Search for High-Mass Dijet Resonances Decaying into Top Quark Pairs using Machine Learning Techniques in the ATLAS Experiment

A thesis submitted for the degree of
Master of Science

Stefio Yosse Andrean

Supervisors:
Torsten Åkesson
Trine Poulsen

LUNDS UNIVERSITET

Particle Physics Division
Department of Physics
Lund University, Sweden

Abstract

This thesis presents an application of machine learning techniques in a search for a high-mass resonance particle decaying into top quark pairs in high energy dijet events using the ATLAS experiment at the Large Hadron Collider. Top tagging is applied to the dijet events to select the events with top signature and suppress the background, increasing the sensitivity of the search. All the data used in this work come from Monte Carlo simulations.

Performance studies are carried out to compare four boosted top taggers available in the ATLAS framework: a conventional 2-variable tagger, two jet substructure-based machine learning tagger using Boosted Decision Tree (BDT) and Deep Neural Network (DNN), and the Topocluster Tagger which uses a DNN to process the kinematics of jets' topoclusters. It is shown that the three machine learning taggers are capable of suppressing more background than the conventional 2-variable tagger by roughly a factor of two at 80% constant signal efficiency. The Topocluster Tagger is chosen to be applied to the dijet mass distribution to be analyzed.

The effect of the tagging is studied by performing Sliding Window Fit (SWIFT) resonance search method to the distribution before and after top tagging. The method scans the dijet mass distribution in the range of 1100 - 6787 GeV, with the assumed integrated luminosity of 100 fb^{-1}. The search is conducted on two distributions: background-only distribution, and signal-injected distribution. The 95% Confidence Level limit plots show an increase in the sensitivity of the search on background-only distribution. This is further confirmed in the signal-injected case where the method manages to pick up significant signal after top tagging, but not before.

Popular Abstract

One of the main goals of the experiments in the Large Hadron Collider is to find phenomena beyond what the Standard Model can explain. There are many theories predicting what phenomena beyond the Standard Model could be circulating in the theory community, and it is the job of the experimentalists to find the truth of these theories. And as you may have already guessed, it is not so easy to find something that can not be explained by the strongest theory in physics.

The challenge in searching for something that exciting is that it happens very rarely (if it doesn't, we would have already found it!). Even in the case that it happens, it is hidden in the abundance of the other not-so-exciting stuff. It is like trying to find a golden egg in a swamp of mud that is full of brown eggs, at night, with only a dim flashlight in your hand. Even if you find it, it would look like a brown egg and chances are, you would throw it back to the mud. So the challenge is, how does one develop a method to distinguish the golden egg and the brown egg? Or to say it in a more technical term: to discriminate the *signal* from the *background*.

In high energy particle collisions, particles are flying out of the collision point, created by transforming energy into mass, following Einstein's $E = mc^2$. Some of these particles cannot exist on their own, so new particles are created from the vacuum to bind with them. The detector would see this as a cone-shaped spray of particles coming from the collision point, called a *jet*. Some theories describing what lies beyond the Standard Model predict that there will be new undiscovered particles that would decay into jets with some particularity. This is why we use jets as our search object – our eggs.

To look for the jets we are interested in, physicists usually use quantities called the *jet substructure variables*. They are variables that describe some particular properties of a jet. Knowing the predicted value of these variables for the signal, we can use that value to make a cut around it to narrow our search. In our egg-finding analogy, if we predict the mass of the golden egg to be m, we can develop an algorithm like, "if the egg is lighter than $m - p$, or heavier than $m + p$, discard it!", where p is how tight you want your search algorithm to be. Make it too loose, you would accept too many brown eggs; make it too tight, you would lose some of the precious golden eggs. One can combine the algorithm with other variables, say, the shininess or the shape of the egg. A good search algorithm is the one that can reject as much background, while at the same time, keeping as much signal as possible.

Now, this is where the machine beats us: while humans can only process a limited number of variables, machine learning algorithm can learn from all the variables there are! Machine learning can extract the information contained in all of the variables and conclude whether the jet is more signal-like or background-like. In the studies in this thesis, machine learning techniques have been shown to perform roughly twice as good compared to a conventional method.

The implementation of machine learning in high energy physics will provide more powerful tools than the ones used in the past searches. These newly acquired tools will allow us to see hidden events that were previously undetectable and increase our chance for new discovery.

We make our world significant by the courage
of our questions and the depth of our answers.

– Carl Sagan, *Cosmos*

Acknowledgements

This thesis is a product from hands and brains of countless people. From physics and coding on the weekdays to fikas and beers on the weekends, those people had contributed in one way or another to the completion of this work. To list everyone on this single piece of paper would be a hopeless endeavor, but I will try my best to mention a few.

Torsten, my supervisor – thank you for all the knowledge you have passed on to me, and the kindness you have shown. Every meeting with you had been so uplifting, I always came out feeling better than when I came in. I have learned so much from you, not only in physics, but also in wisdom and humility. I really look up to you.

Trine; my co-supervisor, friend, and slave-master – thank you for being always available whenever I need help with whatever. If it wasn't for you, I would still be trying to solve a compilation error by now. I've made the other Master students envious because I can always rely on you to help me. I hope the next person you supervise will have as much fun as I did.

Caterina – you are the person I go to when everything else fails. Thank you for always having an answer. I really admire how good you are at what you do, and I strive to at least get close to that. And thank you for getting me a project to do at CERN, I am really grateful for it.

Sam, Stephen, and Ece; mentors from the summer at CERN – the two months I spent at CERN was the best learning experience I ever had in my life. If there is a plot for the amount of new knowledge per second vs my living years, that summer would be the peak. You have my gratitude.

Matteo and the Dijet Group – thank you for being a great audience for every progress-update presentations. The feedbacks, tips, and solutions have been invaluable to the progression of this thesis and also my development as a researcher.

Karishma – this thesis would not get passed Chapter 4 if it wasn't for your package. The support you gave me was more than I deserve. I hope I will be able to do the same to others in the future. Thanks a bunch!

The colleagues at Lund ATLAS Group: Simon, Zhiying, Katja, Eva, Eric, Eleni, (the other) Caterina, Will, Alex, Shi, Saeid, Isabelle, Emelie, Jessamy, and Leonie – thank you for the random talks at the office, both science-related and not. You guys bring so much life to this workplace. The teachers that I got to learn from: Oxana, Else, Anders, Balazs, Florido, and Vytautas – a big part of my development comes from your lectures, thank you.

Silvia and Dimitrios – they say you make good friends at strange places, but I never thought it would be at the Quantum Field Theory lecture. Thank you for the lunches, dinners, and endless lectures on Eurovision. I wish us the best of luck in our next chapter in life and I will make sure we meet again. Maybe, Advanced QFT?

The Indonesian community in Lund – you guys bring the warmth of home thousands of kilometers away to this cold town. Terima kasih.

I owe the opportunity to even be here to the Indonesian Endowment Fund for Education. I per-

sonally believe that geographical boundary and financial condition should not be a limitation to a person's education. Thank you for turning that idea into a reality.

Naris, my beloved – thank you for being patient and trusting. I am very grateful to have you at my side on this journey, I cannot imagine it without you.

And to Mom, Dad, and sis – my dream has always been too strange and difficult to understand, but you trust me anyway. We knew from the start that the path I have chosen is not an easy one, but surprisingly, it has been working out pretty well so far. Thank you for the unyielding love and support, it does not go unnoticed. This one is for you, I hope it makes you proud.

Thesis Overview

Chapter 1 reviews some of the preliminary knowledge that would be required to follow through the text. It begins with a brief summary of the Standard Model (SM), and continues with the speculations surrounding Beyond Standard Model (BSM) Physics. The BSM theories with relevance to the topic are described, supplying the motivation for the work. The search focuses on heavy resonances decaying into top quark pairs. The LHC and the ATLAS experiment are the scientific facilities used in this thesis, and are therefore briefly explained here. Jets as the main object of study are discussed at the end of the chapter.

Chapter 2 explores the boosted top tagging tools that are studied and used in this thesis. It starts with the nature of boosted top quarks, the decay channel that is considered, and the general strategy to capture them. It proceeds with the details of four boosted top taggers: Smoothed Top Tagger, DNN Tagger, BDT Tagger, and Topocluster Tagger. How the taggers operate and how they were developed are explained here. These taggers were developed and trained by the experts of ATLAS.

Chapter 3 takes the four boosted top taggers explained in the previous chapter and test them to see how they compare to each other in their performance. The Monte Carlo samples used are given, along with the event selections applied to them. The taggers are tested by assigning them in two tasks: tagging individual jets and tagging dijet events. The results are shown here.

Chapter 4 explains how to obtain the dijet mass distributions used in the analysis. Since the search is to be sensitive to top quarks, the QCD samples used in the previous chapter needs to be enriched with the SM $t\bar{t}$ productions. The SM $t\bar{t}$ distributions are discussed and added to the QCD dijet mass distributions. The procedure to derive the data-like distributions from the Monte Carlo is explained at the end of the chapter.

Chapter 5 explores the Sliding Window Fit (SWIFT) resonance search method which is the analysis method used in this thesis. The algorithms and calculations done inside the method are briefly explained in the chapter, together with the value of the parameters used in this particular search. The development of SWIFT was done by the experts of ATLAS and not part of this project.

Chapter 6 displays the results of the SWIFT resonance search method when applied the dijet mass distribution. To get a view of the effect of the top tagging to the quality of the search, the analysis is done on the dijet mass distribution before and after top tagging. Discussions on the obtained results are written in the chapter.

Chapter 7 summarizes everything done and learned in this thesis. It also gives an outlook on what the future project can do to improve the work done in here.

Figures with a citation on the captions are taken from outside sources. Those without citation on the captions are original products of this thesis.

Contents

1 Introduction

This chapter guides the reader to some of the basic concepts that are needed in order to proceed to the later chapters while it also gives the motivation behind the work. It starts by briefly describing the Standard Model (SM) of particle physics, followed by what might be hiding in the Beyond Standard Model (BSM) physics. Next, the main experimental instrument in particle physics, the Large Hadron Collider, is introduced. The ATLAS detector is covered to give some understanding of particle detection techniques. Lastly, jets as the main object of the study in this thesis are explained from the physics that originates them to how to construct them from the information given by the detector.

1.1 Standard Model of Particle Physics

Physics, in its attempt to probe the most fundamental nature of the universe, mainly follows the reductionism approach: study the small parts to understand the big ensemble. Started from the ancient Greek by the philosopher Democritus, the atomic theory which says that all matter is made of indivisible discrete units has lead thinkers – ancient and modern alike – to pursue this fundamental constituent of matter that makes up the whole universe. After series of theoretical breakthroughs and experimental discoveries, a collective theory had been developed to encompass all of the known fundamental particles and the forces between them: the Standard Model of particle physics.

Written in the framework of quantum field theory, the Standard Model classifies the fundamental particles based on the quantum numbers they possess [1]. Each particle is uniquely characterized by a set of quantum numbers: mass, spin, and charges. Based on the spin, the particles are divided into two major groups: half-integer spin – *fermions*, and integer spin – *bosons*. The charge determines the type of interaction a particle is subjected to; electric charge for electromagnetic interaction, and color charge for strong interaction. Figure 1.1 shows all the known fundamental particles in the Standard Model and their classifications. In addition to this, there also exists a mirror image of the particles in the Standard Model: the *antiparticles*. They are identical in all regards except that they carry opposite charge sign.

Inside the fermion group, particles are grouped further based on the charges they carry. Fermions that carry color charge are called *quarks* and those do not are called *leptons*. Together, quarks

and leptons take up the role as building blocks of higher structure of matter.

Possibly the most known elementary particle – the *electron* (e) belongs to the charged lepton group, together with its siblings: *muon* (μ) and *tau* (τ). Each of them carries -1 electric charge. Alongside the charged leptons, there also exist the neutral leptons. The *neutrino* (Italian for "little neutral one") was first hypothesized by Wolfgang Pauli in 1930 to answer the missing energy of neutron decays. Neutrons decay into protons by emitting electrons. Careful measurements seem to indicate that the energy of the final products is always less than what it started with. So Pauli suggested there is an invisible neutral particle in the final products that carry that missing energy: the neutrino ν[1]. There are three flavors of neutrino: electron neutrino ν_e, muon neutrino ν_μ, and tau neutrino ν_τ; each forms a doublet with its corresponding charged lepton partners.

Quarks come in six flavors: up (u), down (d), charm (c), strange (s), top (t), and bottom (b). The up, charm, and top have $+\frac{2}{3}$ electric charge and the down, strange, and bottom have $-\frac{1}{3}$ electric charge; they are commonly referred to as "up-type" quarks and "down-type" quarks, respectively. Unlike leptons, quarks carry color charge, which means each of them represents triplet of particles, cleverly named after colors: red, green, and blue[2]. Because of the nature of Quantum Chromodynamics (QCD), quarks are always confined in a "colorless" bound state called *hadrons*; they may not exist by themselves in isolation. For example, a proton – one of the bound state of quarks – is composed of three quarks: two up quarks, and one down quark; each of them can be red, green, and blue as long as they combined into a "white" proton.

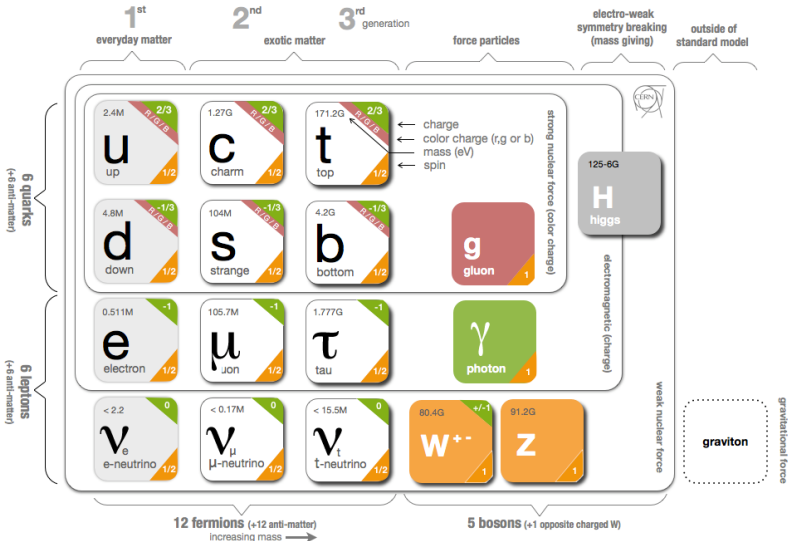


Figure 1.1: Diagram of the Standard Model of particle physics [2].

Bosons take up the role of force mediators. There are four fundamental forces in nature: *strong*, *electromagnetic*, *weak*, and *gravity*; each of them has a boson (or bosons) that is responsible to mediate them. The *strong* force is mediated by gluons, the electromagnetic force by photons, and the weak force by the neutral Z bosons and a pair of charged $W^\pm$ bosons. As of the writing of this thesis, the Standard Model has no explanation for gravity. But it is natural to think that gravity works just like the other forces: carried by a boson, hence the hypothesized *graviton*. Due to its weak strength relative to the other forces, the absence of gravity does not have a

[1] Later it was realized that it is anti-neutrino that plays a role in the neutron decays.

[2] Do not confuse these with actual color. Quarks are too small to be visible, let alone to have color. Color in the literal sense is a property of photon, a whole different particle!

considerable effect on the prediction of the Standard Model in particle physics and can be safely neglected. Table 1.1 lists the four fundamental forces with the corresponding carrier bosons and their relative strength.

Theorized in the 1960s and discovered in 2012, the *Higgs* is somewhat different from the other bosons. The Higgs boson is a product from a mechanism that breaks the electroweak symmetry called the Brout-Englert-Higgs (BEH) mechanism. The BEH mechanism breaks the electroweak symmetry by splitting the bosons of the original symmetry into the massless photon and the massive W and Z bosons, breaking apart the electroweak interaction into electromagnetic and weak interaction. For this to work, the theory requires a field that permeates the whole universe and has non-zero value everywhere: the Higgs field. The stronger a particle couples to the Higgs field, the more mass it gains. Quantum field theory dictates that particles are ripples of fields, so the Higgs boson itself is nothing but a ripple in the Higgs field, just like all other elementary particles are ripples to their corresponding fields.

Table 1.1: The four fundamental forces and the corresponding carriers with their relative strength [3].

Force	Carrier boson	Relative strength
Strong	gluon	1
Electromagnetic	photon	$\frac{1}{137}$
Weak	$W^{\pm}$, Z	10^{-6}
Gravity	graviton[3]	10^{-39}

1.2 Beyond Standard Model Physics

The Standard Model is currently the best theoretical model that describes nature at the most fundamental level. Its predictions agree with every experiment conducted at the currently accessible energy scale. However, it still has not captured the whole picture. Gravity is yet to be incorporated in the framework of the Standard Model; and Dark Matter, which supposedly makes up 27% of our universe [4], is also missing from the table. Furthermore, looking at the diagram of the Standard Model in Figure 1.1, one cannot help but notice a pattern in the arrangement of the particles. The fermions are organized in three generations, each shares exact properties with the other two except differing in mass, which is a free parameter in the Standard Model. There is no explanation on why the particles take the value they have, they are just experimentally measured. This seemingly hidden structure and arbitrariness of the Standard Model hints to a deeper underlying physics that lies beyond.

One of the unsolved problems in particle physics is the *hierarchy problem*. The source of this problem is the quadratically divergent corrections to the Higgs mass. To be consistent with the observation of the Higgs boson mass which is found to be around 125 GeV, it requires an incredible fine-tuning that makes the theory seems unnatural. The problem can also be recognized in the strange huge gap of strength between gravity and the other forces. Many BSM theories offer solutions to the hierarchy problem, e.g Super Symmetric theories, extra dimensions theories, dynamic electroweak symmetry breaking theories; in which many of these predict a heavy particle that couples to SM particles. The high mass of the top quark thus motivates the search in this channel because of the high branching ratio.

[3]The existence of graviton has not been experimentally confirmed.

One possible explanation for the observed weakness of gravity is that there is a hidden extra dimension in spacetime and the force of gravity gets diluted in it. The Randall-Sundrum (RS) Model [5] assumes that the four-dimensional spacetime is just a part of a five-dimensional spacetime with a fifth extra dimension being "warped". The RS model predicts Kaluza-Klein (KK) excited particles as a manifestation of the SM fields that propagate through the extra dimension. The KK particles share the same quantum numbers as the original particles, except heavier in mass [6]. Among other excited states, the KK excited gluons have the largest cross-section and preferentially decay into top quark pairs (92.6% branching ratio [7]).

The Heavy Vector Triplet framework [8] offers a search strategy to look for narrow resonances in a model-independent way. A Heavy Vector Triplet is a set of three high mass spin-1 bosons: a pair of charged bosons, and one neutral boson. The charge eigenstates are superpositions of the field eigenstate V_μ^a with $a = 1, 2, 3$, written as,

$$V_\mu^\pm = \frac{V_\mu^1 \mp i V_\mu^2}{\sqrt{2}}, \text{and} \tag{1.1}$$

$$V_\mu^0 = V_\mu^3 \tag{1.2}$$

Since we are only concerned with top-antitop final states and only neutral charge particles can decay into top pairs due to charge conservation, only the neutral boson V_μ^0 is used in the analysis. Throughout this thesis, the neutral boson will be referred to as HVT. The Z' – a heavier copy of the SM Z boson [9] – is also used as an alternative signal model.

1.3 The Large Hadron Collider

Located 100 m below ground at the border of Switzerland and France, the Large Hadron Collider (LHC) is a 27 km circumference double-ring tunnel where protons and lead ions are accelerated and smashed together. The concept was first made public in a workshop by CERN in Lausanne, Switzerland in 1984; and after years of planning, constructions, and much political effort, the LHC stands today as the largest and most powerful particle accelerator.

Based on what is being accelerated, there are different physics programs conducted at the LHC: proton-proton, lead-lead, and proton-lead collisions. As of 2018, the LHC is operating at a center-of-mass energy ($\sqrt{s}$) of 13 TeV. This thesis focuses on proton-proton collisions at $\sqrt{s}$ = 13 TeV; other types of collision are not discussed further.

The LHC utilizes CERN's older accelerator facilities to accelerate protons from their stationary state in multiple stages. The protons start their flight in LINAC2 (Linear Accelerator 2) where they are accelerated from zero velocity to a third of the speed of light before entering the PS (Proton Synchrotron). From there, the protons go through the SPS (Super Proton Synchrotron) to be further boosted before finally transferred into the LHC ring. The particle journey from the starting point to the LHC can be seen in Figure 1.2.

The double-ring construction of the LHC was meant so that the two beams are accelerated in opposite direction. There are four points where these two rings inter-cross and become the points of collisions for the accelerated particles. At these points, four major particle detectors are placed, each specialized to look for particular physics. LHCb (LHC beauty) specializes in studying the decay of b quarks. ALICE (A Large Ion Collider Experiment) is designed to study

heavy ion collisions to investigate the quark-gluon plasma – a hot and dense state of matter that filled the universe shortly after the Big Bang. ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) are general-purpose detectors dedicated to observe collision events and determine what comes out of the collisions. These two experiments lead the energy frontier of experimental particle physics today and played a central role to one of the largest achievement of particle physics – the discovery of the Higgs boson.

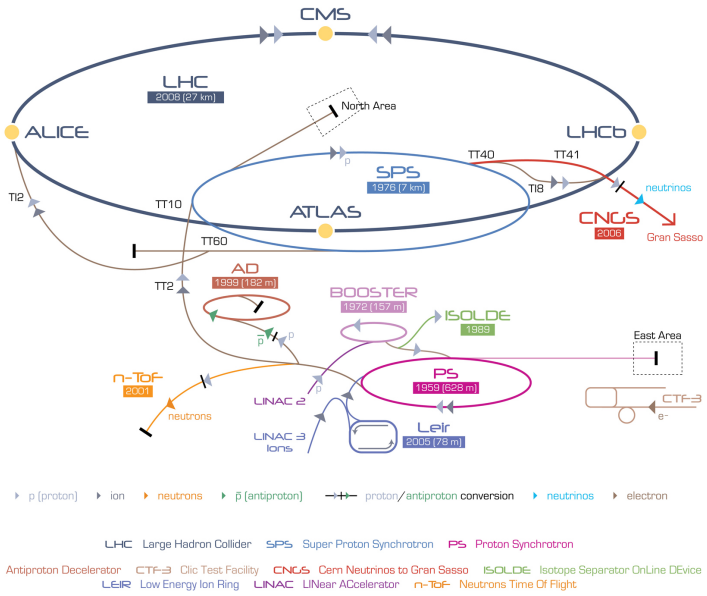


Figure 1.2: Diagram of CERN's accelerator complex [10].

