

Studies on Razor Trigger for Simplified Dark Matter Models
and
Studies on Boosted Top Jet Identification with a Deep
Learning Algorithm
using the ATLAS Detector

Ece Akilli

Master Thesis



Département de Physique Nucléaire et Corpusculaire

Geneva 2015

Abstract

Part I - Studies on Razor Trigger for Simplified Dark Matter Models The Standard Model of particle physics is proven to be a very successful model, however it cannot explain all the observed physical phenomena. Dark matter was hypothesized in order to explain some of these phenomena, specifically the observed anomalies at different astrophysical length scales, and it is one of the keys to understand the physics beyond the Standard Model. If dark matter interacts non-gravitationally with Standard Model particles, it could be produced at the LHC, however it would be invisible to the detectors. Searches for dark matter in the ATLAS experiment at Run-1 have focused until recently on missing transverse energy signals using the effective field theory. However, effective field theory has limitations at the LHC's energy reach and simplified dark matter models have been proposed to overcome these limitations. The so-called razor trigger is an ATLAS Run-2 trigger. The main goal of the razor trigger is to improve the acceptance for analyses which cannot rely on large missing transverse energy. In this thesis, we studied the razor trigger performance using simplified dark matter model samples at 8 TeV center-of-mass energy with the ATLAS detector.

Part II - Studies on Boosted Top Jet Identification with a Deep Learning Algorithm The top quark is the heaviest fundamental particle of the Standard Model. Being a unique particle, top quark plays an important role at studying the Standard Model and constraining models of new physics. LHC, with its designed luminosities and collision energies, is a top quark factory and processes involving top quark can be studied in great accuracy. Hadronically decaying boosted top quarks are particularly interesting for new physics searches as the all-hadronic decay channel of the top quark has the highest branching ratio and the signal to background ratio for such searches improves at high energies. However, conventional particle identification methods are not successful at identifying top quarks at such high energies as the decay products of the boosted top quark are highly collimated. Therefore, dedicated boosted techniques were developed and the resulting substructure variables are used to identify boosted top quarks. Combining the gained information from substructure analyses by using novel techniques can improve the top tagging performance. Deep neural networks are the state-of-the-art machine learning tools which extract more complex features of the input. In this thesis, we studied the performance of top tagging with a deep learning algorithm using simulated samples at 14 TeV center-of-mass energy with the ATLAS detector.

Acknowledgements

This work has been carried out in the Department of Nuclear and Particle Physics of the University of Geneva from September 2014 to 2015 under the direction of Tobias Golling. I would like to thank Giuseppe Iacobucci and my advisor Tobias Golling for accepting me as a student in the University of Geneva ATLAS Group.

I thank Sofia Vallecorsa for helping me on daily basis with the first part of this thesis and I thank the Harvard group, in particular Christopher Rogan, for accepting us to their razor trigger group and providing help, discussions.

I thank Tobias Golling for being a great advisor and Andrea Coccaro for his great supervision. I would like to thank Luke de Oliveira for sharing his expertise. I would also like to thank all the other people who regularly provided insightful discussions: Marie Lanfermann, Olaf Nackenhorts, Michela Paganini and Steven Schramm.

Contents

1	The Standard Model	1
2	ATLAS Experiment	2
2.1	LHC	2
2.2	The ATLAS Detector	2
2.2.1	Trigger System	3
I	Studies on Razor Trigger for Simplified Dark Matter Models	5
1	Theoretical Introduction	6
1.1	Dark Matter	6
1.1.1	Evidence of Dark Matter	6
1.1.2	Dark Matter Candidates and Searches	7
1.1.3	Simplified Dark Matter Models	9
2	Razor Trigger	10
2.1	Motivation	10
2.2	Razor Kinematical Variables	10
2.2.1	Razor Variables	10
2.2.2	Super Razor Variables	12
2.3	Razor Trigger	13
3	Razor Trigger for Simplified Dark Matter Models	14
3.1	Razor Variables	15
3.2	Trigger Overlaps for the Signal Samples	16
3.3	Trigger Efficiencies	18
3.3.1	Trigger Efficiency	18
3.3.2	Trigger Turn-On Curves	20
3.3.3	Trigger Efficiency at the Plateau	21
4	Conclusions	24
II	Studies on Top Jet Identification with a Deep Learning Algorithm	25
1	Theoretical Introduction	26
1.1	Top Quark	26
2	Top Tagging in ATLAS Experiment	28
2.1	Jets in ATLAS Experiment	28
2.1.1	Large Radius Jets	28
2.2	Jet Substructure	29
2.2.1	Substructure Variables	29

2.3	AGILEPack Deep Learning Framework	30
2.3.1	Neural Networks	30
2.3.2	AGILEPack Deep Learning Framework	32
3	Top Tagging with AGILEPack Deep Learning Framework	34
3.1	Training Statistics	36
3.2	Training Weights	38
3.2.1	Performance in p_T Bins	40
3.2.2	p_T Dependence of Performance at 50% Working Point	41
3.2.3	Performance in $ \eta $ Bins	43
3.3	Final results	44
4	Conclusions	45
A	Part I - Sample list	46
B	Part I - Trigger turn-on curves and parameterization	48

Chapter 1

The Standard Model

Our current understanding of particle physics is embodied in the Standard Model [1]. The Standard Model provides information about fundamental particles and their interactions in terms of particle properties and particle exchange between them [2]. In the Standard Model, the matter is made up of two types of particles. These two types of particles are fermions and bosons. In addition, there exists a corresponding antiparticle for each particle. Fermions have half integer spin whereas bosons have integer spin. The fundamental fermions of the Standard Model are quarks and leptons. The fundamental bosons of the Standard Model are the gauge bosons and the Higgs boson. Gauge bosons are the force carriers of the Standard Model and four different gauge bosons are responsible of three different interactions. These three interactions are strong, electromagnetic and weak interactions ¹. At least one spin-0 particle was postulated to be in the Standard Model, which was the only missing piece of it until recently. The Higgs boson, a spin-0 particle, was postulated with the Higgs field being responsible for the masses of the fundamental particles [3], [4] and a new particle compatible with the Standard Model Higgs boson was discovered in 2012 by the ATLAS and CMS collaborations [5].

The precision tests of the Standard Model and the discovery of the Higgs boson validate the Standard Model. Although the Standard Model is very successful, it cannot answer all the questions in particle physics and physics beyond the Standard Model attempts to explain the observations which cannot be explained by the Standard Model. Some of the topics that need explanation beyond the Standard Model are neutrino oscillations, CP violation, matter-antimatter asymmetry in the Universe, existence and nature of dark matter.

¹Gravity, being the weakest of all four interactions in nature, is not included in the Standard Model.

Chapter 2

ATLAS Experiment

2.1 LHC

The Large Hadron Collider (LHC) [6] at CERN is a two-ring superconducting hadron accelerator and collider installed in a 26.7 km tunnel located underground near Geneva. It is designed with the aim of revealing physics beyond the Standard Model by reaching high center-of-mass energies in collisions which have never been reached in colliders. In order to reveal physics beyond the Standard Model, the LHC is designed to collide proton beams with a centre of mass energy of 14TeV and an unprecedented luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. In addition to proton beams, it is designed to collide heavy ions (Pb) with an energy of 2.8TeV per nucleon and a peak luminosity of $10^{27} \text{ cm}^{-2} \text{ s}^{-1}$. Two counter rotating beams are injected into the LHC from the SPS accelerator, afterwards the ejected beams are accelerated while they move around the LHC ring. Finally the beams are collided at four interaction points, corresponding to the positions where four experiments are located. The four experiments are ATLAS, CMS, LHCb and ALICE.

During Run-1, which took place between 2009 and 2012, the LHC had successfully delivered proton-proton collisions at a center-of-mass energies of 7 TeV(in 2011) and 8 TeV(in 2012). The LHC also had delivered lead-lead and proton-lead collisions with a per-nucleon center-of-mass energies of 2.76 TeV and 5.02 TeV respectively. In February 2013, the Long Shutdown 1 of the LHC began. After the Long Shutdown 1, in June 2015 the LHC Run-2 successfully started to deliver proton-proton collisions with a center-of-mass energy of 13 TeV.

2.2 The ATLAS Detector

The ATLAS experiment is one of the two general-purpose, high-luminosity experiments of the LHC [7]. The ATLAS detector [8] is forward-backward symmetric with respect to the interaction point. It has cylindrical geometry and it covers almost the entire solid angle around the interaction point [9]. It has the following four main constituents.

- **Inner detector** is responsible of pattern recognition, momentum and vertex measurements for charged particles, and electron identification. It is immersed in a 2 T solenoidal field.
- **Calorimeter** is responsible of measurement of energy carried by the particles. It consists of distinct parts dedicated to electromagnetic and hadronic calorimeters. Calorimeter depth is chosen carefully so that the calorimeters provide good containment for electromagnetic and hadronic showers.
- **Muon spectrometer** is responsible of identification and measurement of momenta of the muons with high resolution. The muon spectrometer defines the overall dimensions of the ATLAS detector.

- **Magnet system** is responsible of bending charged particles for momentum measurement. It features a hybrid system that consists of four sub-systems(1 central solenoid, 1 barrel toroid, 2 end-cap toroids).

The detector layout is shown in Figure 2.1.

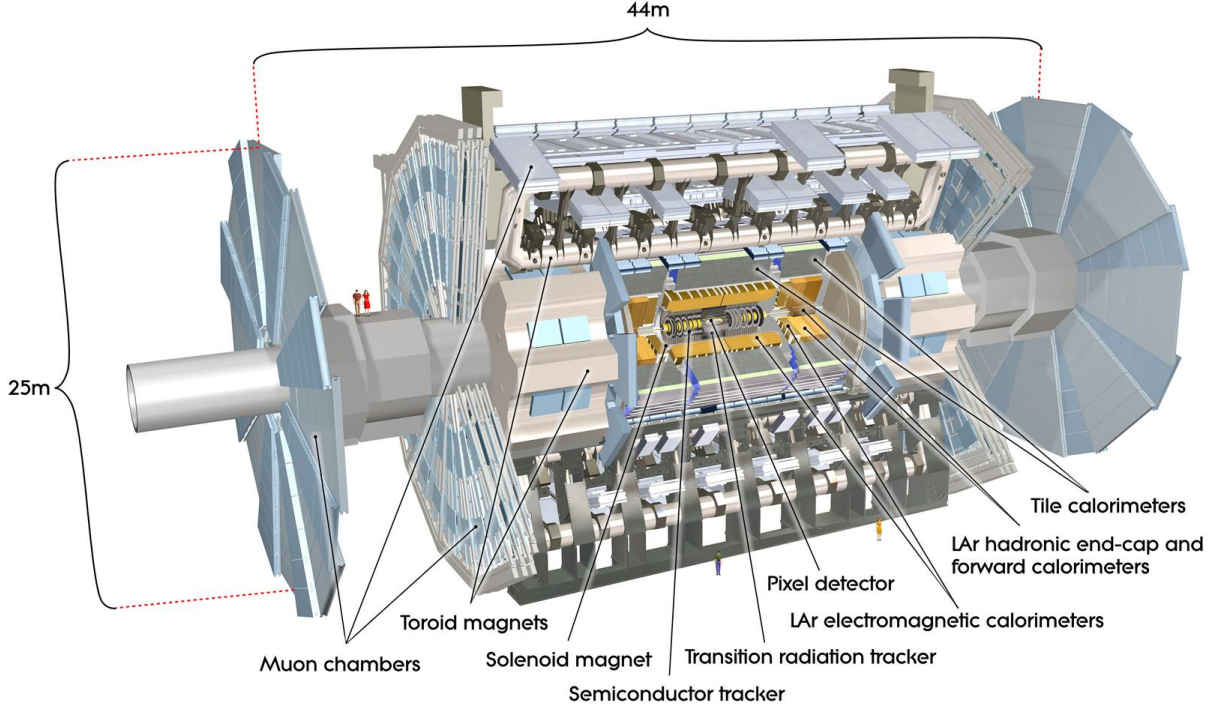


Figure 2.1: Cut-away view of the ATLAS detector [8]

2.2.1 Trigger System

One of the most challenging aspects of a general-purpose detector that captures high-luminosity proton-proton collisions is the trigger system. In the ATLAS detector, efficient triggering with low transverse momenta thresholds on electrons, photons, muons, tau leptons, jets and high flexibility on tagging jets are the criteria that needs to be fulfilled in order to provide high data-taking efficiencies for most physics processes of interest at the LHC [8].

In order to operate at the designed luminosity, LHC has a 40 MHz bunch crossing rate. The ATLAS trigger system is designed to reduce this input rate to an output rate of about 0.5-1 KHz for recording and offline processing [9]. The system has three levels, which are: the first level(L1), the second level(L2) and the third level (Event Filter,EF) triggers. The second level and the third level are called the High Level Trigger (HLT) collectively. The triggers are based on identifying combinations of candidate physics objects. In addition, there are triggers for inelastic proton proton collisions and triggers based on global event properties such as summed transverse energy and missing transverse energy.

The Level 1 Trigger is a hardware based system, it uses the information from the calorimeter and muon subdetectors. Its decision is formed by the Central Trigger Processor (CTP) based on information from the calorimeter trigger towers and dedicated triggering layers in the muon system. In addition to selecting the events which will be seeded to HLT, the L1 triggers also identify Regions of Interest (RoIs) within the detector to be investigated by the HLT. The High Level Trigger is a software based system and it uses the information from all subdetectors,

including inner detector hits, full information from the calorimeter and data from the precision muon detectors. The HLT selection is based on features reconstructed in these systems. Short execution time is a crucial criteria for the HLT algorithms, as the timing constraints need to be satisfied. A schematic of the ATLAS trigger system is shown in Figure 2.2.

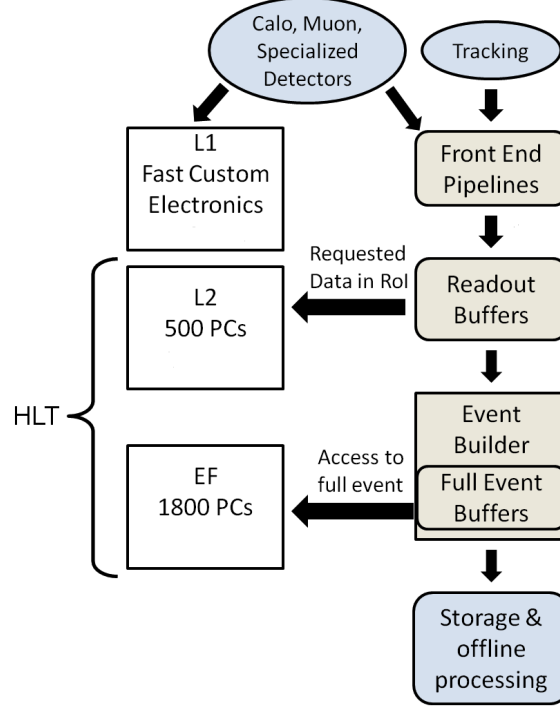


Figure 2.2: Schematic of the ATLAS trigger system [9]

In the first part of this thesis, we are interested in events which involve jets and missing energy. Specifically, the events selected by missing transverse energy(MET or E_T^{miss}) trigger and jet trigger are of interest. Here we provide brief information on these two trigger constituents. A summary of the performance studies that have been made for jet trigger and missing transverse energy trigger with the 7 TeV early LHC collision data can be found in [10].

Jet trigger Jets are the most common objects produced at the LHC. The jet trigger system of the ATLAS experiment is fundamental as jet triggers are the primary means for selecting events containing jets with high transverse momentum and physics analysis topics such as QCD, top quark physics, searches for SUSY, exotic models strongly depend on the jet trigger [11], [12].

Missing transverse energy trigger The weakly interacting particles which are produced at the LHC, such as neutrinos, escape the detectors without leaving a sign. Therefore they cannot be identified by none of the particle identification techniques which are used to identify physics objects such as leptons or jets resulting from hadronized particles. Instead, they are inferred by the significant momentum imbalance transverse to the beamline and the observable is referred as the missing transverse energy. Missing transverse energy triggers [13] are based on calorimetric measurements and they are used to select events with significant transverse momentum imbalance. Many of the beyond the Standard Model searches involve hypothetical weakly interacting particles [14]. Therefore, MET triggers have been used in searches for new phenomena such as processes involving new weakly interacting particles including dark matter [15], [16].

Part I

Studies on Razor Trigger for Simplified Dark Matter Models

Chapter 1

Theoretical Introduction

1.1 Dark Matter

Observations in large astrophysical systems with different scales indicate that the luminous matter is insufficient to account for the observed gravitational effects. Different hypotheses have been formulated to understand the nature of the unexpected observations. One of these approaches suggest that there exists a large amount of dark matter in the Universe. First suggestion for the dark matter which fits our current understanding was made by F. Zwicky. In 1933, after observing the Coma Cluster of galaxies F. Zwicky concluded that there must be invisible or dark matter in the Universe [17]. Another approach suggest that our understanding of the laws of gravitation and the theory of general relativity needs to be modified and the existence of dark matter could be avoided as M. Milgrom discusses in Reference [18]. Currently, the studies are focused mostly to the former approach and there is great effort in the search of dark matter as it is one of the greatest indications of physics beyond the Standard Model.

Additional interesting observations, which could be indicative of the existence of dark matter, come from the study of the positron fraction in the Cosmic Rays. Positron excess in the GeV range has been observed by PAMELA [19], FERMI [20] and AMS-02 [21]. Different scenarios, which include new astrophysical sources and the dark matter annihilation, have been examined to study the origin of the observed excess, studies resulted in limits in dark matter annihilation [22]. AMS-02 recently published the results of the 10.9 million positron and electron events. The observations showed that above ~ 200 GeV, the positron fraction no longer exhibits an increase with energy, which requires collection of more data in higher energies to study the origin of this behavior [23].

1.1.1 Evidence of Dark Matter

Several anomalies at different astrophysical length scales can be interpreted as evidence of dark matter [24].

- **Evidence from the galactic scale**

In 1970 a study on rotation of the Andromeda galaxy was presented by V. C. Rubin and Jr. W. K. Ford, where the rotational velocity curve of Andromeda showed a slow fall in the outer regions, different than our galaxy [25]. In 1980 a publication of a study on rotational properties of Sc galaxies¹ followed, which was conducted by V. C. Rubin, Jr. W. K. Ford and N. Thonnard. During their study, they observed that neither high nor lower luminosity Sc galaxies have falling rotation curves, although a fall in the rotation curves was expected after certain distances from the center of the galaxy, following the Keplerian decline. The study led to the conclusion that Sc galaxies must have significant mass located beyond the optical image [26]. This study had a great impact on astronomers to be convinced that

¹One class of spiral galaxies.

most of the matter in the Universe is dark and the dark matter is clumped around the galaxies. The effect of gravitational lensing also show evidence for dark matter in galaxies.

