SAPIENZA
UNIVERSITÀ DI ROMA

Search for new resonances decaying to pairs of jets in proton-proton collisions at $\sqrt{s} = 13$ TeV

Facoltà di Scienze Matematiche Fisiche e Naturali

Dottorato di Ricerca in Fisica – XXX Ciclo

Candidate

Federico Preiato

ID number 1159893

Thesis Advisor

Dr. Chiara Ilaria Rovelli

A thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Fisica

2018

Thesis defended on February 01, 2018
in front of a Board of Examiners composed by:

Prof. Stefano Giagu

Prof. Marco Serone

Prof. Paolo Aschieri

Search for new resonances decaying to pairs of jets in proton-proton collisions at $\sqrt{s} = 13$ TeV

Ph.D. thesis. Sapienza – University of Rome

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

Author's email: federico.preiato@roma1.infn.it

To my wife,
she gave me the greatest gift
anyone could give another person,
she believed in me.

Contents

Introduction

In this thesis a search for physics beyond the Standard Model (SM) at the Large Hadron Collider (LHC) is presented.

The Standard Model of particle physics is the theory developed throughout the second half of the 20^{th} century describing the elementary particles and their interactions. It summarizes our current understanding of particle interactions and, from an experimental point-of-view, all the measurements performed in the last decades verified at a high precision level its predictions, without evidence of physics beyond the SM. Despite its experimental success, the SM is widely considered an incomplete theory for many reasons. To give some examples: it does not incorporate gravity; it does not explain the asymmetry between matter and antimatter; it does not contain any viable dark matter particle with all the required properties deduced from cosmology.

Several alternative theories have been elaborated to include the SM and to overtake such open issues in a more general theoretical framework. Many of these theories predict the existence of new particles at the TeV scale, which therefore could be accessible at the LHC.

This thesis describes a model independent search for resonances decaying into a pair of quarks and/or gluons which appears in the detector as a pair of hadronic jets. The experimental signature is a bump in the dijet invariant mass distribution close to the resonance mass, over a falling background.

The analysis is based on data recorded by the CMS experiment in 2016, corresponding to an integrated luminosity of 36 fb^{-1} with an energy in the center-of-mass of 13 TeV.

The thesis is organized in the following way. The first two chapters give an overview of the physics motivations, of LHC and CMS. Chapters 3, 4 and 5 summarize my own work and the original contributions I gave.

An overview about the Standard Model, with particular attention to elements of Quantum Chromodynamics, is briefly given in Chapter 1. The main theories beyond the SM that have been probed with this search are summarized as well in this chapter.

Chapter 2 is dedicated to the experimental apparatus, with the description of the LHC and of the CMS experiment with all its subdetectors.

The most important objects for the dijet search are the hadronic jets. The general aspects of the jet reconstruction are described in Chapter 3, where the jet energy calibration procedure used in CMS is also described. Large part of my Ph.D. work was devoted to the jet energy calibration using photon+jet events, crucial to the measurement and the correction of the jet energy scale, and to the energy

corrections for non-standard jets needed for a part of the analysis.

The last two chapters are dedicated to the description of the search for new particles decaying into a pair of jets. In Chapter 4 the analysis strategy is reported. The data and simulation samples, selection criteria and quality checks are described, along the fit strategy in order to estimate the background, that is my main contribute to the analysis. The results and their statistical interpretation are presented in Chapter 5, which also discusses the constraints on dark matter posed by this search.

Chapter 1

The Standard Model and beyond

1.1 The Standard Model

The Standard Model of particle physics is the theory describing the physics of elementary particles in terms of quantum fields. The first step towards the Standard Model was Sheldon Glashow's unification of the electromagnetic and weak interactions [1]. Few years later Steven Weinberg [2] and Abdus Salam [3] incorporated the Higgs mechanism into Glashow's electroweak interaction, giving to the theory its modern form. The SM describes in one coherent framework three of the four fundamental interactions: electromagnetic, weak and strong forces, therefore not including the gravity. The particles described by this theory can be divided in two groups:

- Fermions: Particles with half-integer spin obeying to the Fermi-Dirac statistics. Six leptons and six quarks (in three different colors) compose this group, representing the matter fields.
- Bosons: Particles with integer spin obeying to the Bose-Einstein statistics, mediators of the fundamental forces.

A schematic representation of the SM particles is reported in Figure 1.1.

The Lagrangian of the theory is invariant under the non-Abelian gauge symmetry group:

$$SU(3)_C \times SU(2)_L \times U(1)_Y \tag{1.1}$$

where $SU(3)_C$ indicates the color symmetry, while the invariance under $SU(2)_L \times U(1)_Y$ describes the unification between the electromagnetic and weak forces. The Lagrangian can be written as the sum of two contributes: one describing the strong interaction (Quantum Chromodynamics or QCD), and one the electroweak interactions (EW):

$$L_{SM} = L_{QCD} + L_{EW} \tag{1.2}$$

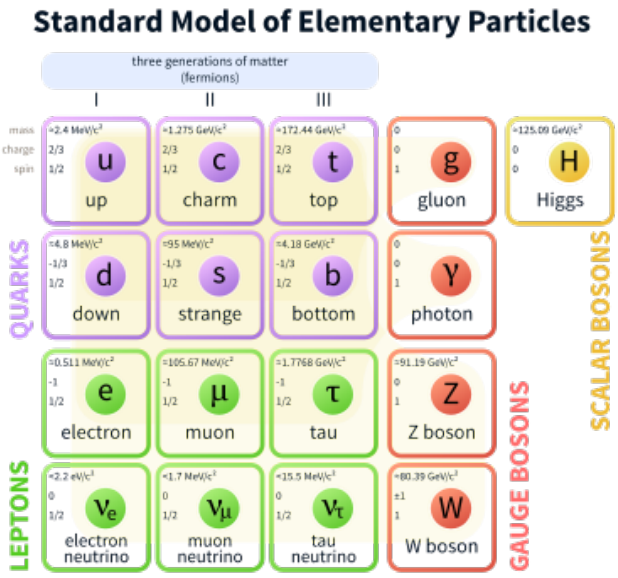

Figure 1.1. The Standard Model of elementary particles, with the three generations of fermions, the gauge bosons in the fourth column, and the Higgs boson in the fifth.

For the search presented in this thesis, the relevant part is the QCD contribution, therefore only this sector will be described in the next paragraphs.

The QCD is the theory which describes the strong interactions between quarks in terms of exchange of gluons which are the force carriers of the theory, like the photons are for the electromagnetic force in quantum electrodynamics.

The gauge invariant QCD Lagrangian can be written as:

$$L_{QCD} = \sum_i \bar{q}_{i,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t^A_{ab} G^A_\mu - m_i \delta_{ab}) q_{i,b} - \frac{1}{4} F^A_{\mu\nu} F^{\mu\nu,A} \tag{1.3}$$

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

$$F^A_{\mu\nu} = \partial_\mu G^A_\nu - \partial_\nu G^A_\mu - g_s f_{ABC} G^B_\mu G^C_\nu \tag{1.4}$$

f_{ABC} are the structure constants satisfying the relation:

$$\left[t^A, t^B\right] = i f_{ABC} t^C \tag{1.5}$$

The QCD is characterized by two peculiar properties: the confinement and the asymptotic freedom. Due to the confinement when two quarks are separated the energy in the gluon field is enough to create another quark pair. For this reason, only color-singlet states can be directly observed and therefore a quark or a gluon cannot exist as free particle. The asymptotic freedom means that the bonds between particles become asymptotically weaker as energy increases and distance decreases. Quarks interact weakly at high energies, allowing perturbative calculations of the

cross sections in deep inelastic processes, and strongly at low energies, preventing the unbinding of baryons.

An exhaustive description of the QCD can be found for example in [4].

1.2 Theories beyond the Standard Model

Many theories beyond the Standard Model predict the existence of new particles at the TeV scale. This thesis focuses on the search for processes producing resonances (denoted by X) decaying to pair of jets as illustrated in Figure 1.2.

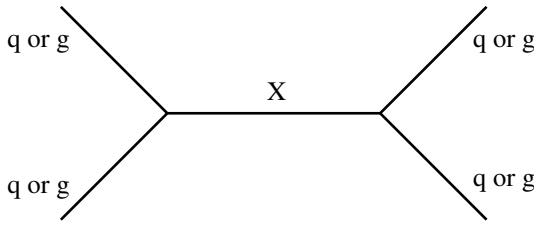

Figure 1.2. Feynman Diagram for a resonance decaying to pair of jets. The initial state and final state both contain two partons (quarks, antiquarks or gluons) and the intermediate state contains an s-channel resonance X.

Despite the approach of the search is basically model-independent, the results are compared with some theoretical models described briefly in the next sections and summarized in Table 1.1.

String resonances

The model with the largest cross section for the production of new particles is a model of string resonances [5] [6], which are Regge excitations of quarks and gluons in string theory decaying predominantly to qg. The superstring theory offers one of the most robust scenarios beyond the Standard Model, provided that its fundamental mass scale is sufficiently "low", i.e. of order TeV. The mass scale M_s of the fundamental strings can be as low as a few TeV provided that space-time extends into large extra dimensions, allowing a novel solution for the hierarchy problem. This mass determines the energy threshold $\sqrt{\hat{s}} \geq M_s$ for the production of Regge resonances in hadron collisions, where $\hat{s}$ is the Mandelstam variable as explained in Section 3.1.1.

In order to rely on the perturbation theory in the computation of the scattering amplitudes, the string coupling is assumed to be small. Under this assumption, the $2 \to 2$ scattering amplitudes involving four gluons as well as those with two gluons plus two quarks do not depend on the compactification details and are completely model-independent (approximately true also for the four-quark scattering). All string effects are encapsulated in these amplitudes in one "form factor" function of Mandelstam variables $\hat{s}$, $\hat{t}$, $\hat{u}$:

$$V(\hat{s},\hat{t},\hat{u}) = \frac{\Gamma(1-\hat{s}/M_s^2)\Gamma(1-\hat{u}/M_s^2)}{\Gamma(1-\hat{t}/M_s^2)} \tag{1.6}$$

The physical content of this form factor becomes clear after using the expansion in terms of s-channel resonances, which shows s-channel poles associated to virtual Regge excitations with masses $\sqrt{n}M_s$.

Only the first-level ($n = 1$) excitation is relevant for the purpose to have resonances in the dijet spectrum, while the string mass M_s is the only free parameter. The exact values of the cross sections also depend on the color factors and the spin values of the excited states.

Excited quarks

Many theories beyond the SM are based on the assumption that the quarks are composite objects [7]. The most compelling evidence of a quark substructure would be provided by the discovery of an excited state of a quark. An excited quark (denoted by q^*) may couple to an ordinary quark and a gauge boson as a gluon via gauge interactions given by the Lagrangian:

$$L = \frac{1}{2\Lambda}\bar{q}^*_R\sigma^{\mu\nu}\left[g_s f_s t_a G^a_{\mu\nu} + gf\frac{\vec{\tau}\cdot\vec{W}_{\mu\nu}}{2} + g'f'\frac{Y}{2}B_{\mu\nu}\right]q_L + h.c. \tag{1.7}$$

where q^*_R is the excited quark field, $\sigma_{\mu\nu}$ is the Pauli spin matrix, q_L is the quark field, $G^a_{\mu\nu}$, $\vec{W}_{\mu\nu}$ and $B_{\mu\nu}$ are the field tensors of SU(3), SU(2) and U(1) gauge fields, t_a and $\vec{\tau}$ are the generators of the color SU(3) and isospin SU(2), Y is the hypercharge and g_s, g and g' are the gauge coupling constants. The compositeness scale, Λ, is the typical energy scale of these interactions, and f_s, f and f' are unknown dimensionless constants determined by the compositeness dynamics, which represent the strengths of the excited quark couplings to the SM partners and are usually assumed to be of order unity.

The production of q^* via gauge interactions would proceed through quark-gluon (qg) annihilation. In the context of this analysis, excited quarks would then decay into a quark and a gluon and they would appear as resonances in the invariant mass distribution of the decay products. The partial width for the decay of an excited quark with mass m^* is given by:

$$\Gamma(q^* \to qg) = \frac{1}{3}\alpha_s f_s^2 \frac{{m^*}^3}{\Lambda^2} \tag{1.8}$$

where α_s is the strong coupling constant.

Colorons and Axigluons

An intriguing possibility for new physics is that an extended color gauge sector may exist. In particular there are models theorizing the existence of new colored states beyond the familiar quarks and gluons [8]. One class of theories beyond the SM is that in which the strong interactions are extended from the standard $SU(3)_{QCD}$ to a larger $SU(3)_1 \times SU(3)_2$ group and in which a spontaneous symmetry breaking reduces the larger group to its diagonal subgroup which is identified with $SU(3)_{QCD}$. Each of these models includes new heavy colored gauge bosons, as colorons or axigluons, transforming as a color octet.

An octet of massive *colorons* [9], together with an octet of massless gluons, is formed by the mixing of the initial gauge bosons in the rotated phase of the physical gauge fields. The Lagrangian of the interaction between the colorons field $C^{a\mu}$ and the quarks is similar to the one of QCD:

$$L_C = -g_s cot\theta \left[\sum_{ij} \bar{q}_i \gamma_\mu t_a q_j\right] C^{a\mu} \tag{1.9}$$

where $g_s = \sqrt{4\pi\alpha_s}$, θ is the gauge boson mixing angle and t_a are the color group generators.

The decay width can be written as:

$$\Gamma_C \approx \frac{N}{6}\alpha_s cot^2\theta M_C \tag{1.10}$$

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

The *axigluons* are predicted by the Chiral Color model [10] which extends the QCD to the gauge group $SU(3)_L \times SU(3)_R$ leading to a second octet of force carriers. The exact implementation of the chiral color group is model dependent, but 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. The axigluon production requires an antiquark in the initial state ($q\bar{q} \to A$), slightly reducing the cross section compared to the excited quarks. The axigluon can not decay to a gluon-gluon pair due to the parity conservation, but it can decay to quark-antiquark pairs, with produce the dijet final state. The axigluon decay to fermions is described by the Lagrangian:

$$L_A = -ig_s \left[\sum_{ij} \bar{q}_i \gamma_5 \gamma_\mu t_a q_j\right] A^{a\mu} \tag{1.11}$$

where A is the axigluon field. The width of the axigluon decay can be written as:

$$\Gamma_A = \frac{N\alpha_s M_A}{6} \tag{1.12}$$

where N indicates the open decay channels and M_A is the axigluon mass, which is a free parameter of the theory.

E6 Diquarks

Within the context of superstrings in ten dimensions, the gauge and gravitational anomaly cancellation (which is necessary for a finite theory) was found to occur only for the gauge groups $SO(32)$ or $E8 \times E8$. This last group leads to chiral fermions (as exist in the SM), whereas $SO(32)$ does not, so $E8 \times E8$ can contain the SM in its usual form and therefore it is the most attractive phenomenologically. To make connection with our four-dimensional world, the extra six dimensions must be compactified on some kind of manifold and several compactification scenarios are possible. The compactification of the additional six dimensions can lead to E6 as an

"effective" Grand Unification Theory (GUT) group. This motivation has inspired the interest in E6 and its subgroups as sources of physics beyond the SM.

The E6 models contain color-triplet scalar diquarks [11], D and D^c, produced with electromagnetic coupling from the valence quarks of the proton ($ud \to D$). The cross section for E6 diquarks is large despite the relatively weak coupling because of the large parton distribution function (PDF) for valence quarks: at high parton momentum the probability of finding a quark in the proton is significantly larger than the probability of finding a gluon or antiquark.

The Lagrangian for this kind of interaction is:

$$L_D = \lambda \epsilon_{ijk} \bar{u}^{ci} \frac{1}{2}(1-\gamma_5) d^j D^k + \frac{\lambda_c}{2} \epsilon_{ijk} \bar{u}^i \frac{1}{2}(1+\gamma_5) d^{cj} D^{ck} + h.c. \tag{1.13}$$

where i, j, k are the color indices, while λ and λ_c are hyper-potential parameters of the E6 model.

The widths of the decays are:

$$\Gamma_D = \alpha_\lambda M_D \tag{1.14}$$

and

$$\Gamma_{D_c} = \alpha_{\lambda_c} M_{D^c} \tag{1.15}$$

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

Color-octet scalars

New physics at the weak scale typically gives rise to unacceptably large flavor changing neutral currents if it can couple at tree level to the quarks. An attractive way to avoid this problem is to impose the principle of minimal flavor violation [12]. Motivated by this principle, some models extend the SM by one family of color-octet scalars [13]. In this thesis the decay channel into a pair of gluons is considered, for which the cross section is relatively large because of the strong coupling.

The Lagrangian of the interaction can be written as:

$$L = g_s \frac{k_s}{\Lambda} d^{abc} S_8^a G_{\mu\nu}^b G^{c\mu\nu} \tag{1.16}$$

where g_s is the strong coupling constant, k_s is the scalar coupling, Λ is the interaction energy scale, d^{abc} are the structure constants of SU(3) and S_8 and $G_{\mu\nu}$ are the color-octet scalar field and the gluon field tensor respectively.

The width of the color octet scalar resonance is given by:

$$\Gamma = \frac{5}{6} \alpha_s k_s^2 \frac{M^3}{\Lambda^2} \tag{1.17}$$

The anomalous coupling of the color-octet scalar model used in this thesis is $k_s^2 = 1/2$.

Z' and W'

New gauge bosons arise from extensions of the electroweak symmetry of the Standard Model in which there is an extra $SU(2)$ gauge group. $SU(2) \times SU(2)$ is spontaneously broken to the diagonal subgroup $SU(2)_L$, which corresponds to the SM group. In these extensions new gauge coupling constants are of the same order as in the $SU(2)_L$ and new gauge bosons are produced. W' and Z' bosons [14], named in analogy with the Standard Model W and Z bosons, refer to these hypothetical gauge bosons.

Assuming that these new gauge bosons couple to the SM particles, their cross sections are calculated by scaling the corresponding cross sections of their SM counterparts. In particular the Fermi constant G_F becomes:

$$G'_F = G_F \left[\frac{M}{M'}\right]^2 \tag{1.18}$$

where M and M' are the masses of SM bosons and new gauge bosons respectively.

Furthermore W' and Z' have electroweak couplings and require antiquarks for their production ($q_1\bar{q}_2 \rightarrow W', q\bar{q} \rightarrow Z'$), giving small cross sections. As in the Standard Model, the Z' production rate is expected to be smaller than the one for W'.

Dark Matter Mediator

The dark matter particles (DM) can avoid observation in direct detection experiments for example at low DM masses ($\mathcal{O}$(GeV)) where the sensitivity is reduced mainly due to the low-energy threshold of the detector (the elastic scattering of DM with mass of 10 GeV would produce nuclear recoils of 1 keV [15]). Furthermore at high DM masses ($\mathcal{O}(10^2$ GeV)) the direct searches lose sensitivity due to the fact that the local dark-matter density is constant (0.3 GeV/cm^3 [16]) and therefore the heavier the individual particles, the less particles are available for the scattering. On the other hand the dark matter search can be strongly constrained, under specific assumptions, by LHC measurements. In this way, an important complementarity is established between searches for DM particles production and searches for the mediator itself, as in the case of the dijet resonance search.

Models in which the dark matter particle χ is a Dirac fermion and the particle mediating the interaction (the "mediator") is exchanged in the s-channel are considered. All the models are characterized by four parameters: the DM mass m_{DM}, the mediator mass M_{med}, the universal mediator coupling to quarks g_q and the mediator coupling to DM g_{DM}.

A simplified model [17, 18] is considered, in which the dark matter mediator is assumed to be a spin-1 particle decaying only to quark-antiquark pairs or DM particles pairs, with unknown mass m_{DM} and with an universal coupling to quarks $g_q = 0.25$ and a DM coupling $g_{DM} = 1.0$.

The two models with a spin-1 mediator Z', vector (V) and axial-vector (AV), have the following interaction Lagrangians:

$$L_V = -g_{DM} Z'_\mu \bar{\chi}\gamma^\mu \chi - g_q \sum_{q=u,d,s,c,b,t} Z'_\mu \bar{q}\gamma^\mu q \tag{1.19}$$

and

$$L_{AV} = -g_{DM} Z'_\mu \bar{\chi}\gamma^\mu\gamma_5\chi - g_q \sum_{q=u,d,s,c,b,t} Z'_\mu \bar{q}\gamma^\mu\gamma_5 q \tag{1.20}$$

The minimal decay width of the mediator is given by the sum of the partial widths for all the decays into DM and quarks that are kinematically accessible. For vector mediator, the partial widths are given by:

$$\Gamma_V^{\chi\bar{\chi}} = \frac{g_{DM}^2 M_{med}}{12\pi}(1-4z_{DM})^{1/2}(1+2z_{DM}) \tag{1.21}$$

$$\Gamma_V^{q\bar{q}} = \frac{g_q^2 M_{med}}{4\pi}(1-4z_q)^{1/2}(1+2z_q) \tag{1.22}$$

where $z_{DM,q} = m_{DM,q}^2/M_{med}^2$ and the two different types of contribution to the width vanish for $M_{med} < 2m_{DM,q}$.

The corresponding expressions for the axial-vector mediator are:

$$\Gamma_{AV}^{\chi\bar{\chi}} = \frac{g_{DM}^2 M_{med}}{12\pi}(1-4z_{DM})^{3/2} \tag{1.23}$$

$$\Gamma_{AV}^{q\bar{q}} = \frac{g_q^2 M_{med}}{4\pi}(1-4z_q)^{3/2} \tag{1.24}$$

For completeness, an alternative method to the simplified model has to be quoted and this is the effective field theory (EFT) approach. It does not include a mediator reducing the interactions between DM and the SM fields down to contact interactions. The advantage of the simplified model with respect to the EFT is that it does not depend on an energy scale, but it is more model-dependent [19].

Randall-Sundrum gravitons

Models with warped extra-dimensions can explain the difference between the Planck scale and the electroweak scale, introducing a non-trivial geometry in the extra-dimensions. Randall and Sundrum (RS) proposed a warped geometry with 5 dimensions with an energy scale of the order of the TeV [20]. In this case, with only one extra-dimension, the metric is given by:

$$ds^2 = e^{-2kr_c\phi}\eta_{\mu\nu}dx^\mu dx^\nu + r_c^2 d\phi^2 \tag{1.25}$$

where k is a scale of the order of the Planck scale, $0 \leq \phi \leq \pi$ is the coordinate for the extra dimension, which is a finite interval whose size is set by the compactification radius r_c, $\eta_{\mu\nu} = diag(-1,1,1,1)$ is the 4D Minkowski metric and x^μ are the usual space-time dimensions. In this model, spin-2 gravitons (denoted by G) are expected, being the Kaluza-Klein (KK) excitations of the gravitational field $h^{\mu\nu}$. The graviton coupling to the Standard-Model fields is given by the interaction Lagrangian:

$$L = -\frac{1}{\Lambda}h^{\mu\nu}T_{\mu\nu} \tag{1.26}$$

where Λ is the energy scale and $T^{\mu\nu}$ is the energy-momentum tensor of the matter fields.

If $\Lambda \sim TeV$, RS gravitons can be produced in hadron collisions and they can decay to gluons or quark-antiquark pairs, leading to the dijet signature. The partial widths are given by:

$$\Gamma(G \to gg) = \frac{x_1^2}{10\pi} \left[\frac{k}{M_{Pl}}\right]^2 m_1 \tag{1.27}$$

and

$$\Gamma(G \to q\bar{q}) = \frac{3x_1^2}{160\pi} \left[\frac{k}{M_{Pl}}\right]^2 m_1 \tag{1.28}$$

where x_1 is the 1st root of the the Bessel function of order 1, m_1 is the mass of the KK excitation and M_{Pl} is the reduced effective 4-D Planck scale.

In this thesis the model with $k/M_{Pl} = 0.1$ is considered.

Model Name	Mediator	Decay Channel
Excited Quark	q*	qg
E_6 Diquark	D	qq
Axigluon	A	$q\bar{q}$
Coloron	C	$q\bar{q}$
Color-Octet Scalar	s_8	gg
Heavy W	W'	$q\bar{q}$
Heavy Z	Z'	$q\bar{q}$
Dark Matter Mediator		$q\bar{q}$
RS Graviton	G	$q\bar{q}, gg$

Table 1.1. Summary of the considered theoretical models predicting a resonance decaying in a pair of quarks or gluons.

Chapter 2

The Large Hadron Collider and the Compact Muon Solenoid

The analysis presented in this thesis was performed with the data collected in 2016 by the Compact Muon Solenoid (CMS) experiment, one of the four major experiments installed at the Large Hadron Collider (LHC). In this chapter, after a short description of the LHC, all the subdetectors of CMS are illustrated.

2.1 The Large Hadron Collider

The Large Hadron Collider at CERN is the largest and most powerful particle accelerator ever built. It is part of a complex system of accelerators, acting as injectors, aiming to accelerate particle beams up to high energy. At LHC heavy ions or protons are accelerated, but for this thesis only the last ones are considered. The protons are extracted from a hydrogen source and injected in the accelerator system that is schematically shown in Figure 2.1 and it is composed by:

- A linear accelerator (called *LINAC 2*) accelerating protons up to an energy of 50 MeV;
- The *Proton Synchrotron Booster* (*PSB*), which increases the proton energy up to 1.5 GeV;
- The *Proton Synchrotron* (*PS*), where the protons reach the energy of 28 GeV;
- The *Super Proton Synchrotron* (*SPS*), a circular accelerator with a radius of 1 km that leads protons to have an energy of 450 GeV;
- The *Large Hadron Collider* (*LHC*), which is the last accelerator of the chain. It is a 27 km long circular accelerator in which the protons reach the energy of 6.5 TeV and they collide with an energy in the center of mass of 13 TeV.

In proton-proton collisions there may be interactions at small distance between the partons, with high transferred transverse momentum. These events, which lead to the creation of new massive particles, are the ones sought but are rare. The majority of the events consists of minimum bias events, which means collisions at

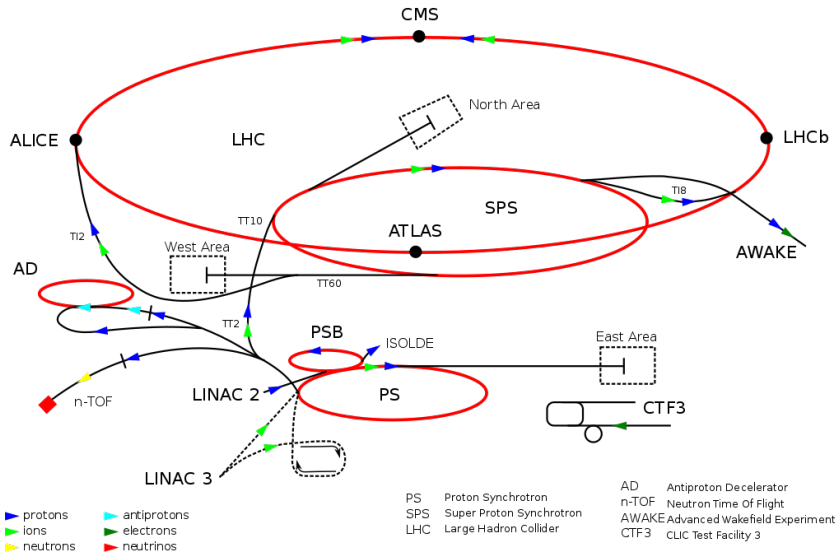

Figure 2.1. Schematic view of the CERN accelerator system.

large distance between the partons and low transferred momentum. For this reason, in addition to a high energy in the center of mass, an important design feature of the LHC is the high *luminosity*, that quantifies the ability of a particle accelerator to produce the required number of interactions. The number of events, N, is in fact the product between the cross section of interest, σ, and the integral over the time of the instantaneous luminosity, L:

$$N = \int L(t)dt \cdot \sigma \tag{2.1}$$

It is easy to see that the luminosity is related to the rate of events for a process i with cross section σ_i by the relation:

$$R_i \equiv \frac{dN_i}{dt} = L \cdot \sigma_i \tag{2.2}$$

The luminosity depends only on the machine parameters and in a circular collider is defined as:

$$L = N_1 N_2 f N_b \int\int\int\int_{-\infty}^{+\infty} \rho_1(x, y, s - s_0)\rho_2(x, y, s + s_0)dxdydsds_0 \tag{2.3}$$

where N_1 and N_2 are the number of protons in a bunch for the two beams, N_b is the number of colliding bunches, f is the revolution frequency and ρ_1 and ρ_2 are the density distribution functions of the two beams. Referring to Figure 2.2, in which two colliding bunches are illustrated, $s_0 = c \cdot t$ is the distance from the central collision point.

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

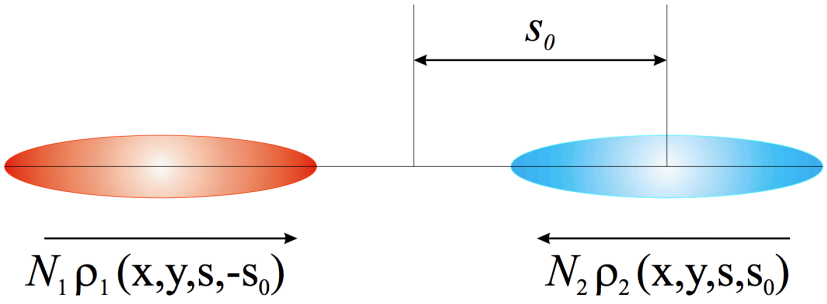

Figure 2.2. Schematic view of two colliding bunches. N_i is the number of particles per bunch, ρ_i is the density function and s_0 is the distance from the central collision point.

$$L = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y} \tag{2.4}$$

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

The design characteristics of the LHC are listed in Table 2.1.

Circumference	27 km
Maximal number of bunches	2808
Number of protons per bunch	$1.05\ 10^{11}$
Bunch spacing	7.48 m = 25 ns
Maximal center of mass energy $\sqrt{s}$	14 TeV
Bunch transverse size $\sigma_x = \sigma_y$	15 µm
Bunch length	7.5 cm
Luminosity	10^{34} cm^{-2}s^{-1}

Table 2.1. Summary of the main LHC technical parameters.

For the first years of data-taking, the so-called "Run1", the center-of-mass energy in the proton-proton collision was 7 and then 8 TeV, while for the current data-taking, the "Run2", the record energy of 13 TeV is reached.

At LHC the beams collide in four interaction points. Two of them are the most general purpose experiments ATLAS (*A Toroidal LHC ApparatuS*) and CMS (*Compact Muon Solenoid*), while the other two host the ALICE (*A Large Ion Collider Experiment*) and LHCb (*Large Hadron Collider beauty*) experiments, respectively focused on heavy ion physics and CP violation measurements. The other smaller experiments at LHC are TOTEM [21] and LHCf [22], which focus on "forward particles", and MoEDAL [23] searching for a hypothetical particle called "magnetic monopole".

2.2 The Compact Muon Solenoid

CMS is a general purpose experiment designed to search for the Higgs boson and for physics beyond the Standard Model in many possible final states and to study

with high precision already known particles and phenomena. The detector has a cylindrical form with radius 7.5 m, length of 22 m and a weight of about 12500 tons. The detector is composed by different subdetectors, each one acting to detect different kinds of particles. A schematic view of the CMS apparatus is shown in Figure 2.3.

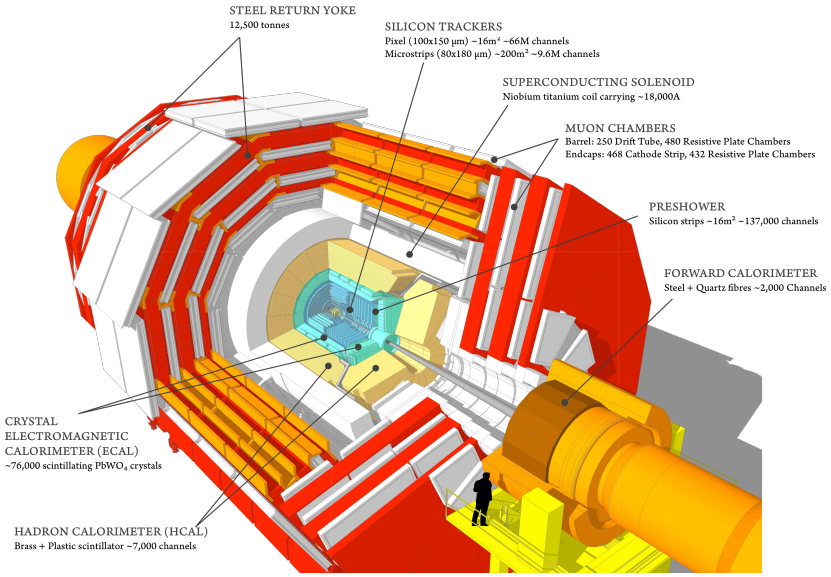

Figure 2.3. Schematic view of the CMS apparatus.

From the interaction point, the particles produced in the collisions go through the following subdetectors:

- the tracking system (*Tracker*) in which charged particles release a track;
- the electromagnetic calorimeter (*ECAL*) where electrons and photons deposit their energy;
- the hadrons continue, almost undisturbed, until the hadronic calorimeter (*HCAL*) where they are absorbed;
- Muons are detected in the chambers (*Muon chambers*) located in the outer part of the detector.

The Tracker and the two calorimeters are located inside a superconducting solenoid providing a uniform magnetic field of 3.8 T at its center.

In Figure 2.4 a transverse slice of the CMS apparatus is shown, as well as the typical subdetector responses to the different particles.

The CMS detector is divided in a central region, coaxial to the beam axis, called *barrel*, and two *endcaps*, which are orthogonal to the beam and close hermetically the cylinder.

The coordinate system adopted by CMS, shown in Figure 2.5, has the origin in the interaction point, the x-axis points to the center of the LHC ring, the y-axis

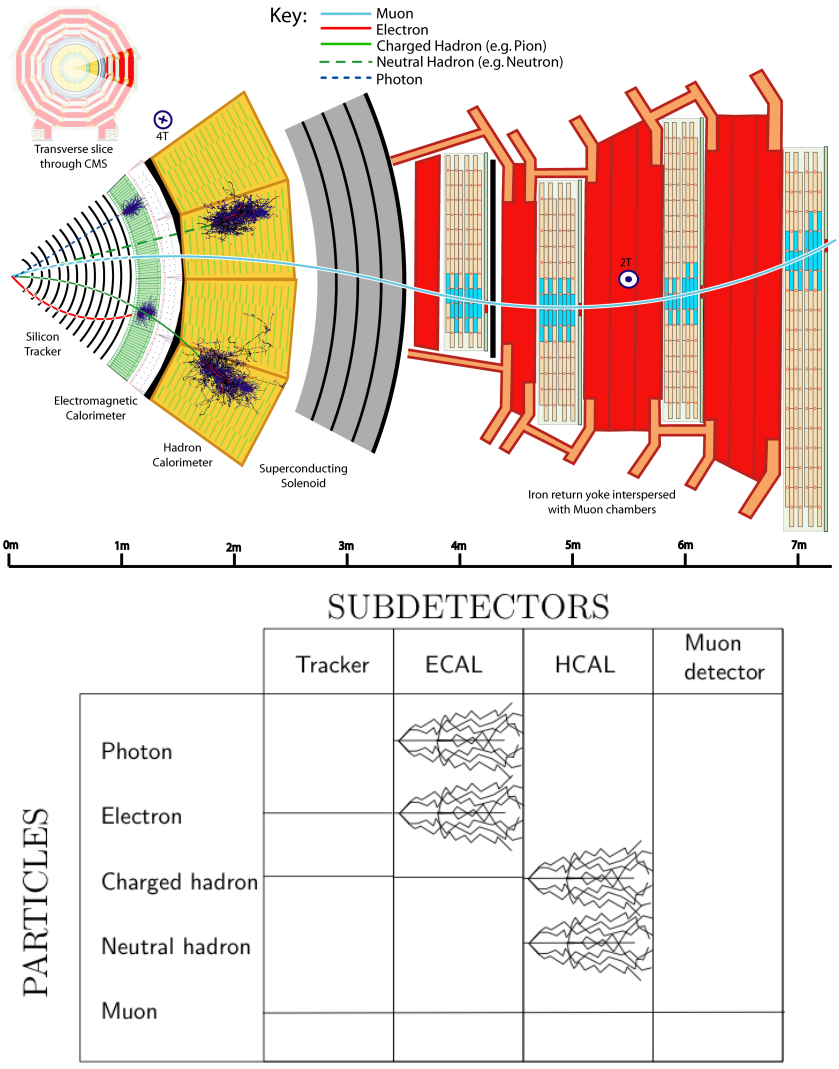

Figure 2.4. (Top) Schematic view of a transverse slice of CMS detector. The sketch reports the signals left in subdetectors by different types of particles. (Bottom) The typical subdetector response to different particles.