1.4 The ATLAS Detector

1.4.1 The Design

ATLAS, shown here by Figure 1.3, is a 25m high and 44m long particle detector placed in one of the collision points of the LHC. The aim of the ATLAS detector is to capture as much information as possible out of particle collision events: identify the outgoing particles and measure their properties. To achieve this, ATLAS was designed to encapsulate the collision point with layers of detectors, each specialized in detecting a certain type of particle. Any particles that are produced at the collision points will travel outward, passing through the layers of detectors until they are stopped or they escape the detector.

Closest to the beam line is the inner detector. Consisting of three sub-detectors: Pixel detector, Semiconductor Tracker (SCT), and Transition Radiation Tracker (TRT), the inner detector is responsible to track the path of the passing charged particles and reconstruct the point they originate from. Being the closest to the collisions, the inner detector is where the most precise measurements with the finest granularity are performed. The inner detector is contained within a strong solenoidal magnet (2 Tesla central field) that provides magnetic field parallel to the beam axis. It utilizes the magnetic field to measure the momentum of charged particles and determine the charge. The momentum is calculated from how much the particles curve due to the magnetic field[4].

[4]The particle's momentum is calculated by using the formula $\vec{F} = q\vec{v} \times \vec{B}$; and the sign of the charge can be

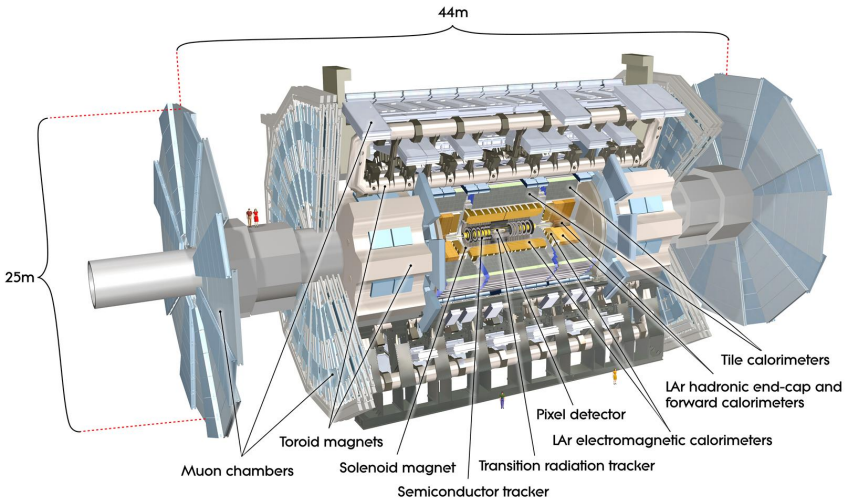


Figure 1.3: Computer generated view of the ATLAS detector [11].

Moving outward, the calorimeter is a sub-detector of ATLAS whose job is to measure the energy of the outgoing particles. The passing particles interact with the nucleus of the calorimeter's materials, creating a shower of particles whose energy can be measured and reconstructed back to the initial energy. The calorimeter consists of hundreds of thousands of fine granular cells to give the needed spatial resolution of particle detection. The calorimeter is radially divided into two parts: electromagnetic and hadronic calorimeter. The electromagnetic calorimeter catches particles that interact electromagnetically, e.g. electrons and photons. Strongly interacting particles passes through the electromagnetic calorimeter, to be later captured by the hadronic calorimeter. Some fraction of the time, strongly interacting particles starts showering in the electromagnetic calorimeter and continue showering to the hadronic calorimeter, in which case the measured energy is summed over the two calorimeters to get the correct energy. The calorimeter is important for the topic of this thesis since it is from the energy of the decay particles that mass of the mother particle is calculated. The jet tagging technique that is discussed in the later chapter also uses energy signature deposited in the calorimeter cells to determine the original particles.

At the outmost layer lies the muon spectrometer. Muons, like electrons, interact with the electromagnetic and weak force. But muons are much heavier; electromagnetically, it is more likely to create electrons than muons, so muons mostly come from the weak interaction. Most of the particles that are created in the collisions are stopped in the calorimeter, only muons pass through the inner detector and the calorimeters without losing significant energy and can reach the muon spectrometer, which is why it is specifically named after the particle. The muon spectrometer consists of gas tubes. When muons pass through the gas tubes, they will ionize the gas, releasing electrons and ions which are collected in the cathode and anode of the tube as signals. This method allows the tracking of the muons' path. The muon spectrometer is permeated by magnetic field from the toroidal magnet system. This is so that momentum of the muons can be measured by the bending of the curved track, similar to what is done in the inner detector.

Not all of the particles can be detected by ATLAS. Neutrinos only interact via weak interaction, which makes them very difficult to catch. This elusiveness forces physicists to think outside

deduced from the direction of the curve: positively charged particles curve according to the right-hand rule, while the negatively charged particles curve according to the left-hand rule.

the box when it comes to detecting neutrinos. Unlike other particles, neutrinos are detected not by detecting the energy signature, but by detecting the absence of it. At the LHC, the colliding protons only have longitudinal momentum along the beam line. Collisions always start with zero transverse momentum, and conservation law dictates that they should end with zero net transverse momentum. So in the events where conservation of transverse momentum is not satisfied, it can be deduced that neutrinos are involved[5]. Figure 1.4 shows the cross-sectional view of the ATLAS detector and how different types of particles interact with the sub-detectors.

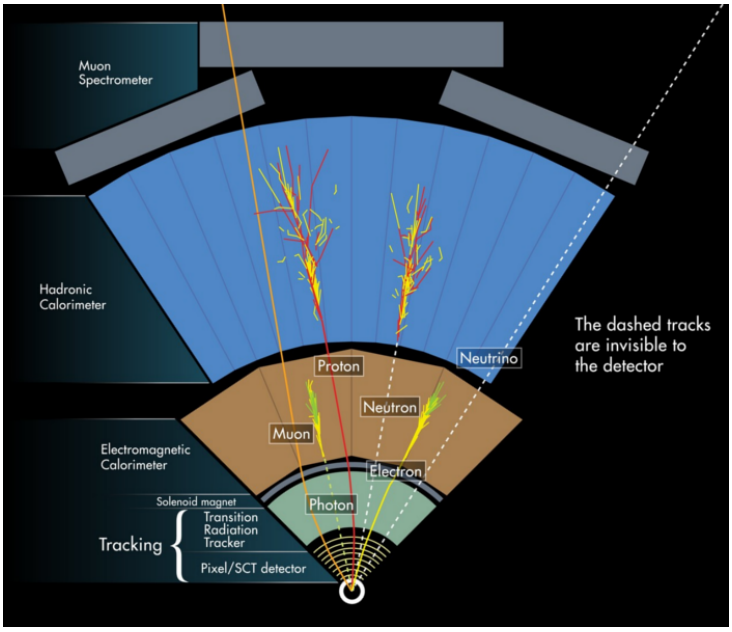


Figure 1.4: Cross-sectional view of the ATLAS detector [12].

1.4.2 The Coordinate System

As shown by Figure 1.5, ATLAS uses a right-handed coordinate system with its origin at the collision point, the positive z-axis points along the beam line, the positive x-axis points to the center of the LHC ring, and the positive y-axis points upward. A spherical coordinate system is also used with ϕ being the azimuthal angle, and θ being the polar angle. The physical observables of particles are based on these coordinates.

One observable that is found often in this thesis is the *rapidity* y, defined as,

$$y \equiv \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) \tag{1.3}$$

where E is the particle's energy and p_z is the longitudinal momentum along the beam axis. Rapidity describes how forward a particle is traveling relative to the beam line. The rapidity is preferred over the polar angle θ to describe forwardness because the difference in rapidity Δy is invariant under Lorentz boost, and also because particle production is independent of rapidity.

One can make an approximation of rapidity by substituting energy E with modulus of the

[5] or, it could be a sign of a new exotic particle with invisible signature.

three-momentum $|\mathbf{p}|$. This approximation is called *pseudorapidity*, defined as,

$$\eta \equiv \frac{1}{2}\ln(\frac{|\mathbf{p}|+p_z}{|\mathbf{p}|-p_z}) = -\ln(\tan\frac{\theta}{2}) \tag{1.4}$$

Notice that at the limit of high energy or in the case of massless particles, pseudorapidity converges with rapidity. Pseudorapidity ranges from $-\infty$ to $+\infty$, corresponds to particles going completely backward and forward along the beam line; and $\eta = 0$ (or $y = 0$) corresponds to particles moving perpendicular to the beam line.

Another important observable is the transverse momentum p_T. It is the modulus of the particle's momentum projected to the transverse (xy) plane. Since the collisions always initiate with purely longitudinal momenta, the transverse momentum must come from the particle interaction during the collisions, which can describe the momentum transfer of the interaction.

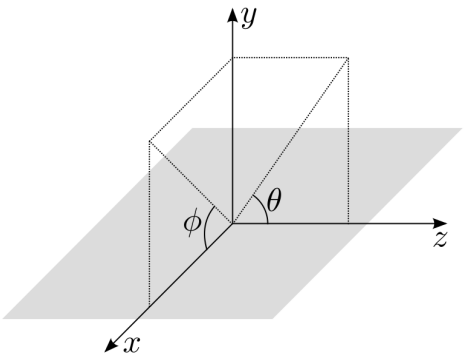


Figure 1.5: The coordinate system used in the ATLAS experiment. The origin is the collision point, the positive z-axis points along the beam line, the positive x-axis points to the center of the LHC, the positive y-axis points upward. ϕ and θ are the azimuthal angle and the polar angle, respectively [13].

1.5 Jets

1.5.1 Quantum Chromodynamics and Color Confinement

QCD is the physics of the strong interaction. One of the uniquenesses of QCD compared to QED is that the force-carrying boson, gluon, can interact with themselves. Figure 1.6 shows the color force lines between a quark and an anti-quark. Just like in QED, the number of force line only depends on the total charge; in this case, color charge, so it remains constant. Due to the gluon self-coupling, the cross-sectional area A too remains constant. Imagine a thought experiment where we can pick the quark and anti-quark by hand and try to separate them. According to the QCD law stated previously, the force lines would stretch with the distance r, just like rubber bands. Thus, the potential energy of two strongly interacting particles, here illustrated by the volume of the flux tube, increases approximately linearly with the separation between them. Consequently, the quark and anti-quark are *confined* as it takes infinite amount energy to separate them. This phenomenon is called the *color confinement*, it prevents experimentalists to observe a stand-alone quark.

The bound state is called a *hadron*. Used in the previous example, a *meson* is a hadron consists

of two constituents: a quark and an anti-quark; and a *baryon* is a hadron made of three quarks. Examples of baryons are protons and neutrons which make up the nucleus of atoms.

When protons collide, their constituent quarks and gluons (partons), interact with each other by exchanging energy. Interactions are happening in the form of momentum transfers between partons from the two colliding protons, moving the colliding partons away from the other partons in their respective protons. In the case of highly energetic collisions, the momentum transfers are high, causing the colliding partons to be strongly accelerated. Since partons carry color charges, they then emit radiation in the form of gluons; just like accelerated electric charges would radiate photons. These emitted gluons may pair-create quark-antiquarks who in their turn also could radiate gluons. The resulting cascades of gluons and quark-antiquark pairs from the interacting partons, are called *parton showers*.

As the distance between the initial bound quarks increase, the potential energy increases due to color confinement; and if it reaches high enough energy, the force string is broken and a pair of quark and antiquark is created, as shown by Figure 1.7. The new antiquark is then bound to the initial quark by a new string, and similarly, the new quark is bound with the initial antiquark. If the energy suffices, the new strings stretch and break again, and so on, until all of the energy is converted into the creation of quark-antiquark pairs, which are recognized as hadrons. Such is the *hadronization* process as modelled by the Lund String Model [14]. Note that this also occurs to the gluons in the parton shower as they also carry color charge and therefore bound by color strings. In the detector frame, an observer would see the accumulation of these processes as a cone-shaped spray of particles going to the same general direction called a *jet*.

Jets become important because they are what the detector sees. When one wants to look for (say) a b quark in the final state, one would look for a b jet in the detector. It is because of this reason that jets become the main object of study for the search in this thesis.

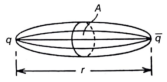


Figure 1.6: Force lines of the strong interaction between a quark and an anti-quark [4].

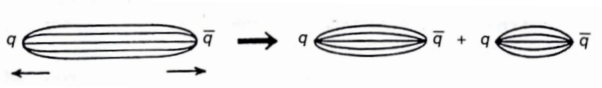


Figure 1.7: String breaking process from a quark-antiquark initial state. The quark from the initial bound state gained momentum and move away from the other quark it was bound with, increasing the potential energy. If the energy is high enough, the force string breaks and quark-antiquark pair is produced. [4]

1.5.2 Topoclustering

To find jets, one starts to look at the energy deposited in the calorimeter. As mentioned before, the calorimeter is an assembly of hundreds of thousands of fine granular cells which absorb and measure energy. To be able to reconstruct physical information about the incoming particles, neighboring cells need to be clustered by a connecting algorithm. These clustered cells are called a *topocluster*. The properties of a topocluster are summed from the constituent cells, resulting in the direction, location, and energy of the cluster. Note that individual topoclusters

do not necessarily contain a full energy of a single particle. Individual topoclusters can contain a full or fractional energy of a single particle, a combined energy of multiple particles, or a full energy of one particle plus some fraction of another particle shower [15].

The clustering algorithm collects the calorimeter signals based on the topology of the calorimeter cells in any given collision event. The goal is to be able to distinguish a localized signal from background electrical noise and pile-up[6]. The clustering starts from the cell that has the highest signal significance and grows by collecting the adjacent cells. The cell significance $\varsigma^{\mathrm{EM}}_{\mathrm{cell}}$ is defined as the ratio of the cell signal $E^{\mathrm{EM}}_{\mathrm{cell}}$ to the average background noise $\sigma^{\mathrm{EM}}_{\mathrm{cell}}$,

$$\varsigma^{\mathrm{EM}}_{\mathrm{cell}} = \frac{E^{\mathrm{EM}}_{\mathrm{cell}}}{\sigma^{\mathrm{EM}}_{\mathrm{cell}}} \tag{1.5}$$

the superscript EM means that the quantities are measured in the electromagnetic (EM) energy scale, which measures correctly the energy deposited by electrons and photons. ATLAS calorimeter is "non-compensating", meaning, it does not correct the signal loss for hadrons. This signal loss can be recovered by further calibration by first determining the hadronic or EM character of the particle shower based on the cluster shapes (depth, location, lateral shapes, etc.) and calibrate it accordingly [16].

The topoclustering algorithm works as follows,

1. Find cells with $\varsigma^{\mathrm{EM}}_{\mathrm{cell}} > 4$ and order them in decreasing $\varsigma^{\mathrm{EM}}_{\mathrm{cell}}$, these cells are called *primary seeds*.

2. The cells neighboring a seed with $\varsigma^{\mathrm{EM}}_{\mathrm{cell}} > 2$ are collected by the seed along with their neighbor, forming a *proto-cluster*.

   - If the neighboring cell is a primary seed, the two proto-clusters are merged.

   - If the neighboring cell passes $\varsigma^{\mathrm{EM}}_{\mathrm{cell}} > 2$ and also belongs to a different proto-cluster, the two proto-clusters are merged.

3. Repeat until the neighboring cells only have signal significance $0 < \varsigma^{\mathrm{EM}}_{\mathrm{cell}} < 2$.

Figure 1.8 shows the result of the topoclustering algorithm in a simulated dijet event[7]. These topoclusters act as inputs to construct the jets.

1.5.3 Jet-finding algorithm

The topoclusters are then put into a machinery called the *jet-finding algorithm*. The purpose of the algorithm is to determine a set of energy deposits that are localized in $\eta - \phi$ space that resembles a cone of collimated particles. There are a number of algorithms that have been developed to do this task, most basically follow the same idea: calculate the distance between two objects and determine if they belong to the same jet. Similar to the topoclustering algo-

[6] Inside accelerators, protons are being collided in bunches, which means every recorded event consists of multiple interactions. In most cases, physicists are only interested in the most energetic collisions in each event, other softer interactions are regarded as "pile-up".

[7] A dijet event is an event that features two jets.

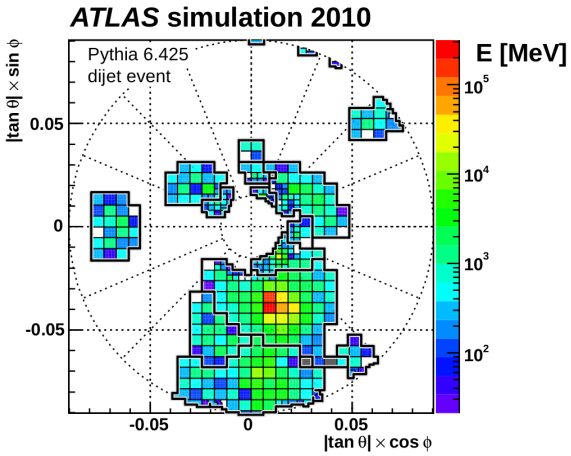


Figure 1.8: Result of the topoclustering algorithm in a simulated dijet event. [15]

rithm, the jet-finding algorithms start with a primary object which iteratively grows in size by including nearby objects.

For a pair of two topoclusters i and j, the relative distance between the two, and relative distance of the object i and the beam d_{iB} are defined as,

$$d_{ij} = \min(p_{T_i}^{2p}, p_{T_j}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \tag{1.6}$$

$$d_{iB} = p_{T_i}^{2p} \tag{1.7}$$

where p_{T_i} is the transverse momentum of object i, $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is the distance between object i and j in $y - \phi$ space, and R is the radius parameter. The choice of p determines the starting object for the jet-finding: $p = 1$ starts the jet-finding at the object with smallest p_T, $p = -1$ starts at the highest p_T object, and $p = 0$ disregards transverse momentum altogether and only looks at angular separation. These different choices of p are named k_T algorithm [17], *anti-k_T* algorithm [18], and *Cambridge/Aachen* algorithm [19], respectively.

The jet-finding algorithm works as follows,

1. For each pair of object i and j, calculate d_{ij}; and for each object i, calculate d_{iB}.

2. Find the minimum $d_{\min}$ of all the d_{ij} and d_{iB}.

   - If $d_{\min} = d_{ij}$, merge i and j into a proto-jet.

   - If $d_{\min} = d_{iB}$, remove i from the list and declare it to be part of the "beam" jet.[8]

3. Repeat until no objects are left.

Moving forward, the definition of a jet departs from a collimated spray of particles, to a set of objects which the jet-finding algorithms determine as a jet, which, if the different jet-finding algorithms work ideally, should be the same.

[8] This follows the "exclusive" formulation of the jet-finding algorithm. In the "inclusive" formulation, i is added to the list of the final inclusive jets.

1.5.4 Dijet Events and Resonance Search

In searches looking for resonances decaying into two quarks, it is instructive to look at events with two jets, or dijet events. The four-vector of the dijet can be obtained by summing the four-vectors of the jet pair, and the invariant mass of the dijet system m_{jj} is calculated as follows[9],

$$m_{jj} = \sqrt{E_{jj}^2 - p_{jj}^2} \tag{1.8}$$

where E_{jj} and p_{jj} are the energy and momentum of the dijet four-vector. Aside from being a direct measure of the energy reach of the experiment, the dijet invariant mass is an important observable because it corresponds to the mass of the mother particle that decays into the two jets. The mother particle could be an undiscovered particle that is produced via resonance that could hint to new physics.

The QCD predicts a smoothly falling invariant mass distribution. Should there be a new unstable particle that decays into two jets, the particle would add to the total cross section that corresponds to two jets final state. Therefore, the presence of a new resonance particle can be observed as a localized excess, or *bump*, in the mass distribution. Looking for this bump is the core principle of resonance searches.

[9]The formula follows a consensus in the field of particle physics to set the speed of light and the Planck constant equal to one, $c = \hbar = 1$. This allows all observables to be expressed in the unit of electron volt (eV).

2 Boosted Top Tagging

The top tagging tools studied in this project are covered in detail in this chapter. The main principle of boosted top tagging, how the taggers work, how they were developed, and the variables they use are explored here. Four taggers are covered: the Smoothed Top Tagger, the Boosted Decision Tree (BDT) Tagger, the Deep Neural Network (DNN) Tagger, and the Topocluster Tagger. The Smoothed Top Tagger uses two jet substructure variables as cuts, while the BDT and DNN Tagger use different machine learning network architectures to process a set of jet substructure variables. Similarly, the Topocluster Tagger also uses a neural network, except that it processes topocluster information instead of jet substructure variables.

2.1 General Description

Top tagging is the term used for an act of determining if a jet originates from a top quark. Due to its heavy mass, top quark has a very short lifetime ($\sim 5 \times 10^{-25}$ s) so it would decay before it reaches the detector [20]. For that reason, top quarks can only be detected indirectly by detecting the decay particles. Since the resonance of interest in this study is in the high mass range (> 1.1 TeV), following conservation of energy, the top quark pair from the resonance will be produced in a boosted state. The decay particles coming from a boosted top are collimated and can be encaptured in a single large-radius jet ("fat" jet), as illustrated by Figure 2.1. Boosted top tagging works by analyzing these fat jets rather than reconstructing the individual decay products. In this chapter, "signal" is defined as jets originated from top (or anti-top) quarks, and "background" is jets originated from the other flavors of quarks or gluons from QCD processes.

Top quark decays into bW^+ almost all the time ($BR(t \to bW^+) = 99.8\%$). This means the top quark decays are mostly determined by how the W boson decays with a signature of an additional b quark. The W boson branching ratio is easy to predict since it couples equally strong with each generation of fermions ($u\bar{d}$, $\nu_e e^+$, etc). Note that the quarks have branching ratios three times bigger than the leptons due to the color charge they carry, and the decay to $t\bar{b}$ is energetically not allowed because the top is heavier than the W. To summarize, the branching ratio of the top quark is,

$$BR(t \to b\nu_e e^+) : BR(t \to b\nu_\mu \mu^+) : BR(t \to b\nu_\tau \tau^+) : BR(t \to bq\bar{q}) = 1 : 1 : 1 : 6 \tag{2.1}$$

with $q\bar{q} = u\bar{d}, c\bar{s}$. The transverse component of the momentum of the neutrinos can be deduced from the assumption that the collision starts with zero transverse momentum. However, the longitudinal component of the neutrinos are not measurable, and therefore, some of the energy is unaccounted. This prevents the invariant mass to be fully constructed in the leptonic channel. Our search focuses on the hadronic decay channel $t \to bW^+ \to bq\bar{q}$ plus the anti-top counterpart, whose diagram is shown by Figure 2.2. Although it promises the largest signal branching ratio and the capability to fully construct the invariant mass, searches in the hadronic channel suffer from the abundant background from the QCD processes. It is the aim of top taggers to select the signal events buried in the background.

