pi)^d} e^{ik \cdot x} f_k(z). \quad (94)$$

Then, the equation of motion becomes:

$$z^{d+1} \partial_z (z^{1-d} \partial_z f_k) - k^2 z^2 f_k - m^2 L^2 f_k = 0. \quad (95)$$

Let us solve (95) near the boundary $z = 0$. We put $f_k \sim z^\beta$ for some exponent β and keep the leading terms near $z = 0$. Then, it is straightforward to find that β must satisfy the following quadratic expression:

$$\beta(\beta - d) - m^2 L^2 = 0, \quad (96)$$

whose solutions are:

$$\beta = \frac{d}{2} \pm \sqrt{\frac{d^2}{4} + m^2 L^2}, \quad (97)$$

Therefore, near $z \sim 0$ the function $f_k(z)$ behaves as :

$$f_k(z) \approx A(k) z^{d-\Delta} + B(k) z^\Delta. \quad (98)$$

where Δ is given by:

$$\Delta = \frac{d}{2} + \nu, \quad \nu = \sqrt{\frac{d^2}{4} + m^2 L^2}. \quad (99)$$

By performing the inverse Fourier transform we can write the expansion near the boundary in position space:

$$\phi(z, x) \approx A(x) z^{d-\Delta} + B(x) z^\Delta, \quad z \rightarrow 0. \quad (100)$$

Notice that Δ is real if $\nu \in \mathbb{R}$, which happens if the mass m satisfies the so-called Breitenlohner-Freedman (BF) bound:

$$m^2 \geq -\left(\frac{d}{2L}\right)^2. \quad (101)$$

This means that m^2 can be negative (and the field ϕ can be tachyonic) but it must satisfy the BF bound. In what follows we will suppose that the BF bound is satisfied. Moreover, notice that:

$$d - \Delta \leq \Delta \iff \nu = 2\Delta - d \geq 0, \quad (102)$$

which is obviously satisfied in the mass is above the BF bound. Then, the term behaving as $z^{d-\Delta}$ in (100) is the dominant one as $z \rightarrow 0$. Let us take the boundary as $z = \varepsilon$ and neglect the subdominant term. We have:

$$\phi(z = \varepsilon, x) \approx \varepsilon^{d-\Delta} A(x). \quad (103)$$

As $d - \Delta$ is negative if $m^2 > 0$, the leading term is typically divergent as we approach the boundary at $z = \varepsilon \rightarrow 0$. In order to identify the QFT source $\varphi(x)$ from the boundary value of the field $\phi(z, x)$ we have to remove the divergences of the latter. We will simply do it by extracting the divergent multiplicative factor from (103), i.e. the QFT source $\varphi(x)$ is identified with $A(x)$. Equivalently, we define:

$$\varphi(x) = \lim_{z \rightarrow 0} z^{\Delta-d} \phi(z, x). \quad (104)$$

Clearly, with this definition $\varphi(x)$ is always finite. Conversely we can write $\phi(z, x) = z^{d-\Delta} \varphi(x)$ at leading order. In order to interpret the meaning of Δ , let us look at the boundary action. If \mathcal{O} is the operator dual to ϕ , this action is given by:

$$S_{\text{bdy}} \sim \int d^d x \sqrt{\gamma_\varepsilon} \phi(\varepsilon, x) \mathcal{O}(\varepsilon, x), \quad (105)$$

where $\gamma_\varepsilon = \left(\frac{L}{\varepsilon}\right)^{2d}$ is the determinant of the induced metric at the $z = \varepsilon$ boundary. Then, by plugging $\phi(\varepsilon, x) = \varepsilon^{d-\Delta} \varphi(x)$ inside S_{bdy} , we get:

$$S_{\text{bdy}} \sim L^d \int d^d x \varphi(x) \varepsilon^{-\Delta} \mathcal{O}(\varepsilon, x). \quad (106)$$

In order to make S_{bdy} finite and independent of ε as $\varepsilon \rightarrow 0$ we should require:

$$\mathcal{O}(\varepsilon, x) = \varepsilon^\Delta \mathcal{O}(x). \quad (107)$$

But, passing from $z = 0$ to $z = \varepsilon$ is a scale transformation in the QFT. Thus, Δ must be interpreted as the mass scaling dimension of the dual operator \mathcal{O} . In other words, the behavior $\mathcal{O}(\varepsilon, x) = \varepsilon^\Delta \mathcal{O}(x)$ is the wave function renormalization of \mathcal{O} as we move into the bulk. Similarly, from the relation $\phi(\varepsilon, x) = \varepsilon^{d-\Delta} \varphi(x)$ it follows that $d - \Delta$ is the mass scaling dimension of the source φ . Remember from (99) that:

$$\Delta = \frac{d}{2} + \sqrt{\left(\frac{d}{2}\right)^2 + m^2 L^2}. \quad (108)$$

Then, if $m^2 \geq -\left(\frac{d}{2L}\right)^2$, the corresponding scaling dimension is real. We consider three different situations, depending on the value of the mass m . The first case we will study is the one in which $m^2 > 0$. In this case $\Delta > d$ and the corresponding

operator \mathcal{O} is called an irrelevant operator. By deforming the CFT with \mathcal{O} , we have the following term in the action:

$$\Delta S = \int d^d x (\text{mass})^{d-\Delta} \mathcal{O} = \int d^d x (\text{mass})^{\text{negative number}} \mathcal{O}. \quad (109)$$

When computing amplitudes with this action one gets that the effect of the interaction depends on $\left(\frac{\text{Energy}}{\text{mass}}\right)^\alpha$, for some $\alpha > 0$. Then, the effects of this interaction can be neglected for low energies since they are suppressed by powers. Therefore, the interaction goes away in the IR but changes completely the UV of the theory.

When $m^2 = 0$ we have from (108) that $\Delta = d$ and the corresponding operator is called marginal. Finally, if m^2 is negative and takes values in the range $-(d/2L)^2 < m^2 < 0$, the dual operator it is called relevant, since it has $\Delta < d$ and changes the IR of the theory.

8.1 Application to the KK Spectrum in $AdS_5 \times \mathbb{S}^5$

We have seen in Sect. 7.1 that a massless scalar in ten-dimensions gives rise to the tower (89) of massive KK scalars in AdS_5 . We will now apply the formula (108) to these five-dimensional scalar fields. First of all, the dimension/mass relation for the case $d = 4$ is:

$$\Delta = 2 + \sqrt{4 + (m L)^2}. \quad (110)$$

Let us first consider a massless scalar, which corresponds to the s-wave ($l = 0$) in (7.1). In this case (110) with $m = 0$ gives $\Delta = 4$. Then, the QFT dual operator should be a scalar operator of dimension 4. Since this s-wave operator is singlet under the $SO(6)$ symmetry of the \mathbb{S}^5 , it must not contain the ϕ_i scalars of the dual $\mathcal{N} = 4$ QFT. The only candidate with these characteristics is the glueball operator:

$$\mathcal{O} = \text{Tr}[F_{\mu\nu} F^{\mu\nu}]. \quad (111)$$

Notice that $\dim[\partial] = \dim[A] = 1$, so, indeed, $\dim(\mathcal{O}) = 4$. For higher order KK modes, the masses are $m^2 L^2 = l(l+4)$ (see 89). Then, the dimensions are:

$$\Delta_l = 2 + \sqrt{4 + l(l+4)} = 4 + l. \quad (112)$$

In this case, the dual operator should transform under the corresponding representation of $SO(6)$ (a symmetric tensor with l indices). One can construct such a tensor by multiplying $\mathcal{N} = 4$ SYM scalar fields ϕ_i (they transform as vectors of $SO(6)$). Then, the natural operator dual to the l th KK mode is:

$$\mathcal{O}_{i_1, \dots, i_l} = \text{Tr}[\phi_{(i_1, \dots, i_l)} F_{\mu\nu} F^{\mu\nu}], \quad (113)$$

with $\phi_{(i_1, \dots, i_l)}$ being the traceless symmetric product of l scalar fields ϕ_i of $\mathcal{N} = 4$ SYM. As $\dim[\phi] = 1$, one can check immediately that the dimension of the operator $\mathcal{O}_{i_1, \dots, i_l}$ in (113) is indeed $4 + l$, in agreement with the AdS/CFT result. It has been checked that this agreement can be extended to all the KK modes of 10d supergravity on $AdS_5 \times \mathbb{S}^5$ (including fermions, forms, \dots).

9 Correlation Functions

Let us now see how one can compute correlation functions in Euclidean space from gravity. The objective is to obtain Euclidean correlation functions of the type:

$$\langle \mathcal{O}(x_1) \dots \mathcal{O}(x_n) \rangle. \quad (114)$$

In field theory these correlators can be calculated from a generating function, which is obtained by perturbing the lagrangian by a source term:

$$\mathcal{L} \rightarrow \mathcal{L} + J(x) \mathcal{O}(x) \equiv \mathcal{L} + \mathcal{L}_J. \quad (115)$$

The generating functional is just:

$$Z_{QFT}[J] = \left\langle \exp\left[\int \mathcal{L}_J\right] \right\rangle_{QFT}. \quad (116)$$

The connected correlators are obtained from the functional derivatives of Z :

$$\left\langle \prod_i \mathcal{O}(x_i) \right\rangle = \prod_i \frac{\delta}{\delta J(x_i)} \log Z_{QFT}[J] \Big|_{J=0} \quad (117)$$

Let us consider now any bulk field $\phi(z, x)$ fluctuating in AdS . Let $\phi_0(x)$ be the boundary value of ϕ :

$$\phi_0(x) = \phi(z=0, x) = \phi|_{\partial AdS}(x). \quad (118)$$

The field ϕ_0 is related to a source for some dual operator \mathcal{O} in the QFT. As we know the actual source is not the value of ϕ at $z = 0$, which is typically divergent, but the limit:

$$\lim_{z \rightarrow 0} z^{\Delta-d} \phi(z, x) = \varphi(x). \quad (119)$$

Then, the AdS/CFT prescription for the generating functional is [2, 3]:

$$Z_{QFT}[\phi_0] = \left\langle \exp \left[\int \phi_0 \mathcal{O} \right] \right\rangle_{QFT} = Z_{gravity}[\phi \rightarrow \phi_0], \quad (120)$$

where $Z_{gravity}[\phi \rightarrow \phi_0]$ is the partition function (i.e. the path integral) in the gravity theory evaluated over all functions which have the value ϕ_0 at the boundary of AdS :

$$Z_{gravity}[\phi \rightarrow \phi_0] = \sum_{\{\phi \rightarrow \phi_0\}} e^{S_{gravity}}. \quad (121)$$

In the limit in which classical gravity dominates, one can substitute the sum by the term corresponding to the classical solution. In this case the generating function becomes:

$$Z_{QFT}[\phi_0] \approx e^{S_{gravity}^{on-shell}[\phi \rightarrow \phi_0]}. \quad (122)$$

One should be careful when evaluating the on-shell gravity action because it typically diverges and has to be renormalized following the procedure of holographic renormalization [19, 20] (see below and the review [21]). Thus, the classical action must be substituted by a renormalized version, which will be denoted by S_{grav}^{ren} and the generating functional becomes:

$$\log Z_{QFT} = S_{grav}^{ren}[\phi \rightarrow \phi_0]. \quad (123)$$

Moreover, the n -point function can be obtained by computing the derivatives with respect to $\varphi = z^{\Delta-d}\phi$:

$$\langle \mathcal{O}(x_1) \dots \mathcal{O}(x_n) \rangle = \frac{\delta^{(n)} S_{grav}^{ren}[\phi]}{\delta \varphi(x_1) \dots \delta \varphi(x_n)} \Big|_{\varphi=0}. \quad (124)$$

9.1 One-Point Function

It is also interesting to compute the one-point function of an operator \mathcal{O} in the presence of the source φ :

$$\langle \mathcal{O}(x) \rangle_\varphi = \frac{\delta S_{grav}^{ren}[\phi]}{\delta \varphi(x)}. \quad (125)$$

Taking into account the relation between ϕ and φ (119), we get:

$$\langle \mathcal{O}(x) \rangle_\varphi = \lim_{z \rightarrow 0} z^{d-\Delta} \frac{\delta S_{grav}^{ren}[\phi]}{\delta \phi(z, x)}. \quad (126)$$

The functional derivative of the classical on-shell action can be computed in closed form. Indeed, let S_{grav} be represented as:

$$S_{grav} = \int_{\mathcal{M}} \int dz d^d x \mathcal{L}[\phi, \partial\phi], \quad (127)$$

with \mathcal{M} being a $(d+1)$ -dimensional manifold whose boundary is located at $z=0$. Under a general change $\phi \rightarrow \phi + \delta\phi$, the classical action S_{grav} varies as:

$$\delta S_{grav} = \int_{\mathcal{M}} \int dz d^d x \left[\frac{\partial \mathcal{L}}{\partial \phi} \delta\phi + \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \delta(\partial_\mu \phi) \right]. \quad (128)$$

Let us now use in (128) that $\delta[\partial_\mu \phi] = \partial_\mu(\delta\phi)$ and let us integrate by parts. We get:

$$\delta S_{grav} = \int_{\mathcal{M}} \int dz d^d x \left[\left(\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \right) \right) \delta\phi + \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \delta\phi \right) \right]. \quad (129)$$

