

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

2011 European School of High-Energy Physics

Cheile Gradistei, Romania
7 – 20 September 2011

Proceedings

Editors: C. Grojean
M. Mulders

ISBN 978-92-9083-401-4

ISSN 0531-4283

Copyright © CERN, 2014

© Creative Commons Attribution 3.0

Knowledge transfer is an integral part of CERN's mission.

CERN publishes this report Open Access under the Creative Commons Attribution 3.0 license (<http://creativecommons.org/licenses/by/3.0/>) in order to permit its wide dissemination and use. The submission of a contribution to a CERN Yellow Report shall be deemed to constitute the contributor's agreement to this copyright and license statement. Contributors are requested to obtain any clearances that may be necessary for this purpose.

This report should be cited as:

Proceedings of the 2011 European School of High-Energy Physics, Cheile Gradistei, Romania, 7–20 September 2011, edited by C. Grojean and M. Mulders, CERN-2014-003 (CERN, Geneva, 2014), DOI: 10.5170/CERN-2014-003

A contribution in this report should be cited as:

[Author name(s)], in Proceedings of the 2011 European School of High-Energy Physics, Cheile Gradistei, Romania, 7–20 September 2011, edited by C. Grojean and M. Mulders, CERN-2014-003 (CERN, Geneva, 2014), pp.[first page]–[last page], DOI: 10.5170/CERN-2014-003.[firstpage]

Abstract

The European School of High-Energy Physics is intended to give young physicists an introduction to the theoretical aspects of recent advances in elementary particle physics. These proceedings contain lecture notes on quantum field theory and the Standard Model, quantum chromodynamics, flavour physics, neutrino physics, physics beyond the Standard Model, cosmology, heavy ion physics, statistical data analysis, as well as an account for the physics results with the data accumulated during the first run of the LHC.

Preface

The nineteenth event in the series of the European School of High-Energy Physics took place in Moeciu, Romania, from 7 to 20 September 2011. It was organized jointly by CERN, Geneva, Switzerland, and JINR, Dubna, Russia, with support from IFIN-HH, Bucharest. The local organization team was chaired by Ioan Ursu from IFIN-HH. The other members of the local committee were: V. Almasan, G. Cata-Danil, C.B. Cizmas, M. Cuciuc, A. Dorobantu, A. Jipa, R. Muresan, B. Popovici, L. Serban, G. Stoicea, and E. Teodorescu.

A total of 99 students coming from 31 different countries attended the school, mainly from member states of CERN and/or JINR, but also a few from other regions. The participants were generally students in experimental High-Energy Physics in the final years of work towards their PhDs.

The School was hosted at the Cheile Gradistei holiday complex and conference centre in Moeciu, a few km outside the town of Bran located about 200 km north of Bucharest. According to the tradition of the school, the students shared twin rooms mixing participants of different nationalities.

A total of 33 lectures were complemented by daily discussion sessions led by six discussion leaders. The students displayed their own research work in the form of posters in an evening session in the first week, and the posters stayed on display until the end of the School. Each discussion group carried out a collaborative project, studying in detail the analysis from a published paper from one of the LHC experiments; a summary was presented by a student representative of each group in an evening session in the second week of the School. The full scientific programme was arranged in the on-site conference facilities.

Our thanks go to the local-organization team and, in particular, to Ioan Ursu for his excellent work and assistance in preparing the School, on both scientific and practical matters, and for his presence throughout the event. Our thanks also go to the efficient and friendly hotel management and staff who assisted the School organizers and the participants in many ways.

Very great thanks are due to the lecturers and discussion leaders for their active participation in the School and for making the scientific programme so stimulating. The students, who in turn manifested their good spirits during two intense weeks, undoubtedly appreciated listening to and discussing with the teaching staff of world renown.

We would like to express our appreciation to Professor Rolf Heuer, Director General of CERN, and Professor Alexander Olchevsky who represented JINR, for their lectures on the scientific programmes of the two organizations and for discussing with the School participants.

In addition to the rich scientific programme, the participants enjoyed numerous sports and leisure activities in and around the Cheile Gradistei complex. Particularly noteworthy were the very nice excursions to Brasov, Sinaia and Bran that were arranged by the local organization team and, in particular, by Andrei Dorobantu.

In parallel with the School, an exhibition about CERN and particle physics in Romania was held in central Bucharest. This was complemented by a public lecture from Professor Rolf Heuer, Director General of CERN, on 20 September, introduced by Dr. Daniel Funeriu, Minister of Education, Research, Youth and Sports. We are grateful for the Romanian sponsorship that funded the board and lodgings for students from the School to remain an extra day in Bucharest to visit the exhibition and attend the public lecture.

We are very grateful to Mrs Helene Haller and Mrs Tatyana Donskova for their untiring efforts in the lengthy preparations for and the day-to-day operation of the School. Their continuous care of the participants and their needs during the School was highly appreciated.

The success of the School was to a large extent due to the students themselves. Their poster session was very well prepared and highly appreciated, their group projects were a great success, and throughout the School they participated actively during the lectures, in the discussion sessions and in the different activities and excursions. Finally, one should not forget the show that they prepared and presented following the farewell banquet.

Nick Ellis
(On behalf of the Organizing Committee)



PHOTOGRAPHS (MONTAGE)



Contents

Preface	
<i>N. Ellis</i>	v
Photograph of participants	vi
Photographs (montage)	vii
Field Theory and the Standard Model	
<i>E. Dudas</i>	1
Basics of QCD for the LHC: $pp \rightarrow H + X$ as a case study	
<i>F. Maltoni</i>	47
Introduction to Flavor Physics	
<i>Y. Grossman</i>	73
Neutrino Oscillation Physics	
<i>B. Kayser</i>	107
Beyond the Standard Model	
<i>B.A. Dobrescu</i>	119
Cosmology	
<i>V.A. Rubakov</i>	151
QCD in Heavy Ions Collisions	
<i>E. Iancu</i>	197
Practical Statistics for the LHC	
<i>K. Cranmer</i>	267
Physics at the LHC – From SM measurements to Searches for New Physics	
<i>K. Jakobs</i>	309
Organizing Committee	359
Local Organizing Committee	359
International Advisors	359
List of Lecturers	359
List of Discussion Leaders	359
List of Students	360
List of Posters	361

Field Theory and the Standard Model

E. Dudas

Theory Group, Physics Department, CERN, Geneva, Switzerland
Centre de Physique Théorique, Ecole Polytechnique, Palaiseau, France
Laboratoire de Physique Théorique, Université de Paris-Sud, Orsay, France

Abstract

This brief introduction to Quantum Field Theory and the Standard Model contains the basic building blocks of perturbation theory in quantum field theory, an elementary introduction to gauge theories and the basic classical and quantum features of the electroweak sector of the Standard Model. Some details are given for the theoretical bias concerning the Higgs mass limits, as well as on obscure features of the Standard Model which motivate new physics constructions.

1 Introduction

The development of Quantum Field Theory and the raise of the Standard Model remains as one of the most fascinating adventures of fundamental science of the twenties century. Indeed, despite the seemingly great difference between the strength, action range and the different role played in the birth and the evolution of our universe by the electromagnetic, weak and strong interactions, we know that all three interactions are based on the gauge principle, which seems to be a fundamental principle of nature. Amazingly enough, gauge theories with or without spontaneous symmetry breaking are also renormalizable, in the leading expansion in the dimension of operators in quantum field theory. There is nothing inconsistent from the modern perspective in non-renormalizable theories, the prominent and most important example of this type being Einstein gravity. However, renormalizability renders a theory highly predictive up to high energy scales. This allowed highly precise tests of quantum electrodynamics (QED) like for example the computation of the electron anomalous magnetic moment or the running with the energy of the fine-structure constant. That's why we can talk today about the precision tests of the Standard Model, possible deviations from it, if found experimentally, having to be interpreted unambiguously as signatures of new physics.

These lectures contain an introduction to the basic features of quantum field theory and the electroweak sector of the Standard Model. They are organized as follows. Section 2 introduces symmetries and the Noether theorem. Section 3 introduces perturbation theory, first time-dependent perturbation theory in quantum mechanics, followed by perturbation theory in quantum field theory. Section 4 is an introduction to abelian and non-abelian gauge theories and elements of their quantization. Section 5 describes spontaneous symmetry breaking, Goldstone theorem and the Higgs mechanism. Section 6 introduces the classical aspects of the electroweak sector of the standard sector. Section 7 discusses renormalizability and examples of energy evolution of couplings in the ϕ^4 scalar theory and QED. Section 8 contains some simple applications and constraints coming from global and gauge anomalies. Section 9 enters into the Higgs physics and some theoretical arguments in favor of a light Higgs boson. As well known, Higgs searches are presently the main goal of the Large Hadron Collider (LHC) at CERN. (Very) preliminary LHC results seem to validate the theoretical picture pioneered long-time ago by Higgs and by Brout–Englert [1] and the more recent theoretical arguments pointing in favor of a light scalar Higgs boson. We end up with brief standard arguments in favor of the Standard Model as an effective theory, to be completed beyond some unknown energy scale with an underlying microscopic theory.

They notes were intended to be necessary ingredients for the lectures on Quantum Chromodynamics (QCD) given by Fabio Maltoni in this School [2], Heavy Ion Collisions lectured by Edmond

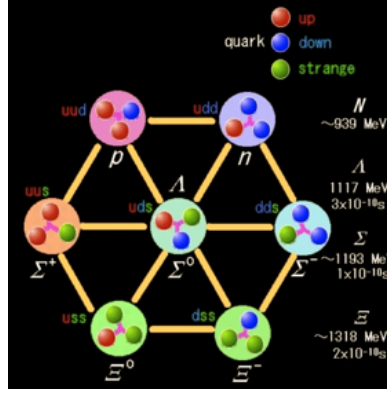


Fig. 1: The flavor $SU(3)_f$ baryon octet. From Ref. [7].

Iancu [3], the lectures on Cosmology by Valery Rubakov [4], on Neutrino Physics by Boris Kayser [5] and Beyond the Standard Model (BSM) by Bogdan Dobrescu [6].

2 Fields, Symmetries and the Noether theorem

Symmetries are fundamental in our understanding in nature. Classic examples are:

- Continuous *spacetime symmetries*, for example space rotations.
- *Discrete symmetries* are fundamental in classification and properties of crystals.
- *Continuous and discrete internal symmetries* in particle physics.

Example the *eightfold way*: Flavor $SU(3)_f$, used by M. Gell-Mann in his famous classification of hadrons which also led to the introduction of color as a new quantum number and to the modern theory of strong interactions, the QCD (see F. Maltoni's lectures [2]).

The importance of symmetries in nature is to a large extent due to the **Noether theorem**: *To any continuous symmetry of a physical system, it corresponds a conserved current and an associate conserved charge.*

Examples of conserved charges associated to continuous symmetries are:

Symmetry	Conserved charge
Time translation	Energy
Space translation	Momentum
Rotations	Angular momentum
Phase rotations wave function	Electric charge

For the case of internal symmetries which are our primary goal here, the proof of the Noether theorem goes as follows. Consider a field theory with ϕ denoting collectively all the fields of the theory, of Lagrangian $\mathcal{L}(\phi, \partial_m \phi)$. The field transformations generated by infinitesimal parameters α_a

$$\delta \phi = \alpha_a(x) T^a \phi, \quad (1)$$

lead to a new Lagrangian

$$\mathcal{L}(\phi, \partial_m \phi) \rightarrow \hat{\mathcal{L}}(\phi, \alpha_a, \partial_m \phi, \partial_m \alpha_a). \quad (2)$$

The variation of the action functional $S(\phi, \partial_m \phi)$ under field variations (1) is

$$\delta S = \int d^4x \left[\frac{\partial \hat{\mathcal{L}}}{\partial \alpha_a} \alpha_a + \frac{\partial \hat{\mathcal{L}}}{\partial (\partial_m \alpha_a)} \partial_m \alpha_a \right] = \int d^4x \left[\frac{\partial \hat{\mathcal{L}}}{\partial \alpha_a} - \partial_m \frac{\partial \hat{\mathcal{L}}}{\partial (\partial_m \alpha_a)} \right] \alpha_a, \quad (3)$$

where to get the result in the last line we performed an integration by parts. By defining the currents

$$J_a^m = \frac{\partial \hat{\mathcal{L}}}{\partial (\partial_m \alpha_a)}, \quad (4)$$

we find that the variation of the action (3) vanishes if

$$\partial_m J_a^m = \frac{\partial \hat{\mathcal{L}}}{\partial \alpha_a}. \quad (5)$$

In the particular case where the field variation is a symmetry of the Lagrangian, we immediately find the conservation law

$$\partial_m J_a^m = 0 \Rightarrow \frac{dQ_a}{dt} = \int d^3\mathbf{x} \partial_m J_a^m = 0, \quad \text{where} \quad Q_a = \int d^3\mathbf{x} J_a^0 \quad (6)$$

is a conserved charge. It is straightforward and is left to the reader as an exercise to show that the conserved current can also be computed according to the following formulae

$$\delta \hat{\mathcal{L}} = J_a^m \partial_m \alpha_a, \quad \text{or} \quad J_a^m = \frac{\partial \hat{\mathcal{L}}}{\partial (\partial_m \phi)} \frac{\delta \phi}{\delta \alpha_a}. \quad (7)$$

Through the Noether theorem, continuous symmetries lead to conserved charges that are manifest in the spectrum and interactions. As known already from quantum mechanics¹ their study greatly simplifies the dynamics.

As we will see later on, *the local (space-time dependent) symmetries* determine the structure of all the fundamental interactions in nature! Indeed, all four fundamental interactions, the electromagnetism, the weak and strong forces and (in a somewhat different way) the gravitational one can be found as consequences of local symmetries called **gauge symmetries**.

3 Quantization and perturbation theory

The second quantization of fields and perturbation theory lead to precise formulae for scattering amplitudes which led to the Feynman diagrams, that are crucial for computing cross sections and other physical observables. The appropriate formalism uses the Heisenberg or interaction picture in quantum mechanics, that we first review, before introducing the corresponding quantum field theory formalism.

3.1 Time-dependent perturbation theory in quantum mechanics

Let's start from Schrodinger versus interaction/Heisenberg picture in Quantum Mechanics.

$$H = H_0 + H_{int}, \quad (8)$$

where H_0 is the free hamiltonian and H_{int} is the interaction. The Schrodinger equation is

$$i \frac{d}{dt} |\Psi_S(t)\rangle = (H_0 + H_{int}) |\Psi_S(t)\rangle. \quad (9)$$

time dep. $\nearrow$ $\nwarrow$ time-indep. operators

¹For example the conservation of angular momentum greatly simplifies the study of hydrogen atom.

In the *interaction (or Heisenberg) picture*

$$|\Psi_I(t)\rangle = e^{iH_0 t} |\Psi_S(t)\rangle \quad , \quad H_{int}(t) = e^{iH_0 t} H_{int}(t) e^{-iH_0 t} \quad (10)$$

the Schrodinger equation becomes (**Exercise:**)

$$i \frac{d}{dt} |\Psi_I(t)\rangle = H_{int}(t) |\Psi_I(t)\rangle. \quad (11)$$

We define the evolution operator $U(t, t_i)$ by

$$|\Psi_I(t)\rangle = U(t, t_i) |\Psi_I(t_i)\rangle \quad , \quad U(t_i, t_i) = 1. \quad (12)$$

Exercise: Check that U satisfies the equation

$$i \frac{\partial U(t, t_i)}{\partial t} = H_{int}(t) U(t, t_i). \quad (13)$$

It can be shown that (**Exercise:**)

$$U(t, t_i) = T e^{-i \int_{t_i}^t dt' H_{int}(t')}, \quad (14)$$

where the *time-ordered product* of the operators A and B is defined as

$$TA(t_1)B(t_2) = \theta(t_1 - t_2) A(t_1)B(t_2) + \theta(t_2 - t_1) B(t_2)A(t_1). \quad (15)$$

The S-matrix is defined as

$$S = \lim_{t \rightarrow \infty, t_i \rightarrow -\infty} U(t, t_i) = T e^{-i \int dt H_{int}(t)} \quad (16)$$

The states in the far past, before the interaction process are free wave packets and are denoted by $|p_1 \cdots p_n, \text{in}\rangle$, where p_i are the momenta of the incident particles. Similarly, the states in the far future, after the interaction process are again free and are denoted by $|p'_1 \cdots p'_m, \text{out}\rangle$, where p'_i are the momenta of the scattered particles. The transition amplitudes passing from the initial to the final state is

$$\begin{aligned} S_{if} &= \langle \Psi_f | S | \Psi_i \rangle = \langle p'_1 \cdots p'_m, \text{in} | S | p_1 \cdots p_n, \text{in} \rangle \\ &= \langle p'_1 \cdots p'_m, \text{out} | p_1 \cdots p_n, \text{in} \rangle \\ &= \text{no interaction term} + i (2\pi)^4 \delta^4 \left(\sum_{j=1}^m p'_j - \sum_{i=1}^n p_i \right) \mathcal{A}_{if}. \end{aligned} \quad (17)$$

The Feynman rules are usually given for the matrix $\mathcal{A}_{if}$.

3.2 Quantization of the scalar theory

Canonical quantization in Quantum field theory uses the Heisenberg (interactive) picture. Let us consider for illustration a scalar theory

$$\begin{aligned} \mathcal{L} &= \frac{1}{2} (\partial \phi)^2 - \frac{m^2}{2} \phi^2 - \frac{\lambda}{4!} \phi^4 = \frac{1}{2} \dot{\phi}^2 - \frac{1}{2} (\nabla \phi)^2 - \frac{m^2}{2} \phi^2 - \frac{\lambda}{4!} \phi^4 \\ &\equiv \mathcal{L}_0 + \mathcal{L}_{int} \quad \text{with} \quad \mathcal{L}_{int} = - \frac{\lambda}{4!} \phi^4. \end{aligned} \quad (18)$$

The metric convention throughout these lectures will be $\eta_{\mu\nu} = \text{diag}(1, -1, -1, -1)$. The conjugate momentum is $\pi = \frac{\partial \mathcal{L}}{\partial \dot{\phi}} = \dot{\phi}$ and the hamiltonian

$$H = \int d^3 \mathbf{x} \left[\dot{\phi} \frac{\partial \mathcal{L}}{\partial \dot{\phi}} - \mathcal{L} \right] = \int d^3 \mathbf{x} \left[\frac{1}{2} \dot{\phi}^2 + \frac{1}{2} (\nabla \phi)^2 + \frac{m^2}{2} \phi^2 + \frac{\lambda}{4!} \phi^4 \right]$$

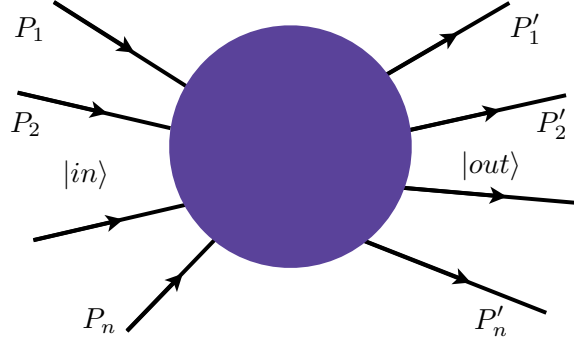


Fig. 2: Scattering amplitude of n initial particles and m final particles. The amplitude of the process is denoted $S_{if} = \langle p'_1 \cdots p'_m, \text{out} | p_1 \cdots p_n, \text{in} \rangle$.

$$\equiv H_0 + H_{int} \quad \text{with} \quad \begin{cases} H_0 = \int d^3\mathbf{x} \left[\frac{1}{2} \dot{\phi}^2 + \frac{1}{2} (\nabla \phi)^2 + \frac{m^2}{2} \phi^2 \right], \\ H_{int} = \int d^3\mathbf{x} \frac{\lambda}{4!} \phi^4. \end{cases} \quad (19)$$

The field equations and the solutions for the free-field theory are:

$$(\square + m^2) \phi(x) = 0 \quad \Rightarrow \quad \phi(x) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2} \sqrt{2\omega_k}} \left(e^{ikx} a_{\mathbf{k}}^\dagger + e^{-ikx} a_{\mathbf{k}} \right), \quad (20)$$

where $k_0 = \omega_k = \sqrt{\mathbf{k}^2 + m^2}$. The solution $\phi(x)$ is the operator in the Heisenberg picture. Quantization proceeds as usual:

$$[a_{\mathbf{k}}, a_{\mathbf{k}'}^\dagger] = \delta^3(\mathbf{k} - \mathbf{k}') \quad \rightarrow \quad [\phi(t, \mathbf{x}, \pi(t, \mathbf{y}))] = i\delta^3(\mathbf{x} - \mathbf{y}). \quad (21)$$

The one-particle states are defined by

$$|\mathbf{k}\rangle = a_{\mathbf{k}}^\dagger |0\rangle \Rightarrow \langle \mathbf{k}' | \mathbf{k} \rangle = \delta^3(\mathbf{k} - \mathbf{k}'), \quad (22)$$

and the energy/hamiltonian is

$$H_0 = \int d^3\mathbf{k} \omega_k (a_{\mathbf{k}}^\dagger a_{\mathbf{k}} + \frac{1}{2}) \quad (23)$$

and is one of a collection of quantum oscillators. Therefore (there is by definition no interaction in the asymptotic past and future)

$$\begin{cases} |\psi_i\rangle = |p_1 p_2 \cdots p_n\rangle = a_{\mathbf{p}_1}^\dagger \cdots a_{\mathbf{p}_n}^\dagger |0\rangle \\ |\psi_f\rangle = |p'_1 p'_2 \cdots p'_m\rangle = a_{\mathbf{p}'_1}^\dagger \cdots a_{\mathbf{p}'_m}^\dagger |0\rangle \end{cases} \quad (24)$$

3.3 Evolution operator and S-matrix in quantum field theory

We define the evolution operator by

$$\phi(x) = U^{-1}(t) \phi_{in}(x) U(t), \quad (25)$$

where $U(t) = U(t, -\infty)$, ϕ_{in} is the incoming (free) field and ϕ is the interacting field. As in quantum mechanics, we separate the interaction from the free hamiltonian

$$H = H_0 + H_{int}(t). \quad (26)$$

The evolution equations for the quantum fields are

$$\frac{\partial \phi(x)}{\partial t} = i [H(\phi), \phi(x)] \quad , \quad \frac{\partial \phi_{in} \phi(x)}{\partial t} = i [H_0(\phi_{in}), \phi_{in}(x)] . \quad (27)$$

By combining (25) and (27), we obtain the equation satisfied by the evolution operator

$$i \frac{dU}{dt} = (H(\phi_{in}) - H_0(\phi_{in})) U = H_I(t) U , \quad (28)$$

where $H_I(t) = H_{int}(\phi_{in}, \pi_{in})$. It is easy to check that the evolution operator satisfies the integral equation

$$U(t) = I - i \int_{-\infty}^t dt_1 H_I(t_1) U(t_1) . \quad (29)$$

This equation can be solved by iteration. It can be shown term by term in the expansion in the interaction that the solution of Eq. (29) can be written in the compact elegant form

$$U(t) = T e^{-i \int_{-\infty}^t dt' H_I(t')} . \quad (30)$$

Consequently, the S-matrix is given by

$$S = \lim_{t \rightarrow \infty} U(t) = T e^{-i \int_{-\infty}^{\infty} dt' H_I(t')} = T e^{i \int d^4x \mathcal{L}_I} . \quad (31)$$

Whereas at first sight the last equality is true only in the absence of derivative interactions $\mathcal{H}_I = -\mathcal{L}_I$, it is actually true in general.

3.4 Reduction formula, perturbation theory and Feynman diagrams

Feynman rules and perturbation theory follow from the expansion in powers of the interaction of S-matrix elements

$$\langle p_1 \cdots p_n, in | S | q_1 \cdots q_l, in \rangle = \langle 0 | a_{\mathbf{p}_1} \cdots a_{\mathbf{p}_n} T e^{i \int d^4x \mathcal{L}_{int}(x)} a_{\mathbf{p}_1}^\dagger \cdots a_{\mathbf{p}_n}^\dagger | 0 \rangle . \quad (32)$$

A very important formula in S-matrix perturbation theory is the reduction or the LSZ (Lehmann–Symanzik–Zimmermann) formula, which relates S-matrix elements to the time-ordered Green functions

$$\begin{aligned} \langle p_1 \cdots p_n, out | q_1 \cdots q_l, in \rangle &= \langle p_1 \cdots p_n, in | S | q_1 \cdots q_l, in \rangle \\ &= \text{disconnected terms} + (iZ^{-1/2})^{n+l} \times \\ &\times \int d^4y_1 \cdots d^4x_l e^{i(\sum_k p_k y_k - \sum_r q_r x_r)} (\square_{y_1} + m^2) \cdots (\square_{x_l} + m^2) \langle 0 | T \phi(y_1) \cdots \phi(x_l) | 0 \rangle \end{aligned} \quad (33)$$

where Z is the wave-function renormalization for the scalar field. The central figures in perturbation theory are therefore the *Green functions*

$$G(x_1 \cdots x_n) = \langle 0 | T \phi(y_1) \cdots \phi(x_n) | 0 \rangle . \quad (34)$$

The Green functions of the interactive field ϕ can be expressed in terms of Green functions of the free-field ϕ_{in} via the crucial formula (see for example Refs. [8, 50])

$$G(x_1 \cdots x_n) = \frac{\langle 0 | T \phi_{in}(x_1) \cdots \phi_{in}(x_n) e^{i \int d^4x \mathcal{L}_{int}[\phi_{in}]} | 0 \rangle}{\langle 0 | T e^{i \int d^4x \mathcal{L}_{int}[\phi_{in}]} | 0 \rangle} . \quad (35)$$

Another important notion is **normal ordering**. A normal ordered operator $: \mathcal{O} :$ is defined such that all creation operators are on the left and all annihilation operators are on the right. By construction then its vev vanishes $\langle : \mathcal{O} : \rangle = 0$. G. Wick found an elegant way to express free-field Green functions in terms

of normal-ordered products, by the so-called **Wick theorem**. The simplest example is the two-point function

$$T\phi_{in}(x)\phi_{in}(y) = : \phi_{in}(x)\phi_{in}(y) : + D_F(x-y) , \quad (36)$$

where $D_F(x-y)$ is the Feynman propagator. An explicit computation gives

$$\begin{aligned} D_F(x-y) &= \int \frac{d^3\mathbf{k}}{(2\pi)^3 2\omega_k} \{ \theta(x_0 - y_0) e^{-ik(x-y)} + \theta(y_0 - x_0) e^{ik(x-y)} \} \\ &= \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2 + i\epsilon} e^{-ik(x-y)} . \end{aligned} \quad (37)$$

The $i\epsilon$ prescription in the Feynman propagator has the property of propagating the positive frequencies into the future and the negative frequencies into the past. This is precisely what will be needed later on in order to capture both particles and antiparticles propagation in a causal way. Wick theorem can be generalized to a time-ordered product of an arbitrary number of fields

$$T\phi_{in}(x_1)\phi_{in}(x_2)\cdots\phi_{in}(x_n) = : \phi_{in}(x_1)\phi_{in}(x_2)\cdots\phi_{in}(x_n) : + \text{all possible contractions} . \quad (38)$$

Let us now discuss the *Feynman diagrams* for the simplest ϕ^4 theory, with

$$\mathcal{L}_{int} = -\frac{\lambda}{4!} \phi^4 . \quad (39)$$

The Feynman rules are usually formulated in the momentum space:

$$G(p_1 \cdots p_n) = \int d^4x_1 \cdots d^4x_n e^{i\sum_i p_i x_i} G(x_1 \cdots x_n) . \quad (40)$$

Applying perturbation theory (33), we obtain the following **Feynman rules**:

- associate to each propagator the factor $\frac{i}{p^2 - m^2 + i\epsilon}$.
- to each vertex the factor $-i\lambda$.
- impose momentum conservation at each vertex.
- integrate over undetermined internal momenta k , $\int \frac{d^4k}{(2\pi)^4}$.
- each diagram is to be divided by a symmetry factor, equal to the number of ways of interchanging components without changing the diagram.
- sum the contributions of all topologically distinct *connected* diagrams.

It can also be shown from Eq. (33) that the denominator cancels precisely all non-connected diagrams in the Feynman diagrams of the Green functions.

Perturbation theory is now one of the cornerstones of QFT. The anomalous magnetic moment of the electron was computed for the first time by Schwinger at one-loop in 1948 [16] (the factor below, $\frac{\alpha}{2\pi}$, is engraved on Schwinger's tombstone). Today it is known up to four-loops!

$$\begin{aligned} a_e &= \frac{g_e - 2}{2} = \frac{\alpha}{2\pi} + \cdots \\ a_e^{\text{exp}} &= (1159652185.9 \pm 3.8) \times 10^{-12} , \\ a_e^{\text{th}} &= (1159652175.9 \pm 8.5) \times 10^{-12} , \end{aligned} \quad (41)$$

where g_e is the gyromagnetic factor of electron coupling to a magnetic field. The theoretical prediction agrees with the experimental measurements in (41) up to the eight digit! There are however *still mysteries* in perturbation theory. For example, for the muon magnetic moment, the measured value at BNL disagrees by 3.4σ from the theoretical SM calculation

$$a_\mu^{\text{th}} = a_\mu^{\text{QED}} + a_\mu^{\text{EW}} + a_\mu^{\text{had}}$$

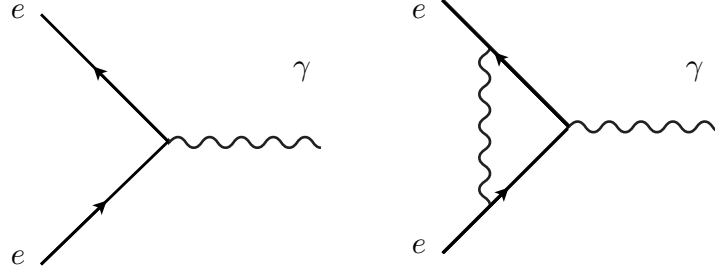


Fig. 3: Simplest Feynman diagrams contributions to the electron magnetic moment. The agreement between perturbative QED computations and the experimentally measured value agree up to the eight digit.

$$a_{\mu}^{\text{exp}} \simeq 0,00116592089.$$

In this case, it is likely that the hadronic contribution is not known accurately enough, since the muon mass is much closer to the hadronic contributions compared to the electron one. This is a very hot research topic nowadays, since any real disagreement could be a hint for new physics contributions coming from virtual loops of new particles.

3.5 Fermions

Relativistic fermions satisfying the Pauli principle are described by *spinors* in quantum field theory. In particular, the relativistic spin 1/2 fermion is described by a four component spinor Ψ via the Dirac equation

$$(i\gamma^m \partial_m - M)\Psi = 0, \quad (42)$$

where γ_m are the 4×4 Dirac matrices satisfying the Clifford algebra

$$\{\gamma^m, \gamma^n\} = 2\eta_{mn}. \quad (43)$$

The Lagrangian giving the Dirac equation is

$$\mathcal{L}_0 = \bar{\Psi}(i\gamma^m \partial_m - M)\Psi, \quad (44)$$

where $\bar{\Psi} = \Psi^\dagger \gamma^0$. A particular role is played by fermions which are eigenstates of the chirality operator, satisfying

$$\gamma^5 = i\gamma_0\gamma_1\gamma_2\gamma_3, \quad (\gamma^5)^2 = 1, \quad \{\gamma^m, \gamma^5\} = 0. \quad (45)$$

It is then possible to define left and right-handed chirality fermions

$$\begin{aligned} \gamma^5 \Psi_L &= -\Psi_L, & \Psi_L &= \frac{1 - \gamma^5}{2} \Psi, \\ \gamma^5 \Psi_R &= \Psi_R, & \Psi_R &= \frac{1 + \gamma^5}{2} \Psi. \end{aligned} \quad (46)$$

In terms of the left/right chirality fermions, the Dirac Lagrangian is written

$$\mathcal{L}_0 = \bar{\Psi}_L i\gamma^m \partial_m \Psi_L + \bar{\Psi}_R i\gamma^m \partial_m \Psi_R - M(\bar{\Psi}_L \Psi_R + \bar{\Psi}_R \Psi_L), \quad (47)$$

whereas the Dirac equation can be split into two equations

$$i\gamma^m \partial_m \Psi_L - M\Psi_R = 0, \quad i\gamma^m \partial_m \Psi_R - M\Psi_L = 0. \quad (48)$$

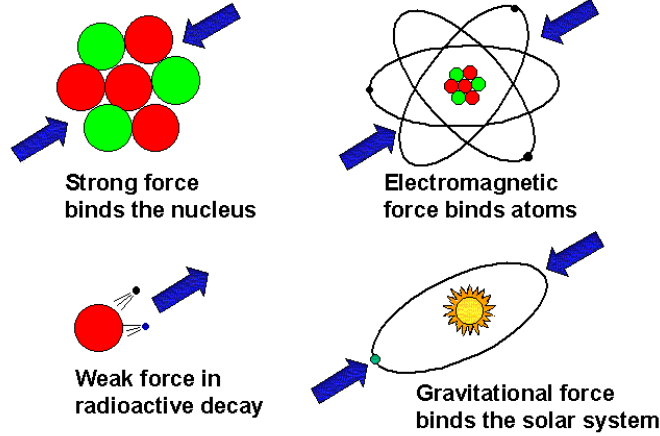


Fig. 4: The four fundamental interactions in nature. From Ref. [17].

As we will see in Section 6, in Nature left and right chirality fermions have different interactions. This is related to the parity violation in the weak interactions and it is at the heart of the construction of the Standard Model.

There are two different type of fermions that could exist in nature. The fermions charged under gauge symmetries are of Dirac type, i.e. their mass (eventually after symmetry breaking, as it will be the case in the Standard Model) is of Dirac type (47). For fermions uncharged under gauge symmetries, they can be of *Majorana* type. In this case, the charge conjugate fermion

$$\Psi^c = C\bar{\Psi}^T, \quad (49)$$

where C is the charge conjugation matrix satisfying

$$C^{-1} \gamma_m C = -\gamma_m^T, \quad C = -C^{-1} = -C^T = -C^\dagger \quad (50)$$

is self-conjugate $\Psi^c = \Psi$, i.e. the fermion is its own antiparticle. In this case, the mass of the fermion can be written as

$$\mathcal{L}_M = -\frac{M}{2} \Psi^T C \Psi + \text{h.c.} \quad (51)$$

It is not yet known if there exist Majorana fermions in nature. One natural possibility are the neutrinos, which will be discussed in detail in the lectures by B. Kayser [5].

4 Gauge theories

The four fundamental interactions in nature have a common feature: they are *gauge interactions*. We will discuss here the internal symmetries which describe the electromagnetic, weak and strong interactions and elements of their quantization.

4.1 Gauge invariance of Schrödinger equation

The simplest example of gauge symmetry arises in the description of particle of mass m and charge q in quantum mechanics. The hamiltonian is

$$H = \frac{1}{2m} (\mathbf{p} - q\mathbf{A})^2 + qV, \quad (52)$$

where the vector $\mathbf{A}$ and the scalar V potential are related to the electric/magnetic fields via

$$\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t} , \quad \mathbf{B} = \nabla \times \mathbf{A} . \quad (53)$$

The Maxwell equations are invariant under the gauge transformations

$$\mathbf{A}' = \mathbf{A} + \nabla \alpha , \quad V' = V - \frac{\partial \alpha}{\partial t} . \quad (54)$$

The Schrödinger equation is *covariant*, with $H = H(\mathbf{A}, V)$, $H' = H(\mathbf{A}', V')$

$$i\hbar \frac{\partial \Psi}{\partial t} = H\Psi \rightarrow i\hbar \frac{\partial \Psi'}{\partial t} = H'\Psi' \quad (55)$$

if under the gauge transformations (52), the wave function transforms as

$$\Psi'(\mathbf{r}, t) = e^{\frac{iq\alpha}{\hbar}} \Psi(\mathbf{r}, t) . \quad (56)$$

Note that the mean value of any physically measurable quantity is *gauge invariant*; for ex. $P(\mathbf{r}) = |\Psi|^2 = |\Psi'|^2$.

Homework: Defining the velocity operator $\mathbf{v} = \frac{1}{m}(\mathbf{p} - q\mathbf{A})$, check that $\langle \Psi | \mathbf{v} | \Psi \rangle = \langle \Psi' | \mathbf{v}' | \Psi' \rangle$.

Gauge principle: Postulate that physical laws are invariant under (54)+ (56). In this case, it can be proven that the hamiltonian is *uniquely determined* to be (52). Equations (56) and (54) define an $U(1)$ transformation. Therefore, $U(1)$ gauge invariance determines the electromagnetic interaction.

4.2 From Dirac and Maxwell equations to QED

Maxwell equations in terms of $A_\mu = (\mathbf{A}, V)$ are invariant under the *gauge transformations*

$$A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu \alpha . \quad (57)$$

Gauge invariance postulate: the physics is invariant under (57), supplemented with the phase transformation

$$\Psi(x) \rightarrow \Psi'(x) = e^{iq\alpha(x)} \Psi(x) . \quad (58)$$

Then Dirac equation is not invariant under (58) unless we replace the derivative with the *covariant derivative*

$$D_\mu \Psi \equiv (\partial_\mu + iqA_\mu) \Psi \rightarrow (D_\mu \Psi)' = (\partial_\mu + iqA'_\mu) \Psi' = e^{iq\alpha(x)} D_\mu \Psi(x) . \quad (59)$$

Dirac equation in an electromagnetic field becomes therefore

$$(i\gamma^\mu D_\mu - M) \Psi = (i\gamma^\mu \partial_\mu - q\gamma^\mu A_\mu - M) \Psi = 0 . \quad (60)$$

The Dirac and Maxwell equations can be derived from the Lagrangian density

$$\mathcal{L}_{QED} = \bar{\Psi}(i\gamma^\mu D_\mu - M)\Psi - \frac{1}{4}F_{\mu\nu}^2 . \quad (61)$$

The coupled Euler–Lagrange field equations are then Eq. (60) and

$$\partial^\mu F_{\mu\nu} = g\bar{\Psi}\gamma_\nu\Psi \equiv j_\nu , \quad (62)$$

where j_ν is the *electromagnetic current* of the charged fermion. From Eq. (62) we can derive the *charge conservation law*

$$\partial^\mu j_\mu = 0 \rightarrow \frac{dQ}{dt} = \int d^3\mathbf{x} \partial^\mu j_\mu = 0 , \quad \text{where} \quad Q = \int d^3\mathbf{x} j_0(\mathbf{x}) . \quad (63)$$

Some comments are in order

- The massless photon has *two propagating degrees of freedom*.
- A photon mass $\mathcal{L}_{mass} = \frac{M_A^2}{2} A_\mu^2$ *breaks gauge invariance* and describes *three degrees of freedom*.
- The propagator of a massive photon is found from inverting the free Lagrangian

$$\begin{aligned} -\frac{1}{4}F_{\mu\nu}^2 + \frac{M_A^2}{2}A_\mu^2 &= \frac{1}{2}A^\mu [g_{\mu\nu}(\square + M_A^2) - \partial_\mu \partial_\nu] A^\nu, \\ \Delta_{\mu\nu}^{-1}(x-y) &= -i(g_{\mu\nu}(\square + M_A^2) - \partial_\mu \partial_\nu) \delta^4(x-y) \end{aligned} \quad (64)$$

Therefore, in momentum space (**Homework**)

$$\Delta^{\mu\nu}(k) = -i \frac{g^{\mu\nu} - \frac{k^\mu k^\nu}{M_A^2}}{k^2 - M_A^2}. \quad (65)$$

Notice that due to the current conservation $\partial^\mu j_\mu = 0$, the longitudinal polarization does not contribute to amplitudes. Therefore, the UV properties of the massless and massive photon theories are *the same*. On the other hand, experimentally the photon is massless to a high accuracy. Indeed, the present experimental limit on the photon mass is $m_\gamma \leq 10^{-18}$ eV.

4.3 Fermions and the quantization of the Dirac field

According to perturbation theory, we start from the free Dirac Lagrangian

$$\mathcal{L}_0 = \bar{\Psi}(i\gamma^\mu \partial_\mu - M)\Psi. \quad (66)$$

Conjugate momentum is $\pi = \frac{\partial \mathcal{L}}{\partial \dot{\Psi}} = i\Psi^\dagger$. The free-field hamiltonian is then

$$H_0 = \int d^3\mathbf{x} \bar{\Psi}(i\gamma^\mu \nabla_\mu + M)\Psi = \int d^3\mathbf{x} \Psi^\dagger(i\alpha \nabla + \beta M)\Psi, \quad (67)$$

where $\gamma = \beta\alpha$, $\gamma_0 = \beta$ and in the last paranthesis we can recognize the Dirac hamiltonian of relativistic quantum mechanics. The solutions of the Dirac equation (66) are of the form

$$\Psi(x) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}\sqrt{2\omega_k}} \sum_{s=1,2} \left[e^{-ikx} a_{\mathbf{k}}^s u^s(k) + e^{ikx} b_{\mathbf{k}}^{s,\dagger} v^s(k) \right], \quad (68)$$

where $u^s(k)$ ($v^s(k)$) are positive (negative) frequency solutions of the Dirac equation

$$(\gamma^\mu k_\mu - M) u^s(k) = 0, \quad (\gamma^\mu k_\mu + M) v^s(k) = 0. \quad (69)$$

The Dirac equation have two independent solutions $s = 1, 2$. The correct quantization for fermions uses anti-commutators

$$\{\Psi_\alpha(t, \mathbf{x}), \Psi_\beta^\dagger(t, \mathbf{y})\} = \delta_{\alpha\beta} \delta^3(\mathbf{x} - \mathbf{y}), \quad (70)$$

all the other anticommutators being zero. This defines the anticomutation relations

$$\{a_{\mathbf{p}}^r, a_{\mathbf{q}}^{s,\dagger}\} = \{b_{\mathbf{p}}^r, b_{\mathbf{q}}^{s,\dagger}\} = \delta^{rs} \delta^3(\mathbf{p} - \mathbf{q}). \quad (71)$$

The vacuum is defined by $a_{\mathbf{p}}^r|0\rangle = b_{\mathbf{p}}^r|0\rangle = 0$, whereas the hamiltonian is given by

$$H_0 = \int d^3\mathbf{k} \sum_s \omega_k \left[a_{\mathbf{k}}^{s,\dagger} a_{\mathbf{k}}^s + b_{\mathbf{k}}^{s,\dagger} b_{\mathbf{k}}^s \right]. \quad (72)$$

Notice that if the theory would have been quantized with commutators, the contribution of the b oscillators would have been of opposite sign and the hamiltonian would have been unbounded from below. The electric charge operator can be defined as in (63) and equals

$$Q = \int d^3\mathbf{k} \sum_s \left[a_{\mathbf{k}}^{s,\dagger} a_{\mathbf{k}}^s - b_{\mathbf{k}}^{s,\dagger} b_{\mathbf{k}}^s \right]. \quad (73)$$

By defining also the helicity operator, it can shown that:

- $a_{\mathbf{k}}^{s,\dagger}$ creates fermions of energy ω_k , momentum $\mathbf{k}$, electric charge $+1$ (in units of the electron electric charge), helicity left (right) for $s = 1$ ($s = 2$).
- $b_{\mathbf{k}}^{s,\dagger}$ creates antifermions of energy ω_k , momentum $\mathbf{k}$, electric charge -1 and helicity right (left) for $s = 1$ ($s = 2$).

Similarly for the scalars case, there is a reduction/LSZ formula (32) and perturbative expansion (33) for the Green functions. The simplest and most important Green function is the fermionic Feynman propagator

$$\begin{aligned} S_F(x-y) &= \langle 0 | T \Psi(x) \bar{\Psi}(y) | 0 \rangle \\ &= \theta(x^0 - y^0) \langle 0 | \Psi(x) \bar{\Psi}(y) | 0 \rangle - \theta(y^0 - x^0) \langle 0 | \bar{\Psi}(y) \Psi(x) | 0 \rangle \\ &= \int \frac{d^4 k}{(2\pi)^4} \frac{i(\gamma^\mu k_\mu + M)}{k^2 - m^2 + i\epsilon} e^{-ik(x-y)}. \end{aligned} \quad (74)$$

4.4 Quantization of the electromagnetic field

The quantization of the free electromagnetic field is subtler than for the case of scalars and fermions. Indeed, starting from the Maxwell Lagrangian

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \quad (75)$$

the conjugate momentum is

$$\pi^\mu = \frac{\partial \mathcal{L}}{\partial \dot{A}_\mu} = F^{\mu 0} \Rightarrow \pi^0 = 0, \quad (76)$$

and we cannot impose canonical commutation relations. The problem can be avoided by using the non-covariant gauges, like the Coulomb gauge ($\text{div} \mathbf{A} = 0$), but it is preferable to maintain manifest Lorentz covariance. The standard option is to modify the Lagrangian by adding a *gauge-fixing term* and changing the Lagrangian to

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2\xi} (\partial_\mu A^\mu)^2, \quad (77)$$

where ξ is a real arbitrary (and unphysical) parameter. In this case the field equations become

$$\square A_\mu - (1 - \frac{1}{\xi}) \partial_\mu (\partial A) = 0, \quad (78)$$

and the canonical momentum $\pi^0 = -\frac{1}{\xi} (\partial_\mu A^\mu)$ does not vanishes anymore.

Observation: the field equations imply $\square (\partial A) = 0$, which suggests that we could impose the Lorentz condition $\partial A = 0$. This is however incompatible with canonical quantization, since $\pi^0 \sim \partial A$. The condition can be only be imposed on physical states $|\psi_{ph}\rangle$

$$\langle \psi_{ph} | \partial_\mu A^\mu | \psi_{ph} \rangle = 0. \quad (79)$$

It can be shown that physical results are independent of ξ . A convenient choice for canonical quantization is $\xi = 1$ (Feynman gauge), in which case the electromagnetic field become a collection of four Klein-Gordon fields $\square A_\mu = 0$. In this case, it can be expanded in plane waves according to

$$A_\mu(x) = \int \frac{d^3 \mathbf{k}}{(2\pi)^{3/2} \sqrt{2\omega_k}} \sum_{r=0}^3 \left[e^{-ikx} a_{\mathbf{k}}^r \epsilon_\mu^r(k) + e^{ikx} a_{\mathbf{k}}^{r,\dagger} \bar{\epsilon}_\mu^r(k) \right], \quad (80)$$

where here $\omega_k = |\mathbf{k}| = k^0$ and $\epsilon_\mu^r(k)$ are the polarization vectors. Canonical quantization in this case goes as follows

$$[A_\mu(t, \mathbf{x}), \pi_\nu(t, \mathbf{y})] = -i\eta_{\mu\nu} \delta^3(\mathbf{x} - \mathbf{y}) \Rightarrow [A_\mu(t, \mathbf{x}), \dot{A}_\nu(t, \mathbf{y})] = -i\eta_{\mu\nu} \delta^3(\mathbf{x} - \mathbf{y}). \quad (81)$$

The commutation relations for the creation/annihilation operators then follow

$$[a_{\mathbf{k}}^r, a_{\mathbf{q}}^{s,\dagger}] = -\eta^{rs} \delta^3(\mathbf{k} - \mathbf{q}) . \quad (82)$$

Finally, the Feynman propagator in this gauge is

$$\langle 0|TA_\mu(x)A_\nu(y)|0\rangle = -\eta_{\mu\nu}D_F(x-y)|_{M=0} = -i\eta_{\mu\nu} \int \frac{d^4k}{(2\pi)^4} \frac{e^{-ik(x-y)}}{k^2 + i\epsilon} . \quad (83)$$

Finally, we can give **the Feynman rules for QED**:

- associate to each fermion propagator of momentum p the factor $\frac{i(\gamma^\mu p_\mu + M)}{p^2 - M^2 + i\epsilon}$.
- to each photon propagator of momentum p (in the $\xi = 1$ gauge) the factor $\frac{-i\eta_{\mu\nu}}{p^2 - M^2 + i\epsilon}$.
- to each vertex the factor $iQe\gamma^\mu$, where $Q = -1$ for the electron.
- to each external initial fermion the factor $u^s(p)$.
- to each external final fermion the factor $\bar{u}^s(p)$.
- to each external initial antifermion the factor $\bar{v}^s(p)$.
- to each external final antifermion the factor $v^s(p)$.
- to each external initial photon the factor $\epsilon_\mu(p)$.
- to each external final photon the factor $\bar{\epsilon}_\mu(p)$.

The reader can find more about the historical rise of QED in Ref. [14].

4.5 Non-abelian gauge theories

The action of $U(1)$ is a particular case of *unitary abelian transformation*. Another case of particular interest are the non-abelian transformations. The non-abelian $SU(n)$ transformations are described by $n \times n$ matrices U , satisfying

$$U^\dagger U = UU^\dagger = I \quad , \quad \det U = 1 . \quad (84)$$

The simplest case is $SU(2)$, proposed by Yang and Mills in 1954 [18]. Its simplest representation is a *doublet*

$$\Psi = \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix} , \quad \Psi' = U(\theta)\Psi , \quad \text{where } U(\theta) = e^{\frac{i}{2}g\theta_a\tau_a} , \quad (85)$$

where τ_a are the Pauli matrices and g is the $SU(2)$ gauge coupling. It turns out that *the number of gauge bosons W_μ^a equals the number of generators* (three for $SU(2)$). The most compact notation introduces a matrix

$$W_\mu = W_\mu^a \frac{\tau_a}{2} = \begin{pmatrix} W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & -W_\mu^3 \end{pmatrix} \equiv \begin{pmatrix} W_\mu^3 & \sqrt{2}W_\mu^+ \\ \sqrt{2}W_\mu^- & -W_\mu^3 \end{pmatrix} \quad (86)$$

Homework: show that

$$D_\mu \Psi \equiv (\partial_\mu - igW_\mu)\Psi \rightarrow (D_\mu \Psi)' = UD_\mu \Psi \quad \text{if } W_\mu \rightarrow W'_\mu = UW_\mu U^{-1} - \frac{i}{g}(\partial_\mu U)U^{-1} . \quad (87)$$

The infinitesimal gauge variation in component form is

$$\delta W_\mu^a = D_\mu \theta^a \equiv \partial_\mu \theta^a + g\epsilon_{abc}W_\mu^b \theta^c . \quad (88)$$

The field strength is built from

$$[D_\mu, D_\nu] = -ig F_{\mu\nu} . \quad (89)$$

$$\begin{aligned}
& \text{Feynman rules for Yang-Mills theories. Solid lines are fermion propagators, wavy lines are gauge propagators, whereas the dotted ones are scalars. The structure constants are defined for an arbitrary gauge group via } [T^a, T^b] = if^{abc}T^c. \text{ From Ref. [15].} \\
& \text{Diagram 1: } \text{Feynman rule for a fermion-gauge vertex.} \quad = \frac{i}{\sqrt{2}}(k_1 - k_2)^\rho \gamma^\mu + \text{cyclic} \\
& \text{Diagram 2: } \text{Feynman rule for a fermion-gauge vertex.} \quad = i\gamma^{\mu\rho}\gamma^{\nu\sigma} - \frac{i}{2}(\gamma^{\mu\nu}\gamma^{\rho\sigma} + \gamma^{\mu\sigma}\gamma^{\nu\rho}) \\
& \text{Diagram 3: } \text{Feynman rule for a scalar-scalar vertex.} \quad = \sqrt{2}f^{abc} \\
& \text{Diagram 4: } \text{Feynman rule for a scalar-gauge vertex.} \quad = -\frac{i}{\sqrt{2}}\delta^{ab}(k_1 - k_2)^\mu \\
& \text{Diagram 5: } \text{Feynman rule for a scalar-gauge vertex.} \quad = \frac{i}{2}\delta^{ab}\gamma^{\mu\nu} \\
& \text{Diagram 6: } \text{Feynman rule for a scalar-gauge vertex.} \quad = -i\delta^{ab}\gamma^{\mu\nu} \\
& \text{Diagram 7: } \text{Feynman rule for a fermion-gauge vertex.} \quad = \frac{i}{\sqrt{2}}\delta^{ab}\gamma^\mu \\
& \text{Diagram 8: } \text{Feynman rule for a fermion-gauge vertex.} \quad = -\frac{i}{\sqrt{2}}\delta^{ab}\gamma^\mu
\end{aligned}$$

Fig. 5: Feynman rules for Yang–Mills theories. Solid lines are fermion propagators, wavy lines are gauge propagators, whereas the dotted ones are scalars. The structure constants are defined for an arbitrary gauge group via $[T^a, T^b] = if^{abc}T^c$. From Ref. [15].

Homework: Show that

$$F_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu - ig[W_\mu, W_\nu], \quad F_{\mu\nu} \rightarrow F'_{\mu\nu} = UF_{\mu\nu}U^{-1}. \quad (90)$$

For $SU(2)$ this implies (**homework:**)

$$F_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g\epsilon_{abc}W_\mu^b W_\nu^c. \quad (91)$$

The Yang–Mills Lagrangian is finally given by

$$\begin{aligned}
\mathcal{L}_{YM} &= -\frac{1}{4}F_{\mu\nu}^a F^{a,\mu\nu} \\
&= -\frac{1}{4}(\partial_\mu W_\nu^a - \partial_\nu W_\mu^a)^2 - \frac{g}{2}\epsilon_{abc}\partial_\mu W_\nu^a W^{b,\mu} W^{c,\nu} - \frac{g^2}{4}\epsilon_{abc}\epsilon_{ade}W_\mu^b W_\nu^c W^{d,\mu} W^{e,\nu}. \quad (92)
\end{aligned}$$

Non-abelian gauge bosons have self-interactions, unlike the photon! The full Lagrangian describing interaction of Yang–Mills fields with charged fermions is then

$$\mathcal{L} = \bar{\Psi}(i\gamma^\mu D_\mu - M)\Psi - \frac{1}{4}F_{\mu\nu}^a F^{a,\mu\nu} + \mathcal{L}_\xi + \mathcal{L}_{\text{ghosts}} = \mathcal{L}_0 + \mathcal{L}_{\text{int}}. \quad (93)$$

In Eq. (93), $\mathcal{L}_\xi$ is a gauge fixing term, whereas $\mathcal{L}_{\text{ghosts}}$ is the Fadeev–Popov ghost Lagrangian [2, 19], coming from the covariant quantization of non-abelian theories.

Homework: show that for an $SU(2)$ doublet

$$\bar{\Psi}(i\gamma^\mu D_\mu - M)\Psi = \bar{\Psi}^k[\delta_{kl}(i\gamma^\mu \partial_\mu - M) + \frac{g}{2}\gamma^\mu W_\mu^a(\tau_a)_{kl}]\Psi^l, \quad (94)$$

whereas the fermion and Yang–Mills field equations are

$$(i\gamma^\mu D_\mu - M) \Psi = 0, \quad \partial^\mu F_{\mu\nu}^a + g\epsilon_{abc}A^{b,\mu}F_{\mu\nu}^c = -g\bar{\Psi}\gamma_\nu \frac{\tau_a}{2}\Psi \quad (95)$$

where on the right hand side, we can identify the $SU(2)$ fermionic current $j_n^a = -g\bar{\Psi}\gamma_\nu \frac{\tau_a}{2}\Psi$.

Notice that the massive Yang–Mills field propagator is

$$\Delta_{\mu\nu}^{ab}(k) = \delta^{ab} \frac{g_{\mu\nu} - \frac{k_\mu k_\nu}{M_A^2}}{k^2 - M_A^2}. \quad (96)$$

Since here $\partial^\mu j_\mu^a \neq 0$, the longitudinal polarization *does contribute* to scattering amplitudes. Therefore, unlike the abelian case, here the UV properties of the massless and massive YM theories are different. This fact has various consequences:

- the theory has bad UV behaviour (uncontrolled UV divergences).
- the amplitude $W_L W_L \rightarrow W_L W_L$, where W_L is the longitudinal component of the W gauge boson, grows with energy and invalides perturbation theory for energies above around 1.2 TeV.

The conclusion of all these problems is that the Yang–Mills boson masses *should not be added by hand*, but be generated in a more subtle way. On the other hand, massless gauge fields (infinite range) cannot describe electroweak interactions, which are short range. We need therefore to give gauge bosons a mass, but we need *need another way* to generate gauge boson masses. This is explained via the spontaneous symmetry breaking and the Higgs mechanism, to which we now turn.

5 Spontaneous symmetry breaking

We already noticed that *Symmetries*, through the Noether theorem, imply the existence of conserved charges. There are however two different ways symmetries are realized in nature:

1. **Weyl–Wigner (WW) realization:** in this case the vacuum state is *invariant* under the symmetry. Then the *symmetry is manifest* in the spectrum and the interactions. Simple examples of this type are: translations (conserved charge: momentum), rotations (conserved charge: angular momentum), $U(1)_{em}$ (conserved charge: electric charge)...
2. **Nambu–Goldstone (NG) realization:** in this case the vacuum state is *not invariant* under the symmetry. In this case the *symmetry is not manifest* in the spectrum. We talk about spontaneous symmetry breaking. Examples of this type include rotation (or parity) symmetry in ferromagnets, $SU(2)_{weak}$, $SU(2)_L \times SU(2)_R$ chiral symmetry of strong interactions, etc

A nice sentence summarizing the outcome of the two realizations of global symmetries is that of S. Coleman in his Erice Lectures [24]: *"the symmetry of the vacuum is the symmetry of the world"*.

The simplest example of the NG realization is the Ising model describing N spins in space dimension d , of hamiltonian

$$H = -J \sum_{(i,j)} S_i S_j - B \sum_i S_i, \quad (97)$$

with $S_i = \pm 1$ labelling the two possible values of the spin " i ". For zero magnetic field $B = 0$ the system has a Z_2 symmetry which reverts the spins $S_i \rightarrow -S_i$. As a consequence, the magnetization defined as

$$M = \lim_{B \rightarrow 0, N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \langle S_k \rangle$$

should therefore vanish. However experimentally it is known that

$$M = 0 \text{ for } T \geq T_c, \quad M \neq 0 \text{ for } T < T_c, \quad \text{where } kT_c = 2dJ. \quad (98)$$

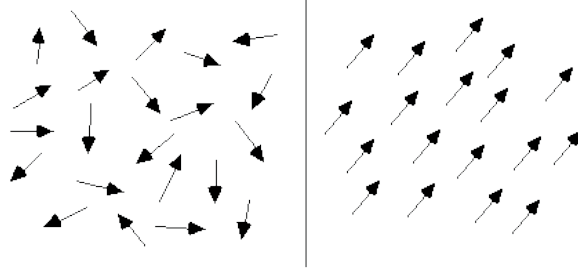


Fig. 6: Magnetization in ferromagnets. At high temperatures, spins orientation are random. At low temperatures, the alignment of spins due to spin-spin interactions breaks the rotational symmetry of the hamiltonian.

The reason for the violation of the Z_2 symmetry is that at low temperatures, due to the spin-spin interactions of strength J , spins tend to align, such that the ground state correspond to a state with all spins aligned. This state does violate the Z_2 symmetry, since the Z_2 transform of this ground state is the state with all spins reversed. Whereas both states (vacua) are equally possible, the transition from one to the other is highly suppressed for large N . So if the system is in one of the two vacua, it will stay there a time that scales as e^N . On the other hand, at high-temperature, spins are oriented arbitrarily in order to increase the entropy (number), which wins over the higher-energy of such configurations. This phenomenon is called spontaneous symmetry breaking, since the hamiltonian of the system respects the Z_2 symmetry, which is broken only by the ground state for $T < T_c$. The field theory analog of this phenomenon is described in the next paragraph.

5.1 The Goldstone theorem

In a theory with continuous symmetry, for every generator which does not annihilate the vacuum $\langle T^a \Phi \rangle \neq 0$ there is a massless, NG particle [20].

Example: One of the simplest examples is the $O(N)$ linear sigma model. Consider a theory with N scalar fields $\Phi = (\phi_1, \phi_2, \dots, \phi_N)$, with Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \Phi)^2 - V(\Phi), \quad V(\Phi) = -\frac{\mu^2}{2}\Phi^2 + \frac{\lambda}{4}(\Phi^2)^2, \quad (99)$$

where in our convention $\mu^2 > 0$ and where $\Phi^2 = \sum_{i=1}^N \phi_i \phi_i$. The model has a continuous $O(N)$ symmetry acting as $\Phi \rightarrow R\Phi$, with R a $N \times N$ rotation matrix. The scalar potential is minimized for

$$\frac{\partial V}{\partial \phi_i} = 0 \quad \Rightarrow \quad \Phi_0^2 = \frac{\mu^2}{\lambda} \equiv v^2. \quad (100)$$

The vacuum manifold is $O(N)$ invariant. By an $O(N)$ rotation, the ground state can be chosen to be

$$\langle \Phi \rangle = \Phi_0 = (0, 0, \dots, v), \quad (101)$$

preserving an $O(N-1)$ subgroup. Goldstone's theorem tells us that we expect the model to have $N-1$ massless particles, corresponding to the number of broken generators of the coset group $O(N)/O(N-1)$. In order to check this, we define a set of shifted fields:

$$\Phi(x) = (\pi^k(x), v + \sigma(x)), \quad k = 1, 2, \dots, N-1, \quad (102)$$

such that $\langle \pi^k \rangle = \langle \sigma \rangle = 0$. The Lagrangian becomes

$$\mathcal{L} = \frac{1}{2}((\partial_\mu \pi)^2 + (\partial_\mu \sigma)^2) - \mu^2 \sigma^2 - \sqrt{\lambda} \mu \sigma^3 - \sqrt{\lambda} \mu \pi^2 \sigma - \frac{\lambda}{4}(\sigma^2 + \pi^2)^2, \quad (103)$$

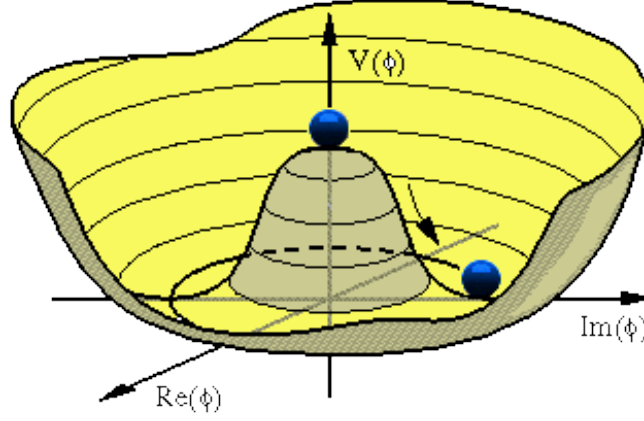


Fig. 7: Spontaneous symmetry breaking. The vacuum manifold minima is symmetric by rotations, but picking up a vacuum breaks the symmetry spontaneously. Figure taken from Ref. [21].

where $\pi^2 = \sum_{k=1}^{N-1} \pi^k \pi^k$. The manifest symmetry is indeed $O(N-1)$, which rotates the "pions" π 's among themselves. The physical masses, visible from (103) are

$$m_\sigma^2 = 2\mu^2, \quad m_{\pi_k}^2 = 0. \quad (104)$$

Therefore we find that *the "pions" are massless*; they are the $N-1$ Nambu–Goldstone (NG) bosons of the broken symmetry. It is said that the unbroken $O(N-1)$ symmetry is realized a la Weyl–Wigner, whereas the original $O(N)$ symmetry is realized a la Nambu–Goldstone.

General (classical) proof of the Goldstone theorem. Consider the scalar theory of Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial_m \Phi_i)^2 - V(\Phi_i), \quad (105)$$

and a *global continuous symmetry group* with generators T^a . The invariance of the scalar potential

$$V(\phi_i + \delta\phi_i) = V(\phi_i) \quad (106)$$

under infinitesimal transformations $\delta\phi_i = i\theta^a T_{ij}^a \phi_j$ of parameters θ^a implies

$$\frac{\partial V}{\partial \phi_i} T_{ij}^a \phi_j = 0. \quad (107)$$

Differentiating again and taking the vacuum expectation value (vev), we get

$$\left\langle \frac{\partial^2 V}{\partial \phi_k \partial \phi_i} T_{ij}^a \phi_j + \frac{\partial V}{\partial \phi_i} T_{ik}^a \right\rangle = 0. \quad (108)$$

Remembering that $\mathcal{M}_{ki}^2 = \langle \frac{\partial^2 V}{\partial \phi_k \partial \phi_i} \rangle$ is the scalar mass matrix, we obtain

$$\mathcal{M}_{ki}^2 (T^a v)_i = 0. \quad (109)$$

We therefore found the general form of the **Goldstone theorem**: *If the vacuum is not invariant under a symmetry generator $T^a v \neq 0$, then $T^a v$ is an eigenvector of the mass matrix $\mathcal{M}^2$ corresponding to a zero eigenvalue.*

Are there known examples of Goldstone bosons in nature? Yes, there are several, but none of them *not corresponding to a fundamental spin 0 particle*! Two well-known examples are

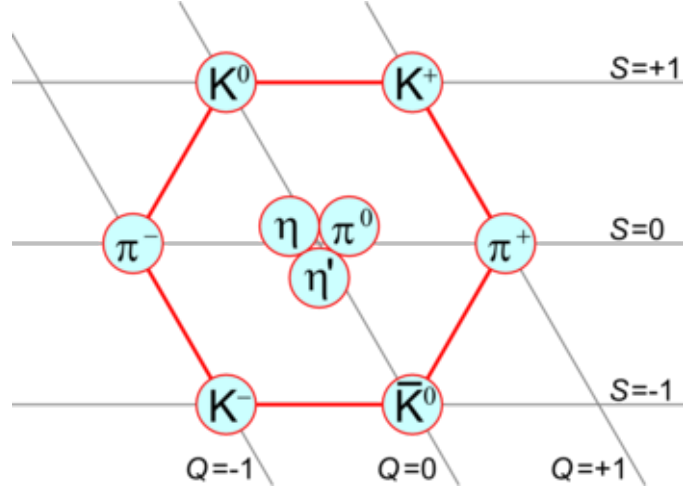


Fig. 8: The meson octet acts as goldstone bosons of the chiral symmetry breaking $SU(3)_L \times SU(3)_R \rightarrow SU(3)_V$. Figure taken from Ref. [22].

- Magnons spin waves in ferromagnets, which are long wavelength collective spin configurations.
- Pions $\pi \sim q \bar{q}$ are *pseudo-Goldstones* for the breaking of the *chiral* $\rightarrow$ vector symmetries $U(3)_L \times U(3)_R \rightarrow SU(3)_V \times U(1)_B$ (see figure 8). They are not exactly massless (therefore the name "pseudo") due to a small explicit breaking coming from quark masses. Pions are (pseudo)scalar particles, but not elementary, they are quark-antiquark bound states. In this case we talk about *dynamical* symmetry breaking.

Observation: The $U(1)_A$ symmetry is broken by quantum anomalies hence there is no corresponding goldstone boson.

A natural question arises: *What happens if the spontaneously broken symmetry is gauged?* The answer is given in the next subsection.

5.2 The Higgs mechanism

Let us start for simplicity with an *abelian gauge theory*

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^2 + |D_\mu \Phi|^2 - V(\Phi), \quad (110)$$

with $D_\mu = \partial_\mu + ieA_\mu$, $\Phi = \frac{1}{\sqrt{2}}(\Phi_1 + i\Phi_2)$ and a scalar potential

$$V = -\mu^2|\Phi|^2 + \lambda(|\Phi|^2)^2 = -\frac{\mu^2}{2}(\Phi_1^2 + \Phi_2^2) + \frac{\lambda}{4}(\Phi_1^2 + \Phi_2^2)^2, \quad (111)$$

invariant under the *local* $U(1)$ gauge transformations

$$\Phi \rightarrow e^{i\alpha(x)}\Phi, \quad A_\mu \rightarrow A_\mu - \frac{1}{e}\partial_\mu\alpha. \quad (112)$$

We expand around the vacuum state

$$\Phi_0 = \sqrt{\frac{\mu^2}{2\lambda}} = \frac{v}{\sqrt{2}}, \quad \Phi(x) = \frac{1}{\sqrt{2}}(v + \phi_1 + i\phi_2). \quad (113)$$

From the quadratic mass terms in the scalar potential we find $m_1^2 = 2\mu^2$, $m_2 = 0$, therefore ϕ_2 is the Goldstone boson. New features appear however from the kinetic term

$$|D_\mu \Phi|^2 = \frac{1}{2}(\partial_\mu \phi_i)^2 + evA_\mu \partial^\mu \phi_2 + \frac{e^2 v^2}{2} A_\mu^2 + \dots \quad (114)$$

Indeed, it is manifest from (114) that *the gauge boson acquired a mass $M_A^2 = e^2 v^2$* . But this can only happen if the gauge field absorbed one degree of freedom, since the massive gauge field has three degrees of freedom, whereas the massless one has only two degrees of freedom. The correct counting of degrees of freedom is

$$A_\mu (M_A = 0) + \phi_2 \rightarrow A_\mu (M_A \neq 0) \quad (115)$$

That this is indeed true can be seen in various ways:

1. The quadratic term in the Lagrangian can be diagonalized by redefining the gauge field

$$-\frac{1}{4}F_{\mu\nu}^2 + \frac{1}{2}(\partial_\mu \phi_2)^2 + \sqrt{2}evA_\mu \partial^\mu \phi_2 + \frac{e^2 v^2}{2} A_\mu^2 = -\frac{1}{4}(\partial_\mu B_\nu - \partial_\nu B_\mu)^2 + \frac{e^2 v^2}{2} B_\mu^2, \quad (116)$$

where $B_\mu = A_\mu + \frac{1}{ev} \partial_\mu \phi_2$. Therefore ϕ_2 *disappeared from the quadratic part, and is "absorbed" into the longitudinal component of the gauge field*.

2. The Goldstone can be *eliminated altogether* from the Lagrangian in the so-called *unitary gauge*. The corresponding parametrization is

$$\Phi(x) = \frac{1}{\sqrt{2}} e^{\frac{i\theta(x)}{v}} (v + \rho(x)) \quad (117)$$

and the Goldstone is removed by the gauge transformation $\Phi \rightarrow \Phi' = e^{-\frac{i\theta}{v}} \Phi$, $A_\mu \rightarrow A'_\mu = A_\mu + \frac{1}{ev} \partial_\mu \theta$. In the unitary gauge, the Lagrangian is (**homework:**)

$$\mathcal{L} = -\frac{1}{4}(F'_{mn})^2 + (\partial_m - ieA'_m)\Phi'(\partial^m + ieA'^m)\Phi' - \mu^2 \Phi'^2 - \lambda \Phi'^4$$

The spectrum of the model contains therefore a massive gauge boson and *the Higgs boson Φ' , of mass $2\mu^2$ [1]*.

The Higgs mechanism, non-abelian case. Consider a gauge group G of rank r and scalar fields in some irreducible n -dimensional representation

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^a F^{a,\mu\nu} + |[(\partial_\mu - igT^a A_\mu^a)\Phi]|^2 - V(\Phi), \quad (118)$$

and $H \in G$ the subgroup of rank s leaving the ground state invariant

$$\begin{aligned} T^a v &= 0, & a &= 1 \dots s, \\ T^a v &\neq 0, & a &= s+1 \dots r. \end{aligned} \quad (119)$$

In the unitary gauge parametrization

$$\Phi(x) = e^{i\sum_{a=s+1}^r T_a \frac{\xi_a(x)}{v}} \frac{\rho(x) + v}{\sqrt{2}}, \quad (120)$$

where ξ_a are the Goldstone bosons, $\rho(x)$ the remaining scalar fields, and $\langle \xi_a \rangle = \langle \rho \rangle = 0$. The gauge transformation

$$\Phi(x) \rightarrow \Phi'(x) = U\Phi, \text{ with } U = e^{-i\sum_{a=s+1}^r T_a \frac{\xi_a(x)}{v}}$$

$$A_\mu \rightarrow A'_\mu = U \left(A_\mu + \frac{i}{g} \partial_\mu \right) U^{-1} \quad (121)$$

eliminates the Goldstone bosons from the Lagrangian. The resulting mass matrix of the vector fields is then

$$M_{ab}^2 = g^2 (T_a v)^\dagger (T_b v) . \quad (122)$$

In this case $r - s$ gauge bosons become massive

$$A_\mu^a \text{ " + " } \xi_a \rightarrow A'^a_\mu = A_\mu^a - \frac{1}{v} D_\mu \xi^a + \dots \quad (123)$$

where A_μ denote the massless gauge fields, containing two degrees of freedom, whereas A'_μ denote the massive gauge fields.

Notice that *the number of physical massive Higgs scalars is equal to the number of original scalars, minus the number of broken gauge generators*. In particular:

- In the Standard Model with *one* Higgs doublet, there is one real Higgs scalar $4 - 3 = 1$, where 4 is the number of initial real degrees of freedom contained into an $SU(2)_L$ scalar doublet, whereas 3 is the number of broken generators in the Standard Model.
- In the Standard Model with *two* Higgs doublets (or the Minimal Supersymmetric Standard Model, MSSM), there are $8 - 3 = 5$ physical Higgs scalars: two neutral scalars h and H^0 , one pseudoscalar A and two charge ones $H^\pm$.

The Higgs mechanism is a elegant and economical way to break electroweak symmetry. *However*, it has its own misteries:

- Elementary scalars were never observed in nature until now.
- It is difficult to keep a scalar light after quantum corrections (*so-called hierarchy problem*). We will come back later on to quantify this problem.

Taken into account these observations, it is reasonable to ask is there are other ways of breaking a gauge (electroweak for our purposes) symmetry. The answer is yes, there are several other options. Some popular ones are:

- A new confining force (*technicolor*) with $\Lambda_{TC} \sim v$. The goldstone bosons "eaten up" by the W and Z gauge bosons are called "technipions" (see Fig. 9).
- composite Higgs models, in which Higgs is a bound state of fermions. One example is the top-antitop condensation, where the Higgs is a top-antitop bound-state $h = \bar{t}_L t_R$.
- symmetry breaking by boundary conditions in extra-dimensional Kaluza–Klein (string) type theories.

6 The electroweak sector of the Standard Model

6.1 Gauge group and matter content

The *Standard model* is a "unified" description of weak and electromagnetic interactions. From the Fermi theory of weak interactions with $G_F/\sqrt{2} = g^2/8M_W^2$, we know that we need a theory that contains at least a charged gauge boson $W_\mu^\pm$ and the photon A_μ .

Experimentally, there also exists *neutral currents* discovered in 1973, mediated by a neutral massive gauge boson, and also *coloured strong interactions*. The SM gauge group is therefore Gauge bosons:

$$G = \begin{matrix} G_\mu^A \\ SU(3)_c \end{matrix} \times \begin{matrix} A_\mu^a \\ SU(2)_L \end{matrix} \times \begin{matrix} B_\mu \\ U(1)_Y \end{matrix} \quad (124)$$

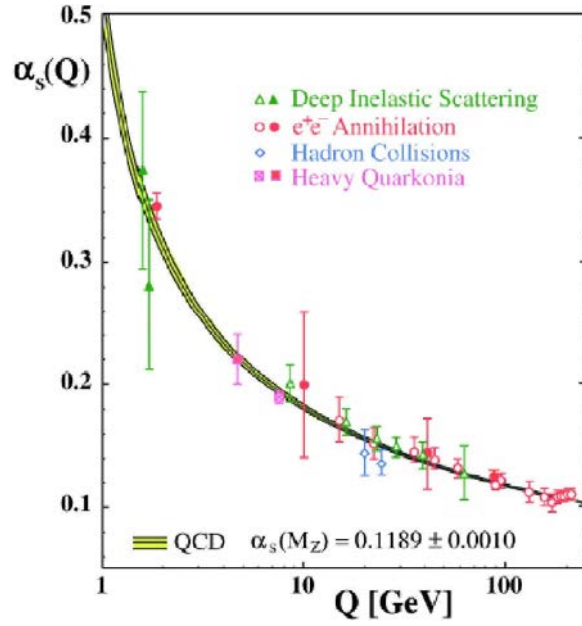


Fig. 9: A confining force similar to QCD called technicolor could be responsible for electroweak symmetry breaking. The electroweak vev is given by a condensate of fermions $\langle \bar{T}_L T_R \rangle \sim v^3 = M_P^3 e^{-3/(2b_0 g_{TC}^2)}$. Taken from Ref. [23].

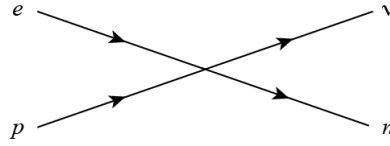


Fig. 10: Fermi theory of weak interactions: the beta decay $n \rightarrow pe^- \bar{\nu}_e$ at low energies $E \ll M_W$ can be described as an effective four-fermion interaction.

In addition to the gauge bosons, the SM contains matter fermions and the Higgs field, in the gauge group representations

- Leptons: $l_i = \begin{pmatrix} \nu_i \\ e_i \end{pmatrix}_L : (\mathbf{1}, \mathbf{2})_{Y=-1}$, and $e_{iR} : (\mathbf{1}, \mathbf{1})_{Y=-2}$
- Quarks: $q_i = \begin{pmatrix} u_i \\ d_i \end{pmatrix}_L : (\mathbf{3}, \mathbf{2})_{Y=1/3}$, $u_{iR} : (\mathbf{3}, \mathbf{1})_{Y=4/3}$, and $d_{iR} : (\mathbf{3}, \mathbf{1})_{Y=-2/3}$
- Higgs field: $\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} : (\mathbf{1}, \mathbf{2})_{Y=1}$.

In the Standard Model, the *Higgs doublet* vev breaks the electroweak gauge sector down to the electric charge $SU(2)_L \times U(1)_Y \rightarrow U(1)_Q$. Notice that only *left-handed* quarks and leptons interact with $SU(2)_L$ gauge fields. The SM Lagrangian has the symbolic form²

$$\mathcal{L}_{SM} = \mathcal{L}_{\text{kin}} - V(\Phi) + \mathcal{L}_{\text{Yuk.}}, \quad (125)$$

²No gluons in what follows. For strong interactions, see the *heavy-ions and QCD lectures* by Edmond Iancu and Fabio Maltoni.

where

$$\mathcal{L}_{\text{kin}} = -\frac{1}{4}F_{\mu\nu}^2 - \frac{1}{4}(F_{\mu\nu}^a)^2 + |D_\mu\Phi|^2 + \bar{\Psi}_L i\gamma^\mu D_\mu \Psi_L + \bar{\Psi}_R i\gamma^\mu D_\mu \Psi_R, \quad (126)$$

and

$$D_\mu \Psi_L = (\partial_\mu - ig\frac{\tau_a}{2}A_\mu - ig'\frac{Y_L}{2}B_\mu)\Psi_L, \quad D_\mu \Psi_R = (\partial_\mu - ig'\frac{Y_R}{2}B_\mu)\Psi_R, \quad (127)$$

whereas the Higgs potential has the form

$$V(\Phi) = -\mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2. \quad (128)$$

The Yukawa sector $\mathcal{L}_{\text{Yuk.}}$ will be discussed later on. With our conventions the electric charge is related to the third component of the isospin T_3 and to the hypercharge Y via

$$Q = T_3 + \frac{Y}{2}. \quad (129)$$

6.2 Weak mixing angles and gauge boson masses

With the help of an $SO(4)$ rotation, the Higgs vev can be written as

$$\Phi = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \text{ where } v^2 = \frac{\mu^2}{\lambda} \simeq (246 \text{ GeV})^2 \quad (\text{from experimental data}). \quad (130)$$

Gauge boson *masses* arise from the covariant derivative (**homework:**)

$$|D_\mu\Phi|^2 \rightarrow \frac{g^2 v^2}{8}|A_\mu^1 - iA_\mu^2|^2 + \frac{v^2}{8}|gA_\mu^3 - g'B_\mu|^2 = \frac{g^2 v^2}{4}W^{+\mu}W_\mu^- + \frac{(g^2 + g'^2)v^2}{8}Z^\mu Z_\mu, \quad (131)$$

where the definitions and the masses of gauge bosons are

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(A_\mu^1 \mp iA_\mu^2), & M_W &= \frac{gv}{2}, \\ Z_\mu &= \frac{gA_\mu^3 - g'B_\mu}{\sqrt{g^2 + g'^2}}, & M_Z &= \frac{v}{2}\sqrt{g^2 + g'^2}, \\ A_\mu &= \frac{g'A_\mu^3 + gB_\mu}{\sqrt{g^2 + g'^2}}, & M_A &= 0. \end{aligned} \quad (132)$$

We now introduce the *electroweak mixing angle*

$$\cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}} = \frac{M_W}{M_Z}, \quad \tan\theta_W = \frac{g'}{g}, \quad (133)$$

that rotates from the weak basis to the mass basis

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} A_\mu^3 \\ B_\mu \end{pmatrix}. \quad (134)$$

Notice that the ratio

$$\rho \equiv \frac{M_W^2}{M_Z^2 \cos^2\theta_W} = 1 \text{ at tree-level in the SM.} \quad (135)$$

The ρ parameter has quantum corrections in the SM, which are dominated by the top quark. Any experimental deviation from the SM value is a possible hint of new physics. Conversely, any model of new physics has to be able to produce a ρ parameter close to one, which is one of the *precision tests* of the Standard Model. This is a killer for most proposals of Beyond the Standard Model physics. For example, technicolor-like theories have difficulties in this respect (although there is no formal proof that they cannot accomodate precision data). Finally, the definition of the electric charge is $e = g \sin\theta_W$.

6.3 Neutral and charged currents

The neutral and charged currents are defined as the fermion bilinears coupling to the charged (W) and neutral (Z) gauge fields. They are worked out starting from the fermionic kinetic terms.

Homework: With the definitions above, show that

$$\begin{aligned} D_\mu &\equiv \partial_\mu - igA_\mu^a \frac{\tau_a}{2} - ig' \frac{Y}{2} B_\mu \\ &= \partial_\mu - ieQA_\mu - \frac{ig}{2\sqrt{2}}(W_\mu^+ \tau_+ + W_\mu^- \tau_-) - \frac{ig}{\cos \theta_W} Z_\mu (T_3 - \sin^2 \theta_W Q). \end{aligned} \quad (136)$$

The fermionic currents are defined as

$$\mathcal{L} \equiv \bar{\Psi}_i i \gamma^\mu \partial_\mu \Psi_i + \frac{g}{\sqrt{2}} (W_\mu^+ J_W^{\mu+} + W_\mu^- J_W^{\mu-}) + \frac{g}{\cos \theta_W} Z_\mu J_Z^\mu + e A_\mu J_{em}^\mu \quad (137)$$

Homework: Using the quantum numbers of the quarks/leptons, show that

$$\begin{aligned} J_W^{\mu,+} &= \bar{v}_L^i \gamma^\mu e_L^i + \bar{u}_L^i \gamma^\mu d_L^i \\ J_W^{\mu,-} &= \bar{e}_L^i \gamma^\mu \nu_L^i + \bar{d}_L^i \gamma^\mu u_L^i, \\ J_{em}^\mu &= -\bar{e}^i \gamma^\mu e^i + \frac{2}{3} \bar{u}^i \gamma^\mu u^i - \frac{1}{3} \bar{d}^i \gamma^\mu d^i, \\ J_Z^\mu &= J^{\mu,3} - \sin^2 \theta_W J_{em}^\mu \\ &= \frac{1}{2} \bar{v}_L^i \gamma^\mu \nu_L^i + \left(-\frac{1}{2} + \sin^2 \theta_W \right) \bar{e}_L^i \gamma^\mu e_L^i + \sin^2 \theta_W \bar{e}_R^i \gamma^\mu e_R^i \\ &\quad + \left(\frac{1}{2} - \frac{2}{3} \sin^2 \theta_W \right) \bar{u}_L^i \gamma^\mu u_L^i - \frac{2}{3} \sin^2 \theta_W \bar{u}_R^i \gamma^\mu u_R^i \\ &\quad + \left(-\frac{1}{2} + \frac{1}{3} \sin^2 \theta_W \right) \bar{d}_L^i \gamma^\mu d_L^i + \frac{1}{3} \sin^2 \theta_W \bar{d}_R^i \gamma^\mu d_R^i. \end{aligned} \quad (138)$$

At low energies $E \ll M_W, M_Z$, the exchange of W and Z bosons leads to the Fermi charged current four-fermion interaction, plus a similar neutral current interaction

$$\mathcal{L}_F = -2\sqrt{2}G_F \left[J_W^{\mu+} J_{\mu-}^W + \rho J_Z^\mu J_\mu^Z \right], \quad (139)$$

where we defined the parameter $\rho = \frac{M_W^2}{M_Z^2 \cos^2 \theta_W}$ which, as we noticed in the previous paragraph, equals one at tree-level in the Standard Model and plays an important role in quantum corrections and constraints on new physics.

6.4 Fermion masses and the CKM matrix

Dirac mass terms in the SM are *not gauge invariant*, due to the chiral nature of electroweak interactions. We can however write Yukawa-type interactions by using the Higgs field

$$-\mathcal{L}_{\text{Yuk}} = h_{ij}^u \bar{q}_L^i u_R^j \tilde{\Phi} + h_{ij}^d \bar{q}_L^i d_R^j \Phi + h_{ij}^e \bar{l}_L^i e_R^j \Phi, \quad (140)$$

where: $\tilde{\Phi} = \begin{pmatrix} \Phi^0 \\ -\Phi^+ \end{pmatrix}$ is the charge-conjugate Higgs field and $i, j = 1, 2, 3$ are flavor indices. The Yukawa couplings generate quarks and lepton masses after the electroweak symmetry breaking:

$$-\mathcal{L}_{\text{mass}} = m_{ij}^u \bar{u}_L^i u_R^j + m_{ij}^d \bar{d}_L^i d_R^j + m_{ij}^l \bar{e}_L^i e_R^j + \text{c.c.}, \quad (141)$$

where $m_{ij}^{u,d,l} = h_{ij}^{u,d,l} v / \sqrt{2}$. We use in what follows for compactness a matrix notation

$$-\mathcal{L}_{\text{mass}} = \bar{u}_L m^u u_R + \bar{d}_L m^d d_R + \bar{e}_L m^l e_R + \text{c.c.}, \quad (142)$$

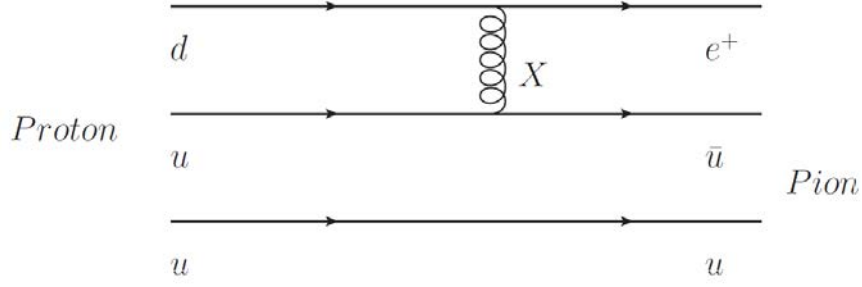


Fig. 11: Diagram leading to proton decay $p \rightarrow \pi^0 e^+$ in unified theories. The superheavy X particle is a GUT gauge boson.

where $m^{u,d,l}$ are 3×3 mass matrices in the flavor space.

Observation: The SM Lagrangian has some automatic (consequences of the gauge symmetries) *global symmetries*:

$$\begin{aligned} \text{baryon number} & \quad U(1)_B \quad , \\ \text{lepton numbers} & \quad U(1)_e \quad , \quad U(1)_\mu \quad , \quad U(1)_\tau \quad . \end{aligned} \quad (143)$$

This is actually very *fortunate* since there are *very strong experimental constraints* on baryon and lepton number violating processes, for example:

$$\begin{aligned} - \text{proton lifetime} \quad \tau_p & \geq 10^{33} \text{ years} \quad , \\ - \text{BR}(\mu \rightarrow e \gamma) & < 2.4 \times 10^{-12} \quad , \quad \text{BR}(\mu^- \rightarrow e^- e^- e^+) < 10^{-10} \quad , \\ - \text{BR}(B \rightarrow X_s \gamma) & \sim 10^{-4} \Rightarrow b \rightarrow s \gamma \text{ should be suppressed} \quad . \end{aligned} \quad (144)$$

These limits constrain seriously any higher-dimensional operator violating flavor, generated by eventual new physics. For example, consider the operator

$$\mathcal{L}_{\text{eff}} \sim \frac{1}{M_X^2} (\bar{q} \gamma_\mu u_R^c) (\bar{l} \gamma^\mu d_R^c) \quad . \quad (145)$$

The bound on the proton lifetime constrains the mass to be heavier than about $\Rightarrow M_X \geq 3 \times 10^{16} \text{ GeV}$. There is a long list of similar effective operators that are tightly constrained by the data. Another simple example is:

$$\mathcal{L}_{\text{eff}} \sim \frac{1}{M^2} (\bar{l}_2 \gamma_m l_1) (\bar{l}_1 \gamma^m l_1) \rightarrow \frac{1}{M^2} (\bar{\mu} \gamma_m e) (\bar{e} \gamma^m e) \quad , \quad (146)$$

that can be generated by a *flavor-dependent Z' gauge boson*. The mass scale M is constrained by the limits on $\mu^- \rightarrow e^- e^- e^+$ to be heavier than $M > 1000 \text{ TeV}$. It turns out that almost all SM extension generates dangerous FCNC and/or proton decay, unless special structure. For example, in *MSSM* we have to impose (i) *R-parity*, (ii) flavor blindness of soft terms.

Observation: With the field content of the SM, there is no operator generating neutrino masses at the renormalizable level. The main effective operator in the SM leading to neutrino masses is dimension five

$$\frac{h_{ij}^\nu}{M} (\bar{l}_i^c \Phi) (l_j \Phi) \Rightarrow m_{ij}^\nu = h_{ij}^\nu \frac{v^2}{M} \quad . \quad (147)$$

Tiny values (of order 10^{-2} eV) neutrino masses ask for $10^{12} \text{ GeV} < M < 10^{15} \text{ GeV}$, see B. Kayser's lectures.

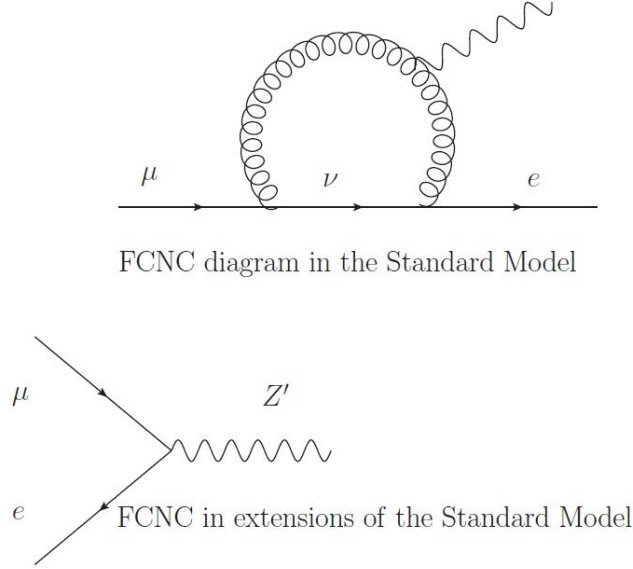


Fig. 12: Typical diagram in theories with a new neutral gauge boson, generating FCNC at tree-level.

Coming back to the quarks and charged leptons masses, we can define the *mass eigenstate basis* (as compared to the *weak eigenstate basis*) with the help of the 3×3 unitary transformations³

$$u_{L,R} = V_{L,R}^u u'_{L,R}, \quad d_{L,R} = V_{L,R}^d d'_{L,R}, \quad e_{L,R} = V_{L,R}^e e'_{L,R}, \quad (148)$$

such that

$$(V_L^u)^\dagger m^u V_R^u = \text{diag}(m_u, m_c, m_t), \quad \text{etc.} \quad (149)$$

In the mass basis, the neutral and the e.m. currents remain the same, whereas the hadronic charged current becomes

$$(J_W^{m,+})_{\text{quarks}} \rightarrow \frac{1}{\sqrt{2}} \bar{u}'_L \gamma^m V_{CKM} d'_L \equiv \frac{1}{\sqrt{2}} \bar{u}'_L \gamma^m \tilde{d}_L, \quad (150)$$

where $V_{CKM} = (V_L^u)^\dagger V_L^d$ is the (unitary) *CKM matrix* [32]. We also defined

$$\tilde{d}_L = V_{CKM} d'_L \leftrightarrow \begin{pmatrix} \tilde{d}_L \\ \tilde{s}_L \\ \tilde{b}_L \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d'_L \\ s'_L \\ b'_L \end{pmatrix}. \quad (151)$$

There are therefore *flavor changing transitions* in the SM: $s \rightarrow uW^-$, etc. Experimental measurements give a *hierarchical form* of V_{CKM} of the type (Wolfenstein parametrization)

$$\begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix}, \quad (152)$$

where $\lambda = \sin \theta_c \simeq .022$ is the Cabibbo angle. N. Cabibbo wrote first in 1962 the 2×2 version of the CKM matrix

$$\begin{pmatrix} \cos \theta_c & \sin \theta_c \\ -\sin \theta_c & \cos \theta_c \end{pmatrix}. \quad (153)$$

³These transformations are not innocent; there is a quantum anomaly that we will discuss later on.

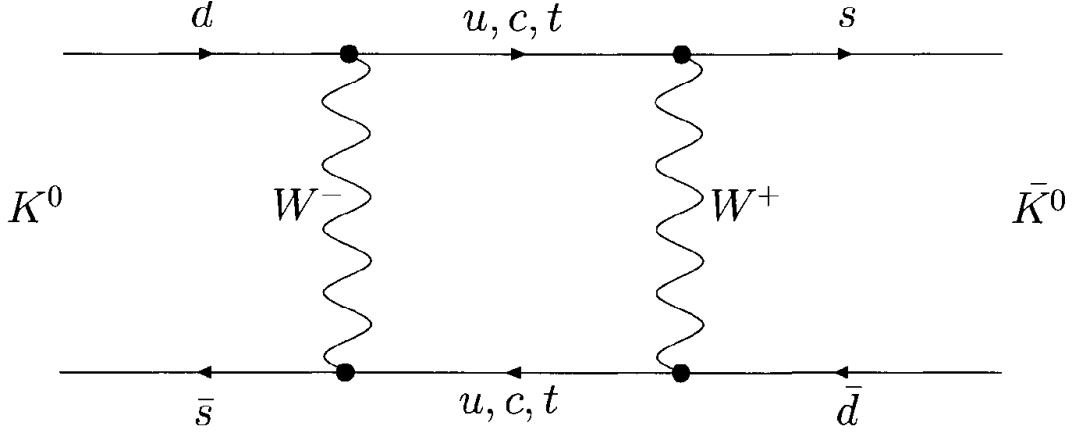


Fig. 13: K^0 - $\bar{K}^0$ mixing generated at loop level in the Standard Model, with quarks $u_i = u, c, t$ running in the loop.

It is simple to check that, after field redefinitions, V_{CKM} contain *three rotation angles* and one *CP violating phase*. Notice also that CP violation in the SM is *suppressed* by λ^3 in V_{CKM} . The *unitarity* of the CKM matrix

$$V_{ik}V_{jk}^* = \delta_{ij} \quad , \quad V_{ki}^*V_{kj} = \delta_{ij} \quad (154)$$

has various important consequences. One of them is *the GIM mechanism* (Glashow–Iliopoulos–Maiani, 1972), to which we now turn.

6.5 The GIM mechanism

The FCNC (flavor changing neutral currents) effects were measured experimentally to be small. This was puzzling in the 1970's, but it was explained in the SM by GIM [26]. Consider for example the $K^0 - \bar{K}^0$ mixing, which can arise at the loop-level. In the limit of equal or vanishing quark masses, the amplitude vanishes due to the unitarity of V_{CKM} :

$$A_{K^0\bar{K}^0} \sim \frac{g^4}{M_W^4} (\sum_i V_{id}^* V_{is}) (\sum_j V_{js} V_{jd}) = 0 . \quad (155)$$

By turning on the quark masses, the main contribution turns out to be proportional to $(m_c^2 - m_u^2)/M_W^4$ and is in excellent agreement with the experimental result.

Historical Remark: In 1972, only the u, d and s quarks were known. The GIM mechanism is considered to be the first proof of the existence of the charm quark.

Homework: Write down explicitly the diagrams for the $K^0 - \bar{K}^0$ mixing in the two generation case, with u and c quarks in the loop.

The unitarity relation

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \quad (156)$$

can be represented geometrically as a *triangle* in a plane, named the *unitarity triangle*. It is customary to rescale the length of one side, i.e. $|V_{cd}V_{cb}^*|$ (well-known), to 1 and to align it along the real axis. The angles are defined as

$$\beta = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right), \quad \gamma = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right), \quad (157)$$

and the lengths are

$$R_t = \left|\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right|, \quad R_u = \left|\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right|. \quad (158)$$

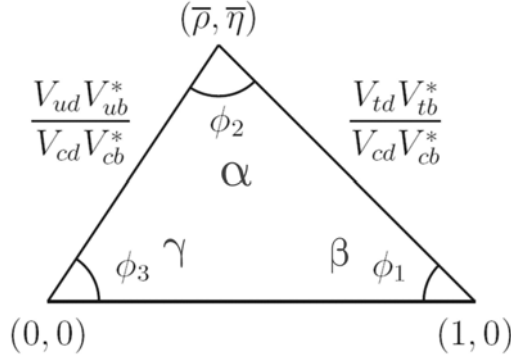


Fig. 14: The unitarity triangle. From Ref. [34].

On the other hand, quarks, leptons masses and the CKM matrix feature strong hierarchies. For example, from neutrino masses to the top mass there are 10^{11} orders of magnitude $m_\nu \sim 10^{-2} \text{ eV} \ll m_e = 0.511 \text{ MeV} \ll m_t \sim 172 \text{ GeV}$.

There is *no hint for a solution of this flavor puzzle* in the SM, since the Yukawa couplings are free-parameters and therefore are not predicted. We are clearly *missing something*: maybe an additional global or gauge symmetry [33] or maybe this comes from an extra dimensional localization or environmental selection.

6.6 The custodial symmetry

The tree-level relation $\rho = M_W^2 / (M_Z^2 \cos^2 \theta_w) = 1$ can be understood as the result of an (*approximate*) *symmetry*, called *custodial symmetry*, proven by Sikivie, Susskind, Voloshin and Zakharov [35].

Theorem: *In any theory of electroweak interactions which conserves the electric charge and has an approximate global $SU(2)$ symmetry under which A_m^a transform as a triplet, $\rho = 1$ at tree-level.*

Here approximate means in the limit of zero hypercharge coupling $g' = 0$ and in the absence of the Yukawa couplings.

Proof: Under this assumption, the gauge boson mass matrix is of the form

$$\begin{pmatrix} M^2 & 0 & 0 & 0 \\ 0 & M^2 & 0 & 0 \\ 0 & 0 & M^2 & m_1^2 \\ 0 & 0 & m_1^2 & m_2^2 \end{pmatrix}. \quad (159)$$

The masslessness of the photon implies $M^2 m_2^2 - m_1^4 = 0$. The resulting $W_3 - A$ mass matrix, written in terms of the W and Z masses, is then of the form:

$$\begin{pmatrix} M_W^2 & \pm M_W \sqrt{M_Z^2 - M_W^2} \\ \pm M_W \sqrt{M_Z^2 - M_W^2} & M_Z^2 - M_W^2 \end{pmatrix}. \quad (160)$$

It is then easy to check that $M_W = \cos \theta_w M_Z$.

On the other hand, in the SM the Higgs potential $V(\Phi^\dagger \Phi)$ is invariant under an $SO(4)$ global symmetry. Indeed, let us write explicitly the four real components of the SM Higgs doublet

$$\Phi = \begin{pmatrix} \Phi_1 + i\Phi_2 \\ \Phi_3 + i\Phi_4 \end{pmatrix}, \text{ then } \Phi^\dagger \Phi = \sum_{i=1}^4 \Phi_i^2. \quad (161)$$

It is then transparent that the Higgs potential and kinetic term have an $SO(4) = SU(2)_L \times SU(2)_R$ symmetry. The Higgs vev, $\langle \Phi \rangle = (0, v/\sqrt{2})$, breaks $SO(4) \rightarrow SO(3) = SU(2)_D$, which corresponds precisely to

the custodial symmetry. From these considerations, it is clear that not any Higgs representations preserve the custodial symmetry. What happens for other Higgs representations? It can be shown that, considering Higgs representations in weak isospin representations of isospin I , the rho parameter is given by

$$\rho = \frac{1}{2} \frac{\sum_i (I_i(I_i + 1) - I_{3i}^2) |\langle 0 | \phi_i | 0 \rangle|^2}{\sum_i I_{3i}^2 |\langle 0 | \phi_i | 0 \rangle|^2} . \quad (162)$$

It is then easy to check that for an arbitrary number of singlet and higgs doublets, $\rho = 1$. On the other hand, for Higgs triplets for example, the higgs vev generate the breaking $SO(3) \rightarrow SO(2)$. In this case there is no custodial symmetry and $\rho \neq 1$.

Strong interactions preserve electric charge and strong isospin. A natural choice for the custodial symmetry is therefore the strong isospin, which then guarantees that $\rho = 1$ to all order in the strong interactions.

A useful parametrization for estimating the violation of the custodial symmetry is:

$$\mathcal{H} = (i\tau_2 \Phi^* \quad \Phi) = \begin{pmatrix} \Phi_0^* & \Phi_+ \\ -\Phi_+^* & \Phi_0 \end{pmatrix}, \quad \Phi^\dagger \Phi = \frac{1}{2} \text{Tr} \mathcal{H}^\dagger \mathcal{H} . \quad (163)$$

The potential $V(\Phi^\dagger \Phi)$ is invariant under $\mathcal{H} \rightarrow U_L \mathcal{H} U_R^\dagger$, with $U_{L,R}$ 2×2 unitary matrices implementing $SU(2)_L \times SU(2)_R$ transformations. The electroweak symmetry breaking pattern is then

$$\langle \mathcal{H} \rangle = \frac{v}{\sqrt{2}} I_{2 \times 2} \quad \text{breaks} \quad SU(2)_L \times SU(2)_R \rightarrow SU(2)_D . \quad (164)$$

As anticipated, the hypercharge gauge interactions $U(1)_Y$ and Yukawa couplings *break* the custodial symmetry. However the particular coupling

$$\mathcal{L}_{\text{Yuk}} = h \begin{pmatrix} \bar{t}_L & \bar{b}_L \end{pmatrix} \mathcal{H} \begin{pmatrix} t_R \\ b_R \end{pmatrix} \quad (165)$$

is invariant under $SU(2)_D$. This corresponds to the limit of equal masses in the quark doublet $h_t = h_b$. On the other hand, W and Z boson masses have quantum corrections that lead to calculable deviations from $\rho = 1$. A one-loop computation in the SM gives

$$\delta\rho = \frac{3g^2(m_t^2 - m_b^2)}{64\pi^2 M_W^2} - \frac{3g'^2}{32\pi^2} \ln \frac{m_H}{M_Z} + \dots \quad (166)$$

where $\dots$ are *subleading* contributions from the SM or from eventual new physics contributions (see the lectures by B. Dobrescu) that have to be smaller than 10^{-3} in order to fit the experimental data [36].

7 Quantum corrections and renormalization

Quantum corrections through loops are subtle to incorporate, due to UV divergences appearing for large momenta of virtual particles running in the loops. Dealing with these divergences is crucial in order to extract physical results. This led to the program of renormalization, which was brilliantly confirmed by various precision measurements, in particular at the LEP collider. The proof of renormalization of the Standard Model led to the 1999 Nobel prize of G. 't Hooft and M. Veltman [37].

7.1 UV divergences and regularization

Perturbation theory in Quantum Field Theory is plagued with *UV divergences*. We have to keep an UV cutoff Λ (which can be implemented in various different ways) in computing physical quantities. There are three cases that arise:

- *Super-renormalizable theories*: In this case only a finite number of Feynman diagrams diverge. Beyond a sufficiently large number of loops, all Green functions are finite.
- *Renormalizable theories*: a finite number of amplitudes/Green functions diverge, with a number of external legs below a maximal value (which is for example four for the ϕ^4 theory, three for QED and four for Yang–Mills theories). For these amplitudes, the UV divergences arise at all orders in perturbation theory.
- *Non-renormalizable theories*: All amplitudes, with an arbitrary number of legs are UV divergent at a certain order in perturbation theory.

In renormalizable and super-renormalizable theories, UV divergences can be *absorbed* into *rescaling of fields* and *redefinitions* of the various couplings and masses. Taking the couplings/masses from experimental data and "hiding" the UV cutoff in their redefinitions, we obtain physical quantities free of UV divergences. In this case, *the theory is predictive at any energy scale*. In non-renormalizable theories, we need an *infinite number* of couplings and masses in order to absorb the UV divergences. We would need an infinite amount of experimental data to determine all these couplings. Therefore, at high-energies $E > \Lambda$ *the theory loses its predictive power*. However, at low-energy the theory is *perfectly predictive*. The typical example of this type is the General Relativity.

7.2 Relevant, marginal and irrelevant couplings

Consider a scalar theory of the form

$$S_\Lambda = \int d^4x \left(\frac{1}{2}(\partial\phi)^2 + \frac{m^2\phi^2}{2} + \sum_n \lambda_n \phi^n \right), \quad (167)$$

where S_Λ is the euclidian action defined with a cutoff Λ . The couplings λ_n have (classical) mass dimensions $[\lambda_n] = 4 - n$. Let us consider the theory with two different maximal euclidian cutoff momenta:

- $0 < p < \Lambda$
- $0 < p < \Lambda' = \varepsilon \Lambda$, where $\varepsilon < 1$.

In case ii) the theory has therefore a *lower cutoff* and it is interpreted as a theory where the high-momenta of theory i) were *integrated out*. The theory i) has the action (167). In the theory ii) the cutoff can be redefined to be the same as in i) with the help of a *scale transformation*

$$x' = \varepsilon x, \quad p' = \varepsilon^{-1} p, \quad \phi' = \varepsilon^{-1} \phi. \quad (168)$$

In terms of the rescaled field and coordinates, the action of theory ii) become

$$S_{\Lambda'} = \int d^4x' \left(\frac{1}{2}(\partial'\phi')^2 + \frac{m'^2(\phi')^2}{2} + \sum_n \lambda'_n (\phi')^n \right), \quad (169)$$

where

$$m'^2 = \frac{1}{\varepsilon^2} m^2, \quad \lambda'_n = \varepsilon^{n-4} \lambda_n. \quad (170)$$

Notice that the new mass and couplings scale with their classical dimension. We see therefore that the mass and couplings with positive dimension *grow* in the IR, whereas couplings with negative dimension *decrease* in the IR. It is said that

$$\begin{aligned} [\lambda_n] > 0 &\Rightarrow \text{relevant couplings,} \\ [\lambda_n] = 0 &\Rightarrow \text{marginal couplings,} \\ [\lambda_n] < 0 &\Rightarrow \text{irrelevant couplings.} \end{aligned} \quad (171)$$

This point of view on renormalization was introduced by K. Wilson and is summarized, for example, in Ref. [38].

7.3 (Non)renormalizability and couplings dimensions

There is a straight connection between renormalizability and the three type of couplings previously defined:

- relevant couplings $\Rightarrow$ super-renormalizability.
- marginal couplings $\Rightarrow$ renormalizability.
- irrelevant couplings $\Rightarrow$ non-renormalizability.

It is easy to argue for this by *dimensional arguments*. Let us consider some simple examples, going back in Minkowski space:

1. Relevant coupling

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 - \frac{m^2\phi^2}{2} - \lambda_3\phi^3. \quad (172)$$

The coupling has dimension $[\lambda_3] = +1$, so it is relevant. At one-loop, the UV divergent terms lead to new terms in the Lagrangian (**homework:**)

$$\delta\mathcal{L}_1 \sim \lambda_3\Lambda^2\phi + \lambda_3^2\phi^2\ln\Lambda, \quad (173)$$

which are both of super-renormalizable type. The first leads to a scalar tadpole, whereas the second leads to a mass renormalization. At two loops, the only UV divergences are a cosmological constant and a scalar tadpole. At three loops, there is only a log UV divergence in the cosmological constant. No UV divergences exist at higher loops. *Dimensional argument:* By dimensional analysis, the highest UV divergent term in the coupling is the three-loop vacuum energy

$$\lambda_3^4 \ln\Lambda. \quad (174)$$

Higher loops have higher powers in λ_3 and cannot contribute to the UV divergent terms in the effective Lagrangian. **Observation:** $1/m^2$ terms are IR, not UV contributions, so they cannot appear in UV divergent terms.

2. Irrelevant coupling

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 - \frac{m^2\phi^2}{2} - \lambda_6\phi^6. \quad (175)$$

The coupling has dimension $[\lambda_6] = -2$, so it is irrelevant. At one-loop, the UV divergent terms in the eight-point amplitude lead to (homework:)

$$\Gamma_{1\text{-loop}}^{(8)}(p_i) \sim c \lambda_6^2 \ln\Lambda + \dots. \quad (176)$$

To cancel this divergence, one has to add a new coupling to the original action

$$\delta\mathcal{L}_1 \sim \lambda_8 \phi^8, \quad (177)$$

and to *adjust the coupling* λ_8 such that

$$\lambda_8 + c \lambda_6^2 \ln\Lambda = \text{finite}. \quad (178)$$

At two-loops, we get new *new UV divergences*, like the one in the six-point amplitude, proportional to

$$\Gamma_{2\text{-loops}}^{(6)}(p_i) \sim c' (p_i p_j) \lambda_6^2 \ln\Lambda, \quad (179)$$

which can be canceled by adding *another coupling*

$$\delta\mathcal{L}_2 \sim \lambda'_8 \phi^4 (\partial\phi)^2, \quad (180)$$

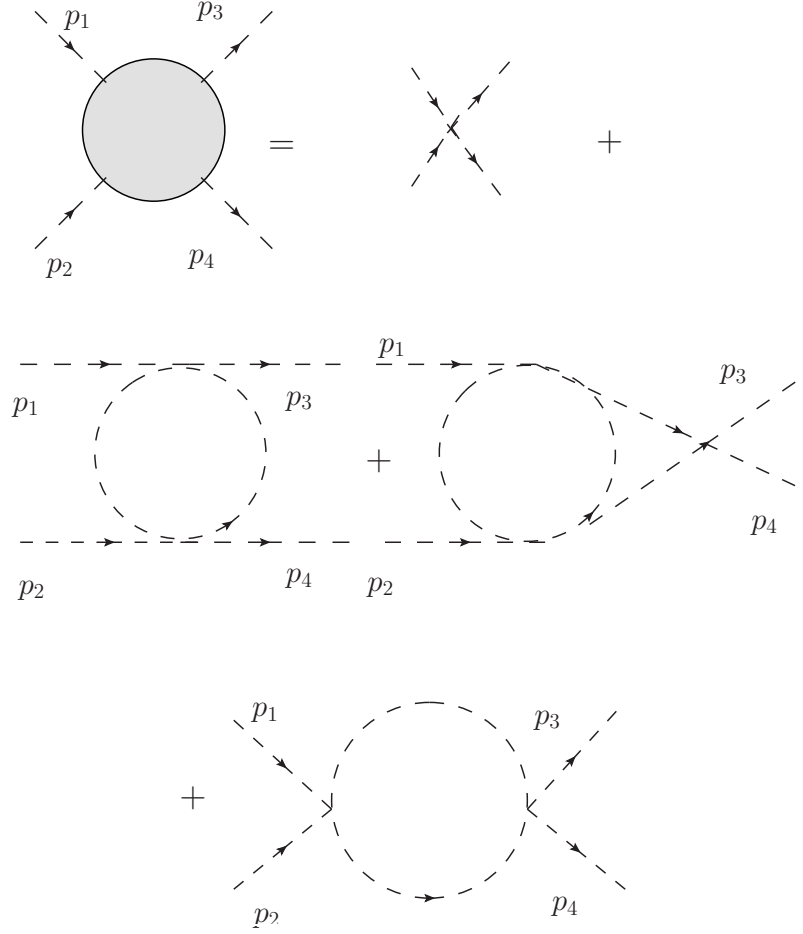


Fig. 15: Diagrams contributing to the renormalization of the self-coupling λ in ϕ^4 theory.

such that

$$\lambda'_g + c' \lambda_g^2 \ln \Lambda = \text{finite} . \quad (181)$$

The UV divergences *proliferate* at higher loop orders, generating an infinite tower of operators of higher and higher dimension. *Dimensional argument:* Terms of the type $\lambda_6^n \phi^{4+2n} \ln \Lambda$, $\lambda_6^n (\partial \phi)^2 \phi^{2n} \ln \Lambda$ have the correct dimension to be generated for any n . Predictivity at high-energy is *lost*. Let us however define $\lambda_6 \sim 1/M^2$. Then: In the IR, $E < M$, the effect of non-renormalizable operators on physical quantities is proportional to some positive power of E/M and/or m/M , so their effects is negligible. Effective theories with cutoff Λ (example General relativity, $\Lambda = M_P$) are therefore predictive at energies $E \ll \Lambda$.

Another viewpoint on this problem is the following: for $\mathcal{L}_{\text{int}} = \sum_n \lambda_n \phi^n$, the leading cross-section for $2 \rightarrow 2$ particle scattering is

$$\sigma = \sum_n c_n \lambda_n^2 E^{2n-10} \sim \frac{1}{E^2} \sum_n c_n \left(\frac{E}{M} \right)^{2n} , \quad (182)$$

for $\lambda_n \sim 1/M^{n-4}$. Therefore the *predictive power is lost* for $E \geq M$.

7.4 Coupling constant renormalization for ϕ^4 theory

Consider the ϕ^4 theory of Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 - \frac{m_0^2}{2}\phi^2 - \frac{\lambda_0}{4!}\phi^4. \quad (183)$$

Let us compute the four-point function at one-loop. By using the Feynman rules for the ϕ^4 theory, we find, according to the figure in the next page

$$\begin{aligned} \Gamma^{(4)}(k_1 k_2 k_3 k_4) &= -i\lambda_0 + \frac{(-i\lambda_0)^2}{2} \int \frac{d^4 p}{(2\pi)^4} \frac{i}{p^2 - m_0^2} \frac{i}{(p - k_1 - k_2)^2 - m_0^2} \\ &\quad + \text{two crossing terms} . \end{aligned} \quad (184)$$

After the Wick rotation to euclidian momenta, the result is given by

$$\begin{aligned} \Gamma^{(4)}(k_1 k_2 k_3 k_4) &= -i\lambda_0 + \frac{i\lambda_0^2}{2} \int \frac{d^4 p}{(2\pi)^4} \frac{1}{p^2 + m_0^2} \frac{1}{(p - k_1 - k_2)^2 + m_0^2} \\ &\quad + \text{two crossing terms} . \end{aligned} \quad (185)$$

The integral is log divergent in the UV. There are various ways to "renormalize" the integral. Here is a simple way. Define

$$V(s) \equiv \int \frac{d^4 p}{(2\pi)^4} \frac{1}{p^2 + m_0^2} \frac{1}{(p - k_1 - k_2)^2 + m_0^2} = \int_{p^2 \geq \mu^2}^{\Lambda} \frac{d^4 p}{(2\pi)^4} \frac{1}{p^4} + \text{finite} , \quad (186)$$

where the energy scale μ is arbitrary. We find (**homework:**)

$$\begin{aligned} \Gamma^{(4)}(k_1 k_2 k_3 k_4) &= -i\lambda_0 + \frac{i\lambda_0^2}{2} [V(s) + V(t) + V(u)] \\ &= -i\lambda_0 + \frac{3i\lambda_0^2}{16\pi^2} \ln \frac{\Lambda}{\mu} + \text{finite} = -i\lambda(\mu) + \text{finite} . \end{aligned} \quad (187)$$

What is the *physical interpretation* of this manipulation? We can separate the answer into two separate steps:

1. λ_0 is *not a physical parameter*. It can be chosen to depend on Λ such that

$$\lambda(\mu) = \lambda_0(\Lambda) - \frac{3\lambda_0^2}{16\pi^2} \ln \frac{\Lambda}{\mu} \quad (188)$$

is independent of Λ .

2. Any value of μ leads to the same physical result. We can find a differential equation for λ by using the fact that λ_0 is independent of μ . We obtain

$$\frac{d\lambda}{d\ln\mu} = \frac{3\lambda^2}{16\pi^2} = \beta(\lambda) , \quad (189)$$

which is called the *renormalization group equation (RGE)* of λ at one-loop, with $\beta(\lambda) = \frac{3\lambda^2}{16\pi^2}$ being the one-loop RG *beta function* coefficient. The solution of (189) is (**homework:**)

$$\lambda(\mu) = \frac{\lambda(\mu_0)}{1 - \frac{3\lambda(\mu_0)}{16\pi^2} \ln \frac{\mu}{\mu_0}} . \quad (190)$$

Notice that there is an equivalent prescription: to add a local "*counterterm*" to the Lagrangian

$$\mathcal{L} + \delta\mathcal{L} = \mathcal{L}_0 , \quad (191)$$

which cancels the UV divergence. The counterterms are treated as interactions in perturbation theory. The two points of view lead to identical results.

In renormalizable theories, a *finite number* of counterterms are needed in order to render the theory UV finite. For the same purpose, in non-renormalizable theories we need an *infinite number* of counterterms.

7.5 QED and the running of fine structure constant

We use here the *counterterm* method for the renormalization of QED. In this case, the initial Lagrangian, the counterterms and their sum is

$$\begin{aligned}\mathcal{L} &= -\frac{1}{4}F_{\mu\nu}^2 + \bar{\Psi}(i\gamma^\mu\partial_\mu - q\gamma^\mu A_\mu - M)\Psi, \\ \delta\mathcal{L} &= -\frac{1}{4}(Z_3 - 1)F_{\mu\nu}^2 + (Z_2 - 1)\bar{\Psi}i\gamma^\mu\partial_\mu\Psi - (Z_1 - 1)q\bar{\Psi}\gamma^\mu A_\mu\Psi - (Z_M - 1)M\bar{\Psi}\Psi, \\ \mathcal{L}_0 &= \mathcal{L} + \delta\mathcal{L} = -\frac{1}{4}(F_{\mu\nu}^0)^2 + \bar{\Psi}_0(i\gamma^\mu\partial_\mu - q_0\gamma^\mu A_\mu^0 - M_0)\Psi_0.\end{aligned}\tag{192}$$

The relations between the *bare and renormalized* quantities are then

$$\begin{aligned}A_\mu^0 &= Z_3^{1/2} A_\mu, \quad \Psi_0 = Z_2^{1/2} \Psi, \\ M_0 &= \frac{Z_M}{Z_2} M, \quad q_0 = \frac{Z_1}{Z_2 Z_3^{1/2}} q,\end{aligned}\tag{193}$$

where Z_1 comes from the one-loop vertex correction, Z_2 (Z_3) is the fermionic (photon) wave function renormalization, whereas Z_M is the mass renormalization. In QED it can be shown that $Z_1 = Z_2$, the so-called Ward identity. Then charge renormalization in QED comes only from vacuum polarization $q_0 = Z_3^{-1/2} q$. The *RG running* can be found from

$$\mu \frac{\partial}{\partial \mu} q_0 = 0 \quad \Rightarrow \quad \beta(q) = \mu \frac{\partial q}{\partial \mu} = q \frac{\partial \ln Z_3^{1/2}}{\partial \ln \mu}.\tag{194}$$

By an explicit computation in QED with just the electron in the loop and by defining the fine-structure constant $\alpha = q^2/(4\pi)$, we find

$$Z_3 = 1 - \frac{\alpha}{3\pi} \ln \frac{\Lambda}{\mu} + \text{finite},\tag{195}$$

where μ is an arbitrary, renormalization scale. We then find

$$\beta(q) = \frac{q^3}{24\pi^2} \quad \Rightarrow \quad \frac{1}{\alpha(Q)} = \frac{1}{\alpha(\mu)} - \frac{1}{3\pi} \ln \frac{Q}{\mu}.\tag{196}$$

We found therefore that the fine structure coupling *increases with energy* ! This can be intuitively interpreted due to the screening of the electric charge by electron-positron pairs from the quantum vacuum (see Figure 16).

The situation for the strong coupling α_3 is different due to the non-abelian nature of the interaction. The result is an *anti-screening* due to gluon self-interactions [40].

There is a tantalizing hint of unification of gauge couplings at high-energy, as seen from figure 17, that could point towards a unified gauge structure at high-energy [41]. Running couplings and renormalization are important everywhere in the SM and its applications. For example:

- In any process the couplings have to be evaluated at the relevant energy scale. Examples:
 - In $\pi^0 \rightarrow \gamma\gamma$, the fine-structure constant has to be evaluated at the pion mass $\alpha(m_\pi)$.
 - Identification of relevant momenta and RGE of operators in QCD is crucial in order to extrapolate perturbative quantities down in energy via the renormalization group.

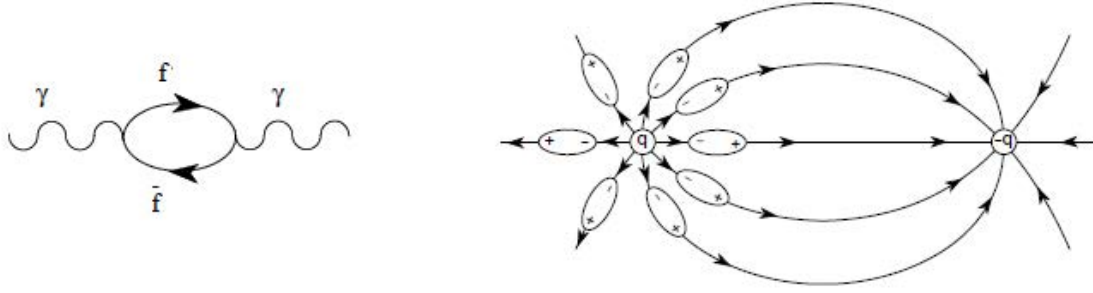


Fig. 16: Screening of electric charges by vacuum polarization provides an intuitive picture of the "running" of electric charge. Figure taken from Ref. [39].

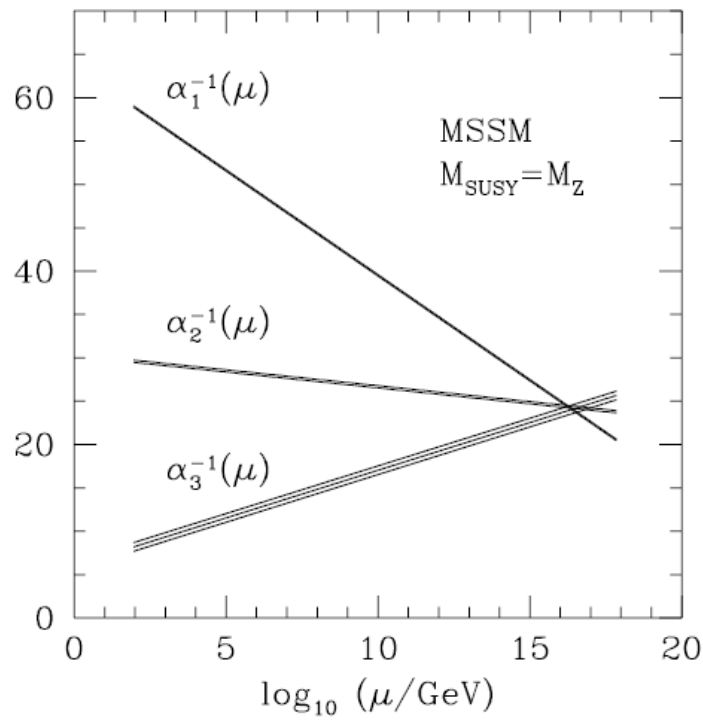


Fig. 17: Extrapolation of gauge couplings in the (minimal supersymmetric extension) of the Standard Model [42]: hint of unification of couplings at high energy?

- In the study of electroweak baryogenesis, for example the scalar potential

$$V(\Phi) \simeq -\mu^2(\langle\Phi\rangle)|\Phi|^2 + \lambda(\langle\Phi\rangle)|\Phi|^4 \quad (197)$$

is to be evaluated at the minimum at the scalar potential.

8 Global and gauge anomalies

Symmetries of the classical action can have *anomalies* at the quantum level, generated by one-loop *triangle diagrams* [27], see Fig. 18. There are two different cases to consider:

- anomalies in the conservation of a global symmetry current,

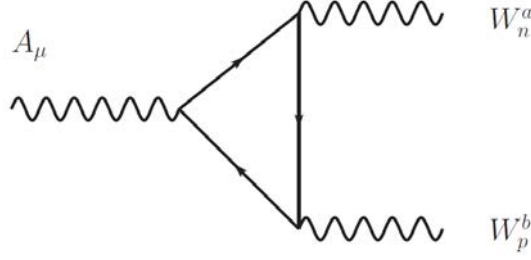


Fig. 18: Adler–Bell–Jackiw triangle anomalies.

- anomalies for gauged symmetries.

8.1 Global anomalies

For global symmetries, this does not create consistency problems and they actually play an important role in QCD and the electromagnetic decay of the π^0 pion. Consider to start with a Dirac fermion coupled to a $U(1)$ gauge field

$$\mathcal{L} = \bar{\Psi} i \gamma^\mu D_\mu \Psi - M \bar{\Psi} \Psi . \quad (198)$$

In the massless limit $M \rightarrow 0$, the model has a vector and an axial symmetry $U(1)_V \times U(1)_A$. The corresponding Noether currents

$$J_m = \bar{\Psi} \gamma_m \Psi \quad , \quad J_m^5 = \bar{\Psi} \gamma_m \gamma_5 \Psi \quad (199)$$

satisfy

$$\partial^m J_m = 0 \quad , \quad \partial^m J_m^5 = 2iM \bar{\Psi} \gamma_5 \Psi - \frac{g^2}{16\pi^2} \epsilon^{mnpq} F_{mn} F_{pq} . \quad (200)$$

The last term is the *quantum anomaly*. Even if both currents are both classically conserved for $M = 0$, there is *no regularization* preserving both the vector and the axial conservation. If $U(1)_V$ is a gauge symmetry (the electromagnetism), we have therefore to choose a regularization preserving the vector current conservation. Then one is forced to accept an anomaly in the axial current. This explains actually why the η' meson is not a pseudo-Goldstone for the dynamical chiral symmetry breaking $U(2)_L \times U(2)_R = SU(2)_L \times SU(2)_R \times U(1)_B \times U(1)_A \Rightarrow SU(2)_V \times U(1)_B$ in QCD. Indeed, in this case the $U(1)_A$ axial current has the QCD anomaly

$$\begin{aligned} J_m^{U(1)_A} &= \bar{u} \gamma_m \gamma_5 u + \bar{d} \gamma_m \gamma_5 d , \\ \partial^m J_m^{U(1)_A} &= 2i(m_u \bar{u} \gamma_5 u + m_d \bar{d} \gamma_5 d) - \frac{g_3^2}{16\pi^2} \epsilon^{mnpq} F_{mn}^A F_{pq}^A , \end{aligned} \quad (201)$$

where F^A is the gluon field strength. Due to the explicit breaking and the nonperturbative instanton effects, the η' gets a mass larger than the other pions, which are the pseudo-goldstones of the axial $SU(2)_A$ symmetry. Another manifestation of the axial anomaly is the decay $\pi^0 \rightarrow \gamma\gamma$. Let us define the $SU(2)$ currents

$$J_m^a = \bar{q} \gamma_m \tau^a q \quad , \quad J_m^{5,a} = \bar{q} \gamma_m \gamma_5 \tau^a q . \quad (202)$$

The fact that the pions are *Goldstone bosons* implies

$$\langle 0 | J_m^{5,a}(x) | \pi^b(p) \rangle = -i p_m f_\pi \delta^{ab} e^{-ipx} . \quad (203)$$

Axial isospin currents have no QCD anomalies, but $J_m^{5,a}$ has an *anomaly* from the electromagnetic coupling

$$\partial^m J_m^{5,3} = -\frac{1}{32\pi^2} \epsilon^{mnpq} F_{mn} F_{pq} \text{tr}(Q^2 \tau_3) = -\frac{N_c e^2}{48\pi^2} \epsilon^{mnpq} F_{mn} F_{pq} , \quad (204)$$

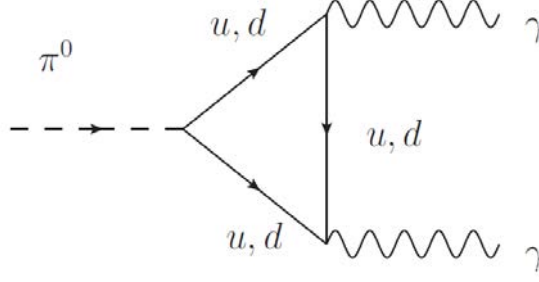


Fig. 19: The electromagnetic pion desintegration $\pi^0 \rightarrow \gamma\gamma$ is related to the axial $U(1)_A$ anomaly.

where Q describes the quark electric charges $Q_u = 2e/3$, $Q_d = -e/3$ and $N_c = 3$ is the number of quark colors.

By using Eqs. (203) and (204) and using that under an axial $SU(2)$ with quarks transforming as $\delta q = i\gamma_5 \tau_3 q$ with $q = (u, d)$, the pion transforms like a Goldstone boson $\delta\pi^0 = \alpha f_\pi$, we obtain that the effect of the anomaly is to generate an effective pion–photon–photon coupling

$$\mathcal{L}_{\text{eff}} = \frac{\pi^0}{f_\pi} \partial^m J_m^{5,3} = - \frac{N_c e^2}{48\pi^2 f_\pi} \pi^0 \epsilon^{mnpq} F_{mn} F_{pq}. \quad (205)$$

Using this effective coupling, the following result is obtained $\Gamma(\pi^0 \rightarrow \gamma\gamma) = \frac{\alpha^2}{64\pi^3} \frac{m_\pi^3}{f_\pi^2}$, which is in excellent agreement with the experimental branching ratio of the pion decay into two photons $\Gamma(\pi^0 \rightarrow \gamma\gamma) = (1.19 \pm 0.08) \times 10^{16} \text{s}^{-1}$.

On the other hand, there is no symmetry principle forbidding the term in the QCD Lagrangian

$$\mathcal{L}_\theta = \theta \frac{g^2}{32\pi^2} \epsilon^{\mu\nu\rho\sigma} \text{Tr}(G_{\mu\nu} G_{\rho\sigma}), \quad (206)$$

where θ is a real (angular) parameter. Even if this can be shown to be a total derivative, instanton configurations in QCD makes this term to have nontrivial consequences. Actually, the θ parameter has another contribution from the unitary redefinitions (148) that we were forced to do in order to diagonalize the quark mass matrices. Indeed, the $U(1)_A$ part of these transformations is anomalous, leading to a change in the theta parameter

$$\theta \rightarrow \theta - \frac{1}{2} \arg \det m^q. \quad (207)$$

The theta parameter violates CP and the gluonic term generates a neutron dipole moment of order $d_n \sim |\theta| \text{ em}_\pi^2 / m_N^3 \sim 10^{-16} |\theta| \text{ ecm}$, in conflict with the experimental data unless $\theta < 10^{-10}$. This leads to the so-called *strong CP problem*. The problem would be absent if the up-quark mass would be zero, since in this case the theta parameter could be shifted to zero by an up-quark chiral redefinition. The masslessness of the up-quark is however excluded by now. One of possible solutions to the strong CP problem is *the axion*, a [28]. If

- there is a new $U(1)_{PQ}$, spontaneously broken global symmetry, with the corresponding pseudo-Goldstone boson a , and the symmetry breaking scale f ,
- which has triangle anomalies with the QCD gauge group $U(1)_{PQ} \times SU(3)_c^2$,

then *the anomaly* generates *new couplings* in the Lagrangian which shift the θ parameter

$$\frac{g^2}{32\pi^2} \xi \frac{a(x)}{f} \epsilon^{\mu\nu\rho\sigma} \text{Tr}(G_{\mu\nu} G_{\rho\sigma}) \Rightarrow \theta_{\text{eff}} = \theta + \xi \frac{a}{f}, \quad (208)$$

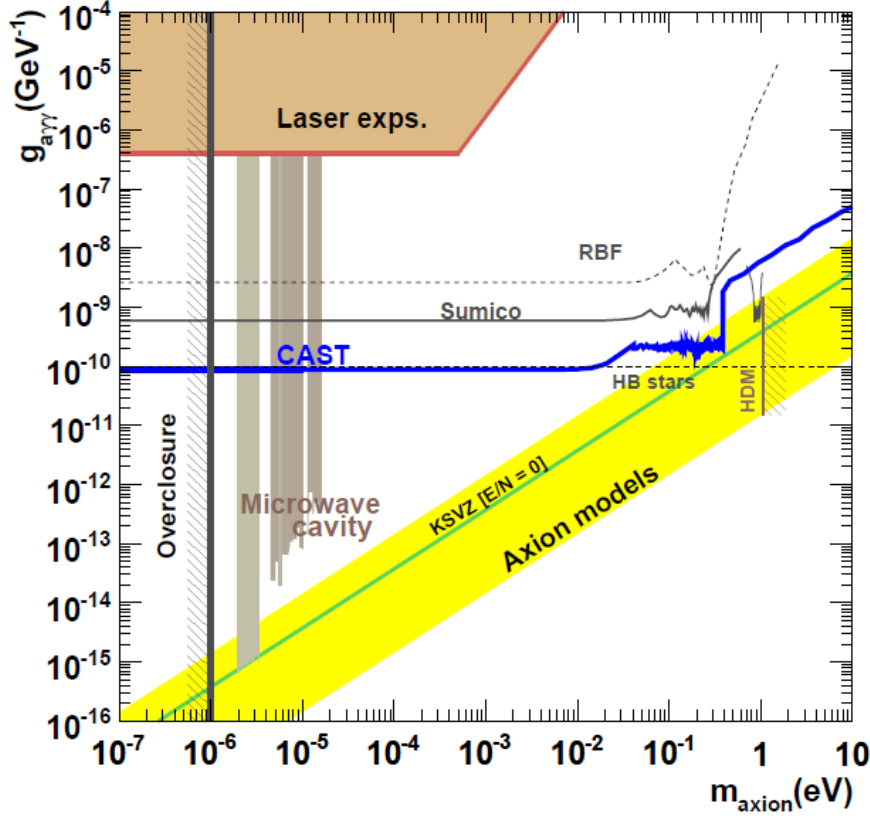


Fig. 20: Axion searches and constraints in the plane $g_{a\gamma}$ (the axion-photon coupling) versus the axion mass. See for example Ref. [43] for more details.

where ξ is a model-dependent parameter parametrizing the strength of the axion couplings to matter. Then *non-perturbative QCD instanton effects* generate an axion potential of the type

$$V(a) \sim \frac{g^2}{32\pi^2} \Lambda_{QCD}^4 \left[1 - \cos \left(\xi \frac{a(x)}{f} + \theta \right) \right]. \quad (209)$$

The minimum of the scalar potential is then at $\theta_{\text{eff}} = 0$ and the axion mass is

$$m_a \sim \xi \frac{g}{4\sqrt{2}\pi} \frac{\Lambda_{QCD}^2}{f}. \quad (210)$$

Axions were intensively searched since the 80's, see Fig. 20 for recent constraints in the plane $(g_{a\gamma}, m_a)$, where $g_{a\gamma}$ is the axion-photon coupling. Axions are also present in *most SUSY and string extensions* of the SM.

Let us finish this presentation with a comment. The axial anomaly is actually a total derivative:

$$\epsilon^{\mu\nu\rho\sigma} \text{Tr}(F_{\mu\nu} F_{\rho\sigma}) = \partial^\mu K_\mu, \quad (211)$$

where

$$K_\mu = 2\epsilon_{\mu\nu\alpha\beta} \left(A^{\nu\alpha} \partial^\alpha A^{\beta\mu} + \frac{1}{3} f^{abc} A^{\nu a} A^{\alpha b} A^{\beta c} \right). \quad (212)$$

Despite this fact, classical configurations generate effects like *the theta angle in QCD and B and L numbers are non-conserved*. Indeed, the *baryonic current* for example has an $U(1)_B \times SU(2)_L^2$ anomaly

$$J_\mu^B = \frac{1}{3} \sum_i \bar{q}_i \gamma_\mu q_i, \quad \partial^\mu J_\mu \sim \frac{g^2}{16\pi^2} \epsilon^{\mu\nu\rho\sigma} \text{Tr}(F_{\mu\nu} F_{\rho\sigma}),$$

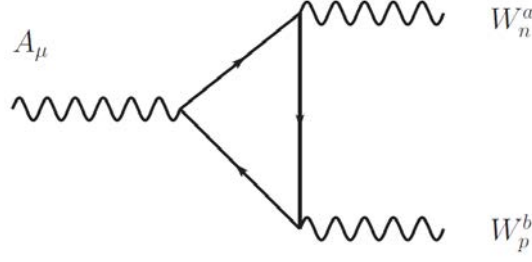


Fig. 21: One-loop gauge anomalies, if present, render the theory *inconsistent* at the quantum level.

$$\Delta B = \int d^4 \partial^\mu J_\mu \sim \int d^4 \epsilon^{\mu\nu\rho\sigma} \text{Tr}(F_{\mu\nu} F_{\rho\sigma}) \sim \oint d\Sigma^\mu dK_\mu \quad (213)$$

which is different from zero for classical gauge field configurations *vanishing slowly* at infinity. The violation of baryonic symmetry of this type is important for generating the observed baryon asymmetry in our universe [29].

8.2 Gauge anomalies

For *gauge symmetries* on the other hand, anomalies, if present as in Fig. 25, generate inconsistencies [30]. Indeed, they would violate *gauge invariance* of the theory since the gauge variation of the Lagrangian from the Noether theorem (5) is

$$\delta \mathcal{L} \sim \alpha_A \partial^\mu J_\mu^A. \quad (214)$$

The corresponding currents are of chiral type

$$J_\mu^A = \bar{\Psi} \gamma_\mu \gamma_5 T^A \Psi = \bar{\Psi}_R \gamma_\mu T^A \Psi_R - \bar{\Psi}_L \gamma_\mu T^A \Psi_L, \quad (215)$$

and their divergences are proportional to

$$\partial^\mu J_\mu^A = - \frac{g_B g_C}{32\pi^2} A^{ABC} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu}^B F_{\rho\sigma}^C. \quad (216)$$

The anomaly coefficients that have to vanish are then

$$A^{ABC} = \text{tr}(\{T^A, T^B\}T^C)_L - \text{tr}(\{T^A, T^B\}T^C)_R = 0, \quad (217)$$

where the trace is taken over *all the fermions* in the theory. For the SM, the only possible anomalies are

$$SU(2)_L^2 U(1)_Y, \quad U(1)_Y^3 \quad \text{and} \quad SU(3)_c^2 U(1)_Y. \quad (218)$$

The results in the SM are

$$\begin{aligned} \text{Tr} \left(\left\{ \frac{\tau^a}{2}, \frac{\tau^b}{2} \right\} Y \right)_L &= \frac{1}{2} \delta^{ab} (\text{Tr } Y)_L = 3 \times (N_c \times \frac{1}{3} - 1) = 0, \\ \text{Tr}(\{Y, Y\}Y)_{L-R} &= \dots = 6(-2N_c + 6) = 0, \\ \text{Tr} \left(\left\{ \frac{\lambda^A}{2}, \frac{\lambda^B}{2} \right\} Y \right)_{L-R} &= \frac{1}{3} \delta^{AB} (\text{Tr } Y)_{L-R} = \dots = 0. \end{aligned} \quad (219)$$

Notice that anomaly cancelation happens *precisely* for $N_c = 3$! This seems to provide a deep *connection between quarks and leptons* in the SM, and a possible *hint towards Grand Unified Theories*. Anomaly cancelation gives strong constraint on *new chiral particles*. For example, it is easy to show that (**home-work:**)

- the only flavor-independent, anomaly-free Z' with the chiral SM spectrum is $U(1)_{B-L}$.
- a fourth lepton generation l_4, E_R alone is inconsistent.

9 The Higgs / Symmetry breaking sector of the Standard Model

The Higgs boson is the last building block of the Standard Model awaiting its experimental discovery. There are good theoretical reasons to and preliminary experimental hints from LHC to hope that this can happen quite soon. We review here some of the theoretical biases which make theorists to favor the existence of a light higgs scalar. The first two, the perturbativity and the stability bound, are obtained by extrapolating the Standard Model to high energy scales and imposing perturbativity of couplings and stability of the SM ground state, respectively. The third one is related to the breakdown of unitarity in the longitudinal WW scattering at high-energy if the higgs is too heavy or if does not exist.

9.1 Perturbativity bounds

As shown in Section 7, in quantum field theory couplings run. The Higgs mass is then obtained by knowing the Higgs self-coupling λ at the electroweak scale $M_h^2 = 2\lambda(v)v^2$. If this coupling is large enough, it will hit a Landau pole at a high-energy scale called Λ in what follows. The RGE for the *Higgs self-coupling* in the SM is

$$16\pi^2 \frac{d\lambda}{d\ln\mu} = 24\lambda^2 - (3g'^2 + 3g^2 - 12h_t^2) \lambda + \frac{3}{8}(g'^4 + 2g^2g'^2 + 3g^4) - 6h_t^4 + \dots, \quad (220)$$

where $\dots$ denote smaller Yukawas. In the *large Higgs mass limit* $\lambda \gg g^2, h_t^2$, this reduces to

$$\frac{d\lambda}{\lambda^2} = \frac{3}{2\pi^2} d\ln\mu \Rightarrow \frac{1}{\lambda(\mu)} = \frac{1}{\lambda(\Lambda)} + \frac{3}{2\pi^2} \ln \frac{\Lambda}{\mu}. \quad (221)$$

This can be interpreted in two alternative ways [44]

1. If the *Higgs mass is known*, the SM has a Landau pole (signal of a non-perturbative regime) $\lambda(\Lambda) \gg 1$ at an energy scale

$$\Lambda = v e^{\frac{2\pi^2}{3\lambda}} = v e^{\frac{4\pi^2 v^2}{3M_h^2}}. \quad (222)$$

2. Conversely, asking for *perturbativity* up to scale Λ (say M_{GUT}), we obtain an *upper bound* on the Higgs mass (**homework:**)

$$M_h^2 \leq \frac{4\pi^2 v^2}{3 \ln \frac{\Lambda}{v}}. \quad (223)$$

9.2 Stability bounds

Standard Model has a potential instability in the *small Higgs mass limit* [45], since if too small at the electroweak scale, λ can *become negative* at high-energy by the RG running. If $\lambda \ll h_t^2$, the relevant leading RGE's are

$$16\pi^2 \frac{d\lambda}{d\ln\mu} = -6h_t^4, \quad 16\pi^2 \frac{dh_t}{d\ln\mu} = \frac{9h_t^3}{2}, \quad (224)$$

which integrate to (**homework:**)

$$\begin{aligned} \lambda(\mu) &= \lambda(\Lambda) + \frac{\frac{3h_t^4(\Lambda)}{8\pi^2} \ln \frac{\Lambda}{\mu}}{1 + \frac{9h_t^2(\Lambda)}{16\pi^2} \ln \frac{\Lambda}{\mu}}, \\ h_t^2(\mu) &= \frac{h_t^2(\Lambda)}{1 + \frac{9h_t^2(\Lambda)}{16\pi^2} \ln \frac{\Lambda}{\mu}}. \end{aligned} \quad (225)$$

As in the perturbativity case limit, this can be interpreted in two ways:

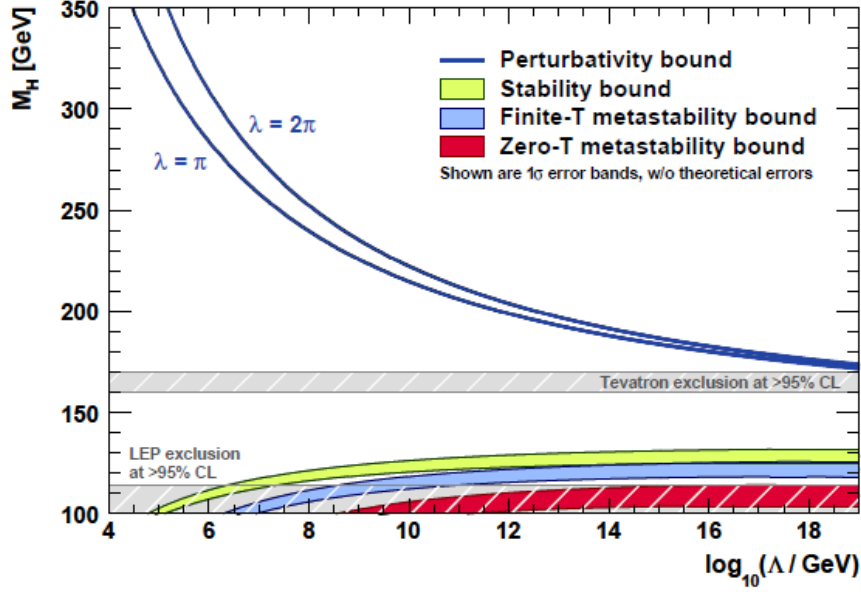


Fig. 22: Perturbativity and stability Higgs mass limits. Λ is the scale of new physics (from Ref. [46]).

1. For a fixed, *known value of the Higgs mass*. Let us take $\mu = v$. Then, new physics should show up before the scale Λ where $\lambda(\Lambda) = 0$,

$$\Lambda \leq v e^{\frac{8\pi^2\lambda}{3h_t^4}} = v e^{\frac{4\pi^2 M_h^2}{3h_t^4 v^2}}. \quad (226)$$

2. Alternatively, for a *fixed* Λ , we get a *lower bound* on the Higgs mass (**homework:**)

$$M_h^2 \geq \frac{3h_t^4 v^2}{4\pi^2} \ln \frac{\Lambda}{v} = \frac{3m_t^4}{\pi^2 v^2} \ln \frac{\Lambda}{v}. \quad (227)$$

These theoretical *Higgs mass limits* are summarized in the plot in Fig. 22 which contains more accurate numerical solution to the RG equations. If the scale Λ is very low, these bounds are very loose. On the other hand, if the SM as an effective theory is valid up to the Planck scale, we obtain a pretty tight mass range $120 \text{ GeV} \lesssim M_h \lesssim 170 \text{ GeV}$.

9.3 $W W$ scattering and unitarity

There is another bound on the higgs mass which does not involve extrapolations of the SM model to very high energies. It is coming from the unitarity of scattering amplitude for the longitudinal $W_L W_L \rightarrow W_L W_L$ scattering [47].

For a massive W gauge particle of momentum k and mass M_W , $A_m = \varepsilon_m e^{ikx}$, the three polarizations satisfy $\varepsilon_m \varepsilon^m = -1$, $k_m \varepsilon^m = 0$. In the rest frame $k^m = (E, 0, 0, k)$, they are

$$\begin{aligned} \text{transverse : } \varepsilon_1^m &= (0, 1, 0, 0) \quad , \quad \varepsilon_2^m = (0, 0, 1, 0) \quad , \\ \text{longitudinal : } \varepsilon_L^m &= \left(\frac{k}{M_W}, 0, 0, \frac{E}{M_W} \right) \sim \frac{k^m}{M} + \mathcal{O}\left(\frac{M_W}{E}\right) \quad , \end{aligned} \quad (228)$$

the last expressions being valid for $k \rightarrow \infty$. Since longitudinal polarization is proportional to the energy, tree-level amplitude behaves as

$$\mathcal{A} = \mathcal{A}^{(4)} \left(\frac{E}{M_W} \right)^4 + \mathcal{A}^{(2)} \left(\frac{E}{M_W} \right)^2 + \dots \quad (229)$$

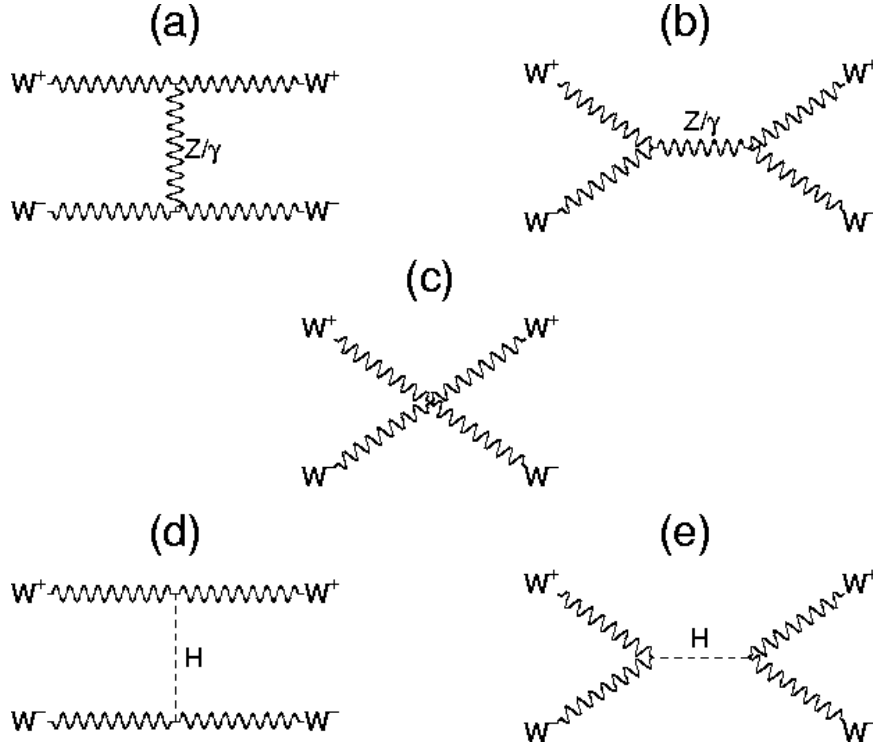


Fig. 23: Tree-level diagrams contributing to WW scattering.

Actually, the diagrams a), b) and c) in Figure 27 give $\mathcal{A} = g^2 \left(\frac{E}{M_W} \right)^2$. On the other hand, *unitarity constrains the amplitude to stay small enough* at any energy. In order to see this, let us consider the unitarity of the S-matrix $S^\dagger S = 1$. Then

$$S = 1 + i\mathcal{A} \quad \Rightarrow \quad i(\mathcal{A} - \mathcal{A}^\dagger) + \mathcal{A}^\dagger \mathcal{A} = 0 \quad (230)$$

By sandwiching this equation between a two-particle state $|i\rangle$:

$$i(\mathcal{A} - \mathcal{A}^\dagger)_{ii} + \sum_f |\mathcal{A}_{fi}|^2 = 0, \quad (231)$$

we find the **optical theorem**: *the imaginary part of the forward amplitude of the process $i \rightarrow i$ is proportional to the total cross section of $i \rightarrow$ anything*. Let us now decompose the scattering amplitude into partial waves

$$\mathcal{A} = \sum_{l=0}^{\infty} (2l+1) P_l(\cos \theta) a_l, \quad (232)$$

where a_l are partial wave amplitudes of the elastic scattering of two particles. Projecting Eq. (231) into the partial wave l gives: $\text{Im } a_l = |a_l|^2$. This is only possible if

$$|\text{Re } a_l| \leq 1/2, \quad 0 \leq \text{Im } a_l \leq 1 \quad \rightarrow \quad |a_l|^2 \leq 5/4, \quad (233)$$

which is the *unitarity bound* we were searching for.

In the case of the SM without the Higgs boson, diagrams a), b) and c) in Fig. 23 lead to

$$a_0 = \frac{g^2 E^2}{M_W^2} \quad \Rightarrow \quad \text{unitarity breaks down for } \sqrt{s} \sim 1.2 \text{ TeV}. \quad (234)$$

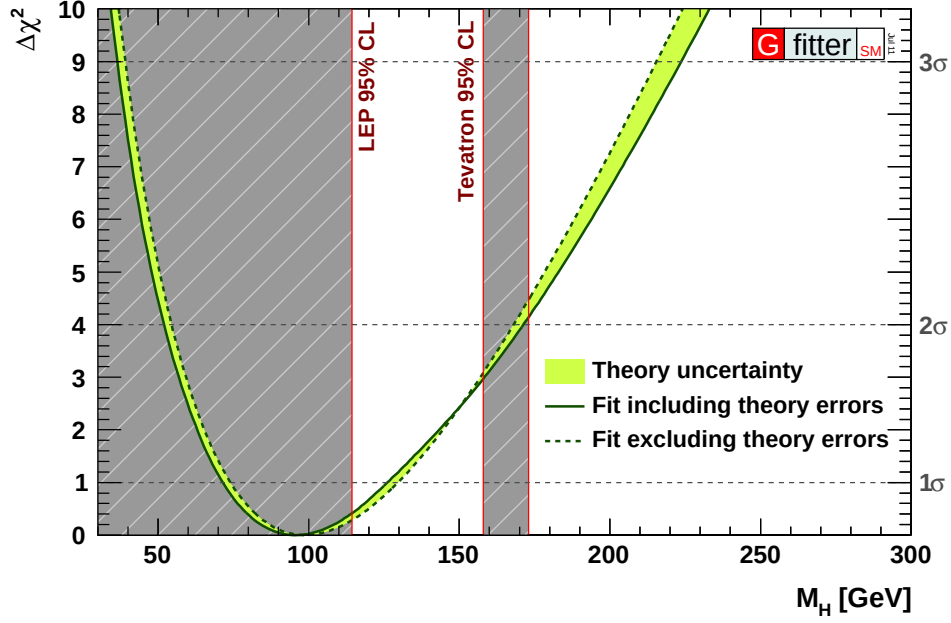


Fig. 24: SM precision tests favor the existence of a light Higgs. Taken from Gfitter webpage [48].

With the Higgs boson present, amplitudes d), e) in Fig. 23 cancel the raising energy term, such that

$$a_0 = \frac{g^2 M_H^2}{4M_W^2} \rightarrow \text{unitarity breaks down unless } M_H \leq 1.2 \text{ TeV} . \quad (235)$$

And by considering other channels, one get the stronger bound $M_H \leq 800 \text{ GeV}$.

Intepretation: If LHC finds no Higgs with a mass $M_H \leq 800 \text{ GeV}$, unitarity of S-matrix will be violated. New light degrees of freedom should exist in order to restore unitarity. Most theorists interpret this result as a *no-loose "theorem"* for LHC: either LHC finds the Higgs, or it should find the degrees of freedom replacing it in order to unitarize the WW scattering.

It is important to keep in mind however that most BSM models have *invisible higgs decays*. For example, dark matter models can have higgs decays into dark matter particles $h \rightarrow DM DM$. In this case, higgs searches are more complicated: the higgs can be "hidden" due to its non-standard decays.

There are other constraints on the Higgs mass that we do not discuss here, coming from precision tests in the Standard Model (see Fig. 24). Most theories have a biased towards a *light Higgs*, since it provides a better fit for the SM precision tests.

9.4 Higgs and the hierarchy problem

Quantum corrections to the Higgs mass in the SM, coming from diagrams in Fig. 25, are quadratically divergent

$$\delta m_h^2 \simeq \frac{3\Lambda^2}{8\pi^2 v^2} (4m_t^2 - 4M_W^2 - 2M_Z^2 - m_h^2) . \quad (236)$$

In a theory including gravity or GUT's, Λ is a physical mass scale $\Lambda = M_P, M_{GUT}$. It is then difficult to understand why

$$m_h^2 = (m_h^0)^2 + \frac{3\Lambda^2}{8\pi^2 v^2} (4m_t^2 - 4M_W^2 - 2M_Z^2 - m_h^2) \sim v^2 \ll \Lambda^2 . \quad (237)$$

This is the *the hierarchy problem* [49].

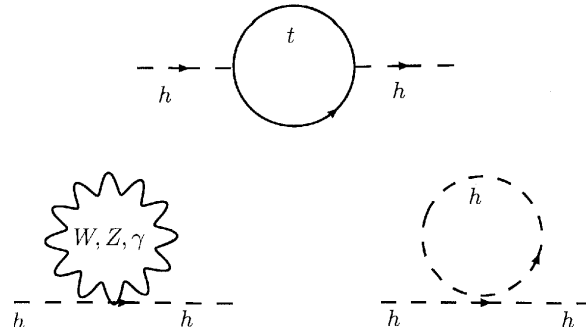


Fig. 25: Quadratic divergences to the Higgs mass in the SM, leading to the *hierarchy problem*.

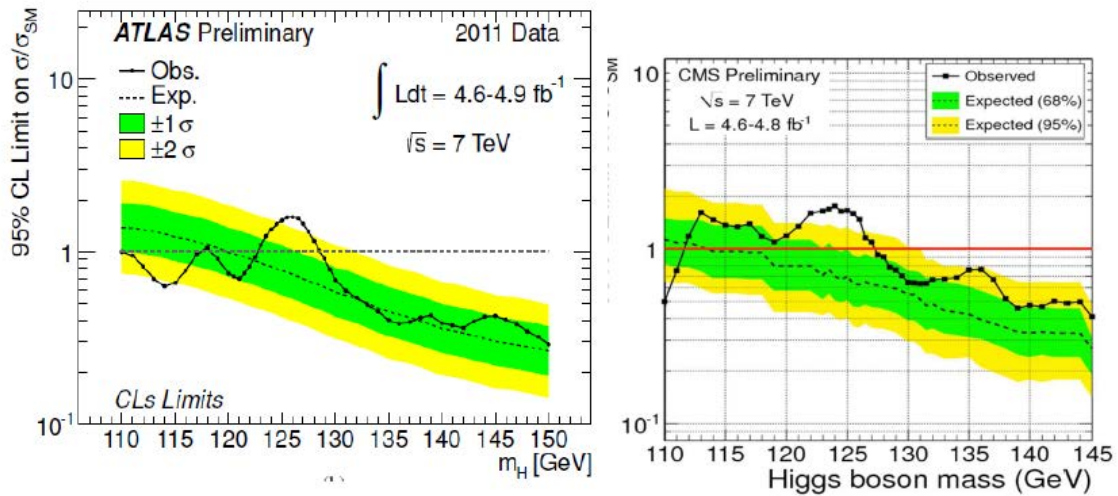


Fig. 26: Possible evidence of a Higgs scalar with mass around 125 GeV from ATLAS (left) and CMS (right), presented at the 2012 Moriond conference [52].

The latest news before this School, from "Lepton-Photon" in August 2011 concerning the Higgs were that both ATLAS and CMS did exclude the SM Higgs at 95 CL for $145 \text{ GeV} \leq M_H \leq 446 \text{ GeV}$ except 288–296 GeV. Before the Christmas 2012 however, some excess in the data, first at ATLAS and then at CMS, has been interpreted as the first possible evidence for a Higgs boson around 125 GeV. Figure 26 summarizes the situation in the Moriond 2012 conference [52].

10 Epilogue: Can Standard Model be the final theory?

Most people believe that Standard Model is just an effective description, for a lot of various reasons:

- There are no *neutrino masses* at the renormalizable level. The neutrino masses and mixings are often considered as a first hint towards a new mass scale beyond the Standard Model. The seesaw mechanism points towards heavy Majorana singlet neutrinos, maybe remnants of Grand Unified Theories.
- The mysterious *hierarchies* in the quarks/lepton masses and mixings. It is likely that quarks and leptons hierarchies hide the existence of new flavor symmetries or of a geometrical origin related to wave functions profiles in a higher-dimensional space.
- Standard Model has no viable *Dark Matter* candidate. This is currently maybe the most pressing

problem: understanding the origin and the properties of the dark matter candidate, which provides about 30% of the energy density of the Universe.

- The problem with the *radiative stability* of the electroweak scale ("the hierarchy problem").
- SM has no accurate *gauge coupling unification*.
- The *strong CP problem*. The most popular solution postulates the existence of new light particles, the axions, which exist in all string theories and often play a central role in their quantum consistency.
- *Gravity is not incorporated* into a renormalizable framework. The only viable well-studied framework of quantum gravity to date is string theory.
- The *cosmological constant* problem $\Lambda \sim 10^{-4} \text{eV}^4 \sim 10^{-120} M_p^4$. This is certainly the biggest mystery in modern physics.

Note that the third, fourth and fifth problems find together a nice solution in low-energy supersymmetry. The elegant embedding of quarks and leptons into complete representations of $SU(5)$ also points out towards a unified gauge group structure.

On the other hand, any theory describing nature has to be validated by experiments. For the time being, LHC found no signal of new physics. It is still early to judge the viability of low-energy supersymmetry or extra dimensional models. There are however preliminary positive LHC hints for a light Higgs boson [51]. But if no SM higgs is discovered by the end of the next year, *something else* must replace it in order to save the unitarity of the S-matrix in the Standard Model. It is likely in this case that new strong forces and resonances exist at TeV energies and LHC should be able to see them after a couple of years of running.

Acknowledgments

I would like to thank the organizers of the CERN summer school 2011 for providing a stimulating and very pleasant environment. This work was partially supported by the European ERC Advanced Grant 226371 MassTeV, by the European Initial Training Network PITN-GA-2009-237920 and by the Agence Nationale de la Recherche.

References

- [1] P. W. Higgs, Phys. Lett. **12** (1964) 132; F. Englert and R. Brout, Phys. Rev. Lett. **13** (1964) 321.
- [2] F. Maltoni, *Basics of QCD for the LHC: $pp \rightarrow H + X$ as a case study*, in these proceedings.
- [3] E. Iancu, *QCD in Heavy Ions Collisions*, in these proceedings.
- [4] V. Rubakov, *Cosmology*, in these proceedings.
- [5] B. Kayser, *Neutrino Oscillation Physics*, in these proceedings.
- [6] B. Dobrescu, *Beyond The Standard Model*, in these proceedings.
- [7] <http://is.gd/E9ntxH>
- [8] C. Itzykson and J. B. Zuber, *Quantum Field Theory*, (Mcgraw-Hill, New York, 1980).
- [9] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*, (Addison-Wesley, Reading, USA 1995).
- [10] S. Weinberg, *The Quantum theory of fields. Vol. 1: Foundations*, (Cambridge Univ. Pr., 1995).
- [11] S. Weinberg, *The quantum theory of fields. Vol. 2: Modern applications*, (Cambridge Univ. Pr., 1996).
- [12] A. Zee, *Quantum field theory in a nutshell*, (Princeton Univ. Pr., 2010).
- [13] P. Ramond, *Field Theory. A Modern Primer*, Front. Phys. **51** (1981) 1.

- [14] S. S. Schweber, *QED and the men who made it: Dyson, Feynman, Schwinger, and Tomonaga*, (Princeton Univ. Pr., 1994).
- [15] Z. Bern, Living Rev. Rel. **5** (2002) 5 [gr-qc/0206071].
- [16] J. S. Schwinger, Phys. Rev. **73** (1948) 416.
- [17] <http://is.gd/LsvXUw>
- [18] C.N. Yang and R. L. Mills, Phys. Rev. **96** (1954) 191.
- [19] L. D. Faddeev and V. N. Popov, Phys. Lett. B **25** (1967) 29.
- [20] Y. Nambu and G. Jona-Lasinio, Phys. Rev. **122** (1961) 345; J. Goldstone, Nuovo Cim. **19** (1961) 154; J. Goldstone, A. Salam and S. Weinberg, Phys. Rev. **127** (1962) 965.
- [21] <http://is.gd/iMIeEn>
- [22] <http://is.gd/jetgWD>
- [23] <http://is.gd/BlA7tb>
- [24] S. R. Coleman, Subnucl. Ser. **11** (1975) 139.
- [25] S. L. Glashow, Nucl. Phys. **22** (1961) 579; S. Weinberg, Phys. Rev. Lett. **19** (1967) 1264; A. Salam, *Elementary Particle Theory*, Ed. W. Svartholm (Almqvist and Wiksell, Stockholm, 1968).
- [26] S. L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. D **2** (1970) 1285; M. K. Gaillard and B. W. Lee, Phys. Rev. D **10** (1974) 897.
- [27] S. L. Adler, Phys. Rev. **177** (1969) 2426; J. S. Bell and R. Jackiw, Nuovo Cim. A **60** (1969) 47; S. L. Adler and W. A. Bardeen, Phys. Rev. **182** (1969) 1517.
- [28] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38** (1977) 1440; S. Weinberg, Phys. Rev. Lett. **40** (1978) 223.
- [29] V. A. Kuzmin, V. A. Rubakov and M. E. Shaposhnikov, Phys. Lett. B **155** (1985) 36.
- [30] C. Bouchiat, J. Iliopoulos and P. Meyer, Phys. Lett. B **38** (1972) 519; D. J. Gross and R. Jackiw, Phys. Rev. D **6** (1972) 477.
- [31] N. Cabibbo, Phys. Rev. Lett. **10** (1963) 531.
- [32] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49** (1973) 652.
- [33] C. D. Froggatt and H. B. Nielsen, Nucl. Phys. B **147** (1979) 277.
- [34] <http://is.gd/003Adl>
- [35] P. Sikivie, L. Susskind, M. B. Voloshin and V. I. Zakharov, Nucl. Phys. B **173** (1980) 189.
- [36] G. Altarelli and R. Barbieri, Phys. Lett. B **253** (1991) 161; M. E. Peskin and T. Takeuchi, Phys. Rev. D **46** (1992) 381; for a review of electroweak precision tests, see e.g. G. Altarelli, R. Barbieri and F. Caravaglios, Int. J. Mod. Phys. A **13** (1998) 1031 [hep-ph/9712368].
- [37] G. 't Hooft, Nucl. Phys. B **35** (1971) 167; G. 't Hooft and M. J. G. Veltman, Nucl. Phys. B **44** (1972) 189, Nucl. Phys. B **50** (1972) 318; B. W. Lee and J. Zinn-Justin, Phys. Rev. D **5** (1972) 3121, Phys. Rev. D **5** (1972) 3137 [Erratum-ibid. D **8** (1973) 4654]; for a book dedicated to the subject, see e.g. J. C. Collins, Cambridge, UK: Univ. Pr. (1984) 380p.
- [38] K. G. Wilson and J. B. Kogut, Phys. Rept. **12** (1974) 75.
- [39] A. Pich, arXiv:1201.0537 [hep-ph].
- [40] D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30** (1973) 1343; H. D. Politzer, Phys. Rev. Lett. **30** (1973) 1346.
- [41] H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32** (1974) 438; H. Georgi, H. R. Quinn and S. Weinberg, Phys. Rev. Lett. **33** (1974) 451.
- [42] S. Dimopoulos, S. Raby and F. Wilczek, Phys. Rev. D **24** (1981) 1681; for a review see e.g. K. R. Dienes, Phys. Rept. **287** (1997) 447 [hep-th/9602045].
- [43] K. Zioutas, M. Tsagri, T. Papaevangelou and T. Dafni, New J. Phys. **11** (2009) 105020 [arXiv:0903.1807 [astro-ph.SR]].

- [44] N. Cabibbo, L. Maiani, G. Parisi and R. Petronzio, Nucl. Phys. B **158** (1979) 295; R. Dashen and H. Neuberger, Phys. Rev. Lett. **50** (1983) 1897; D. Callaway, Nucl. Phys. B **233** (1984) 189.
- [45] N. Cabibbo, L. Maiani, G. Parisi and R. Petronzio, Nucl. Phys. B **158** (1979) 295; P. Q. Hung, Phys. Rev. Lett. **42** (1979) 873; M. Lindner, Z. Phys. C **31** (1986) (1986).
- [46] J. Ellis, J. R. Espinosa, G. F. Giudice, A. Hoecker and A. Riotto, Phys. Lett. B **679** (2009) 369 [arXiv:0906.0954 [hep-ph]].
- [47] B. W. Lee, C. Quigg and H. B. Thacker, Phys. Rev. D **16** (1977) 1519; M. S. Chanowitz and M. K. Gaillard, Phys. Lett. B **142** (1984) 85, Nucl. Phys. B **261** (1985) 379.
- [48] GFitter, <http://is.gd/dxADKg>
- [49] S. Weinberg, Phys. Rev. D **13** (1976) 974, Phys. Rev. D **19** (1979) 1277; E. Gildener, Phys. Rev. D **14** (1976) 1667; L. Susskind, Phys. Rev. D **20** (1979) 2619; G. 't Hooft, in *Proceedings of the NATO Advanced Summer Institute, Cargese 1979* (Plenum, 1980).
- [50] M. E. Peskin, *Summary of Lepton Photon 2011*, arXiv:1110.3805 [hep-ph].
- [51] F. Gianotti (ATLAS) and G. Tonelli (CMS), talks given at the CERN seminar on update on the Standard Model Higgs searches, December 2011, CERN; The ATLAS Collaboration, *Combination of Higgs Boson Searches with up to 4.9fb^{-1} of pp Collisions Data Taken at a center-of-mass energy of 7 TeV with the ATLAS Experiment at the LHC*, ATLAS-CONF-2011-163; The CMS Collaboration, *Combination of SM Higgs Searches*, CMS-PAS-HIG-11-032.
- [52] S. Kortner (ATLAS), M. Pieri (CMS) and J.F. Grivaz, talks at Rencontres de Moriond, March 2012, La Thuile, Italy.

Basics of QCD for the LHC: $pp \rightarrow H + X$ as a case study

F. Maltoni

Université Catholique de Louvain, Louvain-la-Neuve, Belgium

Abstract

Quantum Chromo Dynamics (QCD) provides the theoretical framework for any study of TeV scale physics at LHC. Being familiar with the basic concepts and techniques of QCD is therefore a must for any high-energy physicist. In these notes we consider Higgs production via gluon fusion as an example on how accurate and flexible predictions can be obtained in perturbative QCD. We start by illustrating how to calculate the total cross section at the leading order (yet one loop) in the strong coupling α_S and go through the details of the next-to-leading order calculation eventually highlighting the limitations of fixed-order predictions at the parton level. Finally, we briefly discuss how more exclusive (and practical) predictions can be obtained through matching/merging fixed-order results with parton showers.

1 Introduction

Strongly interacting particles can be described in terms of a $SU(3)$ gauge theory field theory involving gluons and quarks:

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4}G^{\mu\nu,a}G_{\mu\nu}^a + \sum_f \bar{\psi}_i^f i \not{D}_{ij} \psi_j^f, \quad (1)$$

where the sum runs over the quark flavors,

$$\begin{aligned} G_{\mu\nu}^a &= \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - g_s f^{abc} A_\mu^b A_\nu^c, \\ D_{\mu,ij} &= \partial_\mu \delta_{ij} + i g_s t_{ij}^a A_\mu^a, \end{aligned}$$

and t_{ij}^a are the Gell-Mann matrices in the fundamental representation and f^{abc} are the structure functions of $SU(3)$, with

$$[t^a, t^b] = i f^{abc} t^c. \quad (2)$$

Notwithstanding its apparent simplicity, QCD is an amazingly rich theory which is able to account for a wide diversity of phenomena, ranging from really strong (non-perturbative) interactions at low scales, below 1 GeV, to rather weak (perturbative) interactions up to scales of the TeV at colliders, from low density to high density states such as those happening in nuclei collisions or inside stars, from low to high temperatures. For proton-proton collisions at the LHC, where one can consider zero temperature and density, QCD is complicated enough that we have no means available (for the moment!) to solve it exactly and we have to resort to a variety of approximate methods, including perturbation theory (when the coupling is small) and lattice calculations (when the coupling is large). Thanks to the work of theoretical and experimental physicists over the last forty years we are convinced that QCD is a good theory of the strong interactions, of course in the range of energies explored so far and to the level of the theoretical accuracy that can be achieved with current technologies.