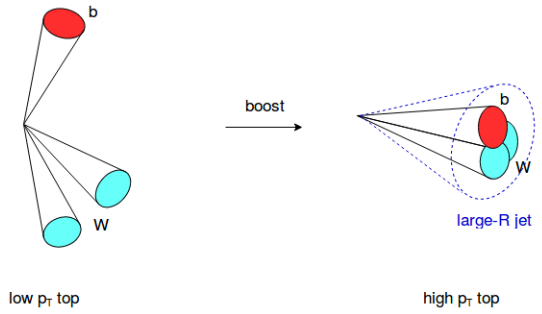


Figure 2.1: Decay particles from a low p_T and a boosted top quark observed in the laboratory frame. The decays of the boosted top quark travel in the same general direction with the boost and become collimated so they can be captured by one large-R ("fat") jet.

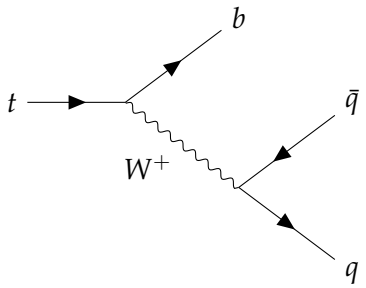


Figure 2.2: Feynman diagram of a hadronically decaying top quark. The three quarks in the final state create jets. In the boosted top scenario, these three jets are compacted together in one fat jet. The diagram for a hadronically decaying anti-top quark is identical with this one but with the charge and fermion line flipped.

2.2 Smoothed Top Tagger

There is a wide range of methods for boosted top tagging. The most common ones utilize the information from jet substructure variables, defined as a set of variables derived from the substructure of the jets to quantify certain property of the jets. If jets of interest can be defined with certain jet substructure variables, the variables can be used to distinguish those jets from other types of jets. The *Smoothed Top Tagger* uses two jet substructure variables, namely: *N-subjettiness* τ_{32} [21] and k_T *splitting scale* $\sqrt{d_{23}}$ [22].

N-subjettiness τ_{32} comes from the idea that the top quark is a three-body decay particle, so the top fat jet should contain three hard sub-jets. To check this, N-subjettiness procedure starts by taking a fat jet and reconstruct it using the exclusive-k_T clustering algorithm and force it to

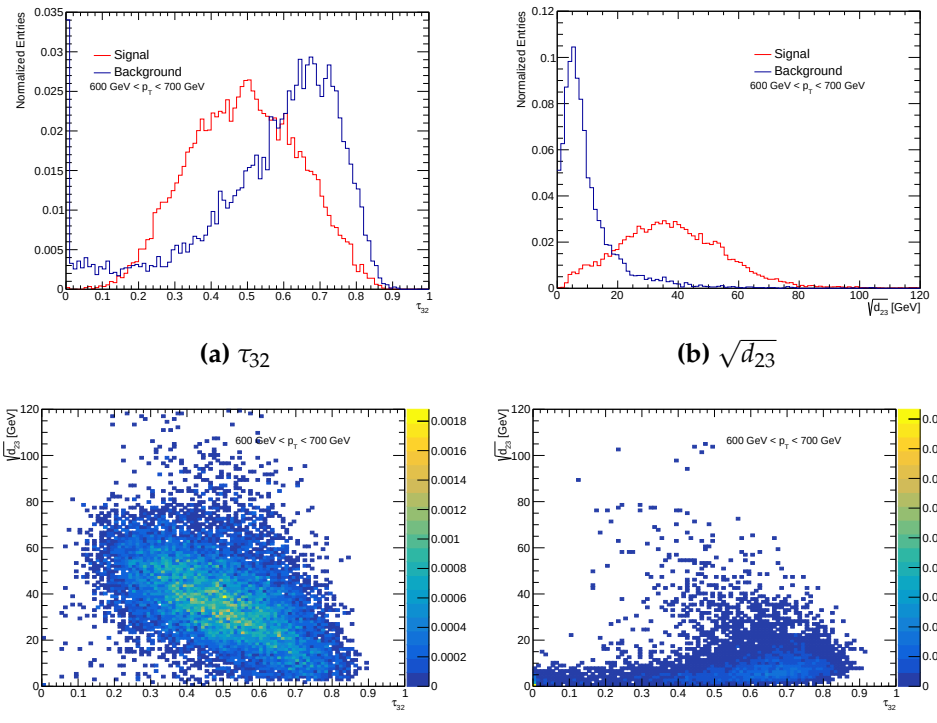
return N sub-jets. Then, it calculates τ_N, defined as,

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k}\, min\{\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k}\} \tag{2.2}$$

where k runs over all the jet's constituents, and $\Delta R_{j,k} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is the distance between a reconstructed sub-jet and a constituent k, and $d_0 = \sum_k p_{T,k} R_{jet}$ is the normalization factor where R_{jet} is the original jet's radius. It can be seen that τ_N measures how aligned the constituents are with the reconstructed sub-jets, or in another word, how well a jet can be regarded as a composition of N sub-jets. It has been shown in Ref. [21] that the ratio $\tau_{32} = \tau_3/\tau_2$ is a better discriminant for three-body jets than τ_3.

The k_T splitting scale $\sqrt{d_{23}}$ is calculated by first reconstructing the jet's constituents using the k_T jet-finding algorithm. The $\sqrt{d_{23}}$ comes from the quantity used in the jet-finding algorithm defined in Eq. 1.6. Because the k_T algorithm starts the clustering from the lowest p_T constituent, d_{23} corresponds to the clustering done in the second to last step, involving the second and third highest p_T particles, which, in the case of a top jet, are often two of the decay particles of the top.

Figure 2.3 shows the distribution of τ_{32} and $\sqrt{d_{23}}$ on signal and background jets in a p_T range of 600 - 700 GeV. It can be seen that signal jets are populating the central region of the $\tau_{32} - \sqrt{d_{23}}$ plane relative to the background jets which are more concentrated in the bottom-right region. The Smoothed Top Tagger chooses a value for τ_{32} as an upper cut and $\sqrt{d_{23}}$ as a lower cut to select the most of the signal while rejecting most of the background.



(a) τ_{32} **(b)** $\sqrt{d_{23}}$

(c) $\tau_{32} - \sqrt{d_{23}}$, signal **(d)** $\tau_{32} - \sqrt{d_{23}}$, background

Figure 2.3: The discrimination power of τ_{32} and $\sqrt{d_{23}}$ is shown individually in (a) and (b). Plot (c) and (d) show the distribution of $\tau_{32} - \sqrt{d_{23}}$ for signal and background. τ_{32} acts as the upper cut, while $\sqrt{d_{23}}$ acts as the lower cut to select most signals and reject most of the background.

2.3 BDT and DNN Tagger

While the previous tagger uses two variables to discriminate signal and background, one may intuitively think that the performance can be improved if more variables are considered. This idea is put into practice by using a machine learning network. Two taggers discussed in this section use two different network architectures, namely Boosted Decision Tree (BDT) and Deep Neural Network (DNN), to process the information from a set of jet substructure variables. Summary of the variables used is shown in Table 2.1. Details of the variables used can be found in Appendix B.

Table 2.1: Summary of input variables used for BDT and DNN top tagger [23].

Variables	BDT	DNN
m^{comb}	✓	✓
p_T	✓	✓
e_3	✓	✓
C_2		✓
D_2	✓	✓
τ_1		✓
τ_2	✓	✓
τ_3		✓
τ_{21}	✓	✓
τ_{32}	✓	✓
$\sqrt{d_{12}}$	✓	✓
$\sqrt{d_{23}}$	✓	✓
Q_W	✓	✓

The network works based on the principle of "supervised training". It learns from a training set that contains objects with known classification, in this case, signal and background jets. The training sets "teach" the network from the properties or variables assigned to each object to best determine the classification. The output is a mathematical function that can be used to predict the degree of classification of an unclassified object. The details of the training phase for the BDT and DNN Tagger: training and testing sets, training weights, etc. can be found in Ref. [24].

Processing all of the variables can be computationally wasteful, and adding uncorrelated variables may negatively affect the performance. Therefore, there needs to be an optimization to determine an optimal set of variables that go into the network. For the BDT Tagger, this is done by sequentially adding the variable into the network, starting from the variable with the highest discrimination. Relative background rejection is calculated at each step. The variables that do not give an improvement to the performance within statistical uncertainties are excluded. Figure 2.4a shows the relative background rejection at each step; the variables left to the vertical dashed line are the optimal variables. The jets used for the training must pass the training criteria: 200 GeV $< p_T <$ 2000 GeV, $m^{\text{comb}} > 40$ GeV[1], and number of constituents (N^{const}) > 2.

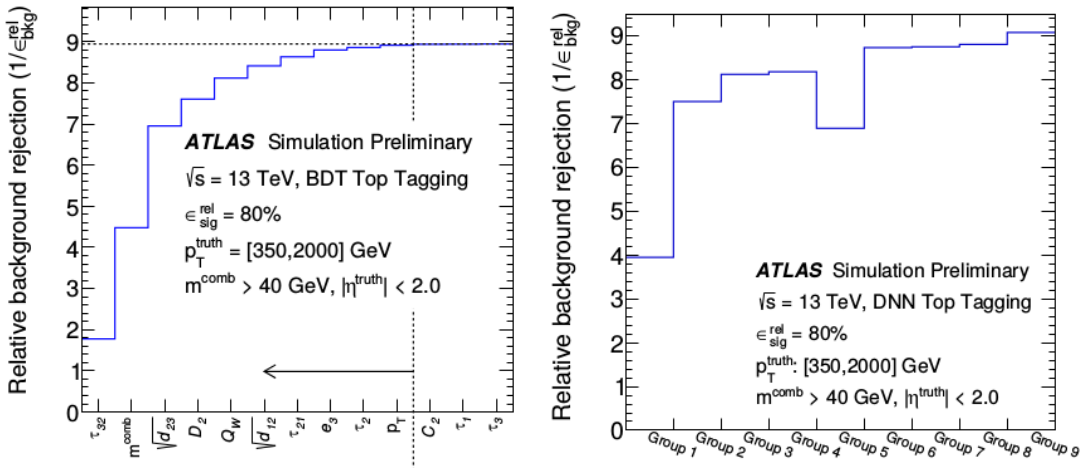
For the DNN Tagger, the optimization is done by training the network using different sets of variables that are grouped together. The grouping is based on the properties that the variables describe and their dependency on other variables, as listed in Table 2.2. The performance is calculated for each group and the relative background rejection is shown in Figure 2.4b. The

[1] After jet calibration, the combined mass m_{comb} is set as jet mass. The calculation of m_{comb} can be found in Appendix B.

group with the best background rejection is chosen, which happens to be the group with all of the starting variables. The same training criterion is applied for the DNN tagger.

Table 2.2: Variables grouping used in the DNN optimization [23].

Group 1	$C_2, D_2, \tau_{21}, \tau_{32}$
Group 2	$C_2, D_2, \tau_{21}, \tau_{32}, m^{\text{comb}}$
Group 3	$C_2, D_2, \tau_{21}, \tau_{32}, m^{\text{comb}}, p_T$
Group 4	$\tau_1, \tau_2, \tau_3, e_3, m^{\text{comb}}, p_T$
Group 5	$C_2, D_2, \tau_{21}, \tau_{32}, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W$
Group 6	$C_2, D_2, \tau_{21}, \tau_{32}, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W, m^{\text{comb}}$
Group 7	$\tau_1, \tau_2, \tau_3, e_3, m^{\text{comb}}, p_T, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W$
Group 8	$C_2, D_2, \tau_{21}, \tau_{32}, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W, m^{\text{comb}}, p_T$
Group 9	$\tau_1, \tau_2, \tau_3, \tau_{21}, \tau_{32}, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W, C_2, D_2, e_3, m^{\text{comb}}, p_T$



(a) BDT optimization scheme **(b)** DNN optimization scheme

Figure 2.4: Relative background rejection plots showing how the optimization procedure is done for the BDT and DNN Tagger. Optimization for the BDT Tagger is done by sequentially adding variables to the network and calculating relative background rejection at each step, the optimal variables are the ones on the left side of the dashed line. For the DNN Tagger, it is done by grouping the variables based on Table 2.2 and calculating relative background rejection on each group, group 9 is chosen since it gives the best performance [23].

2.4 Topocluster Tagger

The Topocluster Tagger [25] applies a similar but slightly different approach. It uses a DNN architecture, but instead of processing jet variables, it processes the kinematics of the jets' constituents/topoclusters. Compared to the jet substructure variables that are derived to quantify certain properties of a jet, the topoclusters promise a more fine-grained information about the structure of jets, and thus, have better resolution. Details about the network can be found in Ref. [25], although the implementation in ATLAS uses a slightly different configuration.

The network takes in kinematics information (p_T, η, ϕ) of the ten highest p_T clusters, ordered in descending p_T. Before being passed into the network, the clusters have to go into preprocessing phase. This is done to reduce the dimensionality of the object that has no physical significance, such as the orientation of the jet.

2.4.1 Preprocessing

The preprocessing phase is done in four steps. Firstly, the p_T of the topoclusters is scaled by 1/1700 GeV so that it falls in the same order of magnitude as the other input observables η and ϕ. Secondly, the topoclusters are translated to position the highest p_T cluster at the origin of the $\eta - \phi$ plane using Eq. 2.3 and 2.4,

$$\eta'_n = \eta_n - \eta_0 \tag{2.3}$$

$$\phi'_n = \phi_n - \phi_0 \tag{2.4}$$

where n denotes the cluster number, and 0 indicates the highest p_T cluster.

Thirdly, a rotation of angle ϑ is performed to align the second highest p_T cluster on the negative y-axis, below the highest p_T cluster. The rotation is done by transforming the p_y and p_z components of the momentum, since direct rotation in $\eta - \phi$ plane is not Lorentz invariant. The calculation of rotation angle ϑ and the transformation of p_y and p_z of the cluster is formulated by Eq. 2.5, 2.6, and 2.7 below.

$$\vartheta = \tan^{-1}\left(\frac{p_{y,1}}{p_{z,1}}\right) + \frac{\pi}{2} \tag{2.5}$$

$$p'_{y,n} = p_{y,n}\cos\vartheta - p_{z,n}\sin\vartheta \tag{2.6}$$

$$p'_{z,n} = p_{y,n}\sin\vartheta + p_{z,n}\cos\vartheta \tag{2.7}$$

Finally, if the average p_T of the clusters is in the left half of the $\eta - \phi$ plane, then flip all the clusters in η,

$$\text{if} \sum_{i=0}^{N} p_{T,i} \cdot \eta'_i < 0, \text{ then } \eta''_i = -\eta'_i \tag{2.8}$$

Figure 2.5 shows the kinematics of topoclusters from ∼400,000 signal and background jets on a $\eta - \phi - p_T$ plot before and after the rotation and flipping. The distinction between signal and background is noticeable in the figure, and it is from this signature that the network learns to discriminate the signal from the background.

The output from the networks is a single quantity called "score" that describes the classification of the jets. Figure 2.6 shows the distributions of the network output score of the BDT, DNN, and Topocluster Tagger; it is on these distributions that the cut is applied by the taggers.

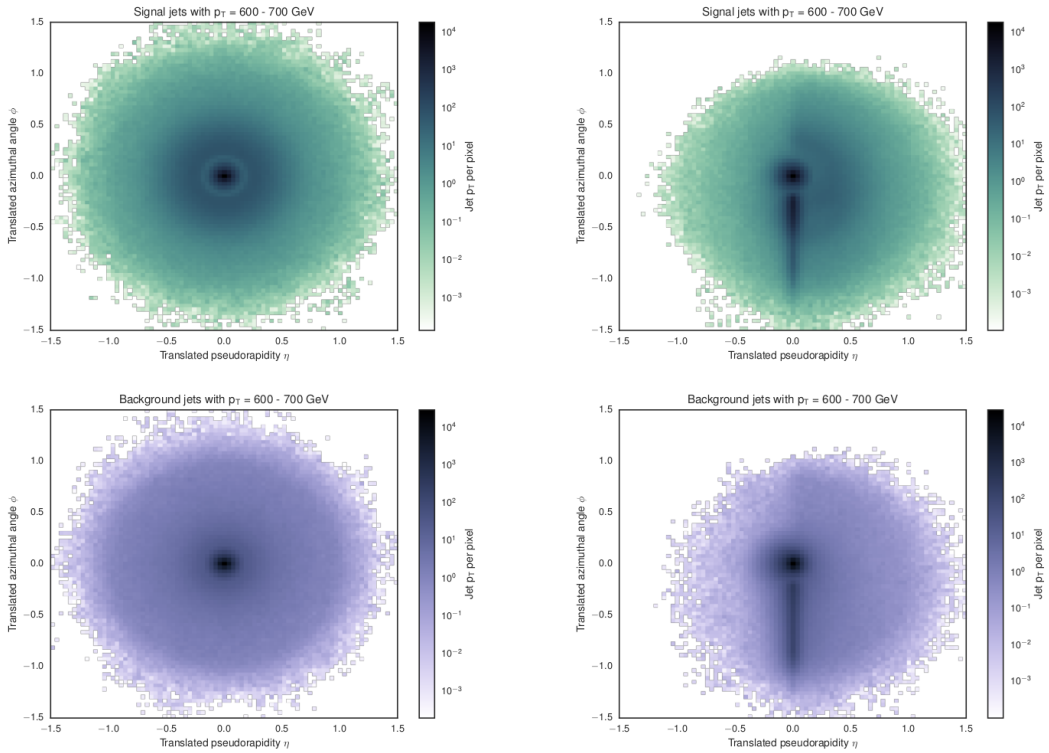
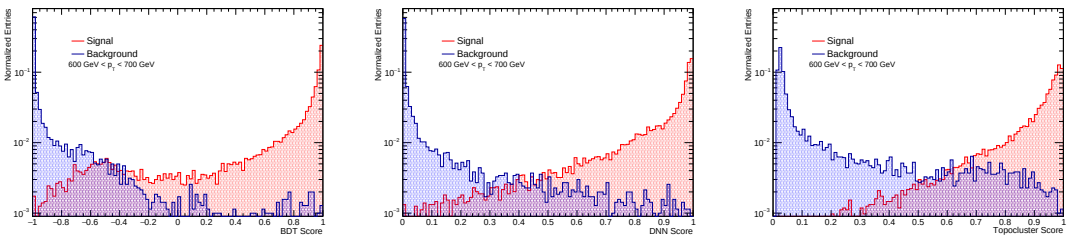


Figure 2.5: Plots on the left (right) column show images of jets before (after) preprocessing. Top (bottom) row shows signal (background) jets. The signature of top jets can be observed as the "shadowy" region on the upper-right plot near the central black circle. [25].



(a) BDT score **(b)** DNN score **(c)** Topocluster score

Figure 2.6: Distribution of the output score of the three machine learning taggers for jet p_T range of 600 - 700 GeV.

3 Performance Studies

This chapter features the performance studies that are done on the four boosted top taggers discussed in the previous chapter. The studies are done to measure the performance of the taggers in two different scenarios: tagging individual jets and tagging dijet events. The Monte Carlo samples used as background and signal are laid out, and the kinematic cuts are given. The performance is measured by comparing the background rejection of each tagger at a mutual signal efficiency level. The Topocluster Tagger is chosen to be the tagger used in the proceeding analysis based on the result and the reasoning explained at the end of this chapter.

3.1 The Monte Carlo Samples

For the background estimates, the samples used are the QCD Dijet datasets generated by PYTHIA8 [26, 27] and EVTGEN [28], known as the "JZXW" samples. In order to generate events with the same statistical power across the p_T range, the QCD Dijet samples are divided into different slices based on the truth[1] jet p_T. Each slice generates the same number of events with different weight for every slice. Inside every slice, each event also has a weight attached to them to make sure the high-end p_T of each slice has the same statistical power as the low-end p_T. The overall weight for the JZXW samples is formulated in Equation 3.1 below. It is necessary to apply this weight to obtain the expected smoothly falling p_T distribution. Full name of the datasets can be found in Appendix A. Note that the JZXW samples only contain dijets originated from QCD light quarks, the top quark contributions are contained by another set of samples discussed in the next chapter. But for the purpose of the performance studies, only the JZXW samples are used, as here, the background is defined simply as non-top jets.

$$\text{weight} = \frac{\sigma \times \text{filter efficiency} \times \text{event weight}}{\text{number of events}} \tag{3.1}$$

For the signal, the neutral boson from the heavy vector triplet (HVT) decaying into $t\bar{t}$ is used. The HVT samples are a good signal candidate for the purpose of the performance studies because the HVT $\rightarrow t\bar{t}$ samples are generated at relatively narrow interval of discrete mass points ranging from 400 GeV to 10 TeV, as shown in Appendix A. This mass range populates the

[1] "Truth" level denotes the information of particles that are generated by the event generator before being reconstructed by the detector simulation.

p_T spectrum from 0 to 6 TeV. The signal samples follow the same weighting procedure as the background JZXW samples, although the event weights are constantly equal to one.

3.2 Event Selection

As described in Chapter 2, boosted top tagging looks into large-R jets and determines if they have top signature. The specification of the jets of interest are topocluster jets reconstructed by the anti-k_T algorithm with radius $R = 1.0$. The jets are trimmed to remove the effect of pileup and other irrelevant processes. In the trimming process, the jet's constituents are reconstructed using the k_T algorithm with radius parameter R_{sub}, and the subjets which carry less than a specific fraction f_{cut} of the original jet p_T are removed. The trimming parameters chosen are $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 5\%$.

In the performance studies, in order to get a well-defined comparison, a more constraining definition of "signal" is required. The top quarks coming from the resonance are required to decay hadronically; events with leptonically decaying top are discarded. The signal jets also need to be ensured to fully contain the decay particles of the top. Containment is defined based on the distance of the truth decay particles to the jet $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$, and is classified into three groups,

1. Top contained
   Require the ΔR of the b quark and the two light quarks from the W to be $\Delta R < 0.75 R_{jet}$.

2. W contained
   Require the ΔR of the two light quarks from the W to be $\Delta R < 0.75 R_{jet}$ and the ΔR the b quark to be $\Delta R > 0.75 R_{jet}$.