The first term in the previous equation vanishes on-shell due to the Euler-Lagrange equations. As the boundary is at $z=\varepsilon \rightarrow 0$, we can write:

$$\delta S_{grav}^{on-shell} = \int_{\varepsilon}^{\infty} \int d^d x \partial_z \left(\frac{\partial \mathcal{L}}{\partial (\partial_z \phi)} \delta\phi \right) = - \int_{\partial M} d^d x \frac{\partial \mathcal{L}}{\partial (\partial_z \phi)} \delta\phi \Big|_{z=\varepsilon}. \quad (130)$$

Let us next define Π as:

$$\Pi = - \frac{\partial \mathcal{L}}{\partial (\partial_z \phi)}, \quad (131)$$

which is the canonical momentum if z is taken as time. Then (130) can be rewritten as:

$$\delta S_{grav}^{on-shell} = \int_{\partial M} d^d x \Pi(\varepsilon, x) \delta\phi(\varepsilon, x). \quad (132)$$

Thus, it follows that:

$$\frac{\delta S_{grav}^{on-shell}}{\delta\phi(\varepsilon, x)} = \Pi(\varepsilon, x) = - \frac{\partial \mathcal{L}}{\partial (\partial_z \phi)}. \quad (133)$$

In general the renormalized action can be written as:

$$S^{ren} = S_{grav}^{on-shell} + S_{ct}, \quad (134)$$

where S_{ct} is the action of counterterms, defined at the boundary $z = \varepsilon$. Let us define the renormalized momentum as:

$$\Pi^{ren}(z, x) = \frac{\delta S^{ren}}{\delta \phi(z, x)}. \quad (135)$$

Clearly, by taking $z = \varepsilon$ in (135) and (134) we have:

$$\Pi^{ren}(\varepsilon, x) = -\frac{\partial \mathcal{L}}{\partial (\partial_z \phi(\varepsilon, x))} + \frac{\delta S_{ct}}{\delta \phi(\varepsilon, x)}. \quad (136)$$

Therefore, from the definition of Π^{ren} in (135) and (126), we obtain the one-point function of \mathcal{O} in presence of the source φ as the following limit in the AdS boundary:

$$\langle \mathcal{O}(x) \rangle_\varphi = \lim_{z \rightarrow 0} z^{d-\Delta} \Pi^{ren}(z, x). \quad (137)$$

9.2 Linear Response Theory

The field theory path integral representation of the one-point function with a source is:

$$\langle \mathcal{O}(x) \rangle_\varphi = \int [D\psi] \mathcal{O}(x) e^{S_E[\psi] + \int d^d y \varphi(y) \mathcal{O}(y)}, \quad (138)$$

where ψ denotes the fields of the QFT. Let us expand the exponent of this expression in a power series of the source φ and let us keep the terms up to linear order:

$$\langle \mathcal{O}(x) \rangle_\varphi = \langle \mathcal{O}(x) \rangle_{\varphi=0} + \int d^d y \langle \mathcal{O}(x) \mathcal{O}(y) \rangle \varphi(y) + \dots \quad (139)$$

Next, we define the euclidean two-point function $G_E(x - y)$ as:

$$G_E(x - y) = \langle \mathcal{O}(x) \mathcal{O}(y) \rangle. \quad (140)$$

Then, (139) can be rewritten as:

$$\langle \mathcal{O}(x) \rangle_\varphi = \langle \mathcal{O}(x) \rangle_{\varphi=0} + \int d^d y G_E(x - y) \varphi(y). \quad (141)$$

We will consider normal-ordered observables such that $\langle \mathcal{O}(x) \rangle_{\varphi=0}$ vanishes. Notice that this always can be achieved by subtracting to \mathcal{O} its vacuum expectation value (VEV) without source. Then, $\langle \mathcal{O}(x) \rangle_\varphi$ measures the fluctuations of the observable away from the expectation value, i.e. the linear response of the system to the external perturbation and we can write:

$$\langle \mathcal{O}(x) \rangle_\varphi = \int d^d y G_E(x - y) \varphi(y). \quad (142)$$

In momentum space this expression can be written as:

$$\langle \mathcal{O}(k) \rangle_\varphi = G_E(k) \varphi(k), \quad (143)$$

and, thus, we can obtain the two-point function in momentum space by dividing the one-point function by the source:

$$G_E(k) = \frac{\langle \mathcal{O}(k) \rangle_\varphi}{\varphi(k)}. \quad (144)$$

In the framework of the AdS/CFT correspondence, since $\varphi = z^{\Delta-d} \phi(z, x)$ with $z \rightarrow 0$, we have the following formula for the two-point function in momentum space:

$$G_E(k) = \lim_{z \rightarrow 0} z^{2(d-\Delta)} \frac{\Pi^{ren}(z, k)}{\phi(z, k)}. \quad (145)$$

9.3 Two-Point Function for a Scalar Field

Let us apply the developments of the previous sections to the case in which the source ϕ is a scalar field in Euclidean AdS_{d+1} . We will assume that the action of ϕ is:

$$S = -\frac{\eta}{2} \int dz d^d x \sqrt{g} \left[g^{MN} \partial_M \phi \partial_N \phi + m^2 \phi^2 \right], \quad (146)$$

where η is a normalization constant. In order to evaluate the on-shell action of ϕ we rewrite S as follows:

$$S = -\frac{\eta}{2} \int dz d^d x \partial_M \left[\sqrt{g} \phi g^{MN} \partial_N \phi \right] + \frac{\eta}{2} \int dz d^d x \phi \sqrt{g} \left[\frac{1}{\sqrt{g}} \partial_M \left(\sqrt{g} g^{MN} \partial_N \phi \right) - m^2 \phi \right]. \quad (147)$$

The second term in (147) vanishes when the equation of motion of ϕ is used. Then, the on-shell action is:

$$S^{on-shell} = -\frac{\eta}{2} \int dz d^d x \partial_M \left[\sqrt{g} \phi g^{MN} \partial_N \phi \right]. \quad (148)$$

Taking into account that the boundary is at $z = \varepsilon$, which is the lower limit of the integration, the on-shell action (148) can be written as:

$$S^{on-shell} = \frac{\eta}{2} \int d^d x \left(\sqrt{g} \phi g^{zz} \partial_z \phi \right)_{z=\varepsilon}. \quad (149)$$

Let us define the “canonical momentum” Π as in (131):

$$\Pi \equiv -\frac{\partial \mathcal{L}}{\partial(\partial_z \phi)} = \eta \sqrt{g} g^{zz} \partial_z \phi. \quad (150)$$

Then the on-shell action of ϕ becomes:

$$S^{on-shell} = \frac{1}{2} \int_{z=\varepsilon} d^d x \Pi(z, x) \phi(z, x). \quad (151)$$

Let us work in momentum space and Fourier transform the field ϕ and the canonical momentum Π :

$$\phi(z, x) = \int \frac{d^d k}{(2\pi)^d} e^{ik \cdot x} f_k(z), \quad \Pi(z, x) = \int \frac{d^d k}{(2\pi)^d} e^{ik \cdot x} \Pi_k(z) \quad (152)$$

Then, Parseval identity allows to rewrite $S^{on-shell}$ as:

$$S^{on-shell} = \frac{1}{2} \int \frac{d^d k}{(2\pi)^d} \Pi_{-k}(z = \varepsilon) f_k(z = \varepsilon). \quad (153)$$

Recalling that the function $f_k(z)$ behaves near $z \sim 0$ as in (98) and the definition of Π , we obtain:

$$\Pi(z, x) \approx \eta L^{d-1} \left[(d - \Delta) A(x) z^{-\Delta} + \Delta B(x) z^{\Delta-d} \right], \quad (z \rightarrow 0), \quad (154)$$

which, in momentum space becomes:

$$\Pi_{-k}(z) \approx \eta L^{d-1} \left[(d - \Delta) A(-k) z^{-\Delta} + \Delta B(-k) z^{\Delta-d} \right], \quad (z \rightarrow 0), \quad (155)$$

Let us use these results to compute the on-shell action, keeping the terms that do not vanish when $\varepsilon \rightarrow 0$. We get:

$$S^{on-shell} = \frac{\eta}{2} L^{d-1} \int \frac{d^d k}{(2\pi)^d} \left[\varepsilon^{-2\nu} (d - \Delta) A(-k) A(k) + d A(-k) B(k) \right]. \quad (156)$$

Notice that the first term in (156) is divergent. We will now find a counterterm to renormalize this divergence of the on-shell action. It must be a local quadratic functional defined at the boundary of AdS . The natural candidate would be a term proportional to:

$$\int_{\partial AdS} d^d x \sqrt{\gamma} \phi^2(\varepsilon, x), \quad (157)$$

where γ is the determinant of the induced metric $\gamma_{\mu\nu}$ on the boundary:

$$ds_{z=\varepsilon}^2 = \gamma_{\mu\nu} dx^\mu dx^\nu = \frac{L^2}{\varepsilon^2} \delta_{\mu\nu} dx^\mu dx^\nu. \quad (158)$$

It is easy to prove that:

$$\int_{\partial AdS} d^d x \sqrt{\gamma} \phi^2(\varepsilon, x) = L^d \int \frac{d^d k}{(2\pi)^d} \left[\varepsilon^{-2\nu} A(-k) A(k) + 2 A(-k) B(k) \right], \quad (159)$$

Let us adjust the coefficient of the counterterm action in such a way that the leading divergence is cancelled. It is immediate to check that this counterterm action must be:

$$S_{ct} = -\frac{\eta}{2} \frac{d-\Delta}{L} \int_{\partial AdS} d^d x \sqrt{\gamma} \phi^2, \quad (160)$$

or, equivalently, in momentum space:

$$S_{ct} = -\frac{\eta}{2} (d-\Delta) L^{d-1} \int \frac{d^d k}{(2\pi)^d} \left[\varepsilon^{-2\nu} A(-k) A(k) + 2 A(-k) B(k) \right]. \quad (161)$$

The renormalized action $S^{ren} = S^{on-shell} + S_{ct}$ can be obtained by adding (156) and (161), with the result:

$$S^{ren} = \frac{\eta}{2} L^{d-1} (2\Delta - d) \int \frac{d^d k}{(2\pi)^d} A(-k) B(k). \quad (162)$$

In order to extract the one-point function from S^{ren} we have to compute the functional derivative with respect to $\varphi(x)$ (recall that $\varphi(x) = A(x)$). However, the coefficient $B(x)$ also depends functionally on $A(x)$. To illustrate this point, let us represent $f_k(z)$ for arbitrary z (not necessarily small) as:

$$f_k(z) = A(k) \phi_1(z, k) + B(k) \phi_2(z, k), \quad (163)$$

where $\phi_1(z, k)$ and $\phi_2(z, k)$ are independent solutions of the equation satisfied by $f_k(z)$, normalized in such a way that for $z \rightarrow 0$ they behave as:

$$\phi_1(z, k) \approx z^{d-\Delta}, \quad \phi_2(z, k) \approx z^\Delta, \quad (164)$$

(the explicit expressions of $\phi_1(z, k)$ and $\phi_2(z, k)$ are given below). To determine completely ϕ we have to impose regularity conditions in the deep IR $z \rightarrow \infty$. As we will see soon, this fixes uniquely the ratio B/A to a value which is independent of the value of the field at the boundary $z = 0$. Let us denote this ratio by χ :

$$\chi = \frac{B}{A}. \quad (165)$$

Clearly, we can write the renormalized action (162) as:

$$S^{ren} = \frac{\eta}{2} L^{d-1} (2\Delta - d) \int \frac{d^d k}{(2\pi)^d} \chi(k) \varphi(k) \varphi(-k), \quad (166)$$

where we have used the fact that $\phi(k) = A(k)$. Then, as the functional derivative $\frac{\delta}{\delta \varphi(x)}$ is equivalent to $(2\pi)^d \frac{\delta}{\delta \varphi(-k)}$ in momentum space, we have:

$$\langle \mathcal{O}(k) \rangle_\varphi = (2\pi)^d \frac{\delta S^{ren}}{\delta \varphi(-k)} = \eta L^{d-1} (2\Delta - d) \chi(k) \varphi(k). \quad (167)$$

Taking into account the definition of χ in (165) and that $2\Delta - d = 2\nu$, we can write (167) as:

$$\langle \mathcal{O}(k) \rangle_\varphi = 2\nu \eta L^{d-1} B(k). \quad (168)$$

Thus, the subleading contribution near the boundary ($B(k)$) determines the VEV of the operator. The two-point function $G_E(k)$ can be obtained by dividing by the source (see 144):

$$G_E(k) = 2\nu \eta L^{d-1} \frac{B(k)}{A(k)}. \quad (169)$$

Let us now calculate explicitly $A(k)$ and $B(k)$. We first define the function $g_k(z)$ as:

$$f_k(z) = z^{\frac{d}{2}} g_k(z). \quad (170)$$

Then, one can easily check from (93) that $g_k(z)$ satisfies the equation:

$$z^2 \partial_z^2 g_k + z \partial_z g_k - (\nu^2 + k^2 z^2) g_k = 0. \quad (171)$$

