Search for narrow resonances in dijet final states at the LHC with $\sqrt{s}=$ 13 TeV

Facoltà di Scienze Matematiche Fisiche e Naturali

Dottorato di Ricerca in Fisica – XXVIII Ciclo

Candidate

Giulia D'Imperio
ID number 1216998

Thesis Advisor

Prof. Shahram Rahatlou

Co-Advisor

Dott. Francesco Santanastasio

Thesis defended on January 22, 2016
in front of a Board of Examiners composed by:

Luciano Maria Barone (Università La Sapienza)
Cinzia Da Già (University of Manchester UK)
Teresa Rodrigo (Universidad de Cantabria ES)

This thesis has been typeset by LaTeX and the Sapthesis class.

Version: January 2016

Author's email: giulia.dimperio@cern.ch

There is nothing like looking, if you want to find something. You certainly usually find something, if you look, but it is not always quite the something you were after.

J.R.R. Tolkien, The Hobbit

Contents

Introduction

In this thesis the very first search of new physics beyond the Standard Model at the Large Hadron Collider (LHC) using data from proton-proton collisions at center-of-mass energy of 13 TeV with the CMS experiment is presented.

The Standard Model, that describes the fundamental particles and their interactions, is a very solid and elegant theory, verified in many independent experiments. However there are still many reasons to believe it is incomplete. To name some: the theory does not incorporate the general relativity; it does not explain the asymmetry of matter and antimatter in the universe; cosmological observations suggest the presence of a dark matter that so far we haven't detected because it interacts weakly with the ordinary matter.

There are several models of new physics that predict the existence of resonances at the TeV scale, and therefore accessible at the proton-proton collisions of the LHC. The object of the search presented in this dissertation are narrow resonances decaying to a pair of quarks and/or gluons, that in the detector will appear as a pair of back-to-back hadronic jets in the transverse plane. If such a resonance exists, the experimental signature will be a bump corresponding to the mass of the resonance over the smoothly falling dijet mass distribution of the QCD processes.

The search is model independent: it does not assume any theoretical model but is sensitive to any resonance that decays to a pair of jets, and the analysis technique is simple and robust. Yet, this is an extremely powerful and promising search in this moment, since the LHC Run 2 has just restarted in Summer 2015, after a long shut down period from the end of 2012, and the new center-of-mass energy of collisions of 13 TeV is almost twice with respect to the past. We will see that such an increase in energy corresponds to much larger cross sections, especially for high mass resonances, and this makes the search competitive with Run 1 results even with a small dataset.

The results presented here make use of 2.4 fb^{-1} of data collected by the CMS experiment in 2015 and these data have already exceeded the sensitivity of the 2012 dijet search for a large part of the spectrum to which this search is sensitive.

Chapter 1 presents a brief introduction of the strong sector of the Standard Model, recalling the elements of Quantum Chromodynamics and its implications for the experimental measurements of hadronic interactions. Also elements of the physics of proton-proton interaction at hadron colliders are introduced.

As we mentioned above, there are several models that predict dijet resonances in the mass region accessible for LHC collisions. The CMS dijet analysis that we are going to present in this thesis compares its results to the expectations of some theoretical models, that are described in Chapt. 2. The chapter presents also a

review of the past dijet resonance searches, focusing on the most recent results at the LHC experiments, and concludes with a quantitative estimate of the gain in sensitivity due to the increase in collision energy from 8 to 13 TeV, that makes this analysis very powerful at the beginning of the LHC Run 2.

Chapter 3 is dedicated to the description of the experimental apparatus: the Large Hadron Collider is introduced, and the Compact Muon Solenoid experiment is described in its various subdetectors.

The main actors of the dijet analysis are hadronic jets. Jets are composite objects and therefore not defined in a unique way: the reconstruction and the procedure to calibrate the jet energy are described in Chapt. 4.

The central subject of this thesis is the dijet resonance search in proton-proton collisions at $\sqrt{s} = 13$ TeV, and the description of the analysis is divided in the last three chapters. Data and simulation samples, selection criteria and quality checks are introduced in Chapt. 5. We proceed describing the signal models and the fit to data in order to estimate the background in Chapt. 6. The fit represents basically a test of smoothness of the observed dijet mass spectrum, and it is the core of the analysis. The studies for the parameterization choice and the fit quality checks, described in detail in this chapter, represent my main contribution to the analysis, together with the development of the analysis framework and the collaboration to the jet energy calibration.

The results and their statistical interpretation are finally presented in Chapt. 7, concluding with a paragraph on future analysis perspectives.

Chapter 1

Elements of Standard Model and collider physics

This first chapter aims to introduce the physics of known Standard Model processes at a hadron collider, that represent the environment where the search for a new resonance is performed.

Section 1.1 introduces the theory and in particular the quantum chromodynamics (QCD) sector of the Standard Model.

The following Section 1.2 presents the effects of the QCD theory in hadron collisions, describing what happens in a proton-proton interaction, the formation of jets, the kinematics of the two-parton scattering, and some useful notions as hadronic and partonic cross sections, parton luminosity factor and luminosity of a collider.

1.1 Standard Model and quantum chromodynamics

The Standard Model (SM) is the theory that describes the fundamental particles and includes in one coherent frame three of the four fundamental interactions in nature: electromagnetic, weak and strong forces. The SM is a quantum field theory, renormalizable and coherent with the Special Relativity. The basis of the theory were posed in the '60 by Glashow, Weinberg and Salam [1–3].

The Lagrangian of the SM is symmetric under the group SU(3) × SU(2) × U(1).

It's possible to divide the model in two sectors: electroweak (EW)and strong (Quantum Chromodynamics, QCD) and express the Lagrangian as the sum of two contributes:

$$\mathcal{L}_{SM} = \mathcal{L}_{EW} + \mathcal{L}_{QCD} \tag{1.1}$$

For the analysis presented in this thesis, the relevant part is only the QCD sector, therefore in this chapter the electroweak part of the model is not described. Some elements of the QCD theory (for an exhaustive discussion see for example [4]) and of hadronic physics at collider [5] are presented to introduce the reader to the subject of this work.

Quantum Chromodynamics is the gauge field theory of the strong interaction between particles that carry the color degree of freedom. The underlying symmetry

group is the $SU(3)_C$, which makes QCD a non-Abelian theory. The profound implication of this property of QCD is that the gauge mediators (gluons) are colored and thus self interacting. The QCD Lagrangian is written as:

$$\mathcal{L}_{QCD} = \sum_i \bar{q}_{i,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s\gamma^\mu t_{ab}^A G_\mu^A - m_i\delta_{ab})q_{i,b} - \frac{1}{4}F_{\mu\nu}^A F^{\mu\nu,A} \qquad (1.2)$$

where $q_{i,b}$ represents the quark spinor of flavor i and color $a = 1 \to 3$, $G_{\mu\nu}^A$ is the gluon field associated with the generator t_{ab}^A ($A = 1 \to 8$), g_s is the gauge coupling and $F_{\mu\nu}^A$ is the gluon field tensor:

$$F_{\mu\nu}^A = \partial_\mu G_\mu^A - \partial_\nu G_\nu^A - g_s f_{ABC}G_\mu^B G_\nu^C \qquad (1.3)$$

The structure constants f_{ABC} satisfy the relation:

$$[t^A, t^B] = if_{ABC}t^C \qquad (1.4)$$

The non-Abelian nature of QCD leads to two remarkable features: the confinement and the asymptotic freedom. As a result of the confinement, only color-singlet states can be directly observed, which means that quarks and gluons cannot be found free. The asymptotic freedom is the property where the running strong coupling constant decreases with increasing momentum transfer between the strongly interacting particles. This in turn means, that the hard-scattering of quarks and gluons can be described in a perturbative way.

1.2   Collider physics

1.2.1   Proton-proton collisions

One of the consequences of the asymptotic freedom is that in the proton-proton collisions at high energy (with respect to the proton mass) the hard scattering interaction happen between the proton constituents. This also means that the actual center-of-mass energy of the collision is only a fraction of the center-of-mass energy in the proton-proton system and it depends on the probability density functions (PDFs) of the proton constituents.

A sketch that illustrates what happens in a high energy proton-proton collision is in Fig. 1.1:

- the green arrows show the two incoming protons

- two of the initial partons (blue lines) interact with hard scattering (red blob)

- the rest of the initial partons interact at very low energy producing the so called "underlying event" (magenta blobs);

- the beam remnants go forward along the beam line (cyan blobs);

- initial (blue) and final (red) partons radiate gluons or split (initial and final state radiation), producing a shower;

- the colored particles in the shower start hadronizing and form unstable hadrons (light green blobs) that finally decay in stable particles (dark green circles).

(Some more details on hadronic showers and hadronization process are in the following section 1.2.2).
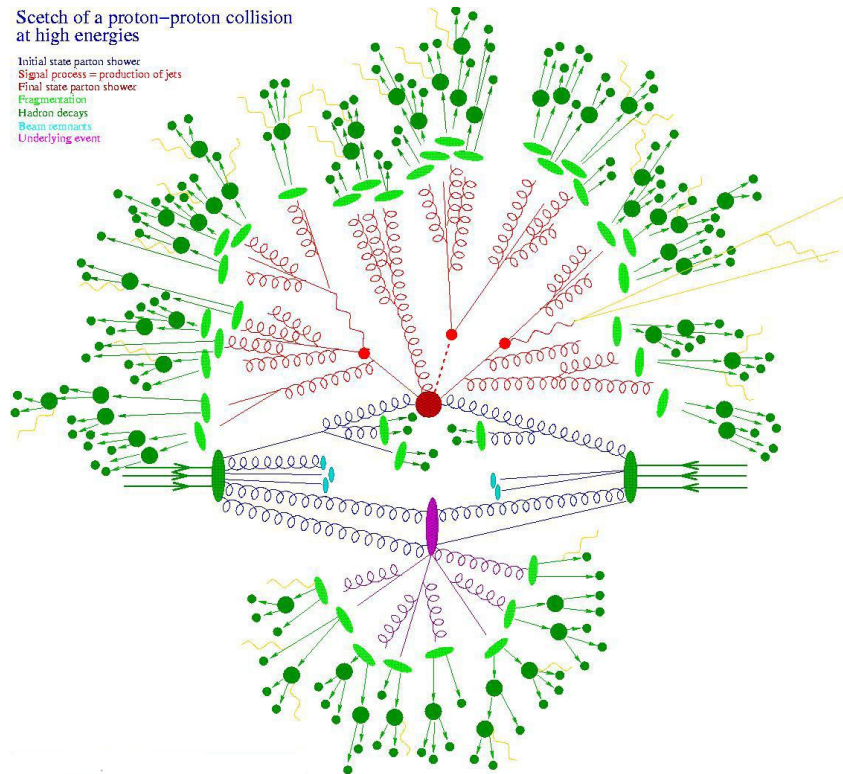
Figure 1.1. Sketch of a proton-proton collision.

Most of the partonic interactions are soft and they are not interesting for the search of resonances at high mass, therefore in experiments it is necessary to select "online", before any sophisticated analysis, the events with a possibly interesting hard scattering interaction. This job is done by the trigger system as we will see with more details further on, in Chapt. 3.

The high intensity of the beams at the LHC results in multiple proton-proton collisions in addition to the interesting one, where hard scattering between the partons takes place. This additional collisions are usually referred to as *pile-up*.

There are two possible contributions to the pile-up:

- one is the energy associated to the proton-proton interactions in the same bunch crossing, called *in-time-pile-up*;

- the other is the due to the finite time of the signal decay time in the detector: if it is comparable with the time between two consecutive bunches (25 ns at the LHC), then the previous and subsequent bunches can contribute to the energy associated to the interesting interaction. This is called *out-of-time-pileup*.

1.2.2 Hadronic jets

We have mentioned in Sec. 1.1 that one of the features of the QCD is the "confinement": because of the confinement, partons cannot be detected free. Instead, the experimental signatures of quarks and gluons are the *jets.* A jet is a "spray" of highly collimated particles, primarily hadrons, but also photons and leptons. It is a not uniquely defined object, but the output of a clustering algorithm, which groups the jet constituents according to their kinematic properties. This procedure is based upon the features of QCD, which describe the transformation of a parton to a set of observable particles. The jet-formation steps are the following:

- **Parton branching:** each parton, whether a gluon or a quark, has a finite probability to split into two partons, which are emitted in small angles with respect to the direction of the initial parton. One feature of the parton branching is that the probability depends on the color factor related to the type of the involved partons. For gluon $\rightarrow$ gluon-gluon, gluon $\rightarrow$ quark-antiquark, and quark $\rightarrow$ gluon-quark splittings, the color factors are $C_{gg} = 3$, $C_{qq} = \frac{4}{3}$, and $C_{qg} = \frac{1}{2}$ respectively. As a result, gluons systematically shower more than quarks. Another implication of parton branching in small angles is that throughout the process, partons are produced close to the direction of the initial partons, which results in a high degree of collimation of the final hadrons. It should be noted, that the parton branching is a perturbative procedure, which can be re-summed to all orders of the perturbation series, under certain assumptions.

- **Hadronization:** when the parton shower has evolved long enough, the energy of the partons is reduced, such that low-momentum transfer occur. In these conditions, the parton interactions become non-perturbative, and the phase of hadronization begins. During the hadronization, partons are combined into color-singlet states, thus forming the hadrons. While the hadronization procedure cannot be described perturbatively, the local parton-hadron duality ensures that the flow of quantum numbers at the hadron level, follows approximately the corresponding flow at parton level.

- **Underlying event and out-of-cone showering:** the term *underlying event* in hadron collisions is used to describe the activity not related to the hard scattering, for example due to multiple parton interactions happening simultaneously. This definition includes the pile-up, that we have defined in the previous section. Since the definition of jets involves the clustering of hadrons which are sufficiently correlated, it can happen that particles originating from the soft interactions are clustered together with those coming from the hard-scattered parton shower. In the opposite direction, partons from the initial shower can be emitted in relatively large angles, and the associated hadrons may not be clustered in the resulting jets. This effect is commonly known as out-of-cone showering.

Despite the fact that the formation of jets is a complicated effect, certain conclusions can be drawn: to first approximation, the kinematical properties of a jet are the same as those of the original parton. However, the various effects involved,

introduce an intrinsic resolution of the hadronic jet properties with respect to the parton properties.

1.2.3   Kinematics of two-parton scattering

Before the details of the strong interaction dynamics are discussed, it is useful to present the kinematical properties of a two-to-two parton scattering. In the topology of the $1+2 \rightarrow 3+4$ scattering (see Fig. 1.2), some general kinematic relations hold, regardless of the details of the interaction.
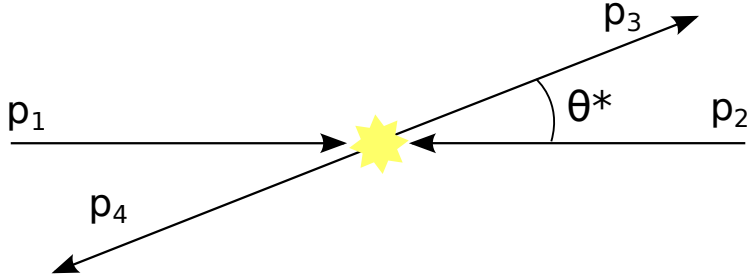
Figure 1.2. Sketch of a two-to-two parton scattering.

The Mandelstam variables of the process are defined as $\hat{s} = (p_1 + p_2)^2$, $\hat{t} = (p_1 - p_3)^2$, and $\hat{u} = (p_2 - p_3)^2$, where p_i are the four-momenta of the partons (see Fig. 1.2). For massless partons, the Mandelstam variables satisfy the relation

$$\hat{t} = -\frac{1}{2}\hat{s}(1 - \cos\theta^*) \tag{1.5a}$$

$$\hat{u} = -\frac{1}{2}\hat{s}(1 + \cos\theta^*) \tag{1.5b}$$

The rapidity of outgoing partons, in the center-of-mass frame, are opposite $(\pm y^*)$, due to transverse momentum conservation, and related to the scattering angle θ^*:

$$\cos\theta^* = \tanh y^* \tag{1.6}$$

The Mandelstam variable $\hat{s}$ can be expressed in terms of the outgoing partons transverse momentum p_T and y:

$$\hat{s} = 4p_T^2 \cosh y^* \tag{1.7}$$

In the laboratory frame, the rapidities $y_{3,4}$ of the outgoing partons are related to the rapidity of the center-of-mass frame $\bar{y}$ and to y as:

$$\bar{y} = \frac{y_3 + y_4}{2} \tag{1.8a}$$

$$y^* = \frac{y_3 - y_4}{2} \tag{1.8b}$$

From the relations above, one can express the scattering angle at the center-of-mass frame as a function of the rapidities of the scattered partons at the laboratory frame:

$$\cos\theta^* = \tanh(\frac{y_3 - y_4}{2}) \tag{1.9}$$

Each initial parton is carrying a fraction x of the hadron momentum and the invariant mass of the two-parton system is expressed as:

$$M^2 = \hat{s} = x_1 x_2 s \tag{1.10}$$

where $x_{1,2}$ are the momentum fractions of the interacting partons and $\sqrt{s}$ is the colliding energy of the hadrons:

$$x_1 = \frac{2p_T}{\sqrt{s}} \cosh y^* \exp \bar{y} \tag{1.11a}$$

$$x_2 = \frac{2p_T}{\sqrt{s}} \cosh y^* \exp -\bar{y} \tag{1.11b}$$

Following from the relation above, the rapidity of the center-of-mass frame $\bar{y}$ can be expressed as a function of the momentum fractions:

$$\bar{y} = \frac{1}{2} = \ln \frac{x_1}{x_2} \tag{1.12}$$

1.2.4 Partonic cross sections

The dynamics of the hard scatter of colliding hadrons is approximately described as a two-to-two process between massless partons. Because of the different structure and color factors of the interaction between the parton types, the matrix elements are different for each subprocess. The leading order (LO) amplitudes can be calculated analytically using the Feynman diagrams at tree level.

Figure 1.3 shows the matrix elements of the various subprocesses, at LO, as a function of $\cos \theta^*$. With the exception of one subprocess ($q_1 \bar{q}_1 \rightarrow q_2 \bar{q}_2$), there is a characteristic $t-$ channel pole which enhances the two-parton scattering at small angles. Another important feature is the fact that, due to the larger color factor of gluons, the matrix element of the subprocesses with gluons in the initial state lead to larger values.

1.2.5 Hadronic cross sections

In a hard scattering process, initiated by colliding hadrons, the experimentally measured cross section can be generally expressed in terms of the parton distribution functions (PDFs) $f(x)$ and the parton-parton scattering cross section $\hat{\sigma}$, summed over all the incoming and outgoing parton types (because the experimentally observed jets cannot distinguish between the parton types):

$$\sigma = \sum_{ij} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij} \left(\alpha_s(\mu_R^2), \frac{Q^2}{\mu_F^2}, \frac{Q^2}{\mu_R^2} \right) \tag{1.13}$$

In the equation above, Q is the characteristic hard scale of the interaction (e.g. the dijet invariant mass in a two-to-two parton scattering), μ_F is the factorization scale, which is of the same order as Q and separates the long-distance, non-perturbative interactions from the hard scattering, and μ_R is the renormalization scale. Both the $\mu_{F,R}$ scales are arbitrary parameters of a fixed-order calculation. At all orders of the perturbative expansion, the cross section should be independent of them
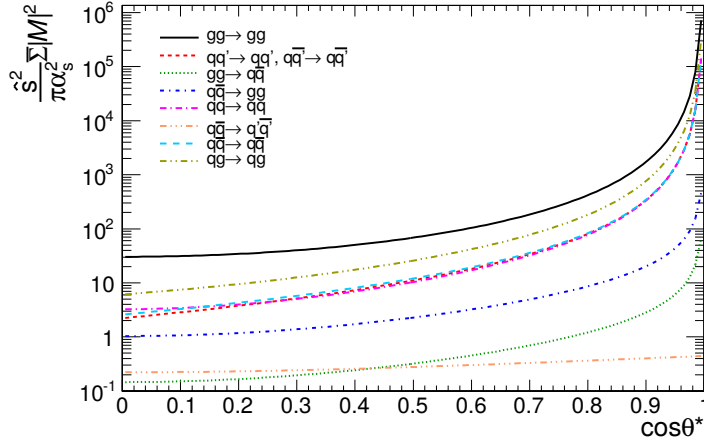
Figure 1.3. Leading order matrix elements for two-to-two massless parton scattering, as a function of $\cos\theta^*$.

$(\partial\sigma/\partial\mu_R = \partial\sigma/\partial\mu_F = 0)$. In all practical calculations of cross sections at a fixed order, it is assumed that $\mu_R = \mu_F = Q$. It should be noted, that the higher the order of the calculation, the weaker is the dependence on $\mu_{R,F}$.

It is often helpful in hadron collisions to quantify the effect of the parton distribution functions by introducing the parton luminosity factor. This is defined as:

$$\frac{dL_{ij}}{d\tau} = \int_0^1 \int_0^1 dx_1 dx_2 f_i(x_1) f_j(x_2) \delta(x_1 x_2 - \tau) \tag{1.14}$$

where

$$\tau = x_1 x_2 = \frac{\hat{s}}{s}. \tag{1.15}$$

In practice, experimental constraints are imposed on the rapidities of the outgoing partons, observed as hadronic jets. It is therefore more convenient to express the parton luminosity as a functions of the variables τ and $\bar{y}$, rather than $x_{1,2}$: $dx_1 dx_2 = \frac{\partial(\tau,\bar{y})}{\partial x_1, x_2} d\tau d\bar{y} = d\tau d\bar{y}$. The parton luminosity then is:

$$\frac{dL_{ij}(\bar{y}_{min}, \bar{y}_{max})}{d\tau} = \int_{\bar{y}_{min}}^{\bar{y}_{max}} f_i\left(\sqrt{\tau}e^{\bar{y}}\right) f_j\left(\sqrt{\tau}e^{-\bar{y}}\right) d\bar{y} \tag{1.16}$$

The hadronic cross section of any process, can be expressed generally as a function of the parton luminosity factor and the partonic cross section:

$$\sigma_{had} = \sum_{i,j} \int \frac{d\tau}{\tau} \left[\frac{1}{s}\frac{dL_{ij}}{d\tau}\right] [\hat{s}\,\hat{\sigma}_{ij}]. \tag{1.17}$$

In the specific case of a two-to-two scattering, resulting in the production of two jets, the differential cross section can be expressed as a function of the diparton invariant mass and the scattering angle at the center-of-mass frame. To first approximation, this cross section is equal to the observed dijet cross section. The matrix elements are folded with the parton distribution functions, giving:

$$\frac{d^2\sigma_{had}}{dm\,d\cos\theta^*} = \frac{\pi\alpha_s^2}{m}\sum_{ij}\left[\frac{1}{s}\frac{dL_{ij}}{d\tau}\right]_{\tau=m^2/s}\hat{\mathcal{F}}_{ij}(\cos\theta^*), \tag{1.18}$$

with

$$\hat{\mathcal{F}}_{ij}(\cos\theta^*) = \sum_{kl}\mathcal{S}(ij\to kl)\frac{1}{1+\delta_{kl}}. \tag{1.19}$$

In the equations above $m\approx\sqrt{\hat{s}}=\sqrt{\tau s}$ is the dijet invariant mass.

1.2.6   Definition of luminosity

The quantity that measures the ability of a particle accelerator to produce the required number of interactions is called the *luminosity*. The number of events, N_{exp}, is the product of the hadronic cross section of interest, σ, and the time integral over the instantaneous luminosity, $\mathcal{L}$:

$$N_{exp} = \int \mathcal{L}(t)dt\cdot\sigma \tag{1.20}$$

The unit of the luminosity is therefore $\text{cm}^2\,\text{s}^{-1}$. Integrated luminosity, on the other hand is usually quoted as the inverse of the standard measures of cross section such as the multiples of barns (picobarns, femtobarns, etc.):

$$1\text{b} = 10^{-24}\text{cm}^{-2} \tag{1.21}$$

In order to compute a luminosity for colliding beams experiment, we have to take into account the properties of the beams, like the number of particles per bunch (N), the number of bunches N_b, the bunch revolution frequency f and the 3D density of particles in the bunches (ρ). A schematic picture is shown in Fig. 1.4.
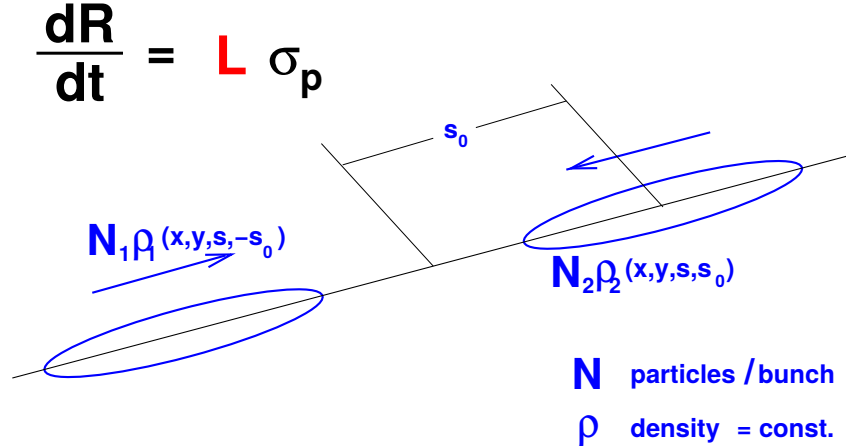
Figure 1.4. Schematic view of a colliding beam interaction.

The luminosity is proportional to the overlap of the bunches, and since the two beams are not stationary but moving through each other, the overlap integral

depends on the longitudinal position of the bunches and therefore on the time as they move towards and through each other. For our integration we use the distance of the two beams to the central collision point $s_0 = c \cdot t$ as the "time" variable (see Fig. 1.4). A priori the two beams have different distribution functions and different number of particles in the beams. Assuming that particles collide head-on with frequency f and that the two bunches meet at $s_0 = 0$, the overlap integral which is proportional to the luminosity ($\mathcal{L}$). We can then write as:

$$\mathcal{L} = 2N_1 N_2 N_b f \int \int \int \int_{-inf}^{+\inf} \rho_1(x, y, s + s_0) \rho_1(x, y, s - s_0) dx dy ds ds_0 \qquad (1.22)$$

Here $\rho_1(x, y, s, s_0)$ and $\rho_2(x, y, s, s_0)$ are the time dependent beam density distribution functions.

Assuming the distributions of the protons in the bunches are identical in the transverse part, gaussian, and uncorrelated in x, y and z, the formula above can be rewritten as:

$$\mathcal{L} = f \frac{N_1 N_2 N_b}{4\pi \sigma_x \sigma_y} \qquad (1.23)$$

where σ_x and σ_y characterize the rms transverse beam sizes in the horizontal and vertical directions.

The revolution frequency in a collider is accurately known and the number of particles or beam intensity is continuously measured with beam current transformers which reach an accuracy of $\sim 1\%$ for LHC nominal beam parameters. The only unknown parameter that needs to be measured is the effective transverse area which depends on the density distribution ρ_1 and ρ_2 of the two beams. This is done with the Van Der Meer technique [6, 7], that consists in scanning the LHC beams through one another to determine the size of the beams at their point of collision. These measurements, when combined with information on the number of circulating protons, allows the determination of an absolute luminosity scale, which in turn is used to calibrate the HF luminometer.

Chapter 2

Hadronic resonances and search at colliders

In presence of a resonance, the cross section of the parton scattering would present a peak in the invariant mass of the final states. The peak would correspond approximately to the resonance mass and would have a finite width that is partly the "natural" width, related to the lifetime, and partly given by instrumental resolution. We will consider narrow resonances, when the natural width is much smaller than the experimental resolution. Basically this search consists in quantifying the deviations of the observed dijet invariant mass from a smoothly falling distribution, expected in the absence of any resonance.

The approach of this thesis is mostly experimental, however it is useful to compare the analysis results to some reference models. The models considered have been previously used in the CMS Run 1 analysis [8], and a brief description of each can be found in Sec. 2.1.

The past dijet searches are discussed in Sec. 2.2, presenting the most recent results of the CMS and ATLAS experiments at the LHC run 1, with $\sqrt{s} = 8$ TeV.

The dijet search is very powerful with the present statistics, even if it is still much smaller than the full 2012 dataset of 20 fb^{-1}, thanks to the increase in the energy of the collisions from 8 to 13 TeV. The potential of the dijet search at 13 TeV is discussed in Sec. 2.3.

2.1 Theoretical models

The results of new physics searches are often presented as a set of upper limits on the cross section times acceptance at 95% CL and compared to new physics models in order to exclude some in a certain range of the explored dijet mass. The models used as "benchmark" for the dijet resonance search at CMS, which is the subject of this thesis, are briefly described in the following paragraphs.

Excited quarks

In various theoretical models, ordinary quarks can be composite objects [9], with a characteristic compositeness scale Λ. As a natural consequence, excited states

are expected, called simply *excited quarks* and denoted by q^*. Depending on the details of the composite models, the excited quarks can have various values of spin and weak isospin. In the simplest case, they take the value of $1/2$. The interaction of excited quarks with the Standard-Model fields is of a "magnetic" type, and the Lagrangian takes the form [10]:

$$\mathcal{L} = \frac{1}{2\Lambda}\bar{q}_R^*\sigma^{\mu\nu}\left(g_s f_s t_a G_{\mu\nu}^a + gf\frac{\vec{\tau}\cdot\vec{W}_{\mu\nu}}{2} + g'f'\frac{Y}{2}B_{\mu\nu}\right)q_L + \text{h.c},\qquad (2.1)$$

where t_a and $\vec{\tau}$ are the generators of the color $SU(3)$ and isospin $SU(2)$, Y is the hypercharge, $G_{\mu\nu}^a$, $\vec{W}_{\mu\nu}$, $B_{\mu\nu}$ are the field tensors, g_s, g, g' are the gauge couplings, and f_s, f, f' are dimensionless constants, accounting for possible deviations from the Standard-Model couplings.

In hadron collisions, the production of an excited quark happens through the quark-gluon fusion. Subsequently, the excited quark decays to an ordinary quark and a gauge boson. The dominant decay channel is $q^* \to qg$, leading to a dijet signature. The partial width for the decay of an excited quark with mass m^* is given by:

$$\Gamma(q^* \to qg) = \frac{1}{3}\alpha_s f_s^2 \frac{m^{*3}}{\Lambda^2} \qquad (2.2)$$

Randall-Sundrum gravitons

The gravity model from Randal and Sundrum [11, 12] (RS) was proposed as a solution to the electroweak vs Planck scale hierarchy problem. In this model the hierarchy is generated by an exponential function of the compactification radius of one extra dimension. The metric in the 5-dimensional space is given by:

$$ds^2 = e^{-2kr_c\phi}\eta_{\mu\nu}x^\mu x^\nu + r_c^2 d\phi^2, \qquad (2.3)$$

where ϕ is the extra dimension with compactification radius r_c, k is a constant of the same order and related to the 5-dimensional Planck scale M, and x^μ are the usual space-time dimensions. The reduced effective 4-D Planck scale $\bar{M}_{Pl}$ is given by:

$$\bar{M}_{Pl}^2 = \frac{M^3}{k}\left(1 - e^{-2kr_c\pi}\right). \qquad (2.4)$$

In this model, spin-2 gravitons appear as the Kaluza-Klein (KK) excitations of the gravitational field $h^{\mu\nu}$, whose coupling to the Standard-Model fields is given by the interaction Lagrangian [13]:

$$\mathcal{L}_I = -\frac{1}{\Lambda_\pi}h^{\mu\nu}T_{\mu\nu}, \qquad (2.5)$$

with $T_{\mu\nu}$ being the energy-momentum tensor of the matter fields. The scale Λ_π and the mass m_n of the KK excitations can be expressed as a function of the fundamental constants of the theory:

$$\Lambda_\pi = \bar{M}_{Pl}e^{-kr_c\pi}, \ \ m_n = kx_ne^{-kr_c\pi}. \qquad (2.6)$$

The coupling constant of the graviton-matter interaction is the inverse of the scale Λ_π:

$$g = \frac{1}{\Lambda_\pi} = x_n \frac{\left(k/\bar{M}_{Pl}\right)}{m_n}, \tag{2.7}$$

where x_n is the n-th root of the the Bessel function of order 1. The phenomenological consequences of the RS-gravitons are essentially determined by their mass, and the ratio $k/\bar{M}_{Pl}$. If the fundamental constants of the model satisfy the relation $kr_c \sim 12$, then $\Lambda_\pi \sim$ TeV, and RS gravitons can be produced in hadron collisions. Through the graviton coupling to the matter fields, RS-gravitons can decay to partons, leading to a dijet signature. The relevant partial widths [14] for the first KK excitation are given by:

$$\Gamma(G \to gg) = \frac{x_1^2}{10\pi} \left(\frac{k}{\bar{M}_{Pl}}\right)^2 m_1, \tag{2.8}$$

and

$$\Gamma(G \to q\bar{q}) = \frac{3x_1^2}{160\pi} \left(\frac{k}{\bar{M}_{Pl}}\right)^2 m_1. \tag{2.9}$$

Axigluons

In the axigluon models [15], the symmetry group of the strong sector is expanded to a chiral color group $SU(3)_L \times SU(3)_R$ which, at some energy, breaks to the diagonal $SU(3)$. Under such a symmetry group, the left-handed and right-handed fermions $\psi_{L,R} = \frac{1}{2}(1 \mp \gamma_5)\psi$ transform differently and the transformations are generated by the $T_{L,R}^a$ generators. Equivalently, the group can be described by a linear transformation of the generators, divided into vectorial $T_V^a = T_L^a + T_R^a$ and axial $T_A^a = T_L^a - T_R^a$ ones. The associated gauge field to the vectorial generators is identified as the usual color field of QCD, while the gauge field associates to the axial generators is called the *axigluon* field. While the exact implementation of the chiral color group is model dependent, there are two universal features: the existence of a massive color octet axigluon field (corresponding to the broken symmetry), and the existence of new particles which are needed to cancel the triangular anomalies.

Axigluons can decay to quark-antiquark pairs, which leads to a dijet experimental signature. Note that due to parity conservation, the axigluon cannot decay to a gluon-gluon pair (all gluon-axigluon vertices must have an even number of axigluons). The axigluon decay to fermions is described by the Lagrangian:

$$\mathcal{L}_A = -ig_s \left(\sum_{ij} \bar{q}_i \gamma_5 \gamma_\mu t_a q_j\right) \mathcal{A}^{a\mu}, \tag{2.10}$$

where $g_s = \sqrt{4\pi\alpha_s}$, $\mathcal{A}$ is the axigluon field, and t^a are the usual color group generators. The width of the axigluon decay can be shown to be [16]:

$$\Gamma_A = \frac{N\alpha_s M_A}{6}, \tag{2.11}$$

where N refers to the open decay channels, and M_A is the axigluon mass. The latter is a free parameter of the theory, determined by the chiral color symmetry breaking scale and the details of the model.

Colorons

Similar to the axial color models, their exist other possibilities to enrich the group structure of the strong sector. Such a model is the flavor-universal coloron [17], where the gauge group is extended to $SU(3)_1 \times SU(3)_2$. The corresponding gauge couplings are denoted as $\xi_{1,2}$. Additionally, the model includes a scalar boson Φ, which develops a non-zero vacuum expectation value and breaks spontaneously the symmetry of the two groups. The diagonal subgroup remains unbroken and is identified as the familiar color group of QCD. In the rotated phase of the physical gauge fields, the initial gauge bosons are mixed, forming an octet of massless gluons and an octet of massive *colorons*. The mass of the colorons is expressed as a function of the fundamental parameters [18]:

$$M_C = \left(\frac{g_s}{\sin\theta\cos\theta} \right) f, \tag{2.12}$$

where θ is the gauge boson mixing angle with $\cot\theta = \frac{\xi_2}{\xi_1}$, and f is the vacuum expectation value of the scalar field. The Lagrangian of the interaction between the colorons field $C^{\mu a}$ and the quarks is similar to QCD:

$$\mathcal{L} = -g_s \cot\theta \left(\sum_{ij} \bar{q}_i \gamma_\mu t_a q_j \right) C^{\mu a}. \tag{2.13}$$

The above interaction predicts the decay of colorons to quarks with kinematically allowed masses. It can be shown that the decay width is:

$$\Gamma_C \approx \frac{N}{6} \alpha_s \cot^2\theta M_C, \tag{2.14}$$

where N is the number of quark flavors with mass less than $M_C/2$.

Color octet scalars

In various theoretical models, bosonic states can arise from gluon-gluon fusion. The *color octet scalar* model (S_8) is one example of exotic color resonances [19]. The coupling of the color octet scalar field with gluons is expressed with the Lagrangian:

$$\mathcal{L} = g_s \frac{\kappa}{\Lambda} d^{abc} S_8^a G_{\mu\nu}^b G^{c,\mu\nu}, \tag{2.15}$$

where g_s is the strong coupling constant, κ is the scalar coupling, Λ is the characteristic scale of the interaction, d^{abc} are structure constants of the $SU(3)$ group defined by the relation $\left\{ t^a, t^b \right\} = \frac{1}{3}\delta^{ab} + d^{abc}t^c$, and S_8, $G_{\mu\nu}$ are the color octet scalar field, and the gluon field tensor, respectively. The width of the color octet scalar resonance is given by:

$$\Gamma = \frac{5}{6} \alpha_s \kappa^2 \frac{M^3}{\Lambda^2}. \tag{2.16}$$

Z' and W'

New gauge bosons arise in models where the symmetry $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ of the electroweak Standard-Model sector is enlarged. Common features in these models are the new gauge coupling constants, which are of the same order as the $\mathrm{SU}(2)_L$ coupling of the Standard Model, and the existence of new gauge bosons, namely W' and Z'. Under the assumption that the new gauge bosons couple to ordinary quarks and leptons similar to their Standard-Model counterparts, the cross sections of these particles are calculated by scaling the the corresponding Standard-Model cross section. In particular, the Fermi constant G_F becomes [20]:

$$G'_F = G_F \left(\frac{M}{M'} \right)^2, \tag{2.17}$$

where M and M' are the masses of W or Z, and W' or Z', respectively.

String resonances

According to the string theory, particles are created by vibrations of relativistic strings, with mass M_s, and they populate Regge trajectories, which relate their spins and masses. In principle, the mass of the fundamental strings is of the order of the Planck scale. However, in some theories with large extra dimensions, it is plausible that M_s lies in the TeV scale. In this case, Regge excited states of quarks and gluons occur in hadron collisions. If the string coupling is small, the basic properties of the Regge excitations (production cross section and width) are model independent (the details of the compactification are irrelevant). This statement is true for parton scattering involving gluons, but only approximately true in the four-quark scattering.

The effect of the Regge excitations can be quantified [21,22] through the presence of a common form factor in the two-to-two parton scattering amplitudes, which is called the *Veneziano form factor* and is written in terms of the Γ-function:

$$V(\hat{s}, \hat{t}, \hat{u}) = \frac{\Gamma(1 - \hat{s}/M_s^2)\Gamma(1 - \hat{u}/M_s^2)}{\Gamma(1 - \hat{t}/M_s^2)}, \tag{2.18}$$

where $\hat{s}, \hat{t}, \hat{u}$ are the usual Mandelstam variables. The physical content of the Veneziano form factor is revealed by an expansion in terms of s-channel poles. Each such pole represents a virtual Regge resonance, with mass $\sqrt{n}M_s$. For the purpose of resonances in the dijet spectrum, only the first-level ($n = 1$) excitation is relevant, while the string mass M_s is the only free parameter. The exact values of the cross sections depend also on the color factors and spin values of the excited states.

E_6 Diquarks

In the context of superstring theory in 10 dimensions, anomaly cancellation requires that the gauge group is $E_8 \times E_8$. Certain models for the compactification of the additional 6 dimensions, predict that the grand unification symmetry group is E_6 [23].

The E_6 models [24] contain color-triplet scalar diquarks, D and D^c with charges $-\frac{1}{3}$ and $\frac{1}{3}$ respectively, which couple to the light quarks u, d.

The interaction Lagrangian between the $E6$ diquarks and the light quarks is given by [25]:

$$\mathcal{L} = \lambda \epsilon_{ijk} \bar{u}^{ci} \frac{1}{2}(1 - \gamma_5) d^j D^k + \frac{\lambda_c}{2} \epsilon_{ijk} \bar{u}^i \frac{1}{2}(1 + \gamma_5) d^{cj} D^{ck} + h.c, \qquad (2.19)$$

where i, j, k are color indices, and λ, λ_c are parameters of the hyper-potential of the general E_6 model. The squared amplitudes for the diquark decays to light quarks are given by [26]:

$$|\mathcal{M}(D \to \bar{u}\bar{d})|^2 = 24\lambda^2 m_D^2, \ \ |\mathcal{M}(D^c \to ud)|^2 = 6\lambda_c^2 m_{D_c}^2. \qquad (2.20)$$

The corresponding widths are:

$$\Gamma_D = \alpha_\lambda M_D, \ \ \Gamma_{D^c} = \frac{1}{4}\alpha_{\lambda_c} M_{D^c}, \qquad (2.21)$$

with $\alpha_\lambda = \lambda^2/4\pi$, $\alpha_{\lambda_c} = \lambda_c^2/4\pi$.