3. Other
   Jets that do not satisfy two previous categories.

Here, R_{jet} is the radius of the jet ($= 1.0$). It is the "top contained" criterion that is taken as the signal in the performance studies. Figure 3.1 shows the distribution of the signal jet mass based on the containment classification divided into two p_T slices: 500 - 1000 GeV and 1000 - 1500 GeV. One can see that the mass distribution for top contained jets peaks at 175 GeV, and the mass distribution for W contained jets peaks at 80 GeV, which agrees with the observed mass of the top quark and W boson. Notice also that the fraction of the top contained jets is increased in the higher p_T slice; as the more boosted are the jets, the more grouped together the decay particles are, making it easier for the jets to contain all three of them.

Kinematic cuts are also applied in the performance studies to mimic the cuts used in the search. To study the boosted top taggers in tagging individual jets, only the highest p_T jet of every event is considered in the performance study of individual jets. Events with leading jet p_T lower than 440 GeV are removed, so that only adequately boosted jets are considered. The jets are required to have more than 2 constituents for the tagger to work effectively due to the fact that the taggers are dependent on the jet substructure variables and the constituents. To ensure that the jet is in the central region of the detector, a $|\eta^{\text{jet}}| < 2.0$ cut is applied.

The cuts used in the dijet study generally follow from the cuts used in individual jets study

with some additions. The sub-leading jet p_T is required to be higher than 350 GeV, note that this is higher than the threshold used in the usual dijet analysis (> 60 GeV) because the taggers require the jets to be considerably boosted. The rapidity difference $y^* = \frac{(y_1 - y_2)}{2}$ is defined for leading and sub-leading jet, and a cut of $|y^*| < 0.6$ is applied to reduce the effect from background processes [29]. Summary of the kinematic cuts used in the performance studies can be found in Table 3.1.

In the individual jets (dijet) performance study, the signal p_T (m_{jj}) distribution is reweighted to match the background p_T (m_{jj}) distribution so that the performance can be continuously measured over the spectrum range. This reweighting procedure is only done in the performance studies and is not applied in the actual analysis.

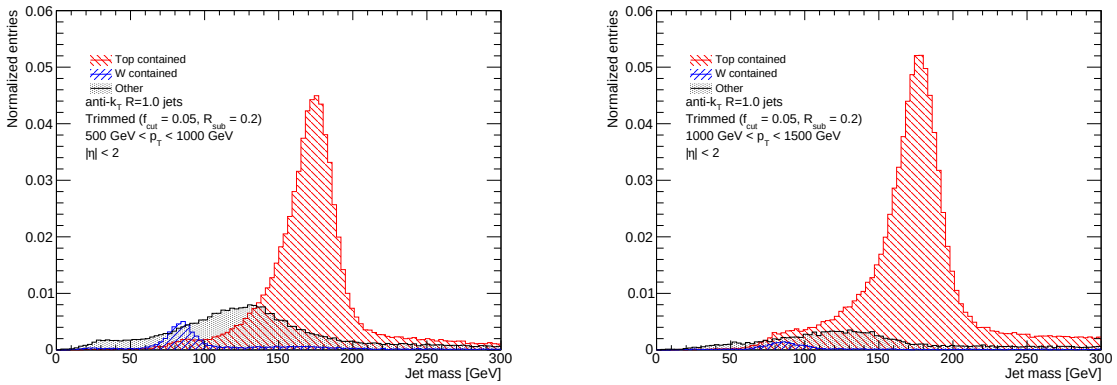


Figure 3.1: Distribution of signal jet mass based on the containment classification for lower p_T [500 - 1000 GeV] (left) and higher p_T [1000 - 1500 GeV] (right). The top contained jets are the ones used as signal in the performance studies.

Table 3.1: Summary of the cuts used in the performance studies.

	Leading jet p_T	Sub-leading jet p_T	$N^{\text{jet}}_{\text{constituents}}$	$\lvert\eta^{\text{jet}}\rvert$	$\lvert y^*\rvert$
Tagging individual jets	> 440 GeV	-	> 2	< 2.0	-
Tagging dijet events	> 440 GeV	> 350 GeV	> 2	< 2.0	< 0.6

3.3 Tagging Individual Jets

To quantify the performance of the taggers, signal efficiency and background rejection is calculated. Signal efficiency is defined as the fraction of the signal jets that are tagged by the tagger, and background rejection is defined as the inverse of background efficiency or the number of background jets that are not tagged per every one background jet that is mis-tagged; written in mathematical form in Equation 3.2 and 3.3. Note that the subscript "total" means the total number of jets that passes the kinematic cuts before the taggers are applied.

$$\text{Signal Efficiency} = \epsilon_{\text{sig}} = \frac{N^{\text{sig}}_{\text{tagged}}}{N^{\text{sig}}_{\text{total}}} \tag{3.2}$$

$$\text{Background Rejection} = \frac{1}{\epsilon_{\text{bkg}}} = \frac{N^{\text{bkg}}_{\text{total}}}{N^{\text{bkg}}_{\text{tagged}}} \tag{3.3}$$

The taggers are set at a constant working point of 80% signal efficiency. For the three machine learning taggers, this is done by finding a cut value that gives 80% signal efficiency for every bin, and fitting over those cut values using a polynomial equation. For the Smoothed Top Tagger, it is done by scanning the $\tau_{32} - \sqrt{d_{23}}$ distribution to look for values that give the desired signal efficiency, and choose the cut values that give the most background rejection for every p_T bin, then finally fit the cut distribution over the jet p_T. The cut functions can be found in Appendix C.

After setting the signal efficiency constantly to 80%, the comparison of the performance is then done by looking at the background rejection. The result is shown in Figure 3.2a and 3.2c. Note that when comparing these plots, one needs to take into consideration that the signal efficiency is not perfectly constant at 80% due to the imperfection of the fit used to get the cut functions. A general trend that can be observed is that the performance of the taggers is declining with increasing p_T. This is due to the fact that the taggers are based on the jet substructure and jet constituent information; as the jets get more boosted, the subjets and constituents get closer together and overlap, resulting in jet substructure variables and constituents to lose some information. It can also be seen in the background rejection plot that the Smoothed Top Tagger is outperformed by the machine learning taggers at most of the p_T range, proving the superiority of the three machine learning taggers.

Another representation of the taggers' performance is shown by the Receiver Operating Characteristic (ROC) curve that plots the signal efficiency vs the background rejection as shown in Figure 3.3 for a p_T slice [1000,1500] GeV. In this p_T range, the three machine learning taggers perform almost identically, and they perform better than the Smoothed Top Tagger by roughly a factor of two, evaluated at the signal efficiency of the Smoothed Top Tagger.

3.4 Tagging Dijet Events

In the case of tagging dijet events, the definition of signal efficiency and background rejection in Equation 3.2 and 3.3 is still used, but now N represents event count instead of jet count. Signal efficiency and background rejection are now calculated over the invariant mass of the dijet system m_{jj}, taking the highest p_T and the second highest p_T jet. To tag dijet events, a new variable is needed to place the cut on. *Dijet score* is simply defined as the average of the leading and sub-leading jet's tagger score,

$$\text{Dijet score} = \frac{\text{score}_1 + \text{score}_2}{2} \tag{3.4}$$

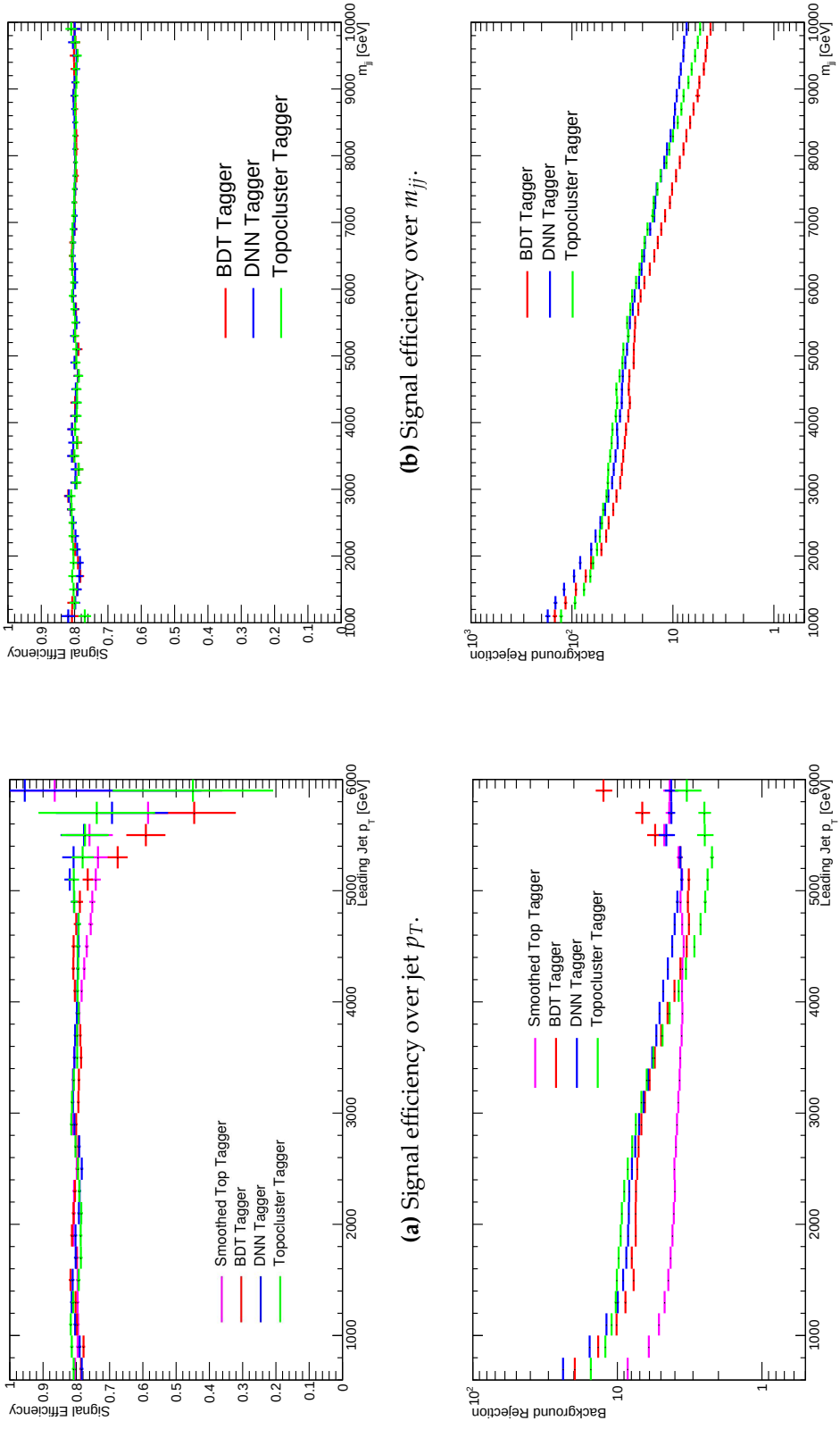
The cut functions derivation follows a similar procedure as before, but now the fit is done over the dijet score distribution over m_{jj}. The cut functions used on dijet score distribution to get constant 80% signal efficiency can be found in Appendix C. The performance of the Smoothed Top Tagger is not studied further in the dijet case as the previous result has shown its relatively poor performance and does not motivate further application.

The result of the performance study in the dijet case is shown by Figure 3.2b and 3.2d. The DNN Tagger performs better compared to the BDT Tagger across the range, this could be due to the extra number of variables that the DNN Tagger uses. The Topocluster Tagger is slightly better in the middle mass range compared to the other two taggers, although being slightly

worse at the lower and higher mass range.

Moving forward, the Topocluster Tagger is chosen to be the tagger used in the dijet resonance search analysis in this project. This is supported by the results shown earlier that the Topocluster Tagger performs best in the mass range of interest. The relatively worse performance in higher mass has less significance since, at the level of integrated luminosity the LHC has so far[2], there are few events at that high mass region. Also, since the Topocluster Tagger uses topocluster kinematics, it processes a more fine-grained information compared to the BDT and DNN Tagger which use jet substructure variables, and thus, has better resolution. This leads to the idea that the Topocluster Tagger has the most potential to improve in the future development.

[2]Integrated luminosity is a measure of the number of collisions at a particle accelerator. The number of event of a certain process is equal to the integrated luminosity multiplied by the corresponding cross section, $N = \int Ldt \times \sigma$.



(a) Signal efficiency over jet p_T.

(b) Signal efficiency over m_{jj}.

(c) Background rejection over jet p_T.

(d) Background rejection over m_{jj}.

Figure 3.2: Performance plots of the boosted top taggers acting on individual jets (left column) and dijet events (right column).

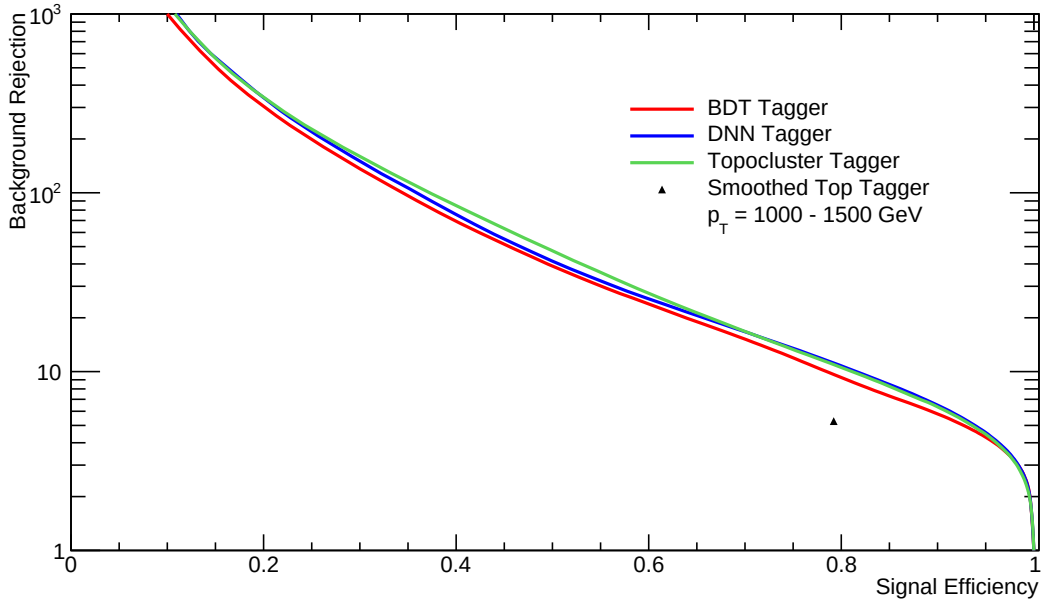


Figure 3.3: ROC Curve showing signal efficiency vs background rejection of the four boosted top taggers in jet p_T range of 1000 - 1500 GeV. The three machine learning taggers are represented by the colored line, and the Smoothed Top Tagger is represented by a single point where it is optimized.

4 Dijet Mass Distribution

After the boosted top taggers are studied and the Topocluster Tagger is picked as the tool of choice, a distribution is needed for the tagger to work on. The Standard Model (SM) $t\bar{t}$ processes are discussed here to be added to the QCD light quarks from the JZXW samples. Together, the QCD light quarks and the SM $t\bar{t}$ makes up the smoothly falling background distribution. After the full QCD mass distribution is obtained and the Topocluster Tagger is applied, the Monte Carlo distribution is transformed so that it has the same statistical character as in real data, the procedure is discussed in the second part of this chapter. The output is the dijet mass distributions that are analyzed by the resonance search method in the next chapter.

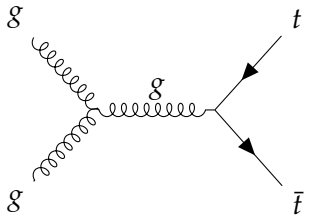
4.1 SM Top Quark Pair Production

Running at the center of mass energy of $\sqrt{s} = 13$ TeV, the LHC is by far the most productive top quark factory in the world. Top quarks can be produced via single top quark or top quark pair productions. Since the presence of top quarks is of interest in this study, it is important to consider the top quarks coming from the SM QCD contribution, especially the top quark pairs. The leading order diagrams of SM $t\bar{t}$ production are shown by Figure 4.1. The gluon-gluon fusion process dominates the $t\bar{t}$ production at the LHC by 90% [30], while the quark-antiquark annihilation is suppressed because the LHC collides pp, which means the antiquark must come from virtual sea quark.

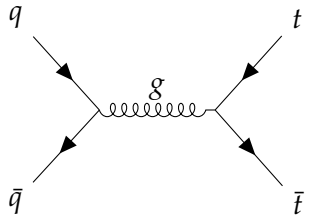
In the context of a BSM resonance search, the SM $t\bar{t}$ production is considered to be part of the background since it originates from known SM processes. The JZXW samples do not contain the QCD top production, therefore, it needs to be added from another dedicated set of samples. The SM $t\bar{t}$ samples are generated by POWHEG [31] + PYTHIA event generator, the list of the full names of the samples can be found in Appendix A. Just like in the JZXW samples, the SM $t\bar{t}$ samples are also sliced into different datasets to make up for the low statistics in the more energetic region. However, in the SM $t\bar{t}$ datasets, samples are sliced based on the invariant mass of the $t\bar{t}$ system $m_{t\bar{t}}$ instead of the truth jet p_T. The slices span from $m_{tt} = 1.1$ TeV to $m_{tt} = 14$ TeV. To populate the $m_{tt} < 1.1$ TeV region, an inclusive sample is used, taking only events with $m_{tt} < 1.1$ TeV. Table 4.1 summarizes the m_{tt} range for each of the SM $t\bar{t}$ samples used. Notice that by applying the cut $m_{tt} < 1.1$ TeV to the inclusive sample, a full SM $t\bar{t}$ distribution across the m_{tt} range is achieved.

The weight used in POWHEG is slightly different from the one used in PYTHIA, as formulated in Equation 4.1 below.

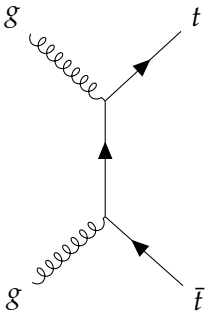
$$\text{weight} = \frac{\sigma \times \text{filter efficiency} \times \text{event weight}}{\text{sum of weight}} \tag{4.1}$$



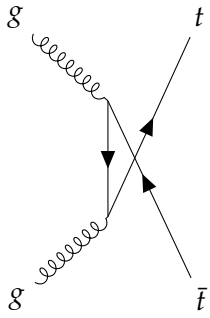
(a) $t\bar{t}$ production via gluon-gluon fusion



(b) $t\bar{t}$ production via quark-antiquark annihilation



(c) $t\bar{t}$ production via gluon scattering



(d) $t\bar{t}$ production via gluon scattering

Figure 4.1: Leading order diagrams of SM $t\bar{t}$ production.

Table 4.1: The SM $t\bar{t}$ samples slices and their corresponding m_{tt} range.

SM $t\bar{t}$ sample	m_{tt} range
Inclusive	0 - 14 TeV
Slice 1	1.1 - 1.3 TeV
Slice 2	1.3 - 1.5 TeV
Slice 3	1.5 - 1.7 TeV
Slice 4	1.7 - 2 TeV
Slice 5	2 - 14 TeV

Figure 4.2 shows the distribution of the SM $t\bar{t}$ with the kinematic cuts applied and scaled by the assumed integrated luminosity of 100 fb^{-1}, the number is an approximation of the current amount of data that has been collected by ATLAS so far.[1] Note that the distribution loses statistical power at $m_{jj} > 4$ TeV, but with the assumed integrated luminosity, this is of little significance because there are very few events beyond that mass point. To get the full QCD background distribution, the distributions from the JZXW samples are then added with the SM $t\bar{t}$ distribution. Figure 4.3 shows the $t\bar{t}$-enriched QCD background mass distribution before and after applying the Topocluster Tagger. Notice that after top tagging, the number of events from the QCD distribution drops at higher magnitude compared to the SM $t\bar{t}$.

[1] By the end of the 2017 run, ATLAS has recorded total integrated luminosity of 86 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions [32].

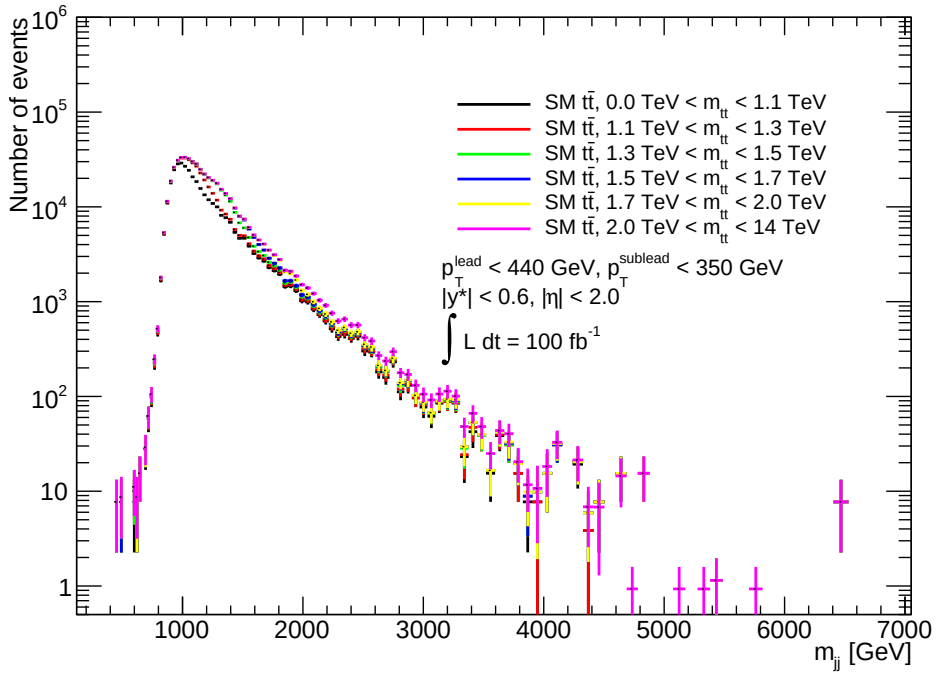


Figure 4.2: Distribution of SM $t\bar{t}$ of different slices scaled by integrated luminosity 100 fb^{-1}. The histograms are stacked, so the total number is represented by the magenta line. The decline at the lower mass region is due to the kinematic cuts removing events with low p_T jets.