Equation (171) is just the modified Bessel equation, whose two independent solutions can be taken to be $g_k = I_{\pm\nu}(kz)$, where $I_{\pm\nu}$ are modified Bessel functions. Thus, the two independent solutions for $f_k(z)$ are:

$$z^{\frac{d}{2}} I_{\pm\nu}(kz). \quad (172)$$

Notice that for $z \rightarrow 0$ the modified Bessel functions behave as:

$$I_{\pm\nu}(z) \approx \frac{1}{\Gamma(1 \pm \nu)} \left(\frac{z}{2}\right)^{\pm\nu}, \quad (z \rightarrow 0). \quad (173)$$

Then, we can take $\phi_1(z, k)$ and $\phi_2(z, k)$ in (163) as:

$$\phi_1(z, k) = \Gamma(1 - \nu) \left(\frac{k}{2}\right)^{\nu} z^{\frac{d}{2}} I_{-\nu}(kz), \quad \phi_2(z, k) = \Gamma(1 + \nu) \left(\frac{k}{2}\right)^{-\nu} z^{\frac{d}{2}} I_{\nu}(kz). \quad (174)$$

One can check that these functions have, indeed, the correct behavior when $z \rightarrow 0$. Then, by using (163) we get:

$$f_k(z) = z^{\frac{d}{2}} \left[\Gamma(1 - \nu) \left(\frac{k}{2}\right)^{\nu} A(k) I_{-\nu}(kz) + \Gamma(1 + \nu) \left(\frac{k}{2}\right)^{-\nu} B(k) I_{\nu}(kz) \right]. \quad (175)$$

Let us now impose that $f_k(z)$ is finite when $z \rightarrow \infty$. This condition determines a precise relation between the coefficients $A(k)$ and $B(k)$, as already mentioned above. When $z \rightarrow \infty$, the functions $I_{\pm\nu}(z)$ behave as:

$$I_{\pm\nu}(z) \approx \frac{e^z}{\sqrt{2\pi z}}, \quad (z \rightarrow \infty). \quad (176)$$

Then, after changing $z \rightarrow kz$, we get for large z :

$$f_k(z) \approx \frac{z^{\frac{d}{2}} e^{kz}}{\sqrt{2\pi kz}} \left[\Gamma(1 - \nu) \left(\frac{k}{2}\right)^{\nu} A(k) + \Gamma(1 + \nu) \left(\frac{k}{2}\right)^{-\nu} B(k) \right], \quad (177)$$

which diverges when $z \rightarrow \infty$ unless the coefficient in brackets vanishes. Then, we must require:

$$\frac{B(k)}{A(k)} = -\frac{\Gamma(1 - \nu)}{\Gamma(1 + \nu)} \left(\frac{k}{2}\right)^{2\nu} = \frac{\Gamma(-\nu)}{\Gamma(\nu)} \left(\frac{k}{2}\right)^{2\nu}. \quad (178)$$

Using this result we can compute the euclidean two-point function. We get:

$$G_E(k) = 2\nu \eta L^{d-1} \frac{\Gamma(-\nu)}{\Gamma(\nu)} \left(\frac{k}{2}\right)^{2\nu}. \quad (179)$$

Let us write this result in position space. The relation between $G_E(x)$ and $G_E(k)$ is:

$$G_E(x) = \int \frac{d^d k}{(2\pi)^d} e^{ikx} G_E(k). \quad (180)$$

We now use the formula:

$$\int \frac{d^d k}{(2\pi)^d} e^{ikx} k^n = \frac{2^n}{\pi^{\frac{d}{2}}} \frac{\Gamma(\frac{d+n}{2})}{\Gamma(-\frac{n}{2})} \frac{1}{|x|^{d+n}}, \quad (181)$$

to obtain the correlator in momentum space:

$$\langle \mathcal{O}(x) \mathcal{O}(0) \rangle = \frac{2\nu\eta L^{d-1}}{\pi^{\frac{d}{2}}} \frac{\Gamma(\frac{d}{2} + \nu)}{\Gamma(-\nu)} \frac{1}{|x|^{2\Delta}}. \quad (182)$$

The behavior $\langle \mathcal{O}(x) \mathcal{O}(0) \rangle \sim |x|^{-2\Delta}$ in (182) confirms that Δ is, indeed, the scaling dimension of the operator $\mathcal{O}(x)$.

10 Quark–Antiquark Potential

Let us consider an external charge moving along a closed curve \mathcal{C} in spacetime in QED. The action for such a charge is:

$$S_{\mathcal{C}} = \oint_{\mathcal{C}} A_\mu dx^\mu. \quad (183)$$

Adding this term to the action is equivalent to insert in the path integral the quantity:

$$e^{iS_{\mathcal{C}}} = e^{i\oint_{\mathcal{C}} A_\mu dx^\mu} \equiv W(\mathcal{C}), \quad (184)$$

where $W(\mathcal{C})$ is the so-called Wilson loop, which is just the holonomy of the gauge field A_μ along the closed curve \mathcal{C} . Notice that $W(\mathcal{C})$ is the Aharonov-Bohm phase factor due to the propagation of a quark along the closed curve \mathcal{C} . The non-abelian analogue of the previous formula is:

$$W(\mathcal{C}) = \text{Tr } P \exp \left[i \oint_{\mathcal{C}} A_\mu dx^\mu \right], \quad (185)$$

where $A_\mu = A_\mu^a T^a$, Tr is the trace over the group indices and P is the path ordering operator. Notice that $\langle W(\mathcal{C}) \rangle$ can be regarded as the amplitude for a creation of a $q\bar{q}$ pair which propagates and is annihilated afterwards.

In particular, we can take a rectangular Wilson loop in Euclidean space. Let T and d be the sizes of the rectangle. In this case the Wilson loop is the amplitude for the propagation of a $q\bar{q}$ pair separated a distance d , which can be calculated by means of the hamiltonian formalism. Actually, when $T \rightarrow \infty$ one has:

$$\lim_{T \rightarrow \infty} \langle W(\mathcal{C}) \rangle \sim e^{-T E(d)}, \quad (186)$$

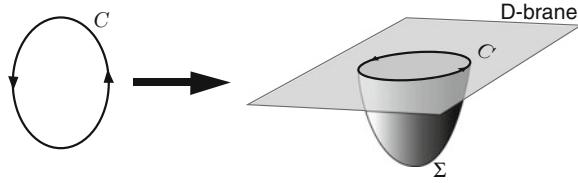


Fig. 10 A Wilson loop along a closed curve \mathcal{C} is holographically realized as the boundary of a surface Σ

where $E(d)$ is the energy of a $q\bar{q}$ pair separated a distance d . In a confining theory the energy grows linearly with the $q\bar{q}$ distance d :

$$E(d) \approx \sigma d, \quad \sigma \rightarrow \text{constant}. \quad (187)$$

Then, the Wilson loop in a confining theory satisfies the so-called area law:

$$\lim_{T \rightarrow \infty} \langle W(\mathcal{C}) \rangle \sim e^{-\sigma T d} \sim e^{-\sigma (\text{area enclosed by the loop})}. \quad (188)$$

Let us see how we can compute VEVs of Wilson loops in the AdS/CFT correspondence [22, 23]. Recall that the endpoint of an open string ending on a D-brane is dual to a quark. Thus, in string theory, the Wilson loop is represented by an open string whose worldsheet Σ has a boundary $\partial\Sigma$ which is in a D-brane and describes the curve \mathcal{C} , as shown in Fig. 10. Then:

$$\langle W(\mathcal{C}) \rangle = Z_{\text{string}}(\partial\Sigma = \mathcal{C}). \quad (189)$$

To have infinitely heavy (non-dynamical) quarks we push the D-brane to the AdS boundary and, therefore, $\partial\Sigma = \mathcal{C}$ lies within the AdS boundary $z = 0$. Then, in the 't Hooft limit, we have:

$$Z_{\text{string}}(\partial\Sigma = \mathcal{C}) = e^{-S(\mathcal{C})}, \quad (190)$$

where $S(\mathcal{C})$ is the on-shell extremal Nambu-Goto action for the string worldsheet satisfying the boundary condition that Σ ends on the curve \mathcal{C} . The Wilson loop becomes simply:

$$\langle W(\mathcal{C}) \rangle = e^{-S(\mathcal{C})}. \quad (191)$$

We will now apply these ideas to the holographic calculation of the rectangular Wilson loop and the corresponding $q\bar{q}$ potential energy in $\mathcal{N} = 4$ SYM. With this purpose, let us consider a string ending on the boundary of AdS_5 , where it extends along one of the gauge theory coordinates (say $x^1 \equiv x$). We shall parameterize the embedding of the string in AdS_5 by a function $z = z(x)$ and we will denote by z' the

derivative of z with respect to x . The induced metric in Euclidean signature on the string worldsheet is:

$$ds^2 = \frac{L^2}{z^2} [dt^2 + (1 + z'^2) dx^2]. \quad (192)$$

Therefore, the Nambu-Goto action for the string is:

$$S = \frac{1}{2\pi\alpha'} \int dt \int dx \sqrt{g} = \frac{TL^2}{2\pi\alpha'} \int dx \frac{\sqrt{1+z'^2}}{z^2}, \quad (193)$$

where $T = \int dt$. The action (193) corresponds to a problem in classical mechanics where x is the “time” and the lagrangian is given by:

$$\mathcal{L} = \frac{TL^2}{2\pi\alpha'} \frac{\sqrt{1+z'^2}}{z^2}. \quad (194)$$

Since this lagrangian does not depend on the “time” x , the “energy” is conserved, which means that:

$$z' \frac{\partial \mathcal{L}}{\partial z'} - \mathcal{L} = \text{constant}. \quad (195)$$

By computing the derivative of \mathcal{L} appearing on the left-hand side of (195), the conservation law written above can be proved to be equivalent to the following first integral of the equation of motion:

$$z^2 \sqrt{1+z'^2} = \text{constant}. \quad (196)$$

For the hanging string configuration we are interested in, the boundary conditions that the function $z(x)$ must satisfy are:

$$z(x = -d/2) = z(x = d/2) = 0, \quad (197)$$

where d is the quark–antiquark separation. Clearly, there is a value of x for which z is maximal. By symmetry this value is just $x = 0$ and, since $z(x)$ has a maximum implies that $z'(x = 0) = 0$. Let z_* be the maximal value of z , i.e. $z_* = z(x = 0)$. Then, the constant appearing in the conservation law is just z_*^2 and we have:

$$z' = \pm \frac{\sqrt{z_*^4 - z^4}}{z^2}, \quad (198)$$

where the two signs correspond to the two sides of the hanging string. The equation written above can be integrated immediately to give x as a function of z :

$$x = \pm \int_{z_*}^z \frac{\xi^2}{\sqrt{z_*^4 - \xi^4}} d\xi. \quad (199)$$

By introducing a new rescaled variable y by means of the relation $\xi = z_* y$, one can write:

$$x = \pm z_* \int_1^{\frac{z}{z_*}} \frac{y^2}{\sqrt{1-y^4}} dy. \quad (200)$$

By imposing the boundary conditions (197), we can obtain the quark–antiquark separation as a function of z_* :

$$\frac{d}{2} = z_* \int_0^1 \frac{y^2}{\sqrt{1-y^4}} dy. \quad (201)$$

The integral over y in (201) can be computed exactly and one gets:

$$z_* = \frac{d}{2\sqrt{2}\pi^{\frac{3}{2}}} \left(\Gamma\left(\frac{1}{4}\right) \right)^2. \quad (202)$$

Then, it follows that d and z_* are proportional. Let us now compute the on-shell action for the string, from which we will obtain the quark–antiquark potential. By plugging the conservation law (196) inside the expression of the action in (193), we get:

$$S = \frac{T L^2 z_*^2}{2\pi\alpha'} \int \frac{dx}{z^4}. \quad (203)$$

Let us change variables in the integral (203) from x to z . The jacobian for this change of variables is:

$$\frac{dx}{dz} = \frac{1}{z'} = \frac{z^2}{\sqrt{z_*^4 - z^4}}, \quad (204)$$

where we have taken into account the value of z' written in (198). Taking into account that the variable z is double-valued, the total action is:

$$S = 2 \times \frac{T L^2 z_*^2}{2\pi\alpha'} \int_{\varepsilon}^{z_*} \frac{dz}{z^2 \sqrt{z_*^4 - z^4}}. \quad (205)$$

Let us next work on a rescaled variable y , related to z as $z = z_* y$. Then, the on-shell action becomes:

$$S = \frac{T L^2}{\pi\alpha' z_*} I_{\varepsilon}, \quad (206)$$

where I_ε is the following integral:

$$I_\varepsilon = \int_{\varepsilon/z_*}^1 \frac{dy}{y^2 \sqrt{1-y^4}}. \quad (207)$$

The integral I_ε in (207) diverges when $\varepsilon \rightarrow 0$. Indeed, one can prove that, for small ε :

$$I_\varepsilon = -\frac{\pi^{\frac{3}{2}} \sqrt{2}}{\left(\Gamma\left(\frac{1}{4}\right)\right)^2} + \frac{z_*}{\varepsilon}. \quad (208)$$