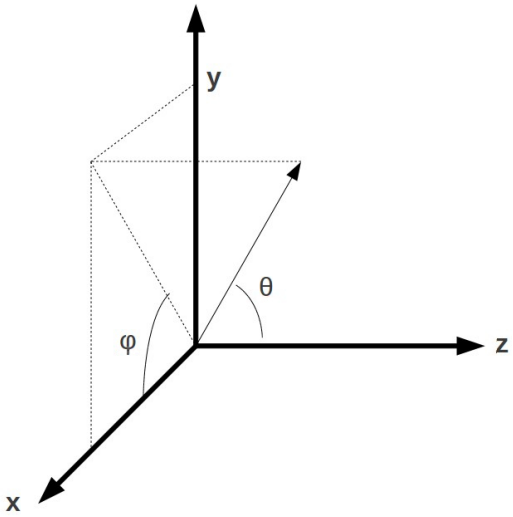

Figure 2.5. Coordinate system in CMS.

points vertically upward and z-axis is in the beam direction. The azimuthal angle ϕ is measured from the x-axis in the x-y plane, while the polar angle θ is measured from the z-axis.

When describing a process the *pseudorapidity*, η, is often used, which is defined as:

$$\eta = -\ln \tan \frac{\theta}{2} \tag{2.5}$$

This variable at high energy tends to the *rapidity* given by:

$$y = \frac{1}{2} \ln \frac{E + p_L}{E - p_L} \tag{2.6}$$

where p_L is the component of the momentum parallel to the beam. The advantage to use the pseudorapidity is that it is a function of the polar angle only, and does not depend from p_L, that for the colliding partons is generally unknown.

2.2.1 Magnet

A magnetic field is needed to bend the path of the charged particles emerging from high-energy collisions in the LHC, allowing the momentum measurement of the tracker thanks to the use of the curvature radius. To achieve the best possible resolution in the muon detection system, CMS chose to have a relatively small magnet producing an intense field. The magnet installed in CMS is a solenoid and it is superconducting allowing electricity to flow almost without resistance. It is 13 m long cylindrical magnet [24] made of Niobium-Titanium, with a diameter of 5.9 m. It provides a uniform magnetic field of 3.8 T at its center, carrying a current of 18 kA. The magnet flux is returned via a 1.5 m thick saturated iron yoke which also acts as a filter, through which only muons and weakly interacting particles such as neutrinos can pass.

2.2.2 Tracker

The momentum of the particles is crucial to build up a picture of events at the heart of the collision. One method to calculate the momentum of a charged particle is to track its path through a magnetic field; the more curved the path, the less momentum the particle had. This is the aim of the CMS tracker system [25]. Its design goal is to reconstruct tracks of high-p_T electrons and muons with an efficiency greater than 95% and greater than 90% for high-p_T charged hadrons.

Furthermore the inner part is dedicated to the reconstruction of the primary vertex and potential secondary vertices.

The CMS tracker system is composed by different layers covering the pseudorapidity region up to $|\eta| < 2.5$. Exploiting the same technologies based on silicon, the barrel and the two endcaps are composed by:

- *Pixel Barrel/Endcap (TPB/TPE)*: Arranged in a 3-layers barrel and a 2-layers endcap on each side, the pixels are used where the particle flow is very large (10^7-10^8 cm^{-2}s^{-1}). An occupancy of about 10^{-4} per pixel per bunch crossing is obtained with pixel dimensions of 100 µm × 150 µm;

- *Inner Barrel/Disks (TIB/TID)*: the particle flow is low enough (10^6 cm^{-2}s^{-1}) to allow the adoption of microstrips with cells of 10 cm × 80 µm. The occupancy in this region is about 10^{-1} per pixel per bunch crossing;

- *Outer Barrel/Endcap (TOB/TEC)* ($r > 55$ cm): the particle flow is smaller (10^5 cm^{-2}s^{-1}) and therefore larger strips can be used (25 cm × 180 µm), keeping the occupancy around percent.

In Figure 2.6 a longitudinal section of the tracking system is shown, highlighting the different layers.

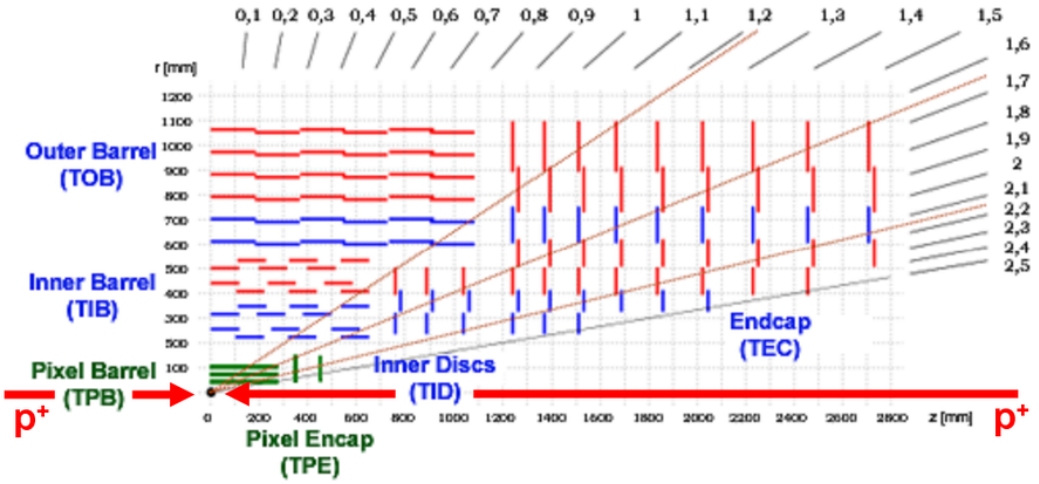

Figure 2.6. Schematic longitudinal section of one quarter of the CMS silicon tracker detector. The distances from the interaction point are reported in millimeters on the left and bottom axes and the pseudorapidity values are reported on the top.

The silicon pixel detector provides a resolution in position of about 10 µm in the transverse plane xy and a precision of about 20 µm for the z-coordinate. The microstrips instead have a resolution that depends on the cell thickness, but it is always better than 55 µm in the transverse plane. The nominal momentum resolution is typically 0.7 (5.0)% at 1 (1000) GeV in the central region.

2.2.3 Electromagnetic calorimeter

The electromagnetic calorimeter (*ECAL*) [26] is an homogeneous and hermetic calorimeter composed by 61200 scintillating crystals in the barrel and 7324 crystals in each endcap. The aim of ECAL is to measure with high precision the energy of photons, electrons and positrons interacting with the crystals. The behavior of the scintillating crystals is determined by the following variables:

- Radiation length, X_0: the mean longitudinal distance over which the energy of an electron is reduced by the factor 1/e. For the typical energies, a calorimeter with a thickness of 25 X_0, such as ECAL, allows a containment of 98% with respect to the total energy of the electromagnetic shower;
- Moliere radius, R_M: it is useful to describe the transverse dimensions of the shower. Defined by:

$$R_M = \frac{X_0 \cdot 21.1[MeV]}{E_C[MeV]} \tag{2.7}$$

  it is the radius of a cylinder containing on average 90% of the shower's energy deposition. In the formula, E_C is the critical energy for which the mean energy loss due to ionization is equal to the loss via Bremsstrahlung;
- Light yield, LY: it is defined as the number of emitted photons per MeV deposited in a crystal;
- τ: it is the scintillation time of a crystal.

Considering these variables, the characteristics of the Lead Tungstate ($PbWO_4$) make it an appropriate choice for operation at CMS. In fact the high density (8.3 g/cm^3) and short radiation length (0.89 cm) of $PbWO_4$ allow the construction of a compact calorimeter inside the magnet. Furthermore the small Moliere radius (2.2 cm) ensures high lateral shower containment and high granularity (needed for π^0-γ separation and angular resolution). The last important characteristic of $PbWO_4$ is the fast light emission, with the 80% of light emitted in 25 ns (same magnitude as the LHC bunch crossing time). A drawback of the lead tungstate is the low LY (100 γ/MeV) compared to other scintillating materials, which forced to use photodetectors with an intrinsic gain. The chosen photodetectors are Avalanche PhotoDiode (APD) in the barrel and Vacuum PhotoTriode (VPT) in the endcaps. The APDs are semiconducting silicon detectors with a strong electric field. When an electron is knocked out from an atom by scintillation light, it is accelerated in the electric field, knocking out electrons from the other atoms. Therefore, APDs are able to produce a very high signal in a short time. The VPTs are used in

the endcaps, where the radiation is too high for silicon photodiodes. The VPTs consist of three electrodes within a vacuum tube: one of them releases an electron when the light arrives, one works as anode producing several electrons that are then accelerated to the third electrode (the dynode), releasing a second batch of electrons.

The barrel part of ECAL (EB) covers the pseudorapidity range up to $|\eta| < 1.479$ and it is made of 36 identical *supermodules*, each one composed by 1700 crystals. The crystals have a trapezoidal shape and a longitudinal dimension of 25.8 X_0. They are organized in a semi-projective geometry, forming a 3^o angle with respect to the line that connects them to the nominal interaction point, to avoid that photons fall in the separation zone between two crystals. The endcaps (EE) cover the pseudorapidity region $1.653 < |\eta| < 3$ and their crystals have a length of 24.7 X_0. Furthermore in front of the endcaps two preshower detectors (ES) are installed. They are additional calorimeters, covering the region $1.7 < |\eta| < 2.6$, made of lead and silicon strips and used to improve the separation between π_0 from γ. In Figure 2.7 the ECAL geometry is schematically shown.

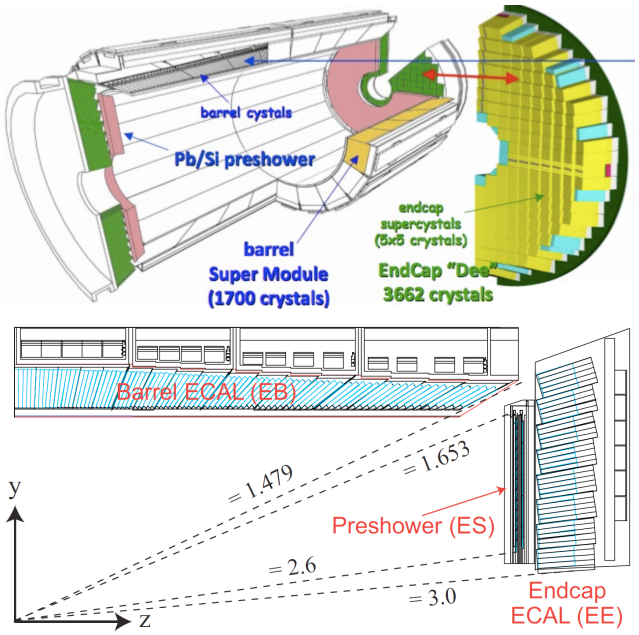

Figure 2.7. Schematic view of electromagnetic calorimeter of CMS. On the top a 3D view is shown, underlying the different part of the detector: the barrel in orange, the endcaps in green and the preshower in pink. On the bottom a section of a quarter of ECAL is reported, with the pseudorapidity values covered by each part.

The electromagnetic calorimeter resolution can be parameterized with the following expression:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{2.8}$$

where E is the energy expressed in GeV. The stochastic term, S, depends on the fluctuations in the number of detected photons, the noise term (N) is related to

the electronics noise and the constant term (C) includes several contributions, such as the lateral containment, the non uniformity of response and the precision of the intercalibration. This constant term is particularly important because it dominates at high energy and for ECAL is around 0.3%.

2.2.4 Hadron calorimeter

The aim of the hadron calorimeter (*HCAL*) [27] is to measure the energy and the direction of the hadrons produced in the collision. Furthermore the combination of information from HCAL and other subdetectors allows to reconstruct the energy of the undetected particles, such as neutrinos, exploiting the energy balancing in the transverse plane. For these reasons the fundamental requirements for HCAL are a good hermerticity, a high transverse granularity and a good containment of the hadron showers.

HCAL is a sampling calorimeter, which means it alternates layers of absorber made of brass and plastic scintillators. It is organized in sections as follows:

- Barrel/Endcap Hadronic Calorimeter (HB/HE): Located inside the magnetic field, the barrel part covers the pseudorapidity region up to $|\eta| < 1.3$, it has a total thickness of 7 interaction lengths (λ_0) and its transverse granularity is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The endcap parts extend the pseudorapidity up to $|\eta| < 3$. The granularity is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ up to $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ in the range $1.6 < |\eta| < 3$;

- Outer Hadronic Calorimeter (HO): Since the depth of the barrel is not enough to ensure a complete containment of the hadronic showers originated from particles with transverse energies above 500 GeV, an additional layer of one λ_0 is present outside the magnet. The granularity and the η covering is the same as HB;

- Forward Hadronic Calorimeter (HF): Covering the very high pseudorapidity region ($3 < |\eta| < 5$), the HF subdetector is a Cerenkov light detector made of quartz fibers embedded within a 165 cm long steel absorber. It measures the energy of particles produced at small angles with respect to the beam line.

A longitudinal view of a quadrant of HCAL is shown in Figure 2.8.
The energy resolution of HCAL can be parameterized as:

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E[GeV]}} \oplus 8\% \tag{2.9}$$

It is larger than the ECAL energy resolution, therefore the final resolution will be dominated by HCAL.

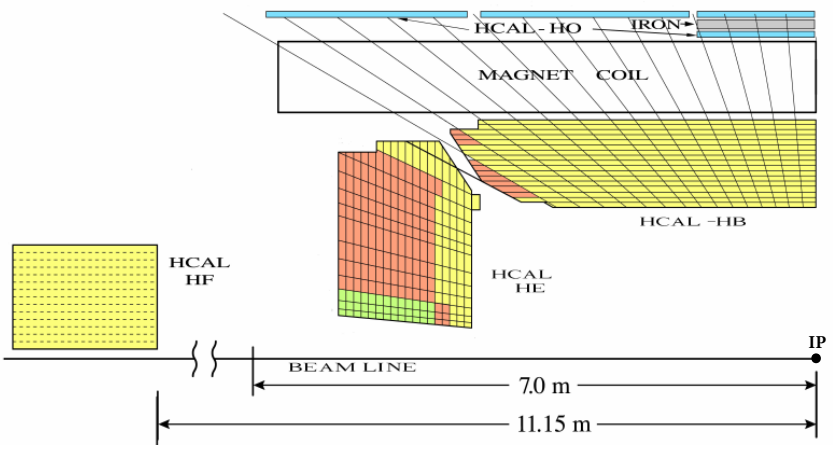

Figure 2.8. View of a section of a quarter of CMS hadron calorimeter. The distances are reported from the interaction point (IP) located to the right end of the beam line.

2.2.5 Muon chambers

The muon detector [28] is the outermost of the CMS subdetectors. It has the role to identify muons, the only charged particles which pass trough the calorimeters without being absorbed and, using also the information from the tracker system, to measure their momentum. It is located outside the magnet and constitutes the return yoke covering the pseudorapidity region up to $|\eta| < 1.4$ with the barrel and up to $|\eta| < 2.4$ with the endcaps. The system relies on three different technologies:

- *Drift Tube* (DT): The DTs are chambers built with 12 layers of drift tubes. Each chamber is composed by parallel plates made of aluminum, filled with a mix of Argon (85%) and CO_2 (15%). They are located in the barrel region, where the relative occupancy is less than 10 $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ and the magnetic field is uniform;

- *Cathode Strip Chambers* (CSC): The CSCs are multi-wire proportional chambers filled with a mixture of CO_2 (50%), Argon (40%) and CF_4 (10%). They are installed in the endcaps where the occupancy is larger (100 $\mathrm{cm}^{-2}\mathrm{s}^{-1}$) and the magnetic field is not uniform;

- *Resistive Plate Chambers* (RPC): The RPCs are chambers made of parallel plates of Bakelite. They are employed in the barrel as well as in the endcaps.

A section of the muon detector is shown in Figure 2.9.

The DT and the CSC provide an excellent spatial resolution (≈100µm) for the measurement of the momentum, while the RPC are mainly used for the bunch crossing identification and for trigger purposes, thanks to their excellent time resolution (3 ns). In Figure 2.10 the resolution in momentum is reported for the muon system only and for the combined information from this detector and the tracker. The combination significantly improves the precision of the measure.

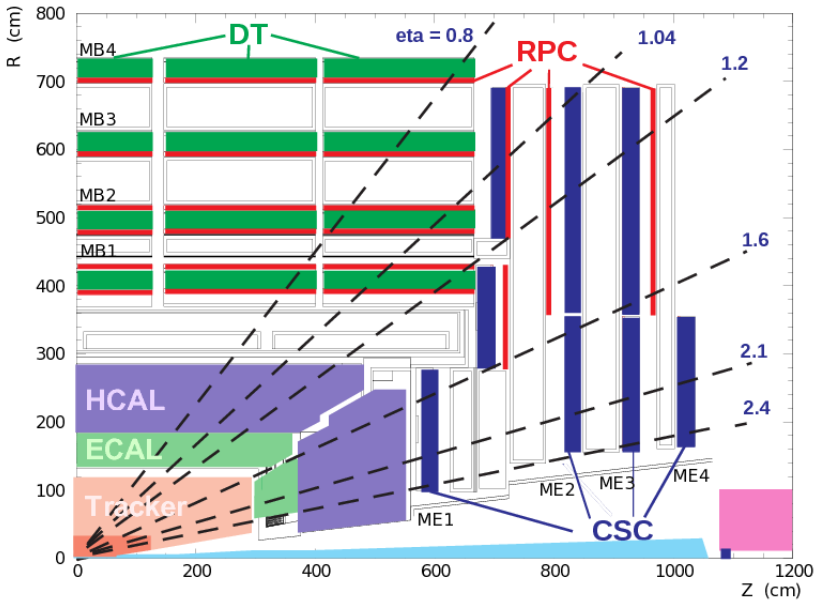

Figure 2.9. View of one quarter of the CMS muon system.

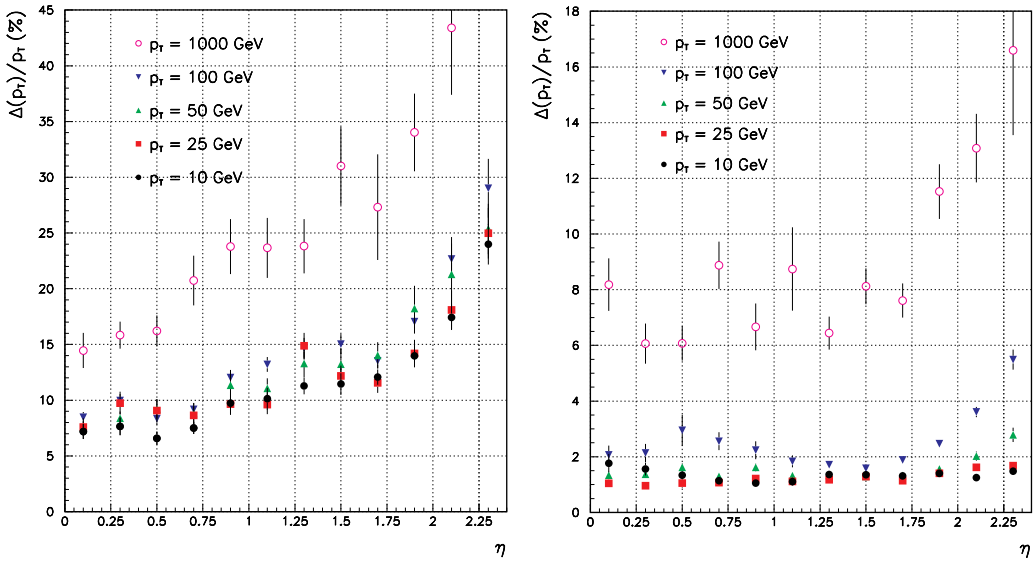

Figure 2.10. The momentum resolution as a function of η for different muon p_T. (Left) The resolution obtained from the muon detector only. (Right) The results from the combination of information from the muon detector and the tracker.

2.2.6 Trigger

The LHC collides proton bunches each 25 ns, that means with a frequency of 40 MHz. Such high frequency, together with a typical event size of about 1 MB, makes it impossible to save the information for all the events. Moreover, saving all the events would not be useful since most of them are soft collisions and therefore not interesting for the CMS physics program. Since the maximum event rate sustainable from CMS is around 1 kHz, a decision-making system (*trigger system*) is in use to decide quickly if an event is interesting and must be saved or if can be discarded. The trigger system in CMS is composed by two levels:

- the first level of the trigger is completely hardware and it is called *Level 1 trigger* or *L1*. Each decision is made in about 3.8 µs and it is based only on the information from the calorimeters and the muon chambers. At this step the frequency of the events is reduced down to about 100 kHz;
- the second level is based on software and it is the *High Level Trigger* or *HLT*. It receives the event from the L1 and the decision is based on various strategies but the guiding principles are a regional reconstruction and a fast event veto. The regional reconstruction tries to avoid the complete event reconstruction, that is too slow for triggering purposes, and focuses on the detector regions where the L1 trigger has found an interesting activity. Non-interesting events are discarded as soon as possible, selecting in the end about 1 kHz of data, that is an acceptable event rate that can be saved on disk.

Chapter 3

Jet reconstruction and calibration

3.1 Collider physics

3.1.1 Parton-Parton collisions

The two peculiar characteristics of the QCD are the quarks confinement and the asymptotic freedom [29]. One of the consequences of the asymptotic freedom is that in proton-proton collisions at high energy, as the ones produced at the LHC, the hard scattering process takes place between the proton constituents, the partons. This also entails that the center-of-mass energy of the collision is a fraction of the center-of-mass energy of the proton-proton system. The fraction of the hadron energy carried by each parton is described by a quantity called Bjorken scaling variable x. Summed over all the partons i inside a hadron it gives:

$$\sum_i x_i = 1 \tag{3.1}$$

A sketch illustrating a typical proton-proton collision is shown in Figure 3.1. The blue arrows are the incoming protons, while the red arrows are the colliding partons (which have the same direction of the protons) and the outgoing partons from the hard interaction. The green arrows represent the initial and final state radiations and the black ones are the other initial partons interacting at low energy producing the so-called "underlying event".

The dynamics of the hard scattering of the colliding hadrons can therefore be approximately described as a two-to-two process between massless partons. Referring to Figure 3.2, the Mandelstam variables $\hat{s}$, $\hat{t}$ and $\hat{u}$ are defined as:

$$\hat{s} = (p_A + p_B)^2 = (p_C + p_D)^2 \tag{3.2a}$$
$$\hat{t} = (p_A - p_C)^2 = (p_B - p_D)^2 \tag{3.2b}$$
$$\hat{u} = (p_A - p_D)^2 = (p_B - p_C)^2 \tag{3.2c}$$

where p_i refers to the four-momentum of each parton.

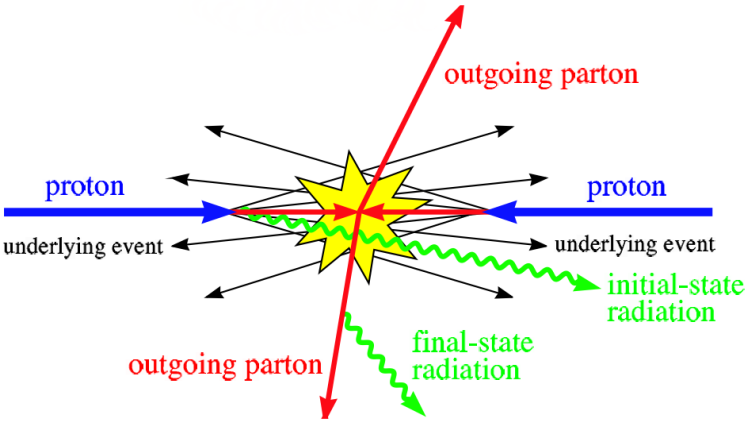

Figure 3.1. Sketch of a proton-proton collision.

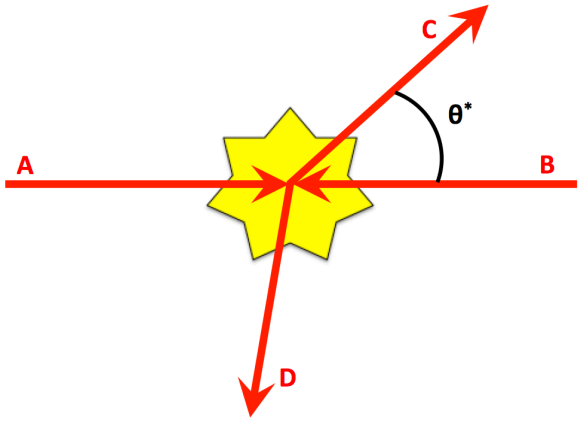

Figure 3.2. Sketch of a two-to-two partons scattering.

The Mandelstam variables are related to the hadron energy and to the scattering angle, θ^*, by the following equations, which are valid for massless partons:

$$\hat{s} = 4x_A x_B E^2 \tag{3.3a}$$

$$\hat{t} = -\frac{\hat{s}}{2}(1 - \cos\theta^*) \tag{3.3b}$$

$$\hat{u} = -\frac{\hat{s}}{2}(1 + \cos\theta^*) \tag{3.3c}$$

where we considered for simplicity symmetric collisions, in which the two incoming beams have the same energy E, as for the LHC.

The transverse momenta of the outgoing partons, p_T, and their rapidities in the laboratory system, y_C and y_D, are connected to the Bjorken-x through:

$$x_A = \frac{p_T}{2E}\left(e^{y_C} + e^{y_D}\right) \text{ and } x_B = \frac{p_T}{2E}\left(e^{-y_C} + e^{-y_D}\right) \tag{3.4}$$

Introducing the variables:

$$\bar{y} = \frac{y_C + y_D}{2} \text{ and } y^* = \frac{y_C - y_D}{2} \tag{3.5}$$

Eq. 3.4 can be expressed as:

$$x_A = \frac{p_T}{E} e^{\bar{y}} \cosh y^* \text{ and } x_B = \frac{p_T}{E} e^{-\bar{y}} \cosh y^* \tag{3.6}$$

In these coordinates:

$$\bar{y} = \frac{1}{2}\log\frac{x_A}{x_B} \tag{3.7}$$

and the invariant mass of the produced system is

$$M = \hat{s} = 4p_T^2 \cosh^2 y^* \tag{3.8}$$

The rapidity of the outgoing partons is therefore related to the scattering angle and it can be written as:

$$\cos\theta^* = \tanh\left(\frac{y_C - y_D}{2}\right) = \tanh y^* \tag{3.9}$$

3.1.2 Hadronic cross section

To derive the cross section of a hadronic scattering process it is useful to introduce the *parton distribution functions* (PDFs), indicated with $f_i(x, \mu^2)$. They allow the calculation of the probability of finding inside a hadron a certain kind of parton i carrying a momentum fraction x at a squared energy scale μ^2. The total cross section can be factorized into a "hard part" expressed by the parton-parton scattering cross section, $\hat{\sigma}(x_A p_A, x_B p_B, \mu^2)$, and a normalization part based on the PDFs. It can be written then as:

$$\sigma = \sum_{ij} \int dx_A dx_B f_i(x_A, \mu^2) f_j(x_B, \mu^2) \hat{\sigma}(x_A p_A, x_B p_B, \mu^2) \tag{3.10}$$

The (arbitrary) factorization scale μ can be thought as the scale which separates the long and short-distance physics. Partons with transverse momentum larger than μ participate in the hard scattering process with a short-distance partonic cross-section $\hat{\sigma}$.

The cross section of the process is, of course, independent of the machine parameters. Nevertheless , for hard scattering processes at a hadron collider, it is helpful to consider the product of a *parton luminosity factor* and the parton-parton cross section $\hat{\sigma}$. Calling $x_A x_B = \tau$, we can define the differential parton luminosity as:

$$\frac{dL_{ij}}{d\tau} = \int_0^1 dx_A dx_B f_i(x_A, \mu^2) f_j(x_B, \mu^2) \delta(x_A x_B - \tau) \tag{3.11}$$

and therefore the hadronic cross section can be written as:

$$\sigma(s) = \sum_{ij} \int \frac{d\tau}{\tau} \left[\frac{1}{s} \frac{dL_{ij}}{d\tau} \right] [\hat{s}\hat{\sigma}_{ij}] \tag{3.12}$$

3.1.3 Hadronization and jets

Particles carrying a color charge, such as quarks or gluons, cannot exist in free form because of QCD confinement which only allows colorless states. For an object containing color charge is therefore energetically convenient to lose its color charge by multiple radiations of gluons which excite the vacuum producing quark-antiquark pairs. They then combine themselves, through a process known as *hadronization*, forming colorless particles. These particles are produced in a collinear direction with respect to the initial partons, due to the quadrimomentum conservation and therefore the ensemble of these particle appears in the detector as a spray of highly collimated particles, called *jet*. Jets are composed by different particles, but on average about 65% of the jet energy is carried by charged particles, about 20% is composed by high-energy photons, while the remaining part is made of neutral hadrons. The photons present in the jet are well measured in ECAL, while the hadrons produce hadronic showers that are absorbed in the calorimeters. A high-energy shower is a very complex process that consists in a cascade of secondary particles produced as the result of a high-energy particle interacting with dense matter. Mainly charged and neutral pions are produced, but, with lower multiplicity, also kaons, nucleons and other hadrons. A large component of the secondary particles in hadron cascades consists of neutral pions, which represent approximately one third of the pions produced in each inelastic collision. Neutral pions decay quickly into two photons, thereby initiating an electromagnetic subshower in the hadron shower. Therefore after the first collision about 1/3 of the energy is deposited in electromagnetic form (f_{em}). Charged pions will produce other three pions if their energy is larger than the energy required for the production. The process then iterates, until all the remaining charged pions have not sufficient energy to produce new particles. The total electromagnetic energy fraction produced in the shower is related to the initial energy E by the power law:

$$f_{em} = 1 - \left(\frac{E}{E_0} \right)^k \tag{3.13}$$

where E_0 is a scale factor and k is related to the average multiplicity and type of particle produced at each step [30]. As we can see from this relation, the higher the energy of the original particle, the higher the electromagnetic energy fraction of shower it originates.

3.2 Jet reconstruction

As already mentioned in the previous paragraph, jets are the experimental signature of quarks and gluons. It is therefore necessary to reconstruct these jets in order to have information on the partons produced in the collision.

For the purpose of this thesis jets reconstructed from different constituents are considered. Also, both jets fully reconstructed offline or jets reconstructed only at trigger level are used. The motivation for this choice lies in the need to probe the lowest possible dijet masses and this can be done with a particular technique, the *data scouting*. A crucial point of this method is that only jets are saved on tape while other information are discarded, which means that the offline reconstruction for this events cannot be performed. This concept and the reasons behind it will be explained in detail in Chapter 4, while the next paragraphs focus on the jet reconstruction.

3.2.1 Clustering algorithms

Since jets are composite objects, a clustering algorithm is needed in order to define them univocally. A good jet algorithm should be collinear and infrared safe, which means that it has to be insensitive to the emission of soft or collinear gluons. If $\vec{p_i}$ is the momentum carried by a jet, the result of the clustering sequence must be invariant under the branching $\vec{p_i} \to \vec{p_j}+\vec{p_k}$ whenever $\vec{p_j}$ and $\vec{p_k}$ are parallel (collinear) or one of them is small (infrared). The main clustering algorithm employed at CMS and satisfying these requirements is the *anti-k_T* algorithm [31]. It takes as input the list of particle candidates and makes use of the distance between two of them (i and j) and of the distance from the beam line, defined as:

$$d_{ij} = min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right)\frac{\Delta_{ij}^2}{R^2} \tag{3.14a}$$

$$d_{iB} = \frac{1}{p_{T,i}^2} \tag{3.14b}$$

$\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ is the geometrical distance between the two particles in the (η, ϕ) plane in which the particles are clustered, $p_{T,i}$, η_i and ϕ_i are respectively the transverse momentum, the pseudorapidity and the azimuth of each particle and the parameter R is radius of the cone in the (η, ϕ) plane. The algorithm finds the minimum between d_{ij} and d_{iB}; if the minimum is d_{iB} the algorithm removes the particle i from the list and call it a "jet", otherwise it recombines the particles i and j into a new particle by summing their quadrimomenta. The process is iterated until only jets remain.

The clustering algorithm can take as input the generator level particles, forming the generator level jets or *GenJets*, or the reconstructed particles, forming the *RecoJets*.

For the jets used in this analysis, the anti-k_T clustering algorithm with parameter R = 0.4 is used.

3.2.2 Particle Flow Jet and CaloJet

The Particle Flow (PF) algorithm [32] is a full event reconstruction which aims at identifying and reconstructing all the particles produced in the collision by combining the information of the different subdetectors. This algorithm collects reconstructed hits in each subdetector independently and creates a list of basic reconstructed elements, namely charged tracks in the tracker and clusters of energy deposits in the calorimeters. For this purpose the CMS experiment can rely on an efficient track reconstruction and on an excellent calorimeter granularity that allows to disentangle overlapping showers. Once the different elements are reconstructed, a link algorithm connects blocks which are topologically compatible, forming a list of particle candidates (*PFCandidates*). For example an **electron** PFCandidate arises from the link between a charged track and one or more energy clusters in ECAL, satisfying some identification requirements. In the same way a charged track linked to energy deposits in the calorimeters (ECAL or HCAL), which is not identified as an electron, is reconstructed as a **charged hadron** candidate. An ECAL energy cluster without any tracks is a **photon** candidate, while HCAL deposits with no tracks are reconstructed as **neutral hadrons**. The formation of the PFCandidate list represents the Particle Flow interpretation of a given proton-proton collision. The PF reconstruction allows a major "cleaning" inside a jet through a method called *charged hadron subtraction* (CHS). The CHS consists in removing the PFCandidates that can be associated to another vertex with respect to the primary hard interaction. In fact the additional proton-proton collisions occurring within the same bunch crossing of the primary hard interaction (*in-time pile-up* or IT PU) produce additional tracks in the tracker and energy deposits in the calorimeters. The charged particles coming from IT PU can be therefore removed before the jet clustering, allowing a better and more precise event reconstruction. The remaining PFCandidates are then used as input for the clustering algorithm, which forms the *PFJets*.

The *CaloJet* components consist of energy deposits in the calorimeter towers, where a calorimeter tower consists of one or more HCAL cells and the geometrically corresponding ECAL crystals. In the forward region, where a different calorimeter technology is employed, the Cerenkov light signals collected by the quartz readout fibers aid the separation of electromagnetic and hadronic signals. A four-momentum is associated to each tower deposit taking as a direction the tower position as seen from the interaction point. Once the jet constituents have been reconstructed, the final jet reconstruction is performed by the clustering algorithm, as explained in the previous section.

3.2.3 Online and Offline reconstruction

Both jets fully reconstructed offline or jets reconstructed only at trigger level are used in this analysis. The reason is explained in detail in Chapter 4, but what is relevant here is that the jet reconstruction is slightly different in the two cases.

The difference in the jet reconstruction algorithm running online and offline is mainly dictated by the different constraint on the average processing time per event.

For example the pulses produced by particles in HCAL are wider than the minimum bunch spacing of 25 ns. In fact, about 95% of the total energy deposition is contained in a 50 ns time interval. Particles from subsequent bunch crossings can lead to overlapping pulses (the so-called *out-of-time pile-up* or OOT PU). Two different algorithms are used to mitigate the OOT PU effect. The algorithm used offline is optimized to give the best estimate of the deposited energy and a complete pile-up removal. The online algorithm is designed to perform the same task but within the time constraints of the online system and it is therefore less performing.

A similar approach is used for the measurement of the electromagnetic energy of the jet reconstructed in ECAL. In this case the optimized algorithm is only used in a fraction of the events where the OOT PU effect is larger.

As consequence of the different algorithms, the CaloJets reconstructed at trigger level contain more pile-up and they are reconstructed with poorer resolution, especially at low energies. These differences result in a difference between the online and offline jet quadrimomenta. It is studied in detail in Section 3.3.4 where jet energy corrections "ad hoc" for the online jets are calculated.

3.3 Jet energy calibration

Like all reconstructed objects, jets need to be calibrated in order to have the correct energy scale and resolution: this is the aim of the jet energy corrections (JEC). The JEC are calculated using a full Monte Carlo (MC) simulation of the detector and are then adjusted for data using a combination of several physics channels and data-driven methods. The scheme employed at CMS is based on a factorized approach [33], in such a way that each step addresses different physical aspects.