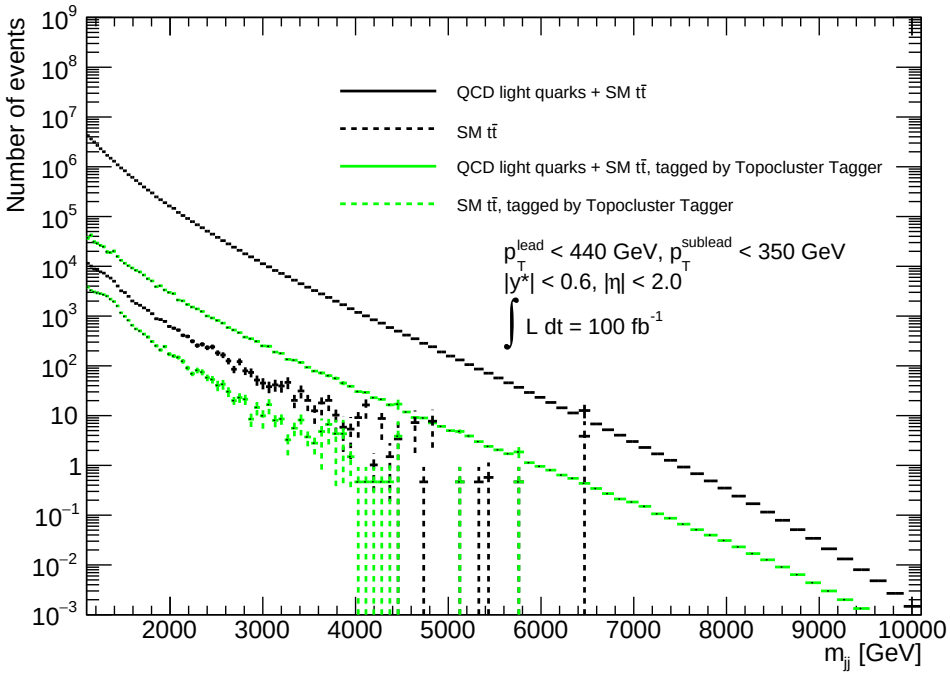


Figure 4.3: QCD dijet mass distribution enriched with SM $t\bar{t}$. Black (green) line represents the distribution before (after) applying the Topocluster Tagger.

4.2 Making Data-like Distributions

In the sliced Monte Carlo simulations, rare events are generated at a much higher rate than what the detector records in reality. The simulated events are then weighted down to get the real-life distribution as previously discussed. This is done to get an optimal statistical power across the energy spectrum, especially at the higher energy region where the occurrence of the events is low. As a result, distributions in Monte Carlo simulations have much less statistical uncertainty then what could be achieved in real experimental data. So in order to study the Monte Carlo distribution as if it was real data, the distribution needs to be transformed so that it would have the same statistical fluctuation as one would find in a real experiment.

To get the *data-like* distribution, first the Monte Carlo distribution is fitted using the SWIFT method (details of the SWIFT method are described in the next chapter). For each bin, the data-like event count is a Poisson random number generated using the fit value evaluated at the center of the bin as the Poisson mean. The uncertainty of each data point is assigned to be $\sqrt{N}$, with N being the data-like event count for the bin. This procedure is done on the $t\bar{t}$-enriched QCD background Monte Carlo distribution, before and after top tagging, as shown in Figure 4.4. These two data-like distributions act as the background mass spectrum in the analysis done in the next chapter.

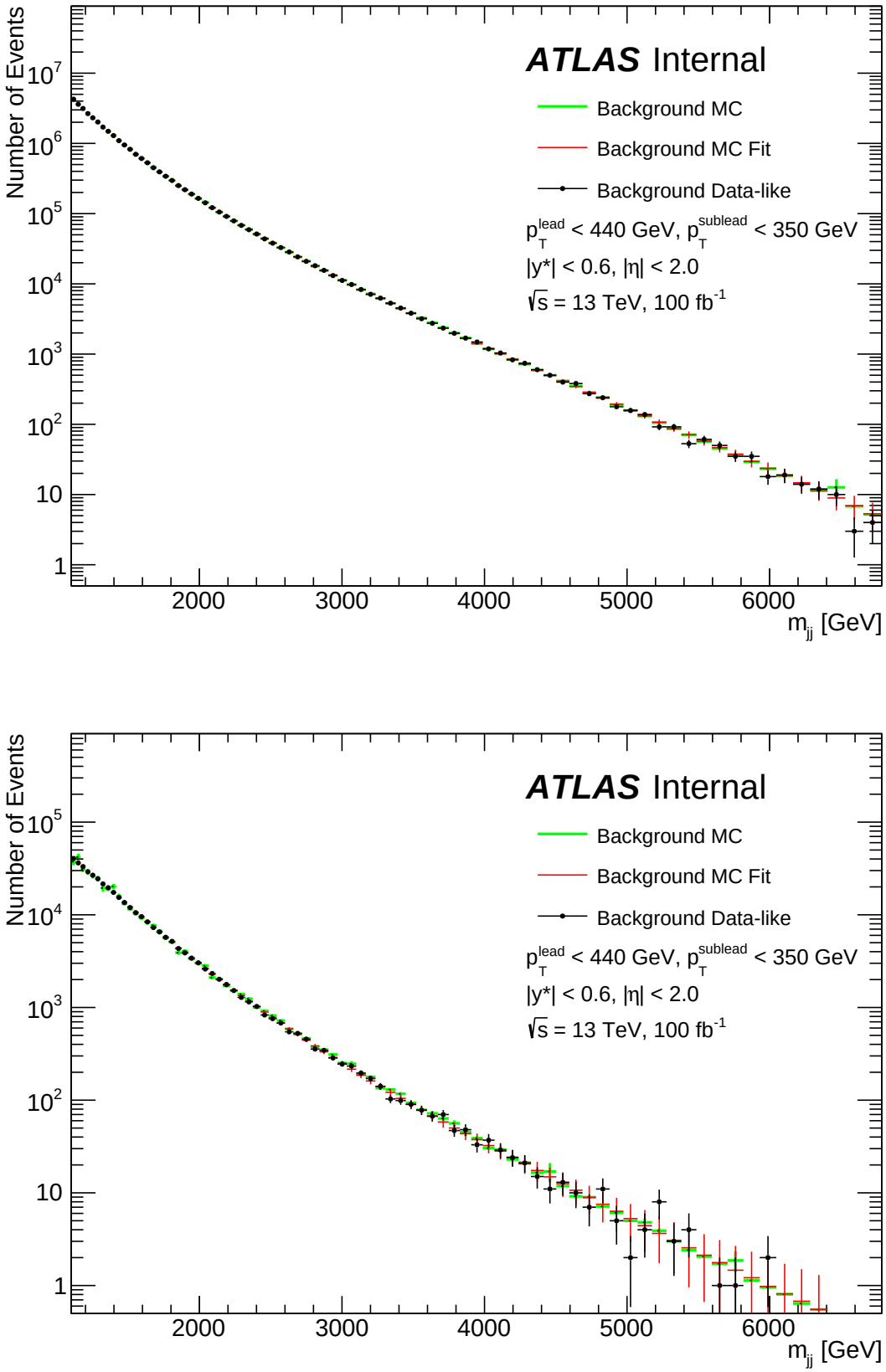


Figure 4.4: Data-like background distribution before top tagging (top plot) and after top tagging (bottom plot), obtained by generating random number according to Poisson distribution using the fit value as the mean for each bin.

5 Sliding Window Fit

The resonance search method used for the analysis in this thesis is called the Sliding Window Fit (SWIFT) [33] method. The main idea of the method is to analyze the mass distribution in smaller windows by doing simultaneous background and background+signal fit, and "slide" the window over the distribution. The motivation behind the method and the general procedure is discussed here along with the search parameters and the chosen values. Readers who are interested in more technical details are referred to the cited source.

5.1 General Description

In the past, resonance searches were done by looking for excess in the dijet mass distribution above a background estimate given by the "dijet function", defined as,

$$f(x) = p_0(1-x)^{p_1}x^{p_2+p_3\ln x} \tag{5.1}$$

where $x = m_{jj}/\sqrt{s}$, and p_i are the parameters. The equation shown above is the 4-parameter variant of the dijet function. By setting $p_3 = 0$, it becomes the 3-parameter dijet function. The dijet function is derived from the leading order matrix element and parton distribution function, and thus, is only an approximation. In the previous dijet searches [34, 35], the dijet function had proven to be effective to fit the mass distribution. However, with the increasing amount of data that the LHC has accumulated and the new mass range it opens, the approximation of the function becomes less efficient at describing the mass distribution, especially at the higher mass range. SWIFT offers a solution to this issue by fitting over smaller windows in which the dijet function can fit relatively easy.

SWIFT essentially is a resonance search method that works by analyzing the data spectrum in many small and overlapping windows. Each window is analyzed by doing various number of fits and extracting some statistical quantities, with the bin edges assigned as the window centers. Generally, the fits performed by SWIFT are divided into two types: nominal and alternate fit. The nominal fit is defined as the 3-parameter dijet function, and the alternate fit is the 4-parameter dijet function.

SWIFT takes in a mass distribution and the signal shape that one looks for. By the end of its

run, SWIFT gives out,

- a local p-value scan based on log-likelihood ratio (LLHR)
- 95% Confidence Level (CL) limits
- a background estimate

The search in this study analyzes the mass range of 1100 GeV to 6787 GeV. The lower limit is motivated by the optimal efficiency of the $|y^*| < 0.6$ selection which starts at 1100 GeV, while the upper limit is motivated by the top-tagged mass distribution having no events beyond 6787 GeV. The first and last window center have to be set at some bins away from the edge to avoid problematic background+signal fit at those areas as the signal component would be cut away by the edge. The first and last window centers are set to be 1493 GeV and 5931 GeV, respectively.

5.2 Signal Parameterization

Prior to the search, SWIFT needs the signal shape to construct the background + signal fits. It requires a signal shape at every window center; unfortunately, signal samples are only generated at discrete mass points, leaving some gaps in between those mass points. In order to circumvent this issue, the "signal morphing" method – included in the SWIFT package – is applied to the signal mass distributions. First, it fits the MC signal mass distributions using a "Gaussian - reverse Landau" (GrL) function, defined as,

$$f(x) = p_0 \times [p_3 \times \text{Gauss}(m_{jj}, p_1, p_2) + (1 - p_3) \times \text{Landau}(-m_{jj}, p_4, p_5)] \tag{5.2}$$

where the parameters p_i are described as follows:

- $p_0 =$ normalization factor
- $p_1 =$ Gaussian mean
- $p_2 =$ Gaussian width
- $p_3 =$ Gaussian/Landau fraction
- $p_4 =$ Landau mean
- $p_5 =$ Landau width

By performing the GrL fit to the MC signal at its generated mass points, the values of the GrL parameters at each generated mass points are obtained. Then, each parameter is interpolated with a cubic spline as a function of mass. As a result, signal shapes at any mass points can be generated by evaluating the parameters of the interpolation functions at the desired mass value.

The signal morphing procedure is done on the HVT and Z' signal shapes before and after top

tagging, the fits and the resulting parameters can be found in Appendix D. For a totally model-independent search, a hypothetical fixed-width Gaussian signal generated at every window center is used.

5.3 Choosing Window Sizes

SWIFT initializes its run by determining the size of each window by performing nominal background-only (3-parameter dijet function) fit in different window sizes. For every window, SWIFT chooses the window size with the best χ^2 p-value.

The window sizes that are allowed are configurable by the user and are measured by a percentage of the window center. For example, for window center at 2 TeV and window size of 30%, the window will span from 2 TeV $\times$ $(1 - 30\%) = 1.4$ TeV to 2 TeV $\times$ $(1 + 30\%) = 2.6$ TeV. SWIFT sets the window edges to the bin edges closest to 1.4 TeV and 2.6 TeV[1]. It is important to make sure that the minimum window size is at least three times the signal width so that the background+signal fit that is performed in the later step can capture the entire signal shape. Minimum and maximum window sizes are set to be 30% and 100%, respectively.

The user also needs to specify the number of window sizes to be tested at the first window center. To test 20 window sizes means that, at the first window center, SWIFT would test 20 window sizes from 30% and 100% with the increment of 3.5%, and choose the window with the best χ^2 p-value. For the next window until the last one, only 5 window sizes are tested: 2 sizes above and below the previous chosen size; this prevents the window to shrink or grow too dramatically. For this analysis, the number of window sizes tested at the first window is chosen to be 20.

5.4 Local p-value Scan

After the window size fixing is done as described in the previous section, SWIFT slides over the spectrum from the first window center to the last, and at each window, two pairs of fits are applied:

1. Nominal
    - Nominal background-only fit
    - Nominal signal+background fit
2. Alternate
    - Alternate background-only fit

[1]The "window center" is not always at the center of the window. For example, for the first window center at 1493 GeV, a 30% window size means that it starts from 1493 GeV $\times$ $(1 - 30\%) = 1045$ GeV, lower than the low edge of 1100 GeV. In this case, the first window is set at 1100 GeV. The same argument applies to the last and close-to-last windows.

- Alternate signal+background fit

The signal+background fit is the dijet function added with the signal shape that is obtained from the signal parametrization step with the window center as the mass of the signal. The normalization factor of the signal is allowed to float to be able to best capture the signal shape in the data if a signal is present. At each window, SWIFT chooses to keep either the nominal or alternate fits based on the χ^2 p-value of the signal+background fits. The size of the signal component determines the "extracted signal" at every window. Note that by this definition, the extracted signal can be negative due to negative fluctuation in the data.

Once the pair of fits is chosen, the log-likelihood ratio (LLHR) is calculated by Equation 5.3 below. LLHR quantifies how much adding the signal component improves the quality of the fit.

$$\text{LLHR} = \ln \frac{\text{LH}_{s+b}}{\text{LH}_b} \tag{5.3}$$

The likelihood (LH) is calculated by summing the Poisson probability of getting the bin content x_i using the estimate from the background fit as the mean λ_i over all bins,

$$\text{LH}(x|\lambda) = \sum_{i=1}^{N} e^{-\lambda_i} \frac{\lambda_i^{x_i}}{x_i!} \tag{5.4}$$

To get the local p-value, Wilk's theorem is applied to convert the LLHR to χ^2 which can be translated to local p-value. According to Wilk's theorem, for a pair of nested models, in the limit of large statistics, χ^2 with degrees of freedom k equals to negative two times the LLHR (Equation 5.5). Wilk's theorem is applicable here because the models used in the ratio use the same underlying function (background-only and signal+background function use the same background function), so they can be considered nested. The LLHR-based local p-value determines the significance of any excess found in the data spectrum.

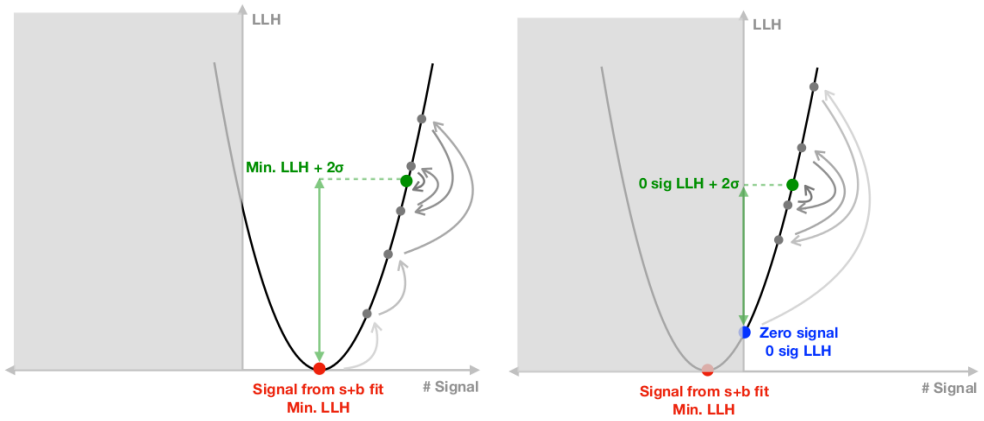
$$\chi^2(k) = -2 \times \text{LLHR} \tag{5.5}$$

5.5 95% Confidence Level Limits

In the case of no significant excess found as determined from the local p-value scan, the 95% Confidence Level (CL) limit is calculated. The 95% CL limit is defined as the rate of signal production that agrees with the data at 95% confidence level or 2σ.

The limit calculation starts with the extracted signal which is derived from the signal component of the signal+background fit. To find the limit, the binary search algorithm is performed as illustrated in Figure 5.1. The red dot in the sketch represents the number of the extracted signal from the signal+background fit and the likelihood it corresponds to (defined as the minimum likelihood), and the green dot represents the number of extracted signal and the likelihood ratio made worse by 2σ from the minimum LLH. The cross section that corresponds to the number of signal of the green dot, by definition, is the 95% CL limit. In the case of a negative extracted signal, the 2σ deviation is calculated from the LLH that corresponds to zero extracted signal (see Figure 5.1b). This comes from the assumption that any signal from BSM processes should not cause negative deviation from the background, so any negative fluctuation is purely statis-

tical. The algorithm tries to find the green dot by injecting or subtracting signal iteratively until it reaches the desired likelihood within a tolerance value (set to be 0.001).



(a) Positive signal case **(b)** Negative signal case

Figure 5.1: Binary search procedure to find the 95% CL limits [33].

5.6 SWIFT Background

To construct the background estimate, after the choice of nominal or alternate pair of fits, SWIFT saves the value of the background-only fit evaluated at the window center of each window to be the background estimate of only that particular bin. This way each window center gets a background estimate. For the bins below and above the first and last window center, the background estimates are obtained by evaluating the background-only fit obtained in the first and last window at those particular close-to-the-edge bins. This completes the background estimation for the whole range; such is the process to produce the fit line in Figure 4.4 shown in the previous chapter.

Figure 5.2 and 5.3 show the performance of the SWiFt background estimate on data-like background dijet mass distribution before and after top-tagging that are used in the analysis (see Figure 4.4), in comparison with the 3-parameter and 4-parameter dijet function. The significance on the bottom pad is defined as $(\text{data} - \text{fit})/\sqrt{\text{data}}$. The significance swing observed on 3-parameter and 4-parameter dijet function fit exposes the approximation nature of the dijet function. The accuracy of the fit is described by χ^2 p-value printed on the plots. It can be seen that the 4-parameter fit is better than 3-parameter fit, and the SWIFT background fit is better than 4-parameter fit.

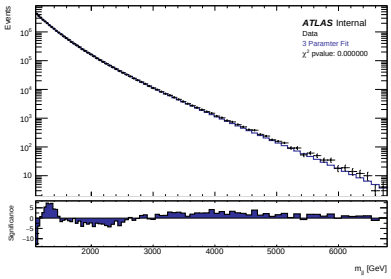
5.7 Signal Subtraction

In the case of a significant excess detected in the data, the background fit could be pulled up by the presence of the signal. This fit bias causes the background estimate to be inaccurate in describing the real background. To handle this issue, SWIFT is equipped with signal subtraction procedure. The procedure is activated if there is a window with local p-value less than a threshold value; here set to be 0.001 which corresponds to approximately 3σ excess. The subtraction is done by extracting the number of signal captured by the signal component of the

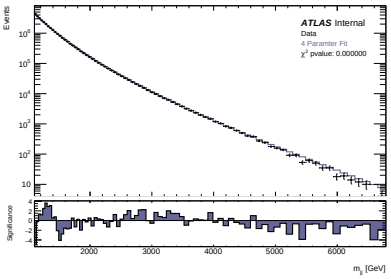
background+signal fit from the data distribution. Note that the signal subtraction is only done at one window: the window with the smallest local p-value. After the signal has been subtracted, the background estimate construction (explained in the previous section) is repeated with the same type of fit (nominal or alternate) and window sizes. The new background estimate can be used to calculate the global p-value or passed to other resonance search method (e.g BUMPHUNTER [36]) as a bias-free background estimate. Table 5.1 summarizes the value of parameters chosen for the SWIFT resonance search performed in this analysis.

Table 5.1: Summary of the parameters used for the SWiFt resonance search in the analysis.

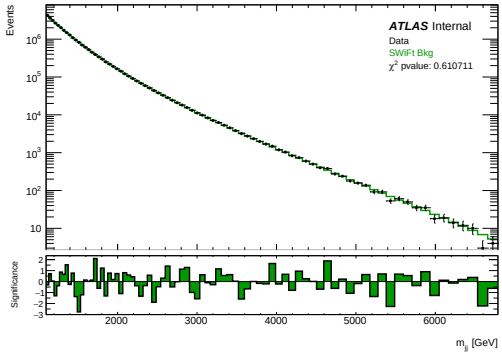
Category	Parameter	Value
Ranges	Low edge of the search	1100 GeV
	High edge of the search	6787 GeV
	First window center	1493 GeV
	Last window center	5931 GeV
Window size	Minimum window size	30%
	Maximum window size	100%
	Number of window size tested	20
Limit calculation	Limit binary search tolerance	0.001
	Integrated luminosity	100 fb^{-1}
Signal subtraction	Local p-value trigger threshold	0.001



(a) 3-parameter dijet function fit

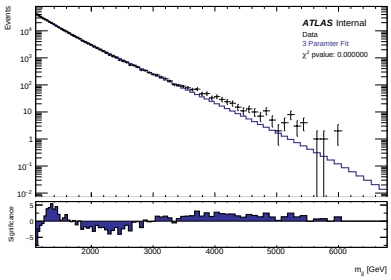


(b) 4-parameter dijet function fit

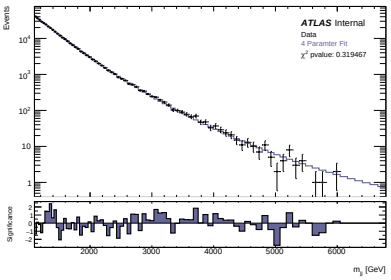


(c) SWiFt background

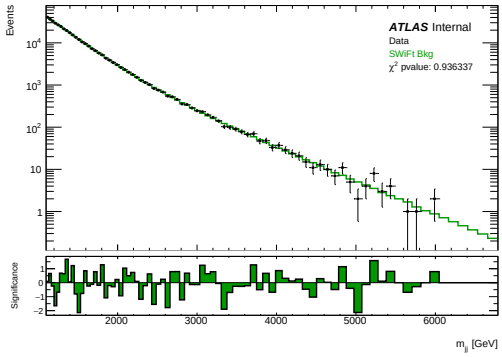
Figure 5.2: Background estimates of the dijet mass distribution using (a) 3-parameter dijet function, (b) 4-parameter dijet function, and (c) SWIFT, before top tagging.



(a) 3-parameter dijet function fit



(b) 4-parameter dijet function fit



(c) SWiFt background

Figure 5.3: Background estimates of the dijet mass distribution using (a) 3-parameter dijet function, (b) 4-parameter dijet function, and (c) SWIFT, after top tagging.

6 Results

This chapter is dedicated to show the results of the resonance search and the analysis surrounding it. For each of the result plots, two cases are displayed: before and after top tagging, to show the effect of the top tagging to the quality of the search. The first part of the chapter discusses the search results where only background processes are considered, and the second part discusses the scenario where signal is injected in the spectrum at various mass points and strengths. The readers are reminded that the data used in the analysis all come from Monte Carlo simulations.