Benchmark Models

The parton-parton resonance models presented in the previous paragraphs involve limited number of free parameters each. The experimental searches traditionally consider benchmark models, with certain parameter assumptions, which are then used to set limits on the masses of the corresponding resonances. Below is a summary of the benchmark models:

- *Axigluons:* the number of quark flavors to which the axigluon can decay is set to $N = 6$, corresponding to the known quarks.

- *Colorons:* the number of quark flavors to which the coloron can decay is set to $N = 6$, and the gauge boson mixing angle is set to $\cot\theta = 1$.

- *Excited quarks:* standard model couplings are assumed ($f_s = f = f' = 1$) and the compositeness scale is set equal to the excited quark mass $\Lambda = M^*$.

- *RS graviton:* the ratio $k/\bar{M}_{Pl}$ is set to $k/\bar{M}_{Pl} = 1$.

- *W', Z':* standard model couplings are assumed.

- *E_6 diquark:* electromagnetic coupling constants are assumed $\alpha_\lambda = \alpha_{\lambda_c} = \alpha_e$.

- *Color octet scalars:* the gauge coupling is set equal to the QCD coupling ($\kappa = 1$), and the characteristic scale of the interaction is set equal to the resonance mass $\Lambda = M$.

Table 2.1 summarizes the basic properties of the resonances discussed in this thesis. In particular, the decay widths are approximate values for the benchmark models, since they also depend on the running of α_s which should be evaluated at a scale equal to the resonance mass. For string resonances the decay width varies significantly, depending on the spin and color quantum numbers of the resonances.

Resonance	Symbol	J^P	Color Multiplet	$\Gamma/(2m_R)$	Decay Channel
excited quark	q^*	$1/2^+$	triplet	0.02	qg
axigluon	A	1^+	octet	0.05	$q\bar{q}$
coloron	C	1^-	octet	0.05	$q\bar{q}$
RS graviton	G	2^-	singlet	0.01	$q\bar{q}, gg$
E_6 diquark	D_6	0^+	triplet	0.004	ud
color octet scalar	S_8	0^+	octet	0.04	gg
heavy W	W'	1^-	singlet	0.01	$q_1\bar{q}_2$
heavy Z	Z'	1^-	singlet	0.01	$q\bar{q}$
string	S	various	various	$0.003 - 0.037$	$q\bar{q},qg,gg$

Table 2.1. Summary of resonances considered in this thesis.

2.1.1 Cross sections

The exact LO calculations of the cross sections and the decay widths of the various resonances involve all the relevant Feynman diagrams associated with each Lagrangian. In practice, the experimental searches presented in this thesis are focused on narrow resonances, which appear as "bumps" on a steeply falling dijet mass spectrum. In all the cases, it is the $s-$channel decay mode of the resonances which produces a "bump".

The cross section of a resonance decaying through the $s-$channel is given by the Breit-Wigner expression:

$$\hat{\sigma}(m)\,(1+2\rightarrow R\rightarrow 3+4) = 16\pi\mathcal{N} \times \frac{\Gamma(1+2\rightarrow R)\times\Gamma(R\rightarrow 3+4)}{\left(m^2-m_R^2\right)^2 + m_R^2\Gamma_R^2}, \quad (2.22)$$

where m_R and Γ_R are the mass and the total width of the resonance, respectively, $\Gamma(1+2\rightarrow R)$ and $\Gamma(R\rightarrow 3+4)$ are the partial widths for the creation and the decay of the resonance to the specific final state. The spin and color multiplicity factor $\mathcal{N}$ is

$$\mathcal{N} = \frac{N_{S_R}}{N_{S_1}N_{S_2}}\frac{C_R}{C_1C_2}, \quad (2.23)$$

where N_{S_R}, $N_{S_{1,2}}$ are the spin multiplicities of the resonance and the initial state particles, while C_R and $C_{1,2}$ are the corresponding color factors. The cross section above arises after integrating over $\cos\theta^*$. The angular dependence of the cross section, in the $s-$channel decay mode, is determined by the spin of the resonance and the spin of the final state particles. It should be noted, that in all resonance cases decaying to two partons, the angular dependence is expressed as a polynomial of $\cos\theta^*$, as opposed to the dominant QCD background, which exhibits a t-channel pole at $\cos\theta^* \rightarrow 1$. More specifically, the angular distributions of the various resonances, in the $s-$channel decay mode, are listed below:

- E_6 *diquark, color octet scalars:* $F(\cos\theta^*) = \text{const.}$

- *excited quark:* $F(\cos\theta^*) \sim 1 + \cos\theta^*$, which becomes $F(|\cos\theta^*|) = \text{const.}$ (odd in $\cos\theta^*$).

- *axigluon, coloron, W', Z':* $F(\cos\theta^*) \sim 1 + \cos^2\theta^*$.

- *RS gravitons:* $F(gg \to G \to q\bar{q}) = F(q\bar{q} \to G \to gg) \sim 1 - \cos^4\theta^*$, $F(gg \to G \to gg) \sim 1 + 6\cos^2\theta^* + \cos^4\theta^*$, and $F(q\bar{q} \to G \to q\bar{q}) \sim 1 - 3\cos^2\theta^* + 4\cos^4\theta^*$.

where $F(\cos\theta^*) \equiv d\hat{\sigma}/d\cos\theta^*$.

It is important to note that, for all these resonance models, the angular distribution of the scattering angle θ^* is significantly different from the QCD interactions that we have described in the previous chapter. In fact the QCD processes are primarily produced via t-channel, with the final jets mostly in the forward region of the detector. The signal processes are instead produced via the s-channel, and have a different distribution in $\cos\theta^*$ depending on the spin of the resonance: for instance the decay products of a spin-0 resonance would have an equal probability of being emitted at every angle θ^* in the center-of-mass reference frame. The distributions of $\cos\theta^*$ for QCD and signal models with different spin are compared in Fig.2.1.
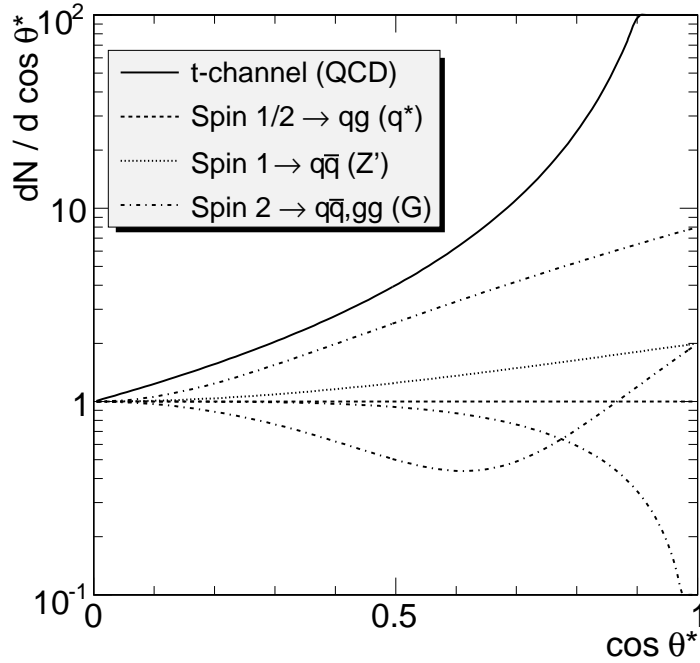
Figure 2.1. Distribution of $\cos\theta^*$ distribution for QCD t-channel processes and resonance with different spin (1/2, 1, 2). The θ^* is the angle between the direction of final partons and the beam line in the center-of-mass reference frame.

For this reason the scattering angle is the most relevant variable for the analysis selection in the past and present dijet resonance searches, as we will see in Sec. 2.2, and than in the detailed description of the Run 2 analysis (Chapt. 5). In practice,

experimental searches impose kinematic constraints on the scattering angle θ^*, such that the QCD background is suppressed. In this case the Breit-Wigner partonic cross section is written as:

$$\hat{\sigma}(m) = \frac{16\pi \times \mathcal{N} \times \mathcal{A}_{\cos\theta^*} \times BR \times \Gamma_R^2}{\left(m^2 - m_R^2\right)^2 + m_R^2 \Gamma_R^2}, \tag{2.24}$$

where BR is the branching fraction of the subprocess, and $\mathcal{A}_{\cos\theta^*}$ is the acceptance after the $\cos\theta^*$ cut. If the resonance is sufficiently narrow ($\Gamma_R << m_R$), the *narrow-width approximation* holds:

$$\frac{1}{\left(m^2 - m_R^2\right)^2 + m_R^2 \Gamma_R^2} \approx \frac{\pi}{m_R \Gamma_R} \delta(m^2 - m_R^2). \tag{2.25}$$

Finally, the hadronic cross section in the narrow-width approximation is derived:

$$\sigma_{had}(m_R) = 16\pi^2 \times \mathcal{N} \times \mathcal{A}_{\cos\theta^*} \times BR \times \left[\frac{1}{s}\frac{dL(\bar{y}_{min}, \bar{y}_{max})}{d\tau}\right]_{\tau = m_R^2/s} \times \frac{\Gamma_R}{m_R}, \tag{2.26}$$

where the parton luminosity $\frac{dL}{d\tau}$ is calculated at $\tau = m_R^2/s$, and constrained in the kinematic range $[\bar{y}_{min}, \bar{y}_{max}]$.

2.1.2 Signal shapes

For narrow resonances the natural width (Γ) is much smaller than the experimental resolution. It has been observed, for these models, that the shape is mostly determined by the final states and independent from the spin. In Fig. 2.2 the simulated signal shapes for three possible combination of the initial and final states (qq → qq, qg → qg, gg → gg) and different resonance masses are shown. In order to separate the contributions to the shape, at each resonance mass are displayed:

- the invariant mass at parton level before hadronization (light green),

- the invariant mass using jets at generator level (blue),

- the invariant mass with reconstructed jets (red).

Looking at Fig. 2.2, we can do some considerations:

- the shapes at parton level are very narrow, symmetric, close to the nominal pole, and similar between the three processes at low mass; at high mass, the processes with gluons are broader and present a significant tail at low mass, far from the resonance pole;

- the shapes at generator level are shifted towards lower mass values and asymmetric: they present a left tail at all masses, but it is larger for processes with gluons;

- the shapes with reconstructed jets are similar to the previous ones, with an additional smearing.
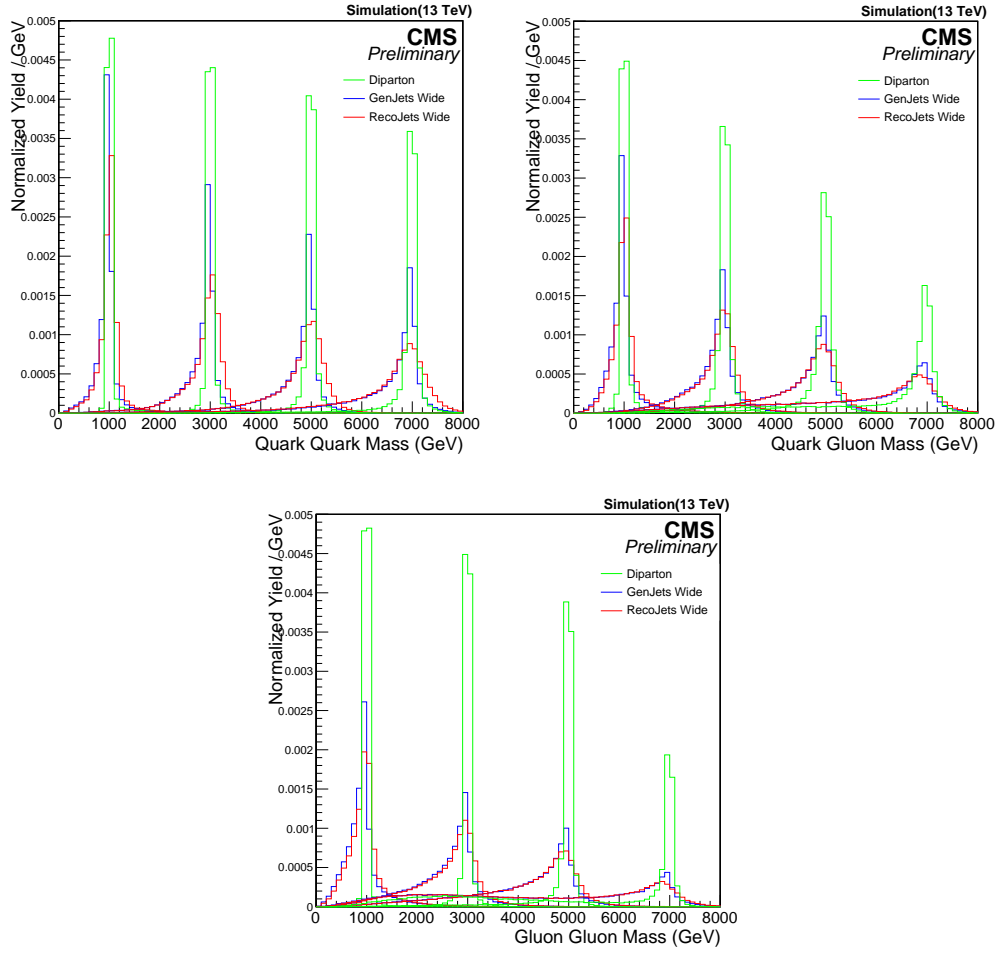
Figure 2.2. Comparison of signal shapes: generator level mass before hadronization, generator level jets, reconstructed jets for qq, qg, and gg resonances.

The tail in the diparton mass, far from the resonance pole, is an effect of the initial state and comes from the parton density functions: the gluons tend to carry a lower energy with respect to quarks, and therefore it is more probable to produce a high mass resonance off shell.

A left tail appears at all masses after the hadronization process: it is mostly due to the final state radiation and the fact that part of the energy carried by the final parton is not well contained in the jet cone. Also these effect is more pronounced for processes with gluons, because gluons have a larger color factor and tend to radiate more than quarks.

An additional symmetric smearing is added from the detector resolution effect.

2.2 Status of dijet searches at $\sqrt{s} = 8$ TeV

The search for a narrow resonances in the dijet mass spectrum is the probably the most natural process to search in a hadron collider. In fact, if the resonance is produced in hadronic collisions via s-channel, that means that it couples to partons and therefore can decay in a pair of jets. We have seen in the previous section that the shape of a narrow resonance is a peak approximately corresponding to the resonance mass, with a smearing caused from different effects: PDFs, final state radiation, and detector resolution. Therefore a narrow resonance would appear as a bump on the smoothly falling dijet mass spectrum predicted by the QCD, and represents a very simple yet striking signature (see Fig. 2.3).
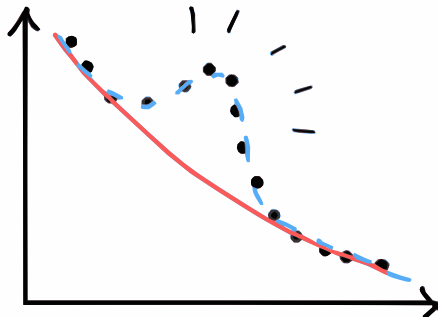
Figure 2.3. Illustration of a dijet resonance appearance over the smooth dijet mass spectrum (red line) of the QCD background.

The dijet mass spectrum has been widely studied at experiments at hadron colliders even before the LHC, with progressively increasing energy collision and integrated luminosity. The technique has remained very similar over the years, and basically consists in a test of smoothness of the observed dijet mass spectrum: once established the background model, data are compared to the background prediction and the significance of observed excesses is studied.

A summary of the dijet searches in the last 30 years is in Tab. 2.2. They are listed in chronological order, along with a summary of the energy and luminosity of the dataset and the reference to the publications. The searches by the UA1 and UA2 experiments used data from the proton anti-proton collisions at the CERN $S\bar{p}pS$ collider at a center-of-mass energy of 0.63 TeV. The searches by the CDF and D0 experiments used data from proton anti-proton collisions at the Fermilab Tevatron at a center-of-mass energy of 1.8 and 1.96 TeV. The searches by the ATLAS and CMS experiments used data from proton proton collisions at the CERN Large Hadron Collider at a center-of-mass energy of 7 TeV, 8 TeV, and recently results at 13 TeV. The latest CMS result is the subject of this dissertation.

Expt.	Yr.	$\sqrt{s}$ (TeV)	$\int L\, dt$ (pb^{-1})	Ref.
UA1	1986	0.63	0.26	[16]
UA1	1988	0.63	0.49	[27]
CDF	1990	1.8	0.026	[28]
UA2	1990	0.63	4.7	[29]
CDF	1993	1.8	4.2	[30]
UA2	1993	0.63	11	[31]
CDF	1995	1.8	19	[32]
CDF	1997	1.8	106	[33]
D0	2004	1.8	109	[34]
CDF	2009	1.96	1130	[35]
ATLAS	2010	7	0.32	[36]
CMS	2010	7	2.9	[37]
ATLAS	2011	7	36	[38]
CMS	2011	7	$1.0\ 10^3$	[39]
ATLAS	2011	7	$1.0\ 10^3$	[40]
CMS	2012	8	$4.0\ 10^3$	[41]
CMS	2014	8	$19.7\ 10^3$	[8]
ATLAS	2014	8	$20.3\ 10^3$	[42]
CMS	Dec. 2015	13	$2.4\ 10^3$	[43]
ATLAS	Dec. 2015	13	$3.6\ 10^3$	[44]

Table 2.2. Searches for dijet resonances at hadron colliders. For each search we list the experiment, year of data publication, center-of-mass energy and integrated luminosity.

In the next sections we will present the results of CMS and ATLAS experiments with the full dataset of 2012, recorded at 8 TeV of center-of-mass energy. These results in fact represent the dijet searches at the highest energy before the recent publication of run 2 data by both collaborations, and still represent the searches with the largest integrated luminosity. The results of CMS and ATLAS are presented respectively in Sec. 2.2.1 and 2.2.2. For a more complete review of the earlier dijet search analyses see [45].

2.2.1   CMS results

As reported also in Tab. 2.2 the final CMS dijet search in 2012 [8] makes use of a dataset of 19.7 fb^{-1} of integrated luminosity.

The analysis technique is almost unchanged in Run 2, therefore for a detailed discussion the reader can refer to Chapter 5. Briefly, the background is modeled fitting the data with an analytic function, as in previous searches in CDF experiment. The signal shapes are, instead, taken from Monte Carlo simulation and studied separately for three final states: quark-quark, quark-gluon and gluon-gluon. The reason of this choice comes from the observations in Sec. 2.1.2:

- narrow resonance shapes are determined mostly by the type of partons they couple to (quarks or gluons);

- they are not symmetric and not easy to model with an analytic function, especially for high masses, where the tail originated by PDF effects is larger;

- the quark-quark, quark-gluon and gluon-gluon final states are representative of the variations in the shapes coming from these effects.

The most relevant selection is on the rapidity separation separation between the two jets, which is related to the scattering angle in the center-of-mass of the interaction. We have seen in Sec. 2.1.1 that the scattering angle distribution of signal and background are very different: the cut used in the analysis correspond to a value of $\cos\theta^* = 0.57$. The other important selection is the dijet mass threshold, necessary to have the full trigger efficiency, which is set to 890 GeV.

In Fig. 2.4 the dijet mass spectrum in data with the fit using a smoothly falling function is shown. Also the QCD MC is superimposed, as a qualitative comparison (it is never really used for the analysis), as well as two shapes for a W' signal hypothesis at 1.9 TeV and for an excited quark of 3.6 TeV. The result is that the data agree very well with the fit parameterization and no significant excesses are observed.

The fit to data starts from the mass value of 890 GeV because this is the lowest energy for which the unprescaled triggers are fully efficient. A search at low mass could be done using *prescaled* triggers, that record only a fraction of the selected events, as ATLAS does (see Sec. 2.2.2), or with another technique called "data scouting" recently introduced in CMS and briefly presented in the next section 2.2.1.

No significant excesses are observed in data and upper limits on the cross sections at 95% Confidence Level (CL) are obtained for quark-quark, quark-gluon and gluon-gluon resonance, summarized in Fig. 2.5. In the same figure, also the cross sections expected from some new physics models are superimposed, in order to produce exclusion limits. When interpreted in the context of specific models the limits exclude: string resonances with masses below 5.0 TeV; excited quarks below 3.5 TeV; scalar diquarks below 4.7 TeV; W bosons below 1.9 TeV or between 2.0 and 2.2 TeV; Z bosons below 1.7 TeV; and Randall-Sundrum gravitons below 1.6 TeV.

A separate search is conducted for narrow resonances that decay to final states including b quarks. The first exclusion limit is set for excited b quarks, with a lower mass limit between 1.2 and 1.6 TeV depending on their decay properties. Searches are also carried out for wide resonances, assuming width-to-mass ratios up to 30%,
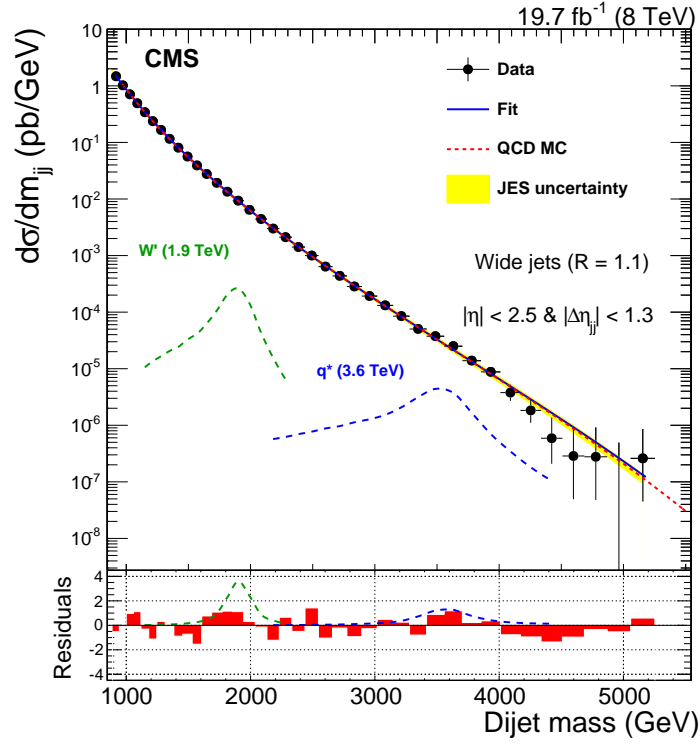
Figure 2.4. Inclusive dijet mass spectrum from wide jets (points) compared to a fit (solid curve) and to predictions including detector simulation of multijet events and signal resonances. The predicted multijet shape (QCD MC) has been scaled to the data. The vertical error bars are statistical only and the horizontal error bars are the bin widths. For comparison, the signal distributions for a W resonance of mass 1.9 TeV and an excited quark of mass 3.6 TeV are shown. The bin-by-bin fit residuals scaled to the statistical uncertainty of the data, $(data - fit)/\sigma_{data}$, are shown at the bottom and compared with the expected signal contributions.
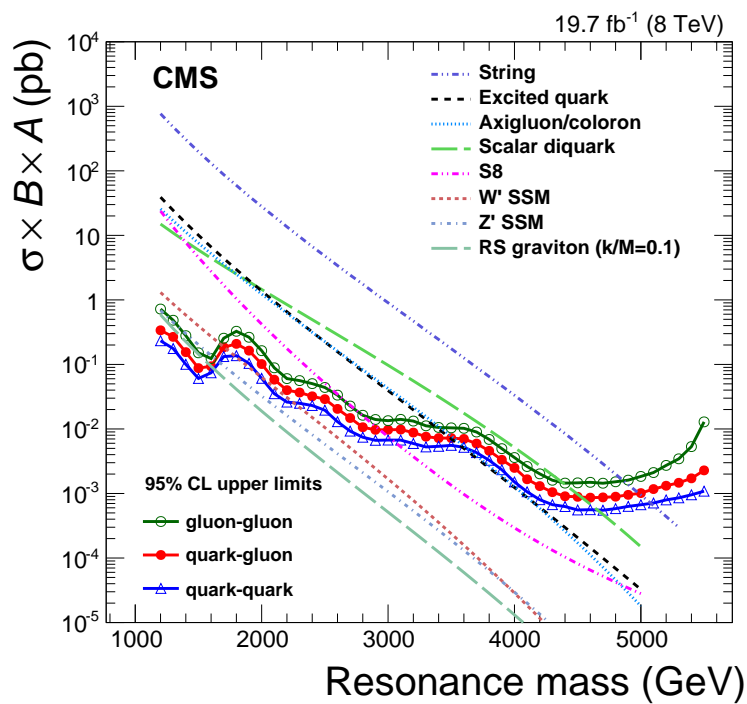
Figure 2.5. The observed 95% CL upper limits on $\sigma \times B \times A$ for narrow dijet resonances. Limits on gluon-gluon, quark-gluon, and quark-quark narrow resonances from the inclusive analysis, compared to LO theoretical predictions for several resonance models.

and for quantum black holes with a range of model parameters. The wide resonance search excludes axigluons and colorons with mass below 3.6 TeV, and color-octet scalars with mass below 2.5 TeV. Lower bounds between 5.0 and 6.3 TeV are set on the masses of quantum black holes.

The analysis presented in this thesis for 2015 dataset (described with all details in Chapters 5-7) does not include the b-quark exclusive search and the wide resonance models. These extensions are planned to be added in the near future also to the 13 TeV search.

Dijet search at low mass

In the standard CMS analysis for the search of dijet resonances the fit to the dijet mass spectrum cannot start before 890 GeV, due to the triggers threshold. Data scouting [46], introduced by CMS as a solution to this problem, is a technique that allows to explore with unprescaled triggers the low dijet mass region from 390 GeV.

Since the ultimate technical limitation is the disk-writing rate, one can write more events reducing the event size. In the case of data scouting, this is achieved limiting the stored event information to the list of four-momenta for jets and leptons in the event. The average event size is reduced to $\approx$10 kB, to be compared to the standard 500 kB/event for ordinary events.

The analysis follows closely the "standard" search performed on the ordinary data stream which is sensitive to resonance masses above $\sim$1 TeV (see previous section). The result is interpreted as a limit on the production cross section for different resonance models (qq, qg and gg final states) shown in Fig. 2.6 as a function of the resonance masses.

Following Ref. [47], the result is translated into an upper limit on the coupling for a Z_B resonance as a function of the resonance mass, and compared to results obtained with similar searches at different collider energies in Fig. 2.7. The result derived from data scouting events translates into the most sensitive results for resonances with mass between ≈ 500 GeV and ≈ 800 GeV.

2.2.2 ATLAS results

The dijet search in ATLAS with proton-proton collisions at a center-of-mass energy $\sqrt{s} = 8$ TeV makes use of the full 2012 dataset, with an integrated luminosity of 20.3 fb^{-1} [42]. The technique will not be described in detail here: it is overall similar to the CMS analysis, with some differences that will be pointed out.

The most relevant selection criteria is the rapidity separation between the two leading jets: events must satisfy $|y| = \frac{1}{2}|y_{\text{lead}} - y_{\text{sublead}}| < 0.6$. This corresponds basically to a $|\Delta\eta_{\text{jj}}| < 1.2$, similar to the cut applied in CMS at 1.3, to reduce QCD background. The invariant mass cut of $m_{\text{jj}} > 250$ GeV is chosen such that the dijet mass spectrum is unbiased by the kinematic selection on p_T. This value is much lower than the 890 GeV threshold applied in CMS search, because combinations of prescaled single-jet triggers are used to reach lower dijet invariant masses. The highest prescale for the trigger combinations used in this analysis is 1/460000 (this trigger combination is used for jets with p_T between 59 and 99 GeV). This effectively means that the search at low dijet mass is not performed with the full dataset of
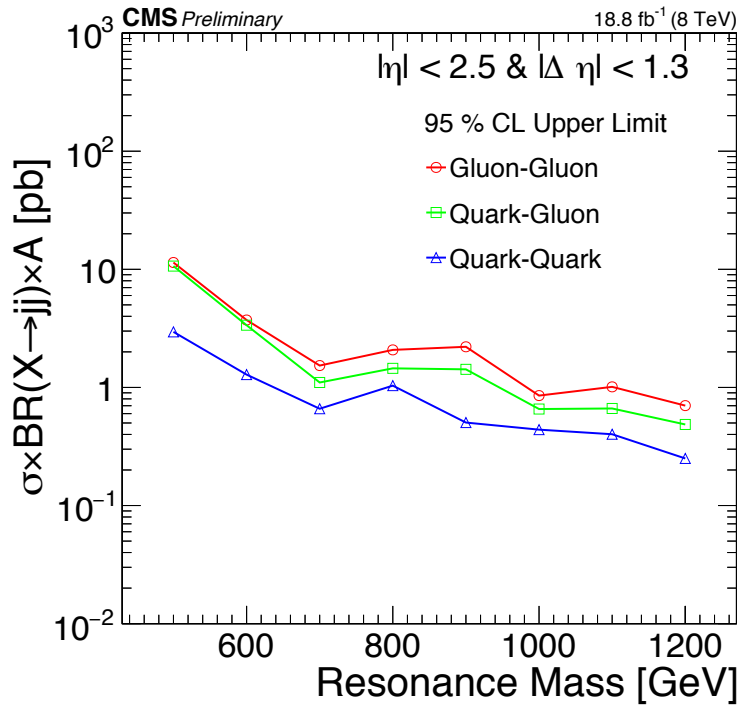
Figure 2.6. 95% CL upper limits on $\sigma \times \mathrm{BR} \times \mathrm{A}$ for a narrow resonance decaying to gg final states (open circles), gq final states (open squares), and qq final states (open triangles).
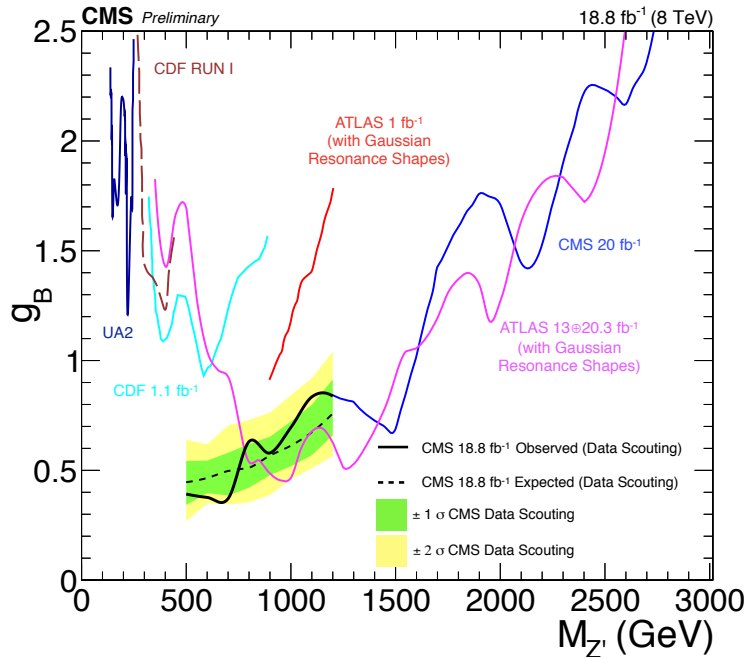
Figure 2.7. Coupling upper limit for a Z'_B resonance, as a function of its mass. The results from this study is compared to results obtained with similar searches at different collider energies.

20.3 fb^{-1}, but only with a fraction of it, depending on the prescale of the trigger active in that m_{jj} region.

To obtain the background model, data are fit with the same function used in CMS and in previous searches at Tevatron. No resonance-like features are observed in the dijet mass spectrum, as it is shown in Fig. 2.8.
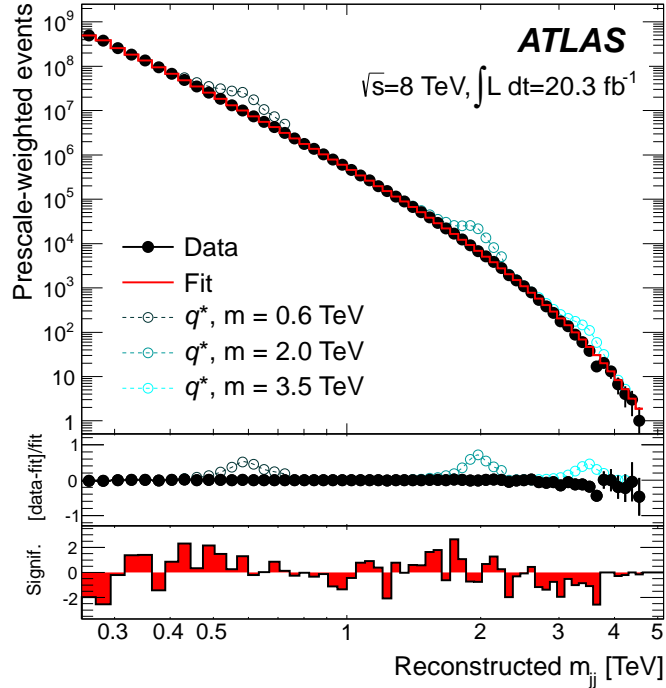
Figure 2.8. The reconstructed dijet mass distribution (filled points) fitted with a smooth functional form (solid line). Predictions for three q* masses are shown above the background. The central panel shows the relative difference between the data and the background fit with overlaid predictions for the same q* masses. The bin-by-bin significance of the data–background difference considering statistical uncertainties only is shown in the bottom panel.

Limits on narrow dijet resonance production are determined using a collection of hypothetical signal shapes that are assumed to be Gaussian-distributed in the dijet mass, unlike in CMS analysis, where the signal templates are taken directly from simulation. Signal shape templates are generated with means (m_G) ranging from 200 GeV to 4.0 TeV and with standard deviations (σ_G) corresponding to the dijet mass resolution estimated from MC simulation and ranging from 7% to 15% of the mean.

For particles with a non-zero natural width generated at masses close to the collision energy, the parton luminosity favors lower-mass collisions. This creates an asymmetric resonance not well represented by a Gaussian distribution. To handle this scenario, Breit-Wigner signals of fixed intrinsic widths (0.5% to 5% of the resonance mass) are generated and multiplied by the parton luminosities for different initial states (qq, qg, gg and $q\bar{q}$). Then parton shower effects and detector resolution

effects are simulated to obtain the final template. The result is then truncated below 250 GeV due to the dijet mass cut.

Limits on the cross section times acceptance are set at the 95% credibility level for various hypotheses of new phenomena in terms of mass or energy scale are shown in Fig. 2.9. This analysis excludes excited quarks with a mass below 4.06 TeV, color-octet scalars with a mass below 2.70 TeV, heavy W bosons with a mass below 2.45 TeV, chiral W bosons with a mass below 1.75 TeV, and quantum black holes with six extra space-time dimensions with threshold mass below 5.66 TeV.
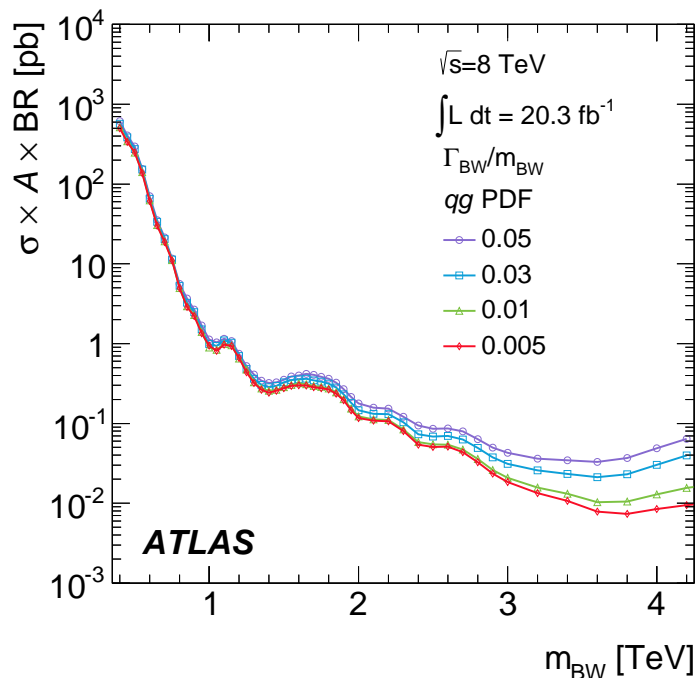
Figure 2.9. The 95% CL upper limits on $\sigma \times A$ for a Breit-Wigner narrow resonance produced by a qg initial state decaying to dijets and convolved with PDF effects, dijet mass acceptance, parton shower and non-perturbative effects and detector resolution, as a function of the mean mass, m_{BW}, for different values of intrinsic width over mass (Γ_{BW}/m_{BW}), taking into account both statistical and systematic uncertainties.

2.3   Potential of LHC at $\sqrt{s} = 13$ TeV

The big jump in the center-of-mass energy from 8 TeV to 13 TeV, makes the Run 2 analysis much more powerful than the previous one, even with a smaller dataset. The reason can be well explained looking at Fig. 2.10.

We have seen in Sec.1.2.4 that the final hadronic cross section can be written in terms of partonic cross section and parton luminosity. The parton luminosity factor in the cross section represents the effect of parton density functions, and it is practically the probability of the hard scattering interaction to happen between two given partons, at a certain center-of-mass energy in the partons' reference frame

$(\sqrt{\hat{s}})$, given the center-of-mass-energy in the protons' reference frame $(\sqrt{s})$. The parton luminosity ratio between 13 TeV and 8 TeV versus M_X shown in Fig. 2.10 represents the gain of LHC Run 2 with respect to Run 1, in terms of probability of producing a resonance at mass M_x. The figure shows that the 13 TeV data have a statistic power of ∼10 times the 8 TeV data for a resonance mass of 2 TeV and of ∼100 for a resonance of 4 TeV.
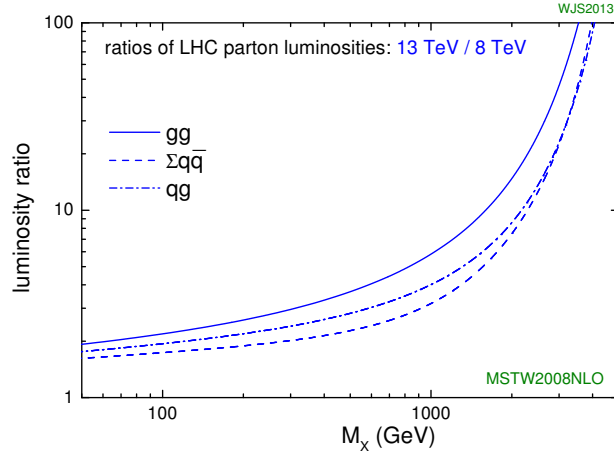
Figure 2.10. The parton luminosity ratio between $\sqrt{s} = 13$ TeV and 8 TeV as a function of the resonance mass M_X for *qq*, *qg* and *gg* final states. The ratio increases quickly at high dijet mass.

It is possible to compare visually the sensitivity of the full dataset of 2012 at 8 TeV and the dataset of 2.4 fb^{-1} presented in this thesis at 13 TeV in Fig. 2.11. The figure is obtained from the tool in [48] and represents an estimate of the system mass that can be probed in BSM searches at one collider setup ("collider 2", e.g. LHC 8 TeV with 20 fb^{-1}) given an established system mass reach of some other collider setup ("collider 1", e.g. LHC 13 TeV with 2.4 fb^{-1}).

The estimate is obtained by determining the system mass at collider 2 for which the number of events is equal to that produced at collider 1, assuming that cross sections scale with the inverse squared system mass and with partonic luminosities. The exact results depend on the relevant partonic scattering channel, as represented by the different lines (q_i), and the band covers the spread of those different partonic channels.

There are quite a few caveats in interpreting the numbers:

- they assume that signal and background are driven by the same partonic scattering channel;

- they assume that reconstruction efficiencies, background rejection rates, etc., all stay reasonably constant as the collider setup changes;

- they assume that the cross-sections are simply proportional to the partonic

luminosity at a given scale (divided by the mass-scale squared), ignoring detailed production spectra, higher-order QCD effects, etc.