The three subsequent steps of the JEC, illustrated in Figure 3.3, are:

- *Level-1 offset correction*: Its purpose is to estimate and subtract the energy not associated with the hard scattering. The energy in excess includes contributions from electronics noise and pile-up;
- *Level-2 Relative* and *Level-3 Absolute correction*: These corrections are MC based and they aim at minimizing the effect of non-uniformities between different CMS subdetectors and the non-compensating nature of the calorimeters. In fact the CMS calorimeters are strongly non-compensating having $R_h < R_\gamma$, where R_h and R_γ are respectively the single-particle response of HCAL and ECAL;
- *Level-2* and *Level-3 Residuals correction*: They aim at correcting the residual difference between data and simulation after the application of the previous levels.

In the next sections the different levels will be briefly described.

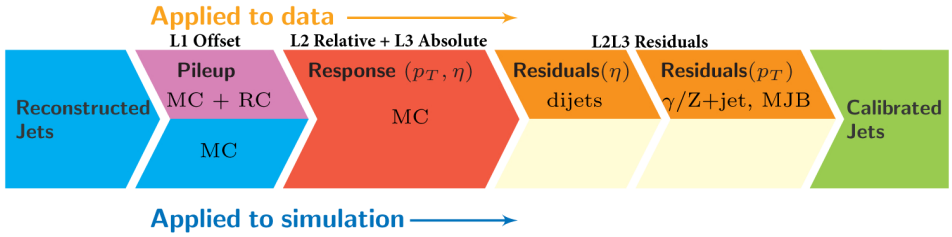

Figure 3.3. Consecutive stages of jet energy corrections, for data (upper row) and simulation (lower row). All the corrections marked with MC are derived from simulation studies, *RC* stands for Random Cone and *MJB* refers to the analysis of multijet events (*MultiJet Balance*).

3.3.1 L1 offset correction

The purpose of the L1 offset correction is to cure the effect of the pile-up and therefore it has a key role since the pile-up increases with the instantaneous luminosity of LHC.

The amount of pile-up present in the event can be estimated from simulation by counting the number of primary vertices or by calculating the diffuse offset p_T density, ρ, in the event. This variable is defined, in each event, as the median of jet momenta p_T divided by their area A_i: $\rho = median(p_T/A_i)$. In order to define the jet area a large number of infinitely soft particles (soft enough not to change the properties of the true jets) are artificially added in the event and clustered by the jet algorithm together with the true jet components. The jet area is then defined as the extent of the region in the (η, ϕ) space occupied by the soft particles clustered in each jet.

The *jet area method*, introduced in [34], uses the area of the jets multiplied by the average p_T density in the event to calculate the offset p_T to be subtracted from the jets. In this method the corrector factor is defined as:

$$C\left(p_T^{raw}, \eta, A_i, \rho\right) = 1 - \frac{\left[\rho_0(\eta) + \rho \cdot \beta(\eta) \cdot \left(1 + \gamma(\eta) \cdot \log(p_T^{raw})\right)\right] \cdot A_i}{p_T^{raw}} \tag{3.15}$$

The input parameters are the uncorrected reconstructed jet transverse momentum p_T^{raw}, the jet pseudorapidity η, the jet area A_j, and the per-event offset p_T density ρ. The parameters $\rho_0(\eta)$, $\beta(\eta)$, and $\gamma(\eta)$, describing the required shaping of the offset versus η, are determined from the simulated particle-level offset.

The pile-up offset in data may be different from the pile-up offset in MC. The differences can arise from the limited accuracy of the description of particle multiplicity and energy spectra in the interactions or it can be associated with the limited accuracy of the detector response for low-p_T particles in simulation. To account for the data/simulation differences, residual L1 offset corrections are derived using data and MC events with no energy deposition from hard interactions (*ZeroBias*) and then they are applied to data. The data to MC scale factor (SF) is defined using the *Random Cone method* as:

$$SF = \frac{O^{RC}_{data}(\eta, \langle\rho\rangle_{data})}{O^{RC}_{MC}(\eta, \langle\rho\rangle_{MC})} \tag{3.16}$$

where

$$O^{RC}(\eta, \langle\rho\rangle) = \langle p_{T,cone}\rangle[\eta, \mu] \tag{3.17}$$

is the average transverse momentum of particles in a randomly placed cone centered at (η, ϕ) and μ is the average number of pile-up interactions per bunch crossing estimated by luminosity monitors.

In Equation 3.17 $(\eta, \langle\rho\rangle)$ explicit marked in the first member are the parameterization variables, while $[\eta, \mu]$ in the second member are the binning variables.

The average p_T offset using the Random Cone method as a function of η is reported for data and MC simulation in Figure 3.4. The offset is normalized by the average number of pile-up interactions and it is separated by the type of PFCandidate. On the bottom panel of the same figure the data/MC scale factor is also reported.

Once jets are corrected using the L1 correction in data and MC, their scale is the same as that for jets without pile-up.

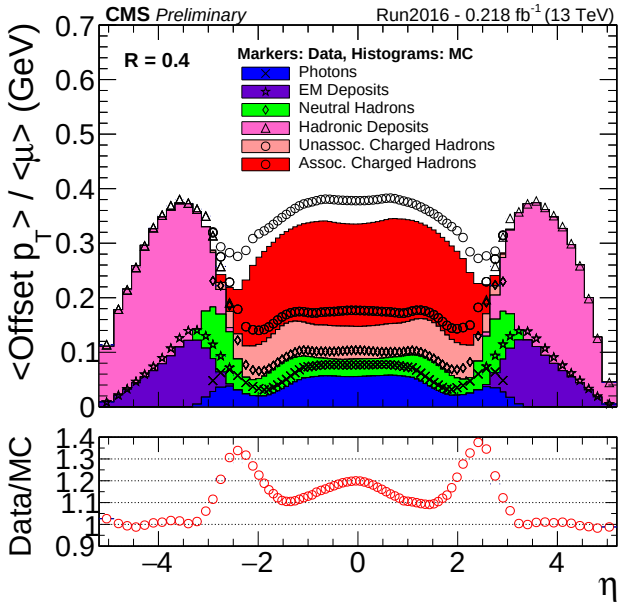

Figure 3.4. The average p_T offset measured in data (markers) and MC simulation (histograms) normalized by the average number of pile-up interactions $\langle\mu\rangle$ and separated by the type of PFCandidate. The portion labeled "charged hadrons" is removed by the charged hadron subtraction. The ratio of data over simulation, representing the scale factor applied for pile-up offset in data, is also shown on the bottom panel.

3.3.2 L2 Relative and L3 Absolute corrections

Jet energy corrections are needed to cure the non-uniformities between different subdetectors and the non-compensating nature of the calorimeters. The L2 Rela-

tive and L3 Absolute corrections are derived from simulation and they are applied to jets which are already corrected for the pile-up offset. The CMS detector simulation contains a detailed model of the detector geometry, data-based alignment and calibration of the detector elements, and emulation of the readout electronics.

The simulated response is defined as the ratio between the average p_T of RecoJets and of the GenJets which are geometrically matched:

$$R(\langle p_{T,reco}\rangle, \eta) = \frac{\langle p_{T,reco}\rangle}{\langle p_{T,gen}\rangle}[p_{T,gen}, \eta] \tag{3.18}$$

A GenJet is matched to the closest RecoJet if it is within half of the cone size $\Delta R < R_{cone}/2$ ($\Delta R < 0.2$ in the case of ak4 jets). The L2 corrections are derived using dijet events. They are called *Relative* because they measure the response of a jet at any η with respect to the response of jets in the region $|\eta| < 1.3$ in order to uniform the response in the detector. After the application of L2 corrections, a similar procedure (L3 Absolute corrections) is applied to uniform the response as a function of the jet p_T.

The simulated jet response as a function of the jet pseudorapidity for different values of the transverse momentum is shown in Figure 3.5.

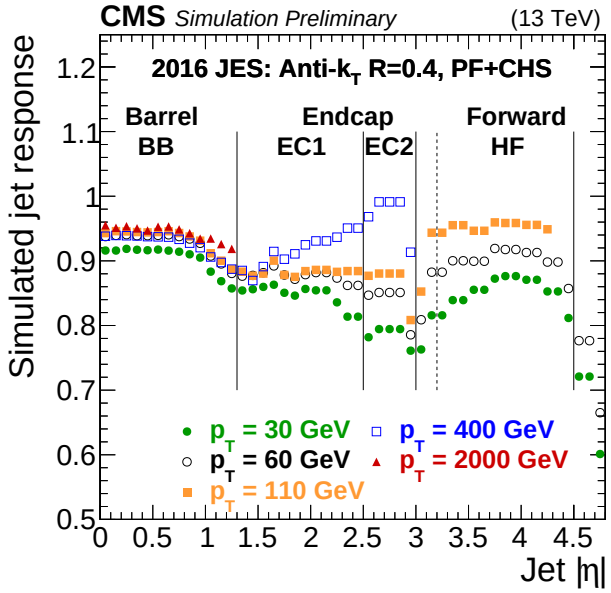

Figure 3.5. The simulated jet response as a function of η for different transverse momenta.

3.3.3 L2L3 Residual corrections

The aim of the L2L3 Residual corrections is to cure the data/MC differences that remain after the previous corrections.

The jet energy response is studied with similar methods of the ones exploited for the L2 Relative and L3 Absolute corrections. The residual corrections are first determined with a sample of dijet events, which have low statistical uncertainty and are used to correct the data to MC ratio as a function of η (L2 Residual JEC). Then

a combination of Z($\rightarrow \mu^+\mu^-$)+jet, Z($\rightarrow e^+e^-$)+jet, γ+jet, and multijet events is used to cure the discrepancy between data and MC as a function of the jet p_T (L3 Residuals).

The basic idea in all the considered topologies is to exploit the transverse momentum balance, at hard-scattering level, between the jet to be calibrated and a reference object: a jet energy response different from unity would generate imbalance at the reconstructed level. Part of the transverse momentum imbalance between the jet to be calibrated and the reference object can also come from the presence of additional jets in the event and therefore this effect is not correlated with the jet energy response. For this reason, all the corrections are studied as a function of the additional jet activity in the event, quantified by the variable α. This is defined as the ratio of the most energetic additional jet divided by the momentum scale of the reference. For example for the γ+jet analysis it is defined as $\alpha = p_{T,jet2}/p_{T,\gamma}$. The corrections are then extrapolated to the value $\alpha = 0$ in order to address only genuine jet energy response effects.

Two methods are employed for the jet energy response calculation: the p_T-balance method in which the jet response is evaluated by comparing the reconstructed jet momentum directly to the momentum of the reference object; and the *missing transverse momentum projection fraction* (MPF) method which considers the response of the whole hadronic activity in the event, recoiling versus the reference object.

The L2 Residual corrections, which fix the data/MC discrepancy as a function of the detector region, are shown in Figure 3.6.

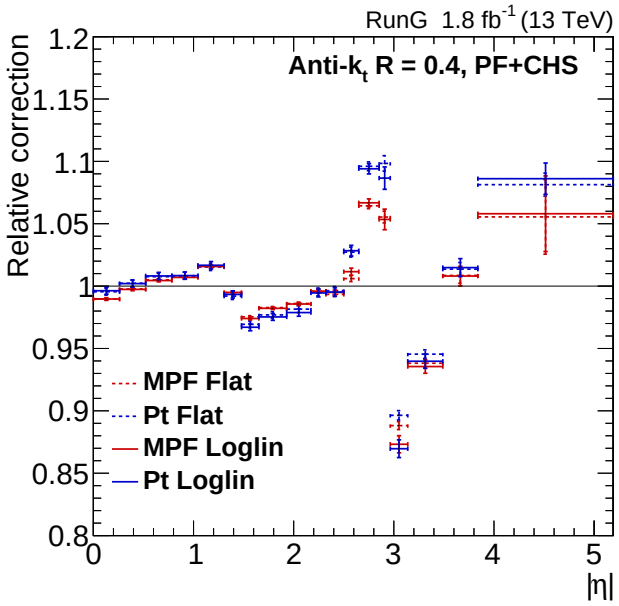

Figure 3.6. L2 Residual JEC as a function of η.

For the L3 Residual correction the ratio between the jet response in data and in MC from Z+jet, γ+jet and multijets events are fitted simultaneously to derive the jet energy correction as a function of the transverse momentum. As an example

the jet energy response calculated with the p_T-balance method using γ+jet events is reported in Figure 3.7. A detailed description of the γ+jet analysis is reported in Appendix A since it was part of my Ph.D. activity.

In Figure 3.8 the data/MC ratios in the jet energy response from the different datasets are reported as well as the fit.

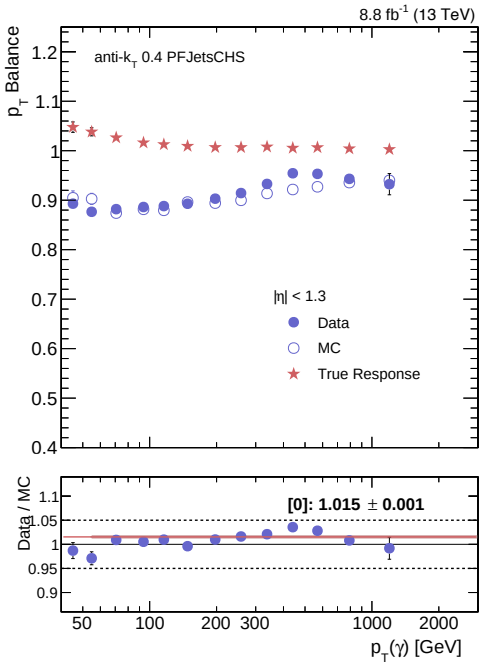

Figure 3.7. Jet energy response obtained from γ+jet events using the p_T-balance method as a function of the p_T. The plot refer to the central region of the detector and the last data period.

The full calibration chain ($L1 + L2 + L3 + L2L3$ *Residuals* JEC) based on 36 fb^{-1} of data is applied to the jets reconstructed offline. In the context of this thesis it is applied to the RECO PFJets.

3.3.4 HLT jet energy scale and resolution

The full set of corrections described in the previous section is derived only for jets reconstructed offline, so another method is necessary in order to correct the energy of jets reconstructed only at trigger level.

The first three levels of corrections ($L1$+$L2$+$L3$), derived from the simulation, are available and are applied to online jets as well. What is missing are JEC for data that bring the energy of the online jets at the same scale of the offline jets. It is desirable therefore to be able to compare HLT data with fully reconstructed data on some sets of events. This allows to compare the energy scale and resolution of the online objects with that of the offline objects, and derive appropriate corrections if needed. For this purpose, a dedicated data stream, called *ParkingScoutingMonitor*, is added to the HLT. This stream contains events for which the same events and

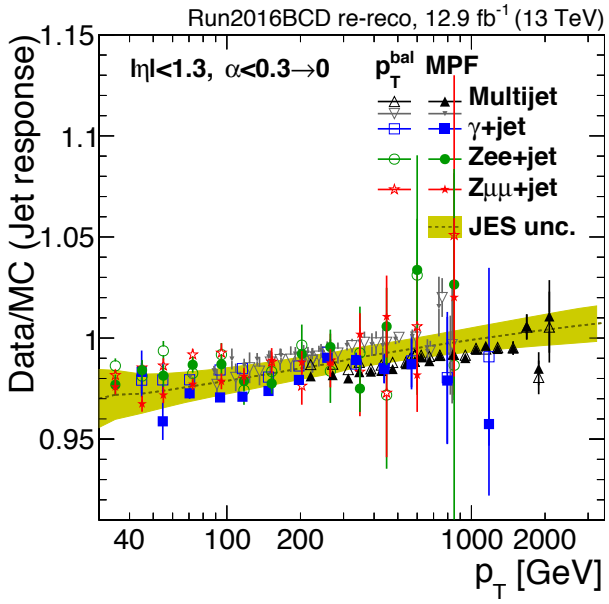

Figure 3.8. Ratio between the jet energy response in data and in MC as a function of the p_T. The results from Z($\rightarrow \mu^+\mu^-$)+jet, Z($\rightarrow e^+e^-$)+jet, γ+jet, and MultiJet analyses are shown. The uncertainty in the data-to-simulation ratio is shown by the shaded region.

objects are saved after both online and offline reconstruction. This allows online objects and offline objects to be compared side-by-side for each event.

The energy resolution for online and offline jets is measured exploiting the p_T balancing in dijet events. The resolution is related to the asymmetry ($\mathcal{A}$) between the transverse momentum of the jets defined as:

$$\mathcal{A} = \frac{p_T^{j_1} - p_T^{j_2}}{p_T^{j_1} + p_T^{j_2}} \tag{3.19}$$

where $p_T^{j_1}$ and $p_T^{j_2}$ refer to the randomly ordered transverse momenta of the two leading jets. The variance of asymmetry can be expressed as:

$$\sigma^2(\mathcal{A}) = \left|\frac{\partial \mathcal{A}}{\partial p_T^{j_1}}\right|^2 \sigma^2(p_T^{j_1}) + \left|\frac{\partial \mathcal{A}}{\partial p_T^{j_2}}\right|^2 \sigma^2(p_T^{j_2}). \tag{3.20}$$

For two jets in the same η region in an ideal dijet topology (with the jets balanced at particle level), the average p_T and widths of their transverse momentum distributions are assumed to be equal:

$$\langle p_T^{j_1} \rangle = \langle p_T^{j_2} \rangle \equiv \langle p_T \rangle \tag{3.21a}$$

$$\sigma(p_T^{j_1}) = \sigma(p_T^{j_2}) \equiv \sigma(p_T) \tag{3.21b}$$

This allows the simplification of Equation 3.20 so that the jet transverse momentum resolution $\sigma(p_T)$ can be written as:

$$\frac{\sigma(p_T)}{p_T} = \sqrt{2} \cdot \sigma(\mathcal{A}) \tag{3.22}$$

In the context of this thesis, the jets p_T asymmetry is not related to the jets resolution, but to the invariant mass of the dijet system, which is the most important variable of the analysis detailed in the next chapters. The outcome is therefore an event correction to be applied at analysis-level, and not a real object calibration.

In general the dijet mass resolution is affected by two terms, one related to the jets energies and one related to the angular separation between them. However the angular term can be assumed negligible for back-to-back objects and the resolution of the invariant mass of two jets, m_{jj}, can be written as:

$$\frac{\sigma(m_{jj})}{m_{jj}} = \frac{1}{2}\sqrt{\left(\frac{\sigma(E_1)}{E_1}\right)^2 + \left(\frac{\sigma(E_2)}{E_2}\right)^2} \tag{3.23}$$

$$= \frac{1}{2}\sqrt{2\left(\frac{\sigma(E)}{E}\right)^2} \tag{3.24}$$

$$= \frac{\sigma(E)}{\sqrt{2}E} \tag{3.25}$$

$$\sim \frac{\sigma(p_T)}{\sqrt{2}p_T} \tag{3.26}$$

Comparing Equations 3.22 and 3.26, the dijet mass resolution is related to the asymmetry resolution as:

$$\sigma(\mathcal{A}) \sim \frac{\sigma(m_{jj})}{m_{jj}} \tag{3.27}$$

In Figure 3.9 two examples of asymmetry distributions, computed in different bins of the dijet invariant mass, are shown.

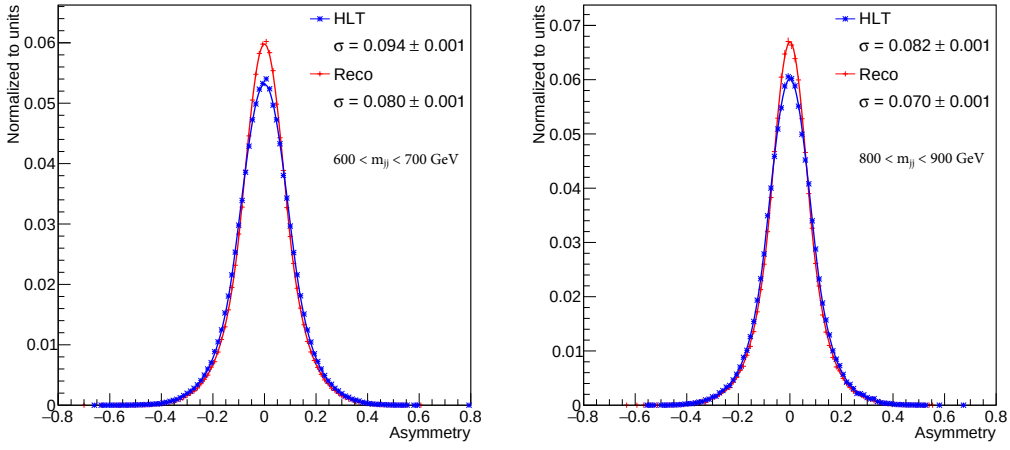

Figure 3.9. The transverse momentum asymmetry computed in dijets events using RECO PFJets (red) and HLT CaloJets (blue). The distributions refer to different m_{jj} bins: on the left $600 < m_{jj} < 700$ GeV and on the right $800 < m_{jj} < 900$ GeV.

The width of the asymmetry as a function of m_{jj} is extracted by a Gaussian fit to these distributions and it is fitted with the function:

$$\sigma(\mathcal{A}) = \sqrt{\frac{N^2}{m_{jj}^2} + \frac{S^2}{m_{jj}} + C^2} \tag{3.28}$$

The trend for HLT and RECO objects as a function of the dijet invariant mass is reported in Figure 3.10 from which it is clear that the HLT and RECO jets have different energy resolution. This will be taken into account in the analysis as explained in detail in Section 4.4.1.

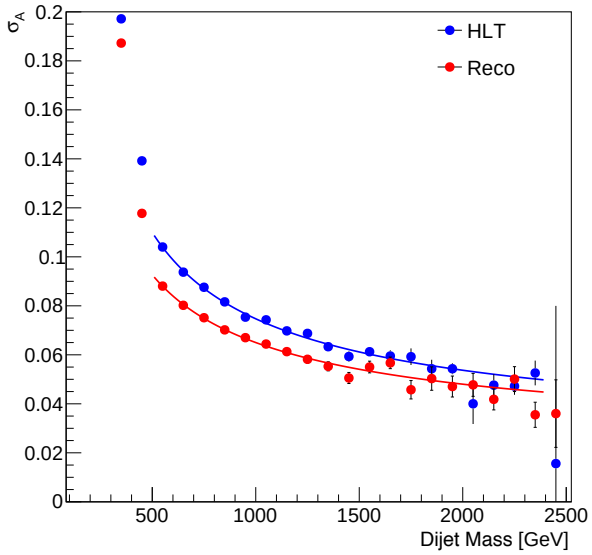

Figure 3.10. The variance of the asymmetry for RECO PFJets (red) and for HLT CaloJets (blue) as a function of the dijet invariant mass.

The comparison between the jet energy scale in the online and offline reconstruction is performed using a "tag-and-probe" method to the events in the overlap sample. The method consists in using the RECO PFJets as reference objects and compare them with the HLT CaloJets. For the RECO jets both permutations of tag and probe are used in turn to increase the statistics, while the HLT jet matching the probe RECO jet is used as a probe. The HLT-to-RECO JEC is defined as the ratio between the transverse momentum of the HLT CaloJet and of the RECO PFJet as a function of the HLT CaloJet p_T. This additional scale correction to be applied on the HLT data is shown in Figure 3.11.

After applying these corrections, the HLT CaloJets have the same energy scale of the RECO PFJets in the whole dijet mass range of interest, as shown in Appendix B.

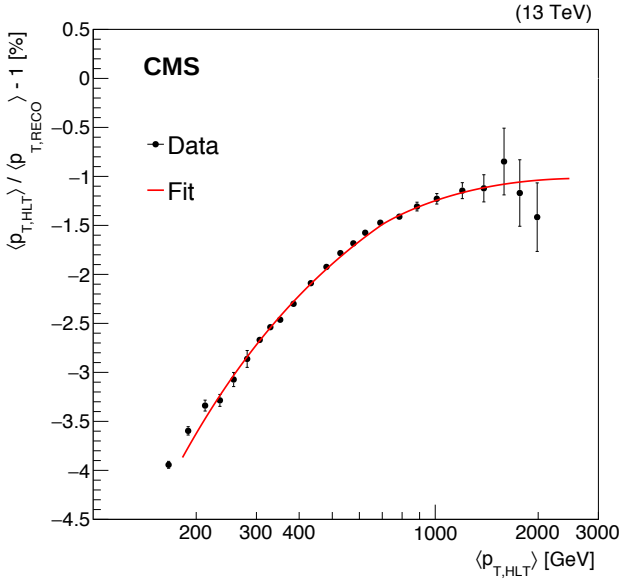

Figure 3.11. The ratio between the transverse momentum of HLT CaloJets and the RECO PFJets as a function of the HLT jet p_T.

Chapter 4

Analysis strategy

The searches in fully-hadronic final states are among the most challenging ones, due to the huge background and the technical difficulty of fitting the large rate of QCD events within the boundaries of the data acquisition system and of the computing and disk resources. Due to these limitations, that become more severe when the collider peak luminosity increases, the thresholds for dijets trigger paths are continuously raised, in order to keep a constant throughput despite the increasing collision rate. As a consequence, the searches with fully hadronic final states can only probe the tails of the kinematic distributions and therefore they leave unexplored territories in the context of the search for beyond Standard Model physics (BSM). This progressive loss of sensitivity at low masses is common to all previous and current experiments at hadron colliders, as discussed in [35]. To compensate for this loss, the CMS experiment can rely on the *data scouting* technique [36]. It allows to reduce significantly the trigger thresholds by reducing the events size compared to the standard CMS data stream, as explained in the next section.

The analysis discussed in this thesis relies both on the standard data stream ("*high-mass*" analysis, $m_{jj} > 1.2$ TeV) and on the data scouting technique ("*low-mass*" analysis, $500 < m_{jj} < 2000$ GeV).

The analysis strategy is common for both analyses and it consists in the search for a peak in the dijet invariant mass distribution over the falling QCD background.

4.1 Data Scouting

The data scouting technique was introduced during LHC Run1 [37] to search for new physics in the dijet final state.

This technique is based on a strong reduction of the event size in order to accept a higher event rate due to lowering the trigger thresholds. The data transfer rate [GB/s] is defined as the event size [GB/event] times the event rate [event/s]. It is clear therefore that if we want a higher event rate we have to lower the event size. In this way it is possible to keep the data transfer rate at a sustainable value for CMS ($\sim$2 GB/s).

The events are written in a slim event content (between 5 and 10 kB per event), tuned on the specific use case of the all-hadronic BSM searches. This small event size compensates the large event rate, resulting in a negligible overhead for the

global rate of data written on disk. The price to pay to have a small event size is the fact that the full event information can not be stored and therefore the object reconstruction is only run online. For this reason the jets used in the *low-mass* analysis are reconstructed at HLT and dedicated energy corrections are needed to bring them to have the same energy scale and resolution as the offline jets, as discussed in Section 3.3.4.

4.2 Data samples

This analysis is based on data collected during the 2016 data taking, with the LHC operating at the bunch spacing of 25 ns and with the center-of-mass energy of 13 TeV.

The events used for the *high-mass* analysis are collected in the primary dataset *JetHT* and the total integrated luminosity is 36 fb^{-1}.

For the *low-mass* analysis the data are collected in the CMS *CaloScouting* stream. Due to some trigger inefficiencies in the last part of the data-taking, the total integrated luminosity used for this part of the analysis is 27 fb^{-1}.

4.3 Triggers

The datasets used in this analysis are based on two kinds of trigger: single-jet triggers, which require at least one jet in the event with p_T above a certain threshold, and triggers based on the variable H_T, defined as the scalar sum of the jets p_T considering all jets in the event with $|\eta| < 3$. The jets considered for trigger purpose can be PFJets or CaloJets depending on the definition of the trigger.

For the *high-mass* analysis we use all data in the JetHT dataset, regardless of which trigger was used to record an event. The main triggers are *PFHT800* and *PFHT900*, which require H_T, calculated with PFJets, greater than 800 GeV and 900 GeV respectively. In the last part of the data-taking, due to some inefficiencies of the above paths, other triggers also became relevant: the *PFJet500*, the *CaloJet500_NoJetID* and the *AK8PFJet450* (the definitions are reported in Table 4.1).

The trigger efficiency is measured with respect to the *SingleMuon45* path (defined in Table 4.1) that is a trigger fully orthogonal to the signal ones in the sense that the selection at HLT is uncorrelated with the selection of the considered triggers and uses a different L1 seed.

Trigger Name	Definition
PFHT800	$H_T > 800$ GeV
PFHT900	$H_T > 900$ GeV
PFJet500	p_T(PFJet) > 500 GeV
CaloJet500_NoJetID	p_T(CaloJet) > 500 GeV
AK8PFJet450	p_T(anti-k_T(R=0.8) PFJet) > 500 GeV
SingleMuon45	one muon with $p_T > 45$ GeV

Table 4.1. Triggers used for the *high-mass* analysis and their definition.

The trigger efficiency is calculated as a function of the dijet mass using:

- as denominator: the dijet mass distribution of the events passing the dijet selection (described in Section 4.5) and the reference trigger *SingleMuon45*;
- as numerator: the dijet mass distribution of the events that satisfy the denominator requirements and the logical *OR* between the signal triggers.

In Figure 4.1 the trigger efficiency as a function of the dijet invariant mass is reported.

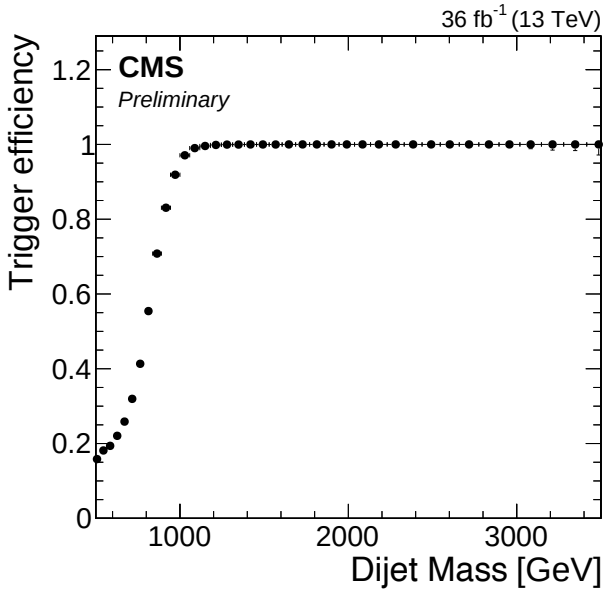

Figure 4.1. Trigger efficiency as a function of the dijet invariant mass for the *high-mass* analysis.

The trigger is considered fully efficient when the measured trigger inefficiency in a bin of the mass spectrum is smaller than the fractional statistics uncertainty on the number of events in the same bin:

$$1 - eff(bin) \leq \frac{1}{\sqrt{N(bin)}} \tag{4.1}$$

This condition is satisfied for masses above 1246 GeV, which is the region probed by the *high-mass* analysis, to ensure that the results are not biased by trigger inefficiency.

For the *low-mass* search the triggers are based on H_T computed using CaloJets with $p_T > 40$ GeV and the main trigger is the *CaloScoutingHT250*. In this case the trigger efficiency is measured with respect to a sample satisfying a CaloScouting trigger which requires only that the event fired a L1 seed based on H_T. The definition of the triggers for the *low-mass* search is reported in Table 4.2.

The trigger efficiency is reported in Figure 4.2 and the fully efficiency is reached at mass 489 GeV. Therefore the data scouting technique allows to probe the mass region between 489 and 1246 GeV, otherwise unattainable with the *high-mass* analysis.

Trigger Name	Definition
CaloScoutingHT250	$H_T(CaloJet) > 250$ GeV
L1HTT	No requirement at HLT;
	at L1 OR between: H_T >200; 240; 270; 280; 300; 320 GeV

Table 4.2. Triggers used for the *low-mass* analysis and their definition.

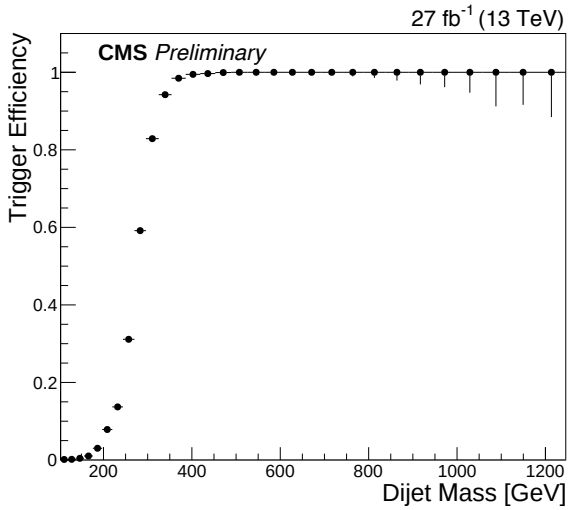

Figure 4.2. Trigger efficiency as a function of the dijet invariant mass for the *low-mass* analysis.

4.4 Monte Carlo simulation

The Monte Carlo samples for signal and background are generated with PYTHIA 8 [38] using the CUETP8M1 tune [39, 40]. The events are passed through a complete simulation of the CMS detector based on GEANT4 [41]. Furthermore pile-up events are injected in the simulation to mimic the LHC operational conditions.

4.4.1 Signal models

The signal samples are produced for different masses and widths of the resonance. For *narrow* resonances, for which the natural width is much smaller than the experimental resolution, the dijet mass distribution is mostly determined by the final state and it is independent from the spin, while for *wider* resonances it also depends on the spin. In both cases the signal models are employed in the search for resonances decaying into two jets and in the dark matter mediator search.

For the narrow resonance and narrow mediator search, the available MC samples cover three combinations of production and decay modes:

- $gg \to G \to gg$
- $qq \to G \to qq$

- $qg \rightarrow q^* \rightarrow qg$

where "g" indicates the gluons, "q" any quark/antiquark, "G" stands for the Randall-Sundrum Graviton and q^* is an excited quark. For each combination 11 samples corresponding to different resonance masses (500 GeV, 750 GeV, and 1 TeV to 9 TeV in 1 TeV spacing) were generated. The binned distribution of the dijet invariant mass from each of these MC samples represents the signal model of the analysis used in the limit setting procedure, as explained in Section 5.1.

In order to produce invariant mass shapes at intermediate resonance masses (every 50/100 GeV for the *low-mass/high-mass* analysis), a vertical interpolation technique is used. First, a parameter X is introduced as $X = M_{jj}/M_{Res}$, where M_{jj} is the dijet mass and M_{Res} is the resonance mass. Some example distributions for X are shown in Figure 4.3.

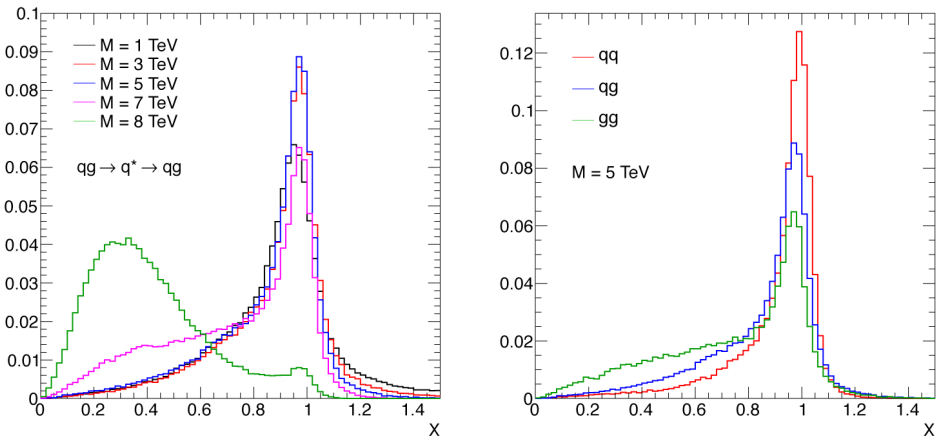

Figure 4.3. (Left) The distributions of X for a qg resonance and various mass points. (Right) Comparison of X distributions at the resonance mass of 5 TeV for qq, qg and gg resonances.

The X distribution is generated for intermediate resonance masses using its neighbor existing X distributions applying the following expression:

$$Prob_{M_i}(X) = Prob_{M_{i-1}}(X) + [Prob_{M_{i+1}}(X) - Prob_{M_{i-1}}(X)] \cdot \frac{M_i - M_{i-1}}{M_{i+1} - M_{i-1}} \quad (4.2)$$

where $Prob_M(X)$ is the probability distribution as a function of X for a resonance with mass M. Finally, the new generated X distribution is converted to dijet invariant mass distribution to get the resonance shape at any resonance masses.