There are many excellent references on QCD with applications to collider physics, from books, (e.g., Ref. [1]) to review articles, to write-up of lectures given in schools, and in particular some of those given at the CERN schools over the years. My lectures at the school were largely based on the inspiring ones by Michelangelo Mangano [2], Paolo Nason [3] and on the most recent ones by Gavin Salam [4], which I warmly recommend. In these notes, I'll present a case study, i.e. how QCD can make accurate

predictions for Higgs production in gluon fusion at the LHC. The aim is to see the basic concepts at work for a realistic and very important process so to verify their understanding and also to have a closer look at the basic techniques used to perform such calculations. When needed and to avoid repetitions, I will refer to specific sections of Ref. [4] as [QCD: Section number] where the reader will find further information on the basic concepts. Links to simple Mathematica® notebooks with the calculations described below can be found at <http://maltoni.home.cern.ch/>.

2 Higgs cross section at the LHC

The factorisation theorem states that the total cross section for the inclusive production of Higgs at the LHC can be written as ¹

$$\sigma(H + X) = \sum_{i,j} \int dx_1 f_i(x_1, \mu_F) \int dx_2 f_j(x_2, \mu_F) \times \hat{\sigma}_{ij \rightarrow H+X}(s, m_H, \mu_F, \mu_R), \quad (3)$$

where the $f_{i/j}(x, \mu_F)$ are the parton distributions functions (long distance term, non-perturbatively calculable) and $\hat{\sigma}$ is the partonic cross section (short distance term, calculable in perturbation theory).

$\hat{\sigma}$ can be written as an expansion in α_S :

$$\begin{aligned} \hat{\sigma}(ij \rightarrow H + x) &= \hat{\sigma}^{(0)}(ij \rightarrow H) \\ &+ \hat{\sigma}^{(1)}(ij \rightarrow H + \text{up to 1 parton}) \\ &+ \hat{\sigma}^{(2)}(ij \rightarrow H + \text{up to 2 partons}) \\ &+ \dots \end{aligned} \quad (4)$$

where the first term gives the leading order (LO) approximation and it is of order α_S^2 , the second next-to-leading (NLO) order (α_S^3) and so on.

It is interesting to know how the Higgs predictions improved and evolved over time. The LO production was considered a long ago [5], the next-to-leading order (NLO) QCD corrections [6–9] were calculated decades ago in the so-called effective field theory (HEFT) approximation (which will be explained in the following) as well in the full SM and found to be very large ($\sigma^{\text{NLO}}/\sigma^{\text{LO}} \sim 2$). This motivated the formidable endeavour of the next-to-next-to-leading order (NNLO) QCD calculations, which have been fully evaluated in HEFT [10–12]. Given that corrections to the HEFT been estimated through a power expansion [13–16] and found to have a negligible impact on total rates, NNLO is the current state of the art for fixed-order predictions.

Before going into the details of the computation of the Higgs cross section, let us remind a few general important points that are relevant for any computation in QCD.

- At LO the factorisation theorem reduces to the parton model: the parton distribution functions $f_i(x)$ are just the probabilities (and therefore positive-definite) of finding a given parton in the initial state hadrons at a given resolution scale μ_F and $\hat{\sigma}$ gives the probability that such partons with a total energy $s = x_1 x_2 S$ will "fuse" into a Higgs.
- Total cross sections are the first and simplest example of a larger class of observables, called Infrared Safe (IS) quantities [QCD:2.3.2], which can be consistently computed in QCD and then compared to experimental data. Such quantities always need to be (at least to some degree) inclusive on possible extra radiation and in particular resilient under soft and/or collinear radiation. The most known example of IS quantities beyond total cross sections are jets [QCD:5]. The constraint of infrared safety becomes non-trivial already at NLO for Eq. (3).

¹Be careful here as for simplicity we adopt the usual pragmatic approach on Higgs production at the LHC and imagine it coming from different channels: gluon-gluon fusion, vector-boson-fusion, vector-boson-associated...and so on. We restrict the discussion to the first one which is the leading mechanism. In fact, various channels overlap if contributions are organized as powers of strong and weak couplings (e.g., $gg \rightarrow H$ appears at the same order in α_S and y_t as $gg \rightarrow t\bar{t}H$) and in general they mix-up once higher-order QCD and EW corrections are included. The separation into channels is anyway useful from the experimental point of view as they typically lead to different final state signatures.

- Total cross sections always inclusive of any possible extra QCD radiation in the event, hereby denoted by X , even when the calculation is performed at LO. In this case, extra radiation up to the scale μ_F is accounted for by the parton distribution function's (PDF), while hard radiation is consistently neglected being of higher order (α_S). Alternatively, one can prove that the total cross section for producing "just a Higgs", i.e., Higgs + no resolvable radiation at an arbitrary small scale is exactly zero at all orders in perturbation theory.
- A very important point to always keep in mind is that the "adjectives" LO, NLO, NNLO need to be always referred to a specific observable, i.e. different observables in a given calculation can be predicted at a different order. For example, when talking about a "NNLO calculation for Higgs production in gluon fusion", what is really meant is that the total inclusive cross section is known at NNLO. The same calculation can predict the rate for Higgs+1 jet (inclusive and exclusive) at NLO and Higgs+2 jets only at LO (where exclusive and inclusive is the same).
- Beyond LO, the separation between long-distance and short-distance physics as described by μ_F (and also μ_R) becomes non-trivial. μ_F and μ_R represent arbitrary scales in the calculation, whose dependence is generated by the truncation of the perturbative expansion at a given order. Exploiting the fact that physical results must be independent on such scales one finds renormalisation-group type equations, such as the β function of QCD [QCD:1.2.3] and the so-called DGLAP evolution equations for the PDF's [QCD:3.2].
- The residual dependence of σ on μ_F and μ_R at any given order in perturbation theory is often used to gauge the accuracy of the predictions [QCD:4.4.1]. This is by itself a very crude approximation, while the towers of leading (subleading,...) log's of the scales can be predicted at all orders in perturbation theory, only an explicit computation is able to provide the finite terms at higher orders. In practice, it is common to choose central scales as the typical hard scale in a process and vary them independently between 1/2 and 2 to identify an uncertainty. However, no solid and unique procedure exists to identify central reference values and variation intervals and to associate a confidence level. However, milder scale dependence of higher-order results compared to lower ones is always used to gauge the improvement on the accuracy of a given prediction.

3 $pp \rightarrow H + X$ at leading order

At LO Eq. 3 can be rewritten as

$$\sigma^{\text{LO}}(H + X) = \int_{\tau_0}^1 dx_1 \int_{\tau_0/x_1}^1 dx_2 f_g(x_1, \mu_F) f_g(x_2, \mu_F) \times \hat{\sigma}^{(0)}(gg \rightarrow H), \quad (5)$$

where $\tau_0 = m_H^2/S$ and $s = x_1 x_2 S$. $\hat{\sigma}$ for a $2 \rightarrow 1$ process can be rewritten as

$$\begin{aligned} \hat{\sigma} &= \frac{1}{2s} \overline{|\mathcal{A}|^2} \frac{d^3 P}{(2\pi)^3 2E_H} (2\pi)^4 \delta^4(p + q - P_H) \\ &= \frac{1}{2s} \overline{|\mathcal{A}|^2} 2\pi \delta(s - m_H^2), \end{aligned} \quad (6)$$

where

$$\tau \equiv x_1 x_2 = \frac{S}{s}, \quad \tau_0 = \frac{m_H^2}{S}. \quad (7)$$

Performing the change of variables $x_1, x_2 \rightarrow \tau, y$ with $x_1 \equiv \sqrt{\tau} e^y$, $x_2 \equiv \sqrt{\tau} e^{-y}$ (verify that the jacobian J is equal to 1) the change of the integration limits and the result becomes

$$\sigma^{\text{LO}}(H + X) = \frac{\pi \overline{|\mathcal{A}|^2}}{m_H^2 S} \int_{\log \sqrt{\tau_0}}^{-\log \sqrt{\tau_0}} dy x g(\sqrt{\tau_0} e^y) g(\sqrt{\tau_0} e^{-y}). \quad (8)$$

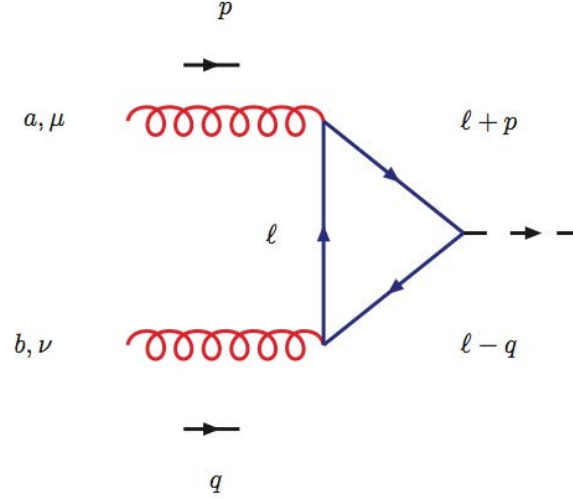


Fig. 1: Representative Feynman diagram for the process $gg \rightarrow H$. Another diagram, the one with the gluons exchanged, contributes to the total amplitude.

This expression shows that for the cross section of a $2 \rightarrow 1$ process at LO, the contribution from the parton distributions (a quantity known as gluon-gluon luminosity) factorises from the dynamics ($|\mathcal{A}|^2$). The gluon-gluon luminosity depends only on the kinematics in the limits of integration and can be computed once for all for each Higgs mass. The problem is therefore reduced to the computation of the amplitude $\mathcal{A}$.

3.1 My first loop (yet finite!) amplitude: $gg \rightarrow H$

Being a color singlet, the Higgs does not couple directly to gluons. However, as no fundamental symmetry forbidding it is present² it can via a loop of a colored and massive particle. In the SM such states are the heavy quarks. Let us consider one quark at the time, i.e., the diagram(s) shown in Fig. 1. The first observation to make, even before starting the calculation, is that even though a triangle loop in general can give rise to divergences, both in the ultra-violet (UV) and in the infrared (IR), in this case we expect a finite result. There are several different ways of convincing that this must be the case. A simple one goes as follows. Divergent terms always factorize over lower order amplitudes. The one-loop amplitude is the first non-zero term contributing to $gg \rightarrow H$ in the perturbative expansion. Therefore there cannot be any divergence. A finite amplitude, however, does not mean that a consistent regularisation procedure is not needed. The reason is that in intermediate steps of the calculation infinities are found that cancel at the end, yet might leave finite terms. As we will see in $gg \rightarrow H$ such finite terms are actually necessary to guarantee the gauge invariance of the result, clearly showing that there is no ambiguity in the procedure.³

To evaluate the diagram of Fig. 1 (there are actually two diagrams, the one shown and another one with the gluons exchanged. They give the same contribution so we'll just multiply our final result by two), we employ use dimensional regularisation in $d = 4 - 2\epsilon$ dimensions.⁴

²In fact, classically, scale invariance would forbid such a coupling. However, scale invariance is broken by renormalisation and therefore it is not a symmetry.

³Less obvious is the case of $\gamma\gamma \rightarrow H$ where the contribution coming from gauge bosons loop has to be done in different gauges (or via low-energy-theorems) to prove the uniqueness and the correctness of dimensional regularisation procedure. Interestingly enough, people seem to forget this fact quite regularly over the years.

⁴Dimensional regularisation comes in several different flavors and attention has to be paid to the details of the implementation. All formulas quoted in the main body of these lecture notes are in the so-called Conventional Dimensional Regularization (CDR) which is the regularisation procedure where the $\overline{\text{MS}}$ scheme is defined. In practice, NLO calculations nowadays are

Using the QCD Feynman rules (QCD: Fig. 3) and the Yukawa interaction, the expression for the amplitude corresponding to the diagram of Fig. 1 reads:

$$i\mathcal{A} = -(-ig_s)^2 \text{Tr}(t^a t^b) \left(\frac{-im_Q}{v} \right) \int \frac{d^d \ell}{(2\pi)^d} \frac{t^{\mu\nu}}{\text{Den}} (i)^3 \epsilon_\mu(p) \epsilon_\nu(q) \quad (9)$$

where the overall minus sign is due to the closed fermion loop.⁵ The denominator is $\text{Den} = (\ell^2 - m_Q^2)[(\ell + p)^2 - m_Q^2][(\ell - q)^2 - m_Q^2]$. Employing the usual Feynman parametrization method to combine the denominators of the loop integral into one:

$$\frac{1}{ABC} = 2 \int_0^1 dx \int_0^{1-x} dy \frac{1}{[Ax + By + C(1-x-y)]^3} \quad (10)$$

one obtains

$$\frac{1}{\text{Den}} = 2 \int dx dy \frac{1}{[\ell^2 - m_Q^2 + 2\ell \cdot (px - qy)]^3}. \quad (11)$$

The next step is to shift the integration momenta to $\ell' = \ell + px - qy$ so the denominator takes the form

$$\frac{1}{\text{Den}} \rightarrow 2 \int dx dy \frac{1}{[\ell'^2 - m_Q^2 + m_H^2 xy]^3}. \quad (12)$$

The numerator of the loop integral in the shifted loop momentum becomes

$$\begin{aligned} t^{\mu\nu} &= \text{Tr} \left[(\ell + m_Q) \gamma^\mu (\ell + \not{p} + m_Q) (\ell - \not{q} + m_Q) \gamma^\nu \right] \\ &= 4m_Q \left[g^{\mu\nu} (m_Q^2 - \ell'^2 - \frac{m_H^2}{2}) + 4\ell'^\mu \ell'^\nu + p^\nu q^\mu \right]. \end{aligned} \quad (13)$$

where we have used the fact that for transverse gluons, $\epsilon(p) \cdot p = 0$ and so terms proportional to the external momenta, p_μ or q_ν , have been dropped. The above expression shows already several interesting aspects.

The first one is that the trace is proportional to the heavy quark mass. This can be easily understood as an effect of the spin-flip coupling of the Higgs. Gluons or photons do not change the spin of the fermion, as vectors map left (right) spinors into left (right) spinors, while the scalars do couple left (right) spinors with right (left) ones. If the quark circulating in the loop is massless then the trace vanishes due to helicity conservation, independently of the actual Yukawa coupling. This is the reason why even when the Yukawa coupling of the light quark and the Higgs is enhanced (such as in SUSY or 2HDM with large $\tan \beta$), the contribution is anyway suppressed by the kinematical mass.

The second point is that simple power counting shows that the terms proportional to the squared loop momentum ℓ'^2 and $\ell'^\mu \ell'^\nu$ give rise to UV divergences. This means that an intermediate and consistent regularisation prescription is needed for intermediate manipulations and that divergences will have to cancel in the final result.

By shifting momenta in the numerator, dropping terms linear in ℓ' and using the relation

$$\int d^d k \frac{k^\mu k^\nu}{(k^2 - C)^m} = \frac{1}{d} g^{\mu\nu} \int d^d k \frac{k^2}{(k^2 - C)^m} \quad (14)$$

to write the amplitude in the form

$$i\mathcal{A} = -\frac{2g_s^2 m_Q^2}{v} \delta^{ab} \int \frac{d^d \ell'}{(2\pi)^d} \int dx dy \left\{ g^{\mu\nu} \left[m_Q^2 + \ell'^2 \left(\frac{4-d}{d} \right) + m_H^2 \left(xy - \frac{1}{2} \right) \right] \right.$$

done in a different scheme which limits the use of the d -dimensional Dirac algebra to the loop computation.

⁵ $\epsilon_\mu(p)$ are the transverse gluon polarizations.

$$+p^\nu q^\mu(1-4xy)\left\}\frac{2dxdy}{(\ell^2-m_Q^2+m_H^2xy)^3}\epsilon_\mu(p)\epsilon_\nu(q).\quad (15)$$

This expression shows that if one computes the integral in $d = 4$, the UV divergent term is absent. For $d = 4 - 2\epsilon$, however, this gives rise to a left-over finite piece, as the scalar integrals are given by

$$\begin{aligned}\int \frac{d^d\ell}{(2\pi)^d} \frac{\ell^2}{(\ell^2-C)^3} &= \frac{i}{32\pi^2} (4\pi)^\epsilon \frac{\Gamma(1+\epsilon)}{\epsilon} (2-\epsilon) C^{-\epsilon} \\ \int \frac{d^d\ell}{(2\pi)^d} \frac{1}{(\ell^2-C)^3} &= -\frac{i}{32\pi^2} (4\pi)^\epsilon \Gamma(1+\epsilon) C^{-1-\epsilon}.\end{aligned}\quad (16)$$

So it is manifest that the divergence $1/\epsilon$ cancels against the $(4-d)/d$ term leaving a finite piece, which in fact ensures that the final result is gauge invariant. By combining it with the other terms in the squared parenthesis we obtain

$$\mathcal{A}(gg \rightarrow H) = -\frac{\alpha_S m_Q^2}{\pi v} \delta^{ab} \left(g^{\mu\nu} \frac{m_H^2}{2} - p^\nu q^\mu \right) \epsilon_\mu(p) \epsilon_\nu(q) \int dxdy \left(\frac{1-4xy}{m_Q^2 - m_H^2 xy} \right). \quad (17)$$

(Note that we have multiplied by 2 in Eq. (17) to include the diagram where the gluon legs are crossed.) The Feynman integral of Eq. (17) can easily be performed to find an analytic result if desired. Note that the tensor structure could have been predicted from the start by imposing gauge invariance, i.e., $p^\mu \mathcal{A}^{\mu\nu} = q^\nu \mathcal{A}^{\mu\nu} = 0$. By defining $I(a)$ as

$$I(a) \equiv \int_0^1 dx \int_0^{1-x} dy \frac{1-4xy}{1-axy}, \quad a = \frac{m_H^2}{m_Q^2}, \quad (18)$$

one can factorise a $1/m_Q^2$ out of the integral and cancel the overall m_Q^2 in front of the amplitude (17). In other terms the heavy quark mass dependence is confined in $I(a)$.

For a light quark, $m_Q \ll m_H$,

$$I(a) \xrightarrow{a \rightarrow \infty} -\frac{1}{2a} \log^2 a = -\frac{m_Q^2}{2m_H^2} \log^2 \frac{m_Q^2}{m_H^2}, \quad (19)$$

showing that in the Standard Model the charm and bottom quark contributions are strongly suppressed by the square of the quark mass over Higgs mass ratio and come with a minus sign (with respect to the top-quark one).

The opposite limit, $m_H \ll m_Q$,

$$I(a) \xrightarrow{a \rightarrow 0} \frac{1}{3}, \quad (20)$$

which is found to be an extremely good approximation even for $m_Q \sim m_H$, is quite surprising at first. In this case the amplitude reads

$$\mathcal{A}(gg \rightarrow H) \xrightarrow{m_Q \gg m_H} -\frac{\alpha_S}{3\pi v} \delta^{ab} \left(g^{\mu\nu} \frac{m_H^2}{2} - p^\nu q^\mu \right) \epsilon_\mu(p) \epsilon_\nu(q). \quad (21)$$

i.e., the amplitude $gg \rightarrow H$ becomes *independent* of the mass of the heavy fermion in the loop. This is a special case of a general low energy theorem (which holds in the $p_H \rightarrow 0$ limit) that states that if the colored particle mass, independently of the other quantum numbers such as its spin acquires (all of) its mass via the Higgs mechanism, it will contribute to the amplitude $gg \rightarrow H$ independently of its mass. In other words $gg \rightarrow H$ acts as a counter of heavy colored particles. In a four generation scenario, for instance, the contribution from the t' and b' would lead to a factor of three increase at the amplitude level, i.e. a factor 9 at the cross section level. Note that this is in an apparent contradiction with our intuition

that heavy particles should decouple and not affect the physics at lower energy. The heavy states would not decouple because of our assumption that their (whole) mass is due to electroweak symmetry breaking and the interaction with the Higgs. Another interesting case is that of SUSY, where down-type and up-type quarks can couple differently to the Higgs(es) and other colored states (squarks) are present in the spectrum. At large $\tan \beta$, i.e. when $m_b \tan \beta \simeq m_t$, the Higgs bottom couplings are enhanced by a factor $\tan \beta$, while those of the top suppressed by a $\cot \beta$. However, the scaling with masses is different in the two limits and the contribution from the bottom anyway suppressed by m_Q/m_H . In addition, the the two contributions will have an opposite sign so that will actually interfere destructively in the amplitude squared. What about the squark contributions? Being heavy scalars and therefore coming with an opposite sign shouldn't the stop cancel exactly the contributions from the top and the others squarks give the dominant contribution? In this case, one has to remember that in (possibly) realistic SUSY models the mass of a squark has two sources: one from the coupling to the Higgs vev, which due to SUSY, it is exactly equal to the SM partner coupling and the other from the SUSY soft-breaking terms. For light quarks the latter are by far dominant giving a scaling for $\mathcal{A}$ of the type $m_q/m_{\tilde{q}}$, so highly suppressed and decoupling. A light stop instead, $m_{\tilde{t}} \simeq m_t$ could lead to a possibly strong suppression of $\mathcal{A}$.

3.2 Total cross section at the LHC at LO

The result can be written as:

$$\sigma^{\text{LO}}(pp \rightarrow H + X) = \frac{\alpha_S^2(\mu_R)}{64\pi v^2} \left| I\left(\frac{m_H^2}{m_Q^2}\right) \right|^2 \tau_0 \int_{\log \sqrt{\tau_0}}^{-\log \sqrt{\tau_0}} dy g(\sqrt{\tau_0} e^y, \mu_F) g(\sqrt{\tau_0} e^{-y}, \mu_F) \quad (22)$$

Using LO PDF's available in public libraries, such as LHAPDF [17] one can easily compute the gluon-gluon luminosity and therefore the LO Higgs cross section at the LHC14, see Fig. 2. An example is given in a Mathematica® notebook that can be found at the web address mentioned at the end of the Introduction. An interesting exercise is to vary the value of the renormalisation and factorisation scales around the natural central choice $\mu_R = \mu_F = m_H$ to try to estimate the unknown higher-orders terms in the perturbative expansion. It has to be noted that at LO, the cross section depends on μ_R only through $\alpha_S(\mu_R)$ which appears in the short distance coefficient and therefore as an overall factor α_S^2 , and depends on μ_F only via the PDF's (both dependences are of logarithmic nature, as the application of the renormalisation group equations easily shows). In other words the dependence on the scales is maximal as there is no explicit dependence on the log of the scales in the short distance coefficients that can compensate those in the coupling and in the PDF's. At this order, this is consistent as scale changes correspond to a change of at least one order in α_S more and in a LO computation only the first term in the perturbative expansion is present. The result of varying the scales independently $1/2 m_H < \mu_R, \mu_F < 2 m_H$ with $1/2 < \mu_F/\mu_R < 2$ in the LO predictions for the LHC is shown in Fig. 9 for different Higgs masses. Result are normalized to the central reference choice $\mu_R = \mu_F = m_H$.

4 Higgs Effective field theory

The main result of the simple calculation $gg \rightarrow H$ is that gluon fusion is basically independent of the heavy quark mass for a light Higgs boson. The result of Eq. (21) can be easily derived starting from the effective vertex,

$$\begin{aligned} \mathcal{L}_{\text{eff}} &= \frac{\alpha_S}{12\pi} G_{\mu\nu}^a G^{a\mu\nu} \left(\frac{H}{v} \right) \\ &= \frac{\beta_F}{g_s} G_{\mu\nu}^a G^{a\mu\nu} \left(\frac{H}{2v} \right) (1 - \delta), \end{aligned}$$

where

$$\beta_F = \frac{g_s^3 N_F}{24\pi^2} \quad (23)$$

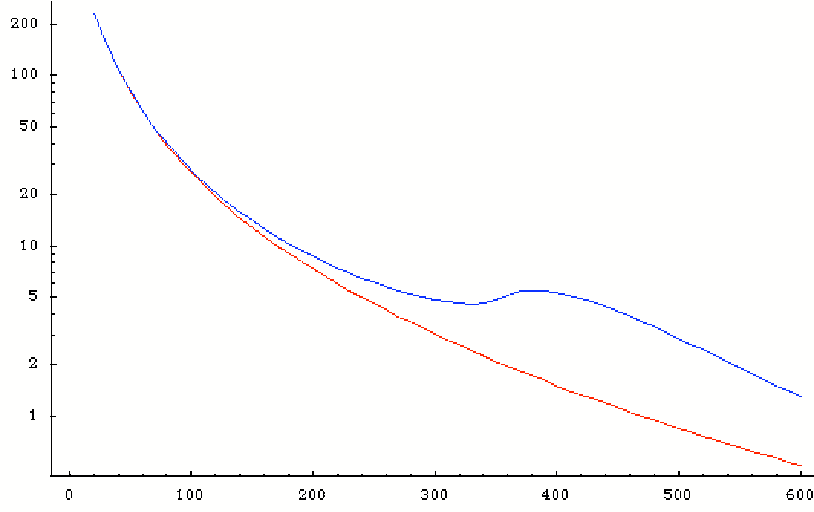


Fig. 2: Example of a plot for the LO cross section for $pp \rightarrow H$ at the LHC14 (pb) as a function of the Higgs mass (GeV) obtained with Mathematica® notebook available from the author (link in the text). The red (lower) curve is the large top-mass limit, while the blue (upper) curve is the result with full top-mass dependence.

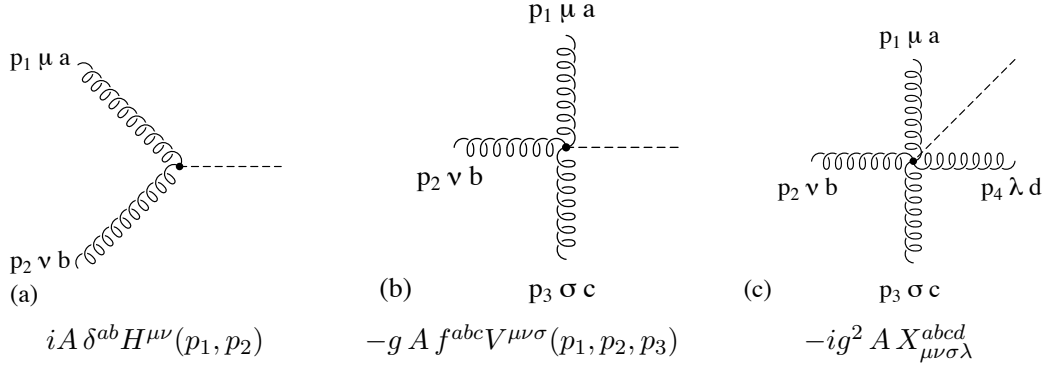


Fig. 3: Feynman rules in the EFT where the top quark is integrated out. Gluon momenta are outgoing.

is the contribution of heavy fermion loops to the $SU(3)$ beta function and $\delta = 2\alpha_S/\pi$.⁶ (N_F is the number of heavy fermions with $m \gg m_H$.) The effective Lagrangian of Eq. (23) gives ggH , $gggH$ and $ggggH$ vertices and can be used to compute the radiative corrections of $\mathcal{O}(\alpha_S^3)$ to gluon production. The correction in principle involve 2-loop diagrams. However, using the effective vertices from Eq. (23), the $\mathcal{O}(\alpha_S^3)$ corrections can be found from a 1-loop calculation. To fix the notation we shall use

$$\mathcal{L}_{\text{eff}} = -\frac{1}{4} A H G_{\mu\nu}^a G^{a,\mu\nu}, \quad (24)$$

⁶The $(1 - \delta)$ term arises from a subtlety in the use of the low energy theorem. Since the Higgs coupling to the heavy fermions is $M_f(1 + \frac{H}{v})\bar{f}f$, the counterterm for the Higgs Yukawa coupling is fixed in terms of the renormalisation of the fermion mass and wavefunction. The beta function, on the other hand, is evaluated at $q^2 = 0$. The $1 - \delta$ term corrects for this mismatch.

where $G_{\mu\nu}^a$ is the field strength of the SU(3) color gluon field and H is the Higgs-boson field. The effective coupling A is given by

$$A = \frac{\alpha_S}{3\pi v} \left(1 + \frac{11}{4} \frac{\alpha_S}{\pi} \right), \quad (25)$$

where v is the vacuum expectation value parameter, $v^2 = (G_F\sqrt{2})^{-1} = (246)^2 \text{ GeV}^2$ and the α_S correction is included, as discussed above. The effective Lagrangian generates vertices involving the Higgs boson and two, three or four gluons. The associated Feynman rules are displayed in Fig. 3. The two-gluon–Higgs-boson vertex is proportional to the tensor

$$H^{\mu\nu}(p_1, p_2) = g^{\mu\nu} p_1 \cdot p_2 - p_1^\nu p_2^\mu, \quad (26)$$

while the vertices involving three and four gluons and the Higgs boson are exactly proportional to their counterparts from pure QCD

$$V^{\mu\nu\rho}(p_1, p_2, p_3) = (p_1 - p_2)^\rho g^{\mu\nu} + (p_2 - p_3)^\mu g^{\nu\rho} + (p_3 - p_1)^\nu g^{\rho\mu}, \quad (27)$$

and

$$\begin{aligned} X_{abcd}^{\mu\nu\rho\sigma} &= f_{abe}f_{cde}(g^{\mu\rho}g^{\nu\sigma} - g^{\mu\sigma}g^{\nu\rho}) + f_{ace}f_{bde}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\sigma}g^{\nu\rho}) \\ &+ f_{ade}f_{bce}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\rho}g^{\nu\sigma}). \end{aligned} \quad (28)$$

5 $gg \rightarrow \text{Higgs}$ @ NLO

The HEFT is clearly a very powerful approximation as it turns a loop computation into a tree-level one. That means that within the HEFT the calculation of the total cross section for Higgs production at NLO will appear as a usual NLO calculation, i.e., involving only one-loop and tree-level diagrams. This is what we describe in this section.

5.1 The NLO computation in a nutshell

At NLO Eq. 3 can be rewritten as

$$\begin{aligned} \sigma^{\text{NLO}}(H + X) &= \int_{\tau_0}^1 dx_1 \int_{\tau_0/x_1}^1 dx_2 f_g(x_1, \mu_F) f_g(x_2, \mu_F) [\hat{\sigma}_B^{(0)}(gg \rightarrow H) + \hat{\sigma}_V^{(1)}(gg \rightarrow H)] \\ &+ \sum_{ijk} \int_{\tau_0}^1 dx_1 \int_{\tau_0/x_1}^1 dx_2 f_i(x_1, \mu_F) f_j(x_2, \mu_F) \times \hat{\sigma}_R^{(1)}(ij \rightarrow H k), \end{aligned} \quad (29)$$

where $\hat{\sigma}^{(0)}(gg \rightarrow H)$ and $\hat{\sigma}_V^{(1)}(gg \rightarrow H)$ denote the Born-level and the virtual cross sections, while $\hat{\sigma}_R^{(1)}(ij \rightarrow H k)$ is the real-emission cross section:

$$\begin{aligned} \hat{\sigma}_{B,V}^{(0,1)}(gg \rightarrow H) &= \frac{1}{2s} |\overline{\mathcal{A}_{B,V}}|^2 d\Phi_B, \\ \hat{\sigma}_R^{(1)}(ij \rightarrow H k) &= \frac{1}{2s} |\overline{\mathcal{A}_R}|^2 d\Phi_R, \end{aligned}$$

In general, the virtual term contains ultraviolet (UV), soft and collinear divergences. The UV divergences are absorbed by a universal redefinition of the couplings entering at the Born amplitude, as dictated by the renormalisation of the SM. When integrated over the full real phase space, the real term generates soft and collinear divergences, too, and only when *infrared(IR)-safe* quantities are computed, these divergences cancel to yield a finite result. IR-safe observables $O(\Phi)$ can be best understood by considering the soft or collinear limit in the real phase space, i.e. when the additional parton has low energy or is parallel to another parton. In this limit, an IR-safe observable yields $\lim O(\Phi_R) = O(\Phi_B)$, where the Born-level

configuration Φ_B is obtained from Φ_R by eliminating the soft particle (in case of soft singularities) or by merging the collinear particles (in case of collinear singularities).

There several ways to handle the cancellation of the singularities, which fall into two large categories, process-dependent and process-independent methods. In the former, one treats each calculation/process independently and performs manipulations of the integrals over the phase space so to obtain analytic or semi-analytic results.

Process independent methods, on the other hand, are based on a very fundamental result, i.e., that the pattern of the soft and collinear divergences is universal and depends only on the quantum numbers of the initial and final state particles in the Born process. That means that given the Born amplitude, one can predict the divergences that will show up in the virtual contributions and will be then cancelled over integration of the extra radiation in the reals. More importantly, such divergences come in just a handful of different types that can be dealt with once and for all.

Let us now rewrite Eq. (29) in a general and short-hand notation

$$\sigma^{\text{NLO}} \equiv \int d\Phi_B [B(\Phi_B) + V(\Phi_B)] O(\Phi_B) + \int d\Phi_R R(\Phi_R) O(\Phi_R) \quad (30)$$

which will be useful in the following. A NLO cross section is written in terms of matrix elements for the Born and virtual integrated over the Born phase space plus the real matrix elements integrated over the real phase space. Within a subtraction method, the real phase space is parametrized in terms of an underlying Born phase space Φ_B and a radiation phase space $\Phi_{R|B}$. A necessary requirement upon this parametrization is that, in the singular limits, by merging collinear partons, or eliminating the soft parton, the real phase becomes equal to the underlying Born one. Then the expectation value of an IR-safe observable reads

$$\begin{aligned} \int d\sigma^{(\text{NLO})} O(\Phi) &= \int d\Phi_B \left[B(\Phi_B) + V(\Phi_B) + \int d\Phi_{R|B} S(\Phi_R) \right] O(\Phi_B) \\ &+ \int d\Phi_R [R(\Phi_R) O(\Phi_R) - S(\Phi_R) O(\Phi_B)] . \end{aligned} \quad (31)$$

The third member of the above equation is obtained by adding and subtracting the same quantity from the two terms of the second member. The terms $S(\Phi_{R|B})$ are the subtraction terms, which contain all soft and collinear singularities of the real-emission term. Using the universality of soft and collinear divergences, they are written in a factorised form as

$$S(\Phi_R) = B(\Phi_B) \otimes \tilde{S}(\Phi_{R|B}) , \quad (32)$$

where the $\tilde{S}(\Phi_{R|B})$ can be composed from universal, process-independent subtraction kernels with analytically known (divergent) integrals. These integral, when summed and added to the virtual term, yield a finite result. The second term of the last member of Eq. (31) is also finite if O is an IR-safe observable, since by construction S cancels all singularities in R in the soft and collinear regions. The most popular subtraction schemes currently used in public NLO codes are based on the dipole subtraction [18] and the so-called FKS scheme [19]. The case of $gg \rightarrow H$ at NLO is particularly simple as the Born amplitude is a $2 \rightarrow 1$ process. This means that the integration over phase space of the real corrections is particularly simple and can therefore be done analytically. This has also the pedagogical advantage that shows explicitly where the divergences come from and to “see” the cancellations term by term. We study the process $gg \rightarrow H$ at NLO, in the large top-quark mass limit. All results given below are in Conventional Dimensional Regularization (CDR), where matrix elements are calculated in d dimensions, including the Born and real contributions, as well as the integration over phase space [6].

5.2 $gg \rightarrow H$: Born in d dimensions

The Born amplitude is calculated via the HEFT feynman rules. The only difference with respect to the previous calculation stems from the fact that now the computation has to be done in $d = 4 - 2\epsilon$ -

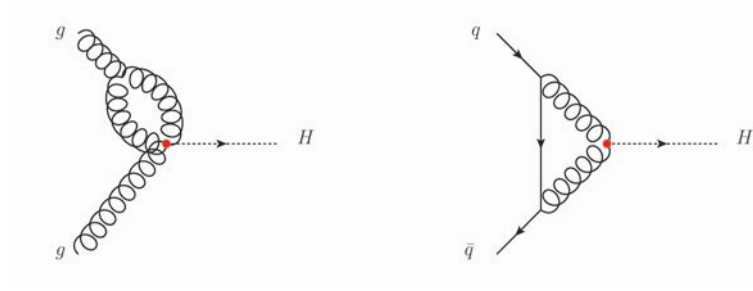


Fig. 4: Example of Feynman diagrams giving null contributions to $ij \rightarrow H$ at one-loop in the HEFT. Bubbles on the gluon legs are zero in dimensional regularisation. $q\bar{q} \rightarrow H$ is zero at all orders in perturbation theory if $m_q = 0$ due to chiral symmetry.

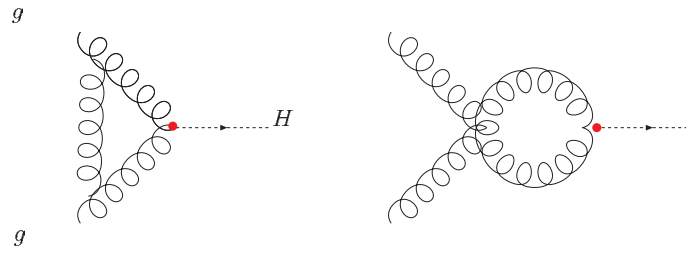


Fig. 5: Feynman diagrams giving non-zero contributions to $gg \rightarrow H$ at one-loop in the HEFT.

dimensions, with ϵ infinitesimal. The phase space does not bring any extra ϵ term. However, the matrix element changes

$$\left(g^{\mu\nu} \frac{m_H^2}{2} - p^\nu q^\mu \right)^2 = \frac{1}{4} (d-2) m_H^4, \quad (33)$$

as well as the average over the initial state gluon polarizations which in d -dimensions are $d-2$. This gives

$$\begin{aligned} \hat{\sigma}_B &= \frac{\alpha_S^2}{\pi} \frac{m_H^2}{576 v^2 s} \frac{\mu^{2\epsilon}}{(1-\epsilon)} \delta(1-z) \\ &\equiv \hat{\sigma}_0 \delta(1-z), \end{aligned} \quad (34)$$

where $z \equiv m_H^2/s$ is the inelasticity of the process, i.e. the fraction of the parton parton energy that goes into the Higgs (for the Born $z = 1$). μ is the usual arbitrary scale that needs to be introduced in dimensional regularisation to correct for the different dimensions and keep the action adimensional ($\hbar = c = 1$). Note that a cross section in d dimensions has dimensions $[\sigma] = M^{2-d}$. Also note that we have defined $\hat{\sigma}_0$ as containing an explicit factor z .

5.3 $gg \rightarrow H$: virtual corrections

There are several diagrams appearing at one-loop. Diagrams involving bubbles on the external gluon legs (with 3-point gluon-gluon-gluon and gluon-gluon-Higgs vertices) give rise to scaleless integrals that are zero in dimensional regularisation, see Fig. 4, left diagram. The $q\bar{q} \rightarrow H$ process, see Fig. 4 right, is proportional to the m_q parton mass which are taken massless and therefore null at all orders. As a result, only two diagrams are non-zero, i.e., the vertex correction and the bubble with the four gluon vertex as

shown in Fig. 5

$$\hat{\sigma}_{\text{tri}} = \hat{\sigma}_0 \delta(1-z) \left[1 + \frac{\alpha_S}{2\pi} C_A \left(\frac{\mu^2}{m_H^2} \right)^\epsilon c_\Gamma \left(-\frac{2}{\epsilon^2} + \frac{10}{3\epsilon} + \frac{179}{36} + \pi^2 \right) \right], \quad (35)$$

$$\hat{\sigma}_{\text{bub}} = \hat{\sigma}_0 \delta(1-z) \left[1 + \frac{\alpha_S}{2\pi} C_A \left(\frac{\mu^2}{m_H^2} \right)^\epsilon c_\Gamma \left(-\frac{10}{3\epsilon} - \frac{179}{36} \right) \right], \quad (36)$$

where

$$c_\Gamma = (4\pi)^\epsilon \frac{\Gamma(1+\epsilon)\Gamma(1-\epsilon)^2}{\Gamma(1-2\epsilon)}. \quad (37)$$

To obtain the results above, one has to write down the loop amplitudes, perform a few simplifications and the decomposition of the tensor integrals appearing in the amplitudes so to express the results in terms of the following two scalar integrals:

$$\begin{aligned} \mu^{2\epsilon} \int \frac{d^d \ell}{(2\pi)^d} \frac{1}{\ell^2(\ell + p_H)^2} &= c_\Gamma \left(\frac{\mu^2}{m_H^2} \right)^\epsilon \left(\frac{1}{\epsilon} + 2 \right), \\ \mu^{2\epsilon} \int \frac{d^d \ell}{(2\pi)^d} \frac{1}{\ell^2(\ell + p_1)^2(\ell + p_2)^2} &= \frac{c_\Gamma}{2m_H^2} \left(\frac{\mu^2}{m_H^2} \right)^\epsilon \left(\frac{2}{\epsilon^2} - \pi^2 \right), \end{aligned} \quad (38)$$

with $p_H = p_1 + p_2$. Summing the contributions of the two diagrams above with the α_S correction from Eq. (25), we obtain

$$\hat{\sigma}_V = \hat{\sigma}_0 \delta(1-z) \left[1 + \frac{\alpha_S}{2\pi} C_A \left(\frac{\mu^2}{m_H^2} \right)^\epsilon c_\Gamma \left(-\frac{2}{\epsilon^2} + \frac{11}{3} + \pi^2 \right) \right], \quad (39)$$

i.e., the total virtual contribution is proportional to the Born amplitude and it contains pole(s) in powers of $1/\epsilon$. The fact that the full virtual amplitude is proportional to the Born is due to the simplicity of a $2 \rightarrow 1$ process. However, in general one can prove that the divergent contributions must be proportional to the Born in the case of collinear (and collinear-soft, the double pole) divergences and to the so-called color-connected Born for the soft ones. Given that the Born amplitude is proportional to α_S^2 and we are calculating QCD corrections, we also expect UV divergences, which are proportional to $1/\epsilon$. The fact that apparently we do not see any pole in $1/\epsilon$ in the result above, it simply means that there is an accidental cancellation between simple poles of IR origin and that of UV origin, as we did not keep them distinct in the calculation. To leave only IR poles in the amplitude to be cancelled with those coming from the real contribution, we therefore proceed here to renormalisation of α_S . This can be attained by the substitution in $\hat{\sigma}_0$, see also [QCD:1.2.3],

$$\alpha_S \rightarrow \alpha_S^{\overline{\text{MS}}}(\mu_R) = \alpha_S \left[1 - \frac{\alpha_S}{2\pi} c_\Gamma \left(\frac{\mu^2}{\mu_R^2} \right)^\epsilon \frac{b_0}{\epsilon} \right], \quad (40)$$

where $b_0 = 11/6 C_A - 2n_f T_F/3$. The UV-renormalized virtual amplitude is

$$\hat{\sigma}_V^{\overline{\text{MS}}}(gg) = \hat{\sigma}_0 \delta(1-z) \left[1 + \frac{\alpha_S}{2\pi} C_A \left(\frac{\mu^2}{m_H^2} \right)^\epsilon c_\Gamma \left(-\frac{2}{\epsilon^2} - \frac{2}{\epsilon} \frac{b_0}{C_A} - 2 \frac{b_0}{C_A} \log \frac{m_H^2}{\mu_R^2} + \frac{11}{3} + \pi^2 \right) \right]. \quad (41)$$

where now the poles in $1/\epsilon^2$, $1/\epsilon$ are only of IR nature. Another important feature which is manifest in the expression above is the appearance of an explicit log of the renormalisation scale in the short distance part. As mentioned before, this is the improvement expected on the scale dependence of a NLO result: the μ_R dependence of the $\alpha_S^2(\mu_R)$ overall coefficient is exactly cancelled by the explicit log up to order α_S^3 .

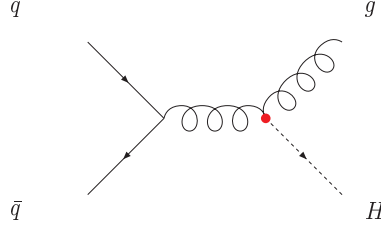


Fig. 6: Feynman diagrams giving $q\bar{q}$ real contributions in the infinite top-quark mass limit. These contributions are finite.

5.4 Real Contributions

Real corrections imply the calculation of $2 \rightarrow 2$ tree-level amplitudes and their integration over phase space in d dimensions. All possible initial and final state partons, gluons, quarks and anti-quarks need to be included,

1. $q\bar{q} \rightarrow Hg$ + crossing (i.e., $\bar{q}q \rightarrow Hg$),
2. $qg \rightarrow Hq$ + crossings (i.e., $\bar{q}g \rightarrow H\bar{q}$, $gq \rightarrow Hq$, $g\bar{q} \rightarrow H\bar{q}$),
3. $gg \rightarrow Hg$.

It is easy to predict which divergences to expect from each of the subprocesses above. The reason is that out of the possible (by Lorentz and color invariance) underlying Born amplitudes, i.e., $q\bar{q} \rightarrow H$ and $gg \rightarrow H$, the only non-zero one is $gg \rightarrow H$. Therefore the first processes must give a finite result when integrated over phase space, the second ones can only contain collinear divergences to be absorbed in quark PDF's, while the last is expected to give rise to soft and collinear divergences, part of which will be absorbed in the gluon PDF's and the rest canceled against those coming from the virtual contributions, Eq. (41).

5.4.1 $q\bar{q} \rightarrow Hg$

This contribution, shown in Fig. 6 is finite and can be calculated directly in four dimensions. A simple calculation gives

$$|\overline{\mathcal{M}}|^2 = \frac{4}{81} \frac{\alpha_S^3}{\pi v^2} \frac{(u^2 + t^2)}{s}, \quad (42)$$

to be integrated over the 4-dimensional phase space

$$d\Phi_2 = \frac{1}{8\pi} (1 - z) dv, \quad (43)$$

where $v = 1/2(1 + \cos \theta)$ and $z = m_H^2/s$ as usual. Using

$$t = -s(1 - z)(1 - v), \quad (44)$$

$$u = -s(1 - z)v, \quad (45)$$

gives

$$\hat{\sigma}_R(q\bar{q}) = \hat{\sigma}_0 \frac{\alpha_S}{2\pi} \frac{64}{27} \frac{(1 - z)^3}{z}. \quad (46)$$

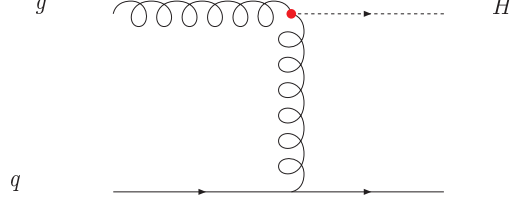


Fig. 7: Feynman diagrams giving qq real contributions in the infinite top-quark mass limit.

5.4.2 $gq \rightarrow Hq$

Let us consider now the contribution from the diagrams with an initial quark, i.e., the process $gq \rightarrow Hq$. The d -dimensional averaged/summed over initial/final state polarizations and colors amplitude is

$$|\overline{\mathcal{M}}|^2 = -\frac{1}{54(1-\epsilon)} \frac{\alpha_S^3}{\pi v^2} \frac{(u^2 + s^2) - \epsilon(u+s)^2}{t}. \quad (47)$$

Integrating it over the d -dimensional phase space

$$d\Phi_2 = \frac{1}{8\pi} \left(\frac{4\pi}{s} \right)^\epsilon \frac{1}{\Gamma(1-\epsilon)} z^\epsilon (1-z)^{1-2\epsilon} v^{-\epsilon} (1-v)^{-\epsilon} dv \quad (48)$$

one gets

$$\hat{\sigma}_R(gq) = \hat{\sigma}_0 \frac{\alpha_S}{2\pi} C_F \left(\frac{\mu^2}{m_H^2} \right)^\epsilon c_\Gamma \left[-\frac{1}{\epsilon} p_{gq}(z) + z - \frac{3}{2} \frac{(1-z)^2}{z} + p_{gq}(z) \log \frac{(1-z)^2}{z} \right], \quad (49)$$

where the $p_{gq}(z)$ color-stripped Altarelli-Parisi splitting function is given in the Appendix, Eqs. (67). We perform the factorisation of the collinear divergences adding the counterterm

$$\sigma_{\text{c.t.}}^{\text{coll.}}(gq) = \sigma_0 \frac{\alpha_S}{2\pi} \left[\left(\frac{\mu^2}{\mu_F^2} \right)^\epsilon \frac{c_\Gamma}{\epsilon} P_{gq}(z) \right]. \quad (50)$$

We note that in fact in CDR the cross section factorises over the d -dimensional splitting functions Eqs. (68). However, the collinear counter-term in $\overline{\text{MS}}$ is defined with the 4-dimensional Altarelli-Parisi splitting functions, Eqs. (67), and that is why we have written the result above in terms of $p_{gq}(z)$ leaving out a finite term z (also note that our definition of σ_0 , Eq. (34), contains a factor z). This gives

$$\begin{aligned} \hat{\sigma}_R^{\overline{\text{MS}}}(gq) &= \hat{\sigma}_R(gq) + \hat{\sigma}_{\text{c.t.}}^{\text{coll.}}(gq) \\ &= \sigma_0 \frac{\alpha_S}{2\pi} C_F \left[p_{gq}(z) \log \frac{m_H^2}{\mu_F^2} + p_{gq}(z) \log \frac{(1-z)^2}{z} + z - \frac{3}{2} \frac{(1-z)^2}{z} \right]. \end{aligned} \quad (51)$$

5.4.3 $gg \rightarrow Hg$

The calculation of the d -dimensional $gg \rightarrow Hg$ amplitude involves the four diagrams shown in Fig. 8 and it is not so trivial to do by hand, yet the final result is very compact:

$$|\overline{\mathcal{M}}|^2 = \frac{1}{24(1-\epsilon)^2} \frac{\alpha_S^3}{\pi v^2} \frac{(m_H^8 + s^4 + t^4 + u^4)(1-2\epsilon) + \frac{1}{2}\epsilon(m_H^4 + s^2 + t^2 + u^2)^2}{stu}. \quad (52)$$

This example is illustrative of the fact that keeping track of the ϵ parts in the amplitude squared makes the calculation significantly more complex for at least two reasons. First the structure of the result itself is more involved. Second, one is forced to work at the squared amplitude level as d dimensional

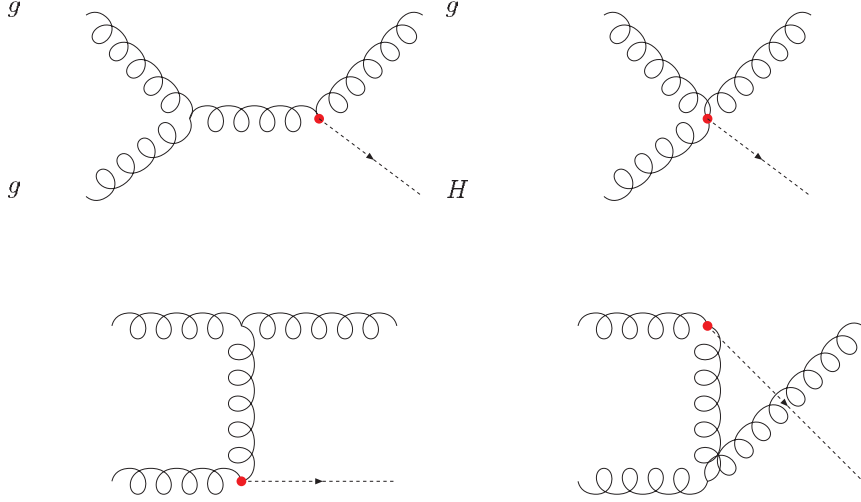


Fig. 8: Feynman diagrams giving gg real contributions in the infinite top-quark mass limit.

contributions come from the $(d - 2)$ dimensional gluon polarizations and therefore cannot exploit the beauty, power and simplicity of helicity amplitude techniques [20, 21]. Computing QCD amplitudes where states have fixed polarizations entails huge simplifications and allows to make predictions for amplitudes with many external partons. For example, tree-level amplitudes in the HEFT involving up to 5 extra partons can be easily obtained automatically using tools such as ALPGEN [22] or MADGRAPH [23]. Fortunately, it turns out that is possible to use a different scheme than CDR and actually perform the computation of the Born and real matrix elements in exactly four dimensions (yet integrate them over the d -dimensional phase space). This involves a different (and a bit tricky) d -dimensional algebra for the loop computations and the introduction of (universal) finite terms for the initial-state counter-terms and UV subtractions, yet with an enormous computational simplification. All public NLO codes for processes at the LHC in practice do use such "maximally four dimensional" d -dimensional regularisation schemes. Integrating the amplitude (52) over the d -dimensional phase space of Eq. (48) gives

$$\begin{aligned} \hat{\sigma}_R(gg) = & \hat{\sigma}_0 \frac{\alpha_S}{2\pi} C_A \left(\frac{\mu^2}{m_H^2} \right)^\epsilon c_\Gamma \left[\left(\frac{2}{\epsilon^2} + \frac{2}{\epsilon} \frac{b_0}{C_A} - \frac{\pi^2}{3} \right) \delta(1-z) \right. \\ & - \frac{2}{\epsilon} p_{gg}(z) - \frac{11}{3} \frac{(1-z)^3}{z} - 4 \frac{(1-z)^2(1+z^2) + z^2}{z(1-z)} \log z \\ & \left. + 4 \frac{1+z^4 + (1-z)^4}{z} \left(\frac{\log(1-z)}{1-z} \right)_+ \right], \end{aligned} \quad (53)$$

where the plus prescription is defined as follows:

$$\int_0^1 dx [h(x)]_+ f(x) = \int_0^1 dx h(x) [f(x) - f(1)]. \quad (54)$$

Note that the $\frac{2}{\epsilon} \frac{b_0}{C_A} \delta(1-z)$ in Eq. (53) comes from reexpressing the divergent term $-\frac{4}{\epsilon} [\frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z)]$ in terms of $-\frac{2}{\epsilon} p_{gg}(z)$, see Eq. (67). The factorisation of the collinear divergence is handled by adding the corresponding counterterm

$$\hat{\sigma}_{\text{c.t.}}^{\text{coll.}}(gg) = 2 \hat{\sigma}_0 \frac{\alpha_S}{2\pi} \left[\left(\frac{\mu^2}{\mu_F^2} \right)^\epsilon \frac{c_\Gamma}{\epsilon} P_{gg}(z) \right], \quad (55)$$

which gives

$$\begin{aligned}
\hat{\sigma}_R^{\overline{\text{MS}}}(gg) &= \hat{\sigma}_R(gg) + \hat{\sigma}_{\text{c.t.}}^{\text{coll.}}(gg) \\
&= \hat{\sigma}_0 \frac{\alpha_S}{2\pi} C_A \left(\frac{\mu^2}{m_H^2} \right)^\epsilon c_\Gamma \left[\left(\frac{2}{\epsilon^2} + \frac{2}{\epsilon} \frac{b_0}{C_A} - \frac{\pi^2}{3} \right) \delta(1-z) \right. \\
&\quad + 2p_{gg} \log \frac{m_H^2}{\mu_F^2} - \frac{11}{3} \frac{(1-z)^3}{z} - 4 \frac{(1-z)^2(1+z^2) + z^2}{z(1-z)} \log z \\
&\quad \left. + 4 \frac{1+z^4 + (1-z)^4}{z} \left(\frac{\log(1-z)}{1-z} \right)_+ \right]. \tag{56}
\end{aligned}$$

We can now recognise that the IR poles match those of the virtual contributions in Eq. (41). Adding up the contributions from real and virtual contributions of the gg channel we obtain (note that our definition of σ_0 , Eq. (34), contains a factor z):

$$\begin{aligned}
\hat{\sigma}^{\overline{\text{MS}}}(gg) &= \hat{\sigma}_R^{\overline{\text{MS}}}(gg) + \hat{\sigma}_V^{\overline{\text{MS}}}(gg) \\
&= \sigma_0 \frac{\alpha_S}{2\pi} C_A \left[\left(\frac{11}{3} + \frac{2}{3} \pi^2 - 2 \frac{b_0}{C_A} \log \frac{m_H^2}{\mu_R^2} \right) \delta(1-z) \right. \\
&\quad - \frac{11}{3} \frac{(1-z)^3}{z} + 2p_{gg} \log \frac{m_H^2}{\mu_F^2} - 4 \frac{(1-z+z^2)^2}{z(1-z)} \log z \\
&\quad \left. + 8 \frac{(1-z+z^2)^2}{z} \left(\frac{\log(1-z)}{1-z} \right)_+ \right]. \tag{57}
\end{aligned}$$

As predicted, the final results for the short distance coefficients is finite (yet scheme dependent) and does contain the necessary $\log$'s of the renormalisation and factorisation scales that compensate up to α_S^3 the corresponding dependences in $\alpha_S^2(\mu_R)$ of the Born amplitude and in the PDF's.

5.5 NLO results: discussion

The expressions above can be easily implemented in a numerical code to perform the convolution integrals with PDF's. A few simple numerical optimizations, such as the choice of integration variables, and a bit of attention to the implementation of the $+$ distributions, that's all is needed. The reader can find a sample implementation in a Mathematica® notebook at the web address mentioned at the end of the Introduction. By running the code with different scale choices, one can associate an uncertainty to the NLO predictions as done at LO. The result, shown in Fig. 9, comes as a big surprise! The NLO calculation predicts a rate twice as large and the respective LO and NLO uncertainty bands do not even overlap. That means that our naive estimate of the uncertainties at LO is totally off and therefore unreliable. It seems also to suggest that perturbation expansion is at stake here. As we had mentioned, this motivated the computation of the NNLO corrections, which are also shown in Fig. 9. Fortunately, NNLO predictions do overlap with NLO and also display a smaller scale dependence, so that the perturbation picture seems safe starting from NLO on. In fact, this behavior is rather special to $pp \rightarrow H + X$ and it is often rephrased by saying that what we call LO (in the perturbative expansion) is not actually the leading one in size and therefore we should not start from that. For instance, in Drell-Yan or VBF this does not happen, and the perturbative expansions (seem to) converge beautifully, see Fig. 10. In any case, the Higgs production reminds us an important fact that we should always keep in mind: scale variation cannot by definition reproduce missing finite terms in the perturbative expansion and as such can only give an indication of what the real uncertainties could be. On the other hand, comparison between predictions from LO and NNLO, their stabilization (or lack thereof) and the use of approximate methods to determine (classes of) higher order terms, all together can provide a rather solid picture on the theoretical uncertainties on a case-by-case basis. We mention, in passing, another important source of uncertainties in making predictions for hadron colliders, i.e., that coming from imperfect knowledge of

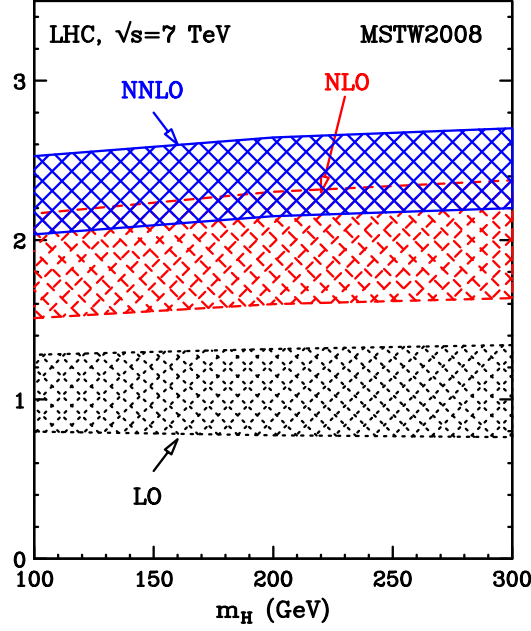


Fig. 9: K-factors for Higgs production from gluon fusion at the LHC. Uncertainty bands are obtained via independent scale variation $1/2 m_H < \mu_R, \mu_F < 2 m_H$ with $1/2 < \mu_F/\mu_R < 2$. The LO and NLO bands can be obtained by implementing the formulas obtained in these notes in a code that performs the numerical integration over the PDF's. Cross-checks and NNLO results can be obtained with HNNLO [24]. (Plot courtesy of M. Grazzini).

the PDF's. Uncertainties are related to unknown higher-order terms in the DGLAP evolution equations that determine as well as from the extraction of the initial condition from experimental data, see [QCD:3] and in particular [QCD:3.3.2].⁷

As far as total cross sections are concerned, the situation is therefore pretty clear. Fixed-order calculations come equipped with self-detecting procedures that can give us information on whether a prediction is reliable or not. If not, it can be systematically improved by including higher-order terms (almost for free nowadays at NLO, yet at a rather high cost at NNLO) and uncertainties can be easily estimated. So it is natural to ask, what about other IR-safe observables?

Let us consider, once again $pp \rightarrow H + X$ as an example, and focus on the Higgs momentum (fully inclusive) distribution, which can be parametrized in terms of only two variables⁸, the rapidity y_H and the transverse momentum p_H^T . At LO (referred to the total cross section), the Higgs can be boosted in the forward or backward directions in the lab system, $y_H = \frac{1}{2} \log \frac{x_1}{x_2}$, yet it has always $p_H^T = 0$, i.e. the distribution in p_H^T is a delta function centered at $p_H^T = 0$. At NLO (again referred to the total cross section), $2 \rightarrow 2$ diagrams enter in the calculation and the Higgs can have a non-zero p_H^T . Since at any point in phase space with $p_H^T \neq 0$ this is the first non-zero contribution, the observable p_H^T of the Higgs is only at LO. In other words if we want to know the p_H^T distribution of the Higgs at NLO over all phase space, we need at least a NNLO prediction for the cross section. Another way of thinking about it is to ask oneself what kind of diagrams are present in the calculation for that observable in a given area of the phase space: if there are only tree-level diagrams then the observable is LO. It is important when working with NLO codes to always think about what kind of observables are actually predicted at NLO, what at LO and what not even at LO. Again, a NNLO computation for the total cross section for $pp \rightarrow H + X$,

⁷The latter does in fact imply also the prediction of experimental observables at the same order in perturbation theory and therefore are also intrinsically also affected by scale dependencies. Such effects are not included normally in the estimation of the uncertainties coming from PDF's.

⁸We do not consider the azimuthal angle ϕ , because for symmetry reasons can only lead to a uniform distribution.

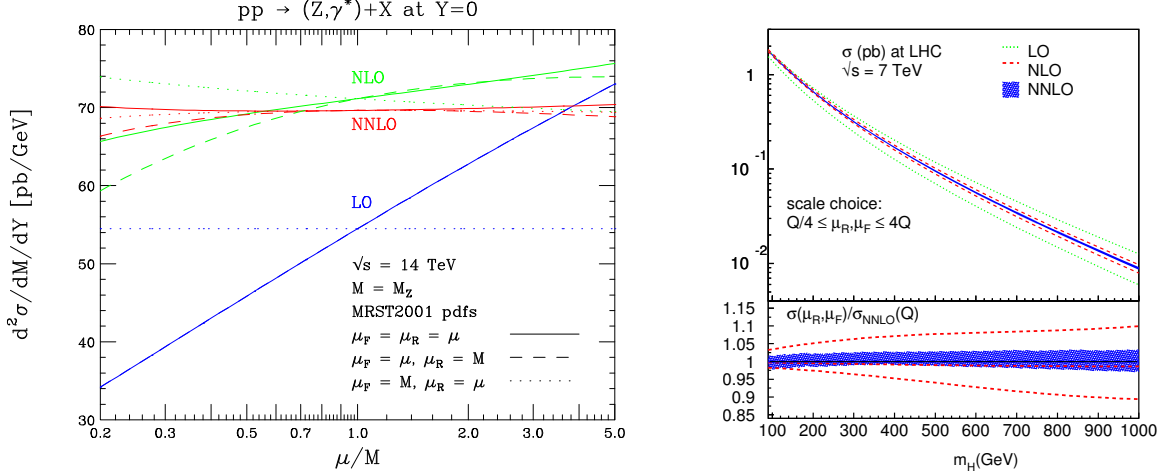


Fig. 10: Examples of improvement in the predictions of processes at LHC in going from LO to NNLO. On the left, scale dependence of the predictions for Z/γ^* production (at $y = 0$) at the LHC14, at fixed order [25]. On the right, Higgs production at the LHC7 via VBF [26] as a function of the Higgs mass. The bands are obtained by independent scale variation in the interval $Q/4 \leq \mu_F, \mu_R \leq 4Q$, Q being the virtuality of the W, Z fusing into the Higgs. In both cases the perturbative expansion behaves extremely well and NNLO predictions overlap with those at LO and NLO and display a much smaller residual uncertainty.

gives NNLO information on the Higgs rapidity distribution, NLO for the Higgs p_H^T and $pp \rightarrow H + 1$ -jet observables, LO for $pp \rightarrow H + 2$ -jets observables and the structure of the jet in $H + 1$ -jet events and no information at all on $pp \rightarrow H + 3$ -jets observables. In short, a fixed-order computation can only make predictions for a finite number of observables, typically with a rather limited number of resolved partons and a very small number of unresolved ones, i.e. just one for a NLO computation and up to two for a NNLO computation. This is the first main limitation of a fixed-order computation. However, it is not the only one.

Consider again the p_H^T distribution of the Higgs as predicted by a NLO computation for the total cross section, Fig. 11. This curve can be easily obtained using the expressions in four dimensions of Eqs. (42, 47, 52), performing the integration over the polar angle together with the PDF's via a Monte-Carlo method and plotting it point-by-point during the integration. The p_H^T distribution is divergent in $p_H^T = 0$ as expected from soft and beam-collinear emissions. As we have learnt such divergences are proportional to $\delta(1 - z)$ where z is the fraction of parton-parton energy taken by the Higgs and are cancelled by the virtual contributions, all of which reside in $p_T = 0$. So the cancellation between real and virtual contributions, all of it happens in the first bin of the histogram. How do we interpret such weird distribution? A useful way is to think about the size of the bin of the distribution as our resolution scale: with a rather coarse binning there is no "going-to-infinity" and predictions are rather stable (this of course includes the total cross section which corresponds to using only one bin), while with thin binning, we start to be sensitive to low energy and virtual emissions which become increasingly important and are not included at all in a fixed-order approach. This is the case where resummed predictions come into rescue: one finds that the leading part of soft emissions (real and virtual) is universal, it can be considered at all orders and included by identifying the log's associated to it and exponentiating them. This can be done either at very high accuracy analytically yet fully inclusively or in a numerical and exclusive way at the leading log with a parton shower (which actually resums both soft and collinear enhancements). The result of including these effects analytically is shown in Fig. 11, red curve. In very crude words, the effect of the resummation is to spread the $\delta(p_T)$ of the virtual contributions over a range of a few tens of GeV with the effect of smoothing out the divergence and producing a "physical" distribution.

In summary, fixed-order calculations in perturbative QCD can be performed in a well-defined

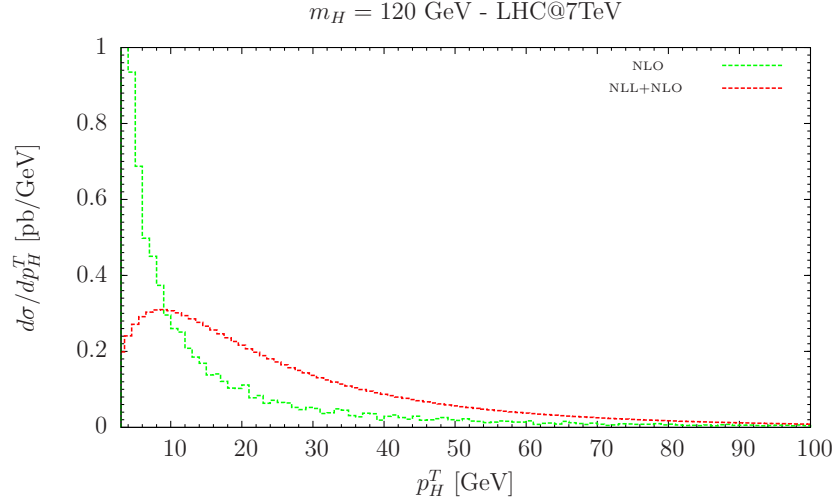


Fig. 11: Higgs p_H^T spectrum for a Higgs of $m_H = 120$ GeV at the LHC7. The labeling NLO and NLL+NLO refer to the total cross section. The curves are normalized to the same value (=total cross section is the same). The green curve is just a LO prediction for the p_H^T of the Higgs. The logarithmic divergence at $p_H^T \rightarrow 0$ is cancelled by the negative infinite virtual contributions at $p_H^T = 0$ (not shown!). The resummed prediction (red curve) features a “physical” smooth behavior at small p_H^T . (The resummed prediction is obtained via HqT [27]).

and quite simple framework, i.e. in the context of the factorization theorem. It is therefore possible to make predictions for inclusive quantities in hadron colliders, which can be systematically improved at the “only” price of an (exponential) increase in the complexity of the calculation. In practice, however, the use of fixed-order predictions is limited by several other important drawbacks. First, only processes with a few resolved partons can be calculated, while in practice we know that hundreds of hadrons can be produced in a single proton-proton interaction of which we are bound to ignore the details. Second, sharp infinities appear in the phase that do cancel between real and virtual contributions if inclusive enough observables are defined, yet lead to unphysical distributions in specific areas of the phase space and/or when the resolved partons become either soft or collinear. Such local positive and negative infinities are unphysical because they appear only due the artificial truncation of the perturbative expansion. Finally, the fact that plus and minus infinities appear locally in phase space also means that fixed order predictions beyond LO cannot be used as probability functions to generate events as distributed in nature. Parton showers, i.e. fully exclusive resummation, and their merging/matching with fixed-order predictions, provide an elegant and powerful way out to all the above limitations.

6 Beyond fixed-order predictions

As we have explicitly verified, fixed-order predictions have important limitations both of principle (areas of phase space and observables, such as jet substructure are poorly described, no hadrons but only partons) and in practice (no event simulation is possible). Fortunately, an alternative approach exists that is based on the fact that the IR structure, soft or collinear, of QCD is universal and contributions can be resummed at all orders. Last but not least, formulas that describe the emission of soft and collinear partons are amenable of a probabilistic interpretation and therefore not only it is possible to perform an explicit resummation but also to associate a full “history” to an hard scattering event, i.e., to associate to every event a full-fledged description of an high-energy event from the two initial protons to the final (possibly hundreds) of hadrons and leptons in the final state. In addition, in the latest years, enormous progress has been achieved in combining the accuracy of fixed-order predictions with the flexibility of

parton showers. These methods are briefly presented here together with their applications to Higgs production. The short presentation below is adapted from Ref. [28]. The reader is also referred to [QCD:4.4] for further details, examples and references.

6.1 Parton Showers

Parton Showers (PS) are able to dress a given Born process with all the dominant (i.e. enhanced by collinear logarithms, and to some extent also soft ones) QCD radiation processes at all orders in perturbation theory. In particular, the dominant contributions, i.e. those given by the leading logarithms, coming from both real and virtual emissions are included. The cross section for the first (which is often also the hardest) emission in a shower reads:

$$d\sigma^{\text{1st step}} = d\Phi_B B(\Phi_B) \left[\Delta(p_\perp^{\min}) + d\Phi_{R|B} \Delta(p_T(\Phi_{R|B})) P(\Phi_{R|B}) \right], \quad (58)$$

where $\Delta(p_T)$ denotes the Sudakov form factor

$$\Delta(p_T) = \exp \left[- \int d\Phi_{R|B} P(\Phi_{R|B}) \Theta(p_T(\Phi_R) - p_T) \right]. \quad (59)$$

This Sudakov form factor can be understood as a no-emission probability of secondary partons down to a resolution scale of p_T . Here $P(\Phi_{R|B})$ is a process-independent universal splitting function that allows to write the PS approximation to the real cross section R^{PS} , typically given schematically by a product of the underlying Born-level term folded with a splitting kernel P

$$R^{\text{PS}}(\Phi) = P(\Phi_{R|B}) B(\Phi_B). \quad (60)$$

In this framework, $\Phi_{R|B}$ is often expressed in terms of three showering variables, like the virtuality t in the splitting process, the energy fraction of the splitting z and the azimuth ϕ . A very simple (and widely used) choice for the splitting function, is

$$P(\Phi_{R|B}) d\Phi_{R|B} = \frac{\alpha_S(t)}{2\pi} P_{a \rightarrow bc}(z) \frac{d\phi}{2\pi} \frac{dt}{t} dz \quad (61)$$

where $P(z)$ are Altarelli-Parisi splitting functions on which any QCD amplitude factorises in the collinear limit $b \parallel c$.

The above definition of the Sudakov form factor, guarantees that the square bracket in Eq. (58) integrates to unity, a manifestation of the probabilistic nature of the parton shower. Thus, integrating the shower cross section over the radiation variables yields the total cross section, given at LO by the Born amplitude. The corresponding radiation pattern consists of two parts: one given by the first term in the square bracket, where no further resolvable emission above the parton-shower cut-off $p_\perp^{\min}$ – typically of the order of 1 GeV – emerges, and the other given by the second term in the square bracket describing the first emission, as determined by the splitting kernel. It is important to stress that the real-emission cross section in a PS generator is only correct in the small angle and/or soft limit, where R^{PS} is a reliable approximation of the complete matrix element.

After the 1st step the process is repeated using the new configuration as the Born one.

While rather crude, the PS approximation is a very powerful one, due mainly to the great flexibility and simplicity in the implementation of $2 \rightarrow 1$ and $2 \rightarrow 2$ high- Q^2 processes. In addition, once augmented with a hadronisation model the simulation can easily provide a full description of a collision in terms of physical final states, i.e., hadrons, leptons and photons. In the current terminology a generic Monte Carlo generator mainly refers to such tools, the most relevant examples of are PYTHIA 6 and PYTHIA 8 [29, 30], HERWIG [31], HERWIG++ [32], and SHERPA [33]. A very clear and exhaustive presentation of parton shower generators can be found in Ref. [34].

6.2 Matrix-element merging (ME+PS)

In parton showers algorithms QCD radiation is generated in the collinear and soft approximation, using Markov chain techniques based on Sudakov form factors. Hard and widely separated jets are thus poorly described in this approach. On the other hand, tree-level fixed order amplitudes can provide reliable predictions in the hard region, while failing in the collinear and soft limits. To combine both descriptions and avoid double counting or gaps between samples with different multiplicity, an appropriate merging method is required.

Matrix-element merging [35] aims at correcting as many large-angle emissions as possible with the corresponding tree-level accurate prediction, rather than only *small-angle* accurate. This is achieved by generating events up to a given (high) multiplicity using a matrix-element generator, with some internal jet-resolution parameter Q_{cut} on the jet separation, such that practically all emissions above this scale are described by corresponding tree-level matrix elements. Their contributions are corrected for running-coupling effects and by Sudakov form factors. Radiation below Q_{cut} on the other hand is generated by a parton-shower program, which is required to veto radiation with separation larger than Q_{cut} . As far as the hardest emission is concerned, matrix-element merging is as accurate as matrix-element corrections (when these are available) or NLO+PS. Since they lack NLO virtual corrections, however, they do not reach NLO accuracy for inclusive quantities. Nevertheless, they are capable to achieve leading-order accuracy for multiple hard radiation, beyond the hardest only, while NLO+PS programs, relying on the parton shower there are only accurate in the collinear and/or soft limit for these quantities.

Several merging schemes have been proposed, which include the CKKW scheme [35–37] and its improvements [38, 39], the MLM matching [40], and the k_T -MLM variation [41]. The MLM schemes have been implemented in several matrix element codes such as ALPGEN [22], MADGRAPH [23], through interfaces to PYTHIA/HERWIG, while SHERPA [33] and HERWIG++ [32] have adopted the CKKW schemes and rely on their own parton showers. In Ref. [42] a detailed, although somewhat outdated description of each method has been given and a comparative study has been performed.

6.3 NLO+PS in a nutshell

Several proposals have been made for the full inclusion of complete NLO effects in PS generators. At this moment, only two of them have reached a mature enough stage to be used in practice: MC@NLO [43] and POWHEG [44]. Both methods correct – in different ways – the real-emission matrix element to achieve an exact tree-level emission matrix element, even at large angle. As we have seen in the previous subsection, this is what is also achieved with matrix-element corrections in parton showers, at least for the simplest processes listed earlier. This, however, is not sufficient for the NLO accuracy, since the effect of virtual corrections also needs to be included. In both methods, the real-emission cross section is split into a singular and non-singular part, $R = R^s + R^f$. One then computes the total NLO inclusive cross section, excluding the finite contribution, at fixed underlying Born kinematics, defined as

$$\bar{B}^s = B(\Phi_B) + \left[V(\Phi_B) + \int d\Phi_{R|B} R^s(\Phi_{R|B}) \right], \quad (62)$$

and uses the formula

$$d\sigma^{\text{NLO+PS}} = d\Phi_B \bar{B}^s(\Phi_B) \left[\Delta^s(p_{\perp}^{\min}) + d\Phi_{R|B} \frac{R^s(\Phi_R)}{B(\Phi_B)} \Delta^s(p_T(\Phi)) \right] + d\Phi_R R^f(\Phi_R) \quad (63)$$

for the generation of the events. In this formula, the term $\bar{B}$ can be understood as a local K -factor reweighting the soft matrix-element correction part of the simulation. Clearly, employing the fact that the term in the first square bracket integrates to unity, as before, the cross section integrates to the full NLO cross section.

In MC@NLO one chooses R^s to be identically equal to the term $B \otimes P$ that the PS generator employs to generate emissions. Within MC@NLO, n -body events are obtained using the $\bar{B}^s$ function,

and then fed to the PS, which will generate the hardest emission according to Eq. (62). These are called $\mathcal{S}$ events in the MC@NLO language. An appropriate number of events are also generated according to the R^f cross section, and are directly passed to the PS generator. These are called $\mathcal{H}$ events. In MC@NLO, $R^f = R - R^s$ is not positive definite, and it is thus necessary to generate negative weighted events in this framework. A library of MC@NLO Higgs processes (gluon fusion, vector-boson associated production, and charged Higgs associated with top) is available at Ref. [45], which is interfaced to HERWIG and HERWIG++. A fully automatized approach, AMC@NLO [46] implemented in the MADGRAPH framework, is now available that allows to compute and combine all necessary ingredients (Born, real, virtual matrix elements plus counterterms) at the user's request.

In POWHEG, one chooses $R^s \leq R$, and in many cases even $R^s = R$, so that the finite cross section R^f vanishes. In this case, the hardest emission is generated within POWHEG itself, and the process is passed to the parton shower only after the hardest radiation is generated. Positive weighted events are obtained, since R^f can always be chosen to be positive definite. In all cases the chosen R^s has exactly the same singularity structure as R , so that R^f always yield a finite contribution to the cross section. Implementations of Higgs production processes with the POWHEG method are available in HERWIG++ [47], in the POWHEGBOX [48] (interfaced to both HERWIG and PYTHIA) and recently in SHERPA [49].

6.4 Improved descriptions of Higgs production

Being of primary importance, Higgs kinematic distributions are now quite well predicted and also available via public codes such as ResBos [50] and HqT [27, 51]. Differential p_H^T distributions accurate to LO yet featuring the exact bottom- and top-quarks mass loop dependence (and therefore can be used also for predictions of scalar Higgs in BSM) can be obtained via HIGLU [52] as well as via HPro [53]. However, in experimental analyses, it is also crucial to get as precise predictions as possible for exclusive observables that involve extra jets, such as the jet p_T spectra and the jet rates, at both parton and hadron level. To optimize the search strategies and in particular to curb the very large backgrounds, current analyses both at Tevatron and at the LHC select 0-, 1- and 2-jet events and perform independent analyses on each sample. The final systematic uncertainties are effected by both the theoretical and experimental ones of such a jet-bin based separation. In the HEFT, fully exclusive parton- and hadron-level calculations can now be performed by Parton Shower (PS) programs or with NLO QCD codes matched with parton showers: via the MC@NLO and POWHEG methods. Beyond the HEFT, fully exclusive predictions ME+PS and NLO+PS techniques has become available only recently [54, 55]. The reason is that one needs to compromise between the validity of HEFT and the complexity of higher loop calculations.

Figure 12 shows a comparison of the predictions of the p^T of the Higgs at LHC7 as obtained in HEFT from:

- a full analytical resummation at NNLL;
- MC@NLO (w/ PYTHIA);
- ME+PS merging (MADGRAPH+PYTHIA);
- POWHEG (w/ PYTHIA).

We first stress (again) that this observable which is at NLO at high- p^T only in the HqT predictions. The ME+PS approach is built to be LO for all observables, while MC@NLO and POWHEG predictions are based on the NLO calculation for the total cross section, the same performed in these notes. Notwithstanding we see that given the expected uncertainties which are quite large above all at high- p^T the shapes are in substantial agreement both in the low and high- p^T ranges. In Fig. 13 the p^T distributions for the first and second jets are shown comparing the ME+PS prediction based on the HEFT and one with the full top-mass dependence and PYTHIA. Even in this case the agreement between the various approaches is extremely good for a light Higgs. For a very heavy Higgs difference in the p_T distributions of the extra jets become visible at quite a high p^T , a region not very relevant phenomenologically.

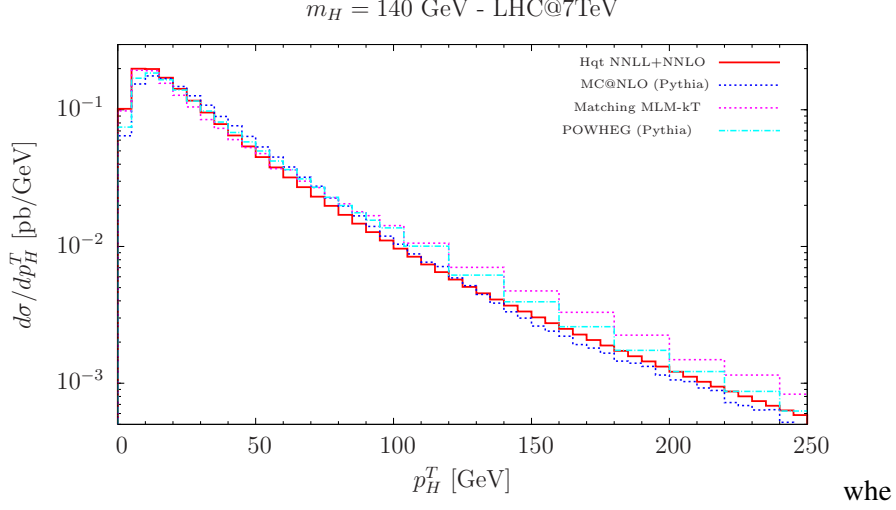


Fig. 12: Higgs p_T^H spectrum for a Higgs of $m_H = 140$ GeV as predicted by a series of improved predictions: NNLL+NNLO resummed (red solid), MC@NLO + Pythia (blue dashes), matrix-element + Pythia merged results (magenta dashes), POWHEG + Pythia (cyan dashes). All predictions display similar features, i.e. a peak between 10–20 GeV and a similar shape at high- p_T^H with differences that lie within their respective uncertainties (not shown).

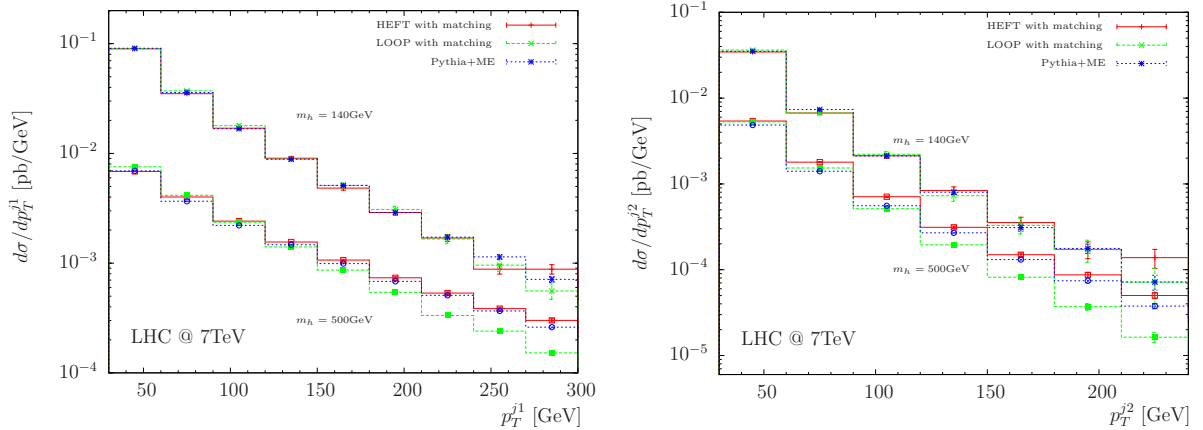


Fig. 13: Jet p_T distributions for associated jets in gluon fusion production of $m_H = 140$ GeV and $m_H = 500$ GeV Higgs bosons at 7 TeV LHC.

7 Conclusions

Progress in the field of QCD predictions for the LHC in the form of MC tools usable by both theorists and experimentalists has made tremendous progress in the last years. It is fair to say that we are now able (or close to be able in some specific very challenging cases) to compute automatically or semi-automatically any interesting cross section for Standard Model and Beyond processes at NLO accuracy and interface it with parton shower programs for event generation. In the LHC era the lowest acceptable accuracy for any serious phenomenological and experimental study is via an NLO event generator. LHC precision physics is now at NNLO in QCD and NLO in EW. Any physicist interested in making discoveries at the LHC needs to be familiar with the ideas, the physics and the reach of the current QCD simulation tools.

To this aim, we have considered $pp \rightarrow H + X$ as a case study. We have illustrated how accurate

and useful predictions for cross sections and other observables can be obtained in QCD, starting from the calculation of Born amplitude (at one loop) and the corresponding hadronic cross section. We have then considered Higgs production at NLO in the HEFT and discussed the limitations of fixed-order predictions. Finally, we have briefly discussed how fully exclusive predictions are obtained with modern tools, that allow to reach the accuracy of NLO predictions together with the full exclusivity of a parton shower approach.

Appendix

Splitting functions and collinear counterterms

We define the 4-dimensional splitting functions as in (4.94) of the ESW book:

$$P_{qq}(z) = C_F p_{qq}(z) = C_F \left[\frac{1+z^2}{(1-z)_+} + \frac{3}{2} \delta(1-z) \right] \quad (64)$$

$$P_{qg}(z) = T_R p_{qg}(z) = T_R [z^2 + (1-z)^2] \quad (65)$$

$$P_{gq}(z) = C_F p_{gq}(z) = C_F \left[\frac{1+(1-z)^2}{z} \right] \quad (66)$$

$$P_{gg}(z) = C_A p_{gg}(z) = 2C_A \left[\frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z) \right] + b_0 \delta(1-z), \quad (67)$$

where $b_0 = 11/6 C_A - 2n_f T_F/3$. We also define the following quantities as the extension of the splitting functions in d -dimensions:

$$P_{ij}^d(z) = P_{ij}(z) + \epsilon P_{ij}^\epsilon(z) \quad (68)$$

where

$$P_{qq}^\epsilon(z) = C_F p_{qq}^\epsilon(z) = -C_F(1-z) \quad (69)$$

$$P_{qg}^\epsilon(z) = T_R p_{qg}^\epsilon(z) = -T_R 2z(1-z) \quad (70)$$

$$P_{gq}^\epsilon(z) = C_F p_{gq}^\epsilon(z) = -C_F z \quad (71)$$

$$P_{gg}^\epsilon(z) = 0 \quad (72)$$

factorisation of the collinear divergences is performed through the addition of the following counterterm for each parton in the initial state:

$$\sigma_{\text{c.t.}}^{\text{CDR}} = \sigma_0^{\text{CDR}} \frac{\alpha_S}{2\pi} \left[\left(\frac{\mu^2}{\mu_F^2} \right)^\epsilon \frac{c_T}{\epsilon} P_{ij}^\epsilon(z) \right] \quad (73)$$

where σ_0^{SCHEME} is the LO cross section and its value depends on the scheme (see the example for Drell-Yan)]. In CDR, when there is a collinear divergence, the cross section behaves as

$$\sigma_R^{\text{coll}} \sim -\frac{1}{\epsilon} P_{ij}^d(z) \sigma_0^{\text{CDR}} + \text{other terms}. \quad (74)$$

Adding the counterterm (73), leaves a finite part

$$\sigma_R^{\overline{\text{MS}}} \sim -P_{ij}^\epsilon(z) (\sigma_0^{\text{CDR}}|_{\epsilon \rightarrow 0}) + \text{other terms}. \quad (75)$$

References

- [1] R. K. Ellis, W. J. Stirling, and B. Webber, *QCD and collider physics*, (Cambridge Univ. Pr., 1996)
- [2] M. L. Mangano, *Introduction to QCD*, in *Proc. European School on High-Energy Physics, St Andrews, UK, 1998*, <http://is.gd/dQkHUw>, CERN-OPEN-2000-255.

- [3] P. Nason, *Introduction to QCD*, in *Proc. European School on High-Energy Physics, Menstrup, DK, 1997*, <http://is.gd/IKo9DC> CERN-1998-003. <http://cds.cern.ch/record/371280?ln=en>
- [4] G. P. Salam, *PoS ICHEP 2010* (2010) 556.
- [5] H. Georgi, S. Glashow, M. Machacek, and D. V. Nanopoulos, *Phys. Rev. Lett.* **40** (1978) 692.
- [6] S. Dawson, *Nucl. Phys. B* **359** (1991) 283.
- [7] A. Djouadi, M. Spira, and P. Zerwas, *Phys. Lett. B* **264** (1991) 440.
- [8] D. Graudenz, M. Spira, and P. Zerwas, *Phys. Rev. Lett.* **70** (1993) 1372.
- [9] M. Spira, A. Djouadi, D. Graudenz, and P. Zerwas, *Nucl. Phys. B* **453** (1995) 17 [[hep-ph/9504378](#)].
- [10] R. V. Harlander and W. B. Kilgore, *Phys. Rev. Lett.* **88** (2002) 201801 [[hep-ph/0201206](#)].
- [11] C. Anastasiou and K. Melnikov, *Nucl. Phys. B* **646** (2002) 220 [[hep-ph/0207004](#)].
- [12] V. Ravindran, J. Smith, and W. L. van Neerven, *Nucl. Phys. B* **665** (2003) 325 [[hep-ph/0302135](#)].
- [13] R. V. Harlander, H. Mantler, S. Marzani, and K. J. Ozeren, *Eur. Phys. J. C* **66** (2010) 359 [[arXiv:0912.2104](#) [[hep-ph](#)]].
- [14] A. Pak, M. Rogal, and M. Steinhauser, *JHEP* **1002** (2010) 025 [[arXiv:0911.4662](#) [[hep-ph](#)]].
- [15] R. V. Harlander, F. Hofmann, and H. Mantler, *JHEP* **1102** (2011) 055 [[arXiv:1012.3361](#) [[hep-ph](#)]].
- [16] A. Pak, M. Rogal, and M. Steinhauser, *JHEP* **1109** (2011) 088 [[arXiv:1107.3391](#) [[hep-ph](#)]].
- [17] M. R. Whalley, D. Bourilkov, and R. C. Group, *The Les Houches Accord PDFs (LHAPDF) and Lhaglu*, [hep-ph/0508110](#).
- [18] S. Catani and M. Seymour, *Nucl. Phys. B* **485** (1997) 291 [Erratum-ibid. *B* **510** (1998) 503] [[hep-ph/9605323](#)].
- [19] S. Frixione, Z. Kunszt, and A. Signer, *Nucl. Phys. B* **467** (1996) 399 [[hep-ph/9512328](#)].
- [20] M. L. Mangano and S. J. Parke, *Phys. Rept.* **200** (1991) 301 [[hep-th/0509223](#)].
- [21] L. J. Dixon, *Calculating scattering amplitudes efficiently*, in *Proc. TASI 1995*, [hep-ph/9601359](#).
- [22] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *JHEP* **0307** (2003) 001 [[hep-ph/0206293](#)].
- [23] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *JHEP* **1106** (2011) 128 [[arXiv:1106.0522](#) [[hep-ph](#)]].
- [24] S. Catani and M. Grazzini, *Phys. Rev. Lett.* **98** (2007) 222002 [[hep-ph/0703012](#)].
- [25] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *Phys. Rev. D* **69** (2004) 094008 [[hep-ph/0312266](#)].
- [26] P. Bolzoni, F. Maltoni, S.-O. Moch, and M. Zaro, *Phys. Rev. D* **85** (2012) 035002 [[arXiv:1109.3717](#) [[hep-ph](#)]].
- [27] D. de Florian, G. Ferrera, M. Grazzini, and D. Tommasini, *JHEP* **1111** (2011) 064 [[arXiv:1109.2109](#) [[hep-ph](#)]].
- [28] S. Dittmaier *et al.* [LHC Higgs Cross Section Working Group Collaboration], *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, [arXiv:1101.0593](#) [[hep-ph](#)].
- [29] T. Sjostrand, S. Mrenna, and P. Z. Skands, *JHEP* **0605** (2006) 026 [[hep-ph/0603175](#)].
- [30] T. Sjostrand, S. Mrenna, and P. Z. Skands, *Comput. Phys. Commun.* **178** (2008) 852 [[arXiv:0710.3820](#) [[hep-ph](#)]].
- [31] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour and B. R. Webber, *JHEP* **0101** (2001) 010 [[hep-ph/0011363](#)].
- [32] M. Bahr, S. Gieseke, M. A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Platzer and P. Richardson *et al.*, *Eur. Phys. J. C* **58** (2008) 639 [[arXiv:0803.0883](#) [[hep-ph](#)]].
- [33] T. Gleisberg, S. Hoeche, F. Krauss, A. Schaliche, S. Schumann and J. -C. Winter, *JHEP* **0402** (2004) 056 [[hep-ph/0311263](#)].

- [34] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Hoche, H. Hoeth, F. Krauss and L. Lonnblad *et al.*, Phys. Rept. **504** (2011) 145 [arXiv:1101.2599 [hep-ph]].
- [35] S. Catani, F. Krauss, R. Kuhn, and B. Webber, JHEP **0111** (2001) 063 [hep-ph/0109231].
- [36] F. Krauss, JHEP **0208** (2002) 015 [hep-ph/0205283].
- [37] L. Lonnblad, JHEP **0205** (2002) 046 [hep-ph/0112284].
- [38] S. Hoeche, F. Krauss, S. Schumann, and F. Siegert, JHEP **0905** (2009) 053 [arXiv:0903.1219 [hep-ph]].
- [39] K. Hamilton, P. Richardson, and J. Tully, JHEP **0911** (2009) 038 [arXiv:0905.3072 [hep-ph]].
- [40] M. L. Mangano, M. Moretti, and R. Pittau, Nucl. Phys. B **632** (2002) 343 [hep-ph/0108069].
- [41] J. Alwall, S. de Visscher, and F. Maltoni, JHEP **0902** (2009) 017 [arXiv:0810.5350 [hep-ph]].
- [42] J. Alwall, S. Hoche, F. Krauss, N. Lavesson, L. Lonnblad, *et al.*, Eur. Phys. J. C **53** (2008) 473 [arXiv:0706.2569 [hep-ph]].
- [43] S. Frixione and B. R. Webber, JHEP **0206** (2002) 029 [hep-ph/0204244].
- [44] P. Nason, JHEP **0411** (2004) 040 [hep-ph/0409146].
- [45] S. Frixione, F. Stoeckli, P. Torrielli, B. R. Webber and C. D. White, *The MCaNLO 4.0 Event Generator*, arXiv:1010.0819 [hep-ph].
- [46] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, R. Pittau and P. Torrielli, Phys. Lett. B **701** (2011) 427 [arXiv:1104.5613 [hep-ph]].
- [47] K. Hamilton, P. Richardson, and J. Tully, JHEP **0904** (2009) 116 [arXiv:0903.4345 [hep-ph]].
- [48] S. Alioli, P. Nason, C. Oleari, and E. Re, JHEP **1006** (2010) 043 [arXiv:1002.2581 [hep-ph]].
- [49] S. Hoche, F. Krauss, M. Schonherr, and F. Siegert, JHEP **1104** (2011) 024 [arXiv:1008.5399 [hep-ph]].
- [50] C. Balazs, J. Huston, and I. Puljak, Phys. Rev. D **63** (2001) 014021 [hep-ph/0002032].
- [51] G. Bozzi, S. Catani, D. de Florian, and M. Grazzini, Nucl. Phys. B **737** (2006) 73 [hep-ph/0508068].
- [52] U. Langenegger, M. Spira, A. Starodumov, and P. Trueb, JHEP **0606** (2006) 035 [hep-ph/0604156].
- [53] C. Anastasiou, S. Bucherer, and Z. Kunszt, JHEP **0910** (2009) 068 [arXiv:0907.2362 [hep-ph]].
- [54] J. Alwall, Q. Li, and F. Maltoni, Phys. Rev. D **85** (2012) 014031 [arXiv:1110.1728 [hep-ph]].
- [55] E. Bagnaschi, G. Degrossi, P. Slavich, and A. Vicini, JHEP **1202** (2012) 088 [arXiv:1111.2854 [hep-ph]].

Introduction to Flavor Physics

Y. Grossman

Cornell University, Ithaca, New York, USA

Abstract

This set of lectures covers the very basics of flavor physics and are aimed to be an entry point to the subject. A lot of problems are provided in the hope of making the manuscript a self study guide.

1 Welcome statement

My plan for these lectures is to introduce you to the very basics of flavor physics. Hopefully, after the lectures you will have enough knowledge, and more importantly, enough curiosity, that you will go on and learn more about the subject.

These are lecture notes and are not meant to be a review. I try to present the basic ideas, hoping to give a clear picture of the physics. Thus, many details are omitted, implicit assumptions are made, and no references are given. Yet details are important: after you go over the current lecture notes once or twice, I hope you will feel the need for more. Then it will be the time to turn to the many reviews [1–11] and books [12, 13] on the subject.

I have tried to include many homework problems for the reader to solve, much more than what I gave in the actual lectures. If you would like to learn the material, I think that the provided problems are the way to start. They force you to fully understand the issues and apply your knowledge to new situations. The problems are given at the end of each section. The questions can be challenging and may take a lot of time. Do not give up after a few minutes!

2 The standard model: a reminder

I assume that you have basic knowledge of Quantum Field Theory (QFT) and that you are familiar with the Standard Model (SM). Nevertheless, I start with a brief review of the SM, not only to remind you, but also since I like to present things in a way that may be different from the way you got to know the SM.

2.1 The basic of model building

In high energy physics, we ask a very simple question: What are the fundamental laws of Nature? We know that QFT is an adequate tool to describe Nature, at least at energies we have probed so far. Thus the question can be stated in a very compact form as: what is the Lagrangian of nature? The most compact form of the question is

$$\mathcal{L} = ? \tag{1}$$

In order to answer this question we need to provide some axioms or “rules.” Our rules are that we “build” the Lagrangian by providing the following three ingredients:

1. The gauge symmetry of the Lagrangian;
2. The representations of fermions and scalars under the symmetry;
3. The pattern of spontaneous symmetry breaking.

The last point practically represented by signs for some parameters. for example the sign of the Higgs mass-squared parameter at the unstable vacuum ($\mu^2 < 0$).

Once we have specified these ingredients, the next step is to write the *most general renormalizable Lagrangian* that is invariant under the gauge symmetry and provides the required spontaneous symmetry breaking pattern. This is far from a trivial statement. The “most general” statement tells us that all terms that satisfy the above conditions must be present in the Lagrangian, even the terms that may be phenomenologically problematic. For example, even though we might not want to include a term that induces proton decay, we cannot simply omit it from our model without some symmetry principle that forbids it.

The require of renormalizability strongly constrains the form of a Lagrangian and, in fact, limits us to only a finite number of terms. This condition should be thinking of an approximation. We think of our model as an effective low energy model and thus write its Lagrangian in power series in term of $1/\Lambda$, where Λ is the UV scale.¹ The requirement of renormalizability is therefore equivalent to saying that we trunk the infinite series of operators.

Few remarks are in order about these starting points.

1. We also impose Poincare invariance. In a way, this can be identified as the gauge symmetry of gravity, and thus can be thought of part of the first postulate.
2. As we already mentioned, we assume QFT. In particular, quantum mechanics is also an axiom.
3. We do not impose global symmetries. They are accidental, that is, they are there only because we do not allow for non renormalizable terms.
4. The basic fermion fields are two component Weyl spinors. The basic scalar fields are complex. The vector fields are introduced to the model in order to preserve the gauge symmetry.
5. Any given model has a finite number of parameters. These parameters need to be measured before the model can be tested.

Note in the last point that the idea of the number of parameters required to define a theory is independent of how they are parametrized. In fact, many not-so-obvious ideas in field theory such as the renormalization group become obvious once one understands that they are reparameterizations of these physical parameters. In practice, if one makes $n > k$ measurements of a theory, one does not first do k parameter measurements and then $(n - k)$ observations of the theory. Instead one would take all n measurements and do a statistical fit for the k parameters to check for self-consistency. Philosophically, however, it is useful and important to remember that a model (a Lagrangian) by itself does not make predictions, it must come with measurements of its parameters.

2.2 An example: the SM

As an example we consider the SM. It is a nice example, mainly because it describes Nature, and also because the tools we use to construct the SM are also those we use when constructing its possible extensions. The SM is defined as follows:

1. The gauge symmetry is

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

2. There are three fermion generations, each consisting of five representations of G_{SM} :

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

Our notations mean that, for example, left-handed quarks, Q_L^I , are triplets of $SU(3)_C$, doublets of $SU(2)_L$ and carry hypercharge $Y = +1/6$. The super-index I denotes gauge interaction eigenstates. The sub-index $i = 1, 2, 3$ is the flavor (or generation) index. There is a single scalar

¹We introduced here some vocabulary. UV is refer to short distance or high energy. While we did us the term IR yet, it worth mentioning that it refers to long distance or low energy.

representation,

$$\phi(1, 2)_{+1/2}. \quad (4)$$

3. The scalar ϕ assumes a VEV,

$$\langle \phi \rangle = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}, \quad (5)$$

which implies that the gauge group is spontaneously broken,

$$G_{\text{SM}} \rightarrow SU(3)_C \times U(1)_{\text{EM}}. \quad (6)$$

This SSB pattern is equivalent to requiring that one parameter in the scalar potential is negative, that is $\mu^2 < 0$, see Eq. (11).

The standard model Lagrangian, $\mathcal{L}_{\text{SM}}$, is the most general renormalizable Lagrangian that is consistent with the gauge symmetry (2) and the particle content (3) and (4). It can be divided to three parts:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{kinetic}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}. \quad (7)$$

We will learn how to count parameters later, but for now we just mention that $\mathcal{L}_{\text{SM}}$ has 18 free parameters² that we need to determine experimentally. Now we talk a little about each part of $\mathcal{L}_{\text{SM}}$.

For the kinetic terms, in order to maintain gauge invariance, one has to replace the derivative with a covariant derivative:

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

Here G_a^μ are the eight gluon fields, W_b^μ the three weak interaction bosons and B^μ the single hypercharge boson. The L_a 's are $SU(3)_C$ generators (the 3×3 Gell-Mann matrices $\frac{1}{2}\lambda_a$ for triplets, 0 for singlets), the T_b 's are $SU(2)_L$ generators (the 2×2 Pauli matrices $\frac{1}{2}\tau_b$ for doublets, 0 for singlets), and the Y 's are the $U(1)_Y$ charges. For example, for the left-handed quarks, Q_L^I , we have

$$\mathcal{L}_{\text{kinetic}}(Q_L) = i\overline{Q_L^I}\gamma_\mu \left(\partial^\mu + \frac{i}{2}g_s G_a^\mu \lambda_a + \frac{i}{2}g W_b^\mu \tau_b + \frac{i}{6}g' B^\mu \right) Q_L^I, \quad (9)$$

while for the left-handed leptons, L_L^I , we have

$$\mathcal{L}_{\text{kinetic}}(L_L) = i\overline{L_L^I}\gamma_\mu \left(\partial^\mu + \frac{i}{2}g W_b^\mu \tau_b - \frac{i}{2}g' B^\mu \right) L_L^I. \quad (10)$$

This part of the Lagrangian has three parameters, g , g' and g_s .

The Higgs potential,³ which describes the scalar self interactions, is given by:

$$\mathcal{L}_{\text{Higgs}} = \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2. \quad (11)$$

This part of the Lagrangian involves two parameters, λ and μ , or equivalently, the Higgs mass and its VEV. The requirement of vacuum stability tells us that $\lambda > 0$. The pattern of spontaneous symmetry breaking, Eq. (5), requires $\mu^2 < 0$. We will not elaborate on this point.

We split the Yukawa part into two, the leptonic and baryonic parts. At the renormalizable level the lepton Yukawa interactions are given by

$$-\mathcal{L}_{\text{Yukawa}}^{\text{leptons}} = Y_{ij}^e \overline{L_{Li}^I} \phi E_{Rj}^I + \text{h.c.} \quad (12)$$

²In fact there is one extra parameter that is related to the vacuum structure of the strong interaction, Θ_{QCD} . Discussing this parameter is far beyond the scope of these lectures, and we only mention it in this footnote in order not to make incorrect statements.

³The Higgs mechanism was not first proposed by Higgs. The first paper suggesting it was by Englert and Brout. It was independently suggested by Higgs and by Guralnik, Hagen, and Kibble.

After the Higgs acquires a VEV, these terms lead to charged lepton masses. Note that the SM predicts massless neutrinos. The Lepton Yukawa terms involve three physical parameters, which are usually chosen to be the three charged lepton masses. We will not discuss the lepton sector in these lectures.

The quark Yukawa interactions are given by⁴

$$-\mathcal{L}_{\text{Yukawa}}^{\text{quarks}} = Y_{ij}^d \overline{Q}_{Li}^I \phi D_{Rj}^I + Y_{ij}^u \overline{Q}_{Li}^I \tilde{\phi} U_{Rj}^I + \text{h.c.} \quad (13)$$

This is the part where quarks masses and flavor arises, and we will spend the rest of the lectures on it. For now, just in order to finish the counting, we mention that the Yukawa interactions for the quarks are described by ten physical parameters. They can be chosen to be the six quark masses and the four parameters of the CKM matrix. We will discuss the CKM matrix at length soon.

2.3 More symmetries

So far we only mentioned the gauge symmetries that we imposed. Before we go on, let me talk about few other type of symmetries.

2.3.1 C, P, T, and their combinations

We start with the discrete symmetries, C, P and T. Any local Lorentz invariant QFT conserves CPT, and in particular, this is also the case in the SM. CPT conservation also implies that T violation is equivalent CP violation. You may wonder why we discuss these symmetries as we are dealing with flavor. It turns out that in Nature, C, P, and CP violation are closely related to flavor physics. There is no reason for this to be the case, but since it is, we study it simultaneously.

In the SM, C and P are “maximally violated.” By that we refer to the fact that both C and P change the chirality of fermion fields. In the SM the left handed and right handed fields have different gauge representations, and thus, independent of the values of the parameters of the model, C and P must be violated in the SM.

The situation with CP is different. The SM can violate CP but it depends on the values of its parameters. It turns out that the parameters of the SM that describe Nature violate CP. The requirement for CP violation is that there is a physical phase in the Lagrangian. In the SM the only place where a complex phase can be physical is in the quark Yukawa interactions. More precisely, in the SM, CP is violated if and only if

$$\text{Im}(\det[Y^d Y^{d\dagger}, Y^u Y^{u\dagger}]) \neq 0. \quad (14)$$

An intuitive explanation of why CP violation is related to complex Yukawa couplings goes as follows. The Hermiticity of the Lagrangian implies that $\mathcal{L}_{\text{Yukawa}}$ has pairs of terms in the form

$$Y_{ij} \overline{\psi}_{Li} \phi \psi_{Rj} + Y_{ij}^* \overline{\psi}_{Rj} \phi^\dagger \psi_{Li}. \quad (15)$$

A CP transformation exchanges the above two operators

$$\overline{\psi}_{Li} \phi \psi_{Rj} \Leftrightarrow \overline{\psi}_{Rj} \phi^\dagger \psi_{Li}, \quad (16)$$

but leaves their coefficients, Y_{ij} and Y_{ij}^* , unchanged. This means that CP is a symmetry of $\mathcal{L}_{\text{Yukawa}}$ if $Y_{ij} = Y_{ij}^*$.

In the SM the only source of CP violation are Yukawa interactions. It is easy to see that the kinetic terms are CP conserving. For the SM scalar sector, where there is a single doublet, this part of the Lagrangian is also CP conserving. For extended scalar sectors, such as that of a two Higgs doublet model, $\mathcal{L}_{\text{Higgs}}$ can be CP violating.

⁴We introduced here $\tilde{\phi} = i\epsilon_{ij}\phi^*$. Basically, this mathematical manipulation is needed so here the neutral component is the upper components.

2.3.2 Global, accidental, and approximate symmetries

One might ask how we can have baryon number conservation if our rules above stated that we may not explicitly impose any global symmetries. The global symmetries of the Standard Model result as outputs of the theory rather than as external constraints. In particular, they come from the structure imposed by renormalizability and gauge invariance. Global symmetries that appear only because non-renormalizable terms are not considered are called accidental symmetries. These are broken explicitly by non-renormalizable terms, but since these terms are small one can often make use of these symmetries. When people ask why the proton does not decay in the Standard Model, then baryon number conservation is only part of the answer; the “fundamental” reason is that baryon-number-violating operators in the Standard Model are non-renormalizable.

The SM has an accidental global symmetry

$$U(1)_B \times U(1)_e \times U(1)_\mu \times U(1)_\tau. \quad (17)$$

where $U(1)_B$ is baryon number and the other three $U(1)$ s are lepton family lepton numbers. The quarks carry baryon number, while the leptons and the bosons do not. We usually normalize it such that the proton has $B = 1$ and thus each quark carries a third unit of baryon number. As for lepton number, in the SM each family carries its own lepton number, L_e , L_μ and L_τ . Total lepton number is a subgroup of this more general symmetry, that is, the sum of all three family lepton numbers. In these lectures we concentrate on the quark sector and therefore we do not elaborate much on the global symmetry of the lepton sector.

In addition to accidental symmetries there are other approximate symmetries which are parametrically small in the sense that they become exact global symmetries when a parameter (or set of parameters) is set to zero. An example is isospin symmetry, which is an approximate global symmetry of the Standard Model. It is broken by the quark masses. Are these parametrically small? We can only say this about dimensionless parameters, and it turns out that the relevant parameter is the ratio of the quark mass splittings to the QCD strong coupling scale,

$$\frac{m_u - m_d}{\Lambda_{\text{QCD}}}. \quad (18)$$

Isospin symmetry is additionally broken by “the most famous small parameter in the world,” α_{EM} since the charge of the u is different from the charge of the d .

There is one more small parameter that you might be familiar with. The Higgs potential obeys a “custodial symmetry.” It is broken at one-loop by the Yukawa couplings. While one might argue that the Yukawas themselves are $\mathcal{O}(1)$ and thus not small, since the breaking only occurs at loop-level the relevant parameter is $y^2/16\pi^2$ which is small.

These accidental and approximate symmetries can be very useful, but we should always remember that they are not as fundamental as the gauge symmetries that we begin with.

2.4 Counting parameters

Before we go on to study the flavor structure of the SM in detail, we explain how to identify the number of physical parameter in any model. The Yukawa interactions of Eq. (13) have many parameters but some are not physical. That is, there is a basis where they are identically zero. Of course, it is important to identify the physical parameters in any model in order to probe and check it.

We start with a very simple example. Consider a hydrogen atom in a uniform magnetic field. Before turning on the magnetic field, the hydrogen atom is invariant under spatial rotations, which are described by the $SO(3)$ group. Furthermore, there is an energy eigenvalue degeneracy of the Hamiltonian: states with different angular momenta have the same energy. This degeneracy is a consequence of the symmetry of the system.

When magnetic field is added to the system, it is conventional to pick a direction for the magnetic field without a loss of generality. Usually, we define the positive z direction to be the direction of the magnetic field. Consider this choice more carefully. A generic uniform magnetic field would be described by three real numbers: the three components of the magnetic field. The magnetic field breaks the $SO(3)$ symmetry of the hydrogen atom system down to an $SO(2)$ symmetry of rotations in the plane perpendicular to the magnetic field. The one generator of the $SO(2)$ symmetry is the only valid symmetry generator now; the remaining two $SO(3)$ generators in the orthogonal planes are broken. These broken symmetry generators allow us to rotate the system such that the magnetic field points in the z direction:

$$O_{xz}O_{yz}(B_x, B_y, B_z) = (0, 0, B'_z), \quad (19)$$

where O_{xz} and O_{yz} are rotations in the xz and yz planes respectively. The two broken generators were used to rotate away two unphysical parameters, leaving us with one physical parameter, the magnitude of the magnetic field. That is, when turning on the magnetic field, all measurable quantities in the system depend only on one new parameter, rather than the naïve three.

The results described above are more generally applicable. Particularly, they are useful in studying the flavor physics of quantum field theories. Consider a gauge theory with matter content. This theory always has kinetic and gauge terms, which have a certain global symmetry, G_f , on their own. In adding a potential that respect the imposed gauge symmetries, the global symmetry may be broken down to a smaller symmetry group. In breaking the global symmetry, there is an added freedom to rotate away unphysical parameters, as when a magnetic field is added to the hydrogen atom system.

In order to analyze this process, we define a few quantities. The added potential has coefficients that can be described by N_{general} parameters in a general basis. The global symmetry of the entire model, H_f , has fewer generators than G_f and we call the difference in the number of generators N_{broken} . Finally, the quantity that we would ultimately like to determine is the number of parameters affecting physical measurements, N_{phys} . These numbers are related by

$$N_{\text{phys}} = N_{\text{general}} - N_{\text{broken}}. \quad (20)$$

Furthermore, the rule in Eq. (20) applies separately for both real parameters (masses and mixing angles) and phases. A general, $n \times n$ complex matrix can be parametrized by n^2 real parameters and n^2 phases. Imposing restrictions like Hermiticity or unitarity reduces the number of parameters required to describe the matrix. A Hermitian matrix can be described by $n(n+1)/2$ real parameters and $n(n-1)/2$ phases, while a unitary matrix can be described by $n(n-1)/2$ real parameters and $n(n+1)/2$ phases.

The rule given by Eq. (20) can be applied to the standard model. Consider the quark sector of the model. The kinetic term has a global symmetry

$$G_f = U(3)_Q \times U(3)_U \times U(3)_D. \quad (21)$$

A $U(3)$ has 9 generators (3 real and 6 imaginary), so the total number of generators of G_f is 27. The Yukawa interactions defined in Eq. (13), Y^F ($F = u, d$), are 3×3 complex matrices, which contain a total of 36 parameters (18 real parameters and 18 phases) in a general basis. These parameters also break G_f down to the baryon number

$$U(3)_Q \times U(3)_U \times U(3)_D \rightarrow U(1)_B. \quad (22)$$

While $U(3)^3$ has 27 generators, $U(1)_B$ has only one and thus $N_{\text{broken}} = 26$. This broken symmetry allows us to rotate away a large number of the parameters by moving to a more convenient basis. Using Eq. (20), the number of physical parameters should be given by

$$N_{\text{phys}} = 36 - 26 = 10. \quad (23)$$

These parameters can be split into real parameters and phases. The three unitary matrices generating the symmetry of the kinetic and gauge terms have a total of 9 real parameters and 18 phases. The symmetry is broken down to a symmetry with only one phase generator. Thus,

$$N_{\text{phys}}^{(r)} = 18 - 9 = 9, \quad N_{\text{phys}}^{(i)} = 18 - 17 = 1. \quad (24)$$

We interpret this result by saying that of the 9 real parameters, 6 are the fermion masses and three are the CKM matrix mixing angles. The one phase is the CP-violating phase of the CKM mixing matrix.

In your homework you will count the number of parameters for different models.

2.5 The CKM matrix

We are now equipped with the necessary tools to study the Yukawa interactions. The basic tool we need is that of basis rotations. There are two important bases. One where the masses are diagonal, called the mass basis, and the other where the $W^\pm$ interactions are diagonal, called the interaction basis. The fact that these two bases are not the same results in flavor changing interactions. The CKM matrix is the matrix that rotates between these two bases.

Since most measurements are done in the mass basis, we write the interactions in that basis. Upon the replacement $\mathcal{R}e(\phi^0) \rightarrow (v + H^0)/\sqrt{2}$ (see Eq. (5)), we decompose the $SU(2)_L$ quark doublets into their components:

$$Q_{Li}^I = \begin{pmatrix} U_{Li}^I \\ D_{Li}^I \end{pmatrix}. \quad (25)$$

and then the Yukawa interactions, Eq. (13), give rise to mass terms:

$$-\mathcal{L}_M^q = (M_d)_{ij} \overline{D}_{Li}^I D_{Rj}^I + (M_u)_{ij} \overline{U}_{Li}^I U_{Rj}^I + \text{h.c.}, \quad M_q = \frac{v}{\sqrt{2}} Y^q, \quad (26)$$

The mass basis corresponds, by definition, to diagonal mass matrices. We can always find unitary matrices V_{qL} and V_{qR} such that

$$V_{qL} M_q V_{qR}^\dagger = M_q^{\text{diag}}, \quad q = u, d, \quad (27)$$

with M_q^{diag} diagonal and real. The quark mass eigenstates are then identified as

$$q_{Li} = (V_{qL})_{ij} q_{Lj}^I, \quad q_{Ri} = (V_{qR})_{ij} q_{Rj}^I, \quad q = u, d. \quad (28)$$

The charged current interactions for quarks are the interactions of the $W_\mu^\pm$, which in the interaction basis are described by Eq. (9). They have a more complicated form in the mass basis:

$$-\mathcal{L}_{W^\pm}^q = \frac{g}{\sqrt{2}} \overline{u}_{Li} \gamma^\mu (V_{uL} V_{dL}^\dagger)_{ij} d_{Lj} W_\mu^+ + \text{h.c.} \quad (29)$$

The unitary 3×3 matrix,

$$V = V_{uL} V_{dL}^\dagger, \quad (V V^\dagger = \mathbf{1}), \quad (30)$$

is the Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix for quarks. As a result of the fact that V is not diagonal, the $W^\pm$ gauge bosons couple to mass eigenstates quarks of different generations. Within the SM, this is the only source of flavor changing quark interactions.

The form of the CKM matrix is not unique. We already counted and concluded that only one of the phases is physical. This implies that we can find bases where V has a single phase. This physical phase is the Kobayashi-Maskawa phase that is usually denoted by δ_{KM} .

There is more freedom in defining V in that we can permute between the various generations. This freedom is fixed by ordering the up quarks and the down quarks by their masses, *i.e.* $(u_1, u_2, u_3) \rightarrow$

(u, c, t) and $(d_1, d_2, d_3) \rightarrow (d, s, b)$. The elements of V are therefore written as follows:

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \quad (31)$$

The fact that there are only three real and one imaginary physical parameters in V can be made manifest by choosing an explicit parametrization. For example, the standard parametrization, used by the Particle Data Group (PDG) [14], is given by

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \quad (32)$$

where $c_{ij} \equiv \cos \theta_{ij}$ and $s_{ij} \equiv \sin \theta_{ij}$. The three $\sin \theta_{ij}$ are the three real mixing parameters while δ is the Kobayashi-Maskawa phase. Another parametrization is the Wolfenstein parametrization where the four mixing parameters are (λ, A, ρ, η) where η represents the CP violating phase. The Wolfenstein parametrization is an expansion in the small parameter, $\lambda = |V_{us}| \approx 0.22$. To $O(\lambda^3)$ the parametrization is given by

$$V = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix}. \quad (33)$$

We will talk in detail about how to measure the CKM parameters. For now let us mention that the Wolfenstein parametrization is a good approximation to the actual numerical values. That is, the CKM matrix is very close to a unit matrix with off diagonal terms that are small. The order of magnitude of each element can be read from the power of λ in the Wolfenstein parametrization.

Various parametrizations differ in the way that the freedom of phase rotation is used to leave a single phase in V . One can define, however, a CP violating quantity in V that is independent of the parametrization. This quantity, the Jarlskog invariant, J_{CKM} , is defined through

$$\mathcal{I}m(V_{ij}V_{kl}V_{il}^*V_{kj}^*) = J_{\text{CKM}} \sum_{m,n=1}^3 \epsilon_{ikm}\epsilon_{jln}, \quad (i, j, k, l = 1, 2, 3). \quad (34)$$

In terms of the explicit parametrizations given above, we have

$$J_{\text{CKM}} = c_{12}c_{23}c_{13}^2 s_{12}s_{23}s_{13} \sin \delta \approx \lambda^6 A^2 \eta. \quad (35)$$

The condition (14) can be translated to the language of the flavor parameters in the mass basis. Then we see that a necessary and sufficient condition for CP violation in the quark sector of the SM (we define $\Delta m_{ij}^2 \equiv m_i^2 - m_j^2$):

$$\Delta m_{tc}^2 \Delta m_{tu}^2 \Delta m_{cu}^2 \Delta m_{bs}^2 \Delta m_{bd}^2 \Delta m_{sd}^2 J_{\text{CKM}} \neq 0. \quad (36)$$

Equation (36) puts the following requirements on the SM in order that it violates CP:

1. Within each quark sector, there should be no mass degeneracy;
2. None of the three mixing angles should be zero or $\pi/2$;
3. The phase should be neither 0 nor π .

A very useful concept is that of the unitarity triangles. The unitarity of the CKM matrix leads to various relations among the matrix elements, for example,

$$\sum_i V_{id}V_{is}^* = 0. \quad (37)$$

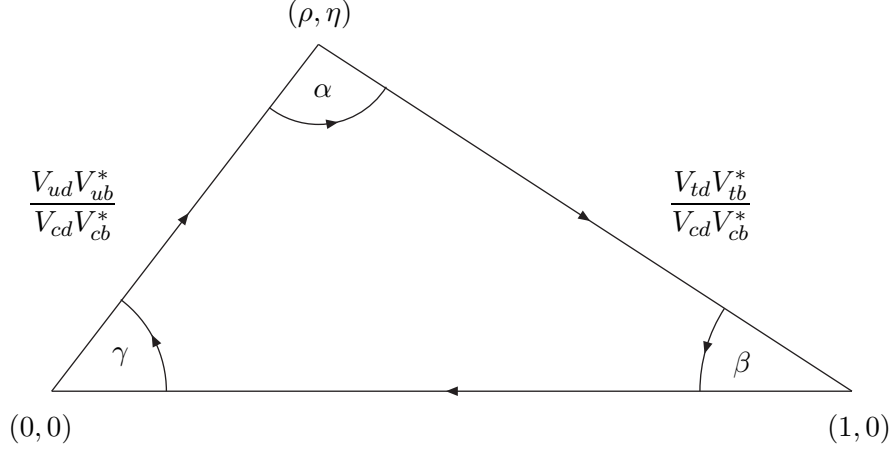


Fig. 1: The unitarity triangle.

There are six such relations and they require the sum of three complex quantities to vanish. Therefore, they can be geometrically represented in the complex plane as a triangle and are called “unitarity triangles”. It is a feature of the CKM matrix that all unitarity triangles have equal areas. Moreover, the area of each unitarity triangle equals $|J_{\text{CKM}}|/2$ while the sign of J_{CKM} gives the direction of the complex vectors around the triangles.

One of these triangles has sides roughly the same length, is relatively easy to probe, and is corresponds to the relation

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \quad (38)$$

For these reasons, the term “the unitarity triangle” is reserved for Eq. (38). We further define the rescaled unitarity triangle. It is derived from Eq. (38) by choosing a phase convention such that $(V_{cd}V_{cb}^*)$ is real and dividing the lengths of all sides by $|V_{cd}V_{cb}^*|$. The rescaled unitarity triangle is similar to the unitarity triangle. Two vertices of the rescaled unitarity triangle are fixed at (0,0) and (1,0). The coordinates of the remaining vertex correspond to the Wolfenstein parameters (ρ, η) . The unitarity triangle is shown in Fig. 1.

The lengths of the two complex sides are

$$R_u \equiv \left| \frac{V_{ud}V_{ub}}{V_{cd}V_{cb}} \right| = \sqrt{\rho^2 + \eta^2}, \quad R_t \equiv \left| \frac{V_{td}V_{tb}}{V_{cd}V_{cb}} \right| = \sqrt{(1 - \rho)^2 + \eta^2}. \quad (39)$$

The three angles of the unitarity triangle are defined as follows:

$$\alpha \equiv \arg \left[-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}} \right], \quad \beta \equiv \arg \left[-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}} \right], \quad \gamma \equiv \arg \left[-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}} \right]. \quad (40)$$

They are physical quantities and can be independently measured, as we will discuss later. Another commonly used notation is $\phi_1 = \beta$, $\phi_2 = \alpha$, and $\phi_3 = \gamma$. Note that in the standard parametrization $\gamma = \delta_{\text{KM}}$.

2.6 FCNCs

So far we have talked about flavor changing charged currents that are mediated by the $W^\pm$ bosons. In the SM, this is the only source of flavor changing interaction and, in particular, of generation changing interaction. There is no fundamental reason why there cannot be Flavor Changing Neutral Currents (FCNCs). After all, two interactions of flavor changing charged current result in a neutral current interaction. Yet, experimentally we see that FCNCs processes are highly suppressed.

This is a good place to pause and open your PDG.⁵ Look, for example, at the rate for the neutral current decay, $K_L \rightarrow \mu^+ \mu^-$, and compare it to that of the charged current decay, $K^+ \rightarrow \mu^+ \nu$. You see that the K_L decay rate is much smaller. It is a good idea at this stage to browse the PDG a bit more and see that the same pattern is found in D and B decays.

The fact that the data show that FCNCs are highly suppressed implies that any model that aims to describe Nature must have a mechanism to suppress FCNCs. The SM's way to deal with the data is to make sure there are no tree level FCNCs. In the SM, FCNCs are mediated only at the loop level and are therefore suppressed (we discuss the exact amount of suppression below). Next we explain why in the SM all neutral current interactions are flavor conserving at the tree level.

Before that, we make a short remark. We often talk about non-diagonal couplings, diagonal couplings and universal couplings. Universal couplings are diagonal couplings with the same strength. An important point to recall is that universal couplings are diagonal in any basis. Non-universal diagonal couplings, in general, become non-diagonal after a basis rotation.

There are four types of neutral bosons in the SM that could mediate tree level neutral currents. They are the gluons, photon, Higgs and Z bosons. We study each of them in turn, explain what is required in order to make their couplings diagonal in the mass basis, and how this requirement is fulfilled in the SM.

We start with the massless gauge bosons: the gluons and photon. For them, tree level couplings are always diagonal, independent of the details of the theory. The reason is that these gauge bosons correspond to exact gauge symmetries. Thus, their couplings to the fermions arise from the kinetic terms. When the kinetic terms are canonical, the couplings of the gauge bosons are universal and, in particular, flavor conserving. In other words, gauge symmetry plays a dual role: it guarantees that the gauge bosons are massless and that their couplings are flavor universal.

Next we move to the Higgs interactions. The reason that the Higgs couplings are diagonal in the SM is because its couplings to fermions are aligned with the mass matrix. The reason is that both the Higgs coupling and the mass matrix are proportional to the same Yukawa couplings. To see that this is the case we consider the Yukawa interactions (13). After inserting $\mathcal{R}e(\phi^0) \rightarrow (v + h)/\sqrt{2}$ and keeping both the fermion masses and Higgs fermion interaction terms we get

$$\begin{aligned} -\mathcal{L}_{\text{Yukawa}}^{\text{quarks}} &= Y_{ij}^d \overline{Q}_{Li}^I \phi D_{Rj}^I + Y_{ij}^u \overline{Q}_{Li}^I \tilde{\phi} U_{Rj}^I \\ &= \frac{Y_{ij}^d}{\sqrt{2}} \left(\overline{D}_{Li}^I D_{Rj}^I \right) (v + h) + \frac{Y_{ij}^u}{\sqrt{2}} \left(\overline{U}_{Li}^I U_{Rj}^I \right) (v + h). \end{aligned} \quad (41)$$

Diagonalizing the mass matrix, we get the interaction in the physical basis

$$M_d \left(\overline{D}_{Li} D_{Rj} \right) (v + h) + M_u \left(\overline{U}_{Li} U_{Rj} \right) (v + h). \quad (42)$$

Clearly, since everything is proportional to $(v + h)$, the interaction is diagonalized together with the mass matrix.

This special feature of the Higgs interaction is tightly related to the facts that the SM has only one Higgs field and that the only source for fermion masses is the Higgs VEV. In models where there are additional sources for the masses, that is, bare mass terms or more Higgs fields, diagonalization of the mass matrix does not simultaneously diagonalize the Higgs interactions. In general, there are Higgs mediated FCNCs in such models. In your homework you will work out an example of such models.

Last, we discuss Z -mediated FCNCs. The coupling for the Z to fermions is proportional to $T_3 - q \sin^2 \theta_W$ and in the interaction basis the Z couplings to quarks are given by

$$-\mathcal{L}_Z = \frac{g}{\cos \theta_W} \left[\overline{u}_{Li}^I \gamma^\mu \left(\frac{1}{2} - \frac{2}{3} \sin^2 \theta_W \right) u_{Li}^I + \overline{u}_{Ri}^I \gamma^\mu \left(-\frac{2}{3} \sin^2 \theta_W \right) u_{Ri}^I + \right.$$

⁵It goes without saying that every student in high energy physics must have the PDG [14]. If, for some reason you do not have one, order it now. It is free and has a lot of important stuff. Until you get it, you can use the online version at pdg.lbl.gov.

$$\overline{d_{Li}^I} \gamma^\mu \left(-\frac{1}{2} + \frac{1}{3} \sin^2 \theta_W \right) d_{Li}^I + \overline{d_{Ri}^I} \gamma^\mu \left(\frac{1}{3} \sin^2 \theta_W \right) d_{Ri}^I \Big] Z_\mu + \text{h.c.} \quad (43)$$

In order to demonstrate the fact that there are no FCNCs let us concentrate only on the left handed up-type quarks. Moving to the mass eigenstates we find

$$\begin{aligned} -\mathcal{L}_Z &= \frac{g}{\cos \theta_W} \left[\overline{u_{Li}} (V_{uL})_{ik} \gamma^\mu \left(\frac{1}{2} - \frac{2}{3} \sin^2 \theta_W \right) (V_{uL}^\dagger)_{kj} u_{Lj} \right], \\ &= \frac{g}{\cos \theta_W} \left[\overline{u_{Li}} \gamma^\mu \left(\frac{1}{2} - \frac{2}{3} \sin^2 \theta_W \right) u_{Li} \right] \end{aligned} \quad (44)$$

where in the last step we used

$$V_{uL} V_{uL}^\dagger = 1. \quad (45)$$

We see that the interaction is universal and diagonal in flavor. It is easy to verify that this holds for the other types of quarks. Note the difference between the neutral and the charged currents cases. In the neutral current case we insert $V_{uL} V_{uL}^\dagger = 1$. This is in contrast to the charged current interactions where the insertion is $V_{uL} V_{dL}^\dagger$, which in general is not equal to the identity matrix.

The fact that there are no FCNCs in Z -exchange is due to some specific properties of the SM. That is, we could have Z -mediated FCNCs in simple modifications of the SM. The general condition for the absence of tree level FCNCs is as follows. In general, fields can mix if they belong to the same representation under all the *unbroken* generators. That is, they must have the same spin, electric charge and $SU(3)_C$ representation. If these fields also belong to the same representation under the *broken* generators their couplings to the massive gauge boson is universal. If, however, they belong to different representations under the broken generators, their couplings in the interaction basis are diagonal but non-universal. These couplings become non-diagonal after rotation to the mass basis.

In the SM, the requirement mentioned above for the absence of Z -exchange FCNCs is satisfied. That is, all the fields that belong to the same representation under the unbroken generators also belong to the same representation under the broken generators. For example, all left handed quarks with electric charge $2/3$ also have the same hypercharge ($1/6$) and they are all an up component of a double of $SU(2)_L$ and thus have $T_3 = 1/2$. This does not have to be the case. After all, $Q = T_3 + Y$, so there are many ways to get quarks with the same electric charge. In your homework, you will work out the details of a model with non-standard representations and see how it exhibits Z -exchange FCNCs.

2.7 Homework

Question 1: Global symmetries

We talked about the fact that global symmetries are accidental in the SM, that is, that they are broken once non-renormalizable terms are included. Write the lowest dimension terms that break each of the global symmetries of the SM.

Question 2: Extra generations counting

Count the number of physical flavor parameters in the quark sector of an extended SM with n generations. Show that such a model has $n(n+3)/2$ real parameters and $(n-1)(n-2)/2$ complex phases. Identify the real parameters as masses and mixing angles and determine how many mixing angles there are.

Question 3: Exotic light quarks

We consider a model with the gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$ spontaneously broken by a single Higgs doublet into $SU(3)_C \times U(1)_{EM}$. The quark sector, however, differs from the standard model one as it consists of three quark flavors, that is, we do not have the c , b and t quarks. The quark representations are non-standard. Of the left handed quarks, $Q_L = (u_L, d_L)$ form a doublet of $SU(2)_L$ while s_L is a singlet. All the right handed quarks are singlets. All color representations and electric charges are the same as in the standard model.

1. Write down (a) the gauge interactions of the quarks with the charged W bosons (before SSB); (b) the Yukawa interactions (before SSB); (c) the bare mass terms (before SSB); (d) the mass terms after SSB.
 2. Show that there are four physical flavor parameters in this model. How many are real and how many imaginary? Is there CP violation in this model? Separate the parameters into masses, mixing angles and phases.
 3. Write down the gauge interactions of the quarks with the Z boson in both the interaction basis and the mass basis. (You do not have to rewrite terms that do not change when you rotate to the mass basis. Write only the terms that are modified by the rotation to the mass basis.) Are there, in general, tree level Z exchange FCNCs? (You can assume CP conservation from now on.)
 4. Are there photon and gluon mediated FCNCs? Support your answer by an argument based on symmetries.
 5. Are there Higgs exchange FCNCs?
 6. Repeat the question with a somewhat different model, where the only modification is that two of the right handed quarks, $Q_R = (u_R, d_R)$, form a doublet of $SU(2)_L$. Note that there is one relation between the real parameters that makes the parameter counting a bit tricky.
-

Question 4: Two Higgs doublet model

Consider the two Higgs doublet model (2HDM) extension of the SM. In this model, we add a Higgs doublet to the SM fields. Namely, instead of the one Higgs field of the SM we now have two, denoted by ϕ_1 and ϕ_2 . For simplicity you can work with two generations when the third generation is not explicitly needed.

1. Write down (in a matrix notation) the most general Yukawa potential of the quarks.
 2. Carry out the diagonalization procedure for such a model. Show that the Z couplings are still flavor diagonal.
 3. In general, however, there are FCNCs in this model mediated by the Higgs bosons. To show that, write the Higgs fields as $\text{Re}(\phi_i) = v_i + h_i$ where $i = 1, 2$ and $v_i \neq 0$ is the VEV of ϕ_i , and define $\tan \beta = v_2/v_1$. Then, write down the Higgs–fermion interaction terms in the mass basis. Assuming that there is no mixing between the Higgs fields, you should find a non-diagonal Higgs fermion interaction terms.
-

3 Probing the CKM

Now that we have an idea about flavor in general and in the SM in particular, we are ready to compare the standard model predictions with data. While we use the SM as an example, the tools and ideas are applicable to a large number of theories.

The basic idea is as follows. In order to check a model we first have to determine its parameters and then we can probe it. When considering the flavor sector of the SM, this implies that we first have to measure the parameters of the CKM matrix and then check the model. That is, we can think about the first four measurements as determining the CKM parameters and from the fifth measurements on we are checking the SM. In practice, however, we look for many independent ways to determine the parameters. The SM is checked by looking for consistency among these measurements. Any inconsistency is a signal of new physics.⁶

There is one major issue that we need to think about: how precisely can the predictions of the theory be tested? Our ability to test any theory is bounded by these precisions. There are two kinds of uncertainties: experimental and theoretical. There are many sources of both kinds, and a lot of research has gone into trying to overcome them in order to be able to better probe the SM and its extensions.

We do not elaborate on experimental details. We just make one general point. Since our goal is to probe the small elements of the CKM, we have to measure very small branching ratios, typically down to $O(10^{-6})$. To do that we need a lot of statistics and a superb understanding of the detectors and the backgrounds.

As for theory errors, there is basically one player here: QCD, or, using its mighty name, “the strong interaction.” Yes, it is strong, and yes, it is a problem for us. Basically, we can only deal with weakly coupled forces. The use of perturbation theory is so fundamental to our way of doing physics. It is very hard to deal with phenomena that we cannot use perturbation theory to describe.

In practice the problem is that our theory is given in terms of quarks, but measurements are done with hadrons. It is far from trivial to overcome this gap. In particular, it becomes hard when we are looking for high precision. There are basically two ways to overcome the problem of QCD. One way is to find observables for which the needed hadronic input can be measured or eliminated. The other way is to use approximate symmetries of QCD, in particular, isospin, $SU(3)_F$ and heavy quark symmetries. Below we only mention how these are used without getting into much detail.

3.1 Measuring the CKM parameters

When we attempt to determine the CKM parameters we talk about two classifications. One classification is related to what we are trying to extract:

1. Measure magnitudes of CKM elements or, equivalently, sides of the unitarity triangle;
2. Measure phases of CKM elements or, equivalently, angles of the unitarity triangle;
3. Measure combinations of magnitudes and phases.

The other classification is based on the physics, in particular, we classify based on the type of amplitudes that are involved:

1. Tree level amplitudes. Such measurements are also referred to as “direct measurements;”
2. Loop amplitudes. Such measurements are also referred to as “indirect measurements;”
3. Cases where both tree level and loop amplitude are involved.

There is no fundamental difference between direct and indirect measurement. We make the distinction since direct measurements are expected to be almost model independent. Most extensions of the

⁶The term “new physics” refers to any model that extends the SM. Basically, we are eager to find indications for new physics and determine what that new physics is. At the end of the lectures we expand on this point.

SM have a special flavor structure that suppresses flavor changing couplings and have a very small effect on processes that are large in the SM, which are tree level processes. On the contrary, new physics can have large effect on processes that are very small in the SM, mainly loop processes. Thus, we refer to loop amplitude measurements as indirect measurements.

3.2 Direct measurements

In order to determine the magnitudes of CKM elements, a number of sophisticated theoretical and experimental techniques are needed, the complete discussion of which is beyond the scope of these lectures. Instead, we give one example, the determination of $|V_{cb}|$ and, hopefully, you will find the time to read about direct determinations of other CKM parameters in one of the reviews such as the PDG or Ref. [7].

The basic idea in direct determination of CKM elements is to use the fact that the amplitudes of semi leptonic tree level decays are proportional to one CKM element. In the case of $b \rightarrow c\ell\bar{\nu}$ it is proportional to V_{cb} . (While the diagram is not plotted here, it is a good time to pause and see that you can plot it and see how the dependence on the CKM element enters.) Experimentally, it turns out that $|V_{ub}| \ll |V_{cb}|$. Therefore we can neglect the $b \rightarrow u\ell\bar{\nu}$ decays and use the full semileptonic B decays data set to measure $|V_{cb}|$ without the need to know the hadronic final state.

The way to overcome the problem of QCD is to use heavy quarks symmetry (HQS). We do not discuss the use of HQS in detail here. We just mention that the small expansion parameter is Λ_{QCD}/m_b . The CKM element $|V_{cb}|$ can be extracted from inclusive and exclusive semileptonic B decays.

In the inclusive case, the problem is that the calculation is done using the b and c quarks. In particular, the biggest uncertainty is the fact that at the quark level the decay rate scales like m_b^5 . The definition of the b quark mass, as well as the measurements of it, is complicated: How can we define a mass to a particle that is never free? All we can define and measure very precisely is the B meson mass.⁷ Using an operator product expansion (OPE) together with the heavy quark effective theory, we can expand in the small parameter and get a reasonable estimate of $|V_{cb}|$. The point to emphasize is that this is a controllable expansion, that is, we know that

$$\Gamma(b \rightarrow c\ell\bar{\nu}) = \Gamma(B \rightarrow X_c\ell\bar{\nu}) \left(1 + \sum_n a_n \right), \quad (46)$$

such that a_n is suppressed by $(\Lambda_{QCD}/m_B)^n$. In principle we can calculate all the a_n and get a very precise prediction. It is helpful that $a_1 = 0$. The calculation has been done for $n = 2$ and $n = 3$.

The exclusive approach overcomes the problem of the b quark mass by looking at specific hadronic decays, in particular, $B \rightarrow D\ell\bar{\nu}$ and $B \rightarrow D^*\ell\bar{\nu}$. Here the problem is that the decay cannot be calculated in terms of quarks: it has to be done in terms of hadrons. This is where using “form factors” is useful as we now explain. The way to think about the problem is that a b field creates a free b quark or annihilates a free anti- b quark. Yet, inside the meson the b is not free. Thus the operator that we care about, $\bar{c}\gamma_\mu b$, is not directly related to annihilating the b quark inside the meson. The mismatch is parametrized by form factors. The form factors are functions of the momentum transfer. In general, we need some model to calculate these form factors, as they are related to the strong interaction. In the B case, we can use HQS, which tells us that in the limit $m_B \rightarrow \infty$ all the form factors are universal, and are given by the (unknown) Isgur-Wise function. The fact that we know something about the form factors makes the theoretical errors rather small, below the 5% level.

Similar ideas are used when probing other CKM elements. For example, in β -decay the $d \rightarrow u\ell\bar{\nu}$ decay amplitude is proportional to V_{ud} . Here the way to overcome QCD is by using isospin where the expansion parameter is m_q/Λ_{QCD} with $q = u, d$. Another example is K -decay, $s \rightarrow u\ell\bar{\nu} \propto V_{us}$. In

⁷There is an easy way to remember the mass of the B meson that is based on the fact that it is easier to remember two things than one. I often ask people how many feet there are in one mile, and they do not know the answer. Most of them also do not know the mass of the B meson in MeV. It is rather amusing to note that the answer is, in fact, the same, 5280.

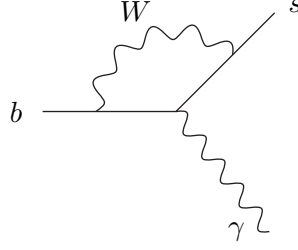


Fig. 2: One of the $b \rightarrow s\gamma$ amplitudes

that case, in addition to isospin, flavor SU(3) is used where we assume that the strange quark is light. In some cases, this is a good approximation, but not as good as isospin.

Direct measurements have been used to measure the magnitude of seven out of the nine CKM matrix components. The two exceptions are $|V_{ts}|$ and $|V_{td}|$. The reason is that the world sample of top decays is very small, and moreover, it is very hard to determine the flavor of the light quark in top decay. These two elements are best probed using loop processes, as we discuss next.

3.3 Indirect measurements

The CKM dependence of decay amplitudes involved in direct measurements of the CKM elements is simple. The amplitudes are tree level with one internal W propagator. In the case of semileptonic decays, the amplitude is directly proportional to one CKM matrix element.

The situation with loop decays is different. Usually we concentrate on FCNC⁸ processes at the one loop level. Since the loop contains an internal W propagator, we gain sensitivity to CKM elements. The sensitivity is always to a combination of CKM elements. Moreover, there are several amplitudes with different internal quarks in the loop. These amplitudes come with different combinations of CKM elements. The total amplitude is the sum of these diagrams, and thus it has a non trivial dependence on combination of CKM elements.

As an example consider one of the most interesting loop induced decay, $b \rightarrow s\gamma$. There are several amplitudes for this decay. One of them is plotted in Fig. 2. (Try to plot the others yourself. Basically the difference is where the photon line goes out.) Note that we have to sum over all possible internal quarks. Each set of diagrams with a given internal up-type quarks, u_i , is proportional to $V_{ib}V_{is}^*$. It can further depend on the mass of the internal quark. Thus, we can write the total amplitude as

$$A(b \rightarrow s\gamma) \propto \sum_{i=u,c,t} f(m_i) V_{ib} V_{is}^*. \quad (47)$$

While the expression in Eq. (47) looks rather abstract, we can gain a lot of insight into the structure of the amplitude by recalling that the CKM matrix is unitary. Using

$$\sum_{i=u,c,t} V_{ib} V_{is}^* = 0, \quad (48)$$

we learn that the contribution of the m_i independent term in f vanishes. Explicit calculation shows that $f(m_i)$ grows with m_i and, if expanding in m_i/m_W , that the leading term scales like m_i^2 .

The fact that in loop decays the amplitude is proportional to m_i^2/m_W^2 is called the GIM mechanism. Historically, it was the first theoretical motivation of the charm quark. Before the charm was

⁸In the first lecture we proved that in the SM there are no tree-level FCNCs. Why do we talk about FCNCs here? I hope the answer is clear.

discovered, it was a puzzle that the decay $K_L \rightarrow \mu^+ \mu^-$ was not observed. The GIM mechanism provided an answer. The fact that the CKM is unitary implies that this process is a one loop process and there is an extra suppression of order m_c^2/m_W^2 to the amplitude. Thus, the rate is tiny and very hard to observe.

The GIM mechanism is also important in understanding the finiteness of loop amplitudes. Any one loop amplitude corresponding to decay where the tree level amplitude is zero must be finite. Technically, this can be seen by noticing that if it were divergence, a counter term at tree-level would be needed, but that cannot be the case if the tree-level amplitude vanishes. The amplitude for $b \rightarrow s\gamma$ it is naively log divergent. (Make sure you do the counting and see it for yourself.) Yet, it is only the m_i independent term that diverges. The GIM mechanism is here to save us as it guarantees that this term is zero. The m_i dependent term is finite, as it should be.

One more important point about the GIM mechanism is the fact that the amplitude is proportional to the mass squared of the internal quark. This implies that the total amplitude is more sensitive to couplings of the heavy quarks. In B decays, the heaviest internal quark is the top quark. This is the reason that $b \rightarrow s\gamma$ is sensitive to $|V_{ts}V_{tb}|$. This is a welcome feature since, as we mentioned before, these elements are hard to probe directly.

In one loop decays of kaons, there is a “competition” between the internal top and charm quarks. The top is heavier, but the CKM couplings of the charm are larger. Numerically, the charm is the winner, but not by a large margin. Check for yourself.

As for charm decay, since the tree level decay amplitudes are large, and since there is no heavy internal quark, the loop decay amplitudes are highly suppressed. So far the experimental bounds on various loop-mediated charm decays are far above the SM predictions. As an exercise, try to determine which internal quark dominates the one loop charm decay.

3.4 Homework

Question 5: Direct CKM measurements from D decays

The ratio of CKM elements

$$r \equiv \frac{|V_{cd}|}{|V_{cs}|} \quad (49)$$

can be estimated assuming SU(3) flavor symmetry. The idea is that in the SU(3) limit the pion and the kaon have the same mass and the same hadronic matrix elements.

1. Construct a ratio of semileptonic D decays that can be use to measure the ratio r .
 2. We usually expect SU(3) breaking effects to be of the order $m_s/\Lambda_{QCD} \sim 20\%$. Compare the observable you constructed to the actual measurement and estimate the SU(3) breaking effect.
-

Question 6: The GIM mechanism: $b \rightarrow s\gamma$ decay

1. Explain why $b \rightarrow s\gamma$ is a loop decay and draw the one loop diagrams in the SM.
2. Naively, these diagrams diverge. Show this.
3. Once we add all the diagrams and make use of the CKM unitarity, we get a finite result. Show that the UV divergences cancel (that is, put all masses the same and show that the answer is zero).
4. We now add a vector-like pair of down type quarks to the SM which we denote by b'

$$b'_R(3, 1)_{-1/3}, \quad b'_L(3, 1)_{-1/3}. \quad (50)$$

Show that in that model Eq. (48) is not valid anymore, that is,

$$\sum_{i=u,c,t} V_{ib} V_{is}^* \neq 0, \quad (51)$$

and that we have a Z exchange tree level FCNCs in the down sector. (The name “vector-like” refers to the case where the left and right handed fields have the same representation under all gauge groups. This is in contrast to a chiral pair where they have different representations. All the SM fermions are chiral.)

5. As we argued, in any model we cannot have $b \rightarrow s\gamma$ at tree level. Thus, in the model with the vector-like quarks, the one loop diagrams must also be finite. Yet, in the SM we used Eq. (48) to argue that the amplitude is finite, but now it is not valid. Show that the amplitude is finite also in this case. (Hint: When you have an infinite result that should be finite the reason is usually that there are more diagrams that you forgot.)

4 Meson mixing

Another interesting FCNC process is neutral meson mixing. Since it is an FCNC process, it cannot be mediated at tree level in the SM, and thus it is related to the “indirect measurements” class of CKM measurements. Yet, the importance of meson mixing and oscillation goes far beyond CKM measurements and we study it in some detail.

4.1 Formalism

There are four neutral mesons that can mix: K , D , B , and B_s .⁹ We first study the general formalism and then the interesting issues in each of the systems. The formalism is that of a two body open system. That is, the system involves the meson states P^0 and $\bar{P}^0$, and all the states they can decay to. Before the meson decays the state can be a coherent superposition of the two meson states. Once the decay happens, coherence is practically lost. This allows us to describe the decays using a non-Hermitian Hamiltonian, like we do for an open system.

We consider a general meson denoted by P . At $t = 0$ is in an initial state

$$|\psi(0)\rangle = a(0)|P^0\rangle + b(0)|\bar{P}^0\rangle, \quad (52)$$

where we are interested in computing the values of $a(t)$ and $b(t)$. Under our assumptions all the evolution is determined by a 2×2 effective Hamiltonian $\mathcal{H}$ that is not Hermitian. Any complex matrix, such as $\mathcal{H}$, can be written in terms of Hermitian matrices M and Γ as

$$\mathcal{H} = M - \frac{i}{2} \Gamma. \quad (53)$$

M and Γ are associated with $(P^0, \bar{P}^0) \leftrightarrow (P^0, \bar{P}^0)$ transitions via off-shell (dispersive) and on-shell (absorptive) intermediate states, respectively. Diagonal elements of M and Γ are associated with the flavor-conserving transitions $P^0 \rightarrow P^0$ and $\bar{P}^0 \rightarrow \bar{P}^0$ while off-diagonal elements are associated with flavor-changing transitions $P^0 \leftrightarrow \bar{P}^0$.

If $\mathcal{H}$ is not diagonal, the meson states, P^0 and $\bar{P}^0$ are not mass eigenstates, and thus do not have well defined masses and widths. It is only the eigenvectors of $\mathcal{H}$ that have well defined masses and decay widths. We denote the light and heavy eigenstates as P_L and P_H with $m_H > m_L$. (Another possible

⁹You may be wondering why there are only four meson mixing systems. If you do not wonder and do not know the answer, then you should wonder. We will answer this question shortly.

choice, which is standard for K mesons, is to define the mass eigenstates according to their lifetimes: K_S for the short-lived and K_L for the long-lived state. The K_L is experimentally found to be the heavier state.) Note that since $\mathcal{H}$ is not Hermitian, the eigenvectors do not need to be orthogonal to each other. Due to CPT, $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$. Then when we solve the eigenvalue problem for $\mathcal{H}$ we find that the eigenstates are given by

$$|P_{L,H}\rangle = p|P^0\rangle \pm q|\bar{P}^0\rangle, \quad (54)$$

with the normalization $|p|^2 + |q|^2 = 1$ and

$$\left(\frac{q}{p}\right)^2 = \frac{M_{12}^* - (i/2)\Gamma_{12}^*}{M_{12} - (i/2)\Gamma_{12}}. \quad (55)$$

If CP is a symmetry of $\mathcal{H}$ then M_{12} and Γ_{12} are relatively real, leading to

$$\left|\frac{q}{p}\right| = 1, \quad (56)$$

where the phase of q/p is unphysical. In that case the mass eigenstates are orthogonal

$$\langle P_H | P_L \rangle = |p|^2 - |q|^2 = 0. \quad (57)$$

The real and imaginary parts of the eigenvalues of $\mathcal{H}$ corresponding to $|P_{L,H}\rangle$ represent their masses and decay-widths, respectively. The mass difference Δm and the width difference $\Delta\Gamma$ are defined as follows:

$$\Delta m \equiv M_H - M_L, \quad \Delta\Gamma \equiv \Gamma_H - \Gamma_L. \quad (58)$$

Note that here Δm is positive by definition, while the sign of $\Delta\Gamma$ is to be determined experimentally. (Alternatively, one can use the states defined by their lifetimes to have $\Delta\Gamma \equiv \Gamma_S - \Gamma_L$ positive by definition.) The average mass and width are given by

$$m \equiv \frac{M_H + M_L}{2}, \quad \Gamma \equiv \frac{\Gamma_H + \Gamma_L}{2}. \quad (59)$$

It is useful to define dimensionless ratios x and y :

$$x \equiv \frac{\Delta m}{\Gamma}, \quad y \equiv \frac{\Delta\Gamma}{2\Gamma}. \quad (60)$$

We also define

$$\theta = \arg(M_{12}\Gamma_{12}^*). \quad (61)$$

Solving the eigenvalue equation gives

$$(\Delta m)^2 - \frac{1}{4}(\Delta\Gamma)^2 = (4|M_{12}|^2 - |\Gamma_{12}|^2), \quad \Delta m \Delta\Gamma = 4\mathcal{R}e(M_{12}\Gamma_{12}^*). \quad (62)$$

In the limit of CP conservation, Eq. (62) is simplified to

$$\Delta m = 2|M_{12}|, \quad |\Delta\Gamma| = 2|\Gamma_{12}|. \quad (63)$$

4.2 Time evolution

We move on to study the time evolution of a neutral meson. For simplicity, we assume CP conservation. Later on, when we study CP violation, we will relax this assumption, and study the system more generally. Many important points, however, can be understood in the simplified case when CP is conserved and so we use it here.

In the CP limit $|q| = |p| = 1/\sqrt{2}$ and we can choose the relative phase between p and q to be zero. In that case the transformation from the flavor to the mass basis, Eq. (54), is simplified to

$$|P_{L,H}\rangle = \frac{1}{\sqrt{2}} (|P^0\rangle \pm |\bar{P}^0\rangle). \quad (64)$$

We denote the state of an initially pure $|P^0\rangle$ after an time t as $|P^0(t)\rangle$ (and similarly for or $|\bar{P}^0\rangle$). We obtain

$$|P^0(t)\rangle = \cos\left(\frac{\Delta E t}{2}\right) |P^0\rangle + i \sin\left(\frac{\Delta E t}{2}\right) |\bar{P}^0\rangle, \quad (65)$$

and similarly for $|\bar{P}^0(t)\rangle$. Since flavor is not conserved, the probability to measure a specific flavor, that is P or $\bar{P}$, oscillates in time, and it is given by

$$\begin{aligned} \mathcal{P}(P \rightarrow P)[t] &= |\langle P^0(t)|P^0\rangle|^2 = \frac{1 + \cos(\Delta E t)}{2}, \\ \mathcal{P}(P \rightarrow \bar{P})[t] &= |\langle P^0(t)|\bar{P}^0\rangle|^2 = \frac{1 - \cos(\Delta E t)}{2}, \end{aligned} \quad (66)$$

where $\mathcal{P}$ denotes probability.

A few remarks are in order:

- In the meson rest frame, $\Delta E = \Delta m$ and $t = \tau$, the proper time.
- We learn that we have flavor oscillation with frequency Δm . This is the parameter that eventually gives us the sensitivity to the weak interaction and to flavor.
- We learn that by measuring the oscillation frequency we can determine the mass splitting between the two mass eigenstates. One way this can be done is by measuring the flavor of the meson both at production and decay. It is not trivial to measure the flavor at both ends, and we do not describe it in detail here, but you are encouraged to think and learn about how it can be done.

4.3 Time scales

Next, we spend some time understanding the different time scales that are involved in meson mixing. One scale is the oscillation period. As can be seen from Eq. (66), the oscillation time scale is given by Δm .¹⁰

Before we talk about the other time scales we have to understand how the flavor is measured, or as we usually say it, tagged. By “flavor is tagged” we refer to the decay as a flavor vs anti-flavor, for example b vs $\bar{b}$. Of course, in principle, we can tag the flavor at any time. In practice, however, the measurement is done for us by Nature. That is, the flavor is tagged when the meson decays. In fact, it is done only when the meson decays in a flavor specific way. Other decays that are common to both P and $\bar{P}$ do not measure the flavor. Such decays are also very useful as we will discuss later. Semi-leptonic decays are very good flavor tags:

$$b \rightarrow c\mu^-\bar{\nu}, \quad \bar{b} \rightarrow \bar{c}\mu^+\nu. \quad (67)$$

The charge of the lepton tells us the flavor: a μ^+ tells us that we “measured” a b flavor, while a μ^- indicates a $\bar{b}$. Of course, before the meson decays it could be in a superposition of a b and a $\bar{b}$. The decay acts as a quantum measurement. In the case of semileptonic decay, it acts as a measurement of flavor vs anti-flavor.

Aside from the oscillation time, one other time scale that is involved is the time when the flavor measurements is done. Since the flavor is tagged when the meson decays, the relevant time scale is the decay width, Γ . We can then use the dimensionless quantity, $x \equiv \Delta m/\Gamma$, defined in Eq. (60), to understand the relevance of these two time scales. There are three relevant regimes:

¹⁰What we refer to here is, of course, $1/\Delta m$. Yet, at this stage of our life as physicists, we know how to match dimensions, and thus I interchange between time and energy freely, counting on you to understand what I am referring to.

1. $x \ll 1$. We denote this case as “slow oscillation”. In that case the meson has no time to oscillate, and thus to good approximation flavor is conserved. In practice, this implies that $\cos(\Delta mt) \approx 1$ and using it in Eq. (66) we see that $\mathcal{P}(P \rightarrow P) \approx 1$ and $\mathcal{P}(P \rightarrow \bar{P}) \rightarrow 0$. In this case, an upper bound on the mass difference is likely to be established before an actual measurement. This case is relevant for the D system.
2. $x \gg 1$. We denote this case as “fast oscillation”. In this case the meson oscillates many times before decaying, and thus the oscillating term practically averaged out to zero.¹¹ In practice in this case $\mathcal{P}(P \rightarrow P) \approx \mathcal{P}(P \rightarrow \bar{P}) \approx 1/2$ and a lower bound on Δm can be established before a measurement can be done. This case is relevant for the B_s system.
3. $x \sim 1$. In this case the oscillation and decay times are roughly the same. That is, the system has time to oscillate and the oscillation are not averaged out. In a way, this is the most interesting case since then it is relatively easy to measure Δm . Amazingly, this case is relevant to the K and B systems. We emphasize that the physics that leads to Γ and Δm are unrelated, so there is no reason to expect $x \sim 1$. Yet, Nature is kind enough to produce $x \sim 1$ in two out of the four neutral meson systems.

It is amusing to point out that oscillations give us sensitivity to mass differences of the order of the width, which are much smaller than the mass itself. In fact, we have been able to measure mass differences that are 14 orders of magnitude smaller than the corresponding masses. It is due to the quantum mechanical nature of the oscillation that such high precision can be achieved.

In some cases there is one more time scale: $\Delta\Gamma$. In such cases, we have one more relevant dimensionless parameter $y \equiv \Delta\Gamma/(2\Gamma)$. Note that y is bounded, $-1 \leq y \leq 1$. (This is in contrast to x which is bounded by $x > 0$.) Thus, we can talk about several cases depending on the values of y and x .

1. $|y| \ll 1$ and $y \ll x$. In this case the width difference is irrelevant. This is the case for the B system.
2. $y \sim x$. In this case the width difference is as important as the oscillation. This is the case in the D system where $y \ll 1$ and for the K system with $y \sim 1$.
3. $|y| \sim 1$ and $y \ll x$. In this case the oscillation averages out and the width difference shows up as a difference in the lifetime of the two mass eigenstates. This case may be relevant to the B_s system, where we expect $y \sim 0.1$.

There are few other limits (like $y \gg x$) that are not realized in the four meson systems. Yet, they might be realized in some other systems yet to be discovered.

To conclude this subsection we summarize the experimental data on meson mixing

$$\begin{aligned}
 x_K &\sim 1, & y_K &\sim 1, \\
 x_D &\sim 10^{-2}, & y_D &\sim 10^{-2}, \\
 x_d &\sim 1, & y_d &\lesssim 10^{-2}, \\
 x_s &\sim 10, & y_s &\lesssim 10^{-1}.
 \end{aligned} \tag{68}$$

Note that y_d and y_s have not been measured and all we have are upper bounds.

4.4 Calculation of the mixing parameters

We now explain how the calculation of the mixing parameters is done. We only briefly remark on $\Delta\Gamma$ and spend some time on the calculation of Δm . As we have done a few times, we will do the calculation in the SM, keeping in mind that the tools we develop can be used in a large class of models.

¹¹This is the case we are very familiar with when we talk about decays into mass eigenstates. There is never a decay into a mass eigenstate. Only when the oscillations are very fast and the oscillatory term in the decay rate averages out, the result seems like the decay is into a mass eigenstate.

In order to calculate the mass and width differences, we need to know the effective Hamiltonian, $\mathcal{H}$, defined in Eq. (53). For the diagonal terms, no calculations are needed. CPT implies $M_{11} = M_{22}$ and to an excellent approximation it is just the mass of the meson. Similarly, $\Gamma_{11} = \Gamma_{22}$ is the average width of the meson. What we need to calculate is the off diagonal terms, that is M_{12} and Γ_{12} .

We start by discussing M_{12} . For the sake of simplicity we consider the B meson as a concrete example. The first point to note is that M_{12} is basically the transition amplitude between a B and a $\bar{B}$ at zero momentum transfer. In terms of states with the conventional normalization we have

$$M_{12} = \frac{1}{2m_B} \langle B | \mathcal{O} | \bar{B} \rangle. \quad (69)$$

We emphasize that we should not square the amplitude. We square amplitudes to get transition probabilities and decay rates, which is not the case here.

The operator that appears in Eq. (69) is one that can create a B and annihilate a $\bar{B}$. Recalling that a B meson is made of a $\bar{b}$ and d quark (and $\bar{B}$ from b and $\bar{d}$), we learn that in terms of quarks it must be of the form

$$\mathcal{O} \sim (\bar{b} d)(\bar{b} d). \quad (70)$$

(We do not explicitly write the Dirac structure. Anything that does not vanish is possible.) Since the operator in Eq. (70) is an FCNC operator, in the SM it cannot be generated at tree level and must be generated at one loop. The one loop diagram that generates it is called “a box diagram”, because it looks like a square. It is given in Fig. 3. The calculation of the box diagram is straightforward and we end up with

$$M_{12} \propto \frac{g^4}{m_W^2} \langle B | (\bar{b}_L \gamma_\mu d_L)(\bar{b}_L \gamma^\mu d_L) | \bar{B} \rangle \sum_{i,j} V_{id}^* V_{ib} V_{jd}^* V_{jb} F(x_i, x_j), \quad (71)$$

such that

$$x_i \equiv \frac{m_i^2}{m_W^2}, \quad i = u, c, t, \quad (72)$$

and the function F is known, but we do not write it here.

Several points are in order

1. The box diagram is second order in the weak interaction, that is, it is proportional to g^4 .
2. The fact that the CKM is unitary (in other words, the GIM mechanism) makes the m_i independent term vanish and to a good approximation $\sum_{i,j} F(x_i, x_j) \rightarrow F(x_t, x_t)$. We then say that it is the top quark that dominates the loop.
3. The last thing we need is the hadronic matrix element, $\langle B | (\bar{b}_L \gamma_\mu d_L)(\bar{b}_L \gamma^\mu d_L) | \bar{B} \rangle$. The problem is that the operator creates a free b and $\bar{d}$ quark and annihilates a free $\bar{b}$ and a d . This is not the same as creating a $\bar{B}$ meson and annihilating a B meson. Here, lattice QCD helps and by now a good estimate of the matrix element is available.
4. Similar calculations can be done for the other mesons. Due to the GIM mechanism, for the K meson the function F gives an extra m_c^2/m_W^2 suppression.
5. Last we mention the calculation of Γ_{12} . An estimate of it can be made by looking at the on-shell part of the box diagram. Yet, once particle goes on shell, QCD becomes important, and the theoretical uncertainties in the calculation of Γ_{12} are larger than that of M_{12} .

Putting all the pieces together we see how the measurement of the mass difference is sensitive to some combination of CKM elements. Using the fact that the amplitude is proportional to the heaviest internal quark we get from Eqs. (71) and (63)

$$\Delta m_B \propto |V_{tb} V_{td}|^2, \quad (73)$$

where the proportionality constant is known with an uncertainty at the 10% level.

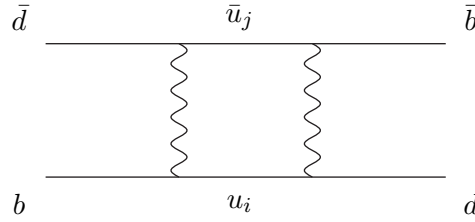


Fig. 3: A box diagram that generate an operators that can lead to $B \leftrightarrow \bar{B}$ transition.

4.5 Homework

Question 7: The four mesons

It is now time to come back and ask why there are only four mesons that we care about when discussing oscillations. In particular, why we do not talk about oscillation for the following systems

1. $B^+ - B^-$ oscillation
2. $K - K^*$ oscillation
3. $T - \bar{T}$ oscillation (a T is a meson made out of a t and a $\bar{u}$ quarks.)
4. $K^* - \bar{K}^*$ oscillation

Hint: The last three cases all have to do with time scales. In principle there are oscillations in these systems, but they are irrelevant.

Question 8: Kaons

Here we study some properties of the kaon system. We did not talk about it at all. You have to go back and recall (or learn) how kaons decay, and combine that with what we discussed in the lecture.

1. Explain why $y_K \approx 1$.
 2. In a hypothetical world where we could change the mass of the kaon without changing any other masses, how would the value of y_K change if we made m_K smaller or larger.
-

Question 9: Mixing beyond the SM

Consider a model without a top quark, in which the first two generations are as in the SM, while the left-handed bottom (b_L) and the right-handed bottom (b_R) are $SU(2)$ singlets.

1. Draw a tree-level diagram that contributes to $B - \bar{B}$ mixing in this model.
 2. Is there a tree-level diagram that contributes to $K - \bar{K}$ mixing?
 3. Is there a tree-level diagram that contributes to $D - \bar{D}$ mixing?
-

5 CP violation

As we already mentioned, it turns out that in Nature CP violation is closely related to flavor. In the SM, this is manifested by the fact that the source of CP violation is the phase of the CKM matrix. Thus, we will spend some time learning about CP violation in the SM and beyond.

5.1 How to observe CP violation?

CP is the symmetry that relates particles with their anti-particles. Thus, if CP is conserved, we must have

$$\Gamma(A \rightarrow B) = \Gamma(\bar{A} \rightarrow \bar{B}), \quad (74)$$

such that A and B represent any possible initial and final states. From this we conclude that one way to find CP violation is to look for processes where

$$\Gamma(A \rightarrow B) \neq \Gamma(\bar{A} \rightarrow \bar{B}). \quad (75)$$

This, however, is not easy. The reason is that even when CP is not conserved, Eq. (74) can hold to a very high accuracy in many cases. So far there are only very few cases where Eq. (74) does not hold to a measurable level. The reason that it is not easy to observe CP violation is that there are several conditions that have to be fulfilled. CP violation can arise only in interference between two decay amplitudes. These amplitudes must carry different weak and strong phases (we explain below what these phases are). Also, CPT implies that the total width of a particle and its anti-particle are the same. Thus, any CP violation in one channel must be compensated by CP violation with an opposite sign in another channel. Finally, it happens that in the SM, which describes Nature very well, CP violation comes only when we have three generations, and thus any CP violating observable must involve all the three generations. Due to the particular hierarchical structure of the CKM matrix, all CP violating observables are proportional to very small CKM elements.

In order to show this we start by defining weak and strong phases. Consider, for example, the $B \rightarrow f$ decay amplitude A_f , and the CP conjugate process, $\bar{B} \rightarrow \bar{f}$, with decay amplitude $\bar{A}_{\bar{f}}$. There are two types of phases that may appear in these decay amplitudes. Complex parameters in any Lagrangian term that contributes to the amplitude will appear in complex conjugate form in the CP-conjugate amplitude. Thus, their phases appear in A_f and $\bar{A}_{\bar{f}}$ with opposite signs and these phases are CP odd. In the SM, these phases occur only in the couplings of the $W^\pm$ bosons and hence CP odd phases are often called “weak phases.”

A second type of phases can appear in decay amplitudes even when the Lagrangian is real. They are from possible contributions of intermediate on-shell states in the decay process. These phases are the same in A_f and $\bar{A}_{\bar{f}}$ and are therefore CP even. One type of such phases is easy to calculate. It comes from the trivial time evolution, $\exp(iEt)$. More complicated cases are where there is rescattering due to the strong interactions. For this reason these phases are called “strong phases.”

There is one more kind of phases in addition to the weak and strong phases discussed here. These are the spurious phases that arise due to an arbitrary choice of phase convention, and do not originate from any dynamics. For simplicity, we set these unphysical phases to zero from now on.

It is useful to write each contribution a_i to A_f in three parts: its magnitude $|a_i|$, its weak phase ϕ_i , and its strong phase δ_i . If, for example, there are two such contributions, $A_f = a_1 + a_2$, we have

$$\begin{aligned} A_f &= |a_1|e^{i(\delta_1+\phi_1)} + |a_2|e^{i(\delta_2+\phi_2)}, \\ \bar{A}_{\bar{f}} &= |a_1|e^{i(\delta_1-\phi_1)} + |a_2|e^{i(\delta_2-\phi_2)}. \end{aligned} \quad (76)$$

Similarly, for neutral meson decays, it is useful to write

$$M_{12} = |M_{12}|e^{i\phi_M}, \quad \Gamma_{12} = |\Gamma_{12}|e^{i\phi_\Gamma}. \quad (77)$$

Each of the phases appearing in Eqs. (76) and (77) is convention dependent, but combinations such as $\delta_1 - \delta_2$, $\phi_1 - \phi_2$, and $\phi_M - \phi_\Gamma$ are physical. Now we can see why in order to observe CP violation we need two different amplitudes with different weak and strong phases. It is easy to show and I leave it for the homework.

A few remarks are in order:

1. The basic idea in CP violation research is to find processes where we can measure CP violation. That is, we look for processes with two decay amplitudes that are roughly of the same size with different weak and strong phases.
2. In some cases, we can get around QCD. In such cases, we get sensitivity to the phases of the unitarity triangle (or, equivalently, of the CKM matrix). These cases are the most interesting ones.
3. Some observables are sensitive to CP phases without measuring CP violation. That is like saying that we can determine the angles of a triangle just by knowing the lengths of its sides.
4. While we talk only about CP violation in meson oscillations and decays, there are more types of CP violating observables. In particular, triple products and electric dipole moments (EDMs) of elementary particles encode CP violation. They are not directly related to flavor, and are not covered here.
5. So far CP violation has been observed only in meson decays, particularly, in K_L , B_d and $B^\pm$ decays. In the following, we concentrate on the formalism relevant to these systems.

5.2 The three types of CP violation

When we consider CP violation in meson decays there are two types of amplitudes: mixing and decay. Thus, there must be three ways to observe CP violation, depending on which type of amplitudes interfere. Indeed, this is the case. We first introduce the three classes and then discuss each of them in some length.

1. CP violation in decay, also called direct CP violation. This is the case when the interference is between two decay amplitudes. The necessary strong phase is due to rescattering.
2. CP violation in mixing, also called indirect CP violation. In this case the absorptive and dispersive mixing amplitudes interfere. The strong phase is due to the time evolution of the oscillation.
3. CP violation in interference between mixing and decay. As the name suggests, here the interference is between the decay and the oscillation amplitudes. The dominant effect is due to the dispersive mixing amplitude (the one that gives the mass difference) and a leading decay amplitude. Also here the strong phase is due to the time evolution of the oscillation.

In all of the above cases the weak phase comes from the Lagrangian. In the SM these weak phases are related to the CKM phase. In many cases, the weak phase is one of the angles of the unitarity triangle.

