Università di Genova

Scuola di Scienze Matematiche, Fisiche e Naturali

Dipartimento di Fisica

Master Degree Thesis in Physics

Trackless flavour-tagging at high jet transverse momentum and its impact on the physics reach of the ATLAS experiment at LHC

CERN-THESIS-2020-315
29/10/2020

Supervisors:
Dott. Andrea Coccaro
Prof. Carlo Schiavi

Co-Supervisor:
Prof. Flavio Fontanelli

Candidate:
Martino Tanasini

Academic Year 2019-2020

Contents

Acknowledgement

This thesis has been with me during all this year. The idea, that was born at CERN during my stay as a Summer Student, grew up in the intimacy of many, many houses; lots located in different countries. From the smallest room in Paris, to the vast yards of an Oxford college, to the blossoming spring of the countryside of Piemonte, it has been a demanding companion of many adventures and tribulations. It gave me an occupation during the obscure times of the lockdown, provoked by the pandemics that cast its shadow over the whole planet. To go through all of this with my only efforts would have been impossible. Luckily, I was not alone. This dedication goes to those that were there during all the story: to my supervisors in particular, but to all the others as well. Thank you very much.

Introduction

"What is the world made of?"

The Standard Model of particle physics is humans' most ambitious attempt to answer this question. A wonderfully complex attempt, with an awe-inspiring power of predicting the results of all the experiments performed to test it in the last century; but still - an attempt. Many pieces of the puzzle are still missing, and in the probably long long path that lays ahead of thousands of particle physicists to come, the next steps to be made are all concerned with the quest of these missing pieces. Do neutrinos have mass? Why is it so small? Why are the galaxies shaped as we observe them to be? Where has all the antimatter gone? Who carries gravity? Is it a particle? Is it possible to predict theoretically the values of the free parameters of the Standard Model? We do not know yet.

But in many collaborations around the world, and in particular at the Large Hadron Collider, the most powerful particle accelerator ever built, physicists are doing their best to answer these questions. The present thesis, in a nutshell, deals with trying to improve the instruments used at the ATLAS experiment to identify one particle that behaves very particularly, the b-quark, when it is produced at high momentum.

Quarks are particles that interact by means of all the interactions we know of. They have a *mass*, hence they feel gravity. They carry an *electric* charge, and thus they are subject to the familiar electromagnetic interactions. They have a *flavour*: hence, they interact *weakly* as well. Finally, they are *coloured*, which means they carry a non null colour charge, the charge of Quantum Chromodynamics. Thus, they are subject to the *strong* interaction.

Being coloured, quarks are always observed as *jets* in particle colliders. Jets are sprays of collimated particles that are generated by the *hadronisation* of the coloured parent particle. Due to a law of nature named *colour confinement*, no colour charge can be observed in a free state. When one of such charges is produced in a collision at a particle collider, it undergoes a cascade particle production process that leads to a colourless final state[1]. The process of identifying the flavour (i.e. the identity) of the quark from which a hadronic jet originated is called *flavour tagging*.

Efficient flavour tagging algorithms are crucial tools for high-energy physics experiments at the Large Hadron Collider. Particularly important is the development of tools discriminating jets containing b- and c-hadrons[2] (b-jets and c-jets) from the others (*light*-flavour jets[3]). The flavour of jets in ATLAS is identified using deep learning techniques[4] combining reconstructed quantities related to distinctive features of the hadrons. These signatures are connected to the relatively long lifetime of b- and c- hadrons and include the presence of the hadrons' decay vertex, separated from the primary collision, and a corresponding high impact parameter of *tracks*. Tracks are trajectories of charged particles as reconstructed by the tracking detectors (the innermost components of the ATLAS detector, immersed in a magnetic field). The impact parameter is the distance of the point of closest approach of a track to the primary collision. Tracks are currently the basic ingredients of the ATLAS flavour

[1] Actually, the t-quark, a fundamental particle that carries colour charge, decays so fast into other quarks that it does not have the time to trigger the hadronisation process. Thus, it is always observed through the jets generated by its decay products.

[2] Hadrons are composite particles made of more than one quark. b- and c-hadrons are thus respectively particles that contain at least a b or c quark.

[3] The suffix *light* collectively refers to the u,d and s quarks

[4] Deep learning is a branch of machine learning that uses multi-layered Neutral Networks.

tagging algorithms, since they are used both directly as inputs for the flavour identification task and for reconstructing the possible presence of secondary vertexes in the jets, that are then used in the vertex-based algorithms.

Unfortunately, problems arise when studying jets originated by very *boosted* (i.e. impulsive) quarks. All the track-based flavour tagging algorithms decrease their performance when exposed to such cases indeed. One of the reasons behind this behaviour is that, with a lot of available energy, the number of particles of the hadronisation process increases swiftly. Furthermore, the jet becomes very collimated. The combination of these two factors gives rise to a densely populated environment inside the jet, negatively affecting the overall performance of the tracking algorithms. Furthermore, because of their boost and of relativistic time dilation, impulsive quarks fly a long distance before decaying. Thus, they can traverse one or more *layers* of the tracking detector. As it will be explained during the development of this thesis, this fact concurs to the *deterioration of the available track information.*

Since many ATLAS analyses extensively rely on an efficient identification of the b-jets, improving the performance of the b-tagging algorithms (i.e. flavour tagging algorithms for the identification of b-jets) would be very beneficial for the future scientific program of the ATLAS collaboration. Examples are the study of the $H \to b\bar{b}$ decay in the VH production mode ([39]) and the search for new particles decaying in pairs of jets ([50]).

A possible approach would be to refine the existing tracking and track-based flavour tagging algorithms, and indeed many improvements have been made on the subject in the last years. The aim of this thesis is to study a different way of approaching the problem that does not rely on the track information: making a direct use of the tracking hit *cluster* information. Clusters are group of sensors located in the innermost region of the ATLAS detector that activate when traversed by electrically charged particles. They are the inputs of the tracking algorithms, that group together those produced by the same particles traversing different layers of the Inner Detector to reconstruct their trajectories. Thus, in a sense, clusters are already used for the flavour tagging task, since they are the basic elements of tracking. But in the cases in which the track information deteriorates, designing the flavour tagging algorithms to make a *direct* use of the cluster information could help enhancing their performance.

Preliminary studies on this cluster-based approach showed how the information on the number of clusters per layer collected in each jet owns a discriminating power which is maintained at impulses up $\mathcal{O}(1)$ TeV when comparing b-jets with the others ([69],[70]). This fact can be understood by considering that the relatively long lifetime of a boosted b-hadron can cause it to decay in a region of the Inner Detector more external than one or more of the pixel layers. In such cases, a *multiplicity jump* is expected in the number of clusters per layer between those located more internally than the decay point and those more external; since those hit before the decay will collect only a cluster from the b-hadron, while those hit after the decay will ideally collect all the clusters left by the multiple decay products. Since displaced decays of this type are typical of b-quarks (*light*-quarks decaying always *inside* the innermost layer of the Inner Detector) the presence of a multiplicity jump can help distinguishing the associated jets.

In this thesis, we will study the performance of an *ad-hoc* designed machine learning algorithm in discriminating b-jets from *light*-flavour jets. The study will be performed using simulated Monte Carlo samples that loyally mimic the ATLAS detector geometry and the physical processes that happen inside it when the collisions occur. We will characterise the extra performance achievable by introducing the cluster information by providing it with various sets of discriminating variables (i.e. features of the jets that should differ in b- and *light*- jets), both track and cluster-based, to see whether the expected benefits at high impulse extend to the realistic case of the examined Monte Carlo simulation[5].

If benefits are observed indeed, our results will be like another drop in the filling jar of the *trackless studies* for flavour tagging: we wish it shall soon be poured in the vast sea of the state-of-the-art algorithms that are used in ATLAS in the investigation of the mysteries of Nature.

The structure of this thesis is the following. In Chapters 1, 2 and 3, I will provide the general background needed to understand the flavour tagging process. I will present the Large Hadron Col-

[5]The previous studies on the use of the cluster information in the flavour tagging task relying on machine learning were made on more simplified models of the detector.

lider and the ATLAS detector (Chapter 1), I will give a general overview of the Standard Model of particle physiscs (Chapter 2), and I will present jets both theoretically as shower of particles belonging to the same hadronisation process and experimentally through the explanation of the anti-k_T jet reconstruction algorithm. Then, in Chapter 4, I will take a closer look to the b-quark, presenting its features and discovery, and I will introduce the state-of-the-art b-tagging algorithms used in ATLAS. To do so, I will also need to define the key ingredients used by such algorithms: hits, clusters, space points, tracks, and vertexes. In Chapter 5 I will present the Monte Carlo sample used for our reaserch, I will elaborate on the deterioration of the track information in highly boosted jets, I will explain the adopted machine learning tools, and I will present their design and application in the flavour tagging task, characterising the benefits gained by the direct use of the cluster information. Chapter 6 is dedicated to the presentation of two ATLAS analyses that extensively relied on good b-tagging performance, namely the study of the $H \to b\bar{b}$ decay in the VH production mode ([38]) and the search for new particles decaying in pairs of jets ([50]). Then, I will elaborate on how much an increase of the b-tagging performance of the scale of those observed in the previous Chapters would lower the exclusion upper limits on some features of one of the signal searched for in [50], if such increases were gained on the ATLAS state-of-the-art b-tagging algorithm named DL1r. Finally, in Chapter 7, I will talk about the further possible studies.

Chapter 1

The Large Hadron Collider

The Large Hadron Collider (LHC) is the largest and most powerful particle accelerator in the world [26]. It was switched on for the first time on the 10th of September 2008, and it is still the latest addition to CERN's accelerator complex. It has undergone two major shutdown periods to upgrade its utilities during its lifetime: the Long Shutdown 1, that took place between 2013 and 2015; and the Long Shutdown 2, that was meant to cover the 2018 - 2021 period but has been prolonged due to the Covid19 pandemics. Furthermore, on the 19 September 2008, an incident consisting in a large helium leak caused an interruption of the activities of the accelerator that lasted several months [19]. The LHC consists of a 26.7-kilometre ring of superconducting magnets with some accelerating structures to increase the energy of the particles along the way. The LHC was constructed by the European Organization for Nuclear Research (CERN) in the same tunnel that housed its Large Electron-Positron Collider (LEP). The tunnel is circular and is located 50–175 metres underground, on the border between France and Switzerland near the city of Geneva (Fig 1.1)

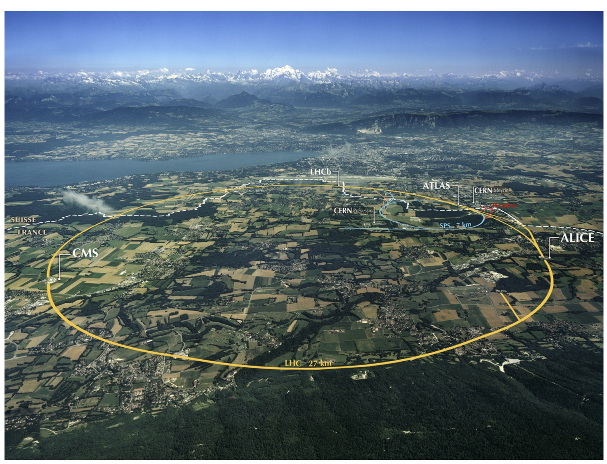


Figure 1.1: LHC: a top view

Inside the accelerator, two high-energy particle beams are made to collide at four locations corresponding to the positions of four particle detectors – ATLAS (A Toroidal LHC ApparatuS), CMS (Compact Muon Solenoid), ALICE (A Large Ion Collider Experiment) and LHCb (LHC-beauty) (Fig 1.2). The beams travel in opposite directions in separate beam pipes – two tubes kept at ultrahigh vacuum. They are guided around the accelerator ring by a strong magnetic field maintained by superconducting electromagnets. The electromagnets are built from coils of special electric cable that operates in a superconducting state. This requires chilling the magnets to -271.3°– a temperature colder than outer space. For this reason, much of the accelerator is connected to a distribution system of liquid helium, which cools the magnets, as well as to other supply services.

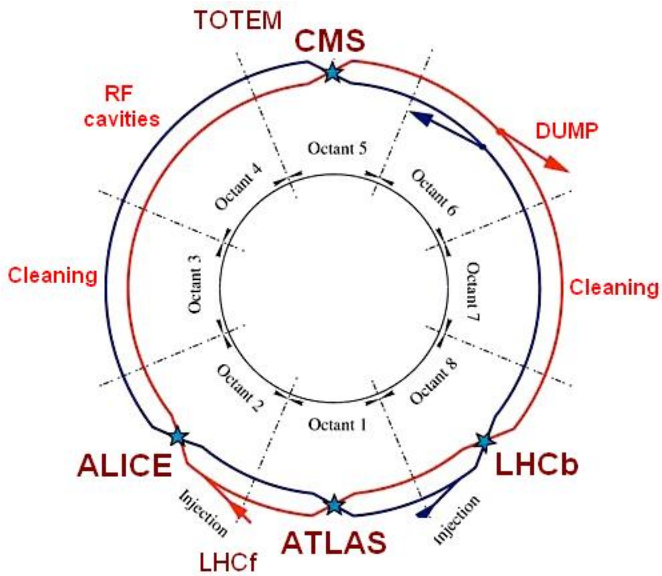


Figure 1.2: Layout of the LHC. The two beams are shown in red and blue. The positions of the four particle detectors are shown, as well as the location of the point of injection and dumping of the beams.

The most important parameters for the LHC are the energy reached in the center of mass reference frame and the collision event rate [14]. The LHC accelerates proton beams up to energies of 6.5 TeV to be collided in ATLAS, CMS and LHCb; and heavy ion beams up to energies of 2.76 TeV to be collided in ALICE. Considering that the energy of the proton is distributed among the quarks and gluons that constitute it, a proton beam energy of 6.5 TeV means that the average centre-of-mass deep-scattering collision energies are greater than 1 TeV. To be able to observe an high number of interesting collision products with the detector, an high collision rate is required. The number of collision events that the LHC can provide is given by the product of the cross section of the event and the machine luminosity L, which is determined by the proton beams parameters:

$$L = \frac{f_{\mathrm{rev}} n_b N^2}{\sigma_x \sigma_y} F(\Phi, \sigma_{x,y}, \sigma_s) \tag{1.1}$$

where f_{rev} is the revolution frequency; n_b the number of bunches; N the number of particles within each bunch; F a geometric factor that depends on the crossing angle of the two beans (Φ), the transverse root mean squared (r.m.s) beam size ($\sigma_{x,y}$) and the r.m.s bunch length (σ_s). At the LHC, the nominal value of L is set to $10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, thus requiring 2808 bunches each containing 1.15×10^{11} protons, a $\sigma_{x,y}$ of $16\,\mu$m , a σ_s of 7.5 cm and an angle of $320\,\mu$rad at the interaction points.

The protons that make up the beams go through many accelerators before being injected in the LHC tunnel (Fig 1.3). Until the current 2019-2020 shutdown, the protons were produced into the linear accelerator LINAC2 [21] by ionization of an hydrogen gas through an electric field, and then accelerated up to 50 MeV before being injected in the Proton Synchroton Booster (PBS) [24]. When the shutdown is over, LINAC2 will be replaced by the linear accelerator LINAC4 [22], which will produce H^- ions and accelerate them up to 160 MeV. The ions will then be stripped of their two electrons during injection from LINAC4 into the PBS to leave only protons. The PBS, made up of four superimposed synchrotron rings, further accelerates the protons up to 1.4 GeV. The particles are then injected in the Proton Synchrotron (PS) [23], a 628 m diameter accelerator which once was CERN's flagship machine and now is used in the accelerator chain to increase the energy of the protons up to

25 GeV. In the PS, the low energy bunches are splitted to generate the bunch pattern and spacing needed for the LHC. The bunches are then injected into the Super Proton Synchrotron (SPS) [25], the 7-km long accelarator responsible for the Nobel-prize winning discovery of the W and Z bosons, which accelarates the protons up to 450 GeV. At this stage, the already high energy beams are finally injected in the LHC, in which they are accelerated up to the energy of 6.5 Tev, and collided in the four detectors.

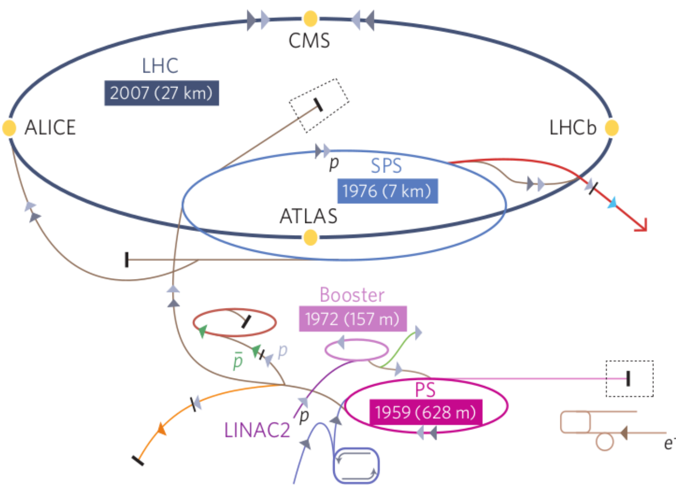


Figure 1.3: Acceleration chain of the proton beams.

The four detectors are used by researchers to address wide and independent physics programs, each aiming at testing various aspects of the Standard Model (SM). ALICE [5] is a 26 m long, 16 m high, and 16 m wide general purpose detector designed to study the physics of strongly interacting matter and of the Quark-Gluon Plasma (QGP) in the collisions of Pb ions. Its scientific program consists in studying the phase diagram of strongly interacting matter, the Quantum Chromodynamics (QCD) phase transistion and the physics of the QGP - a state of matter in which an interacting assembly of quarks and gluons admits free colour charges. Such a state is accessed when a critical energy density is reached. The required condition lasts shortly, and the out-of-equilibrium system created in the collisions undergoes dynamical evolution to a final hadronic state. The understanding of this fast-evolving system is a scientific playground in which many effective theories have been tested, providing the opportunity to apply in concert many concepts from apparently distant fields such as elementary-particle physics, nuclear physics, equilibrium and non-equilibrium thermodynamics, and hydrodynamics.

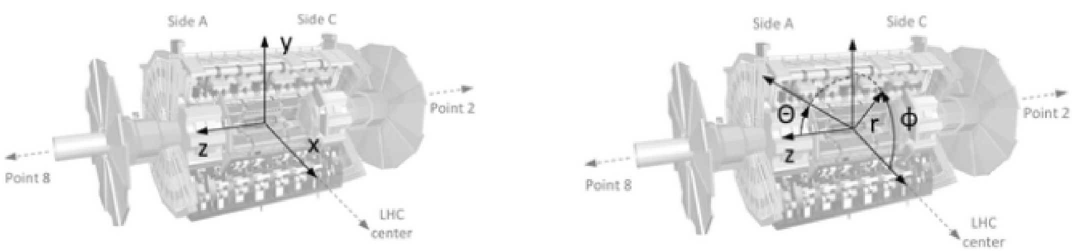
LHCb [62] is a 21 metres long, 10 metres high and 13 metres wide detector. Unlike other hermetic detectors, that surround the entire collision point, LHCb is designed to detect mainly forward particles. Its aim is to look for new physics in CP violation - a very fundamental symmetry that reflects the possible invariance of a theory under the inversion of the spatial axes and of the sign of the charges - and to study rare decays of the b- and c-hadrons. Actually, CP violation is thought to be related to the puzzle of the asymmetries between matter and antimatter observed in our Universe. It is predicted by the SM, but to an amount which cannot account for the scope of the observed asymmetry: researchers at the LHCb collaboration hope to find new sources of violation that could be accounted for in the frameworks of theories beyond the SM.

CMS [54] is a 21 metres long, 15 metres wide and 15 metres wide general-purpose detector. Its physics programme addresses topics ranging from high precision measurements of the SM to search for new physics beyond the SM. It is world wide known for being the first observer of the Higgs boson together with the ATLAS detector, which will be the focus of the next Chapter.

1.1 ATLAS detector

ATLAS [35] is a general purpose detector designed to provide as many signatures as possible using electron, photon, muon, jet and missing energy measurements. Being 46 metres long, 25 metres high and 25 metres wide, it is the largest volume particle detector ever constructed [17]. Its design was conceived to meet the following requirements [43]:

- Good electromagnetic calorimetry to identify electrons and photons and measure their energies, complemented by full coverage hadronic calorimetry to effectuate measurements of jets and missing transverse energy.
- High precision muon momentum measurements.
- Large acceptance in pseudorapidity (η) and almost full azimuthal angle (ϕ) coverage (Fig 1.4).
- Fast selection and measurements of particles from low to high-p_T, to provide high efficiencies for most of the physics at the energy scale of the collisions at the LHC.



(a) Cartesian coordinates

(b) Polar coordinates

Figure 1.4: ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point in the centre of the detector. The positive x-axis is defined by the direction from the interaction point to the centre of the LHC ring, with the positive y-axis pointing upwards, while the beam direction defines the z-axis. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity η is defined in terms of the polar angle θ by $\eta = -\ln\tan(\theta/2)$. The angular distance is defined as $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$. Rapidity is defined as $y = 0.5\ln[(E + p_z)/(E - p_z)]$ where E denotes the energy and p_z is the component of the momentum along the beam direction.

These requirements were implemented by building a cylyndrical particle detector with a near 4π coverage in solid angle and a forward-backward symmetry, consisting of an inner detector (ID) for tracking, surrounded by a superconducting solenoid providing a 2T axial magnetic field, electromagnetic and electronic calorimeters, and a muon spectrometer (Fig 1.5). In the following sections, I will give a brief overview of each component of the detector.

Magnet System

The ATLAS Magnet System [33] is made of four superconducting magnets - the two End-cap Toroids, the Barrel Toroid and the Central Solenoid - together with a magnet power system, controls, cryogenics and the refrigeration plant. The Magnet System bends particles around the detector, making it easy to contain their tracks and providing an important tool for particle identification.

Barrel Toroid The Barrel Toroid is the largest component of the Magnetic System. It is composed of eight coils assembled radially and symmetrycally around the beam axis.The toroidal field it generates

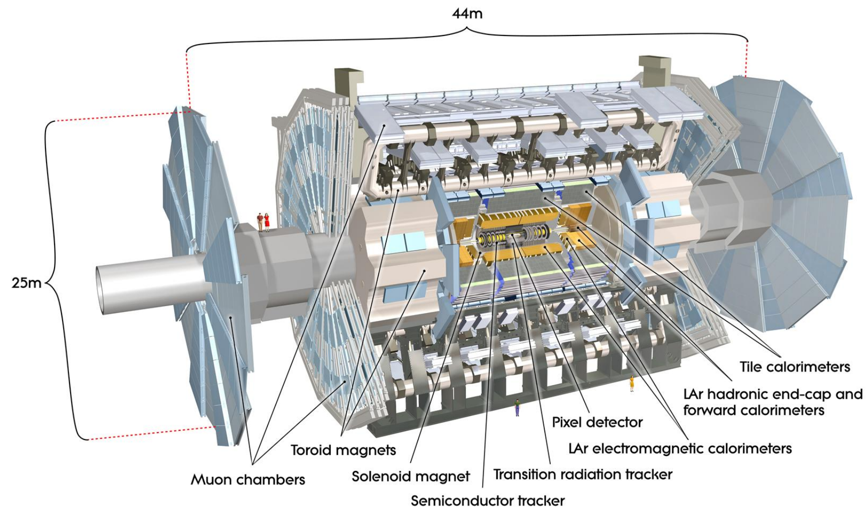


Figure 1.5: Overview of the ATLAS detector and its sub-systems. The central cylyndrical components are called *barrel*, while the two sides of the cylinder perpendicular to the beam axis are called *end-caps*.

is almost perpendicular to the track of the particles that move radially outwards from the interaction point, so it bends the charged ones in the planes perpendicular to the x-y plane. It generates a non-uniform magnetic field that peaks at 3.9 T working at a temperature of 4.7 K. The field is contained in the $0 < |\eta| < 2.7$

End-cap Toroids The End-cap Toroids, like the Barrel Toroid, are composed of eight coils radially and symmetrically assembled around the beam axis. They are rotated of 22.5° with respect to the Barrel Toroid coil system, to optimize the bending power in the interface region between each End-cap Toroid and the Barrel Toroid. The End-cap Toroids generate non-uniform magnetic fields between 4 to 8 Tm at a working point of 4.7 K covering the pseudorapidity range from 1.6 to 1.7.

Central Solenoid The Central Solenoid provides an axial field of 2T at the center of the detector, with the field lines in the beam direction. Thus, it bends the particles in the x-y plane. Since it is in front of the Electromagnetic Calorimeter, it is designed to minimize the material that composes it, while maximizing the field strength and uniformity to guarantee the performance of the inner tracker. It operates at 4.5 K.

1.1.1 Inner Detector

The Inner Detector (ID) [30] is the first ATLAS detector encountered by the collision products in their trajectories. It is immersed in the 2T solenoidal field of the Central Solenoid. It is responsible for particle tracking in ATLAS; it achieves pattern recognition, momentum measurements, vertex reconstruction and electron identification tasks, thanks to the combination of the three subdetectors that compose it: the Pixel and Semi-Conductor (SCT) trackers, and the Transition Radiation Tracker (TRT) (Fig 1.6).

The performance of the tracking system are characterized by the resolution of primary vertex (PV) positions in the z direction, which is $\mathcal{O}(100\,\mu m)$ for vertexes with more than 3 tracks; and by the resolution of the impact parameter, which is measured to be $\mathcal{O}(10\,\mu\text{m})$ [52][1].

[1] See Sec. 4.2 for the definitions of the impact parameter, the PV and the hard scatter vertex.

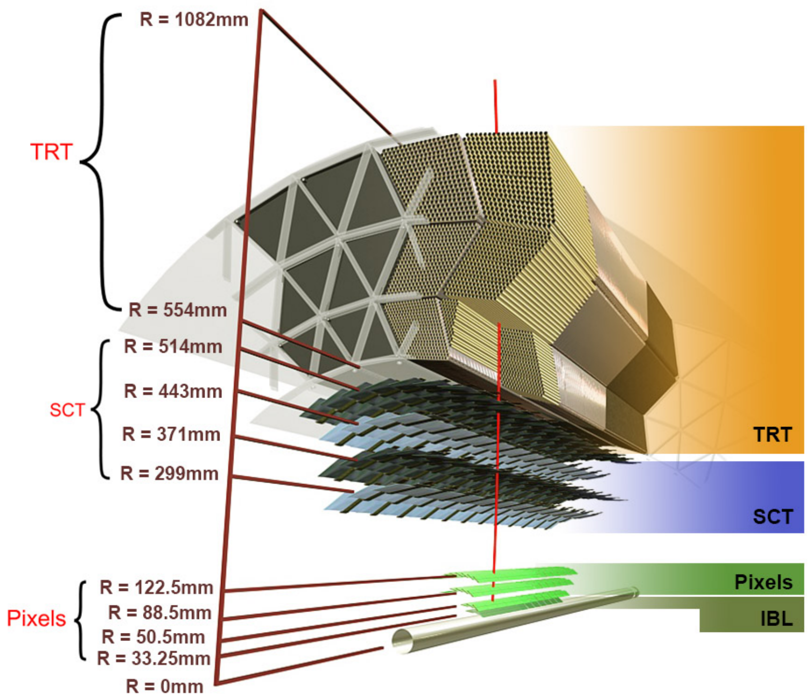


Figure 1.6: Schematic view of the Inner Detector

The given targets in momentum and vertex resolution require high-precision measurements which are made with the two innermost subdetectors thanks to their fine granularity. Highest granularity is achieved in the semiconductor pixel detectors. Unfortunately, the total number of high precision layers has an upper limit set by their high cost and mass density. So each track crosses a maximum of four pixel layers and four strip layers. The biggest number of tracking points is therefore obtained in the TRT (typically 36 per track), which requires less material and has a lower cost.

The Pixel Detector The Pixel Detector ([9],[31]) is the innermost element of the ID. It is the most important for the identification of the hard scatter vertex and possible secondary vertexes from decays in flight and for quark jet flavour identification, and provides excellent spatial resolution for the reconstruction of PVs. It is designed to provide at least four points on charged tracks emanating from the collision, and spans a pseudorapidity range $|\eta| < 2.5$. Its active region is composed of four barrel layers, the IBL, that was added during the first long shutdown of the LHC [31], the B-Layer, and the first and second Layer (Fig **??**); and of four end-cap disks per side. It covers a sensitive area of $\sim 1.7\,\mathrm{m}^2$. The building blocks of the Pixel Detector are modules, that contain the pixels and the related electronics: there are approximately 2250 of them, each one containing about 50000 pixels. In the IBL, the pixels have a size of $\sim 50 \times 250\,\mu\mathrm{m}^2$; in the remaining layers, they are $\sim 50 \times 400\,\mu\mathrm{m}^2$.

