FERMILAB-THESIS-2020-19

Search for resonance in inclusive and b-tagged dijet mass spectra in proton-proton collision at $\sqrt{s} = 13$ TeV and constraints on dark matter and other models

by

Zhixing Wang,

A Dissertation

In

Physics

Submitted to the Graduate Faculty of
Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

DOCTOR OF PHILOSOPHY

Approved

Sung-Won Lee
Chair of Committee

Robert Harris

Shuichi Kunori

Nural Akchurin

Junyang Huang

Mark A. Sheridan
Dean of the Graduate School

December, 2020

ACKNOWLEDGMENTS

Foremost, I want to express my sincere gratitude to Prof. Sung-Won Lee, who guided my Ph.D. study and research with his patience, enthusiasm, motivation, and immense knowledge. Dr. Lee distinguishes himself as an energetic and driven person who contributes both to the physics and astronomy department and the High Energy Physics (HEP) group at Texas Tech University (TTU). The thing I learned from him is not only about particle physics, but also the way of treating my life of being optimistic, energetic, and always be wise.

The members of my committee have my utmost gratitude and respect for their support and guidance: Dr. Nural Akchurin, Dr. Shuichi Kunori, Dr. Robert Harris, Dr. Junyang Huang, and the aforementioned Dr. Lee. for their encouragement, insightful comments. They help me in all directions.

I very much appreciate the HEP group as a whole. These projects I worked on with CMS(Compact Muon Solenoid) bring me a lot of opportunities and is advantageous to my career aspirations. I had a lot of great opportunities to attend and give a talk at conferences. I can also keep my mind sharp and allow me to have the chance to work in the world's top research facility, where I was able to work with the world's best scientists from who I can learn so much.

My sincere thanks also go to Dr. Federico De Guio whose help cannot be overestimated, for the stimulating, inspiring discussions, for the sleepless nights we were working together before deadlines. All these works can't be accomplished without his guidance, patience, and assistance. Even though I know I might have pissed him off sometimes, but I'm happy we archived the paper and all the work.

I also want to thank Alberto Zucchetta and Manuel Sommerhalder, who work with us on publishing the paper. They are all cool co-workers. I would like to extend my thanks to the people in the Dijet group as well, I get so many help and deep thought and patient instruction from you.

I would like to extend my sincere thanks to my colleagues and also my friends, Kamal Lamichhane, Tielige Menkge, Samila Muthumuni, Dr. Vinay Hegde, Dr. Andrew Whitbeck, and Dr. Jordan Damgov. They have been kind and helpful to me all the time in these years. I can always count on your hands when troubles pop up. In particular, I am grateful to Dr. James Faulkner for enlightening me on the first glance of particle physics research. It has been a pleasure to work with you.

Additionally, I would like to thank my friend and my family, some of who might be thousands of miles away from me. Your relentless support and encouragement, even just a small talk or a dinner together outside, have pulled me up from depression and low mood countless times. Thanks, I love you all.

TABLE OF CONTENTS

ABSTRACT

Two analyses searching for new heavy resonances and interpreting the result with dark matter are presented in this dissertation. The first analysis searches for resonances decaying into dijet final states using the proton-proton collision data at center mass of $\sqrt{s} = 13$ TeV with the CMS detector at Large Hadron Collider in 2016, corresponding to a total integrated luminosity of 36 fb^{-1}. No significant evidence for the production of new particles is observed. A simplified model of interactions between quarks and dark matter is used to interpret the result. The second analysis searches for resonances decaying into dijet final states using the proton-proton collision data at center mass of $\sqrt{s} = 13$ TeV with the CMS detector at Large Hadron Collider in 2016, 2017, and 2018, corresponding to a total integrated luminosity of 137.5 fb^{-1}. A b-tagger named DeepJet is used to examine the events with jets containing B hadrons. No evidence of new heavy resonances is observed, as well. The 95% confidence level cross-section upper limit is set on excited b-quark b^* and Coloron models. Limits are also set on a simplified model of interactions between b-quarks and dark matter.

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Time has passed around 100 years since the first subatomic particle, electron, was discovered by J.J. Thomson [28] in 1897. After the discovery of a 125 GeV new particle was announced by both CMS Collaboration [29] and ATLAS Collaboration [30], in 2012, which is the final piece of the Standard Model (SM) of particle physics, new particles had been expected by physicists to lie over the horizon. Higgs boson obeyed every rule theorists predicted. That is both good and bad news. The good news is that it proves we are walking in the right direction. The bad news is that if it had acted just slightly differently it would raise many questions, but now it raised few questions and gave us no new clue about where to go.

So what is next for particle physics? Just like what Lord Kelvin said a hundred years ago, "The beauty and clearness of the dynamical theory, which asserts heat and light to be modes of motion, is at present obscured by two clouds." [31], even though the Standard Model has been highly successful at describing the behavior of the elementary particles of nature and the forces that act on these particles, there is still much-unrevealed mysteriousness. For example, even though neutrinos have all been discovered, however in the SM, they should all be massless, but neutrino oscillation experiments have shown that neutrinos do have mass, with more new theoretical problems have been raised for the mass terms for neutrinos. In addition, Dark matter (DM) has attracted many people's interests because it's still an enigma, despite extensive astronomical confirmation of its existence [32; 33; 34] and showing that 26% of the observed universe should be dark matter. To explain the deficiencies of the Standard Model, many models have been raised up. The theoretical developments that try to explain the deficiencies of the Standard Model are called physics beyond the Standard Model (BSM). People are trying all they can with all reasonable models we currently have, that's what we call "searching for every corner of the room." The high energy, high luminosity collider is our most powerful weapon in the "war" against the unknown.

Many BSM models predict the existence of new states coupling to quarks and gluons, which would manifest themselves as resonances in the dijet mass spectrum.

Dijet spectrum is the distribution of the invariant mass of the pair of jets having the largest values of p_T, which is expected to decrease smoothly with increasing mass, within the framework of the standard model. However, if BSM models exist, the signal of them will appear either as a resonant peak in the dijet invariant mass spectrum or through a peculiar shape of the two jets angular correlation. As a result, the dijet search strategy is to distinguish the peculiar signature from an overwhelming background. This kind of search can aim at detecting a signal from q*, W$'$, Z$'$, Quantum Black Holes, or Randall-Sundrum graviton benchmark models.

Some special algorithms can be used to increase the sensitivity of the dijet search, and b-tagging is one of them. B-tagging is a method to help identify jets that originate from one special quark name b quark. B-tagging is a very important method not only in the dijet search but many other searches because some important high-mass particles (both recently discovered and hypothetical) decay into bottom quarks, such as the Higgs boson is more likely to decay into b quark than any other particle. If the identifying of b quark can be done more accurately, then it will help find the events like the decay of the Higgs boson. In the dijet search, with the help of b-tagging, the sensitivity of models that can decay into b quark can also be improved, by applying b-tagging to one or both of the jets.

Among all BSM models, dark matter is probably the most popular topic. There are three classes of searching strategy to search for dark matter: (i) direct detection in shielded underground detectors; (ii) indirect detection with satellites, balloons, and ground-based telescopes looking for signals of DM annihilation; (iii) particle colliders aiming at direct DM production. The method of searching for resonance in the dijet final state can also be used to search for them, which belongs to the collider search. Even though we haven't observed any interaction between dark matter and standard model particles, but some models predict dark matter might be produced in some heavy mediators decay and these mediators can be produced by standard model particles in high energy. The mediator can be either a known particle, such as the Z boson or the Higgs boson, or an unknown particle. Simplified Model introduced by Dark Matter Working Group (DMWG) is one of them. It offers people a benchmark to compare with not only other searches at the LHC but also other direct and indirect detect experiment.

The simplified model allows the mediator to decay to the quark-quark final state, which means it can be b and anti-b quark in the final state. As we just mentioned, the b-tagging algorithm can significantly increase the sensitivity to those searches with b quark in the final state. It's more interesting to apply the b-tagging algorithm to the b-jet from the simplified model mediator decay.

CHAPTER 2

THEORETICAL OVERVIEW

Search for new physics is highly depends on the understanding of all observable phenomenon that is known as background, which can be removed to separate the excess signals. The SM has done a good job of providing predictions for the most observable background in particle physics, though with an exception like DM and neutrino. Still, many analyses spanning several decades have continued to probe and further corroborate the SM picture. This dissertation will focus first on the searches for BSM particles, and then the interpretation of the result with DM.

2.1 The Standard Model

The standard Model (SM) is the name given in the 1970s, for a theory of fundamental particles and how they interact. For now, there are 17 named particles in the SM, which is shown in Fig 2.1 It addresses three of the four known forces in nature: electromagnetic (EM) force, the weak nuclear force, and the strong nuclear

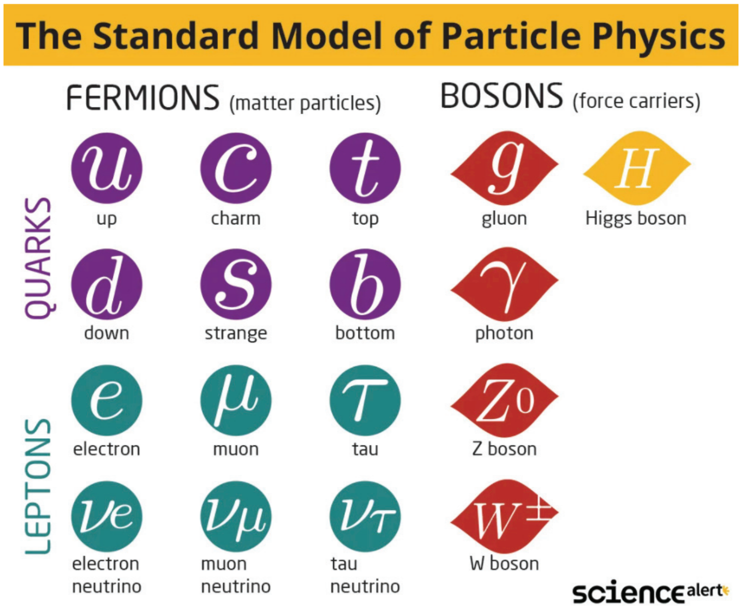


Figure 2.1: The SM categorizes fundamental particles into related groups [1]

force, excluding gravity, in terms of the properties of a limited number of elementary particles. Back in 1954, Chen Ning Yang and Robert Mills extended the concept of gauge theory [35], which explains strong interaction. Seven years later, Sheldon Glashow combined EM and the weak interactions [60]. Later in 1967, the Higgs mechanism is incorporated by Steven Weinberg and Abdus Salam into Glashow's electroweak interaction [36; 37; 38; 39], giving it its modern form.

2.1.1 Particles

All matter in the world is made of elementary particles. these particles have two basic types named quarks and leptons. Each group has 6 particles, which are related in pairs, or generations. The lightest is the most stable particles, they form the first generation. This pair has up and down quarks. The next-generation class of particles is less stable but heavier. This pair has charm and strange quarks. The third generation is even heavier and not stable at all, which are bottom and top quarks. For the leptons, the electron and electron neutrino are the first generation, the muon and muon neutrino are the next generation. Tau and the tau neutrino belong to the last generation.

These 12 particles are all fermions and have spin 1 or 1/2. According to spin-statistics theorem, fermions respect the Pauli exclusion principle. Each fermion has a corresponding antiparticle. The defining property of quarks is they can carry color charge and can interact in the strong interaction. This phenomenon makes quarks strongly bound to another, forming color-neutral composite particles called hadrons. Hadrons contain multiple quarks and/or anti-quarks, either two or three. We name two quarks hadron as mesons and three quarks as baryons. The most popular hadron is the lightest baryons named proton and neutron. Meanwhile, quarks carry electric charge and weak isospin, which allows them to interact with other fermions via EM and the weak interaction. The particle carries no color-charge are name leptons. Three of them are neutrinos that don't carry electric charges as well, which is only directly influenced by the weak nuclear force.

Except for the 12 fermions, gauge bosons are defined as force carriers that mediate the strong, weak, and EM fundamental interactions. At a macroscopic level, EM allows particles to attract or repel one another via electric and magnetic fields, and gravitation allows particles with mass to attract another one with mass. The SM explains that these forces are the result of matter particles exchanging other particles, which are force mediating particles. The gauge boson particles have spin of 1, as the result, they don't obey the Pauli exclusion principle. In a summary, Photons mediate the EM force between electrically charged particles. W^+, W^-, and Z gauge bosons mediate the weak interactions between particles of different flavors. $W^\pm$ can only act in left-handed particles and right-handed antiparticles. Z boson interacts with both left-handed particles and antiparticles. The eight gluons mediate the strong interactions between color charged particles. They are massless, which are labeled by a combination of color and anticolor charge. They can also interact with themselves, which is described in Quantum Chromodynamics (QCD).

The Higgs boson is a massive scalar elementary particle theorized by Peter Higgs [40; 41] Robert Brout, and Francois Englert [42] in 1964. It plays a very unique role in the SM by explaining why other particles, except neutrinos and gluons, have mass. Elementary-particle masses, and the differences between EM (mediated by the photon) and the weak force (mediated by the W and Z bosons), are critical to many aspects of the structure of microscopic (and hence macroscopic) matter. Because the Higgs boson is very massive and can decay very quickly, only the very high energy accelerator can observe and record it. On 4 July 2012, both CMS and ATLAS reported the new particle with a mass of 125 GeV/c^2, is consistent with the theorized particle named Higgs boson. In the next year, it was confirmed as the Higgs boson.

2.1.2 Interactions

The gauge bosons are force carriers, as mentioned in the previous section. They are 8 gluons, $W^\pm$, and Z. The fundamental force interact strength is quantized by a coupling constant α. It's different in each interaction and its magnitude affects physical observables, for example, cross-section. To be precise, weak interaction has $\alpha_{week} \approx 10^{-6}$, EM interactions has $\alpha_{EM} \approx 1/137$, and $\alpha_s \approx 1$. Though SM doesn't include gravity force, it has $\alpha_g \approx 10^{-39}$. The three fundamental interactions, along with a combination interaction between EM and weak interaction, will be discussed in this part.

The strong force, as its name shows, is a very strong force that acts between subatomic particles. It was named as color, which is analogous to electric charge, since both of them can't be visually seen in nature, while electric charge is the source of EM, and color is the source of the strong force. It binds quarks together to form baryons and mesons, such as proton and neutron. Gluons, as its name indicates, exchange in the process of the strong interaction, and 'glue' quarks together. The strong force is a very short-range interaction, which is about 10^{-15} meter, roughly the same as the diameter of proton and neutron. Unlike other forces, strong forces become stronger with distances. As the distance increases, the force increases as the tension does in a piece of elastic as its two ends separate. Confinement is the phenomenon for the combination of quarks. The strong interaction is 100 times stronger than the EM force. But in the even smaller distance, the strong force will become even weaker and act like independent particles.

The electromagnetic force, explains how both moving and stationary electric charged particles interact. It combines electric force and magnetic force because they are considered to be a different force. It has an infinite influence range so it can affect everything in the universe. EM force is also the most common force which is responsible for events beyond the nuclear scale in daily life. The electromagnetic force can be generated by three types of fields known as the electrostatic field, magnetostatics field, and the electromagnetic field. The electromagnetic force is carried by electromagnetic fields composed of electric fields and magnetic fields, and it is responsible for electromagnetic radiation such as light.

The weak interaction is not the weakest force in the universe. It gets its name because it's the weakest force in the SM. It is like the strong force which is only effective at very short distances. Weak interaction acts on the subatomic level and plays a crucial role in powering stars and creating elements. The weak force is carried by the W and Z bosons. W bosons are charged particles and designated by their symbols. By emitting a W boson, the weak force changes the flavor of a quark. This's one of the key steps that make sun burn. The Z boson is an electrically neutral particle and carries a weak neutral current. Weak interactions also appear in events in daily life, such as beta decay, electron capture, and Nuclear fusion.

However, electromagnetism and the weak interaction are modeled as two different aspects of the same force. Above certain energy, 246 GeV, they would merge into one force. In the early stage of the universe, when it's very hot, the electromagnetic force and weak force merge into a combined electroweak force. It broke into two forces during the quark epoch. Mathematically EM and weak interaction are unified via a Yang-Mills field with a SU(2) $\otimes$ U(1) gauge group. These fields are weak isospin fields W_1, W_2, and W_3, as well as the weak hypercharge field B. This invariance is known as electroweak symmetry. In the SM, the $W^{\pm}$ and Z and the photon are produced through the spontaneous symmetry breaking of the electroweak symmetry SU(2) $\otimes$ $U(1)_\gamma$ to $U(1)_{EM}$, effected by the Higgs mechanism.

2.1.3 QCD

Quantum Chromodynamics(QCD) is the gauge theory to explain the strong interaction, with symmetry group SU(3), which makes QCD a non-Abelian theory.

The Lagrangian is written as

$$\mathcal{L}_{QCD} = \sum_i \bar{q}_{i,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t^A_{ab} G^A_{\mu\nu} - m_i \delta_{ab}) q_{i,b} - \frac{1}{4} F^A_{\mu\nu} F^{\mu\nu A} \tag{2.1}$$

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

$$F^A_{\mu\nu} = \partial_\mu G^A_\mu - \partial_\nu G^A_\nu - g_s f_{ABC} G^B_\mu G^C_\nu \tag{2.2}$$

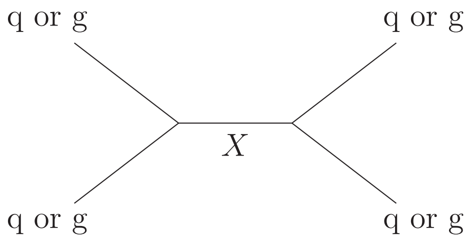
The structure constants f_{ABC} satisfy the relation:

$$[t^A, t^B] = i f_{ABC} t^C \tag{2.3}$$

The non-Abelian nature of QCD has two remarkable features: confinement and asymptotic freedom. As a result of the confinement, only color-singlet states can be directly observed, which means that gluons and quarks cannot be seen free and hard-scattering of quarks and gluons can be described in a perturbative way.

2.2 Beyond Standard Model

Many interesting BSM models have possible signals appearing either as a resonant peak in the dijet invariant mass spectrum, where dijet event is a collision between subatomic particles that produces two particle jets. The following diagram show dijet resonance in the s-channel. The initial state and final states contain two partons (quarks, anti-quarks, or gluons) and the intermediate state contains a resonance X.



In this section, several models that decay to dijet final state will be described. They are q*, Coloron, simplified dark matter model, and other models. However, q* will only have one quark and one gluon in the final state, Coloron and simplified dark matter model will have two quarks in the final state because this analysis is only focused on the above models. Other minor models will just be described briefly.

2.2.1 Excited quark (q*)

Excited states of composite quarks [2] are strongly produced giving large cross-sections (qg → q*). Excited quark is the most common benchmark model in resonances search in dijet final state, while the dominant decay model is qg. It's very popular in recent decades, because of their potential of explaining the quark-lepton generation structure and the observed pattern of fermion masses and mixing angles. If we discover excited quarks, then it'll be convincing evidence of the substructure of quarks and leptons. The mass of excited fermions are expected to be at least a few hundred GeV because the substructure scale Λ can't be too much smaller than 1 TeV, and the excited state should not be much lighter than Λ.

The mass of excited quarks arises prior to $SU(2) \otimes U(1)$ breaking, member of excited weak doublet should be almost degenerate in mass. The effective Lagrangians [43; 44; 45] is given by :

Table 2.1: Decay branching ratio (BR) of excited u and d quark

Decay Mode	BR [%]	Decay Mode	BR [%]
$u^* \to ug$	83.4	$d^* \to dg$	83.4
$u^* \to u\gamma$	2.2	$u^* \to d\gamma$	0.5

$$\mathcal{L}_{eff} = \frac{1}{2M^*}\overline{q^*}\sigma^{\mu\nu}[g_s f_s \frac{\lambda^a}{2}F^a_{\mu\nu} + gf\frac{\overrightarrow{\tau}}{2}\overrightarrow{W}_{\mu\nu} + g'f'\frac{Y}{2}B_{\mu\nu}]q_L + h.c.. \tag{2.4}$$

Here the q* and q$_L$ denote the isospin coublets of excited and lefthanded ground state quarks, $V_{\mu\nu}, V = F^a, \overrightarrow{W}, B$ is the field strength tensor for the gluon the SU(2) and the U(1) gauge fields. Y is 1/3, which is the weak hypercharge. In the end, g_s, g and g' are the gauge coupling constants, and f_s, f, and f' are free parameters determined by the composite dynamics.

The rate for q* model can be calculated from the Eq. 2.4. Assuming $M^* > m^*_{W,Z}$ and neglecting ordinary quark masses [44; 45], we get:

$$\Gamma(q^* \to gq) = \frac{1}{3}\alpha_s f_s^2 M^* \tag{2.5}$$

$$\Gamma(q^* \to \gamma q) = \frac{1}{4}\alpha f_\gamma^2 M^* \tag{2.6}$$

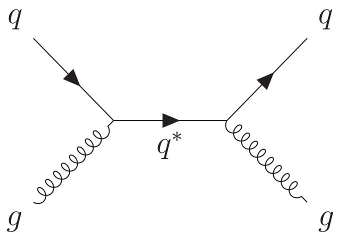
Here

$$f_\gamma = fT_3 + f'\frac{Y}{2} \tag{2.7}$$

According to Eq. 2.5, an excited quark will decay predominantly via strong interactions into ordinary a quark and a gluon. Radiative transitions and decays into quarks and a weak boson will also appear at $O(\alpha/\alpha_s)$, i.e. at the few % levels. As an example, the decay branching ratio (BR) of excited u and d quark is shown in Table. 2.1

In hadronic collision, excited quarks can be produced either pairwise or singly. Pairwise production can through $q\bar{q}$ and gluon fusion, but the rate is quite small.

On the other hand, the rate of single production of q^* via quark-gluon fusion can be highlighted at high energy provided that f_s is not small. The subsequent decay into quark and gluon leads to a peak in the dijet final state at $m=M^*$, as shown in the Feynman diagram below. With not overwhelming background, this is pretty clean and simply signal for q^*. The single excited quark production cross-section in jet-jet (solid line) are shown in Fig. 2.2. The number around the line show the $\sqrt{s}$ value in TeV



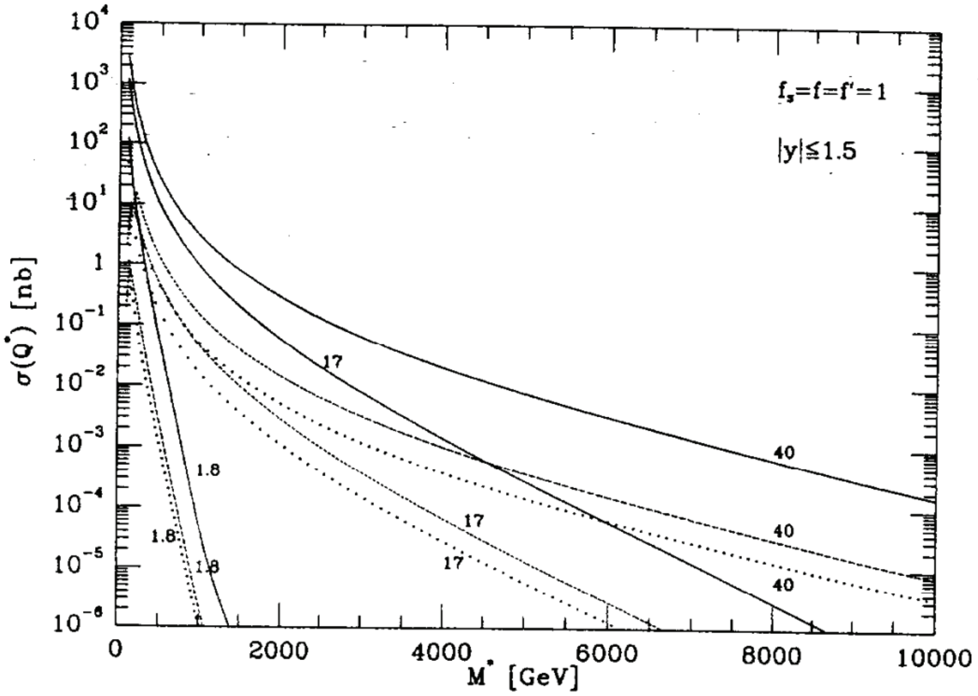
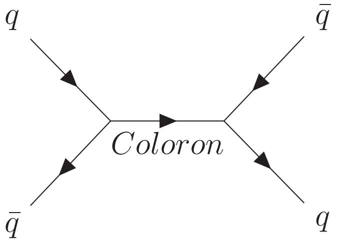


Figure 2.2: The single excited quark production cross-section in jet-jet (solid line). The number around the line show the $\sqrt{s}$ value in TeV [2]

In this analysis, we use the canonical model in which the compositeness scale Λ is equal to the resonance mass, all couplings are equal to them in the SM, and we also consider b^* and their decay to bg in the final state.

2.2.2 Coloron

The result from CDF [46] indicates that the inclusive cross-section for jets with $E_T > 200$ GeV is significantly higher than that predicted by QCD. This excess can be explained by a phenomenological model of quark substructure, or a model containing a new strong-coupled Z' gauge boson [47; 48], which Coloron is a simply flavor-universal variant of the coloron model of Hill and Parke [49]. Colorons are from additional color interaction and can be strongly produced, but as it requires an antiquark in the initial state, as shown in the following Feynman diagram, it will slightly reduce the cross-section compared to excited. [50].



As Hill and Parke [49], the QCD gauge ground is extended to $\mathrm{SU}(3)_1 \bigotimes US(3)_2$, with gauge couplings ξ_1 and ξ_2, with $\xi_1 \leq \xi_2$. In this coloron model, all quarks are assigned to triplet representations of the strong $\mathrm{SU}(3)_2$ group. The symmetry is broken to its diagonal subgroup at a scale f by introducing a Φ which transforms as a $(3, \bar{3})$ under the two SU(3) groups. The overall constant is adjusted so that the minimum of U is 0. The potential can be calculated as:

$$U(\Phi) = \lambda_1 Tr(\Phi\Phi^\dagger - f^2 I)^2 + \lambda_2 Tr(\Phi\Phi^\dagger - \frac{I}{3}(Tr\Phi\Phi^\dagger))^2 \tag{2.8}$$

The coloron-exchange may be approximated by the effective four-fermion interaction:

$$\mathcal{L}_{eff} = -\frac{g_3^2 cot^2\theta}{2!M_C^2} J_\mu^a J^{\mu a} \tag{2.9}$$

2.2.3 Simplified Dark Matter

There are 3 ways of searching for dark matter: direct detection, indirect detection, and indirect search in the collider. Utilizing indirect search for dark matter in LHC, the minimum evidence of dark matter production has an excess event with a single final state object X recoiling against large amounts of missing transverse momentum or energy (MET). As illustrated in Fig. 2.3, we can construct a large number of qualitatively different dark matter models. At the end of the dark matter forum's activity, a formal LHC Dark Matter Working Group (LHCDMWG) was created. LHCDMWG recommends s-channel models [17] and how to compare these results to Direct (DD) and indirect detection (ID) experiments.

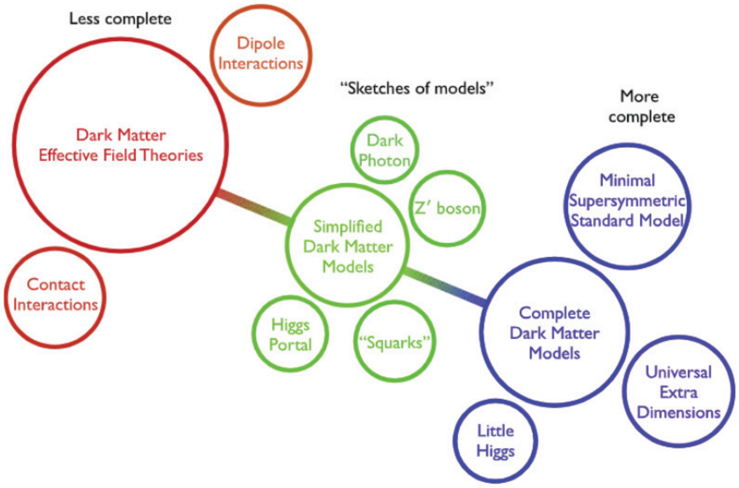


Figure 2.3: Artistic view of the dark matter theory space.

In this analysis, we only use leptophobic vector and axial-vector models, which are with a spin-1 mediator Z', and have the following Lagrangians:

$$\mathcal{L}_{vector} = -g_{DM} Z'_{\mu}\bar{\chi}\gamma^{\mu}\chi - g_q \sum_{q=u,d,s,c,b,t} Z'_{\mu}\bar{q}\gamma^{\mu}q \tag{2.10}$$

$$\mathcal{L}_{axial-vector} = -g_{DM} Z'_{\mu}\bar{\chi}\gamma^{\mu}\gamma_5\chi - g_q \sum_{q=u,d,s,c,b,t} Z'_{\mu}\bar{q}\gamma^{\mu}\gamma_5 q \tag{2.11}$$

Note that the universality of the coupling g_q guarantees that the above spin-1 simplified models are minimal flavor violating (MFV) [51]. We only want to use dark matter models whose dark matter mediator is a Dirac fermion χ, and the mediator mediating the interaction is exchanged in s-channel. As the assumption, each model is characterized by four parameters: the dark matter mass M$_{DM}$, the mediator mass M$_{med}$, the universal mediator coupling to quarks g$_q$, and the mediator coupling to dark matter g$_{DM}$. Mediator χ has zero couplings to lepton to avoid the stringent LHC bounds from di-lepton searches.

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

$$\Gamma^{\chi\bar{\chi}}_{vector} = \frac{g^2_{DM} M_{med}}{12\pi}(1 - 4z_{DM})^{1/2}(1 + 2z_{DM}) \tag{2.12}$$

$$\Gamma^{q\bar{q}}_{vector} = \frac{g^2_q M_{med}}{4\pi}(1 - 4z_q)^{1/2}(1 + 2z_q) \tag{2.13}$$

where $z_{DM,q} = m^2_{DM,q}/M^2_{med}$ and strictly restrict $M_{med} < 2m_{DM,q}$, The corresponding expressions for the axial-vector mediator are

$$\Gamma^{\chi\bar{\chi}}_{axial-vector} = \frac{g^2_{DM} M_{med}}{12\pi}(1 - 4z_{DM})^{3/2} \tag{2.14}$$

$$\Gamma^{q\bar{q}}_{axial-vector} = \frac{g^2_q M_{med}}{4\pi}(1 - 4z_q)^{3/2} \tag{2.15}$$

For now, it's recommended that mono-jet-like searches produce limits for a single choice of couplings. the values of the coupling to produce the limit on signal strengths are:

Vector mediator : $g_{DM} = 1$ and $g_q = 0.25$

Axial-vector mediator : $g_{DM} = 1$ and $g_q = 0.25$

The quark couplings g_q should be universal in all cases and the minimum width should be set to all the mediators so that the mediator has no couplings to other particles except quarks and dark matter. This choice provides consistent comparison across collider results, which ensure that the mediator has $\Gamma_{med}/M_{med} \lessapprox 10\%$ and is far from the strong coupling regime. It can furthermore be motivated by the requirement to avoid dijet constraints from the LHC and earlier hadron colliders.

2.2.4 Other Models

1. Diquarks [14] from superstring-inspired E_6 grand unified models are produced with electromagnetic coupling from the valence quarks of the proton (ud $\rightarrow D$). The cross-section for E_6 diquarks is large despite the relatively weak-coupling because of the large parton distribution function 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.

2. Heavy W and Z bosons [15] (called henceforth W' and Z') inspired by left-right symmetric grand unified models, have electroweak couplings and require antiquarks for their production($q_1\bar{q}_2 \rightarrow W'$, $q_1\bar{q}_2 \rightarrow W'$), giving small cross sections. As in the SM the Z' production rate is expected to be smaller than W'.

CHAPTER 3
LHC COLLIDER AND CMS DETECTOR

The data used in this dissertation was obtained by the contribution of thousands of people over the world. To explore the unknown field of physics, people have built the largest machine in human history, named Large Hadron Collider (LHC), to provide the highest energy we can ever achieve [52]. The Compact Muon Solenoid (CMS) is one of the two general-purpose detectors to study the Higgs boson and other new physics at LHC. CMS involves 5000 particle physicists, engineers, technicians, students, and support staff from 200 institutes in 50 countries [53]. The following section will describe the detail of components of the LHC collider and CMS detector.

3.1 Large Hadron Collider

The LHC is the world's largest and most powerful particle accelerator. It is 27 km long in circumference and lies 175 m down underground, located right on the border between France and Switzerland near Geneva. The depth is to protect the detector from the background sources, such as cosmic rays, and the length is to allow the proton to run very close to the speed of light under 7.7 Teslas. It can accelerate a proton to 6500 GeV and an ion to 2560 GeV per nucleon, which means the energy at the center of the collision can go up to 13 TeV. The maximum taemperature is about 5.5 trillion degrees Celsius. This catastrophic collision, which can break atoms into quarks and gluons, happens every 25 ns when LHC is in running mode. To power the machine, LHC needs to consume 0.75 TWh per year, while the whole Geneva only consumes 3 TWh a year [3]. The first concept of LHC appeared in the early 1980s and its construction was approved by the CERN council 10 years later. It was constructed between 1998 and 2008.

There are four main detectors, and 3 additional smaller detectors on the ring. The 4 main detectors are known as ATLAS, CMS, ALICE, and LHCb. CMS and ALTAS are general-purpose detectors which are to study the Higgs boson and look for new physics, while ALICE is to study quark-gluon plasma and LHCb is for the study of the b-physics.

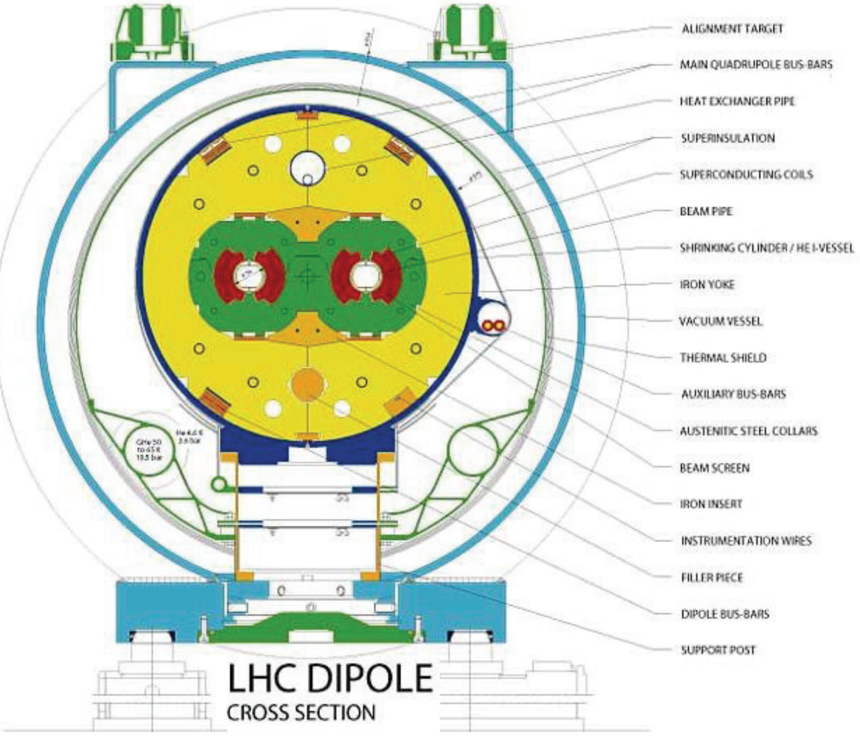


Figure 3.1: Cross section of a dipole [4]

There are 1232 main dipoles, each is 15 m long and weighs 35 tonnes, to bend the paths of the particles, as shown in Fig. 3.1. 392 main quadrupoles help to keep the particles in a tight beam. All these dipoles and quadrupoles need to stay in 1.9 K (-271.3 °C) with liquid helium. It's to keep superconducting magnets that guide and focus the proton beams to stay in a superconductive state and current can run in it without resistance. In this state, the magnets can be relatively small but still produce up to 9 Tesla.

Technically, Particles don't gain 6.5 TeV energy in one single accelerator. As shown in Fig 3.2, particles need to go through a complex succession inside the machine to achieve higher and higher energy. Proton was extracted from hydrogen atoms before injected into Linac 2 and stripped away its electrons via a strong electric field. It'll have 50 GeV after coming out of Linac 2 and will be directed into Proton Synchrotron Booster (PSB), which is an intermediate step to increase the number of protons in multi-stage and accelerate them up to 1.4 TeV. When particles

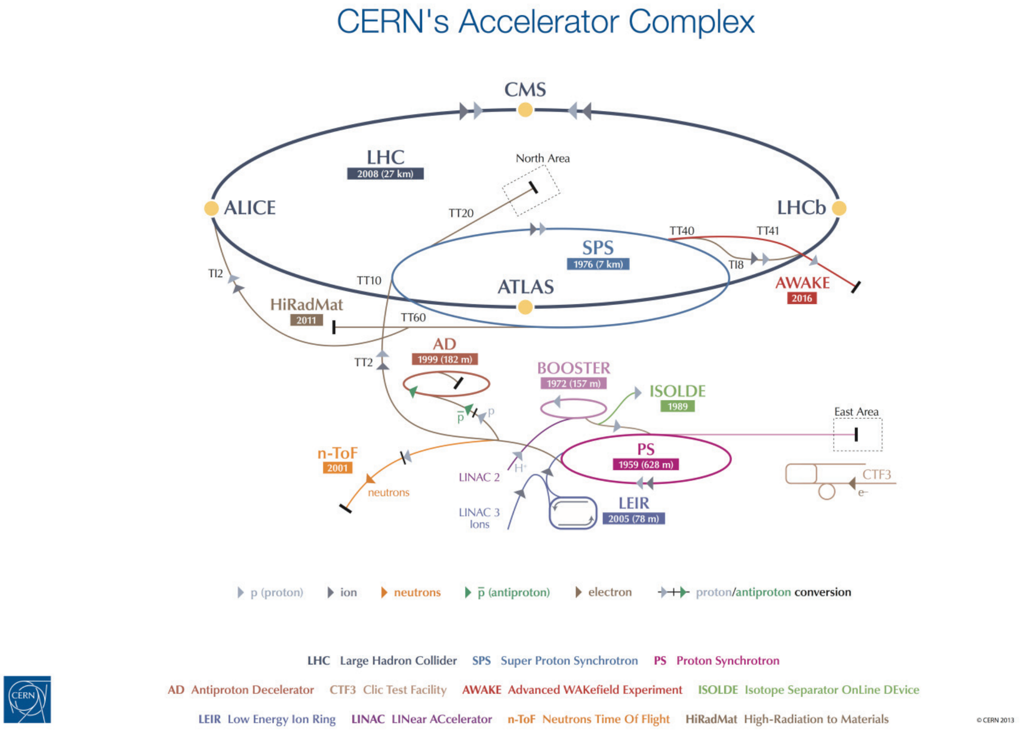


Figure 3.2: Schematic of the CERN facilities that contribute to the LHC beam preparation and collisions [3]

come out, they have been coupled into a more condensed bunch and received by Proton Synchrotron (PS). In PS, particles will be pushed to 25 GeV and will be directed to Super Proton Synchrotron (SPS) for a further boost. SPS is the second-largest accelerator at CERN which is 7 km in circumference. It not only boosts the particle to 450 GeV for LHC but also provide high energy beams to other experiments like COMPASS, NA61/SHINE, and NA62. Notably, SPS provided a neutrino stream to Gran Sasso Laboratory in Italy, 730 km from CERN for the CNGS experiment. At last, LHC receives the 450 GeV particles from SPS and is ready for the final boost and squeeze. The particles will be split into two bunches and inserted into two pipes of opposite directions around the LHC ring. The electric field will accelerate the beam to 6.5 TeV and the strong magnetic field will keep the beam running in a circle. After beams reach their maximum energy, there is one more step, which is to squeeze the beam into a more condensed bunch. All these setup steps will take hours. Beams of two directions will cross with each other every 25 ns, with up to 70 pileups. After dozens of hours of running, the beam bunch will become sparse and will be dumped. Two beams will be dumped at different locations, where beam1 in point 3 and beam 2 in point 6.

An important measurement done by both LHC and each detector is the instantaneous luminosity, which is defined as the ability of a collider can produce the required number of interactions, and the integrated luminosity is described as the total data is recorded throughout time [54]. The instantaneous luminosity is given by

$$\mathcal{L} = \frac{N_1 N_2 N_b f}{4\pi \sigma_x \sigma_y} \tag{3.1}$$

where N_1 and N_2 is the number of particles of two bunches and N_b is the number of bunches in one beam, f is the revolution frequency, σ_x and σ_y are horizontal and vertical beam sizes at collision points, respectively.

During Run II, LHC has delivered total integrated luminosity of 162.85 fb^{-1} and CMS has recored 150.26 fb^{-1}, as shown in Fig 3.3. The online luminosity measurements are provided by the Pixel Luminosity Telescope (PLT), the Hadronic Forward (HF) calorimeter, the Beam Conditions Monitor-Fast (BCM1F), and the Drift Tubes (DT). The offline measurement is based on the Pixel Cluster Counting (PCC) algorithm.

3.2 The CMS Detector

CMS is a general-purpose detector. It's located at the LHC interaction point P5, situated 100 m underground near the village of Cessy, France. The original goal of CMS is to investigate new physics including searching for the Higgs boson, particles that could make dark matter, and new particles. CMS is 21 m long and 15 m in diameter and weighs 14000 tones. It gets its name from its huge solenoid magnet built around it, which is a cylindrical shape of a coil of superconducting cable make by "yoke". The CMS detector consists of layers that exploit the different properties of particles to catch and measure the energy and momentum of each one. They are pixels, strips, ECAL, HCAL, and Muon chambers, sequentially from the inner side to the outer side. A comprehensive diagram of the CMS experiment is shown in

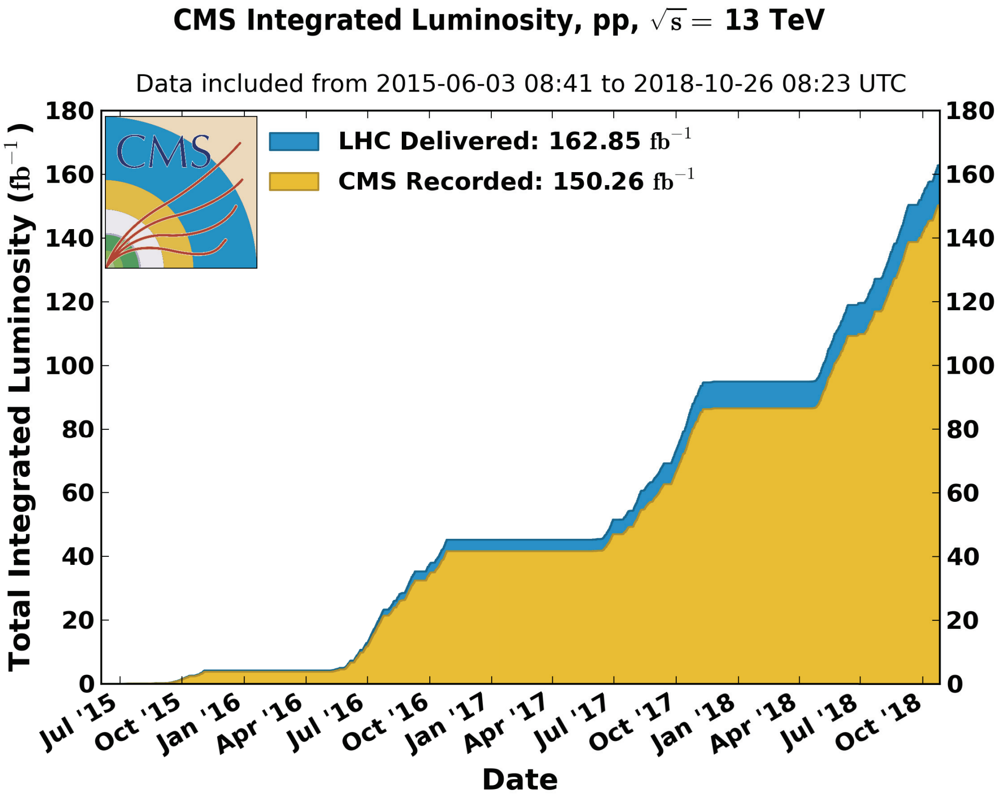


Figure 3.3: Cumulative delivered and recorded luminosity versus time for 2015-2018 (pp data only) with $\sqrt{s} = 13$ TeV data [5]

Fig. 3.4 and 3.5.

At design energies and luminosities, on average around 50 inelastic collisions (or pileup collisions) were expected to be superimposed on the primary hard interaction of interest. A single bunch crossing every 25 ns generates about 1000 charged particles in the detector. To accurately measure the energy of particles and distinguish the product(s) of the hard interaction from the pileup, CMS employs a high-resolution tracking system and high granularity calorimeters with good time resolution, all enclosed in a 3.8 Tesla magnetic field. This results in millions of electronic channels which are synchronized to read out data at every bunch crossing.

