Search for Dark Matter in Dijet Events using Quark/Gluon Tagging

Neža Ribarič
Supervising Professor: Iain Bertram

September 2019

Abstract

This thesis explores the possibility of improving the limits set on the Dark Matter mediator Z' obtained from the dijet invariant mass spectrum using a technique called quark/gluon tagging. The main discriminator between quark and gluon initiated jets is the number of tracks within the jet, however both linear and logarithmic functions of jet p_T and invariant mass with various starting parameters were tested. The new limits obtained from the sub-samples were compared to the un-tagged dijet data sample generated to mimic 2015/2016 data of 37fb^{-1} collected with the ATLAS detector at the LHC. The samples are generated using Pythia 8 with the leading-order NNPDF 2.3 PDFs and A14 tune and the detector response is modeled in Geant 4. It was shown that for an ideal separation (based on parton truth ID), the improvement in significance can range up to 50%. However, the best separation functions based on jet p_T and m_{jj} show improvements in the range of 2-5%. The dependence on m_{jj} shows potential for significant improvements in the mass range lower than the one analyzed in this thesis.

Declaration of Authorship

This thesis is a presentation of original research performed by the author. It has not been previously submitted for the award of a higher degree. The thesis would not be possible without the Quark/Gluon tagger group, dijet and di-b-jet group and exotics group from the ATLAS collaboration. The majority of the plots shown in this thesis are taken from reports of their work and are rightfully cited. My work was solely on the optimization of the Q/G tagger for the Z' signal with $g_q = 0.2$ which included the significance plots and final limit plots on the signal.

Acknowledgements

I thank my mentor, Prof. Bertram, for his courage of accepting me as his student, but mostly for the vasts amounts of knowledge I have gained. There have been situations I wished I could have just recorded the whole conversation or presentation, because even the simplest of question were answered with things I didn't even know I needed. I am extraordinarily grateful to have visited Geneva and learned all the ATLAS software while there. Besides the help in my academic growth, I deeply appreciate the trust he put in me to finish my dissertation while allowing me to be an athlete.

There are a lot of people that helped me with various things during the year at the Particle Physics department and in the office, I thank you. Either for cases where my code wasn't working or the algorithms were not working or just for the daily conversation, I appreciate it all.

As far as venting out goes, I thank my parents for listening to me when I made no sense at all and yet somehow are still so deeply proud of me for prolonging to face the real world of job hunting for another couple of years. You're the real MVPs. I love you both so much. I have to place my coach in this sector as well, because he's been like a big brother to me, showing support in all aspects of my life and adjusting the training to it.

-"Še dalje brez medalje"

Lastly, I thank Florjan and Nadja for their kind words and support in times of darkness. Sometimes loving me takes a hell of a lot of effort and I would not have made it this far if I did not have you two at my side.

Contents

Preface

Based on a variety of evidence, some of which is presented in this thesis, the majority of the Universe is composed of matter named "Dark Matter" after it's non-luminous nature. A lot of effort has been given into searching for this new type of matter with various experiments including ATLAS. Apart from the standard way of searching for the dark matter particle itself, this thesis focuses on the search for the dark matter mediator Z' allowing interactions between the Standard Model and this new Dark Sector. Such new heavy particles that couple to partons could be produced directly in proton-proton collisions and decay back to Standard Model partons, which will shower and hadronize, creating sprays of particles known as jets. The production of jet pairs in hadronic colliders primarily result from $2 \to 2$ parton scattering processes via strong interactions, described by Quantum Chromodynamics and provides the background to the majority of new physics searches. The theory predicts a smooth monotonically falling dijet invariant mass function, meaning that a new resonant state would manifest itself as a local excess near the mass of the resonance. This makes the dijet final state very appealing. However, the QCD background is still dominant although various cuts are applied to suppress it. In this thesis quark/gluon tagging is explored as a potential to improve the signal to background ratio and consequently the ability of finding physics beyond the standard model. It exploits the difference in the number of tracks associated with jets initiated by either quarks or gluons. The lower mass regions of the QCD background are dominated by gluon-gluon dijet final states, however many of the exotic signals, like the lepthophobic Z', decay to quark-quark final states, thus we expect dijet tagging to improve signal significance. In this thesis the dependence of the number of tracks on the jet p_T and dijet invariant mass is explored in order to find an optimal separation function. The functions are tested on a MC generated data sample by comparing the limits of the whole dijet data sample with the newly acquired sub-samples.

The thesis start with a brief description of the Standard Model with a focus on proton-proton collisions and formation of jets. This is followed by a few pieces of evidence for Dark Matter including a discussion of why the answer should lie in the Physics beyond the Standard Model. Chapter 3 focuses on the LHC and ATLAS detector. Finally Chapter 4 describes the data modeling and event reconstruction, followed by the analysis chapter which explores the Q/G tagger in more detail.

1 The Standard Model of Particle Physics

The Standard Model (SM) [1] is currently the best description of the fundamental particles and their interactions we have, withstanding all of the tests conducted so far at the LHC and its predecessors. The fundamental particles, which make up all of the matter in the Universe, are called fermions. The group contains all particles with half-integer spin. Fermions can be divided into two groups: *quarks and leptons*. Unlike leptons, quarks are always found in bound states (*hadrons*) either as mesons $q\bar{q}$ or baryons qqq. Fermions are the particles lying at the outermost part of the circle in Figure 1 with the quarks colored orange and the leptons in green.

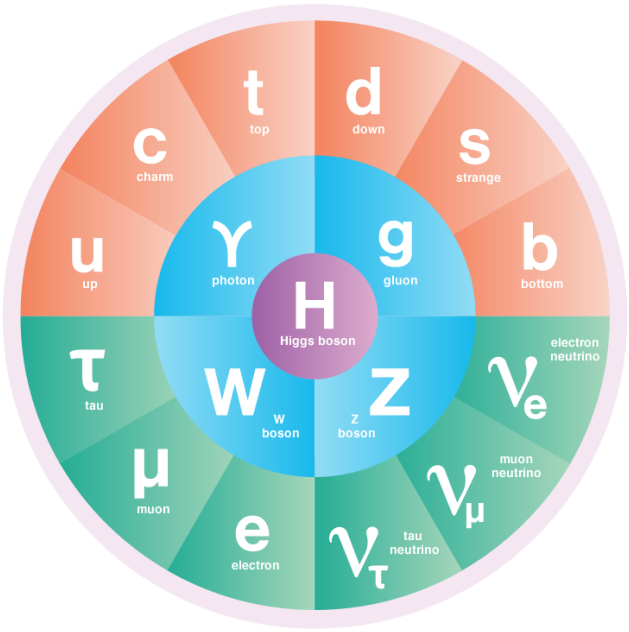


Figure 1: The Standard Model particles. Image taken from [2].

Mathematically, the SM is a gauge field theory based on the symmetry group $SU(3) \times SU(2) \times U(1)$, containing 12 generators with a nontrivial commutator algebra, making it "non-Abelian". Associated with each of those generators is a vector (*spin 1*) boson with the same quantum numbers as the generator. Each of the SM fundamental forces acting on a particle can be described as an exchange of one (or more) of these bosons. For example the interaction between two electrically charged particles in Quantum Electrodynamics is mediated by an exchange of a photon emitted from one particle and reabsorbed by the other. There are three fundamental forces in the SM: electromagnetism, the weak and strong nuclear forces, but it does not yet contain the fourth fundamental force; gravity. Each of these forces "couples" to a certain charge, for example the electromagnetic force only "sees" particles that are electrically charged. The force carrier of electromagnetism is the photon and carriers of the weak force are the $W^+, W^-$and Z^0 bosons. SU(3) is the group describing the strong interactions with the remaining eight bosons called gluons associated with "color" charge. The bosons are colored blue in Figure 1.

Out of all of the bosons, only gluons and photons are massless and the others acquire mass

through the breaking of the gauge symmetry. The symmetry breaking in the electroweak theory is of a particular type called spontaneous symmetry breaking where from the continuum of degenerate vacuum (minimal energy) states, the system is found in one particular vacuum state resulting in the violation of the symmetry of the states spectrum. In the SM this spontaneous symmetry breaking is realized by the Brout-Englert-Higgs mechanism, requiring an additional boson with zero spin, which is colored purple in Figure 1. The Higgs boson was discovered in 2012 by the ATLAS and CMS detectors at the LHC [3].

1.1 Quantum Chromodynamics

QCD, the theory of strong interactions is a SU(3) unbroken gauge theory with massless mediators, the gluons, coupling to three "color" charges. The color charge is analogous to charge in electromagnetic interactions, with a noticeable difference: there are only two electric charges, positive and negative, but there are three color charges (red, green and blue). This choice is unique, because of the following constraints:

1. It must be able to distinguish a quark from and antiquark, because there are meson states of q and $\overline{q}$ but not two quarks [4].

2. The group must produce a completely anti-symmetric color singlet baryon state made from three quarks in order to obey the Fermi exclusion principle

3. The choice must agree with processes that directly measure the number of color charges, for example in e^+e^- annihilation, the rate of production of hadrons is directly proportional to N_c and the data indicates the value 3 [5].

Consider a simple interaction, where a quark radiates a gluon, the quark's color (but not flavor) may change. As an example, a blue up-quark converts into a red up-quark. Since color charge must be conserved, the difference must be carried by the radiated gluon. Gluons, unlike photons in QED, carry color charge, specifically they are "bicolored". The radiated gluon in the example must carry a unit of blueness and a unit of anti-redness. This gives QCD a very simple structure, but a very rich and dynamical content, because apart from the fundamental quark-gluon vertex, there are also gluon-gluon vertices. Because the number of colors is 3, one would expect to find 9 gluons. Since, we do not observe a color singlet state, there are in fact only eight gluons [6].

1.1.1 Asymptotic Freedom

Two of the most important properties of QCD are asymptotic freedom and confinement. Both result from the strength of the force (α_s) being dependent on the square of the four-momenta transfer Q^2 among the particles involved in the interaction. The coupling constant α_s is not constant, but rather runs with Q as can be seen in Figure 2.

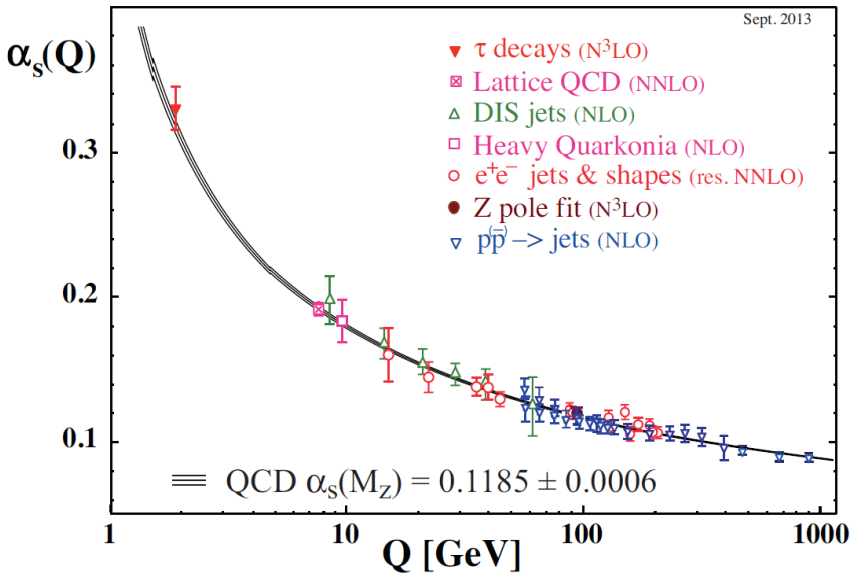


Figure 2: The strong coupling constant as a function of the momentum transfer in the interaction Q, with experimentally obtained data plotted with colored points and the fit shown with black lines. Figure is taken from [7].

The fact that the coupling parameter decreases for increasing Q^2 and vanishes asymptotically is called asymptotic freedom. Starting with deep inelastic scattering (DIS) experiments, $ep \longrightarrow e+X$, where X is a hadronic state composed of multiple hadrons, electrons scattered off the proton at low energies. However, with increasing energy the electron does not scatter of the proton as a whole, but off individual charged constituents inside it: this is the QCD parton model[11, 12]. The deep-inelastic cross section becomes identical to the scattering of an electron on point-like particles, carrying a fraction x of the proton's momentum, multiplied by the probability of finding that point-like particle $f(x, Q^2)$. The hadronic cross section, σ, in a pp collider can now be computed as a sum over all parton-parton cross sections containing the same final state, weighted by the probability of finding each parton in the proton carrying a given momentum fraction. For example, partons i and j in the two protons interact to create particles c and d in the final state:

$$\sigma(p + p \to c + d + X) = \Sigma_{ij} \int dx_1 dx_2 f_i(x_1) f_j(x_2) \sigma(i + j \to c + d) \tag{1}$$

where the "X" is all additional by products produced besides c and d. The relationship between the proton and parton momentum is embodied in the parton distribution functions (PDFs), f_i, which are non-perturbative and need to be extracted from hard-scatter experiments. The functions are depicted in Figure 3 for two values of Q^2.

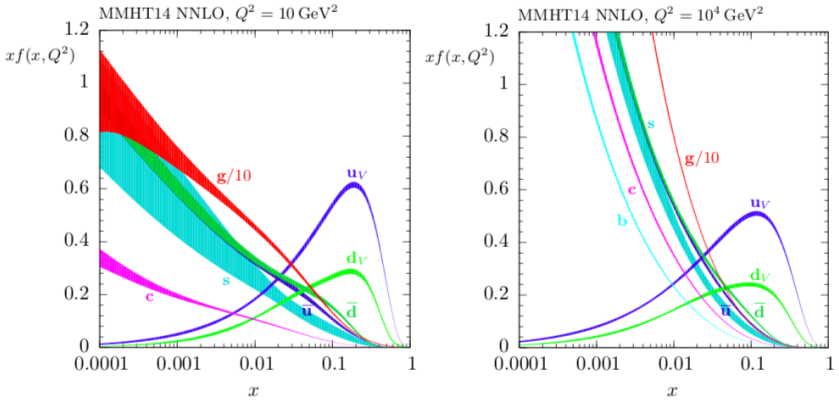


Figure 3: Next-to-next-to-leading order PDFs at $Q^2 = 10$ GeV^2 and $Q^2 = 104$ GeV^2, with associated 68% confidence-level uncertainty bands. Figures taken from [14].

Measurements of these functions show that the sea of gluons and quark-antiquark pairs contains all quark flavors with masses smaller than the four-momentum transfer in the collision. Gluons are continuously splitting and recombining into quark-antiquark pairs and both of the partons can radiate other partons. With increasing four-momentum transfer, more gluons are radiated from the original quark which can then split into quark-antiquark pairs, resulting in decreasing the initial quark momentum fraction and the growth of gluon density and the quark-antiquark sea. These splittings depend on so called ("Casimir") color-factors: $C_F = (N_c^2 - 1)/(2N_c)$, $C_A = N_c = 3$ and $T_R = 1/2$ and are associated with the probabilities of a gluon emission from a quark, gluon emission from a gluon and a gluon splitting into a $q\overline{q}$ pair. The equations describing the evolution of a parton distribution due to gluon radiation and splittings are known as the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [8].

1.1.2 Confinement

Contrary to Asymptotic Freedom, the interaction strength becomes large for small momentum transfers resulting in hadrons being tightly bound states of quarks with no overall color charge. In an effort to separate a meson, the interaction energy grows as the distance between the pair until it is energetically more efficient to create a $q\overline{q}$ pair from the vacuum. The large centre-of-mass energy in a proton-proton (pp) collider results in the final state partons having high energies and in turn moving apart very fast, creating new pairs between them. This results in observing back-to-back jets (sprays of particles) of colorless hadrons. The partons created in these splittings are typically produced at very small angles to the initial parton and the resulting shower is highly collimated in the direction of the initial parton. After the showering process reaches lower energies the hadronisation process begins, which is non-perturbative and described using models. The schematic view of this process is depicted in Figure 4.

1.1.3 Jet formation

In a pp collider, the large centre-of-energy results in a high Q interaction, a hard scattering process, between two partons. In such a scattering, a short lived resonance could be created or more likely, the standard QCD process, such as a quark-quark scattering, occurs. In either case, the scattering can result in a production of partons in the final state. Partons not involved in the hard scattering could also interact creating an Underlying Event, which is depicted in red in Figure 4. In addition to the hard scattering process, both the initial and final state partons can radiate gluons via the process described above. This is called inital state radiation (ISR) and final state radiation (FSR) respectively. After parton showering reaches lower energies, hadronzation occurs which results in collimated groups of hadrons that we label as a jet. Understanding the jet formation is essential in modeling the QCD background, since it is the dominant background in the majority of searches at the LHC. However a precise definition of a jet is needed for any quantitative analysis.

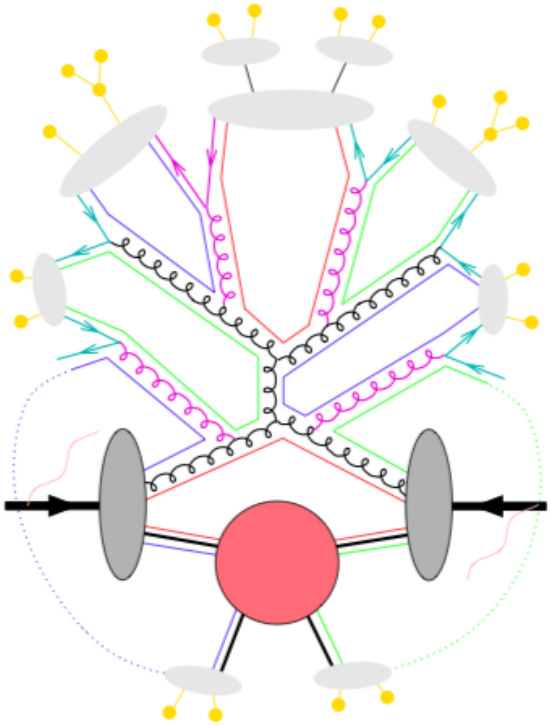


Figure 4: Visual schematic of the structure of a pp-collision taken from [15]. The incoming protons are represented with the solid black lines indicating their momenta is opposite. At collision, the partons not involved in the hard scattering interact in the underlying event which is depicted as a red circle. Shower evolvement is depicted with multiple colors representing gluon radiation or splitting into $q\overline{q}$ pairs. Hadronisation is depicted with gray and the subsequent decay of the hadrons is depicted with yellow.

1.2 Dijet Resonance Search and Motivation

There are several reasons why the dijet final state is an appealing final state for searches at the Large Hadron Collider (LHC). Firsty, the LHC is a hadron collider meaning that any new particles produced in the collisions could also decay back to quarks, producing jets in the final state. From the predictions of QCD, the invariant mass spectrum is a smoothly falling spectrum, where new particles would manifest themselves as a local excess over the SM background. The background can be easily modeled through data-driven techniques and can provide strong limits in the absence of any new physics. The very appealing aspect of dijet searches is the increase of sensitivity to small cross-section with the rise of data sizes as well as centre-of-mass energies. This can be seen in Figure 5, which compares the limits obtained from different experiments varying in centre-of-mass energies and luminosity. This highly motivates future dijet analysis as the LHC almost doubled the centre-of-mass energy to 13 TeV in 2015 and is now working on increasing the centre-of-mass energy to 14 TeV then improving the luminosity by a factor of 10 by 2025. This also presents an excellent opportunity for further improving the limits with a Q/G tagger, since the QCD background is now only constrained with an angular cut of the final state jets and lower limit to their invariant mass.

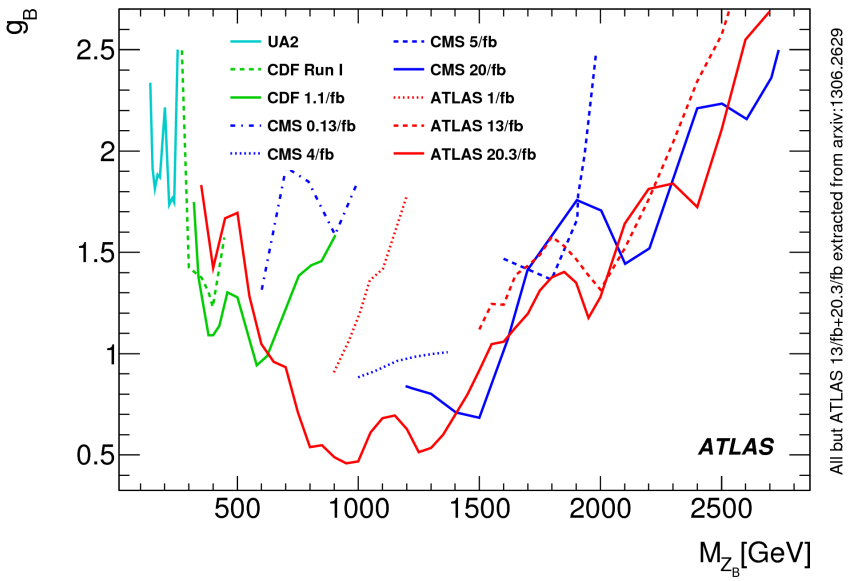


Figure 5: Comparison plot of limits obtained from various experiments differing in data size and centre-of-mass energy. The limits are plotted in the plane of quark coupling g_B versus the mass of the dark matter mediator Z_B. Figure taken from [16]

2 Dark Matter: Evidence and Theory

2.1 Evidence

Our observation of the night sky was through a sole medium: light. However, it soon became apparent that some gravitational effects can not be accounted for solely with observed luminous matter.

2.1.1 Rotational velocities of galaxies

The most well known evidence for DM is Vera Rubin's observation of the radial velocities of spiral galaxies [18, 19]. If one assumes the spherical distribution of matter in the galaxy and applies the gravitational analogue to Gauss' Law, we get a simple relation between the radial velocity of objects in the galaxy and the distance to the galactic centre. Towards the outskirts of the galaxy, the mass does not increase anymore and we expect the velocity to decrease as $v_r \propto \sqrt[2]{r}$. Because stars of different types emit light at different wavelengths, the mass of the galaxy can be calculated by measuring the light emission of the galaxy at that wavelength by simply inferring how many stars of that type are needed to produce such light emission. Using information previously determined about the masses of different types of stars, the mass contribution of stars to the galaxy is obtained [20]. Vera Rubin's observations showed a very slow decrease of the velocities with the distance from the centre and it was later calculated by Bosma [21], van Albada and Sancisi [22] that this flatness can not be fully accounted for by just modifying the weight of individual components (disc, gas) as can be seen in the plot below.

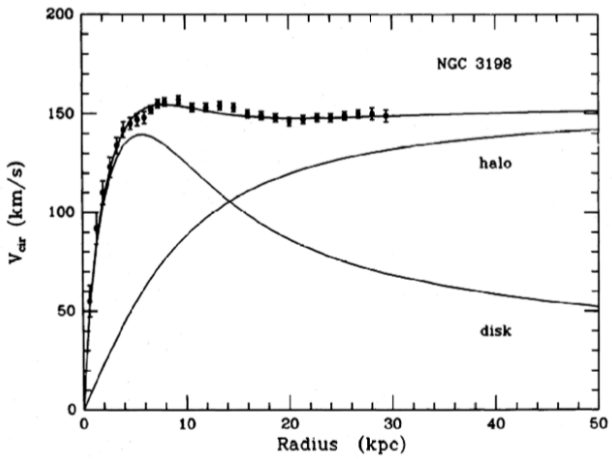


Figure 6: Rotation curve of a spiral galaxy NGC 3198, where the contribution from the luminous disc and dark matter halo is shown by solid lines. Comparison plot obtained from [22].