5.3 CP violation in decay

We first talk about CP violation in decay. This is the case when

$$|A(P \rightarrow f)| \neq |A(\bar{P} \rightarrow \bar{f})|. \quad (78)$$

The way to measure this type of CP violation is as follows. We define

$$a_{CP} \equiv \frac{\Gamma(\bar{B} \rightarrow \bar{f}) - \Gamma(B \rightarrow f)}{\Gamma(\bar{B} \rightarrow \bar{f}) + \Gamma(B \rightarrow f)} = \frac{|\bar{A}/A|^2 - 1}{|\bar{A}/A|^2 + 1}. \quad (79)$$

Using Eq. (76) with ϕ as the weak phase difference and δ as the strong phase difference, we write

$$A(P \rightarrow f) = A(1 + r \exp[i(\phi + \delta)]), \quad A(\bar{P} \rightarrow \bar{f}) = A(1 + r \exp[i(-\phi + \delta)]), \quad (80)$$

with $r \leq 1$. We get

$$a_{CP} = r \sin \phi \sin \delta. \quad (81)$$

This result shows explicitly that we need two decay amplitudes, that is, $r \neq 0$, with different weak phases, $\phi \neq 0, \pi$ and different strong phases $\delta \neq 0, \pi$.

A few remarks are in order:

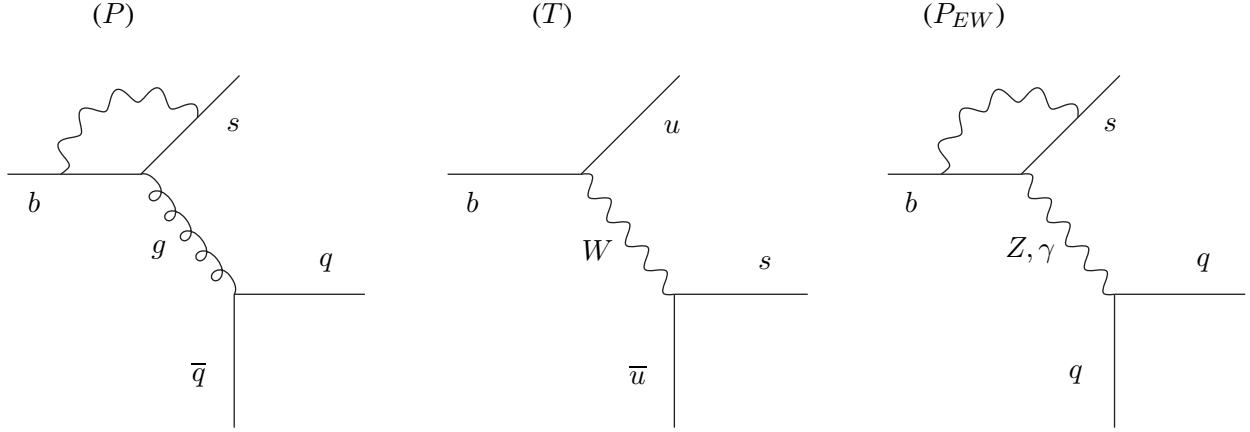


Fig. 4: The $B \rightarrow K\pi$ amplitudes. The dominant one is the strong penguin amplitude (P), and the sub-dominant ones are the tree amplitude (T) and the electroweak penguin amplitude (P_{EW}).

1. In order to have a large effect we need each of the three factors in Eq. (81) to be large.
2. CP violation in decay can occur in both charged and neutral mesons. One complication for the case of neutral meson is that it is not always possible to tell the flavor of the decaying meson, that is, if it is P or $\bar{P}$. This can be a problem or a virtue.
3. In general the strong phase is not calculable since it is related to QCD. This may not be a problem if all we are after is to demonstrate CP violation. In other cases the phase can be independently measured, eliminating this particular source of theoretical error.

5.3.1 $B \rightarrow K\pi$

Our first example of CP violation in decay is $B^0 \rightarrow K^+\pi^-$. At the quark level the decay is mediated by $b \rightarrow s\bar{u}u$ transition. There are two dominant decay amplitudes, tree level and one loop penguin diagrams.¹² Two penguin diagrams and the tree level diagram are plotted in Fig. 4.

Naively, tree diagrams are expected to dominate. Yet, this is not the case here. The reason is that the tree diagram is highly CKM suppressed. It turns out that this suppression is stronger than the loop suppression such that $r = |P/T| \sim 0.3$. (Here we use P and T to denote the penguin and tree amplitudes.) In terms of weak phases, the tree amplitude carries the phase of $V_{ub}V_{us}^*$. The dominant internal quark in the penguin diagram is the top quark and thus to first approximation the phase of the penguin diagram is the phase of $V_{tb}V_{ts}^*$, and to first approximation $\phi = \alpha$. As for the strong phase, we cannot calculate it, and there is no reason for it to vanish since the two amplitudes have different structure. Experimentally, CP violation in $B \rightarrow K\pi$ decays has been established. It was the first observation of CP violation in decay.

We remark that $B \rightarrow K\pi$ decays have much more to offer. There are four different such decays, and they are all related by isospin, and thus many predictions can be made. Moreover, the decay rates are relatively large and the measurements have been performed. The full details are beyond the scope of these lectures, but you are encouraged to go and study them.

¹²This is the first time we introduce the name penguin. It is just a name, and it refers to one loop amplitude of the form $f_1 \rightarrow f_2 B$ where B is a neutral boson that can be on-shell or off-shell. If the boson is a gluon we may call it QCD penguin. When it is a photon or a Z boson it is called electroweak penguin.

5.3.2 $B \rightarrow DK$

Our second example is $B \rightarrow DK$ decay. This decay involves only tree level diagrams, and is sensitive to the phase between the $b \rightarrow c\bar{u}s$ and $b \rightarrow u\bar{c}s$ decay amplitude, which is γ . The situation here is involved as the D further decays and what is measured is $B \rightarrow f_D K$, where f_D is a final state that comes from a D or $\bar{D}$ decay. This “complication” turns out to be very important. It allows us to construct theoretically very clean observables. In fact, $B \rightarrow DK$ decays are arguably the cleanest measurement of a CP violation phase in terms of theoretical uncertainties.

The reason for this theoretical cleanliness is that all the necessary hadronic quantities can be extracted experimentally. We consider decays of the type

$$B \rightarrow D(\bar{D}) K(X) \rightarrow f_D K(X), \quad (82)$$

where f_D is a final state that can be accessible from both D and $\bar{D}$ and X represents possible extra particles in the final state. The crucial point is that in the intermediate state the flavor is not measured. That is, the state is in general a coherent superposition of D and $\bar{D}$. On the other hand, this state is on-shell so that the $B \rightarrow D$ and $D \rightarrow f_D$ amplitudes factorize. Thus, we have quantum coherence and factorization at the same time. The coherence makes it possible to have interference and thus sensitivity to CP violating phases. Factorization is important since then we can separate the decay chain into stages such that each stage can be determined experimentally. The combination is then very powerful, we have a way to probe CP violation without the need to calculate any decay amplitude.

To see the power of the method, consider using $B \rightarrow DKX$ decays with n different X states, and $D \rightarrow f_D$ with k different f_D states, one can perform $n \times k$ measurements. Because the B and D decay amplitude factorize, these $n \times k$ measurements depend on $n + k$ hadronic decay amplitudes. For large enough n and k , there is a sufficient number of measurements to determine all hadronic parameters, as well as the weak phase we are after. Since all hadronic matrix elements can be measured, the theoretical uncertainties are much below the sensitivity of any foreseeable future experiment.

5.4 CP violation that involves mixing

We move on to study CP violation that involves mixing. This kind of CP violation is the one that was first discovered in the kaon system in the 1960s, and in the B system more recently. They are the ones that shape our understanding of the picture of CP violation in the SM, and thus, they deserve some discussion.

We start by re-deriving the oscillation formalism in a more general case where CP violation is included. Then we will be able to construct some CP violating observables and see how they are related to the phases of the unitarity triangle. For simplicity we concentrate on the B system. We allow the decay to be into an arbitrary state, that is, a state that can come from any mixture of B and $\bar{B}$. Consider a final state f such that

$$A_f \equiv A(B \rightarrow f), \quad \bar{A}_f \equiv A(\bar{B} \rightarrow f). \quad (83)$$

We further define

$$\lambda_f \equiv \frac{q}{p} \frac{\bar{A}_f}{A_f}. \quad (84)$$

We consider the general time evolution of a P^0 and $\bar{P}^0$ mesons. It is given by

$$\begin{aligned} |P^0(t)\rangle &= g_+(t) |P^0\rangle - (q/p) g_-(t) |\bar{P}^0\rangle, \\ |\bar{P}^0(t)\rangle &= g_+(t) |\bar{P}^0\rangle - (p/q) g_-(t) |P^0\rangle, \end{aligned} \quad (85)$$

where we work in the B rest frame and

$$g_{\pm}(t) \equiv \frac{1}{2} \left(e^{-im_H t - \frac{1}{2}\Gamma_H t} \pm e^{-im_L t - \frac{1}{2}\Gamma_L t} \right). \quad (86)$$

We define $\tau \equiv \Gamma t$ and then the decay rates are

$$\begin{aligned}\Gamma(B \rightarrow f)[t] &= |A_f|^2 e^{-\tau} \left\{ (\cosh y\tau + \cos x\tau) + |\lambda_f|^2 (\cosh y\tau - \cos x\tau) \right. \\ &\quad \left. - 2\text{Re} [\lambda_f (\sinh y\tau + i \sin x\tau)] \right\}, \\ \Gamma(\bar{B} \rightarrow f)[t] &= |\bar{A}_f|^2 e^{-\tau} \left\{ (\cosh y\tau + \cos x\tau) + |\lambda_f|^{-2} (\cosh y\tau - \cos x\tau) \right. \\ &\quad \left. - 2\text{Re} [\lambda_f^{-1} (\sinh y\tau + i \sin x\tau)] \right\},\end{aligned}\quad (87)$$

where $\Gamma(B \rightarrow f)[t]$ ($\Gamma(\bar{B} \rightarrow f)[t]$) is the probability for an initially pure B ($\bar{B}$) meson to decay at time t to a final state f .

Terms proportional to $|A_f|^2$ or $|\bar{A}_f|^2$ are associated with decays that occur without any net oscillation, while terms proportional to $|\lambda|^2$ or $|\lambda|^{-2}$ are associated with decays following a net oscillation. The $\sinh(y\tau)$ and $\sin(x\tau)$ terms in Eqs. (87) are associated with the interference between these two cases. Note that, in multi-body decays, amplitudes are functions of phase-space variables. The amount of interference is in general a function of the kinematics, and can be strongly influenced by resonant substructure. Equations. (87) are much simplified in the case where $|q/p| = 1$ and $|A_f/\bar{A}_f| = 1$. In that case $|\lambda| = 1$ is a pure phase.

We define the CP observable of asymmetry of neutral meson decays into final state f

$$\mathcal{A}_f(t) \equiv \frac{\Gamma[\bar{B}(t) \rightarrow \bar{f}] - \Gamma[B(t) \rightarrow f]}{\Gamma[\bar{B}(t) \rightarrow \bar{f}] + \Gamma[B(t) \rightarrow f]}.\quad (88)$$

We restrict ourself to the case where f is a CP eigenstates, and then $f = \bar{f}$. Also consider the case where there is no CP violation in the decay, and then the decay amplitudes fulfill $|\bar{A}_f| = |A_f|$. We further consider the case where $\Delta\Gamma = 0$ and $|q/p| = 1$, as expected to a good approximation for B system. In that case the interference between decays with and without mixing is the only source of the asymmetry and we get

$$\mathcal{A}_f(t) = \text{Im}(\lambda_f) \sin(x\tau) = \sin[\arg(\lambda_f)] \sin(\Delta m t).\quad (89)$$

where in the last step we used $|\lambda| = 1$. We see that once we know Δm , and if the above conditions are satisfied, we have a clean measurement of the phase of λ_f . This phase is directly related to an angle in the unitarity triangle, as we discuss shortly.

It is instructive to describe the effect of CP violation in decays of the mass eigenstates. For cases where the width difference is negligible, this is usually not very useful. It is not easy to generate, for example, a B_H mass eigenstate. When the width difference is large, like in the kaon system, this representation can be very useful as we do know how to generate K_L states. We assume $|q/p| = 1$ and then the decay into CP eigenstates is given by

$$\Gamma(P_L \rightarrow f_{CP}) = 2|A_f|^2 e^{-\Gamma_L t} \cos^2 \theta, \quad \Gamma(P_H \rightarrow f_{CP}) = 2|A_f|^2 e^{-\Gamma_H t} \sin^2 \theta.\quad (90)$$

where

$$\theta \equiv \frac{\arg \lambda}{2}.\quad (91)$$

In this case, CP violation, that is $\theta \neq 0$, is manifested by the fact that two non-degenerate states can decay to the same CP eigenstate final state.

We also consider decays into a pure flavor state. In that case $\lambda = 0$, and we can isolate the effect of CP violation in mixing. CP violation in mixing is defined by

$$|q/p| \neq 1.\quad (92)$$

This is the only source of CP violation in charged-current semileptonic neutral meson decays $P, \bar{P} \rightarrow \ell^\pm X$. This is because we use $|A_{\ell^+ X}| = |\bar{A}_{\ell^- X}|$ and $A_{\ell^- X} = \bar{A}_{\ell^+ X} = 0$, which, to lowest order in G_F , is the case in the SM and in most of its extensions, and thus $\lambda = 0$.

This source of CP violation can be measured via the asymmetry of “wrong-sign” decays induced by oscillations:

$$\mathcal{A}_{\text{SL}}(t) \equiv \frac{\Gamma[\bar{B}(t) \rightarrow \ell^+ X] - \Gamma[B(t) \rightarrow \ell^- X]}{\Gamma[\bar{B}(t) \rightarrow \ell^+ X] + \Gamma[B(t) \rightarrow \ell^- X]} = \frac{1 - |q/p|^4}{1 + |q/p|^4}. \quad (93)$$

Note that this asymmetry of time-dependent decay rates is actually time independent.

We are now going to give few examples of cases that are sensitive to CP violation that involve mixing.

5.4.1 $B \rightarrow \psi K_S$

The “golden mode” with regard to CP violation in interference between mixing and decays is $B \rightarrow \psi K_S$. It provides a very clean determination of the angle β of the unitarity triangle.

As we already mentioned we know that to a very good approximation in the B system $|q/p| = 1$. In that case we have

$$\lambda_f = e^{-i\phi_B} \frac{\bar{A}_f}{A_f}, \quad (94)$$

where ϕ_B refers to the phase of M_{12} (see Eq. (77)). Within the SM, where the top diagram dominates the mixing, the corresponding phase factor is given to a very good approximation by

$$e^{-i\phi_B} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}. \quad (95)$$

For $f = \psi K$, which proceeds via a $\bar{b} \rightarrow \bar{c} c \bar{s}$ transition, we can write

$$A_{\psi K} = (V_{cb}^* V_{cs}) T_{\psi K} \quad (96)$$

where $T_{\psi K}$ is the magnitude of the tree amplitude. In principle there is also a penguin amplitude that contributes to the decay. The leading penguin carries the same weak phase as the tree amplitude. The one that carries a different weak phase is highly CKM suppressed and we neglect it. This is a crucial point. Because of that we have to a very good approximation

$$|\lambda_{\psi K_S}| = 1, \quad \text{Im}(\lambda_{\psi K_S}) = \sin 2\beta, \quad (97)$$

and Eq. (89) can be used for this case. We conclude that a measurement of the CP asymmetry in $B \rightarrow \psi K_S$ gives a very clean determination of the angle β of the unitarity triangle. Here we were able to overcome QCD by the fact that the decay is dominated by one decay amplitude that cancels once the CP asymmetry is constructed. As for the strong phase it arises due to the oscillation and it is related to the known Δm . This CP asymmetry measurement was done and is at present the most precise measurement of any angle or side of the unitarity triangle.

A subtlety arises in this decay that is related to the fact that B^0 decays into ψK^0 while $\bar{B}^0$ decays into $\psi \bar{K}^0$. A common final state, e.g. ψK_S , is reached only via $K^0 - \bar{K}^0$ mixing. We do not elaborate on this point.

There are many more decay modes where a clean measurement of angles can be performed. In your homework, you will work out one more example and even try to get an idea of the theoretical errors.

5.4.2 K decays

CP violation was discovered in K decays, and until recently, it was the only meson where CP violation had been measured. CP violation were first observed in $K_L \rightarrow \pi\pi$ decays in 1964, and later in semileptonic K_L decays in 1967. Beside the historical importance, kaon CP violation provides important bounds on the unitarity triangle. Moreover, when we consider generic new physics, CP violation in kaon decays provides the strongest bound on the scale of the new physics. This is a rather interesting result based on the amount of progress that has been made in our understanding of flavor and CP violation in the last 45 years.

While the formalism of CP violation is the same for all mesons, the relevant approximations are different. For the B system, we neglected the width difference and got the very elegant formula, Eq. (89). For the B mesons it is easy to talk in terms of flavor (or CP) eigenstates, and use mass eigenstates only as intermediate states to calculate the time evolutions. For kaons, however, the width difference is very large

$$\frac{\Gamma_S}{\Gamma_L} \sim 600. \quad (98)$$

This implies that, to very good approximation, we can get a state that is pure K_L . All we have to do is wait. Since we do have a pure K_L state, it is easy to talk in terms of mass eigenstates. Note that it is not easy to get a pure K_S state. At short times we have a mixture of states and, only after the K_S part has decayed, we have a pure K_L state.

In terms of mass eigenstates, CP violation is manifested if the same state can decay to both CP even and CP odd states. This should be clear to you from basic quantum mechanics. Consider a symmetry, that is, an operator that commutes with the Hamiltonian. In the case under consideration, if CP is a good symmetry it implies $[\text{CP}, H] = 0$. When this is the case, any non-degenerate state must be an eigenstate of CP. In a CP-conserving theory, any eigenstate of CP must decay to a state with the same CP parity. In particular, it is impossible to observe a state that can decay to both CP even and CP odd states. Thus, CP violation in kaon decays was established when $K_L \rightarrow \pi\pi$ was observed. K_L decays dominantly to three pions, which is a CP odd state. The fact that it decays also to two pions, which is a CP even state, implies CP violation.

We do not get into the details of the calculations, but it must be clear at this stage that the rate of $K_L \rightarrow \pi\pi$ must be related to the values of the CKM parameters and, in particular, to its phase. I hope you will find the time to read about it in one of the reviews I mentioned.

Before concluding, we remark on semileptonic CP violation in kaons. When working with mass eigenstates, CP conservation implies that

$$\Gamma(K_L \rightarrow \pi^- \ell^+ \nu) = \Gamma(K_L \rightarrow \pi^+ \ell^- \bar{\nu}). \quad (99)$$

(If it is not clear to you why CP implies the above, stop for a second and convince yourself.) Experimentally, the above equality was found to be violated, implying CP violation.

In principle, the CP violation in $K_L \rightarrow \pi\pi$ and in semileptonic decay are independent observables. Yet, when all the decay amplitudes carry the same phase, these two are related. This is indeed the case in the kaon system, and thus we talk about one parameter that measure kaon CP violation, which is denoted by ε_K . (Well, there is one more parameter called ε'_K , but we will not discuss it here.)

5.5 Homework

Question 10: Condition for CP violation

Using Eq. (76), show that in order to observe CP violation, $\Gamma(B \rightarrow f) \neq \Gamma(\bar{B} \rightarrow \bar{f})$, we need two amplitudes with different weak and strong phases.

Question 11: Mixing formalism

In this question, you are asked to develop the general formalism of meson mixing.

1. Show that the mass and width differences are given by

$$4(\Delta m)^2 - (\Delta \Gamma)^2 = 4(4|M_{12}|^2 - |\Gamma_{12}|^2), \quad \Delta m \Delta \Gamma = 4\text{Re}(M_{12}\Gamma_{12}^*), \quad (100)$$

and that

$$\left| \frac{q}{p} \right| = \left| \frac{\Delta m - i\Delta \Gamma/2}{2M_{12} - i\Gamma_{12}} \right|. \quad (101)$$

2. When CP is a good symmetry all mass eigenstates must also be CP eigenstates. Show that CP invariance requires

$$\left| \frac{q}{p} \right| = 1. \quad (102)$$

3. In the limit $\Gamma_{12} \ll M_{12}$ show that

$$\Delta m = 2|M_{12}|, \quad \Delta \Gamma = 2|\Gamma_{12}| \cos \theta, \quad \left| \frac{q}{p} \right| = 1. \quad (103)$$

4. Derive Eqs. (87).
5. Derive Eq. (89).
6. Show that when $\Delta \Gamma = 0$ and $|q/p| = 1$

$$\begin{aligned} \Gamma(B \rightarrow X \ell^- \bar{\nu})[t] &= e^{-\Gamma t} \sin^2(\Delta m t/2), \\ \Gamma(B \rightarrow X \ell^+ \nu)[t] &= e^{-\Gamma t} \cos^2(\Delta m t/2). \end{aligned} \quad (104)$$

Question 12: $B \rightarrow \pi^+ \pi^-$ and CP violation

One of the interesting decays to consider is $B \rightarrow \pi\pi$. Here we only briefly discuss it.

1. First assume that there is only tree level decay amplitude (that is, neglect penguin amplitudes). Draw the Feynman diagram of the amplitude, paying special attention to its CKM dependence.
 2. In that case, which angle of the unitarity triangle is the time dependent CP asymmetry, Eq. (89), sensitive to?
 3. Can you estimate the error introduced by neglecting the penguin amplitude? (Note that one can use isospin to reduce this error. Again, you are encouraged to read about it in one of the reviews.)
-

Question 13: B decays and CP violation

Consider the decays $\bar{B}^0 \rightarrow \psi K_S$ and $B^0 \rightarrow \phi K_S$. Unless explicitly noted, we always work within the framework of the standard model.

1. $\bar{B}^0 \rightarrow \psi K_S$ is a tree-level process. Write down the underlying quark decay. Draw the tree level diagram. What is the CKM dependence of this diagram? In the Wolfenstein parametrization, what is the weak phase of this diagram?
2. Write down the underlying quark decay for $B^0 \rightarrow \phi K_S$. Explain why there is no tree level diagram for $B^0 \rightarrow \phi K_S$.

3. The leading one loop diagram for $B^0 \rightarrow \phi K_S$ is a gluonic penguin diagram. As we have discussed, there are several diagrams and only their sum is finite. Draw a representative diagram with an internal top quark. What is the CKM dependence of the diagram? In the Wolfenstein parametrization, what is the weak phase of the diagram?
4. Next we consider the time dependent CP asymmetries. We define as usual

$$\lambda_f \equiv \frac{\bar{A}_f q}{A_f p}, \quad A_f \equiv A(B^0 \rightarrow f), \quad \bar{A}_f \equiv A(\bar{B}^0 \rightarrow f). \quad (105)$$

In our case we neglect subleading diagrams and then we have $|\lambda| = 1$ and thus

$$a_f \equiv \frac{\Gamma(\bar{B}^0(t) \rightarrow f) - \Gamma(B^0(t) \rightarrow f)}{\Gamma(\bar{B}^0(t) \rightarrow f) + \Gamma(B^0(t) \rightarrow f)} = -\text{Im}\lambda_f \sin(\Delta m_B t) \quad (106)$$

Both $a_{\psi K_S}$ and $a_{\phi K_S}$ measure the same angle of the unitarity triangle. That is, in both cases, $\text{Im}\lambda_f = \sin 2x$ where x is one of the angles of the unitarity triangle. What is x ? Explain.

5. Experimentally,

$$\text{Im}\lambda_{\psi K_S} = 0.68(3), \quad \text{Im}\lambda_{\phi K_S} = 0.47(19). \quad (107)$$

Comment about these two results. In particular, do you think these two results are in disagreement?

6. Assume that in the future we will find

$$\text{Im}\lambda_{\psi K_S} = 0.68(1), \quad \text{Im}\lambda_{\phi K_S} = 0.32(3). \quad (108)$$

That is, that the two results are not the same. Below are three possible “solutions”. For each solution explain if you think it could work or not. If you think it can work, show how. If you think it cannot, explain why.

- (a) There are standard model corrections that we neglected.
- (b) There is a new contribution to $B^0 - \bar{B}^0$ mixing with a weak phase that is different from the SM one.
- (c) There is a new contribution to the gluonic penguin with a weak phase that is different from the SM one.

Question 14: Decay of mass eigenstates

Derive Eq. (90). The idea is to understand that when we talk about mass eigenstates, we are talking about “late times,” $t \gg x\Gamma$ so that the $\sin(\Delta mt)$ term can be averaged out.

6 Putting it all together: The big picture

After we got a taste of how to probe the flavor sector, we are ready to ask: what are the results? That is, how compatible are all the measurements? As we explained, in principle we need four measurements to define the flavor parameters and the rest are checks on the model. Then we can ask what are the implications of these results on the big question of physics, that is, what is the fundamental Lagrangian of Nature.

6.1 The current status of the SM flavor sector

Out of the four flavor parameters of the SM, two are known to high accuracy, and there is a very good agreement between the various ways they are determined. These two parameters are λ and A of the

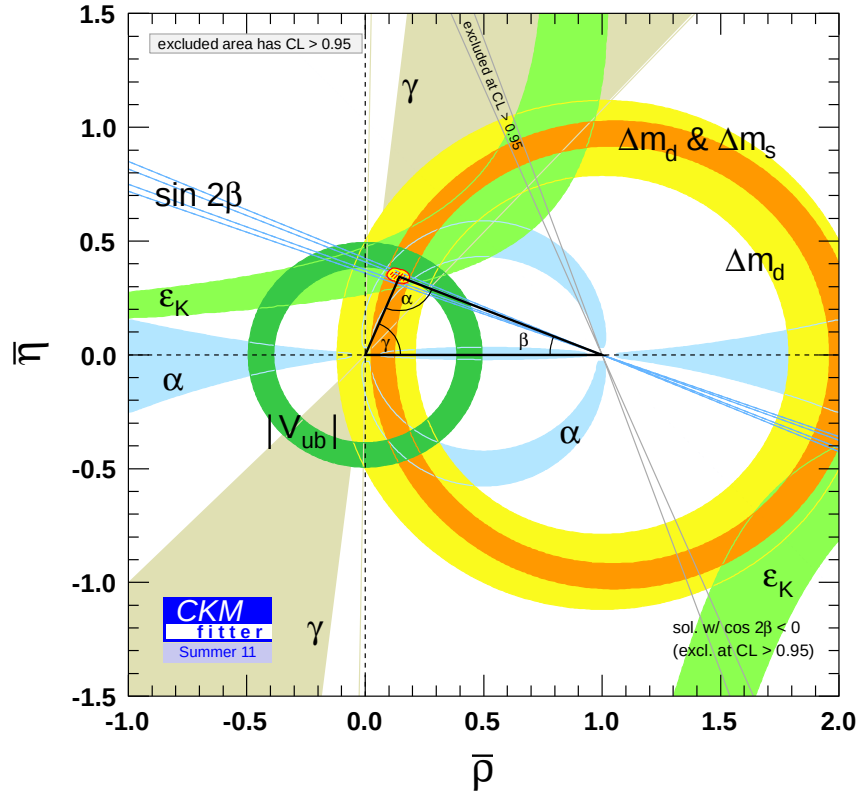


Fig. 5: Global fit to the unitarity triangle based on all available data. (Taking from the CKMfitter group website, ckmfitter.in2p3.fr.)

Wolfenstein parametrization of the CKM matrix. Thus, it is customary to plot all the other measurements as bounds on the rescaled unitarity triangle, that depend only on two parameters, ρ and η . There are many measurements and bounds and within the errors each of them gives an allowed range in the $\rho - \eta$ plane for the one undetermined apex of the rescaled unitarity triangle.

What I am about to show you (in the next page, please do not peek!) is the most recent compilation of all these results. I am sure you have seen it before. Yet, I hope that now you appreciate it much more. In class, I always stop at this point. I like to make sure the students do not miss the big moment, and that they see how amazing physics is. If I knew how to play a trumpet, this would be the moment that I would use it. In a written version, it is a bit harder. I cannot stop you from looking at the next page and I certainly cannot play the trumpet here. Still I do ask you to take a break here. Make sure you fully understand what you are going to see.

Now, that you are ready, take your time and look at Fig. 5. What you see are many bounds all overlapping at one small area in the $\rho - \eta$ plane. That is, taking the two measurements as determining ρ and η , all the rest are checks on the SM. You see that the flavor sector of the SM passes all its tests. Basically, all the measurements agree.

The most important implication of this triumph of theoretical and experimental physics is the following statement: The Cabibbo-Kobayashi-Maskawa mechanism is the dominant source of flavor and CP violation in low-energy flavor-changing processes. This is a very important statement. Indeed the Nobel prize was awarded to Kobayashi and Maskawa in 2008 because it is now experimentally proven that the KM phase is the one which explains the observed CP violation in Nature.

6.2 Instead of a summary: The NP flavor problem

The success of the SM can be seen as a proof that it is an effective low energy description of Nature. There are, however, many reasons to suspect that the SM has to be extended. A partial list includes the hierarchy problem, the strong CP problem, baryogenesis, gauge coupling unification, the flavor puzzle, neutrino masses, and gravity. We are therefore interested in probing the more fundamental theory. One way to go is to search for new particles that can be produced at yet unreached energies. Another way to look for new physics is to search for indirect effects of heavy unknown particles. Flavor physics is used to probe such indirect signals of physics beyond the SM.

In general, flavor bounds provide strong constraints on new physics models. This fact is called “the new physics flavor problem”. The problem is actually the mismatch between the new physics scale that is required in order to solve the hierarchy problem and the one that is needed in order to satisfy the experimental bounds from flavor physics.

In order to understand what the new physics flavor problem is, let us first recall the hierarchy problem. In order to prevent the Higgs mass from getting a large radiative correction, new physics must appear at a scale that is a loop factor above the weak scale

$$\Lambda \lesssim 4\pi m_W \sim 1 \text{ TeV}. \quad (109)$$

Here, and in what follows, Λ represents the new physics scale. Note that such TeV new physics can be directly probed in collider searches.

While the SM scalar sector is unnatural, its flavor sector is impressively successful.¹³ This success is linked to the fact that the SM flavor structure is special. As we already mentioned, the charged current interactions are universal (in the mass basis, this is manifest through the unitarity of the CKM matrix) and FCNCs are highly suppressed: they are absent at the tree level and at the one loop level they are further suppressed by the GIM mechanism. These special features are important in order to explain the observed pattern of weak decays. Thus, any extension of the SM must conserve these successful features.

Consider a generic new physics model, where the only suppression of FCNCs processes is due to the large masses of the particles that mediate them. Naturally, these masses are of the order of the new physics scale, Λ . Flavor physics, in particular measurements of meson mixing and CP violation, puts severe constraints on Λ . In order to find these bounds we take an effective field theory approach. At the weak scale we write all the non-renormalizable operators that are consistent with the gauge symmetry of the SM. In particular, flavor-changing four Fermi operators of the form (the Dirac structure is suppressed)

$$\frac{q_1 \bar{q}_2 q_3 \bar{q}_4}{\Lambda^2}, \quad (110)$$

are allowed. Here q_i can be any quark as long as the electric charges of the four fields in Eq. (110) sum up to zero. We emphasize that there is no exact symmetry that can forbid such operators. This is in contrast to operators that violate baryon or lepton number that can be eliminated by imposing symmetries like $U(1)_{B-L}$ or R-parity. The strongest bounds are obtained from meson mixing and CP-violation measurements. Depending on the mode we find bounds of the order

$$\Lambda \gtrsim \text{few} \times 10^4 \text{ TeV}. \quad (111)$$

There is tension between the new physics scale that is required in order to solve the hierarchy problem, Eq. (109), and the one that is needed in order not to contradict the flavor bounds, Eq. (111). The hierarchy problem can be solved with new physics at a scale $\Lambda \sim 1 \text{ TeV}$. Flavor bounds, on the other hand, require $\Lambda \gtrsim 10^4 \text{ TeV}$. This tension implies that any TeV scale new physics cannot have a generic flavor structure. This is the new physics flavor problem.

¹³The flavor structure of the SM is interesting since the quark masses and mixing angles exhibit hierarchy. These hierarchies are not explained within the SM, and this fact is usually called “the SM flavor puzzle”. This puzzle is different from the new physics flavor problem that we are discussing here.

Flavor physics has been mainly an input to model building, not an output. The flavor predictions of any new physics model are not a consequence of its generic structure but rather of the special structure that is imposed to satisfy the severe existing flavor bounds. It is clearly a very interesting open question to determine the NP model and how it deals with flavor.

6.3 Concluding remarks

This is a good time to finish the lectures. I hope that you gained some understanding of flavor physics, how it was used to shape the SM as we know it, and why it is so important in our quest to find the theory that extends the SM. In the near future we expect more data in the energy frontier, as well as more flavor data. It can be really fun to see how the two can work together to show us what Nature is at really short distances, that is, to help us in getting a better answer to the fundamental question of physics

$$\mathcal{L} = ? \quad (112)$$

Acknowledgements

I thank Joshua Berger, Mario Martone, and Dean Robinson for comments on the manuscript. The work of YG is supported by NSF grant number PHY-0757868 and by a U.S.-Israeli BSF grant.

References

- [1] O. Gedalia and G. Perez, *TASI 2009 Lectures - Flavor Physics*, in *Proc. TASI 2009*, arXiv:1005.3106 [hep-ph].
- [2] G. Isidori, *B physics in the LHC era*, in *Proc. 65th Scottish Universities Summer School in Physics: LHC Physics*, arXiv:1001.3431 [hep-ph].
- [3] A. J. Buras, PoS EPS **HEP2009** (2009) 024, arXiv:0910.1032 [hep-ph].
- [4] U. Nierste, *Three Lectures on Meson Mixing and CKM phenomenology*, in *Proc. Helmholtz International Summer School on Heavy Quark Physics, Dubna, 2008*, arXiv:0904.1869 [hep-ph].
- [5] M. Artuso, E. Barberio and S. Stone, PMC Phys. **A3**, 3 (2009) [arXiv:0902.3743 [hep-ph]].
- [6] Y. Nir, *Probing new physics with flavor physics (and probing flavor physics with new physics)*, in *Proc. Prospects in Theoretical Physics summer program on the Standard Model and Beyond, Princeton, 2007*, arXiv:0708.1872 [hep-ph].
- [7] A. Hocker and Z. Ligeti, Ann. Rev. Nucl. Part. Sci. **56**, 501 (2006) [arXiv:hep-ph/0605217].
- [8] M. Neubert, *Effective field theory and heavy quark physics*, in *Proc. TASI 2004*, arXiv:hep-ph/0512222.
- [9] Y. Nir, *CP violation in meson decays*, in *Proc. Les Houches summer school on Theoretical Physics, Les Houches, 2005*, arXiv:hep-ph/0510413.
- [10] A. J. Buras, *Flavour physics and CP violation*, in *Proc. European School on High-Energy Physics, Sant Feliu de Guixols, Spain, 2004*, arXiv:hep-ph/0505175.
- [11] Z. Ligeti, *Introduction to heavy meson decays and CP asymmetries*, in *Proceedings of 30th SLAC Summer Institute on Particle Physics, Menlo Park, 2002*, arXiv:hep-ph/0302031.
- [12] I. I. Y. Bigi and A. I. Sanda, *CP violation*, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **9**, 1 (2000).
- [13] G. C. Branco, L. Lavoura and J. P. Silva, *CP violation*, Int. Ser. Monogr. Phys. **103**, 1 (1999).
- [14] The Review of Particle Physics C. Amsler et al. (Particle Data Group), Physics Letters **B667** (2008) 1. Recent updates can be found online at pdg.lbl.gov.

Neutrino Oscillation Physics

B. Kayser

Fermilab, Batavia, IL, USA

Abstract

We present a description of the physics of neutrino oscillation. This description is centered on a new way of deriving the oscillation probability. We also provide a brief guide to references relevant to topics other than neutrino oscillation.

1 Introduction

If someone asks you what neutrinos are good for, then you can point out that, if neutrinos did not exist, the chain of nuclear reactions that powers the sun would be impossible. The reaction that initiates this chain is the fusion process

$$\text{Spin : } \begin{array}{ccccccc} p & + & p & \rightarrow & d & + & e^+ & + & \nu \\ 1/2 & & 1/2 & & 1 & & 1/2 & & 1/2 \end{array},$$

where we have indicated below each particle its intrinsic spin. Obviously, if a neutrino were not emitted, this process would not conserve angular momentum, so it would be forbidden. Then the chain of reactions that powers the sun could not even get started, and we humans on planet Earth would not exist.

Neutrinos and photons are by far the most abundant elementary particles in the universe. Thus, if we would like to comprehend the universe, we must understand the neutrinos. Of course, studying the neutrinos is challenging, since the only known forces through which these electrically-neutral leptons interact are the weak force and gravity. Consequently, interactions of neutrinos in a detector are very rare events, so that very large detectors and intense neutrino sources are needed to make experiments feasible. Nevertheless, we have confirmed that the weak interactions of neutrinos are correctly described by the Standard Model (SM) of elementary particle physics. Moreover, in the last 14 years, we have discovered that neutrinos have nonzero masses, and that leptons mix. These discoveries have been based on the observation that neutrinos can change from one "flavor" to another — the phenomenon known as neutrino oscillation. We shall explain the physics of neutrino oscillation, deriving the probability of oscillation in a new way. We shall also provide a very brief guide to references that can be used to study some major neutrino-physics topics other than neutrino oscillation.

2 Physics of Neutrino Oscillation

2.1 Preliminaries

There are three known flavors of neutrinos: ν_e , ν_μ , and ν_τ . We shall define the neutrino of a given flavor in terms of leptonic W -boson decay. This decay produces a charged lepton, which may be an e , μ , or τ , plus a neutrino. We define the ν_e as the neutrino produced when the charged lepton is an e , the ν_μ as the neutrino produced together with a μ , and the ν_τ as the one that accompanies a τ .

Suppose a neutrino ν_α of flavor α ($= e, \mu, \text{ or } \tau$), born in a W decay, interacts in a detector *immediately*, before it has time to evolve into something else. Suppose further that, by exchanging a W boson with its target in the detector, this neutrino turns into a charged lepton. Then, as far as we know, this charged lepton will always be of the same flavor as the neutrino. Thus, it will be of the same flavor as the charged lepton with which the neutrino was born.

Now imagine that we send a neutrino on a *long* journey, say from your present location straight downward to a detector on the opposite side of the Earth. Suppose that this neutrino is created in the pion decay $\pi \rightarrow \text{Virtual } W \rightarrow \mu + \nu_\mu$, so that at birth it is a ν_μ . Imagine that this neutrino interacts via

W exchange in the distant detector, turning into a charged lepton. If neutrinos have masses and leptons mix, then this charged lepton need not be a μ , but could be, say, a τ . Since it is only a ν_τ that can turn into a τ , the appearance of this τ would imply that during its journey, our neutrino has evolved from a ν_μ into a ν_τ , or at least into a neutrino with a nonzero ν_τ component. The last 14 years have brought us compelling evidence that such changes of neutrino flavor actually occur. As we shall see, the probability of flavor change in vacuum has an oscillatory character, so flavor change is commonly referred to as neutrino oscillation.

That neutrinos have masses means that there is some spectrum of neutrino mass eigenstates ν_i , whose masses m_i^ν we would like to determine. That leptons mix means that the neutrinos of definite flavor, ν_e, ν_μ , and ν_τ , are not the mass eigenstates ν_i . Instead, the neutrino state $|\nu_\alpha\rangle$ of flavor α , which is the neutrino state that is created in leptonic W decay together with the charged lepton of the same flavor, is a quantum superposition

$$|\nu_\alpha\rangle = \sum_i U_{\alpha i}^* |\nu_i\rangle \quad (1)$$

of the mass eigenstates $|\nu_i\rangle$. (From now on, a ν with a Greek subscript such as α or β will denote a neutrino of definite flavor, while one with a Latin subscript such as i or j will denote a neutrino of definite mass.) In the superposition of Eq. (1), the coefficients $U_{\alpha i}^*$ are (complex conjugates of the) elements of the leptonic mixing matrix U — the leptonic analogue of the quark mixing matrix. Now, there are at least 3 neutrinos ν_α of definite flavor, and they must be orthogonal to one another, or a neutrino of one flavor, interacting via W exchange, would sometimes turn into a charged lepton of a *different* flavor. Out of these 3 orthogonal ν_α , we can form 3 orthogonal linear combinations that will be neutrino mass eigenstates ν_i . (The mass eigenstates must be orthogonal because they are eigenstates of a Hermitean operator, the Hamiltonian.) For all we know, there are more than 3 neutrino mass eigenstates. However, if there are only 3, then U is a 3×3 matrix, and, being the matrix that transforms the states of one quantum basis into those of another, it is unitary. The matrix U is sometimes referred to as the Maki-Nakagawa-Sakata (MNS) matrix, or as the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, to honor several pioneering contributors to the physics of mixing and oscillation.

Mixing is readily incorporated into the SM description of the coupling of the leptons to the W . For this coupling, we have in the SM Lagrangian density the term

$$\mathcal{L}_{\ell\nu W} = -\frac{g}{\sqrt{2}} \sum_{\alpha=e,\mu,\tau} (\bar{\ell}_{L\alpha} \gamma^\lambda \nu_{L\alpha} W_\lambda^- + \bar{\nu}_{L\alpha} \gamma^\lambda \ell_{L\alpha} W_\lambda^+) . \quad (2)$$

Here, g is the semi-weak coupling constant, ν_α is the neutrino of flavor α as before, and ℓ_α is the charged lepton of flavor α . That is, $\ell_e \equiv e$, $\ell_\mu \equiv \mu$, and $\ell_\tau \equiv \tau$. The subscript L denotes a left-handed chiral projection: $\ell_{L\alpha} = [(1 - \gamma_5)/2]\ell_\alpha$, and similarly for $\nu_{L\alpha}$. Note from Eq. (2) that, in conformity with the rule quoted earlier, the neutrino of flavor α couples only to the charged lepton of the same flavor. To explicitly incorporate mixing into the $\ell\nu W$ coupling, we insert Eq. (1) into Eq. (2), so that the latter becomes

$$\mathcal{L}_{\ell\nu W} = -\frac{g}{\sqrt{2}} \sum_{\substack{\alpha=e,\mu,\tau \\ i=1,2,3}} (\bar{\ell}_{L\alpha} \gamma^\lambda U_{\alpha i} \nu_{Li} W_\lambda^- + \bar{\nu}_{Li} \gamma^\lambda U_{\alpha i}^* \ell_{L\alpha} W_\lambda^+) . \quad (3)$$

Here, as before, ν_i is a neutrino mass eigenstate, and we have taken into account the fact that the field operator which absorbs the state $\sum_j U_{\alpha j}^* |\nu_j\rangle$ of Eq. (1) is not $\sum_i U_{\alpha i}^* \nu_i$, but $\sum_i U_{\alpha i} \nu_i$.

From Eq. (3), we see that, apart from the factor of $g/\sqrt{2}$ out front and kinematical factors, the amplitude for $W^+ \rightarrow \ell_\alpha^+ + \nu_i$ or for $\ell_\alpha^- + W^+ \rightarrow \nu_i$ is just $U_{\alpha i}^*$, while that for $\nu_i \rightarrow \ell_\beta^- + W^+$ is just $U_{\beta i}$. Writing out the mixing matrix explicitly, it is

$$U = \begin{matrix} & \begin{matrix} \nu_1 & \nu_2 & \nu_3 \end{matrix} \\ \begin{matrix} e \\ \mu \\ \tau \end{matrix} & \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \end{matrix} . \quad (4)$$

From Eq. (3), we see that, for example, the e (top) row of U tells us what linear combination of the neutrino mass eigenstates ν_1 , ν_2 , and ν_3 couples to an e and a W . Similarly, the ν_1 (first) column of U tells us what linear combination of the charged lepton mass eigenstates e , μ , and τ couples to a ν_1 and a W . Similarly for the other rows and columns.

2.2 Probability of Neutrino Oscillation in Vacuum

Let us now find the probability $P(\nu_\alpha \rightarrow \nu_\beta; L, E)$ that a neutrino born as a ν_α — a neutrino of flavor α — will then behave like a ν_β — a neutrino of flavor β — after traveling through vacuum for a distance L with energy E . The conventional derivation of this probability may be found in many places in the literature [1]. Rather than reproduce that derivation, or present the one we gave in the lectures (which may be found in Ref. [2]), here we present a new approach [3]. We apply this approach to neutrinos produced in the pion decay $\pi \rightarrow \mu + \nu$, which we view in the pion rest frame, as in Fig. 1. To illustrate

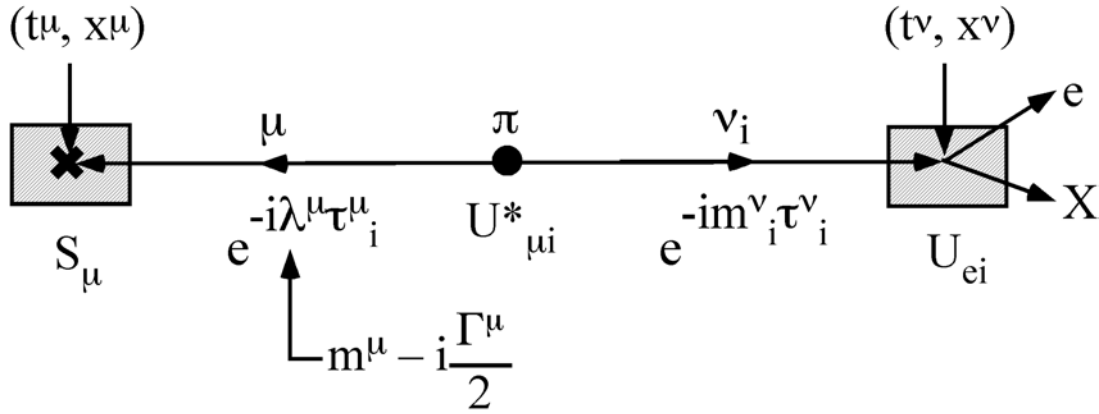


Fig. 1: The pion-rest-frame view of a pion decay followed by interactions of the daughter neutrino and muon. The amplitudes shown for the various parts of this scenario are explained in the text.

how the approach works, we consider a scenario in which the neutrino, having been created in the π decay at a spacetime point $(0, 0)$, then interacts at the spacetime point (t^ν, x^ν) whose pion-rest-frame coordinates are t^ν and x^ν . The interaction is via W exchange with a target in a neutrino detector, and, for illustration, we suppose that the charged lepton into which the neutrino is converted by this interaction is an electron. Of course, the interaction will also produce a recoil X . As part of the same overall scenario, the muon created together with the neutrino in the π decay interacts at the spacetime point (t^μ, x^μ) whose pion-rest-frame coordinates are t^μ and x^μ . We imagine that the interaction is with matter that surrounds the π decay region. The full scenario is pictured in Fig. 1.

We shall find the amplitude for the entire scenario, including the π decay and the ν and μ interactions [4]. To do this, we shall use the fact that if a particle has mass m and width Γ , adding up to a complex mass $\lambda = m - i\Gamma/2$, then the amplitude for this particle to propagate for a proper time τ in its own rest frame is $\exp(-i\lambda\tau)$. (If this propagation is through a distance x during a time t in some frame in which the particle has momentum p and energy E , then $\exp(-im\tau)$, the non-decaying part of $\exp(-i\lambda\tau)$, is simply the familiar quantum-mechanical plane-wave factor $\exp[i(px - Et)]$.)

In the neutrino mass eigenstate basis, the neutrino that travels from the π decay point to the interaction point in the neutrino detector will be one or another of the neutrino mass eigenstates ν_i . Since we cannot make measurements that determine which ν_i was involved in any given event without destroying the oscillation pattern, we must add the amplitudes for the contributions of the different ν_i coherently.

From Eq. (3) and the discussion that follows it, the relevant factor in the amplitude for the decay $\pi \rightarrow \mu + \nu$ to yield, in particular, the neutrino mass eigenstate ν_i is $U_{\mu i}^*$, as shown in Fig. 1. Similarly, the relevant factor for the charged lepton created when the ν_i interacts in the detector to be, in particular, an e is U_{ei} . The amplitude for the muon to interact in matter we shall call S_μ .

From the discussion of the amplitude for a particle to propagate, we see that the amplitude for the neutrino mass eigenstate ν_i , of mass m_i^ν , to propagate is $\exp(-im_i^\nu \tau_i^\nu)$. Here, we are neglecting the extremely small neutrino decay width, and τ_i^ν is the proper time that elapses in the ν_i rest frame while the ν_i travels from the spacetime point $(0, 0)$ where it was born to the given interaction point (t^ν, x^ν) in the detector. The index i on τ_i^ν is present because the time that elapses in the ν_i rest frame during the journey to the given point (t^ν, x^ν) in the pion rest frame depends on the ν_i energy in the latter frame, hence on the ν_i mass, and consequently on which ν_i is involved.

Similarly, the amplitude for the muon to propagate is $\exp(-i\lambda^\mu \tau_i^\mu) = \exp[-i(m^\mu - i\Gamma^\mu/2)\tau_i^\mu]$. Here we are taking into account the decay of the muon by including in its complex mass λ^μ its decay width Γ^μ . The quantity τ_i^μ is the proper time that elapses in the μ rest frame while the μ travels from the π decay point $(0, 0)$ to the μ interaction point (t^μ, x^μ) . The index i on τ_i^μ is present because the muon and the neutrino are kinematically entangled. The energies of both of these particles in the pion rest frame depend on the mass of the emitted neutrino, so that they depend on which ν_i it is. Consequently, the proper times that elapse in the rest frames of the muon and the neutrino both depend on which ν_i is emitted, at least in principle.

Multiplying together the amplitudes for the various parts of the scenario pictured in Fig. 1, and coherently adding the contributions of the different ν_i , we find that the amplitude Amp for the entire scenario is given by

$$Amp = \sum_{i=1,2,3} S_\mu e^{-i(m^\mu - i\Gamma^\mu/2)\tau_i^\mu} U_{\mu i}^* e^{-im_i^\nu \tau_i^\nu} U_{ei} . \quad (5)$$

We note that this amplitude is Lorentz invariant.

How do the muon and neutrino propagation amplitudes $\exp[-i(m^\mu - i\Gamma^\mu/2)\tau_i^\mu]$ and $\exp(-im_i^\nu \tau_i^\nu)$ actually depend on i ? From the Lorentz transformation, the proper time τ_i^μ in the muon propagation amplitude is given by

$$\tau_i^\mu = \frac{1}{m^\mu} (E_i^\mu t^\mu - p_i^\mu x^\mu) . \quad (6)$$

Here, E_i^μ and p_i^μ are, respectively, the muon energy and momentum in the pion rest frame when the emitted neutrino is ν_i . In evaluating the right-hand side of Eq. (6), we choose a muon interaction point at which x^μ is related to t^μ by

$$x^\mu = v_0^\mu t^\mu = \frac{p_0^\mu}{E_0^\mu} t^\mu . \quad (7)$$

Here, v_0^μ , p_0^μ , and E_0^μ are, respectively, the velocity, momentum, and energy that the muon would have in the pion rest frame if neutrinos were massless. The spacetime points passed by the peak of the quantum-mechanical wave packet that describes the muon propagation in greater detail than is needed here would satisfy Eq. (7) to an excellent approximation.

From Eqs. (6) and (7), the muon in $\pi \rightarrow \mu + \nu_i$ and that in $\pi \rightarrow \mu + \nu_j$ have travel proper times that differ by

$$\tau_i^\mu - \tau_j^\mu = \frac{t^\mu}{m^\mu} \left[(E_i^\mu - E_j^\mu) - (p_i^\mu - p_j^\mu) \frac{p_0^\mu}{E_0^\mu} \right] . \quad (8)$$

For a given neutrino mass m^ν , the pion-rest-frame energy of the muon in $\pi \rightarrow \mu + \nu$ is

$$E^\mu = \frac{(m^\pi)^2 + (m^\mu)^2 - (m^\nu)^2}{2m^\pi} , \quad (9)$$

where m^π is the pion mass. Thus, in Eq. (8),

$$E_i^\mu - E_j^\mu = -\frac{\Delta m_{ij}^2}{2m^\pi} , \quad (10)$$

where $\Delta m_{ij}^2 \equiv (m_i^\nu)^2 - (m_j^\nu)^2$. Moreover, for given muon energy E^μ , $(p^\mu)^2 = (E^\mu)^2 - (m^\mu)^2$, so that

$$\frac{dp^\mu}{d[(m^\nu)^2]} = \frac{E^\mu}{p^\mu} \frac{dE^\mu}{d[(m^\nu)^2]} . \quad (11)$$

Consequently, to lowest order in the squares of the neutrino masses,

$$p_i^\mu - p_j^\mu = \frac{E_0^\mu}{p_0^\mu} \left[-\frac{\Delta m_{ij}^2}{2m^\pi} \right] . \quad (12)$$

Inserting Eqs. (10) and (12) into Eq. (8), we find that to lowest (i.e. first) order in the squares of the neutrino masses,

$$\tau_i^\mu - \tau_j^\mu = \frac{t^\mu}{m^\mu} \left[-\frac{\Delta m_{ij}^2}{2m^\pi} \right] \left[1 - \frac{E_0^\mu}{p_0^\mu} \frac{p_0^\mu}{E_0^\mu} \right] = 0 . \quad (13)$$

Thus, to lowest order, the muon propagation amplitude $\exp(-im^\mu \tau_i^\mu) \exp[-(\Gamma^\mu/2)\tau_i^\mu]$ actually does not depend on which ν_i is emitted [5]. The factor $\exp(-im^\mu \tau_i^\mu)$ will have no significant effect at all on the absolute square of the amplitude of Eq. (5) for the scenario in Fig. 1. The factor $\exp[-(\Gamma^\mu/2)\tau_i^\mu]$ will lead to an overall decay of this amplitude with muon travel time, reflecting the obvious fact that the probability for the muon to remain present (i.e., not yet decayed), so that it may interact, decays with time. But this overall decay of the amplitude for the scenario in Fig. 1 will not affect the neutrino oscillation pattern. Since it is only that pattern in which we are ultimately interested, we can drop the entire muon propagation amplitude from the amplitude of Eq. (5).

Turning to the propagation amplitude $\exp(-im_i^\nu \tau_i^\nu)$ for the neutrino ν_i , we have from the Lorentz transformation the relation

$$m_i^\nu \tau_i^\nu = E_i^\nu t^\nu - p_i^\nu x^\nu . \quad (14)$$

Here, E_i^ν and p_i^ν are, respectively, the ν_i energy and momentum in the pion rest frame. Since, in practice, neutrinos are ultra-relativistic, we choose a neutrino interaction point at which t^ν is related to x^ν by $t^\nu = x^\nu \equiv L^0$. Then the propagation phases for neutrino mass eigenstates ν_i and ν_j differ by

$$m_i^\nu \tau_i^\nu - m_j^\nu \tau_j^\nu = [(E_i^\nu - E_j^\nu) - (p_i^\nu - p_j^\nu)]L^0 . \quad (15)$$

In analogy to Eq. (9), for given neutrino mass m^ν , the pion-rest-frame energy of the neutrino, E^ν , is given by

$$E^\nu = \frac{(m^\pi)^2 + (m^\nu)^2 - (m^\mu)^2}{2m^\pi} . \quad (16)$$

Thus, the energies of two different neutrino mass eigenstates ν_i and ν_j differ by

$$E_i^\nu - E_j^\nu = \frac{\Delta m_{ij}^2}{2m^\pi} . \quad (17)$$

In addition, for given neutrino energy E^ν , $(p^\nu)^2 = (E^\nu)^2 - (m^\nu)^2$. From this relation and conservation of energy in $\pi \rightarrow \mu + \nu$, one easily finds that

$$\left. \frac{dp^\nu}{d[(m^\nu)^2]} \right|_{m^\nu=0} = -\frac{E_0^\mu}{E_0^\nu} \frac{1}{2m^\pi} , \quad (18)$$

where $E_0^\nu = [(m^\pi)^2 - (m^\mu)^2]/2m^\pi$ is the pion-rest-frame energy that the neutrino would have if it were massless. It follows that, to lowest order in Δm_{ij}^2 , the momenta of ν_i and ν_j differ by

$$p_i^\nu - p_j^\nu = -\frac{E_0^\mu}{E_0^\nu} \frac{\Delta m_{ij}^2}{2m^\pi} . \quad (19)$$

Inserting Eqs. (17) and (19) into Eq. (15), we find that to lowest order,

$$m_i^\nu \tau_i^\nu - m_j^\nu \tau_j^\nu = \frac{\Delta m_{ij}^2}{2m^\pi} \left[1 + \frac{E_0^\mu}{E_0^\nu} \right] L^0 = \Delta m_{ij}^2 \frac{L^0}{2E_0^\nu} . \quad (20)$$

From this result, we see that we may take the neutrino propagation amplitude $\exp(-im_i^\nu \tau_i^\nu)$ to be

$$e^{-i(m_i^\nu)^2 \frac{L^0}{2E_0^\nu}} , \quad (21)$$

and all the relative phases in the amplitude of Eq. (5) will be correct. Then, if we delete from Eq. (5) the muon interaction and propagation amplitudes, which do not affect the neutrino oscillation pattern because they are i -independent, Eq. (5) yields

$$Amp = \sum_{i=1,2,3} U_{\mu i}^* e^{-i(m_i^\nu)^2 \frac{L^0}{2E_0^\nu}} U_{ei} . \quad (22)$$

A neutrino flavor-change experiment will carry out its work in the rest frame of some neutrino detector — the laboratory frame. However, until now we have been viewing our illustrative process of interest from the rest frame of the pion whose decay creates our neutrino. Now, as we shall see momentarily, for given neutrino energy, the probability of flavor change oscillates as a function of the distance L that the neutrino travels in the laboratory frame. If we are to observe this oscillation, then obviously the neutrino source — the pion in our example — must be spatially localized to within an oscillation wavelength. But then, by the uncertainty principle $\Delta p \Delta x \geq \hbar$, there must be some uncertainty in the lab-frame pion momentum [6]. The pions whose decays produce the neutrinos of an oscillation experiment cannot be known to be precisely at rest. Thus, we must find the amplitude for the scenario in Fig. 1 when the pion is moving in the lab frame — the rest frame of the neutrino detector. In addition, we must express this amplitude in terms of lab-frame variables. Accomplishing these goals is easy. First, we recall that the amplitude Amp of Eqs. (5) and (22) is Lorentz invariant. It is valid both in the pion rest frame and in the lab frame, in which in general the pion is moving. Secondly, to express the amplitude of Eq. (22) in terms of lab-frame, rather than pion-rest-frame, variables, we note that, as already remarked, neutrinos are ultra-relativistic. Thus, in the pion rest frame, the travel time of one of them is equal to its travel distance L^0 . Then, by the Lorentz transformation, its travel distance L in the lab frame is given by

$$L = \gamma_\pi (1 + \beta_\pi) L^0 , \quad (23)$$

where β_π is the velocity of the pion in the lab, and $\gamma_\pi = 1/\sqrt{1 - \beta_\pi^2}$. Similarly, in the pion rest frame, the momentum p_0^ν of a massless neutrino is equal to its energy E_0^ν . Thus, its energy E in the lab frame is given by

$$E = \gamma_\pi (1 + \beta_\pi) E_0^\nu . \quad (24)$$

We see that

$$\frac{L}{E} = \frac{L^0}{E_0^\nu} , \quad (25)$$

so that we may write the amplitude of Eq. (22) as

$$Amp = \sum_{i=1,2,3} U_{\mu i}^* e^{-i(m_i^\nu)^2 \frac{L}{2E}} U_{ei} . \quad (26)$$

As explained in Section 2.1, the neutrino produced in $\pi \rightarrow \mu + \nu$ is by definition a ν_μ . In the calculation above, we have worked in neutrino mass eigenstate basis, so, in effect, we have broken the ν_μ down into its mass eigenstate components. For purposes of illustration, we have considered a scenario in which the neutrino interaction in the detector yields an electron. Since, in neutrino flavor basis, it is only a ν_e that can yield an electron, the sequence of events pictured in Fig. 1 is what would commonly be called $\nu_\mu \rightarrow \nu_e$ oscillation, with the addition of an interaction between matter and the muon that is produced together with the neutrino in the pion decay. We are interested mainly in the probability for the $\nu_\mu \rightarrow \nu_e$ oscillation, integrated over all the possible fates of the muon. Apart from a possible overall normalization factor, this muon-integrated $\nu_\mu \rightarrow \nu_e$ oscillation probability, $P(\nu_\mu \rightarrow \nu_e; L, E)$, will be given by the absolute square of the amplitude Amp of Eq. (26), from which the muon interaction and propagation amplitudes have been removed.

Generalizing to a scenario in which the neutrino is born together with a charged lepton of flavor α ($= e, \mu$, or τ), and then interacts in a detector and makes a charged lepton of flavor β (not necessarily different from α), we see from Eq. (26) that the amplitude would be

$$Amp(\nu_\alpha \rightarrow \nu_\beta; L, E) = \sum_{i=1,2,3} U_{\alpha i}^* e^{-i(m_i^\nu)^2 \frac{L}{2E}} U_{\beta i} . \quad (27)$$

Apart from a possible overall normalization factor, the probability $P(\nu_\alpha \rightarrow \nu_\beta; L, E)$ of the $\nu_\alpha \rightarrow \nu_\beta$ oscillation will then be the absolute square of this amplitude. Assuming that the mixing matrix U is unitary, we find from Eq. (27) that this absolute square, summed over all possible final flavors β , including $\beta = \alpha$, is

$$\begin{aligned} \sum_{\text{All } \beta} |Amp(\nu_\alpha \rightarrow \nu_\beta; L, E)|^2 &= \sum_{\beta} \left(\sum_i U_{\alpha i}^* e^{-i(m_i^\nu)^2 \frac{L}{2E}} U_{\beta i} \right) \left(\sum_j U_{\alpha j} e^{i(m_j^\nu)^2 \frac{L}{2E}} U_{\beta j}^* \right) \\ &= \sum_{i,j} U_{\alpha i}^* e^{-i(m_i^\nu)^2 \frac{L}{2E}} U_{\alpha j} e^{i(m_j^\nu)^2 \frac{L}{2E}} \delta_{ij} \\ &= \sum_i |U_{\alpha i}|^2 = 1 . \end{aligned} \quad (28)$$

Thus, the amplitude $Amp(\nu_\alpha \rightarrow \nu_\beta; L, E)$ of Eq. (27) is a properly normalized probability amplitude. It needs no additional normalization factor. The probability $P(\nu_\alpha \rightarrow \nu_\beta; L, E)$ of $\nu_\alpha \rightarrow \nu_\beta$ oscillation is simply its absolute square. Taking this absolute square, and making use of the assumed unitarity of the mixing matrix, we find that

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_\beta; L, E) &= \delta_{\alpha\beta} - 4 \sum_{i>j} \Re(U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*) \sin^2(\Delta m_{ij}^2 \frac{L}{4E}) \\ &\quad + 2 \sum_{i>j} \Im(U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*) \sin(\Delta m_{ij}^2 \frac{L}{2E}) . \end{aligned} \quad (29)$$

In deriving this expression for the oscillation probability, we have assumed that the neutral lepton in Fig. 1 is a *neutrino*, not an *antineutrino*. The factors $U_{\mu i}^*$ and U_{ei} that we took from Eq. (3) and incorporated into the amplitude of Eq. (5) depended on this assumption. As we see from Eq. (3), if the neutral lepton had been an antineutrino, then $U_{\mu i}^*$ and U_{ei} would have been replaced, respectively, by $U_{\mu i}$ and U_{ei}^* . In addition, the amplitude S_μ for the matter interaction of the μ^+ from the reaction $\pi^+ \rightarrow \mu^+ + \nu$ that produces a neutrino would have been replaced by a different amplitude S'_μ for the matter interaction of the μ^- from the reaction $\pi^- \rightarrow \mu^- + \bar{\nu}$ that produces an antineutrino. However, as we have seen, the muon-matter interaction amplitude is ultimately irrelevant. Moreover, so long as CPT invariance holds, a particle and its antiparticle have the same mass and the same width. Thus, the

muon and neutrino propagation amplitudes in Eq. (5) would be unchanged if the μ^+ and neutrino from $\pi^+ \rightarrow \mu^+ + \nu$ were replaced by the μ^- and antineutrino from $\pi^- \rightarrow \mu^- + \bar{\nu}$. We conclude from this $\pi \rightarrow \mu + \nu$ example that, completely generally,

$$P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta; L, E) = P(\nu_\alpha \rightarrow \nu_\beta; L, E | U \rightarrow U^*) . \quad (30)$$

That is, the probability for the antineutrino oscillation $\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta$ is the same as for the corresponding neutrino oscillation $\nu_\alpha \rightarrow \nu_\beta$, except that the mixing matrix U in the latter is replaced by U^* in the former. From Eq. (29), we then have

$$\begin{aligned} P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta; L, E) &= \delta_{\alpha\beta} - 4 \sum_{i>j} \Re(U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*) \sin^2(\Delta m_{ij}^2 \frac{L}{4E}) \\ &\quad + 2 \sum_{i>j} \Im(U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*) \sin(\Delta m_{ij}^2 \frac{L}{2E}) . \end{aligned} \quad (31)$$

We see that if U is not real, then the $\nu_\alpha \rightarrow \nu_\beta$ and $\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta$ oscillation probabilities can differ.

The coupling of the leptons to the W boson has a parity-violating, chirally left-handed structure, as described by Eq. (3) or Eq. (2). Moreover, the neutrinos we study experimentally are ultra-relativistic, and for ultra-relativistic fermions there is essentially no difference between chirality and helicity. As a result, the neutrinos we study experimentally, which are produced by the chirally left-handed coupling of Eq. (3), are essentially always of left-handed (i.e., negative) helicity. It is easy to show that, in contrast, the antineutrinos produced by this coupling are essentially always of right-handed (i.e., positive) helicity. By $\nu_\alpha \rightarrow \nu_\beta$, we mean the oscillation of neutrinos of left-handed helicity, and by $\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta$ we mean the oscillation of antineutrinos of right-handed helicity. Now, the particle-antiparticle symmetry operation that turns a neutrino of left-handed helicity into an antineutrino of right-handed helicity is CP. The charge conjugation operation C turns the neutrino into an antineutrino with no change of kinematical variables, and the parity operation P reverses the helicity. Thus, if, owing to a nonvanishing value of the last term in Eq. (31), the probabilities for $\nu_\alpha \rightarrow \nu_\beta$ and $\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta$ should differ, this difference would be a violation of CP invariance. To date, CP violation has been observed only in the quark sector, and its observation in the neutrino sector would establish that the leptons violate CP as well. This observation in the latter sector would also make it more plausible that the baryon-antibaryon asymmetry of the universe arose, at least in part, through a scenario called leptogenesis that involves hypothesized very heavy neutrinos [7, 8].

We see from Eq. (31) that the oscillation probability oscillates as a function of L/E , justifying our calling neutrino flavor change “oscillation”. We also see from Eq. (31) that oscillation from one flavor α into a different one β implies nonzero mass splittings Δm_{ij}^2 , hence nonzero neutrino masses. Similarly, such oscillation implies that U is not diagonal, which is to say that there is nontrivial leptonic mixing. Inserting into the quantity $\Delta m_{ij}^2 L/4E$, on which the oscillation in Eq. (31) depends, the so-far-omitted factors of $\hbar$ and c , we find that

$$\Delta m_{ij}^2 \frac{L}{4E} = 1.27 \Delta m_{ij}^2 (\text{eV}^2) \frac{L (\text{km})}{E (\text{GeV})} . \quad (32)$$

Thus, the factor $\sin^2(\Delta m_{ij}^2 \frac{L}{4E})$ in Eq. (31) becomes $\sin^2[1.27 \Delta m_{ij}^2 (\text{eV}^2) \frac{L (\text{km})}{E (\text{GeV})}]$. This factor is appreciable when its argument is $\geq \mathcal{O}(1)$. Thus, an oscillation experiment with given L/E is sensitive to squared-mass splittings $\Delta m_{ij}^2 (\text{eV}^2) \gtrsim E (\text{GeV})/L (\text{km})$. For example, if $E = 1 \text{ GeV}$, and L is the diameter of the Earth, $\sim 10^4 \text{ km}$, values of E and L that are encountered in studies of the neutrinos made in the Earth’s atmosphere by cosmic rays, then there will be sensitivity to $\Delta m_{ij}^2 \gtrsim 10^{-4} \text{ eV}^2$. As this illustrates, neutrino oscillation experiments can be sensitive to very tiny mass splittings. We note, however, that oscillation depends *only* on these splittings, and not on the individual neutrino masses. Determining those individual masses will require another approach.

Neutrino flavor change can be sought in two ways. In a beam of neutrinos born with a known flavor α , one can look for the appearance of neutrinos of a different flavor β . This is referred to as an appearance experiment. Alternatively, in a known flux of neutrinos ν_α of a given flavor α , one can look for the disappearance of some of this known flux, due to the oscillation of some of the ν_α into neutrinos of other flavors. This is referred to as a disappearance experiment.

In Eq. (28), we have confirmed that the probability of oscillation, $P(\nu_\alpha \rightarrow \nu_\beta; L, E)$, summed over all possible final flavors β , including $\beta = \alpha$, is unity. That is, the probability that a neutrino changes flavor, plus the probability that it does not change flavor, is unity. This statement remains true even if there are more than the three flavors of neutrino (ν_e , ν_μ , and ν_τ) that we have been taking into account. However, from experimental studies of the decays $Z \rightarrow \nu_\alpha \bar{\nu}_\alpha$ of the Z boson, we know that any additional flavors of neutrino beyond ν_e , ν_μ , and ν_τ do not couple to the Z . The SM then implies that these additional flavors do not couple to the W either. Neutrinos that do not couple to the SM W or Z bosons, and consequently do not participate in any known interaction other than gravity, are called sterile neutrinos. Such neutrinos may participate in some as-yet-unknown interaction that lies beyond the SM. However, any such interaction is invisible at presently accessible neutrino energies, so sterile neutrinos will leave no trace in a neutrino detector. Thus, if we start with a beam of neutrinos of one of the active flavors (i.e., ν_e , ν_μ , or ν_τ), and some of the neutrinos in this beam oscillate into sterile neutrinos, an experiment that can measure the *total* active flux in the beam (i.e., the sum of the ν_e , ν_μ , and ν_τ fluxes) will find that some of the active flux has vanished.

Among the special cases of the oscillation probability formula of Eq. (31), the best known is the one that describes oscillation when only two mass eigenstates are important. Let us call these mass eigenstates ν_1 and ν_2 , and the two neutrinos of definite flavor that we can construct as superpositions of ν_1 and ν_2 , ν_α and ν_β . There is only one squared-mass splitting, $m_2^2 - m_1^2 \equiv \Delta m^2$, in this physical system, and the mixing matrix U is 2×2 . It can be shown that, as far as neutrino oscillation is concerned, if U is unitary, it may be taken to be given by

$$U \equiv \begin{pmatrix} U_{\alpha 1} & U_{\alpha 2} \\ U_{\beta 1} & U_{\beta 2} \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \quad (33)$$

In this expression, the angle θ is referred to as the mixing angle. Inserting this mixing matrix and the single Δm^2 into Eq. (31), we find immediately that for $\beta \neq \alpha$

$$P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) = \sin^2 2\theta \sin^2\left(\Delta m^2 \frac{L}{4E}\right). \quad (34)$$

For no flavor change, we find that

$$P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\alpha) = 1 - \sin^2 2\theta \sin^2\left(\Delta m^2 \frac{L}{4E}\right). \quad (35)$$

2.3 Neutrino Flavor Change in Matter

Many of the experimental studies of neutrino flavor change that have been carried out have involved neutrinos that travel through matter. In some cases, interaction of the neutrinos with electrons in the matter significantly modifies the flavor content of the beam, relative to what it would be in vacuum. Treatments of the effect of matter on neutrino flavor change may be found, for example, in Refs. [1] and [9]. Here we shall make just a few brief comments.

Coherent forward scattering of an electron neutrino ν_e from electrons in matter, caused by W -boson exchange, gives the ν_e an extra interaction potential energy

$$V = +\sqrt{2}G_F N_e. \quad (36)$$

Here, G_F is the Fermi coupling constant of the weak interaction, and N_e is the number of electrons per unit volume. Correspondingly, an electron antineutrino $\bar{\nu}_e$ traveling through matter has an extra interaction potential energy

$$\bar{V} = -\sqrt{2}G_F N_e . \quad (37)$$

These extra energies raise the effective mass of a ν_e in matter, and lower that of a $\bar{\nu}_e$. A useful measure of the fractional importance of this matter effect on an oscillation involving a vacuum mass splitting Δm^2 is given by the parameter

$$x = \frac{\sqrt{2}G_F N_e}{\Delta m^2/2E} . \quad (38)$$

On the right-hand side of this relation, the numerator is the extra energy of a ν_e due to matter interaction, and the denominator is the quantity with dimensions of energy that occurs in the relative phase of two interfering terms in the vacuum oscillation amplitude of Eq. (27).

We see from Eq. (38) that the matter effect grows with the neutrino energy E . As Eq. (38) suggests, the matter effect is sensitive to the sign of Δm^2 . That is, it can be used to determine which of two mass eigenstates with known couplings to the various charged leptons is the heavier one. Owing to the fact that $\bar{V} = -V$, matter affects antineutrinos differently than it affects neutrinos. As a result, an observed difference between the oscillation in matter of antineutrinos and neutrinos can have two sources: 1) CP violation coming from a mixing matrix U that is not real, as may be seen from Eq. (31), and 2) the matter effect. Experiments seeking to demonstrate that neutrino oscillation violates CP will have to disentangle these two effects.

3 A Brief Guide to References

The lectures presented at the 2011 European School of High Energy Physics, and the 2011 International School on Astro Particle Physics devoted to Neutrino Physics and Astrophysics, covered a number of topics in addition to the physics of neutrino oscillation. One may study those other topics in the following references:

- History of neutrino oscillation results: Refs. [1] and [10].
- Physics of Majorana neutrinos and neutrinoless double beta decay: Refs. [11–13].
- Leptogenesis: Refs. [7, 8].
- Recent experimental results, and the status of our knowledge: The original papers in this fast-moving field, and Ref. [14].

References

- [1] See, for example, C. Giunti and C. Kim, *Fundamentals of Neutrino Physics and Astrophysics* (Oxford University Press, Oxford, 2007).
- [2] B. Kayser, *Proceedings of the SLAC Summer Institute of 2004*, eConf **C040802**, L004 (2004), hep-ph/0506165.
- [3] B. Kayser, *Proceedings of the 9th Int. Conf. on Flavor Physics and CP Violation*, Ed. A. Soffer, eConf **C110523**, arXiv: 1110.3047. What follows is based on this reference, and quotes a bit of its text.
- [4] This treatment of a problem involving particle mixing follows a method described in B. Kayser and L. Stodolsky, *Phys. Lett.* **B359**, 343 (1995).
- [5] This was first noticed by E. Akhmedov and A. Smirnov, *Found. Phys.* **41**, 1279 (2011), arXiv: 1008.2077.
- [6] B. Kayser, *Phys. Rev.* **D24**, 110 (1981).

- [7] S. Davidson, E. Nardi, and Y. Nir, *Phys. Rept.* **466**, 105 (2008), arXiv: 0802.2962.
- [8] B. Kayser, *Proceedings of the 22nd Rencontres de Blois*, Eds. L. Celniker, J. Dumarchez, B. Klima, and J. Trân Thanh Vân (Gioi Publishers, Vietnam, 2011), arXiv: 1012.4469.
- [9] J. Bahcall, *Neutrino Astrophysics* (Cambridge University Press, Cambridge, 1989).
- [10] M. Fukugita and T. Yanagida, *Physics of Neutrinos and Applications to Astrophysics* (Springer, Berlin, 2003).
- [11] S. Elliott and P. Vogel, *Ann. Rev. Nucl. Part. Sci.* **52**, 115 (2002), hep-ph/0202264.
- [12] W. Rodejohann, *Int. J. Mod. Phys.* **E20**, 1833 (2011), arXiv: 1106.1334.
- [13] B. Kayser, *Proceedings of the Carolina International Symposium on Neutrino Physics*, Eds. F. Avignone, R. Creswick, K. Kubodera, and M. Purohit (IOP Publishing, Bristol, 2009), arXiv: 0903.0899.
- [14] Particle Data Group (K. Nakamura, *et al.*), *J. Phys. G* **37**, 1 (2010), accessible at pdg.lbl.gov, and its periodic updates.

Beyond the Standard Model

B.A. Dobrescu

Theoretical Physics Department, Fermilab, Batavia, IL, USA

Abstract

Despite the success of the standard model in describing a wide range of data, there are reasons to believe that additional phenomena exist, which would point to new theoretical structures. Some of these phenomena may be discovered in particle physics experiments in the near future. These lectures overview hypothetical particles, solutions to the hierarchy problem, theories of dark matter, and new strong interactions.

1 What is the standard model?

The Standard Model of particle physics is an $SU(3)_c \times SU(2)_W \times U(1)_Y$ gauge theory with quarks transforming in the $q_L^i \sim (3, 2, 1/6)$, $u_R^i \sim (3, 1, 2/3)$ and $d_R^i \sim (3, 1, -1/3)$ representations of the gauge group, and leptons transforming as $L_L^i \sim (1, 2, -1/2)$ and $e_R^i \sim (1, 1, -1)$. The index $i = 1, 2, 3$ labels the generations of fermions. The Standard Model also includes a single Higgs doublet [1] transforming as $H \sim (1, 2, 1/2)$. The vacuum expectation value (VEV) of the Higgs doublet, $\langle H \rangle = (v_H, 0)$, breaks the electroweak $SU(2)_W \times U(1)_Y$ symmetry down to the gauge symmetry of electromagnetism, $U(1)_{\text{em}}$. With these fields of spin 1 (gauge bosons), 1/2 (quarks and leptons) and 0 (Higgs doublet), the Standard Model is remarkably successful at describing a tremendous amount of data [2] in terms of a single mass parameter (the electroweak scale $v_H \approx 174 \text{ GeV}$) and 18 dimensionless parameters [3].

If one requires that the Lagrangian contains only renormalizable interactions, then the Standard Model cannot accommodate gravity or neutrino masses. However, the requirement of renormalizability goes beyond the sound comparison of theory and experiment, by imposing theoretical constraints at energy scales well above those accessible in current experiments.

Gravity is nicely imbedded in an extension of the Standard Model through the inclusion of a graviton (massless spin-2 particle) with dimension-5 couplings to the stress-energy tensor [4] suppressed by the Planck scale $M_P = 2 \times 10^{19} \text{ GeV}$. Graviton exchange reproduces (up to order $1/M_P$ effects) general relativity, so this theory describes gravitational interactions with sufficient accuracy for practical purposes (issues related to quantum gravity are not experimentally accessible in the foreseeable future).¹

Neutrino masses can be included in a couple of ways which may be differentiated in principle through future experiments [5]. An important dichotomy is whether the neutrino masses are of Majorana or Dirac type. Majorana masses may be obtained from dimension-5 operators of the type

$$\frac{c_{ij}}{M_N} H H \left(\overline{L}_L^i L_L^j \right) . \quad (1)$$

The dimensionless coefficients c_{ij} then determine the elements of the neutrino mass matrix through $m_{\nu}^{ij} = c_{ij} v^2 / M_N$. Imposing $c_{ij} < O(1)$, the measured atmospheric neutrino mass-squared difference $|\Delta m_{\text{atm}}^2| \approx (0.05 \text{ eV})^2$ requires the mass scale where the description in terms of dimension-5 operators breaks down to satisfy $M_N \lesssim 10^{14} \text{ GeV}$. The nonrenormalizable operators (1) may be generated by the tree-level exchange of gauge-singlet fermions (commonly called "right-handed neutrinos"² and labelled by ν_R) or $SU(2)_W$ -triplet particles, or by loops involving various new particles.

¹General relativity includes an additional mass scale, the cosmological constant, which given current knowledge appears to be independent of the Planck scale. The accelerated expansion of the Universe indicates that the cosmological constant minus the vacuum energy density is tiny (but nonzero). Ignoring issues about fine-tuning (discussed in Section 2), this may be a Standard Model effect because all existing particles contribute to the vacuum energy density.

²"Right-handed neutrino" is a potentially confusing name given that left- and right-handed fermions may be interchanged by a charge conjugation operation.

If the neutrino masses are of the Dirac type, then there are right-handed neutrinos which combine with the left-handed ones through Yukawa terms in the Lagrangian:

$$-\lambda_{ij}^\nu \bar{\nu}_R^i H L_L^j \quad (2)$$

The Yukawa couplings λ_{ij} must be very small, $\lambda_{ij} \lesssim 10^{-12}$, which implies that the ν_R 's cannot be observed directly (unless they happen to have sizable interactions with other new particles [6]). The observation of atmospheric and solar neutrino oscillations has established that at least two of the neutrino masses are nonzero, and therefore there is need for at least two ν_R 's. More complicated origins of neutrino masses can also be imagined: for example, a collection of higher-dimensional operators can contribute to both Majorana and Dirac masses [6].

I will refer to the Standard Model plus the graviton and neutrino masses (either Majorana or Dirac) as the $\text{SM}\nu\mathcal{G}$, and I will refer to it loosely as the standard model. In the case of only two ν_R 's and purely Dirac masses the $\text{SM}\nu\mathcal{G}$ has 27 parameters: 19 from the SM, the Planck scale, the cosmological constant, two ν masses, three ν mixing angles, and a ν CP-violating phase; only the latter parameter is experimentally unconstrained for now, and all the parameters from the neutrino sector are dimensionless.

Exercise 1.1: Count the parameters of the $\text{SM}\nu\mathcal{G}$ if the neutrino masses are purely of the Majorana type as in Eq. (1).

2 Evidence for physics beyond the standard model

Let us overview the existing experimental and observational information regarding physics beyond the $\text{SM}\nu\mathcal{G}$. The only robust piece of information (as of May 2012) comes from various observations of galaxy clusters, the galactic rotation curves, and cosmological data: there is need for at least one new electrically-neutral and stable particle to play the role of dark matter [2]. We do not know what are the spin, mass or interactions of this particle, and in fact we do not know whether there is a single particle or a complicated hidden sector including various kinds of particles. Section 4 reviews some popular theories of dark matter particles.

Currently there are no other convincing experimental results which cannot be explained within the $\text{SM}\nu\mathcal{G}$. There are, however, theoretical reasons to expect that there is new physics beyond the $\text{SM}\nu\mathcal{G}$. The most often invoked one is the so-called hierarchy problem, discussed in Section 3. Other theoretical problems of the standard model include the following:

- The pattern of measured quark and lepton masses suggests the existence of some underlying mechanism that generates the hierarchies between various fermion masses. Proposed solutions include, for example, fermion masses generated by higher-dimensional operators with coefficients controlled by discrete symmetries, loop-induced masses in the presence of some new particles, or exponential suppressions due to different wave functions along extra dimensions.
- The strong CP problem is the question of why QCD does not lead to large CP violation. The nonobservation of a neutron dipole electric moment implies that the operator $\epsilon_{\mu\nu\lambda\tau} G^{\mu\nu} G^{\lambda\tau}$, with $G^{\mu\nu}$ being the gluon field strength, has a coefficient smaller than 10^{-12} . A nice theoretical explanation for this seemingly unnatural coefficient is the presence of an global $U(1)$ symmetry explicitly broken by a QCD anomaly. This solution implies the existence of an axion field, of spin 0 and very small mass, which although is still allowed by all experimental constraints, is increasingly constrained.
- The $SU(3) \times SU(2) \times U(1)$ charges of the quarks and leptons are strange enough (at least at first sight) to require some explanation in a deeper theory. A beautiful explanation is that the six fields ($q_L, u_R, d_R, l_L, e_R, \nu_R$) of each generation are exactly the components of a single $SO(10)$ representation. Furthermore, the $SU(3) \times SU(2) \times U(1)$ gauge couplings appear to unify at a scale of about

10^{16} GeV, suggesting an underlying Grand Unified Theory. It is conceivable though that the SM fermions just happen to form full GUT representations without an actual unification gauge group; after in any self-consistent theory the fermion charges are highly restricted by the requirement of gauge anomaly cancellation.

- The asymmetry between matter and antimatter in the observable universe suggests a mechanism for baryogenesis which requires some new sources of CP violation. There is however enough leeway to engineer such a mechanism, for example through the neutrino sector, so that it might not have observable effects in upcoming experiments.

Overall, one should keep in mind that theoretical reasons for new physics are usually based on esthetic prejudice and do not necessarily need to be valid, given the self-consistency of the $\text{SM}\nu\mathcal{G}$ up to scales near M_P . Only experimental or observational data that is in disagreement with $\text{SM}\nu\mathcal{G}$ (such as that pertaining to dark matter) may provide conclusive evidence for physics beyond the standard model.

Exercise 2.1: *The $SO(10)$ grand unified gauge group includes an $SU(5) \times U(1)$ subgroup. Assign the SM fermions of one generation to the $10 + \bar{5}$ representation of $SU(5)$.*

3 Hierarchy problem

The $\text{SM}\nu\mathcal{G}$ has two seemingly independent mass scales (ignoring for now the cosmological constant¹): the electroweak scale v_H and the Planck scale M_P . The ratio of these is tiny, 10^{-17} , which raises the question of whether there is some underlying dynamics that links the two mass scales. Furthermore, quantum effects tend to push the electroweak scale towards the Planck scale, so there is need for some stabilizing mechanism if one does not want to rely on extreme tuning of parameters in the underlying theory. Perturbatively, this is the statement that the squared-mass of the Higgs doublet receives quadratic divergences from loops, primarily involving the top quark but also the Higgs doublet itself and the electroweak bosons.

There are various ideas about how to solve the hierarchy problem, so let us discuss them briefly, in turn. The most explored ones are dynamically broken supersymmetry (Section 3.1) and a warped extra dimension (Section 3.2). However, several other ideas have been proposed:

- *Higgs as a pseudo Nambu-Goldstone boson.* A global continuous symmetry which is spontaneously broken implies the existence of a massless spin-0 particle, called Nambu-Goldstone boson (NGB). If the global symmetry is also explicitly broken, then the NGB acquires a mass, and is referred to as pseudo NGB. In the limit where the explicit breaking is controlled by a mass parameter much smaller than the VEV responsible for spontaneous breaking, the pseudo NGB is much lighter than other spin-0 particles expected to have mass of the order of the VEV. If the SM is embedded in a theory in which the Higgs doublet arises as a pseudo NGB, then the hierarchy problem is solved. However, it is rather difficult to design such a mechanism without introducing additional elementary scalar fields and thereby reintroducing a hierarchy problem. A well known proposal [7, 8] is to have a new strongly interacting gauge theory acting on some new fermions such that their chiral symmetry is dynamically broken. Some of the ensuing pseudo NGBs (analogues of the pions in QCD) may carry the same quantum numbers as the Higgs doublet. For more recent studies of related theories, see *e.g.* Ref. [9].

- *Composite Higgs boson from top condensation.* The large mass of the top quark suggests that it is involved in electroweak symmetry breaking. If some new interaction binds a top quark and a top anti-quark within a composite spin-0 field, then at sufficiently strong coupling the composite field acquires a VEV that breaks the electroweak symmetry. The large coupling of the composite field to its constituents (*i.e.*, a $t\bar{t}$ pair) then leads to a large top mass. It turns out, however, that the top is not heavy enough

to accommodate this mechanism unless the scale of the new interaction is many orders of magnitude above the weak scale, which would require fine-tuning. Nevertheless, if there is a new quark with mass of order 1 TeV which mixes with the top quark, then a bound state of χ with t behaves at low energy exactly as the SM Higgs doublet [10]. To solve the hierarchy problem, it is then necessary to have a new asymptotically-free gauge interaction (‘topcolor’ [11]) which is spontaneously broken near the scale (in the TeV range) where it becomes strong.

- *Technicolor.* Instead of being broken by the VEV of a Higgs doublet, the electroweak symmetry may be broken by the VEV of a fermion-antifermion pair (usually called a ‘fermion bilinear’). This phenomenon actually exists within QCD, where the chiral symmetry breaking is triggered by a $\bar{q}q$ condensate. That breaking however, is three orders of magnitude weaker than that required to fit the W and Z masses. Technicolor is a new gauge theory in which some new gauge interactions breaks the chiral symmetry of some new fermions, called techni-fermions [12]. Given that there is no Higgs doublet or any other elementary scalar field, technicolor solves beautifully the hierarchy problem. However, technicolor faces various phenomenological hurdles.

- *Large extra dimensions (ADD).* If the SM is localized on a 3-dimensional wall embedded in a 3+n-dimensional space and only the graviton propagates in the full space, then the coupling of the graviton to SM particles is suppressed by volume of the extra dimensions [13]. This allows the observed strength of the gravitational interactions to be very small, suppressed by M_P , while the fundamental mass M_* that sets the scale of strongly coupled gravity to be near 1 TeV. As a result, the ADD scenario replaces the hierarchy problem by the question of what makes the volume of the 3+n dimensional space so large in $1/M_*$ units. One could imagine various ways of tackling the latter question. Measurements of gravity are sensitive to the presence of the extra dimensions, and the current upper limit on their radius is around 4×10^{-5} m [14]. The collider implications of this scenario involve emission of KK gravitons which appear as missing transverse momentum, and virtual exchange of KK gravitons which induce nonresonant modifications of various cross sections (the relevant Feynman rules are derived in Ref. [15]).

- *Little Higgs theories.* The 1-loop quadratic divergences in the Higgs self-energy may be cancelled by a set of new particles having the same spin as their SM partners. The divergence due to the top-quark loop is cancelled by the divergence of a top-prime quark loop. The coupling of the top-prime quark to the Higgs boson is given by a dimension-5 operator whose coefficient is linked to the top Yukawa coupling by a symmetry [16]. The extension of this mechanism to more than one loop is not obvious. An interesting version of the Little Higgs mechanism is provided by theories with T parity [17]. These are discussed further in Section 4.

- *Twin Higgs.* Twin Higgs is a mechanism for canceling the 1-loop quadratic divergences similar with Little Higgs in that there is a partner of same spin for each SM particle that has a large contribution to the Higgs self-energy. There is an essential peculiarity of the Twin-Higgs mechanism though: the partners do not carry SM gauge charges [18]. Consequently, there are no easy-to-find signals at the LHC.

Some of the solutions outlined above may also be combined to obtain different mechanisms. For example, the Twin Higgs mechanism may cancel the leading 1-loop divergences, and then the compositeness of the Higgs doublet (for example in a top condensation model) may be natural at a scale in the several TeV range. Given the relatively large number of proposed solutions to the hierarchy problem, and the possibility that many other solutions might be identified in the future, it is difficult to draw firm conclusions based on the requirement of no excessive fine-tuning.

3.1 Dynamically-broken supersymmetry

Supersymmetry is an extension of Lorentz invariance to a superspace that includes an extra dimension with an anticommuting ("fermionic") coordinate. Exact supersymmetry implies that each field has a fermionic and a bosonic component of equal mass, such that loops involving fermions contributing to the squared-mass of the Higgs doublet exactly cancel loops involving bosons.

We know that supersymmetry is not an exact symmetry of nature because, for example, there is no bosonic partner of the electron of mass equal with that of the electron. It is possible, however, that supersymmetry is an exact symmetry of the underlying theory but the ground state of the universe breaks supersymmetry (this is equivalent with a nonzero vacuum energy). If that is the case then only certain terms in the Lagrangian break supersymmetry. These so-called soft susy-breaking terms lead to modified masses for superpartners (and some trilinear interactions between scalar fields). As a result the Higgs quadratic divergences are suppressed at a scale given by the superpartner masses.

To make these statements more precise, let us first discuss the particle content of the supersymmetric theory that includes the $SM \nu \mathcal{G}$, usually called the Minimal Supersymmetric Standard Model (MSSM) [19]. For each SM quark, which includes a left-handed and a right-handed component, there are two complex scalar fields called squarks. Likewise, for each SM left- or right-handed lepton there is a complex scalar field called slepton. For each $SU(3)_c \times SU(2)_W \times U(1)_Y$ gauge boson there is a spin-1/2 field (a Majorana fermion) generically called gaugino; gluino refers specifically to the $SU(3)_c$ gauge superpartner while wino, zino, and bino (or photino) are the superpartners of the electroweak gauge bosons. The graviton has a spin-3/2 partner called gravitino. Finally, there are two Higgs doublets (in order to have no gauge anomaly), H_u and H_d , and each has a spin-1/2 partner called higgsino.

Supersymmetry ensures that the quadratically divergent part of the top loop contribution to the Higgs mass is cancelled by the loops involving top squarks ("stops"), as shown in Fig. 1. The remnant of this cancellation is of the order of the stop masses times a loop factor, so that the fine-tuning is not worrisome provided the stop masses are not larger than order 1 TeV (more precise statements imply upper bounds of about 500–700 GeV [20]). A light left-handed stop implies that the other scalar from the same $SU(2)_W$ doublet, namely the superpartner of the left-handed bottom quark ("left-handed sbottom"), should also be light [21]. Similarly, loops with Higgs doublets are cancelled by loops with higgsinos, loops with W and Z bosons are cancelled by loops with winos and zinos, and so on for all the particles.

Supersymmetry with soft breaking terms enforces these cancellations of quadratic divergences at any numbers of loops. For practical purposes, corrections involving more than two loops are less relevant for inferring an upper limit on superpartner masses. The gluino contributes at two loops to the Higgs mass, so that one obtains an upper limit of the gluino mass of about 1.5 TeV. This conclusion is more model dependent than the upper limit on the stop masses. For example, in models beyond the MSSM where the gauginos are Dirac particles [22], the two loop contributions are automatically suppressed, so significantly heavier gluinos are natural.

The presence of two Higgs doublets in the MSSM allows a mass term for the Higgs superfields, of the type $\mu H_u H_d$. The parameter μ is a supersymmetric mass, so that its contribution to the mass-squared

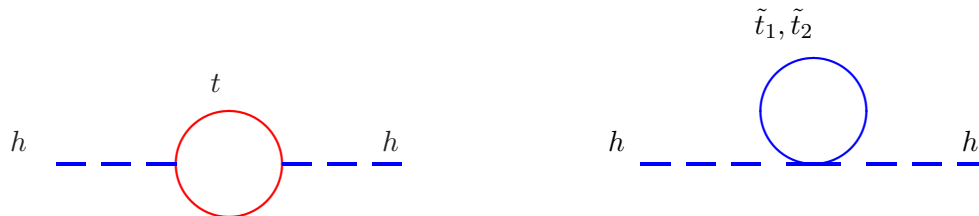


Fig. 1: Quadratic divergences to the self-energy of the Higgs doublet. First diagram is the dominant contribution in the SM, while the second one arises in the MSSM and cancels the quadratic divergence due to the SM top loop.

of each Higgs doublet is $|\mu|^2$ while the higgsino mass is $|\mu|$. This implies that the higgsinos cannot be heavier than the electroweak scale without fine-tuning. The relation between the various contributions to the Higgs square masses can be written as

$$M_Z^2 = -2 \left(M_{H_u}^2 + \delta M_{H_u}^2 + |\mu|^2 \right) , \quad (3)$$

where $\delta M_{H_u}^2$ is the 1-loop correction to the squared mass of H_u :

$$\delta M_{H_u}^2 \approx -3 \frac{\lambda_t^2}{8\pi^2} \left(m_{\tilde{t}_1}^2 + m_{\tilde{t}_2}^2 \right) \ln \left(\frac{\Lambda}{m_{\tilde{t}}} \right) . \quad (4)$$

where Λ is the scale where the soft-susy terms are generated. Thus, the hierarchy problem requires the MSSM to have soft-susy breaking terms such that the stops have masses near the electroweak scale, and a μ term such that the higgsinos are even lighter. This raises two questions: why are the soft susy masses so much smaller than the Planck scale, and why is the supersymmetric μ term close in size to the soft susy-breaking masses?

There is a nice answer to the first question: a gauge theory with a gauge coupling of order one at the Planck scale may become nonperturbative at a much smaller scale due to the logarithmic running of the gauge coupling. If that type of dynamics leads to a nonzero vacuum energy, then soft susy breaking terms are generated and a large hierarchy between the Planck scale and the scale of supersymmetry breaking may be naturally generated. The gauge theory that breaks supersymmetry is called the dynamically susy-breaking (DSB) sector. Supersymmetry breaking can be transmitted from the DSB sector to the MSSM by gauge interactions (gauge mediation, with $\Lambda \sim O(10^4)$ TeV), Planck scale suppressed operators ('supergravity mediation', with $\Lambda \sim O(10^7)$ TeV), or other mechanisms.

Given that the μ -term has nothing to do with supersymmetry breaking, there is need for some additional connection between the DSB sector and the generation of the μ term. Although this is highly non-trivial, various realistic models of dynamically broken supersymmetry have been constructed (for example, see Ref. [23] for models of gauge mediation). Nevertheless, these are theories with a rich field content and rather complicated dynamics at various scales between the electroweak and Planck scales. The majority of the phenomenological studies concentrate on the MSSM without being concerned with the exact mechanism of DSB. The phenomenology of the MSSM is discussed in Section 4.

Exercise 3.1: *Check the cancellation of quadratic divergences in Fig. 1.*

3.2 A warped extra dimension: RS1

Consider the existence of an extra spatial dimension of coordinate y , transverse to the usual 3+1 space-time dimensions of coordinates x^μ , such that the line element is given by

$$ds^2 = e^{-2ky} \eta_{\mu\nu} x^\mu x^\nu - y^2 . \quad (5)$$

The 5D Einstein equations need to have a solution given certain boundary conditions. Here k is the AdS_5 curvature, and has dimensions of mass. The unit of length along the 4D Minkowski space depends on the position along $y \equiv x^4$. Randall and Sundrum [24] have shown that a slice of AdS_5 (space exists only for $0 \leq y \leq L$) is a solution to the 5D Einstein equations provided the vacuum energy $V(y)$ satisfies

$$V(0) = -V(L) = \frac{\Lambda}{k} , \quad (6)$$

where Λ is the 5D cosmological constant. They proposed the following set-up ("RS1"): gravity propagates in the 5D bulk, while all standard model fields are localized at $y = L$. The points $y = 0$ and $y = L$ are usually referred to as the Planck brane and the standard model brane, respectively. It is assumed that

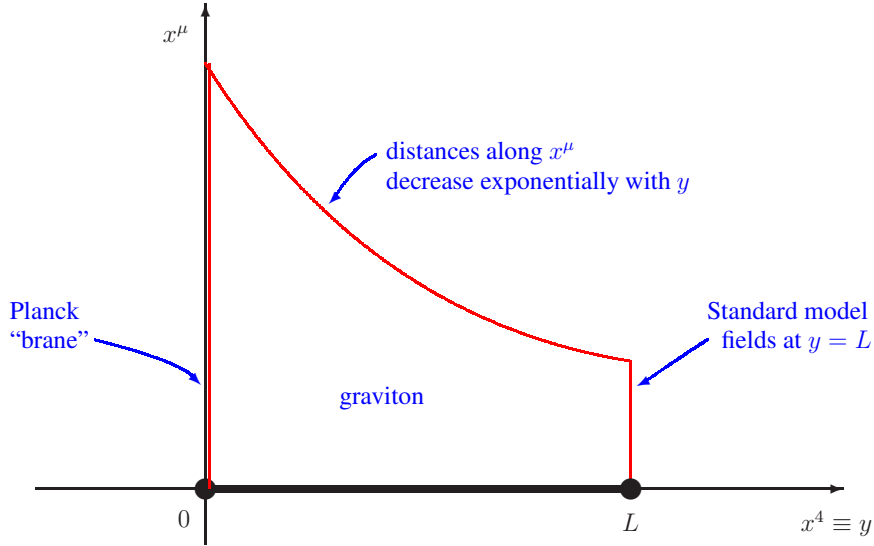


Fig. 2: RS1 model: only the graviton propagates in the warped bulk.

the curvature k is a couple of orders of magnitude below the fundamental 5D scale M_5 , where 5D gravity becomes strongly coupled, so that the effective theory has a range of validity. The reduced Planck mass can be derived in terms of M_5 and $k \gg 1/L$:

$$\overline{M}_{\text{Pl}} \approx \sqrt{\frac{M_5^3}{k}}, \quad (7)$$

and the 5D cosmological constant is given by $\Lambda = -24M_5^3 k^2$.

As a consequence of the metric, any mass parameter m_0 that appears in some terms of the 5D action which are localized at $y = L$ corresponds to a mass m in the effective 4D theory obtained after integrating over y , with

$$m = m_0 e^{-kL}. \quad (8)$$

Thus, if the VEV that breaks the $SU(2)_W \times U(1)_Y$ symmetry in the 5D theory is $v_0 \sim O(k)$, then the corresponding VEV in the 4D effective theory is $v = v_0 e^{-kL}$. Taking $M_5 \approx 10k$ one finds $v \simeq 174 \text{ GeV}$ for $kL \approx 34$. This is a remarkable result: the huge ratio $\overline{M}_{\text{Pl}}/v$ arises from a theory in which there are no large hierarchies between the input mass parameters ($M_5, k, 1/L, v_0$).

Exercise 3.2: Change the Minkowski coordinates in Eq. (5) so that in terms of the new coordinates all distances are given as measured by an observer localized at $y = L$. Show that all input mass parameters are of order the TeV scale in this case, and that the observed weakness of gravity is due to the small wavefunction of the 0-mode graviton on the standard model brane.

The gravity action [4] in the warped bulk is given by

$$S_{\text{gravity}} = 2M_5^3 \int d^4x \int_0^L dy \sqrt{\det G} \left[G^{\alpha\beta} \left(\partial_\beta \Gamma_{\alpha\delta}^\delta - \partial_\delta \Gamma_{\alpha\beta}^\delta \right. \right. \\ \left. \left. + \Gamma_{\alpha\delta}^\eta \Gamma_{\eta\beta}^\delta - \Gamma_{\alpha\beta}^\eta \Gamma_{\eta\delta}^\delta \right) + 12k^2 \right], \quad (9)$$

where $G^{\alpha\beta}$ is the 5D metric, $\det G$ is the determinant of the metric, and $\Gamma_{\alpha\beta}^\delta$ is the 5D connection:

$$\Gamma_{\alpha\beta}^\delta = \frac{1}{2} G^{\delta\eta} (\partial_\alpha G_{\beta\eta} + \partial_\beta G_{\eta\alpha} - \partial_\eta G_{\alpha\beta}). \quad (10)$$

Letters from the beginning of the Greek alphabet label the 5D coordinates ($\alpha, \beta, \delta, \eta, \dots = 0, 1, 2, 3, 4$), and letters from the middle of the Greek alphabet label the Minkowski coordinates ($\mu, \nu, \rho, \sigma, \dots = 0, 1, 2, 3$).

Small gravitational fluctuations are described by expanding the 5D metric about the warped background:

$$G_{\alpha\beta} = \begin{pmatrix} e^{-2ky} \eta_{\mu\nu} & 0 \\ 0 & -1 \end{pmatrix} + \frac{2}{M_5^{3/2}} \begin{pmatrix} e^{-2ky} h_{\mu\nu}(x^\rho, y) & h_{\mu 4}(x^\rho, y) \\ h_{4\nu}(x^\rho, y) & h_{44}(x^\rho, y) \end{pmatrix}. \quad (11)$$

Here $h_{\mu\nu}$ is the 5D graviton polarized along the Minkowski coordinates, while $h_{\mu 4}$ and h_{44} are the polarizations along the extra dimension of the 5D graviton; $h_{\mu 4}$ form a 5D spin-1 field called graviphoton, and h_{44} is a 5D spin-0 field called the graviscalar. The 5D graviton $h_{\alpha\beta}$ is a symmetric tensor field, and therefore it has 15 components. However, not all of its components are physical. Imposing a gauge fixing condition that $h_{\alpha\beta}$ is traceless and transverse,

$$h^\alpha_\alpha = 0, \quad \partial^\alpha h_{\alpha\beta} = 0, \quad (12)$$

eliminates 10 of the components. The remaining five components are physical degrees of freedom, divided as follows: two in $h_{\mu\nu}$ (a traceless and transverse 4×4 symmetric tensor), two in $h_{\mu 4}$ (a massless gauge field with 4-components), and one in h_{44} .

Exercise 3.3: Keeping only the quadratic terms in $h^{\alpha\beta}$, derive the Lagrangian that describes the propagation of the graviton in the warped background.

The graviton must have a massless 0-mode, $h_{\mu\nu}^{(0)}$, in order to generate the observed long-range gravitational effects. Hence, the graviton is an even field, *i.e.*, its derivative with respect to y must vanish at $y = 0$ and $y = L$. The KK decomposition for the tensor components of the 5D graviton field is

$$h_{\mu\nu}(x^\rho, y) = \frac{1}{\sqrt{L}} \left[h_{\mu\nu}^{(0)}(x^\rho) + \sum_{j \geq 1} h_{\mu\nu}^{(j)}(x^\rho) \chi_j(y) \right]. \quad (13)$$

The graviton KK modes, $h_{\mu\nu}^{(j)}(x)$, are spin-2 particles in 4D. For $j \geq 1$, they are massive and therefore they have 5 degrees of freedom (corresponding angular momenta $\pm 2, \pm 1$ and 0). Two of these degrees of freedom originate in $h_{\mu\nu}(x^\rho, y)$, while the other three are given at each KK level by the graviphoton and graviscalar, which in the unitary gauge disappear from the spectrum. General coordinate invariance requires $h_{\mu 4}$ to have odd boundary conditions, and h_{44} to have even boundary conditions [25]. As a result, the 0-mode of h_{44} remains as physical spin-0 particle, which is referred to as the radion. Its properties are analyzed in Ref. [26].

Plugging the KK decomposition (13) into the 5D kinetic terms for $h_{\mu\nu}(x^\rho, y)$ gives the following differential equation for the KK functions [27]:

$$\left[\frac{d}{dy} \left(e^{-4ky} \frac{d}{dy} \right) + m_j^2 e^{-4ky} \right] \chi_j(y) = 0, \quad (14)$$

where m_j is the mass of the j th KK mode of spin-2. Solving this equation with the Neumann boundary conditions

$$\frac{d}{dy} \chi_j(y)|_{y=0} = \frac{d}{dy} \chi_j(y)|_{y=L} = 0 \quad (15)$$

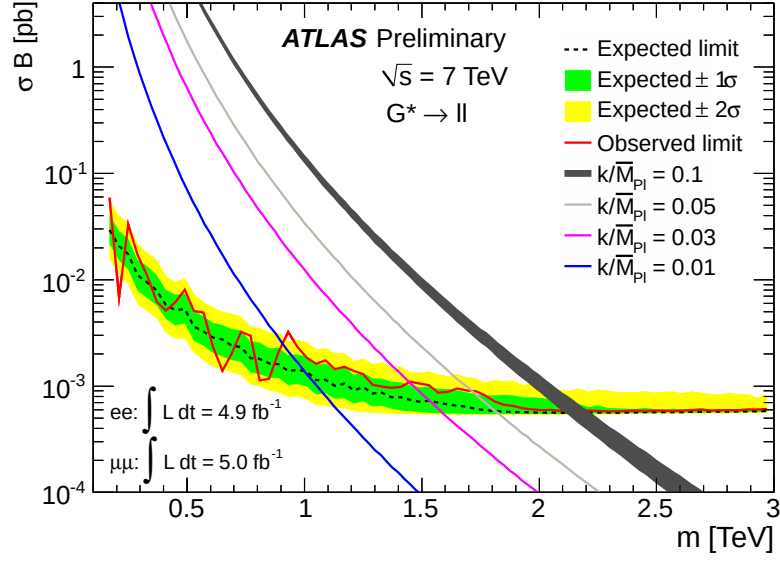


Fig. 3: Limits on the cross section times branching fraction for the level-1 graviton of the RS1 model, obtained by the ATLAS Collaboration by searching for a narrow e^+e^- or $\mu^+\mu^-$ resonance. This plot is reproduced from Ref. [28].

leads to KK functions given in terms of the J_2 and Y_2 Bessel functions:

$$\chi_j(y) = \frac{e^{2ky}}{N_j} \left[J_2 \left(\frac{m_j}{k} e^{ky} \right) + \alpha_j Y_2 \left(\frac{m_j}{k} e^{ky} \right) \right], \quad (16)$$

where N_j is a normalization constant, and α_j are constants determined by the boundary conditions (15). The mass of the level-1 KK modes is at the TeV scale, $m_1 \simeq 3.8ke^{-kL}$, and the higher modes have masses $m_2 \simeq 1.8m_1$, $m_3 \simeq 2.7m_1$, $m_4 \simeq 3.5m_1$, ...

The couplings of the graviton KK modes to standard model fields are identical to those of the graviton 0-mode except for an overall factor of e^{kL} . Therefore, both the masses of the graviton KK modes and the suppression of the couplings are of order the TeV scale. The collider signals of these spin-2 particles are heavy resonances that decay into neutral pairs of standard model particles. The production cross section at the Tevatron or the LHC may be computed from the following couplings to gluons

$$\frac{\sqrt{2}}{\overline{M}_{\text{Pl}}} e^{kL} \left(h_{\rho}^{(j)\rho} \eta_{\mu\nu} - 4h_{\mu\nu}^{(j)} \right) G_{\sigma}^{\mu} G^{\nu\sigma}, \quad (17)$$

and to quarks

$$\frac{4}{\overline{M}_{\text{Pl}}} e^{kL} \left[\left(h_{\sigma}^{(j)\sigma} \eta_{\mu\nu} - h_{\mu\nu}^{(j)} \right) ((\partial^{\nu} \bar{q}) \gamma^{\mu} q - \bar{q} \gamma^{\mu} \partial^{\nu} q) - m_q h_{\sigma}^{(j)\sigma} \bar{q} q \right]. \quad (18)$$

Exercise 3.4: Compute the parton-level cross sections $gg \rightarrow h_{\mu\nu}^{(1)}$ and $q\bar{q} \rightarrow h_{\mu\nu}^{(1)}$, in the narrow width approximation, as a function of the curvature k and the $h_{\mu\nu}^{(1)}$ mass.

The most useful decays of the graviton KK modes are into $\gamma\gamma$, e^+e^- and $\mu^+\mu^-$, due to the small backgrounds. Figure 3 shows the current Tevatron limit on the RS1 model. Figure 4 shows the predicted cross section for $pp \rightarrow h_{\mu\nu}^{(j)} \rightarrow \ell^+\ell^-$ at the LHC.

Standard model in a warped extra dimension. It is interesting to study what are the effects of the warped metric on the propagation along the extra dimension of fields other than the graviton. Even though this scenario is not motivated by the hierarchy problem, the remainder of this Section is focused on this highly active topic.

Boundary conditions at $y = 0$ and $y = L$ must be specified for each field propagating in the bulk. A gauge field propagating in a warped extra dimension, with Neumann boundary conditions, has a y -independent 0-mode [30]. This flat profile, forced by gauge invariance, is in stark contrast to the exponential profile of the graviton 0-mode. The Kaluza-Klein decomposition for the gauge field is

$$A_\mu(x^\nu, y) = \frac{1}{\sqrt{2L}} \left[A_\mu^{(0)}(x^\nu) + \sum_{j \geq 1} A_\mu^{(j)}(x^\nu) f_j(y) \right]$$

where the KK functions are given by

$$f_j(y) = \frac{e^{ky}}{N_j} \left[J_1\left(\frac{m_j}{k} e^{ky}\right) - \frac{J_0(m_j/k)}{Y_0(m_j/k)} Y_1\left(\frac{m_j}{k} e^{ky}\right) \right]$$

Here J and Y are Bessel functions, N_j is a normalization constant, M_j is the mass of the j th KK mode of spin-1. The level-1 gauge boson has a mass $M_1 = 2.5k e^{-kL}$, so at the scale associated with the standard model brane, 1.5 times lighter than the level-1 graviton. The higher spin-1 KK modes have masses $M_2 = 2.3M_1$, $M_3 = 3.6$, $M_4 = 4.8$, ... The couplings of the higher spin-1 KK modes to the fermions localized on the standard model brane are all given by the 4D gauge coupling (*i.e.*, the coupling of the 0-mode gauge boson) times a “volume” factor of $\sqrt{2kL} \approx 8.2$. Therefore, if the standard model gauge bosons propagate in the warped bulk while the quark and leptons are localized at $y = L$, then the spin-1 KK modes are strongly coupled to the fermions. The 4-fermion effective interactions induced by KK exchange are ruled out unless $M_1 > O(20)$ TeV (this is based on the assumption that single boson exchange is a good approximation of the effects induced by the rather strongly coupled KK modes).

If fermions are also propagating in the warped bulk, then their 0-modes have an exponential profile [31]. In addition, the exponential profile may be tuned independently for each fermion flavor, because

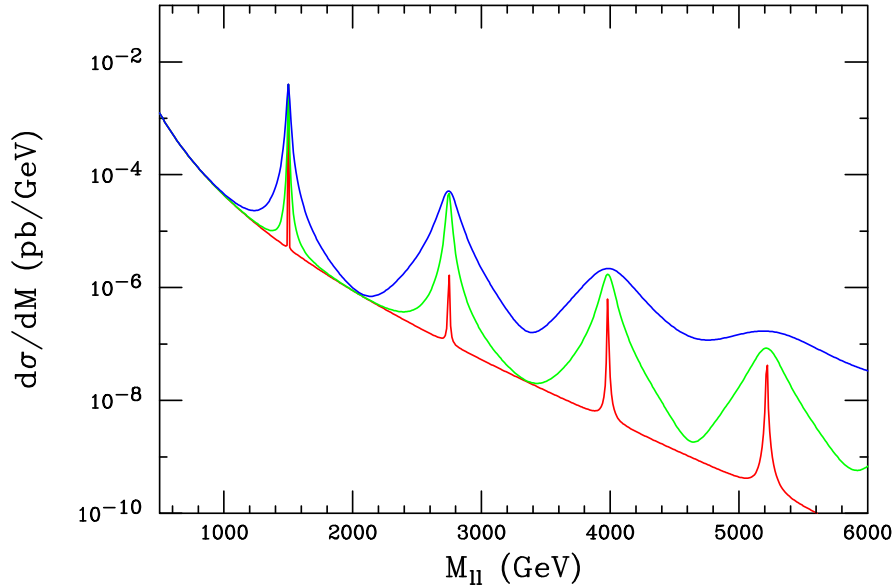


Fig. 4: Theoretical cross section for $pp \rightarrow h_{\mu\nu}^{(j)} \rightarrow \ell^+\ell^-$ at the 14 TeV LHC [29]. The three curves correspond, from top to bottom, to $k/\overline{M}_{Pl} = 0.1, 0.05$ and 0.01 . The dilepton resonance searches at the 7 TeV LHC already exclude a KK graviton of 1.5 TeV with $k/\overline{M}_{Pl} > 0.03$, as shown in Fig. 3.

the fermions may have bulk masses. The fermionic part of the 5D Lagrangian takes the following form:

$$\sqrt{g} \left(i\bar{\psi}\gamma^\mu\partial_\mu\psi - c_\psi k\bar{\psi}\psi \right) , \quad (19)$$

where g is the determinant of the 5D metric. The ensuing y -dependence of the 0-mode is

$$f_\psi(y) = f_\psi(L)e^{-(2-c_\psi)k(L-y)/2} . \quad (20)$$

This result may be important for explaining the hierarchies among the standard model fermion masses. The set of lectures [32] is devoted to model building in a warped extra dimension and their holographic interpretation.

Exercise 3.5: *Show that the fermion bulk mass does not prevent the existence of a massless 0-mode (when the Higgs VEV is neglected).*

4 Theories of dark matter and their collider signatures

The total mass of dark matter is roughly five times larger than the mass of luminous matter. The most studied and searched for type of dark matter is made of electrically-neutral and stable particles with mass of the order of the electroweak scale. The latter property is motivated by the so-called WIMP miracle: a particle of mass $\sim v_H$ and coupling ~ 1 has a relic abundance in agreement with the observed dark matter density.

In order to ensure the stability of the dark matter particle, there is need for a symmetry. The simplest choice is a discrete symmetry such as Z_2 (a parity). A discrete symmetry may be imposed on many models, leaving many options for the types of WIMPs. There are three popular theories that include a Z_2 symmetry for dark matter: supersymmetry (see Section 4.1), universal extra dimensions (see Section 4.2) and Little Higgs with T parity [17]. It turns out that there are some deep theoretical connections between these theories: supersymmetry implies the existence of a fermionic extra dimension, Little Higgs with T parity involves a deconstructed extra dimension³, while UED involves real extra dimensions. At the same time, the properties of the dark matter candidates are completely different in these three theories. For example the spin of the dark matter particle is a fermion in supersymmetry, a boson of spin 1 in minimal UED, and a spin-0 particle in Little Higgs with T parity (as well as in the case of two UEDs).

4.1 Supersymmetric standard model

The Minimal Supersymmetric Standard Model contains many new particles (33) and many new parameters (105), so that it is not surprising that it leads to many interesting signatures. In order to solve the hierarchy problem, the masses of some superpartners must be near the electroweak scale, as discussed in Section 3.1.

Supersymmetry by itself does not guarantee the presence of a dark matter candidate. However, in order to make the MSSM viable, it is necessary to include some discrete symmetry that prevents proton decays. The usual choice (albeit there are many alternatives) for that symmetry is the so called R -parity, under which the SM particles are even while their superpartners are odd. This eliminates the dangerous QQd and QLu operators (when both are present, they lead to fast proton decay). Furthermore, R -parity ensures that the lightest superpartner (LSP) is a stable particle. The mass spectrum of the superpartners depends on the mechanism through which supersymmetry breaking is communicated to the SM fields.

³The construction of 4D theories which are equivalent to 5D theories involving gauge fields and fermions in the bulk is detailed in Ref. [33].

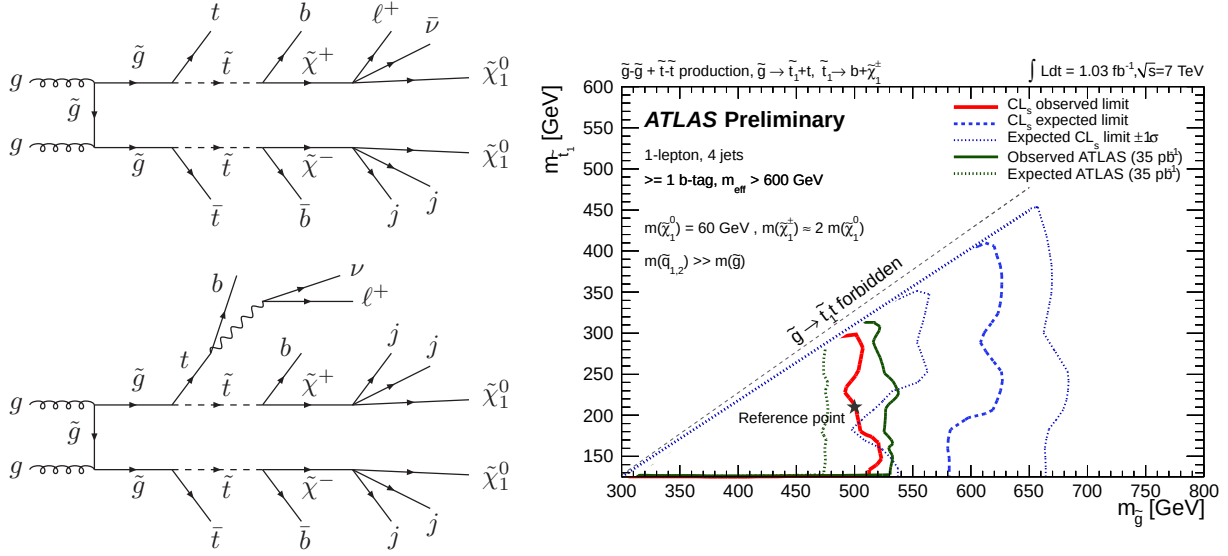


Fig. 5: Left: Typical diagrams for gluino pair production followed by cascade decays involving stops. Right: limits from ATLAS [35] on the stop and gluino masses using the $\ell + b$ jets $+\cancel{E}_T$.

In many instances the lightest superpartner is the bino, but there are well studied models where it is the gravitino. Both of these particles are electrically-neutral and represent viable dark matter candidates.

R-parity implies that superpartners can be produced only in pairs and that they undergo a series of decays that end with the LSP. The gluino is a color-octet fermion, so that QCD requires an interaction of a gluon to two gluinos with a coupling set by the strong gauge coupling $g_s = \sqrt{4\pi\alpha_s}$. As a result, at hadron colliders the production cross section is rather large for gluino pair production. Thus, the most common signatures involve the cascade decays of two gluinos, which necessarily involve a number of jets and either missing energy due to the LSPs escaping the detector or charged stable particles if their lifetime is long enough to have displaced vertices. The squarks are color-triplet scalars, so that there are interactions of two squarks with one or two gluons, but the cross section for pair production of squarks is an order of magnitude smaller than that of gluinos of comparable mass.

A popular assumption about the mass spectrum is that the gauginos have a unified mass at the GUT scale, and also the squarks and sleptons have the same mass at the GUT scale, which leads to a certain mass spectrum at the TeV scale. The typical signature is jets plus missing E_T plus leptons [19]. Given the nonobservation of this signature implies that the superpartner masses are above the TeV scale, which makes this scenario less appealing.

A less constrained scenario is that where the third generation squarks are the only light squarks (see *e.g.* Ref. [34]). Their decays into a quark and the LSP then lead to the final states as those shown in Fig. 5. Note that the production mechanism in Fig. 5 relies on not too heavy gluinos. A more convincing test of MSSM is direct $\tilde{t}\tilde{t}^*$ production. However, this has small rate, large background, and there is model dependence in the final state: for example, $t\bar{t} + \cancel{E}_T$ if the LSP is the bino, or $t\bar{t}\tau^+\tau^-\tau^+\tau^- + \cancel{E}_T$ in gauge-mediation where the LSP is the gravitino, as shown in Fig. 6.

Another important test of the MSSM is direct production of higgsinos, because their mass is set by the electroweak scale, as explained in Section 3.1. To be more precise, the physical states are not just higgsinos, but rather admixtures of higgsinos and electroweak gauginos called charginos (charge ± 1) or neutralinos (electrically neutral). An off-shell W boson can produce a chargino and a neutralino, leading for example to relatively clean $3\ell + \cancel{E}_T$ signatures, but the rate for this process is an order of magnitude smaller than squark pair production.

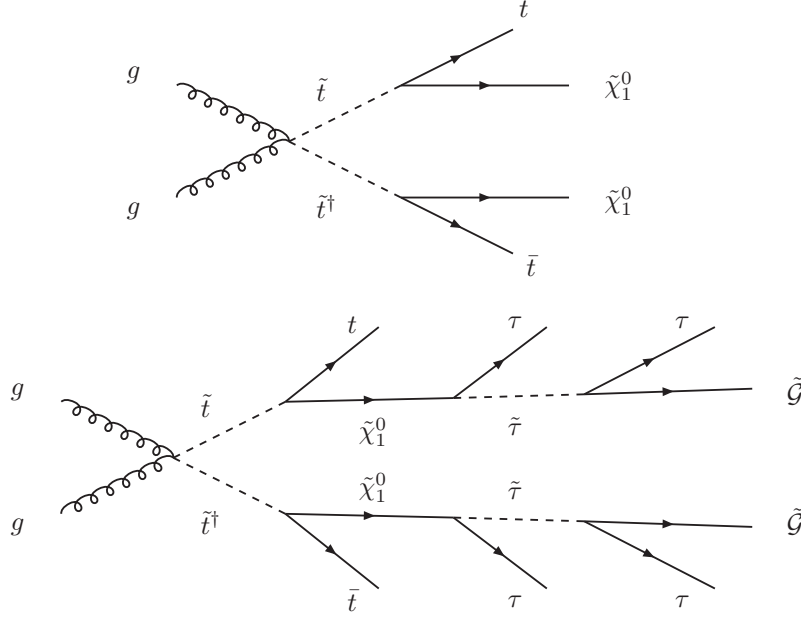


Fig. 6: Direct production of a stop pair leading to the $t\bar{t} + \cancel{E}_T$ or $t\bar{t} + 4\tau + \cancel{E}_T$ final states.

Exercise 4.1a: Draw Feynman diagrams for stop pair production followed by cascade decays involving the chargino ($\chi^\pm$) decay to a $W^\pm$ boson and a neutralino (χ_1^0). Propose search strategies of these processes at the LHC.

Exercise 4.1b: Find the final states arising from chargino pair production (through an off-shell Z or photon), due to the chargino interactions with a W and a neutralino, or with a lepton and a slepton.

4.2 Universal extra dimensions

Any particle propagating through extra dimensions, whether compactified on a circle or an interval, would appear in experiments as a tower of massive particles in 3 spatial dimensions. The presence of these KK modes can be easily understood based on the usual particle-in-a-box problems in quantum mechanics: given that space along the extra dimensions is compact, the energy states are quantized. The kinetic energy due to motion along the extra dimensions manifests itself as mass in the usual 3 spatial dimensions. For an interval of length L , the mass of the lightest KK modes is $(\pi/L)\hbar/c$ (the natural unit system, $\hbar = c = 1$, is used in what follows). This mass is the compactification scale, and its inverse $R \equiv L/\pi$ is the ‘radius’ of the extra dimension.

Thus, any extra-dimensional field theory is equivalent to a 4-dimensional one that includes a series of heavy particles. The spectrum and interactions of these KK particles depend on the boundary conditions and metric. Circle compactification would not allow any particle discovered so far to propagate along the extra dimension. The reason is that any gauge field, such as the photon, would include a spin-0 partner of equal mass and couplings as the spin-1 particle. Furthermore, any fermion that propagates along the extra dimension would be a vectorlike fermion: its left- and right-handed components would have the same gauge charges, which is not true for any of the elementary fermions discovered so far. However, if the compactification is on an interval, then the unwanted vectorlike partners of the observed fermions and the spin-0 partners of the gauge fields may be eliminated by the boundary conditions at the end of the interval.

If all bosons propagate in extra dimensions while the fermions are localized at the end points of

an interval [36, 37], the KK modes of the standard model gauge bosons appear as s -channel resonances at the LHC so that one can set limits on $1/R$ of several TeV.

Universal extra dimensions (UED) are arguably the simplest kind of extra dimensions: all particles propagate along some flat compact extra dimensions. The remarkable feature of UED is that a remnant of translational invariance along the extra dimensions is preserved such that a single KK mode cannot couple at tree-level to zero modes [38]. As a result, the limits are relaxed by an order of magnitude compared to the extra dimensions accessible only to bosons. Furthermore, UED lead to dramatically different phenomenological implications. The lightest KK particle is a dark matter candidate, and the collider signals include cascade decays involving leptons, jets and missing energy, as well as narrow resonances.

Field theories in extra dimensions are strongly coupled in the ultraviolet, so that their study could shed light on nonperturbative phenomena, with possible applications to dynamical electroweak symmetry breaking and compositeness. A useful review of strongly coupled theories and their relation to extra dimensions is given in Ref. [12]. A thorough presentation of the phenomenology of various extra dimensional models can be found in the TASI lectures of G. Kribs [39]. The implications of universal extra dimensions for dark matter, which continue to be analyzed by many groups, are reviewed in Ref. [40].

4.2.1 Field theory on a flat compact dimension

Before discussing the phenomenology of extra dimensions, it is necessary to study the general features of quantum field theory in a flat extra dimension. The cases of spin 0, 1/2 and 1 are analyzed in turn.

Scalar field on the interval. Let us consider a five-dimensional (5D) spacetime: four spacetime dimensions of coordinates x^μ , $\mu = 0, 1, 2, 3$, form the usual Minkowski spacetime, and one transverse spatial dimension of coordinate x^4 is flat and compact, with $0 \leq x^4 \leq L$. Thus the extra dimension is an interval (see Fig. 7), and the boundary conditions at its end points determine the spectrum of KK modes.

Free scalar fields, $\Phi(x^\mu, x^4)$, are described by the following action:

$$S_\Phi = \int d^4x \int_0^L dx^4 \left(\partial_\alpha \Phi^\dagger \partial^\alpha \Phi - M_0^2 \Phi^\dagger \Phi \right). \quad (1)$$

The parameter M_0 is the 5D mass of Φ . We use letters from the beginning of the Greek alphabet to label the 5D coordinates $\alpha, \beta, \dots = 0, 1, 2, 3, 4$, and letters from the middle of the Greek alphabet to label the Minkowski coordinates $\mu, \nu, \dots = 0, 1, 2, 3$. Given that the action is dimensionless, and that the coordinates have mass dimension -1 , the 5D bosons have mass dimension $+3/2$.

Under a variation of the field, $\delta\Phi(x^\mu, x^4)$, the variation of the action is given by

$$\delta S_\Phi = \delta S_\Phi^v + \delta S_\Phi^s, \quad (2)$$

where the first term is a ‘volume’ integral,

$$\delta S_\Phi^v = - \int d^4x \int_0^L dx^4 \left(\partial^\alpha \partial_\alpha \Phi^\dagger + M_0^2 \Phi^\dagger \right) \delta\Phi, \quad (3)$$



Fig. 7: The extra dimension of coordinate x^4 extends from $x^4 = 0$ to $x^4 = L$, and is transverse to the usual three spatial dimensions.

and the second term is a ‘surface’ integral,

$$\delta S_{\Phi}^s = \int d^4x \left(\partial_4 \Phi^\dagger \delta \Phi \Big|_{x^4=L} - \partial_4 \Phi^\dagger \delta \Phi \Big|_{x^4=0} \right) . \quad (4)$$

Here we have assumed as usual that the field vanishes at $x^\mu \rightarrow \pm\infty$. Given that the action has to be stationary with respect to *any* variation of the field, the volume and surface terms must vanish independently. Requiring $\delta S_{\Phi}^v = 0$ implies that Φ is a solution to the 5D Klein-Gordon equation,

$$(\partial^\mu \partial_\mu - \partial_4^2 + M_0^2) \Phi = 0 , \quad (5)$$

while $\delta S_{\Phi}^s = 0$ forces the boundary conditions that can be imposed on Φ to obey

$$(\partial_4 \Phi^\dagger) \delta \Phi \Big|_{x^4=L} = (\partial_4 \Phi^\dagger) \delta \Phi \Big|_{x^4=0} . \quad (6)$$

Given that the values of $\delta \Phi(x^\mu, x^4)$ at $x^4 = 0$ and $x^4 = L$ are in general not correlated (unless the two points are identified, which would not allow chiral fermions in the 4D theory), and Eq. (6) must be valid for any $\delta \Phi$, both the left- and right-handed sides of Eq. (6) must vanish. Therefore,

$$\partial_4 \Phi \Big|_{x^4=0} = 0 \quad \text{or} \quad \Phi(x^\mu, 0) = 0 \quad (7)$$

and

$$\partial_4 \Phi \Big|_{x^4=L} = 0 \quad \text{or} \quad \Phi(x^\mu, L) = 0 . \quad (8)$$

We now solve the 5D Klein-Gordon equation,

$$(\partial^\mu \partial_\mu - \partial_4^2 + M_0^2) \Phi = 0 , \quad (9)$$

subject to the boundary conditions (7) and (8). Since the boundary conditions are independent of x^μ , then Φ can be decomposed in Fourier modes as follows:

$$\Phi(x^\mu, x^4) = \sum_j \Phi^{(j)}(x^\mu) f^j(x^4) . \quad (10)$$

The 4D scalar fields $\Phi^{(j)}$ (‘KK modes’ or excitations), satisfy

$$(\partial^\mu \partial_\mu + M_0^2 + M_j^2) \Phi^{(j)}(x^\mu) = 0 , \quad (11)$$

where M_j^2 is a positive eigenvalue. The f^j functions are solutions to the one-dimensional equation,

$$(\partial_4^2 + M_j^2) f^j(x^4) = 0 . \quad (12)$$

A general solution to the above equation is

$$f^j(x^4) = C_+ e^{ijx^4/R} + C_- e^{-ijx^4/R} , \quad (13)$$

where $C_\pm$ are complex coefficients, and j is a real number such that

$$M_j = \frac{j}{R} , \quad (14)$$

and we defined the ‘compactification radius’

$$R \equiv \frac{L}{\pi} . \quad (15)$$

The boundary conditions (7) and (8) impose a relation between the two coefficients, $C_- = \pm C_+$, and also restrict the values of j : $e^{4ij\pi} = 1$. Furthermore, the normalization condition,

$$\int_0^L dx^4 [f^j(x^4)]^* f^{j'}(x^4) = \delta_{jj'} , \quad (16)$$

determines the last coefficient up to a phase factor which we choose to be one. Explicitly, the solutions to Eq. (12) can be written as

$$f_0^j(x^4) = \frac{1}{\sqrt{L}(1 + \delta_{j,0})} \cos\left(\frac{jx^4}{R}\right) \quad (17)$$

for $\partial_4 \Phi|_{x^4=0} = \partial_4 \Phi|_{x^4=L} = 0$ (Neumann boundary conditions),

$$f_1^j(x^4) = \frac{1}{\sqrt{L}} \sin\left(\frac{jx^4}{R}\right) \quad (18)$$

for $\Phi(x^\mu, 0) = \Phi(x^\mu, L) = 0$ (Dirichlet boundary conditions),

$$f_2^j(x^4) = \frac{1}{\sqrt{L}} \sin\left(\frac{(j - 1/2)x^4}{R}\right) \quad (19)$$

for $\Phi(x^\mu, 0) = \partial_4 \Phi|_{x^4=L} = 0$ (mixed boundary conditions),

$$f_3^j(x^4) = \frac{1}{\sqrt{L}} \cos\left(\frac{(j - 1/2)x^4}{R}\right) \quad (20)$$

for $\partial_4 \Phi|_{x^4=0} = \Phi(x^\mu, L) = 0$ (mixed boundary conditions),

with j an integer called ‘KK number’.

The functions f_n^j form a complete orthonormal set on the interval if

$$\sum_j [f_n^j(x^4)]^* f_n^j(x'^4) = \delta(x'^4 - x^4) . \quad (21)$$

The allowed values for j must be chosen such that the above completeness condition is satisfied. It is straightforward to check that $j \geq 0$ for $n = 0$, and $j \geq 1$ for $n = 1, 2, 3$. For $n = 0$ there is a state ($j = 0$) of zero momentum (‘zero mode’) along the compact dimension.

Exercise 4.2: *Integrate the action (1) over x^4 and show that the 4D particles $\Phi^{(j)}(x^\mu)$ have masses*

$$M^{(j)} = \sqrt{M_0^2 + \frac{j^2}{R^2}} , \quad (22)$$

if the boundary conditions are of the type $n = 0$ or 1, and

$$M^{(j)} = \sqrt{M_0^2 + \frac{(j - 1/2)^2}{R^2}} , \quad (23)$$

if $n = 2$ or 3.

In what follows we will concentrate on the KK functions f_0 and f_1 , which are usually referred to as even and odd, respectively; the corresponding boundary conditions represent the so called S^1/Z_2 orbifold.

Fermions on the interval: chiral boundary conditions. We now turn to free spin-1/2 fields in five dimensions. The Clifford algebra is generated by five anti-commuting matrices: Γ^α , $\alpha = 0, 1, 2, 3, 4$. The minimal dimensionality of these matrices is 4×4 . The Γ^α matrices can be used to construct a spinor representation of the $SO(1, 4)$ Lorentz group, with the generators explicitly given by

$$\frac{\Sigma^{\alpha\beta}}{2} = \frac{i}{4} [\Gamma^\alpha, \Gamma^\beta] . \quad (24)$$

The spin-1/2 fermions in five dimensions have four components. In terms of the usual γ^μ matrices used in 4D field theory, one may take $\Gamma^\mu = \gamma^\mu$ and $\Gamma^4 = i\gamma_5$. The fact that $\Gamma^4 \propto \Gamma^0\Gamma^1\Gamma^2\Gamma^3$ implies that the $SO(1, 4)$ Lorentz group has a single spin-1/2 representation, and thus the 5D fermions are vectorlike.

Upon compactification of x^4 , the $SO(1, 3)$ Lorentz symmetry generated by $\Sigma^{\mu\nu}/2$, $\mu, \nu = 0, 1, 2, 3$, remains unbroken. There are two chiralities under $SO(1, 3)$, labeled as usual by L and R . These are projected by

$$P_{L,R} = \frac{1}{2} (1 \pm i\Gamma^4) . \quad (25)$$

A 5D fermion, Ψ , decomposes into two fermions of definite chirality under $SO(1, 3)$:

$$\Psi(x^\mu, x^4) = \Psi_L(x^\mu, x^4) + \Psi_R(x^\mu, x^4) , \quad (26)$$

where

$$\Psi_{L,R} \equiv P_{L,R} \Psi . \quad (27)$$

As in Section 2, we consider the compactification on an interval: $0 \leq x^4 \leq L$. The action for a free 5D field of spin 1/2 and mass zero is

$$S_\Psi = \int d^4x \int_0^L dx^4 \frac{i}{2} [\bar{\Psi} \Gamma^\alpha \partial_\alpha \Psi - (\partial_\alpha \bar{\Psi}) \Gamma^\alpha \Psi] . \quad (28)$$

Note that the 5D fermions have mass dimension +2. Under an arbitrary variation of the field, $\delta\Psi(x^\mu, x^4)$, the action has to be stationary both inside the interval and on its boundary:

$$\begin{aligned} \delta S_\Psi^v &= - \int d^4x \int_0^L dx^4 i (\partial_\alpha \bar{\Psi}) \Gamma^\alpha \delta\Psi = 0 , \\ \delta S_\Psi^s &= \frac{i}{2} \int d^4x \int_0^L dx^4 (\bar{\Psi} \Gamma^4 \delta\Psi|_{x^4=L} - \bar{\Psi} \Gamma^4 \delta\Psi|_{x^4=0}) \\ &= 0 . \end{aligned} \quad (29)$$

The first equation implies that Ψ is a solution to the 5D Weyl equation, which can be decomposed into two equations:

$$\begin{aligned} \Gamma^\mu \partial_\mu \Psi_L &= -\Gamma^4 \partial_4 \Psi_R , \\ \Gamma^\mu \partial_\mu \Psi_R &= -\Gamma^4 \partial_4 \Psi_L . \end{aligned} \quad (30)$$

The second equation (29) restricts the values of Ψ on the boundary.

In the case of a fermion whose zero-mode is left-handed, the boundary conditions are as follows:

$$\begin{aligned} \partial_4 \Psi_L(x^\mu, 0) &= \partial_4 \Psi_L(x^\mu, L) = 0 , \\ \Psi_R(x^\mu, 0) &= \Psi_R(x^\mu, L) = 0 . \end{aligned} \quad (31)$$

The ensuing KK decomposition is given by

$$\Psi = \frac{1}{\sqrt{L}} \left\{ \Psi_L^{(0)}(x^\mu) + \sqrt{2} \sum_{j \geq 1} \left[\Psi_L^{(j)}(x^\mu) \cos\left(\frac{jx^4}{R}\right) + \Psi_R^{(j)}(x^\mu) \sin\left(\frac{jx^4}{R}\right) \right] \right\} . \quad (32)$$

All fermion KK modes for $j \geq 1$ pair up to form vectorlike fermions of Dirac masses $M^{(j)}$ as given in Eq. (22). In the case of a fermion whose zero-mode is right-handed, the above equations apply with left- and right-handed labels interchanged. The conclusion is that the boundary conditions for the left- and right-handed fermions are forced by the stationarity of the action to eliminate the zero mode for one of the chiralities. This is true for the interval compactification discussed here, while the compactification on a circle would preserve both the left- and right-handed zero modes.

Exercise 4.3: Write down an Yukawa interaction between a bulk fermion and a bulk scalar, and show that the 4D Yukawa coupling λ (obtained after integration over x^4) is given in terms of the 5D Yukawa coupling λ_5 by

$$\lambda = \frac{\hat{\lambda}}{\sqrt{L}} . \quad (33)$$

Gauge fields on the interval. A 5D gauge boson has five components: $A_\mu(x^\nu, x^4)$, $\mu, \nu = 0, 1, 2, 3$, and $A_4(x^\nu, x^4)$ which corresponds to the polarization along the extra dimension. From the point of view of the 4D theory, A_4 is a tower of spinless KK modes.

Exercise 4.4: Write down the action for a 5D gauge boson, and then imposing the stationarity of the action under an arbitrary field variation, derive the field equations.

The boundary conditions consistent with gauge invariance are given by

$$\begin{aligned} \partial_4 A_\mu(x^\nu, 0) &= \partial_4 A_\mu(x^\nu, L) = 0 , \\ A_4(x, 0) &= A_4(x, L) = 0 . \end{aligned} \quad (34)$$

Solving the field equations with these boundary conditions yields the following KK expansions:

$$\begin{aligned} A_\mu &= \frac{1}{\sqrt{L}} \left[A_\mu^{(0)}(x^\nu) + \sqrt{2} \sum_{j \geq 1} A_\mu^{(j)}(x^\nu) \cos\left(\frac{jx^4}{R}\right) \right] , \\ A_4 &= \sqrt{\frac{2}{L}} \sum_{j \geq 1} A_4^{(j)}(x^\nu) \sin\left(\frac{jx^4}{R}\right) . \end{aligned} \quad (35)$$

The zero-mode $A_\mu^{(0)}(x^\nu)$ is the massless gauge boson associated with the 4D gauge transformations. Note that A_4 does not have a zero-mode. In the unitary gauge, the $A_4^{(j)}(x^\nu)$ KK modes are the longitudinal components of the heavy spin-1 KK modes $A_\mu^{(j)}(x^\nu)$.

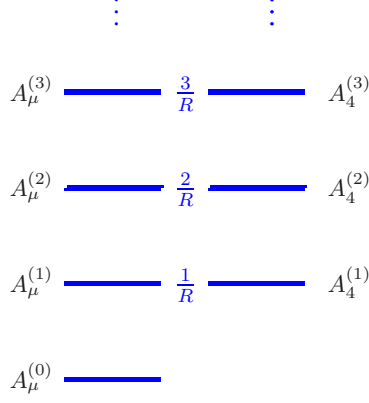


Fig. 8: The spectrum of KK modes for the gauge boson associated with an unbroken gauge symmetry.

Assuming that the gluon is a 5D field and that the quarks are localized at $x^4 = 0$, the terms of the 4D Lagrangian describing the interactions between gluon KK modes and quarks are given by

$$\begin{aligned}
 \mathcal{L}_{4D} &= \int_0^L dx^4 \hat{g} G_\mu^a(x^\nu, x^4) [\delta(x^4) \bar{q}(x^\nu) \gamma^\mu T^a q(x^\nu)] \\
 &= g_s \left(G_\mu^{(0)a} + \sqrt{2} \sum_{j \geq 1} G_\mu^{(j)a}(x) \right) \bar{q} \gamma^\mu T^a q
 \end{aligned} \tag{36}$$

From the above equation follows that the 4D gauge coupling g_s is given in terms of the 5D gauge coupling $\hat{g}$ by $g_s = \hat{g}/\sqrt{L}$, and that the coupling of any gluon KK mode to quarks is larger than the QCD coupling by a factor of $\sqrt{2}$. This factor is a consequence of the normalization of the KK functions, which in turn follows from the canonical normalization of the kinetic terms.

If the gauge symmetry is broken by the vacuum expectation value, v , of a bulk scalar field Φ with Neumann boundary conditions, then the KK modes of the CP-odd components of Φ mix with the corresponding $A_4^{(j)}$. At each KK level one linear combination becomes the longitudinal degree of freedom of $A_\mu^{(j)}$ and the orthogonal one appears as a spin-0 particle. The masses of both $A_\mu^{(j)}$ and of the spin-0 particles are given by

$$M_A^{(j)} = \sqrt{g^2 v^2 + \frac{j^2}{R^2}}. \tag{37}$$

where g is the 4D gauge coupling.

4.2.2 One universal extra dimension

Being equipped with the basics of field theory in five dimensions, we can now discuss the case where all standard model particles propagate along one flat extra dimension compactified on an interval, *i.e.*, one universal extra dimension.

Ignoring electroweak-symmetry breaking effects, the tree-level spectrum consists of equally spaced KK levels (of mass j/R), and on each level the KK modes for all standard model particles are degenerate. Each standard model chiral fermion has a tower of vectorlike modes. The KK spectrum of the $(t, b)_L$ doublet and t_R is illustrated in Fig. 9.

To understand the effects of electroweak symmetry breaking on the KK fermion spectrum, let us analyze the case of the top quark (the same applies to the other standard model fermions, except that

electroweak symmetry breaking effects are suppressed by their small Yukawa couplings). Let us denote the bulk fermion whose zero mode is t_L (which is part of a weak doublet $\mathcal{Q}_3$) by $\mathcal{Q}_t(x^\mu, x^4)$, and the bulk fermion whose zero mode is t_R (weak singlet) by $\mathcal{U}_3(x^\mu, x^4)$. The terms of the 4D Lagrangian responsible for top KK masses come from the kinetic terms along the extra dimension and from the Yukawa couplings to the Higgs doublet:

$$\mathcal{L}_{4D} = \int_0^L dx^4 \left[\overline{\mathcal{Q}}_t i \gamma^4 (-\partial_4) \mathcal{Q}_t + \overline{\mathcal{U}}_3 i \gamma^4 (-\partial_4) \mathcal{U}_3 + \left(-\hat{\lambda}_t H \overline{\mathcal{Q}} \mathcal{U}_3 + \text{H.c.} \right) \right] \quad (38)$$

The bulk Higgs doublet, H , which is an even field, has a negative-squared 5D mass; this leads to a vacuum expectation value $v \simeq 174$ GeV only for the zero mode $H^{(0)}$. Inserting now the KK decompositions for $\mathcal{Q}_t$ and $\mathcal{U}_3$, which are analogous to Eq. (32), we obtain the following terms responsible for level- j top masses:

$$\begin{aligned} & \frac{-2}{L} \int_0^L dx^4 \left\{ \overline{\mathcal{Q}}_{tR}^{(j)} \mathcal{Q}_{tL}^{(j)} \sin \frac{jx^4}{R} \partial_4 \cos \frac{jx^4}{R} + \overline{\mathcal{U}}_{3R}^{(1)} \mathcal{U}_{3L}^{(j)} \cos \frac{jx^4}{R} \partial_4 \sin \frac{jx^4}{R} \right. \\ & \left. + \lambda_t v \left[\overline{\mathcal{Q}}_{tR}^{(j)} \mathcal{U}_{3L}^{(j)} \sin^2 \left(\frac{jx^4}{R} \right) + \overline{\mathcal{U}}_{3R}^{(j)} \mathcal{Q}_{tL}^{(j)} \cos^2 \left(\frac{jx^4}{R} \right) \right] \right\} + \text{H.c.} \end{aligned} \quad (39)$$

where $\lambda_t = \hat{\lambda}_t / \sqrt{L} \approx 1$ is the standard model top Yukawa coupling. From the above equation it is clear that electroweak symmetry breaking leads to mixing between the towers of vectorlike fermions that have t_L and t_R as zero modes. After integration over x^4 we get the mass terms for the top quark KK modes at level j :

$$\left(\overline{\mathcal{Q}}_{tR}^{(j)}, \overline{\mathcal{U}}_{3R}^{(j)} \right) \begin{pmatrix} -\frac{j}{R} & \lambda_t v \\ \lambda_t v & \frac{j}{R} \end{pmatrix} \begin{pmatrix} \mathcal{Q}_{tL}^{(j)} \\ \mathcal{U}_{3L}^{(j)} \end{pmatrix}. \quad (40)$$

Due to the minus sign in the 11 element, the two eigenvalues are equal at tree level:

$$M_{t_1}^{(j)} = M_{t_2}^{(j)} = \sqrt{\frac{j^2}{R^2} + m_t^2}. \quad (41)$$

The mass degeneracy among the level- j modes of various standard model particles is further lifted by loop corrections [42]. An important property of any type of interaction in a 5D theory is that it grows with the energy, such that it becomes nonperturbative at a certain energy scale Λ . In the equivalent 4D description that includes a tower of KK modes, the number of KK modes grows with the energy

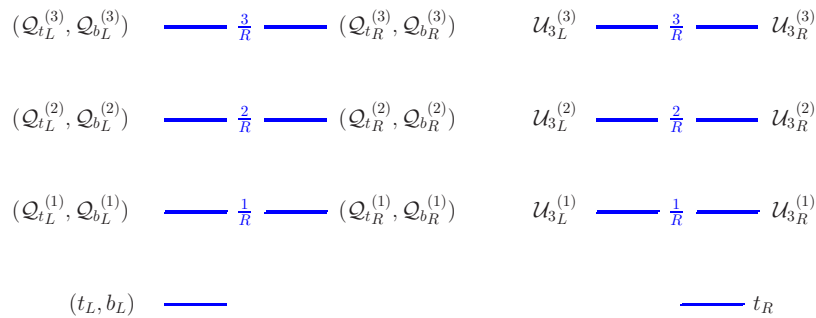


Fig. 9: The tree-level spectrum of top-quark KK modes, ignoring electroweak symmetry breaking.

scale such that the loop corrections are divergent; they need to be cutoff at a scale Λ , which means that the tower of modes is truncated. In practice, the QCD interactions become nonperturbative in the UV at scales roughly two orders of magnitude above the compactification scale. Hence, one can study perturbatively the effective theory below the scale Λ , but higher-dimensional operators suppressed by that scale are likely to be generated by physics at scales above Λ . Working within this effective theory, the loop corrections to KK masses are logarithmically dependent on Λ . For fermions, the leading 1-loop corrections may be obtained by substituting

$$\frac{j}{R} \rightarrow \frac{j}{R} \left[\sum_{i=1}^3 9C_i \alpha_i - (3 - 2C_2) \frac{\lambda_f^2}{4\pi} \right] \frac{1}{4\pi} \ln(\Lambda R/j) \quad (42)$$

in the tree-level mass formulas. Here α_i for $i = 1, 2, 3$ are the $U(1)_Y$, $SU(2)_W$ and $SU(3)_c$ coupling constants, $C_1 = y_f^2$ for fermions of hypercharge y_f (using the normalization where y_f is the electric charge for weak singlets), $C_2 = 3/4$ for $SU(2)_W$ doublets and 0 for singlets, $C_3 = 4/3$ for quarks and 0 for leptons, and λ_f is the fermion Yukawa coupling. Note that the $SU(2)_W$ loop corrections to the top KK masses split the diagonal elements in Eq. (40), and therefore $t_1^{(j)}$ and $t_2^{(j)}$ (which are the physical Dirac fermions representing the top modes at each level) end up with different masses.

For gauge bosons, the leading loop corrections are given by the following substitution in the tree-level masses:

$$\frac{j^2}{R^2} \rightarrow \frac{j^2}{R^2} C'_i \frac{\alpha_i}{4\pi} \ln(\Lambda R/j), \quad (43)$$

where C'_i equals to 23 for the gluon, to 15 for the $SU(2)_W$ bosons, and to $-1/3$ for the hypercharge boson B_μ . The different loop contributions to the W_μ^3 and B_μ KK masses have a dramatic effect on their mixing due to electroweak symmetry breaking: the mixing vanishes in the limit where $1/R \gg v$. For this reason, the KK modes of the photon are labelled in the literature by either $\gamma^{(j)}$ or $B^{(j)}$, and the Z modes are labelled by either $Z^{(j)}$ or $W^{3(j)}$.

Exercise 4.5: Compute the masses of $\gamma^{(j)}$ and $Z^{(j)}$ as a function of $1/R$, v , and Λ .

Exercise 4.6: Write down a Higgs mass term localized at both $x^4 = 0$ and $x^4 = L$, and determine its effects on the KK Higgs masses.

It is often assumed that higher-dimensional operators (in particular, kinetic terms localized on the boundary) and localized Higgs mass terms may be ignored at the scale Λ [42]. The lightest KK particle is then the first KK mode of the photon, and the heaviest particles at each level are the KK modes of the gluon and quarks. The mass spectrum of all level-1 particles is shown in Fig. 10. If the unspecified UV completion of this low-energy effective theory gives rise to operators localized at the ends of the interval, then the KK spectrum may change; this possibility is not considered in what follows.

Momentum conservation along the extra dimension is broken by the boundary conditions, but a remnant of it is left intact. This is reflected in a selection rule for the KK-numbers of the particles participating in any interaction. A vertex with particles of KK numbers $j_1, \dots, j_p$ exists at tree level only if $j_1 \pm \dots \pm j_p = 0$ for a certain choice of the $\pm$ signs. This selection rule has important phenomenological implications. First, it is not possible to produce only one KK 1-mode at colliders. Second, tree-level exchange of KK modes does not contribute to currently measurable quantities. Therefore, the corrections to electroweak observables are loop suppressed, and the limit on $1/R$ from electroweak measurements is rather weak, of the order of the electroweak scale [38].

To derive the interactions among KK modes, one should start with the 5D Lagrangian, replace the 5D fields by their KK decomposition, integrate over x^4 , and replace the 5D coupling parameters by their 4D counterparts which are identified by inspecting the interactions among 0-modes. The interactions

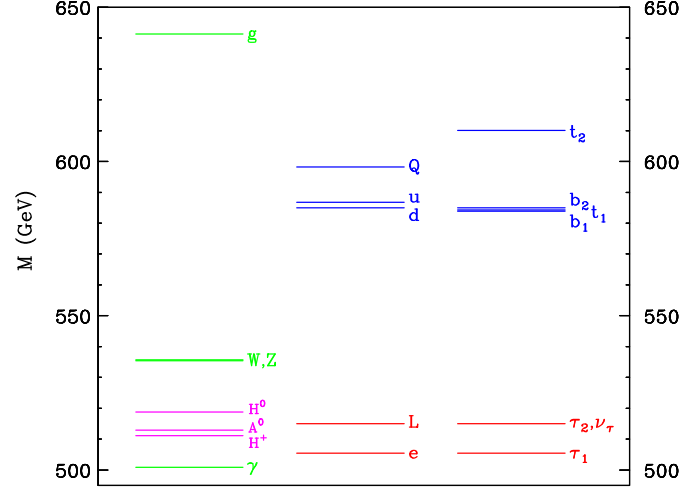


Fig. 10: Mass spectrum of level-1 KK modes for a compactification scale of $1/R = 500$ GeV and $\Lambda R = 20$, from Ref. [42].

$$\begin{aligned}
 & \begin{array}{c} \text{Diagram 1: } G_\mu^{(j)a} \text{ (wavy line) splits into } Q^{(j)} \text{ and } (u, d). \\ \text{Vertex factor: } = i g_s \gamma^\mu T^a P_L \end{array} & \begin{array}{c} \text{Diagram 2: } G_\mu^{(j)a} \text{ (wavy line) splits into } \mathcal{U}^{(j)}; \mathcal{D}^{(j)} \text{ and } u; d. \\ \text{Vertex factor: } = i g_s \gamma^\mu T^a P_R \end{array} \\
 & \begin{array}{c} \text{Diagram 3: } G_\nu^b \text{ (wavy line) and } G_\mu^{(j)a} \text{ (wavy line) meet at a vertex, with } G_\rho^{(j)c} \text{ (wavy line) and } q \text{ (wavy line) as external lines.} \\ \text{Vertex factor: } = g_s f^{abc} [(k-p)_\lambda g_{\mu\nu} + (p-q)_\mu g_{\nu\rho} + (q-k)_\nu g_{\mu\rho}] \end{array} \\
 & \begin{array}{c} \text{Diagram 4: } G_\mu^a \text{ (wavy line) and } G_\nu^b \text{ (wavy line) meet at a vertex, with } G_\rho^{(j)c} \text{ (wavy line) and } G_\sigma^{(j)d} \text{ (wavy line) as external lines.} \\ \text{Vertex factor: } = -i g_s^2 [f^{abe} f^{cde} (g^{\mu\rho} g^{\nu\sigma} - g^{\mu\sigma} g^{\nu\rho}) + f^{ace} f^{bde} (g^{\mu\nu} g^{\rho\sigma} - g^{\mu\sigma} g^{\nu\rho}) \\ + f^{ade} f^{bce} (g^{\mu\nu} g^{\rho\sigma} - g^{\mu\rho} g^{\nu\sigma})] \end{array}
 \end{aligned}$$

Fig. 11: Feynman rules for vertices involving gluon KK modes ($G_\mu^{(j)a}$) and standard model particles. The couplings to quarks are chiral. G_μ^a is the massless gluon, $Q^{(j)}$ are the $SU(2)_W$ -doublet KK quarks, and $\mathcal{U}^{(j)}, \mathcal{D}^{(j)}$ are the $SU(2)_W$ -singlet KK quarks.

of the KK modes with the massless gluon are fixed by the $SU(3)_c$ gauge invariance. For example, the coupling of a gluon to a pair of KK quarks is the same as for any quark pair, and the quartic coupling of two gluon 0-modes and two level- j KK gluons is the same as the quartic gluon coupling of usual QCD. However, the interactions of gluon KK modes with quarks are peculiar. As opposed to the 4D QCD interactions, which are vectorlike, the interactions of higher gluon modes distinguish between left- and right-handed quarks. This follows from the fact that the couplings of a spin-1 particle to fermions do not change the chirality, and thus a right-handed standard model quark couples only to the right-handed components of its KK modes (which are $SU(2)_W$ singlets, $\mathcal{U}^{(j)}$ or $\mathcal{D}^{(j)}$), while a left-handed standard

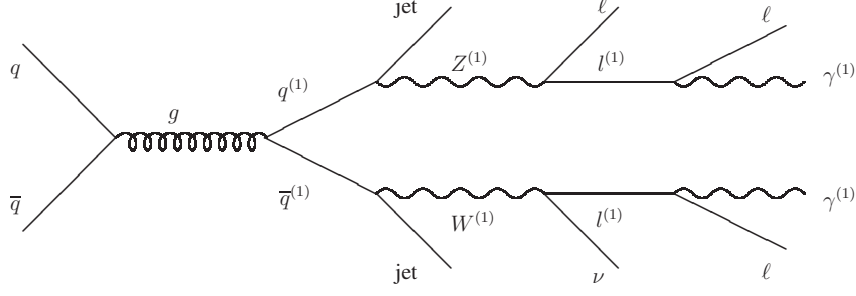


Fig. 12: $3\ell + \cancel{E}_T$ signal from UED. $\gamma^{(1)}$ is the dark matter candidate which escapes the detector.

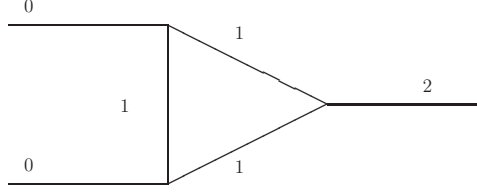


Fig. 13: One-loop induced coupling of a 2-mode to two zero-modes.

model quark couples only to the left-handed components of its KK modes ($Q^{(j)}$). These interactions are contained in the following terms of the 4D Lagrangian:

$$g_s G_\mu^{(j)a} \left(\overline{Q}_L^{(j)} \gamma^\mu T^a (u_L, d_L)^\top + \overline{U}_R^{(j)} \gamma^\mu T^a u_R + \overline{D}_R^{(j)} \gamma^\mu T^a d_R + \text{H.c.} \right), \quad (44)$$

where g_s is the QCD gauge coupling. The Feynman rules describing the interactions of gluon KK modes to 0-modes are shown in Fig. 11.

The 1-modes may be produced in pairs at colliders. At the Tevatron and the LHC, pair production of the colored KK modes has large cross sections [43, 44] as long as $1/R$ is not too large. The colored KK modes suffer cascade decays [45] like the ones shown in Fig. 12. Note that at each vertex the KK-number is conserved, and the $\gamma^{(1)}$ escapes the detector. The signal is $\ell^+ \ell^- \ell^\pm + 2j + \cancel{E}_T$. However, the approximate degeneracy of the KK modes implies that the jets may be relatively soft, and it could be challenging to distinguish them from the background. The leptons are also soft (with energies of a few percent of the compactification scale, as can be seen from the mass differences displayed in Fig. 10), but usually pass some reasonably chosen cuts.

A search for KK modes in events with two muons of same charge, performed by the D0 Collaboration [46], has set a limit of $1/R > 260$ GeV at the 95% CL in the minimal UED model. Using data collected by the CDF Collaboration during the Run I of the Tevatron, Ref. [47] has set a limit of $1/R > 280$ GeV based on the $3\ell + \cancel{E}_T$ signature. Very recently, the ATLAS Collaboration [48] has searched in 1 fb^{-1} of data for the KK modes of UED using final states with jets and $\cancel{E}_T$, and has set a strong limit of $1/R > 600$ GeV. Larger LHC data sets can improve this limit, or alternatively, may lead to a discovery.

If a signal is seen at the LHC, then it is important to differentiate the UED models from alternative explanations, such as superpartner cascade decays [45]. Measuring the spins at the LHC would provide an important discriminant, but such measurements are challenging [44, 50]. A more promising way is to look for second level KK modes. These can be pair produced as the first level modes. However, unlike the first level modes, the second level modes may decay into standard model particles. Such decays occur at one loop, via diagrams having the structure shown in Fig. 13.

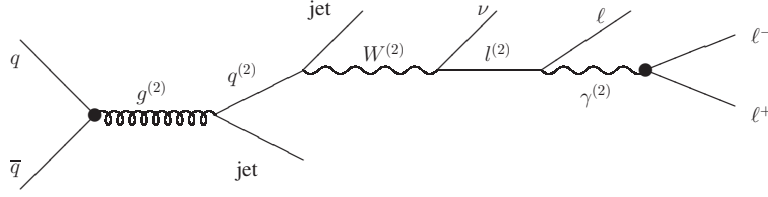


Fig. 14: s -channel production of the level-2 gluon followed by cascade decay, and $\gamma^{(2)}$ decays to e^+e^- and $\mu^+\mu^-$. The $\bullet$ represents an 1-loop effective vertex, as in Fig. 13.

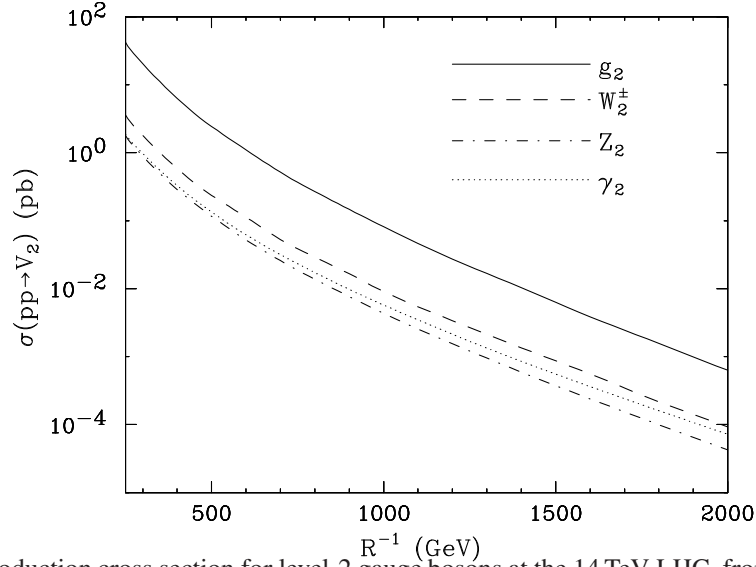


Fig. 15: Production cross section for level-2 gauge bosons at the 14 TeV LHC, from Ref. [50].

This diagram demonstrates that in the presence of loop corrections, the selection rule for KK numbers of the particles interacting at a vertex becomes

$$j_1 \pm \dots \pm j_p = 0 \bmod 2 . \quad (45)$$

This implies the existence of an exact Z_2 symmetry: the KK parity $(-1)^j$ is conserved. Its geometrical interpretation is invariance under reflections with respect to the middle of the $[0, L]$ interval. Given that the lightest particle with j odd is stable, the $\gamma^{(1)}$ is a promising dark matter candidate. For $1/R$ in the 0.5 to 1.5 TeV range the $\gamma^{(1)}$ relic density fits nicely the dark matter density [49]. This whole range of compactification scales will be probed at the LHC [45].

Another consequence of the loop-induced coupling of a 2-mode to two zero-modes is that the 2-mode can be singly produced in the s -channel. The typical signal will be the cascade decay shown in Fig. 14, followed by $\gamma^{(2)}$ decay into hard leptons. The cross section in this channel at the LHC is substantial for compactification scales up to 2 TeV (see Fig. 15).

4.2.3 More dimensions

Gauge theories in more than four spacetime dimensions are nonrenormalizable. This is not a problem as long as there is a range of scales where the higher-dimensional field theory is valid. For gauge couplings of order unity, as in the Standard model, the range of scales is of the order of $(4\pi)^{2/n}$, so that only low values of n are interesting. Furthermore, the low energy observables get corrections from loops with KK modes. The leading corrections are finite in the $n = 1$ case and logarithmically divergent for $n = 2$, while for $n \geq 3$ they depend quadratically or stronger on the cut-off. Therefore, the effects of the unknown physics above the cut-off scale can be kept under control only for $n = 1$ and $n = 2$.

The case of two universal extra dimensions has been analyzed less extensively compared to $n = 1$ UED. The general features of the standard model in $n = 2$ UED are presented in Ref. [52]. The hadron collider phenomenology of (1,0) modes, which are the lightest KK particles, has been explored in Ref. [53]. Cascade decays of spinless adjoints proceed through tree-level 3-body decays involving leptons as well as one-loop 2-body decays involving photons. As a result, spectacular events with as many as six charged leptons, or one photon plus four charged leptons are expected to be observed at the LHC. Unusual events with relatively large branching fractions include three leptons of same charge plus one lepton of opposite charge, or one photon plus two leptons of same charge.

The cascade decays of the (1,1) modes [52], which are heavier than the (1,0) modes by a factor of $\sqrt{2}$, generate a series of closely-spaced narrow resonances in the $t\bar{t}$ invariant mass distribution.

5 New particles, one or two at a time

There may exist various types of new particles, including vectorlike quarks or leptons, new gauge bosons (G' , Z' , W' , γ'), new scalars (neutral or charged ‘Higgs’ bosons; diquarks, leptoquarks; color-octets transforming under $SU(2)_W$ as singlets, doublets, or triplets), or even more exotic ones. For illustration, in Section 5.1 we discuss spin-1 electrically-neutral bosons, usually known as Z' bosons, in a rather model-independent way. A similar approach may be applied to any other new particle, assuming that it can be produced and studied somewhat in isolation of other new particles (unlike the superpartners of KK modes of UED which are produced in cascade decays with many other new particles).

5.1 Z' boson

The couplings of a Z' boson to the first-generation fermions are given by

$$Z'_\mu (g_u^L \bar{u}_L \gamma^\mu u_L + g_d^L \bar{d}_L \gamma^\mu d_L + g_u^R \bar{u}_R \gamma^\mu u_R + g_d^R \bar{d}_R \gamma^\mu d_R + g_\nu^L \bar{\nu}_L \gamma^\mu \nu_L + g_e^L \bar{e}_L \gamma^\mu e_L + g_e^R \bar{e}_R \gamma^\mu e_R), \quad (1)$$

where u, d, ν and e are the quark and lepton fields in the mass eigenstate basis, and the coefficients $g_u^L, g_d^L, g_u^R, g_d^R, g_\nu^L, g_e^L, g_e^R$ are real dimensionless parameters. If the Z' couplings to quarks and leptons are generation-independent, then these seven parameters describe the couplings of the Z' boson to all Standard Model fermions. More generally, however, the Z' couplings to fermions are generation-dependent, in which case Eq. (1) may be written with generation indices $i, j = 1, 2, 3$ labeling the quark and lepton fields, and with the seven coefficients promoted to 3×3 Hermitian matrices (e.g., $g_{eij}^L \bar{e}_L^i \gamma^\mu e_L^j$, where e_L^2 is the left-handed muon, etc.).

These parameters describing the Z' boson interactions with quarks and leptons are subject to some theoretical constraints. Quantum field theories that include a heavy spin-1 particle are well behaved at high energies only if that particle is a gauge boson associated with a spontaneously broken gauge symmetry. Quantum effects preserve the gauge symmetry only if the couplings of the gauge boson to fermions satisfy anomaly cancellation conditions. Furthermore, the fermion charges under the new gauge symmetry are constrained by the requirement that the quarks and leptons get masses from gauge-invariant interactions with Higgs doublets or whatever else breaks the electroweak symmetry.

The relation between the couplings displayed in Eq. (1) and the gauge charges z_{fi}^L and z_{fi}^R of the fermions $f = u, d, \nu, e$ involves the unitary 3×3 matrices V_f^L and V_f^R that transform the gauge eigenstate fermions f_L^i and f_R^i , respectively, into the mass eigenstates. In addition, the Z' couplings are modified if the new gauge boson in the gauge eigenstate basis ($\tilde{Z}'_\mu$) has a kinetic mixing $(-\chi/2)B^{\mu\nu}\tilde{Z}'_{\mu\nu}$ with the hypercharge gauge boson B^μ (due to a dimension-4 or 6 operator, depending on whether the new gauge symmetry is Abelian or not), or a mass mixing $\delta M^2 \tilde{Z}^\mu \tilde{Z}'_\mu$ with the linear combination ($\tilde{Z}_\mu$) of neutral bosons which has same couplings as the Standard Model Z^0 . Both the kinetic and mass mixings shift the

mass and couplings of the Z boson, such that the electroweak measurements impose upper limits on χ and $\delta M^2/(M_{Z'}^2 - M_Z^2)$ of the order of 10^{-3} [2]. Keeping only linear terms in these two small quantities, the couplings of the mass-eigenstate Z' boson are given by

$$g_{fij}^L = g_z V_{fii'}^L z_{fi'}^L (V_f^L)_{i'j}^\dagger + \frac{e}{\alpha_W} \left(\frac{s_W \chi M_{Z'}^2 + \delta M^2}{2s_W (M_{Z'}^2 - M_Z^2)} \sigma_f^3 - \epsilon Q_f \right), \quad (2)$$

$$g_{fij}^R = g_z V_{fii'}^R z_{fi'}^R (V_f^R)_{i'j}^\dagger - \frac{e}{c_W} \epsilon Q_f, \quad (3)$$

where g_z is the new gauge coupling, Q_f is the electric charge of f , e is the electromagnetic gauge coupling, s_W and c_W are the sine and cosine of the weak mixing angle, $\sigma_f^3 = +1$ for $f = u, \nu$ and $\sigma_f^3 = -1$ for $f = d, e$, and

$$\epsilon = \frac{\chi (M_{Z'}^2 - c_W^2 M_Z^2) + s_W \delta M^2}{M_{Z'}^2 - M_Z^2}. \quad (4)$$

A simple origin of a Z' boson is a new $U(1)'$ gauge symmetry. In that case, the matricial equalities $z_u^L = z_d^L$ and $z_\nu^L = z_e^L$ are required by the $SU(2)_W$ gauge symmetry. Given that the $U(1)'$ interaction is not asymptotically free, the theory may be well-behaved at high energies (for example, by embedding $U(1)'$ in a non-Abelian gauge group) only if the Z' couplings are commensurate numbers, *i.e.* any ratio of couplings is a rational number. Satisfying the anomaly cancellation conditions (which include an equation cubic in charges) with rational numbers is highly nontrivial, and in general new fermions charged under $U(1)'$ are necessary.

Consider first the case where the couplings are generation-independent (the V_f matrices then disappear from Eq. (3), so that there are five commensurate couplings: $g_q^L, g_u^R, g_d^R, g_l^L, g_e^R$. Four sets of charges are displayed in Table Charge, each of them spanned by one free parameter, x [54]. The first set, labelled $B - xL$, has charges proportional to the baryon number minus x times the lepton number. These charges allow all Standard Model Yukawa couplings to a Higgs doublet which is neutral under $U(1)_{B-xL}$, so that there is no tree-level $\tilde{Z} - \tilde{Z}'$ mixing. For $x = 1$ one recovers the $U(1)_{B-L}$ group, which is non-anomalous in the presence of one “right-handed neutrino” (a chiral fermion that is a singlet under the Standard Model gauge group) per generation. For $x \neq 1$, it is necessary to include some fermions that are vector-like (*i.e.* their mass terms are gauge invariant) with respect to the electroweak gauge group and chiral with respect to $U(1)_{B-xL}$. In the particular cases $x = 0$ or $x \gg 1$ the Z' is leptophobic or quark-phobic, respectively.