In the coordinate system of CMS, the positive direction of the x-axis is pointed from the interaction point (IP) to the center of the LHC ring, the positive direction of the y-axis is pointed to straight to the top. and the positive direction of the z-axis is pointed along the beam counter-clockwise along the LHC ring. For Angular coordinates, while the origin is still the center of the IP, the azimuthal angle ϕ is in the x-y plane along the positive x-axis, and the polar angle θ is measured from the positive z-axis. However, pseudorapidity, η, is used more often than polar angle θ, because θ is not invariant under the interactions in longitudinally boosted collisions. It's the same reason we used conserved transverse momentum p_T instead of momentum p. The pseudorapidity is defined as

$$\eta = -\ln tan(\frac{\theta}{2}) \tag{3.2}$$

and inversely:

$$\theta = 2\arctan(e^{-\eta}) \tag{3.3}$$

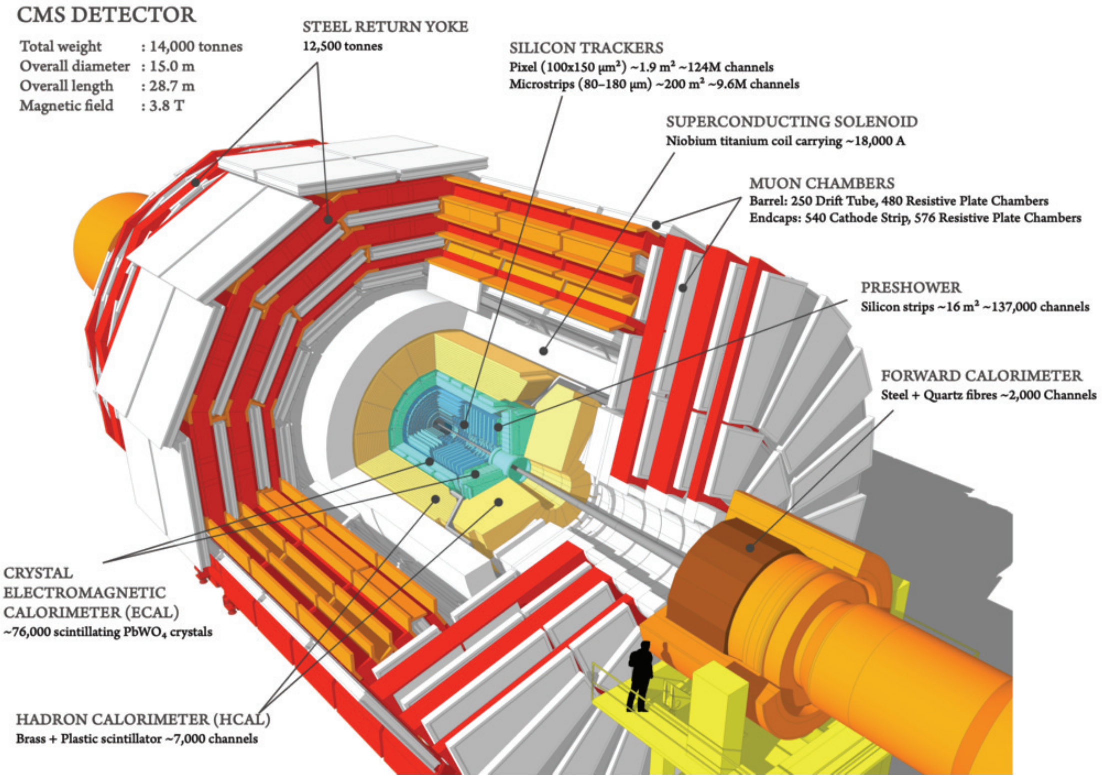


Figure 3.4: Sectional view of the CMS detector.

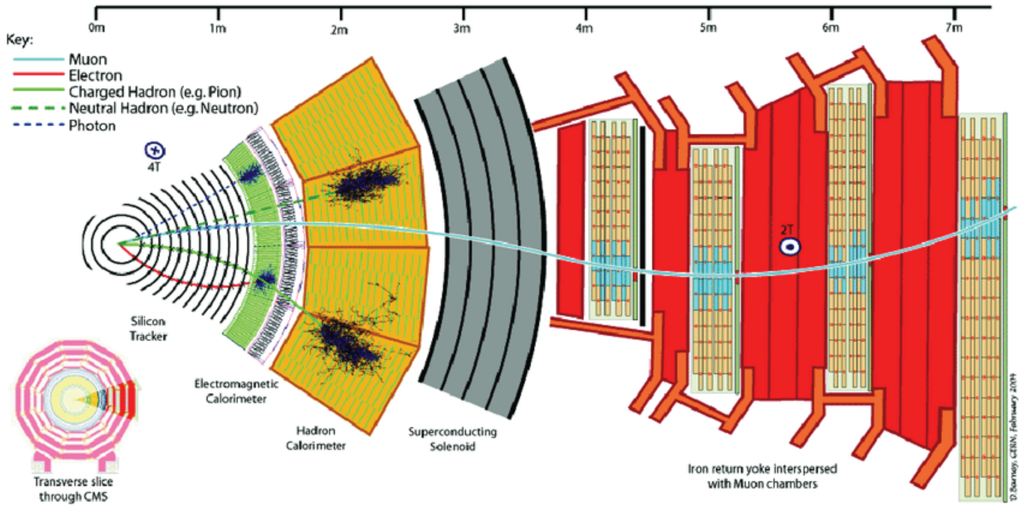


Figure 3.5: Transverse slice of CMS and how different particles intersect with each layer.

3.2.1 Tracking Systems

Tracker is the first subsystem particles come through after interaction. Tracker is the most inner part of CMS to “track” the paths of high energy charged particles like muons, electrons, and hadrons, as well as some short short-lived particles such as b quarks. It can provide high precision measurement points in three dimensions along the curved trajectories of those particles up to pseudorapidities $|\eta| < 2.5$, to calculate the momentum of a particle. The less curved the path through the magnetic field, the more momentum the particle had. The barrel region ($|\eta| < 0.9$) has the best efficiency. The structure of the tracker is shown in Fig. 3.6. Tracker detector has two sub-detectors, both of which are read out via a chain of analog electronics and optical links. They can transmit absolute pulse height. The pixel detector can transmit pixel coordinates and all data-processing of strip detector happens in off-detector electronics.

The detector must build the precise measurement of the path because we need it to reconstruct the position of the primary interaction and secondary decay vertices, which will be used for pileup events and particle identification. LHC was designed to have 25 pileups for Run II. But in the 2017 and 2018 pileup scenarios, the

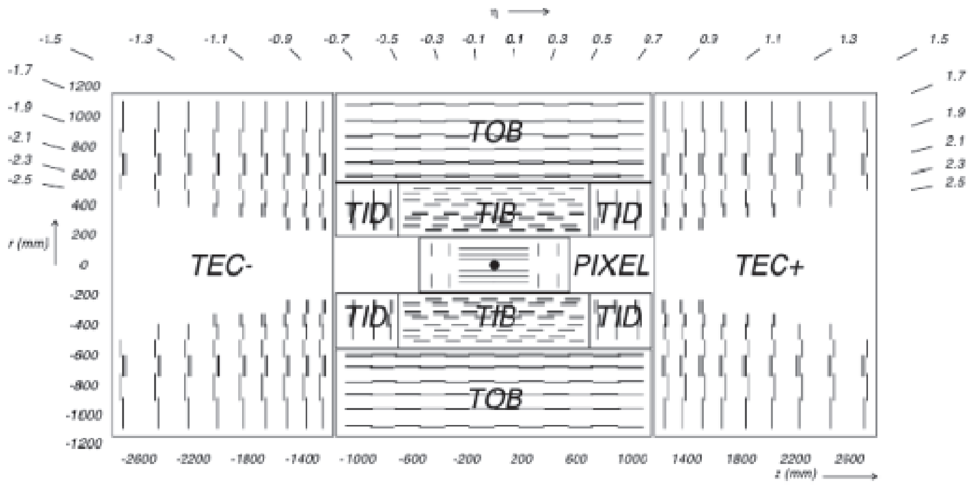


Figure 3.6: Layout of CMS tracker on r-z slice plane before the phase I upgrade. The horizontal axis pointing along z and the vertical axis pointing to the radius r. [6]

average number of pileups is 32 and with more than 50 interactions in short periods. Additionally, there will be an issue called out of time pileup. The high pileup and out of time pileup will challenge tracker occupancy. People have exchanged pixel detectors to solve this during Year-End Technical Stop (YETS) in 2016/2017, which is called the Phase I upgrade.

3.2.1.1 Pixel Detector

The pixel detector [55] is located at the innermost part of the CMS tracker detector, which covers the region of $|\eta| < 2.5$. Pixel detector consists of a barrel part (BPIX) and endcap discs (FPIX). The current pixel detector has one additional tracking layer than before both in the barrel and endcap region, which offer full 4-hit tracking coverage in all $|\eta| < 2.5$ region, with more redundancy compared to the older system. Even though the new pixel detector has a new μTCA-based data acquisition system (DAQ), replaced the old VME-based DAQ system, both old and new Phase-I upgraded system use the same type of n^+-in-n sensors, which covers active area of 16.2×64.8 mm^2 and thickness of 285 μm. The radii of the four layers have been changed to 29, 68, 109, 160 mm from the IP in the BPIX, and 291, 396, and 516 mm in FPIX. Benefiting from the change in layout, the total number of channels has almost doubled from 66 million channels, which is 48 million for BPIX and 18 million for FPIX, to 124 million channels, which is 79 million for BPIX and 45 million for FPIX. Even though one more layer and more channels mounted, the pixel detector becomes even lighter. The reduction of the system benefits from the ultra-lightweight CO_2 evaporative cooling, which can also remove heat more effectively, and relocating much of the passive material. With the improvement of the fake rate of primary vertices and secondary vertices of the high pileup, this upgrade can greatly improve tracking efficiency and b-tagging performance. The normal pattern recognition and track reconstruction work in a way that better tracks constructed first and then removes the hits of this track from the remaining hits, iteratively, because the “best” track in each iteration is less likely to be the fake tracks. In the old system, this procedure starts with 2 (a pair seed) or 3 (triplet seed) “seed” pixel hits with minimum track p_T. But with the additional barrel and endcap disk, and their full-coverage over the η region, it allows perfect 4 (quadruplet seed) “seed” pixel hits, which made fake rate intrinsically lower than before. As a result, b-tagging algorithms can be benefited from the upgrade, either the old CSV and new DeepJet, which highly rely on the lifetime information, like secondary vertices (SV), impact parameter (IP) significance, and decay lengths. Details of the b-tag algorithms will be talked about in the section 4.7, because of the low fake rate, and higher resolution measurement on the SV, IP, and the decay length [7].

The pixel upgrade during 2016/2017 YETS is shown in Fig. 3.7.

3.2.1.2 Strip Detector

The tracking detector in the central part of the CMS experiment is a very unique instrument, in both size and complexity. It consists of two systems using silicon sensor technology: one is the silicon pixels and the other is silicon microstrips. The pixel system, which is described in the previous section, is surrounded by the Silicon Strip Tracker (SST), which will be the subject of this section. The mechanical layout and mounting of the bottom half part of tracker subdetectors are shown in the Fig 3.8.

The purpose of the SST is to provide efficient and precise measurements of

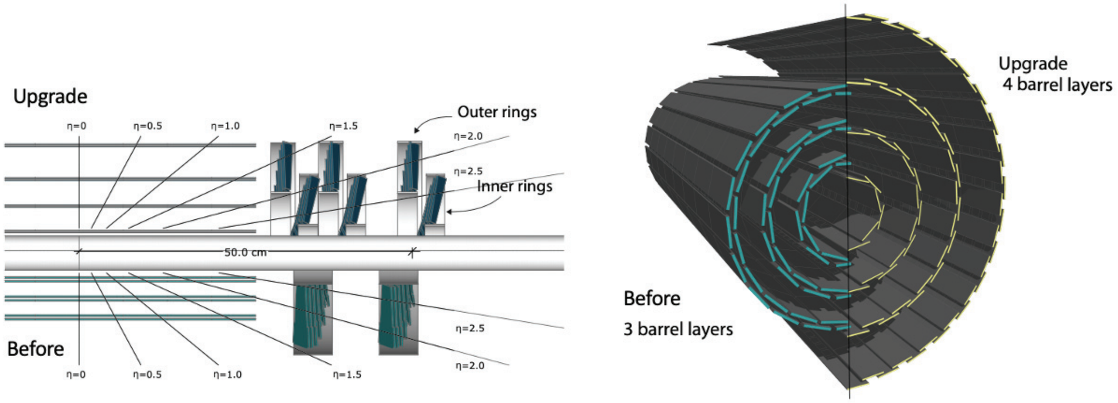


Figure 3.7: Left: Comparison of the geometrical layouts of the old (bottom) and upgraded (top) CMS pixel detectors in Phase-I between 2016/2017. Right: Transverse-oblique view comparison between pixel barrel layers. [7]

trajectories of a charged particle in a solenoidal 3.8 T magnetic field with a p_T accuracy of 1-10% in the range of $1 - 1000$ GeV/c in the central region. The SST has five main subdetectors: the tracker inner barrel with inner disks (TIB and TID), the tracker outer barrel (TOB), and the tracker endcaps on positive and negative sides (TECP and TECM). Sometimes CMS combines the two tracker endcaps and calls them TEC. Each of the TID disks has 3 disk rings of modules, while TEC disks have 7 rings. Overall, the tracker cylinder is 5.6 m long and 2.4 m in diameter, with a total active area of 206 m^2, with 15148 detector modules and comprising 9.3 million detector channels. The TIB extends in the z-direction to ± 70 cm and 55 cm in radius. TIB consists of two half-length barrels with 4 detector layers, attached with 3 TID disks at each end. The TID disk is equipped with wedge-shaped silicon detectors with radial stripes. The TOB part surrounds the TIB and TID, which has an outer radius of 116 cm, ranges in $|z|$ up to 118 cm, with 6 barrel layers in total. The barrel part of strips tracker orients along the z-direction, except for the double-sided stereo modules in the first two layers of TIB and TOB, where they have been rotated by 100 mrad, which provides all particles reconstruction of z coordinate. At the very outside, TECP and TECM cover the region $124 < |\mathrm{z}| < 282$ cm and $22.5 < \mathrm{r} < 113.5$ cm. Each of the two subdetectors is composed of 9 disks, carrying up to 7 rings of wedge-shaped silicon detectors with radial strips, similar to

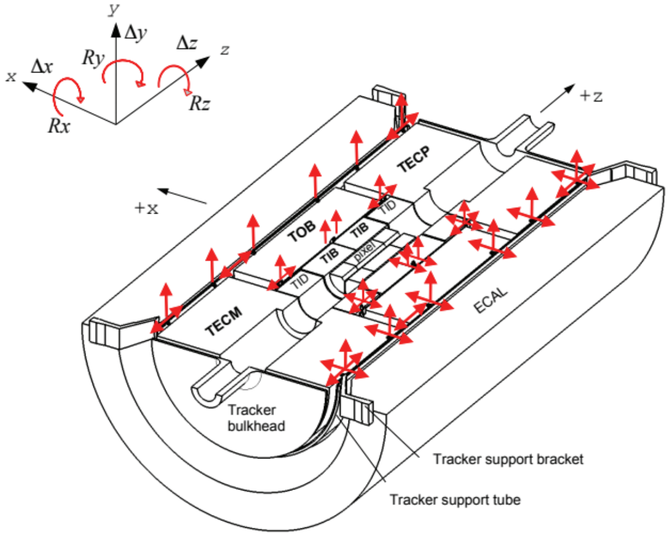


Figure 3.8: Mechanical layout and mounting of the tracker subdetectors (bottom half is shown). The TIB+TID are mounted inside the TOB, while the TOB, TECP, and the TECM are mounted inside the TST. The red arrows indicate the connection points and their kinematic constraints [8]

TID. Rings 1, 2, and 5 are also equipped with stereo modules for the reconstruction of the r-coordinate.

All modules in the silicon strip detector have one or two silicon sensors that are glued on the carbon fiber (CF) frame together with a ceramic readout hybrid, with a mounting precision of 10 μm. There are overall 26 different module designs, to optimize for different positions in the tracker. In detail, the TIB is split into 2 halves, negative and positive in z-coordinates, which allow easy insertion into the TOB. TIB substructures have 16 CF half-cylinders or shells. The modules are assembled in rows that overlap like roof tiles for better coverage and compensation for the Lorentz angle. An Aluminium cooling tube, with 0.3 mm wall thickness and 4×1.5 mm^2 rectangular profile is glued to the mounts of the detector modules. Each row has three modules on one cooling loop and each cooling pipe is connected at the edges of the shells to the circular collector pipe that gives extra rigidity to the whole TIB mechanical structure. The overall positional accuracy of the assembly of all shells is about 500 μm.

The TOB main structure consists of six cylindrical layers supported by four disks, two at the ends, and two in the middle of the TOB structure. The disks are made of 2 mm thick CF laminate and are connected by cylinders at the inner and outer diameters. The cylinders are produced from 0.4 mm CF skins glued onto two sides of a 20 mm thick aramid-fiber honeycomb core. The detector modules are mounted on 688 substructures called rods. The rods are inserted into openings on the disks, such that each rod is supported by two disks. The accuracy of mounting the rods is about 140 μm in $r - \phi$ and 500 μm in z. Each rod has 6 or 12 (for rods with double-sided modules) silicon modules mounted in a row. A 2 mm diameter copper-nickel cooling pipe is attached to the CF frame of the rod. Each module is mounted on the rod with an accuracy of 30 μm by two precision inserts connected to the cooling pipes, and two springs.

Each TEC side consists of nine disks with 16 wedge-shaped substructures on each disk, called petals. Overall there are 144 petals with different layouts, depending on the disk location. The petals are made of CF skins with a honeycomb structure inside. The wedge-shaped detector modules are mounted on the petals with an accuracy of 20 μm using four Aluminium inserts that are connected to the cooling pipe. A titanium cooling pipe of about 7 m in length, 3.9 mm in diameter, with 0.25 mm wall thickness is integrated into the petal honeycomb structure and is bent to connect all heat sink inserts. The petals are mounted on the CF disks with a precision of 70 μm. All nine disks of each TEC are connected with eight CF bars forming a rigid structure. These bars are also used to hold service cables and cooling pipes. The overall accuracy of the disks assembly in the TEC subdetector is about 150 μm in all coordinates.

The temperature of different mechanical parts inside the strip depends on the distribution of heat sources and heat sink. It’s monitored by dedicated sensors that are directly mounted on readout hybrids, silicon sensors, and mechanical structures, distributed throughout the detector volume. Meanwhile, the non-uniformity of heat dissipation and heat removal results in significant temperature variations inside the tracker. The temperature measured in SST in all subdetectors running with a cooling plant at an operating temperature of $-5°$C, which are shown in Fig. 3.9. The white areas represent detectors that are non-operational detectors, which is about 2.5% of the total area. The red areas are hot parts that are five cooling loops that are closed because of leaks and bad cooling contacts even in thermal equilibrium.

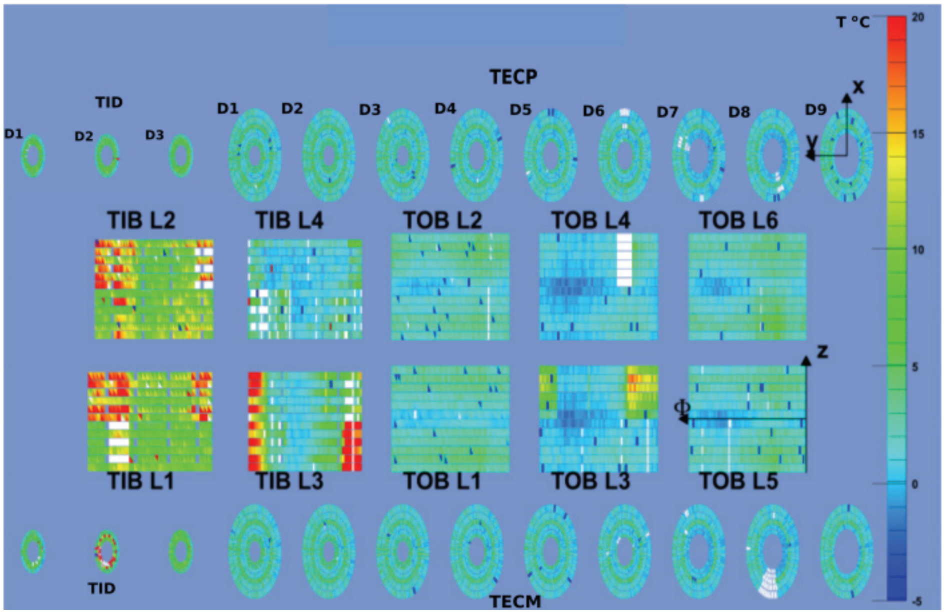


Figure 3.9: Example of the temperature distribution, shown as a color palette (°C), measured on silicon sensors in the TIB (L1-L4), TOB layers (L1-L6), and the TEC (D1-D6), TED (D1-D3) with the cooling plant operating at T= -5°C. The white spots correspond to nonoperational detectors, and red spots are the closed cooling loops and bad cooling contacts.

3.2.2 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) is one the most important components of the CMS detector, whose is designed to identify and measure photons and electrons in stringent requirement on energy resolution, in order to be sensitive to the two photons from Higgs boson decay. ECAL is also designed to provide fast response, radiation tolerance and excellent energy resolution [56]. It consists of 75848 lead tungstate ($PBWO_4$) crystal. The barrel region of ECAL extending to a pseudorapidity $|\eta|$ of 1.48, while the two endcaps extending to cover $|\eta| < 3.0$. Scintillation light from the crystals is detected by Avalanche Photodiodes (APDs) in the barrel region and by Vacuum Phototriodes (VPTs) in the endcaps. The layout of ECAL, including barrel region, endcap region and preshower, is demonstrated in the Fig. 3.10.

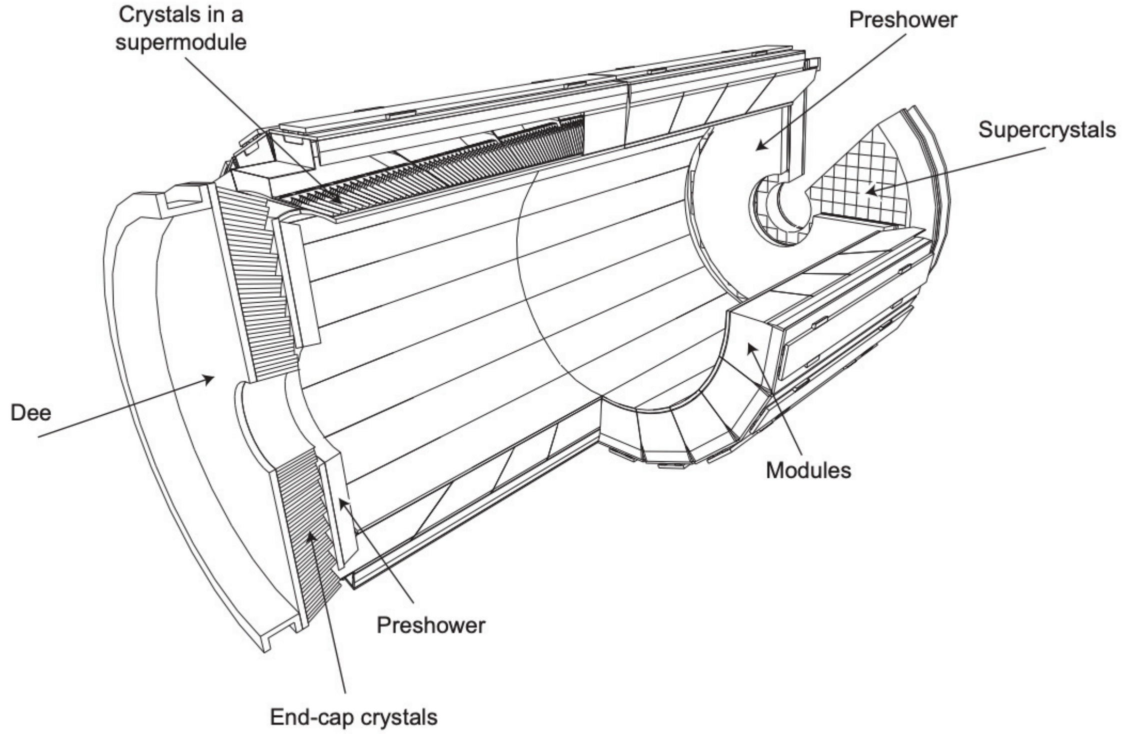


Figure 3.10: Layout of the CMS electromagnetic calorimeter, showing the barrel supermodules, the two endcaps and the preshower detectors. [9]

In the need of desired energy resolution, it is necessary to maintain the stability of the per-channel energy calibration over time. These strict requirements on the stability of the temperature of the ECAL and the high voltage applied to the APDs. This is due to the temperature dependence of the crystal light yield, as well as the sensitivity of the APD gains to variations in both temperature and high voltage (the VPT response is much less sensitive to temperature and high voltage variations). Also, changes in crystal transparency under irradiation must be tracked and corrected.

ECAL Barrel (EB) has 36 supermodules, each of which is composed of 1700 tapered $PbWO_4$ crystals with a frontal area of approximately 2.2×2.2 cm^2 and a length of 23 cm (corresponding to 25.8 radiation lengths). All crystal axes are inclined of angle 3°, which is relative to the direction of the nominal interaction point, in both the ϕ and η projections. Scintillation light from the crystals will be detected by two Hamamatsu S8148 5×5 mm^2 APDs, which are specifically designed for CMS and operate at a gain of 50. All of them are connected in parallel to the on-detector readout electronics, which are put in units of 5×5 crystals, each unit corresponding to a trigger tower. Each trigger tower has five Very Front End (VFE) cards, each of which can accept data from 5 APD pairs. APD signals will be pre-amplified and shaped by Multiple Gain Pre-Amplifier (MGPA) ASICs located on the VFE boards, which consist of three parallel amplification stages [57]. The output signal will be digitalized by a 12-bit ADC, which runs at a frequency of 40 MHz, The data from 5 VFE cards are transferred to a signal front-end card, which can generate the trigger primitive data [58], and then transmits to the dedicated off-detector trigger electronics. The energy resolution is defined as

$$\frac{\sigma}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{\sqrt{E}} \oplus 0.3\% \tag{3.4}$$

Two ECAL endcaps composed of 4 half-disk “dees”, each of them consists of 3662 tapered crystal with a frontal area of 2.68×2.68 cm^2 and a length of 22 cm. arranged in a quasi-projective geometry. The crystal is focussed at a point which is 1.3 m away from the nominal interaction point along the beamline, with off-pointing angles between 2° and 8°. Crystal in each “dee” has 138 standard 5×5 super crystal units and 18 special shaped super crystals that are located at the inner and outer radii. Scintillation light is detected by VPTs produced by NRIE with an active area of 280 mm^2 and operating at gains of 8-10, which are glued to the rear face of the crystals.

The silicon pre-shower detectors, which are located in front of the ECAL endcaps, have been fully commissioned before LHC operation in late 2009, which is intended to prevent false signals caused by short-lived particles called neutral pions. Those neutral pions can also decay into high energy photons, which eventually will decay into two closely-spaced lower energy photons.

3.2.3 Hadron Calorimeter

The Hadron Calorimeter (HCAL) measures the energy of hadrons, which are particles made of quarks and gluons like pions and kaons. Additionally, it provides an indirect measurement of the presence of non-interacting, uncharged particles such as neutrinos. It's important to measure those particles because it indicates evidence of the presence of new particles like the Higgs boson or SUSY particles. To make sure we don't miss any evidence, layers of the HCAL were built in a staggered fashion so that there is a gap in direct lines. HCAL is a sampling calorimeter, which can find the position, energy, and arrival time of particles, using alternating layers of the absorber and fluorescent scintillator materials that produce a rapid light pulse, when the particle passes through. HCAL has 4 major sections, HCAL Barrel (HB), HCAL Endcap (HE), HCAL Outer (HO), and HCAL Forward (HF). The overall location of them in CMS is shown in Fig. 3.11.

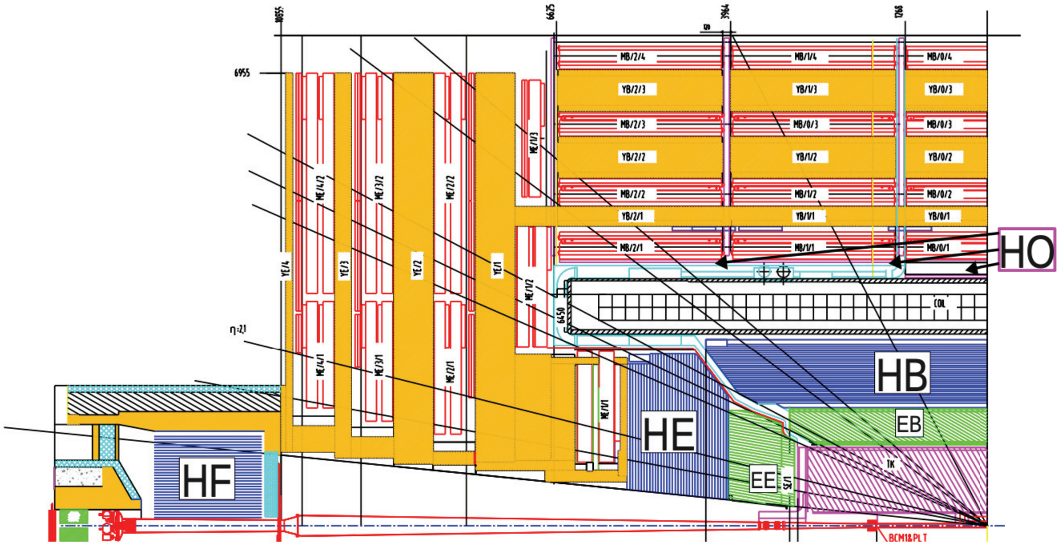


Figure 3.11: An r-z schematic drawing of a quarter of the CMS detector showing the location of the HB, HE, HO, and HF calorimeters in CMS. [10]

After Phase I upgrade, SiPM is installed to replace HPDs in HB and HE. The SiPM is an array of Geiger-mode operated avalanche photodiodes, in the size of microns. It only has an area of a few square millimeters but has a very high gain, in the order or 10^4, while the recovery time is less than 10 ns, to ensure minimal shifts in response from pileup events. Because of the improvement of performance in SiPM, coupled with new developments in the data link technology, a significant increase in depth segmentation in the HB and HE calorimeters becomes possible. The increased segment of HB and HE are shown in Fig. 3.12, indicated by color/shading of the tile structure. This structure will allow better tracking of the hadronic shower. It can also manage radiation damage better in the high $|\eta|$ region in HE, reducing the response of the individual tiles. The longitudinal segmentation of the HCAL can provide shower profile information which is used to verify the electromagnetic particles identified in the ECAL have little impact on the HCAL. But the segment only suppresses the influence of pileup particle that contributes to the first layer of HCAL. In the same way, the deeper layer of HCAL can also be used to identify prompt muon and rejection of muons produced in the decay of hadrons in flight with isolation requirements. The HCAL is readout in individual towers with a cross-section $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and 0.17×0.17 at larger pseudorapidities. The energy resolution of combined calorimeters was measured in pion test beam:

$$\frac{\sigma}{E} = \frac{110\%}{\sqrt{E}} \oplus 9\% \tag{3.5}$$

where E is in the unit of GeV

HF plays a very important role in identifying forward jets, determining missing transverse energy, and measuring the luminosity since it extends the coverage of HCAL to high pseudorapidities ($3.0 < |\eta| < 5.0$). However, when LHC entering the phase of 25 ns, anomalous signals in HF becomes more difficult to reject as the luminosity increases, as well as the topological requirements, become harder to apply. After Phase I upgrade, the Photomultiplier Tubes (PMTs) are replaced with Hamamatsu R7600U-200-M4 multi-anode tubes, which has thinner optical windows and metal envelopes rather than glass. These changes can significantly reduce the rate of anomalous signals, though not to suppress them entirely. Another change made is upgrading the TDC capability in the ADC chip, allowing the determination of pulse arrival times and pulse widths for signals.

The HCAL Outer calorimeter (HO) is working as a tail-catcher for hadronic

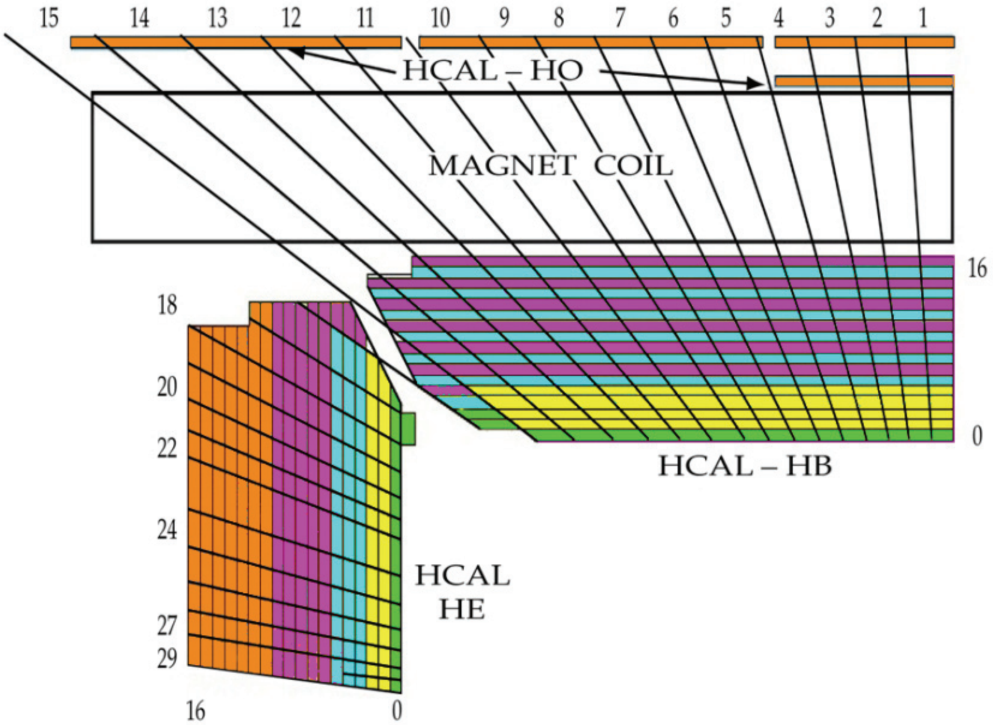


Figure 3.12: Depth segmentation structure of HB and HE after Phase I upgrade, using SiPM photodetectors

showers and it's very useful in muon identification. However, HO is not part of the Phase I upgrade, because the HPDs are already installed during the first long shutdown of LHC in 2013 and 2014.

3.2.4 Muon Detects

As the name "Compact Muon Solenoid" impacts, it's one of the most important tasks of CMS to detect muons and record their energy and trajectory. Muon is a charged particle and can penetrate several meters of iron with no interaction. As a result, it's hard for CMS to stop it in calorimeters. Therefore, muon detectors are placed at the very outside of CMS and find the signal of muons.

There are 1400 chambers in total and they are made up of 3 types. The first type is Drift Tube (DT), which sits only in the barrel region. The goal of DT is to find the drift time in an anode wire of a cell with a shaped electric field, to determine the position of muons. The second type is the Cathode Strip Chamber (CSC), which is located at the endcap. The CSC is named by its finely segmented cathode strip readout. The readout can measure the position of muon accurately in the bending plane ($R-\phi$) coordinate when muon passes through the gas volume. The last is named Resistive Plate Chamber (RPC), which is placed in both the barrel and endcap region. It is a double-gap chamber operated in the avalanche mode and is primarily designed to provide timing information for the muon trigger.

The R-z schematic diagram of CMS, which introducing details about muon detectors is shown in Fig 3.13.

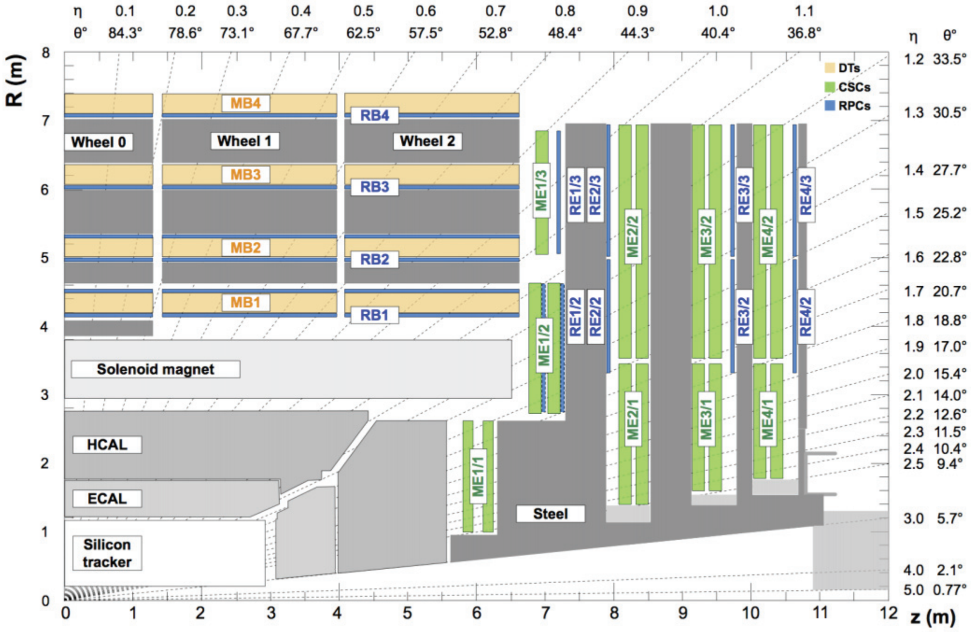


Figure 3.13: Schematic of the CMS detector from R-z side view. The location of each muon stations and the steel flux-return disks (dark areas) are shown. The DTs are labeled as MB (Muon Barrel), CSCs are labeled as ME (Muon Endcap) and RPCs are labeled as RB and RE, because they are mounted in both barrel and endcaps of CMS.

3.3 Trigger System

3.3.1 Trigger system

The CMS experiment has a two-level trigger system. The first level is the firmware based Level-1-Trigger (L1T), which can reduce the event rate from 40 MHz to 100 kHz. The Next level is the High-Level Trigger (HLT), which can further reduce the rate to 1 kHz.

3.3.2 L1 trigger

L1T receives information from the calorimeter (ECAL, HCAL) and the muon chambers (CSC, DT, RPC), which is named trigger primitives (TP), while no tracker information is taken into consideration in this step. Calorimeter TPs are then combined to form calorimetric towers and then like to the compatible muon hits. Afterward, they are used to form the L1 trigger objects, such as muons, jets, and electrons/photons. In the end, the global decision is made based on the presence of energy deposits compatible with these physics objects. The schematic of the L1 trigger is shown in Fig. 3.14.

3.3.3 High Level Trigger

The HLT is a streamlined version of the CMS offline reconstruction software running on a cluster of commercial rack-mounted computers, which has about 26000 CPUs. As the output rate from L1T is about 100 kHz, the HLT is designed to further reduce this rate to 1 kHz. This limitation is from the computing resources involved in the offline reconstruction because the space on the tape would allow us to record up to 2 kHz. Since the limitation applies to the average HLT rate, we want to have a short period with a peak rate of 1 kHz. HLT consists of hundreds of HLT paths, which are sequences of reconstructing and filtering modules. if an event is rejected by a filter, the subsequent modules will not run. The idea is to reconstruct quick objects first and then reconstruct slower objects, to keep a low average computing time. For example, the selections used in an HLT path with b tagging uses the following information: L1 seed, calorimetric objects, b tagging with regional tracking, and PF objects.

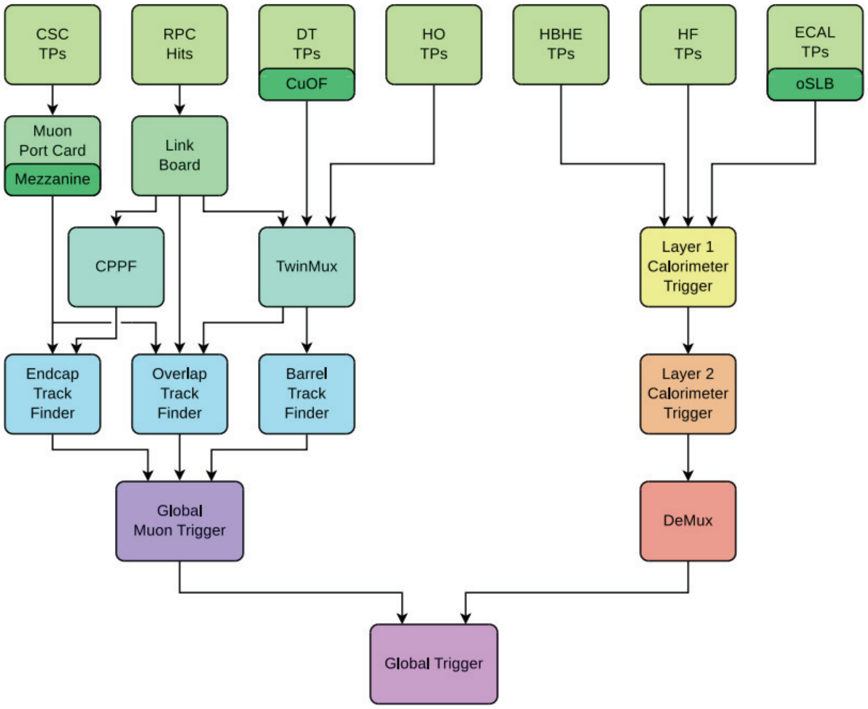


Figure 3.14: Schematic of the L1 trigger

CHAPTER 4
SIMULATION AND JET RECONSTRUCTION

Identifying the fundamental particles and describing their interactions from the first principles is the goal, not only for the accelerator experiments but also for many other experiments, such as dark matter detection and neutrino experiments underground or up in space. Discoveries from these experiments as well as their interpretation will rely on our ability to perform accurate simulations for both the signals and their backgrounds. At the LHC, there are a few challenges to extract physics from the data. The first is that LHC is a very high energy pp collider which involves a large number of jets and high MET, in which there is an overwhelming background in many BSM searches. It needs to precisely predict the wide classes of those models, as well as to help build up an effective field theory from the data. It's also required to accurately measure the corresponding parameters, production rate, and/or BR.

Meanwhile, hadronic jets are the crucial component of the CMS physics program. It's required to have the precise determination of momentum and momentum resolution of jets for searching for new physics. At the same time, the study for jet substructures, which can improve the identification of heavy flavor particles decaying into single jets, whereas quark-gluon separation can be used for the reduction of QCD background and signal enhancement.

In this context, every study at the LHC all require good simulation, reconstruction tools, and algorithms. As a result, some important advances are made in this field in recent years.

4.1 Simulation and Generation

All Monte Carlo datasets are generated by `MadGraph5` (`MG5`) [59] and `PYTHIA` [60], which is widely used in the particle physics field.

`MG5` is open-source software written in Python and features a collaborative development structure. It can generate matrix elements in the tree-level for any Lagrangian based model. FeynRules [61; 62; 63] are implemented via the UFO interface. and automatic generation of the corresponding helicity amplitude subroutines via the ALOHA package [64]. When using `MG5`, the user specifies a process in terms of initial and final state particles. As a result, MG generates all Feynman diagrams for the process and outputs the computer code necessary to evaluate the matrix element at a given phase space point. Calls to helicity wave functions and amplitudes are used to do matrix element evaluation, which has been implemented in the HELAS package. `MG5` can also produce a pictorial output of the Feynman diagrams for the process. Cross-section and decay width can also be calculated using the computer code produced by `MG5`.

`PYTHIA` program is a standard tool for the generation of the event in high-energy collisions between elementary particles comprising a coherent set of physics models for the evolution from a few-body hard-scattering process to a complex multi-particles final state. Part of the models is derived from the theory, while the others are based on existing phenomenological models., with parameters to be determined by data. The main task of `PYTHIA` including exploration of experimental consequences of theoretical models, the development of search strategies, and interpretation of data, and studying the detector.

4.2 CMSSW Event Jet Reconstruction

As a general-purpose detector at the LHC, CMS is designed on the concept of cylindrical detection layers, nested around the beam axis. From the primary interaction point, the first subdetector particle encounter is the track, where the trajectories (tracks) and origins (vertices) will be reconstructed by the signals (hits) in the sensitive layers. Meanwhile, the trajectories in the track will be bent when they are immersed in a magnetic field so that the electric charge and the momenta can be archived. Afterward, electrons and photons will be absorbed in the ECAL, the electromagnetic showers from the hits can be detected as clusters of energy recorded in the neighbor cells. The shower from charged and neutral hadrons begins in the ECAL, which will be then fully absorbed in the HCAL, the energies and directions can be clustered. Muons and neutrinos have no interaction when passing through the previous subdetectors. However, the muon detector at the very outside will keep the tracks of muons and record their energy. The tagging of the jet, like the decays from the b quark. The simplified view is graphically summarized in Fig. 3.5

The particle-flow (PF) reconstruction is developed to significantly improve event description combining all basic elements from all detector layers, and the corresponding measurements to reconstruct the particle properties based on this identification [65]. This allows the cluster energies to be calibrated more accurately under either the photon or the hadron hypothesis.