6.1 Background Only

The results shown in this section are the SWIFT resonance search results using only the dijet mass contribution from the background processes, before and after top tagging by the Topocluster Tagger. The background processes considered are the QCD light jets plus the SM $t\bar{t}$ production, as shown by Figure 4.4. The SWIFT procedure is applied to the two data-like mass distribution and the relevant outputs are shown here. Multiple signal models are used for the resonance search: hypothetical Gaussian signal with various widths ($\sigma_G/m_G = 3\%$, 5%, and 7%), HVT $\rightarrow t\bar{t}$, and Z′ $\rightarrow t\bar{t}$.

Figure 6.1 shows the local p-value scan as a function of the window center for the dijet mass distribution before and after top tagging. Since no local p-value is lower than the detection threshold (< 0.001), no significant excess is found in the local p-value scan, as expected from background-only distributions. Any excesses observed (dips in the local p-value plots) can be attributed to statistical fluctuations. It can be concluded that the method does not mistake any statistical fluctuation as a significant excess. Note that since no local p-value falls below the signal subtraction threshold for all signal models, both before and after top tagging, the signal subtraction is not triggered.

After the local p-value scan has proven no significant excesses observed, SWIFT calculates the 95% CL limits. The limit plots for the HVT and Z′ model are shown by Figure 6.2, along with the theory line for each model. The expected limit is the median of the 95% CL limits calculated from 1000 pseudo-experiments, and the green and yellow band is where 68% and 95% of the limits of the pseudo-experiments lies around the median, respectively. The theory line is calculated by the product of cross section, branching ratio, and selection efficiency at

each generated mass points. Note that the selection efficiency is affected by the top tagging, so the line is slightly lower for the top-tagged case. The mass points below the theory line are excluded at 95% confidence level. By comparing the limit plots before and after top tagging, it can be observed that the act of top tagging decreases the cross section limit on both models by roughly one order of magnitude. Before top tagging, there are no mass points excluded for the two signal models. After top tagging, HVT $\rightarrow t\bar{t}$ is excluded at mass points [1533, 1698] GeV and 2600 GeV, and Z′ $\rightarrow t\bar{t}$ is excluded at mass points [1533, 1698] GeV, 1968 GeV, and [2542, 2600] GeV. This signifies that the act of top tagging increases the sensitivity of the search, although, a better tagger is desired for more exclusion at the given cross sections.

The limit plots for the Gaussian signal models are shown by Figure 6.3, with the bands drawn for the Gaussian model with width $\sigma_G/m_G = 3\%$ and the limits from the other widths are overlaid. Here, the sensitivity increase after top tagging can also be observed by the decrease in the cross section limits. Notice that the limits from wider signals are generally worse. As the signal gets wider, it requires a higher cross section to leave a noticeable "bump" on the mass distribution, thus, the worse limits.

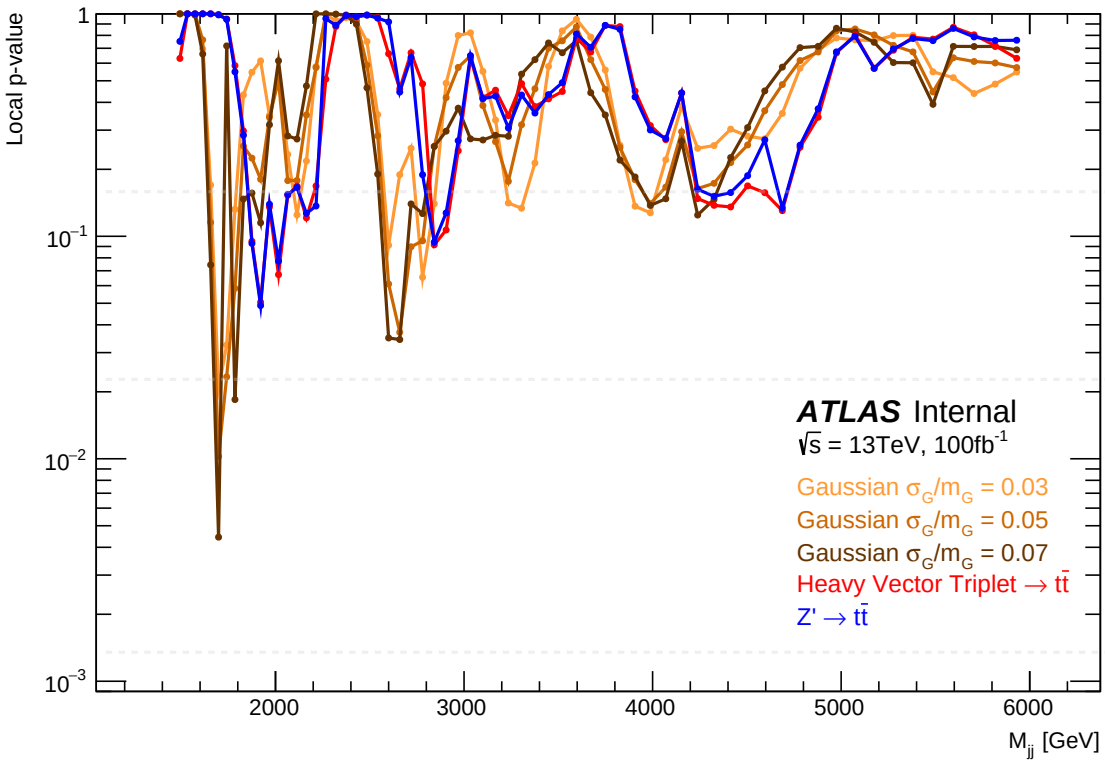
6.2 Injected Signal

This section presents the resonance search results of dijet mass distributions from background processes with a signal presence. The Z′ $\rightarrow t\bar{t}$ signal of mass 2 TeV, 3 TeV, and 4 TeV are injected separately to the background mass distribution. The signal strength for each injected signal mass points is varied by multiplying the predicted cross section by a factor of 1, 3, and 5; creating multiple mass distributions containing signals at different mass points and strengths.

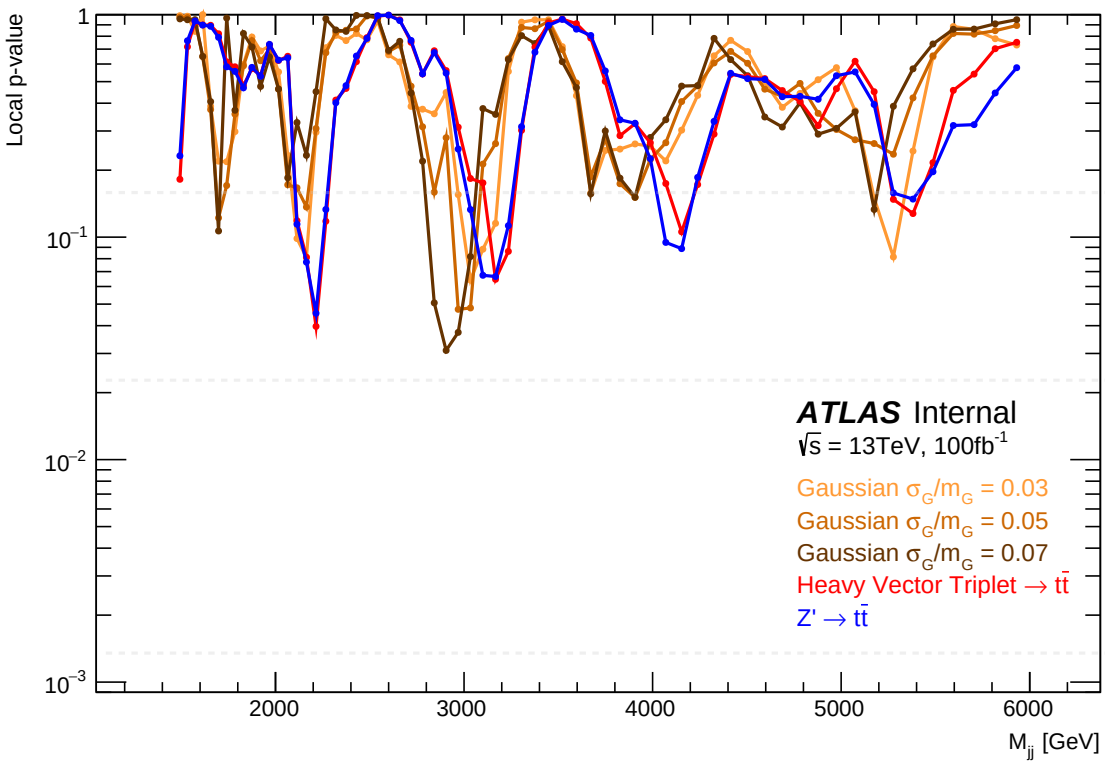
The results from the local p-value scan for each of the signal injected mass distributions are shown in Figure 6.4 (note the different ranges of the y-axis). A general trend can be observed: local p-value around each injected mass points after top tagging is lower than before top tagging, which implies that the presence of the signal is more pronounced and becomes relatively easy to detect after top tagging. The detection is most recognized in case of injected signal at 2 TeV mass point, where the local p-value reaches 9×10^{-5} (4×10^{-10}) for $\sigma \times 3(5)$ at 1968 GeV. For the injected signal at 3 TeV mass point, detection is only observed in the $\sigma \times 5$ case at 3100 GeV; and is not observed for the injected signal at 4 TeV mass point. The lost of sensitivity at higher mass is due to the small predicted cross section at higher mass and the insufficient integrated luminosity to produce a significant amount of signal event at the higher mass range.

As a final proof of the benefit of top tagging in the resonance search, the BUMPHUNTER [36] algorithm is employed to analyze the dijet mass distribution with Z′ $\rightarrow t\bar{t}$ injected at mass 2 TeV with strength $\sigma \times 3$, before and after top tagging, as shown by Figure 6.5. The background line is the SWIFT background estimate for each distribution. Note that since the local p-value reaches < 0.001 after top tagging (see Figure 6.4d), the background estimate for the top-tagged distribution is the one from the signal subtracted distribution. Before top tagging, BUMPHUNTER fails to detect any significant excess; and after top tagging, it manages to identify the signal correctly at the range of 1785-2164 GeV, matching the injected signal at mass 2 TeV.

All of these results are proofs that the quality of the search is improved by applying top tagging. The fact that one might see something that was previously hidden encourages the application of this tool and method on real data, and see what might have been hiding all along.

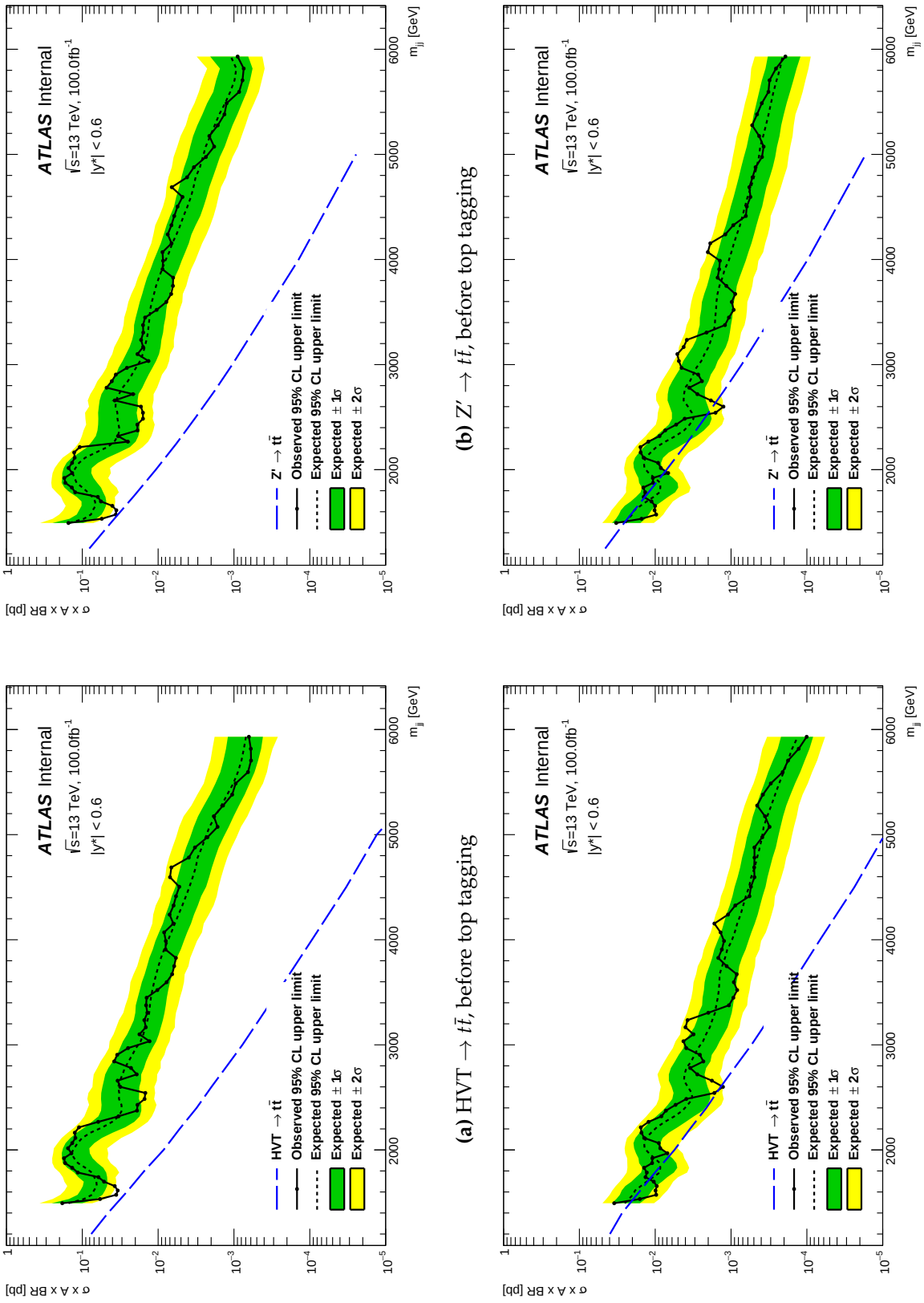


(a) before top tagging



(b) after top-tagging

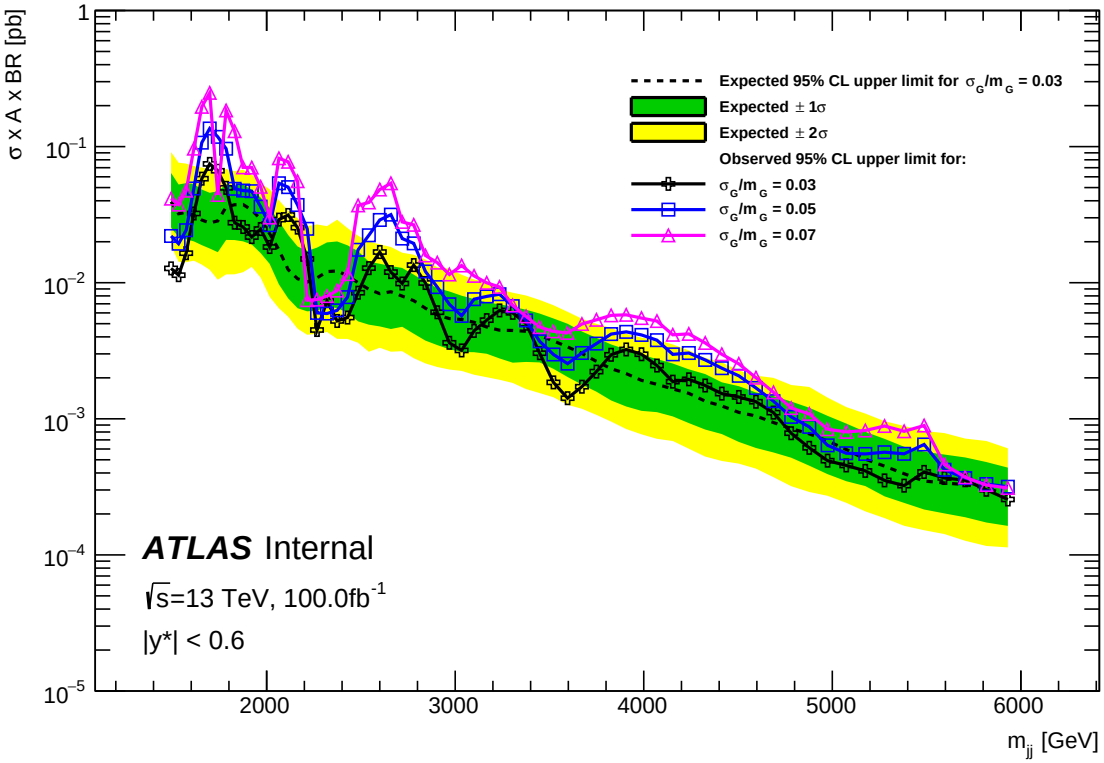
Figure 6.1: Local p-value plotted over the window center for background-only dijet mass distribution (a) before and (b) after top tagging.



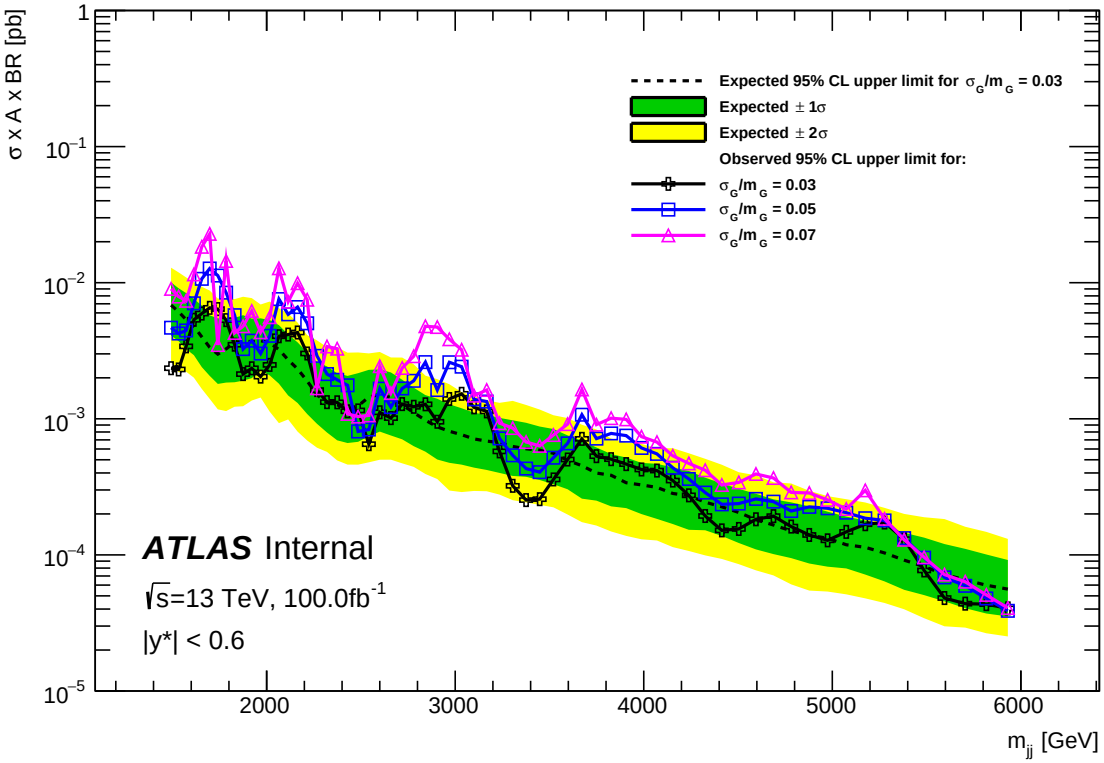
(a) HVT $\rightarrow t\bar{t}$, before top tagging **(b)** $Z' \rightarrow t\bar{t}$, before top tagging

(c) HVT $\rightarrow t\bar{t}$, after top tagging **(d)** $Z' \rightarrow t\bar{t}$, after top tagging

Figure 6.2: The 95% CL limit plots for HVT (left column) and Z' (right column) signal models, before (top row) and after (bottom row) the top tagging. The decrease in the cross section limit after the top tagging signifies the sensitivity improvement of the search.

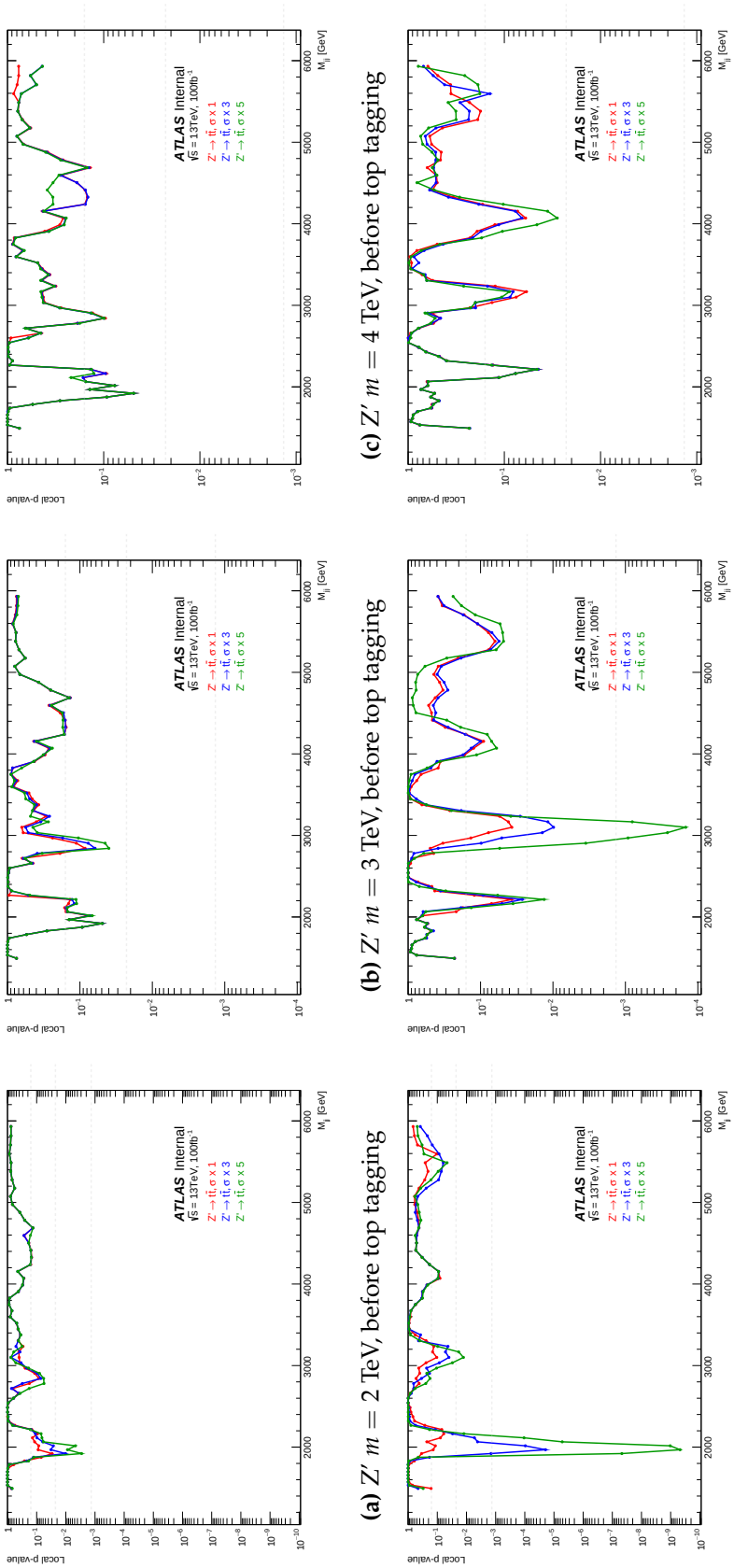


(a) before top tagging



(b) after top tagging

Figure 6.3: The 95% CL limit plots for a hypothetical Gaussian signal with various widths (a) before and (b) after the top tagging. The band is drawn for the $\sigma_G/m_G = 0.03$ Gaussian signal.