The on-shell action S is just TE , where E is the energy of the quark–antiquark. It follows that E is given by:

$$E = -\frac{4\pi^2 L^2}{\left(\Gamma\left(\frac{1}{4}\right)\right)^4 \alpha'} \frac{1}{d} + \frac{L^2}{\pi \alpha'} \frac{1}{\varepsilon}. \quad (209)$$

The divergent term in (209) corresponds to the quark and antiquark masses (which are considered to be infinitely large in the static limit). To check this, let us compute the euclidean action of a single string which goes straight from the boundary $z = \varepsilon$ to $z = \infty$ at fixed x . This configuration can be described by using t and z as worldvolume coordinates. The induced metric is obtained by keeping constant the x coordinate and is given by:

$$ds^2 = \frac{L^2}{z^2} (dt^2 + dz^2). \quad (210)$$

The on-shell Nambu-Goto action for this configuration is:

$$S_{|} = \frac{T}{2\pi \alpha'} \int_{\varepsilon}^{\infty} \frac{L^2}{z^2} dz^2 = \frac{TL^2}{2\pi \alpha' \varepsilon}. \quad (211)$$

Therefore, the energy for a configuration of two straight parallel strings in AdS_5 is:

$$E_{||} = 2 \times \frac{L^2}{2\pi \alpha' \varepsilon}, \quad (212)$$

which is equal to the divergent term in (209), as claimed. The quark–antiquark potential $V_{q\bar{q}}$ is obtained by subtracting the divergent contribution due to the quark masses:

$$V_{q\bar{q}} = E - E_{||}. \quad (213)$$

By subtracting (209) and (212), one gets:

$$V_{q\bar{q}} = -\frac{4\pi^2 L^2}{\left(\Gamma\left(\frac{1}{4}\right)\right)^4 \alpha'} \frac{1}{d}. \quad (214)$$

Let us write this result in terms of gauge theory quantities. Recall that $L^2 = \sqrt{N^2 g_{YM}} \alpha'$ or, equivalently $L^2 = \sqrt{\lambda} \alpha'$. Then, (214) can be rewritten as:

$$V_{q\bar{q}} = -\frac{4\pi^2 \sqrt{\lambda}}{\left(\Gamma\left(\frac{1}{4}\right)\right)^4} \frac{1}{d}. \quad (215)$$

The Coulombic $1/d$ dependence in (215) is a consequence of the conformal invariance of the theory. The non-analytic dependence on the coupling λ is a non-perturbative effect. In field theory this non-analyticity results from the summation of an infinite number of Feynman diagrams. It is interesting to compare (215) with the perturbative $q\bar{q}$ potential, valid for small λ , which is given by:

$$V_{q\bar{q}} = -\frac{\pi\lambda}{d}. \quad (216)$$

Remarkably we have been able to find a non-perturbative result in a interacting quantum field theory by studying the catenary curve for a hanging string in classical gravity.

In the next two sections we will generalize the result (215) to the case of $\mathcal{N} = 4$ SYM at finite temperature and to the gravity dual of a theory in which the quarks are confined. In the former case we have to study a string hanging from the boundary of an anti-de-Sitter black hole, whereas in the latter we must deal with a geometry obtained from the AdS black hole by a double analytic continuation.

10.1 Quark–Antiquark Potential at Finite Temperature

Let us suppose that we put our gauge theory in a thermal bath at a temperature T . In the holographic correspondence the heat bath is dual to a bulk geometry with an event horizon (a black hole) and the Hawking temperature of the black hole is identified as the temperature of heat bath. In particular, the dual of $\mathcal{N} = 4$ SYM at finite temperature is a black hole in AdS_5 , whose metric in euclidean space is (see below):

$$ds^2 = \frac{L^2}{z^2} \left[f(z) dt^2 + d\mathbf{x}^2 + \frac{dz^2}{f(z)} \right], \quad (217)$$

with $f(z)$ being the following function:

$$f(z) = 1 - \frac{z^4}{z_0^4}. \quad (218)$$

The parameter z_0 corresponds to the position of the horizon and is related to the black hole temperature T as $T = (\pi z_0)^{-1}$ (see Sect. 11.2). In the context of the holographic duality one naturally identifies the black hole temperature T with the temperature of the gauge theory.

As in the zero temperature case, we will parametrize the embedding of the string by a function $z = z(x)$. The induced metric is:

$$ds^2 = \frac{L^2}{z^2} \left[f dt^2 + \left(1 + \frac{z'^2}{f} \right) dx^2 \right]. \quad (219)$$

The Nambu-Goto action for this embedding becomes:

$$S = \frac{\tau L^2}{2\pi\alpha'} \int dx \frac{\sqrt{f(z) + z'^2}}{z^2}, \quad (220)$$

where $\tau = \int dt$. Following the same steps as in the zero temperature case, it is straightforward to derive the following first integral of the equation of motion:

$$\frac{z^2 \sqrt{f(z) + z'^2}}{f(z)} = \text{constant} = \frac{z_*^2}{\sqrt{f(z_*)}}. \quad (221)$$

From (221) we obtain readily the value of z' :

$$z' = \pm \sqrt{\frac{f(z)}{f(z_*)}} \frac{\sqrt{z_*^4 - z^4}}{z^2}. \quad (222)$$

Let us now define the constant ρ as follows:

$$\rho \equiv \left(\frac{z_0}{z_*} \right)^4. \quad (223)$$

Then, we have:

$$x = \pm z_* \sqrt{\rho - 1} \int_1^{\frac{z}{z_*}} \frac{y^2 dy}{\sqrt{(1 - y^4)(\rho - y^4)}}, \quad (224)$$

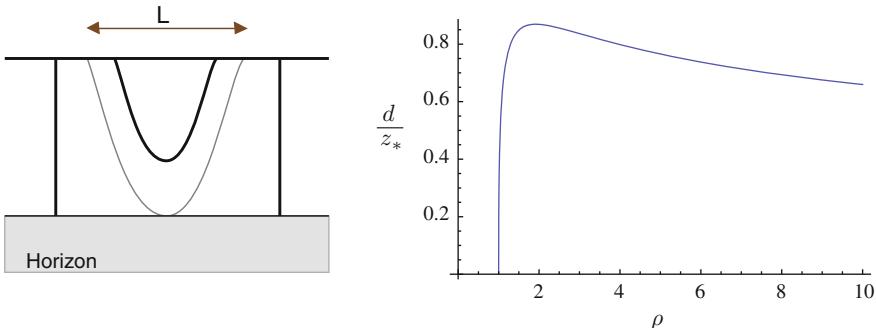


Fig. 11 In a *black hole* metric, the string ends on the horizon when the $q\bar{q}$ separation is large enough. Correspondingly, the distance d reaches a maximum value, as the plot on the right shows

and the quark–antiquark distance d is equal to:

$$d = 2z_* \sqrt{\rho - 1} \int_0^1 \frac{y^2 dy}{\sqrt{(1 - y^4)(\rho - y^4)}}. \quad (225)$$

Notice that these expressions become the ones at zero temperature as $\rho \rightarrow \infty$, as it should. Moreover, as $z_* \rightarrow z_0$, $\rho \rightarrow 1$ and the distance $d \rightarrow 0$ (see Fig. 11). Actually, by varying ρ one can see that there is a maximal value of d ($d_{\max} \sim z_0$) and the $q\bar{q}$ bound state becomes unbound due to thermal screening, in agreement with the behavior expected from the point of view of the gauge theory at non-zero temperature.

10.2 Quark–Antiquark Potential in a Confining Background

Let us consider the AdS black hole in Euclidean signature and let us go back to Minkowski signature by analytic continuation along a different direction, namely by making $x_3 \rightarrow it$. If we call u to the original euclidean time, we arrive at the following metric:

$$ds^2 = \frac{L^2}{z^2} \left[-dt^2 + dx_1^2 + dx_2^2 + f(z)du^2 + \frac{dz^2}{f(z)} \right]. \quad (226)$$

In this case the space ends smoothly at $z = z_0$, which should be thought as a IR mass scale. We will now compute the $q\bar{q}$ potential for the metric (226) and we will verify that it corresponds to a confining background. With this purpose, let us consider a fundamental string in the euclidean version of the above metric and let us use (t, x) as worldsheet coordinates. The induced metric is:

$$ds^2 = \frac{L^2}{z^2} \left[dt^2 + \left(1 + \frac{z'^2}{f}\right) dx^2 \right], \quad (227)$$

and the Nambu-Goto action becomes:

$$S = \frac{\tau L^2}{2\pi\alpha'} \int \frac{dx}{z^2} \sqrt{1 + \frac{z'^2}{f(z)}}, \quad (228)$$

where $\tau = \int dt$. The first integral corresponding to the action (228) is:

$$\frac{z^2}{\sqrt{f(z)}} \sqrt{f(z) + z'^2} = z_*^2, \quad (229)$$

from which we get:

$$z' = \pm \sqrt{f(z)} \frac{\sqrt{z_*^4 - z^4}}{z^2}. \quad (230)$$

This equation can be integrated as:

$$x = \pm z_* \sqrt{\rho} \int_1^{\frac{z}{z_*}} \frac{y^2 dy}{\sqrt{(1-y^4)(\rho-y^4)}}, \quad (231)$$

where the constant ρ is the same as in (223). It follows that the $q\bar{q}$ distance d is:

$$d = 2z_* \sqrt{\rho} \int_0^1 \frac{y^2 dy}{\sqrt{(1-y^4)(\rho-y^4)}}. \quad (232)$$

In this case, d grows without limit as $\rho \rightarrow 1$, or equivalently as the turning point approaches the end of the space (i.e. when $z_* \rightarrow z_0$) (see Fig. 12). In the large d limit the profile of the hanging string is approximately rectangular. The energy due to the vertical parts of the profile can be identified with the masses of the static quarks, which have to be subtracted to get the potential energy. The $q\bar{q}$ potential is just due to the horizontal part of the profile. Since in this part z is approximately constant and equal to z_0 , we get that the contribution to the euclidean action is:

$$S_{horizontal} = \frac{\tau L^2}{2\pi\alpha'} \frac{d}{z_0^2}, \quad (233)$$

which corresponds to an area law (it is proportional to τd) and gives rise to a confining potential of the type:

$$V = \sigma_s d, \quad (234)$$

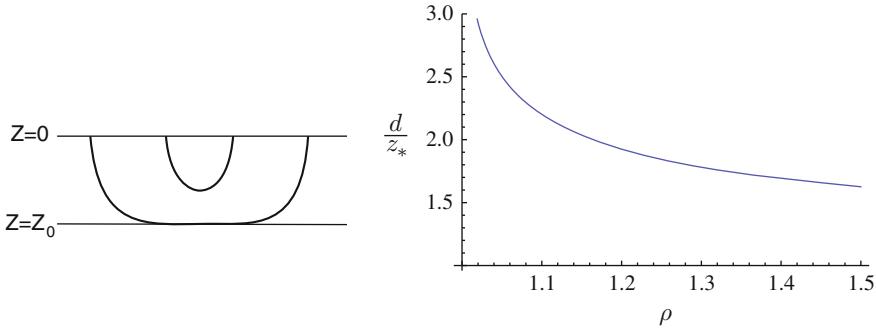


Fig. 12 In a confining metric the space ends smoothly and the profile of the hanging string becomes almost rectangular for large $q\bar{q}$ separation (left). The distance d is not bounded, as shown on the plot on the right

with σ_s being the effective string tension, given by:

$$\sigma_s = \frac{L^2}{2\pi\alpha'} \frac{1}{z_0^2}. \quad (235)$$

In terms of gauge theory quantities, σ_s can be written as:

$$\sigma_s = \frac{\sqrt{\lambda}}{2\pi z_0^2}. \quad (236)$$

Notice that $M \sim 1/z_0$ is an IR scale that can be identified with the mass gap of the theory (and z_0 with the glueball size). The previous formula for the tension is simply:

$$\sigma_s \sim \sqrt{\lambda} M^2. \quad (237)$$

11 Black Hole Thermodynamics

The partition function in statistical mechanics in the canonical ensemble is given by:

$$Z = \text{Tr } e^{-\frac{H}{T}}, \quad (238)$$

where, H is the hamiltonian operator, T is the temperature and we are taking the Boltzmann constant $k_B = 1$. The thermal average of an operator \mathcal{O} at the temperature T is:

$$\langle \mathcal{O} \rangle_T = \frac{\text{Tr}[\mathcal{O} e^{-\frac{H}{T}}]}{Z}. \quad (239)$$

In the path integral approach, the average $\langle \mathcal{O} \rangle_T$ can be written as:

$$\langle \mathcal{O} \rangle_T \sim \int [D\psi] \langle \psi(x), t | \mathcal{O} e^{-\frac{H}{T}} | \psi(x), t \rangle, \quad (240)$$

where the expectation value is taken between the same initial and final state $|\psi(x), t\rangle$ (as it corresponds to a trace in Hilbert space). Equivalently, since the hamiltonian operator implements time evolution, we can rewrite (240) as:

$$\langle \mathcal{O} \rangle_T \sim \int [D\psi] \langle \psi(x), t | \mathcal{O} | \psi(x), t + \frac{i}{T} \rangle. \quad (241)$$

Then, to perform thermal averages we have to consider imaginary time evolution and we have to impose periodic boundary conditions in the Hilbert space (antiperiodic for fermions). Then, the euclidean time t_E must be periodically identified

$$t_E \equiv t_E + \frac{1}{T}. \quad (242)$$

Thus, the compactification of Euclidean time is equivalent to having $T \neq 0$.