QCD multijet events samples are used to qualify the jet performance while jets are reconstructed with the anti-k_t algorithm with radius parameter R = 0.4. The algorithm clusters either all particles reconstructed by the PF algorithm (PF jets), or the sum of the ECAL and HCAL energies deposited in the calorimeter towers, or all stable particles produced by the event generator excluding neutrinos (reference jets) Details of AK4 jets will be described in the Sec. 4.3. Fig. 4.1 show the simulation and reconstruction of a dijet event.

4.3 Anti-k_t Jet Clustering Algorithms and AK4 Jet

The distance d_{ij} is defined as the distance between entities, like particles and pseudojets, i and j. While d_{iB} is the distance between entity i and beam B. The inclusive clustering proceeds by identifying the smallest of the distances. If it is a d_{ij} which recombines the entities i and j, while if it is the d_{iB} calling i a jet and removing it from the list of entities. The distances are recalculated and the procedure repeated until no entities are left.

The equation of d_{ij} and d_{iB} is defined in Eq. 4.1 and Eq. 4.2

$$d_{ij} = min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta_{tj}^2}{R^2} \tag{4.1}$$

$$d_{Bj} = k_{ti}^{2p} \tag{4.2}$$

Here $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and k_{ti},y_i and ϕ_i are the transverse momentum, rapidity and azimuth of particle i, respectively. Addition to the normal way of using radius parameter R, we add a parameter named p to govern the relative power of the energy versus geometrical Δ_{ij} scales.

When p = 1, we can recover the inclusive k_t algorithm. When p = 0, then we have the inclusive Cambridge/Aachen algorithm [66]. However, when p = -1, it is

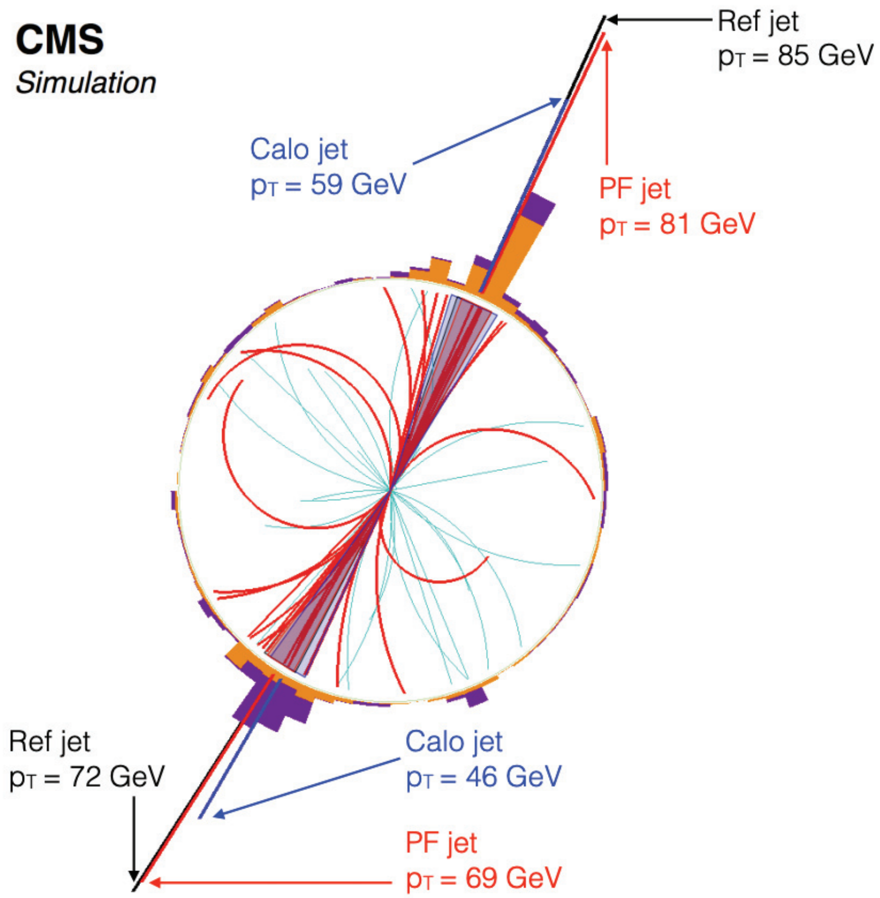


Figure 4.1: Jet reconstruction in a simulated dijet event [11].

referred to the “anti-k_t” jet-clustering algorithm [67].

The general functionality of anti-k_t algorithm is understood by the following. If an event contains a few well-separated hard particles with transverse momenta $k_{t1}, k_{t2}...$, and some other soft particles.the distance $d_{1i} = min(1/k_{t1}^2, 1/k_{ti}^2)\Delta_{1i}^2/R^2$ between a hard particle 1 and a soft particle i is exclusively determined by the transverse momentum of the hard particle and the Δ_{1i} separation. In this way the soft particles will tend to cluster with hard ones before they cluster themselves, which means it there no another hard particle with 2R of a hard particle, then this one will cluster all soft particles with a radius of R. However, if there is another one with the $R < \Delta_{12} < 2R$, then both particles can't be partly conical. The overlapping part will be simply divided by a straight line equally between the two hard particles if $k_{t1} \approx k_{t2}$. If $k_{t1} \gg k_{t2}$, the jet of the first particle will be conical and the other will be partly conical. When the distance between two hard particles $\Delta_{12} < R$, then the two hard particles 1 and 2 will form a single jet.if the $k_{t1} \gg k_{t2}$ then it will be a conical jet centered at k_1. If not, the shape will be a union of the cones around each hard particle plus a cone centered on the final jet.

In the PF reconstruction, the particles produced in pileup interactions give rise to additional charged hadrons, photons, and neutral hadrons. If a jet is identified as coming from pileup, these charged hadrons are removed from the list of reconstructed particles used to form physics objects. This widely used algorithm is called pileup charged-hadron subtraction and denoted as CHS jet.

And at last, AK4 implies a jet clustered with the anti-k_T algorithm with distance parameter $R = 0.4$ in the above Eq. 4.1.

4.4 Wide Jet

Geometrically close jets are combined into "wide jets" and used to determine the dijet mass, as in our previous dijet searches. The wide-jet algorithm, designed for dijet resonance event reconstruction, reduces the analysis sensitivity to gluon radiation from the final state partons. An AK4 jet and the four-vectors of all other jets, if they are within $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 1.1$ of the seed jet, and are added to the nearest leading jet to obtain a wide jet. The above requirements maximize the search sensitivity for isotropic decays of dijet resonances in the presence of a QCD dijet background.

4.5 Jet Energy Correction

The jet energy correction (JEC) is determined by the detailed MC simulation and then adjust for data using a data-driven method applied on several samples [68]. The JEC is extracted by jets applying cut on the following selection. The jet p_T ¿ 10 GeV and $|\eta| < 5.2$ with uncertainty $\leq 3\%$ and performed in the following state: First, the pileup offset and noise correction is determined, then the simulation-based response is established (both stages applied to MC and data), and finally absolute and relative residual corrections are applied to data only. Optional flavor corrections can also be applied.

CMS has a factorized solution to deal with this problem, containing multiple levels of processes, which are all listed in the Fig. 4.2. Each level of these corrections is to scale the jet four momentum with scale factor.

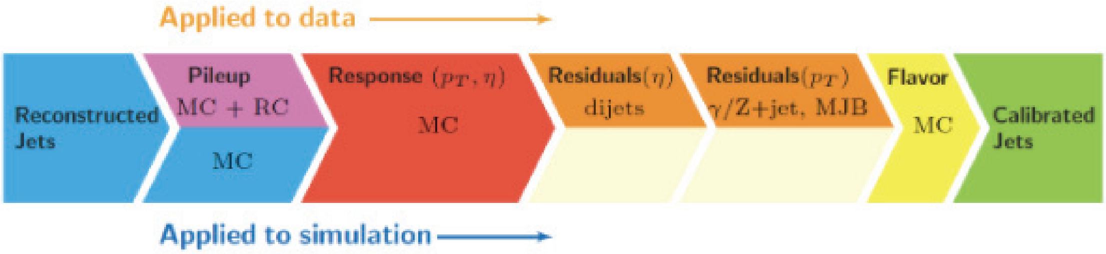


Figure 4.2: Factorized jet energy correction steps to change the jet energy from measured jet energy deposition in the detectors to the corrected particle-level jet energy

The first level is to remove the energy coming from pile-up events. The second and the third level correction is to correct the jet p_T from the reconstructed one to the particle-level one, which is determined based on the p_T and η. Also, after the L2 and L3 correction, there are still remaining small differences. L2L3Residuals is meant to deal with this. These corrections include a p_T dependence of the JES relative to the JES of the barrel jet.

The jet energy resolution is defined as $\sigma(< p_T > / < p_{T,part} >)$ after applying the JEC. It's determined with p_T asymmetry in dijet data and p_T-balance in Z/γ+jets, ad it is parametrized as a function of particle-level jet $p_{T,part}$ and the average number μ of pileup interactions in the bins of jet η. The resolution is stable against pileup above $p_T = 100$ GeV and measures about 10% (5%) above 100 GeV (1 TeV).

4.6 Dijet Resonance Shapes

The dijet system is composed of the two wide jets with the highest p_T (leading jets) in an event. The dijet mass is given by $m = \sqrt{(E_1 + E_2)^2 - (\overrightarrow{p_1} + \overrightarrow{p_2})^2}$. The heart of the search for dijet resonances is the measurement of the dijet mass distribution and the estimation of the background. Unlike many other searches in high energy physics, the search for dijet resonances is completely dominated by a single background. The observed dijet mass distribution comes from the dominant process in hadronic collisions: 2 → 2 scattering of partons predicted by perturbative QCD.

4.6.1 Angular Requirement

In CMS dijet search, we want to limit all jets to stay in the range of tracker, which means all of them should stay in the region where $|\eta| < 2.5$. In addition, in order to suppress the QCD background, another requirement is needed on the η between the two jets. This is because QCD production of dijets at high mass is dominated by t-channel production, with an angular distribution that is approximately Rutherford scattering, as shown below, which peaking strongly when $cos\theta^* = 1$, where θ^* is the scattering angle.

$$\frac{dN}{dcos\theta^*} \sim \frac{1}{\hat{t}^2} \sim \frac{1}{(1 - cos\theta)^2} \tag{4.3}$$

On the other hand, however, in the s-channel, angular distributions of dijet resonances are much flatter in $cos\theta^*$ than in QCD, with the exact angular distribution depending on the spin of the resonance and the final state partons. To optimize the $|cos\theta^*|$ cut, the integrated signal, S, over the square root of the integrated background, $\sqrt{B}$ should be maximized. However, there is another way to optimize the cut, which is using Eq. 4.3 as background and signal flat in $|cos\theta^*|$ which result in the optimal cut $|cos\theta^*_{MAX}|$=$1/\sqrt{3}$. In out dijet search, this cut on the scattering angle leads to the cut on the $|\Delta\eta|$

4.7 b-tagging Algorithm

Bottom quark (b quark) identification plays a very important role at the LHC and many analyses highly rely on the accuracy and efficiency of the identification. A b-tagging can be used in suppressing a contribution from background processes with non-b jets in the final state and enhance the sensitivity with signals with b jet in the final state.

These algorithms utilize the properties of b hadron decays, such as the presence of displaced vertex in a jet, and the values of transverse impact parameters of tracks associated with this jet. Fig 4.3 show the decay of a heavy-flavor jet with a secondary vertex(SV) from the decay of a b hadron and resulting in a charged particle track, which displaced with respect to the primary interaction vertex (PV), resulting in a large impact parameter (IP) value. The displacement of tracks with respect to the primary vertex is characterized by their impact parameter, which can be defined as the distance between the primary vertex and the tracks at their points of closest approach. The vector pointing to the point of closest approach from the PV is referred to as the IP vector. It can be defined as 3D or in the plane transverse to the beamline as a 2D vector. The longitudinal IP is defined along the beamline as 1D. The IP can be positive, means the track is upstream and the angle between the IP and the jet axis is smaller than $\pi/2$, or negative. Additionally, the b quarks are quite massive compare to lighter jets and massless gluons, which make the b jets have larger p_T relative to the jet axis than the other jets. Meanwhile, there is about 20% of the cases, the jet from b hadron decay contains a muon. Hence, the changed lepton presence can also be used to identify b-jet.

There properties make b jets discriminated from the other jets are listed bellow:

- **Long lifetime:** $\tau \approx 1.5\ ps, c\tau \approx 450\ \mu m, \gamma\beta c\tau \approx 1.8\ mm$ @ 20 GeV
- **Large mass:** around 5 GeV
- **High track multiplicity:** around 4 or 5
- **Large semileptonic branching fraction:** for electrons and muons, 20% each
- **Hard fragmentation function:** a large fraction of the original b quark momentum carried by the b hadron

From which, the b-tagging algorithms can use the above properties and set a likelihood of a jet contains a b hadron, which can be based on

- track information
- secondary vertex information

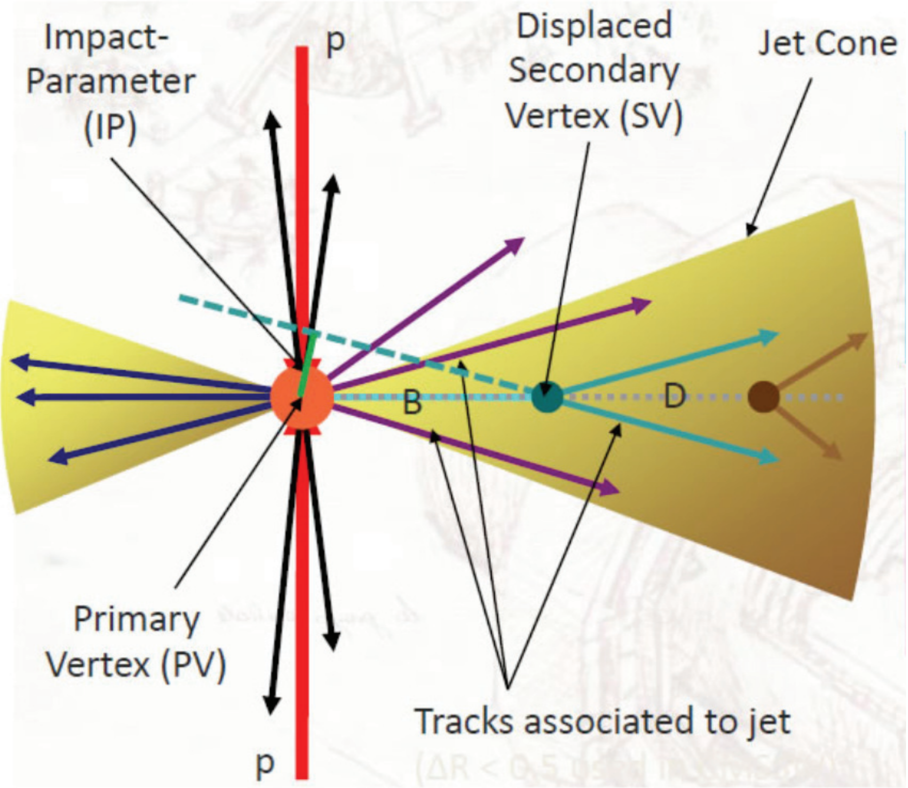


Figure 4.3: Illustration of a heavy-flavor jet with a secondary vertex(SV) from the decay of of a b hadron and resulting in charged particle track.

- soft lepton information
- some combination of the above

In addition, the performance of b-tagging should also be corrected by the difference in the tagging efficiency between MC and real data, by measuring the corresponding data-to-MC b-tagging efficiency scale factor (SF). The misidentification probability of a light flavor jet being identified as b jets should be measured as well, as the associated SF.

Besides identifying AK4 b-jets originated from the hadronization of b quarks, novel b-tagging also includes tagging subjects as wide-area jets(fat jets) to identify boosted hadronically decaying top quarks and Higgs bosons. It will retain a good tagging efficiency of the signal by rejecting a large amount of the background. In addition, in order to train a supervised machine to learn the algorithm, it's very important to have the label of each event. Simulated QCD and $t\bar{t}$ MC datasets are used for training.

During the early state of LHC Run II, CMS delivered about the total integrated luminosity of 36 fb^{-1} data. There were a few b-tagging algorithms are developed to improve the sensitivity of identifying b jets. The first tagger used was CSVv2 since Run II, which is based on CSV tagger in Run I [69]. After that, a new b-tagging algorithm that uses a deep learning algorithm was developed named DeepCSV [70]. These two taggers will be described in detail in Section 4.7.1. In Section 4.7.2, a new structure of the deep neural network (DNN) b-tagging algorithm will be described and have a better performance than the first two.

4.7.1 Previous b-tagger, CSVv2 and DeepCSV

The first b-tagger used in Run II was named CSVv2, which is developed based on the old CSV [71; 69] b-tagger used in Run I, and combined the information like displaced tracks, secondary vertices associated with the jet using a multivariate technique. There are two types of vertices used to decide if the CSVv2 algorithm exists. The inclusive vertex finding (IVF) will be considered first, otherwise, adaptive vertex reconstruction (AVR) [72] algorithm will be referred. At least two tracks are required in each jet. When calculating the values of track variables, it's required to have an angular distance with respect to the jet axis of $\Delta R < 0.3$. In addition, any combination of two tracks compatible with the mass of the K_S^0 meson in a window of 30 MeV is rejected. A default output discriminator value of -1 will be outputted if the jet has neither a selected track nor a secondary vertex. There are few steps in training CSVv2 algorithm. The first step, which is called 'training' of the algorithm, is to learn the features and combine them into discriminator value. It's important to ensure that no unwanted behavior is learned during this step. Also, to avoid the discrimination between jet flavors caused by different jet p_T and η distributions, distributions will be reweighted to obtain the same spectrum for all jet flavors. The training is performed on inclusive multijet events in three independent vertex categories:

- **RecoVertex:** one or more SV

- **Pseudo Vertex:** no SV, but has two tracks with a 2D IP and satisfy some other requirement.
- NoVertex: not assigned to the previous two categories.

When jets are in the RecoVertex category, the SV will be obtained using the IVF algorithm.

The following variables will be combined into the algorithm: ① SV 2D flight distance significance, ② number of SV, ③ track η_{rel}, ④ corrected SV mass, ⑤ number of tracks from SV, ⑥ SV energy ratio, ⑦ $\Delta R(SV, jet)$, ⑧ 3D IP significance of the first four tracks, ⑨ track $p_{T,rel}$ ratio, ⑩ track distance, ⑪ track decay length, ⑫ summed tracks E_T ratio, ⑬ $\Delta R(summedtracks, jet)$, ⑭ first track 2D IP significance above c threshold, ⑮ The number of selected tracks, and ⑯ The jet p_T and η.

The above variables will in each category be feed into a feed-forward one hidden layer multilayer perceptron. The number of nodes equals twice the number of inputs. The output from each category will be combined into a likelihood with account the fraction of jets of each flavor expected in $t\bar{t}$ events.

In a summary, the main improvement of CSVv2 from CSV is the 1. The IVF secondary vertex reconstruction algorithm, more input variables, multilayer perceptron (MLP), and jet p_T and η dependence. Fig. 4.4 show the workflow of CSVv2. Firstly, tracks will be associated with a jet inside a certain ΔR, then jets after selection criteria will be passed to track-based tagging algorithm. These selected tracks will be passed to the vertex reconstructor as previously mentioned to reconstruct secondary vertices. Together with the tracks, these secondary vertices will be combined to train combined tagging algorithms, which is CSVv2.

After the success of CSVv2, a new tagger was developed named DeepCSV to utilize the advantage of model DNN structure and extended the range of maximum considered tracks per jets. The same tracks and IVF SVs are used in the DeepCSV, as CSVv2 tagger, as well as the same input variables. Only the rack-based variables are set up to six tracks in the training. Also, the distribution of all input variables is preprocessed to center the mean of each distribution around zero and to obtain a root-mean-square value of unity, as known as normalized, which can speed up the train speed and avoid the gradient vanish and exploit during training. The training

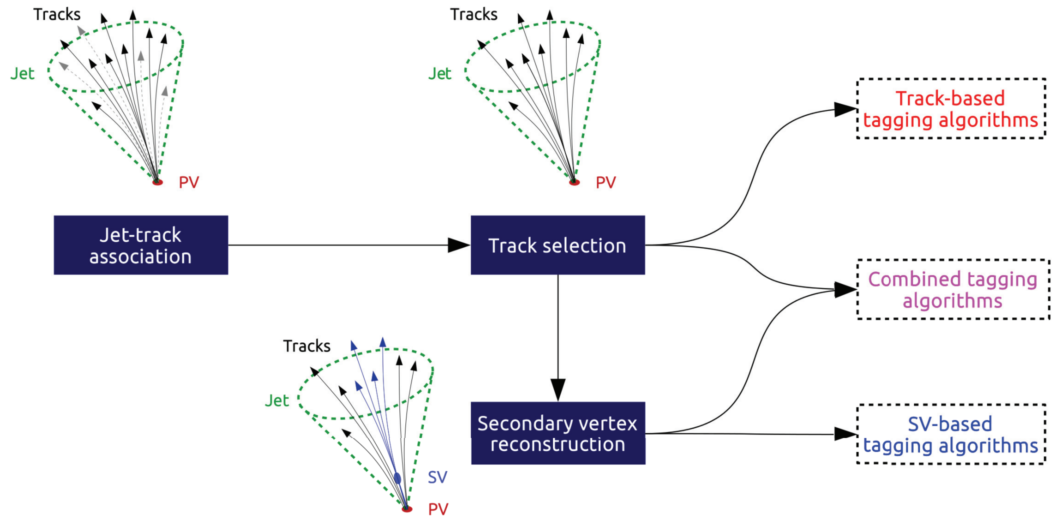


Figure 4.4: Work flow of several b-taggers

of DeepCSV used Keras [73] deep learning library, interfaced with the TensorFlow. The neural network used in the CVSv2 extends to 100 nodes with four hidden layers instead of one. Each layer uses ReLU as the activation function. The output layer of the neural network uses normalized exponential function as activation function so we can interpret the result to likelihood. The output has five nodes, corresponding to 5 jet flavor categories used in the training. The five categories are exact one b hadron, two b hadrons, no b hadron, exact one c hadron, and two c hadrons. Each category is independent of each other. In the analysis, we use P(b)+P(bb) discriminator to tag b jet. It has been checked that summing the probabilities for these two categories is equivalent to using a combined training for these categories.

4.7.2 DeepJet b-tagger

As DeepCSV greatly improve the performance of CSVv2 b-tagger. However, it's just a new version of CSVv2, with a small adjustment. After that, people start to study applying different types of DNN to identify the particle in a jet. Some applied convolutional or dense networks to study the analogy of the calorimeter cells to pixels [74; 75; 76; 77]. Some proposed recurrent neural networks [70; 78]. CMS [79] and ATLAS [80] released their own analysis note to apply vast DNNs in the context of flavor tagging. Based on these studies, a new b-tagger is developed in the of jet tagging and regression for jets with radii of 0.4 and 0.8. In this analysis and this section, only the algorithm with 0.4 radii is considered.

QCD and $t\bar{t}$ datasets are used as training samples. POWHEG [81] are used to generate $t\bar{t}$ samples, and PYTHIA are usd for both QCD and $t\bar{t}$ sample generation. In the dataset, heavy flavor hadrons, which are scaled to negligible transverse momentum so that it has no impact on the final properties of jets, are added up to the list of stable particles. Jet containing one or more heavy flavor objects in the constituents are assigned one of the major heavy flavor labels, afterwards, they are subdivided into minor labels to separate different decays or the number of heavy-flavor hadrons in a jet. Jet with no heavy flavor objects is assigned to an independent category. The scale of each feature will be normalized as previously discussed, while p_T and η are direct feed into the neural network.

Features of DeepJet algorithm training are separated into 4 parts.

The first part is the 15 global features, which are with per-jet values, with no upper or lower bounds are applied. They are: ① p_T, ② η, ③ N_{cPF}, ④ N_{nPF}, ⑤ N_{SV}, ⑥ N_{PV}, ⑦ trackSumJetEtRatio, ⑧ trackSumJetDeltaR, ⑨ vertexCategory, ⑩ trackSip2dValAboveCharm, ⑪ trackSip2dSigAboveCharm, ⑫ trackSip3dValAboveCharm, ⑬ trackSip3dSigAboveCharm, ⑭ jetNSelectedTracks, and ⑮ jetNTracksEtaRel,

The second part is the 18 charged PF candidate features, some with upper bounds and lower bounds, but some are not: ① trackEtaRel, ② trackPtRel, ③ trackPPar, ④ trackDeltaR, ⑤ trackPParRatio, ⑥ trackSip2dVal, ⑦ trackSip2dSig, ⑧ trackSip3dVal, ⑨ trackSip3dSig, ⑩ trackJetDistVal, ⑪ trackJetDistSig, ⑫ $p_T(cPF)/p_T(j)$, ⑬ $\Delta R(cPF, SV)$, ⑭ fromPV, ⑮ VTXass, ⑯ $\omega_p(cPF)$, ⑰ χ^2, and ⑱ quality,

The third part is the 6 neutral PF candidate features: ① $p_T(nPF)/p_T(j)$, ② $\Delta R(nPF, SV)$ ③ isGamma, ④ hadFrac, ⑤ $\Delta R(nPF)$, ⑥ $\omega_p(cPF)$.

The forth part is the 12 secondary vertex features: ① $p_T(SV)$, ② $\Delta R(SV)$, ③ m_{SV}, ④ $N_{track}(SV)$, ⑤ $\chi^2(SV)$ ⑥ $\chi^2_n(SV)$, ⑦ $d_{xy}(SV)$, ⑧ $S_{xy}(SV)$, ⑨ $d_{3D}(SV)$, ⑩ $cos\theta(SV)$, ⑪ $E_{rel}(SV)$, and ⑫ $S_{3D}(SV)$.

In these events, the particles and vertices are sorted using a hierarchical sorting algorithm. Charged candidates and secondary vertices are sorted by impact parameter significance.

Multiple types of DNN, including MLP, Convolutional Neural Network(CNN), and Recurrent Neural Network(RNN). CNN is spread in image recognition, where they effectively summarize small regions of the image and build more useful features than the raw pixels, like edges or alike, which then are feed to the following layers. Meanwhile, RNN has very good performance in processing dynamic features, especially dynamic in time in this case.

In the structure of DeepJet training, the first layer is convolutional layers and in order to allow non-linearities, four layers are used. The convolution is done using a 1 * 1 filter to reduce computational cost if applied to individual particle candidates. In these layers, a rectified linear unit(ReLU) is used as an activation function. The output from convolutional layers is a sequence of features of particle candidates, which is ordered from the input particle candidate sorting by displacement significance. This sequence of features will be then feed into long-short-memory RNN and then be compressed into a single vector per sequence. the output from the RNN is then combined with global variables and feed into MLP with 8 layers, whose first layer has 200 nodes, and the rest has 100. Like the previous, ReLU is used as an activation function.

Between each MLP layers, dropout layers with 0.1 rate and batch normalization layers are used each layer. For the final layer, a softmax function is used as an activation function and cross-entropy as the loos for minimization. For each minimization, adam optimizers are used and train for 50 epochs. The full architecture is illustrated in Fig. 4.5

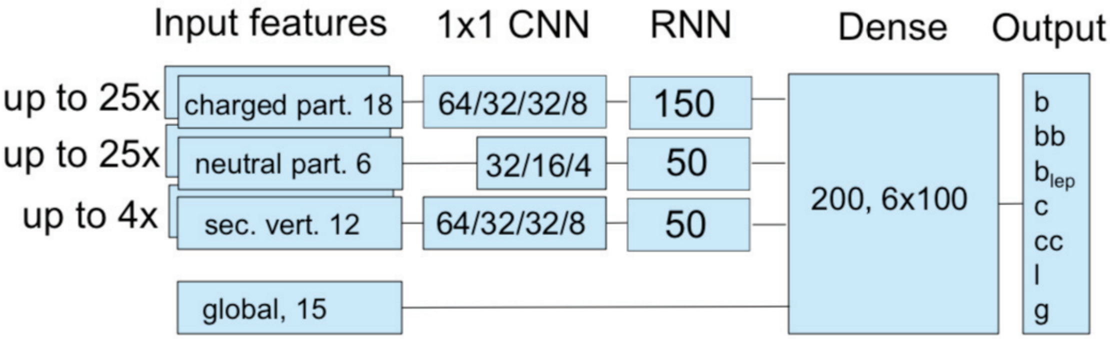


Figure 4.5: Architecture of AK4 b-tag Deep Neural Network, except for the Dropout and batch normalization layers, while the number in each box indicates the number of nodes per layer

CHAPTER 5

SEARCH FOR HEAVY RESONANCES IN PP COLLISION AT $\sqrt{s} = 13$ TEV USING 2016 DATA

This analysis focus on the search for dijet resonances in the high mass region, using a data sample of pp collisions at $\sqrt{s} = 13$ TeV collected in 2016, corresponding to an integrated luminosity of 36 fb^{-1}. All Jets are required to have $p_T > 30$ GeV and $\eta < 2.5$. Spatially close jets are combined into "wide jets" and used to determine the dijet mass. The background from t-channel dijet events peaks at large values of $|\Delta\eta_{jj}|$ and is suppressed by requiring the pseudorapidity separation of the two wide jets to satisfy $|\Delta\eta_{jj}| < 1.3$. The above requirements maximize the search sensitivity for isotropic decays of dijet resonances in the presence of QCD dijet background.

The published analysis search for resonance in both low and high mass search The low mass search uses data collected in the CMS Calo Scouting stream in 2016, which uses a set of scouting triggers and reconstructed in the HLT level, and high mass search uses RECO data. The low mass searches are not the target of this analysis, details can be found in [82], They will not be talked about too much.

5.1 Data and Monte Carlo Samples

Many standard MINIAOD data samples of a different run period of 2016 are used in this analysis, named B,C,E,D,F,G, and H, summarized in Table. 5.1. Some of the data sets listed may not contain certified luminosity and thus do not contribute to the result. The total integrated luminosity calculated with normtag_DATACERT.json as of 23 Jan 2017 was 36.8 fb^{-1}. The total integrated luminosity after correction for the 10 Feb 2017 re-evaluation is 35.9 fb^{-1}.

The signal samples were produced with narrow resonances (i.e., the resonance width is much smaller than the experimental mass resolution) and listed in table. 5.2. Even though these samples are not the final set of the BSM models, but they are the generic shape of the qq final state.

Table 5.1: Data sets used in the high-mass RECO Analysis.

Dateset
/JetHT/Run2016B-23Sep2016-v1/MINIAOD
/JetHT/Run2016B-23Sep2016-v3/MINIAOD
/JetHT/Run2016C-23Sep2016-v1/MINIAOD
/JetHT/Run2016D-23Sep2016-v1/MINIAOD
/JetHT/Run2016E-23Sep2016-v1/MINIAOD
/JetHT/Run2016F-23Sep2016-v1/MINIAOD
/JetHT/Run2016G-23Sep2016-v1/MINIAOD
/JetHT/Run2016H-PromptReco-v1/MINIAOD
/JetHT/Run2016H-PromptReco-v2/MINIAOD
/JetHT/Run2016H-PromptReco-v3/MINIAOD

5.2 Trigger

5.2.1 High-mass RECO Analysis

We use all jet data in the JetHT dataset for the analysis, regardless of what trigger is used to record an event. In the full 2016 dataset, our main triggers PFHT800 and PFHT900 suffer from the HLT trigger inefficiencies of post-ICHEP data taking. Thus for studying the trigger efficiency of the JetHT dataset for our analysis we include all the un-prescaled triggers to the study, which are PFHT800, PFHT900, PFJet500, CaloJet500_NoJetID, and AK8PFJet450. We measure trigger efficiency mainly with respect to SingleMuon45, as the low-energy PFHT triggers are also affected by the efficiency problems and thus are not considered reliable. Turn-on studies with respect to PFHT475 are shown in Fig. 5.2 as supporting material.

In the 2016 ICHEP analysis [65] the trigger was fully efficient for dijet masses above 1058 GeV. For the full 2016 analysis we observe the trigger turn-on to be significantly affected by the HLT trigger issues (see Fig. 5.1). The bin where the statistical uncertainty of dijet mass spectrum and trigger efficiency is of similar magnitude starts at 1246 GeV. The statistical uncertainty in the bin 1246 – 1313 GeV is 0.08% and the trigger inefficiency at the center of this bin is 0.05%.

Table 5.2: MC datasets for dijet resonances in RunIISpring16 with number of events.

Dataset	Events
/RSGravitonToQuarkQuark_kMpl01_M_500_TuneCUETP8M1_13TeV_pythia8/ RunIISpring16MiniAODv2-PUSpring16_80X_mcRun2_asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM	99 004
/RSGravitonToQuarkQuark_kMpl01_M_750_TuneCUETP8M1_13TeV_pythia8/ RunIISpring16MiniAODv2-PUSpring16_80X_mcRun2_asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM	99 598
/RSGravitonToQuarkQuark_kMpl01_M_1000_TuneCUETP8M1_13TeV_pythia8/ RunIISpring16MiniAODv2-PUSpring16_80X_mcRun2_asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM	99 106
/RSGravitonToQuarkQuark_kMpl01_M_2000_TuneCUETP8M1_13TeV_pythia8/ RunIISpring16MiniAODv2-PUSpring16_80X_mcRun2_asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM	
/RSGravitonToQuarkQuark_kMpl01_M_3000_TuneCUETP8M1_13TeV_pythia8/ RunIISpring16MiniAODv2-PUSpring16_80X_mcRun2_asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM	100 000
/RSGravitonToQuarkQuark_kMpl01_M_4000_TuneCUETP8M1_13TeV_pythia8/ RunIISpring16MiniAODv2-PUSpring16_80X_mcRun2_asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM	99 817
/RSGravitonToQuarkQuark_kMpl01_M_5000_TuneCUETP8M1_13TeV_pythia8/ RunIISpring16MiniAODv2-PUSpring16_80X_mcRun2_asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM	100 000
/RSGravitonToQuarkQuark_kMpl01_M_6000_TuneCUETP8M1_13TeV_pythia8/ RunIISpring16MiniAODv2-PUSpring16_80X_mcRun2_asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM	99 901
/RSGravitonToQuarkQuark_kMpl01_M_7000_TuneCUETP8M1_13TeV_pythia8/ RunIISpring16MiniAODv2-PUSpring16_80X_mcRun2_asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM	99 632
/RSGravitonToQuarkQuark_kMpl01_M_8000_TuneCUETP8M1_13TeV_pythia8/ RunIISpring16MiniAODv2-PUSpring16_80X_mcRun2_asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM	99 956
/RSGravitonToQuarkQuark_kMpl01_M_9000_TuneCUETP8M1_13TeV_pythia8/ RunIISpring16MiniAODv2-PUSpring16_80X_mcRun2_asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM	100 000

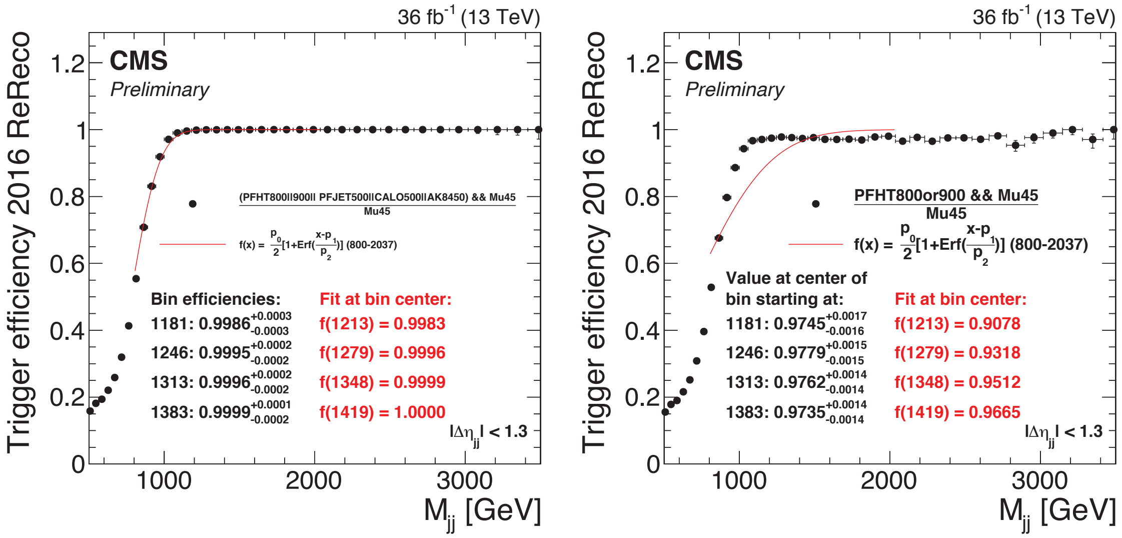


Figure 5.1: Efficiency of analysis triggers is measured against SingleMuon45 trigger. Turn-on is studied using all the un-prescaled triggers in the JetHT dataset, i.e. PFHT800, PFHT900, PFJet500, CaloJet500_NoJetID and AK8PFJet450. We consider triggers fully efficient starting from 1246 GeV. For showing the effect of the back-up triggers, efficiency is measured with only PFHT800 and PFHT900 on the right.

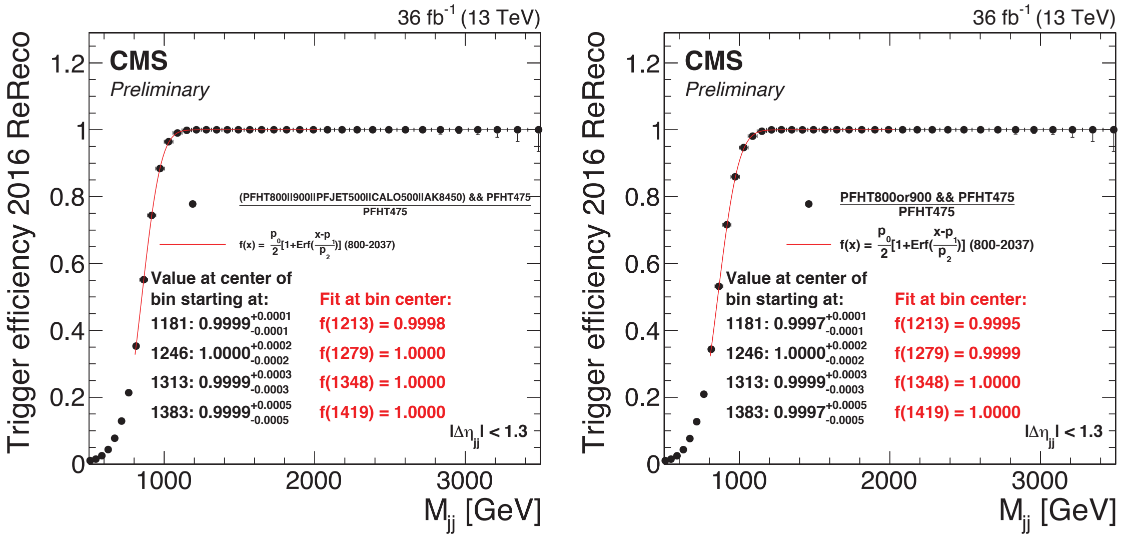


Figure 5.2: Efficiency of all the unprescaled triggers is measured w.r.t to PFHT475 on the left and PFHT800+PFHT900 w.r.t PFHT475 on the right. As the the turn-on is sharp in both cases, comparison with the study w.r.t SingleMuon45 trigger shows that the PFHT475 is not entirely reliable trigger to study the efficiency with.

5.3 Event Reconstruction and Selection

The baseline event selection follows the standard of previous CMS dijet searches [65].

The jet momenta and energies are corrected using calibration constants described in Section 5.4. All jets are required to have $p_T > 30$ GeV and $|\eta| < 2.5$. The two jets with the largest p_T are defined as the leading jets. In this analysis, we use the standard PF jet ID.

Geometrically close jets are combined into "wide jets" and used to determine the dijet mass, as in our previous dijet searches. The wide-jet algorithm, designed for dijet resonance event reconstruction, reduces the analysis sensitivity to gluon radiation from the final state partons. The two leading jets are used as seeds and the four-vectors of all other jets, if they are within $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 1.1$ of the seed jet, and are added to the nearest leading jet to obtain two wide jets. The above requirements maximize the search sensitivity for isotropic decays of dijet resonances in the presence of a QCD dijet background.

The pseudorapidity separation of the two wide jets is required to be $|\Delta\eta_{jj}| < 1.3$.

This requirement suppresses the QCD t-channel, while the signal contribution significantly comes from the s-channel. It also restricts the region of our measurement predominantly within the barrel region.

5.4 Jet Energy Corrections

For the high-mass RECO analysis, we use the most recent corrections from the JEC group, provided in the release Spring16_23Sep2016V1_DATA, consisting of L1+L2+L3 MC truth based corrections and the L2L3residual corrections derived from in-situ measurements of RECO data. The corrections are applied in four different intervals of validity (IOV) as recommended by the JEC group.

The JEC layers used are L1FastJet_AK4PFchs, L2Residual_AK4PFchs, L3Absolute_AK4PFchs, and L2L3Residual_AK4PFchs, and the IOV versions used are BCDV1, for runs BCD, EV1 for run E, FV1 for run F early and GHV1 for runs F late, G and H.

5.5 Dijet Data and QCD Background Predictions

Since QCD production to multiple jets is the dominant background for this analysis, it's needed to perform a comparison between data to QCD background predictions for the dijet events. The QCD samples are generated with `PYTHIA` program with the CUETP8M1 tune and simulated by GEANT4. These samples are normalized to the data by multiplying by a factor 0.87 so that the total number of events in the samples agrees with the observed number in the data, with very few events in the low $\Delta\phi$ region. It shows that the data sample is dominated by genuine parton-parton scattering, along with large backgrounds from detector noise or other nonphysical sources.

Fig 5.3 shows the distribution of azimuthal angle $\Delta\phi$ distribution of the two wide jets. In QCD production, the two jets are expected to be "back-to-back", which means $\Delta\phi = \pi$. We observed a similar peak in Fig 5.3.

The $|\Delta\eta|$ distribution of data and QCD samples are shown in Fig 5.4. It shows that $|\Delta\eta|$ is also dominated by t-channel QCD production of two jets, whose production rate increases with increasing $|\Delta\eta|$, whereas s-channel signals from most models of dijet resonances would decrease with increasing $|\Delta\eta|$.

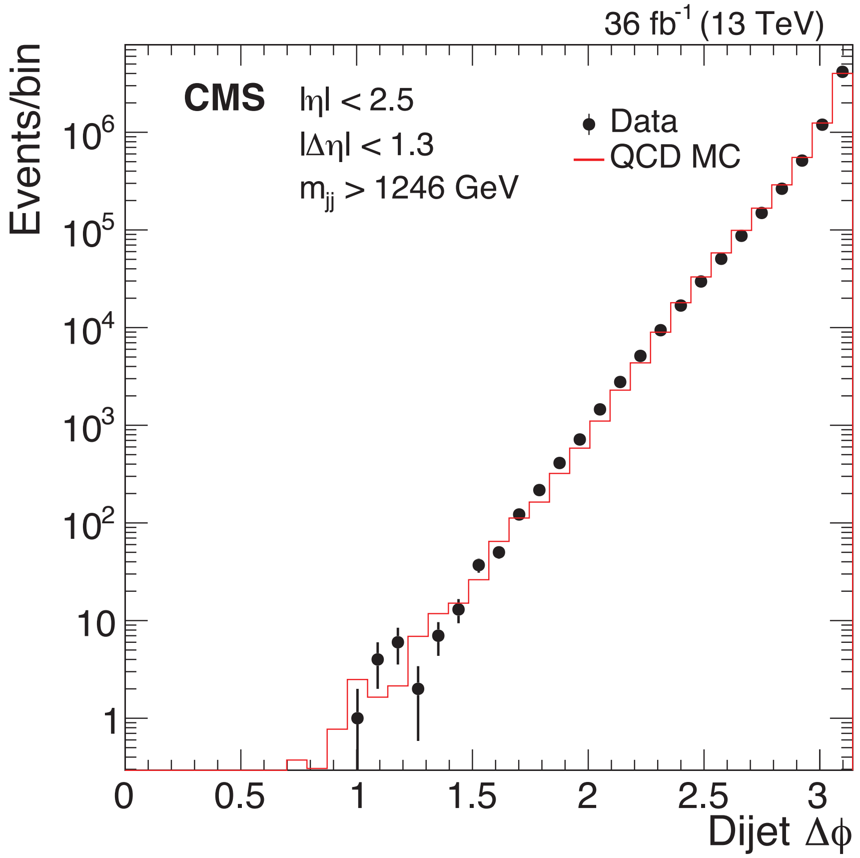


Figure 5.3: The azimuthal angular separation between the two wide jets (in radians) from the high-mass search. Data (points) are compared to QCD predictions from the `PYTHIA 8` MC including detector simulation (histogram) normalized to the data.