In the dijet search case these assumptions are reasonable.
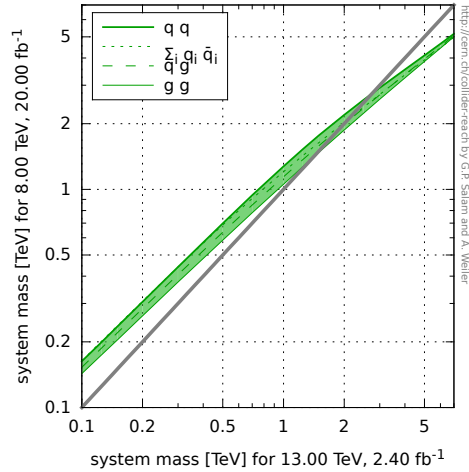
Figure 2.11. Estimate of the system mass that can be probed in BSM searches at one collider setup ("collider 2" on y axis) given an established system mass reach of some other collider setup ("collider 1" on x axis). The different green lines indicates the different final states (qq, qg, gg, and the sum of the three). Collider 2 is the LHC setup of Run 1: 8 TeV of center-of-mass energy with an integrated luminosity of 20 fb^{-1}; collider 1" is the LHC at 13 TeV with 2.4 fb^{-1} of integrated luminosity.

The mass where the grey diagonal crosses the green lines, indicates the point where the two setups have the same sensitivity. Where the green lines go below the grey, the 13 TeV dataset starts to be more statistically powerful than the 8 TeV full dataset. This happens around 2 TeV if we consider an average of all partonic channels. Considering just qq final states, the run 2 analysis is expected to be more sensitive than run 1 above 2.5 TeV, while considering just gg final states it happens already at 1.5 TeV.

Chapter 3

The Large Hadron Collider and the CMS experiment

3.1   The Large Hadron Collider

The Large Hadron Collider (LHC) at CERN is the largest and most energetic particle accelerator ever built. It occupies the 27 km long tunnel previously hosting the LEP collider, about 100 meters underneath the surface across the French-Swiss national border near Geneva. It is a superconducting proton-proton collider, capable of producing collisions at a center of mass energy of up to 14 TeV, and a maximal instantaneous luminosity of 10^{34} cm^{-2}s^{-1}. Its first research run took place from March 2010 to February 2013 at an initial energy of 3.5 TeV per beam (7 TeV total), almost 4 times more than the previous world record for a collider, rising to 4 TeV per beam (8 TeV total) from 2012. On February 2013 the LHC's Run 1 officially ended, and it was shut down for planned upgrades. Test collisions restarted in the upgraded collider on April 2015, reaching 6.5 TeV per beam on May 2015 (13 TeV total, the current world record for particle collisions). Its second research run commenced on schedule, on 3 June 2015.

The LHC proton injection chain is schematically shown in Figure 3.1

The LHC can cross its beams in four interaction points. Two of them have high luminosity and are dedicated to the general purpose experiments ATLAS [49] and CMS [50]. The other two, at lower luminosity, serve the ALICE [51] and LHCb experiments [52], respectively focused on heavy ion physics and CP violation measurements.

Protons are accelerated three times before they enter the LHC ring: the LINAC brings them to 50 MeV, the Proton Synchrotron (PS) to 1.4 GeV, and finally the Super Proton Synchrotron (SPS) injects them into the LHC at 450 GeV.

The LHC then completes the acceleration by bringing them to 13 TeV with its 400 MHz radiofrequency cavities, capable of 'kicks' which result in increases of the proton energy of 0.5 MeV per turn.

Since the collisions occur between particles of the same electrical charge, two separate acceleration cavities and two different magnetic field configurations are required. The LHC is equipped with 1232 superconducting 14.2 m long Niobium-Titanium dipole magnets, cooled down to 1.9 K by means of super-fluid Helium,
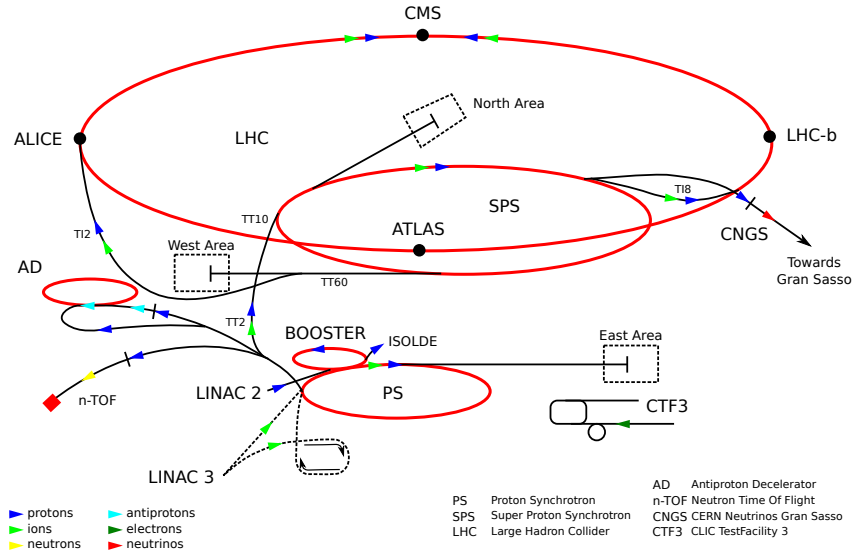
Figure 3.1. LHC injection scheme.

that create a bending magnetic field of about 8.3 T. The magnets are placed in the eight curved sections which connect the straight sections of the LHC ring.

The high luminosity of the LHC is obtained by a high frequency bunch crossing and a high number of protons per bunch: two beams of protons with an energy of up to 7 TeV (6.5 TeV in the initial physics runs of Run 2), circulating in two different vacuum chambers, contain each up to 2808 bunches (typical number of bunches around 1500-2000). The bunches, with a nominal number of 10^{11} protons each, have a very small transverse spread (about 15 μm in the transverse directions) and are about 7.5 cm long in the beam direction at the interaction points. During the first three weeks of data taking, in July 2015, the temporal bunch spacing was 50 ns, while for the rest of the year the LHC operated at a nominal frequency of 40 MHZ (one collision each 25 ns).

A summary of the principal LHC technical parameters is given in Table 3.1.

3.2   The CMS experiment

The Compact Muon Solenoid (CMS) experiment [50] is one of the two general-purpose detectors which take data at the LHC. Its physics goals range from the search for the Higgs boson to the searches for physics beyond the Standard Model, to the precision measurements of already known particles and phenomena.

The overall layout of CMS is shown in Fig. 3.2. From the interaction point, the outgoing particles encounter the following subdetectors:

- the tracking system,

- the electromagnetic calorimeter,

- the hadronic calorimeter,

Circumference [km]	27
Number of magnet dipoles	1232
Dipolar magnetic field [T]	8.33
Maximal crossing rate [MHz]	40
Maximal number of bunches	2808
Actual number of bunches	$\sim$2300
Magnet temperature [K]	1.9
Maximal beam energy [TeV]	7
Actual beam energy [TeV]	6.5
Maximal luminosity [cm^{-2} s^{-1}]	10^{34}
Actual peak luminosity [cm^{-2} s^{-1}]	$\sim 5 \cdot 10^{33}$
Protons per bunch	$1.1 \cdot 10^{11}$
Bunch time separation [ns]	25
Bunch lenght [cm]	7.5
Bunch transverse size [µm]	15
Crossing angle [rad]	$2 \cdot 10^{-4}$
Beam lifetime [h]	7
Luminosity lifetime [h]	10

Table 3.1. Summary of the principal LHC technical parameters.

- the return yoke of the superconductive magnet (that is not a "subdetector", but serves also as a filter, allowing only muons and weakly interacting particles such as neutrinos to pass through and reach the muon chambers),

- the muon system.

The components of the CMS detector are described in the following sections, after an introduction of the coordinate system convention in Sec.sec:coordinates.

3.2.1 Coordinate system

CMS uses a right-handed cartesian coordinate system, with the origin at the nominal interaction point at the center of the detector. The x-axis points to the center of the LHC, the y-axis points upwards, perpendicular to the LHC plane, and the z-axis along the anticlockwise-beam direction.

Given the cylindrical system, it is also useful to define some variables in a angular reference system. The polar angle, θ, is measured from the positive z-axis and the azimuthal angle, ϕ, lies in the x-y plane. Transverse energy and momentum (p_T and E$_T$), i.e. the particle's energy (E) and momentum (p) in the transverse plane, are therefore defined as $p_T = p \cdot \sin(\theta)$ and E$_T$ = E $\cdot \sin(\theta)$.

A frequently-used variable, at hadron colliders, is the *pseudorapidity*, defined as:

$$\eta = -\ln \tan \frac{\theta}{2} \qquad (3.1)$$

CMS DETECTOR

Total weight	: 14,000 tonnes
Overall diameter	: 15.0 m
Overall length	: 28.7 m
Magnetic field	: 3.8 T

STEEL RETURN YOKE
12,500 tonnes

SILICON TRACKERS
Pixel (100x150 µm) ~16m² ~66M channels
Microstrips (80x180 µm) ~200m² ~9.6M channels

SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000A

MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers

PRESHOWER
Silicon strips ~16m² ~137,000 channels

FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels

CRYSTAL ELECTROMAGNETIC CALORIMETER (ECAL)
~76,000 scintillating PbWO₄ crystals

HADRON CALORIMETER (HCAL)
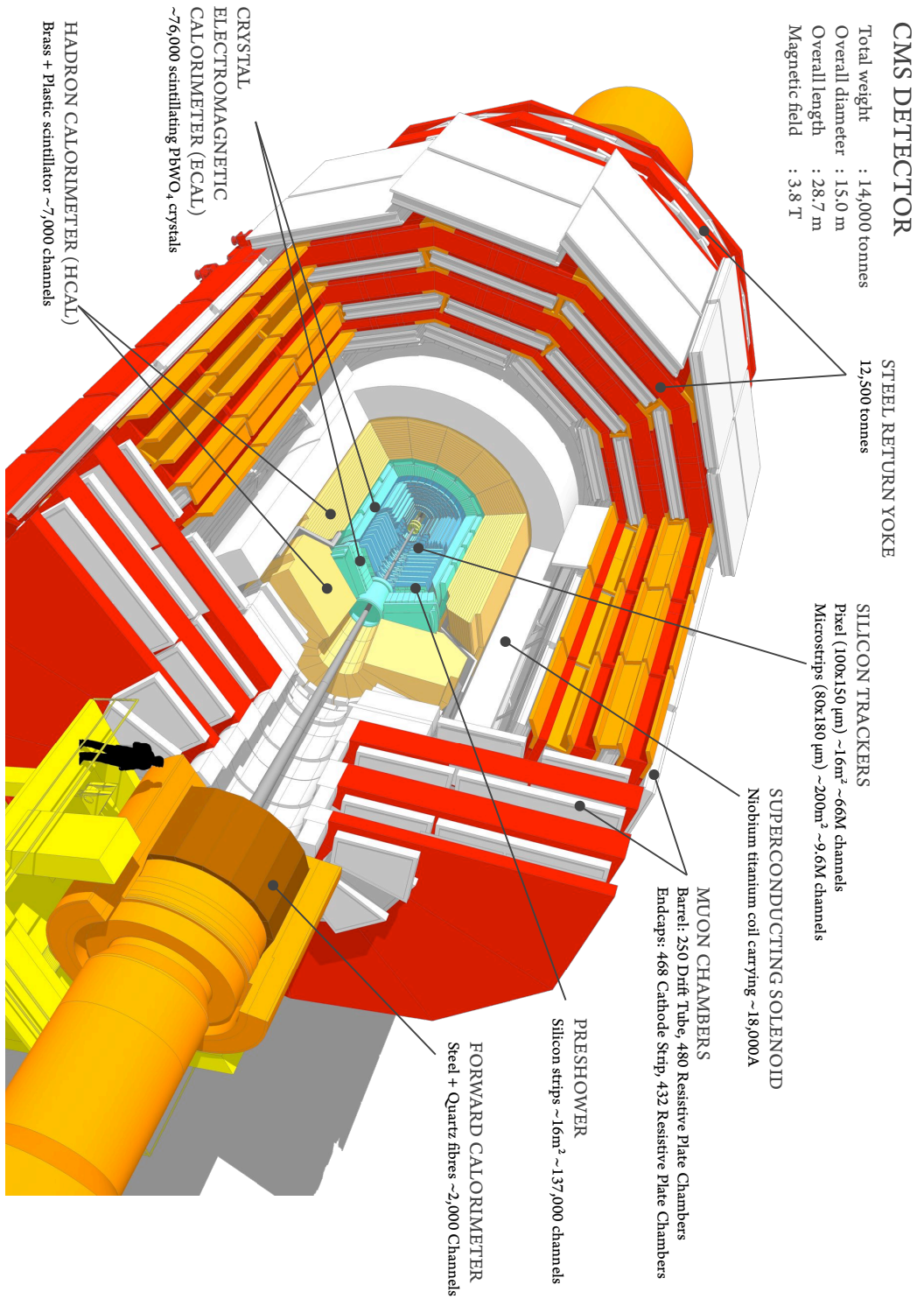Brass + Plastic scintillator ~7,000 channels

Figure 3.2. Inner structure of the CMS detector.

As a function of three-momentum p, pseudorapidity can be written as

$$\eta = \frac{1}{2}\ln(\frac{|p| + p_L}{|p| - p_L}) = \text{artanh}(\frac{p_L}{|p|}) \tag{3.2}$$

where p_L is the component of the momentum along the beam axis. The pseudorapidity can be rewritten as a function of the polar angle θ and by definition it is zero for $\theta = \pi/2$ (central part of the detector), and $\pm \inf$ for $\theta = 0, \pi$ (see Fig. 3.3):

$$\eta = -\ln\tan(\frac{\theta}{2}) \tag{3.3}$$

In the limit where the particle is traveling close to the speed of light, or equivalently in the approximation that the mass of the particle is negligible, one can make the substitution $m \ll p \Rightarrow E \approx p \Rightarrow \eta \approx y$, and hence the pseudorapidity converges to the definition of rapidity used in experimental particle physics:

$$y \equiv \frac{1}{2}\ln\left(\frac{E + p_{\text{L}}}{E - p_{\text{L}}}\right) \tag{3.4}$$

The advantage to use pseudorapidity is that it is function of the polar angle only, and does not depend from the longitudinal component of the momentum p_L, that for the colliding partons is generally unknown.
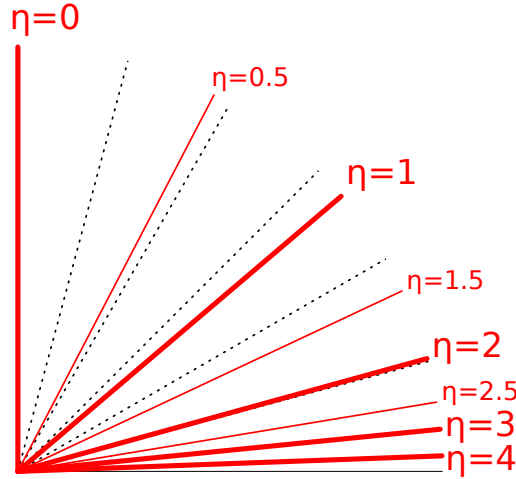
Figure 3.3. Pseudorapidity values shown on a polar plot. The angle of $\theta = 0$ is along the beam axis which lies on the horizontal line and corresponds to $\eta = \infty$.

Moreover in the high energy approximation, as for rapidity, intervals of pseudorapidity are invariant under Lorentz boosts along the beam direction. The Lorentz invariant angular distance $\Delta\eta$ can be expressed also as a function of the scattering angle (in the partons center-of-mass reference frame) θ^*:

$$\cos\theta^* = \tanh(\frac{\Delta\eta}{2}) \tag{3.5}$$

Pseudorapidity can also be used to define a Lorentz-invariant measure of angular separation between particles, $\Delta R^2 \equiv (\Delta\eta)^2 + (\Delta\phi)^2$.

3.2.2   Magnet

In order to achieve a compact and high-resolution muon detection system, a large bending power is required. This can be achieved by a relatively small solenoid, provided that an intense magnetic field is produced, as the bending starts at the collision vertex. A large enough length/radius ratio is also demanded for, in order to ensure good momentum resolution in the forward region as well.

These considerations led to the choice of a 13 m long superconducting cylindrical Niobium-Titanium coil, with a diameter of 5.9 m [53]. It provides a uniform magnetic field of 3.8 T at its center, carrying a current of 18 kA and a total stored magnetic energy of 2.4 GJ. The magnet flux is returned by a saturated iron yoke, which also works as mechanical support structure of the detector.

3.2.3   Tracker

Momentum of particles is crucial to build up a picture of events at the heart of the collision. One method to calculate the momentum of a particle is to track its path through a magnetic field; the more curved the path, the less momentum the particle had. The CMS tracker records the paths taken by charged particles by finding their positions at a number of key points. The design goal of the inner tracking system is to reconstruct isolated, high-p_T electrons and muons with efficiency greater than 95%, and tracks of particles within jets with efficiency greater than 90%, within a pseudorapidity coverage of $|\eta| < 2.4$.

The CMS tracker is entirely based on silicon detector technology. It is in fact composed in three detectors one inside the other, all built with similar technology and geometry. Starting from the center we have:

- Pixel Detector based on silicon pixels.

- Inner Tracker based on silicon microstrips.

- Outer Tracker based on silicon microstrips.

The two outer parts form together the Silicon Strip Tracker and have a sensitive area of 206 m^2 with about 9.3 million channels where each channel is a strip. Instead the Pixel Detector is based on square pixels 150 μm $\times$ 150 μm .Their total number is around 45 millions.

The material budget in units of radiation lenght is shown in Fig. 3.5 as a function of pseudorapidity. The tracker adds up from less than one radiaton lenght in the barrel to a maximum of two in the transition region between barrel and endcaps.

3.2.4   Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) [54] is the detector dedicated to the measurement of electrons and photons energy. It is of course very important also in the measurement of hadronic jets, the most relevant physics objects for the analysis presented in this thesis, since a significant fraction of the jets energy is electromagnetic. In particular, CMS ECAL have been designed for the measurement of a low-mass Higgs boson in the decay channel $H \to \gamma\gamma$, and it played a fundamental role in
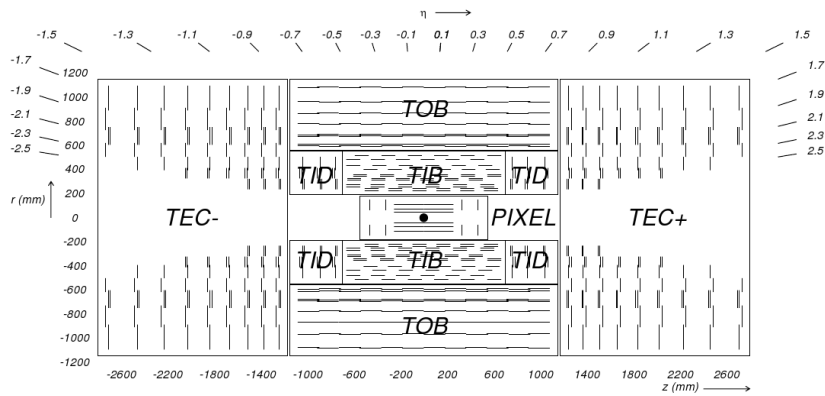
Figure 3.4. View of the CMS tracker in the rz-plane. Each line in the strip tracker represents a silicon strip detector, whereas lines in the pixel detector represent ladders and petals on which the detectors are mounted in the barrel and endcaps,
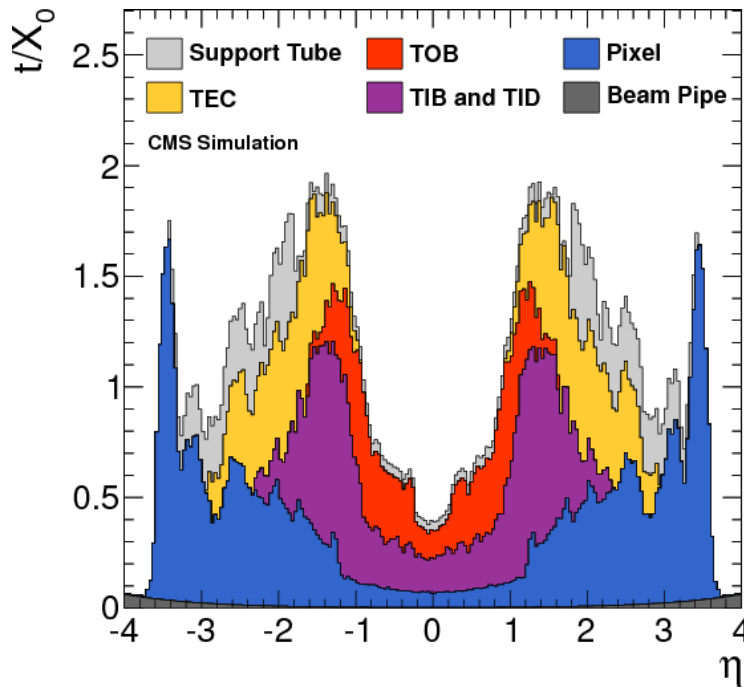
Figure 3.5. Silicon tracker material budget as a function of pseudorapidity, expressed in units of radiation lengths (X_0). Different material categories are shown: beam pipe, pixels, different parts of the strip system (TIB, TID, TOB and TEC) and support tubes

its discovery. The chosen solution to achieve the necessary energy resolution was an hermetic, homogeneous electromagnetic calorimeter, made of 75848 scintillating crystals of lead tungstate ($PbWO_4$).

The structure is divided in a central cylindrical part (the barrel) and two end-caps. In front of the endcaps an additional *preshower* detector is placed (see next section). A 3D view of the detector is in Fig. 3.6 on top: barrel, endcap and preshower are visible, and a $H \to \gamma\gamma$ event is shown for illustrative purpose. Figure 3.6 bottom shows a quadrant of the ECAL in the longitudinal section.
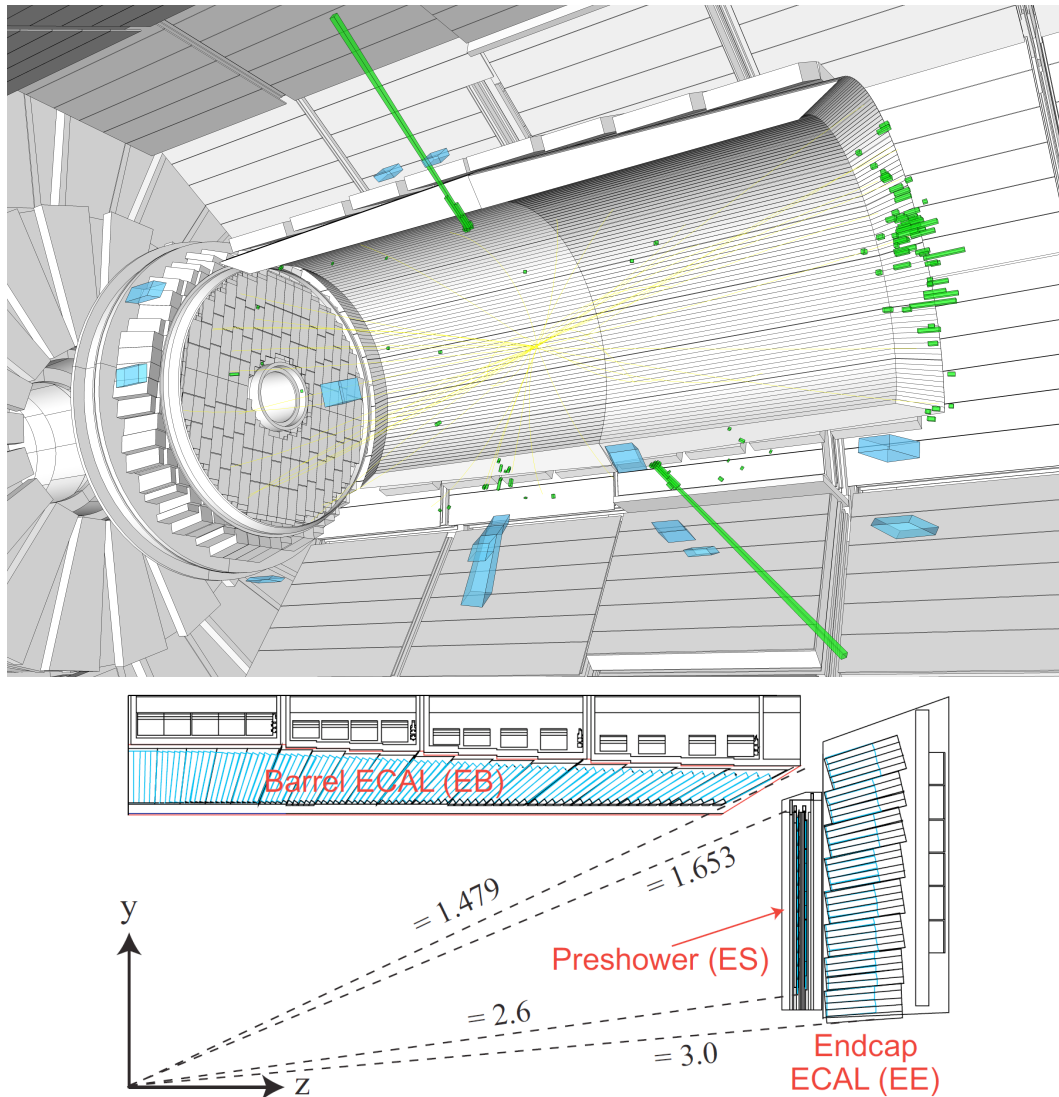
Figure 3.6. Top: CMS ECAL in a 3D view with candidate Higgs boson decaying into two photons (large green towers). Bottom: Longitudinal view of one quarter of the ECAL.

Lead tungstate crystals

The characteristics of the Lead Tungstate crystals ($PbWO_4$) make them an appropriate choice for operation at LHC.

- High density (8.3g/cm^3) and short radiation lenght (0.89 cm) allow the construction of a compact calorimeter inside the magnet.

- Small Molière radius (2.2 cm) ensures lateral shower containment and, therefore, high granularity. High granularity is needed for $\pi^0 - \gamma$ separation and angular resolution.

- Fast light emission: the scintillation decay time is of the same order of magnitude as the LHC bunch crossing time. About 80% of the light is emitted in 25 ns.

The light output is relatively low: about 4.5 photoelectrons per MeV are collected in both the avalanche photodiodes (APDs) and the vacuum phototriodes (VPTs), where the higher APD quantum efficiency is balanced by their smaller surface coverage on the back face of the crystal. The crystals emit blue-green scintillation light with a broad maximum at 420 nm.

Geometry

The barrel part of the ECAL (EB) covers the pseudorapidity range $|\eta| < 1.479$ (see Fig. 3.6 bottom). The front face of the crystals is at a radius of 1.29 m and each crystal has a square cross-section of 22×22 mm^2 and a length of 230 mm corresponding to 25.8 X$_0$. The truncated pyramid-shaped crystals are mounted in a a semi-projective geometry forming a 3° angle with respect to the nominal interaction point (this avoids that photons from primary interaction fall in the separation zone of two crystals). The crystal cross-section corresponds to $\Delta\eta \times \Delta\phi = 0.0175 \times 0.0175$ (1°). The barrel has a granularity of 360 crystals in the ϕ-direction and 2×85 crystals in the η-direction, resulting in a total number of 61200 crystals. The crystal volume in the barrel amounts to 8.14 m^3 (64.7 t). Crystals for each half-barrel are grouped in 18 supermodules each subtending 20° in ϕ. Each supermodule comprises four modules with 500 crystals in the first module and 400 crystals in each of the remaining three modules.

Two endcaps (EE) are placed at a distance of 3 m from the nominal interaction point, covering up to $\eta = 3$. They are made of crystals with a length of 24.7 X$_0$. The mechanical design of the endcap calorimeter is based on an off-pointing pseudo-projective geometry using tapered crystals of the same shape and dimensions (24.7 $\times$ 24.7 $\times$ 220 mm^3) grouped together into units of 36, referred to as supercrystals. A total of 268 identical supercrystals is used to cover each endcap with a further 64 sectioned supercrystals used to complete the inner and outer perimeter. Each endcap contains 14648 crystals, corresponding to a volume of 1.52 m^3 (12.6 t).

ECAL preshower

The endcaps are equipped with a preshower detector (ES), covering the region $1.7 < \eta < 2.6$. It is a two-layer sampling calorimeter made of lead and silicon strips

detectors. The thickness of the two lead absorbers is respectively 2 X_0 and 1 X_0. The preshower is used to obtain a better spatial resolution in the endcaps in order to separate photons from π^0s.

It is placed in front of the crystals, and contains two lead converters of a total thickness of 2 X_0 and 1 X_0 respectively, followed by detector planes of silicon strips with a pitch of < 2 mm. The impact position of the electromagnetic shower is determined by the center-of-gravity of the deposited energy. The accuracy is typically 300 μm at 50 GeV. In order to correct for the energy deposited in the lead converter, the energy measured in the silicon is used to apply corrections to the energy measurement in the crystal. The fraction of energy deposited in the preshower (tipically 5% at 20 GeV) decreases with increasing incident energy.

Energy resolution

The energy resolution of a homogeneous calorimeter for electrons and photons of energy E can be parameterized as a squared sum of three terms:

$$(\frac{\sigma_E}{E})^2 = (\frac{S}{\sqrt{E}})^2 + (\frac{N}{E})^2 + C^2 \tag{3.6}$$

where E is the energy expressed in GeV.

The stochastic term (S) depends on fluctuations of the number of detected photons, the noise term (N) is the term due to the electronics noise and the constant term (C) depends on lateral containment, non uniformity of response and intercalibration.

The values of these parameters have been measured at a beam test at a single-crystal level and were found to be $S = 2.8\%$ GeV$^{1/2}$, $N = 124$ MeV and $C = 0.3\%$.

The energy resolution for photons with E$_T \simeq 60$ GeV varies between 1.1% and 2.6% over the solid angle of the ECAL barrel, and from 2.2% to 5% in the endcaps. The ECAL energy resolution for electrons with E$_T \simeq 45$ GeV from Z $\to$ ee decays is better than 2% in the central region of the ECAL barrel ($|\eta| < 0.8$), and is between 2% and 5% elsewhere. For low-bremsstrahlung electrons, where 94% or more of their energy is contained within a 3 $\times$ 3 array of crystals, the energy resolution improves to 1.5% for $|\eta| < 0.8$.

3.2.5 Hadronic calorimeter

The hadronic calorimeter (HCAL) [55] is the detector devoted to the energy measurement of hadrons and the missing transverse energy of events. To accomplish this, it has to contain the *showers* generated from hadronic particles fitting inside the magnet coil.

The HCAL is a sampling calorimeter, meaning it finds a particle's position, energy and arrival time using alternating layers of absorber and fluorescent scintillator materials that produce a rapid light pulse when the particle passes through. Special optic fibres collect up this light and feed it into readout boxes where photodetectors amplify the signal. When the amount of light in a given region is summed up over many layers of tiles in depth, called a *tower*, this total amount of light is a measure of a particle's energy. It reaches from the outer surface of the ECAL at 1.77 m to the inner surface of the magnet at 2.95 m radially

To accomplish the most important requirements (hermeticity, good transverse granularity, good containment of the hadron shower), the HCAL is organized into barrel (HB and HO), endcap (HE) and forward (HF) sections:

- There are 36 barrel wedges, each weighing 26 tonnes. These form the last layer of detector inside the magnet coil whilst a few additional layers, the outer barrel (HO), sit outside the coil, ensuring no energy leaks out the back of the HB undetected.

- Similarly, 36 endcap wedges form HE, and measure particle energies as they emerge through the ends of the solenoid magnet.

- Lastly, the two hadronic forward calorimeters (HF) are positioned at either end of CMS, to pick up the myriad particles coming out of the collision region at shallow angles relative to the beam line. These receive the bulk of the particle energy contained in the collision so must be very resistant to radiation and use different materials to the other parts of the HCAL.

A longitudinal view of a quadrant of HCAL is shown in Fig. 3.7. All the HCAL component are visible: the barrel (HB, HO) and endcap (HE) sections, plus the forward detector (HF).
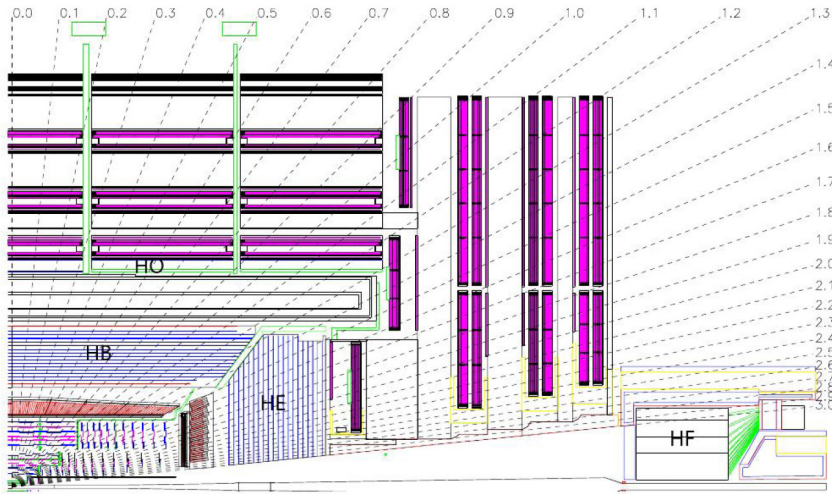
Figure 3.7. Longitudinal view of one quarter of the detector in the rη-plane, showing the positions of the HCAL parts: hadron barrel (HB), hadron outer (HO), hadron endcap (HE) and hadron forward (HF).

The hadron barrel part of the HCAL (HB) covers a region of $|\eta| < 1.4$ and consists of 2304 towers, resulting in a segmentation of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. Each tower is made up of alternating layers of non-magnetic brass absorber and plastic scintillator material. The reason for the absorber material to be non-magnetic is that it must not affect the magnetic field. The two hadron endcaps cover a region of $1.3 < |\eta| < 3.0$. They are positioned in the end parts of the CMS detector and thus are allowed to contain magnetic material. Here iron is used as the absorber

material. The granularity begins from $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ at $|\eta| < 1.6$ up to $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ at $1.6 < |\eta| < 3.0$.

The outer barrel (HO), placed outside of the magnet, has the task to absorb escaping hadron showers from particles with transverse energies above 500 GeV. Without the outer barrel, these particles would cause a large missing transverse energy which is not convenient for many physics analysis purposes. The granularity and η range of outer barrel is the same as the hadron barrel. The forward region of $3.0 < |\eta| < 5.0$ is covered by the steel/quartz fiber hadron forward calorimeter. The whole coverage of the HCAL is almost hermetic; $0 < \phi < 2\pi$ in azimuth and $0 < |\eta| < 5.0$ in pseudorapidity.

The energy resolution of the HCAL is parameterized as

$$\frac{\sigma}{E} = \frac{100\%}{E(GeV)} + 8\% \tag{3.7}$$

It is much worse than ECAL resolution, therefore the energy resolution of the combined calorimeters is dominated by HCAL.
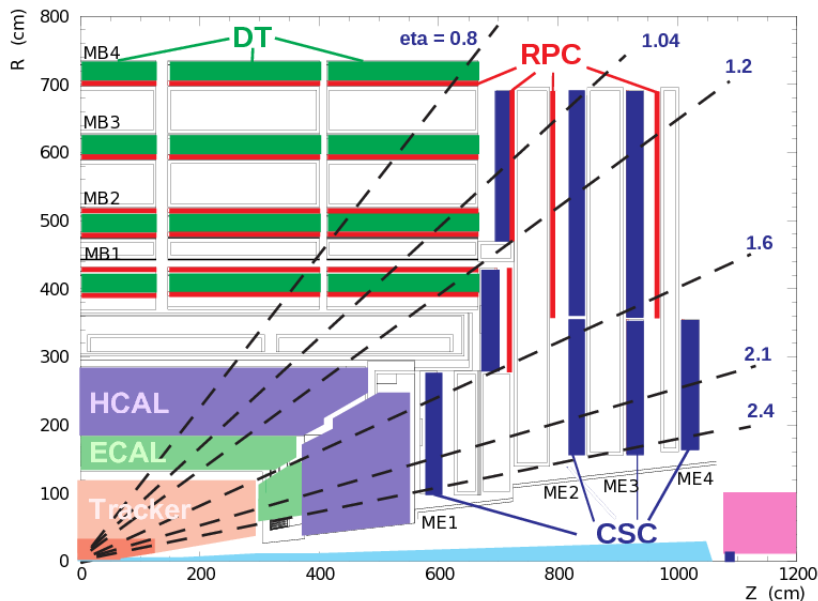
3.2.6   Muon system

The muon system [56] is the outermost of the CMS subdetectors. It has the aim of detecting muons, the only charged particles which are able to pass through the calorimeters without being absorbed. It is placed outside the magnet coil, and it has a pseudorapidity reach of $|\eta| < 2.4$. It is subdivided in a barrel and two endcaps:the barrel covers the region of $0 < |\eta| < 1.2$, and the endcaps the region $1.2 < |\eta| < 2.4$. Both regions are made of four layers of measuring stations, imbedded in the iron of the magnet return yoke, where the return field of the solenoid is about 1.5 T. A sketch of the muon system is shown in Fig. 3.8.

Different experimental techniques in different regions of the detector are used:

- Drift tubes (DT): This kind of detectors is used in the central part of the muon system. Each chamber is made of twelve 4-cm-wide tubes containing a stretched wire within a gas volume.

- Cathode strip chambers (CSC): In the endcap region, where particle multiplicity is higher, arrays of anode wires, crossed with cathode strips, within a gas volume are used for muon detection.

- Resistive Plate Chambers (RPC): Both barrel and endcaps are equipped with this fast gaseous detectors. They consist of two parallel plates separated by a gas volume. Their excellent time resolution (3 ns) makes them suitable to be used also as fast high-efficiency triggers.

3.2.7   Trigger and data acquisition

At the design luminosity of 10^{34} cm^{-2}s^{-1}, each bunch, traveling oppositely inside the beam pipes, encounters the other 40 million times per second and they passes through each other (in other words the bunch crossing frequency is 40 MHz). The average number of proton-proton collision per each bunch crossing is approximately

Figure 3.8. Layout of one quarter of section the CMS muon system.

20. This corresponds to a data acquisition rate which is practically impossible to store in any tape or disk. Moreover, saving all events would not be useful since most of them are soft collisions, not interesting for the CMS physics program.

A "trigger system" is therefore needed in order to rapidly select potentially interesting events. The maximum amount of data that can be recorded is ~ 2 GB/s, that for an event size of ~ 2 MB corresponds to an event rate of 1 kHz. The aim of the trigger system is to lower the event rate from 10^9 Hz to the level of 1 kHz, the maximum rate that can be archived by the on-line computer farm.

The Trigger system is structured in two levels: a Level-1 Trigger (L1) and a high-level trigger (HLT) [57].

- The L1 system is made of a series of hardware processors able to do simple logical operations directly on detector signals. This allows fast decisions (maximum decision time is 3.2 μs) and the rate is reduced to 100 kHz.

- The high level triggers (HLT) is a software system implemented in a multi-processor computer farm. It takes decisions analyzing the reconstructed quantities for a given object and it reduces the rate of an additional factor 10^2, bringing the frequency to about 1 kHz.

To study the physics of processes with very high rate, such that the detector information for *all* the events is impossible to be written on disk and stored, even using the two-levels trigger system described in this section, one possibility is to save only a fraction of events introducing a *prescale*. Prescaled triggers are commonly used in many analyses, by the way they are not the optimal solution for a search, since they collect only a fraction of the data, corresponding to a lower luminosity with respect to the total sample.

The analysis presented in this thesis makes use of data where the *unprescaled* triggers are fully efficient, following the approach used during Run 1 (see Sec. 2.2.1), and starting therefore the search from masses around 1.2 TeV as it is explained in detail in Chapter 5. The possibilities to extend the search at low mass are at least two: one is to use prescaled trigger as ATLAS does (see Sec. 2.2.2), the other is to use the "data scouting" technique, as it was done for Run 1 and briefly described in Sec. 2.2.1.

Level-1 Trigger

The Level-1 is the first trigger that the signal from detectors encounter, and the decision whether taking or discarding data from a particular bunch crossing has to be taken in 3.2 μs; if the L1 accepts the event, the data are moved to be processed by the HLT.

To deal with the high bunch crossing rate, the L1 trigger has to take a decision in a time too short to read data from the whole detector, therefore it employs the calorimetric and muon information only, since the tracker algorithms are too slow for this purpose. The L1 trigger is organized into a Calorimeter Trigger and a Muon Trigger, whose informations are transferred to the Global Trigger which takes the final accept-reject decision. The Calorimeter Trigger is based on trigger towers, 5 × 5 matrices of ECAL crystals, which match the granularity of HCAL cells. The trigger towers are grouped in 4 × 4 squares. The Calorimeter Trigger identifies the best four candidates of each of the following classes: electrons and photons, central jets, forward jets and τ-jets (identified from the shape of the deposited energy). The information of these objects is passed to the Global Trigger, together with the measured calorimetric missing transverse energy. The Muon Trigger is ran separately for each muon detector. The information is then merged and the best four muon candidates are transferred to the Global Trigger. The Global Trigger takes the accept-reject decision exploiting both the characteristic of the single objects and of combinations of them.