The flatness can be achieved by imposing a new mass component with a distribution proportional to r, which corresponds to a self-gravitational gas of non-interacting particles.

2.1.2 Light to Mass ratio

Exploring the gravitational effects on an even larger scale was Fritz Zwicky by observing the Coma cluster [23]. The kinetic energy of objects inside the cluster was determined by measuring the Doppler shifts of the galaxies in the cluster. Assuming that the galaxy cluster is an isolated system, the virial theorem can be used to relate the average velocity of objects inside the cluster and the potential energy (mass of the system) [24]. This method of determining the mass of the system is insensitive to whether the objects emit light or not. By comparing results with another luminosity dependent method Zwicky's calculations showed a large light-to-mass ratio, indicating that the majority of mass in the cluster was actually non-luminous. While Zwicky failed to account for the roughly 10% of the mass contained in the intracluster gas, there was still a tremendous amount of "missing" matter and it was henceforth named "dark" [25, 26].

2.1.3 Bullet Cluster

The Bullet Cluster consists of two clusters which underwent a collision [28]. By observing the magnification, distortion or shifting of light emitted by background objects the image in Figure 7 was obtained by combining observations from Chandra-X ray telescope and data from the Hubble Space Telescope. The mass distribution obtained through weak lensing clearly indicates that the visible matter is displaced in regards to the invisible component. The visible matter displays a characteristic shock wave, but the dark matter components crossed with almost no interaction. The Bullet Cluster is also the best argument against Modified Newtonian Dynamics (MOND) [27] theories and sets an upper limit to the self-interaction strength of DM.

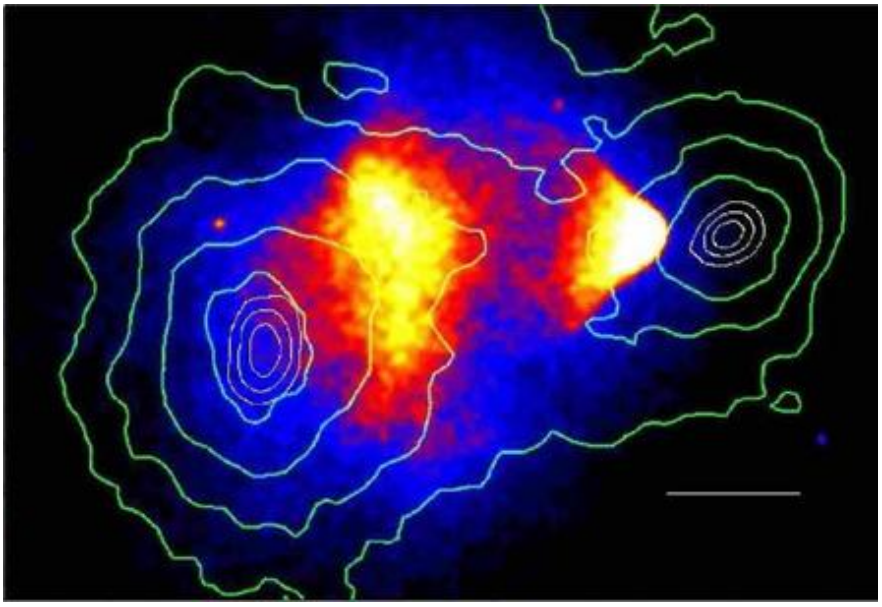


Figure 7: Deep Chandra image of the Bullet cluster with green lines representing mass contours. Image is taken from [28].

2.1.4 Cosmic Microwave Background

The Cosmic Microwave Background (CMB), discovered by Penzias and Wilson in 1964 [29], is an excess background temperature of approximately 2.73 K coming from the "last scattering" of photons right before the universe became transparent to electromagnetic interactions. Moments after the Big Bang, the universe was a dense plasma of charged particles and photons, which firstly went through a rapid expansion then the rate of expansion slowly decreased and the plasma was cooling down. At a point, known as the epoch of recombination, neutral atoms started to form and photons, previously trapped inside this plasma due to constant interactions with charged particles, were now released. COBE [30] satellite measured fluctuations in CMB that can be attributed to two effects: low energy photons being released from areas that were more dense at the time of the last scattering and acoustic oscillations. These fluctuations are incredibly small, in fact, too small to have an impact on the structure and formation of our Universe. Ordinary matter only becomes electrically neutral at the epoch of recombination, before that, the electrostatic forces are too big for the matter to clump into gravitational wells. The fluctuations in CMB measured by COBE show a need for a charge neutral form of matter that started the process of structure formation before recombination occurred. The CMB fluctuations also offer a method of calculating the density parameters of types of matter. The abundance of all matter, baryonic matter and dark matter, presented in Table 1, was precisely measured by the Planck mission providing a shocking result that our Universe is composed of only a few percent of ordinary (baryonic) matter [31].

type	density parameter value
matter	$\Omega_m h^2 = 0.315 \pm 0.007$
baryonic matter	$\Omega_b h^2 = 0.0224 \pm 0.0001$
dark matter	$\Omega_{dm} h^2 = 0.120 \pm 0.001$

Table 1: Types of matter with their corresponding density parameters obtained from [31].

.

2.2 Physics Beyond the Standard Model

Although the above observations may not give answers to what Dark Matter is, they constrain what it is not. The biggest constraint comes from the fact that it is in fact dark, meaning it doesn't scatter light. This sole constraint limits [32, 33, 34, 35] DM to not interact electromagnetically and removes the majority of SM particles from the list of viable candidates. Since all of the SM quarks are charged, this also implies that interactions via the strong force are not possible. What we are left with from the Standard Model are three weakly interacting neutrinos. Neutrinos have been considered as the most excellent candidate, because of the "undisputed virtue of being known to exist" [36]. Considering their relic density, calculated by

$$\Omega_\nu h^2 = \Sigma_i^3 \frac{m_i}{93 \text{ eV}} \tag{2}$$

with h being the Planck constant and inserting the best estimate for individual neutrino masses m_i, we arrive to the result $\Omega_\nu h^2 \lesssim 0.07$. This means that, while neutrinos could be a viable candidate, they are simply not abundant enough to be the sole contributor. By now we have fully exhausted the SM list of particles, meaning that we have to resort to physics beyond the SM to solve the DM problem.

2.2.1 Analysis specific model constraints

If non-gravitational interactions exist between DM and SM, then DM can be produced at particle accelerators, but would not interact with the detector. One way to measure the DM particle is through observations involving SM particles X ($= g, q, \gamma, Z, W$ or h). In such reactions, called "mono-X" or $E_T^{miss} + X$, the particles or jets recoil against an invisible state and the missing momenta E_T^{miss} is calculated. The telltale missing energy signature is, however, not enough to determine if such a particle is a viable DM candidate. This conclusion can only be drawn from combining LHC searches with astrophysical observations [37, 38, 39] and direct detection experiments [45, 46, 47, 48].

The rich structure of the SM and the fact that DM is almost five times more abundant indicates that the structure of DM may be more complex than a single particle. The simple signal models considered as benchmarks for the LHC Run-2 were presented in the ATLAS-CMS Dark Matter Forum [49]. Alongside the usual DM searches based on missing energy, this analysis is based on searching for *the mediator particle* itself. This method neatly takes advantage of the fact that if a DM particle can be produced from SM via this new mediator, then that mediator can also decay back into SM particles.

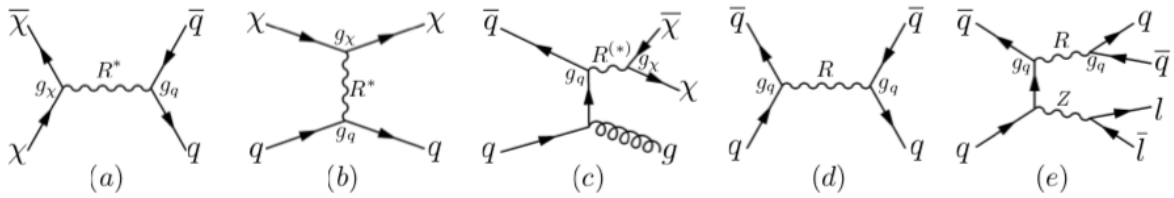


Figure 8: Processes involving visible sector quarks $q, \overline{q}$, DM particles $\chi, \overline{\chi}$ and on-shell (off-shell) mediator $R(R^*)$: a) DM annihilation, b) DM scattering, c) monojet signature, d) dijet signature, e) dijet associated production. The Feynman diagrams are taken from [39].

In the last few years several B-physics experiments have reported deviations from the SM prediction of flavor-changing transitions such as $b \longrightarrow s$. The anomalies were first observed by LHC's B-factory experiment LHCb [40] in $B_s \longrightarrow \phi\mu^+\mu^-$ [41] and $B \longrightarrow K\mu^+\mu^-$ [42], but have been confirmed by the Belle experiment [43]. The statistical significance of these anomalies, 2.2σ and 2.6σ respectively, are too low to claim a discovery, but are enough to have motivated speculations of possible mechanisms beyond the SM that could be responsible for them. One of the possible explanations is the existence of a new gauge boson (Z') that arises from a simple extension of the SM with a new U(1) gauge symmetry [50, 51]. This highly motivated the search for a new gauge boson at other experiments like ATLAS, but also gives a possible solution of the DM problem. Assuming that the DM Dirac fermion particle has charges only under this group and that some SM particles also contain this charge, the interaction between the two sectors can be mediated via the exchange of this new mediator with potentially large couplings [52, 53, 45, 54, 55, 56, 57]. Some of the interactions are depicted in Figure 8. The coupling to leptons is strongly constrained by dilepton resonance searches, which can be seen in Figure 9, so in this analysis we consider a leptophobic mediator, where decays to SM leptons give no relevant contribution. In case the interaction is vector like, the interaction is strongly constrained by direct detection experiments and searches at the LHC would give no relevant bounds, which can be seen from Figure 10, so an axial-vector coupling is assumed.

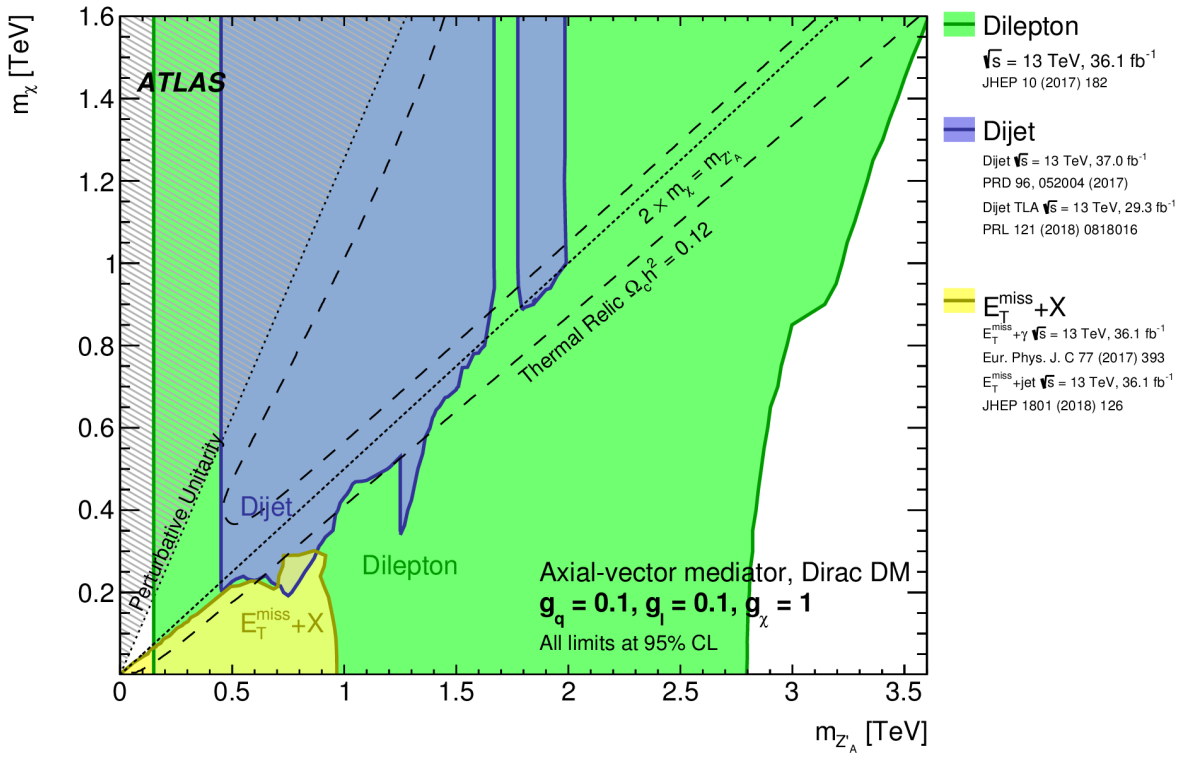


Figure 9: Regions in a (mediator-mass, DM-mass) plane excluded by different searches for a leptophilic axial-vector mediator models with $g_q = 0.1$. Plot from [44]

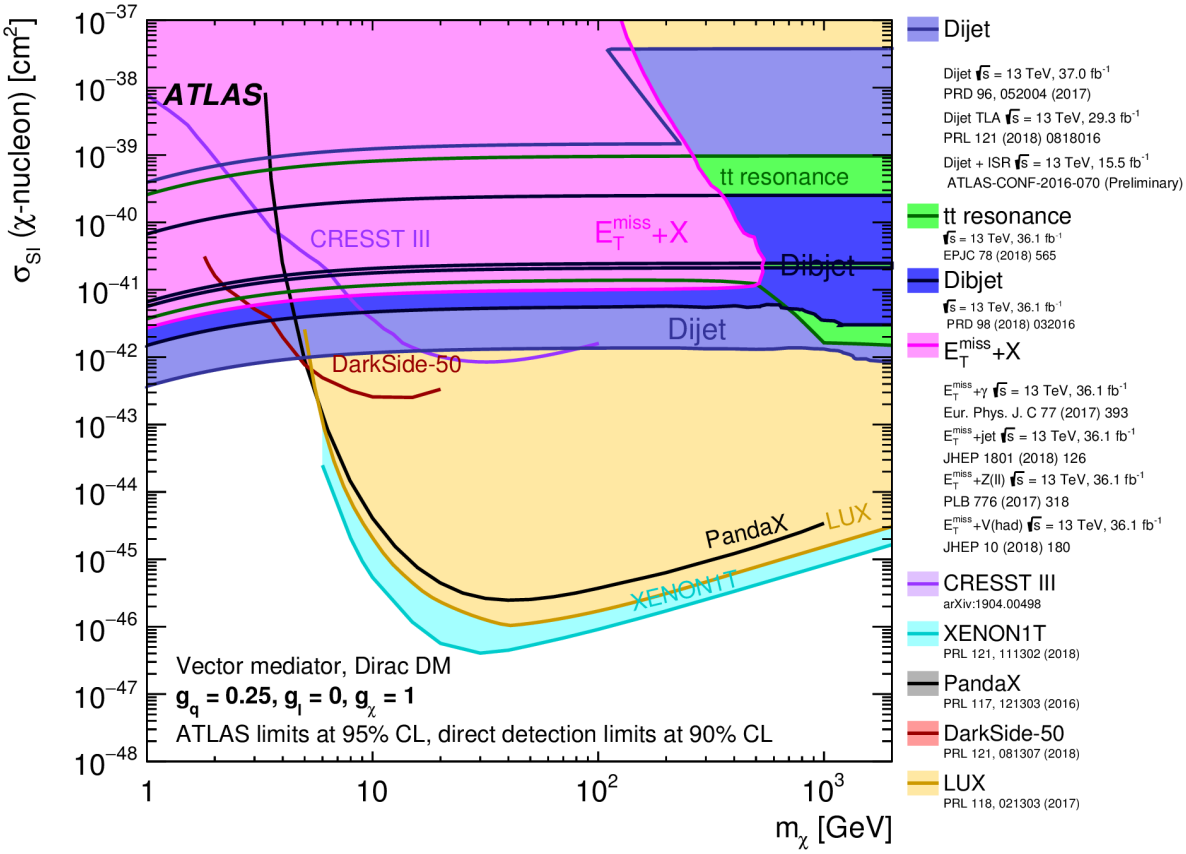


Figure 10: Plot comparing the inferred limits to constraints gained by direct detection experiments on the spin-dependent WIMP-proton scattering cross-section in the context of the vector leptophobic Z' model with $g_q = 0.25$. Plot from [44]

The corresponding interaction Lagrangian between the DM and SM quarks, where the coupling to quarks is assumed to be universal, is written as

$$L_{axial-vector} = g_q \Sigma_q Z'_\mu \overline{q} \gamma^\mu \gamma^5 q \; + \; g_\chi Z'_\mu \overline{\chi} \gamma^\mu \gamma^5 \chi \tag{3}$$

and g_q, g_χ are the coupling strengths to SM quarks and the DM particle respectively. If the mediator does not have any other decay channels, visible or invisible, then the minimal mediator width is:

$$\Gamma^A_{min} = \frac{g_\chi^2 m(Z')}{12\Pi} \beta^3_{DM} \Theta(m(Z') - 2m_\chi) \; + \; \Sigma_q \frac{3 g_q^2 m(Z')}{12\Pi} \beta^3_q \Theta(m(Z') - 2m_q) \tag{4}$$

and is fixed by the coupling choice. $\Theta(x)$ denotes the heaviside step function and

$$\beta_{DM,q} = \sqrt{1 - \frac{4m^2_{DM,q}}{m(Z')^2}} \tag{5}$$

is the velocity of the particle in the mediators rest frame. The limits become less stringent as the resonance intrinsic width increases, so we search for so called narrow resonances, meaning the intrinsic width of the resonance is small compared to the experimental dijet mass resolution. The mediator width as a function of its mass is shown in Figure 11 for the simple choice of $g_\chi = g_q = 1$.

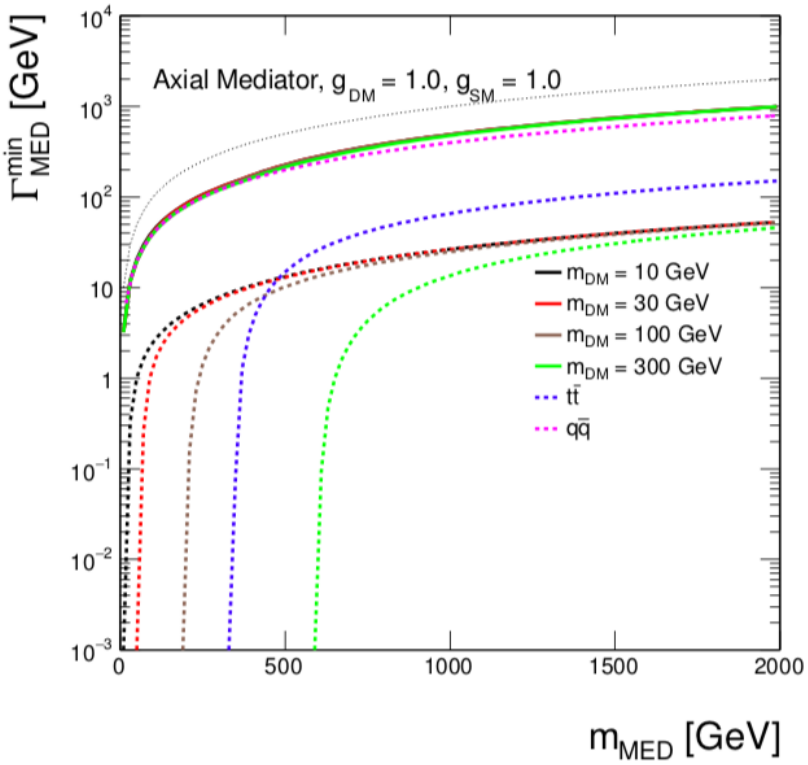


Figure 11: Minimal width, Γ^{min}_{MED}, as a function of mediator mass,m_{MED}, for an axial-vector mediator with $g_q = 1$.

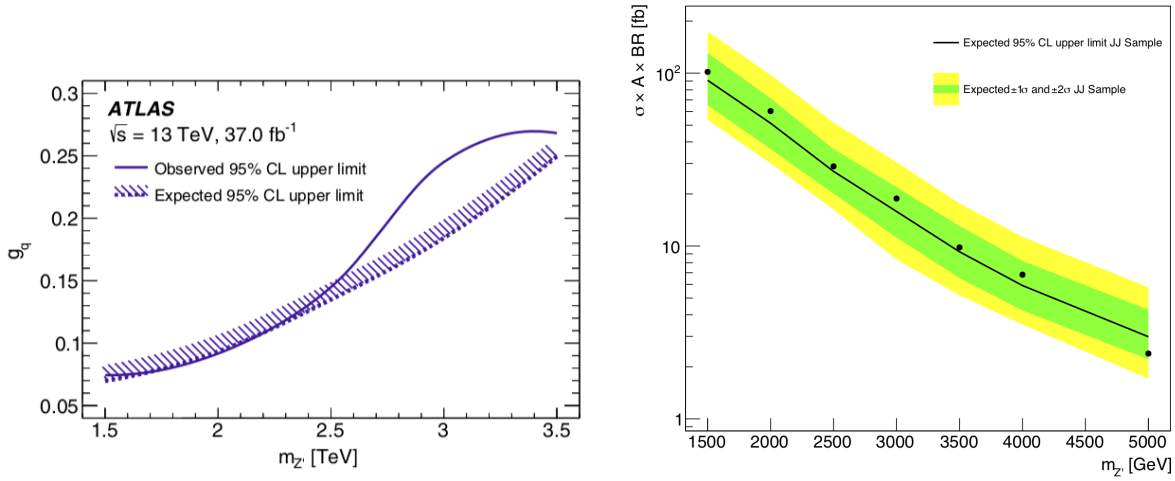
The parameters in this model are thus the coupling of the mediator Z' to SM quarks g_q, coupling of the mediator to the DM fermion g_χ, mass of the DM fermion m_χ and the mediator mass $m(Z')$. The two most relevant parameters in this dijet resonance analysis are the SM coupling strength, because it drives the intrinsic width of the resonance, and the mediator mass. Recent observations at the XENON-1T [58] experiment that could possibly correspond to a observation of a DM signal show a slight excess of events that point to a DM particle consistent with the TeV scale. For this search the DM mass and coupling of the mediator to DM fermion were fixed to 10 TeV and 1.5 respectively and the parameters were scanned to obtain a signal grid used for limit setting. The parameters obtained are shown in Table 2.

The limits set on the Z' signal in the (couling strength, mediator mass) plane gained in the 2015/16 Run-2 analysis of untagged data is shown below. The additional plot on the right shows the limits of cross section times acceptance times branching fraction set on the specific signal with $g_q = 0.2$ with the untagged data, as this will be the comparison basis of this analysis.

g_q	$m(Z')$ $[TeV]$								
0.05	1.5								
0.1	1.5	2.0	2.5						
0.15	1.5	2.0	2.5	3.0	3.5	4.0			
0.2	1.5	2.0	2.5	3.0	3.5	4.0	5.0		
0.3			2.5	3.0	3.5	4.0	5.0	6.0	
0.4						4.0	5.0	6.0	7.0
0.5						4.0	5.0	6.0	7.0

Table 2: Parameter grid in g_q and $m(Z')$ scanned in full simulation.