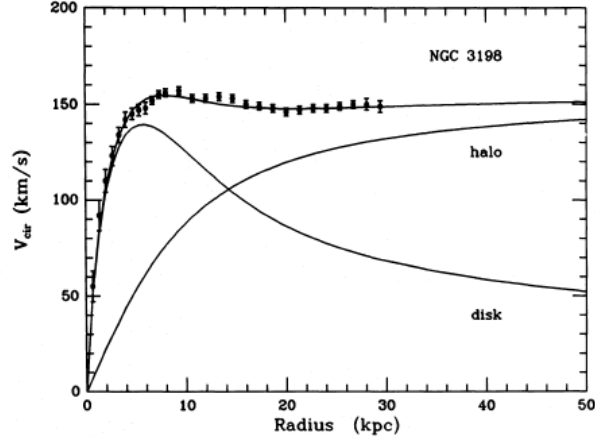


Figure 1.1: Rotation curve of NGC 3198 [27]

- **The scale of galaxy clusters**

As mentioned previously, first evidence of dark matter in our current understanding originated from the observation of dispersion of velocities in the Coma cluster. F. Zwicky discussed the observed dispersion of velocities under different considerations and he inferred the presence of large amount of dark matter [17].

- **Cosmological scales**

Cosmic Microwave Background (CMB) radiation is one of the predictions of the Big Bang model and it was first observed in 1965 by Arno Penzias and Robert Wilson [28]. CMB temperature distribution over all the sky is very uniform as it was predicted by the Big Bang model. However, there are very small fluctuations in the temperature distribution and these fluctuations give information about the origin, evolution, and content of the Universe. Full mission Planck observations of temperature and polarization anisotropies of the CMB radiation data gives the most precise results on the cosmological parameters which were published by ESA and the Planck collaboration. The Planck collaboration determined the parameters for the standard spatially-flat six-parameter Λ CDM cosmology, by analyzing the Planck temperature data combined with Planck lensing. The analysis results in a matter density parameter $\Omega_m = 0.308 \pm 0.012$ ² where the matter content is dominated by the cold dark matter³ [29].

1.1.2 Dark Matter Candidates and Searches

The studies on dark matter predict the dark matter to be non-baryonic, non-relativistic, stable and electrically neutral [30]. There are several candidates which satisfy the dark matter criteria. Some candidates of dark matter are: sterile neutrinos, axions, weakly interacting massive particles (WIMPs), superheavy dark matter(WIMPzillas). Amongst these candidates, WIMPs have been widely studied.

WIMPs are particles which have masses between approximately 10 GeV and a few TeV, and with interaction strength in the order of the weak interaction. These two features with addition of the correct abundance make WIMPs good dark matter candidates [31]. WIMPs as dark matter candidates are considered to be produced in the hot early Universe and then annihilated in pairs. Due to this scenario, the interactions of WIMPs with Standard Model particles kept them in thermal equilibrium at the high temperatures in the early Universe. As the Universe

²Parameter in 68 % confidence limit.

³ $\Omega_b h^2 = 0.02226 \pm 0.00023$, $\Omega_c h^2 = 0.1186 \pm 0.0020$ in 68 % confidence limit.

expanded and cooled, the rate of formation and annihilation eventually became too low, and the WIMP abundance froze out [30]. The WIMPs which survive to the present are known as thermal relics. Such particles are generically predicted in models of physics beyond the Standard Model, including models with supersymmetry or extra spatial dimensions [32]. Some of the weakly interacting massive particles which are predicted in physics beyond the Standard Model are:

- Supersymmetric (SUSY) candidates such as neutralinos, sneutrinos, gravitinos, axinos,
- Dark matter from Little Higgs models,
- Light scalar dark matter,
- The lightest Kaluza-Klein particle in extra dimensions.

Dark Matter Searches

Although the effects of dark matter’s gravitational interaction with baryonic matter have been observed by different experiments, non-gravitational interaction of dark matter with particles has not been observed yet. Consequently, whether the dark matter interacts with other particles non-gravitationally or not is still an unanswered question. In order to detect and understand the nature of the dark matter, the effort should include different sets of experiments so that possible interactions with different particles can be spanned. The three main groups of such experiments are direct searches, indirect searches and collider searches. The direct searches of dark matter are looking for the elastic scattering of target nucleus by the dark matter particles traveling in the Universe. These searches are mainly carried out in the underground detectors, in order to avoid the background. The indirect searches of dark matter are searching for the annihilation products of dark matter. Observations of gamma ray photons and cosmic rays by telescopes, satellites, neutrino satellites are good means of studying dark matter indirectly. The collider searches are based on the assumption that the dark matter interacts with the Standard Model particles and therefore can be produced at high energy collisions.

WIMP Dark Matter Searches at the LHC and ATLAS

If the dark matter interacts non-gravitationally with the Standard Model particles and if it is sufficiently light, dark matter particles should be produced at the LHC. As the WIMPs are neutral, non-baryonic, stable and weakly interacting, the produced WIMP dark matter would not be visible at the detectors. Therefore, the missing energy signals are the key to search for the dark matter signals at the particle colliders. At the ATLAS experiment many of the WIMP dark matter searches are using the *Mono-X* events. *Mono-X* events consists of a particle X and the consequent missing energy, where the X stands for photons, jets or a decaying boson. Examples of such studies of the ATLAS collaboration can be found in [33], [15] and [34] respectively. There are also studies which are looking for mono-lepton events, for example [35]. In this project, we focused to events where the dark matter pair production is accompanied by a jet,

$$pp \rightarrow \chi + \bar{\chi} + jet$$

where χ represents the dark matter particle. The mono-jets signals at the LHC have been extensively studied in the scope of Effective Field Theory(EFT) which resulted in constraints on the models [36], [16], [34], [15].

In the EFT framework at the colliders, WIMP is the only new particle that can be produced within the LHC’s energy reach. The EFT of dark matter describes the WIMP pair production accompanied by a SM particle radiation, where the partons interact via a heavy mediator which is heavier than the typical momentum exchanged in the process and cannot be produced directly at the LHC [37], [16]. In this scenario, the heavy mediator can be integrated out and the process can be described by contact operators. However, this approach has limitations at the LHC as the energies involved can be very high and the processes can occur at energies beyond the validity of the EFT itself [38]. Therefore, the validity of EFT has been studied⁴ and the simplified dark

⁴See, for example References [16], [38], [39], [40].

matter models have been proposed and studied where a light mediator is introduced, rather than being integrated out [16], [41].

1.1.3 Simplified Dark Matter Models

The simplified dark matter model samples which we used in this study are based on the Reference [16] where the dark matter is assumed to be a Dirac fermion χ . In the framework of simplified dark matter model, a light propagating Z' like particle is introduced to mediate the interaction of matter and dark matter. The parameters of such an interaction include the dark matter mass(m_χ), mediator mass(m_M) and the couplings of the mediator to the Standard Model particles(g_χ) and the dark matter (g_{SM}). As these two couplings are in principle constrained by m_M and the mediator width(Γ), Γ is used to tag the samples.

LHC Run-2 is reaching very high energies, consequently the possible limitations of EFT are becoming more significant. In addition to overcoming the limitation due to high energies, the comparison studies on the EFT and the simplified dark matter models suggest possible improvements by the usage of simplified models [16]. Recently the analyses which use the simplified models have gained momentum [41]. Therefore, we studied the razor trigger for simplified dark matter models.

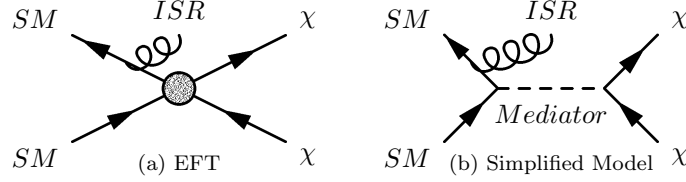


Figure 1.2: Feynman diagrams for mono-jet events where the jet is an ISR particle

Chapter 2

Razor Trigger

Razor trigger in the ATLAS Experiment is dedicated to exotics searches and it is based on the super razor variables [42]. The idea behind the razor trigger was to implement a new trigger to improve the acceptance for analyses that cannot rely on large missing transverse energy or very high transverse momentum jet selection and to improve the event selection flexibility compared to Run-1 in the ATLAS experiment. Razor triggers by the CMS collaboration has been used successfully in analyses such as Reference [43].

2.1 Motivation

As emphasized previously, searches for the phenomena beyond the Standard Model which involve new weakly interacting particles in general use the missing transverse energy triggers. However, the searches depending on the missing transverse energy have limitations. Not only measuring the missing transverse energy is very challenging, but also interpreting a signal with missing transverse energy is very challenging as most observables related to detailed features of the events are missing [44]. As a result of these, there have been strong efforts to interpret these signals by development of kinematical variables [14], [45]. Razor variables and the super razor variables are such kinematical variables and the super razor variables are used as selection criteria at the high level trigger of the razor trigger in ATLAS.

2.2 Razor Kinematical Variables

2.2.1 Razor Variables

Two kinematical variables, M_R and M_{R^*} , have been introduced in Reference [14] by Christopher Rogan to distinguish new massive strongly interacting particles from QCD background, which were predicted to be helpful to also reduce the dominant electroweak background $Z(\nu\nu)+\text{di-jets}$. These variables' distributions contain information about the masses of the pair produced particles and the weakly interacting particles resulting from produced particles' decays. Additionally, the dimensionless variables R and R^* have been introduced, which can be used to select interesting events in the presence of large backgrounds.

The variable M_R is defined by considering the following process, which is shown in figure 3.1. Two massive particles S_1 and S_2 (both with mass M_S) are produced by a hard partonic subprocess in a hadron hadron collision, the decay of these massive particles into respectively Q_1, χ_1 and Q_2, χ_2 follow. The particles S_i 's decay products Q_i are massless, stable particles which can be detected by the detector, whereas χ_i are massive (with mass m_χ), stable particles which escape the detector without leaving a sign.

Different reference frames, different particle pairs' center of mass frames, the individual particles' rest frames and the lab frame, are used to get more information about the system. The extracted information from these reference frames, helps to reduce the missing degrees of freedom

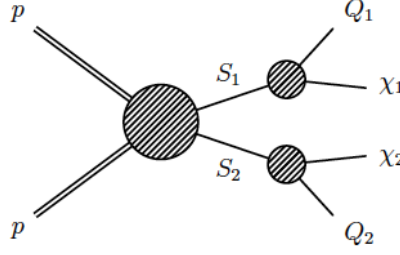


Figure 2.1: \mathbb{Z}_2 symmetry motivated event signature [14]

in the event. Although in this section the considered process involves the visible decays each consist of a single effectively massless particle, the decays may include more than one visible particle. In that case, the visible objects are grouped into two visible objects called megajets by summing the visible decay products' four momenta.

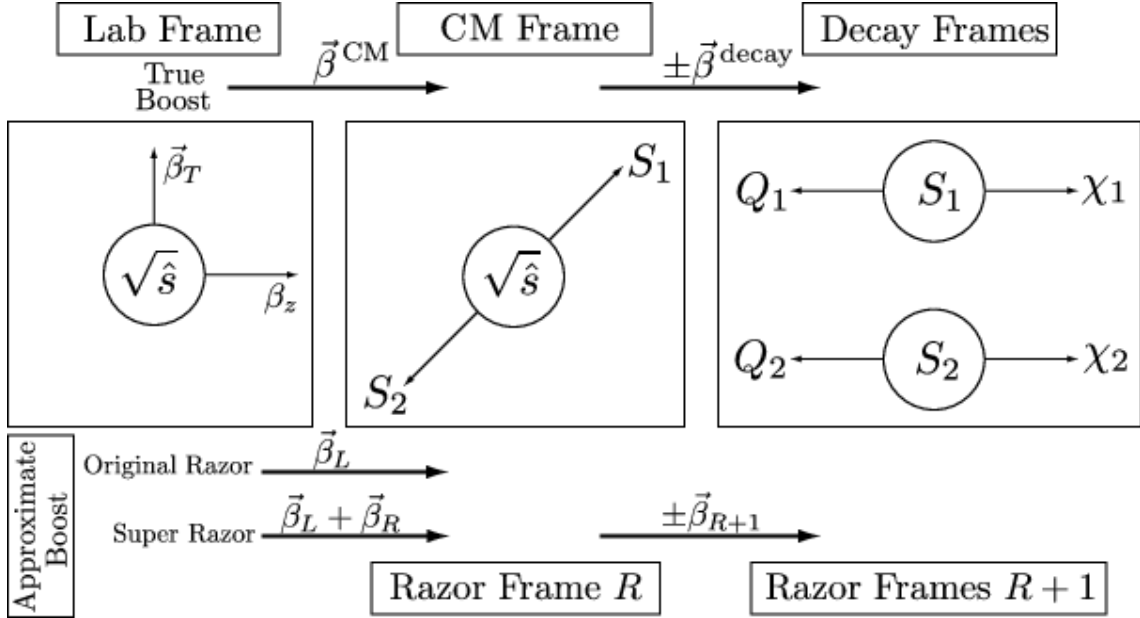


Figure 2.2: The three sets of frames relevant to the razor reconstruction [42]

If it was possible to reconstruct the rest frames of the S_i decays, the corresponding energies (E_i) of the visible particles Q_i in the rest frames of the S_i 's decays would be given by

$$2E_1 = 2E_2 = \frac{m_s^2 - m_\chi^2}{m_s} = M_\Delta$$

where M_Δ is the mass splitting. However, as at least one of the decay products escape the detector, there is not enough information to reconstruct the decay frames. The approach of the razor is to construct variables based on assumptions to approximate reference frames. One of the assumptions is, the heavy particles are generally produced near the threshold. This assumption results in the approximation that the CM of the hard process is also the decay frames of the particles S_i . Consequently, we can move from the laboratory frame to the CM frame by finding the longitudinal boost to a reference frame where Q_i have equal and opposite z-component of momentum (q_i^z). This reference frame is called the razor frame and the relevant boost is $\beta_L = \frac{q_1^z + q_2^z}{E_1 + E_2}$, where E_i are the energies of the decay products. Just like in the in the rest frames of the S_i 's decays, the energies of the particles in this razor frame are expected to be $2E_{R1} \sim 2E_{R2} \sim M_\Delta$.

The variable M_R , which carries information about the mass splitting, is a longitudinally boost invariant mass defined in the R-frame as

$$M_R^2 = (E_1 + E_2)^2 - (q_1^z + q_2^z)^2.$$

A second mass variable M_T^R , the transverse mass term, is defined by using the visible and invisible transverse momenta in the event such that

$$(M_T^R)^2 = \frac{1}{2}[E_T^{miss}(q_{1T} + q_{2T}) - \vec{E}_T^{miss}(\vec{q}_{1T} + \vec{q}_{2T})].$$

The dimensionless variable that is used with M_R to distinguish the signal from the background is the ratio of these two mass terms,

$$R^2 = \left(\frac{M_T^R}{M_R}\right)^2.$$

The assumptions involved in defining the R -frame, M_R and R variables may not always hold. The studies show that the useful properties of the variable M_R are robust against considered variations [14]. However, some deviations from the assumptions related to the reference frames may result in unphysical boosts, which are greater than 1, resulting in an ill defined R -frame and M_R . Therefore, the variables M_R^* and R^* are introduced by using the always well defined R^* -frames¹.

There have been studies² that used razor variables. These studies showed that the distribution of razor variables for signal and the Standard Model background samples show different characteristics, leading to good discrimination powers. For example the study on razor variables and dark matter parameter space at the LHC reported in [47], used razor variables with the objective to discriminate the kinematics of beyond the Standard Model heavy pair production from the Standard Model backgrounds without any strong assumptions about the missing transverse energy spectrum or the details of the underlying decay chain. This study has showed that the signal efficiency was improved by the usage of razor variables which resulted in an improvement of about 40% in the direct detection cross section.

2.2.2 Super Razor Variables

The super razor variables [42], which contain information about the ratio of mass scales of the particles in the event, are introduced as improvements to the razor variables by introducing an additional class of visible particles to the objects Q_i . The additional class of particles are the particles that may come from the initial state radiation (ISR) or different processes irrelevant to the heavy particle decays. Introducing this class of particles results in a shift of the production frame by an additional boost that was not taken into account by the original longitudinal razor boost. In this case, the boost to the razor frame R is built by the summation of the longitudinal boost($\vec{\beta}_L$) and the additional boost($\vec{\beta}_R$).

\vec{J} is defined such that it is the sum of the momenta of all particles in this class

$$\vec{J}_T = -\vec{E}_T^{miss} - \vec{q}_{1T} - \vec{q}_{2T}$$

and the $\vec{\beta}_R$ is constructed as

$$\vec{\beta}_R = \frac{\{-\vec{J}_T, p_z^R\}}{\sqrt{|\vec{J}_T|^2 + |p_z^R|^2 + \hat{s}_R}}$$

where \hat{s}_R is the approximate center-of-mass energy. \hat{s}_R is constructed based on the assumptions that the invariant mass of the visible system is equal to the invariant mass of the invisible system and that the constructed variables do not depend on the unknown p_z^R .

¹For details please see Reference [14].

²See References [46], [47].

$$\frac{\hat{s}_R}{4} = \frac{1}{2}(M_R^2 + \vec{J}_T(\vec{q}_1 + \vec{q}_2) + M_R \sqrt{M_R^2 + |\vec{J}_T|^2 + 2\vec{J}_T(\vec{q}_1 + \vec{q}_2)}).$$

The variable \hat{s}_R carries information about the mass splitting M_Δ , just like the old razor variable M_R^2 , and the pair production scale $\sqrt{\hat{s}}$.

Next, the approximate boost to the decay frames of the S_i particles $\vec{\beta}_{R+1}$ is constructed. As the two decay frames have equal and opposite boosts from the pair production frame, $\vec{\beta}_{R+1}$ is defined as

$$\vec{\beta}_{R+1} = \frac{\vec{q}_{R1} - \vec{q}_{R2}}{E_{R1} + E_{R2}}$$

where \vec{q}_{Ri} are the 4-momenta of Q_i in the razor frame R so that the boost has the correct symmetry.

A variable which approximates the M_Δ for signal events is defined by using the \hat{s}_R and the Lorentz factor γ_{R+1} associated with the boosts $\vec{\beta}_{R+1}$

$$M_\Delta^R = \frac{\sqrt{\hat{s}_R}}{2\gamma_{R+1}}.$$

Studies which compares the true values and the approximated values and how the variables can be used to distinguish signal and background can be found in Reference [42].

In addition to the mass variables, two variables $\Delta\phi_R^\beta$ and $|\cos\theta_{R+1}|$ related to the kinematic angles are introduced in order to capture more information about the event. In this study we have not used the angle variables and only their physical meanings are given for completeness. $\Delta\phi_R^\beta$ is the azimuthal angle between the razor boost $\vec{\beta}_R$ and the sum of visible momenta calculated in the razor frame R, it carries information about ratio of masses of the pair produced particles and their daughters. $|\cos\theta_{R+1}|$ is a dimensionless variable that carries information about the energy difference between two visible particles.