(a) Z' $m = 2$ TeV, before top tagging **(b)** Z' $m = 3$ TeV, before top tagging **(c)** Z' $m = 4$ TeV, before top tagging

(d) Z' $m = 2$ TeV, after top tagging **(e)** Z' $m = 3$ TeV, after top tagging **(f)** Z' $m = 4$ TeV, after top tagging

Figure 6.4: Results of local p-value scan on various signal injected mass distributions.

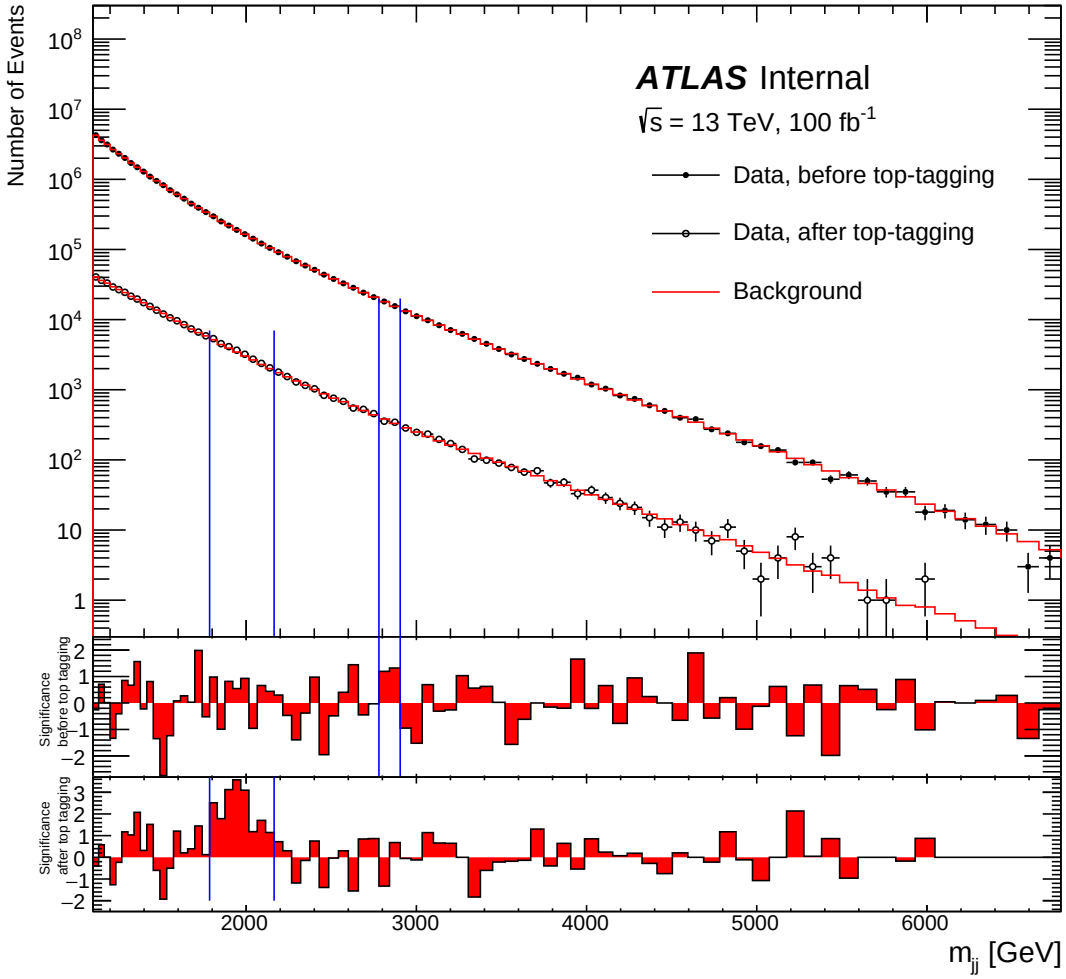


Figure 6.5: The BUMPHUNTER result for the mass spectrum with injected Z′ m = 2000 GeV ($\sigma \times 3$). The background line is the background estimate from SWIFT. The increase of sensitivity by top tagging let the BUMPHUNTER pick up the presence of the signal that is not detected before top tagging.

7 Summary and Outlook

This thesis is motivated by the potential of discovering new physics in the top channel, specifically in the hadronic final state. By looking at the hadronic channel, the invariant mass of the resonance particle can be fully reconstructed and it also promises the largest fraction of the signal. However, this comes with a price of an overwhelming amount of background from QCD processes. To handle this problem, boosted top tagging is applied to keep most of the signal, while rejecting as much background as possible.

Four boosted top taggers are studied: the Smoothed Top Tagger, the DNN Tagger, the BDT Tagger, and the Topocluster Tagger. The performances of the taggers are evaluated by applying them in two different tasks: tagging individual jets, and tagging dijet events. At the p_T range of 1000 - 1500 GeV, the Smoothed Top Tagger is outperformed by the three machine learning-based taggers by roughly a factor of two. By comparing the background rejection at a constant working point of 80% signal efficiency in tagging dijet events, the Topocluster Tagger is observed to be slightly better than the BDT and DNN Tagger in the mass range of interest. Since the topoclusters contain lower level informations compared to the jet substructure variables, the Topocluster Tagger arguably has better resolution to resolve jets at higher p_T, and thus, promises potential for further improvement in the future. Following this argument, and supported by the result of the performance study, the Topocluster Tagger is chosen to be the tagger used in the resonance search in the thesis.

The SWIFT resonance search method is employed to analyze the MC-derived dijet mass distributions using various signal models: HVT $\rightarrow t\bar{t}$, $Z' \rightarrow t\bar{t}$, and Gaussian signal shapes of width σ_G/m_G = 3%, 5%, and 7%. To see the effect of top tagging on the quality of the search, the method is performed before and after top tagging.

First, the method is tested on the background-only mass distribution. The local p-value scan returns no significant excess from the distributions both before and after top tagging, as one would expect from a background-only distribution. The 95% CL limits are calculated for each of the signal models. It can be concluded from the decrease of the limits for each signal models after top tagging, that the search gains sensitivity from adapting top tagging.

Second, a signal with various masses and strengths are injected separately to the distributions to see how well the method can detect the signal. The signal used for injection is $Z' \rightarrow t\bar{t}$ at mass 2 TeV, 3 TeV, and 4 TeV; with strengths of $\sigma \times 1$, $\sigma \times 3$, and $\sigma \times 5$. Confirming the

observation from the background-only test, the sensitivity improvement is noticeable from the local p-value after top tagging that falls lower around the injected mass points compared to the one before top tagging. To better visualize the benefit of top tagging, the mass distribution injected with Z' (m = 2 TeV, $\sigma \times 3$) before and after top tagging are passed to BUMPHUNTER together with the background estimate from SWIFT. Before top tagging, the background is too overwhelming compared to the signal, resulting in BUMPHUNTER's failure in detecting the signal. After top tagging, the background is suppressed and BUMPHUNTER manages to identify the signal around the injected mass point.

There are several modifications that one can do to improve the analysis done in this project. The tagger cut functions can be altered to find a more optimal working point that can give better background rejection, which does not have to be at constant signal efficiency. Further study on the machine learning network can prove fruitful in improving the performance of the tagger (which is demonstrated in Ref [37], where the Long Short-Term Memory network improves the performance of the Topocluster Tagger by a factor of two). The parameters of the SWIFT resonance search can be adjusted to suit a more specific model. But of course, the ultimate interest lies in applying the tools and methods demonstrated in this work to real data, and see what might surface.

Appendices

A

Lists of Monte Carlo Samples

Table A.1: Datasets used for background estimates

Sample	Name
QCD Dijet	mc15_13TeV.361023.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ3W.merge.DAOD_JETM8.e3668_s2576_s2132_r7267_r6282_p2528 mc15_13TeV.361024.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ4W.merge.DAOD_JETM8.e3668_s2576_s2132_r7267_r6282_p2528 mc15_13TeV.361025.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ5W.merge.DAOD_JETM8.e3668_s2576_s2132_r7267_r6282_p2528 mc15_13TeV.361026.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ6W.merge.DAOD_JETM8.e3569_s2608_s2183_r7267_r6282_p2528 mc15_13TeV.361027.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ7W.merge.DAOD_JETM8.e3668_s2608_s2183_r7267_r6282_p2528 mc15_13TeV.361028.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ8W.merge.DAOD_JETM8.e3569_s2576_s2132_r7267_r6282_p2528 mc15_13TeV.361029.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ9W.merge.DAOD_JETM8.e3569_s2576_s2132_r7267_r6282_p2528 mc15_13TeV.361030.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ10W.merge.DAOD_JETM8.e3569_s2576_s2132_r7267_r6282_p2528 mc15_13TeV.361031.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ11W.merge.DAOD_JETM8.e3569_s2608_s2183_r7267_r6282_p2528 mc15_13TeV.361032.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ12W.merge.DAOD_JETM8.e3668_s2608_s2183_r7267_r6282_p2528
SM $t\bar{t}$ production	mc15_13TeV.410007.PowhegPythiaEvtGen_P2012_ttbar_hdamp172p5_allhad.merge.DAOD_JETM6.e4135_s2608_s2183_r7725_r7676_p3297 mc15_13TeV.303722.PowhegPythiaEvtGen_P2012_ttbar_hdamp172p5_allhad_mtt_1.merge.DAOD_JETM6.e4321_s2608_r7772_r7676_p3297 mc15_13TeV.303723.PowhegPythiaEvtGen_P2012_ttbar_hdamp172p5_allhad_mtt_2.merge.DAOD_JETM6.e4321_s2608_r7772_r7676_p3297 mc15_13TeV.303724.PowhegPythiaEvtGen_P2012_ttbar_hdamp172p5_allhad_mtt_3.merge.DAOD_JETM6.e4321_s2608_r7772_r7676_p3297 mc15_13TeV.303725.PowhegPythiaEvtGen_P2012_ttbar_hdamp172p5_allhad_mtt_4.merge.DAOD_JETM6.e4321_s2608_r7772_r7676_p3297 mc15_13TeV.303726.PowhegPythiaEvtGen_P2012_ttbar_hdamp172p5_allhad_mtt_5.merge.DAOD_JETM6.e4321_s2608_r7772_r7676_p3297

Table A.2: Datasets used for signal estimates

Sample	Name
HVT	mc15_13TeV.307382.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m0400.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307383.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m0600.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307384.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m0800.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307385.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m1000.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307386.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m1200.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307387.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m1400.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307388.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m1600.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307389.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m1800.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307390.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m2000.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307391.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m2200.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307392.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m2400.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307393.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m2600.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307394.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m2800.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307395.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m3000.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307396.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m3500.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307397.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m4000.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307398.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m4500.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307399.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m5000.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307400.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m5400.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307401.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m5800.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307402.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m6200.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307403.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m6600.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307404.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m7000.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307405.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m7400.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307406.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m7800.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307407.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m8200.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307408.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m8600.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307409.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m9000.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307410.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m9400.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307411.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m9800.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297 mc15_13TeV.307412.MadGraphPythia8EvtGen_A14NNPDF23LO_HVT_Agv1_Vztt_all_m10000.merge.DAOD_JETM6.e5624_s2726_r7772_r7676_p3297
Z′	mc15_13TeV.301323.Pythia8EvtGen_A14NNPDF23LO_zprime500_tt.merge.DAOD_JETM8.e4061_s2608_s2183_r7772_r7676_p2794 mc15_13TeV.301324.Pythia8EvtGen_A14NNPDF23LO_zprime750_tt.merge.DAOD_JETM8.e4061_s2608_s2183_r7772_r7676_p2794 mc15_13TeV.301325.Pythia8EvtGen_A14NNPDF23LO_zprime1000_tt.merge.DAOD_JETM8.e4061_s2608_s2183_r7772_r7676_p2794 mc15_13TeV.301326.Pythia8EvtGen_A14NNPDF23LO_zprime1250_tt.merge.DAOD_JETM8.e4061_s2608_s2183_r7725_r7676_p2794 mc15_13TeV.301327.Pythia8EvtGen_A14NNPDF23LO_zprime1500_tt.merge.DAOD_JETM8.e4061_s2608_s2183_r7772_r7676_p2794 mc15_13TeV.301328.Pythia8EvtGen_A14NNPDF23LO_zprime1750_tt.merge.DAOD_JETM8.e4061_s2608_s2183_r7772_r7676_p2794 mc15_13TeV.301329.Pythia8EvtGen_A14NNPDF23LO_zprime2000_tt.merge.DAOD_JETM8.e4061_s2608_s2183_r7772_r7676_p2794 mc15_13TeV.301330.Pythia8EvtGen_A14NNPDF23LO_zprime2250_tt.merge.DAOD_JETM8.e4061_s2608_s2183_r7772_r7676_p2794 mc15_13TeV.301331.Pythia8EvtGen_A14NNPDF23LO_zprime2500_tt.merge.DAOD_JETM8.e4061_s2608_s2183_r7772_r7676_p2794 mc15_13TeV.301332.Pythia8EvtGen_A14NNPDF23LO_zprime2750_tt.merge.DAOD_JETM8.e4061_s2608_s2183_r7772_r7676_p2794 mc15_13TeV.301333.Pythia8EvtGen_A14NNPDF23LO_zprime3000_tt.merge.DAOD_JETM8.e3723_s2608_s2183_r7725_r7676_p2794 mc15_13TeV.301334.Pythia8EvtGen_A14NNPDF23LO_zprime4000_tt.merge.DAOD_JETM8.e3723_s2608_s2183_r7772_r7676_p2794 mc15_13TeV.301335.Pythia8EvtGen_A14NNPDF23LO_zprime5000_tt.merge.DAOD_JETM8.e3723_s2608_s2183_r7772_r7676_p2794

B

Jet Substructure Variables

B.1 Combined Mass

The combined mass m^{comb} [38] is a linear combination of the nominal jet mass m^{calo} which is reconstructed from the topoclusters and the track-assisted jet mass m^{TA}, defined as,

$$m^{\text{comb}} = w^{\text{calo}} m^{\text{calo}} + w^{\text{TA}} m^{\text{TA}} \tag{B.1}$$

The track-assisted jet mass m^{TA} is calculated by reconstructing the tracks in the inner detector, defined as,

$$m^{\text{TA}} = m_{\text{tracks}} \times \frac{p_T^{\text{calo}}}{p_T^{\text{tracks}}} \tag{B.2}$$

where m_{tracks} and p_T^{tracks} are the invariant mass and transverse momentum calculated from the tracks inside the jets and p_T^{calo} is the original jet transverse momentum obtained from the calorimeter.

The weights w^{TA} and w^{calo} are defined as,

$$w^{\text{TA}} = \frac{\sigma_{\text{TA}}^{-2}}{\sigma_{\text{calo}}^{-2} + \sigma_{\text{TA}}^{-2}} \tag{B.3}$$

$$w^{\text{calo}} = \frac{\sigma_{\text{calo}}^{-2}}{\sigma_{\text{calo}}^{-2} + \sigma_{\text{TA}}^{-2}} \tag{B.4}$$

$$\tag{B.5}$$

where σ_{TA}^{-2} and $\sigma_{\text{calo}}^{-2}$ are the m^{TA} and calorimeter mass resolutions, respectively.

B.2 Energy Correlation Function

The N-point energy correlation function (ECF) [39] is defined as,

$$ECF_N(\beta) = \sum_{i_1<i_2<...<i_N \epsilon J} \Big(\prod_{a=1}^{N} p_{T,i_a}\Big)\Big(\prod_{b=1}^{N-1}\prod_{c=b+1}^{N} R_{i_b i_c}\Big)^{\beta} \tag{B.6}$$

where β is the angular exponent (set to 1) and R_{ij} is the distance between the i and j constituents in the $y-\phi$ plane. Only up to 3-point correlation functions are considered:

$$ECF_1(\beta) = \sum_{i \epsilon J} p_{T,i} \tag{B.7}$$

$$ECF_2(\beta) = \sum_{i<j \epsilon J} p_{T,i} p_{T,j} (\Delta R_{ij})^{\beta} \tag{B.8}$$

$$ECF_3(\beta) = \sum_{i<j<k \epsilon J} p_{T,i} p_{T,j} p_{T,k} (\Delta R_{ij} \Delta R_{ik} \Delta R_{jk})^{\beta} \tag{B.9}$$

Using Equation B.7 to B.9, one can construct ratios of ECF, as follows,

$$e_2^{(\beta)} = \frac{ECF_2(\beta)}{ECF_1(\beta)^2} \tag{B.10}$$

$$e_3^{(\beta)} = \frac{ECF_3(\beta)}{ECF_1(\beta)^3} \tag{B.11}$$

Furthermore, double ratios of ECF can be defined, as follows,

$$C_2(\beta) = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^2} \tag{B.12}$$

$$D_2(\beta) = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^3} \tag{B.13}$$

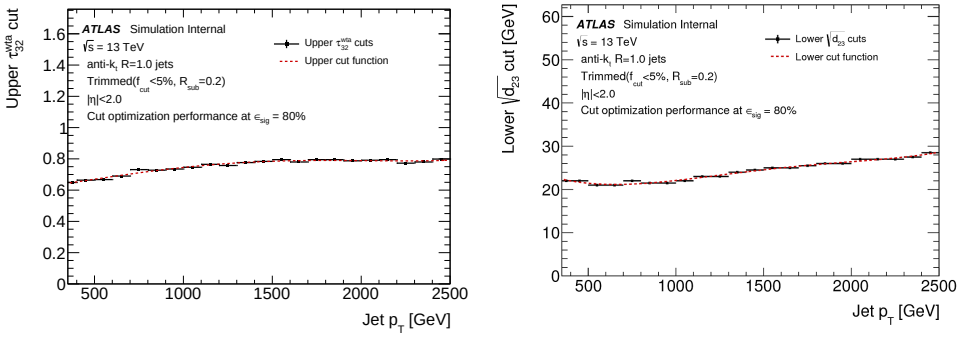
B.3 Minimum Invariant Mass

Q_W [40] is simply defined as the minimum of the invariant masses of the combinations of three strongest subjets, as shown by Equation B.14. The subjets are obtained by deconstructing the jet using the k_T jet-finding algorithm.

$$Q_W = \min(m_{12}, m_{23}, m_{31}) \tag{B.14}$$

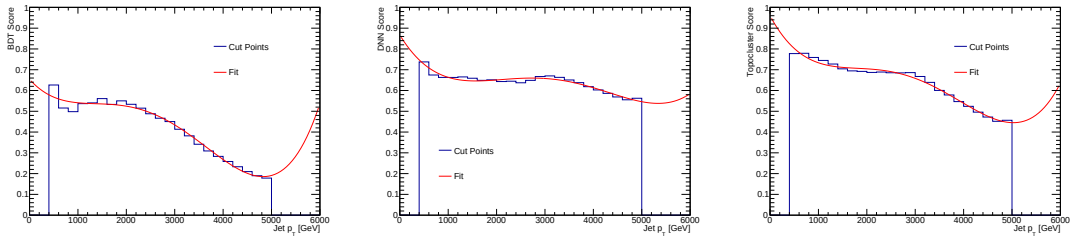
C

Tagger Cut Functions

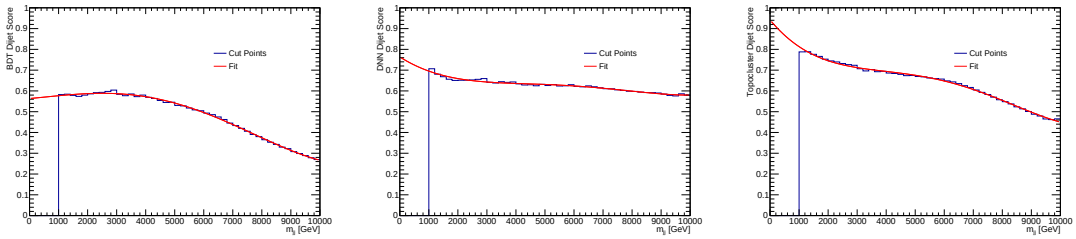


(a) Upper τ_{32} cut **(b)** Lower $\sqrt{d_{23}}$ cut

Figure C.1: Upper and lower cut functions used by the Smoothed Top Tagger on tagging individual jets to get constant working point of 80% signal efficiency over jet p_T.



(a) BDT Tagger **(b)** DNN Tagger **(c)** Topocluster Tagger



(d) BDT Tagger **(e)** DNN Tagger **(f)** Topocluster Tagger

Figure C.2: Cut functions used by the three machine learning taggers on tagging individual jets (top) and tagging dijet events (bottom) to get constant working point of 80% signal efficiency over p_T and m_{jj}.