The Silicon Microstrip Trackers The SCT [44] is made of four cylinders in the barrel and nine disks in each of the two end-caps, covering a pseudorapidity range $|\eta| < 2.5$. Its active part is made of microstrip silicon sensors with a pitch of $\sim 80\,\mu$m and a length of $\sim 12\,$cm grouped in modules. The SCT provide four precision points per track in the r, ϕ and z coordinates, using small angle stereo between the strips in subsequent layers to provide the z measurement.

The Transition Radiation Tracker The TRT [53] consists of a barrel part and two end-caps. Its basic elements are drift-tubes called straws, which measure the passage of charged particles in the x-y plane. Each straw has a diameter of 4 mm. It provides both a large number of track measurements, typically 36, and electron identification capability, thanks to the use of Xenon and Argon gas to detect transition-radiation photons created in a radiator between the straws. The TRT contains 420000 electronic channels, each providing a drift-time measurement that give a spatial resolution of $\sim 170\,\mu$m per straw, and two independent thresholds that allow the detector to discriminate between tracking hits - the ones passing the lower thresholds - and transition-radiation hits, which pass the higher threshold and are the ones associated to electrons.

1.1.2 Calorimeters

Calorimeters [18] measure the energy lost by a particle when it passes trough the detector. When the particle interacts with the calorimeter, it generates a cascade of other particles, the so called *shower*. By absorbing as much of the shower as possible, it is possible to reconstruct the energy of the parent particle that generated it. To do so, calorimeters are made of layers of passive high density materials such as lead to contain the shower, interleaved with layers of active medium such as liquid argon, in which the energy deposed by the shower is measured. Electromagnetic (EM) calorimeters measure the energy of photons and electrons, while hadronic calorimeters sample the energy of hadrons. Calorimeters can stop most of the known particles except muons, which are thus measured in the Muon Spectrometer (Subsection 1.1.3) and neutrinos, that interact so weakly with matter to be reconstructable only by the missing energy in the collision products.

The ATLAS calorimeter system [29] (Figure 1.7) consists of EM and hadronic calorimeters, the Liquid Argon (LAr) Calorimeter and the Tile Hadronic Calorimeter; and covers the range $|\eta| < 4.9$.

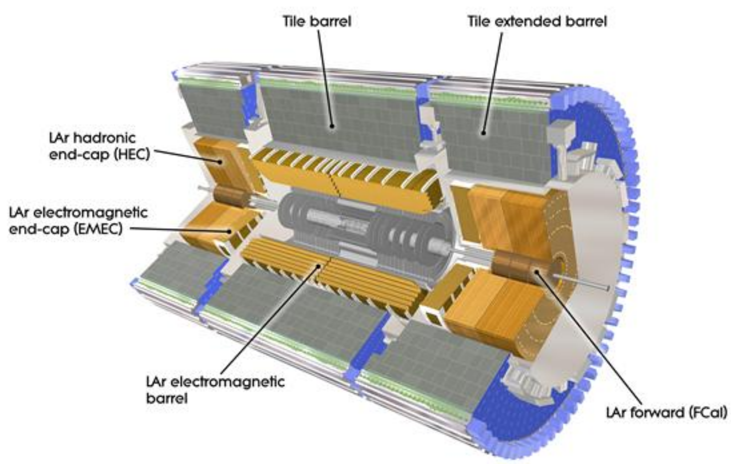


Figure 1.7: ATLAS calorimetry schematized

LAr calorimeter The LAr calorimeter [32] uses Liquid Argon as an active material due to the linearity of its response, its stability, uniformity in the detector, its radiation hardness and the possibility of shaping it into an highly granular configuration. It is divided into a barrel part, covering the $|\eta| < 1.475$ region, and two end-cap components covering $1.375 < |\eta| < 3.2$. The endcaps consist of a forward calorimeter, EM and hadronic endcaps. Over the η region corresponding to the ID, the EM calorimeter is fine-granular for precision measurements of electrons and photons. The coarse granularity of the rest of the calorimeter is sufficient for jet reconstruction and missing energy measurements.

Since calorimeters must provide good containment for the showers, their depth is an important design parameter. The EM calorimeter is > 22 radiation lengths (X_0) thick in the barrel and $> 24X_0$ in the end-caps, where a radiation length X_0 is approximately the length in which the number of particles in the shower doubles and the energy of the particles halves. The physical principle on which the LAr calorimeter is based can be summarised as follows: particles of the shower are mainly produced in the lead passive layers by the interaction of the parents with the Pb atoms. The charged ones ionize Ar atoms in the active LAr layers. Her an electron field is applied that drifts the electrons towards read-out electrodes, in which they are collected for the measurement.The design energy resolution for the EM calorimeter is a function of the energy, and is given by $\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c$, where $a = 10\%\sqrt{\text{GeV}}$ is a stochastic term related to the fluctuation of the number of particles generated at each step of the shower, $b = 200\text{MeV}$ is a noise term and $c = 0.2\%$ in the barrel is an offset term [4].

Tile Hadronic Calorimeter The Tile Hadronic Calorimeter [34] is a cylindrical structure covering the $|\eta| < 4.9$ region, with a depth of about 2 m. It is divided in 3 sections along the beam direction: the Tile Barrel and the Tile Extended Barrel. These sections are segmented in 64 modules each, constructed of iron plates (the passive layers) and of plastic scintillator tiles[2]. The produced photons are collected and multiplied for the measurement. The design energy resolution of the Tile Hadronic Calorimeter is given by $\frac{\sigma(E)}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\%$

1.1.3 Muon Spectrometer

Muons are particles that usually pass through the ID and the calorimeters losing an amount of energy small enough not to be absorbed. The Muon Spectrometer [75] is the outermost and largest part of the ATLAS detector, used to identify the muons and measure their momenta. It measures the momentum of the muons in the pseudorapidity range of $|\eta| < 2.7$, and can be used for triggering in the $|\eta| < 2.4$ region. It has a momentum resolution of $\frac{\sigma_{p_T}}{p_T} = 10\%$ at $p_T = 1\,\text{TeV}$. It is divided into four subsections named chambers: the Resistive Plate Chambers (RPC) and the Thin Gap Chambers (TGC), which provide the so called Level-1 Muon trigger; the Monitored Drift Tubes (MDT), and the Cathode Strip Chambers (CSC), in which the precision measurements of muons momentum are performed (Figure 1.8). The Level-1 Trigger is an hardware filter that aims at selecting collisions with experimental features compatible with the presence of interesting physical objects, like either high-p_T muons (identified by the above mentioned Muon trigger) or high-p_T electrons and photons reconstructed in the calorimeters.

RPC chambers Resistive Plate chambers[3] are used in the Barrel. Each chamber uses two gas volumes, Bakelite plates, and four planes read-out strips. Two layers of chambers are used to trigger on a low-p_T threshold, while an outer layer triggers on an high-p_T threshold.

TGC chambers Thin Gap Chamber provide the Level-1 muon trigger in the End-Cap. TGC are multi-wire chambers[4] operated in saturated mode (i.e. the strenght of the signal is independent of the number of ionization charges), with a mixture of carbon dioxide and n-pentane as a gas chamber. They are also used for 2nd coordinate measurements ($z - r$ plane).

[2] A scintillator is a material that radiates light when interacting with a charged particle

[3] RPCs consist of a positively charged plate, the cathode, and a negatevely charged one, the anode, both made of a very high resistivity plastic material and separated by a gas volume. Charges passing trough the gas ionize it, causing an avalanche of electrons.

[4] Multi-wire chambers consist in a set of parallel anode wires, placed between two cathode planes. The volume is filled with a gas chosen for its ionization proprieties. The charged particles that pass through the gas ionize it, and the electrons are drifted by the field to the read-out channels.

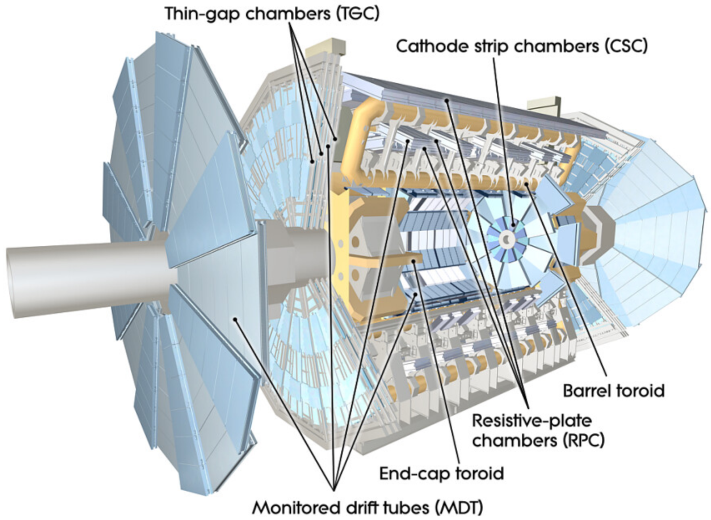


Figure 1.8: ATLAS Muon Spectrometer schematized

MDT Chambers The precision measurement of muon momentum is performed almost everywhere by the Monitored Drift Tube chambers[5]. These are drift chambers formed by aluminium tubes arranged in two multy-layers. The gas used is a mixture of Ar-CO_2. The spatial resolution of a multilayer is approximately equal to 50 μm. MDT measures the curves of tracks, from which the information on the momentum is extracted. They are also used for 2nd coordinate measurements.

CSC Chambers Cathode Strip Chambers[6] are used to perform precision measurements of the trajectory of the muons in the $2 < |\eta| < 2.7$ region of the End-Cap. The obtained spatial resolution in the bending plane is 60 μm

1.2 Highlights of LHC discoveries

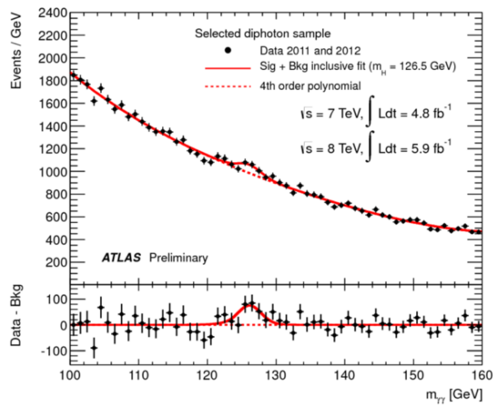
In 12 years of efforts and data taking, physicists at the LHC have been making very important new observations, while probing and excluding over a broad range of final-state phase space some overwhelmingly popular theoretical models proposed to go beyond the Standard Model [78]. They have tested many features of the Standard Model behaviour at the highest energies ever reached in a particle collider, with an unprecedented instantaneous luminosity. In this section, I will give brief overview of the main findings at the LHC, and of what someone hoped it to find but which is still missing, waiting to be discovered in the next Runs, in some future accelerator, or to disappear in the oblivion of the proposed and never observed hypotheses.

The Higgs Boson On the 4th July 2012, the ATLAS and CMS experiments announced they had both observed a new particle in the mass region around 125 GeV [47],[55] (Figure 1.9). It was the

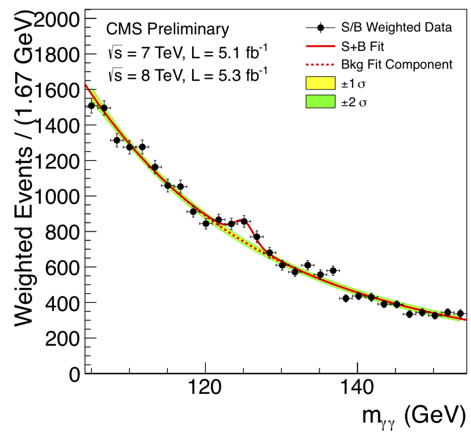
[5] Drift Chambers are multiwire chambers in which spatial resolution is achieved by measuring the time electrons need to reach the anode wire

[6] CSCs consist of arrays of positively-charged anode wires crossed with negatively-charged copper cathode strips within a gas volume. When muons pass through, they ionize gas atoms, which drift to the anode wires creating an avalanche of electrons. Positive ions move away from the wire and towards the copper cathode, also inducing a charge pulse in the strips, at right angles to the wire direction. Because the strips and the wires are perpendicular, two position coordinates for each passing particle are obtained.

first fundamental scalar particle ever discovered, the first with spin = 0, the first with a rest energy of ~ 126 GeV, ant the last missing piece of the Standard Model.



(a) ATLAS



(b) CMS

Figure 1.9: The invariant mass from pairs of photons selected in the Higgs boson to $\gamma\gamma$ analysis, as shown at the seminar at CERN on July 4th, 2012. The excess of events over the background prediction around 125 GeV is consistent with the Higgs boson predicted by the SM.

Tests of the SM LHC activity has contributed to push the limits of validity of the SM: up to now, everything we have been observing is in agreement with it. No exotic decay has been observed, no fundamental rules have been violated, no BSM particles have popped out from the data.

New bound states of exotic particles QCD imposes that composite particles must be colourless. Baryons (3 quarks), antibaryons (3 antiquarks) and mesons (quark/antiquark pairs) have long been known, but they do not exhaust all the combinations admitted by Quantum Chromodynamics. At LHCb, tetraquark and pentaquark states have been observed and measured, some of them with improved precision with respect to previous observations of other collaborations [61], some of them for the first time ([60], [57], [58]), hence corroborating QCD predictions.

Supersymmetry, Extra Dimensions, Dark Matter Particles Many hoped the LHC to find some hints pointing toward these once very popular theories. Nothing has showed up yet. Supersymmetric particles, extra dimensions or particle-based dark matter might still exist, but the most popular versions of these theories have been ruled out by the lack of evidence at the LHC.

CP-violating decays CP-violating decays had been already observed before, but the LHC is finding evidence of additional CP-violation in composite particles involving the b,s and c-quarks [59]. In particular, the amount of CP-violation in b-quarks is larger then expected; a fact that could be relevant for explaining the observed asymmetry of matter/antimatter in the Universe.

Neutral Current Conservation A prediction of the SM that constrains many BSM extensions is that it is not possible to change the flavour of a particle trough the exchange of a neutral boson. Flavour changing neutral currents are predicted by theories that add additional particles and interactions to the SM, such as Grand Unified Theories. So far, all neutral currents observed at LHC are shown to be conserved, further corroborating the SM.

Chapter 2

Theoretical background

The Standard Model of particle physics is currently the best theory to predict the outcomes of the experiments that have been performed in the last century. It is built on the framework of Quantum Field Theory (QFT), a theory born from the unification of Quantum Mechanics (QM) and Special Relativity (SR); and predicts the existence of 12 spin-1/2 elementary particles and 12 spin-1/2 antiparticles, that interact through the exchange of 9 spin-1 gauge bosons, plus a scalar field which corresponds to the mechanism by which these elementary particles acquire a non-zero mass. In this Chapter, I will give a brief overview of the SM, of the basic tools of QFT which are used to make the predictions tested at the particle colliders; and of what we know not to know: the phenomena not accounted for by our state-of-the-art theory of the nature of matter. Such an overview is given in Natural Units, in which $\hbar = c = k_B = G = 1$: thus, momentum, mass, and energy are all expressed in terms of the same unit, while length and time have the units of the inverse of an energy, and the electric charge is dimensionless.

2.1 The Standard Model

The SM is a non-abelian gauge field theory locally invariant under the action of the global simmetry group:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y . \tag{2.1}$$

In a single framework, it describes the electormagnetic, weak and strong interactions; and the fundamental particles which interact through them. Particles in the SM are divided in categories, which specify internal structures of the model built from phenomenological affinities of its constituents. The first distinction to be made is between gauge bosons, the spin-1 force carrying gauge particles, and fermions, the spin-1/2 particles that constitute ordinary matter. The spin-0 Higgs boson stands alone.

Gauge bosons, which obey the Bose-Einstein statistics, are the mediators of three out of the four fundamental interactions we know about: the strong, electromagnetic, and weak interaction. Actually, QFT explains the macroscopic interactions as resulting from particles exchanging other particles. Through this exchange, momentum is transferred. Macroscopically, the observed effect is of a force at distance influencing both the interacting particles. In this sense the exchanged particle is refereed to as *force mediating particle*, and accounts for the quite unsatisfactory ancient concept of action at distance that used to disturb Isaac Newton, about which he wrote: *"It is inconceivable that inanimate brute matter should, without the mediation of something else which is not material, operate upon and affect other matter without mutual contact"*.

The gauge bosons of the SM are:

- **Photons**: the massless mediators of the electromagnetic interaction between electrically charged particles.

- $\mathbf{W^+}, \mathbf{W^-}, \mathbf{Z}$: the massive mediators of the weak interactions, that affects all fermions.
- **Gluons**: the eight massless mediators of the strong interaction between particles carrying *colour charge*, the charge of QCD - the theory of strong interactions. Being colour-charged, gluons interact with themselves. This fact is responsible for many of the peculiar features of QCD.

Fermions obey the Fermi-statistics, which implies the esclusion principle stating that two of them cannot occupy the same quantum state. Each fermion has a corresponding antiparticle, i.e. its charge-conjugated state. In Figure 2.1 fermions of the SM are divided into 6 quarks, that carry colour charge, and 6 leptons, that are colourless. They are divided into three *generations* as well (generations being the vertical columns in the fermionic sector of Fig. 2.1). Masses of corresponding particles that belong to different generations differ by various orders of magnitude.

Figure 2.1: The Standard Model

Quarks (u, d, c, s, t, b) carry electrical and colour charge. Furthermore, they have a weak isospin: thus they interact via electromagnetic, strong and weak interaction. Quarks exhibit the phenomenon of colour confinement, explained in the framework of QCD: they are always observed in colour-neutral composite particles called *hadrons*, and never as free-particle final states. Leptons don't carry colour charge. Electrons, Muons and Taus are electrically charged and have a weak isospin, thus they interact via electromagnetic and weak interaction; while neutrinos are electrically neutral and interact only via weak interaction.

Generations differ because of phenomenological proprieties of their constituents. Each member of a generation has a greater mass than the corresponding members of the previous generations. There are no lighter particles for the first generation members to decay to: thus they compose all the stable matter we know of. Neutrino masses represent an open problem. Thanks to the observed phenomenon of neutrino flavour oscillations, it is an established fact that they are not massless. Nevertheless, up to now we have only been able to put upper limits on their masses, of $\sim$ 1eV: many orders of magnitude less than any other massive particle of the SM.

Finally, the Higgs boson is a spinless massive boson that carries no electric or colour charge. Its existence is connected to the spontaneous breaking of SU(2)$\otimes$SU(1) gauge symmetry, generating via the Brout-Englert-Higgs mechanism the masses of the massive particles. The observed fact that the Higgs is massive means that it auto-interacts.

Figure 2.2 schematizes all particle interactions in the SM.

2.2 Gauge symmetries in the SM

In QFT, a particle is described as an excitation of an underlying quantized field [82]. The dynamic equations of the fields are encoded in scalar functions called *Lagrangian densities*, from which they

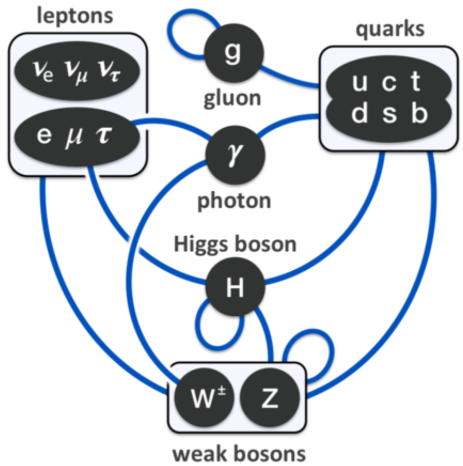


Figure 2.2: Interactions in the SM.

are extracted through the Euler-Lagrange equations. Lagrangian densities can exhibit symmetries, i.e. transformations on the fields that leave the lagrangian density invariate. Through the Nöether theorem, each symmetry is connected to a conserved quantity. Charges and interactions of the SM are theoretically obtained by imposing the Lagrangian density to be locally invariant under the gauge group simmetry $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. $SU(3)_C$ describes the colour symmetry of quarks and is related to strong interactions. $SU(2)_L \otimes U(1)_Y$ is the symmetry structure of the electroweak theory: $U(1)_Y$ is the hypercharge symmetry group and $SU(2)_L$ is related to the weak interactions. The complete Lagrangian density of the SM is the sum of two terms individually invariant under the subscripted symmetries:

$$\mathcal{L}_{SM} = \mathcal{L}_{SU(3)} + \mathcal{L}_{SU(2)\times SU(1)} \tag{2.2}$$

2.2.1 Lagrangian and Lagrangian density

In classical dynamics, the Lagrangian function is

$$L(q_i, \dot{q}_i) = T(q_i, \dot{q}_i) - V(q_i, \dot{q}_i), \tag{2.3}$$

where T is the kinetic and V is the potential energy of the considered system[1]. q_i are the generalised coordinates that, depending on the considered system, can stand for spatial coordinates, displacements, angles, etc.; and $\dot{q}_i$ are their time derivatives. The equations of motion of the system are entirely specified by the Lagrangian through the Euler-Lagrange equations,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0 \tag{2.4}$$

which, for classical mechanics, can be shown to be equivalent to the Newton's equations.

For continuous systems such as fields (i.e. continuous functions ϕ of the space-time coordinates x^μ), the discrete treatment given above is extended by replacing the Lagrangian L with a Lagrangian density $\mathcal{L}$,

$$L(q_i, \dot{q}_i) \longrightarrow \mathcal{L}(\phi_i, \partial_\mu \phi_i);$$

[1] For simplicity's sake, let's consider a case in which the Lagrangian doesn't depend explicitly on time (as it could do, for example, through a time evolving electric field).

in which the generalised coordinates are replaced by the fields $\phi_i(t,x,y,z)$ and the time derivatives are replaced by

$$\partial_\mu \phi_i \equiv \frac{\partial \phi_i}{\partial x^\mu}$$

L is obtained by $\mathcal{L}$ through the integration over the spatial coordinates

$$L = \int \mathcal{L} \mathrm{d}^3\mathbf{x}.$$

The *action* S is obtained by further integration of the Lagrangian on time

$$S = \int L \mathrm{dt}.$$

Action is important because in the discrete case it can be proved that solutions of the Euler-Lagrange equation are also solutions of the principle of Least Action equation,

$$\delta S = \delta \int_{t_1}^{t_2} L(\mathbf{q}(\mathrm{t}), \dot{\mathbf{q}}(\mathrm{t}))\mathrm{dt} = 0 \tag{2.5}$$

which is solved by the trajectory in the generalised coordinates space between the configurations at time t_1 and t_2 that minimizes the action. The equivalent of the Euler-Lagrange equations for the continuous case are hence obtained by imposing the Least Action principle to the action $S(\phi_i, \partial_\mu \phi_i)$ obtained via the above-mentioned substitutions of the generalised coordinates with the fields:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)} \right) - \frac{\partial \mathcal{L}}{\partial \phi_i} = 0 \tag{2.6}$$

2.2.2 Nöether's Theorem

Let's go back for a second to the discrete case to better illustrate an important concept. If the Lagrangian of a system doesn't depend on a particular coordinate q_i, Eq. 2.4 implies

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial \dot{q}_i} \right) = 0.$$

Consequently,

$$Q = \left(\frac{\partial L}{\partial \dot{q}_i} \right)$$

is a conserved quantity of the system. It means that a symmetry of the Lagrangian (i.e. the invariance of L under changes of the coordinate q_i) is connected to a conserved quantity. This is a simple example of the Nöther's theorem, that states that every differentiable symmetry of the action of a physical system has a corresponding conservation law.

In field theory, Noöether's theorem connects symmetries of the Lagrangian to conserved currents j^μ that satisfy continuity equations $\partial_\mu j^\mu = 0$, thus implying the existence of conserved charges.

An example: global and local Gauge invariance in Quntum Electrodynamics The Lagrangian density for a free spin-1/2 particle is

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi, \tag{2.7}$$

which is left unchanged by a global U(1) phase transformation of the fields:

$$\psi \to \psi' = e^{i\theta}\psi. \tag{2.8}$$

It can be shown that the conserved current corresponding to this symmetry is the four-vector current

$$j^\mu = \bar{\psi}\gamma^\mu\psi. \tag{2.9}$$

and the related conserved charge is the electric charge. If one requires the Lagrangian density to be invariant under *local* phase transformations

$$\psi(x) \to \psi'(x) = e^{iq\chi(x)}\psi(x), \tag{2.10}$$

something interesting happens. The dependence on x of the phase transformation in Eq. 2.10, responsible for the name *local* given to the transformation, implies that the derivatives acting on the fields act on the phase as well. Thus the Lagrangian density is not left invariant by the transformation:

$$\begin{aligned} \mathcal{L} \to \mathcal{L}' &= i\bar{\psi}e^{-iq\chi}\gamma^\mu\left[e^{iq\chi}\partial_\mu\psi + iq(\partial_\mu\chi)e^{iq\chi}\psi\right] - m\bar{\psi}e^{-iq\chi}e^{iq\chi}\psi \\ &= \mathcal{L} - q\bar{\psi}\gamma^\mu(\partial_\mu\chi)\psi \neq \mathcal{L}. \end{aligned} \tag{2.11}$$

The Lagrangian density can be made invariant by replacing the derivative ∂_μ in Eq. 2.10 with the *covariant derivative* D_μ,

$$\partial_\mu \to D_\mu = \partial_\mu + iqA_\mu,$$
$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - q\bar{\psi}\gamma^\mu A_\mu\psi,$$

where A_μ is a new field. If A_μ transforms as

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu\chi, \tag{2.12}$$

when $\mathcal{L}$ is transformed the extra term $q\bar{\psi}\gamma^\mu(\partial_\mu\chi)\psi$ of Eq. 2.11 is cancelled by a corresponding term of the opposite sign produced by the covariant derivative. Eq. 2.12 shows that the transformation of A_μ has the same form of the gauge transformation of the electromagnetic vector potential in classical EM. Thus, the local U(1) phase transformation is referred to as a *local gauge transformation* and A_μ is identified as the photon field.

Requiring the local gauge invariance has forced us to introduce a new term of the form $qj^\mu A_\mu$, with j^μ given by Eq. 2.9, that describes the interaction between a current and a new field. This is a general feature of field theories: *imposing a local invariance of the Lagrangian density causes the emergence of an interaction term.* The full Lagrangian density of Qunatum Electrodynamics (QED) is obtained by adding the kinetic term $-\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$ describing the free photons , in which $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$; and by choosing $q = -e$:

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu\partial_\mu - m_e)\psi + e\bar{\psi}\gamma^\mu A_\mu\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{2.13}$$

The kinetic term $-\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$ is automatically locally gauge invariant. From $\mathcal{L}_{QED}$ the covariant form of the Maxwell equations $\partial^\mu F^{\mu\nu} = j^\nu$ is obtained as one of the Euler-Lagrange equations.

2.2.3 SU(3)

Just like QED is the QFT associated to a U(1) local gauge symmetry of the Universe, QCD, the theory of strong interactions between quarks, is associated to an invariance of the Yang-Mills Lagrangian density

$$\mathcal{L}_{YM} = \bar{\Psi}^{\alpha,A}(i\partial\!\!\!/ - m_A)\bar{\Psi}^{\alpha,A} + gA^a_\mu\bar{\Psi}^{\alpha,A}\gamma^\mu T^a_{\alpha\beta}\Psi^{\beta,A} - \frac{1}{4}F^{\mu\nu a}F^a_{\mu\nu} \tag{2.14}$$

under SU(3) local phase transformations:

$$\psi(x) \to \psi'(x) = \exp[ig_s\boldsymbol{\alpha}(x)\cdot\hat{\mathbf{T}}]\psi(x). \tag{2.15}$$

In Eq. 2.14- 2.15:

- $\Psi^{\alpha,A}$ are the quark matter fields, which live in the vector space acted-on by the fundamental representation of the gauge group. $A = u, d, c, s, t, b$ is the flavour index and $\alpha = 1, 2, 3$ is the colour index.
- $\partial\!\!\!/ = \gamma^\mu \partial_\mu$, with γ^μ the four 4×4 Dirac matrices.
- T^a are the eight generators of SU(3) in the fundamental representation
- $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g\epsilon_{akj} A^j_\mu A^k_\nu$, where A^a_μ are the eight gluon gauge fields and ϵ_{akj} is the Levi-Civita tensor.
- g is the gauge coupling constant
- $\alpha(x)$ are eight functions which specify the local phase at each point in space-time

The term $g\epsilon_{akj} A^j_\mu A^k_\nu$ in $F^a_{\mu\nu}$, that is not present in the EM tensor of QED, accounts for the gluon auto-interaction, which is involved in many phenomenological features of the QCD, e.g. hadronisation and colour confinement.