The super razor variables M_Δ^R , \hat{s}_R , $\vec{\beta}_R$, $\vec{\beta}_{R+1}$, $\Delta\phi_R^\beta$ and $|\cos\theta_{R+1}|$ can be thought of as a kinematic basis. Excluding the explicit requirements on MET and making selections based on this kinematic basis can help to increase sensitivity to models with small values of M_Δ .

2.3 Razor Trigger

The razor trigger that we studied uses two super-razor variables which are \hat{s}_R and γ_{R+1} in their forms used to define M_Δ^R . If we look at the two dimensional distributions of $\sqrt{\hat{s}_R}$ and $\frac{1}{\gamma_{R+1}}$ for signal and background, the shapes of the distributions are expected to be different for WIMP pair production signals and the Standard Model backgrounds. A study carried out by the Harvard group has shown that this is the case for several pair production events such as di-squark production with direct decays to quarks and the lightest supersymmetric particles, the dark matter pair production modeled by the Effective Field Theory through D5 operator. Following the observation of the two dimensional $\sqrt{\hat{s}_R}$ and $\frac{1}{\gamma_{R+1}}$ distributions of the simulated WIMP pair productions with a center-of-mass energy of 8 TeV and the collected data at the same energy, the Harvard group suggested a hyperbolic razor trigger on this distribution. The razor trigger applies a hyperbolic cut on the $\sqrt{\hat{s}_R}$ and $\frac{1}{\gamma_{R+1}}$ distribution at the HLT which is seeded by different first level jet and missing transverse energy requirements. The hyperbolic cut is in the following form,

$$\Pi = (\sqrt{\hat{s}_R} + C_2)\left(\frac{1}{\gamma_{R+1}} + C_2\right)$$

where the parameters Π , C_1 and C_2 can be tuned.

Promising results of the studies motivated the performance studies of this razor trigger on the dark matter pair production signals based on simplified dark matter models.

Chapter 3

Razor Trigger for Simplified Dark Matter Models

In this project, we studied the performance of the razor trigger on dark matter pair production signals based on simplified dark matter models with a Z' like mediator which couples the Standard Model particles to the dark sector. In order to do so, we used the corresponding simulated MC12 mono-jet samples with a center-of-mass energy of 8 TeV¹. This set of samples covers different dark matter and mediator mass (m_χ and m_M respectively) points. In addition to different mass points, for each mass point two different mediator widths(Γ) are covered. Vector coupling scenario is covered for each mass point whereas axial coupling scenario is covered for selected mass points. In addition, the samples were produced with two different matching scale cuts(QCUT80 and QCUT300) which we combined according to the leading jet p_T in the event² for this analysis. The generator used for the simulations was MadGraph, Pythia was used for parton showering and hadronization. The paper that presents the theories and corresponding models can be found in Reference [16]. The trigger information is retrieved from the simulated samples, skimmed trigger ntuples are produced from SUSY D3PDs, razor variables are constructed at the HLT by using Harvard group's trigger code [49] with minor changes. In this thesis, we focus to the samples with $\sqrt{s} = 8\text{TeV}$ although we also studied several samples with $\sqrt{s} = 14\text{TeV}$. We observed that in 14 TeV samples, we loose sensitivity as large fraction of the phase-space we were hoping to gain is not included in these samples. The reason of the loss in 14 TeV samples is, in 14 TeV samples there is a minimum leading parton p_T requirement, which is set to be 250 GeV. This requirement results in an enhanced leading jet p_T and MET distribution relative to an inclusive sample. On the contrary, 8 TeV samples do not have this strong generator-level requirement. Therefore, the results obtained by using 8 TeV samples are expected to be more indicative of potential gains.

A summarized list of the 52 QCUT-combined samples' properties is given below:

- **Vector coupling:** $\Gamma = \{\text{W}3, \text{W}8\pi\}$; $m_\chi = \{50, 400\}\text{GeV}$,
 $m_M = \{50, 100, 300, 600, 1000, 3000, 6000, 10000, 30000\}\text{GeV}$
- **Axial coupling:** $\Gamma = \{\text{W}3, \text{W}8\pi\}$; $m_\chi = \{50, 400\}\text{GeV}$,
 $m_M = \{50, 100, 300, 600\}\text{GeV}$

The full list of used samples is presented in Appendix A.

During this project, two different razor HLT parameter sets³ acting on several possible Run-2 L1 seeds⁴ were compared with the other MET and jet triggers. The razor triggers that we

¹These samples include some of the samples which were used in recent sensitivity studies, see Reference [48].

²If the leading parton p_T is below 350 GeV the events are taken from the QCUT80 samples, if it is above 350 GeV events are taken from QCUT300 samples.

³ c_1, c_2, Π

⁴Some of the considered L1 seeds were *2J15_XE55*, *2J10_XE60*, *XE70*, *4J20*.

studied were optimized taking into account rate(~ 35 Hz for $2 \times 10^{24} \text{cm}^{-2} \text{s}^{-1}$), background rejection and signal efficiency studies for several processes. Two razor triggers resulting from two different parameter sets acting on L1 seeds were *ProdR200-L1* with $c_1 = 85$, $c_2 = 0.048$, $\Pi = 200$ and *ProdR170-L1* with $c_1 = 200$, $c_2 = 0$, $\Pi = 170$. We analyzed the promising *ProdR200-L1*, *ProdR170-L1* with *2J15-XE55* and *XE70* L1 seeds in more detail and compared these triggers with the favored MET trigger *xe100-XE70*. We observed that although the two event filters(*ProdR200*, *ProdR170*) acting on the same L1 seed gave very similar results, event selection based on *ProdR170-L1* resulted in a slightly better signal efficiency. For simplicity, only the results of *ProdR170-L1* are presented here.

3.1 Razor Variables

In this section, two dimensional distributions of the variables $\sqrt{\hat{s}_R}$, $\frac{1}{\gamma_{R+1}}$ and the hyperbolic cut corresponding to the razor *ProdR170*⁵ are presented for two different samples where 2 jets with $p_T \geq 30$ GeV are required⁶. Two different mass points are chosen as examples of samples with vector coupling and samples with axial coupling. Although the distribution of the variables differ depending on the m_χ , m_M , Γ and slightly on the interaction type, in general the razor accepts most of the interesting events in the $\sqrt{\hat{s}_R}$, $\frac{1}{\gamma_{R+1}}$ phase space. We observed that when the difference between the dark matter mass and the mediator mass is smaller, the distribution is populated near smaller values of $\sqrt{\hat{s}_R}$ and near larger values of $\frac{1}{\gamma_{R+1}}$ as expected.

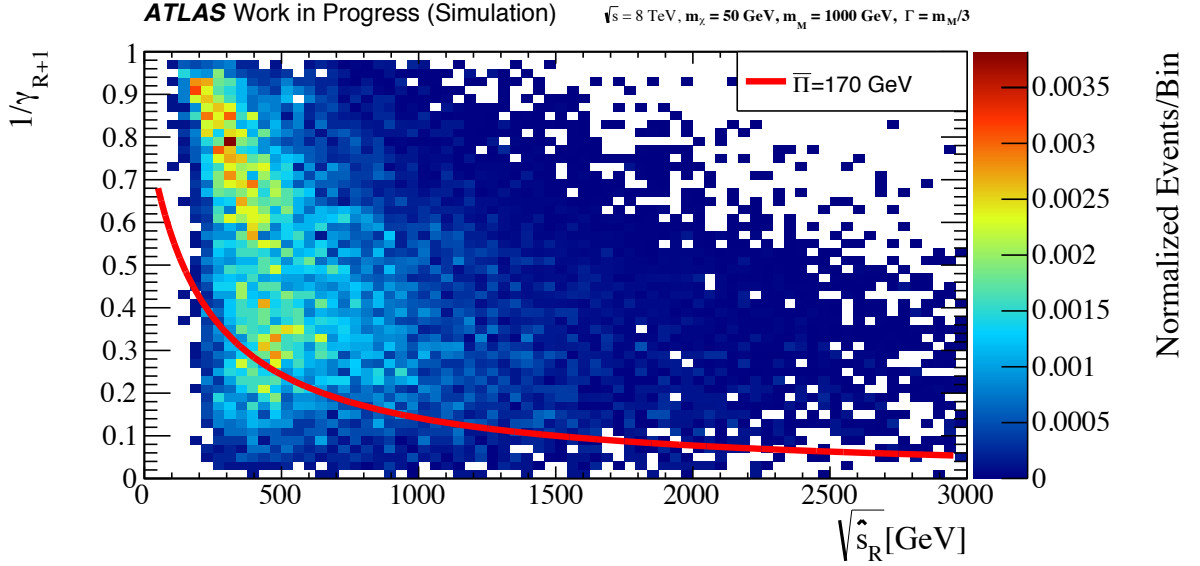


Figure 3.1: 2D Razor variables distribution for a vector coupling sample, *ProdR170*: $c_1 = 200$, $c_2 = 0$, $\Pi = 170$

⁵Without any L1 seed requirement.

⁶Razor variables are defined when there are at least 2 jets with $p_T \geq 30$ GeV in the event.

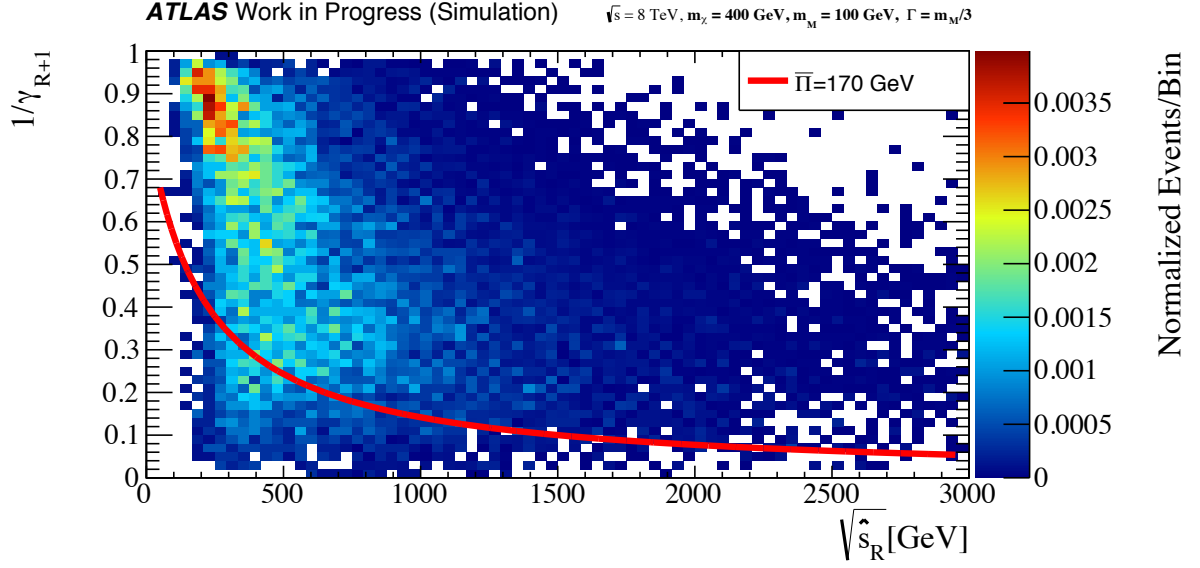


Figure 3.2: 2D Razor variables distribution for an axial coupling sample, ProdR170: $c_1 = 200$, $c_2 = 0$, $\Pi = 170$

3.2 Trigger Overlaps for the Signal Samples

Trigger overlap tables show what fraction of the events each Run-2 trigger can uniquely select relative to other triggers. Trigger overlaps which are presented in this section only consider the events which satisfy the criteria of having minimum 2 jets with $p_T \geq 30$ GeV and show percentage of events ($z \times 100$) which fail Trigger X and pass Trigger Y.

$$z = \frac{N_{Y,filtered}}{N_{filtered}}$$

where $N_{filtered}$ represents total number of events which satisfy the criteria of having minimum 2 jets with $p_T \geq 30$ GeV and $N_{Y,filtered}$ represents number of events which fail trigger X, pass trigger Y in addition to satisfying the same criteria as $N_{filtered}$. As an example, trigger overlaps of two chosen samples from the previous section are shown in the below figures.

We observed that the triggers *ProdR170_2J15_XE55*⁷ and *xe100_XE70*⁸ unique events relative to other razor and MET triggers from the signal samples. In particular, *ProdR170_2J15_XE55* selects up to 12% unique signal events relative to all jet and MET triggers, *xe100_XE70* selects up to 14% unique signal events relative to *ProdR170_2J15_XE55* trigger where only the *filtered* events are considered. Therefore, *ProdR170_2J15_XE55* appeared to be the favorable razor trigger for simplified dark matter model signals. In addition, studies by other groups showed that due to its desirable background rejection, signal efficiencies and rate of *ProdR170_2J15_XE55* for several other processes, *ProdR170_2J15_XE55* was a good candidate.

⁷Trigger *ProdR170_2J15_XE55* is the razor trigger which denotes 2 jets with the transverse energy threshold of 15 GeV accompanied with MET with the 55 GeV threshold at level 1, and with the HLT razor configuration *ProdR170*.

⁸Trigger *xe100_XE70* is the MET trigger which denotes MET trigger at level 1 with 70 GeV threshold, confirmed at high level trigger with 100 GeV threshold.

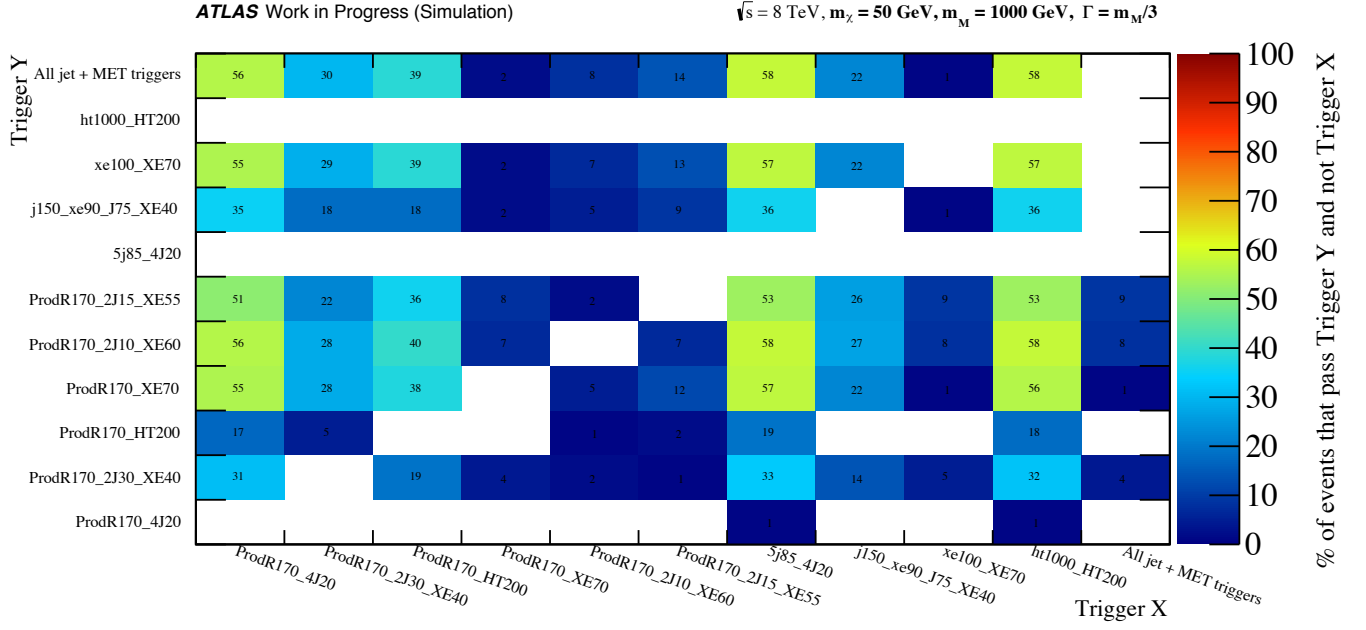


Figure 3.3: Trigger overlap for a vector coupling sample

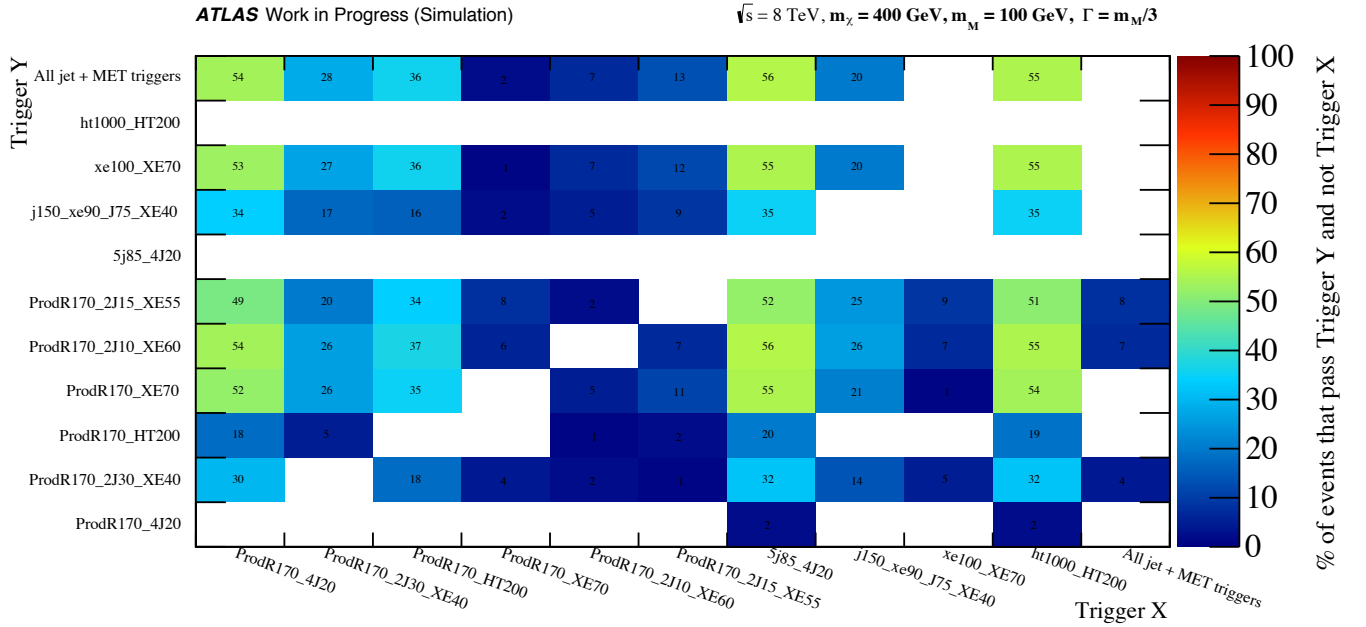


Figure 3.4: Trigger overlap for an axial coupling sample

3.3 Trigger Efficiencies

In previous sections we observed that the razor trigger *ProdR170_2J15_XE55* selects unique events compared to other triggers, including MET triggers which most of dark matter searches in the ATLAS experiment use. In order to study efficiency gained by adding razor trigger to the proposed Run-2 trigger menu, we studied the signal efficiencies of two promising triggers: *xe100_XE70* and *ProdR170_2J15_XE55*. As it was expected, we observed that razor trigger selects fewer number of events compared to MET trigger. Therefore, considering that razor trigger selects unique events relative to MET trigger, we studied the trigger efficiency of the MET alone (*xe100_XE70*) and the 'OR' combination of the razor and MET triggers (*ProdR170_2J15_XE55||xe100_XE70*). We calculated these trigger efficiencies separately for our 52 samples. However, for simplicity, here only the trigger efficiencies of samples with a vector coupling and with the mediator width $\Gamma = m_M/3$ are presented. These samples are chosen to be presented as the vector coupling samples have higher number of mass points compared to axial coupling samples and the efficiency gains of the samples with different couplings, mediator widths are only slightly different.