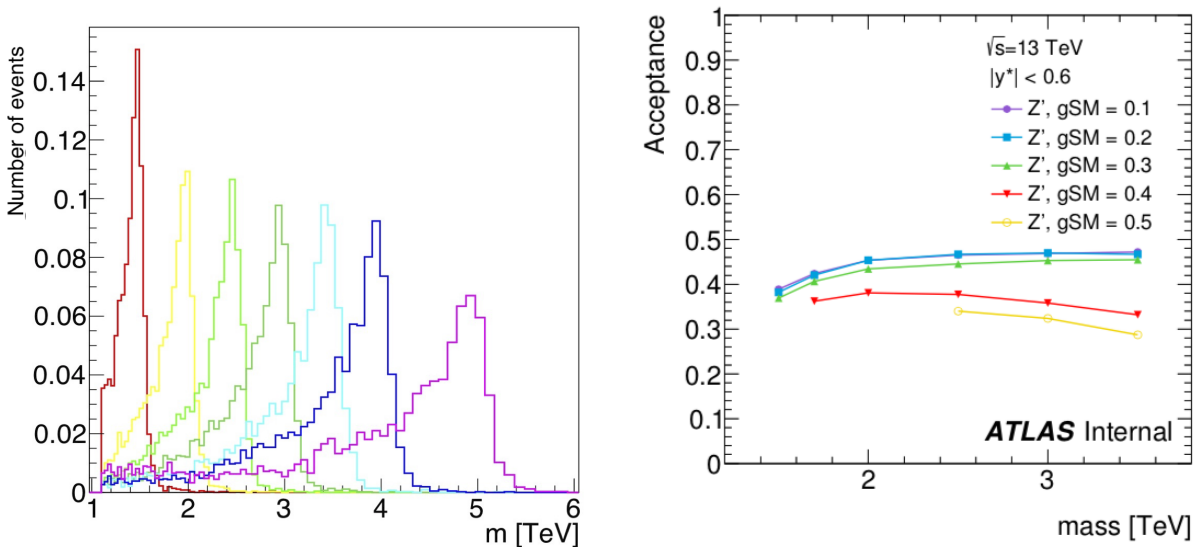
.



(a) (b)

Figure 12: On the left plot: exclusion limits of the Z' model as a function of g_q obtained by the un-tagged dijet analysis on 37fb^{-1} of data collected in 2015 and 2016. Figure obtained from [59]. On the right: limits set on the Z' signal model with $g_q = 0.2$.

The specific focus of this thesis was the signal model with $g_q = 0.2$, because it allows study of the Q/G tagger in a broad mass range, but also has the highest detector acceptance. The geometrical acceptance of the detector is determined by Monte Carlo Simulations and in general depends on the polar and azimuthal angle of the emitted particle. Both the mass range of the mediator and the detector acceptance for the medaitor with coupling strength to quarks $g_q = 0.2$ can be seen in Figure 12.



(a) (b)

Figure 13: Comparison of normalized Z' signal shapes for $g_{SM} = 0.2$ and varying mediator mass in a) and detector acceptance as a function of signal mass with varying coupling strengths in b).

3 The LHC and ATLAS Detector

3.1 CERN's Large Hadron Collider

The Large Hadron Collider (LHC) is currently the world's most powerful particle accelerator, situated 100 m underground in the tunnel previously used for the LEP experiments, and consisting of a 27 km ring used to collide two proton beams of opposite momentum.

The LHC is supplied with protons by a chain of accelerators that boost the particles before injecting them into the large ring. The process starts with a bottle of hydrogen gas at a linear accelerator called LINAC2. The gas is passed through an electric field, stripping it of electrons so that only protons enter the accelerator. Using radiofrequency cavities (RF cavities) to charge cylindrical conductors, the protons are accelerated by the same charged conductor behind them and by opposite charged conductor before them. Alternating the conductors between positive and negative charge, gives the protons a constant acceleration. As the protons reach the end of LINAC2, they reach energies up to 50 MeV and are passed onto the Proton Synchrotron Booster (PSB), which is the first circular accelerator in the injection chain. It consists of four superimposed rings with a diameter of 25 m that requires a special construction to split and combine the proton beams coming in and out of the PSB. A so-called proton distributor splits the beam by deflecting parts of it at different angles as it travels through successive pulsed magnets. Similarly, after being accelerated to 1.4 GeV, the beams are then merged in the manner that they were split. They are passed on to the Proton Synchrotron (PS), which gives the final acceleration up to 25 GeV before they are finally injected into the large LHC ring. Interestingly, when the LHC is running, it uses less than 0.1% of the protons prepared by the injector chain. This is mainly because the LHC is a storage ring, meaning the beams circulate the ring for hours, colliding at every intersection. There are four such intersection points, each with a detector to precisely measure the collision results: ATLAS, ALICE, LHCb and CMS. The injector chain and the detectors of the LHC can be seen in a schematic in Figure 14.

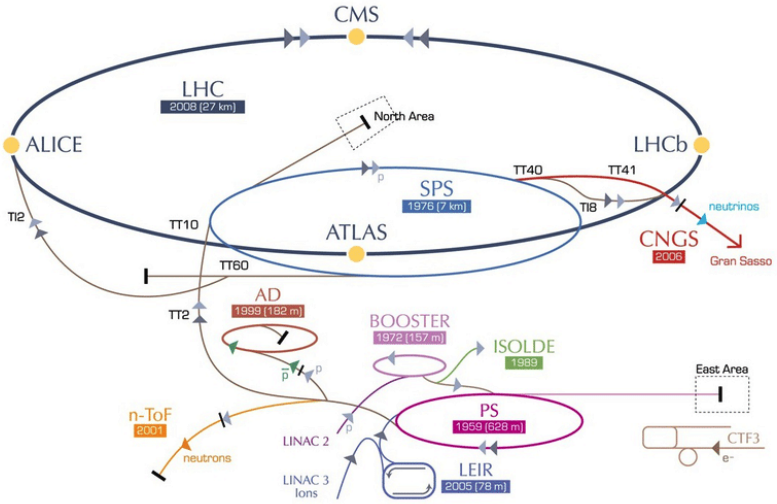


Figure 14: Diagram of the CERN Accelerator complex [60].

The LHC started operating in 2010 with the centre-of-mass energy $\sqrt{s} = 7$ TeV and delivered a total integrated luminosity of 5 fb^{-1} by the end of the year. The centre-of-mass energy reached in Run-2 is $\sqrt{s} = 13$ TeV with a staggering integrated luminosity of 140 fb^{-1}.

3.2 Overview of the ATLAS detector

The ATLAS detector(A Toroidal LHC Apparatus) [61] is a 46 m long cylinder with a 25 m diameter lying approximately 93 m underground to reduce the flux of cosmic rays from reaching the detector, but also to shield the outside from high energy radiation produced in the collisions. It consists of several subsystems laying concentrically around the collision point in order to measure the trajectories, momentum and energy of the outgoing particles, which in return enables us to identify and measure their properties. The layout of the detector's sub-components cane be viewed in Figure 15.

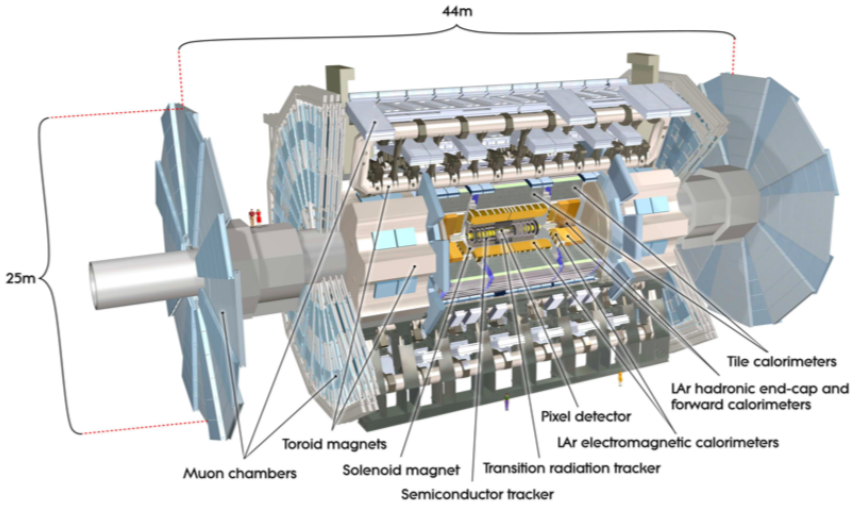


Figure 15: Diagram of the ATLAS detector and its sub-components [61].

When describing the detector and the physics inside it, a right-handed coordinate system is used with the z-axis defined by the beam direction and the origin of the coordinate system placed at the nominal interaction point. The positive x-axis direction points towards the centre of the LHC ring and the positive y-axis direction points upwards along the zenith. Angles used to describe more complex variables are the azimuthal angle ϕ and polar angle θ, where ϕ is measured around the beam axis, in the x-y plane, and θ is measured from the z-axis. The pseudorapidity is defined as

$$\eta = -\ln\tan\left(\theta/2\right), \tag{6}$$

which is equal to rapidity if the object is approximately massless. When the mass can not be neglected, like in the case of jets, rapidity y is more appropriate

$$y = \frac{1}{2}\ln(\frac{E+p_z}{E-p_z}), \tag{7}$$

where E is the energy and p_z is the momentum projected in the z-direction. Rapidity differences are invariant with respect to Lorentz boosts along the beam axis, that is why it is frequently used in accelerator physics. However, the laboratory frame rarely coincides with the center of mass frame of the particles, because of their composite nature, and the complexity of the physics means that η is easier to estimate than y. In the high energy regime ($pc \gg mc^2$) the two quantities are almost identical [63]. To limit the number of events we record, jets in consideration must pass analysis specific cuts. For a dijet analysis the two most important factors are the rapidity difference $|y*| = \frac{1}{2}|y_1 - y_2|$ and the dijet invariant mass, calculated as:

$$m_{jj} = \sqrt{(E_1^2 + E_2^2 - |\overrightarrow{p_1}^2 + \overrightarrow{p_2}^2|)} = \sqrt{\hat{s}} = 2p_T\cosh y^* \tag{8}$$

Apart from a large enough centre-of-mass energy to produce a heavy resonance, the determining factor in a discovery is also it's significance. The interaction rate is dependent on the particle's

cross section and on the luminosity, L, of the LHC, defined as:

$$L = \frac{f}{4\pi} \frac{N_a N_b n_b \Im}{\sqrt{\beta_x \varepsilon_x \beta_y \varepsilon_y}}, \tag{9}$$

where N_a, N_b are the number of protons in the bunch, n_b is the number of bunches in the ring, f is the revolution frequency, $\beta_{x,y}$ are the amplitude functions at the bunch crossing, $\varepsilon_{x,y}$ is the emittance and $\Im$ represents the geometrical correction factor.

3.2.1 Inner Detector

Data acquisition starts with the Inner Detector composed of a Pixel Detector (PD), Semiconductor Tracker (SCT) and Transition Radiation Tracker (TRT) providing the direction, momentum and charge of electrically charged outgoing particles. The arrangement in the Inner Detector can be viewed in Figure 16, while the specific layout of the sensors and structural elements can be seen in Figure 17. High-resolution as well as fine-granularity detectors at inner radii with continuous tracking elements at the outer radii, all contained in a solenoidal magnet, are required to make high-precision measurements of the very large number of tracks at the LHC. The system provides full coverage in ϕ with the acceptance in pseudorapidity $|\eta| < 2.5$ for particles coming from the beam intersection. A diagram of the ATLAS Inner detector can be seen in Figure 15 as well as in Figure 16. The later shows the detector's components in the central region with the radial distances of the sub-components from the intersection point.

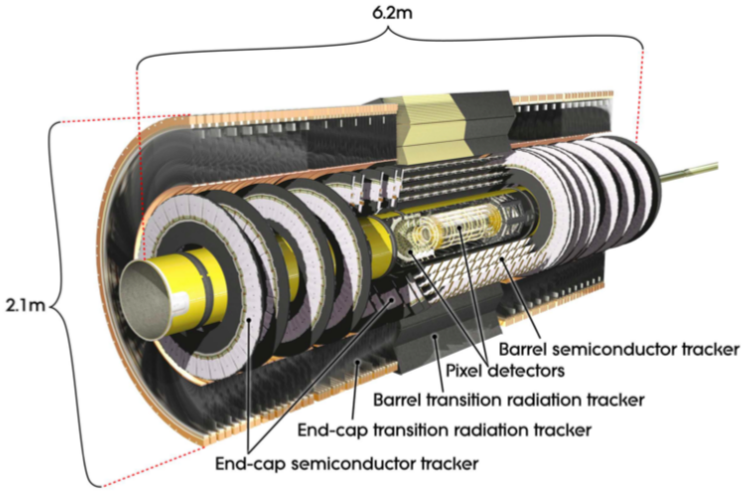


Figure 16: Diagram of the ATLAS Inner Detector and its sub-components[61].

3.2.1.1 Pixel Detector

To provide measurements as close to the interaction point as possible, the Pixel detector consists of 1744 silicon pixel modules [64] arranged in three concentric barrel layers. Arranged in rings at the ends of pixel barrels are the end-caps made of end-cap pixel modules. Sensitive elements cover radial distances 50.5 mm $<$ r $<$ 150 mm and typically provide three measurement points. The module's active area of 16.4 mm$\times$60.8 mm contains approximately 47232 pixels the size of 50 μm$\times$400 μm with the shorter side of the module defining the local x-coordinate and the longer side the y-coordinate. The local x-coordinate provides position measurements in the $R\phi$ plane with the y-coordinate orientated along the z direction of the detector. Each barrel module contains read-out chips with individual circuits for each pixel element including buffering to store the data while the level-1 trigger decision is made (explained later on). Pixel hits are read out if the signal exceeds a threshold. The pulse height is measured using the Time-over-Threshold technique [62].

3.2.1.2 The Semiconductor tracker

Like the PD, the Semiconductor Tracker (SCT) is composed of four barrel detectors arranged in cylindrical layers and two symmetric end-cap detectors containing nine disks each. It typically provides eight strip measurements (four space-points), with the barrel strips being orientated parallel to the solenoid field and beam axis, while the end-cap strip direction is radial. Strips on each side of the module have a small angle between them to perform stereo angle measurement covering the radial region of 299 mm $<$ r $<$ 514 mm.

3.2.1.3 Transition Radiation Tracker

The detector consists of 298304 drift tubes (straws) covering the radial distance of 563 mm $<$ r $<$1066 mm. The straws are 4 mm in diameter and arranged in three cylindrical layers and 32ϕ sectors. Each of them has split anodes and is read from each side [65]. Straws in the end-cap region are arranged in wheel-like modular structures with radial orientation [66]. The layout ensures that charged particles with $p_T > 0.5$ GeV and $|\eta| < 2.0$ cross more than 30 straws. Electron identification is provided by the polypropylene fibres between the straws in the barrel region or foils between the straws in the end-cap region.

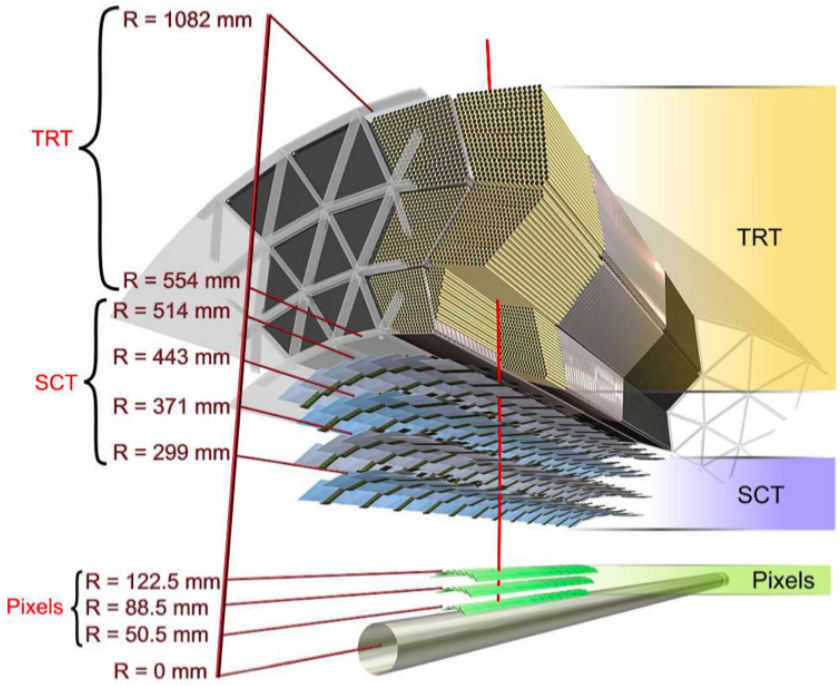


Figure 17: Diagram of the Inner detector showing the sensors and structural elements [61].

3.2.2 The Calorimeters

The ATLAS calorimeters cover the range $|\eta| < 4.9$ and are designed to measure the energy (and position/direction) of both charged and neutral particles as well as jets with the material and techniques differing to suit the physics process of interest. An important design consideration is the thickness of the detector, because it must contain the electromagnetic and hadronic showers, but also limit punch-through to the muon system. The whole system is schematically shown in Figure 18.

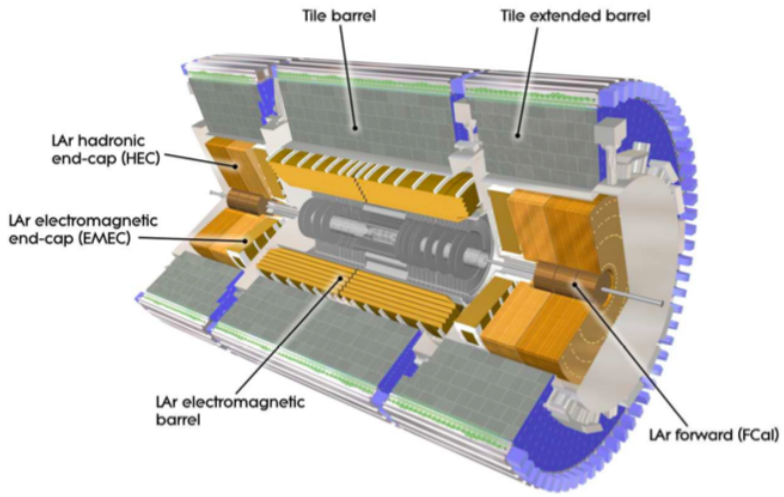


Figure 18: ATLAS calorimetry system[61].

3.2.2.1 Electromagnetic Calorimeter

The Lead Liquid Argon (LAr) detector is used for the EM and hadronic calorimetry and shares a vacuum vessel with the solenoidal magnet, eliminating the need for two vacuum walls. The EM calorimeter is composesd of two barrel parts and two end-cap components, each in its own cyrostat. The accordion geometry of the absorbers and electrodes in the barrels and end-caps provide a full coverage in ϕ without any cracks as well as a fast signal extraction. The waves of the accordion shape are axial and run in ϕ in the barrels and are parallel to the radial direction and run axially in the end-caps. An accurate position measurement is achieved by finely segmenting the first layer ($0 < |\eta| < 2.5$) in η.

3.2.2.2 Hadronic Calorimeter

The hadronic calorimeter is composed of three parts: the tile calorimeter, the liquid-argon hadronic end-cap calorimeter (HEC) and the liquid-argon forward calorimeter (FCal).

3.2.2.3 Tile Calorimeter

The tile calorimeter surrounding the EM calorimeter covering the region $|\eta| < 2.7$, uses steel as an absorber and scintillator as an active medium. It's sub-devided into three barrels, each containing 64 steel-scintillator modules further divided radially into three layers: the tiles, the fibres and the photomultipliers (PMT). When ionising particles cross the tiles, they induce a production of ultraviolet scintillation light in the polystyrene material. This light is subsequently converted into visible light by flour molecules doped into the polystyrene in the wavelength-shifting fibres located at the tile edges. Two wavelength-shifting fibres at the edge of each tile collect light and transmit them to the PMT's housed at the outer edge of the module.

3.2.2.4 Hadronic end-cap Calorimeter and Forward Calorimeter

Using a flat-plate design, the HEC module is a copper/liquid-argon calorimeter covering the range $1.5 < |\eta| < 3.2$, sharing the two end-cap cryostats with the EM end-cap calorimeters and FCal. The end-caps are cylindrically shaped, consisting of two wheels each with an outer radius of 2030 mm and consisting of 32 identical modules. Covering the range of $3.1 < |\eta| < 4.9$ is the FCal. A very hermetic design allows the FCal to sit in the same cyostats as the end-cap calorimeters, providing minimal energy loss into cracks between the system and excess into the muon system. The system is split into three modules: one electromagnetic, with copper absorber, and two hadronic with tungsten absorbers. A shield of copper alloy has been mounted to the back of the last module to reduce background in the end-cap muon system.

3.2.3 Muon Spectrometer

The outermost part of ATLAS is the Muon Spectrometer, built to measure the momentum of charged particles that have escaped detection in the calorimeters.The MS contains a variety of sub-detector components to provide high precision tracking measurements of particles bent by the toroidal field in the range $|\eta| < 2.7$, but also fast read-outs for triggering in the $|\eta| < 2.4$ region. The whole system reflects the symmetry of the toroidal magnet, consisting of eight octants. Each is subdivided in to two sectors in the azimuthal direction with different lateral extensions to provide an overlap in ϕ minimalism the gaps in detector coverage. Precision-tracking chambers in the barrel region are situated between and on the eight coils of the magnet, while the end-cap chambers are in front and behind the end-cap magnets. The barrel chambers are concentrically shaped while the end-cap chambers are wheel shaped, both sub-divided into three layers. Monitored Drift Tube chambers (MDT) consisting of drift tubes containing argon/carbon-dioxide mixture at 3bar are used for precision momentum measurements, which result in a 80 μm per tube and 35 μm resolution per tube. At the tube ends, the electrons resulting from ionisation are collected with a tungsten-rhenium wire. Because of the multi-tube design, high reliability is provided as the failure of a signal tube does not affect the performance of the surrounding tubes. However the disadvantage of the radial drift geometry causes a track crossing close to the wire to create a pulse train. The pulse duration is proportional to its order and may mimic several threshold crossings (hits) per track. To compensate for this, an adjustable dead-time has been implemented into the front-end of the readout chain. The MDT's are replaced by Cathode Drift Chambers (CDC) in the innermost layer ($2 < |\eta| < 2.7$) to cope with counting rates up to ~ 1000 Hz/cm^2. Each track is measured in four

consecutive CDC planes in comparison to the MDT's where a track typically crosses 2×4 inner, 2×3 middle and 2×3 outer layers of MDT tubes, however the faster response time restores the tracking efficiency.

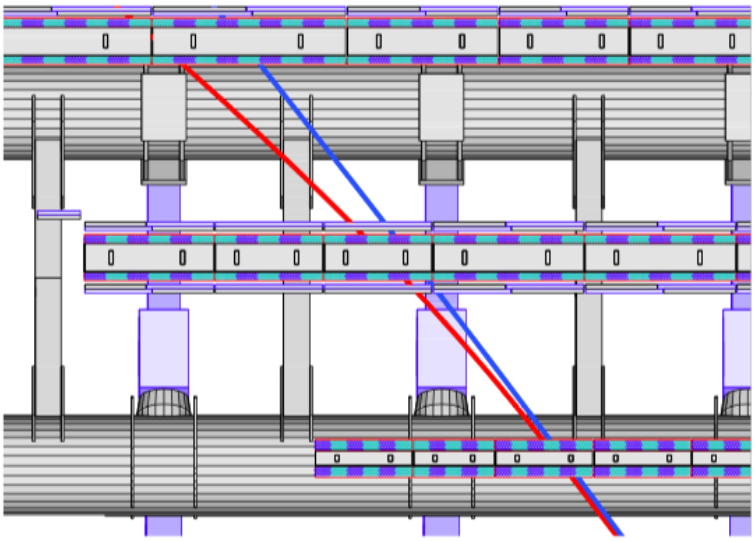


Figure 19: Trajectories of muons with momenta of 4GeV and 20GeV in the bending plane of the barrel muon spectrometer [61].