2.2.4 SU(2)⊗ U(1)

Mathematically, electromagnetism and weak interactions are unified as a Yang-Mills theory with a $SU(2) \otimes U(1)$ local gauge symmetry, in which fermion fields interact with 4 gauge fields. These gauge fields give rise to massless bosons, the three W bosons of weak isosospin W_1,W_2 and W_3; and the B boson of weak hypercharge. These gauge fields are not the physical fields that we observe in Nature: due to the Higgs mechanism, a *spontaneous symmetry breaking*

$$SU_L(2) \otimes U(1)_Y \to U(1)_{EM}$$

occurs [83] that mixes the W_3 and B bosons into two bosons with different masses, the photon and the Z^0:

$$\begin{pmatrix} \gamma \\ Z_0 \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix} \tag{2.16}$$

where θ_W is the *weak mixing angle*. While an illustration of the spontaneous symmetry breaking is beyond the scope of this thesis, in the following I will give a brief overview of $\mathcal{L}_{SU(2)\otimes U(1)}$ before the symmetry breaking occurs.

Before the electroweak symmetry breaking, the Lagrangian density of the electroweak interaction can be divided into four parts:

$$\mathcal{L}_{SU(2)\otimes U(1)} = \mathcal{L}_f + \mathcal{L}_g + \mathcal{L}_h + \mathcal{L}_y. \tag{2.17}$$

$\mathcal{L}_f$ is the Dirac-like lagrangian density for the fermions of the SM

$$\mathcal{L}_f = \bar{Q}_i i D\!\!\!\!/\, Q_i + \bar{u}_i i D\!\!\!\!/\, u_i + \bar{L}_i i D\!\!\!\!/\, L_i + \bar{e}_i i D\!\!\!\!/\, e_i,$$

where the subscript i runs over the tree generations of fermions; Q,u and d are the left-handed doublet, right-handed singlet up and right-handed singlet down quark fields; and L the left handed leptonic fields; leaving in the vector space acted-on by the fundamental representation of $SU(2)$[2]. $\mathcal{L}$ has been made invariant with respect to $SU(2)\otimes U(1)$ local phase transformations by substitution of the derivative with the covariant derivative

$$\partial_\mu \to D_\mu \equiv \partial_\mu + i\frac{g'}{2} Y B_\mu + i g_W \mathbf{T} \cdot W_\mu, \tag{2.18}$$

[2]See [82] for an overview of the important role of chirality in the electroweak theory.

where $\mathbf{T}$ are the three generators of SU(2) in the fundamental representation, B_μ and $\mathbf{W}_\mu$ are the gauge fields, Y is the hypercharge and g_W and g' are the gauge couplings respectively for SU(2) and U(1).

The $\mathcal{L}_g$ term

$$\mathcal{L}_g = -\frac{1}{4}W^{\mu\nu a}W^a_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu},$$

in which $W^{\mu\nu a}$ and $B^{\mu\nu}$ are the gauge tensors analogous to those of QCD and QED, is the free term for the gauge fields. The $\mathcal{L}_h$ and $\mathcal{L}_y$ terms describe the Higgs field, a scalar field necessary to explain the observed masses of the weak gauge bosons, and its interaction with gauge bosons and fermions. To describe them, I will briefly introduce the concept of spontaneous symmetry breaking.

2.2.5 Spontaneous symmetry breaking in the $\lambda\phi^4$ theory and the Higgs field

Consider a scalar field ϕ with a Lagrangian density given by a kinetic and a potential term:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - V(\phi). \tag{2.19}$$

If

$$V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4, \tag{2.20}$$

the theory obtained is the famous $\lambda\phi^4$ theory. In Eq. 2.20 the term $\frac{1}{2}\mu^2\phi^2$ is the mass term of the particle[3] and the $\frac{1}{4}\lambda\phi^4$ term is a self-interaction term corresponding to the four-point interaction of Fig 2.3. The Lagrangian density of Eq. 2.19 is symmetric for the transformation

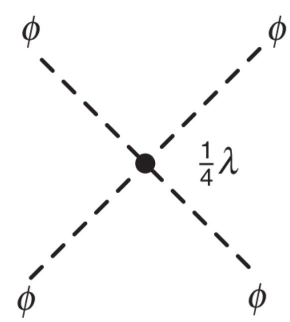


Figure 2.3: The four point interaction for a scalar field with the potential in Eq. 2.20

$$\phi \to \phi' = -\phi.$$

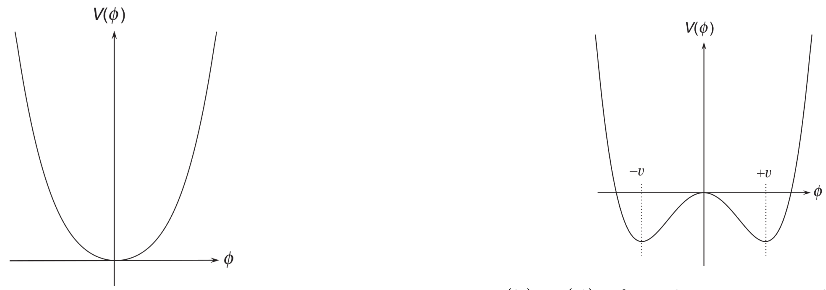
The lowest energy state of the field ϕ is called the *vacuum state* and corresponds to the minimum of the potential of Eq. 2.20. In this model, if $\mu^2 > 0$ the potential has a minimum at $\phi = 0$, and thus the vacuum state is the non-degenerate state $\phi = 0$ (Fig. 2.4a); while if $\mu^2 < 0$ the potential acquires the shape of Fig. 2.4b. In this second case *the vacuum state is degenerate*, because the potential has *two* minima:

$$\phi = \pm v = \left|\sqrt{\frac{-\mu^2}{\lambda}}\right|.$$

The vacuum state of the field will be $\phi = +v$ or $\phi = -v$. Interestingly, the actual state breaks the symmetry of the Lagrangian. This is a simple example of *spontaneous symmetry breaking*: the potential

[3] In QFT the coefficient of a term of the Lagrangian density quadratic in a field is the mass of the excitations associated to that field.

of Fig. 2.4b preserves the symmetry of the Lagrangian density, but, due to the degeneration of the ground state, the actual physical state does not. Spontaneous symmetry breaking is ubiquitous in physics (e.g. the direction of the net magnetisation of a ferromagnet breaks the rotational symmetry of the related Lagrangian).



(a) $V(\phi)$ before the symmetry breaking

(b) $V(\phi)$ after the symmetry breaking

Figure 2.4

In the SM, $SU(2) \otimes U(1)$ is a spontaneously broken symmetry of the $\mathcal{L}_{SU(2)\otimes U(1)}$ Lagrangian. The rupture is explained by the Higgs mechanism, in which a weak isospin doublet of complex scalar fields

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.21}$$

is added to the model with the Lagrangian density

$$\mathcal{L}_h = (\partial_\mu \phi^\dagger)(\partial^\mu \phi) - V(\phi), \tag{2.22}$$

where

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2. \tag{2.23}$$

For $\mu^2 < 0$, the potential has an infinite set of degenerate minima, ultimately causing the simmetry breaking.

No mass term is explicitly present in the SM Lagrangian, as it would break the $U(2) \otimes U(1)$ gauge symmetry . Nevertheless, through spontaneous symmetry breaking, a process known as Brout-Englert-Higgs mechanism explains the emergence of the masses of the massive gauge bosons from their coupling with the Higgs field that compares in Eq. 2.4b.

Last is the $\mathcal{L}_y$ term, which accounts for the interaction of the fermions with the Higgs field. Before simmetry breaking, it can be written as

$$\mathcal{L}_y = -g_i[L_i \phi R_i + \bar{R}_i \phi^\dagger L_i], \tag{2.24}$$

where R_i are the leptonic singlet fields and g_i are the *Yukawa couplings* of the leptons to the Higgs field. A generalized Brout-Englert-Higgs mechanism can be shown to generate the mass terms of the fermions as well, their values being proportional to their Yukawa couplings with the Higgs field.

2.3 Phenomena not explained and theoretical issues in the Standard Model

The SM has been tested for decades, and not one measure has ever been found to deviate from it significantly. Despite its power of predictivity and outstanding successes, the SM is not the end of the story. It comes with many free parameters, a set of theoretical conundrums yet to be resolved, and it is unable to explain various observed phenomena. In the following, I will give an overview of its main known issues.

Gravity The SM does not include gravity. The approach of adding a *graviton* to the SM does not yield to predictions that fit experimental data without further modification of the SM. Furthermore, a quantized field theory of gravity is still searched by the theorists, string theory being a controversial possibility.

Dark matter and dark energy Cosmological observations of the velocity distribution of stars in galaxies and of the rate of expansion of the observable Universe tell us that the SM explains about 5% of the energy present in the Universe [20]. From these observations 26% of the matter in the Universe could be made by an unknown weak interacting type of matter only sensible to gravitational forces - the so-called dark matter - whose hypothetical particles have never been observed in colliders.

Dark energy would make up the remaining 68% of energy of the Universe. It seems to be evenly distributed throughout the Universe, meaning that it has only a global effect on the Universe as a whole. The even distribution of dark energy in the Universe suggests a link with the vacuum. Interestingly enough, attempts to explain it in terms of the vacuum energy of the SM lead to a spectacular mismatch of 120 orders of magnitude with the observed values [2].

Neutrino masses According to the SM, neutrinos could be massless. Neutrino oscillation experiments have proved that they are massive instead. Neutrino mass terms can be added to the SM, but this leads to other problems: for example, it is not clear if they would arise with the Higgs-mechanism like it happens for the other massive particles in the SM.

Matter-antimatter asymmetry The predicted amount of CP-violation in the SM is not enough to explain the survival of matter over anti-matter in the early stages of the Universe as pictured by the Big-Bang theory.

Hierarchy problem In the SM, the loop corrections to the mass of the Higgs diverge at very large energies. To cancel out this divergence, the values of some of the free parameters of the SM must be accurately chosen - a process known as *fine-tuning*. A very small displacement from the fine-tuned values is expected to imply a very different mass of the Higgs, and a very different Universe from the one we know. Were we this lucky? Do we live in the only zone of the Universe in which these values get the right number to allow life to develop (anthropic principle)? Or is there a deeper and more elegant explanation?

Strong CP problem The SU(3) local gauge symmetry of QCD allows for a term that would not conserve CP symmetry, but no CP-violation QCD process has ever been observed. It seems that the coefficient of this term is exactly zero: another exemple of fine-tuning.

Chapter 3

Jets

A jet is a narrow cone of hadrons and other particles all originating from the hadronization of a quark or a gluon produced in a collision at a particle collider. Jets are very important instruments and objects of study both for experimental and phenomenological particle physics, and are of primary importance for the research conducted and presented in this thesis. They are a key object in both theoretical and experimental QCD: their emergence can be theoretically explained by the process of hadronisation of colour-charged particles and their production cross section can be computed by theorists; while experimentalists have developed techniques to reconstruct them in the detector and use them to test the predictions. In the first part of this Chapter, I will talk about jets from both the points of view - the theorist's and the experimentalist's - and I will focus on the reconstruction algorithms employed at the LHC.

3.1 From colour confinement to hadronisation to jets

Despite the various experimental attempts in the past decades, no one has ever observed a free quark directly. This experimental fact can be explained by *colour confinement*, which states that coloured objects are always confined to colour singlet states. It means that no object with a non-zero colour charge can propagate as a free particle. This feature of QCD is believed to arise because of the non-zero colour charge of gluons and their consequent self interaction. A qualitative understanding of how the interacting gluons could be responsible for the confinement can be obtained by thinking about what happens when two quarks are pulled apart. Quarks interact via exchange of virtual gluons. Because these gluons carry colour charge, they themselves interact with each other via the exchange of virtual gluons, which cause an attractive force between gluon and gluon. These newly exchanged gluons, carrying colour charge, interact via the exchange of coloured gluons... and so on (Fig 3.1a). All these self-interactions result in a squeezing of the colour field between the quarks into a tube (Fig 3.1b).

At relatively large distances, the energy density contained in the tube is constant, giving rise to an effective potential $V(\mathbf{r}) \sim kr$, proportional to the distance between the quarks. k is experimentally measured to be $\sim 1\text{GeV/fm}$, corresponding to a force of the order $\mathcal{O}(10^5)$ N between two unconfined quarks. The form of the potential also implies that to separate two quarks to infinity and ultimately set them free, one would need an infinite amount of energy.

When a quark or a gluon is produced in a collision, it undergoes a process known as *hadronisation* (Fig 3.2).

Here is an example of it: initially, the quark and antiquark separate at high velocities. At a certain separation, the energy stored in the colour field, which increases almost linearly with separation, is sufficient to produce a new $q\bar{q}$ pair that breaks the initial tube of colur field into two smaller strings. This process continues and new pairs are produced until all the quarks and antiquarks have sufficiently low energy to combine into colourless hadrons. The result of this process are two "hadron showers"

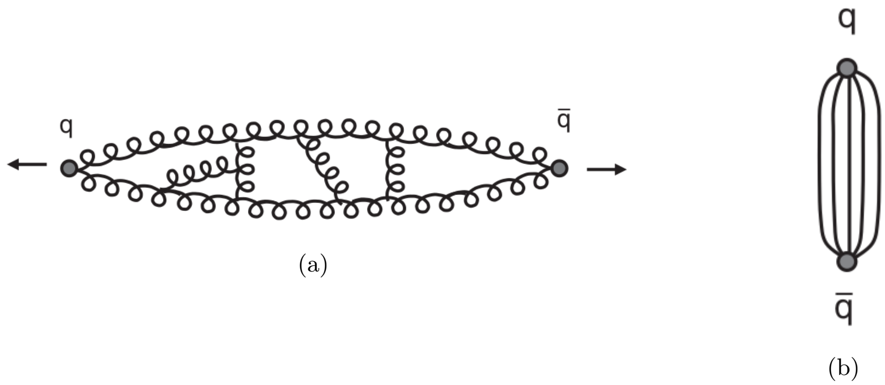


Figure 3.1: Qualitative illustration of the effect of the gluon self-interactions. Since gluons exchanged by coloured particles attract each other (a), the colour field is squeezed in a tube between the interacting quarks (b).

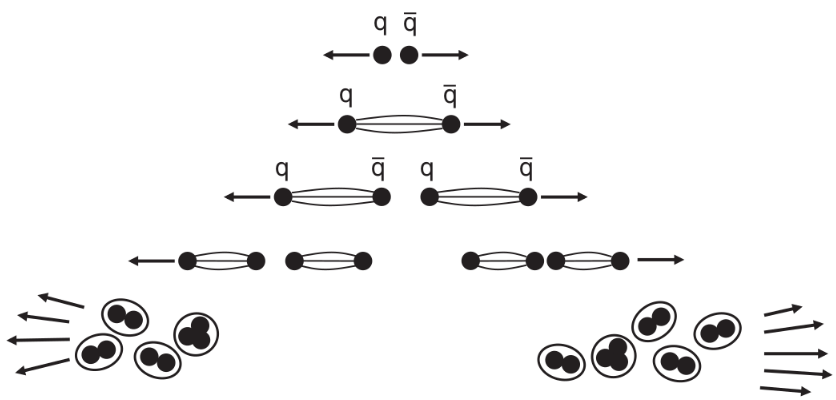


Figure 3.2: Hadronisation process step-by-step. The colour field between the separating quarks stretches until it stores enough energy to create a new pair. The process continues until all the quarks and antiquarks have sufficiently low energy to combine into colourless hadrons.

called *jets*, one in the first quark direction, one in the first antiquark's. Thus, in the colliders, quarks and gluons produced in the hard scattering process are *always observed as jets of hadrons.*

3.2 Jet reconstruction algorithms

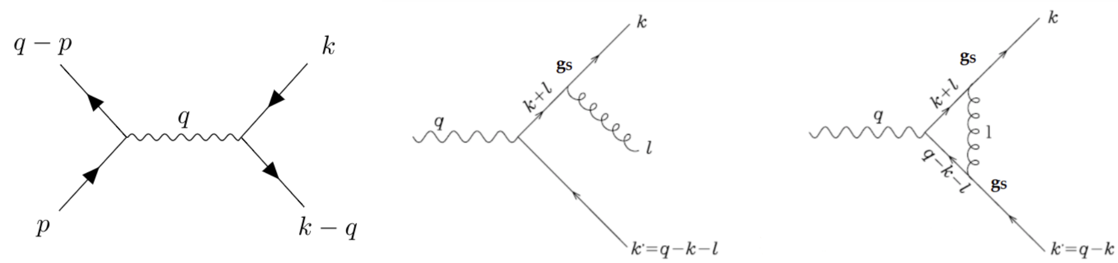
As a consequence of the above described effect, we are not able to detect directly the hard scattering products of a QCD process. What we do observe are energy deposits in the calorimeters, and hits in the tracking detectors left by the charged colourless final states of the hadronisation process. To reconstruct the kinematic properties of the initial quark/gluon, it is thus necessary to study together the signals left in the detector by the particles originating from the same parton[1]. Jet reconstruction algorithms are clustering algorithms that use the calorimetry, the tracking, or both the information to define jets [8]. An ideal algorithm is able to reconstruct the physical jet of the hadronisation of the underlying hard-scattering product. It is able to reconstruct the kinematics of the jet (and thus the

[1] Parton is the name that addresses both gluons and quarks.

energy of the parent parton), its charge and its structure.

There are some aspects that need to be considered when building a jet reconstruction algorithm: jet size and infra-red and collinear (IRC) safety. For the sake of simplicity, let's imagine the reconstructed jets to be perfectly cone-shaped, with their origins located at the collision point. The size of the jet is the maximum angular opening allowed for the cone. A large-enough jet radius is important because it allows the jet to capture enough of the hadrons to accurately compute the jet mass and energy, while an exceedingly big radius increases the risk of including in the jet large amounts of objects extraneous to the particular hadron-shower in exam, leading to an overestimation of the jet mass and energy.

Infrared and collinear safety are important requirements to match the previsions of theorists due to a feature of the cross sections of QCD processes. To understand it, consider the Feynman diagrams of Figure 3.3.



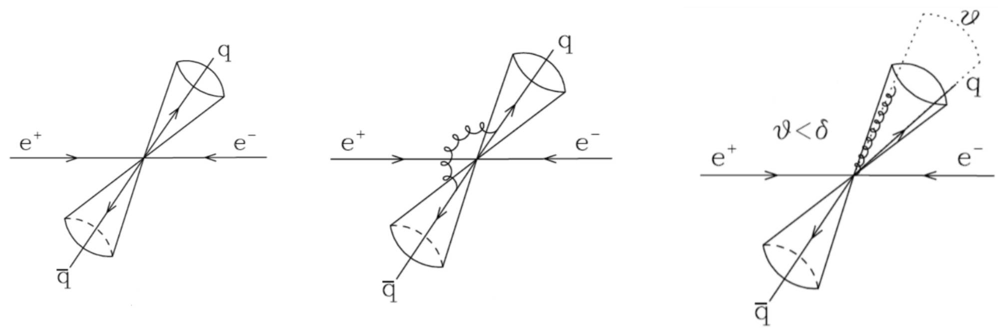
(a) Born Level

(b) First order corrections

Figure 3.3: Feynman diagram of the $e^+e^- \to q\bar{q}$ EM interaction (a) and first order QCD corrections (b).

Figure 3.3a shows the Feynman diagram of the $e^+e^- \to q\bar{q}$ interaction. This diagram, called the Born level diagram, is purely electromagnetic. Its amplitude probability can be computed fairly simply with the techniques of QFT. The problem arises when considering the QCD quantum corrections to the Born level. As it is known from the QCD lagrangian (Eq. 2.14), quarks can emit gluons. Hence, the process illustrated in the right diagram of Fig. 3.3b, in which a virtual gluon is exchanged between the separating quark and antiquark is admitted by QCD. It ends up in a final state indistinguishable from the Born level: the rules of QFT state that, in such cases, the two amplitudes must be added when computing the probability for the process to occur. Unfortunately, the contribution of these quantum corrections diverges. Since a physical probability cannot be divergent, a cancellation must be introduced by some other process. By computing the amplitude of the left diagram of Fig. 3.3b, in which a real gluon in radiated by one of the quarks, it can be shown that when the gluon is emitted collinearly to the quark or when it is emitted with vanishing energy, the amplitude probability diverges. The first scenario is called *collinear* divergence, while the second is callled *infrared.* These two divergences combined are exactly the opposite of the divergence encountered when the virtual gluon is exchanged between the quark and antiquark. The infrared and collinear gluon radiation scenarios represent the cases in which the final state of the left of Fig. 3.3b becomes indistinguishable from that of Fig. 3.3a. Thus, the amplitude probabilities of infrared and collinear gluon emissions must be added to those of the Born Level and of the first order loop correction when computing the probability: hence, all the divergences cancel out.

When computing the cross section of jet productions, theorists, well aware of the problem illustrated above, to guarantee the cancellation of the divergences define the jets so that the Born level process, the virtual correction and the infrared and collinear radiation all end up in the same configuration (Fig. 3.4). If they failed to do so, e.g. if they defined jets so that the collinear gluon in Fig. 3.4c ended up in a different jet from the nearest quark, the three cases in figure would not cancel out the infinity when computing the cross section. An observable that succeed in dodging the problems of infrared and collinear divergence is called *infrared and collinear safe.*



(a) Born Level

(b) Virtual gluon

(c) Gluon radiation

Figure 3.4: IRC safe jets: the Born Level (a) and the first order diverging QCD corrections of the emission of a virtual gluon and of the radiation of a collinear gluon all result in the same jets. When computing the cross section of the production of the jets in the figures, thus, the divergence is cancelled out. If the situation of (b) or (c) ended up in different jet configurations (e.g. in three jets for (c)), the amplitudes of (b) and (c) could not be summed, and the divergences would not be cancelled out perturbatively.

In order to match the predictions, jet reconstruction algorithms must be programmed to build IRC-safe jets. in this thesis, we will adopt a jet reconstruction algorithm called anti-K_t [15], an IRC-safe sequential clustering algorithm belonging to the K_t family. Sequential clustering algorithms build jets with a bottom-up approach: they combine particles starting from closest ones with respect to an ad-hoc notion of distance. K_t family algorithms assume that particles belonging to the same jet have similar transverse momenta. Hence, they group particles based on proximity with respect to a distance d_{ij} that takes into account the transverse momenta. In the anti-K_t case it is

$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2}) \frac{\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2}{R}, \tag{3.1}$$

where $\Delta\eta_{ij}$ and $\Delta\phi_{ij}$ are the angular distances between the particles, and R is a configuration parameter ($R = 0.4$ in this thesis).

A distance in the transverse momentum space between each particle i and the beam axis d_{iB} is also defined to be

$$d_{iB} = p_{Ti}^{-2}. \tag{3.2}$$

The recipe is the following: for each particle i, the minimum of the entire set $\{d_{ij}, d_{iB}\}$ is found. If d_{ij} is the minimum, then the particles i and j are combined into one by summation of their four-momenta, and i and j are removed from the set of particles. If d_{iB} is found to be the minimum instead, i is treated as the constituent of a new jet. The definition of d_{ij} in Eq. 3.1 states that for the anti-k_T algorithm two low-p_T particles tend to be far away. Thus, the jets grow around high-momentum core particles. The result are jets with sizes that only fluctuate slightly and almost perfectly cone shaped (Fig. 3.5).

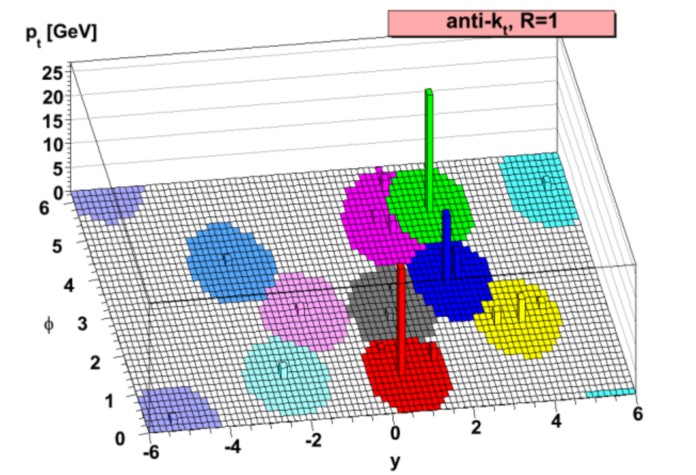


Figure 3.5: Transverse momentum and spatial distribution in the ϕ and y coordinates (Fig. 1.4) of the location and width of the jets reconstructed with the anti-k_T algorithm. The areas are very similar, and the shapes are almost perfectly circular.

Chapter 4

b-tagging in ATLAS

A key role to make jets an effective tool to study the fundamental interactions at colliders is played by flavour tagging: the process of identifying the flavour of the quark from which a hadronic jet originated. The present thesis focuses on b-tagging algorithms - those that try to discriminate jets originated by b-quarks from those originated by c- or u, s, d-quarks. In this Chapter I will present the b-quark, its discovery and main features; I will give an overview of the observables used for b-tagging and I will present the state-of-the-art algorithms used in the ATLAS collaboration.

4.1 The beauty quark

The beauty quark is the negatively charged third generation quark. It was first introduced theoretically in 1973 by Makoto Kobayashi and Toshihide Maskawa, who, in a Nobel Prize-winning paper, proved that a new coloured fermionic field was needed in addition to the up, down, strange and charm fields to explain the observed phenomena of CP-violation [72]. Such field was discovered in 1977 by the Fermilab E288 experiment, who observed three resonances in the combined mass of $\mu^+\mu^-$ pairs in the 9-10 GeV region produced by colliding 400 GeV beams of protons on a copper target (Fig. 4.1) [63],[67]. The three observed resonances were bound states of $b\bar{b}$ that have been called Υ(1S), Υ(2S) and Υ(3S).

The b-quark's bare mass is around 4.18 GeV [3] and its electric charge is -1/3 e. It has an average lifetime of $\tau_0 = (1.57 \pm 0.01) \times 10^{-12}$ s, which corresponds to a flight length d for a b-hadron in the laboratory frame:

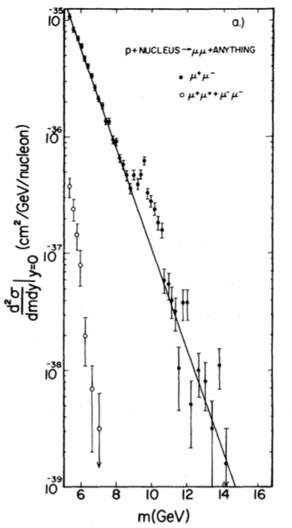
$$d \sim \gamma\beta c\tau_0; \tag{4.1}$$

where β is the relativistic factor and γ the Lorentz factor. The value of τ_0 is relatively high with respect to the other quarks. Thus the b-hadron decay can occur at a point sufficiently displaced to be resolved from the primary vertex, leading to a very specific signature of the quark.

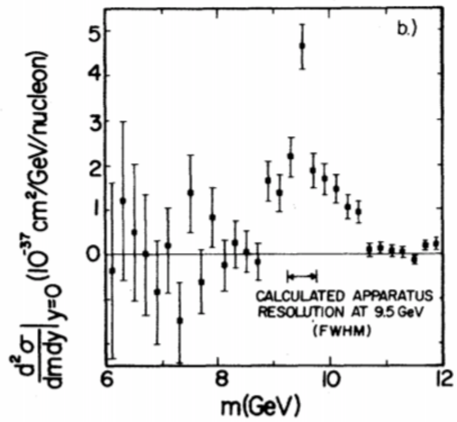
The b-quark can decay via the weak interaction to a u or d quark. The matrix that contains the information about the weak decay rates is the Cabibbo-Kobayashi-Maskawa (CKM) matrix:

$$\begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} = \begin{bmatrix} 0.97401^{+0.00011}_{-0.00011} & 0.22650^{+0.00048}_{-0.00048} & (3.61^{+0.11}_{-0.09}) \times 10^{-3} \\ 0.22636^{+0.00048}_{-0.00048} & 0.97320^{+0.00022}_{-0.00022} & (40.53^{+0.83}_{-0.61}) \times 10^{-3} \\ (8.54^{+0.23}_{-0.16}) \times 10^{-3} & (39.78^{+0.82}_{-0.60}) \times 10^{-3} & 0.999172^{+0.000024}_{-0.000035} \end{bmatrix} \tag{4.2}$$

Each entry represents the symmetric contribution of the weak interaction to the decay rate between the two subscripted quarks. To compute from the CKM the decay rate of the quarks, also the mass difference between the examined pair must be taken in consideration; thus, for example, the V_{tb} term



(a)



(b)

Figure 4.1: Excess in the $m_{\mu\mu}$ mass distribution as first observed by the Fermilab E288 experiment (a) before and (b) after background subtraction [63],[67]

forces the t-quark to decay almost always to a b-quark, while the b cannot decay to a t, the t being more than 40 times heavier than the b. The low values of the V_{ub} and V_{cb} terms are responsible for the long lifetime of the b.