In Fig. 5.5, we observe the dijet mass distribution of data is has a similar shape as QCD samples generated by `PYTHIA`, which includes leading order (LO) QCD calculation and effects from parton shower. The number of dijet falls steeply and smoothly as a function of dijet mass.

we also compare the dijet mass data to next-to-leading order (NLO) QCD prediction from `POWHEG` 2.0 normalized to the data, shown in Fig.5.6.

It is shown that the dijet mass of data has the same approximate behavior as expected from QCD predictions. However, the QCD calculations' intrinsic uncertainties make them unreliable estimators of the backgrounds in dijet resonance searches. Instead, we will use the dijet data to estimate the background.

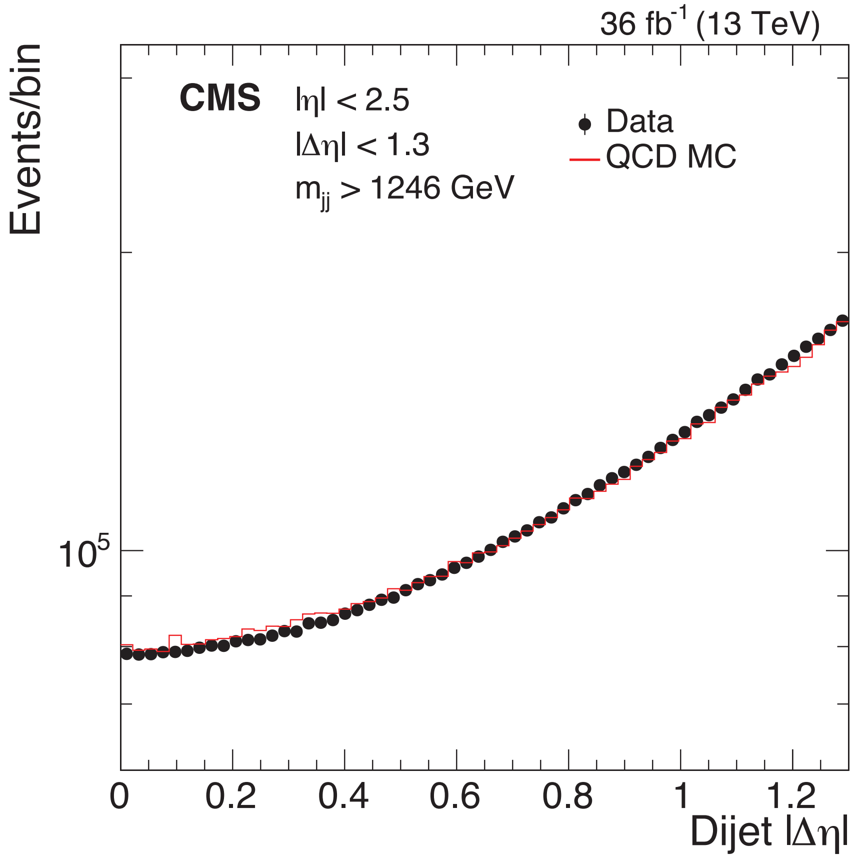


Figure 5.4: The pseudorapidity separation between the two wide jets from the high-mass search. Data (points) are compared to QCD predictions from the `PYTHIA 8` MC including detector simulation (histogram) normalized to the data.

5.6 Dijet Mass Spectrum and Background Fit

The measured dijet mass spectrum is defined by the following function:

$$\frac{d\sigma}{dm} = \frac{1}{\int Ldt} \frac{N_i}{\Delta m_i} \tag{5.1}$$

where m is the dijet mass; N_i is the number of events in the i-th dijet mass bin, Δm_i is the width of the i-th dijet mass bin, and the integrated luminosity is $\int Ldt$. The bin width is approximately the dijet mass resolution, and gradually increases as a function of mass.

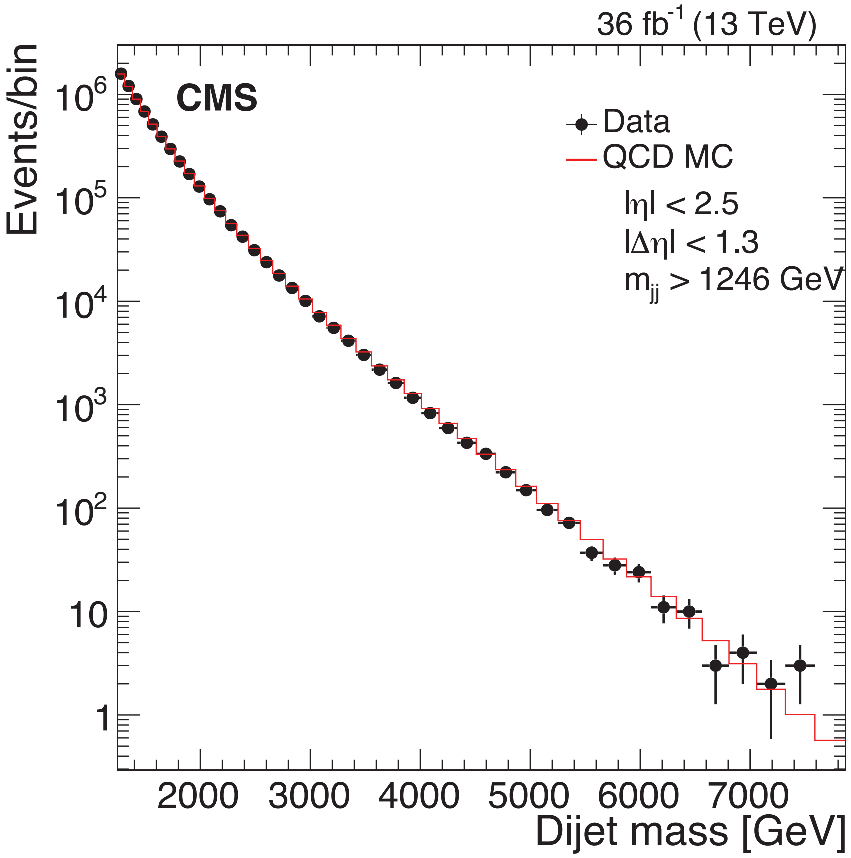


Figure 5.5: The dijet mass of the two wide jets from the high-mass search. Data (points) are compared to QCD predictions from the `PYTHIA 8` MC including detector simulation (histogram) normalized to the data. The horizontal lines on the data points show the variable bin sizes.

5.6.1 Fit Method

To test the smoothness, we fit the dijet mass spectrum separately for the high-mass search with a background-only function. The fit function for the high-mass analysis, defined as

$$\frac{\mathrm{d}\sigma}{\mathrm{d}m_{jj}} = \frac{p_0(1 - m_{jj}/\sqrt{s})^{p_1}}{(m_{jj}/\sqrt{s})^{p_2 + p_3 \log(m_{jj}/\sqrt{s})}} \ , \tag{5.2}$$

is a four-parameter continuous function that has been used extensively in previous searches to fit the falling dijet spectrum [83; 84].

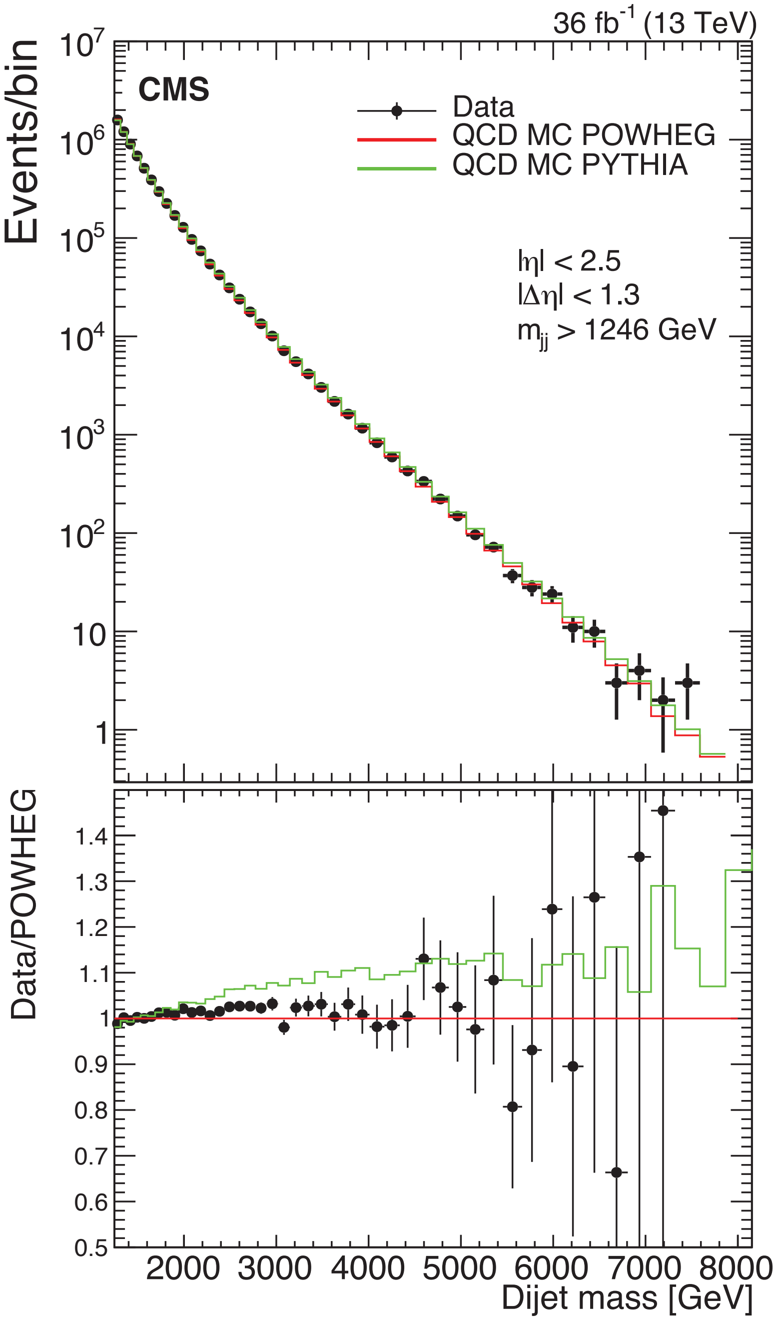


Figure 5.6: The dijet mass distribution of the two wide jets from the high-mass search. (Upper) Data (points) are compared to predictions from the `POWHEG` MC in red (darker) and the `PYTHIA 8` MC in green (lighter), including detector simulation, each normalized to the data. (Lower) The ratio of data to the `POWHEG` prediction, compared to unity and compared to the ratio of the `PYTHIA 8` MC to the `POWHEG` prediction. The horizontal lines on the data points show the variable bin sizes.

We perform an extended, background-only, binned, maximum likelihood fit to the 2016 data with the following likelihood,

$$\mathcal{L}(\text{data}|\boldsymbol{\theta}) = \prod_{i=1}^{n_b} \text{Poisson}(x_i|b_i(\boldsymbol{\theta})) = \prod_{i=1}^{n_b} \frac{b_i(\boldsymbol{\theta})^{x_i} e^{-b_i(\boldsymbol{\theta})}}{x_i!} \ , \tag{5.3}$$

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 expected events N_b,

$$b_i(\boldsymbol{\theta}) = N_b \int_{m_{jj}^{\min,i}}^{m_{jj}^{\max,i}} \mathrm{d}m_{jj} \ p(m_{jj}) \ , \tag{5.4}$$

where $p(m_{jj})$ is the fit function in equations 5.2 normalized to unity.

To test the goodness-of-fit (GOF), we define two test statistics. One is a modified chi-square,

$$\chi^2 = \sum_{i=1}^{n_b} \left(\frac{x_i - b_i}{\sigma_{x_i}} \right)^2 \ , \tag{5.5}$$

where we define the "uncertainty" σ_{x_i} in terms of the 68% CL region of a Poisson distribution. Technically, the definition is as follows, setting $\alpha = 1 - 0.687$,

$$\sigma_{x_i} = \begin{cases} D_c^{-1}(\alpha/2, x_i + 1) \ , & \text{if } b_i > x_i \\ D^{-1}(\alpha/2, x_i) \ , & \text{if } b_i < x_i \end{cases} \tag{5.6}$$

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

$$D(\alpha/2, x_i) = \int_{-\infty}^{\alpha/2} \frac{1}{\Gamma(x_i)} z^{x_i - 1} e^{-z} \ \mathrm{d}z \ , \tag{5.7}$$

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

$$D_c(\alpha/2, x_i + 1) = \int_{\alpha/2}^{+\infty} \frac{1}{\Gamma(x_i + 1)} z^{x_i} e^{-z} \ \mathrm{d}z \ . \tag{5.8}$$

Similarly, an alternative test statistic, which is a likelihood ratio with respect to the *saturated model* can also be defined,

$$-2\log\lambda(\boldsymbol{\theta}) = -2\log\frac{\prod_{i=1}^{n_b}\text{Poisson}(x_i|b_i(\boldsymbol{\theta}))}{\prod_{i=1}^{n_b}\text{Poisson}(x_i|x_i)} = 2\sum_{i=1}^{n_b}\left[b_i(\boldsymbol{\theta}) - x_i + x_i\log\frac{x_i}{b_i(\boldsymbol{\theta})}\right]. \tag{5.9}$$

In both cases, we generate 10,000 pseudodatasets from the best-fit model parameters on data, refit each pseudodataset with a maximum likelihood fit, and save the test statistic value.

5.6.2 Fit Results

The result of the background-only fit to the data is shown in Fig. 5.7. defined as the observed number of events in each bin divided by the integrated luminosity and the bin width, with predefined bins of a width corresponding to the dijet mass resolution. The chi-squared per number of degrees of freedom of the fit is χ^2/NDF = 38.9/39. Fig. 5.7 also shows examples of dijet mass distributions for signal events generated with the `PYTHIA` program with the CUETP8M1 tune and including a GEANT4-based simulation of the CMS detector. The quark-quark (qq) resonances are modeled by qq → G → qq, the quark-gluon (qg) resonances are modeled by qg → q^* → qg, and the gluon-gluon (gg) resonances are modeled by gg → G → gg. The predicted mass distributions have Gaussian cores from jet energy resolution, and tails towards lower mass values primarily from QCD radiation. The contribution of the low mass tail to the lineshape depends on the parton content of the resonance (qq, qg, or gg). Resonances containing gluons, which emit more QCD radiation than quarks, are wider and have a more pronounced tail. The signal distributions are shown in Fig. 5.7 are for qq, qg, and gg resonances with signal cross sections corresponding to the limits at 95% confidence level (CL) obtained by this analysis. There is no evidence for a narrow resonance in the data. The lower panel shows the difference between the data and the fitted parametrization, divided by the statistical uncertainty of the data.

In Fig. 5.8, we show the distribution of both test statistics from these pseudo experiments as well as the value observed in data. For the first test, the observed value has a p-value around 38% and the effective number of degrees of freedom (from fitting a χ^2 distribution) is 37.1±0.1, while for the second test the p-value is

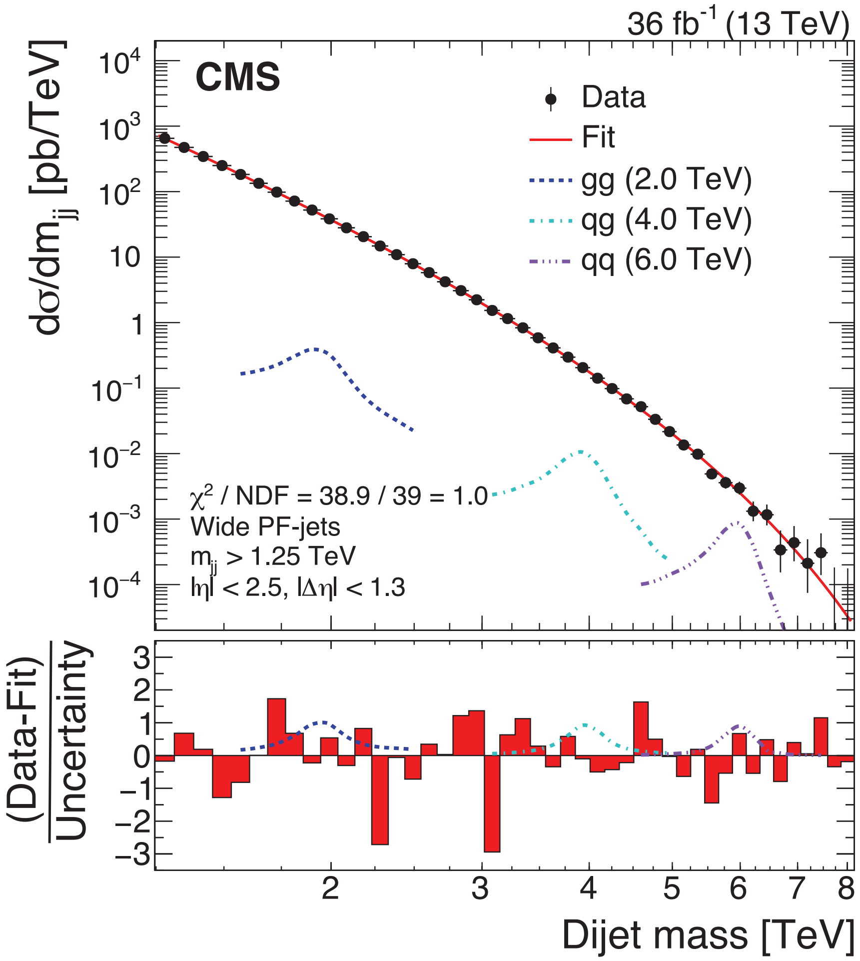


Figure 5.7: The binned fit for RECO performed in the range 1246 ¡ m_{jj} ¡ 8152 GeV. The signal shapes shown are normalized to the excluded cross section for each resonance type at that mass.

around 43% and the effective number of degrees of freedom is 42.4±0.1.

5.7 Signal Shapes

5.7.1 High-mass RECO analysis

The signal resonance shape depends on the decay mode (gg, qg, qq, or Gaussian lineshape). Because we only consider narrow resonances, where the natural

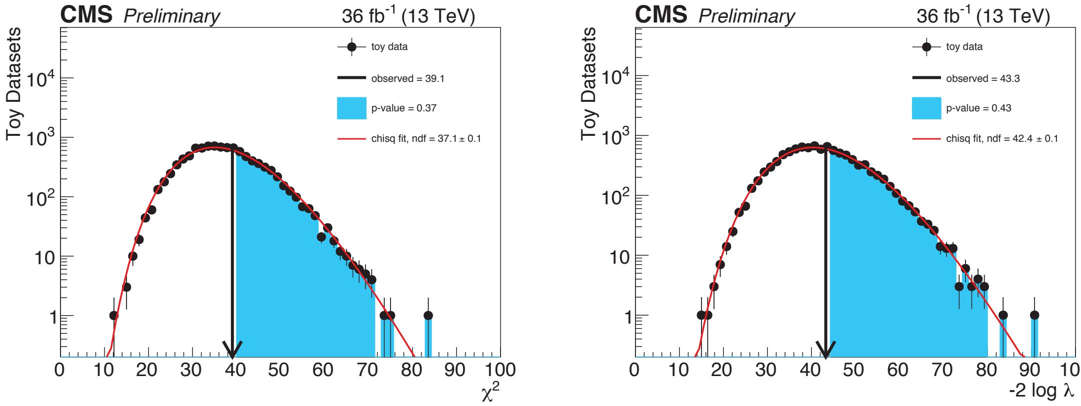


Figure 5.8: Toy distribution for the goodness-of-fit study using the χ^2 test statistic (left) and the $-2\log\lambda$ test statistic (right) as defined in the text at sec.5.6.1

resonance width is negligible compared to the dijet mass resolution at CMS, the natural width does not affect the resonance shape. The generic shapes for the two types of decay modes were produced from samples of the processes gg → G → gg, qg → q* → qg, and qq → G → qq for 11 mass points: 500 GeV, 750 GeV, and 1 TeV to 9 TeV in 1 TeV spacing. Some of the resonance shapes are shown in Fig. 5.9.

The shapes of resonances with gg, qg, and qq final states differ due to the differences between quarks and gluons. Gluons emit more radiation than quarks, which casuses the resonance width to increase with the number of gluons in the final state. The low-mass tail of the resonance shape is due to the effect of final state radiation and the parton distribution function (PDF) having higher parton luminosity at low mass than at high mass. The high-mass tail is smaller and is caused by initial state radiation. These resonance shapes are approximately valid for any model of resonances involving the same pairs of partons, assuming the relative half-width of the model $((\Gamma/2)/M)$ is small compared to the dijet mass resolution. In this analysis, only the signal shape of quark-quark final state is used.

5.8 Systematic Uncertainties

5.8.1 Jet Energy Scale

The uncertainty on the Jet Energy Scale (JES) is the uncertainty that comes from the difference between simulation and data on the scale. If the simulated scale

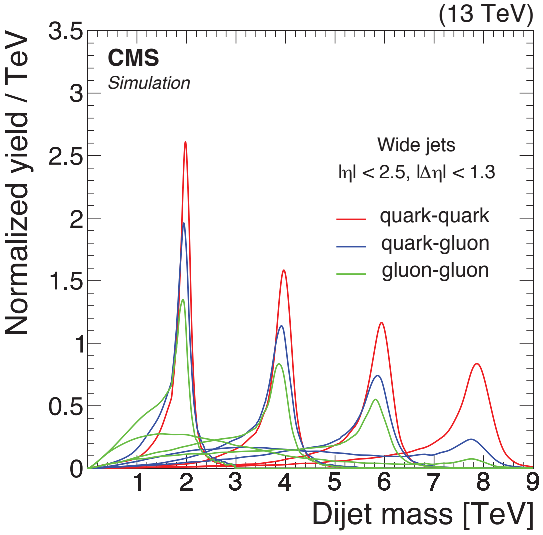


Figure 5.9: Signal shapes for gg → G → gg, qg → q^* → qg, and qq → G → qq resonances. The integral of the shapes has been normalized to unity.

has a high response, then the position of simulated resonance will be placed at the higher mass region than it should be placed. The JES uncertainty is only on the simulated signal but the background since the background is coming from the only fit of data in our process. So this uncertainty is the uncertainty about the position of the simulated resonance signal. The 2011 official CMS JES uncertainty is studied and provided with Jet energy correction.

5.8.2 Jet Energy Resolution

The uncertainty on the Jet Energy Resolution (JER) is the uncertainty which also comes from the difference between simulation and data on the jet energy resolution. If the simulated JER is bigger than the JER from data, the width of the simulated resonance is broader than the real resonance. The JER uncertainty is only on a simulated signal like JES uncertainty. The uncertainty on JER at high dijet mass is ±10%. The tails of the resolution function are in good agreement between data and MC.

5.8.3 Luminosity Uncertainty

This is the uncertainty which shows the error of measuring the total luminosity

5.8.4 Total Systematic Uncertainty

The systematic uncertainties propagated to the signal prediction are shown in table 5.3.

Table 5.3: Summary of systematic uncertainties propagated to the signal prediction in RECO analysis.

Systematic Uncertainty Source	Nominal Value	Uncertainty
Jet Energy Scale	no shift	±2% shift of m_{jj}
Jet Energy Resolution	no smearing	10% of RECO resolution
Luminosity	137.5 fb^{-1}	±2.6%

5.9 Result

The CLs criterion [85; 86] is used to determine the 95% CL limit on the signal contribution in the data, using the `RooStats` package implemented in the CMS `combine` tool. The `Asymptotic` method [87] is used to calculate preliminary 95% C.L. upper limits with 1σ and 2σ bands using the CLs frequentist calculation currently recommended by the LHC Higgs Combination Group [88]. The `ProfileLikelihood` method is used for significance and the background p-value; finally, the `MaxLikelihoodFit` method allows to get the signal *best fit ratio*, the fit pulls and the pre/post fit distributions.

Results are obtained from a combined signal and background fit the binned m_X distribution, based on a profile likelihood defined as

$$\mathcal{L} = \prod_i \frac{\mu_i^{n_i} \cdot e^{-\mu_i}}{n_i!} \qquad \text{with} \qquad \mu_i = \sigma N_i(S) + N_i(B)$$

where $N_i(S)$ and $N_i(B)$ are the i-th signal and background events, and σ is the signal strength modifier parameter. While the signal shape parameters are held constant after the fit to the Monte Carlo signal, the background shape, and normalization parameters are still left free to float in the combined fit. Systematic uncertainties are treated as nuisance parameters and are profiled in the statistical interpretation. The background-only hypothesis is tested against the signal in the considered categories, and with no evidence of significant deviations from background expectation, the asymptotic modified frequentist method is used to determine the limit at the 95% CL on the signal contribution. The 95% CL upper limit, σ_{95}, is calculated from the posterior probability density P_{POST} as follows:

$$\frac{\int_0^{\sigma_{95}} P_{POST}(\sigma)d\sigma}{\int_0^{\inf} P_{POST}(\sigma)d\sigma} = 0.95 \tag{5.10}$$

Fig. 5.10 shows the limits from both the low mass search and the high mass search together. The boundary between the low mass and high mass search is shown by a vertical dashed line at a resonance mass of 1.6 TeV. Limits for resonances with mass less than or equal to 1.6 TeV are shown from the low mass search and limits for resonances with mass greater than or equal to 1.6 TeV are shown from the high mass search. The resonance mass value of 1.6 TeV was chosen to maintain a reasonable acceptance for the 1.246 TeV dijet mass requirement imposed by the high mass search to have full trigger efficiency. The observed 95% CL upper limits on the product of the cross-section, branching fraction, and acceptance for dijet resonances decaying to quark-quark. The corresponding expected limits (dashed) and their variations at the 1 and 2 standard deviation levels (shaded bands) are also shown. Limits are compared to predicted cross-sections for axigluons [12], colorons [13], scalar diquarks [14], new gauge bosons W' and Z' with SM-like couplings [15], and dark matter mediators for mDM= 1 GeV [16; 17].

The expected and observed exclusion are summarized in the table 5.4

In this analysis, searches for resonances decaying into a pair of jets have been performed using proton-proton collisions at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of up to 35.9 fb^{-1} has been performed. We set expected and observed limit on multiple models at 95% CL. The dijet mass spectra are observed to be smoothly falling distributions. In the analyzed data samples, there is no evidence for resonant particle production.

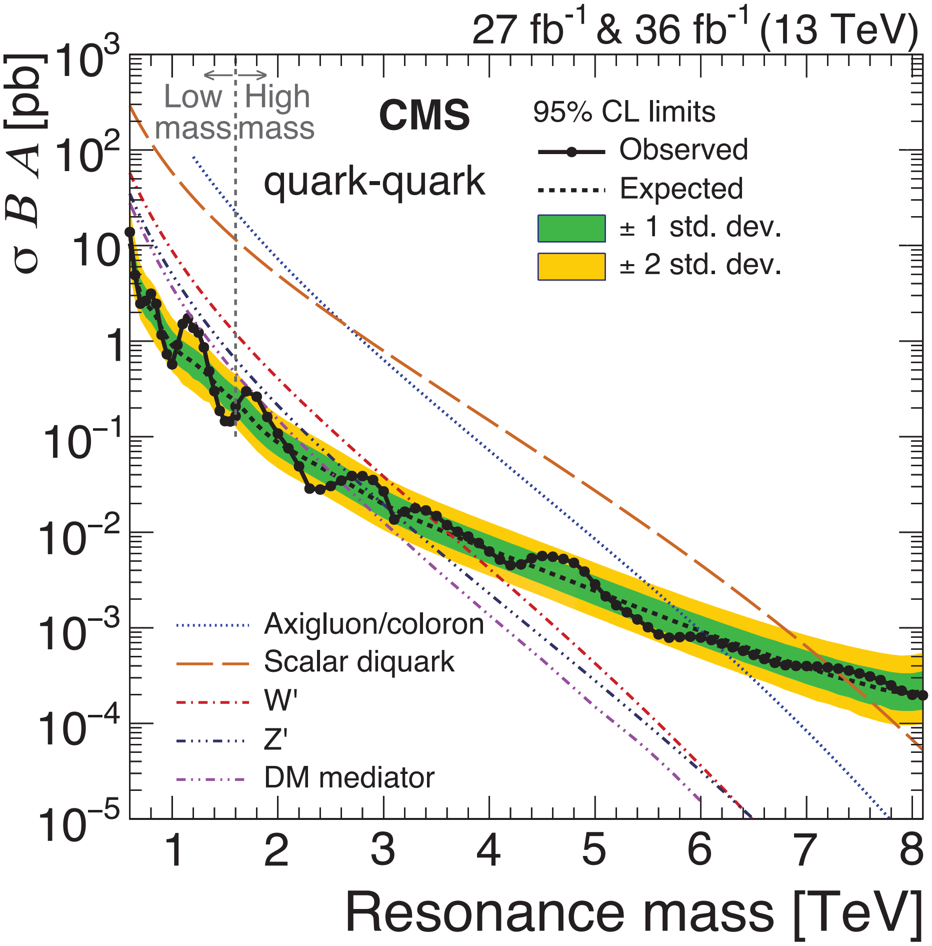


Figure 5.10: The observed 95% CL upper limits from both the low mass search and the high mass search on the product of the cross section, branching fraction, and acceptance for dijet resonances decaying to quark-quark. The dash line at 1.6 TeV indicates the starting mass of high mass search. The corresponding expected limits (dashed) and their variations at the 1 and 2 standard deviation levels (shaded bands) are also shown. Limits are compared to predicted cross sections for axigluons [12], colorons [13], scalar diquarks [14], new gauge bosons W' and Z' with SM-like couplings [15], and dark matter mediators for mDM= 1 GeV [16; 17].

5.10 Dark Matter Interpretations

In the simplified model which is recommended in [16], mediators only decay into qq and pairs of dark matter particles, with unknown mass m_{DM}. In this case, we set

Table 5.4: Summary of systematic uncertainties propagated to the signal prediction in RECO analysis.

Model	Final State	Observed (expected) Mass Limit [TeV]
Scalar Diquark	qq	7.2 (7.4)
Axigluon/Coloron	$q\bar{q}$	6.1 (6.0)
Z'	$q\bar{q}$	2.7 (2.9)
W'	$q\bar{q}$	3.3 (3.6)
Dark Matter Mediator ($m_{\mathrm{DM}} = 1$ GeV)	$q\bar{q}$	2.6 (2.5)

the universal quark coupling $g_q = 0.25$ and $g_{\mathrm{DM}} = 1.0$. In order to find which mass of dark matter and mediator is excluded, we do a scan over dark matter mass and mediator mass plane, using MadGraph5. The lowest dark matter mass is 1 GeV, indistinguishable from zero. We scan the dark matter mass between zero and 1.7 TeV in 50 GeV steps. We scan the mediator mass between 500 GeV and 4 TeV in 0.5 TeV steps. In the critical mass range between 1800 GeV and 2800 GeV, we use 50 GeV steps. We find the cross-section at the parton level from the `MG5` simulation. The production of the cross-section requires the following kinematic cuts $|\Delta\eta_{jj}| < 1.3$, $|\eta| < 2.5$, and the two jet mass (M_{jj}) within a 500 GeV mass window of the resonances mass for narrow resonances. The cross-sections as a function of dark matter mass from this `MG5` scan behaves as expected from the partial width equations 2.12-2.15.

Fig. 5.11 shows the cross-section as a function of mediator mass at four critical values of dark matter mass for both types of a mediator. All the cross-section values above the upper limit points are excluded by this search. The mediator cross-section increases with increasing dark matter mass. The curve dips where the mediator starts to decay into dark matter, which reduces the cross-section. Furthermore, the dip is more pronounced for the vector model where the dark matter decay channel becomes faster as a function of increasing mediator mass.

For each value of dark matter mass, we compare the cross-section as a function of

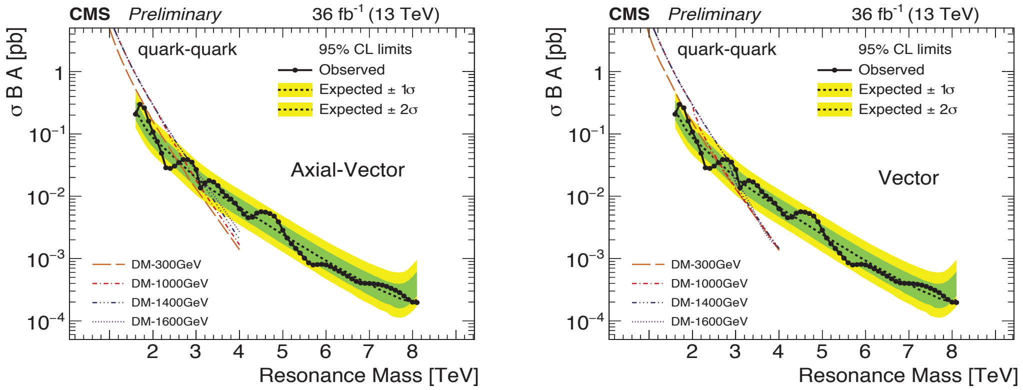


Figure 5.11: The 95% CL uppper limits on the product of the cross section, branching fraction, and acceptance for quark-quark type dijet resonances, compared with predicted cross sections of dark matter mediators for $m_{\mathrm{DM}} = 0.3, 1.0, 1.4, 1.6$ TeV, for both Vector and Axial-Vector model.

mediator matter mass to the 95% CL cross-section limit, to obtain a 2D exclusion area on the dark matter mass vs. mediator mass, for the axial-vector mediator and vector mediator, as shown in Fig. 5.12. To present the smoothest exclusions as a function of dark matter mass, the exclusions in Fig. 5.12 use the cross-sections calculated starting from `MG5` with $m_{\mathrm{DM}} = 1$ GeV, and the dark matter mass dependence of the mediator cross-section is determined analytically from the partial width equations 2.12-2.15.

Fig. 5.12 shows the the excluded values of mediator mass as a function of m_{DM} for both types of mediators, and Fig. 5.13 compares them with the exclusions in the mono-X channels from MET + X searches [89]. For $m_{\mathrm{DM}} = 1$ GeV, indistinguishable from zero, the excluded range of mediator mass (M_{Med}) is between 0.6 and 2.6 TeV. If $m_{\mathrm{DM}} > M_{\mathrm{Med}}/2$, the mediator cannot decay to dark matter particles, and the dijet cross section from the mediator models becomes identical to that in a leptophbic Z' model without any dark matter used in Fig. 5.14 with a coupling $g'_q = g_q = 0.25$. Therefore for these values of m_{DM} the limits on the mediator mass in Fig. 5.12 are identical to the limits on the Z' mass at g'_q=0.25 in Fig. 5.14. Similarly, if $m_{\mathrm{DM}} = 0$, the limits on the mediator mass in Fig. 5.12 are identical to the limits on the Z' mass at $g'_q = g_q/\sqrt{1 + 16/(3N_f)} \approx 0.182$ in Fig. 5.14

, where N_f is the effective number of quark flavors contributing to the width of the resonance, $N_f = 5 + \sqrt{1 - 4m_t^2/M_{\mathrm{Med}}^2}$.

Fig. 5.15 shows the limits on the universal quark coupling g_q as a function of mediator mass and dark matter mass for a value of dark matter coupling fixed at $g_{\mathrm{DM}} = 1$. Couplings g_q as small as 0.07 are excluded at lower values of mediator mass and higher values of dark matter mass, similar to the exclusions in Fig. 5.14 for g'_q. For $m_{\mathrm{DM}} > M_{\mathrm{Med}}/2$ the coupling limits on g_q in Fig. 5.15 are independent of dark matter mass and identical to the coupling limits on g'_q in Fig. 5.14. This is because for $m_{\mathrm{DM}} > M_{\mathrm{Med}}/2$ the dijet cross-sections from the dark matter mediator model are identical to that for a leptophobic Z' which only decays to quarks.

As outlined in detail in Ref. [16] these results can also be compared with results from direct detection experiments.

The limits in Fig. 5.12 are first re-calculated at 90% CL and shown in Fig. 5.16. The 90% CL limits in Fig. 5.16 are then translated into the plane of the dark matter mass versus the DM-nucleon interaction cross section, shown in Fig. 5.17. The translation is done using the predicted relation between the interaction cross section and the mediator mass. An axial-vector mediator leads to a spin-dependent cross section, σ^{SD}, and a vector mediator leads to a spin-independent cross section, σ^{SI}. Fig. 5.17 shows the results from this search and compares them with dark matter searches by direct detection experiments [19; 23; 21; 22; 20; 24; 25; 26; 27].

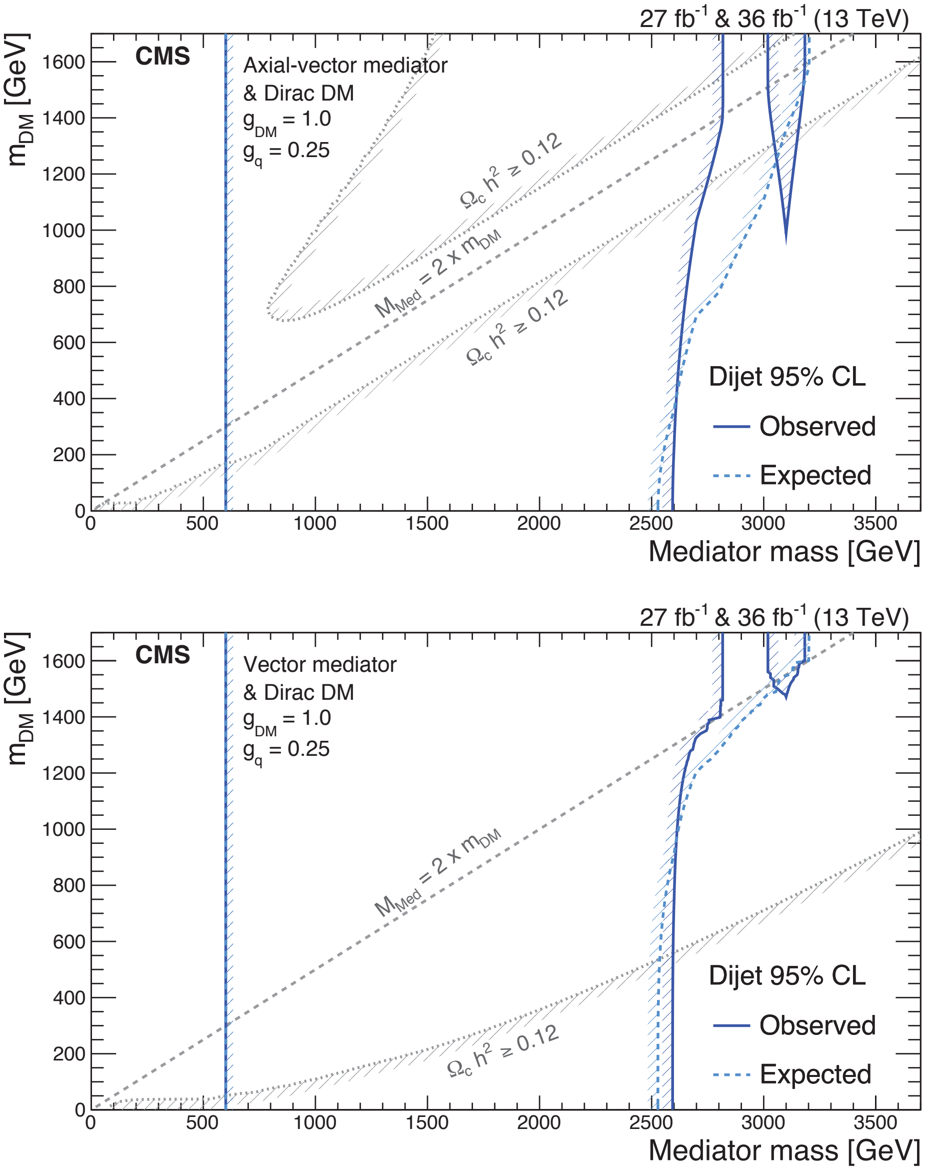


Figure 5.12: The 95% CL observed (solid) and expected (dashed) excluded regions in the plane of dark matter mass vs. mediator mass, for an axial-vector mediator (upper) and a vector mediator (lower), compared to the excluded regions where the abundance of DM exceeds the cosmological relic density (light gray). Following the recommendation of the LHC DM working group [16; 17], the exclusions are computed for Dirac DM and a universal quark coupling g_q= 0.25 and for a DM coupling of g_{DM}= 1.0. It should also be noted that the excluded region strongly depends on the chosen coupling and model scenario. Therefore, the excluded regions and relic density contours are shown in this plot do not apply to other choices of coupling values or models.

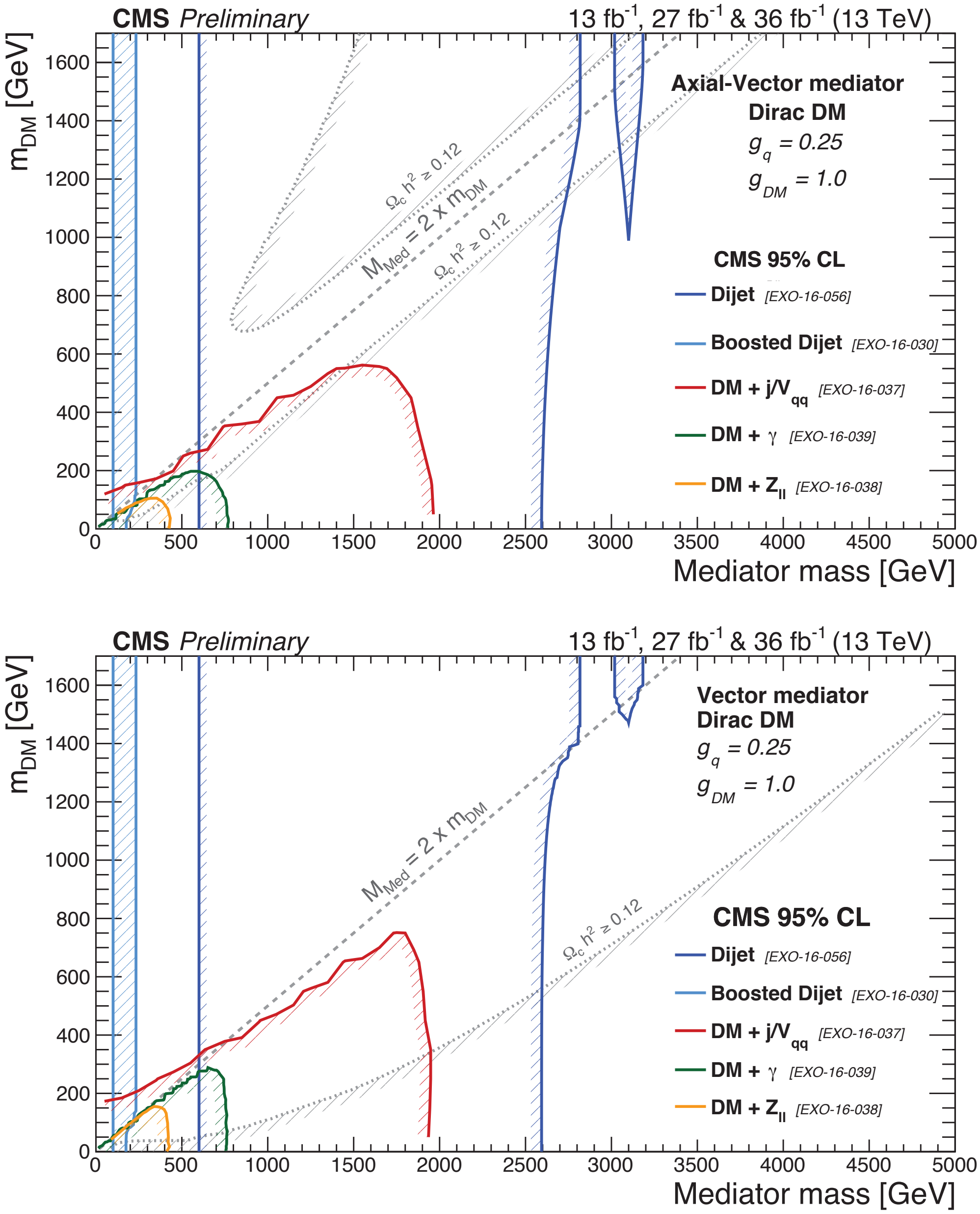


Figure 5.13: The 95% CL observed excluded regions in the plane of dark matter mass vs. mediator mass, for an axial-vector mediator (top) and a vector mediator (bottom), are shown for this analysis using 27 fb^{-1} and 36 fb^{-1} in the dijet channel (dark blue) in comparison with the results using 13 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), and are compared to the constraints from the cosmological relic density of dark matter (light gray) determined from astrophysical measurement and M$_{AD}$DM version 2.0.6 as described in [18].