Let us apply these ideas to obtain the Hawking temperature of a black hole. We will assume that we have an euclidean metric of the type:

$$ds^2 = g(r) [f(r) dt_E^2 + d\mathbf{x}^2] + \frac{1}{h(r)} dr^2, \quad (243)$$

where the functions $f(r)$ and $h(r)$ have a first-order zero at $r = r_0$, which is the location of the horizon, and $g(r_0) \neq 0$. Then, for $r \approx r_0$

$$f(r) \approx f'(r_0)(r - r_0), \quad h(r) \approx h'(r_0)(r - r_0), \quad (244)$$

while we can take $g(r) = g(r_0)$. Then, the near-horizon euclidean metric can be written as:

$$ds^2 \approx g(r_0) [f'(r_0)(r - r_0) dt_E^2 + d\mathbf{x}^2] + \frac{1}{h'(r_0)} \frac{dr^2}{r - r_0}. \quad (245)$$

Let us define a new radial variable ρ such that:

$$\frac{1}{h'(r_0)} \frac{dr^2}{r - r_0} = d\rho^2, \quad (246)$$

which can be integrated to give the following relation between ρ and r :

$$\rho = 2 \sqrt{\frac{r - r_0}{h'(r_0)}}. \quad (247)$$

Next, we define an angular coordinate θ such that

$$g(r_0) f'(r_0) (r - r_0) dt_E^2 = \rho^2 d\theta^2, \quad (248)$$

which is equivalent to defining θ as:

$$\theta = \frac{1}{2} \sqrt{g(r_0) f'(r_0) h'(r_0)} t_E. \quad (249)$$

In the new variables, the (t_E, r) part of the metric takes the form $d\rho^2 + \rho^2 d\theta^2$, which is locally like the flat metric of a plane. In order to have $\rho = 0$ (i.e. the horizon) as a regular point without any curvature singularity, the variable θ must be a periodic variable with period 2π . Otherwise we would have a conical singularity at the origin, due to the deficit angle. We have argued above for a general system that the compactification of the euclidean time is equivalent to having a non-zero temperature T . Thus, it follows that we can assign a temperature T to a black hole (the Hawking temperature). In order to find the value of T , let us notice that the periodicity under $\theta \rightarrow \theta + 2\pi$ is equivalent to periodicity under $t_E \rightarrow t_E + \frac{1}{T}$, where T is the Hawking temperature given by:

$$\frac{1}{T} = \frac{4\pi}{\sqrt{g(r_0) f'(r_0) h'(r_0)}}. \quad (250)$$

In the next subsections we will apply this formula to determine the Hawking temperature of a couple of black holes.

11.1 Application to the Schwarzschild Black Hole

Let us consider the ordinary Schwarzschild metric in four dimensions:

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \frac{dr^2}{1 - \frac{2GM}{r}} + r^2 d\Omega_2^2, \quad (251)$$

which is a particular case of the general expression written in (243), with the functions f , h and g being given by:

$$g(r) = 1, \quad f(r) = h(r) = 1 - \frac{2GM}{r}. \quad (252)$$

In (251) G is the four-dimensional Newton constant and M is the mass of the black hole. The horizon in the geometry (251) is located at $r = r_0 = 2GM$. Since:

$$f'(r_0) = h'(r_0) = \frac{2GM}{r_0^2} \quad (253)$$

it follows that the Hawking temperature for the Schwarzschild black hole is given by:

$$T = \frac{1}{8\pi GM}. \quad (254)$$

Let us use the expression of T in (254) to obtain the black hole entropy S of the Schwarzschild black hole. We will identify the mass M with the internal energy and we will make use of the first law of thermodynamics:

$$dM = TdS = \frac{1}{8\pi GM} dS, \quad (255)$$

which can be integrated to give the entropy S :

$$S = 4\pi GM^2, \quad (256)$$

The horizon area A_H for the metric (251) is the area of the surface $t = \text{constant}$, $r = r_0$. It is straightforward to demonstrate that A_H is:

$$A_H = 4\pi r_0^2 = 16 G^2 M^2. \quad (257)$$

It is now immediate to check that the relation between S and A_H is:

$$S = \frac{A_H}{4G}, \quad (258)$$

which is noting but the celebrated Bekenstein-Hawking entropy formula which relates the entropy of a black hole with the area of its horizon (see 6).

II.2 AdS Black Hole

The metric of a black hole in AdS_{d+1} is:

$$ds^2 = \frac{L^2}{z^2} \left[f(z) dt_E^2 + d\mathbf{x}^2 + \frac{dz^2}{f(z)} \right], \quad (259)$$

where $f(z)$ is the following function:

$$f(z) = 1 - \frac{z^d}{z_0^d}, \quad (260)$$

with z_0 being a constant ($z = z_0$ is the location of the horizon). This metric is just of our general form (243) with $r \rightarrow z$, f given as above and g and h being:

$$g = \frac{L^2}{z^2}, \quad h = \frac{z^2}{L^2} f. \quad (261)$$

In this case the derivatives of f and h at the horizon are:

$$f'(z_0) = -\frac{d}{z_0}, \quad h'(z_0) = -\frac{dz_0}{L^2}, \quad (262)$$

and, therefore, we have:

$$g(z_0) f'(z_0) h'(z_0) = \frac{d^2}{z_0^2}. \quad (263)$$

Thus, the Hawking temperature is related to z_0 by means of the relation:

$$T = \frac{d}{4\pi z_0}. \quad (264)$$

The horizon is the hypersurface $z = z_0$ and t constant, whose area in the metric (259) is:

$$A_H = \left(\frac{L}{z_0}\right)^{d-1} V_{d-1}, \quad (265)$$

where V_{d-1} is the volume along the Minkowski spacial directions. In terms of the temperature, A_H can be written as:

$$A_H = \left(\frac{4\pi}{d}\right)^{d-1} L^{d-1} T^{d-1} V_{d-1}, \quad (266)$$

and the entropy can be computed from the Bekenstein-Hawking formula, with the result:

$$S = \frac{A_H}{4G_{d+1}} = \frac{1}{4G_{d+1}} \left(\frac{4\pi}{d}\right)^{d-1} L^{d-1} T^{d-1} V_{d-1}. \quad (267)$$

We now define the entropy density s as:

$$s = \frac{S}{V_{d-1}}. \quad (268)$$

Let us write s in terms of the QFT central charge $c_{QFT} = \frac{1}{4}(L/l_P)^{d-1}$. We get:

$$s = \left(\frac{4\pi}{d}\right)^{d-1} c_{QFT} T^{d-1}. \quad (269)$$

In the case of $\mathcal{N} = 4$ SYM, we take $d = 4$ and $c_{SYM} = N^2/2\pi$ and the entropy density is given by:

$$s_{SYM} = \frac{\pi^2}{2} N^2 T^3. \quad (270)$$

From the expression of the entropy density in (270) we can obtain the value of the pressure by means of the thermodynamic relation:

$$s = \frac{\partial p}{\partial T}. \quad (271)$$

We get:

$$p = \frac{\pi^2}{8} N^2 T^4. \quad (272)$$

Moreover, the energy density ε can be obtained from p and s by means of the standard thermodynamic relation:

$$\varepsilon = -p + Ts. \quad (273)$$

Using the values of p and s computed from holography, we arrive at the following value of the energy density:

$$\varepsilon = \frac{3\pi^2}{8} N^2 T^4. \quad (274)$$

11.3 Comparison with Field Theory

Let us compare the strong coupling values of s , p and ε found above with those corresponding the $\mathcal{N} = 4$ SYM at zero coupling. The partition function in the canonical ensemble for a gas of non-interacting relativistic bosons and fermions is

$$\log Z = \mp V_3 \int \frac{d^3 p}{(2\pi)^3} \log \left(1 \mp e^{-\frac{\omega(p)}{T}} \right), \quad (275)$$

where the upper minus (lower plus) signs corresponds to bosons (respectively, fermions) and $\omega(p) = \sqrt{\mathbf{p}^2 + m^2}$. In the massless case, we just take $\omega(p) = |\mathbf{p}|$ and we get:

$$\frac{\log Z}{V_3} = \mp \int_0^\infty \frac{dp}{2\pi^2} p^2 \log \left(1 \mp e^{-\frac{p}{T}} \right). \quad (276)$$

Let us change variables in the integral (276) from p to $x = p/T$. After integrating by parts we find:

$$\frac{\log Z}{V_3} = \frac{T^3}{6\pi^2} \int_0^\infty dx \frac{x^3}{e^x \mp 1}. \quad (277)$$

To calculate these integrals we use the general results, valid for $n \in \mathbb{Z}$:

$$\int_0^\infty dx \frac{x^{2n-1}}{e^x + 1} = \frac{2^{2n-1} - 1}{2n} \pi^{2n} B_n, \quad \int_0^\infty dx \frac{x^{2n-1}}{e^x - 1} = \frac{(2\pi)^{2n} B_n}{4n}, \quad (278)$$

where B_n denotes the Bernoulli numbers. In particular for $n = 2$, since $B_2 = 1/30$, we have:

$$\int_0^\infty dx \frac{x^3}{e^x - 1} = \frac{\pi^4}{15}, \quad \int_0^\infty dx \frac{x^3}{e^x + 1} = \frac{7\pi^4}{120}. \quad (279)$$

Thus, for bosons:

$$\frac{\log Z}{V_3} = \frac{\pi^2}{90} T^3, \quad (\text{bosons}), \quad (280)$$

while for fermions:

$$\frac{\log Z}{V_3} = \frac{7\pi^2}{720} T^3, \quad (\text{fermions}). \quad (281)$$

The entropy density can be obtained from the partition function by means of the standard statistical mechanics relation:

$$s = \frac{\partial}{\partial T} \left[T \frac{\log Z}{V_3} \right] = 4 \frac{\log Z}{V_3}. \quad (282)$$

Then, it follows that:

$$s_{\text{boson}} = \frac{2\pi^2}{45} T^3, \quad s_{\text{fermion}} = \frac{7\pi^2}{180} T^3. \quad (283)$$

In $\mathcal{N} = 4$ SYM the number of bosons we have is:

$$[2(\text{gauge field}) + 6(\text{scalar field})] N^2 = 8 N^2, \quad (284)$$

while the number of fermions is:

$$[2 \times 4(\text{Weyl spinors})] N^2 = 8 N^2, \quad (285)$$

which, due to the supersymmetry, is the same as the number of bosons. Therefore, the total entropy density of $\mathcal{N} = 4$ SYM when the coupling constant is zero is:

$$s_{\mathcal{N}=4 \text{ free gas}} = 8 N^2 \left[\frac{2\pi^2}{45} + \frac{7\pi^2}{180} \right] T^3 = \frac{2\pi^2}{3} N^2 T^3. \quad (286)$$

By using the thermodynamic formulas (273) and (271), we obtain the values of the pressure and energy density:

$$p_{\mathcal{N}=4 \text{ free gas}} = \frac{\pi^2}{6} N^2 T^4, \quad \varepsilon_{\mathcal{N}=4 \text{ free gas}} = \frac{\pi^2}{2} N^2 T^4. \quad (287)$$

Then, by comparing with the result given by the black hole, we obtain:

$$s_{\text{black hole}} = \frac{3}{4} s_{\mathcal{N}=4 \text{ free gas}}, \quad (288)$$

and similarly for the pressure and energy density:

$$p_{\text{black hole}} = \frac{3}{4} p_{\mathcal{N}=4 \text{ free gas}}, \quad \varepsilon_{\text{black hole}} = \frac{3}{4} \varepsilon_{\mathcal{N}=4 \text{ free gas}}. \quad (289)$$

Therefore, the AdS/CFT correspondence predicts that the values of s , p and ε at infinite coupling differ from their values at zero coupling by a multiplicative factor $3/4$. These reductions of the entropy, pressure and energy density as the coupling is increased are in very good agreement with the results obtained in lattice simulations for different gauge theories.