4.2 Key ingredients for b-jet identification

Jets in ATLAS are labelled according to the heaviest quark of $p_T \geq 5$ GeV they contain within a cone of size $\Delta R = 0.3$ around the jet axis [46]. Thus, if a jet contains a b-hadron it is labelled as b-jet; as c-jet if it contains a c-hadron and no b-hadrons; otherwise it is labelled as *light*-flavour jet. τ-jets break this rule, because they are labelled with the name of a lepton: a τ-jet contains a τ lepton and no b- or c- hadrons. t-quarks, being the heaviest of the SM and due to the very high value of the V_{tb} term of Eq. 4.2, have an average lifetime of $\mathcal{O}(10^{-25}\,\text{s})$, which is about a twentieth of the timescale for strong interactions. Hence, t-quarks do not have the time to hadronize and are never found in jets.

Unfortunately, since particles in the colliders do not wander around shouting out-loud their own names, the labelling procedure described above can be applied only to particles originating from simulated collisions. When facing real data, sophisticated techniques are needed to identify the flavours of reconstructed jets: such techniques are collectively referred to as flavour tagging techniques. In Flavour Tagging, one uses the observables available in the collider to best guess the flavours of the observed jets. To quantify the performance of flavour-tagging algorithms, they are run on the simulated data, which actually contains labels for the simulated particles - the so called *truth information*. The mismatch of the jet-flavours as identified by the flavour tagging algorithm and as labelled by the *truth* labelling procedure described above can then be studied, defining the overall performance of the algorithm. In the following, I will give an overview of the objects reconstructed in the ATLAS detector which are used, directly or indirectly, by b-tagging algorithms.

4.2.1 Hits, Clusters and Space Points

Charged particles flying through the ID deposit energy in the pixel layers. This energy excites electrons in the valence band of the silicon semiconductors to the conduction band. The excited electrons can be collected by the application of electric potentials, producing an electric current that signals that an energy deposition occurred in the pixel. Individual pixels are read out if the amount of energy collected in that pixel excites a tunable threshold of 3500 electrons [28]. If that happens, a *hit* is said to have been measured in the pixel. The hit provides a 2D spatial information (η,ϕ) about the interaction of the charge with the pixel layer. The missing information about the third spatial dimension (r) is deduced by knowing the location of the pixel layers in the detector.

When a charged particle traverses a pixel sensor, charge is typically collected in more than one pixel. Neighbouring activated pixels are grouped in *clusters* through a procedure called *clustering*. Clusters are called *single-particle* clusters if they are produced by the charge deposit of one particle, while they are called *merged* if they are produced by multiple particles. Clusters are the basic blocks from which tracks are reconstructed. Clusters can be *split* if they are divided by a splitting algorithm into sub-clusters, each one used as an individual measurement for a different track; and they can be *shared* if they are not split by the splitting algorithm but are associated to more than one track candidate.

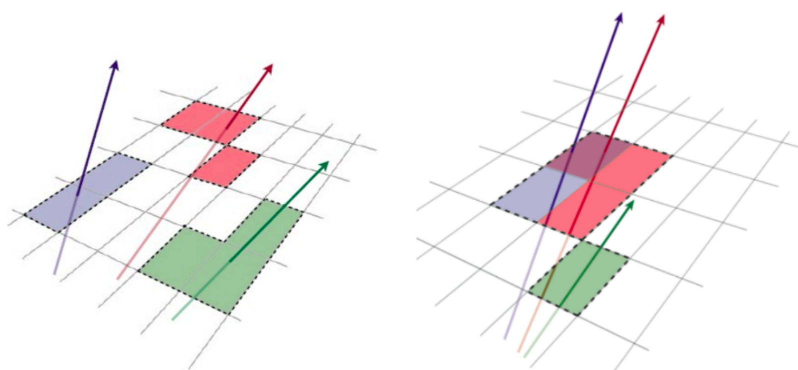


Figure 4.2: Single-particle clusters (left), produced by the charge deposit of one particle each; and split clusters (right), divided into sub-clusters, each one used as an individual measurement for a different track.

From each cluster, the *space point*, i.e. the point of intersection of the particle with the detector layer, is estimated through a charge interpolation technique based on the charge distributed in the different hits. Space points are then used to reconstruct the trajectories (tracks) of the charged particles. Tracks are then used as inputs for many b-tagging algorithms in ATLAS. Aim of this thesis is to also study the impact of the raw cluster information for its direct use in the b-tagging task; as it will be widely described in Chapter 5.

4.2.2 Tracks

Tracking is the process of reconstructing the trajectories of the electrically charged particles from the space points they left in the tracking layers. It can be divided into two subsequent steps, *track finding* and *track fitting*. Track finding is performed through a pattern recognition procedure that tries to group together space points originated from the same particle in subsequent detector layers. Through track fitting, i.e. the interpolation of the curve connecting the space points grouped together in the finding stage, the parameters of the track are reconstructed. Particles flying through a transverse uniform magnetic field follow an helicoidal trajectory that can be parametrized by five numbers. The same number of parameters is used for particles flying trough the ATLAS tracker system (Fig. 4.3), even if the magnetic field they are subjected to is inhomogeneous. These parameters are:

- d_0 : the transverse impact parameter, i.e. the distance of the point of closest approach of the curve to the z axis.

- z_0 : the longitudinal impact parameter, i.e. the distance of the point of closest approach defined for d_0 to the x-y plane containing the collision point.

- ϕ_0: the azimuthal angle at the point of closest approach.

- θ: the polar angle at the point of closest approach.

- q/p_T: the ratio between the charge and the transverse momentum of the track. This quantity, if the gradient of the magnetic field B is sufficiently smooth, is related to the curvature of the track through $p_T = qBR$, where R is the radius of the osculating circle of the curve.

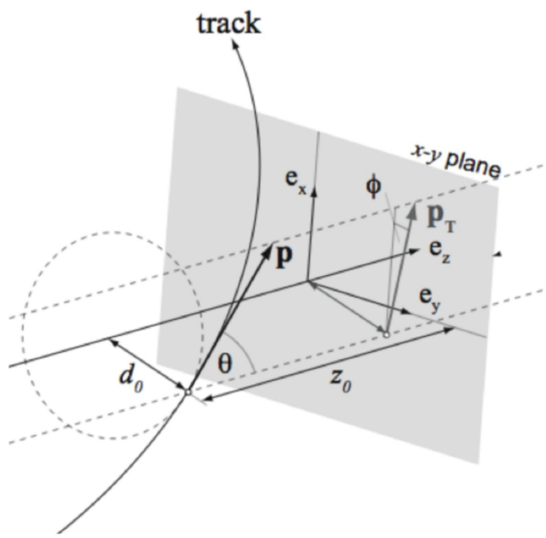


Figure 4.3: Parameterization of the reconstructed trajectory of a charged particle traversing the inner detector.

The accuracy of the track reconstruction is limited by several reasons, such as the resolution of the elements of the detector, the uncertainty on the location of the space points and on the point-per-point value of the magnetic field. For an exhaustive description of the track-related issues in ATLAS, see [77].

4.2.3 Primary Vertex

A well performing PV reconstruction is central for b-tagging since the PV defines the reference point from which track impact parameters and vertex displacements are computed (Fig 4.4). As for track reconstruction, the procedure for PV reconstruction is divided in a finding and a fitting stage [48]. The inputs of the algorithm are tracks passing a quality criterion selection. A pattern recognition process is performed on them to associate the tracks to vertex candidates. A fitting procedure follows to estimate the best vertex position based on the track parameters. All vertices with at least two associated tracks with p_T greater than 0.5 GeV are retained as valid PV. More than one PV is expected for every event because many protons of the colliding beans do interact at the same time for the typical instantaneous luminosities at the LHC. A longitudinal vertex position resolution of $\sim$30 μm is achieved for events with a high multiplicity of reconstructed tracks; and the transverse resolution is $\sim$10 μm. The PV associated to the group of tracks with the highest sum of squared transverse momenta is selected as the primary interaction point (also known as hard scatter vertex).

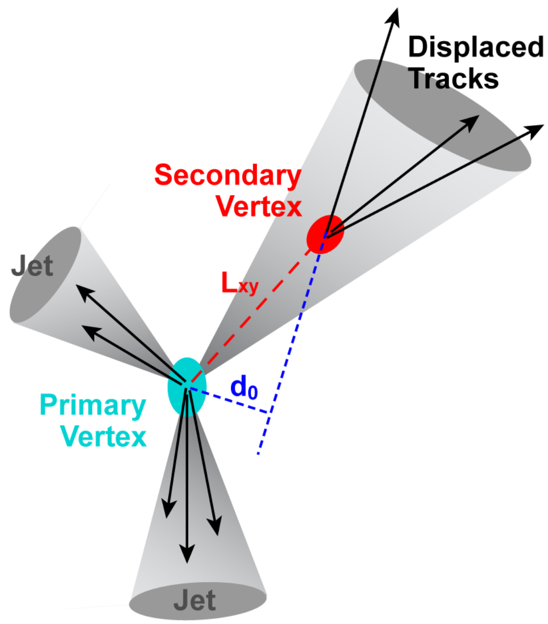


Figure 4.4: Schematic representation, not to scale, of hadronic jets originating from different quark flavours. In particular, two *light*-flavour and a b-jet are depicted, the former with associated tracks pointing back to the position of the primary interaction while the latter with a displaced decay originating a secondary vertex and associated tracks with high impact parameters.

4.3 Algorithms for b-jet identification

In this section I will present the different algorithms used for b-jet identification in ATLAS. They can be broadly divided into the low-level and high-level categories. Among the low-level algorithms, two approaches are followed: the first is based on the large impact parameters of the tracks coming from the decay of the b-hadron; while the second is about explicitly reconstructing displaced vertexes. Among the different algorithms falling into these categories, the IP2D and IP3D ones will be described in greater details. The reason is that, in this research, the IP3D tagging performance has been chosen as the benchmark to quantify the benefits of directly employing the information extracted from clusters, in addition to that extracted from tracks as it is done in the established flavour-tagging strategies at hadron colliders (Sec. 5.5). Low-level algorithms results are combined using multivariate classifiers in the high-level algorithms. Two high-level algorithms have been developed in ATLAS: MV2, which is based on a BDT; and DL1, a deep feed-forward neural network.

4.3.1 IP2D and IP3D

IP2D and IP3D algorithms exploit the long lifetime of b-hadrons, that results in a decay displaced from the PV. The impact parameter of the tracks is given by $\beta\gamma c\tau \sin\alpha$, where α is the decay product's angle of emission with respect to the decayed particle's direction. Hence, a long lifetime τ result in a large impact parameter. The IP algorithms actually use the impact parameters *significances*, that are defined as the ratio of the impact parameter and its uncertainty. Because of their displacement, tracks generated from b-hadron decays are expected to have impact parameter significances that differ significantly from zero, whereas those from tracks in *light*-flavour jets are typically consistent with zero within the resolution of the experimental measurement [40](Fig. 4.5). Not all the tracks assigned to a jet[1] are used by the IP algorithms. A selection is applied based on the following requirements:

- $p_T > 1$ GeV;
- $|d_0| < 1$ mm and $|z_0 \sin\theta| < 1.5$ mm;
- Seven or more silicon hits with at most two silicon holes, at most one of which is in the pixel detector[2].

The selected tracks are then divided into categories (*track grades*) based on criteria of increasing quality that depend on the hit pattern of the track [41]. As it is clear from Fig. 4.5, impact parameters are assigned a sign[3], positive if the position of the crossing of the track with the jet axis is located upstream to the PV along the jet axis, and negative otherwise.

The IP2D tagger uses the transverse impact parameter significance d_0/σ_{d_0} as a discriminating variable, whereas IP3D uses the longitudinal impact parameter significance $z_0 \sin\theta/\sigma_{z_0 \sin\theta}$ as well. Both the algorithms are based on a Log Likelihood Ratio (LLR) discriminant. Based on its impact parameter, each track is assigned a probability of belonging to a b-jet (p_b) or a *light*-flavour jet (p_u). p_b and p_u are extracted from reference histograms of d_0/σ_{d_0} and z_0/σ_{z_0} derived from MC simulations. Different histograms are made for the different track grades: Fig. 4.5 shows those for the highest-quality tracks.

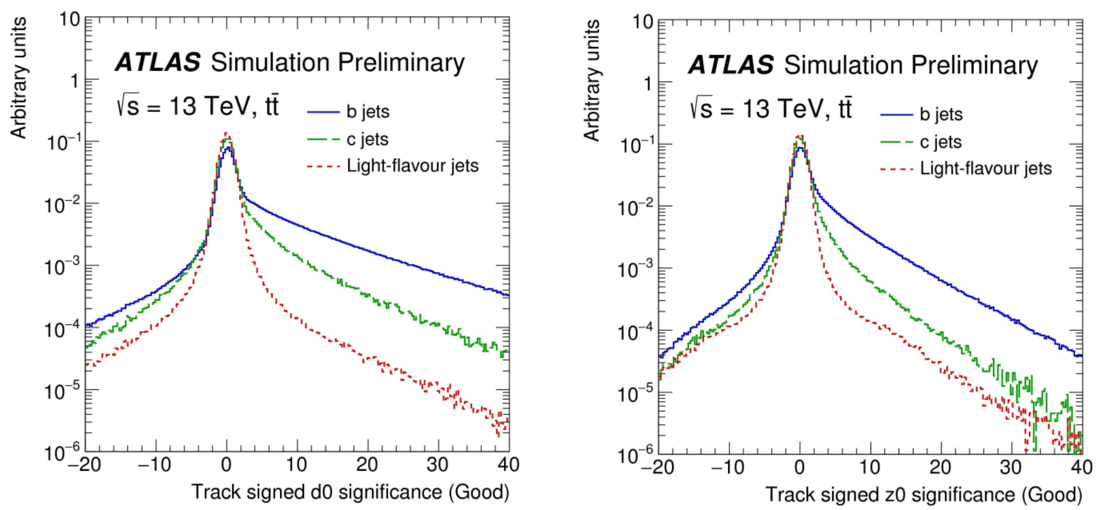


Figure 4.5: Transverse (left) and longitudinal (right) signed impact parameter significance of *Good* tracks in a Monte Carlo simulation of $t\bar{t}$ events. The histograms are used to extract the likelihood of each track of belonging to a b- or *light*-flavour jet. E.g., a track with a transverse impact parameter of 20 is more likely to belong to a b- than to a *light*-flavour jet.

The LLR discriminant of a jet is then computed as the sum of the per-track log-likelihood ratio

$$\text{LLR} = \sum_{i=1}^{N} \log\left(\frac{p_b}{P_u}\right), \tag{4.3}$$

[1]Tracks are associated to the nearest jet by exploiting the angular separation ΔR between the track and the jet axis. The ΔR requirement varies as a function of jet p_T, being wider for low p_T values (0.45 for jet p_T of 20 GeV) and narrower for high values (0.26 for jet p_T of 150 GeV) since decay products from energetic heavy-flavoured hadrons are more collimated.

[2]Silicon hits are hits in the Pixel Layers and SCT. A silicon hole is a hit expected to be associated with the track but not present.

[3]Being distances, impact parameters are always positive unless a sign convention is defined.

where N is the number of tracks associated to a given jet. No correlation among the tracks contributing to the sum is assumed. The LLR distributions of the IP2D and IP3D are shown in Fig. 4.6. The big difference between the distributions for *light*-flavour jets and *b*-jets is what accounts for the power of discrimination of the taggers.

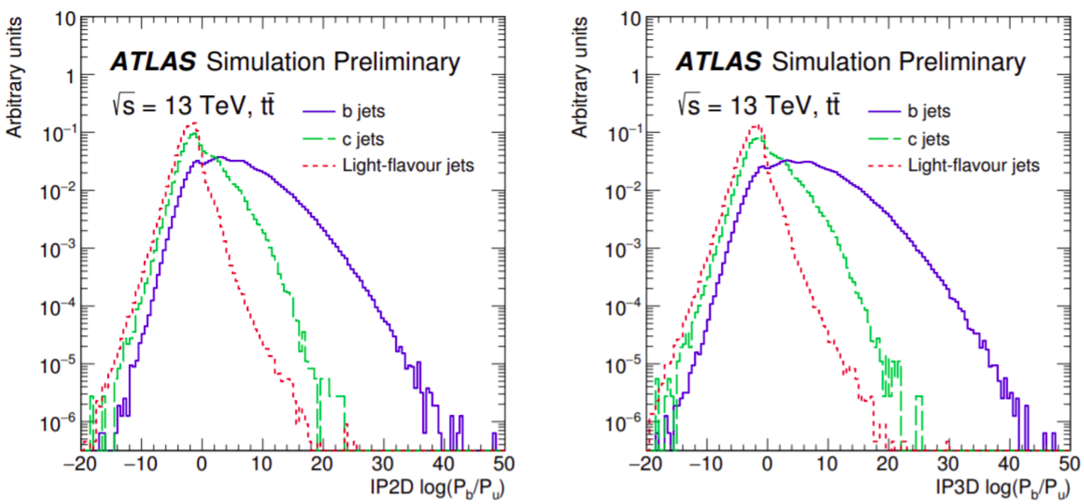


Figure 4.6: LLR of the IP2D (left) and IP3D (right) algorithms. The differences between the blue and red lines in the plots account for the discrimination power of the algorithms.

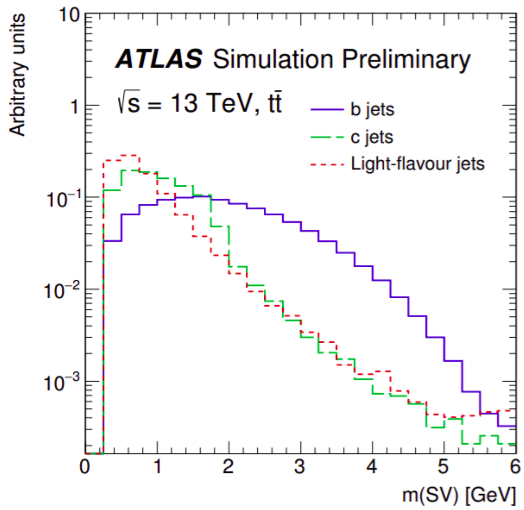
4.3.2 Recurrent and Deep Sets Neural Network Track-based taggers

The impact parameters of the tracks associated to a jet are intrinsically correlated. If one track with a large impact parameter is found, then maybe a displaced decay occurred: thus, it is likely than other tracks with a big impact parameter will be found in the jet. The IP2D and IP3D algorithms do not exploit the correlations between tracks because filling histograms such as those shown in Fig. 4.5 while taking into account the correlation between different tracks and between tracks and other variables would require a sample of oversized dimension; a fact known as *the curse of dimensionality* [36]. Neural networks such as IPRNN [37] and DIPS [36] allow to take advantage of the correlations between the tracks without needing to deal with oversized samples. IPRNN and DIPS differ in the way the set of tracks associated to each jet is treated, the former requiring an ordering of the tracks by d_0/σ_{d_0} and the second treating the elements as a set without a specific ordering. They both outperform the IP3D and IP2D algorithm, DIPS being the best among all the track based algorithms. Further information about the architecture of IPRNN, DIPS, and the comparison of their performance can be found in [36] and [37].

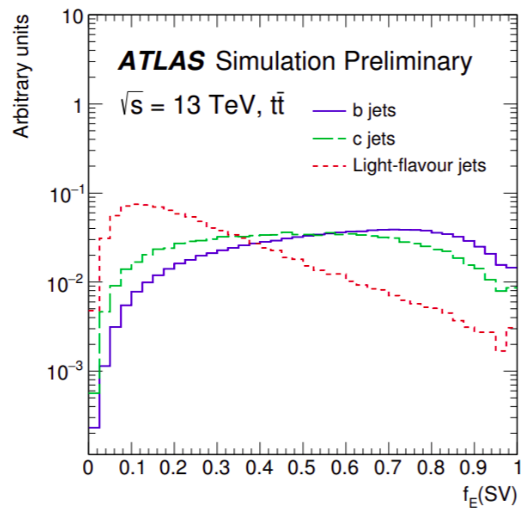
4.3.3 Secondary vertex finding algorithm

The secondary vertex finding algorithm (SV) [42] reconstructs a single displaced secondary vertex in a jet following a procedure analogous to the one described in Sec. 4.2.3, but limited to the tracks associated to the jet. Further attention is payed to the removal from the vertex candidates of those compatible with known objects, such as long-lived particles, photon conversions or interactions with the detector material.

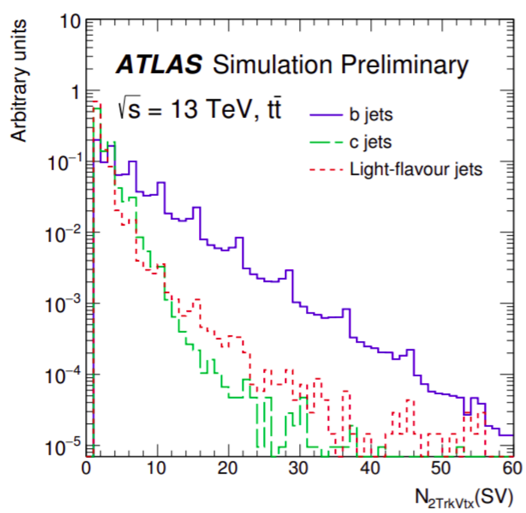
Because of the long lifetime of the *b*-hadrons, secondary vertexes are expected to be more displaced in *b*-jets than in *light*-flavour jets (Fig. 4.4); allowing to use some features of the SV for the *b*-tagging task. The *b*-tagging standalone performance of the SV algorithm is evaluated using a LLR discriminant computed from histograms built from three of the SV output variables: the vertex mass, reconstructed from the energies of the tracks associated to the secondary vertex; its energy fraction, defined as the total energy of the tracks associated to the secondary vertex over the total energy of the tracks in the jet; and the number of two tracks secondary vertex candidates in the jet (Fig. 4.7 and 4.12). More SV output variables are used by the high level taggers. They are listed in Fig. 4.10.



(a)



(b)



(c)

Figure 4.7: Output variables of the SV algorithm used for the evaluation of its standalone performance: invariant mass associated to the secondary vertex (a), fraction of energy (b), number of 2-tracks secondary vertex candidates (c).

4.3.4 Topological multi-vertex finding algorithm

The topological multi-vertex algorithm JetFitter [45] tries to reconstruct the full b-hadron decay chain exploiting the topological structure of the weak b- and c-hadron decays. It adopts a different strategy from the SV algorithm (which in the finding stage inclusively searches for all the secondary vertexes candidates): JetFitter admits the simultaneous reconstruction of several vertices along a line that approximates the b- and c-hadrons flight directions (Fig 4.8).

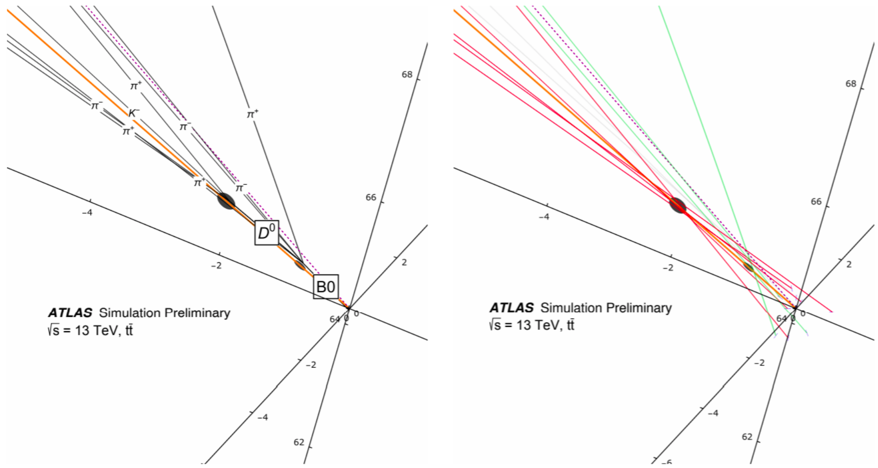


Figure 4.8: Event display of a simulated B^0 decay showing generated (left) and reconstructed (right) tracks. JetFitter performs a multi-vertex fit constraining all vertices to lie on the flight direction of the B^0 meson (orange line in both panels).

During a vertex finder stage, JetFitter tries to find the correct number of vertices in the jet and the tracks associated to each vertex; during the vertex fitter stage the positions of the vertexes are reconstructed. The main underlying topological assumption of the JetFitter algorithm is to consider the b- and c-hadron decay vertexes aligned on a line defined by the b-hadron flight path. The assumption comes with some advantages (see [45] for an exhaustive list), the main one being the possibility of reconstructing vertexes with only one associated track.

Eight discriminating variables are then extracted from the reconstructed topological structure of the jet. They are listed in Fig. 4.10. The b-tagging standalone performance of the JetFitter algorithm is evaluated using a LLR discriminant computed from histograms built from three of the JetFitter output variables: vertexes mass, energy fraction and decay length significance (Fig. 4.9 and 4.12).

4.3.5 MV2

MV2 [40] is an high-level tagging algorithm consisting in a BDT that combines the outputs of the low-level tagging algorithm in a single multivariate discriminant. BDTs are algorithms that combine a group of classifiers called Decision Trees in a single classifier that estimates the probability of an input to belong to one of the categories that are being classified (see Sec. 5.4 for my explanation of the functioning principles of BDTs). Fig. 4.10 lists all the variables used by the MV2 algorithm[4]. Fig. 4.11a shows the distribution of the MV2 output discriminant variable for the different flavoured jets. Fig. 4.12 shows its performance in the tagging task: MV2 outperforms all the low-level tagging algorithms, but is outperformed by DL1, the latest high-level algorithm in ATLAS.

4.3.6 DL1

DL1 is the second high-level b-tagging algorithm [40]. It is based on a deep feed-forward neural network that combines all the discriminating variables listed in Fig. 4.10. Neural networks (NNs) are machine learning algorithms built connecting together various building blocks called *neurons*. A neuron is an object that takes some weighted inputs and combines them through an *activation function* into a single

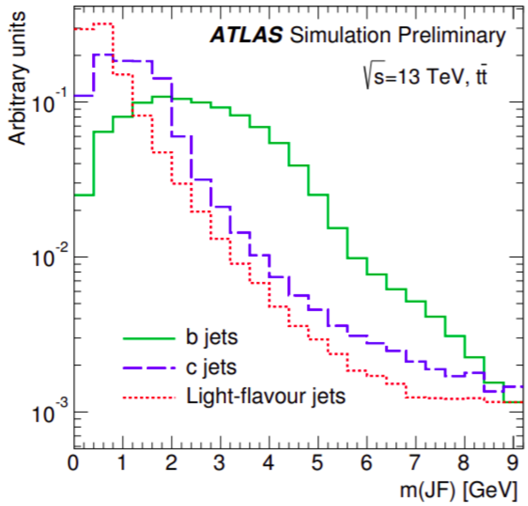
[4]The JetFitter c-tagging variables at the bottom of the Table are used only by DL1.

weighted output. Feed-forward neural networks combine together different *layers* of nodes: the outputs of the nodes in the n_{th} layer are passed as inputs to the $n_{th}+1$ layer, until the output layer is reached. In DL1, the output layer has three nodes corresponding to the estimated probabilities p_b,p_c and p_u for the jet of being respectively a b-, c- or *light*-flavour jet. The final DL1 b-tagging discriminant is defined as

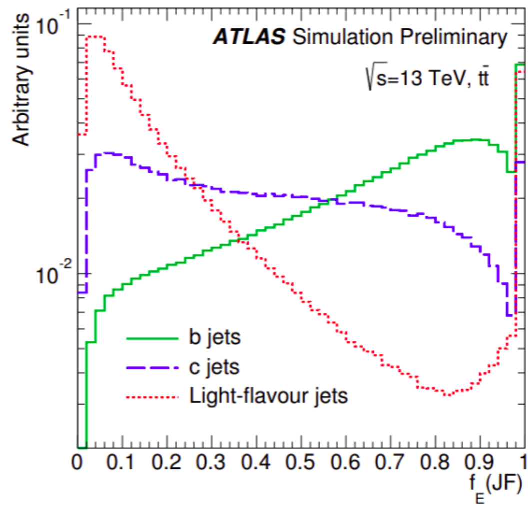
$$D_{\mathrm{DL1}} = \ln\left(\frac{p_b}{f_c p_c + (1-f_c)p_{\mathrm{light}}}\right), \tag{4.4}$$

where f_c represents the c-jet fraction in the background training sample. Each node in the network is assigned a weight, whose value is tuned during a training phase through a regression process that aims at minimizing a *loss function* that quantifies the performance of the network in the classification task. See [84] for an introduction to NNs.

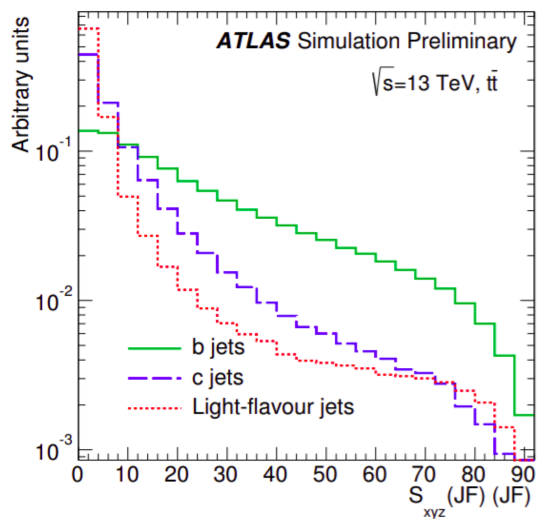
The output discriminanants of the DL1 algorithm are shown in Fig.. 4.11b. Fig. 4.12 shows that the DL1 performance in discriminating *light*- from b-jets are the best in ATLAS.