HLT Trigger

The High-Level Trigger reduces the output rate to about 1 kHz. It is a highly-customizable software system, in which flexibility is maximized because there is complete freedom in deciding which data to access, as well as the sophistication of the adopted algorithms. The HLT software is organized in a set of algorithms (known as HLT "paths") which are designed to select specific event topologies. The guiding principles are regional reconstruction and fast event veto. Regional reconstruction tries to avoid the complete event reconstruction, which would take time, but rather focuses on the detector regions close to where the L1 trigger has found interesting activity. Fast event veto means that uninteresting events are discarded as soon as possible, therefore freeing the processing power for the next events in line. This has led to the development of three virtual trigger levels: the first level accesses only the muon and calorimetric data, the second level adds the data of the pixel seeds, the final step reads the full event information.

3.2.8   Software

The goals of the CMS software are to process end select events inside the HLT farm, to deliver the processed results to the experimenters within the CMS collaboration and to provide tools for them to analyze the processed information and produce physics results. The overall collection of software, now referred to as `CMSSW`, is built around a Framework, an Event Data Model, and Services needed by the simulation. The physics and utility modules are written by detector groups. The modules can be plugged into the application framework at run time, independently of the computing environment. The software should be developed keeping in mind not only performance but also modularity, flexibility, maintainability, quality assurance and documentation. CMS has adopted an object-oriented development methodology, based primarily on the `C++` programming language.

The primary goal of the CMS Framework and Event Data Model (EDM) is to facilitate the development and deployment of reconstruction and analysis software. The EDM is centered around the Event class, which holds all data that was taken during a triggered physics event as well as all data derived from the data taking (e.g. calibration and alignment constants).

The detailed CMS detector and physics simulation is currently based on the `GEANT` 4 [58] simulation toolkit and the CMS object-oriented framework and event model. `GEANT` 4 provides a rich set of physics processes describing electromagnetic and hadronic interactions in detail. It also provides tools for modeling the full CMS detector and geometry and the interfaces required for retrieving information from particle tracking through these detectors and the magnetic field. The validation of `GEANT` 4 in the context of CMS is described in detail in [59]. The CMS `GEANT` 4-based simulation program uses the standard CMS software framework and utilities, as used by the reconstruction programs. The simulation is implemented for all CMS subdetectors in both the central and forward region, including the field map of the 3.8 T solenoid. In addition, several test-beam prototypes and layouts have been simulated. The full simulation program implements the sensitive detector behavior, track selection mechanism, hit collection and digitization (i.e. detector electronic response). The detailed simulation workflow is as follows:

- a physics group configures an appropriate Monte Carlo event generator (several are used) to produce the data samples of interest;

- the production team/system runs the generator software to produce generator event data files;

- the physics group validates the generator data samples and selects a configuration for the `GEANT`-based simulation of CMS, with generator events as input, to produce (using the standard CMS framework) persistent hits in the detectors;

- the physics group validates these hit data which are then used as input to the subsequent digitization step, allowing for pile-up to be included. This step converts hits into digitizations which correspond to the output of the CMS electronics.

The digitization step, following the hit creation step, constitutes the simulation of the electronic readout used to acquire data by the detector. It starts from the hit positions and simulated energy losses in the detectors and produces an output that needs to be as close as possible to real data coming from CMS. Information from the generation stage (e.g. particle type and momentum) is preserved in the digitization step. The output of this step has the same format of real collision events, and therefore can be fed to the same reconstruction software chain.

Collision events are reconstructed and stored if they satisfy at least one of the High Level Trigger paths employed online. Depending on the type of HLT path which was fired, an event is stored in a given Primary Dataset (PD), which will therefore collect events with similar topologies. One example of Primary Dataset is the `JetHT`, used for the dijet analysis. It is defined by hadronic triggers that require of a threshold on the scalar sum of transverse momentum of all the jets in the event (H_T).

Chapter 4

Jet recostruction and calibration at CMS

This chapter is dedicated to the jet reconstruction in CMS. What we call "jet" is a spray of particles originated by a quark or glob that, having color charge, need to recombine in hadrons. The process of hadronization and the interaction with the detector material, as we have seen in Chapter 1, generates many particles that lose their energy in a cascade of decays inside the CMS calorimeters. As described in detail in Chapter 3, the electromagnetic and hadronic calorimeters can measure the energy released by the particles in the shower. Charged particles are also detected by the silicon tracker that provides a measurement of their transverse momentum.

The measurements of the various CMS sub-detectors go through the Particle Flow (PF) reconstruction algorithm, that makes use of all those informations in order to identify the particles. The PF is described in Sec. 4.1. Since jets are composite objects they are not defined in a unique way: the PF candidates are used as inputs for the jet clustering algorithms, which are discussed in Sec. 4.2.

Then the energy of the final reconstructed jets has to be corrected for different effects: one is the pile-up, because the energy coming from the non-primary vertices enters in the reconstruction cone, and is wrongly associated to the jet in raw reconstruction. Other effects that need a correction are the non-uniformity of the jet energy response across the detector, and the *non-compensating* nature of the calorimeters. For a compensating calorimeter [60] the response of electromagnetic-interacting particles (let's call it R_γ) and hadronic particles (R_h) are the same : $R_h = R_\gamma$. The CMS calorimeters are instead strongly non-compensating: the CMS HCAL has been measured [61] to have $R_h/R_\gamma \approx 0.7$, and for the crystal ECAL lower values, of the order of 0.45-0.5, are assumed. The jet energy calibration is crucial for the dijet analysis and will be discussed in detail in Sec. 4.3.

4.1 Particle Flow reconstruction

The Particle Flow [62] is a full event reconstruction technique which aims to reconstruct all stable particles produced in a given proton-proton collision. To do so it exploits all CMS subdetectors to their full granularity and correlates information between them in order to optimize particle reconstruction and identification

performance.

The design of the CMS detector proves to be well-suited for this type of event reconstruction: its large silicon tracker and the 3.8 T magnetic field in which it is immersed allow precise and efficient charged particle detection for transverse momenta as low as 150 MeV, and its crystal electromagnetic calorimeter allows excellent resolution in the measurement of photon and electron energies. Charged particles and photons make up on average about 85% of a jet's energy, so only 15% of it will be reconstructed in the hadronic calorimeter alone.

The Particle Flow algorithm first collects reconstructed hits in each subdetector independently and creates a list of basic reconstructed elements (blocks), namely charged tracks in the tracker, clusters of energy deposits in the calorimeters. Once blocks are formed, a link algorithm connects blocks which are topologically compatible, giving way to particle flow particle candidates (PFCandidates). PFCandidates may be of different types, depending on the type of blocks involved in their reconstruction:

- **electrons** arise from the link between a charged track and one or more ECAL clusters, provided an electron identification set of criteria is satisfied;

- charged tracks linked to any number of calorimeter (ECAL or HCAL) clusters, and which are not identified as electrons, are reconstructed as **charged hadron** candidates;

- ECAL energy deposits not compatible with charged tracks give way to **photon** candidates;

- energy deposits in the HF calorimeters are reconstructed as **HF hadronic** or **electromagnetic** particle candidates, depending on the depth at which the energy is released in the HF quartz fibres;

- finally **muons** are reconstructed and identified with very large efficiency and purity from a combination of the tracker and muon chamber information.

The formation of the PFCandidate list represents the Particle Flow interpretation of a given proton-proton collision in CMS, as it attempts to mirror the true particle composition of the event to the best of our knowledge. Particle flow jet reconstruction (PFJets) is then just a matter of choosing the jet algorithm with which the PFCandidates are to be clustered, as it is discussed in the following Sec. 4.2.

The PF jet momentum and spatial resolution are greatly improved with respect to calorimeter jets,as the use of the tracking detectors and high granularity of ECAL allow resolution and measurement of charged hadron and photons inside a jet, which together constitute $\sim 85\%$ of the jet energy.

But the non uniformity of the jet response as a function of p_T and η makes necessary the use of jet energy corrections, as described in Sec. 4.3.

4.2 Jet clustering algorithms

We have seen in Chapter 1 that scattered partons from the hard subprocess eventually turn into a spray of hadrons due to color confinement. This spray of particles

can be identified as jet objects in the detector through the application of a clustering algorithm using PFCandidates as inputs.

What we call "jet" therefore is not defined in a unique way, but depends on the chosen algorithm used to recluster the PFCandidates. In the ideal case the four-vector associated to the jet object is exactly identical to the parton that generated the shower. Of course it is true with some approximation, but there are some theoretical and experimental criteria that a "good" clustering algorithm should have:

- be simple to implement in an experimental analysis;

- be simple to implement in the theoretical calculation;

- be defined at any order of perturbation theory;

- yield finite cross sections at any order of perturbation theory;

- yield a cross section that is relatively insensitive to hadronization;

- satisfy infrared and collinear safety.

Infrared and collinear (IRC) safety are fundamental requirements for jet algorithms. "Infrared safety" means that adding a soft gluon should not change the results of the jet clustering. An algorithm is "collinear safe" if, splitting one parton into two partons (e.g. a gluon splitting in two quarks), the results of the jet clustering does not change. A sketch showing the effects of non-infrared and non-collinear safe algorithm is represented in Fig. 4.1.

A whole family of infrared and collinear-safe algorithms are the k_T-algorithms [63]. The generalized k_T-algorithm is based on a pair-wise recombination and it combines two particles if their *distance* (opportunely defined as a function of the transverse momentum (k_T) is less than a given threshold. They work as follows:

- define the "distance" between two input particles $d_{ij} = \min(k_{Ti}^{2p}, k_{Tj}^{2p})\Delta R_{ij}^2/R^2$, with $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$.

- define the "distance" d_{iB} between the particle i and the beams $d_{iB} = k_{Ti}^{2p}$

- find the minimum of d_{ij} and d_{iB}

- if the minimum is d_{iB}, remove particle i from the list and call it a "jet"; if it's d_{ij} , recombine particles i and j into a new particle by summing their quadrimomenta;

- iterate until only jets are left.

Depending on the value of p it is possible to obtain the anti-k_T algorithm ($p = -1$), the Cambridge-Aachen ($p = 0$) and the k_T ($p = 1$). With the k_T algorithm the softer particles are clustered first, and then added the others in increasing energy. With the anti-k_T algorithm hard particles are clustered initially instead than soft particles. The Cambridge-Aachen algorithm, having $p = 0$, is a purely geometric clustering algorithm (no dependence from the transverse momentum k_T).
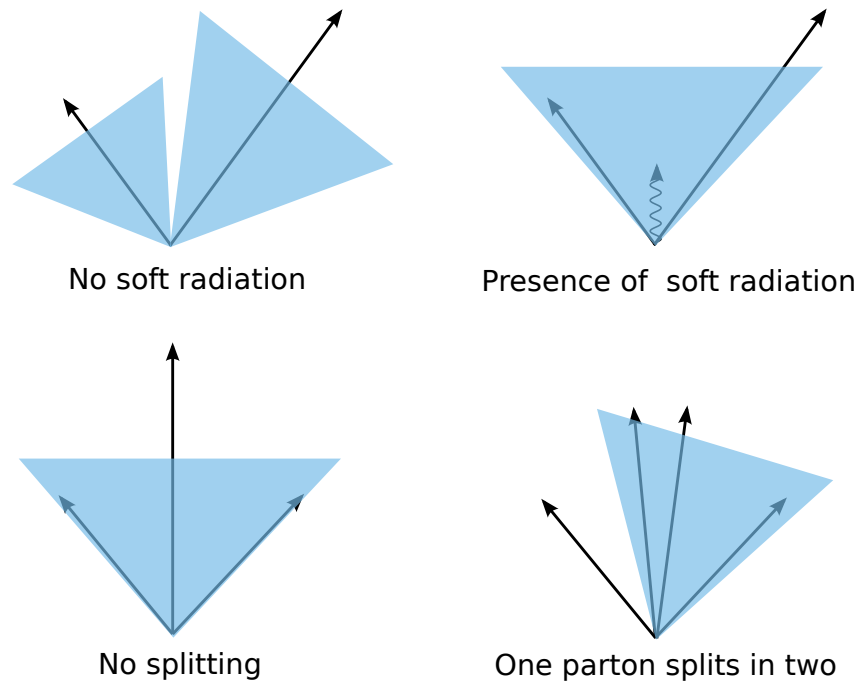
Figure 4.1. Illustration of the infrared sensitivity of a cursory designed jet algorithm: the emission of a soft gluon changes the number of jets (top). Illustration of the product of a collinear unsafe jet algorithm: the collinear splitting changes the direction of the jet (bottom).
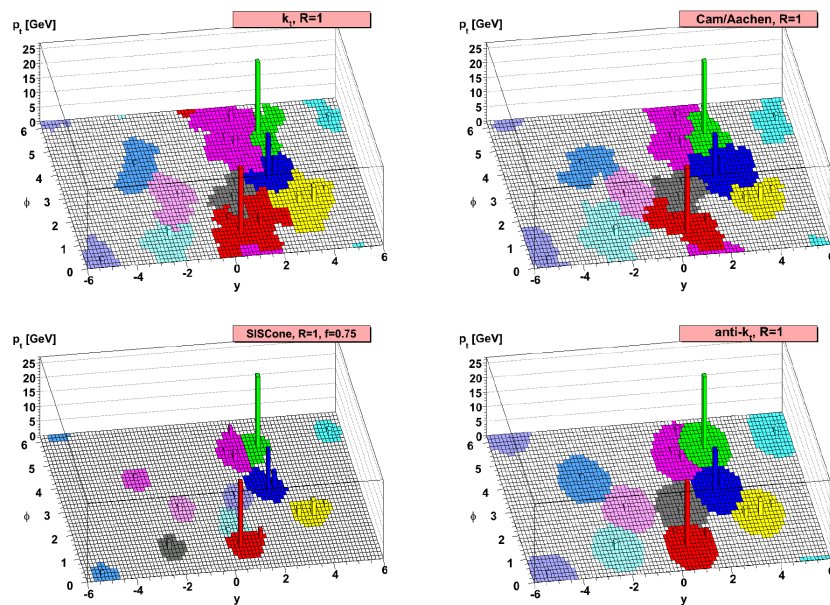
Figure 4.2. Illustration of different jet algorithms at parton level.

The behaviors of different jet algorithms are illustrated in Figure 4.2: the anti-k_T jet algorithm gives the jets with the most definite shape.

The jets used in this analysis are reconstructed with the anti-k_T clustering algorithm with size parameter $R = 0.4$ in the $\eta - \phi$ plane (PF ak4 jets). The anti-k_T algorithm is interfaced to the CMS software framework through the FastJet package. At generator level, the list of final state particles produced in the hadronization process constitute the list of objects that will be clustered; at reconstruction level a list of particle candidates is passed to the algorithm, produced by the full event reconstruction technique known as the CMS Particle Flow, which is illustrated in Sec. 4.1.

The additional pp collisions occurring within the same bunch crossings of the primary hard interaction produce additional tracks in the tracker and deposit energy in the calorimeters. This contribution is called *in-time pile-up* (IT PU), as we have seen also in Sec. 1.2.1. The IT PU from charged particles is reduced by identifying which vertex the charged PFCandidates originate from and removing those unambiguously associated to pile-up vertices before clustering jets. This method is referred to as *charged hadron subtraction* (CHS).

The pp collisions occurring in the previous and subsequent bunch crossings also contribute calorimetric energy to the same time window a the primary interaction due to the finite signal decay time in the calorimeters. This contribution is called *out-of-time pile-up* (OOT PU).

The amount of OOT PU can be reduced at the detector level shortening the signal time integration window and using more advanced techniques that use the signal timing and pulse shape to subtract a varying pedestal and fit in-time and out-of-time pulses simultaneously. Both ECAL and HCAL use pulse fitting to remove OOT PU.

The leading primary vertex (PV) is chosen based on the largest sum of squares of the track transverse momenta ($\sum |p_T^{\mathrm{track}}|^2$) associated to the vertex. Subleading PVs are classified as pile-up vertices. The CHS can remove approximately 50% of IT PU, within tracker coverage.
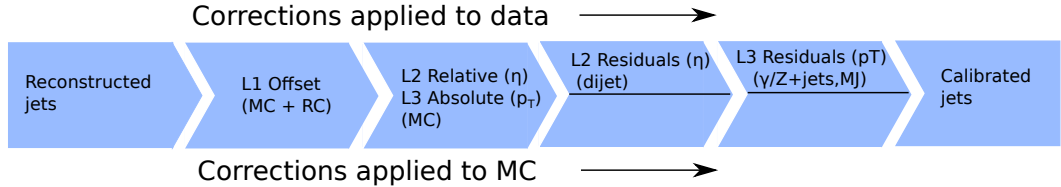
4.3 Jet energy calibration

The jet energy correction (JEC) scheme employed at CMS is based on a factorized approach [64]. Different sets of corrections are defined, in such a way that they address different physical aspects. They are derived in a consecutive way, and need to be applied in the correct order, since in principle they could not be completely uncorrelated.

The levels of corrections defined in CMS are:

- **Level 1 offset corrections**, which corrects jets for the effect of overlapping diproton collisions (pile-up);

- **Level 2 relative corrections**, which minimizes the effect of non-uniformities between different CMS subdetectors;

- **Level 3 absolute corrections**, which addresses the fundamental *non-compensating* nature of the CMS calorimetric system.

- **L2L3 residual corrections**, which correct the residual difference between data and simulation after the application of L1L2L3.

Corrections applied to data ⟶

| Reconstructed jets | L1 Offset (MC + RC) | L2 Relative (η) L3 Absolute (p_T) (MC) | L2 Residuals (η) (dijet) | L3 Residuals (pT) (γ/Z+jets,MJ) | Calibrated jets |

Corrections applied to MC ⟶

Figure 4.3. Scheme of the JEC method in CMS. Starting from reconstructed raw jets, it is shown the sequence of corrections.

A scheme that summarize the jet energy correction technique in CMS is in Fig. 4.3.

First level are L1 offset corrections, which are derived from MC and applied to both MC and data. For the data, the L1 corrections contain an additional scale factor derived from data driven techniques. The L1, L2 and L3 corrections are derived from MC, represent the bulk of the JECs and are applied both to data and MC. The benefit of relying heavily on simulation to derive the jet energy response is that we are not sensitive to many of the biases intrinsic in the data-based methods and can cover corners of phase space that are not easily accessible in data. A more detailed description of MC corrections, and some recent results can be found in Sec. 4.3.1.

After these corrections are applied, the comparison between data and simulation shows a residual disagreement. Therefore an additional set of "Residual corrections" to be applied only to data is needed, in order to obtain a good data-MC closure. These corrections are derived using different techniques and different datasets (dijet, multijets, photon + jets, Z + jets). The results are then combined with a global fit, to extract the final residual correction factors as a function of p_T and η.

The residual corrections are also referred to as L2 relative and L3 absolute, since, like L2 and L3 MC corrections, they aim to correct respectively the uniformity of response in the detector and the absolute energy scale. More details on residual JEC are in Sec. 4.3.2.

4.3.1   Corrections from simulation

Level 1 Offset corrections

The most important in terms of "size" of the corrections are the pile-up corrections. The aim of these correction is to subtract the additional energy which is irradiated inside a jet cone by secondary proton-proton collisions (the pile-up), on an event-by-event basis. The method used in CMS subtracts also the additional energy coming from the remnants of the colliding protons, what is called *underlying event* (UE).

In order to do so, a technique called *jet area method* is employed. This algorithm uses the effective area of the jets multiplied by the average energy density to calculate the offset energy to be subtracted by the jets.

In addition to the original method that used only the jet area A_j and the energy density ρ, η-dependent and $\log(p_T)$-dependent terms have been added to the correction to model better the pile-up and underlying event contribution in different parts of the detector and for different jet energies. The full correction formula used at CMS is

$$C_{L1}(p_T^{\mathrm{uncorr}}, \eta, A_j\rho) = 1 - \frac{[\rho_0(\eta) + \rho \cdot \beta(\eta) \cdot (1 + \gamma(\eta) \cdot \log(p_T^{\mathrm{uncorr}}))] \cdot A_j}{p_T^{\mathrm{uncorr}}} \quad (4.1)$$

The input parameters are the uncorrected reconstructed jet transverse momentum p_T^{uncorr}, jet pseudorapidity η, jet area A_j and the per-event p_T offset density ρ.

The correction for the underlying event density is not explicitly present in Eq. 4.1 but it is effectively absorbed in $\rho_0(\eta)$. The multiplicative factor $\beta(\eta)$ corrects for the non-uniformity of the offset versus η and the residual correction factor $\gamma(\eta)$ adds the logarithmic dependence from jet p_T.

The parameters $\rho_0(\eta)$, $\beta(\eta)$ and $\gamma(\eta)$ are determined from simulated true offset. The "effective" energy density in the square brackets in Eq. 4.1 is then multiplied by the jet *active area* (equal to πR^2 in the case of anti-k_T jets) and the result is subtracted from the energy of the jet.

The pile-up offset (in GeV units) before and after L1-Fastjet corrections for the PF ak4 jets with CHS is shown in Fig. 4.4.
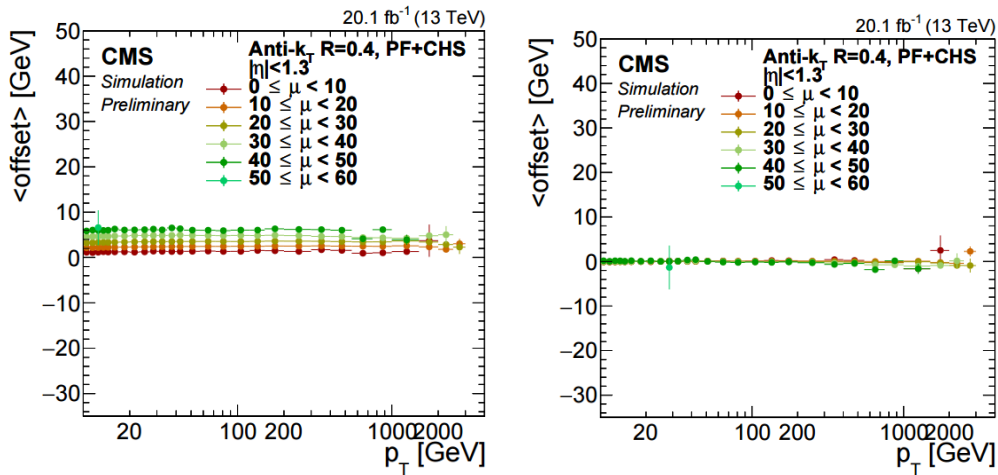
Figure 4.4. Simulated pile-up offset for PF ak4 jets with CHS in the barrel ($|\eta| < 1.3$) before (left) and after (right) L1 corrections.

Then a scale factor data/simulation is added for the corrections to be applied to data. This scale factor is obtained building a *Random Cone* centered at (η,ϕ) and dividing the average energy density inside the cone in data (using a *Zero Bias* sample, that has no energy deposition from hard interactions) by the true average offset in simulation (Eq. 4.2).

$$SF = \frac{\mathcal{O}_{\mathrm{data}}^{RC}(\eta, \langle\rho\rangle_{\mathrm{data}})}{\mathcal{O}_{\mathrm{MC}}^{RC}(\eta, \langle\rho\rangle_{\mathrm{MC}})} \quad (4.2)$$

Level 2 relative and Level 3 absolute corrections

The simulated true response is defined as the ratio of arithmetic means of matched reconstructed and generated jets

$$R(p_{T,\text{ptcl}},\eta) = \frac{\langle p_{T,\text{reco}}\rangle}{\langle p_{T,\text{ptcl}}\rangle}[p_{T,\text{ptcl}},\eta] \tag{4.3}$$

binned in particle level $p_{T,\text{ptcl}}$ and reconstructed η. The jet matching is done by radial distance, with maximum at half the con size ($DeltaR = 0.2$ for $R = 0.4$). This ensures highest matching efficiency with unique matches for anti-k_T jets.

The corrections are derived first versus η, to uniform the response in the detector: they are called Level 2 *relative* corrections (L2). Then, after the application of L2 corrections, Level 3 *absolute* corrections (L3) are derived versus jet p_T.

The correction factors of MC corrections are shown in Fig. 4.5. The closure, i.e. the response after the application of L1+L2+L3 corrections for anti-k_T jets with $R = 0.4$ and CHS is shown in Fig. 4.6.
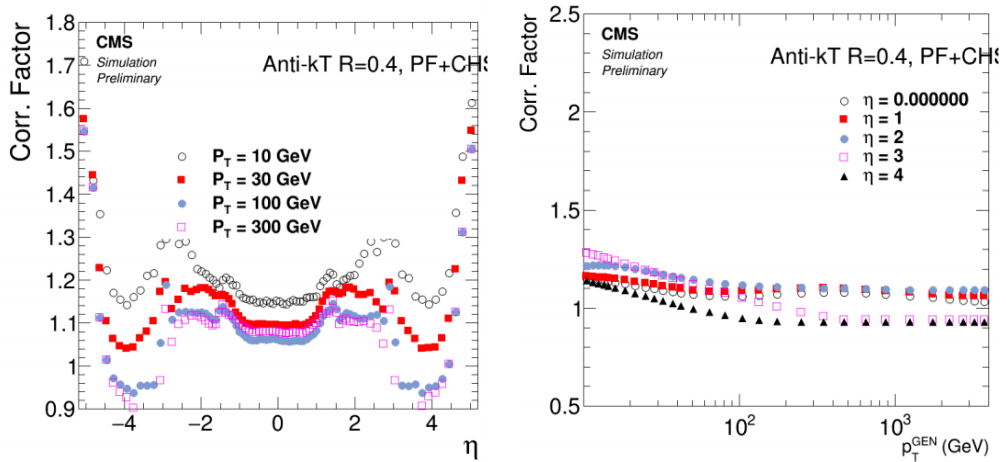

Figure 4.5. Relative (left) and absolute (right) jet energy corrections as a function of pseudorapidity.

4.3.2   Corrections from data

The residual data/simulation scale factors for JEC are determined after correcting jets for pile-up and simulated true response, and propagating the corrections to the missing-E_T. These corrections are first determined with a high statistics dijet sample relative to $|\eta| < 1.3$ over a wide range of p_T: this gives a *relative* scale factor, and therefore represents the L2 relative residual corrections, as a function of jet pseudorapidity.

The absolute scale (L3 absolute residual corrections) is obtained from a combination of Z$\to \mu\mu$+jet, Z$\to ee$+jet, γ+jet and multijet events at $\eta < 1.3$ in a more limited range of p_T from about 30 GeV and 1 TeV.
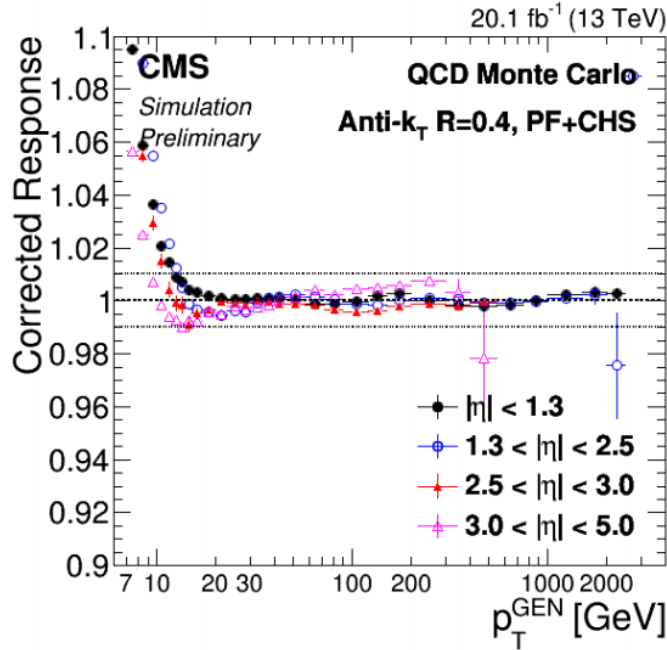
Figure 4.6. Corrected jet response versus jet p_T for anti-k_T jets with $R = 0.4$ and CHS.

It is beyond the purpose of this thesis to give a detailed description of all these methods, since to all effects, each of these techniques is a stand-alone analysis. For a more complete treatment of the jet energy corrections derivation in CMS the reader can refer to [64]. The γ+jets method will be instead described in detail since was part of my PhD activity in the CMS collaboration. A dedicated section with recent results can be found in App. A.

The residual JECs have turned out to be very important in particular in the first phase of data taking of LHC Run 2. In fact, some reconstruction issues in the calorimeters made the data/MC ratio of responses p_T dependent with a non negligible slope.This is visible in Fig. 4.7, where the global fit to the results of the various analysis that contribute to the L3 resudual corrections is shown for central jets ($|\eta| < 1.3$). The result is compared to the global fit obtained at the end of Run 1, that is more flat and closer to 1.

4.3.3 Jet energy scale uncertainties

The systematic uncertainties for the relative η-dependent and the absolute p_T-dependent corrections are summarized in Fig. 4.8.

The jets relevant for this analysis are restricted in the central region $\eta < 2.5$ therefore, from Fig. 4.8 left, the uncertainty on the L2 Residual correction is less than 4%. The requirement of the dijet mass to be greater than 1.2 TeV (see Sec. 5.3.2) constrain also the jet p_T to be greater than a few 100 GeV. From the right figure, the uncertainty of the L3 Residual correction is below 2% in all the spectrum. It must be noted that in the region $p_T > 1$ TeV there are no data avail-
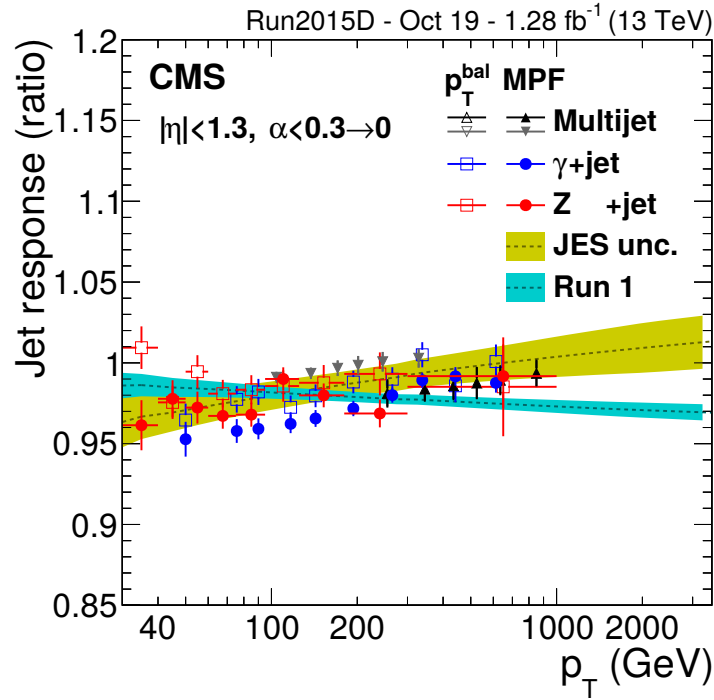
Figure 4.7. Global fit to the jet response (line with yellow band) using the contribution of gamma+jets (blue markers), Z+jets (red markers), and multijets (black markers). The line with light blue uncertainty band shows the Run 1 global fit, for comparison.
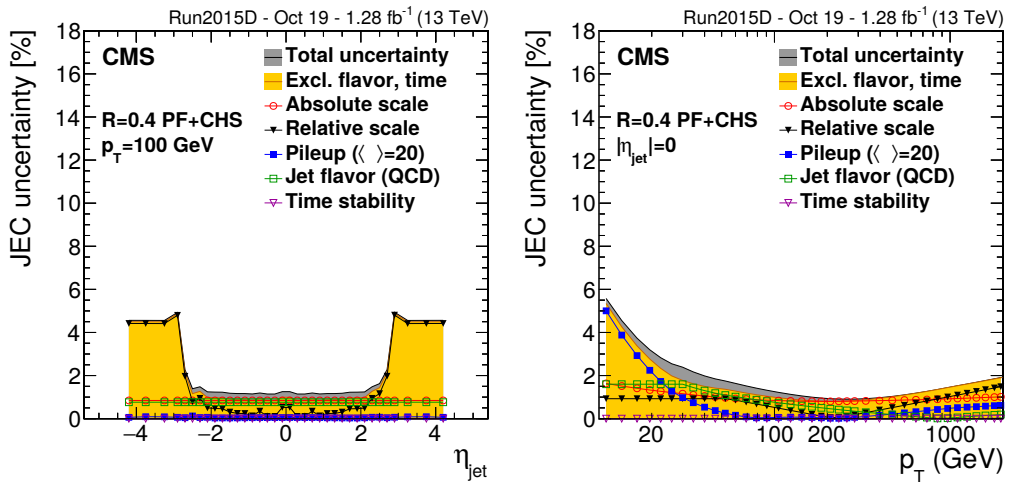
Figure 4.8. The uncertainties of L2 Residuals corrections versus η at $p_T = 100$ GeV (left), and L3 Residual corrections versus p_T at $\eta = 0$ (right).

able in the γ+jets sample, then both the JECs and the uncertainties in the high mass region are obtained with an extrapolation.

Chapter 5

Data sample and selection criteria

5.1 Dataset and trigger selection

The data presented in this thesis have been collected mostly on September and October 2015, with the LHC operating at the bunch spacing of 25 ns.

The most of high dijet mass events are collected in the primary dataset `JetHT` and this is the one used for this search. The dataset is splitted in different periods of the data taking and the sub-datasets with the relative integrated luminosity are summarized in Table 5.1

Table 5.1. JetHT sub-datasets and their integrated luminosity

Dataset name	Run range	Integrated luminosity (pb^{-1})
`JetHT Run2015D-PromptReco-v3`	256630-258158	887.308
`JetHT Run2015D-PromptReco-v4`	258159-260627	1557.399

The trigger strategy is similar to the one used in the dijet analysis at $\sqrt{s} = 8$ TeV. The unprescaled triggers that collects most of the events at high dijet mass, interesting for this analysis are the `PFHT800` and `PFJet500` paths, based respectively on the scalar sum of transverse momentum on all jets in the event (H_T), and on the transverse momentum of the leading jet (p_T). This are basic triggers used for many searches of physics beyond Standard Model in hadronic final states in CMS. Another trigger is used to perform efficiency study, as described below. The triggers used in this studies, with the corresponding selection and their prescales, are summarized in Tab. 5.2.

The study of the trigger efficiency curve is important to decide where we can start to fit the dijet mass distribution in data, without having distortions in the low mass region due to the trigger turn on. It is also important to check if there is any failure at high mass, in order not to lose any interesting dijet event.

The trigger system, as explained in Sec.3.2.7, is made of two levels: Level 1 (L1) and High Level Trigger (HLT), the total efficiency is the product of two efficiencies

| trigger name	definition	Prescale
HLT_PFHT800_v*	$H_T > 800$ GeV	1
HLT_PFJet500_v*	$p_T(\text{jet}) > 500$ GeV	1
HLT_Mu45Eta2p1_v*	$p_T(\mu) > 45$ GeV and $\eta < 2.1$	1

Table 5.2. Triggers and their prescales.

$\epsilon_{\text{tot}} = \epsilon_{\text{L1}} \cdot \epsilon_{\text{HLT}}$. To measure the total ϵ_{tot} one can study the L1 (ϵ_{HLT}) and HLT efficiencies (ϵ_{L1}) separately, or directly the total, using a reference trigger that is uncorrelated with the selection of the H_T trigger and uses a different L1 seed. Lepton triggers, for example, are good candidates for this scope: the request of one lepton is completely independent from the jets in the event, and also the L1 seed is not correlated.

The trigger efficiency has been studied in `SingleMuon` dataset and the reference trigger used is `HLT_Mu45Eta2p1` that requires one muon with $p_T > 45$ GeV and $\eta < 2.1$. We calculate the trigger efficiency as a function of the dijet mass using

- as denominator, the dijet mass distribution of events that pass the dijet standard selection (see Sec. 5.3.2) and the reference trigger `HLT_Mu45Eta2p1_v*`;

- as numerator, the dijet mass distribution of events that satisfy the denominator requirements *and* the `PFHT800` ⊕ `PFJet500` trigger requirements (where we have represented the logical operator *or* between the two sets with ⊕).

The turn on curve as a function of m_{jj}^{Wide} is calculated using the histograms in Fig. 5.1, and it is shown in Fig. 5.2.
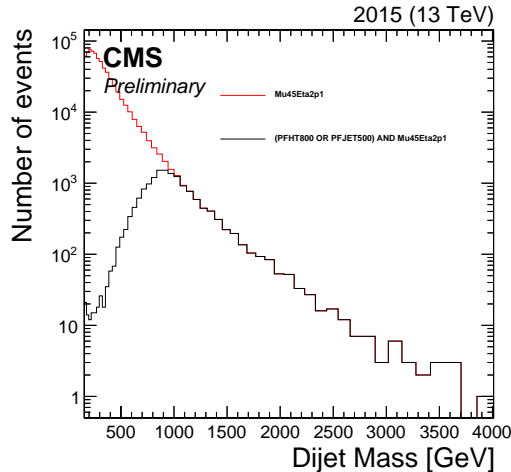
Figure 5.1. The m_{jj}^{Wide} distributions of events in data passing the full dijet event selection plus the trigger selection. All events are required to pass at least the `HLT_Mu45Eta2p1` monitoring trigger.
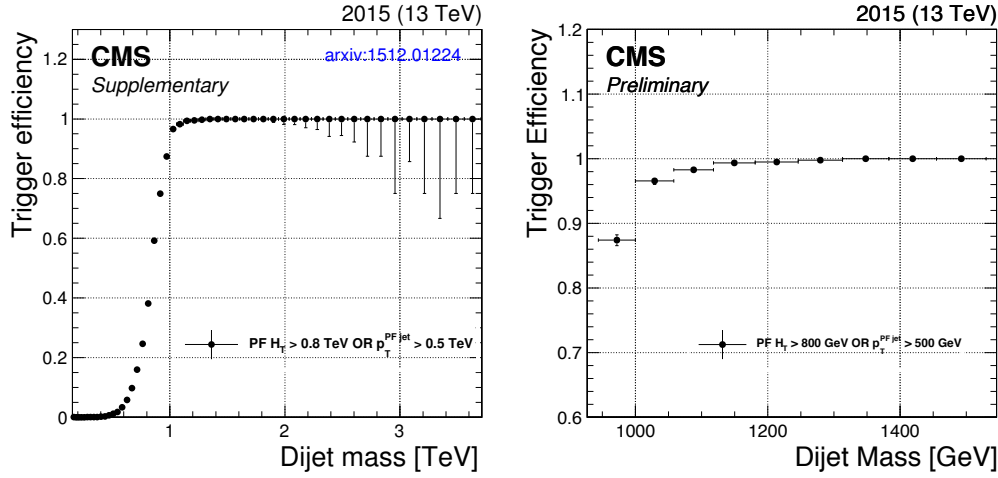
Figure 5.2. Trigger absolute efficiency as function of m_{jj}^{Wide}. A zoomed version is shown on the right.

The turn on curve of the trigger efficiency is complete for m_{jj}^{Wide} greater than about 1.2 TeV, therefore the fit to the dijet mass distribution for the background estimate can start around this point. This is effectively the lowest energy that this analysis can explore to search new physics in dijet final state at center-of-mass energy of 13 TeV. As we will see later in Chapt. 7, the calculation of upper limits on the cross sections starts from resonance masses of 1.5 TeV, in order to use the complete shape in the fit, and avoid biases that could arise fitting a truncated shape next to the low mass boundary.

5.2 Monte Carlo simulation

5.2.1 Signal

We consider different possible sources of narrow resonance signal, meaning with *narrow* that the relative resonance width is much smaller than the experimental mass resolution [37]. It is known from previous studies that the shape of a narrow resonance is mainly determined by the combination of partons in the initial and final states.

There are different combinations of the possible production and decay modes, and in the analysis we used MC samples for:

- gg → RS Graviton → gg

- qq → RS Graviton → qq

- qg → excited quark → qg

where we indicated the gluons with "g" and any quark/antiquark with "q". For each of these processes we have 9 samples at 9 different mass points from 1 to 9 TeV, with about 100k events for each sample.

The Monte Carlo simulation (MC) signal samples are generated with `PYTHIA` 8 [65] using `CUETP8M1` tune and they are produced with the following LHC conditions: 25 ns bunch-crossing spacing and average number of pile-up interactions of 20. In addition, the shapes are passed through a complete simulation of the CMS detector based on `GEANT4`.

5.2.2   Background

The MC sample used for background studies is the QCD multijet production generated with `PYTHIA` 8 with `CUETP8M1` tune. The samples are binned in the transverse momentum of the hard scattering process $\hat{p}_T$ and they are produced using bunch spacing at 25 ns and average number of pile-up interactions equal to 20. The simulated events also pass through a complete `GEANT4` simulation of the CMS detector.