Similarly to the MDT's, two disks with eight chambers compose the CDC's system and are sub-divided into small and large chambers in ϕ. Each chamber has four individual planes providing four independent measurements in η and ϕ along the track. The wires are oriented in the radial direction, with both cathodes segmented in order for the one with strips perpendicular to the wires to provide precision coordinate and the other with parallel wires to provide the transverse coordinate. The track position is then interpolated between charge induced by neighboring cathode strips.

3.2.4 Data Aquisition and Trigger System

In Run 2, the LHC was operating at a centre-of-mass energy of 13TeV with an instantaneous luminosity of $1.2 \cdot 10^{34}$ cm^{-2}s^{-1}, with a bunch-crossing rate of approximately 40MHz. The ATLAS trigger system has to cope with these challenges in order to efficiently select the relevant physics processes and reduce the rate to around 1000 Hz. The whole trigger system is composed of a hardware Level-1 (L1) and a software based high-level trigger (HLT). The L1 trigger is implemented in a fast custom-made electronics to cope with new incoming data every 2.5 μs and reduces the interaction rate to about 100 kHz. A decision to record the event is made based on regions of interest in η/ϕ retrieved from the L1 Muon and L1 Calorimeter systems. In the HLT a decision is formed typically within 300 ms. It typically provides 2500 independent trigger chains executing offline-like algorithms within the L1 regions of interests and higher precision reconstruction is performed later

in the sequence. Events accepted by the HLT are written into different data streams depending on their usage (e.g. for physics analysis or detector calibration), which allows high rates or incoming data without consuming significant amount of available bandwidth.

4 Event Simulation, Reconstruction and Selection

4.1 Jet Reconstruction

Jets are the the outcome of a $2 \rightarrow 2$ parton scattering by strong interactions and are frequently observed at ATLAS, however the precise definition of a jet is needed in order to conduct any analysis. The idea of clustering final state hadrons into jets is to reduce the number of complex particle tracks to a final state involving just a few jets. There are a number of properties that are desired by the jet algorithms described by the *"Snowmass accord"*[69, 70], which were further supplemented with experiment specific details. IRC (infrared and collinear) safe algorithms ensure the parameters used in the analysis are independent of unpredictable dynamics within the jet formation and results can be compared between experiments like ATLAS and CMS, which use different tracking and calorimeter equipment. Schematic views of IR and coliinear safe algorithms are seen in Figure 20 and 21 respectively.

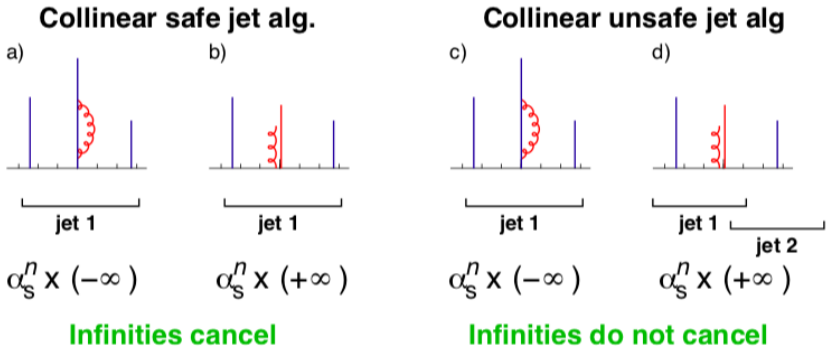


Figure 20: Illustration of collinear safe (upper left) and unsafe (upper right) jet algorithms. The vertical lines represent partons with height proportional to their momentum and the vertical line represents rapidity. Image taken from [73].

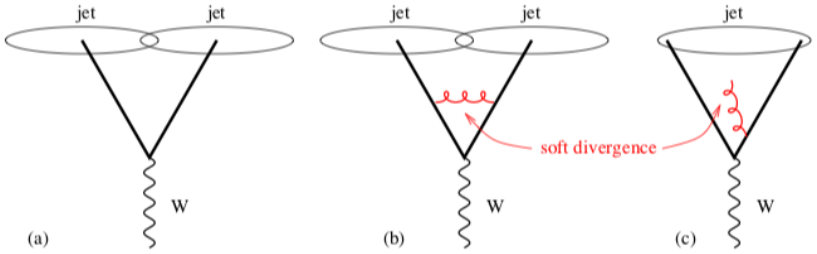


Figure 21: Illustration of IR unsafe algorithm, where the additional soft gluon emission changes a two jet event to a single jet event, with the explicit angular distribution shown. Illustration obtained from [73].

4.1.1 Topo-cluster formation

To try and extract a significant signal from electronic noise and other sources of fluctuations, such as pile-up, the calorimeter cells are clustered into topologically connected cell signals. The individual topo-clusters reconstructed are not expected to contain the detector response to a single particle all the time, but rather the fractional response to a single particle (full shower or shower fragment), the merged response of several particles or a combination of merged full and partial showers.

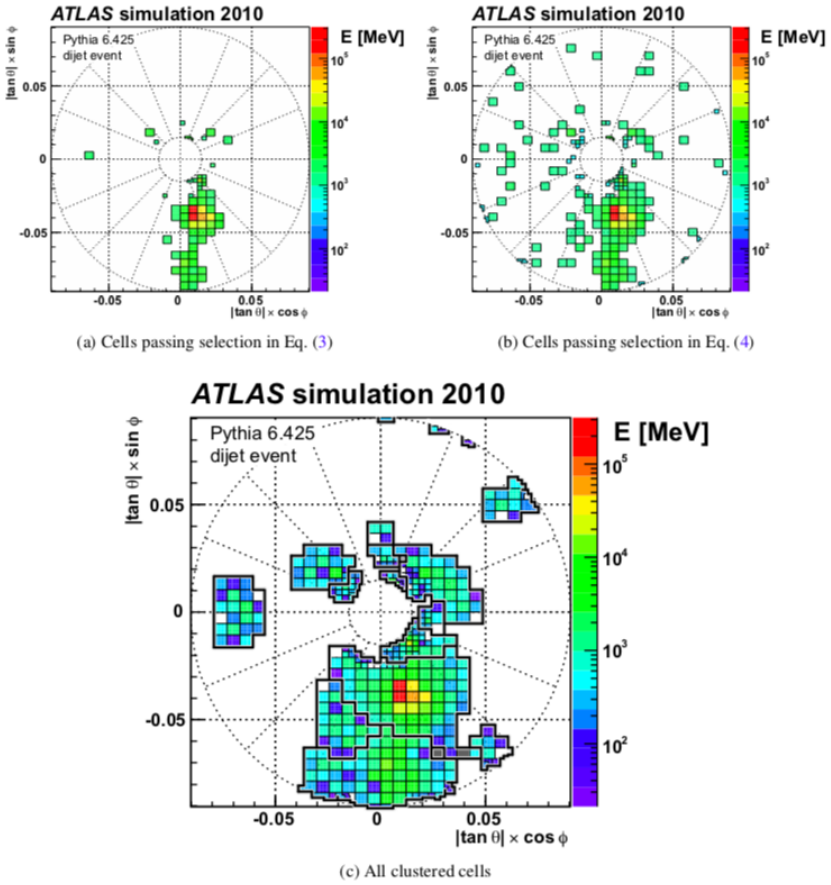
The observable dictating the cluster formation is the cell signal significance, defined as the ratio between the cell signal,E^{EM}_{cell}, and the expected noise in the cell, $\sigma^{EM}_{noise,cell}$. Both are measured on the electromagnetic scale. The algorithm starts with a cell with highly significant signal and formates cluster in "seed and collect" steps, which are repeated until all the connected cells pass the following criteria:

$$|\varsigma^{EM}_{cell}| = \frac{E^{EM}_{cell}}{\sigma^{EM}_{noise,cell}} > S = 4 \tag{10}$$

$$|\varsigma^{EM}_{cell}| > N = 2 \tag{11}$$

$$|\varsigma^{EM}_{cell}| > P = 0 \tag{12}$$

Each seed cell satisfying the first equation above forms a proto-cluster and is added to a list of decreasing ς^{EM}_{cell}. Next, the neighboring cells are evaluated. Neighbors are considered as two cells directly adjacent in the sampling layer or in adjacent layers, having at least partial overlap in (η,ϕ) plane. This way the sampling of cells can span over modules within the same calorimeter as well as the sub-detector transition regions. The cells that neighboring the seed cell and satisfy the second and third equations above are collected into the corresponding proto-cluster. If a neighboring cell's significance passes the second equation, then it's neighboring cells are also added to the proto-cluster. If a neighbor is a seed cell, then the two proto-clusters are merged, the same process applies if two proto-clusters share a neighboring cell pasing the second equation. The formations stops when the neighboring cells pass the threshold P, but no longer threshold N. The resulting proto-clusters are characterized by a core of highly significant cells that are enclosed by an envelope of less significant cells. These clusters are then inputted into a jet reconstruction algorithm. The procedure of topo-cluster formation can be seen in Figure 21.



(a) Cells passing selection in Eq. (3)

(b) Cells passing selection in Eq. (4)

(c) All clustered cells

Figure 22: Illustration of topo-cluster formation, with the cells passing the seeding criteria shown in a), cells passing the growth criteria shown in b) and finally the last picture depicts the cells that were clustered into topo-clusters with their outline. Illustration obtained from [71].

4.1.2 Anti-k_t algorithm

The algorithm used in this analysis is a so-called sequential clustering algorithm, meaning the process starts from a single calorimetric cluster, called $anti-k_t$ with the R-parameter set to 0.4. It analyses two distances; the distance between two particles, d_{ij}, and the distance between the particle and the beam axis, d_{iB}, defined as:

$$d_{ij} = k_{ti}^{2p}(k_{tj}^{2p})\frac{\Delta_{ij}^2}{R^2} \tag{13}$$

$$d_{iB} = k_{ti}^{2p} \tag{14}$$

where $\Delta_{ij}^2 = (y_i + y_j)^2 + (\phi_i + \phi_j)^2$ and k_{ti}, y_i and ϕ_i are the transverse momentum, rapidity and azimuth of particle i. The parameter p describes the relative power of energy versus geometrical scales. The values of p = 1 or 0 correspond to the inclusive k_t algorithms and Cambridge/Aachen algorithm [72], where the value of -1 refers to the $anti-k_t$ algorithm.

For a given final state particle, both of those distances are calculated and compared. If d_{ij} is smaller, then the particles i and j are combined using the summation of four vectors and removed from the list of particles. If d_{iB} is smaller, then i is labelled as a final jet and removed from the list of particles. The functionality of the algorithm can be best seen in application. Considering an event with a few well separated hard particles with transverse momenta k_{t1}, $k_{t2}\ldots$ and many soft particles. The d_{i1} between a hard particle 1 and a soft particle i will be determined by the transverse momentum of the hard particle and geometrical separation. The d_{ij} between two soft particles will be much larger, therefore the soft particles will tend to cluster with hard particles long before they cluster with themselves. In the case of a hard particle with no hard particles in the radius of 2R, it will accumulate all the soft particles within that circle resulting in a perfectly conical jet. If there is another hard particle such that $R < \Delta_{12} < 2R$ then there will be two jets, but only one the jet with larger transverse momenta will be perfectly conical and the smaller one will miss the part in which it's overlapping with the larger jet. In case both hard particles have equal transverse momenta, the overlapping area is split equally by a straight line and neither of the jets will be canonical. Similarly if $\Delta_{12} < R$ for the two particles, they will be clustered into a single canonical jet centered on the particle with the larger transverse momenta.

The key feature outlined above is that soft particles do not change the shape of the jet, meaning the algorithm is resilient with respect to soft radiation, but flexible towards hard radiation. Comparison of multiple jet algorithms (varying p-value) is shown in Figure 23 on a parton-level event with approximately 10^2 random soft *"ghost"* particles. The area of which the random ghosts are clustered within a jet is shown in color and one can see that for the $anti - k_t$ algorithm the hard jets are circular whereas the soft jets have more complex shape. The best comparison of the algorithms can be seen for $\varphi = 5$ and $y = 2$ [74].

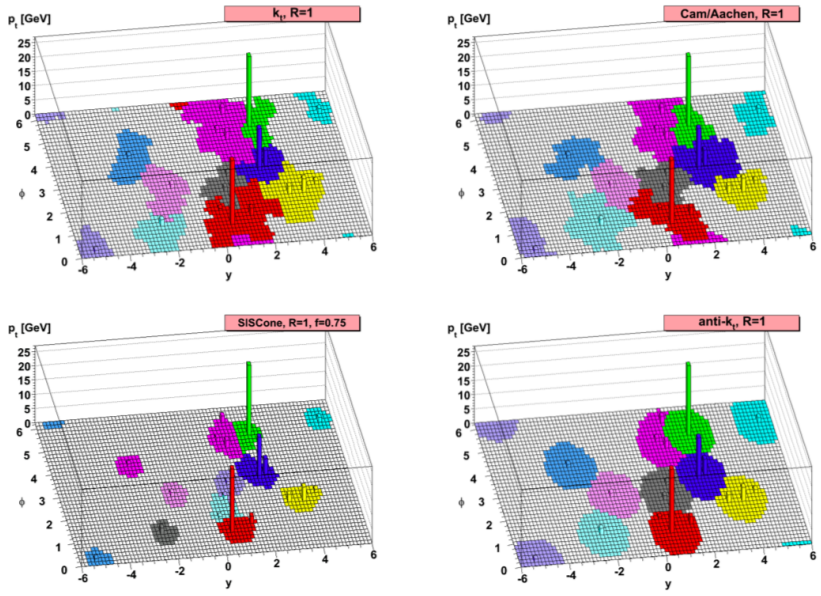


Figure 23: A sample parton-level event generated with Herwig and clustered with different jet alogithms. Image from [74].

4.2 Event Simulation

While the search for physics beyond the SM is a very exciting task, as physicists we must be able to use the SM to provide accurate predictions of the experiments we are doing. We do so by generating simulated events with generators like Pythia, which play the role of the accelerators like the LHC and put them through detector simulations like GEANT that play the role of the ATLAS detector. The generated data goes through the exact same reconstruction stages as the collected data and allows us to apply the same physical cuts. Monte Carlo simulations can be used for various aspects of the analysis from detector calibration to determining the systematic uncertainties. In this analysis it is used to model the Z' signal as well as the data sample. This provides a chance to fine tune the parameters and functions used in differentiating quark and gluon jets, because unlike the data, MC provides the "truth" information of each event.

The main objective of the MC simulation is to provide us with a calculation of the cross-section of the process in question, reflecting the probability of the process to occur. The calculation relies on the QCD factorization theorem allowing us to separate the hard-scatter partonic cross-section from the non-perturbative low momentum interactions. The latter are given by the Parton Distriution Functions (PDFs) relating the momentum of the parton to the momentum of the incoming proton. Because of their non-perturbative nature, they are derived empirically and are accessible in a library of PDF functions, but some are also built-in to Pythia.

The partonic cross-section is proportional to the square of the matrix element multiplied by the phase space part of the cross-section. To know the probability of finding final state particles 4-momenta in a certain part of available phase space, we have to integrate their probability density functions. If the number of final state particles is very large, then this integral is over several dimensions and impossible to calculate with the standard methods. Such integrals are evaluated by Monte Carlo methods by randomly generating 4-momenta of the final state particles with a random number generator. The "Hit and Miss" [75] integration is the simplest MC method with the idea to generate random points in some region in space with known volume, which also encloses the volume of interest. If the external volume is V_e and the fraction of hits that hit the volume we want to compute is f_h, then the region we are trying to evaluate is simply $V = V_e f_h$. While in the simplest example our final state is composed of two partons, the detector observes jets, meaning a multi-hadron final state.

The trick to handling multi-particle final states is to factorize a complex $2 \to n$ process to a simple $2 \to 2$, convoluted with showers [76, 77]. The probabilities of a splitting to occur are summarized by the DGLAP equations that can be combined to allow for successive emissions and model the development of a parton shower in several steps. Here the Sudakov factor is used to express the shower evolution to smaller and smaller Q^2 instead of evolving to later and later times. The inclusion of the Sudakov factor ensures the probability branch never exceeds unity. The simplest way to imagine the MC method is considering a gluon with probability of emitting another gluon, $P_{g\to gg}$, probability of splitting to a $q\overline{q}$ pair, $P_{g\to q\overline{q}}$, and probability of nothing happening $P_{nothing} = 1 - (P_{g\to gg} + P_{g\to q\overline{q}})$. At this point, the MC generates a random number r. If $r < P_{g\to gg}$, then the gluon radiates another gluon, the additional gluon is added to the list of particles and the momenta of the initial gluon is reduced. If $P_{g\to gg} < r < P_{g\to gg} + P_{g\to q\overline{q}}$, then the gluon splits into a $q\overline{q}$ pair. Finally if $r > P_{g\to gg} + P_{g\to q\overline{q}}$, then nothing happens.

Additional to the final state radiation, the initial partons could have radiated partons. The

picture inside the proton is not static, gluons are constantly splitting and are emitted and absorbed by the quarks, however this is done in a non-perturbative region. The before mentioned DGLAP equations emerge in a conditional probability of the PDFs: if parton b is present, what is the probability that it came from a previous branching $a \rightarrow b\ c$. This is known as the "backwards way" point of view, where one starts with the hard scattering process and tried to reconstruct the picture before.

Up to this point, everything is said to be at parton level, however our experiments only observe hadrons. The shower generators typically also model the hadronisation process, with the specific generator used in this analysis utilising the Lund string model. Here the quarks are joined by strings that stretch as the quarks separate, creating a $q\overline{q}$ pair as the strings snap. As mentioned previously, this process continues until the energy is no longer high enough. Hadronization can be further divided into fragmentation and decay, where the hadrons produced in the hadronization process decay into lower mass hadrons. Apart from the partonic hard-scattering, the remaining partons not involved in this process can also radiate and create final-state jets, making pp collision simulations significantly more complicated. Modeling this underlying event can be very complex and is also performed by the shower generator. The modeling itself can be improved by tunable parametes with an optimised set of these parameters called the Monte Carlo tunes.

The final step involves modeling the detector responses. Until this point, the MC is said to be truth level, referring to an output we would have gained in an ideal detector. To obtain the reconstucted MC, the truth level is paired with additonal pile-up interactions and put through a detector simulation called Geant. This creates the sought after energy deposits left in the simulated detector that can be now put through event selection and jet reconstruction algorithms specific to the analysis and described below.

The data sample used in this analysis is modeled with Pythia 8 [78] using the A14 set of tuned parameters for the underlying event and the leading-order NNPDF2.3 parton distribution functions [79] with the detector effects then modeled by Geant 4 [80, 81].

4.3 Event selection

The ATLAS detector is a complex multi-layered system, so the ATLAS Data Quality group labels luminosity blocks as either good, flawed or bad to mark the operation of several subsystems in that period and the information is stored in Good Run List (GRL). Luminosity blocks which are not listed in the GRL are considered bad, and should not be used for analysis. The GRL removes entire luminosity blocks, which can include thousands of events. To avoid discarding good events among them, there are additional detector-level flags that label individual events as appropriate for physics. The baseline selection cuts applied to the data in this analysis are:

1. GRL: Requirement that all relevant detectors were in good shape and ready for physics
2. LAr: Liquid Argon calorimeter error rejected(errorState(xAOD::EventInfo::LAr))
3. Tile: Tile Calorimenter error rejected (errorState(xAOD::EventInfo::Tile))
4. SCT: SCT single event upsets rejected(errorState(xAOD::EventInfo::SCT))
5. Core: Incomplete event build rejected (isEventFlagBitSet(xAOD::EventInfo::Core))
6. Primary vertex: the highest $\Sigma p_T^2(trk)$ vertex has at least two tracks associated with it (xAOD::VxType::VertexType::PriVtx)
7. Trigger: passes the lowest unprescaled single-jet trigger HLT_j380
8. All jets with $p_T > 150$ GeV pass LooseBad cleaning cuts
9. Leading jet $p_T > 440$ GeV
10. Subleading jet $p_T > 60$ GeV

Where the selections 6-10 are specific to analysis considering jet final states. To increase the sensitivity, two additional cuts specific for resonance analysis in the higher mass range are applied:

1. $|y*| < 0.6$
2. $m_{jj} > 1100$ GeV

with $|y*| = \frac{1}{2}|y_1 - y_2|$ being the Lorentz invariant rapidity difference in the two leading jets. This cut selects central jets to limit the QCD background. The dijet rapidity difference is related to the polar scattering angle θ, which is sensitive to the dynamics of the underlying process. The QCD processes are dominated by t-channel gluon exchanges, leading to a Rutherford-like distribution, while new physics processes mainly go through s-channel decays, making the angular distribution well defined by the spins of the resonance and the initial and final state partons. Applying a rapidity difference cut increases the signal significance.

4.3.1 Jet cleaning

The LooseBad selection was designed to produce high good jet efficiency while maintaining a high fake jet rejection. To reject fake jets due to significant coherent noise in the LAr calorimeters, a characteristic ionization signal is gained from simulated electronic responses. The quadratic difference between the measured signal in a given LAr calorimeter cell and the expected pulse shape,Q_{cell}^{LAr}, is used to discriminate between the fake and real energy deposits. From this cell-level quantity (Q_{cell}^{LAr}), several jet-level quantities are derived. $\langle Q \rangle$ is the average jet quality, defined as the energy squared average of the pulse quality in the calorimeter cells in the jet and is normalized such that $0 < \langle Q \rangle < 1$. f_Q^{LAr} is the fraction of energy deposited in the LAr cells of a jet with poor quality defined as $Q_{cell}^{LAr} > 4000$ and similarly the f_Q^{HEC} is defined as a fraction of energy deposited in the hadronic end-cap (HEC) calorimeter cells of a jet with poor signal quality ($Q_{cell}^{LAr} > 4000$). Since noisy cells in the calorimeters can generate a large fake energy deposit as well as a large fake negative energy deposit, the sum of all cells with negative energy is introduced as a new variable E_{neg}. The presence of negative energy in a good jet happens due to electronic and pile-up noise. Most of the jets have a low value of E_{neg}, except for fake jets with $E_{neg} > 60$ GeV that also have large values of f_Q^{HEC}. Comparison of good jet and bad jet enriched sample plots for the four derived variables can be seen in Figure 24.