The signal models for the *low-mass* analysis are obtained smearing the invariant mass shapes used in the *high-mass* analysis and described so far to match the resolution in HLT data. In fact, as discussed in Section 3.3.4, the HLT CaloJets have different resolution with respect to the RECO PFJets due to the different reconstruction and algorithm and this difference has to be taken into account for the resonance shapes. The smearing factor is extracted randomly from a Gaussian with mean 1 and width equal to the difference in quadrature between the measured energy resolution for HLT CaloJets and RECO PFJets:

$$\sigma(Gaussian) = \sqrt{\left(\frac{\sigma(m_{jj})}{m_{jj}}\right)^2_{RECO} - \left(\frac{\sigma(m_{jj})}{m_{jj}}\right)^2_{HLT}} \tag{4.3}$$

Some of the signal resonance shapes are shown in Figure 4.4 as well as the effect of the smearing procedure. The shapes for resonances decaying into gg, qg and qq differ due to the differences between quarks and gluons:

- gluons emit more radiation than quarks, which causes the resonance width to increase with the number of gluons in the final state;
- the low-mass tail is due to the combined effect of the final state radiation and the parton distribution function since the parton luminosity is higher at low-mass than at high-mass;
- the high-mass tail is smaller than the low-mass tail and it is caused by the initial state radiation.

Even if several theories are used in the results interpretation, the invariant mass shapes obtained with the RS Graviton and excited quark models are approximately valid for any other model of narrow resonance involving the same pair of partons.

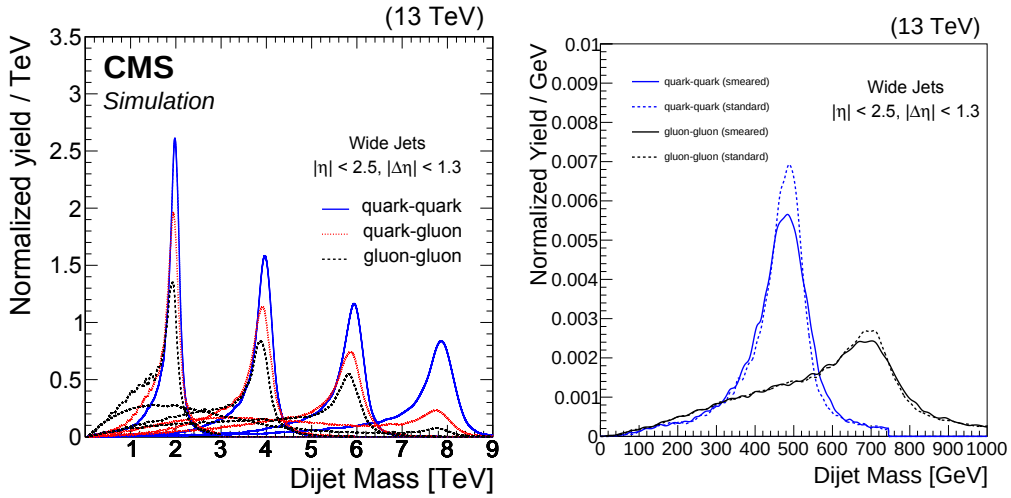

Figure 4.4. (Left) Signal shapes for $gg \to \mathrm{G} \to gg$, $qq \to \mathrm{G} \to qq$, $qg \to q^* \to qg$ resonances with a mass of 2, 4, 6 and 8 TeV. (Right) The signal shape for qq and gg resonance with a mass of 500 and 750 GeV before (dotted) and after (line) the additional smearing applied for the *low-mass* analysis.

For the wide resonance search, for which the resonance spin has to be taken into account, the signal shapes are modeled on a spin-2 RS Graviton decaying to qq and gg for the resonance search and on a spin-1 dark matter mediator for the dark matter mediator search. The signal samples are generated varying the ratio of the 5-dimensional curvature to the reduced Planck mass, $k/\overline{M}_{\mathrm{Pl}}$, which acts as the coupling constant of the model [20]. The coupling parameter is related to the resonance width by the relation [42]:

$$\Gamma/M \approx 1.4(k/\overline{M}_{\mathrm{Pl}})^2 \tag{4.4}$$

The samples are produced at the same masses as for the narrow resonance case, which means from 1 to 9 TeV in bins of 100 GeV, and with Γ/M equal to 0.00001, 0.01, 0.05, 0.1, 0.15, 0.2, 0.3.

The dijet mass distributions for the resonance search are shown in Figure 4.5 for some masses and some width-to-mass ratios as an example.

For the wide dark matter mediator search, the signal shapes for different values of the resonance mass M and Γ/M for a vector dark matter mediator are shown in Figure 4.6. These shapes are indistinguishable from those for a wide axial-vector mediator model.

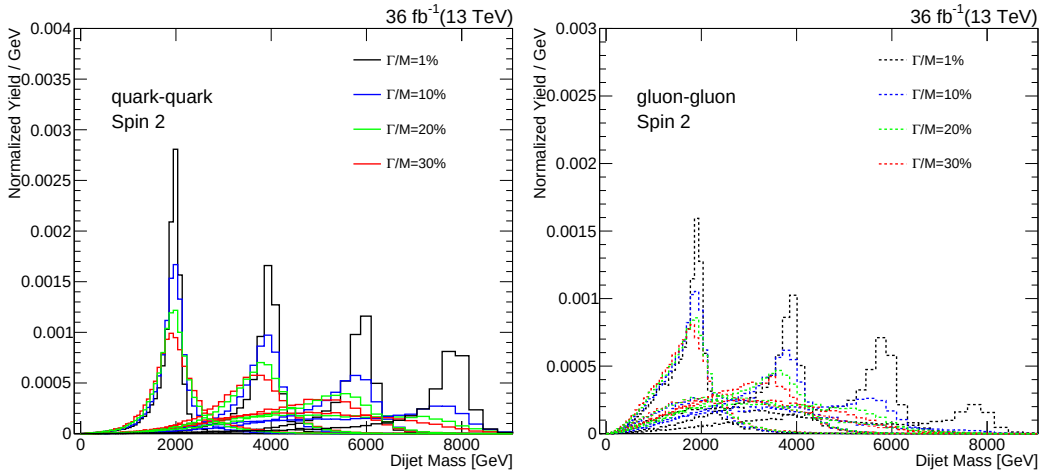

Figure 4.5. Signal shapes for qq (left) and gg (right) resonances with masses of 2, 4, 6 and 8 TeV and Γ/M = 1%, 10%, 20% and 30%.

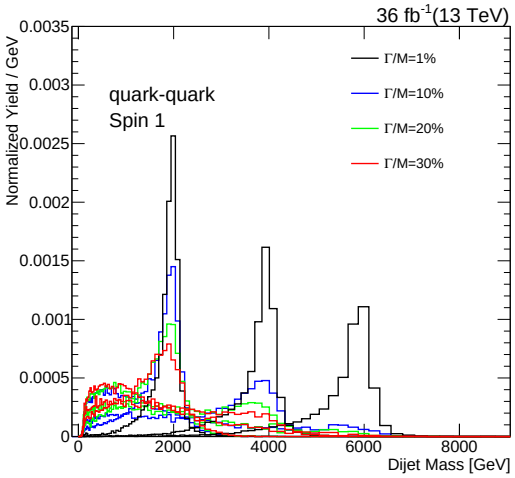

Figure 4.6. Signal shapes for vector dark matter mediator with masses of 2, 4 and 6 TeV and Γ/M = 1%, 10%, 20% and 30%.

4.4.2 Background samples

The dominant background for this analysis is the QCD production of two or more jets. The MC samples for the QCD multijet background are not used in the analysis to produce physics results, but they are used to check the data quality.

The samples are binned in the transverse momentum of the hard scattering process $\hat{p}_T$ starting from 50 GeV. The list of datasets with the corresponding cross section, number of events, and equivalent integrated luminosity is reported in Table 4.3.

QCD $\hat{p}_T$-binned [GeV]	Cross section [pb]	Events	Equivalent luminosity [fb^{-1}]
$50 < \hat{p}_T < 80$	19204300	9,968,410	5.191×10^{-4}
$80 < \hat{p}_T < 120$	2762530	9,968,410	0.003608
$120 < \hat{p}_T < 170$	471100	6,863,827	0.01457
$170 < \hat{p}_T < 300$	117276	6,914,086	0.05896
$300 < \hat{p}_T < 470$	7823	5,970,600	0.7632
$470 < \hat{p}_T < 600$	648.2	3,928,870	6.061
$600 < \hat{p}_T < 800$	186.9	3,959,768	21.19
$800 < \hat{p}_T < 1000$	32.293	3,924,080	121.5
$1000 < \hat{p}_T < 1400$	9.4183	2,999,069	318.4
$1400 < \hat{p}_T < 1800$	0.84265	396,409	470.4
$1800 < \hat{p}_T < 2400$	0.114943	396,100	3.446×10^3
$2400 < \hat{p}_T < 3200$	0.00682981	399,226	5.845×10^4
$\hat{p}_T > 3200$	0.000165445	383,926	2.321×10^6

Table 4.3. Background MC samples binned in $\hat{p_T}$ with the corresponding cross section, number of events, and equivalent integrated luminosity for each bin.

4.5 Event selection criteria

The analysis uses Particle Flow jets and CaloJets reconstructed with the anti-k_T algorithm with a distance parameter of $R = 0.4$ and calibrated as discussed in Chapter 3.3. The PFJets are used for the *high-mass* analysis and are reconstructed offline. The CaloJets are used for the *low-mass* analysis and are reconstructed only at trigger level. In both cases, the jets are required to have $p_T > 30$ GeV and $|\eta| < 2.5$. Also, they are required to pass a set of jet quality criteria ("Jet ID"), to remove spurious jets associated with calorimeter and/or readout electronics noise. The Jet ID is mainly based on the jet energy composition and takes into account the energy fraction from charged hadrons (CHF), neutral hadrons (NHF), muons (MuF), neutral electromagnetic deposits (NEF) and charged electromagnetic deposits (CEF). Its performances are studied in pure noise non-collision data samples, such as cosmics or data from triggers on empty bunches during LHC operation. For PFJets we use a tight working point corresponding to the requirements listed in Table 4.4. It has an efficiency of ∼99% and a background rejection of ∼100% in the region of interest. For online CaloJets the track reconstruction is not performed and therefore the charged and neutral energy fractions can not be distinguished. For this reason

the Jet ID for CaloJets only requires only that the electromagnetic and the hadronic energy fractions are between 5% and 95% of the total jet energy. For both analyses an event is rejected if one of the two jets with the highest transverse momentum (*leading jets*) fails the Jet ID criteria.

Variable	Cut
NHF	< 0.90
NEF	< 0.90
MuF	< 0.8
# of constituents	> 1
and in addition for jets $\|\eta\| < 2.4$:	
CHF	> 0
CEF	< 0.90
Charged multiplicity	> 0

Table 4.4. The requirements of the "Tight" Jet ID applied on the RECO PFJets.

The jets passing the selection are used as input to an additional reclustering, with a larger cone, performed to recover the energy loss due to the final state radiation (*FSR*). The two leading jets are used as seeds to build larger jets that collect any other jet geometrically close to one of the seeds within a distance of $\Delta R = 1.1$. The outputs of the reclustering are two so-called "WideJets" for which a sketch is reported in Figure 4.7. The WideJets radius $\Delta R = 1.1$ has been optimized minimizing the expected upper limit on the cross section. The improvement in the dijet mass reconstruction is shown in Figure 4.8, where we compare the dijet invariant mass reconstructed using ak4 jets and WideJets. An excited quark resonance with mass $M_{q*} = 4.0$ TeV that decays in one quark and one gluon is shown as an example signal.

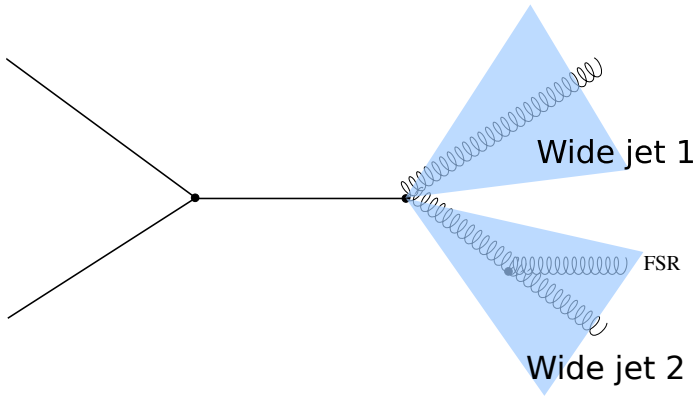

Figure 4.7. The energy loss due to the FSR is recovered and included into the "Wide jet 2" thanks to the reclustering.

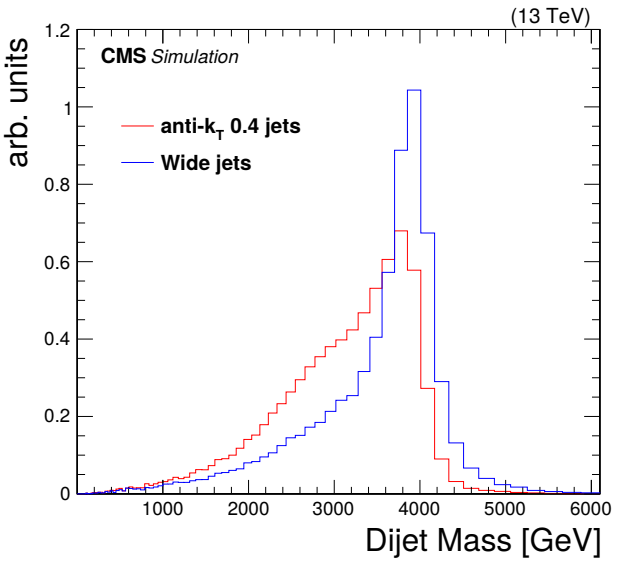

Figure 4.8. Comparison between the dijet mass distribution calculated with the ak4 jets (red) and with the WideJets (blue). An excited quark resonance with $M_{q*} = 4.0$ TeV decaying in one quark and one gluon is simulated.

In order to suppress the QCD background it is possible to exploit the angular distribution of the jets. As shown in Figure 4.9, the resonant signal is produced via a s-channel diagram, while the background via the t-channel.

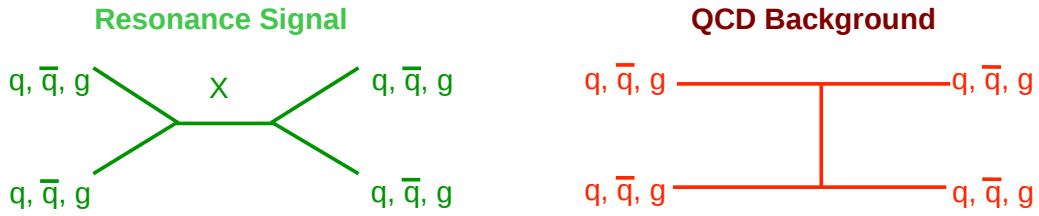

Figure 4.9. The Feynman diagram for a resonant signal and for the background. The signal is produced via the s-channel diagram, while the background is in t-channel.

This implies that the distribution of the scattering angle θ^* is significantly different between the QCD interactions and the signal processes. In the QCD events the jets are mostly produced in the forward regions of the detector exhibiting a pole at $\cos(\theta^*) \to 1$. The signal processes instead have different distributions in $\cos(\theta^*)$ depending on the spin of the resonance. For example, as shown in Figure 4.10, the decay products of a half-integer spin resonance would have an equal probability of being emitted at every angle θ^* showing a flat distribution as a function of $\cos(\theta^*)$.

To enhance the analysis sensitivity it is possible to exploit this difference selecting a range of $\cos(\theta^*)$ or, equivalently, of $\Delta\eta_{jj}$ thanks to the relation:

$$\cos(\theta^*) = \tan(\Delta\eta_{jj}/2) \tag{4.5}$$

where $\Delta\eta_{jj}$ is the difference between the pseudorapidity of the two jets ($\Delta\eta_{jj} = \eta_{j_1} - \eta_{j_2}$). The optimal cut value, chosen maximizing the signal over background ratio, is $|\cos\theta^*| < 0.57$ or equivalently $|\Delta\eta_{jj}| < 1.3$.

Finally the m_{jj} cuts ($m_{jj} > 489$ GeV for the *low-mass* analysis, $m_{jj} > 1246$ GeV for the *high-mass* analysis) derived from the trigger studies presented in Section 4.3, are applied.

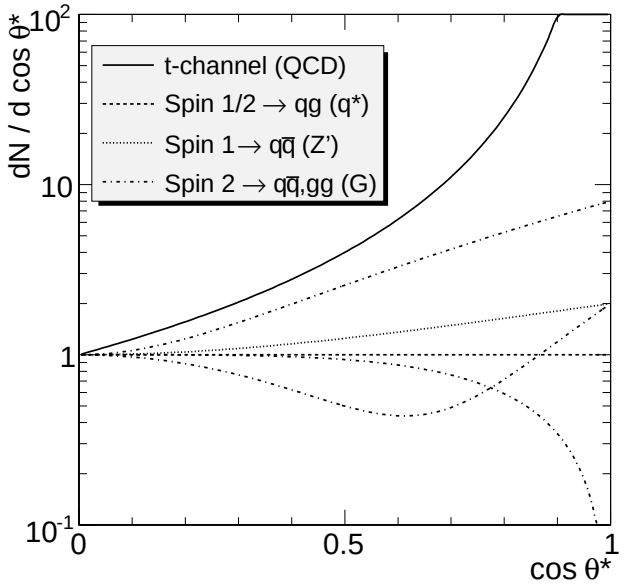

Figure 4.10. Comparison between the angular distribution of the jets for the t-channel QCD background (solid line) and for s-channel resonant processes with different spin and final states (dotted lines).

4.6 Data quality

A detailed comparison between data and QCD MC simulation is performed with the purpose to examine and study the quality of data and how robust the event selection is against beam and detector related noise, detector pathologies, reconstruction failures, etc. In this section the QCD prediction is shown normalized to the number of events in data by multiplying it by a factor of 0.87 (data/MC) for the *high-mass* analysis and a factor of 0.96 for the *low-mass* search.

In Figure 4.11 the azimuthal separation of the two WideJets, $\Delta\phi = |\phi_1 - \phi_2|$, is shown. It displays the typical "back-to-back" distribution expected for the QCD production of two jets, with good agreement between data and prediction over six orders of magnitude in the number of events. The strong peak at $\Delta\phi = \pi$, with very few events in the region $\Delta\phi = 0$, shows that the data are clean. Such distribution is physically produced from parton-parton scattering, while background from detector noise or other non-physical sources would produce events more isotropic in $\Delta\phi$.

The distribution of $|\Delta\eta| = |\eta_1 - \eta_2|$, reported in Figure 4.12, shows how forward the dijet production is. It can be noticed that the distribution is dominated by a t-channel parton exchange as expected for the QCD production of two jets, with a

production rate that increases with increasing $|\Delta\eta|$, while a s-channel signal from a dijet resonance would decrease with increasing $|\Delta\eta|$.

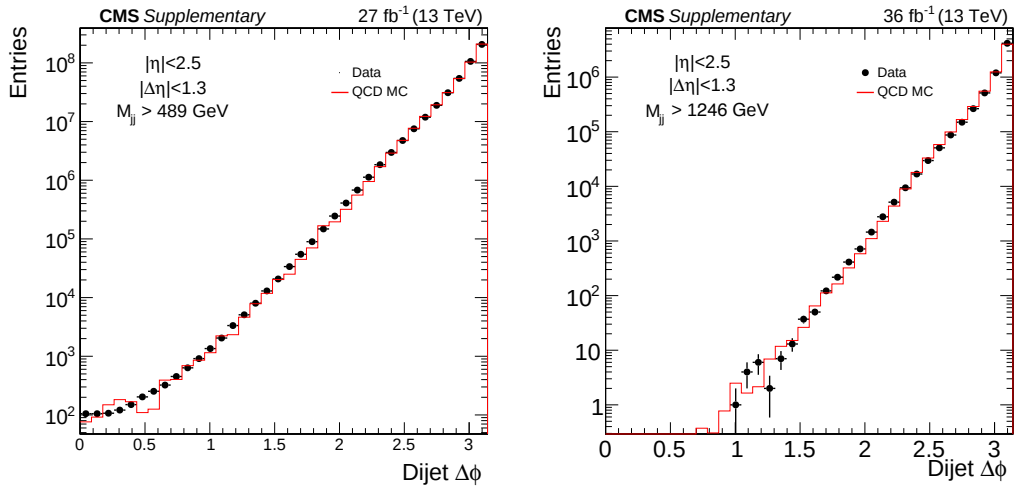

Figure 4.11. The azimuthal angular separation between the two selected WideJets, for the *low-mass* search (left) and the *high-mass* search (right). Data (points) are compared to QCD predictions from PYTHIA 8 (histogram) normalized to the data.

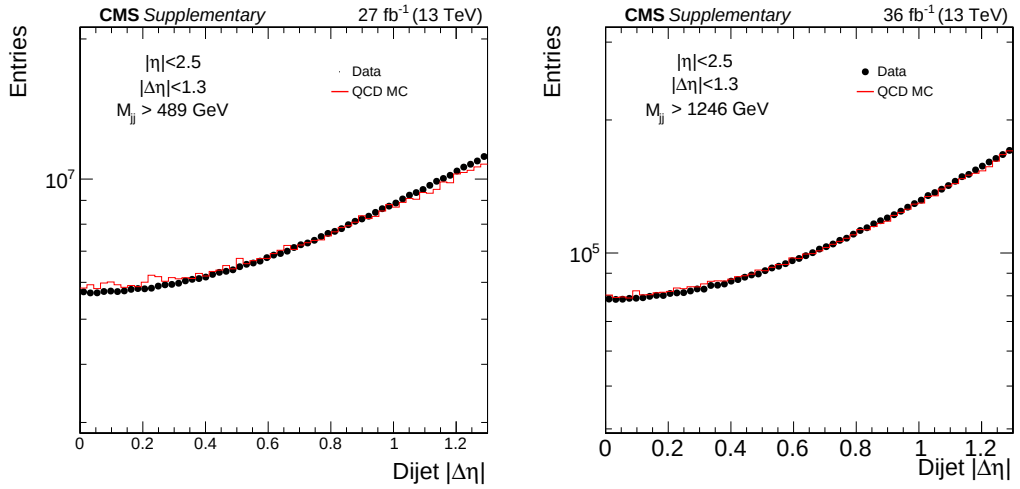

Figure 4.12. The $\Delta\eta$ of the two selected WideJets, for the *low-mass* search (left) and the *high-mass* search (right). Data (points) are compared to QCD predictions from PYTHIA 8 (histogram) normalized to the data.

The last comparison reported in this section is the dijet mass distribution. It can be observed from Figure 4.13 that the number of produced dijet events falls steeply and smoothly as a function of dijet mass. The observed dijet mass distributions are similar to the QCD prediction from PYTHIA, which includes a leading order QCD calculation and parton shower. In Figure 4.14 we compare the dijet mass data to the QCD prediction from POWHEG [43], which is a next to leading order QCD calculation and parton shower. It is clear from these comparisons that the

dijet mass spectrum behaves approximately as expected from QCD predictions. However, even if the POWHEG prediction models the data better than PYTHIA, intrinsic uncertainties within QCD calculations make it challenging to use them to accurately predict the background in a dijet resonance search. For this reason only data are used to estimate the background.

I show here only a few examples quantities, but many others were compared (e.g. the jet kinematic variables), all leading to the same conclusion: data and MC are in good agreement, showing that the data sample is clean with no pathologies and no indication of noise present after the dijet selection.

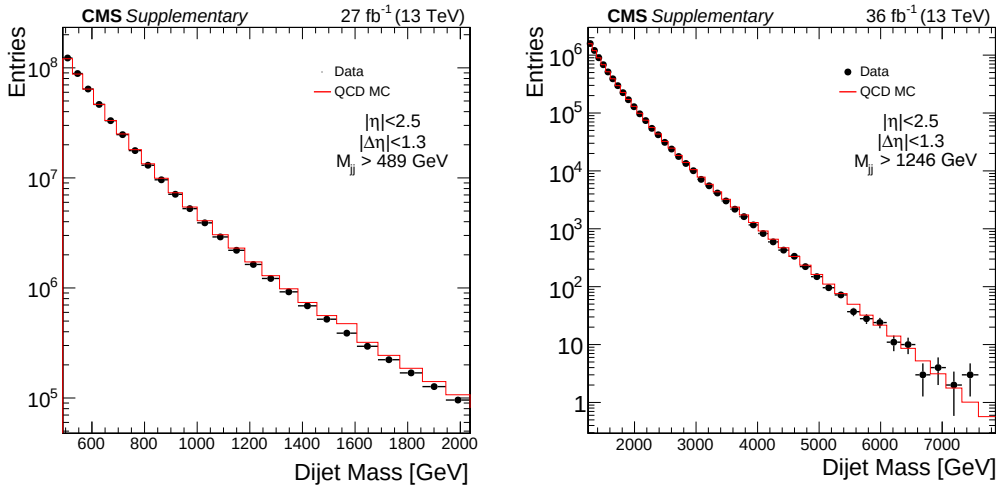

Figure 4.13. The invariant mass spectrum of the two selected WideJets, for the *low-mass* search (left) and the *high-mass* search (right). Data (points) are compared to QCD predictions from PYTHIA 8 (histogram) normalized to the data.

4.7 Background estimate

The background estimate is obtained directly from data and does not rely on Monte Carlo simulations. It is obtained performing a likelihood fit to the observed dijet mass distribution in the background-only hypothesis. The dijet mass spectra are fit separately for the *low-mass* and *high-mass* searches.

The fit function for the *high-mass* analysis, defined as:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}m_{jj}} = \frac{p_0(1-m_{jj}/\sqrt{s})^{p_1}}{(m_{jj}/\sqrt{s})^{p_2+p_3\log(m_{jj}/\sqrt{s})}} \tag{4.6}$$

is a four-parameter continuous function that has been used extensively in previous searches to fit the falling dijet spectrum [44, 45].

The fit function for the *low-mass* analysis is a five-parameter continuous function, defined as:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}m_{jj}} = \frac{p_0(1-m_{jj}/\sqrt{s})^{p_1}}{(m_{jj}/\sqrt{s})^{p_2+p_3\log(m_{jj}/\sqrt{s})+p_4\log^2(m_{jj}/\sqrt{s})}} \tag{4.7}$$

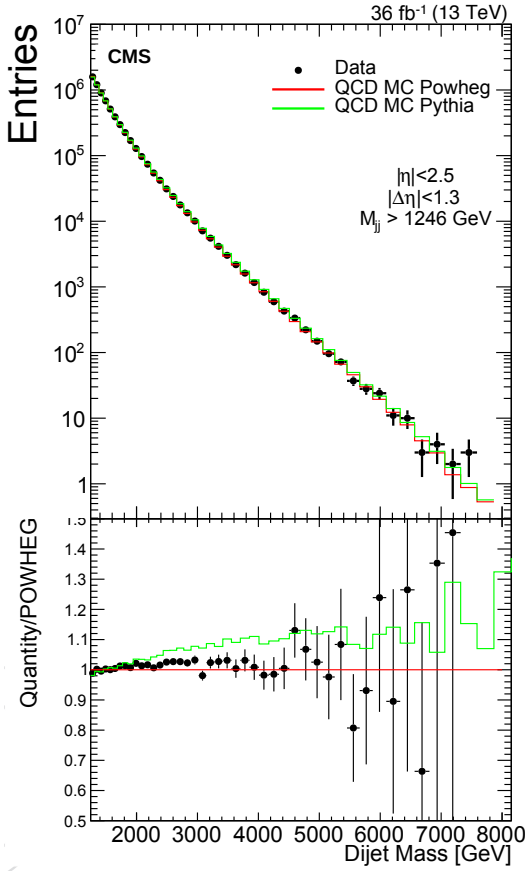

Figure 4.14. The invariant mass spectrum of the two selected WideJets for the *high-mass* search. Data (points) are compared to predictions from the POWHEG (red) and the PYTHIA 8 (green) generators, after the full detector simulation and normalized to the data. On the bottom panel the ratio of each distribution to the POWHEG prediction is shown.

The choice of this function, with an additional parameter with respect to the one for the *high-mass* analysis, is justified by a Fisher Test (or *F-Test*) [46]. This test allows to probe the hypothesis that one function with $n+1$ parameters does not provide a significantly better fit than another function with n parameters. The hypothesis is rejected if the observed confidence level is less than a desired probability (in this analysis set to 0.05). The details of the test and the results over four different function "families" with different number of parameters (13 functions in total) can be found in Appendix C.

We perform an extended, background-only, binned, maximum likelihood fit to data with the following likelihood:

$$\mathcal{L}(\text{data}|b(\boldsymbol{\theta})) = \prod_{i=1}^{n_b} \text{Poisson}(x_i|b_i(\boldsymbol{\theta})) = \prod_{i=1}^{n_b} \frac{b_i(\boldsymbol{\theta})^{x_i} e^{-b_i(\boldsymbol{\theta})}}{x_i!} \tag{4.8}$$

where n_b is the number of bins, $\boldsymbol{\theta}$ is the vector of nuisance parameters (p_0, p_1, p_2, p_3), x_i is the data yield in bin i, and b_i is the integral of the fit function in bin i multiplied

by the total number of background events N_b:

$$b_i(\boldsymbol{\theta}) = N_b \int_{m_{jj}^{\min,i}}^{m_{jj}^{\max,i}} \mathrm{d}m_{jj}\ p(m_{jj}) \tag{4.9}$$

where $p(m_{jj})$ is the fit function in Equations 4.6 and 4.7 normalized to unity. For the *low-mass* analysis $\boldsymbol{\theta}$ becomes $\boldsymbol{\theta}(p_0, p_1, p_2, p_3, p_4)$.

The fits to data are shown in Figure 4.15, which gives the measured differential cross section times the branching ratio times the acceptance ($\frac{d\sigma \times B \times A}{dm_{jj}}$ [pb/TeV]) as a function of the dijet mass. Variable bins corresponding approximately to the dijet mass resolution are used. The difference between the data and the fit, normalized to the statistical uncertainty of the data in each bin, is also shown on the bottom of Figure 4.15.

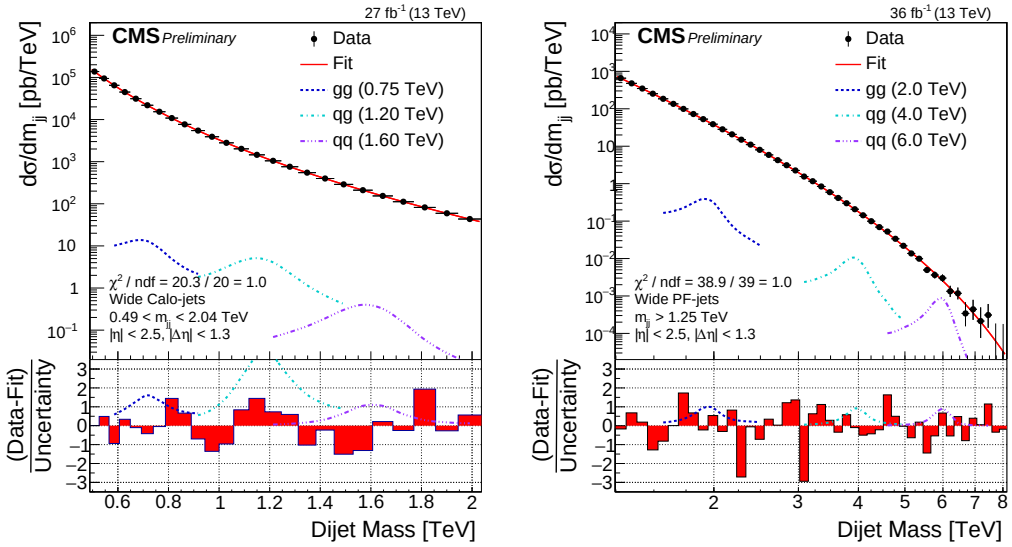

Figure 4.15. Dijet mass spectrum (points) compared to a fitted parameterization of the background (solid curve) for the *low-mass* search (left) and the *high-mass* search (right). The lower panel in each plot shows the difference between the data and the fitted parameterization, divided by the statistical uncertainty of the data. Examples of predicted signals from narrow gluon-gluon, quark-gluon, and quark-quark resonances are shown with cross sections equal to the observed upper limits at 95% CL.

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

To test the goodness-of-fit, we define two test statistics. One is a modified chi-square, defined as:

$$\chi^2 = \sum_{i=1}^{n_b} \left(\frac{x_i - b_i}{\sigma_{x_i}} \right)^2 \tag{4.10}$$

where the "uncertainty" σ_{x_i} is defined in terms of the 68% CL region of a Poisson distribution. Setting $\alpha = 1 - 0.687$,

$$\sigma_{x_i} = \begin{cases} D_c^{-1}(\alpha/2, x_i + 1) & \text{if } b_i > x_i \\ D^{-1}(\alpha/2, x_i) & \text{if } b_i < x_i \end{cases} \tag{4.11}$$

$D^{-1}(\alpha/2, x_i)$ is the quantile function of the gamma distribution, which is the inverse of the cumulative distribution function (lower tail) of the gamma distribution:

$$D(\alpha/2, x_i) = \int_{-\infty}^{\alpha/2} \frac{1}{\Gamma(x_i)} z^{x_i - 1} e^{-z} \, \mathrm{d}z \tag{4.12}$$

$D_c^{-1}(\alpha/2, x_i + 1)$ is the inverse of the cumulative distribution function (upper tail) of the gamma distribution:

$$D_c(\alpha/2, x_i + 1) = \int_{\alpha/2}^{+\infty} \frac{1}{\Gamma(x_i + 1)} z^{x_i} e^{-z} \, \mathrm{d}z \tag{4.13}$$

Similarly, an alternative test statistic can be performed. It is a likelihood ratio with respect to a so-called *saturated model* because contains as many parameters as the data points, defined as:

$$-2 \log \lambda(\boldsymbol{\theta}) = -2 \log \frac{\prod_{i=1}^{n_b} \mathrm{Poisson}(x_i | b_i(\boldsymbol{\theta}))}{\prod_{i=1}^{n_b} \mathrm{Poisson}(x_i | x_i)} = 2 \sum_{i=1}^{n_b} \left[b_i(\boldsymbol{\theta}) - x_i + x_i \log \frac{x_i}{b_i(\boldsymbol{\theta})} \right] \tag{4.14}$$

The two tests are redundant checks of the goodness-of-fit, but the second test is more robust when we have small statistics (like at very *high-mass*). In both cases, we generate pseudodatasets from the best-fit model obtained on data and refit each pseudodataset with a maximum likelihood fit. In Figure 4.16, we show the distribution of both test statistics from these pseudoexperiments as well the value observed in data. The goodness of the results are quantified by the *p-value* defined as the probability of the observed result assuming the background-only hypothesis true. For the first test, the observed p-value is 37% and the effective number of degrees of freedom (from fitting a chi-square distribution) is 37.1 ± 0.1. For the second test the p-value is around 43% and the effective number of degrees of freedom is 42.4 ± 0.1. In Figure 4.17 the same results for the *low-mass* analysis are reported. In this case the observed p-value is around 61% and the effective number of degrees of freedom is 22.8 ±0.1 for both the statistics tests.

Since the fit for the *low-mass* analysis shows a kind of modulation in its residual an additional check for the goodness of fit is performed for this analysis: the Wald-Wolfowitz runs test. This test allows to quantify the probability that a sequence of a two-valued data is compatible with the hypothesis of a random sequence. In this way the randomness of the positive and negative sign sequence of the fit residuals is tested. A bad fit, in fact, could still have a good χ^2 but a long sub-sequences of adjacent positive or negative residuals would indicate that the data are not well described by the fitting function. A "run" of a sequence is defined as the maximal non-empty segment of the sequence consisting of adjacent equal elements. The run test is based on the null hypothesis that each element in the sequence is independently drawn from the same distribution. Under the null hypothesis, the number

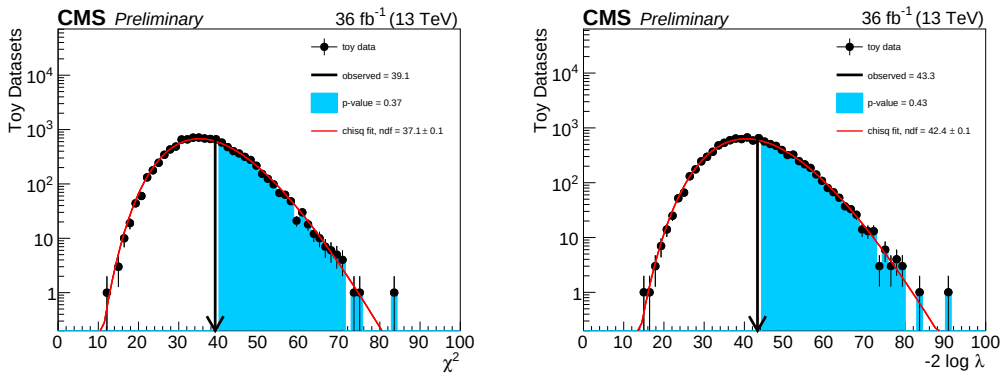

Figure 4.16. Toy distribution for the goodness-of-fit study for the *high-mass* analysis using the χ^2 test statistic (left) and the $-2\log\lambda$ test statistic (right) as defined in the text.