(a)



(b)

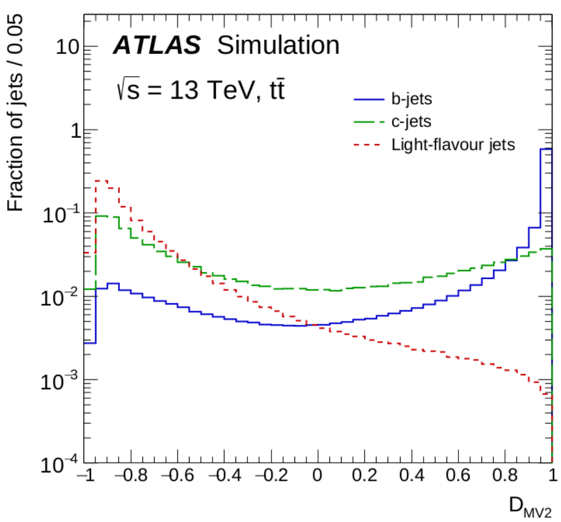


(c)

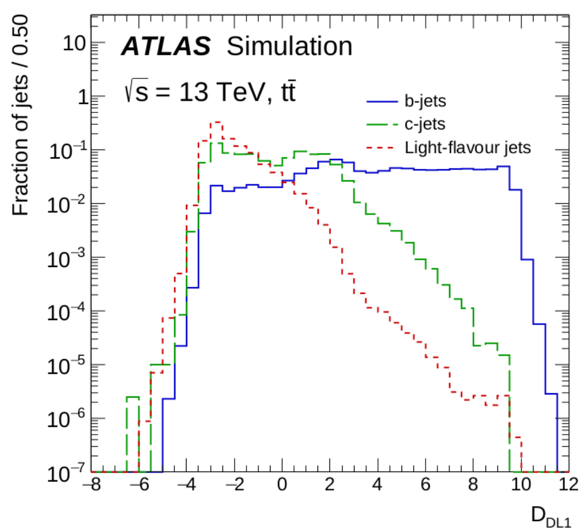
Figure 4.9: Output variables of the JetFitter algorithm used for the evaluation of its standalone performance: invariant mass associated to the vertexes (a), fraction of energy (b), decay length significance (c).

Input	Variable	Description
Kinematics	p_{T}	Jet p_{T}
	η	Jet $\|\eta\|$
IP2D/IP3D	$\log(P_b/P_{\mathrm{light}})$	Likelihood ratio between the b-jet and light-flavour jet hypotheses
	$\log(P_b/P_c)$	Likelihood ratio between the b- and c-jet hypotheses
	$\log(P_c/P_{\mathrm{light}})$	Likelihood ratio between the c-jet and light-flavour jet hypotheses
SV1	$m(\mathrm{SV})$	Invariant mass of tracks at the secondary vertex assuming pion mass
	$f_E(\mathrm{SV})$	Energy fraction of the tracks associated with the secondary vertex
	$N_{\mathrm{TrkAtVtx}}(\mathrm{SV})$	Number of tracks used in the secondary vertex
	$N_{\mathrm{2TrkVtx}}(\mathrm{SV})$	Number of two-track vertex candidates
	$L_{xy}(\mathrm{SV})$	Transverse distance between the primary and secondary vertex
	$L_{xyz}(\mathrm{SV})$	Distance between the primary and the secondary vertex
	$S_{xyz}(\mathrm{SV})$	Distance between the primary and the secondary vertex divided by its uncertainty
	$\Delta R(\vec{p}_{\mathrm{jet}}, \vec{p}_{\mathrm{vtx}})(\mathrm{SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.
JetFitter	$m(\mathrm{JF})$	Invariant mass of tracks from displaced vertices
	$f_E(\mathrm{JF})$	Energy fraction of the tracks associated with the displaced vertices
	$\Delta R(\vec{p}_{\mathrm{jet}}, \vec{p}_{\mathrm{vtx}})(\mathrm{JF})$	ΔR between the jet axis and the vectorial sum of momenta of all tracks attached to displaced vertices
	$S_{xyz}(\mathrm{JF})$	Significance of the average distance between PV and displaced vertices
	$N_{\mathrm{TrkAtVtx}}(\mathrm{JF})$	Number of tracks from multi-prong displaced vertices
	$N_{\mathrm{2TrkVtx}}(\mathrm{JF})$	Number of two-track vertex candidates (prior to decay chain fit)
	$N_{\mathrm{1\text{-}trk\ vertices}}(\mathrm{JF})$	Number of single-prong displaced vertices
	$N_{\geq 2\text{-}\mathrm{trk\ vertices}}(\mathrm{JF})$	Number of multi-prong displaced vertices
JetFitter c-tagging	$L_{xyz}(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Distance of 2^{nd} or 3^{rd} vertex from PV
	$L_{xy}(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Transverse displacement of the 2^{nd} or 3^{rd} vertex
	$m_{\mathrm{Trk}}(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Invariant mass of tracks associated with 2^{nd} or 3^{rd} vertex
	$E_{\mathrm{Trk}}(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Energy fraction of the tracks associated with 2^{nd} or 3^{rd} vertex
	$f_E(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Fraction of charged jet energy in 2^{nd} or 3^{rd} vertex
	$N_{\mathrm{TrkAtVtx}}(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Number of tracks associated with 2^{nd} or 3^{rd} vertex
	$Y_{\mathrm{trk}}^{\mathrm{min}}, Y_{\mathrm{trk}}^{\mathrm{max}}, Y_{\mathrm{trk}}^{\mathrm{avg}}(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Min., max. and avg. track rapidity of tracks at 2^{nd} or 3^{rd} vertex

Figure 4.10: List of the discriminating variables used by the high-level algorithms MV2 (all but the JetFitter c-tagging variables) and DL1 (all the variables).



(a)



(b)

Figure 4.11: Multivariate discriminating variables of the MV2 (a) and DL1 (b) algorithms.

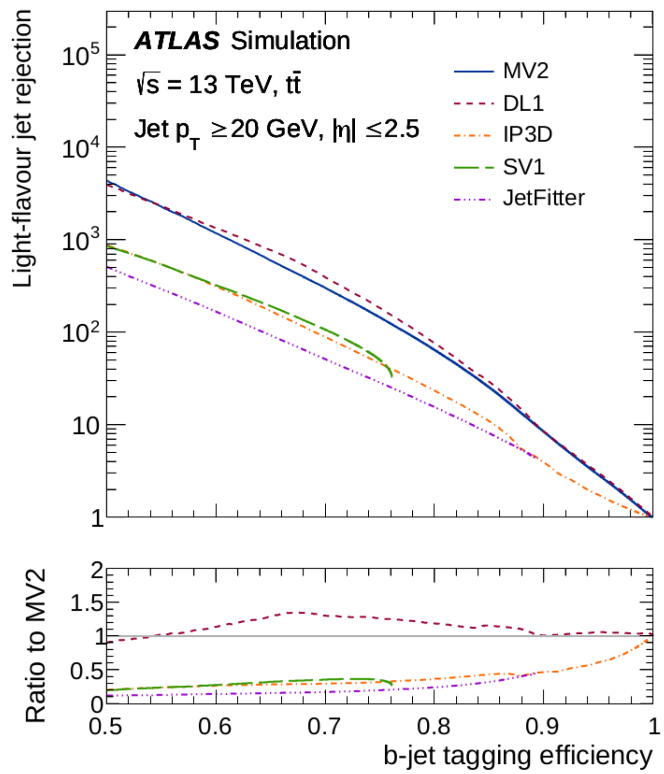


Figure 4.12: ROC curves of the performance of the described b-tagging algorithms compared. ROC curves plot the rejection (i.e. the inverse of the probability of wrongly classifying a jet) of a given b-tagging algorithm as a function of its efficiency (i.e. the probability of correctly classifying a b-jet). Thus, the higher is the area under the curve, the better is the performance of the algorithm. The ratio of the DL1 ROC curve over the MV2's shows how the former outperforms the second in the 0.6-0.9 efficiency range.

Chapter 5

Trackless *b*-tagging

In this Chapter, I will present my work to study the impact of the cluster information in the tracking detectors for tagging high p_T jets. All the plots that follow, unless the one explicitly referenced, have been obtained during my work described in this thesis using a Monte Carlo simulation containing a heavy Z' boson [73] decaying democratically into $b\bar{b}$, $c\bar{c}$ and $l\bar{l}$ pairs, where l is a *light-flavour* quark (u,d and s).

In general, the b-tagging algorithms based on tracking and secondary vertexing information described in Chapter 4 decrease their performance when applied to high-p_T jets. This is due to a deterioration of the track information, to the increased fraction of fragmentation tracks in the jet, and to the long flight lenght of impulsive b-hadrons, that can cause them to decay more externally than one or more of the innermost layers of the ID, their decay products thus leaving less hits in the detector. These issues are the focus of Sec. 5.2. Discriminant information based on clusters is not employed in any of the state-of-the-art b-tagging algorithms in ATLAS, but it can help improve their discrimination power for jets with sufficiently high transverse momentum. In Sec. 5.3, I will explain why, referring to previous studies on the subject.

In this study, I have designed and trained a BDT based on the output of the IP3D algorithm plus simple cluster-based variables. BDTs have been widely adopted for flavour-tagging classification in ATLAS by the MV-based family of algorithms. In Sec. 5.4, I will explain how they work.

The training and application of the BDT to the sample is described in Sec. 5.5; where I present the results and prove that the algorithm can learn the distinctive features of late-decaying b-hadrons described in Sec. 5.3.

5.1 The Sample

The simulation used to conduct this research employs the decays of Z' bosons with a mass spectrum ranging from 0 to 12 TeV (Fig. 5.1) to generate high energy jets. Z' is an hypothetical boson that would be included in the weak sector by a modified $U'(1)$ local gauge symmetry. As bizarre as it may seem to use an hypothetical particle to study jet reconstruction and tagging, this is a standard choice for jet flavour studies at the TeV scale. The reason is that the jets resulting from the Z' decay have a flat-enough p_T spectrum up to momenta of about 6 TeV (Fig. 5.2). This allows to access a sufficient amount of jets in the high-p_T region (i.e. $p_T > 100$ GeV) without generating an unreasonable amount of them in the low-p_T region (as it would be necessary if jets were generated by decaying $t\bar{t}$, the standard for jet flavour studies below the TeV scale).

Z's in this sample decay to hadronic jet pairs with a modified branching ratio that sets the probabilities of $Z' \to b\bar{b}, c\bar{c}$ and $l\bar{l}$ all to one third.

The interaction of particles with the detector and the consequent showering are simulated with GEANT4 [56], the standard toolkit in ATLAS; whereas PYTHIA8 [80] is used o simulate pp collisions with center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and it is interfaced to MADGRAPH5 [6] to simulate parton showers and hadronisation. EVTGEN [74] is used to simulate the decays of the b- and c-hadrons. Jets are reconstructed from clusters in the calorimeters using the anti-k_T algorithm with radius parameter R=0.4. More information about the simulation can be found in [40].

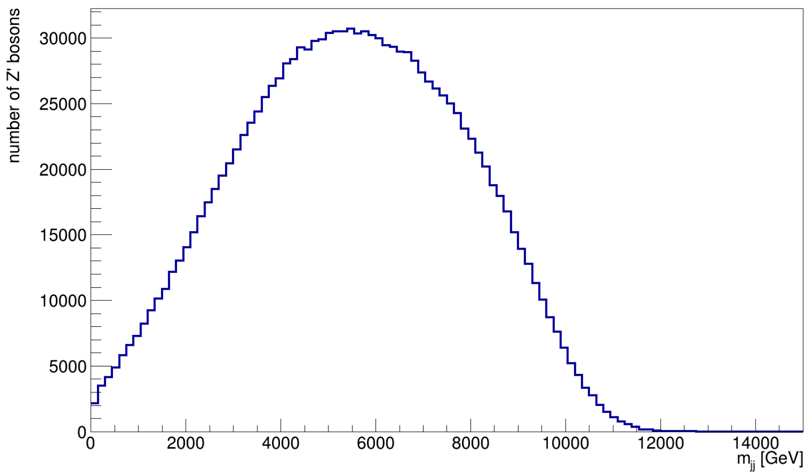


Figure 5.1: Mass spectrum of the resonance of the highest (leading) and second-to-highest (subleading) jets when ordered in descending order according to their p_T values. Since the most impulsive jets in the sample come from the decay of the Z' boson, the mass distribution represents, modulo detector-induced effects, the mass of the simulated Z' boson.

Some fiducial cuts were made to select the jets used in this research. For each collision event, only the jet with the highest pT (leading) and the second-to-highest pT (subleading) were considered. In Sec. 5.2 only the jets with p_T between 0 and 650 GeV are shown in order to best illustrate the deterioration of the track information with the increase of p_T. In all the other sections, the admitted p_T was chosen to be between 100 GeV and 2.5 TeV, the range that contains the region that benefits from the addition of the cluster information to the BDT of Sec. 5.5. To be considered in this analysis, jets must have a value $|\eta| < 1.0$, meaning that the analysis is performed in the barrel, thus avoiding the geometrical complications that would arise from considering the endcaps as well[1]. To reduce the number of jets with large energy fractions from *pile-up* collision vertices, the Jet Vertex Tagger [81] algorithm is used. Pile-up collisions are all the multiple collisions between protons of the same two colliding beams, that happen from the same bunch crossing. They represent a strong experimental challenge at the LHC, since the decay products from pile-up collisions are an important source of background in the study of the primary collision. In our simulation, they are generated by PYTHIA8. The average number of pile-up events per bunch crossing in the simulation is 24.4. The JVT procedure builds a multivariate discriminant going form 0 to 1 for each jet which is likely to get a small value for those with a large fraction of high-pT tracks from pile-up. Thus, jets with $p_T < 120$ GeV and JVT < 0.59 are rejected. The rate of pile-up jets with $p_T > 120$ GeV is known to be small enough to remove the JVT requirement above this threshold. For the sake of simplicity, only b-jets containing one and just one b-hadron were considered. Finally, only b-jets in which the b-hadron decays *inside* the region delimited by the last pixel layer are admitted. For the others, the variables described in Sec. 5.3 could not be applied without considering also the SCT-layers.

[1] Such complications would consist in less uniformity in the distribution of jets and number of cluster per layer due to the geometrical shape of the detector. To restrict our research to the barrel does not imply a loss of generality: it could be extended to the excluded $|\eta|$ region by means of a more careful treating of the η information during the training.

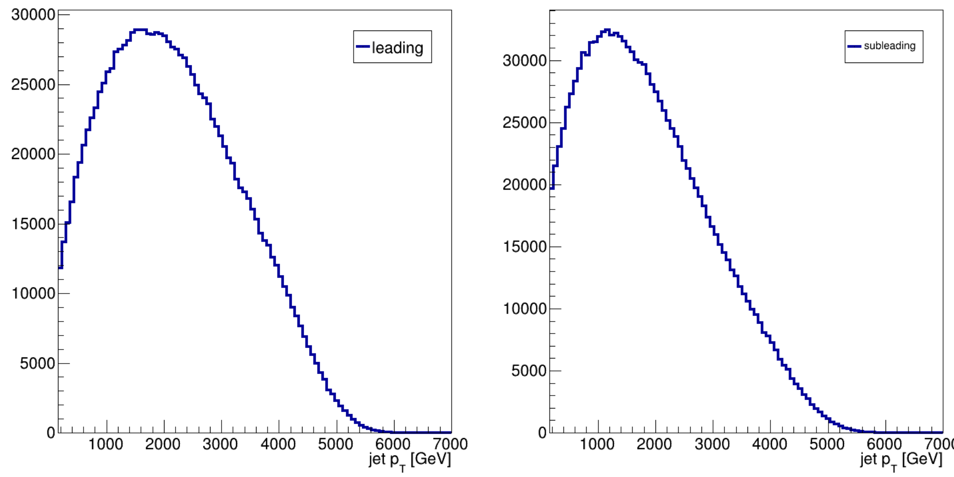


Figure 5.2: p_T spectrum of the highest (leading) and the second-to-highest (subleading) jets when ordered in descending order according to their p_T values.

The described cuts applied to the available statistics end up in a sample of jets composed by ~ 230000 b-jets, 340000 c-jets and 780000 *light*-flavour jets. Of these, only the b- and *light*-flavour jets were used for the analysis of Sec. 5.5.

5.2 Decrease of the b-tagging performance at high-p_T

All the track based algorithms described in Chapter 4 decrease their performance at high-p_T. For example, in a very recent analysis [50] which used for the first time at ATLAS the state-of-the-art b-tagging algorithm (DL1r, a version of DL1 that uses the RNNIP discriminant variables among its inputs), the efficiency of the adopted selection was shown to drop from 65% for a b-jet p_T of around 500 GeV to 10% for a p_T of around 2 TeV (Fig. 5.3), to be compared with an approximately constant rejection of *light*-flavour jets. .

This behaviour can be understood by considering that the decay products of high-p_T b-hadrons are very collimated (Fig. 5.4), due to the Lorentz transformation of angles

$$\tan(\theta) = \frac{|\vec{p^*}|\sin(\theta^*)}{\gamma(|\vec{p^*}|\cos(\theta^*) + \beta\epsilon^*)}, \tag{5.1}$$

where θ and θ^* are the angles between the decay product's initial trajectory and the b-hadron's, respectively in the laboratory and b-hadron's reference frame; $\vec{p^*}$ and ϵ^* are the decay product's 3-impulse and energy in the b-hadron's reference frame, and γ and β are the b-hadron's relativistic and Lorentz factor as measured from the laboratory. To gain a more qualitative feeling of the effect of Eq. 5.1, consider that, in the case in which the decay product is emitted at $\theta^* = \Pi/2$ at high energy (i.e. $|\vec{p^*}| \sim \epsilon^*$), it becomes $\sin\theta \sim 1/\gamma$.

In addition, high-pT jets are also accompanied by a larger number of tracks from fragmentation, as can be seen in Fig. 5.5. In such a dense environment, the merging of hits produced in the tracking

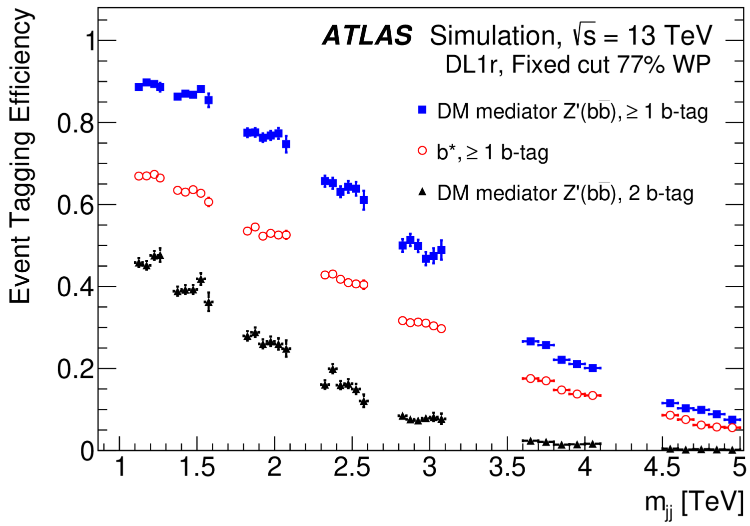


Figure 5.3: A drop in the efficiency in the reconstruction of mass resonances decaying in $b\bar{b}$ pairs and in jet pairs containing at least a b) is observed for energies > 500 GeV (b). The plot is taken from [50]: see Sec. 6.2 for an extended explanation of the analysis.

detectors by different charged particles leads to a reduced efficiency for track and vertex reconstruction algorithms.

Furthermore, boosted b-hadrons often decay beyond the innermost layers of the tracking detector (Fig. 5.6.), leading to a significant loss of reconstructed tracks (Fig. 5.7).

5.3 The Idea

The many issues encountered while trying to employ the track information for tagging jets with high p_T suggest to study ideas that do not rely on track reconstruction to improve the performance in the high energy region. The one I am going to illustrate is very intuitive; and relies on the characteristic long lifetime of b-hadrons. As already mentioned in Eq. 4.1, when the b-hadron is impulsive enough, it can decay after traversing one or more layers of the detector, producing a *jump* in the number of clusters between the layers before and after the decay (Fig. 5.8).

The use of the jump as a discriminating variable to identify b-jets has been studied and proven worthy in [69] and [70]. In my research, I use the information on the number of clusters per layer in the Pixel Detectors of the ID to train a machine learning algorithm. To optimise the information content, not all the clusters are considered: as it is clear from Fig. 5.9, the vast majority of clusters coming from the decay of the b-hadron falls inside a cone of $dR = 0.04$ around the jet-axis. Since the jump is expected to be coming from the clusters generated by b-hadron decay products, to consider clusters in a wider cone would would only dilute the available information in useless noise.

To check whether restricting even further the selection of the "core" clusters closest to the jet axis is beneficial to the performance of the BDT, I consider the four $dR = 0.01, 0.02, 0.03$ and 0.04 cases. In addition, the $dR = 0.4$ case is studied. Fig. 5.10 shows the distributions of the number of clusters for each pixel layers for the background *light*-flavour jets and for b-jets in the 0.04 cone, the one with which we observe the best performance of the BDT in the tagging task.

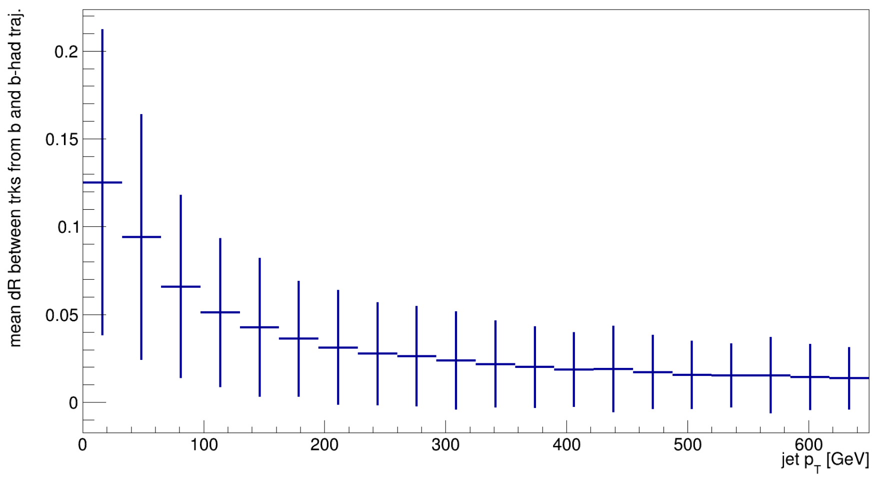


Figure 5.4: The tracks originating from the b-hadron are increasingly collimated when the jet p_T rises. The vertical error lines, here and everywhere, show the Root Mean Square of the distribution.

5.4 Boosted Decision Trees

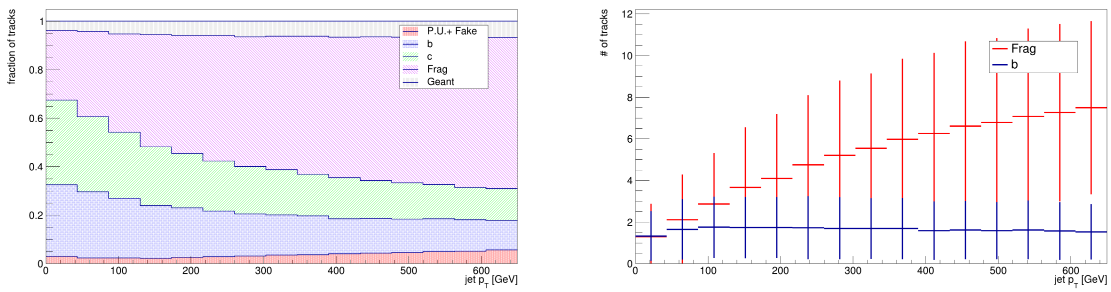
BDTs are machine learning algorithms that use the decisions taken by a group of *decision trees* to classify input objects into distinct categories [66]. Decision trees are predictive models that use a set of discriminating variables about an object to predict the value of another target variable whose value is unknown. Decision trees in which the target variable can take a discrete set of values are called *classification* trees; while they are called *regression* trees if the target variable can take continuous values. An algorithm that classifies the inputs based on the majority vote of a group of decision trees is called a *random forest*. Boosting algorithms add the option to define procedures to give different weights to the vote of individual trees.

BDTs have a wide range of applicability that goes from online advertising to document retrieval to spam recognition. In the present thesis, a BDT is used to discriminate b-jets from *light*-flavour jets based on a set of discriminating variables that includes the information on the number of clusters per layer.

In this section I will describe how to build a single decision tree and a random forest. Then, I will illustrate *AdaBoost*, the boosting algorithm used to build the classifier developed in this thesis.

5.4.1 Decision Trees

A decision tree is an algorithm that asks questions about the values of the variables characterising an input object, and classifies the input based on the answers. The path followed by the input through the tree is determined by the answers given at each step. Consider the tree in Fig. 5.11, taken as an example by the set of trees built by the BDT I will describe in Sec. 5.5. The coloured rectangles are called *nodes*, the one at the top being the *root node* and the ones not connected to lower nodes being the *leaf nodes*. The path connecting the root to a leaf is called *branch*, and the *depth* of the tree is the number of nodes of the longest branch, root node excluded. At each node, a test is made upon the value of a specific variable. If the test answer is found to be positive, the input proceeds down the branch to the left, while it goes to the right if the answer is negative. The phase space is thus split in many regions that are classified as signal or background, depending on the majority of the *true* values of the target variables (which are known in a training sample) that end up in the final leaf node. In our example, a sample of jets - for which the target variable is the value of a label stating whether the



(a) Fraction of tracks.

(b) Number of tracks.

Figure 5.5: With the increase of the jet p_T, the number of tracks from fragmentation dominates the number coming from the b-hadron. (a) shows the *truth* fraction of tracks coming from different sources in the jets as a function of the jet p_T: Pile-up tracks, fake tracks (i.e. tracks erroneously reconstructed by the tracking algorithms from group of clusters that were not originated by the same charged particle), tracks from the fragmentation process, tracks from the b and c-hadrons decays and tracks generated by the products of the interactions with the detector. (b) shows the mean number of tracks from the b-hadron and from fragmentation as a function of the jet p_T.

jet is a b- or a *light*-flavour jet - is used for the training, b being considered the signal and *light* the background.

The first step in growing the tree is the definition of the root node. To perform this task, for each variable a scan is made on all the possible cut values[2] to find which one best discriminates signal from background. In our example the discrimination power is quantified using the Gini Index:

$$\text{Gini Index} = \frac{SB}{(S+B)^2}, \tag{5.2}$$

where S and B are the number of signal and background jets that would end up in the nodes below the root if the cut in exam were chosen. The Gini Index for the cut is the average of the Gini Index of the left and right node, weighted with their number of jets. The Gini Index is a *purity* quantifier, which has a maximum when $S = B$ and has its minimum value 0 if either S or B is zero: thus, the better is the capability of the cut to identify a region of signal or background in the sample phase space, the lower is its Gini Index value (Fig. 5.12).

The value of the Gini Index is computed for each possible cut on each discriminating variable, and the one that gets the lowest value is chosen as the root node. This process is iterated for all the nodes until the maximum depth chosen for the decision tree is reached, until a node is found for which every further possible cut would end up decreasing its Gini Index, or until the number of jets in a node reaches a given minimum size. A ranking of the discriminating variables is derived by counting how often the variables are used to split the nodes, and by weighting each split occurence by the square of its Gini Index and by the number of jets in the node [68].

Decision trees are simple classifiers easy to interpret, but they lack in accuracy. To cite from "The Elements of Statistical Learning" [66]: *"Trees have one aspect that prevents them from being the ideal tool for predictive learning, namely inaccuracy"*. When the tree of Fig. 5.11 classifies an input jet as b or *light*, four different scenarios are possible for the outcome (Fig. 5.13):

- **True Positive (TP)**: the jet is a b-jet and it is classified as such.
- **True Negative (TN)**: the jet is a *light*-flavour jet and it is classified as such.

[2]The granularity of the cut is often defined by the user to reduce the computing time. The optimal cut for a given binning of the input variable distribution is surely found only if *every* possible cut is checked.