D

Signal Morphing

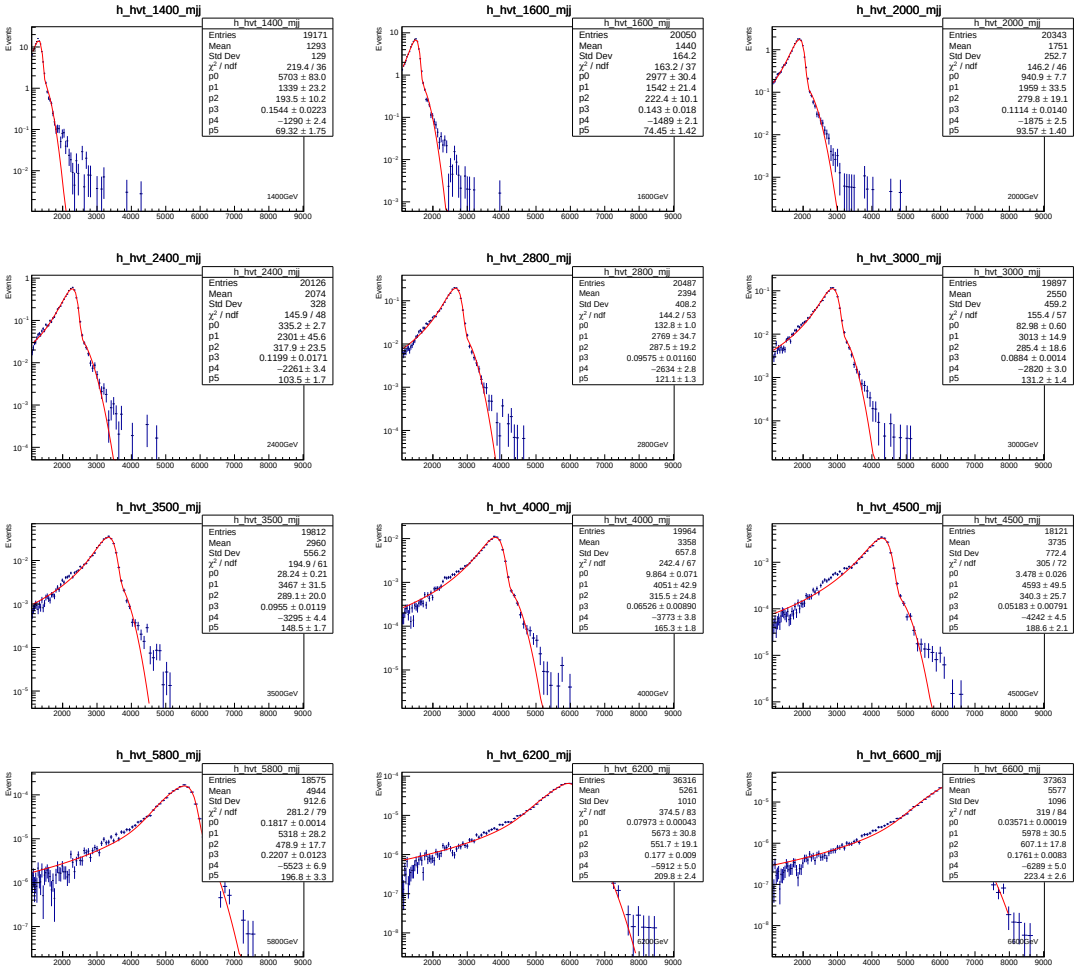


Figure D.1: GrL fits on the HVT signal before top-tagging at various mass points.

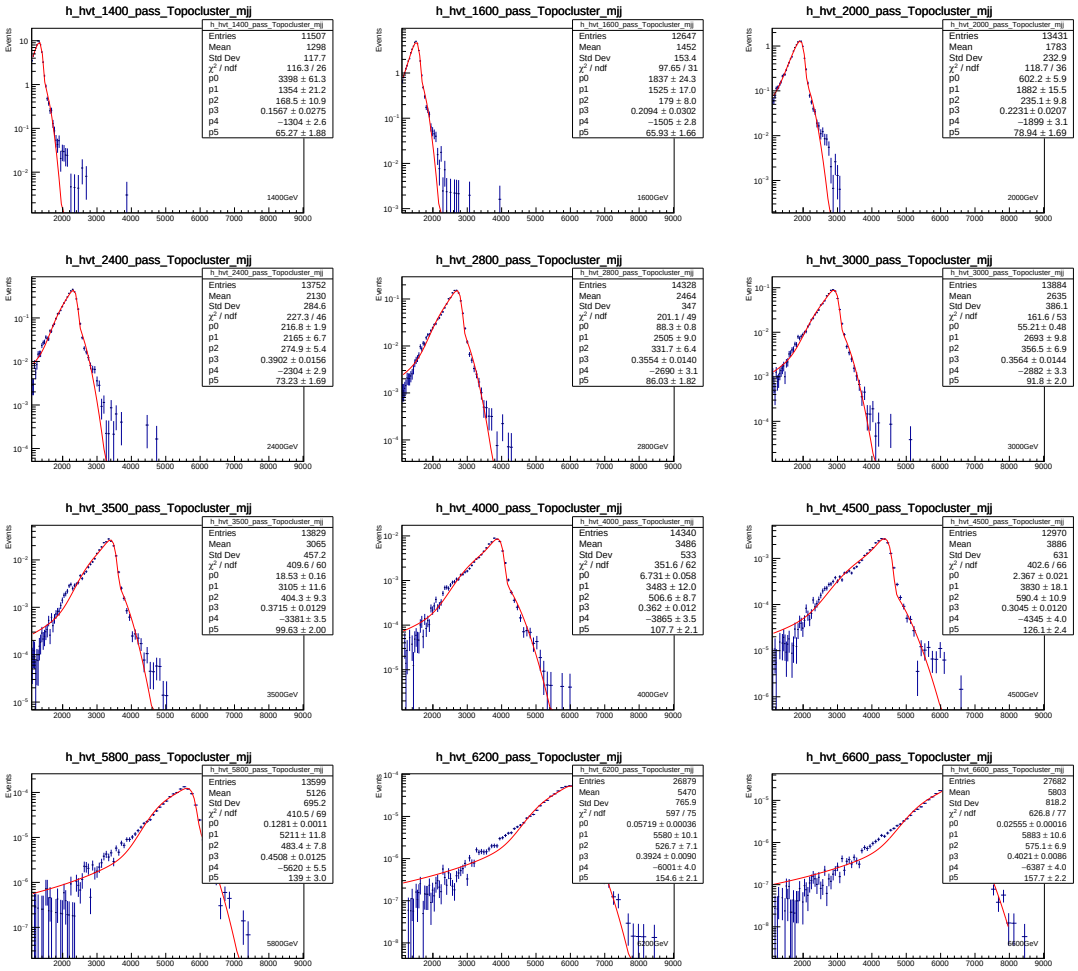


Figure D.2: GrL fits on the HVT signal after top-tagging at various mass points.

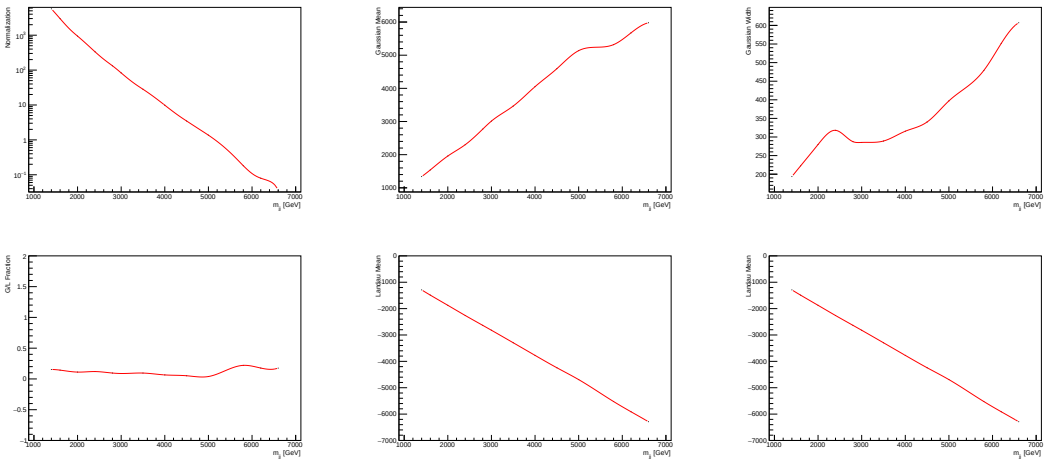


Figure D.3: Cubic spline interpolation of the GrL parameters corresponding to Figure D.1

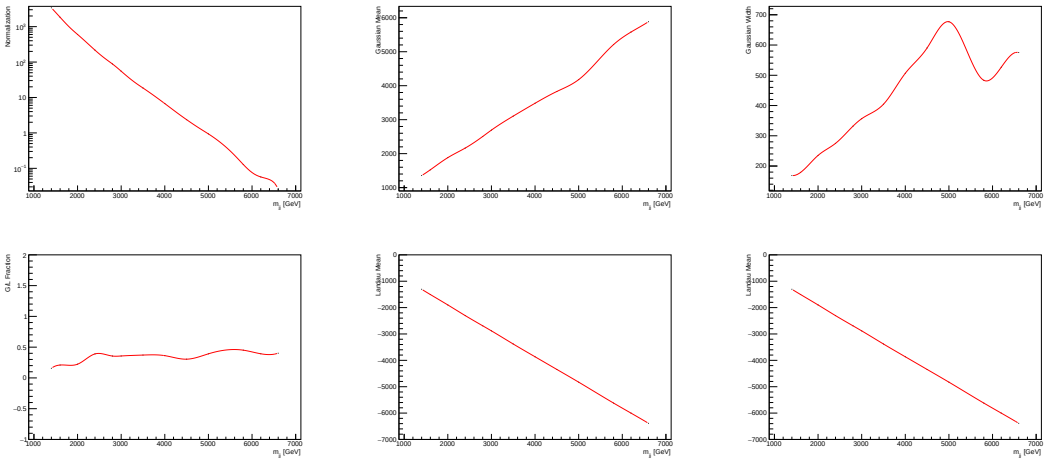


Figure D.4: Cubic spline interpolation of the GrL parameters corresponding to Figure D.2

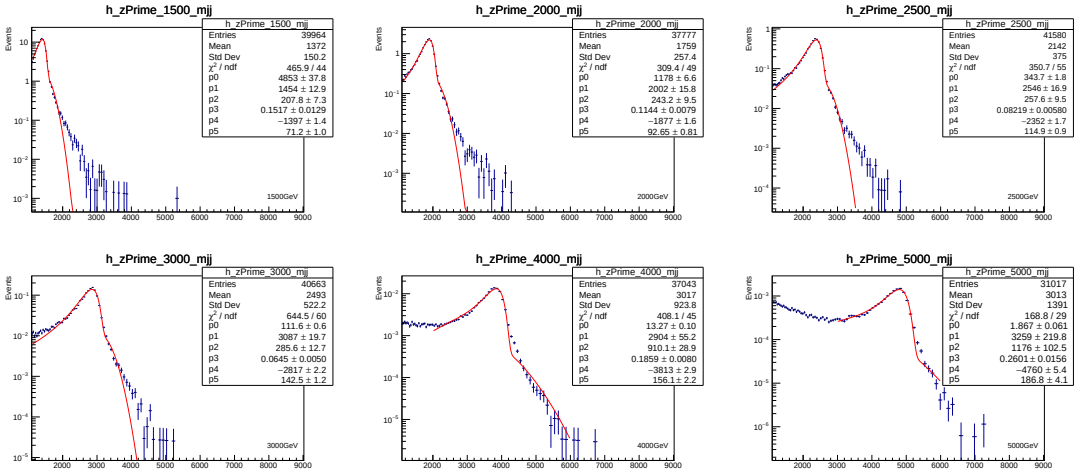


Figure D.5: GrL fits on the Z' signal before top-tagging at various mass points.

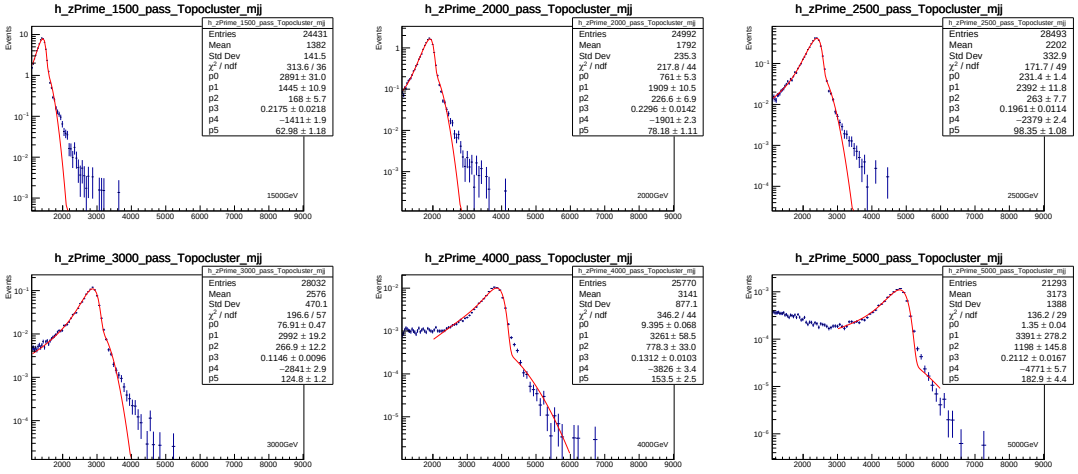


Figure D.6: GrL fits on the Z' signal after top-tagging at various mass points.

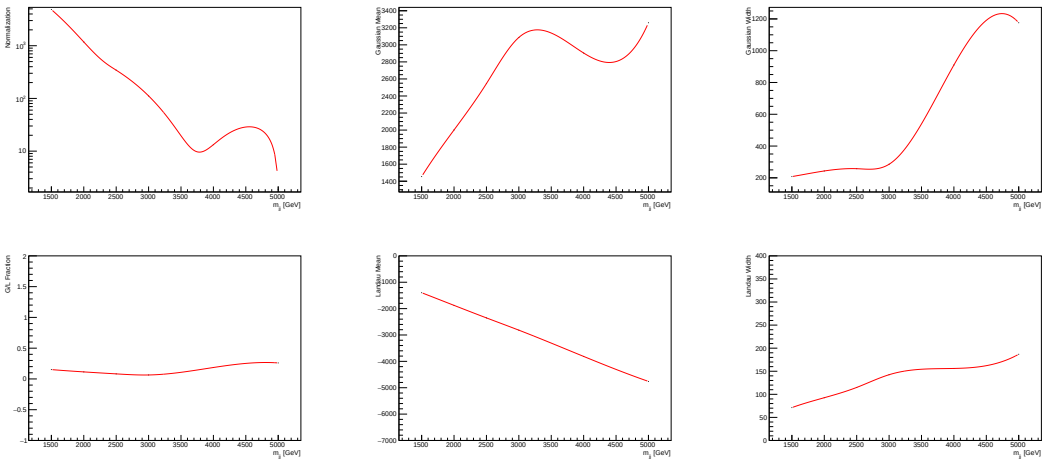


Figure D.7: Cubic spline interpolation of the GrL parameters corresponding to Figure D.5

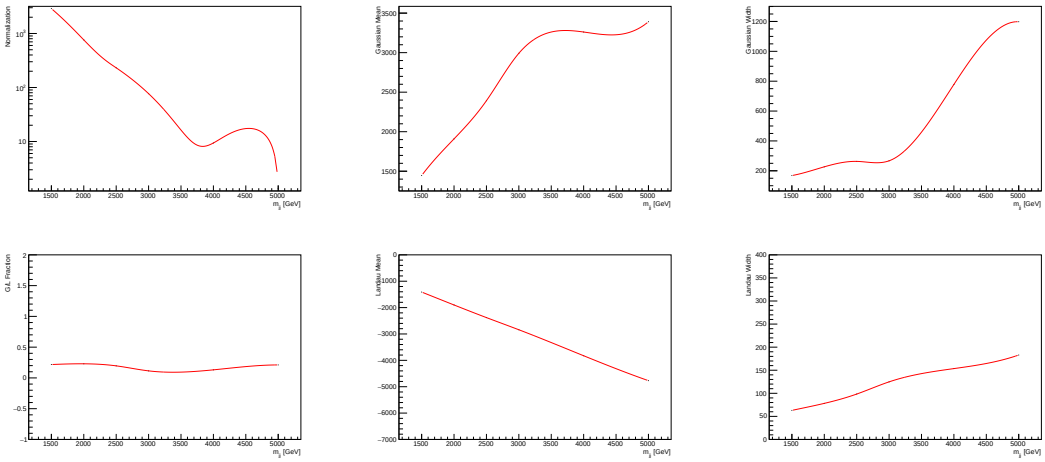


Figure D.8: Cubic spline interpolation of the GrL parameters corresponding to Figure D.6

Bibliography

[1] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995.

[2] D. Galbraith and C. Burgard, "Diagram of standard model of particle physics." http://united-states.cern/physics/standard-model-and-beyond. Accessed: 2018-04-03.

[3] K. A. Olive *et al.*, "Review of Particle Physics," *Chin. Phys.*, vol. C38, p. 090001, 2014.

[4] G. Kane, *Modern Elementary Particle Physics*. Cambridge University Press, 2 ed., 2017.

[5] L. Randall and R. Sundrum, "Large mass hierarchy from a small extra dimension," *Phys. Rev. Lett.*, vol. 83, pp. 3370–3373, Oct 1999.

[6] T. Farooque and P. Sinervo, "Search for Heavy Resonances Decaying to Top Quark Pairs in the Boosted All-hadronic Decay Channel," Jun 2013. Presented 23 May 2013.

[7] B. Lillie, J. Shu, and T. M. P. Tait, "Kaluza-Klein gluons as a diagnostic of warped models," *Phys. Rev. D.*, vol. 76, p. 115016, Dec. 2007.

[8] D. Pappadopulo, A. Thamm, R. Torre, and A. Wulzer, "Heavy vector triplets: bridging theory and data," *Journal of High Energy Physics*, vol. 2014, p. 60, Sep 2014.

[9] G. Altarelli, B. Mele, and M. Ruiz-Altaba, "Searching for New Heavy Vector Bosons in $p\bar{p}$ Colliders," *Z. Phys.*, vol. C45, p. 109, 1989. [Erratum: Z. Phys.C47,676(1990)].

[10] "Diagram of CERN's accelerator complex." http://public-archive.web.cern.ch/public-archive/en/research/AccelComplex-en.html. Accessed: 2018-04-03.

[11] J. Pequenao, "Computer generated image of the whole ATLAS detector." Mar 2008.

[12] "Cross-sectional view of the atlas detector." http://atlasexperiment.org/multimedia/. Accessed: 2018-04-03.

[13] P. Pasuwan, "Searches for low-mass dijet resonances in trijet final states using Monte Carlo samples," May 2016.

[14] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics," *Phys. Rept.*, vol. 97, pp. 31–145, 1983.

[15] ATLAS Collaboration, "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1," *ArXiv e-prints*, Mar. 2016.

[16] L. Bryngemark, *Search for new Phenomena in Dijet Angular Distributions at* $\sqrt{s} = 8$ *and 13 TeV*. PhD thesis, Lund University, 2016.

[17] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions," *Physical Review D*, vol. 48, pp. 3160–3166, Oct. 1993.

[18] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *Journal of High Energy Physics*, vol. 4, p. 063, Apr. 2008.

[19] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms," *Journal of High Energy Physics*, vol. 8, p. 001, Aug. 1997.

[20] E. E. Boos, O. E. Brandt, D. S. Denisov, S. P. Denisov, and P. D. Grannis, "The top quark (20 years after its discovery)," *Physics Uspekhi*, vol. 58, pp. 1133–1158, Dec. 2015.

[21] J. Thaler and K. Van Tilburg, "Identifying boosted objects with N-subjettiness," *Journal of High Energy Physics*, vol. 3, p. 15, Mar. 2011.

[22] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins, and et al., "Measurement of k_T splitting scales in $W \rightarrow lv$ events at $\sqrt{s} = 7$ TeV with the ATLAS detector," *European Physical Journal C*, vol. 73, p. 2432, May 2013.

[23] "Performance of Top Quark and W Boson Tagging in Run 2 with ATLAS," Tech. Rep. ATLAS-CONF-2017-064, CERN, Geneva, Aug 2017.

[24] "Identification of Hadronically-Decaying W Bosons and Top Quarks Using High-Level Features as Input to Boosted Decision Trees and Deep Neural Networks in ATLAS at $\sqrt{s}$ = 13 TeV," Tech. Rep. ATL-PHYS-PUB-2017-004, CERN, Geneva, Apr 2017.

[25] J. Pearkes, W. Fedorko, A. Lister, and C. Gay, "Jet Constituents for Deep Neural Network Based Top Quark Tagging," *ArXiv e-prints*, Apr. 2017.

[26] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *Journal of High Energy Physics*, vol. 5, p. 026, May 2006.

[27] T. Sjöstrand, S. Mrenna, and P. Skands, "A brief introduction to PYTHIA 8.1," *Computer Physics Communications*, vol. 178, pp. 852–867, June 2008.

[28] D. J. Lange, "The EvtGen particle decay simulation package," *Nucl. Instrum. Meth.*, vol. A462, pp. 152–155, 2001.

[29] 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," *ArXiv e-prints*, Mar. 2017.

[30] P. de Jong, "Top physics at the lhc." Heavy Quarks and Leptons, Melbourne, 2008.

[31] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method," *JHEP*, vol. 0711, p. 070, 2007.

[32] "Luminosity public results run2." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2. Accessed: 2018-04-20.

[33] K. Sekhon and D. Amidei, "Search for New Particles Decaying to Hadronic Jets in Proton-Proton Collision at $\sqrt{s} = 13$ TeV," May 2018. Presented 04 May 2018.

[34] ATLAS Collaboration, "Search for New Particles in Two-Jet Final States in 7 TeV Proton-Proton Collisions with the ATLAS Detector at the LHC," *Physical Review Letters*, vol. 105, p. 161801, Oct. 2010.

[35] ATLAS Collaboration, "Search for new phenomena in dijet mass and angular distributions from pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector," *Physics Letters B*, vol. 754, pp. 302–322, Mar. 2016.

[36] G. Choudalakis, "On hypothesis testing, trials factor, hypertests and the BumpHunter," *ArXiv e-prints*, Jan. 2011.

[37] S. Egan, W. Fedorko, A. Lister, J. Pearkes, and C. Gay, "Long Short-Term Memory (LSTM) networks with jet constituents for boosted top tagging at the LHC," *ArXiv e-prints*, Nov. 2017.

[38] "Jet mass reconstruction with the ATLAS Detector in early Run 2 data," Tech. Rep. ATLAS-CONF-2016-035, CERN, Geneva, Jul 2016.

[39] A. J. Larkoski, G. P. Salam, and J. Thaler, "Energy correlation functions for jet substructure," *Journal of High Energy Physics*, vol. 6, p. 108, June 2013.

[40] J. Thaler and L.-T. Wang, "Strategies to identify boosted tops," *Journal of High Energy Physics*, vol. 7, p. 092, July 2008.