12 Transport Coefficients

Let us begin by writing the linear response formulas of Sect. 9.2 in real time. We consider a system in QFT to which we couple a source $\varphi(x)$ to a local operator $\mathcal{O}(x)$:

$$S = S_0 + \int d^d x \mathcal{O}(x) \varphi(x), \quad (290)$$

where the d -dimensional spacetime has Minkowski signature. We will assume that the unperturbed VEV of the operator \mathcal{O} vanishes. Then, the one-point function of the operator \mathcal{O} in the presence of the source is given by:

$$\langle \mathcal{O}(x) \rangle_\varphi = - \int G_R(x - y) \varphi(y) dy , \quad (291)$$

where $G_R(x - y)$ is the retarded Green's function, defined as:

$$i G_R(x - y) \equiv \theta(x^0 - y^0) \langle [\mathcal{O}(x), \mathcal{O}(y)] \rangle . \quad (292)$$

The fact that the linear response is determined by the retarded correlator is a consequence of causality since the source can influence the system only after it has been turned on. In momentum space the relation (291) between the one-point function and the source becomes:

$$\langle \mathcal{O}(\omega, \mathbf{k}) \rangle_\varphi = -G_R(\omega, \mathbf{k}) \varphi(\omega, \mathbf{k}) . \quad (293)$$

We are interested in analyzing the long wavelength hydrodynamic limit. In this case one can take the zero spatial momentum and zero frequency limit of the retarded correlator. In this limit one studies the response of the system to a time varying source $\varphi(t)$, which can be approximately represented as:

$$\langle \mathcal{O} \rangle_\varphi \approx -\chi \partial_t \varphi , \quad (294)$$

where the real constant χ is the so-called transport coefficient. In frequency space, for $\omega \rightarrow 0$, the relation (294) is given by:

$$\langle \mathcal{O} \rangle_\varphi \approx i \omega \chi \varphi(\omega) . \quad (295)$$

The linear response equation for this quantity is obtained from the long wavelength limit $\mathbf{k} \rightarrow 0$ of (293):

$$\langle \mathcal{O} \rangle_\varphi = -G_R(\omega, \mathbf{k} = 0) \varphi(\omega) . \quad (296)$$

By comparing (295) and (296) we get:

$$G_R(\omega, \mathbf{k} = 0) = -i\omega \chi , \quad (\omega \rightarrow 0) , \quad (297)$$

or, taking the imaginary part:

$$\text{Im } G_R(\omega, \mathbf{k} = 0) = -\omega \chi , \quad (\omega \rightarrow 0) . \quad (298)$$

From (298) we immediately get the so-called Kubo formula for χ :

$$\chi = - \lim_{\omega \rightarrow 0} \lim_{\mathbf{k} \rightarrow 0} \frac{1}{\omega} \text{Im} G_R(\omega, \mathbf{k}) . \quad (299)$$

Let us now see how we can make use of (299) to compute the transport coefficient χ by using holographic methods. Let us consider a $(d+1)$ -dimensional metric of the form:

$$ds^2 = g_{tt} dt^2 + g_{zz} dz^2 + g_{xx} \delta_{ij} dx^i dx^j . \quad (300)$$

We will assume the metric (300) has an horizon at $z = z_0$ and that g_{tt} and that g_{zz} behave near $z = z_0$ as:

$$g_{tt} \approx -c_0(z_0 - z) , \quad g_{zz} \approx \frac{c_z}{z_0 - z} , \quad z \rightarrow z_0 , \quad (301)$$

with c_0 and c_z being constants. Let us consider a massless scalar field ϕ in this metric such that its action is:

$$S = -\frac{1}{2} \int d^{d+1} x \sqrt{-g} \frac{\partial_M \phi \partial^M \phi}{q(z)} , \quad (302)$$

where the function $q(z)$ is an effective coupling of the mode. The Euler-Lagrange equation of motion derived from (302) is:

$$\partial_M \left(\frac{\sqrt{-g}}{q} g^{MN} \partial_M \phi \right) = 0 . \quad (303)$$

In terms of the canonical momentum Π , defined as (see 131):

$$\Pi \equiv -\frac{\partial \mathcal{L}}{\partial(\partial_z \phi)} = \frac{\sqrt{-g}}{q} g^{zz} \partial_z \phi , \quad (304)$$

the equation of motion (303) becomes:

$$\partial_z \Pi = -\frac{\sqrt{-g}}{q} \left(\frac{\partial_t^2 \phi}{g_{tt}} + \frac{\partial_i^2 \phi}{g_{xx}} \right) . \quad (305)$$

Let us now Fourier transform ϕ and Π in the variables t, \mathbf{x} :

$$\begin{aligned} \phi(z, t, \mathbf{x}) &= \int \frac{d\omega d^{d-1}k}{(2\pi)^d} e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} \phi(z, \omega, \mathbf{k}) , \\ \Pi(z, t, \mathbf{x}) &= \int \frac{d\omega d^{d-1}k}{(2\pi)^d} e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} \Pi(z, \omega, \mathbf{k}) . \end{aligned} \quad (306)$$

Since in this massless case the scaling dimension Δ is equal to d (see 108 for $m = 0$), the one-point function is just the limit of the momentum Π at the boundary $z = 0$ and the boundary field φ is just obtained by taking the limit $z \rightarrow 0$ of the bulk field ϕ , without any multiplicative factor (see 137 and 119). It follows from (295) that the transport coefficient χ is obtained as:

$$\chi = \lim_{k_\mu \rightarrow 0} \lim_{z \rightarrow 0} \text{Im} \left[\frac{\Pi(z, k_\mu)}{\omega \phi(z, k_\mu)} \right] = \lim_{k_\mu \rightarrow 0} \lim_{z \rightarrow 0} \frac{\Pi(z, k_\mu)}{i\omega \phi(z, k_\mu)}, \quad (307)$$

where, in the last step, we used the fact that $\text{Re}\Pi/\omega \rightarrow 0$ as $\omega \rightarrow 0$ (see below). It turns out that evaluating $\Pi/(\omega \phi)$ at the boundary is equivalent to evaluate it at the horizon. In order to prove this, let us consider the general equation:

$$\partial_z [A(z) \partial_z \phi] = B(z) \phi(z). \quad (308)$$

Let us write (308) in hamiltonian form. We first define:

$$P(z) \equiv A(z) \partial_z \phi(z). \quad (309)$$

Then, (308) can be written as:

$$\partial_z P(z) = B(z) \phi(z). \quad (310)$$

Then, one can readily prove that (308) and (310) can be combined in the following first-order Riccati equation:

$$\partial_z \left(\frac{P(z)}{\phi(z)} \right) = B(z) - \frac{1}{A(z)} \left(\frac{P(z)}{\phi(z)} \right)^2. \quad (311)$$

For the equation of motion of the scalar field ϕ in momentum space, the functions $A(z)$ and $B(z)$ are given by:

$$A(z) = \frac{\sqrt{-g}}{q} g^{zz}, \quad B(z) = \frac{\sqrt{-g}}{q} \left[\frac{\omega^2}{g_{tt}} + \frac{\mathbf{k}^2}{g_{xx}} \right], \quad (312)$$

and $P(z) = \Pi(z)$ (see 304). Then, we can write:

$$\partial_z \left[\frac{\Pi}{\omega \phi} \right] = -\omega \left[\frac{q g_{zz}}{\sqrt{-g}} \left(\frac{\Pi}{\omega \phi} \right)^2 - \frac{\sqrt{-g}}{q g_{tt}} \left(1 + \frac{g_{tt}}{g_{xx}} \frac{\mathbf{k}^2}{\omega^2} \right) \right]. \quad (313)$$

The right-hand side of (313) vanishes when the ordered limit $\lim_{\omega \rightarrow 0} \lim_{\mathbf{k} \rightarrow 0}$ is taken. It follows that $\Pi/(\omega \phi)$ is independent of z in this limit and, as claimed above, it can be evaluated at the horizon. Thus, we can write the transport coefficient χ as:

$$\chi = \lim_{k_\mu \rightarrow 0} \lim_{z \rightarrow z_0} \frac{\Pi(z, k_\mu)}{i\omega \phi(z, k_\mu)}. \quad (314)$$

In order to evaluate the right-hand side of (314), let us study the equation of motion near $z = z_0$, where Π becomes:

$$\Pi \approx \frac{1}{c_z} \frac{\sqrt{-g(z_0)}}{q(z_0)} (z_0 - z) \partial_z \phi, \quad (315)$$

and the equation of motion takes the form:

$$\partial_z \left[(z_0 - z) \partial_z \phi(z, k_\mu) \right] + c_z \left[\frac{\omega^2}{c_0(z_0 - z)} - \frac{k^2}{g_{xx}(z_0)} \right] \phi(z, k_\mu) = 0. \quad (316)$$

To find an approximate solution of (316) near the horizon $z = z_0$, we neglect the last term in the previous equation and try to find a solution of the type:

$$\phi = (z_0 - z)^\beta. \quad (317)$$

Plugging the ansatz (317) in (316), we get that the exponent β can take the following two values:

$$\beta = \pm i \sqrt{\frac{c_z}{c_0}} \omega, \quad (318)$$

which correspond to the following two solutions:

$$\phi_\pm \sim (z_0 - z)^{\pm i \sqrt{\frac{c_z}{c_0}} \omega}. \quad (319)$$

Only one of the two solutions in (319) is compatible with causality. Indeed, let us define a new variable r as $z_0 - z = e^r$. In the r variable the horizon is located at $r \rightarrow -\infty$. By inserting the t dependence, the two solutions ϕ_\pm are:

$$\phi_\pm \sim e^{-i(\omega t \mp \Omega r)}, \quad (320)$$

where $\Omega = \omega \sqrt{c_z/c_0}$. Clearly, ϕ_- is an incoming wave at the horizon since if we increase $t \rightarrow t + \varepsilon$, we must decrease r as $r \rightarrow r - \varepsilon \Omega/\omega$ to keep ϕ_- constant. Then, the wave ϕ_- moves towards the horizon $r \rightarrow -\infty$ and is an infalling wave at the horizon. Similarly, ϕ_+ is an outgoing wave at the horizon. Causality on the gravity side, is implemented if we choose our solution to be the infalling one ϕ_- . One can show that this corresponds to having retarded Green's functions on the field theory side. This infalling solution near $z = z_0$ satisfies:

$$\partial_z \phi_- = \sqrt{\frac{g_{zz}(z_0)}{-g_{tt}(z_0)}} i\omega \phi_-, \quad (321)$$

and, therefore, one has:

$$\left. \frac{\Pi}{i\omega\phi_-} \right|_{z_0} = \frac{1}{q(z_0)} \sqrt{\frac{g}{g_{zz}g_{tt}}} \Big|_{z_0}. \quad (322)$$

It follows that the transport coefficient χ is given by:

$$\chi = \frac{1}{q(z_0)} \sqrt{\frac{g}{g_{zz}g_{tt}}} \Big|_{z_0}. \quad (323)$$

Notice that the square root in this last equation is just the area of the horizon A_H divided by the spatial volume V . Then, we can alternatively write the transport coefficient χ as:

$$\chi = \frac{1}{q(z_0)} \frac{A_H}{V}. \quad (324)$$

In particular, for an AdS black hole, the previous expression becomes:

$$\chi = \frac{1}{q(z_0)} \left(\frac{L}{z_0} \right)^{d-1}, \quad (\text{AdS black hole}). \quad (325)$$

More interestingly, one can compare the general formula (324) with the entropy density, as given by the Bekenstein-Hawking formula $s = A_H/(4G_N V)$. By computing the ratio χ/s , we get the simple result:

$$\frac{\chi}{s} = \frac{4G_N}{q(z_0)}. \quad (326)$$

13 Holographic Viscosities

Hydrodynamics can be thought as an effective theory describing the dynamics of a continuous system at large distances and time scales. In order to study the dynamics of the system we have to analyze the energy-momentum tensor $T^{\mu\nu}$ which is conserved ($\partial_\mu T^{\mu\nu} = 0$). We will assume that the system is at local thermal equilibrium and that the state of the system at a given time is determined by the local temperature $T(x)$ and the local fluid velocity $u^\mu(x)$, which satisfies the condition $u_\mu u^\mu = -1$.

The relation between $T^{\mu\nu}$ and u^μ and the thermodynamic functions is expressed by means of the so-called constitutive relations, which for isotropic fluids takes the form:

$$T^{\mu\nu} = (\varepsilon + p) u^\mu u^\nu + p g^{\mu\nu} - \sigma^{\mu\nu}, \quad (327)$$

where ε is the energy density and p is the pressure, while $\sigma^{\mu\nu}$ is the so-called dissipative part of $T^{\mu\nu}$ and depends on the derivatives of $T(x)$ and u^μ . In order to parametrize $\sigma_{\mu\nu}$ at first-order in the derivatives, let us consider a local rest frame in which $u^i(x) = 0$ (and $u^\mu = (1, 0, 0, 0)$). One can choose this frame in such a way that the dissipative corrections to the energy-momentum tensor components $T^{0\mu}$ vanish. Then, $\sigma^{00} = \sigma^{0i} = 0$ or, equivalently, $T^{00} = \varepsilon$, $T^{0i} = 0$. The only non-zero elements of the dissipative energy-momentum tensor are σ_{ij} . At first order in derivatives σ_{ij} can be written as:

$$\sigma_{ij} = \eta(\partial_i u_j + \partial_j u_i - \frac{2}{3} \delta_{ij} \partial_k u_k) + \zeta \delta_{ij} \partial_k u_k, \quad (328)$$

where η is the so-called shear viscosity and ζ is the bulk viscosity.