3.3.1 Trigger Efficiency

Efficiency of each trigger is defined as

$$\epsilon_{trigger} = \frac{N_{trigger}}{N_{total}}$$

where N_{total} represent the total number of events in the sample and $N_{trigger}$ represents the number of events which pass a specific trigger or pass at least one of the two triggers, *xe100_XE70* and (*ProdR170_2J15_XE55||xe100_XE70*) respectively. The errors on the efficiencies are calculated by error propagation as presented here.

$$\sigma_{\epsilon_{trigger}}^2 = \sigma_{N_{trigger}}^2 \left(\frac{\partial \epsilon_{trigger}}{\partial N_{trigger}} \right)^2 + \sigma_{N_{total}}^2 \left(\frac{\partial \epsilon_{trigger}}{\partial N_{total}} \right)^2$$

$$\sigma_{\epsilon_{trigger}} = \sqrt{\sigma_{N_{trigger}}^2 \frac{1}{N_{total}^2} + \sigma_{N_{total}}^2 \frac{N_{trigger}^2}{N_{total}^4}}$$

The samples that we used had event weights, consequently $N_{trigger}$, N_{total} are the weighted total number of events and corresponding errors are given by

$$\sigma_{N_{total}} = \sqrt{\sum w_i^2},$$

$$\sigma_{N_{trigger}} = \sqrt{\sum w_{i,trigger}^2}$$

where w_i and $w_{i,trigger}$ are the relevant event weights. As we are interested in events with minimum two jets with the razor approach and super razor variables are only defined for events which involve minimum 2 jets with $p_T \geq 30$ GeV, it is fair to compare the efficiencies with a filter imposing this condition. In this case, the efficiency is defined as

$$\epsilon_{trigger,filtered} = \frac{N_{trigger,filtered}}{N_{filtered}}$$

with the errors defined accordingly. Resulting efficiencies and relative errors are presented in tables below.

Table 3.1: Vector coupling, $\Gamma = m_M/3$, $\sqrt{s} = 8 \text{ TeV}$

m_M [GeV]	Efficiency of (<i>xe100_XE70</i>) [%]	
	m_χ [GeV]	
	50	400
50	33.7 ± 0.5	48.7 ± 0.6
100	31.7 ± 0.5	49.2 ± 0.6
300	42.1 ± 0.5	52.6 ± 0.6
600	47.1 ± 0.6	52.3 ± 0.6
1000	51.0 ± 0.6	53.2 ± 0.6
3000	50.8 ± 0.6	55.0 ± 0.6
6000	49.3 ± 0.6	55.0 ± 0.6
10000	49.2 ± 0.6	55.0 ± 0.6
30000	49.2 ± 0.6	54.7 ± 0.6

Table 3.2: Vector coupling, $\Gamma = m_M/3$, $\sqrt{s} = 8 \text{ TeV}$

m_M [GeV]	Efficiency of (<i>ProdR170_2J15_XE55</i> <i>xe100_XE70</i>) [%]	
	m_χ [GeV]	
	50	400
50	39.5 ± 0.5	53.0 ± 0.6
100	37.5 ± 0.5	53.6 ± 0.6
300	46.8 ± 0.6	56.9 ± 0.6
600	51.5 ± 0.6	56.7 ± 0.6
1000	55.3 ± 0.6	57.1 ± 0.6
3000	55.3 ± 0.6	58.7 ± 0.7
6000	54.0 ± 0.6	59.0 ± 0.7
10000	53.8 ± 0.6	59.3 ± 0.7
30000	53.8 ± 0.6	59.1 ± 0.7

Table 3.3: Vector coupling, $\Gamma = m_M/3$, $\sqrt{s} = 8 \text{ TeV}$, Filter: 2 jets with $p_T \geq 30 \text{ GeV}$

m_M [GeV]	Efficiency of <i>xe100_XE70</i> [%]	
	m_χ [GeV]	
	50	400
50	38.8 ± 0.8	55.3 ± 0.9
100	35.8 ± 0.8	56.2 ± 0.9
300	48.5 ± 0.9	60.7 ± 1.0
600	55.0 ± 1.0	60.3 ± 1.0
1000	57.7 ± 1.0	59.8 ± 1.0
3000	56.9 ± 1.0	62.3 ± 1.0
6000	56.9 ± 1.0	62.5 ± 1.0
10000	56.1 ± 0.9	64.0 ± 1.0
30000	56.0 ± 0.9	61.7 ± 1.0

Table 3.4: Vector coupling, $\Gamma = m_M/3$, $\sqrt{s} = 8 \text{ TeV}$, Filter: 2 jets with $p_T \geq 30 \text{ GeV}$

m_M [GeV]	Efficiency of $(ProdR170_2J15_XE55 xe100_XE70)[\%]$	
	m_χ [GeV]	
	50	400
50	52.3 ± 1.0	64.3 ± 1.0
100	49.6 ± 0.9	65.4 ± 1.0
300	59.8 ± 1.0	70.1 ± 1.1
600	65.1 ± 1.1	70.0 ± 1.1
1000	67.3 ± 1.1	68.7 ± 1.1
3000	67.2 ± 1.1	70.8 ± 1.1
6000	67.3 ± 1.1	71.4 ± 1.1
10000	66.3 ± 1.1	72.3 ± 1.1
30000	66.2 ± 1.1	71.3 ± 1.1

3.3.2 Trigger Turn-On Curves

Trigger efficiencies depend on constituents and physical properties of events. It is important to identify which offline conditions result in a stable trigger efficiency as it is essential for physics analyses to use triggers in regions where the triggers have stable efficiencies. Trigger turn-on curves show the trigger efficiency dependence on the related variables and they are used to identify the points where trigger efficiency reaches a *plateau*. In other words, they are used to determine the offline cuts imposed on events that pass the trigger. The MET trigger turn on is studied against the same quantity reconstructed offline while the razor trigger turn on is studied against $\sqrt{\hat{s}_R}, \frac{1}{\gamma_{R+1}}$. The desired trigger efficiency plateau is at 99% efficiency for MET and 93%⁹ for razor trigger. Statistics of 52 samples were not sufficient to study the trigger turn-ons individually. Fortunately, trigger turn-on curves for individual samples showed similar characteristics and we combined the 52 samples with respect to their cross sections in order to have higher statistics and evaluated turn-on curves with the combined sample. After identifying the trigger turn-ons globally, we cross checked and confirmed that the global turn-on curve models well the individual samples' turn-on curves.

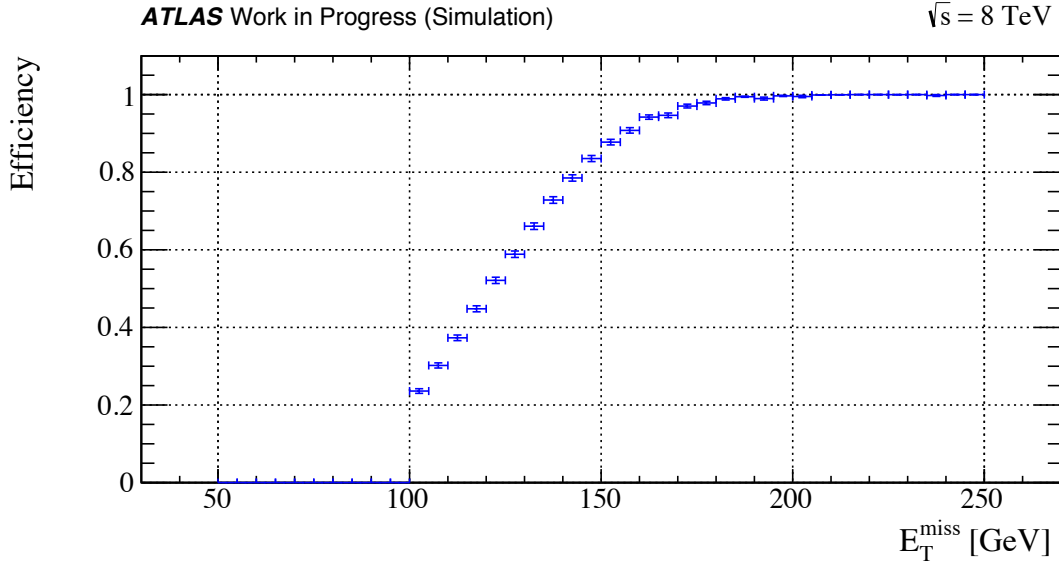


Figure 3.5: Efficiency of $xe100_XE70$

MET trigger curve for all events reaches the 99% efficiency plateau at $E_T^{miss} = 200 \text{ GeV}$. As

⁹Due to Gaussian distribution, 93% efficiency in 2D corresponds to 99% efficiency in 1D.

we studied the combination of MET and razor triggers, MET trigger turn-on is also evaluated for events where the razor variables are not defined due to the condition of having 2 jets with $p_T \geq 30 \text{ GeV}$ in the event and it behaves similar to global turn-on curve, please see Appendix B for this turn-on curve.

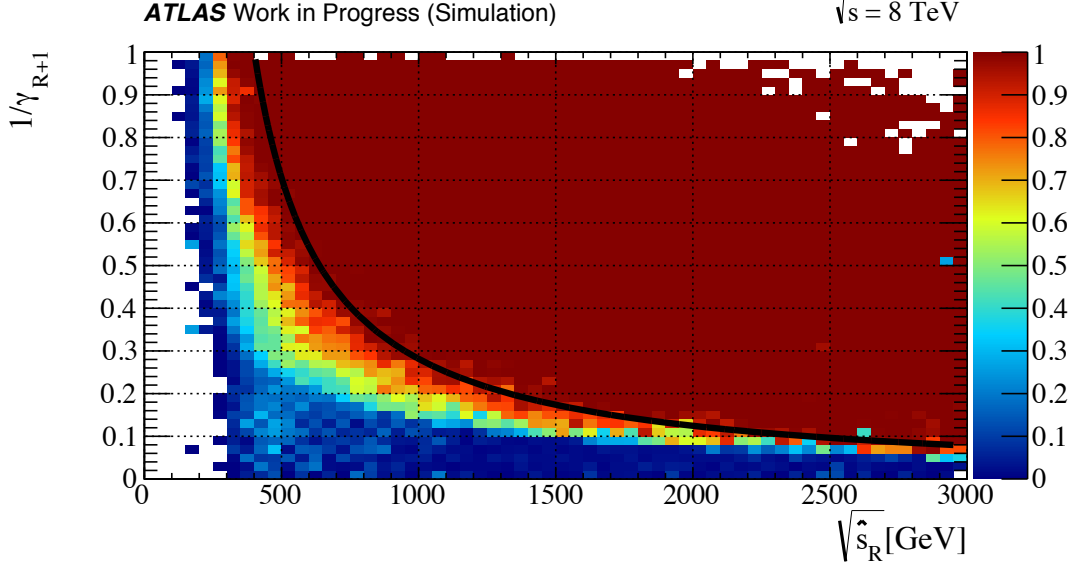


Figure 3.6: Efficiency of $(ProdR170_2J15_XE55 || xe100_XE70)$ and the resulting hyperbolic offline cut at $230 = (\sqrt{\hat{s}_R} - 170) + (\frac{1}{\gamma_{R+1}})$

Razor trigger efficiency per bin in $\hat{s}_R, \frac{1}{\gamma_{R+1}}$ space is shown in the figure above. We observed that the trigger turn-on curve follows an hyperbolic distribution. We followed the following empirical approach to parametrize the trigger turn-on curve in two dimensions:

1. Impose thresholds on $\frac{1}{\gamma_{R+1}}$,
2. Evaluate 1D efficiency and inefficiency curves in \hat{s}_R bins where the events pass the imposed thresholds on $\frac{1}{\gamma_{R+1}}$,
3. Retrieve the points which reaches %99 efficiency,
4. Fit the retrieved points to a hyperbolic function $p_0 = (\sqrt{\hat{s}_R} + p_1)(\frac{1}{\gamma_{R+1}} + p_2)$, fix p_2 to 0 in order to decrease the number of free parameters.

We observed that the efficiency reaches a plateau where the function is $230 = (\sqrt{\hat{s}_R} - 170) + (\frac{1}{\gamma_{R+1}})$ Please see Appendix B for more details and the related figures.

3.3.3 Trigger Efficiency at the Plateau

Following the trigger turn-on curves, offline cuts which let us work in the trigger efficiency plateau region were defined as $y \geq 230$ for the razor trigger where $y = (\sqrt{\hat{s}_R} - 170) + (\frac{1}{\gamma_{R+1}})$ and as $E_T^{miss} > 200 \text{ GeV}$ for the MET trigger. These offline cuts could be imposed on the 'OR' combination of the triggers $ProdR170_2J15_XE55$ and $xe100_XE70$ in several ways. We studied the following offline cut combinations.

1. Razor trigger offline cut condition is imposed on the trigger combination $(ProdR170_2J15_XE55 || xe100_XE70)$ resulting in selection $(ProdR170_2J15_XE55 || xe100_XE70)_{off}$

2. Razor trigger offline cut condition is imposed on $ProdR170_2J15_XE55$ and MET trigger offline cut condition is imposed on $xe100_XE70$, resulting in $(ProdR170_2J15_XE55)_{off}$, $(xe100_XE70)_{off}$ respectively and the combined 'OR' selection $\{(ProdR170_2J15_XE55)_{off} \parallel (xe100_XE70)_{off}\}$
3. If the razor variables are defined, razor trigger offline cut condition is imposed on the filtered trigger combination. If the razor variables are not defined, MET trigger offline condition is imposed on $xe100_XE70$. Resulting in $(ProdR170_2J15_XE55 \parallel xe100_XE70)_{filter,off}$, $(xe100_XE70)_{no\ razor,off}$ respectively and the combined 'OR' selection $\{(ProdR170_2J15_XE55)_{filter,off} \parallel (xe100_XE70)_{no\ razor,off}\}$

We observed that offline cut 2 is the one which results in the maximum efficiency gain at the plateau and consequently, results presented here are based on $\{(ProdR170_2J15_XE55)_{off} \parallel (xe100_XE70)_{off}\}$. Please keep in mind that, we didn't consider the background rejection of the trigger as this study is focused to signal efficiency of the trigger. However, in order to have a final decision on which offline cuts to apply, the background rejection of the trigger must be considered.

Table 3.5: Vector coupling, $\Gamma = m_M/3$, $\sqrt{s} = 8\ TeV$

m_M [GeV]	Efficiency of $(xe100_XE70)$ at the Plateau [%]	
	m_χ [GeV]	
	50	400
50	6.8 ± 0.2	21.7 ± 0.3
100	5.2 ± 0.2	22.4 ± 0.3
300	11.6 ± 0.2	23.5 ± 0.3
600	17.1 ± 0.3	22.8 ± 0.3
1000	21.0 ± 0.3	23.4 ± 0.3
3000	20.2 ± 0.3	25.0 ± 0.4
6000	19.7 ± 0.3	25.3 ± 0.4
10000	19.7 ± 0.3	25.3 ± 0.4
30000	19.8 ± 0.3	24.9 ± 0.4

Table 3.6: Vector coupling, $\Gamma = m_M/3$, $\sqrt{s} = 8\ TeV$

m_M [GeV]	Efficiency of $(ProdR170_2J15_XE55 \parallel xe100_XE70)$ at the Plateau [%]	
	m_χ [GeV]	
	50	400
50	9.5 ± 0.2	24.7 ± 0.4
100	7.6 ± 0.2	25.7 ± 0.4
300	14.9 ± 0.3	26.8 ± 0.4
600	20.6 ± 0.3	26.1 ± 0.4
1000	24.1 ± 0.4	26.6 ± 0.4
3000	23.6 ± 0.3	28.3 ± 0.4
6000	23.2 ± 0.3	28.4 ± 0.4
10000	22.9 ± 0.3	28.7 ± 0.4
30000	23.1 ± 0.3	28.4 ± 0.4

Table 3.7: Vector coupling, $\Gamma = m_M/3$, $\sqrt{s} = 8\text{ TeV}$, Filter: 2 jets with $p_T \geq 30\text{ GeV}$

	Efficiency of <i>xe100_XE70</i> at the Plateau [%]	
m_M [GeV]	m_χ [GeV]	
	50	400
50	10.0 ± 0.3	28.1 ± 0.5
100	7.4 ± 0.3	28.8 ± 0.6
300	16.1 ± 0.4	31.4 ± 0.6
600	23.7 ± 0.5	30.9 ± 0.6
1000	27.7 ± 0.6	30.7 ± 0.6
3000	26.2 ± 0.5	32.2 ± 0.6
6000	26.9 ± 0.6	34.0 ± 0.6
10000	26.8 ± 0.6	33.9 ± 0.6
30000	26.9 ± 0.6	32.8 ± 0.6

Table 3.8: Vector coupling, $\Gamma = m_M/3$, $\sqrt{s} = 8\text{ TeV}$, Filter: 2 jets with $p_T \geq 30\text{ GeV}$

	Efficiency of (<i>ProdR170_2J15_XE55</i> <i>xe100_XE70</i>) at the Plateau [%]	
m_M [GeV]	m_χ [GeV]	
	50	400
50	16.2 ± 0.5	34.6 ± 0.6
100	13.1 ± 0.4	35.8 ± 0.7
300	23.9 ± 0.6	38.8 ± 0.7
600	31.6 ± 0.6	38.3 ± 0.7
1000	34.6 ± 0.7	37.9 ± 0.7
3000	33.9 ± 0.7	39.6 ± 0.7
6000	34.7 ± 0.7	40.8 ± 0.7
10000	33.8 ± 0.7	41.3 ± 0.7
30000	34.2 ± 0.7	40.5 ± 0.7

Chapter 4

Conclusions

We studied the signal efficiency of the proposed razor triggers for simplified dark matter model samples by comparing the razor trigger and MET trigger combined efficiency (*ProdR170_2J15_XE55||xe100_XE70*) with the only MET trigger *xe100_XE70* efficiency. We observed that the razor trigger selects unique events relative to the MET trigger and the addition of the razor trigger results in a gain. The average gain without the offline cut applied is approximately 5% where all events are considered and approximately 10% where only the events with minimum 2 jets which satisfy $p_T \geq 30$ GeV are considered. The average gain at the 99% efficiency plateau is approximately 3% where all events are considered and approximately 7% where only the events with minimum 2 jets which satisfy $p_T \geq 30$ GeV are considered.