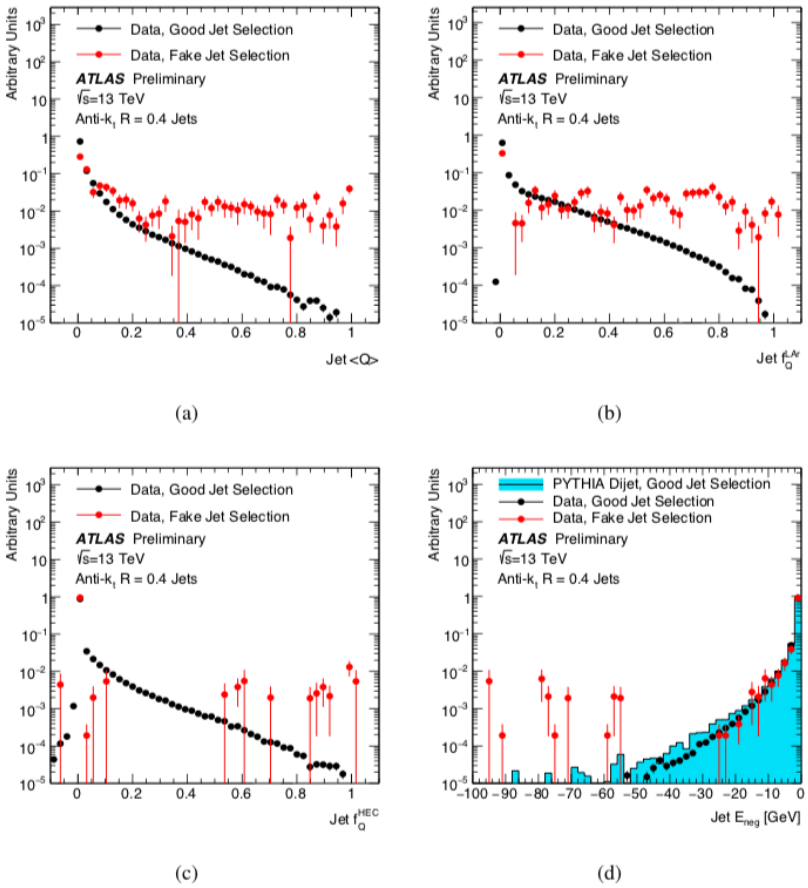


Figure 24: Distributions of f_Q^{LAr} in a), f_Q^{HEC} in b), $\langle Q \rangle$ in c) and E_{neg} in d). The distribution of a good jet enriched sample is shown in black, distribution coming from the fake jets in red and the simulated data sample only shown in d) as the blue histogram. The plots were taken from [83].

Calorimeter noise and beam-induced background show more as localized excess oriented longitudinally in the calorimeter cells, so two additional variables limiting the fake jets are introduced: the electromagnetic fraction (f_{EM}) and the hadronic end-cap fraction (f_{HEC}), both defined as the ratio of energy deposit in the specific calorimeter and the total energy. This way, the jet energy deposit in the expected direction of parton showering is used to discriminate between real and fake jets. Additionally, since the ATLAS calorimeters are segmented in depth, the maximal energy deposit in any layer (f_{max}) can be used. Good jets have a smooth distribution for all three variables, which can be seen in Figure 25 with a comparison to a bad jet enriched sample.

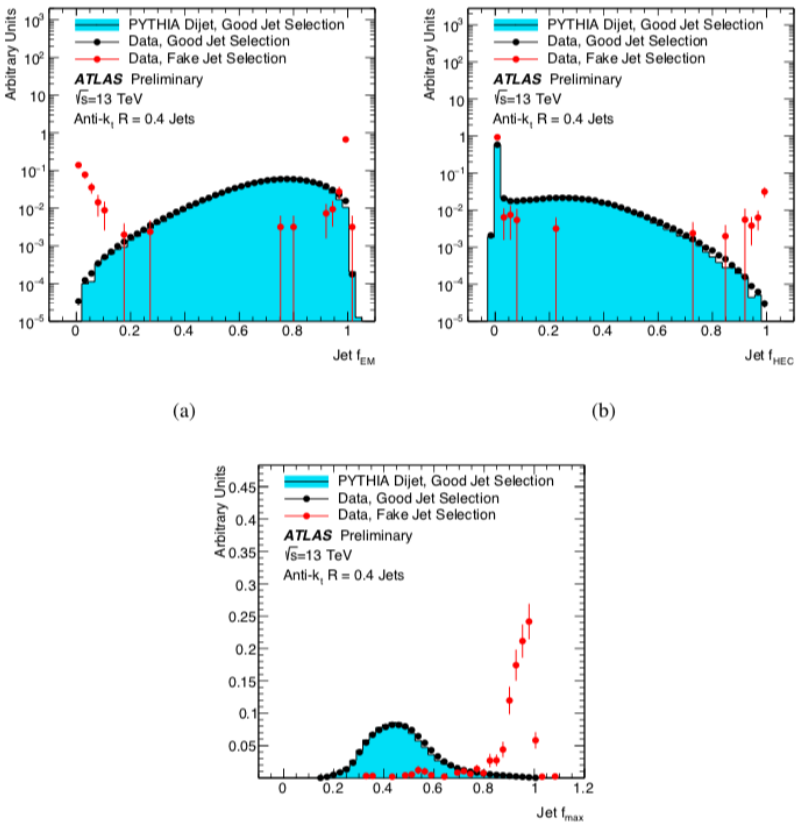
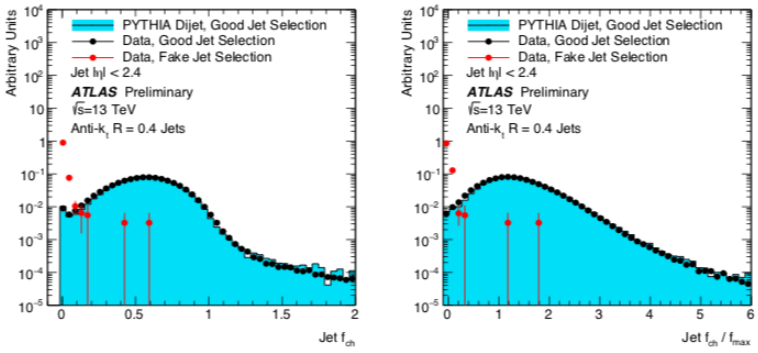


Figure 25: Distributions of f_{EM} in a), f_{HEC} in b) and f_{max} in c). The distribution of a good jet enriched sample is shown in black, distribution coming from the fake jets in red and the simulated data sample shown as the blue histogram. The plots were taken from [83].

The last variable and very important for this analysis is the jet charged fraction (f_{ch}), that takes into account the fact that real jets contain charged hadrons which can be reconstructed by the ID tracking system. It is calculated as a fraction of a scalar sum of the p_T of tracks originating from the primary vertex and the jet p_T [83]. Oly a small number of good jets have small values of f_{ch} and f_{ch}/f_{max}, which can be seen in Figure 26, therefore these variables can serve as additional cuts for bad jet rejection.



(a) (b)

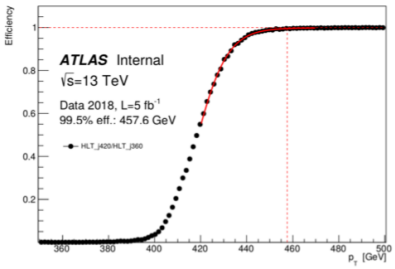
Figure 26: Distributions of f_{ch} in a) and f_{ch}/f_{max} in b) for $\mid \eta \mid < 2.4$. The distribution of a good jet enriched sample is shown in black, distribution coming from the fake jets in red and the simulated data sample shown as the blue histogram. The plots were taken from [83].

In summary, a jet passes the LooseBad cut, if it satisfies at least one of the following conditions and is taken from the data sample appropriate for analysis:

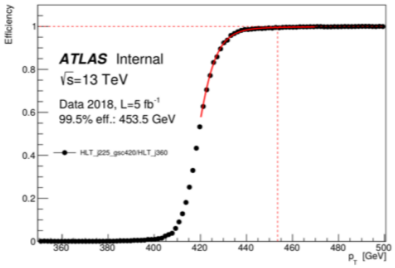
1. $f_{HEC} > 0.5$ and $\mid f_Q^{HEC} \mid > 0.5$ *and* $\langle Q \rangle > 0.8$
2. $\mid E_{neg} \mid > 60$ GeV
3. $f_{EM} > 0.95$ and $f_Q^{LAr} > 0.8$and $\langle Q \rangle > 0.8$ and $\mid \eta \mid < 2.8$
4. $f_{max} > 0.99$ and $\mid \eta \mid < 2$
5. $f_{EM} < 0.05$ and $f_{ch} < 0.05$ and $\mid \eta \mid < 2$
6. $f_{EM} < 0.05$ and $\mid \eta \mid \geq 2$

4.3.2 Triggering

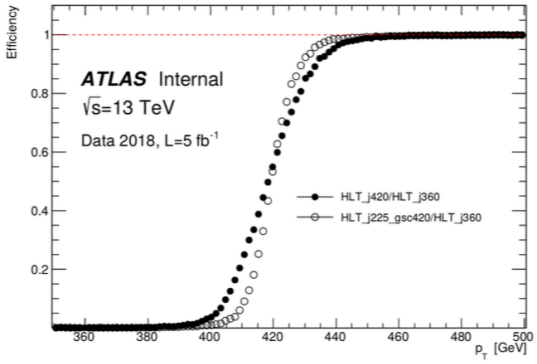
For single-jet triggers, the naming convention follow either 'Jnnn' for L1 triggers or 'jnnn' for HLT trigger, where 'nnn' denotes the p_T threshold for the trigger in GeV. The energy scale used for the L1 triggers is the EM scale, while for HLT triggers a similar calibration sequence to that used for offline jets is applied to the HLT jets, bringing their scale to the hadronic scale. For the complete Run-2 dataset, two single-jet unprescaled triggers were used, the HLT_j420 and HLT_j225_gsc420_boffperf_split. While both search for jets with $p_T > 420$ GeV, the later also applies the global sequential calibration to improve trigger turn-on. The triggers full efficiency, plotted in Figure 24, provides a lower bound for the jet invariant mass at 1.1 TeV, while the upper bound is unconstrained.



(a) HLT_j420



(b) HLT_j225_gsc420_boffperf_split

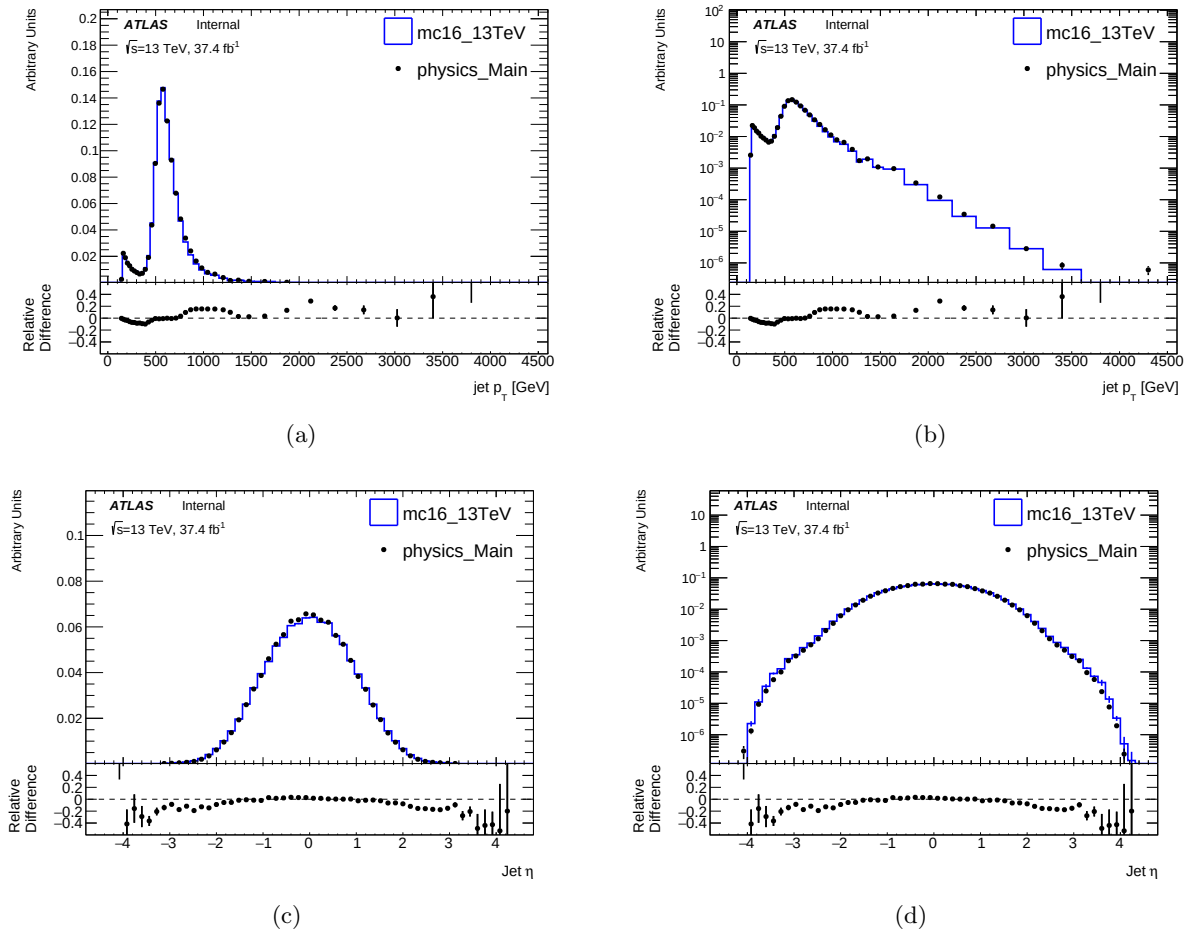


(c) Comparison

Figure 27: Trigger efficiency as a function of jet p_T for both HLT triggers in the first two plots and their comparison in the plot below.

4.4 Data and MC comparison

The baseline selection and specific selections for dijet resonances are now applied to both collected data and MC simulated data. Before conducting any further analysis, both samples are compared as a cross-check in case the samples show discrepancies in the phase space of the analysis. The plots shown in Figures 26-29 show good agreement between the jet basic kinematic variables as well as the invariant mass of the whole sample and sub-samples after applying the Q/G tagger.



(a) (b)

(c) (d)

Figure 28: Comparison plots of data collected in 2015/16, labeled as Physics_ Main, and MC generated data, labeled as mc16_ 13TeV, for the basic kinematic variables: jet p_T and jet η with the plots on the right shown in logarithmic scale.

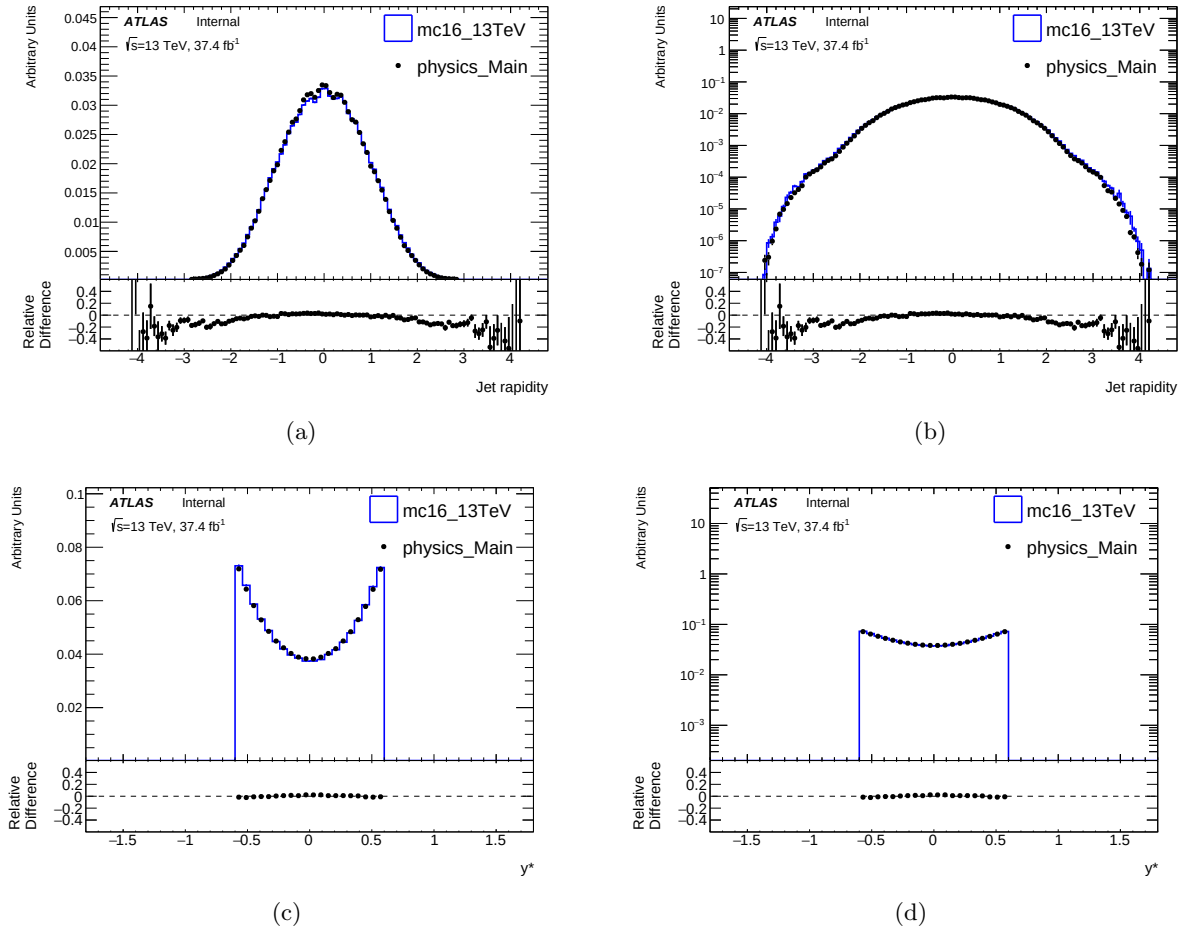


Figure 29: Comparison plots of data collected in 2015/16, labeled as Physics_ Main, and MC generated data, labeled as mc16_ 13TeV, for the basic kinematic variables: jet rapidity and y^* with the plots on the right shown in logarithmic scale.

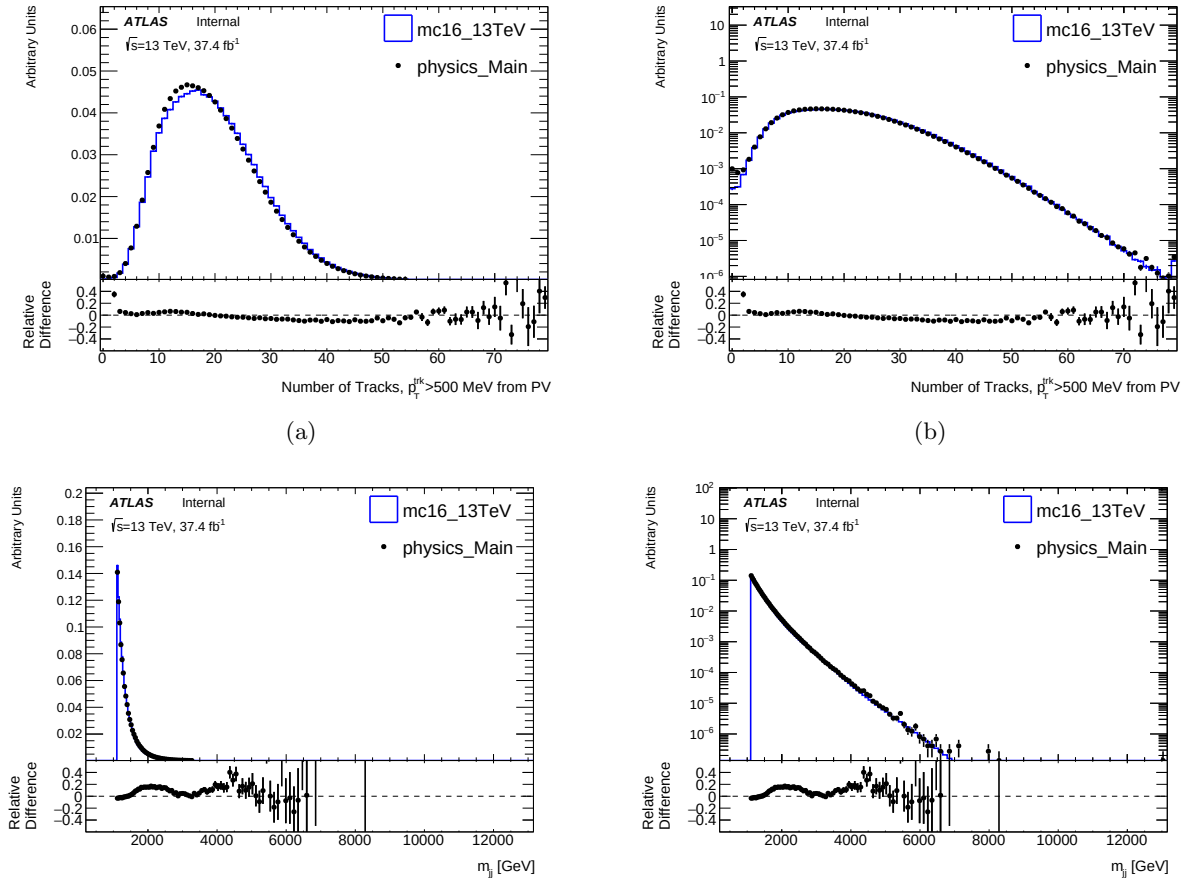
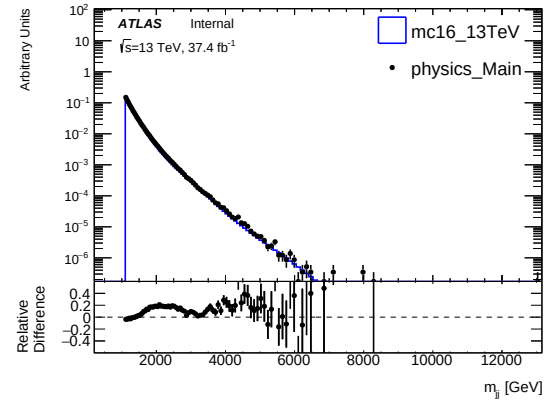
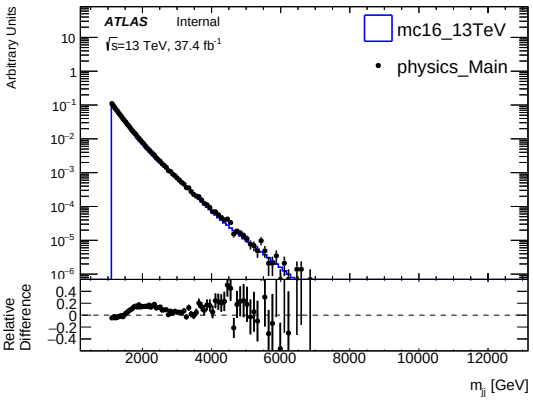


Figure 30: Comparison plots of jet track multiplicity and whole data sample m_{jj} with their logarithmic scale plots on the right for data collected in 2015/16, labeled as Physics_ Main, and MC generated data, labeled as mc16_ 13TeV.

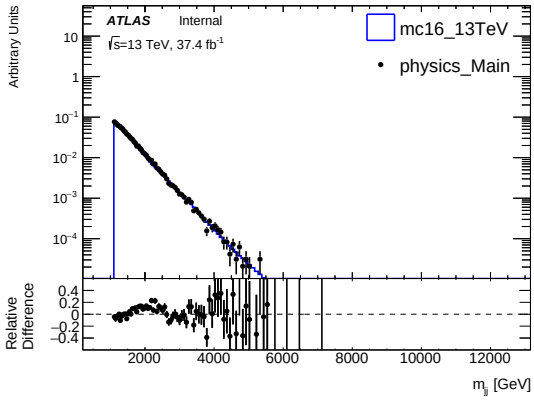
There is a slight displacement in the peaks in plot a) of Figure 31 where the number of jets is plotted against the number of tracks. There are more jets with higher number of tracks in simulation indicating that this is due to the reconstruction efficiency. In reality reconstructing particle tracks is very difficult, but this is also very hard to simulate as you would need the right number of charged particles produced in the collision and then the right track reconstruction to match that number. Even with the data and simulation showing slight differences, the difference between quark and gluon jets multiplicities are much larger, so we would not expect this to affect the analysis further. The dependency of jet track multiplicity on the simulation software is also being studied and will be included in the uncertainties of the Q/G tagger.