The list of datasets for background with the corresponding cross section, number of events, and equivalent integrated luminosity is reported in Tab. 5.3.

dataset $\hat{p}_T$-binned (GeV)	cross section (pb)	number of events	equivalent lumi (fb^{-1})
QCD $300 < \hat{p}_T < 470$	7823	2930578	0.375
QCD $470 < \hat{p}_T < 600$	648.2	1939229	2.991
QCD $600 < \hat{p}_T < 800$	186.9	1890256	10.11
QCD $800 < \hat{p}_T < 1000$	32.293	1911296	59.19
QCD $1000 < \hat{p}_T < 1400$	9.4183	1461216	155.1
QCD $1400 < \hat{p}_T < 1800$	0.84265	197959	234.9
QCD $1800 < \hat{p}_T < 2400$	0.114943	194924	$1.696 \cdot 10^3$
QCD $2400 < \hat{p}_T < 3200$	0.0068298	198383	$29.05 \cdot 10^3$
QCD $\hat{p}_T > 3200$	0.000165445	188696	$1140 \cdot 10^3$

Table 5.3. Leading order cross sections for Monte Carlo datasets of QCD multijet processes generated with `PYTHIA8`, tune `CUETP8M1`.

5.3   Selection criteria

5.3.1   Jets selection and wide jet algorithm

The dijet analysis uses Particle Flow jets with anti-k_T algorithm and width parameter $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ (PF ak4 jets), that have been discussed in detail in Chapt.4. As we have seen, these definition satisfies the requirements of infrared and collinear safety, and the jet energy is corrected using MC and data-based techniques in order to take into account the pile-up extra energy, the non-uniformity of the response across the detector and the residual difference in the absolute scale of the energy between data and MC.

CMS has developed jet quality criteria ("Jet ID") for PF jets which are found to retain the vast majority of real jets in the simulation while rejecting most fake jets arising from calorimeter and/or readout electronics noise. These are studied in pure noise non-collision data samples such as cosmic trigger data or data from triggers on empty bunches during LHC operation. Jets used in the analysis are required to satisfy the "Tight" PF Jet ID criteria:

- neutral hadron fraction less than 0.9

- neutral electromagnetic energy fraction less than 0.9

- muon fraction < 0.8

- number of constituents > 1

and in addition for jets in $-2.4 \geq \eta \geq 2.4$ requirements that involve the tracks information

- charged hadron fraction > 0

- charged multiplicity > 0

- charged electromagnetic fraction < 0.9

This criteria have an efficiency of 99% and a background rejection of 84%, according to studies on early data at 13 TeV in the Minimum-Bias dataset.

Events are rejected if at least one of the two leading jets does not pass the tight jet ID, or if it is outside of the tracker acceptance: $|\eta| < 2.5$. Events are also rejected if at least one of the two leading jets has Muon Energy Fraction greater than 0.8. This cut was introduced in the run 1 analysis to remove mis-reconstructed muons identified as jets and causing large nonphysical tails in the dijet mass distribution. The inefficiency of this cut for the resonance signal samples under study is found to be negligible.

The leading (second leading) jet is required to have a minimum $p_T > 60$ (30) GeV and to be in the tracker coverage region $|\eta| < 2.5$. The two leading PF ak4 jets are required to to have $p_T > 30$ GeV, $|\eta| < 2.5$ and to pass the tight PF Jet ID criteria.

The dijet analysis choice, as in the past, is to recluster in a larger cone the corrected PF ak4 jets that pass the selection described above, and use *wide jets* to reconstruct the invariant mass of the dijet system. This allows to contain better the energy of the hadrons in presence of final state radiation (FSR), and thus improves the dijet mass resolution with the resonance peak resulting both closer to the nominal mass and narrower. An illustration of the wide jet is in Fig. 5.3. The improvement of the dijet resolution is shown clearly in Fig.5.4, where we compare the dijet mass reconstructed with the PF jets 0.4 and the wide jets used in this analysis ($\Delta R = 1.1$), for an excited quark resonance with mass $M_{q*} = 4.0$ TeV that decays in one quark and one gluon.

The wide jets are obtained using the PF ak4 jets as inputs, following this procedure:

- take the two PF ak4 jets with the highest transverse momentum (leading jets) as "seeds" of the wide jets, forming two "temporary" wide jets wj$_1$ and wj$_2$;

- for each PF ak4 jet j_i, see if it is geometrically closer to the first (wj$_1$) or second temporary wide jet (wj$_2$), comparing $\Delta R_{i1,2}$ between j_i and wj$_{1,2}$;

- if $\Delta R_{i1} < \Delta R_{i2}$ and $\Delta R_{i1} < 1.1$ sum the j_i to the first temporary wide jet wj$_1$;

- else if $\Delta R_{i2} < \Delta R_{i1}$ and $\Delta R_{i2} < 1.1$ sum the j_i to the second temporary wide jet wj$_2$;
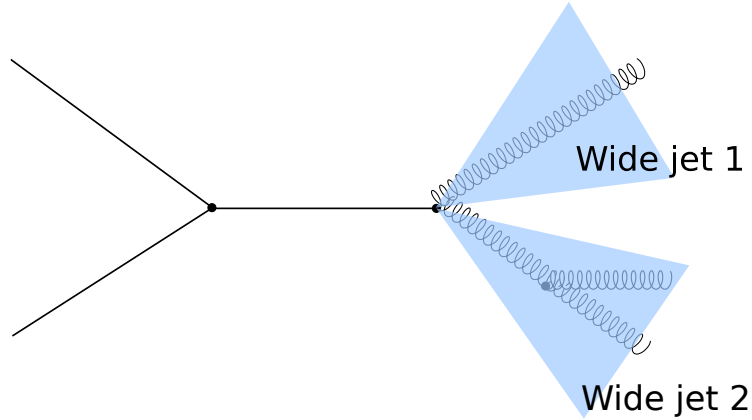
Figure 5.3. Illustration that show an event with final state radiation, where the use of the wide jets allows to contains the radiated energy and improves the dijet mass reconstruction.

At the end of this iterative clustering algorithm, the two final wide jets are reordered in p_T so that the first is the most energetic one: $p_T(\text{WJ}_1) > p_T(\text{WJ}_2)$. The resulting wide jets are by construction infrared and collinear safe and have the the proper jet energy corrections, as the parent anti-jets. In addition, the selection in p_T on the input jets acts like a sort of "pruning" algorithm that reduces the contribute of softer jets and pile-up.

In the phase of the analysis preparation, in 2014 and early 2015, we have studied the optimization of the cone width in order to minimize the expected upper limits on the cross sections, and the value $\Delta R = 1.1$ is found to be optimal.

5.3.2   Event Selection

Events are selected requiring at least one reconstructed vertex within $|z| < 24$ cm around the nominal interaction point. The most relevant selection criteria are:

  i a threshold in the dijet mass calculated using wide jets $m_{\text{jj}}^{Wide} > 1181 GeV$.

 ii a cut on the angular separation between the two wide jets $|\Delta\eta_{\text{jj}}^{Wide}| < 1.3$

The reason of (i) is that, as we have seen in Sec. 5.1, the trigger turn-on curve is complete around 1.2 TeV. This requirement assures that the dijet mass spectrum is not distorted by trigger inefficiencies close to the low mass edge. The specific value of 1181 GeV is used for the threshold corresponds to the low edge of the bin around 1.2 TeV, in the variable-size binning used in this analysis. The binning has been fixed in the previous dijet analysis in CMS and the bin size corresponds roughly to the experimental resolution at that mass.

The requirement (ii) is a cut on the $|\Delta\eta_{\text{jj}}^{Wide}|$ between the jets. This quantity is related to the emission angle of the final partons with respect to the beam line in the center-of-mass reference frame (the scattering angle θ^*):

$$\cos\theta^* = \tanh(\frac{\Delta\eta}{2}) \qquad\qquad (5.1)$$
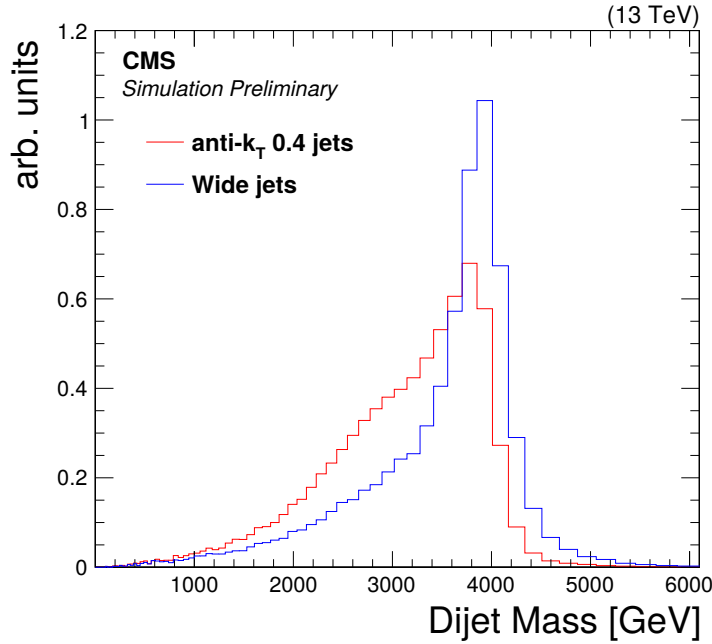
Figure 5.4. Dijet mass distribution of an excited quark resonance with $M_{q*} = 4.0$ TeV decaying in one quark and one gluon.

and the cut $|\Delta\eta_{jj}^{Wide}| < 1.3$ corresponds to $\cos\theta^* < 0.57$. This criterion is introduced to improve the signal over background ratio, excluding the region close to $\cos\theta^* = 1$ where most of the QCD processes concentrates as it was shown in Fig 2.1 in Chapt. 2. From the figure it is visible that the cut at $\cos\theta^* < 0.57$ ($|\Delta\eta_{jj}^{Wide}| < 1.3$) suppresses the background and the enhances signal for all the considered models. The analysis, with this choice, remains inclusive with respect to different new physics hypotheses.

The threshold of the cut has been re-optimized last year during the analysis preparation with Monte Carlo simulation studies and $|\Delta\eta_{jj}^{Wide}| < 1.3$ is found to be the optimal cut for the analysis at $\sqrt{s} = 13$ TeV. The expected upper limit on the cross section of an excited quark (q*) resonance as a function of the cut in $\Delta\eta$ is shown in Fig. 5.5.

5.4 Data quality studies

For events passing the main selection criteria, we have performed detailed data quality checks. These include comparisons between data and Monte Carlo as well as examination of the stability of reconstructed quantities as a function of time, and as a function of the number of reconstructed vertices to check dependence from pile-up.

5.4.1 Comparisons between data and simulation

The QCD background distributions from simulation are never used in the analysis to produce physics results, but they are shown for a qualitative comparison and
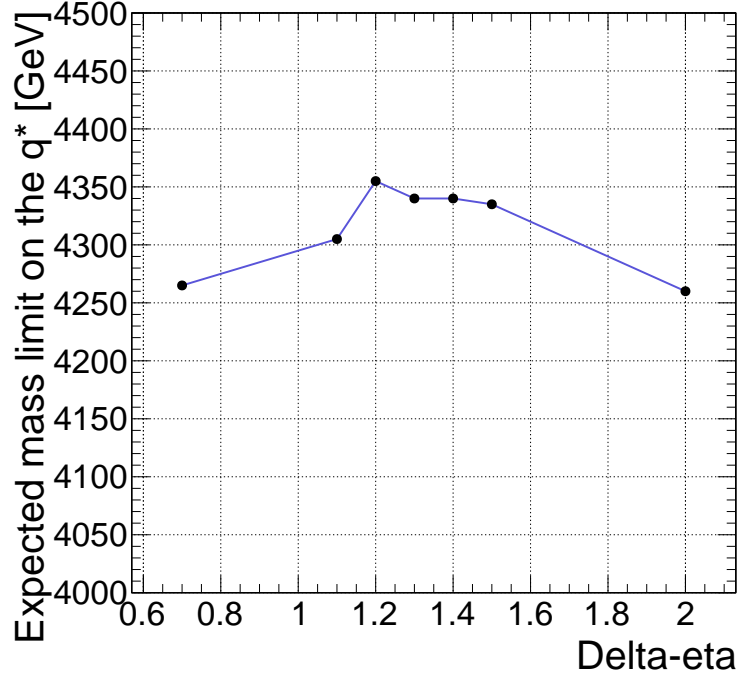
Figure 5.5. Expected upper limit on the cross section of an excited quark (q*) resonance as a function of the cut in $\Delta\eta$.

used as a check of the data quality. In all the plots presented in this section the Monte Carlo simulation is scaled to the integral of data, and the ratio of data/MC events is approximately 0.88.

First of all we compare the dijet mass spectrum in data and Monte Carlo (Fig. 5.6). The dijet mass distribution is overall in good agreement between data and simulation, it extends up to $m_{\mathrm{jj}}^{Wide} = 6.1$ TeV, that is significantly higher than the highest mass event at $\sqrt{s} = 8$ TeV, which has $m_{\mathrm{jj}}^{Wide} = 5.2$ TeV.

The other event related distributions examined are:

- the ratio of the missing transverse energy in the event to the total transverse energy, $\frac{E_T^{miss}}{\sum E_T}$;

- the angle in the transverse plane between the two leading jets, $\Delta\phi = \phi_1 - \phi_2$;

- the angle between the colliding partons and the scattered partons at the center-of-mass frame $\cos(\theta^*) = \tanh(\frac{\Delta\eta_{jj}}{2})$

The first variable, $\frac{E_T^{miss}}{\sum E_T}$, is sensitive to the presence of noise which would tend to create a significant energy imbalance, and thus fake missing energy, with respect to the total energy in the event. Events with presence of noise would therefore tend to populate higher values of this ratio than expected. The second variable, $\Delta\phi = \phi_1 - \phi_2$, is also sensitive to noise, since fake jets would populated regions away from the $\Delta\phi = \pi$ peak of dijet signature. Finally , the third variable, $\cos(\theta^*)$, which
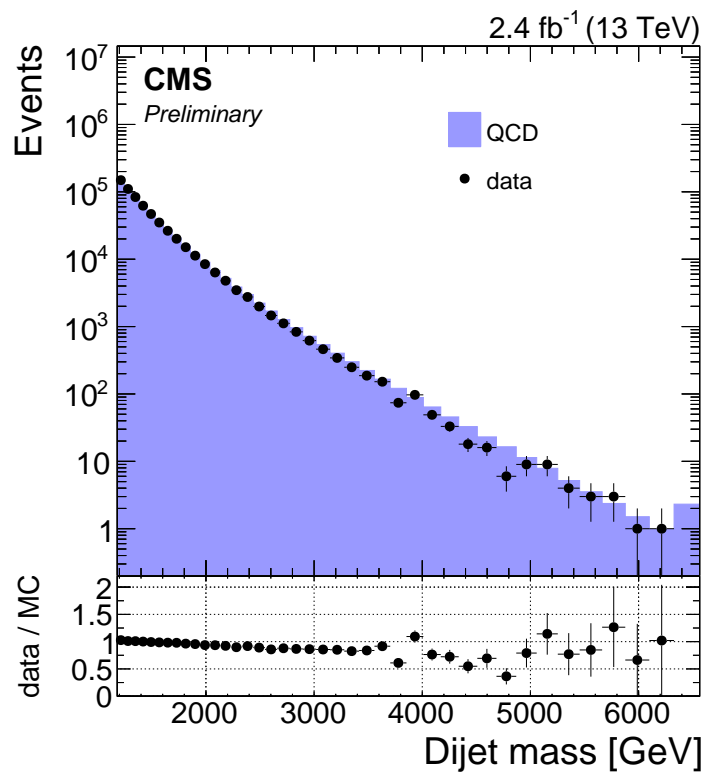
Figure 5.6. Dijet mass spectrum after full analysis selection for data (black markers) and QCD LO simulation (filled histogram). The binning has a variable size corresponding approximately to the dijet mass resolution and the last bin shows the overflow.

represents the scattering angle of the final partons in the center-of-mass frame, is also sensitive to noise and any deviation from the expectation would be an indication of data pathologies. In addition and related to this variable, the absolute difference in pseudorapidity between the two leading wide-jets is also examined.

In Fig. 5.7-5.9 we show the comparisons between data and simulated events for the variables discussed previously. They show a good agreement and no sign of noise.
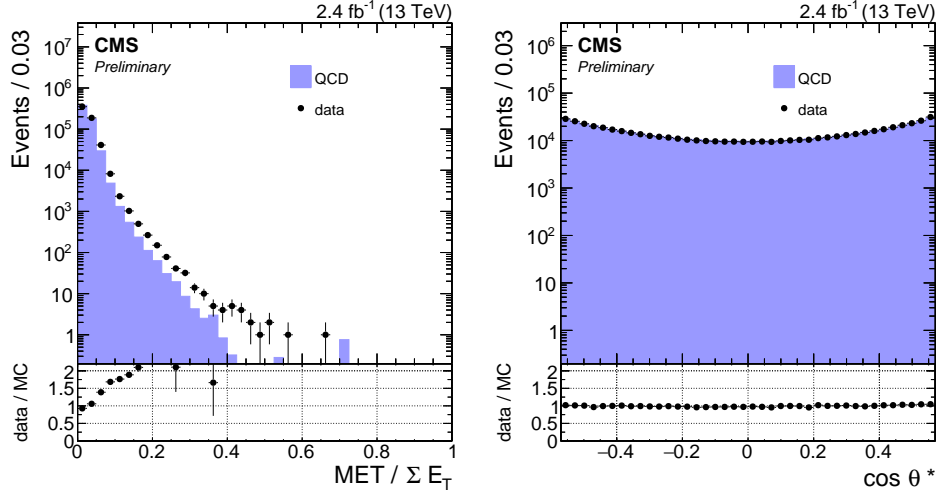

Figure 5.7. Left: Ratio of the transverse missing energy to the total transverse energy of the event. Right: The angle between beam axis and the dijet system at the center-of-mass frame, $\cos(\theta^*)$ of the event, after all selection criteria are applied, for data (points) and simulated (continuous histogram) events. The bottom pad in the plots show the data/MC ratio.

We have described the jet identification criteria in Sec. 5.3.1. They are based on the jet energy components:

- The charged hadron fraction, representing mostly the charged pions ($\pi^{\pm}$) jet content.

- The neutral hadron fraction, representing mostly the long lived neutral hadrons like Λ^0, neutrons and kaons jet content.

- The neutral electromagnetic fraction, representing mostly the neutral pions that quickly decays in photons ($\pi^0 \rightarrow \gamma\gamma$).

If Hadronic Calorimeter (HCAL) noise is present, we expect an excess of the jets neutral hadron fraction in data with respect to simulated (QCD) events. Similarly, if Electromagnetic Calorimeter (ECAL) noise is present, we expect an excess of the jets neutral electromagnetic fraction in data with respect to simulated events. In Fig. 5.9-5.11 we show the three jet energy fractions after all selection criteria are applied. These distributions show non-perfect agreement due to some issues in calorimeters reconstruction and calibration in the early running, but they don't show suspect spikes that would indicate noise. Moreover, except for the jet ID these
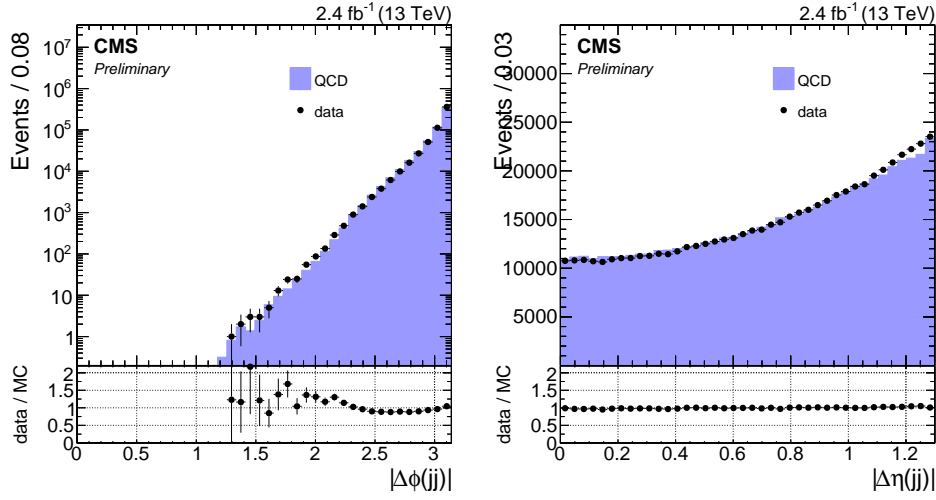
Figure 5.8. Left: The angle between the two final wide-jets of the event, $|\Delta\phi_{jj}^{Wide}|$. Right: The difference in pseudorapidity between the two final wide-jets of the event, $|\Delta\eta_{jj}^{Wide}|$, after all selection criteria are applied, for data (points) and simulated (filled histogram) events. The bottom pad in the plots show the data/MC ratio.

variables are not used in the analysis, therefore the small disagreement observed is irrelevant for the search.
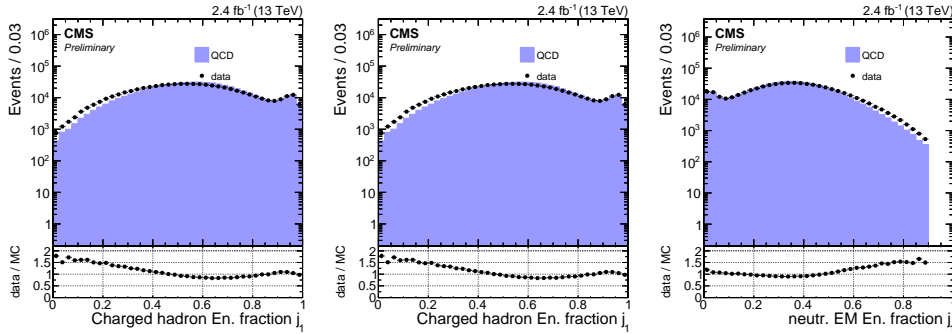
Figure 5.9. The charged hadron (left), neutral hadron (middle) and neutral EM (right) fraction of the leading jet after all selection criteria are applied, for data (points) and simulated (filled histogram) events. The bottom pad in the plots show the data/MC ratio.

In addition to the energy fractions, also the agreement between data and simulated events for the jet kinematic quantities is examined: p_T, η and ϕ, are shown in Fig. 5.12 after all selection criteria are applied, showing good agreement overall.
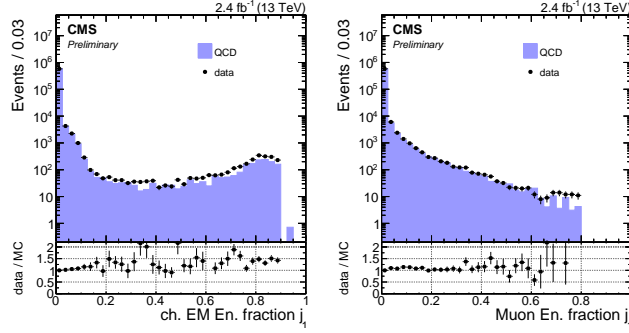
Figure 5.10. The charged electromagnetic (left) and muon (right) fraction of the leading jet after all selection criteria are applied, for data (points) and simulated (filled histogram) events. The bottom pad in the plots show the data/MC ratio.
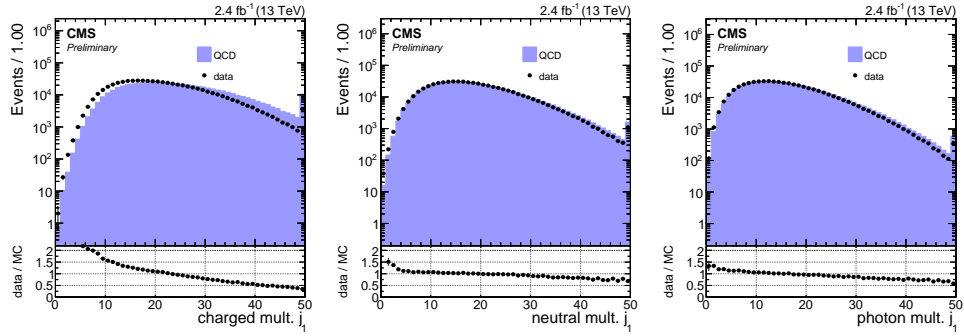
Figure 5.11. The charged particle (left), neutral particle (middle) and photon (right) multiplicity of the leading wide jet after all selection criteria are applied, for data (points) and simulated (filled histogram) events. The bottom pad in the plots show the data/MC ratio.

Figure 5.12. From left to right the p_T, η, and ϕ of the two leading jets wide-jets after all selection criteria are applied, for data (points) and simulated (filled histogram) events. The bottom pad in the plots show the data/MC ratio.
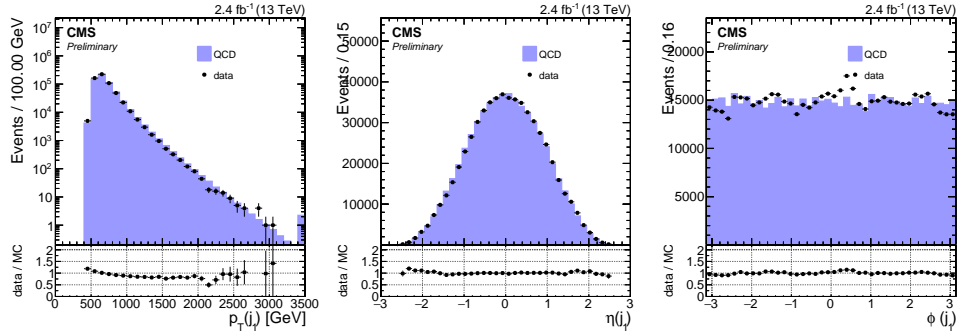
5.4.2 Event Displays

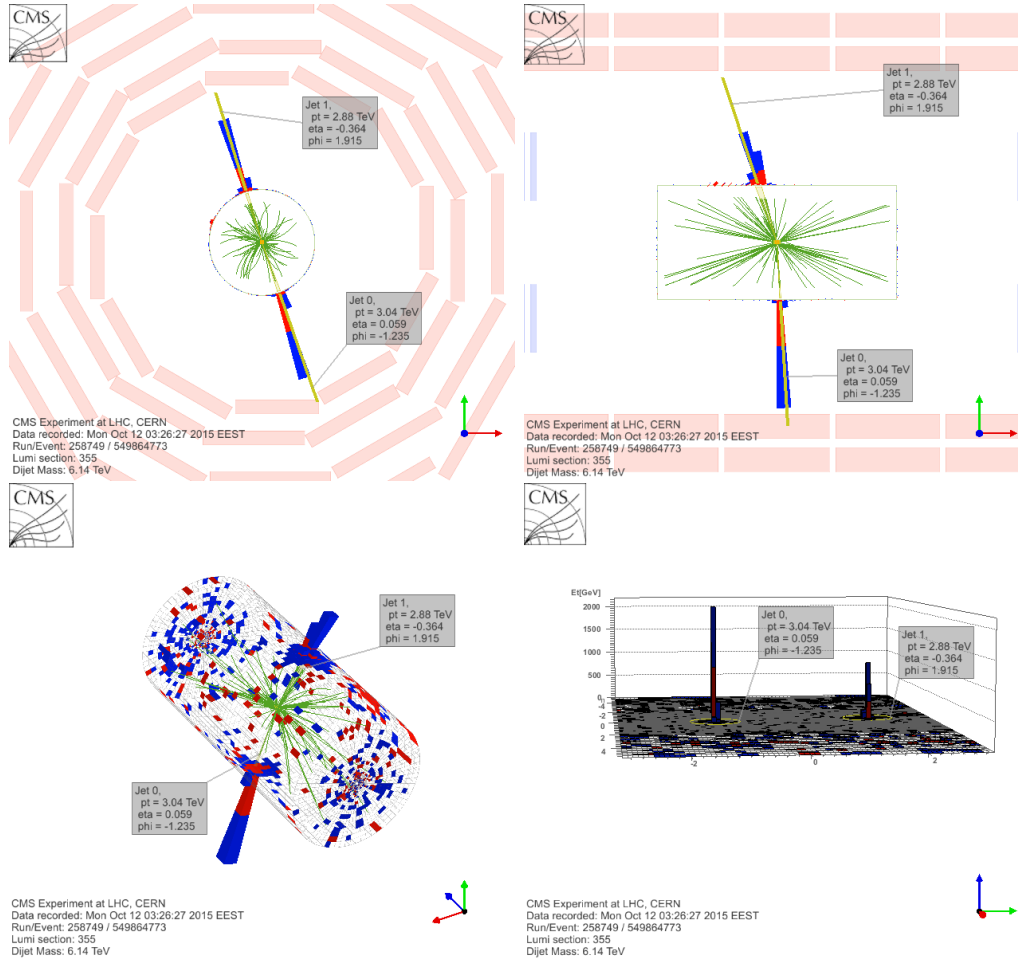The highest dijet mass event satisfying all selection criteria, visible also in Fig. 5.6, is shown in Figs. 5.13 with a mass of 6.1 TeV.


Figure 5.13. Event display of the highest dijet mass event at 6.1 TeV satisfying analysis selection criteria.

5.4.3   Time and pile-up dependence effects

One additional important study to ensure the quality of the recorded and selected data is to examine the behavior of basic data quantities as a function of time and pileup. To quantify the stability of event and jet related characteristics, they have been studied as a function of run number, and as a function of the primary vertices in the event.

In Figs. 5.14-5.15 the stability of the dijet mass, missing energy over total energy, $\cos(\theta^*) = \tanh([y_1 - y_2]/2)$ and $\Delta\phi = \phi_1 - \phi_2$ as a function of time is examined. Event related quantities are stable as a function of time, as indicated by the absence of any significant slope in the distributions.
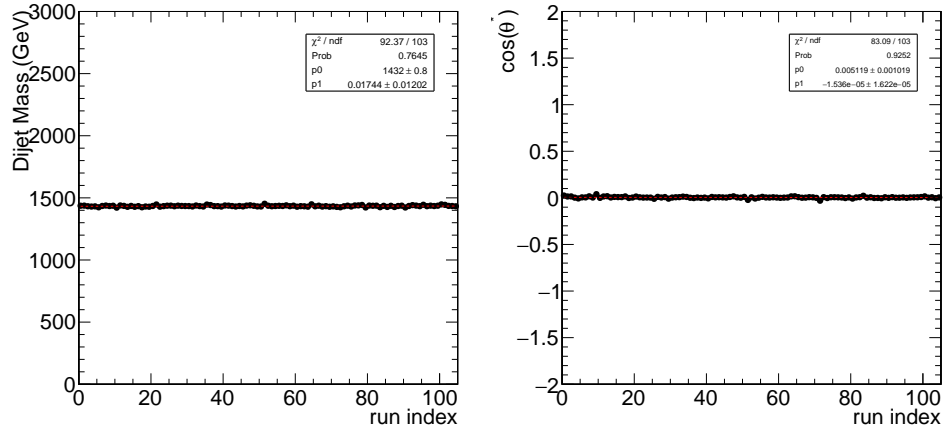
Figure 5.14. The dijet mass (left) and $\cos(\theta^*)$ (right) distribution after all selection criteria are applied as a function of time.
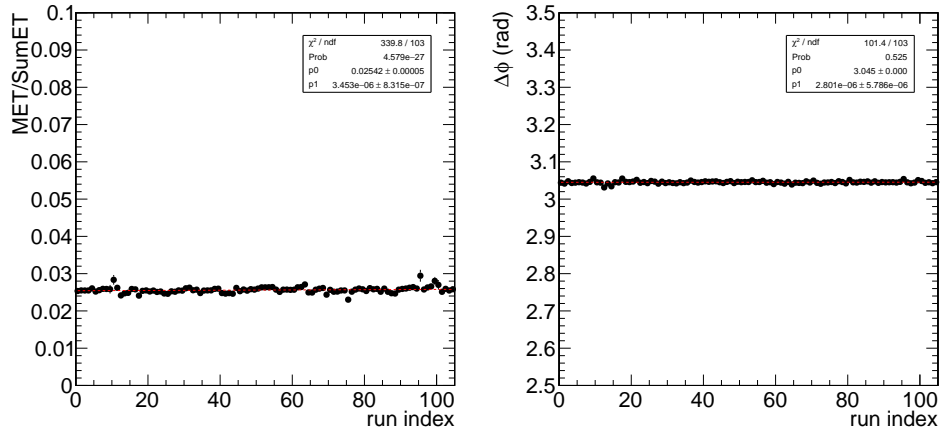
Figure 5.15. The transverse missing energy to the total transverse energy (left) and $\Delta\phi$ (right) distribution after all selection criteria are applied as a function of time.

The jet energy fractions and the jet kinematic distributions as a function of time

are shown in Figs. 5.16-5.17. Again the absence of any significant slope indicates that jet related quantities are stable as a function of time.
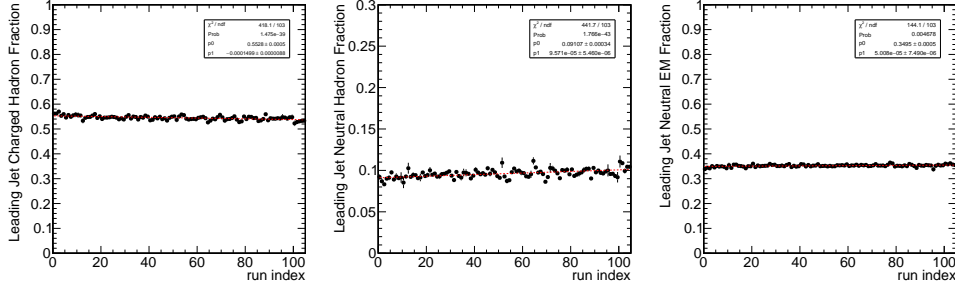
Figure 5.16. The jet charged (left), neutral hadron (middle), and neutral EM (right) energy fractions after all selection criteria are applied as a function of time.

Figure 5.17. The jet p_T (right), η (middle), and ϕ (right) energy fractions after all selection criteria are applied as a function of time.

Figs. 5.18-5.19 show the stability of the dijet mass, missing energy over total energy, $\cos(\theta^*)$ and $\Delta\phi = \phi_1 - \phi_2$ as a function of the number of reconstructed vertices. Event related quantities are stable as a function of the number of reconstructed vertices, as indicated by the absence of any significant slope in the distributions, with the only exception of the transverse missing energy to the total transverse energy (Fig. 5.19 left). This is expected and in any case this variable is never used in the analysis but only monitored to spot noise or detector problems.

The jet energy fractions and the jet kinematic distributions as a function of the number of reconstructed vertices are shown in Figs. 5.20-5.21.

Finally Fig. 5.22 shows the stability of the dijet cross section (after all analysis selections) as a function of time.

Figure 5.18. The dijet mass (left) and $\cos(\theta^*)$ (right) distribution after all selection criteria are applied as a function of the number of reconstructed vertices.
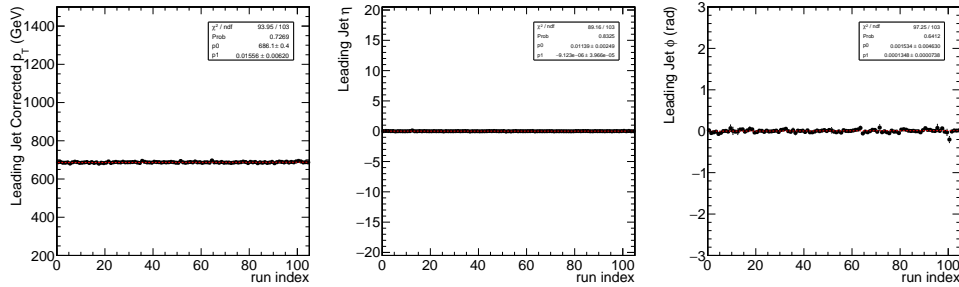

Figure 5.19. The transverse missing energy to the total transverse energy (left) and $\Delta\phi$ (right) distribution after all selection criteria are applied as a function of the number of reconstructed vertices.

Figure 5.20. The jet charged (left), neutral hadron (middle), and neutral EM (right) energy fractions after all selection criteria are applied as a function of the number of reconstructed vertices.
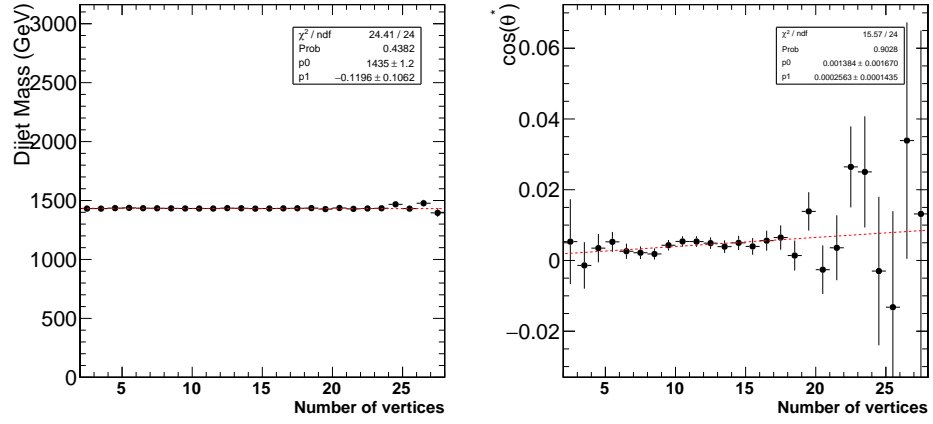

Figure 5.21. The jet p_T (right), η (middle), and ϕ (right) energy fractions after all selection criteria are applied as a function of the number of reconstructed vertices.
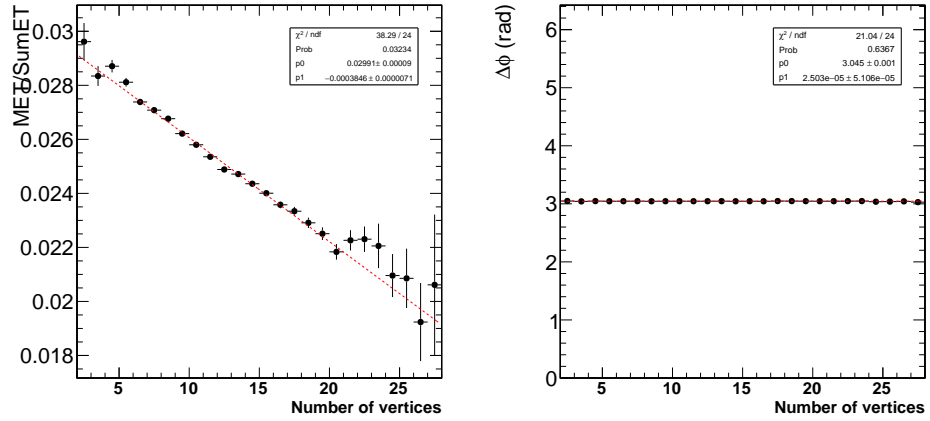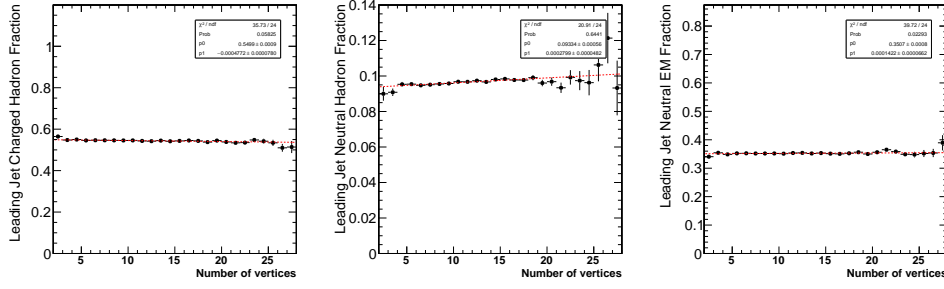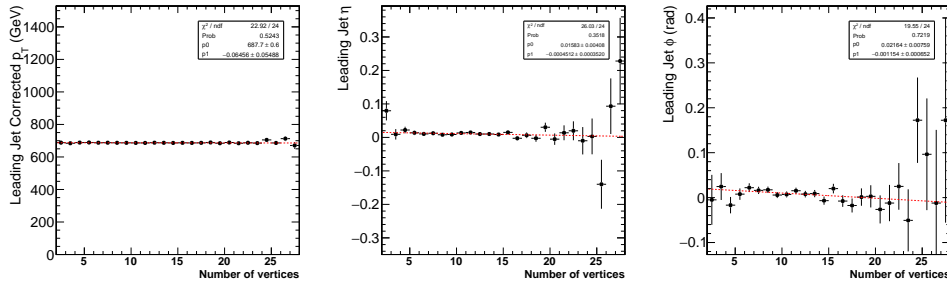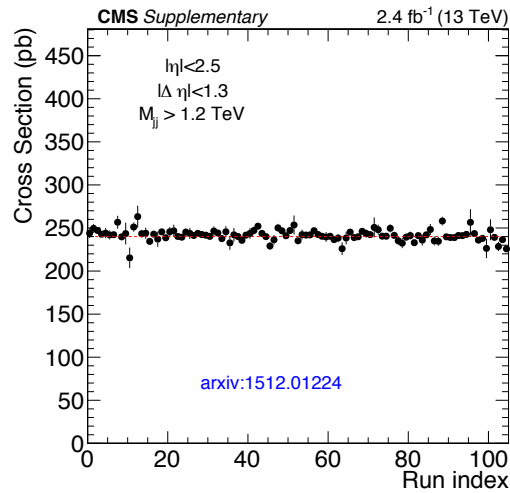

Figure 5.22. Stability of the dijet cross section as a function of time, and after all selection criteria are applied.