The second set, $U(1)_{10+x5}$, has charges that commute with the representations of the $SU(5)$ grand unified group. Here x is related to the mixing angle between the two $U(1)$ bosons encountered in the $E_6 \rightarrow SU(5) \times U(1) \times U(1)$ symmetry breaking patterns of grand unified theories. This set leads to $\tilde{Z} - \tilde{Z}'$ mass mixing at tree level, such that for a Z' mass close to the electroweak scale, the measurements at the Z -pole require some fine tuning between the charges and VEVs of the two Higgs doublets. Vector-like fermions charged under the electroweak gauge group and also carrying color are required (except for $x = -3$) to make this set anomaly free. The particular cases $x = -3, 1, -1/2$ are usually labelled $U(1)_\chi$, $U(1)_\psi$, and $U(1)_\eta$, respectively. Under the third set, $U(1)_{d-xu}$, the weak-doublet quarks are neutral, and the ratio of u_R and d_R charges is $-x$. For $x = 1$ this is the “right-handed” group $U(1)_R$. For $x = 0$, the charges are those of the E_6 -inspired $U(1)_I$ group, which requires new quarks and leptons. Other generation-independent sets of $U(1)'$ charges are given in Ref. [55].

In the absence of new fermions charged under the SM gauge group, the most general generation-independent charge assignment is $U(1)_{q+xu}$, which is a linear combination of hypercharge and $B - L$. Many other anomaly-free solutions exist if generation-dependent charges are allowed. Table ChargeGen shows such solutions that depend on two free parameters, x and y , with generation dependence only in the lepton sector, which includes one right-handed neutrino per generation. The charged-lepton masses may be generated by Yukawa couplings to a single Higgs doublet. These are forced to be flavor diagonal

by the generation-dependent $U(1)'$ charges, so that there are no tree-level flavor-changing neutral current (FCNC) processes involving electrically-charged leptons. For the “leptocratic” set, neutrino masses are induced by operators of high dimensionality that may explain their smallness [6].

If the $SU(2)_W$ -doublet quarks have generation-dependent $U(1)'$ charges, then the mass eigenstate quarks have flavor off-diagonal couplings to the Z' boson (see Eq. (1), and note that $V_u^L (V_d^L)^\dagger$ is the CKM matrix). These are severely constrained by measurements of FCNC processes, which in this case are mediated at tree-level by Z' boson exchange [57]. The constraints are relaxed if the first and second generation charges are the same, although they are increasingly tightened by the measurements of B meson properties. If only the $SU(2)_W$ -singlet quarks have generation-dependent $U(1)'$ charges, there is more freedom in adjusting the flavor off-diagonal couplings because the $V_{u,d}^R$ matrices are not observable in the Standard Model.

The anomaly cancellation conditions for $U(1)'$ could be relaxed only if at scales above $4\pi M_{Z'}/g_z$ there is an axion which has certain dimension-5 couplings to the gauge bosons. However, such a scenario violates unitarity unless the quantum field theory description breaks down at a scale near $M_{Z'}$ [58].

Z' bosons may also arise from larger gauge groups. These may be orthogonal to the electroweak group, as in $SU(2)_W \times U(1)_Y \times SU(2)'$, or may embed the electroweak group, as in $SU(3)_W \times U(1)$ [59]. If the larger group is spontaneously broken down to $SU(2)_W \times U(1)_Y \times U(1)'$ at a scale $v_\star \gg M_{Z'}/g_z$, then the above discussion applies up to corrections of order $M_{Z'}^2/(g_z v_\star)^2$. For $v_\star \sim M_{Z'}/g_z$, additional gauge bosons have masses comparable to $M_{Z'}$, including at least a W' boson [59]. If the larger gauge group breaks together with the electroweak symmetry directly to the electromagnetic $U(1)_{\text{em}}$, then the left-handed fermion charges are no longer correlated ($z_u^L \neq z_d^L$, $z_\nu^L \neq z_e^L$) and a $Z'W^+W^-$ coupling is induced.

The couplings shown in Eq. (1) lead to Z' production in the s -channel at colliders, and to Z' decays to fermion pairs. The decay into e^+e^- has a width

$$\Gamma(Z' \rightarrow e^+e^-) \simeq \left[(g_e^L)^2 + (g_e^R)^2 \right] \frac{M_{Z'}}{24\pi}. \quad (5)$$

The decay width into $q\bar{q}$ is similar, except for an additional color factor of 3, QCD radiative corrections. Fermion mass corrections are important for the $t\bar{t}$ final state. Thus, one may compute the Z' branching fractions in terms of the couplings of Eq. (1), but other decay channels, such as WW or a pair of new particles, could also have large widths that should be added to the total decay width.

Z' bosons with couplings to quarks (see Eq. (1)) may be produced at hadron colliders in the s channel, and would show up as resonances in the invariant mass distribution of the decay products. The cross section for producing a Z' boson at the LHC which then decays to some $f\bar{f}$ final state takes the form

$$\sigma(pp \rightarrow Z'X \rightarrow f\bar{f}X) \simeq \frac{\pi}{48s} \sum_q c_q^f w_q(s, M_{Z'}^2) \quad (6)$$

for flavor-diagonal couplings to quarks. Here we have neglected the interference with the Standard Model contribution to $f\bar{f}$ production, which is a good approximation for a narrow Z' resonance. The coefficients

$$c_q^f = \left[(g_q^L)^2 + (g_q^R)^2 \right] B(Z' \rightarrow f\bar{f}) \quad (7)$$

contain all the dependence on the Z' couplings, while the functions w_q include all the information about parton distributions and QCD corrections [54,55]. This factorization holds exactly to NLO, and the deviations from it induced at NNLO are very small. Note that the w_u and w_d functions are substantially larger than the w_q functions for the other quarks. Eq. (6) also applies to the Tevatron, except for changing the pp initial state to $p\bar{p}$, which implies that the $w_q(s, M_{Z'}^2)$ functions are replaced by some other functions $\bar{w}_q((1.96 \text{ TeV})^2, M_{Z'}^2)$.

It is common to present results of Z' searches as limits on the cross section versus $M_{Z'}$. An alternative is to plot exclusion curves for fixed $M_{Z'}$ values in the $c_u^f - c_d^f$ planes, allowing a simple derivation of the mass limit within any Z' model (see e.g. Ref. [62]).

The observation of a dilepton resonance at the LHC would determine the Z' mass and width. A measurement of the total cross section would define a band in the $c_u^\ell - c_d^\ell$ plane. Angular distributions can be used to measure several combinations of Z' parameters (an example of how angular distributions improve the Tevatron sensitivity is given in Ref. [63]). Even though the original quark direction in a pp collider is unknown, the leptonic forward-backward asymmetry A_{FB}^ℓ can be extracted from the kinematics of the dilepton system, and is sensitive to parity-violating couplings. A fit to the Z' rapidity distribution can distinguish between the couplings to up and down quarks. These measurements, combined with off-peak observables, have the potential to differentiate among various Z' models [64]. For example, the couplings of a Z' boson with mass below 1.5 TeV can be well determined with 100 fb^{-1} of data at $\sqrt{s} = 14 \text{ TeV}$. With this amount of data, the spin of the Z' boson may be determined for $M_{Z'} \leq 3 \text{ TeV}$ [65], and the expected sensitivity extends to $M_{Z'} \sim 5 - 6 \text{ TeV}$ for many models [66].

The Z' decays into e^+e^- and $\mu^+\mu^-$ are useful due to relatively good mass resolution and large acceptance. The Z' decays into $e\mu$ and $\tau^+\tau^-$, along with $t\bar{t}$, $b\bar{b}$ and jj which suffer from larger backgrounds, are also important as they probe various combinations of Z' couplings to fermions. The $pp \rightarrow Z'X \rightarrow W^+W^-X$ process may also be explored at the LHC, and is important for disentangling the origin of electroweak symmetry breaking. The Z' boson may be produced in this process through its couplings to either quarks [67] or W bosons [68].

6 Conclusions

So far the only robust evidence for physics beyond the standard model (defined as in Section 1) is that various astronomical and cosmological measurements require dark matter. This implies the existence of at least one new particle and some new symmetry that keeps it stable. It remains to be seen whether that particle has sufficiently strong interactions with the SM particles to be produced at the LHC or to be observed in direct detection experiments.

The Higgs sector is the least explored part of the standard model. There are hints of a Higgs boson at a mass of about 125 GeV from the ATLAS and CMS experiments, as well as from the Tevatron. Many precise measurements are required before assessing whether that is indeed the SM Higgs boson. If it will turn out to be a Higgs boson with non-standard interactions, then an important window to new physics will be opened.

Besides theories that address the hierarchy problem and include a dark matter candidate (such as the MSSM), there are many possibilities for new particles. Among those there are spin-0 bosons not necessarily associated with extended Higgs sectors (such as color-octets and lepto-quarks), spin-1/2 fermions (such as vectorlike quarks or leptons), spin-1 bosons (e.g., Z' and W' bosons arising from extensions of the electroweak symmetry, or colorons arising from extensions of QCD), and higher spin particles. Each of these particles would lead to a variety of signals at colliders, so that the LHC experiments should search for these in as many final states as possible.

It should also be emphasized that physics at the TeV scale could involve new strong interactions, and that some if not all SM particles may be composite. Given the current poor understanding of strongly coupled field theory, the measurement of various differential cross sections at the LHC could uncover truly unexpected phenomena.

Acknowledgments. I would like to thank the organizers of ESHEP for creating a pleasant and stimulating environment at Cheile Gradistei. I am grateful to the graduate student participants for asking many insightful questions; I trust that they will use data to answer the toughest questions.

References

- [1] S. Weinberg, Phys. Rev. Lett. **19** (1967) 1264.
- [2] K. Nakamura et al. (Particle Data Group), J. Phys. **G37** (2010) 075021 and 2011 partial update for the 2012 edition, <http://pdg.lbl.gov/>.
- [3] E. Dudas, *Field Theory and the Standard Model*, in these proceedings.
- [4] J. F. Donoghue, Phys. Rev. D **50** (1994) 3874 [gr-qc/9405057]; *Introduction to the Effective Field Theory Description of Gravity*, arXiv:gr-qc/9512024.
- [5] B. Kayser, *Neutrino Oscillation Physics*, in these proceedings.
- [6] M. -C. Chen, A. de Gouvea and B. A. Dobrescu, Phys. Rev. **D75** (2007) 055009 [hep-ph/0612017].
- [7] D. B. Kaplan and H. Georgi, Phys. Lett. **B136** (1984) 183. M. J. Dugan, H. Georgi and D. B. Kaplan, Nucl. Phys. **B254** (1985) 299.
- [8] For a set of lectures, see R. S. Chivukula, *Models of electroweak symmetry breaking: Course*, hep-ph/9803219.
- [9] R. Contino, Y. Nomura and A. Pomarol, Nucl. Phys. **B671** (2003) 148 [hep-ph/0306259].
G. F. Giudice, C. Grojean, A. Pomarol and R. Rattazzi, JHEP **0706** (2007) 045 [hep-ph/0703164].
- [10] R. S. Chivukula, *et al*, Phys. Rev. D **59** (1999) 075003 [hep-ph/9809470]. B. A. Dobrescu and C. T. Hill, Phys. Rev. Lett. **81** (1998) 2634 [hep-ph/9712319].
- [11] C. T. Hill, Phys. Lett. **B266** (1991) 419.
- [12] For a review, see C. T. Hill and E. H. Simmons, Phys. Rept. **381** (2003) 235 [Erratum-ibid. **390** (2004) 553] [arXiv:hep-ph/0203079].
- [13] N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, Phys. Lett. **B429** (1998) 263 [arXiv:hep-ph/9803315].
- [14] D. J. Kapner, *et al*, Phys. Rev. Lett. **98** (2007) 021101 [arXiv:hep-ph/0611184].
- [15] T. Han, J. D. Lykken and R. J. Zhang, Phys. Rev. **D59** (1999) 105006 [arXiv:hep-ph/9811350];
G. F. Giudice, R. Rattazzi and J. D. Wells, Nucl. Phys. **B544** (1999) 3 [arXiv:hep-ph/9811291].
- [16] N. Arkani-Hamed, A. G. Cohen and H. Georgi, Phys. Lett. **B513** (2001) 232 [hep-ph/0105239].
N. Arkani-Hamed, A. G. Cohen, E. Katz and A. E. Nelson, JHEP **0207** (2002) 034 [hep-ph/0206021].
N. Arkani-Hamed, *et al*, JHEP **0208** (2002) 021 [hep-ph/0206020].
- [17] H. -C. Cheng and I. Low, JHEP **0309** (2003) 051 [hep-ph/0308199].
- [18] Z. Chacko, H. -S. Goh and R. Harnik, Phys. Rev. Lett. **96** (2006) 231802 [hep-ph/0506256].
- [19] S. P. Martin, *A Supersymmetry primer*, hep-ph/9709356.
- [20] See, *e.g.*, M. Papucci, J. T. Ruderman and A. Weiler, JHEP **1209** (2012) 035 [arXiv:1110.6926 [hep-ph]].
- [21] H. M. Lee, V. Sanz and M. Trott, JHEP **1205** (2012) 139 [arXiv:1204.0802 [hep-ph]].
- [22] P. J. Fox, A. E. Nelson and N. Weiner, JHEP **0208** (2002) 035 [hep-ph/0206096].
- [23] M. Dine, A. E. Nelson, Y. Nir and Y. Shirman, Phys. Rev. **D53** (1996) 2658 [hep-ph/9507378].
- [24] L. Randall and R. Sundrum, Phys. Rev. Lett. **83** (1999) 3370 [arXiv:hep-ph/9905221].
- [25] M. S. Carena, J. D. Lykken and M. Park, Phys. Rev. **D72** (2005) 084017 [arXiv:hep-ph/0506305].
- [26] W. D. Goldberger and M. B. Wise, Phys. Rev. Lett. **83** (1999) 4922 [arXiv:hep-ph/9907447].
- [27] H. Davoudiasl, J. L. Hewett and T. G. Rizzo, Phys. Rev. Lett. **84** (2000) 2080 [arXiv:hep-ph/9909255].
- [28] ATLAS Collaboration, *Search for high-mass dilepton resonances with 5/fb*, ATLAS-2012-007.
- [29] H. Davoudiasl, J. L. Hewett, T. G. Rizzo, Phys. Rev. **D63** (2001) 075004 [hep-ph/0006041].
- [30] H. Davoudiasl, J. L. Hewett and T. G. Rizzo, Phys. Lett. **B473** (2000) 43 [hep-ph/9911262]; A. Pomarol, Phys. Lett. **B486** (2000) 153 [arXiv:hep-ph/9911294].

- [31] S. Chang, J. Hisano, H. Nakano, N. Okada and M. Yamaguchi, Phys. Rev. **D62** (2000) 084025 [arXiv:hep-ph/9912498]; T. Gherghetta and A. Pomarol, Nucl. Phys. **B586** (2000) 141 [arXiv:hep-ph/0003129].
- [32] T. Gherghetta, *Warped models and holography*, in *Proc. Les Houches summer school on Theoretical Physics, Les Houches, 2005*, arXiv:hep-ph/0601213.
- [33] H. C. Cheng, C. T. Hill, S. Pokorski and J. Wang, Phys. Rev. **D64** (2001) 065007 [arXiv:hep-th/0104179].
- [34] C. Brust, A. Katz, S. Lawrence and R. Sundrum, JHEP **1203** (2012) 103 [arXiv:1110.6670 [hep-ph]].
- [35] ATLAS Collaboration, *Search for supersymmetry in final states with missing transverse momentum, b-jets and one lepton*, ATLAS-CONF-2011-130.
- [36] I. Antoniadis, Phys. Lett. **B246** (1990) 377.
- [37] K. R. Dienes, E. Dudas and T. Gherghetta, Nucl. Phys. **B537** (1999) 47 [arXiv:hep-ph/9806292].
- [38] T. Appelquist, H. C. Cheng and B. A. Dobrescu, Phys. Rev. **D64** (2001) 035002 [arXiv:hep-ph/0012100].
- [39] G. D. Kribs, *TASI 2004 lectures on phenomenology of extra dimensions*, in *Proc. TASI 2004*, arXiv:hep-ph/0605325.
- [40] D. Hooper and S. Profumo, Phys. Rept. **453** (2007) 29 [arXiv:hep-ph/0701197].
- [41] H. Georgi, A. K. Grant and G. Hailu, Phys. Lett. **B506** (2001) 207 [arXiv:hep-ph/0012379].
- [42] H. C. Cheng, K. T. Matchev and M. Schmaltz, Phys. Rev. **D66** (2002) 036005 [hep-ph/0204342].
- [43] C. Macesanu, C. D. McMullen and S. Nandi, Phys. Rev. **D66** (2002) 015009 [hep-ph/0201300].
- [44] J. M. Smillie and B. R. Webber, JHEP **0510** (2005) 069 [arXiv:hep-ph/0507170].
- [45] H. C. Cheng, K. T. Matchev, M. Schmaltz, Phys. Rev. **D66** (2002) 056006 [hep-ph/0205314].
- [46]
- [46] V. M. Abazov *et al.* [D0 Collaboration], Phys. Rev. Lett. **108** (2012) 131802 [arXiv:1112.4092 [hep-ex]].
- [47] C. Lin, *A search for universal extra dimensions in the multi-lepton channel from proton antiproton collisions at $\sqrt{s} = 1.8$ TeV*, CDF Thesis, 2005.
- [48] ATLAS Collaboration, *Search for supersymmetry with jets and missing transverse momentum: Additional model interpretations*, ATLAS-CONF-2011-155.
- [49] G. Servant and T. M. P. Tait, Nucl. Phys. **B650** (2003) 391 [hep-ph/0206071].
- H. C. Cheng, J. L. Feng and K. T. Matchev, Phys. Rev. Lett. **89** (2002) 211301 [arXiv:hep-ph/0207125].
- [50] A. Datta, K. Kong and K. T. Matchev, Phys. Rev. **D72** (2005) 096006 [Erratum-ibid. **D72** (2005) 119901] [arXiv:hep-ph/0509246].
- [51] D. E. Kaplan and T. M. P. Tait, JHEP **0111** (2001) 051 [arXiv:hep-ph/0110126].
- [52] G. Burdman, B. A. Dobrescu and E. Ponton, Phys. Rev. **D74** (2006) 075008 [hep-ph/0601186].
- [53] B. A. Dobrescu, K. Kong, R. Mahbubani, JHEP **0707** (2007) 006 [hep-ph/0703231].
- [54] M.S. Carena *et al*, Phys. Rev. **D70** (2004) 093009 [arXiv:hep-ph/0408098]
- [55] E. Accomando *et al* Phys. Rev. **D83** (2001) 075012 [arXiv:1010.6058 [hep-ph]].
- [56] See, *e.g.*, F. Del Aguila, M. Cvetič, and P. Langacker, Phys. Rev. **D52** (1995) 37 [arXiv:hep-ph/9501390]
- [57] P. Langacker and M. Plumacher, Phys. Rev. **D62** (2000) 013006; [arXiv:hep-ph/0001204] R.S. Chivukula and E.H. Simmons, Phys. Rev. **D66** (2002) 015006 [arXiv:hep-ph/0205064].
- [58] L.E. Ibanez and G.G. Ross, Phys. Lett. **B332** (1994) 100 [arXiv:hep-ph/9403338].
- [59] See the Section on W' searches in Ref. [2].

- [60] M. Bando, T. Kugo, and K. Yamawaki, Phys. Rept. **164** (1988) 217.
- [61] J. Alcaraz *et al* [ALEPH, DELPHI, L3, OPAL Collaborations, LEP Electroweak Working Group], *A combination of preliminary electroweak measurements and constraints on the standard model*, hep-ex/0612034.
- [62] A. Abulencia *et al.* [CDF Collaboration], Phys. Rev. Lett. **95** (2005) 252001 [arXiv:hep-ex/0507104].
- [63] A. Abulencia *et al* [CDF Collaboration], Phys. Rev. Lett. **96** (2006) 211801 [hep-ex/0602045]
- [64] F. Petriello and S. Quackenbush, Phys. Rev. **D77** (2008) 115004 [arXiv:0801.4389 [hep-ph]].
- [65] P. Osland *et al*, Phys. Rev. **D79** (2009) 115021 [arXiv:0904.4857 [hep-ph]].
- [66] G. L. Bayatian *et al* [CMS Collaboration], J. Phys. **G34** (2007) 995.
- [67] A. Alves *et al*, Phys. Rev. **D80** (2009) 073011 [arXiv:0907.2915 [hep-ph]].
- [68] H. J. He *et al.*, Phys. Rev. **D78** (2008) 031701 [arXiv:0708.2588 [hep-ph]].

Cosmology

V.A. Rubakov

Institute for Nuclear Research of the Russian Academy of Sciences, Moscow, Russia

Abstract

In these lectures we first concentrate on the cosmological problems which, hopefully, have to do with the new physics to be probed at the LHC: the nature and origin of dark matter and generation of matter-antimatter asymmetry. We give several examples showing the LHC cosmological potential. These are WIMPs as cold dark matter, gravitinos as warm dark matter, and electroweak baryogenesis as a mechanism for generating matter-antimatter asymmetry. In the remaining part of the lectures we discuss the cosmological perturbations as a tool for studying the epoch preceding the conventional hot stage of the cosmological evolution.

1 Introduction

The more we learn about our Universe, the better we understand that it is full of mysteries. These fall into three broad classes. One major mystery is dark energy, which deserves a separate class. We briefly discuss dark energy in Section 5, although, honestly speaking, we do not have much to say about it. The second class most likely has to do with the early hot epoch of the cosmological evolution, and the third one with an even earlier stage which preceded the hot epoch. Along with dark energy, we encounter mysteries of the second class when studying the present composition of the Universe. It hosts matter but not antimatter, and after 40 years after it was understood that this is a problem, we do not have an established theory explaining this asymmetry. The Universe hosts dark matter, and we do not know what it is made of. In this context, one of the key players is the Large Hadron Collider. Optimistically, the LHC experiments may discover dark matter particles and their companions, and establish physics behind the matter-antimatter asymmetry. Otherwise they will rule out some very plausible scenarios; this will also have profound impact on our understanding of the early Universe. Let us mention also exotic hypotheses on physics beyond the Standard Model, like TeV scale gravity; their support by the LHC will have an effect on the early cosmology, which is hard to overestimate.

In the first part of these lectures we concentrate on a few examples showing the LHC cosmological potential. Before coming to that, we briefly introduce the basic notions of cosmology that are useful for our main discussion. We then turn to dark matter, and present the WIMP scenario for cold dark matter, which is currently the most popular one — for good reason. We also consider light gravitino scenario for warm dark matter. Both are probed by the LHC, as they require rather particular new physics in the LHC energy range. We then discuss electroweak baryogenesis — a mechanism for the generation of matter-antimatter asymmetry that may have operated at temperature of order 100 GeV in the early Universe. This mechanism also needs new physics at energies 100 – 300 GeV, so it will be confirmed or ruled out by the LHC.

The third class of mysteries is related to cosmological perturbations, i.e., inhomogeneities in energy density and associated gravitational potentials and, possibly, relic gravity waves. As we explain in the second part of these lectures, the observed properties of density perturbations show that they were generated at some epoch that preceded the hot stage of the cosmological evolution. Obviously, the very fact that we are confident about the existence of such an epoch is a fundamental result of theoretical and observational cosmology. The most plausible hypothesis on that epoch is cosmological inflation, though the observational support of this hypothesis is presently not particularly strong, and alternative scenarios have not been ruled out. We will briefly discuss the potential of future cosmological observations in discriminating between different options.

These lectures are meant to be self-contained, but we necessarily omit numerous details, while trying to make clear the basic ideas and results. More complete accounts of particle physics aspects of cosmology may be found in reviews [1]. Dark matter, including various hypotheses about its particles, is reviewed in Ref. [2]. Electroweak baryogenesis is discussed in detail in reviews [3]. For reviews on dark energy, see, e.g., Ref. [4]. The theory and observations of cosmological perturbations are presented in Ref. [5].

2 Homogeneous isotropic Universe

2.1 Friedmann–Lemaître–Robertson–Walker metric

Two basic facts about our visible Universe are that it is *homogeneous and isotropic* at large spatial scales, and that it *expands*.

There are three types of homogeneous and isotropic three-dimensional spaces. These are¹ three-sphere, flat (Euclidean) space and three-hyperboloid. Accordingly, one speaks about closed, flat and open Universe; in the latter two cases the spatial size of the Universe is infinite, whereas in the former the Universe is compact.

The homogeneity and isotropy of the Universe mean that its hypersurfaces of constant time are either three-spheres or Euclidean spaces or three-hyperboloids. The distances between points may (and, indeed, do) depend on time, i.e., the interval has the form

$$ds^2 = dt^2 - a^2(t)dx^2, \quad (1)$$

where dx^2 is the distance on unit three-sphere/Euclidean space/hyperboloid. The metric (1) is usually called Friedmann–Lemaître–Robertson–Walker (FLRW) metric, and $a(t)$ is called the scale factor. In our Universe $\dot{a} \equiv \frac{da}{dt} > 0$, which means that the distance between points of fixed spatial coordinates $\mathbf{x}$ grows, $dl^2 = a^2(t)dx^2$. The space stretches out; the Universe expands.

The coordinates $\mathbf{x}$ are often called comoving coordinates. It is straightforward to check that $\mathbf{x} = \text{const}$ is a time-like geodesic, so a galaxy put at a certain $\mathbf{x}$ at zero velocity will stay at the same $\mathbf{x}$. Furthermore, as the Universe expands, non-relativistic objects lose their velocities $\dot{\mathbf{x}}$, i.e., they get frozen in the comoving coordinate frame.

Observational data set strong constraints on the spatial curvature of the Universe. They tell that to a very good approximation our Universe is spatially flat, i.e., our *3-dimensional space is Euclidean*. In what follows dx^2 is simply the line interval in the Euclidean 3-dimensional space.

2.2 Redshift

Like the distances between free particles in the expanding Universe, the photon wavelength increases too. We will always label the present values of time-dependent quantities by subscript 0: the present wavelength of a photon is thus denoted by λ_0 , the present time is t_0 , the present value of the scale factor is $a_0 \equiv a(t_0)$, etc. If a photon was emitted at some moment of time t in the past, and its wavelength at the moment of emission was λ_e , then we receive today a photon whose physical wavelength is longer,

$$\frac{\lambda_0}{\lambda_e} = \frac{a_0}{a(t)} \equiv 1 + z.$$

Here we introduced the redshift z . The redshift of an object is directly measurable. λ_e is fixed by physics of the source, say, it is the wavelength of a photon emitted by an excited hydrogen atom. So,

¹Strictly speaking, this statement is valid only locally: in principle, homogeneous and isotropic Universe may have complex global properties. As an example, spatially flat Universe may have topology of three-torus. There is some discussion of such a possibility in literature, and fairly strong limits have been obtained by the analyses of cosmic microwave background radiation [6].

one identifies a series of emission or absorption lines, thus determining λ_e , and measures their actual wavelengths λ_0 . These spectroscopic measurements give accurate values of z even for distant sources. On the other hand, the redshift is related to the time of emission, and hence to the distance to the source.

Let us consider a “nearby” source, for which $z \ll 1$. This corresponds to relatively small $(t_0 - t)$. Expanding $a(t)$, one writes

$$a(t) = a_0 - \dot{a}(t_0)(t_0 - t) . \quad (2)$$

To the leading order in z , the difference between the present time and the emission time is equal to the distance to the source r (the speed of light is set equal to 1). Let us define the Hubble parameter

$$H(t) = \frac{\dot{a}(t)}{a(t)}$$

and denote its present value by H_0 . Then Eq. (2) takes the form $a(t) = a_0(1 - H_0 r)$, and we get for the redshift, again to the leading non-trivial order in z ,

$$1 + z = \frac{1}{1 - H_0 r} = 1 + H_0 r .$$

In this way we obtain the Hubble law,

$$z = H_0 r , \quad z \ll 1 . \quad (3)$$

Traditionally, one tends to interpret the expansion of the Universe as runaway of galaxies from each other, and redshift as the Doppler effect. Then at small z one writes $z = v$, where v is the radial velocity of the source with respect to the Earth, so H_0 is traditionally measured in units “velocity per distance”. Observational data give [8]

$$H_0 = [71.0 \pm 2.5] \frac{\text{km/s}}{\text{Mpc}} \approx (14 \cdot 10^9 \text{ yrs})^{-1} , \quad (4)$$

where $1 \text{ Mpc} = 3 \cdot 10^6 \text{ light yrs} = 3 \cdot 10^{24} \text{ cm}$ is the distance measure often used in cosmology. Traditionally, the present value of the Hubble parameter is written as

$$H_0 = h \cdot 100 \frac{\text{km}}{\text{s} \cdot \text{Mpc}} . \quad (5)$$

Thus $h \approx 0.71$. We will use this value in further estimates.

Let us point out that the interpretation of redshift in terms of the Doppler shift is actually not adequate, especially for large enough z . In fact, there is no need in this interpretation at all: the “radial velocity” enters neither theory nor observations, so this notion may be safely dropped. Physically meaningful quantity is redshift z itself.

A final comment is that H_0^{-1} has dimension of time, or length, as indicated in Eq. (4). Clearly, this quantity sets the cosmological scales of time and distance at the present epoch.

2.3 Hot Universe

Our Universe is filled with cosmic microwave background (CMB). Cosmic microwave background as observed today consists of photons with excellent black-body spectrum of temperature

$$T_0 = 2.726 \pm 0.001 \text{ K} . \quad (6)$$

The spectrum has been precisely measured by various instruments and does not show any deviation from the Planck spectrum [7].

Thus, the present Universe is “warm”. Earlier Universe was warmer; it cooled down because of the expansion. While the CMB photons freely propagate today, it was not so at early stage. When the Universe was hot, the usual matter (electrons and protons with rather small admixture of light nuclei) was in the plasma phase. At that time photons strongly interacted with electrons due to the Thomson scattering and protons interacted with electrons via Coulomb force, so all these particles were in thermal equilibrium. As the Universe cooled down, electrons “recombined” with protons into neutral hydrogen atoms, and the Universe became transparent to photons. The temperature scale of recombination is, very crudely speaking, determined by the ionisation energy of hydrogen, which is of order 10 eV. In fact, recombination occurred at lower temperature², $T_{rec} \approx 3000$ K. An important point is that the duration of the period of recombination was considerably shorter than the Hubble time at that epoch; to a reasonable approximation, recombination occurred instantaneously.

The importance of the recombination epoch (more precisely, the epoch of photon last scattering; we will use the term recombination for brevity) is that the CMB photons travel freely after it: the density of hydrogen atoms was so small (about 250 cm^{-3} right after recombination) that the gas was transparent to photons. So, CMB photons give the photographic picture of the Universe at recombination, i.e., at redshift and age

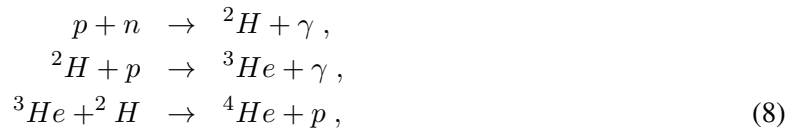
$$z_{rec} = 1090, \quad t_{rec} = 370\,000 \text{ years}. \quad (7)$$

It is worth noting that even though after recombination photons no longer were in thermal equilibrium with anything, the shape of the photon distribution function has not changed, except for overall redshift. Indeed, the thermal distribution function for *ultra-relativistic* particles, the Planck distribution, depends only on the ratio of frequency to temperature, $f_{Planck}(p, T) = f(\omega_p/T)$, $\omega_p = |p|$. As the Universe expands, the photon momentum gets redshifted, $p(t) = p(t_{rec}) \cdot \frac{a(t_{rec})}{a(t)}$, the frequency is redshifted in the same way, but the shape of the spectrum remains Planckian, with redshifted temperature. Hence, the Planckian form of the observed spectrum is no surprise. Generally speaking, this property does not hold for massive particles.

At even earlier times, the temperature of the Universe was even higher. The earliest time at the hot stage which has been observationally probed so far is the Big Bang Nucleosynthesis epoch; that epoch began at temperature of order 1 MeV, when the lifetime of the Universe was about 1 s. At that time the weak processes like

$$e^- + p \longleftrightarrow n + \nu_e$$

switched off, and the comoving number density of neutrons froze out. Somewhat later, these neutrons combined with protons into light elements in thermonuclear reactions



etc., up to ${}^7\text{Li}$. Comparison of the Big Bang Nucleosynthesis theory with the observational determination of the composition of cosmic medium gives us confidence that we understand the Universe at that epoch. Notably, we are convinced that the cosmological expansion was governed by General Relativity.

2.4 Properties of components of cosmic medium

Let us come back to photons. Their effective temperature after recombination scales as

$$T(t) \propto a^{-1}(t). \quad (9)$$

²The reason is that the number density of electrons and protons is small compared to the number density of photons. At temperature above 3000 K, a hydrogen atom formed in an electron-proton encounter is quickly destroyed by absorbing a photon from the high energy tail of the Planck distribution, and after that the electron/proton lives long time before it meets proton/electron and forms a hydrogen atom again. In thermodynamical terms, at temperatures above 3000 K there is large entropy per electron/proton, and recombination is not thermodynamically favourable because of entropy considerations.

This behaviour is characteristic of *ultra-relativistic* free species (at zero chemical potential). The same formula is valid for ultra-relativistic particles (at zero chemical potential) which are in thermal equilibrium. Thermal equilibrium means adiabatic expansion; during adiabatic expansion, the temperature of ultra-relativistic gas scales as the inverse size of the system, according to usual thermodynamics. The energy density of ultra-relativistic gas scales as $\rho \propto T^4$, and pressure is $p = \rho/3$.

Both for free photons, and for photons in thermal equilibrium, the number density behaves as follows,

$$n_\gamma = \text{const} \cdot T^3 \propto a^{-3} ,$$

and the energy density is given by the Stefan–Boltzmann law,

$$\rho_\gamma = \frac{\pi^2}{30} \cdot 2 \cdot T^4 \propto a^{-4} , \quad (10)$$

where the factor 2 accounts for two photon polarizations. The present number density of relic photons is

$$n_{\gamma,0} = 410 \text{ cm}^{-3} , \quad (11)$$

and their energy density is

$$\rho_{\gamma,0} = 2.7 \cdot 10^{-10} \frac{\text{GeV}}{\text{cm}^3} . \quad (12)$$

An important characteristic of the early Universe is the entropy density of cosmic plasma in thermal equilibrium. It is given by

$$s = \frac{2\pi^2}{45} g_* T^3 , \quad (13)$$

where g_* is the number of degrees of freedom with $m \lesssim T$, that is, the degrees of freedom which are relativistic at temperature T ; each spin state counts as an independent degree of freedom, and fermions contribute to g_* with a factor of 7/8. The point is that the entropy density scales *exactly* as a^{-3} ,

$$sa^3 = \text{const} , \quad (14)$$

while temperature scales *approximately* as a^{-1} . The property (14) is nothing but the reflection of the fact that the Universe expands relatively slowly, and the evolution is adiabatic (barring fairly exotic scenarios with strong entropy generation at some early cosmological epoch). The temperature would scale as a^{-1} if the number of relativistic degrees of freedom would be independent of time. This is not the case, however. Indeed, the value of g_* depends on temperature: at $T \sim 10 \text{ MeV}$ relativistic species are photons, neutrinos, electrons and positrons, while at $T \sim 1 \text{ GeV}$ four flavors of quarks, gluons, muons and τ -leptons are relativistic too. The number of degrees of freedom in the Standard Model at $T \gtrsim 100 \text{ GeV}$ is

$$g_*(100 \text{ GeV}) \approx 100 .$$

The present value of the entropy density (taking into account neutrinos as if they were massless) is

$$s_0 \approx 3000 \text{ cm}^{-3} . \quad (15)$$

The parameter g_* determines not only the entropy density but also the energy density of the cosmic plasma in thermal equilibrium. The Stefan–Boltzmann law gives

$$\rho_{rad} = \frac{\pi^2}{30} g_* T^4 , \quad (16)$$

where subscript *rad* indicates that we are talking about the relativistic component (radiation in broad sense).

Let us now turn to non-relativistic particles: baryons, massive neutrinos, dark matter particles, etc. If they are not destroyed during the evolution of the Universe (that is, they are stable and do not annihilate), their number density merely gets diluted,

$$n \propto a^{-3} . \quad (17)$$

This means, in particular, that the baryon-to-photon ratio stays constant in time (we consider for definiteness the late Universe, $T \lesssim 100$ keV),

$$\eta_B \equiv \frac{n_B}{n_\gamma} = \text{const} \approx 6.1 \cdot 10^{-10} . \quad (18)$$

The numerical value here is determined by two independent methods: one is Big Bang Nucleosynthesis theory and measurements of the light element abundances, and another is the measurements of the CMB temperature anisotropy. It is reassuring that these methods give consistent results (with comparable precision).

The energy density of non-relativistic particles scales as

$$\rho(t) = m \cdot n(t) \propto a^{-3}(t) , \quad (19)$$

in contrast to more rapid fall off (10) characteristic of relativistic species.

Finally, dark energy density does not decrease in time as fast as in Eqs. (10) or (19). In fact, to a reasonable approximation dark energy density does not depend on time at all,

$$\rho_\Lambda = \text{const} . \quad (20)$$

Dark energy with exactly time-independent energy density is the same thing as the cosmological constant, or Λ -term.

2.5 Composition of the present Universe

The basic equation governing the expansion rate of the Universe is the Friedmann equation, which we write for the case of spatially flat Universe,

$$H^2 \equiv \left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi}{3} G \rho , \quad (21)$$

where dot denotes derivative with respect to time t , ρ is the *total* energy density in the Universe and G is Newton's gravity constant; in natural units $G = M_{Pl}^{-2}$ where $M_{Pl} = 1.2 \cdot 10^{19}$ GeV is the Planck mass. The Friedmann equation is nothing but the (00)-component of the Einstein equations of General Relativity,

$$R_{00} - \frac{1}{2} g_{00} R = 8\pi T_{00} ,$$

specified to FLRW metric.

Let us introduce the parameter

$$\rho_c = \frac{3}{8\pi G} H_0^2 \approx 5 \cdot 10^{-6} \frac{\text{GeV}}{\text{cm}^3} . \quad (22)$$

According to Eq. (21), it is equal to the sum of all forms of energy density in the present Universe. As a side remark, we note that the latter statement would not be true if our Universe were not spatially flat. However, according to observations, spatial flatness holds to a very good precision, corresponding to less than 1 per cent deviation of the total energy density from ρ_c [9].

As we now discuss, the cosmological data correspond to a very weird composition of the Universe.

Before proceeding, let us introduce a notion traditional in the analysis of the composition of the present Universe. For every type of matter i with the present energy density $\rho_{i,0}$, one defines the parameter

$$\Omega_i = \frac{\rho_{i,0}}{\rho_c} .$$

Then Eq. (21) tells that $\sum_i \Omega_i = 1$ where the sum runs over all forms of energy. Let us now discuss contributions of different species to this sum.

We begin with **baryons**. The result (18) gives

$$\rho_{B,0} = m_B \cdot n_{B,0} \approx 2.4 \cdot 10^{-7} \frac{\text{GeV}}{\text{cm}^3} . \quad (23)$$

Comparing this result with the value of ρ_c given in Eq. (22), one finds

$$\Omega_B = 0.045 .$$

Thus, baryons constitute rather small fraction of the present energy density in the Universe.

Photons contribute even smaller fraction, as is clear from Eq. (12), namely $\Omega_\gamma \approx 5 \cdot 10^{-5}$. From electric neutrality, the number density of **electrons** is about the same as that of baryons, so electrons contribute negligible fraction to the total mass density. The remaining known stable particles are **neutrinos**. Their number density is calculable in Hot Big Bang theory and these calculations are nicely confirmed by Big Bang Nucleosynthesis. The present number density of each type of neutrinos is

$$n_{\nu_\alpha,0} = 110 \text{ cm}^{-3} ,$$

where $\nu_\alpha = \nu_e, \nu_\mu, \nu_\tau$ (more appropriately, ν_α are neutrino mass eigenstates). Direct limit on the mass of electron neutrino, $m_{\nu_e} < 2 \text{ eV}$, together with the observations of neutrino oscillations suggest that every type of neutrino has mass smaller than 2 eV (neutrinos with masses above 0.05 eV must be degenerate, according to neutrino oscillation data). The energy density of all types of neutrinos is thus smaller than ρ_c :

$$\rho_{\nu,\text{total}} = \sum_\alpha m_{\nu_\alpha} n_{\nu_\alpha} < 3 \cdot 2 \text{ eV} \cdot 110 \frac{1}{\text{cm}^3} \sim 6 \cdot 10^{-7} \frac{\text{GeV}}{\text{cm}^3} ,$$

which means that $\Omega_{\nu,\text{total}} < 0.12$. This estimate does not make use of any cosmological data. In fact, cosmological observations give stronger bound

$$\Omega_{\nu,\text{total}} \lesssim 0.014 . \quad (24)$$

This bound is mostly due to the analysis of the structures at relatively small length scales, and has to do with streaming of neutrinos from the gravitational potential wells at early times when neutrinos were moving fast. In terms of the neutrino masses the bound (24) reads [10, 11]

$$\sum m_{\nu_\alpha} < 0.6 \text{ eV} ,$$

so every neutrino must be lighter than 0.2 eV . It is worth noting that the atmospheric neutrino data, as well as K2K, Minos and T2K experiments tell us that the mass of at least one neutrino must be larger than about 0.05 eV . Comparing these numbers, one sees that it may be feasible to measure neutrino masses by cosmological observations (!) in the future.

Coming back to our main topic here, we conclude that most of the energy density in the present Universe is not in the form of known particles; most energy in the present Universe must be in “something unknown”. Furthermore, this “something unknown” has two components: clustered (dark matter) and unclustered (dark energy).

Clustered dark matter consists presumably of new stable massive particles. These make clumps of energy (mass) which constitute most of the mass of galaxies and clusters of galaxies. There are various ways of estimating the contribution of non-baryonic dark matter into the total energy density of the Universe (see Ref. [2] for details):

- Composition of the Universe affects the angular anisotropy of cosmic microwave background. Quite accurate measurements of the CMB anisotropy, available today, enable one to estimate the total mass density of dark matter.

- Composition of the Universe, and especially the density of non-baryonic dark matter, is crucial for structure formation of the Universe. Comparison of the results of numerical simulations of structure formation with observational data gives reliable estimate of the mass density of non-baryonic clustered dark matter.

The bottom line is that the non-relativistic component constitutes about 27 per cent of the total present energy density, which means that non-baryonic dark matter has

$$\Omega_{DM} \approx 0.22 , \quad (25)$$

the rest is due to baryons.

There is direct evidence that dark matter exists in the largest gravitationally bound objects – clusters of galaxies. There are various methods to determine the gravitating mass of a cluster, and even mass distribution in a cluster, which give consistent results. To name a few:

- One measures velocities of galaxies in galactic clusters, and makes use of the gravitational virial theorem,

$$\text{Kinetic energy of a galaxy} = \frac{1}{2} \text{Potential energy} .$$

In this way one obtains the gravitational potential, and thus the distribution of the total mass in a cluster.

- Another measurement of masses of clusters makes use of intracluster gas. Its temperature obtained from X-ray measurements is also related to the gravitational potential.

- Fairly accurate reconstruction of mass distributions in clusters is obtained from the observations of gravitational lensing of background galaxies by clusters.

These methods enable one to measure mass-to-light ratio in clusters of galaxies. Assuming that this ratio applies to all matter in the Universe³, one arrives at the estimate for the mass density of clumped matter in the present Universe. Remarkably, this estimate agrees with Eq. (25).

Finally, dark matter exists also in galaxies. Its distribution is measured by the observations of rotation velocities of distant stars and gas clouds around a galaxy.

Thus, cosmologists are confident that much of the energy density in our Universe consists of new stable particles. We will see that there is good chance for the LHC to produce these particles.

Unclustered dark energy. Non-baryonic clustered dark matter is not the whole story. Making use of the above estimates, one obtains an estimate for the energy density of all particles, $\Omega_\gamma + \Omega_B + \Omega_{\nu, total} + \Omega_{DM} \approx 0.27$. This implies that 73 per cent of the energy density is unclustered. This component is called dark energy; it has the properties similar to those of vacuum. We will briefly discuss dark energy in Section 5.

All this fits nicely all cosmological observations, but does not fit to the Standard Model of particle physics. It is our hope that the LHC will shed light at least on some of the properties of the Universe.

³This is a fairly strong assumption, since only about 10 per cent of galaxies are in clusters.

2.6 Regimes of cosmological expansion

The cosmological expansion at the present epoch is determined mostly by dark energy, since its contribution to the right hand side of the Friedmann equation (21) is the largest,

$$\Omega_\Lambda = 0.73 .$$

Non-relativistic matter (dark matter and baryons) is also non-negligible,

$$\Omega_M = 0.27 , \quad (26)$$

while the energy density of relativistic matter (photons and neutrinos, if one of the neutrino species is massless or very light) is negligible today. This was not always the case. Making use of Eq. (10) for photons and relativistic neutrinos, Eq. (17) for non-relativistic matter, and assuming for definiteness that dark energy density is constant in time, we can rewrite the Friedmann equation (21) in the following form

$$\begin{aligned} H^2(t) &= \frac{8\pi}{3M_{Pl}^2} [\rho_\Lambda + \rho_M(t) + \rho_{rad}(t)] \\ &= H_0^2 \left[\Omega_\Lambda + \Omega_M \left(\frac{a_0}{a(t)} \right)^3 + \Omega_{rad} \left(\frac{a_0}{a(t)} \right)^4 \right] . \end{aligned} \quad (27)$$

It is appropriate for our purposes to treat neutrinos as massless particles; including their contribution to Ω_{rad} one has

$$\Omega_{rad} = 8.4 \cdot 10^{-5} . \quad (28)$$

Equation (27) tells that at early times, when the scale factor $a(t)$ was small, the expansion was dominated by relativistic matter (“radiation”), later on there was long period of domination of the non-relativistic matter, and in future the expansion will be dominated by dark energy,

$$\dots \implies \text{Radiation domination} \implies \text{Matter domination} \implies \Lambda\text{-domination} .$$

Dots here denote some cosmological epoch preceding the hot stage of evolution; as we discuss in Section 6, we are confident that such an epoch existed, but do not quite know what it was. Making use of (26) and (28), it is straightforward to find the redshift at radiation–matter equality, when the first two terms in (27) are equal,

$$1 + z_{eq} = \frac{a_0}{a(t_{eq})} = \frac{\Omega_M}{\Omega_{rad}} \approx 3000 ,$$

and using the Friedmann equation one finds the age of the Universe at equality

$$t_{eq} \approx 60\,000 \text{ years} .$$

Note that recombination occurred at matter domination, but rather soon after equality, see (7).

It is useful for what follows to find the evolution of the scale factor at the radiation domination epoch. At that time the energy density is given by Eq. (16), so that the Friedmann equation can be written as follows

$$H = \frac{T^2}{M_{Pl}^*} , \quad (29)$$

where $M_{Pl}^* = M_{Pl}/(1.66\sqrt{g_*})$. Now, we neglect for simplicity the dependence of g_* on temperature, and hence on time, and recall that in this case the temperature scales as a^{-1} , see Eq. (14). Hence, we obtain

$$\frac{\dot{a}}{a} = \frac{\text{const}}{a^2} .$$

This gives the desired evolution law

$$a(t) = \text{const} \cdot \sqrt{t} . \quad (30)$$

The constant here does not have physical significance, as one can rescale the coordinates x at some fixed moment of time, thus changing the normalization of a .

There are several points to note regarding the result (30). First, the expansion *decelerates*:

$$\ddot{a} < 0 .$$

This property holds also for the matter dominated epoch, but, as we see momentarily, it does not hold for domination of the dark energy.

Second, time $t = 0$ is the Big Bang singularity (assuming erroneously that the Universe starts being radiation dominated). The expansion rate

$$H(t) = \frac{1}{2t}$$

diverges as $t \rightarrow 0$, and so does the energy density $\rho(t) \propto H^2(t)$ and temperature $T \propto \rho^{1/4}$. Of course, the classical General Relativity and usual notions of statistical mechanics (e.g., temperature itself) are not applicable very near the singularity, but our result suggests that in the picture we discuss (hot epoch right after the Big Bang), the Universe starts its classical evolution in a very hot and dense state, and its expansion rate is very high in the beginning. It is customary to assume for illustrational purposes that the relevant quantities in the beginning of the classical expansion take the Planck values, $\rho \sim M_{Pl}^4$, $H \sim M_{Pl}$, etc.

Third, at a given moment of time the size of a causally connected region is finite. Consider signals emitted right after the Big Bang and travelling with the speed of light. These signals travel along the light cone with $ds = 0$, and hence $a(t)dx = dt$. So, the coordinate distance that a signal travels from the Big Bang to time t is

$$x = \int_0^t \frac{dt}{a(t)} \equiv \eta . \quad (31)$$

In the radiation dominated Universe

$$\eta = \text{const} \cdot \sqrt{t} .$$

The physical distance from the emission point to the position of the signal is

$$l_{H,t} = a(t)x = a(t) \int_0^t \frac{dt}{a(t)} = 2t .$$

As expected, this physical distance is finite, and it gives the size of a causally connected region at time t . It is called the horizon size (more precisely, the size of particle horizon). A related property is that an observer at time t can see only the part of the Universe whose current physical size is $l_{H,t}$. Both at radiation and matter domination one has, modulo numerical constant of order 1,

$$l_{H,t} \sim H^{-1}(t) .$$

To give an idea of numbers, the horizon size at the present epoch is

$$l_{H,t_0} \approx 15 \text{ Gpc} \simeq 4.5 \cdot 10^{28} \text{ cm} .$$

One property of the Universe that starts its expansion from radiation domination is puzzling. Using Eq. (31) one sees that the size of the observable Universe increases in time. For example, the coordinate size of the present horizon is about 50 times larger than the coordinate size of the horizon at recombination. Hence, when performing CMB observations we see 50^2 regions on the sphere of last scattering

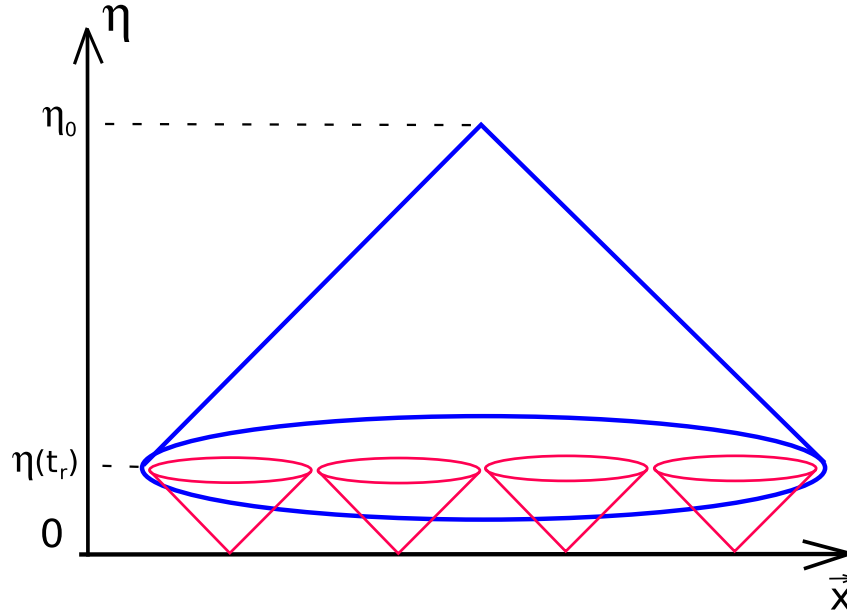


Fig. 1: Causal structure of space-time in hot Big Bang theory

which were causally disconnected at the recombination epoch, see Fig. 1. Yet they look exactly the same! Clearly, this is a problem for the hot Big Bang theory, which is called horizon problem. We will see in Section 6 that this problem has a somewhat different side, which unambiguously shows that the hot Big Bang theory is not the whole story: the hot epoch was preceded by some other, very different epoch of the cosmological evolution.

To end up this section, let us note that the properties of the Universe dominated by dark energy are quite different. Assuming for definiteness that ρ_Λ is independent of time, we immediately find the solution to the Friedmann equation for the Λ -dominated Universe:

$$a(t) = \text{const} \cdot e^{H_\Lambda t}, \quad (32)$$

where $H_\Lambda = \sqrt{8\pi\rho_\Lambda/3M_{Pl}^2}$. The cosmological expansion *accelerates*,

$$\ddot{a} > 0.$$

The dark energy was introduced precisely for explaining the accelerated expansion of the Universe at the present epoch.

3 Dark matter

Dark matter is characterized by the mass-to-entropy ratio,

$$\left(\frac{\rho_{DM}}{s}\right)_0 = \frac{\Omega_{DM}\rho_c}{s_0} \approx \frac{0.22 \cdot 5 \cdot 10^{-6} \text{ GeV} \cdot \text{cm}^{-3}}{3000 \text{ cm}^{-3}} = 4 \cdot 10^{-10} \text{ GeV}. \quad (33)$$

This ratio is constant in time since the freeze out of dark matter density: both number density of dark matter particles n_{DM} (and hence their mass density $\rho_{DM} = m_{DM}n_{DM}$) and entropy density dilute exactly as a^{-3} .

Dark matter is crucial for our existence, for the following reason. Density perturbations in baryon-electron-photon plasma before recombination do not grow because of high pressure, which is mostly due to photons; instead, perturbations are sound waves propagating in plasma with time-independent

amplitudes. Hence, in a Universe without dark matter, density perturbations in baryonic component would start to grow only after baryons decouple from photons, i.e., after recombination. The mechanism of the growth is pretty simple: an overdense region gravitationally attracts surrounding matter; this matter falls into the overdense region, and the density contrast increases. In the expanding matter dominated Universe this gravitational instability results in the density contrast growing like $(\delta\rho/\rho)(t) \propto a(t)$. Hence, in a Universe without dark matter, the growth factor for baryon density perturbations would be at most⁴

$$\frac{a(t_0)}{a(t_{rec})} = 1 + z_{rec} = \frac{T_{rec}}{T_0} \approx 10^3. \quad (34)$$

The initial amplitude of density perturbations is very well known from the CMB anisotropy measurements, $(\delta\rho/\rho)_i = 5 \cdot 10^{-5}$. Hence, a Universe without dark matter would still be pretty homogeneous: the density contrast would be in the range of a few per cent. No structure would have been formed, no galaxies, no life. No structure would be formed in future either, as the accelerated expansion due to dark energy will soon terminate the growth of perturbations.

Since dark matter particles decoupled from plasma much earlier than baryons, perturbations in dark matter started to grow much earlier. The corresponding growth factor is larger than Eq. (34), so that the dark matter density contrast at galactic and sub-galactic scales becomes of order one, perturbations enter non-linear regime and form dense dark matter clumps at $z = 5 - 10$. Baryons fall into potential wells formed by dark matter, so dark matter and baryon perturbations develop together soon after recombination. Galaxies get formed in the regions where dark matter was overdense originally. The development of perturbations in our Universe is shown in Fig. 2. For this picture to hold, dark matter particles must be non-relativistic early enough, as relativistic particles fly through gravitational wells instead of being trapped there. This means, in particular, that neutrinos cannot constitute a considerable part of dark matter, hence the bound (24).

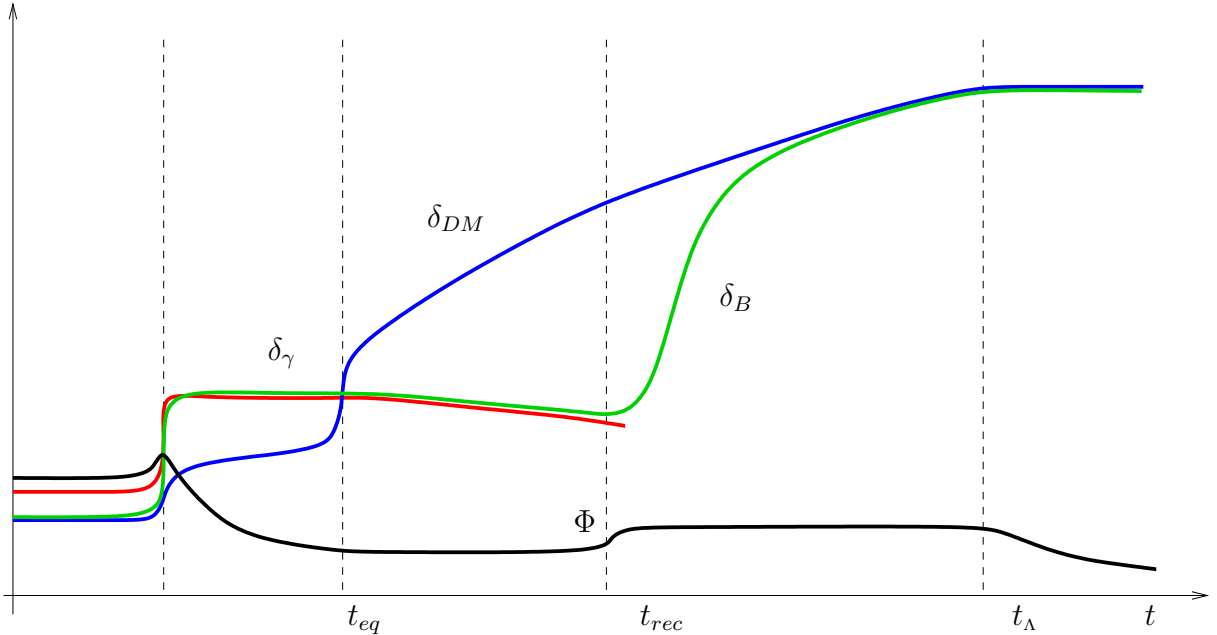


Fig. 2: A sketch of the time dependence, in the linear regime, of density contrasts of dark matter, baryons and photons, $\delta_{DM} \equiv \delta\rho_{DM}/\rho_{DM}$, δ_B and δ_γ , respectively, as well as the Newtonian potential Φ . t_{eq} and t_Λ correspond to the transitions from radiation domination to matter domination, and from decelerated expansion to accelerated expansion, t_{rec} refers to the recombination epoch.

⁴Because of the presence of dark energy, the growth factor is even somewhat smaller.

Depending on the mass of the dark matter particles and mechanism of their production in the early Universe, dark matter may be *cold* (CDM) and *warm* (WDM). Roughly speaking, CDM consists of heavy particles, while the masses of WDM particles are smaller,

$$\text{CDM : } m_{DM} \gtrsim 100 \text{ keV} , \quad (35a)$$

$$\text{WDM : } m_{DM} = 3 - 30 \text{ keV} . \quad (35b)$$

This assumes that the dark matter particles were in thermal (kinetic) equilibrium at some early times, or, more generally, that their kinetic energy was comparable to temperature. This need not be the case for very weakly interacting particles; a well known example is axions which are *cold* dark matter candidates despite their very small mass. Likewise, very weakly interacting warm dark matter particles may be much heavier than Eq. (35b) suggests.

We will discuss warm dark matter option later on, and now we move on to CDM.

3.1 WIMPS: best guess for cold dark matter

There is a simple mechanism of the dark matter generation in the early Universe. It applies to *cold* dark matter. Because of its simplicity and robustness, it is considered by many as a very likely one, and the corresponding dark matter candidates — weakly interacting massive particles, WIMPs — as the best candidates. Let us describe this mechanism in some detail.

Let us assume that there exists a heavy stable neutral particle Y , and that Y -particles can only be destroyed or created via their pair-annihilation or creation, with annihilation products being the particles of the Standard Model. We will see that the overall cosmological behaviour of Y -particles is as follows. At high temperatures, $T \gg m_Y$, the Y -particles are in thermal equilibrium with the rest of cosmic plasma; there are lots of Y -particles in the plasma, which are continuously created and annihilate. As the temperature drops below m_Y , the equilibrium number density decreases. At some “freeze-out” temperature T_f the number density becomes so small, that Y -particles can no longer meet each other during the Hubble time, and their annihilation terminates. After that the number density of survived Y ’s decreases like a^{-3} , and these relic particles contribute to the mass density in the present Universe. Our purpose is to estimate the range of properties of Y -particles, in which they serve as dark matter.

Assuming thermal equilibrium, elementary considerations of mean free path of a particle in gas give for the lifetime of a non-relativistic Y -particle in cosmic plasma, τ_{ann} ,

$$\sigma_{ann} \cdot v \cdot \tau_{ann} \cdot n_Y \sim 1 ,$$

where v is the velocity of Y -particle, σ_{ann} is the annihilation cross section at velocity v and n_Y is the equilibrium number density given by the Boltzmann law at zero chemical potential,

$$n_Y = g_Y \cdot \left(\frac{m_Y T}{2\pi} \right)^{3/2} e^{-\frac{m_Y}{T}} ,$$

where g_Y is the number of spin states of Y -particle. Note that we consider non-relativistic regime, $m_Y \ll T$. Let us introduce the notation

$$\sigma_{ann} v = \sigma_0$$

(in fact, the left hand side is to be understood as thermal average). If the annihilation occurs in s -wave, then σ_0 is a constant independent of temperature, for p -wave it is somewhat suppressed at $T \ll m_Y$. One should compare the lifetime with the Hubble time, or annihilation rate $\Gamma_{ann} \equiv \tau_{ann}^{-1}$ with the expansion rate H . At $T \sim m_Y$, the equilibrium density is of order $n_Y \sim T^3$, and $\Gamma_{ann} \gg H$ for not too small σ_0 . This means that annihilation (and, by reciprocity, creation) of Y -pairs is indeed rapid, and Y -particles are indeed in thermal equilibrium with the plasma. At very low temperature, on the other hand, the equilibrium number density $n_Y^{(eq)}$ is exponentially small, and the equilibrium rate is small,

$\Gamma_{ann}^{(eq)} \ll H$. At low temperatures we cannot, of course, make use of the equilibrium formulas: Y -particles no longer annihilate (and, by reciprocity, are no longer created), there is no thermal equilibrium with respect to creation–annihilation processes, and the number density n_Y gets diluted only because of the cosmological expansion.

The freeze-out temperature T_f is determined by the relation

$$\tau_{ann}^{-1} \equiv \Gamma_{ann} \sim H ,$$

where we can still use the equilibrium formulas, as Y -particles are in thermal equilibrium (with respect to annihilation and creation) just before freeze-out. Making use of the relation (29) between the Hubble parameter and temperature at radiation domination, we obtain

$$\sigma_0(T_f) \cdot n_Y(T_f) \sim \frac{T_f^2}{M_{Pl}^*} , \quad (36)$$

or

$$\sigma_0(T_f) \cdot g_Y \cdot \left(\frac{m_Y T_f}{2\pi} \right)^{3/2} e^{-\frac{m_Y}{T_f}} \sim \frac{T_f^2}{M_{Pl}^*} .$$

The latter equation gives the freeze-out temperature, which, up to loglog corrections, is

$$T_f \approx \frac{m_Y}{\ln(M_{Pl}^* m_Y \sigma_0)} \quad (37)$$

(the possible dependence of σ_0 on temperature is irrelevant in the right hand side: we are doing the calculation in the leading-log approximation anyway). Note that this temperature is somewhat lower than m_Y , if the relevant microscopic mass scale is much below M_{Pl} . This means that Y -particles freeze out when they are indeed non-relativistic, hence the term “cold dark matter”. The fact that the annihilation and creation of Y -particles terminate at relatively low temperature has to do with rather slow expansion of the Universe, which should be compensated for by the smallness of the number density n_Y .

At the freeze-out temperature, we make use of Eq. (36) and obtain

$$n_Y(T_f) = \frac{T_f^2}{M_{Pl}^* \sigma_0(T_f)} .$$

Note that this density is inversely proportional to the annihilation cross section (modulo logarithm). The reason is that for higher annihilation cross section, the creation–annihilation processes are longer in equilibrium, and less Y -particles survive.

Up to a numerical factor of order 1, the number-to-entropy ratio at freeze-out is

$$\frac{n_Y}{s} \simeq \frac{1}{g_*(T_f) M_{Pl}^* T_f \sigma_0(T_f)} . \quad (38)$$

This ratio stays constant until the present time, so the present number density of Y -particles is $n_{Y,0} = s_0 \cdot (n_Y/s)_{freeze-out}$, and the mass-to-entropy ratio is

$$\frac{\rho_{Y,0}}{s_0} = \frac{m_Y n_{Y,0}}{s_0} \simeq \frac{\ln(M_{Pl}^* m_Y \sigma_0)}{g_*(T_f) M_{Pl}^* \sigma_0(T_f)} \simeq \frac{\ln(M_{Pl}^* m_Y \sigma_0)}{\sqrt{g_*(T_f)} M_{Pl} \sigma_0(T_f)} ,$$

where we made use of Eq. (37). This formula is remarkable. The mass density depends mostly on one parameter, the annihilation cross section σ_0 . The dependence on the mass of Y -particle is through the logarithm and through $g_*(T_f)$; it is very mild. The value of the logarithm here is between 30 and 40, depending on parameters (this means, in particular, that freeze-out occurs when the temperature drops 30 to 40 times below the mass of Y -particle). Plugging in other numerical values ($g_*(T_f) \sim 100$,

$M_{Pl}^* \sim 10^{18}$ GeV), as well as numerical factor omitted in Eq. (38), and comparing with Eq. (33) we obtain the estimate

$$\sigma_0(T_f) \equiv \langle \sigma v \rangle(T_f) = (1 \div 2) \cdot 10^{-36} \text{ cm}^2. \quad (39)$$

This is a weak scale cross section, which tells us that the relevant energy scale is TeV. We note in passing that the estimate (39) is quite precise and robust.

If the annihilation occurs in s -wave, the annihilation cross section may be parametrized as $\sigma_0 = \alpha^2/M^2$ where α is some coupling constant, and M is a mass scale (which may be higher than m_Y). This parametrization is suggested by the picture of Y pair-annihilation via the exchange by another particle of mass M . With $\alpha \sim 10^{-2}$, the estimate for the mass scale is roughly $M \sim 1$ TeV. Thus, with very mild assumptions, we find that the non-baryonic dark matter may naturally originate from the TeV-scale physics. In fact, what we have found can be understood as an approximate equality between the cosmological parameter, mass-to-entropy ratio of dark matter, and the particle physics parameters,

$$\text{mass-to-entropy} \simeq \frac{1}{M_{Pl}} \left(\frac{\text{TeV}}{\alpha_W} \right)^2.$$

Both are of order 10^{-10} GeV, and it is very tempting to think that this is not a mere coincidence. If it is not, the dark matter particle should be found at the LHC.

Of course, the most prominent candidate for WIMP is neutralino of the supersymmetric extensions of the Standard Model. The situation with neutralino is somewhat tense, however. The point is that the pair-annihilation of neutralinos often occurs in p -wave, rather than s -wave. This gives the suppression factor in $\sigma_0 \equiv \langle \sigma_{ann} v \rangle$, proportional to $v^2 \sim T_f/m_Y \sim 1/30$. Hence, neutralinos tend to be overproduced in most of the parameter space of MSSM and other models. Yet neutralino remains a good candidate, especially at high $\tan \beta$.

3.2 Warm dark matter: light gravitinos

The cold dark matter scenario successfully describes the bulk of the cosmological data. Yet, there are clouds above it. First, according to numerical simulations, CDM scenario tends to overproduce small objects — dwarf galaxies: it predicts hundreds of satellite dwarf galaxies in the vicinity of a large galaxy like Milky Way, whereas only dozens of satellites have been observed so far. Second, again according to simulations, CDM tends to produce too high densities in galactic centers (cusps in density profiles); this feature is not confirmed by observations either. There is no strong discrepancy yet, but one may be motivated to analyse a possibility that dark matter is not that cold.

An alternative to CDM is warm dark matter whose particles decouple being relativistic. Let us assume for definiteness that they are in kinetic equilibrium with cosmic plasma when their number density freezes out (thermal relic). After kinetic equilibrium breaks down, and WDM particles decouple completely, their spatial momenta decrease as a^{-1} , i.e., the momenta are of order T all the time after decoupling. WDM particles become non-relativistic at $T \sim m$, where m is their mass. Only after that the WDM perturbations start to grow⁵: as we mentioned above, relativistic particles escape from gravitational potentials, so the gravitational potentials get smeared out instead of getting deeper. Before becoming non-relativistic, WDM particles travel the distance of the order of the horizon size; the WDM perturbations therefore are suppressed at those scales. The horizon size at the time t_{nr} when $T \sim m$ is of order

$$l(t_{nr}) \simeq H^{-1}(T \sim m) = \frac{M_{Pl}^*}{T^2} \sim \frac{M_{Pl}^*}{m^2}.$$

Due to the expansion of the Universe, the corresponding length at present is

$$l_0 = l(t_{nr}) \frac{a_0}{a(t_{nr})} \sim l(t_{nr}) \frac{T}{T_0} \sim \frac{M_{Pl}}{m T_0}, \quad (40)$$

⁵The situation in fact is somewhat more complicated, but this simplified picture will be sufficient for our estimates.

where we neglected (rather weak) dependence on g_* . Hence, in WDM scenario, structures of sizes smaller than l_0 are less abundant as compared to CDM. Let us point out that l_0 refers to the size of the perturbation as if it were in the linear regime; in other words, this is the size of the region from which matter collapses into a compact object.

The present size of a dwarf galaxy is a few kpc, and the density is about 10^6 of the average density in the Universe. Hence, the size l_0 for these objects is of order $100 \text{ kpc} \simeq 3 \cdot 10^{23} \text{ cm}$. Requiring that perturbations of this size, but not much larger, are suppressed, we obtain from Eq. (40) the estimate (35b) for the mass of WDM particles.

Among WDM candidates, light gravitino is probably the best motivated. The gravitino mass is of order

$$m_{3/2} \simeq \frac{F}{M_{Pl}} ,$$

where $\sqrt{F}$ is the supersymmetry breaking scale. Hence, gravitino masses are in the right ballpark for rather low supersymmetry breaking scales, $\sqrt{F} \sim 10^6 - 10^7 \text{ GeV}$. This is the case, e.g., in gauge mediation scenario. With so low mass, gravitino is the lightest supersymmetric particle (LSP), so it is stable in many supersymmetric extensions of the Standard Model. From this viewpoint gravitinos can indeed serve as dark matter particles. For what follows, important parameters are the widths of decays of other superpartners into gravitino and the Standard Model particles. These are of order

$$\Gamma_{\tilde{S}} \simeq \frac{M_{\tilde{S}}^5}{F^2} \simeq \frac{M_{\tilde{S}}^5}{m_{3/2}^2 M_{Pl}^2} , \quad (41)$$

where $M_{\tilde{S}}$ is the mass of the superpartner.

One mechanism of the gravitino production in the early Universe is decays of other superpartners. Gravitino interacts with everything else so weakly, that once produced, it moves freely, without interacting with cosmic plasma. At production, gravitinos are relativistic, hence they are indeed *warm* dark matter candidates. Let us assume that production in decays is the dominant mechanism and consider under what circumstances the present mass density of gravitinos coincides with that of dark matter.

The rate of gravitino production in decays of superpartners of the type $\tilde{S}$ in the early Universe is

$$\frac{d(n_{3/2}/s)}{dt} = \frac{n_{\tilde{S}}}{s} \Gamma_{\tilde{S}} ,$$

where $n_{3/2}$ and $n_{\tilde{S}}$ are number densities of gravitinos and superpartners, respectively, and s is the entropy density. For superpartners in thermal equilibrium, one has $n_{\tilde{S}}/s = \text{const} \sim g_*^{-1}$ for $T \gtrsim M_{\tilde{S}}$, and $n_{\tilde{S}}/s \propto \exp(-M_{\tilde{S}}/T)$ at $T \ll M_{\tilde{S}}$. Hence, the production is most efficient at $T \sim M_{\tilde{S}}$, when the number density of superpartners is still large, while the Universe expands most slowly. The density of gravitinos produced in decays of $\tilde{S}$'s is thus given by

$$\frac{n_{3/2}}{s} \simeq \left(\frac{d(n_{3/2}/s)}{dt} \cdot H^{-1} \right)_{T \sim M_{\tilde{S}}} \simeq \frac{\Gamma_{\tilde{S}}}{g_*} H^{-1}(T \sim M_{\tilde{S}}) \simeq \frac{1}{g_*} \cdot \frac{M_{\tilde{S}}^5}{m_{3/2}^2 M_{Pl}^2} \cdot \frac{M_{Pl}^*}{M_{\tilde{S}}^2} .$$

This gives the mass-to-entropy ratio today:

$$\frac{m_{3/2} n_{3/2}}{s} \simeq \sum_{\tilde{S}} \frac{M_{\tilde{S}}^3}{g_*^{3/2} M_{Pl} m_{3/2}} , \quad (42)$$

where the sum runs over all superpartner species *which have ever been relativistic in thermal equilibrium*. The correct value (33) is obtained for gravitino masses in the range (35b) at

$$M_{\tilde{S}} = 100 - 300 \text{ GeV} . \quad (43)$$

Thus, the scenario with gravitino as warm dark matter particle requires light superpartners, which are to be discovered at the LHC.

A few comments are in order. First, decays of superpartners is not the only mechanism of gravitino production: gravitinos may also be produced in scattering of superpartners. To avoid overproduction of gravitinos in the latter processes, one has to assume that the maximum temperature in the Universe (reached after post-inflationary reheating stage) is quite low, $T_{max} \sim 1 - 10$ TeV. This is not a particularly plausible assumption, but it is consistent with everything else in cosmology and can indeed be realized in some models of inflation. Second, existing constraints on masses of strongly interacting superpartners (gluinos and squarks) suggest that their masses exceed Eq. (43). Hence, these particles should not contribute to the sum in Eq. (42), otherwise WDM gravitinos would be overproduced. This is possible, if masses of squarks and gluinos are larger than T_{max} , so that they were never abundant in the early Universe. Third, gravitino produced in decays of superpartners is *not* a thermal relic, as it was never in thermal equilibrium with the rest of cosmic plasma. Nevertheless, since gravitinos are produced at $T \sim M_{\tilde{g}}$ and at that time have energy $E \sim M_{\tilde{g}} \sim T$, our estimate (40) does apply. Finally, the decay into gravitino and the Standard Model particles is the only decay channel for the next-to-lightest superpartner (NLSP). Hence, the estimate for the total width of NLSP is given by Eq. (41), so that

$$c\tau_{NLSP} = \text{a few} \cdot \text{mm} - \text{a few} \cdot 100 \text{ m}$$

for $m_{2/3} = 3 - 30$ keV and $M_{NLSP} = 100 - 300$ GeV. Thus, NLSP should either be visible in a detector, or fly it through.

Needless to say, the warm gravitino scenario is a lot more contrived than the WIMP option. It is reassuring, however, that it can be ruled out or confirmed at the LHC.

3.3 Discussion

If dark matter particles are indeed WIMPs, and the relevant energy scale is of order 1 TeV, then the Hot Big Bang theory will be probed experimentally up to temperature of (a few) $\cdot (10 - 100)$ GeV and down to age $10^{-9} - 10^{-11}$ s in relatively near future (compare to 1 MeV and 1 s accessible today through Big Bang Nucleosynthesis). With microscopic physics to be known from collider experiments, the WIMP density will be reliably calculated and checked against the data from observational cosmology. Thus, WIMP scenario offers a window to a very early stage of the evolution of the Universe.

If dark matter particles are gravitinos, the prospect of probing quantitatively so early stage of the cosmological evolution is not so bright: it would be very hard, if at all possible, to get an experimental handle on the gravitino mass; furthermore, the present gravitino mass density depends on an unknown reheat temperature T_{max} . On the other hand, if this scenario is realized in Nature, then the whole picture of the early Universe will be quite different from our best guess on the early cosmology. Indeed, gravitino scenario requires low reheat temperature, which in turn calls for rather exotic mechanism of inflation.

The mechanisms discussed here are by no means the only ones capable of producing dark matter, and WIMPs and gravitinos are by no means the only candidates for dark matter particles. Other dark matter candidates include axions, sterile neutrinos, Q-balls, very heavy relics produced towards the end of inflation, etc. Hence, even though there are grounds to hope that the dark matter problem will be solved by the LHC, there is no guarantee at all.

4 Baryon asymmetry of the Universe

In the present Universe, there are baryons and almost no antibaryons. The number density of baryons today is characterized by the ratio η_B , see Eq. (18). In the early Universe, the appropriate quantity is

$$\Delta_B = \frac{n_B - n_{\bar{B}}}{s},$$

where $n_{\bar{B}}$ is the number density of antibaryons, and s is the entropy density. If the baryon number is conserved, and the Universe expands adiabatically, Δ_B is constant, and its value, up to a numerical factor, is equal to η (cf. Eqs. (11) and (15)). More precisely,

$$\Delta_B \approx 0.8 \cdot 10^{-10} .$$

At early times, at temperatures well above 100 MeV, cosmic plasma contained many quark-antiquark pairs, whose number density was of the order of the entropy density,

$$n_q + n_{\bar{q}} \sim s ,$$

while the baryon number density was related to densities of quarks and antiquarks as follows (baryon number of a quark equals 1/3),

$$n_B = \frac{1}{3}(n_q - n_{\bar{q}}) .$$

Hence, in terms of quantities characterizing the very early epoch, the baryon asymmetry may be expressed as

$$\Delta_B \sim \frac{n_q - n_{\bar{q}}}{n_q + n_{\bar{q}}} .$$

We see that there was one extra quark per about 10 billion quark-antiquark pairs! It is this tiny excess that is responsible for the entire baryonic matter in the present Universe: as the Universe expanded and cooled down, antiquarks annihilated with quarks, and only the excessive quarks remained and formed baryons.

There is no logical contradiction to suppose that the tiny excess of quarks over antiquarks was built in as an initial condition. This is not at all satisfactory for a physicist, however. Furthermore, inflationary scenario does not provide such an initial condition for the hot Big Bang epoch; rather, inflation theory predicts that the Universe was baryon-symmetric just after inflation. Hence, one would like to explain the baryon asymmetry dynamically.

The baryon asymmetry may be generated from initially symmetric state only if three necessary conditions, dubbed Sakharov's conditions, are satisfied. These are

- (i) baryon number non-conservation;
- (ii) C- and CP-violation;
- (iii) deviation from thermal equilibrium.

All three conditions are easily understood. (i) If baryon number were conserved, and initial net baryon number in the Universe was zero, the Universe today would still be symmetric. (ii) If C or CP were conserved, then the rate of reactions with particles would be the same as the rate of reactions with antiparticles, and no asymmetry would be generated. (iii) Thermal equilibrium means that the system is stationary (no time dependence at all). Hence, if the initial baryon number is zero, it is zero forever, unless there are deviations from thermal equilibrium.

There are two well understood mechanisms of baryon number non-conservation. One of them emerges in Grand Unified Theories and is due to the exchange of super-massive particles. It is similar, say, to the mechanism of charm non-conservation in weak interactions, which occurs via the exchange of heavy W -bosons. The scale of these new, baryon number violating interactions is the Grand Unification scale, presumably of order $M_{GUT} \simeq 10^{16}$ GeV. It is rather unlikely that the baryon asymmetry was generated due to this mechanism: the relevant temperature would be of order M_{GUT} , while so high reheate temperature after inflation is difficult to obtain.

Another mechanism is non-perturbative [12] and is related to the triangle anomaly in the baryonic current (a keyword here is "sphaleron" [13, 14]). It exists already in the Standard Model, and, possibly with slight modifications, operates in all its extensions. The two main features of this mechanism, as

applied to the early Universe, is that it is effective over a wide range of temperatures, $100 \text{ GeV} < T < 10^{11} \text{ GeV}$, and that it conserves $(B - L)$.

Let us pause here to discuss the physics behind electroweak baryon and lepton number non-conservation in little more detail, though still at a qualitative level. A detailed analysis can be found in the book [15] and in references therein.

The first object to consider is the baryonic current,

$$B^\mu = \frac{1}{3} \cdot \sum_i \bar{q}_i \gamma^\mu q_i ,$$

where the sum runs over quark flavors. Naively, the baryonic current is conserved, but at the quantum level its divergence is non-zero, the effect called triangle anomaly (similar effect goes under the name of axial anomaly in the context of QED and QCD),

$$\partial_\mu B^\mu = \frac{1}{3} \cdot 3_{\text{colors}} \cdot 3_{\text{generations}} \cdot \frac{g^2}{32\pi^2} \epsilon^{\mu\nu\lambda\rho} F_{\mu\nu}^a F_{\lambda\rho}^a ,$$

where $F_{\mu\nu}^a$ and g are the field strength of the $SU(2)_W$ gauge field and the $SU(2)_W$ gauge coupling, respectively. Likewise, each leptonic current ($n = e, \mu, \tau$) is anomalous,

$$\partial_\mu L_n^\mu = \frac{g^2}{32\pi^2} \cdot \epsilon^{\mu\nu\lambda\rho} F_{\mu\nu}^a F_{\lambda\rho}^a .$$

A non-trivial fact is that there exist large field fluctuations, $F_{\mu\nu}^a(\mathbf{x}, t) \propto g^{-1}$ which have

$$Q \equiv \int d^3x dt \frac{g^2}{32\pi^2} \cdot \epsilon^{\mu\nu\lambda\rho} F_{\mu\nu}^a F_{\lambda\rho}^a \neq 0 .$$

Furthermore, for any such fluctuation the value of Q is integer. Suppose now that a fluctuation with non-vanishing Q has occurred. Then the baryon numbers in the end and beginning of the process are different,

$$B_{fin} - B_{in} = \int d^3x dt \partial_\mu B^\mu = 3Q . \quad (44)$$

Likewise

$$L_{n, fin} - L_{n, in} = Q . \quad (45)$$

This explains the selection rule mentioned above: B is violated, $(B - L)$ is not.

At zero temperature, the large field fluctuations that induce baryon and lepton number violation are vacuum fluctuations, called instantons, which to a certain extent are similar to virtual fields that emerge and disappear in vacuum of quantum field theory at the perturbative level. The peculiarity is that instantons are *large* field fluctuations. The latter property results in the exponential suppression of the probability of their emergence, and hence the rate of baryon number violating processes, by a factor

$$e^{-\frac{16\pi^2}{g^2}} \sim 10^{-165} .$$

On the other hand, at high temperatures there are large *thermal* fluctuations (“sphalerons”) whose rate is not necessarily small. And, indeed, B -violation in the early Universe is rapid as compared to the cosmological expansion at sufficiently high temperatures, provided that

$$\langle \phi \rangle_T < T , \quad (46)$$

where $\langle \phi \rangle_T$ is the Higgs expectation value at temperature T .

One may wonder how baryon number may be not conserved even though there are no baryon number violating terms in the Lagrangian of the Standard Model. To understand what is going on, let us consider a massless *left handed* fermion field in the background of the $SU(2)$ gauge field $\mathbf{A}(\mathbf{x}, t)$, which depends on space-time coordinates in a non-trivial way. As a technicality, we set the temporal component of the gauge field equal to zero, $A_0 = 0$, by the choice of gauge. One way to understand the behavior of the fermion field in the gauge field background is to study the system of eigenvalues of the Dirac Hamiltonian $\{\omega(t)\}$. The Hamiltonian is defined in the standard way

$$H_{Dirac}(t) = i\alpha^i (\partial_i - igA_i(\mathbf{x}, t)) \frac{1 - \gamma_5}{2} ,$$

where $\alpha^i = \gamma^0 \gamma^i$, so that the Dirac equation has the Schrödinger form,

$$i \frac{\partial \psi}{\partial t} = H_{Dirac} \psi .$$

We are going to discuss the eigenvalues $\omega_n(t)$ of the operator $H_{Dirac}(t)$, treating t as a parameter. These eigenvalues are found from

$$H_{Dirac}(t)\psi_n = \omega_n(t)\psi_n .$$

At $\mathbf{A} = 0$ the system of levels is shown schematically in Fig. 3. Importantly, there are both positive- and negative-energy levels. According to Dirac, the lowest energy state (Dirac vacuum) has all negative energy levels occupied, and all positive energy levels empty. Occupied positive energy levels (three of them in Fig. 3) correspond to real fermions, while empty negative energy levels describe antifermions (one in Fig. 3). Fermion-antifermion annihilation in this picture is a jump of a fermion from a positive energy level to an unoccupied negative energy level.

As a side remark, this original Dirac picture is, in fact, equivalent to the more conventional (by now) procedure of the quantization of fermion field, which does not make use of the notion of negative energy levels. The discussion that follows can be translated into the conventional language; however, the original Dirac picture turned out to be a lot more transparent in our context. This is a nice example of the complementarity of various approaches in quantum field theory.

Let us proceed with the discussion of the fermion energy levels in gauge field backgrounds. In weak background fields, the energy levels depend on time (move), but nothing dramatic happens. For adiabatically varying background fields, the fermions merely sit on their levels, while fast changing fields generically give rise to jumps from, say, negative- to positive-energy levels, that is, creation of fermion-antifermion pairs. Needless to say, fermion number ($N_f - N_{\bar{f}}$) is conserved.

The situation is entirely different for the background fields with non-zero Q . The levels of left-handed fermions move as shown in the left panel of Fig. 4. Some levels necessarily cross zero, and the net number of levels crossing zero from below equals Q . This means that the number of left-handed fermions is not conserved: for adiabatically varying gauge field $\mathbf{A}(\mathbf{x}, t)$ the motion of levels shown in the left panel of Fig. 4 corresponds to the case in which the initial state of the fermionic system is vacuum (no fermions at all) whereas the final state contains Q real fermions (two in the particular case shown). If the evolution of the gauge field is not adiabatic, the result for the fermion number non-conservation is the same: there may be jumps from negative energy levels to positive energy levels, or vice versa. These correspond to creation or annihilation of fermion-antifermion pairs, but the net change of the fermion number (number of fermions minus number of antifermions) remains equal to Q . Importantly, the initial and final field configurations of the gauge field may be trivial, $\mathbf{A} = 0$ (up to gauge transformation), so that fermion number non-conservation may occur due to a fluctuation that begins and ends in the gauge field vacuum⁶. This is precisely an instanton-like vacuum fluctuation. At finite temperatures, processes of this type occur due to thermal fluctuations, sphalerons.

⁶A subtlety here is that in four-dimensional gauge theories, this is impossible for *Abelian* gauge fields, so fermion number non-conservation is inherent in *non-Abelian* gauge theories only.

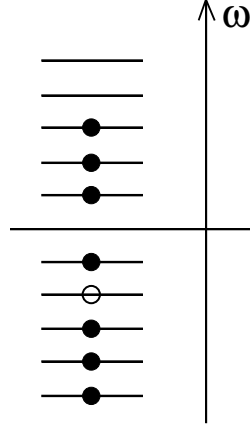


Fig. 3: Fermion energy levels at zero background gauge field.

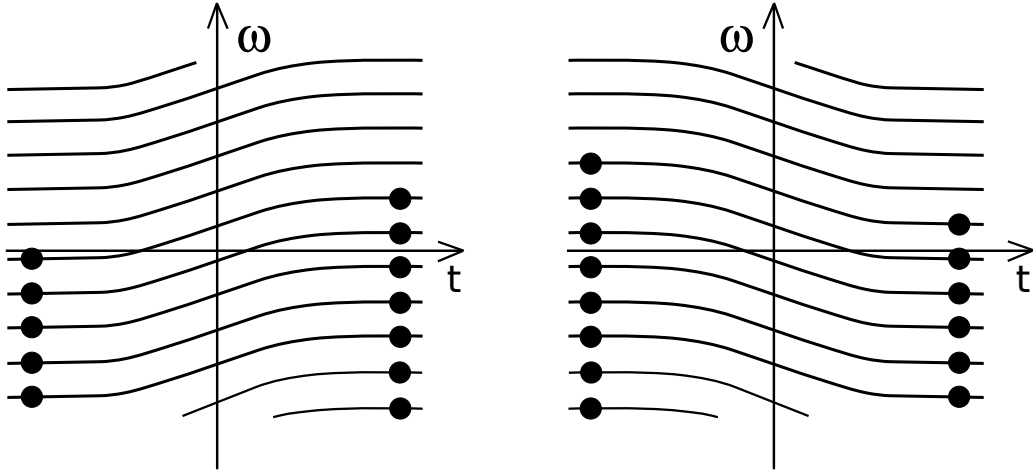


Fig. 4: Motion of fermion levels in background gauge fields with non-vanishing Q (shown is the case $Q = 2$). Left panel: left-handed fermions. Right panel: right-handed fermions.

If the same gauge field interacts also with right-handed fermions, the motion of the levels of the latter is opposite to that of left-handed fermions. This is shown in the right panel of Fig. 4. The change in the number of right-handed fermions is equal to $(-Q)$. So, if the gauge interaction is vector-like, the total fermion number $N_{left} + N_{right}$ is conserved, while chirality $N_{left} - N_{right}$ is violated even for massless fermions. This explains why there is no baryon number violation in QCD. On the other hand, non-perturbative violation of chirality in QCD in the limit of massless quarks has non-trivial consequences, which are indeed confirmed by phenomenology. In this sense anomalous non-conservation of fermion quantum numbers is an experimentally established fact.

In electroweak theory, right-handed fermions do not interact with $SU(2)_W$ gauge field, while left-handed fermions do. Therefore, fermion number is not conserved. Since fermions of each $SU(2)_W$ -doublet interact with the $SU(2)_W$ gauge bosons (essentially W and Z) in one and the same way, they are equally created in a process involving a gauge field fluctuation with non-zero Q . This again leads to the relations (44) and (45), i.e., to the selection rules $\Delta B = \Delta L$, $\Delta L_e = \Delta L_\mu = \Delta L_\tau$.

It is tempting to use this mechanism of baryon number non-conservation for explaining the baryon asymmetry of the Universe. There are two problems, however. One is that CP-violation in the Standard Model is too weak: the CKM mechanism alone is insufficient to generate the realistic value of the baryon

asymmetry. Hence, one needs extra sources of CP-violation. Another problem has to do with departure from thermal equilibrium that is necessary for the generation of the baryon asymmetry. At temperatures well above 100 GeV electroweak symmetry is restored, the expectation value of the Higgs field ϕ is zero⁷, the relation (46) is valid, and the baryon number non-conservation is rapid as compared to the cosmological expansion. At temperatures of order 100 GeV the relation (46) may be violated, but the Universe expands very slowly: the cosmological time scale at these temperatures is

$$H^{-1} = \frac{M_{Pl}^*}{T^2} \sim 10^{-10} \text{ s} , \quad (47)$$

which is very large by the electroweak physics standards. The only way in which strong departure from thermal equilibrium at these temperatures may occur is through the first order phase transition.

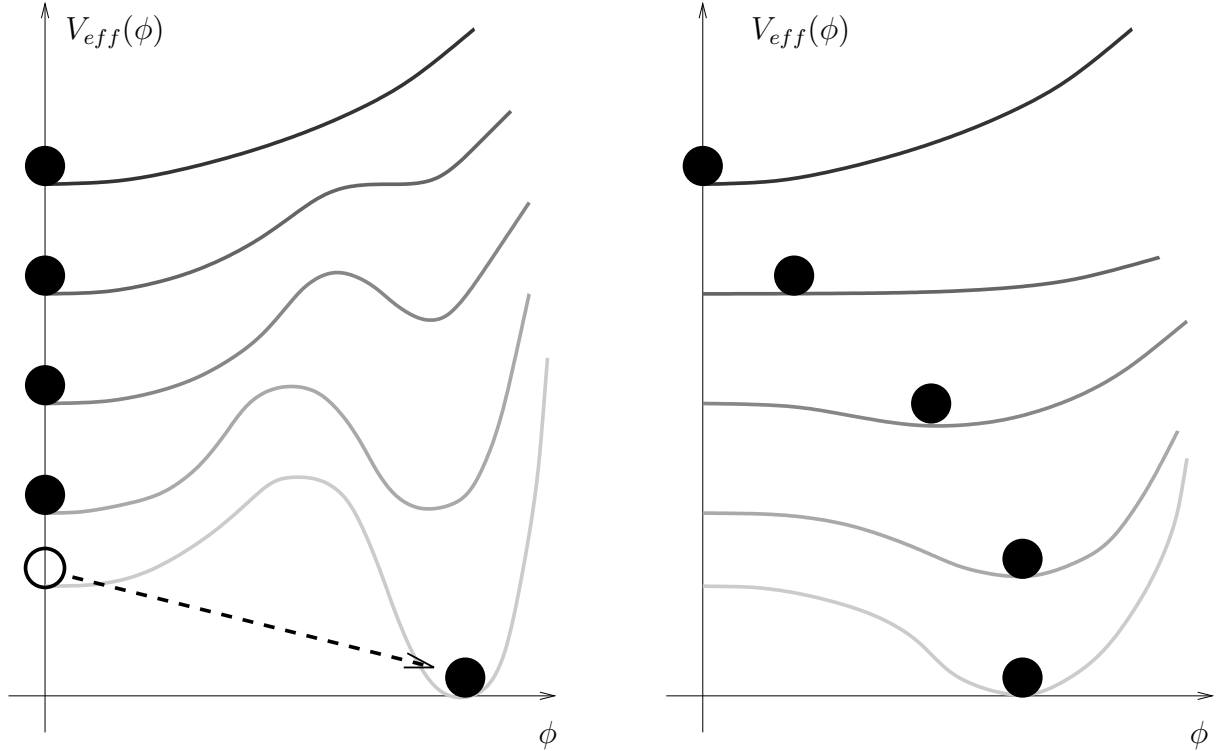


Fig. 5: Effective potential as function of ϕ at different temperatures. Left: first order phase transition. Right: second order phase transition. Upper curves correspond to higher temperatures.

The property that at temperatures well above 100 GeV the expectation value of the Higgs field is zero, while it is non-zero in vacuo, suggests that there may be a phase transition from the phase with $\langle\phi\rangle = 0$ to the phase with $\langle\phi\rangle \neq 0$. The situation is pretty subtle here, as ϕ is not gauge invariant, and hence cannot serve as an order parameter, so the notion of phases with $\langle\phi\rangle = 0$ and $\langle\phi\rangle \neq 0$ is vague. In fact, neither electroweak theory nor most of its extensions have a gauge-invariant order parameter, so there is no real distinction between these “phases”. This situation is similar to that in liquid-vapor system, which does not have an order parameter and may or may not experience vapor-liquid phase transition as temperature decreases, depending on other parameters characterizing this system, e.g., pressure. In the Standard Model the role of such a parameter is played by the Higgs self-coupling λ or, in other words, the Higgs boson mass.

Continuing to use somewhat sloppy terminology, we observe that the interesting case for us is the first order phase transition. In this case the effective potential (free energy density as function of

⁷There are subtleties at this point, see below.

ϕ) behaves as shown in the left panel of Fig. 5. At high temperatures, there exists one minimum of V_{eff} at $\phi = 0$, and the expectation value of the Higgs field is zero. As the temperature decreases, another minimum appears at finite ϕ , and then becomes lower than the minimum at $\phi = 0$. However, the probability of the transition from the phase $\phi = 0$ to the phase $\phi \neq 0$ is very small for some time, so the system gets overcooled. The transition occurs when the temperature becomes sufficiently low, as shown schematically by an arrow in Fig. 5. This is to be contrasted to the case, e.g., of the second order phase transition with the behavior of the effective potential shown in the right panel of Fig. 5. In the latter case, the field slowly evolves, as the temperature decreases, from zero to non-zero vacuum value, and the system remains very close to the thermal equilibrium at all times.

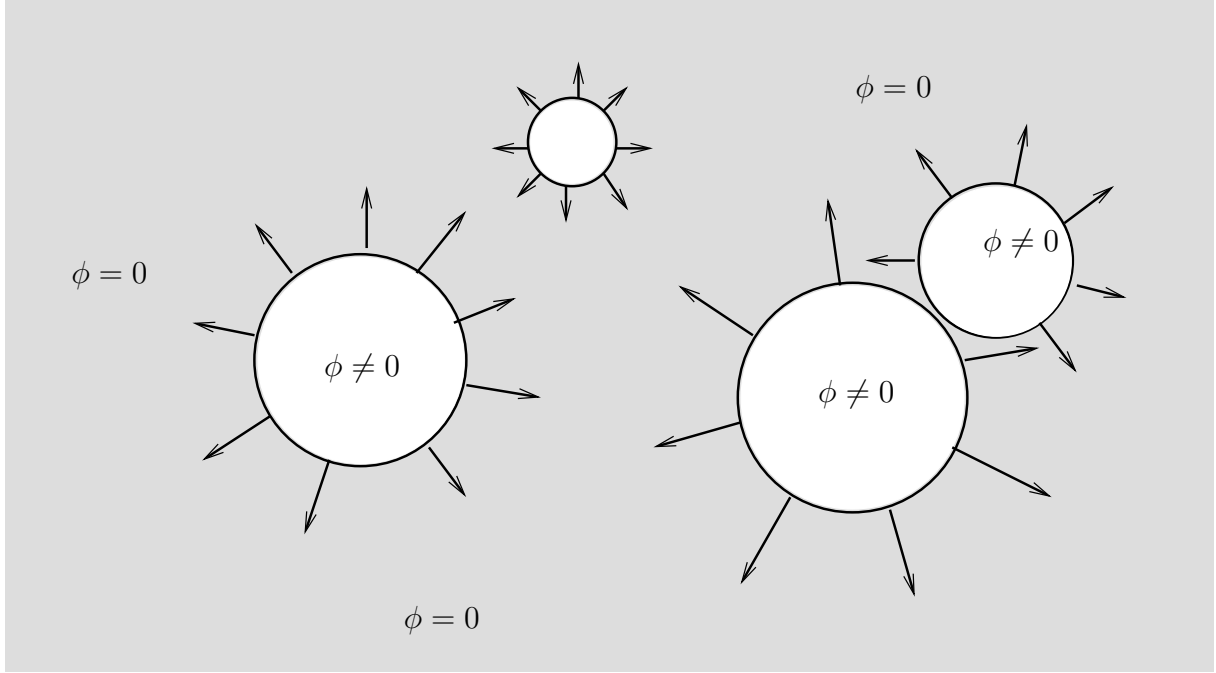


Fig. 6: First order phase transition: boiling Universe.

The first order phase transition occurs via spontaneous creation of bubbles of the new phase inside the old phase. These bubbles then grow, their walls eventually collide, and the new phase finally occupies entire space. The Universe boils, as shown schematically in Fig. 6. In the cosmological context, this process happens when the bubble nucleation rate per Hubble time per Hubble volume is of order 1, $\Gamma_{nuc} \sim H^{-4}$. The velocity of the bubble wall in the relativistic cosmic plasma is roughly of the order of the speed of light (in fact, it is somewhat smaller, from $0.1 c$ to $0.01 c$), simply because there are no relevant dimensionless parameters characterizing the system. Hence, the bubbles grow large before their walls collide: their size at collision is roughly of order of the Hubble size. While at nucleation the bubble is microscopic — its size is dictated by the electroweak scale and is roughly of order $(100 \text{ GeV})^{-1} \sim 10^{-16} \text{ cm}$ — its size at collision of walls is macroscopic, $H^{-1} \sim \text{a few cm}$, as follows from Eq. (47). Clearly, boiling is a highly inequilibrium process, and one may hope that the baryon asymmetry may be generated at that time. And, indeed, there exist mechanisms of the generation of the baryon asymmetry, which have to do with interactions of quarks and leptons with moving bubble walls. The value of the resulting baryon asymmetry may well be of order 10^{-10} , as required by observations, provided that there is enough CP-violation in the theory.

A necessary condition for the electroweak generation of the baryon asymmetry is that the inequality (46) must be violated *just after* the phase transition. Indeed, in the opposite case the electroweak baryon number violating processes are fast after the transition, and the baryon asymmetry, generated

during the transition, is washed out afterwards. Hence, the phase transition must be of strong enough first order. This is *not* the case in the Standard Model. To see why this is so, and to get an idea in which extensions of the Standard Model the phase transition may be of strong enough first order, let us consider the effective potential in some detail. At zero temperature, the Higgs potential has the standard form,

$$V(\phi) = -\frac{m^2}{2}|\phi|^2 + \frac{\lambda}{4}|\phi|^4.$$

Here

$$|\phi| \equiv (\phi^\dagger \phi)^{1/2} \quad (48)$$

is the length of the Higgs doublet ϕ , $m^2 = \lambda v^2$ and $v = 247 \text{ GeV}$ is the Higgs expectation value in vacuo. The Higgs boson mass is related to the latter as follows,

$$m_H = \sqrt{2\lambda}v. \quad (49)$$

Now, to the leading order of perturbation theory, the finite temperature effects modify the effective potential into

$$V_{eff}(\phi, T) = \frac{\alpha}{2}|\phi|^2 - \frac{\beta}{3}T|\phi|^3 + \frac{\lambda}{4}|\phi|^4, \quad (50)$$

with $\alpha(T) = -m^2 + \hat{g}^2 T^2$, where $\hat{g}^2$ is a positive linear combination of squares of coupling constants of all fields to the Higgs field (in the Standard Model, a linear combination of g^2 , g'^2 and y_i^2 , where g and g' are gauge couplings and y_i are Yukawa couplings), while β is a positive linear combination of cubes of coupling constants of all *bosonic* fields to the Higgs field. In the Standard Model, β is a linear combination of g^3 and g'^3 , i.e., a linear combination of M_W^3/v^3 and M_Z^3/v^3 ,

$$\beta = \frac{1}{2\pi} \frac{2M_W^3 + M_Z^3}{v^3}. \quad (51)$$

The cubic term in Eq. (50) is rather peculiar: in view of Eq. (48) it is not analytic in the original Higgs field ϕ . Yet this term is crucial for the first order phase transition: for $\beta = 0$ the phase transition would be of the second order. The origin of the non-analytic cubic term can be traced back to the enhancement of the Bose–Einstein thermal distribution at low momenta, $p, m \ll T$,

$$f_{Bose}(p) = \frac{1}{e^{\frac{\sqrt{p^2+m_a^2}}{T}} - 1} \simeq \frac{T}{\sqrt{p^2+m_a^2}},$$

where $m_a \simeq g_a|\phi|$ is the mass of the boson a that is generated due to the non-vanishing Higgs field, and g_a is the coupling constant of the field a to the Higgs field. Clearly, at $p \ll g_a|\phi|$ the distribution function is non-analytic in ϕ ,

$$f_{Bose}(p) \simeq \frac{T}{g_a|\phi|}.$$

It is this non-analyticity that gives rise to the non-analytic cubic term in the effective potential. Importantly, the Fermi–Dirac distribution,

$$f_{Fermi}(p) = \frac{1}{e^{\frac{\sqrt{p^2+m_a^2}}{T}} + 1},$$

is analytic in m_a^2 , and hence $\phi^\dagger \phi$, so fermions do not contribute to the cubic term.

With the cubic term in the effective potential, the phase transition is indeed of the first order: at high temperatures the coefficient α is positive and large, and there is one minimum of the effective

potential at $\phi = 0$, while for α small but still positive there are two minima. The phase transition occurs at $\alpha \approx 0$; at that moment

$$V_{eff}(\phi, T) \approx -\frac{\beta T}{3}|\phi|^3 + \frac{\lambda}{4}|\phi|^4.$$

We find from this expression that immediately after the phase transition the minimum of V_{eff} is at

$$\phi \simeq \frac{\beta T}{\lambda}.$$

Hence, the necessary condition for successful electroweak baryogenesis, $\phi > T$, translates into

$$\beta > \lambda. \quad (52)$$

According to Eq. (49), λ is proportional to m_H^2 , whereas in the Standard Model β is proportional to $(2M_W^3 + M_Z^3)$. Therefore, the relation (52) holds for small Higgs boson masses only; in the Standard Model one makes use of Eqs. (49) and (51) and finds that this happens for $m_H < 50$ GeV, which is ruled out⁸.

This discussion indicates a possible way to make the electroweak phase transition strong. What one needs is the existence of new bosonic fields that have large enough couplings to the Higgs field(s), and hence provide large contributions to β . To have an effect on the dynamics of the transition, the new bosons must be present in the cosmic plasma at the transition temperature, $T \sim 100$ GeV, so their masses should not be too high, $M \lesssim 300$ GeV. In supersymmetric extensions of the Standard Model, the natural candidate for long time has been stop (superpartner of top-quark) whose Yukawa coupling to the Higgs field is the same as that of top, that is, large. The light stop scenario for electroweak baryogenesis would indeed work, as has been shown by the detailed analysis in Ref. [16].

Yet another issue is CP-violation, which has to be strong enough for successful electroweak baryogenesis. As the asymmetry is generated in the interactions of quarks and leptons (and their superpartners in supersymmetric extensions) with the bubble walls, CP-violation must occur at the walls. Recall now that the walls are made of the Higgs field(s). This points towards the necessity of CP-violation in the Higgs sector, which may only be the case in a theory with more than one Higgs fields.

To summarize, electroweak baryogenesis requires a considerable extension of the Standard Model, with masses of new particles in the range 100 – 300 GeV. Hence, this mechanism will definitely be ruled out or confirmed by the LHC. We stress, however, that electroweak baryogenesis is not the only option at all: an elegant and well motivated competitor is leptogenesis [17]; several other mechanisms have been proposed that may be responsible for the baryon asymmetry of the Universe.

5 Dark energy

Dark energy, the famous “substance”, does not clump, unlike dark matter. It gives rise to the accelerated expansion of the Universe. As we see from Eq. (32), the Universe with constant energy density should expand exponentially; if the energy density is almost constant, the expansion is almost exponential. Let us make use of the first law of thermodynamics, which for the adiabatic expansion reads

$$dE = -pdV,$$

and apply it to comoving volume, $E = \rho V$, $V = a^3$. We obtain for dark energy

$$d\rho_\Lambda = -3\frac{da}{a}(\rho_\Lambda + p_\Lambda),$$

⁸In fact, in the Standard Model with $m_H > 114$ GeV, there is no phase transition at all; the electroweak transition is smooth crossover instead. The latter fact is not visible from the expression (50), but that expression is the lowest order perturbative result, while the perturbation theory is not applicable for describing the transition in the Standard Model with large m_H .

or

$$\frac{d\rho_\Lambda}{\rho_\Lambda} = -3\frac{da}{a}(1+w) ,$$

where we introduced the equation of state parameter w such that

$$p_\Lambda = w\rho_\Lambda .$$

Thus, (almost) time-independent dark energy density corresponds to $w \approx -1$, i.e., effective pressure of dark energy is negative. We emphasize that pressure is by definition a spatial component of the energy-momentum tensor, which in the homogeneous and isotropic situation has the general form

$$T_{\mu\nu} = \text{diag} (\rho, p, p, p) .$$

Dark energy density does not depend on time at all, if $p_\Lambda = -\rho_\Lambda$, i.e.,

$$T_{\mu\nu} = \rho_\Lambda \eta_{\mu\nu} ,$$

where $\eta_{\mu\nu}$ is the Minkowski tensor. This is characteristic of vacuum, whose energy-momentum tensor must be Lorentz-covariant. Observationally, w is close to -1 to reasonably good precision. The most accurate determination, which, however, does not include systematic errors in supernovae data and possible time-dependence of w , is [9]

$$w = -0.98 \pm 0.05 . \quad (53)$$

So, the dark energy density is almost time-independent, indeed.

The problem with dark energy is that its present value is extremely small by particle physics standards,

$$\rho_{DE} \approx 4 \text{ GeV/m}^3 = (2 \times 10^{-3} \text{ eV})^4 .$$

In fact, there are two hard problems. One is that particle physics scales are much larger than the scale relevant to the dark energy density, so the dark energy density is zero to an excellent approximation. Another is that it is non-zero nevertheless, and one has to understand its energy scale. To quantify the first problem, we recall the known scales of particle physics and gravity,

$$\begin{array}{ll} \text{Strong interactions :} & \Lambda_{QCD} \sim 1 \text{ GeV} , \\ \text{Electroweak :} & M_W \sim 100 \text{ GeV} , \\ \text{Gravitational :} & M_{pl} \sim 10^{19} \text{ GeV} . \end{array}$$

In principle, vacuum should contribute to ρ_Λ , and there is absolutely no reason for vacuum to be as light as it is. The discrepancy here is huge, as one sees from the above numbers.

To elaborate on this point, let us note that the action of gravity plus, say, the Standard Model has the general form

$$S = S_{EH} + S_{SM} - \rho_{\Lambda,0} \int \sqrt{-g} d^4x ,$$

where $S_{EH} = -(16\pi G_N)^{-1} \int R \sqrt{-g} d^4x$ is the Einstein–Hilbert action of General Relativity, S_{SM} is the action of the Standard Model and $\rho_{\Lambda,0}$ is the bare cosmological constant. In order that the vacuum energy density be almost zero, one needs fantastic cancellations between the contributions of the Standard Model fields into the vacuum energy density, on the one hand, and $\rho_{\Lambda,0}$ on the other. For example, we know that QCD has a complicated vacuum structure, and one would expect that the energy density of QCD combined with $\rho_{\Lambda,0}$ should be of order $(1 \text{ GeV})^4$. Nevertheless, it is not, so at least for QCD, one needs a cancellation on the order of 10^{-44} . If one goes further and considers other interactions, the numbers get even worse.

What are the hints from this “first” cosmological constant problem? There are several options, though not many. One is that the Universe could have a very long prehistory. Extremely long. This option has to do with relaxation mechanisms. Suppose that the original vacuum energy density is indeed large, say, comparable to the particle physics scales. Then there must be a mechanism which can relax this value down to an acceptably small number. It is easy to convince oneself that this relaxation could not happen in the history of the Universe we know of. Instead, the Universe should have a very long prehistory during which this relaxation process might occur. At that prehistoric time, the vacuum in the Universe must have been exactly the same as our vacuum, so the Universe in its prehistory must have been exactly like ours, or almost exactly like ours. Only in that case could a relaxation mechanism work. There are concrete scenarios of this sort [18]. However, at the moment it seems that these scenarios are hardly testable, since this is prehistory.

Another possible hint is towards anthropic selection. The argument that goes back to Weinberg and Linde [19, 20] is that if the cosmological constant were larger, say, by a factor of 100, we simply would not exist: the stars would not have formed because of the fast expansion of the Universe. So, the vacuum energy density may be selected anthropically. The picture is that the Universe may be much, much larger than what we can see, and different large regions of the Universe may have different properties. In particular, vacuum energy density may be different in different regions. Now, we are somewhere in the place where one can live. All the rest is empty of human beings, because there the parameters such as vacuum energy density are not suitable for their existence. This is disappointing for a theorist, as this point of view allows for arbitrary tuning of fundamental parameters. It is hard to disprove this option, on the other hand. We do exist, and this is an experimental fact. The anthropic viewpoint may, though hopefully will not, get more support from the LHC, if no or insufficient new physics is found there. Indeed, another candidate for an environmental quantity is the electroweak scale.

Let us recall in this regard the gauge hierarchy problem: the electroweak scale $M_W \sim 100$ GeV is much lower than the natural scale in gravitational physics, the Planck mass, $M_{Pl} \sim 10^{19}$ GeV. The electroweak scale in the Standard Model is unprotected from large contributions due to high energy physics, and in this sense it is very similar to the cosmological constant. There are various anthropic arguments showing that the electroweak scale must be small. A simple example is that if one makes it larger without touching other parameters, then quarks would be too heavy. Neutron would be the lightest baryon, and proton would be unstable. There would be no stable hydrogen, and that is presumably inconsistent with our existence. Hence, one of the “solutions” to the gauge hierarchy problem is anthropic.

An interesting part of the story is that unlike the cosmological constant, there are natural ways to make the electroweak scale small and render it small in extensions of the Standard Model, like low energy supersymmetry. All these extensions require new physics at TeV energies. So we are in a situation where the experiment has to say its word. If it says that none of these extensions is there in Nature, then we will have to take the anthropic viewpoint much more seriously than before.

Turning to the “second” cosmological constant problem, we note that the scale 10^{-3} eV may be associated with some new light field(s), rather than with vacuum. This implies, in general, that ρ_Λ depends on time, i.e., $w \neq -1$ and w may well depend on time itself. “Normal” field (called quintessence in this context) has $w > -1$, but there are examples (rather contrived) of fields with $w < -1$ (called phantom fields). Current data are compatible with time-independent w equal to -1 , but their precision is not particularly high. We conclude that future cosmological observations may shed new light on the field content of fundamental theory.

6 Cosmological perturbations and the very early Universe

With Big Bang nucleosynthesis theory and observations, we are confident of the theory of the early Universe at temperatures up to $T \simeq 1$ MeV, that corresponds to age of $t \simeq 1$ second. With the LHC, we hope to be able to go up to temperatures $T \sim 100$ GeV and age $t \sim 10^{-10}$ second. The question is: are

we going to have a handle on even earlier epoch?

The key issue in this regard is cosmological perturbations. These are inhomogeneities in the energy density and associated gravitational potentials, in the first place. This type of inhomogeneities is called scalar perturbations, as they are described by 3-scalars. There may exist perturbations of another type, called tensor; these are primordial gravity waves. We will mostly concentrate on scalar perturbations, since they are observed; tensor perturbations are important too, and we comment on them later on. It is worth pointing out that perturbations of the present size below ten Megaparsec have large amplitudes today and are non-linear, but in the past their amplitudes were small, and they can be described within the linearized theory. Indeed, CMB temperature anisotropy tells us that the perturbations at recombination epoch were roughly at the level

$$\delta \equiv \frac{\delta\rho}{\rho} = 10^{-4} - 10^{-5} .$$

Thus, the linearized theory works very well before recombination and somewhat later.

Properties of scalar perturbations are measured in various ways. Perturbations of large spatial scales leave their imprint in CMB temperature anisotropy and polarization, so we have very detailed knowledge of them. Shorter wavelength perturbations are studied by analysing distributions of galaxies and quasars at present and in relatively near past. There are several other methods, some of which can probe even shorter wavelengths. As we discuss in more detail below, scalar perturbations in the linear regime are actually Gaussian random field, and the first thing to measure is its power spectrum. Overall, independent methods give consistent results, see Fig. 7.

Cosmic medium in our Universe has several components that interact only gravitationally: baryons, photons, neutrinos, dark matter. Hence, there may be and, in fact, there are perturbations in each of these components. As we pointed out in the beginning of Section 3, electromagnetic interactions between baryons, electrons and photons were strong before recombination, so these species made single fluid, and it is appropriate to talk about perturbations in this fluid. After recombination, baryons and photons evolved independently.

The main point of this part of lectures is that by analysing the density perturbations, we have already learned a number of very important things. To appreciate what they are, it is instructive to consider first the baryon-electron-photon fluid before recombination. Perturbations in this fluid are nothing but sound waves; they obey a wave equation. So, let us turn to the wave equation in the expanding Universe.

6.1 Wave equation in expanding Universe. Subhorizon and superhorizon regimes.

The actual system of equations for density perturbations in the baryon-electron-photon fluid and associated gravitational potentials is fairly cumbersome. So, let us simplify things. Instead of writing and then solving the equations for sound waves, let us consider a toy example, the case of massless scalar field. The general properties of density perturbations are similar to this case, although there are a few places in which they differ; we comment on the differences in due course.

The action for the massless scalar field is

$$S = \frac{1}{2} \int d^4x \sqrt{-g} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi = \int d^3x dt a^3 \left[\frac{1}{2} \dot{\phi}^2 - \frac{1}{2a^2} (\partial_i \phi)^2 \right] ,$$

where we specified to FLRW metric in the second expression. The field equation thus reads:

$$-\frac{d}{dt}(a^3 \dot{\phi}) + a \partial_i \partial_i \phi = 0,$$

i.e.,

$$\ddot{\phi} + 3H\dot{\phi} - \frac{1}{a^2} \partial_i^2 \phi = 0 , \tag{54}$$

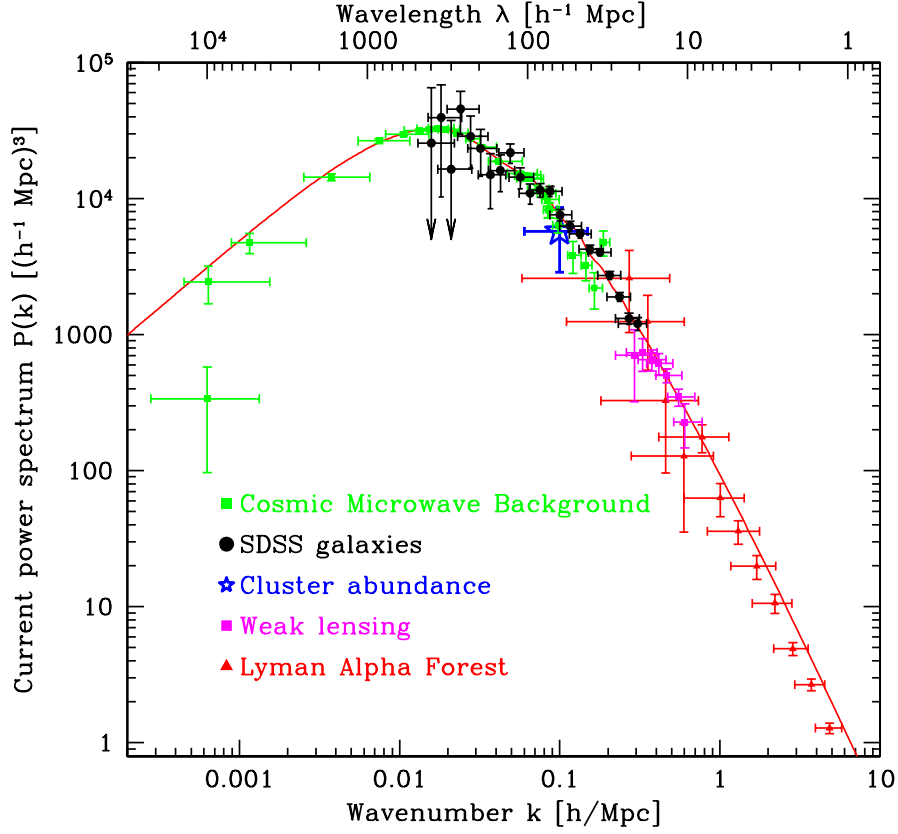


Fig. 7: Power spectrum of density perturbations, measured by various methods and translated to the present epoch by using the linearized theory [21]. k is the present wavenumber, and $h \approx 0.7$ is the dimensionless Hubble parameter at the present epoch.

where $H \equiv \dot{a}/a$ is again the Hubble parameter. This equation is linear in ϕ and homogeneous in space, so it is natural to represent ϕ in terms of the Fourier harmonics,

$$\phi(\mathbf{x}, t) = \int e^{i\mathbf{k}\mathbf{x}} \phi_{\mathbf{k}}(t) d^3k.$$

Clearly, the value of k for a given Fourier mode is constant in time. However, k is not the physical wavenumber (physical momentum), since x is not the physical distance. k is called conformal momentum, while physical momentum equals $q \equiv 2\pi/\lambda = 2\pi/(a(t)\Delta x) = k/a(t)$. Δx here is time-independent comoving wavelength of perturbation, and λ is the physical wavelength; the latter grows due to the expansion of the Universe. Accordingly, as the Universe expands, the physical momentum of a given mode decreases (gets redshifted), $q(t) \propto a^{-1}(t)$. For a mode of given conformal momentum k , Eq. (54) gives:

$$\ddot{\phi} + 3H\dot{\phi} + \frac{k^2}{a^2}\phi = 0. \quad (55)$$

Besides the redshift of momentum, the cosmological expansion has the effect of inducing the second term, “Hubble friction”.

Equation (55) has two time-dependent parameters of the same dimension: k/a and H . Let us consider two limiting cases: $k/a \ll H$ and $k/a \gg H$. In cosmological models with conventional equation of state of the dominant component (e.g., matter-dominated or radiation-dominated Universe),

H^{-1} is of the order of the size of the cosmological horizon, see Section 2.6. So, the regime $k/a \ll H$ is the regime in which the physical wavelength $\lambda = 2\pi a/k$ is greater than the horizon size (this is called superhorizon regime), while for $k/a \gg H$ the physical wavelength is smaller than the horizon size (subhorizon regime). The time when the wavelength of the mode coincides with the horizon size is called horizon crossing. In what follows we denote this time by the symbol $\times$. Both at radiation- and matter-dominated epochs, the ratio $k/(aH)$ grows. Indeed, in the radiation-dominated epoch $a \propto \sqrt{t}$, while $H \propto t^{-1}$, so $k/(aH) \propto \sqrt{t}$. This means that every mode was at some early time superhorizon, and later on it becomes subhorizon, see Fig. 8. It is straightforward to see that for all cosmologically interesting wavelengths, horizon crossing occurs much later than 1 s after the Big Bang, i.e., at the time we are confident about. So, there is no guesswork at this point.

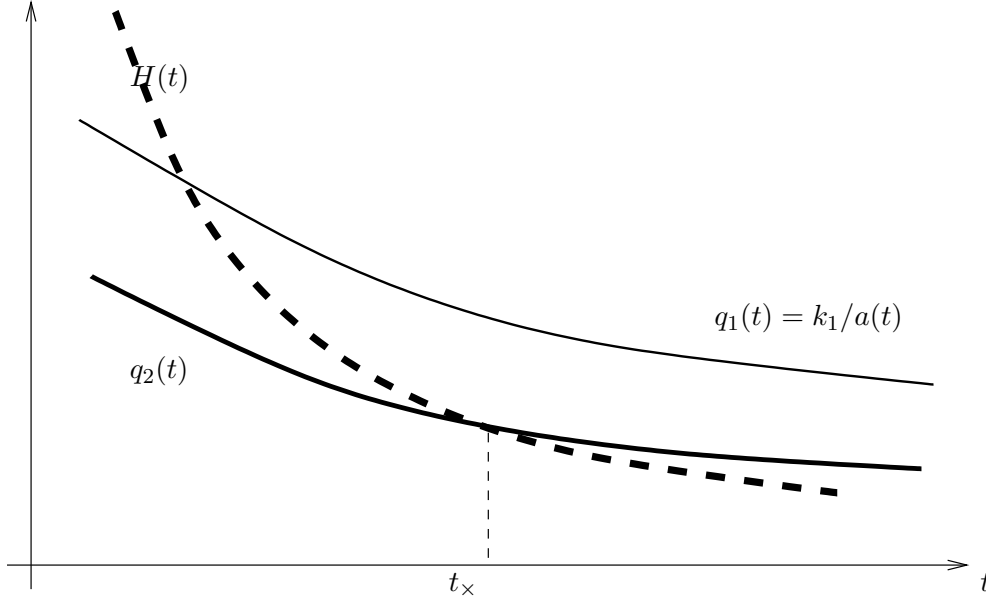


Fig. 8: Physical momenta (solid lines, $k_2 < k_1$) and Hubble parameter (dashed line) at radiation- and matter-dominated epochs. $t_\times$ is the horizon entry time.

Now we can address the question of the origin of density perturbations. By causality, any mechanism of their generation that operates at the radiation- and/or matter-dominated epoch, can only work after the horizon entry time $t_\times$. Indeed, no physical process can create a perturbation whose wavelength exceeds the size of an entire causally connected region. So, in that case the perturbation modes were never superhorizon. On the other hand, if modes were ever superhorizon, they have to exist already in the beginning of the hot epoch. Hence, in the latter situation one has to conclude that there existed another epoch *before* the hot stage: that was the epoch of the generation of primordial density perturbations.

Observational data, notably (but not only) on CMB temperature anisotropy and polarization, disentangle these two possibilities. They unambiguously show that density perturbations *were* superhorizon at radiation and matter domination!

To understand how this comes about, let us see what is special about a perturbation which was superhorizon at the hot stage. For a superhorizon mode, we can neglect the term $\phi \cdot k^2/a^2$ in Eq. (54). Then the field equation, e.g., in the radiation-dominated Universe ($a \propto t^{1/2}$, $H = 1/2t$), becomes

$$\ddot{\phi} + \frac{3}{2t} \dot{\phi} = 0. \quad (56)$$

The general solution to this equation is

$$\phi(t) = A + \frac{B}{\sqrt{t}}, \quad (57)$$

where A and B are constants. This behavior is generic for all cosmological perturbations at the hot stage: there is a constant mode (A in our case) and a mode that decays in time. If we extrapolate the decaying mode $B/\sqrt{t}$ back in time, we get very strong (infinite in the limit $t \rightarrow 0$) perturbation. For density perturbations (and also tensor perturbations) this means that this mode corresponds to strongly inhomogeneous early Universe. Therefore, the consistency of the cosmological model dictates that the decaying mode has to be absent for actual perturbations. Hence, for given $\mathbf{k}$, the solution is determined by a single parameter, the initial amplitude A of the mode $\phi_{\mathbf{k}}$.

After entering the subhorizon regime, the modes oscillate — these are the analogs of conventional sound waves. In the subhorizon regime one makes use of the WKB approximation to solve the complete equation

$$\ddot{\phi} + \frac{3}{2t}\dot{\phi} + \frac{k^2}{a^2(t)}\phi = 0. \quad (58)$$

The general solution in the WKB approximation reads

$$\phi(t) = \frac{A'}{a(t)} \cos \left(\int_0^t \frac{k}{a(t')} dt' + \psi_0 \right), \quad (59)$$

with the two constants being the amplitude A' and the phase ψ_0 . The amplitude A' of these oscillations is determined by the amplitude A of the superhorizon initial perturbation, while *the phase ψ_0 of these oscillations is uniquely determined by the condition of the absence of the decaying mode, $B = 0$.* Imposing this condition yields

$$\phi(t) = cA \frac{a_{\times}}{a(t)} \sin \left(\int_0^t \frac{k}{a(t')} dt' \right), \quad (60)$$

where the constant c is of order 1 and can be evaluated by solving the complete equation (58). The decreasing amplitude of oscillations $\phi(t) \propto 1/a(t)$ and the particular phase $\psi_0 = -\pi/2$ in Eq. (59) are peculiar properties of the wave equation (54), as well as the radiation-dominated cosmological expansion. However, *the fact that the phase of oscillations is uniquely determined by the requirement of the absence of the superhorizon decaying mode is generic.*

The perturbations in the baryon–photon medium before recombination — sound waves — behave in a rather similar way. Their evolution is as follows:

$$\delta_{\gamma} \equiv \frac{\delta\rho_{\gamma}}{\rho_{\gamma}} = \begin{cases} \text{const}, & \text{outside horizon,} \\ \text{const} \cdot \cos \left(\int_0^t v_s \frac{k}{a(t')} dt' \right), & \text{inside horizon,} \end{cases} \quad (61)$$

where $v_s \equiv \sqrt{dp/d\rho}$ is the sound speed. The baryon–photon medium before recombination is almost relativistic⁹, since $\rho_B < \rho_{\gamma}$. Therefore, $v_s \approx 1/\sqrt{3}$. Let us reiterate that the phase of the oscillating solution in (61) is uniquely defined.

6.2 Oscillations in CMB angular spectrum

CMB gives us the photographic picture of the Universe at recombination (photon last scattering), see Fig. 9. Waves of different momenta k are at different phases at recombination. At that epoch, oscillations in time in Eq. (61) show up as oscillations in momentum. This in turn gives rise to the observed oscillations in the CMB angular spectrum.

⁹This does not contradict the statement that the Universe is in matter-dominated regime at recombination. The dominant component at this stage is dark matter.

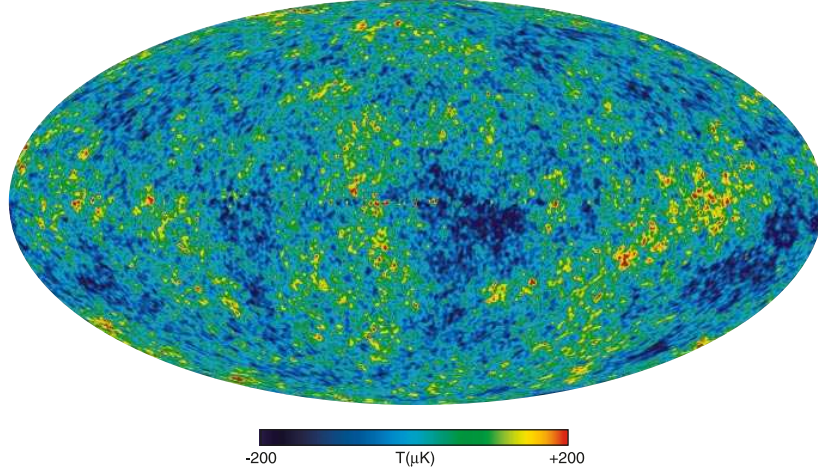


Fig. 9: CMB sky as seen by WMAP experiment.

In more detail, at the time of last scattering t_{rec} we have

$$\delta_\gamma \equiv \frac{\delta\rho_\gamma}{\rho_\gamma} = A(k) \cdot \cos\left(\int_0^{t_{rec}} v_s \frac{k}{a(t')} dt'\right) = A(k) \cdot \cos kr_s, \quad (62)$$

where $A(k)$ is linearly related to the initial amplitude of the superhorizon perturbation and is a non-oscillatory function of k , and

$$r_s = \int_0^{t_{rec}} v_s \frac{dt'}{a(t')}$$

is the comoving size of the sound horizon at recombination, while its physical size equals $a(t_{rec})r_s$. So, we see that the density perturbation at recombination indeed oscillates as a function of wavenumber. The period of this oscillation is determined by r_s , which is a straightforwardly calculable quantity.

Omitting details, the fluctuation of the CMB temperature is partially due to the density perturbation in the baryon-photon medium at recombination. The relevant place is the point where the photons last scatter before coming to us. This means that the temperature fluctuation of photons coming from the direction $\mathbf{n}$ in the sky is, to a reasonable accuracy,

$$\delta T(\mathbf{n}) \propto \delta_\gamma(\mathbf{x}_\mathbf{n}, \eta_{rec}) + \delta T_{smooth}(\mathbf{n}),$$

where $T_{smooth}(\mathbf{n})$ corresponds to the non-oscillatory part of the CMB angular spectrum, and

$$\mathbf{x}_\mathbf{n} = -\mathbf{n}(\eta_0 - \eta_{rec}).$$

Here the variable η is defined in (31), and η_0 is its present value, so that $(\eta_0 - \eta_{rec})$ is the coordinate distance to the sphere of photon last scattering, and $\mathbf{x}_\mathbf{n}$ is the coordinate of the place where the photons coming from the direction $\mathbf{n}$ scatter last time. $T_{smooth}(\mathbf{n})$ originates from the gravitational potential generated by the dark matter perturbation; dark matter has zero pressure at all times, so there are no sound waves in this component, and there are no oscillations at recombination as a function of momentum.

One expands the temperature variation on celestial sphere in spherical harmonics:

$$\delta T(\mathbf{n}) = \sum_{lm} a_{lm} Y_{lm}(\theta, \phi).$$

The multipole number l characterizes the temperature fluctuations at the angular scale $\Delta\theta = \pi/l$. The sound waves of momentum k are seen roughly at an angle $\Delta\theta = \Delta x/(\eta_0 - \eta_{rec})$, where $\Delta x = \pi/k$ is coordinate half-wavelength. Hence, there is the correspondence

$$l \longleftrightarrow k(\eta_0 - \eta_{rec}) .$$

Oscillations in momenta in (62) thus translate into oscillations in l , and these are indeed observed, see Fig. 10.

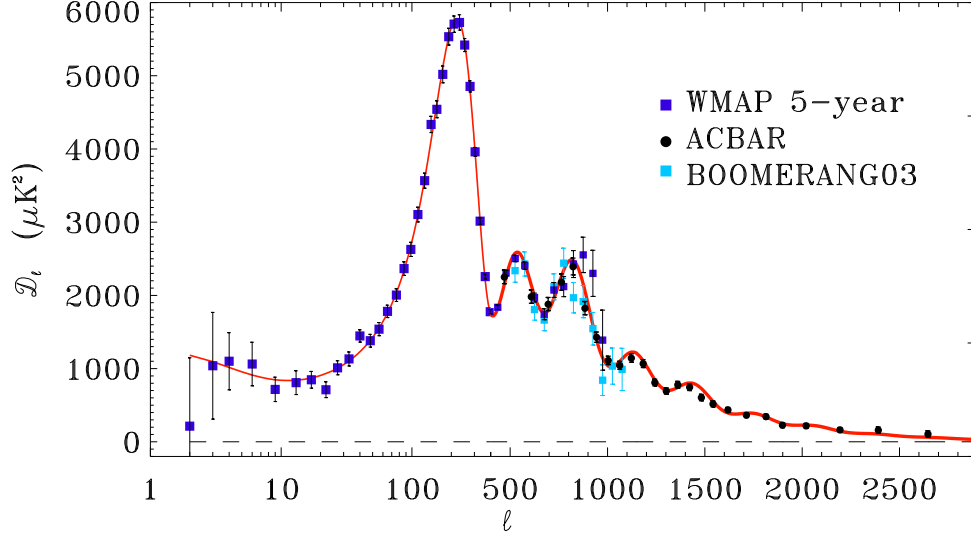


Fig. 10: The angular spectrum of the CMB temperature anisotropy [22]. The quantity in vertical axis is D_l defined in Eq. (65).

To understand what is shown in Fig. 10, we note that all observations today support the hypothesis that a_{lm} are independent Gaussian random variables. Gaussianity means that

$$P(a_{lm})da_{lm} = \frac{1}{\sqrt{2\pi C_l}} e^{-\frac{a_{lm}^2}{2C_l}} da_{lm}, \quad (63)$$

where $P(a_{lm})$ is the probability density for the random variable a_{lm} . For a hypothetical ensemble of Universes like ours, the average values of products of the coefficients a_{lm} would obey

$$\langle a_{lm} a_{l'm'} \rangle = C_l \delta_{ll'} \delta_{mm'}. \quad (64)$$

This gives the expression for the temperature fluctuation:

$$\langle [\delta T(\mathbf{n})]^2 \rangle = \sum_l \frac{2l+1}{4\pi} C_l \approx \int \frac{dl}{l} D_l ,$$

where

$$D_l = \frac{l(l+1)}{2\pi} C_l . \quad (65)$$

It is the latter quantity that is usually shown in plots, in particular, in Fig. 10. Note the unconventional scale on the horizontal axis, aimed at showing both small l region (large angular scales) and large l region.

The fact that the CMB angular spectrum has oscillatory behavior unambiguously tells us that *density perturbations were indeed superhorizon* at hot cosmological stage. If these perturbations were

generated by some causal mechanism after horizon entry, there would be no reason for the phase ψ_0 in (59) (better to say, in the analog of (59) for density perturbations) to take a very definite value. Instead, one would expect that this phase is a random function of $\mathbf{k}$, so there would be no oscillations in l in the CMB angular spectrum at all. This is indeed the case in concrete causal models aimed at generating the density perturbations at the hot stage, which make use, e.g., of topological defects (strings, textures, etc.), see Fig. 11.

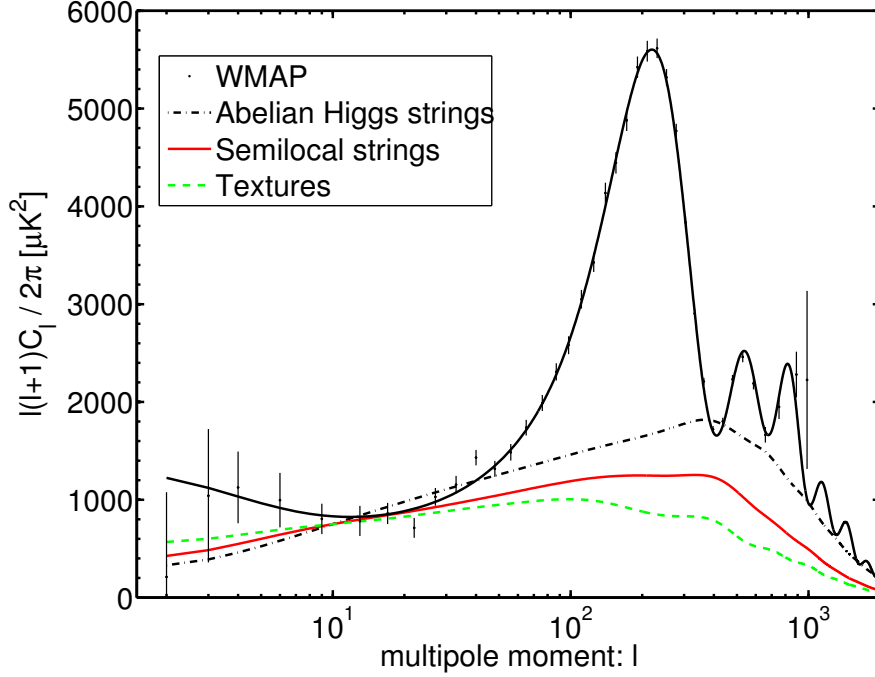


Fig. 11: The angular spectrum of the CMB temperature anisotropy in causal models that generate the density perturbations at the hot stage (non-oscillatory lines) versus data (sketched by oscillatory line) [23].

Another point to note is that the CMB measurements show that at recombination, there were density perturbations which were still superhorizon at that time. These correspond to low multipoles, $l \lesssim 50$. Perturbations of these wavelengths cannot be produced at the hot stage before recombination, and, indeed, causal mechanisms produce small power at low multipoles. This is also seen in Fig. 11.

6.3 Baryon acoustic oscillations

Another manifestation of the well defined phase of sound waves in baryon-photon medium before recombination is baryon acoustic oscillations. Right after recombination, baryons decouple from photons, the sound speed in the baryon component becomes essentially zero, and the spatial distribution of the baryon density freezes out. Since just before recombination baryons, together with photons, have energy distribution (61) which is oscillatory function of k , there is oscillatory component in the Fourier spectrum of the total matter distribution after recombination. This oscillatory component persists until today, and shows up as oscillations in the matter power spectrum $P(k)$. This is a small effect, since the dominant component at late times is dark matter,

$$\rho_M(\mathbf{k}) = \rho_{DM}(\mathbf{k}) + \rho_B(\mathbf{k}) ,$$

and only ρ_B oscillates as function of k

$$\delta\rho_B(\mathbf{k}) \approx \rho_B \delta_\gamma(\mathbf{k}) = \rho_B \cdot A(k) \cdot \cos kr_s \quad (66)$$

(as we already noticed, dark matter has zero pressure at all times, so there are no sound waves in this component). Nevertheless, this effect has been observed in large galaxy surveys, see Fig. 12.

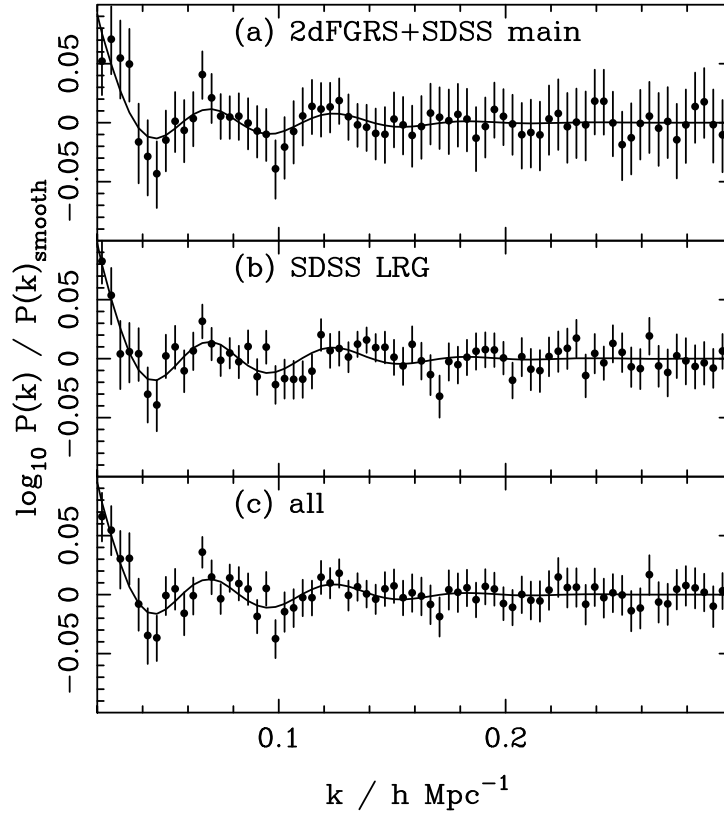


Fig. 12: Baryon acoustic oscillations in matter power spectrum detected in galaxy surveys [24].

There is a simple interpretation of the effect. As we discuss below, the overdensities in the baryon-photon medium and in the dark matter are at the same place before horizon entry (adiabatic mode). But before recombination the sound speed in baryon-photon plasma is of the order of the speed of light, while the sound speed in dark matter is basically zero. So, the overdensity in baryons generates an outgoing density wave after horizon crossing. This wave propagates until recombination, and then freezes out. On the other hand, the overdensity in the dark matter remains in its original place. The current distance from the overdensity in dark matter to the front of the baryon density wave equals 150 Mpc. Hence, there is an enhanced correlation between matter perturbations at this distance scale, which shows up as a feature in the correlation function¹⁰, see Fig. 13. In the Fourier space, this feature produces oscillations (66).

6.4 “Side” remarks

Before proceeding to further discussion of primordial perturbations, let us make a couple of miscellaneous remarks.

¹⁰Notice that the separation at Fig. 13 is given in h^{-1} Mpc, where $h = H_0/100 \text{ km}/(\text{s} \cdot \text{Mpc}) \approx 0.7$. Hence, $100 h^{-1}$ Mpc roughly corresponds to 150 Mpc.

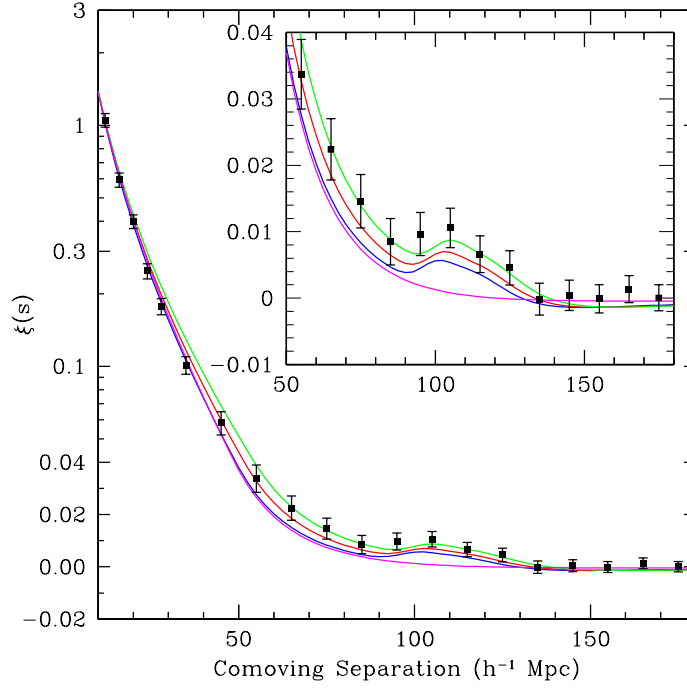


Fig. 13: Baryon acoustic oscillations in galaxy correlation function [25].

6.4.1 Cosmic variance.

We can measure only one Universe, and the best one can do is to *define* the angular spectrum C_l obtained from the data as

$$C_l = \frac{1}{2l+1} \sum_{m=-l}^l |a_{lm}|^2.$$

This is not the same thing as C_l defined in (64), as the latter definition involves averaging over an ensemble of Universes. For given l , there are $(2l+1)$ independent coefficients a_{lm} only, so there exists an irreducible statistical uncertainty of order $\delta C_l / C_l \sim 1/\sqrt{2l+1}$, called cosmic variance. It is particularly pronounced at small l and, indeed, it is much larger than the experimental errors in this part of the angular spectrum (as an example, error bars in the left part of Fig. 10 are precisely due to the cosmic variance).

6.4.2 Measuring the cosmological parameters.

The angular spectrum of CMB temperature anisotropy and polarization, as well as other cosmological data, encodes information on the cosmological parameters. As an example, the sound horizon at recombination is a good standard ruler back at that epoch. It is seen at an angle that depends on the geometry of 3-dimensional space (an interval is seen at larger angle on a sphere than on a plane) and on the dark energy density (since dark energy affects the distance to the sphere of photon last scattering). This is shown in Fig. 14.

Likewise, the baryon acoustic oscillations provide a standard ruler at relatively late times (low redshifts $z \sim 0.2 - 0.4$). A combination of their measurement with CMB anisotropy give quite precise determination of both spatial curvature (which is found to be zero within error bars) and dark energy density. Notably, this determination of ρ_Λ is in good agreement with various independent data, notably, with the data on SNe 1a, which were the first unambiguous evidence for dark energy [27, 28].

There are many other ways in which the cosmological parameters, including Ω_B and Ω_{DM} , affect

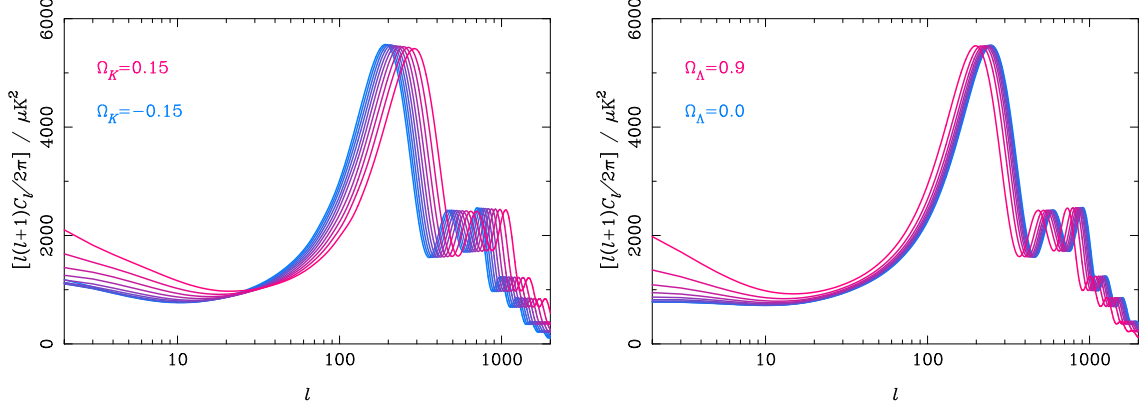


Fig. 14: Effect of spatial curvature (left) and dark energy (right) on the CMB temperature angular spectrum [26]. $\Omega_k = \pm(RH_0)^2$ is the relative contribution of spatial curvature to the Friedmann equation, with R being the radius of spatial curvature. Negative sign corresponds to 3-sphere. As Ω_k decreases, the curves in the left plot move left. Likewise, the curves on the right plot move left as Ω_Λ increases.

the CMB anisotropies. In particular, the heights of the acoustic peaks in the CMB temperature angular spectrum are very sensitive to the baryon-to-photon ratio η_B (and hence to Ω_B), the overall shape of the curve in Fig. 10 strongly depends on Ω_{DM} , etc. By fitting the CMB data and combining them with the results of other cosmological observations, one is able to obtain quite precise knowledge of our Universe.

6.5 Properties of primordial density perturbations — hints about the earliest cosmological epoch

As we emphasized above, the density perturbations were generated at a very early, pre-hot epoch of the cosmological evolution. Obviously, it is of fundamental importance to figure out what precisely that epoch was. One of its properties is clear right away: it must be such that the cosmologically relevant wavelengths, including the wavelengths of the present horizon scale, were subhorizon early at that epoch. Only in that case the perturbations of these wavelengths could be generated in a causal manner at the pre-hot epoch. Notice that this is another manifestation of the horizon problem discussed in Section 2.6: we know from the observational data on density perturbations that our entire visible Universe was causally connected by the beginning of the hot stage.

An excellent hypothesis on the pre-hot stage is inflation, the epoch of nearly exponential expansion,

$$a(t) = e^{\int H dt}, \quad H \approx \text{const}.$$

Originally [29], inflation was designed to solve the problems of the hot Big Bang cosmology, such as the horizon problem, as well as the flatness, entropy and other problems. It does this job very well: the horizon size at inflation is at least

$$l_H(t) = a(t) \int_{t_i}^t \frac{dt'}{a(t')} = H^{-1} e^{H(t-t_i)},$$

where t_i is the time inflation begins, and we set $H = \text{const}$ for illustrational purposes. This size is huge for $t - t_i \gg H^{-1}$, so the entire visible Universe is naturally causally connected.

From the viewpoint of perturbations, the physical momentum $q(t) = k/a(t)$ decreases (gets redshifted) at inflation, while the Hubble parameter stays almost constant. So, every mode is first subhorizon ($q(t) \gg H(t)$), and later superhorizon ($q(t) \ll H(t)$) at inflation. This situation is opposite to what happens at radiation and matter domination, see Fig. 15; this is precisely the pre-requisite for generating the density perturbations. In fact, inflation does generate primordial density perturbations [30], whose properties are consistent with everything we know about them.

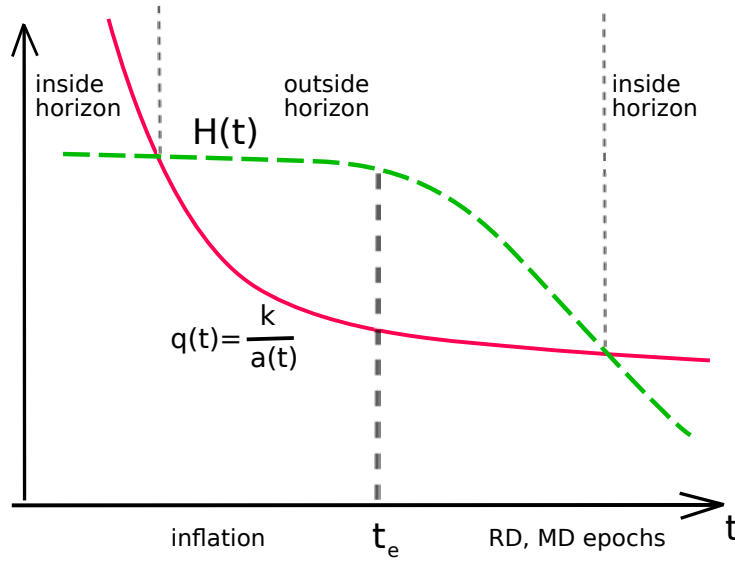


Fig. 15: Physical momentum and Hubble parameter at inflation and later. t_e is the time of the inflation end.

Inflation is not the only hypothesis proposed so far, however. One option is the bouncing Universe scenario, which assumes that the cosmological evolution begins from contraction, then the contracting stage terminates at some moment of time (bounce) and is followed by expansion. A version is the cycling Universe scenario with many cycles of contraction–bounce–expansion. Another scenario is that the Universe starts out from nearly flat and static state and then speeds up its expansion. Theoretical realizations of these scenarios are more difficult than inflation, but they are not impossible, as became clear recently. So, one of the major purposes of cosmology is to choose between various hypotheses on the basis of observational data. The properties of cosmological perturbations are the key issue in this regard.

There are several things which we already know about the primordial density perturbations. By “primordial” we mean the perturbations deep in the superhorizon regime at the radiation-domination epoch. As we already know, perturbations are time-independent in this regime. They set the initial conditions for further evolution, and this evolution is well understood, at least in the linear regime. Hence, using observational data, one is able to measure the properties of primordial perturbations. Of course, since the properties we know of are established by observations, they are valid within certain error bars. Conversely, deviations from the results listed below, if observed, would be extremely interesting.

First, density perturbations are **adiabatic**. This means that there are perturbations in the energy density, but *not in composition*. More precisely, the baryon-to-entropy ratio and dark matter-to-entropy ratio are constant in space,

$$\delta\left(\frac{n_B}{s}\right) = \text{const} , \quad \delta\left(\frac{n_{DM}}{s}\right) = \text{const} . \quad (67)$$

This is consistent with the generation of the baryon asymmetry and dark matter at the hot cosmological epoch: in that case, all particles were at thermal equilibrium early at the hot epoch, the temperature completely characterized the whole cosmic medium at that time, and as long as physics behind the baryon asymmetry and dark matter generation is the same everywhere in the Universe, the baryon and dark matter abundance (relative to the entropy density) is necessarily the same everywhere. In principle, there may exist *entropy* (or isocurvature) perturbations, such that at the early hot epoch energy density (dominated by relativistic matter) was homogeneous, while the composition was not. This would give initial conditions for the evolution of density perturbations, which would be entirely different from those characteristic of the adiabatic perturbations. As a result, the angular spectrum of the CMB temperature

anisotropy would be entirely different, see Fig. 16. No admixture of the entropy perturbations have been detected so far, but it is worth emphasizing that even small admixture will show that the most popular mechanisms for generating dark matter and/or baryon asymmetry (including those discussed in Sections 3 and 4) have nothing to do with reality. One would have to think, instead, that the baryon asymmetry and/or dark matter were generated before the beginning of the hot stage.

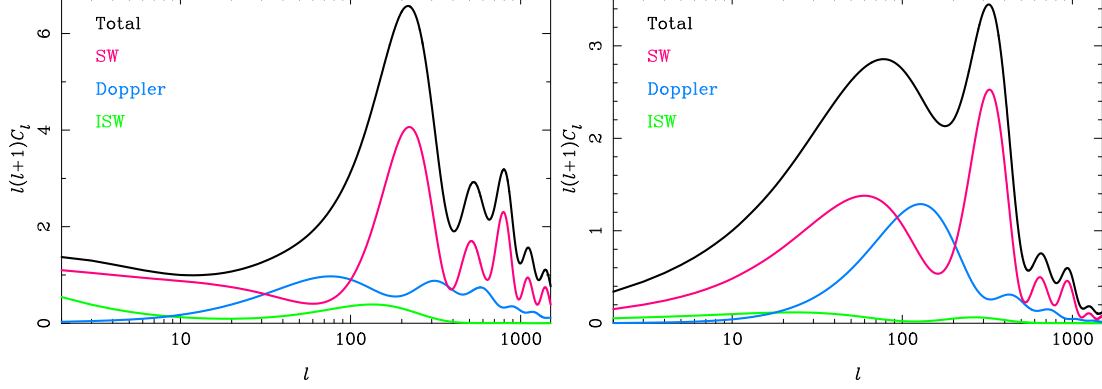


Fig. 16: Angular spectrum of the CMB temperature anisotropy for adiabatic perturbations (left) and entropy perturbations (right) [26].

Second, the primordial density perturbations are **Gaussian random field**. Gaussianity means that the three-point (and all odd) correlation function vanishes, while the four-point function and all higher order even correlation functions are expressed through the two-point function via Wick's theorem:

$$\begin{aligned}\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle &= 0 \\ \langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\delta(\mathbf{k}_4) \rangle &= \langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2) \rangle \cdot \langle \delta(\mathbf{k}_3)\delta(\mathbf{k}_4) \rangle \\ &\quad + \text{permutations of momenta} ,\end{aligned}$$

while all odd correlation functions vanish. A technical remark is in order. As a variable characterizing the primordial adiabatic perturbations we use here

$$\delta \equiv \delta\rho_{\text{rad}}/\rho_{\text{rad}} = \delta\rho/\rho$$

deep at the radiation-dominated epoch. This variable is not gauge-invariant, so we implicitly have chosen the conformal Newtonian gauge. In cosmological literature, other, gauge-invariant quantities are commonly in use, ζ and $\mathcal{R}$. In the conformal Newtonian gauge, there is a simple relationship, valid in the superhorizon regime at radiation domination:

$$\mathcal{R} = \zeta = \frac{3}{4}\delta .$$

We will continue to use δ as the basic variable.

Coming back to Gaussianity, we note that this property is characteristic of *vacuum fluctuations of non-interacting (linear) quantum fields*. Hence, it is quite likely that the density perturbations originate from the enhanced vacuum fluctuations of non-interacting or weakly interacting quantum field(s). Free quantum field has the general form

$$\phi(\mathbf{x}, t) = \int d^3k e^{-i\mathbf{k}\mathbf{x}} \left(f_{\mathbf{k}}^{(+)}(t) a_{\mathbf{k}}^{\dagger} + e^{i\mathbf{k}\mathbf{x}} f_{\mathbf{k}}^{(-)}(t) a_{\mathbf{k}} \right) ,$$

where $a_{\mathbf{k}}^\dagger$ and $a_{\mathbf{k}}$ are creation and annihilation operators. For the field in Minkowski space-time one has $f_{\mathbf{k}}^{(\pm)}(t) = e^{\pm i\omega_{\mathbf{k}}t}$, while enhancement, e.g. due to the evolution in time-dependent background, means that $f_{\mathbf{k}}^{(\pm)}$ are large. But in any case, Wick's theorem is valid, provided that the state of the system is vacuum, $a_{\mathbf{k}}|0\rangle = 0$.

Inflation does the job very well: fluctuations of all light fields get enhanced greatly due to the fast expansion of the Universe. This is true, in particular, for the field that dominates the energy density at inflation, called inflaton. Enhanced vacuum fluctuations of the inflaton are nothing but perturbations in the energy density at inflationary epoch in the simplest inflationary models, which are reprocessed into perturbations in the hot medium after the end of inflation. The generation of the density perturbations is less automatic in scenarios alternative to inflation, but there are various examples showing that this is not a particularly difficult problem.

Non-Gaussianity is an important topic of current research. It would show up as a deviation from Wick's theorem. As an example, the three-point function (bispectrum) may be non-vanishing,

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) G(k_i^2; \mathbf{k}_1\mathbf{k}_2; \mathbf{k}_1\mathbf{k}_3) \neq 0 .$$

The shape of $G(k_i^2; \mathbf{k}_1\mathbf{k}_2; \mathbf{k}_1\mathbf{k}_3)$ is different in different models, so this shape is a potential discriminator. In some models the bispectrum vanishes, e.g., due to symmetries. In that case the trispectrum (connected 4-point function) may be measurable instead. Non-Gaussianity is very small in the simplest inflationary models, but it can be sizeable in more contrived models of inflation and in alternatives to inflation. It is worth emphasizing that non-Gaussianity has not been detected yet.

Another important property is that the primordial power spectrum of density perturbations **is flat** (or almost flat). A convenient definition of the power spectrum for homogeneous and anisotropic Gaussian random field is¹¹

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle = \frac{1}{4\pi k^3} \mathcal{P}(k) \delta(\mathbf{k} + \mathbf{k}') . \quad (68)$$

The power spectrum $\mathcal{P}(k)$ defined in this way determines the fluctuation in a logarithmic interval of momenta,

$$\langle \delta^2(\mathbf{x}) \rangle = \int_0^\infty \frac{dk}{k} \mathcal{P}(k) .$$

By definition, the flat spectrum is such that $\mathcal{P}$ is independent of k . It is worth noting that the flat spectrum was conjectured by E. Harrison [31] and Ya. Zeldovich [32] in the beginning of 1970's, long before realistic mechanisms of the generation of density perturbations have been proposed.

In view of the approximate flatness, a natural parametrization is

$$\mathcal{P}(k) = A_s \left(\frac{k}{k_*} \right)^{n_s-1} , \quad (69)$$

where A_s is the amplitude, (n_s-1) is the tilt and k_* is a fiducial momentum, chosen at one's convenience. The flat spectrum in this parametrization has $n_s = 1$. Cosmological data favor the value $n_s \approx 0.96$ (i.e., slightly smaller than 1), see below, but it is fair to say that $n_s = 1$ is still consistent with observations.

The flatness of the power spectrum calls for some symmetry behind this property. In inflationary theory this is the symmetry of the de Sitter space-time, which is the space time of constant Hubble rate,

$$ds^2 = dt^2 - e^{2Ht} d\mathbf{x}^2 , \quad H = \text{const} .$$

This metric is invariant under spatial dilatations supplemented by time translations,

$$\mathbf{x} \rightarrow \lambda \mathbf{x} , \quad t \rightarrow t - \frac{1}{2H} \log \lambda .$$

¹¹Note that the the definition of the power spectrum used in Figs. 7 and 12 is different from (68).

At inflation, H is almost constant in time, and the de Sitter symmetry is an approximate symmetry. For this reason inflation automatically generates nearly flat power spectrum.

The de Sitter symmetry is not the only candidate symmetry behind the flatness of the power spectrum. One possible alternative is conformal symmetry [33, 34]. The point is that the conformal group includes dilatations, $x^\mu \rightarrow \lambda x^\mu$. This property indicates that the relevant part of the theory possesses no scale, and has good chance for producing the flat spectrum. Model-building in this direction has begun recently [34].

6.6 What's next?

Thus, only very basic facts about the primordial density perturbations are observationally established. Even though very suggestive, these facts by themselves are not sufficient for unambiguously establishing the properties of the Universe at the pre-hot epoch of its evolution. In coming years, new properties of cosmological perturbations will hopefully be discovered, which will shed much more light on this pre-hot epoch. Let us discuss some of the potential observables.

6.6.1 Tensor perturbations = relic gravity waves

The simplest, and hence most plausible models of inflation predict sizeable tensor perturbations, which are perturbations of the metric independent of perturbations in the energy density. After entering the horizon, tensor perturbations are nothing but gravity waves. The reason for their generation at inflation is that the exponential expansion of the Universe enhances vacuum fluctuations of all fields, including the gravitational field itself. In inflationary theory, the primordial tensor perturbations are Gaussian random field with nearly flat power spectrum

$$\mathcal{P}_T = A_T \left(\frac{k}{k_*} \right)^{n_T}, \quad (70)$$

where the inflationary prediction is $n_T \approx 0$ (the reason for different definitions of the tensor spectral index n_T in (70) and scalar spectral index n_s in (69) is purely historical).

On the other hand, there seems to be no way of generating nearly flat tensor power spectrum in alternatives to inflation. In fact, most, if not all, alternative scenarios predict unobservably small amplitude of tensor perturbations. Thus, the discovery of tensor modes would be the strongest possible argument in favor of inflation. It is worth noting that *non-observation* of tensor perturbations would not rule inflation out: there are numerous models of inflation which predict tensor modes of very small amplitude.

The tensor power is usually characterized by the tensor-to-scalar ratio

$$r = \frac{A_T}{A_s}.$$

The simplest inflationary models predict, roughly speaking, $r \sim 0.1 - 0.3$. The current situation is summarized in Fig. 17. Clearly, there is an indication for the negative scalar tilt ($n_s - 1$) or non-zero tensor amplitude, or both, though it is premature to say that the flat scalar spectrum with no tensor modes (the Harrison–Zeldovich point) is ruled out.

For the time being, the most sensitive probe of the tensor perturbations is the CMB temperature anisotropy. However, the most promising tool is the CMB polarization. The point is that a certain class of polarization patterns (called B-mode) is generated by tensor perturbations, while scalar perturbations are unable to create it. Hence, the Planck experiment, and especially dedicated experiments aiming at measuring the CMB polarization may well discover the tensor perturbations, i.e., relic gravity waves. Needless to say, this would be a profound discovery. To avoid confusion, let us note that the CMB polarization has been already observed, but it belongs to another class of patterns (so called E-mode) and is consistent with the existence of the scalar perturbations only.

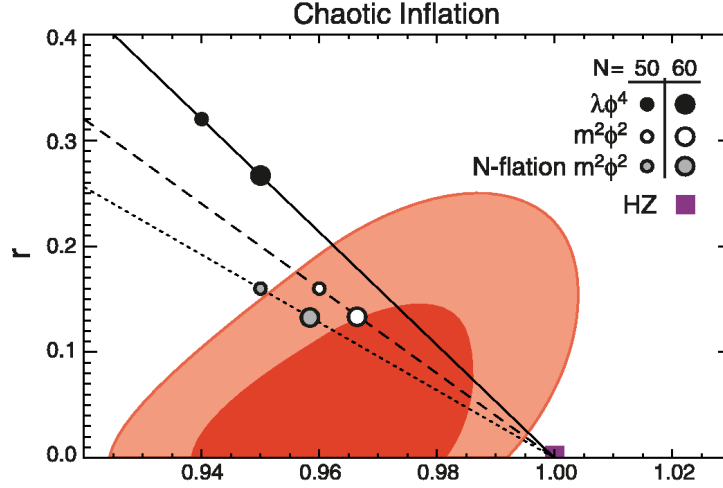


Fig. 17: Allowed regions (at 68% and 95% CL) in the plane (n_s, r) , where n_s is the scalar spectral index and r is the tensor-to-scalar ratio. Rectangle shows the Harrison–Zeldovich point (flat scalar spectrum, no tensor modes). Points on the lines show predictions of popular inflationary models.

6.6.2 Scalar tilt.

Inflationary models and their alternatives will be constrained by the precise determination of the scalar tilt $(n_s - 1)$ and its dependence on momentum k . It appears, however, that the information on $n_s(k)$ that will be obtained in reasonably near future will be of limited significance from the viewpoint of discriminating between different (and even grossly different) scenarios.

6.6.3 Non-Gaussianity.

As we pointed out already, non-Gaussianity of density perturbations is very small in the simplest inflationary models. Hence, its discovery will signalize that either inflation and inflationary generation of density perturbations occurred in a rather complicated way, or an alternative scenario was realized. Once the non-Gaussianity is discovered, and its shape is revealed even with moderate accuracy, many concrete models will be ruled out, while at most a few will get strong support.

6.6.4 Statistical anisotropy.

In principle, the power spectrum of density perturbations may depend on the direction of momentum, e.g.,

$$\mathcal{P}(\mathbf{k}) = \mathcal{P}_0(k) \left(1 + w_{ij}(k) \frac{k_i k_j}{k^2} + \dots \right)$$

where w_{ij} is a fundamental tensor in our part of the Universe (odd powers of k_i would contradict commutativity of the Gaussian random field $\delta(\mathbf{k})$, see Eq. (68)). Such a dependence would definitely imply that the Universe was anisotropic at the pre-hot stage, when the primordial perturbations were generated. This statistical anisotropy is rather hard to obtain in inflationary models, though it is possible in inflation with strong vector fields [35]. On the other hand, statistical anisotropy is natural in some other scenarios, including conformal models [36]

The statistical anisotropy would show up in correlators [37]

$$\langle a_{lm} a_{l'm'} \rangle \quad \text{with } l' \neq l \text{ and/or } m' \neq m$$

At the moment, the situation with observational data is controversial [38], and the new data, notably from the Planck experiment, will hopefully clear it up.

6.6.5 *Admixture of entropy perturbations.*

As we explained above, even small admixture of entropy perturbations would force us to abandon the most popular scenarios of the generation of baryon asymmetry and/or dark matter, which assumed that it happened at the hot epoch. The WIMP dark matter would no longer be well motivated, while other, very weakly interacting dark matter candidates, like axion or superheavy relic, would be preferred. This would make the direct searches for dark matter rather problematic.

7 Conclusion

We are at the eve of new era not only in particles physics, but also in cosmology. There is reasonably well justified expectation that the LHC will shed light on long-standing cosmological problems of the origin of the baryon asymmetry and nature of dark matter in our Universe. The ideas we discussed in these lectures in this regard may well be not the right ones: we can only hypothesize on physics beyond the Standard Model and its role in the early Universe.

In fact, the TeV scale physics may be dramatically different from physics we get used to. As an example, it is not excluded that TeV is not only electroweak, but also gravitational scale. This is the case in models with large extra dimensions, in which the Planck scale is related to the fundamental gravity scale in a way that involves the volume of extra dimensions, and hence the fundamental scale can be much below M_{Pl} (for a review see, e.g., Ref. [39]). If the LHC will find that, indeed, the fundamental gravity scale is in the TeV range, this would have most profound consequences for both microscopic physics and cosmology. On the microscopic physics side, this would enable one to study at colliders quantum gravity and its high-energy extension — possibly string theory, while on the cosmological side, the entire picture of the early Universe would have to be revised. Inflation, if any, would have to occur either at low energy density or in the regime of strong quantum gravity effects. The highest temperatures in the usual expansion history would be at most in the TeV range, so dark matter and baryon asymmetry would have to be generated either below TeV temperatures or in quantum gravity regime. Even more intriguing will be the study of quantum gravity cosmological epoch, with hints from colliders gradually coming. This, probably, is too bright a prospective to be realistic.

It is more likely that the LHC will find something entirely new, something theorists have not thought about. Or, conversely, find so little that one will have to get serious about anthropic principle. In any case, the LHC results will definitely change the landscape of fundamental physics, cosmology included.

The observational data unequivocally tell us that the hot stage of the cosmological evolution was preceded by some other epoch, at which the cosmological perturbations were generated. The best guess for this epoch is inflation, but one should bear in mind that there are alternative possibilities. It is fascinating that with new observational data, there is good chance to learn what precisely that pre-hot epoch was. It may very well be that in this way we will be able to probe physics at the energy, distance and time scales well beyond the reach of the LHC.