(a)



(b)



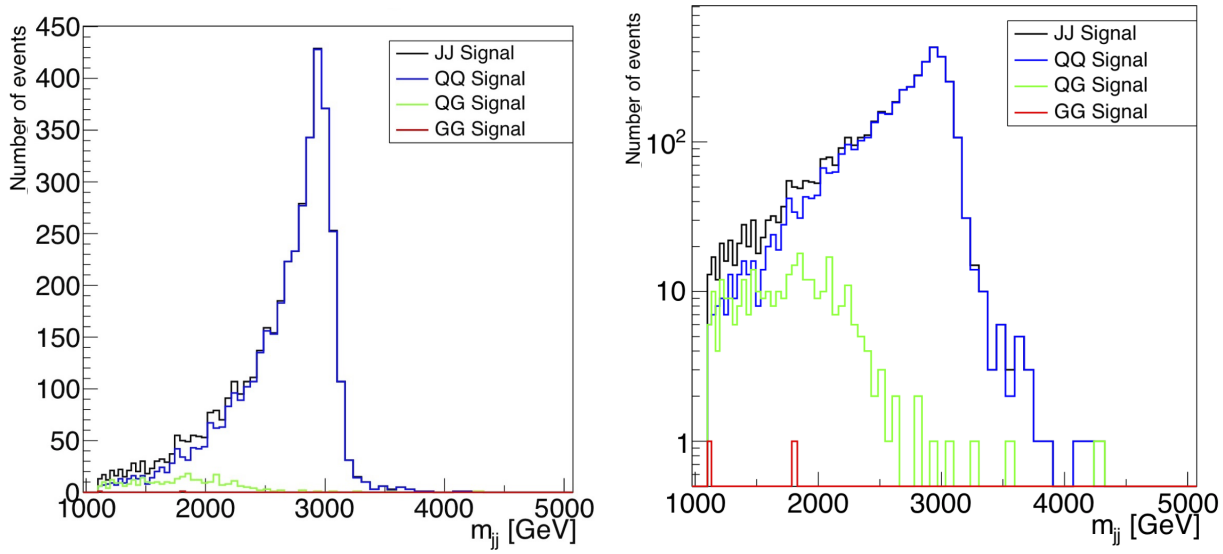
(c)

Figure 31: Comparison plots of a random selection sub-samples: QQ,QG and GG shown in a),b) and c) respectfully, for data collected in 2015/16, labeled as Physics_ Main, and MC generated data, labeled as mc16_ 13TeV.

5 Analysis

5.1 Jet Tagging

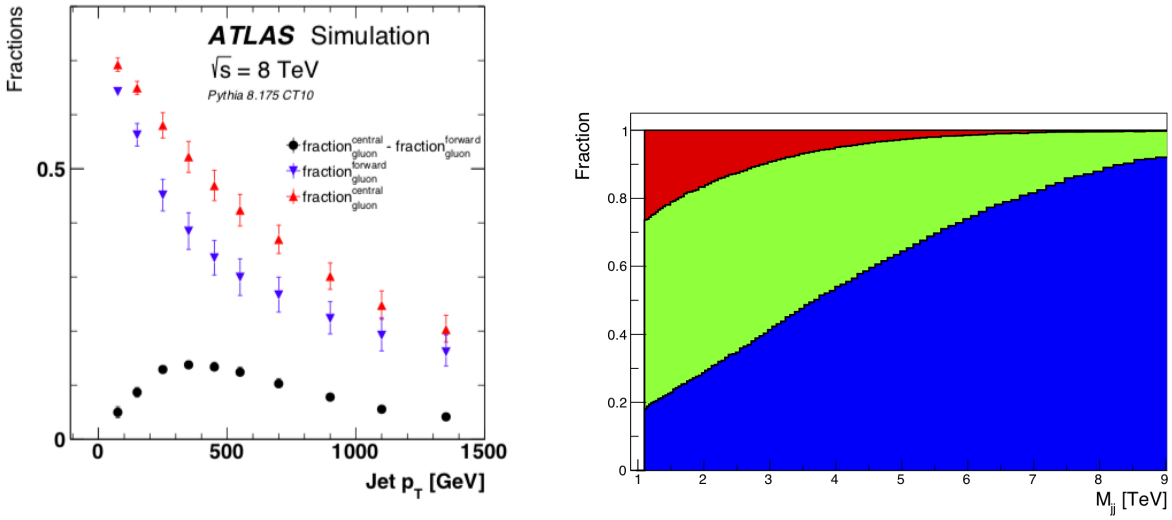
In order to increase the sensitivity for the mediator search, or other searches for physics beyond the SM in general, the fact that the main decay channel of the Z' to SM particles is to quarks is exploited.



(a) (b)

Figure 32: Plots showing the compositions of a 3 TeV simulated signal as functions of m_{jj} with the quark dijets colored blue, a mixed dijet event (quark and gluon) in green, gluon events in red and the untagged dijet signal in black. For a clearer view, the right plot is plotted as $\log(y)$.

The ability to distinguish between quark and gluon initiated jets has a long history [84, 85, 86, 87, 88] in new physics searches as well as measurements that test QCD processes. In pp-collisions, the dijet final states are a mixture of QQ, QG and GG events, however, because of the parton density functions, plotted in Figure 3, a vast amount of background events are expected to be GG at lower p_T and QQ in the higher regions. As can be seen from the figures below, the number of gluon initiated jets decrease with increasing p_T, however, the signal is mostly made of QQ dijet events. With a good method of differentiating the two jets, one can expect to limit the background in the low m_{jj} region and subsequently improve the search sensitivity.



(a) (b)

Figure 33: In a): fraction of jets originating from gluons as a function of jet p_T for the more forward jet (down triangle), the more central jet (up triangle), and the difference between these two fractions (circle). Plot taken from [89]. In b): the fractions of QQ, QG and GG dijet events calculated from the MC generated data at truth level.

There are a number of theoretical predictions from QCD that predict subtle differences between quark and gluon jets. Gluons carry both a color and anti-color charge, therefore a higher all together color charge than quarks. This is the reason we expect differences in fragmentation for quark and gluon jets, leading to gluon jets having a higher charged particle multiplicity, softer hadron spectrum and a broader p_T spectrum. As described in the jet reconstruction and SM sector, in the process of fragmentation (showering), the jet initiating parton radiates another parton in a cascade manner until the partons hadronize. More importantly, the Altarelli-Parisi splitting functions contain the Casimir factors of C_A and C_F for gluon radiating a gluon and gluon radiating of a quark ($C_A/C_F = 9/4$), which results in the difference in particle multiplicities inside quark and gluon initiated jets. Initially, the track multiplicity was studied with Monte Carlo simulations as a function of $|\eta|$, because in a dijet event well balanced in p_T, the jet with higher η (forward jet) is more likely to be a result of scattering off a valance quark. The filled circles and squares in the figure below represent the value of charged particle multiplicities for forward/central jets, open red and black point show charged particle multiplicities in quark/gluon gained directly from MC simulations for forward/central jets, while the open blue points are calculated by using the set of equations

$$\langle n^{q,g}_{charged} \rangle = f^{f,c}_q \langle n^q_{charged} \rangle \; + \; f^{f,c}_g \langle n^g_{charged} \rangle \tag{15}$$

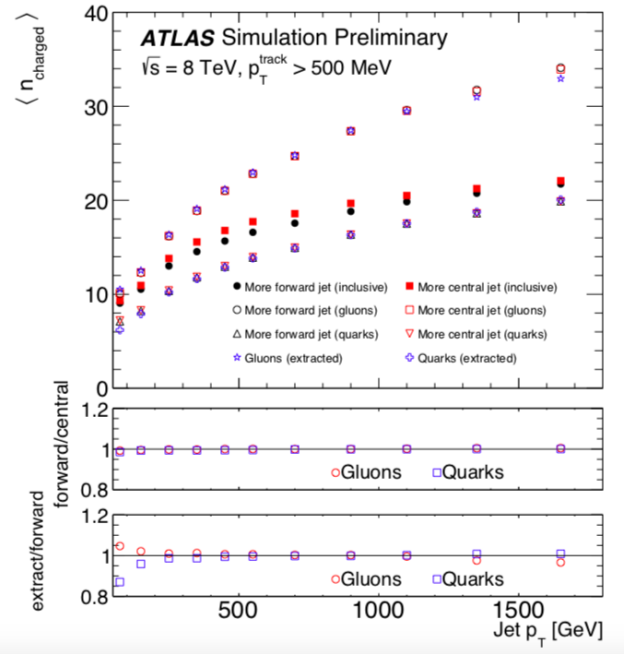


Figure 34: Graph demonstrating that the charged particle multiplicity inside a jet only depends on jet p_T, because the open stars, circles and up triangles as well as open crosses, squared and down triangles are stacked on top of each other. Plot taken from [89].

The circles and squares are almost perfectly stacked on top of each other meaning that the jet multiplicity shows negligible dependence on $|\eta|$, but rises with rising jet p_T. The sole discriminant used in this analysis, and described in [89], is the number of tracks inside a jet with the charged particle transverse momentum above 500 MeV. The discriminant functions tested in this analysis were linear and logarithmic functions of jet, p_T, or invariant mass, m_{jj}, with two parameters, offset and slope, written as

$$n_{track} = k \times p_T(m_{jj}) \; + \; n_0 \tag{16}$$

where k and n_0 are the two varying parameters. If the number of tracks measured in the jet is larger than the number of tracks calculated by the above equation, then the jet was marked as gluon initiated. Similarly for a lower track number the jet was marked as quark initiated. This separated the whole dijet sample into three sub-samples: QQ and GG, where both jets are either quark or gluon initiated, and QG sample with one of each. An example of a separated dijet spectrum can be seen in Figure 33.

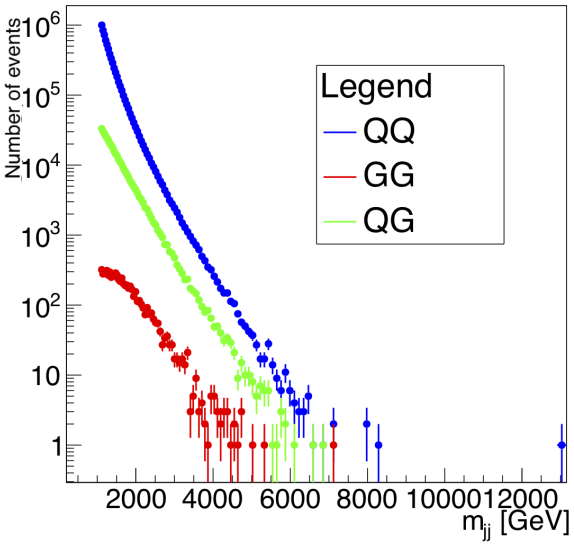


Figure 35: Plot showing a simulated dijet sample divided into sub-samples of only quark jets, QQ, only gluon jets GG and a mixture of both, QG, based on a logarithmic function of jet p_T.

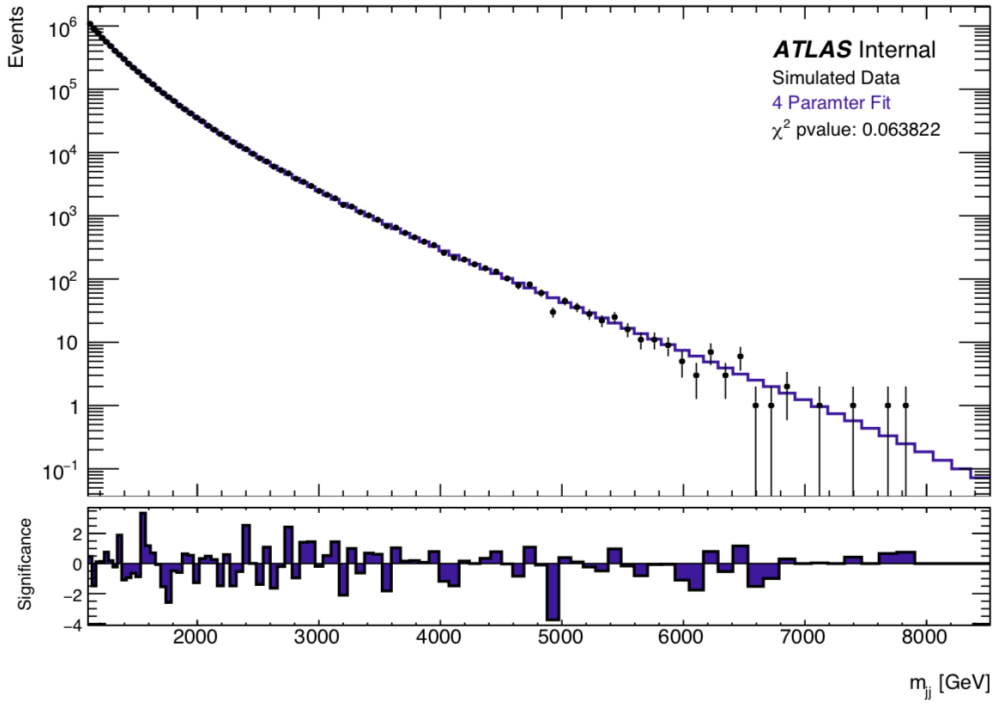
5.2 Background estimation

To obtain a background estimation on the data samples the Sliding Window Fits (SWiFt) algorithm was used as opposed to a global fit. SWiFt algorithm is a resonance search method that searches for data excess by sliding over the data in small overlapping windows which are automatically determined by the algorithm. It starts by placing a window center at the bin edge near the low mass regions. The first few and last few bin edges of the histogram are not used as window centers to avoid edge-effects. The window size is chosen by performing multiple nominal background-only fits with different window sizes and choosing the one with the best χ^2 p-value. The window sizes are specified by the user, for example 20 windows ranging from 30% to 100% of the window center. Scanning over different window sizes takes into account the "window choice" uncertainty. After performing an analysis in the chosen window size, SWiFt slides the window center one bin to the right and picks a new window size around the new center. For the first window, all possible sizes are scanned. For all other windows, a total of 5 windows are scanned: two below and two above the previous window size. This allows the window size to grow or shrink in a smooth and controlled manner.

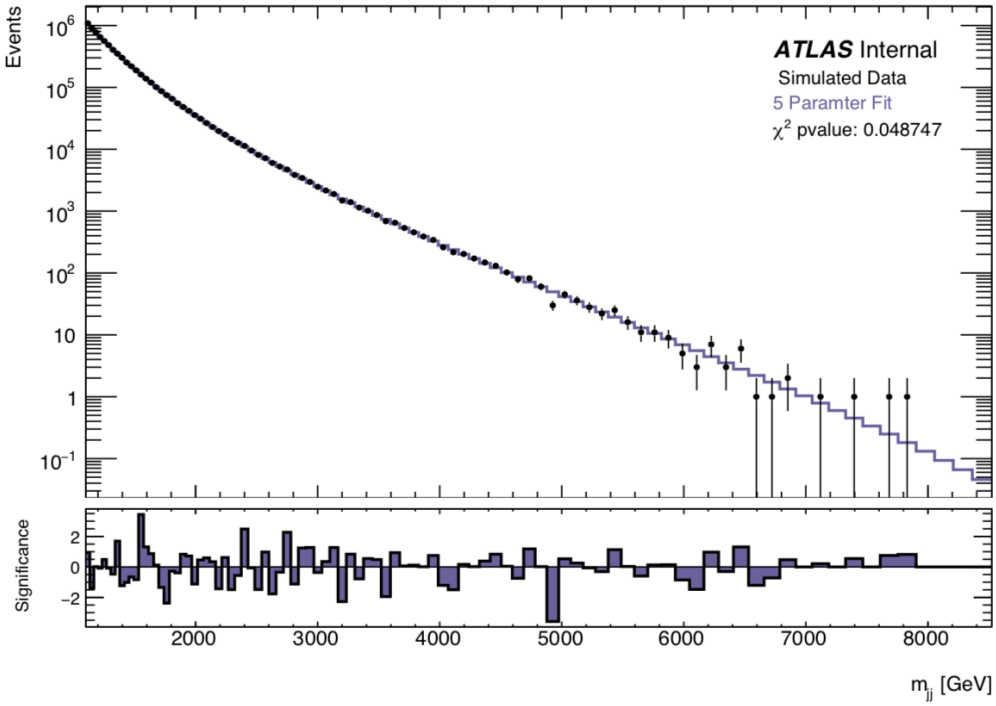
After the window size is fixed at a certain window center, SWiFt performs the background-only fits. The background-only functions are the "dijet functions" defined as,

$$f(x) = p_0(1-x)^{p_1} x^{p_2+p_3 \ln x + p_4 (\ln x)^2} \tag{17}$$

where p_i are the fit parameters and $x = m_{jj}/\sqrt{s}$. The 5 and 4-parameter version, obtained by setting $p_4 = 0$, are used in this analysis as the nominal and alternate functions. Historically the *"dijet functions"* were very successful in global fits, however with increasing the span of the data and data itself, it became too challenging to perform global adhoc fits.



(a)



(b)

Figure 36: Figure showing the SWiFt background using only 5 and 4 parameter functions.

Depending on which fit has a better χ^2 p-value, calculated by

$$\chi^2 = \Sigma_{i_0}^{i_n} \frac{(x_i - \lambda_i)^2}{\lambda_i}, \tag{18}$$

in the window, with x_i and λ_i representing bin content and background fit prediction respectively, SWiFt chooses to either keep the nominal or alternate function. Using the χ^2 p-value takes into account the number of degrees of freedom, which allows for a fair comparison between the functions and prevents the highest-order function being favored (hence preventing over-fitting). The ability to swap between the functions takes into account the background function choice uncertainty. In each window the background-only fit is evaluated at the window center and as the window slides across the distribution bin-by-bin the SWiFt background is constructed by stitching together the background-only fits at the window centers. A SWiFt background fit to a simulated data sample can be viewed in Figure 35. For the very first and last windows, an addition to evaluating the background-only fit at the window center, the bins that are below (for first window) or above (for last window) are evaluated. This gives the SWiFt background estimations for the bins at the edges of the mass distribution.

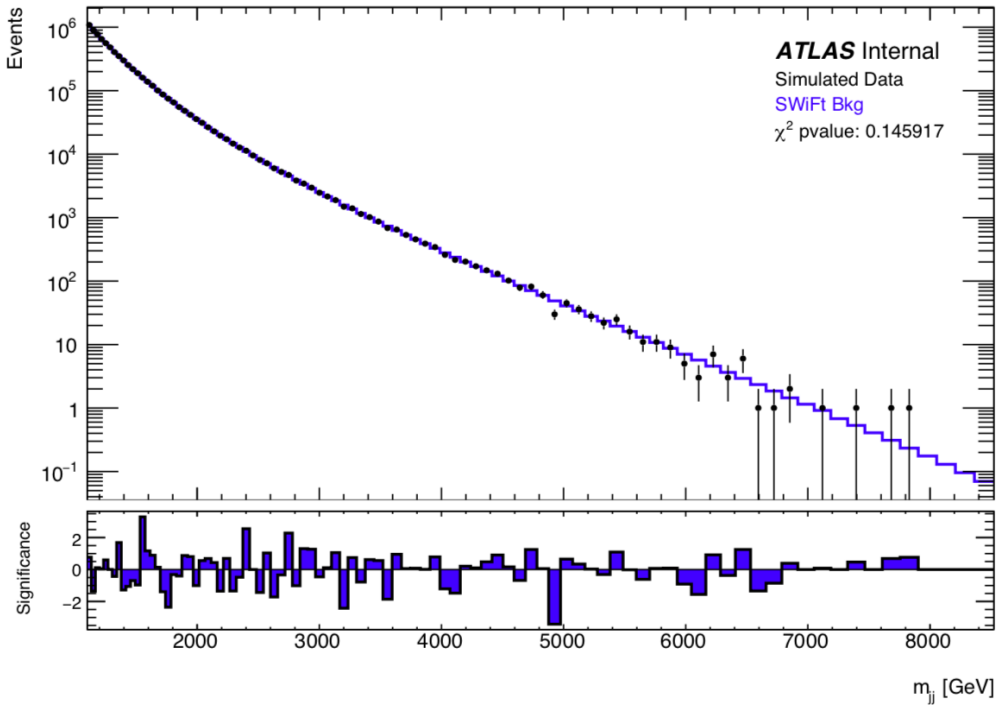


Figure 37: Figure showing the SWiFt background fit, using 4 and 5 parameter functions as the nominal an alternate background functions.

5.3 Limit Setting

This analysis is conducted on a MC generated data with no injected signal, so the main focus was on the comparison of the tagged and un-tagged dijet limits. HistFitter [90] begins with the HistFactory package to construct a parametric probability density function that describes the input data histogram and contains the parameters of interest, in our case the rate of a signal process, as well as the the nuisance parameters, which describe the systematic uncertainties, and are further denoted as θ_i, but were not included at this point.

The signal and background histograms are used to construct probability density functions $f(x_i)$ via

$$f_s(x_i) = \frac{\nu_i^{sig}}{S\Delta_i} \;\; and \;\; f_b(x_i) = \frac{\nu_i^{bkg}}{B\Delta_i} \tag{19}$$

with ν being the signal (background) value in the i-th bin, Δ being the bin width and $S(B) = \Sigma_i \nu_i^{sig(bkg)}$ being the total number of events. The expected value of the parameter of interest n_i in the data histogram can be expressed in terms of the signal and background histograms as:

$$E[n_i] = \mu s_i + b_i \tag{20}$$

where s_i and b_i represent the mean events from signal and background contributions in the i-th bin and μ is the signal strength parameter, with $\mu = 0$ corresponding to the background-only hypothesis and $\mu = 1$ to the nominal signal hypothesis. The signal and background contributions are calculated by integrating the signal and background probability density functions in the i-th bin and multiplying the result with the total mean number of signal or background events. From the above parameters, the likelihood function is constructed as the product of Poisson probabilities for all bins:

$$L(\mu, \theta) = \Pi_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \tag{21}$$

where n_j is the value in the j-th bin. The measure of incompatibility of the hypothesis and the data is quantified with the likelihood ratio

$$\lambda(\mu) = \frac{L(\mu, \widehat{\widehat{\theta}})}{L(\widehat{\mu}, \widehat{\theta})} \tag{22}$$

The $\widehat{\widehat{\theta}}$ denotes the value of θ for which the likelihood function is maximized with a fixed μ parameter and is therefore a function of μ. The denominator in the above equation is the maximized unconditional likelihood function, with $\widehat{\mu}$ and $\widehat{\theta}$ as the maximum-likelihood estimators.

The first part of the statistical analysis is trying to reject the background only hypothesis. The lower limit for the rejection corresponds to a p-value of 2.87×10^{-7}, where as the p-value necessary to exclude a signal hypothesis is the p-value of 0.05 (95% confidence level). Once it is established that, unfortunately, we have not discovered any new physics, upper limits on the given signal are set. For this, HistFitter constructs an additional test statistic, defined as

$$q_\mu = -2 \ln \lambda(\mu) \; for \; \widehat{\mu} \leq \mu \tag{23}$$

$$q_\mu = 0 \; for \; \widehat{\mu} > \mu \tag{24}$$

Data with $\widehat{\mu} > \mu$ represents less compatibility with μ than the data obtained and is therefore taken out of the rejection region of the test. In the limit of a large-sample [91], Wilk's theorem holds and the probability density function behaves as a χ^2 distribution with one degree of freedom $f(q_\mu \mid \mu) \sim \chi^2$. The p-value can be then calculated as

$$p_\mu = 1 - F(q_\mu \mid \mu) = 2(1 - \Phi(\sqrt{q_\mu})) \tag{25}$$

where Φ is the cumulative distribution of the standard (zero mean, unit variance) Gaussian. The confidence interval at CL=1-α consists of all the values of the parameter that are not rejected. The upper limit on the parameter is the greatest value for which $p_\mu \geq \alpha$. In practice the upper limit on the signal is obtained by setting $p_\mu = \alpha$ and solving for the parameter.