Let us write $\sigma^{\mu\nu}$ in covariant form. We first define the projector onto the directions perpendicular to u^μ , as:

$$P^{\mu\nu} = g^{\mu\nu} + u^\mu u^\nu. \quad (329)$$

Then, in an arbitrary curved metric $\sigma_{\mu\nu}$ can be written as:

$$\sigma^{\mu\nu} = P^{\mu\alpha} P^{\nu\beta} \left[\eta(\nabla_\alpha u_\beta + \nabla_\beta u_\alpha) + \left(\zeta - \frac{2}{3}\eta \right) g_{\alpha\beta} \nabla \cdot u \right], \quad (330)$$

where the covariant derivatives of the vectors u_β are defined as:

$$\nabla_\alpha u_\beta = \partial_\alpha u_\beta - \Gamma_{\alpha\beta}^\mu u_\mu, \quad (331)$$

with $\Gamma_{\alpha\beta}^\mu$ being the Christoffel symbols, defined as:

$$\Gamma_{\alpha\beta}^\mu = \frac{1}{2} g^{\mu\lambda} \left[\frac{\partial g_{\lambda\alpha}}{\partial x^\beta} + \frac{\partial g_{\lambda\beta}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\lambda} \right]. \quad (332)$$

We will find $\sigma^{\mu\nu}$ as (minus) the one-point function of $T^{\mu\nu}$ in the presence of a metric perturbation of the type $g^{\mu\nu} \rightarrow \eta^{\mu\nu} + h^{\mu\nu}$. By using linear response theory (291), we obtain:

$$\sigma^{\mu\nu}(x) = \int G_R^{\mu\nu,\alpha\beta}(x-y) h_{\alpha\beta}(y) dy, \quad (333)$$

where the retarded correlator is just:

$$i G_R^{\mu\nu,\alpha\beta}(x-y) = \theta(x^0 - y^0) \langle [T^{\mu\nu}(x), T^{\alpha\beta}(y)] \rangle. \quad (334)$$

Let us now consider the following metric perturbation:

$$g_{00}(t, \mathbf{x}) = -1, \quad g_{0i}(t, \mathbf{x}) = 0, \quad g_{ij}(t, \mathbf{x}) = \delta_{ij} + h_{ij}(t). \quad (335)$$

with $h_{ij} \ll 1$ and such that is traceless ($h_{ii} = 0$). We will assume that h_{ij} is a function of t and that this variation with t is slow. The inverse metric is just:

$$g^{00}(t, \mathbf{x}) = -1, \quad g^{0i}(t, \mathbf{x}) = 0, \quad g^{ij}(t, \mathbf{x}) = \delta_{ij} - h_{ij}(t), \quad (336)$$

from which we get the different components of the projector $P^{\mu\nu}$:

$$P^{00} = 0, \quad P^{0i} = 0, \quad P^{ij} = \delta_{ij} - h_{ij}. \quad (337)$$

At first order in the perturbation, the Christoffel symbols are given by:

$$\Gamma_{00}^0 = \Gamma_{0i}^0 = 0, \quad \Gamma_{ij}^0 = \frac{1}{2} \partial_0 h_{ij}. \quad (338)$$

Therefore the covariant derivatives of the velocity are:

$$\nabla_0 u_0 = \nabla_0 u_i = 0, \quad \nabla_i u_j = \frac{1}{2} \partial_0 h_{ij}. \quad (339)$$

From these values and the hypothesis that the metric perturbation is traceless it follows that the covariant divergence of the velocity vanishes:

$$\nabla \cdot u = \frac{1}{2} \partial_0 h_{ii} = 0. \quad (340)$$

Let us assume that the only non-zero value of h_{ij} is h_{12} . Then, the linear response value of σ^{12} in frequency space is:

$$\sigma^{12}(\omega) = G_R^{12,12}(\omega, \mathbf{k} = 0) h_{12}(\omega). \quad (341)$$

Moreover, by using the values of the covariant derivatives of the velocity we get

$$\sigma^{12}(t) = \eta \partial_0 h_{12}(t), \quad (342)$$

or, in frequency space (for low ω):

$$\sigma^{12}(\omega) = -i\eta \omega h_{12}(\omega). \quad (343)$$

By comparing these two expressions of σ^{12} , we obtain again Kubo formula for the shear viscosity, namely:

$$\eta = - \lim_{\omega \rightarrow 0} \left[\frac{1}{\omega} \text{Im} G_R^{12,12}(\omega, \mathbf{k} = 0) \right]. \quad (344)$$

In order to compute holographically the retarded correlator of T_{12} , let us consider a general $d + 1$ -dimensional diagonal metric and let us perturb it by adding a non-

diagonal element along $x^1 x^2$:

$$ds^2 = g_{tt} dt^2 + g_{zz} dz^2 + g_{xx} (\delta_{ij} dx^i dx^j + 2\phi dx^1 dx^2), \quad (345)$$

where ϕ is small and independent of x^1 and x^2 . Notice that, in the perturbed metric at first order in ϕ , one has:

$$g_{12} = g_{xx} \phi, \quad g^1_2 = \phi. \quad (346)$$

Let us write the $x^1 x^2$ part of the metric (at first order) as:

$$g_{xx} (dx^1)^2 + g_{xx} (dx^2 + \phi dx^1)^2, \quad (347)$$

which is clearly similar to the ansatz corresponding to a Kaluza-Klein reduction along x^2 with KK gauge field connection:

$$A = \phi dx^1. \quad (348)$$

Therefore, we can use the known results of the KK reduction to write the quadratic action for ϕ . Indeed, the Einstein-Hilbert action for the metric leads to the following expression for the action of the gauge connection A (or equivalently the metric perturbation ϕ):

$$S_\phi = -\frac{1}{16\pi G_N} \int d^{d+1}x \sqrt{-g} g_{xx} \frac{F^2}{4}, \quad (349)$$

where g is the determinant of the unperturbed metric (with $\phi = 0$), $F = dA$ is the field strength of A and in $F^2 = F_{\mu\nu} F^{\mu\nu}$ the indices are raised with the unperturbed metric. As:

$$F = \partial_t \phi dt \wedge dx^1 + \partial_3 \phi dx^3 \wedge dx^1 + \partial_z \phi dz \wedge dx^1, \quad (350)$$

we have:

$$F^2 = \frac{2}{g_{xx}} \left[g^{tt} (\partial_t \phi)^2 + g^{xx} (\partial_3 \phi)^2 + g^{zz} (\partial_z \phi)^2 \right], \quad (351)$$

or, taking into account that $\partial_1 \phi = \partial_2 \phi = 0$:

$$F^2 = \frac{2}{g_{xx}} g^{MN} \partial_M \phi \partial_N \phi. \quad (352)$$

Thus, the action for the perturbation ϕ is:

$$S_\phi = -\frac{1}{16\pi G_N} \int d^{d+1}x \sqrt{-g} \frac{1}{2} g^{MN} \partial_M \phi \partial_N \phi, \quad (353)$$

which is the canonical form of the action for a scalar field, with normalization constant

$$q = 16\pi G_N . \quad (354)$$

It follows from our general calculation of Sect. 12 of the transport coefficients for a scalar field that the shear viscosity is given by:

$$\eta = \frac{1}{16\pi G_N} \frac{A_H}{V} , \quad (355)$$

and, therefore, the ratio η/s is just [24, 25]:

$$\frac{\eta}{s} = \frac{1}{4\pi} . \quad (356)$$

In ordinary units the ratio η/s is given by:

$$\frac{\eta}{s} = \frac{\hbar}{4\pi k_B} , \quad (357)$$

where k_B is the Boltzmann constant. Notice that this result does not depend on the metric chosen. It is valid for any theory with a gravity dual given by Einstein gravity coupled to matter fields. In this sense it is a universal result valid at $\lambda \rightarrow \infty$ (infinite coupling limit). The finite coupling corrections can also be calculated. For $\mathcal{N} = 4$ SYM one gets:

$$\frac{\eta}{s} = \frac{1}{4\pi} \left(1 + \frac{15\zeta(3)}{\lambda^{\frac{3}{2}}} + \dots \right) , \quad (358)$$

where $\zeta(x)$ is the Riemann zeta function ($\zeta(3) = 1.2020$). In general, η/s at $\lambda \rightarrow \infty$ is very small:

$$\frac{\eta}{s} = 0.07957 . \quad (359)$$

The finite coupling corrections make η/s increase. It is interesting to compare with the weak coupling calculation, valid when $\lambda \rightarrow 0$:

$$\frac{\eta}{s} = \frac{A}{\lambda^2 \log\left(\frac{B}{\sqrt{\lambda}}\right)} , \quad (360)$$

where A and B are constant coefficients that depend on the theory. Notice that in (360) $\eta/s \rightarrow \infty$ as $\lambda \rightarrow 0$. This is because a weakly coupled gauge theory is a gas with strong dissipative effects, in which momentum can be transported over long distances due to the long free path. In contrast a strongly coupled plasma is an almost perfect fluid in which momentum is rapidly transferred between layers of sheared fluid.

Kovtun, Son and Starinets (KSS) [24] conjectured that $1/(4\pi)$ is the lower bound for η/s . The lowest values of the η/s ratio found experimentally occur in two physical systems: the quark-gluon plasma created in heavy ion collisions at RHIC and the ultracold atomic Fermi gases at very low temperature. Both systems have η/s which is slightly above $1/(4\pi)$.

Changing the gravity theory the KSS bound is violated. For example, by adding higher curvature terms as in the Gauss-Bonnet gravity, whose action is given by:

$$S_{GB} = \frac{1}{16\pi G_5} \int d^5x \sqrt{-g} \left[R - 2\Lambda - \frac{3}{\Lambda} \lambda_{GB} \left(R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} \right) \right]. \quad (361)$$

For this gravity theory, the following η/s ratio is obtained:

$$\frac{\eta}{s} = \frac{1 - 4\lambda_{GB}}{4\pi}, \quad (362)$$

which violates the KSS bound if $\lambda_{GB} > 0$.

14 Final Remarks

In these lecture notes we have reviewed the basic features of the AdS/CFT duality, focusing on its conceptual foundations and on some particular applications. We have only scratched the surface of the subject, which in the last years has become a highly diversified field with many ramifications and connections. For the reader interested in knowing more on some of these applications of the holographic duality, let us quote some review articles, where the reader can find detailed accounts and the original references.

In this review we have mostly dealt with the gravitational description of $\mathcal{N} = 4$ $SU(N)$ gauge theory, which only contains fields transforming in the adjoint representation of the gauge group. In this sense $\mathcal{N} = 4$ SYM is a theory of pure glue. In order to extend the duality to theories closer to particle physics phenomenology one should be able to include flavor fields transforming in the fundamental representation of the gauge group (i.e. quarks). This can be done by adding the so-called flavor branes, as reviewed in [8, 26]. Moreover, the gauge/gravity duality can be extended to include less supersymmetric theories exhibiting confinement (see [27]) and one can construct holographic duals of quantum chromodynamics [8, 9, 28, 29]. The holographic methods can also be used to study dynamical electroweak symmetry breaking, in the framework of walking technicolor models [30]. On the other hand, the AdS/CFT correspondence has unveiled the integrable character of planar $\mathcal{N} = 4$ SYM and has allowed to extend the duality beyond the supergravity regime [31].

One of the more interesting recent developments of the gauge/gravity duality is the application of the string theoretical ideas to the down-to-earth problems of condensed matter physics. Indeed, condensed matter physics is full of strongly-

coupled systems which display quantum criticality with specific scaling laws. The gauge/gravity duality allows to map this scaling behavior to the general covariance of a gravity theory, for which one can apply the calculational tools and physical intuition of general relativity. In this way one can model the behavior of unusual phases of matter, such as strange metals or unconventional superconductors [6, 32–34]. In this context holography is emerging as a new tool to understand the collective quantum behavior not explained by the conventional paradigms, such as the Fermi-liquid theory.

It is also interesting to point out the connection between the AdS/CFT correspondence and quantum information theory. In particular, there is a holographic proposal for the entanglement entropy [35, 36], which allows a simple geometrical calculation of the latter as the area of a minimal surface. Let us finally mention that holography has also been applied to the study of strongly-coupled hydrodynamics. This particular version of the duality is called the fluid-gravity correspondence [37, 38].

Hopefully these lectures will stimulate the reader to explore some of the topics listed above.

Acknowledgments I am grateful to Yago Bea, Niko Jokela, Javier Mas and Ricardo Vázquez for their comments and help in the preparation of these lecture notes. I also thank Carlos Merino for his invitation to deliver the course on the AdS/CFT correspondence at the third IDPASC school. This work is funded in part by the Spanish grant FPA2011-22594, by Xunta de Galicia (Consellería de Educación, grant INCITE09 206 121 PR and grant PGIDIT10PXIB206075PR), by the Consolider-Ingenio 2010 Programme CPAN (CSD2007-00042), and by FEDER.