References

- [1] R. H. Brandenberger, *Particle physics aspects of modern cosmology*, arXiv:hep-ph/9701276; W. L. Freedman and M. S. Turner, *Rev. Mod. Phys.* **75** (2003) 1433 [arXiv:astro-ph/0308418]; V. Rubakov, *PoS RTN2005* (2005) 003; J. A. Peacock, *Cosmology and particle physics*, in *Proc. 1998 European School of High-Energy Physics, St. Andrews, Scotland*; M. Shaposhnikov, *Cosmology and astrophysics*, in *Proc. 2000 European School of High-Energy Physics, Caramulo, Portugal*; I. I. Tkachev, *Astroparticle physics*, in *Proc. 2003 European School on High-Energy Physics, Tsakhkadzor, Armenia*, arXiv:hep-ph/0405168.
- [2] G. Jungman, M. Kamionkowski and K. Griest, *Phys. Rept.* **267** (1996) 195 [arXiv:hep-ph/9506380]; A. Bottino and N. Fornengo, *Dark matter and its particle candidates*, arXiv:hep-

- ph/9904469; K. A. Olive, *Dark matter*, arXiv:astro-ph/0301505; G. Bertone, D. Hooper and J. Silk, Phys. Rept. **405** (2005) 279. [arXiv:hep-ph/0404175].
- [3] A. D. Dolgov, *Baryogenesis, 30 years after*, arXiv:hep-ph/9707419 V. A. Rubakov and M. E. Shaposhnikov, Usp. Fiz. Nauk **166** (1996) 493 [Phys. Usp. **39** (1996) 461] [arXiv:hep-ph/9603208]; A. Riotto and M. Trodden, Ann. Rev. Nucl. Part. Sci. **49** (1999) 35 [arXiv:hep-ph/9901362]; M. Trodden, Rev. Mod. Phys. **71** (1999) 1463 [arXiv:hep-ph/9803479].
- [4] S. Weinberg, Rev. Mod. Phys. **61** (1989) 1; V. Sahni and A. A. Starobinsky, Int. J. Mod. Phys. D **9** (2000) 373 [arXiv:astro-ph/9904398]; P. J. E. Peebles and B. Ratra, Rev. Mod. Phys. **75** (2003) 559 [arXiv:astro-ph/0207347]; T. Padmanabhan, Phys. Rept. **380** (2003) 235 [arXiv:hep-th/0212290]; E. J. Copeland, M. Sami and S. Tsujikawa, Int. J. Mod. Phys. D **15** (2006) 1753 [arXiv:hep-th/0603057].
- [5] V. F. Mukhanov, H. A. Feldman and R. H. Brandenberger, Phys. Rept. **215** (1992) 203; A. R. Liddle and D. H. Lyth, Phys. Rept. **231** (1993) 1 [arXiv:astro-ph/9303019]; D. H. Lyth and A. Riotto, Phys. Rept. **314** (1999) 1 [arXiv:hep-ph/9807278].
- [6] A. de Oliveira-Costa, G. F. Smoot and A. A. Starobinsky, *Constraining topology with the CMB*, arXiv:astro-ph/9705125; N. G. Phillips and A. Kogut, Astrophys. J. **645**, 820 (2006) [arXiv:astro-ph/0404400].
- [7] E. Gawiser and J. Silk, Phys. Rept. **333**, 245 (2000) [arXiv:astro-ph/0002044].
- [8] D. Larson *et al.*, Astrophys. J. Suppl. **192** (2011) 16 [arXiv:1001.4635 [astro-ph.CO]].
- [9] E. Komatsu *et al.* [WMAP Collaboration], Astrophys. J. Suppl. **192** (2011) 18 [arXiv:1001.4538 [astro-ph.CO]].
- [10] E. Komatsu *et al.* [WMAP Collaboration], Astrophys. J. Suppl. **180** (2009) 330 [arXiv:0803.0547 [astro-ph]].
- [11] U. Seljak *et al.* [SDSS Collaboration], Phys. Rev. D **71**, 103515 (2005) [arXiv:astro-ph/0407372].
- [12] G. 't Hooft, Phys. Rev. Lett. **37** (1976) 8.
- [13] F. R. Klinkhamer and N. S. Manton, Phys. Rev. D **30**, 2212 (1984).
- [14] V. A. Kuzmin, V. A. Rubakov and M. E. Shaposhnikov, Phys. Lett. B **155** (1985) 36.
- [15] V. A. Rubakov, *Classical theory of gauge fields*, Princeton, USA: Univ. Pr. (2002) 444 p.p.
- [16] M. S. Carena, M. Quiros and C. E. M. Wagner, Phys. Lett. B **380** (1996) 81 [arXiv:hep-ph/9603420]; M. S. Carena, M. Quiros, A. Riotto, I. Vilja and C. E. M. Wagner, Nucl. Phys. B **503** (1997) 387 [arXiv:hep-ph/9702409]; M. S. Carena, M. Quiros, M. Seco and C. E. M. Wagner, Nucl. Phys. B **650** (2003) 24 [arXiv:hep-ph/0208043].
- [17] B. Kayser, *Neutrino Oscillation Physics*, in these proceedings.
- [18] V. A. Rubakov, Phys. Rev. D **61**, 061501 (2000) [arXiv:hep-ph/9911305]; P. J. Steinhardt and N. Turok, Science **312**, 1180 (2006) [arXiv:astro-ph/0605173].
- [19] S. Weinberg, Phys. Rev. Lett. **59**, 2607 (1987).
- [20] A. D. Linde, *Inflation And Quantum Cosmology*, in: *Three hundred years of gravitation*. Cambridge Univ. Press, Eds. Hawking, S.W. and Israel, W., 604-630 (1987).
- [21] M. Tegmark *et al.* [SDSS Collaboration], Astrophys. J. **606** (2004) 702 [arXiv:astro-ph/0310725].
- [22] C. L. Reichardt *et al.*, Astrophys. J. **694** (2009) 1200 [arXiv:0801.1491 [astro-ph]].
- [23] J. Urrestilla, N. Bevis, M. Hindmarsh, M. Kunz and A. R. Liddle, JCAP **0807**, 010 (2008) [arXiv:0711.1842 [astro-ph]].
- [24] W. J. Percival, S. Cole, D. J. Eisenstein, R. C. Nichol, J. A. Peacock, A. C. Pope and A. S. Szalay, Mon. Not. Roy. Astron. Soc. **381** (2007) 1053 [arXiv:0705.3323 [astro-ph]].
- [25] D. J. Eisenstein *et al.* [SDSS Collaboration], Astrophys. J. **633** (2005) 560 [arXiv:astro-ph/0501171].

- [26] A. Challinor, Lect. Notes Phys. 2004. **653**. 71 [arXiv:astro-ph/0403344].
- [27] A. G. Riess *et al.* [Supernova Search Team Collaboration], Astron. J. **116** (1998) 1009 [arXiv:astro-ph/9805201].
- [28] S. Perlmutter *et al.* [Supernova Cosmology Project Collaboration], Astrophys. J. **517** (1999) 565 [arXiv:astro-ph/9812133].
- [29] A. A. Starobinsky, JETP Lett. **30** 682 (1979) [Pisma Zh. Eksp. Teor. Fiz. **30**, 719 (1979)] Phys. Lett. B **91**, 99 (1980); A. H. Guth, Phys. Rev. D **23**, 347 (1981); A. D. Linde, Phys. Lett. B **108**, (1982) 389; Phys. Lett. B **129** 177 (1983); A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
- [30] V. F. Mukhanov and G. V. Chibisov, JETP Lett. **33**, (1981) 532 [Pisma Zh. Eksp. Teor. Fiz. **33**, 549 (1981)]; S. W. Hawking, Phys. Lett. B **115**, 295 (1982); A. A. Starobinsky, Phys. Lett. B **117**, 175 (1982); A. H. Guth and S. Y. Pi, Phys. Rev. Lett. **49**, 1110 (1982); J. M. Bardeen, P. J. Steinhardt and M. S. Turner, Phys. Rev. D **28**, 679 (1983).
- [31] E. R. Harrison, Phys. Rev. D **1**, 2726 (1970).
- [32] Y. B. Zeldovich, Mon. Not. Roy. Astron. Soc. **160**, 1P (1972).
- [33] I. Antoniadis, P. O. Mazur and E. Mottola, Phys. Rev. Lett. **79** (1997) 14 [arXiv:astro-ph/9611208].
- [34] V. A. Rubakov, JCAP **0909** (2009), 030; [arXiv:0906.3693 [hep-th]]; P. Creminelli, A. Nicolis and E. Trincherini, JCAP **1011**, 021 (2010); [arXiv:1007.0027 [hep-th]]; K. Hinterbichler and J. Khoury, JCAP **1204** (2012) 023 [arXiv:1106.1428 [hep-th]].
- [35] M. A. Watanabe, S. Kanno and J. Soda, Phys. Rev. Lett. **102** (2009) 191302 [arXiv:0902.2833 [hep-th]]; T. R. Dulaney and M. I. Gresham, Phys. Rev. D **81** (2010) 103532 [arXiv:1001.2301]; A. E. Gumrukcuoglu, B. Himmetoglu and M. Peloso, Phys. Rev. D **81** (2010) 063528 [arXiv:1001.4088 [astro-ph.CO]].
- [36] M. Libanov and V. Rubakov, JCAP **1011** (2010) 045 [arXiv:1007.4949]; M. Libanov, S. Ramazanov and V. Rubakov, JCAP **1106** (2011) 010 [arXiv:1102.1390 [hep-th]].
- [37] L. Ackerman, S. M. Carroll and M. B. Wise, Phys. Rev. D **75** (2007) 083502 [Erratum-ibid. D **80** (2009) 069901] [arXiv:astro-ph/0701357]; A. R. Pullen and M. Kamionkowski, Phys. Rev. D **76** (2007) 103529 [arXiv:0709.1144 [astro-ph]].
- [38] N. E. Groeneboom and H. K. Eriksen, Astrophys. J. **690** (2009) 1807 [arXiv:0807.2242 [astro-ph]]; D. Hanson and A. Lewis, Phys. Rev. D **80** (2009) 063004 [arXiv:0908.0963 [astro-ph.CO]]; C. L. Bennett *et al.*, Astrophys. J. Suppl. **192** (2011) 17 [arXiv:1001.4758 [astro-ph.CO]].
- [39] V. A. Rubakov, Phys. Usp. **44** (2001) 871 [Usp. Fiz. Nauk **171** (2001) 913] [arXiv:hep-ph/0104152].

QCD in Heavy Ion Collisions

E. Iancu

Institut de Physique Théorique de Saclay, Gif-sur-Yvette, France

Abstract

These lectures provide a modern introduction to selected topics in the physics of ultrarelativistic heavy ion collisions which shed light on the fundamental theory of strong interactions, the Quantum Chromodynamics. The emphasis is on the partonic forms of QCD matter which exist in the early and intermediate stages of a collision — the colour glass condensate, the glasma, and the quark-gluon plasma — and on the effective theories that are used for their description. These theories provide qualitative and even quantitative insight into a wealth of remarkable phenomena observed in nucleus-nucleus or deuteron-nucleus collisions at RHIC and/or the LHC, like the suppression of particle production and of azimuthal correlations at forward rapidities, the energy and centrality dependence of the multiplicities, the ridge effect, the limiting fragmentation, the jet quenching, or the dijet asymmetry.

1 Introduction

With the advent of the high-energy colliders RHIC (the Relativistic Heavy Ion Collider operating at RHIC since 2000) and the LHC (the Large Hadron Collider which started operating at CERN in 2008), the physics of relativistic heavy ion collisions has entered a new era: the energies available for the collisions are high enough — up to 200 GeV per interacting nucleon pair at RHIC and potentially up to 5.5 TeV at the LHC (although so far one has reached ‘only’ 2.76 TeV) — to ensure that *new forms of QCD matter*, characterized by high parton densities, are being explored by the collisions. These new forms of matter refer to both the wavefunctions of the incoming nuclei, prior to the collision, which develop high gluon densities leading to *colour glass condensates*, and the partonic matter produced in the intermediate stages of the collision, which is expected to form a *quark-gluon plasma*. The asymptotic freedom property of QCD implies that these high-density forms of matter are *weakly coupled* (at least in so far as their bulk properties are concerned) and hence can be studied via controlled calculations within perturbative QCD. But such studies remain difficult and pose many challenges to the theorists: precisely because of their high density, these new forms of matter are the realm of *collective, non-linear phenomena*, whose mathematical description often transcends the ordinary perturbation theory. Moreover, there are also phenomena (first revealed by the experiments at RHIC) which seem to elude a weak-coupling description and call for non-perturbative techniques.

These challenges stimulated new ideas and the development of new theoretical tools aiming at a fundamental understanding of *QCD matter under extreme conditions* : high energy, high parton densities, high temperature. The ongoing experimental programs at RHIC and the LHC provide a unique and timely opportunity to test such new ideas, constrain or reject models, and orient the theoretical developments. Over the last decade, the experimental and theoretical efforts have gone hand in hand, leading to a continuously improving physical picture, which is by now well rooted in QCD. The purpose of these lectures is to provide an introduction to this physical picture, with emphasis on those aspects of the dynamics for which we are confident to have a reasonably good (although still far from perfect) understanding from first principles, i.e. from the Lagrangian of Quantum Chromodynamics. These aspects concern the *partonic stages* of a heavy ion collision, at sufficiently early times. These are also the stages to which refers most of the experimental and theoretical progress over the last decade.

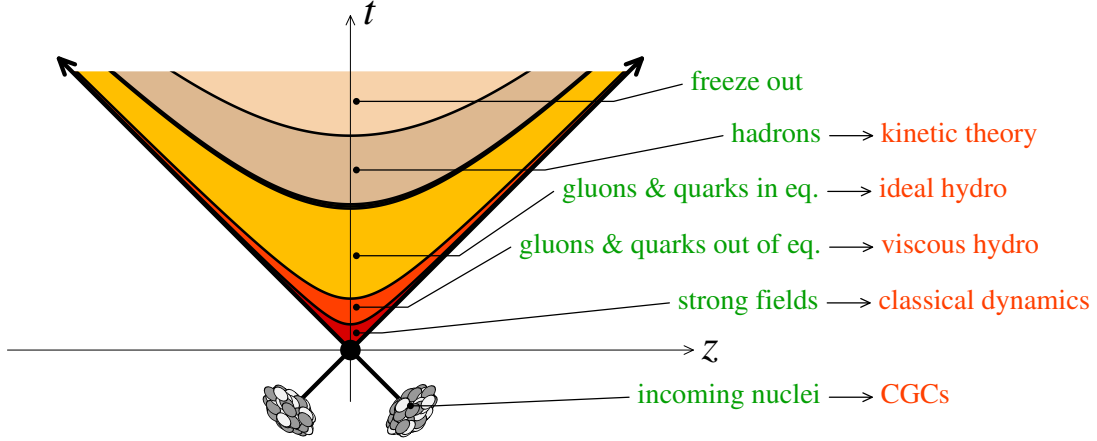


Fig. 1: Schematic representation of the various stages of a HIC as a function of time t and the longitudinal coordinate z (the collision axis). The ‘time’ variable which is used in the discussion in the text is the *proper time* $\tau \equiv \sqrt{t^2 - z^2}$, which has a Lorentz-invariant meaning and is constant along the hyperbolic curves separating various stages in this figure.

2 Stages of a heavy ion collision: the case for effective theories

The theoretically motivated space-time picture of a heavy ion collision (HIC) is depicted in Fig. 1. This figure illustrates the various forms of QCD matter intervening during the successive phases of the collision:

1. Prior to the collision, and in the center-of-mass frame (which at RHIC and the LHC is the same as the laboratory frame), the two incoming nuclei look as two Lorentz-contracted ‘pancakes’, with a longitudinal extent smaller by a factor $\gamma \sim 100$ (the Lorentz boost factor) than the radial extent in the transverse plane. As we shall see, these ‘pancakes’ are mostly composed with *gluons* which carry only tiny fractions $x \ll 1$ of the longitudinal momenta of their parent nucleons, but whose density is rapidly increasing with $1/x$. By the uncertainty principle, the gluons which make up such a high-density system carry *relatively large transverse momenta*. A typical value for such a gluon in a Pb or Au nucleus is $k_\perp \simeq 2 \text{ GeV}$ for $x = 10^{-4}$. By the ‘asymptotic freedom’ property of QCD, the gauge coupling which governs the mutual interactions of these gluons is *relatively weak*. This gluonic form of matter, which is dense and weakly coupled, and dominates the wavefunction of any hadron (nucleon or nucleus) at sufficiently high energy, is *universal* — its properties are the same form all hadrons. It is known as the *colour glass condensate* (CGC).
2. At time $\tau = 0$, the two nuclei hit with each other and the interactions start developing. The ‘hard’ processes, i.e. those involving relatively large transferred momenta $Q \gtrsim 10 \text{ GeV}$, are those which occur faster (within a time $\tau \sim 1/Q$, by the uncertainty principle¹). These processes are responsible for the production of ‘hard particles’, i.e. particles carrying transverse energies and momenta of the order of Q . Such particles, like (hadronic) jets, direct photons, dilepton pairs, heavy quarks, or vector bosons, are generally the most striking ingredients of the final state and are often used to characterize the topology of the latter — e.g., one speaks about ‘a dijet event’, cf. Fig. 2 left, or ‘a photon-jet’ event, cf. Fig. 2 right.
3. At a time $\tau \sim 0.2 \text{ fm/c}$, corresponding to a ‘semi-hard’ transverse momentum scale $Q \sim 1 \text{ GeV}$, the bulk of the partonic constituents of the colliding nuclei (meaning the gluons composing the respective CGCs) are liberated by the collision. This is when most of the ‘multiplicity’ in the final

¹Throughout these notes, we shall generally use the natural system of units $\hbar = c = k_B = 1$, so in particular there is no explicit factor $\hbar$ in the uncertainty principle. Yet, in some cases, we shall restore this factor for more clarity.

state is generated; that is, most of the hadrons eventually seen in the detectors are produced via the fragmentation and the hadronisation of the initial-state gluons liberated at this stage. But before ending up in the detectors, these partons undergo a complex evolution. Just after being liberated, they form a relatively dense medium, whose average density energy in Pb+Pb collisions at the LHC is estimated as $\varepsilon \gtrsim 15 \text{ GeV/fm}^3$; this is about 10 times larger than the density of nuclear matter and 3 times larger than in Au+Au collisions at RHIC. This non-equilibrium state of partonic matter, which besides its high density has also other distinguished features to be discussed later, is known as the *glasma*.

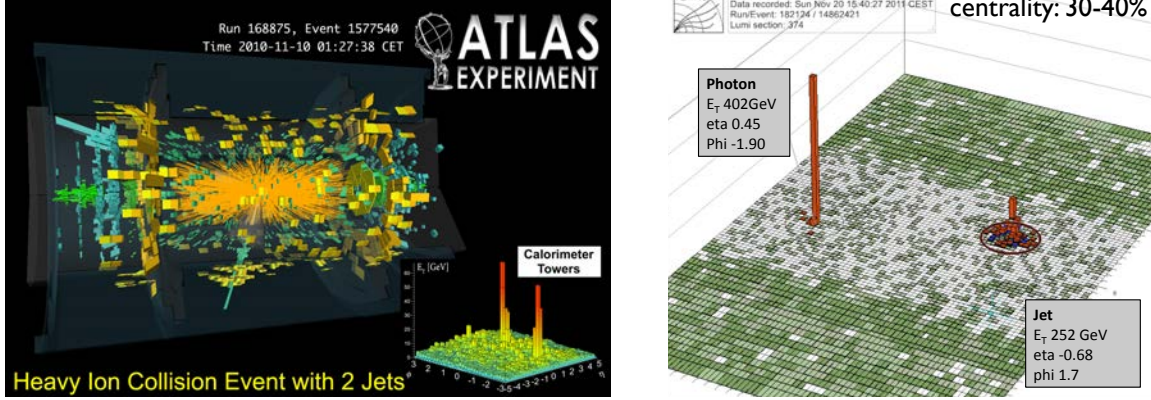


Fig. 2: A couple of di-jets events in Pb+Pb collisions at ATLAS (left) and CMS (right).

4. If the produced partons did not interact with each other, or if these interactions were negligible, then they would rapidly separate from each other and independently evolve (via fragmentation and hadronization) towards the final-state hadrons. This is, roughly speaking, the situation in proton-proton collisions. But the data for heavy ion collisions at both RHIC and the LHC exhibit collective phenomena (like the ‘elliptic flow’ to be discussed later) which clearly show that the partons liberated by the collision *do* actually interact with each other, and quite strongly. A striking consequence of these interactions is the fact that this partonic matter rapidly approaches towards *thermal equilibrium* : the data are consistent with a relatively short thermalization time, of order $\tau \sim 1 \text{ fm/c}$. This is striking since it requires rather strong interactions among the partons, which can compete with the medium expansion: these interactions have to redistribute energy and momentum among the partons, in spite of the fact that the latter separate quite fast away from each other. Such a rapid thermalization seems incompatible with perturbative calculations at weak coupling and represents a main argument in favour of a new paradigm: the dense partonic matter produced in the intermediate stages of a HIC may actually be a *strongly coupled fluid*.
5. The outcome of this thermalization process is the high-temperature phase of QCD matter known as the *quark-gluon plasma*. The abundant production and detailed study of this phase is the Holy Grail of the heavy ion programmes at RHIC and the LHC. The existence of this phase is well established via theoretical calculations on the lattice, but its experimental production within a HIC is at best ephemeral: the partonic matter keeps expanding and cooling down (which in particular implies that the temperature is space and time dependent, i.e. thermal equilibrium is reached only *locally*) and it eventually hadronizes — that is, the ‘coloured’ quark and gluons get trapped within colour-singlet hadrons. Hadronization occurs when the (local) temperature becomes of the order of the critical temperature T_c for deconfinement, which from lattice QCD studies is known to be within the range $T_c \simeq 150 \div 180 \text{ MeV}$. In Pb+Pb collisions at the LHC, this is estimated to happen around a time $\tau \sim 10 \text{ fm/c}$.

6. For larger times $10 \lesssim \tau \lesssim 20$ fm/c, this hadronic system is still relatively dense, so it preserves local thermal equilibrium while expanding. One then speaks of a *hot hadron gas*, whose temperature and density are however decreasing with time.
7. Around a time $\tau \sim 20$ fm/c, the density becomes so low that the hadrons stop interacting with each other. That is, the collision rate becomes smaller than the expansion rate. This transition between a fluid state (where the hadrons undergo many collisions) and a system of free particles is referred to as the *freeze-out*. From that moment on, the hadrons undergo free streaming until they reach the detector. One generally expects that the momentum distribution of the outgoing particles is essentially the same as their thermal distribution within the fluid, towards the late stages of the expansion, just before the freeze-out. This assumption appears to be confirmed by the data: the particle spectra as measured by the detectors can be well described as *thermal* (Maxwell–Boltzmann) distributions, with only few free parameters, like the fluid temperature and velocity at the time of freeze-out. This is generally seen as an additional argument in favour of thermalization, but one must be cautious on that, since the mechanism of hadronisation itself can lead to spectra which are apparently thermal. As a matter of fact, the freeze-out temperature extracted from the ratios of particle abundances at RHIC appears to be the same, $T_f \simeq 170$ MeV, in both Au+Au and p+p collisions, while of course no QGP phase is expected in p+p.

Although extremely schematic, this simple enumeration of the various stages of a HIC already illustrates the variety and complexity of the forms of matter traversed by the QCD matter liberated by the collision on its way to the detectors. In principle, all these forms of matter and their mutual transformations admit an unambiguous theoretical description in the framework of Quantum Chromodynamics, which is the fundamental theory of strong interactions. But although this theory exists with us for about 40 years, it is still far from having delivered all its secrets. Indeed, in spite of the apparent simplicity of its Lagrangian, which looks hardly more complicated than that of the quantum electrodynamics (the theory of photons and electrons), the QCD dynamics is considerably richer and more complicated — which is why it can accommodate so many phases! What renders the theoretical study of HIC’s so difficult is the extreme complexity of the relevant forms of hadronic matter, characterized by high (parton or hadron) densities and strong collective phenomena. For a theorist, the most efficient way to try and organize this complexity is to build *effective theories*.

An ‘effective theory’ should not be confused with a ‘model’: its main purpose is not to provide a heuristic description of the data using some physical guidance together with a set of free parameters. Rather, it aims at a fundamental understanding and its construction is always guided by the underlying fundamental theory — here, QCD. Specifically, an effective theory is a simplified version of the fundamental theory which includes the ‘soft’ (*i.e.* low energy and momentum) degrees of freedom (d.o.f.) required for the description of the physical phenomena occurring at a relatively large space-time scale, but ignores the ‘hard’ d.o.f. with higher energies and momenta. More precisely, the hard modes cannot be totally ignored — they interact with the soft modes and thus affect the properties of the effective theory —, rather they are ‘integrated out’ via some coarse-graining (or ‘renormalization group’) procedure, which can be perturbative or non-perturbative.

If the coupling is weak ($g \ll 1$), the ‘hard-soft’ interactions can be treated in perturbation theory and then the effective theory emerges as a controlled approximation to the original theory. This generally amounts to computing Feynman graphs with hard loop momenta and soft external legs. By the uncertainty principle, the hard modes are localized on short space-time distances, so their net effect is to provide *quasi-local vertices*, or ‘parameters’ — like effective masses and couplings — in the effective Lagrangian for the soft modes. But even at weak coupling, one often has to deal with a large, or even infinite, number of Feynman graphs *at any given order* in g , because the contributions due to individual graphs are enhanced by the large disparity of scales between the hard and soft d.o.f. and/or by the high density of medium constituents. This is where the effective theory is most useful: it allows us to ‘re-sum’ (modulo some approximations) a large number of Feynman graphs of the original field theory and

replace their effects by a small number of ‘parameters’ in the effective Lagrangian.

When the coupling is relatively strong, $g \gtrsim 1$, standard perturbation theory (the expansion in powers of g) is bound to fail and the construction of effective theories becomes more problematic. So long as the coupling is just *moderately* strong, say $g \sim \mathcal{O}(1)$, there is still hope that some insightful resummations of the perturbation theory, as based on the proper identification of the relevant d.o.f., may reasonably work — we shall later encounter some examples in that sense. If, in some regime, the coupling happens to be even stronger, perturbation theory brings no guidance anymore, and there is no systematic method to construct effective theories. They can merely be postulated on the basis of general physical considerations, like the *symmetries* of the fundamental theory. In such a case, the effective masses or coupling constants are generally treated as free parameters, to be matched against the data or, in some cases, against *lattice QCD* calculations. Effective theories may also emerge for rather deep and unexpected reasons, as we shall see on the example of the *gauge/string duality* later on.

If in the previous discussion we mentioned both *weak* and *strong* coupling scenarios, is because in QCD — and indeed in any of the fundamental field theories in Nature — the coupling ‘constant’ is *not* fixed: it ‘runs’ with the typical momenta exchanged in the interactions, meaning that it is different when probing the physics on different space-time scales. What is essential about QCD is the property of *asymptotic freedom*: the fact that the coupling becomes weaker on shorter distances, or with increasing momentum transfer. Given our experience with electromagnetism, this property may look counterintuitive. In QED, the electric charge of the nucleus inside an atom is well known to be *screened* by the surrounding electron cloud, so that the atom appears electrically neutral from far away. Similarly, the electric charge of an electron is screened by electron-positron virtual pairs which pump up from the vacuum, with the result that the *effective* charge $\alpha_{\text{e.m.}}(R) \equiv e^2(R)/4\pi$ decreases with the distance R from the electron. But in QCD, there is *anti-screening*: the effective colour charge of a quark or gluon, as measured by its coupling ‘constant’ $\alpha_s(R) \equiv g^2(R)/4\pi$, increases with R or with decreasing the transferred momentum Q . (Recall that $Q \sim 1/R$ by the uncertainty principle.) Specifically, one has $\alpha_s(Q^2) \simeq 0.1$ when $Q = 100 \text{ GeV}$ (the typical scale for electroweak physics and also for hadronic jets at the LHC). This ‘asymptotic freedom’ is, of course, the ultimate reason behind the success story of perturbative QCD in relation with ‘hard’ processes. It also justifies the use of perturbation theory for integrating out the ‘hard’ d.o.f. in the construction of effective theories for the ‘soft’ ones.

But there is also the reverse of the medal: with decreasing Q below 100 GeV, the QCD coupling is increasing, albeit slowly, according to

$$\alpha_s(Q^2) \equiv \frac{g^2(Q^2)}{4\pi} = \frac{4\pi N_c}{(11N_c - 2N_f) \ln(Q^2/\Lambda_{\text{QCD}}^2)}, \quad (1)$$

so that e.g. $\alpha_s(Q^2) \simeq 0.4$ when $Q = 2 \text{ GeV}$. Formally, Eq. (1) predicts that the coupling diverges when $Q = \Lambda_{\text{QCD}}$, but this equation cannot be trusted for $Q \lesssim 1 \text{ GeV}$, as it has been obtained in perturbation theory. The fate of the QCD coupling for $Q \sim \Lambda_{\text{QCD}}$ is still under debate, but various non-perturbative approaches suggest that $\alpha_s(Q^2)$ should (roughly) saturate at a value close to one. For all purposes, this is very strong coupling (e.g. it corresponds to $g \simeq 3$).

After this digression through the general scope of an effective theory and the QCD running coupling, let us return to the main stream of our presentation, namely, the phases of QCD as probed in a HIC. Some key ideas, that will be succinctly mentioned here and developed in more detail in the remaining part of these lectures, are as follows:

(i) The different stages of a HIC involve different forms of hadronic matter with specific active degrees of freedom. Their theoretical description requires different effective theories.

(ii) During the early stages of the collision — the colour glass condensate and the glasma — the parton density is very high, the typical transverse momenta are semi-hard (a few GeV), and the QCD coupling is moderately weak, say $\alpha_s \sim 0.3$. In this case, perturbation theory is (at least, marginally)

valid, but it goes beyond a straightforward expansion in powers of α_s . To construct the corresponding effective theory, one needs to resum an infinite class of Feynman graphs which are enhanced by high-energy and high gluon density effects. This has been done in the recent years, led to a formalism — *the CGC effective theory* — which offers a unified description from first principles for both the nuclear wavefunctions prior to the collision and the very early stages of the collision. A key ingredient in this construction is the proper recognition of the relevant d.o.f. : *quasi-classical colour fields*. The concept of *field* is indeed more useful in this high-density environment than that of *particle*, since the phase-space occupation numbers are large ($\gg 1$), meaning that the would-be ‘particles’ overlap with each other and thus form coherent states, which are more properly described as classical field configurations.

(iii) At later stages, the partonic matter expands, the phase-space occupation numbers decrease, and the concept of *particle* becomes again meaningful: the classical fields break down into particles. If these particles are weakly coupled (as one may expect by continuity with the previous stages), then their subsequent evolution can be described by *kinetic theory*. This is an effective theory which emerges under the assumption that the mean free path between two successive collisions is much longer than any other microscopic scale (like the duration of a collision or the Compton wavelength $\lambda = 1/k_\perp$ of a particle). Over the last years, kinetic theory has been extensively derived from QCD at weak coupling, but the results appear to be deceiving: for instance, they cannot explain the rapid thermalization suggested by the data at RHIC and the LHC. (The thermalization times predicted by perturbative QCD are much larger, $\tau \gtrsim 10$ fm/c.) Several alternative solutions have been proposed so far, but the final outcome is still unclear. One of these proposals is that the softer modes, which keep large occupation numbers and should be better described as classical fields, become *unstable* due to the anisotropy in the momentum distribution of the harder particles (which in turns follows from the disparity between longitudinal and transverse expansions). But numerical simulations of the coupled system soft fields-hard particles leads to thermalization times which are still too large. Another suggestion is that the partonic matter is (moderately) *strongly coupled* — the QCD coupling could indeed become larger, because of the system being more dilute. In such a scenario, a candidate for an effective theory is the *AdS/CFT correspondence*, to be discussed later.

(iv) Assuming (local) thermal equilibrium, and hence the formation of a *quark-gluon plasma* (QGP), the question is whether this plasma is weakly or strongly coupled. The maximal temperature of this plasma, as estimated from the average energy density, should be around $T \sim 500 \div 600$ MeV; so the respective coupling is moderately strong: $\alpha_s \sim 0.3 \div 0.4$ or $g \sim 1.5 \div 2$. The thermodynamic properties (like pressure or energy density) of a QGP at *global* thermal equilibrium within this range of temperatures are by now well known from numerical calculations on a lattice and can serve as a baseline of comparison for various effective theories. If the coupling is weak, one has to use the *Hard Thermal Loop effective theory* (HTL), a version of the kinetic theory which describes the long-range (or ‘soft’) excitations of the QGP. This effective theory lies at the basis of a physical picture of the QGP as a gas of *weakly-coupled quasi-particles* — quarks or gluons with temperature-dependent effective masses and couplings. Using this picture as a guideline for reorganizations of the perturbation theory, one has been able to reproduce the lattice data quite well. Thus, the thermodynamics appears to be consistent with a weak-coupling picture for the QGP, although this picture is considerably more complicated than that emerging from naive perturbation theory (the strict expansion in powers in g). Yet, this is not the end of the problem, as we shall see.

(v) The QGP created in the intermediate stages of a HIC is certainly not in *global* thermal equilibrium, but only in a *local* one: it keeps expanding. Under very general assumptions, the effective theory describing this flow is *hydrodynamics*. The corresponding equations of motion are simply the conservation laws for energy, momentum, and other conserved quantities (like the electric charge or the baryonic number), and as such they are universally valid. But these equations also involve ‘parameters’, like the *viscosities*, which describe dissipative phenomena occurring during the flow and which depend upon the specific microscopic dynamics. The values of these parameters are very different at weak vs.

strong coupling. A meaningful way to characterize the strength of dissipation is via the dimensionless viscosity-over-entropy-density ratio η/s . (This ratio is dimensionless when using natural units; otherwise it has the dimension of $\hbar$.) Remarkably, the *elliptic flow* data at RHIC and the LHC to be later discussed suggest a very small value for this ratio, which is inconsistent with the present calculations at weak coupling, based on kinetic theory. On the other hand, such a small ratio is naturally emerging at strong coupling, as shown by calculations within the AdS/CFT correspondence. The smallness of η/s represents so far the strongest argument in favour of a *strongly coupled quark-gluon plasma* (sQGP). This may look contradictory with the previous conclusions drawn from thermodynamics. But one should remember that the QCD coupling depends upon the relevant space-time scale and that hydrodynamics refers to the long-range behaviour of the fluid, as encoded in its softest modes. By contrast, thermodynamics is rather controlled by the hardest modes — those with typical energies and momenta of the order of the (local) temperature. So, it is not inconceivable that a same system look effectively weakly coupled for some phenomena and strongly coupled for some others.

(vi) Another strategy for studying the hadronic matter produced in a HIC refers to the use of *hard probes*. These are particles with large transverse energies (say, $E_\perp \gtrsim 20$ GeV at the LHC), which are produced in the very first instants of a collision and then cross the QCD matter liberated at later stages along their way towards the detector. Some of these particles, like the (direct) photons and the dilepton pairs, do not interact with this matter and hence can be used as a baseline for comparisons. But other particles, like quarks, gluons, and the jets initiated by them, *do* interact, and by measuring the effects of these interactions — say, in terms of energy loss, or the suppression of multi-particle correlations — one can infer informations about the properties of the matter they crossed. The RHIC data have demonstrated that semi-hard partons can lose a substantial fraction of the transverse energy via interactions in the medium (‘jet quenching’), thus suggesting that these interactions can be quite *strong*. These results have been confirmed at the LHC, which moreover found that even *very* hard jets ($E_\perp \gtrsim 100$ GeV) can be strongly influenced by the medium, in the sense that they get strongly defocused : the energy distribution in the polar angle with respect to the jet axis becomes much wider after having crossed the medium. This is visible for the photon-hadron di-jet event in the right panel of Fig. 2 : the photon and the parton which has initiated the hadronic jet have been created by a hard scattering, so they must have been balanced in transverse momentum at the time of their creation. Yet, the central peak in the hadronic jet, which represents the final ‘jet’ according to the conventional definition, carries much less energy than the photon jet. This is interpreted as the result of energy transfer to large polar angles (outside the conventional ‘jet’ definition) via in-medium interactions. In order to study such interactions, in particular high-density effects like *multiple scattering* and *coherence*, it is again useful to build effective theories. In that case too, it is not so clear whether the physics is controlled by *mostly weak coupling*, or a *mostly strong* one. (By itself, the jet is hard, but its coupling to the relatively soft constituents of the medium may be still governed by a moderately strong coupling.) In fact, the unexpectedly strong jet quenching observed at RHIC is sometimes interpreted as another evidence for strong coupling behaviour. *Moderately* strong coupling turns out to be the most difficult situation to deal with, so in these lectures we shall rather describe the effective theories proposed in the limiting situations of weak and, respectively, strong coupling. Within perturbative QCD, this is known as *medium-induced gluon radiation*. At strong coupling, it again relies on AdS/CFT.

3 The Color Glass Condensate

This chapter is devoted to the *early stages* of an ultrarelativistic heavy ion collision (HIC), that is, the wavefunctions of the energetic nuclei prior to the collision and the partonic matter liberated by the collision. As already mentioned, these early stages are the realm of high-density, coherent, forms of QCD matter, characterized by high gluon occupation numbers. Such forms of matter can be described in terms of *strong, semi-classical, colour fields*. In what follows, we shall explain this theoretical description, starting with the perhaps more familiar *parton picture* of QCD scattering at high energy.

3.1 The QCD parton picture

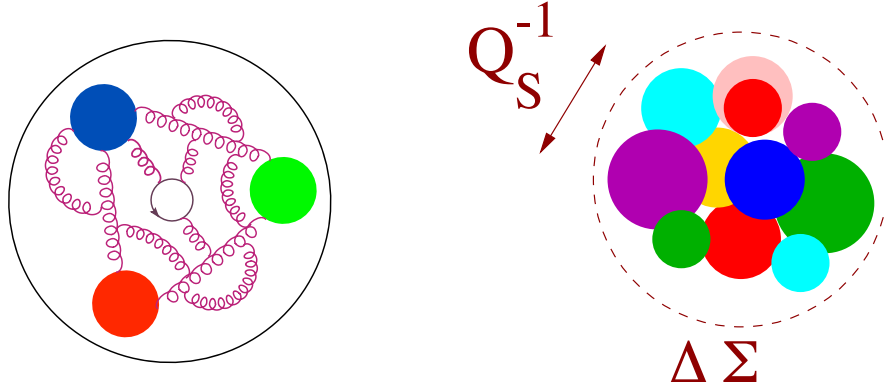


Fig. 3: Left: a cartoon of the proton structure in its rest frame. Right: cartoon of one saturation disk in the infinite momentum frame (this will be discussed in Section 3.7).

The microscopic structure of a hadron depends upon the resolution scales which are used to probe it, that is, upon the kinematics of the scattering process. It furthermore depends upon the Lorentz frame in which the hadron is seen: unlike physical observables, like cross-sections, which are boost invariant, the physical interpretation of these observables in terms of partons depends upon the choice of a frame. This is best appreciated by first looking at a hadron (say, a proton) in its rest frame (RF), where the proton 4-momentum reads $P_0^\mu = (M, 0, 0, 0)$. The proton has the quantum numbers of a system of three quarks — the ‘valence quarks’ — which are bound by confinement in a colour singlet state. But this binding proceeds via the exchange of gluons, which in turn can generate additional quark-antiquark pairs (see Fig. 3). All these partons are ‘virtual’, meaning that they keep appearing and disappearing, and have typical energies and momenta of order Λ_{QCD} , since this is the scale where the QCD coupling becomes of $\mathcal{O}(1)$ and thus the binding is most efficient. Clearly, such fluctuations are non-perturbative. Λ_{QCD} is also the typical scale for vacuum fluctuations, like a quark-antiquark pair pumping up from the vacuum and then being reabsorbed. By the uncertainty principle, such fluctuations have lifetimes and sizes of order $1/\Lambda_{\text{QCD}}$, of the same order as the proton size itself. Under these conditions, it makes no sense to speak about ‘hadronic substructure’: the hadronic fluctuations are ephemeral, delocalized over the whole proton volume, and cannot be distinguished from the vacuum fluctuations having the same kinematics and quantum numbers.

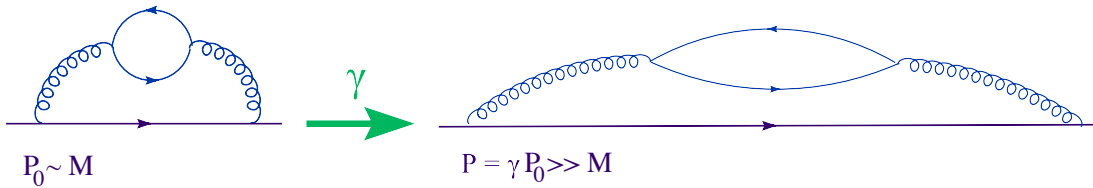


Fig. 4: A hadronic fluctuation in the hadron rest frame (left) and in the infinite momentum frame (right).

However, the situation changes if one observes the same hadron in a frame which is boosted by a large Lorentz factor $\gamma \gg 1$ w.r.t. the rest frame. Then the hadron 4-momentum reads $P^\mu = (E, 0, 0, P)$ with $E = \sqrt{P^2 + M^2} \simeq P$. (We have chosen the boost along the z axis and denoted $P_z = P$.) In this boosted frame, conventionally referred to as the *infinite momentum frame* (IMF), the lifetime of the hadronic fluctuations is enhanced by Lorentz time dilation (see Fig. 4),

$$\Delta t_{\text{IMF}} = \gamma \Delta t_{\text{RF}} \sim \frac{\gamma}{\Lambda_{\text{QCD}}}, \quad (2)$$

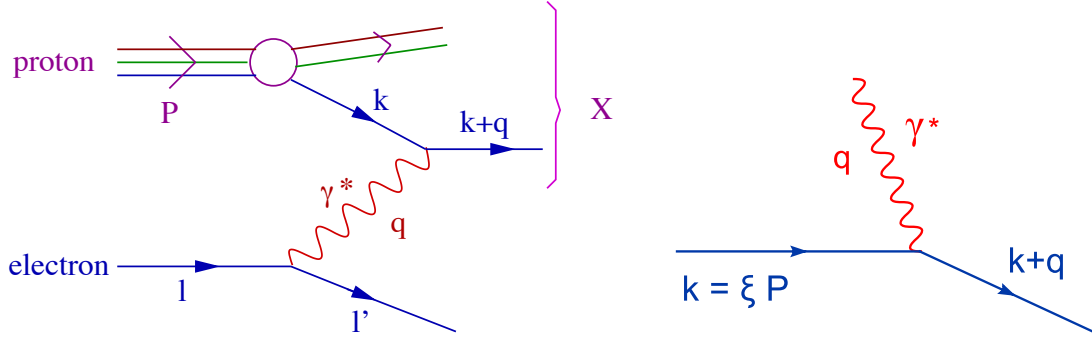


Fig. 5: Left: the DIS process. Right: the absorption of the virtual photon by a quark.

so these fluctuations are now well separated from the those of the vacuum (which have a lifetime $\sim 1/\Lambda_{\text{QCD}}$ in any frame, since the vacuum is boost invariant). The lifetime (2) is much larger than the duration of a typical collision process (see below); so, for the purpose of scattering, the hadronic fluctuations can be viewed as *free, independent quanta*. These quanta are the *partons* (a term coined by Feynman). It then becomes possible to *factorize* the cross-section (say, for a hadron-hadron collision) into the product of *parton distribution functions* (one for each hadron partaking in the collision), which describe the probability to find a parton with a given kinematics inside the hadronic wavefunction, and *partonic cross-sections*, which, as their name indicates, describe the collision between subsets of partons from the target and the projectile, respectively. If the momentum transferred in the collision is hard enough, the partonic cross-sections are computable in perturbation theory. The parton distributions are *a priori* non-perturbative, as they encode the information about the binding of the partons within the hadron. Yet, there is much that can be said about them within perturbation theory, as we shall explain. To that aim, one needs to better appreciate the role played by the *resolution* of a scattering process. In turn, this can be best explained on the example of a simpler process: the electron-proton *deep inelastic scattering* (DIS).

The DIS process is illustrated in Fig. 5 (left): an electron with 4-momentum ℓ_μ scatters off the proton by exchanging a virtual photon (γ^*) with 4-momentum q_μ and emerges after scattering with 4-momentum $\ell'_\mu = \ell_\mu - q_\mu$. The exchanged photon is *space-like* :

$$q^2 = (\ell - \ell')^2 = -2\ell \cdot \ell' = -2E_\ell E_{\ell'}(1 - \cos \theta_{\ell\ell'}) \equiv -Q^2 \quad \text{with } Q^2 > 0, \quad (3)$$

with $E_\ell = |\ell|$, $E_{\ell'} = |\ell'|$, and $\theta_{\ell\ell'} = \angle(\ell, \ell')$. The positive quantity Q^2 is referred to as the ‘virtuality’. The *deeply inelastic* regime corresponds to $Q^2 \gg M^2$, since in that case the proton is generally broken by the scattering and its remnants emerge as a collection of other hadrons (denoted by X in Fig. 5). The (inclusive) DIS cross-section involves the sum over all the possible proton final states X for a given ℓ' .

A space-like probe is very useful since it is well localized in space and time and thus provides a snapshot of the hadron substructure on controlled, transverse and longitudinal, scales, as fixed by the kinematics. Specifically, we shall argue that, when the scattering is analyzed in the proton IMF, the virtual photon measures partons which are localized in the transverse plane within an area $\Sigma \sim 1/Q^2$ and which carry a longitudinal momentum $k_z = xP$, where x is the *Bjorken variable* :

$$x \equiv \frac{Q^2}{2(P \cdot q)} = \frac{Q^2}{s + Q^2 - M^2}, \quad (4)$$

where $s \equiv (P + q)^2$ is the invariant energy squared of the photon+proton system. That is, the two kinematical invariants Q^2 and x , which are fixed by the kinematics of the initial state (ℓ , P) and of the scattered electron (ℓ'), completely determine the transverse size ($\sim 1/Q$) and the *longitudinal momentum fraction* (x) of the parton that was involved in the scattering. This parton is necessarily a quark (or

antiquark), since the photon does not couple directly to gluons. But the DIS cross-section allows us to indirectly deduce also the gluon distribution, as we shall see.

As a first step in our argument, consider a quark excitation of the hadron, viewed in the IMF. This quark is a virtual fluctuation which has been boosted together with the proton, so its virtuality and its transverse momentum are both small as compared to its longitudinal momentum $k_z = \xi P$. (We temporarily denote with ξ the fraction of the proton longitudinal momentum which is carried by the quark.) So, for most purposes, one can treat the quark as a nearly on-shell excitation with 4-momentum as $k^\mu \simeq \xi P^\mu = (\xi P, 0, 0, \xi P)$ and $k^2 \equiv k_\mu k^\mu \simeq 0$. (More precisely, $k^2 \simeq k_\perp^2 \sim \Lambda_{\text{QCD}}^2$.) Such an excitation has a relatively large lifetime, which can be estimated as in Eq. (2) :

$$\Delta t_{\text{fluct}} = \gamma \Delta t_{\text{RF}} \simeq \frac{2k_z}{k_\perp^2} = \frac{2\xi P}{k_\perp^2}, \quad (5)$$

where $\Delta t_{\text{RF}} \sim 2/k_\perp$ is the lifetime of the fluctuation in the hadron rest frame and $\gamma = k_z/k_\perp$ is the boost factor from the RF to the IMF.

Consider now the absorption of the virtual photon by the quark, cf. Fig. 5 right. The quark is liberated by this collision, meaning that it is put on shell; so we can write

$$(k + q)^2 = 0 \implies -Q^2 + 2\xi P \cdot q = 0 \implies \xi = \frac{Q^2}{2(P \cdot q)} = x, \quad (6)$$

where we have also used $k^2 \approx 0$, as discussed before. We see that the collision identifies the longitudinal momentum fraction ξ of the participating quark with the Bjorken- x kinematical variable, as anticipated. From now on, we shall use the notation x for both quantities.

To also clarify the *transverse resolution* of the virtual photon, we first need an estimate for the *collision time*. This is the typical duration of the partonic process $q + \gamma^* \rightarrow q$ (cf. Fig. 5 right) and is given by the uncertainty principle: $\Delta t_{\text{coll}} \sim 1/\Delta E$, where $\Delta E = q_0 + |\mathbf{k} + \mathbf{q}| - |\mathbf{k}|$ is the energy difference at the photon emission vertex. To estimate ΔE , it is convenient to choose a space-like photon with zero energy and only transverse momentum: $q^\mu = (0, \mathbf{q}_\perp, 0)$. Then

$$\Delta E = |\mathbf{k} + \mathbf{q}| - |\mathbf{k}| = \sqrt{(xP)^2 + q_\perp^2} - xP \simeq \frac{q_\perp^2}{2xP} \implies \Delta t_{\text{coll}} \simeq \frac{2xP}{Q^2}. \quad (7)$$

(Note that $Q^2 = q_\perp^2$ for the virtual photon at hand.) In order to be ‘found’ by the photon, a quark excitation must have a lifetime larger than this collision time:

$$\Delta t_{\text{fluct}} \simeq \frac{2xP}{k_\perp^2} > \Delta t_{\text{coll}} \simeq \frac{2xP}{Q^2} \implies k_\perp^2 < Q^2. \quad (8)$$

Hence, the virtual photon can discriminate only those partons having transverse momenta smaller than its virtuality Q . By the uncertainty principle, such partons are localized within a transverse area $\sim 1/Q^2$, as anticipated after Eq. (4).

The previous considerations motivate the following formula for the DIS cross-section :

$$\sigma_{\gamma^* p}(x, Q^2) = \frac{4\pi^2 \alpha_{\text{em}}}{Q^2} F_2(x, Q^2), \quad (9)$$

where the first factor in the r.h.s. is the elementary cross-section for the photon absorption by a quark (or an antiquark), whereas the second factor — the *structure function* $F_2(x, Q^2)$ — is the sum of the *quark and antiquark distribution functions*, weighted by the respective electric charges squared

$$F_2(x, Q^2) = \sum_f e_f^2 [xq_f(x, Q^2) + x\bar{q}_f(x, Q^2)],$$

$$q_f(x, Q^2) \equiv \frac{dN_f}{dx}(Q^2) = \int^Q d^2\mathbf{k}_\perp \frac{dN_f}{dx d^2\mathbf{k}_\perp}. \quad (10)$$

That is, $q_f(x, Q^2)dx$ is the number of quarks of flavor f with longitudinal momentum fraction between x and $x + dx$ and which occupy a transverse area $1/Q^2$.

One may naively think that the condition $k_\perp^2 < Q^2$ is trivially satisfied, since the partons confined inside the hadron have transverse momenta $k_\perp \sim \Lambda_{\text{QCD}}$, whereas $Q^2 \gg \Lambda_{\text{QCD}}^2$ by the definition of DIS. If that were the case, the structure function $F_2(x, Q^2)$ would be independent of Q^2 — a property known as *Bjorken scaling*. However, the DIS data show that Bjorken scaling holds only approximately and only in a limited range of values for x , namely for $x \gtrsim 0.1$. This can be understood as follows : the typical transverse momenta are $\sim \Lambda_{\text{QCD}}$ only for the valence quarks and, more generally, for the partons with relatively large longitudinal momentum fractions. But virtual quanta with much larger values for $k_\perp$ can be generated via radiative processes like *bremssstrahlung*. Such quanta have very short lifetimes, but so long as $k_\perp^2 < Q^2$, they can still contribute to DIS. Also, they generally have small values of x , as they share all together the longitudinal momenta of their parents partons. Hence, we expect the *parton evolution* via *bremssstrahlung* to lead to an increase in the parton distributions at large values of $k_\perp$ and small values of x . This evolution is responsible for the violations of the Bjorken scaling seen in the data and, more generally, for the *DGLAP evolution* [1, 2] of the parton distribution functions with increasing Q^2 . It is furthermore responsible for the rapid growth in the gluon distribution with decreasing x and the formation of a *colour glass condensate* at high energy. This will be further discussed in Section 3.3.

3.2 Particle production at the LHC: why small x ?

Before we turn to a discussion of parton evolution, let us explain here why we shall be mostly interested in partons with small longitudinal momentum fractions $x \ll 1$. As it should be clear from Eq. (4), small values of x correspond to the *high-energy* regime at $s \gg Q^2$. The conceptual importance of this regime will be explained later, but for the time being let us discuss it from the experimental point of view. Very small values of x , as low as $x = 10^{-5} \div 10^{-4}$, have been already reached in the e+p collisions at HERA, but in that context they were associated (because of the experimental constraints) with rather small values of the transferred momentum Q^2 . Namely, the HERA data at $x \leq 10^{-4}$ correspond to values $Q^2 < 1 \text{ GeV}^2$ which are only marginally under control in perturbation theory. Because of that, the DIS data at HERA remained inconclusive for a check of our theoretical understanding of the physics at small x .

But the situation has changed with the advent of the new hadron-hadron colliders, RHIC and, especially, the LHC. Given the much higher available energies, the bulk of the particle production (with semi-hard transverse momenta) in these experiments is controlled by partons with $x \leq 10^{-3}$. Moreover, for special kinematical conditions to be shortly specified, one can probe values as low as $x \sim 10^{-6}$ with truly hard momentum transfers, such as $Q^2 = 10 \text{ GeV}^2$.

To describe the kinematics of particle production, it is useful to introduce a new kinematical variable, the *rapidity* y , which is an alternative for the longitudinal momentum. For an on-shell particle with 4-momentum $p^\mu = (E, \mathbf{p}_\perp, p_z)$, the rapidity is defined as

$$y \equiv \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \implies E = m_\perp \cosh y, \quad p_z = m_\perp \sinh y, \quad (11)$$

where $m_\perp \equiv \sqrt{m^2 + p_\perp^2}$ is the ‘transverse mass’ and $E^2 = m_\perp^2 + p_z^2$. Note that y is positive for a ‘right-mover’ ($p_z > 0$) and negative for a ‘left-mover’ ($p_z < 0$). In fact, one has $v_z = p_z/E = \tanh y$, so y is simply related to the longitudinal boost factor: $\gamma = \cosh y$. A similar quantity which is perhaps more useful in the experiments (since easier to measure) is the *pseudo-rapidity*

$$\eta \equiv \frac{1}{2} \ln \frac{p + p_z}{p - p_z} = -\ln \tan \frac{\theta}{2}, \quad (12)$$

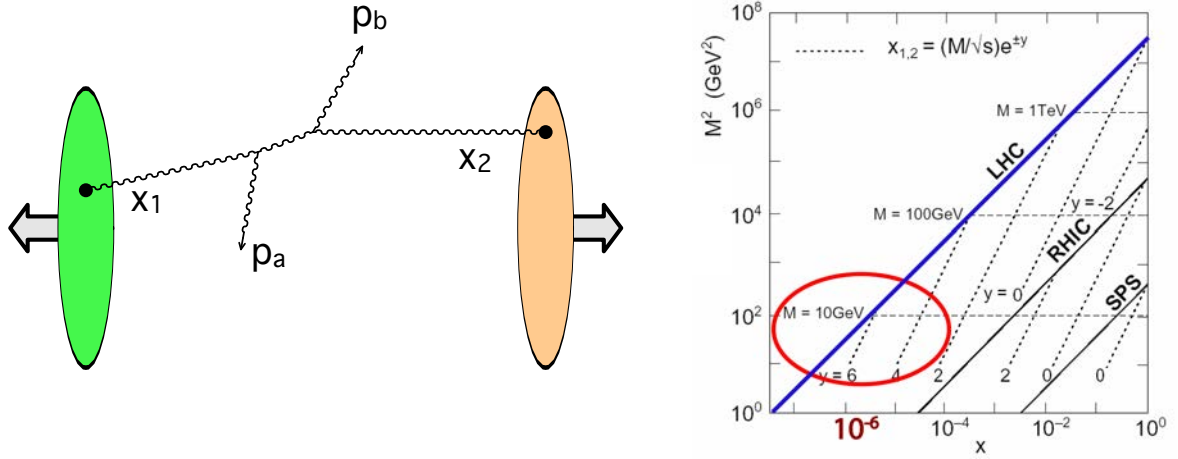


Fig. 6: Left: two particle production in hadron-hadron collision. Right: the kinematical domain accessible at the LHC, as compared to RHIC and the SPS; the small- x region is highlighted.

where $p = |\mathbf{p}|$ is the magnitude of the 3-momentum vector. As shown by the second equality above, η is directly related to the polar angle θ made by the particle with the longitudinal axis ($\cos \theta = p_z/p$). For massless particles or for ultrarelativistic ones (whose masses can generally be ignored), the two rapidities coincide with each other, as manifest by comparing Eqs. (11) and (12).

Consider now the process illustrated in Fig. 6 (left), i.e. the production of a pair of particles in a partonic subcollision of a hadron-hadron scattering. In the center-of-mass (COM) frame, the two partons partaking in the collision have 4-momenta $k_i^\mu = x_i P_i^\mu + k_{i\perp}^\mu$ where $i = 1, 2$, $P_1^\mu = (P, 0, 0, P)$, $P_2^\mu = (P, 0, 0, -P)$, and $k_{i\perp}^\mu = (0, \mathbf{k}_{i\perp}, 0)$. Notice that $P = \sqrt{s}/2$. The two outgoing particles will be characterized by the respective transverse momenta, $\mathbf{p}_{a\perp}$ and $\mathbf{p}_{b\perp}$, and rapidities, y_a and y_b . Energy-momentum conservation implies $\mathbf{p}_{a\perp} + \mathbf{p}_{b\perp} = \mathbf{k}_{1\perp} + \mathbf{k}_{2\perp}$ and

$$x_1 = \frac{p_{a\perp}}{\sqrt{s}} e^{y_a} + \frac{p_{b\perp}}{\sqrt{s}} e^{y_b}, \quad x_2 = \frac{p_{a\perp}}{\sqrt{s}} e^{-y_a} + \frac{p_{b\perp}}{\sqrt{s}} e^{-y_b}. \quad (13)$$

For particle production at RHIC or the LHC, the average transverse momentum of a hadron in the final state is below 1 GeV ; moreover, 99% of the ‘multiplicity’ (i.e. of the total number of produced hadrons) has $p_\perp \leq 2$ GeV (see Fig. 7). For $p_{a,b\perp} = 1$ GeV and central rapidities $y_{a,b} \simeq 0$, Eq. (13) implies

$$x_i \simeq 10^{-2} \text{ at RHIC } (\sqrt{s} = 200 \text{ GeV}), \quad x_i \simeq 4 \times 10^{-4} \text{ at the LHC } (\sqrt{s} = 2.76 \text{ TeV}), \quad (14)$$

where in the case of the LHC we have chosen the maximal COM energy *per nucleon pair* that has been reached so far in Pb+Pb collisions. Thus, the bulk of the particle production is initiated by partons carrying small values of x , as anticipated. Moreover, one of these values (x_1 or x_2) can be made much smaller by studying particle production at either *forward*, or *backward*, rapidities. The rapidities are ‘forward’ when both y_a and y_b are positive and relatively large, that is, the final particles propagate essentially along the same direction as the original hadron ‘1’ ; then their production probes very small values of x_2 in the wavefunction of the hadron ‘2’ and comparatively large values of x_1 . At the LHC, one can probe values as small as $x_2 \sim 10^{-6}$ for $p_\perp \sim 10$ GeV, as indicated in the r.h.s. of Fig. 6.

We finally discuss the cross-section for the production of a pair of hadrons. When the transverse momenta $p_{a,b\perp}$ are large enough, one can ignore the ‘intrinsic’ transverse momenta $\mathbf{k}_{1,2\perp}$ of the colliding partons. Then the transverse momentum conservation $\mathbf{p}_{a\perp} + \mathbf{p}_{b\perp} \simeq 0$ implies that the outgoing particles propagate back-to-back in the transverse plane, i.e. they make an azimuthal angle $\Delta\phi \simeq \pi$. The

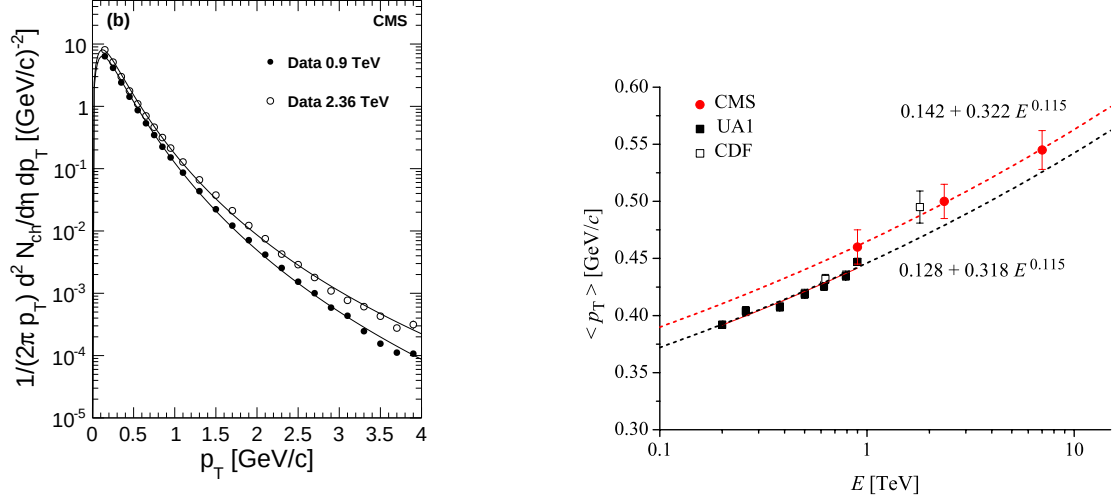


Fig. 7: Transverse momentum dependence of the single inclusive particle production in hadron-hadron collisions. Left: Pb+Pb collisions at the LHC. Right: p+p collisions at the LHC and p+ $\bar{p}$ collisions at SPS and CDF.

associated cross-section admits the following *collinear factorization*, analogous to Eq. (10) for DIS

$$\frac{d\sigma}{dp_\perp^2 dy_1 dy_2} = \sum_{ij} x_1 f_i(x_1, \mu^2) x_2 f_j(x_2, \mu^2) \frac{d\hat{\sigma}_{ij}}{dp_\perp^2}, \quad (15)$$

where $x f_i(x, \mu^2)$ are parton distributions for all species of partons ($i = q, \bar{q}, g$), μ^2 is the factorization scale, and $d\hat{\sigma}_{ij}/dp_\perp^2$ is the cross-section for the (relatively hard) partonic process $i + j \rightarrow a + b$. Leading-order perturbative QCD yields $d\hat{\sigma}/dp_\perp^2 \propto \alpha_s^2/p_\perp^4$ at high energy. So, if one tries to compute the total multiplicity by integrating over all values of $p_\perp^2$ (say, for $y_a \sim y_b \sim 0$), then one faces a quadratic infrared divergence from the limit $p_\perp^2 \rightarrow 0$. One may think that this divergence is cut off at $p_\perp \sim \Lambda_{\text{QCD}}$, since this is the typical value expected for the intrinsic momenta $k_{1,2\perp}$. But then one would conclude that the bulk of the particle production, even at very high energy, is concentrated at very soft transverse momenta, of the order of the confinement scale Λ_{QCD} . Moreover, the average $p_\perp$ would be independent of the energy (since of $\mathcal{O}(\Lambda_{\text{QCD}})$). These conclusions are however contradicted by the data in Fig. 7, which rather show that $\langle p_\perp \rangle \simeq 0.5 \text{ GeV}$ is about 2 to 3 times larger than Λ_{QCD} at the LHC energies and, remarkably, it clearly rises with the COM energy $E = \sqrt{s}$. This conflict between the data and the prediction (15) of collinear factorization clearly shows that the latter cannot be extrapolated down to lower values for $p_\perp$, say of order 1 GeV. The proper way to describe this *semi-hard* region within (perturbative) QCD will be explained in the next subsection. The main outcome of that analysis will be to introduce a new infrared cutoff in the problem, which is dynamically generated — via gluon evolution with decreasing x — and rises as a power of the energy. This is the *saturation momentum*.

3.3 Gluon evolution at small x

In perturbative QCD, parton evolution proceeds via bremsstrahlung, which favors the emission of *soft* and *collinear* gluons, i.e. gluons which carry only a small longitudinal momentum fraction $x \ll 1$ and a relatively small transverse momentum $k_\perp$. Fig. 8 illustrates one elementary step in this evolution: the emission of a gluon which carries a fraction $x = k_z/p_z$ of the longitudinal momentum of its parent parton (quark or gluon). For $x \ll 1$ and to lowest order in α_s , the differential probability for this emission (obtained as the modulus squared of the amplitude represented in Fig. 8) reads

$$dP_{\text{Brem}} \simeq C_R \frac{\alpha_s(k_\perp^2)}{\pi^2} \frac{d^2 k_\perp}{k_\perp^2} \frac{dx}{x}, \quad (16)$$

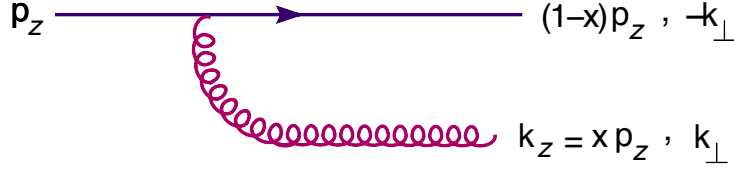


Fig. 8: Gluon bremsstrahlung out of a parent quark to lowest order in pQCD.

where C_R is the $SU(N_c)$ Casimir in the colour representation of the emitter: $C_A = N_c$ for a gluon and $C_F = (N_c^2 - 1)/2N_c$ for a quark. (N_c is the number of colours, which is equal to 3 in real QCD, but it is often kept as a free parameter in theoretical studies, because many calculations simplify in the formal limit $N_c \gg 1$. The results obtained in this limit provide insightful, qualitative and semi-quantitative informations about real QCD.) Eq. (16) exhibits the collinear ($k_\perp \rightarrow 0$) and soft ($x \rightarrow 0$) singularities mentioned above, which result in the enhancement of gluon emission at small $k_\perp$ and/or x . If the emitted parton with small x were a quark instead a gluon, there would be no small x enhancement, only the collinear one. This asymmetry, due to the spin-1 nature of the gluon, has the remarkable consequence that the small- x part of the wavefunction of any hadron is built mostly with gluons.

As manifest on Eq. (16), parton branching is suppressed by a power of $\alpha_s(k_\perp^2)$, which is small when $k_\perp \gg \Lambda_{\text{QCD}}$. But this suppression can be compensated by the large phase-space available for the emission, which equals $\ln(Q^2/\Lambda_{\text{QCD}}^2)$ for the emission of a parton (quark or gluon) with transverse momentum $k_\perp \ll Q$ and, respectively, $\ln(1/x)$ for that of a gluon with longitudinal momentum fraction ξ within the range $x \ll \xi \ll 1$. Hence, for large $Q^2 \gg \Lambda_{\text{QCD}}^2$ and/or small $x \ll 1$, such radiative processes are not suppressed anymore and must be resummed to all orders. Depending upon the relevant values of Q^2 and x , one can write down evolution equations which resum either powers of $\alpha_s \ln(Q^2/\Lambda_{\text{QCD}}^2)$, or of $\alpha_s \ln(1/x)$, to all orders. The coefficients in these equations represent the elementary splitting probability and can be computed as power series in α_s , starting with the leading-order result in Eq. (16).

The evolution with increasing Q^2 is described by the DGLAP equation (from Dokshitzer, Gribov, Lipatov, Altarelli and Parisi) [1, 2]. This evolution mixes quarks and gluons (see Fig. 9.a), which in particular allows us to reconstruct the gluon distribution from the experimental results for F_2 . The small- x evolution, on the other hand, involves only gluons and corresponds to resumming ladder diagrams like those in Fig. 9.b in which successive gluons are strongly ordered in x (see below). Both evolutions lead to an increase in the number of partons at small values of x (and a decrease at large values $x \gtrsim 0.1$), but the physical consequences are very different in the two cases:

(i) When increasing Q^2 , one emits partons which occupy a smaller transverse area $\sim 1/Q^2$, as shown in Fig. 11 (right). The decrease in the area of the individual partons is much stronger than the corresponding increase in their number. Accordingly, the *occupation number* in the transverse plane *decreases* with increasing Q^2 , meaning that the partonic system becomes *more and more dilute*. Accordingly, the partons may be viewed as *independent*. This observation lies at the basis of the conventional parton picture, which applies for sufficiently high Q^2 (at a given value of x).

The parton occupation number mentioned above yields the proper measure of the parton density in the hadron. It can be estimated as [*the number of partons with a given value of x*] $\times$ [*the area occupied by one parton*] divided by [*the transverse area of the hadron*], that is (for gluons, for definiteness),

$$n(x, Q^2) \simeq \frac{xg(x, Q^2)}{Q^2 R^2}, \quad \text{with} \quad g(x, Q^2) \equiv \frac{dN_g}{dx}(Q^2) = \int^Q d^2\mathbf{k}_\perp \frac{dN_g}{dx d^2\mathbf{k}_\perp}, \quad (17)$$

where R is the hadron radius in its rest frame (so its transverse area is $\sim \pi R^2$ in any frame). The numerator in the above definition of the occupation number, that is

$$xg(x, Q^2) \equiv x \frac{dN_g}{dx} = k_z \frac{dN_g}{dk_z} \simeq \frac{\Delta N_g}{\Delta z \Delta k_z}, \quad (18)$$

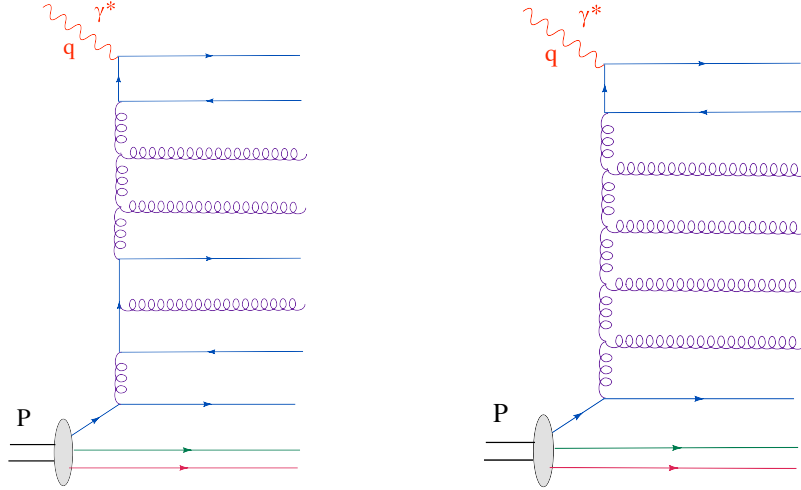


Fig. 9: Parton evolution in perturbative QCD. The parton cascade on the right involves only gluons (at intermediate stages) and is a part of the BFKL resummation at small x (see the discussion around Eq. (22)).

is known as the *gluon distribution*. The last estimate above follows from the uncertainty principle: partons with longitudinal momentum $k_z = xP$ are delocalized in z over a distance $\Delta z \simeq 1/k_z$. Hence, the gluon distribution yields the number of gluons *per unit of longitudinal phase-space*, which is indeed the right quantity for computing the occupation number. Note that gluons with $x \ll 1$ extends in z over a distance $\Delta z \sim 1/xP$ which is much larger than the Lorentz contracted width of the hadron, $R/\gamma \sim 1/P$. This shows that the image of an energetic hadron as a ‘pancake’, that would be strictly correct if the hadron was a classical object, is in reality a bit naive: it applies for the valence quarks with $x \sim \mathcal{O}(1)$ (which carry most of the total energy), but not also for the small- x partons (which are the most numerous, as we shall shortly see).

(ii) When decreasing x at a fixed Q^2 , one emits mostly gluons which have smaller longitudinal momentum fractions, but which occupy, roughly, the same transverse area as their parent gluons (see Fig. 11 right). Then the gluon occupation number, Eq. (17), *increases*, showing that *the gluonic system evolves towards increasing density*. As we shall see, this evolution is quite fast and eventually leads to a breakdown of the picture of independent partons.

In order to describe the small- x evolution, let us start with the gluon distribution generated by a single valence quark. This can be inferred from the bremsstrahlung law in Eq. (16) (the emission probability is the same as the number of emitted gluons) and reads

$$x \frac{dN_g}{dx}(Q^2) = \frac{\alpha_s C_F}{\pi} \int_{\Lambda_{\text{QCD}}^2}^{Q^2} \frac{dk_{\perp}^2}{k_{\perp}^2} = \frac{\alpha_s C_F}{\pi} \ln \left(\frac{Q^2}{\Lambda_{\text{QCD}}^2} \right), \quad (19)$$

where we have ignored the running of the coupling — formally, we are working to leading order (LO) in pQCD where the coupling can be treated as fixed — and the ‘infrared’ cutoff Λ_{QCD} has been introduced as a crude way to account for confinement: when confined inside a hadron, a parton has a minimum virtuality of $\mathcal{O}(\Lambda_{\text{QCD}}^2)$. In Eq. (19) it is understood that $x \ll 1$. In turn, the soft gluon emitted by the valence quark can radiate an even softer gluon, which can radiate again and again, as illustrated in Fig. 10. Each emission is formally suppressed by a power of α_s , but when the final value of x is tiny, the smallness of the coupling constant can be compensated by the large available phase-space, of order $\ln(1/x)$ per gluon emission. This evolution leads to an increase in the number of gluons with $x \ll 1$.

For a quantitative estimate, consider the first such correction, that is, the two-gluon diagram in Fig. 10 left: the region in phase-space where the longitudinal momentum fraction x_1 of the intermediate

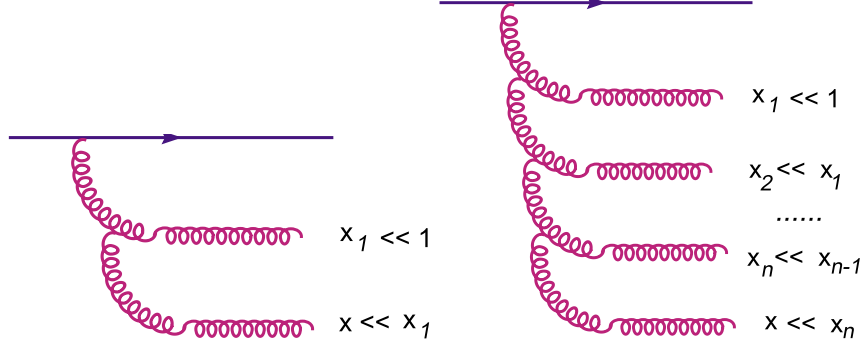


Fig. 10: Gluon cascades produced by the high-energy (BFKL) evolution of the proton wavefunction.

gluon obeys $x \ll x_1 \ll 1$ provides a contribution of relative order

$$\frac{\alpha_s N_c}{\pi} \int_x^1 \frac{dx_1}{x_1} = \bar{\alpha}_s \ln \frac{1}{x}, \quad \bar{\alpha}_s \equiv \frac{\alpha_s N_c}{\pi}. \quad (20)$$

When $\bar{\alpha}_s \ln(1/x) \sim 1$, this becomes of $\mathcal{O}(1)$, meaning that this two-gluon diagram contributes on the same footing as the single gluon emission in Fig. 8. A similar conclusion holds for a diagram involving n intermediate gluons *strongly ordered* in x , cf. Fig. 10 right, which yields a relative contribution of order

$$\bar{\alpha}_s^n \int_x^1 \frac{dx_n}{x_n} \int_{x_n}^1 \frac{dx_{n-1}}{x_{n-1}} \dots \int_{x_2}^1 \frac{dx_1}{x_1} = \frac{1}{n!} \left(\bar{\alpha}_s \ln \frac{1}{x} \right)^n. \quad (21)$$

When $\bar{\alpha}_s \ln(1/x) \gtrsim 1$, the correct result for the gluon distribution at leading order is obtained by summing contributions from all such ladders. As clear from Eq. (21), this sum *exponentiates*, modifying the integrand of Eq. (19) into

$$x \frac{dN_g}{dx dk_\perp^2} \sim \frac{\alpha_s C_F}{\pi} \frac{1}{k_\perp^2} e^{\omega \bar{\alpha}_s Y}, \quad Y \equiv \ln \frac{1}{x}, \quad (22)$$

where ω is a number of order unity which cannot be determined via such simple arguments. The variable Y is the rapidity difference between the final gluon and the original valence quark and it is often simply referred to as ‘the rapidity’. The quantity in the l.h.s. of Eq. (22) is the number of gluons per unit rapidity and with a given value $k_\perp$ for the transverse momentum, a.k.a. the *unintegrated gluon distribution*².

To go beyond this simple power counting argument, one must treat more accurately the kinematics of the ladder diagrams and include the associated virtual corrections. The result is the *BFKL equation* (from Balitsky, Fadin, Kuraev, and Lipatov) [3] for the evolution of the unintegrated gluon distribution with Y . The solution of this equation, which resums perturbative corrections $(\bar{\alpha}_s Y)^n$ to all orders, confirms the exponential increase in Eq. (22), albeit with a $k_\perp$ -dependent exponent and modifications to the $k_\perp^{-2}$ -spectrum of the emitted gluons.

An important property of the BFKL ladder is its *coherence in time*: the lifetime of a parton being proportional to its value of x , $\Delta t \simeq 2k_z/k_\perp^2 \propto x$, cf. Eq. (5), the ‘slow’ gluons at the lower end of the cascade have a much shorter lifetime than the preceding ‘fast’ gluons. Therefore, for the purposes of small- x dynamics, fast gluons with $x' \gg x$ act as *frozen colour sources emitting gluons at the scale x* . Because these sources may overlap in the transverse plane, their colour charges add coherently, giving rise to a large colour charge density. The *average* colour charge density is zero by gauge symmetry but *fluctuations* in the colour charge density — as measured in particular by the unintegrated gluon distribution — are nonzero and increase rapidly with $1/x$, cf. Eq. (22).

²The occupation number (17) is more correctly defined as the unintegrated gluon distribution per unit transverse area: $n(Y, \mathbf{k}_\perp) = dN_g/(dY d^2 \mathbf{k}_\perp d^2 \mathbf{b}_\perp)$ where $\mathbf{b}_\perp$ (the ‘impact parameter’) is the transverse position of a gluon with respect to the center of the hadron.

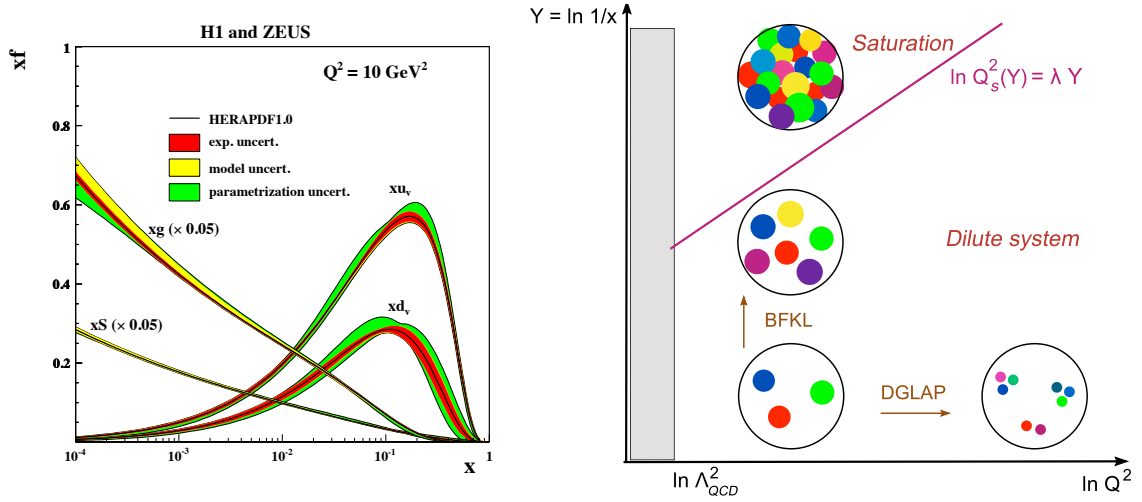


Fig. 11: Left: the $1/x$ -evolution of the gluon, sea quark, and valence quark distributions for $Q^2 = 10 \text{ GeV}^2$, as measured at HERA (combined H1 and ZEUS analysis [4]). Note that the gluon and sea quark distributions have been reduced by a factor of 20 to fit inside the figure. Right: the ‘phase-diagram’ for parton evolution in QCD; each coloured blob represents a parton with transverse area $\Delta x_\perp \sim 1/Q^2$ and longitudinal momentum $k_z = xP$. The straight line $\ln Q_s^2(x) = \lambda Y$ is the saturation line, cf. Eq. (25), which separates the dense and dilute regimes.

This growth is indeed seen in the data: e.g., the HERA data for DIS confirm that the proton wavefunction at $x < 0.01$ is totally dominated by gluons (see Fig. 11 left). However, on physical grounds, such a rapid increase in the gluon distribution cannot go on for ever (that is, down to arbitrarily small values of x). Indeed, the BFKL equation is *linear* — it assumes that the radiated gluons do not interact with each other, like in the conventional parton picture. While such an assumption is perfectly legitimate in the context of the Q^2 -evolution, which proceeds towards increasing diluteness, it eventually breaks down in the context of the Y -evolution, which leads to a larger and larger gluon density. As long as the gluon occupation number (17) is small, $n \ll 1$, the system is dilute and the mutual interactions of the gluons are negligible. When $n \sim \mathcal{O}(1)$, the gluons start overlapping, but their interactions are still weak, since suppressed by $\alpha_s \ll 1$. The effect of these interactions becomes of order one only when n is as large as $n \sim \mathcal{O}(1/\alpha_s)$. When this happens, non-linear effects (to be shortly described) become important and stop the further growth of the gluon distribution. This phenomenon is known as *gluon saturation* [5–7]. An important consequence of it is to introduce a new transverse-momentum scale in the problem, the *saturation momentum* $Q_s(x)$, which is determined by Eq. (17) together with the condition that $n \sim 1/\alpha_s$:

$$n(x, Q^2 = Q_s^2(x)) \sim \frac{1}{\alpha_s} \implies Q_s^2(x) \simeq \alpha_s \frac{xg(x, Q_s^2(x))}{R^2}. \quad (23)$$

Except for the factor α_s , the r.h.s. of Eq. (23) is recognized as the density of gluons per unit transverse area, for gluons localized within an area $\Sigma \sim 1/Q_s^2(x)$ set by the saturation scale. Gluons with $k_\perp \leq Q_s(x)$ are at saturation: the corresponding occupation numbers are large, $n \sim 1/\alpha_s$, but do not grow anymore when further decreasing x . Gluons with $k_\perp \gg Q_s(x)$ are still in a dilute regime: the occupation numbers are relatively small $n \ll 1/\alpha_s$, but rapidly increasing with $1/x$ via the BFKL evolution. The separation between the saturation (or dense, or CGC) regime and the dilute regime is provided by the *saturation line* in Fig. 11 right, to be further discussed below.

The microscopic interpretation of Eq. (23) can be understood with reference to Fig. 12 (left) : gluons which have similar values of x (and hence overlap in the longitudinal direction) and which occupy a same area $\sim 1/Q^2$ in the transverse plane can *recombine* with each other, with a cross-section $\sigma_{gg \rightarrow g} \simeq \alpha_s/Q^2$. After taking also this effect into account, the change in the gluon distribution in one step of the

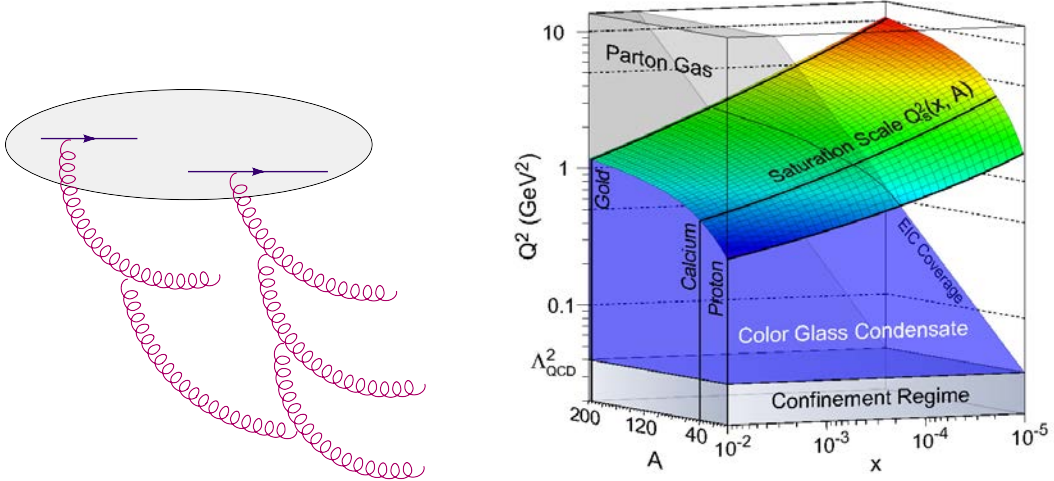


Fig. 12: Left: $gg \rightarrow g$ recombination process leading to saturation. Right: the saturation momentum $Q_s(x, A)$ as a function of the longitudinal momentum fraction x and of the atomic number A .

small- x evolution (i.e. under a rapidity increment $Y \rightarrow Y + dY$) can be schematically written as

$$\frac{\partial}{\partial Y} xg(x, Q^2) = \omega \bar{\alpha}_s xg(x, Q^2) - \bar{\alpha}_s \frac{\alpha_s}{Q^2 R^2} [xg(x, Q^2)]^2. \quad (24)$$

The overall factor of $\bar{\alpha}_s$ in the r.h.s. comes from the differential probability $\propto \bar{\alpha}_s dY$ to emit one additional gluon in this evolution step, cf. Eq. (16). The first term, linear in $xg(x, Q^2)$, represents the BFKL evolution; by itself, this would lead to the exponential growth with Y shown in Eq. (22). The second term, quadratic in $xg(x, Q^2)$, is the rate for recombination. This is formally suppressed by one factor α_s , but it becomes as important as the first term when Q^2 is of the order of the saturation momentum $Q_s^2(x)$ introduced in Eq. (23). When that happens, the r.h.s. of Eq. (24) vanishes, and then the gluon distribution stops growing with Y . The above argument, due to Gribov, Levin and Ryskin back in 1983 [5], is a bit oversimplified (and the actual evolution equation is considerably more complicated than Eq. (24); see the review papers [8–15] and the discussion in Section 3.4 below), but it has the merit to illustrate in a simple way the physical mechanism at work: the gluon occupation numbers saturate because the non-linear effects associated with the high gluon density compensate the bremsstrahlung processes.

Remarkably, Eq. (23) implies that the saturation momentum increases with $1/x$, since so does the gluon distribution for $k_\perp \gtrsim Q_s(x)$, cf. Eq. (22). So, for sufficiently small values of x (say, $x \leq 10^{-4}$ in the case of a proton), one expects $Q_s^2(x) \gg \Lambda_{\text{QCD}}^2$. In that case, the (semi)hard scale $Q_s(x)$ supplants Λ_{QCD} as an infrared cutoff for the calculation of physical observables like the multiplicity (cf. the discussion at the end of Section 3.2). This has the remarkable consequence that, for sufficiently high energy, the bulk of the particle production can be computed in perturbation theory. But the proper framework to perform this calculation is not standard pQCD as based on the collinear factorization, but the CGC effective theory which includes the non-linear physics of gluon saturation. This will be discussed in the next subsection.

Gluon occupancy is further amplified if instead of a proton we consider a large nucleus with atomic number $A \gg 1$. The corresponding gluon distribution $xg_A(x, Q^2)$ scales like A , since gluons can be radiated by any of the $3A$ valence quarks of the A nucleons. Since the nuclear radius scales like $R_A \sim A^{1/3}$, Eq. (17) implies that the gluon occupation number scales as $A^{1/3}$. This factor is about 6 for the Au and Pb nuclei respectively used at RHIC and the LHC. Thus, for a large nucleus, saturation effects become important at larger values of x than for a proton. This explains why ultrarelativistic heavy ion collisions represent a privileged playground for observing and studying the effects of saturation.

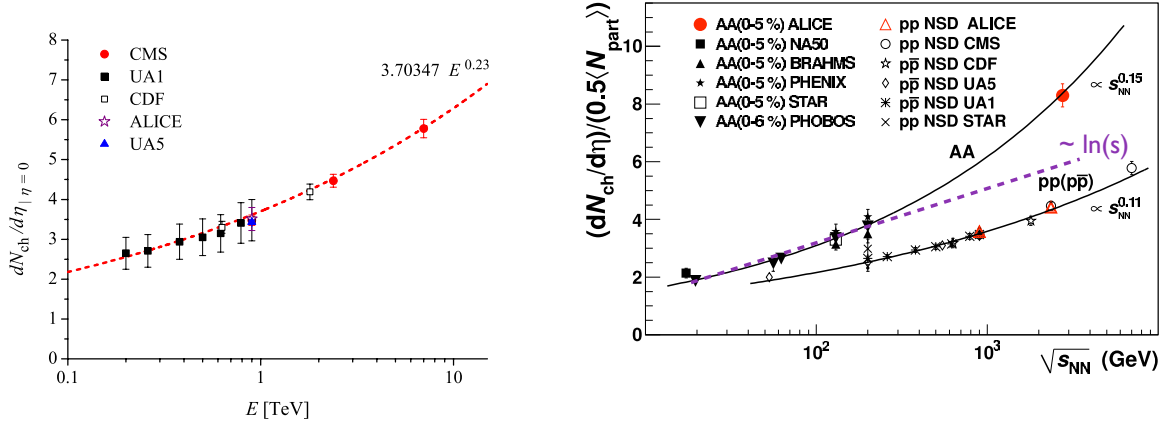


Fig. 13: Charged particles multiplicity at central (pseudo)rapidity $\eta = 0$ as a function of the COM energy for p+p and nucleus-nucleus collisions. The data are consistent with a power-law increase with $\sqrt{s}$ (with an exponent which is slightly larger for A+A than for p+p), but appear to exclude a logarithmic law $\propto \ln s$.

Figure 12 (right) summarizes our current expectations for the value and the variation of the saturation momentum. The dependence upon x is by now known to next-to-leading-order (NLO) accuracy [16] — that is, by resumming radiative corrections $\alpha_s [\alpha_s \ln(1/x)]^n$ to all orders together with non-linear effects. The result can be roughly expressed as

$$Q_s^2(x, A) \simeq Q_0^2 A^{1/3} \left(\frac{x_0}{x} \right)^\lambda, \quad \text{with } \lambda = 0.2 \div 0.3, \quad (25)$$

with the power λ known as the *saturation exponent*. The overall scale Q_0^2 , which has the meaning of the proton saturation scale at the original value x_0 , is non-perturbative and cannot be computed within the CGC effective theory. (The latter governs only the evolution from x_0 down to $x \ll x_0$.) In practice, this is treated as a free parameter which is fitted from the data. The fits yield $Q_0 \simeq 0.5 \text{ GeV}$ for $x_0 = 10^{-2}$. Figure 12 shows that for $x = 10^{-5}$ (a typical value for forward particle production at the LHC), $Q_s \simeq 1 \text{ GeV}$ for the proton, while $Q_s \simeq 3 \text{ GeV}$ for the Pb nucleus. This difference is significant: while 1 GeV is only marginally perturbative, 3 GeV is sufficiently ‘hard’ to allow for controlled perturbative calculations. This confirms the usefulness of HIC as a laboratory to study saturation.

Before we conclude this subsection, let us notice some robust predictions of the saturation physics, which do not require a detailed theory and can be directly checked against the data. One of them refers to the energy-dependence of the average transverse momentum of the produced particles: as shown in Fig. 7, this grows like a power of $E = \sqrt{s}$, with an exponent which is fitted from the data as 0.115. This is consistent with expectations based on gluon saturation [17]. Indeed, prior to the collision, the gluon distribution inside the hadron wavefunction is peaked at $k_\perp \sim Q_s$ (see Fig. 16 below and the related discussion) and these gluons are then released in the final state. We thus expect the average $p_\perp$ of the produced hadrons to scale like $Q_s(x)$ evaluated at the appropriate value of x , that is, $x = p_\perp/E$ (cf. Eq. (13)). This argument implies $\langle p_\perp \rangle \propto E^{\lambda/2}$, which is indeed consistent with the data in Fig. 7 (right) together with the estimate in (25) for the saturation exponent. Another prediction of this kind refers to the particle multiplicity in the final state $dN/d\eta$, say, at central (pseudo)rapidity $\eta = 0$. By the above argument, this is dominated by gluons with $Q^2 \simeq Q_s^2(x)$ and hence it is proportional to the respective gluon distribution, that is, to $Q_s^2(x)$ itself (cf. Eq. (23)) : $dN/d\eta \propto Q_s^2(E) \sim E^\lambda$. Once again, this appears to be consistent with the data for both p+p and A+A collisions, as shown in Fig. 13.

3.4 The CGC effective theory

The partonic form of matter made with the saturated gluons is known as the *colour glass condensate* [8–15].

- This is *coloured* since gluons carry the ‘colour’ charge of the non-Abelian group $SU(3)$.
- It is a *glass* because of the separation in time scales, due to Lorentz time dilation, between the ‘slow’ gluons at small x and their ‘fast’ sources at larger x . The sources appear as ‘frozen’ over the characteristic time scales for the dynamics at small x , but they can vary over much larger time scales, as set by their own, comparatively large, longitudinal momenta. A system which behaves as a solid on short time scales and as a fluid on much longer ones, is a glass.
- It is a *condensate* because the saturated gluons and their sources have high occupation numbers $n(x, k_\perp) \sim 1/\alpha_s$ and their colour charges add coherently to each other, as explained in Section 3.3 in relation with the BFKL ladder. A coherent quantum state with high occupancy can be in a first approximation described as a *classical field* (here, a colour field), which is the most generic example of a condensate.

Because of its high density, the CGC is *weakly coupled* and thus it can be studied within perturbative QCD. This is strictly correct for sufficiently small values of x , such that $Q_s^2(x) \gg \Lambda_{\text{QCD}}^2$ and hence $\alpha_s(Q_s^2) \ll 1$, but it remains marginally true for the phenomenology at RHIC and, especially, the LHC, where the saturation momentum is semi-hard, cf. Fig. 12 (right). Based on that, an *effective theory* has been explicitly constructed, which resums an infinite series of Feynman graphs of the ordinary perturbation theory — those which are enhanced by either the large logarithm $\ln(1/x)$, or by the high gluon density. This theory governs the dynamics of the gluons with a given, small, value of x , while the gluons at larger values $x' \gg x$ have been ‘integrated out’ in perturbation theory. In order to describe its mathematical structure, it is useful to recall that the gluon field in QCD is represented by a non-Abelian vector potential $A_a^\mu(x)$ where the upper index μ refers to the 4 Minkowski coordinates and the subscript a is a colour index in the adjoint representation of $SU(N_c)$ and can take $N_c^2 - 1 = 8$ values.

The CGC effective theory may be viewed as a non-linear generalization of the BFKL evolution, but in fact it is much more complex than just a non-linear evolution equation (say, like that in Eq. (24)). The BFKL equation applies to the unintegrated gluon distribution (or occupation number), which is a Fourier transform of the 2-point function³ $\langle A_a^i(x) A_a^i(y) \rangle$ of the colour fields within the hadron (The average refers to the hadron wavefunction and the upper index i with $i = 1, 2$ indicates the transverse directions.) This quantity offers more information than the standard parton distributions like $xg(x, Q^2)$ — it also describes the distribution of gluons in transverse momentum, and not only in x —, but it still does not probe *many-body correlations* in the gluon distribution, as the higher n -point functions with $n \geq 4$ would do. The restriction to the 2-point function is justified so long as the system is dilute and gluons do not interact with each other. But this cannot encode the non-linear physics of saturation, which is sensitive to higher n -point functions and hence to correlations. In fact, to correctly describe gluon saturation, one needs to control n -point functions with *arbitrarily high* n . This can be understood as follows: the fact that the occupation numbers are $n \sim \mathcal{O}(1/\alpha_s)$ at saturation, means that the colour field strengths are as large as $A_a^i \sim \mathcal{O}(1/g)$, and then there is no penalty for inserting arbitrary powers of A_a^i . Indeed, any such an insertion is accompanied by a factor of g . (Recall that interactions in QCD enter via the covariant derivative $D^\mu = \partial^\mu - igA^\mu$.) So, the CGC effective theory is truly *an infinite hierarchy of coupled evolution equations* describing the simultaneous evolution of all the n -point functions. Remarkably enough, this hierarchy can be summarized into a single, *functional*, evolution equation for the *CGC weight function* — a functional generalization of the ‘unintegrated’ gluon distribution that will be shortly discussed.

The key ingredient for such an economical description is the proper choice of the relevant degrees of freedom: as already mentioned, the small- x gluons with high occupation numbers $n \sim 1/\alpha_s$ can be

³See Eq. (35) for a more precise definition of the unintegrated gluon distribution in the presence of non-linear effects.

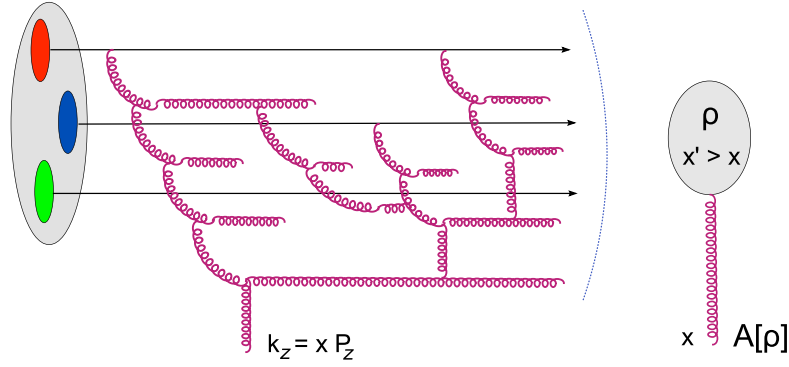


Fig. 14: Schematic representation of the d.o.f. involved in the CGC effective theory and of the quantum evolution which is taken into account in this theory. A newly emitted gluon with a small longitudinal momentum fraction $x \ll 1$ rescatter off the gluon field $A_a^\mu[\rho]$ created in the previous steps by the gluons with larger values $x' \gg x$, effectively represented by their global colour charge density ρ_a .

treated *semi-classically* to leading order in α_s — that is, they can be described as *classical colour fields* $A_a^\mu(x)$ radiated by *colour sources* representing the faster gluons with $x' \gg x$. This distinction between ‘classical fields’ (= the small- x gluons for which the effective theory is built) and their ‘sources’ (= the large- x gluons which are integrated out in the construction of the effective theory) is illustrated in Fig. 14. The effective theory based on this separation is valid to LO in α_s , but to all orders in $\alpha_s \ln(1/x)$ and in the classical field $A_a^\mu \sim \mathcal{O}(1/g)$.

The mathematical structure of the CGC theory is rather complex and it will be only schematically described here. To that aim, it is convenient to switch to *light-cone vector notations*. Namely, for any 4-vector such as x^μ , p^μ , A_a^μ etc. we shall define its light-cone (LC) components as

$$x^+ \equiv \frac{1}{\sqrt{2}} (x^0 + x^3), \quad x^- \equiv \frac{1}{\sqrt{2}} (x^0 - x^3), \quad x_{\text{LC}}^\mu = (x^+, x^-, \mathbf{x}_\perp). \quad (26)$$

In LC notations, the scalar product reads $k \cdot x \equiv k_\mu x^\mu = k^+ x^- + k^- x^+ - \mathbf{k}_\perp \cdot \mathbf{x}_\perp$.

To see the usefulness of these notations, consider a right-moving ultrarelativistic hadron, with $P^\mu \simeq (P, 0, 0, P)$: this propagates at nearly the speed of light along the trajectory $x^3 = t$. In LC notations, the 4-momentum $P_{\text{LC}}^\mu \simeq (\sqrt{2}P, 0, 0, 0)$ has only a ‘plus’ component, while the trajectory reads simply $x^- = 0$. The same holds for any of the large- x partons which move quasi-collinearly with the hadron and serve as sources for the small- x gluons that we are interested in. In the semi-classical approximation, these small- x gluons are described as the solution to the Yang–Mills equations (the non-Abelian generalization of the Maxwell equations) having these ‘fast’ gluons as sources:

$$D_\nu^{ab} F_b^{\nu\mu}(x) = \delta^{\mu+} \rho^a(x^-, \mathbf{x}_\perp). \quad (27)$$

In this equation, the l.h.s. features the covariant derivative $D_\nu^{ab} = \partial_\nu - g f^{abc} A_c^\nu$ and the field strength tensor $F_a^{\nu\mu} = \partial^\nu A_a^\mu - \partial^\mu A_a^\nu - g f^{abc} A_b^\nu A_c^\mu$ associated with the classical colour field, while the r.h.s. is the *colour current* of the ‘fast’ gluons: $J_a^\mu = \delta^{\mu+} \rho_a$, with $\rho^a(x^-, \mathbf{x}_\perp)$ their colour charge density. The latter is localized in x^- near $x^- = 0$ and is independent of time (hence of x^+), because these fast charges are ‘frozen’ by Lorentz time dilation. But the distribution of these charges in transverse space is *random*, since the fast gluons can be in any of the quantum configurations produced at the intermediate stages of the gluon evolution down to x . The proper way to describe this randomness is to give the *probability* to find a specific configuration $\rho_a(x^-, \mathbf{x}_\perp)$ of the colour charge density. This probability is a functional of $\rho_a(x^-, \mathbf{x}_\perp)$, known as the *CGC weight function* and denoted as $W_Y[\rho]$, with $Y = \ln(1/x)$. This functional is *gauge-invariant*, which in particular ensures that $\langle \rho_a(x^-, \mathbf{x}_\perp) \rangle = 0$, as it should.

To the accuracy of interest, all the observables relevant for the scattering off the small- x gluons are represented by gauge-invariant operators built with the classical field A_a^μ . If $O[A]$ is such an operator, then its hadron expectation value is computed by averaging over all the configurations of ρ with the CGC weight function:

$$\langle O[A] \rangle_Y \equiv \int \mathcal{D}\rho W_Y[\rho] O[A[\rho]], \quad (28)$$

where $A_a^\mu[\rho]$ is the solution to Eq. (27).

The expectation value (28) depends upon the rapidity $Y = \ln(1/x)$ via the corresponding dependence of the weight function $W_Y[\rho]$. The latter is obtained by successively integrating the quantum gluon fluctuations in layers of x , down to the value of interest. One step in this evolution corresponds to the emission of a new gluon (with a probability $\mathcal{O}(\alpha_s)$ per unit rapidity) out of the preexisting ones. But unlike in the BFKL evolution, where gluons with different rapidities do not ‘see’ each other, in the context of the CGC evolution, the newly emitted gluon is allowed to interact with the strong colour field radiated by ‘sources’ (gluons and valence quarks) with higher values of x (see Fig. 14). Accordingly, the change in the CGC weight function in one evolution step is *non-linear* in the background field $A_a^\mu[\rho]$, and hence in the colour charge density ρ_a . This procedure generates a *functional evolution equation* for $W_Y[\rho]$ with the schematic form (see Refs. [8, 10] for details)

$$\frac{\partial W_Y[\rho]}{\partial Y} = H_{\text{JIMWLK}} \left[\rho, \frac{\delta}{\delta \rho} \right] W_Y[\rho], \quad (29)$$

where the JIMWLK Hamiltonian (from Jalilian-Marian, Iancu, McLerran, Weigert, Leonidov, and Kovner [18, 19]) H_{JIMWLK} is non-linear in ρ to all orders (thus encoding the rescattering effects in the emission vertex) but quadratic in the functional derivatives $\delta/\delta\rho$ (corresponding to the fact that there is only one new gluon emitted in each step in the evolution). In the dilute regime, where parametrically $g\rho \ll 1$, the non-linear effects are negligible, the JIMWLK Hamiltonian can be expanded to quadratic order in ρ , and then it describes the BFKL evolution. But for $g\rho \sim 1$, the non-linear effects encoded in H_{JIMWLK} prevent the emission of new gluons; this is *gluon saturation*.

Equations (27)–(29) are the central equations of the CGC effective theory. When completed with an initial condition at the rapidity Y_0 at which one starts the high-energy evolution, they fully specify the gluon distribution in the hadron wavefunction, including all its correlations. The initial condition $W_{Y_0}[\rho]$ is not determined by the effective theory itself, rather one must resort on some model. For a large nucleus ($A \gg 1$) and for $Y_0 = 4 \div 5$ (corresponding to $x_0 \sim 0.01$), a reasonable initial condition is provided by the McLerran–Venugopalan (MV) model [7], which assumes that the ‘fast’ colour sources are the $N_c \times A$ valence quarks, which radiate independently from each other (since they are typically confined within different nucleons). The corresponding weight function is a Gaussian in ρ_a .

By taking a derivative w.r.t. Y in Eq. (28) and using Eq. (29) for W_Y , one can deduce evolution equations for all the observables of interest. In general, these equations do not form a closed set; rather, they form an infinite hierarchy (originally derived by Balitsky [20]) which couples n -point functions with arbitrarily large values of n . In practice, this hierarchy can be truncated via mean field approximations [8, 21, 22], leading to closed but non-linear equations, in particular the Balitsky–Kovchegov equation [23], that can be explicitly solved. It is also possible to solve the functional JIMWLK equation (29) via numerical simulations on a lattice [24–26], although in practice this is quite cumbersome.

In order to describe a scattering cross-section, the CGC effective theory developed so far must be combined with a factorization scheme. This will be described in the next subsection.

3.5 Particle production from the CGC

Let us start with some general remarks on factorization in scattering at high energies: this is a generic consequence of *causality*. For a hadron-hadron collision in the COM frame, the collision time $\Delta t_{\text{coll}} \sim$

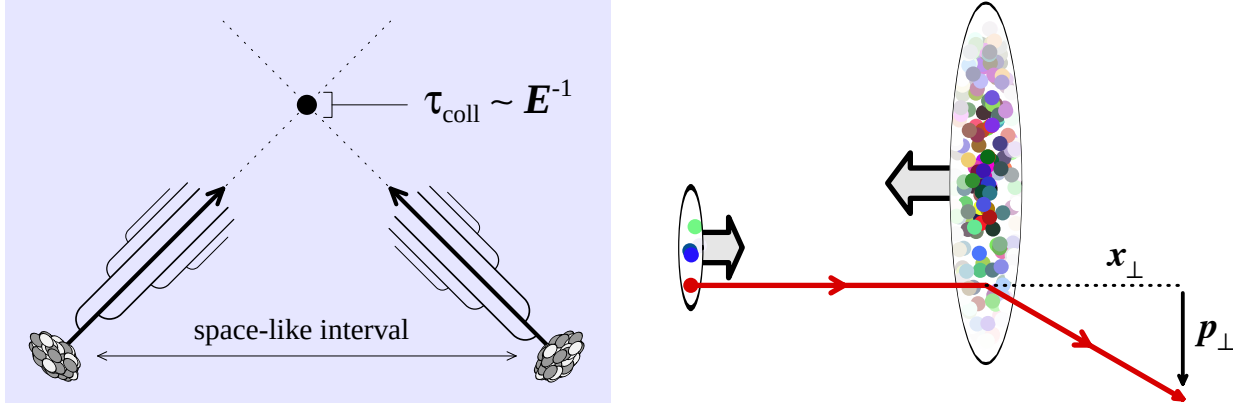


Fig. 15: Left: The space-time picture of a high-energy collisions, illustrating the factorization of the cross-section. Right: particle production in a proton-nucleus (or ‘dilute-dense’) collision.

$1/\sqrt{s}$ is much shorter than the lifetime (5) of the partons participating in the collisions, which is proportional to the parton longitudinal momentum $xP \sim \sqrt{s}$. Hence, these partons have been produced long time before the collision, at a time where the two incoming hadrons were causally disconnected from each other (see Fig. 15 left). Accordingly, the respective parton distributions have evolved *independently* from each other and thus they are *universal* — i.e. independent of the scattering process that is used to probe them. This argument is purely kinematic and hence it remains true in the presence of QCD interactions leading to *parton evolution* or *gluon saturation*. However, the precise form of the factorization formula depends upon the kinematics and the structure of the process at hand and it is different when probing dense or dilute parts of the hadron wavefunction.

(i) In the *dilute regime*, which corresponds to the situation where the transverse momenta $p_{\perp}$ of the produced partons are significantly larger than the saturation momenta in the two hadrons as evaluated at the relevant values of x , cf. Eq. (13), the partonic subprocess involves merely a *binary collision* (cf. Fig. 6 left) : one parton in one projectile interacts with one parton in the other projectile, to produce the final state. Then, the cross-section depends only upon the *parton densities* (the 2-point correlations of the quark and gluon fields) in the incoming hadrons and the factorization formula takes a rather simple form: the hadronic cross-section is the convolution of two parton distribution functions (one for each hadron) times the cross-section for the partonic subprocess.

Even in this case, one needs to distinguish between two types of factorizations, depending upon the kinematics of the final state:

- If the relevant values of x are not that small (say $x \gtrsim 0.01$), then the parton evolution with decreasing x can be neglected and one can use the *collinear factorization* : the partons are assumed to move collinearly with the incoming hadrons (that is, one neglects their ‘intrinsic’ transverse momenta $\sim \Lambda_{\text{QCD}}$) and the parton distributions like $xg(x, \mu^2)$ depend only upon the longitudinal momentum fractions and upon the transverse resolution μ^2 (the ‘factorization scale’) of the hard, partonic, subprocess. The dependence upon μ^2 reflects the DGLAP evolution with increasing virtuality Q^2 . Eq. (15) provides an example of collinear factorization.
- At smaller values of x , such that $\alpha_s \ln(1/x) \gtrsim 1$, the small- x evolution becomes important, leading to an increase in the density of gluons and in their average transverse momentum. Because of that, the gluons cannot be considered as ‘collinear’ anymore : their distribution in transverse momenta must be explicitly taken into account. That is, one has to use the ‘*unintegrated*’ *gluon distribution*, whose evolution with $1/x$ is described by the BFKL equation, cf. Eq. (22). The corresponding factorization formula, known as *k_T -factorization*, involves a convolution over the transverse momenta of the participating gluons. This is in fact a limiting case of the CGC factor-

ization to be described shortly — namely its dilute limit, in which the saturation effects can be neglected.

(ii) The *dense regime* corresponds to collisions which probe saturation effects in at least one of the incoming hadrons. This happens when the transverse momenta of some of the produced particles are comparable with the saturation momentum at the relevant values of x . In such a case, the partons from one projectile scatter off a dense gluonic system (a CGC) in the other projectile, so they typically undergo *multiple scattering*. This is a non-linear effect similar to saturation: each additional scattering represents a correction of order $\alpha_s n$ to the cross-section, which for $n \sim 1/\alpha_s$ is an effect of order one. (As usual, $n = n(x, k_\perp)$ denotes the gluon occupation number in the dense projectile and is of $\mathcal{O}(1/\alpha_s)$ when $k_\perp \lesssim Q_s(x)$.) So, when a parton scatters off a CGC, the multiple scattering series must be resummed to all orders. This resummation involves arbitrarily many insertions of the strong colour field A_a^μ which represents the CGC, which implies that the associated cross-section is sensitive to multi-gluon correlations (n -point functions of the field A_a^μ with $n \geq 2$). Clearly, the multiple scattering cannot be encoded in the (collinear or k_T) factorization schemes alluded to above, which involve only the respective 2-point functions — the parton distributions.

Fortunately, there is an important simplification which occurs at high energy and which permits to compute multiple scattering to all orders: an energetic parton is not significantly deflected by the scattering and thus can be assumed to preserve a straight line trajectory throughout the collision. This is known as the *eikonal approximation*. To explain it in a simple, but phenomenologically relevant, setting, consider first *proton-nucleus ($p+A$) collisions*, cf. Fig. 15 right. This is an example of a *dense-dilute scattering*, that is, a collision in which a projectile which is relatively dilute (the ‘proton’) and can therefore be treated in the collinear or the k_T factorization, scatters off a dense target (the ‘nucleus’). Then the partons from the dilute projectile undergo multiple scattering off the strong colour field of the dense target. For the kinematical conditions at RHIC or the LHC, this ‘dense-dilute’ scenario is optimally realized in the case of particle production at *forward rapidities*, that is, in the fragmentation region of the proton (deuteron at RHIC).

To be specific, consider the production of a light quark with rapidity $y > 0$ and semi-hard transverse momentum $p_\perp$. The production mechanism is as follows: a quark from the proton, with relatively large longitudinal momentum fraction $x_1 = (p_\perp/\sqrt{s})e^y$ and negligible transverse momentum, scatters off the gluons with $x_2 \simeq (p_\perp/\sqrt{s})e^{-y}$ and $k_{2\perp} \sim Q_s(x_2)$ from the nucleus — which form a dense system because $x_2 \ll 1$ and $A \gg 1$ — and thus accumulates a final transverse momentum $p_\perp \sim Q_s(x_2)$. Within the CGC effective theory, the nucleus in a given scattering event is described as a classical colour field A_a^μ , off which the quark scatters with the S -matrix

$$S_{\alpha\beta} = \langle \beta | \text{T exp} \left\{ i \int d^4y \mathcal{L}_{\text{int}}(y) \right\} | \alpha \rangle \quad \text{with} \quad \mathcal{L}_{\text{int}}(y) = j_a^\mu(y) A_\mu^a(y) \quad \text{and} \quad j_a^\mu(y) = g \bar{\psi} \gamma^\mu t^a \psi. \quad (30)$$

Here $\mathcal{L}_{\text{int}}(y)$ is the Lagrangian density for the interaction between the colour current j_a^μ of the quark and the colour field A_μ^a of the target, and α (β) represents the ensemble of the quantum numbers characterizing the state of the quark prior to (after) the collision. The quark deflection angle reads $\theta \simeq p_\perp/E_1$ with $E_1 = x_1\sqrt{s}/2$ (the quark energy). For the kinematics of interest we have $p_\perp \ll E_1$, so this angle is small, $\theta \ll 1$, as anticipated. Hence, one can assume that the quark keeps a *fixed transverse coordinate* $\mathbf{x}_\perp$ while crossing the nucleus. This is the eikonal approximation. In this approximation, $\alpha = (\mathbf{x}_\perp, i)$ and $\beta = (\mathbf{x}_\perp, j)$, where i and j are colour indices in the fundamental representation of $\text{SU}(N_c)$ which indicate the quark colour states before and respectively after the scattering. Also, assuming the quark to be a left-mover (and hence the nucleus to be a right mover, as in Eq. (27)), one can write $j_a^\mu(y) \simeq \delta^{\mu-} g t^a \delta(y^+) \delta^{(2)}(\mathbf{y}_\perp - \mathbf{x}_\perp)$. Then Eq. (30) reduces to

$$\langle \mathbf{x}_\perp, j | S | \mathbf{x}_\perp, i \rangle \simeq V_{ij}(\mathbf{x}_\perp) \quad \text{with} \quad V(\mathbf{x}_\perp) \equiv \text{T exp} \left\{ ig \int dx^- A_a^+(x^-, \mathbf{x}_\perp) t^a \right\}, \quad (31)$$

where the ‘path-ordering’ symbol T denotes the ordering of the colour matrices $A_a^+(x^-, \mathbf{x}_\perp) t^a$ in the exponent, from right to left, in increasing order of their x^- arguments. The integration runs formally over all the values of x^- , but in reality this is restricted to the longitudinal extent of the nucleus, which is localized near $x^- = 0$ because of Lorentz contraction⁴. The path-ordered exponential V is a colour matrix in the fundamental representation, also known as a *Wilson line*. It shows that the only effect of the scattering in the high energy limit is to ‘rotate’ the colour state of the quark while the latter is crossing the nucleus. If instead of a quark, one would consider the scattering of a gluon, the corresponding S -matrix would be again a Wilson line, but in the *adjoint* representation ($t^a \rightarrow T^a$). When the target field is weak, $gA^+ \ll 1$, one can expand the exponential in Eq. (31) in powers of gA^+ , thus generating the multiple scattering series. But when $gA^+ \sim 1$, as is the case for a target where the gluons are at saturation, such an expansion becomes useless, since all the terms count on the same order. In such a case, one has to work with the all-order result, as compactly encoded in the Wilson line.

The above considerations also show that multiple scattering at high energy is most conveniently treated in the transverse *coordinate* representation: the successive collisions modify the transverse momentum of the partonic projectile, but do not significantly alter its transverse coordinate (or ‘impact parameter’). But the interesting observable is the cross-section for producing a quark (or gluon) with a given transverse momentum $p_\perp$ and rapidity y . This is obtained by multiplying the amplitude $V_{ij}(\mathbf{x}_\perp)$ with the complex conjugate amplitude $V_{ji}^\dagger(\mathbf{y}_\perp)$ for a quark at a *different* impact parameter $\mathbf{y}_\perp$, and then taking the Fourier transform $\mathbf{x}_\perp - \mathbf{y}_\perp \rightarrow \mathbf{p}_\perp$. This yields (for the forward kinematics of interest here)

$$\frac{dN_q}{dy d^2\mathbf{p}_\perp} \simeq x_1 f_q(x_1, p_\perp^2) \int d^2\mathbf{r}_\perp e^{-i\mathbf{r}_\perp \cdot \mathbf{p}_\perp} \frac{1}{N_c} \left\langle \text{tr} V(\mathbf{x}_\perp) V^\dagger(\mathbf{y}_\perp) \right\rangle_Y. \quad (32)$$

The quark distribution $x_1 f_q(x_1, p_\perp^2)$ gives the probability to find a quark ‘collinear’ with the proton, with longitudinal momentum fraction x_1 , on the resolution scale $p_\perp^2$ set by the partonic scattering. Within the integral, we have defined $\mathbf{r}_\perp \equiv \mathbf{x}_\perp - \mathbf{y}_\perp$. Furthermore, the colour trace has been generated by summing over the final colour indices ($\sum_j$) and averaging over the initial ones ($(1/N_c) \sum_i$), and the brackets denote the average over the target wavefunction, evolved up to the rapidity $Y = \ln(1/x_2)$. In the CGC formalism, this target average is computed according to Eq. (28), that is, as an average over the colour charge density of the ‘fast’ sources which are responsible for the target field $A_a^+(x^-, \mathbf{x}_\perp)$ via Eq. (27):

$$S_Y(\mathbf{x}_\perp, \mathbf{y}_\perp) \equiv \frac{1}{N_c} \left\langle \text{tr} V(\mathbf{x}_\perp) V^\dagger(\mathbf{y}_\perp) \right\rangle_Y = \int \mathcal{D}\rho W_Y[\rho] \frac{1}{N_c} \text{tr}(V(\mathbf{x}_\perp) V^\dagger(\mathbf{y}_\perp)). \quad (33)$$

This 2-point function of the Wilson lines is recognized as the S -matrix for the scattering between a *colour dipole* (a quark-antiquark pair in an overall colour singlet state) and the CGC. The non-linear effects enter this S -matrix at two levels: (i) via *multiple scattering* for the quark and the antiquark, as described by the respective Wilson lines, and (ii) via *gluon saturation* in the target wavefunction, as encoded in the CGC weight function. In this context, the target saturation momentum $Q_s(A, Y)$ also plays the role of the *unitarization scale* for the colour dipole: the dipole scattering becomes strong (meaning that $|S_Y(r_\perp)| \ll 1$) when the dipole size $r_\perp$ is of order $1/Q_s$ or larger. Indeed, so long as the dipole is relatively small, such that $r_\perp \ll 1/Q_s$, it predominantly scatters off the gluon modes with $k_\perp \sim 1/r_\perp \gg Q_s$, which are dilute. The corresponding target field is weak ($gA^+ \ll 1$), the Wilson lines are close to one, and so is the dipole S -matrix. Namely, for $r_\perp Q_s \ll 1$ one finds $1 - S_Y(r_\perp) \propto (r_\perp^2 Q_s^2(Y))^{\gamma_s}$ with $\gamma_s \simeq 0.63$. On the other hand, a large dipole with $r_\perp \gtrsim 1/Q_s(Y)$ probes the high-density gluon modes with $k_\perp \lesssim Q_s(Y)$; the associated colour fields are strong, $gA^+ \sim 1$, so the Wilson lines are rapidly oscillating and their product averages out to a very small value: $|S_Y(r_\perp)| \ll 1$ when $r_\perp Q_s \gg 1$.

The dipole S -matrix provides a convenient framework to study high energy evolution and saturation since, in the limit where the number of colours is large $N_c \gg 1$, it obeys a relatively simple

⁴More precisely, the small- x gluons which participate in the scattering are delocalized within a distance $\Delta x^- \sim 1/(x_2 P)$ around $x^- = 0$, as explained after Eq. (18).

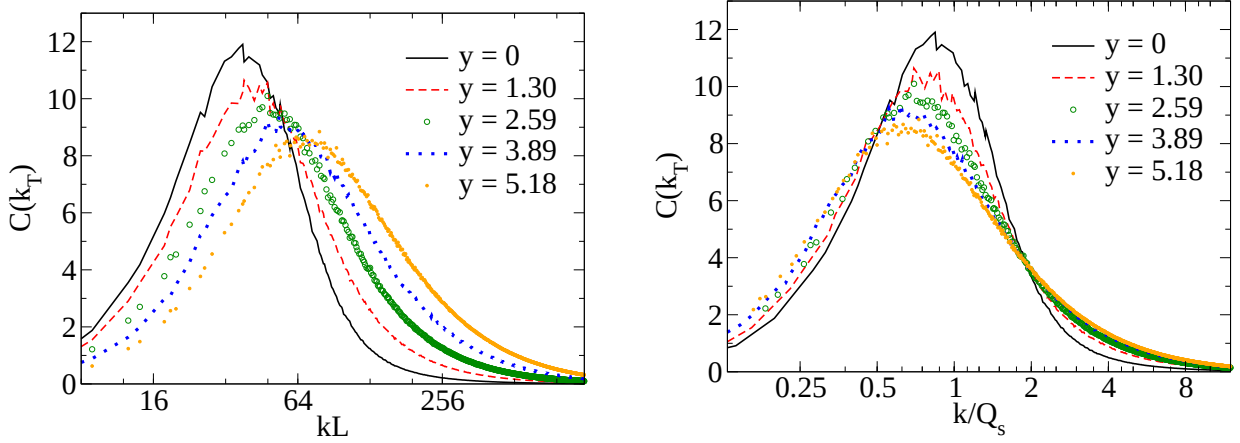


Fig. 16: The results of the JIMWLK evolution for the generalized gluon distribution in Eq. (35), parameterized as $\Phi(Y, \mathbf{k}_\perp) \equiv (1/\alpha_s)C(Y, \mathbf{k}_\perp)$. Left: as a function of $k_\perp L$, with L the transverse size of the system. Right: as a function of the scaling variable $k_\perp/Q_s(Y)$, to exhibit geometric scaling (the curves corresponding to different values of Y fall approximately on top of each other). From Ref. [25].

equation for the evolution with Y — a non-linear generalization of the BFKL equation known as the Balitsky–Kovchegov (BK) equation [23]. Originally deduced within Mueller’s ‘dipole picture’ [9, 27] (an insightful reformulation of the BFKL evolution valid at large N_c), the BK equation also emerges from the large- N_c limit of the Balitsky–JIMWLK hierarchy. Remarkably, the BK equation is presently known to next-to-leading order (NLO) accuracy [28–30]. This is important in view of phenomenological studies of both deep inelastic scattering — at high energy, the DIS cross-section (9) can be related to the dipole S -matrix [31] and the NLO corrections are essential in order to achieve a good description of the HERA data at small x [32–36] — and of particle production in dense-dilute scattering, which is the topics of interest for us here.

Specifically, Eqs. (32)–(33) express the *CGC factorization* for inclusive quark production in dense-dilute scattering [11, 37–39]. Eq. (32) is usually written as

$$\frac{dN_q}{dy d^2\mathbf{p}_\perp} \simeq \frac{\alpha_s}{p_\perp^2} x_1 f_q(x_1, p_\perp^2) \Phi_A(Y, \mathbf{p}_\perp, \mathbf{b}_\perp), \quad (34)$$

where the first factor $\alpha_s/p_\perp^2$ is the cross-section for the elementary $q + g \rightarrow q$ scattering (the would-be partonic subprocess in the single scattering limit) while

$$\Phi_A(Y, \mathbf{p}_\perp, \mathbf{b}_\perp) \equiv \frac{p_\perp^2}{\alpha_s} \int d^2\mathbf{r}_\perp e^{-i\mathbf{r}_\perp \cdot \mathbf{p}_\perp} \frac{1}{N_c} \left\langle \text{tr} V(\mathbf{x}_\perp) V^\dagger(\mathbf{y}_\perp) \right\rangle_Y, \quad (35)$$

plays the role of a *generalized unintegrated gluon distribution* (here, for the nucleus) at impact parameter $\mathbf{b}_\perp = (\mathbf{x}_\perp + \mathbf{y}_\perp)/2$. When $p_\perp \gg Q_s(A, Y)$, Eq. (35) can be evaluated in the single scattering approximation, as obtained by expanding the product of Wilson lines to quadratic order in A^+ . In that (dilute) regime, Φ_A reduces indeed to the usual unintegrated gluon distribution — the one which enters the k_T -factorization and obeys the BFKL equation. But for lower momenta $p_\perp \lesssim Q_s(A, Y)$, the non-linear effects become essential, as already discussed in relation with the dipole scattering.

Figure 16 shows the CGC prediction for the generalized gluon distribution (35), as obtained via the numerical resolution of the JIMWLK equation with initial conditions at $Y = 0$ of the McLerran–Venugopalan type (and with a running coupling) [25]. The two plots exhibit a pronounced peak at a special value of the transverse momentum $p_\perp$ which increases with Y : this special value is, of course, the saturation momentum $Q_s(A, Y)$. In fact, it is quite easy to understand these plots in the light of our previous discussion of the dipole scattering. The Fourier transform in Eq. (35) is controlled by the

competition between the complex phase $e^{-ir_\perp \cdot p_\perp}$ and the dipole S -matrix $S_Y(r_\perp)$. For large momenta $p_\perp \gg Q_s$, the complex exponential limits the integration to small dipole sizes $r_\perp \lesssim 1/p_\perp \ll 1/Q_s$, for which the scattering is weak: $1 - S_Y(r_\perp) \sim (r_\perp^2 Q_s^2(Y))^{\gamma_s}$. Then the Fourier transform yields

$$\Phi_A(Y, p_\perp) \simeq \frac{1}{\alpha_s} \left(\frac{Q_s^2(A, Y)}{p_\perp^2} \right)^{\gamma_s} \quad \text{for } p_\perp \gg Q_s(A, Y). \quad (36)$$

The difference $1 - \gamma_s \simeq 0.37$ is an *anomalous dimension* introduced by the high energy evolution. (Without this evolution, one would have the bremsstrahlung spectrum $\propto 1/p_\perp^2$, cf. Eq. (22).) For lower momenta $p_\perp < Q_s$, the integral over $r_\perp$ in Eq. (35) is limited by the dipole S -matrix to values $r_\perp \lesssim 1/Q_s < 1/p_\perp$. (Recall that $S_Y(r_\perp) \ll 1$ when $r_\perp \gg 1/Q_s$.) Then

$$\Phi_A(Y, p_\perp) \simeq \frac{p_\perp^2}{\alpha_s} \int d^2 r_\perp \Theta(1/Q_s - r_\perp) \simeq \frac{1}{\alpha_s} \frac{p_\perp^2}{Q_s^2(A, Y)} \quad \text{for } p_\perp \lesssim Q_s(A, Y). \quad (37)$$

Equations (36) and (37) explain the pronounced peak in the gluon distribution at $p_\perp \simeq Q_s(Y)$, as visible in Fig. 16. This in turn implies that the cross-section for particle production is dominated by semi-hard gluons with $p_\perp \sim Q_s(Y)$ and hence can be computed in perturbation theory. Another important consequence of saturation, which is visible too in Eqs. (36) and (37) (and is numerically tested in the right panel of Fig. 16), is *geometric scaling* [40–43]: the unintegrated gluon distribution depends upon the two kinematical variables $p_\perp$ and Y only via the dimensionless ratio $p_\perp/Q_s(Y)$. This scaling has important consequences for the phenomenology, to be discussed in the next section.

Note also that, as a result of the non-linear physics, the generalized ‘gluon distribution’ (35) is a *process-dependent quantity*: it depends not only upon the gluon density in the target, but also upon the nature of the partonic subcollision. For instance, if instead of the quark production, one would consider the production of a gluon, the ‘fundamental’ Wilson lines in Eq. (35) would be replaced by ‘adjoint’ ones [37–39]. Also, if one considers a more complicated final state — say, the production of a pair of partons — then the analog of Eq. (35) will involve a pair of Wilson lines for each of the partons partaking in the collision (one such a line in the direct amplitude and another one in the complex conjugate amplitude). More examples in that sense can be found in Refs. [44–49]. See also Ref. [50] for a recent, comprehensive, discussion of the relation between the CGC factorization and the k_T -factorization. These considerations show that, in the presence of non-linear effects, the notion of ‘parton distribution’ ceases to be useful: the observables involve higher n -point functions of the gluon fields (generally, via the product of Wilson lines) which moreover couple with each other under the non-linear, JIMWLK, evolution. The complete information about the gluon correlations and their evolution with Y (to leading-logarithmic accuracy at least) is encoded in the CGC weight function.

We now turn to *nucleus-nucleus* ($A + A$) collisions, which for the typical kinematical conditions at RHIC and the LHC represents an example of *dense-dense scattering*: the wavefunctions of both nuclei develop saturation effects which influence the production of particles with semi-hard transverse momenta. Hence, in order to compute the bulk of particle production, one must take into account the many-body correlations associated with gluon saturation in both nuclei, together with the multiple scattering between these two saturated gluon distributions. Once again, this complex problem can be addressed within the CGC formalism and in the eikonal approximation. Not surprisingly, the treatment of the two nuclear projectiles is now *symmetric*: they are both described as colour glass condensates, with weight functions $W_{Y_1}[\rho_1]$ and, respectively, $W_{Y_2}[\rho_2]$. And their collision is described as the scattering between two classical distributions of colour charges moving against each other.

To be more specific, consider inclusive gluon production in the COM frame, where the nuclei have rapidities $\pm y_{\text{beam}}$. If the produced gluon has rapidity y , it will probe the evolutions of the two nuclear wavefunctions up to rapidities $Y_1 = y_{\text{beam}} - y$ and $Y_2 = y_{\text{beam}} + y$, respectively. This motivates the following CGC factorization for the spectrum of the produced gluons [51]

$$\left\langle \frac{dN}{dy d^2 p_\perp} \right\rangle = \int [D\rho_1 D\rho_2] W_{y_{\text{beam}}-y}[\rho_1] W_{y_{\text{beam}}+y}[\rho_2] \frac{dN}{dy d^2 p_\perp} \Big|_{\text{class}}, \quad (38)$$

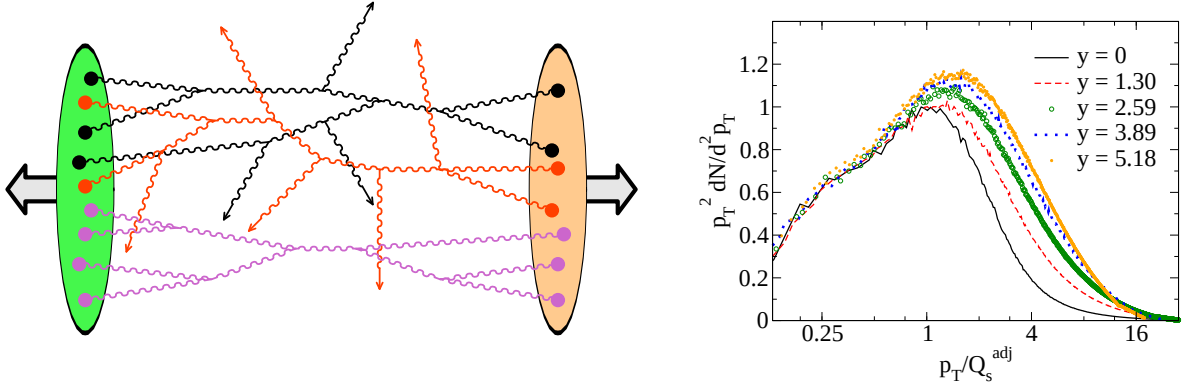


Fig. 17: Left panel: a cartoon of the ‘dense-dense’ collision between two heavy nuclei, which illustrates the complexity of the process. Non-linear effects enter via gluon saturation (in both nuclei) and multiple scattering. Right panel: the gluon spectrum generated by this collision within the CGC formalism (from Ref. [25]).

where the last factor inside the integrand, $(dN/dy d^2p_\perp)|_{\text{class}}$, represents the spectrum produced in the scattering between two given configurations of classical colour sources (the ‘fast partons’) — one for each nucleus. More precisely, as discussed in relation with Eq. (27), the right-moving nucleus is described in a given event as a colour current having only a ‘plus’ component: $J_1^{\mu,a} = \delta^{\mu+} \rho_1^a$, with the charge density ρ_1^a localized near $x^- = 0$ (due to Lorentz contraction) and independent of x^+ (by Lorentz time dilation). Similarly, the left-moving nucleus is represented by a colour current with only a ‘minus’ component: $J_2^{\mu,a} = \delta^{\mu-} \rho_2^a$, with ρ_2^a localized near $x^+ = 0$ and independent of x^- . At a classical level, the ‘scattering’ between these two currents is described by the solution A_a^μ to the Yang–Mills equation including both types of sources (compare to Eq. (27)) :

$$D_\nu^{ab} F_b^{\nu\mu}(x) = \delta^{\mu+} \rho_1^a(x^-, \mathbf{x}_\perp) + \delta^{\mu-} \rho_2^a(x^+, \mathbf{x}_\perp). \quad (39)$$

This equation describes multiple scattering because it is non-linear: the collision begins at $x^+ = x^- = 0$ (i.e. $t = z = 0$) and for positive values of x^+ and x^- , the solution A_a^μ is non-linear to all orders in both ρ_1 and ρ_2 . This solution cannot be computed analytically, but numerical solutions are by now available [25, 52–54]. The cross-section $(dN/dy d^2p_\perp)|_{\text{class}}$ for particle production is obtained via the Fourier transform of this classical solution, that is, by projecting the field A_a^μ onto modes with transverse momentum $p_\perp$. Finally, the average over the CGC weight functions of the two nuclei, cf. Eq. (38), is numerically performed. It is this last procedure which introduces the dependence of the cross-section upon the rapidity y , via the corresponding dependence of the two weight functions. Note that, in line with the general philosophy of the CGC formalism, the only quantum effects to be included in the calculation are those associated with the high-energy evolution of the projectile wavefunctions prior to scattering, which are enhanced by the large logarithms $Y_i = \ln(1/x_i)$, with $i = 1, 2$. The final outcome of this calculation is the gluon spectrum displayed in Fig. 17 (right panel) [25]. This is very similar to the ‘unintegrated’ gluon distribution in any of the incoming nuclei (compare to Fig. 16), in particular, it is peaked at a value of $p_\perp$ of the order of the saturation momentum and which increases with y . Some further consequences of the solution to Eq. (39) will be discussed in Section 3.7.

Note finally an important difference between p+A (dilute-dense) and A+A (dense-dense) collisions: in the former, the particles produced by the collision do not interact with each other, but merely evolve via fragmentation and hadronisation towards the final hadrons observed in the detectors; by contrast, in A+A collisions the partonic medium created in the early stages of the collisions is very dense, so these partons keep interacting with each other — one then speaks about *final state* interactions, as opposed to the *initial state* ones, which were associated with high density effects like saturation in the

incoming wavefunctions. These ‘final state’ interactions redistribute the partons in energy and momentum, which makes it difficult to verify the early-stages spectrum, as given by the aforementioned CGC calculations, against the measured hadron yield. Yet, the CGC formalism has the virtue to provide the *initial conditions* for the subsequent dynamics, at a time $\tau \simeq 1/Q_s \sim 0.2$ fm/c. So, its predictions can be at least indirectly tested, via calculations of the final state effects which include initial conditions of the CGC type. We shall return to such issues later on.

3.6 Some experimental signatures of the CGC in HIC

Let us consider now some phenomenological applications of the previous results for ‘dilute-dense’ scattering [10, 11, 14, 15] (and Refs. therein). We start with the R_{pA} ratio, defined as the ratio between particle production in p+A collisions and that in p+p collisions for the same kinematics; schematically,

$$R_{pA}(\eta, p_\perp) \equiv \frac{1}{N_{coll}} \frac{\left. \frac{dN_h}{d^2p_\perp d\eta} \right|_{pA}}{\left. \frac{dN_h}{d^2p_\perp d\eta} \right|_{pp}}, \quad (40)$$

where the subscript h denotes the hadron species and N_{coll} is the number of binary proton-nucleon collisions in the p+A scattering at a given impact parameter, as computed under the assumption that the various nucleons inside the nucleus scatter independently from each other. For relatively central collisions (i.e. small impact parameters; see Fig. 23), one has $N_{coll} \simeq A^{1/3}$. The normalization in Eq. (40) is such that R_{pA} would be equal to one if the p+A collision was a superposition of A incoherent p+p collisions. Conversely, any deviation in R_{pA} from unity is an indication of coherence (high-density) effects in the nuclear wavefunction. Such a deviation is clearly seen in the respective RHIC data at *forward rapidities* ($\eta > 0$). More precisely, RHIC performed deuteron-gold (d+Au) collisions⁵ at $\sqrt{s} = 200$ GeV per nucleon pair and measured the ratio R_{d+Au} for semi-hard momenta $p_\perp = 1 \div 5$ GeV and for rapidities $\eta = 0 \div 4$ in the deuteron fragmentation region [55, 56].

The results of the corresponding analysis by BRAHMS are shown in Fig. 18. For $p_\perp \gtrsim 2$ GeV, they show an *enhancement* ($R_{d+Au} > 1$) for $\eta = 0$, known as the ‘Cronin peak’, which however disappears when increasing η , leading to *suppression* ($R_{d+Au} < 1$) at $\eta > 1$. E.g. for $\eta = 3.2$, one finds $R_{d+Au} \simeq 0.6 \div 0.8$ for $p_\perp = 2 \div 4$ GeV. This behaviour can be understood in terms of saturation in the nuclear gluon distribution and its evolution with Y [57–60]. For central rapidities $\eta \simeq 0$ and $p_\perp = 2$ GeV, one probes $x_2 \sim 10^{-2}$, which is large enough for the high-energy evolution to be negligible. Then the gluon density in the nucleus is large just because of the many ($3A$) valence quarks acting as sources for gluon radiation. An incoming parton from the deuteron scatters off this dense gluonic system and thus acquires an additional transverse momentum of order $Q_s(A, x_2) \sim 1$ GeV. This yields a shift in the spectrum of the produced particles towards higher values of $p_\perp$, leading to the Cronin peak. For forward rapidities, say $\eta = 3.2$ and $p_\perp = 2$ GeV, one has $x_2 \sim 10^{-4}$ and then the high-energy evolution with $Y = \ln(1/x_2)$ is important. In that case, one can use the factorization (34) for both d+Au and for the p+p collision which serves as a benchmark, leading to

$$R_{d+Au} \simeq \frac{1}{A^{1/3}} \frac{\Phi_A(Y, p_\perp)}{\Phi_p(Y, p_\perp)}. \quad (41)$$

For this kinematics, the saturation effects are important in the gold nucleus (since the nuclear saturation scale $Q_s(A, Y) \sim 2$ GeV is comparable with $p_\perp$) and lead to a slow down of the evolution with Y . On the other hand, saturation is still negligible for the proton, so the corresponding unintegrated gluon distribution $\Phi_p(Y, p_\perp)$ rises rapidly with Y , according to the BFKL evolution. Hence, when increasing η (and thus Y), the denominator in Eq. (41) rises much faster than the numerator there, leading to a decrease in the ratio. This is precisely the trend seen in the data.

⁵For d+Au collisions the number of binary collisions in Eq. (40) should be evaluated as $N_{coll} \simeq 2A^{1/3}$ since the projectile deuteron involves 2 nucleons, i.e. twice as much as the proton.

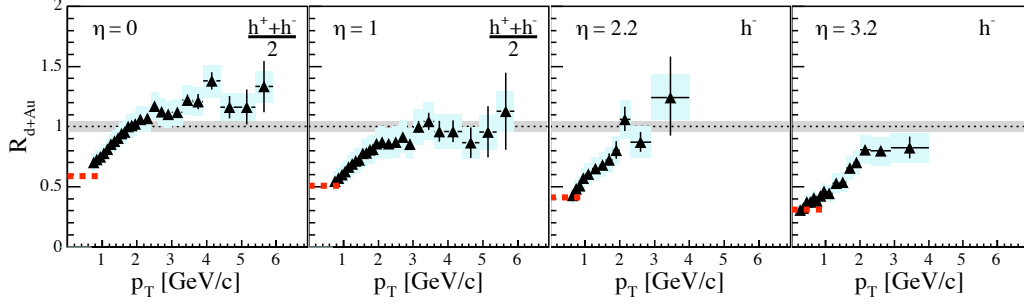


Fig. 18: BRAHMS results for d+Au collisions at 200 GeV/nucleon [55]: the ratio R_{dAu} for charged hadrons as a function of $p_{\perp}$ for central and forward pseudorapidities.

In particular, for sufficiently high Y , such that both the nucleus and the proton in the ratio in Eq. (41) are at saturation, and for transverse momenta comparable or slightly larger to the nuclear saturation momentum ($p_{\perp} \gtrsim Q_s(A, Y) \simeq A^{1/3} Q_s(p, Y)$), one can use Eq. (36) for the unintegrated gluon distributions in both targets. This yields a particularly simple result,

$$R_{d+Au} \simeq \frac{1}{A^{1/3}} \left(\frac{Q_s^2(A, Y)}{Q_s^2(p, Y)} \right)^{\gamma_s} \simeq A^{-\frac{1-\gamma_s}{3}}, \quad (42)$$

which is independent of both the transverse momentum (within a limited range above the nuclear saturation momentum) and the rapidity. It would be interesting to test this prediction in p+A collisions at the LHC, where the coverage in both Y and $p_{\perp}$ will be larger than at RHIC.

Another important consequence of saturation, which is manifest in Eqs. (36)–(37) and has been implicitly used in deriving Eq. (42), is geometric scaling in the gluon distribution. Via Eq. (34), this suggests a similar scaling in the spectrum of the produced particles. This prediction has been tested against the LHC data for p+p collisions [61], with the results shown in Fig. 19. There one can see the ratio

$$R_{E_1/E_2}(p_{\perp}, Y) = \frac{(dN/d^2p_{\perp}d\eta)|_{E_1}}{(dN/d^2p_{\perp}d\eta)|_{E_2}} \quad (43)$$

between the measured spectra for single-inclusive charged hadron production at two COM energies, $E_1 = \sqrt{s_1}$ and $E_2 = \sqrt{s_2}$, and for midrapidities: $|\eta| \leq 2.4$. More precisely, one displays two such ratios, as obtained by combining the LHC data for three different energies: $\sqrt{s} = 0.9, 2.36$, and 7 GeV. (Note that these energies are high enough for the saturation effects to be important in the proton wavefunction for $\eta \simeq 0$ and semi-hard transverse momenta.) If the spectrum $(dN/d^2p_{\perp}d\eta)$ scales as a function of $\tau \equiv p_{\perp}/Q_s(Y)$, then the ratio (43) should be equal to one when plotted as a function of τ . This expectation is indeed met by the data, as shown in the r.h.s of Fig. 19. At this point, it is worth noting that geometric scaling has been first observed in the HERA data for DIS [40] and that the experimental search for this remarkable behaviour has been inspired by the theoretical ideas about gluon saturation.

Another remarkable regularity in the data which is naturally explained by the CGC is the *limiting fragmentation*: when the charged particle rapidity distribution $dN/d\eta$ is plotted as a function of the variable $\eta' = \eta - y_{\text{beam}}$ — the rapidity difference between the produced particle and the dilute projectile (say, the deuteron in the case of d+Au collisions) —, then the distribution turns out to be independent of the collision energy over a wide range around $\eta' = 0$, whose extent is increasing with $\sqrt{s}$ (see Fig. 20). Note that $\eta' \simeq 0$ corresponds to forward rapidities ($\eta > 0$) according to our previous terminology, to which Eq. (34) applies. Moreover, for such rapidities, even a nucleus-nucleus collision may be viewed as ‘dilute-dense’, in the sense that one of the nuclei is probed at large x_1 , where it looks dilute.

To understand limiting fragmentation on the basis of Eq. (34), notice that (i) $\eta - y_{\text{beam}} \simeq \ln x_1$ (the rapidity of the produced particle is roughly equal to that of the fast parton which initiated the scatter-

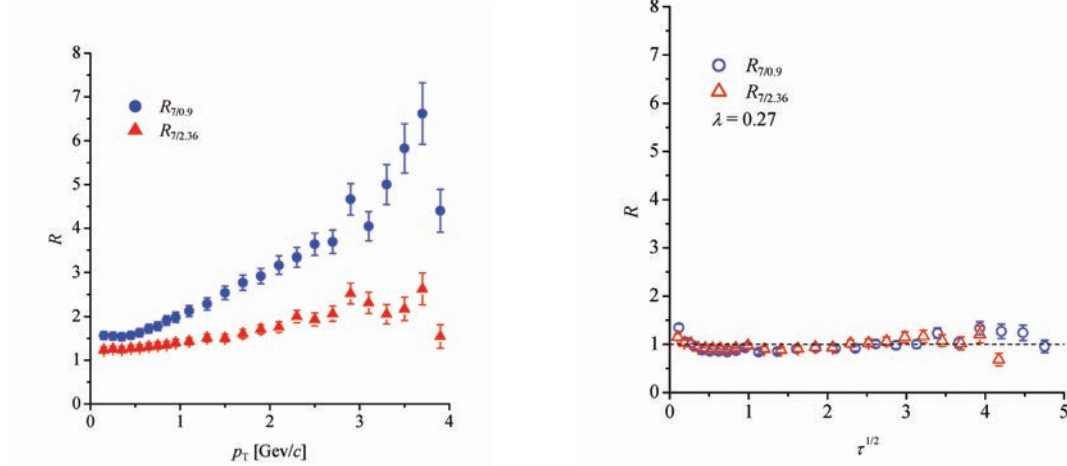


Fig. 19: The ratio R_{E_1/E_2} between particle spectra at energies E_1 and E_2 , as measured for three energies at the LHC, is plotted as a function of $p_\perp$ (left panel) and of the scaling variable $\tau = p_\perp/Q_s(Y)$ (right panel). From Ref. [61].

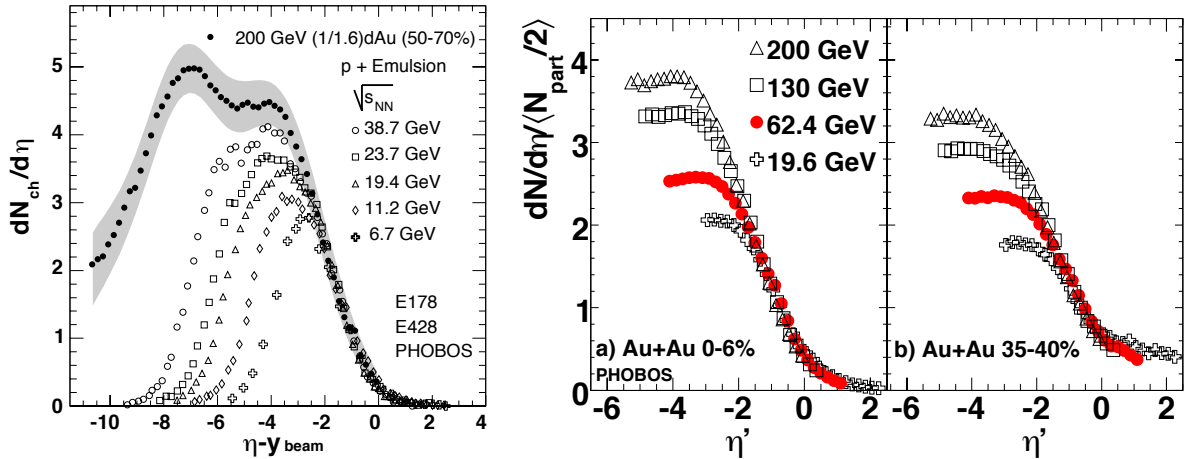


Fig. 20: Pseudorapidity density distributions of charged particles emitted in d+Au collisions (left panel) and Au+Au collisions (right panel, separately for central and peripheral impact parameters), at various energies, as a function of the $\eta' = \eta - y_{\text{beam}}$ variable. In the right panel, the impact parameter is represented by the centrality bins: 0-6% (i.e. the 6% most central data) and respectively 35-40%. From Ref. [63].

ing), (ii) the ‘multiplicity’ $dN/d\eta$ is obtained by integrating the spectrum (34) over all values of $p_\perp$, (iii) this integral is dominated by $p_\perp \sim Q_s(Y)$ (since the gluon distribution is strongly peaked at Q_s , cf. Fig. 16), and (iv) the result of the integration is very weakly dependent upon $Y = \ln(1/x_2)$, by geometric scaling (indeed, $\Phi_A(Y, p_\perp)$ is roughly a function of $\tau \equiv p_\perp/Q_s(Y)$, which is evaluated at $\tau = 1$ when performing the integral). These arguments imply that $dN/d\eta$ depends upon $x_1 \simeq \exp(\eta - y_{\text{beam}})$ but is approximately independent of x_2 (and hence of the total collision energy) within the range of ‘forward rapidities’. This is precisely the property called limiting fragmentation, as visible in Fig. 20 for both d+Au and Au+Au collisions at RHIC (PHOBOS) [62, 63].

As a final application for ‘dilute-dense’ scattering, we shall consider the production of a pair of hadrons at forward rapidities. The kinematics has been already explained in relation with Eq. (13) : when both η_a and η_b are positive and large, the produced hadrons explore small values $x_2 \ll 1$ in the

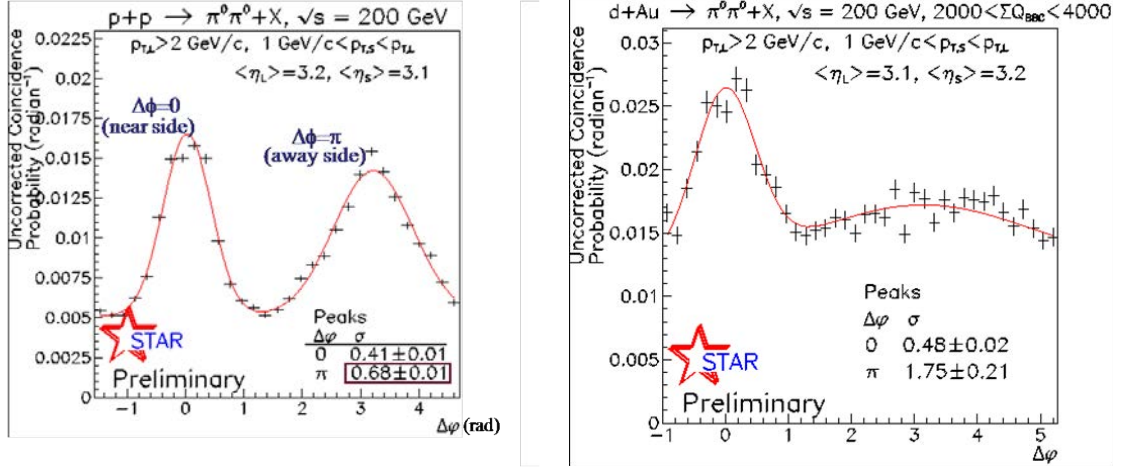


Fig. 21: Di-hadron azimuthal correlations at forward rapidities ($\eta_a, \eta_b \simeq 3$) and semi-hard transverse momenta as measured in p+p (left panel) and d+Au (right panel) collisions at RHIC, for $\sqrt{200}$ GeV per nucleon pair.

target wavefunction and hence they can experience high density effects — multiple scattering and gluon saturation. As before, these effects are important when the transverse momenta $p_{a\perp}$ and $p_{b\perp}$ of the two hadrons are comparable to the target saturation momentum. The respective cross-section admits a factorization similar to Eq. (32), where however the target expectation value now involves the trace of the product of *four* Wilson lines: two for the produced partons in the direct amplitude, and two for the same partons in the complex conjugate amplitude. Hence, the generalized ‘gluon distribution’ of Eq. (35) gets now replaced by a 4-point function, known as a *colour quadrupole* [22, 46–49].

A convenient way to study high-density effects in this setup is to measure *di-hadron correlations in the azimuthal angle*, which is the angle indicating the direction of propagation in the transverse plane. If the medium effects are negligible, two relatively hard particles ($p_\perp \gg \Lambda_{\text{QCD}}$) are produced back-to-back ($\mathbf{p}_{a\perp} + \mathbf{p}_{b\perp} \simeq 0$). So, if the trigger detects one of these particles together with its fragmentation products (a ‘jet’), then by measuring the particle distribution in the same event one should find another ‘jet’ at a relative angle $\Delta\Phi \simeq \pi$. On the other hand, if the target looks dense on the transverse resolution scale of the produced particles, then the $p_\perp$ -distribution gets broadened via multiple scattering and the peak corresponding to the ‘away jet’ at $\Delta\Phi \simeq \pi$ gets smeared, or it even disappears.

This scenario has been indeed confirmed by the RHIC data as measured by STAR [64, 65]. Figure 21 shows the experimental results for di-hadron production at forward rapidities ($\eta_{a,b} \simeq 3$) and semi-hard transverse momenta in both p+p and d+Au collisions. As already discussed in relation with Eq. (40), for this kinematics one expects the saturation effects to be negligible for a proton target, but important for a gold nucleus. And indeed, the data for azimuthal correlations in Fig. 21 show a pronounced ‘away’ peak in the p+p collisions (the left panel), but a strongly suppressed one in the d+Au ones (the right panel). Such a suppression was actually predicted by the CGC effective theory [48, 66] and its experimental observation at RHIC is one of the most compelling evidences in favour of gluon saturation available so far.

Turning to A+A (or ‘dense-dense’) collisions, we recall that, in that case, the CGC framework provides the initial conditions for the subsequent evolution of the liberated partonic matter, but not also the spectrum of the produced hadrons. Still, predictions can be more for more inclusive quantities like the total multiplicity, which are expected to be less affected by ‘final state’ interactions. In particular, Fig. 22 exhibits the centrality dependence of the multiplicity of the charged particles as measured at the LHC in Pb+Pb collisions at $\sqrt{s} = 2.76$ TeV/nucleon pair, together with the respective predictions of various theoretical models. The centrality of a collision refers to the relative impact parameter $b_\perp$ of the two projectiles in the transverse plane (see Fig. 23). For A+A collisions, this is often parameterized

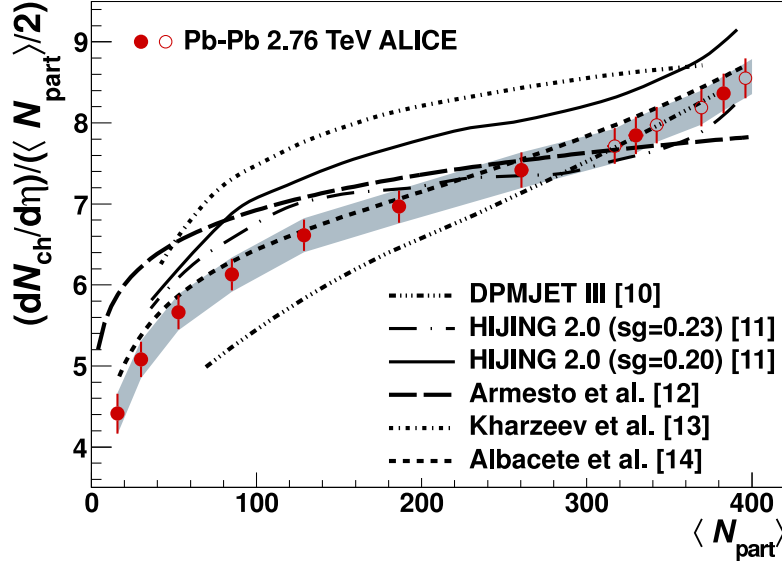


Fig. 22: The LHC data (ALICE [67]) for the midrapidity ($\eta = 0$) charged particle multiplicity in Pb+Pb collisions at $\sqrt{s} = 2.76$ TeV/nucleon pair, normalized by the number of participants N_{part} and plotted as a function of N_{part} — hence of the centrality of the collisions, which increases (meaning that the impact parameter $b_{\perp}$ decreases) with increasing N_{part} . The predictions of six theoretical calculations are also shown. The three lowest ones, marked as ‘Armesto et al’, ‘Kharzeev et al’, and ‘Albacete et al’, are inspired by the CGC effective theory and explicitly include gluon saturation. The three other models include some of the effects of saturation, in the form of an energy-dependent ‘infrared’ cutoff at low transverse momenta, $p_{\perp}^{\text{min}} \sim E^{\lambda}$, which effectively plays the role of Q_s .

in terms of the ‘number of participants’ — the number of incoming nucleons from the two nuclei in the region where the nuclei overlap with each other (the ‘interaction region’). Clearly, central collisions ($b_{\perp} \simeq 0$) involve more participants than the peripheral (large $b_{\perp}$) ones. Although one cannot compute $dN_{\text{ch}}/d\eta$ fully from first principles (as this also requires some information about the distribution of nucleons within the nuclear disk), one can easily estimate the dependence of this quantity upon N_{part} within the framework of the CGC effective theory. As discussed in relation with Fig. 17, the partons produced in the early stages of a A+A collision are typically gluons with transverse momenta $p_{\perp} \sim Q_s$, which have been liberated by the collision. So the multiplicity $dN_{\text{ch}}/d\eta$ near $\eta = 0$ is proportional to the number of such gluons which were present in the initial nuclei, within their region of overlapping:

$$\left. \frac{dN_{\text{ch}}}{d\eta} \right|_{\eta=0} \propto S \frac{xg_A(x, Q_s^2)}{\pi R_A^2} \propto \frac{1}{\alpha_s(Q_s^2)} S Q_s^2(A, E) \sim N_{\text{part}} E^{\lambda} \ln \frac{Q_s^2(A, E)}{\Lambda_{\text{QCD}}^2}. \quad (44)$$

Here, S is the transverse area of the interaction region and we have used Eq. (23) for the (nuclear) saturation momentum together with the fact that $S Q_s^2(A, E)$ is proportional to N_{part} and it grows with the COM energy $E = \sqrt{s}$ like E^{λ} (cf. Eq. (25) and the discussion of Fig. 13). Hence, the ratio $(dN_{\text{ch}}/d\eta)/N_{\text{part}}$ is expected to be only weakly dependent upon N_{part} , via the corresponding dependence of the running coupling: $1/\alpha_s(Q_s^2) \sim \ln(Q_s^2/\Lambda_{\text{QCD}}^2) \sim \ln N_{\text{part}}$. This is in good agreement with the data, as shown in Fig. 22. (For the most refined calculation to date, whose results are indicated in Fig. 22 by the curve denoted as ‘Albacete et al’, see Ref. [68].)

3.7 The Glasma

By the uncertainty principle, it takes a time $\Delta t \sim 1/Q_s$ to liberate a particle with transverse momentum $p_{\perp} \sim Q_s$ at midrapidities ($\eta \simeq 0$). So, by that time, the small- x gluons which were originally confined

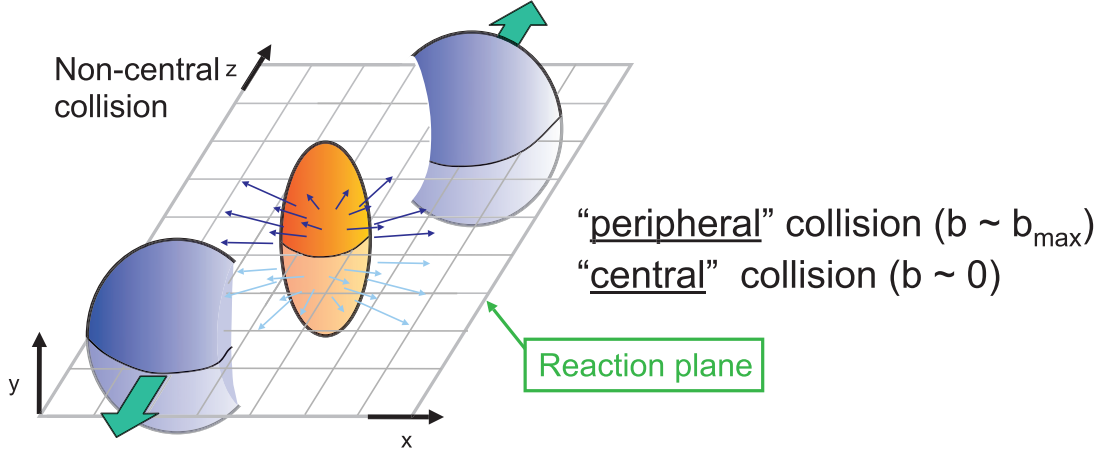


Fig. 23: A sketch of the geometry of a heavy ion collision. The collision (or longitudinal) axis is denoted as z , while x and y are the transverse coordinates. The interaction region (the almond-shape region where the two nuclei overlap which each other) is singled out. N_{part} is the number of nucleons in this region. The interaction region is horizontally cut by the reaction plane (x, z) .

within the wavefunctions of the incoming nuclei, are released by the collision. What is the subsequent evolution of these gluons? To answer this question one needs a better understanding of their configuration at the time of emission. This can be inferred from the solution A_a^μ to the Yang–Mills equation (39), or, more precisely of the associated chromo-electric and magnetic fields, $\mathbf{E}_a$ and $\mathbf{B}_a$.

Prior to the collision ($t < 0$), these fields describe the two ‘colour glass condensates’ of the incoming nuclei, which by causality are independent of each other. Due to the high-energy kinematics, the CGC fields turn out to be quite simple (see Fig. 24): for each nucleus, the respective vectors $\mathbf{E}_a$ and $\mathbf{B}_a$ have only transverse components, E_a^i and B_a^i with $i = 1, 2$, meaning that they are orthogonal to the collision axis x^3 . Besides, they are also orthogonal to each other, $\mathbf{E}_a \cdot \mathbf{B}_a = 0$ (for each of the $N_c^2 - 1 = 8$ values of the colour index a), and they have equal magnitudes:

$$\mathbf{E}_a \perp \mathbf{B}_a \perp x^3, \quad |\mathbf{E}_a| = |\mathbf{B}_a| \quad (\text{prior to the collision}). \quad (45)$$

These initial electric and magnetic fields are localized near $x^- = 0$ for the left-moving nucleus and, respectively, $x^+ = 0$ for the right-moving one, like the respective colour charges. In a given event, their values and orientations can randomly vary from one point to the other in the transverse plane. But on the average, the fields at different points are correlated due to ‘memory’ effects in the high energy evolution, in particular, due to saturation. The correlations, which are encoded in the respective CGC weight functions, are typically restricted to a *saturation disk*, i.e. to transverse areas with radius $\sim 1/Q_s$: domains separated by transverse distances $\Delta x_\perp \gg 1/Q_s$ evolve independently from each other, since saturation prohibits the emission of gluons with momenta $k_\perp \ll Q_s$. Within a saturation disk, gluons arrange themselves in such a way to shield their colour charges and thus minimize their mutual repulsion; accordingly, a saturation disk has zero overall colour charge (see the right figure in Fig. 3). Also, gluons can be correlated with each other *in rapidity*, due to the fact that they have common ancestors, i.e. they belong to the same parton cascade. Such correlations extend over a rapidity interval $\Delta Y \sim 1/\alpha_s$, since this is the typical value of Y which is required to build parton cascades according to the BFKL evolution. These correlations, which were built in the initial wavefunctions via the high-energy evolution, get transmitted to the gluons liberated by the collision and thus have consequences for the distribution of the particles in the final state.

In view of the above, a A+A collision can be viewed as the scattering between two sheets of coloured glass, as illustrated in Fig. 24. Incidentally, a similar structure for the incoming fields — electric

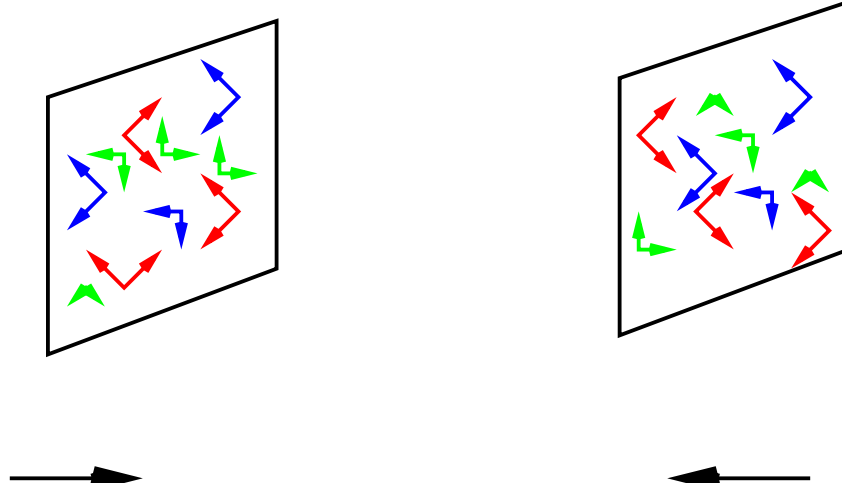


Fig. 24: The chromo-electric and chromo-magnetic field configurations in the two nuclei prior to the collision.

and magnetic fields which are orthogonal to the beam axis and to each other, and which are localized near the respective light-cone and frozen by Lorentz time dilation — would also hold if the nuclei were made with *electric* (rather than *colour*) charges. In both QED and QCD, such field configurations — known as the Weizsäcker–Williams fields — represent the boosted version of the Coulomb fields created by the ensemble of (electric or colour) charges in their rest frame. However, the non-Abelian structure of QCD is essential for having a *collision*: the Abelian version of Eq. (39) would be linear and it would not describe a scattering process. (The total electromagnetic field at $t > 0$ would be simply the sum of the individual fields of the two nuclei.) The non-linear effects encoded in the Yang–Mills equation (39) describe the scattering between the small- x gluons in the two CGC’s. The corresponding solution at $t > 0$ (more precisely, in the forward light cone at $x^+ > 0$ and $x^- > 0$, which is the space-time region causally connected to the collision) represents the gluonic matter produced by the collision.

This solution exhibits a very interesting structure: in addition to the *transverse* fields on the two sheets, which after the collision are separating from each other, there are also *longitudinal*, electric and magnetic fields, E_a^3 and B_a^3 , which extend along the collision axis. The latter give rise to *colour flux tubes* (or ‘strings’) with the endpoints on the two sheets and a typical transverse radius $1/Q_s$ (see Fig. 25 left). Right after the collision ($t = 0_+$), these fields are quite strong, $E^3 \sim B^3 \sim 1/g$, since they carry most of the energy of the original CGC fields. At such early times, the gluonic matter is still in a high-density, coherent state, for which a description in terms of classical fields is better suited than one in terms of particles. But with increasing time, the system expands, its density decreases, and so does the strength of the fields. After a time $t \sim 1/Q_s$, the magnitudes of all the fields (transverse and longitudinal) becomes of order one, meaning that, from now on, these fields can be also interpreted as incoherent superpositions of *particles*. The spectrum of these particles (mostly gluons) is obtained from the Fourier modes of the colour fields at time $t \gtrsim 1/Q_s$. These particles can interact with each other, as their density is still quite high (albeit decreasing with time, due to expansion). As we shall see, these interactions are expected to lead to a phase of local thermal equilibrium — the quark-gluon plasma (QGP).

The intermediate, non-equilibrium, form of matter, which interpolates between the CGC in the initial wavefunctions and the QGP at later stages is known as the *glasma* (a name coined as a combination of ‘glass’ and ‘plasma’) [69]. The main, qualitative, feature of the glasma is the presence of longitudinal colour flux tubes with transverse area $\sim 1/Q_s^2$, which are *boost invariant*: the colour fields depend upon the *proper time* τ but not also upon the *space-time rapidity* η_s . The variables τ and η_s , which can be used

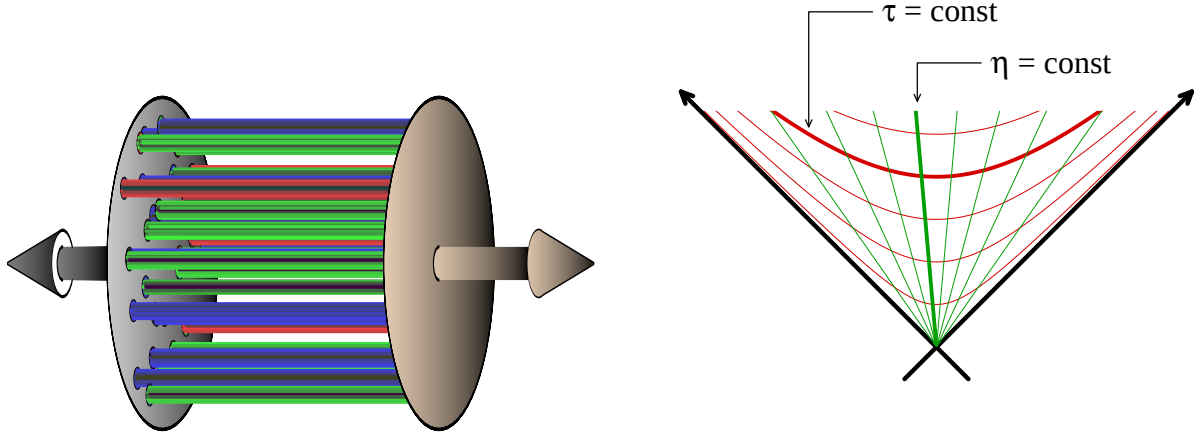


Fig. 25: Left: the longitudinal colour flux tubes which develop in between the remnants of the two nuclei in the early stages ($\tau \lesssim 1/Q_s$) of a A+A collision. Right: various choices of coordinates in the future light cone.

instead of t and $z \equiv x^3$ in the forward LC, are defined as

$$\tau = \sqrt{t^2 - z^2}, \quad \eta_s = \frac{1}{2} \ln \frac{t+z}{t-z} \implies \tanh \eta_s = \frac{z}{t}. \quad (46)$$

Under a boost along the z axis, τ is invariant while η_s is shifted by a constant. Lines of constant τ and of constant η_s are shown in Fig. 25 (right). The fact that the glasma fields depend upon τ but not upon η_s is a consequence of the symmetries of the collision, as encoded in the classical field equations (39). This is also consistent with the hypothesis of *uniform longitudinal expansion*, as originally formulated by Bjorken. Specifically, Bjorken has assumed that (i) after being produced at $t \simeq z \simeq 0$, the particles undergo free longitudinal streaming, meaning that they keep a constant velocity along the z axis; accordingly, the particles that can be found at some later time t at point z are those with a longitudinal velocity $v_z = z/t$; (ii) the distribution of the produced particles is uniform in v_z . Together, (i+ii) imply that the distribution at time t is independent of z/t , hence of η_s . Note that this argument identifies the *momentum* rapidity y (cf. Eq. (11)) of the produced particles with their *space-time* rapidity η_s : $\tanh y \equiv v_z = z/t \equiv \tanh \eta_s$. Hence, the boost invariance of the glasma fields implies that the distribution of the particles produced by the decay of these fields is independent of y . This is a generic feature of the particle production at the *classical* level, that is, on an event-by-event basis: the associated spectra are boost invariant. But the *physical* spectra, as obtained after averaging the classical results with the CGC weight functions of the incoming nuclei, cf. Eq. (38), *are* rapidity-dependent, because of the respective dependencies of the weight functions, as introduced by the quantum evolution with Y .

An interesting consequence of the above considerations, which might be related to a remarkable phenomenon seen in the RHIC [70–73] and the LHC data [74–76] and known as the *ridge*, refers to the rapidity dependence of the two particle correlation. The latter is defined as

$$C_2(\eta_a, \mathbf{p}_{a\perp}; \eta_b, \mathbf{p}_{b\perp}) = \left\langle \frac{dN_2}{d\eta_a d^2\mathbf{p}_{a\perp} d\eta_b d^2\mathbf{p}_{b\perp}} \right\rangle - \left\langle \frac{dN}{d\eta_a d^2\mathbf{p}_{a\perp}} \right\rangle \left\langle \frac{dN}{d\eta_b d^2\mathbf{p}_{b\perp}} \right\rangle, \quad (47)$$

and what is generally plotted is the ratio $\mathcal{R}$ between this correlation and its disconnected part (below, N_a denotes the number of particles of type a in a given bin in pseudo-rapidity and azimuthal angle),

$$\mathcal{R} \equiv \frac{\langle N_a N_b \rangle - \langle N_a \rangle \langle N_b \rangle}{\langle N_a \rangle \langle N_b \rangle}, \quad (48)$$

as a function of the rapidity and the azimuthal separations between the two particles, $\Delta\eta = \eta_a - \eta_b$ and $\Delta\phi = \phi_a - \phi_b$. Remarkably, the data for A+A collisions at both RHIC and the LHC show the existence

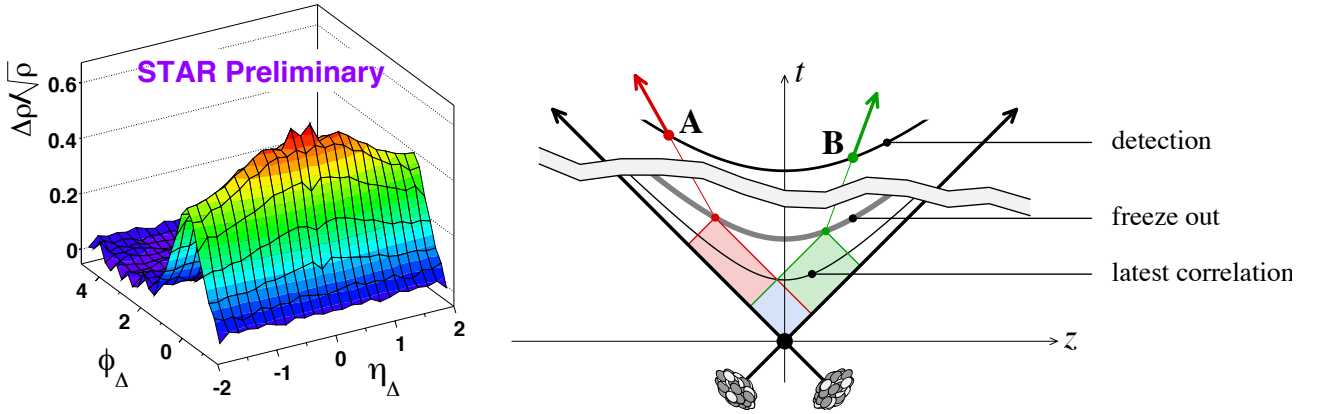


Fig. 26: Left: the ‘ridge’ in the di-hadron correlations as measured by RHIC (STAR). Right: the causal structure of the di-hadron rapidity correlations; the 2 hadrons A and B propagating with different rapidities were causally connected in the space-time region where the respective backward light-cones with vertices on the freeze-out surface were overlapping with each other.

of correlations which extend over a large rapidity interval $\Delta\eta \simeq 4 \div 8$, but restricted to small azimuthal separations $\Delta\phi \simeq 0$ (see Fig. 26 left). This means that particles which propagate along very different directions with respect to the collision axis preserve nevertheless a common direction of motion in the transverse plane. By causality, such a correlation must have been produced at early times, when these particles — which rapidly separate from each other — were still causally connected (see the right panel of Fig. 26). A simple estimate gives

$$\tau_{\max} = \tau_{\text{freeze-out}} e^{-\frac{|\Delta\eta|}{2}}, \quad (49)$$

for the latest time at which these particles could have been correlated. For a freeze-out time $\tau_{\text{freeze-out}} \approx 10 \text{ fm/c}$, and rapidity separations $\Delta\eta \geq 4$, one sees that these correlations must have been generated before 1 fm/c .