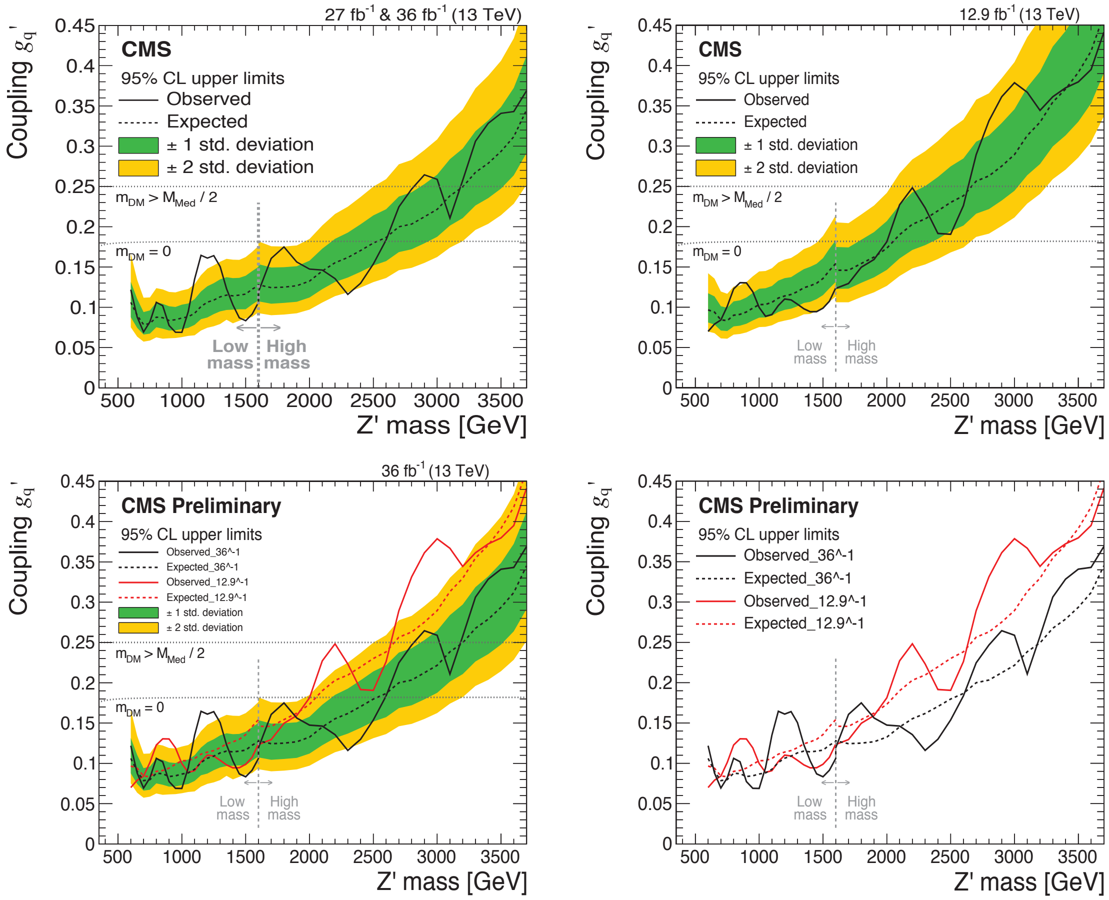
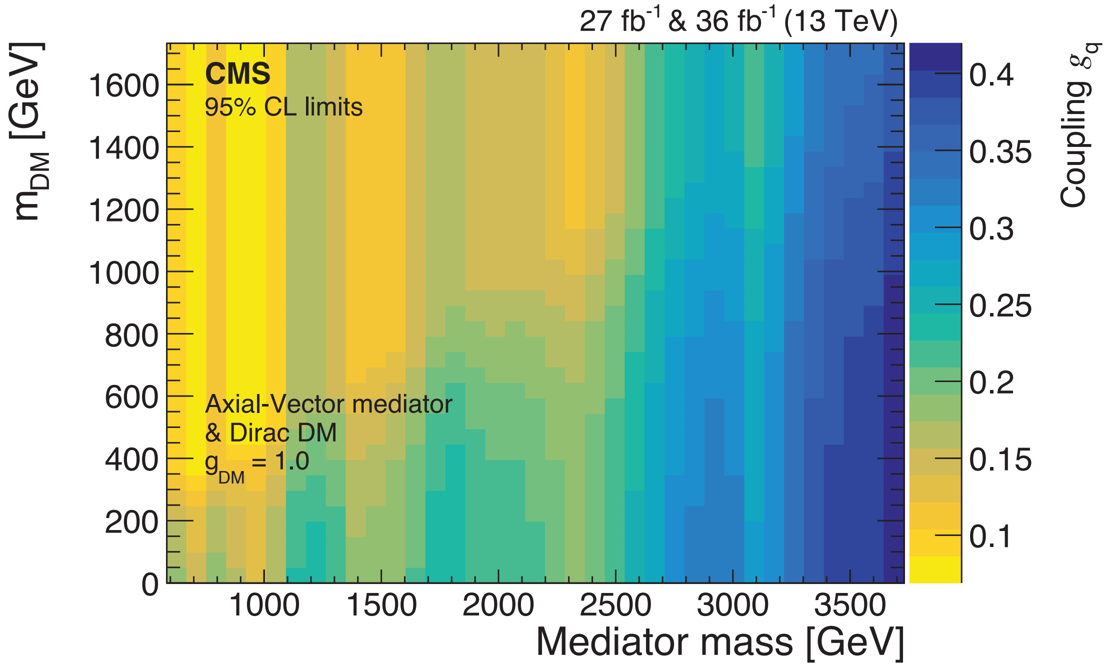


Figure 5.14: The 95% Cl upper limits on the universal quark coupling g'$_q$ as a function of resonance mass for a leptonphobic Z' resonance that only couples to quarks. This show the observed limits (solid), expected limits (dashed) and their variation at the 1 and 2 standard deviation levels (shaded bands). Dotted horizontal lines are also shown as the coupling strength for which the cross section for dijet production in this model is the same as for a dark matter mediator. Top left) This analysis with up to 36 fb^{-1}, Top right) previous published analysis with 12.9 fb^{-1}, bottom) comparisons between this analysis with up to 36 fb^{-1} and previous analysis with 12.9 fb^{-1}.



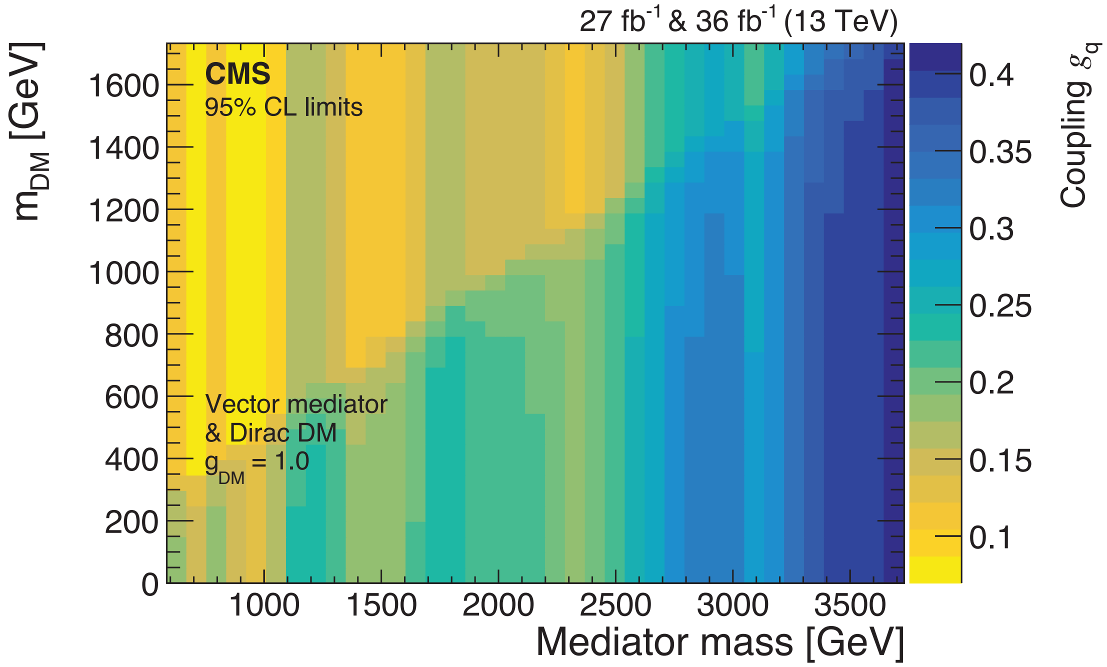
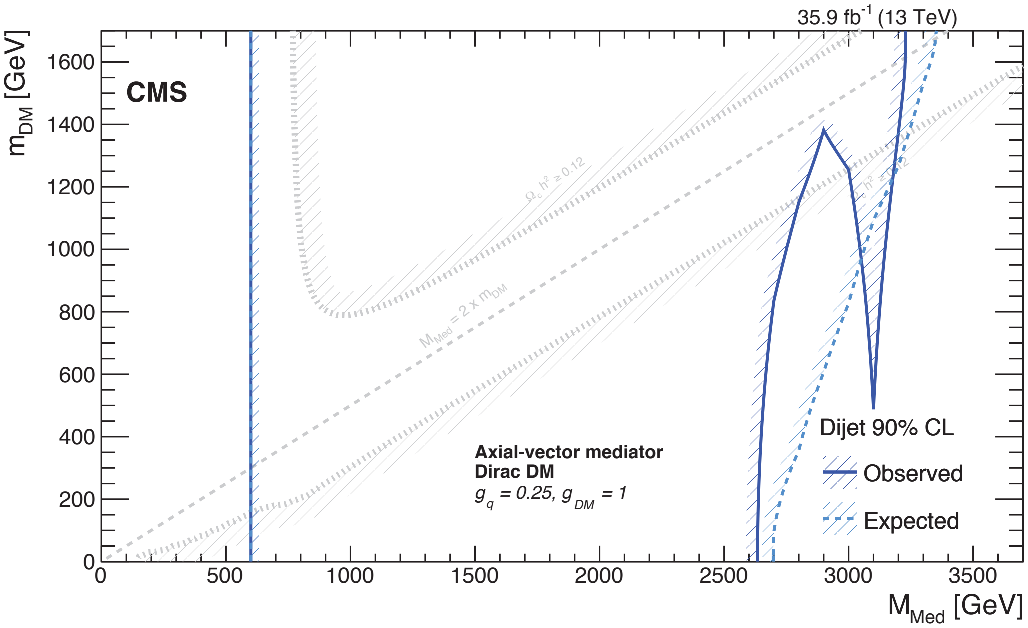


Figure 5.15: The 95% CL observed upper limits on a universal quark coupling g_q (color scale at right) in the plane of the dark matter particle mass versus mediator mass for an axial-vector mediator (upper) and a vector mediator (lower).



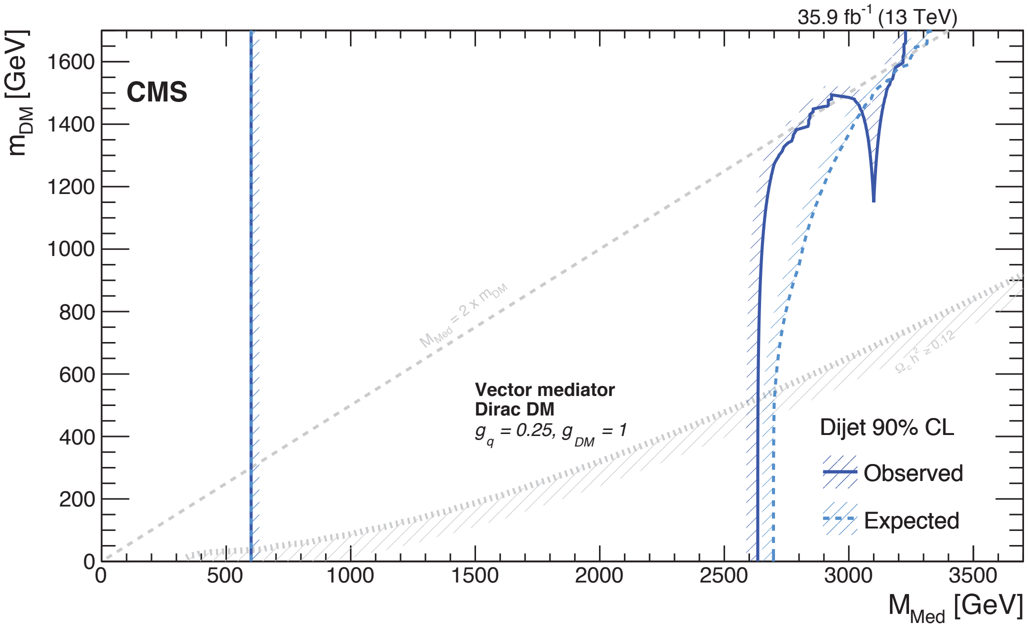


Figure 5.16: Same as Fig. 5.12 but with dijet limits at 90% CL. Cosmological relic density is old in this figure but is up to dated in Fig. 5.12.

.

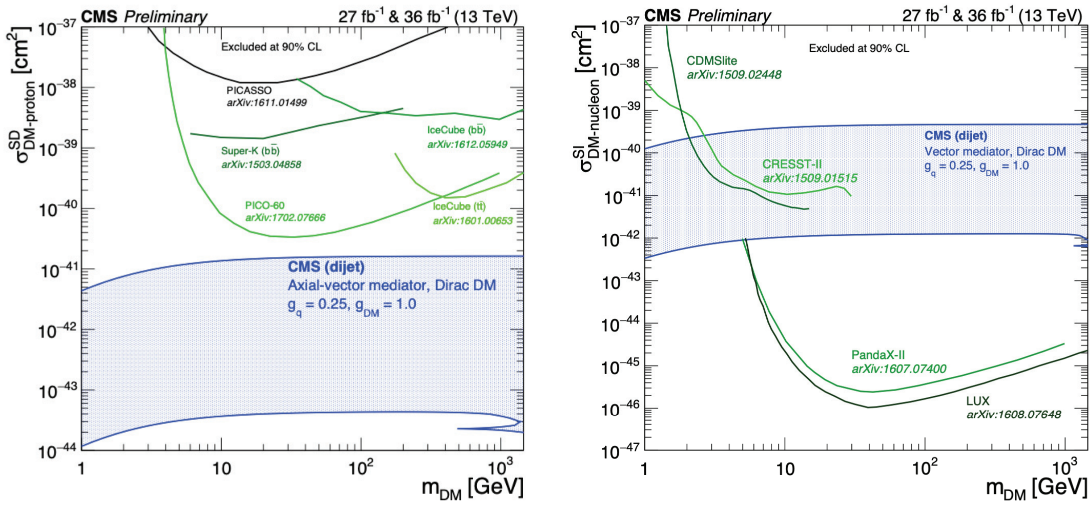


Figure 5.17: Excluded regions at 90% CL in the plane of dark matter nucleon interaction cross section vs. dark matter mass. (left) The CMS exclusion of a spin-dependent cross section (shaded) from an axial-vector mediator decaying to dijets is compared with limits from PICASSO [19], Super-Kamiokande [20], IceCube [21; 22] and PICO-60 [23]. (right) The CMS exclusion of a spin-independent cross section (shaded) from a vector mediator decaying to dijets is compared with limits from LUX [24], PandaX-II [25], CDMSLite [26], and CRESST-II [27]. These direct detection experiments have documented the most constraining results in the mass range shown. The CMS exclusions are for Dirac dark matter and couplings $g_q = 0.25$ and $g_{\mathrm{DM}} = 1$, 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.

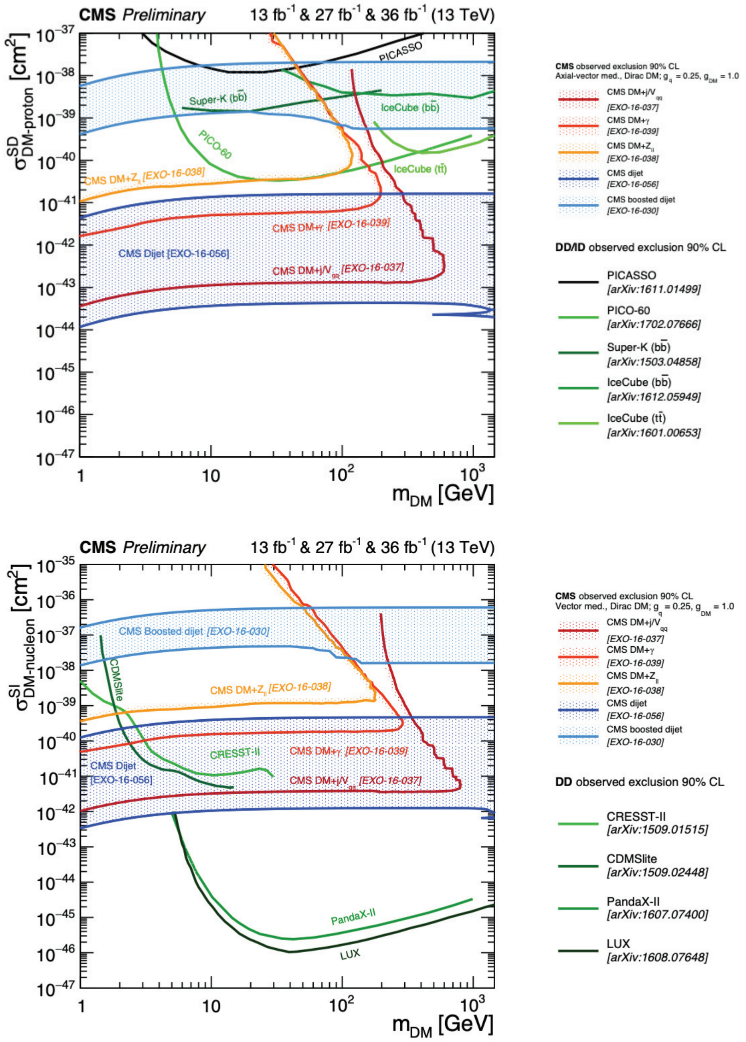


Figure 5.18: Excluded regions at 90% CL in the plane of dark matter nucleon interaction cross section vs. dark matter mass. The CMS exclusions using 27 and 36 fb^{-1} from dijets (dark blue shaded) and using 13 fb^{-1} from boosted dijets (light blue and shaded), mono-jets (red), mono-photons (orange), and mono-Z (yellow) are compared with direct detection experiments. (top) The exclusion of a spin-dependent cross section from an axial-vector mediator compared with limits from PICASSO [19], Super-Kamiokande [20], IceCube [21; 22] and PICO-60 [23]. (bottom) The exclusion of a spin-independent cross section from a vector mediator compared with limits from LUX [24], PandaX-II [25], CDMSLite [26], and CRESST-II [27]. These direct detection experiments have documented the most constraining results in the mass range shown. The CMS exclusions are for Dirac dark matter and couplings $g_q = 0.25$ and $g_{\mathrm{DM}} = 1$, 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

CHAPTER 6

SEARCH FOR HEAVY RESONANCES IN THE B-TAGGED DIJET MASS SPECTRUM IN PP COLLISION AT $\sqrt{s} = 13$ TEV USING FULL RUN II DATA BTAG

This analysis focus on the search for dijet resonances in the high mass region, using pp collisions at $\sqrt{s} = 13$ TeV collected in 2016, 2017, and 2018, corresponding to an integrated luminosity of 137.5 fb^{-1}. All Jets are required to have $p_T > 30$ GeV and $\eta < 2.5$. Spatially close jets are combined into “wide jets” and used to determine the dijet mass. The background from t-channel dijet events peaks at large values of $|\Delta\eta_{jj}|$ and is suppressed by requiring the pseudorapidity separation of the two wide jets to satisfy $|\Delta\eta_{jj}| < 1.1$. The above requirements maximize the search sensitivity for isotropic decays of dijet resonances in the presence of a QCD dijet background. In order to increase the sensitivity to signals with b quark in the final state, an analysis searching for heavy resonance in the b-tagged dijet final state is presented in this section.

6.1 Data and Signal Samples

6.1.1 Collision Datasets

The primary dataset used for this analysis is JetHT for the whole Run II which doesn’t include any jet tagging requirement at the trigger level. The total integrated luminosity sums up to 137.5 fb^{-1} when summing the 2016, 2017 and 2018 datasets, which are shown in table 6.1 - 6.3 respectively. For the 2016 dataset 07Aug2017_V11 JEC is used, Fall17_17Nov2017_V6 is used for the 2017 dataset while Autumn_18_V10 is used for the 2018 dataset.

6.1.2 MC Samples Datasets

For all signal samples considered, the assumption is that the resonance intrinsic width is much smaller than the dijet mass resolution. Two sets of signal shapes are generated accordingly to the signal model considered:

- bg → b^* → bg
- qq → Z’ → $\text{b}\bar{\text{b}}$

Table 6.1: 2016 Datasets used in the high-mass RECO analysis.

Dateset
/JetHT/Run2016B-07Aug17_ver1-v1/MINIAOD
/JetHT/Run2016B-07Aug17_ver2-v1/MINIAOD
/JetHT/Run2016C-07Aug17-v1/MINIAOD
/JetHT/Run2016D-07Aug17-v1/MINIAOD
/JetHT/Run2016E-07Aug17-v1/MINIAOD
/JetHT/Run2016F-07Aug17-v1/MINIAOD
/JetHT/Run2016G-07Aug17-v1/MINIAOD
/JetHT/Run2016H-07Aug17-v1/MINIAOD

Table 6.2: 2017 Datasets used in the high-mass RECO analysis.

Dateset
/JetHT/Run2017B-17Nov2017-v1/MINIAOD
/JetHT/Run2017C-17Nov2017-v1/MINIAOD
/JetHT/Run2017D-17Nov2017-v1/MINIAOD
/JetHT/Run2017E-17Nov2017-v1/MINIAOD
/JetHT/Run2017F-17Nov2017-v1/MINIAOD

- qq → DM mediator → $b\bar{b}$

All MC samples are generated and processed via the `MINIAOD` and JEC in the table 6.4, in order to have a good agreement with inclusive full Run II data analysis.

6.1.2.1 $b^* \to bg$

The b^* signal samples are generated privately with `PYTHIA` using the following GEN fragment:

```
PythiaParameters = cms.PSet(
    parameterSets = cms.vstring('pythia8CommonSettings',
        'pythia8CUEP8M1Settings',
```

Table 6.3: 2018 Datasets used in the high-mass RECO analysis.

Dateset
/JetHT/Run2018A-17Sep2018-v1/MINIAOD
/JetHT/Run2018B-17Sep2018-v1/MINIAOD
/JetHT/Run2018C-17Sep2018-v1/MINIAOD
/JetHT/Run2018D-PromptReco-v2/MINIAOD

Table 6.4: CMSSW release and GT used for `miniAOD` private production of MC samples for the 2016 (left), 2017 (middle) and 2018 (bottom)considered in the analysis.

Year	CMSSW release	Global Tag
2016	CMSSW_8_0_21	80X_mcRun2_asymptotic_2016_TrancheIV_v6
2017	CMSSW_9_4_7	94X_mc2017_realistic_v14
2018	CMSSW_10_2_5	102X_upgrade2018_realistic_v15

```
    'processParameters'),
processParameters = cms.vstring('ExcitedFermion:bg2bStar = on',
    '4000005:m0 = X',
    '4000005:onMode = off',
    '4000005:onIfMatch = 21 5',
    'ExcitedFermion:Lambda = X',
    'ExcitedFermion:coupFprime = 1.',
    'ExcitedFermion:coupF = 1.',
    'ExcitedFermion:coupFcol = 1.'),
pythia8CUEP8M1Settings = cms.vstring('Tune:pp 14',
    'Tune:ee 7',
    'MultipartonInteractions:pT0Ref=2.4024',
    'MultipartonInteractions:ecmPow=0.25208',
    'MultipartonInteractions:expPow=1.6'),
pythia8CommonSettings = cms.vstring('Tune:preferLHAPDF = 2',
    'Main:timesAllowErrors = 10000',
    'Check:epTolErr = 0.01',
```

```
        'Beams:setProductionScalesFromLHEF = off',
        'SLHA:keepSM = on',
        'SLHA:minMassSM = 1000.',
        'ParticleDecays:limitTau0 = on',
        'ParticleDecays:tau0Max = 10',
        'ParticleDecays:allowPhotonRadiation = on')
)
```

where X indicates both the b* and the compositeness scale Λ. The b* is forced to decay into a bg final state. Private samples have been produced in 2019 with the same release and conditions used for the official q* samples employed in the dijet inclusive analysis. b* samples are located on EOS and available at the path indicated in table 6.5.

Table 6.5: Signal datasets of b* sample for the three years considered in the analysis.

Dataset
/eos/cms/store/group/phys_exotica/dijet/Dijet13TeV/deguio/sampleGen_2016/BstarToJJ/*
/eos/cms/store/group/phys_exotica/dijet/Dijet13TeV/deguio/sampleGen_2017/BstarToJJ/*
/eos/cms/store/group/phys_exotica/dijet/Dijet13TeV/deguio/sampleGen_2018/BstarToJJ/*

6.1.2.2 Z' → $b\bar{b}$

The Z' signal samples are generated privately with `PYTHIA` using the following GEN fragment:

```
PythiaParameters = cms.PSet(
        parameterSets = cms.vstring('pythia8CommonSettings',
            'pythia8CUEP8M1Settings',
            'processParameters'),
        processParameters = cms.vstring('NewGaugeBoson:ffbar2gmZZprime = on',
            'Zprime:gmZmode = 3',
            '32:m0 = X',
            '32:onMode = off',
            '32:onIfAny = 5'),
```

```
        pythia8CUEP8M1Settings = cms.vstring('Tune:pp 14',
            'Tune:ee 7',
            'MultipartonInteractions:pT0Ref=2.4024',
            'MultipartonInteractions:ecmPow=0.25208',
            'MultipartonInteractions:expPow=1.6'),
        pythia8CommonSettings = cms.vstring('Tune:preferLHAPDF = 2',
            'Main:timesAllowErrors = 10000',
            'Check:epTolErr = 0.01',
            'Beams:setProductionScalesFromLHEF = off',
            'SLHA:keepSM = on',
            'SLHA:minMassSM = 1000.',
            'ParticleDecays:limitTau0 = on',
            'ParticleDecays:tau0Max = 10',
            'ParticleDecays:allowPhotonRadiation = on')
    )
```

where X indicates both the Z' and the compositeness scale Λ. The Z' is forced to decay into a bg final state. Private samples have been produced in 2019 with the same release and conditions used for the official Z' samples employed in the dijet inclusive analysis. Z' samples are located on EOS and available at the path indicated in table 6.6.

Table 6.6: Signal datasets of Z' samples for the Full Run II considered in the analysis.

Dataset
/eos/cms/store/group/phys_exotica/dijet/Dijet13TeV/deguio/sampleGen_2016/ZprimeToBBbar/*
/eos/cms/store/group/phys_exotica/dijet/Dijet13TeV/deguio/sampleGen_2017/ZprimeToBBbar/*
/eos/cms/store/group/phys_exotica/dijet/Dijet13TeV/deguio/sampleGen_2018/ZprimeToBBbar/*

6.1.2.3 Simplified Dark Matter Mediator

The simplified Dark Matter models, with mediator decaying to b$\bar{\text{b}}$ final state, are generated via `MG5` from the GEN level. Recommended couplings with $g_q = 0.25$ and $g_{DM} = 1.0$ are used. The resonance mass of the dark matter mediator equals 1 GeV and Z' masses vary from 1 TeV to 9 TeV with 1 TeV step, plus 500 GeV to cover the low mass part. The original cross-section of each sample can be obtained from the LHE files from the `MG5` generator. In order to match the analysis requirement, the usual kinematic constraints ($\Delta\eta_{jj} < 1.1$ and $|\eta| < 2.5$) are applied for each event, A ±500 GeV mass window selection is also applied to the dijet invariant mass of each event in order to keep it in the narrow resonance. The theoretical limit obtained from the `MG5` is listed in the following table 6.7 and 6.8

Table 6.7: Cross-section of vector leptophobic dark matter mediator Z' decay to $b\bar{b}$ and $q\bar{q}$ final state, obtained from `MG5`

mass	Xsec of qq	Xsec of bb
1000	3.79	0.678
2000	0.15	0.0275028735
3000	0.013	0.00237807168
4000	0.00137	2.5e-4
5000	1.4e-4	2.9e-5
6000	1.47e-5	3e-6
7000	1.23e-6	3.1e-7
8000	7.31e-8	2.05e-8
9000	2.56e-9	2.6e-9

Table 6.8: Cross-section of axial-vector leptophobic dark matter mediator Z' decay to $b\bar{b}$ and $q\bar{q}$ final state, obtained from `MG5`

mass	Xsec of qq	Xsec of bb
1000	3.79	0.678
1000	3.99	0.67946086
2000	0.15	0.02688567
3000	0.014	0.0023953606
4000	0.00145	2.5e-4
5000	1.46e-4	2.8e-5
6000	1.52e-5	3e-6
7000	1.29e-6	3.1e-7
8000	7.37e-8	2.1e-8
9000	2.65e-9	2.65e-9

The generated MINIAOD files of each year are listed in the tables 6.9 − 6.14.

Table 6.9: Signal datasets of vector dark matter mediator for the 2016

Leptophobic Vector Dark matter mediator Dateset in 2016
/DMAVGeneration2016_2020/zhixing-DMVgenerationZ0500DM125MINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMVgenerationZ0750DM138MINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMVgenerationZ1000DM151MINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMVgenerationZ2000DM1102MINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMVgenerationZ3000DM1154MINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMVgenerationZ4000DM1205MINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMVgenerationZ5000DM1256MINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMVgenerationZ6000DM1308MINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMVgenerationZ7000DM1359MINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMVgenerationZ8000DM1411MINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMVgenerationZ9000DM1462MINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER

Table 6.10: Signal datasets of axial-vector dark matter mediator for the 2016

Leptophobic Axial-Vector Dark Matter Mediator Dateset in 2016
/DMAVGeneration2016_2020/zhixing-DMAVgenerationZ1000DM151bbMINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMAVgenerationZ2000DM1102bbMINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMAVgenerationZ3000DM1154bbMINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMAVgenerationZ4000DM1205bbMINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMAVgenerationZ5000DM1256bbMINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMAVgenerationZ500DM125bbMINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMAVgenerationZ6000DM1308bbMINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMAVgenerationZ7000DM1359bbMINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMAVgenerationZ8000DM1411bbMINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER
/DMAVGeneration2016_2020/zhixing-DMAVgenerationZ9000DM1462bbMINIAODSIM-bd3e7bcff6c9bcad356ea4ed7e4f08b4/USER

6.2 Trigger, Event Reconstruction and Selections

The baseline event selection follows the standard of previous CMS dijet searches [90]. All AK4PFCHS jets are required to have $p_T > 30$ GeV and $|\eta| < 2.5$. The two jets with the largest p_T are defined as the leading jets. In the standard RECO analysis, we use the standard PF jet ID.

Geometrically close jets are combined into “wide jets” and used to determine the dijet mass, as in our previous dijet searches. The wide-jet algorithm, designed for dijet resonance event reconstruction, reduces the analysis sensitivity to gluon radiation from the final state partons. The two leading jets are used as seeds and the four-vectors of all other jets, if they are within $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 1.1$ of the seed jet, and are added to the nearest leading jet to obtain two wide jets. The above requirements maximize the search sensitivity for isotropic decays of dijet

Table 6.11: Signal datasets of vector dark matter mediator for the 2017

Leptophobic Vector Dark matter mediator Dateset in 2016
/DMAVGeneration2017_2020/zhixing-DMVgenerationZ0500DM125MINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMVgenerationZ1000DM151MINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMVgenerationZ2000DM1102MINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMVgenerationZ3000DM1154MINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMVgenerationZ4000DM1205MINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMVgenerationZ5000DM1256MINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMVgenerationZ6000DM1308MINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMVgenerationZ7000DM1359MINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMVgenerationZ8000DM1411MINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMVgenerationZ9000DM1462MINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER

Table 6.12: Signal datasets of axial-vector dark matter mediator for the 2017

Leptophobic Axial-Vector Dark Matter Mediator Dateset in 2016
/DMAVGeneration2017_2020/zhixing-DMAVgenerationZ1000DM151bbMINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMAVgenerationZ2000DM1102bbMINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMAVgenerationZ3000DM1154bbMINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMAVgenerationZ4000DM1205bbMINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMAVgenerationZ5000DM1256bbMINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMAVgenerationZ500DM125bbMINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMAVgenerationZ6000DM1308bbMINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMAVgenerationZ7000DM1359bbMINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMAVgenerationZ8000DM1411bbMINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER
/DMAVGeneration2017_2020/zhixing-DMAVgenerationZ9000DM1462bbMINIAODSIM-5f646ecd4e1c7a39ab0ed099ff55ceb9/USER

resonances in the presence of a QCD dijet background.

The pseudorapidity separation of the two wide jets is required to be $|\Delta\eta_{jj}| < 1.1$ for the signal region. This requirement suppresses the QCD t-channel, while the signal contribution significantly comes from the s-channel. It also restricts the region of our measurement predominantly within the barrel region. The corresponding one for the control regions, enhanced in QCD and with smaller signal contamination, is $1.1 < |\Delta\eta_{jj}| < 2.6$.

As in previous dijet searches, hadronic triggers based on the scalar sum of the jet transverse momenta are used. Depending on the year considered, the trigger thresholds are changing. For each year, a lower cut on the dijet invariant mass is applied to ensure full efficiency on the trigger turn-on. Trigger efficiency studies have been presented in AN-2017-348 [91] and AN-2019-073 [92].

Table 6.13: Signal datasets of vector dark matter mediator for the 2018

Leptophobic Vector Dark matter mediator Dateset in 2016
/DMAVGeneration2018_2020/zhixing-DMVgenerationZ0500DM125MINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMVgenerationZ1000DM151MINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMVgenerationZ2000DM1102MINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMVgenerationZ3000DM1154MINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMVgenerationZ4000DM1205MINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMVgenerationZ5000DM1256MINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMVgenerationZ6000DM1308MINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMVgenerationZ7000DM1359MINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMVgenerationZ8000DM1411MINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMVgenerationZ9000DM1462MINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER

Table 6.14: Signal datasets of axial-vector dark matter mediator for the 2018

Leptophobic Axial-Vector Dark Matter Mediator Dateset in 2016
/DMAVGeneration2018_2020/zhixing-DMAVgenerationZ500DM125bbMINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMAVgenerationZ1000DM151bbMINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMAVgenerationZ2000DM1102bbMINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMAVgenerationZ3000DM1154bbMINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMAVgenerationZ4000DM1205bbMINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMAVgenerationZ5000DM1256bbMINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMAVgenerationZ6000DM1308bbMINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMAVgenerationZ7000DM1359bbMINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMAVgenerationZ8000DM1411bbMINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER
/DMAVGeneration2018_2020/zhixing-DMAVgenerationZ9000DM1462bbMINIAODSIM-3ee3afd6b5a1410aea6d0b4d52723d06/USER

In accordance with other dijet searches, a veto is applied to events with a high-fat jet (slimmed AK8 jets) mass or the ones containing isolated electrons or muons. The fat jet veto is applied to all events in which both primary and secondary fat jet have a PUPPI [93] corrected soft drop mass higher than 65 GeV. The isolated electron veto occurs whenever there is at least one reconstructed electron in the acceptance region with $p_T > 50$ GeV and a cut-based ID higher or equal to 2. The isolated muon veto is applied in the case of at least one reconstructed muon in the acceptance region with $p_T > 50$ GeV, a high-p_T cut-based ID higher or equal to 2, and tracker-based relative isolation smaller than 0.1.

6.2.1 Trigger Efficiency

The trigger efficiency is derived via the SingleMuon dataset and is reported for each year in Fig 6.1. The turn-on curves become fully efficient above approximately 1.5 TeV. Thus, a dijet mass threshold of 1.53 TeV is used in the following analysis.

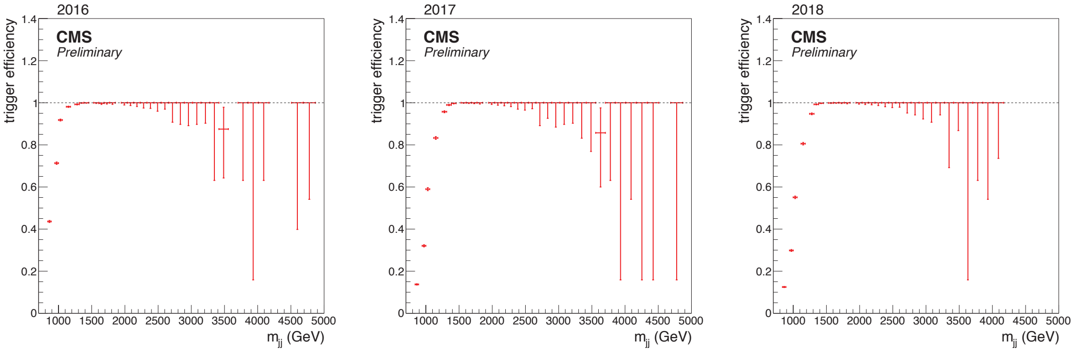


Figure 6.1: Trigger efficiency from the SingleMuon dataset in 2016 (left), 2017 (center) and 2018 (right).

6.2.2 Performance of the b-tagger at High-p_T

The b-tagger used in this analysis exploits features such as the impact parameters of tracks and the displaced vertices to identify jets originating from the decay of b hadrons. The b-tagging algorithm performance has been optimized separately for the three years considered in this analysis and the WPs (loose, medium, tight) are defined to provide a specific b-jet efficiency on an inclusive $t\bar{t}$ sample [94]. In this analysis different categories requiring exactly one, at least one, or exactly two b-tagged jets are defined depending on the signal model considered.

The reconstruction and identification of secondary vertices become particularly challenging when considering the decay of heavy resonances and the performance of the b-tagger becomes worse as a function of the jet p_T. As the jet energy increases, the track multiplicity in the core of the jet increases and the jet becomes more and more collimated making the pattern recognition increasingly more difficult.

In Fig 6.2 an RS graviton signal sample has been used to study the b-tagging performance as a function of the transverse momentum for high p_T jets. The probability for a non-b jet (jets originating from uds quarks or a gluon are considered) to be identified as a b jet is reported as a function of the b-tagging efficiency for different tagging algorithms: CSVv2, DeepCSV, DeepJet. The DeepJet algorithm clearly outperforms the other centrally available taggers and it has been selected for this analysis.

In Fig 6.3 the DeepJet performance is reported for different years and different p_T bins. The visible performance improvement in 2017 and 2018 with respect to 2016 is due to the upgrade of the inner tracking detector which gained one layer of pixels as part of the CMS phase 1 upgrade.

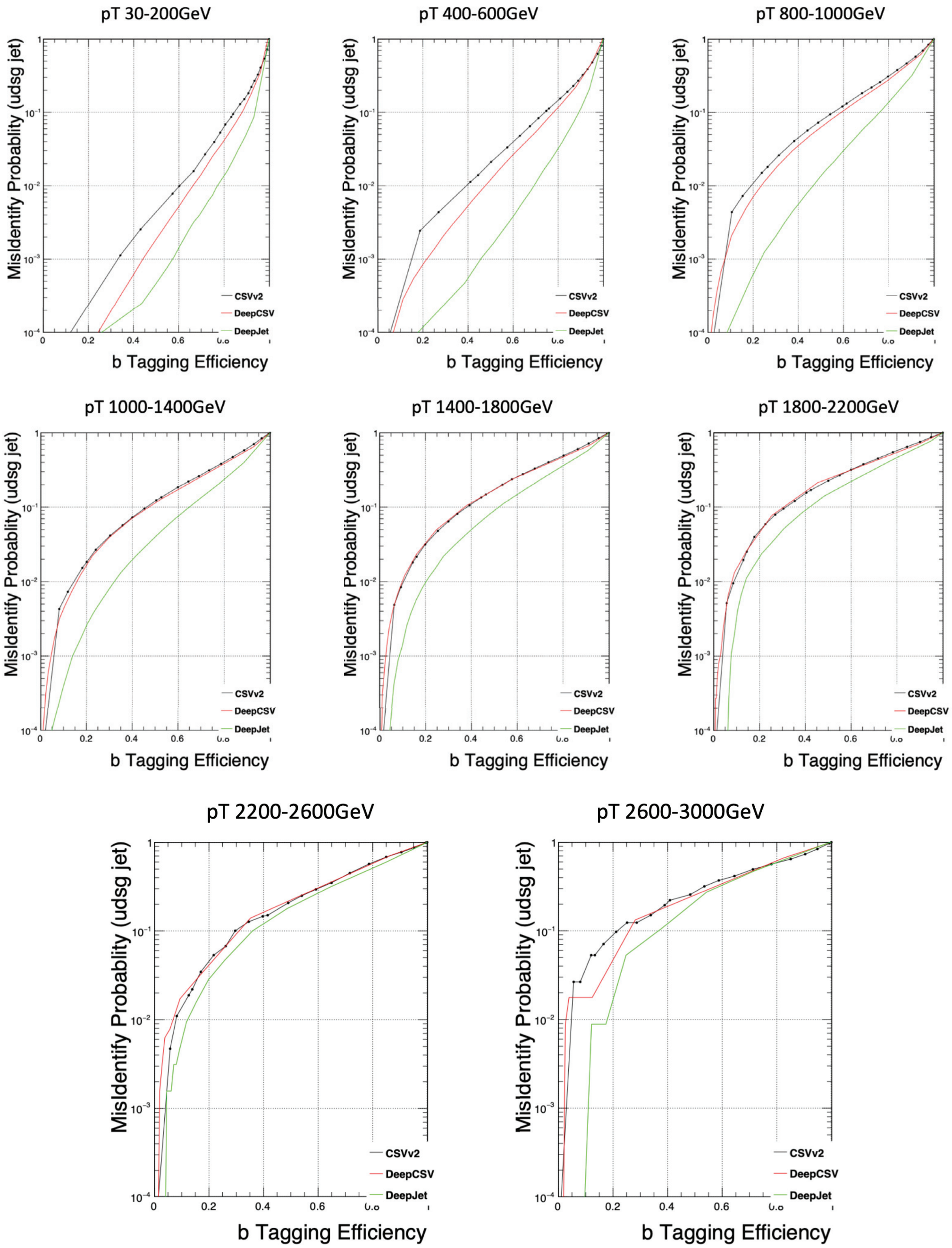
6.2.3 Signal Efficiency

Theoretically, the b-jet identification and the misidentification probability to identify a non-b jet as a b jet will not be the same in data and MC samples. It's needed to apply a scale factor to the simulation sample in order to correct for the difference. The BTV POG measures the DeepJet efficiency on b and the misidentification probability on lighter flavor jets on 2016, 2017, and 2018 data and make recommended discriminator for analysis, named loose, medium, and tight working point (WP). The loose, medium and tight WP is defined as the values of the discriminator cut for which the rate for misidentifying a light jet as a b jet is 10%, 1%, and 0.1%, respectively. The recommended WP of each year is listed in table 6.2.3

WP	2016	2017	2018
loose	0.0614	0.0521	0.0494
medium	0.3093	0.3033	0.2770
tight	0.7221	0.7489	0.7264

Meanwhile, we have four categories we can take into consideration:

- **0b** none of the two jets pass the selection
- **1b** only one of the two jets pass the selection
- **2b** both of the two jets pass the selection
- **le1b** at least one of the two jets pass the selection


pT 30-200GeV

pT 400-600GeV

pT 800-1000GeV

pT 1000-1400GeV

pT 1400-1800GeV

pT 1800-2200GeV

pT 2200-2600GeV

pT 2600-3000GeV

Figure 6.2: Performance comparison for different b-tagging algorithms in different p_T bins. A RS graviton sample from the 2017 production has been used.

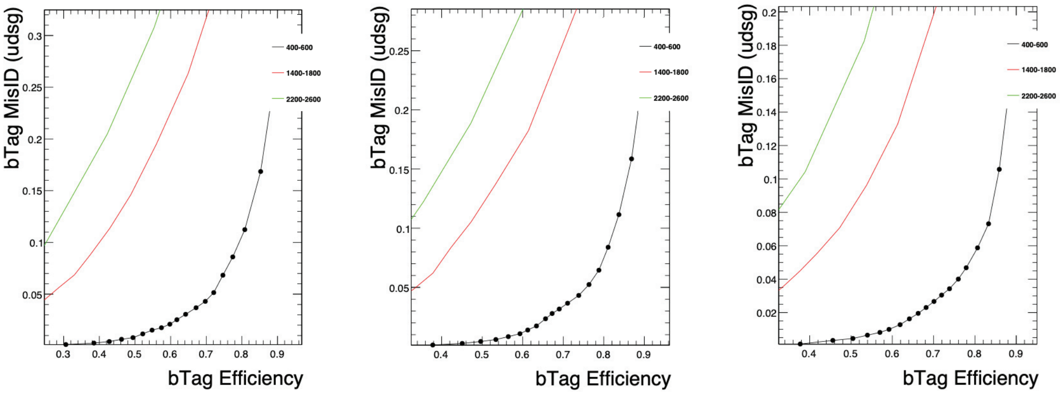


Figure 6.3: DeepJet performance (mis-identification rate vs efficiency) in three p_T bins for the three years considered in this analysis.