This study is focused to MC12 signal samples with 8 TeV center-of-mass energy. The reason for focusing to 8 TeV samples is the sensitivity loss due to the minimum leading parton p_T requirement in the 14 TeV MC12 samples. Please also note that we carried out and completed the study before the final decision on the razor trigger was made. Therefore, the results presented here are only indications of potential gains and a study with final razor triggers, new 14 TeV samples is required to have a final conclusion.

Part II

Studies on Top Jet Identification with a Deep Learning Algorithm

Chapter 1

Theoretical Introduction

1.1 Top Quark

Fundamental fermions of the Standard Model are categorized into groups by the types of interactions that they experience [1]. Quarks are one of the two fundamental fermions of the Standard Model, they carry non-integer electric charge and they are the only fermions that carry the *color charge*¹. Consequently, they are the only fermions which experience strong interaction in addition to weak and electromagnetic interactions. Due to color confinement of QCD, particles which carry color charges are not observed individually but observed in hadrons such as proton. Although they can't exist individually in nature, quarks and gluons can be produced freely at colliders. The top quark is one of the six quarks of the Standard Model and its discovery in proton-proton collisions was announced in 1995 by Fermilab [50], [51]. Since its discovery, the properties of the top quark are studied in detail². The top quark is the heaviest fundamental particle of the Standard Model with 173 GeV mass [55] and it carries +2/3 electric charge in addition to the color charge. Hence, top quark is very different than the other quarks. It is heavier than the W boson of the Standard Model, and it decays by first order weak interaction, corresponding to a very short lifetime, via

$$t \rightarrow W^+ + q$$

where q stands for *down*(d), *strange*(s), *bottom*(b) quarks and where the W boson then decays to a quark, antiquark pair or lepton accompanied by a neutrino [2] via

$$W^+ \rightarrow \bar{l} + \nu,$$

$$W^+ \rightarrow \bar{q} + q'.$$

The decay mode with a W boson and a b quark is the only significant decay mode with a branching ratio of 99% and the relevant Feynman diagrams are shown below.

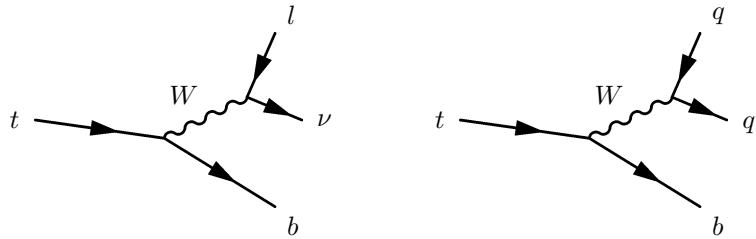


Figure 1.1: Feynman diagrams of significant top quark decays

QCD is the dominant process at the hadron colliders and the collisions at the LHC produce many quarks and gluons. Collision energies at LHC result in both high top pair production and

¹Color charge is the QCD equivalent of the electric charge of QED.

²For examples of such studies please see References [52], [53], [54], [55].

high single top quark production cross sections. Therefore, with its designed luminosities, LHC is a true top quark factory and processes involving top quark can be studied in great accuracy. Being a unique particle, top quark plays an important role at studying the Standard Model and constraining models of new physics. In particular, it has a special role in the context of electroweak symmetry breaking within and beyond the Standard Model [56].

Precise measurement of the top quark mass(m_t) and mass difference between top and antitop are very important to test the Standard Model as the former is a fundamental parameter of the Standard Model and latter is a test of the CPT invariance [57]. Some other tests of Standard Model with top quarks are on W polarization in top quark decays, perturbative QCD, chiral anomalies, flavor changing neutral currents, top-Higgs Yukawa coupling [58]. In addition to studying and testing the Standard Model, top quark plays an important role both as background and as the associated particle in various beyond the standard model searches. Supersymmetry(SUSY) is one of these extensions of the Standard Model. In SUSY searches at hadron colliders, the top quark plays an important role as top quark production is a dominant background for such searches and top quark can also be part of the signal, as the decay product of third generation squarks [59]. Other important beyond the standard model searches are $t\bar{t}$ resonances [60], vector like quarks [61], invisible exotic particles [62] searches.

Chapter 2

Top Tagging in ATLAS Experiment

Top quark is produced predominantly in hadron-hadron collisions through strong interaction and it decays rapidly. Identifying top quark by conventional particle identifications techniques fail at high energies as the decay products of top quark are highly collimated at such energies. In identifying the semi-leptonically decaying top quarks, one suffers from a lower branching ratio compared to hadronically decaying top quarks and the lepton isolation with respect to the b jet makes identification of semi-leptonic decaying top quarks challenging. Top tagging is a relatively recent approach to identify boosted hadronically decaying top quarks, which avoids reconstructing individual top decay products and instead uses large- R jets to reconstruct the entire top decay chain.

2.1 Jets in ATLAS Experiment

Quarks and gluons produced at the LHC are not what we observe in the LHC detectors due to color confinement and hadronization [63]. The gluon and quarks(except the top quark) hadronize, forming baryons and mesons which afterwards decay in many stages. Due to the high energy of the initially produced partons, decay products of the formed hadrons are highly boosted together and they form *jets*. Therefore, each parton¹ at the ATLAS detector is observed as a jet and the jets' measured properties can be linked to corresponding partons. Jets deposit large amount of energy to the calorimeters and they are usually defined by the jet recombination algorithms such as Cambridge-Aachen(C/A) [64], [65] and anti- k_T [66], amongst which anti- k_T is the default jet recombination algorithm in ATLAS.

Different than the other quarks, produced top quarks at the LHC decay in the ATLAS detector before hadronization, due to the short lifetime of the top quark. As it was emphasized before, the top quark almost exclusively decays through the single mode $t \rightarrow W^+ + b$ ². Hence, top quark is not observed as a jet but its decay products' signatures, specifically a b jet and decay products of W are observed.

2.1.1 Large Radius Jets

The center-of-mass energy of the LHC collisions has reached an unprecedented energy in the TeV scale. Such collision energies produce large samples of heavy particles with a transverse momentum that exceeds their rest mass, resulting in very high Lorentz boosts. Decay products of such boosted heavy particles are highly collimated at the detectors that classical reconstruction algorithms which rely on a one-to-one jet-to-parton assignment are often not successful, in particular for hadronic decays of such boosted objects [69]. Some examples of these decays are

¹Except top quark.

²For recent related measurements please see References [67], [68].

W,Z bosons' and top quark's hadronic decays. Consequently, techniques have been developed to overcome the limitations arising from these boosted objects at the LHC. In the context of large radius (large-R or fat) jets, jets are reconstructed with a larger radius parameter in order to capture the energy of the complete hadronic decay in a single jet. The substructure information of these large-R jets is used to identify heavy boosted object among the abundant jet production at the LHC.

In addition to reaching unprecedented energies, LHC is reaching high-luminosity phase and these two together increase number of pile-up events per bunch crossing. Therefore, dedicated jet grooming techniques have been developed to separate QCD jets from jets of boosted object decays and remove the pile up, soft radiation contamination in order to see the substructure of the large-R jets clearly in such environments.

2.2 Jet Substructure

As it was presented in the previous section, jet grooming techniques and jet substructure analysis is used to identify decay product jets in a large-R jet. In the ATLAS experiment three grooming techniques are used: filtering [70], trimming [71] and pruning [72]. Brief information about trimming is given below as trimmed large-R jets were used in this project.

Trimming The trimming procedure uses the constituents of the large anti- k_T jet formed with the large radius R and reclusters its constituents into smaller jets (subjets) with radius R_{sub} by using the k_T algorithm. The ratio of the p_T^i of subjets to the p_T of the large jet (p_T^{jet}) $f = p_T^i / p_T^{jet}$, is used as a selection criterion. If the ratio f is smaller than a chosen value f_{cut} , the subjet is rejected. The subjets which pass the selection criterion are recombined into a groomed jet. The trimming process is shown figuratively below.

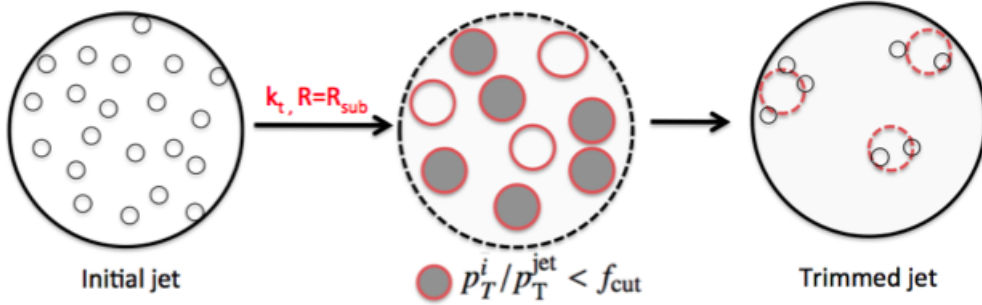


Figure 2.1: Trimming [73]

2.2.1 Substructure Variables

Substructure variables provide information about the characteristics and the substructure of the large-R jets. Boosted object tagging algorithms use different variables depending on the problem. Substructure variables which are generally used for top tagging in ATLAS experiment [74] are the trimmed mass of the large-R jet, N-subjettiness, k_T splitting scales.

Trimmed mass The mass of the trimmed jet is more pile-up and radiation safe compared to mass of the large-R jet without trimming. In addition, it provides good discrimination between hadronic top quark jets and QCD jets. Therefore, it is used in top tagging algorithms.

k_T Splitting scales The k_T sequential recombination algorithm's splitting scales are determined by clustering objects together according to their distance from each other. The substructure variables which are used for top tagging are determined by reclustering the constituents of

the trimmed large- R jet with the k_T algorithm and are defined by

$$\sqrt{d_{ij}} = \min(p_{Ti}, p_{Tj}) \times \Delta R_{ij}$$

where ΔR_{ij} is the distance between two subjets in (η, ϕ) space and p_{Ti}, p_{Tj} are transverse momenta of these two subjets. The k_T splitting scales give information about the last recombination steps and whether last recombination steps merge the decay products of massive particles.

N-subjettiness N-subjettiness variables are observables related to the subjet multiplicity. Each τ_N is a measure of how well jets can be described as containing N or fewer subjets. τ_N are defined by

$$\tau_N = \frac{1}{d_0} \sum_k p_{Tk} \times \Delta R_k^{min}, \quad d_0 \equiv \sum_k p_{Tk} \times R$$

where R is the jet radius parameter in the exclusive k_T algorithm, p_{Tk} is the p_T of constituent k and ΔR_k^{min} is the distance from constituent k to the axis of the closest subjet³.

Ratios of τ_N ,

$$\tau_{NM} = \frac{\tau_N}{\tau_M}$$

are also generally used in order to discriminate hadronic top jets from QCD jets as they give information about how likely the large- R jet has N or M subjets.

Top tagging by using substructure variables in the ATLAS experiment is proven to be successful during Run-1⁴. In particular, tagging algorithms which use set of substructure variables perform very well. There has been interest in combining gained information from substructure analyses in more complex ways in order to provide a better discriminant for boosted top quarks. In addition to existing algorithms, several multivariate analysis(MVA) based taggers are being studied. Some of these taggers use Toolkit for Multivariate Analysis(TMVA) [75] methods such as boosted decision trees, neural networks. In this project, we studied the performance of a deep neural network top tagger.

2.3 AGILEPack Deep Learning Framework

Machine learning techniques are used in different branches of science to handle complex problems and large datasets. There has been great interest in machine learning over the past decade which has led to highly optimized advanced algorithms. In particular, deep neural networks are studied extensively and now they are considered to be the state of the art tools which extract more complex features of the input.

As high energy physics requires analysis of huge amount of data, application of machine learning techniques can be useful at many different levels. There have been studies on applications of neural networks in high energy physics⁵ and flavor tagging algorithms based on neural networks are already being used in the ATLAS experiment [78]. It was shown that deep learning algorithms improve the discrimination between signal and background in physics analysis [79], however the usage of deep neural networks are not yet common in high energy physics. The AGILEPack deep learning framework applies modern deep learning techniques at a different level, which is the particle identification at the ATLAS experiment. Top tagging algorithms based on a previous version of AGILEPack were studied during Run-1 and showed significant performance improvements.

2.3.1 Neural Networks

A standard neural network (NN) consists of simple, connected processors that are called neurons [80] and the basic units of neural networks are defined by the activation functions acting on

³The axes within the jet are defined by exclusive k_T clustering algorithm.

⁴See Reference [74] for studies on the performance of used top tagging algorithms.

⁵See References [76], [77] for examples.

each neuron($f^{(i)}()$). Each neuron in the *input layer* is connected to the following layer's neurons by weights(W_0) and biases(b_0) of the neural network. Recursively, for each middle layer, the neurons are connected to following middle layer's neurons and the output of the current layer is received as input at the following layer, finally reaching the last layer (*output layer*) of the neural network. For any layer i , this can be shortly represented by

$$\mathbf{a}^{(i+1)} = f^{(i)}(W_i \mathbf{a}^{(i)} + b_i)$$

where $\mathbf{a}^{(i)}$ is the input received at layer i , (W_i, b_i) are neural network's weights and bias, $f^{(i)}$ is the activation function of layer i , $\mathbf{a}^{(i+1)}$ is the output delivered to layer $i+1$. Several properties which characterize neural networks are presented below.

- **Problem type: classification or regression**

Neural networks can be used for two different problems which are predicting a continuous valued output or predicting a discrete valued output. Former is called the regression problem, the second is called the classification problem. Top tagging is a classification problem as we want to predict if a large-R jet is a top quark (1) or not (0). Therefore, the neural networks we use for top tagging are classifiers.

- **Learning procedure**

If the neural network is trained to find weights which make the neural net map to the desired values (*target*) by minimizing the error between the output and the target, the training procedure is called supervised learning. Whereas, if no information about the target is used during the training, the training procedure is called unsupervised learning. Unsupervised learning [81] can be used as a first step (*unsupervised pre-training*) for the neural network's actual training via supervised learning.

- **Architecture**

Architecture of a neural network defines how many hidden layers it has, and how many neurons each layer consists of. Neural networks are categorized in two groups depending on how many hidden (middle) layers they have. Shallow networks have *few* hidden layers, whereas deep networks have *many* hidden layers. Definitions of *few* and *many* hidden layers are not strictly defined as they are changing fast in time following the fast evolving machine learning techniques and tools. In this study we used a neural network with 3 hidden layers and we refer to it as a deep network.

Deep Networks and Backpropagation Algorithm

Deep networks which are trained with supervised learning generally use backpropagation algorithms. Backpropagation algorithm uses the method of gradient descent. A backpropagation algorithm consists of two steps which is the feedforward step and the backpropagation of errors. The backpropagation algorithm can be represented by the following mapping and error minimizing steps.

The goal of the algorithm is to map the input layer represented by the vector \mathbf{x} to the target \mathbf{t} by the Neural Network \mathcal{N} :

$$\mathbf{x} \xrightarrow{\mathcal{N}} \mathbf{t}$$

where \mathcal{N} is

$$\mathbf{a}^{(i+1)} = f^{(i)}(W_i \mathbf{a}^{(i)} + b_i)$$

applied recursively at each layer at the feedforward step. The neural network's output results in $\mathbf{y} \equiv \mathcal{H}(\mathbf{x}, \mathcal{N})$ with the hypothesis \mathcal{H} . As the neural network cannot map the input to the target perfectly, the mapping results in an error $E(\mathbf{t}, \mathbf{y})$, which the neural network tries to minimize. In order to minimize the error, the algorithm first calculates the gradient of a cost function $(\partial\Phi/\partial\theta)$ ⁶ with respect to all the weights of the network. The gradient is then used to update the weights

⁶Cost function Φ is a chosen function of \mathbf{y} and \mathbf{t} , which gives information about the resulting cumulative error of up-to-now considered training instances.

aiming to minimize the cost function. Let us refer to the set of weights at time step t (W_t, b_t) as θ_t , then the update rule of a generic network is

$$\theta_{t+1} = \theta_t + \Delta\theta_t$$

with

$$\Delta\theta_t = (\mu - 1)\gamma \frac{\partial\Phi}{\partial\theta_t} + \mu\Delta\theta_{t-1}$$

where γ, μ are the training hyper-parameters, called the learning rate and the momentum respectively. The learning defines the step size of the update, and the addition of the momentum helps neural network to be more resistant to noise in the training sample. These hyper-parameters of the neural network need to be optimized depending on the problem and the statistics in order to achieve the best training of the neural network. Another important parameter, which in general is embedded in the cost function, is the regularizer. The regularizing term is a function of the regularizer and the training weights, it is added to the cost function in order to avoid very large weights and consequently overfitting.

Other variables of a training procedure are the number of epochs, which defines the number of passes over each event and the mini batch size which is the number of events to be considered before the algorithm updates the gradient.

In summary, the performance of a neural network strongly depends on the architecture, basic units of the neural network and hyper-parameters, regularizer, number of epochs, mini batch size of the training procedure.

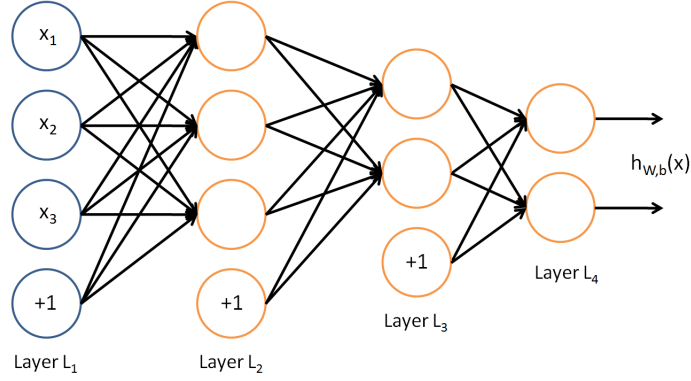


Figure 2.2: Example of a Deep Network

2.3.2 AGILEPack Deep Learning Framework

Algorithms for Generalized Inference, Learning, and Extraction Package (AGILEPack) is a C++ framework for deep learning which is designed for HEP purposes by Luke de Oliveira [82]. AGILEPack includes rectified linear units, sigmoid units and linear units as basic units. Like many other deep networks, AGILEPack uses a backpropagation algorithm.

The top tagging algorithm which is provided by the AGILEPack framework is called AGILETopTagger. A previously studied AGILETopTagger was the result of an unsupervised pretraining followed by a supervised training, by stacked denoising autoencoders. The unsupervised pretraining helped the neural network to be optimized to avoid getting stuck in a local extremum. The AGILEPack version that we used in this project is a new version of AGILEPack and it makes use of a recent, well studied optimization algorithm called Adam: A Method for Stochastic Optimization [83]. In addition, the rectifier linear units are recently included in AGILEPack, which helps networks to be trained efficiently without unsupervised pretraining. Therefore, only supervised learning with the new version of AGILEPack is used.