Long-range rapidity correlations are natural in the glasma picture, where all the spectra (in particular, the 2-particle ones) are independent of y — at least, at the classical level. After averaging with the CGC weight functions, as in Eq. (38), the 2-particle spectrum acquires a dependence upon the *average* rapidity of the two particles $(\eta_a + \eta_b)/2$, but not upon their difference $\Delta\eta$. This last argument remains correct so long as one can neglect quantum corrections due to soft gluon emissions within the rapidity interval $\Delta\eta$, which in turn requires $\Delta\eta \lesssim 1/\alpha_s$. Concerning the azimuthal collimation of the ridge, this is not a consequence of the glasma — the particles produced via the decay of the classical fields are emitted isotropically in the transverse plane in the rest frame of the medium —, but can be generated via *radial flow*: the local fluid element has some transverse velocity $\mathbf{v}_\perp$, which introduces a bias in the azimuthal distribution of the particles produced by the decay of a same flux tube (the final correlation is peaked around the direction of $\mathbf{v}_\perp$). Flow phenomena will be discussed in more detail in the next section.

For sufficiently high energy, the long-range rapidity correlations invoked for A+A collisions should be also present in the p+p collision. In that case, one expects no flow, as there are fewer produced particles and the freeze-out time is shorter. Yet, a small ridge has been measured in p+p collisions by the CMS collaboration at the LHC [77], but only in the high-multiplicity events (which are believed to be more central) and within a limited range in $p_\perp$ (from 1 to 3 GeV), which is in the ballpark of the proton saturation momentum at the LHC. In that case, the azimuthal collimation could be explained by an intrinsic angular correlation between the emission of two particles from the glasma field [78].

Let us also note another possible consequence of the glasma flux tubes: the presence of longitu-

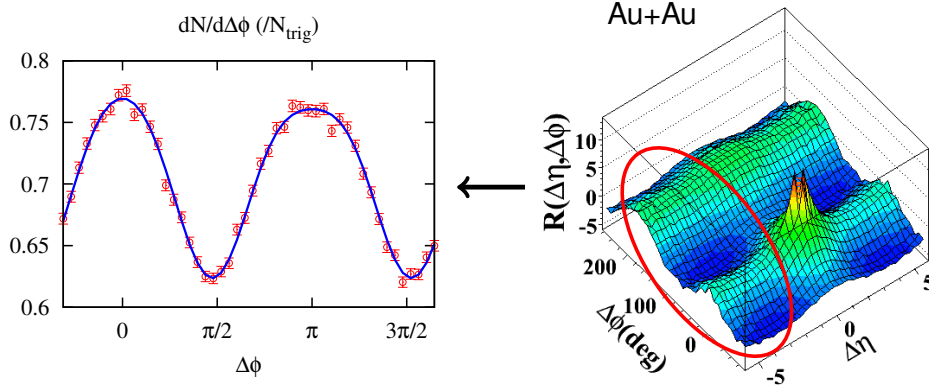


Fig. 27: The double-peak structure characteristic of elliptic flow: two wide peaks at $\Delta\phi = 0$ and $\Delta\phi = \pi$ which extend over a wide interval $\Delta\eta$ in rapidity. The strong correlation peak visible at $(\Delta\phi, \Delta\eta) \simeq (0, 0)$ is associated with the fragmentation of the trigger jet.

dinal electric and magnetic fields implies the existence of topologically non-trivial configurations, characterized by a large density of *Chern–Simons topological charge*. Such configurations are interesting in that they break the charge-parity (CP) symmetry: via the chiral anomaly, they generate a difference between the number of quarks with right-handed and respectively left-handed helicity. In the context of HIC, they may generate a new phenomenon, known as the *chiral magnetic effect* [79]: quarks with opposite helicities can be separated by the ultra strong magnetic fields ($B \sim 10^{18}$ Gauss) created in the peripheral ultrarelativistic A+A collisions, thus leading to a charge asymmetry between the two sides of the reaction plane. Since the direction of the magnetic field varies from one collision to another, this effect leads to fluctuations in the distribution of the electric charge of the final hadrons. These theoretical expectations appear to be supported by measurements at RHIC (STAR) [80].

4 The Quark Gluon Plasma

The main topic of this chapter is the *quark-gluon plasma* (QGP) — the partonic form of QCD matter in thermal equilibrium which exists for sufficiently large temperatures, as demonstrated by numerical calculations in lattice QCD. This form of matter is expected to be created during the intermediate stages of a ultrarelativistic HIC, albeit only for a short lapse of time and in a state of only *local* thermal equilibrium. We shall first review the main experimental evidence in favour of QGP in a HIC, namely the observation of *flow* in the particle production and its successful description in terms of *hydrodynamics*. Then we shall discuss the QGP *thermodynamics* from the viewpoints of lattice theory and perturbative QCD. We shall also mention the difficulty of perturbation theory to describe the dynamics *out-of-equilibrium* (in particular, the transport coefficients and the process of thermalization). Then we shall consider the phenomenon of *jet quenching* (the energy loss by an energetic parton via interactions in the plasma), which is an important tool for exploring the deconfined matter produced in HIC's. At several places in what follows, we shall encounter situations where perturbation theory appears to be insufficient and which may signal a regime of *strong coupling*. To address such situations from the opposite limit — that of a coupling which is *arbitrarily* strong —, one can rely on techniques borrowed from string theory, via the *AdS/CFT correspondence*. This will be briefly discussed (in relation with the physics of HIC's) in the last section of these lectures.

4.1 Correlations and flow in HIC

In our previous discussion of the ‘ridge’, in Section 3.7, we focused on the long-range rapidity correlations which are rather strongly peaked in $\Delta\phi = \phi_a - \phi_b$ near $\Delta\phi = 0$, leading to the resemblance

with a mountain ridge. But as a matter of fact, the di-hadron correlations measured in A+A collisions at RHIC and the LHC show an even more pronounced double-peak structure, visible in Fig. 27, with large but wider peaks at $\Delta\phi = 0$ and $\Delta\phi = \pi$. (In the analysis leading to Fig. 26, this structure has been subtracted away to render the ridge effect visible.) This is known as *elliptic* (or ‘transverse’) *flow* [81]. As also visible in Fig. 27, this double peak structure can be well parameterized as

$$\left\langle \frac{dN_{pairs}}{d\Delta\phi} \right\rangle \propto v_2^2 \cos(2\Delta\phi), \quad (50)$$

with v_2 the ‘magnitude of the elliptic flow’.

The explanation of this phenomenon turns out to be quite simple: it reflects the anisotropy of the interaction region — the almond-shape region where the two nuclei overlap with each other; see Fig. 23 — for non-central collisions. This anisotropy entails a pressure gradient in the initial conditions: the pressure is larger along the minor axis of the ellipse (the x axis in the left panel of Fig. 28) rather than along the major one; accordingly, more particles will be emitted in the direction of the largest gradient. This ultimately generates an anisotropy in the azimuthal distribution of the produced particles, which for symmetry reasons is of the form shown in Eq. (50):

$$\frac{dN}{d\phi} \propto 1 + 2v_2 \cos 2\phi. \quad (51)$$

This argument looks simple, as anticipated, but there is something deep about it: the role played by *collective phenomena* like pressure gradients or flow. Such phenomena are natural for many-body systems which relatively *strong interactions* — sufficiently strong to be able to transmit the asymmetry of the initial geometry into properties of the final state. This is an important point to which we shall return.

The qualitative arguments above suggests that the coefficient v_2 characterizing the strength of the anisotropy should increase with centrality. This trend is indeed seen in the data (at least, for not too peripheral collisions, for which the interaction region becomes tiny and dilute). Also, particles which experience a stronger pressure gradient are expected to have a larger transverse momentum, as they inherit the velocity of the fluid; so, v_2 should rise with $p_\perp$. This expectation, too, is confirmed by the data, at least for not too large $p_\perp \lesssim 5$ GeV: very hard particles cannot be driven by the medium, so for them v_2 is naturally small. The measurements of v_2 (say, via 2-hadron correlations) yield very similar results at RHIC and the LHC (see the left panel of Fig. 30), showing that v_2 is roughly independent of the COM energy. Note also that the typical values of v_2 for semi-hard momenta are relatively large, $v_2 \sim 0.2$, meaning that the collective phenomena alluded to above are indeed quite strong.

But the elliptic flow and the ridge are not the only collective phenomena hidden in the di-hadron correlations illustrated in Fig. 27. By looking at the most central collisions where v_2 is relatively small, one sees not only the narrow ‘ridge’ at $\Delta\phi \simeq 0$, but also a ‘double-hump’ on the away side, at $|\Delta\phi - \pi| \simeq 1.1$, which extends too over a large interval $\Delta\eta$ (see Fig. 29 left). The harmonic decomposition of this signal reveals higher Fourier modes with significant strengths, as illustrated in the right panel of Fig. 29. This leads to the following generalization of Eq. (51) :

$$\left\langle \frac{dN_{pairs}}{d\Delta\phi} \right\rangle \propto 1 + 2 \sum_{n=1}^{\infty} \langle v_n^2 \rangle \cos(n\Delta\phi) \quad (52)$$

where the various coefficients v_n up to v_6 have been extracted from the LHC data and they are compared to v_2 in the right panel of Fig. 30. (All these coefficients are roughly independent of η , meaning that they describe correlations over a wide interval $\Delta\eta$.)

What is the physics of such higher harmonics ? It is generally believed that they are the consequence of *fluctuations* in the distribution of nucleons within the interaction region, as illustrated in Fig. 28 (right panel) [82]. Namely, even though the overlapping region between the two nuclei has an

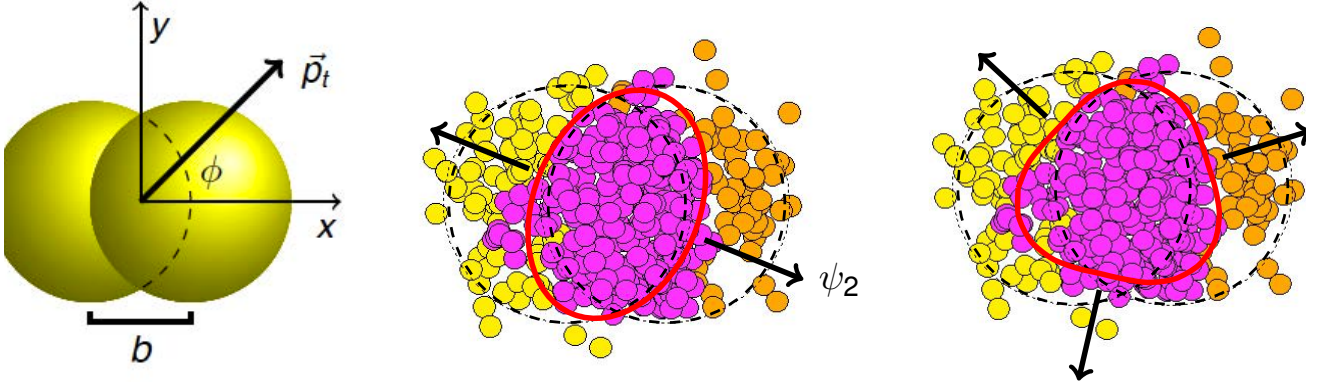


Fig. 28: Left: a frontal view of a HIC, illustrating the impact parameter $b_{\perp}$, the almond-shape interaction region, and the azimuthal angle ϕ . Central and right: the distribution of nucleons within the interaction region in a given event can include elliptic (v_2), triangular (v_3), or even higher harmonic modes. The reference angles ψ_2, ψ_3 show the tilt of the interaction region with respect to the geometrical ‘reaction plane’ (cf. Fig. 23).

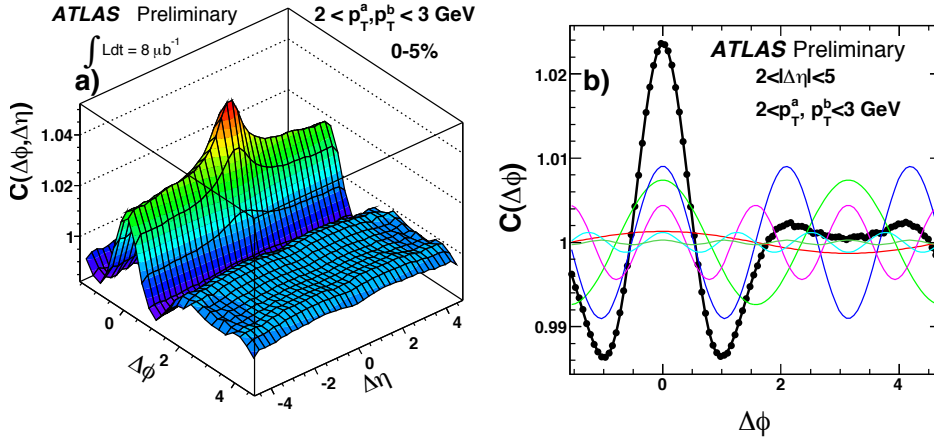


Fig. 29: The ‘ridge’ at $\Delta\phi \simeq 0$ and the ‘double-hump’ at $|\Delta\phi - \pi| \simeq 1.1$ as visible in the LHC data for di-hadron correlations in the 5% most central collisions (left panel), together with the harmonic decomposition of these correlations, cf. Eq. (52) (right panel) [83].

elliptic shape, the nuclear matter inside it is neither homogeneous, nor strictly ellipsoidal, because of fluctuations in the particle distribution. The azimuthal distribution in a given event can be decomposed into harmonics, with coefficients v_n and reference angles ψ_n :

$$\frac{dN}{d\phi} \propto 1 + \sum_{n=1}^{\infty} 2v_n \cos n(\phi - \psi_n). \quad (53)$$

The reference angles ψ_n are generally different from the conventional reaction plane ($\psi_{RP} = 0$) and are difficult to measure, but they drop out in the 2-particle correlations, as manifest in Eq. (52).

The fact that the initial geometry of the interaction region can have complicated fluctuations is not necessarily a surprise, given the granularity of the nucleons. What is remarkable though is the ability of the system to transmit these fluctuations into the distribution of the produced particles, via *transverse flow*. The effective theory for flow is *hydrodynamics* and will be succinctly discussed in the next section.

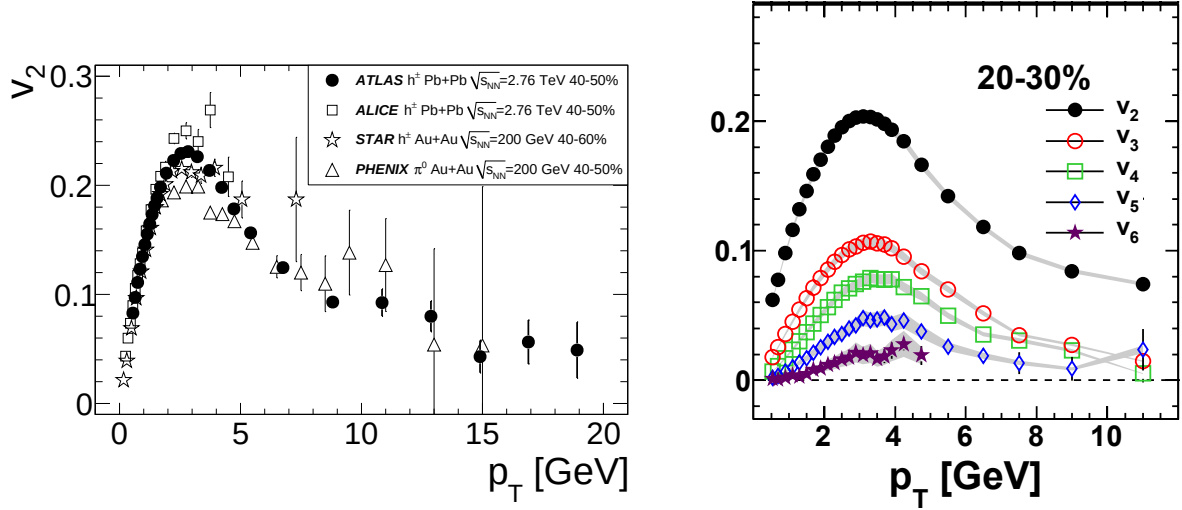


Fig. 30: Left: the RHIC and LHC results for v_2 as a function of $p_\perp$, for collisions within the centrality bin 40% – 50%; from Ref. [84]. Right: the LHC results (ATLAS) for $v_n(p_\perp)$, for $n = 2, 3, 4, 5$ and 6, and the centrality bin 20% – 30% [83].

4.2 Hydrodynamics and kinetic theory

Hydrodynamics is the theory which describes the flow of a fluid independently of its detailed microscopic structure. More precisely, the equations of hydrodynamics have an universal form (at least, for a given underlying fundamental theory, like QCD), but they involve a few ‘parameters’ which depend upon the nature of the fluid and can in principle be computed via microscopic calculations. The scope of hydrodynamics can be most easily explained with reference to thermodynamics. The latter describes a many-body system in *global thermal equilibrium*, in which the intensive quantities, like temperature, pressure and the chemical potentials associated with the various conserved charges (electric charge, baryonic charge, etc) are *time-independent* and *uniform* throughout the volume V of the system. Hydrodynamics can be viewed as a generalization of this picture towards a state of *local* equilibrium: the intensive quantities alluded to above can vary in space and time, but they do that so slowly that one can still assume thermal equilibrium to hold locally, in the vicinity of any point. Gradients of pressure and thermodynamics naturally lead to *flow*, with a local fluid velocity $\mathbf{v}$ which is itself slowly varying in space and time.

The equations of hydrodynamics are simply the ensemble of the relevant *conservation laws* — for the energy, momentum and the other conserved charges:

$$\partial_\mu T^{\mu\nu} = 0, \quad \partial_\mu J_B^\mu = 0, \quad \dots \quad (54)$$

where $T^{\mu\nu}$ is the energy-momentum tensor, J_B^μ is the density of the baryonic current (the volume integral of J_B^0 is the difference between the number of baryons and the number of antibaryons), and the dots stand for other conserved charges. These densities depend upon the intensive (local) quantities describing the state of the fluid: the energy density $\varepsilon = E/V$, the pressure P , the 4-velocity $u^\mu = \gamma(1, \mathbf{v})$ (with $\gamma = 1/\sqrt{1 - v^2}$), and a set of ‘friction coefficients’ known as *viscosities*, which characterize the dissipative properties of the medium.

The relations between the densities of the conserved charges ($T^{\mu\nu}$, J_B^μ , ...) and the intensive quantities are obtained via a *gradient expansion* with respect to the slow space-time variations of the latter. More precisely, this amounts to an expansion in powers of ℓ/R , where R is a characteristic size of the system (in a HIC, R is the transverse size of the interaction region) and ℓ is the *mean free path* of the particles composing the fluid (the typical distance between two successive collisions). This quantity will

play an important role in what follows, so let us open here a parenthesis and discuss it in more detail.

At least for sufficiently weak coupling, the mean free path ℓ can be estimated using *kinetic theory*. This is an effective theory too, but it applies at shorter, microscopic, scales: in that context, the mean free path is typically the *largest* scale in the problem (it is much larger than the Compton wavelength $\lambda \sim 1/k$ of a particle or the typical duration of a scattering processes). Kinetic theory allows one to follow the evolution of the particle distributions in *phase-space* — i.e. in space-time and in momentum space. To that aim, this theory involves more information about the microscopic dynamics, like cross-sections for the particles interactions described via the ‘collision term’ in the *Boltzmann equation* — the central equation of kinetic theory. But even without solving that equation, one can deduce an estimate for ℓ via simple considerations: the *collision rate* (the inverse of the typical time τ_{coll} between two successive collisions) scales like $\tau_{\text{coll}}^{-1} \sim nv_{\text{rel}}\sigma$, where n is the particle density, v_{rel} is their average relative velocity, and σ is the cross-section for their mutual interactions. The mean free path is then obtained as

$$\ell \sim v\tau_{\text{coll}} \sim \frac{v}{v_{\text{rel}}n\sigma} \sim \frac{1}{n\sigma}, \quad (55)$$

where v is the average velocity of the particles, so v/v_{rel} is a number of order one. Since σ is naturally proportional to some power of the coupling constant, Eq. (55) shows that the mean free path becomes smaller — meaning that *the hydrodynamical description works better* — when the coupling is *strong*.

To be more specific, consider a system that will play an important role in what follows: a *weakly coupled quark-gluon plasma* with (local) temperature T . This is a nearly ideal gas of ultrarelativistic particles, so the particle densities scale like $n \sim T^3$ separately for quarks and gluons. To leading order in α_s , scattering is controlled by the $2 \rightarrow 2$ elastic collisions shown in Fig. 31, where the external lines represent thermal particles with typical energies and momenta of order T . These processes yield $\sigma \propto \alpha_s^2$. However, for the processes involving the exchange of a gluon in the t channel, there is a logarithmic enhancement associated with the singularity of the Coulomb scattering at small angles: the Rutherford formula reads $d\sigma/d\Omega \propto \alpha_s^2/(T^2 \sin^4 \theta)$, with θ the scattering angle, and it is strongly divergent when $\theta \rightarrow 0$. The cross-section σ which is relevant for computing the mean free path (55) is not the *total* cross-section $\sigma_{\text{tot}} = \int d\Omega (d\sigma/d\Omega)$, but rather the *transport cross-section* :

$$\sigma = \int d\Omega (1 - \cos \theta) \frac{d\sigma}{d\Omega} \propto \int d\theta \sin \theta (1 - \cos \theta) \frac{\alpha_s^2}{T^2 \sin^4 \theta} \sim \frac{\alpha_s^2}{T^2} \int_g \frac{d\theta}{\theta} \sim \frac{\alpha_s^2}{T^2} \ln \frac{1}{\alpha_s}, \quad (56)$$

which more properly characterizes the efficiency of the interactions in redistributing energy and momentum. The factor $1 - \cos \theta$, which vanishes as $\theta^2/2$ at small angles, accounts for the fact that the small-angle scattering is inefficient in that sense, as intuitive from the fact that one cannot equilibrate an anisotropic energy-momentum distribution via collinear scattering. Due to this factor, the integral in Eq. (56) is only *logarithmically* divergent as $\theta \rightarrow 0$ (unlike σ_{tot} , which would be *quadratically* divergent). In reality, this divergence is screened by plasma effects which occur at the momentum scale gT (see the discussion in Section 4.3). This implies that the minimal collision angles are $\theta \sim gT/T \sim g$, corresponding to transferred momenta of order gT . Hence, to leading logarithmic accuracy, the integral can be estimated as shown in the r.h.s. of Eq. (56). In turn, this implies the following estimate for the mean free path in a weakly-coupled QGP (cf. Eq. (55))

$$\ell \sim \frac{1}{T} \frac{1}{\alpha_s^2 \ln(1/\alpha_s)}, \quad (57)$$

which as long as $g \ll 1$ is indeed much larger than both the Compton wavelength $\lambda \sim 1/T$ of the thermal particles and the typical duration $\sim 1/gT$ of a scattering process.

We now close the parenthesis dedicated to the mean free path and return to the discussion of hydrodynamics. As already mentioned, this is a legitimate effective theory for flow when $\ell \ll R$. The *constitutive relation* allows one to relate the energy-momentum tensor to the velocity field u^μ and its

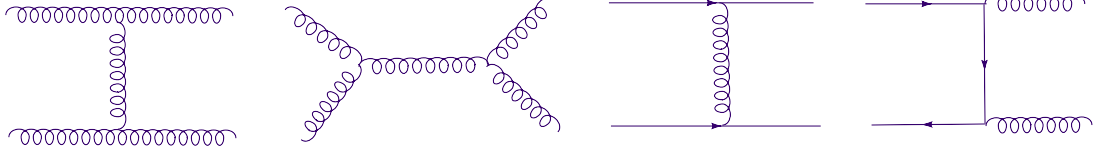


Fig. 31: $2 \rightarrow 2$ partonic processes dominating the cross-section for (elastic) scattering at weak coupling.

gradients, via an expansion in powers of ℓ/R . The powers of ℓ are associated with dissipative phenomena, while those of $1/R$ with the gradients in the fluid. To zeroth order in this gradient expansion one obtains the *ideal hydrodynamics*. This is ‘ideal’ in the sense that there is no dissipation. The corresponding structure of $T^{\mu\nu}$ follows entirely from the assumption of local thermal equilibrium. Namely, in the *local rest frame* of a fluid element ($u_{RF}^\mu = (1, 0, 0, 0)$), the energy-momentum tensor has the diagonal structure familiar from the thermodynamics: $T_{RF}^{\mu\nu} = \text{diag}(\varepsilon, P, P, P)$. Boosting to the laboratory frame, where the fluid 4-velocity is u^μ , this yields

$$T^{\mu\nu} = (\varepsilon + P)u^\mu u^\nu - P g^{\mu\nu}, \quad (58)$$

where $g^{\mu\nu} = \text{diag}(1, -1, -1, -1)$ is the Minkowski metric tensor. The r.h.s. of Eq. (58) involves 5 independent quantities: ε , P , and the 3 components v^i of the (local) velocity. Their space-time evolution is determined by the respective conservation law in Eq. (54), which yields 4 equations, together with the assumed *equation of state*, which specifies the functional relation $\varepsilon(P)$ between the energy density and the pressure (e.g., $\varepsilon = 3P$ for an ideal gas of massless particles).

Since ideal hydrodynamics ignores dissipation, one might think that it corresponds to a situation where the coupling is weak, but that would be wrong: it rather corresponds to *strong coupling*. This may seem counterintuitive but it can be understood as follows: the dissipative phenomena are proportional to the ability of the system to transfer momentum in a direction *perpendicular* to the fluid velocity (since such a transfer results in slowing down the flow). Within kinetic theory at least, this transfer is realized by particles moving throughout the fluid in between successive collisions. Hence, the rate for transfer is proportional to the mean free path (55) and thus to the inverse of the coupling. As an example, consider the *shear viscosity*: this characterizes the friction force between two neighboring layers of fluid which propagate, say, along the x axis, but at slightly different velocities (so there is a non-zero gradient $\partial u_x / \partial y$). There is friction because some longitudinal momentum p_x gets transferred from the faster layer to the slower one, at a rate proportional to the velocity gradient:

$$\frac{1}{A} \frac{dp_x}{dt} = -\eta \frac{\partial u_x}{\partial y}, \quad (59)$$

where A is the contact area between the two layers and η is the shear viscosity. Within kinetic theory, $\eta \simeq \ell p v \sim (\rho/n)(v/\sigma)$ where ρ is the *mass density* in the fluid and the second estimate follows after using Eq. (55). For a non-relativistic fluid $\rho/n = m$, so $\eta \sim m v / \sigma$, whereas for the weakly-coupled QGP, $\rho = \varepsilon \simeq 3nT$ and $v = 1$ and therefore

$$\eta \sim \frac{T}{\sigma} \sim \frac{T^3}{\alpha_s^2 \ln(1/\alpha_s)}. \quad (60)$$

In both cases, the fluid density has canceled in the ratio ρ/n , so the viscosity is *independent of the density*, or, equivalently, of the *pressure*. (Recall that $P = n k_B T$, with k_B the Boltzmann constant.) This remarkable conclusion has been first derived by Maxwell in 1860 via kinetic theory, and then confirmed by him experimentally.

The l.h.s. of Eq. (59) represents a contribution to the component T_{xy} of the energy-momentum (or ‘shear’) tensor : the flux of the x component of the momentum vector across a surface with constant y .

So, Eq. (59) displays a dissipative correction to $T_{\mu\nu}$; as expected, this is of linear order in the gradient expansion and it scales like ℓ/R (since $\eta \sim \ell$ and $\partial_y \sim 1/R$). While first-order gradient corrections (leading to the Navier–Stokes equation) are sufficient to describe dissipation for a non-relativistic fluid, this is not true anymore for a *relativistic* one: to be consistent with causality and Lorentz invariance, one must use a *second-order formalism*, which also includes *quadratic* terms in the gradient expansion.

To summarize, for the problem of hydrodynamics to be well defined, one needs to specify (i) the equation of state $\varepsilon(P)$ (this is generally taken from lattice QCD calculations; see below), (ii) the time τ_0 at which hydrodynamical evolution can be turned on (meaning that local thermal equilibrium has been reached), (iii) the initial conditions at τ_0 for the energy density $\varepsilon(\mathbf{x})$ and the velocity $\mathbf{v}(\mathbf{x})$ fields, and (iv) the various viscosities like η which characterize the dissipative properties of the medium. Note that, in this context, the ‘initial time’ τ_0 is not the same as the time $\tau_s \sim 1/Q_s$ at which the CGC formalism provides the ‘initial conditions’ (in the sense of the discussion in Section 3.7), but it is the *a priori* larger *equilibration time* τ_{eq} . Within the hydro simulations, this is a free parameter, like the viscosities or the parameters which enter the equation of state. These parameters are fixed *a posteriori*, by matching the results of the hydro evolution at the time of freeze-out onto some of the experimental results for particle production, like the centrality dependence of the particle multiplicities and of their average transverse momentum.

For quite some time, roughly until 2007, it seemed that the RHIC data can be well accounted for (within the error bars) by ideal hydrodynamics [85]. This led to the conjecture that the deconfined matter produced at the intermediate stages of a HIC might be *strongly interacting* (‘strongly coupled quark-gluon plasma’ or sQGP). In order to test this conjecture, and also to describe the more accurate, recent data, it became necessary to include dissipative effects, within the second-order formalism. Full calculations in that sense, including comparison with RHIC data, became available only recently [86–88] (and refs. therein). They are all consistent with a non-zero, albeit small, *relative* value of the viscosity, as measured by the ratio η/s . Here, s is the entropy density, and the ratio η/s is dimensionless in natural units (in general, it has the dimension of $\hbar$). This ratio is a natural measure of the deviations from ideal hydro, as we explain now. The entropy density s is proportional to the particle density; e.g., $s = 4n$ for an ideal gas of massless particles. Thus, $\eta/s \sim \ell v(\rho/n) \sim \ell/\lambda$, where λ is the Compton wavelength of a particle in the fluid: $\lambda = 1/(mv)$ in the non-relativistic case and $\lambda \sim 1/T$ for a weakly coupled QGP. By the uncertainty principle, the ratio ℓ/λ cannot be smaller than $\hbar$ times a number of $\mathcal{O}(1)$. So, the ratio η/s cannot become arbitrary small, even when increasing the coupling. In that sense, a physical fluid can never be ideal.

An additional argument in that sense comes from the study of a strongly-coupled theory via the AdS/CFT correspondence. At least for the more symmetric, conformal, field theories to which it applies, this formalism predicts a lower bound on the ratio η/s , namely [89, 90] (the subscript ‘CFT’ refers to a conformal field theory; see Section 4.5 for details)

$$\left. \frac{\eta}{s} \right|_{\text{CFT}} \geq \frac{\hbar}{4\pi}, \quad (61)$$

with the lower bound being reached in the limit of an infinitely strong coupling (in a sense to be characterized in Section 4.5). One remarkable thing about the heavy-ion data at RHIC and the LHC is that they seem to require a value η/s almost as small as this absolute lower bound: $\eta/s \simeq 0.08 \div 0.20$ depending upon the details of the ‘initial conditions’ at time τ_0 . (For instance, the analysis in Ref. [86] favors a value $\eta/s \simeq 0.16$ for initial conditions of the CGC type, as shown in the left panel of Fig. 32.) The other remarkable thing is that, in order to be successful, the hydro descriptions of the data must assume a very small equilibration time $\tau_0 \lesssim 1 \text{ GeV}/c$. These are both hallmarks of a system with *strong interactions*. Indeed, the particles thermalize by exchanging energy and momentum (and other quantum numbers) with each other, via their mutual collisions. So, we expect the thermalization time to be shorter for strongly interacting systems. This expectation is supported by kinetic theory, which yields a thermalization time $\tau_{\text{eq}} \simeq \ell/v \propto [T\alpha_s^2 \ln(1/\alpha_s)]^{-1}$ to leading-order at weak coupling. For realistic values of

α_s , this perturbative estimate is too large to be consistent with the data (even when corrected for inelastic processes like $2 \rightarrow 3$, which turn out to be important [91]).

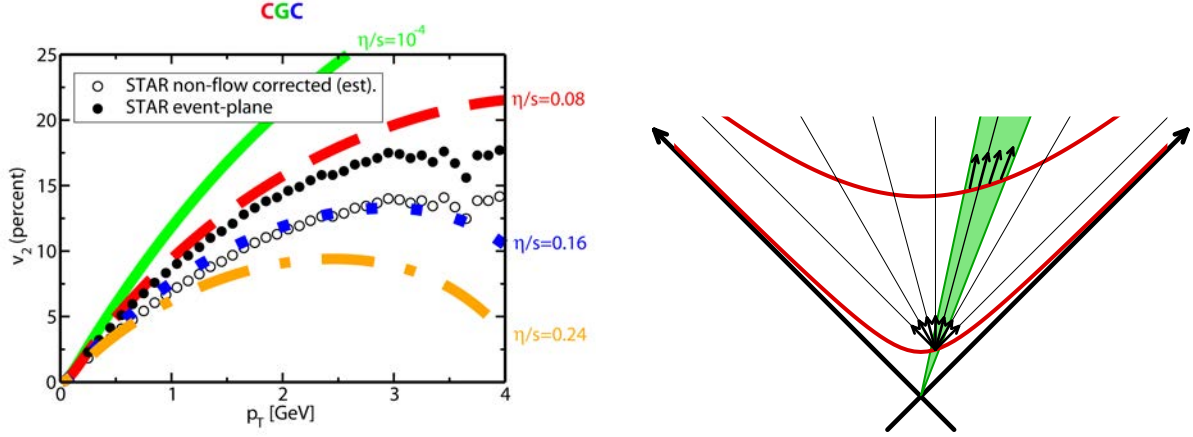


Fig. 32: Left panel: the v_2 results of hydro calculations using the second order formalism with CGC initial conditions and various values of η/s [86]. The comparison with the RHIC data favors $\eta/s \simeq 0.16$, which is about twice the lower limit (61) predicted by AdS/CFT at infinitely strong coupling. Right label: the longitudinal expansion has the tendency to collimate nearby particles, thus opposing to the evolution towards isotropy.

In the context of heavy ion collisions, the evolution towards (local) thermal equilibrium is furthermore hindered by the extreme anisotropy of the initial conditions and also by the anisotropy of the early-time expansion, which is predominantly longitudinal. Recall the glasma picture of the initial conditions (at times $\tau_s \sim 1/Q_s$), which is that of colour flux tubes extending along the collision x , cf. Fig. 25. Flux tubes have an internal tension opposing to their longitudinal extension, like a string. Accordingly, the longitudinal component of the energy-momentum tensor associated with the glasma is *negative*; one finds $T_{\text{glasma}}^{\mu\nu} = \text{diag}(\varepsilon, \varepsilon, \varepsilon, -\varepsilon)$ at such early times, which is extremely anisotropic, as anticipated. Subsequent interactions among particles are supposed to restore isotropy, but this is rendered difficult by the *longitudinal expansion*, as illustrated in the right panel of Fig. 32. Namely, even if the particle distribution turns out to be locally isotropic at a given position in space and time, the subsequent anisotropic expansion rapidly separates the particles from each other according to the directions of their velocities: only those particles remain close to each other which had nearly parallel velocities. In other terms, by itself, the longitudinal expansion would naturally build a particle distribution in which nearby particles move along quasi-parallel directions, thus opposing isotropy. To beat this tendency and ensure isotropy, one needs strong interactions which continuously randomize the directions of motion of the particles. This would be natural at strong coupling, as alluded to above. But there are also other scenarios which are currently explored, including weak-coupling ones. One promising mechanism in that sense refers to *plasma (Weibel) instabilities*: due to the anisotropy of the expanding parton distribution, the soft colour fields radiated by these partons can develop unstable modes, that is, modes whose amplitudes grow exponentially with time (at least, during a limited time interval). So far, it is not clear whether this mechanism can lead to rapid isotropisation in the presence of longitudinal expansion, but its studies are under way (see Refs. [92–97] for recent work and related references). Recent developments include a calculation (similar to previous work in inflationary dynamics) of the spectrum of initial quantum fluctuations in the glasma [94], a parametric analysis of the interplay between plasma instabilities and Bjorken expansion in the weak-coupling limit [95, 96], and an interesting scenario (still at weak coupling) in which the elastic scattering between the highly occupied glasma fields leads to the formation of a transient Bose–Einstein condensate [98].

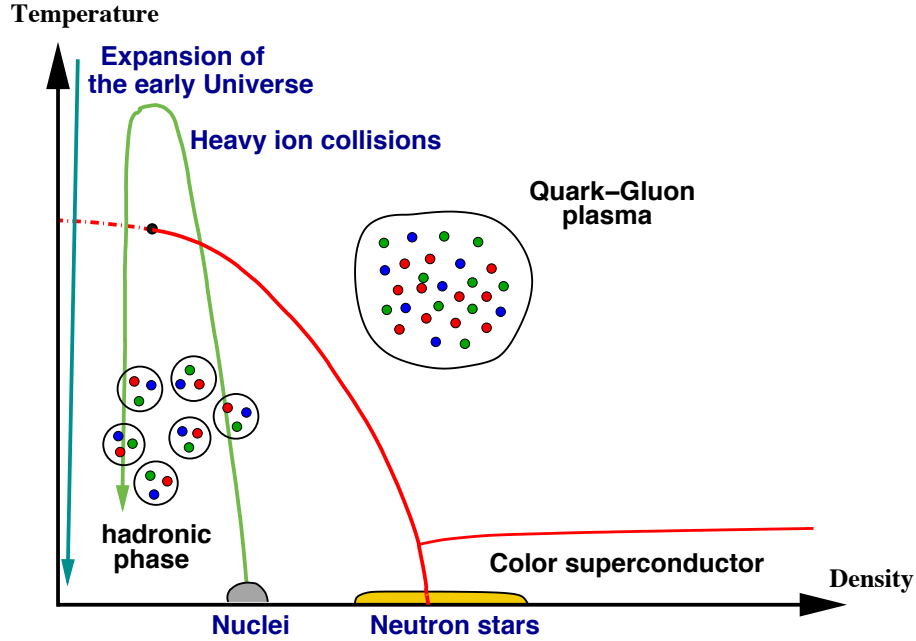


Fig. 33: Schematic representation of the phase-diagram in QCD at finite temperature and non-zero quark density, as emerging from lattice calculations at zero (or small) quark density and from various theoretical considerations (like pQCD) in the other domains.

4.3 QGP: Thermodynamics and collective excitations

In this section, we shall deviate from the experimental situation in HIC's, where the partonic medium is rapidly expanding, and focus on a quark-gluon plasma at rest, in thermal and chemical equilibrium. The existence of such a deconfined phase in QCD at finite temperature has been unambiguously demonstrated via numerical calculations on a lattice, which have also given a lot of information about the thermodynamics of this system. Some of this information has been corroborated via analytic calculations at weak coupling, which turned out to be very non-trivial. The analytic methods become essential when one is interested in *real-time* phenomena, like the response of the system to time-dependent external perturbations, as characterized by *transport coefficients*. Indeed, real-time phenomena cannot be (easily) studied via lattice calculations⁶, which are *a priori* formulated in a space-time with Euclidean signature ('imaginary time'). It should be also stressed that, even though weak-coupling techniques appear to be quite successful in reproducing the lattice results for the QGP thermodynamics, the hypothesis that the coupling be *strong* is not yet totally excluded (within the temperature range relevant for the phenomenology at RHIC and the LHC): indeed, weak-coupling calculations seem unable to explain the small η/s ratio supported by the data (cf. Section 4.2). In what follows, all that will be explained in some detail.

Figure 33 shows a cartoon of the phase-diagram expected in QCD when varying the temperature T and the net quark density (or the quark chemical potentials μ_f), by which one means the difference between the density of quarks and that of the antiquarks. (For simplicity, Fig. 33 treats the three light quark flavors — the only ones to be relevant for the phase diagram — on the same footing.) This diagram has been actually *demonstrated* only in special corners, like the *deconfinement phase transition* with increasing T at zero (or small) fermionic density, that has been established on the lattice, and the islands denoted as 'nuclei' or 'neutron stars', which are rather well understood within nuclear theory. The 'colour superconductivity' phase at high quark density, which is predicted by pQCD (at least for $\mu \gg \Lambda_{\text{QCD}}$), will not be discussed here, as it is not expected to play any role in the ultrarelativistic

⁶There is some recent progress in computing transport coefficients on the lattice (see e.g. the review paper [99] and references therein), but although promising, this method is still inaccurate and very fastidious.

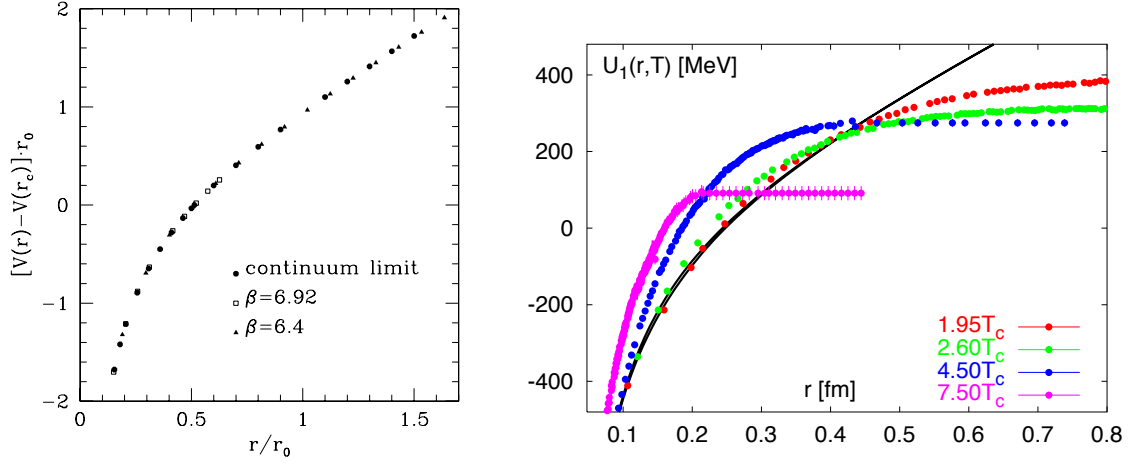


Fig. 34: The quark-antiquark potential between two heavy quarks, as computed in QCD on the lattice. Left panel: $T = 0$. Right panel: various temperatures, which are all larger than the critical temperature for deconfinement T_c (for comparison, the $T = 0$ potential is also shown, as the continuous line which keeps rising) [103].

HIC's. (See the review papers [100, 101] and references therein for detailed discussions of this phase.)

As also illustrated in Fig. 33, the deconfinement phase transition has been first explored during the expansion of the Early Universe: the high temperature ‘soup’ of matter created right after the Big Bang was originally in the deconfined, QGP, phase; due to its rapid expansion, this matter has cooled down and thus crossed into the confined, hadronic, phase, at a very short time $\sim 10^{-5}$ seconds after the Big Bang. In the context of HIC's, this transition is being probed the other way around: to start with, the partons are confined within the nucleons composing the two nuclei; the collision liberates these partons and, if their energy density is high enough, they can thermalize at a temperature superior to the critical temperature for deconfinement. If so, they form a transient QGP phase which cools down via expansion and eventually ‘evaporates’ into hadrons. In both scenarios, the net quark density is small and plays no role for the transition. In the Early Universe, the excess in the number of quarks over antiquarks was negligible (if any !) [102]. In HIC's, there is of course a net baryon number, due to the $2A \simeq 400$ nucleons within the incoming nuclei; however, this excess is small compared to the number of hadrons (a few thousand) produced in the final state. This implies that most of the partons which exist in the intermediate stages of the collision are actually gluons or ‘sea’ quark-antiquark pairs.

The fundamental property of the QGP is, of course, *deconfinement* : quarks and gluons can move (more or less freely, depending upon their mutual interactions) throughout the whole volume of the plasma, without being confined within hadrons with radii $\sim 1/\Lambda_{\text{QCD}}$. How is this possible ? A quark and an antiquark in isolation (i.e. at zero temperature) attract each other via a force which becomes roughly constant — corresponding to a linear potential; see Fig. 34 left — at distances $r \gtrsim 1/\Lambda_{\text{QCD}}$. Due to this force, the $q\bar{q}$ pair is tightly bound (‘confined’) into a meson. This meson can be broken, say, via a hard scattering, but only at the expense of producing additional gluons and $q\bar{q}$ pairs which ‘glue’ to the original quark and antiquark, in such a way to form colour singlet states (new hadrons). This is the situation in the ‘usual’ hadronic processes, including p+p collisions at the LHC, where the parton density right after the collision is not very high — so, these partons can evolve and eventually hadronize independently of each other. But in HIC's, the density of the liberated partons is such that the typical interparticle separation is much shorter than $1/\Lambda_{\text{QCD}}$. At such short distances, the attraction force between these partons is smaller than their kinetic energy, so the partons can move around each other and arrange themselves in such a way to minimize their mutual repulsion. The net result is that the colour charge gets *screened* over relatively short distances $r \ll 1/\Lambda_{\text{QCD}}$, thus preventing the development of confining forces at larger distances.

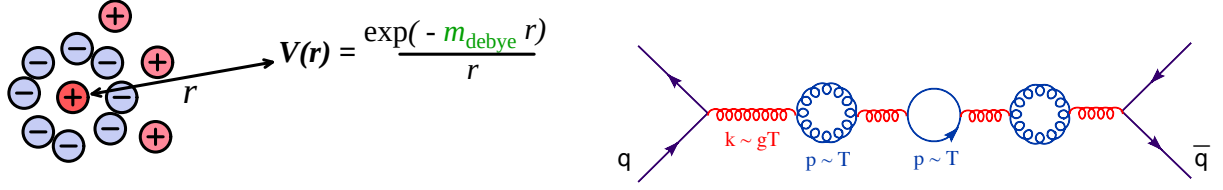


Fig. 35: Left panel: illustration of the Debye screening in a QED plasma; the screening length is the inverse of the Debye ‘mass’: $r_D = 1/m_D$. A similar mechanism is active in the QGP. Right panel: the Debye screening in QCD is the result of one-loop corrections due to the ‘hard’ ($k \sim T$) thermal particles (quark and gluons) which are resummed in the propagator of the ‘soft’ ($k \sim gT$) electric gluon exchanged between the quark and the antiquark.

This colour screening in the QCD plasma is very similar to electric (Debye) screening in ordinary, electromagnetic, plasmas, or in electrolytes. Ions with positive electric charge attract ions or electrons with negative charge, in such a way to form clouds of particles which look electrically neutral when seen from far away: the net charge decreases exponentially with the distance from the central charge (see Fig. 35 left). In the context of QCD, the ‘positive and negative electric charges’ are replaced by the $N_c^2 - 1 = 8$ ‘colour’ charges carried by quarks and gluons, but the exponential screening of the chromo-electric charges works in a similar way. As a consequence of that, the quark-antiquark potential flattens out (meaning that there is no attraction force) at large distances $r \gtrsim r_D$ with r_D the Debye radius. This flattening is clearly seen in the lattice calculations at finite temperature (see e.g. the discussion in Ref. [103]), as illustrated in the right panel of Fig. 34. This also suggests that in a finite temperature plasma one cannot have quarkonia (bound states made with a heavy quark (Q) and a heavy antiquark ($\bar{Q}$), like J/ψ , with size $r_{Q\bar{Q}} > r_D$. This observation [104] led to the fertile idea of *quarkonia melting* in a quark-gluon plasma, a very active field of research for both theoretical (including lattice) and experimental studies of HIC’s. (See Ref. [105] for a recent overview of the theory and more references.)

The lattice calculations also allow one to study the deconfinement phase transition with increasing temperature. The respective results for the pressure and energy density are illustrated in Fig. 36. They show a sudden increase around a *critical temperature* $T_c \simeq 160 \div 180$ MeV, interpreted as the result of the rise in the number of degrees of freedom (d.o.f.), due to the liberation of quarks and gluons. For $T \leq T_c$, the only ‘thermodynamically active’, hadronic d.o.f. (those whose masses are not much higher than T) are the 3 pions: π^0 and $\pi^\pm$. For $T > T_c$, this number jumps from 3 to 52: the gluons, which appear in 8 colours and 2 transverse polarizations ($8 \times 2 = 16$ d.o.f.), and the 3 light quarks and antiquarks, each of them having 2 spin states and 3 possible colours ($3 \times 3 \times 2 \times 2 = 36$ d.o.f.). Lattice calculations become more tedious for light quark masses and the extrapolation to physical quark masses has become possible only recently [106, 107]. This is important since both the actual value of the critical temperature and the nature of the phase transition are strongly influenced by the values of these masses. A phase transition is said to be of n th order if it involves a discontinuity in the derivative of order $(n - 1)$ of the pressure. For instance, if the QCD phase transition was of *first-order*, then it would proceed via a mixed phase where hadronic bubbles coexist with regions of QGP. But this is not what happens in QCD: recent lattice calculations [108] show that, for physical quark masses, the deconfinement phase transition is truly a *cross-over*, that is, a relatively smooth process during which the pressure and all its derivatives remain continuous across the transition.

But albeit smooth, the phase transition represents a genuinely non-perturbative phenomenon, which cannot be described within perturbative QCD. This can be appreciated e.g. by inspection of the lattice results for the *trace anomaly* $(\varepsilon - 3P)/T^4$, which exhibit a sharp peak around T_c , as visible in the left panel of Fig. 37. As mentioned in Section 4.2, $\varepsilon = 3P$ for an ideal gas of massless particles, meaning that the corresponding energy-momentum tensor $T^{\mu\nu} = \text{diag}(\varepsilon, P, P, P)$ is traceless: $T^\mu_\mu = 0$. This property is in fact a general consequence of *conformal symmetry*: it holds for any theory which

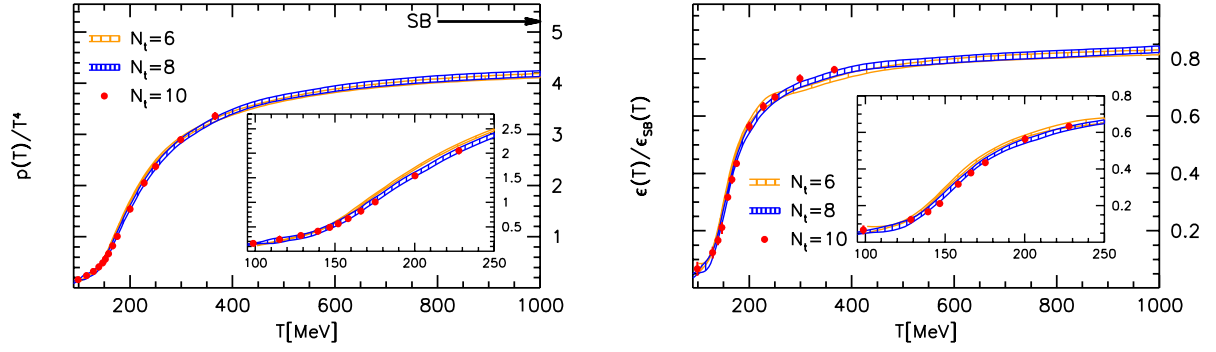


Fig. 36: The lattice results (as obtained by the Budapest–Wuppertal collaboration [109]) for the QCD pressure and energy density as a function of the temperature.

involves no intrinsic mass parameter and hence is invariant under dilations. This is in particular the case for QCD with massless quarks at the *classical* level. However, at the *quantum* level, conformal symmetry in QCD is broken by the radiative corrections responsible for the running of the coupling, which introduce the mass scale Λ_{QCD} . So, not surprisingly, the corresponding ‘trace anomaly’ (the deviation of T^μ_μ from zero) is proportional to the β -function, which measures the running of the coupling :

$$T^\mu_\mu = \varepsilon - 3P = \beta(g) \frac{\partial P}{\partial g}, \quad \beta(g) \equiv \frac{\partial g}{\partial \ln \mu_{\text{ren}}}. \quad (62)$$

Here μ_{ren} is the renormalization scale, as introduced by the subtraction of the ultraviolet divergences. In perturbation theory at finite temperature, it is convenient to choose $\mu_{\text{ren}} = 2\pi T$ as the central value and study the dependence of the results upon variations in μ_{ren} (typically by a factor of 2) around this central value. These variations measure the stability of the calculations against higher order corrections and thus are indicative of the theoretical uncertainties. They are shown as ‘error bands’ for the theoretical results in Figs. 37 and 39.

The peak in the l.h.s. of Fig. 37 is a hallmark of the phase transition and is clearly non-perturbative. But for temperatures above T_c , the ‘trace anomaly’ is rapidly decreasing (its relative strength becomes of order 10% for $T \gtrsim 3T_c$), thus suggesting that a perturbative approach may become viable. This is furthermore indicated by the fact that, for $T \gtrsim 3T_c$, the pressure and the energy density reach about 80% of the respective values for an ideal gas of quarks and gluons, denoted as ‘SB’ (from ‘Stefan–Boltzmann’) in Fig. 36. A deviation of 20% may seem sufficiently small to be easily accommodated in perturbation theory, but this turns out not to be the case. There are two main reasons for that. First, unlike what happens at $T = 0$, where perturbation theory in QCD is an expansion in powers of $\alpha_s \equiv g^2/4\pi$, at finite temperature this is rather an expansion in powers of g , for reasons to be shortly explained. Second, the relevant values of the QCD running coupling are not that small: for $T \simeq 3T_c$ and hence $2\pi T \simeq 2 \text{ GeV}$, one has $\alpha_s \simeq 0.25$ and hence $g = 1.5 \div 2$. For such large value of g , there is no reason why an expansion in powers of g should converge, and indeed it does not: as visible in the right panel of Fig. 37, the successive corrections of $\mathcal{O}(g^2)$, $\mathcal{O}(g^3)$, $\mathcal{O}(g^4)$, and $\mathcal{O}(g^5)$ jump up and down, without any sign of convergence. (The weak-coupling expansion of the pressure in QCD is known to $\mathcal{O}(g^6 \ln(1/g))$ [111], which is the highest order that can be computed in perturbation theory: the corrections of $\mathcal{O}(g^6)$ and higher are afflicted with severe infrared divergences due to magnetic gluons; see below.)

The reason why, at finite temperature, perturbation theory is an expansion in powers of g rather than α_s is because the quantum corrections associated with *soft gluons* — those with momenta k much smaller than T — are amplified by the Bose–Einstein thermal distribution function:

$$n_B(k) = \frac{1}{e^{\beta E_k} - 1} \simeq \frac{T}{E_k} \gg 1 \quad \text{when} \quad E_k = |\mathbf{k}| \ll T. \quad (63)$$

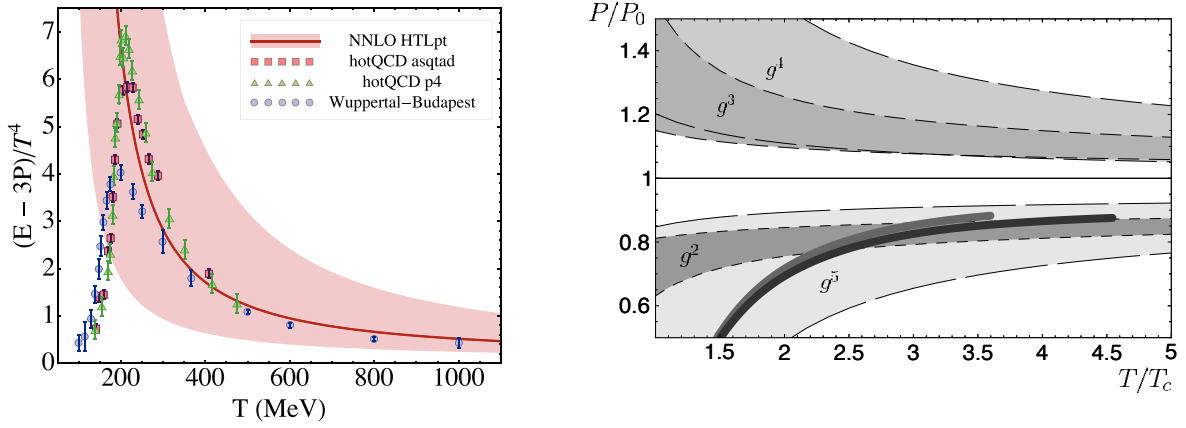


Fig. 37: Left panel: the trace anomaly in QCD, as numerically computed on the lattice [106, 107], together with the respective predictions of a ‘HTL-resummed’ perturbation theory [110] (see the text for details), shown as a band (due to the uncertainty in the choice of the renormalization scale μ_{ren}). Right panel: the predictions of (strict) perturbation theory for the pressure, as computed up to order g^2 , g^3 , g^4 , and g^5 , respectively. Whereas the $\mathcal{O}(g^2)$ -result appears to match the lattice results (the dark grey band) rather closely, this agreement is spoiled after including higher order corrections; the ensuing series in powers of g shows no sign of convergence.

This property is generic: it holds for any field theory which involves massless bosons (e.g., it holds for photons in a QED plasma). When $k \rightarrow 0$, the thermal factors $n_B(k) \simeq T/k$ lead to infrared divergences in the calculation of Feynman graphs, which are regulated by plasma effects, like Debye screening. These effects typically enter at the ‘soft’ scale⁷ gT . For instance, the *Debye mass* $m_D \equiv 1/r_D$ which characterizes the exponential screening of the electric colour charge by the plasma constituents, is generated by the one-loop diagrams illustrated in the right panel of Fig. 35, which yield $m_D \sim gT$. The resummation of these diagrams within the propagator of the exchanged gluon, as also illustrated in Fig. 35, renders the (electric) gluons effectively massive⁸: $E_k = \sqrt{k^2 + m_D^2}$. Hence, when $k \rightarrow 0$, the Bose–Einstein occupation number remains finite, but it is parametrically large: $n_B(k) \simeq T/m_D \sim 1/g$. This inverse power of g changes the perturbative order of the 2-loop correction to the pressure with one ‘hard’ loop ($k \sim T$) and one ‘soft’ ($k \lesssim gT$) from $\alpha_s^2 \sim g^4$ to $g^4 n_B(k) \sim g^3$. This is the origin of the odd powers of g in the perturbative expansion.

One should also mention here that Debye screening, as illustrated in the left panel of Fig. 35, is operational for the *electric* gluons (i.e. for the Coulomb interactions), but not also for the *magnetic* ones — those having transverse polarizations. For non-relativistic plasmas, magnetic interactions are suppressed by powers of the velocities, but for (ultra)relativistic plasmas, like the QGP, they are as important as the electric ones. One expects magnetic interactions in the QGP to be screened at the ‘ultrasoft’ scale $m_{\text{mag}} \sim g^2 T$, but the associated physics — in particular, the contribution of the ‘ultrasoft’ magnetic gluons to thermodynamics, which starts at $\mathcal{O}(g^6)$ — cannot be computed in perturbation theory. Indeed, each additional ‘ultrasoft’ loop is accompanied by a factor $\sim g^2 n_B(k) \sim g^2 (T/m_{\text{mag}}) \sim 1$, meaning that diagrams with arbitrarily many such loops contribute at the same order in g . This explains why the corrections of $\mathcal{O}(g^6)$ are non-perturbative, as alluded to above.

In QCD, the Debye mass is only one example of a class of one-loop ‘corrections’ which are non-

⁷This scale gT is truly ‘soft’ only so long as $g \ll 1$; as we shall shortly argue, results obtained under the assumption that the coupling is weak can be extrapolated towards $g \sim 1$ provided the plasma effects are properly taken into account.

⁸There is strictly speaking a difference between the *Debye mass* m_D , which governs the infrared ($k \rightarrow 0$) limit of the *static* ($k_0 = 0$) propagator for the electric gluons, and the *thermal mass* m_g , which enters the dispersion relation for the *on-shell* gluons; but these quantities are proportional with each other and are both of order gT ; see e.g. Refs. [112–114].

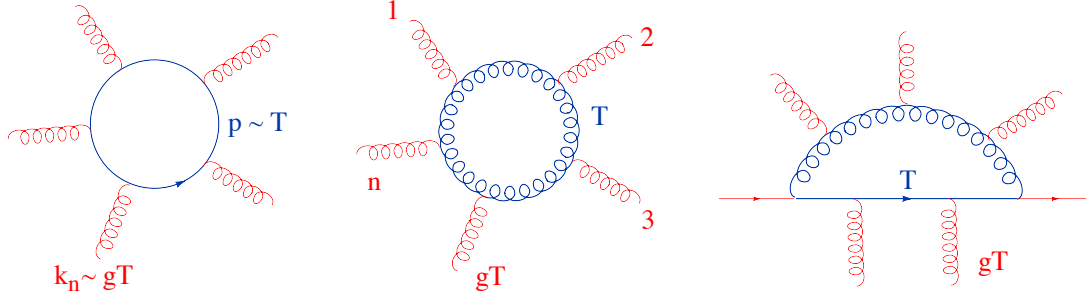


Fig. 38: A few examples of hard thermal loops (HTL's). All the external momenta are soft ($k_i \sim gT$) and the loop integrations are dominated by relatively hard modes with $p \sim T$.

perturbative at the ‘soft’ scale gT and should be viewed as a part of the *leading-order theory* at that scale, and not as corrections [112, 115]. These diagrams are known as *hard thermal loops*, since the typical momenta within the loop are of order T (the value preferred by the statistical, Bose–Einstein and Fermi–Dirac factors) and thus are hard compared to the soft ($k \sim gT$) momenta flowing along the external legs. There are HTL’s with any number n of external gluons lines and with either zero, or two, quark external lines (see Fig. 38 for some examples). They are generally *non-local*, that is, they depend upon the external, soft, momenta. In particular, the HTL’s for the 2-point functions (the quark and gluon self-energies) encode phenomena like Debye screening for the electric gluons, dynamical screening (or Landau damping) for the magnetic gluons, and the dispersion relations for on-shell quanta with momenta of order gT . Such quanta have wavelengths $\lambda \sim 1/gT$ which are parametrically larger than the typical separation $\sim 1/T$ between the typical plasma constituents — quarks and gluons with momenta of order T . Accordingly, the soft modes are truly *collective excitations* (or ‘plasma waves’), with either quark or gluon quantum numbers.

As already mentioned, the HTL’s are of the same order as the respective tree-level amplitudes with ‘soft’ external legs, so they cannot be expanded out in perturbation theory. Rather, one needs to perform a *reorganization* of the perturbation theory in which the HTL’s are viewed as a part of the leading-order theory for the soft modes. Roughly speaking, this amounts to expanding around a gas of *dressed quasi-particles* whose zeroth-order properties (propagators and interaction vertices) are encoded in the HTL’s. In practice, there are various ways to perform such reorganizations and it is quite reassuring that all the methods that have been proposed so far [110, 111, 113, 114, 116, 117] appear to be successful in describing the lattice data (although with considerably different amounts of efforts).

One of these methods, known as ‘HTL perturbation theory’ (HTLpt) [117], consists in including the HTL’s in the ‘tree-level’ effective theory, by adding and subtracting $\mathcal{L}_{\text{HTL}}$ (the sum of the HTL amplitudes) to/from the original Lagrangian:

$$\mathcal{L}_{\text{QCD}} = \mathcal{L}_0 + \mathcal{L}_{\text{int}} = (\mathcal{L}_0 + \mathcal{L}_{\text{HTL}}) + (-\mathcal{L}_{\text{HTL}} + \mathcal{L}_{\text{int}}) = \mathcal{L}'_0 + \mathcal{L}'_{\text{int}}. \quad (64)$$

In this equation, $\mathcal{L}_0$ is the free ($g = 0$) piece of the QCD Lagrangian, $\mathcal{L}_{\text{int}}$ is the respective interaction piece, $\mathcal{L}'_0 \equiv \mathcal{L}_0 + \mathcal{L}_{\text{HTL}}$ represents the new ‘tree-level Lagrangian’ which defines the Feynman rules (HTL-resummed propagators and vertices) of HTLpt, and, finally, $\mathcal{L}'_{\text{int}} \equiv \mathcal{L}_{\text{int}} - \mathcal{L}_{\text{HTL}}$ is the new ‘interaction Lagrangian’. The subtracted piece $-\mathcal{L}_{\text{HTL}}$ within $\mathcal{L}'_{\text{int}}$ acts as a ‘counterterm’ to prevent double counting (the HTL’s have been already included in the zeroth-order theory and they should not be regenerated via loop corrections in HTLpt) and also to correct for the fact that, within $\mathcal{L}'_0$, the HTL’s are used for *all* the modes, including the hard modes to which they do not really apply (this introduces spurious contributions at lower orders which are compensated by the ‘counterterm’ only in higher orders). In order for such compensations to efficiently work, one needs to go up to relatively high orders in HTLpt, which involve very tedious calculations (due to the non-local nature of the HTL’s). It was only recently, after pushing such calculations up to three loop order [110], that one has finally reached a good

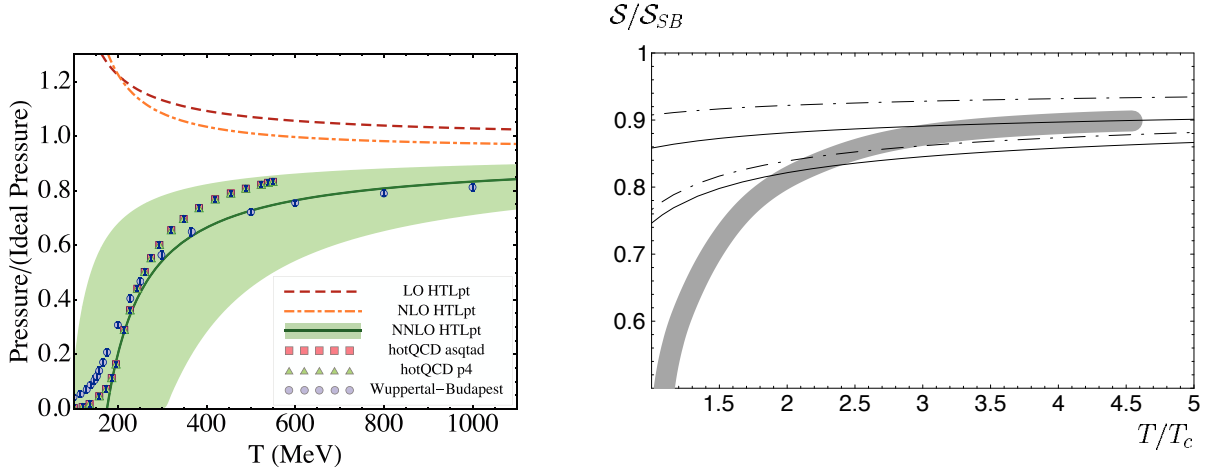


Fig. 39: Comparison between the predictions of two versions of HTL-resummed perturbation theory for thermodynamics and the respective lattice results. Left panel: the predictions of HTLpt [117] for the pressure at one-loop (LO), 2-loop (NLO) and, respectively, 3-loop (NNLO) order [110] vs. the lattice results (small cercles, triangles or squares) from two different collaborations. Right panel: the 2-loop result of the ‘2-particle irreducible resummation’ of the entropy [116] — solid and dotted lines correspond to two successive approximations for the thermal masses — vs. the lattice results shown as the grey band. In both cases, a good agreement with the lattice results is observed for temperatures $T \gtrsim 2.5T_c$. The theoretical ‘error bands’ follows from varying the renormalization scale in the range $\pi T \leq \mu_{\text{ren}} \leq 4\pi T$.

agreement with the respective lattice results for $T \gtrsim 2.5T_c$. This agreement is visible in the left panel of Fig. 37 for the trace anomaly and in the left panel of Fig. 39 for the pressure.

A more economical approach, in which a similarly good agreement with the lattice results (see the right panel of Fig. 39) has been obtained via a simpler, 2-loop, calculation, is the ‘2-particle irreducible (2PI) resummation’ of the entropy [113, 116]. In that approach, most of the difference between the 2PI result for S and the corresponding result S_{SB} for the ideal gas comes from the thermal masses $m_q, m_g \sim gT$ acquired by the *hard* ($k \sim T$) quarks and gluons via interactions in the plasma. Albeit formally small ($m_{q,g} \sim gT \ll k \sim T$), these masses cannot be expanded in perturbation theory, since such an expansion would generate powers of m^2/k^2 leading to infrared divergences in the integral over k . The success of the 2PI description supports the physical picture of the QGP in terms of *quasi-particles* — quarks and gluons with typical momenta $k \sim T$, which are dressed by the medium (in particular, in the sense of acquiring thermal masses), but whose residual interactions are relatively small. Moreover, a substantial part of these residual interactions can be associated with collective excitations and screening effects at the ‘soft’ scale gT , as encoded in the HTL-resummed propagators.

It is furthermore interesting to notice that the HTL resummation is based on the separation of scales $gT \ll T$ which is *a priori* valid at weak coupling ($g \ll 1$), yet this turns out to rather successfully describe the lattice data in a temperature range where $g = 1.5 \div 2$. This confirms that the failure of ordinary perturbation theory (cf. Fig. 37, right panel) is not imputable to the fact that the coupling is relatively strong, but rather it is a consequence of expanding out the medium effects in powers of g in a kinematical domain where they are truly non-perturbative.

Yet, the issue of the strength of the coupling remains open in so far as the study of *dynamical phenomena* is concerned. These phenomena refer to non-trivial evolutions in time, say off-equilibrium deviations in response to external perturbations. As long as the perturbations are small, their effects can be computed within the *linear response theory*, via the Kubo formula: the response of the plasma is linear in the strength of the perturbation, with a proportionality, or ‘transport’, coefficient which represents a

correlation function in thermal equilibrium. For instance, a constant electric field $\mathbf{E}$ acting on the quark constituents of the plasma (which carry electric charge) induces an electromagnetic current with density $\langle j_{\text{em}}^i \rangle = \sigma E^i$, where σ is the *electrical conductivity*. The respective Kubo formula relates σ to the long-wavelength ($k^i \rightarrow 0$) and zero-frequency ($\omega \rightarrow 0$) limit of the current-current correlator in thermal equilibrium. Other transport coefficients include the (quark) flavor diffusion coefficients and the shear viscosity η introduced in Eq. (59). The use of Kubo formulæ for perturbative calculations at weak coupling turns out to be quite tedious, because of the need to perform sophisticated resummations [118]. However, these formulæ are very useful for non-perturbative calculations of the transport coefficients, either on the lattice [99], or via the AdS/CFT correspondence at strong coupling [119].

For a weakly coupled QGP and to leading order in the coupling, the transport coefficient can be alternatively, and more efficiently, computed from the Boltzmann equation (linearized with respect to the off-equilibrium perturbation). This amounts to solving a linear integral equation which effectively resums an infinite number of diagrams of the ordinary perturbation theory in thermal equilibrium. These diagrams describe multiple scattering via soft gluon exchanges and can be generated by iterating the $2 \rightarrow 2$ elastic processes shown in Fig. 31 arbitrarily many times. The ensuing transport coefficients are of the parametric form anticipated (on the example of the shear viscosity) in Eq. (60), but the use of the Boltzmann equation allows one to obtain more precise results, which are complete to leading order in α_s [120, 121]. Yet, these results are deceiving with respect to the heavy-ion phenomenology: as already mentioned in Section 4.2, the leading order estimate for η/s is too large to be consistent with the hydrodynamical description of the data.

The last observation raises the question of the next-to-leading order corrections. Their calculation is extremely complicated and so far this has been accomplished for just one quantity: the diffusion coefficient D for a heavy quark with mass $M \gg T$. In the context of HIC, this quantity controls the collisional energy loss and the thermalization of heavy quarks like the charm or the bottom. Once again, the LO perturbative estimate for D [122] appears to be too large to be consistent with the data. The NLO correction to D is of relative order g and has been computed in Ref. [123]. This appears to go in the right direction (it diminishes the value of D), but the effect is extremely large for realistic values of g — the NLO ‘correction’ is almost an order of magnitude larger than the respective LO result ! —, thus rising doubts about the reliability of the whole scheme. It looks like the perturbative series suffers from a lack-of-convergence problem similar to that noticed for the pressure. It might be that this problem too will be cured by all-orders resummations of the HTL’s; but this issue is still open since such resummations have not yet been performed for *dynamical* quantities. Alternatively, there is the possibility that the transport phenomena, which involve long-range dynamics, be sensitive to rather large values of the QCD running coupling, which exclude weak-coupling techniques. If so, one could search for physical guidance in the corresponding results at strong coupling, as obtained via the AdS/CFT correspondence (see Section 4.5 below). Finally, let us notice that the first lattice results for the transport coefficients have started to emerge, although the current errors bars are still quite large. These calculations are very difficult as they require to numerically perform an analytic continuation (from imaginary time to real time), which in turns requires very precise numerical data. In view of that, it is quite encouraging that the recent lattice results for the heavy quark diffusion coefficient [124, 125] appear to be consistent with the heavy-ion phenomenology, within the (lattice and experimental) errors bars.

4.4 Jet quenching

In Section 3.6 we have mentioned two interesting phenomena occurring in ‘dense-dilute’ (p+A or d+A) collisions — the suppression of particle production and that of azimuthal di-hadron correlations *at forward rapidities* —, which in that context have been interpreted as consequences of gluon saturation in the wavefunction of the nuclear target: the larger the rapidity, the smaller the values of the longitudinal momentum fraction that are probed in the nucleus, and hence the stronger the saturation effects. On the other hand, the RHIC data for d+Au collisions at *central* rapidities ($\eta \leq 1$) show no similar suppression

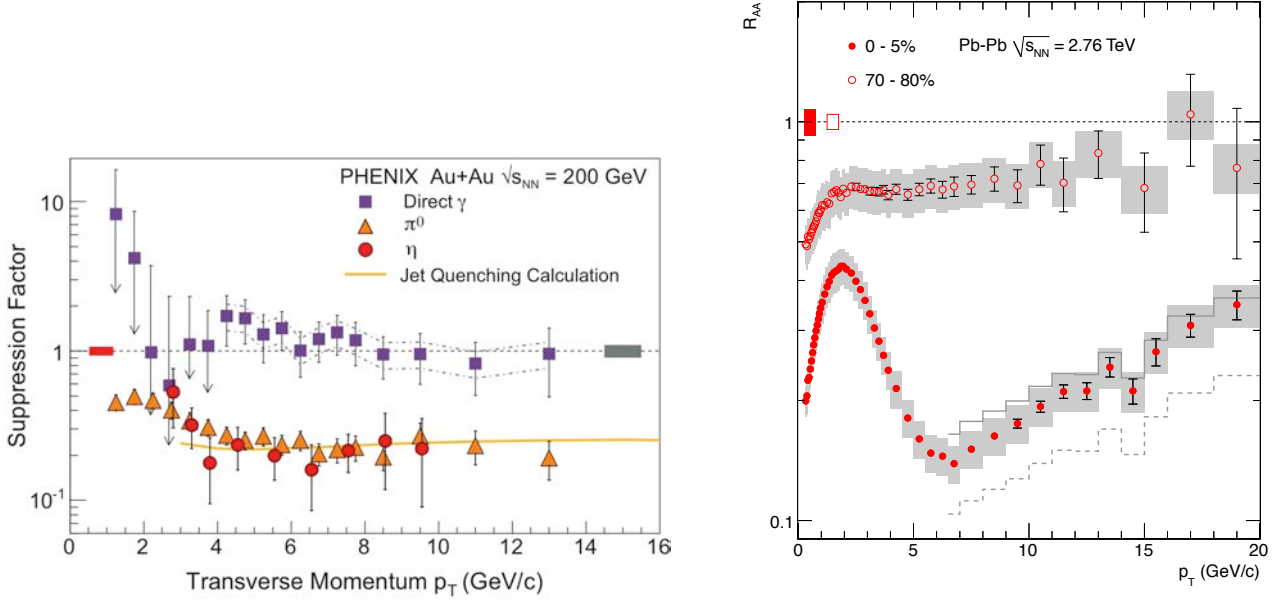


Fig. 40: The nuclear modification factor R_{AA} , as measured in central Au+Au collisions at RHIC (PHENIX, left panel) and in Pb+Pb collisions at the LHC (CMS, right panel). For comparison, one also shows the data for the production of direct photons, which show no nuclear effect as expected (left panel) and for hadron production in peripheral collisions, for which the nuclear effects are quite small (right panel).

(see the 2 left-most plots in Fig. 18 and also the corresponding data in the right plot in Fig. 41 below), which implies that nuclear saturation effects are not important in the central-rapidity kinematics at RHIC. (But this is likely to change at the LHC; see e.g. Refs. [14, 68, 126].) The situation is however different for the ‘dense-dense’ A+A collisions: the respective data at RHIC and the LHC show a strong suppression of particle production and of the azimuthal di-hadron correlations *already for central rapidities* (and for relatively hard transverse momenta). These phenomena cannot be related to ‘initial-state’, saturation, effects in the incoming nuclei (at least not fully), since they do not show up in the mid-rapidity d+Au data at RHIC. Rather, they must correspond to interactions in the *final state*, that is, interactions with the dense partonic medium (the glasma and the quark-gluon plasma) which exists at intermediate stages. The change in the properties of a hard particle or of the associated jet induced by its interactions in the medium is generally referred to as *jet quenching* (see e.g. Refs. [127–130] for recent reviews).

The suppression of particle production in A+A collisions is best characterized by the ratio R_{AA} (the ‘nuclear modification factor’), defined by analogy with Eq. (40), that is

$$R_{AA}(\eta, p_{\perp}) \equiv \frac{1}{N_{coll}} \frac{\left. \frac{dN_h}{d^2p_{\perp} d\eta} \right|_{AA}}{\left. \frac{dN_h}{d^2p_{\perp} d\eta} \right|_{pp}}, \quad (65)$$

where the number N_{coll} of binary collisions at a given impact parameter scales like $A^{4/3}$ (for relatively central collisions): indeed, there is a factor $A^{1/3}$ associated with the longitudinal width of each of the two nuclei and an additional factor of $R_A^2 \propto A^{2/3}$ coming from the integral over all the impact parameters. The experimental results for R_{AA} (at mid-rapidities) in Au+Au collisions at RHIC ($\sqrt{s_{NN}} = 200$ GeV) and in Pb+Pb collisions at the LHC ($\sqrt{s_{NN}} = 2.76$ TeV) are shown in the left and right panels of Fig. 40, respectively. As anticipated, they show a substantial suppression of the hadron production as compared to p+p collisions, which persists up to $p_{\perp} \simeq 20$ GeV, at least. The interpretation of this suppression as a dense-medium effect is furthermore supported by the fact that (i) the direct photons (which do not interact with the hadronic matter) show indeed no suppression (cf. the left figure), and (ii) even

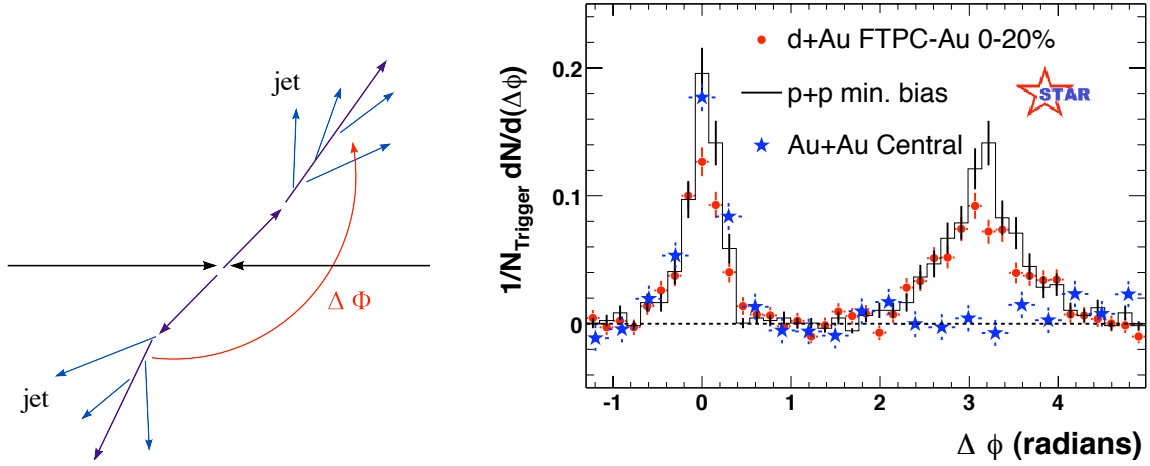


Fig. 41: Left: cartoon of a typical di-jet event as produced by a hard scattering. Right: azimuthal distribution of hadrons with $p_{\perp} \geq 2$ GeV relative to a trigger hadron with $p_{\perp} \geq 4$ GeV, as measured at RHIC (STAR), in p+p, d+Au and central Au+Au collisions [131].

for hadrons, the suppression is considerably smaller in the peripheral collisions (cf. the right figure), in agreement with the fact that the density and the size of the produced medium are much smaller in that setup. Note also that the suppression at intermediate values $p_{\perp} \simeq 6 \div 7$ GeV is stronger at the LHC than at RHIC, indicating that the medium produced there is denser, as expected.

Concerning the suppression of azimuthal correlations in the di-hadron production, this is clearly visible in the right plot in Fig. 41, which shows data taken at RHIC for hadrons with $p_{\perp} \gtrsim 2$ GeV : unlike for p+p and d+Au collisions, where one can see a peak at $\Delta\Phi = \pi$, as expected for a pair of hadrons which are produced back-to-back, there is no such a peak in the central Au+Au collisions. This is interpreted as the consequence of the interactions suffered by the ‘away’ particle (the one that would have normally emerged at $\Delta\Phi = \pi$) while propagating through the medium. Via such interactions, the particle transverse momentum has been degraded and/or the particle has been deviated towards different directions, so it will not show up around $\Delta\Phi = \pi$, nor in the original bin in $p_{\perp}$.

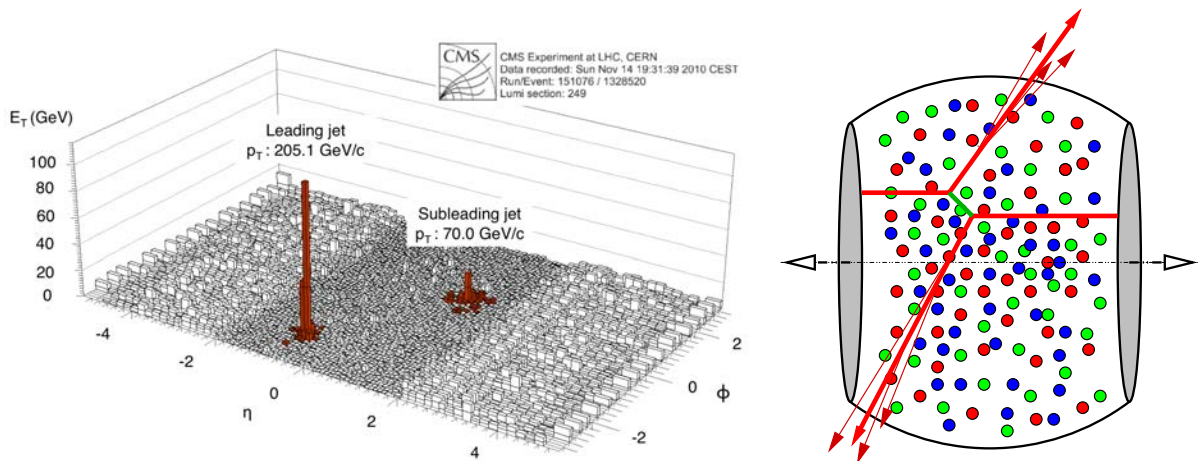


Fig. 42: Left panel: a highly asymmetric di-jet event in Pb+Pb collisions at the LHC as measured by CMS. Right panel: a cartoon of an asymmetric di-jet event in A+A collisions. The hard scattering producing the jets occurs near the edge of the fireball. One of the jets (the ‘trigger jet’) leaves the medium soon after its formation and thus escapes unscattered, while the other one (the ‘away jet’) crosses the medium and is strongly modified by the latter.

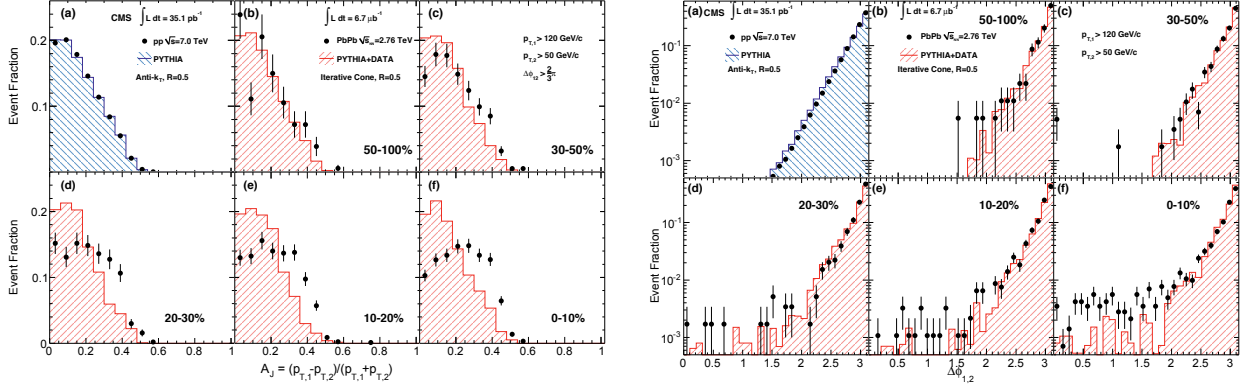


Fig. 43: The distribution of the CMS data for p+p collisions ($\sqrt{s} = 7\text{ GeV}$) and Pb+Pb collisions ($\sqrt{s_{NN}} = 2.76\text{ TeV}$) as a function of A_J (left panel) and respectively of the azimuthal separation $\Delta\Phi_{12}$ w.r.t. the leading-jet axis (right panel) in different bins of centrality: (a) refers to p+p, (f) to the 10% most central Pb+Pb, (e) to the centrality bin 10-20%, etc. The histograms show the respective predictions of PYTHIA, which agree well with the data for p+p and for peripheral Pb+Pb, but not also for central Pb+Pb. From Ref. [133].

Besides confirming and sharpening the discoveries at RHIC, the first heavy ion data at the LHC revealed a new phenomenon, whose observation was possible because of the unprecedented ability of the detectors there (notably the calorimeters at ATLAS and CMS, with a wide coverage in rapidity) to reconstruct jets: the *di-jet asymmetry* [132, 133]. Namely, a significant fraction of the di-jet events in Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76\text{ TeV}$ shows a large transverse energy imbalance between the trigger (or ‘leading’) jet and the away (or ‘subleading’) jet. (The left panel in Fig. 42 and the right panel in Fig. 2 display such asymmetric di-jet events, as measured by CMS.) One should stress that the criterion used to define a ‘jet’ — the value of the product $R = \Delta\Phi \times \Delta\eta$ between the spreadings of the hadron yield in azimuthal angle and in pseudo-rapidity — is the same for the leading and subleading jets. Moreover, a substantial asymmetry between the two jets exists already in p+p collisions, because of the bias introduced by the trigger process (see the histograms in Fig. 43). But the heavy ion collisions show a significant increase in this asymmetry, which becomes more pronounced with increasing centrality. A quantitative way to characterize this asymmetry is via the *transverse energy imbalance*,

$$A_J = \frac{p_{\perp 1} - p_{\perp 2}}{p_{\perp 1} + p_{\perp 2}}, \quad (66)$$

where $p_{\perp 1}$ ($p_{\perp 2}$) is the transverse momentum of the leading (subleading) jet. The normalization in Eq. (66) is useful for removing uncertainties in the overall jet energy scale. Figure 43 (left panel) shows the distribution of the Pb+Pb events as a function of A_J , in different bins of centrality: for the most peripheral collisions, this is quite similar to the respective distribution for p+p collisions, as shown in figure (a). But for the more central collisions, there is an increase in the fraction of events with relatively large $A_J = 0.3 \div 0.4$, which significantly exceeds the respective prediction of the PYTHIA Monte-Carlo event generator (which neglects the medium effects). This demonstrates that there is additional energy loss by the jet, estimated as 20 to 30 GeV, due to its interactions in the medium. But this energy loss does not lead to significant angular decorrelations: as visible in the right panel of Fig. 43, the distribution of the subleading jet is still peaked at $\Delta\Phi_{12} = \pi$, like in p+p collisions (and in good agreement with the PYTHIA simulations). This implies that the additional energy loss is not to be attributed to rare, hard, emissions (which would typically lead to 3-jet events). A careful analysis of the background around the away side jet allowed one to establish that the missing energy is in fact associated with relatively *soft particles* ($p_{\perp} \leq 2\text{ GeV}$) emitted at *large angles* with respect to the jet axis [133]. This rises the question about the physical mechanism which is responsible for such in-medium emissions at large angles.

As we shall now explain, a natural mechanism in that sense exists in QCD at weak coupling: this is

medium-induced gluon radiation, that is, the emission of gluons stimulated by the interactions between the partons composing the jet and the constituents of the medium. The most interesting situation — originally studied by Baier, Dokshitzer, Mueller, Peigné, and Schiff [134, 135] and independently by Zakharov [136, 137], following pioneering work by Gyulassy and Wang [138] (see also Refs. [139–145]) —, is when the medium is so dense that the gluon *formation time* is much larger than the *mean free path* of a parton propagating through the medium. In that case, there are many collisions which coherently contribute to the emission of a single gluon (see Fig. 44), leading to a suppression of the radiation spectrum as compared to the (Bethe–Heitler) spectrum that would be produced via independent multiple scattering. This suppression is known as the Landau–Pomeranchuk–Migdal (LPM) effect.

The ‘gluon formation time’ is the typical time that it takes in order to emit a gluon with a given kinematics. This concept is quite similar to the ‘fluctuation lifetime’ introduced in Eq. (5), but it is instructive to present here an alternative derivation for it, which is adapted to the problem at hand. To be specific, consider the emission of a gluon with momentum k and energy $\omega = |k|$ by an energetic quark propagating along the z axis. Even though the $q\bar{q}g$ vertex in QCD is local, the emission process is truly non-local, as it takes some time for the emitted gluon to lose coherence w.r.t. its parent quark. Namely, when the gluon starts being emitted, its wavefunction is still overlapping with that of the quark, so the two quanta cannot be distinguished from each other. But with increasing time, the gluon separates from the quark and their quantum coherence gets progressively lost. When the quark is very energetic, the gluon is typically emitted at a very small angle $\theta \simeq k_{\perp}/\omega \ll 1$ and the coherence between the two quanta is measured by their overlap in the transverse space. The gluon is considered as being ‘formed’ (or ‘fully emitted’) when its transverse separation $b_{\perp} \simeq \theta \Delta t$ from the quark becomes larger than its transverse Compton wavelength $\lambda_{\perp} = 1/k_{\perp} \simeq 1/(\omega\theta)$. This condition is satisfied after a time

$$\Delta t_{\text{form}} \simeq \frac{2\omega}{k_{\perp}^2} \simeq \frac{2}{\omega\theta^2}. \quad (67)$$

(The factor of 2 in the numerator is conventional.) The above argument is completely general: it holds for gluon emissions in the medium or in the vacuum. What is different, however, in the two cases is the mechanism causing the radiation and the associated gluon spectrum.

To better appreciate this difference, remember first that an on-shell quark cannot radiate: it can produce virtual fluctuations and thus develop a partonic substructure, as discussed in Section 3, but energy-momentum conservation prevent these quanta to become on-shell, and hence to separate from the parent quark. For the radiation to be possible, the quark and/or the emitted gluon must suffer additional interactions, which provide the energy deficit.

Consider first the situation in the *vacuum*: the quark is produced in an off-shell state via a hard scattering and then evacuates its virtuality via bremsstrahlung. Namely, it emits a gluon with energy ω and transverse momentum $k_{\perp}$ within a time interval $\Delta t \sim 2\omega/k_{\perp}^2$ after the original scattering. These values ω and $k_{\perp}$ are arbitrary (subjected to energy-momentum conservation) and independent of each other. However the bremsstrahlung spectrum, Eq. (16), favors the emission of soft ($x \ll 1$, or relatively small ω) and nearly collinear ($\theta \rightarrow 0$, or small $k_{\perp}$) gluons, for which the formation time is long.

The situation is very different for a quark propagating through a *dense medium*: the quark undergoes collisions with the medium constituents, with a typical distance ℓ (the mean free path) between two successive collisions. Any such a collision provides a small acceleration, thus allowing the quark to radiate. Accordingly, the initial virtuality of the quark is not essential anymore: the quark can now radiate *anywhere* within the medium, and not only within a distance $\sim \Delta t_{\text{form}}$ after the original hard scattering. This implies that the phase-space for in-medium emissions is *enhanced* by a factor $L/\Delta t_{\text{form}}$ with respect to emissions in the vacuum. Here L is the longitudinal extent of the medium as crossed by the quark and is typically much larger than Δt_{form} , as we shall see. Moreover, the emission mechanism and the associated formation time are influenced by the *gluon* interactions in the medium, which destroy the coherence between the gluon and the parent quark and thus facilitate the radiation.

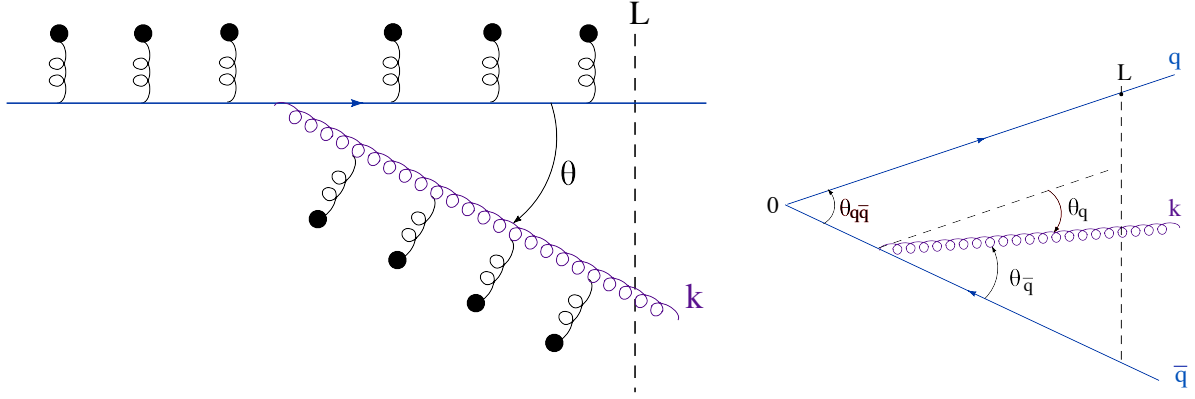


Fig. 44: Right: A cartoon illustrating medium-induced gluon radiation by an energetic quark propagating through a dense QCD medium. Both the quark and the emitted gluon undergo multiple scattering off the medium constituents (represented by the black blobs). Left: A quark-antiquark antenna emitting a gluon (here, from the antiquark leg) within the medium with size L . The interactions with the medium are not explicitly shown.

To estimate Δt_{form} for medium-induced emissions, we also need the rate at which the (virtual) gluon accumulates transverse momentum via rescattering in the medium. We shall later check that the successive collisions proceed independently from each other and thus provide transverse momenta which are randomly oriented and add in quadrature. This implies that the *average* transverse momentum *squared* grows linearly with time: $\langle k_{\perp}^2 \rangle \simeq \hat{q} \Delta t$, where Δt is the lifetime of the virtual gluon, as measured from the emission vertex, and $\hat{q}$ is a medium-dependent transport coefficient known as the *jet quenching parameter*, to be specified later. Hence, during its formation, the gluon acquires a typical transverse momentum squared $k_f^2 \simeq \hat{q} \Delta t_{\text{form}}$ via scattering within the medium. On the other hand, the condition for quantum decoherence requires the relation (67) between Δt_{form} and k_f^2 . Together, these two conditions determine *both* the formation time and the typical transverse momentum of the gluon at the time of emission:

$$\Delta t_{\text{form}} \simeq \frac{2\omega}{k_f^2} \quad \& \quad k_f^2 \simeq \hat{q} \Delta t_{\text{form}} \quad \implies \quad \Delta t_{\text{form}} \simeq \sqrt{\frac{2\omega}{\hat{q}}} \quad \& \quad k_f^2 \simeq (2\omega\hat{q})^{1/2}. \quad (68)$$

We thus see that, for medium-induced radiation, $k_{\perp}$ and ω are *not* independent kinematical variables anymore: the transverse momentum of the emitted gluon is acquired via interactions in the medium during the formation time which grows with the energy like $\Delta t_{\text{form}} \propto \omega^{1/2}$.

This mechanism for gluon production is operational provided the formation time is much larger than the mean free path ℓ , but smaller than the size L of the medium which is available for the emission process (the distance traveled through the plasma by the parent quark):

$$\ell \ll \Delta t_{\text{form}} \leq L \quad \implies \quad \omega_{\min} \equiv \frac{1}{2} \hat{q} \ell^2 \ll \omega \leq \omega_c \equiv \frac{1}{2} \hat{q} L^2. \quad (69)$$

These arguments imply that the typical emission angle at the time of formation, $\theta_f \simeq k_f/\omega$, cannot be arbitrarily small:

$$\theta_f \simeq \frac{k_f}{\omega} \simeq \left(\frac{2\hat{q}}{\omega^3} \right)^{1/4} \quad \implies \quad \theta_c \equiv \frac{2}{\sqrt{\hat{q}L^3}} \leq \theta_f \ll \theta_{\max} \equiv \frac{2}{\sqrt{\hat{q}\ell^3}}. \quad (70)$$

Unlike bremsstrahlung, the in-medium radiation does *not* favour collinear radiation. In fact, the smaller is the gluon energy ω , the larger is the emission angle θ_f and the shorter the emission time Δt_{form} . The final spectrum favors indeed the emission of *relatively soft* gluons with $\omega \ll \omega_c$, for which the

formation time is much smaller than the size of the medium, $\Delta t_{\text{form}} \ll L$, and the emission angle is quite large: $\theta_f \gg \theta_c$. Moreover, after being emitted, the gluons keep interacting with the medium and thus get deflected at even larger angles: their average transverse momentum can rise up to a final value $\langle k_{\perp}^2 \rangle \simeq \hat{q}(L - t_0) \sim \hat{q}L$, where t_0 is the time at the emission vertex. This phenomenon is known as *transverse momentum broadening*.

The above considerations show that the medium-induced radiation is very efficient in broadening the jet energy in the transverse plane, via the emission of soft ($\omega \ll \omega_c$) gluons, in qualitative agreement with the LHC data for di-jet asymmetry. On the other hand, the R_{AA} data for ‘high- $p_{\perp}$ suppression’ are probably more sensitive to the emission of harder gluons, with $\omega \sim \omega_c$, which dominate the energy loss by the ‘leading particle’ (the parton which has initiated the jet). Accordingly, the total energy loss $\Delta E \sim \omega_c \sim \hat{q}L^2$ scales like the *square* of the medium length L , and not like L (as one would expect for a mechanism where the energy is lost *locally*, say via elastic collisions in the medium). The reason for this scaling with L^2 is, of course, the fact that the actual mechanism at work is *non-local*: it takes a time $\Delta t_{\text{form}}(\omega)$ to emit a gluon and for $\omega \sim \omega_c$ this time is of the order of L . Since moreover the emission can be initiated at any point within L , the overall energy loss scales like L^2 . Reversing the argument, one concludes that the *stopping length* for a particle which loses all its energy inside the medium scales like $L_{\text{stop}} \sim E^{1/2}$, where E is the initial energy of that particle.

For more quantitative studies and applications to phenomenology, one still needs an estimate for the jet quenching parameter $\hat{q}$. Let us assume, for definiteness, that the medium is a quark-gluon plasma with (local) temperature T and weak coupling. The typical collisions which matter for the problem at hand are *soft*, in the sense that the typical transferred momentum is of the order of the Debye mass $m_D \sim gT$ introduced in Section 4. This is so since the soft collisions occur much more frequently than the hard ones: the corresponding cross-section is not the *transport* cross-section evaluated in Eq. (56) (that was the cross-section for scattering at *large angles*), but rather the *total* cross-section

$$\sigma_{\text{tot}} = \int d\Omega \frac{d\sigma}{d\Omega} \propto \int d\theta \sin\theta \frac{\alpha_s^2}{T^2 \sin^4\theta} \sim \frac{\alpha_s^2}{T^2} \int_g \frac{d\theta}{\theta^3} \sim \frac{\alpha_s}{T^2} \ln \frac{1}{\alpha_s}. \quad (71)$$

This is dominated by small-angle scattering — the integral over θ is cut off at $\theta \sim g$ by the plasma effects — and is larger by a factor $1/\alpha_s$ than the transport cross-section (56). The relevant mean free path is obtained by inserting the total cross-section in Eq. (55): $\ell \sim 1/(n\sigma_{\text{tot}}) \sim [T\alpha_s \ln(1/\alpha_s)]^{-1}$. As anticipated, this is parametrically larger than the interaction range $1/m_D \sim 1/gT$, meaning that successive collisions can be treated as independent. Since on the average there is a transfer $\Delta k_{\perp}^2 \sim m_D^2$ of transverse momentum squared per collision, we finally conclude that

$$\hat{q} \equiv \frac{d\langle k_{\perp}^2 \rangle}{dt} \simeq \frac{m_D^2}{\ell} \sim \alpha_s^2 T^3 \ln(1/\alpha_s), \quad (72)$$

for a weakly-coupled QGP. This is merely a parametric estimate, valid to leading logarithmic accuracy, and as such it suffers from the same lack-of-accuracy drawback as discussed for other transport coefficients towards the end of Section 4.3: it cannot be trusted for phenomenological applications. In fact, the whole set-up above described is a bit too idealized to correspond to the actual experimental situation. To cope with that, more sophisticated, phenomenological models have been proposed which treat the geometry of the collision in a more realistic way (finite volume, longitudinal expansion, time and point dependent jet quenching parameter) and include also free parameters. Such models are quite successful in describing the data, for both the R_{AA} ratio (65) and the di-jet asymmetry (66) (see Ref. [146] for a recent discussion and more references), but at the expense of using a rather large (average) value for $\hat{q}$ — considerably larger than the corresponding perturbative estimate to leading logarithmic accuracy. This situation is sometimes viewed as an argument in favor of the strong coupling scenario, but it might simply reflect the inaccuracy of the current perturbative results.

Note finally that the theory discussed above has addressed the (medium-induced) emission of a single gluon, whereas in reality one expects the in-medium evolution of a jet to involve several succes-

sive emissions — both by the hard parton which has initiated the jet and by its descendants. (Multiple emissions become important when the quantity $\alpha_s(L/\Delta t_{\text{form}})$ — which is roughly the probability for one gluon emission — becomes of order one.) Phenomenological models generally assume that successive emissions proceed independently from each other, but this is still to be demonstrated: *a priori*, there could be interference effects between emissions by different sources (the various partons forming the jet). For jet evolution in the *vacuum*, one knows that such interference effects are indeed important: they lead to *angular ordering* of the subsequent emissions — the successive emission angles are smaller and smaller [2, 147]. For the case of *in-medium* radiation, there is so far no explicit calculation of two (or more) successive emissions, but there are studies of interference effects in the emission by two sources: a quark and an antiquark forming a ‘colour antenna’ (see the right panel of Fig. 44) [145, 148, 149]. In particular, the analysis in Ref. [145] shows that the interference effects are negligible so long as the antenna opening angle (the angle $\theta_{q\bar{q}}$ in Fig. 44) is much larger than the minimal angle θ_c introduced in Eq. (70). As previously explained, the typical emission angles obey this condition already at the time of formation and they become even larger at later times, due to the momentum broadening by the medium. This suggests that successive medium-induced gluon emissions can be effectively treated as *independent*, thus justifying a *probabilistic* approach to in-medium jet evolution, like in Refs. [150–152].

4.5 The AdS/CFT correspondence: insights at strong coupling

At several points in the previous presentation, we pointed out observables whose values as extracted from the heavy-ion data seem difficult to understand if the coupling is weak, but would be more naturally accommodated at strong coupling. These observables include the viscosity-over-entropy ratio η/s , the thermalization time τ_{eq} , the jet quenching parameter $\hat{q}$, and the heavy quark diffusion coefficient D . In all these cases, the hypothesis of a strong coupling must be subjected to caution. First, these quantities are measured only *indirectly*, that is, they are extracted from fits to the data based on complex analyses which involve theoretical prejudices (notably, on the overall physics scenario), various assumptions which are difficult to check, and a considerable amount of model-building (concerning e.g. the initial conditions for hydrodynamics, the geometry of the collision, the theoretical description of multi-particle interactions). So, it is fair to say that the systematic uncertainties on these observables are still quite large (even though, within a *given* scenario, they might be strongly constrained by the data). Second, the weak-coupling results which serve as benchmarks for comparison are generally leading-order results in the perturbative expansion. But, as emphasized in Section 4.3, the standard perturbation theory (i.e. the strict expansion in powers of the coupling) is not reliable for the description of transport phenomena, even if the coupling is weak. This is so because of the need to resum finite-density effects like the Hard Thermal Loops to all orders, and this has not been done so far for dynamical quantities. The situation becomes even more complicated for the far-from-equilibrium situations, as relevant for the phenomenology, where the medium effects are not well understood.

This being said, the hypothesis of a strong coupling is both interesting and intriguing, and not easy to refute on the basis of asymptotic freedom or of the current lattice data. Indeed, as already noted in Section 4.3, the QCD coupling g is quite large when evaluated for temperatures a few times T_c (the critical temperature for deconfinement): $g = 1.5 \div 2$. As also mentioned there, the perturbative series at finite temperature is truly an expansion in powers of g (and not of $\alpha_s = g^2/4\pi$), so for that purpose the coupling is moderately strong. Reorganizing the perturbation theory via appropriate resummations of medium effects is one of the possible strategies to cope with this problem. But performing fully non-perturbative calculations, whenever possible, is clearly interesting. For thermodynamics, lattice QCD is the obvious and pertinent non-perturbative tool. As discussed in Section 4.3, its results are roughly consistent with those of HTL resummations at weak coupling. Yet, as we shall later argue, the lattice QCD results for the pressure do not totally exclude a strong coupling scenario. For real-time quantities and the non-equilibrium evolution, lattice methods become unapplicable (or, at least, inefficient), so it has become common practice to rely on the *AdS/CFT correspondence* for guidance as to general properties

of strongly coupled field theories at finite temperature. (See Ref. [153] for a general review on AdS/CFT and Refs. [119, 154–156] for recent reviews of its applications to a finite-temperature plasma.)

The AdS/CFT correspondence (or ‘gauge/string duality’) does not apply to QCD, but to a ‘cousin’ of it, the $\mathcal{N} = 4$ supersymmetric Yang–Mills (SYM) theory, which has a non-Abelian *gauge* symmetry with the ‘colour’ group $SU(N_c)$, like QCD, but also additional *global* symmetries (notably, supersymmetry), which strongly constrain the dynamics. These additional symmetries ensure that the *conformal invariance* of the classical Lagrangian is preserved after including quantum corrections — meaning that, unlike in QCD, the coupling is fixed and there is no confinement. Accordingly, this theory has probably little to say about the zero-temperature, hadronic, phase of QCD, where the non-perturbative aspects of QCD are controlled by confinement. Moreover, this is probably not a good model for the QCD dynamics in the vicinity of the deconfinement phase transition, where the running coupling effects are known to be important, as shown by the lattice results for the ‘trace anomaly’ in Fig. 37 (left). However, as also manifest in that figure, the relative ‘anomaly’ $(\epsilon - 3P)/\epsilon$ decreases very fast with increasing T above T_c and becomes unimportant (smaller than 10%) for $T \gtrsim 2T_c \simeq 400$ MeV. Hence, there is a hope that, within the intermediate range of temperatures at $2T_c \lesssim T \lesssim 5T_c$, which is the relevant range for heavy ion collisions at RHIC and the LHC, the dynamics in QCD may be at least qualitatively understood by analogy with the $\mathcal{N} = 4$ SYM theory at strong coupling.

Specifically, the AdS/CFT correspondence is a *duality*, that is, an equivalence between two theories which *a priori* look very different from each other: (i) the $\mathcal{N} = 4$ SYM gauge theory mentioned above (the ‘conformal field theory’, or CFT) and (ii) a special, ‘type II B’, string theory, leaving in a curved 10 dimensional space-time with Anti-de-Sitter (AdS) geometry⁹. This duality is interesting in that it maps the strong coupling sector of $\mathcal{N} = 4$ SYM onto the weak coupling sector of the string theory. Accordingly, it allows one to compute observables in the CFT at strong coupling via perturbative calculations in the string theory. More precisely, the ‘strong coupling limit’ to which refers the duality is the special limit (g denotes the gauge coupling in $\mathcal{N} = 4$ SYM)

$$\lambda \equiv g^2 N_c \rightarrow \infty \quad \text{with} \quad g^2 \ll 1, \quad (73)$$

that is, the limit of a large number of colours ($N_c \rightarrow \infty$) taken for a fixed, and relatively small, value of the gauge coupling g . This defines indeed a regime of strong coupling (despite g being small) because when $N_c \gg 1$ the effective coupling in the gauge theory is the ‘t Hooft coupling $\lambda = g^2 N_c$. (This is also true for QCD with colour group $SU(N_c)$.) For instance, the perturbation theory in the multi-colour limit is dominated by planar Feynman graphs which are such that each additional loop brings a factor of λ . As long as $\lambda \ll 1$, the large- N_c limit of the theory can be studied in a perturbative expansion in powers of λ . In the opposite limit $\lambda \rightarrow \infty$ (but with $g \ll 1$), one can rely on the AdS/CFT correspondence. In that limit, the dual string theory reduces to ‘supergravity’ (or SUGRA) — a classical field theory in a curved space-time with 10 dimensions. From the solutions to the classical equations of motion (e.g., Einstein equations), one can unambiguously construct, via the AdS/CFT dictionary, the correlations in the $\mathcal{N} = 4$ SYM theory at infinitely strong coupling.

The $\mathcal{N} = 4$ SYM theory at finite temperature and $\lambda \gg 1$ provides a model for the strongly-coupled quark-gluon plasma (sQGP). The corresponding string-theory dual is obtained by adding a *black hole* into the AdS space-time. This is somewhat natural, since, as we know from Hawking, a black hole has entropy and generates black-body radiation, so in that sense it behaves indeed like a thermal system. The entropy of a black hole is proportional to the area of its event horizon. (No information can escape from the volume inside the horizon, so this volume cannot contribute to the entropy.) The corresponding,

⁹More precisely, this 10 dimensional space-time is the direct product $\text{AdS}_5 \times S^5$, where AdS_5 is the 5-dimensional Anti-de-Sitter space-time, with constant negative curvature, and S^5 is the 5-dimensional sphere, with constant positive curvature; see e.g. Ref. [153] for details.

Bekenstein–Hawking, formula can be adapted to supergravity, to yield

$$S_{\text{BH}} = \frac{\text{Horizon area}}{4G_{10}} \implies s \equiv \frac{S_{\text{BH}}}{V_3} = \frac{\pi^2}{2} N_c^2 T^3 = \frac{3}{4} s_0, \quad (74)$$

where G_{10} is Newton constant in 10 dimensions and V_3 is the volume of the physical 3-dimensional space (as usual, we set $\hbar = c = k_B = 1$). The last equality in Eq. (74) shows that the entropy density s of the $\mathcal{N} = 4$ SYM plasma at *infinitely strong* coupling is 3/4 of the corresponding quantity s_0 at *zero* coupling ! Hence, in spite of the interactions being so strong, the entropy does not deviate strongly from that of an ideal gas. The first correction to this result at strong coupling, of order $1/\lambda^{3/2}$, is also known [153] and it is positive — meaning that the NLO result for s is even closer to the respective Stefan–Boltzmann limit. Even though such results cannot be directly applied to QCD, they nevertheless suggest that the relatively small deviations — about 20% in the temperature range relevant for HIC’s, cf. Fig. 36 — between the lattice results for the pressure in QCD and the respective ideal gas limit are not necessarily in contradiction with a strong coupling scenario.

We have previously mentioned that one important prediction of AdS/CFT is the limiting value (61) for the ratio η/s , which has been conjectured to be a lower bound of nature [90] (as it holds at infinitely strong coupling for all the gauge theories having a gravity dual). Let us sketch here the derivation of this result [89, 90]. As explained around Eq. (59), the shear viscosity η describes the response of the plasma to ‘shear forces’ (its ability to transfer momentum p_x along the y direction). This is made precise by the Kubo formula which expresses η as a 2-point function of the shear tensor T_{xy} in thermal equilibrium:

$$\eta = \lim_{\omega \rightarrow 0} \frac{1}{2\omega} \int dt d^3\mathbf{x} e^{-i\omega t} \langle [T_{xy}(t, \mathbf{x}), T_{xy}(0, \mathbf{0})] \rangle_T. \quad (75)$$

This representation is exact, so in particular it holds at strong coupling. In that case, the string-theory dual of the 2-point function in the r.h.s. is the cross-section for the absorption of a soft *AdS graviton* (the supergravity field dual to the energy-momentum tensor in the CFT) by the black hole. This cross-section is known from general relativity: it is proportional to the area of the event horizon, like the entropy. Thus, in this context, η and s are naturally proportional with each other. The proportionality coefficient can be explicitly computed, with the result that $\eta/s = 1/(4\pi)$. It is remarkable that the heavy ion data seem to favour a value which is close to this conjectured lower bound.

From Section 4.1 we recall that the shear viscosity enters the equations of hydrodynamics at *linear* order in the gradient expansion. This corresponds to the fact that, in the respective Kubo formula (75), η is extracted from the term linear in ω in the small frequency expansion of the 2-point function of the shear tensor. By going up to the second order in this expansion, one can similarly extract the transport coefficients for the *second-order* formalism (which, we recall, are essential to provide a consistent formulation of relativistic hydrodynamics). Interestingly, the calculation of these coefficients turns out to be simpler at *strong coupling*, where one can rely on AdS/CFT for that purpose, than at *weak coupling*, where the respective calculations (say, using kinetic theory) require the resummation of infinitely many Feynman graphs. And as a matter of fact, the general structure of the second-order terms in the equations of hydrodynamics has been clarified only recently, via AdS/CFT calculations at strong coupling [157, 158]. Thus, thanks to AdS/CFT, one can study the emergence of hydrodynamics from the underlying fundamental field theory in a controlled way (at least on the example of $\mathcal{N} = 4$ SYM).

Since the string theory dual of a finite- T plasma is a black hole in AdS_5 , it is natural that the phenomenon of thermalization at strong coupling can be studied as the emergence of an event horizon in the solution to the Einstein equations. We more precisely mean the equations describing the deviation in the AdS_5 metric generated by some matter distribution, which is initially out of equilibrium. For instance, the collision between two heavy ions can be modeled as the scattering between two gravitational shock waves — two Lorentz contracted shells of matter which propagate against each other and scatter via gravitational interactions. This problem has been addressed within the supergravity context, via analytic

approximations [159–161] and via exact, numerical, calculations [162], with the conclusion that the system evolves rather fast towards a (locally) isotropic distribution. Interestingly, it appears that the only acceptable solution to the Einstein equations which is boost invariant is the one describing the emergence of perfect hydrodynamics (in the sense of Eq. (58)) at asymptotically large times [163, 164].

Another phenomenon which is interesting to study at strong coupling is jet quenching — the energy loss by a ‘hard probe’ (energetic parton) propagating through a strongly coupled plasma. The corresponding AdS/CFT calculations have been performed both for a very heavy quark (which loses only a tiny fraction of its total energy) and for a light parton (quark, gluon, or virtual photon), which can be totally stopped in the medium [119, 155, 156] (and references therein). Here we shall focus on the second case — that of a light, but very energetic, parton with original energy $E \gg T$. The corresponding ‘dual’ object on the supergravity side can be a semi-classical string falling into AdS_5 (in the case of a light quark), a pair of such strings (to describe a gluon), or a falling wavepacket carrying the photon quantum numbers (for a virtual photon). The respective AdS/CFT calculations [165–168] revealed that, in all such cases, the stopping distance over which the light parton loses most of its energy through interactions in the medium scales like

$$L_{\text{stop}} \sim \frac{1}{T} \left(\frac{E}{T} \right)^{1/3}. \quad (76)$$

Note the difference w.r.t. the corresponding result at weak coupling, which in Section 4.4 has been found to scale like $E^{1/2}$. This reflects the difference between the respective mechanisms for energy loss, that we shall now explain [155, 165, 169].

From Section 4.4, we recall that the mechanism at work at weak coupling is *medium-induced radiation* — the emission of gluons stimulated by the interactions between the radiating system (the ‘hard probe’ and its partonic descendants) and the individual constituents of the medium. In general, several such interactions can contribute to the emission of a single gluon (the LPM effect), but the role of the individual interactions is nevertheless well identified: they provide transverse momentum kicks at a rate measured by the jet quenching parameter, Eq. (72). This is in agreement with the fact that, at weak coupling, the plasma is a collection of elementary constituents, or ‘quasi-particles’ (cf. Section 4.3), which are pointlike and quasifree. But at strong coupling, we do not expect such a quasi-particle picture to hold anymore — rather, the plasma should look homogeneous, without any microscopic substructure. And indeed, the AdS/CFT results like Eq. (76) can be understood by assuming that the plasma acts on the external probe with a *uniform force* $F_T \sim T^2$. This is like a gravitational force in the sense that it is fully determined by the local energy density $\sim T^4$ in the plasma, irrespective of its microscopic nature. The effect of this force on a virtual parton (the ‘hard probe’) is to stimulate gluon emission, via *medium-induced parton branching* [155, 165, 169].

Specifically, a partonic fluctuation with energy ω and virtuality Q can decay under the action of the plasma force F_T provided the mechanical work $W = L F_T$ furnished by this force over a distance L of the order of the lifetime of the fluctuation ($L \sim \omega/Q^2$) is large enough to compensate the parton virtuality. This condition implies

$$\frac{\omega}{Q^2} T^2 \sim Q \implies Q = Q_s(\omega) \sim (\omega T^2)^{1/3} \quad \& \quad L \simeq \frac{\omega}{Q_s^2(\omega)} \sim \frac{1}{T} \left(\frac{\omega}{T} \right)^{1/3}, \quad (77)$$

in agreement with Eq. (76). More precisely, the above argument provides the typical distance L for the occurrence of one branching, but this is of the same order of magnitude as the overall stopping distance; indeed, the subsequent branchings involve gluons which are softer and softer, and thus proceed faster and faster. For a given energy $\omega \gg T$, any parton with initial virtuality $Q_0 \leq Q_s$ can decay in this way, including the *space-like* photon exchanged in DIS (cf. Section 3.1). Accordingly, the quantity $Q_s(\omega)$ plays also the role of the *saturation momentum* for the finite- T plasma at strong coupling. However, unlike at weak coupling, where the phenomenon of saturation requires large gluon occupation numbers

$n \sim 1/\alpha_s$, cf. Eq. (23), at strong coupling one can argue that it occurs for occupation numbers of order one [155, 170].

In the previous discussion we have implicitly assumed the plasma to be infinite (or, at least, much larger than the stopping distance (76)). This means that, on the supergravity side, one has studied the propagation of the ‘dual’ objects in a metric describing a black hole into AdS_5 . The corresponding calculations for a *finite-size* medium are more difficult, in particular because the corresponding metric is more complicated. But one can at least heuristically revert the logic leading to Eq. (76) and conclude that, if an energetic parton propagates through the medium over a finite distance L without being stopped, then the amount of energy lost by the particle scales like $\Delta E \sim L^3$ [169]. Interestingly, this scaling appears to be supported by some of the RHIC data [171]. This result at strong coupling should be contrasted with the corresponding scaling-law at weak coupling, namely $\Delta E \sim \hat{q}L^2$ (cf. Section 4.4). This difference reflects the fact that the medium-induced parton branching is not a *local* phenomenon (unlike transverse momentum broadening at weak coupling), but is delocalized over a distance of the order of the lifetime E/Q_s^2 of the decaying parton, which in turn is commensurable with its stopping distance.

The above picture of medium-induced parton branching can also explain the AdS/CFT results for the energy loss and the transverse momentum broadening of a *heavy quark* propagating through a strongly coupled plasma [119, 156] (and Refs. therein). In that case, the variables ω and Q which appear in Eq. (77) refer to any of the quanta emitted by the heavy quark: among all the virtual fluctuations of the latter, the only ones which can decay (and thus take away energy and momentum) are those which, for a given energy ω , have a relatively small virtuality $Q \lesssim Q_s(\omega)$. (Quanta with $Q \gg Q_s(\omega)$ cannot significantly interact with the plasma and hence they are reabsorbed by the heavy quark.) The energy loss is dominated by the most energetic among the emitted quanta — those having a boost factor ω/Q comparable to that (denoted as γ) of the heavy quark. These two conditions, $\omega/Q \simeq \gamma$ and $Q \lesssim Q_s(\omega)$, imply the following upper limits on the energy and transverse momentum that can be taken away by one emitted parton: $\omega \leq \omega_{max}$ with $\omega_{max} \sim \gamma Q_s(\omega_{max}) \sim \gamma^{3/2}T$ and, respectively, $\Delta k_\perp \lesssim Q_s(\omega_{max}) \sim \gamma^{1/2}T$. These maximal values control the energy loss and the transverse momentum broadening of the heavy quark. By also taking into account the typical duration ω/Q^2 of an emission, one finally deduces the following expressions

$$-\frac{dE}{dt} \simeq \sqrt{\lambda} \frac{\omega}{(\omega/Q_s^2)} \Big|_{\omega_{max}} \simeq \sqrt{\lambda} Q_s^2 \sim \sqrt{\lambda} \gamma T^2. \quad (78)$$

$$\frac{d\langle k_\perp^2 \rangle}{dt} \sim \frac{\sqrt{\lambda} Q_s^2}{(\omega/Q_s^2)} \sim \sqrt{\lambda} \frac{Q_s^4}{\gamma Q_s} \sim \sqrt{\lambda} \sqrt{\gamma} T^3, \quad (79)$$

for the respective rates. The factor $\sqrt{\lambda}$ in the r.h.s.’s of these equations appears because the heavy quark is a semi-classical object which acts as a colour source with a strength of order $\sqrt{\lambda}$ at strong coupling — meaning that it emits a number of quanta (with given ω and Q) of order $\sqrt{\lambda}$ during the formation time ω/Q^2 of one such a quanta. As anticipated, Eqs. (78)–(79) agree at parametric accuracy with the respective results of the AdS/CFT calculations [119, 156].

Acknowledgments

I would like to thank the organizers of the 2011 European School of High-Energy Physics, notably Nick Evans, Christophe Grojean and Ioan Ursu, for inviting me to present this series of lectures. During the preparation of the lecture notes, I have benefited from many exchanges and related discussions. I am particularly grateful to my colleagues who patiently read these pages and helped me improved my presentation and correct the misprints: Fabio Dominguez, Hanna Grönqvist, Jean-Yves Ollitrault, Dionysis Triantafyllopoulos, and Raju Venugopalan. Also, many of the figures which appear in these notes have been graciously provided by my colleagues in Saclay, notably François Gelis (Figs. 1, 6 left, 15, 17 left, 25, 26 right, 32 right, and 42 right) and Matt Luzum (Fig. 28).

References

- [1] V.N. Gribov and L.N. Lipatov, *Sov. Journ. Nucl. Phys.* **15** (1972), 438; G. Altarelli and G. Parisi, *Nucl. Phys.* **B126** (1977), 298; Yu. L. Dokshitzer, *Sov. Phys. JETP* **46** (1977), 641.
- [2] R. K. Ellis, W. J. Stirling and B. R. Webber, *QCD and collider physics, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **8** (1996) 435 p.
- [3] L.N. Lipatov, *Sov. J. Nucl. Phys.* **23** (1976) 338; E.A. Kuraev, L.N. Lipatov and V.S. Fadin, *Zh. Eksp. Teor. Fiz* **72**, 3 (1977) (*Sov. Phys. JETP* **45** (1977) 199); Ya.Ya. Balitsky and L.N. Lipatov, *Sov. J. Nucl. Phys.* **28** (1978) 822.
- [4] HERA Combined Structure Functions Working Group Collaboration, C. Gwenlan, *Nucl. Phys. Proc. Suppl.* **191** (2009) 5, arXiv:0902.1807.
- [5] L. V. Gribov, E. M. Levin and M. G. Ryskin, *Phys. Rept.* **100** (1983) 1–150.
- [6] A. H. Mueller and J.-w. Qiu, *Nucl. Phys.* **B268** (1986) 427.
- [7] L. D. McLerran and R. Venugopalan, *Phys. Rev.* **D49** (1994) 2233, hep-ph/9309289.
- [8] E. Iancu, A. Leonidov and L. McLerran, *The colour glass condensate: An introduction*, hep-ph/0202270.
- [9] A. H. Mueller, *Parton saturation: An Overview*, hep-ph/0111244.
- [10] E. Iancu and R. Venugopalan, *The color glass condensate and high energy scattering in QCD*, hep-ph/0303204.
- [11] J. Jalilian-Marian and Y. V. Kovchegov, *Prog. Part. Nucl. Phys.* **56** (2006) 104, hep-ph/0505052.
- [12] H. Weigert, *Prog. Part. Nucl. Phys.* **55** (2005) 461, hep-ph/0501087.
- [13] D. N. Triantafyllopoulos, *Acta Phys. Polon.* **B36** (2005) 3593, hep-ph/0511226.
- [14] F. Gelis, E. Iancu, J. Jalilian-Marian and R. Venugopalan, *Ann. Rev. Nucl. Part. Sci.* **60** (2010) 463, arXiv:1002.0333.
- [15] T. Lappi, *Int. J. Mod. Phys.* **E20** (2011) 1, arXiv:1003.1852.
- [16] D. N. Triantafyllopoulos, *Nucl. Phys.* **B648** (2003) 293, hep-ph/0209121.
- [17] L. McLerran and M. Praszalowicz, *Acta Phys. Polon.* **B41** (2010) 1917, arXiv:1006.4293.
- [18] J. Jalilian-Marian, A. Kovner, A. Leonidov and H. Weigert, *Nucl. Phys.* **B504** (1997) 415; *Phys. Rev.* **D59** (1999) 014014; J. Jalilian-Marian, A. Kovner and H. Weigert, *Phys. Rev.* **D59** (1999) 014015; A. Kovner, J. G. Milhano and H. Weigert, *Phys. Rev.* **D62** (2000) 114005; H. Weigert, *Nucl. Phys.* **A703** (2002) 823.
- [19] E. Iancu, A. Leonidov and L. McLerran, *Nucl. Phys.* **A692** (2001) 583; *Phys. Lett.* **B510** (2001) 133; E. Ferreira, E. Iancu, A. Leonidov and L. McLerran, *Nucl. Phys.* **A703** (2002) 489; E. Iancu and L. McLerran, *Phys. Lett.* **B510** (2001) 145.
- [20] I. Balitsky, *Nucl. Phys.* **B463** (1996) 99, hep-ph/9509348.
- [21] Y. V. Kovchegov, J. Kuokkanen, K. Rummukainen and H. Weigert, *Nucl. Phys.* **A823** (2009) 47, arXiv:0812.3238.
- [22] E. Iancu and D. Triantafyllopoulos, *JHEP* **1204** (2012) 025, arXiv:1112.1104.
- [23] Y. V. Kovchegov, *Phys. Rev.* **D60** (1999) 034008, hep-ph/9901281.
- [24] K. Rummukainen and H. Weigert, *Nucl. Phys.* **A739** (2004) 183, hep-ph/0309306.
- [25] T. Lappi, *Phys. Lett.* **B703** (2011) 325, arXiv:1105.5511.
- [26] A. Dumitru, J. Jalilian-Marian, T. Lappi, B. Schenke and R. Venugopalan, *Phys. Lett.* **B706** (2011) 219, arXiv:1108.4764.
- [27] A. H. Mueller, *Nucl. Phys.* **B415** (1994) 373.
- [28] I. Balitsky, *Phys. Rev.* **D75** (2007) 014001, hep-ph/0609105.
- [29] Y. V. Kovchegov and H. Weigert, *Nucl. Phys.* **A784** (2007) 188, hep-ph/0609090.

- [30] I. Balitsky and G. A. Chirilli, *Phys. Rev.* **D77** (2008) 014019, arXiv:0710.4330.
- [31] K. J. Golec-Biernat and M. Wusthoff, *Phys. Rev.* **D59** (1998) 014017, hep-ph/9807513.
- [32] E. Iancu, K. Itakura and S. Munier, *Phys. Lett.* **B590** (2004) 199, hep-ph/0310338.
- [33] H. Kowalski, L. Motyka and G. Watt, *Phys. Rev.* **D74** (2006) 074016, hep-ph/0606272.
- [34] V. Goncalves, M. Kugeratski, M. Machado and F. Navarra, *Phys. Lett.* **B643** (2006) 273, hep-ph/0608063.
- [35] P. Tribedy and R. Venugopalan, *Nucl. Phys.* **A850** (2011) 136, arXiv:1011.1895.
- [36] J. L. Albacete, N. Armesto, J. G. Milhano, P. Quiroga-Arias and C. A. Salgado, *Eur. Phys. J.* **C71** (2011) 1705, arXiv:1012.4408.
- [37] Y. V. Kovchegov and A. H. Mueller, *Nucl. Phys.* **B529** (1998) 451, hep-ph/9802440.
- [38] Y. V. Kovchegov and K. Tuchin, *Phys. Rev.* **D65** (2002) 074026, hep-ph/0111362.
- [39] J. P. Blaizot, F. Gelis and R. Venugopalan, *Nucl. Phys.* **A743** (2004) 13, hep-ph/0402256.
- [40] A. M. Stasto, K. J. Golec-Biernat and J. Kwiecinski, *Phys. Rev. Lett.* **86** (2001) 596, hep-ph/0007192.
- [41] E. Iancu, K. Itakura and L. McLerran, *Nucl. Phys.* **A708** (2002) 327, hep-ph/0203137.
- [42] A. H. Mueller and D. N. Triantafyllopoulos, *Nucl. Phys.* **B640** (2002) 331, hep-ph/0205167.
- [43] S. Munier and R. B. Peschanski, *Phys. Rev. Lett.* **91** (2003) 232001, hep-ph/0309177.
- [44] A. Kovner and U. A. Wiedemann, *Phys. Rev.* **D64** (2001) 114002, hep-ph/0106240.
- [45] J. P. Blaizot, F. Gelis and R. Venugopalan, *Nucl. Phys.* **A743** (2004) 57, hep-ph/0402257.
- [46] J. Jalilian-Marian and Y. V. Kovchegov, *Phys. Rev.* **D70** (2004) 114017, hep-ph/0405266.
- [47] R. Baier, A. Kovner, M. Nardi and U. A. Wiedemann, *Phys. Rev.* **D72** (2005) 094013, hep-ph/0506126.
- [48] C. Marquet, *Nucl. Phys.* **A796** (2007) 41, arXiv:0708.0231.
- [49] F. Dominguez, C. Marquet, B.-W. Xiao and F. Yuan, *Phys. Rev.* **D83** (2011) 105005, arXiv:1101.0715.
- [50] E. Avsar, *TMD factorization and the gluon distribution in high energy QCD*, arXiv:1203.1916.
- [51] F. Gelis, T. Lappi and R. Venugopalan, *Phys. Rev.* **D78** (2008) 054019, arXiv:0804.2630.
- [52] A. Krasnitz, Y. Nara and R. Venugopalan, *Phys. Rev. Lett.* **87** (2001) 192302, hep-ph/0108092.
- [53] A. Krasnitz, Y. Nara and R. Venugopalan, *Nucl. Phys.* **A717** (2003) 268, hep-ph/0209269.
- [54] T. Lappi, *Phys. Rev.* **C67** (2003) 054903, hep-ph/0303076.
- [55] BRAHMS Collaboration, I. Arsene *et al.*, *Phys. Rev. Lett.* **93** (2004) 242303, nucl-ex/0403005.
- [56] STAR Collaboration, J. Adams *et al.*, *Phys. Rev. Lett.* **97** (2006) 152302, nucl-ex/0602011.
- [57] D. Kharzeev, E. Levin and L. McLerran, *Phys. Lett.* **B561** (2003) 93, hep-ph/0210332.
- [58] J. L. Albacete, N. Armesto, A. Kovner, C. A. Salgado and U. A. Wiedemann, *Phys. Rev. Lett.* **92** (2004) 082001, hep-ph/0307179.
- [59] E. Iancu, K. Itakura and D. Triantafyllopoulos, *Nucl. Phys.* **A742** (2004) 182, hep-ph/0403103.
- [60] D. Kharzeev, Y. V. Kovchegov and K. Tuchin, *Phys. Lett.* **B599** (2004) 23, hep-ph/0405045.
- [61] M. Praszalowicz, *Acta Phys. Polon.* **B42** (2011) 1557, arXiv:1104.1777.
- [62] PHOBOS Collaboration, B. Back *et al.*, *Nucl. Phys.* **A757** (2005) 28, nucl-ex/0410022.
- [63] G. I. Veres, *Nucl. Phys.* **A774** (2006) 287, nucl-ex/0511037.
- [64] STAR Collaboration, E. Braidot *et al.*, *Nucl. Phys.* **A854** (2011) 168, arXiv:1008.3989.
- [65] E. Braidot, *Two-particle azimuthal correlations at forward rapidity in STAR*, arXiv:1102.0931.
- [66] J. L. Albacete and C. Marquet, *Phys. Rev. Lett.* **105** (2010) 162301, arXiv:1005.4065.
- [67] ALICE Collaboration, K. Aamodt *et al.*, *Phys. Rev. Lett.* **106** (2011) 032301, arXiv:1012.1657.

- [68] J. L. Albacete and A. Dumitru, *A model for gluon production in heavy-ion collisions at the LHC with rcBK unintegrated gluon densities*, arXiv:1011.5161.
- [69] T. Lappi and L. McLerran, *Nucl. Phys.* **A772** (2006) 200, hep-ph/0602189.
- [70] STAR Collaboration, J. Adams *et al.*, *Phys. Rev. Lett.* **95** (2005)152301, nucl-ex/0501016. Data tables available at <http://www.star.bnl.gov/central/publications>.
- [71] PHENIX Collaboration, A. Adare *et al.*, *Phys. Rev.* **C78** (2008) 014901, arXiv:0801.4545.
- [72] STAR Collaboration, B. Abelev *et al.*, *Phys. Rev.* **C80** (2009) 064912, arXiv:0909.0191.
- [73] PHOBOS Collaboration, B. Alver *et al.*, *Phys. Rev. Lett.* **104** (2010) 062301, arXiv:0903.2811.
- [74] J. Jia, *J. Phys.* **G38** (2011) 124012, arXiv:1107.1468.
- [75] CMS Collaboration, W. Li, *J.Phys.G* **G38** (2011) 124027, arXiv:1107.2452.
- [76] J. F. Grosse-Oetringhaus, *J. Phys.* **G38** (2011) 124028, arXiv:1107.0556.
- [77] CMS Collaboration, V. Khachatryan *et al.*, *JHEP* **1009** (2010) 091, arXiv:1009.4122.
- [78] A. Dumitru, K. Dusling, F. Gelis, J. Jalilian-Marian, T. Lappi and R. Venugopalan, *Phys. Lett.* **B697** (2011) 21, arXiv:1009.5295.
- [79] D. E. Kharzeev, L. D. McLerran and H. J. Warringa, *Nucl. Phys.* **A803** (2008) 227, arXiv:0711.0950.
- [80] STAR Collaboration, B. Abelev *et al.*, *Phys. Rev.* **C81** (2010) 054908, arXiv:0909.1717.
- [81] J.-Y. Ollitrault, *Phys. Rev.* **D46** (1992) 229.
- [82] B. Alver and G. Roland, *Phys. Rev.* **C81** (2010) 054905, arXiv:1003.0194.
- [83] J. Jia, *J. Phys.* **G38** (2011) 124012, arXiv:1107.1468.
- [84] ATLAS Collaboration, G. Aad *et al.*, *Phys. Lett.* **B707** (2012) 330, arXiv:1108.6018.
- [85] P. F. Kolb and U. W. Heinz, *Hydrodynamic description of ultrarelativistic heavy ion collisions*, nucl-th/0305084. In *Quark Gluon Plasma 3*, Eds: R.C. Hwa and X.N. Wang, World Scientific, Singapore.
- [86] M. Luzum and P. Romatschke, *Phys. Rev.* **C78** (2008) 034915, arXiv:0804.4015.
- [87] J.-Y. Ollitrault, *J. Phys. Conf. Ser.* **312** (2011) 012002, arXiv:1008.3323.
- [88] M. Luzum, *J. Phys.* **G38** (2011) 124026, arXiv:1107.0592.
- [89] G. Policastro, D. T. Son and A. O. Starinets, *Phys. Rev. Lett.* **87** (2001) 081601, hep-th/0104066.
- [90] P. Kovtun, D. Son and A. Starinets, *Phys. Rev. Lett.* **94** (2005) 111601, hep-th/0405231.
- [91] R. Baier, A. H. Mueller, D. Schiff and D. Son, *Phys. Lett.* **B502** (2001) 51, hep-ph/0009237.
- [92] A. Ipp, A. Rebhan and M. Strickland, *Phys. Rev.* **D84** (2011) 056003, arXiv:1012.0298.
- [93] T. Epelbaum and F. Gelis, *Nucl. Phys.* **A872** (2011) 210, arXiv:1107.0668.
- [94] K. Dusling, F. Gelis and R. Venugopalan, *Nucl. Phys.* **A872** (2011) 161, arXiv:1106.3927.
- [95] A. Kurkela and G. D. Moore, *JHEP* **12** (2011) 044, arXiv:1107.5050.
- [96] A. Kurkela and G. D. Moore, *JHEP* **11** (2011) 120, arXiv:1108.4684.
- [97] J. Berges, K. Boguslavski and S. Schlichting, *Phys. Rev.* **D85** (2012) 076005 [arXiv:1201.3582 [hep-ph]].
- [98] J.-P. Blaizot, F. Gelis, J.-F. Liao, L. McLerran and R. Venugopalan, *Nucl. Phys.* **A873** (2012) 68, arXiv:1107.5296.
- [99] H. B. Meyer, *Eur. Phys. J.* **A47** (2011) 86, arXiv:1104.3708.
- [100] K. Rajagopal and F. Wilczek, *The condensed matter physics of QCD*, hep-ph/0011333.
- [101] M. G. Alford, A. Schmitt, K. Rajagopal and T. Schafer, *Rev. Mod. Phys.* **80** (2008) 1455, arXiv:0709.4635.

- [102] V. Rubakov, *Cosmology*, in these proceedings.
- [103] F. Karsch, *J. Phys.* **G30** (2004) S887, hep-lat/0403016.
- [104] T. Matsui and H. Satz, *Phys. Lett.* **B178** (1986) 416.
- [105] A. Mocsy, *Nucl. Phys.* **A830** (2009) 411C, arXiv:0908.0746.
- [106] Y. Aoki, S. Borsanyi, S. Durr, Z. Fodor, S. D. Katz, *et al.*, *JHEP* **0906** (2009) 088, arXiv:0903.4155.
- [107] M. Cheng, S. Ejiri, P. Hegde, F. Karsch, O. Kaczmarek, *et al.*, *Phys. Rev.* **D81** (2010) 054504, arXiv:0911.2215.
- [108] Y. Aoki, G. Endrodi, Z. Fodor, S. Katz and K. Szabo, *Nature* **443** (2006) 675, hep-lat/0611014.
- [109] Z. Fodor and S. Katz, *The Phase diagram of quantum chromodynamics*, arXiv:0908.3341.
- [110] J. O. Andersen, L. E. Leganger, M. Strickland, and N. Su, *JHEP* **1108** (2011) 053, arXiv:1103.2528.
- [111] K. Kajantie, M. Laine, K. Rummukainen and Y. Schroder, *Phys. Rev.* **D67** (2003) 105008, hep-ph/0211321.
- [112] J.-P. Blaizot and E. Iancu, *Phys. Rept.* **359** (2002) 355, hep-ph/0101103.
- [113] J.-P. Blaizot, E. Iancu and A. Rebhan, *Thermodynamics of the high temperature quark gluon plasma*, hep-ph/0303185.
- [114] U. Kraemmer and A. Rebhan, *Rept. Prog. Phys.* **67** (2004) 351, hep-ph/0310337.
- [115] E. Braaten and R. D. Pisarski, *Nucl. Phys.* **B337** (1990) 569.
- [116] J. Blaizot, E. Iancu and A. Rebhan, *Phys. Rev. Lett.* **83** (1999) 2906, hep-ph/9906340.
- [117] J. O. Andersen, E. Braaten and M. Strickland, *Phys. Rev. Lett.* **83** (1999) 2139, hep-ph/9902327.
- [118] S. Jeon, *Phys. Rev.* **D52** (1995) 3591, hep-ph/9409250.
- [119] J. Casalderrey-Solana, H. Liu, D. Mateos, K. Rajagopal and U. A. Wiedemann, *Gauge/String Duality, Hot QCD and Heavy Ion Collisions*, arXiv:1101.0618.
- [120] P. B. Arnold, G. D. Moore and L. G. Yaffe, *JHEP* **0011** (2000) 001, hep-ph/0010177.
- [121] P. B. Arnold, G. D. Moore and L. G. Yaffe, *JHEP* **0305** (2003) 051, hep-ph/0302165.
- [122] G. D. Moore and D. Teaney, *Phys. Rev.* **C71** (2005) 064904, hep-ph/0412346.
- [123] S. Caron-Huot and G. D. Moore, *JHEP* **0802** (2008) 081, arXiv:0801.2173.
- [124] A. Francis, O. Kaczmarek, M. Laine and J. Langelage, *PoS LATTICE2011* (2011) 202, arXiv:1109.3941.
- [125] D. Banerjee, S. Datta, R. Gavai and P. Majumdar, *Phys. Rev.* **D85** (2012) 014510, arXiv:1109.5738.
- [126] J. L. Albacete and C. Marquet, *Phys. Lett.* **B687** (2010) 174, arXiv:1001.1378.
- [127] J. Casalderrey-Solana and C. A. Salgado, *Acta Phys. Polon.* **B38** (2007) 3731, arXiv:0712.3443.
- [128] D. d’Enterria, *Jet quenching*, arXiv:0902.2011.
- [129] U. A. Wiedemann, *Jet Quenching in Heavy Ion Collisions*, arXiv:0908.2306.
- [130] A. Majumder and M. Van Leeuwen, *Prog. Part. Nucl. Phys.* **A66** (2011) 41, arXiv:1002.2206.
- [131] STAR Collaboration, J. Adams *et al.*, *Nucl. Phys.* **A757** (2005) 102, nucl-ex/0501009.
- [132] ATLAS Collaboration, G. Aad *et al.*, *Phys. Rev. Lett.* **105** (2010) 252303, arXiv:1011.6182.
- [133] CMS Collaboration, S. Chatrchyan *et al.*, *Phys. Rev.* **C84** (2011) 024906, arXiv:1102.1957.
- [134] R. Baier, Y. L. Dokshitzer, A. H. Mueller, S. Peigne and D. Schiff, *Nucl. Phys.* **B483** (1997) 291, hep-ph/9607355.
- [135] R. Baier, Y. L. Dokshitzer, A. H. Mueller, S. Peigne and D. Schiff, *Nucl. Phys.* **B484** (1997) 265, hep-ph/9608322.

- [136] B. Zakharov, *JETP Lett.* **63** (1996) 952, hep-ph/9607440.
- [137] B. Zakharov, *JETP Lett.* **65** (1997) 615, hep-ph/9704255.
- [138] M. Gyulassy and X.-n. Wang, *Nucl. Phys.* **B420** (1994) 583, nucl-th/9306003.
- [139] U. A. Wiedemann, *Nucl. Phys.* **B588** (2000) 303, hep-ph/0005129.
- [140] M. Gyulassy, P. Levai and I. Vitev, *Nucl. Phys.* **B594** (2001) 371, nucl-th/0006010.
- [141] X.-N. Wang and X.-f. Guo, *Nucl. Phys.* **A696** (2001) 788, hep-ph/0102230.
- [142] P. B. Arnold, G. D. Moore and L. G. Yaffe, *JHEP* **0206** (2002) 030, hep-ph/0204343.
- [143] C. A. Salgado and U. A. Wiedemann, *Phys. Rev. Lett.* **93** (2004) 042301, hep-ph/0310079.
- [144] P. B. Arnold, *Phys. Rev.* **D79** (2009) 065025, arXiv:0808.2767.
- [145] J. Casalderrey-Solana and E. Iancu, *JHEP* **1108** (2011) 015, arXiv:1105.1760.
- [146] T. Renk, *Phys. Rev.* **C85** (2012) 064908 arXiv:1202.4579.
- [147] Y. L. Dokshitzer, V. A. Khoze, A. H. Mueller and S. I. Troian, in *Basics of perturbative QCD*, Gif-sur-Yvette, France. Ed. Frontieres (1991) 274 p.
- [148] Y. Mehtar-Tani, C. A. Salgado and K. Tywoniuk, *Phys. Rev. Lett.* **106** (2011) 122002, arXiv:1009.2965.
- [149] Y. Mehtar-Tani, C. Salgado and K. Tywoniuk, *Phys. Lett.* **B707** (2012) 156, arXiv:1102.4317.
- [150] K. Zapp, J. Stachel and U. A. Wiedemann, *Phys. Rev. Lett.* **103** (2009) 152302, arXiv:0812.3888.
- [151] B. Schenke, C. Gale and S. Jeon, *Phys. Rev.* **C80** (2009) 054913, arXiv:0909.2037.
- [152] K. C. Zapp, J. Stachel and U. A. Wiedemann, *JHEP* **1107** (2011) 118, arXiv:1103.6252.
- [153] O. Aharony, S. S. Gubser, J. M. Maldacena, H. Ooguri and Y. Oz, *Phys. Rept.* **323** (2000) 183, hep-th/9905111.
- [154] D. T. Son and A. O. Starinets, *Ann. Rev. Nucl. Part. Sci.* **57** (2007) 95, arXiv:0704.0240.
- [155] E. Iancu, *Acta Phys. Polon.* **B39** (2008) 3213, arXiv:0812.0500.
- [156] S. S. Gubser and A. Karch, *Ann. Rev. Nucl. Part. Sci.* **59** (2009) 145, arXiv:0901.0935.
- [157] R. Baier, P. Romatschke, D. T. Son, A. O. Starinets and M. A. Stephanov, *JHEP* **0804** (2008) 100, arXiv:0712.2451.
- [158] V. E. Hubeny, S. Minwalla and M. Rangamani, *The fluid/gravity correspondence*, arXiv:1107.5780.
- [159] J. L. Albacete, Y. V. Kovchegov and A. Taliotis, *JHEP* **0905** (2009) 060, arXiv:0902.3046.
- [160] S. S. Gubser, S. S. Pufu and A. Yarom, *JHEP* **0911** (2009) 050, arXiv:0902.4062.
- [161] S. Lin and E. Shuryak, *Phys. Rev.* **D79** (2009) 124015, arXiv:0902.1508.
- [162] P. M. Chesler and L. G. Yaffe, *Phys. Rev. Lett.* **106** (2011) 021601, arXiv:1011.3562.
- [163] R. A. Janik and R. B. Peschanski, *Phys. Rev.* **D73** (2006) 045013, hep-th/0512162.
- [164] G. Beuf, M. P. Heller, R. A. Janik and R. Peschanski, *JHEP* **0910** (2009) 043, arXiv:0906.4423.
- [165] Y. Hatta, E. Iancu and A. Mueller, *JHEP* **0805** (2008) 037, arXiv:0803.2481.
- [166] S. S. Gubser, D. R. Gulotta, S. S. Pufu and F. D. Rocha, *JHEP* **0810** (2008) 052, arXiv:0803.1470.
- [167] P. M. Chesler, K. Jensen, A. Karch and L. G. Yaffe, *Phys. Rev.* **D79** (2009) 125015, arXiv:0810.1985.
- [168] P. Arnold and D. Vaman, *JHEP* **1010** (2010) 099, arXiv:1008.4023.
- [169] F. Dominguez, C. Marquet, A. Mueller, B. Wu, and B.-W. Xiao, *Nucl. Phys.* **A811** (2008) 197, arXiv:0803.3234.
- [170] Y. Hatta, E. Iancu and A. Mueller, *JHEP* **0801** (2008) 063, arXiv:0710.5297.

- [171] C. Marquet and T. Renk, *Phys. Lett.* **B685** (2010) 270, [arXiv:0908.0880](#).

Practical Statistics for the LHC

K. Cranmer

Center for Cosmology and Particle Physics, New York University, USA

Abstract

This document is a pedagogical introduction to statistics for particle physics. Emphasis is placed on the terminology, concepts, and methods being used at the Large Hadron Collider. The document addresses both the statistical tests applied to a model of the data and the modeling itself.

1 Introduction

It is often said that the language of science is mathematics. It could well be said that the language of experimental science is statistics. It is through statistical concepts that we quantify the correspondence between theoretical predictions and experimental observations. While the statistical analysis of the data is often treated as a final subsidiary step to an experimental physics result, a more direct approach would be quite the opposite. In fact, thinking through the requirements for a robust statistical statement is an excellent way to organize an analysis strategy.

In these lecture notes¹ I will devote significant attention to the strategies used in high-energy physics for developing a statistical model of the data. This modeling stage is where you inject your understanding of the physics. I like to think of the modeling stage in terms of a conversation. When your colleague asks you over lunch to explain your analysis, you tell a story. It is a story about the signal and the backgrounds – are they estimated using Monte Carlo simulations, a side-band, or some data-driven technique? Is the analysis based on counting events or do you use some discriminating variable, like an invariant mass or perhaps the output of a multivariate discriminant? What are the dominant uncertainties in the rate of signal and background events and how do you estimate them? What are the dominant uncertainties in the shape of the distributions and how do you estimate them? The answer to these questions forms a *scientific narrative*; the more convincing this narrative is the more convincing your analysis strategy is. The statistical model is the mathematical representation of this narrative and you should strive for it to be as faithful a representation as possible.

Once you have constructed a statistical model of the data, the actual statistical procedures should be relatively straight forward. In particular, the statistical tests can be written for a generic statistical model without knowledge of the physics behind the model. The goal of the RooStats project was precisely to provide statistical tools based on an arbitrary statistical model implemented with the RooFit modeling language. While the formalism for the statistical procedures can be somewhat involved, the logical justification for the procedures is based on a number of abstract properties for the statistical procedures. One can follow the logical argument without worrying about the detailed mathematical proofs that the procedures have the required properties. Within the last five years there has been a significant advance in the field's understanding of certain statistical procedures, which has led to some commonalities in the statistical recommendations by the major LHC experiments. I will review some of the most common statistical procedures and their logical justification.

¹These notes borrow significantly from other documents that I am writing contemporaneously; specifically Ref. [1], documentation for HistFactory [2] and the ATLAS Higgs combination.

2 Conceptual building blocks for modeling

2.1 Probability densities and the likelihood function

This section specifies my notations and conventions, which I have chosen with some care.² Our statistical claims will be based on the outcome of an experiment. When discussing frequentist probabilities, one must consider ensembles of experiments, which may either be real, based on computer simulations, or mathematical abstraction.

Figure 1 establishes a hierarchy that is fairly general for the context of high-energy physics. Imagine the search for the Higgs boson, in which the search is composed of several “channels” indexed by c . Here a channel is defined by its associated event selection criteria, not an underlying physical process. In addition to the number of selected events, n_c , each channel may make use of some other measured quantity, x_c , such as the invariant mass of the candidate Higgs boson. The quantities will be called “observables” and will be written in roman letters e.g. x_c . The notation is chosen to make manifest that the observable x is frequentist in nature. Replication of the experiment many times will result in different values of x and this ensemble gives rise to a *probability density function* (PDF) of x , written $f(x)$, which has the important property that it is normalized to unity

$$\int f(x) dx = 1 .$$

In the case of discrete quantities, such as the number of events satisfying some event selection, the integral is replaced by a sum. Often one considers a parametric family of PDFs

$$f(x|\alpha) ,$$

read “ f of x given α ” and, henceforth, referred to as a *probability model* or just *model*. The parameters of the model typically represent parameters of a physical theory or an unknown property of the detector’s response. The parameters are not frequentist in nature, thus any probability statement associated with α is Bayesian.³ In order to make their lack of frequentist interpretation manifest, model parameters will be written in greek letters, e.g.: μ, θ, α, ν .⁴ From the full set of parameters, one is typically only interested in a few: the *parameters of interest*. The remaining parameters are referred to as *nuisance parameters*, as we must account for them even though we are not interested in them directly.

While $f(x)$ describes the probability density for the observable x for a single event, we also need to describe the probability density for a dataset with many events, $\mathcal{D} = \{x_1, \dots, x_n\}$. If we consider the events as independently drawn from the same underlying distribution, then clearly the probability density is just a product of densities for each event. However, if we have a prediction that the total number of events expected, call it ν , then we should also include the overall Poisson probability for observing n events given ν expected. Thus, we arrive at what statisticians call a marked Poisson model,

$$\mathbf{f}(\mathcal{D}|\nu, \alpha) = \text{Pois}(n|\nu) \prod_{e=1}^n f(x_e|\alpha) , \quad (1)$$

where I use a bold $\mathbf{f}$ to distinguish it from the individual event probability density $f(x)$. In practice, the expectation is often parametrized as well and some parameters simultaneously modify the expected rate and shape, thus we can write $\nu \rightarrow \nu(\alpha)$. In RooFit both f and $\mathbf{f}$ are implemented with a RooAbsPdf; where `RooAbsPdf::getVal(x)` always provides the value of $f(x)$ and depending on `RooAbsPdf::extendMode()` the value of ν is accessed via `RooAbsPdf::expectedEvents()`.

²As in the case of relativity, notational conventions can make some properties of expressions manifest and help identify mistakes. For example, $g_{\mu\nu}x^\mu y^\nu$ is manifestly Lorentz invariant and $x^\mu + y_\nu$ is manifestly wrong.

³Note, one can define a conditional distribution $f(x|y)$ when the joint distribution $f(x, y)$ is defined in a frequentist sense.

⁴While it is common to write s and b for the number of expected signal and background, these are parameters *not* observables, so I will write ν_S and ν_B . This is one of few notational differences to Ref. [1].

The *likelihood function* $L(\alpha)$ is numerically equivalent to $f(x|\alpha)$ with x fixed – or $\mathbf{f}(\mathcal{D}|\alpha)$ with $\mathcal{D}$ fixed. The likelihood function should not be interpreted as a probability density for α . In particular, the likelihood function does not have the property that it normalizes to unity

$$\int L(\alpha) d\alpha = 1. \quad \text{Not True!}$$

It is common to work with the log-likelihood (or negative log-likelihood) function. In the case of a marked Poisson, we have what is commonly referred to as an extended maximum likelihood [3]

$$-\ln L(\alpha) = \underbrace{\nu(\alpha) - n \ln \nu(\alpha)}_{\text{extended term}} - \sum_{e=1}^n \ln f(x_e) + \underbrace{\ln n!}_{\text{constant}}.$$

To reiterate the terminology, *probability density function* refers to the value of f as a function of x given a fixed value of α ; *likelihood function* refers to the value of f as a function of α given a fixed value of x ; and *model* refers to the full structure of $f(x|\alpha)$.

Probability models can be constructed to simultaneously describe several channels, that is several disjoint regions of the data defined by the associated selection criteria. I will use e as the index over events and c as the index over channels. Thus, the number of events in the c^{th} channel is n_c and the value of the e^{th} event in the c^{th} channel is x_{ce} . In this context, the data is a collection of smaller datasets: $\mathcal{D}_{\text{sim}} = \{\mathcal{D}_1, \dots, \mathcal{D}_{c_{\text{max}}}\} = \{\{x_{c=1,e=1} \dots x_{c=1,e=n_c}\}, \dots, \{x_{c=c_{\text{max}},e=1} \dots x_{c=c_{\text{max}},e=n_{c_{\text{max}}}}\}\}$. The index c , in RooFit, is referred to as a RooCategory and it is used to inside the dataset to differentiate events associated to different channels or categories. The class RooSimultaneous associates the dataset $\mathcal{D}_c$ with the corresponding marked Poisson model. The key point here is that there are now multiple Poisson terms. Thus we can write the combined (or simultaneous) model

$$\mathbf{f}_{\text{sim}}(\mathcal{D}_{\text{sim}}|\alpha) = \prod_{c \in \text{channels}} \left[\text{Pois}(n_c|\nu(\alpha)) \prod_{e=1}^{n_c} f(x_{ce}|\alpha) \right], \quad (2)$$

remembering that the symbol product over channels has implications for the structure of the dataset.

2.2 Auxiliary measurements

Auxiliary measurements or control regions can be used to estimate or reduce the effect of systematic uncertainties. The signal region and control region are not fundamentally different. In the language that we are using here, they are just two different channels.

A common example is a simple counting experiment with an uncertain background. In the frequentist way of thinking, the true, unknown background in the signal region is a nuisance parameter, which I will denote ν_B .⁵ If we call the true, unknown signal rate ν_S and the number of events in the signal region n_{SR} then we can write the model $\text{Pois}(n_{\text{SR}}|\nu_S + \nu_B)$. As long as ν_B is a free parameter, there is no ability to make any useful inference about ν_S . Often we have some estimate for the background, which may have come from some control sample with n_{CR} events. If the control sample has no signal contamination and is populated by the same background processes as the signal region, then we can write $\text{Pois}(n_{\text{CR}}|\tau\nu_B)$, where n_{CR} is the number of events in the control region and τ is a factor used to extrapolate the background from the signal region to the control region. Thus the total probability model can be written $\mathbf{f}_{\text{sim}}(n_{\text{SR}}, n_{\text{CR}}|\nu_S, \nu_B) = \text{Pois}(n_{\text{SR}}|\nu_S + \nu_B) \cdot \text{Pois}(n_{\text{CR}}|\tau\nu_B)$. This is a special case of Eq. 2 and is often referred to as the ‘on/off’ problem [4].

Based on the control region alone, one would estimate (or ‘measure’) $\nu_B = n_{\text{CR}}/\tau$. Intuitively the estimate comes with an ‘uncertainty’ of $\sqrt{n_{\text{CR}}}/\tau$. We will make these points more precise in Section 3.1,

⁵Note, you can think of a counting experiment in the context of Eq. 1 with $f(x) = 1$, thus it reduces to just the Poisson term.

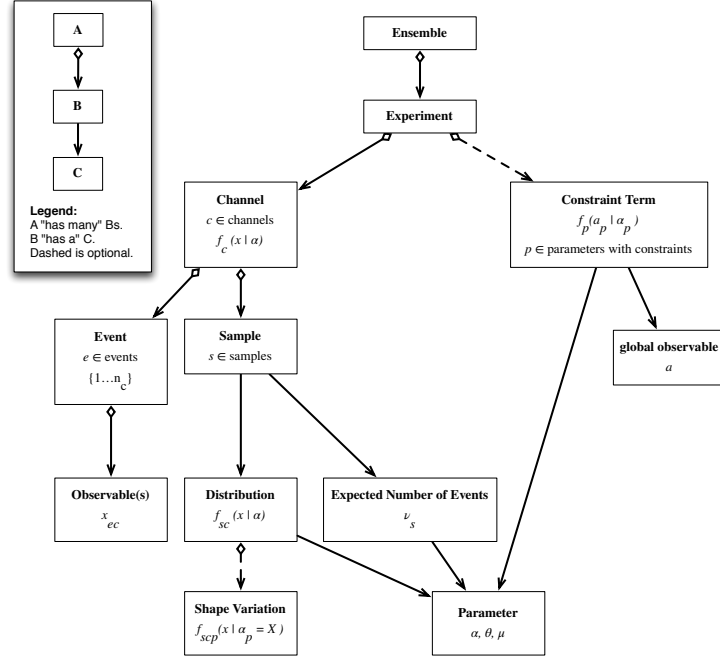


Fig. 1: A schematic diagram of the logical structure of a typical particle physics probability model and dataset structures.

but the important lesson here is that we can use auxiliary measurements (i.e. n_{CR}) to describe our uncertainty on the nuisance parameter ν_B statistically. Furthermore, we have formed a statistical model that can be treated in a frequentist formalism – meaning that if we repeat the experiment many times n_{CR} will vary and so will the estimate of ν_B . It is common to say that auxiliary measurements ‘constrain’ the nuisance parameters. In principle the auxiliary measurements can be every bit as complex as the main signal region, and there is no formal distinction between the various channels.

The use of auxiliary measurements is not restricted to estimating rates as in the case of the on/off problem above. One can also use auxiliary measurements to constrain other parameters of the model. To do so, one must relate the effect of some common parameter α_p in multiple channels (i.e. the signal region and a control regions). This is implicit in Eq. 2.

2.3 Frequentist and Bayesian reasoning

The intuitive interpretation of measurement of ν_B to be $n_{\text{CR}}/\tau \pm \sqrt{n_{\text{CR}}}/\tau$ is that the parameter ν_B has a distribution centered around n_{CR}/τ with a width of $\sqrt{n_{\text{CR}}}/\tau$. With some practice you will be able to immediately identify this type of reasoning as Bayesian. It is manifestly Bayesian because we are referring to the probability distribution of a parameter. The frequentist notion of probability of an event is defined as the limit of its relative frequency in a large number of trials. The large number of trials is referred to as an ensemble. In particle physics the ensemble is formed conceptually by repeating the experiment many times. The true values of the parameters, on the other hand, are states of nature, not the outcome of an experiment. The true mass of the Z boson has no frequentist probability distribution. The existence or non-existence of the Higgs boson has no frequentist probability associated with it. There is a sense in which one can talk about the probability of parameters, which follows from Bayes’s theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (3)$$

Bayes's theorem is a theorem, so there's no debating it. It is not the case that Frequentists dispute whether Bayes's theorem is true. The debate is whether the necessary probabilities exist in the first place. If one can define the joint probability $P(A, B)$ in a frequentist way, then a Frequentist is perfectly happy using Bayes theorem. Thus, the debate starts at the very definition of probability.

The Bayesian definition of probability clearly can't be based on relative frequency. Instead, it is based on a degree of belief. Formally, the probability needs to satisfy Kolmogorov's axioms for probability, which both the frequentist and Bayesian definitions of probability do. One can quantify degree of belief through betting odds, thus Bayesian probabilities can be assigned to hypotheses on states of nature. In practice human's bets are generally not 'coherent', thus this way of quantifying probabilities may not satisfy the Kolmogorov axioms.

Moving past the philosophy and accepting the Bayesian procedure at face value, the practical consequence is that one must supply prior probabilities for various parameter values and/or hypotheses. In particular, to interpret our example measurement of n_{CR} as implying a probability distribution for ν_B we would write

$$\pi(\nu_B|n_{\text{CR}}) \propto f(n_{\text{CR}}|\nu_B)\eta(\nu_B), \quad (4)$$

where $\pi(\nu_B|n_{\text{CR}})$ is called the *posterior* probability density, $f(n_{\text{CR}}|\nu_B)$ is the likelihood function, and $\eta(\nu_B)$ is the *prior* probability. Here I have suppressed the somewhat curious term $P(n_{\text{CR}})$, which can be thought of as a normalization constant and is also referred to as the *evidence*. The main point here is that one can only invert 'the probability of n_{CR} given ν_B ' to be 'the probability of ν_B given n_{CR} ' if one supplies a prior. Humans are very susceptible to performing this logical inversion accidentally, typically with a uniform prior on ν_B . Furthermore, the prior degree of belief cannot be derived in an objective way. There are several formal rules for providing a prior based on formal rules (e.g. the Jefferey's prior and Reference priors), though these are not accurately described as representing a degree of belief. Thus, that style of Bayesian analysis is often referred to as objective Bayesian analysis.

Some useful and amusing quotes on Bayesian and Frequentist reasoning:

*"Using Bayes's theorem doesn't make you a Bayesian, **always** using Bayes's theorem makes you a Bayesian."* – unknown

"Bayesians address the questions everyone is interested in by using assumptions that no one believes. Frequentist use impeccable logic to deal with an issue that is of no interest to anyone." – Louis Lyons

2.4 Consistent Bayesian and Frequentist modeling of constraint terms

Often a detailed probability model for an auxiliary measurement are not included directly into the model. If the model for the auxiliary measurement were available, it could and should be included as an additional channel as described in Section 2.2. The more common situation for background and systematic uncertainties only has an estimate, "central value", or best guess for a parameter α_p and some notion of uncertainty on this estimate. In this case one typically resorts to including idealized terms into the likelihood function, here referred to as "constraint terms", as surrogates for a more detailed model of the auxiliary measurement. I will denote this estimate for the parameters as a_p , to make it manifestly frequentist in nature. In this case there is a single measurement of a_p per experiment, thus it is referred to as a "global observable" in RooStats. The treatment of constraint terms is somewhat *ad hoc* and discussed in more detail in Section 4.1.6. I make it a point to write constraint terms in a manifestly frequentist form $f(a_p|\alpha_p)$.

Probabilities on parameters are legitimate constructs in a Bayesian setting, though they will always rely on a prior. In order to distinguish Bayesian PDFs from frequentist ones, greek letters will be used for their distributions. For instance, a generic Bayesian PDF might be written $\pi(\alpha)$. In the context of

a main measurement, one might have a prior for α_p based on some estimate a_p . In this case, the prior $\pi(\alpha_p)$ is really a posterior from some previous measurement. It is desirable to write with the help of Bayes theorem

$$\pi(\alpha_p|a_p) \propto L(\alpha_p)\eta(\alpha_p) = f(a_p|\alpha_p)\eta(\alpha_p) , \quad (5)$$

where $\eta(\alpha_p)$ is some more fundamental prior.⁶ By taking the time to undo the Bayesian reasoning into an objective PDF or likelihood and a prior we are able to write a model that can be used in a frequentist context. Within RooStats, the care is taken to separately track the frequentist component and the prior; this is achieved with the `ModelConfig` class.

If one can identify what auxiliary measurements were performed to provide the estimate of α_p and its uncertainty, then it is not a logical fallacy to approximate it with a constraint term, it is simply a convenience. However, not all uncertainties that we deal result from auxiliary measurements. In particular, some theoretical uncertainties are not statistical in nature. For example, uncertainty associated with the choice of renormalization and factorization scales and missing higher-order corrections in a theoretical calculation are not statistical. Uncertainties from parton density functions are a bit of a hybrid as they are derived from data but require theoretical inputs and make various modeling assumptions. In a Bayesian setting there is no problem with including a prior on the parameters associated to theoretical uncertainties. In contrast, in a formal frequentist setting, one should not include constraint terms on theoretical uncertainties that lack a frequentist interpretation. That leads to a very cumbersome presentation of results, since formally the results should be shown as a function of the uncertain parameter. In practice, the groups often read Eq. 5 to arrive at an effective frequentist constraint term.

I will denote the set of parameters with constraint terms as $\mathbb{S}$ and the global observables $\mathcal{G} = \{a_p\}$ with $p \in \mathbb{S}$. By including the constraint terms explicitly (instead of implicitly as an additional channel) we arrive at the total probability model, which we will not need to generalize any further:

$$\mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{sim}}, \mathcal{G}|\alpha) = \prod_{c \in \text{channels}} \left[\text{Pois}(n_c|\nu_c(\alpha)) \prod_{e=1}^{n_c} f_c(x_{ce}|\alpha) \right] \cdot \prod_{p \in \mathbb{S}} f_p(a_p|\alpha_p) . \quad (6)$$

3 Physics questions formulated in statistical language

3.1 Measurement as parameter estimation

One of the most common tasks of the working physicist is to estimate some model parameter. We do it so often, that we often don't realize it. For instance, the sample mean $\bar{x} = \sum_{e=1}^n x_e/n$ is an estimate for the mean, μ , of a Gaussian probability density $f(x|\mu, \sigma) = \text{Gauss}(x|\mu, \sigma)$. More generally, an *estimator* $\hat{\alpha}(\mathcal{D})$ is some function of the data and its value is used to estimate the true value of some parameter α . There are various abstract properties such as variance, bias, consistency, efficiency, robustness, etc [5]. The bias of an estimator is defined as $B(\hat{\alpha}) = E[\hat{\alpha}] - \alpha$, where E means the expectation value of $E[\hat{\alpha}] = \int \hat{\alpha}(x)f(x)dx$ or the probability-weighted average. Clearly one would like an unbiased estimator. The variance of an estimator is defined as $\text{var}[\hat{\alpha}] = E[(\alpha - E[\hat{\alpha}])^2]$; and clearly one would like an estimator with the minimum variance. Unfortunately, there is a tradeoff between bias and variance. Physicists tend to be allergic to biased estimators, and within the class of unbiased estimators, there is a well defined minimum variance bound referred to as the Cramér–Rao bound (that is the inverse of the Fisher information, which we will refer to again later).

The most widely used estimator in physics is the maximum likelihood estimator (MLE). It is defined as the value of α which maximizes the likelihood function $L(\alpha)$. Equivalently this value, $\hat{\alpha}$, maximizes $\log L(\alpha)$ and minimizes $-\log L(\alpha)$. The most common tool for finding the maximum likelihood estimator is `Minuit`, which conventionally minimizes $-\log L(\alpha)$ (or any other function) [6]. The jargon is that one ‘fits’ the function and the maximum likelihood estimate is the ‘best fit value’.

⁶Glen Cowan has referred to this more fundamental prior as an ‘urprior’, which is based on the German use of ‘ur’ for forming words with the sense of ‘proto-, primitive, original’.

When one has a multi-parameter likelihood function $L(\boldsymbol{\alpha})$, then the situation is slightly more complicated. The maximum likelihood estimate for the full parameter list, $\hat{\boldsymbol{\alpha}}$, is clearly defined. The various components $\hat{\alpha}_p$ are referred to as the *unconditional maximum likelihood estimates*. In the physics jargon, one says all the parameters are ‘floating’. One can also ask about maximum likelihood estimate of α_p is with some other parameters α_o fixed; this is called the *conditional maximum likelihood estimate* and is denoted $\hat{\alpha}_p(\alpha_o)$. These are important quantities for defining the profile likelihood ratio, which we will discuss in more detail later. The concept of variance of the estimates is also generalized to the covariance matrix $\text{cov}[\alpha_p, \alpha_{p'}] = E[(\hat{\alpha}_p - \alpha_p)(\hat{\alpha}_{p'} - \alpha_{p'})]$ and is often denoted $\Sigma_{pp'}$. Note, the diagonal elements of the covariance matrix are the same as the variance for the individual parameters, i.e. $\text{cov}[\alpha_p, \alpha_p] = \text{var}[\alpha_p]$.