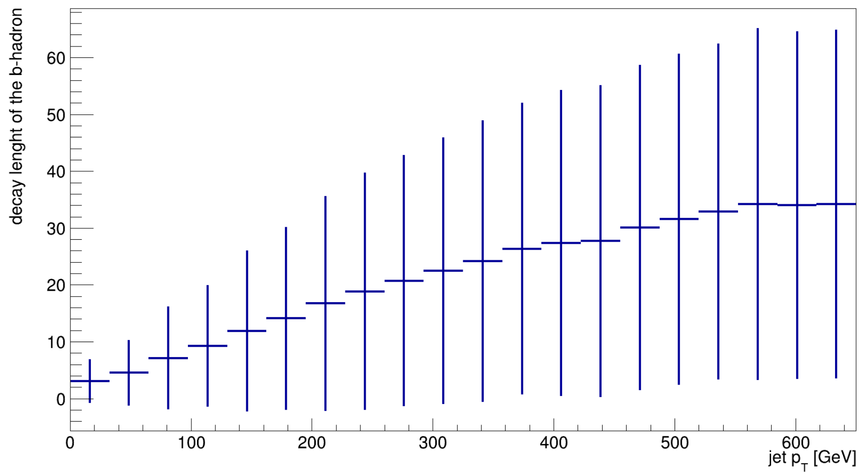
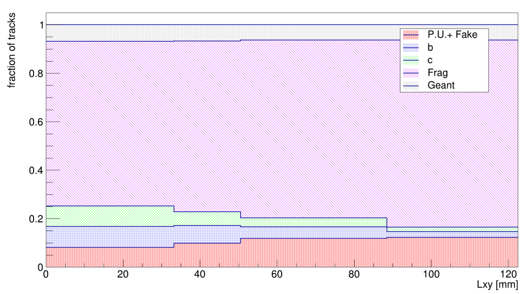
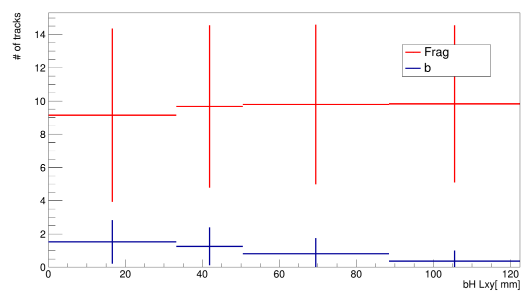


Figure 5.6: When the b-hadron becomes increasingly impulsive, its flight length becomes sufficient to traverse the innermost layers of the ID.



(a) Fraction of tracks



(b) Number of tracks

Figure 5.7: With the increase of flight length of the b-hadron, the number of tracks from fragmentation remains constant, while the number of tracks coming from the b-hadron drops.

- **False Positive (FP)**: the jet is a *light*-flavour jet and it is classified as b-jet.
- **False Negative (FN)**: the jet is a b-jet and it is classified as a *light*-flavour jet.

The accuracy of a classifier is defined as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}. \tag{5.3}$$

The highest is the accuracy, the better is the classifier in labelling the input. Decision trees seldom provide predictive accuracy comparable to the one estimated in the training process. In other words, they are fine tuned to describe the data of the training sample, but lack in flexibility when tested on a different samples. Random Forests and BDTs improve their accuracy while maintaining most of their desirable proprieties such as simplicity. Some advantages, nevertheless, are sacrificed: some of the interpretability by an human's eye, speed, and, for *AdaBoost*, robustness against the overlapping in the distributions of the discriminating variables and mislabeling of the training sample jets [66].

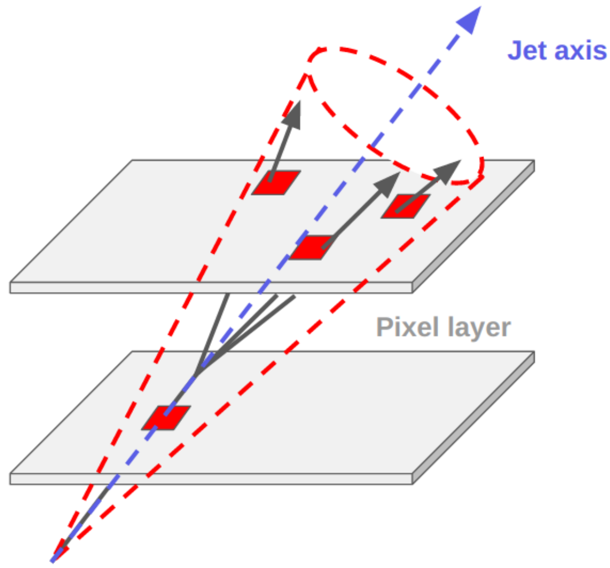


Figure 5.8: The decay of a boosted b-hadron produces a jump in the number of clusters when it occurs between two subsequent layers.

5.4.2 Random Forests

Random Forests combine the simplicity of decision trees with flexibility, resulting in a vast improvement in accuracy. Even if they are not used in this research, an illustration of their functioning procedure is beneficial for introducing *AdaBoost*. As suggested by their name, Random Forests are classifiers that use the majority vote of a group of independent, randomly grown decision trees to classify their input (Fig. 5.14).

Each decision tree of the forest is grown with the following procedure. First of all, a *bootstrapped dataset* is built by randomly picking elements of the original sample, allowing to pick the same element more than once. A decision tree is then grown from the bootstrapped dataset following the procedure described in Sec. 5.4.1 with a difference: the discriminant variables candidate to become the cut of each node are chosen among a fixed size *random* subset of the available variables. Furthermore, no maximum depth is set for the trees.

The procedure is repeated for each tree of the forest. The combined use of bootstrapped dataset and random choice of the variables end up in a wide variety of trees. Ultimately, this is the feature of the Random Forest that accounts for its flexibility[3]. When every tree has been grown, the accuracy of the temporary forest is checked as follows. To classify an input object, it is passed to every tree of the forest, and the most voted classification is chosen. The procedure of bootstrapping the data and then using the aggregate to classify an input is called *bagging*. Since the bootstrapped datasets have sizes equal or smaller then the original's, and since repetition of previous choices of elements is allowed in bootstrapping, for each specific boostrapped dataset an *out-of-bag* dataset can be defined, which contains the elements of the original dataset that were never chosen during the specific bootstrapping. The elements of the out-of-bag samples are given as inputs to the trees that were grown without using them, and are classified according to their majority vote. The accuracy of the temporary forest is defined by the proportion of out-of-bag elements correctly classified. The proportion of them incorrectly classified is called *out-of-bag* error.

All the procedure for growing the forest is repeated many times, each time choosing a different size for the set of variables used to build the nodes of the trees. Each iteration produces a different forest.

[3]See [66] for a complete mathematical proof ot this statement.

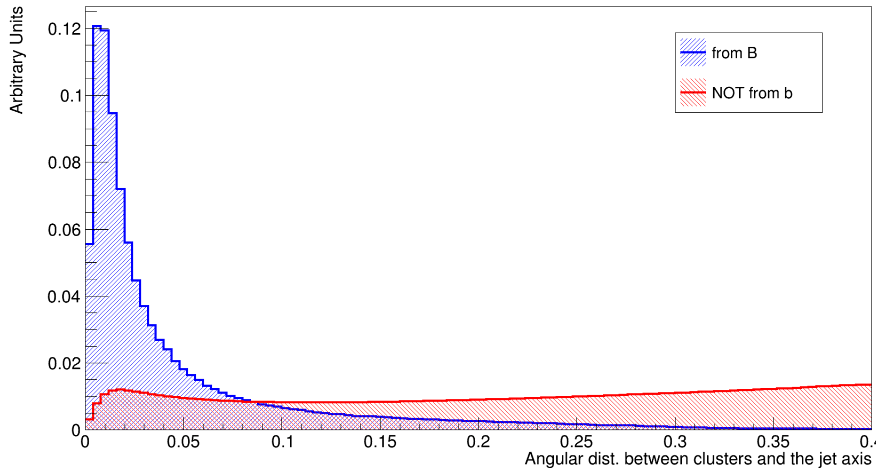


Figure 5.9: Angular distance of b-jets' clusters with respect to the jet direction. It is observed that about 90% of the clusters coming from the b-hadron fall in a cone of $dR = 0.04$.

The one that minimizes the out-of-bag error is the final Random Forest.

5.4.3 Boosted Decision Trees with *AdaBoost*

Boosting procedures combine the outputs of many "weak" classifiers to produce a powerful "committee". If the base classifiers are decision trees, they get the name of *Boosted Decision Trees*. BDTs are in some sense similar to Random Forests, but, as we shall see, they carry some deep differences as well. *AdaBoost* [64] is the boosting procedure applied to build the BDT used in this research. It builds a forest which is different from the Random Forest described in Sec. 5.4.2 in three main aspects:

- A maximum depth is fixed for each tree in the forest, while the Random Forest uses full grown trees.
- Some trees get more say than others in the final classification: the *amount of say* of each tree is weighted by the algorithm; while in a Random Forest the vote of each tree is treated equally.
- The trees are not independent, since the growth of each tree is influenced by the errors made by the previously grown tree. In a Random Forest, each tree grows independently from the others.

Because of their fixed maximum depth, *AdaBoost*'s trees are *weak classifiers*[4]. Nevertheless, in many applications, *AdaBoost* was proven to be optimized by the use of simple "stumps", i.e. trees composed by a single root node and two leaves; meaning that the fixed maximum depth is generally not an handicap for the algorithm. Fig. 5.15 , taken by [66], shows the power of AdaBoost of improving the performance of a stump very close to random guessing (which would have a probability of 0.5 of mistaking the classification) over a set of simulated data. A single decision tree with 244 nodes is outperformed after about 20 iterations of the *AdaBoost* procedure. Each iteration represents the addition to the BDT of a new stump. To understand how such an improvement is achieved, let us take a closer look to the process of growing the BDT.

To grow a BDT, the first step is to weight each element of the training sample equally with a weight $w_i^0 = 1/N_{tot}$, where N_{tot} is the number of elements in the training sample, the superscript

[4]They are given this name because they are not grown to the extent of reaching the maximum discriminating power they could achieve individually.

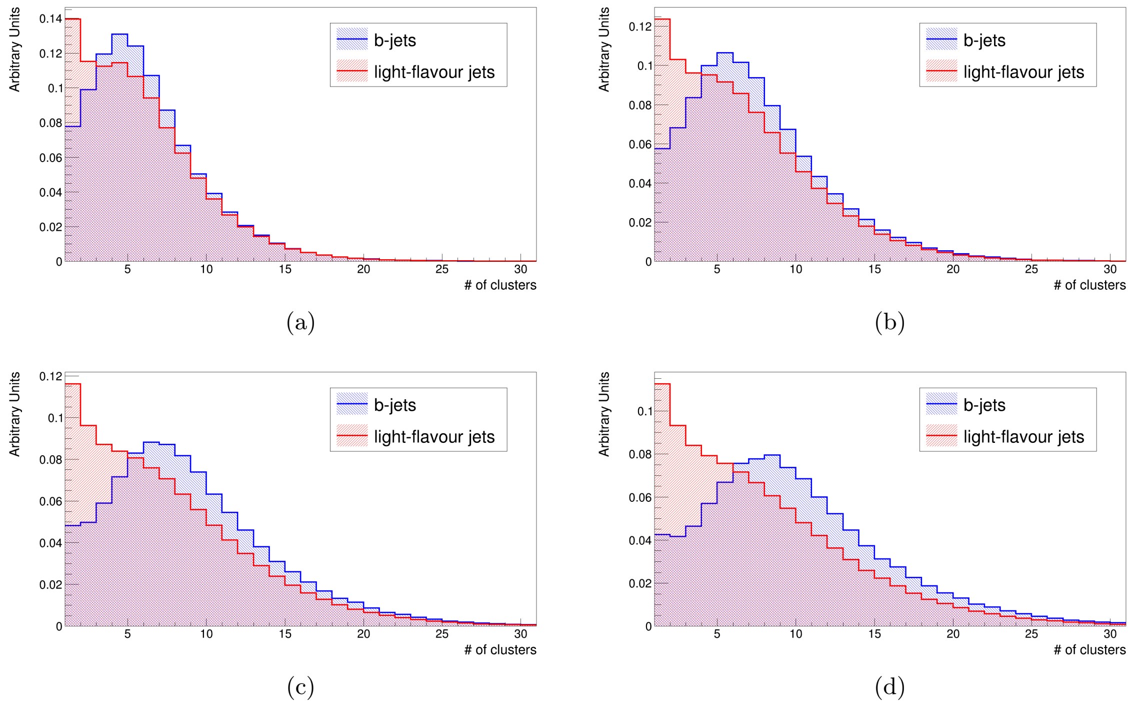


Figure 5.10: Distributions of the number of clusters in the IBL (a), B-layer (b), PIX1 (c) and PIX2 (d).

keeps track of the number of iterations made, and the subscript indicates the element of the sample. A decision tree with a fixed maximum length is then grown using the technique described in Sec. 5.4.1 on a bootstrapped sample. A Total Error is defined as the ratio between the number of false positives and false negatives that the tree makes on the out-of-bag sample and its N_{tot}. The Amount of Say is then computed for the tree, where

$$\text{Amount of Say} = \ln\left[\left(\frac{1-\text{Tot Err}}{\text{Tot Err}}\right)^{\beta}\right], \tag{5.4}$$

where β is a tunable parameter.

As can be seen in Fig. 5.16, Eq. 5.4 assigns relatively high positive values of the Amount of Say to trees with a low Tot Err, no Amount of Say to trees that have a Tot Err of 0.5 and are thus identical to random guessing, and relatively high negative values to trees with a low Tot Err. This last feature is justified by observing that a decision tree that always misclassifies the input, if "inverted", becomes very good in correctly classifying the sample elements.

The Amount of Say of the tree is used to reweight the sample elements w_i^0 according to the following formula:

$$w_i^1 = w_i^0 e^{h_i \times \text{Amount of Say}}, \tag{5.5}$$

where h_i is a value for each element that equals to +1 if it was incorrectly classified by the tree and -1 otherwise. After the application of Eq. 5.5, the new weights w_i^1 are normalized to add up to 1. Hence, the algorithm *reduces* the weights of the elements that were *correctly* classified, and *increases* the weights of the elements that were *incorrectly* classified. The goal of the reweigthing is to emphasize to the following tree the need to correctly classify the elements that are incorrectly classified by the preceeding. Actually, once the reweighting has been performed, the training sample is bootstrapped

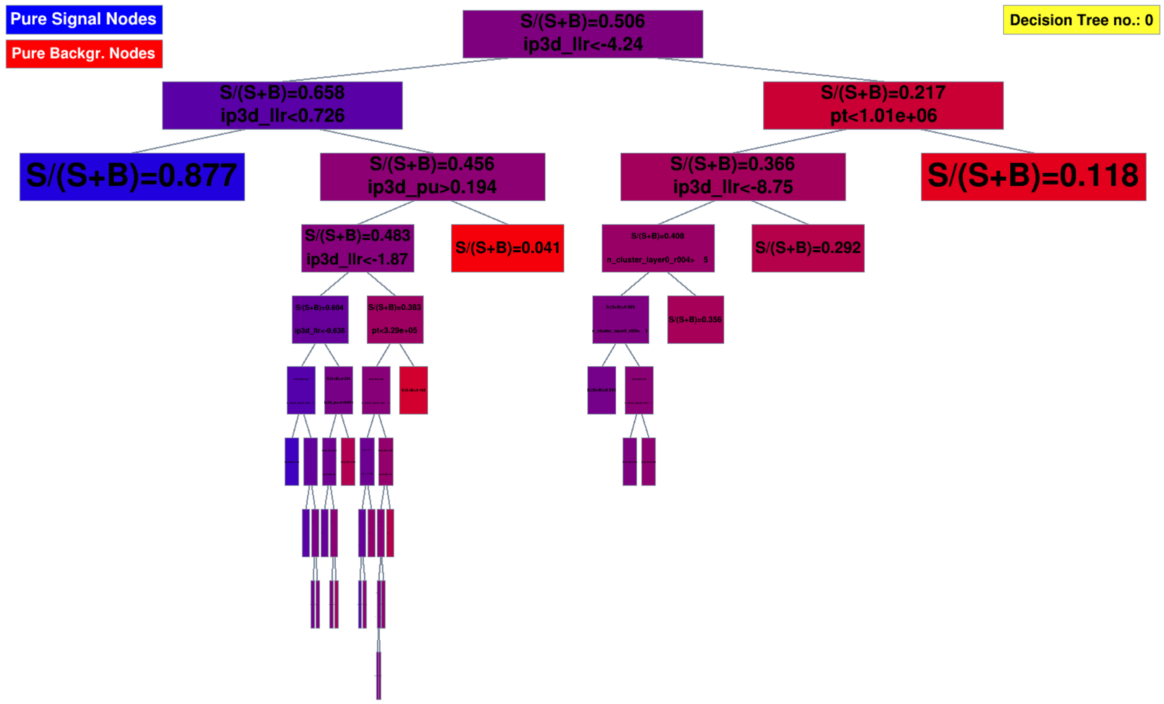


Figure 5.11: One of the decision trees built by the BDT during the training on a training sample containing part of the simulated *light*- and *b*-jets.

accordingly to the probability distribution determined by the elements weights, resulting in a sample in which the occurrence of the elements of the original sample is proportional to their weights. Each element of the bootstrapped sample is then reweighted to $w_i^1 = 1/N_{tot-bootstrap}$ and the process is iterated to build the following trees. With the described bootstrapping procedure, the algorithm succeeds in using the errors made by the n_{th} tree to influence the growth of the $n + 1_{th}$. Thus, the trees are not independent, each one being influenced by the preceeding. The Amount of Say is also used in the classification when the BDT is ready, which is performed trough the weighted average:

$$\text{MVA}_{\text{Boost}}(\mathbf{x}) = \frac{1}{N_{\text{trees}}} \sum_{i}^{N_{\text{trees}}} \text{Amount of Say}_i \times c_i(\mathbf{x}), \tag{5.6}$$

where $\mathbf{x}$ is the set of realizations of the discriminating variables for the input object, $c_i(\mathbf{x})$ are the results of the individual classifiers, encoded for signal and background as $c_i(\mathbf{x}) = +1$ and -1 respectively; and N_{trees} is the number of trees in the BDT. $\text{MVA}_{\text{Boost}}(\mathbf{x})$ is thus the multivariate boosted event classification: large values of it indicate signal like inputs, and small values indicate background-like.

5.5 Application

The BDT used for this research was designed using the Tool for Multivariate Analysis (TMVA) [68] integrated in the ROOT analysis framework [13]. In this section, I will describe its design and the chosen parameters for the training procedure; and I will present the resulting performance in discriminating *b*-jets from *light*-flavour jets before and after the inclusion of the information on the number of clusters per layer.

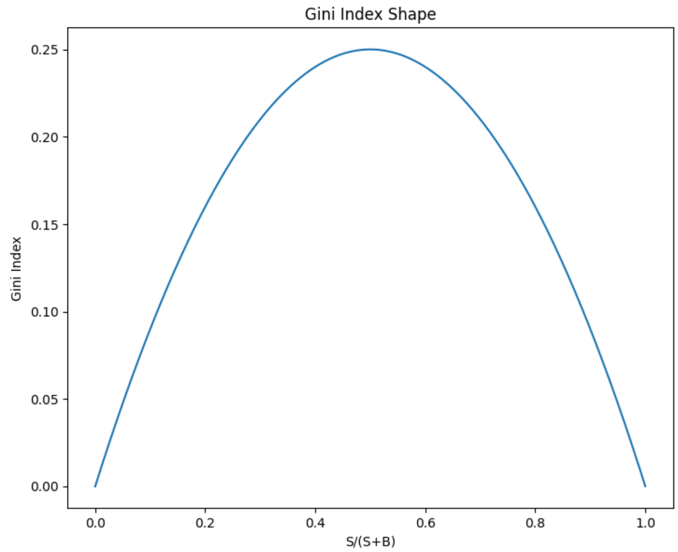


Figure 5.12: The higher is the density of Signal or Background jets in the node, the closer to 0 is the Gini Index.

5.5.1 BDT design

The training of the BDT is performed over half of the jet sample, resulting in a training sample of 113428 b-jets and 390077 *light*-flavour jets randomly picked by the TMVA. The *light*-flavour jets in the training sample are then globally reweighted to the number of b-jets to prevent the BDT from learning fake features from the different population of regions of the phase space for signal and background provoked by the unbalance in the number of b-jets and *light*-flavour jets. Even if they are not considered discriminating variables per se,, p_T and η of jets are included in the training, to allow the BDT to take advantage of their eventual correlations with the other input variables. So, to avoid differences between signal and background in these distributions from being interpreted as discriminating by the BDT, the signal jets are reweighted in p_T to match the spectrum of the *light* background jet. We assumed the differences in the η distributions to be small enough to avoid a reweighting on this variable as well. No reweighting is applied at the evaluation stage of the BDT.

The values of the parameters of the BDT must be carefully chosen to guarantee its optimal performance in the classification of the evaluation sample from the set of discriminating variables. In the following, such performance are analyzed over three sets of variables: IP3D, IP3D+NCLUS, IP3D+NCLUS+NCLUSDIFF (Tab. 5.1).

The parameters to tune are:

- Number of trees in the BDT,
- Maximum depth of the trees,
- Minimum Size of the nodes,
- Number of cuts to evaluate the Gini Index,
- Size of the bootstrapped samples,
- Exponent β of the Amount of Say of the *AdaBoost* algorithm.

A carefully chosen number of trees in the BDT is needed to guarantee the convergence of the classification performance to its best value on the sample. It could not be obtained by a undersized forest

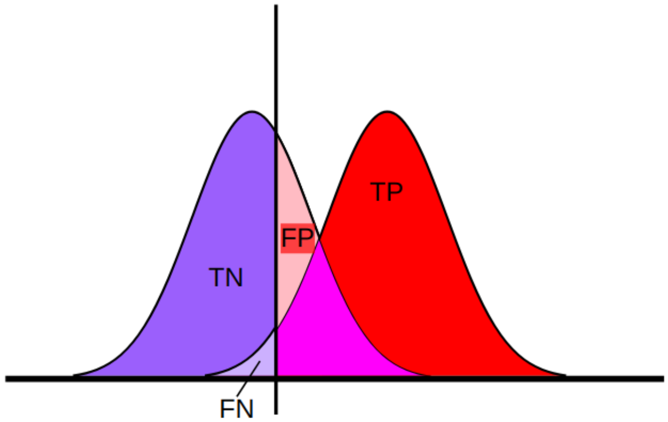


Figure 5.13: Four possible outcomes for the classification of an input by a specific cut: True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN).

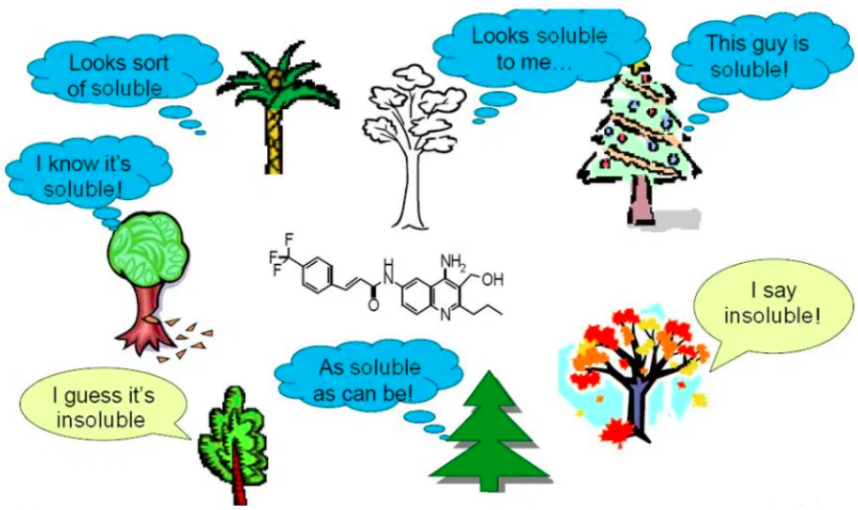


Figure 5.14: A Random Forest trying to decide wether a molecule is soluble.

(Fig. 5.15), while an oversized one would be uselessly time consuming. The tuning of the maximum depth of the trees is related to the risk of over-training the BDT, i.e. to optimize it on the specific training sample by learning also its casual features that would not be present in any other random sample. Furthermore, *AdaBoost* is known in literature to prefer weak classifiers, so we expect shallow trees to be the good choice. The minimum size of the nodes is correlated to the maximum depth of the trees, since deeper ones will divide the phase space in less populated leaves. The risk of admitting very small nodes is to grow a lot of over-trained insignificant leaves, while forcing them to be too big could prevent the BDT from reaching optimality. The number of cuts to evaluate the Gini Index, instead, is exclusively a matter of computational power. The best cut at each node is surely found only if every possible cut is checked, but it is possible that coarser granularity does not affect the overall performance while making the training significantly faster. To tune the size of the bootstrapped sample, the pay-off of two opposite tendencies must be evaluated. If the bootstrapped sample is too small, the tree will not have enough statistics to grow properly on; while an imprecise evaluation of the Tot Err rate on a undersized out-of-bag sample would worsen the weighting of the sample for the following

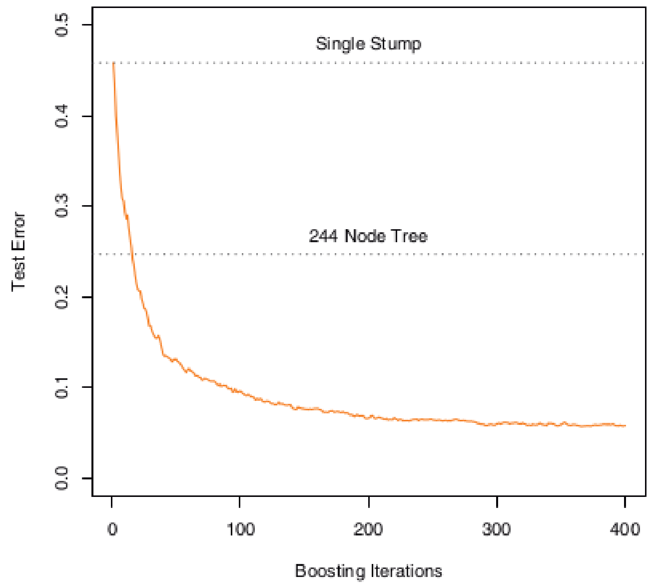


Figure 5.15: Test error rate for boosting with stumps as a function of the number of iterations.

iteration of *AdaBoost*. Last is the exponent β of the *AdaBoost* algorithm. It affects how much each tree is influenced by the errors of the previous tree. A low value of β is related to a "slow training" of the algorithm; but it has also an effect on the final classification performance that must be optimized.

It is not possible to predict from first principles which configuration of the BDT will perform best. Some mathematical reasoning can be done to guide an initial choice of the parameters [66], but a trial-and-error procedure is always needed to tune the classifier. The procedure we adopted was to adapt the parameters used for the optimization of the MV2 [46] tagger of ATLAS to our sample, and make the obtained parameter oscillate around their values while monitoring the final classification performance of the BDT. The results are listed in Tab. 5.2.

5.5.2 IP3D

We begin by training the BDT with the IP3D and kinematic variables (Tab. 5.2). Thus, we set the Maximum Depth to 5 (Sec. 5.5.1). The distributions of the outputs of the IP3D algorithm in our sample are shown in Fig. 5.18 and 5.19. The distributions show the expected behaviour, i.e. a higher likelihood of being b- jets for signal jets with respect to background jets (Fig. 5.18a), a higher likelihood of being *light*-flavour jets for background jets with respect to signal jets (Fig. 5.18b), and the shape of the log-likelihood ratio distribution expected from Sec. 4.3.1.

Then, to evaluate the performance of the BDT, we compute its Receiver Operating Curve (ROC), both inclusive and in bins of jet p_T. A ROC curve is a plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The distribution of the MVA output of the BDT is evaluated on the evaluation sample (Fig. 5.20a). Then, a scan is made on the possible *efficiency working points*, i.e. values of the cut between signal and background that give a chosen area of True Positive (Fig. 5.13). The *rejection*, defined as 1/False Negative, is then computed for each working point. The ROC is the curve resulting by plotting the rejection over the efficiency working point (Fig. 5.20b). Since in the best scenario the MVA distributions for signal and background jets do not overlap (resulting in a value of FN equal to zero for every working point) the bigger is the area under the ROC curve the better is the BDT in the classification task.

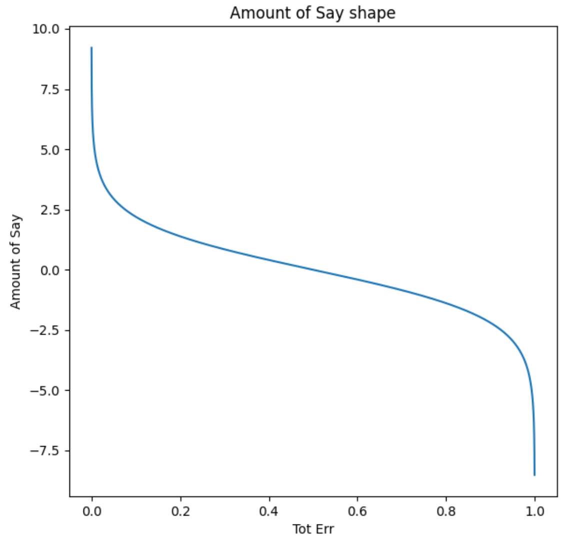


Figure 5.16: Shape of the Amount of Say for $\beta = 1$. Trees that make a very low Tot Err have a relatively large Amount of Say; trees whose Tot Error is equal to random guessing (i.e. trees that have Tot Err = 0.5)) have an Amount of Say = 0; trees that have a very high Tot Err have a relatively high negative value of the Amount of Say. β is a parameter related to the learning rate of the algorithm that have to be tuned to optimize the training.