Training weights c_w are used in AGILEPack to define how much each jet updates the neural

network weights by acting as a coefficient to the learning rate, resulting in an effective step size by

$$\gamma \rightarrow \gamma^{eff} = c_w \gamma \quad \Rightarrow \quad \Delta \theta_t \rightarrow \Delta \theta_t^{eff}.$$

Choosing the optimal set of training weights is important as aggressively varying training steps would harm the training of a deep learning network.

Adaptive Moment Estimation(Adam): A Method for Stochastic Optimization [83]

Adam is an algorithm for first order gradient based optimization of stochastic cost functions, based on adaptive estimates of lower order moments. The method is aimed towards machine learning problems with large datasets or high dimensional parameter spaces. It only requires first order gradients and it computes individual adaptive learning rates for different parameters $(\theta : (W, b))$ from estimates of the first and second moment of the gradients. It naturally performs a form of step size annealing and it adjusts the two important training hyper-parameters: learning rate and momentum. Adam was tested on different machine learning problems and it provides good default settings for its hyper-parameters.

Chapter 3

Top Tagging with AGILEPack Deep Learning Framework

In this project we used simulated signal and background samples to study the performance of AGILETopTaggers. The usage of MC simulations, rather than the collected data, is necessary for purely supervised learning because the neural network needs to be trained with labeled large-R jets and it is challenging to extract a pure set of top quarks from data.

Samples In order to study the identification of high transverse momentum(boosted) top quarks with AGILEPack, we used the anti- k_T large-R jets with large radius parameter $R = 1.0$, subjet radius parameter $R_{sub}=0.3$ and $f_{cut} = 5\%$ trimming. The trimmed anti- k_T jets were geometrically matched with the condition $\Delta R \leq 0.75$ and they satisfied the criteria $p_T > 300$ GeV, $|\eta| < 2$. We used the corresponding jets from signal and background MC12 samples with a center-of-mass energy of 14 TeV. The signal sample that we used was $Z' \rightarrow t\bar{t}$, the background samples we used were JZ3W, JZ4W dijet samples¹. We applied the following pre-processing to this set of jets before training and testing the neural network. First, we prepared a full working sample by picking equal number of signal and background jets which satisfied the criteria $300 < p_T < 1000$ GeV². It is important to train the neural network with equal number of signal and background jets to avoid biases which may occur during the training due to statistical differences in samples. Afterwards, in order to train and test the neural network in separate samples, we used 70% of the total number of jets for training and 30% of the total number of jets for testing in each sample(resulting in total ~ 590000 training and ~ 250000 testing jets). It is important to point out that it is not always possible to construct all substructure variables with meaningful values, consequently there are some jets which carry unphysical substructure variables. These jets need to be taken care of during training and testing. In our samples, this was a small fraction of jets and we never included these jets in the training samples whereas we included them in the testing sample and introduced them as a source of inefficiency. As a last step of the pre-processing, we calculated the testing and training weights of the samples³. Testing sample weights were calculated by combining the background samples'(JZ3W and JZ4W) MC event weights with respect to their cross sections and filter efficiencies and fitting the signal sample's(Z') p_T distribution to the resulting combined p_T distribution of QCD background. The 1D reweighting of the signal sample to the background sample makes sure that the results on the performance won't be dependent on a particular signal shape. Resulting weighted number of testing jets were ~ 5450000 . Training sample weights(c_w) were calculated by applying several different reweightings as the optimal set of training sample weights were an open question. It is important to emphasize here that, different than the testing weights, the training weights don't affect the statistics of the training sample. Instead, they affect the update rule of the training

¹The initial samples were provided by Johannes Erdmann and Lily Asquith.

²We imposed this additional criteria to avoid very small or large weights.

³During the reweighting of the testing and the training samples, total number of weighted signal and background jets are kept equal.

procedure. Therefore, the effects of training with different reweighting scenarios were studied in this project.

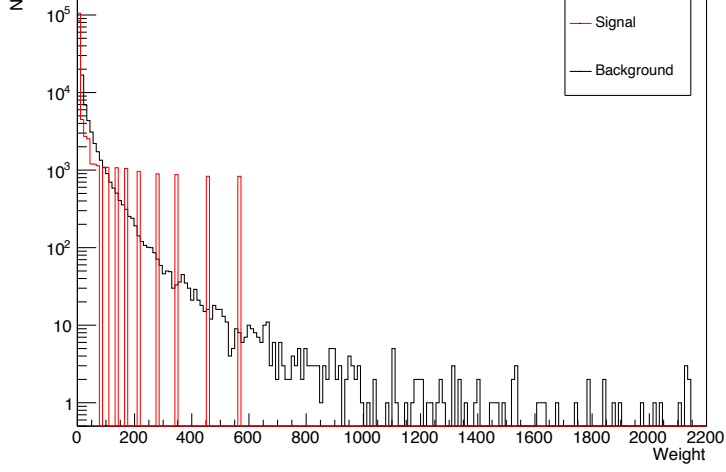


Figure 3.1: Testing weights for signal and background

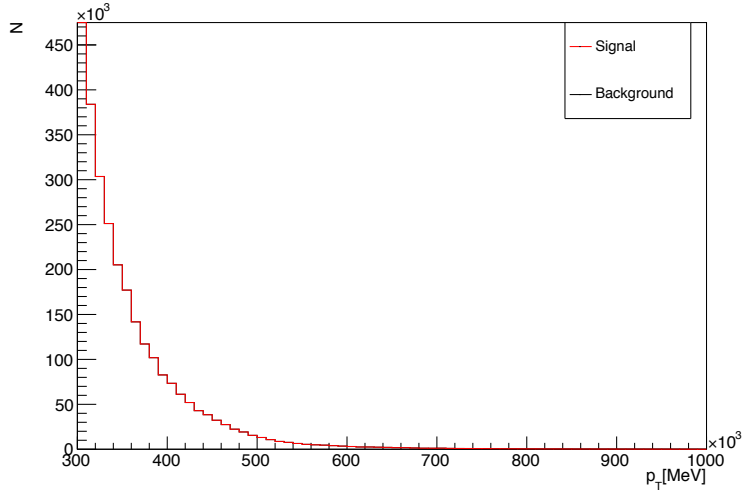


Figure 3.2: p_T distributions of signal and background

Training Top tagging with AGILEPack is based on the substructure variables. The choice of the input variables are crucial for the deep learning algorithm as noise or any artificial properties in the input variables of simulated samples can harm the deep learning algorithm. Therefore, it is very important to use well defined and safe variables which provide good discrimination. We used the variables:

- Trimmed mass
- k_T splittings: $\sqrt{d_{12}}, \sqrt{d_{23}}, \sqrt{d_{34}}$
- N-subjettiness: $\tau_1, \tau_2, \tau_3, \tau_{21}, \tau_{32}$

which provide an input layer with 9 neurons. Next, the architecture of the deep network should be defined by considering the number of inputs, statistics and the complexity of the input variables. The neural network was trained to predict if the jet under consideration was a top quark or a jet from the QCD background, which is represented by a single neuron at the output layer. Once the input layer size and the output layer size are defined by the user, AGILEPack suggests a heuristic architecture if the architecture is not defined by the user. Throughout this project, we used the architecture suggested by AGILEPack which was a neural network with 3 hidden layers. The hidden layers had 10, 6, 2 neurons respectively. The hyper-parameters were set to default values suggested by Adam($\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$), which adjusts the learning rate and momentum by computing individual adaptive learning rates for different parameters⁴. In addition, we trained the neural network with the regularizer=0.000001 in order to avoid overfitting. As it was mentioned before, only supervised learning is used for training in this project as recent studies showed that unsupervised learning is not needed if the neural network is well optimized. We chose the number of supervised epochs to be 20. The mini batch size was chosen after comparing the performance of the neural networks trained with 1, 5, 10 mini batch sizes. The neural network trained with batch size 1 performed slightly better and the mini batch size was fixed to 1 for the results presented here. Neural networks studied in this project are always trained in the p_T range $300 < p_T < 1000$ GeV, and η range $|\eta| < 2$. Training in p_T , η bins were not possible due to low statistics.

Performance evaluation After freezing the architecture, hyper parameters and the training procedure, we studied the performance of the taggers in terms of background rejection($1/\epsilon_b$) and signal efficiency(ϵ_s). Background rejection $1/\epsilon_b$, is defined as the ratio of total number of background jets in the testing sample to number of background jets which pass the signal selection criteria. Signal efficiency is defined as the ratio of number of signal jets which pass the selection criteria to total number of signal jets in the testing sample and the chosen range.

$$\frac{1}{\epsilon_{QCD}} = \frac{N_{total}^{QCD}}{N_{tagged}^{QCD}}, \epsilon_t = \frac{N_{tagged}^{top}}{N_{total}^{top}}$$

The jets with unphysical substructure variables were only included in N_{total}^{QCD} and N_{total}^{top} as they were always rejected. This rejection had a very small effect on our samples as the total weighted number of jets didn't change significantly⁵.

3.1 Training Statistics

Deep neural networks are designed to be trained with at least half a million jets and they achieve their best performance with millions of events. We had limited statistics with the samples that we used(~ 590000 training jets). Hence, we first studied the neural network's performance dependence on the training statistics in order to make sure that the neural network is reaching a stable performance with our training samples. In order to study the dependence on training statistics, while varying the training sample size, we used a common sample⁶ to test the performance of each neural network and compared their performances. The training sample sizes that we used are shown in Table 3.1. We compared the performance of neural networks which were trained with and without training weights⁷. The results are shown in Figure 3.3 and Figure 3.4.

We observed that the neural network which was trained with the least number of jets performs worst, and the neural network which was trained with the greatest number of jets performs best in general, although we see some irregularities in the performance of neural networks.

If the neural network is trained without weights, the neural network's performance is not always

⁴We also tested the performance with several different learning rates and the default value given by Adam indeed performed better.

⁵After this rejection, total weighted number of jets in signal reduces to 2724250 from 2724760, whereas in background reduces to 2715590 from 2724760

⁶With 254400(unweighted) testing jets

⁷The training weights which are based on $(norm_mcevt)^{0.8}$ will be explained in the next section.

Table 3.1: Training samples

	Number of jets
Sample 1	74200
Sample 2	148400
Sample 3	222600
Sample 4	296800
Sample 5	371000
Sample 6	445200
Sample 7	519400
Sample 8	593600

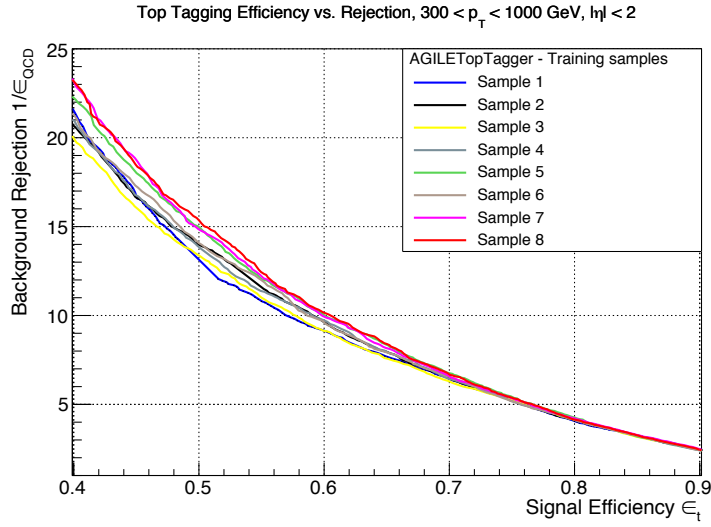


Figure 3.3: Performance comparison of neural networks which were trained with different sample sizes and trained without training weights

improved with increasing the number of training jets until it reaches a certain value. The performance difference between training Sample 7 and Sample 8 appears to reach a more stable performance. However, in order to have a final conclusion we would need to test the performance with at least one other sample which has higher statistics than sample 8.

If the neural network is trained with weights, the neural network's performance is not always improved with increasing the number of training jets, however the difference in the performance decreases when the training samples have equal or greater than 445200 jets.

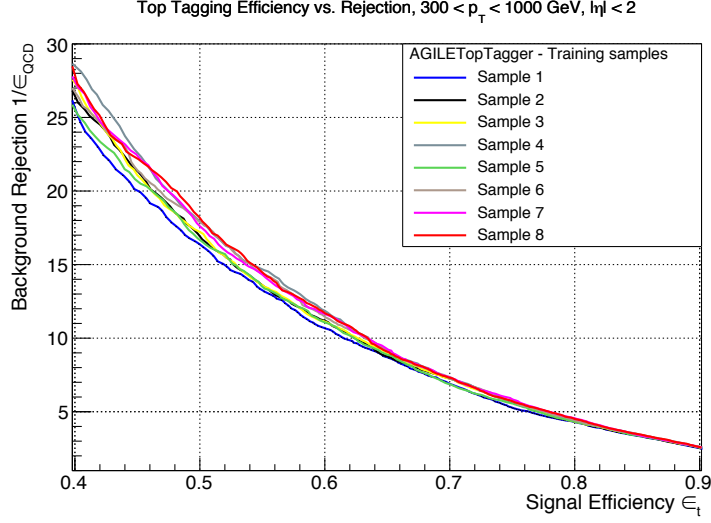


Figure 3.4: Performance comparison of neural networks which were trained with different sample sizes and with training weights based on $(norm_mcevt)^j$ with $j = 0.8$

3.2 Training Weights

As it was mentioned earlier, training weights affect how effectively each jet updates the neural network's weights and biases. Optimal set of training weights would be a good proxy for the testing sample, avoid biases and would not use extremely small or large weights. We studied the effect of the training weights by comparing the performance of the neural networks which were trained with different set of training weights but with the same sample and training procedure. The set of training weights we studied had different prioritized goals amongst the above features of an optimal set. We compared the performances of the neural networks trained with the following weights.

- **No weight** Training a neural network without applying training weights and comparing it with the other neural networks is crucial to make sure that the training of the neural network achieves moderate performance independent of the training weights and to understand the effect of the training weights.
- **2D reweighting in $p_T - \eta$ with a flat p_T distribution** This reweighting is applied with the goal of being independent of the statistical differences across the p_T spectrum. We applied a 2D reweighting in order to not change the η distribution significantly while reweighting and we made sure that the reweighting did not result in very small or large weights. Resulting weight range is $[0.34, 16]$.
- **1D reweighting in p_T : $(norm_mcevt)^j$ with $j = \{0.5, 0.8, 1\}$** We used these sets of weights to find an optimal weight distribution which train the neural network with a distribution that mimics the QCD background by using relative MC event weights while avoiding a very wide weight range. As we wanted to avoid very large weights, we first normalized the weighted number of jets to the total number of unweighted jets, which are referred to as $(norm_mcevt)$. We compared the resulting p_T distributions and performance differences between three smooth functions of $(norm_mcevt)$ in order to find which set of weights performed best. The resulting weight ranges are given below.

1. $(norm_mcevt)$: Resulting weight range = $[0.002, 98]$
2. $(norm_mcevt)^{0.8}$: Resulting weight range = $[0.005, 40]$
3. $(norm_mcevt)^{0.5}$: Resulting weight range = $[0.03, 10]$

As a first comparison, performances of the neural networks in the whole p_T region of our samples in terms of signal efficiency and background rejection are shown in Figure 3.5. We also included two widely used basic substructure variable taggers in our performance plots in order to compare the performance of the neural networks with these taggers. We observed that all AGILETopTaggers significantly perform better than the substructure variable cuts. Additionally, the neural networks trained with $(norm_mcevt)^j$ where $j = \{0.5, 0.8, 1\}$ outperform the neural networks which were trained without weight and with a 2D reweighting in $p_T - \eta$.

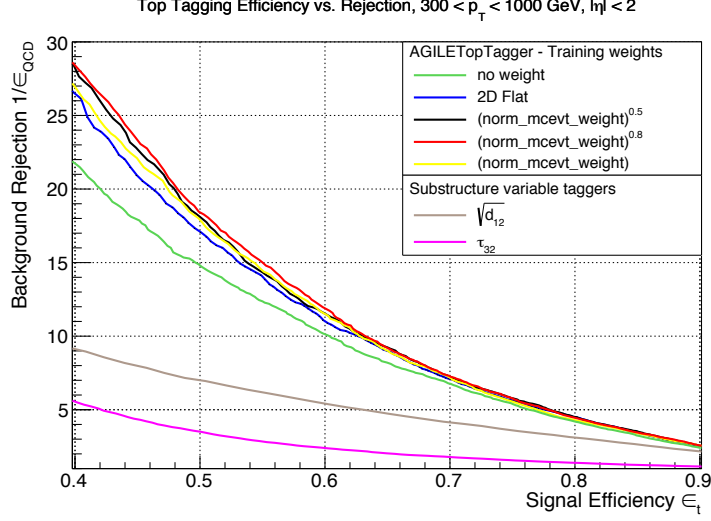


Figure 3.5: Performance comparison of neural networks which were trained with different training weights. The neural network is trained and tested in the same p_T window

As we applied different reweightings which result in significantly different p_T distributions, it was important to evaluate the performance differences in high and low p_T bins in order to make sure that we were not improving the performance only in a certain region while worsening the performance in other regions by applying the training weights.

3.2.1 Performance in p_T Bins

We evaluated and compared the performances of the neural networks⁸ in the low ($300 < p_T < 500$ GeV) and high ($500 < p_T < 1000$ GeV) p_T bins. We observed that all AGILETopTaggers perform better in the high p_T region despite the fact that the training distribution in p_T changes significantly. Amongst the $(norm_mcevt)^j$ training weight based taggers, $(norm_mcevt)^{0.8}$ generally performs slightly better. The performances in low and high p_T bins are presented in Figure 3.6 and Figure 3.7.