In the case of a Poisson model $\text{Pois}(n|\nu)$ the maximum likelihood estimate of ν is simply $\hat{\nu} = n$. Thus, it follows that the variance of the estimator is $\text{var}[\hat{\nu}] = \text{var}[n] = \nu$. Thus if the true rate is ν one expects to find estimates $\hat{\nu}$ with a characteristic spread around ν ; it is in this sense that the measurement has an estimate has some uncertainty or ‘error’ of $\sqrt{n}$. We will make this statement of uncertainty more precise when we discuss frequentist confidence intervals.

When the number of events is large, the distribution of maximum likelihood estimates approaches a Gaussian or normal distribution.⁷ This does not depend on the PDF $f(x)$ having a Gaussian form. For small samples this isn’t the case, but this limiting distribution is often referred to as an *asymptotic distribution*. Furthermore, under most circumstances in particle physics, the maximum likelihood estimate approaches the minimum variance or Cramér–Rao bound. In particular, the inverse of the covariance matrix for the estimates is asymptotically given by

$$\Sigma_{pp'}^{-1}(\boldsymbol{\alpha}) = E \left[- \frac{\partial^2 \log f(x|\boldsymbol{\alpha})}{\partial \alpha_p \partial \alpha_{p'}} \middle| \boldsymbol{\alpha} \right], \quad (7)$$

where I have written explicitly that the expectation, and thus the covariance matrix itself, depend on the true value $\boldsymbol{\alpha}$. The right side of Eq. 7 is called the (expected) Fisher information matrix. Remember that the expectation involves an integral over the observables. Since that integral is difficult to perform in general, one often uses the observed Fisher information matrix to approximate the variance of the estimator by simply taking the matrix of second derivatives based on the observed data

$$\tilde{\Sigma}_{pp'}^{-1}(\boldsymbol{\alpha}) = - \frac{\partial^2 \log L(\boldsymbol{\alpha})}{\partial \alpha_p \partial \alpha_{p'}}. \quad (8)$$

This is what Minuit’s Hesse algorithm⁸ calculates to estimate the covariance matrix of the parameters.

3.2 Discovery as hypothesis tests

Let us examine the statistical statement associated to the claim of discovery for new physics. Typically, new physics searches are looking for a signal that is additive on top of the background, though in some cases there are interference effects that need to be taken into account and one cannot really talk about ‘signal’ and ‘background’ in any meaningful way. Discovery is formulated in terms of a hypothesis test where the background-only hypothesis plays the role of the null hypothesis and the signal-plus-background hypothesis plays the role of the alternative. Roughly speaking, the claim of discovery is a statement that the data are incompatible with the background-only hypothesis. Consider the simplest scenario where one is counting events in the signal region, n_{SR} and expects ν_B events from background and ν_S events from the putative signal. Then we have the following hypotheses:

⁷There are various conditions that must be met for this to be true, but skip the fine print in these lectures. There are two conditions that are most often violated in particle physics, which will be addressed later.

⁸The matrix is called the Hessian, hence the name.

symbol	statistical name	physics name	probability model
H_0	null hypothesis	background-only	$\text{Pois}(n_{SR} \nu_B)$
H_1	alternate hypothesis	signal-plus-background	$\text{Pois}(n_{SR} \nu_S + \nu_B)$

In this simple example it's fairly obvious that evidence for a signal shows up as an excess of events and a reasonable way to quantify the compatibility of the observed data n_{CR}^0 and the null hypothesis is to calculate the probability that the background-only would produce at least this many events; the p -value

$$p = \sum_{n=n_{SR}^0}^{\infty} \text{Pois}(n|\nu_B) . \quad (9)$$

If this p -value is very small, then one might choose to reject the null hypothesis.

Note, the p -value is *not* to be interpreted as the probability of the null hypothesis given the data – that is a manifestly Bayesian statement. Instead, the p -value is a statement about the probability to have obtained data with a certain property assuming the null hypothesis.

How do we generalize this to more complicated situations? There were really two ingredients in our simple example. The first was the proposal that we would reject the null hypothesis based on the probability for it to produce data at least as extreme as the observed data. The second ingredient was the prescription for what is meant by more discrepant; in this case the possible observations are ordered according to increasing n_{SR} . One could imagine using difference between observed and expected, $n_{SR} - \nu_B$, as the measure of discrepancy. In general, a function that maps the data to a single real number is called a *test statistic*: $T(\mathcal{D}) \rightarrow \mathbb{R}$. How does one choose from the infinite number of test statistics?

Neyman and Pearson provided a framework for hypothesis testing that addresses the choice of the test statistic. This setup treats the null and the alternate hypotheses in an asymmetric way. First, one defines an *acceptance region* in terms of a test statistic, such that if $T(\mathcal{D}) < k_\alpha$ one accepts the null hypothesis. One can think of the $T(\mathcal{D}) = k_\alpha$ as defining a contour in the space of the data, which is the boundary of this acceptance region. Next, one defines the *size of the test*, α ,⁹ as the probability the null hypothesis will be rejected when it is true (a so-called Type-I error). This is equivalent to the probability under the null hypothesis that the data will not be found in this acceptance region, i.e. $\alpha = P(T(\mathcal{D}) \geq k_\alpha | H_0)$. Note, it is now clear why there is a subscript on k_α , since the contour level is related to the size of the test. In contrast, if one accepts the null hypothesis when the alternate is true, it is called a Type-II error. The probability to commit a Type-II error is denoted as β and it is given by $\beta = P(T(\mathcal{D}) < k_\alpha | H_1)$. One calls $1 - \beta$ the *power* of the test. With these definitions in place, one looks for a test statistic that maximizes the power of the test for a fixed test size. This is a problem for the calculus of variations, and sounds like it might be very difficult for complicated probability models.

It turns out that in the case of two simple hypotheses (probability models without any parameters), there is a simple solution! In particular, the test statistic leading to the most powerful test is given by the likelihood ratio $T_{NP}(\mathcal{D}) = f(\mathcal{D}|H_1)/f(\mathcal{D}|H_0)$. This result is referred to as the Neyman–Pearson lemma, and I will give an informal proof. We will prove this by considering a small variation to the acceptance region defined by the likelihood ratio. The solid red contour in Fig. 2 represents the rejection region (the complement to the acceptance region) based on the likelihood ratio and the dashed blue contour represents a small perturbation. If we can say that any variation to the likelihood ratio has less power, then we will have proved the Neyman–Pearson lemma. The variation adds (the left, blue wedge) and removes (the right, red wedge) rejection regions. Because the Neyman–Pearson setup requires that both tests have the same size, we know that the probability for the data to be found in the two wedges must be the same under the null hypothesis. Because the two regions are on opposite sides of the contour defined by $f(\mathcal{D}|H_1)/f(\mathcal{D}|H_0)$, then we know that the data is less likely to be found in the small region that we added than the small region we subtracted assuming the alternate hypothesis. In other words, there is

⁹Note, α is the conventional notation for the size of the test, and has nothing to do with a model parameter in Eq. 2.

less probability to reject the null when the alternate is true; thus the test based on the new contour is less powerful.

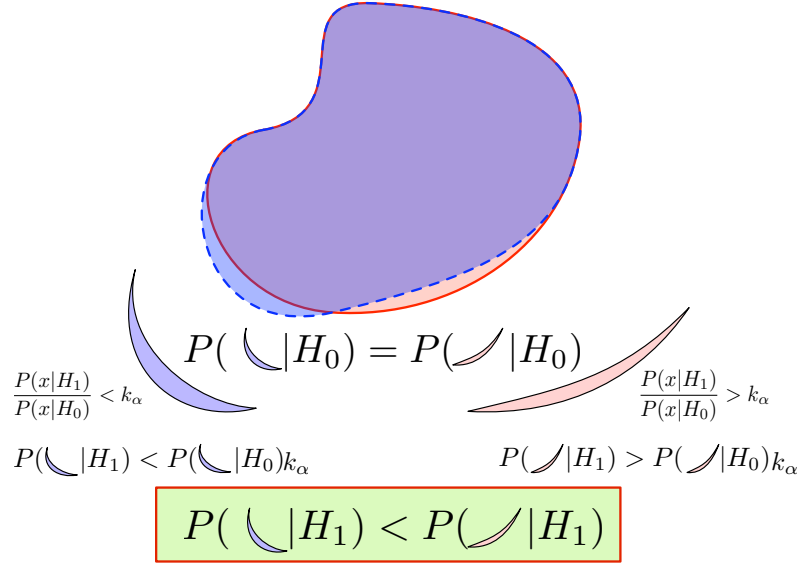


Fig. 2: A graphical proof of the Neyman–Pearson lemma.

How does this generalize for our most general model in Eq. 6 with many free parameters? First one must still define the null and the alternate hypotheses. Typically is done by saying some parameters – the parameters of interest α_{poi} – take on specific values takes on a particular value for the signal-plus-background hypothesis and a different value for the background-only hypothesis. For instance, the signal production cross-section might be singled out as the *parameter of interest* and it would take on the value of zero for the background-only and some reference value for the signal-plus-background. The remainder of the parameters are called the *nuisance parameters* α_{nuis} . Unfortunately, there is no equivalent to the Neyman–Pearson lemma for models with several free parameters – so called, composite models. Nevertheless, there is a natural generalization based on the profile likelihood ratio.

Remembering that the test statistic T is a real-valued function of the data, then any particular probability model $\mathbf{f}_{\text{tot}}(\mathcal{D}|\alpha)$ implies a distribution for the test statistic $f(T|\alpha)$. Note, the distribution for the test statistic depends on the value of α . Below we will discuss how one constructs this distribution, but lets take it as given for the time being. Once one has the distribution, then one can calculate the p -value is given by

$$p(\alpha) = \int_{T_0}^{\infty} f(T|\alpha) dT = \int \mathbf{f}(\mathcal{D}|\alpha) \theta(T(\mathcal{D}) - T_0) d\mathcal{D} = P(T \geq T_0|\alpha), \quad (10)$$

where T_0 is the value of the test statistic based on the observed data and $\theta(\cdot)$ is the Heaviside function.¹⁰ Usually the p -value is just written as p , but I have written it as $p(\alpha)$ to make its α -dependence explicit.

Given that the p -value depends on α , how does one decide to accept or reject the null hypothesis? Remembering that α_{poi} takes on a specific value for the null hypothesis, we are worried about how the p -value changes as a function of the nuisance parameters. It is natural to say that one should not reject the null hypothesis if the p -value is larger than the size of the test *for any value of the nuisance parameters*. Thus, in a frequentist approach one should either present p -value explicitly as a function of α_{nuis} or take

¹⁰The integral $\int d\mathcal{D}$ is a bit unusual for a marked Poisson model, because it involves both a sum over the number of events and an integral over the values of x_e for each of those events.

its maximal (or supremum) value

$$p_{\text{sup}}(\alpha_{\text{poi}}) = \sup_{\alpha_{\text{nuis}}} p(\alpha_{\text{nuis}}) . \quad (11)$$

As a final note it is worth mentioning that the size of the test, which serves as the threshold for rejecting the null hypothesis, is purely conventional. In most sciences conventional choices of the size are 10%, 5%, or 1%. In particle physics, our conventional threshold for discovery is the infamous 5σ criterion – which is a conventional way to refer to $\alpha = 2.87 \cdot 10^{-7}$. This is an incredibly small rate of Type-I error, reflecting that claiming the discovery of new physics would be a monumental statement. The origin of the 5σ criterion has its roots in the fact that traditionally we lacked the tools to properly incorporate systematics, we fear that there are systematics that may not be fully under control, and we perform many searches for new physics and thus we have many chances to reject the background-only hypothesis. We will return to this in the discussion of the look-elsewhere effect.

3.3 Excluded and allowed regions as confidence intervals

Often we consider a new physics model that is parametrized by theoretical parameters. For instance, the mass or coupling of a new particle. In that case we typically want to ask what values of these theoretical parameters are allowed or excluded given available data. Figure 3 shows two examples. Figure 3(a) shows an example with $\alpha_{\text{poi}} = (\sigma/\sigma_{SM}, M_H)$, where σ/σ_{SM} is the ratio of the production cross-section for the Higgs boson with respect to its prediction in the standard model and M_H is the unknown Higgs mass parameter in the standard model. All the parameter points above the solid black curve correspond to scenarios for the Higgs boson that are considered ‘excluded at the 95% confidence level’. Figure 3(b) shows an example with $\alpha_{\text{poi}} = (m_W, m_t)$ where m_W is the mass of the W -boson and m_t is the mass of the top quark. We have discovered the W -boson and the top quark and measured their masses. The blue ellipse ‘is the 68% confidence level contour’ and all the parameter points inside it are considered ‘consistent with data at the 1σ level’. What is the precise meaning of these statements?

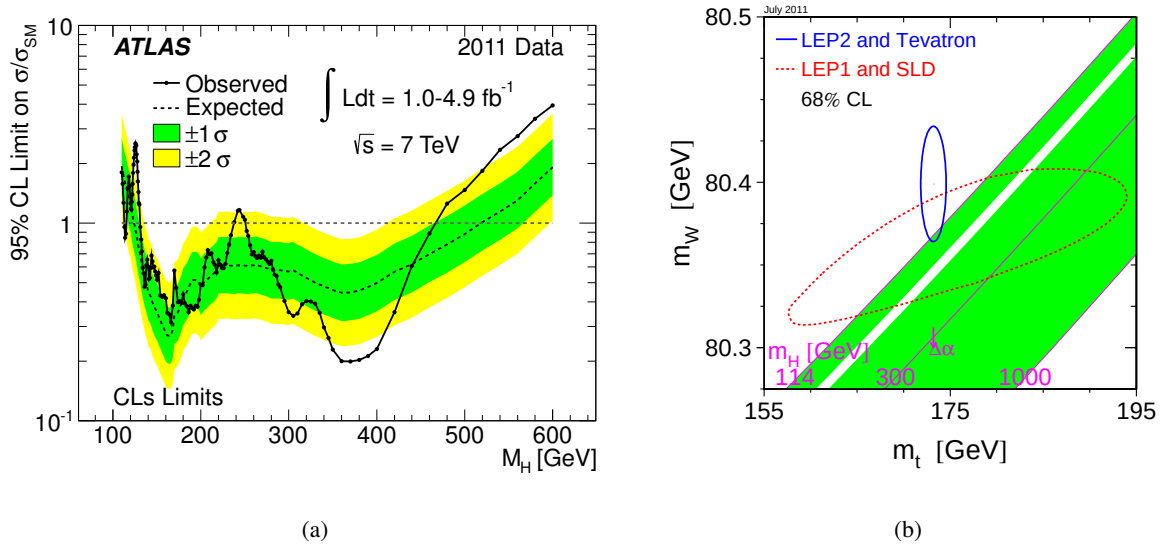


Fig. 3: Two examples of confidence intervals.

In a frequentist setting, these allowed regions are called *confidence intervals* or *confidence regions*, and the parameter points outside them are considered excluded. Associated with a confidence interval

is a confidence level, i.e. the 95% and 68% confidence level in the two examples. If we repeat the experiments and obtain different data, then these confidence intervals will change. It is useful to think of the confidence intervals as being random in the same way the data are random. The defining property of a 95% confidence interval is that it *covers* the true value 95% of the time.

How can one possibly construct a confidence interval has the desired property, that it *covers* the true value with a specified probability, given that we don't know the true value? The procedure for building confidence intervals is called the Neyman Construction [7], and it is based on 'inverting' a series of hypothesis tests (as described in Section 3.2). In particular, for each value of α in the parameter space one performs a hypothesis test based on some test statistic where the null hypothesis is α . Note, that in this context, the null hypothesis is changing for each test and generally is not the background-only. If one wants a 95% confidence interval, then one constructs a series of hypothesis test with a size of 5%. The confidence interval $I(\mathcal{D})$ is constructed by taking the set of parameter points where the null hypothesis is accepted.

$$I(\mathcal{D}) = \{\alpha \mid P(T(\mathcal{D}) > k_\alpha \mid \alpha) < \alpha\} , \quad (12)$$

where the final α and the subscript k_α refer to the size of the test. Since a hypothesis test with a size of 5% should accept the null hypothesis 95% of the time if it is true, confidence intervals constructed in this way satisfy the defining property. This same property is usually formulated in terms of *coverage*. Coverage is the probability that the interval will contain (cover) the parameter α when it is true,

$$\text{coverage}(\alpha) = P(\alpha \in I \mid \alpha) . \quad (13)$$

The equation above can easily be misinterpreted as the probability the parameter is in a fixed interval I ; but one must remember that in evaluating the probability above the data $\mathcal{D}$, and, thus, the corresponding intervals produced by the procedure $I(\mathcal{D})$, are the random quantities. Note, that coverage is a property that can be quantified for any procedure that produces the confidence intervals I . Intervals produced using the Neyman Construction procedure are said to "cover by construction"; however, one can consider alternative procedures that may either under-cover or over-cover. Undercoverage means that $P(\alpha \in I \mid \alpha)$ is smaller than desired and over-coverage means that $P(\alpha \in I \mid \alpha)$ is larger than desired. Note that in general coverage depends on the assumed true value α .

Since one typically is only interested in forming confidence intervals on the parameters of interest, then one could use the supremum p -value of Eq. 11. This procedure ensures that the coverage is at least the desired level, though for some values of α it may over-cover (perhaps significantly). This procedure, which I call the 'full construction', is also computationally very intensive when α has many parameters as it require performing many hypothesis tests. In the naive approach where each α_p is scanned in a regular grid, the number of parameter points tested grows exponentially in the number of parameters. There is an alternative approach, which I call the 'profile construction' [8,9] and which statisticians call an 'hybrid resampling technique' [10,11] that is approximate to the full construction, but typically has good coverage properties. We return to the procedures and properties for the different types of Neyman Constructions later.

Figure 4 provides an overview of the classic Neyman construction corresponding to the left panel of Fig. 5. The left panel of Fig. 5 is taken from the Feldman and Cousins's paper [12] where the parameter of the model is denoted μ instead of θ . For each value of the parameter μ , the acceptance region in x is illustrated as a horizontal bar. Those regions are the ones that satisfy $T(\mathcal{D}) < k_\alpha$, and in the case of Feldman–Cousins the test statistic is the one of Eq. 52. This presentation of the confidence belt works well for a simple model in which the data consists of a single measurement $\mathcal{D} = \{x\}$. Once one has the confidence belt, then one can immediately find the confidence interval for a particular measurement of x simply by drawing a vertical line for the measured value of x and finding the intersection with the confidence belt.

Unfortunately, this convenient visualization doesn't generalize to complicated models with many channels or even a single channel marked Poisson model where $\mathcal{D} = \{x_1, \dots, x_n\}$. In those more

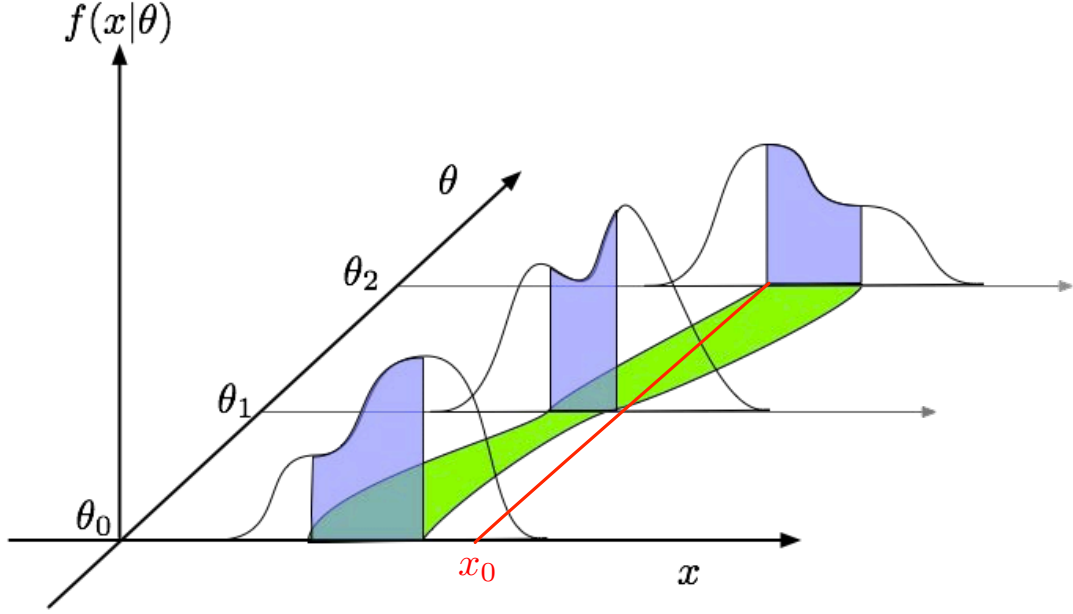


Fig. 4: A schematic visualization of the Neyman Construction. For each value of θ one finds a region in x that satisfies $\int f(x|\theta)dx$ (blue). Together these regions form a confidence belt (green). The intersection of the observation x_0 (red) with the confidence belt defines the confidence interval $[\theta_1, \theta_2]$.

complicated cases, the confidence belt can still be visualized where the observable x is replaced with T , the test statistic itself. Thus, the boundary of the belt is given by k_α vs. μ as in the right panel of Fig. 5. The analog to the vertical line in the left panel is now a curve showing how the observed value of the test statistic depends on μ . The confidence interval still corresponds to the intersection of the observed test statistic curve and the confidence belt, which clearly satisfies $T(\mathcal{D}) < k_\alpha$. For more complicated models with many parameters the confidence belt will have one axis for the test statistic and one axis for each model parameter.

Note, a 95% confidence interval *does not* mean that there is a 95% chance that the true value of the parameter is inside the interval – that is a manifestly Bayesian statement. One can produce a Bayesian *credible interval* with that interpretation; however, that requires a prior probability distribution over the parameters. Similarly, for any fixed interval I one can compute the Bayesian credibility of the interval

$$P(\alpha \in I|\mathcal{D}) = \frac{\int_I \mathbf{f}(\mathcal{D}|\alpha)\pi(\alpha)d\alpha}{\int \mathbf{f}(\mathcal{D}|\alpha)\pi(\alpha)d\alpha}. \quad (14)$$

4 Modeling and the Scientific Narrative

Now that we have established a general form for a probability model Eq. (2) and we have translated the basic questions of measurement, discovery, and exclusion into the statistical language we are ready to address the heart of the statistical challenge – building the model. It is difficult to overestimate how important the model building stage is. So many of the questions that are addressed to the statistical experts in the major particle physics collaborations are not really about statistics *per se*, but about model building. In fact, the first question that you are likely to be asked by one of the statistical experts is “what is your model?”

Often people are confused by the question “what is your model?” or simply have not written it down. You simply can’t make much progress on any statistical questions if you haven’t written down a model. Of course, people do usually have some idea for what it is that they want to do. The process of

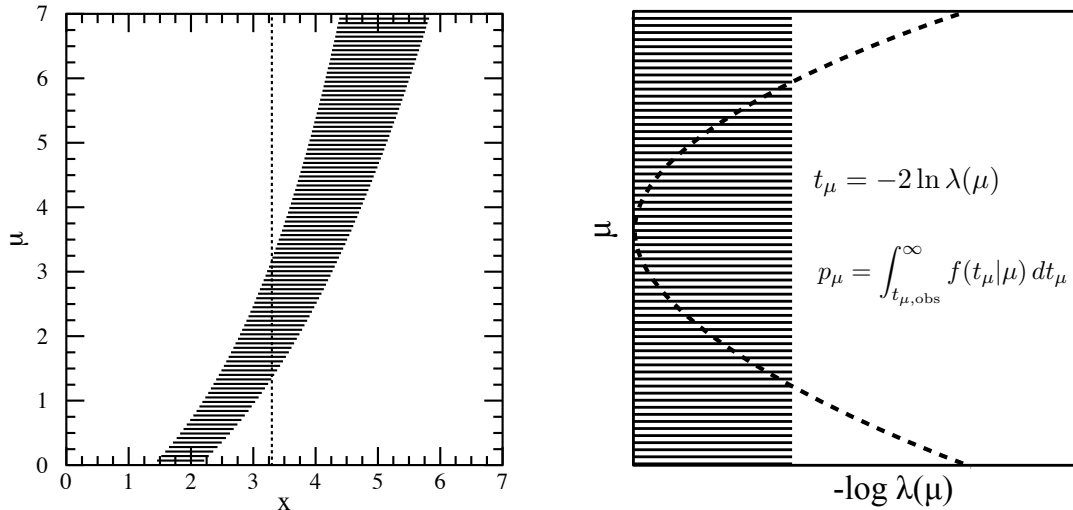


Fig. 5: Two presentations of a confidence belt (see text). Left panel taken from Ref. [12]. Right panel shows a presentation that generalizes to more complicated models.

writing down the model often obviates the answer to the question, reveals some fundamental confusion or assumption in the analysis strategy, or both. As mentioned in the introduction, writing down the model is intimately related with the analysis strategy and it is a good way to organize an analysis effort.

I like to think of the modeling stage in terms of a *scientific narrative*. I find that there are three main narrative elements, though many analyses use a mixture of these elements when building the model. Below I will discuss these narrative elements, how they are translated into a mathematical formulation, and their relative pros and cons.

4.1 Simulation Narrative

The simulation narrative is probably the easiest to explain and produces statistical models with the strongest logical connection to physical theory being tested. We begin with an relation that every particle physicists should know for the rate of events expected from a specific physical process

$$\text{rate} = (\text{flux}) \times (\text{cross section}) \times (\text{efficiency}) \times (\text{acceptance}) , \quad (15)$$

where the cross section is predicted from the theory, the flux is controlled by the accelerator¹¹, and the efficiency and acceptance are properties of the detector and event selection criteria. It is worth noting that the equation above is actually a repackaging of a more fundamental relationship. In fact the fundamental quantity that is predicted from first principles in quantum theory is the *scattering probability* $P(i \rightarrow f) = |\langle i|f \rangle|^2 / (\langle i|i \rangle \langle f|f \rangle)$ inside a box of size V over some time interval T , which is then repackaged into the Lorentz invariant form above.

In the simulation narrative the efficiency and acceptance are estimated with computer simulations of the detector. Typically, a large sample of events is generated using Monte Carlo techniques. The Monte Carlo sampling is performed separately for the hard (perturbative) interaction (e.g. MadGraph), the parton shower and hadronization process (e.g. Pythia and Herwig), and the interaction of particles with the detector (e.g. Geant). Note, the efficiency and acceptance depend on the physical process considered, and I will refer to each such process as a *sample* (in reference to the corresponding sample of events generated with Monte Carlo techniques).

¹¹In some cases, like cosmic rays, the flux must be estimated since the accelerator is quite far away.

To simplify the notation, I will define the effective cross section, $\sigma_{\text{eff.}}$ to be the product of the total cross section, efficiency, and acceptance. Thus, the total number of events expected to be selected for a given scattering process, ν , is the product of the time-integrated flux or time-integrated luminosity, λ , and the effective cross section

$$\nu = \lambda \sigma_{\text{eff.}} . \quad (16)$$

I use λ here instead of the more common L to avoid confusion with the likelihood function and because when we incorporate uncertainty on the time-integrated luminosity it will be a parameter of the model for which I have chosen to use greek letters.

If we did not need to worry about detector effects and we could measure the final state perfectly, then the distribution for any observable x would be given by

$$\text{(idealized)} \quad f(x) = \frac{1}{\sigma_{\text{eff.}}} \frac{d\sigma_{\text{eff.}}}{dx} . \quad (17)$$

Of course, we do need to worry about detector effects and we incorporate them with the detector simulation discussed above. From the Monte Carlo sample of events¹² $\{x_1, \dots, x_N\}$ we can estimate the underlying distribution $f(x)$ simply by creating a histogram. If we want we can write the histogram based on B bins centered at x_b with bin width w_b explicitly as

$$\text{(histogram)} \quad f(x) \approx h(x) = \sum_{i=1}^N \sum_{b=1}^B \frac{\theta(|x_i - x_b|/w_b)}{N} \frac{\theta(|x - x_b|/w_b)}{w_b} , \quad (18)$$

where the first Heaviside function accumulates simulated events in the bin and the second selects the bin containing the value of x in question. Histograms are the most common way to estimate a probability density function based on a finite sample, but there are other possibilities. The downside of histograms as an estimate for the distribution $f(x)$ is that they are discontinuous and have dependence on the location of the bin boundaries. A particularly nice alternative is called kernel estimation [13]. In this approach, one places a kernel of probability $K(x)$ centered around each event in the sample:

$$\text{(kernel estimate)} \quad f(x) \approx \hat{f}_0(x) = \frac{1}{N} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right) . \quad (19)$$

The most common choice of the kernel is a Gaussian distribution, and there are results for the optimal width of the kernel h . Equation 19 is referred to as the fixed kernel estimate since h is common for all the events in the sample. A second order estimate or adaptive kernel estimation provides better performance when the distribution is multimodal or has both narrow and wide features [13].

4.1.1 The multi-sample mixture model

So far we have only considered a single interaction process, or sample. How do we form a model when there are several scattering processes contributing to the total rate and distribution of x ? From first principles of quantum mechanics we must add these different processes together. Since there is no physical meaning to label individual processes that interfere quantum mechanically, I will consider all such processes as a single sample. Thus the remaining set of samples that do not interfere simply add incoherently. The total rate is simply the sum of the individual rates

$$\nu_{\text{tot}} = \sum_{s \in \text{samples}} \nu_s \quad (20)$$

¹²Here I only consider unweighted Monte Carlo samples, but the discussion below can be generalized for weighted Monte Carlo samples.

and the total distribution is a weighted sum called a *mixture model*

$$f(x) = \frac{1}{\nu_{\text{tot}}} \sum_{s \in \text{samples}} \nu_s f_s(x) , \quad (21)$$

where the subscript s has been added to the equations above for each such sample. With these two ingredients we can construct our marked Poisson model of Eq. 1 for a single channel, and we can simply repeat this for several disjoint event selection requirements to form a multi-channel simultaneous model like Eq. 2. In the multi-channel case we will give the additional subscript $c \in \text{channels}$ to ν_{cs} , $f_{cs}(x)$, $\nu_{c,\text{tot}}$, and $f_c(x)$. However, at this point, our model has no free parameters α .

4.1.2 Incorporating physics parameters into the model

Now we want to parametrize our model in terms of some physical parameters α , such as those that appear in the Lagrangian of a some theory. Changing the parameters in the Lagrangian of a theory will in general change both the total rate ν and the shape of the distributions $f(x)$. In principle, we can repeat the procedure above for each value of these parameters α to form $\nu_{cs}(\alpha)$ and $f_{cs}(x|\alpha)$ for each sample and selection channel, and, thus, from $\mathbf{f}_{\text{sim}}(\mathcal{D}|\alpha)$. In practice, we need to resort to some interpolation strategy over the individual parameter points $\{\alpha_i\}$ where we have Monte Carlo samples. We will return to these interpolation strategies later.

In some case the only effect of the parameter is to scale the rate of some scattering process $\nu_s(\alpha)$ without changing its distribution $f_s(x|\alpha)$. Furthermore, the scaling is often known analytically, for instance, a coupling constant produces a linear relationship like $\nu(\alpha_p) = \xi\alpha_p + \nu_0$. In such cases, interpolation is not necessary and the parametrization of the likelihood function is straightforward.

Note, not all physics parameters need be considered parameters of interest. There may be a free physics parameter that is not directly of interest, and as such it would be considered a nuisance parameter.

4.1.2.1 An example, the search for the standard model Higgs boson

In the case of searches for the standard model Higgs boson, the only free parameter in the Lagrangian is m_H . Once m_H is specified the rates and the shapes for each of the scattering processes (combinations of production and decay modes) are specified by the theory. Of course, as the Higgs boson mass changes the distributions do change so we do need to worry about interpolating the shapes $f(x|m_H)$. However the results are often presented as a *raster scan* over m_H , where one fixes m_H and then asks about the rate of signal events from the Higgs boson scattering process. With m_H fixed this is really a simple hypothesis test between background-only and signal-plus-background¹³, but we usually choose to construct a parametrized model that does not directly correspond to any theory. In this case the parameter of interest is some scaling of the rate with respect to the standard model prediction, $\mu = \sigma/\sigma_{\text{SM}}$, such that $\mu = 0$ is the background-only situation and $\mu = 1$ is the standard model prediction. Furthermore, we usually use this global μ factor for each of the production and decay modes even though essentially all theories of physics beyond the standard model would modify the rates of the various scattering processes differently. Figure 3 shows confidence intervals on μ for fixed values of m_H . Values below the solid black curve are not excluded (since an arbitrarily small signal rate cannot be differentiated from the background-only and this is a one-sided confidence interval).

4.1.3 Incorporating systematic effects

The parton shower, hadronization, and detector simulation components of the simulation narrative are based on phenomenological models that have many adjustable parameters. These parameters are nui-

¹³Note that $H \rightarrow WW$ interferes with “background-only” WW scattering process. For low Higgs boson masses, the narrow Higgs width means this interference is negligible. However, at high masses the interference effect is significant and we should really treat these two processes together as a single sample.

sance parameters included in our master list of parameters α . The changes in the rates $\nu(\alpha)$ and shapes $f(x|\alpha)$ due to these parameters lead to systematic uncertainties¹⁴. We have already eluded to how one can deal with the presence of nuisance parameters in hypothesis testing and confidence intervals, but here we are focusing on the modeling stage. In principle, we deal with modeling of these nuisance parameters in the same way as the physics parameters, which is to generate Monte Carlo samples for several choices of the parameters $\{\alpha_i\}$ and then use some interpolation strategy to form a continuous parametrization for $\nu(\alpha)$, $f(x|\alpha)$, and $\mathbf{f}_{\text{sim}}(\mathcal{D}|\alpha)$. In practice, there are many nuisance parameters associated to the parton shower, hadronization, and detector simulation so this becomes a multi-dimensional interpolation problem¹⁵. This is one of the most severe challenges for the simulation narrative.

Typically, we don't map out the correlated effect of changing multiple α_p simultaneously. Instead, we have some nominal settings for these parameters α^0 and then vary each individual parameter 'up' and 'down' by some reasonable amount $\alpha_p^\pm$. So if we have N_P parameters we typically have $1 + 2N_P$ variations of the Monte Carlo sample from which we try to form $\mathbf{f}_{\text{sim}}(\mathcal{D}|\alpha)$. This is clearly not an ideal situation and it is not hard to imagine cases where the combined effect on the rate and shapes cannot be factorized in terms of changes from the individual parameters.

What is meant by "vary each individual parameter 'up' and 'down' by some reasonable amount" in the paragraph above? The nominal choice of the parameters α^0 is usually based on experience, test beam studies, Monte Carlo 'tunings', etc. These studies correspond to auxiliary measurements in the language used in Section 2.2 and Section 2.4. Similarly, these parameters typically have some maximum likelihood estimates and standard uncertainties from the auxiliary measurements as described in Section 3.1. Thus our complete model $\mathbf{f}_{\text{tot}}(\mathcal{D}|\alpha)$ of Eq. 6 should not only deal with parametrizing the effect of changing each α_p but also include either a constraint term $f_p(a_p|\alpha_p)$ or an additional channel that describes a more complete probability model for the auxiliary measurement.

Below we will consider a specific interpolation strategy and a few of the most popular conventions for constraint terms. However, before moving on it is worth emphasizing that while, naively, the matrix element associated to a perturbative scattering amplitude has no free parameters (beyond the physics parameters discussed above), fixed order perturbative calculations do have residual scale dependence. This type of *theoretical uncertainty* has no auxiliary measurement associated with it even in principle, thus it really has no frequentist description. This was discussed briefly in Section 2.4. In contrast, the parton density functions are the results of auxiliary measurements and the groups producing the parton density function sets spend time providing sensible multivariate constraint terms for those parameters. However, those measurements also have uncertainties due to parametrization choices and theoretical uncertainties, which are not statistical in nature. In short we must take care in ascribing constraint terms to theoretical uncertainties and measurements that have theoretical uncertainties¹⁶.

4.1.4 Tabulating the effect of varying sources of uncertainty

The treatment of systematic uncertainties is subtle, particularly when one wishes to take into account the correlated effect of multiple sources of systematic uncertainty across many signal and background samples. The most important conceptual issue is that we separate the source of the uncertainty (for instance the uncertainty in the calorimeter's response to jets) from its effect on an individual signal or background sample (e.g. the change in the acceptance and shape of a W +jets background). In particular, the same source of uncertainty has a different effect on the various signal and background samples. The effect of these 'up' and 'down' variations about the nominal predictions $\nu_s(\alpha^0)$ and $f_{sb}(x|\alpha^0)$ is quantified by dedicated studies. The result of these studies can be arranged in tables like those below. The main purpose of the HistFactory XML schema is to represent these tables. And HistFactory is a tool that can convert these tables into our master model $\mathbf{f}_{\text{tot}}(\mathcal{D}|\alpha)$ of Eq. 6 implemented as a RooAbsPdf

¹⁴Systematic uncertainty is arguably a better term than systematic *error*.

¹⁵This is sometimes referred to as 'template morphing'

¹⁶"Note that I deliberately called them theory *errors*, not uncertainties."— Tilman Plehn

with a `ModelConfig` to make it compatible with `RooStats` tools. The convention used by `HistFactory` is related to our notation via

$$\nu_s(\alpha) f_s(x|\alpha) = \eta_s(\alpha) \sigma_s(x|\alpha) \quad (22)$$

where $\eta_s(\alpha)$ represents relative changes in the overall rate $\nu(\alpha)$ and $\sigma_s(x|\alpha)$ includes both changes to the rate and the shape $f(x|\alpha)$. This choice is one of convenience because histograms are often not normalized to unity, but instead in code rate information. As the name implies, `HistFactory` works with histograms, so instead of writing $\sigma_s(x|\alpha)$ the table is written as $\sigma_{sb}(\alpha)$, where b is a bin index. To compress the notation further, $\eta_{p=1,s=1}^+$ and $\sigma_{psb}^\pm$ represent the value of when $\alpha_p = \alpha_p^\pm$ and all other parameters are fixed to their nominal values. Thus we arrive at the following tabular form for models built on the simulation narrative based on histograms with individual nuisance parameters varied one at a time:

Syst	Sample 1	...	Sample N
Nominal Value	$\eta_{s=1}^0 = 1$	...	$\eta_{s=N}^0 = 1$
$p=\text{OverallSys } 1$	$\eta_{p=1,s=1}^+, \eta_{p=1,s=1}^-$	...	$\eta_{p=1,s=N}^+, \eta_{p=1,s=N}^-$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$p=\text{OverallSys } M$	$\eta_{p=M,s=1}^+, \eta_{p=M,s=1}^-$	...	$\eta_{p=M,s=N}^+, \eta_{p=M,s=N}^-$
Net Effect	$\eta_{s=1}(\alpha)$	...	$\eta_{s=N}(\alpha)$

Table 1: Tabular representation of sources of uncertainties that produce a correlated effect in the normalization individual samples (e.g. `OverallSys`). The η_{ps}^+ represent histogram when $\alpha_s = 1$ and are inserted into the `High` attribute of the `OverallSys` XML element. Similarly, the η_{ps}^- represent histogram when $\alpha_s = -1$ and are inserted into the `Low` attribute of the `OverallSys` XML element. Note, this does not imply that $\eta^+ > \eta^-$, the $\pm$ superscripts correspond to the variation in the source of the systematic, not the resulting effect.

Syst	Sample 1	...	Sample N
Nominal Value	$\sigma_{s=1,b}^0$	...	$\sigma_{s=N,b}^0$
$p=\text{HistoSys } 1$	$\sigma_{p=1,s=1,b}^+, \sigma_{p=1,s=1,b}^-$	...	$\sigma_{p=1,s=N,b}^+, \sigma_{p=1,s=N,b}^-$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$p=\text{HistoSys } M$	$\sigma_{p=M,s=1,b}^+, \sigma_{p=M,s=1,b}^-$	...	$\sigma_{p=M,s=N,b}^+, \sigma_{p=M,s=N,b}^-$
Net Effect	$\sigma_{s=1,b}(\alpha)$	...	$\sigma_{s=N,b}(\alpha)$

Table 2: Tabular representation of sources of uncertainties that produce a correlated effect in the normalization and shape individual samples (e.g. `HistoSys`). The σ_{psb}^+ represent histogram when $\alpha_s = 1$ and are inserted into the `HighHist` attribute of the `HistoSys` XML element. Similarly, the σ_{psb}^- 's represent histogram when $\alpha_s = -1$ and are inserted into the `LowHist` attribute of the `HistoSys` XML element.

4.1.5 Interpolation Conventions

For each sample, one can interpolate and extrapolate from the nominal prediction $\eta_s^0 = 1$ and the variations $\eta_{ps}^\pm$ to produce a parametrized $\eta_s(\alpha)$. Similarly, one can interpolate and extrapolate from the nominal shape σ_{sb}^0 and the variations $\sigma_{psb}^\pm$ to produce a parametrized $\sigma_{sb}(\alpha)$. We choose to parametrize α_p such that $\alpha_p = 0$ is the nominal value of this parameter, $\alpha_p = \pm 1$ are the “ $\pm 1\sigma$ variations”. Needless to say, there is a significant amount of ambiguity in these interpolation and extrapolation procedures and they must be handled with care. Bellow are some of the interpolation strategies supported by `HistFactory`. These are all ‘vertical’ style interpolation treated independently per-bin. Four interpolation strategies are described below and can be compared in Fig. 6. The interested reader is invited to look at alternative ‘horizontal’ interpolation strategies, such as the one developed by Alex Read in Ref. [14]

(the RooFit implementation is called RooIntegralMorph) and Max Baak's RooMomentMorph. These horizontal interpolation strategies are better suited for features moving, such as the location of an invariant mass bump changing with the hypothesized mass of a new particle..

Piecewise Linear (InterpCode=0)

The piecewise-linear interpolation strategy is defined as

$$\eta_s(\alpha) = 1 + \sum_{p \in \text{Syst}} I_{\text{lin.}}(\alpha_p; 1, \eta_{sp}^+, \eta_{sp}^-) \quad (23)$$

and for shape interpolation it is

$$\sigma_{sb}(\alpha) = \sigma_{sb}^0 + \sum_{p \in \text{Syst}} I_{\text{lin.}}(\alpha_p; \sigma_{sb}^0, \sigma_{psb}^+, \sigma_{psb}^-) \quad (24)$$

with

$$I_{\text{lin.}}(\alpha; I^0, I^+, I^-) = \begin{cases} \alpha(I^+ - I^0) & \alpha \geq 0 \\ \alpha(I^0 - I^-) & \alpha < 0 \end{cases} \quad (25)$$

PROS: This approach is the most straightforward of the interpolation strategies.

CONS: It has two negative features. First, there is a kink (discontinuous first derivative) at $\alpha = 0$ (see Fig. 6(b-d)), which can cause some difficulties for numerical minimization packages such as Minuit. Second, the interpolation factor can extrapolate to negative values. For instance, if $\eta^- = 0.5$ then we have $\eta(\alpha) < 0$ when $\alpha < -2$ (see Fig. 6(c)).

Note that one could have considered the simultaneous variation of α_p and $\alpha_{p'}$ in a multiplicative way. The multiplicative accumulation is not an option currently.

Note that this is the default convention for $\sigma_{sb}(\alpha)$ (i.e. HistoSys).

Piecewise Exponential (InterpCode=1)

The piecewise exponential interpolation strategy is defined as

$$\eta_s(\alpha) = \prod_{p \in \text{Syst}} I_{\text{exp.}}(\alpha_p; 1, \eta_{sp}^+, \eta_{sp}^-) \quad (26)$$

and for shape interpolation it is

$$\sigma_{sb}(\alpha) = \sigma_{sb}^0 \prod_{p \in \text{Syst}} I_{\text{exp.}}(\alpha_p; \sigma_{sb}^0, \sigma_{psb}^+, \sigma_{psb}^-) \quad (27)$$

with

$$I_{\text{exp.}}(\alpha; I^0, I^+, I^-) = \begin{cases} (I^+/I^0)^\alpha & \alpha \geq 0 \\ (I^-/I^0)^{-\alpha} & \alpha < 0 \end{cases} \quad (28)$$

PROS: This approach ensures that $\eta(\alpha) \geq 0$ (see Fig. 6(c)) and for small response to the uncertainties it has the same linear behavior near $\alpha \sim 0$ as the piecewise linear interpolation (see Fig. 6(a)).

CONS: It has two negative features. First, there is a kink (discontinuous first derivative) at $\alpha = 0$, which can cause some difficulties for numerical minimization packages such as Minuit. Second, for large uncertainties it develops a different linear behavior compared to the piecewise linear interpolation. In particular, even if the systematic has a symmetric response (i.e. $\eta^+ - 1 = 1 - \eta^-$) the interpolated response will develop a kink for large response to the uncertainties (see Fig. 6(c)).

Note that the one could have considered the simultaneous variation of α_p and $\alpha_{p'}$ in an additive way, but this is not an option currently.

Note, that when paired with a Gaussian constraint on α this is equivalent to linear interpolation and a log-normal constraint in $\ln(\alpha)$. This is the default strategy for normalization uncertainties $\eta_s(\alpha)$ (i.e. OverallSys) and is the standard convention for normalization uncertainties in the LHC Higgs Combination Group. In the future, the default may change to the Polynomial Interpolation and Exponential Extrapolation described below.

Polynomial Interpolation and Exponential Extrapolation (InterpCode=4)

The strategy of this interpolation option is to use the piecewise exponential extrapolation as above with a polynomial interpolation that matches $\eta(\alpha = \pm\alpha_0)$, $d\eta/d\alpha|_{\alpha=\pm\alpha_0}$, and $d^2\eta/d\alpha^2|_{\alpha=\pm\alpha_0}$ and the boundary $\pm\alpha_0$ is defined by the user (with default $\alpha_0 = 1$).

$$\eta_s(\alpha) = \prod_{p \in \text{Sys}} I_{\text{poly|exp.}}(\alpha_p; 1, \eta_{sp}^+, \eta_{sp}^-, \alpha_0) \quad (29)$$

with

$$I_{\text{poly|exp.}}(\alpha; I^0, I^+, I^-, \alpha_0) = \begin{cases} (I^+/I_0)^\alpha & \alpha \geq \alpha_0 \\ 1 + \sum_{i=1}^6 a_i \alpha^i & |\alpha| < \alpha_0 \\ (I^-/I_0)^{-\alpha} & \alpha \leq -\alpha_0 \end{cases} \quad (30)$$

and the a_i are fixed by the boundary conditions described above.

PROS: This approach avoids the kink (discontinuous first and second derivatives) at $\alpha = 0$ (see Fig. 6(b-d)), which can cause some difficulties for numerical minimization packages such as Minuit. This approach ensures that $\eta(\alpha) \geq 0$ (see Fig. 6(c)).

Note: This option is not available in ROOT 5.32.00, but is available for normalization uncertainties (OverallSys) in the subsequent patch releases. In future releases, this may become the default.

4.1.6 Consistent Bayesian and Frequentist modeling

The variational estimates $\eta^\pm$ and $\sigma^\pm$ typically correspond to so called “ $\pm 1\sigma$ variations” in the source of the uncertainty. Here we are focusing on the source of the uncertainty, not its affect on rates and shapes. For instance, we might say that the jet energy scale has a 10% uncertainty.¹⁷ This is common jargon, but what does it mean? The most common interpretation of this statement is that the uncertain parameter α_p (e.g. the jet energy scale) has a Gaussian distribution. However, this way of thinking is manifestly Bayesian. If the parameter was estimated from an auxiliary measurement, then it is the PDF for that measurement that we wish to include into our probability model. In the frequentist way of thinking, the jet energy scale has an unknown true value and upon repeating the experiment many times the auxiliary measurements estimating the jet energy scale would fluctuate randomly about this true value. To aid in this subtle distinction, we use greek letters for the parameters (e.g. α_p) and roman letters for the auxiliary measurements a_p . Furthermore, we interpret the “ $\pm 1\sigma$ ” variation in the frequentist sense, which leads to the constraint term $f_p(a_p|\alpha_p)$. Then, we can pair the resulting likelihood with some prior on α_p to form a Bayesian posterior if we wish according to Eq. 5.

It is often advocated that a “log-normal” or “gamma” distribution for α_p is more appropriate than a gaussian constraint [15]. This is particularly clear in the case of bounded parameters and large uncertainties. Here we must take some care to build a probability model that can maintain a consistent interpretation in Bayesian a frequentist settings. Table 3 summarizes a few consistent treatments of the frequentist PDF, the likelihood function, a prior, and the resulting posterior.

Finally, it is worth mentioning that the uncertainty on some parameters is not the result of an auxiliary measurement – so the constraint term idealization, it is not just a convenience, but a real conceptual

¹⁷Without loss of generality, we choose to parametrize α_p such that $\alpha_p = 0$ is the nominal value of this parameter, $\alpha_p = \pm 1$ are the “ $\pm 1\sigma$ variations”.

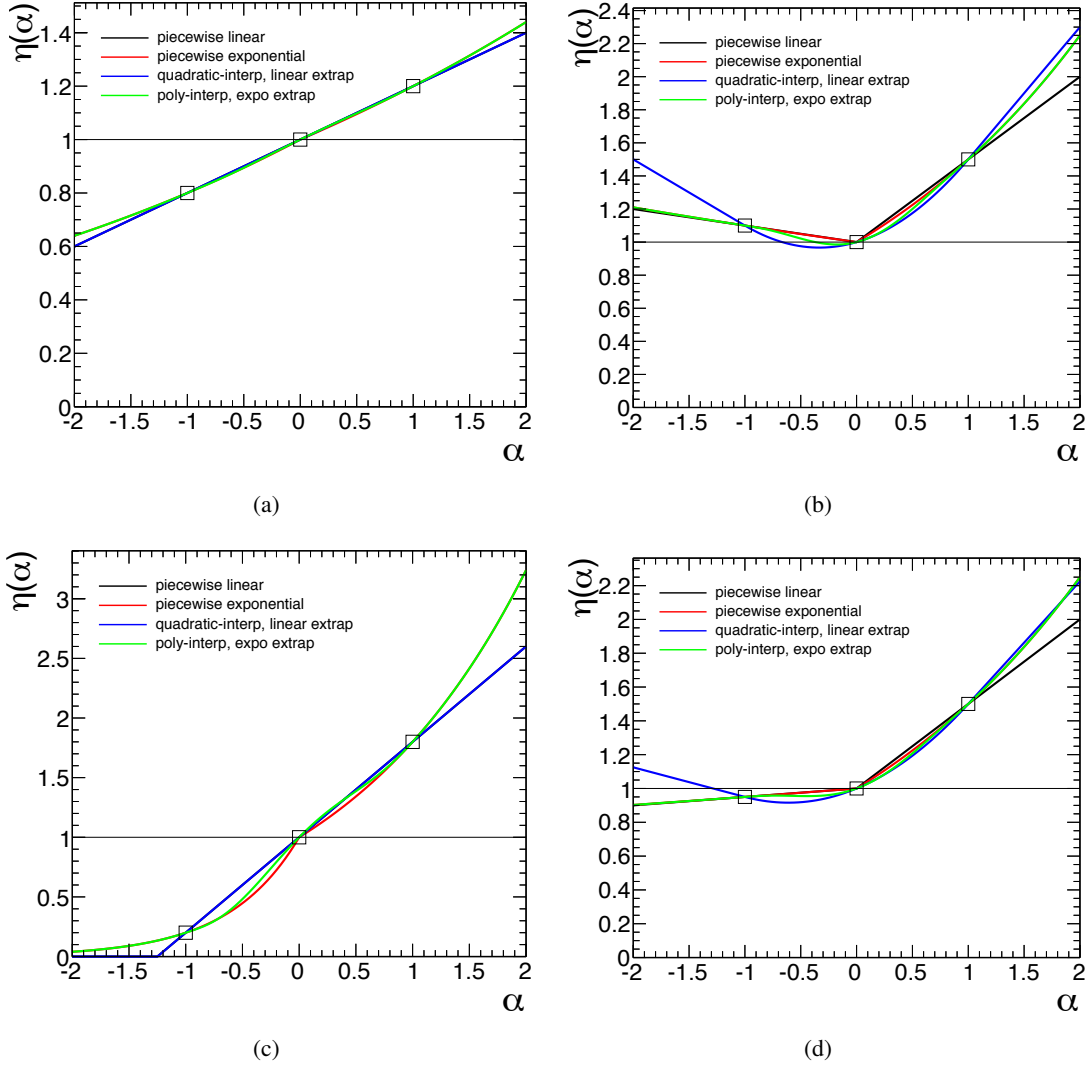


Fig. 6: Comparison of the three interpolation options for different $\eta^\pm$. (a) $\eta^- = 0.8$, $\eta^+ = 1.2$, (b) $\eta^- = 1.1$, $\eta^+ = 1.5$, (c) $\eta^- = 0.2$, $\eta^+ = 1.8$, and (d) $\eta^- = 0.95$, $\eta^+ = 1.5$

leap. This is particularly true for theoretical uncertainties from higher-order corrections or renormalization and factorization scale dependence. In these cases a formal frequentist analysis would not include a constraint term for these parameters, and the result would simply depend on their assumed values. As this is not the norm, we can think of reading Table 3 from right-to-left with a subjective Bayesian prior $\pi(\alpha)$ being interpreted as coming from a fictional auxiliary measurement.

4.1.6.1 Gaussian Constraint

The Gaussian constraint for α_p corresponds to the familiar situation. It is a good approximation of the auxiliary measurement when the likelihood function for α_p from that auxiliary measurement has a Gaussian shape. More formally, it is valid when the maximum likelihood estimate of α_p (e.g. the best fit value of α_p) has a Gaussian distribution. Here we can identify the maximum likelihood estimate of α_p with the global observable a_p , remembering that it is a number that is extracted from the data and thus

PDF	Likelihood $\propto$	Prior π_0	Posterior π
$G(a_p \alpha_p, \sigma_p)$	$G(\alpha_p a_p, \sigma_p)$	$\pi_0(\alpha_p) \propto \text{const}$	$G(\alpha_p a_p, \sigma_p)$
$\text{Pois}(n_p \tau_p\beta_p)$	$P_\Gamma(\beta_p A = \tau_p; B = 1 + n_p)$	$\pi_0(\beta_p) \propto \text{const}$	$P_\Gamma(\beta_p A = \tau_p; B = 1 + n_p)$
$P_{\text{LN}}(n_p \beta_p, \sigma_p)$	$\beta_p \cdot P_{\text{LN}}(\beta_p n_p, \sigma_p)$	$\pi_0(\beta_p) \propto \text{const}$	$P_{\text{LN}}(\beta_p n_p, \sigma_p)$
$P_{\text{LN}}(n_p \beta_p, \sigma_p)$	$\beta_p \cdot P_{\text{LN}}(\beta_p n_p, \sigma_p)$	$\pi_0(\beta_p) \propto 1/\beta_p$	$P_{\text{LN}}(\beta_p n_p, \sigma_p)$

Table 3: Table relating consistent treatments of PDF, likelihood, prior, and posterior for nuisance parameter constraint terms.

its distribution has a frequentist interpretation.

$$G(a_p|\alpha_p, \sigma_p) = \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp \left[-\frac{(a_p - \alpha_p)^2}{2\sigma_p^2} \right] \quad (31)$$

with $\sigma_p = 1$ by default. Note that the PDF of a_p and the likelihood for α_p are positive for all values.

4.1.6.2 Poisson (“Gamma”) constraint

When the auxiliary measurement is actually based on counting events in a control region (e.g. a Poisson process), it is more accurate to describe the auxiliary measurement with a Poisson distribution. It has been shown that the truncated Gaussian constraint can lead to undercoverage (overly optimistic) results, which makes this issue practically relevant [4]. Table 3 shows that a Poisson PDF together with a uniform prior leads to a gamma posterior, thus this type of constraint is often called a “gamma” constraint. This is a bit unfortunate since the gamma distribution is manifestly Bayesian and with a different choice of prior, one might not arrive at a gamma posterior. When dealing with the Poisson constraint, it is no longer convenient to work with our conventional scaling for α_p which can be negative. Instead, it is more natural to think of the number of events measured in the auxiliary measurement n_p and the mean of the Poisson parameter. This information is not usually available, instead one usually has some notion of the relative uncertainty in the parameter σ_p^{rel} (e.g. the jet energy scale is known to 10%). In order to give some uniformity to the different uncertainties of this type and think of relative uncertainty, the nominal rate is factored out into a constant τ_p and the mean of the Poisson is given by $\tau_p\alpha_p$.

$$\text{Pois}(n_p|\tau_p\alpha_p) = \frac{(\tau_p\alpha_p)^{n_p} e^{-\tau_p\alpha_p}}{n_p!} \quad (32)$$

Here we can use the fact that $\text{Var}[n_p] = \sqrt{\tau_p\alpha_p}$ and reverse engineer the nominal auxiliary measurement

$$n_p^0 = \tau_p = (1/\sigma_p^{\text{rel}})^2. \quad (33)$$

where the superscript 0 is to remind us that n_p will fluctuate in repeated experiments but n_p^0 is the value of our measured estimate of the parameter.

One important thing to keep in mind is that there is only one constraint term per nuisance parameter, so there must be only one σ_p^{rel} per nuisance parameter. This σ_p^{rel} is related to the fundamental uncertainty in the source and we cannot infer this from the various response terms $\eta_{ps}^\pm$ or $\sigma_{pub}^\pm$.

Another technical difficulty is that the Poisson distribution is discrete. So if one were to say the relative uncertainty was 30%, then we would find $n_p^0 = 11.11\dots$, which is not an integer. Rounding n_p to the nearest integer while maintaining $\tau_p = (1/\sigma_p^{\text{rel}})^2$ will bias the maximum likelihood estimate of α_p away from 1. To avoid this, one can use the gamma distribution, which generalizes more continuously with

$$P_\Gamma(\alpha_p|A = \tau_p, B = n_p - 1) = A(A\alpha_p)^B e^{-A\alpha_p} / \Gamma(B). \quad (34)$$

This approach works fine for likelihood fits, Bayesian calculations, and frequentist techniques based on asymptotic approximations, but it does not offer a consistent treatment of the PDF for the global observable n_p that is needed for techniques based on Monte Carlo sampling.

4.1.6.3 Log-normal constraint

From Eadie et al.: “The log-normal distribution represents a random variable whose logarithm follows a normal distribution. It provides a model for the error of a process involving many small multiplicative errors (from the Central Limit Theorem). It is also appropriate when the value of an observed variable is a random proportion of the previous observation.” [15, 16]. This logic of multiplicative errors applies to the measured value, not the parameter. Thus, it is natural to say that there is some auxiliary measurement (global observable) with a log-normal distribution. As in the gamma/Poisson case above, let us again say that the global observable is n_p with a nominal value

$$n_p^0 = \tau_p = (1/\sigma_p^{\text{rel}})^2. \quad (35)$$

Then the conventional choice for the corresponding log-normal distribution is

$$P_{\text{LN}}(n_p|\alpha_p, \kappa_p) = \frac{1}{\sqrt{2\pi \ln \kappa}} \frac{1}{n_p} \exp \left[-\frac{\ln(n_p/\alpha_p)^2}{2(\ln \kappa_p)^2} \right] \quad (36)$$

while the likelihood function is (blue curve in Fig. 7(a)).

$$L(\alpha_p) = \frac{1}{\sqrt{2\pi \ln \kappa}} \frac{1}{n_p} \exp \left[-\frac{\ln(n_p/\alpha_p)^2}{2(\ln \kappa_p)^2} \right]; \quad (37)$$

To get to the posterior for α_p given n_p we need an ur-prior $\eta(\alpha_p)$

$$\pi(\alpha_p) \propto \eta(\alpha_p) \frac{1}{\sqrt{2\pi \ln \kappa}} \frac{1}{n_p} \exp \left[-\frac{\ln(n_p/\alpha_p)^2}{2(\ln \kappa_p)^2} \right] \quad (38)$$

If $\eta(\alpha_p)$ is uniform, then the posterior looks like the red curve in Fig. 7(b). However, when paired with an “ur-prior” $\eta(\alpha_p) \propto 1/\alpha_p$ (green curve in Fig. 7(b)), this results in a posterior distribution that is also of a log-normal form for α_p (blue curve in Fig. 7(b)).

4.1.7 Incorporating Monte Carlo statistical uncertainty on the histogram templates

The histogram based approaches described above are based Monte Carlo simulations of full detector simulation. These simulations are very computationally intensive and often the histograms are sparsely populated. In this case the histograms are not good descriptions of the underlying distribution, but are estimates of that distribution with some statistical uncertainty. Barlow and Beeston outlined a treatment of this situation in which each bin of each sample is given a nuisance parameter for the true rate, which is then fitted using both the data measurement and the Monte Carlo estimate [17]. This approach would lead to several hundred nuisance parameters in the current analysis. Instead, the HistFactory employs a lighter weight version in which there is only one nuisance parameter per bin associated with the total Monte Carlo estimate and the total statistical uncertainty in that bin. If we focus on an individual bin with index b the contribution to the full statistical model is the factor

$$\text{Pois}(n_b|\nu_b(\alpha) + \gamma_b \nu_b^{\text{MC}}(\alpha)) \text{Pois}(m_b|\gamma_b \tau_b), \quad (39)$$

where n_b is the number of events observed in the bin, $\nu_b(\alpha)$ is the number of events expected in the bin where Monte Carlo statistical uncertainties need not be included (either because the estimate is data driven or because the Monte Carlo sample is sufficiently large), $\nu_b^{\text{MC}}(\alpha)$ is the number of events estimated using Monte Carlo techniques where the statistical uncertainty needs to be taken into account. Both expectations include the dependence on the parameters α . The factor γ_b is the nuisance parameter reflecting that the true rate may differ from the Monte Carlo estimate $\nu_b^{\text{MC}}(\alpha)$ by some amount. If the total statistical uncertainty is δ_b , then the relative statistical uncertainty is given by $\nu_b^{\text{MC}}/\delta_b$. This corresponds to a total Monte Carlo sample in that bin of size $m_b = (\delta_b/\nu_b^{\text{MC}})^2$. Treating the Monte Carlo

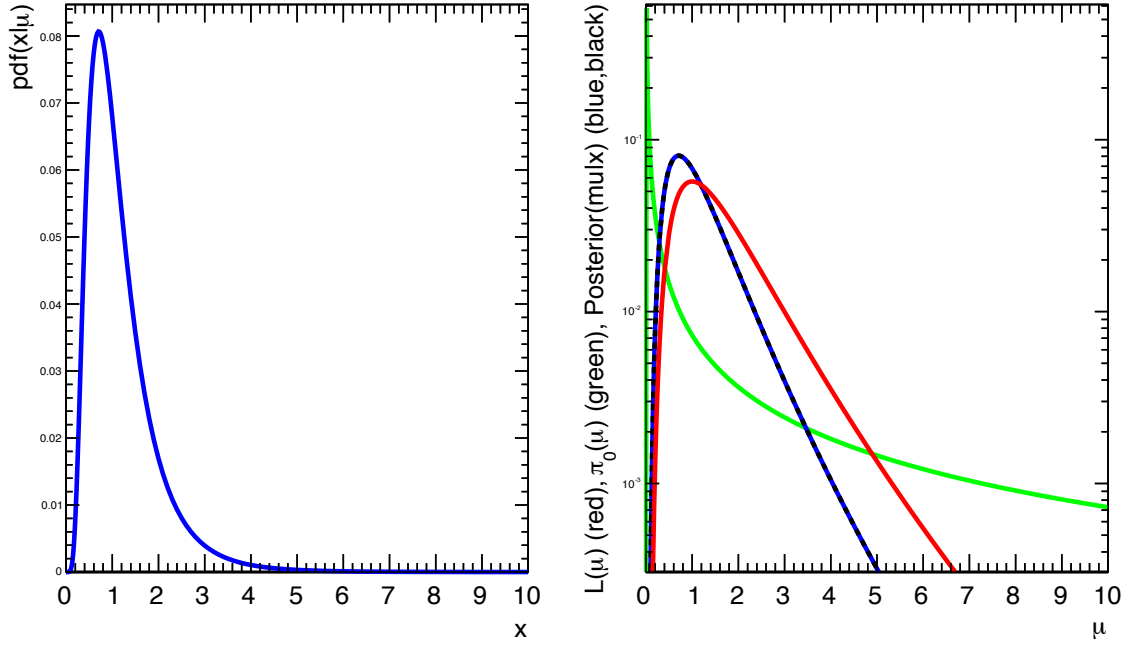


Fig. 7: The lognormal constraint term: (left) the PDF for the global observable a_p and (right) the likelihood function, the posterior based on a flat prior on α_p , and the posterior based on a $1/\alpha_p$ prior.

estimate as an auxiliary measurement, we arrive at a Poisson constraint term $\text{Pois}(m_b|\gamma_b\tau_b)$, where m_b would fluctuate about $\gamma_b\tau_b$ if we generate a new Monte Carlo sample. Since we have scaled γ to be a factor about 1, then we also have $\tau_b = (\nu_b^{\text{MC}}/\delta_b)^2$; however, τ_b is treated as a fixed constant and does not fluctuate when generating ensembles of pseudo-experiments.

It is worth noting that the conditional maximum likelihood estimate $\hat{\gamma}_b(\alpha)$ can be solved analytically with a simple quadratic expression.

$$\hat{\gamma}_b(\alpha) = \frac{-B + \sqrt{B^2 - 4AC}}{2A}, \quad (40)$$

with

$$A = \nu_b^{\text{MC}}(\alpha)^2 + \tau_b\nu_b^{\text{MC}}(\alpha) \quad (41)$$

$$B = \nu_b(\alpha)\tau + \nu_b(\alpha)\nu_b^{\text{MC}}(\alpha) - n_b\nu_b^{\text{MC}}(\alpha) - m_b\nu_b^{\text{MC}}(\alpha) \quad (42)$$

$$C = m_b\nu_b(\alpha). \quad (43)$$

In a Bayesian technique with a flat prior on γ_b , the posterior distribution is a gamma distribution. Similarly, the distribution of $\hat{\gamma}_b$ will take on a skew distribution with an envelope similar to the gamma distribution, but with features reflecting the discrete values of m_b . Because the maximum likelihood estimate of γ_b will also depend on n_b and $\hat{\alpha}$, the features from the discrete values of m_b will be smeared. This effect will be more noticeable for large statistical uncertainties where τ_b is small and the distribution of $\hat{\gamma}_b$ will have several small peaks. For smaller statistical uncertainties where τ_b is large the distribution of $\hat{\gamma}_b$ will be approximately Gaussian.

4.2 Data-Driven Narrative

The strength of the simulation narrative lies in its direct logical link from the underlying theory to the modeling of the experimental observations. The weakness of the simulation narrative derives from the

weaknesses in the simulation itself. Data-driven approaches are more motivated when they address specific deficiencies in the simulation. Before moving to a more abstract or general discussion of the data-driven narrative, let us first consider a few examples.

The first example we have already considered in Section 2.2 in the context of the “on/off” problem. There we introduced an auxiliary measurement that counted n_{CR} events in a control region to estimate the background ν_B in the signal region. In order to do this we needed to understand the ratio of the number of events from the background process in the control and signal regions, τ . This ratio τ either comes from some reasonable assumption or simulation. For example, if one wanted to estimate the background due to jets faking muons $j \rightarrow \mu$ for a search selecting $\mu^+\mu^-$, then one might use a sample of $\mu^\pm\mu^\pm$ events as a control region. Here the motivation for using a data-driven approach is that modeling the processes that lead to $j \rightarrow \mu$ relies heavily on the tails of fragmentation functions and detector response, which one might reasonably have some skepticism. If one assumes that control region is expected to have negligible signal in it, that backgrounds that produce $\mu^+\mu^-$ other than the jets faking muons, and that the rate for $j \rightarrow \mu^-$ is the same¹⁸ as the rate for $j \rightarrow \mu^+$, then one can assume $\tau = 1$. Thus, this background estimate is as trustworthy as the assumptions that went into it. In practice, several of these assumptions may be violated. Another approach is to use simulation of these background processes to estimate the ratio τ ; a hybrid of the data-driven and simulation narratives.

Let us now consider the search for $H \rightarrow \gamma\gamma$ shown in Fig. 8 [18,19]. The right plot of Fig. 8 shows the composition of the backgrounds in this search, including the continuum production of $pp \rightarrow \gamma\gamma$, the γ +jets process with a jet faking a photon $j \rightarrow \gamma$, and the multi-jet process with two jets faking photons. The continuum production of $\gamma\gamma$ has a theoretical uncertainty that is much larger than the statistical fluctuations one would expect in the data. Similarly, the rate of jets faking photons is sensitive to fragmentation and the detector simulation. These uncertainties are large compared to the statistical fluctuations in the data itself. Thus we can use the distribution in Fig. 8 to measure the total background rate. Of course, the signal would also be in this distribution, so one either needs to apply a mass window around the signal and consider the region outside of the window as a sideband control sample or model the signal and background contributions to the distribution. In the case of the $H \rightarrow \gamma\gamma$ shown in Fig. 8 [18,19] the modeling of the distribution signal and background distributions is not based on histograms from simulation, but instead a continuous function is used as an effective model. I will discuss this effective modeling narrative below, but point out that here this is another example of a hybrid narrative.

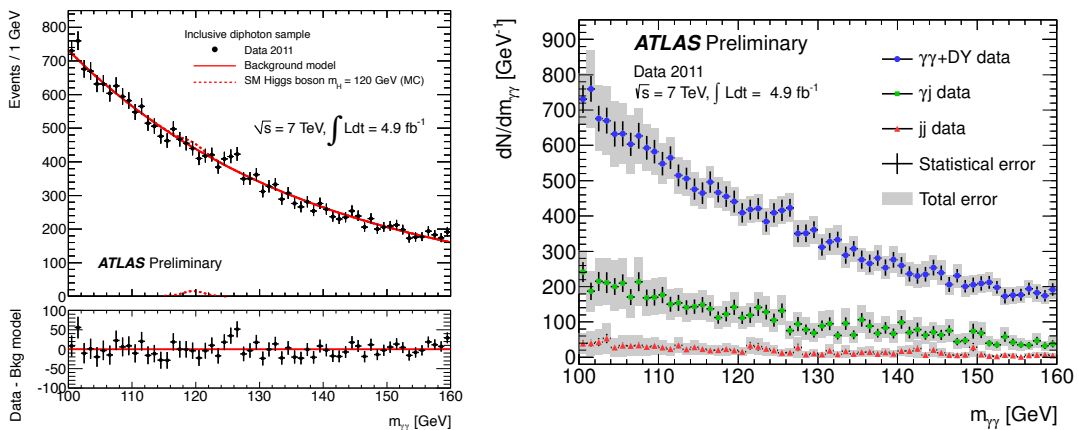


Fig. 8: Distribution of diphoton invariant mass distributions in the ATLAS $H \rightarrow \gamma\gamma$ search. The left plot shows a fit of an effective model to the data and the right plot shows an estimate of the $\gamma\gamma$, γ +jet, and dijet contributions.

¹⁸Given that the LHC collides pp and not $p\bar{p}$, there is clearly a reason to worry if this assumption is valid.

The final example to consider is an extension of the ‘on/off’ model, often referred to as the ‘ABCD’ method. Let us start with the ‘on/off’ model: $\text{Pois}(n_{\text{SR}}|\nu_S + \nu_B) \cdot \text{Pois}(n_{\text{CR}}|\tau\nu_B)$. As mentioned above, this requires that one estimate τ either from simulation or through some assumptions. The ABCD method aims to introduce two new control regions that can be used to measure τ . To see this, let us imagine that the signal and control regions correspond to requiring some continuous variable x being greater than or less than some threshold value x_c . If we could introduce a second discriminating variable y such that the distribution for background factorizes $f_B(x, y) = f_B(x)f_B(y)$, then we have a handle to measure the factor τ . Typically, one introduces a threshold y_c so that the signal contribution is small below this threshold¹⁹. Figure 9 shows an example where $x_c = y_c = 5$. With these two thresholds we have four regions that we can schematically refer to as A, B, C, and D. In the case of simply counting events in these regions we can write the total expectation as

$$\begin{aligned}\nu_A &= 1 \cdot \mu + \nu_A^{MC} + 1 \cdot \nu_A \\ \nu_B &= \epsilon_B \mu + \nu_B^{MC} + \tau_B \nu_B \\ \nu_C &= \epsilon_C \mu + \nu_C^{MC} + \tau_C \nu_C \\ \nu_D &= \epsilon_D \mu + \nu_D^{MC} + \tau_D \nu_D = \epsilon_D \mu + \nu_D^{MC} + \tau_B \tau_C \nu_D\end{aligned}\tag{44}$$

where μ is the signal rate in region A, ϵ_i is the ratio of the signal in the regions B, C, D with respect to the signal in region A, ν_i^{MC} is the rate of background in each of the regions being estimated from simulation, ν_i is the rate of the background being estimated with the data driven technique in the signal region, and τ_i are the ratios of the background rates in the regions B, C, and D with respect to the background in region A. The key is that we have used the factorization $f_B(x, y) = f_B(x)f_B(y)$ to write $\tau_D = \tau_B \tau_C$. The right panel of Fig. 9 shows a more complicated extension of the ABCD method from a recent ATLAS SUSY analysis [20].

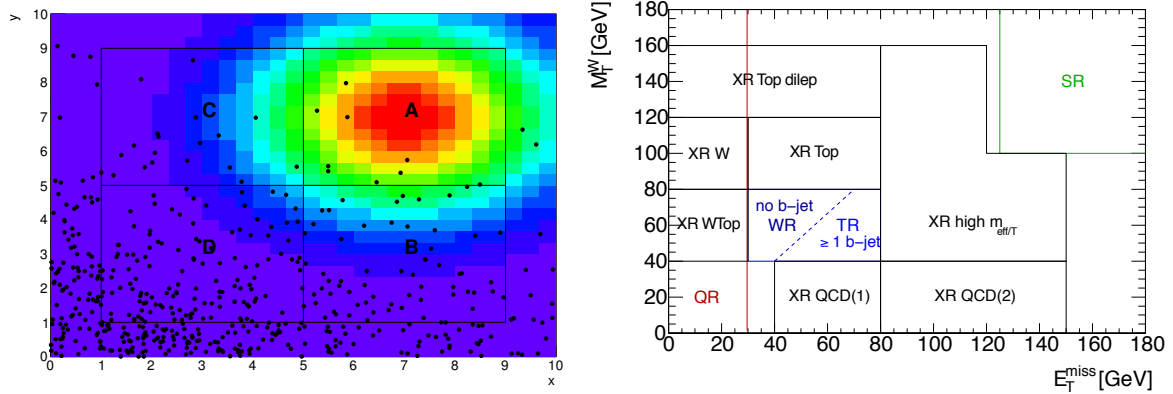


Fig. 9: An example of ABCD (from Alex Read) in the $x - y$ plane of two observables x and y (left). A more complex example with several regions in the $M_T^W - E_T^{\text{miss}}$ plane [20].

4.3 Effective Model Narrative

In the simulation narrative the model of discriminating variable distributions $f(x|\alpha)$ is derived from discrete samples of simulated events $\{x_1, \dots, x_N\}$. We discussed above how one can use histograms or kernel estimation to approximate the underlying distribution and interpolation strategies to incorporate systematic effects. Another approach is to assume a parametric form for the distribution to serve as an

¹⁹The relative sign of the cut is not important, but has been chosen for consistency with Fig. 9.

effective model. For example, in the $H \rightarrow \gamma\gamma$ analysis shown in Fig. 8 a simple exponential distribution was used to model the background. The state-of-the-art theoretical predictions for the continuum $\gamma\gamma$ background process do not predict exactly an exponentially falling distribution, and the analysis must (and does) incorporate the systematic associated to the effective model. Similarly, it is common to use a polynomial in some limited sideband region to estimate backgrounds under a peak. These effective models can range from very ad hoc²⁰ to more motivated. For instance, one might use knowledge of kinematics and phase space and/or detector resolution to construct an effective model that captures the relevant physics. The advantage of a well motivated effective model is that few nuisance parameters may describe well the relevant family of probability densities, which is the challenge for generic (and relatively unsophisticated) interpolation strategies usually employed in the simulation narrative.

4.4 The Matrix Element Method

Ideally, one would not use a single discriminating variable to distinguish the process of interest from the other background processes, but instead would use as much discriminating power as possible. This implies forming a probability model over a multi-dimensional discriminating variable (i.e. a multivariate analysis technique). In principle, both the histogram-based and kernel-based approach generalize to distributions of multi-dimensional discriminating variables; however, in practice, they are limited to only a few dimensions. In the case of histograms this is particularly severe unless one employs clever binning choices, while in the kernel-based approach one can model up to about 5-dimensional distributions with reasonable Monte Carlo sample sizes. In practice, one often uses multivariate algorithms like Neural Networks or boosted decision trees²¹ to map the multiple variables into a single discriminating variable. Often these multivariate techniques are seen as somewhat of a black-box. If we restrict ourselves to discriminating variables associated with the kinematics of final state particles (as opposed to the more detailed signature of particles in the detector), then we can often approximate the detailed simulation of the detector with a parametrized detector response. If we denote the kinematic configuration of all the final state particles in the Lorentz invariant phase space as Φ , the initial state as i , the matrix element (potentially averaged over unmeasured spin configurations) as $\mathcal{M}(i, \Phi)$, and the probability due to parton density functions for the initial state i going into the hard scattering as $f(i)$, then we can write that the distribution of the, possibly multi-dimensional, discriminating variable x as

$$f(x) \propto \int d\Phi f(i) |\mathcal{M}(i, \Phi)|^2 W(x|\Phi), \quad (45)$$

where $W(x|\Phi)$ is referred to as the transfer function of x given the final state configuration Φ . It is natural to think of $W(x|\Phi)$ as a conditional distribution, but here I let W encode the efficiency and acceptance so that we have

$$\frac{\sigma_{\text{eff.}}}{\sigma} = \frac{\int dx \int d\Phi |\mathcal{M}(i, \Phi)|^2 W(x|\Phi)}{\int d\Phi |\mathcal{M}(i, \Phi)|^2}. \quad (46)$$

Otherwise, the equation above looks like another application one Bayes's theorem where $W(x|\Phi)$ plays the role of the PDF/likelihood function and $\mathcal{M}(i, \Phi)$ plays the role of the prior over the Φ . It is worth pointing out that this is a frequentist use of Bayes's theorem since $d\Phi$ is the Lorentz invariant phase space which explicitly has a measure associated with it.

4.5 Event-by-event resolution, conditional modeling, and Punzi factors

In some cases one would like to provide a distribution for the discriminating variable x based conditional on some other observable in the event y : $f(x|\alpha, y)$. For instance, one might want to say that the energy resolution for electrons depends on the energy itself through a well-known calorimeter resolution

²⁰For instance, the modeling of $H \rightarrow ZZ^{(*)} \rightarrow 4l$ described in Ref. [21] (see Eq. 2 of the corresponding section).

²¹A useful toolkit for high-energy physics is TMVA, which is packaged with ROOT [22].

parametrization like $\sigma(E)/E = A/\sqrt{E} + B$. These types of conditional distributions can be built in RooFit. A subtle point studied by Punzi is that if $f(y|\alpha)$ depends on α the inference on α can be biased [23]. In particular, if one is trying to estimate the amount of signal in a sample and the distribution of y for the signal is different than for the background, the estimate of the signal fraction will be biased. This can be remedied by including terms related to $f(y|\alpha)$, colloquially called ‘Punzi Factors’. Importantly, this means one cannot build conditional models like this without knowing or assuming something about $f(y|\alpha)$.

5 Frequentist Statistical Procedures

Here I summarize the procedure used by the LHC Higgs combination group for computing frequentist p -values uses for quantifying the agreement with the background-only hypothesis and for determining exclusion limits. The procedures are based on the profile likelihood ratio test statistic.

The parameter of interest is the overall signal strength factor μ , which acts as a scaling to the total rate of signal events. We often write $\mu = \sigma/\sigma_{SM}$, where σ_{SM} is the standard model production cross-section; however, it should be clarified that the same μ factor is used for all production modes and could also be seen as a scaling on the branching ratios. The signal strength is called so that $\mu = 0$ corresponds to the background-only model and $\mu = 1$ is the standard model signal. It is convenient to separate the full list of parameters α into the parameter of interest μ and the nuisance parameters θ : $\alpha = (\mu, \theta)$.

For a given data set $\mathcal{D}_{\text{sim}}$ and values for the global observables $\mathcal{G}$ there is an associated likelihood function over μ and θ derived from combined model over all the channels including all the constraint terms in Eq. 6

$$L(\mu, \theta; \mathcal{D}_{\text{sim}}, \mathcal{G}) = \mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{sim}}, \mathcal{G}|\mu, \theta) . \quad (47)$$

The notation $L(\mu, \theta)$ leaves the dependence on the data implicit, which can lead to confusion. Thus, we will explicitly write the dependence on the data when the identity of the dataset is important and only suppress $\mathcal{D}_{\text{sim}}, \mathcal{G}$ when the statements about the likelihood are generic.

We begin with the definition of the procedure in the abstract and then describe three implementations of the method based on asymptotic distributions, ensemble tests (Toy Monte Carlo), and importance sampling.

5.1 The test statistics and estimators of μ and θ

These definitions in this section are all relative to a given dataset $\mathcal{D}_{\text{sim}}$ and value of the global observables $\mathcal{G}$, thus we will suppress their appearance. The nomenclature follows from Ref. [1].

The maximum likelihood estimates (MLEs) $\hat{\mu}$ and $\hat{\theta}$ and the values of the parameters that maximize the likelihood function $L(\mu, \theta)$ or, equivalently, minimize $-\ln L(\mu, \theta)$. The dependence of the likelihood function on the data propagates to the values of the MLEs, so when needed the MLEs will be given subscripts to indicate the data set used. For instance, $\hat{\theta}_{\text{obs}}$ is the MLE of θ derived from the observed data and global observables.

The conditional maximum likelihood estimate (CMLEs) $\hat{\hat{\theta}}(\mu)$ is the value of θ that maximizes the likelihood function with μ fixed (see Fig. 10); it can be seen as a multidimensional function of the single variable μ . Again, the dependence on $\mathcal{D}_{\text{sim}}$ and $\mathcal{G}$ is implicit. This procedure for choosing specific values of the nuisance parameters for a given value of μ , $\mathcal{D}_{\text{sim}}$, and $\mathcal{G}$ is often referred to as “profiling”. Similarly, $\hat{\hat{\theta}}(\mu)$ is often called “the profiled value of θ ”.

Given these definitions, we can construct the profile likelihood ratio

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})} , \quad (48)$$

which depends explicitly on the parameter of interest μ , implicitly on the data $\mathcal{D}_{\text{sim}}$ and global observables $\mathcal{G}$, and is independent of the nuisance parameters θ (which have been eliminated via “profiling”).

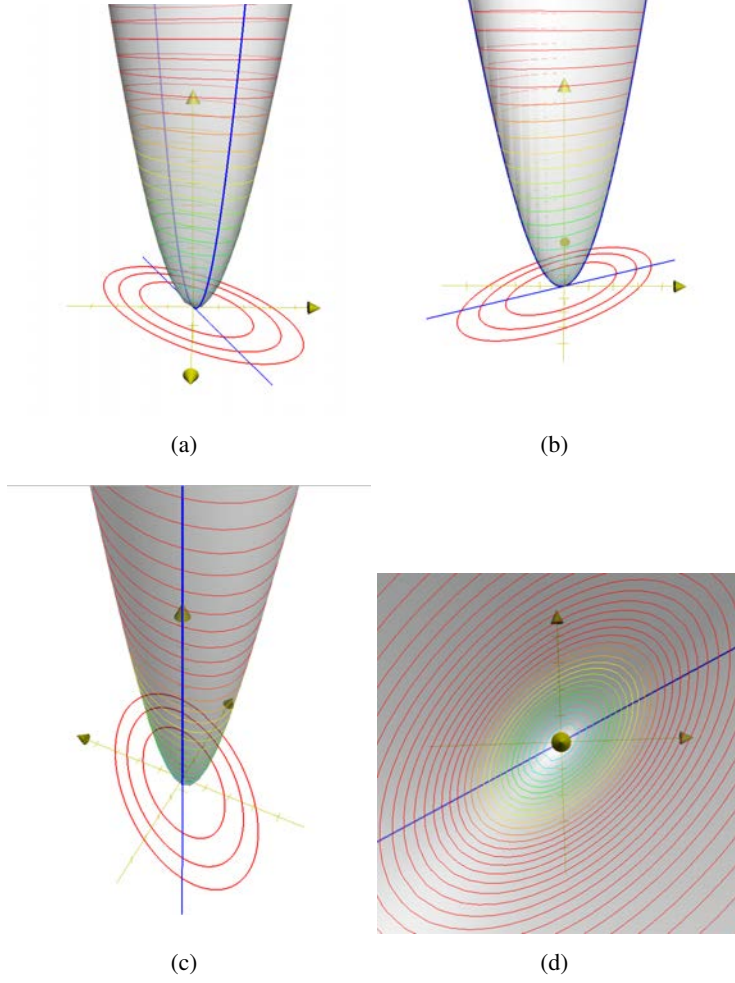


Fig. 10: Visualization of a two dimensional likelihood function $-2 \ln L(\mu, \theta)$. The blue line in the plane represents the profiling operation $\hat{\theta}(\mu)$ and the blue curve along the likelihood surface represents $-2 \ln \lambda(\mu)$. Note it is was to show that the blue line exits the contours of $-2 \ln L(\mu, \theta)$ when they are perpendicular to the μ axis, which provides the correspondence between the profile likelihood ratio and the description of the Minos algorithm.

In any physical theory the rate of signal events is non-negative, thus $\mu \geq 0$. However, it is often convenient to allow $\mu < 0$ (as long as the PDF $f_c(x_c|\mu, \theta) \geq 0$ everywhere). In particular, $\hat{\mu} < 0$ indicates a deficit of events signal-like with respect to the background only and the boundary at $\mu = 0$ complicates the asymptotic distributions. Reference [1] uses a trick that is equivalent to requiring $\mu \geq 0$ while avoiding the formal complications of a boundary, which is to allow $\mu < 0$ and impose the constraint in the test statistic itself. In particular, one defines $\tilde{\lambda}(\mu)$

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})} & \hat{\mu} \geq 0, \\ \frac{L(\mu, \hat{\theta}(\mu))}{L(0, \hat{\theta}(0))} & \hat{\mu} < 0 \end{cases} \quad (49)$$

This is not necessary when ensembles of pseudo-experiments are generated with “Toy” Monte Carlo techniques, but since they are equivalent we will write $\tilde{\lambda}$ to emphasize the boundary at $\mu = 0$.

For discovery the test statistic $\tilde{q}_0$ is used to differentiate the background-only hypothesis $\mu = 0$ from the alternative hypothesis $\mu > 0$:

$$\tilde{q}_0 = \begin{cases} -2 \ln \tilde{\lambda}(\mu) & \hat{\mu} > 0 \\ 0 & \hat{\mu} \leq 0 \end{cases} \quad (50)$$

Note that $\tilde{q}_0$ is test statistic for a one-sided alternative. Note also that if we consider the parameter of interest $\mu \geq 0$, then it is equivalent to the two-sided test (because there are no values of μ less than $\mu = 0$).

For limit setting the test statistic $\tilde{q}_\mu$ is used to differentiate the hypothesis of signal being produced at a rate μ from the alternative hypothesis of signal events being produced at a lesser rate $\mu' < \mu$:

$$\tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} = \begin{cases} -2 \ln \frac{L(\mu, \hat{\theta}(\mu))}{L(0, \hat{\theta}(0))} & \hat{\mu} < 0, \\ -2 \ln \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})} & 0 \leq \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu. \end{cases} \quad (51)$$

Note that $\tilde{q}_\mu$ is a test statistic for a one-sided alternative; it is a test statistic for a one-sided upper limit.

The test statistic $\tilde{t}_\mu$ is used to differentiate signal being produced at a rate μ from the alternative hypothesis of signal events being produced at a lesser or greater rate $\mu' \neq \mu$.

$$\tilde{t}_\mu = -2 \ln \tilde{\lambda}(\mu) . \quad (52)$$

Note that $\tilde{t}_\mu$ is a test statistic for a two-sided alternative (as in the case of the Feldman–Cousins technique, though this is more general as it incorporates nuisance parameters). Note that if we consider the parameter of interest $\mu \geq 0$ and we test at $\mu = 0$ then there is no “other side” and we have $\tilde{t}_{\mu=0} = \tilde{q}_0$. Finally, if one relaxes the constraint $\mu \geq 0$ then the two-sided test statistic is written t_μ or, simply, $-2 \ln \lambda(\mu)$.

5.2 The distribution of the test statistic and p -values

The test statistic should be interpreted as a single real-valued number that represents the outcome of the experiment. More formally, it is a mapping of the data to a single real-valued number: $\tilde{q}_\mu : \mathcal{D}_{\text{sim}}, \mathcal{G} \rightarrow \mathbb{R}$. For the observed data the test statistic has a given value, e.g. $\tilde{q}_{\mu, \text{obs}}$. If one were to repeat the experiment many times the test statistic would take on different values, thus, conceptually, the test statistic has a distribution. Similarly, we can use our model to generate pseudo-experiments using Monte Carlo techniques or more abstractly consider the distribution. Since the number of expected events $\nu(\mu, \theta)$ and the distributions of the discriminating variables $f_c(x_c|\mu, \theta)$ explicitly depend on θ the distribution of the test statistic will also depend on θ . Let us denote this distribution

$$f(\tilde{q}_\mu|\mu, \theta) , \quad (53)$$

and we have analogous expressions for each of the test statistics described above.

The p -value for a given observation under a particular hypothesis (μ, θ) is the probability for an equally or more ‘extreme’ outcome than observed assuming that hypothesis

$$p_{\mu, \theta} = \int_{\tilde{q}_{\mu, \text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu, \theta) d\tilde{q}_\mu . \quad (54)$$

The logic is that small p -values are evidence against the corresponding hypothesis. In Toy Monte Carlo approaches, the integral above is really carried out in the space of the data $\int d\mathcal{D}_{\text{sim}} d\mathcal{G}$.

The immediate difficulty is that we are interested in μ but the p -values depend on both μ and θ . In the frequentist approach the hypothesis $\mu = \mu_0$ would not be rejected unless the p -value is sufficiently

small *for all* values of θ . Equivalently, one can use the supremum p -value for over all θ to base the decision to accept or reject the hypothesis at $\mu = \mu_0$.

$$p_\mu^{\text{sup}} = \sup_{\theta} p_{\mu, \theta} . \quad (55)$$

The key conceptual reason for choosing the test statistics based on the profile likelihood ratio is that asymptotically (i.e. when there are many events) the distribution of the profile likelihood ratio $\lambda(\mu = \mu_{\text{true}})$ is independent of the values of the nuisance parameters. This follows from Wilks's theorem. In that limit $p_\mu^{\text{sup}} = p_{\mu, \theta}$ for all θ .

The asymptotic distributions $f(\lambda(\mu)|\mu, \theta)$ and $f(\lambda(\mu)|\mu', \theta)$ are known and described in Section 5.5. For results based on generating ensembles of pseudo-experiments using Toy Monte Carlo techniques does not assume the form of the distribution $f(\tilde{q}_\mu|\mu, \theta)$, but knowing that it is approximately independent of θ means that one does not need to calculate p -values for all θ (which is not computationally feasible). Since there may still be some residual dependence of the p -values on the choice of θ we would like to know the specific value of θ^{sup} that produces the supremum p -value over θ . Since larger p -values indicate better agreement of the data with the model, it is not surprising that choosing $\theta^{\text{sup}} = \hat{\theta}(\mu)$ is a good estimate of θ^{sup} . This has been studied in detail by statisticians, and is called the Hybrid Resampling method and is referred to in physics as the ‘profile construction’ [8, 11, 24].

Based on the discussion above, the following p -value is used to quantify consistency with the hypothesis of a signal strength of μ :

$$p_\mu = \int_{\tilde{q}_{\mu, \text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}(\mu, \text{obs})) d\tilde{q}_\mu . \quad (56)$$

A standard 95% confidence-level, one-sided frequentist confidence interval (upper limit) is obtained by solving for $p'_{\mu_{\text{up}}} = 5\%$. For downward fluctuations the upper limit of the confidence interval can be arbitrarily small, though it will always include $\mu = 0$. This feature is considered undesirable since a physicist would not claim sensitivity to an arbitrarily small signal rate. The feature was the motivation for the modified frequentist method called CL_s [25–27], and the alternative approach called power-constrained limits [28].

To calculate the CL_s upper limit, we define p'_μ as a ratio of p -values,

$$p'_\mu = \frac{p_\mu}{1 - p_b} , \quad (57)$$

where p_b is the p -value derived from the same test statistic under the background-only hypothesis

$$p_b = 1 - \int_{\tilde{q}_{\mu, \text{obs}}}^{\infty} f(\tilde{q}_\mu|0, \hat{\theta}(\mu = 0, \text{obs})) d\tilde{q}_\mu . \quad (58)$$

The CL_s upper-limit on μ is denoted μ_{up} and obtained by solving for $p'_{\mu_{\text{up}}} = 5\%$. It is worth noting that while confidence intervals produced with the CL_s method over cover, a value of μ is regarded as excluded at the 95% confidence level if $\mu < \mu_{\text{up}}$. The amount of over coverage is not immediately obvious; however, for small values of μ the coverage approaches 100% and for large values of μ the coverage is near the nominal 95% (due to $\langle p_b \rangle \approx 0$).

For the purpose of discovery one is interested in the compatibility of the data with the background-only hypothesis. Statistically, a discovery corresponds to rejecting the background-only hypothesis. This compatibility is based on the following p -value

$$p_0 = \int_{\tilde{q}_{0, \text{obs}}}^{\infty} f(\tilde{q}_0|0, \hat{\theta}(\mu = 0, \text{obs})) d\tilde{q}_0 . \quad (59)$$

This p -value is also based on the background-only hypothesis, but the test statistic $\tilde{q}_0$ is suited for testing the background-only while the test statistic $\tilde{q}_\mu$ in Eq. 58 is suited for testing a hypothesis with signal.

It is customary to convert the background-only p -value into the quantile (or “sigma”) of a unit Gaussian. This conversion is purely conventional and makes no assumption that the test statistic q_0 is Gaussian distributed. The conversion is defined as:

$$Z = \Phi^{-1}(1 - p_0); \quad (60)$$

where Φ^{-1} is the inverse of the cumulative distribution for a unit Gaussian. One says the significance of the result is $Z\sigma$ and the standard discovery convention is 5σ , corresponding to $p_0 = 2.87 \cdot 10^{-7}$.

5.3 Expected sensitivity and bands

The expected sensitivity for limits and discovery are useful quantities, though subject to some degree of ambiguity. Intuitively, the expected upper limit is the upper limit one would expect to obtain if the background-only hypothesis is true. Similarly, the expected significance is the significance of the observation assuming the standard model signal rate (at some m_H). To find the expected limit one needs a distribution $f(\mu_{\text{up}}|\mu = 0, \theta)$. To find the expected significance one needs the distribution $f(Z|\mu = 1, \theta)$ or, equivalently, $f(p_0|\mu = 1, \theta)$. We use the median instead of the mean, as it is invariant to the choice of Z or p_0 . More important is that the expected limit and significance depend on the value of the nuisance parameters θ , for which we do not know the true values. Thus, the expected limit and significance will depend on some convention for choosing θ . While many nuisance parameters have a nominal estimate (i.e. the global observables in the constraint terms), others do not (e.g. the exponent in the $H \rightarrow \gamma\gamma$ background model). Thus, we choose a convention that treats all of the nuisance parameters consistently, which is the profiled value based on the observed data. Thus for the expected limit we use $f(\mu_{\text{up}}|0, \hat{\theta}(\mu = 0, \text{obs}))$ and for the expected significance we use $f(p_0|\mu = 1, \hat{\theta}(\mu = 1, \text{obs}))$. An unintuitive and possibly undesirable feature of this choice is that the expected limit and significance depend on the observed data through the conventional choice for θ .

With these distributions we can also define bands around the median upper limit. Our standard limit plot shows a dark green band corresponding to $\mu_{\pm 1}$ defined by

$$\int_0^{\mu_{\pm 1}} f(\mu_{\text{up}}|0, \hat{\theta}(\mu = 0, \text{obs})) d\mu_{\text{up}} = \Phi^{-1}(\pm 1) \quad (61)$$

and a light yellow band corresponding to $\mu_{\pm 2}$ defined by

$$\int_0^{\mu_{\pm 2}} f(\mu_{\text{up}}|0, \hat{\theta}(\mu = 0, \text{obs})) d\mu_{\text{up}} = \Phi^{-1}(\pm 2) \quad (62)$$

5.4 Ensemble of pseudo-experiments generated with “Toy” Monte Carlo

The p -values in the procedure described above require performing several integrals. In the case of the asymptotic approach, the distributions for $\tilde{q}_\mu$ and $\tilde{q}_0$ are known and the integral is performed directly. When the distributions are not assumed to take on their asymptotic form, then they must be constructed using Monte Carlo methods. In the “toy Monte Carlo” approach one generates pseudo-experiments in which the number of events in each channel n_c , the values of the discriminating variables $\{x_{ec}\}$ for each of those events, and the auxiliary measurements (global observables) a_p are all randomized according to $\mathbf{f}_{\text{tot}}$. We denote the resulting data $\mathcal{D}_{\text{toy}}$ and global observables $\mathcal{G}_{\text{toy}}$. By doing this several times one can build an ensemble of pseudo-experiments and evaluate the necessary integrals. Recall that Monte Carlo techniques can be viewed as a form of numerical integration.

The fact that the auxiliary measurements a_p are randomized is unfamiliar in particle physics. The more familiar approach for toy Monte Carlo is that the nuisance parameters are randomized. This requires a distribution for the nuisance parameters, and thus corresponds to a Bayesian treatment of the

nuisance parameters. The resulting p -values are a hybrid Bayesian-Frequentist quantity with no consistent definition of probability. To maintain a strictly frequentist procedure, the corresponding operation is to randomize the auxiliary measurements.

While formally this procedure is well motivated, as physicists we also know that our models can have deficiencies and we should check that the distribution of the auxiliary measurements does not deviate too far from our expectations.

Technically, the pseudo-experiments are generated with the RooStats `ToyMCSampler`, which is used by the higher-level tool `FrequentistCalculator`, which is in turn used by `HypoTestInverter`.

5.5 Asymptotic Formulas

The following has been extracted from Ref. [1] and has been reproduced here for convenience. The primary message of Ref. [1] is that for a sufficiently large data sample the distributions of the likelihood ratio based test statistics above converge to a specific form. In particular, Wilks's theorem [29] can be used to obtain the distribution $f(\lambda(\mu)|\mu)$, that is the distribution of the test statistic $\lambda(\mu)$ when μ is true. Note that the asymptotic distribution is independent of the value of the nuisance parameters. Wald's theorem [30] provides the generalization to $f(\lambda(\mu)|\mu', \theta)$, that is when the true value is not the same as the tested value. The various formulae listed below are corollaries of Wilks's and Wald's theorems for the likelihood ratio test statistics described above. The Asimov data described immediately below was a novel result of Ref. [1].

5.5.1 The Asimov data and $\sigma = \text{var}(\hat{\mu})$

The asymptotic formulae below require knowing the variance of the maximum likelihood estimate of μ

$$\sigma = \text{var}[\hat{\mu}] . \quad (63)$$

One result of Ref. [1] is that σ can be estimated with an artificial dataset referred to as the *Asimov* dataset. The Asimov dataset is defined as a binned dataset, where the number of events in bin b is exactly the number of events expected in bin b . Note, this means that the dataset generally has non-integer number of events in each bin. For our general model one can write

$$n_{b,A} = \int_{x \in \text{bin } b} \nu(\alpha) f(x|\alpha) dx \quad (64)$$

where the subscript A denotes that this is the Asimov data. Note, that the dataset depends on the value of α implicitly. For a model of unbinned data, one can simply take the limit of narrow bin widths for the Asimov data. We denote the likelihood evaluated with the Asimov data as $L_A(\mu)$. The important result is that one can calculate the expected Fisher information of Eq. 7 by computing the observed Fisher information on the likelihood function based on this special Asimov dataset.

A related and convenient way to calculate the variance of $\hat{\mu}$ is

$$\sigma \sim \frac{\mu}{\sqrt{\tilde{q}_{\mu,A}}} . \quad (65)$$

where $\tilde{q}_{\mu,A}$ is to use the $\tilde{q}_{\mu}$ test statistic based on a background-only Asimov data (i.e. the one with $\mu = 0$ in Eq. 64). It is worth noting that higher-order corrections to the formulae below are being developed to address the case when the variance of $\hat{\mu}$ depends strongly on μ .

5.5.2 Asymptotic Formulas for $\tilde{q}_0$

For a sufficiently large data sample, the PDF $f(\tilde{q}_0|\mu')$ is found to approach

$$f(q_0|\mu') = \left(1 - \Phi\left(\frac{\mu'}{\sigma}\right)\right) \delta(q_0) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{q_0}} \exp\left[-\frac{1}{2} \left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)^2\right] . \quad (66)$$

For the special case of $\mu' = 0$, this reduces to

$$f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2}. \quad (67)$$

That is, one finds a mixture of a delta function at zero and a chi-square distribution for one degree of freedom, with each term having a weight of 1/2. In the following we will refer to this mixture as a half chi-square distribution or $\frac{1}{2}\chi_1^2$.

From Eq. (66) the corresponding cumulative distribution is found to be

$$F(q_0|\mu') = \Phi\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right). \quad (68)$$

The important special case $\mu' = 0$ is therefore simply

$$F(q_0|0) = \Phi\left(\sqrt{q_0}\right). \quad (69)$$

The p -value of the $\mu = 0$ hypothesis is

$$p_0 = 1 - F(q_0|0), \quad (70)$$

and therefore the significance follows from the simple formula

$$Z = \Phi^{-1}(1 - p_0) = \sqrt{q_0}. \quad (71)$$

5.5.3 Asymptotic Formulas for $\tilde{q}_\mu$

For a sufficiently large data sample, the PDF $f(\tilde{q}_\mu|\mu)$ is found to approach

$$\begin{aligned} f(\tilde{q}_\mu|\mu') &= \Phi\left(\frac{\mu' - \mu}{\sigma}\right)\delta(\tilde{q}_\mu) \\ &+ \begin{cases} \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\tilde{q}_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{\tilde{q}_\mu} - \frac{\mu - \mu'}{\sigma}\right)^2\right] & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2 \\ \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\frac{(\tilde{q}_\mu - (\mu^2 - 2\mu\mu')/\sigma^2)^2}{(2\mu/\sigma)^2}\right] & \tilde{q}_\mu > \mu^2/\sigma^2 \end{cases}. \end{aligned} \quad (72)$$

The special case $\mu = \mu'$ is therefore

$$f(\tilde{q}_\mu|\mu) = \frac{1}{2}\delta(\tilde{q}_\mu) + \begin{cases} \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\tilde{q}_\mu}}e^{-\tilde{q}_\mu/2} & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2 \\ \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\frac{(\tilde{q}_\mu + \mu^2/\sigma^2)^2}{(2\mu/\sigma)^2}\right] & \tilde{q}_\mu > \mu^2/\sigma^2. \end{cases} \quad (73)$$

The corresponding cumulative distribution is

$$F(\tilde{q}_\mu|\mu') = \begin{cases} \Phi\left(\sqrt{\tilde{q}_\mu} - \frac{\mu - \mu'}{\sigma}\right) & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2, \\ \Phi\left(\frac{\tilde{q}_\mu - (\mu^2 - 2\mu\mu')/\sigma^2}{2\mu/\sigma}\right) & \tilde{q}_\mu > \mu^2/\sigma^2. \end{cases} \quad (74)$$

The special case $\mu = \mu'$ is

$$F(\tilde{q}_\mu|\mu) = \begin{cases} \Phi\left(\sqrt{\tilde{q}_\mu}\right) & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2, \\ \Phi\left(\frac{\tilde{q}_\mu + \mu^2/\sigma^2}{2\mu/\sigma}\right) & \tilde{q}_\mu > \mu^2/\sigma^2. \end{cases} \quad (75)$$

The p -value of the hypothesized μ is as before given by one minus the cumulative distribution,

$$p_\mu = 1 - F(\tilde{q}_\mu|\mu) . \quad (76)$$

As when using q_μ , the upper limit on μ at confidence level $1 - \alpha$ is found by setting $p_\mu = \alpha$ and solving for μ , which reduces to the same result as found when using q_μ , namely,

$$\mu_{\text{up}} = \hat{\mu} + \sigma \Phi^{-1}(1 - \alpha) . \quad (77)$$

Note that because σ depends in general on μ , Eq. (77) must be solved numerically.

5.5.4 Expected CL_s Limit and Bands

For the CL_s method we need distributions for $\tilde{q}_\mu$ for the hypothesis at μ and $\mu = 0$. We find

$$p'_\mu = \frac{1 - \Phi(\sqrt{q_\mu})}{\Phi(\sqrt{q_{\mu,A}} - \sqrt{q_\mu})} . \quad (78)$$

The median and expected error bands will therefore be

$$\mu_{\text{up}+N} = \sigma(\Phi^{-1}(1 - \alpha\Phi(N)) + N) \quad (79)$$

with

$$\sigma^2 = \frac{\mu^2}{q_{\mu,A}} , \quad (80)$$

and μ can be taken as $\mu_{\text{up}}^{\text{med}}$ in the calculation of σ . Note that for $N = 0$ we find the median limit

$$\mu_{\text{up}}^{\text{med}} = \sigma \Phi^{-1}(1 - 0.5\alpha) . \quad (81)$$

The fact that σ (the variance of $\hat{\mu}$) defined in Eq. 65 in general depends on μ complicates situations and can lead to some discrepancies between the correct value of the bands and those obtained with the equation above. The bands tend to be too narrow. A modified treatment of the bands taking into account the μ dependence of σ is under development.

5.6 Importance Sampling

[The following section has been adapted from a text written primarily by Sven Kreiss, Alex Read, and myself for the ATLAS Higgs combination. It is reproduced here for convenience.]

To claim a discovery, it is necessary to populate a small tail of a test statistic distribution. Toy Monte Carlo techniques use the model $\mathbf{f}_{\text{tot}}$ to generate toy data $\mathcal{D}_{\text{toy}}$. For every pseudo-experiment (toy), the test statistic is calculated and added to the test statistic distribution. Building this distribution from toys is independent of the assumptions that go into the asymptotic calculation that describes this distribution with an analytic expression. Recently progress has been made using Importance Sampling to populate the extreme tails of the test statistic distribution, which is much more computationally intensive with standard methods. The presented algorithms are implemented in RooStats ToyMCSampler.

5.6.1 Naive Importance Sampling

An ensemble of "standard toys" is generated from a model representing the Null hypothesis with $\mu = 0$ and the nuisance parameters θ fixed at their profiled values to the observed data θ_{obs} , written as $\mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{sim}}, \mathcal{G}|\mu = 0, \theta_{\text{obs}})$. With importance sampling however, the underlying idea is to generate toys from a different model, called the importance density. A valid importance density is for example the

same model with a non-zero value of μ . The simple Likelihood ratio is calculated for each toy and used as a weight:

$$\text{weight} = \frac{\mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{toy}}, \mathcal{G}_{\text{toy}} | \mu = 0, \boldsymbol{\theta}_{\text{obs}})}{\mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{toy}}, \mathcal{G}_{\text{toy}} | \mu = \mu', \boldsymbol{\theta}_{\text{obs}})}$$

The weighted distribution is equal to a distribution of unweighted toys generated from the Null. The choice of the importance density is a delicate issue. Michael Woodroffe presented a prescription for creating a well behaved importance density [31]. Unfortunately, this method is impractical for models as large as the combined Higgs models. An alternative approach is shown below.

5.6.2 Phase Space Slicing

The first improvement from naive importance sampling is the idea of taking toys from both, the null density and the importance density. There are various ways to do that. Simply stitching two test statistic distributions together at an arbitrary point has the disadvantage that the normalizations of both distributions have to be known.

Instead, it is possible to select toys according to their weights. First, toys are generated from the Null and the simple Likelihood ratio is calculated. If it is larger than one, the toy is kept and otherwise rejected. Next, toys from the importance density are generated. Here again, the simple Likelihood ratio is calculated but this time the toy is rejected when the Likelihood ratio is larger than one and kept when it is smaller than one. If kept, the toy's weight is the simple Likelihood ratio which is smaller than one by this prescription.

In the following section, this idea is restated such that it generalizes to multiple importance densities.

5.6.3 Multiple Importance Densities

The above procedure for selecting and reweighting toys that were generated from both densities can be phrased in the following way:

- A toy is generated from a density with $\mu = \mu'$ and the Likelihoods $\mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{toy}}, \mathcal{G}_{\text{toy}} | \mu = 0, \boldsymbol{\theta}_{\text{obs}})$ and $\mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{toy}}, \mathcal{G}_{\text{toy}} | \mu = \mu', \boldsymbol{\theta}_{\text{obs}})$ are calculated.
- The toy is vetoed when the Likelihood with $\mu = \mu'$ is not the largest. Otherwise, the toy is used with a weight that is the ratio of the Likelihoods.

This can be generalized to any number of densities with $\mu_i = \{0, \mu', \mu'', \dots\}$. For the toys generated from model i :

$$\text{veto: if } \mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{toy}}, \mathcal{G}_{\text{toy}} | \mu = \mu_i, \boldsymbol{\theta}_{\text{obs}}) \neq \max \{ \mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{toy}}, \mathcal{G}_{\text{toy}} | \mu = \mu_j, \boldsymbol{\theta}_{\text{obs}}) : \mu_j = \{0, \mu', \mu'', \dots\} \} \quad (82)$$

$$\text{weight} = \frac{\mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{toy}}, \mathcal{G}_{\text{toy}} | \mu = 0, \boldsymbol{\theta}_{\text{obs}})}{\mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{toy}}, \mathcal{G}_{\text{toy}} | \mu = \mu_i, \boldsymbol{\theta}_{\text{obs}})} . \quad (83)$$

The number of importance densities has to be known when applying the vetos. It should not be too small to cover the parameter space appropriately and it should not be too large, because too many importance densities lead to too many vetoed toys which decreases overall efficiency. The value and error of $\hat{\mu}$ from a fit to data can be used to estimate the required number of importance densities for a given target overlap of the distributions. See Fig. 11 for an example.

The sampling efficiency in the tail can be further improved by generating a larger number of toys for densities with larger values of μ . For example, for n densities, one can generate $2^k / 2^n = 2^{k-n}$ of the overall toys per density k with $k = 0, \dots, n-1$. The toys have to be re-weighted for example

by $2^{n-1}/2^k$ resulting in a minimum re-weight factor of one. The current implementation of the error calculation for the p -value is independent of an overall scale in the weights.

The method using multiple importance densities is similar to Michael Woodroffe’s [31] prescription of creating a suitable importance density with an integral over μ . In the method presented here, the integral is approximated by a sum over discrete values of μ . Instead of taking the sum, a mechanism that allows for multiple importance densities is introduced.

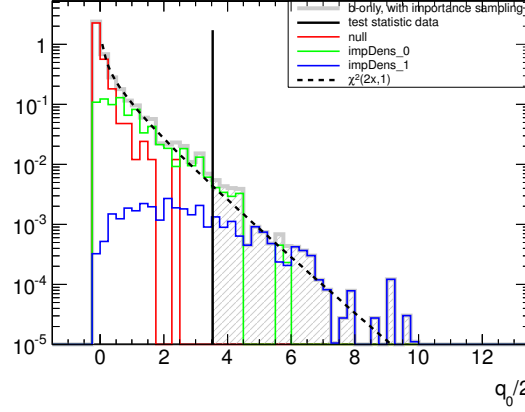


Fig. 11: An example sampling of a test statistic distribution using three densities, the original null density and two importance densities.

5.7 Look-elsewhere effect, trials factor, Bonferoni

Future versions of this document will discuss the so-called look-elsewhere effect in more detail. Here we point to the primary development recently: Refs. [32, 33].

5.8 One-sided intervals, CLs, power-constraints, and Negatively Biased Relevant Subsets

Particle physicists regularly set upper-limits on cross sections and other parameters that are bounded to be non-negative. Standard frequentist confidence intervals should nominally cover at the stated value. The implication that a 95% confidence level upper-limit covers the true value 95% of the time is that it doesn’t cover the true value 5% of the time. This is true no matter how small the cross section is. That means that if there is no signal present, 5% of the time we would be excluding any positive value of the cross-section. Experimentalists do not like this since we would not consider ourselves sensitive to arbitrarily small signals.

Two main approaches have been proposed to protect from excluding signals to which we do not consider ourselves sensitive. The first is the CL_s procedure introduced by Read [25–27] and described above. The CL_s procedure produces intervals that over-cover – meaning that the intervals cover the true value more than the desired level. The coverage for small values of the cross-section approaches 100%, while for large values of the cross section, where the experiment does have sensitivity, the coverage converges to the nominal level (see Fig. 12). Unfortunately, the coverage for intermediate values is not immediately accessible without more detailed studies. Interestingly, the modified frequentist CL_s procedure reproduces the one-sided upper limit from a Bayesian procedure with a uniform prior on the cross section for simple models like number counting analyses. Even in very complicated models we see very good numerical agreement between CL_s and the Bayesian approach, even though the interpretation of the numbers is different.

An alternate approach called power-constrained limits (PCL) is to leave the standard frequentist procedure unchanged while adding an additional requirement for a parameter point to be considered

‘excluded’. The additional requirement is directly a measure of the sensitivity of to that parameter point based on the notion of power (or Type II error). This approach makes the coverage of the procedure manifest [28].

Surprisingly, one-sided upper limits on a bounded parameter are a subtle topic that has led to debates among the experts of statistics in the collaborations and a string of interesting articles from statisticians. The discussion is beyond the scope of the current version of these notes, but the interested reader is invited and encouraged to read Ref. [34] and the responses from notable statisticians on the topic. More recently Cousins tried to formalize the sensitivity problem in terms of a concept called Negatively Biased Relevant Subsets (NBRS) [35]. While the power-constrained limits do not formally emit NBRS, it is an interesting insight. Even more recently, Vitells has found interesting connections with CL_s and the work of Birnbaum [27, 36]. This connection is significant since statisticians have primarily seen CL_s as an ad hoc procedure mixing the notion of size and power with no satisfying properties.

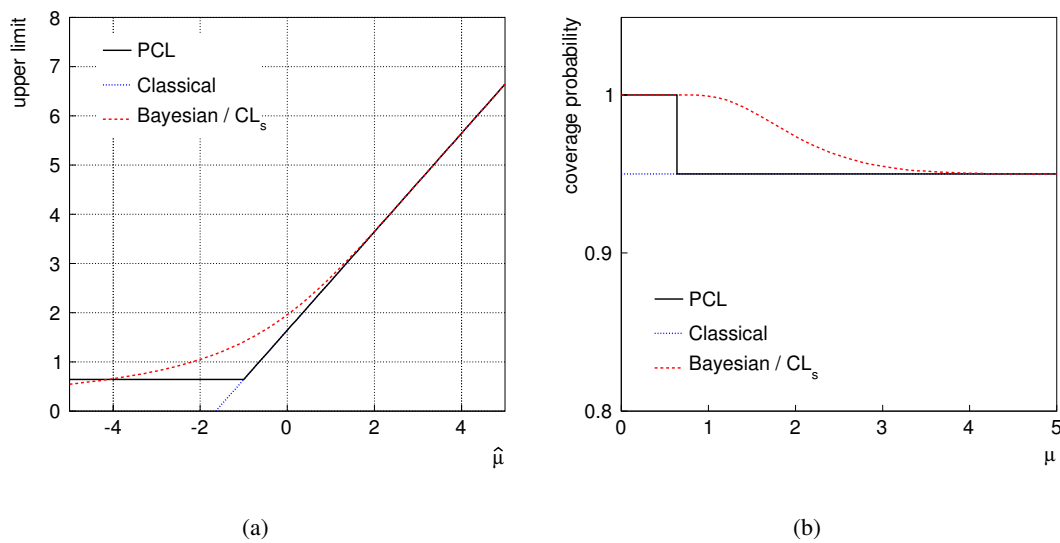


Fig. 12: Taken from Fig.3 of [28]: (a) Upper limits from the PCL (solid), CL_s and Bayesian (dashed), and classical (dotted) procedures as a function of μ , which is assumed to follow a Gaussian distribution with unit standard deviation. (b) The corresponding coverage probabilities as a function of μ .

6 Bayesian Procedures

[This section is far from complete. Some key practical issues and references to other literature are given.]

Unsurprisingly, Bayesian procedures are based on Bayes’s theorem as in Eq. 3 and Eq. 5. The Bayesian approach requires one to provide a prior over the parameters, which can be seen either as an advantage or a disadvantage [37, 38]. In practical terms, one typically wants to build the posterior distribution for the parameter of interest. This typically requires integrating, or *marginalizing*, over all the nuisance parameters as in Eq. 14. These integrals can be over very high dimensional posteriors with complicated structure. One of the most powerful algorithms for this integration is Markov Chain Monte Carlo, described below. In terms of the prior one can either embrace the subjective Bayesian approach [39] or take a more ‘objective’ approach in which the prior is derived from formal rules. For instance, Jeffreys’s Prior [40] or their generalization in terms of Reference Priors [41].

Given the logical importance of the choice of prior, it is generally recommended to try a few options to see how the result numerically depends on the choice of priors (i.e. sensitivity analysis). This leads me to a few great quotes from prominent statisticians:

“Sensitivity analysis is at the heart of scientific Bayesianism”– Michael Goldstein

“Perhaps the most important general lesson is that the facile use of what appear to be uninformative priors is a dangerous practice in high dimensions”– Brad Efron

“Meaningful prior specification of beliefs in probabilistic form over very large possibility spaces is very difficult and may lead to a lot of arbitrariness in the specification”– Michael Goldstein

“Objective Bayesian analysis is the best frequentist tool around”– Jim Berger

6.1 Hybrid Bayesian-Frequentist methods

It is worth mentioning that in particle physics there has been widespread use of a hybrid Bayesian-Frequentist approach in which one marginalizes nuisance parameters. Perhaps the most well known example is due to a paper by Cousins and Highland [42]. In some instances one obtains a Bayesian-averaged model that depends only on the parameters of interest

$$\bar{\mathbf{f}}(\mathcal{D}|\alpha_{\text{poi}}) = \int \mathbf{f}_{\text{tot}}(\mathcal{D}|\alpha) \eta(\alpha_{\text{nuis}}) d\alpha_{\text{nuis}} \quad (84)$$

and then proceeds with the typical frequentist methodology for calculating p -values and constructing confidence intervals. Note, in this approach the constraint terms that are appended to $\mathbf{f}_{\text{sim}}$ of Eq. 2 to obtain $\mathbf{f}_{\text{tot}}$ of Eq. 6 are interpreted as in Eq. 5 and $\eta(\alpha_{\text{nuis}})$ is usually a uniform prior. Furthermore, the global observables or auxiliary measurements a_p are typically left fixed to their nominal or observed values and not randomized. In other variants the full model without constraints $\mathbf{f}_{\text{sim}}(\mathcal{D}|\alpha)$ is used to define the test statistic but the distribution of the test statistic is obtained by marginalizing (or randomizing) the nuisance parameters as in Eq. 5. See the following references [4, 43–49] for more details.

The shortcomings of this approach are that the coverage is not guaranteed and the method uses an inconsistent notion of probability. Thus it is hard to define exactly what the p -values and intervals mean in a formal sense.

6.2 Markov Chain Monte Carlo and the Metropolis–Hastings Algorithm

The Metropolis–Hastings algorithm is used to construct a Markov chain $\{\alpha_i\}$, where the samples α_i are proportional to the target posterior density or likelihood function. The algorithm requires a proposal function $Q(\alpha|\alpha')$ that gives the probability density to propose the point α given that the last point in the chain is α' . Note, the density only depends on the last step in the chain, thus it is considered a Markov process. At each step in the algorithm, a new point in parameter space is proposed and possibly appended to the chain based on its likelihood relative to the current point in the chain. Even when the proposal density function is not symmetric, Metropolis–Hastings maintains ‘detailed balance’ when constructing the Markov chain by counterbalancing the relative likelihood between the two points with the relative proposal density. That is, given the current point α , proposed point α' , likelihood function L , and proposal density function Q , we visit α' if and only if

$$\frac{L(\alpha')}{L(\alpha)} \frac{Q(\alpha|\alpha')}{Q(\alpha'|\alpha)} \geq \text{Rand}[0, 1] \quad (85)$$

Note, if the proposal density is symmetric, $Q(\alpha|\alpha') = Q(\alpha'|\alpha)$, then the ratio of the proposal densities can be neglected (which can be computationally expensive). Above we have written the algorithm to sample the likelihood function $L(\alpha)$, but typically one would use the posterior $\pi(\alpha)$. Within RooStats the Metropolis–Hastings algorithm is implemented with the `MetropolisHastings` class, which returns a `MarkovChain`. Another powerful tool is the Bayesian Analysis Toolkit (BAT) [50]. Note, one can use a RooFit / RooStats model in the BAT environment.

An alternative to Markov Chain Monte Carlo is the nested sampling approach of Skilling [51] and the MultiNest implementation [52].

Lastly, we mention that sampling algorithms associated to Bayesian belief networks and graphical models may offer enormous advantages to both MCMC and nested sampling due to the fact that they can take advantage of the conditional dependencies in the model.

6.3 Jeffreys's and Reference Prior

One of the great advances in Bayesian methodology was the introduction of Jeffreys's rule for selecting a prior based on a formal rule [40]. The rule selects a prior that is invariant under reparametrization of the observables and covariant with reparametrization of the parameters. The rule is based on information theoretic arguments and the prior is given by the square root of the determinant of the Fisher information matrix, which we first encountered in Eq. 7.

$$\pi(\alpha) = \sqrt{\det \Sigma_{pp'}^{-1}(\alpha)} = \sqrt{\det \left[\int \mathbf{f}_{\text{tot}}(\mathcal{D}|\alpha) \frac{-\partial^2 \log \mathbf{f}_{\text{tot}}(\mathcal{D}|\alpha)}{\partial \alpha_p \partial \alpha_{p'}} d\mathcal{D} \right]} \quad (86)$$

While the right-most form of the prior looks daunting with complex integrals over partial derivatives, the Asimov data described in Section 5.5.1 and Ref. [1] provide a convenient way to calculate the Fisher information. Figures 13 and 14 show examples of RooStats numerical algorithm for calculating Jeffreys's prior compared to analytic results on a simple Gaussian and a Poisson model.

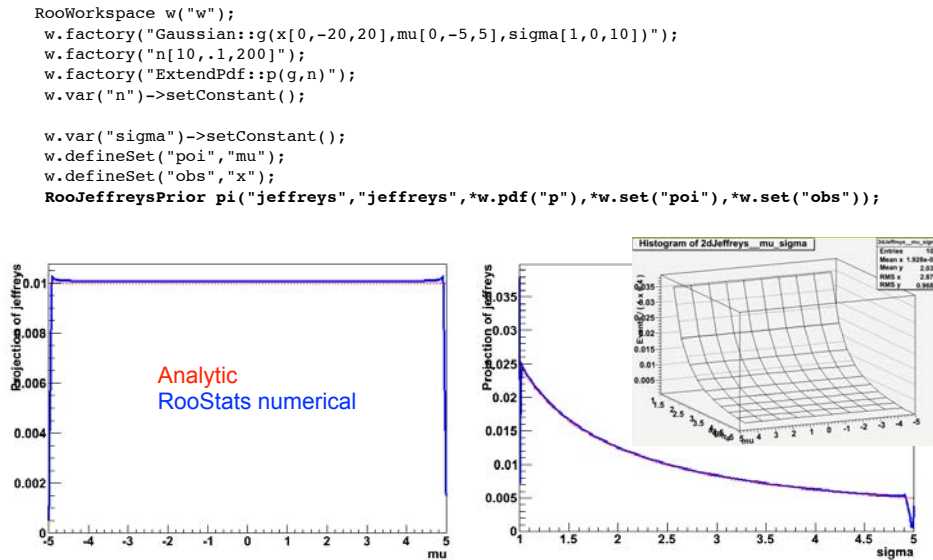


Fig. 13: Example code making a Gaussian distribution (with 10 events expected) and the Jeffreys Prior for μ and σ calculated numerically in RooStats and compared to the analytic result.

Unfortunately, Jeffreys's prior does not behave well in multidimensional problems. Based on a similar information theoretic approach, Bernardo and Berger have developed the Reference priors [53–56] and the associated Reference analysis. While attractive in many ways, the approach is fairly difficult to implement. Recently, there has been some progress within the particle physics context in deriving the reference prior for problems relevant to particle physics [41, 57].

6.4 Likelihood Principle

For those interested in the deeper and more philosophical aspects of statistical inference, the likelihood principle is incredibly interesting. This section will be expanded in the future, but for now I simply suggest searching on the internet, the Wikipedia article, and Ref. [36]. In short the principle says that

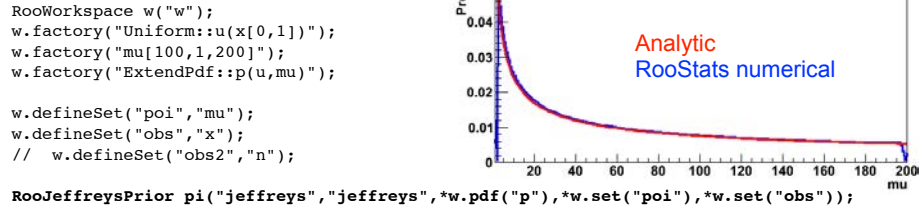


Fig. 14: Example code making a Poisson distribution (with 100 replications expected) and the Jeffreys Prior for μ calculated numerically in RooStats and compared to the analytic result.

all inference should be based on the likelihood function of the observed data. Frequentist procedures violate the likelihood principle since p -values are tail probabilities associated to hypothetical outcomes (not the observed data). Generally, Bayesian procedures and those based on the asymptotic properties of likelihood tests do obey the likelihood principle. Somewhat ironically, the objective Bayesian procedures such as Reference priors and Jeffreys's prior can violate the likelihood principle since the prior is based on expectations over hypothetical outcomes.

7 Unfolding

Another topic for the future. The basic aim of unfolding is to try to correct distributions back to the true underlying distribution before detector 'smearing'. For now, see Refs. [58–64].

8 Conclusions

It was a pleasure to lecture at the Cheile Gradistei school on High-Energy Physics and present the recent developments and practice in statistics for particle physics. Quite a bit of progress has been made in the last few years in terms of statistical methodology, in particular the formalization of a fully frequentist approach to incorporating systematics, a deeper understanding of the look-elsewhere effect, the development of asymptotic approximations of the distributions important for particle physics, and in roads to Bayesian reference analysis. Furthermore, most of these developments are general purpose and can be applied across diverse models. While those developments are interesting, the most important area for most physicists to devote their attention in terms of statistics is to improve the modeling of the data for his or her individual analysis.

References

- [1] G. Cowan, K. Cranmer, E. Gross, O. Vitells, *Eur. Phys. J.* **C71** (2011) 1554.
- [2] K Cranmer and G. Lewis, *The histfactory users guide*, 2011, <https://twiki.cern.ch/twiki/pub/RooStats>.
- [3] R. Barlow, *Nucl. Instrum. Meth.* **A297** (1990) 496.
- [4] R. D. Cousins, J. T. Linnemann, and J. Tucker, *Nucl. Instrum. Meth.* **A595** (2008) 480.
- [5] F. James and M. Roos, *Nucl. Phys.* **B172** (1980) 475.
- [6] F. James and M. Roos, *Comput. Phys. Commun.* **10** (1975) 343.
- [7] J. Neyman, *Phil. Trans. Royal Soc. London* **A236** (1937) 333.
- [8] G. Feldman, *Multiple measurements and parameters in the unified approach*, Technical report, 2000.

- [9] K. Cranmer, *Statistical challenges for searches for new physics at the LHC*, arXiv:physics/0511028.
- [10] C. Chuang and T. L. Lai, *Statist. Sinica* **10** (2001) 1.
- [11] M. Walker, B. Sen and M. Woodroffe, On the unified method with nuisance parameters. *Statist. Sinica* **19** (2009) 301.
- [12] G. J. Feldman and R. D. Cousins, *Phys. Rev.* **D57** (1998) 3873.
- [13] K. S. Cranmer, *Comput. Phys. Commun.* **136** (2001) 198.
- [14] A. L. Read, *Nucl. Instrum. Meth.* **A425** (1999) 357.
- [15] R. Cousins, *Probability density functions for positive nuisance parameters*, http://www.physics.ucla.edu/~cousins/stats/cousins_lognormal_prior.pdf.
- [16] W.T Eadie, D Drijard, F.E. James, M Roos, and B Sadoulet, *Statistical methods in experimental physics*, American Elsevier Pub. Co, 1997.
- [17] R. J. Barlow and C. Beeston, *Comput. Phys. Commun.* **77** (1993) 219.
- [18] G. Aad *et al.* [ATLAS Collaboration], *Search for the standard model higgs boson in the diphoton decay channel with 4.9fb^{-1} of ATLAS data at $\sqrt{s} = 7\text{ TeV}$* , ATLAS-CONF-2011-161 (2011).
- [19] G. Aad *et al.* [ATLAS Collaboration], *Phys. Rev. Lett.* **108** (2012) 111803.
- [20] G. Aad *et al.* [ATLAS Collaboration], *Phys. Rev.* **D85** (2012) 012006.
- [21] G. Aad *et al.* [ATLAS Collaboration], *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, arXiv:0901.0512 [hep-ex].
- [22] A. Hocker, J. Stelzer, F. Tegenfeldt, H. Voss, K. Voss, *et al.*, *PoS ACAT* (2007) 040,
- [23] G. Punzi, *eConf C* **030908** (2003) WELT002.
- [24] K. Cranmer, *Statistics for the LHC: Progress, Challenges and Future*, Proc. PHYSTAT-LHC Workshop on Statistical Issues for LHC Physics, 2008.
- [25] A. L. Read, *Modified frequentist analysis of search results (the CL_s method)*, Proc. First Workshop on Confidence Limits, CERN, Geneva, Switzerland, 2000.
- [26] A. L. Read, Presentation of search results: the CL_s technique. *J. Phys.* **G28** (2002) 2693.
- [27] Wikipedia, *CL_s upper limits*, http://en.wikipedia.org/wiki/CLs_upper_limits.
- [28] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Power-Constrained Limits*, arXiv:1105.3166 [physics.data-an].
- [29] S.S. Wilks, *Ann. Math. Statist.* **9** (1938) 60.
- [30] A. Wald, *Transactions of the American Mathematical Society* **54** (1943) 426.
- [31] M Woodroffe, *Importance Sampling and Error Probabilities*, (2010), <http://people.stat.sfu.ca/~lockhart/richard/banff2010/woodroffe.pdf>
- [32] E. Gross and O. Vitells, *EPJC* **70** (2010) 525.
- [33] ATLAS Collaboration, *Procedure for the lhc higgs boson search combination in summer 2011*, ATL-PHYS-PUB-2011-011.
- [34] M. Mandelkern, *Statistical Science*, **17** (2002) 149.
- [35] R. D. Cousins, *Negatively Biased Relevant Subsets Induced by the Most-Powerful One-Sided Upper Confidence Limits for a Bounded Physical Parameter*, arXiv: 1109.2023 [physics].
- [36] A. Birnbaum, *Journal of the American Statistical Association* **57** (1962) 269.
- [37] G D'Agostini, *Reports on Progress in Physics* **66** (2003) 1383.
- [38] R. D. Cousins, *Am. J. Phys.* **63** (1995) 398.
- [39] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.
- [40] H. Jeffreys, *Proc. of the Royal Society of London* **186** (1946) 453.
- [41] L. Demortier, S. Jain, and H. B. Prosper, *Phys. Rev.* **D82** (2010) 034002.
- [42] , R.D. Cousins, and V.L. Highland, *Nucl. Instrum. Meth.* **A320** (1992) 331.

- [43] J. Conrad and F. Tegenfeldt, *Likelihood ratio intervals with Bayesian treatment of uncertainties: Coverage, power and combined experiments*, arXiv:physics/0511055.
- [44] F. Tegenfeldt and J. Conrad, *Nucl. Instrum. Meth.* **A539** (2005) 407.
- [45] J. Conrad, O. Botner, A. Hallgren, and C.P. de los Heros, *Coverage of confidence intervals for Poisson statistics in presence of systematic uncertainties*, arXiv:hep-ex/0206034.
- [46] J. Conrad, O. Botner, A. Hallgren, and C. P. de los Heros, *Phys. Rev.* **D67** (2003) 012002.
- [47] W. A. Rolke, A. M. Lopez, and J. Conrad, *Nucl. Instrum. Meth.* **A551** (2005) 493.
- [48] G. C. Hill, *Phys. Rev.* **D67** (2003) 118101.
- [49] L. Demortier, *P values and nuisance parameters*, CERN-2008-001.
- [50] A Caldwell, D. Kollar, and K. Kroeninger, *Comput. Phys. Commun.* **180** (2009) 2197.
- [51] J. Skilling, *AIP Conference Proceedings* **735** (2004) 1.
- [52] F. Feroz, M.P. Hobson, and M. Bridges, *Mon. Not. Roy. Astron. Soc.* **398** (2009) 1601.
- [53] J.O. Berger and J. M. Bernardo, *Bayesian Statistics 4*, Oxford University Press, 1992.
- [54] J.O. Berger and J. M. Bernardo, *Biometrika* **79** (1992) 25.
- [55] J.O. Berger and J. M. Bernardo, *J. Amer. Statist. Assoc.* **84** (1989) 200.
- [56] J. M. Bernardo, *J. R. Statist. Soc.* **B41** (1979) 113.
- [57] D. Casadei, *JINST* **7** (2012) P01012.
- [58] H. B. Prosper, *Experience from Searches at the Tevatron*, in *Proc. PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, CERN, 2011
- [59] G. D'Agostini, *Nuclear Instruments and Methods in Physics Research* **A362** (1995) 487.
- [60] T. Adye, *Unfolding algorithms and tests using RooUnfold*, arXiv:1105.1160 [physics.data-an].
- [61] B. Malaescu, *An Iterative, Dynamically Stabilized(IDS) Method of Data Unfolding*, arXiv:1106.3107 [physics.data-an].
- [62] V. Blobel, *An Unfolding method for high-energy physics experiments*, arXiv:hep-ex/0208022.
- [63] A Hocker and V Kartvelishvili. *Nucl. Instrum. Meth.* **A372** (1996) 469.
- [64] A. N. Tikhonov. *Sov. Math.* **5** (1963) 1035.

Physics at the LHC – From Standard Model Measurements to Searches for New Physics

K. Jakobs

Physikalisches Institut, University of Freiburg, Germany

Abstract

The successful operation of the *Large Hadron Collider* (LHC) during the past two years allowed to explore particle interaction in a new energy regime. Measurements of important Standard Model processes like the production of high- p_T jets, W and Z bosons and top and b -quarks were performed by the LHC experiments. In addition, the high collision energy allowed to search for new particles in so far unexplored mass regions. Important constraints on the existence of new particles predicted in many models of physics beyond the Standard Model could be established. With integrated luminosities reaching values around 5 fb^{-1} in 2011, the experiments reached as well sensitivity to probe the existence of the Standard Model Higgs boson over a large mass range. In the present report the major physics results obtained by the two general-purpose experiments ATLAS and CMS are summarized.

1 Introduction

In March 2010 the Large Hadron Collider started its operation at the highest centre-of-mass energy ever reached and delivered first proton-proton collisions at 7 TeV. The years 2010 and 2011 showed a very successful operation of both the collider and the associated experiments. During the start-up year 2010 data corresponding to an integrated luminosity of about 48 pb^{-1} could be delivered. This successful operation was followed by an even more successful year 2011 where the collider delivered data corresponding to an integrated luminosity of 5.5 fb^{-1} and exceeded the original design goal of 1 fb^{-1} by far. In April 2011 the world record on the instantaneous luminosity was reached with a luminosity of $4.7 \cdot 10^{32} \text{ cm}^{-2}\text{sec}^{-1}$. Meanwhile luminosities beyond $3 \cdot 10^{33} \text{ cm}^{-2}\text{sec}^{-1}$ have been reached.

However, not only the accelerator but also the experiments showed an extremely successful operation. They were able to record the delivered luminosity with efficiencies of the order of 94%. All detector subsystems worked well with a high number of functioning channels, typically exceeding 99%.

The data were used to test the Standard Model [1–6] of particle physics in the new energy regime. At the LHC as a hadron collider the production of particles via the strong interaction is dominating. Therefore, the test of Quantum Chromodynamics (QCD) [4–6], the theory of strong interactions, was in the focus during the early phase. Tests of QCD can be performed at small distances or for processes with large momentum transfer. Among them the production of jets with large transverse momenta (p_T) has the largest cross section. The investigation of the production of W and Z bosons, their associated production with jets and the production of top quarks constitute other important tests of QCD in the new energy regime.

Due to the high centre-of-mass energy of 7 TeV the LHC has a large discovery potential for new particles with masses beyond the limits set by the Tevatron experiments, even already in the initial phase of operation. Due to the dominating strong production, this holds in particular for particles that carry colour charge, like e.g. the supersymmetric partners of quarks and gluons. Due to the excellent luminosity performance of the LHC in 2011 the sensitivity for many models of new physics were pushed far beyond the mass range explored so far.

In this review the physics motivation for the LHC is briefly recalled in Section 2. The phenomenology of proton-proton collisions and the calculation of cross sections is briefly reviewed in Section 3. The

measurement of important Standard Model processes at the LHC is discussed in Section 4. The status of the search for the Standard Model Higgs boson is summarized in Section 5. It should be noted that in this paper the status of the Higgs boson search as of March 2012 is presented. Given the large increase in the integrated luminosity during the second half of 2011, these results supersede by far those presented at the school in September 2011. In the remaining sections of the paper the searches for physics beyond the Standard Model are discussed. The search for supersymmetric particles is described in Section 6, the search for other scenarios is summarized in Section 7.

2 The Physics Questions at the LHC

The Standard Model is a very successful description of the interactions of particles at the smallest scales (10^{-18} m) and highest energies accessible to current experiments. It is a quantum field theory which describes the interactions of spin-1/2 pointlike fermions whose interactions are mediated by spin-1 gauge bosons. The bosons are a consequence of local gauge invariance of the underlying Lagrangian under the symmetry group $SU(3) \times SU(2) \times U(1)$ [1–3].

The $SU(2) \times U(1)$ symmetry group, which describes the electroweak interactions, is spontaneously broken by the existence of a postulated scalar field, the so-called Higgs field, with a non-zero vacuum expectation value [7–12]. This leads to the emergence of massive vector bosons, the W and Z bosons, which mediate the weak interaction, while the photon of electromagnetism remains massless. One physical degree of freedom remains in the Higgs sector which should manifest as a neutral scalar boson H which is so far unobserved. The description of the strong interaction (Quantum Chromodynamics or QCD) is based on the gauge group $SU(3)$ [4–6]. Eight massless gluons mediate this interaction.

All experimental particle physics measurements performed up to date are in excellent agreement with the predictions of the Standard Model. The only noticeable exception is the evidence for non-zero neutrino masses observed in neutrino-oscillation experiments [13]. There remain, however, many key questions open and it is generally believed that the Standard Model can only be a low energy effective theory of a more fundamental underlying theory. One of the strongest arguments for an extension of the Standard Model is the existence of Dark Matter [14] in the universe. There is no explanation for such a type of matter in the Standard Model.

The key questions can be classified to be linked to mass, unification and flavour:

- Mass: What is the origin of mass?
How is the electroweak symmetry broken? Is the solution, as implemented in the Standard Model, realized in Nature, and linked to this, does the Higgs boson exist?
- Unification: What is the underlying fundamental theory?
Can the three interactions which are relevant for particle physics be unified at larger energy and are there new symmetries found towards unification? Are there new particles, e.g. supersymmetric particles, at higher energy scales? And finally, it must also be answered how gravity can eventually be incorporated.
- Flavour: Why are there three generations of matter particles? What is the origin of CP violation in the weak interaction? What is the origin of neutrino masses and mixings?

The high energy and luminosity of the LHC offers a large range of physics opportunities. The primary role of the LHC is to explore the TeV-energy range where answers to at least some of the aforementioned questions are expected to be found. In the focus is certainly the search for the Higgs boson. The LHC experiments have the potential to explore the full relevant mass range and either to discover the Standard Model Higgs boson or to exclude its existence.

Another focus area constitutes the search for supersymmetric particles which can be carried out at the LHC up to the masses of a few TeV. If such particles are discovered, their link to the dark matter in the universe must be investigated. This can only be done in conjunction with experiments on direct dark matter detection [15].

However, it is important to stress that the remit of the LHC is not only to look for the *expected* or theoretically favoured models, but to carry out a thorough investigation of as many final states as possible. It is important to search for any deviation from the Standard Model predictions. This implies that the Standard Model predictions must be reliably tested in the new energy domain. In particular during the early phase of experimentation at the LHC, detailed measurements of Standard Model processes must be carried out. Some of these processes can as well be used for the understanding of the detector response and its calibration.

Finally, with increasing precision of the Standard Model measurements, it is also important to test the consistency of the model via quantum corrections. Important contributions from the LHC in this area will be precise measurements of the W mass and of the top-quark mass, which can be used to constrain the Higgs boson mass. A direct confrontation of this prediction to a direct Higgs boson mass measurement may constitute the ultimate test of the Standard Model at the LHC.

3 Phenomenology of proton-proton collisions

Scattering processes at high-energy hadron colliders can be classified as either hard or soft. Quantum Chromodynamics is the underlying theory for all such processes, but the approach and level of understanding is very different for the two cases. For hard processes, e.g. high- p_T jet production or W and Z production, the rates and event properties can be predicted with good precision using perturbation theory. For soft processes, e.g. the total cross section, the underlying event etc., the rates and properties are dominated by non-perturbative QCD effects, which are less well understood. An understanding of the rates and characteristics of predictions for hard processes, both signals and backgrounds, using perturbative QCD (pQCD) is crucial for tests of the theory and for searches for new physics.

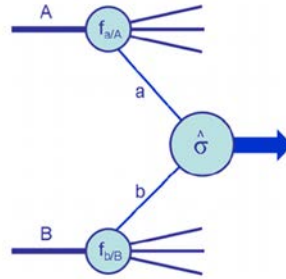


Fig. 1: Diagrammatic structure of a generic hard-scattering process (from Ref. [16]).

The calculation of a hard-scattering process for two hadrons A and B can be illustrated as displayed in Fig. 1. Two partons of the incoming hadrons undergo a hard scattering process characterized by the cross section $\hat{\sigma}$. The structure of the incoming hadrons is described by the parton density functions (PDFs) $f_{a/A}(x_a, \mu_F^2)$ (see Section 3.1), i.e. the probability to find a parton a in hadron A with a momentum fraction x_a at the energy scale μ_F^2 . To obtain the hadron-hadron cross section, a summation over all possible parton-parton scattering processes and an integration over the momentum fractions has to be performed [16]:

$$\sigma_{AB} = \sum_{a,b} \int dx_a \cdot dx_b f_{a/A}(x_a, \mu_F^2) f_{b/B}(x_b, \mu_F^2) \hat{\sigma}_{ab}(x_a, x_b, \alpha_s(\mu_R^2)) . \quad (1)$$

The calculations of the hard scattering process $\hat{\sigma}_{ab}$ are performed in perturbative QCD and the results depend on the strong coupling constant α_s and its renormalization scale μ_R . The scale μ_F that appears in the parton density functions is the so-called factorization scale, which can be thought of as the scale that separates long- and short-distance physics [16]. Large logarithms related to gluons emitted collinear with incoming quarks can be absorbed in the definition of the parton densities, giving rise to logarithmic scaling violations which can be described via the DGLAP¹ evolution equations [17]. The perturbative calculation can be written as

$$\hat{\sigma}_{ab}^{[n]} = \hat{\sigma}_{ab}^{[0]} + \sum_{j=k+1}^{k+n} c_j \cdot \alpha_s^j, \quad (2)$$

where $\hat{\sigma}_{ab}^{[0]}$ denotes the leading order (LO) cross section and n denotes the perturbative order of the calculation. The index k denotes the order of α_s appearing in the leading order calculation, which might as well be 0, like for the Drell-Yan production of W and Z bosons, as discussed below. The cross sections at higher orders, which are usually denoted as next-to-leading order (NLO) and next-to-next-to-leading order (NNLO) etc., are often parametrized in terms of total K factors, defined at each perturbative order $[n]$ as the ratio of the cross section computed to that order normalized to the Born level cross section:

$$\sigma^{[n]} = \sigma^{[0]} \cdot K_{\text{tot}}^{[n]}. \quad (3)$$

As discussed above, the scale μ is an arbitrary parameter, which in general is, however, chosen to be of the order of the energy characterizing the parton-parton interaction, like, for example, the mass of the vector bosons or the transverse momenta of outgoing jets. The more orders are included in the perturbative expansion, the weaker the dependence on μ . As an example, the production of W and Z bosons is discussed in Section 3.3.

Those partons which do not take part in the hard scattering process will produce what is generally called the ‘underlying event’. Finally, it should be stressed that Eq. 2 does not describe the bulk of the events which occur at a hadron collider. It can only be used to describe the most interesting classes of events which involve a hard interaction. Most events result from elastic and soft inelastic interactions generally called ‘minimum bias’ events. In the following a few specific examples of hard scattering processes are discussed.

3.1 Parton Distribution Functions

The parton distribution functions (PDFs) $f(x, Q^2)$ for a hadron provide the probability density of finding a parton with momentum fraction x at momentum transfer Q^2 which defines the energy scale of the process. The Q^2 dependence is induced by the usage of perturbation theory and the resulting higher order corrections. It is described by the DGLAP evolution equations [17]. However, the functional form of the PDFs is not predicted by perturbative QCD and has to be measured experimentally.

Various classes of experiments are sensitive to the proton PDFs, such as deep inelastic scattering at fixed target experiments with electron, muon or neutrino beams, and electron-proton scattering at the HERA collider. Also experiments at pure hadronic colliders such as the Tevatron ($p\bar{p}$) and the LHC (pp) can yield valuable information.

In order to determine the parton distributions from the measurements, a parametrization is assumed to be valid at some starting value $Q^2 = Q_0^2$. The DGLAP evolution functions are used to evolve the PDFs to a different Q^2 where predictions of the measured quantities (e.g. structure functions) are obtained. The predictions are then fitted to the measured datasets, thus constraining the parameters (typically 10 to 20) of the parametrisation [18]. Various collaborations performed fits to the available datasets and provided PDF sets for the proton, for instance the groups ABKM [19], CTEQ [20], CT10 [21], HERAPDF [22,23], JR [24], MSTW [25] and NNPDF [26, 27]. The NNPDF collaboration has already included the first

¹Dokshitzer-Gribov-Lipatov-Altarelli-Parisi

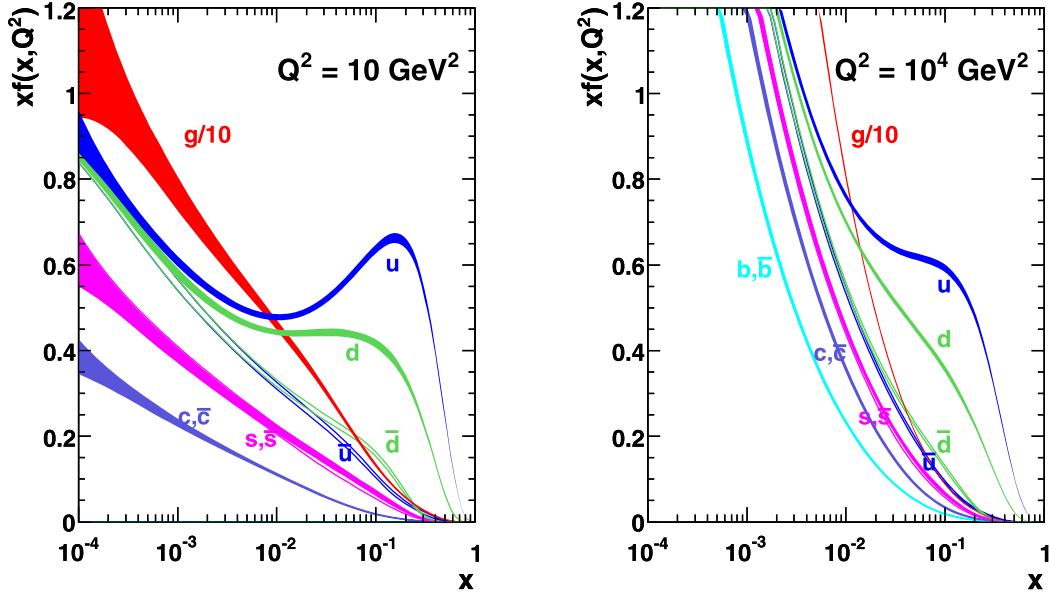


Fig. 2: Parton distribution functions of the proton as determined for the MSTW08 PDF set for (left) $Q^2 = 10 \text{ GeV}^2$ and (right) $Q^2 = 10^4 \text{ GeV}^2$. The bands reflect the uncertainties at the 68% confidence level (from Ref. [25]).

lepton charge asymmetry measurements in the W boson production by the ATLAS [28] and CMS [29] experiments in their fit [30]. The PDFs determined by MSTW are shown in Fig. 2 at two different Q^2 scales. For example, it can be observed that gluons dominate the low x region and the contributions from sea quarks become more dominant at higher Q^2 .

3.2 Jet Production via QCD scattering processes

Two-jet events result when an incoming parton from one hadron scatters off an incoming parton from the other hadron to produce two high transverse momentum partons which are observed as jets. The parton processes that contribute at leading order are shown in Fig. 3. The matrix elements have been calculated at leading order [31] and next-to-leading order [32,33]. At the LHC, terms involving gluons in the initial state are dominant at low p_T . Unlike in lowest order, where a direct correspondence between a jet cross

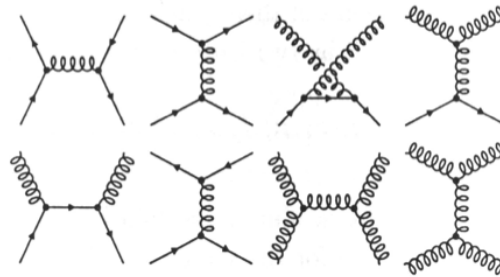


Fig. 3: Leading order diagrams for the production of high- p_T jets.

section and the parton cross section can be made, a prescription is needed to derive jet cross sections in next-to-leading order. When such prescriptions are applied, the next-to-leading order cross sections show substantially smaller sensitivities to variations of the renormalization scale than at lowest order.

3.3 W and Z Production

In leading order the production of the vector bosons W and Z is described by the Drell-Yan process, where a quark and an antiquark from the incoming hadrons annihilate. This process has been calculated up to next-to-next-to-leading order in the strong coupling constant α_S [34–37]. Some of the relevant Feynman diagrams are given in Fig. 4. When going from LO to NLO the cross sections increase by

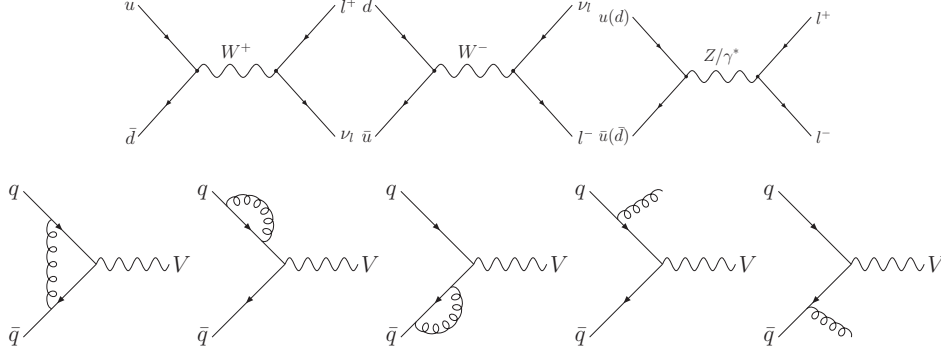


Fig. 4: Leading order (top) and some next-to-leading order diagrams (bottom) for the production of W and Z bosons.

about 20% and the factorization and renormalization scale uncertainties decrease. This is nicely shown in Fig. 5. Including NNLO contributions slightly decreases the cross-sections but the result is consistent

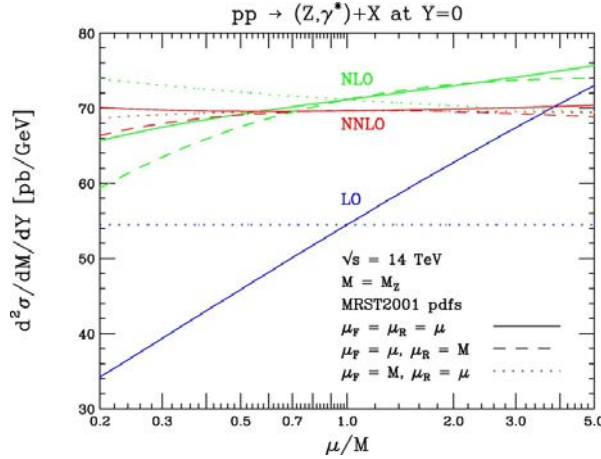


Fig. 5: Dependence of the production cross section of on-shell Z bosons at rapidity $y = 0$ on the choice of the renormalization and factorization scales. For each order in perturbation theory (LO, NLO, NNLO), three curves are shown. The solid curve represents the results obtained under a common variation of $\mu_R = \mu_F = \mu$ over the range $M/5 < \mu < 5M$. The dashed (dotted) curves represent the results obtained under variations of the factorization (renormalization) scale alone, holding the other scale fixed (from Ref. [37]).

with the NLO prediction within the NLO scale uncertainty, indicating that the perturbative expansion converges. The impact of higher order corrections on the predicted rapidity² distributions of the W and Z bosons is shown in Fig. 6 for proton-proton collisions with a centre-of-mass energy of $\sqrt{s} = 14$ TeV. This figure also illustrates that larger cross sections for W^+ production than for W^- production are

²The rapidity y of a particle is related to its energy E and the projection of its momentum on the beam axis p_z by $y = \frac{1}{2} \ln[(E + p_z)/(E - p_z)]$. The pseudorapidity η is defined as $\eta = -\ln \tan \frac{\theta}{2}$, where θ is the polar angle.

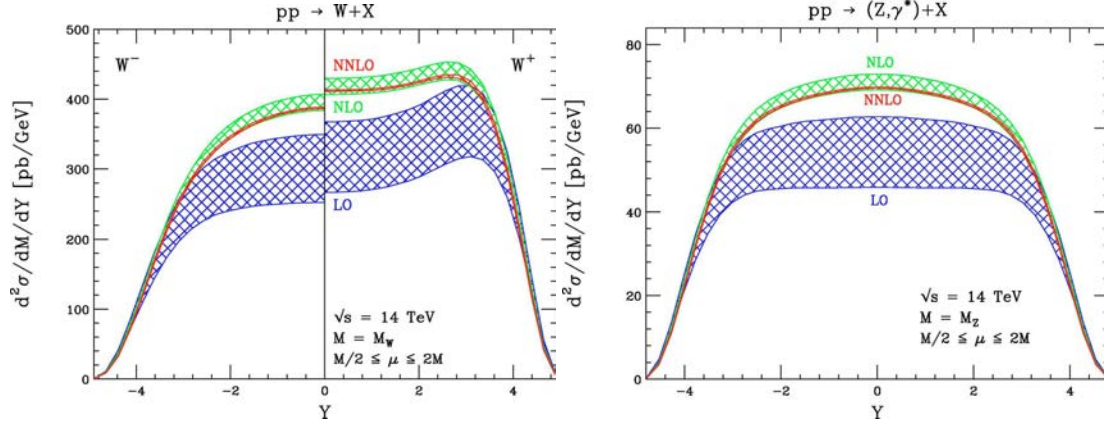


Fig. 6: Theory predictions at LO, NLO and NNLO of the rapidity distributions for W (left) and Z (right) boson production in proton-proton collisions at $\sqrt{s} = 14$ TeV. The bands indicate the factorization and renormalization scale uncertainties, obtained by scale variations in the range $M_{W/Z}/2 \leq \mu \leq 2M_{W/Z}$ (from Ref. [37]).

expected at the LHC. This asymmetry results from the dominance of u over d valence quarks in the incoming protons (see also Fig. 2).

Electroweak radiative corrections for the W and Z boson production have been computed up to next-to-leading order [38, 39]. These corrections change the production cross sections and affect kinematic properties like the lepton transverse momenta, lepton rapidities and the transverse and invariant masses of the lepton pairs. In particular for precision measurements like that of the W boson mass, it is therefore important to take electroweak radiative corrections into account.

3.4 Cross Sections at the LHC

An overview of cross sections of some benchmark processes at proton-proton and proton-antiproton colliders as a function of the centre-of-mass energy is shown in Fig. 7. The total inelastic proton-proton cross section is dominant and reaches a huge value of about 70 mb at the LHC. Processes which can proceed via the strong interaction have a much larger cross section than electroweak processes. The dominant electroweak process, the production of W and Z bosons is found to be about six and seven orders of magnitude smaller than the total inelastic proton-proton cross section. However, this process constitutes the most copious source of prompt high- p_T leptons, which are important for many physics measurements and searches for new physics at the LHC. The production processes of the Standard Model Higgs boson and of other non-coloured heavy new particles have small cross sections and therefore require a correspondingly high integrated luminosity for their detection. The Higgs boson production cross section is found to be ten to eleven orders of magnitude smaller than the inelastic pp cross section. The exact values depend strongly on the mass of the Higgs boson, as further discussed in Section 5.1.

4 Measurement of Standard Model processes

4.1 The production of high- p_T jets

A measurement of the production of high- p_T jets constitutes an important test of QCD in the new energy regime of the LHC. Events with two high transverse momentum jets (dijets) arise from parton-parton scattering where the outgoing scattered partons manifest themselves as hadronic jets. The measurements of the inclusive jet production cross section or the dijet production cross section are therefore also sensitive to the structure of the proton and may lead to further constraints on the PDFs. In addition, the precise measurement of jet production is important for searches of physics beyond the Standard Model. New

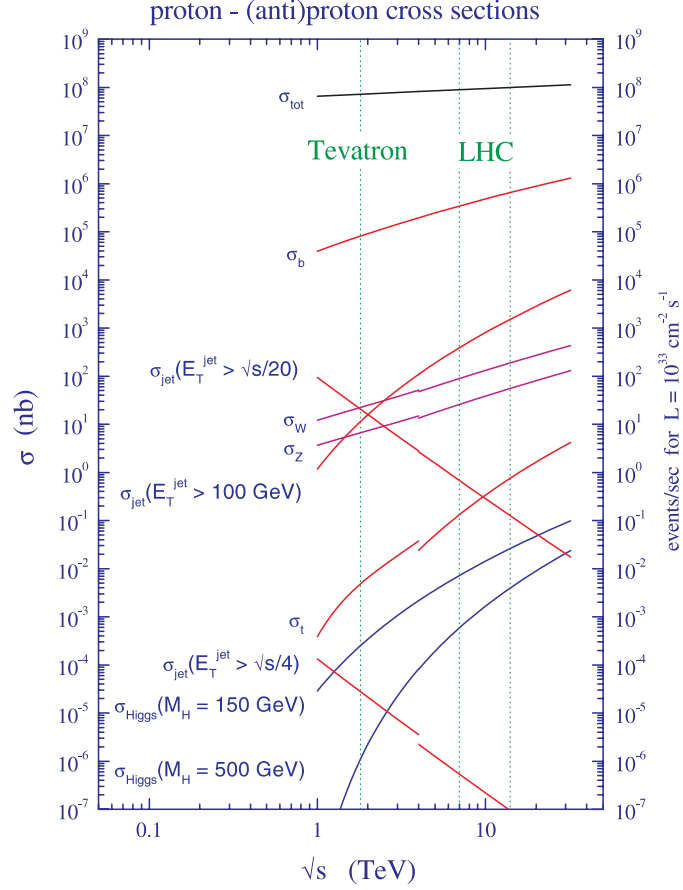


Fig. 7: Standard Model cross sections at the Tevatron and LHC colliders as function of the centre-of-mass energy $\sqrt{s}$. In case of the LHC both the energy during data taking in 2010/2011 ($\sqrt{s} = 7$ TeV) and the nominal energy ($\sqrt{s} = 14$ TeV) are marked (adapted from Ref. [16]).

physics may lead to significant deviations from the expected QCD behaviour. For example, a substructure of quarks may manifest itself in deviations of the measured inclusive jet-production cross section from the expected behaviour at high transverse momenta. The measurement of the dijet cross section as a function of the dijet mass m_{jj} allows for a sensitive search for physics beyond the Standard Model, such as dijet resonances or contact interactions of composite quarks.

The production of multijets provides as well an important background in the search for physics beyond the Standard Model. In many cases, leptons and missing transverse energy are used as final states signatures. Although they do not appear at first place in QCD jet production, they might originate from decays of heavy quarks, from mis-measurements of jet energies or from mis-identification of jets as leptons. Although the probability for this to happen is small, the contributions to the background can be sizeable, give the huge jet production cross sections.

4.1.1 Jet reconstruction and calibration

For the reconstruction of jets both the ATLAS and CMS experiments use the infrared- and collinear-safe *anti- k_T* jet clustering algorithm [40] with distance parameters $0.4 \leq R \leq 0.7$. The inputs are either topological calorimeter cluster energies [41,42] in the ATLAS experiment or particle flow objects [43,44] in the CMS experiment. For the theoretical comparison the input can also be four-vectors from stable particles in generator-level simulations. In all cases residual jet-level corrections are needed to account

for energy losses not detectable on cluster or particle flow level. These jet-level calibrations are Monte Carlo based correction functions in pseudorapidity $|\eta|$ and p_T . The jet energy scale and the attached uncertainties are validated with in-situ methods using the balance of transverse momenta in dijet and γ -jet events. The systematic jet energy scale uncertainties are found to be typically in the range of $\pm(3 - 6)\%$ over a large range of η and p_T . The larger values are reached at large $|\eta|$ as well as at very low and very high p_T .

4.1.2 Jet cross section measurements

The inclusive jet production cross section has been measured by both the ATLAS [45] and CMS [46] experiments as a function of the jet transverse momentum (p_T) and jet rapidity (y). In addition, double differential cross sections in the maximum jet rapidity $y_{\max}$ and dijet mass m_{jj} for dijet events are measured [45,47]. The data are corrected for migration and resolution effects due to the steeply falling spectra in p_T and mass. The NLO perturbative parton-level QCD predictions are corrected for hadronisation and the underlying event activity. Figure 8 (left) shows the inclusive jet cross-section measurement for jets with size $R = 0.4$ as a function of jet transverse momentum from the ATLAS collaboration [45], based on the total data set collected in 2010 corresponding to an integrated luminosity of 37 pb^{-1} . The experimental systematic uncertainties are dominated by the jet energy scale uncertainty. There is an additional overall uncertainty of $\pm 3.4\%$ due to the luminosity measurement. The theoretical uncertainties result mainly from the choice of the renormalization and factorization scales, parton distribution functions, $\alpha_s(m_Z)$ and the modelling of non-perturbative effects.

The cross section measurement as a function of the dijet invariant mass from the CMS collaboration [47] is shown in Fig. 8 (right). Like for the inclusive jet cross section measurements, the experimental uncertainties are in the range 10-20% and are dominated by uncertainties on the jet energy scale and resolution.

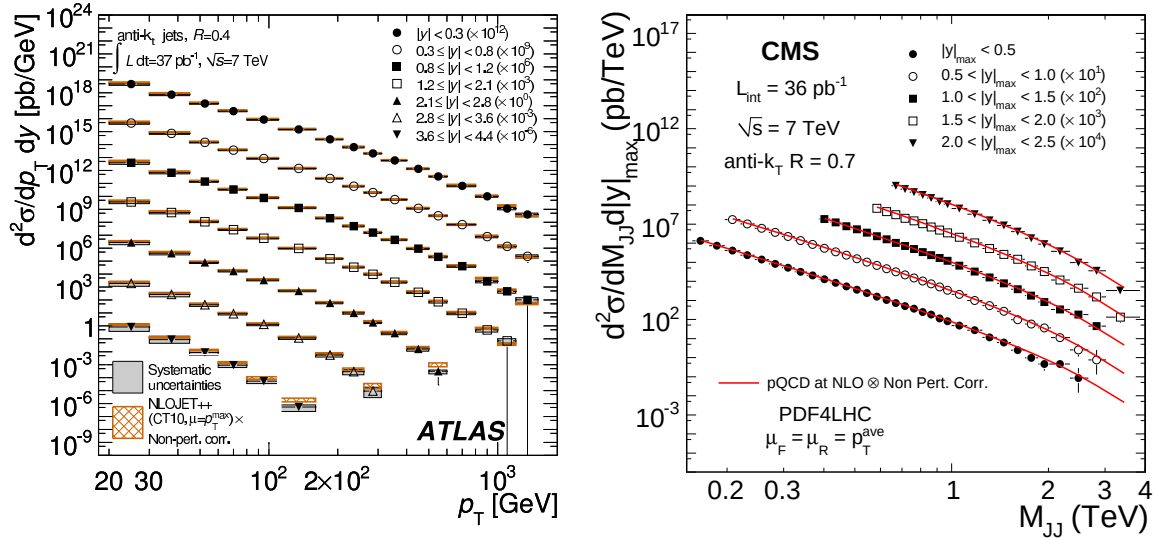


Fig. 8: (Left): Inclusive jet double-differential cross section as a function of jet p_T in different regions of $|y|$ from the ATLAS collaboration. (Right): Measured double-differential dijet cross sections (points) as a function of the dijet invariant mass m_{jj} in bins of the variable $y_{\max}$ from the CMS collaboration. The data are compared in both cases to NLO pQCD calculations to which non-perturbative corrections have been applied. The error bars indicate the statistical uncertainty on the measurement. The dark-shaded band indicates the quadratic sum of the experimental systematic uncertainties, excluding the uncertainties from the luminosity. The theory uncertainty is shown as the light, hatched band (from Refs. [45, 47]).

Different NLO pQCD predictions, using different PDF sets, are compared to the data and the corresponding ratios of data to the NLO predictions. Figure 9 shows an example from the ATLAS collaboration [45]. Within the experimental and theoretical uncertainties the data are well described by the predictions, although they are found to be systematically higher than the data. The deviations become larger at large $|y|$ and p_T . However, it is impressive to see that the QCD calculations are able to describe the data over many orders of magnitude and up to the highest values of p_T and mass ever observed.

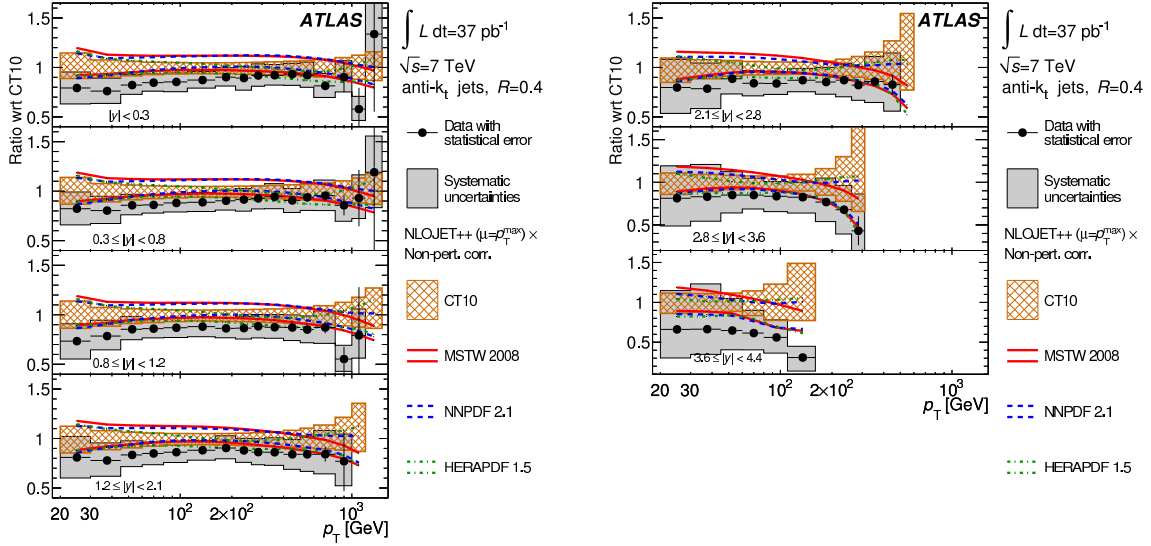


Fig. 9: Ratios of inclusive jet double-differential cross sections to the theoretical predictions. The ratios are shown as a function of jet p_T in different regions of $|y|$. The theoretical error bands obtained by using NLOJET++ with different PDF sets (CT10, MSTW 2008, NNPDF 2.1, HERAPDF 1.5) are shown (from Ref. [45]).

The ATLAS and CMS collaborations have performed many further studies on jet production, including the measurement of dijet angular distributions [48] and dijet angular decorrelations [45, 49]. At Born level, dijets are produced with equal transverse momenta p_T and back-to-back in the azimuthal angle ($\Delta\phi_{\text{dijet}} = |\phi_{\text{jet1}} - \phi_{\text{jet2}}|$). Gluon emission will decorrelate the two highest p_T jets and cause smaller angular separations. The measurement of the angular distribution between the highest p_T jets is therefore also a sensitive test of perturbative QCD with the advantage that the measurement is not strongly affected by the dominant systematic uncertainty on the jet energy scale. The predictions from NLO pQCD are found to be in reasonable agreement with the measured distributions [45, 49].

In addition, the production of multijets was studied [50, 51]. In a data sample corresponding to an integrated luminosity of 2.4 pb^{-1} the ATLAS collaboration has identified 115 events with more than six jets. One such event is shown in Fig. 10. The transverse energy deposition in the calorimeter is shown as a function of η and ϕ . For this event the six jets are well separated spatially. Leading-order Monte Carlo simulations have been compared to the measured multi-jet inclusive and differential cross sections. For events containing two or more jets with $p_T > 60 \text{ GeV}$, of which at least one has $p_T > 80 \text{ GeV}$, a reasonable agreement is found between data and leading-order Monte Carlo simulations with parton-shower tunes that describe adequately the ATLAS $\sqrt{s} = 7 \text{ TeV}$ underlying event data. The agreement is found after the predictions of the Monte Carlo simulations are normalized to the measured inclusive two-jet cross section.

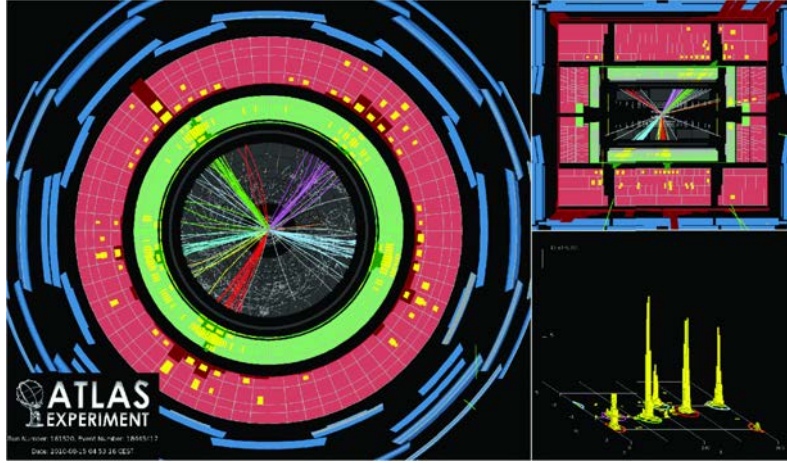


Fig. 10: Event display of a six-jet event passing the ATLAS multijet selection requirements. The towers in the bottom right figure represent transverse energy deposited in the calorimeter projected on a grid of η and ϕ . Jets with transverse momenta ranging from 84 to 203 GeV are measured in this event (from Ref. [50]).