Fig. 5.21 clearly shows the deterioration of the discrimination power of the BDT with the increase of the jets transverse momenta.

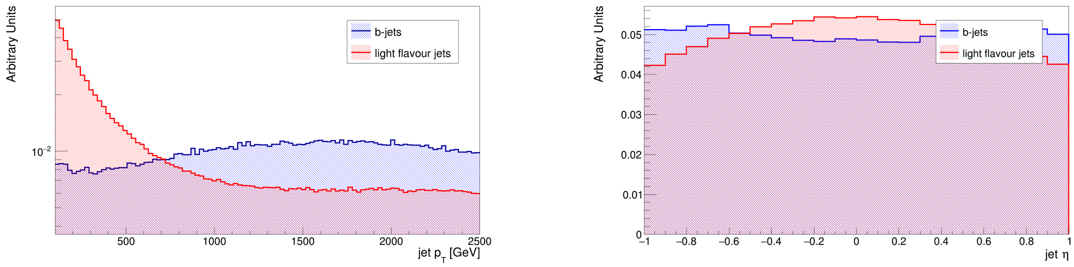
5.5.3 IP3D+NCLUS

We proceed by training the BDT with the NCLUS, IP3D and kinematic variables. Thus, we set the Maximum Depth to 9. The distributions of the number of clusters per layer have already been shown in Fig. 5.10, and the kinematic and IP3D variable distributions are unchanged from the previous sections. Fig. 5.22a and Fig. 5.22b show the output distributions of the BDT on the signal and background jet and the consequent ROC curve; including all the jet p_T values. Fig. 5.23 shows the ROC curves in increasing bins of jet p_T.

IP3D+NCLUS vs IP3D

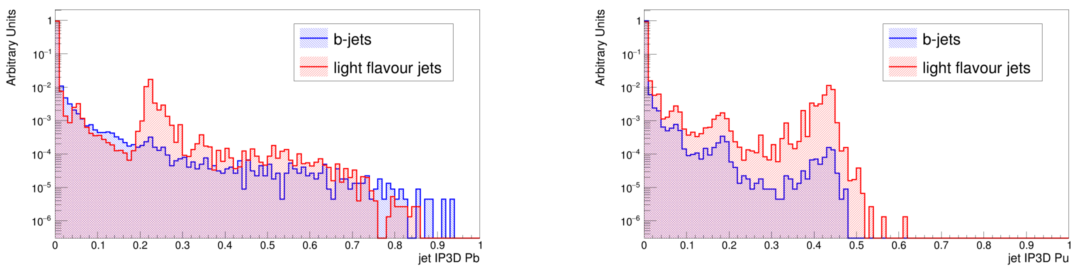
The comparison of the performance of the BDT with and without feeding it with the cluster information clearly shows how the addition is beneficial over all the p_T range studied in this research. The improvement can be seen in Fig. 5.24, in which the ratios of the ROC curves of the IP3D+NCLUS over the IP3D variable sets are shown to be bigger than one for every efficiency working point and in every jet p_T bin. While looking at this plot, it is beneficial to keep in mind that the usual efficiency working points for b-tagging algorithms in ATLAS are around the 77 % value.

A nice way to visualize the improvement of the classification performance for different p_T regions is shown in Fig. 5.25. There, the rejection of *light*-flavour jets is shown for increasing p_T, for a fixed efficiency working point of 77%. The p_T region between 250 and 650 GeV, in which the cluster information is more beneficial, is clearly visible from the ratio plot. Peaks of 30 % of improvements in the rejection performance of the BDT are reached.



(a) (b)

Figure 5.17: Kinematic distributions of jets in the sample: jet p_T (a) and jet η (b).



(a) P_b (b) P_{light}

Figure 5.18: Likelihood P_b (a) and P_u (b) of the *b*- and *light*-flavour jet hypotheses.

5.5.4 IP3D+NCLUS+NCLUSDIFF

We finally train the BDT with the NCLUSDIFF, NCLUS, IP3D and kinematic variables. Thus, we set the Maximum Depth to 12. The distributions of the difference in the number of clusters per layer for subsequent layers is shown in Fig. 5.26. In order to be able to distinguish the cases in which the difference of two subsequent layers is zero because no cluster was found in them from the cases in which it is zero because their non-zero number of clusters is equal, we define NCLUSDIFF variables as follows:

Input	Variable	Description
Kinematics	p_T	jet p_T
	η	jet η
IP3D	P_b	Likelihood of the b-jet hypothesis
	P_{light}	Likelihood of the *light*-flavour jet hypothesis
	$\log(P_b/P_{light})$	Log-Likelihood ratio between the b- and *light*-flavour jet hypothesis
NCLUS	n-clus-layer0	Number of clusters in the IBL in a cone of $dR = 0.04$ around the jet axis
	n-clus-layer1	Number of clusters in the B-layer in a cone of $dR = 0.04$ around the jet axis
	n-clus-layer2	Number of clusters in the PIX1 layer in a cone of $dR = 0.04$ around the jet axis
	n-clus-layer3	Number of clusters in the PIX2 layer in a cone of $dR = 0.04$ around the jet axis
NCLUSDIFF	n-clus-diff-layer01	Difference between the number of clusters in the B-layer and the IBL in a cone of $dR = 0.04$ around the jet axis
	n-clus-diff-layer12	Difference between the number of clusters in the PIX1 and B-layer in a cone of $dR = 0.04$ around the jet axis
	n-clus-diff-layer23	Difference between the number of clusters in the PIX2 and PIX1 in a cone of $dR = 0.04$ around the jet axis

Table 5.1: Input variables used by the BDT

$$
\text{n-clus-diff-layer(n)(n+1)} = \begin{cases} \text{n-clus-layer(n+1)} - \text{n-clus-layer(n)} & \text{if both layer(n+1) and layer(n) have at least one cluster in the cone.} \\ -100 & \text{if both layer(n+1) and layer(n) have zero clusters.} \\ -100 - \text{n-clus-layer(n)} & \text{if only layer(n+1) have zero clusters.} \\ \text{n-clus-layer(n+1)} - 100 & \text{if only layer(n) have zero clusters.} \end{cases}
$$

Parameter	Tuned Value
Number of trees	850
Maximum depth	Number of input variables
Minimum Node Size	0.5 %
Number of cuts	20
Bagged Sample Fraction	0.3
β	0.5

Table 5.2: Tuned Parameters of the BDT

The resulting distributions are shown in Fig. 5.26.

Fig. 5.27a and Fig. 5.27b show the output distributions of the BDT on the signal and background jets and the consequent ROC curve, including all the jet p_T values. Fig. 5.28 shows the ROC curves in increasing bins of jet p_T.

IP3D+NCLUS+NCLUSDIFF vs IP3D+NCLUS

Comparing the performance of the BDT with and without the information of the jump between subsequent layers shows how the addition does not add discrimination power (Fig. 5.29 and 5.30). This observation does not represent a surprise: it reflects the fact that the information on the jump *is already intrinsic to the information on the number of clusters per layer*, and that the machine is learning what can be learned about the jump from the raw number of clusters and their correlation. Furthermore, the differences actually contain *less* information than that contained in the number of cluster: looking at the second, the BDT can also learn about the absolute value of the numbers, whereas the differences do not distinguish between pairs of numbers separated by the same difference.

5.5.5 Summary

In this section, the procedure to choose the parameters of a BDT, the checks on the stability of the BDT's performance to their changes and the application of the BDT to the task of tagging b-jets with three different sets of discriminating variables were illustrated. The BDT have proven to be very stable to changes in the parameter values, saving us from the need of doing a proper automated hyperparameter optimization [71]. The stability is to be expected from the scarcity of discriminating variables that we are using, making our BDT a very simple one. As a comparison, the MV2 BDT is trained on 30 discriminating variables with various amounts of internal correlations, while our most complicated design included 12 variables, 7 of them being very correlated (the cluster information in the various layers). The BDT was shown to benefit from the addition of the NCLUS set of discriminating variables, while the further addition of the NCLUSDIFF set was found to be redundant. We interpreted this observation by inferring that the BDT is able to learn the information on the jump from the raw number of clusters, and argued why this was expected. Fig. 5.31 is a way of visualizing the jump "hidden" in the raw number of clusters.

ROC ratios regarding different angular openings of the jets were computed during this research, the $dR = 0.04$ configuration resulting the better performing among those that we checked.

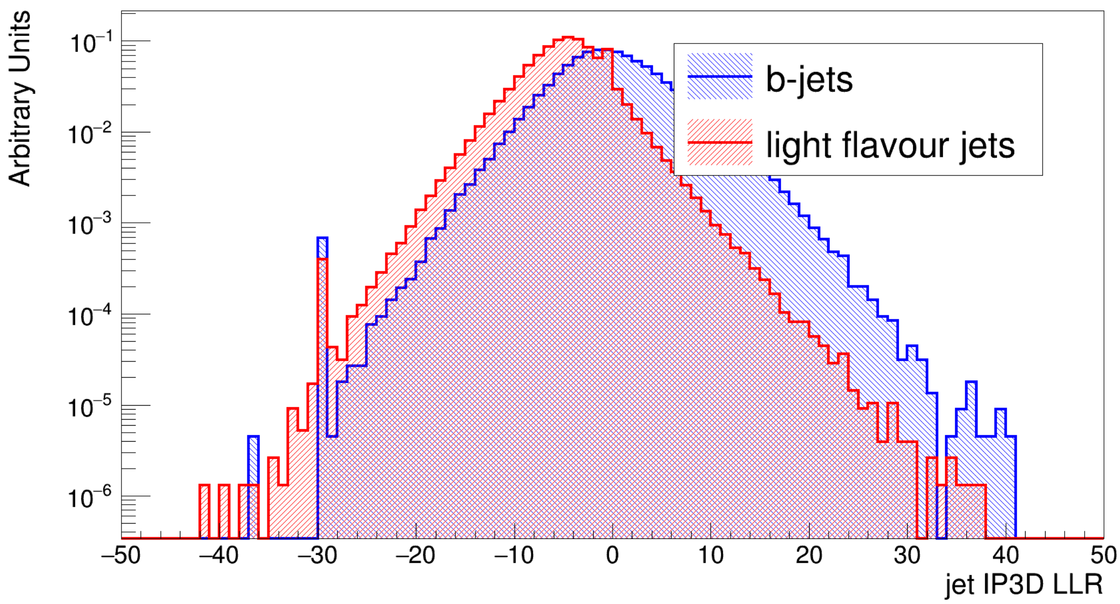
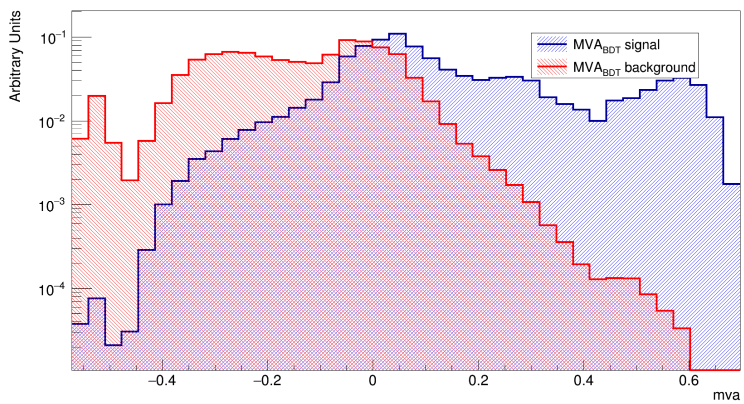
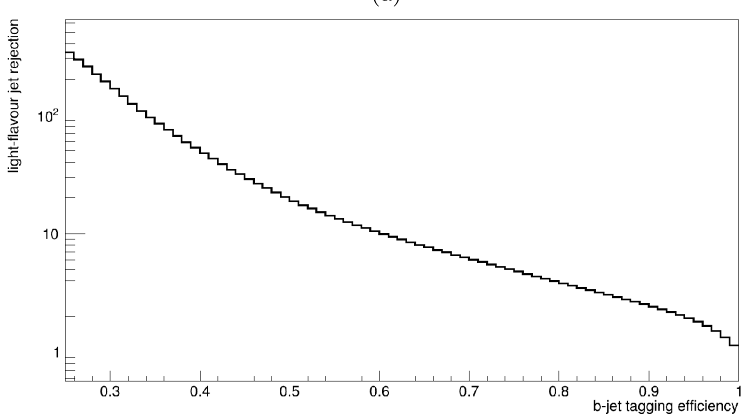


Figure 5.19: Loglikelihood ratio between the b- and *light*-flavour jet hypotheses for signal and background jets.



(a)



(b)

Figure 5.20: MVA output distribution (a) and ROC curve (b) of the BDT with the IP3D and kinematic sets of variables, considering all the pT values.

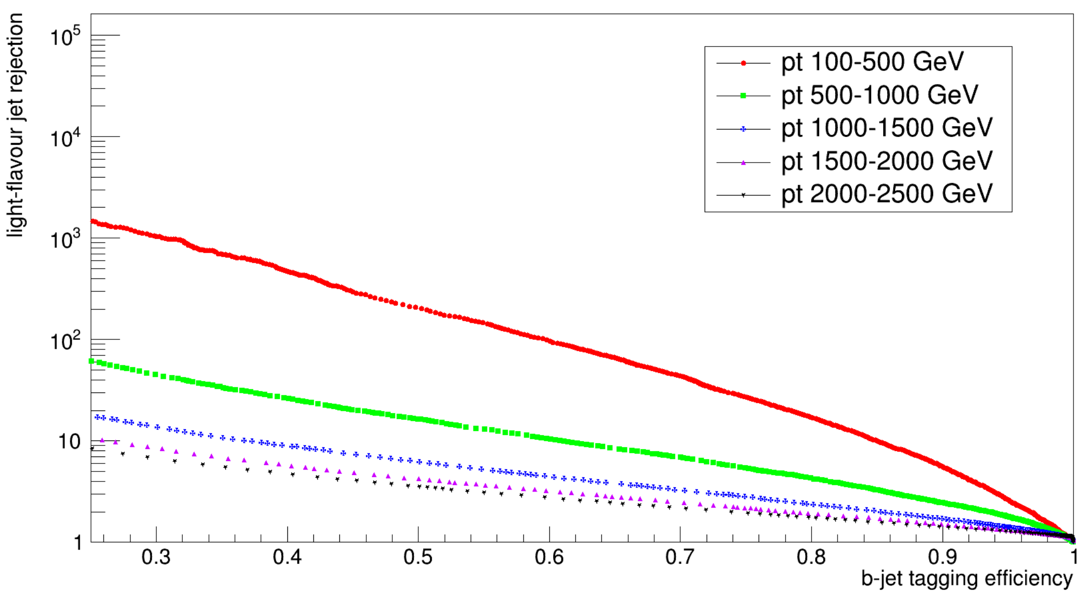
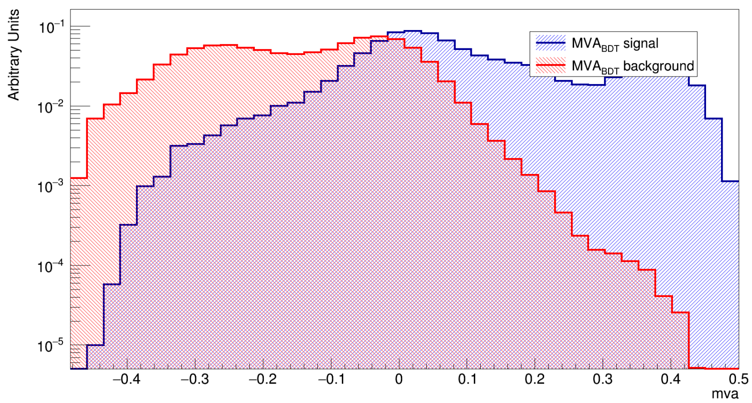
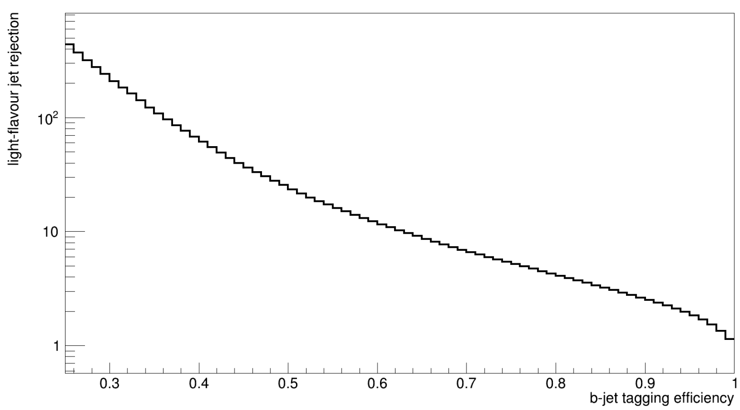


Figure 5.21: ROC curves for increasing jet pT with the IP3D and kinematic sets of variables. The discrimination deteriorates with the increase of the jet p_T.



(a)



(b)

Figure 5.22: MVA output distribution (a) and ROC curve (b) of the BDT with the the NCLUS, IP3D and kinematic sets of variables, considering all the pT values.

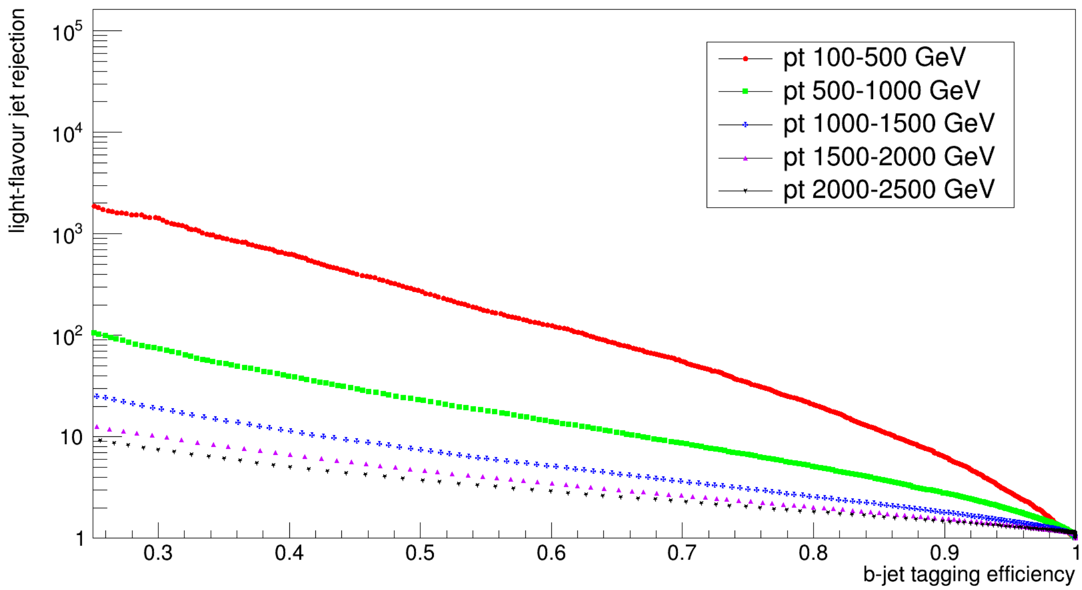


Figure 5.23: ROC curves for increasing jet pT with the NCLUS, IP3D and kinematic sets of variables. The discrimination deteriorates with the increase of the jet p_T.

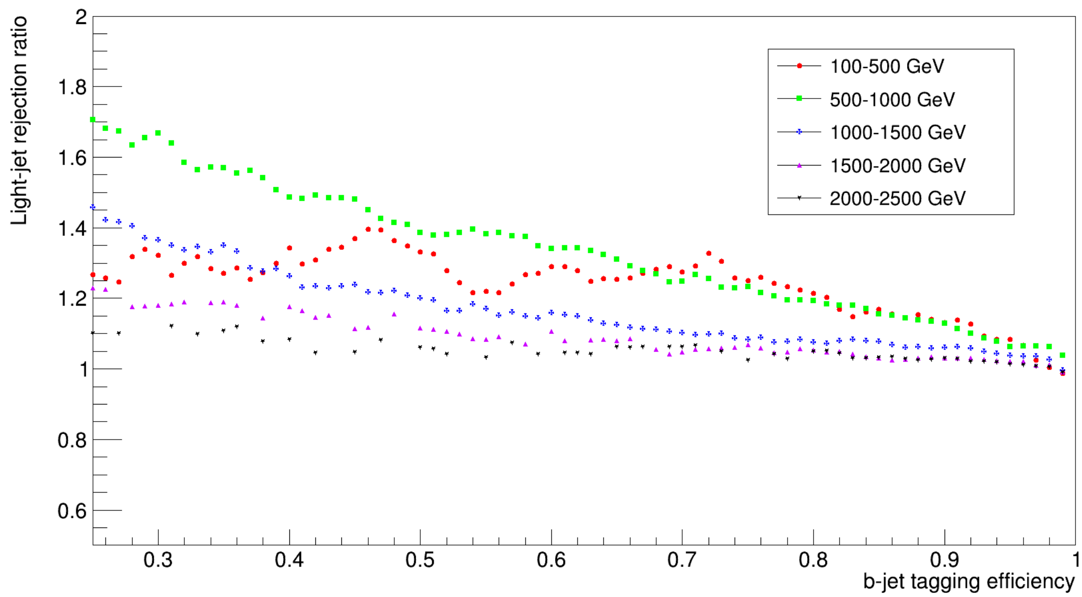


Figure 5.24: Ratios of the NCLUS+IP3D/IP3D ROC curves in the various p_T bins.

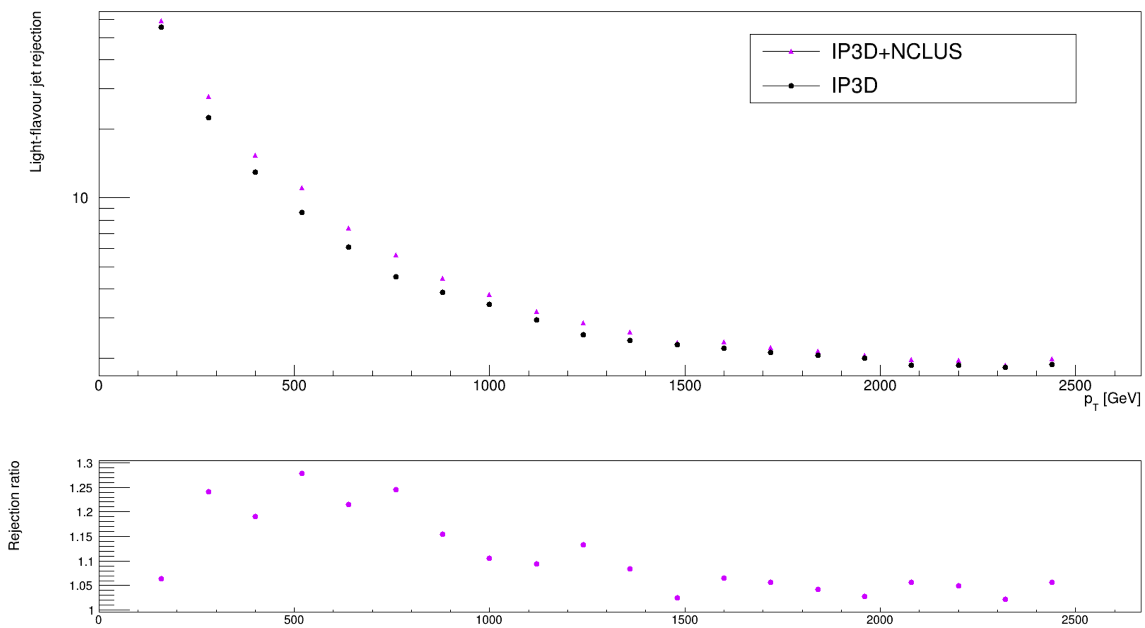


Figure 5.25: *light*-flavour jet rejection performance over pT for a fixed efficiency working point of 77 % for IP3D+NCLUS (Violet) and for IP3D (Black). The ratio shows the p_T regions in which the best improvement is reached, peaking at 30 %.

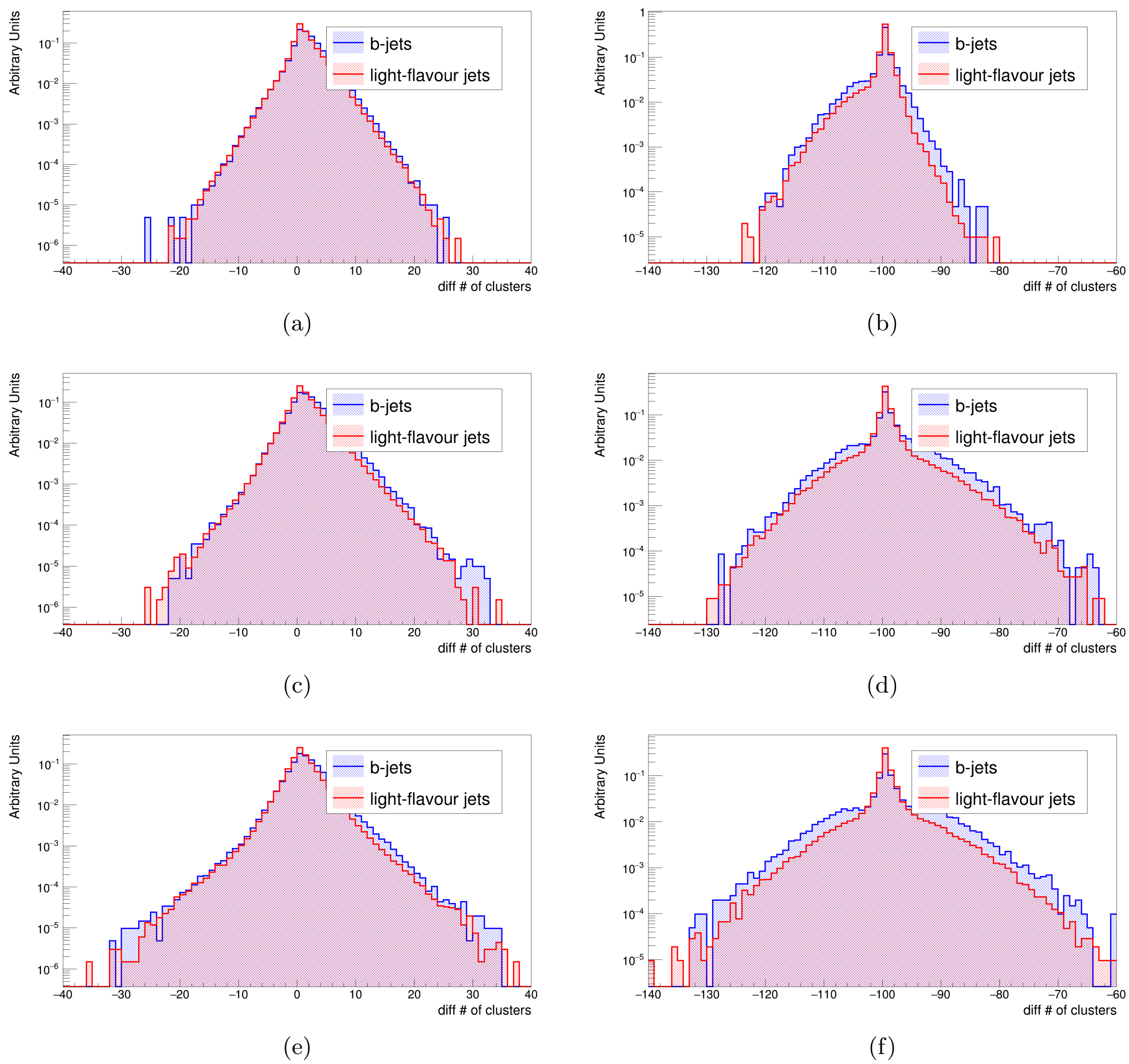
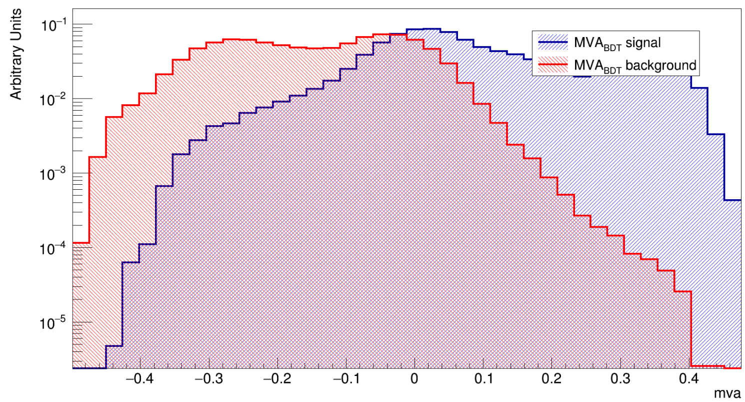
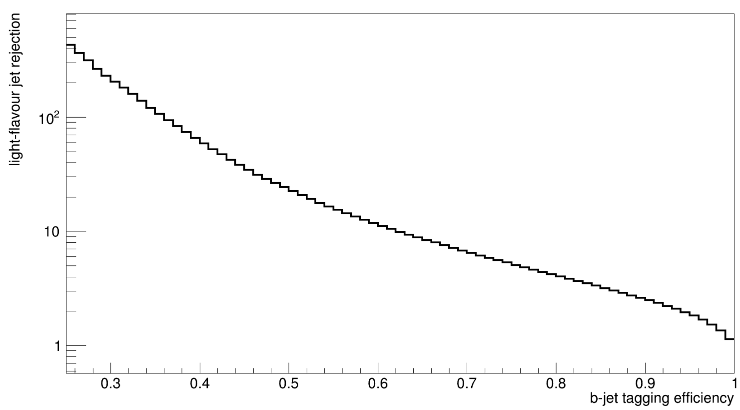


Figure 5.26: Distributions of the differences of the number of clusters in subsequent pixel layers. The two different regions of the variable obtained by distinguishing the cases in which none of the two subsequent layers have zero clusters from those in which at least one of them has zero clusters are shown in different pads. B-layer − IBL, first case scenario (a) and second case scenario (b). PIX1 − B-layer, first case scenario (c) and second case scenario (d). PIX2 − PIX1, first case scenario (e) and second case scenario (f).