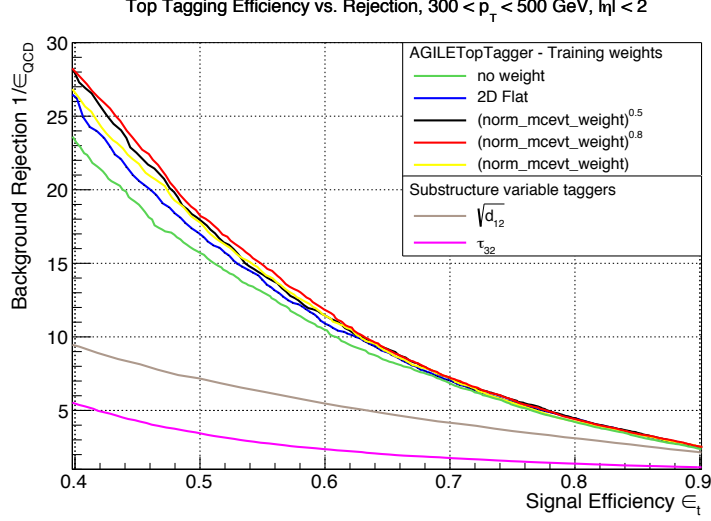


Figure 3.6: Performance comparison of neural networks which were trained with different training weights. The neural network is tested in the low p_T bin

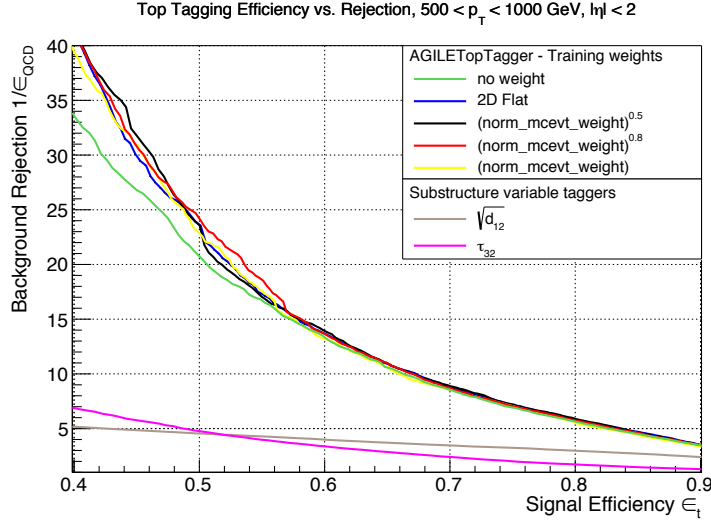


Figure 3.7: Performance comparison of neural networks which were trained with different training weights. The neural network is tested in the high p_T bin

⁸Please keep in mind that the neural networks are always trained in the following p_T and η ranges $300 < p_T < 1000$ GeV, $|\eta| < 2$.

3.2.2 p_T Dependence of Performance at 50% Working Point

We observed significant differences in background rejections which correspond to 50% signal efficiency between substructure variable taggers, neural networks trained with the weights based on $(norm_mcevt)^j$ and neural networks trained with the weights based on the 2D reweighting in $p_T - \eta$ with a flat p_T distribution. In this section we compare the p_T dependence of signal efficiencies and background rejections at 50% working points(WP) of low and high p_T bins⁹. Although there are minor differences in p_T dependence of top taggers, we didn't observe any significant difference that could be harmful.

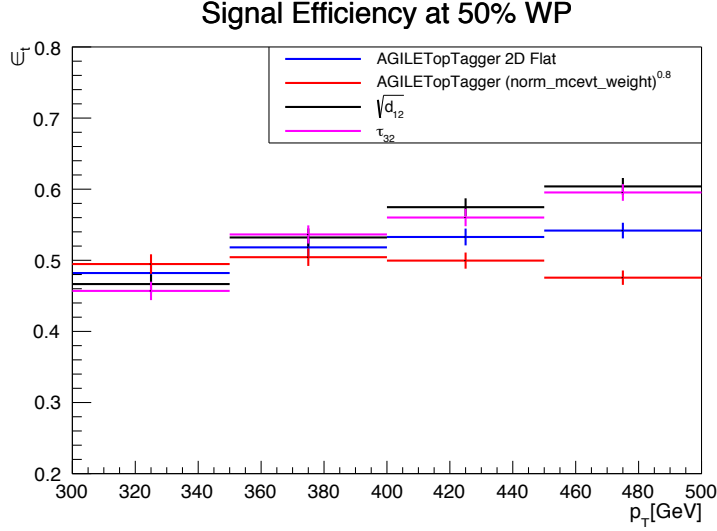


Figure 3.8: Comparison of signal efficiency dependence on p_T in the $300 < p_T < 500$ GeV bin

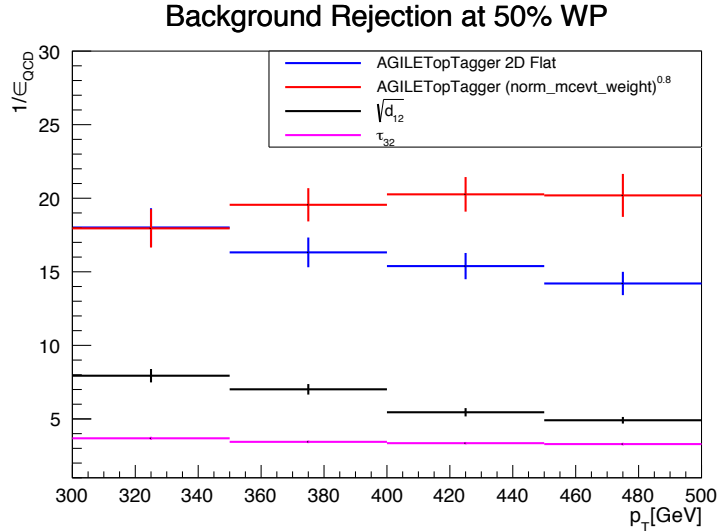


Figure 3.9: Comparison of background rejection dependence on p_T in the $300 < p_T < 500$ GeV bin

⁹50% WP is the operating point corresponding to an overall efficiency of 50% in the considered p_T bin.

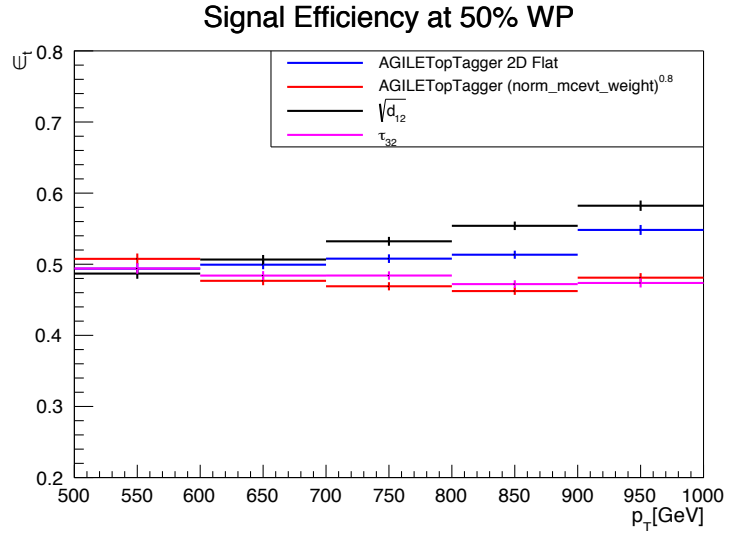


Figure 3.10: Comparison of signal efficiency dependence on p_T in the $500 < p_T < 1000$ GeV bin

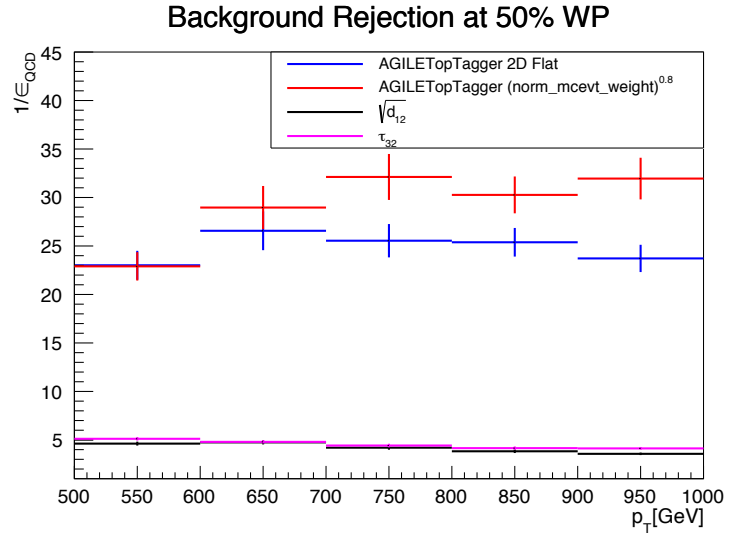


Figure 3.11: Comparison of background rejection dependence on p_T in the $500 < p_T < 1000$ GeV bin

3.2.3 Performance in $|\eta|$ Bins

As a last step, we compared the performance in two $|\eta|$ bins ($|\eta| < 0.7$ and $0.7 < |\eta| < 2$) because in only one training reweighting η is taken into account. Including η in the reweighting in addition to p_T is expected to be a secondary effect. We observed that AGILETopTaggers perform better in high $|\eta|$ bin and AGILETopTagger trained with weights based on $(norm_mcevt)^{0.8}$ is again the most performant tagger in both bins.

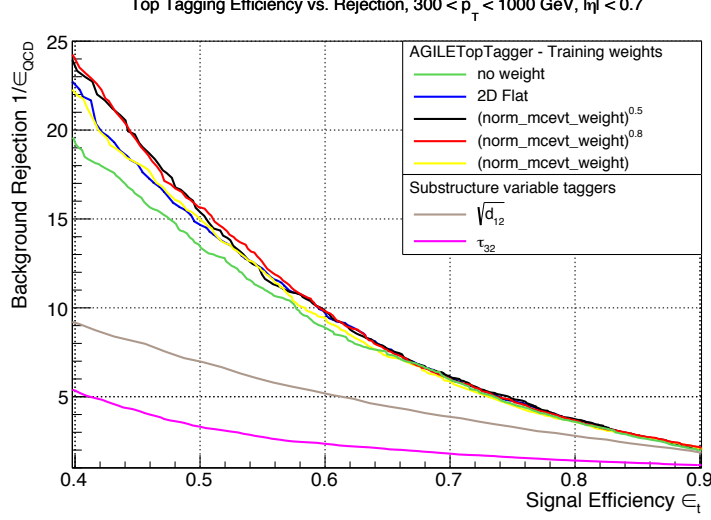


Figure 3.12: Performance comparison of neural networks which were trained with different training weights. The neural network is tested in the low η bin

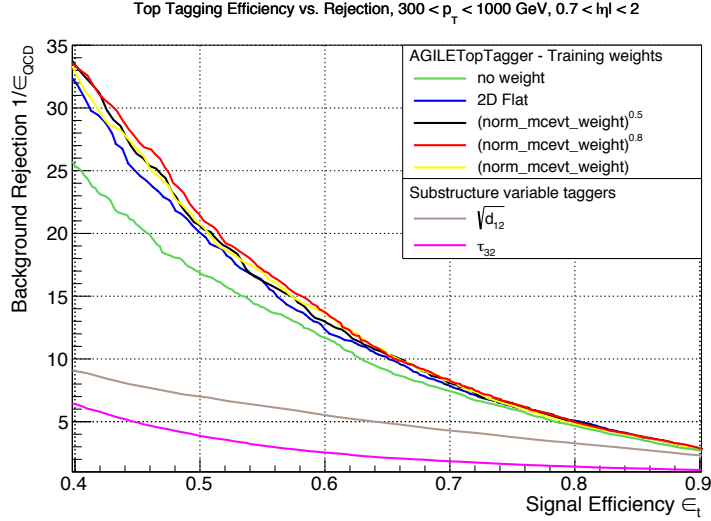


Figure 3.13: Performance comparison of neural networks which were trained with different training weights. The neural network is tested in the high η bin

3.3 Final results

We observed that AGILETopTagger that is the result of the neural network trained with weights based on $(norm_mcevt)^8$ performs best in the full p_T and η spectrum as it provides a good balance between having a reasonable training weight range for the deep learning algorithm and providing a good proxy for the testing sample. Hence, we chose it as the benchmark AGILETopTagger for final comparison with other top tagging algorithms. We compared its performance with the basic N-subjettiness τ_{32} tagger and early Run-2 pre-recommendation top taggers based on τ_{32} , and the calibrated trimmed mass. We evaluated the performance of the Run-2 pre-recommendation top tagger by varying the cut on τ_{32} while applying the recommended cuts on the trimmed mass. Please note that in our samples we don't have the information about the calibrated trimmed mass and the evaluated performance of this tagger is only an indication of its performance in our samples.

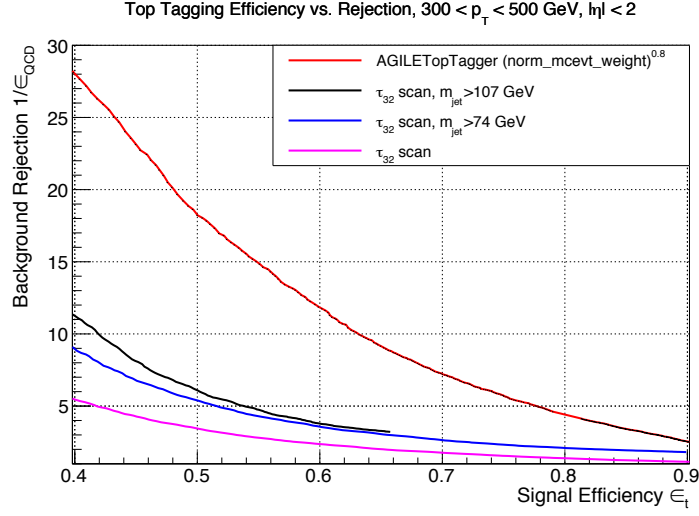


Figure 3.14: Performance comparison of AGILETopTagger with the basic substructure variable taggers in the low p_T bin

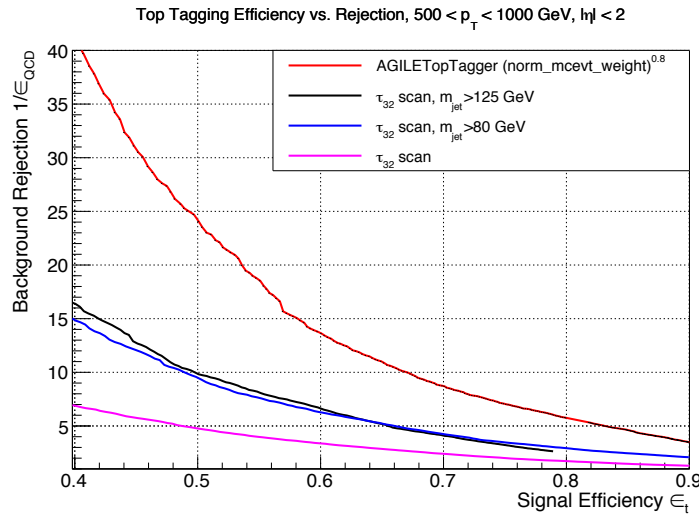


Figure 3.15: Performance comparison of AGILETopTagger with the basic substructure variable taggers in the high p_T bin

Chapter 4

Conclusions

As a first step, we studied the effect of the training sample size. Although the training of the deep network resulted in relatively stable top tagging performance with our training sample, one would need higher statistics in order to study the deep networks in more detail and optimize further the AGILETopTagger.

We focused in the studies on the performances of the neural networks trained with different sets of weights and we observed differences in the performances of these neural networks. All AGILEPack neural networks outperformed the basic substructure taggers. In addition, all AGILEPack neural networks performed better in high p_T ($300 < p_T < 1000$ GeV) and high $|\eta|$ ($0.7 < |\eta| < 2$) regions independent of the training weights. The AGILETopTagger based on $(norm_mcevt)^{0.8}$ training weights showed the overall best performance in addition to not having a strong p_T dependence, which clarifies that the training weights do not create a bias in certain regions of the p_T spectrum.

Outlook and plans for future Studies carried out in this project were a small part of a much wider effort on top tagging with advanced machine learning techniques. Adopting advanced machine learning techniques in object identification require extensive studies to adapt these techniques for high energy physics specific problems.

In order to make AGILETopTagger one of the top tagging algorithms in the ATLAS experiment, first of all AGILETopTagger should be trained and tested with MC15 simulated samples which provide higher statistics and a better description of the high-luminosity, high energy environment of the LHC. In particular, it is crucial to study the top tagging performance as a function of pile-up with these samples. Moreover, in this study AGILETopTagger's performance was only compared with the basic substructure variable cut based taggers, and the fair comparison of its performance with other MVA methods' are necessary to study the benefits or the hazards of employing a deep learning algorithm in top tagging.

In addition to necessary tests with new simulated samples, the performance on data should be evaluated and the performance agreement between data and MC simulation should be studied in order to use AGILETopTagger in physics analyses.