The test-statistic q_μ is constructed to increase monotonically for increasing signal-like (decreasing background-like) experiments. The confidence in the signal+background hypothesis is given by the probability that the the test-statistic is less or equal to the observed value in the experiment:

$$CL_{s+b} = P_{s+b}(q_\mu \leq q_{\mu,obs}) \tag{26}$$

where

$$P_{s+b}(q_\mu \leq q_{\mu,obs}) = \int_{-\infty}^{q_{\mu,obs}} \frac{dP_{s+b}}{dq_\mu} dq_\mu \tag{27}$$

Small values of CL_{s+b} indicate poor compatibility with the signal+background hypothesis and favor the background only hypothesis. Similarly, the confidence on the background hypothesis is given by

$$CL_b = P_b(q_\mu \leq q_{\mu,obs}) \tag{28}$$

and similarly

$$P_b(q_\mu \leq q_{\mu,obs}) = \int_{-\infty}^{q_{\mu,obs}} \frac{dP_b}{dq_\mu} dq_\mu \tag{29}$$

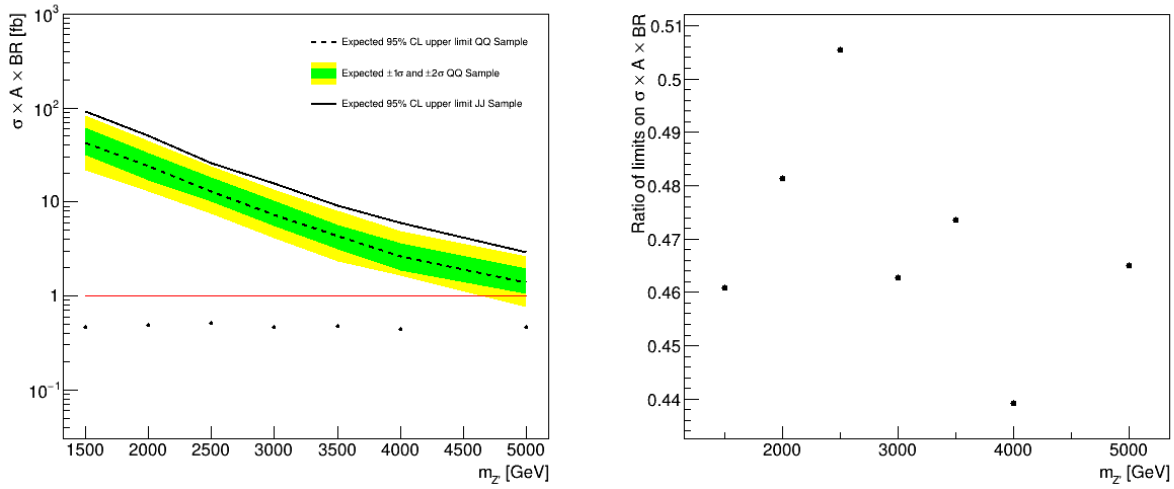
Values of CL_b close to 1 indicate a poor compatibility with the background only hypothesis and favor the signal+background hypothesis. It has been shown in [92] for the case of the Higgs mass, that for large cross-sections, the signal+background and background distributions are well separated and therefore lead to strong confirmations of either hypothesis. However, as the mass of the Higgs increased, the cross-section fell and the overlap of the distributions grew from where one is no longer able to say one of the hypothesis is more strongly supported than the other. A widely accepted solution to this problem is the modified frequentist or CL_s procedure that normalizes the confidence level observed in the signal+background hypothesis, CL_{s+b}, to the cofidence level obtained by the background only hypothesis, CL_b:

$$CL_s \equiv \frac{CL_{s+b}}{CL_b} \tag{30}$$

The CL_s is not a true confidence, it is a ratio of confidences, where the signal hypothesis is considered excluded at the confidence level CL when $1 - CL_s \leq CL$. The consequence of the CL_s method is that the hypothetical false exclusion rate is generally less than the nominal rate of 1-CL, the difference in fact grows with the increase of similarity between the two hypothesis. This avoids a flaw of the CL_{s+b} method that would report a better expected performance of the experiment containing a larger background compared to a similar experiment with the same expected signal [92]. The CL_s method is employed in the HistFitter software framework and used to present the results of this analysis.

5.4 Results

Conducting this analysis on generated data gives a unique opportunity to see what the upper limit on improvement might be, as we have access to the parton's truth ID. The MC generated dijet sample was separated into three sub-samples on an event to event basis solely by reading the final state parton truth ID's. This was then repeated for the signal samples and the individual samples have been run through SWiFt and HistFitter. The limits obtained for the QQ sub-sample are compared to the un-tagged dijet limits and plotted in Figure 36.



(a) (b)

Figure 38: Limit comparison plot of the Z' signal model with $g_q = 0.2$ based on MC simulated data and parton's true ID information. The plot on the left shows the limits on cross section times acceptance times branching ratio for the untagged simulated dijet sample with the solid line, while the dotted line represents the limits obtained for a QQ sub-sample with the green and yellow 2σ variations. Points marked with black dots represent the ratio of tagged and un-tagged data. Any improvement is signified by the ratio lying below the red line. The plot on the right shows the ratio of the untagged dijet sample compared to the QQ sub-sample for all masses of the Z' signal.

Because there are a total of two true GG events in the signal sample with a mass of 3 TeV in comparison to a simulated 450 events at 3 TeV just from the GG sample alone, the increase of efficiency needed for improvement in signal significance is unobtainable. On the other hand, the signal is mostly composed of QQ dijet events and the background is composed of mostly QG and GG events in the lower mass end, improving the limits for this sub-sample is achievable, if the efficiency of the signal does not drop significantly. Of course the improvements shown in the above truth plots are not accessible realistically. In reality implementing a Q/G tagger means that there will be an acceptance and rejection rate for quark and gluon jets. While this means limiting the background, it also means you cut into the signal. To quantify our sensitivity to a future discovery and also the required performance of the Q/G tagger, the signal significance is plotted, and shown

in Figure 37, for various rejection and acceptance values for all the Z' signal masses with $g_q = 0.2$.

$$S_{QQ} = \Sigma_i \frac{N_{QQ_i}\varepsilon_q^2}{\sqrt{B_{QQ_i}\varepsilon_q^2 + B_{QG_i}\varepsilon_q\varepsilon_g + B_{GG_i}\varepsilon_g^2}} \tag{31}$$

In each bin, the number of QQ, QG and GG events is multiplied by the efficiency. For a required 90% acceptance of quark jets (denoted as ε_q), only 81% of the true QQ background sub-sample B_{QQ} will be present. For a 60% gluon rejecton rate $(1 - \varepsilon_g)$, this means that 54% of the true QG sample and 36% of the true GG samples will be tagged as QQ dijet events. Since the Z' signal is mostly composed of the QQ dijet final states, the other two sub-samples in the signal contribute a negligible amount. The ratio is calculated and summed over all background bins. As can be seen from the plot, both values need to be significantly high for improving the significance and that improvement might only be accessible in the lower mass region. This gives the basis for finding the optimal parameter values for the discriminating functions.

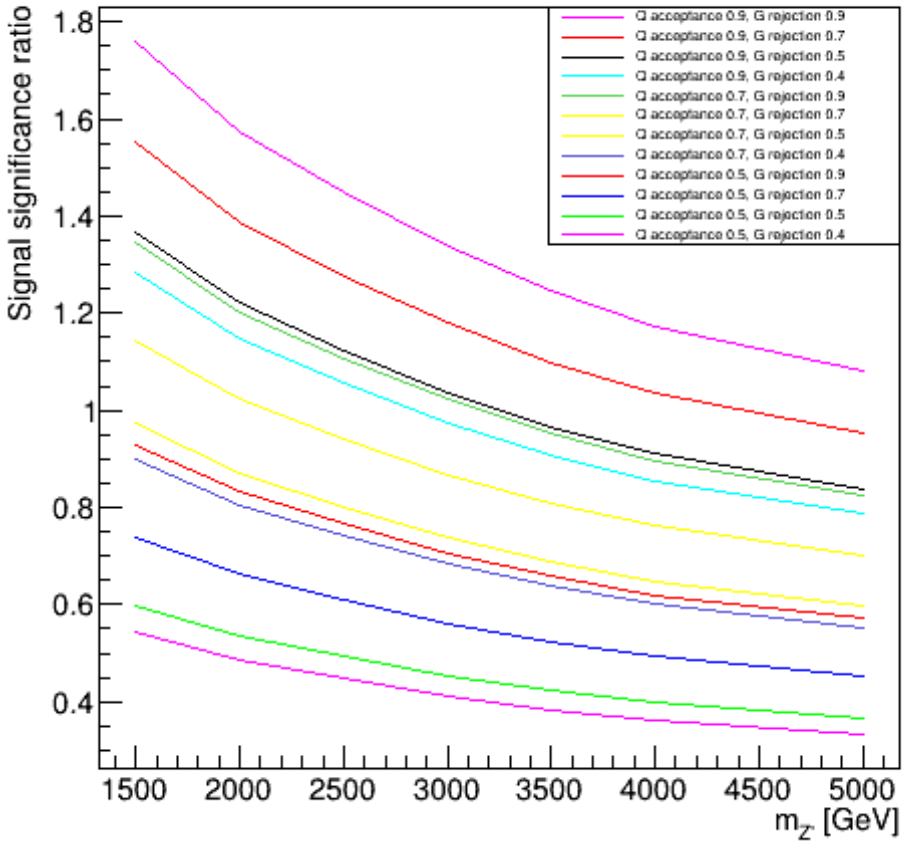


Figure 39: Signal significance as a function of mass for various acceptance and rejection values for quark and gluon jets.

5.4.1 Gaining the starting set of parameters

To determine a set of initial parameters (offset and slope) for both m_{jj} and p_T discriminant functions, the truth information in the MC produced signals with the coupling strength of 0.2 is used as this subset contains the most signals with different masses and provides the opportunity to test the Q/G tagger through a broad mass range. Each combination of offset, slope and function type (log or linear) is labeled as a selection.

In the MC generated signal, the whole set of dijet events is divided into smaller sections in m_{jj}. For a true QQ event in that section, the jet with the higher number of tracks is taken and histograms showing the dependence of the number of jets versus the number of tracks are plotted. This is labeled as a selection $Signal\ selection1 - m_{jj}$. $Signal\ selection2 - m_{jj}$ followed the same procedure, but now both jets in a true QQ event filled the histogram. For functions dependent on p_T, differentiating between higher and lower number of tracks between the two jets in an event does not bring any advantage, so that was omitted and only one selection was gained $Signal\ selection1 - p_T$. At this point, each selection has several histograms depicting the number of jets versus the number of tracks for a invariant mass or p_T range. Every non-zero histogram was integrated with the track range increasing in steps of one to gain the number of tracks for which we keep 90% of jets. This gives a dependence of the minimal track number on the $m_{jj}(p_T)$. These data points were then fitted with linear and logarithmic functions for each signal mass with the 0.2 coupling strength and the mean was taken.

In the data sample, there are more events than just QQ to consider, the data was again divided into smaller ranges of $m_{jj}(p_T)$ and the same three selections as above were gained. Further selections were gained by also using QG and GG events. Data selections 3 and 4 in m_{jj} use true GG events, but selection 3 only uses the jet with the lowest number of jets and the integration procedure yields the number of tracks necessary to reject 90% of jets. $Data\ selection2 - p_T$ similarly uses both jets in a true GG event. Further $m_{jj}(p_T)$ selections use all dijet events, QQ, QG and GG to fill the appropriate histograms, after fitting the optimal rejection and acceptance function, the mean of them is taken. Selections labeled as A are gained by taking the mean from only Q histograms, while selections labeled as B take the mean from considering both Q and G histograms.

This gives the following functions that are either linear and logarithmic in $m_{jj}(p_T)$ with varying offset (n_0) and slope (k) parameters:

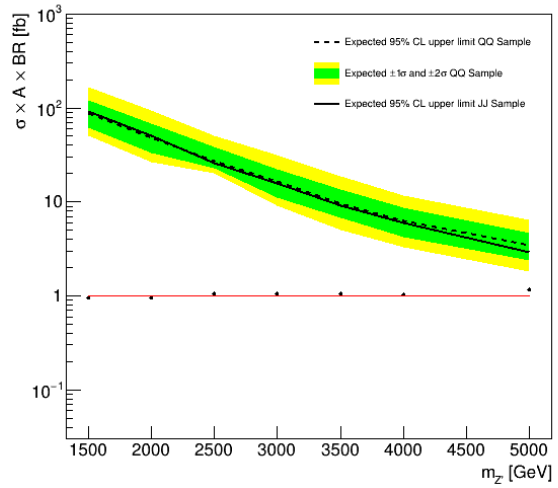
	Selection 1A		Selection 1B	
	Linear	Logarithmic	Linear	Logarithmic
OFFSET	31.7	0.86	25.58	9.16
SLOPE	0.0008	4.25	0.0004	1.98
	Selection 2A		Selection 2B	
	Linear	Logarithmic	Linear	Logarithmic
OFFSET	31.46	-0.57	23.13	4.82
SLOPE	0.0009	4.37	0.0005	2.5
	Selection 3A		Selection 3B	
	Linear	Logarithmic	Linear	Logarithmic
OFFSET	27.84	3.87	23.46	9.26
SLOPE	0.0007	3.36	0.0003	1.74
	Selection 4A		Selection 4B	
	Linear	Logarithmic	Linear	Logarithmic
OFFSET	28.2	5.05	21.5	8.0
SLOPE	0.0006	3.22	0.0004	1.89

Table 3: Table of parameter values for the Q/G separation functions based on dijet invariant mass.

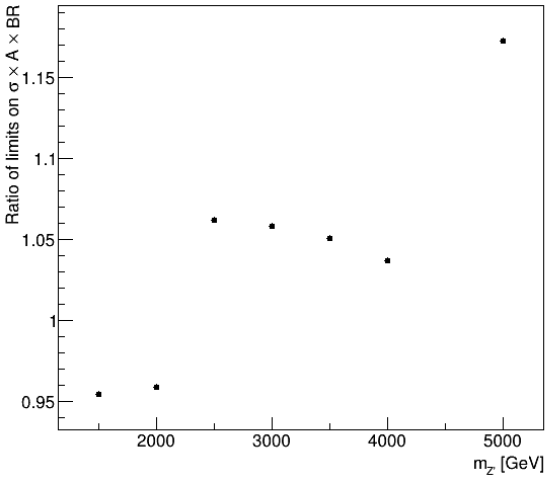
	Selection 1A		Selection 1B	
	Linear	Logarithmic	Linear	Logarithmic
OFFSET	31.53	28.24	24.5	24.25
SLOPE	1.1	0.06	0.000057	0.09
	Selection 2A		Selection 2B	
	Linear	Logarithmic	Linear	Logarithmic
OFFSET	31.42	27.5	24.15	21.45
SLOPE	0.000029	0.16	1.37	0.08

Table 4: Table of parameter values for the Q/G separation functions based on jet p_T.

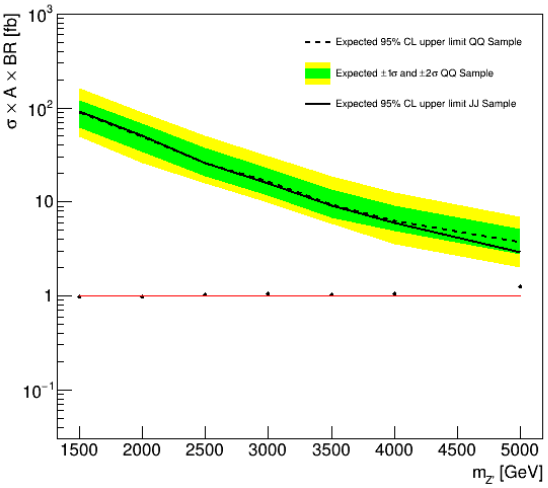
The plots in Figures 38 and 39 show selections with the best limit improvements for both linear and logarithmic functions of $m_{jj}(p_T)$. A logarithmic function models the track multiplicity better and therefore yields a more consistent improvement through the whole range of signal masses, however that means that the improvement is smaller, approximately 2% in comparison to the improvement gained in linear dependency of 5%. A similar thing can be observed in choosing the discriminating function to be dependent on either m_{jj} or p_T. The later calculates the number of tracks independently for both jets and again yields a more consistent improvement through the whole range, while tagging jets depending on the invariant mass, in the majority of times, tags both jets at once, since if the jet with the lower number of tracks is labeled as gluon initiated, so will the other. Therefore the best option for future analysis is a logarithmic dependence on jet p_T in this mass range. However, the linear dependency on m_{jj} looks promising for the mass range below the one analyzed in this analysis.



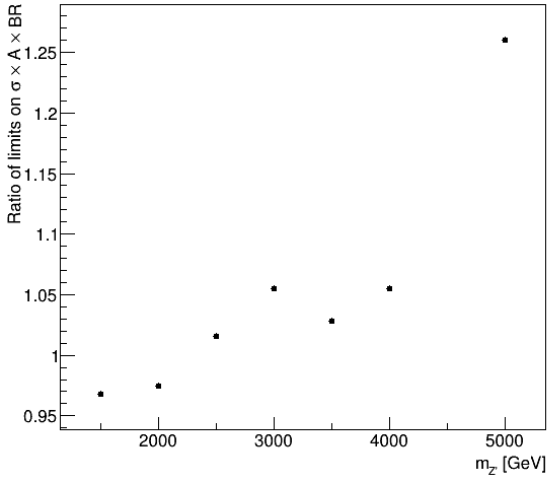
(a)



(b)

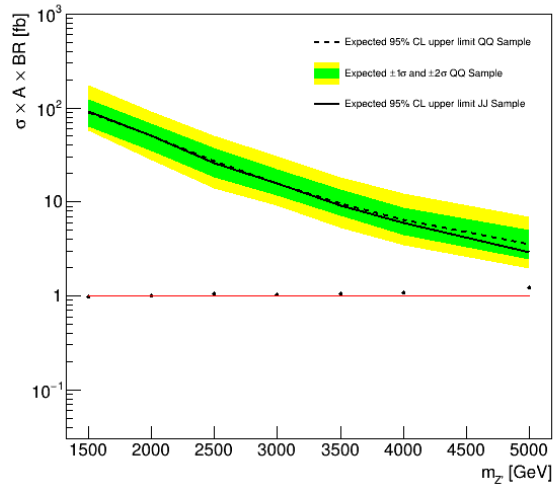


(c)

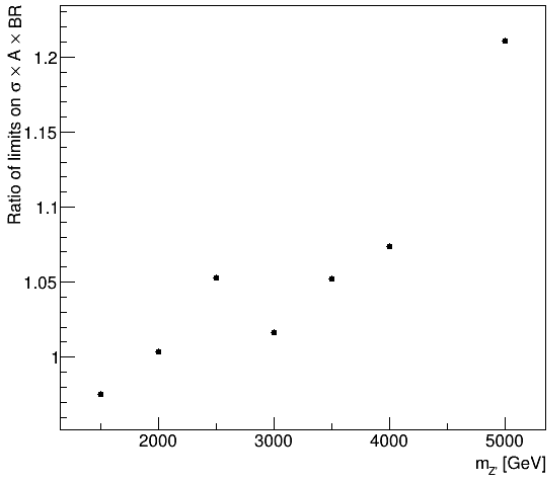


(d)

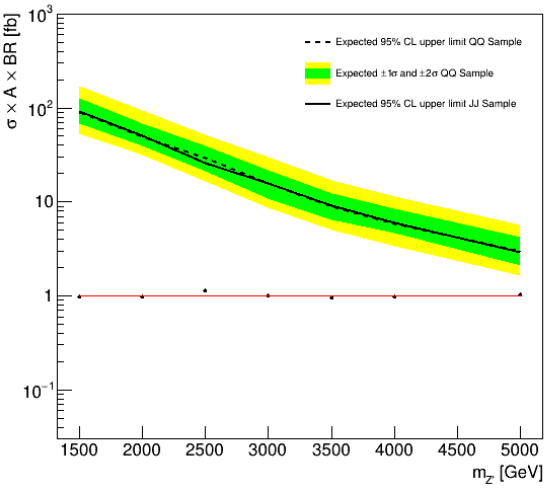
Figure 40: Limits on cross section times acceptance times branching ratio on simulated data and signal Z'. The solid line represents the un-tagged dijet simulated sample limits and the dotted line is the limits obtained using the Q/G tagger based on dijet invariant mass. Points marked with black dots represent the ratio of tagged and un-tagged data. Any improvement is signified by the ratio lying below the red line. The ratio of both limits is plotted both in the lower part of the figure on the left (limit improvement is below the red line) and a detailed plot is shown in figured on the right. The first two plots are obtained using a Q/G separation function based on linear dependency on m_{jj}, while the lower two plots use a logarithmic function.



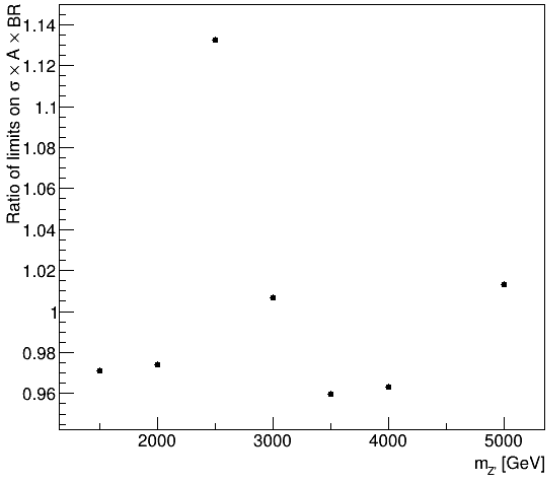
(a)



(b)



(c)



(d)

Figure 41: Limits on cross section times acceptance times branching ratio on simulated data and signal Z'. The solid line represents the un-tagged dijet simulated sample limits and the dotted line is the limits obtained using the Q/G tagger based on jet p_T. Points marked with black dots represent the ratio of tagged and un-tagged data. Any improvement is signified by the ratio lying below the red line. The ratio of both limits is plotted both in the lower part of the figure on the left (limit improvement is below the red line) and a detailed plot is shown in figured on the right. The first two plots are obtained using a Q/G separation function based on linear dependency on p_T, while the lower two plots use a logarithmic function.

5.5 Uncertainties

No systematic uncertainties have been applied to the results at this stage, however it is necessary to approximate them in order to see if the limit improvement by Q/G jet tagging is significant enough.

5.5.1 Signal modeling

There are two potential impacts on Q/G tagging on the signal model: signal shape variation and efficiency. The figure below shows a comparison of the JJ signal to the QQ signal gained through tagging. The comparison of signal shapes with and without Q/G tagging shows good agreement for the Z' prime signal and does not noticeably affect the signal resonance shape.