In order to preview if the WP can give benefit to the result, we need to utilized the formula $\frac{\text{number of signal}}{\sqrt{\text{number of bkg}}}$. If we can calculate the rate of how many percent of the events can go into the b-tag category in both data and signal, then we can see how much the b-tagger can benefit the limit if it's beyond 1. Instead of tagging efficiency, which is defined as the rate of b-jets passing the tagger selection, we define a new property named the tagging rate, which is defined as the rate of an event passing a specific category under a b-tagger selection.

$$taggingrate = \frac{\#passbTag \quad \&\& \quad passDijetSelection}{\#passDijetSelection} \tag{6.1}$$

And we can defined the tagging power as following:

$$taggingpower = \frac{signalTagRate}{\sqrt{BkgTagRate}} \tag{6.2}$$

The signal tagging rates are shown in Fig 6.4, 6.5 and 6.6. Meanwhile, background tagging rates are shown in Fig 6.7. At the end, the tagging power is shown in Fig 6.8.

In the Fig 6.4, showing the DeepJet tagging rate of b* sample. Comparing the tagging rate between 1b category, 2b category, and le1b category in each plot. With the loose WP, the tagging rate is stable among all mass points because it's too loose to select b-quark from the background. when the WP becomes tighter, the tagging rate starts to fall down when the mass goes up. because b-tagger is only trained at the low jet p_T, specifically between 0 and 1 TeV, it's not expected to have good behavior beyond 2 TeV. When the mass goes beyond 6 TeV, the tagging rate starts to rise up again. This's due to more QCD events that will be in the tail of samples than the true b* decay events. That's why we stop the search up to 6 TeV. In addition, the same check on the Z' samples using the medium WP is shown in Fig 6.5

Figs. 6.7 shows that the tagging rate in the real collision data. They are similar among all three years. The loose WP always increases with the invariant mass because there are more physics events happen instead of simply an elastic collision event. The medium and tight WP, however, is stable in all bins. The rate fluctuates in the high mass bin because of the low statistic.

Figs. 6.8 show that the tagging power in each year using each WP in the le1b category. These plots show that the medium and tight WP will have tagging power beyond 1 around 2 TeV, where we expect to see an exclusion limit.

Figs. 6.9 show the scan over several discrete discriminator cut, using 1b and le1b category, while the gray dash line indicates the limit of applying no b-tagger. In all plots, we can see that the le1b category can give a better limit than the 1b category, and both categories have a few discriminator cut which benefits the limit. In 2016, the best limit can be get at 0.25, in 2017 it's 0.65, and in 2018 it's 0.418.

In the data analysis, we want to choose one of the WPs as the discriminator cut, because we can take the advantage of measured b-tag systematic, and easy to compare with other results. From this perspective, and also from the limit scan, it is shown that 2016 and 2018 are very close to the medium WP, while the best cut happens to be close to the tight WP. Since in the 2017 case, the limit of medium and tight WP only has few tens of GeV differences. In order to be consistent with the other two years. We decided to choose the medium WP as the discriminator cut for all 3 years.

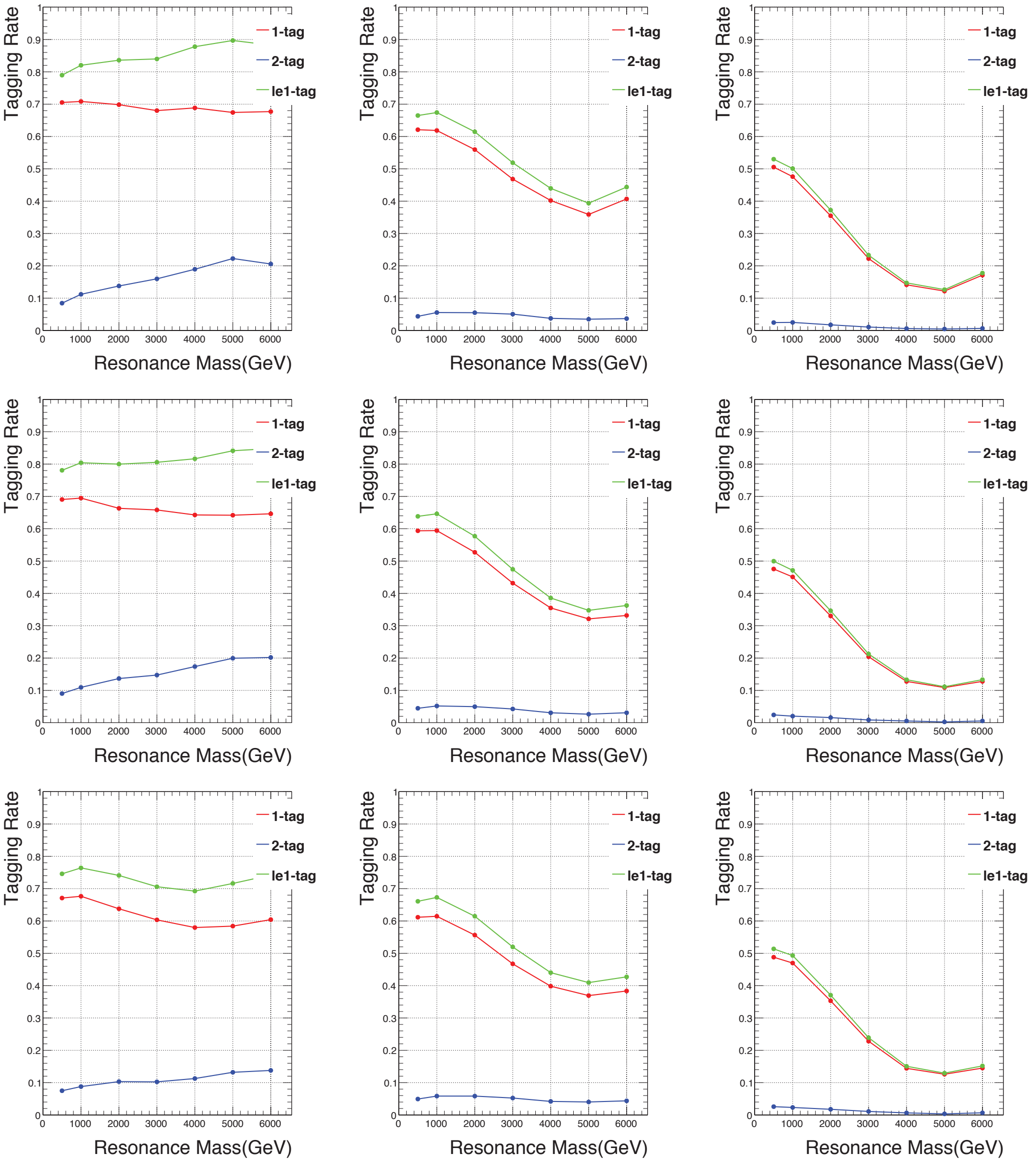


Figure 6.4: DeepJet tagging rate of b* sample with only one b jet in the final state, applying loose (left), medium (central) and tight (right) WP at 2016 (top), 2017 (middle) and 2018 (bottom) dataset.

6.3 Data Quality

In this section, a number of kinematic quantities are shown to examine the agreement between data and simulated events. The background is composed by the

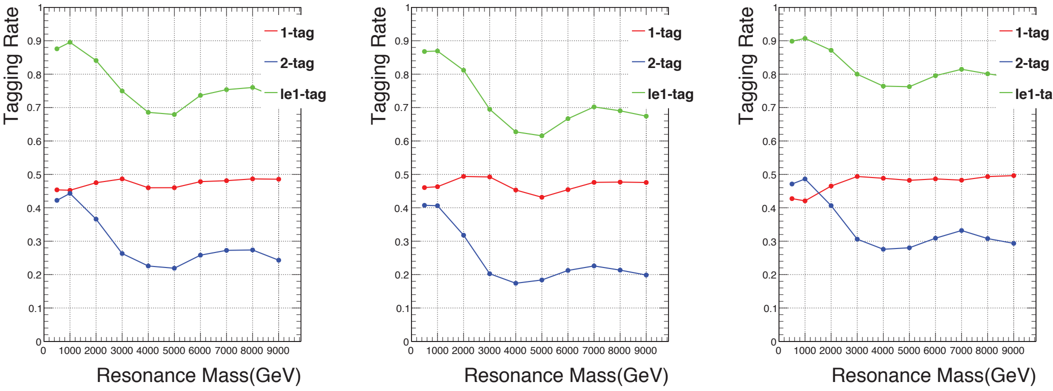


Figure 6.5: DeepJet tagging rate of Z' sample with two b jets in the final state, applying loose (left), medium (central) and tight (right) WP at 2016 (top), 2017 (middle) and 2018 (bottom) dataset.

$t\bar{t}$ production and QCD. Regardless of the final selection, the $t\bar{t}$ component is always negligible and does never exceed a few percent of the total background.

In the following figures, MC samples are normalized to the luminosity without accounting for further rescaling factors.

6.4 Background Estimation

The main background of this analysis consists of multijet processes, which may contain genuine or fake b quarks. Even after the application of b-tagging selections, the $t\bar{t}$ background consists only of 1% or less of the total background, depending on the category, and thus it requires a special treatment neither for the background prediction nor for dedicated selections.

The background, considered as the sum of the main QCD multijet and the minor $t\bar{t}$ background, is estimated directly from data. The method chosen has been widely used in CMS searches and is commonly referred to as "dijet fit" or "smoothness test" method. The smoothly falling background is parametrized by analytic functions with a variable number of parameters fitted directly to the data in the signal regions. This method offers a simple and robust estimation, does not depend on auxiliary control regions, and has no extrapolation systematic uncertainties. On the other hand, it is required that the background should be smooth, *e.g.* does not

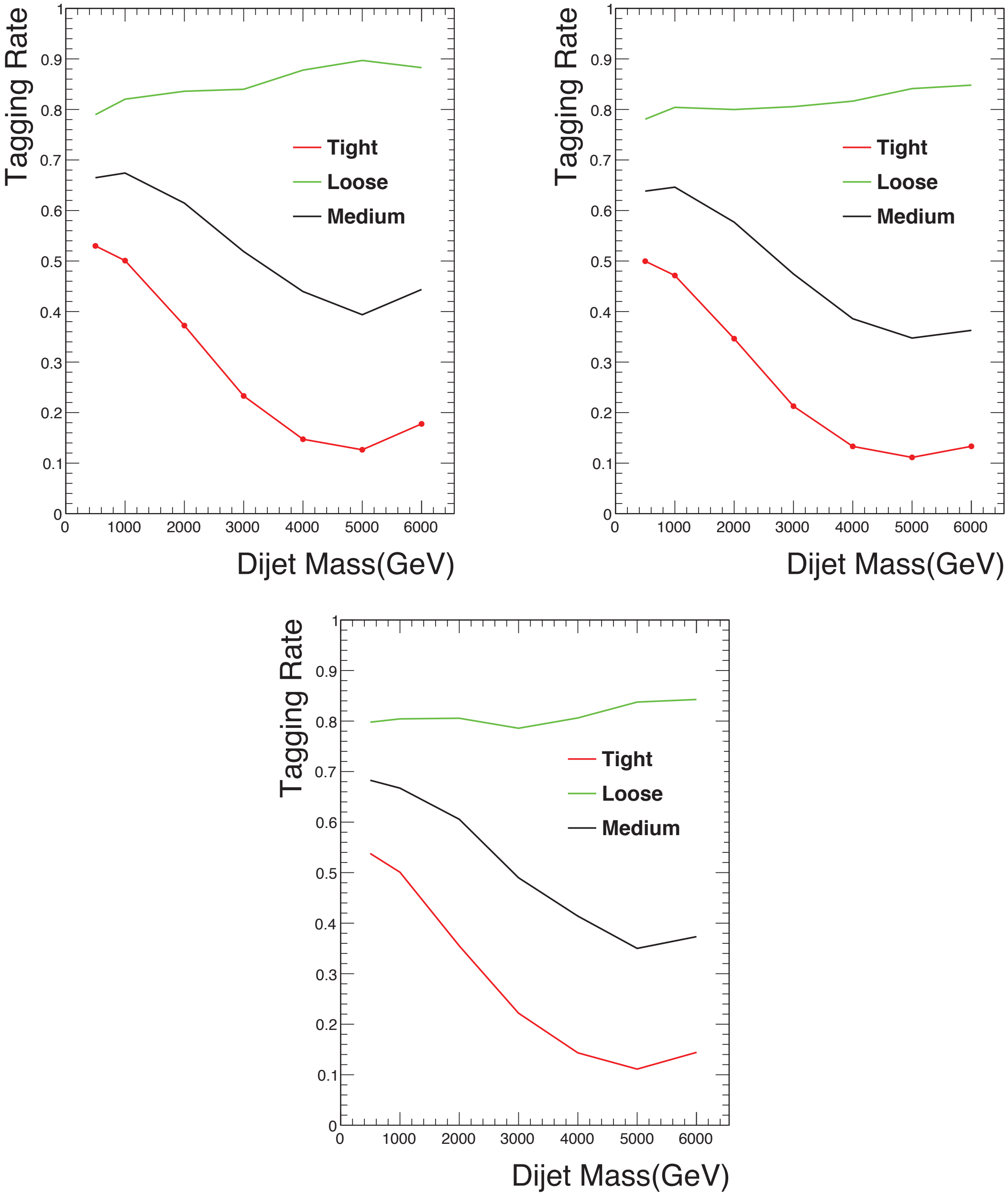


Figure 6.6: Loose (red), medium (black) and tight (green) WP of DeepJet tag rate of b* sample with only one b jet in the final state, showing 2016 (top left) and 2017 (top right) on top and 2018 on the bottom

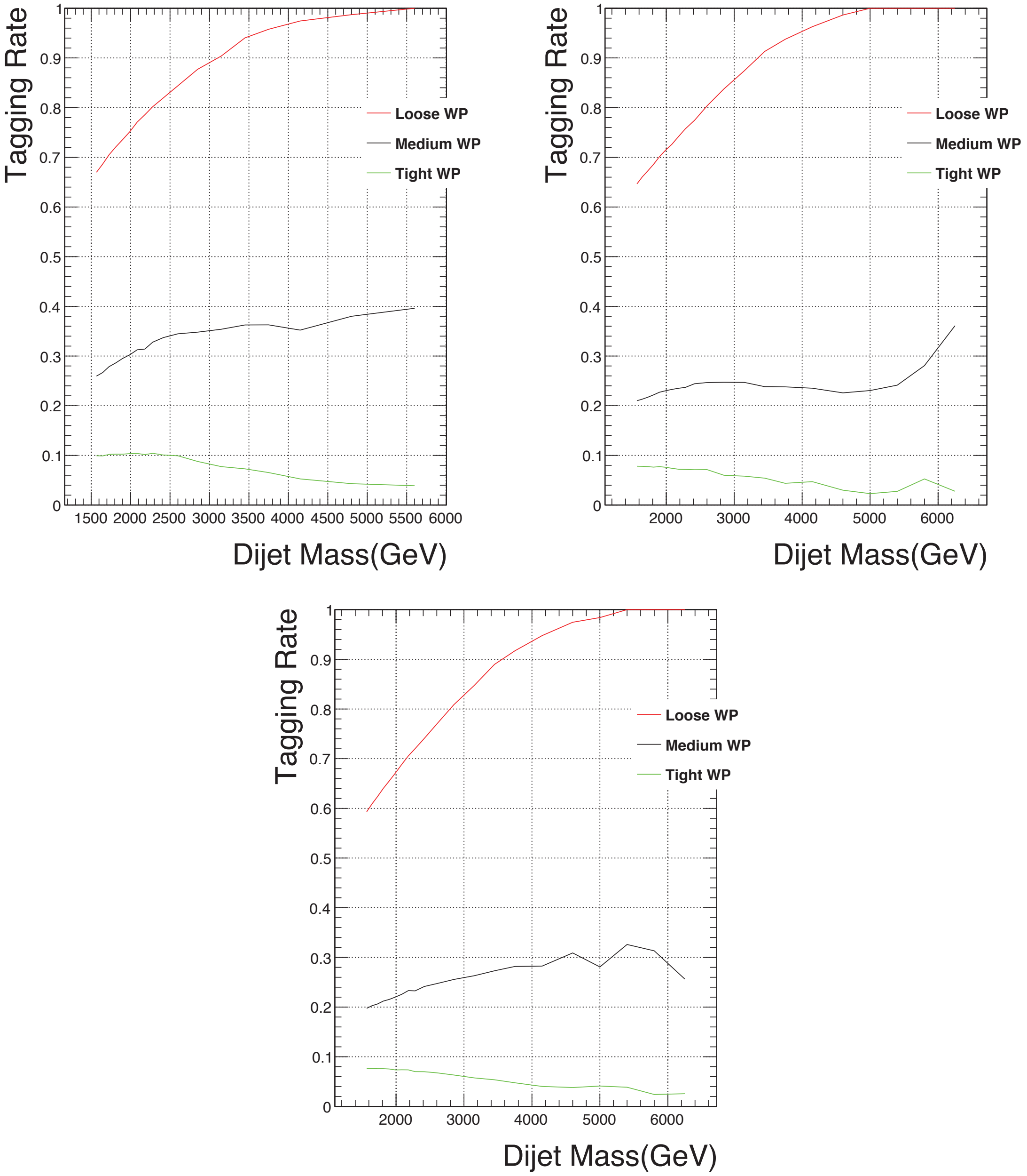


Figure 6.7: Loose (red), medium (black) and tight (green) WP of DeepJet tag rate of real collision data with only one b jet in the final state, showing 2016 (top left) and 2017 (top right) on top and 2018 on the bottom

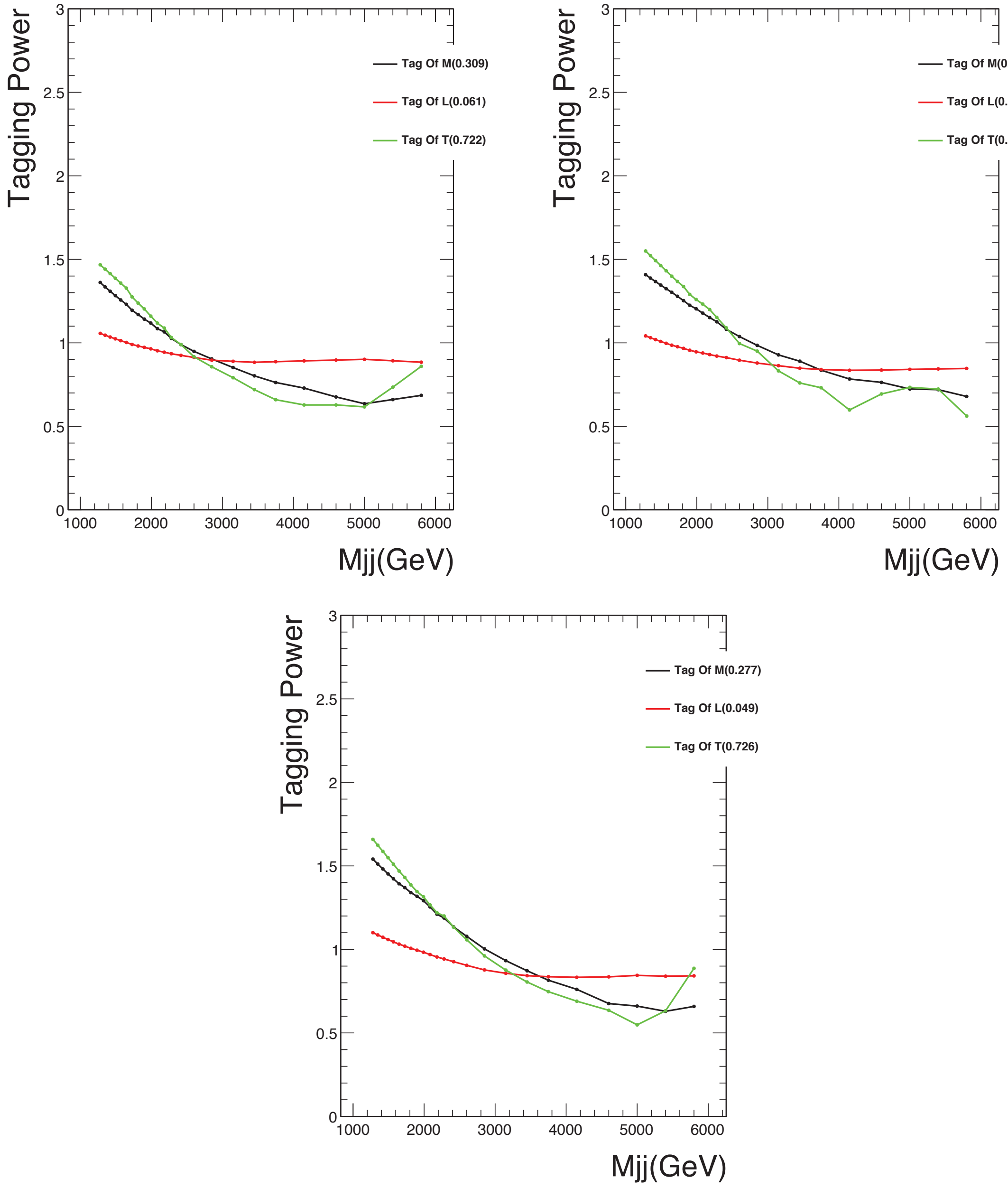


Figure 6.8: Tagging power of each year, 2016 (top left) and 2017 (top right) are on the top, and 2018 in the bottom.

have “bumps” that are not modeled by the analytic functions. The same method is used also in Ref. [95].

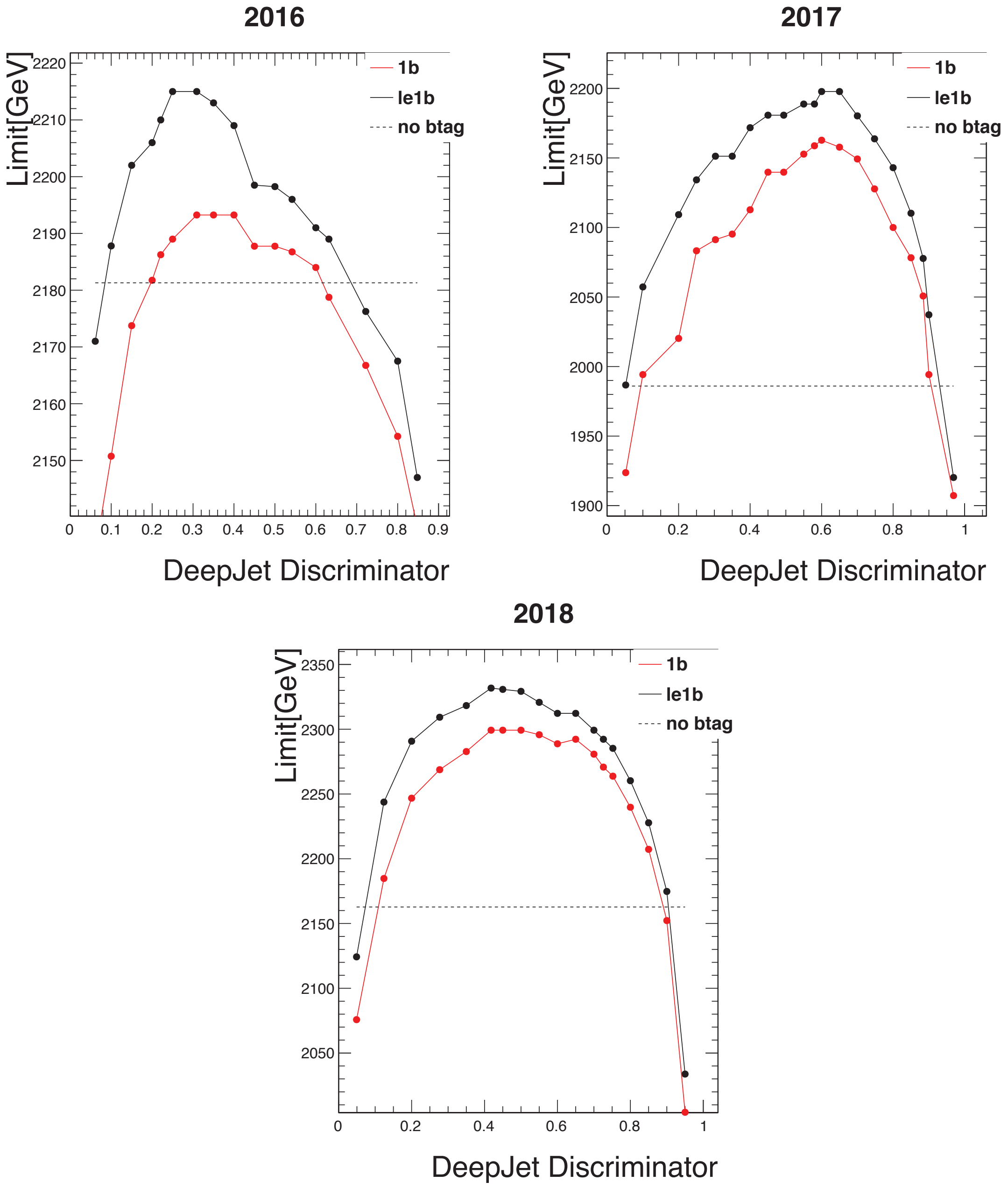


Figure 6.9: Limit scan of 2016 (top left), 2017 (top right) and 2018 (bottom), since no expect limit can be obtain, using 2 times of the final cross section compare with the theoretical cross section

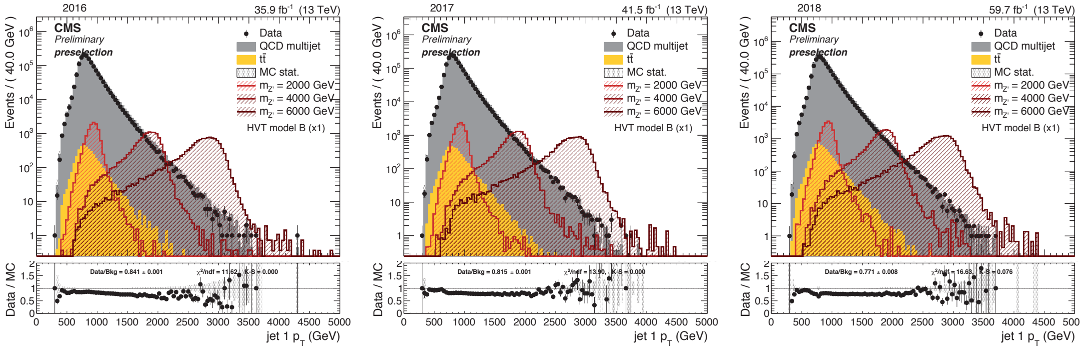


Figure 6.10: Leading jet p_T in 2016 (left), 2017 (center), and 2018 (right).

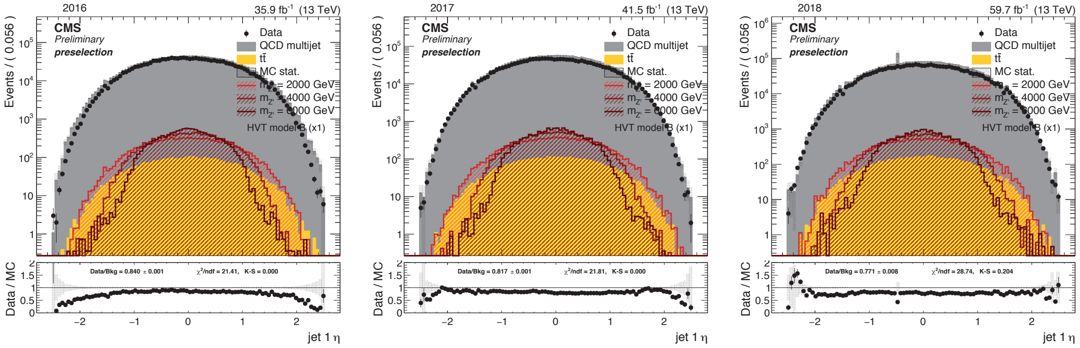


Figure 6.11: Leading jet η in 2016 (left), 2017 (center), and 2018 (right).

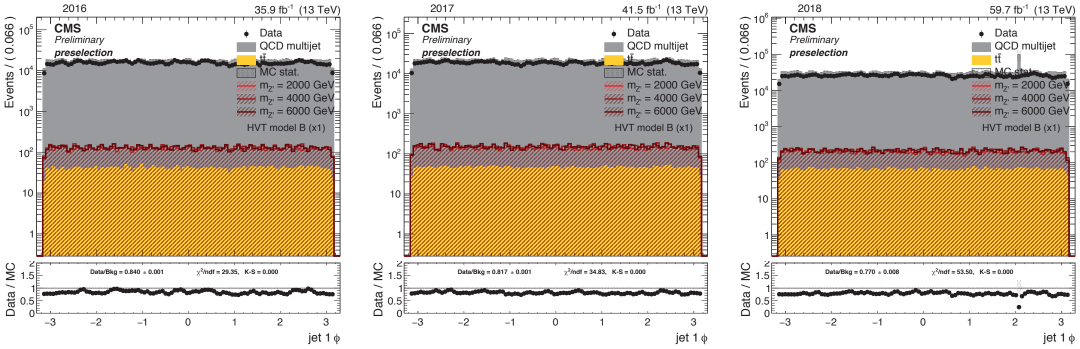


Figure 6.12: Leading jet ϕ in 2016 (left), 2017 (center), and 2018 (right).

The functions considered are power laws of the variable $m_X/\sqrt{s}$, where $\sqrt{s} = 13$ TeV, which defined as the observed number of events in each bin divided by the

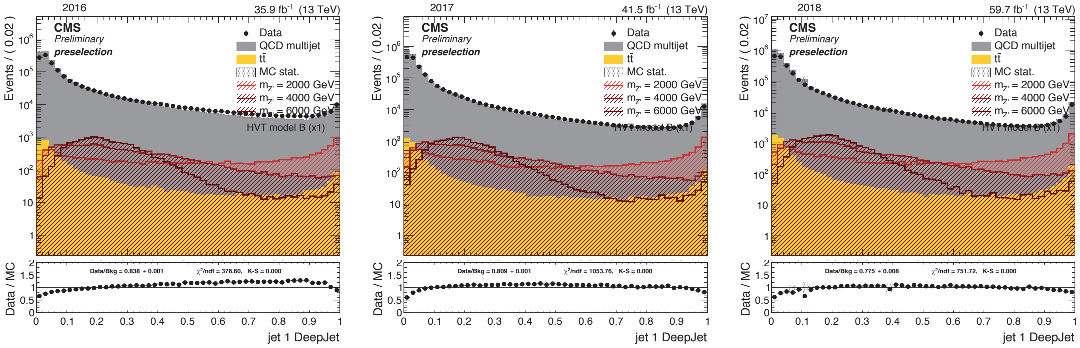


Figure 6.13: Leading jet DeepJet discriminator in 2016 (left), 2017 (center), and 2018 (right).

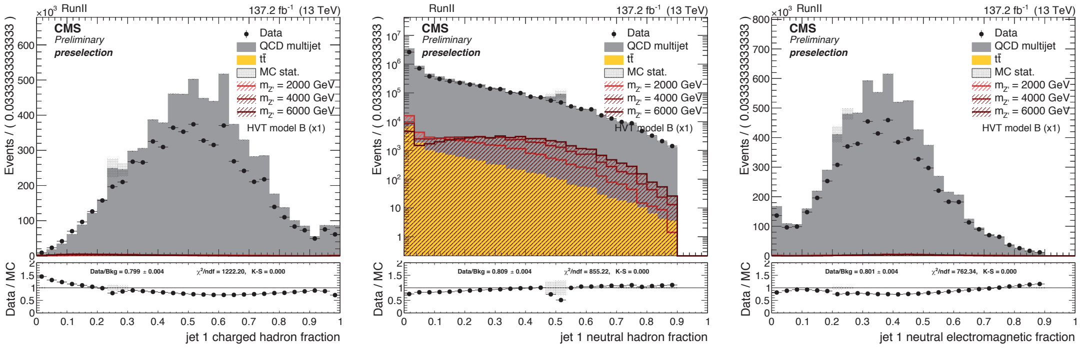


Figure 6.14: Leading jet charged hadron fraction (left), neutral hadron fraction (center), neutral electromagnetic fraction (right) in Run 2.

integrated luminosity and the bin width, with predefined bins of a width corresponding to the dijet mass resolution, and the number of parameters (including the normalization) spans between 4 and 6:

- **4 parameters :** $p_0 \cdot \dfrac{\left(1-\frac{m_X}{\sqrt{s}}\right)^{p_1}}{\left(\frac{m_X}{\sqrt{s}}\right)^{p_2+p_3\cdot\log\left(\frac{m_X}{\sqrt{s}}\right)}}$
- **5 parameters :** $p_0 \cdot \dfrac{\left(1-\frac{m_X}{\sqrt{s}}\right)^{p_1}}{\left(\frac{m_X}{\sqrt{s}}\right)^{p_2+p_3\cdot\log\left(\frac{m_X}{\sqrt{s}}\right)+p_4\cdot\log^2\left(\frac{m_X}{\sqrt{s}}\right)}}$

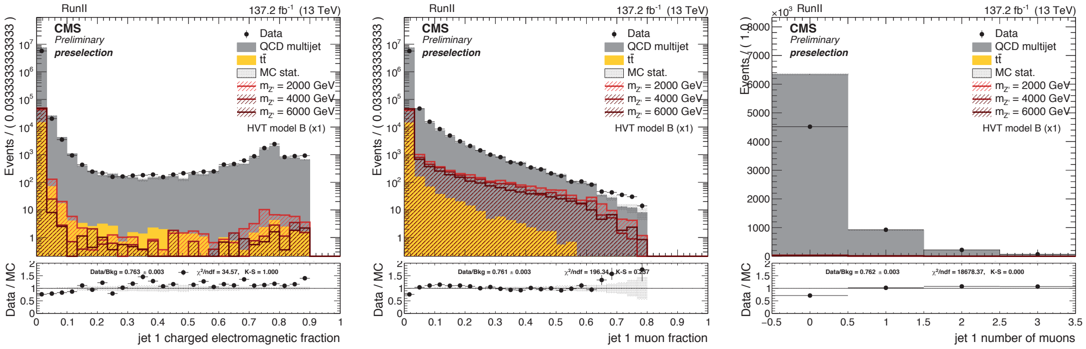


Figure 6.15: Leading jet charged electromagnetic fraction (left), muon fraction (center), muon multiplicity (right) in Run 2.

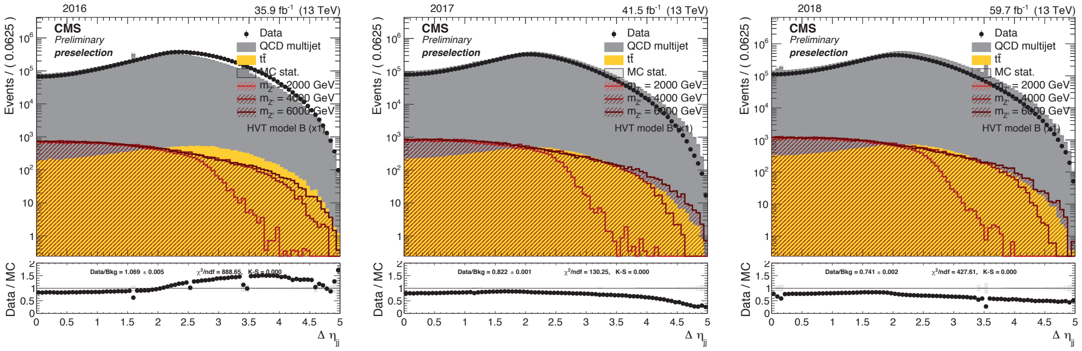


Figure 6.16: $\Delta\eta_{jj}$ between the two jets in 2016 (left), 2017 (center), and 2018 (right). The pre-selection on $\Delta\eta_{jj}$ has been omitted in this case in order to show the full spectrum.

- **6 parameters :** $p_0 \cdot \dfrac{\left(1-\frac{m_X}{\sqrt{s}}\right)^{p_1}}{\left(\frac{m_X}{\sqrt{s}}\right)^{p_2+p_3\cdot\log\left(\frac{m_X}{\sqrt{s}}\right)+p_4\cdot\log^2\left(\frac{m_X}{\sqrt{s}}\right)+p_5\cdot\log^3\left(\frac{m_X}{\sqrt{s}}\right)}}$

The functions, and the corresponding number of parameters, are chosen independently in each control and signal region with a Fisher test [96]. The F-test starts from basic fit quantities, the RSS (squared sum of the residuals of the fit), and the number of degrees of freedom (DOF). In this implementation, only bins with a non-zero number of entries are considered in the RSS. The test uses these quantities to compare one reference function and the function with one more parameter than the reference function to determine whether or not the additional

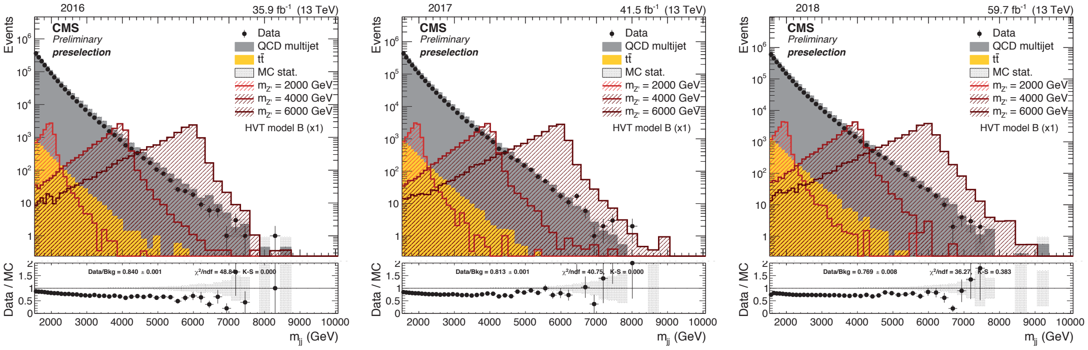


Figure 6.17: Dijet invariant mass in 2016 (left), 2017 (center), and 2018 (right).

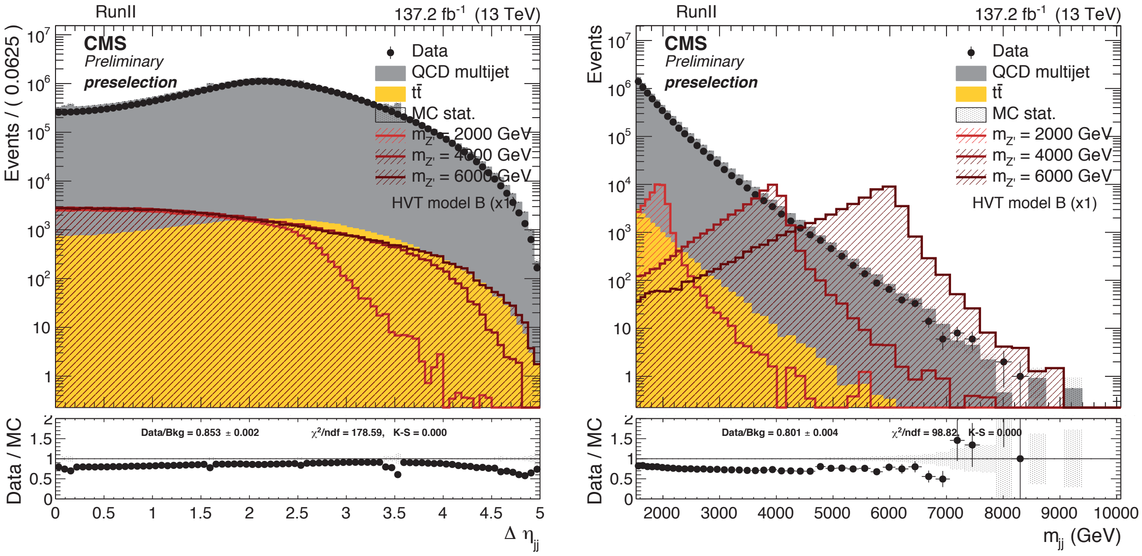


Figure 6.18: $\Delta\eta$ (left) and dijet invariant mass (right) between the two jets in the full Run 2 (2016, 2017, 2018). The pre-selection on $\Delta\eta_{jj}$ has been omitted in the left hand side plot in order to show the full spectrum.

parameter is needed. If the simpler fit function is correct, the relative increase in the sum of squares is expected to be smaller than the relative increase in DOF:

$$\frac{RSS_1 - RSS_2}{RSS_2} \lesssim \frac{n1 - n2}{n2}$$

where n_1 and n_2 are the DOF of function 1 and function 2, respectively. If the more complex model is correct:

$$\frac{RSS_1 - RSS_2}{RSS_2} > \frac{n1 - n2}{n2}$$

The Fisher test is based on the quantity F, defined as:

$$F = \left(\frac{RSS_1 - RSS_2}{n_2 - n_1}\right) / \left(\frac{RSS_2}{N - n_2}\right)$$

The F value is used to determine a CL after integrating the Fisher distribution from 0 to F, and N the number of bins:

$$CL = 1 - \int_0^F \mathcal{F} df$$

which gives the CL under the null hypothesis of the simpler function being sufficient. If $CL > 10\%$, the simpler function is sufficient, otherwise, more parameters are needed. In the latter case, the Fisher test is repeated by increasing the number of parameters of both functions, until a function with a sufficient number of parameters is found.

In order for the triggers to be fully efficient, the fits start from 1530 GeV. In most categories, a 3-parameter function describes the background sufficiently well, while in a few cases a 4-parameter function is necessary. Each ratio plot also reports the χ^2 and the number of degrees of freedom, calculated without considering the empty bins. The χ^2/NDF is always close or less than 1, certifying the goodness of the fits.

The measured dijet mass spectrum is defined by the following function:

$$\frac{d\sigma}{dm} = \frac{1}{\int Ldt} \frac{N_i}{\Delta m_i} \tag{6.3}$$

where m the dijet mass; N_i is the number of events in the i-th dijet mass bin; Δm_i is the width of the i-th dijet mass bin; and the integrated luminosity is $\int Ldt$. The bin width is approximately the dijet mass resolution, and gradually increases as a function of mass.

In Fig 6.19, 6.20, and 6.21 the fit to the dijet invariant mass distribution in data is reported for the $n_{b-tag} \geq 1$ category and separately for each years considered in this analysis. We perform an extended, background-only, binned, maximum likelihood fit to the data where events are required to have at least one b-tagged jets satisfying the medium WP. The result of the fit is presented using a 4, 5, and 6 parameters function separately. In all cases, a minimum cut on the dijet mass ($m_{jj} > 1.53$ TeV) is applied to ensure full trigger efficiency in the fitted range. In order to test the goodness-of-fit (GOF), we define a modified chi-square test statistics as illustrated in AN-16-202 [97] and we generate 10,000 pseudodatasets from the best-fit model parameters on data, refit each pseudodataset with a maximum likelihood fit, and save the test statistic value. Further Fisher tests are done to decide the parameter of function to use, results of which are shown in tables 6.15 for 2016 fits, 6.16 for 2017 fits and 6.17 for 2018 fits.

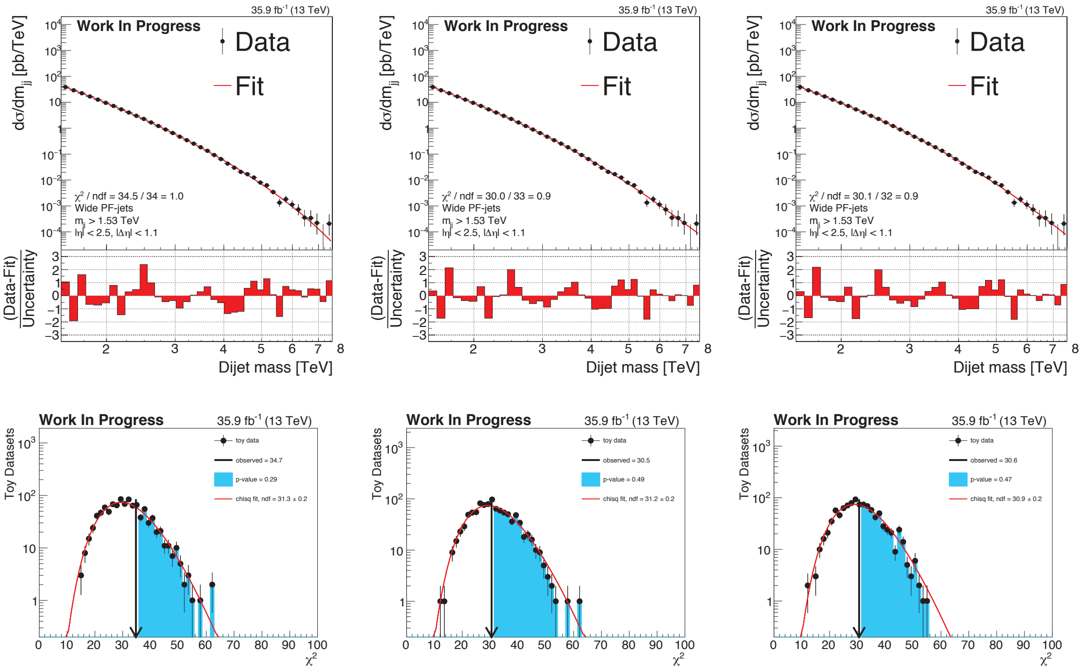


Figure 6.19: Fit to the dijet invariant mass distribution for the 2016 dataset using a 4, 5 or 6 parameter function (upper row). Toy distribution for the goodness-of-fit study using the chi-square test statistic (bottom row).