References

1. J.M. Maldacena, The large N limit of superconformal field theories and supergravity. *Adv. Theor. Math. Phys.* **2**, 231 (1998). [arXiv:hep-th/9711200](https://arxiv.org/abs/hep-th/9711200)
2. S.S. Gubser, I.R. Klebanov, A.M. Polyakov, Gauge theory correlators from noncritical string theory. *Phys. Lett. B* **428**, 105 (1998). [arXiv:hep-th/9802109](https://arxiv.org/abs/hep-th/9802109)
3. E. Witten, Anti-de Sitter space and holography. *Adv. Theor. Math. Phys.* **2**, 253 (1998). [arXiv:hep-th/9802150](https://arxiv.org/abs/hep-th/9802150)
4. O. Aharony, S.S. Gubser, J.M. Maldacena, H. Ooguri, Y. Oz, Large N field theories, string theory and gravity. *Phys. Rep.* **323**, 183 (2000). [arXiv:hep-th/9905111](https://arxiv.org/abs/hep-th/9905111)
5. E. D’Hoker, D.Z. Freedman, Supersymmetric gauge theories and the AdS/CFT correspondence, in Strings, Brane and extra dimensions: TASI 2001, ed. by Steven S. Gubser, Joseph D. Lykken (River Edge, N.J., World Scientific, 2004) [arXiv:hep-th/0201253](https://arxiv.org/abs/hep-th/0201253)
6. S.A. Hartnoll, Lectures on holographic methods for condensed matter physics. *Class. Quantum Gravity* **26**, 224002 (2009). [arXiv:0903.3246](https://arxiv.org/abs/0903.3246) [hep-th]
7. J. McGreevy, Holographic duality with a view toward many-body physics. *Adv. High Energy Phys.* **2010**, 723105 (2010). [arXiv:0909.0518](https://arxiv.org/abs/0909.0518) [hep-th]
8. J. Casalderrey-Solana, H. Liu, D. Mateos, K. Rajagopal, U.A. Wiedemann, Gauge/string duality, hot QCD and heavy ion collisions. (Cambridge University Press, Cambridge, UK, 2014)
9. Y. Kim, I.J. Shin, T. Tsukioka, Holographic QCD: past, present, and future. *Prog. Part. Nucl. Phys.* **68**, 55 (2013). [arXiv:1205.4852](https://arxiv.org/abs/1205.4852) [hep-ph]
10. A. Adams, L.D. Carr, T. Schäfer, P. Steinberg, J.E. Thomas, Strongly correlated quantum fluids: ultracold quantum gases, quantum chromodynamic plasmas, and holographic duality. *New J. Phys.* **14**, 115009 (2012). [arXiv:1205.5180](https://arxiv.org/abs/1205.5180) [hep-th]

11. M.B. Green, J.H. Schwarz, E. Witten, *Superstring Theory*, vol. 2 (Cambridge University Press, Cambridge, 1987)
12. J. Polchinski, *String Theory*, vol. 2 (Cambridge University Press, Cambridge, 1998)
13. K. Becker, M. Becker, J.H. Schwarz, *String Theory and M-Theory* (Cambridge University Press, Cambridge, 2007)
14. E. Kiritsis, *String Theory in a Nutshell* (Princeton University Press, Princeton, 2007)
15. B. Zwiebach, *A First Course in String Theory* (Cambridge University Press, Cambridge, 2009)
16. L.E. Iban  z, A. Uranga, *String Theory and Particle Physics* (Cambridge University Press, Cambridge, 2012)
17. G. 't Hooft, A planar diagram theory for strong interactions. *Nucl. Phys. B* **72**, 461 (1974)
18. A.V. Manohar, Large N QCD. in *Probing the Standard Model of Particle Interactions*, ed. by R. Gupta, A. Morel, E. De Rafael, F. David. Amsterdam, (Elsevier, The Netherlands, 1999) [arXiv:hep-ph/9802419](https://arxiv.org/abs/hep-ph/9802419)
19. M. Henningson, K. Skenderis, The holographic Weyl anomaly. *JHEP* **9807**, 023 (1998). [arXiv:hep-th/9806087](https://arxiv.org/abs/hep-th/9806087)
20. S. de Haro, S.N. Solodukhin, K. Skenderis, Holographic reconstruction of space-time and renormalization in the AdS/CFT correspondence. *Commun. Math. Phys.* **217**, 595 (2001). [arXiv:hep-th/0002230](https://arxiv.org/abs/hep-th/0002230)
21. K. Skenderis, Lecture notes on holographic renormalization. *Class. Quantum Gravity* **19**, 5849 (2002). [arXiv:hep-th/0209067](https://arxiv.org/abs/hep-th/0209067)
22. J.M. Maldacena, Wilson loops in large N field theories. *Phys. Rev. Lett.* **80**, 4859 (1998). [arXiv:hep-th/9803002](https://arxiv.org/abs/hep-th/9803002)
23. S.J. Rey, J.T. Yee, Macroscopic strings as heavy quarks in large N gauge theory and anti-de Sitter supergravity. *Eur. Phys. J. C* **22**, 379 (2001). [arXiv:hep-th/9803001](https://arxiv.org/abs/hep-th/9803001)
24. P. Kovtun, D.T. Son, A.O. Starinets, Viscosity in strongly interacting quantum field theories from black hole physics. *Phys. Rev. Lett.* **94**, 111601 (2005). [arXiv:hep-th/0405231](https://arxiv.org/abs/hep-th/0405231)
25. D.T. Son, A.O. Starinets, Viscosity, black holes, and quantum field theory. *Ann. Rev. Nucl. Part. Sci.* **57**, 95 (2007). [arXiv:0704.0240](https://arxiv.org/abs/0704.0240) [hep-th]
26. J. Erdmenger, N. Evans, I. Kirsch, E. Threlfall, Mesons in gauge/gravity duals—a review. *Eur. Phys. J. A* **35**, 81 (2008). [arXiv:0711.4467](https://arxiv.org/abs/0711.4467) [hep-th]
27. J.D. Edelstein, R. Portugues, Gauge/string duality in confining theories. *Fortsch. Phys.* **54**, 525 (2006). [arXiv:hep-th/0602021](https://arxiv.org/abs/hep-th/0602021)
28. K. Peeters, M. Zamaklar, The string/gauge theory correspondence in QCD. *Eur. Phys. J. ST* **152**, 113 (2007). [arXiv:0708.1502](https://arxiv.org/abs/0708.1502) [hep-th]
29. D. Mateos, String theory and quantum chromodynamics. *Class. Quantum Gravity* **24**, S713 (2007). [arXiv:0709.1523](https://arxiv.org/abs/0709.1523) [hep-th]
30. M. Piai, Lectures on walking technicolor, holography and gauge/gravity dualities. *Adv. High Energy Phys.* **2010**, 464302 (2010). [arXiv:1004.0176](https://arxiv.org/abs/1004.0176) [hep-ph]
31. N. Beisert, C. Ahn, L.F. Alday, Z. Bajnok, J.M. Drummond, L. Freyhult, N. Gromov, R.A. Janik et al., Review of AdS/CFT integrability: an overview. *Lett. Math. Phys.* **99**, 3 (2012). [arXiv:1012.3982](https://arxiv.org/abs/1012.3982) [hep-th]
32. S. Sachdev, Condensed matter and AdS/CFT. *Lect. Notes Phys.* **828**, 273 (2011). [arXiv:1002.2947](https://arxiv.org/abs/1002.2947) [hep-th]
33. S. Sachdev, What can gauge-gravity duality teach us about condensed matter physics? *Ann. Rev. Condens. Matter Phys.* **3**, 9 (2012). [arXiv:1108.1197](https://arxiv.org/abs/1108.1197) [cond-mat.str-el]
34. A.G. Green, An introduction to gauge gravity duality and its application in condensed matter. *Contemp. Phys.* **54**(01), 33–48 (2013). [arXiv:1304.5908](https://arxiv.org/abs/1304.5908) [cond-mat.str-el]
35. T. Nishioka, S. Ryu, T. Takayanagi, Holographic entanglement entropy: an overview. *J. Phys. A* **42**, 504008 (2009). [arXiv:0905.0932](https://arxiv.org/abs/0905.0932) [hep-th]
36. T. Takayanagi, Entanglement entropy from a holographic viewpoint. *Class. Quantum Gravity* **29**, 153001 (2012). [arXiv:1204.2450](https://arxiv.org/abs/1204.2450) [gr-qc]
37. M. Rangamani, Gravity and hydrodynamics: lectures on the fluid-gravity correspondence. *Class. Quantum Gravity* **26**, 224003 (2009). [arXiv:0905.4352](https://arxiv.org/abs/0905.4352) [hep-th]
38. V.E. Hubeny, S. Minwalla, M. Rangamani, The fluid/gravity correspondence in String Theory and Its Applications: From meV to the Planck Scale, ed. by M. Dine, T. Banks, S. Sachdev, (World Scientific, 2012) [arXiv:1107.5780](https://arxiv.org/abs/1107.5780) [hep-th]

Welcome Speech by the Rector



The Rector of the University of Santiago de Compostela, Professor Juan J. Casares Long giving his welcome speech during the official reception held at the Pazo de San Xerome to the participants at the Third IDPASC School, with professors Mário Pimenta and Carlos Merino (picture by A. Ponte)

Queimada



Professor Mário Pimenta stirring the queimada under the supervision of professors Alessandro De Angelis and Carlos Merino, during the banquet of the Third IDPASC School, held at the Hotel-Monumento San Francisco, in Santiago de Compostela, on Wednesday, January 30th (picture from the C. Merino collection)

List of Participants

1. Paula Álvarez Cartelle, Universidade de Santiago de Compostela, Galiza, Spain.
2. Carlota Andrés Casas, Universidade de Santiago de Compostela, Galiza, Spain.
3. Liliana Apolinario, Laboratório de Instrumentação e Física Experimental de Partículas de Lisboa, Portugal.
4. Néstor Armesto Pérez, Universidade de Santiago de Compostela, Galiza, Spain.
5. Yago Bea Besada, Universidade de Santiago de Compostela, Galiza, Spain.
6. Cristian Bosch Serrano, IFIC—Universitat de Valencia, Spain.
7. Joao Carvalho, LIP—Universidade de Coimbra, Portugal.
8. Gustavo Castelo-Branco, Instituto Superior Técnico de Lisboa, Portugal.
9. Ruben Conceição, Laboratório de Instrumentação e Física Experimental de Partículas de Lisboa, Portugal.
10. Alessandro De Angelis, Università di Udine, Italy.
11. Francisco Diogo, Laboratório de Instrumentação e Física Experimental de Partículas de Lisboa, Portugal.
12. Álvaro Dosil Suárez, Universidade de Santiago de Compostela, Galiza, Spain.
13. David Emmanuel—Costa, Instituto Superior Técnico de Lisboa, Portugal.
14. Joao Espadanal, Laboratório de Instrumentação e Física Experimental de Partículas de Lisboa, Portugal.
15. Xabier García Feal, Universidade de Santiago de Compostela, Galiza, Spain.
16. Daniel García Fernández, Universidade de Santiago de Compostela, Galiza, Spain.
17. Mauel García Fernández, Universidade de Santiago de Compostela, Galiza, Spain.
18. María José Herrero, Universidad Autónoma de Madrid-IFT, Spain.
19. María Luisa López Ibañez, IFIC—Universitat de Valencia, Spain.
20. Antonio López Maroto, Universidad Complutense de Madrid, Spain.
21. Jesús Martínez Asencio, Universitat de Valencia, Spain.
22. Carlos Merino Gayoso, Universidade de Santiago de Compostela, Galiza, Spain.
23. Alexis Moscoso Rial, Universidade de Santiago de Compostela, Galiza, Spain.
24. Carlos Naya Rodríguez, Universidade de Santiago de Compostela, Galiza, Spain.

25. Micael Oliveira, Universidade de Coimbra, Portugal.
26. Daniel Pablos Alfonso, Universitat de Barcelona, Spain.
27. Carlos Pajares Vales, Universidade de Santiago de Compostela, Galiza, Spain.
28. Gonzalo Parente Bermúdez, Universidade de Santiago de Compostela, Galiza, Spain.
29. Mário Pimenta, Instituto Superior Técnico and Laboratório de Instrumentação e Física Experimental de Partículas, Lisboa, Portugal.
30. Emiliano Pinto, Universidade de Lisboa, Portugal.
31. Sylvestre Pires, Laboratoire de Physique Nucléaire et des Hautes Énergies, Paris, France.
32. Mateusz Płoskon, Lawrence Berkeley Laboratory—ERN, USA—Switzerland.
33. Cristian Redondo Lourés, Universidade de Santiago de Compostela, Galiza, Spain.
34. Manoel Rodríguez Calvo, Universidade de Santiago de Compostela, Galiza, Spain.
35. Manoel Anxo Rodríguez-Moldes Díaz, Universidade de Santiago de Compostela, Galiza, Spain.
36. Pablo Ruíz Valls, IFIC—Universitat de Valencia, Spain.
37. Cibrán Santamarina Ríos, Universidade de Santiago de Compostela, Galiza, Spain.
38. Yuli M. Shabelski, Petersburg Nuclear Physics Institute, St.Petersburg, Russia.
39. Mário Sousa, Universidade de Lisboa, Portugal.
40. Carlo Stella, Università di Udine, Italy.
41. Guillermo Torralba Elipe, Universidade de Santiago de Compostela, Galiza, Spain.
42. Ricardo Vázquez López, Universidade de Santiago de Compostela, Galiza, Spain.
43. Alfonso Vázquez Ramallo, Universidade de Santiago de Compostela, Galiza, Spain.
44. Carlos Vázquez Sierra, Universidade de Santiago de Compostela, Galiza, Spain.
45. Pablo Villanueva Pérez, IFIC—Universitat de Valencia, Spain.
46. Michael Williams, Massachusset Institute of Technology—LHCb, USA—Switzerland.