Chapter 6

Likelihood fit method

We describe here the likelihood fit method used to estimate the smooth background distribution and test the signal hypotheses for the observed dijet mass spectrum. The fit studies for the choice of the background parameterization and the evaluation of the bias in signal extraction described in Secs. 6.4-6.5 have been my specific tasks within the analysis team, together with a significant contribution to the analysis framework development.

6.1 Signal and background model

We have seen in Sec. 2.1.2 that the type of parton pairs in the resonance decay does affect the resonance shape. To obtain generic shapes for the three types of parton pairings (quark-quark, quark-gluon, gluon-gluon), we use excited qurk and RS graviton simulated shapes (see Sec. 5.2.1).

Some resonance shapes are shown in Fig. 6.1. They are approximately valid for any model of resonance involving these pairs of partons, assuming the relative half-width of the model $((\Gamma/2)/M)$ is small compared to the dijet mass resolution.

In order to produce mass points at intermediate masses (every 100 GeV), an interpolation technique is used. Basically, the method is relied on vertical interpolation. First, a new parameter, X, is introduced as $X = \frac{M_{jj}}{M_{Res}}$, where M_{jj} is dijet mass and M_{Res} is resonance mass. Then the new X distribution of any resonance mass are generated using neighbor existing Monte Carlo samples. For example, if we want to generate the X distribution of resonances with a mass at 4.5 TeV, we use the equation below.

$$Prob_{4.5TeV}(x) = Prob_{4TeV}(x) + \left[Prob_{5TeV}(x) - Prob_{4TeV}(x) \right] \cdot \frac{4.5 - 4}{5 - 4} \qquad (6.1)$$

Since 4 TeV and 5 TeV Monte Carlo samples are the neighbor resonance mass points of 4.5 TeV for which we have the simulation samples, they are used as input in the method. The generalization formula can be written as below.

$$Prob_M(x) = Prob_{M_1}(x) + \left[Prob_{M_2}(x) - Prob_{M_1}(x) \right] \cdot \frac{M - M_1}{M_2 - M_1} \qquad (6.2)$$
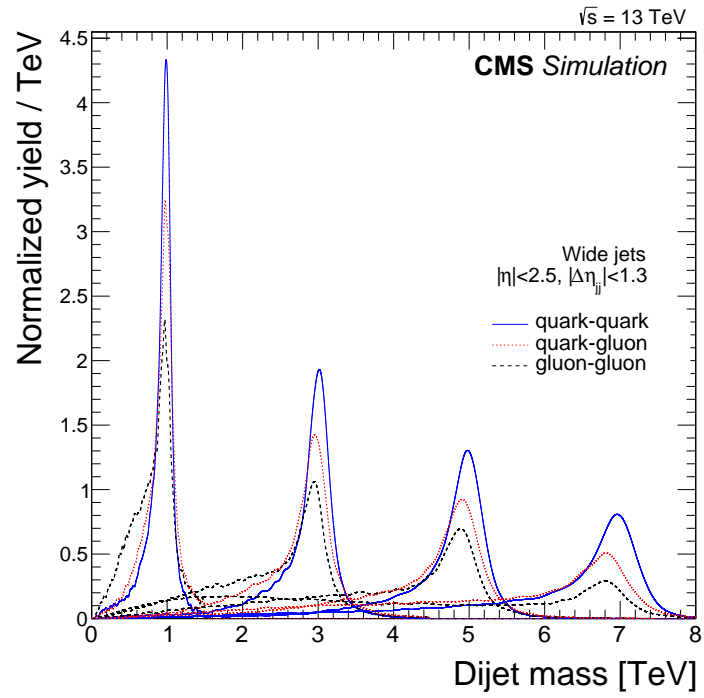
Figure 6.1. Dijet mass shapes for different mass points for the model gg → RS Graviton → gg, qq → RS Graviton → qq, qg → excited quark → qg, reconstructed with wide jets. All shapes are normalized to the same integral.

where M is the resonance mass point which is wanted to generate and M_1 and M_2 are the neighbor existing Monte Carlo mass points of M.

Finally, the generated new X distribution is converted to variable dijet mass bins to get resonance shape at any resonance masses. The distributions of X variable is shown in Fig. 6.2 for various resonance mass points and resonance types.
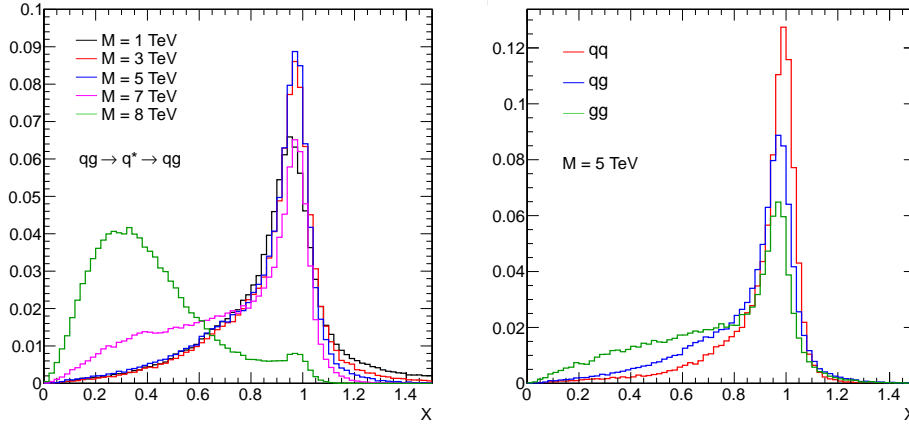
Figure 6.2. The distributions of X for qg resonance at various resonance mass points are shown on left. Comparison of X distributions at the resonance mass of 5 TeV for qq, qg and gg resonances is presented on right.

The estimate of background for the dijet analysis is obtained directly from data and does not rely on Monte Carlo simulation. The analysis does not have a blinding policy, therefore the fit technique, which is the same as Run 1, has been fixed before looking at data and tested using QCD Monte Carlo samples. The fit function used for the background-only hypothesis is the 4-parameter function in equation 6.3, used in many previous dijet searches [8, 37, 41, 66].

$$\frac{d\sigma}{dm_{jj}} = \frac{p_0(1-x)^{p_1}}{x^{p_2+p_3\ln(x)}} \tag{6.3}$$

where $x = m_{jj}^{Wide}/\sqrt{s}$ (m_{jj}^{Wide} is measured in GeV and $\sqrt{s} = 13000$ GeV).

The fit to data in order to obtain the background estimation is made maximizing the likelihood function, that is therefore introduced in the next section 6.2, before proceeding with the fit result in Sec. 6.3. The chapter concludes with Sec. 6.4, where the background parameterization choice is compared to some possible alternatives with different number of free parameters, with the use of a Fisher `F-test`.

6.2 Likelihood

Once we have defined the signal and background models in the previous section, we can define the likelihood function that is used for the background fit, the computa-

tion of significance of observed excesses in data, and the calculation of upper limits on the dijet cross sections.

We use a variable-size binning for the dijet mass, with the bin width corresponding approximately to the resolution. The binned likelihood L can be written as:

$$L = \prod_k \frac{\mu_k^{n_k} e^{-\mu_k}}{n_k!} \qquad (6.4)$$

where

$$\mu_k = N_k(B) + N_k(S) \qquad (6.5)$$

n_k is the measured number of events in the kth dijet mass bin and μ_k is the expected number of events in the same bin. The expected number of events is made of two terms:

- $N_k(B) = N_B \int_{m_{k,\text{low}}}^{m_{k,\text{high}}} B(m)dm$, where $N_k(B)$ is the number of expected background events in the kth bin, and it is obtained with an integral to the background pdf B in the dijet mass bin range, normalized to N_B;

- $N_k(S) = \mathcal{L} \cdot \sigma \cdot \int_{m_{k,\text{low}}}^{m_{k,\text{high}}} S(m)dm$ is the number of signal events (where $\mathcal{L}$ is the luminosity, σ is the signal cross section times branching ratio times acceptance and S is the signal pdf).

6.3   Dijet mass fit

The background estimation obtained from a likelihood fit to the observed dijet mass distribution in the background-only hypothesis: the signal cross section σ is set to zero in Eq. 6.5 and the 4 parameters of Eq. 6.3 are varied in order to maximize the likelihood.

Fig. 6.3 shows the measured differential cross section times branching ratio times acceptance $\frac{d\sigma \times B \times A}{dm_{jj}}$ (pb TeV^{-1}) for the dataset of 2.4 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV as a function of the dijet mass. Variable bins corresponding approximately to the dijet mass resolution are used in this analysis. The observed differential cross section in each of the dijet mass variable bins is reported in the Tab. 6.1.
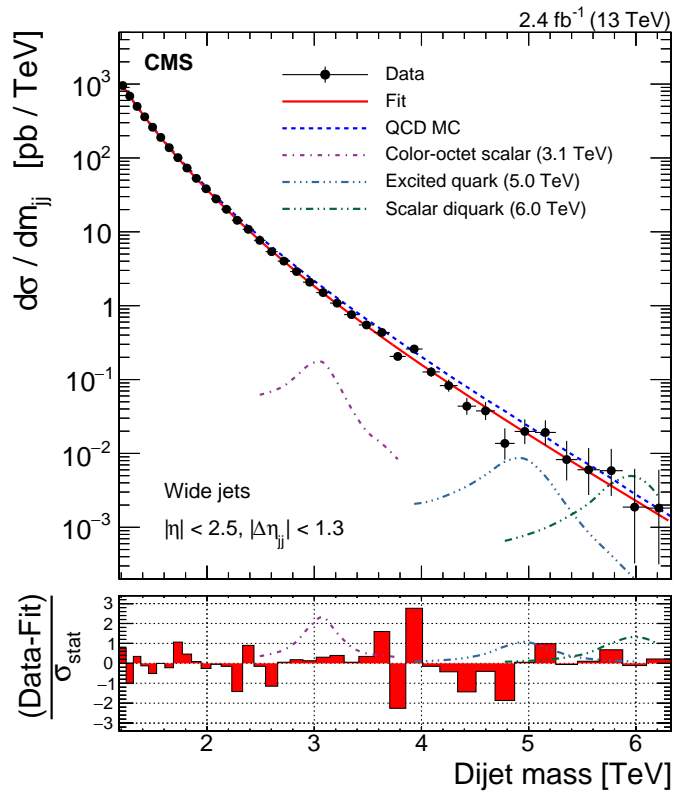
Figure 6.3. Dijet mass spectrum using wide jets (points) compared to a smooth fit (solid line) and to predictions [65] including detector simulation of QCD (dashed line) and expectations of narrow resonance signals in three models (dot-dash) at the values of resonance mass excluded by this analysis with 95% CL. The QCD prediction has been normalized to the data. The error bars are statistical only. The bin-by-bin fit residuals divided by the uncertainty of the data are shown at the bottom, and compared to the expected residuals from the same signals as above.

Bin low edge (TeV)	Bin up edge (TeV)	$\frac{d(\sigma \times B \times A)}{dm}$ (pb / TeV)	Err. low (pb / TeV)	Err. up (pb / TeV)
1.181	1.246	938	2.43	2.44
1.246	1.313	674	2.03	2.03
1.313	1.383	489	1.69	1.7
1.383	1.455	353	1.42	1.42
1.455	1.530	256	1.18	1.19
1.530	1.607	186	0.994	1.00
1.607	1.687	135	0.831	0.836
1.687	1.770	99.1	0.699	0.704
1.770	1.856	71.7	0.584	0.589
1.856	1.945	51.9	0.488	0.493
1.945	2.037	37.6	0.409	0.413
2.037	2.132	27.4	0.343	0.347
2.132	2.231	19.8	0.286	0.29
2.231	2.332	14.1	0.239	0.243
2.332	2.438	10.6	0.202	0.206
2.438	2.546	7.52	0.169	0.173
2.546	2.659	5.29	0.138	0.142
2.659	2.775	3.93	0.118	0.121
2.775	2.895	2.85	9.85e-02	0.102
2.895	3.019	2.04	8.2e-02	8.54e-02
3.019	3.147	1.48	6.87e-02	7.19e-02
3.147	3.279	1.06	5.74e-02	6.05e-02
3.279	3.416	0.743	4.71e-02	5.02e-02
3.416	3.558	0.539	3.94e-02	4.23e-02
3.558	3.704	0.426	3.45e-02	3.74e-02
3.704	3.854	0.202	2.34e-02	2.63e-02
3.854	4.010	0.254	2.58e-02	2.85e-02
4.010	4.171	0.124	1.77e-02	2.04e-02
4.171	4.337	8.13e-02	1.41e-02	1.68e-02
4.337	4.509	4.28e-02	9.99e-03	1.27e-02
4.509	4.686	3.70e-02	9.15e-03	1.17e-02
4.686	4.869	1.34e-02	5.32e-03	8.01e-03
4.869	5.058	1.95e-02	6.37e-03	8.89e-03
5.058	5.253	1.89e-02	6.17e-03	8.62e-03
5.253	5.455	8.10e-03	3.88e-03	6.4e-03
5.455	5.663	5.90e-03	3.21e-03	5.74e-03
5.663	5.877	5.73e-03	3.12e-03	5.58e-03
5.877	6.099	1.84e-03	1.52e-03	4.24e-03
6.099	6.328	1.79e-03	1.48e-03	4.11e-03

Table 6.1. Observed differential cross section $d\frac{(\sigma \times B \times A)}{dm}$.

The maximum-likelihood fit to the data in the range ∼1.2 to ∼6.3 TeV is shown in the same figure and gives the following result:

```
p0 =  1.121e-03 +/-   0.019e-03
p1 =  6.960      +/-   0.062
p2 =  6.676      +/-   0.007
p3 =  0.270      +/-   0.003
```

The exact starting point for the fit (1181 GeV) comes from the dijet pre-defined variable-size binning: the threshold is chosen to avoid bias from trigger inefficiency, as described in Sec.5.1, and the fit range is extended to the last populated bin.

The QCD MC simulation, scaled to the area of data (data divided MC is approximately 0.88), is shown for comparison.

The difference between the data and the fit is also shown at the bottom of Fig. 6.3, and that difference is normalized to the statistical uncertainty of the data in each bin. Asymmetric Poissonian errors are used and, to calculate the normalized residuals, the upper error band is taken if fit>data and the lower band otherwise.

The data are well described by the smooth fit and there is no evidence for a dijet resonance.

Finally, in the same figure, we also show example signals of the three final states (qq, qg, and gg) at the resonance mass values (6.0, 5.0, and 3.1 TeV) for which this search excludes the corresponding models (scalar diquark, excited quark, and color octet scalar) as discussed in more detail in the next chapter.

To give an estimation of the goodness of fit we calculated the chi-squared (χ^2): it is 31 for 35 degrees of freedom.

The fit function for background-only hypothesis describes well the data and we do not observe any significant excess in the dijet mass spectrum. Details on the significance of the observed excesses in data compared to the background estimation are discussed in next chapter, where we will also describe the procedure to obtain upper limits on the dijet resonance cross sections and present the results.

6.4   Background parameterization choice

The background fit to data using a function with 4 free parameters, gives a good result in terms of χ^2, as we have seen in the previous section. For a previous version of this analysis, with a much smaller dataset of 42 pb^{-1} [67], it was found with a `F-test` [68] that the adequate parameterization was a function with only 3 parameters, which could be obtained from Eq.6.3 setting p_3 to zero. However, from simulation studies with an equivalent integrated luminosity of 1 fb^{-1} (less than half the current dataset) we had the indication that the 4$^{\text{th}}$ parameter was necessary to describe that sample. Therefore we expected that at least 4 parameters are needed with our sample.

The `F-test` has been repeated with the current dataset of 2.4 fb^{-1}. The procedure is the following: we take the m_{jj}^{Wide} distributions in data after the final selection and we perform 4 different set of fits with different number of floating parameters:

- 2 parameters $f_2 = \frac{p_0}{x^{p_2}}$

- 3 parameters $f_3 = \frac{p_0(1-x)^{p_1}}{x^{p_2}}$

- 4 parameters $f_4 = \frac{p_0(1-x)^{p_1}}{x^{p_2+p_3\ln(x)}}$

- 5 parameters $f_5 = \frac{p_0(1-x)^{p_1}}{x^{p_2+p_3\ln(x)+p_4\ln(x)^2}}$

where $x = m_{\mathrm{jj}}^{Wide}/\sqrt{s}$ (m_{jj}^{Wide} is measured in TeV and $\sqrt{s} = 13$ TeV).

The fit results to the data for the 4 parameterizations are shown in Fig.6.4.
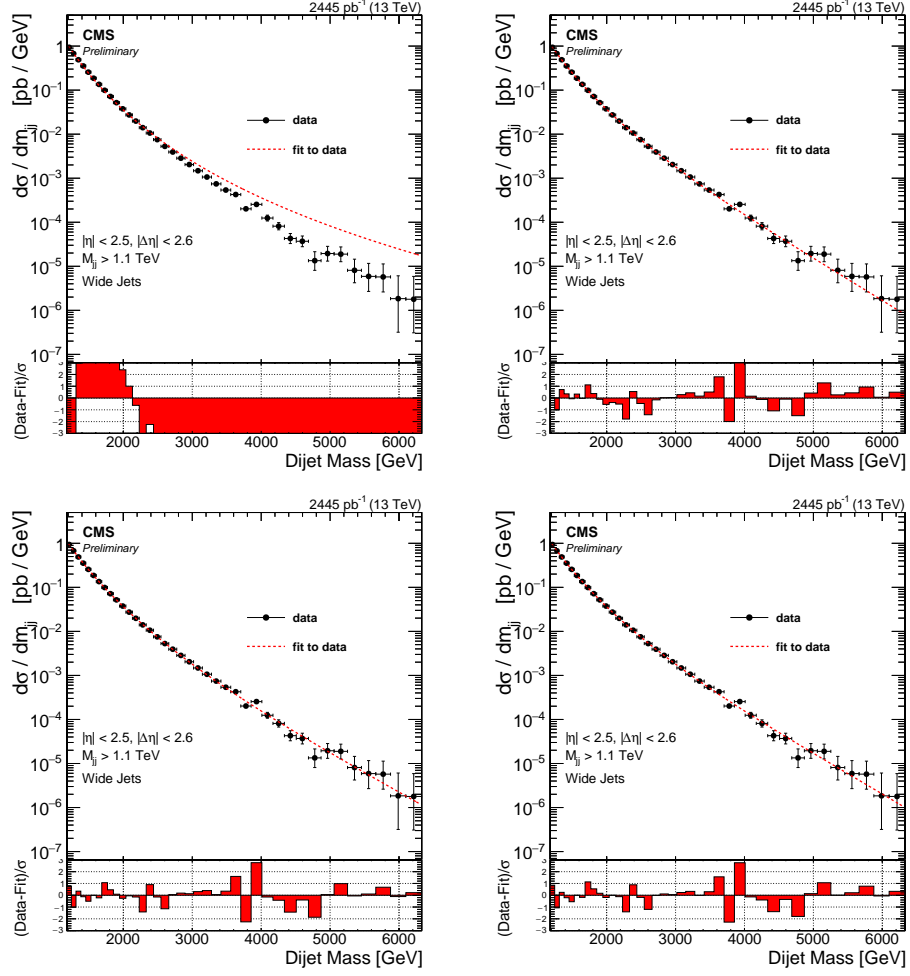
Figure 6.4. Fit to data using functions with different number of parameters: 2 (top left), 3 (top right), 4 (bottom left), and 5 (bottom right). In each plot the black points are the data and the line is the fit. At the bottom of each fit the residuals normalized to the error are shown ($\frac{\mathrm{data}-\mathrm{fit}}{\mathrm{error}}$).

If there are N data points to estimate parameters of 2 models (for example f_1 and f_2), the F-statistic is given by

$$F_{21} = \frac{\frac{RSS_1 - RSS_2}{n2 - n1}}{\frac{RSS_2}{N - n2}} \tag{6.6}$$

where model f_1 has $n1$ parameters, and model f_2 has $n2$ parameters ($n2 > n1$), and f_1 can be obtained fixing one of the parameters of model f_2. RSS_i is the residual sum of squares of model i ($\sum_{bins}(\text{data}_{bin} - \text{fit}_{bin})^2$, skipping bins with 0 entries, and with fit_{bin} value calculated integrating the function over the bin range and dividing by the bin width).

Under the null hypothesis that *model 2 does not provide a significantly better fit than model 1*, the F variable will have and F-distribution with $(n2 - n1, N - n2)$ degrees of freedom. The null hypothesis is rejected if the observed confidence level $CL_{21} = 1 - \int_{-\inf}^{F_{21}} F - \text{distribution}(x; n2 - n1, N - n2)dx$ is smaller than the desired probability α (set to 0.05). The $F - \text{distribution}$ is defined as follows:

$$F - \text{distribution}(x; d1, d2) = \frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}}}}{x B(\frac{d_1}{2}, \frac{d_2}{2})} \tag{6.7}$$

for real $x \geq 0$. Here B is the beta function.

In the particular case of this analysis, we use the `F-test` recursively to understand how many parameters are needed to fit the data: we calculate the F-statistic and the associated CL comparing f_2 and f_3, than f_3 and f_4, f_4 and f_5. The results of the `F-test` are summarized in Table 6.2.

functions	F	CL
f_2, f_3	8196.9	5.33e-15
f_3, f_4	-8.41	1
f_4, f_5	-5.01	1

Table 6.2. Results of the Fisher test for different models from 2 to 5 free parameters

Applying the `F-test` method to this specific dataset, it seems that we do not need 4 parameters to fit the data; we have however indications that 4 parameters are suitable for a dataset of this size.

- In a previous, about twice smaller, dataset the result of the same test was that the 4 parameter fit was giving a significantly better fit than the 3 parameter one. This can be attributed to statistical fluctuations. The current dataset is in fact particularly smooth, as one can see from the χ^2 values, and this might explain the result that, in this specific case, only 3 parameters are needed.

- Simulation studies during analysis preparation also suggest that using 4 parameters is adequate for a background dataset of about 1 fb^{-1}.

- Just looking at χ^2 values (in Tab. 6.3) and the residual of Figure 6.4, we can see that there is a visible improvement adding a 4$^\text{th}$ parameter. Instead, using 5 parameters instead of 4 does not change much the global picture.

In conclusion, taking all these inputs into consideration, we decided to use 4 parameters for the background function.

function	χ^2 / dof
f_2	2075.7 / 37
f_3	33.1 / 36
f_4	30.8 / 35
f_5	31.1 / 34

Table 6.3. Table of the χ^2 and number of degrees of freedom for the dijet mass fit using functions with different number of free parameters, from 2 to 5.

6.5 Signal bias evaluation

The background estimation is obtained with a fit to data and, in presence of signal, the fit function could absorb the bump, introducing a bias in the signal extraction. When the dataset is fit in the signal-plus-background hypothesis to set upper limits and calculate the significance of excesses in data, the results would be biased by the fact that part of the real signal is included in the background estimation.

It is clear that, for a new physics search, it is very important to have under control and quantify this effect. This can be done with a "signal bias study" with toy-experiments, with the following procedure:

- We generate 1000 toy experiments, with an equivalent integrated luminosity of 2 fb^{-1} and with a resonance at different masses (from 1.4 to 8.8 TeV in steps of 200 GeV).

- Then we perform a maximum likelihood fit of signal-plus-background to the dijet mass spectrum, leaving the shape parameters of the background and the signal normalization free to vary in a very large interval with flat probability.

- Finally we compare the extracted number of signal events with the injected one, and we evaluate the "signal bias".

We define the bias as the difference between the signal events obtained from the fit $N_{\mathrm{sig}}^{\mathrm{fit}}$ and the injected value $N_{\mathrm{sig}}^{\mathrm{inj}}$, normalized by the fit error:

$$\mathrm{bias} = \frac{N_{\mathrm{sig}}^{\mathrm{fit}} - N_{\mathrm{sig}}^{\mathrm{inj}}}{N_{\mathrm{sig}}^{\mathrm{err}}} \tag{6.8}$$

A simple bias study can be done using the same background parameterization (in Eq. 6.3) for generation and fit, and this is used as a "closure" test for the standard background fit function: this is the best case scenario, assuming that the real data follow exactly the model used for the fit. To give a more realistic estimation of the amount of bias, we repeat the study generating toys with a template from QCD Monte Carlo and fitting with the usual 4-parameter function. This will give an estimation of the bias assuming that data follow QCD simulation prediction.

The distribution of the signal bias at the mass points of 2.0, 3.0, 4.0, 5.0, 6.0 and 7.0 TeV for the closure method are shown in Fig. 6.5.

The bias as a function of the mass resonance for the two cases (closure test on the left, and QCD simulation template on the right) are shown in Fig. 6.6. The
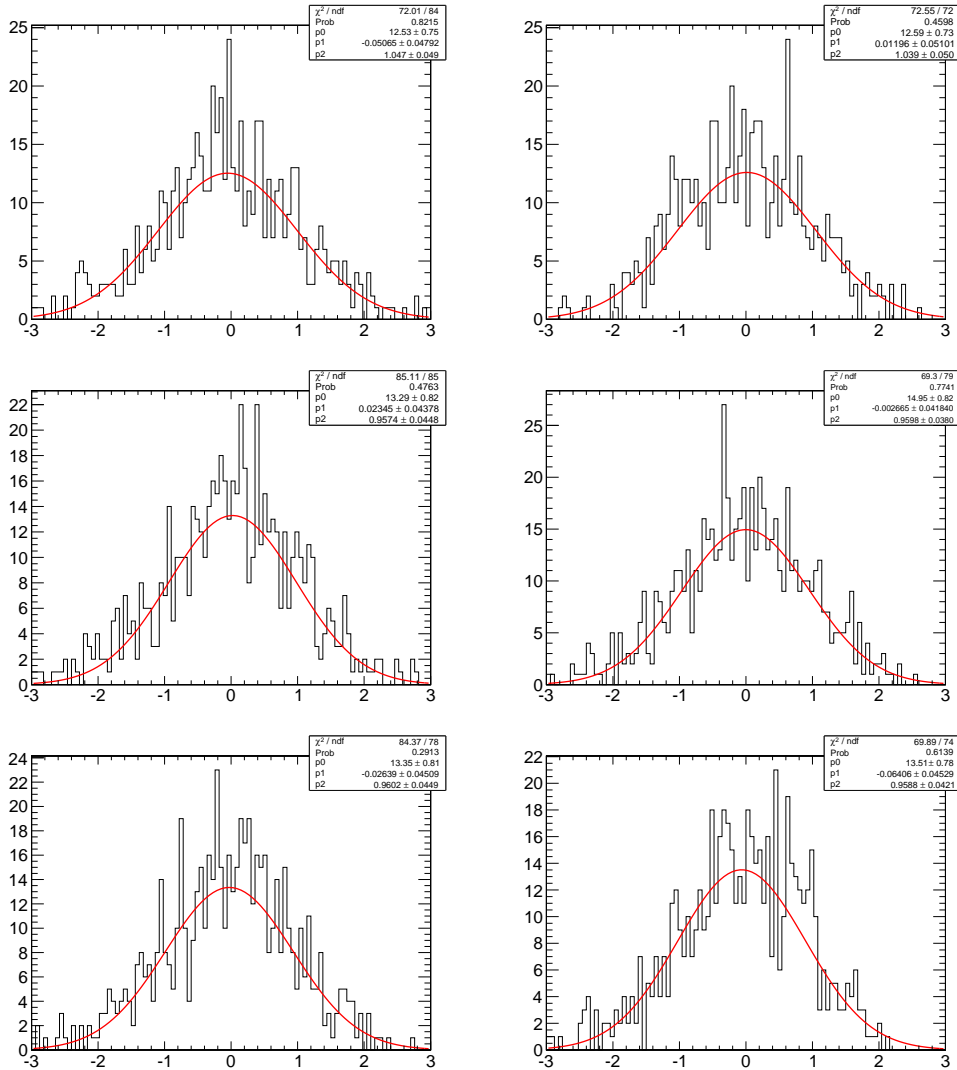
Figure 6.5. Distribution of the signal bias at different mass hypotheses: from top-left 2.0, 3.0, 4.0, 5.0, 6.0 and 7.0 TeV mass points are shown.

figure shows that in both cases the deviation between the fitted and the injected signal, normalized with the error, is small compared to 0.5 of the statistical error (dashed line), except for one the mass point at 1.6 TeV.
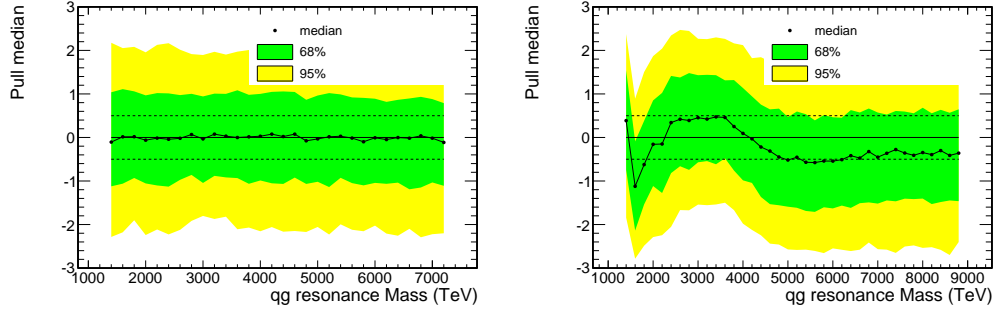
Figure 6.6. Measured signal bias as a function of the resonance mass is shown. The background is generated using the 4-parameter function in eq.6.3 for the left plot (closure test), and using a template from QCD Monte Carlo for the right plot. The points correspond to the median bias on the number of signal events and the green and yellow bands correspond respectively to 68% and 95% quantiles.

We consider good the fits that give a bias smaller than 0.5 of statistical error. If one treats the a bias as a systematic error of 50% of the statistical uncertainty the effect is, in fact, small. Adding the bias to the statistical error in quadrature one obtains:

$$\text{err}_{\text{tot}} = \sqrt{\text{err}_{\text{stat}}^2 + \text{bias}^2} = \sqrt{(1 + 0.25) \cdot \text{err}_{\text{stat}}^2} \approx 1.12 \cdot \text{err}_{\text{stat}} \qquad (6.9)$$

meaning that the bias affects the total error for $\approx 12\%$ only.

Chapter 7

Results and statistical interpretation

This chapter is dedicated to the statistical analysis, that allows to quantify the smoothness of the observed dijet mass spectrum shown in the previous chapter in Fig. 6.3. The methods for the significance calculation and the procedure to obtain upper limits on the dijet cross section are described and results presented.

7.1   Significance of observed excesses

In this section we present a method used to estimate the significance of local excesses in data.

We use a likelihood-based significance estimator defined as

$$\mathrm{Sig} = \mathrm{sgn}(S)\sqrt{-2\ln\left(\frac{L_B}{L_{S+B}}\right)}, \tag{7.1}$$

where L_B and L_{S+B} are maximum likelihoods from the best background-only and signal-plus-background fits to the data, respectively. The likelihood function is the same defined in Eq. 6.4 for the signal-plus-background hypothesis. For background-only hypothesis σ is set to zero.

Since the signal strength in the signal-plus-background fit is allowed to be negative, the significance estimator is signed accordingly. However, the significance values will only be reported for upward fluctuations (positive signal strengths).

The significance calculation has been implemented in the limit setting framework and so the exact same background parameterization, likelihood function, signal shapes, pseudo-experiment generation procedure, etc. are used (more details in the next section). It should also be noted that the look-elsewhere effect and systematic uncertainties are not included in the results presented in this section.

The results of a significance scan as a function of the resonance mass for the three resonance type hypotheses (qq, qg, and gg) are shown in Fig. 7.1. The most significant excess ($\sim 1.7\sigma$) occurs for the qq resonance hypothesis at 4.0 TeV of mass.
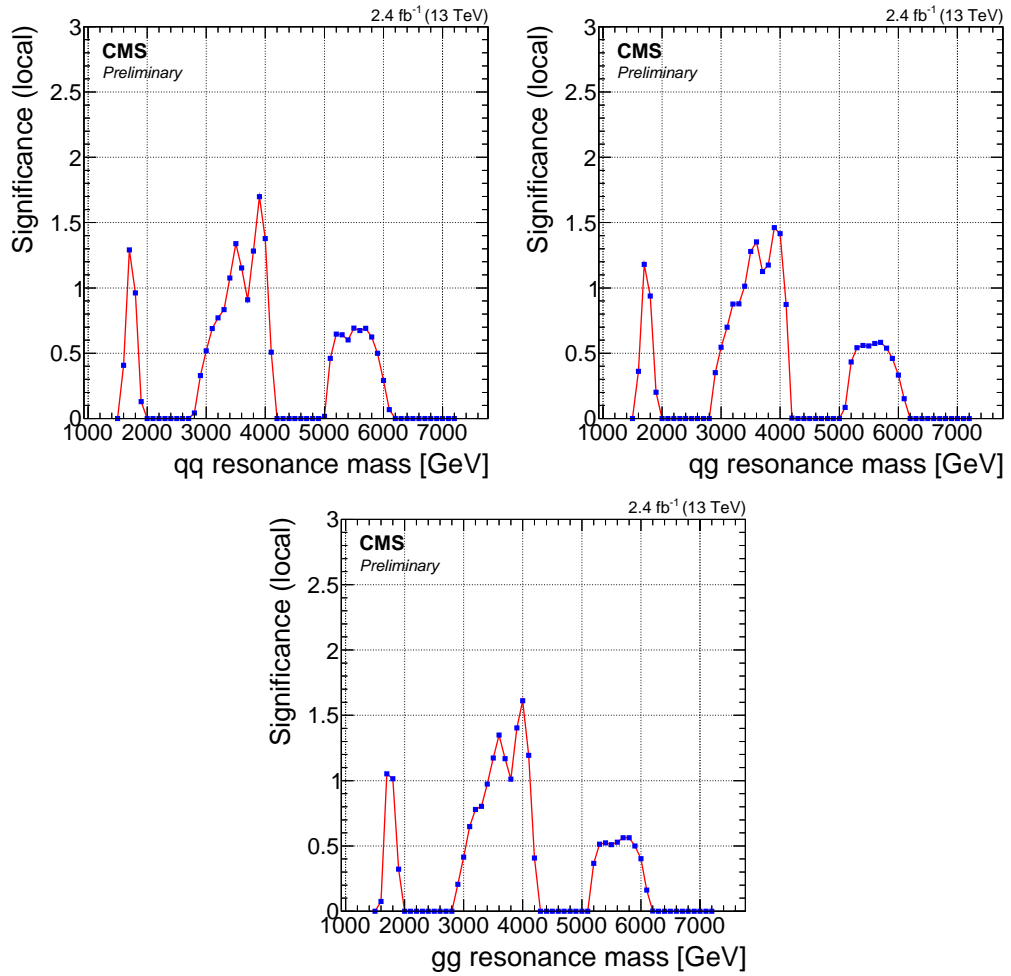
Figure 7.1. Significance scan as a function of the resonance mass for qq (top left), qg (top right) and gg (bottom) resonances.

7.2   Calculation of exclusion limits

We do not observe any significant excess in data, therefore we proceed to set upper limits on the cross sections times branching ratio of a dijet resonance.

In order to achieve proper coverage for the credibility intervals in the presence of a signal that is not yet strong enough to be evinced, the data are fit to the background function plus a signal line shape with the signal cross section treated as a free parameter.

We determine background shapes from a signal plus background fit to the data for each signal hypothesis (i.e. for each mass point considered): the resulting fit function with the signal cross section set to zero is used as the background hypothesis. The integral of the B distribution function (normalized to the total N_B) provides the expected number of background events in each bin:

$$N_k(B) = N_B \int_{m_{k,\text{low}}}^{m_{k,\text{high}}} B(m) dm \tag{7.2}$$

The number of signal events in the kth dijet mass bin,

$$N_k(S) = \mathcal{L} \cdot \sigma \cdot \int_{m_{k,\text{low}}}^{m_{k,\text{high}}} S(m) dm \tag{7.3}$$

comes from integral of the signal histogram template scaled with luminosity $\mathcal{L}$ times cross section σ. We assume a flat prior for the resonance cross section σ.

With the assumptions above, the likelihood normalized to unity is equivalent to a posterior probability density, and can be used to set limits: we calculate the posterior probability density as a function of signal cross section $P_{\text{POST}}(\sigma)$, for resonances with mass from 1.5 TeV to 7.2 TeV in 0.1 TeV steps. The reason to start the limit computation from 1.5 TeV (we have seen in Chapt. 6 that the background fit can start around 1.2 TeV) is that we want to include the full signal shape without truncating, in order to avoid bias in the signal extraction.

The 95% confidence level upper limit σ_{95} is calculated from P_{POST} as follows:

$$\frac{\int_0^{\sigma_{95}} P_{\text{POST}}(\sigma) d\sigma}{\int_0^{\infty} P_{\text{POST}}(\sigma) d\sigma} = 0.95 \tag{7.4}$$

The methodology to determine the *expected* limits in absence of signal makes use of toy experiments as described below:

- we determine background shape from a signal plus background fit to the data for each signal hypothesis;

- the number of events expected in each bin from the background function are used to create 200 pseudo-datasets that includes only the background component;

- we produce a distribution of the upper limits on cross section for each toy model at each mass point with the same procedure used for observed limits (described above);

- then the expected limit at a given mass point is derived from the median of each distribution and the uncertainty bands are obtained by using four quantiles corresponding to 1σ down and up (probability equal to 0.159 and 0.841) and 2σ down and up (probability equal to 0.021 and 0.979).

The procedure described here does not include the systematic uncertainties. The sources of systematics are described in the following section 7.3. The method to incorporate their effect in the likelihood and proceed with the limit calculation is discussed in Sec. 7.4.1 and then the results are presented in Sec. 7.4.2.

7.3   Systematic uncertainties

The source of systematic uncertainties considered in this analysis are:

- Jet Energy Scale (JES)

- Jet Energy Resolution (JER)

- Background Shape

- Luminosity

Since the background shape is derived from data, the JES, JER, and luminosity uncertainties are only considered for the resonance signal.

Jet Energy Scale (JES)

The JES uncertainty is only considered for the simulation of the resonance, as the background component comes directly from the data via the signal plus background fit and therefore has the same JES as the data. Given that the background shape is well defined and derived from data, shifting the resonance dijet mass to lower values by the jet energy scale (JES) uncertainty gives more SM background from QCD in a window around the resonance mass, and therefore a larger upper limit on the resonance cross section (the opposite happens for an increase in the JES).

These uncertainties on the jet p_T propagate almost linearly to the dijet mass. The 2% value is used for all the mass range considered in this analysis and propagated to the limit setting procedure.

A typical uncertainty of 1-2% on the jet energy scale was found for the entire η and p_T range of the reconstructed jets in run 1 [64].

Jet Energy Resolution (JER)

The uncertainty on the jet energy resolution translates into an uncertainty of 10% on the resolution of the dijet mass [64]. This uncertainty is propagated to the search by changing the width of the resonance shape by $\pm 10\%$, which results in slight stretching or shrinking of the resonance shape itself.

Background Shape

The background parameters are all considered as nuisance parameters distributed with flat prior around the best fit values in a sufficiently large range, for which the limit is found to be stable.

Luminosity

The uncertainty on the integrated luminosity is set to the value of 5% based on the result of the Van Der Meer scans. The uncertainty is propagated changing the signal normalization up and down by 5%.

7.4 Exclusion limits on dijet resonances

7.4.1 Effect of systematic uncertainties

The likelihood definition when including the systematics is

$$L(\vec{x}|a, \vec{\Delta}) = \prod_k \frac{\mu_k(a, \vec{\Delta})^{n_k} e^{-\mu_k(a, \vec{\Delta})}}{n_k!} \tag{7.5}$$

where $\vec{x}$ is the observed data, $\vec{\Delta} \equiv (\Delta_m, \Delta_\sigma, \mathcal{L}, \vec{B})$ are the *nuisance* parameters, and a is the variable of interest, that is the cross section times acceptance times branching fraction of the resonance signal. The nuisance parameters in this measurement correspond to the jet-energy scale (Δ_m), the jet-energy resolution (Δ_σ), the luminosity ($\mathcal{L}$), and the background shape ($\vec{B}$), described in the previous paragraphs. Note that we represent the background nuisance as a vector itself, since each of the 4 parameters of the background model are treated as nuisances.