4.2 The production of W and Z bosons

W and Z bosons are expected to be produced abundantly at the LHC. The large dataset and the high LHC energy allow for detailed measurements of their production properties in a previously unexplored kinematic domain. These conditions, together with the proton-proton nature of the collisions, provide new constraints on the parton distribution functions and allow for precise tests of perturbative QCD. Besides the measurements of the W and Z boson production cross sections, the measurement of their ratio R and of the asymmetry between the W^+ and W^- cross sections (see Section 3.3) constitute important tests of the Standard Model. This ratio R can be measured with a higher relative precision because both experimental and theoretical uncertainties partially cancel. With larger data sets this ratio can be used to provide constraints on the W -boson width Γ_W [52].

4.2.1 Inclusive cross-section measurements

Measurement of the W^+ , W^- and Z/γ^* boson inclusive production cross sections are performed using the leptonic decay modes $W \rightarrow \ell\nu$ and $Z \rightarrow \ell\ell$. Already in 2010, the two collaborations published first measurements in the electron and muon decay modes [53, 54]. They were updated with the full data sample taken in 2010 corresponding to an integrated luminosity of 36 pb^{-1} [55, 56]. In this data sample the ATLAS experiment has observed a total of about 270.000 $W \rightarrow \ell\nu$ decays and a total of about 24.000 $Z/\gamma^* \rightarrow \ell\ell$ decays. The measurements in the electron and muon channels were found to give consistent results and were combined to obtain a single joint measurement taking into account the statistical and systematic uncertainties and their correlations. The results are displayed in Fig. 11 together with previous measurements of the total W and Z production cross sections by the UA1 [57] and UA2 [58] experiments at $\sqrt{s} = 0.63 \text{ TeV}$ at the CERN Sp $\bar{p}$ S and by the CDF [52] and DØ [59] experiments at $\sqrt{s} = 1.8 \text{ TeV}$ and $\sqrt{s} = 1.96 \text{ TeV}$ at the Fermilab Tevatron collider and by the PHENIX [60] experiment in proton-proton collisions at $\sqrt{s} = 0.5 \text{ TeV}$ at the RHIC collider. These measurements are compared to the NNLO theoretical predictions for proton-proton and proton-antiproton collisions. All measurements are in good agreement with the theoretical predictions and the energy dependence of the total W and Z production cross sections is well described. The precision of the integrated W and Z/γ^* cross sections in the fiducial regions is $\sim \pm 1.2\%$ with an additional uncertainty of $\pm 3.4\%$ resulting from the knowledge of the luminosity. It should be noted that the experimental uncertainties are already dominated by systematic uncertainties. The total integrated cross sections are obtained from an extrapolation of the measurement

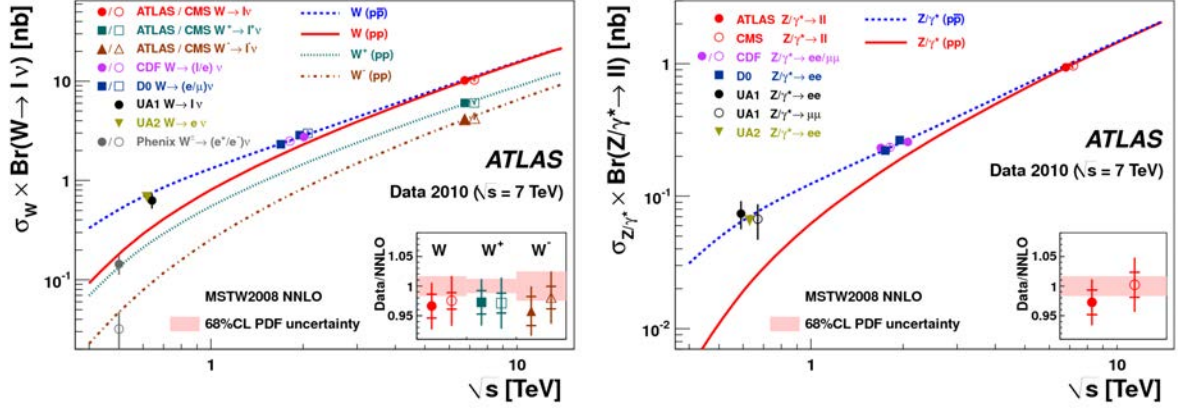


Fig. 11: The measured values of $\sigma_W \cdot \text{Br}(W \rightarrow \ell\nu)$ for W^+ , W^- and for their sum (left) and of $\sigma_{Z/\gamma^*} \times \text{Br}(Z/\gamma^* \rightarrow \ell\ell)$ (right) compared to the theoretical predictions based on NNLO QCD calculations. Results are shown for the combined measurements of electron and muon final states. The predictions are shown for both proton-proton (W^+ , W^- and their sum) and proton-antiproton colliders (W) as a function of $\sqrt{s}$. In addition, previous measurements at proton-antiproton and proton-proton colliders are shown. The data points at the various energies are staggered to improved readability. The data points are shown with their total uncertainty. The theoretical uncertainties are not shown in this figure (from Ref. [56]).

in the fiducial regions to the full acceptance. Due to uncertainties on the acceptance corrections, the uncertainties on the total cross sections are about twice as large.

A summary of the ratios of the measured total W^+ , W^- , W and Z/γ^* cross sections by the CMS collaboration to the theoretical NNLO calculations is shown in Fig. 12 (left). Within the experimental and theoretical uncertainties there is excellent agreement. This figure also includes a comparison of the measured ratios $R_{W/Z} = \sigma_W \cdot \text{Br}(W \rightarrow \ell\nu)/\sigma_Z \cdot \text{Br}(Z \rightarrow \ell\ell)$ and $R_{+/-} = \sigma_{W^+} \cdot \text{Br}(W^+ \rightarrow \ell^+\nu)/\sigma_{W^-} \cdot \text{Br}(W^- \rightarrow \ell^-\nu)$. Due to the cancellation of uncertainties, most notably the luminosity uncertainty, the precision of these ratio measurements is more precise. Also the measured ratios are well described by the theoretical NNLO calculations.

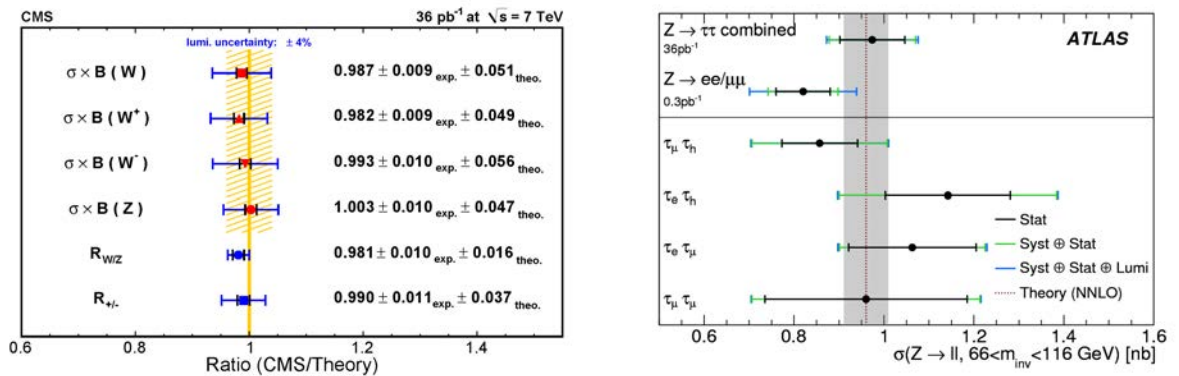


Fig. 12: (Left): Ratio of CMS measurement of W and Z cross sections to theory expectations. The experimental uncertainty is the sum in quadrature of the statistical and the systematic uncertainties not including the uncertainty on the extrapolation to the full acceptance due to parton density functions (taken from Ref. [55]). (Right): Measurements of the $Z \rightarrow \tau\tau$ cross sections from the ATLAS experiment in various τ decay modes and comparison to theoretical predictions and to the measurements in the electron and muon channels (taken from Ref. [61]).

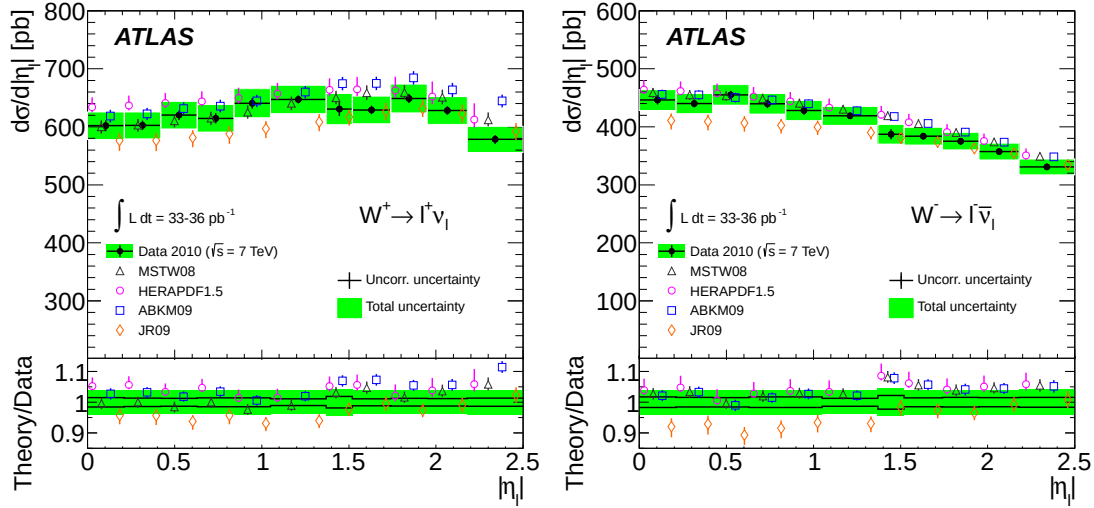


Fig. 13: Differential cross-section measurement $d\sigma/d|\eta_\ell|$ for W^+ (left) and W^- (right) for $W \rightarrow \ell\nu$ from the ATLAS collaboration compared to the NNLO theory predictions using various PDF sets. The kinematic requirements are $p_T(\ell) > 20$ GeV, $p_T(\nu) > 25$ GeV and $m_T > 40$ GeV. The ratio of theoretical predictions to data is also shown. Theoretical points are displaced for clarity within each bin (from Ref. [56]).

Meanwhile the cross sections have also been measured in the $W \rightarrow \tau\nu$ [62] and $Z \rightarrow \tau\tau$ [61, 63] decay modes, where hadronically decaying τ leptons are identified and reconstructed. The results obtained by the ATLAS collaboration in various τ decay modes are displayed in Fig. 12 (right). They are found to be in good agreement with theoretical predictions and with the results obtained in the $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu\mu$ final states.

4.2.2 Differential cross section measurements

With the complete data set collected in 2010 more differential cross-section measurements became possible. Both the ATLAS and CMS collaborations have performed measurements as a function of lepton pseudorapidity η_ℓ , for the $W^\pm$ cross sections, and of the boson rapidity, y_Z , for the Z/γ^* cross section [55, 56]. For the Z/γ^* case, all values refer to dilepton mass windows from 66–116 GeV and 60–110 GeV for the ATLAS and CMS analyses, respectively. The cross sections are measured in well-defined kinematic regions within the detector acceptance, defined by the pseudorapidity of the charged lepton and the transverse momentum of the neutrino. The differential W^+ and W^- cross sections as measured by the ATLAS collaboration are shown in Fig. 13. The measurements for the electron and muon final states were found to be in good agreement with each other and were combined. These data are compared with the theoretical NNLO predictions using various NNLO PDF sets (JR09, ABKM09, HERAPDF1.5 and MSTW08). The differential Z/γ^* cross section as a function of the boson rapidity as measured by the ATLAS and CMS collaborations are shown in Fig. 14. Although the gross features of these differential W and Z cross-section measurements are well described by the theoretical calculations, the (pseudo)rapidity dependence shows some disagreement which carries important information on the underlying parton density functions. It is expected that these differential measurements will reduce the uncertainties on the parton density functions. Very recently, these data have been used together with the ep scattering data from HERA to extract the ratio or the strange-to-down sea quark density at Bjorken x values around 0.01. The ratio is found to be consistent with 1 and supports the hypothesis that the density of the light sea quarks is flavour independent [64]. The general agreement between theory and experiment is remarkable and provides evidence for the universality of the PDFs and the reliability of perturbative QCD calculations in the kinematic regime of the LHC.

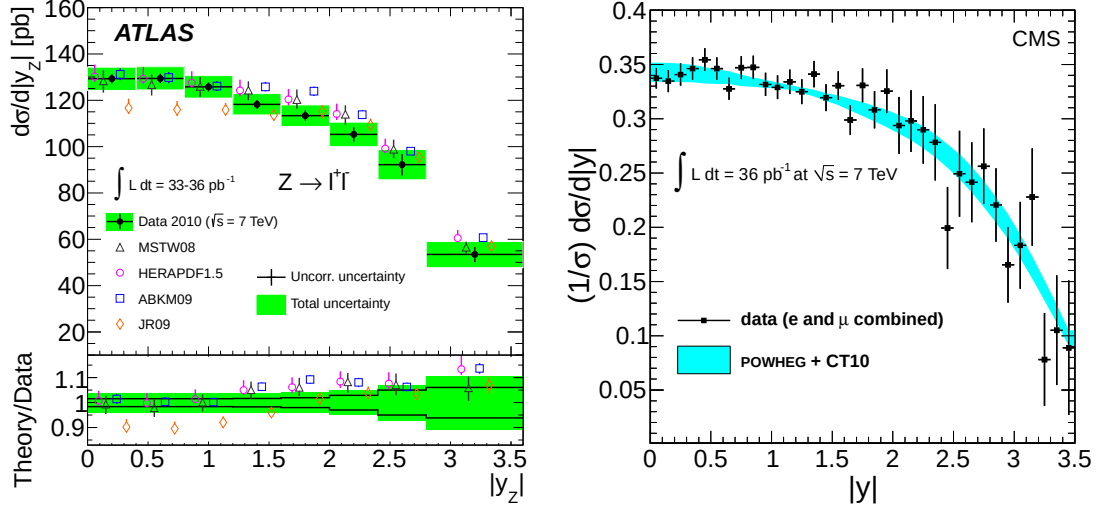


Fig. 14: Differential cross-section measurements $d\sigma/d|y_Z|$ for $Z \rightarrow \ell\ell$ from the ATLAS (left) and CMS (right) collaborations compared to NNLO theory predictions using various PDF sets. The kinematic requirements are $66 < m(\ell\ell) < 116 \text{ GeV}$ and $p_T(\ell) > 20 \text{ GeV}$. The ratio of theoretical predictions to data is also shown for the ATLAS measurements. Theoretical points are displaced for clarity within each bin (from Refs. [55,56]).

4.2.3 Measurements of the associated W and Z + jet production

The study of the associated production of vector bosons with high- p_T jets constitutes another important test of the perturbative QCD. In addition, these final states are a significant background to studies of other Standard Model processes, such as $t\bar{t}$ or diboson production, as well as for searches for the Higgs boson and for physics beyond the Standard Model.

The ATLAS and CMS collaborations have presented detailed measurements of these processes based on the complete dataset from 2010, corresponding to an integrated luminosity of 36 pb^{-1} [65–68]. Cross sections have been determined for the associated W and Z +jet production as a function of inclusive jet multiplicity, N_{jet} , for up to five jets. At each multiplicity, the cross sections have also been presented as a function of jet transverse momenta of all jets. The results, corrected for all detector effects and for all backgrounds such as diboson and top quark pair production, are compared with particle-level predictions from perturbative QCD. As an example, the W +jets cross-section measurements as a function of jet multiplicity are shown in Fig. 15 (left) and as a function of the p_T of the first jet in the event in Fig. 15 (right). Leading-order multiparton event generators like *ALPGEN* [69] or *SHERPA* [70], normalized to the NNLO total cross section for inclusive W -boson production, describe the data reasonably well for all measured inclusive jet multiplicities. Next-to-leading-order calculations from *MCFM* [71], studied for $N_{\text{jet}} \leq 2$, and *BlackHat-Sherpa* [72], studied for $N_{\text{jet}} \leq 4$, are found to be mostly in good agreement with the data. This also holds for the measurement of the transverse momentum distributions of the W and Z boson, which are correlated to the jet activity in the W and Z events.

4.3 Production of top quarks

4.3.1 Measurement of production cross sections

The top quark is the heaviest known elementary particle with a mass of about 173 GeV . Due to its high mass it is believed to play a special role in the electroweak symmetry breaking and possibly in models of new physics beyond the Standard Model. In that context it is remarkable that its Yukawa coupling λ_t is close to one. We still know little about the properties of the top quark, like spin, charge, lifetime, decay properties (rare decays) or the gauge and Yukawa couplings. Another important parameter is the mass

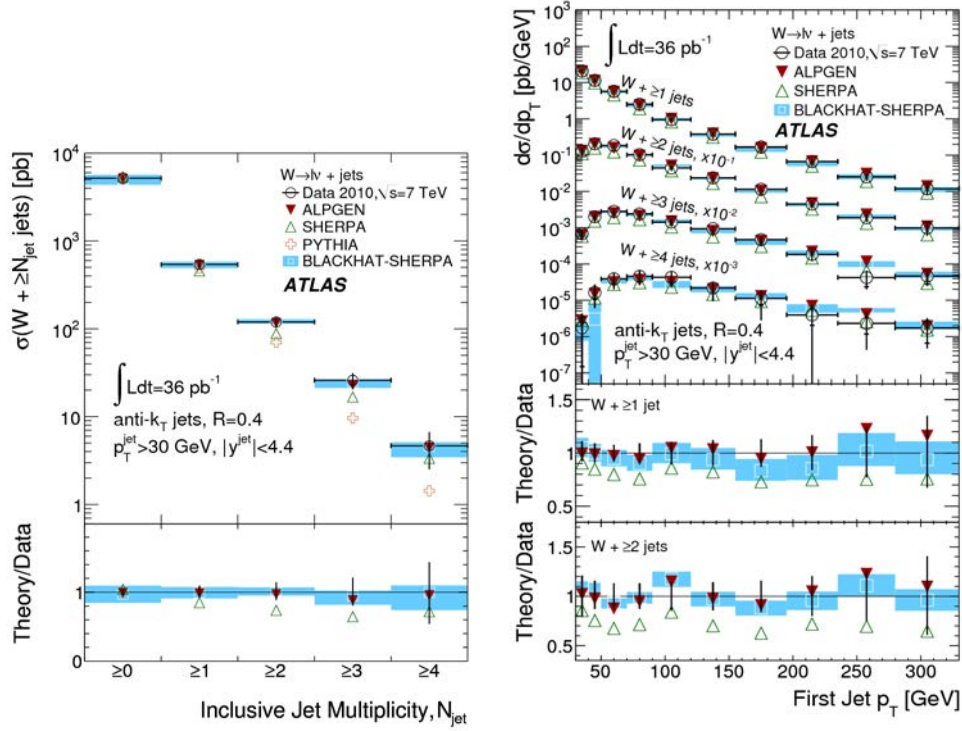


Fig. 15: The measured W +jets cross sections as a function of jet multiplicity (left) and as a function of the p_T of the first jet in the event (right). The p_T of the first jet is shown separately for events with ≥ 1 jet to ≥ 4 jets. Shown are predictions from ALPGEN, SHERPA and BlackHat-SHERPA, and the ratio of theoretical predictions to data (from Ref. [68]).

of the top quark, which has, however, been relatively precisely measured at the Tevatron collider to be $m_{\text{top}} = 173.2 \pm 0.90 \text{ GeV}$, i.e. with a precision of 0.5%. A further improvement here is important for a precise test of electroweak radiative corrections.

Due to the high mass, the top quarks decays mainly via $t \rightarrow Wb$ before it hadronizes. The production of $t\bar{t}$ pairs at the LHC proceeds mainly via gluon initiated processes and is expected to be a factor of 20 larger at the LHC with $\sqrt{s} = 7 \text{ TeV}$ than at the Tevatron. The decays studied are characterized by the number of charged leptons in the final state. A large fraction of the branching ratio is devoted to lepton-hadron decays, where one of the W s decays leptonically and the other one into a pair of jets, i.e. $t\bar{t} \rightarrow Wb Wb \rightarrow \ell\nu b qqb$. The final state in this case consists of a lepton, missing transverse momentum (carried away by the neutrino) and four jets out of which two are originating from b -quarks. The complementary dilepton decay mode is also important for top-quark physics at hadron colliders. The fully hadronic channel is more difficult to trigger on and shows a worse signal-to-background ratio, but despite this has also been measured at the LHC.

Both the ATLAS and CMS collaborations measured the production cross section for the pair production of top quarks in all above-mentioned final states [73–77]. The results are displayed in Fig. 16. The most precise measurement comes from the single lepton channel, which shows already a precision of the order of $\pm 7\%$. In this channel the cross sections are measured with and without the requirement of a b -tagged jet. The results obtained in the dilepton channel are consistent with these results. The measurements are found to be in good agreement with the approximate NNLO calculations [78, 79], although the experimental values are found to be about 1σ higher in each experiment. The experimental measurement is already limited by the experimental systematic uncertainties (jet energy scale, b -tagging,

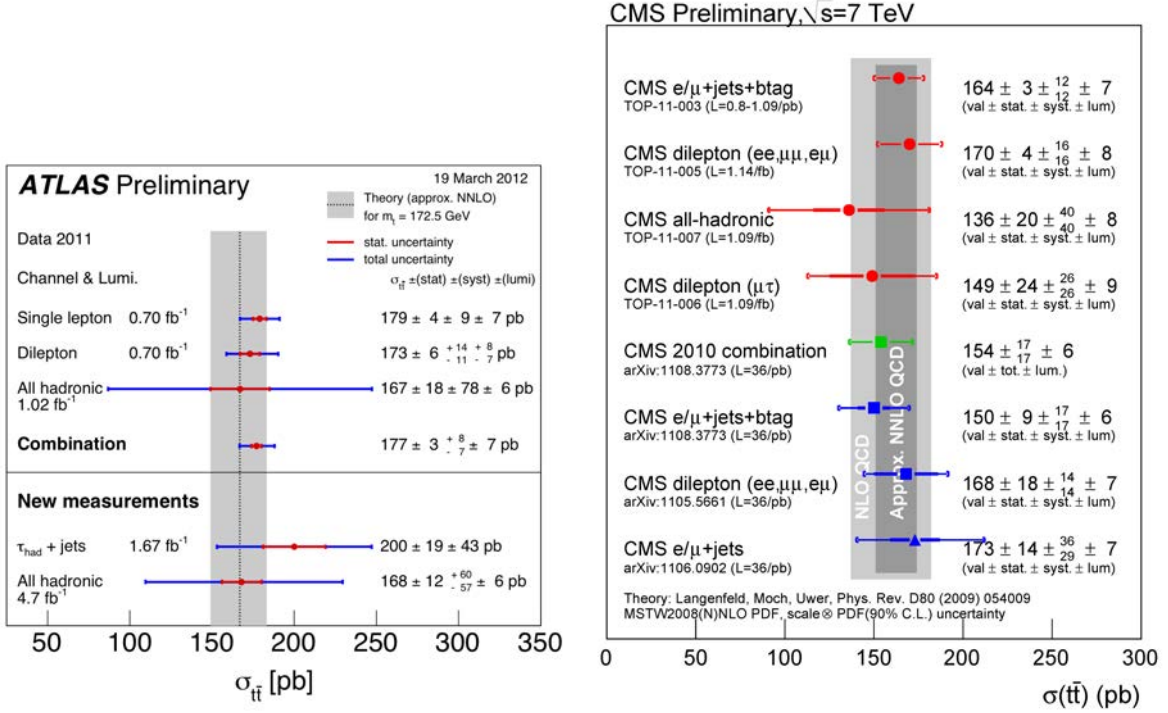


Fig. 16: The measured value of $\sigma_{t\bar{t}}$ in the various decay modes and the combination of these measurements from the ATLAS (left) and CMS (right) experiments. The approximate NNLO prediction with its uncertainty is also shown (from Refs. [74, 77]).

...) and by the uncertainty on the luminosity.

Single top quarks can be produced at the LHC via electroweak processes. The t -channel production of single top-quarks has been measured by both the ATLAS [80] ($L = 0.7 \text{ fb}^{-1}$) and CMS [81] ($L = 1.5 \text{ fb}^{-1}$) collaborations. The results are found to be consistent with the Standard Model expectations. The measured cross section value from the CMS collaboration is shown in Fig. 17 in comparison to the theoretical expectation and the measurements at the Tevatron.

4.3.2 Measurement of the top-quark mass

Among the various top-quark properties, the ATLAS and CMS collaborations have already presented first measurements on the top-quark mass m_t in several final states [82–85]. The most precise result was presented recently as a preliminary result by the CMS collaboration [84]. The top-quark mass has been measured using a sample of $t\bar{t}$ candidate events with one muon and at least four jets in the final state. The full 2011 data set corresponding to an integrated luminosity of 4.7 fb^{-1} was used. In this sample 2391 candidate events were selected and using a likelihood method the top-quark mass was measured from fits to kinematic distributions, simultaneously with the jet energy scale (JES). The result of $m_t = 172.6 \pm 0.6$ (stat + JES) ± 1.2 (syst) GeV is consistent with the Tevatron result (see Fig. 17 (right)). It is impressive that such a precision, in particular on the systematic uncertainty, can be claimed after only two years of operation of the LHC. The dominant contribution to this systematic uncertainty results from uncertainties on the b -jet energy scale and from modelling uncertainties estimated via changes of the factorization scale [84]. However, it remains to be seen whether the small overall uncertainty can be confirmed by the ATLAS experiment. The results of the present measurements at the LHC are summarized in Fig. 17 together with the Tevatron results.

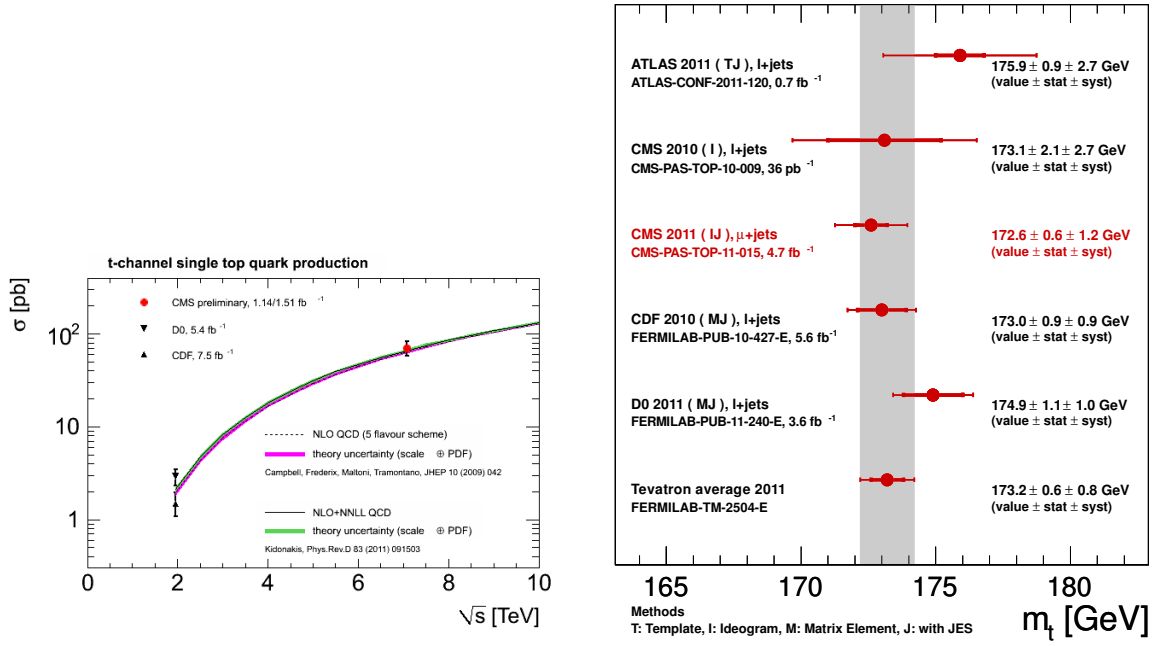


Fig. 17: (Left): Measured cross section for single-top quark production via the t -channel process in the CMS experiment in comparison to the theoretical expectation and the measurements at the Tevatron (from Ref. [81]). (Right): Compilation of the top-quark mass measurements from the ATLAS, CDF, CMS and DØ experiments (from Ref. [84]).

4.4 The production of diboson pairs

The production of pairs of bosons ($W\gamma$, WW , WZ , ZZ) at the LHC is of great interest since it provides an excellent opportunity to test the predictions on the structure of the gauge couplings of the electroweak sector of the Standard Model at the TeV energy scale. In addition, WW and ZZ pairs constitute the irreducible background in important Higgs boson search channels like $H \rightarrow WW$ and $H \rightarrow ZZ$.

The dominant Standard Model W^+W^- production mechanisms are s -channel and t -channel quark-antiquark annihilation. The s -channel production occurs only through the triple gauge coupling vertex and accounts for $\sim 10\%$ of the full W^+W^- production cross section. The leading-order Feynman diagrams for the dominant $q\bar{q}' \rightarrow W^+W^-$ production mechanisms at the LHC are shown in the left and middle diagrams of Fig. 18. The total cross section $\sigma(q\bar{q}, q\bar{q}') \rightarrow W^+W^-$ are known at next-to-leading order. The gluon fusion through quark loops, shown in the right diagram of Fig. 18, contributes an additional 2.9%.

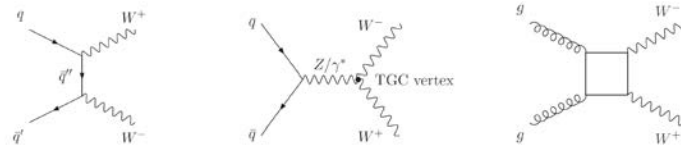


Fig. 18: (Left): The Standard Model tree-level Feynman diagram for W^+W^- production through the $q\bar{q}$ initial state in the t -channel. (Middle): The corresponding Standard Model tree-level diagram in the s -channel, which contains the WWZ and $WW\gamma$ triple gauge boson coupling (TGC) vertices. (Right): The gluon fusion process, mediated by quark loops.

The ZZ production proceeds at leading order via t -channel quark-antiquark interactions. The ZZZ and $ZZ\gamma$ triple gauge boson couplings (nTGCs) are absent. Hence there is no contribution from s -channel $q\bar{q}$ annihilation at tree level. Many models of physics beyond the Standard Model predict values of nTGCs at the level of 10^{-4} to 10^{-3} [86]. The signature of nonzero nTGCs is an increase of the ZZ cross section at high ZZ invariant mass and high transverse momentum of the Z bosons [87].

The ATLAS and CMS collaborations have measured the cross sections for all diboson production processes, $W\gamma$ [88, 89], WW [90–92], WZ [91, 93], ZZ [91, 94]. Several analyses are already based on the full data set from 2011. The results are found to be in good agreement with the predictions from the Standard Model and first constraints on anomalous triple gauge boson couplings have been set. The agreement between the measurements and the Standard Model predictions is shown for the WW and ZZ production in Fig. 19. The constraints on the triple gauge boson couplings are still limited by the number of observed diboson events.

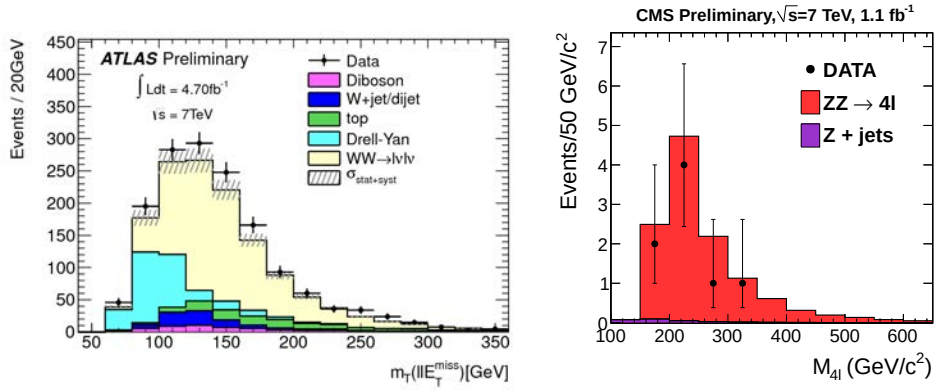


Fig. 19: (Left): The distributions of m_T of the dilepton+ E_T^{miss} system for the W^+W^- candidates in the ATLAS experiment (from Ref. [92]). (Right): The distribution of the four-lepton invariant mass for the ZZ candidate events in the CMS experiment (from Ref. [91]).

4.5 Summary

As discussed in the previous subchapters, the first two years have seen a very successful operation of the LHC collider and of the experiments. The data collected have been used to extract precise measurements of many Standard Model processes. They range from the measurement of W and Z production with cross sections in the order of picobarns via the measurement of $t\bar{t}$ production to the measurement of important diboson processes. The summary of all measured cross sections by the ATLAS collaboration is shown in Fig. 20 together with the theoretical predictions. Within the uncertainties, excellent agreement is found for all the processes considered. This is a remarkable achievement of the Standard Model and the underlying theoretical concepts, including QCD and factorization. The smallest cross sections measured, the diboson production cross sections, are extremely relevant for the Higgs boson search, as discussed in the next section.

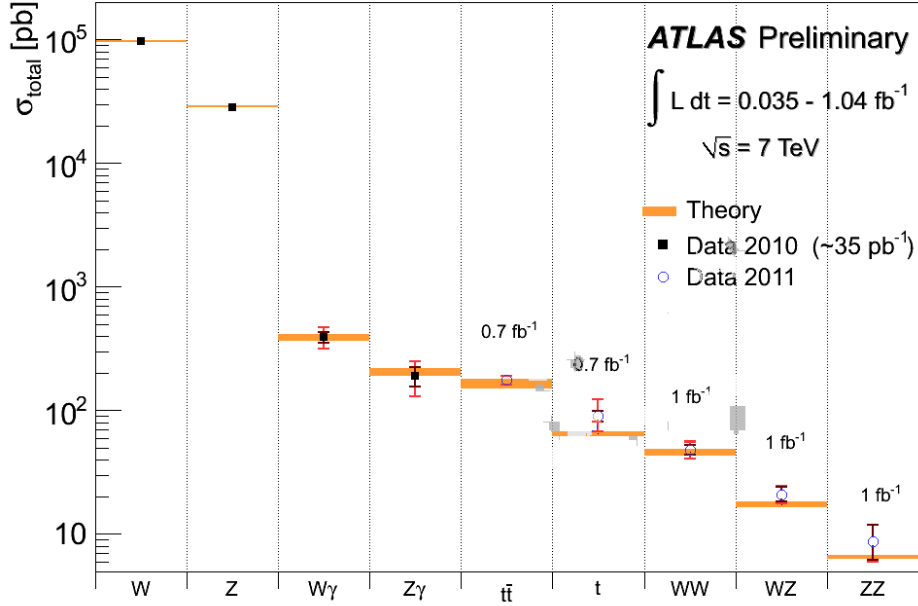


Fig. 20: Summary of several Standard Model total production cross-section measurements from the ATLAS collaboration compared to the corresponding theoretical expectations. The integrated luminosities used for the measurements are indicated on the figure. The dark error bars represent the statistical uncertainties. The red error bars represent the full uncertainties, including systematics and luminosity uncertainties. All theoretical expectations were calculated at NLO or higher.

5 Search for the Higgs boson

The Higgs boson is the only Standard Model particle that has not been discovered so far. Indirectly, high precision electroweak data constrain its mass via their sensitivity to radiative corrections. Assuming the overall validity of the Standard Model, a global fit [95] to all electroweak data leads to $m_H = 94^{+29}_{-24}$ GeV. On the basis of the present theoretical knowledge, the Higgs sector in the Standard Model remains largely unconstrained. While there is no direct prediction for the mass of the Higgs boson, an upper limit of ~ 1 TeV can be inferred from unitarity arguments [96].

Direct searches at the e^+e^- collider LEP has led to a lower bound on its mass of 114.4 GeV [97]. Before the LHC started its operation, the Fermilab Tevatron $p\bar{p}$ collider with a centre-of-mass energy of 1.96 TeV was the leading accelerator exploring the energy frontier. Until the end of data taking in September 2011, the two experiments CDF and DØ have collected data corresponding to an integrated luminosity of 11.9 fb^{-1} . During the past years, they were able to exclude Higgs boson masses around 160 GeV, mainly based on the search for the $H \rightarrow WW^{(*)} \rightarrow \ell\nu\ell\nu$ decay mode. In Summer 2011, the combination of the results from the two experiments, based on data corresponding to an integrated luminosity of 8.6 fb^{-1} , excluded a mass range from 156 - 177 GeV [98]. At the same time, the first exclusions from the ATLAS and CMS experiments, based on a data corresponding to an integrated luminosity of up to 2.3 fb^{-1} , were presented. The ATLAS experiment excluded mass ranges from 146–230 GeV, 256–282 GeV and 296–459 GeV [99]. The CMS analysis was based on data corresponding to an integrated luminosity of up to 1.7 fb^{-1} and the Higgs boson was excluded in the mass ranges from 145–216 GeV, 226–288 GeV and 310–400 GeV [100].

Since then the full data set of the LHC taken until the end of 2011 has been analyzed. Preliminary results were presented in a special seminar at CERN in December 2011 and updates were presented at the

Moriond conference 2012. They created a lot of attention and excitement in the community since the data show tantalizing hints of a possible Higgs boson signal in the low mass region around 125 GeV. However, it must be stressed that the background-only probability still shows acceptable values, in particular if the look-elsewhere effect [101] is taken into account.

In the following, these results are summarized since they supersede those presented at the CERN school in September 2011. Before entering the discussion, the Higgs boson production at hadron colliders and the Higgs boson decay properties as well as a few statistical issues are briefly summarized.

5.1 Higgs boson production at the LHC

At hadron colliders, Higgs bosons can be produced via four different production mechanisms:

- gluon fusion, $gg \rightarrow H$, which is mediated at lowest order by a heavy quark loop;
- vector boson fusion (VBF), $qq \rightarrow qqH$;
- associated production of a Higgs boson with weak gauge bosons, $qq \rightarrow W/Z H$ (Higgs strahlung, Drell-Yan like production);
- associated Higgs boson production with heavy quarks, $gg, qq \rightarrow t\bar{t}H, gg, qq \rightarrow b\bar{b}H$ (and $gb \rightarrow bH$).

The dominant production mode is the gluon-fusion process, followed by the vector boson fusion. In the low mass region it amounts at leading order to about 20% of the gluon-fusion cross section, whereas it reaches the same level for masses around 800 GeV. The associated WH , ZH and $t\bar{t}H$ production processes are relevant only for the search of a light Standard Model Higgs boson with a mass close to the LEP limit.

The Higgs boson production cross sections are computed up to next-to-next-to-leading order (NNLO) [102] in QCD for the gluon-fusion process. In addition, QCD soft-gluon resummations up to next-to-next-to-leading log (NNLL) improve the NNLO calculation [103]. The next-to-leading order (NLO) electroweak corrections [104] are also applied, assuming factorization between QCD and electroweak corrections. The cross sections for the VBF process are calculated with full NLO QCD and electroweak corrections [105], and approximate NNLO QCD corrections are available [106]. The $W/Z H$ processes are calculated at NLO [107] and at NNLO [108], and NLO electroweak radiative corrections [109] are applied. Also for the associated $t\bar{t}H$ production the full NLO QCD corrections are available [110]. For a more detailed review of the theoretical aspects of Higgs boson production the reader is referred to Ref. [111]. The results of the calculations and the estimated theoretical uncertainties are shown for the different production processes in Fig. 21 (left) as a function of the Higgs boson mass [111].

5.2 Higgs boson decays

The branching ratios and the total decay width of the Standard Model Higgs boson are shown in Fig. 21 (right) as a function of the Higgs boson mass. They have been calculated taking into account both electroweak and QCD corrections [112, 113]. When kinematically accessible, decays of the Standard Model Higgs boson into vector boson pairs WW or ZZ dominate over all other decay modes. Above the kinematic threshold, the branching fraction into $t\bar{t}$ can reach up to 20%. All other fermionic decays are only relevant for Higgs boson masses below ~ 140 GeV, with $H \rightarrow b\bar{b}$ dominating. The branching ratios for both $H \rightarrow \tau\tau$ and $H \rightarrow gg$ reach up to about 8% at Higgs boson masses between 100 and 120 GeV. Decays into two photons, which are of interest due to their relatively clean experimental signature, can proceed via charged fermion and W loops with a branching ratio of up to $2 \cdot 10^{-3}$ at low Higgs boson masses.

Compared to the mass resolution of hadron collider experiments, the total decay width of the Standard Model Higgs boson is small at low masses and becomes significant only above the threshold for

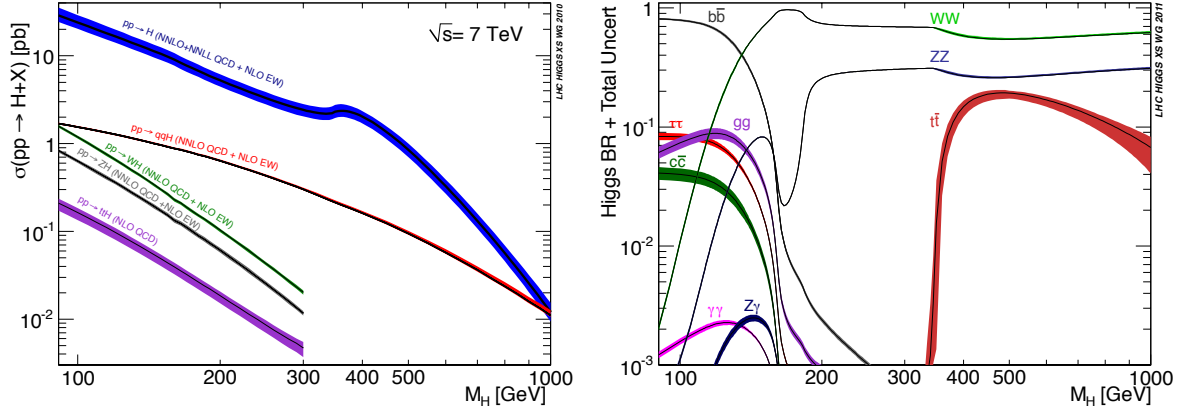


Fig. 21: (Left): Production cross sections for the different production processes for a Standard Model Higgs boson as a function of the Higgs boson mass at the LHC. (Right): Branching ratios of the Standard Model Higgs boson as a function of Higgs boson mass (from Ref. [111]).

decays into ZZ . For a Higgs boson with a mass of ~ 1 TeV the resonance is broad with a width of about 600 GeV. In this mass regime, the Higgs field is coupling strongly, resulting in large corrections [111, 114].

5.3 Search for the Standard Model Higgs boson at the LHC

The Standard Model Higgs boson is searched for at the LHC in various decay channels, the choice of which is given by the signal rates and the signal-to-background ratios in the different mass regions. The search strategies and background rejection methods have been established in many studies over the past years [115]. Among the most important channels are the inclusive $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^{(*)} \rightarrow \ell\ell\ell\ell$ decay channels. These channels are characterized by a very good mass resolution. In the low mass region, the Higgs boson appears as a sharp resonance, the width of which is dominated by the detector resolution, on top of flat backgrounds which are dominated by $\gamma\gamma$ and ZZ continuum production, respectively. In addition, the $H \rightarrow WW$ decay mode contributes in particular in the mass region around 160 GeV, however, due to the neutrinos in the final state, no mass peak can be reconstructed. Evidence for Higgs boson production is given by a broad peak in the transverse mass distribution (see below). From the fermionic decays, only the modes $H \rightarrow \tau\tau$ and $H \rightarrow b\bar{b}$ have a chance to be detected. For the $H \rightarrow \tau\tau$ decays the selection of the vector boson fusion production mode is important to improve the signal-to-background ratio by exploiting forward jet tagging [116]. The $b\bar{b}$ decay mode is searched for in the associated production of the Higgs boson with a vector boson or with a $t\bar{t}$ pair [117, 118].

In the following the present status (March 2012) of the Higgs boson search in the various final states at the LHC is described. Before the individual channels are discussed some common issues on the statistical treatment are given. At the end of this section the combination of the results is presented for both the ATLAS and CMS collaborations.

5.3.1 Limit setting, statistical issues

In the following the distributions of reconstructed masses or other distributions as measured in data are compared to the expectations from Standard Model background processes. In order to test the compatibility of the data with the background-only hypothesis a so-called p_0 probability value is calculated. It quantifies the probability that a background-only experiment is more signal-like than that observed. The *local* p_0 probability is assessed for a fixed m_H hypothesis and the equivalent formulation in terms of number of standard deviations is referred to as the *local* significance. Since fluctuations of the back-

ground could occur at any point in the mass range the results have to be corrected for the look-elsewhere effect [101]. The probability for a background-only experiment to produce a local significance of this size or larger anywhere in a given mass region is referred to as the *global* p_0 . The corresponding reduction in the significance is estimated using the prescription described in Ref. [119].

The data can also be used to set 95% confidence level (C.L.) upper limits (σ_{95}) on the cross section for Higgs boson production. These cross sections are usually normalized to the expected Standard Model value (σ_{95}/σ_{SM}). All exclusion limits quoted in the following have been calculated using the CL_s method [120]. In addition to the *observed* limits based on the observed data, also the *expected* limits are calculated. They are calculated as a function of m_H and are based on the central values of the expected background in case no Higgs signal is present. The 1σ and 2σ fluctuations around the expected exclusion limits are calculated as well.

Excluded mass regions are determined from a comparison of the observed cross-section limit to the Standard Model cross-section value. If the observed value of σ_{95}/σ_{SM} is smaller than 1 (Standard Model cross-section expectation) for a particular hypothetical Higgs boson mass, this mass value can be excluded with a confidence level of 95%. Systematic uncertainties are incorporated by introducing nuisance parameters with constraints. Asymptotic formulae [121] are used to derive the limits and p_0 values. This procedure has been validated using ensemble tests and a Bayesian calculation of the exclusion limits with a uniform prior on the signal cross section. These approaches to the limits typically agree with the asymptotic median results to within a few percent [122].

5.3.2 Search for $H \rightarrow \gamma\gamma$ decays

The decay $H \rightarrow \gamma\gamma$ is a rare decay mode, which is only detectable in a limited Higgs boson mass region between 100 and 150 GeV, where both the production cross section and the decay branching ratio are relatively large. Excellent energy and angular resolution are required to observe the narrow mass peak above the irreducible prompt $\gamma\gamma$ continuum. In addition, there is a reducible background resulting from direct photon production or from two-jet production via QCD processes. These processes contribute if one or both jets are misidentified as a photon. The background can be determined from a fit to the data (sidebands) such that uncertainties from Monte Carlo predictions or uncertainties due to normalizations in control regions are avoided. Due to the rather large amount of material in the tracking detectors of the LHC experiments there is a high probability for a photon to undergo conversion and therefore both unconverted and converted photons need to be reconstructed.

Both collaborations have presented results on the $H \rightarrow \gamma\gamma$ search in a mass range between 110 and 150 GeV based on the full data set collected until the end of 2011 [123, 124]. The ATLAS analysis [123] separates events into nine independent categories. The categorisation is based on the direction of each photon and whether it was reconstructed as a converted or unconverted photon, together with the momentum component of the diphoton system transverse to the thrust axis. The distribution of the invariant mass of the diphoton events, $m_{\gamma\gamma}$, summed over all categories, is shown in Fig. 22 (left). The fit to the background is performed separately in each category in the mass range 100 - 160 GeV by using an exponential function with free slope and normalization parameters. The result for the total sample is superimposed in Fig. 22. The signal expectation for a Higgs boson with $m_H = 120$ GeV is also shown. The mass resolution depends on the classification of the photon (calorimeter η region, conversion status) and is found to be 1.4 GeV in the best category and 1.7 GeV on average. The residuals of the data with respect to the total background as a function of $m_{\gamma\gamma}$ is also shown in Fig. 22. Around a mass of 126 GeV an excess of events above the background is seen (see discussion below). The 95% C.L. upper limits on the cross section for Higgs boson production normalized to the Standard Model value, σ_{95}/σ_{SM} , are shown in Fig. 22 (right). The observed exclusion limits follow well the expectations over a large mass range, except in the region around 126 GeV. The ATLAS data allow for a 95% C.L. exclusion of a Standard Model Higgs boson in the mass ranges between 113–115 GeV and 134.5–136 GeV.

The analysis of the CMS collaboration [124] is done in a similar way. Diphoton events are split

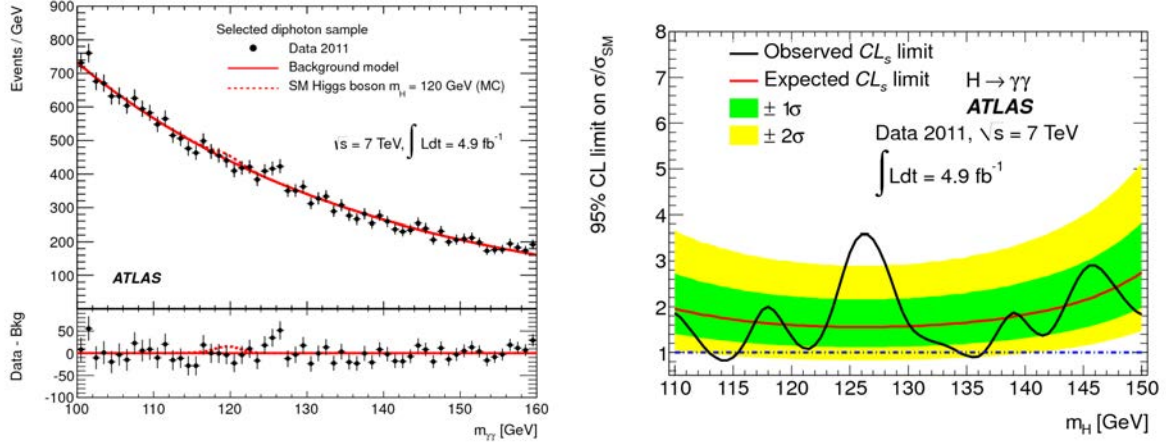


Fig. 22: (Left): Invariant mass distribution for the selected data sample in the ATLAS experiment, overlaid with the total background (see text). The bottom inset displays the residual of the data with respect to the total background. The Higgs boson expectation for a mass hypothesis of 120 GeV corresponding to the Standard Model cross section is also shown. (Right): Observed and expected 95% C.L. limits on the Standard Model Higgs boson production cross section normalized to the predicted one as a function of m_H (from Ref. [123]).

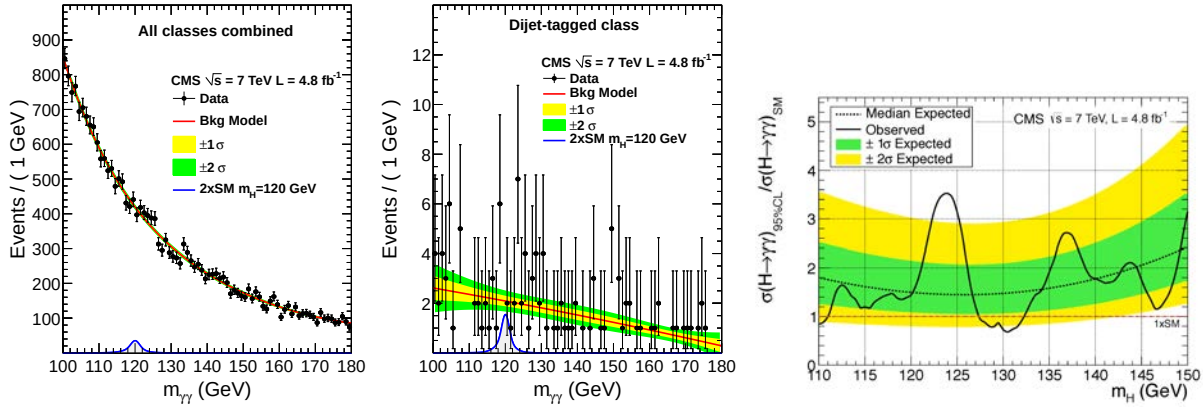


Fig. 23: (Left): Invariant mass distribution for the selected data sample in the CMS experiment, overlaid with the total background. (Middle): The invariant mass distribution for diphotons fulfilling the VBF selection (see text). The Higgs boson expectation for a mass hypothesis of 120 GeV corresponding to the Standard Model cross section multiplied by a factor of two is also shown. (Right): Observed and expected 95% C.L. limits on the Standard Model Higgs boson production cross section normalized to the predicted one as a function of m_H (from Ref. [124]).

into four categories depending on their η value and shower shape characteristics (to distinguish converted from unconverted photons). The diphoton mass resolution is best for the class of two central unconverted photons and reaches a value of 1.2 GeV (full width at half maximum divided by 2.35) for a Higgs boson mass of 120 GeV. Including the other classes a weighted average resolution of ~ 1.8 GeV is found. A further class of events is introduced to select the vector boson fusion topology (VBF topology). By requiring two jets with a large separation in pseudorapidity, a class of events is defined for which the expected signal-to-background ratio is about an order of magnitude larger than for the events in the four classes defined by photon properties. The $m_{\gamma\gamma}$ distributions observed in the data for the sum of the five event classes and for the VBF topology separately are shown in Fig. 23 (left, middle) together with the background fits based on polynomial functions. The uncertainty bands shown are computed from the

fit uncertainty on the background yields. The limit set on the cross section of a Higgs boson decaying to two photons normalized to the Standard Model value is shown in Fig. 23 (right). The CMS analysis excludes at the 95% C.L. the Standard Model Higgs boson decaying into two photons in the mass range 128 to 132 GeV. However, it should be noted that this exclusion, as well as the ATLAS exclusions in this channel, are *lucky* since the expected sensitivities are larger than one and the observed values of $\sigma_{95}/\sigma_{\text{SM}}$ are at the edges of the 2σ bands. The fluctuations of the observed limit about the expected limit are consistent with statistical fluctuations to be expected in scanning the mass range. The largest deviation in the CMS experiment is seen at $m_{\gamma\gamma} = 124$ GeV.

In order to quantify the fluctuations seen in both experiments, the probabilities for the background-only hypothesis have been calculated. The observed and expected local p_0 values obtained are displayed in Fig. 24 for the ATLAS (left) and CMS (right) data. Before considering the uncertainty on the signal mass position, the largest excess with respect to the background-only hypothesis in the mass range 110–150 GeV is observed at 126.5 GeV in the ATLAS data with a local significance of 2.9σ . The uncertainty on the mass position (± 0.7 GeV) due to the imperfect knowledge of the photon energy scale has a small effect on the significance. When this uncertainty is taken into account, the significance is slightly reduced to 2.8σ . The local p_0 value corresponding to the largest upwards fluctuation in the CMS data at 124 GeV has a significance of 3.1σ . The observed significances reduce to 1.5σ for ATLAS and 1.8σ for CMS, when the look-elsewhere effect is taken into account over the mass range 110–150 GeV.

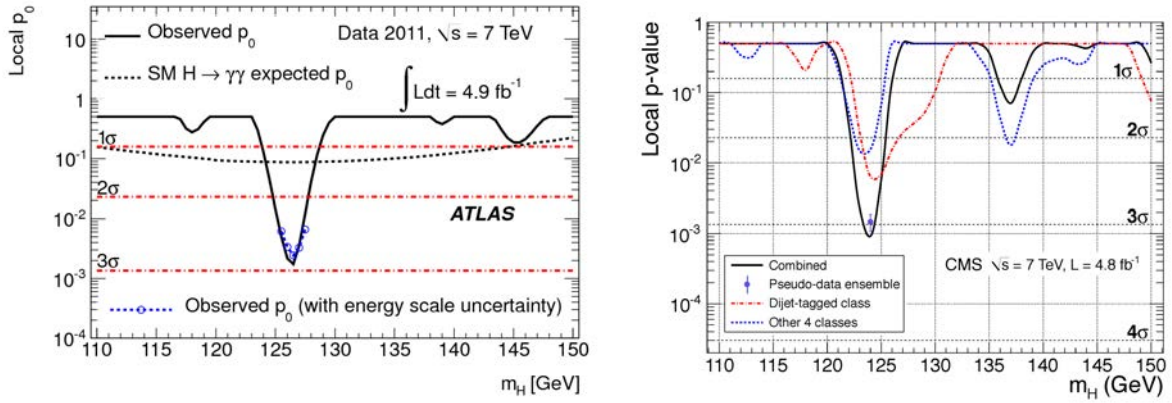


Fig. 24: The observed local p_0 , the probability that the background fluctuates to the observed number of events or higher, for the ATLAS (left) and CMS (right) data. In the ATLAS case, the open points indicate the observed local p_0 value when energy scale uncertainties are taken into account. The dotted line shows the expected median local p_0 for the signal hypothesis when tested at m_H . In the CMS case, the p_0 values are shown for the VBF-tagged class separately (from Refs. [123, 124]).

5.3.3 Search for $H \rightarrow ZZ^{(*)} \rightarrow \ell\ell\ell\ell$ decays

The decay channel $H \rightarrow ZZ^{(*)} \rightarrow \ell\ell\ell\ell$ provides a rather clean signature in the mass range $115 \text{ GeV} < m_H < 2 m_Z$. In addition to the irreducible backgrounds from ZZ^* and $Z\gamma^*$ production, there are large reducible backgrounds from $t\bar{t}$ and $Zb\bar{b}$ production. Due to the large production cross section, the $t\bar{t}$ background dominates at production level, whereas the $Zb\bar{b}$ events contain a genuine Z boson in the final state and are therefore more difficult to reject. In addition, there is background from ZZ continuum production, where one of the Z bosons decays into a τ pair, with subsequent leptonic decays of the τ leptons, and the other Z decays into an electron or muon pair.

Both collaborations have performed the $H \rightarrow ZZ^{(*)} \rightarrow \ell\ell\ell\ell$ search for m_H hypotheses in the full 110 to 600 GeV mass range using data corresponding to an integrated luminosity of $\sim 4.8 \text{ fb}^{-1}$ [125,

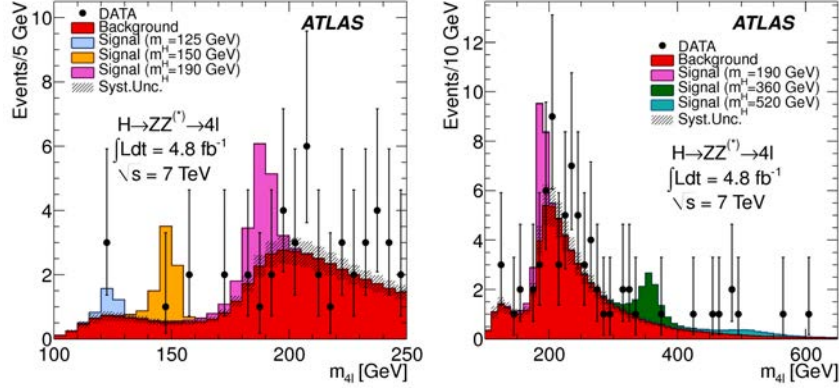


Fig. 25: Distributions of the four-lepton invariant mass, $m_{4\ell}$ of the selected candidates, compared to the background expectation for the 100–250 GeV mass range (left) and the full mass range (right) in the ATLAS experiment. The signal expectations for several m_H hypotheses are also shown (from Ref. [125]).

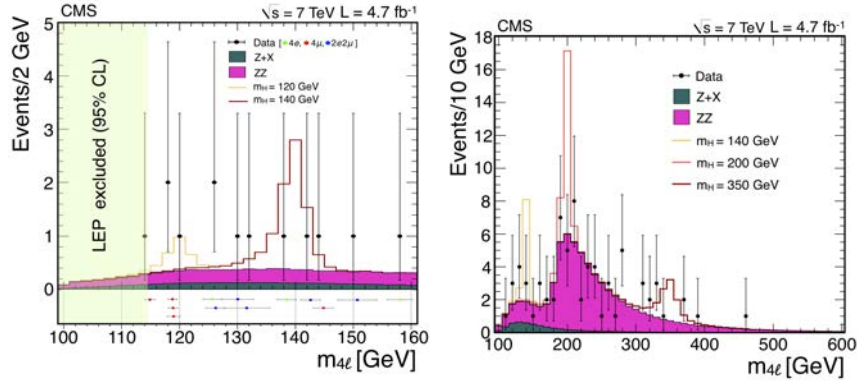


Fig. 26: Distributions of the four-lepton invariant mass, $m_{4\ell}$ of the selected candidates, compared to the background expectation for the 100–160 GeV mass range (left) and the full mass range (right) in the CMS experiment. The signal expectations for several m_H hypotheses are also shown (from Ref. [126]).

[126]. It has been shown that in both experiments calorimeter and track isolation requirements together with impact parameter requirements can be used to suppress the irreducible background well below the irreducible ZZ^* continuum background. The residual irreducible Z +jets and $t\bar{t}$ backgrounds, which have an impact mostly for low invariant four-lepton masses, are estimated from control regions in the data. The irreducible ZZ^* background is estimated using Monte Carlo simulation. The events are categorised according to the lepton flavour combinations. Mass resolutions of approximately 1.5% in the four-muon channel and 2% in the four-electron channel are achieved at $m_H \sim 120$ GeV [125]. The four-lepton invariant mass is used as a discriminant variable. The observed and expected mass distributions for events selected after all cuts are displayed in Figs. 25 and 26 for the ATLAS and CMS experiments, respectively.

The measured mass distributions are again confronted to the background-only hypotheses. The corresponding p_0 values are shown in Fig. 27 for the two experiments. In the ATLAS experiment large upward deviations from the background-only hypothesis are observed for $m_H = 125$ GeV, 244 GeV and 500 GeV with local significances of 2.1σ , 2.2σ and 2.1σ , respectively. After accounting for the look-elsewhere effect none of these excesses is significant. The CMS collaboration observes excesses of events around 119 GeV, 126 GeV and 320 GeV. The most significant excess for a mass value near 119 GeV

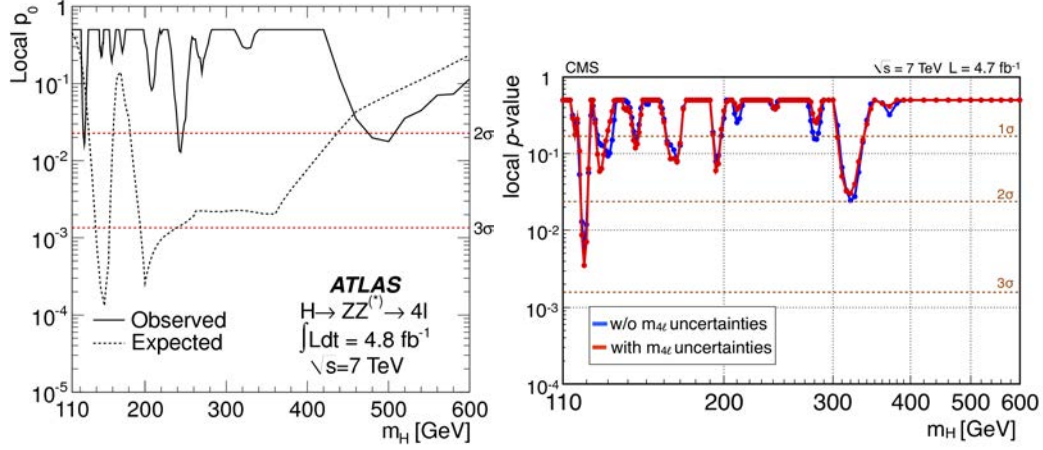


Fig. 27: The observed local p_0 values as a function of the Higgs boson mass in the $H \rightarrow ZZ^{(*)} \rightarrow \ell\ell\ell\ell$ channel in the ATLAS (left) and CMS (right) experiments. The dashed curve shows the expected median local p_0 . The horizontal lines indicate values of constant significance of 1σ , 2σ and 3σ (from Refs. [125, 126]).

corresponds to a local (global) significance of about 2.5σ (1.0σ).

5.3.4 Search for $H \rightarrow WW^{(*)} \rightarrow \ell\nu\ell\nu$ decays

The decay mode $H \rightarrow WW^{(*)} \rightarrow \ell\nu\ell\nu$ has the highest sensitivity for Higgs boson masses around 170 GeV. Based on searches in this channel, mass regions could be excluded by both the Tevatron and the LHC experiments already in Summer 2011 [98–100]. However, this channel is more challenging in the low mass region around 125 GeV since due to the reduced $H \rightarrow WW$ branching ratio the expected signal rates are small. Due to the presence of neutrinos it is not possible to reconstruct a Higgs boson mass peak and evidence for a signal must be extracted from an excess of events above the expected backgrounds. Usually, the WW transverse mass (m_T), computed from the leptons and the missing transverse momentum,

$$m_T = \sqrt{(E_T^{\ell\ell} + E_T^{\text{miss}})^2 - |\mathbf{p}_T^{\ell\ell} + \mathbf{p}_T^{\text{miss}}|^2},$$

where $E_T^{\ell\ell} = \sqrt{|\mathbf{p}_T^{\ell\ell}|^2 + m_{\ell\ell}^2}$, $|\mathbf{p}_T^{\text{miss}}| = E_T^{\text{miss}}$ and $|\mathbf{p}_T^{\ell\ell}| = p_T^{\ell\ell}$, is used to discriminate between signal and background. The WW , $t\bar{t}$ and single-top production processes constitute severe backgrounds and the signal significance depends critically on their absolute knowledge.

The analyses of the ATLAS and CMS collaborations are based on the full data set ($\sim 4.7 \text{ fb}^{-1}$) [127, 128]. In order to optimize the sensitivity, the analyses are split into different lepton final states (ee , $e\mu$ and $\mu\mu$) and different jet multiplicities. In addition, they have been optimized for different mass regions (low and high mass). Typical selection cuts require the presence of two isolated high p_T leptons with a significant missing transverse energy and a small azimuthal angular separation. The latter requirement is motivated by the decay characteristics of a spin-0 Higgs boson decaying into two W bosons with their subsequent $W \rightarrow \ell\nu$ decay [129]. The various jet categories are sensitive to different Higgs boson production mechanisms and have very different background compositions. The 0-jet category is mainly sensitive to the gluon-fusion process and has the non-resonant WW production as major background. The 2-jet category is more sensitive to the vector-boson fusion process, with $t\bar{t}$ as dominant background. As a final discriminant the WW transverse mass distribution is used. This distribution is shown in Fig. 28 (left) for events passing the 0-jet selection in ATLAS. The observed data are well described by the expected background contributions which are dominated by the WW production. As another example, Fig. 29 (left) shows the distribution of the azimuthal angle difference

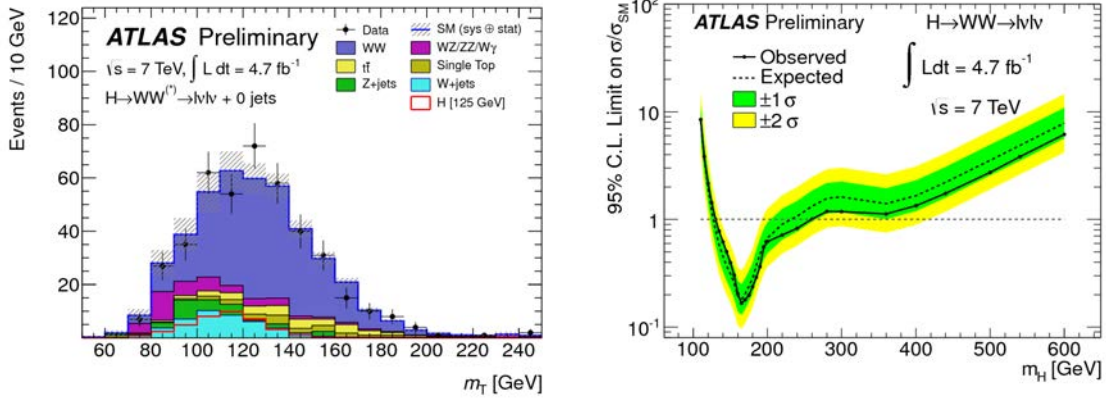


Fig. 28: (Left): The distribution of the transverse mass m_T in the $H+0$ jet channel of the ATLAS analysis. The expected signal for a Standard Model Higgs boson with $m_H = 125$ GeV is superimposed. (Right): Expected (dashed) and observed (solid) 95% C.L. upper limits on the cross section, normalized to the Standard Model cross section, as a function of m_H . The results at neighbouring mass points are highly correlated due to the limited mass resolution in this final state (from Ref. [127]).

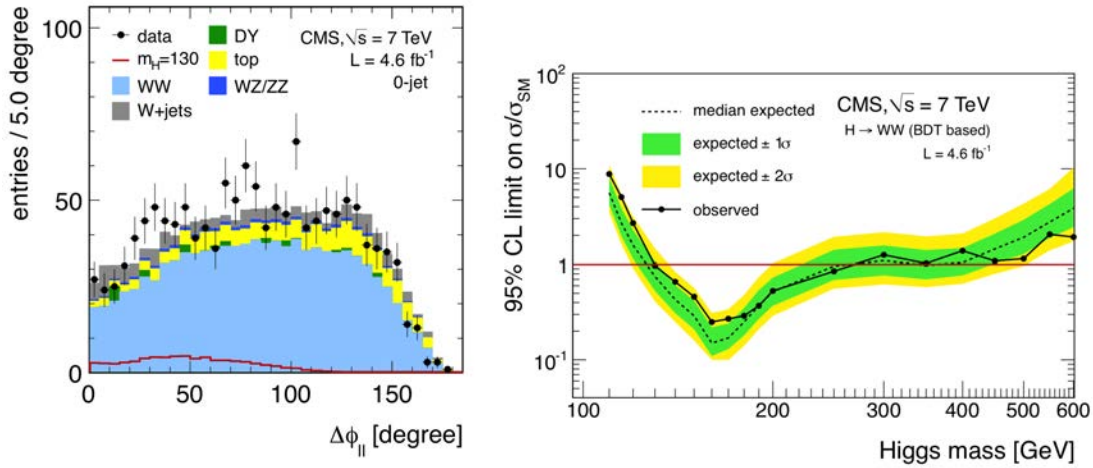


Fig. 29: (Left): The distribution of the azimuthal angle separation $\Delta\phi_{\ell\ell}$ in the $H+0$ jet channel of the CMS analysis. The expected signal for a Standard Model Higgs boson with $m_H = 130$ GeV is superimposed. (Right): Expected (dashed) and observed (solid) 95% C.L. upper limits on the cross section, normalized to the Standard Model cross section, as a function of m_H . The results at neighbouring mass points are highly correlated due to the limited mass resolution in this final state (from Ref. [128]).

($\Delta\phi_{\ell\ell}$) between the two selected leptons for events in the 0-jet category in the CMS experiment. Also this distribution is well described by the expected background processes. To enhance the sensitivity, the CMS experiment exploits two different analysis strategies for the 0-jet and 1-jet categories, the first one using a cut-based approach and the second one using a multivariate technique [128].

Since no significant excesses of events are found in any of the event categories in both the ATLAS and CMS experiments, upper limits on the production cross section are set, as shown in Figs. 28 and 29. The ATLAS experiment excludes the existence of a Standard Model Higgs boson over a mass range from 130 - 260 GeV, while the expected exclusion, in case no Higgs boson is present, is $127 \leq m_H \leq$

234 GeV. The CMS experiments excludes a mass range from 129–270 GeV, with an expected range from 127–270 GeV.

5.3.5 Search for $H \rightarrow \tau\tau$ and $H \rightarrow b\bar{b}$ decays

In addition to the searches described above, the search for the Higgs boson has also been performed in the $H \rightarrow \tau\tau$ [130, 131] and $H \rightarrow b\bar{b}$ final states [132, 133]. These searches do not yet reach the sensitivity of the others described above. They are, however, included in the overall combination of the results of the two collaborations [134, 135]. The observed and expected cross section limits are included in Fig. 31.

5.3.6 Search for the Higgs boson in the high mass region

For higher Higgs boson masses ($m_H > 2m_Z$) the decays $H \rightarrow WW$ and $H \rightarrow ZZ$ dominate. Due to the higher mass and improved signal-to-background conditions, also the decays $H \rightarrow ZZ \rightarrow \ell\ell\nu\nu$ [136, 137], $H \rightarrow ZZ \rightarrow \ell\ell qq$ [138, 139], $H \rightarrow ZZ \rightarrow \ell\ell\tau\tau$ [140] and $H \rightarrow WW \rightarrow \ell\nu qq$ [141] provide additional sensitivity.

The $H \rightarrow ZZ \rightarrow \ell\ell\nu\nu$ is the most sensitive channel. The selection of two leptons and large missing transverse energy gives rather good signal-to-background conditions. The dominant backgrounds are from diboson and $t\bar{t}$ production. Also in this case the transverse mass m_T of the $\ell\ell - E_T^{\text{miss}}$ system is used as discriminating variable. The distributions are shown in Fig. 30 for the ATLAS (left) and CMS (right) experiments together with expected signals at 400 GeV. No indications for excesses are seen and upper limits on the Higgs boson production cross sections are set. They are included as well in Fig. 31.

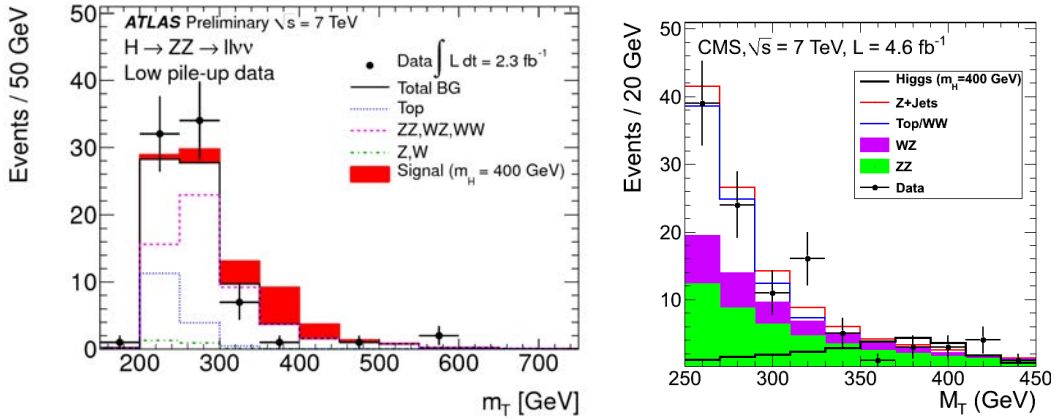


Fig. 30: The distributions of the transverse mass of $H \rightarrow ZZ \rightarrow \ell\ell\nu\nu$ candidates in the ATLAS (left) and CMS (right) experiments. Expected signals for a Higgs boson with a mass of 400 GeV are superimposed (from Refs. [136, 137]).

5.4 Combination results of searches for the Standard Model Higgs boson

5.4.1 Excluded mass ranges

The ATLAS and CMS experiments have combined their respective search results on the Standard Model Higgs boson [134, 135]. The combination procedure is based on the profile likelihood ratio test statistic $\lambda(\mu)$ [121], which extracts the information on the signal strength $\mu = \sigma/\sigma_{\text{SM}}$ from a full likelihood including all the parameters describing the systematic uncertainties and their correlations. More details on the statistical procedure used are described in Ref. [119].

In Fig. 31 the expected and observed 95% C.L. limits are shown from the individual channels entering this combination, separately for the ATLAS and CMS experiments. The combined 95% C.L.

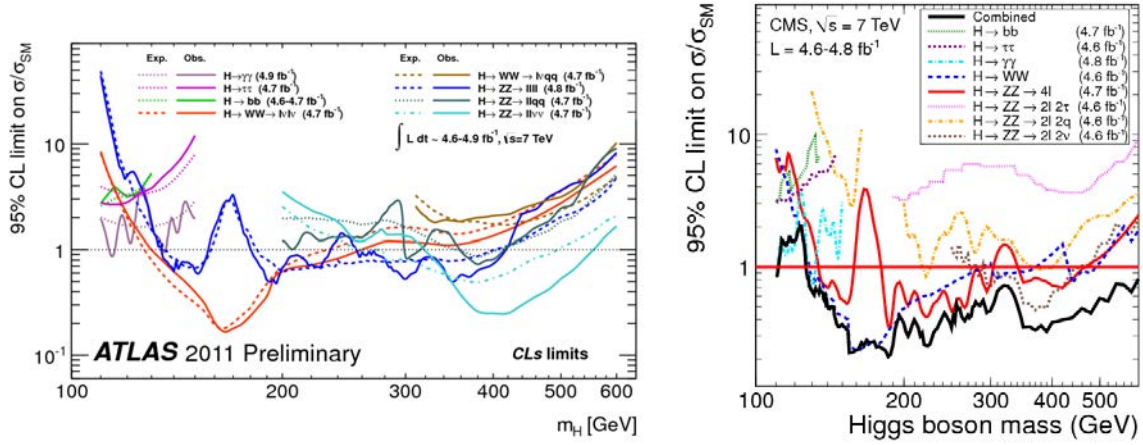


Fig. 31: The observed (solid) and expected (dashed) 95% C.L. cross section upper limits for the individual search channels in the ATLAS (left) and CMS (right) experiments, normalized to the Standard Model Higgs boson production cross section, as a function of the Higgs boson mass. The expected limits are those for the background-only hypothesis, i.e. in the absence of a Higgs boson signal (from Refs. [134, 135]).

exclusion limits are shown in Fig. 32 as a function of m_H for the full mass range and for the low mass range. The combined expected 95% C.L. exclusion regions for the two experiments are very similar and cover the m_H range from 120 to 555 GeV for the ATLAS and from 118 to 543 GeV for the CMS experiment. Based on the observed limit, the ATLAS experiment excludes at the 95% C.L. the Standard Model Higgs boson in three mass ranges: from 110.0–117.5 GeV, from 118.5 to 122.5 GeV and from 129 to 539 GeV. The 95% C.L. CMS exclusion covers the range 127–600 GeV. The observed exclusion covers a large part of the expected exclusion range, with the exception of the low mass region where an excess of events above the expected background is observed. It is striking that both experiments are not able to cover the mass window from about 118 to 129 GeV, despite their sensitivity in this range.

5.4.2 Compatibility with the background-only hypothesis

Excesses of events are observed in the ATLAS experiment near 126 GeV in the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^{(*)} \rightarrow \ell\ell\ell\ell$ channels, both of which provide fully reconstructed candidates with high-resolution in invariant mass. The CMS experiment observes two localized excesses, one at 119.5 GeV associated with three $Z \rightarrow 4\ell$ events and the other one at 124 GeV, arising mainly from the $\gamma\gamma$ channel. In addition, a broad offset of about one standard deviation is seen for the low resolution channels $H \rightarrow WW$, $H \rightarrow \tau\tau$ and $H \rightarrow b\bar{b}$. The observed local p_0 values, calculated using the asymptotic approximation, as a function of m_H and the expected value in the presence of a Standard Model Higgs boson signal are shown in Fig. 33 in the low mass region for the two experiments.

In the ATLAS data the local significance for the combined result reaches 2.6σ for $m_H=126$ GeV with an expected value in the presence of a signal at that mass of 2.9σ . The local significance for the combination of the CMS channels at $m_H = 124$ GeV amounts to 3.1σ .

The significance of the excesses is mildly sensitive to energy scale systematic (ESS) uncertainties and the resolution for photons and electrons. The observed effect of the ESS uncertainty is small and reduces the maximum local significance in the ATLAS experiment from 2.6σ to 2.5σ .

The global p_0 for local excesses depends on the range of m_H and the channels considered. The global probability for an excess as large as the one observed in the ATLAS combination at 126 GeV to occur anywhere in the mass range 110–600 GeV is estimated to be approximately 30%, decreasing to 10% in the range 110–146 GeV, which is not excluded at the 99% confidence level by the LHC combined Standard Model Higgs boson search [142]. The global significance for the CMS excess is estimated to

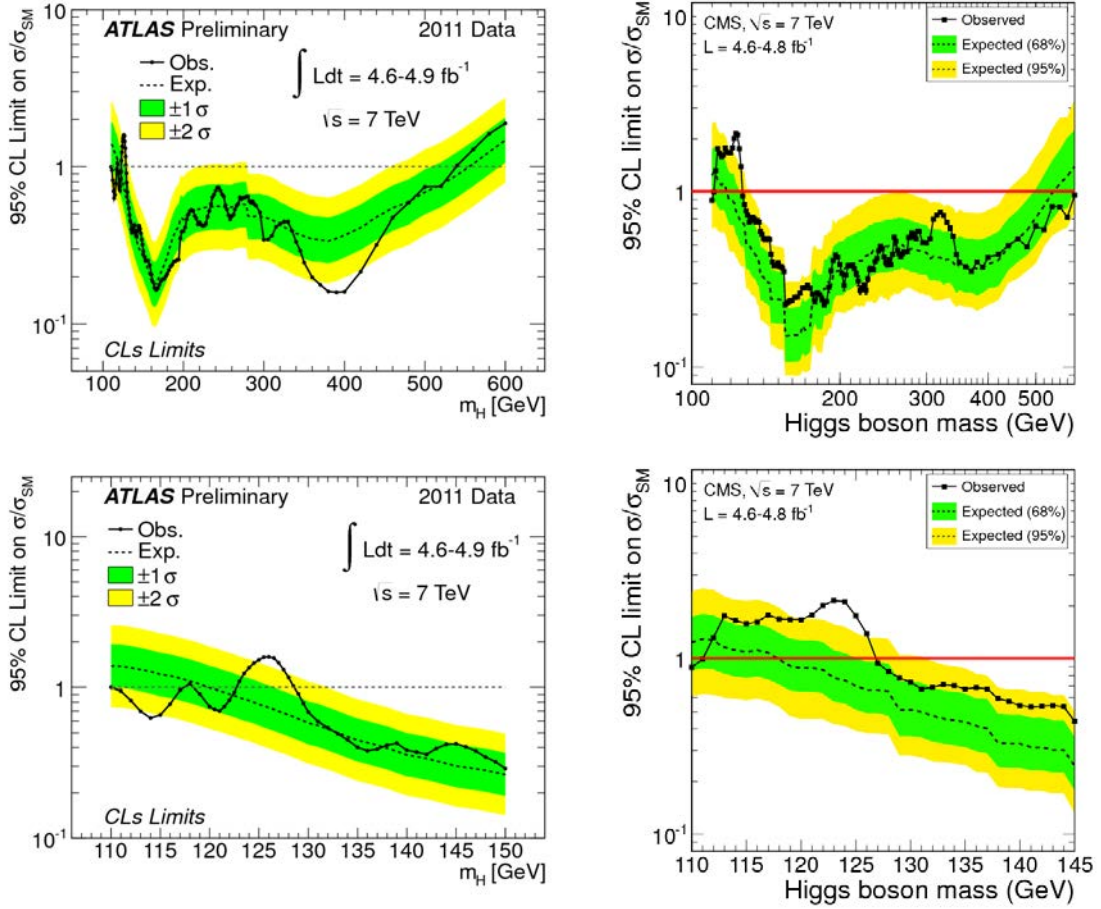


Fig. 32: The observed (full line) and expected (dashed line) 95% C.L. combined upper limits on the Standard Model Higgs boson production cross section divided by the Standard Model expectation as a function of m_H in the full mass range considered in the analyses (top) and in the low mass range (bottom) for the ATLAS (left) and CMS (right) experiments. The dotted curves show the median expected limit in the absence of a signal and the green and yellow bands indicate the corresponding 68% and 95% intervals (from Refs. [134, 135]).

be 1.5σ for the full search range from 110–600 GeV and 2.1σ for the restricted search range from 110–145 GeV.

The best-fit value of μ , denoted $\hat{\mu}$, is displayed for the combination of all channels for the two experiments in Fig. 34. The bands around $\hat{\mu}$ illustrate the μ interval corresponding to $-2 \ln \lambda(\mu) < 1$ and represent an approximate $\pm 1\sigma$ variation. The excess observed for $m_H = 126$ GeV in the ATLAS experiment corresponds to $\hat{\mu}$ of approximately $0.9^{+0.4}_{-0.3}$, which is compatible with the signal strength expected from a Standard Model Higgs boson at that mass ($\mu = 1$). Also for the CMS experiment the $\hat{\mu}$ values are within one sigma of unity in the mass range from 117–126 GeV.

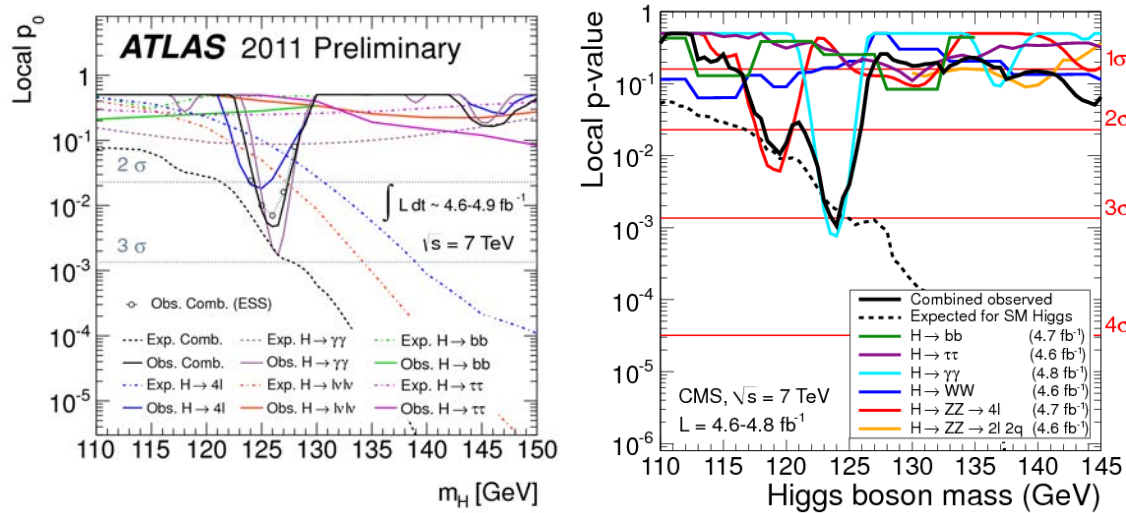


Fig. 33: The local probability p_0 for a background-only experiment to be more signal-like than the observation in the low mass range as a function of m_H for the ATLAS (left) and CMS (right) experiments. The p_0 values are shown for individual channels as well as for the combination. The dashed curves show the median expected local p_0 under the hypothesis of a Standard Model Higgs boson production signal at that mass. The horizontal dashed lines indicate the p values corresponding to significances of 1σ to 5σ (from Refs. [134, 135]).

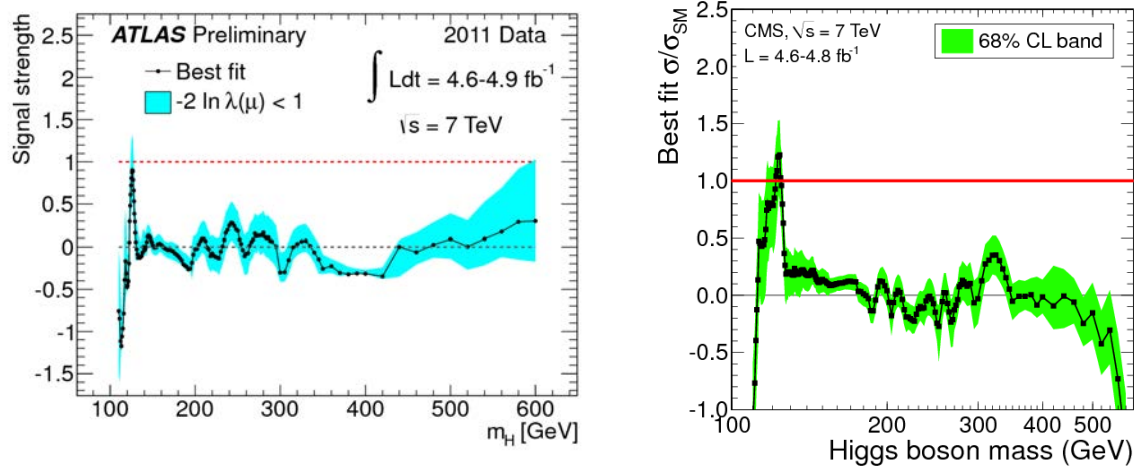


Fig. 34: The best-fit signal strength $\hat{\mu}$ as a function of the Higgs boson mass hypothesis in the full mass range for the combination of the ATLAS (left) and CMS (right) analyses. The μ value indicates by what factor the Standard Model Higgs boson cross section would have to be scaled to best match the observed data. The band shows the interval around $\hat{\mu}$ corresponding to a variation of $-2 \ln \lambda(\mu) < 1$ (from Refs. [134, 135]).

6 Search for Supersymmetric Particles

Due to the high centre-of-mass energy of 7 TeV, the LHC has a large discovery potential for new heavy particles beyond the Tevatron limits. This holds in particular for particles with colour charge, such as squarks and gluinos in supersymmetry (SUSY) [143, 144]. However, due to the excellent luminosity performance of the LHC in 2011, sensitivity also exists for electroweak production of charginos and neutralinos, the supersymmetric partners of the electroweak gauge bosons and the Higgs boson. In the following a few results of the searches by the ATLAS and CMS collaborations for supersymmetry with up to 2 fb^{-1} of LHC pp data at $\sqrt{s} = 7 \text{ TeV}$ are summarized. Since none of the analyses have observed any excess above the Standard Model expectations, limits on SUSY parameters or masses of SUSY particles are set. The discussion presented here follows largely the review of Ref. [145] on the results from the ATLAS collaboration.

6.1 Searches with jets and missing momentum

Assuming conservation of R-parity, the lightest supersymmetric particle (LSP) is stable and weakly interacting, and will typically escape detection. If the primary produced particles are squarks or gluinos (and assuming a negligible lifetime of these particles), they will decay to final states with energetic jets and significant missing transverse momentum. This final state can be produced in a large number of R-parity conserving models [146], in which squarks, $\tilde{q}$, and gluinos, $\tilde{g}$, can be produced in pairs as $\tilde{g}\tilde{g}$, $\tilde{g}\tilde{q}$, or $\tilde{q}\tilde{q}$. They can decay via $\tilde{q} \rightarrow q\tilde{\chi}_1^0$ and $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_1^0$ to weakly interacting neutralinos, $\tilde{\chi}_1^0$. However, also charginos or heavier neutralinos might appear in the decay cascade and these particles may produce high transverse momentum leptons in their decays into the LSP.

The ATLAS and CMS collaborations have carried out analyses with a lepton veto [147, 148], requiring one isolated lepton [149, 150], or requiring two or more leptons [151, 152]. In addition, a dedicated search was performed for events with high jet multiplicity with six or more jets [153]. Data samples corresponding to integrated luminosities between 1.0 and 1.3 fb^{-1} were used. Events are triggered either on the presence of a jet plus large missing momentum, or on the presence of at least one high- p_T lepton. Backgrounds to the searches arise from Standard Model processes such as vector boson production plus jets (W + jets, Z + jets), top quark pair production and single top production, QCD multijet production, and diboson production. They are estimated in a semi-data-driven way, using control regions in combination with a transfer factor obtained from simulation. The results are interpreted in the MSUGRA/CMSSM model [154], and in particular as limits in the plane spanned by the common scalar mass parameter at the GUT scale m_0 and the common gaugino mass parameter at the GUT scale $m_{1/2}$, for values of the common trilinear coupling parameter $A_0 = 0$, Higgs mixing parameter $\mu > 0$, and ratio of the vacuum expectation values of the two Higgs doublets $\tan \beta = 10$. Figure 35 (left) shows the results for the analyses of the ATLAS collaboration with ≥ 2 , ≥ 3 or ≥ 4 jets plus missing transverse momentum, and the multijets plus missing momentum analysis. For a choice of parameters leading to equal squark and gluino masses, squark and gluino masses below approximately 1 TeV are excluded. The 1-lepton and 2-lepton results are less constraining in MSUGRA/CMSSM for this choice of parameters, but these analyses are complementary, and therefore no less important. The exclusion contours obtained by the CMS collaboration in different final states, including the lepton channels, are shown in Fig. 36.

6.2 Simplified model interpretation

The various analyses have also been interpreted in simplified models assuming specific production and decay modes. The constraints implied by the MSUGRA/CMSSM models [154] are relaxed, leaving more freedom for the variation of particle masses and decay modes. Interpretations in simplified models thus show better the limitations of the analyses as a function of the relevant kinematic variables.

Inclusive search results with jets and missing momentum are interpreted using simplified models with either pair production of squarks or of gluinos, or production of squark-gluino pairs. Direct squark

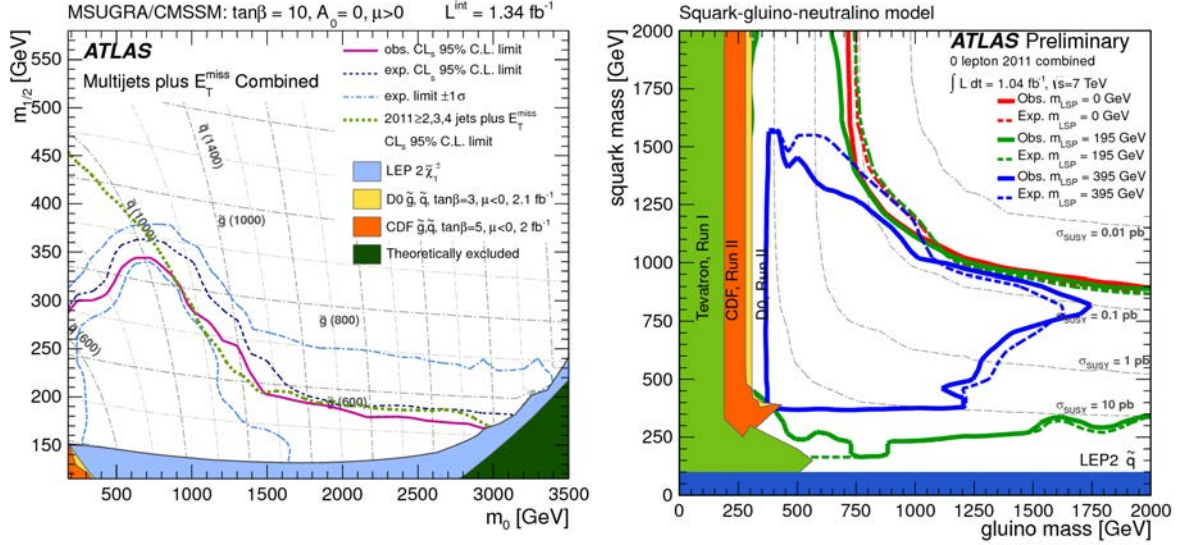


Fig. 35: (Left): Exclusion contours in the MSUGRA/CMSSM ($m_0 - m_{1/2}$)-plane for $A_0 = 0$, $\tan\beta = 10$ and $\mu > 0$, arising from the analysis of the ATLAS collaboration with ≥ 2 , ≥ 3 or ≥ 4 jets plus missing transverse momentum, and the multijets plus missing momentum analysis (from Ref. [153]). (Right): Exclusion contours from the ATLAS analyses in the squark-gluino mass plane for three values of the LSP mass using the simplified model description (see text) (from Ref. [155]).

decays ($\tilde{q} \rightarrow q\tilde{\chi}_1^0$) or direct gluino decays ($\tilde{g} \rightarrow q\tilde{q}\tilde{\chi}_1^0$) are dominant if all other particle masses have multi-TeV values, so that those do not play a role. Using these assumptions, the excluded mass regions are sensitive to the mass of the LSP ($\tilde{\chi}_1^0$). Figure 35 (right) shows the ATLAS results interpreted in terms of limits on (first and second generation) squark and gluino masses, for three values of the LSP ($\tilde{\chi}_1^0$) mass, and assuming that all other SUSY particles are very massive [155]. Further interpretations are also done in terms of limits on gluino mass versus LSP mass assuming high squark masses, or in terms of limits on squark mass vs LSP mass assuming large gluino masses [149, 155].

The results of the inclusive jets plus missing momentum searches, interpreted in these simplified models, indicate that masses of first and second generation squarks and of gluinos must be above approximately 750 GeV. An important caveat in this interpretation is the fact that this is only true for neutralino LSP masses below approximately 250 GeV (as in MSUGRA/CMSSM [154] for values of $m_{1/2}$ below ~ 600 GeV). For higher LSP masses, the squark and gluino mass limits are significantly less restricting. It will be a challenge for further analyses to extend the sensitivity of inclusive squark and gluino searches to the case of heavy neutralinos. If the LSP is heavy, events are characterized by less energetic jets and less missing transverse momentum. This will be more difficult to trigger on, and lead to higher Standard Model backgrounds in the analysis.

6.3 Search for stop and sbottom production

Important motivations for electroweak-scale supersymmetry are the facts that SUSY might provide a natural solution to the hierarchy problem by preventing ‘unnatural’ fine-tuning of the Higgs sector, and that the lightest stable SUSY particle is an excellent dark matter candidate. It is instructive to consider what such a motivation really requires from SUSY: a relatively light top quark partner (the stop, $\tilde{t}$ and an associated sbottom-left, $\tilde{b}_L$), a gluino not much heavier than about 1.5 TeV to keep the stop light, given that it receives radiative corrections from loops like $\tilde{t} \rightarrow \tilde{g}t \rightarrow \tilde{t}$, and electroweak gauginos below the TeV scale [156]. There are no strong constraints on first and second generation squarks and sleptons; in fact heavy squarks and sleptons make it easier for SUSY to satisfy the strong constraints from flavour

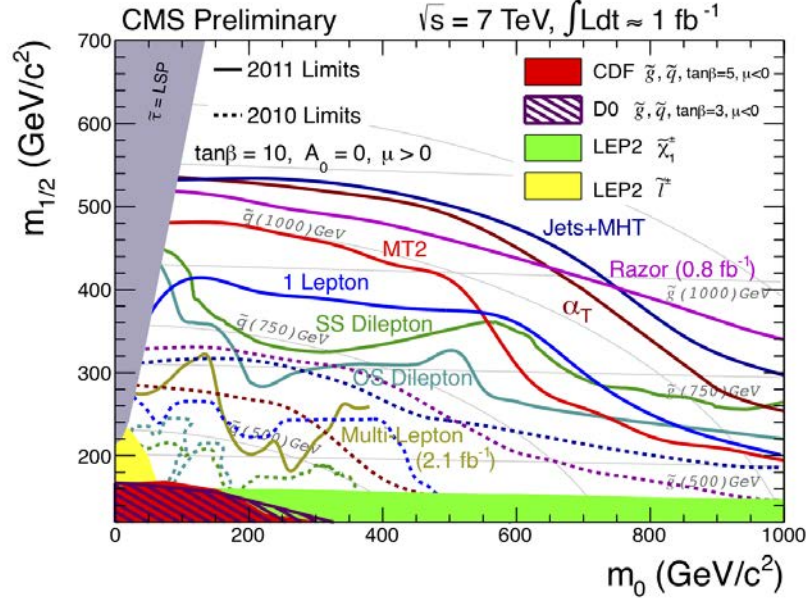


Fig. 36: Summary of exclusion contours in the MSUGRA/CMSSM ($m_0 - m_{1/2}$)-plane for the parameters $A_0 = 0$, $\tan \beta = 10$ and $\mu > 0$ for various analyses and different final states from the CMS collaboration (from Ref. [150]).

physics.

Motivated by these considerations, the ATLAS and CMS collaborations have also carried out a number of searches for supersymmetry with b -tagged jets, which are sensitive to sbottom and stop production, either in direct production or in production via gluino decays. Jets are tagged as originating from b -quarks by an algorithm that exploits both track impact parameter and secondary vertex information.

Direct sbottom pair production is searched for in a data sample corresponding to an integrated luminosity of 2 fb^{-1} by requiring two b -tagged jets with $p_T > 130, 50 \text{ GeV}$ and significant missing transverse momentum of more than 130 GeV [157]. The final discriminant in the ATLAS analysis is the boost-corrected contranverse mass m_{CT} [158], and signal regions with $m_{CT} > 100, 150, 200 \text{ GeV}$ are considered. No excesses are observed above the expected backgrounds from top, W +heavy flavour and Z +heavy flavour production. Figure 37 (left) shows the resulting limits in the sbottom-neutralino mass plane, assuming sbottom pair production and sbottom decays into a b -quark plus a neutralino (LSP) with a 100% branching fraction. Under these assumptions, sbottom masses up to 390 GeV are excluded for neutralino masses below 60 GeV .

The ATLAS collaboration has searched for stop quark production in gluino decays [159] using an analysis requiring at least four high- p_T jets of which at least one should be b -tagged, one isolated lepton, and significant missing transverse momentum. Since the number of observed events agrees with the expectations from Standard Model processes, limits are set in the gluino-stop mass plane, assuming the gluino to decay as $\tilde{g} \rightarrow \tilde{t}t$, and the stop quark to decay as $\tilde{t} \rightarrow b\tilde{\chi}_1^\pm$. The obtained mass limits are shown in Fig. 37 (right).

Further searches for direct stop pair production are in progress. These searches are challenging due to the similarity with the top-quark pair-production final state for stop masses similar to the top mass, and due to the low cross section for the production of stops with high mass. The ATLAS collaboration has searched for signs of new phenomena in events passing a top-quark pair selection with large missing transverse momentum [160]. Such an analysis is sensitive to pair production of massive partners of the top quark, decaying to a top quark and a long-lived undetected neutral particle. No excess above background was observed, and limits on the cross section for pair production of top quark partners are

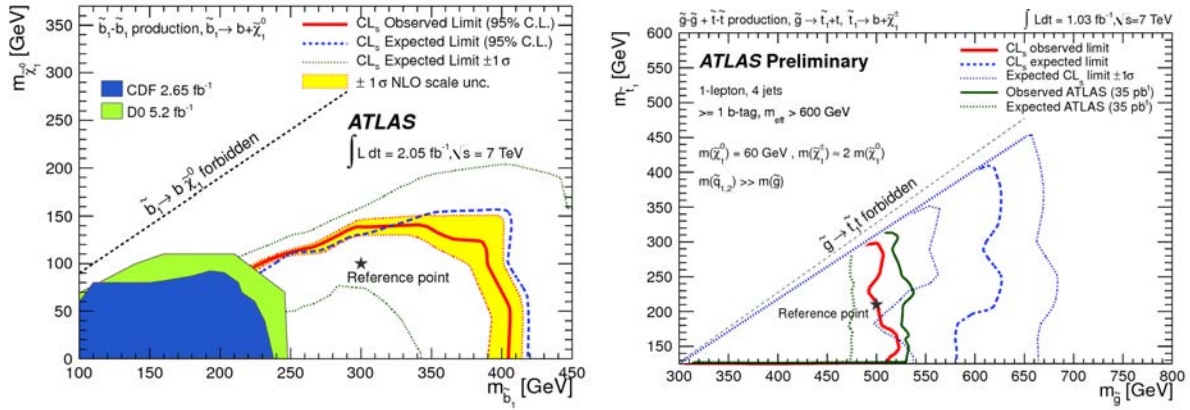


Fig. 37: (Left): Exclusion contours from the ATLAS analyses in the sbottom-neutralino mass plane resulting from the analysis searching for sbottom pair production assuming $\tilde{b}_1 \rightarrow b \tilde{\chi}_1^0$ decays (from Ref. [157]). (Right): Exclusion contours from the ATLAS analyses in the gluino-stop mass plane resulting from the analysis searching for stop production via gluino decays. The assumptions made to derive the limits are given in the figure (from Ref. [159]).

set. These limits constrain fermionic exotic fourth generation quarks, but not yet scalar partners of the top quark, such as the stop quark [160].

6.4 Search for supersymmetry in multilepton final states

The search for final states with several leptons and missing transverse momentum are sensitive to the production of charginos and/or heavier neutralinos (other than the LSP), decaying leptonically into the LSP. These analyses comprise the golden search modes at the Tevatron, but are also rapidly gaining relevance at the LHC and both the ATLAS and CMS collaborations have performed corresponding analyses [151, 152]. The ATLAS collaboration has published results of various analyses searching for dilepton events plus missing momentum in data corresponding to an integrated luminosity of 1.0 fb^{-1} [151]. Three searches are performed for new phenomena in final states with opposite-sign and same-sign dileptons and missing transverse momentum. These searches also include signal regions that place requirements on the number and p_T of energetic jets in the events. For all signal regions good agreement is found between the numbers of observed events and the predictions of expected events from Standard Model processes. Additionally, in opposite-sign events, a search is made for an excess of same-flavour over different-flavour lepton pairs. Effective production cross sections in excess of 9.9 fb for opposite-sign events with missing transverse momentum greater than 250 GeV are excluded at 95% C.L. For same-sign events with missing transverse momentum greater than 100 GeV , effective production cross sections in excess of 14.8 fb are excluded at 95% C.L. The latter limit is interpreted in a simplified electroweak gaugino production model excluding chargino masses up to 200 GeV , under the assumption that slepton decays are dominant [151].

The CMS collaboration has presented preliminary results, based on data corresponding to an integrated luminosity of 2.1 fb^{-1} , on the search for supersymmetric particles in three- and four-lepton final states, including hadronic decays of τ leptons [152]. The backgrounds from Standard Model processes are suppressed by requiring missing transverse energy, Z-mass vetos of the invariant dilepton mass or high jet activity. Control samples in data are used to obtain reliable background estimates. Within the statistical and systematic uncertainties the numbers of observed events are consistent with the expectations from Standard Model processes. These results are used to exclude previously unexplored regions of the supersymmetric parameter space assuming R-parity conservation with the lightest supersymmet-

ric particle being a neutralino. The corresponding exclusion contours in the MSUGRA/CMSSM [154] interpretation are shown in Fig. 38 in the $(m_0 - m_{1/2})$ plane for $A_0 = 0$, $\mu > 0$, and for $\tan\beta$ values of 3 and 10. They extend significantly the regions excluded by the CDF [161] and DØ [162] experiments and those excluded with previous searches at the LHC [163].

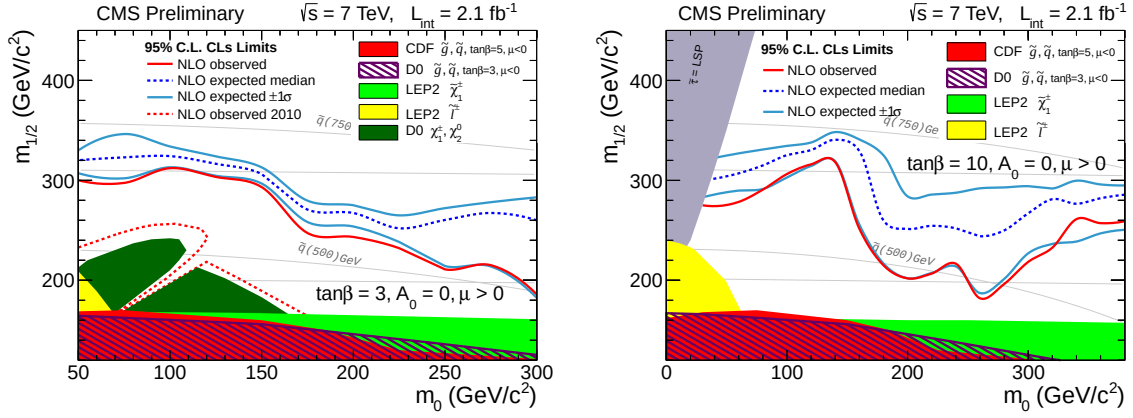


Fig. 38: Exclusion contours in the MSUGRA/CMSSM $(m_0 - m_{1/2})$ -plane for the parameters $A_0 = 0$, $\mu > 0$ and $\tan\beta = 3$ (left) and $\tan\beta = 10$ (right) obtained from searches for SUSY production in final states with multileptons by the CMS collaboration (from Ref. [152]).

6.5 Summary and outlook on SUSY searches

Many different searches for the production of supersymmetric particles have been performed in a large variety of final states by the ATLAS and CMS collaborations at the LHC. Data corresponding to integrated luminosities in the range between 1.0 and 4.7 fb^{-1} taken during the year 2011 have been analyzed. In all channels, the number of observed events is in agreement with the expectations from Standard Model processes and no evidence for the production of supersymmetric particles has been found so far. The data have been used to set already rather strong limits on the masses of possible supersymmetric particles. A summary of the most important mass limits is given in Fig. 39.

In addition to the analyses summarized here, many other analyses have been performed and many different final states have been explored. There are investigations of SUSY searches in gauge mediated supersymmetry breaking models, by using final states with photons or multileptons. In addition, in many models (split SUSY, R-hadrons, anomaly-mediated SUSY breaking and in certain parts of the phase space of gauge-mediated SUSY breaking scenarios) SUSY particles may be long-lived either because their decay is kinematically suppressed or due to very small couplings, e.g. in R-parity violating models. Many of these scenarios have already been explored and the reader is referred to the corresponding publications of the ATLAS and CMS collaborations. Also in all these searches for more exotic SUSY scenarios the number of observed events is in agreement with the expectations from background from Standard Model processes.

Although no signs of supersymmetry have been found so far, it is important to realize that actual tests of ‘natural’ supersymmetry are only just beginning. In this respect, the LHC run of 2012, with an expected luminosity of more than 10 fb^{-1} , possibly at $\sqrt{s} = 8 \text{ TeV}$, will be very important. However, experimentally there will be considerable challenges in triggering and in dealing with high pile-up conditions. In the longer term, increasing the LHC beam energy to $> 6 \text{ TeV}$ will again enable the crossing of kinematical barriers and open the way for multi-TeV SUSY searches.

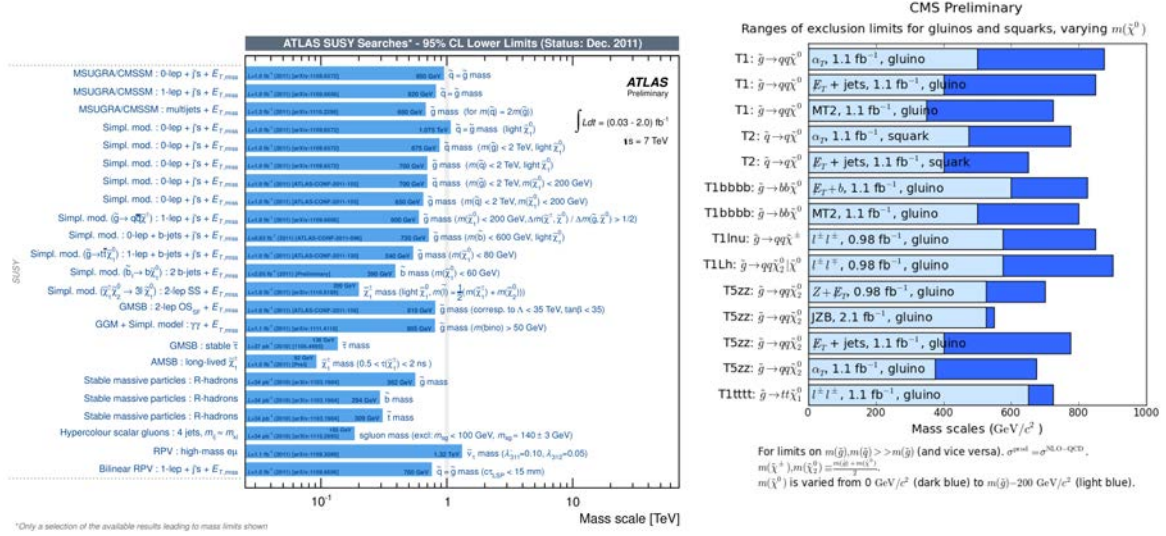


Fig. 39: Summary of excluded mass ranges from a variety of searches for the production of supersymmetric particles from the ATLAS (left) and CMS (right) collaborations. Only a representative selection of available results is shown. The CMS results indicate the change of the limits under variation of the neutralino mass from 0 to 200 GeV.

7 Search for other Physics Scenarios Beyond the Standard Model

As already mentioned in Section 2, the Standard Model is an extremely successful effective theory which has been extensively tested over the past forty years. However, a number of fundamental questions are left unanswered. Many models for physics Beyond the Standard Model (BSM) have been proposed and the ATLAS and CMS experiments have used the data collected in 2010 and 2011 to search for indications of new physics. An impressive list of analyses has been performed. So far, no indications for deviations from the Standard Model have been found. The event numbers and kinematical distributions in all final states considered agree with the expectations from Standard Model processes. Therefore, these analyses have been used to constrain the parameter space of many BSM models.

Since it is impossible to present and discuss all analyses in such a summary paper, a few benchmark processes are selected and the search results are presented in the following. This concerns the search for new vector bosons, or more general the search for heavy dilepton resonances, the search for compositeness and the search for dijet resonances. Finally the results from other searches are briefly summarized.

7.1 Search for heavy dilepton resonances

The ATLAS and CMS collaborations have performed searches for narrow high-mass neutral and charged resonances decaying into e^+e^- or $\mu^+\mu^-$ pairs or $e\nu$ or $\mu\nu$, respectively. In several extensions of the Standard Model new heavy spin-1 neutral gauge bosons such as Z' [164–166], technimesons [167–169], as well as spin-2 Randall-Sundrum gravitons, G^* , [170] are predicted. Additional heavy charged gauge bosons appear e.g. in left-right-symmetric models [171].

The benchmark models considered in the analyses for the Z' are the Sequential Standard Model [164], with the same couplings to fermions as the Z boson, and the E_6 grand unified symmetry group [166], broken into $SU(5)$ and two additional $U(1)$ groups, leading to new neutral gauge fields ψ and χ . The particles associated with the additional fields can mix in a linear combination to form the Z' candidate: $Z'(\theta_{E_6}) = Z'_\psi \cos \theta_{E_6} + Z'_\chi \sin \theta_{E_6}$, where θ_{E_6} is the mixing angle between the two gauge bosons. The

pattern of spontaneous symmetry breaking and the value of θ_{E_6} determine the Z' couplings to fermions.

Other models predict additional spatial dimensions as a possible explanation for the gap between the electroweak symmetry breaking scale and the gravitational energy scale. The Randall-Sundrum (RS) model [170] predicts excited Kaluza-Klein modes of the graviton, which appear as spin-2 resonances. These modes have a narrow intrinsic width when $k/\bar{M}_{\text{Pl}} < 0.1$, where k is the spacetime curvature in the extra dimension, and $\bar{M}_{\text{Pl}} = M_{\text{Pl}}/\sqrt{8\pi}$ is the reduced Planck scale.

The search performed by the ATLAS experiment [172] is based on a dataset corresponding to an integrated luminosity of up to 1.2 fb^{-1} . The observed invariant mass spectrum is shown in Fig. 40 (left) for the e^+e^- final state after final selections. The backgrounds from Drell-Yan, $t\bar{t}$, diboson and W +jets production are determined from Monte Carlo simulation after normalization to the respective (N)NLO cross sections. The background from QCD multijet production is estimated using data-driven methods with the inversion of lepton identification criteria. The simulated backgrounds are rescaled so that the total sum of the backgrounds matches the observed number of events observed in data in the 70–110 GeV mass interval. The scaling factor is within 1% of unity. The advantage of this approach is that the uncertainty on the luminosity and any mass independent uncertainties on efficiencies, cancel between the Z' (G^*) and the Z boson. The dilepton invariant mass distributions are well described by the prediction from Standard Model processes. Figure 40 (left) also displays the expected Z' signals in the Sequential Standard Model for three mass hypotheses.

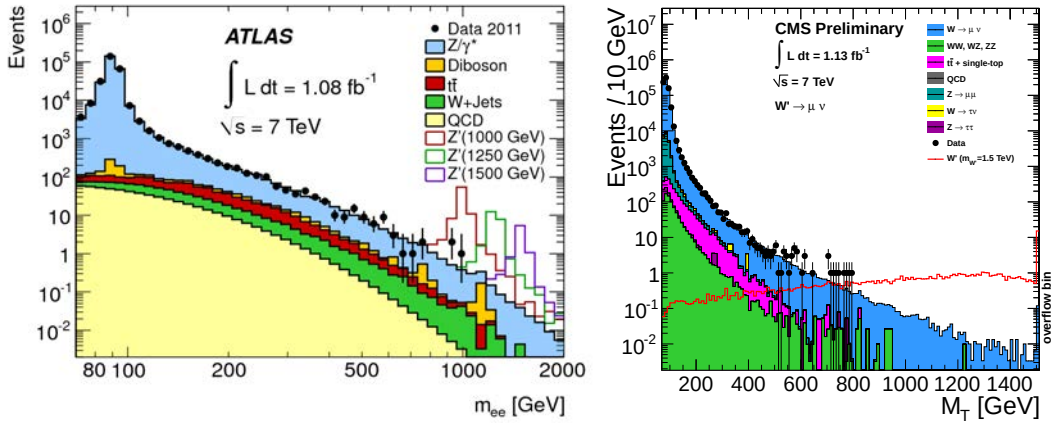


Fig. 40: (Left): Distribution of the dielectron invariant mass after final selections in the ATLAS experiment, compared to the stacked sum of all expected backgrounds, with three example Z'_{SSM} signals overlaid. The bin width is constant in $\log m_{\ell\ell}$ (from Ref. [172]). (Right): Distribution of the $\mu - E_{\text{T}}^{\text{miss}}$ transverse mass after final selections in the CMS experiment. The expected signal from a hypothetical W' boson with a mass of 1.5 TeV is superimposed (from Ref. [173]).

Given the good agreement between the data and the Standard Model expectations, limits are set on the cross section times branching ratio for the different Z' models. The resulting mass limits are 1.83 TeV for the Sequential Standard Model Z' boson, 1.49–1.64 TeV for various E_6 -motivated Z' bosons, and 0.71–1.63 TeV for a Randall–Sundrum graviton with couplings $(k/\bar{M}_{\text{Pl}})$ in the range 0.01–0.1. Similar analyses have been performed by the CMS collaboration [174] and comparable limits have been extracted. They are included in the summary of results from different experiments for various physics models in Table 1.

The benchmark model considered in the search for the W' is the Sequential Standard Model [164], with the same couplings to fermions as the W boson. In this case the transverse mass of the lepton and $E_{\text{T}}^{\text{miss}}$ system is used as discriminating variable. As an example, the measured transverse mass distribution in the muon final state in the CMS experiment [173] is shown in Fig. 40 (right). The expectation

Table 1: Observed 95% C.L. mass lower limits on Z' , G^* gravitons and W' resonances obtained for various models in the ATLAS and CMS experiments. The results from searches at the Tevatron are included for comparison.

Model	Experiment	L_{int} (fb ⁻¹)	95% C.L. limits			Ref.
			e^+e^- (TeV)	$\mu^+\mu^-$ (TeV)	$\ell^+\ell^-$ (TeV)	
Z'_{SSM}	CDF/DØ	5.5			1.07	[176]
	ATLAS/CMS	0.036	0.96	0.83	1.05/1.14	[177]
	ATLAS	1.1 / 1.2	1.70	1.61	1.83	[172]
	CMS	1.1			1.94	[174]
$Z' E_6$ models	ATLAS	1.1 / 1.2			1.49 - 1.64	[172]
	CMS	1.1			1.62	[178]
$G^* k/\bar{M}_{\text{Pl}} = 0.01$	ATLAS	1.1 / 1.2			0.71	[172]
$G^* k/\bar{M}_{\text{Pl}} = 0.03$		1.1 / 1.2			1.03	
$G^* k/\bar{M}_{\text{Pl}} = 0.05$		1.1 / 1.2			1.33	
$G^* k/\bar{M}_{\text{Pl}} = 0.10$		1.1 / 1.2			1.63	
$G^* k/\bar{M}_{\text{Pl}} = 0.05$		1.1			1.45	
$G^* k/\bar{M}_{\text{Pl}} = 0.10$	CMS	1.1			1.78	[174]
W'_{SSM}	ATLAS	1.04	2.08	1.98	2.15	[175]
	CMS	1.1			2.27	[173]

for a W' signal with a mass of 1.5 TeV is superimposed. Also the transverse mass distributions measured by the LHC experiments are well described by the prediction from Standard Model processes and the data allow to exclude heavy W' bosons with masses below 2.15 TeV (ATLAS) [175] and 2.25 TeV (CMS) [173] at the 95% C.L.

The mass limits obtained at the LHC are the most stringent to date, including indirect limits set by LEP2. It is striking to see how fast the LHC experiments have superseded the limits obtained with much higher luminosity at the Tevatron. The analyses based on the data from 2010 ($L_{\text{int}} = 36 \text{ pb}^{-1}$) resulted in comparable limits to those obtained at the Tevatron based on an integrated luminosity of 5.5 fb^{-1} (see Table 1).

7.2 Limits on new physics from jet production

The measurements on inclusive and dijet production, as discussed in Section 4.1, can also be used to constrain contributions from new physics that would modify the expected QCD behaviour in the jet production cross sections. Two examples are discussed in the following.

7.2.1 Substructure of quarks

Both collaborations have searched for quark compositeness by investigating the angular distribution of jet events [179, 180]. At small scattering angles in the centre-of-mass system of the two partons, the angular distribution is expected to be proportional to the Rutherford cross section, $d\hat{\sigma}/d\cos\theta^* \sim 1/(1-\cos\theta^*)^2$. For the scattering of massless partons, which are assumed to be collinear with the beam protons, the longitudinal boost of the parton-parton centre-of-mass frame with respect to the proton-proton centre-of-mass frame, y_{boost} , and θ^* are obtained from the rapidities y_1 and y_2 of the jets from the two scattered partons by $y_{\text{boost}} = \frac{1}{2}(y_1 + y_2)$ and $|\cos\theta^*| = \tanh y^*$, where $y^* = \frac{1}{2}|y_1 - y_2|$ and where $\pm y^*$ are the rapidities of the two jets in the parton-parton centre-of-mass frame. The variable $\chi_{\text{dijet}} = e^{2y^*}$ is used to measure the dijet angular distribution, which for collinear massless-parton scattering takes the form $\chi_{\text{dijet}} = (1 + |\cos\theta^*|)/(1 - |\cos\theta^*|)$. This choice of χ_{dijet} , rather than θ^* , is motivated by the fact that $d\sigma_{\text{dijet}}/d\chi_{\text{dijet}}$ is flat for Rutherford scattering.

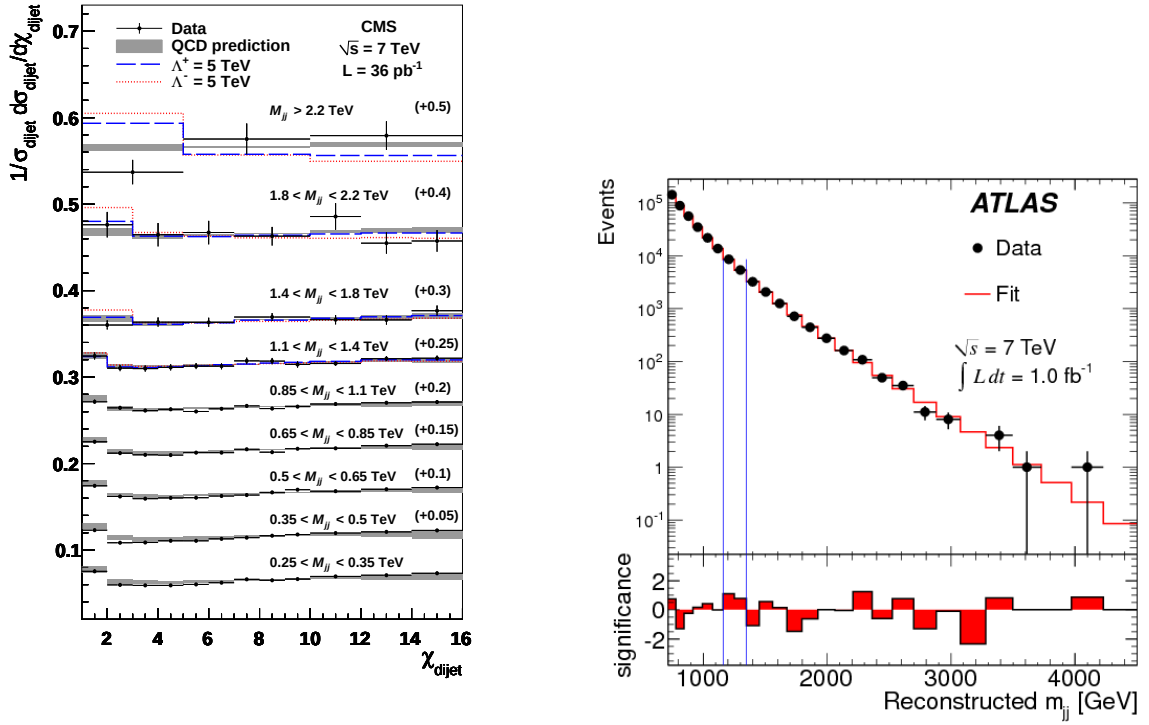


Fig. 41: (Left): Normalized dijet angular distributions in several ranges of the dijet mass as measured by the CMS collaboration. The data points include statistical and systematic uncertainties. The results are compared with the predictions of pQCD at NLO (shaded bands) and with the predictions including a contact interaction term of compositeness scale $\Lambda^+ = 5 \text{ TeV}$ (dashed histogram) and $\Lambda^- = 5 \text{ TeV}$ (dotted histogram). The shaded bands show the effect on the NLO pQCD predictions due to μ_R and μ_F scale variations and PDF uncertainties, as well as the uncertainties from the non-perturbative corrections added in quadrature (from Ref. [180]). (Right): The reconstructed dijet mass distribution (filled points) measured by the ATLAS collaboration fitted with a smooth functional form describing the QCD background. The bin-by-bin significance of the data-background difference is shown in the lower panel. Vertical lines show the most significant excess found (from Ref. [181]).

The differential dijet angular distributions for different m_{jj} ranges, and corrected for detector effects as measured by the CMS experiment using the 2010 data ($L_{\text{int}} = 36 \text{ pb}^{-1}$) are shown in Fig. 41 (left). The data are found to be in good agreement with pQCD predictions at NLO calculated with NLOJET++ [32, 33], which are superimposed on the figure. The measured dijet angular distributions can be used to set limits on quark compositeness parametrized by a four-fermion contact interaction term in addition to the QCD Lagrangian. The value of the mass scale Λ characterizes the strengths of the quark substructure binding interactions and the physical size of the composite states. A color- and isospin-singlet contact interaction (CI) of left-handed quarks gives rise to an effective Lagrangian term [182, 183]

$$L_{qq} = \eta_0 \frac{2\pi}{\Lambda^2} (\bar{q}_L \gamma^\mu q_L) (\bar{q}_L \gamma_\mu q_L), \quad (4)$$

where $\eta_0 = +1$ corresponds to destructive interference between the QCD and the new physics term, and $\eta_0 = -1$ to constructive interference. From the measured χ_{dijet} distribution, lower limits on the contact interaction scale of $\Lambda^+ = 5.6 \text{ TeV}$ and $\Lambda^- = 6.7 \text{ TeV}$ for destructive and constructive interference, respectively, have been set by the CMS collaboration at the 95% C.L. [180]. The expected limits in case of no substructure are 5.0 TeV and 5.8 TeV, respectively. The ATLAS collaboration has performed a similar analysis and excludes at the 95% C.L. quark contact interactions with a scale $\Lambda < 9.5 \text{ TeV}$ [179].

Table 2: The 95% C.L. mass lower limits on dijet resonance models.

Model	Experiment	L_{int} (fb ⁻¹)	95% C.L. limits		Ref.
			Expected (TeV)	Observed (TeV)	
Excited quark q^*	ATLAS	1.0	2.81	2.99	[181]
	CMS	1.0	2.68	2.49	[185]
Axigluon	ATLAS	1.0	3.07	3.32	[181]
	CMS	1.0	2.66	2.47	[185]
Colour Octet Scalar	ATLAS	1.0	1.77	1.92	[181]
E_6 diquarks	CMS	1.0	3.28	3.52	[185]

However, it should be noted that this observed limit is significantly above the expected limit of 5.7 TeV for the data sample corresponding to an integrated luminosity of 36 pb⁻¹. Very recently, the CMS collaboration has published the results of an updated analysis based on data corresponding to an integrated luminosity of 2.2 fb⁻¹ [184] and taking NLO calculations for the QCD predictions into account. Also this larger data set has been found to be in good agreement with the QCD expectations. For the contact interaction model described above, 95% C.L. limits of $\Lambda^+ = 7.5$ TeV and $\Lambda^- = 10.5$ TeV have been set. The expected limits are 7.0 TeV and 9.7 TeV, respectively.

7.2.2 Dijet resonances

The ATLAS and CMS collaborations have also examined the dijet mass spectrum for resonances due to new phenomena localised near a given mass, employing data-driven background estimates that do not rely on detailed QCD calculations [181, 185]. The searches are based on data corresponding to an integrated luminosity of 1.0 fb⁻¹. As an example, the observed dijet mass distribution measured in the ATLAS experiment, which extend up to masses of ~ 4 TeV is displayed in Fig. 41 (right). It is found to be in good agreement with a smooth function representing the Standard Model expectation. Since no evidence for the production of new resonances is found, 95% C.L. mass limits have been set in the context of several models of new physics: excited quarks (q^*) [186, 187], axigluons [188–190], scalar colour octet states [191] and scalar diquarks predicted in Grand Unified Theories based on the E_6 gauge group [192]. The results are summarized in Table 2. Also these limits are the most stringent ones to date.

7.3 Summary of results on other searches

Many different searches for the Beyond the Standard Model processes have been performed by the ATLAS and CMS collaborations at the LHC. Data corresponding to integrated luminosities in the range between 1.0 and 4.7 fb⁻¹ taken during the year 2011 have been analyzed and many different final states have been investigated. So far, no indications for deviations from the Standard Model have been found. The event numbers and kinematical distributions in all final states considered agree with the expectations from Standard Model processes. Therefore, these analyses have been used to constrain the parameter space of many BSM models. A summary of the most important limits from the ATLAS collaboration is given in Fig. 42. Comparable limits have been set by the CMS collaboration.

8 Conclusions

With the start-up of the operation of the LHC at high energies particle physics has entered a new era. Both the accelerator and the detectors have worked magnificently. Until the end of 2011 data corresponding to an integrated luminosity of 5.5 fb⁻¹ have been recorded with high efficiency by the LHC experiments. Based on these data, many tests of the predictions of the Standard Model and searches for physics Beyond the Standard Model have been performed in the new energy regime. So far, all measurements have been found to be in good agreement with the predictions from the Standard Model. Towards the end of 2011,

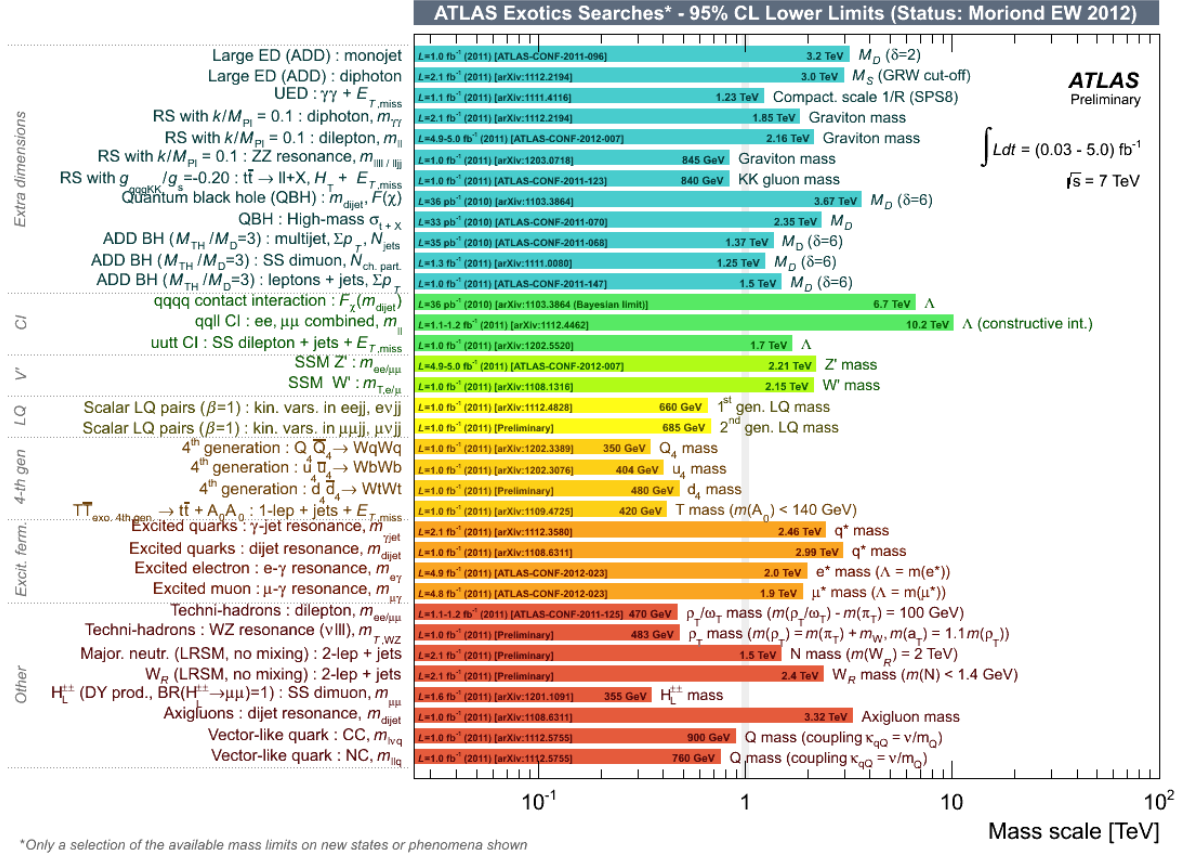


Fig. 42: Summary of excluded mass ranges from a variety of searches from the ATLAS experiment for Beyond the Standard Model physics processes. Only a representative selection of available results is shown.

the experiments have reached sensitivity for the Standard Model Higgs boson. A large fraction of the possible Higgs boson mass range has already been excluded by the ATLAS and CMS experiments with a confidence level of 95%. However, it is striking that both experiments are not able to exclude the existence of the Higgs boson in the mass range from 118–129 GeV, despite their sensitivity in this range. In addition, tantalizing hints for a Higgs boson signal have been seen by both experiments in the two high resolution channels $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^{(*)} \rightarrow llll$. However, the statistical significance is not sufficient to claim evidence. More data are needed to clarify the situation. With a successful run of the LHC in 2012 a final conclusion on the existence of the Standard Model Higgs boson might be reached and the year 2012 might enter as the “Year of the Higgs Boson” into the history of Physics.

References

- [1] S. Glashow, Nucl. Phys. **22** (1961) 579.
- [2] S. Weinberg, Phys. Rev. Lett. **19** (1967) 1264.
- [3] A. Salam, in *Elementary Particle Theory*, Ed. W. Svartholm (Almqvist and Wiksell, Stockholm, 1968).
- [4] H. D. Politzer, Phys. Rev. Lett. **30** (1973) 1346.
- [5] D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30** (1973) 1343.
- [6] H. Fritzsch and M. Gell-Mann, in *Proc. XVI Int. Conf. on High Energy Physics* (Fermilab, 1972).

- [7] P. Higgs, Phys. Lett. **12** (1964) 132.
- [8] F. Englert and R. Brout, Phys. Rev. Lett. **13** (1964) 321.
- [9] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Phys. Rev. Lett. **13** (1964) 585.
- [10] P. W. Higgs, Phys. Rev. Lett. **13** (1964) 508.
- [11] P. W. Higgs, Phys. Rev. **145** (1966) 1156.
- [12] T. W. B. Kibble, Phys. Rev. **155** (1967) 1554.
- [13] W. Alberico and S. Bilenky, Phys. Part. Nucl. **35** (2004) 297, and references therein.
- [14] For a review, see: M. Drees and G. Garbier, *Dark matter*, in 2011 Review of Particle Physics, K. Nakamura *et al.* (Particle Data Group), J. Phys. **G37**, 075021 (2010).
- [15] For a review, see: R. Gaitskell, Annual Review of Nuclear and Particle Science **54** (2004) 315, and references therein.
- [16] J. M. Campbell, J. W. Huston, and W. J. Stirling, Reports on Progress in Physics **70** (2007) 89.
- [17] L.N. Lipatov, Sov. J. Nucl. Phys. **20** (1975) 95; V.N. Gribov, L.N. Lipatov, Sov. J. Nucl. Phys. **15** (1972) 438; G. Altarelli, G. Parisi, Nucl. Phys. B **126** (1977) 298; L. Yu Dokshitzer, Sov. Phys. - JETP **46** (1977) 641.
- [18] R. Devenish and A. Cooper-Sarkar, *Deep inelastic scattering* (Oxford University Press, 2004).
- [19] S. Alekhin, J. Blümlein, S. Klein, and S. Moch, Phys. Rev. **D81** (2010) 014032, arXiv:0908.2766 [hep-ph].
- [20] P. M. Nadolsky *et al.*, Phys. Rev. **D78** (2008) 013004, arXiv:0802.0007 [hep-ph].
- [21] H.-L. Lai *et al.*, Phys. Rev. **D82** (2010) 074024, arXiv:1007.2241 [hep-ph].
- [22] The H1 and ZEUS Collaboration, F. D. Aaron *et al.*, JHEP **01** (2010) 109, arXiv:0911.0884 [hep-ex].
- [23] The H1 and ZEUS Collaboration, V. Radescu, *HERA Precision Measurements and Impact for LHC Predictions*, arXiv:1107.4193 [hep-ex].
- [24] P. Jimenez-Delgado and E. Reya, Phys. Rev. **D79** (2009) 074023, arXiv:0810.4274 [hep-ph].
- [25] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, Eur. Phys. J. **C63** (2009) 189, arXiv:0901.0002 [hep-ph].
- [26] R. D. Ball *et al.*, Nucl. Phys. **B838** (2010) 136, arXiv:1002.4407 [hep-ph].
- [27] R. D. Ball *et al.*, Nucl. Phys. **B849** (2011) 296, arXiv:1101.1300 [hep-ph].
- [28] The ATLAS Collaboration, Phys. Lett. **B701** (2011) 31, arXiv:1103.2929 [hep-ex].
- [29] The CMS Collaboration, S. Chatrchyan *et al.*, JHEP **04** (2011) 050, arXiv:1103.3470 [hep-ex].
- [30] R. D. Ball *et al.*, Nucl. Phys. B **855** (2012) 608, arXiv:1108.1758 [hep-ph].
- [31] B.L. Combridge, J. Kripfganz, and J. Ranft, Phys. Lett. **B70** (1977) 234.
- [32] Z. Nagy, Phys. Rev. Lett. **88** (2002) 122003, arXiv:hep-ph/0110315 [hep-ph].
- [33] Z. Nagy, Phys. Rev. D **68** (2003) 094002, arXiv:hep-ph/0307268 [hep-ph].
- [34] R. Hamberg, W. L. van Neerven, and T. Matsuura, Nucl. Phys. **B359** (1991) 343.
- [35] W. L. van Neerven and E. B. Zijlstra, Nucl. Phys. **B382** (1992) 11.
- [36] P. J. Rijken and W. L. van Neerven, Phys. Rev. **D51** (1995) 44, arXiv:hep-ph/9408366.
- [37] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, Phys. Rev. **D69** (2004) 094008, arXiv:hep-ph/0312266.
- [38] S. Dittmaier and M. Krämer, Phys. Rev. **D65** (2002) 073007, arXiv:hep-ph/0109062 [hep-ph].
- [39] S. Dittmaier and M. Huber, JHEP **01** (2010) 060, arXiv:0911.2329 [hep-ph].
- [40] M. Cacciari, G. Salam, and G. Soyez, JHEP **04** (2008) 063, arXiv:0802.1189.

- [41] The ATLAS Collaboration, *Calorimeter Clustering Algorithms: Description and Performance*, ATLAS-LARG-PUB-2008-002.
- [42] The ATLAS Collaboration, *Expected Performance of the ATLAS Experiment – Detector, Trigger and Physics*, arXiv:0901.0512.
- [43] The CMS Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus and MET*, CMS-PAS-PFT-09-001.
- [44] The CMS Collaboration, *Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector*, CMS-PAS-PFT-10-001.
- [45] The ATLAS Collaboration, Phys. Rev. D **86** (2012) 014022, arXiv:1112.6297.
- [46] The CMS Collaboration, Phys. Rev. Lett. **107** (2011) 132001, arXiv:1106.0208.
- [47] The CMS Collaboration, Phys. Rev. Lett. **B700** (2011) 187, arXiv:1104.1693.
- [48] The CMS Collaboration, Phys. Rev. Lett. **106** (2011) 201804, arXiv:1102.2020.
- [49] The CMS Collaboration, Phys. Rev. Lett. **106** (2011) 122003, arXiv:1101.5029.
- [50] The ATLAS Collaboration, Eur. Phys. J. **C71** (2011) 1763, arXiv:1107.2092 [hep-ex].
- [51] The CMS Collaboration, Phys. Lett. **B702** (2011) 336, arXiv:1106.0647 [hep-ex].
- [52] The CDF Collaboration, A. Abulencia *et al.*, J. Phys. **G34** (2007) 2457.
- [53] The ATLAS Collaboration, JHEP **1012** (2010) 060, arXiv:1010.2130 [hep-ex].
- [54] The CMS Collaboration, JHEP **1101** (2011) 080, arXiv:1012.2466 [hep-ex].
- [55] The CMS Collaboration, JHEP **1110** (2011) 132, arXiv:1107.4789 [hep-ex].
- [56] The ATLAS Collaboration, Phys. Rev. D **85** (2012) 072004, arXiv:1109.5141 [hep-ex].
- [57] The UA1 Collaboration, C. Albajar *et al.*, Phys. Lett. **B198** (1987) 271.
- [58] The UA2 Collaboration, J. Alitti *et al.*, Phys. Lett. **B276** (1992) 365.
- [59] The DØ Collaboration, *Measurement of the Cross Section for W and Z Production to Electron Final States with the DØ Detector at $\sqrt{s} = 1.96$ TeV*, 4403-CONF; *Measurement of the Cross-section for Inclusive W Production in the Muon Channel at $\sqrt{s} = 1.96$ TeV using the DØ Detector*, 4750-CONF.
- [60] PHENIX Collaboration, A. Adare *et al.*, Phys. Rev. Lett. **106** (2011) 062001, arXiv:1009.0505 [hep-ex].
- [61] The ATLAS Collaboration, Phys. Rev. **D84** (2011) 112006, arXiv:1108.2016 [hep-ex].
- [62] The ATLAS Collaboration, Phys. Lett. **B706** (2012) 276–294, arXiv:1108.4101 [hep-ex].
- [63] The CMS Collaboration, JHEP **1108** (2011) 117, arXiv:1104.1617 [hep-ex].
- [64] The ATLAS Collaboration, Phys. Rev. Lett. **109** (2012) 012001, arXiv:1203.4051 [hep-ex].
- [65] The CMS Collaboration, JHEP **01** (2012) 010, arXiv:1110.3226 [hep-ex].
- [66] The CMS Collaboration, Phys. Rev. **D 85** (2012) 032002, arXiv:1110.4973 [hep-ex].
- [67] The ATLAS Collaboration, Phys. Rev. **D85** (2012) 032009, arXiv:1111.2690 [hep-ex].
- [68] The ATLAS Collaboration, Phys. Rev. D **85** (2012) 092002, arXiv:1201.1276 [hep-ex].
- [69] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, JHEP **0307** (2003) 001, arXiv:hep-ph/0206293 [hep-ph].
- [70] T. Gleisberg, S. Hoeche, F. Krauss, M. Schönherr, S. Schumann *et al.*, JHEP **0902** (2009) 007, arXiv:0811.4622 [hep-ph].
- [71] J. M. Campbell and R. K. Ellis, Phys. Rev. **D60** (1999) 113006, arXiv:hep-ph/9905386.
- [72] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, T. Gleisberg, H. Ita, D. A. Kosower, and D. Maître, Phys. Rev. Lett. **106** (2011) 092001. [arXiv:1009.2338 [hep-ph]].
- [73] The CMS Collaboration, JHEP **1107** (2011) 049, arXiv:1105.5661 [hep-ex].
- [74] The CMS Collaboration, Phys. Rev. **D84** (2011) 092004, arXiv:1108.3773 [hep-ex].

- [75] The ATLAS Collaboration, Phys. Lett. **B707** (2012) 459–477, arXiv:1108.3699 [hep-ex].
- [76] The ATLAS Collaboration, Phys. Lett. **B 711** (2012) 244, arXiv:1201.1889 [hep-ex].
- [77] The ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross section in the final state with a hadronically decaying tau lepton and jets using the ATLAS detector*, ATLAS-CONF-2012-032.
- [78] S. Moch and P. Uwer, Phys. Rev. **D78** (2008) 034003, arXiv:0804.1476 [hep-ph].
- [79] U. Langenfeld, S. Moch, and P. Uwer, *New results for $t\bar{t}$ production at hadron colliders*, arXiv:0907.2527 [hep-ph].
- [80] The ATLAS Collaboration, *Measurement of the t-channel single top-quark production cross section in 0.70/fb of pp collisions at $\sqrt{s}=7$ TeV*, ATLAS-CONF-2011-101.
- [81] The CMS Collaboration, Phys. Rev. Lett. **107** (2011) 091802, arXiv:1106.3052 [hep-ex].
- [82] The ATLAS Collaboration, *Measurement of the top quark mass from 2011 ATLAS data using the template method*, ATLAS-CONF-2011-120.
- [83] The ATLAS Collaboration, *Determination of the Top Quark Mass with a Template Method in the All Hadronic Decay Channel using 2.04 fb⁻¹ of ATLAS data*, ATLAS-CONF-2012-030.
- [84] The CMS Collaboration, *Measurement of the top mass by using the muon+jets channel*, CMS-PAS-TOP-11-015.
- [85] The CMS Collaboration, *Measurement of the top quark mass in the dilepton channel in pp collisions at $\sqrt{s}=7$ TeV*, CMS-PAS-TOP-11-016.
- [86] J. Ellison and J. Wudka, Ann. Rev. Nucl. Part. Sci. **48** (1998) 33, arXiv:hep-ph/9804322 [hep-ph].
- [87] U. Baur and D. Rainwater, Phys. Rev. **D62** (2000) 113011.
- [88] The CMS Collaboration, Phys. Lett. **B701** (2011) 535, arXiv:1105.2758 [hep-ex].
- [89] The ATLAS Collaboration, JHEP **1109** (2011) 072, arXiv:1106.1592 [hep-ex].
- [90] The ATLAS Collaboration, Phys. Rev. Lett. **107** (2011) 041802, arXiv:1104.5225 [hep-ex].
- [91] The CMS Collaboration, *Measurement of the WW, WZ and ZZ cross sections at CMS*, CMS-PAS-EWK-11-010.
- [92] The ATLAS Collaboration, *Measurement of the W^+W^- production cross section in proton-proton collisions with the ATLAS Detector*, ATLAS-CONF-2012-025.
- [93] The ATLAS Collaboration, Phys. Lett. **B709** (2012) 341, arXiv:1111.5570 [hep-ex].
- [94] The ATLAS Collaboration, Phys. Rev. Lett. **108** (2012) 041804, arXiv:1110.5016 [hep-ex].
- [95] LEP Electroweak Group, SLD Electroweak Group, Updated numbers March 2012, <http://lepwwg.web.cern.ch/LEPEWWG>.
- [96] B. W. Lee, C. Quigg, and H. B. Thacker, Phys. Rev. Lett. **38** (1977) 883.
M. Quiros, *Constraints on the Higgs boson properties from the effective potential*, arXiv:hep-ph/9703412 [hep-ph], in *Perspectives on Higgs Physics II*, Ed. G. L. Kane (World Scientific, Singapore).
A. Ghinculov and T. Binoth, Acta Phys. Polon. **B30** (1999) 99, arXiv:hep-ph/9807227 [hep-ph].
- [97] LEP Working Group for Higgs boson searches [ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration], Phys. Lett. **B565** (2003) 61, arXiv:hep-ex/0306033 [hep-ex].
- [98] The Tevatron New Phenomina and Higgs Working Group [CDF and DØ Collaborations], *Combined CDF and DØ Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb⁻¹ of Data*, arXiv:1107.5518 [hep-ex].
- [99] The ATLAS Collaboration, *Update of the Combination of Higgs Boson Searches in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Experiment at the LHC*, ATLAS-CONF-2011-135.

- [100] The CMS Collaboration, *Combination of Higgs Searches*, CMS-PAS-HIG-11-022.
- [101] E. Gross, O. Vitells, Eur. Phys. J. **C70** (2010) 525.
- [102] A. Djouadi, M. Spira, P. Zerwas, Phys. Lett. **B264**, 440 (1991); S. Dawson, Nucl. Phys. **B359**, 283 (1991); M. Spira, A. Djouadi, D. Graudenz, P. Zerwas, Nucl. Phys. **B453**, 17 (1995); R. Harlander, W. Kilgore, Phys. Rev. Lett. **88**, 201801 (2002); C. Anastasiou, K. Melnikov, Nucl. Phys. **B646**, 220 (2002); V. Ravindran, J. Smith, W. van Neerven, Nucl. Phys. **B665**, 325 (2003).
- [103] S. Catani, D. de Florian, M. Grazzini, P. Nason, JHEP **07** (2003) 028.
- [104] U. Aglietti, R. Bonciani, G. Degrossi, A. Vicini, Phys. Lett. **B595** (2004) 432; S. Actis, G. Passarino, C. Sturm, S. Uccirati, Phys. Lett. **B670** (2008) 12.
- [105] M. Ciccolini, A. Denner, S. Dittmaier, Phys. Rev. Lett. **99**, 161803 (2007); M. Ciccolini, A. Denner, S. Dittmaier, Phys. Rev. **D77**, 013002 (2008); K. Arnold *et al.*, Comput. Phys. Commun. **180**, 1661 (2009).
- [106] P. Bolzoni, F. Maltoni, S. Moch, M. Zaro, Phys. Rev. Lett. **105** (2010) 011801.
- [107] T. Han, S. Willenbrock, Phys. Lett. **B273** (1991) 167.
- [108] O. Brein, A. Djouadi, R. Harlander, Phys. Lett. **B579** (2004) 149.
- [109] M. L. Ciccolini, S. Dittmaier, M. Kramer, Phys. Rev. **D68** (2003) 073003.
- [110] W. Beenakker *et al.*, Phys. Rev. Lett. **87**, 201805 (2001); W. Beenakker *et al.*, Nucl. Phys. **B653**, 151 (2003); S. Dawson, L. H. Orr, L. Reina, D. Wackeroth, Phys. Rev. **D67**, 071503 (2003); S. Dawson, C. Jackson, L. Orr, L. Reina, D. Wackeroth, Phys. Rev. **D68**, 034022 (2003).
- [111] LHC Higgs Cross Section Working Group, Eds. S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka, CERN-2011-002 (2011) arXiv:1101.0593 and arXiv:1201.3084.
- [112] A. Djouadi, J. Kalinowski, and M. Spira, Comput. Phys. Commun. **108** (1998) 56, arXiv:hep-ph/9704448 [hep-ph].
- [113] A. Djouadi, M. Spira, and P. Zerwas, Z. Phys. **C70** (1996) 427, arXiv:hep-ph/9511344 [hep-ph].
- [114] A. Ghinculov, T. Binoth, and J. van der Bij, Phys. Lett. **B427** (1998) 343, arXiv:hep-ph/9802367 [hep-ph].
- [115] The ATLAS Collaboration, *Detector and physics performance technical design report*, CERN/LHCC/99-15 (1999);
The CMS Collaboration, J. Phys. **G34** (2007) 995.
The ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics, Volume I: Performance*, CERN-OPEN-2008-20.
- [116] D.L. Rainwater, D. Zeppenfeld, JHEP **12** (1997) 5; D.L. Rainwater, D. Zeppenfeld, K. Hagiwara, Phys. Rev. **D59** (1999) 014037; T. Plehn, D.L. Rainwater, D. Zeppenfeld, Phys. Rev. **D61** (2000) 093005; D.L. Rainwater, D. Zeppenfeld, Phys. Rev. **D60** (1999) 113004 [Erratum Phys. Rev. **D61** (2000) 099901]; N. Kauer, T. Plehn, D.L. Rainwater, D. Zeppenfeld, Phys. Lett. **B503** (2001) 113.
- [117] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, Phys. Rev. Lett. **100** (2008) 242001, arXiv:0802.2470 [hep-ph].
- [118] The ATLAS Collaboration, *ATLAS Sensitivity to the Standard Model Higgs in the HW and HZ Channels at High Transverse Momenta*, ATL-PHYS-PUB-2009-088 (2009)
- [119] The ATLAS and CMS Collaborations, LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in summer 2011*, ATL-PHYS-PUB-2011-011 (2011).
- [120] A. L. Read, J. Phys. **G28** (2002) 2693.
- [121] G. Cowan, K. Cranmer, E. Gross, O. Vitells, Eur. Phys. J. **C71** (2011) 1554.
- [122] The ATLAS Collaboration, Phys. Lett. **B 710** (2012) 49, arXiv:1202.1408 [hep-ex].
- [123] The ATLAS Collaboration, Phys. Rev. Lett. **108** (2012) 111803, arXiv:1202.1414 [hep-ex].

- [124] The CMS Collaboration, Phys. Lett. B **710** (2012) 403, arXiv:1202.1487 [hep-ex].
- [125] The ATLAS Collaboration, Phys. Lett. B **710** (2012) 383, arXiv:1202.1415 [hep-ex].
- [126] The CMS Collaboration, Phys. Rev. Lett. **108** (2012) 111804 arXiv:1202.1997 [hep-ex].
- [127] The ATLAS Collaboration, *Search for the Standard Model Higgs boson in the $H \rightarrow WW^{(*)} \rightarrow \ell\nu\ell\nu$ decay mode with 4.7 fb^{-1} of ATLAS data at $\sqrt{s} = 7\text{ TeV}$* , ATLAS-CONF-2012-012.
- [128] The CMS Collaboration, *Search for the Standard Model Higgs boson decaying to a W pair in the fully leptonic final state in pp collisions at $\sqrt{s} = 7\text{ TeV}$* , CERN-PH-EP-2012-018 (2012).
- [129] M. Dittmar and H. Dreiner, Phys. Rev. **D55** (1997) 167.
- [130] The ATLAS Collaboration, *Search for the Standard Model Higgs boson in the $H \rightarrow \tau^+\tau^-$ decay mode with 4.7 fb^{-1} of ATLAS data at 7 TeV* , ATLAS-CONF-2012-014 (2012).
- [131] The CMS Collaboration, Phys. Lett. B **713** (2012) 68 arXiv:1202.4083 [hep-ex].
- [132] The ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a b -quark pair using up to 4.7 fb^{-1} of pp collision data at $\sqrt{s} = 7\text{ TeV}$ with the ATLAS detector at the LHC*, ATLAS-CONF-2012-015.
- [133] The CMS Collaboration, Phys. Lett. **B710** (2012) 284.
- [134] The ATLAS Collaboration, *An update to the combined search for the Standard Model Higgs boson with the ATLAS detector at the LHC using up to 4.9 fb^{-1} of pp collision data at $\sqrt{s} = 7\text{ TeV}$* , ATLAS-CONF-2012-019.
- [135] The CMS Collaboration, Phys. Lett. B **710** (2012) 26, arXiv:1202.1488 [hep-ex].
- [136] The CMS Collaboration, JHEP **1203** (2012) 040, arXiv:1202.347 [hep-ex].
- [137] The ATLAS Collaboration, *Search for a Standard Model Higgs in the $H \rightarrow ZZ \rightarrow \ell^+\ell^-\nu\bar{\nu}$ decay channel with 4.7 fb^{-1} with the ATLAS detector*, ATLAS-CONF-2012-016.
- [138] The ATLAS Collaboration, *Search for a Standard Model Higgs in the mass range $200\text{--}600\text{ GeV}$ in the channel $H \rightarrow ZZ \rightarrow \ell\ell qq$ with the ATLAS detector*, ATLAS-CONF-2012-017.
- [139] The CMS Collaboration, JHEP **1204** (2012) 036, arXiv:1202.141 [hep-ex].
- [140] The CMS Collaboration, JHEP **1203** (2012) 081 arXiv:1202.3617 [hep-ex].
- [141] The ATLAS Collaboration, *Search for the Higgs boson in the $H \rightarrow WW \rightarrow \ell\nu jj$ decay channel using 4.7 fb^{-1} of pp collisions at $\sqrt{s} = 7\text{ TeV}$ with the ATLAS detector*, ATLAS-CONF-2012-018.
- [142] The ATLAS and the CMS Collaborations, *Combined Standard Model Higgs boson searches with up to 2.3 fb^{-1} of pp collisions at $\sqrt{s}=7\text{ TeV}$ at the LHC*, ATLAS-CONF-2011-157.
- [143] For a review, see: S. P. Martin, *A Supersymmetry primer*, arXiv:hep-ph/9709356 [hep-ph].
- [144] H. Nilles, Phys. Rep. **110** (1984) 1, and references therein.
- [145] P. de Jong, *SUSY Searches at ATLAS*, arXiv:1201.4548 [hep-ex].
- [146] P. Fayet, Phys. Lett. **B69** (1977) 489;
G. R. Farrar and P. Fayet, Phys. Lett. **B76** (1978) 575.
- [147] The ATLAS Collaboration, Phys. Lett. B **710** (2012) 67, arXiv:1109.6572 [hep-ex].
- [148] The CMS Collaboration, Phys. Rev. Lett. **107** (2011) 221804, arXiv:1109.2352 [hep-ex].
- [149] The ATLAS Collaboration, Phys. Rev. **D85** (2012) 012006, arXiv:1109.6606 [hep-ex],
- [150] The CMS Collaboration, *Search for new physics with single-leptons at the LHC*, CMS-PAS-SUS-11-015.
- [151] The ATLAS Collaboration, Phys. Lett. **B709** (2012) 137, arXiv:1110.6189 [hep-ex].
- [152] The CMS Collaboration, *Multileptonic SUSY searches*, CMS-PAS-SUS-11-013.
- [153] The Atlas Collaboration, JHEP **1111** (2011) 099, arXiv:1110.2299 [hep-ex].
- [154] A. H. Chamseddine *et al.*, Phys. Rev. Lett. **49** (1982) 970; R. Barbieri *et al.*, Phys. Lett. **B119**

- (1982) 343; L. E. Ibanez, Phys. Lett. **B118** (1982) 73; L. J. Hall *et al.*, Phys. Rev. **D27** (1983) 2359; N. Ohta, Prog. Theor. Phys. **70** (1983) 542; G. L. Kane *et al.*, Phys. Rev. **D49** (1994) 6173.
- [155] The ATLAS Collaboration, *Search for supersymmetry with jets and missing transverse momentum: Additional model interpretations*, ATLAS-CONF-2011-155.
- [156] R. Barbieri, *SUSY phenomenology*, talk at the Hadron Collider Symposium 2011, Paris.
- [157] The ATLAS Collaboration, Phys. Rev. Lett. **108** (2012) 181802, arXiv:1112.3832 [hep-ex].
- [158] G. Polesello and D. R. Tovey, JHEP **1003** (2010) 030, arXiv:0910.0174 [hep-ph].
- [159] The ATLAS Collaboration, *Search for supersymmetry in pp collisions at $\sqrt{s} = 7$ TeV in final states with missing transverse momentum, b -jets and one lepton with the ATLAS detector*, ATLAS-CONF-2011-130.
- [160] The ATLAS Collaboration, Phys. Rev. Lett. **108** (2012) 041805, arXiv:1109.4725 [hep-ex].
- [161] The CDF Collaboration, Phys. Rev. Lett. **98** (2007) 131804.
- [162] The DØ Collaboration, V. Abazov *et al.*, Phys. Lett. **B680** (2009) 34.
- [163] The CMS Collaboration, S. Chatrchyan *et al.*, Phys. Lett. **B704** (2011) 411, arXiv:1106.0933 [hep-ex].
- [164] P. Langacker, Rev. Mod. Phys. **81** (2009) 1199.
- [165] J. Erler, P. Langacker, S. Munir, and E. Rojas, JHEP **0908** (2009) 017, arXiv:0906.2435 [hep-ph].
- [166] D. London and J. L. Rosner, Phys. Rev. **D34** (1986) 1530.
- [167] K. Lane and E. Eichten, Phys. Lett. **B222** (1989) 274.
- [168] K. Lane and S. Mrenna, Phys. Rev. **D67** (2003) 115011.
- [169] A. Belyaev, R. Foadi, M. T. Frandsen, M. Järvinen, F. Sannino, and A. Pukhov, Phys. Rev. **D79** (2009) 035006.
- [170] L. Randall and R. Sundrum, Phys. Rev. Lett. **83** (1999) 3370.
- [171] J.C. Pati and A. Salam, Phys. Rev. **D10** (1974) 275; R. Mohapatra and J.C. Pati, Phys. Rev. **D11** (1975) 2558; G. Senjanovic and R.N. Mohapatra, Phys. Rev. **D12** (1975) 1502.
- [172] The ATLAS Collaboration, Phys. Rev. Lett. **107** (2011) 272002, arXiv:1108.1582 [hep-ex].
- [173] The CMS Collaboration, *Search for W' in the leptonic channels in pp Collisions at $\sqrt{s} = 7$ TeV*, CMS-PAS-EXO-11-024.
- [174] The CMS Collaboration, *Search for Resonances in the Dilepton Mass Distribution in pp Collisions at $\sqrt{s} = 7$ TeV*, CMS-PAS-EXO-11-019.
- [175] The ATLAS Collaboration, Phys. Lett. **B705** (2011) 28, arXiv:1108.1316 [hep-ex].
- [176] The DØ Collaboration, Phys. Lett. **B695** (2011) 88.
The CDF Collaboration, Phys. Rev. Lett. **106** (2011) 121801.
- [177] The ATLAS Collaboration, Phys. Lett. **B700** (2011) 163, arXiv:1103.6218 [hep-ex].
The CMS Collaboration, JHEP **1105** (2011) 093, arXiv:1103.0981 [hep-ex].
- [178] The CMS Collaboration, *Search for W' (or techni- ρ) to WZ* , CMS-PAS-EXO-11-041.
- [179] The ATLAS Collaboration, New J. Phys. **13** (2011) 053044, arXiv:1103.3864 [hep-ex].
- [180] The CMS Collaboration, Phys. Rev. Lett. **106** (2011) 201804, arXiv:1102.2020 [hep-ex].
- [181] The ATLAS Collaboration, Phys. Lett. **B708** (2012) 37–54, arXiv:1108.6311 [hep-ex].
- [182] E. J. Eichten, K. D. Lane, and M. E. Peskin, Phys. Rev. Lett. **50** (1983) 811.
- [183] E. Eichten, I. Hinchliffe, K. Lane, and C. Quigg, Rev. Mod. Phys. **56** (1984) 579.
- [184] The CMS Collaboration, JHEP **1205** (2012) 055, arXiv:1202.5535 [hep-ex].
- [185] The CMS Collaboration, Phys. Lett. **B704** (2011) 123, arXiv:1107.4771 [hep-ex].
- [186] U. Baur, I. Hinchliffe, and D. Zeppenfeld, Int. J. Mod. Phys. **A2** (1987) 1285.

- [187] U. Baur, M. Spira, and P. M. Zerwas, Phys. Rev. **D42** (1990) 815.
- [188] P. H. Frampton and S. L. Glashow, Phys. Lett. **B190** (1987) 157.
- [189] P. H. Frampton and S. L. Glashow, Phys. Rev. Lett. **58** (1987) 2168.
- [190] J. Bagger, C. Schmidt, and S. King, Phys. Rev. **D37** (1988) 1188.
- [191] T. Han, I. Lewis, and Z. Liu, JHEP **1012** (2010) 085, [arXiv:1010.4309](#) [hep-ph].
- [192] J. L. Hewett and T. G. Rizzo, Phys. Rept. **183** (1989) 193.

Organizing Committee

T. Donskova (JINR, Dubna, Russia)
N. Ellis (CERN, Geneva, Switzerland)
C. Grojean (CERN, Geneva, Switzerland)
H. Haller (CERN, Geneva, Switzerland)
A. Olchevsky (JINR, Dubna, Russia)
M. Spiropulu (CERN, Geneva, Switzerland)

Local Organizing Committee

V. Almasan (ITIM Cluj-Napoca, Romania)
G. Cata-Danil (Politehnica Univ. Bucharest, Romania)
C.B Cizmas (Transilvania Univ. Brasov, Romania)
M. Cuciuc (IFIN-HH, Romania)
A. Dorobantu (IFIN-HH, Romania)
A. Jipa (Univ. of Bucharest, Romania)
R. Muresan (IFIN-HH, Romania)
B. Popovici (IFIN-HH, Romania)
L. Serban (IFIN-HH, Romania)
G. Stoicea (IFIN-HH, Romania)
E. Teodorescu (IFIN-HH, Romania)
I. Ursu (chair) (IFIN-HH, Romania)

International Advisors

A. Skrinsky (BINP Novosibirsk, Russia)
N. Tyurin (IHEP Protvino, Russia)

Lecturers

K. Cranmer (New York University, USA)
B. Dobrescu (Fermilab, USA)
E. Dudas (Ecole Polytechnique, France)
Y. Grossman (Cornell University, USA)
E. Iancu (CEA-Saclay, France)
K. Jacobs (Freiburg University, Germany)
B. Kayser (Fermilab, USA)
F. Maltoni (Université Catholique de Louvain-la-Neuve, Belgium)
V. Rubakov (INR RAS, Russia)

Discussion Leaders

C. Alexa (IFIN-HH, Romania)
A. Bednyakov (JINR, Russia)
C. Biggio (IFAE, Spain)
D. Gorbunov (INR RAS, Russia)
D. Ghilencea (IFIN-HH, Romania)
G. Soyez (CEA-Saclay, France)

Students

Marco AGUSTONI
Lee ALLISON
Valerio ALTINI
Susana AMOR SANTOS
Philipp ANGER
Lucia BATKOVA
Kathrin BECKER
Moritz BECKMANN
Eleonora BENHAR NOCCIOLI
Agni BETHANI
Alexey BOLDYREV
Nicolas BOUSSON
Michal BROZ
Ion BURDUCEA
Robert CANTRILL
Roberta CARDINALE
Julien CHASSERAT
Robert Andrew CURRIE
Yuriy DAVYGORA
Antonia DI CRESCENZO
Thi Kieu Oanh DOAN
James William DOLEN
Otilia DUCU
Jan EBR
Andrea FERRETTO PARODI
Grzegorz GACH
Valentina GALLO
Paolo GANDINI
Mikhail GAVRILENKO
nicolin GOVENDER
Marco GRASSI
Sam GREGSON
Andrey GRINYUK
Zoe HATHERELL
Sarah HEIM
Hauke HELD
Dennis HELLMICH
Otto HINDRICHS
Annika HOLLNAGEL
Valerio IPPOLITO
Adam JINARU
Daniel JOHNSON
Tatjana JOVIN
Christian JUNG
Janos KARANCSI
Shahla KHALILOVA
Palash KHAN
Benjamin KLEIN
Robert Jan KNEGJENS
Alexandr KOZLINSKIY

Nadiia KRUPINA
Christopher LESTER
Karina MARTIN
Stewart MARTIN-HAUGH
Nicolas MERIC
Ljiljana MORVAJ
Elena OLIVER GARCIA
George - Octavian PADURARIU
Christopher John PARKINSON
Remus-Andrei PAUN
Michele PINAMONTI
Alexander PINDER
Mihail-Ioan POP
Kilian ROSBACH
Anthony ROSE
Valerio ROSSETTI
Marina ROTARU
Nick RYDER
Luka SANTELJ
Jean-Baptiste SAUVAN
Steffen SCHAEPE
Christian SCHROEDER
Thomas SCHWINDT
Sercan SEN
Daniel SHORT
Sascha STAHL
Jonathan STAHLMAN
Simona Maria STELLACCI
Mihai STRATICIUC
Cecilia TACCINI
Nancy TANNOURY
Andres TIKO
Maciej TRZEBINSKI
Alexandra TUDORACHE
Valentina TUDORACHE
Ruggero TURRA
Michal VAJZER
Lizardo VALENCIA PALOMO
Marco VANADIA
Jakub VICHA
Shannon WALCH
Sinead WALSH
Tan WANG
Hendrik WEBER
Jordan WEBSTER
Liv WIJK
Martin WILDT
Ivan YELETSKIKH
Andy YEN

Posters

Author	Poster title
K. BECKER	t-channel single top quark cross section measurement with the ATLAS detector
M. BECKMANN & J. LIST	ILC: spins running the Gauntlet
A. BOLDYREV	Di-muon trigger efficiency studies for rare decay $B_s^0 \rightarrow \mu^+ \mu^-$
I. BOLDYREV & M. STRATICIUC	Strangeness production at LHCb
A. DI CRESCENZO	Search for short-lived particle decays in the OPERA experiment
T.K.O. DOAN	The transverse momentum measurement of Z/γ^* production in the ATLAS experiment in 2011
J. EBR	Some attempts to improve the agreement between air shower simulations and data
G. GACH	Exclusive diffractive di-jet production in $e - p$ collisions
P. GANDINI	The hunt for $b \rightarrow u$ transitions in $B^\pm \rightarrow DK^\pm$ decays
S. HEIM	Search for high-mass dielectron resonances at ATLAS
D. HELLMICH	Measuring the cross section of the Z boson with two additional b -jets in ATLAS
A. HOLLNAGEL	The OPERA experiment
D. JOHNSON	Study of the decays $D^0 \rightarrow K_s^0 K^\pm \pi^\mp$ at CLEO
R.J. KNEGJENS	In pursuit of new physics with B_s mesons
L. ALLISON	Study of $\psi(2S)$, focusing on $\psi(2S) \rightarrow J/\psi(\mu^+ \mu^-) \pi^+ \pi^-$
L. VALENCIA PALOMO	J/ψ measurement in Pb-Pb collisions in the ALICE experiment at the LHC
K. MARTIN	Particle identification in the SND detector at VEPP-2000 collider
N. MERIC	A dive into electroweak and quantum chromodynamics processes
M. PINAMONTI	$t\bar{t}$ production cross section measurement at ATLAS
L. MORVAJ	Searching for the small mass-splitting supersymmetry
V. ROSSETTI	Search for new phenomena in events with a monojet and large missing transverse momentum with the ATLAS detector

Author	Poster title
M. ROTARU	General search for new phenomena in ep and pp collisions
J.B. SAUVAN	Probing the jet energy measurement using Z +jet events in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector
S. STAHL	CP violation in B_s^0 mixing
R. TURRA	Measurements of isolated prompt photons in pp collisions with the ATLAS detector
J. VICHA	Energy calibration at the Pierre Auger observatory with respect to mass composition of cosmic rays
S. WALCH	Search for the Standard Model Higgs boson in the $H \rightarrow WW \rightarrow \ell\nu\ell\nu$ channel with ATLAS
I. YELETSKIKH & I. BOYKO & V.A. BEDNYAKOV & M. DEMICHEV & M. CHIZHOV	Search for Z^* in dimuon channel in ATLAS 2011 data