Part I - Sample list

[illegible]

mc12.8TeV.182420.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM50_W3_QCUT300.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182417.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM100_W3_QCUT80.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182421.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM100_W3_QCUT300.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182418.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM300_W3_QCUT80.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182422.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM300_W3_QCUT300.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182419.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM600_W3_QCUT80.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182423.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM600_W3_QCUT300.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182408.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM50_MM50_W8pi_QCUT80.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182412.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM50_MM50_W8pi_QCUT300.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182409.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM50_MM100_W8pi_QCUT80.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182413.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM50_MM100_W8pi_QCUT300.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182410.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM50_MM300_W8pi_QCUT80.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182414.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM50_MM300_W8pi_QCUT300.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182411.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM50_MM600_W8pi_QCUT80.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182415.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM50_MM600_W8pi_QCUT300.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182424.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM50_W8pi_QCUT80.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182428.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM50_W8pi_QCUT300.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182425.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM100_W8pi_QCUT80.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182429.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM100_W8pi_QCUT300.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182426.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM300_W8pi_QCUT80.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182430.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM300_W8pi_QCUT300.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182427.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM600_W8pi_QCUT80.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328
mc12.8TeV.182431.MadgraphPythia_AUET2BCTEQ6L1_dmA_DM400_MM600_W8pi_QCUT300.merge.NTUP_SUSY.e2219_a188_a205_r4540_p1328

Appendix B

Part I - Trigger turn-on curves and parameterization

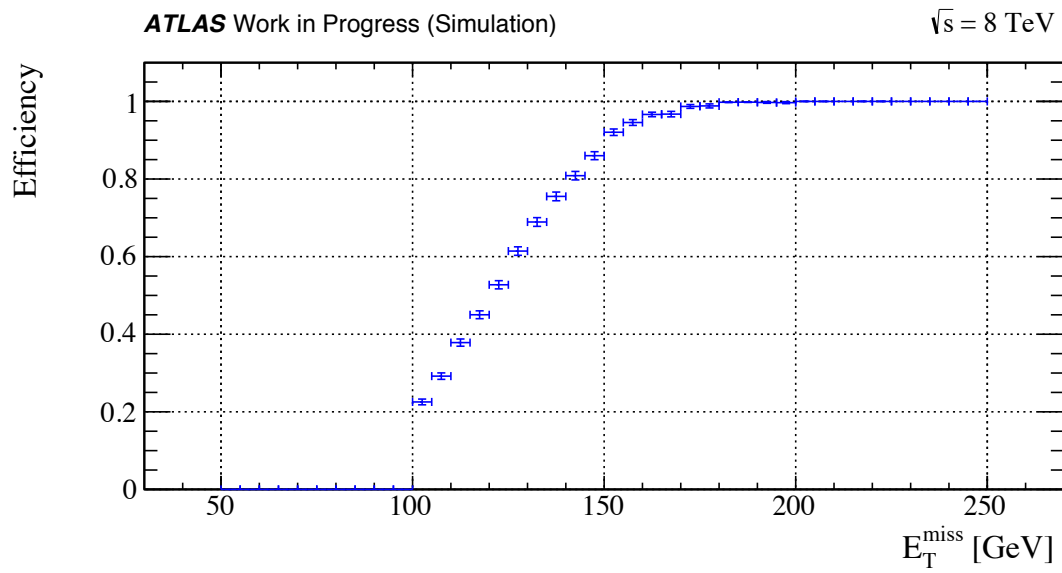


Figure B.1: Efficiency of *xe100_XE70* where there are less than 2 jets with $p_T \geq 30 \text{ GeV}$

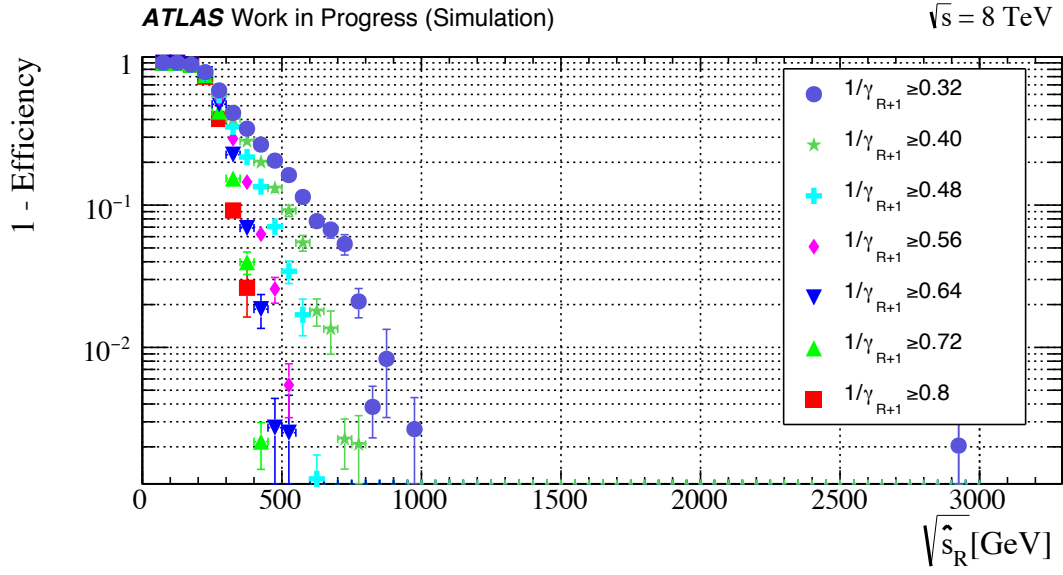


Figure B.2: Inefficiency of $(ProdR170_2J15_XE55||xe100_XE70)$ in \hat{s}_R bins where the events pass the imposed thresholds on $\frac{1}{\gamma_{R+1}}$

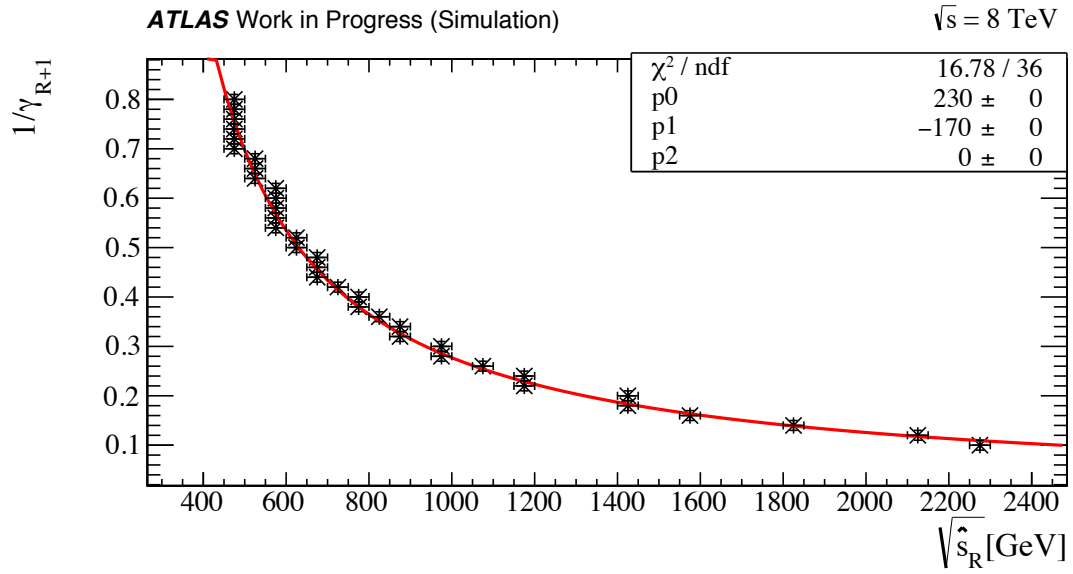


Figure B.3: Parametrizing the turn-on curve: Fit the retrieved points from the previous figure to a hyperbolic function. Shown parameters are imposed after rounding the original resulting fit parameters

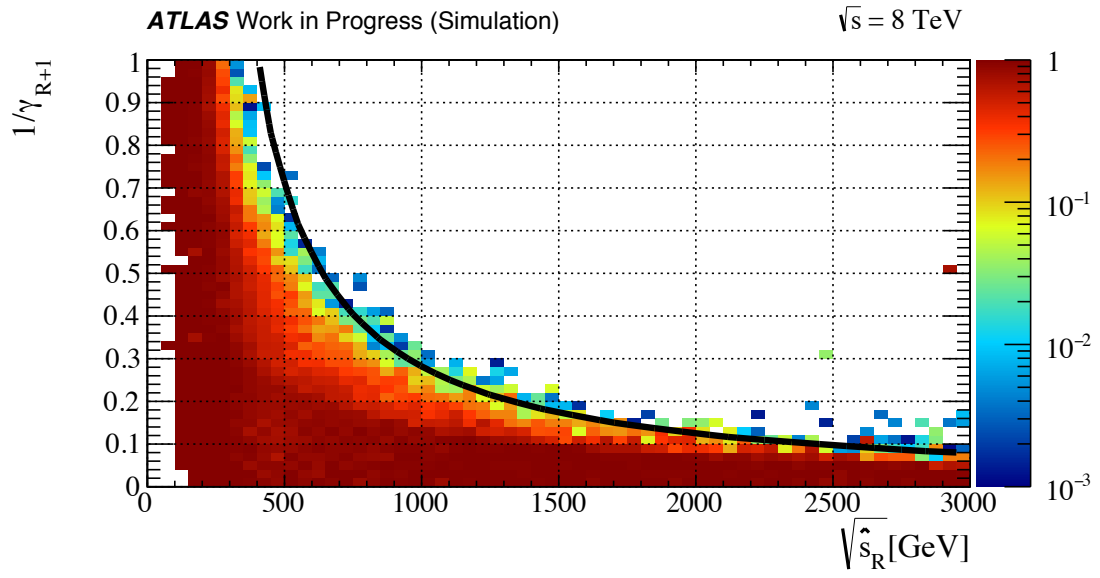


Figure B.4: Inefficiency of $(ProdR170_2J15_XE55||xe100_XE70)$ and the hyperbolic offline cut at $230 = (\sqrt{\hat{s}_R} - 170) + (\frac{1}{\gamma_{R+1}})$

Bibliography

- [1] Mark Thomson. *Modern particle physics*. Cambridge University Press, Cambridge, Sep 2013.
- [2] B R Martin and G Shaw. *Particle Physics; 3rd ed.* The Manchester Physics Series. Wiley, New York, NY, 2008.
- [3] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.
- [4] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.
- [5] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.
- [6] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.
- [7] Oliver S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, et al. LHC Design Report. 1. The LHC Main Ring. 2004.
- [8] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.
- [9] Georges Aad et al. Performance of the ATLAS Trigger System in 2010. *Eur.Phys.J.*, C72:1849, 2012.
- [10] L. Zhao and PH Beachemin. Commissioning of the ATLAS Jet and Missing Energy Triggers with Beam collisions at the LHC. 2010.
- [11] Performance of the ATLAS Jet Trigger in the Early $\sqrt{s}=7$ TeV Data. 2010.
- [12] L. Lopes. Performance of the ATLAS jet trigger. 2012.
- [13] D. Casadei, I. Aracena, S. Banerjee, P-H Beauchemin, S. Calvet, et al. The implementation of the ATLAS missing Et triggers for the initial LHC operation. 2011.
- [14] Christopher Rogan. Kinematical variables towards new dynamics at the LHC. 2010.
- [15] Aad, Georges and others. Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Eur. Phys. J.*, C75(7):299, 2015.
- [16] Patrick J. Fox, Roni Harnik, Joachim Kopp, and Yuhsin Tsai. Missing Energy Signatures of Dark Matter at the LHC. *Phys.Rev.*, D85:056011, 2012.
- [17] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helv.Phys.Acta*, 6:110–127, 1933.
- [18] M. Milgrom. A Modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. *Astrophys.J.*, 270:365–370, 1983.

- [19] Oscar Adriani et al. An anomalous positron abundance in cosmic rays with energies 1.5-100 GeV. *Nature*, 458:607–609, 2009.
- [20] M. Ackermann et al. Measurement of separate cosmic-ray electron and positron spectra with the Fermi Large Area Telescope. *Phys.Rev.Lett.*, 108:011103, 2012.
- [21] M. Aguilar and Alberti et al. First result from the alpha magnetic spectrometer on the international space station: Precision measurement of the positron fraction in primary cosmic rays of 0.5–350 gev. *Phys. Rev. Lett.*, 110:141102, Apr 2013.
- [22] Lars Bergström, Torsten Bringmann, Ilias Cholis, Dan Hooper, and Christoph Weniger. New limits on dark matter annihilation from alpha magnetic spectrometer cosmic ray positron data. *Phys. Rev. Lett.*, 111:171101, Oct 2013.
- [23] L. Accardo et al. High Statistics Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–500 GeV with the Alpha Magnetic Spectrometer on the International Space Station. *Phys.Rev.Lett.*, 113:121101, 2014.
- [24] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: Evidence, candidates and constraints. *Phys.Rept.*, 405:279–390, 2005.
- [25] V. C. Rubin and W. K. Ford, Jr. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *Astrophys.J.*, 159:379, February 1970.
- [26] V. C. Rubin, W. K. J. Ford, and N. . Thonnard. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/. *Astrophys.J.*, 238:471–487, June 1980.
- [27] T. S. van Albada, J. N. Bahcall, K. Begeman, and R. Sancisi. Distribution of dark matter in the spiral galaxy NGC 3198. *Astrophys.J.*, 295:305–313, August 1985.
- [28] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *Astrophys.J.*, 142:419–421, July 1965.
- [29] P.A.R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. 2015.
- [30] T. A. Porter, R. P. Johnson, and P. W. Graham. Dark Matter Searches with Astroparticle Data. *Annual Review of Astronomy and Astrophysics*, 49:155–194, September 2011.
- [31] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.
- [32] Daniel Bauer et al. Dark Matter in the Coming Decade: Complementary Paths to Discovery and Beyond. *Phys.Dark Univ.*, 7-8:16–23, 2013.
- [33] Aad, Georges and others. Search for new phenomena in events with a photon and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev.*, D91(1):012008, 2015.
- [34] Aad, Georges and others. Search for dark matter in events with a Z boson and missing transverse momentum in pp collisions at $\sqrt{s}=8$ TeV with the ATLAS detector. *Phys.Rev.*, D90(1):012004, 2014.
- [35] Yang Bai and Tim M. P. Tait. Searches with Mono-Leptons. *Phys. Lett.*, B723:384–387, 2013.
- [36] Arvind Rajaraman, William Shepherd, Tim M. P. Tait, and Alexander M. Wijangco. LHC Bounds on Interactions of Dark Matter. *Phys. Rev.*, D84:095013, 2011.
- [37] Maria Beltran, Dan Hooper, Edward W. Kolb, Zosia A. C. Krusberg, and Tim M. P. Tait. Maverick dark matter at colliders. *JHEP*, 09:037, 2010.

- [38] Giorgio Busoni, Andrea De Simone, Enrico Morgante, and Antonio Riotto. On the Validity of the Effective Field Theory for Dark Matter Searches at the LHC. *Phys.Lett.*, B728:412–421, 2014.
- [39] Giorgio Busoni, Andrea De Simone, Johanna Gramling, Enrico Morgante, and Antonio Riotto. On the Validity of the Effective Field Theory for Dark Matter Searches at the LHC, Part II: Complete Analysis for the s -channel. *JCAP*, 1406:060, 2014.
- [40] Giorgio Busoni, Andrea De Simone, Thomas Jacques, Enrico Morgante, and Antonio Riotto. On the Validity of the Effective Field Theory for Dark Matter Searches at the LHC Part III: Analysis for the t -channel. *JCAP*, 1409:022, 2014.
- [41] Ruth Pöttgen. Exotics Searches at ATLAS. 2014.
- [42] Matthew R. Buckley, Joseph D. Lykken, Christopher Rogan, and Maria Spiropulu. Super-Razor and Searches for Sleptons and Charginos at the LHC. *Phys.Rev.*, D89:055020, 2014.
- [43] Serguei Chatrchyan et al. Inclusive search for supersymmetry using the razor variables in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. Lett.*, 111(8):081802, 2013.
- [44] Partha Konar, Kyoungchul Kong, and Konstantin T. Matchev. $\sqrt{\hat{s}_{min}}$: A Global inclusive variable for determining the mass scale of new physics in events with missing energy at hadron colliders. *JHEP*, 0903:085, 2009.
- [45] Jay Hubisz, Joseph Lykken, Maurizio Pierini, and Maria Spiropulu. Missing energy look-alikes with 100 pb^{-1} at the LHC. *Phys.Rev.*, D78:075008, 2008.
- [46] Serguei Chatrchyan et al. Inclusive search for squarks and gluinos in pp collisions at $\sqrt{s} = 7$ TeV. *Phys.Rev.*, D85:012004, 2012.
- [47] Patrick J. Fox, Roni Harnik, Reinard Primulando, and Chiu-Tien Yu. Taking a Razor to Dark Matter Parameter Space at the LHC. *Phys.Rev.*, D86:015010, 2012.
- [48] Sensitivity to WIMP Dark Matter in the Final States Containing Jets and Missing Transverse Momentum with the ATLAS Detector at 14 TeV LHC. Technical Report ATL-PHYS-PUB-2014-007, CERN, Geneva, Jun 2014.
- [49] Harvard University. Harvardtriggercode, <https://github.com/etolley/harvardtriggercode>.
- [50] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.
- [51] F. Abe et al. Observation of top quark production in $\bar{p}p$ collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995.
- [52] M. Beneke et al. Top quark physics. In *1999 CERN Workshop on standard model physics (and more) at the LHC, CERN, Geneva, Switzerland, 25-26 May: Proceedings*, 2000.
- [53] V. M. Abazov et al. Observation of Single Top Quark Production. *Phys. Rev. Lett.*, 103:092001, 2009.
- [54] Richard Hawkings. Top quark physics at the LHC. *Comptes Rendus Physique*, 16:424–434, 2015.
- [55] Sven Menke. Measurements of the top quark mass using the ATLAS and CMS detectors at the LHC. In *23rd International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2015) Dallas, Texas, United States, April 27-May 1, 2015*, 2015.
- [56] Gilad Perez. Top quark theory and the new physics searches frontier. *Phys. Scripta*, T158:014008, 2013.
- [57] Stijn Blyweert. Top-quark mass measurements at the LHC. In *Proceedings, 47th Rencontres de Moriond on QCD and High Energy Interactions*, 2012.

- [58] E. Boos, O. Brandt, D. Denisov, S. Denisov, and P. Grannis. The Top Quark. 2015.
- [59] Jean-François Arguin. Experimental interplay between the top quark and Supersymmetry at the LHC and Tevatron. *J. Phys. Conf. Ser.*, 452(1):012005, 2013.
- [60] Georges Aad et al. A search for $t\bar{t}$ resonances using lepton-plus-jets events in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 08:148, 2015.
- [61] Georges Aad et al. Search for production of vector-like quark pairs and of four top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 08:105, 2015.
- [62] Georges Aad et al. Search for invisible particles produced in association with single-top-quarks in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Eur. Phys. J.*, C75(2):79, 2015.
- [63] Tilman Plehn. *Lectures on LHC physics; 2nd ed.* Lecture Notes in Physics. Springer, Cham, 2015.
- [64] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better jet clustering algorithms. *JHEP*, 08:001, 1997.
- [65] M. Wobisch and T. Wengler. Hadronization corrections to jet cross-sections in deep inelastic scattering. In *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, 1998.
- [66] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti- $k(t)$ jet clustering algorithm. *JHEP*, 04:063, 2008.
- [67] Vardan Khachatryan et al. Measurement of the ratio $B(t \rightarrow Wb)/B(t \rightarrow Wq)$ in pp collisions at $\sqrt{s} = 8$ TeV. *Phys. Lett.*, B736:33–57, 2014.
- [68] Georges Aad et al. Comprehensive measurements of t -channel single top-quark production cross sections at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys. Rev.*, D90(11):112006, 2014.
- [69] A. Altheimer et al. Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23rd-27th of July 2012. *Eur. Phys. J.*, C74(3):2792, 2014.
- [70] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet substructure as a new Higgs search channel at the LHC. *AIP Conf. Proc.*, 1078:189–191, 2009.
- [71] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jet Trimming. *JHEP*, 02:084, 2010.
- [72] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches. *Phys. Rev.*, D81:094023, 2010.
- [73] Georges Aad et al. Performance of jet substructure techniques for large- R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *JHEP*, 09:076, 2013.
- [74] Identification of high transverse momentum top quarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. Technical Report ATLAS-CONF-2015-036, CERN, Geneva, Aug 2015.
- [75] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.
- [76] Krzysztof M. Graczyk and Cezary Juszczak. Applications of Neural Networks in Hadron Physics. *J. Phys.*, G42(3):034019, 2015.
- [77] Leandro G. Almeida, Mihailo Backović, Mathieu Cliche, Seung J. Lee, and Maxim Perelstein. Playing Tag with ANN: Boosted Top Identification with Pattern Recognition. *JHEP*, 07:086, 2015.

- [78] Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb^{-1} of Data from the ATLAS Detector. 2012.
- [79] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. Searching for Exotic Particles in High-Energy Physics with Deep Learning. *Nature Commun.*, 5:4308, 2014.
- [80] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015. Published online 2014; based on TR arXiv:1404.7828 [cs.NE].
- [81] Yoshua Bengio, Aaron C. Courville, and Pascal Vincent. Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR*, abs/1206.5538, 2012.
- [82] Luke de Oliveira. Algorithms for generalized inference, learning, and extraction package, <https://github.com/lukedeo/agilepack>.
- [83] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.