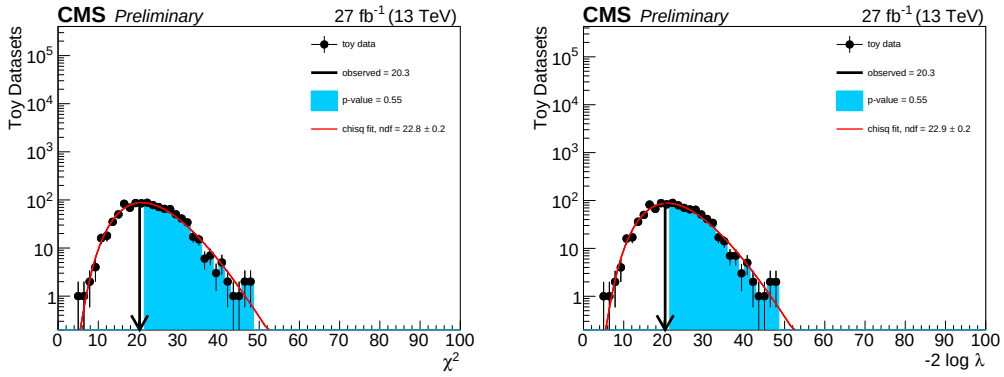

Figure 4.17. Toy distribution for the goodness-of-fit study for the *low-mass* analysis using the χ^2 test statistic (left) and the $-2\log\lambda$ test statistic (right) as defined in the text.

of runs, N_{runs}, in a sequence of N elements is a random variable whose conditional distribution is approximately normal, with mean μ and variance σ given by:

$$\mu = \frac{2N_+N_-}{N} + 1 \tag{4.15a}$$

and

$$\sigma^2 = \frac{2N_+N_-(2N_+N_- - N)}{N^2(N-1)} = \frac{(\mu-1)(\mu-2)}{N-1} \tag{4.15b}$$

where N_+ and N_- are the positive and negative values observed respectively ($N_+ + N_- = N$). Therefore the probability associated with the observed value of N_{runs} is given by integrating this Gaussian distribution:

$$P(N_{runs}) = \int_{-\inf}^{N_{runs}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4.16}$$

For uncorrelated residuals the number of runs should not stay in the tails of the Gaussian. In fact if the number of runs is significantly higher or lower than

expected, the hypothesis of statistical independence of the elements may be rejected. Furthermore the test does not assume that the positive and negative elements have equal probabilities of occurring, but only assumes that the elements are independent and identically distributed.

In our case N_{runs} indicates how many times the residuals change sign, N_+ is the number of positive residuals and N_- the number of the negatives. The background fit gives the values $N_- = 13$ and $N_+ = 12$. Therefore, if the null hypothesis is true, the number of runs is a random variable extracted from the Gaussian distribution with $\mu = 13.48$ and $\sigma = 2.44$. The observed number of runs is $N_{runs} = 13$ and then the probability to obtain this value is $P(13) = 42\%$ which demonstrates that the sequence of the residuals for the *low-mass* analysis is well compatible with the non-correlation hypothesis.

From these tests we conclude that the fit function for background-only hypothesis describes well the data and we do not observe any significant excess in the dijet mass spectrum.

Chapter 5

Results and Interpretation

Since no excess of events has been observed in the dijet mass spectra, exclusion limits on the production cross section of new particles decaying to pairs of quarks and gluons are set. The results are interpreted in the context of the BSM theories presented in Section 1.2 and discussed in terms of:

- Resonances decaying into two partons;
- Dark Matter mediator models.

Both cases of narrow and wide resonances are considered in the interpretation.

5.1 Narrow resonances decaying into two partons

The background parameterization and the signal models are used to set exclusion limits on the production cross section of new particles. A separate limit is determined for each final state (qq, qg, and gg) because of the dependence of the dijet resonance shape on the two final state partons shown in Figures 4.4.

The modified frequentist method (or CL_s method) [47, 48], in conjunction with the use of asymptotic formulas [49], is utilized to set upper limits on the signal cross sections. We use a multi-bin counting experiment likelihood, which is a product of Poisson distributions corresponding to different bins defined as:

$$\mathcal{L}(\text{data}|\mu s + b) = \prod_{i=1}^{n_b} \text{Poisson}(x_i|\mu s_i + b_i) = \prod_{i=1}^{n_b} \frac{(\mu s_i + b_i)^{x_i} e^{-(\mu s_i + b_i)}}{x_i!} \tag{5.1}$$

where μ is the signal strength, n_b is the number of bins, x_i is the data yield in bin i, and s_i and b_i are the integrated signal and background rates in bin i. This likelihood is used to define the test statistic following the LHC CLs procedure [50]:

$$q_\mu = -2 \log \frac{\mathcal{L}(\text{data}|\mu s + b)}{\mathcal{L}(\text{data}|\hat{\mu} s + b)}, \; 0 \leq \hat{\mu} \leq \mu \tag{5.2}$$

where $\hat{\mu}$ maximizes the likelihood $\mathcal{L}(\text{data}|\mu s + b)$. We can therefore derive the observed 95% CL upper limit on the signal strength by computing the value of μ that satisfies the condition:

$$CL_s \equiv \frac{CL_{s+b}}{CL_b} = \frac{1-\Phi(\sqrt{q_\mu})}{\Phi(\sqrt{q_{\mu,A}}-\sqrt{q_\mu})} = 0.05 \tag{5.3}$$

where $q_{\mu,A}$ is the test statistic evaluated on a dataset corresponding exactly to the expected background and $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. Similar expressions can be used to derive the expected 95% CL upper limit:

$$\frac{1-\Phi(\sqrt{q_{\mu,A}})}{0.5} = 0.05 \tag{5.4}$$

and to find the $\pm N\sigma$ uncertainty band around the expected limit:

$$\frac{1-\Phi(\sqrt{q_{\mu,A}} \mp N)}{\Phi(\pm N)} = 0.05 \tag{5.5}$$

Systematic uncertainties are related to nuisance parameters that are incorporated into the model. The effect of the systematic uncertainties on the limits is studied in detail in Section 5.1.1.

For narrow resonances, we evaluate the likelihood independently at each value of resonance pole mass from 0.6 to 1.6 TeV in 50 GeV steps in the *low-mass* search, and from 1.6 to 8.1 TeV in 100 GeV steps in the *high-mass* search.

The resonance mass boundary of 1.6 TeV between the two analyses is chosen such that the dijet mass requirement $m_{jj} > 1246$ GeV is reasonably efficient for all the considered final states. For a resonance with a mass of 1.6 TeV, the acceptance of this requirement is:

- 57% for a gluon-gluon resonance;
- 76% for a quark-gluon resonance;
- 85% for a quark-quark resonance.

Figure 5.1 shows the 95% CL asymptotic LHC CL_s limits on the cross section times the branching ratio times the acceptance ($\sigma \times BR \times A$) for a gg resonance, a qg resonance, and a qq resonance for the *high-mass* and *low-mass* analyses. The bottom right figure shows the limits for the RS Graviton model, which foresees the decay of the resonance to both qq and gg final states and for which the cross section upper limits is obtained from the branching fraction weighted average of the limits on quark-quark and gluon-gluon resonances. The observed and the expected limits are reported (points and dashed line) as well as the $\pm 1\sigma$ (green) and $\pm 2\sigma$ (yellow) bands corresponding to the variation of the expected limits by one or two standard deviations respectively.

The acceptance of the minimum dijet mass requirement in each search ($m_{jj} > 489$ GeV for the *low-mass* analysis, $m_{jj} > 1246$ GeV for the *high-mass* analysis) is evaluated separately for qq, qg, and gg resonances. It is taken into account by correcting the limits, and therefore it does not appear in the acceptance A.

Table 5.1 and Figure 5.2 summarize the limits for the different final states.

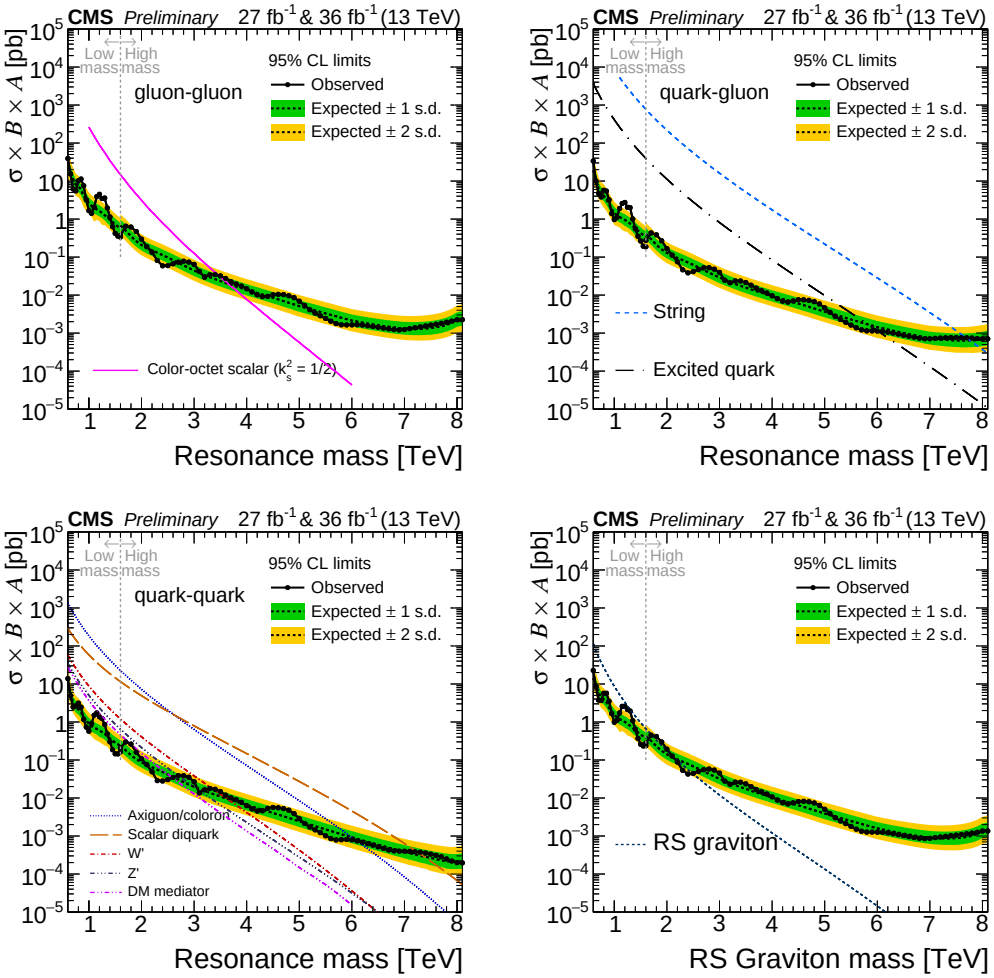

Figure 5.1. 95% CL asymptotic LHC CL_s limits on $\sigma \times BR \times A$ as a function of the resonance mass for a gg resonance (top left), a qg resonance (top right), and a qq resonance (bottom left). Bottom right plot shows the gg and qq resonance limits weighted for the branching ratios of a Randall-Sundrum Graviton. The results from both the *low-mass* and the *high-mass* analyses are reported.

Mass [TeV]	95% CL upper limit [pb]							
	gg		qg		qq		qq-gg for RSG	
	Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
Low-Mass search								
0.50	2.81e+01	3.56e+01	3.33e+01	3.88e+01	4.65e+01	4.51e+01	4.02e+01	4.19e+01
0.55	4.09e+01	3.17e+01	3.73e+01	2.72e+01	1.74e+01	1.34e+01	2.55e+01	1.97e+01
0.60	3.93e+01	2.10e+01	3.37e+01	1.90e+01	1.38e+01	1.05e+01	2.27e+01	1.41e+01
0.65	1.55e+01	1.77e+01	1.01e+01	1.14e+01	4.92e+00	5.15e+00	8.62e+00	9.57e+00
0.70	6.14e+00	1.12e+01	4.73e+00	6.32e+00	2.47e+00	3.16e+00	3.77e+00	6.01e+00
0.75	5.50e+00	8.13e+00	3.82e+00	4.68e+00	2.64e+00	2.49e+00	3.66e+00	4.52e+00
0.80	1.02e+01	7.15e+00	5.73e+00	4.06e+00	3.14e+00	2.14e+00	5.70e+00	3.96e+00
0.85	1.13e+01	5.93e+00	5.45e+00	3.33e+00	2.46e+00	1.79e+00	5.71e+00	3.31e+00
0.90	7.56e+00	4.04e+00	3.21e+00	2.42e+00	1.17e+00	1.36e+00	3.53e+00	2.35e+00
0.95	3.23e+00	3.32e+00	1.40e+00	1.86e+00	7.30e-01	1.10e+00	1.66e+00	1.93e+00
1.00	1.66e+00	2.60e+00	9.67e-01	1.45e+00	5.72e-01	9.06e-01	9.83e-01	1.55e+00
1.05	1.41e+00	2.22e+00	1.11e+00	1.26e+00	9.11e-01	7.90e-01	1.10e+00	1.33e+00
1.10	2.06e+00	1.96e+00	1.90e+00	1.13e+00	1.51e+00	7.11e-01	1.72e+00	1.19e+00
1.15	3.90e+00	1.79e+00	2.58e+00	1.04e+00	1.74e+00	6.55e-01	2.58e+00	1.10e+00
1.20	4.49e+00	1.63e+00	2.74e+00	9.49e-01	1.38e+00	6.00e-01	2.60e+00	1.01e+00
1.25	3.48e+00	1.45e+00	2.04e+00	8.64e-01	1.23e+00	5.40e-01	2.12e+00	9.00e-01
1.30	3.58e+00	1.26e+00	2.00e+00	7.60e-01	8.61e-01	4.85e-01	1.95e+00	7.97e-01
1.35	1.96e+00	1.11e+00	1.01e+00	6.62e-01	4.85e-01	4.24e-01	1.08e+00	7.03e-01
1.40	1.14e+00	9.55e-01	5.56e-01	5.71e-01	3.00e-01	3.69e-01	6.44e-01	6.09e-01
1.45	6.32e-01	8.33e-01	3.52e-01	4.97e-01	1.86e-01	3.27e-01	3.70e-01	5.35e-01
1.50	4.20e-01	7.23e-01	2.66e-01	4.30e-01	1.45e-01	2.84e-01	2.59e-01	4.67e-01
1.55	3.57e-01	6.38e-01	1.93e-01	3.81e-01	1.44e-01	2.59e-01	2.34e-01	4.18e-01
1.60	3.37e-01	5.58e-01	1.87e-01	3.45e-01	1.64e-01	2.35e-01	2.37e-01	3.72e-01
High-Mass search								
1.6	3.72e-01	6.72e-01	2.74e-01	4.08e-01	2.07e-01	2.38e-01	2.77e-01	4.22e-01
1.7	6.50e-01	5.02e-01	4.33e-01	2.96e-01	2.99e-01	1.79e-01	4.50e-01	3.18e-01
1.8	6.17e-01	3.55e-01	3.86e-01	2.10e-01	2.62e-01	1.34e-01	4.17e-01	2.30e-01
1.9	4.71e-01	2.63e-01	2.69e-01	1.60e-01	1.61e-01	1.06e-01	2.98e-01	1.76e-01
2.0	2.97e-01	2.07e-01	1.67e-01	1.29e-01	1.08e-01	8.71e-02	1.93e-01	1.41e-01
2.1	1.88e-01	1.74e-01	1.12e-01	1.10e-01	7.56e-02	7.44e-02	1.27e-01	1.20e-01
2.2	1.34e-01	1.50e-01	7.53e-02	9.49e-02	4.90e-02	6.43e-02	8.82e-02	1.04e-01
2.3	8.15e-02	1.30e-01	4.62e-02	8.32e-02	2.86e-02	5.57e-02	5.33e-02	9.04e-02
2.4	5.89e-02	1.13e-01	3.84e-02	7.21e-02	2.80e-02	4.82e-02	4.26e-02	7.88e-02
2.5	5.96e-02	9.73e-02	4.15e-02	6.23e-02	3.05e-02	4.16e-02	4.44e-02	6.82e-02
2.6	6.67e-02	8.32e-02	4.71e-02	5.33e-02	3.47e-02	3.58e-02	5.01e-02	5.87e-02
2.7	7.32e-02	7.09e-02	5.22e-02	4.55e-02	3.88e-02	3.08e-02	5.56e-02	5.03e-02
2.8	7.79e-02	6.04e-02	5.26e-02	3.91e-02	3.87e-02	2.63e-02	5.79e-02	4.30e-02
2.9	7.37e-02	5.18e-02	4.82e-02	3.35e-02	3.53e-02	2.28e-02	5.43e-02	3.71e-02
3.0	6.42e-02	4.43e-02	3.96e-02	2.90e-02	2.68e-02	1.96e-02	4.54e-02	3.19e-02
3.1	4.20e-02	3.86e-02	2.46e-02	2.53e-02	1.36e-02	1.74e-02	2.79e-02	2.80e-02
3.2	2.95e-02	3.37e-02	2.11e-02	2.24e-02	1.64e-02	1.54e-02	2.30e-02	2.46e-02
3.3	3.41e-02	2.96e-02	2.36e-02	1.96e-02	1.78e-02	1.36e-02	2.61e-02	2.17e-02
3.4	3.47e-02	2.63e-02	2.34e-02	1.75e-02	1.69e-02	1.22e-02	2.60e-02	1.94e-02
3.5	3.19e-02	2.33e-02	2.14e-02	1.58e-02	1.48e-02	1.10e-02	2.36e-02	1.73e-02
3.6	2.74e-02	2.08e-02	1.82e-02	1.41e-02	1.19e-02	9.81e-03	1.99e-02	1.55e-02
3.7	2.25e-02	1.87e-02	1.52e-02	1.27e-02	1.01e-02	8.86e-03	1.66e-02	1.40e-02
3.8	1.96e-02	1.68e-02	1.31e-02	1.16e-02	9.02e-03	8.03e-03	1.45e-02	1.26e-02
3.9	1.72e-02	1.53e-02	1.13e-02	1.05e-02	7.72e-03	7.25e-03	1.27e-02	1.15e-02
4.0	1.47e-02	1.37e-02	9.57e-03	9.45e-03	6.29e-03	6.57e-03	1.07e-02	1.03e-02
4.1	1.21e-02	1.25e-02	8.06e-03	8.67e-03	5.17e-03	5.98e-03	8.86e-03	9.41e-03
4.2	1.02e-02	1.14e-02	6.93e-03	7.89e-03	4.52e-03	5.40e-03	7.54e-03	8.57e-03
4.3	9.12e-03	1.03e-02	6.55e-03	7.20e-03	4.61e-03	4.91e-03	7.01e-03	7.78e-03
4.4	9.27e-03	9.35e-03	7.01e-03	6.57e-03	5.35e-03	4.46e-03	7.45e-03	7.07e-03
4.5	1.02e-02	8.47e-03	7.52e-03	5.98e-03	5.65e-03	4.04e-03	8.09e-03	6.42e-03
4.6	1.05e-02	7.69e-03	7.51e-03	5.44e-03	5.55e-03	3.65e-03	8.19e-03	5.83e-03
4.7	1.03e-02	6.96e-03	7.27e-03	4.91e-03	5.26e-03	3.31e-03	7.97e-03	5.28e-03
4.8	9.62e-03	6.27e-03	6.72e-03	4.46e-03	4.79e-03	2.99e-03	7.40e-03	4.77e-03
4.9	8.56e-03	5.69e-03	5.86e-03	4.04e-03	3.88e-03	2.70e-03	6.42e-03	4.32e-03
5.0	6.90e-03	5.10e-03	4.62e-03	3.67e-03	2.85e-03	2.44e-03	5.05e-03	3.89e-03
5.1	5.34e-03	4.70e-03	3.53e-03	3.33e-03	2.14e-03	2.22e-03	3.89e-03	3.57e-03
5.2	4.11e-03	4.28e-03	2.77e-03	3.04e-03	1.73e-03	2.01e-03	3.03e-03	3.25e-03
5.3	3.35e-03	3.94e-03	2.28e-03	2.77e-03	1.45e-03	1.81e-03	2.49e-03	2.98e-03
5.4	2.85e-03	3.60e-03	1.92e-03	2.50e-03	1.22e-03	1.64e-03	2.12e-03	2.72e-03
5.5	2.43e-03	3.28e-03	1.62e-03	2.29e-03	1.01e-03	1.50e-03	1.79e-03	2.48e-03
5.6	2.05e-03	3.02e-03	1.38e-03	2.08e-03	8.54e-04	1.36e-03	1.51e-03	2.27e-03
5.7	1.78e-03	2.77e-03	1.22e-03	1.90e-03	7.88e-04	1.23e-03	1.33e-03	2.08e-03
5.8	1.65e-03	2.53e-03	1.15e-03	1.73e-03	8.00e-04	1.11e-03	1.27e-03	1.89e-03
5.9	1.64e-03	2.33e-03	1.14e-03	1.58e-03	8.09e-04	1.02e-03	1.27e-03	1.74e-03
6.0	1.64e-03	2.13e-03	1.13e-03	1.43e-03	7.91e-04	9.25e-04	1.26e-03	1.59e-03
6.1	1.66e-03	2.01e-03	1.11e-03	1.34e-03	7.45e-04	8.39e-04	1.25e-03	1.49e-03
6.2	1.63e-03	1.89e-03	1.06e-03	1.24e-03	6.84e-04	7.66e-04	1.21e-03	1.39e-03
6.3	1.56e-03	1.79e-03	1.00e-03	1.16e-03	6.26e-04	6.99e-04	1.15e-03	1.30e-03
6.4	1.49e-03	1.69e-03	9.41e-04	1.08e-03	5.75e-04	6.44e-04	1.08e-03	1.23e-03
6.5	1.42e-03	1.61e-03	8.82e-04	1.00e-03	5.21e-04	5.89e-04	1.02e-03	1.16e-03
6.6	1.36e-03	1.53e-03	8.26e-04	9.37e-04	4.72e-04	5.40e-04	9.66e-04	1.09e-03
6.7	1.29e-03	1.47e-03	7.79e-04	8.82e-04	4.30e-04	4.91e-04	9.13e-04	1.04e-03
6.8	1.24e-03	1.41e-03	7.35e-04	8.27e-04	4.06e-04	4.55e-04	8.71e-04	9.88e-04
6.9	1.21e-03	1.36e-03	7.11e-04	7.78e-04	4.00e-04	4.18e-04	8.54e-04	9.45e-04
7.0	1.24e-03	1.32e-03	7.08e-04	7.29e-04	3.98e-04	3.81e-04	8.69e-04	9.09e-04
7.1	1.31e-03	1.30e-03	7.27e-04	7.05e-04	3.94e-04	3.57e-04	9.06e-04	8.85e-04
7.2	1.38e-03	1.30e-03	7.40e-04	6.81e-04	3.86e-04	3.27e-04	9.43e-04	8.72e-04
7.3	1.46e-03	1.30e-03	7.53e-04	6.62e-04	3.74e-04	3.02e-04	9.81e-04	8.62e-04
7.4	1.54e-03	1.32e-03	7.61e-04	6.50e-04	3.57e-04	2.84e-04	1.02e-03	8.68e-04
7.5	1.62e-03	1.36e-03	7.62e-04	6.38e-04	3.33e-04	2.66e-04	1.06e-03	8.80e-04
7.6	1.71e-03	1.42e-03	7.59e-04	6.38e-04	3.10e-04	2.47e-04	1.10e-03	9.07e-04
7.7	1.81e-03	1.51e-03	7.53e-04	6.38e-04	2.84e-04	2.29e-04	1.14e-03	9.47e-04
7.8	1.93e-03	1.65e-03	7.43e-04	6.44e-04	2.50e-04	2.17e-04	1.19e-03	1.02e-03
7.9	2.06e-03	1.87e-03	7.19e-04	6.56e-04	2.20e-04	2.11e-04	1.25e-03	1.15e-03
8.0	2.25e-03	2.24e-03	7.03e-04	6.93e-04	1.99e-04	2.11e-04	1.35e-03	1.35e-03
8.1	2.26e-03	2.41e-03	7.05e-04	7.35e-04	1.97e-04	2.17e-04	1.36e-03	1.45e-03

Table 5.1. Observed and expected upper limits at 95% CL on $\sigma \times BR \times A$ for a gg resonance, a qg resonance and a qq resonance as a function of the resonance mass. For the RS Graviton the limits are obtained from the branching fraction weighted average of the limits on qq and gg resonances.

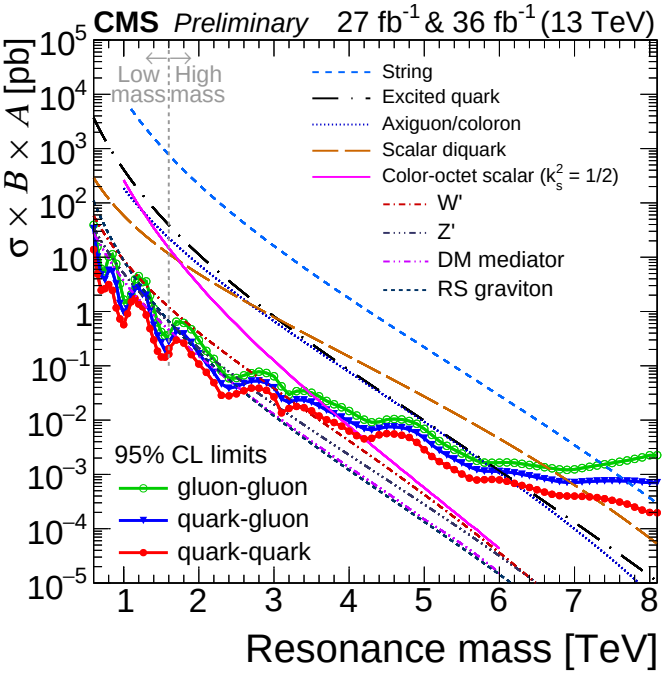

Figure 5.2. Summary of 95% CL asymptotic LHC CL_s limits on $\sigma \times BR \times A$ for a gg resonance, a qg resonance, and a qq resonance. The results from both the *low-mass* and the *high-mass* analyses are reported.

The strength of a possible excess of events in the dijet mass spectrum can be quantified by the *significance*. Technically, the measure of the local significance is obtained imposing $\mu = 0$ in Equation 5.2:

$$q_0 = -2 \log \frac{\mathcal{L}(\text{data}|b)}{\mathcal{L}(\text{data}|\hat{\mu}s + b)} \tag{5.6}$$

and therefore the observed local significance, Z, is simply:

$$Z = \sqrt{q_0} \tag{5.7}$$

The significances shown in this section are "local" since the look-elsewhere effect (LEE) [51] is not taken into account. The LEE relies on the fact that, when searching for a new resonance somewhere in a possible mass range, the significance of observing a local excess of events should take into account the probability of observing such an excess anywhere in the range. The look-elsewhere effect would lower the already low local significance observed.

The observed local significance as a function of the resonance mass for the three considered final states (qq, qg, and gg) is shown in Figure 5.3 for the *high-mass* analysis and in Figure 5.4 for the *low-mass* analysis. The most significant local excess (1.9σ) occurs for the gg resonance hypothesis at mass 850 GeV.

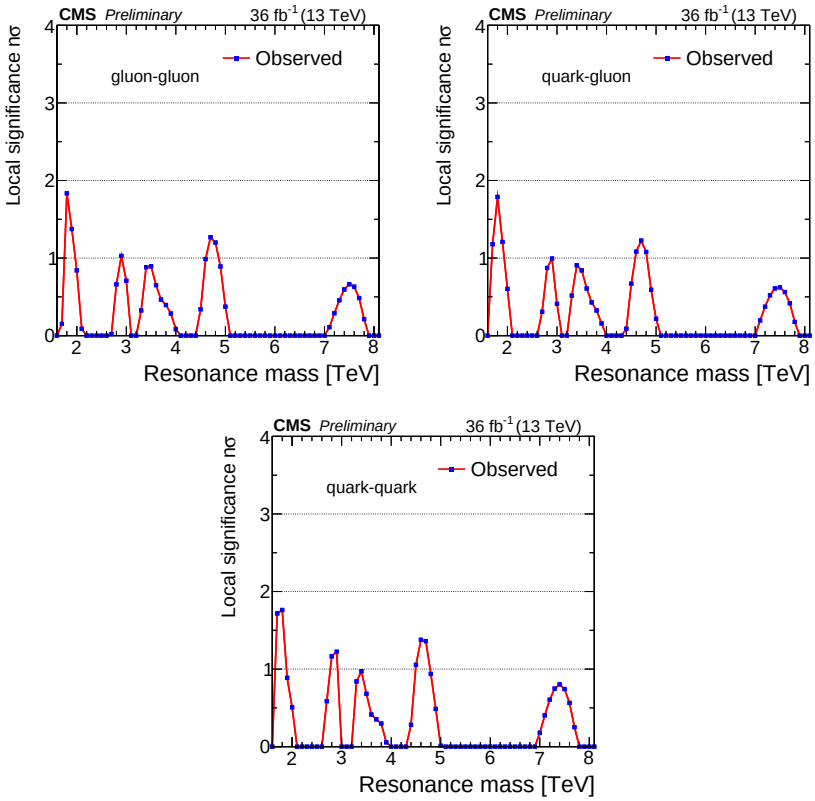

Figure 5.3. Local significance as a function of the resonance mass for a gg resonance (top left), a qg resonance (top right), and a qq resonance (bottom) observed in the *high-mass* analysis.

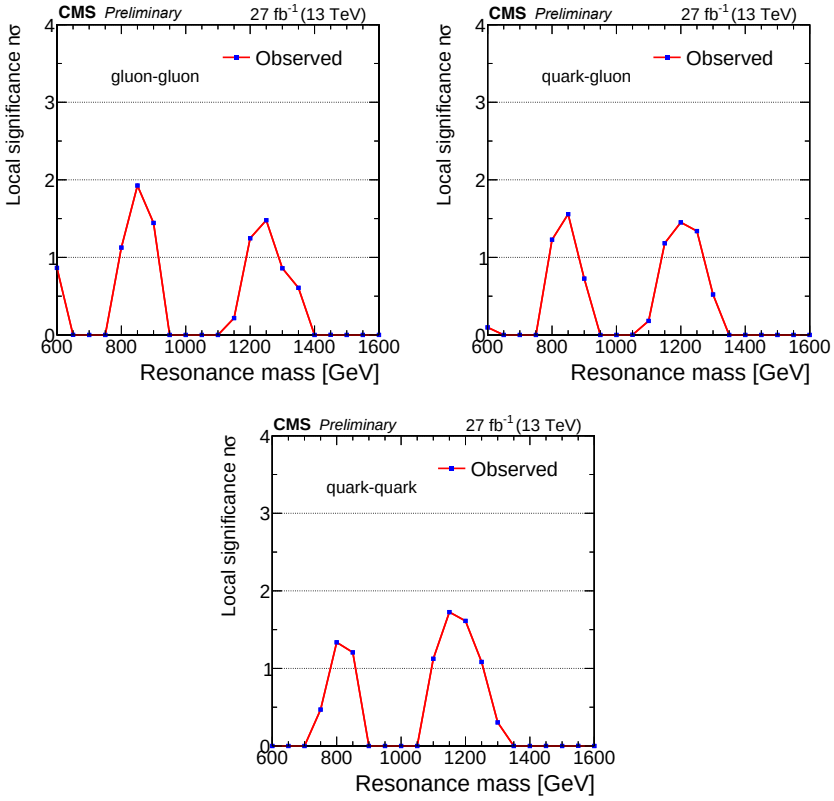

Figure 5.4. Local significance as a function of the resonance mass for a gg resonance (top left), a qg resonance (top right), and a qq resonance (bottom) observed in the *low-mass* analysis.

5.1.1 Systematic uncertainties

The systematic uncertainties are taken into account as nuisance parameters in the limit setting procedure. The dominant sources of systematic uncertainty are the jet energy scale and resolution, integrated luminosity, and the exact parameterization of the background shape in the dijet mass distribution. The uncertainty in the jet energy scale in both the *low-mass* and the *high-mass* search is 2% and it is determined from $\sqrt{s} = 13$ TeV data using the methods described in [52]. This uncertainty is propagated to the limits by shifting the dijet mass shape for signal by 2%. The uncertainty in the jet energy resolution translates into an uncertainty of 10% in the resolution of the dijet mass [40], and it is propagated to the limits by observing the effect of increasing and decreasing by 10% the reconstructed width of the dijet mass shape for signal. The uncertainty in the integrated luminosity is 2.5% [53], and it is propagated to the normalization of the signal. Changes in the values of the parameters describing the background introduce a change in the signal strength, which is accounted for as a systematic uncertainty. It is automatically evaluated via profiling, effectively refitting for the optimal values of the background parameters for each value of the resonance cross section. The extent to which the background uncertainty affects the limit depends significantly on the amount of the background. Its effect is larger when the background is larger, which happens for wide resonances and low resonance masses. The largest effect occurs for the qg resonances because they are wider, while the smallest effect is for the qq resonances. The systematic uncertainties are summarized in Table 5.2.

The effect of the systematics on the exclusion limit is reported for qq resonances in Figure 5.5 in which the ratio of the limits with and without systematics applied is shown. The systematic uncertainties reflect in a weakening of the upper limit on μ by a factor of 2-3 for the *low-mass* analysis and a factor around 1.5 for the *high-mass* analysis.

Systematic Uncertainty source	Uncertainty
Jet energy resolution	10% of resolution
Jet energy scale	$\pm$ 2% shift of m_{jj}
Luminosity	$\pm$ 2.5% on the normalization
Background parameterization	depends on the signal shape and the resonance mass

Table 5.2. Summary of the systematic uncertainties propagated to the limit setting procedure.

5.1.2 Theoretical interpretation

The exclusion limits on the production cross-section for new resonances, presented in the previous sections, can be compared to some theoretical predictions to set limits on the mass of the resonance. The considered theoretical models are discussed in Section 1.2. The model predictions shown in Figures 5.1 and 5.2 are calculated in the narrow-width approximation using PYTHIA 8 and the CTEQ6L1 [54] PDF at leading order, with a next-to-leading order correction factor. The branching frac-

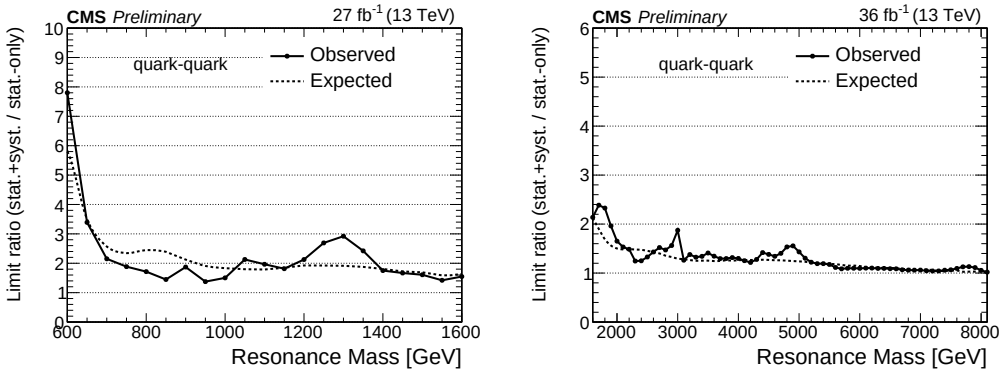

Figure 5.5. The ratio between the observed limits on the cross section including statistical+systematic uncertainties and the limits including statistical uncertainties only (points) for a resonance decaying to quark-quark in the *low-mass* search (left) and the *high-mass* search (right). The ratio between the expected limits is also reported (dashed line).

tion includes the direct decays of the resonance into the five light quarks and gluons only, excluding the top quark from the decay, although top quarks are included in the calculation of the resonance width. The acceptance is evaluated at the parton level for the resonance decay to two partons. In the case of isotropic decays, the acceptance is $A \sim 0.6$ and it is independent of the resonance mass. For a given model, new particles are excluded at 95% CL in the mass region where the theoretical prediction lies at or above the observed upper limit for the appropriate final state in Figures 5.1 and 5.2. The exclusion limits for the nine considered benchmark models are summarized in Table 5.3. They are more stringent than the exclusion limits in the dijet channel previously reported by CMS [44, 45, 55].