We see the 6 parameters fitting function is better than the 5 parameters fitting function, while the 5 parameters fitting function is better than the 4 parameters fitting function, for 2016 m$_{jj}$. But, according to the F-score credit level, the credit

Table 6.15: F-test to fitting fucntions on 2016 data

	RSS1	n1	RSS2	n2	F-score	CL
4para-5para	4.88e-8	4	5.52e-8	5	3.8	5.98%
5para-6para	5.52e-8	5	5.66e-8	6	0.8	37.78%

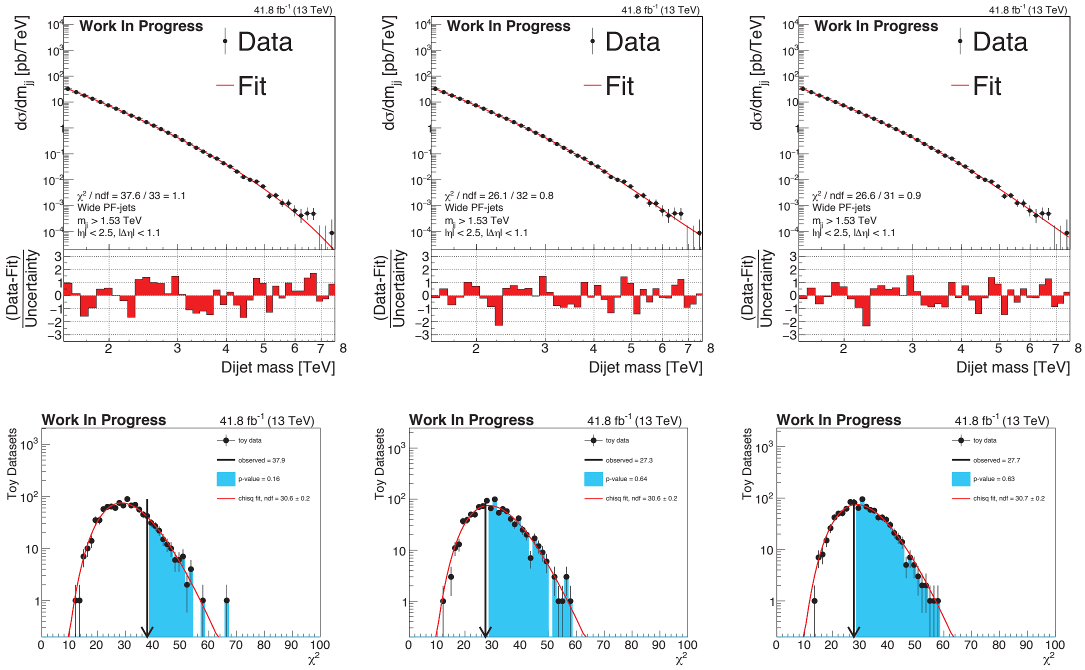


Figure 6.20: Fit to the dijet invariant mass distribution for the 2017 dataset using a 4, 5 or 6 parameter function (upper row). Toy distribution for the goodness-of-fit study using the chi-square test statistic (bottom row).

level for 5-para against 4-para is lower than 10% while 6-para against 5-para is beyond 10%. The 5 parameter function is decided to be used for fitting the 2016 data.

In summary, a 5 parameter function is used for the 2016 and 2017 datasets, while for the 2018 dataset a 4 parameter function is used.

Base on the best fit of each year, the χ^2 per number of degrees of freedom of the fit is $\chi^2/NDF = 39/33$ for 2016, 26.1/32 for 2017, and 20.2/31 for 2018. There is no evidence for a narrow resonance in the data. The lower panel shows the difference between the data and the fitted parametrization, divided by the statistical

Table 6.16: F-test to fitting fucntions on 2017 data

	RSS1	n1	RSS2	n2	F-score	CL
4para-5para	4.88e-8	4	5.52e-8	5	3.8	5.98%
5para-6para	5.52e-8	5	5.66e-8	6	0.8	37.78%

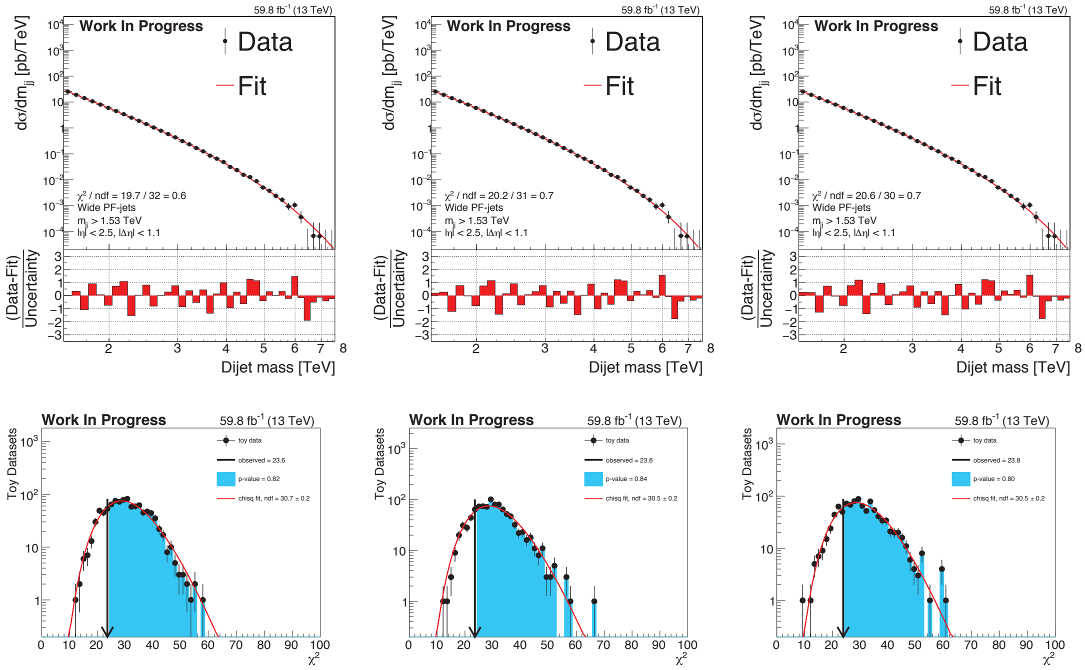


Figure 6.21: Fit to the dijet invariant mass distribution for the 2018 dataset using a 4, 5 or 6 parameter function (upper row). Toy distribution for the goodness-of-fit study using the chi-square test statistic (bottom row).

uncertainty of the data.

6.5 Signal Modeling

In order to see the signal sample quality, the signal sample needs to be checked. This analysis searches for narrow dijet resonances whose natural resonance width is small compared to the CMS dijet mass resolution. For such resonances, the resonance shape is dominated by the experimental resolution and to a large extent depends only on the type of parton pairs in the resonance decay.

Table 6.17: F-test to fitting fucntions on 2018 data

	RSS1	n1	RSS2	n2	F-score	CL
4para-5para	1.97e-8	4	1.77e-8	5	-3.4	null
5para-6para	1.77e-8	5	1.725e-8	6	-0.87	null

For the case of the b*, Z', and dark matter mediator model, the signal shapes produced for resonance masses up to 6 TeV are reported in Fig 6.22 and 6.23. The intrinsic width of the resonances is narrow when compared to the detector resolution. The contribution of the low mass tail to the line shape becomes more important as we go higher in mass and depend on the parton content of the resonance. For high-mass resonances, there is a significant contribution that depends both on the parton distribution functions and on the natural width of the Breit-Wigner distribution as described in [98].

More checks on the signal shape after applying the btag of each category and each WP are shown in the Fig 6.24, 6.25, and 6.26. It's shown that the signal shape doesn't change after applying each category b-tag. Except for the tight WP, and the changes are due to the low statistics because they cut it too tight. In addition, the same check on the dark matter mediator samples are shown in 6.27, applying medium WP.

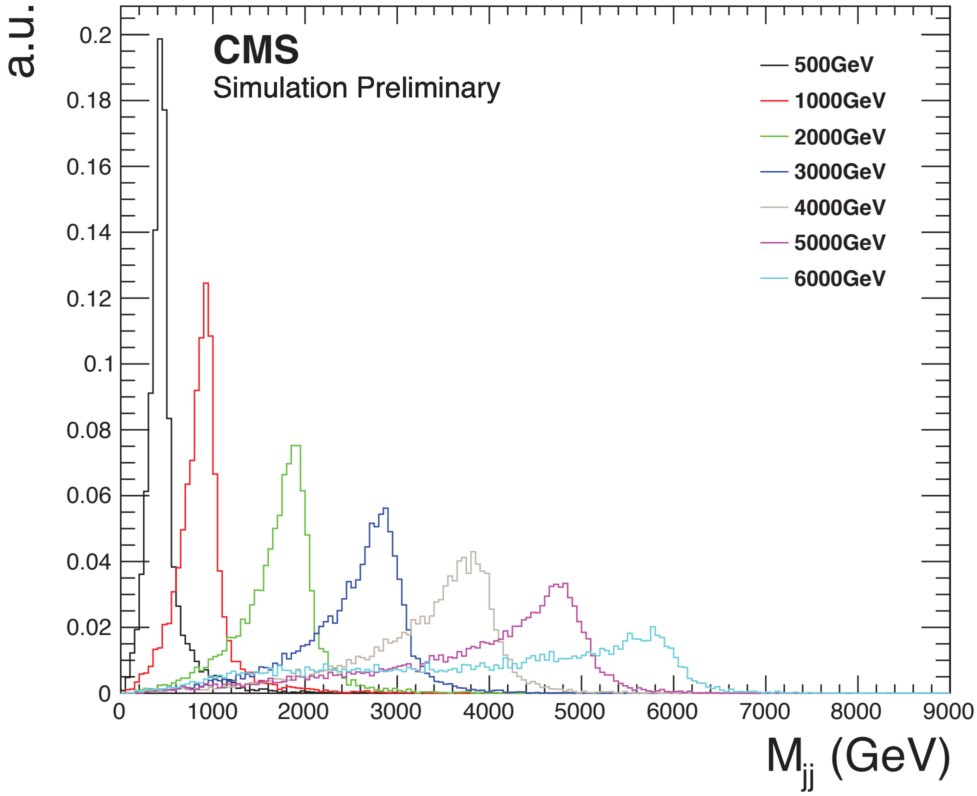


Figure 6.22: b* model signal shapes from reconstructed dijet mass spectra from the `PYTHIA` 8 MC event generator including simulation of the CMS detector.

Figs. 6.24-6.26 show the comparison between signal with only one b jet in the final state, apply no b-tag and applying exact one, exact two DeepJet b-tagger with loose, medium and tight WP of each year, while Fig 6.27 show the comparison Z' signal shapes, with two b jets in the final state.

6.6 Systematics

6.6.1 Jet Energy Scale

The uncertainty on the JES is the uncertainty that comes from the difference between simulation and data on the scale. If the simulated scale has a high response, then the position of simulated resonance will be placed at the higher mass region than it should be placed. The JES uncertainty is only on the simulated signal but the background since the background is coming from the only fit of data

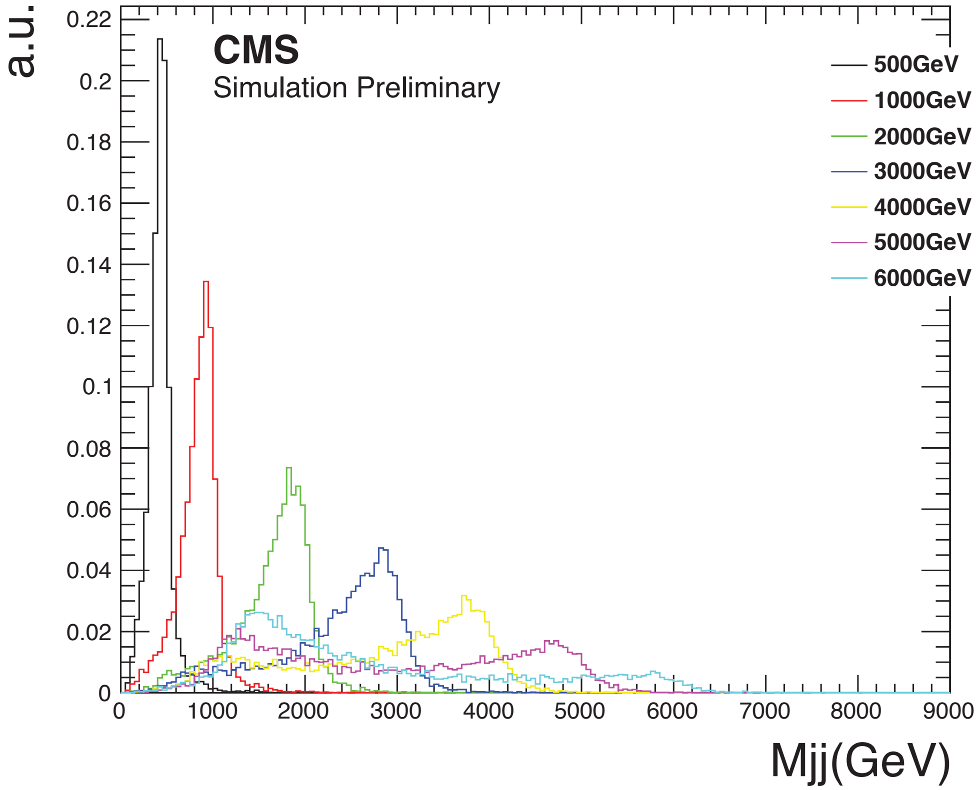


Figure 6.23: Z' signal shapes from reconstructed dijet mass spectra from the `PYTHIA` 8 MC event generator including simulation of the CMS detector.

in our process. So this uncertainty is the uncertainty about the position of the simulated resonance signal. Official CMS JES uncertainty is studied and provided with Jet energy correction [99].

6.6.2 Jet Energy Resolution

The uncertainty on the JER is the uncertainty which also comes from the difference between simulation and data on the jet energy resolution. If the simulated JER is bigger than the JER from data, the width of the simulated resonance is broader than the real resonance. The JER uncertainty is only on a simulated signal like JES uncertainty. The uncertainty on JER at high dijet mass is ±10%. The tails of the resolution function are in good agreement between data and MC.

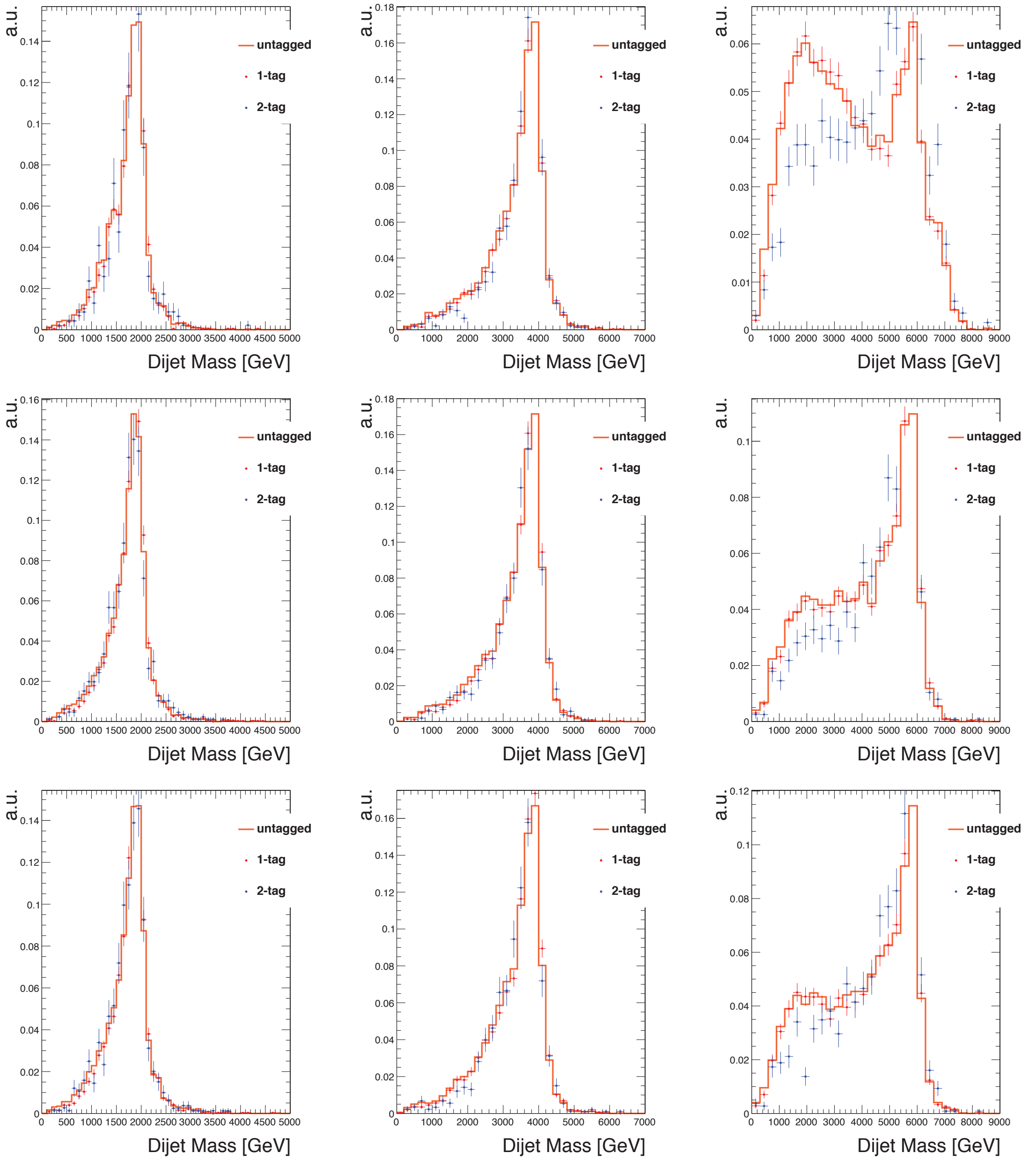


Figure 6.24: 2016 (top), 2017 (middle) and 2018 (bottom) b* signal shape applying loose WP DeepJet b-tagger at 2 TeV (left), 4 TeV (central) and 6 TeV (right)

6.6.3 Luminosity Uncertainty

This is the uncertainty which shows the error of measuring the total luminosity

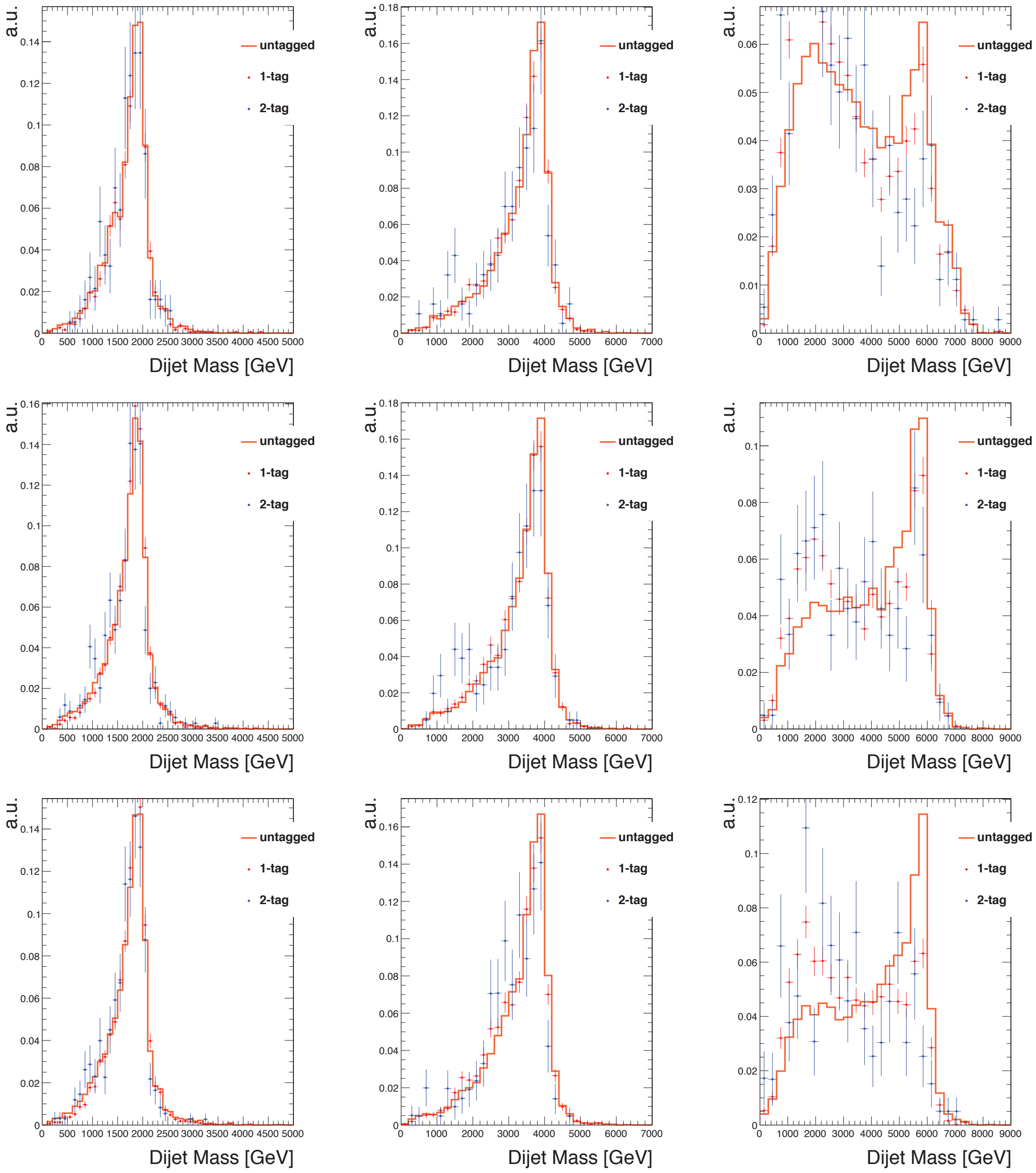


Figure 6.25: 2016 (top), 2017 (middle) and 2018 (bottom) b* signal shape applying medium WP DeepJet b-tagger at 2 TeV (left), 4 TeV (central) and 6 TeV (right)

6.6.4 b-tag Uncertainty

To deduce the uncertainty on the b-tagging efficiency, we followed the method of re-weighting events using scale factors and MC b-tagging efficiencies as

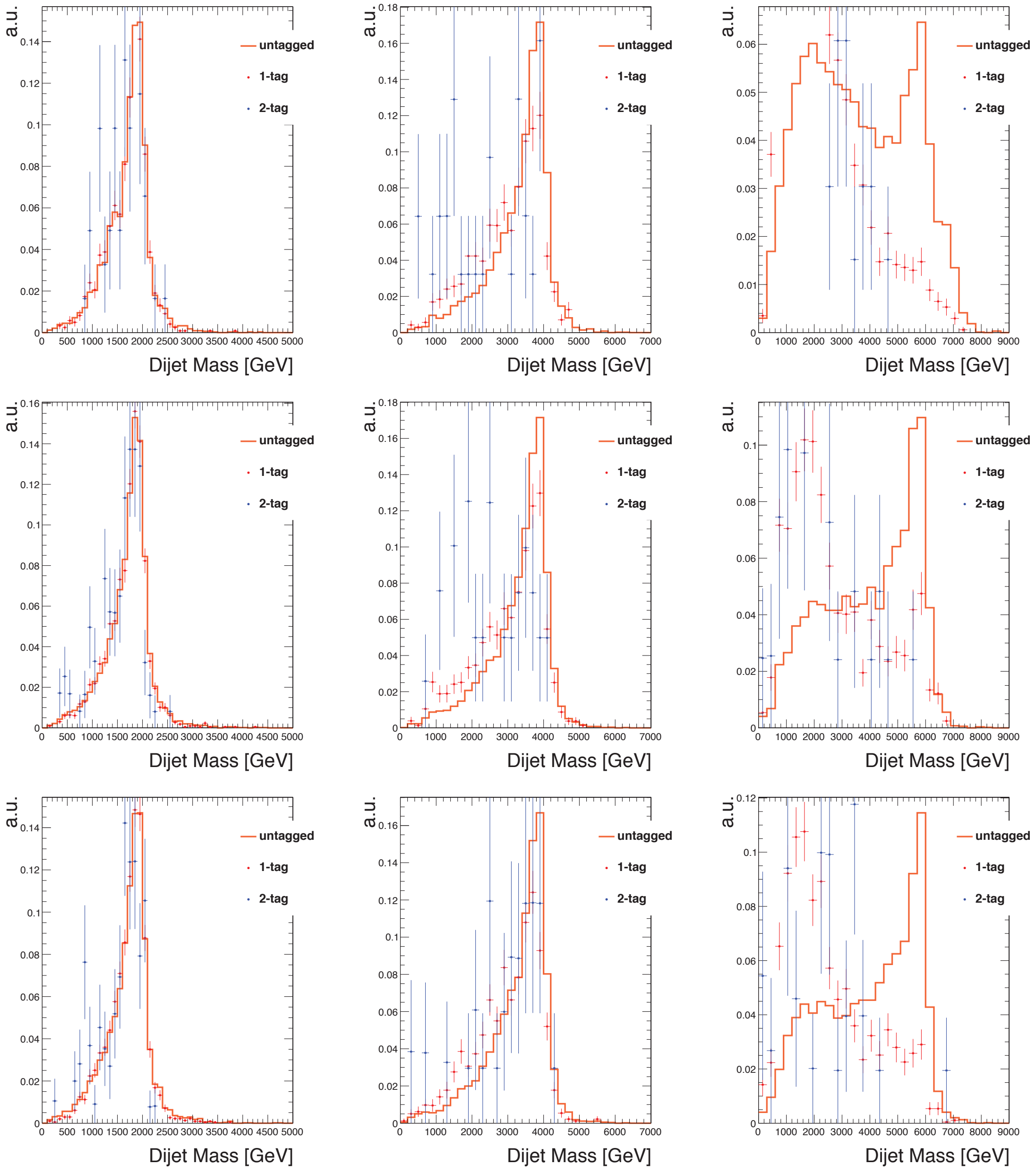


Figure 6.26: 2016 (top), 2017 (middle) and 2018 (bottom) b* signal shape applying tight WP DeepJet b-tagger at 2 TeV (left), 4 TeV (central) and 6 TeV (right)

recommended by the b-tagging physics object group [100]. This method consists of changing the weights of the selected MC events. A given configuration of jets in MC and data has the following probability:

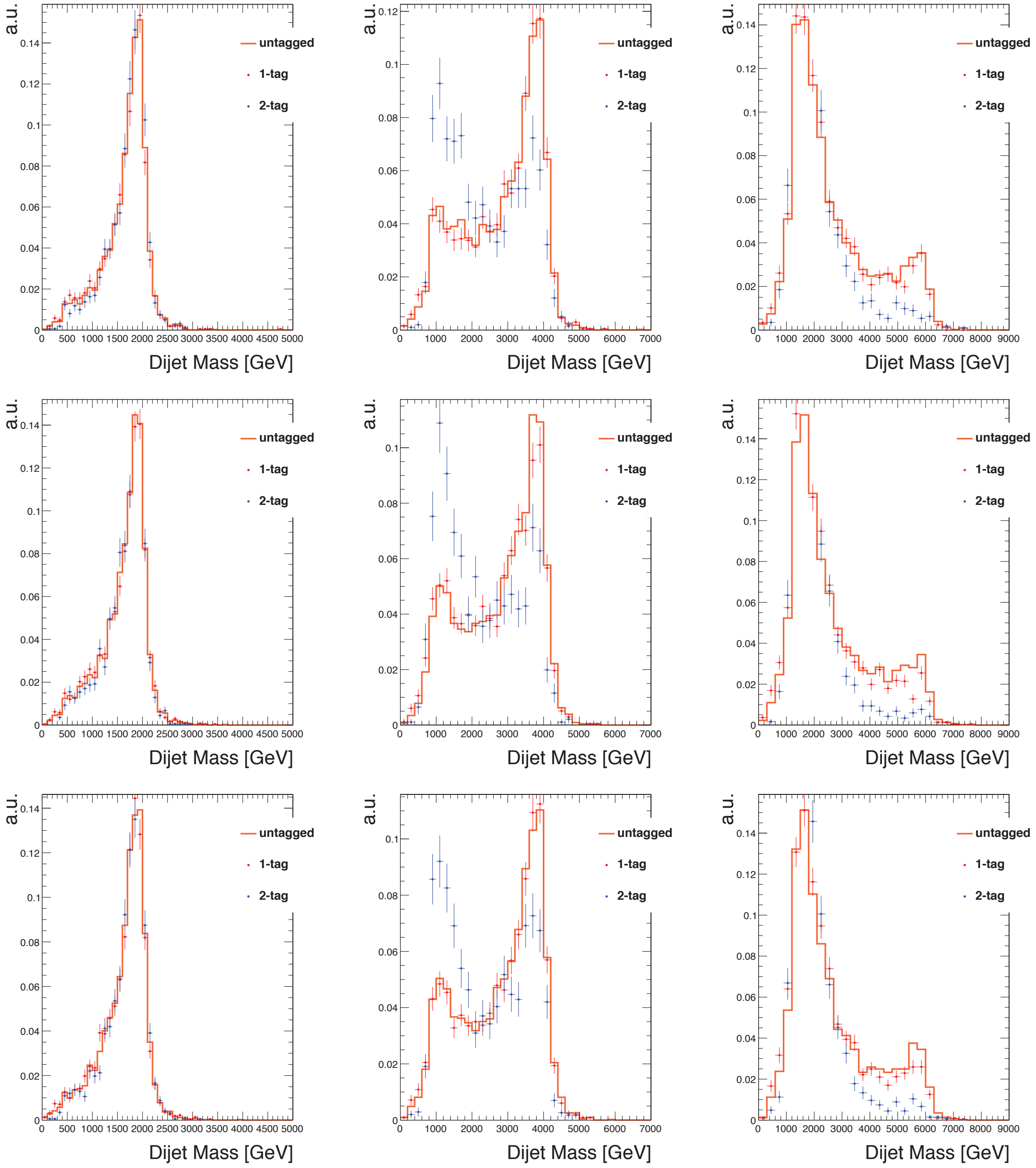


Figure 6.27: 2016 (top), 2017 (middle) and 2018 (bottom) Z' signal shape applying tight WP DeepJet b-tagger at 2 TeV (left), 4 TeV (central) and 6 TeV (right)

$$P(\mathrm{MC}) = \prod_{i=\mathrm{tagged}} \varepsilon \prod_{j=\mathrm{untagged}} (1-\varepsilon_j) \qquad (6.4)$$

$$P(\text{DATA}) = \prod_{i=\text{tagged}} \text{SF}_i\varepsilon_i \prod_{j=\text{untagged}} (1 - \text{SF}_j\varepsilon_j) \tag{6.5}$$

with the MC b-tagging efficiencies ε_i and scale factors SF_i, which both are functions of the jet flavor, jet p_T and jet η. The MC event weight is then computed as

$$w = \frac{P(\text{DATA})}{P(\text{MC})} \tag{6.6}$$

The systematic uncertainty on the b-tagging efficiency is deduced via the uncertainty on the scale factor measurement. These are evaluated separately for each year and mass point. In accordance with the prescription of the b-tagging physics object group, the b-tagging uncertainty for jets with p_T >1 TeV is doubled because the scale factors have not been measured beyond this threshold. Figs. 6.28 show the SF v.s. AK4 jet p_T, it has no dependence on η

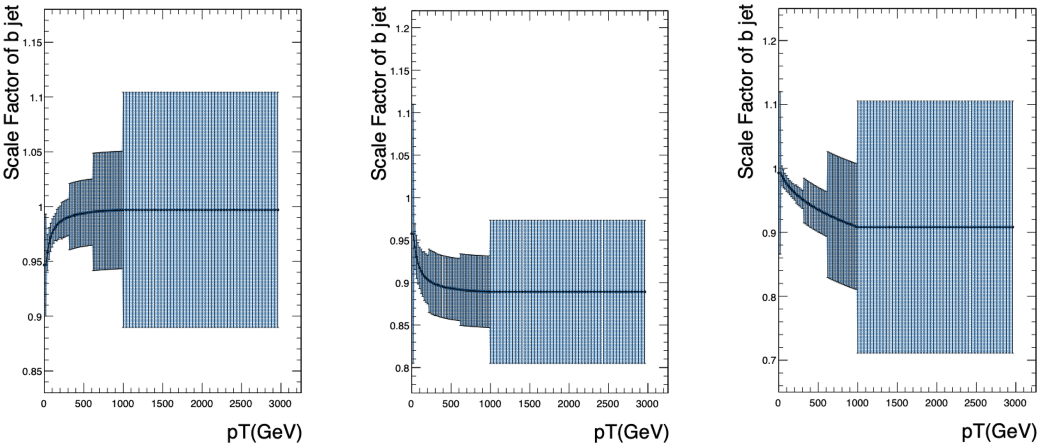


Figure 6.28: Exctracted DeepJet SF of 2016 (left), 2017 (central) and 2018 (bottom) v.s. AK4 jet p_T, it has no dependance on η.

6.6.5 Total Systematic Uncertainty

The systematic uncertainties on the background estimation are accounted for by fitting directly to the data. For the signal estimation, we included four main sources of systematic uncertainties: the uncertainty on the luminosity of 2.6%, the jet energy scale uncertainty of 2%, the jet energy resolution uncertainty of 10%, as well as the b-tagging efficiency uncertainty. The uncertainty on the jet energy scale and resolution is applied to the mean and the width of the signal fit model respectively.

The total systematic uncertainty is shown the table 6.18

Table 6.18: Summary of systematic uncertainties propagated to the signal prediction in RECO analysis.

Systematic Uncertainty Source	Nominal Value	Uncertainty
Jet Energy Scale	no shift	±2% shift of m_{jj}
Jet Energy Resolution	no smearing	10% of RECO resolution
btag Uncertainty	no shift	Depend on jet p_T
Luminosity	137.5 fb^{-1}	±2.6%

6.7 Results

The CLs criterion [85; 86] is used to determine the 95% confidence-level limit on the signal contribution in the data, using the `RooStats` package implemented in the CMS `combine` tool. The `Asymptotic` method [87] is used to calculate preliminary 95% C.L. upper limits with 1σ and 2σ bands using the CLs frequentist calculation currently recommended by the LHC Higgs Combination Group [88]. The `ProfileLikelihood` method is used for significance and the background p-value; finally, the `MaxLikelihoodFit` method allows to get the signal *best fit ratio*, the fit pulls and the pre/post fit distributions.

Results are obtained from a combined signal and background fit the binned m_X distribution, based on a profile likelihood defined as

$$\mathcal{L} = \prod_i \frac{\mu_i^{n_i} \cdot e^{-\mu_i}}{n_i!} \qquad \text{with} \qquad \mu_i = \sigma N_i(S) + N_i(B)$$

where $N_i(S)$ and $N_i(B)$ are the i-th signal and background events, and σ is the signal strength modifier parameter. While the signal shape parameters are held constant after the fit to the Monte Carlo signal, the background shape, and normalization parameters are still left free to float in the combined fit. Systematic uncertainties are treated as nuisance parameters and are profiled in the statistical interpretation. The background-only hypothesis is tested against the signal in the considered categories, and with no evidence of significant deviations from background expectation, the asymptotic modified frequentist method is used to determine the limit at the 95% CL on the signal contribution. The 95% CL upper limit, σ_{95}, is calculated from the posterior probability density P_{POST} as follows:

$$\frac{\int_0^{\sigma_{95}} P_{POST}(\sigma)d\sigma}{\int_0^{\inf} P_{POST}(\sigma)d\sigma} = 0.95 \tag{6.7}$$

In the absence of excesses, a 95% CL limit on the b* and Coloron model is set by looking at the category where at least one of the two leading jets is b-tagged. In this case, upper limits are expressed in terms of $\sigma \times BR \times A$ where A is the acceptance accounting for the kinematic requirements $\Delta\eta_{jj} < 1.1$ and $|\eta| < 2.5$.

Upper limits of b* and Coloron are reported for the full Run II dataset in Fig 6.29 and Fig 6.31, while in Fig 6.30 and Fig 6.32 for each year separately. The expected limit excludes b* between 1667 GeV and 2100 GeV The observed limit excluded b* before 1904 GeV and between 2175 GeV and 2351 GeV. The expected limit excludes Coloron before 4958 GeV. The observed limit excluded Coloron before 4707 GeV.

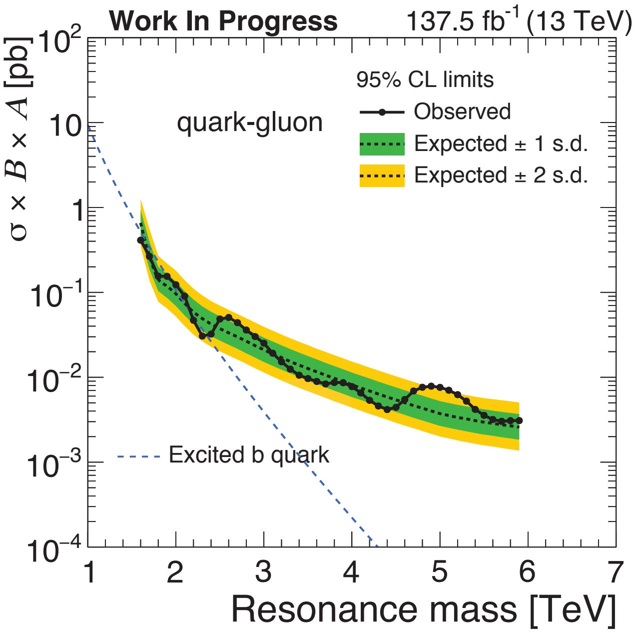


Figure 6.29: The observed 95% CL upper limits on the product of the cross section, branching fraction, and acceptance for dijet resonances decaying to bg. The corresponding expected limits (dashed) and their variations at the 1 and 2 standard deviation levels (shaded bands) are also shown. Limits are compared to predicted cross sections for a b* model. The expected limit excludes b* between 1667 GeV and 2100 GeV The observed limit excluded b* before 1904 GeV and between 2175 GeV and 2351 GeV

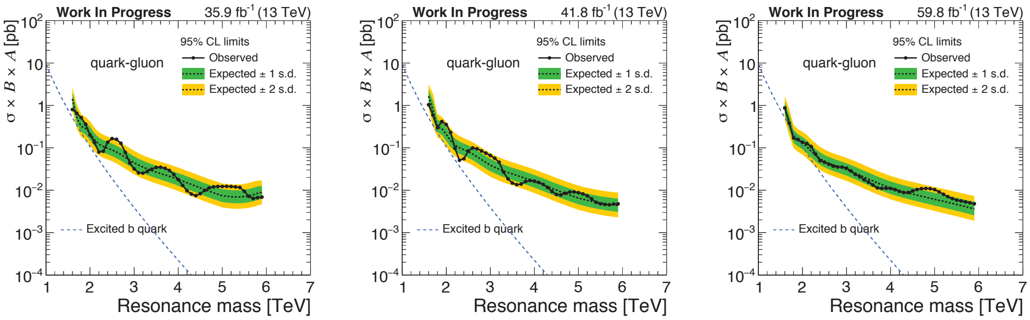


Figure 6.30: Similar to Fig 6.29, but this time limits are reported separately for each year.

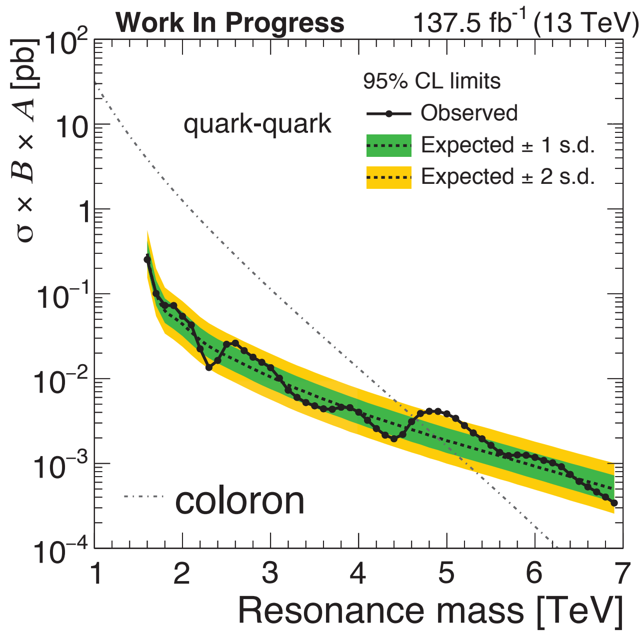


Figure 6.31: The observed 95% CL upper limits on the product of the cross section, branching fraction, and acceptance for dijet resonances decaying to bg. The corresponding expected limits (dashed) and their variations at the 1 and 2 standard deviation levels (shaded bands) are also shown. Limits are compared to predicted cross sections for a Coloron model. The expected limit excludes Coloron before 4958 GeV. The observed limit excluded Coloron before 4707 GeV.

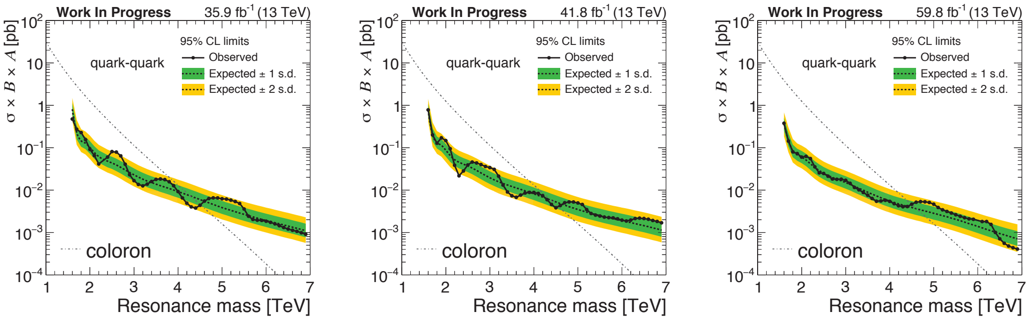


Figure 6.32: Similar to Fig 6.31, but this time limits are reported separately for each year.

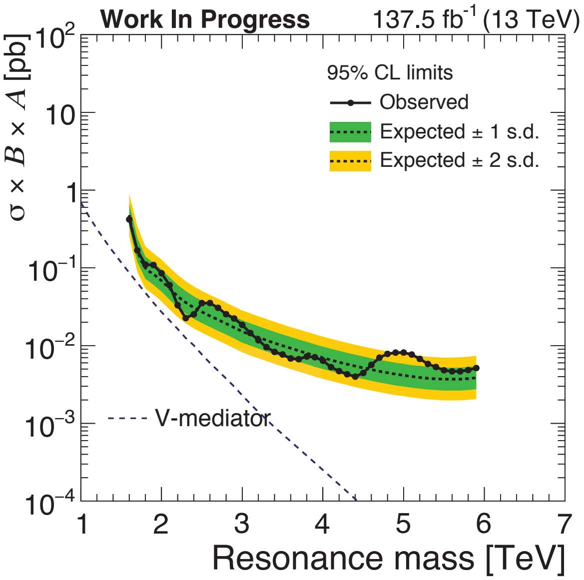


Figure 6.33: The observed 95% CL upper limits on the product of the cross section, branching fraction, and acceptance for dijet resonances decaying to bb. The corresponding expected limits (dashed) and their variations at the 1 and 2 standard deviation levels (shaded bands) are also shown. Limits are compared to predicted cross sections for a leptophobic vector dark matter mediator model. Unfortunately, we can not get any observed or expected limit.

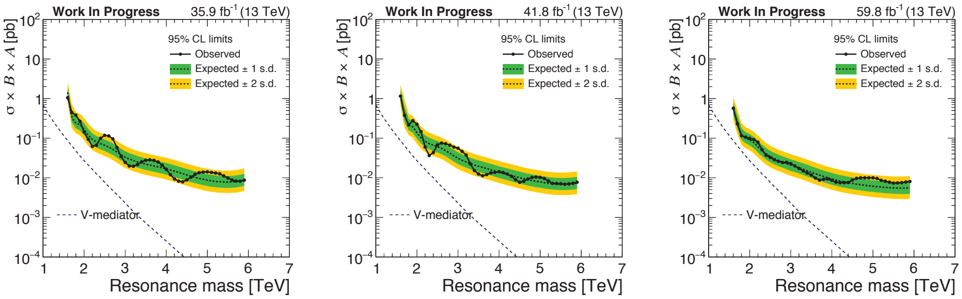


Figure 6.34: Similar to Fig 6.33, but this time limits are reported separately for each year.