The parameters a and $\vec{\Delta}$ enter through the mean number of events μ_k according to

$$\mu_k(a, \vec{\Delta}) = \mathcal{L} \cdot a \cdot S'(m_k, \Delta_m, \Delta_\sigma) dm_k + N_B \cdot B(m_k) dm_k \tag{7.6}$$

where the first and second terms in the expression are the number of signal and background events in bin k.

These are calculated for each bin k^{th} by evaluating the continuous probability distribution functions (pdfs) S' and B in the bin center m_k and multiplying by the bin width dm_k. The expressions $\mathcal{L} \cdot a$ and N_B are effectively normalizations respectively of the signal and background pdfs.

We incorporate the jet-energy-scale and jet-energy-resolution nuisance parameters in the signal pdf construction by the following transformation:

$$S'(m_k, \Delta_m, \Delta_\sigma) = S(\Delta_m \cdot [\Delta_\sigma \cdot (m_k - m_0) + m_0]) \tag{7.7}$$

where m_0 is the theoretical resonance mass. Here, $S(m)$ is the original pdf measured in simulation, while variations in Δ_m and Δ_σ result in a "shifting" and "stretching" of the pdf, respectively. This transformation therefore captures the essence of these systematic uncertainties.

To set limits on the unknown parameter a, the likelihood function needs to be integrated over the nuisance parameters (*marginalization*). The integration is formally expressed as:

$$L'(m|a) = \int L(m|a; \vec{\Delta}) \cdot \pi(\vec{\Delta})d\vec{\Delta} \tag{7.8}$$

where $\pi(\vec{\Delta})$ is the pdf of the nuisance parameters. We assume that the pdf is factorizable (uncorrelated nuisance parameters)

$$\pi(\vec{\Delta}) = \pi(\Delta_m) \cdot \pi(\Delta_\sigma) \cdot \pi(\vec{B}) \cdot \pi(\mathcal{L}) \tag{7.9}$$

(note that the background nuisance is a vector itself, since each of the 4 free parameters is treated as a separate nuisance parameter).

We choose a lognormal distribution to describe the pdfs for Δ_m, Δ_σ, and $\mathcal{L}$, where the median of the distribution is chosen to be the best estimate of the nuisance parameter, and the shape parameter is chosen to be $\log(\delta + 1)$, where δ is the uncertainty on the nuisance parameter.

The 4 parameters of the pdf $\pi(\vec{B})$ are treated as nuisances and, following the Bayesian approach, should be integrated over the entire space of parameters values to extract the posterior distribution of the signal cross section. Technically,

- a first signal-plus-background fit to the data is performed to identify a reasonable starting point for the parameter values;

- covariance matrix of the background parameters is diagonalized and the variations of the original parameters along the eigenvectors of the covariance matrix are introduced as nuisance parameters (with uniform priors as opposed to log-normal priors used for all the other nuisance parameters);

- the background nuisance parameters are integrated over a sufficiently large range around the best-fit values such that the results are found to be stable.

A ratio between limits with and without systematic uncertainties included is shown in Fig. 7.2. It can be seen that the ratio increases with decreasing resonance mass with the ratio for observed limits additionally susceptible to fluctuations in data.

To better understand the size of the contribution from different sources of systematic uncertainties, limit ratios for qg resonances for individual sources of uncertainty as well as combinations of some of them are shown in Fig. 7.3. For practical reasons, the ratio is produced for observed limits where running a large number of pseudo-experiments is not necessary. As can be seen, the dominant source of systematic uncertainty is the background uncertainty with the jet energy scale uncertainty being the second most important source. The rising trend with decreasing resonance mass is a consequence of the fact that the background shape is less constrained in the signal plus background fit for resonance shapes close to the lower edge of the considered dijet mass range.
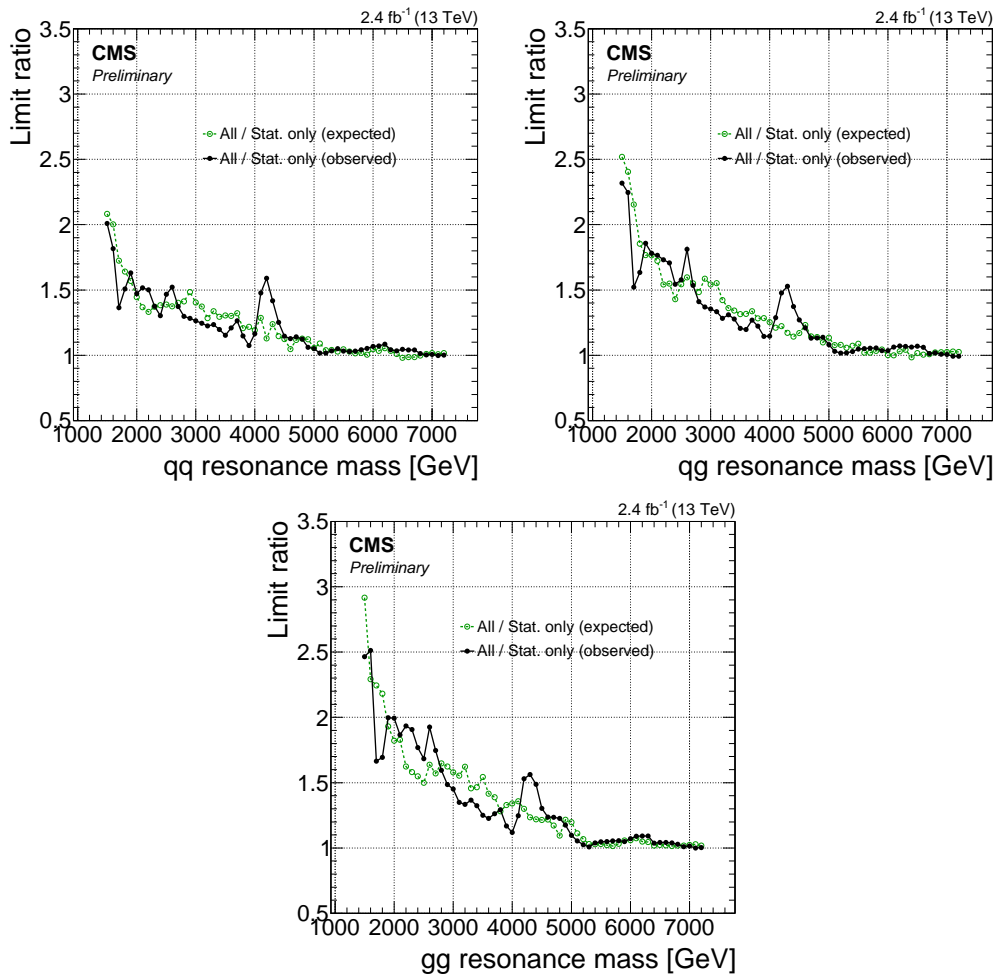
Figure 7.2. The ratio of the cross section upper limits with systematic uncertainties included and the statistical only limits as a function of the resonance mass for narrow resonances decaying into qq (top left), qg (top right), and gg (bottom) final states.
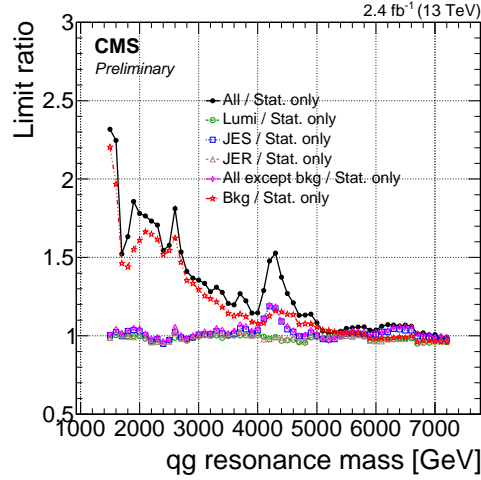
Figure 7.3. Ratios of the cross section upper limits with different sources of systematic uncertainties included and the statistical only limits as a function of the resonance mass for narrow resonances decaying into qg final states.

7.4.2 Results

No significant excess is observed in data with respect to the background prediction. As we have seen in Sec. 7.1 the largest observed excess in data occurs for the qq resonance hypothesis at a mass of 4.0 TeV and has a local significance of $\sim 1.7\sigma$.

In this section we present the final results and the exclusion limits that the dijet analysis can set with the dataset of 2.4 fb^{-1} presented in this thesis.

Figure 7.4 shows the expected limits and their uncertainty bands for qq, qg, and gg resonances compared to both observed limits and model predictions. In the plots the black points represent the observed upper limits on cross section times branching ratio times acceptance ($\sigma \times B \times A$) of the full selection described in Section 5.3.2, for narrow qq, qg, and gg resonances at 95% credibility level. The blue dashed line represent the expected upper limits in the hypothesis of no signal. Finally the theoretical expectation of $\sigma \times B \times A$ for the benchmark models described in Chapt. 2 are superimposed.

The upper limits curves represent the maximum signal cross section for which the observed (expected) data can be interpreted in background-plus-signal hypothesis, assuming a credibility level of 95%. The observed (expected) curves are below the signal models up to certain masses, depending on the model, meaning that in those regions the corresponding models are excluded by the observation (or expected to be excluded). When the observed (expected) curves cross the signal models, those masses mark the point where the analysis becomes not sensitive and the observation (expectation) is still compatible with the presence of a signal predicted from the corresponding model.

Figure 7.5 summarizes the observed model-independent upper limits at the 95% confidence level on $\sigma \times B \times A$ for narrow qq, qg, and gg resonances in one single plot. They are compared to parton level predictions of $\sigma \times B \times A$, without any detector simulation, in order to determine mass limits on new particles. The observed upper
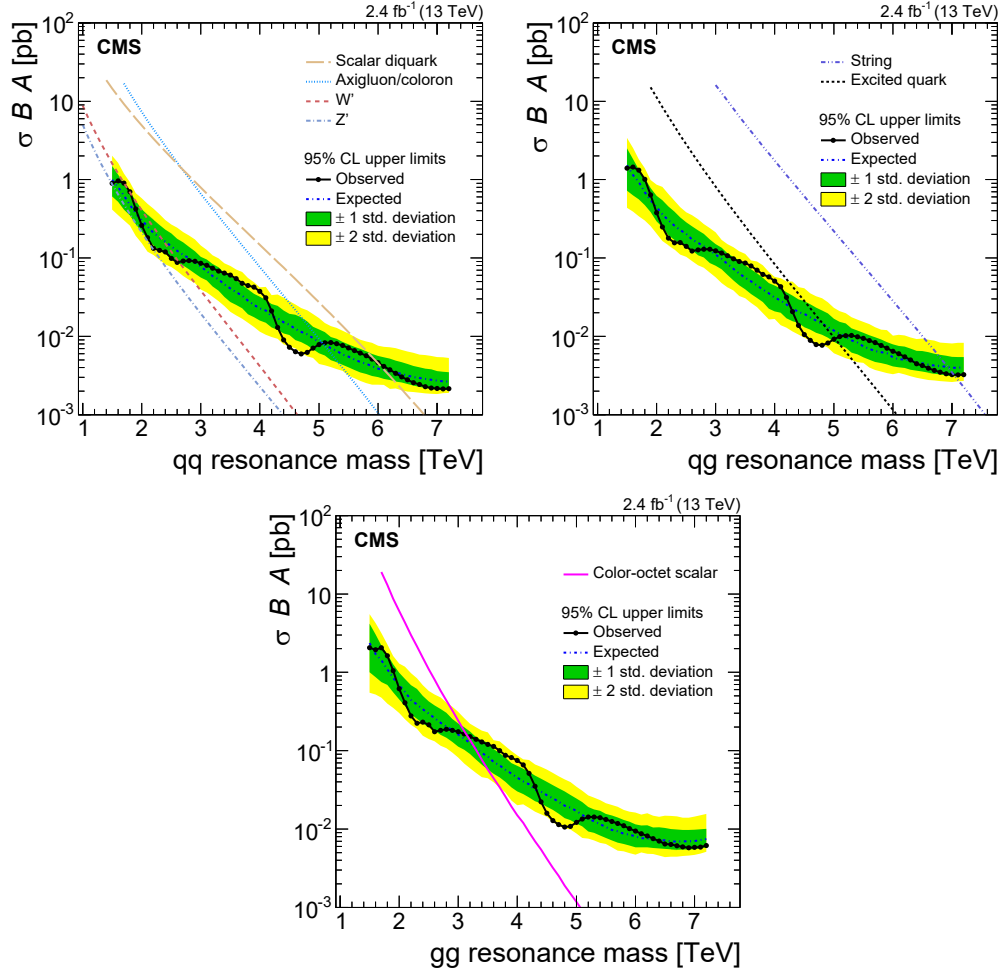
Figure 7.4. The observed 95% upper limits on for narrow resonances decaying into qq (top left), qg (top right) and gg (bottom) final states. The limits are shown as points and solid lines. Also shown are the expected limits (dashed blue lines) and their variation at the 1σ and 2σ levels (shaded bands). Predicted cross sections calculated for various narrow resonances are also shown. For the limits shown here systematic uncertainties are included in their calculation.

limits at each mass are also reported in Table 7.1. The observed and expected mass limits are reported in Table 7.2 for various models. For 6 of the 8 considered models, this first analysis with 2.4 fb^{-1} of run 2 [43] data extends the CMS previous mass limits from the full run 1 statistics.
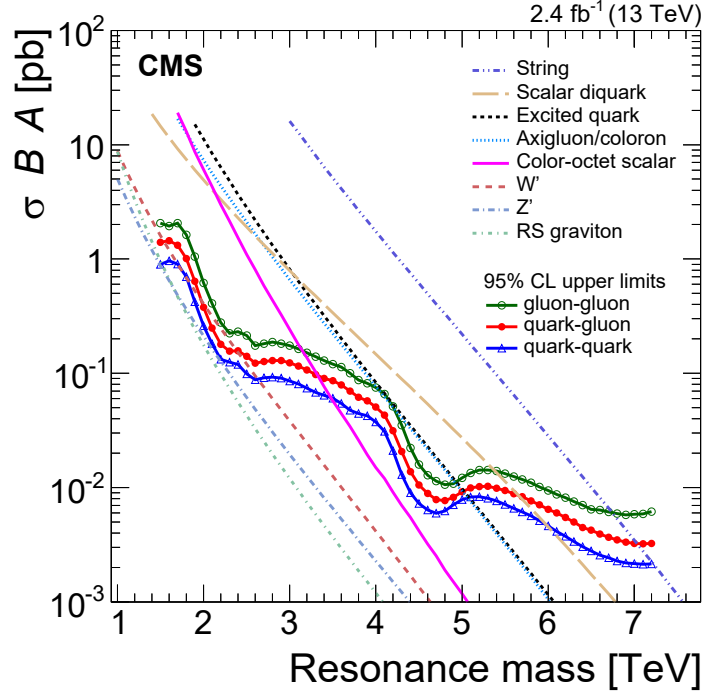
Figure 7.5. The observed 95% CL upper limits on $\sigma \times B \times A$ for dijet resonances of the type gluon-gluon, quark-gluon, and quark-quark, compared to theoretical predictions for string resonances [69,70], scalar diquarks [71], excited quarks [72,73], axigluons [74,75], colorons [75,76], color octet scalars S8 [77], new gauge bosons W$'$ and Z$'$ [78], and RS gravitons [79].

Table 7.1. Observed upper limits at the 95% CL on $\sigma \times B \times A$ for resonances decaying to , , and final states as a function of the resonance mass.

Mass (TeV)	Upper limit (pb)		
1.5	0.821	1.139	1.969
1.6	1.197	1.836	2.438
1.7	1.099	1.712	2.766
1.8	0.803	1.262	2.046
1.9	0.630	0.924	1.464
2.0	0.462	0.683	1.123
2.1	0.302	0.433	0.726
2.2	0.210	0.295	0.487
2.3	0.169	0.222	0.338
2.4	0.160	0.201	0.293
2.5	0.140	0.186	0.270
2.6	0.115	0.160	0.240
2.7	0.093	0.132	0.203
2.8	0.077	0.112	0.169
2.9	0.068	0.099	0.139
3.0	0.067	0.094	0.129
3.1	0.069	0.094	0.122
3.2	0.070	0.096	0.133
3.3	0.068	0.093	0.131
3.4	0.064	0.089	0.124
3.5	0.063	0.087	0.121
4.0	0.059	0.079	0.111
4.5	0.015	0.023	0.038
5.0	0.007	0.009	0.013
5.5	0.007	0.009	0.013
6.0	0.006	0.008	0.012
6.5	0.004	0.007	0.010
7.0	0.004	0.006	0.011

Model	Final State	Obs. Mass Limit [TeV]	Exp. Mass Limit [TeV]
String Resonance (S)	qg	7.0	6.9
Scalar Diquark (D)	qq	6.0	6.1
Axigluon (A)/Coloron (C)	q$\bar{\text{q}}$	5.1	5.1
Excited Quark (q*)	qg	5.0	4.8
Color Octet Scalar (S8)	gg	3.1	3.3
Heavy W (W′)	q$\bar{\text{q}}$	2.6	2.3
Heavy Z (Z′)	q$\bar{\text{q}}$	–	–
RS Graviton (G)	q$\bar{\text{q}}$, gg	–	–

Table 7.2. Observed and expected 95% CL limits on the mass of various resonances. This analysis excludes the resonances listed at 95% CL between a mass of 1.5 TeV and the limits shown.

7.5 Future analysis perspectives

With the present dataset the dijet search is more sensitive than the previous analysis of 20 fb^{-1} of data collected at LHC with 8 TeV proton-proton collisions for masses greater than 2 TeV.

This result is very important, but there is still room for improvements in sensitivity (and perhaps a discovery) at masses between 1.5 and 2 TeV with only few more fb^{-1} of luminosity. Therefore the dijet analysis will be repeated next year adding new data.

The dijet analysis on early data presented here considers only narrow resonances, and the same approach of Run 1 analysis has been chosen, in order to rely on a robust technique and provide results fast. However, for the future, some possible improvements and refinements can be added. In the next paragraphs are shortly reported some ideas for the future analysis upgrades.

Wide resonances

The narrow resonance models represent only a class of possible new physics phenomena. It could also be possible that new resonances are *wide*, i.e. the natural width is comparable to or larger than the experimental dijet mass resolution. The signature for a wide resonance would be a broad enhancement in the dijet mass distribution.

The sensitivity of the search to wide resonances was studied in Run 1, generating one particular model with different coupling parameters, in order to obtain different relative widths Γ/M and study the upper limits versus the resonance width. The study demonstrated that the analysis was sensitive to resonances with Γ/M up to $\approx 30\%$.

A similar study could be repeated for the dijet search in Run 2 data.

Smearing technique to obtain dijet signal shapes

If we want to consider wide resonances, and repeat the study of the analysis sensitivity versus resonance width, many signal samples are necessary. The same would happen if we want to interpret the observed limits in the context of many models, since in principle, when the resonances are wide, the shape could be model dependent.

The full simulation of a large number of MC samples is time consuming and could represent a limit for these studies. To overcome this problem, a technique to obtain the reconstructed shapes without running the full detector simulation is under development. This could be done using generator level shapes and studying the smearing effect of detector reconstruction in one particular case, for which the full simulated sample is available. One can obtain the smearing function to pass from the simulated to the reconstructed shape in that particular model, and test if this transfer distribution works for other models.

This method is now under study, but so far it seems very powerful and promising.

Background control sample

The background in dijet analysis is estimated directly from a fit to data. Other data driven techniques could be used to crosscheck the background model, such as the use of a control region.

The "sidebands" of the signal region could be obtained inverting the selection on (or equivalently $\cos\theta^*$) and using the events forward in the detector. A control region defined in this way would contain much more QCD events and a comparable number of signal events, depending on the model, with respect to the signal region (as we discussed extensively in Sec. 2.1 and illustrated in Fig. 2.1).

A transformation function from the background shape in the control region $B_{\mathrm{c}}^{\mathrm{data}}$ to the shape in the signal region $B_{\mathrm{s}}^{\mathrm{data}}$ can be obtained using the simulation:

$$B_{\mathrm{s}}^{\mathrm{data}} = B_{\mathrm{c}}^{\mathrm{data}} \times \frac{B_{\mathrm{s}}^{\mathrm{MC}}}{B_{\mathrm{c}}^{\mathrm{MC}}} \tag{7.10}$$

The assumption here is that the ratio of signal over control region shapes is the same in data and MC.

Of course a systematic uncertainty should be associated to this method, taking into account the uncertainties of MC simulation, but the advantage is that in the ratio many of them cancel out.

This technique was never used before in the dijet analysis, but is commonly used in other new physics searches, and is currently under study with promising results.

Data scouting and search at low mass

The finite speed to record data on disk represents a limit to the sustainable event rate that can be recorded by the experiments at the LHC, as we discussed in Sec. 3.2.7. Since at low dijet mass the event rates are too high to save all the events, the possibilities to extend the search below the trigger thresholds are two:

- use prescaled triggers, which save only a fraction of the total selected events;

- reduce the size of the the event, saving only the necessary informations for the dijet search in a special data format and with dedicate triggers.

The second approach is the one of the "data scouting" technique, that have been used for the dijet search in Run 1, as briefly described in Sec. 2.2.1. The plan is to repeat the data scouting analysis also with 13 TeV data in the next future.

Angular analysis

The current dijet selection allows only events with $|\Delta\eta_{\mathrm{jj}}| < 1.3$, restricting to central jets. This cut is introduced to suppress the QCD background and enhance the significance of a possible signal. However, depending on the model, also the forward region could contain signal (see discussion in Sec. 2.1 and Fig. 2.1).

A possible extension of the analysis consists in dividing the analysis in categories of $|\Delta\eta_{\mathrm{jj}}|$ (or $\cos\theta^*$) with different signal purity and combining the results, instead of introducing the angular cut. This could allow to improve the global analysis sensitivity and to extend the analysis to models that predict a resonance out of the current signal region.

Conclusions

In this dissertation the first search of new phenomena beyond Standard Model using data from proton-proton collisions at $\sqrt{s}$ =13 TeV recorded with the CMS experiment is presented. The dijet narrow resonances represent a simple and striking signature, and the search is extremely powerful at the start-up of LHC Run 2 with a new energy scale, as discussed extensively.

The data have been analyzed in parallel with the data taking since the very beginning of the run in June; first public results have been presented in the most important summer conferences of the year: some performance plots and event displays have been shown in July at the EPS conference, with a dataset of 37 pb^{-1} [80] and a complete analysis documentation has been released for the LHCP conference in August with a dataset of 42 pb^{-1} [67]. The dataset presented here amounts to 2.4 fb^{-1} of integrated luminosity collected from the CMS experiment mostly on September and October 2015, and represents the full dataset recorded with 25 ns bunch spacing configuration at $\sqrt{s} = 13$ TeV. A paper based on this data sample has been submitted to the Physical Review Letters journal on December 3, 2015 [43], and this represents the first CMS publication of a new physics search at $\sqrt{s} = 13$ TeV.

The observed dijet mass spectrum is smooth and no significant excesses are found: the largest excess is measured in the quark-quark final state, at a mass of 4.0 TeV and has a local significance of $\sim 1.7\sigma$. Upper limits on the cross sections of dijet resonances are set in three different final states (quark-antiquark, quark-gluon and gluon-gluon) starting from 1.5 TeV of mass. The observed limits are then compared with the theoretical expectation for 8 models, in order to exclude these hypotheses at 95% of credibility level for some mass ranges.

The analysis with 2.4 fb^{-1} of data is more sensitive than the search with full Run 1 dataset at $\sqrt{s}$ =8 TeV for resonance masses above 2 TeV, almost all the dijet spectrum to which this search is sensitive. This is due to the large increase of the parton luminosity factor from 8 to 13 TeV collisions, and it is confirmed by the fact that for 6 resonance models (of the 8 considered) the present analysis can set significantly more stringent limits on the cross sections.

Appendices

Appendix A

Analysis of γ + jets events for jet calibration

The measurement of the jet energy scale at CMS is done with the combination of different methods. The most important one, especially at the very beginning of LHC Run 2, when a small dataset is available, makes use of the events with one isolated photon and one jet back-to-back in the transverse plane. In this appendix we present the analysis of photon plus jet events used to calculate the scale of jet energy in data. The development of this technique in CMS started at the beginning of run 1 and is described in this thesis [81] and in this paper [64] based on 7 TeV results.

The results shown here are obtained for Particle flow jets clustered with the anti-k_T algorithm with cone ΔR=0.4 (PF ak4 jets), that are the ones used in the dijet search analysis, and a dataset of 1.3 fb^{-1} collected in 2015 at $\sqrt{s} = 13$ TeV. The dominant production diagrams of the γ + jet final state at a proton-proton collider are shown in Fig. A.1.
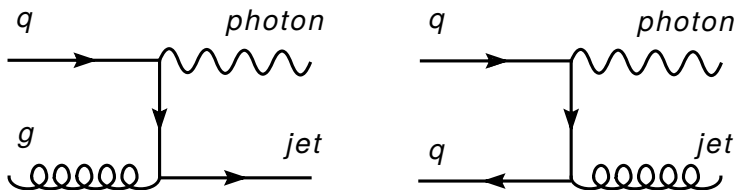
Figure A.1. Dominant photon+jet production diagrams at a proton-proton collider.

At leading order, in these events the photon and jet are balanced in the transverse plane, hence the precision with which the photon is measured in the ECAL can be exploited to infer the true jet energy momentum.

A.1 Dataset and trigger

The results shown in this appendix for the γ+jets analysis makes use of a dataset of 1.3 fb^{-1} collected in 25 ns LHC bunch spacing configuration. This sample is the

one used to derive the jet energy corrections applied in the dijet analysis described in the main body of this dissertation.

The result of the global fit combining all the methods to obtain the L3 Residual corrections as a function of the jet p_T have been shown in Sec. 4.3.2, Fig. 4.7.

Dataset	Luminosity [pb^{-1}]
`SinglePhoton Run2015D 05Oct ReReco`	553
`SinglePhoton -PromptReco-v4`	711

Table A.1. Analyzed data for the photon plus jet analysis associated to their corresponding integrated luminosity.

Signal events are stored in the `SinglePhoton` Primary Dataset after firing the single photon high level triggers. These triggers require the presence of an energy deposit in the ECAL, to which a transverse momentum requirement and other identification requirement are applied. The names of the analyzed datasets, together with their corresponding integrated luminosity, are reported in Tab. A.1.

The level of prescales introduced in the lower transverse momentum paths in principle has changed during the data taking. The presence of different prescale levels in neighboring transverse momentum ranges can create biases in the response estimation, as migrations from higher-p_T/less prescaled trigger paths can pollute lower-p_T events. In order to avoid these biases, an explicit requirement of specific triggers has been introduced in different ranges of the photon p_T.

The triggers used in the different (offline) photon p_T regions are summarized in Tab. A.2 together with the online trigger selection and the order of the prescale. The trigger selection, in addition to a photon p_T threshold, includes several other identification criteria, in order to select events where the photon is clean and isolated:

- R_9 variable is calculated taking the ECAL supercluster associated to the photon, and building a 3×3 matrix around the most energetic crystal. R_9 is defined as the energy of the highest energy crystal divided by the total energy in the 3×3 matrix. This variable is close to 1 when the photon is clean and well isolated. Two photons from a boosted $\pi^0 \to \gamma\gamma$ decay could fake one single photon in the reconstruction, but would also have a broader cluster, with R_9 well below 1.

- H/E variable is the ratio between the energy in the HCAL calorimeter geometrically close to the photon, divided by the energy of the ECAL cluster associated to the photon candidate. For a true photon this variable is close to zero, since most of the energy is deposited in the ECAL calorimeter.

- Isolation variable is the sum of the energy in the tracker (charged isolation), ECAL (electromagnetic isolation), or HCAL (neutral hadron isolation) inside a cone with $\Delta R = 0.4$ around the photon candidate, and excluding its energy. For a good γ plus jet event it is close to zero.

The photon p_T-binning used in the analysis is chosen in order to have the same edges of the triggers, to avoid any other bias coming from the changing of prescales.

photon p_T [GeV]	trigger path	online selection	prescale
$40 < p_T < 60$	HLT_Photon30_R9Id90_HE10_IsoM	$p_T > 30$ GeV && R9 > 0.9 && H/E < 0.1 && Iso < 10 GeV	$\approx$300
$60 < p_T < 85$	HLT_Photon50_R9Id90_HE10_IsoM	$p_T > 50$ GeV && R9 > 0.9 && H/E < 0.1 && Iso < 10 GeV	$\approx$150
$85 < p_T < 100$	HLT_Photon75_R9Id90_HE10_IsoM	$p_T > 75$ GeV && R9 > 0.9 && H/E < 0.1 && Iso < 10 GeV	$\approx$3
$100 < p_T < 130$	HLT_Photon90_R9Id90_HE10_IsoM	$p_T > 90$ GeV && R9 > 0.9 && H/E < 0.1 && Iso < 10 GeV	$\approx$1.5
$130 < p_T < 175$	HLT_Photon120_R9Id90_HE10_IsoM	$p_T > 120$ GeV && R9 > 0.9 && H/E < 0.1 && Iso < 10 GeV	1
$p_T > 175$	HLT_Photon165_R9Id90_HE10_IsoM	$p_T > 165$ GeV && R9 > 0.9 && H/E < 0.1 && Iso < 10 GeV	1

Table A.2. Summary of the HLT trigger paths used in the analysis in different regions of the offline photon p_T. The path is reported together with the selection applied to the online photon object, and the order of prescale.

The analysis is performed separately also in regions of pseudorapidity of the jets. The binning in photon transverse momentum (p_T^γ) and pseudorapidity of the leading jet $\eta(j_1)$ is summarized in Tab. A.3.

variable	analysis bin edges
p_T^γ [GeV]	40, 60, 85, 100, 130, 175, 250, 300, 400, 500, 1500
$\eta(j_1)$	0, 1.3, 2.0, 2.5, 3.0, 3.2, 5.2

Table A.3. Analysis binning in in photon transverse momentum $(p_T{}^\gamma)$ and pseudorapidity of the leading jet $\eta(j_1)$.

A.2   Monte Carlo Simulation dataset

The data are compared to Monte Carlo events generated with `PYTHIA8` with tune `CUETP8M1` and passed through a full simulation of the CMS detector, implemented in the `GEANT 4` software framework.

The analyzed sample (see Tab. A.4) has been generated approximately flat in the transverse momentum of the hard scattering interaction $(\hat{p}_T)$, in order to have statistics in the high energy tail. The per-event generator weight is stored in order to obtain the correct $\hat{p}_T$ spectrum.

Physical process	Simulated events
Photon plus jets, 15 $\hat{p}_T$ 6000 flat	9799030

Table A.4. Table containing the process and the number of generated events. Since the events are simulated with an approximately flat distribution in order to have a good statistic significance in the high p_T tail, a per-event generator weight is also stored in the sample. This allows to re-obtain the correct event characteristics distributions.

The Monte Carlo events have been reweighed in order to match the number of vertexes observed in the data.

A.3   Selection

The analysis selection criteria for the photon are the following:

- Transverse momentum of the photon $p_T^\gamma > 14$ GeV. This cut is introduced to match the pre-cut in the CMS analysis format of data.

- Restriction of the photon to the central region of pseudorapidity $|\eta(\gamma)| < 1.3$, where the energy resolution is best.

- Photon identification (PhotonID) selection includes requirements on several cluster shape variable in the ECAL and on the isolation of the photon candidate. The cut-based photon selection has been studied by the Physics Object Group (POG) of CMS in 13 TeV data and they provide centrally a recipe for

the analyses that make use of photons. Here the tight working point is used, with an efficiency of $\approx 75\%$.

The requirement applied to jets are

- transverse momentum of all the jets in the event $p_T > 15$ GeV;

- fraction of the transverse momentum of the subleading jet with respect to the photon $\alpha < 0.3$, with $\alpha \equiv p_T(j_2)/p_T^\gamma$);

Finally the event selection requires

- exactly one "good" photon: we reject all events with more than one photon satisfying all the criteria listed above;

- events well balanced in the transverse plane: $\Delta\phi(j_1, \gamma) > 2.8$, being $\Delta\phi(j_1, \gamma)$ the angle in radiants in the transverse plane between the leading jet and the photon.

The comparison of data to simulation for some representative variable of the analysis after the full selection is shown in Fig. A.2. The agreement between data and simulation is overall reasonably good.

A.4 Jet Transverse Momentum response measurement

The aim of jet energy corrections is to bring the jet response to 1 homogeneously in all the detector and set the absolute scale (i.e. obtain scale factors as a function of η and p_T). Two methods are used to quantify the jet energy response: the jet balancing and the Missing-E_T projection fraction method (MPF).

A.4.1 Photon-jet balancing

The p_T balancing is simply defined as the ratio between the measured transverse momentum of the leading jet (j_1) and the transverse momentum of the photon, which serves as reference object.

$$R_{\text{balancing}} = \frac{p_T^{j_1}}{p_T^\gamma} \tag{A.1}$$

It is always possible to factorize the Eq. A.1 in the following manner:

$$R_{\text{balancing}} = \frac{p_T^{\text{recoJet1}}}{p_T^\gamma} = \frac{p_T^{\text{recoJet1}}}{p_T^{\text{genJet1}}} \cdot \frac{p_T^{\text{genJet1}}}{p_T^\gamma} \tag{A.2}$$

where we have introduced the transverse momentum of the generator jet matched to the reconstructed jet.

The new expression presents two factors. By comparing to Eq. A.1 one can easily recognize the true response variable in the first ratio. We will define this ratio as the *intrinsic response*, and it depends on the chosen jet reconstruction scheme and on the jet transverse momentum. It is the object of the jet energy scale
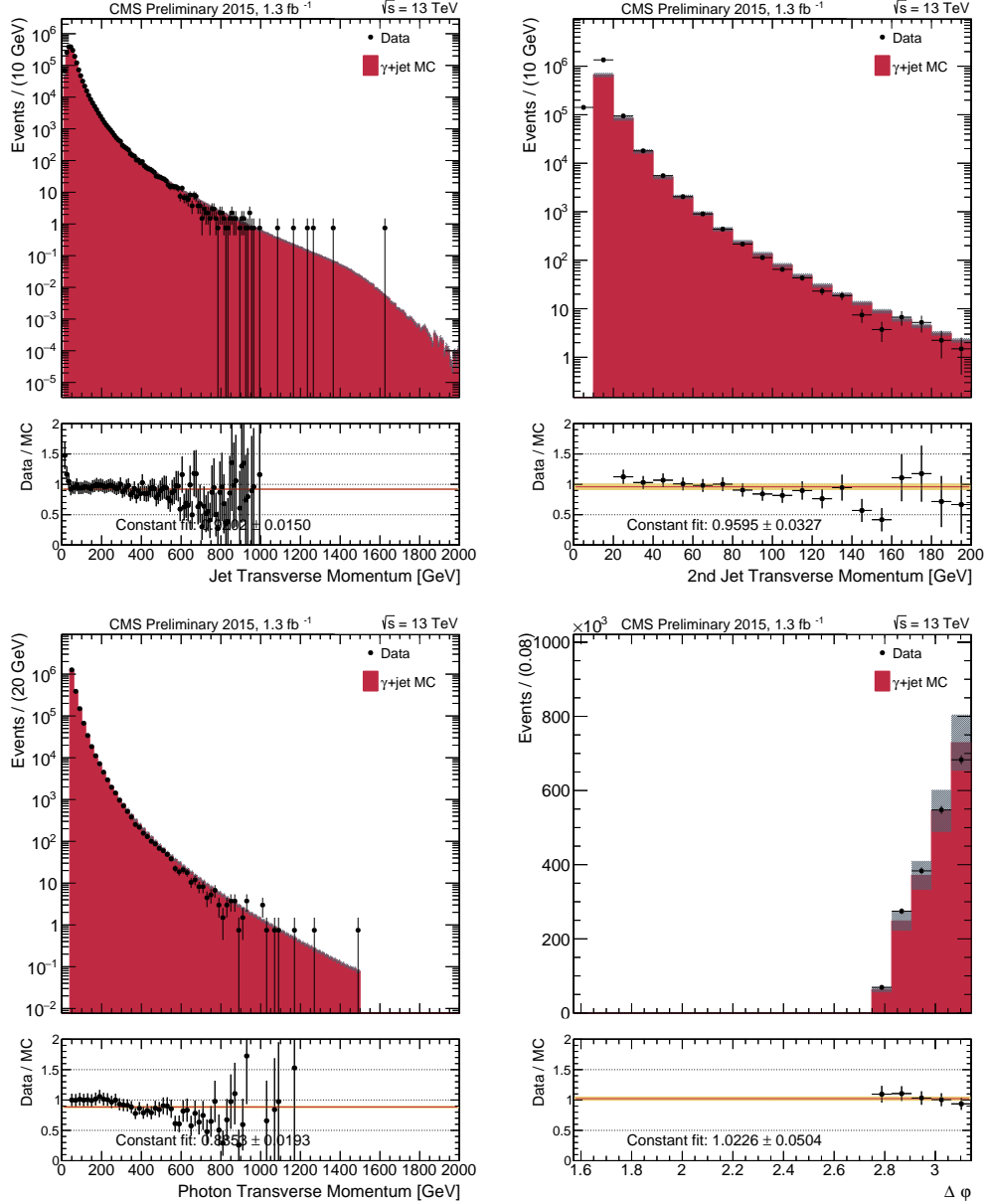
Figure A.2. Data-MC comparison of some kinematic distribution of γ+jets events: the transverse momenta of the leading and second leading jet -when present- are shown on top; photon p_T is on bottom left; the angle between the photon and the leading jet in the transverse plane $\Delta\phi$ is on bottom right. Monte Carlo simulation is scaled to the luminosity of the data sample.

measurement. The second ratio, on the other hand, $\frac{p_T^{\text{genJet}}}{p_T^\gamma}$, is a measure of the imbalance at generator level between the photon and the leading jet. It depends on the amount of additional event activity, and on the efficiency of the chosen jet algorithm. We will call it generically *imbalance.*

Imbalance is the main source of bias in estimating the jet energy scale with photon+jet balancing. In order to reduce its effects a requirement on the transverse momentum of the subleading jet is introduced. We define α as the ratio between the transverse momentum of the second leading jet to the photon's one.

$$\alpha \equiv p_T^{j_2}/p_T^\gamma \qquad (A.3)$$

This variable is required to be less than a threshold, that is set to $\alpha < 0.3$.

The requirement on the second jet p_T does not eliminate completely the bias. In order to correct this effect a more sophisticated approach is needed: the jet response has to be extrapolated to the ideal event topology with one photon and only one jet.

A sketch of a γ+jets event with a second jet caused by the radiation of a gluon is shown in Fig. A.3.
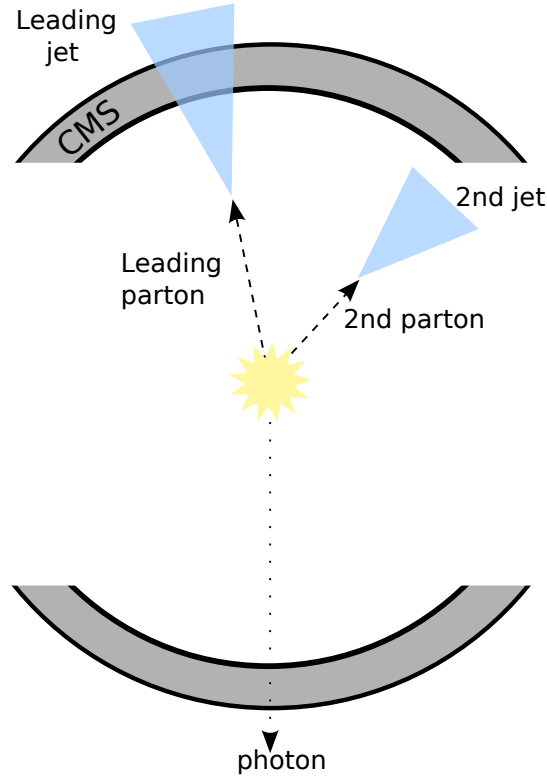
Figure A.3. Schematic view of a γ+jet event in the transverse plane. The second jet from ISR/FSR causes an "imbalance" between the leading jet and photon p_T.