		Observed (expected) exclusion limits on the resonance mass [TeV]			
Model	Final State	36 fb^{-1} 13 TeV	12.9 fb^{-1} 13 TeV	2.4 fb^{-1} 13 TeV	20 fb^{-1} 8 TeV
String	qg	7.7 (7.7)	7.4 (7.4)	7.0 (6.9)	5.0 (4.9)
Scalar diquark	qq	7.2 (7.4)	6.9 (6.8)	6.0 (6.1)	4.7 (4.4)
Axigluon/coloron	$q\bar{q}$	6.1 (6.0)	5.5 (5.6)	5.1 (5.1)	3.7 (3.9)
Excited quark	qg	6.0 (5.8)	5.4 (5.4)	5.0 (4.8)	3.5 (3.7)
Color-octet scalar ($k_s^2 = 1/2$)	gg	3.4 (3.6)	3.0 (3.3)	—	—
W'	$q\bar{q}$	3.3 (3.6)	2.7 (3.1)	2.6 (2.3)	2.2 (2.2)
Z'	$q\bar{q}$	2.7 (2.9)	2.1 (2.3)	—	1.7 (1.8)
RS Graviton	$q\bar{q}$, gg	1.7* (2.1)	1.9 (1.8)	—	1.6 (1.3)
DM Mediator ($m_{\mathrm{DM}} = 1$ GeV)	$q\bar{q}$	2.6 (2.5)	2.0 (2.0)	—	—

Table 5.3. Observed and expected exclusion limits on the resonance mass at 95% CL obtained in this analysis and compared to previous results from CMS with 12.9 fb^{-1} and 2.4 fb^{-1} at $\sqrt{s} = 13$ TeV [44, 55] and with 20 fb^{-1} at $\sqrt{s} = 8$ TeV [45]. For the latest analysis the listed models are excluded between 0.6 TeV and the indicated exclusion limit on the resonance mass. *For the RS Graviton model, in addition to the observed exclusion limit listed above, this analysis also excludes the mass interval between 2.1 and 2.5 TeV.

5.1.3 Bias Study

One of the most important questions to answer is if the chosen signal+background fit would be able to measure the signal correctly. This can be checked performing a "bias test": we inject a signal and compare the outcome of the fit with the injected signal, i.e. measure the "bias" on the signal strength.

In order to perform this bias test, pseudodatasets are generated from different functions and then they are fitted with the functions chosen to describe the background (Equations 4.6 and 4.7). The bias test result depends on the choice of the background model used for generation and fitting of the toy model, which is not obvious and unique because we don't know the "true" background distribution in data. A simple bias study can be done using the same background parameterization to generate and fit the pseudodataset, and it is used as a "closure test". However the closure bias study 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 using different functions and fitting with the functions used in the analysis. The alternative functions for generation are reported in Equations 5.8 and 5.9 for the *high-mass* and *low-mass* search respectively.

$$\frac{\mathrm{d}\sigma}{\mathrm{d}m_{jj}} = p_0 \exp[p_1 x^{p_2} + p_1(1-x)^{p_3}] \text{ (4-par. modified exponential)} \tag{5.8}$$

$$\frac{\mathrm{d}\sigma}{\mathrm{d}m_{jj}} = p_0 \exp[p_1 x^{p_2} + p_3(1-x)^{p_4}] \text{ (5-par. modified exponential)} \tag{5.9}$$

We define the measure of the bias as the difference between the fitted signal strength $\hat{\mu}$ and the injected signal strength μ divided by the statistical+systematic uncertainty σ_μ:

$$\text{Bias} = \frac{\hat{\mu} - \mu}{\sigma_\mu} \tag{5.10}$$

In the following, all of the signal templates are normalized so that $\mu = 1$ corresponds to $\sigma \times B \times A = 1$ pb. We perform the bias study using the qq resonance signal shapes, with injected signal strength μ corresponding to the 95% CL limit obtained previously in Section 5.1. For example, Figure 5.6 shows the distributions of the measured bias for a qq resonance with mass 1350 GeV (*low-mass* analysis): pseudodatasets are generated either with the 5 parameters dijet function (Eq. 4.7) and with the 5 parameters modified exponential function (Eq. 5.9) and then they are fitted with the 5 parameters dijet function. From a Gaussian fit is possible therefore to extract the bias at that resonance mass.

This study is performed covering the mass range considered in this analysis: from 600 GeV to 1600 GeV by 50 GeV steps for the *low-mass* analysis and from 1600 GeV to 8100 GeV for the *high-mass* search with steps of 100 GeV. Figure 5.7 shows the bias as a function of the resonance mass (qq final state) for the *high-mass* analysis. In this case the bias observed is always lower than 20%, that is negligible compared to other systematic uncertainties.

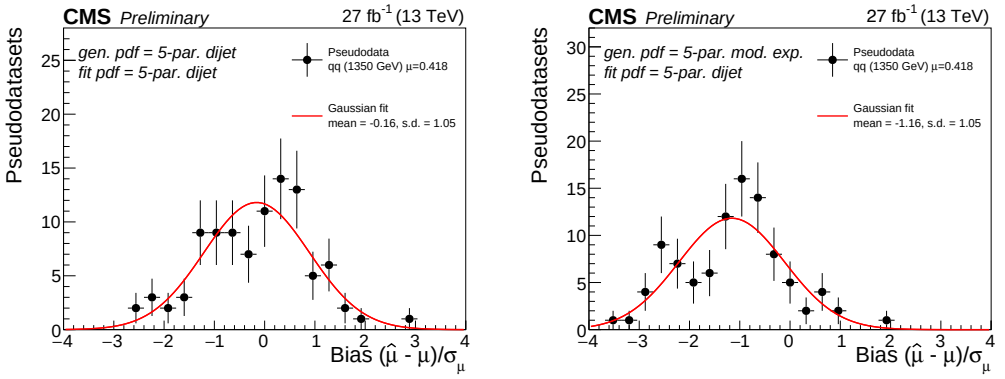

Figure 5.6. Bias on the signal strength for a qq resonance with mass 1350 GeV and $\sigma \times BR \times A = 0.418$ pb. The pseudodatasets are generated with the 5 parameters dijet function (left) or with the 5 parameters modified exponential function (right) and fitted with the 5 parameters dijet function.

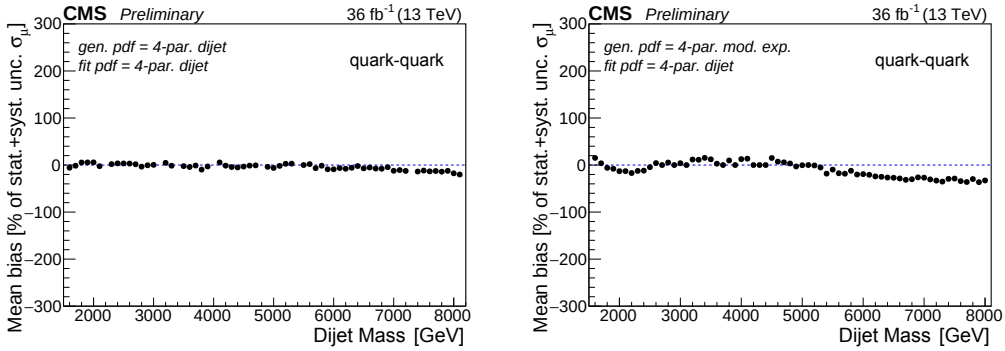

Figure 5.7. Bias on the signal strength in percentage of σ_μ as a function of the qq resonance mass. The pseudodatasets are generated with the 4 parameters dijet function (left) or with the 4 parameters modified exponential function (right) and fitted with the 4 parameters dijet function.

In Figure 5.8 the bias as a function of the qq resonance mass for the *low-mass* analysis is shown. In this case we observe up to 150% bias as a percentage of σ_μ.

In order to quantify the effect of this high bias on the limits, an additional study known as "discrete profiling" or "envelope" method [56] is performed. This method investigates the effect on the 95% CL limits when profiling a discrete nuisance parameter, corresponding to the choice of background parameterization. In Figure 5.9 the ratio of the limits with discrete profiling to the nominal limits is reported and it can be noticed that adding discrete profiling has small impact. The expected limit with discrete profiling is always within $\pm 1\sigma$ of the nominal expected limit without discrete profiling.

A further study is based on another definition of bias:

$$\text{Bias}_2 = (\hat{\mu} - \mu)/\mu \tag{5.11}$$

in order to measure the effect of the bias in terms of the injected signal strength μ. In this case, the observed bias, reported in Figure 5.10, is up to 50% and it is

small compared to the effect of the already existing systematic uncertainties. In fact, such uncertainties weaken the upper limit on μ by a factor of 2-3, as shown in Figure 5.5.

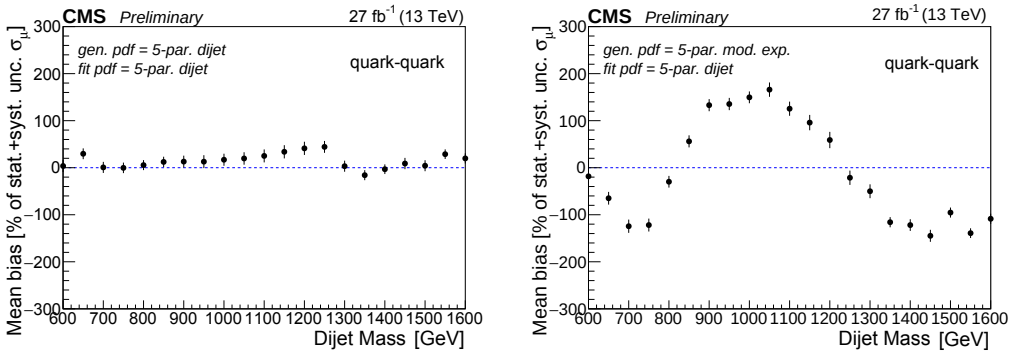

Figure 5.8. Bias on the signal strength in percentage of σ_μ as a function of the qq resonance mass. The pseudodatasets are generated with the 5 parameters dijet function (left) or with the 5 parameters modified exponential function (right) and fitted with the 5 parameters dijet function.

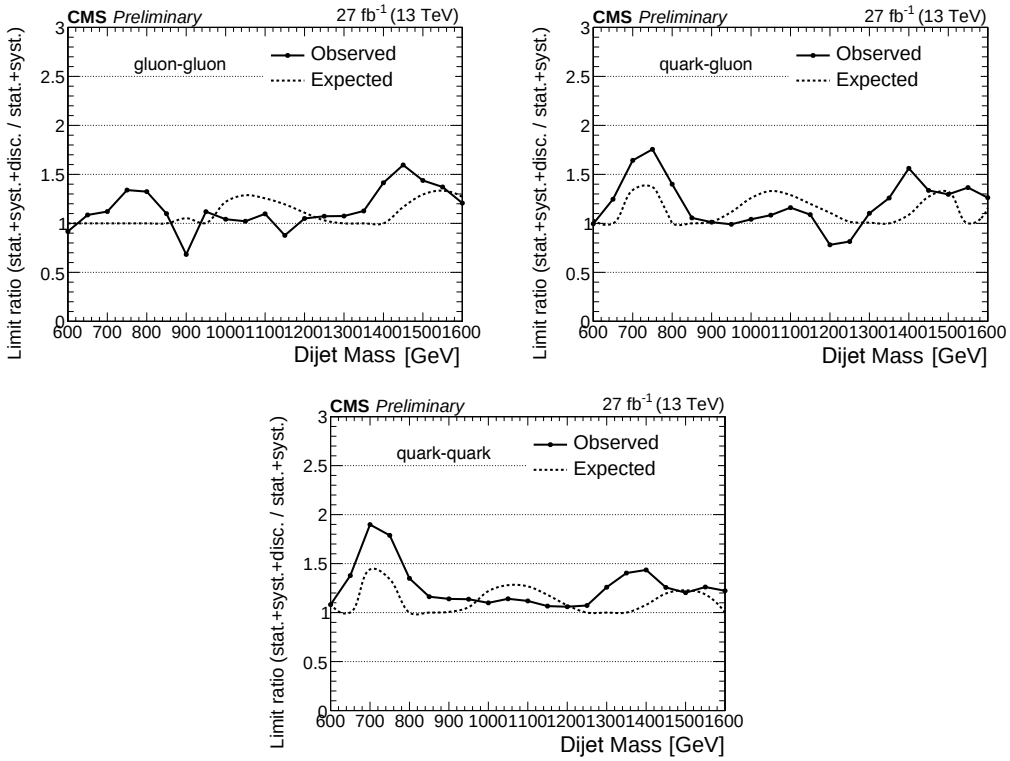

Figure 5.9. The ratio between the limits with discrete profiling and the nominal limits for a gg resonance (top left), a qg resonance (top right), and a qq resonance (bottom).

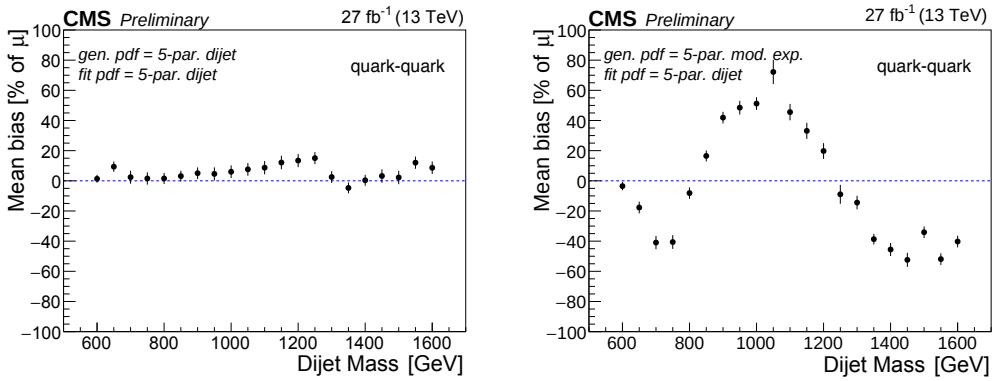

Figure 5.10. Bias on the signal strength in percentage of μ as a function of the qq resonance mass when fitting with the 5 parameters dijet function and generating with the 5 parameters dijet function (left) and 5 parameters modified exponential function (right).

5.2 Wide Resonance decaying into two partons

The results presented in the previous section only refer to narrow-width resonances. In this section the search in the *high-mass* region is extended to the case of wide resonances, with width over mass ratio between 0.00001 and 30%.

The limits calculation follows the same procedure employed with the narrow resonances.

In order to take into account the fact that the resonances are wider, the limits are quoted only for a range of masses and widths that satisfies two conditions:

- at low resonance mass, the core of the signal shape is preserved after the trigger selection $m_{jj} > 1246$ GeV;
- at high resonance mass, the presence of the low-mass tails in the signal shape does not significantly affect the limit value.

The latter condition is enforced defining a truncated shape by a cut off at 70% of the nominal resonance mass:

$$m_{jj}^{truncated} \geq 70\% \cdot m_{jj}^{nominal} \tag{5.12}$$

and requiring that the expected limit derived for the truncated signal shape and corrected for the difference in acceptance because of the truncation is close to that derived for the full shape, within the typical uncertainty of 15% in the expected limits:

$$\frac{\text{Limits}^{\text{full shape}} - \text{Limits}^{\text{truncated}}}{\text{Limits}^{\text{full shape}}} \leq 15\% \tag{5.13}$$

This condition does not affect the maximum allowed resonance mass for qq resonance, for which the limits are quoted up to 8 GeV for all the widths, while it does restrict the maximum mass for a gg resonance. The acceptance due to

the truncation as a function of the gg resonance mass and the width is shown in Figure 5.11. These values are used to correct the truncated limits.

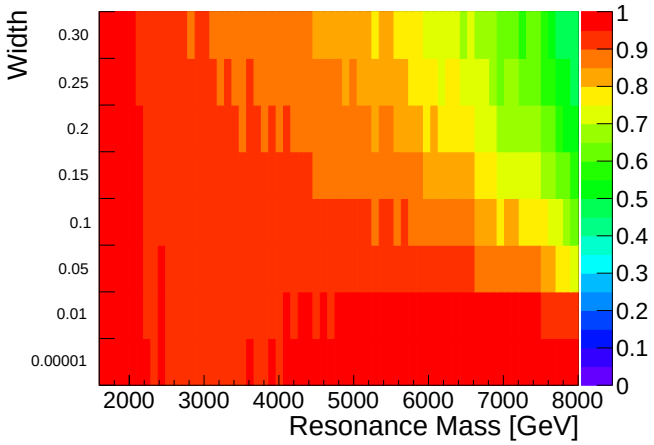

Figure 5.11. The acceptance due to the truncation in the width versus mass plane for gg resonances.

Figure 5.12 shows the quantity in Eq. 5.13. The highest resonance mass for gg final state that satisfies that condition is:

- M = 5 TeV for a resonance with 30% width;
- M = 5.5 TeV for 25% width;
- M = 6 TeV for 20% width;
- M = 7.5 TeV for 15% width;
- M = 8 TeV for 10%, 5% and narrow resonances.

The truncated shapes are used only for this study in order to decide the maximum mass for which the limits can be set, while the limits calculation relies on the full shapes.

The excluded cross section times branching ratio times acceptance at 95% CL as a function of the resonance mass and for different values of Γ/M are shown in Fig. 5.13 for qq resonances and in Figure 5.14 for gg resonances.

The limits for the different widths for the two possible final states are compared in Figure 5.15. The limits get worse as the resonance intrinsic width increases. As for the narrow-width case, the limits are less stringent in the gg case due to the more pronounced tail in the shapes at low mass.

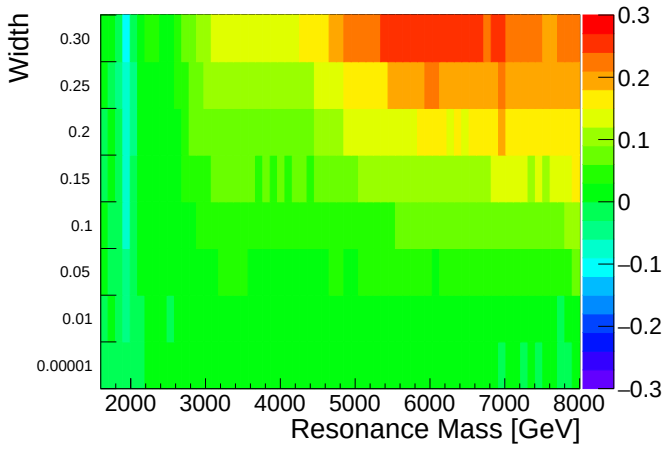

Figure 5.12. The difference between the limits with the truncated shapes and the nominal limits over the nominal limits in the width versus mass plane for gg resonances. The limits are not quoted for the masses and the widths with more than 15% of difference (red-ish zone).

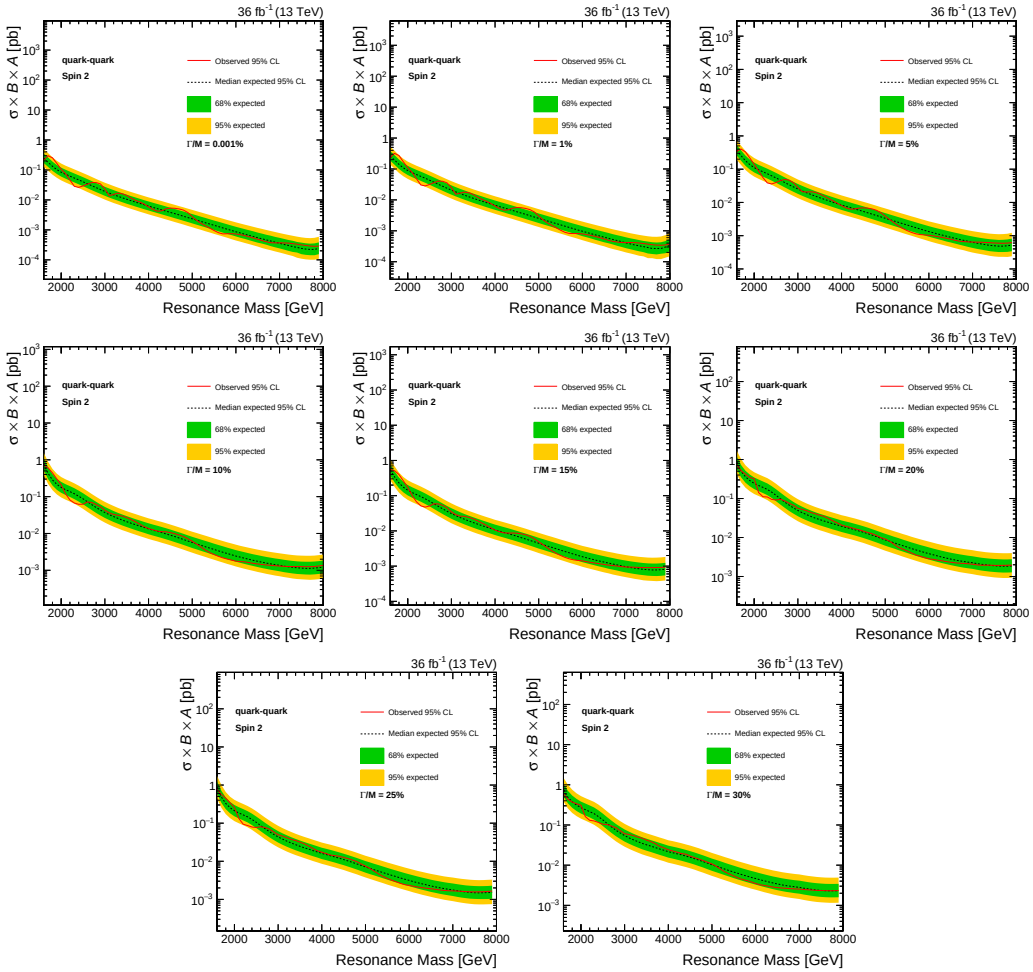

Figure 5.13. The 95% CL upper limits on the production of spin-2 resonances decaying to qq as a function of the resonance mass for the values of Γ/M considered in the analysis.

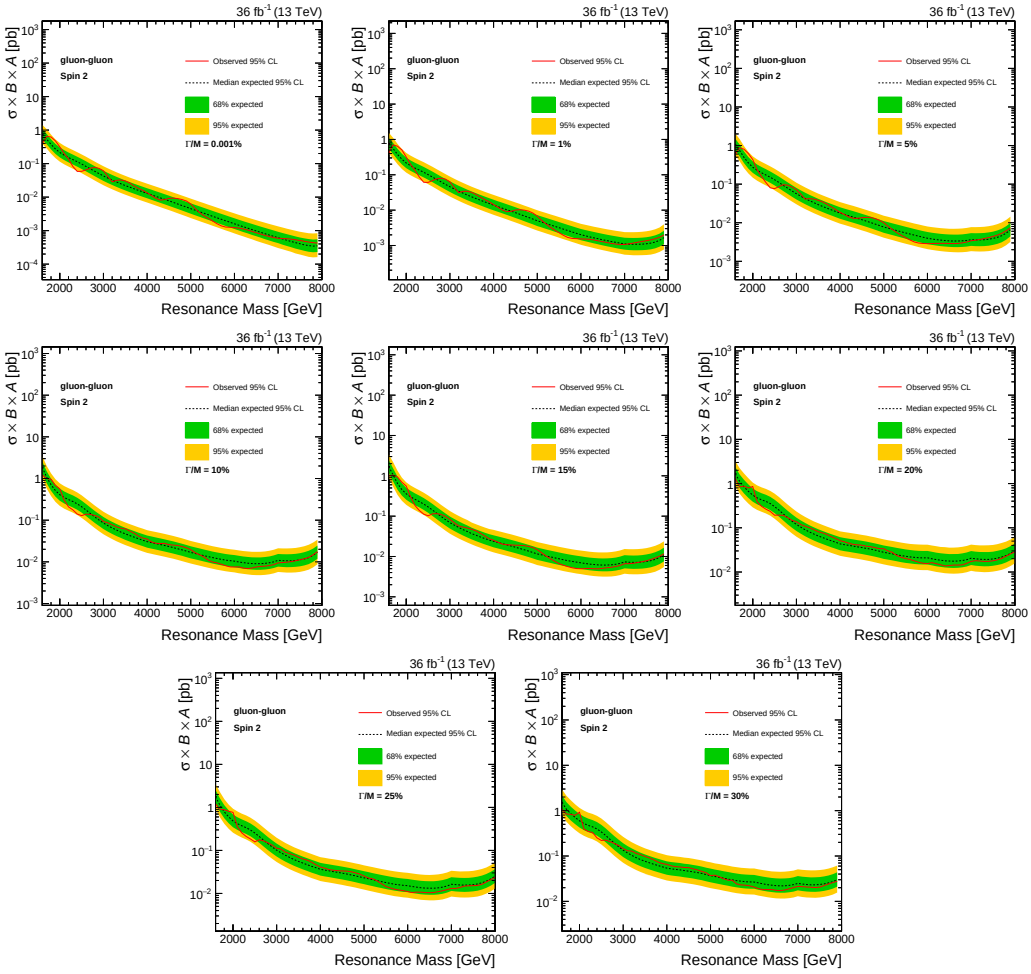

Figure 5.14. The 95% CL upper limits on the production of spin-2 resonances decaying to gg as a function of the resonance mass for the values of Γ/M considered in the analysis.

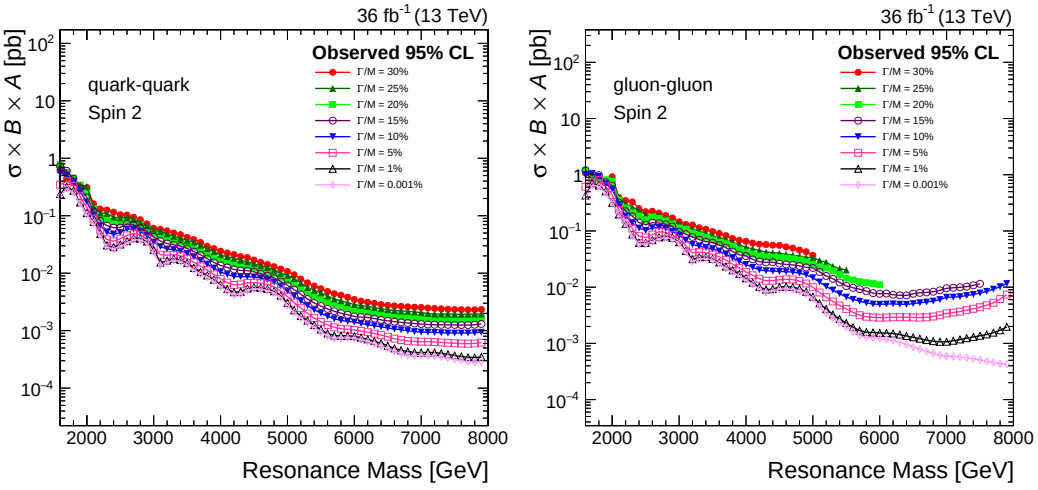

Figure 5.15. Comparison of the observed 95% CL upper limits on $\sigma \times BR \times A$ for spin-2 qq (left) and gg (right) resonances as a function of the resonance mass for the values of Γ/M considered in the analysis.

5.3 Dark matter mediator

The dark matter (DM) is an unidentified type of matter distinct from ordinary matter, dark energy and neutrinos. The name indicates that it does not interact or emit any electromagnetic radiation, which means it is invisible to the entire electromagnetic spectrum. Although it has not been observed directly, there are abundant astrophysical evidences of its presence, such as the gravitational lensing and its gravitational effects on visible matter. The standard model of cosmology indicates that 26.8% of the universe are dark matter, 68.3% are dark energy and only 4.9% are ordinary matter, which means that the majority of the universe still remain unseen. The most widely accepted hypothesis on the form for dark matter is that it is composed of weakly interacting massive particles (WIMPs) that interact only through gravity and the weak force. These reasons motivate the search for dark matter both in astrophysical experiments and at colliders.

The model considered in this analysis is the simplified model explained in Section 1.2. In this model, the DM particle is assumed to be a Dirac Fermion χ and the particle mediating the interaction is exchanged in the s-channel, as shown in Figure 5.16.

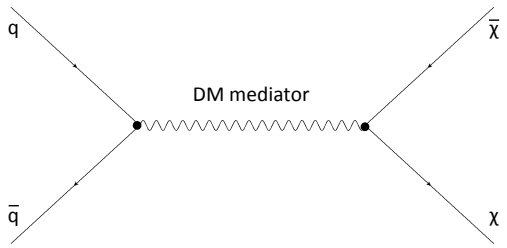

Figure 5.16. Feynman diagram for the process $q\bar{q}$ →DM mediator→ $\chi\bar{\chi}$.

The model is characterized by four parameters:

- the DM mass, m_{DM};
- the mediator mass, M_{med};
- the universal mediator coupling to quarks, g_q;
- the mediator coupling to DM, g_{DM}.

The DM mediators only decays into qq and pairs of DM particles (*leptophobic*), with unknown mass m_{DM} and the coupling set at $g_q = 0.25$ and $g_{DM} = 1.0$.

In order to find the excluded region in the dark matter mass vs. mediator mass plane, the theoretical cross sections are calculated as a function of the mediator mass for different values of m_{DM}. Dark matter masses between 1 GeV (indistinguishable from zero) and 1.7 TeV in 50 GeV steps were considered. The cross sections at parton level are calculated from MadGraph 5 simulation, which is consistent with the partial width Equations 1.21 and 1.24. Figure 5.17 shows the upper limits on the cross section as a function of the mediator mass for both types of mediator, vector

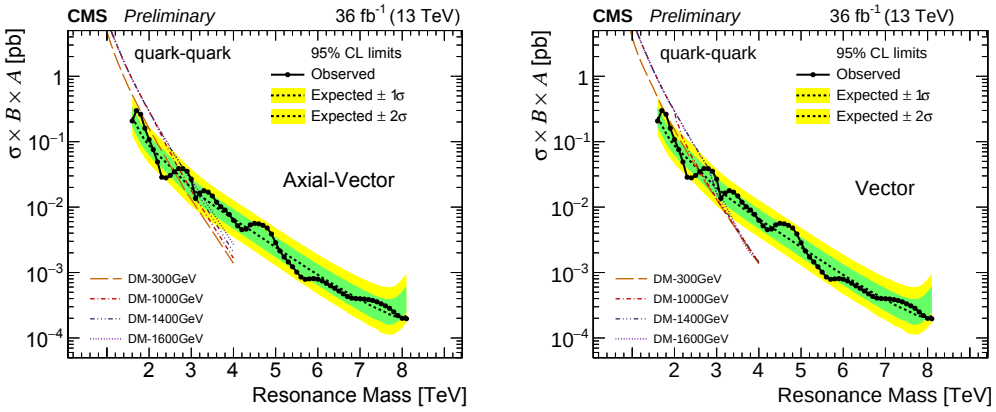

Figure 5.17. The 95% CL upper limits on the product of the cross section, branching fraction and acceptance for qq resonances, compared with the predicted cross sections of dark matter mediators for m_{DM} = 0.3, 1, 1.4, 1.6 TeV, for both axial-vector (left) and vector (right) mediator model.

and axial-vector. The theoretical cross sections for a DM mediator are reported in the same Figure for four values of dark matter mass as an example.

As for the other models seen in the previous sections, all the cross section values above the observed upper limits are excluded by this search.

Comparing the theoretical cross sections for each value of dark matter mass to the 95% CL cross section limits, we can obtain a 2D exclusion area on the dark matter mass vs. mediator mass plane. The cross sections are calculated starting from MadGraph 5 with m_{DM} = 1 GeV, and the dark matter mass dependence of the mediator cross section is determined analytically from the partial width equations 1.21 and 1.24. The exclusion regions for a DM mediator are compared with the results from other CMS analyses as the mono-X analyses (searches for missing transverse energy + X with X = jet, photon or a vector boson [57, 58, 59]) and a boosted dijet search [60]. The results of this analysis in terms of DM exclusion are also compared to constraints from the cosmological relic density of DM, as described in [61]. They are determined from the tool *MadDM* [62], which gives a numerical estimate of the expected relic density based on the standard model of cosmology for any model containing a DM candidate. The curve corresponding to $\Omega_c \times h^2 = 0.12$, which is the latest results from Planck satellite observations [63], is reported. The comparison of the exclusion regions for vector and axial-vector mediator in the dark matter mass vs. mediator mass plane is shown in Figure 5.18.

As outlined in [17] these results can be compared with results from direct detection experiments. For this purpose the limits in the dark matter mass vs. mediator mass plane are re-calculated at 90% CL and then translated into the plane of the DM mass versus the DM-nucleon interaction cross section. The translation is done using the predicted relation between the DM-nucleon interaction cross section and the mediator mass:

$$\sigma^{\mathrm{SI}}_{\mathrm{DM-N}} \simeq 6.9 \times 10^{-41} cm^2 \cdot \left(\frac{g_q g_{DM}}{0.25}\right)^2 \left(\frac{1\ \mathrm{TeV}}{M_{med}}\right)^4 \left(\frac{\mu_{n\chi}}{1\ \mathrm{GeV}}\right)^2 \tag{5.14}$$

$$\sigma^{\mathrm{SD}}_{\mathrm{DM-N}} \simeq 2.4 \times 10^{-42} cm^2 \cdot \left(\frac{g_q g_{DM}}{0.25}\right)^2 \left(\frac{1\ \mathrm{TeV}}{M_{med}}\right)^4 \left(\frac{\mu_{n\chi}}{1\ \mathrm{GeV}}\right)^2 \tag{5.15}$$

where $\mu_{n\chi} = m_n m_{DM}/(m_n + m_{DM})$ with $m_n \simeq 0.939$ GeV the nucleon mass. An axial-vector mediator leads to a spin-dependent cross section, σ^{SD}, while a vector mediator leads to a spin-independent cross section, σ^{SI}. Figure 5.19 shows the results from the CMS searches previously quoted and compares them with dark matter search by direct detection experiments [64, 65, 66, 67, 68, 69, 70, 71, 72].

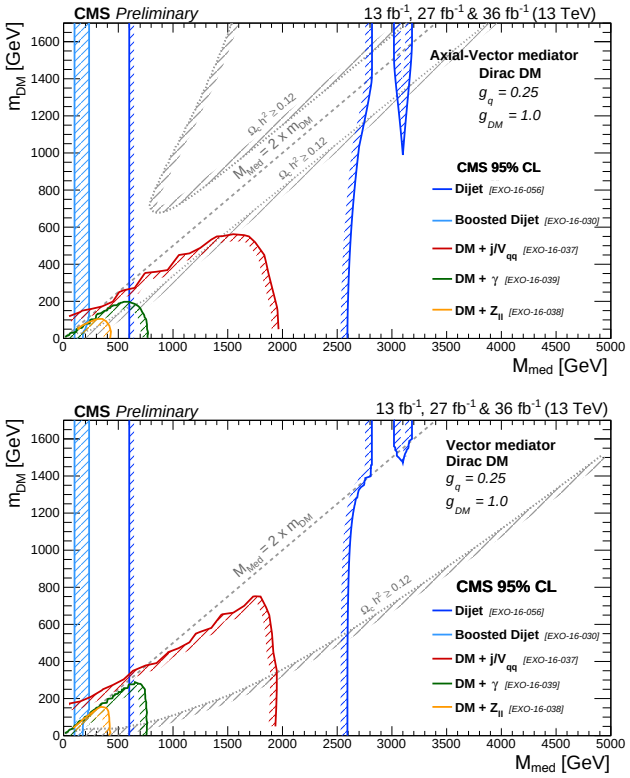

Figure 5.18. The 95% CL observed excluded regions in the dark matter mass vs. mediator mass plane, for an axial-vector mediator (top) and a vector mediator (bottom). The results using 27 fb^{-1} and 36 fb^{-1} in the dijet channel (dark blue) are compared with the results using 13 fb^{-1} from boosted dijets (light blue) and from MET + X searches in the mono-jet channel (red), mono-photon channel (green) and mono-Z channel (yellow). The constraints from the cosmological relic density of DM are reported as well (light gray).