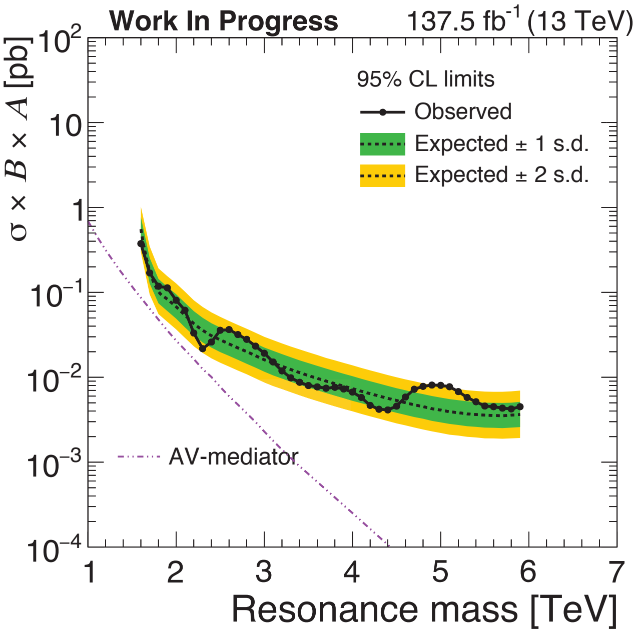


Figure 6.35: The observed 95% CL upper limits on the product of the cross section, branching fraction, and acceptance for dijet resonances decaying to bb. The corresponding expected limits (dashed) and their variations at the 1 and 2 standard deviation levels (shaded bands) are also shown. Limits are compared to predicted cross sections for a Leptophobic Axial-Vector Dark Matter mediator model. Unfortunately, we can not get any observed or expected limit.

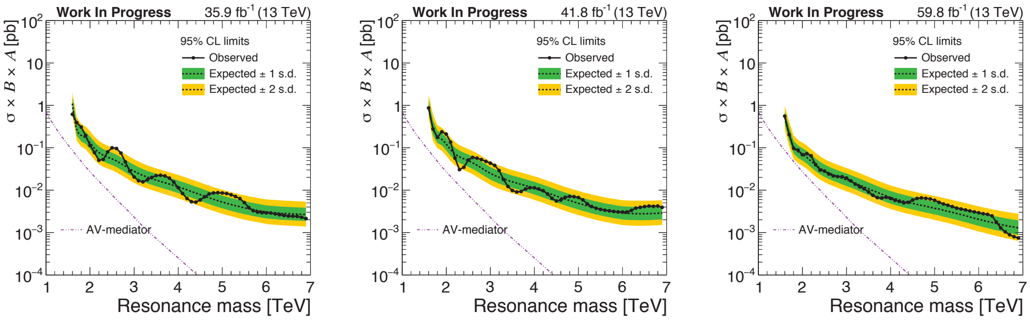


Figure 6.36: Similar to Fig 6.35, but this time limits are reported separately for each year.

The 95% Upper limit on the vector leptophobic dark model mediator is shown as Fig 6.33 and Fig 6.34. The 95% Upper limit on the axial-vector leptophobic dark model mediator is shown as Fig 6.35 and Fig 6.36.

Searches for resonances decaying into a pair of jets have been performed using proton-proton collisions at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of up to 137.5 fb^{-1} applying b-tag algorithm has been performed. We set expected and observed limit on b* and Coloron model at 95% CL. The dijet mass spectra are observed to be smoothly falling distributions. In the analyzed data samples, there is no evidence for resonant particle production.

6.8 Dark Matter Interpretation

Following the approach from previous publications [82], we search for Dark Matter (DM) and refer to simplified models where DM mediators are leptophobic vector and axial-vector mediators that decay only into quark pairs and DM particles with unknown mass m_{DM}. In the model we use, the DM particle is assumed to be a Dirac Fermion and the particle mediating the interaction is exchanged in the s-channel. 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, and the mediator coupling to DM g_{DM}.

Following the prescriptions of the Dark Matter Working Group [101] we set the universal quark coupling $g_q = 0.25$ and $g_{DM} = 1.0$ while m_{DM} is set to 1 GeV. In order to find which mass of the mediator is excluded, we perform a scan over the mediator mass between 1 TeV and 8 TeV in 1 TeV steps using `MG5` which allows us to obtain the cross-section at the parton level. Also in this case the usual kinematic constraints are applied ($\Delta\eta_{jj} < 1.1$ and $|\eta| < 2.5$) and the dijet mass is required to be within a 500 GeV mass window of the resonances mass.

Similarly to what has been done for the b* search described earlier, events are required to have at least one b-tagged jet among the two leading jets.

The 95% CL upper limits on the universal quark coupling g'_q as a function of resonance mass for a leptophobic Z' resonance that only couples to quarks are reported in Fig 6.37. This plots agrees with the dark matter model limit in 6.35 and 6.33. We don't have limit when dark matte mass is close to 0. But if we want to have the same limit from Z' which only couple to quark at this dark matter mass, the coupling between Z' and quark is around. When dark matter mass is to high and can't be produced (m$_{DM}$ > M$_{med}$/2 dash line), the coupling is the same as the coupling between dark matter mediator and quark, which is 0.25.

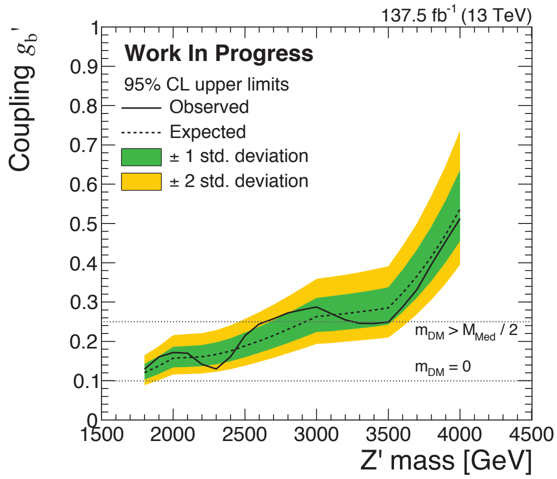


Figure 6.37: Upper limits on the coupling as a function of mass for a model of a leptophobic Z' resonance with a universal quark coupling, g'_q

6.9 Bias study

In order to estimate the potential bias introduced by the choice of the functional background model, Monte Carlo dijet invariant mass distributions were produced under the signal hypothesis and using the best fitting background model. A background model function with one more parameter was then used to fit this toy distribution. The difference between the measured and the injected signal strength Δr divided by the fit uncertainty σ_r yields an approximately Gaussian distribution when repeated on multiple toy mass spectra. The mean of this distribution is interpreted as the potential bias of choosing one model when in fact the other one would be correct.

The results of a bias study when considering the following two functions:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}m_{jj}} = \frac{p_0 \mathrm{exp}[-p_2 x - p_3 x^2]}{x^{p_1}} \qquad (4 - \text{parameter ATLAS}) \ , \tag{6.8}$$

$$\frac{\mathrm{d}\sigma}{\mathrm{d}m_{jj}} = \frac{p_0 \mathrm{exp}[-p_2 x - p_3 x^2 - p_4 x^3]}{x^{p_1}} \qquad (5 - \text{parameter ATLAS}) \ . \tag{6.9}$$

Figs. 6.38, 6.39, and 6.40 show the fits on m_{jj} of 2016, 2017, and 2018 respectively. Same parameter functions are used for each year, 5 parameter function for 2016 and 2017 data, while 4 parameter function is used for 2018 data. Clearly, we can see that there is little difference between the two function family. Fig 6.41 proves this statement. the two fitting functions are overlayed and the ratio between them is calculated and shown in the box below. Despite in high m_{jj}, the ratio goes off of 1, which is due to the lower statistics.

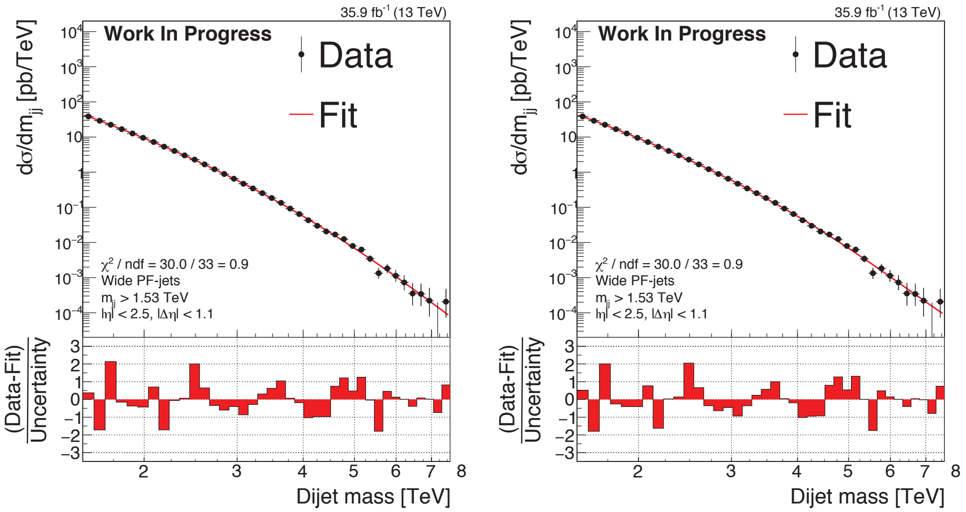


Figure 6.38: Fit to the dijet invariant mass distribution for 2016 data using 5 parameter CMS function (left) and 5 parameter ATLAS function (left)

We define the following way to measure the bias:

$$Bias = (\hat{\mu} - \mu)/\sigma_{\mu} \tag{6.10}$$

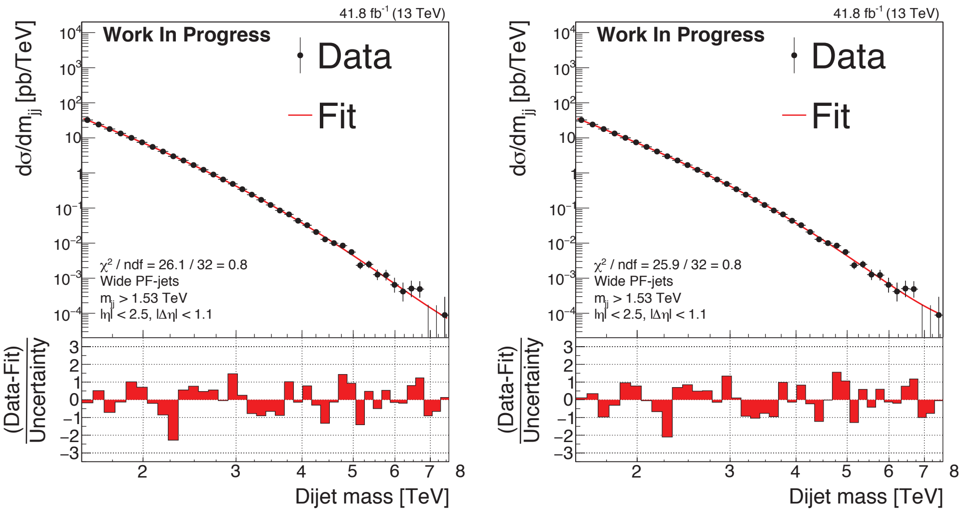


Figure 6.39: Fit to the dijet invariant mass distribution for 2017 data using 5 parameter CMS function (left) and 5 parameter ATLAS function (left)

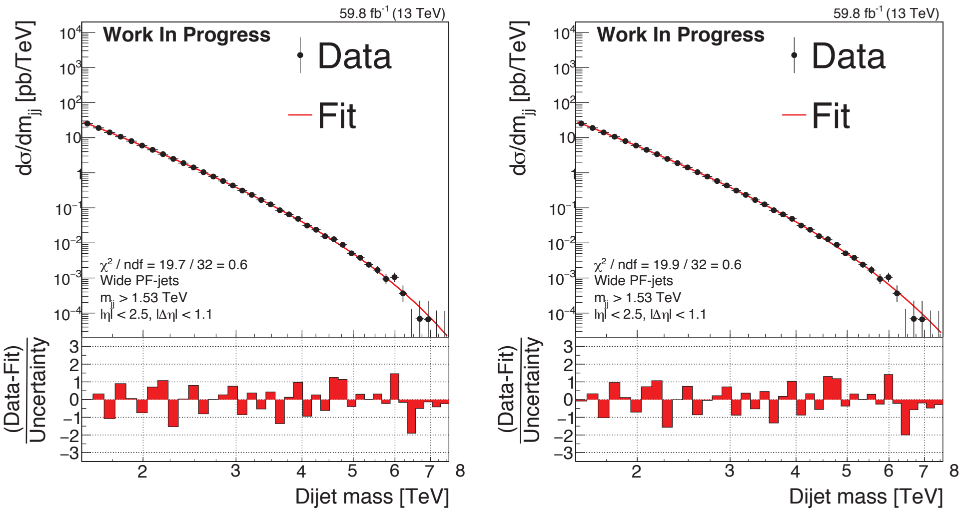


Figure 6.40: Fit to the dijet invariant mass distribution for 2018 data using 5 parameter CMS function (left) and 5 parameter ATLAS function (left)

All of the signal templates are normalized to 0.1 so that $\mu = 1$ corresponding to $\sigma \times B \times A = 0.1$ pb.The bias is checked under no injected signal, 2σ and 5σ, which

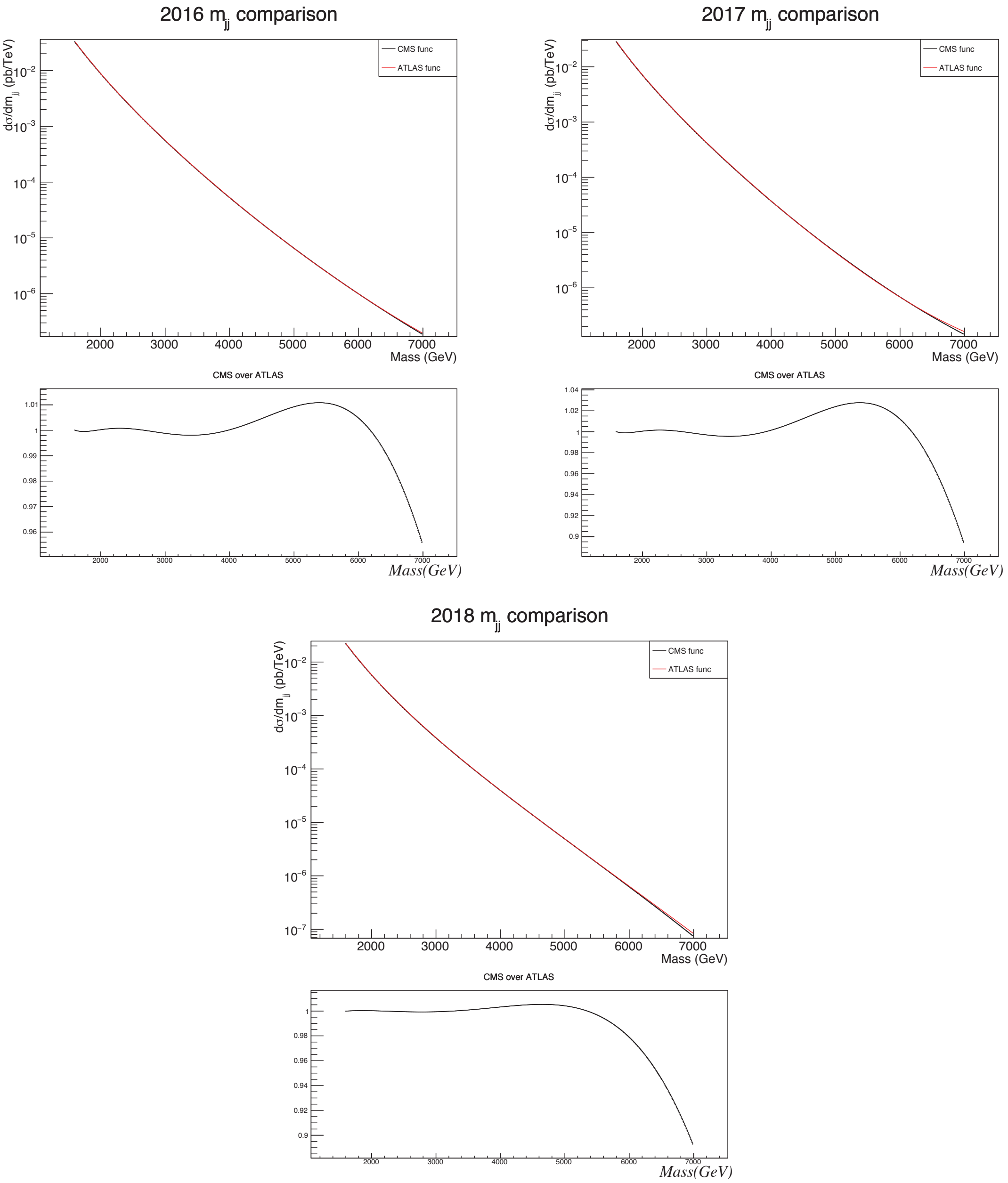


Figure 6.41: CMS fit function (black) compare with same parameter ATLAS fit function, with ratio shown in the bottom box.

σ equals to the 95% CL expected limit. In fig 6.42, the distribution of bias when generating toy datasets with the 5-parameter CMS dijet function, and then fitting them with the 5-parameter ATLAS dijet function. The error distributions are all nice following Gaussian shape, proving no significant bias need to be worried about. Fig 6.43 show the bias as a function of bg mass for all injected signal scenario. We observed up to around 50% of the bias as a percentage of σ_μ, another sanity check with the generating with 5 parameter ATLAS function, and fitting with 5 parameter CMS function is shown on the bottom. The two plots are mirrored.

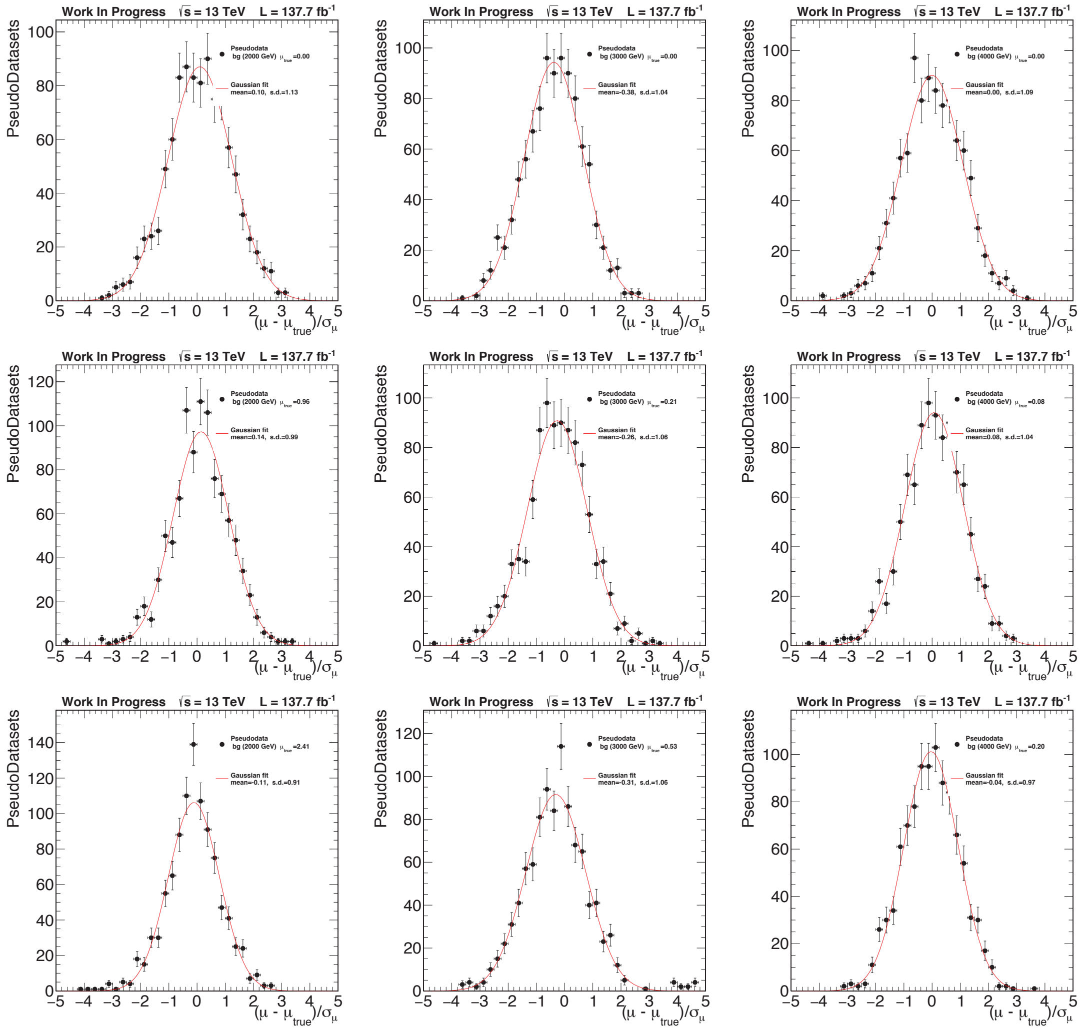
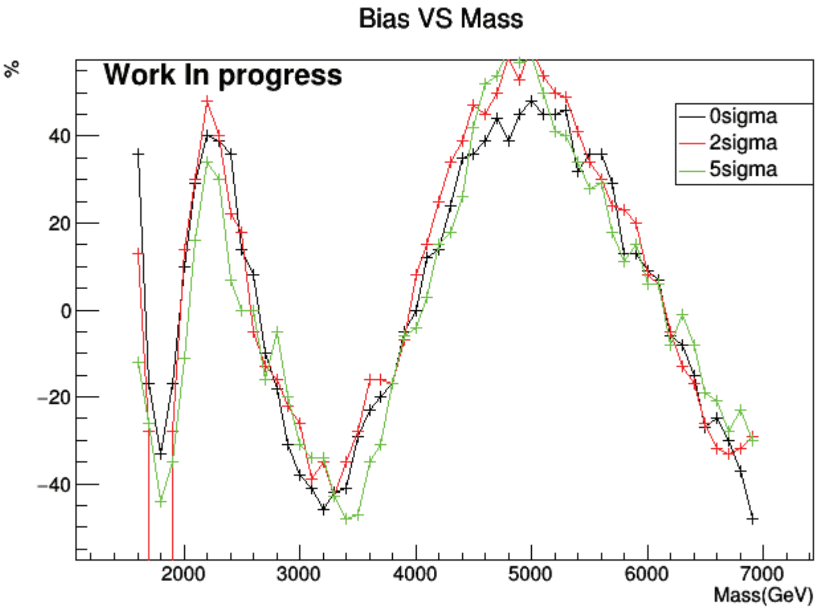


Figure 6.42: Distribution of the signal strength pulls with no injected signal (top), 2 sigma (middle) and 5 sigma (bottom), defined as the difference between measured and generated signal strength Δr divided by the uncertainty σ_r, using 1000 generated toys for Z' masses of 2 (left), 3 (central), and 4 (right) TeV along with a gaussian fit (red line),



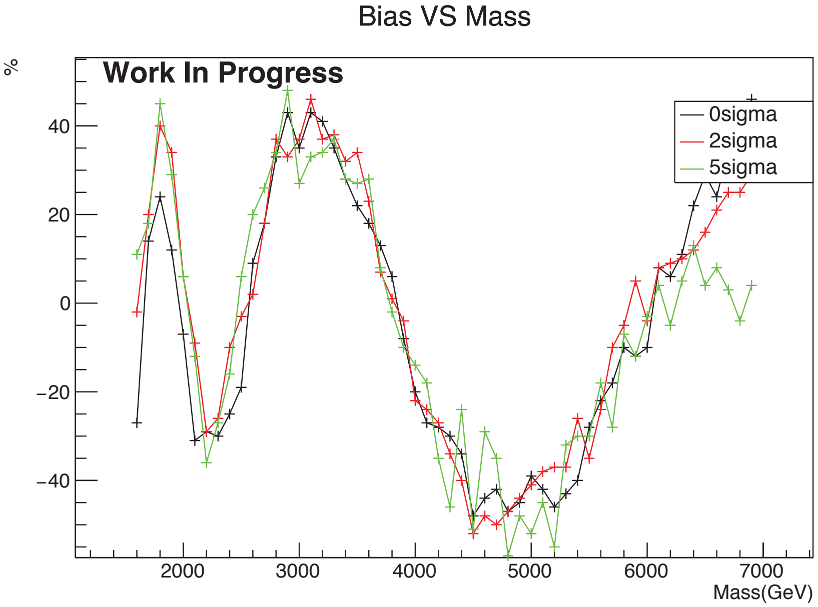


Figure 6.43: Bias as a function of qg resonance mass when fitting with the 5-parameter ATLAS function and generating with the 5-parameter function (top), or fitting with the 5-parameter dijet function and generating with the 5-parameter ATLAS function (bottom)

CHAPTER 7
CONCLUSION

In the first part of the dissertation, the 2016 collision data is used for qq resonance search and DM interpretation. We scan the mediator mass between 500 GeV and 4 TeV in 0.5 TeV steps. In the critical mass range between 1800 GeV and 2800 GeV, we use 50 GeV steps. Fig. 5.12 shows the excluded values of mediator mass as a function of m_{DM} for both types of mediators, and Fig 5.13 compares them with the exclusions in the mono-X channels from MET + X searches. The 90% CL limits in Fig. 5.17 are then translated into the plane of the DM mass versus the DM-nucleon interaction cross-section, shown in Fig. 5.18.

In the second part of the dissertation, the whole RunII collision data is used, assist with the DeepJet b-tagger. In the absence of excesses, a 95% CL limit on the b^* model is set by looking at the category where at least one of the two leading jets is b-tagged. In this case, upper limits are expressed in terms of $\sigma \times BR \times A$ where A is the acceptance accounting for the kinematic requirements $|\Delta\eta_{jj}| < 1.1$ and $|\eta| < 2.5$. The expected limit was excluded between 1660 and 2100 GeV, the observed limit was excluded bellow 1900 GeV, and between 2175 and 2351 GeV. In addition, the interpretation of leptophobic vector and axial-vector dark matter mediator was set, with coupling $\mathrm{g}_q = 0.25$ and $\mathrm{g}_{DM} = 1.0$. The 95% CL upper limits on the universal quark coupling g'_q as a function of resonance mass for a leptophobic Z' resonance that only couples to quarks are reported.

BIBLIOGRAPHY

[1] What Is The Standard Model of Particle Physics? https://www.sciencealert.com/the-standard-model.

[2] Baur U, Hinchliffe I, Zeppenfeld D. 1987 Excited Quark Production at Hadron Colliders. *Int. J. Mod. Phys. A* (doi:10.1142/S0217751X87000661).

[3] Facts and figures about the LHC. https://home.cern/resources/faqs/facts-and-figures-about-lhc.

[4] High vacuum, Taking a closer look at LHC. https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.high_vacuum.

[5] CMS Luminosity - Public Results. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[6] Elashri M. 2017 Strip hit resolution of CMS Tracker analysis. *10.13140/RG.2.2.11136.84480* .

[7] CMS Collaboration CMS Technical Design Report for the Pixel Detector Upgrade. Tech. Rep. CERN-LHCC-2012-016. CMS-TDR-11, FERMILAB-DESIGN-2012-02.

[8] CMS Collaboration Mechanical stability of the CMS strip tracker measured with a laser alignment system. *arXiv:1701.02022* (doi:10.1088/1748-0221/12/04/P04023).

[9] CMS Collaboration Performance and Operation of the CMS Electromagnetic Calorimeter. *arXiv:0910.3423* (doi:10.1088/1748-0221/5/03/T03010).

[10] 2012 CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter , CERN-LHCC-2012-015, CMS-TDR-010. (doi:10.2172/1151651).

[11] Sirunyan AM, *et al.* 2017 Particle-flow reconstruction and global event description with the CMS detector. *JINST* , CMS-PRF-14-001. CMS-PRF-14-001-004. 10. (doi:10.1088/1748-0221/12/10/P10003).

[12] Bagger J, Schmidt C, King S. 1988 Axigluon production in hadronic collisions. *Phys. Rev. D* (doi:10.1103/PhysRevD.37.1188).

[13] Chivukula RS, Cohen AG, Simmons EH. 1996 New Strong Interactions at the Tevatron? *Phys. Lett. B* (doi:10.1016/0370-2693(96)00464-9).

[14] Hewett JL, Rizzo TG. 1989 Low-Energy Phenomenology of Superstring Inspired E(6) Models. *Phys. Rept.* (doi:10.1016/0370-1573(89)90071-9).

[15] Eichten E, Hinchliffe I, Lane K, Quigg C. 1984 Supercollider physics. *Rev. Mod. Phys.* (doi:10.1103/RevModPhys.56.579).

[16] Recommendations on presenting LHC searches for missing transverse energy signals using simplified s-channel models of dark matter. (doi:https://doi.org/10.1016/j.dark.2019.100365).

[17] Abdallah J, *et al.* 2015 Simplified models for dark matter searches at the LHC. *Phys. Dark Univ.* (doi:10.1016/j.dark.2015.08.001).

[18] T du Pree, K Hahn, P Harris, and C Roskas. 2016 Cosmological constraints on Dark Matter models for collider searches .

[19] Behnke E, *et al.* 2017 Final Results of the PICASSO Dark Matter Search Experiment. *Astropart. Phys.* (doi:10.1016/j.astropartphys.2017.02.005).

[20] Choi K, *et al.* 2015 Search for Neutrinos from Annihilation of Captured Low-Mass Dark Matter Particles in the Sun by Super-Kamiokande. *Phys. Rev. Lett.* (doi:10.1103/PhysRevLett.114.141301).

[21] Aartsen MG, *et al.* 2017 Search for annihilating dark matter in the Sun with 3 years of IceCube data. *Eur. Phys. J.* (doi:10.1140/epjc/s10052-017-4689-9).

[22] Aartsen MG, *et al.* 2016 Improved limits on dark matter annihilation in the Sun with the 79-string IceCube detector and implications for supersymmetry. *JCAP* (doi:10.1088/1475-7516/2016/04/022).

[23] Amole C, *et al.* 2017 Dark Matter Search Results from the PICO-60 C_3F_8 Bubble Chamber.

[24] Akerib D, Alsum S, Araujo H, Bai X, Bailey A, et al. 2017 Results from a Search for Dark Matter in the Complete LUX Exposure. *Physical Review Letters* , 2. (doi:10.1103/physrevlett.118.021303).

[25] Tan A, *et al.* 2016 Dark Matter Results from First 98.7 Days of Data from the PandaX-II Experiment. *Phys. Rev. Lett.* (doi:10.1103/PhysRevLett.117.121303).

[26] Agnese R, *et al.* 2016 New Results from the Search for Low-Mass Weakly Interacting Massive Particles with the CDMS Low Ionization Threshold Experiment. *Phys. Rev. Lett.* (doi:10.1103/PhysRevLett.116.071301).

[27] Angloher G, *et al.* 2016 Results on light dark matter particles with a low-threshold CRESST-II detector. *Eur. Phys. J. C* (doi:10.1140/epjc/s10052-016-3877-3).

[28] Thomson JJ. 1897 Cathode rays. *Phil. Mag. Ser. 5* (doi:10.1080/14786449708621070).

[29] CMS Collaboration Global conservation. *Phys. Lett. B 716 (2012) 30* (doi:10.1016/j.physletb.2012.08.021).

[30] ATLAS Collaboration Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys.Lett. B716 (2012) 1-29* (doi:10.1016/j.physletb.2012.08.020).

[31] Right Hon LK, *et al.* I. Nineteenth century clouds over the dynamical theory of heat and light. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* .

[32] Feng JL. 2010 Dark Matter Candidates from Particle Physics and Methods of Detection. *Annu. Rev. Astron. Astrophys.* (doi:10.1146/annurev-astro-082708-101659).

[33] B Gianfranco HD, Joseph S. 2005 Particle dark matter: evidence, candidates and constraints. *Physics Reports* (doi:10.1016/j.physrep.2004.08.031).

[34] Porter, Troy A and Johnson, Robert P and Graham, Peter W. 2011 Dark Matter Searches with Astroparticle Data. *Annu. Rev. Astron. Astrophys.* (doi:10.1146/annurev-astro-081710-102528).

[35] Yang CN, Mills RL. 1954 Conservation of Isotopic Spin and Isotopic Gauge Invariance. *Physical Review* (doi:10.1103/PhysRev.96.191).

[36] Weinberg S. 1967 A Model of Leptons. *Phys. Rev. Lett.* (doi:10.1103/PhysRevLett.19.1264).

[37] Englert F, Brout R. 1964 Broken Symmetry and the Mass of Gauge Vector Mesons (doi:10.1103/PhysRevLett.13.321).

[38] Higgs PW. 1964 Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.* **13**. (doi:10.1103/PhysRevLett.13.508).

[39] Guralnik G, Hagen C, Kibble T. 1964 Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.* (doi:10.1103/PhysRevLett.13.585).

[40] Higgs PW. 1964 Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.* (doi:10.1103/PhysRevLett.13.508).

[41] Higgs PW. 1964 Broken symmetries, massless particles and gauge fields. *Phys. Lett.* (doi:10.1016/0031-9163(64)91136-9).

[42] Englert F, Brout R. 1964 Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.* (doi:10.1103/PhysRevLett.13.321).

[43] Cabibbo N, Maiani L, Srivastava Y. 1984 Anomalous Z Decays: Excited Leptons. *Phys. Lett. B* (doi:10.1016/0370-2693(84)91850-1).

[44] De Rujula A, Maiani L, Petronzio R. 1984 Search for Excited Quarks. *Phys. Lett. B* (doi:10.1016/0370-2693(84)90930-4).

[45] Kuhn JH, Zerwas PM. 1984 Excited Quarks and Leptons. *Phys. Lett. B* (doi:10.1016/0370-2693(84)90618-X).

[46] Abe F, *et al.* 1996 Inclusive jet cross section in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. Lett.* (doi:10.1103/PhysRevLett.77.438).

[47] Chiappetta P, Layssac J, Renard F, Verzegnassi C. 1996 Hadrophilic Z-prime: A Bridge from LEP-1, SLC and CDF to LEP-2 anomalies. *Phys. Rev. D* (doi:10.1103/PhysRevD.54.789).

[48] Altarelli G, Bartolomeo ND, Feruglio F, Gatto R, Mangano ML. 1996 Rb, Rc and jet distributions at the Tevatron in a model with an extra vector boson. *Phys. Lett. B* (doi:10.1016/0370-2693(96)00237-7).

[49] Hill CT, Parke SJ Top quark production: Sensitivity to new physics. *Phys. Rev. D* (doi:10.1103/PhysRevD.49.4454).

[50] Sirunyan AM, *et al.* 2020 Search for high mass dijet resonances with a new background prediction method in proton-proton collisions at $\sqrt{s} = 13$ TeV (doi:10.1007/JHEP05(2020)033).

[51] D'Ambrosio G, Giudice G, Isidori G, Strumia A. 2002 Minimal flavour violation: an effective field theory approach. *Nuclear Physics B* (doi:10.1016/S0550-3213(02)00836-2).

[52] Large Hadron Collider. https://en.wikipedia.org/wiki/Large_Hadron_Collider.

[53] CMS Detector. https://home.cern/science/experiments/cms#:~:text=The%20Compact%20Muon%20Solenoid%20(CMS,could%20make%20up%20dark%20matter.

[54] Herr W, Muratori B. 2006 Concept of luminosity. *CERN Document* (doi:10.5170/CERN-2006-002.361).

[55] Saha A. 2017 Phase 1 upgrade of the CMS pixel detector. *J. Instrum.* (doi:10.1088/1748-0221/12/02/c02033).

[56] Gratta G, Newman H, and Zhu RY Crystal Calorimeters in Particle Physics. *Annual Review of Nuclear and Particle Science 44 (1994), no. 1, 453-500. doi:10.1146/annurev.ns.44.120194.002321.* (doi:10.1088/1748-0221/5/03/T03010).

[57] Raymond M and et al The MGPA Electromagnetic Readout Chip for CMS. *CERN/LHCC 2003-055 (2003)* (doi:10.1088/1748-0221/5/03/T03010).

[58] Paganini P CMS Electromagnetic Trigger commissioning and first operation experiences. *J. Phys. Conf. Ser. 160 (2009) 012062.* (doi:10.1088/1748-0221/5/03/T03010).

[59] Alwall J, Frederix R, Frixione S, *et al.* 2014 The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP* (doi:10.1007/JHEP07(2014)079).

[60] Sjostrand T, Ask S, Christiansen, *et al.* 2015 An introduction to PYTHIA 8.2. *Comput. Phys. Commun.* (doi:10.1016/j.cpc.2015.01.024).

[61] Christensen ND, Duhr C, Fuks B, Reuter J, Speckner C. 2012 Introducing an interface between FeynRules and WHIZARD. *Eur. Phys. J. C* (doi:10.1140/epjc/s10052-012-1990-5).

[62] Christensen N, de Aquino P, Degrande C, Duhr C, Fuks B, Herquet M, Maltoni F, Schumann S. 2011 A comprehensive approach to new physics simulations. *Eur. Phys. J. C* (doi:10.1140/epjc/s10052-011-1541-5).

[63] Adam A, Christensen ND, Degrande C, Duhr C, Fuks B. 2014 FeynRules 2.0-A complete toolbox for tree-level phenomenology. *Comput. Phys. Commun.* (doi:10.1016/j.cpc.2014.04.012).

[64] de Aquino P, Link W, Maltoni F, Mattelaer O, Stelzer T. 2012 ALOHA: Automatic Libraries Of Helicity Amplitudes for Feynman Diagram Computations. *Comput. Phys. Commun.* (doi:10.1016/j.cpc.2012.05.004).

[65] Sirunyan A, Tumasyan A, Adam W, Asilar E, Bergauer T, Brandstetter J, Brondolin E, Dragicevic M, Ero J, Flechl M, *et al.* 2017 Particle-flow reconstruction and global event description with the CMS detector. *J. Instrum.* (doi:10.1088/1748-0221/12/10/p10003).

[66] Wobisch M, Wengler T. 1998 Hadronization corrections to jet cross-sections in deep inelastic scattering. In: *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)*.

[67] Matteo Cacciari and Gavin P Salam and Gregory Soyez. 2008 The anti-ktjet clustering algorithm. *JHEP* (doi:10.1088/1126-6708/2008/04/063).

[68] CMS Collaboration. 2017 Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *J. Instrum.* (doi:10.1088/1748-0221/12/02/p02014).

[69] CMS Collaboration. 2013 Identification of b-quark jets with the CMS experiment. *J. Instrum.* (doi:10.1088/1748-0221/8/04/p04013).

[70] Guest D, et al Jet Flavor Classification in High-energy Physics with Deep Neural Networks. *Phys. RevCSM. D94 (2016), no. 11, 112002* (doi:10.1103/PhysRevD.94.112002).

[71] Weiser C. 2006 A Combined Secondary Vertex Based B-Tagging Algorithm in CMS. Tech. Rep. CMS-NOTE-2006-014, CERN.

[72] Waltenberger W. 2008 Adaptive Vertex Reconstruction. Tech. Rep. CMS-NOTE-2008-033, CERN, Geneva.

[73] Chollet F, *et al.* 2015. Keras. https://github.com/fchollet/keras.

[74] Cogan J, Kagan M, Strauss E, Schwarztman A. 2015 Jet-images: computer vision inspired techniques for jet tagging. *JHEP* (doi:10.1007/JHEP02(2015)118).

[75] Komiske PT, Metodiev EM, Schwartz MD. 2017 Deep learning in color: towards automated quark/gluon jet discrimination. *Phys. Rev. D 93, 094034* (doi:10.1007/JHEP01(2017)110).

[76] Baldi P, Bauer K, Eng C, Sadowski P, Whiteson D. 2016 Jet substructure classification in high-energy physics with deep neural networks. *Phys. Rev. D* (doi:10.1103/PhysRevD.93.094034).

[77] Cogan J, Kagan M, Strauss E, Schwartzman A. 2015 Jet-images: computer vision inspired techniques for jet tagging. *JHEP* (doi:10.1007/JHEP02(2015)118).

[78] Louppe G, Cho K, Becot C, Cranmer K. 2019 QCD-Aware Recursive Neural Networks for Jet Physics. *JHEP* **2019**. (doi:10.1007/jhep01(2019)057).

[79] CMS Collaboration CMS Phase 1 heavy flavour identification performance and developments .

[80] ATLAS Collaboration. 2017 Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment .

[81] Nason P. 2004 A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP* (doi:10.1088/1126-6708/2004/11/040).

[82] Sirunyan AM, Tumasyan A, Adam W, Ambrogi, The CMS Collaboration Search for narrow and broad dijet resonances in proton-proton collisions at $\sqrt{s}$ = 13 TeV and constraints on dark matter mediators and other new particles. *JHEP* (doi:10.1007/JHEP08(2018)130).

[83] CMS Collaboration. 2016 Search for narrow resonances decaying to dijets in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.* , 7. (doi:10.1103/PhysRevLett.116.071801).

[84] CMS Collaboration. 2015 Search for resonances and quantum black holes using dijet mass spectra in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Phys. Rev.* (doi:10.1103/PhysRevD.91.052009).

[85] Junk T. 1999 Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A* (doi:10.1016/S0168-9002(99)00498-2).

[86] Read AL. 2002 Presentation of search results: the CL_s technique. *J. Phys. G* (doi:10.1088/0954-3899/28/10/313).

[87] Cowan G, Cranmer K, Gross E, Vitells O. 2011 Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C* (doi:10.1140/epjc/s10052-011-1554-0).

[88] CMS and ATLAS Collaborations. 2011 Procedure for the LHC Higgs boson search combination in Summer 2011. CMS Note CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, CERN.

[89] Sirunyan AM, Tumasyan A, Adam W, Asilar E, Bergauer T, Brandstetter J, Brondolin E, Dragicevic M, Ero J, et al. 2017 Search for dark matter produced with an energetic jet or a hadronically decaying W or Z boson at $\sqrt{s}$ = 13 TeV. *JHEP.* (doi:10.1007/jhep07(2017)014).

[90] CMS Collaboration. 2011 Search for resonances in the dijet mass spectrum from 7 TeV pp collisions at CMS. *Phys. Lett. B* (doi:10.1016/j.physletb.2011.09.015).

[91] Saoulidou N Searches for dijet resonances in pp collisions at $\sqrt{s} = 13$ TeV using the 2016 and 2017 datasets . Tech. Rep. AN-2017-348.

[92] Sirunyan AM. 2020 Search for high mass dijet resonances with a new background prediction method in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP* (doi:10.1007/jhep05(2020)033).

[93] Bertolini D, Harris P, Low M, Tran N. 2014 Pileup per particle identification. *JHEP* (doi:10.1007/jhep10(2014)059).

[94] CMS Collaboration. 2018 Performance of the DeepJet b tagging algorithm using 41.9 fb^{-1} of data from proton-proton collisions at 13 TeV with Phase 1 CMS detector , CMS-DP-2018-058.

[95] Sirunyan A, Tumasyan A, Adam W, Ambrogi F, et al. 2016 Search for massive resonances decaying into WW, WZ, ZZ, qW and qZ in the dijet final state at $\sqrt{s} = 13$ TeV using 2016 data. Tech. Rep. CMS-PAS-B2G-16-021, CERN.

[96] Hand DJ. 2012 Statistical Concepts: A Second Course, Fourth Edition by Richard G. Lomax, Debbie L. Hahs-Vaughn. *International Statistical Review* .

[97] Sirunyan A, Tumasyan A, Adam W, Asilar E, Bergauer T, et al. 2017 Search for dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV and

constraints on dark matter and other models. *Phys. Lett. B* (doi:10.1016/j.physletb.2017.02.012).

[98] Sjöstrand T, Mrenna S, Skands P. 2008 A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.* (doi:10.1016/j.cpc.2008.01.036).

[99] Jec Energy Correction twiki. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECDataMC.

[100] Methods to apply b-tagging efficiency scale factors. https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods.

[101] Boveia A, Buchmueller, *et al.* 2016 Recommendations on presenting LHC searches for missing transverse energy signals using simplified *s*-channel models of dark matter. *Phys. Dark Univ.* , CERN-LPCC-2016-001. CERN-LPCC-2016-001. (doi:10.1016/j.dark.2019.100365).

[102] Harris R, Kousouris K. 2011 Searches for dijet resonances at hadron colliders. *Int. J. Mod. Phys. A* 10.1142/S0217751X11054905.

[103] The dramatic rise of the term deep learning in research. https://theatlas.com/charts/ByhdcCsp7.

[104] Pumplin J, Stump D, Huston J, Lai H, Nadolsky PM, Tung W. 2002 New generation of parton distributions with uncertainties from global QCD analysis. *JHEP* (doi:10.1088/1126-6708/2002/07/012).

[105] Glashow SL. 1961 Partial-symmetries of weak interactions. *Nuclear Physics* (doi:10.1016/0029-5582(61)90469-2).

[106] Strologas J, CMS Collaboration. 2016 Performance of Jet reconstruction in CMS at 13 TeV. *PoS* (doi:10.22323/1.282.0736).