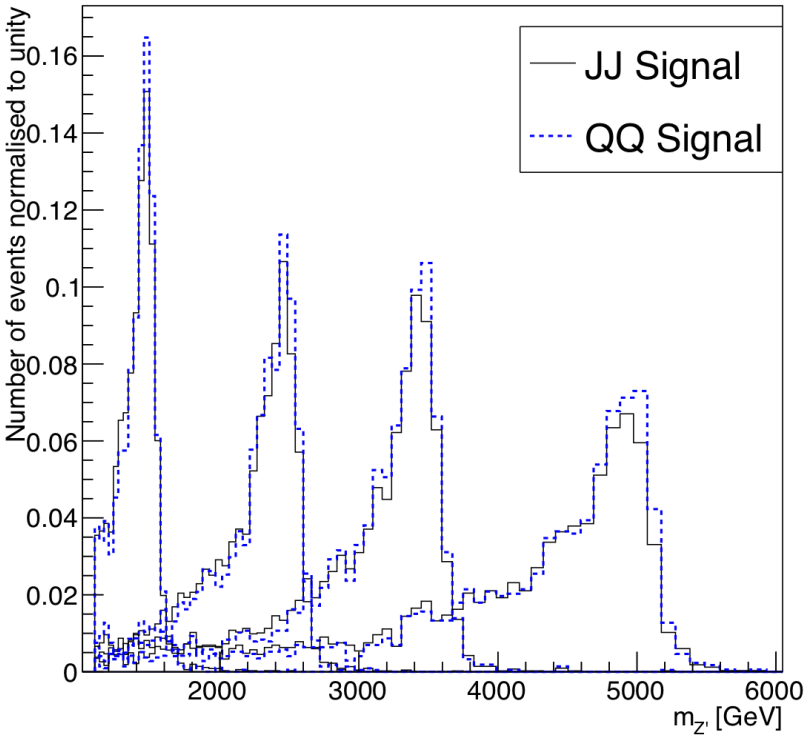


Figure 42: Signal shape comparison for a tagged and untagged signal for varying masses. The histograms are normalized to unity.

Another impact that the implemented tagger can bring is the change of signal efficiency. The below plot shows a comparison of a tagged and truth signal efficiency. A high and steady efficiency across the whole mass range is achieved and therefore should not impact the analysis significantly.

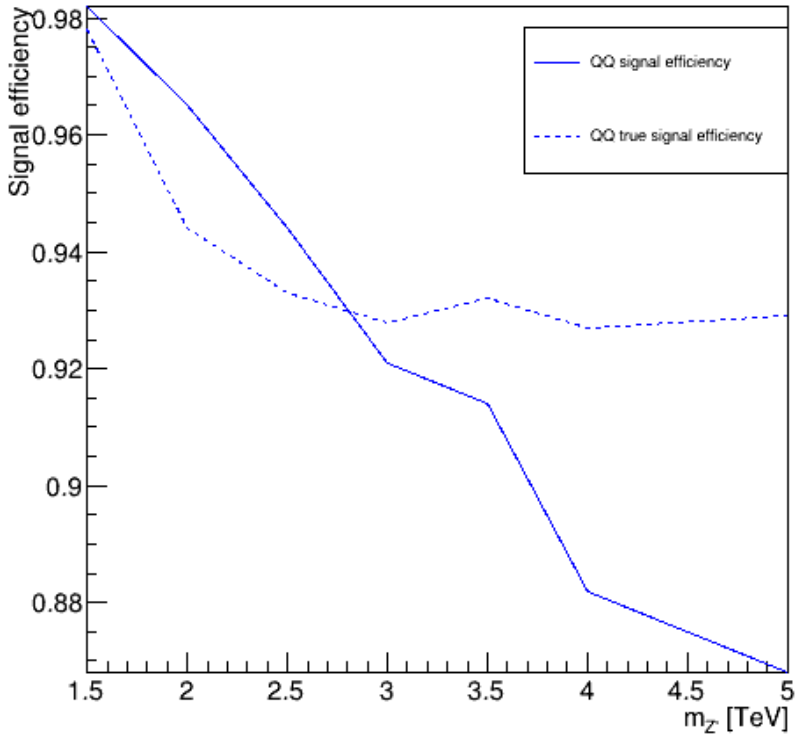


Figure 43: Plot comparing efficiencies of tagged and untagged signals with the 0.2 coupling strength and varying masses.

5.5.2 Systematic uncertainties

A luminosity uncertainty is applied as a scale factor to the normalization of signal samples. For the combined Run-2 dataset, a 1.7% luminosity uncertainty is used. An additional 1% flat systematic is applied to account for the PDF and scale uncertainties.

The additional systematics and uncertainties are calculated and will be applied at the HistFitter stage of the analysis. The systematic uncertainties gained through a data-driven background estimation are the fit function uncertainty and the uncertainty of the parameters of the fitting function. The former is calculated on a bin to bin basis and is non-symmetrical. A collection of pseudodata is thrown from the nominal background result and from each a nominal and alternate background estimations are obtained. The mean difference between the nominal and alternative background fits are recorded in each bin to define the size of the function choice uncertainty. For the $139fb^{-1}$ Run-2 data the uncertainty ranges from far below 1% to the maximum of 2% at the highest m_{jj} bins. The second uncertainty is associated with the quality of the fit itself. From the nominal background fit to data, a large number of pseudodata sets are generated using Poisson statistics and each fitted with the same starting conditions as the observed data. The error of the fit is recorded in each bin and defines the root mean square of the function value in that bin for all

the pseudo-experiments.

The Jet Energy Scale uncertainty varies with m_{jj} and is obtained by fitting the benchmark signal with a Gaussian and it's $\pm 1\sigma$ variation. The shifts of the Gaussian are summed in quadrature to define a single nuisance parameter. For the JES a set of 7 nuisance parameters is used to fit the signal templates. The procedure is repeated for the Jet Energy Resolution uncertainty with 8 nuisance parameters. All the four vectors of all the jets are the shifted by 1σ in the JES uncertainty and the result is fitted to determine the shift in peak location as a percentage of signal mass. To give an approximation, both of the effects were calculated for an excited quark signal model, presented in the ATLAS Exotics publication note and were found to be less than 1%.

To make a rough approximation, any improvement lower than 5% would be overshadowed by the uncertainties.

6 Conclusion

In this thesis, the implementation of a Q/G tagger is tested on a MC simulated data sample mimicking the 2015 and 2016 Run-2 collected data with a centre-of-mass energy of 13 TeV. From the collected data, dijet events with invariant mass above 1.1 TeV were picked to produce a smooth invariant mass spectrum. Firstly, the required acceptance and rejection rates for individual quark or gluon initiated jets were tested to obtained the desired performance of the Q/G tagger. It was shown that a minimum of 60% acceptance and 70% rejection rate is needed for improving the significance compared to the whole dijet analysis. This was used to analyze the staring parameters in linear and logarithmic functions of both jet p_T and m_{jj}, which were then used to divide the dijet sample into three sub-samples: QQ, QG and GG. In each dijet event, the tagger is used to label both jets independently as either quark or gluon initiated, based on the number of tracks in the jet and the calculated number of tracks from the discriminating function. The newly obtained sub-samples were run through SWiFt and HistFitter to obtain limit plots on the cross section times acceptance times branching fraction of the signal. The limits were then compared to the full dijet sample limits. It was shown that the logarithmic function of jet p_T works significantly well across the whole mass range resulting in a steady improvement of about 2%, while functions dependent on the invariant mass might present limit improvements in ranges lower than 1.1TeV as the limits improved in the range of 5% in the lower mass end, but did not show any improvement in the higher mass range.

It was shown that we can expect large improvements in signal significance with a precise definition of quark and gluon initiated jets, however a more complex tagger will be needed in achieving them. Since the Q/G tagger based on track multiplicity keeps the signal efficiency high, it presents a good starting point for further analysis.

References

[1] *M. Herrero, "The Standard model", NATO Sci. Ser. C* 534 *(1999) 1 doi* : 10.1007/978 − 94 − 011 − 4689 − 0_1 *[hep-ph/9812242].*

[2] *Standard Model of Particle Physics. General Photo. URL: https://www.symmetrymagazine.org/standard-model/*

[3] *G. Aad et al., [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. B* **716** *(2012) 1 doi:10.1016/j.physletb.2012.08.020 [arXiv:1207.7214 [hep-ex]].*

[4] *Gell-Mann, M.: Phys. Lett. 8, 214 (1964); Zweig, G.: CERN TH 401 and 412 (1964); Greenberg, O.W.: Phys. Rev. Lett. 13, 598 (1964)*

[5] *Particle Data group, Yao, W.-M., et al.: J. Phys. G 33, 1 (2006)*

[6] *D. Griffiths, "Introduction to elementary particles", Weinheim, Germany: Wiley-VCH (2008) 454 p*

[7] *Particle data Group, G. Dissertori et al.* Quantum chromodynamics. *URL: http* : *//pdg.lbl.gov/*2014/$reviews/rpp$2014 − rev − *qcd.pdf*

[8] *Altarelli, Guido et al., Nucl.Phys. B126 (1977) 298-318 LPTENS-77-6*

[9] *S. Forte et al., Nucl. Phys. B602, 585 (2001).*

[10] *M. Anselmino et al., Z. Phys. C64, 267 (1994).*

[11] *J.D. Bjorken and E.A. Paschos, Phys. Rev. B72, 195 (1977).*

[12] *R.P. Feynman, Photon Hadron Interactions (Benjamin, New York, 1972).*

[13] *M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018).*

[14] *L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, "Parton distributions in the LHC era: MMHT 2014 PDFs", Eur. Phys. J. C* **75** *(2015) no.5, 204 doi:10.1140/epjc/s10052-015-3397-6 [arXiv:1412.3989 [hep-ph]].*

[15] S. Gieseke, "Event generators: New developments", hep-ph/0210294.

[16] ATLAS Collaboration. Baryonic Z' Summary Plot. URL:*http* : *//atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/EXOT* − 2013 − 11/$figaux_1$0.*png*

[17] J.D. Bjorken, Phys. Rev. 179, 1547 (1969).

[18] V.C.Rubin and J.Ford, W.Kent, Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions, Astrophys. J. 159 (1970) 379.

[19] V.C.Rubin,N.Thonnard,andJ.Ford,W.K.,Rotational properties of 21SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/, Astrophys. J. 238 (1980) 471.

[20] Dark and Luminous Matter in Bright Spiral Galaxies, p2. S. Kassin. Ohio State University, Ph.D Dissertation (2004).

[21] A.Bosma, 21-cm line studies of spiral galaxies.II. The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types., Astron. J. 86 (1981) 1825.

[22] T.S. van Albada, J.N.Bahcall, K.Begeman, and R.Sancisi, Distribution of dark matter in the spiral galaxy NGC 3198, Astrophys. J. 295 (1985) 305.

[23] Dark Matter: A Primer, p1. K. Garrett, G. Duda. Advances in Astronomy (2011), Vol. 2011.

[24] The Virial Theorem in Stellar Astrophysics, pp8-10. G.W. Collins. Pachart Publishing House(1978). Made available by NASA ADS at http://ads.harvard.edu/books/1978vtsa.book/.

[25] F.Zwicky, Die Rotverschiebung von extragalaktischen Nebeln, Helv.Phys.Acta6(1933).

[26] F.Zwicky, On the Masses of Nebulae and of Clusters of Nebulae, Astrophys.J.86(1937)217.

[27] Bekenstein, J., "The modified Newtonian dynamics — MOND and its implications for new physics", Contemp. Phys., 47, 387–403, (2006). [DOI], [ADS], [arXiv:astro-ph/0701848].

[28] D. C. Dai, R. Matsuo and G. Starkman, "Gravitational Lenses in Generalized Einstein-Aether theory: The Bullet Cluster", Phys. Rev. D **78** (2008) 104004 doi:10.1103/PhysRevD.78.104004 [arXiv:0806.4319 [gr-qc]]

[29] Penzias, A.A., and Wilson, R.W., 1965 ApJ, 142, 419

[30] John C. Mather and Gary F Hinshaw (2008) Cosmic background explorer. Scholarpedia, 3(3):4732.

[31] N. Aghanim *et al.* [Planck Collaboration],"Planck 2018 results. VI. Cosmological parameters", arXiv:1807.06209 [astro-ph.CO].

[32] S. D. McDermott, H.-B. Yu, and K. M. Zurek, Turning off the Lights: How Dark is Dark Matter?, Phys. Rev. D 83 (2010) 063509, [arXiv:1011.2907].

[33] C.Kouvaris,CompositeMillichargedDarkMatter,Phys.Rev.D88(2013)015001, [arXiv:1304.7476].

[34] A.D.Dolgov,S.L.Dubovsky,G.I.Rubtsov,andI.I.Tkachev,Constraintsonmillicharged particles from Planck, Phys. Rev. D 88 (2013) 117701, [arXiv:1310.2376].

[35] E. Del Nobile, M. Nardecchia, and P. Panci, Millicharge or Decay: A Critical Take on Minimal Dark Matter,[arXiv:1512.05353].

[36] A. and G. Sharma, B. C. Chauhan, "Dark matter and Neutrinos", arXiv:1711.10564

[37] M. Beltran, D. Hooper, E. W. Kolb, and Z. C. Krusberg, Deducing the nature of dark matter from direct and indirect detection experiments in the absence of collider signatures of new physics, Phys.Rev. D80 (2009) 043509, [0808.3384].

[38] M. Beltran, D. Hooper, E. W. Kolb, Z. A. Krusberg, and T. M. Tait, Maverick dark matter at colliders, JHEP 1009 (2010) 037, [1002.4137].

[39] M. Chala, F. Kahlhoefer, M. McCullough, G. Nardini and K. Schmidt-Hoberg, "Constraining Dark Sectors with Monojets and Dijets", JHEP **1507** (2015) 089 doi:10.1007/JHEP07(2015)089 [arXiv:1503.05916 [hep-ph]].

[40] LHCb Collaboration, R. Aaij et al., Angular analysis of the $B^0 \longrightarrow K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity, JHEP 02 (2016) 104, [arXiv:1512.04442].

[41] LHCb Collaboration, R. Aaij et al., Angular analysis and differential branching fraction of the decay $B_s^0 \longrightarrow \phi\mu^+\mu^-$, JHEP 09 (2015) 179, [arXiv:1506.08777].

[42] LHCb Collaboration, R. Aaij et al., Differential branching fractions and isospin asymmetries of $B \longrightarrow K^\star\mu^+\mu^-$ decays, JHEP 06 (2014) 133, [arXiv:1403.8044].

[43] Belle Collaboration, S. Wehle et al., Lepton-Flavor-Dependent Angular Analysis of $B \longrightarrow K^\star l^+ l^-$, Phys. Rev. Lett. 118 (2017), no. 11 111801, [arXiv:1612.05014].

[44] M. Aaboud *et al.* [ATLAS Collaboration], "Constraints on mediator-based dark matter and scalar dark energy models using $\sqrt{s} = 13$ TeV pp collision data collected by the ATLAS detector", JHEP **1905** (2019) 142 doi:10.1007/JHEP05(2019)142 [arXiv:1903.01400 [hep-ex]].

[45] M. T. Frandsen, F. Kahlhoefer, A. Preston, S. Sarkar, and K. Schmidt-Hoberg, LHC and Tevatron Bounds on the Dark Matter Direct Detection Cross-Section for Vector Mediators, JHEP 1207 (2012) 123, [1204.3839].

[46] O. Buchmueller, M. J. Dolan, and C. McCabe, Beyond Effective Field Theory for Dark Matter Searches at the LHC, JHEP 1401 (2014) 025, [1308.6799].

[47] O. Buchmueller, M. J. Dolan, S. A. Malik, and C. McCabe, Characterising dark matter searches at colliders and direct detection experiments: Vector mediators, JHEP 1501 (2015) 037, [1407.8257].

[48] P. Harris, V. V. Khoze, M. Spannowsky, and C. Williams, Constraining Dark Sectors at Colliders: Beyond the Effective Theory Approach, Phys.Rev. D91 (2015), no. 5 055009, [1411.0535].

[49] D. Abercrombie et al., Dark matter benchmark models for early LHC Run-2 searches: Report of the ATLAS/CMS Dark Matter Forum, arXiv:1507.00966.

[50] B. Holdom, Two $U(1)$'s and Epsilon Charge Shifts, Phys.Lett. B166 (1986) 196.

[51] K. Babu, C. F. Kolda, and J. March-Russell, Implications of generalized $Z - Z'$mixing, Phys.Rev. D57 (1998), [hep-ph/9710441].

[52] E. Dudas, Y. Mambrini, S. Pokorski, and A. Romagnoni, (In)visible Z' and dark matter, JHEP 0908 (2009) 014, [0904.1745].

[53] P. J. Fox, J. Liu, D. Tucker-Smith, and N. Weiner, An Effective Z', Phys.Rev. D84 (2011) 115006, [1104.4127].

[54] V. Martin-Lozano, M. Peiro and P. Soler, Isospin violating dark matter in Stuckelberg portal scenarios, 1503.01780.

[55] A. Alves, S. Profumo, and F. S. Queiroz, The dark Z portal: direct, indirect and collider searches, JHEP 1404 (2014) 063, [1312.5281].

[56] G. Arcadi, Y. Mambrini, M. H. G. Tytgat, and B. Zaldivar, Invisible Z' and dark matter: LHC vs LUX constraints, JHEP 1403 (2014) 134, [1401.0221].

[57] O. Lebedev and Y. Mambrini, Axial dark matter: The case for an invisible Z, Phys.Lett. B734 (2014), [1403.4837].

[58] XENON Collaboration, E. Aprile et al., Dark Matter Search Results from a One Ton-Year Exposure of XENON1T, Phys. Rev. Lett. 121 (2018), no. 11 111302, [arXiv:1805.12562].

[59] M. Aaboud *et al.* [ATLAS Collaboration], "Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s}$ =13 TeV with the ATLAS detector", Phys. Rev. D **96** (2017) no.5, 052004 doi:10.1103/PhysRevD.96.052004 [arXiv:1703.09127 [hep-ex]].

[60] Fabienne Marcastel. CERN's Accelerator Complex. La chaÃ®ne des accrateurs du CERN. Oct 2013. General Photo. URL: https://cds.cern.ch/record/1621583

[61] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. JINST, 3:S08003, 2008, doi:10.1088/1748-0221/3/08/S08003

[62] T. Akesson *et al.*, "Particle identification using the time-over-threshold method in the ATLAS Transition Radiation Tracker", Nucl. Instrum. Meth. A **474** (2001) 172. doi:10.1016/S0168-9002(01)00878-6

[63] E. Daw, Lecture 7 - Rapidity and Pseudorapidity,23 March 2012. [Online]. Available: http://$www.hep.shef.ac.uk/edaw/PHY206/Site/2012_course_files/phy206rlec7.pdf$

[64] G. Aad *et al.*, ATLAS pixel detector electronics and sensors, JINST 3 (2008) P07007

[65] H. Abat *et al.*, The ATLAS TRT Barrel Detector, JINST 3 (2008) P02014.

[66] E. Abat *et al.*, The ATLAS TRT end-cap detectors, JINST 3 (2008) P1003.

[67] F. Ceradine, in Proceedings of the XXIII International Confer ence on High Energy Physics, Berkeley, California, 1986, edit- ed by Stewart C. Loken (World Scientific, Singapore, 1987), Vol. II, p. 1051

[68] R. M. Harris and K. Kousouris, Searches for dijet resonances at hadron colliders, Int. J. Mod. Phys. A 26 (2011) 5005, arXiv: 1110.5302 [hep-ex].

[69] J. E. Huth *et al.*, "Toward a standardization of jet definitions", In *Snowmass 1990, Proceedings, Research directions for the decade* 134-136 and Fermilab Batavia - FERMILAB-Conf-90-249 (90/12,rec.Mar.91) 6 p. (105313)

[70] S. Ellis, Z. Kunszt, and D. Soper, Phys. Rev. D. 40 2188 (1989).

[71] G. Aad *et al.* [ATLAS Collaboration], "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1", Eur. Phys. J. C **77** (2017) 490 doi:10.1140/epjc/s10052-017-5004-5 [arXiv:1603.02934 [hep-ex]].

[72] [CMS Collaboration], "A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging", CMS-PAS-JME-09-001.

[73] G. P. Salam,Towards Jetography, Eur. Phys. J. C **67** (2010) 637 doi:10.1140/epjc/s10052-010-1314-6 [arXiv:0906.1833 [hep-ph]].

[74] M. Cacciari, G. P. Salam and G. Soyez, The anti-k_t jet clustering algorithm, JHEP **0804** (2008) 063 doi:10.1088/1126-6708/2008/04/063 [arXiv:0802.1189 [hep-ph]].

[75] R. Bala, "An introduction to Pythia: The Event Generator", URL: *http* : *//people.du.ac.in/ pmehta/FinalSem/Rockyfinal.pdf*

[76] T. Sjostrand, "Monte Carlo Generators", hep-ph/0611247.

[77] M. H. Seymour and M. Marx, "Monte Carlo Event Generators", $doi : 10.1007/978-3-319-05362-2_8$ arXiv:1304.6677 [hep-ph].

[78] T. Sjöstrand, S. Mrenna and P. Z. Skands, A brief introduction to PYTHIA 8.1, Comput. Phys. Commun. 178 (2008) 852, arXiv: 0710.3820 [hep-ph]

[79] R. D. Ball et al., Parton distributions with LHC data, Nucl. Phys. B 867 (2013) 244, arXiv: 1207.1303 [hep-ph]

[80] ATLAS Collaboration, The ATLAS Simulation Infrastructure, Eur. Phys. J. C 70 (2010) 823, arXiv: 1005.4568 [physics.ins-det]

[81] S. Agostinelli et al., GEANT4: A Simulation toolkit, Nucl. Instrum. Meth. A 506 (2003) 250

[82] S. Höche, "Introduction to parton-shower event generators", $doi : 10.1142/9789814678766_0005$ arXiv:1411.4085 [hep-ph].

[83] The ATLAS collaboration,Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector, ATLAS-CONF-2015-029.

[84] H. P. Nilles and K. H. Streng, Phys. Rev. D 23, 1944 (1981) doi:10.1103/PhysRevD.23.1944; L. M. Jones, Phys. Rev. D 39, 2550 (1989) doi:10.1103/PhysRevD.39.2550; Z. Fodor, Phys. Rev. D 41, 1726 (1990) doi:10.1103/PhysRevD.41.1726; L. Jones, Phys. Rev. D 42, 811 (1990) doi:10.1103/PhysRevD.42.811.

[85] L. Lonnblad, C. Peterson and T. Rognvaldsson, Nucl. Phys. B 349, 675 (1991) doi:10.1016/0550-3213(91)90392-B; H. Luo, M. x. Luo, K. Wang, T. Xu and G. Zhu, arXiv:1712.03634 [hep-ph].

[86] M. S. Alam et al. [CLEO Collaboration], Phys. Rev. D 46, 4822 (1992) doi:10.1103/PhysRevD.46.4822.

[87] G. Alexander et al. [OPAL Collaboration], Phys. Lett. B 388, 659 (1996) doi:10.1016/S0370-2693(96)01319-6.

[88] P. Abreu et al. [DELPHI Collaboration], Phys. Lett. B 449, 383 (1999) doi:10.1016/S0370-2693(99)00112-4 [hep-ex/9903073].

[89] ATLAS Collaboration. Quark versus Gluon Jet Tagging Using Charged-Particle Constituent Multiplicity with the ATLAS Detector. "ATL-PHYS-PUB-2017-009", Available: "http://cds.cern.ch/record/2263679"

[90] M. Baak, G. J. Besjes, D. Cô̂te, A. Koutsman, J. Lorenz and D. Short, "HistFitter software framework for statistical data analysis", Eur. Phys. J. C **75** (2015) 153 doi:10.1140/epjc/s10052-015-3327-7 [arXiv:1410.1280 [hep-ex]].

[91] G. Cowan, K. Cranmer, E. Gross and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", Eur. Phys. J. C **71** (2011) 1554 Erratum: [Eur. Phys. J. C **73** (2013) 2501] doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z [arXiv:1007.1727 [physics.data-an]].

[92] A. L. Read, "Presentation of search results: The CL(s) technique", J. Phys. G **28** (2002) 2693. doi:10.1088/0954-3899/28/10/313