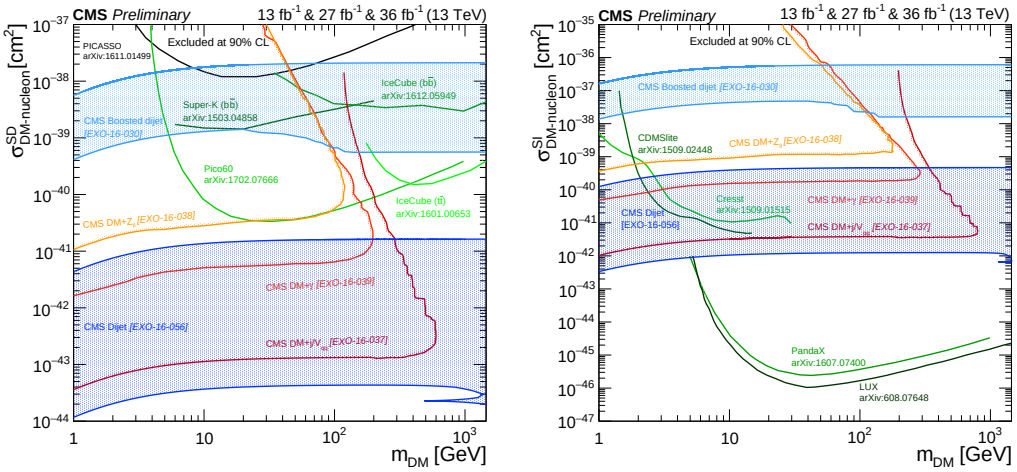

Figure 5.19. Excluded regions at 90% CL in the dark matter-nucleon interaction cross section vs. dark matter mass plane. The CMS exclusions using 27 fb^{-1} and 36 fb^{-1} from dijets channel (dark blue shaded) and using 13 fb^{-1} from boosted dijets (light blue shaded), mono-jets (red), mono-photons (orange), and mono-Z (yellow) are compared with direct detection experiments. (Left) The CMS exclusion of a spin-dependent cross section for an axial-vector mediator compared with limits from PICASSO [64], PICO-60 [65], IceCube [66, 67] and Super-Kamiokande [68]. (Right) The CMS exclusion of a spin-independent cross section for a vector mediator compared with limits from LUX [69], PandaX-II [70], CDMSLite [71], and CRESST-II [72]. The CMS exclusions are set for Dirac DM particle and couplings $g_q = 0.25$ and $g_{DM} = 1$, for leptophobic axial-vector and vector mediators, and they strongly depend on these choices and are not applicable to other choices of coupling values or models. The CMS limits do not include a constraint on the relic density.

5.3.1 Wide dark matter mediator

In the previous section we fixed the couplings g_q and g_{DM} and we performed a scan over the masses m_{DM} and M_{med}. This allowed to set limits in the dark matter mass vs. mediator mass plane.

In this section limits in the quark coupling vs. mediator mass plane are set in order to constrain another parameter of the phase space. For this purpose we fix the dark matter parameters $m_{DM} = 1$ and $g_{DM} = 1$ and we perform a scan over the other two parameters. For this particular choice of the parameters, the relation between the resonance width and the quark coupling is:

$$\Gamma \approx \frac{(18g_q^2 + 1)M_{med}}{12\pi} \tag{5.16}$$

Considering wide dark matter mediators (high Γ/M_{med}), it therefore means putting constraints on high quark coupling.

The following results are shown assuming the model of vector dark matter mediator, since the shapes and limits are found indistinguishable from the axial-vector mediator model and therefore the conclusions are valid for both models.

The study with the truncated shapes explained in the previous sections is performed also in this case. The highest mediator mass for which the limits can be set is:

- M = 5 TeV for a resonance with 30% width;
- M = 5.5 TeV for a resonance with 20% width;
- M = 6 TeV for 20% width;
- M = 6.5 TeV for 15% width;
- M = 7 TeV for 10% width;
- M = 7.5 TeV for 5% width;
- M = 8 TeV for 1% width.

The observed exclusion limits on the cross section times the branching ratio times the acceptance at 95% CL as a function of the resonance mass for different values of Γ/M is shown in Fig. 5.20. The results taking into account the truncation study are summarized in Figure 5.21.

The theoretical cross sections for the production of a DM mediator in the model with $m_{DM} = 1$ GeV and $g_{DM} = 1$ are calculated from MadGraph 5 for different quark coupling between $0.1 < g_q < 1.0$ in 0.1 steps, as shown in Fig. 5.22.

In order to compare the observed limits on the cross section (reported for different width-to-mass ratio Γ/M) and the theoretical cross section (reported in terms of coupling g_q) we used the relationship 5.16. Using this relation, the MadGraph predictions for the mediator cross section can be converted to a function of width and then they can be compared to the observed cross section limits. The conversion is performed for mediator mass in the range $1.6 < M_{med} < 4.7$ TeV in 100 GeV

steps because the analysis has no sensitivity to higher mass. The comparison is shown for some of the masses considered as an example in Figure 5.23.

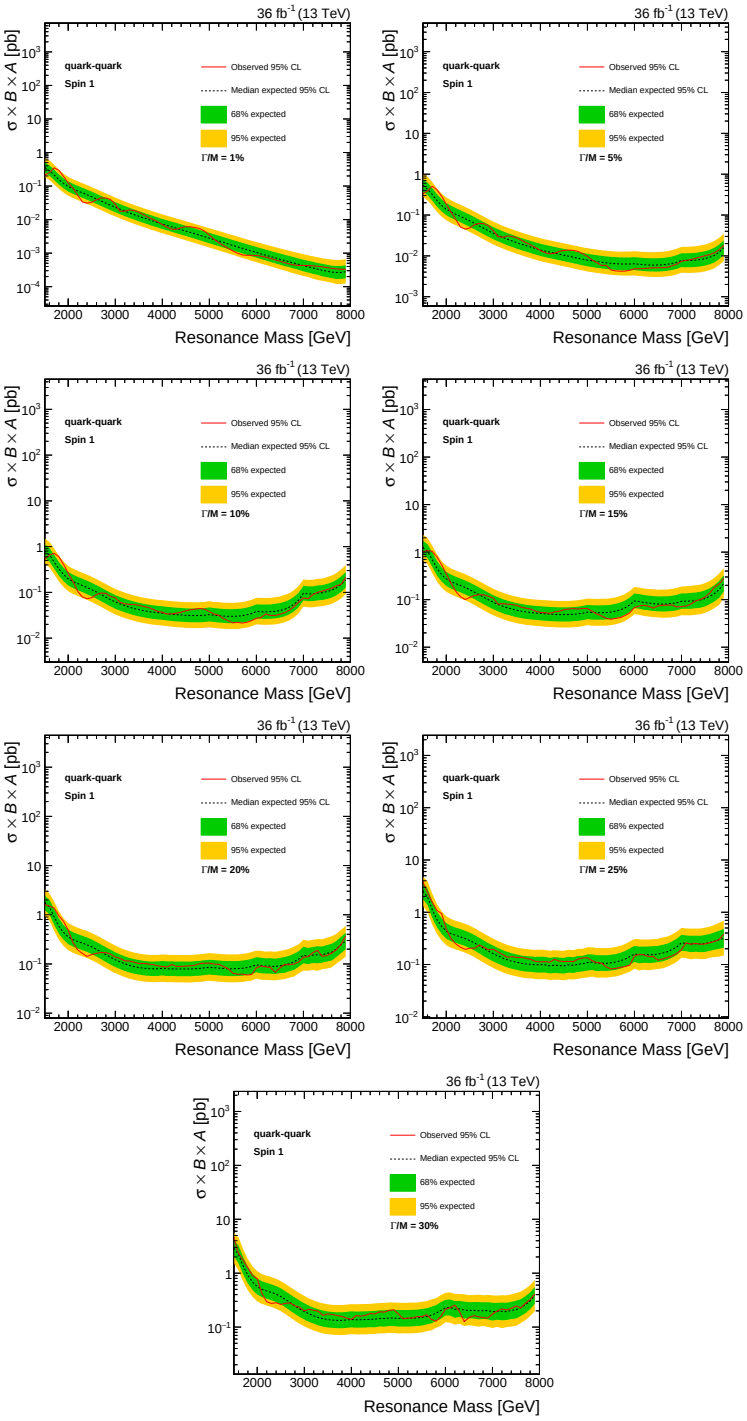

Figure 5.20. The 95% CL upper limits on the production of a spin-1 dark matter mediator decaying to qq as a function of the mediator mass for the values of Γ/M considered in the analysis.

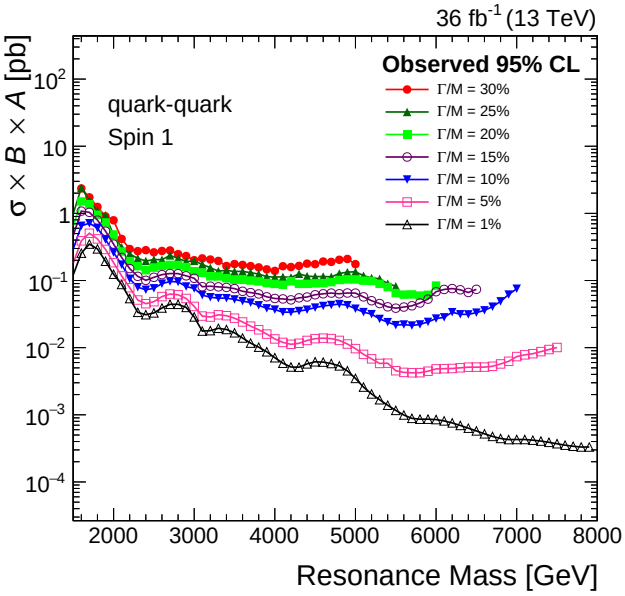

Figure 5.21. Comparison of the observed 95% CL upper limits on $\sigma \times BR \times A$ for a spin-1 dark matter mediator decaying to qq as a function of the mediator mass for the values of the width-to-mass ratio Γ/M considered in the analysis.

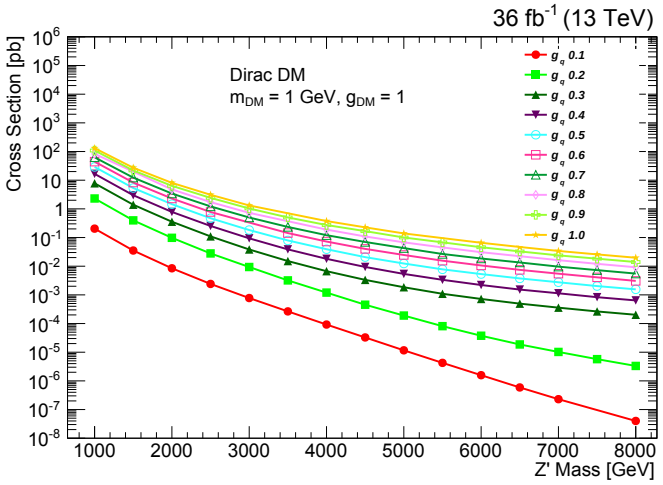

Figure 5.22. The theoretical dark matter mediator cross section as a function of the mediator mass for $m_{DM} = 1$ GeV, $g_{DM} = 1$ and quark coupling between $0.1 < g_q < 1.0$.

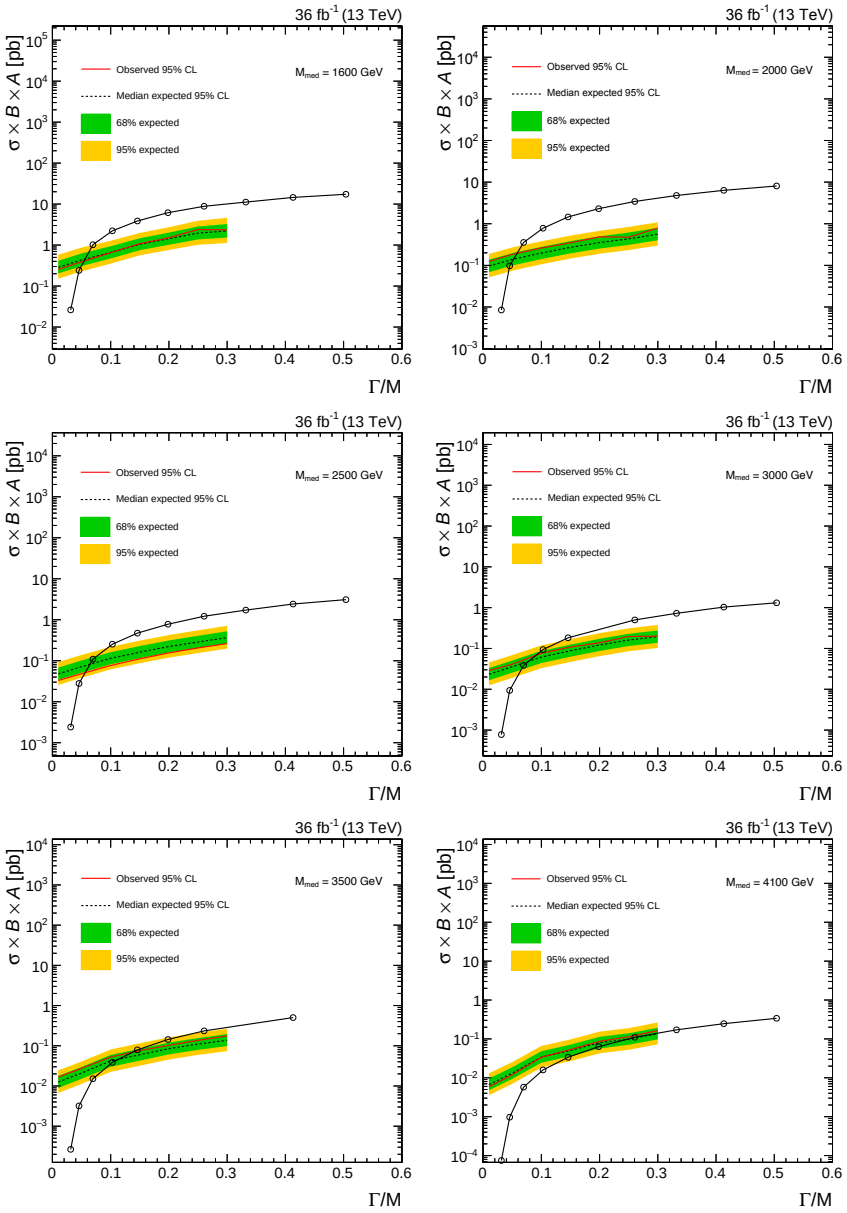

Figure 5.23. The 95% CL upper limits on the production cross section times the branching ratio times the acceptance for a dark matter mediator as a function of Γ/M. Some of the mediator masses considered in the analysis are shown as an example. Limits are compared to cross sections of dark matter mediator for $m_{DM} = 1$ GeV and $g_{DM} = 1$.

All the cross section values above the observed upper limits on the cross section are excluded by this search as well as the corresponding values of Γ/M. Finally we can revert back to the quark coupling using Equation 5.16 to find the excluded values of g_q as a function of the mediator mass. This result is shown in Fig. 5.24 where we can see that the observed upper limit on g_q varies from approximately 0.2 to 0.8. These values correspond to natural widths of 5% (narrow resonance) and 30% (wide resonance) respectively, demonstrating that sensitivity to higher values of g_q are obtained from this wide resonance search. In fact the limits from the narrow resonance search are approximately valid up to a quark coupling of about 0.4, corresponding to a width of 10%, for a vector dark matter mediator up to a mass of 3.7 TeV. The limits taking into account the natural width of the resonance improve the accuracy of the narrow width limits, probing up to a larger coupling of 0.76, corresponding to a natural width of 30%, for a dark matter mediator up to a higher mass of 4.1 TeV.

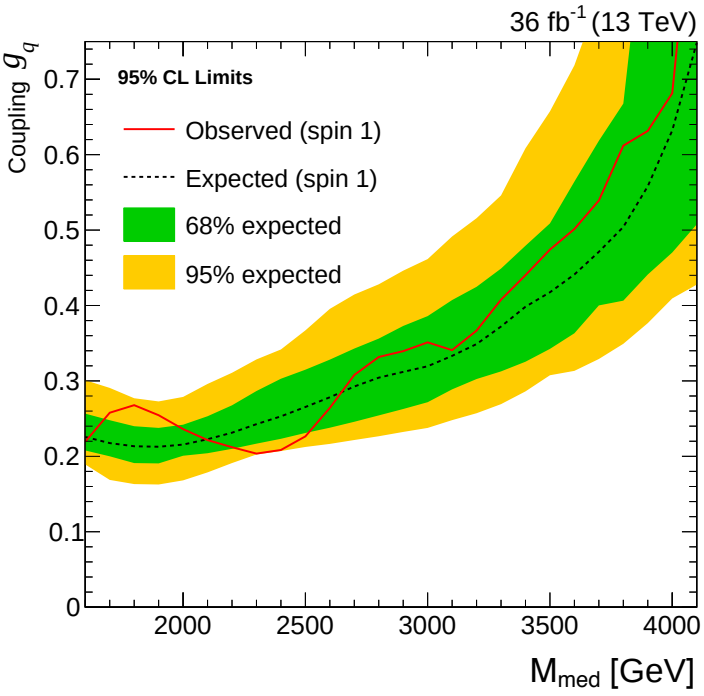

Figure 5.24. The 95% CL upper limits on the quark coupling g_q as a function of the dark matter mediator mass. The observed limits (solid), expected limits (dashed) and their variation at the 1 and 2 standard deviation levels (shaded bands) are shown.

Conclusions

In this thesis a model independent search for physics beyond the Standard Model with a dijet final state was presented. The analyzed dataset consists of 36 fb^{-1} of proton-proton collisions delivered by the LHC with a center-of-mass energy of 13 TeV, and collected by the CMS experiment in 2016. This dataset overcame in few weeks the data collected in 2015 (corresponding to an integrated luminosity of 2.4 fb^{-1}) and therefore the analysis of these data was fundamental since the beginning of 2016 data-taking.

The search was performed in the region of masses between ∼500 GeV and 8 TeV. Due to trigger and disk-space limitations, the lower part of this range was not investigated in Run2 and could be probed by this analysis thanks to the data-scouting technique.

The analysis strategy consists in searching for a peak in the dijet mass distribution over a falling background spectrum. The observed dijet mass spectrum is smooth and no significant excess was found. Upper limits on the production cross section of new resonances are set for three different final states (quark-antiquark, quark-gluon and gluon-gluon). Either narrow and wide resonances with a width up to 30% of the resonance mass are considered. The observed limits are then compared with the theoretical expectation for 9 benchmark models. Such models consider either the production of a massive resonance directly coupling to quarks or gluons, or the presence of a vector or axial-vector mediator that couples to quarks and dark matter particles. The exclusion limits set on the production of new resonances are among the most stringent ones set at colliders. Dark matter mediator masses from 600 GeV and up to ∼2.6 TeV (largely independent of the DM particle mass) are excluded for the benchmark choice of the mediator couplings. This is a more stringent limit with respect to those set by the traditional E_T^{miss} + X searches at the LHC. In the plane of the dark matter-nucleon interaction cross section versus dark matter mass, the dijet search is also more sensitive than direct detection experiments for spin-dependent cross sections.

This analysis is and will remain in the future among the most powerful ones to search for new physics and dark matter at LHC. In fact any hypothetical new particle that might be produced originates from the colliding protons and therefore it must couple to quarks and/or gluons.

For the future some improvements to the analysis are possible. The background could be estimated using a control sample and not from a direct fit to the data. With this approach many uncertainties would cancel out in the ratio between the control and the signal region, reducing the systematics on the background. A possible extension of the analysis consists in dividing the analysis in categories of $|\Delta\eta|$ with

different signal purity and combining the results, instead of the angular cut currently applied. This would improve the global analysis sensitivity and would extend the analysis to models that predict a resonance out of the current signal region.

Appendix A

Jet energy calibration using γ + jet events

Jets are composite objects coming from the hadronization of outgoing partons in LHC collisions. Jet clustering algorithm are designed to group in a single object all the stable particles coming from the parton shower. Ideally, the reconstructed jet energy should be the same as the showering parton, which is the information needed from the analyses. However the measured energy is affected by detector reconstruction and physics effects like initial and final state radiation and pile-up. To correct the jet energy, a factorized approach with several independent correction levels is employed in CMS. Each level aims at correcting for different effects, as explained in Section 3.3. One promising event topology useful to calculate the jet energy corrections is the photon + jet process in which one isolated photon is back-to-back in the transverse plane to one jet. The photon energy is in fact measured with very high precision, with a negligible resolution with respect to the jet's one, and can be assumed as the "true" energy. Therefore, comparing the photon and jet energies, the jet energy scale can be obtained. The photon + jet calibration is part of the last level of jet energy corrections (*L2L3Residual JEC*). At this point the jets in the MC simulation are fully calibrated, while the jets in data have to be corrected to match the same jet energy scale of MC.

A.1 Datasets and triggers

The jet energy calibration is based on the entire 2016 dataset which consists in 36 fb^{-1} of data collected by CMS experiment in 2016 at $\sqrt{s} = 13$ TeV. Due to the different detector conditions the dataset is split in four periods for which different JEC are derived. The datasets used, with corresponding integrated luminosities collected in 2016 are listed in A.1.

Signal events are stored in the *SinglePhoton* dataset after having passed one of the *SinglePhoton* triggers. These triggers require the presence of an energy deposit in ECAL, with requirements on transverse momentum and other identification variables. In detail, the variables used for photon identification in the trigger are:

- R9: The ratio between the energy contained in a 3×3 crystals matrix around

Datasets	Integrated luminosity (fb^{-1})
First period	
SinglePhoton Run2016B	~ 5.8
SinglePhoton Run2016C	~ 2.6
SinglePhoton Run2016D	~ 4.3
	B + C + D ~ 12.7
Second period	
SinglePhoton Run2016E	~ 4.0
SinglePhoton Run2016F	~ 2.7 (run < 278802)
	E + Early F ~ 6.7
Third period	
SinglePhoton Run2016F	~ 0.4 (run > 278802)
SinglePhoton Run2016G	~ 7.6
	Late F + G ~ 8.0
Fourth period	
SinglePhoton Run2016H	~ 8.8

Table A.1. 2016 data taking divided in the four different periods and the corresponding integrated luminosity.

the most energetic crystal and the energy of the whole photon supercluster. This variable makes a good separation between unconverted photons (energy not spread in tracker) and converted photons (energy spread by magnetic field before reaching ECAL);

- H/E: the ratio between the energy in HCAL geometrically close to the photon, divided by the energy deposit in ECAL 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 (Charged/Neutral/Electromagnetic): The scalar sum of the energy of the Charged/Neutral/Electromagnetic PF Candidates inside a cone of radius 0.4 around the photon candidate (excluding its energy). For a clean γ+jet event it is close to zero.

The triggers used with their definition and the photon p_T range in which are required are listed in Table A.2.

The analysis makes use of a MC sample generated with PYTHIA 8 and CUETP8M1 tune, including the simulation of the LHC operational conditions, reported in Table A.3.

Trigger Name	Online selection	photon p_T range [GeV]
HLT_Photon30_R9Id90_HE10_IsoM	$\boldsymbol{p_T > 30}$; $H/E < 0.1$ $R9 > 0.9$; $Iso < 10GeV$	40-60
HLT_Photon50_R9Id90_HE10_IsoM	$\boldsymbol{p_T > 50}$; $H/E < 0.1$ $R9 > 0.9$; $Iso < 10GeV$	60-85
HLT_Photon75_R9Id90_HE10_IsoM	$\boldsymbol{p_T > 75}$; $H/E < 0.1$ $R9 > 0.9$; $Iso < 10GeV$	85-105
HLT_Photon90_R9Id90_HE10_IsoM	$\boldsymbol{p_T > 90}$; $H/E < 0.1$ $R9 > 0.9$; $Iso < 10GeV$	105-130
HLT_Photon120_R9Id90_HE10_IsoM	$\boldsymbol{p_T > 120}$; $H/E < 0.1$ $R9 > 0.9$; $Iso < 10GeV$	130-175
HLT_Photon165_R9Id90_HE10_IsoM	$\boldsymbol{p_T > 165}$; $H/E < 0.1$ $R9 > 0.9$; $Iso < 10GeV$	> 175

Table A.2. Triggers used in the γ+jet analysis with their definition and the corresponding photon p_T range in which they are used.

Dataset	Cross section (pb)
/GJet_Pt-15To6000_TuneCUETP8M1-Flat_13TeV_pythia8_20M	365896

Table A.3. MC sample used in the analysis with the corresponding cross section.

A.2 Event selection

The event reconstruction is performed using the Particle Flow algorithms that reconstructs all the particles in an event and identifies them as muons, electrons, photons, charged and neutral hadrons. The hadrons are then merged together with the anti-k_T algorithm with cone ΔR=0.4 (PF ak4 jets), as for the jets used in the *high-mass* dijet analysis.

The γ+jet process can be produced from the Feynman diagram reported in Figure A.1.

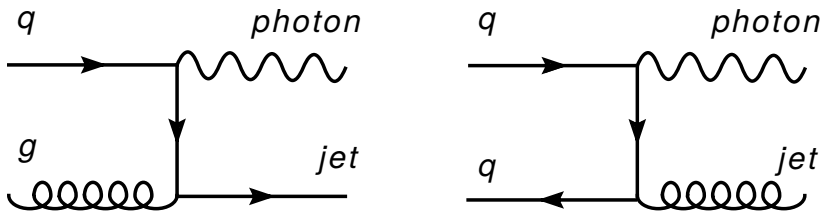

Figure A.1. Dominant photon + jet production diagrams.

Its main background is represented by QCD events in which a jet is misidentified as a photon. In order to keep this background at reasonably low level, a stringent photon identification is required. This analysis makes use of CMS cut-based photon identification for 2016 data, using the tightest working point with an efficiency of 71% and a background rejection of about 89%. Furthermore the photon, after having passed the trigger selection described in the previous section, have to pass

an offline selection based on the variables explained in the previous section and on the following additional variables:

- $\sigma_{i\eta i\eta}$: the spatial second order moment of the photon candidate in the η direction, computed as

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_{i\in 5\times 5}(\eta_i - \bar{\eta})^2 w_i}{\sum_{i\in 5\times 5} w_i}}, \qquad w_i = \max(0, 4.7 + \log(E_i/E_{5\times 5})) \quad \text{(A.1)}$$

- "Conversion safe" electron veto which removes the photon candidate if its supercluster is matched to an electron track with no missing hits in the innermost tracker layers.

The final photon selection applied is reported in Table A.4.

Variable	Cut
H/E	< 0.05
$\sigma_{i\eta i\eta}$	< 0.01
charged hadron iso	< 0.76
neutral hadron iso	$< 0.97 + 0.014 \cdot p_T + 0.000019 \cdot p_T^2$
electromagnetic iso	$< 0.08 + 0.0053 \cdot p_T$
conv. safe electron veto	yes
R9	> 0.90

Table A.4. Selection applied to the photon.

In addition to these requirements, only photons in the pseudorapidity region of $|\eta| < 1.3$ are considered since they have the best energy measurement in this region. The photon is also required outside the jet cone requiring $\Delta R(jet, \gamma) > R_{cone}$, where R_{cone} is 0.4 in the case of ak4 jets and ΔR is the geometrical distance defined as $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

The jets are identified applying the CMS cut-based jet identification for 2016 data, using the loosest working point. It is based on the same variables used for the jet identification in the *high-mass* dijet analysis, reported in Section 4.5. The selection is applied on the jets with $p_T > 15$ GeV and it is reported in Table A.5.

The jets are corrected with the first levels of correction up to the L2Residual JEC and these corrections are propagated to the missing transverse energy calculation that will be used in a jet energy response definition.

Besides the mentioned requirements concerning the object's attributes, two further criteria related to the event topology are crucial for this analysis:

- event balancing in the transverse plane: $\Delta\phi(j_1, \gamma) > 2.8$, being $\Delta\phi(j_1, \gamma)$ the angle in the transverse plane between the leading jet and the photon;

- fraction of the transverse momentum of the subleading jet with respect to the photon $\alpha < 0.3$, with $\alpha = p_T(j_2)/p_T(\gamma)$ (see Section A.3.3).

Variable	Cut
For $\|\eta(jet)\| \leq 2.7$:	
NHF	< 0.99
NEF	< 0.99
# of constituents	> 1
and in addition for $\|\eta(jet)\| \leq 2.4$:	
CHF	> 0
CEF	< 0.99
Charged multiplicity	> 0
For $2.7 < \|\eta(jet)\| \leq 3.0$:	
NEF	<0.90
# of neutral particles	> 2
For $\|\eta(jet)\| > 3.0$:	
NEF	<0.90
# of neutral particles	> 10

Table A.5. The requirements of the "Loose" Jet ID applied on the γ+jet analysis.

A.3 Jet energy response measurement

Two methods can be used to quantify the jet energy response: one is based on the transverse momentum balancing, while the other is the missing transverse energy projection fraction (MPF) method. In both cases the jet energy response is calculated in bins of $p_T(\gamma)$ and $\eta(j_1)$, where j_1 is the leading jet, which are reported in Table A.6.

In the next sections both the methods are illustrated showing the results for the central region of the detector as an example.

$\|\eta(j_1)\|$ bins						
<1.3	1.3-2.0	2.0-2.5	2.5-3.0	3.0-3.2	3.2-5.2	
$p_T(\gamma)$ bins						
40-50	50-60	60-85	85-105	105-130	130-175	175-230
230-300	300-400	400-500	500-700	700-1000	>1000	

Table A.6. Bins in $p_T(\gamma)$ and $\eta(j_1)$ in which the jet energy responses are calculated.

A.3.1 Transverse momentum balancing

In γ+jet events, the photon and the jet are back-to-back in the transverse plane and the transverse momenta are balanced. It is therefore possible defined the jet energy response exploiting this balancing. In this first definition the jet energy response (R_{Bal}) is defined as the ratio between the transverse momentum of the leading jet (j_1) and the transverse momentum of the photon:

$$R_{Bal} = \frac{p_T^{j_1}}{p_T^{\gamma}} \tag{A.2}$$

The mean of this distribution in each p_T and $|\eta|$ bin is taken as jet energy response in that bin.

A.3.2 Missing energy projection fraction

The missing energy projection fraction (MPF) is an alternative method that uses the missing transverse energy ($\vec{E}_T^{miss}$) instead of the jet transverse momentum. In an ideal case, 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 process:

$$\vec{p_T}^{\gamma} + \vec{p_T}^{recoil} = 0 \tag{A.3}$$

On the detector level (*reco*) instead, these transverse momenta are scaled by the detector responses R_γ and R_{recoil} of the respective quantity:

$$\begin{aligned} p_T^{reco\ \gamma} &= R_\gamma \cdot p_T^{\gamma} \\ p_T^{reco\ recoil} &= R_{recoil} \cdot p_T^{recoil} \end{aligned} \tag{A.4}$$

Deviations of the detector responses from one lead therefore to some reconstructed missing energy:

$$\vec{p_T}^{reco\ \gamma} + \vec{p_T}^{reco\ \gamma} = -\vec{E_T}^{miss} \tag{A.5}$$

that can be written, substituting Equation A.4 in Equation A.5, as:

$$R_\gamma \vec{p_T}^{\gamma} + R_{recoil} \vec{p_T}^{recoil} = -\vec{E_T}^{miss} \tag{A.6}$$

Multiplying both sides of Equation A.6 by $\frac{\vec{p_T}^{\gamma}}{|p_T^{\gamma}|^2}$ and substituting $\vec{p_T}^{recoil}$ with Equation A.3, the definition of R_{MPF} is obtained as:

$$R_{MPF} \equiv R_{recoil} = 1 + \frac{\vec{E_T}^{miss} \cdot \vec{p_T}^{\gamma}}{|p_T^{\gamma}|^2} \tag{A.7}$$

where it has been assumed a response for the photon equal to 1 ($R_\gamma = 1$) since the excellent photon reconstruction.

In Figure A.2 is reported an example of the jet energy response distribution using the two different methods in the bin $|\eta| < 1.3$, $105 < p_T < 130$ and for the last data-taking period.

A.3.3 Extrapolation

The jet energy response is well defined when only one jet is present in the event, exactly back-to-back with respect to the photon. However if the recoiling parton radiates a gluon by final state radiation some of the energy could not be contained in the leading jet cone, generating an additional jet in the event. A sketch of a γ+jet event with a second jet caused by the radiation of a gluon is shown in Figure A.3.

To mitigate the effect of FSR the results have to be extrapolated to the ideal event topology with one photon and only one jet. For this purpose, the variable α

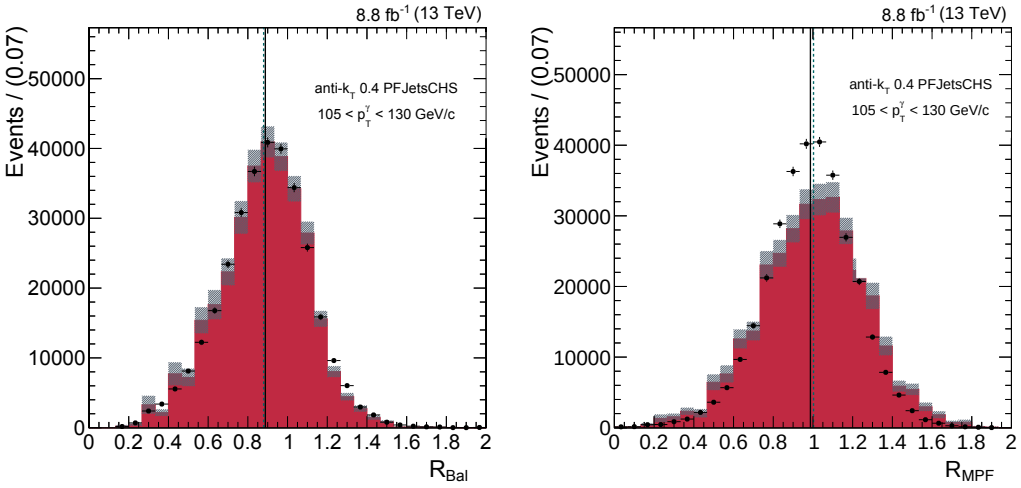

Figure A.2. The jet energy response using the p_T balancing response (left) and the MPF method (right) for the photon p_T range 105-130 GeV and $|\eta|$ <1.3.

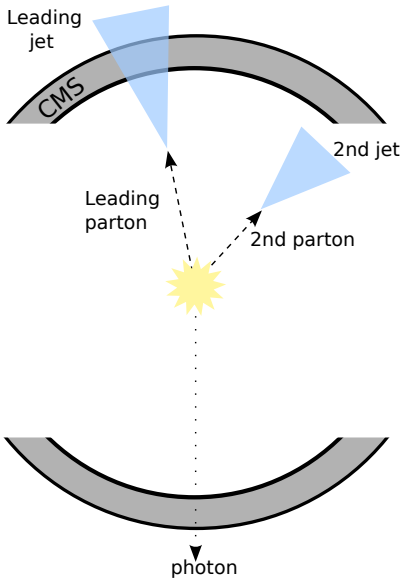

Figure A.3. Schematic view of a γ+jet event in the transverse plane. The second jet from ISR/FSR causes an imbalance between the leading jet and photon p_T.

is defined as the ratio between the transverse momentum of the second leading jet to the photon's one:

$$\alpha \equiv \frac{p_T^{j_2}}{p_T^{\gamma}} \tag{A.8}$$

This variable takes into account the secondary activity in the event, in fact, in the ideal case with no secondary activity ($p_T^{j_2} = 0$), α will be 0. A cut on this variable is imposed at 0.3, as reported in section A.2, in order to select events with low additional activity. The jet energy response is therefore calculated in bins of α, from 0.03 to 0.3 by step of 0.03, and the response is extrapolated to $\alpha \to 0$ for each photon p_T and jet η bin of the analysis. In Figure A.4 is reported the response (for both methods) as a function of α in a particular (η, p_T) bin.

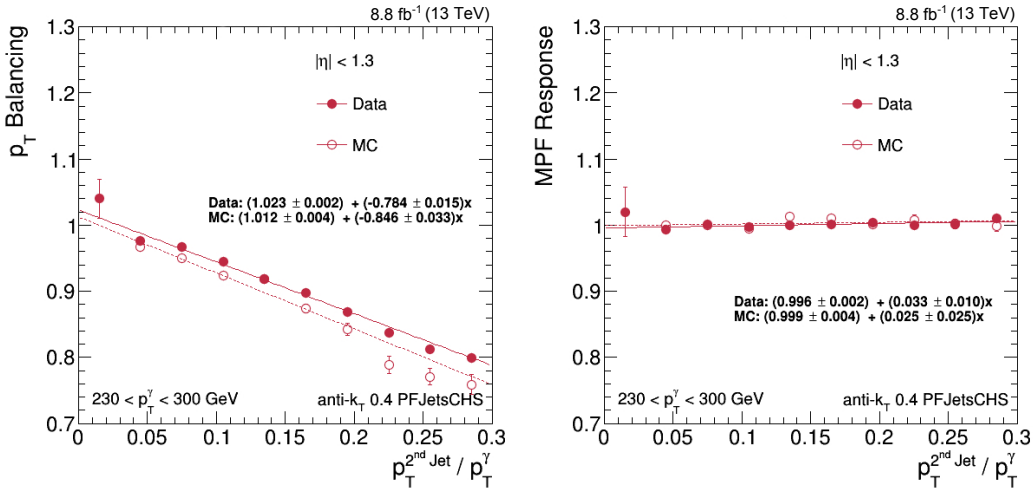

Figure A.4. Jet energy response using the p_T balancing method (left) and MPF method (right) are shown as a function of α and extrapolated to $\alpha \to 0$ for the photon p_T range 230-300 GeV and $|\eta| < 1.3$.