A.4.2 Missing-E_T Projection Fraction

The MPF is another jet energy response that is defined using the missing transverse energy instead of the jet transverse momentum. In the γ+jet topology, the transverse momentum of the photon is perfectly balanced by a hadronic recoil supposed to originate from the outgoing parton of the hard subprocess.

$$\vec{p}_T^{\gamma} + \vec{p}_T^{recoil} = 0 \tag{A.4}$$

On the detector level, these transverse momenta are scaled by the detector responses R_γ and R_{recoil} of the respective quantity. Deviations of the detector responses from one lead to some reconstructed missing energy, E_T^{miss}.

$$R_\gamma \vec{p}_T^{\gamma} - R_{recoil} \vec{p}_T^{\gamma} = -\vec{E}_T^{miss} \tag{A.5}$$

Multiplying both sides of Eq. A.5 by $\frac{\vec{p}_T^{\gamma}}{|p_T^{\gamma}|^2}$ and substituting Eq. A.4.2 the definition of R_{MPF} is obtained:

$$R_{MPF} \equiv R_{recoil} = \frac{\vec{E}_T^{miss} \cdot \vec{p}_T^{\gamma}}{|p_T^{\gamma}|^2} + R_\gamma \tag{A.6}$$

The jet response with the two methods in three different p_T (γ) bins and for the jet in the barrel $|\eta| < 1.3$ is shown in Fig. A.4.

In each photon transverse momentum bin the response estimate can be derived as the mean of the truncated response shape, in order to minimize the effects of rare outliers. The truncation is a two-step procedure: first the mode of the distribution is found through an iterative gaussian fit, extended only to ± 1.5 standard deviations about the gaussian mean; once the bin in which the mode is included is found, bins are iteratively added, symmetrically about the mode bin, until 99% of the histogram's integral is reached. The response estimator is then defined as the average of this truncated distribution, and the resolution as its RMS, divided by the mean.

The response as a function of the photon p_T is shown in Fig.A.5 for the balancing method (left), and the MPF method (right).

The simple balancing method, in general, gives a response value more distant from 1 then the MPF. This is more pronounced at low p_T, and it is mostly due to the the presence of other jets from radiation, that introduce an "imbalance" in the transverse plane. The MPF method has instead a flatter response, since by definition it is not sensitive to the leakage of energy in the leading jet's cone.

An extrapolation procedure is introduced, in order to estimate the balancing response also in presence of more than one jet. The extrapolation method is described in detail in the following section.

A.4.3 Extrapolation method

In order to minimize the imbalance bias, that is important especially for the balancing response, an extrapolation method is introduced. Basically the trend of the response is studied as a function of the subleading jet's transverse momentum. The
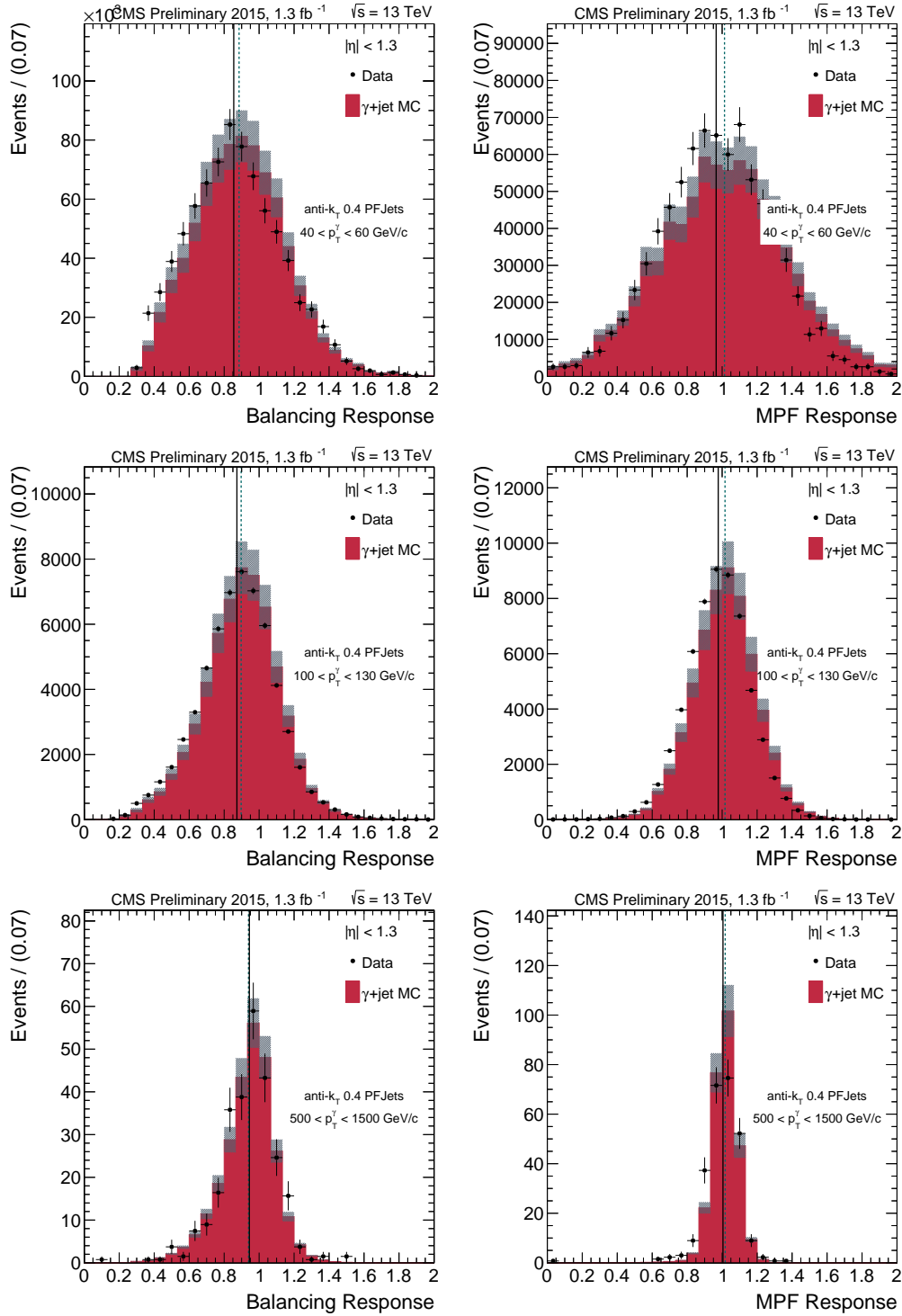
Figure A.4. Balancing (left) and MPF (right) response distributions in 1.3 fb^{-1} of 2015 data, in three representative transverse momentum ranges, for PF ak4 jets reconstructed in the CMS barrel. The MC distributions are normalized to the luminosity of the data sample.
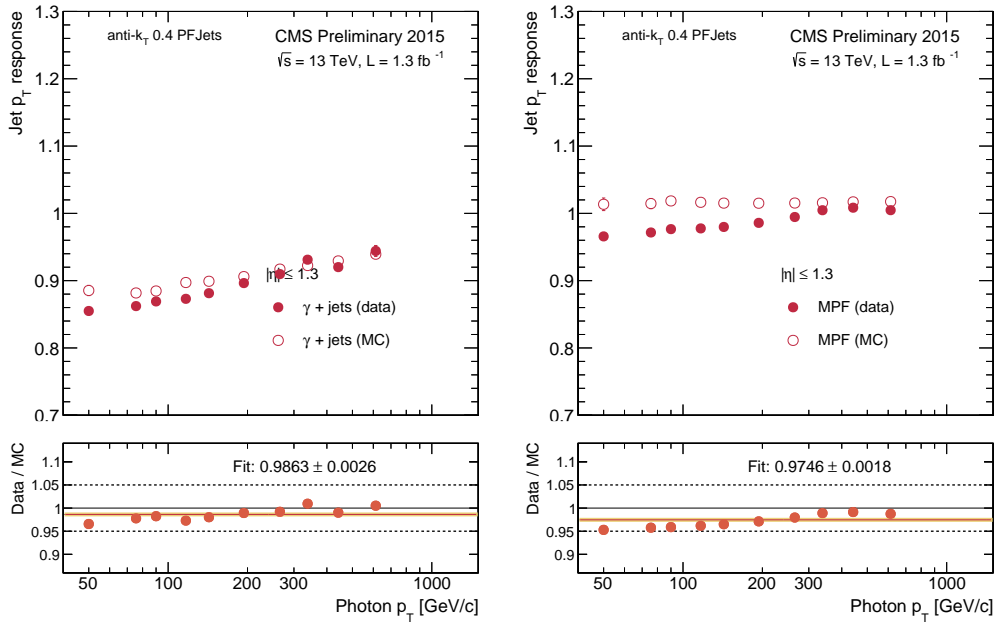
Figure A.5. Measurement of the response of PF ak4 jets in the CMS barrel ($|\eta| < 1.3|$). On the left, balancing response as a function of photon transverse momentum in data (filled circles) and simulation (empty circles). Data and simulation response versus the photon p_T for the MPF response is on the right. In both figures the data/MC ratio with a constant fit is shown at the bottom.

trend is then extrapolated to the ideal case of no secondary jet activity, with photon and leading jet perfectly balanced in the transverse plane.

In a given photon transverse momentum range, recalling expression A.2, which we may rewrite as

$$R_{\text{balancing}} = R_{\text{intr}} \cdot R_{\text{imb}} \tag{A.7}$$

we expect:

- the intrinsic response R_{intr} to be independent from the subleading jet (as long as it is 'reasonably' small), as it concerns only the leading jet;

- the imbalance R_{imb} to have a strong dependence on the subleading jet.

Our assumption is that these two effects are not correlated, so that they factorize, and therefore the response will have the simple expression:

$$\langle R_{\text{balancing}} \rangle = \langle R_{\text{intr}} \rangle \cdot \langle R_{\text{imb}} \rangle \tag{A.8}$$

Empirically we find that the functional dependence of R_{imb} on the subleading jet p_T is of quadratic form. Therefore, in a given photon p_T bin we will have:

$$\langle R_{\text{balancing}} \rangle (p_T^{\text{2ndJet}}) = c \cdot [1 - q - m(p_T^{\text{2ndJet}})^2] \tag{A.9}$$

therefore c is the object of this measurement, m describes the dependence of the imbalance on the subleading jet, and q quantifies the amount of irreducible imbalance between the photon and the leading jet.

The method's operation is shown in Fig. A.6, where the trends of the different contributions are shown as a function of the relative subleading jet transverse momentum (α), in four representative p_γ ranges. In each graph, the intrinsic response (black squares) and the imbalance (red triangles) can be seen, together with their fit functions. The product of these two functions is shown with a grey line, and, if the made assumptions are correct, should constitute the predicted trend for MC points (open red markers). The overall good agreement between the two is a confirmation of the validity of the method in simulation.

The measured trends in the data are also shown in each graph with solid red markers.The effect of the irreducible imbalance cannot be measured on data but must be accounted for, therefore the function used in the fit to the data has the functional form defined in Eq.A.9, but with the q parameter fixed to the value obtained on the simulation.

The summary of p_T balancing response measurements as a function of photon transverse momentum are shown in Fig. A.7. The plot shows the response values, in the data and in the simulation, for non-extrapolated (grey) and extrapolated balancing method (red). The true response is also shown (black markers).

A.5   Jet Transverse Momentum resolution measurement

The jet p_T resolution (JER) is relatively poor compared to resolution of many other physics objects (electron, muons, photons) and the biases caused by smearing can be important for steeply falling spectra and for resonance decays.
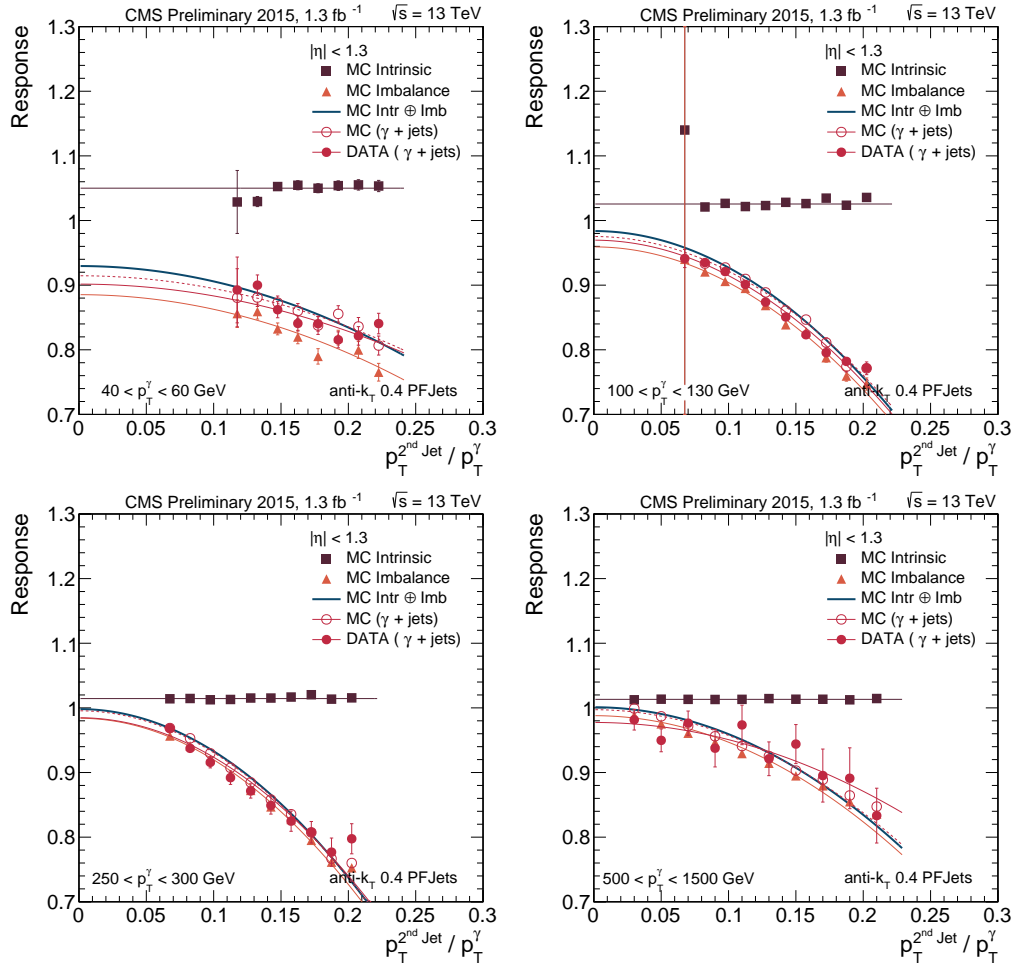
Figure A.6. Balancing response extrapolation in four representative transverse momentum ranges, for PF ak4 jets reconstructed in the barrel.
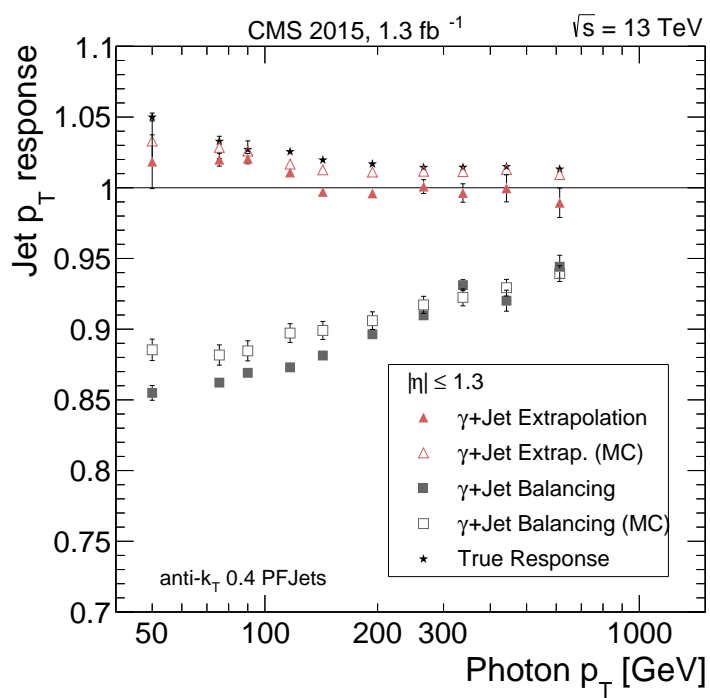
Figure A.7. Corrected response measurement, as a function of photon transverse momentum, for PF ak4 jets reconstructed in the barrel. Results for balancing before (grey) and after extrapolation (red) are shown both for data (solid) and the Monte Carlo simulation (hollow). A comparison to the true response (black markers) is also shown.

In this section we present a measurement of the JER obtained from γ+jet data. Measurement of jet p_T resolution is an extension of the methods used for measuring the jet energy scale, but instead of looking at the mean of the response distribution, we are interested in its width. The jets are corrected for JEC before deriving JER.

Recalling Eq.A.7, our assumptions are that, in a given p_T^γ bin, the intrinsic resolution is independent of p^{2ndJet}, whereas the imbalance effect to be linear. In formulas:

$$\sigma_{\mathrm{intr}}(p_T^{\mathrm{2ndJet}}) = c' \tag{A.10a}$$

$$\sigma_{\mathrm{imb}}(p_T^{\mathrm{2ndJet}}) = q' + m' \cdot p_T^{\mathrm{2ndJet}} \tag{A.10b}$$

$$\sigma_{\mathrm{balancing}}(p_T^{\mathrm{2ndJet}}) = \sqrt{c'^2 + q'^2 + q'm'p_T^{\mathrm{2ndJet}} + m'^2} \tag{A.10c}$$

being c', m' and q' the free parameters of the fit.

The performance of the method is shown in Fig. A.8, for the data and the simulation, in four representative p_T^γ bins. The color coding is the same as in the response case. Again, the good agreement between the "predicted" trend (grey line) and the reconstructed MC estimates (open red circles) proves the internal consistency of the method. The data points are fitted with the expected functional form, and, similarly as in the response case, the contribution of the irreducible imbalance (q') is fixed to the value fitted in the MC.

The results of the corrected jet p_T resolution as a function of transverse momentum are shown in Fig A.9. The plot shows the resolution before (grey squares) and after extrapolation (red circles), in data (solid) and MC (hollow), and compares them to the true resolution. The effect of the extrapolation is to improve significantly the measured resolution, bringing it close to the true resolution measured on Monte Carlo.
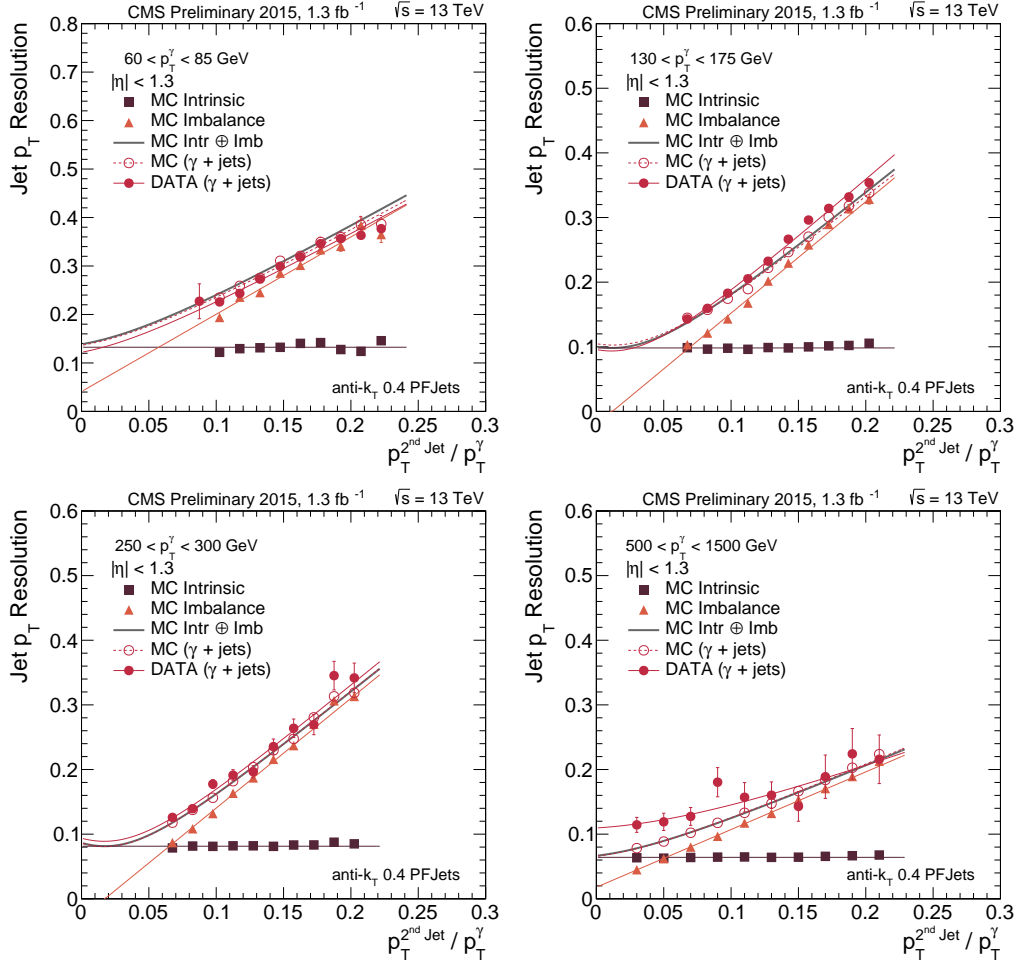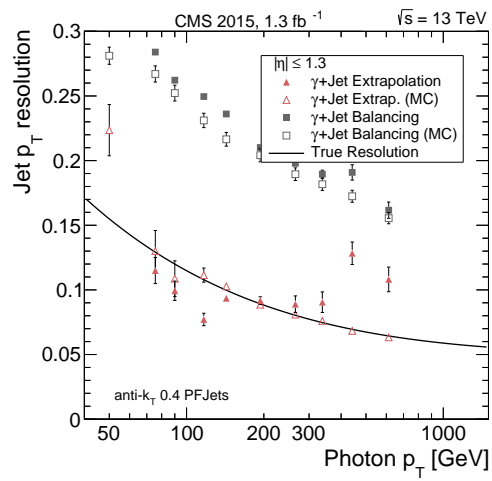
Figure A.8. Response extrapolation in four representative transverse momentum ranges, for PF ak4 jets reconstructed in the barrel.

Figure A.9. Resolution measurement for PF ak4 jets reconstructed in the barrel. The plot shows the resolution before (grey squares) and after extrapolation (red circles), in data (solid) and MC (hollow). Also the true resolution is shown for comparison.

Bibliography

[1] S. L. Glashow, "Partial-symmetries of weak interactions," *Nuclear Physics*, vol. 22, no. 4, pp. 579 – 588, 1961.

[2] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967.

[3] A. Salam, "Weak and Electromagnetic Interactions," *Conf. Proc.*, vol. C680519, pp. 367–377, 1968.

[4] I. G. Knowles, M. Schmelling, and G. Dissertori, *Quantum chromodynamics high energy experiments and theory.* Oxford Scholarship Online; International series of monographs on physics, Oxford Oxford Univ. Press, 2009.

[5] J. M. Campbell, J. W. Huston, and W. J. Stirling, "Hard Interactions of Quarks and Gluons: A Primer for LHC Physics," *Rept. Prog. Phys.*, vol. 70, p. 89, 2007.

[6] S. White, "Luminosity Scans at the LHC. Luminosity Scans at LHC," 2011.

[7] The CMS Collaboration, "Measurement of CMS Luminosity," *CMS Physics Analysis Summary*, vol. CMS-PAS-EWK-10-004, 2010.

[8] V. Khachatryan *et al.*, "Search for resonances and quantum black holes using dijet mass spectra in proton-proton collisions at $\sqrt{s} = 8$ TeV," *Phys. Rev. D*, vol. 91, p. 052009, 2015.

[9] U. Baur, I. Hinchliffe, and D. Zeppenfeld, "Excited Quark Production at Hadron Colliders," *Int. J. Mod. Phys.*, vol. A2, p. 1285, 1987.

[10] U. Baur, M. Spira, and P. M. Zerwas, "Excited-quark and -lepton production at hadron colliders," *Phys. Rev. D*, vol. 42, pp. 815–824, Aug 1990.

[11] L. Randall and R. Sundrum, "A Large mass hierarchy from a small extra dimension," *Phys. Rev. Lett.*, vol. 83, no. MIT-CTP-2860, PUPT-1860, BUHEP-99-9, pp. 3370–3373, 1999.

[12] L. Randall and R. Sundrum, "An alternative to compactification," *Phys. Rev. Lett.*, vol. 83, pp. 4690–4693, Dec 1999.

[13] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo, "Phenomenology of the randall-sundrum gauge hierarchy model," *Phys. Rev. Lett.*, vol. 84, pp. 2080–2083, Mar 2000.

[14] J. Bijnens, P. Eerola, M. Maul, A. Mansson, and T. Sjostrand, "QCD signatures of narrow graviton resonances in hadron colliders," *Phys. Lett.*, vol. B503, pp. 341–348, 2001.

[15] P. H. Frampton and S. L. Glashow, "Chiral color: An alternative to the standard model," *Physics Letters B*, vol. 190, no. 1–2, pp. 157 – 161, 1987.

[16] J. Bagger, C. Schmidt, and S. King, "Axigluon production in hadronic collisions," *Phys. Rev. D*, vol. 37, pp. 1188–1196, Mar 1988.

[17] R. S. Chivukula, A. G. Cohen, and E. H. Simmons, "New strong interactions at the Tevatron?," *Phys. Lett.*, vol. B380, pp. 92–98, 1996.

[18] E. H. Simmons, "Coloron phenomenology," *Phys. Rev.*, vol. D55, pp. 1678–1683, 1997.

[19] T. Han, I. Lewis, and Z. Liu, "Colored resonant signals at the lhc: largest rate and simplest topology," *Journal of High Energy Physics*, vol. 2010, no. 12, 2010.

[20] E. Eichten, I. Hinchliffe, K. Lane, and C. Quigg, "Supercollider physics," *Rev. Mod. Phys.*, vol. 56, pp. 579–707, Oct 1984.

[21] L. A. Anchordoqui, H. Goldberg, D. Lüst, S. Nawata, S. Stieberger, and T. R. Taylor, "Dijet signals for low mass strings at the large hadron collider," *Phys. Rev. Lett.*, vol. 101, p. 241803, Dec 2008.

[22] S. Cullen, M. Perelstein, and M. E. Peskin, "TeV strings and collider probes of large extra dimensions," *Phys. Rev.*, vol. D62, p. 055012, 2000.

[23] P. Candelas, G. T. Horowitz, A. Strominger, and E. Witten, "Vacuum configurations for superstrings," *Nuclear Physics B*, vol. 258, pp. 46 – 74, 1985.

[24] J. L. Hewett and T. G. Rizzo, "Low-energy phenomenology of superstring-inspired {E6} models," *Physics Reports*, vol. 183, no. 5-6, pp. 193 – 381, 1989.

[25] G. Katsilieris, O. Korakianitis, and S. Vlassopulos, "Two-jet rates affected by exotic particles at hadron colliders," *Physics Letters B*, vol. 288, no. 1-2, pp. 221 – 226, 1992.

[26] V. Angelopoulos, J. Ellis, H. Kowalski, D. Nanopoulos, N. Tracas, and F. Zwirner, "Search for new quarks suggested by the superstring," *Nuclear Physics B*, vol. 292, pp. 59 – 92, 1987.

[27] C. Albajar, M. Albrow, and O. Allkofer, "Two-jet mass distributions at the {CERN} proton-antiproton collider," *Physics Letters B*, vol. 209, no. 1, pp. 127 – 134, 1988.

[28] F. Abe *et al.*, "The Two jet invariant mass distribution at $\sqrt{s} = 1.8$ TeV," *Phys. Rev.*, vol. D41, pp. 1722–1725, 1990.

[29] J. Alitti *et al.*, "A Measurement of two jet decays of the W and Z bosons at the CERN $\bar{p}p$ collider," *Z. Phys.*, vol. C49, pp. 17–28, 1991.

[30] F. Abe *et al.*, "Search for quark compositeness, axigluons and heavy particles using the dijet invariant mass spectrum observed in $p\bar{p}$ collisions," *Phys. Rev. Lett.*, vol. 71, pp. 2542–2546, 1993.

[31] J. Alitti *et al.*, "A Search for new intermediate vector mesons and excited quarks decaying to two jets at the CERN $\bar{p}p$ collider," *Nucl. Phys.*, vol. B400, pp. 3–24, 1993.

[32] F. Abe *et al.*, "Search for new particles decaying to dijets in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV," *Phys. Rev. Lett.*, vol. 74, p. 3538, 1995.

[33] F. Abe *et al.*, "Search for new particles decaying to dijets at CDF," *Phys. Rev. D*, vol. 55, p. 5263, 1997.

[34] V. Abazov *et al.*, "Search for new particles in the two-jet decay channel with the dø detector," *Phys. Rev. D*, vol. 69, p. 111101, Jun 2004.

[35] T. Aaltonen *et al.*, "Search for new particles decaying into dijets in proton-antiproton collisions at $\sqrt{s} = 1.96$ tev," *Phys. Rev. D*, vol. 79, p. 112002, 2009.

[36] G. Aad *et al.*, "Search for New Particles in Two-Jet Final States in 7 TeV Proton-Proton Collisions with the ATLAS Detector at the LHC," *Phys. Rev. Lett.*, vol. 105, p. 161801, 2010.

[37] V. Khachatryan *et al.*, "Search for Dijet Resonances in 7 TeV pp Collisions at CMS," *Phys. Rev. Lett.*, vol. 105, p. 211801, 2010.

[38] G. Aad *et al.*, "Search for New Physics in Dijet Mass and Angular Distributions in pp Collisions at $\sqrt{s} = 7$ TeV Measured with the ATLAS Detector," *New J. Phys.*, vol. 13, p. 053044, 2011.

[39] S. Chatrchyan *et al.*, "Search for Resonances in the Dijet Mass Spectrum from 7 TeV pp Collisions at CMS," *Phys. Lett. B*, vol. 704, p. 123, 2011.

[40] G. Aad *et al.*, "Search for New Physics in the Dijet Mass Distribution using 1 fb^{-1} of pp Collision Data at $\sqrt{s} = 7$ TeV collected by the ATLAS Detector," *Phys. Lett. B*, vol. 708, p. 37, 2012.

[41] S. Chatrchyan *et al.*, "Search for narrow resonances using the dijet mass spectrum in pp collisions at $\sqrt{s} = 8$ TeV," *Phys. Rev. D*, vol. 87, p. 114015, 2013.

[42] G. Aad *et al.*, "Search for new phenomena in the dijet mass distribution using pp collision data at $\sqrt{s} = 8$ TeV with the ATLAS detector," *Phys. Rev. D*, vol. 91, p. 052007, 2015.

[43] V. Khachatryan *et al.*, "Search for narrow resonances decaying to dijets in proton-proton collisions at sqrt(s) = 13 TeV," 2015.

[44] "Search for New Phenomena in Dijet Mass and Angular Distributions with the ATLAS Detector at $\sqrt{s} = 13$ TeV," 2015.

[45] R. M. Harris and K. Kousouris, "Searches for Dijet Resonances at Hadron Colliders," *Int. J. Mod. Phys.*, vol. A26, pp. 5005–5055, 2011.

[46] T. C. Collaboration, "Search for Resonances Decaying to Dijet Final States at $\sqrt{s} = 8$ TeV with Scouting Data," 2015.

[47] B. A. Dobrescu and F. Yu, "Coupling-mass mapping of dijet peak searches," *Phys. Rev.*, vol. D88, no. 3, p. 035021, 2013. [Erratum: Phys. Rev.D90,no.7,079901(2014)].

[48] G. Salam and A. Weiler, "Collider reach." http://collider-reach.web.cern.ch/collider-reach/.

[49] The ATLAS Collaboration, "The atlas experiment at the cern large hadron collider," *Journal of Instrumentation*, vol. 3, no. 08, p. S08003, 2008.

[50] The CMS Collaboration, "The cms experiment at the cern lhc," *Journal of Instrumentation*, vol. 3, no. 08, p. S08004, 2008.

[51] The ALICE Collaboration, "The alice experiment at the cern lhc," *Journal of Instrumentation*, vol. 3, no. 08, p. S08002, 2008.

[52] The LHCb Collaboration, "The lhcb detector at the lhc," *Journal of Instrumentation*, vol. 3, no. 08, p. S08005, 2008.

[53] The CMS Collaboration, *The CMS magnet project: Technical Design Report.* Technical Design Report CMS, Geneva: CERN, 1997.

[54] The CMS Collaboration, *The CMS electromagnetic calorimeter project: Technical Design Report.* Technical Design Report CMS, Geneva: CERN, 1997.

[55] The CMS Collaboration, *The CMS hadron calorimeter project: Technical Design Report.* Technical Design Report CMS, Geneva: CERN, 1997.

[56] T. C. Collaboration, "CMS, the Compact Muon Solenoid. Muon technical design report," 1997.

[57] S. Dasu *et al.*, "CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems," 2000.

[58] S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit," *Nucl. Instrum. Meth.*, vol. A506, pp. 250–303, 2003.

[59] The CMS Collaboration, "CMS physics Technical Design Report, Volume II: Physics Performance.," *J. Phys. G*, vol. 34, no. CERN-LHCC-2006-021. CMS-TDR-008-2, pp. 995–1579. 669 p, 2006. revised version submitted on 2006-09-22 17:44:47.

[60] R. Wigmans, *Energy Measurements in Particle Physics.* Oxford Science Publications, 2000.

[61] S. Abdullin, V. Abramov, and Acharya, "The cms barrel calorimeter response to particle beams from2to350 gev/c," *The European Physical Journal C*, vol. 60, no. 3, pp. 359–373, 2009.

[62] The CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET," Tech. Rep. CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr 2009.

[63] M. Cacciari, G. P. Salam, and G. Soyez, "The anti- k t jet clustering algorithm," *Journal of High Energy Physics*, vol. 2008, no. 04, p. 063, 2008.

[64] S. Chatrchyan *et al.*, "Determination of jet energy calibration and transverse momentum resolution in cms," *JINST*, vol. 6, p. P11002, 2011.

[65] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1," *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008.

[66] S. Chatrchyan *et al.*, "Search for narrow resonances and quantum black holes in inclusive and *b*-tagged dijet mass spectra from pp collisions at $\sqrt{s} = 7$ TeV," *JHEP*, vol. 01, p. 013, 2013.

[67] The CMS Collaboration, "Search for Narrow Resonances using the Dijet Mass Spectrum with 42 pb^{-1} of pp Collisions at $\sqrt{s} = 13$ TeV," Tech. Rep. CMS-PAS-EXO-15-001, CERN, Geneva, 2015.

[68] R. G. Lomax and D. L. Hahs-Vaughn, "Statistical concepts: A second course," *Routledge Academic, London, 2007.*

[69] L. A. Anchordoqui, H. Goldberg, D. Lust, S. Nawata, S. Stieberger, *et al.*, "Dijet signals for low mass strings at the LHC," *Phys. Rev. Lett.*, vol. 101, p. 241803, 2008.

[70] S. Cullen, M. Perelstein, and M. E. Peskin, "TeV strings and collider probes of large extra dimensions," *Phys. Rev. D*, vol. 62, p. 055012, 2000.

[71] J. L. Hewett and T. G. Rizzo, "Low-Energy Phenomenology of Superstring Inspired E(6) Models," *Phys. Rept.*, vol. 183, p. 193, 1989.

[72] U. Baur, I. Hinchliffe, and D. Zeppenfeld, "Excited quark production at hadron colliders," *Int. J. Mod. Phys. A*, vol. 2, p. 1285, 1987.

[73] U. Baur, M. Spira, and P. Zerwas, "Excited quark and lepton production at hadron colliders," *Phys. Rev. D*, vol. 42, p. 815, 1990.

[74] P. H. Frampton and S. L. Glashow, "Chiral color: An alternative to the standard model," *Phys. Lett. B*, vol. 190, p. 157, 1987.

[75] R. S. Chivukula, A. Farzinnia, J. Ren, and E. H. Simmons, "Hadron Collider Production of Massive Color-Octet Vector Bosons at Next-to-Leading Order," *Phys. Rev. D*, vol. 87, p. 094011, 2013.

[76] E. H. Simmons, "Coloron phenomenology," *Phys. Rev. D*, vol. 55, p. 1678, 1997.

[77] T. Han, I. Lewis, and Z. Liu, "Colored resonant signals at the lhc: Largest rate and simplest topology," *JHEP*, vol. 12, p. 085, 2010.

[78] E. Eichten, I. Hinchliffe, K. D. Lane, and C. Quigg, "Super collider physics," *Rev. Mod. Phys.*, vol. 56, p. 579, 1984.

[79] L. Randall and R. Sundrum, "An alternative to compactification," *Phys. Rev. Lett.*, vol. 83, p. 4690, 1999.

[80] The CMS Collaboration, "First Results from Dijet Resonance Search using 37 pb^{-1} of proton-proton collisions at $\sqrt{s}$=13 TeV," Jul 2015.

[81] F. Pandolfi, *Search for the Standard Model Higgs Boson in the $H \rightarrow ZZ \rightarrow l^+l^-q^+q^-$ Decay Channel at CMS*. PhD thesis, Università La Sapienza, Roma, 2011.

Acknoledgements

È giunto il momento di ringraziare le persone che hanno contribuito più o meno direttamente, più o meno consapevolmente al compimento di questo lavoro.

Il primo grazie lo devo al mio relatore, Shahram, e ai "senior" del gruppo CMS Roma con cui mi sono trovata a lavorare più da vicino in questi 3 anni: Chiara e Paolo che mi hanno seguito all'inizio del mio lavoro di dottorato, Francesca che mi ha trasmesso una piccola parte della sua sapienza su ECAL, Daniele e in particolare Francesco S., che hanno guidato e seguito il mio lavoro più da vicino nell'ultimo anno. Con il continuo confronto (e le numerose richieste) di tutti voi ho imparato molto su cosa significa fare fisica e lavorare in gruppo.

Devo ringraziare l'INFN per l'opportunità che ho avuto di passare periodi più o meno lunghi negli anni scorsi e l'ultimo anno interamente al CERN (e, devo sottolinearlo, sempre adeguatamente retribuita). L'esperienza di lavorare da vicino all'esperimento a cui collaboro, insieme a fisici di tutto il mondo, e in un momento abbastanza *unico* per la fisica ad LHC, è stata un'occasione di crescita umana e professionale che non avrei avuto modo di fare altrimenti.

Ringrazio anche il Prof. Longo per avermi dato l'opportunità di fare una piccola esperienza nella didattica, come assistente al suo corso di Fisica Nucleare e Subnucleare. È stato divertente e anche molto istruttivo passare dall'altra parte della cattedra.

Grazie ai miei genitori, che mi hanno sempre sostenuta e incoraggiata nelle mie scelte, senza mai giudicare né interferire.

Grazie speciale a Vi, che è presenza costante e importante nella mia vita, qualunque cosa io faccia, dovunque io mi trovi.

Grazie a Livia, che mi ha fatto da guida fin dall'inizio della mia avventura al CERN, quando sono arrivata la prima volta come Summer Student nel lontano 2011. Da allora si sono susseguite una serie innumerevole di cene, ospitate, gite fuori porta, e anche nottate di lavoro insieme. Il tempo che ho passato al CERN negli ultimi 4 anni sarebbe stato vuoto senza tutto questo. Grazie!

Grazie a Francesco M. per avermi ispirato nell'attitudine verso il lavoro, soprattutto al CERN, dove la tentazione di identificare sé stessi con la propria analisi di fisica e il rischio di alienazione certe volte diventano forti. Quindi grazie per le cene a casa di amici non-fisici, i giri nei negozi di fumetti e cd usati, e i concerti improbabili a Ginevra.

Grazie a ivanotalamo (minuscolo e tutt'attaccato) per un sacco di cose. Intanto per avermi pazientemente aiutato con le mie infinite richieste sul tier 2: alla fine ho imparato ad essere una brava utente, e senza il tuo pronto intervento in molte occasioni avrei sprecato un sacco di lavoro e di tempo. E poi grazie per i pranzi al 32, le

chiacchierate sui libri "con i disegnetti", i fumetti, l'illustrazione, e per la testimonianza che si può continuare a perseguire seriamente un'aspirazione professionale in questo campo pur facendo un lavoro completamente scorrelato.

Infine condenso in queste righe i ringraziamenti alle altre persone con cui ho condiviso parte di questo percorso in vari modi: gli amici che conosco da anni, quelli conosciuti da poco, quelli che ho continuato a sentire spesso, e quelli che sento poco ma appena capita è come se ci fossimo salutati il giorno prima. Non vi elenco ad uno ad uno perché sarebbe lungo, e perché finirei sicuramente per dimenticarmi qualcuno, ma penso che possiate riconoscervi. Grazie collettivo anche a voi!

For even the very wise cannot see all ends.
J.R.R. Tolkien, The Fellowship of the Ring