A.4 Results

As described in Section A.3, the p_T balancing and MPF methods are used to estimate the jet response. The means of the jet balancing and MPF distributions for data and for MC (see Figure A.2) in each $p_T(\gamma)$ and $\eta(j_1)$ bin represent the jet energy responses, R_{Bal} and R_{MPF}, for that bin.

In figure A.5 these responses as a function of the p_T are reported for both methods. The jet energy responses are reported as an example for the first η bin and before the extrapolation to zero secondary activity.

After the extrapolation to $\alpha \to 0$, performed as explained in Section A.3.3, the extrapolated jet energy responses as a function of the photon p_T in the first η bin are reported in Figure A.6.

The *L3Residual* jet energy corrections needed to cure the difference between data and MC as a function of the p_T are calculated as the ratio between the jet

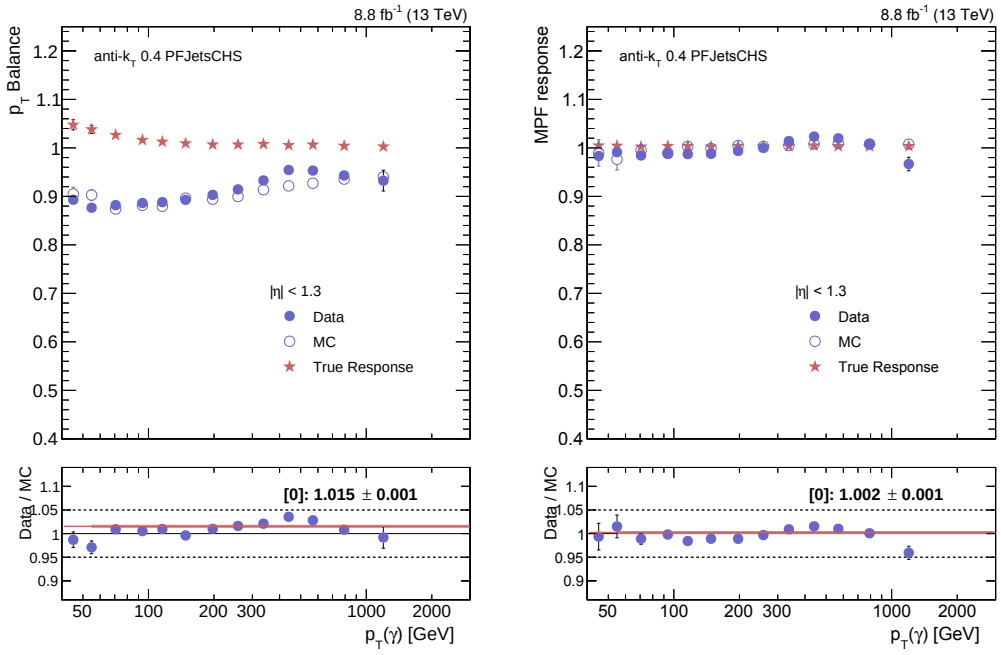

Figure A.5. Jet energy response calculated using the p_T-balancing method (left) and the MPF method (right) as a function of photon p_T in the region $|\eta| < 1.3$ and before the extrapolation.

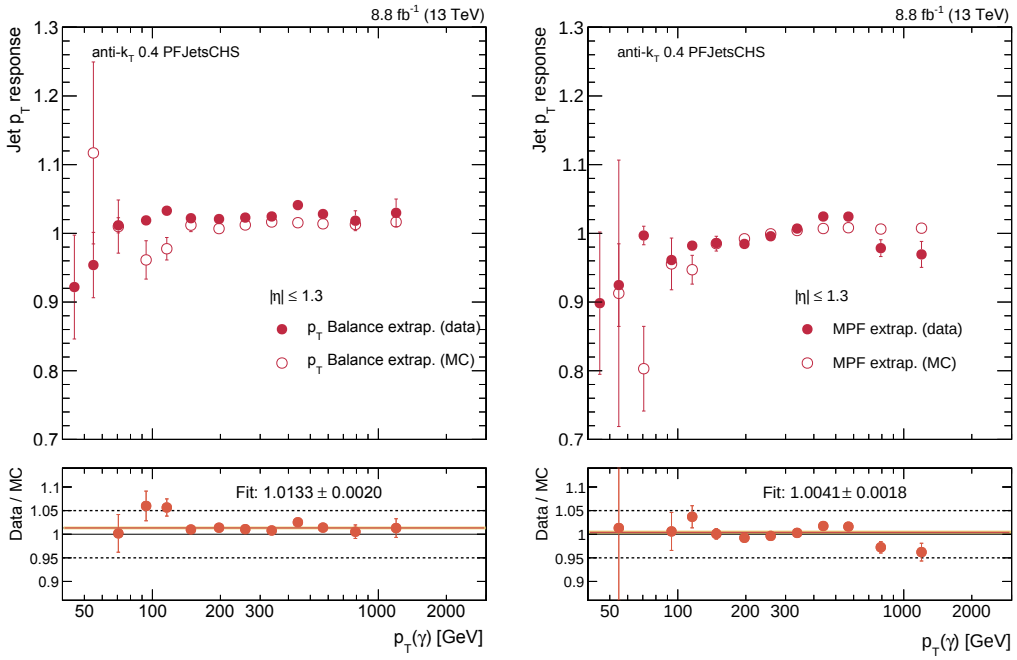

Figure A.6. Jet energy response calculated using the p_T-balancing method (left) and the MPF method (right) as a function of photon p_T in the region $|\eta| < 1.3$ and after the extrapolation to $\alpha \to 0$.

energy response in data and in MC using the two methods described. The results from the γ+jet analysis are combined with the results from Z+jet analysis and Multijets analysis as explained in detail in Section 3.3.

Appendix B

Closure test for the HLT-to-RECO JEC

The differences in the jet reconstruction between the online and the offline jets lead to differences in the jet energy. The HLT-to-RECO corrections described in Section 3.3.4 are applied on the HLT CaloJets employed in the *low-mass* analysis to cure these differences. In this section we want to check the level of agreement between HLT CaloJets and RECO PFJets after having applied these corrections. The test is performed applying the HLT-to-RECO corrections to the HLT CaloJets and comparing the HLT dijet mass with the one calculated with the RECO PFJets. Using the same dataset used to compute the corrections, in which both online and offline objects are saved, we calculate the ratio between the dijet invariant mass from the corrected HLT CaloJets and the one from the RECO PFJets:

$$R = \frac{m_{jj,HLT}^{corr}}{m_{jj,RECO}} \tag{B.1}$$

This ratio is studied as a function of the dijet invariant mass chosen randomly between HLT Calo and RECO PF in order to reduce the effect due to the different energy resolutions. The results of this test is reported in Figure B.1 in which it can be observed that the mean of the ratio, reported in black markers, is flat and around 1 for the whole dijet mass range.

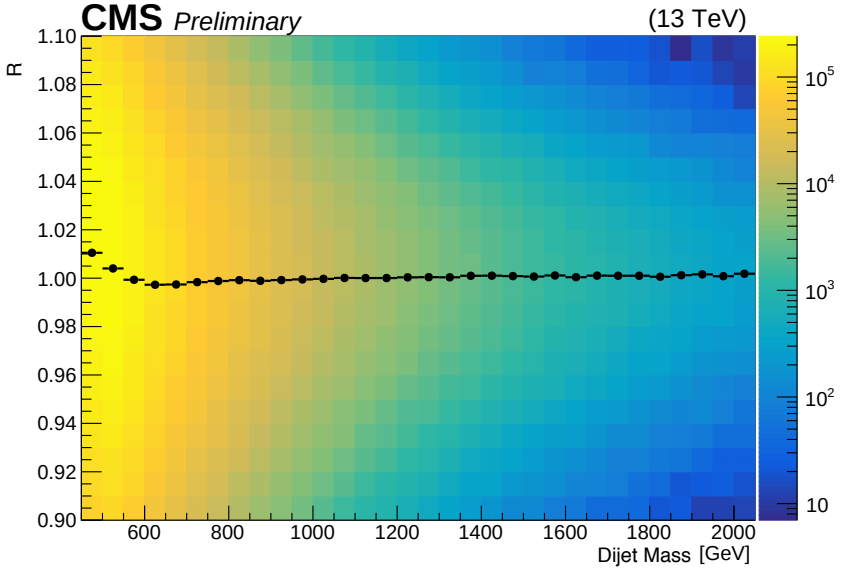

Figure B.1. The ratio between the invariant mass from the HLT CaloJets and the RECO PFJets as a function of m_{jj} randomly chosen between HLT or RECO. The colored scale on z-axis represents the number of events, while the average response of HLT jets relative to RECO jets is shown in black. It is within 1% of unity and not dependent on the mass.

Appendix C

Fisher test for the *low-mass* analysis

In this section a study on the choice of the function for the dijet invariant mass fit for the *low-mass* analysis is reported. It is performed using 4 different function families, and for each family a different number of parameters are used. The different functions used are reported in Table C.1 where $x = \frac{m_{jj}}{\sqrt{s}}$.

Standard Dijet Family	
4 Par	$\frac{d\sigma}{dm_{jj}} = \frac{p_0(1-x)^{p_1}}{x^{p_2+p_3 log(x)}}$
5 Par	$\frac{d\sigma}{dm_{jj}} = \frac{p_0(1-x)^{p_1}}{x^{p_2+p_3 log(x)+p_4 log^2(x)}}$
6 Par	$\frac{d\sigma}{dm_{jj}} = \frac{p_0(1-x)^{p_1}}{x^{p_2+p_3 log(x)+p_4 log^2(x)+p_5 log^3(x)}}$
Polynomial extension Family	
5 Par	$\frac{d\sigma}{dm_{jj}} = \frac{p_0(1-x)^{p_1}(1+p_4 x)}{x^{p_2+p_3 log(x)}}$
6 Par	$\frac{d\sigma}{dm_{jj}} = \frac{p_0(1-x)^{p_1}(1+p_4 x+p_5 x^2)}{x^{p_2+p_3 log(x)}}$
7 Par	$\frac{d\sigma}{dm_{jj}} = \frac{p_0(1-x)^{p_1}(1+p_4 x+p_5 x^2+p_6 x^3)}{x^{p_2+p_3 log(x)}}$
8 Par	$\frac{d\sigma}{dm_{jj}} = \frac{p_0(1-x)^{p_1}(1+p_4 x+p_5 x^2+p_6 x^3+p_7 x^4)}{x^{p_2+p_3 log(x)}}$
UA2/ATLAS Family	
4 Par	$\frac{d\sigma}{dm_{jj}} = \frac{p_0 exp[-p_2 x-p_3 x^2]}{x^{p_1}}$
5 Par	$\frac{d\sigma}{dm_{jj}} = \frac{p_0 exp[-p_2 x-p_3 x^2-p_4 x^3]}{x^{p_1}}$
6 Par	$\frac{d\sigma}{dm_{jj}} = \frac{p_0 exp[-p_2 x-p_3 x^2-p_4 x^3-p_5 x^4]}{x^{p_1}}$
7 Par	$\frac{d\sigma}{dm_{jj}} = \frac{p_0 exp[-p_2 x-p_3 x^2-p_4 x^3-p_5 x^4-p_6 x^5]}{x^{p_1}}$
Modified Exponential Family	
3 Par	$\frac{d\sigma}{dm_{jj}} = p_0 exp[p_1 x^{p_2}]$
4 Par	$\frac{d\sigma}{dm_{jj}} = p_0 exp[p_1 x^{p_2} + p_1(1-x)^{p_3}]$
5 Par	$\frac{d\sigma}{dm_{jj}} = p_0 exp[p_1 x^{p_2} + p_3(1-x)^{p_4}]$

Table C.1. The different function families tested by the Fisher Test.

The fits for the so-called Standard Dijet Family are reported in figure C.1, for the Polynomial Extension family in figure C.2, for the UA2/ATLAS family in figure C.3, and for the Modified Exponential Family in figure C.4.

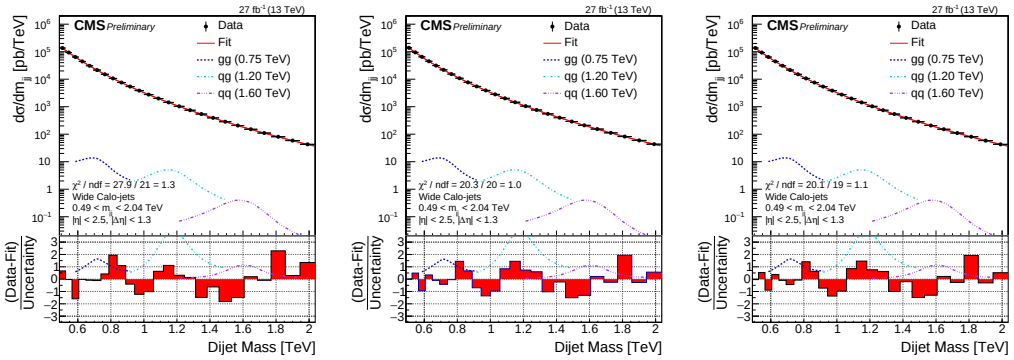

Figure C.1. Fit to the dijet mass spectrum (points) using the functions in the "Standard Dijet Family": (Left) 4 parameters function; (Middle) 5 parameters function; (Right) 6 parameters function.

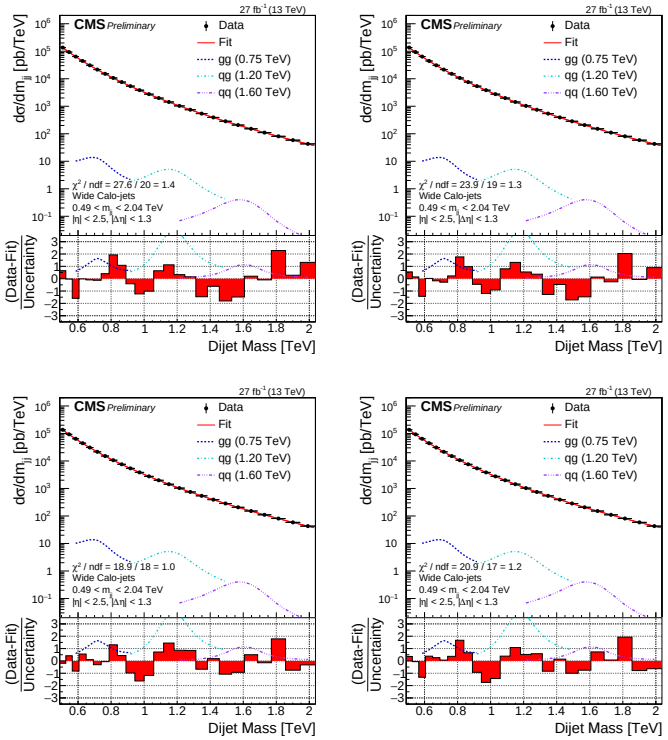

Figure C.2. Fit to the dijet mass spectrum (points) using the functions in the "Polynomial extension family": (Top left) 5 parameters function; (Top right) 6 parameters function; (Bottom left) 7 parameters function; (Bottom right) 8 parameters function.

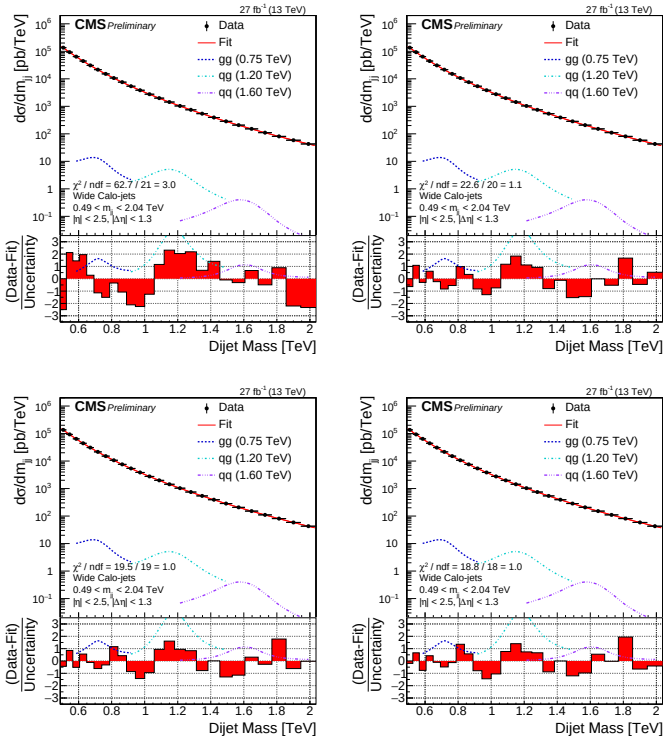

Figure C.3. Fit to the dijet mass spectrum (points) using the functions in the "UA2/ATLAS family": (Top left) 4 parameters function; (Top right) 5 parameters function; (Bottom left) 6 parameters function; (Bottom right) 7 parameters function.

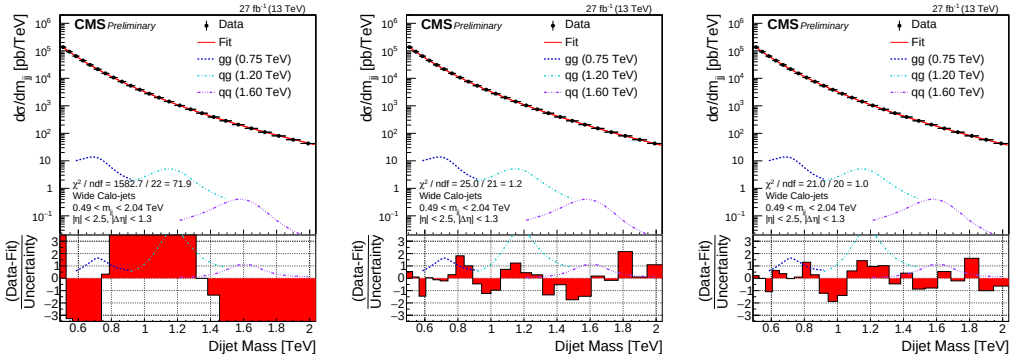

Figure C.4. Fit to the dijet mass spectrum (points) using the functions in the "Modified Exponential Family": (Left) 3 parameters function; (Middle) 4 parameters function; (Right) 5 parameters function.

In order to derive how many free parameters are effectively needed in the fit function a Fisher Test [46] can be performed. The procedure is described below.

- Given two models (for example f_1 and f_2), the F-value is given by:

$$F_{21} = \frac{\frac{RSS_1 - RSS_2}{n_2 - n_1}}{\frac{RSS_2}{N - n_2}}$$

  where N is the number of data points, n_1 and n_2 are the number of parameters for the models f_1 and f_2 (with $n_2 > n_1$), RSS_i is the residual sum of squares of model i:

$$\sum_{bins} (data_{bin} - fit_{bin})^2 \tag{C.1}$$

  skipping bins with 0 entries and fit_{bin} is 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 an F-distribution with ($n_2 - n_1$, $N - n_2$) degrees of freedom. The observed confidence level CL_{21} can be defined as:

$$CL_{21} = 1 - \int_{-inf}^{F_{21}} F - distribution(n_2 - n_1, N - n_2)$$

- The null hypothesis is rejected if CL_{21} is smaller than the desired probability α (set to 0.05).

The F-values for each family are reported in Table C.2 comparing the function with n parameters with the one with $n + 1$ parameters. As it can be observed from this table the best functions for the background parameterization are:

- the 5 parameters function for the Standard Dijet Family;
- the 7 parameters function for the Polynomial Extension Family;
- the 6 parameters function for the UA2/ATLAS Family;
- the 5 parameters function for the Modified Exponential Family;

This study justifies the choice to use the 5 parameters function from the Standard Dijet Family.

	F-value	CL
Standard Dijet Family		
4Par-5Par	21.2	0.0002
5Par-6Par	0.24	0.62
Polynomial extension Family		
5Par-6Par	4.1	0.057
6Par-7Par	13.49	0.001
7Par-8Par	-5.94	nan
UA2/ATLAS Family		
4Par-5Par	113.7	0
5Par-6Par	6.87	0.016
6Par-7Par	3.56	0.07
Modified Exponential Family		
3Par-4Par	3560	0
4Par-5Par	8.8	0.008

Table C.2. The F-value and the corresponding confidence level from the comparison between the function with n parameters with the one with $n + 1$ parameters for each family.

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", *Physical Review Letters*, vol. 19, no. 21, pp. 1264–1266, 1967.

[3] A. Salam, "Elementary particle physics: Relativistic groups and analyticity", in *Eighth Nobel Symposium. Stockholm: Almquvist and Wiksell*, p. 367, 1968.

[4] G. Dissertori, I. G. Knowles, and M. Schmelling, *Quantum chromodynamics: high energy experiments and theory*, vol. 115. Oxford University Press, 2003.

[5] L. A. Anchordoqui, H. Goldberg, D. Lust, S. Nawata, S. Stieberger, and T. R. Taylor, "Dijet signals for low mass strings at the large hadron collider", *Physical review letters*, vol. 101, no. 24, p. 241803, 2008.

[6] S. Cullen, M. Perelstein, and M. E. Peskin, "TeV strings and collider probes of large extra dimensions", *Physical Review D*, vol. 62, no. 5, p. 055012, 2000.

[7] U. Baur, I. Hinchliffe, and D. Zeppenfeld, "Excited quark production at hadron colliders", *International Journal of Modern Physics A*, vol. 2, no. 04, pp. 1285–1297, 1987.

[8] R. S. Chivukula, A. Farzinnia, E. H. Simmons, and R. Foadi, "Production of massive color-octet vector bosons at next-to-leading order", *Physical Review D*, vol. 85, no. 5, p. 054005, 2012.

[9] E. H. Simmons, "Coloron phenomenology", *Physical Review D*, vol. 55, no. 3, p. 1678, 1997.

[10] 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.

[11] 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.

[12] G. D'Ambrosio, G. Giudice, G. Isidori, and A. Strumia, "Minimal flavour violation: An effective field theory approach", *Nuclear Physics B*, vol. 645, no. 1, pp. 155–187, 2002.

[13] 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, p. 85, 2010.

[14] E. Eichten, I. Hinchliffe, K. Lane, and C. Quigg, "Supercollider physics", *Reviews of Modern Physics*, vol. 56, no. 4, p. 579, 1984.

[15] J. Lewin and P. Smith, "Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil", *Astroparticle Physics*, vol. 6, no. 1, pp. 87–112, 1996.

[16] A. M. Green, "Astrophysical uncertainties on direct detection experiments", *Modern Physics Letters A*, vol. 27, no. 03, p. 1230004, 2012.

[17] A. Boveia, O. Buchmueller, G. Busoni, F. D'Eramo, A. De Roeck, A. De Simone, C. Doglioni, M. J. Dolan, M.-H. Genest, K. Hahn, *et al.*, "Recommendations on presenting LHC searches for missing transverse energy signals using simplified s-channel models of dark matter", *arXiv preprint arXiv:1603.04156*, 2016.

[18] J. Abdallah, H. Araujo, A. Arbey, A. Ashkenazi, A. Belyaev, J. Berger, C. Boehm, A. Boveia, A. Brennan, J. Brooke, *et al.*, "Simplified models for dark matter searches at the LHC", *Physics of the Dark Universe*, vol. 9, pp. 8–23, 2015.

[19] A. De Simone and T. Jacques, "Simplified models vs. effective field theory approaches in dark matter searches", *The European Physical Journal C*, vol. 76, no. 7, pp. 1–27, 2016.

[20] L. Randall and R. Sundrum, "An alternative to compactification", *Physical Review Letters*, vol. 83, no. 23, p. 4690, 1999.

[21] TOTEM Collaboration, "The TOTEM experiment at the CERN Large Hadron Collider", *Journal of Instrumentation*, vol. 3, no. 08, p. S08007, 2008.

[22] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *Journal of Instrumentation*, vol. 3, no. 08, p. S08006, 2008.

[23] MoEDAL Collaboration, "The MoEDAL experiment at the LHC: status and results", *arXiv preprint arXiv:1703.07141*, 2017.

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

[25] CMS Collaboration, "The CMS tracker system project: Technical Design Report", tech. rep., CMS-TDR-005, 1997.

[26] CMS Collaboration, "The CMS Electromagnetic Calorimeter Technical Design Report", 1997.

[27] CMS Collaboration, "CMS: The hadron calorimeter technical design report", *CERN/LHCC*, pp. 97–31, 1997.

[28] CMS Collaboration, "The muon project technical design report", *CERN/LHCC*, vol. 32, no. 404, p. 1997, 1997.

[29] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and collider physics*. Cambridge university press, 2003.

[30] C. Grupen and B. Shwartz, *Particle detectors*, vol. 26. Cambridge university press, 2008.

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

[32] CMS Collaboration, "Particle-flow event reconstruction in CMS and performance for jets, taus and MET", 2009.

[33] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *Journal of Instrumentation*, vol. 6, no. 11, p. P11002, 2011.

[34] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Physics Letters B*, vol. 659, no. 1, pp. 119–126, 2008.

[35] B. A. Dobrescu and F. Yu, "Coupling-mass mapping of dijet peak searches", *Physical Review D*, vol. 88, no. 3, p. 035021, 2013.

[36] CMS Collaboration, "Data parking and data scouting at the CMS experiment", *CMS Physics Analysis Summary CMS-DP-2012-022*, 2012.

[37] CMS Collaboration, "Search for narrow resonances in dijet final states at $\sqrt{s} =$ 8 TeV with the novel CMS technique of data scouting", *Physical review letters*, vol. 117, no. 3, p. 031802, 2016.

[38] T. Sjöstrand, S. Mrenna, and P. Skands, "A brief introduction to PYTHIA 8.1", *Computer Physics Communications*, vol. 178, no. 11, pp. 852–867, 2008.

[39] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *The European Physical Journal C*, vol. 76, no. 3, p. 155, 2016.

[40] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 tune", *The European Physical Journal C*, vol. 74, no. 8, p. 3024, 2014.

[41] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, *et al.*, "GEANT4 - a simulation toolkit", *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003.

[42] J. Bijnens, P. Eerola, M. Maul, A. Månsson, and T. Sjöstrand, "QCD signatures of narrow graviton resonances in hadron colliders", *Physics Letters B*, vol. 503, no. 3, pp. 341–348, 2001.

[43] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *Journal of High Energy Physics*, vol. 2007, no. 11, p. 070, 2007.

[44] CMS Collaboration, "Search for narrow resonances decaying to dijets in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Physical review letters*, vol. 116, no. 7, p. 071801, 2016.

[45] CMS Collaboration, "Search for resonances and quantum black holes using dijet mass spectra in proton-proton collisions at $\sqrt{s} = 8$ TeV", *Physical Review D*, vol. 91, no. 5, p. 052009, 2015.

[46] R. G. Lomax and D. L. Hahs-Vaughn, *Statistical concepts: A second course*. Routledge, 2013.

[47] T. Junk, "Confidence level computation for combining searches with small statistics", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 434, no. 2-3, pp. 435–443, 1999.

[48] A. L. Read, "Presentation of search results: the CLs technique", *Journal of Physics G: Nuclear and Particle Physics*, vol. 28, no. 10, p. 2693, 2002.

[49] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *The European Physical Journal C*, vol. 71, no. 2, p. 1554, 2011.

[50] ATLAS Collaboration, "Procedure for the LHC Higgs boson search combination in summer 2011", tech. rep., ATL-PHYS-PUB-2011-011, 2011.

[51] E. Gross and O. Vitells, "Trial factors for the look elsewhere effect in high energy physics", *The European Physical Journal C-Particles and Fields*, vol. 70, no. 1, pp. 525–530, 2010.

[52] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST*, vol. 12, no. CERN-PH-EP-2015-305, p. P02014, 2016.

[53] CMS Collaboration, "CMS luminosity measurements for the 2016 data taking period", Tech. Rep. CMS-PAS-LUM-17-001, CERN, Geneva, 2017.

[54] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. Nadolsky, and W.-K. Tung, "New generation of parton distributions with uncertainties from global QCD analysis", *Journal of High Energy Physics*, vol. 2002, no. 07, p. 012, 2002.

[55] CMS Collaboration, "Search for dijet resonances in proton-proton collisions at and constraints on dark matter and other models", *Physics Letters B*, vol. 769, pp. 520–542, 2017.

[56] P. Dauncey, M. Kenzie, N. Wardle, and G. Davies, "Handling uncertainties in background shapes: the discrete profiling method", *Journal of Instrumentation*, vol. 10, no. 04, p. P04015, 2015.

[57] CMS Collaboration, "Search for dark matter produced with an energetic jet or a hadronically decaying W or Z boson at $\sqrt{s} = 13$ TeV", *Journal of High Energy Physics*, vol. 2017, no. 7, p. 14, 2017.

[58] CMS Collaboration, "Search for new physics in the monophoton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", tech. rep., 2017.

[59] CMS Collaboration, "Search for dark matter in Z+E_T^{miss} events using 12.9 fb^{-1} of 2016 data", tech. rep., CMS-PAS-EXO-16-038, 2016.

[60] CMS Collaboration, "Search for low mass vector resonances decaying to quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. Lett.*, vol. 119, no. 11, 2017.

[61] T. d. Pree, K. Hahn, P. Harris, and C. Roskas, "Cosmological constraints on dark matter models for collider searches", *arXiv preprint arXiv:1603.08525*, 2016.

[62] M. Backović, K. Kong, and M. McCaskey, "MadDM v. 1.0: Computation of dark matter relic abundance using MadGraph 5", *Physics of the Dark Universe*, vol. 5, pp. 18–28, 2014.

[63] P. A. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Banday, R. Barreiro, J. Bartlett, N. Bartolo, *et al.*, "Planck 2015 results. XIII. Cosmological parameters", *Astronomy & Astrophysics*, vol. 594, p. A13, 2016.

[64] E. Behnke, M. Besnier, P. Bhattacharjee, X. Dai, M. Das, A. Davour, F. Debris, N. Dhungana, J. Farine, M. Fines-Neuschild, *et al.*, "Final results of the PICASSO dark matter search experiment", *Astroparticle Physics*, vol. 90, pp. 85–92, 2017.

[65] C. Amole, M. Ardid, I. J. Arnquist, D. M. Asner, D. Baxter, E. Behnke, P. Bhattacharjee, H. Borsodi, M. Bou-Cabo, P. Campion, *et al.*, "Dark matter search results from the PICO-60 C_3F_8 Bubble Chamber", *Physical review letters*, vol. 118, no. 25, p. 251301, 2017.

[66] M. Aartsen, M. Ackermann, J. Adams, J. Aguilar, M. Ahlers, M. Ahrens, D. Altmann, K. Andeen, T. Anderson, I. Ansseau, *et al.*, "Search for annihilating dark matter in the *sun* with 3 years of IceCube data", *The European Physical Journal C*, vol. 77, no. 3, p. 146, 2017.

[67] M. Aartsen, K. Abraham, M. Ackermann, J. Adams, J. Aguilar, M. Ahlers, M. Ahrens, D. Altmann, T. Anderson, I. Ansseau, *et al.*, "Improved limits on dark matter annihilation in the Sun with the 79-string IceCube detector and implications for supersymmetry", *Journal of Cosmology and Astroparticle Physics*, vol. 2016, no. 04, p. 022, 2016.

[68] K. Choi, K. Abe, Y. Haga, Y. Hayato, K. Iyogi, J. Kameda, Y. Kishimoto, M. Miura, S. Moriyama, M. Nakahata, *et al.*, "Search for neutrinos from annihilation of captured low-mass dark matter particles in the Sun by Super-Kamiokande", *Physical review letters*, vol. 114, no. 14, p. 141301, 2015.

[69] D. Akerib, S. Alsum, H. Araújo, X. Bai, A. Bailey, J. Balajthy, P. Beltrame, E. Bernard, A. Bernstein, T. Biesiadzinski, *et al.*, "Results from a search for dark matter in LUX with 332 live days of exposure", *arXiv preprint arXiv:1608.07648*, 2016.

[70] A. Tan, M. Xiao, X. Cui, X. Chen, Y. Chen, D. Fang, C. Fu, K. Giboni, F. Giuliani, H. Gong, *et al.*, "Dark matter results from first 98.7 days of data from the PandaX-II experiment", *Physical review letters*, vol. 117, no. 12, p. 121303, 2016.

[71] R. Agnese, A. Anderson, T. Aramaki, M. Asai, W. Baker, D. Balakishiyeva, D. Barker, R. B. Thakur, D. Bauer, J. Billard, *et al.*, "New results from the search for low-mass weakly interacting massive particles with the CDMS low ionization threshold experiment", *Physical review letters*, vol. 116, no. 7, p. 071301, 2016.

[72] G. Angloher, A. Bento, C. Bucci, L. Canonica, X. Defay, A. Erb, F. von Feilitzsch, N. F. Iachellini, P. Gorla, A. Gütlein, *et al.*, "Results on light dark matter particles with a low-threshold CRESST-II detector", *The European Physical Journal C*, vol. 76, no. 1, p. 25, 2016.

Ringraziamenti

Nella mia tesi magistrale scrivevo che "il futuro non è cosi spaventoso. Può essere quello che vuoi tu". Con questa idea voglio affrontare la fine di questo arduo percorso che è il dottorato. Percorso che però non ho compiuto da solo, in molti mi hanno accompagnato più o meno consapevolmente e a questi vanno i miei ringraziamenti.

Il primo grazie va alla mia relatrice Chiara per la pazienza e gli sforzi nelle correzioni di questa tesi, l'aiuto nel lavoro e le chiacchiere informali sul futuro. Vorrei ringraziare poi il mio secondo "unofficial" supervisor Francesco per avermi seguito e consigliato in molti step dell'analisi. Grazie a tutti i "senior" del gruppo di CMS Roma da cui ho imparato molto, sulla fisica e non.

Ringrazio l'INFN per avermi permesso di lavorare in un grande esperimento come CMS e per avermi dato l'opportunità di vivere a Ginevra nell'ultimo anno di dottorato. È stata un'esperienza fondamentale nella mia vita e motivo di una crescita personale e lavorativa.

Un grazie va poi ai miei genitori, le mie sorelle e mio cognato per avermi sostenuto, aver creduto in me e avermi fatto compagnia durante quest'anno a Ginevra.

Grazie a Miriam per i consigli "medici" e a Claudio per essersi fatto N ore di macchina nei viaggi Roma-Ginevra (A/R!). Insieme a loro ringrazio Stefano e Zia Marina (e Dodi e i mesoni) per le cene in compagnia.

Grazie ai miei amici di sempre che da parecchi anni sono un'altra famiglia per me: a Claudia, Dario, Jessica, Marco e Giada che mi hanno visto crescere ed erano presenti a tutti gli altri traguardi raggiunti.

Grazie a loro sono entrati nella mia "famiglia allargata" anche altre persone che voglio ringraziare semplicemente perché sono come sono: a Massi e Sabrina perché sono due persone solari e al mio per-sempre collega 244 per la sua semplicitá e perché tutto a questo mondo costa 5 euro adesso.

Grazie a MariaRoberta, amica-rivelazione dell'anno, a cui faccio un in bocca al lupo visto che anche lei sta finendo il suo percorso.

Ringrazio i Pippons per avermi accolto tra di loro, per le uscite, i barbecue, le cene insieme e le arrampicate. Un ringraziamento speciale va a Simone che più di tutti ha condiviso con me le difficoltà del Ph.D (ma anche le gioie come Corfù), alla "Laureanda" Chiara per le chiacchierate sceme su Skype (team-Fede sta vincendo!) e a Giulia che mi ha aiutato molto con l'analisi.

Voglio ringraziare anche i miei colleghi-amici che hanno cominciato tutto ciò insieme a me anni fa e che ancora vedo/sento nonostante ognuno abbia preso la sua strada. Grazie a Giorgio e Natalia perché è sempre un piacere vederli e brindare con loro, a Donny per cui provo un affetto smisurato anche a distanza e a Federico perché certi compagni di laboratorio non si scordano mai.

Infine voglio dire il grazie più grande alla persona a cui è dedicata questa tesi. Grazie mille a Daniela che è la mia socia in tutto e lo è stata anche in questo. In questi 3 anni di dottorato ha sempre creduto in me più di quanto abbia fatto io, si è sorbita tutti i racconti folli di questo folle lavoro, tutti i nervosismi (inevitabili e a volte necessari) e tutti i miei viaggi con periodi di assenza più o meno lunghi fino al mio anno all'estero. Grazie perché se questa tesi esiste è anche merito suo.