(a)



(b)

Figure 5.27: MVA output distribution (a) and ROC curve (b) of the BDT with the NCLUSDIFF, NCLUS, IP3D and kinematic sets of variables, considering all the pT values.

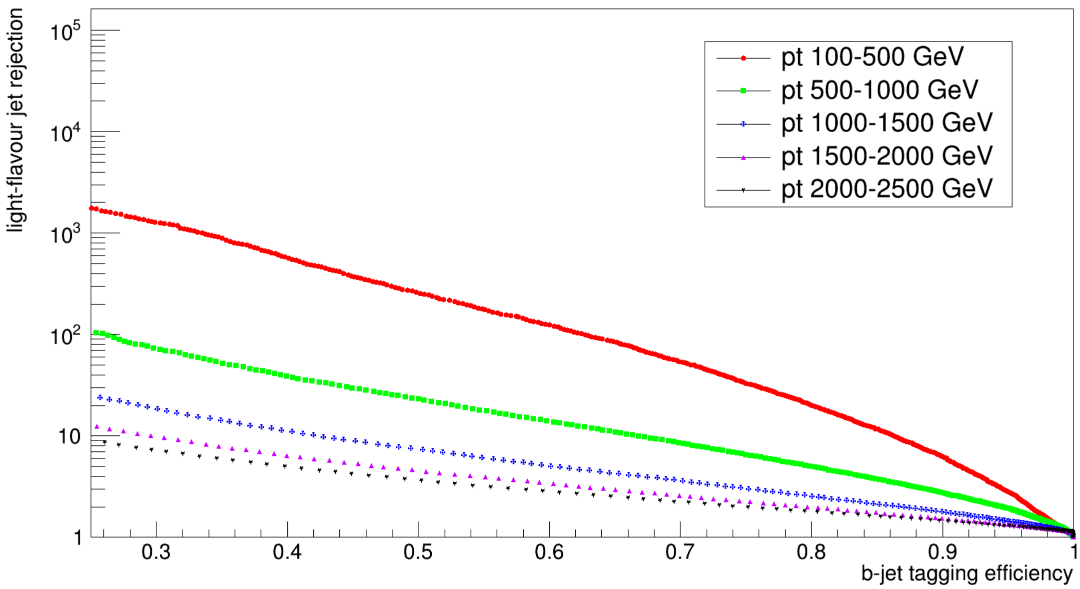


Figure 5.28: ROC curves for increasing jet p_T with the NCLUSDIFF, NCLUS, IP3D and kinematic sets of variables. The discrimination deteriorates with the increase of the jet p_T.

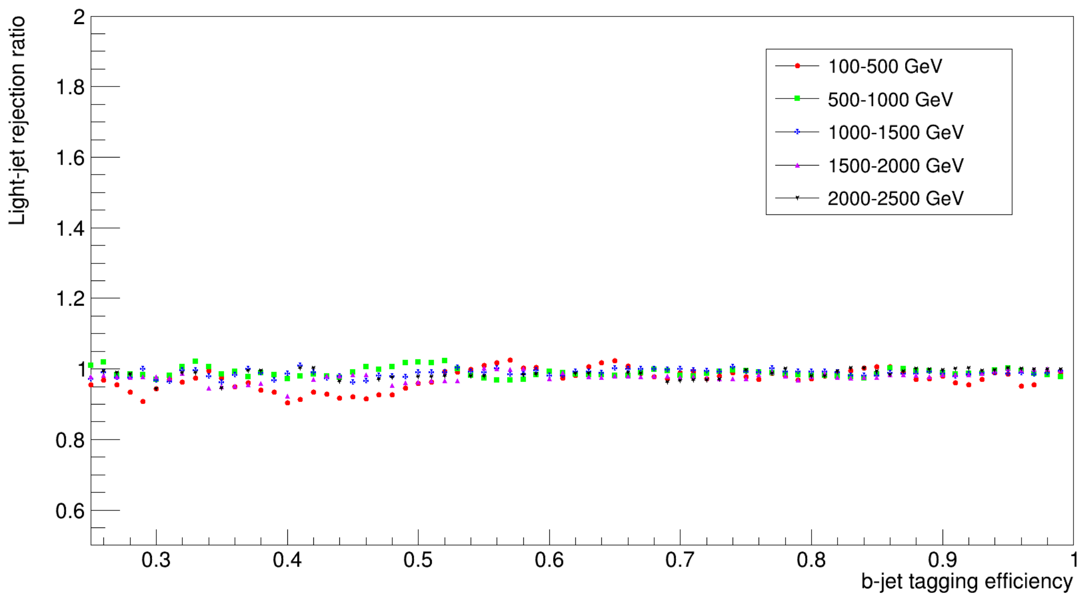


Figure 5.29: Ratios of the NCLUSDIFF+NCLUS+IP3D/NCLUS+IP3D ROC curves in the various p_T bins.

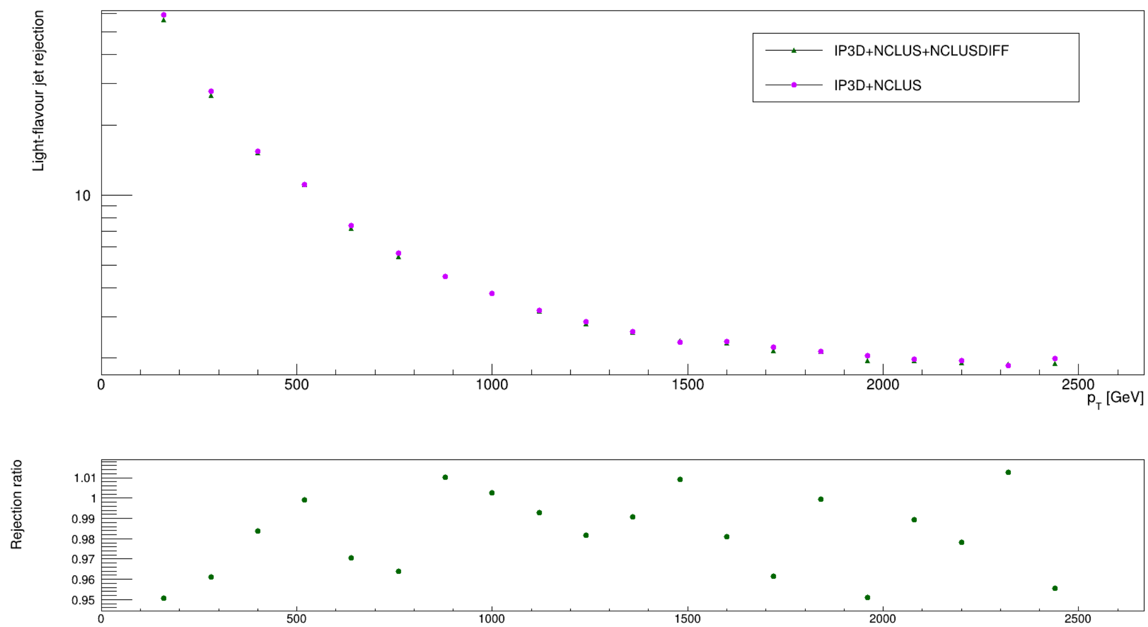


Figure 5.30: *light*-flavour jet rejection performance over pT for a fixed efficiency working point of 77 % for IP3D+NCLUS+NCLUSDIFF (Dark Green) and for IP3D+NCLUS (Violet). No significant gain in the rejection performance can be seen.

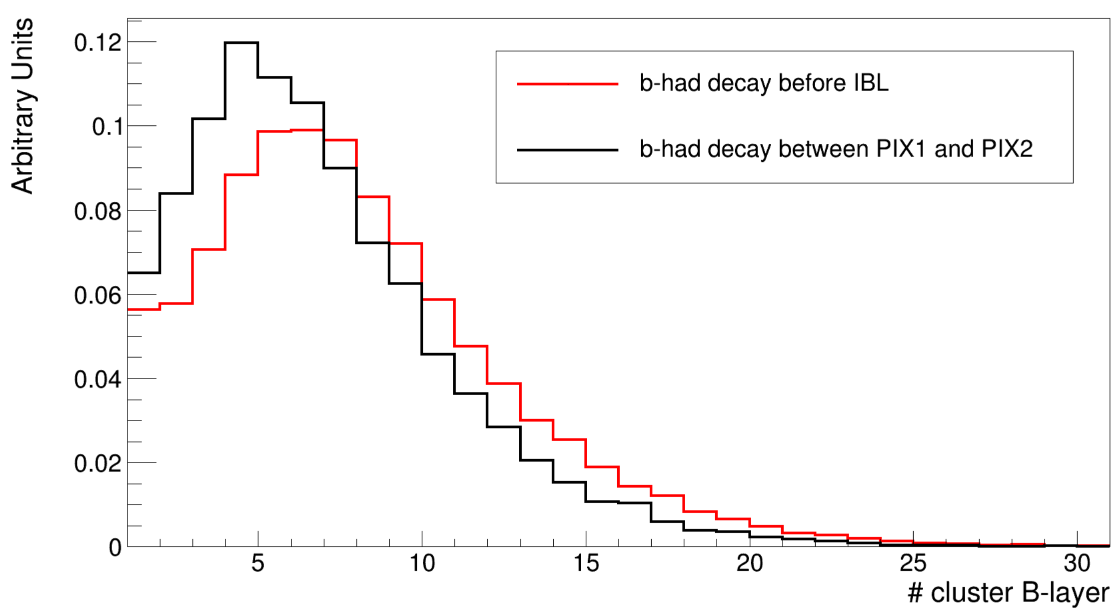


Figure 5.31: A visualization of the jump information in the number of clusters. Here the distribution of the number of clusters in the B-layer is shown for b-jets in which the b decayed before the IBL (Red) and after PIX1 (Black). The former having an higher mean means that the number of clusters increases for the layers more external than the b-hadron decay.

Chapter 6

Impact on physics

Many searches for new physics beyond the SM and many studies of the SM itself benefit from well performing b-tagging algorithms. Examples range from the study of the SM Higgs boson pair production in the $HH \to b\bar{b}b\bar{b}$ decay channel ([11], [10]) to searches of new physics such as graviton and radion particles decaying to heavy fermions or bosons in warped extra-dimension models [65], supersymmetric particles coupled to third-generation quarks [7]; and ultimately any search for new unforeseen physics with couplings to heavy SM particles and b-quarks in particular. In this Chapter, I will introduce the analysis of the Higgs boson decay channel $H \to b\bar{b}$ and the most recent result of searching new heavy resonances decaying to pairs of jets (dijets) with the ATLAS detector. Then, I will elaborate on how improvements of the b-tagging performance at the scale of those obtained in this work and illustrated in Chapter 5 would result in a better sensitivity when searching for Gaussian-shaped resonances decaying to jet pairs.

6.1 $H \to b\bar{b}$

An important class of physics domains in which the b-quark is involved is the Higgs sector: the $b\bar{b}$ decay of the Higgs boson directly tests its coupling to fermions. Such decay channel was observed for the first time in 2018 independently by the ATLAS [39] and CMS [79] collaborations[1].

The $H \to b\bar{b}$ decay mode is by far the dominant at the measured mass of 125.09 ± 0.21 GeV for the Higgs mass: from Fig. 6.1, it can be seen that it accounts for almost 60% of all possible decays. Notwithstanding its abundance, the $H \to b\bar{b}$ was the latest of the Higgs decay modes to be observed, six years later than the much rarer $H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$ discovery channels.

The main reason is that the most copious production process of the Higgs (Fig. 6.2a), the so called *gluon fusion* (Fig. 6.2b), leads to just a pair of b-jets after the decay of the Higgs in $b\bar{b}$. These pairs are almost impossible to distinguish from the background of the b-jets produced via the QCD process $pp \to b\bar{b}$.

To solve this problem, the more distinguishable decays of Higgs bosons produced in a less common production mode had to be considered. The production of the Higgs in association with a W or Z vector boson and its subsequent decay in a pair of $b\bar{b}$ ($VH \to b\bar{b}$) (Fig. 6.3) is the most convenient, since the leptonic decays $W \to l\nu$, $Z \to ll$ and $Z \to \nu\nu$ (where l stands for an electron or a muon) provide signatures that allow to reject much of the QCD background.

Even in the $VH \to b\bar{b}$ decay channels, the challenges posed by the background remain demanding, since top quarks and vector bosons can decay to final states with the same particles as those in Fig. 6.3. For instance, Fig. 6.4 shows a decay chain of a $t\bar{t}$ pair that ends up in a final state containing an electron or a muon and two b-quarks, as it is the case for Fig. 6.3b.

[1] The experiments running at the Tevatron accelerator complex, instead, reported a mild excess of events in 2012 when analysing the full dataset collected at Bevatron and when searching for the SM Higgs boson in the $b\bar{b}$ decay channel[16].

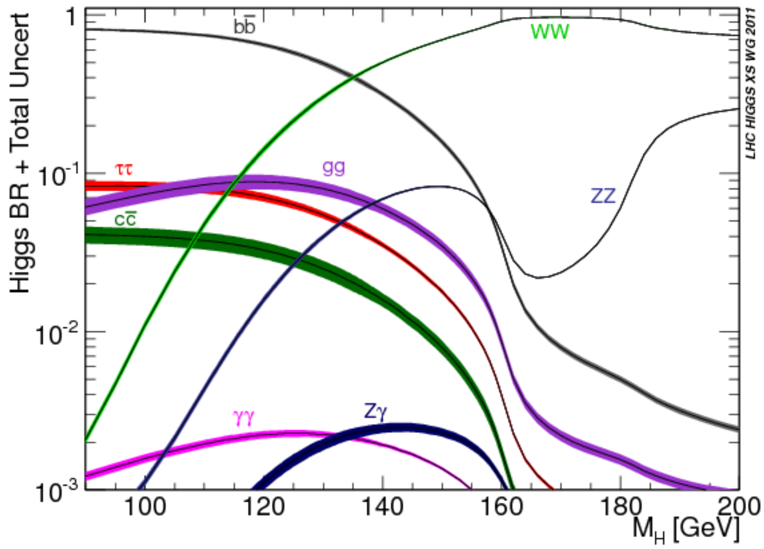


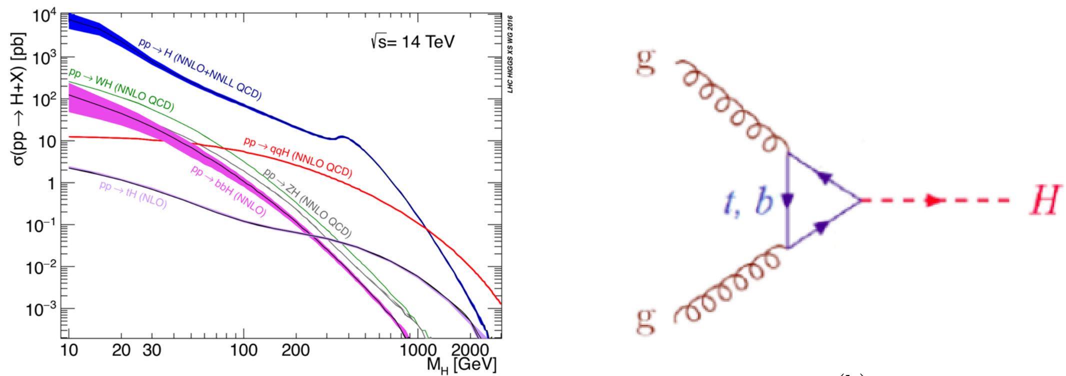
Figure 6.1: Branching Ratios of the Higgs decay channels as they would be for various masses of the Higgs boson. The real value is currently known to be 125.09 ± 0.21 GeV: there, the $H \to b\bar{b}$ branching ratio is of ~ 58 %.

The most important variable for the discrimination of the signal from such backgrounds is the invariant mass m_{bb} of the pairs of b-jets in the data. The centrality of an efficient b-tagging for such an analysis lays in the fact that better performance of the algorithm translate in a higher rejection of jets misidentified as b-jets, thus reducing the background. Hence, better b-tagging performance allow a signal of the same strength to stand out more clearly from the background, the second being reduced. Fig. 6.5 shows the expected distribution of m_{bb} and the observed data summed over all channels and regions of the ATLAS $VH \to b\bar{b}$ latest analysis [38] after the subtraction of all the backgrounds apart from the WZ and ZZ productions. These two are intentionally kept to show the grey peak arising from the Z boson decays to $b\bar{b}$ pairs being at the right mass value, hence validating the analysis procedure. The red shoulder around the 125 GeV value is consistent in shape and rate with the expectation of Higgs boson production.

m_{bb} is not the only discriminating variable used in [38], since other kinematic variables such as the angular separation of the b-jets or the p_T of the associated vector bosons can discriminate between the signal and the various backgrounds. A multivariate analysis based on a BDT is therefore performed: from its output, the analysis of the 139 fb^{-1} of data collected by the LHC led to the observation of $VH \to b\bar{b}$ with a significance of 4.0 σ in the WH and 5.3 σ in the ZH production modes.

6.2 New resonances in the mass distributions of jet pairs

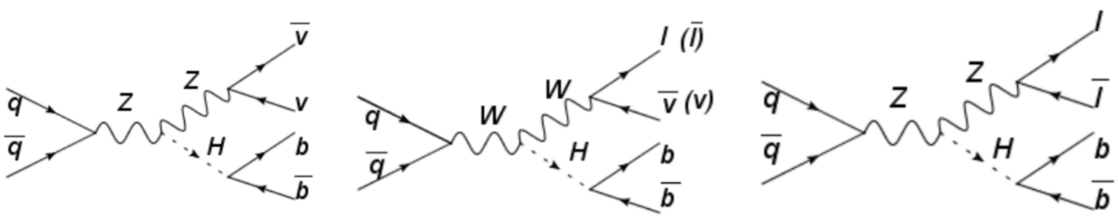
In the SM, dijet events are produced mainly by QCD processes. The invariant mass distribution m_{jj} of such events is predicted to be smootlhy decreasing. Thus, a new particle decaying into quarks or gluons would emerge as a resonance in the m_{jj} spectrum. The data collected by the ATLAS detector at the centre-of-mass energy of $\sqrt{s} = 13$TeV corresponding to 139 fb^{-1} were analyzed in 2019 to search for any evidence of one or more of these resonances [50]. The m_{jj} spectrum ranging from 1.1 TeV to 8 TeV was probed. No excess of events above the predicted background was observed, and hence the results were interpreted in terms of exclusion in the context of several BSM physics scenarios, from excited quarks to quantum black holes and Kaluza-Klein gravitons. Furthermore, generic Gaussian-shaped



(a)

(b)

Figure 6.2: Cross sections for the different Higgs production modes (a) and Feynman diagram of the gluon fusion mode (b), that corresponds to the blue line of the left panel.



(a) $ZH \to \nu\bar{\nu}b\bar{b}$

(b) $WH \to l\nu b\bar{b}$

(c) $ZH \to Vb\bar{b}$

Figure 6.3: Feynman diagrams of the three channels used in the $VH \to b\bar{b}$ observation.

narrow resonance were looked for[2].

In the analysis, b-jets were identified using DL1r at the efficiency working point of 77%. For each event, only the leading and subleading p_T jets were considered. To maximise the sensitivities to different types of resonances, events were classified into three categories based on their b-jets content: one inclusive with no b-tagging requirement, a one-b-tagged category ($1b$), requiring at least one of the two leading jets to be b-tagged, and a two-b-tagged category ($2b$), with both the leading and subleading jets being tagged as b-jets. Events were further divided into categories based on various kinematic and spatial requirements on the jets and on the mass of the resonance m_{jj}.

Once divided the observed data, the m_{jj} distributions of each of the categories were analyzed to see if they were compatible with the expected sources of background processes. The expected background is computed from the observed data by a fitting method known as sliding-window fitting [49], based on a parametric function $f(x) = p_1(1-x)^{p_2}x^{p_3+p_4 \ln x}$, where $x = m_{jj}/\sqrt{s}$ and $p_{1,2,3,4}$ are the four fitting parameters.

The expected background is then compared to the observed data to check if any localised excess in the observed m_{jj} distribution is significantly different from the background only hypothesis. The statistical significance of the eventual excesses is quantified using the BUMPHUNTER test [27]. An evidence of new signal is claimed if the probability for a poissonian fluctuation of the background to produce the observed excess is less than 0.003. No significant deviation from the expected background was observed.

The analysis does not end here, since when the observed data are found to be compatible with the background only hypothesis, constraints on the various signal models that would produce resonances in

[2]This last scenario is the one on which we will estimate the impact of an improved b-tagging in Sec. 6.3

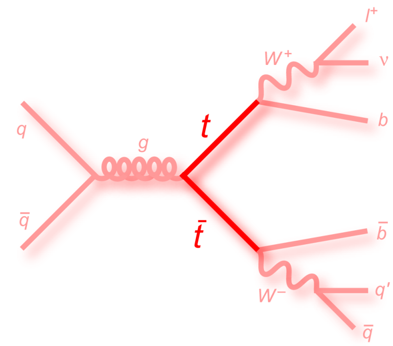


Figure 6.4: An example of a $t\bar{t}$ decay chain that ends up in a final state which is a background for the $VH \to b\bar{b}$ analysis.

the dijet invariant mass distributions can still be computed. This type of test, based on the CLs method [76], tries to quantify how unlikely it is that the m_{jj} distribution arising from a sample *containing* some new particles (the *signal plus background hypothesis*) would fluctuate to become compatible with the background only hypothesis. To do so, MC simulations are performed to produce the expected dijet distributions that would arise from the decays of the new resonances that are being looked for. For instance, the expected m_{jj} spectrum of data containing Z's with masses of 2.5 TeV decaying to pairs of jets would show a resonance around their mass value. Then, the 95% *exclusion confidence level (CL)* of the signal plus background hypothesis is computed. Thus, for example, when the statement is made that "*Z' masses below 2.7 TeV are excluded at the 95% CL*" [50], what is meant is that the probability for the distribution that would arise from the presence of Z's with masses smaller than 2.7 TeV to produce a m_{jj} distribution compatible with the actual observations is less than 5 %.

Constraints were posed on all the studied models: an example can be seen in Fig. 6.6. There, the upper limits posed on the combined cross section σ of the production of Gaussian shaped signals times the kinematic acceptance of the detector (i.e. the percentage of jets the detector is able to collect due to its intrinsic features) times the Branching Ratio for the signal to decay to two jets belonging to the specific categories are shown as a function of the mass of the signal and of its width. Values of this combined quantities higher than their upper limits would have produced m_{jj} distributions compatible with the observed data in less than the 5% of the cases, if the experiment could be reproduced in the limit of infinite statistics.

The relevance of a well performing b-tagging for the analysis described in this section lays in the flavour composition of the jets that pass the $1b$ and $2b$ selections. Since the b-tagging algorithms have non infinite rejection rates of jets not coming from b-quarks; some jet pairs that should not pass the b-tagged selections are mislabelled and access the tagged categories. When this happens, background is introduced in the $1b$ and $2b$ categories. The impact of this fact on the flavour composition of the distributions can be studied in the background MC samples, as it is shown in Fig. 6.7. There, it can be seen the dependence of the background flavours on the resonance mass, and thus on the p_T of the jets. In particular, the facts that *light-light* (ll) become as represented as *b-light* (bl) dijets at ~ 3 TeV in the $1b$ category and that bb jet pairs decrease with the increase of m_{jj} in every category are worth noting.

If the amount of misclassified jets were to diminish in the $1b$ and $2b$ categories, resonances coupling to b-quarks would emerge from the background distribution with fewer events, hence with less integrated luminosity. Similarly, in case of no excess in the data, stricter upper limits could be set in the cases in which no signal were to be found in the data, as it is shown in Fig. 6.8. There, the expected 95 % CL upper limits on the cross section times branching ratio times acceptance times b-tagging efficiency are shown as a function of a Dark Matter Z' mediator mass [1] ; for the latest analysis, and for the previous result obtained with $36\,\mathrm{fb}^{-1}$ and a different, less powerful b-tagging strategy. The previous

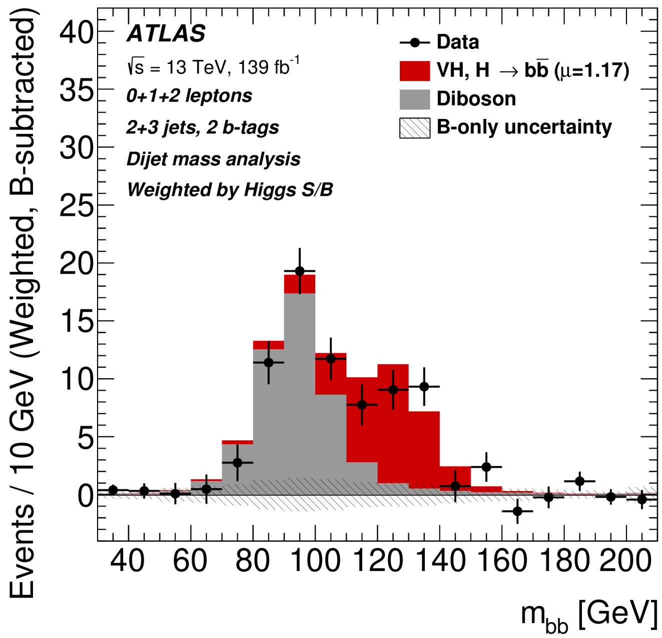


Figure 6.5: Distribution of m_{bb} from all search channel combined after the subtraction of all the background except for WZ and ZZ. The separate peaks of the Z and H productions are clearly visible.

result is scaled to the integrated luminosity of the current result to show the impact of an improved b-tagging on lowering the upper limits set on the signal model. A factor of up to 3.5 improvement beyond that expected from the increase of the integrated luminosity in the expected upper limits is observed.

6.3 Estimate of the impact of the cluster information

In this section, I will elaborate on how much the improvements gained on the b-tagging performance by the cluster information would affect the exclusion upper limits on the Gaussian-shaped signal of the analysis of Sec. 6.2 in the $1b$ category, by means of a reduction of the background entering the m_{jj} spectrum. To do so, I will use the same spectrum used in the analysis, that can be found in [51]. In Sec. 5.5.3, I showed how the cluster information can improve the b-tagging performance of a BDT trained with the outputs of the IP3D algorithms. There, the improvements of the *light*-flavour jet rejection as a function of the jet p_T were shown for an efficiency working point of 77 % (the same adopted in [50]) (Fig. 5.24).

If the improvements gained by the cluster information to DL1r were of the same amount of those gained with respect to the IP3D output, more stringent upper limits could be determined for the cross sections of the studied signals in the b-tagged categories[3]. This is because the fractional contribution

[3]It can be objected that the gains *would not* be the same, because DL1r is a much more complex and performing algorithm that the IP3D algorithm developed in this thesis for studying the discriminating properties of adding cluster information. Still DL1r is constructed from a number of discriminating variables extracted from the reconstructed tracks associated to jets. The higher the transverse momentum of the jet, the more likely that tracks loose key discriminant information which could be, at least partially, recovered by constructing new variables based on clusters. For this reason, while there is no evidence that the improvements when adding cluster-based information to IP3D will translate as they are when adding the same cluster-based information to DL1r, it is expected that some orthogonality in the discriminant

of the mislabelled dijets containing *light*-flavour jets in the *b*-tagged categories would diminish. Let me explain why.

Consider the 1-*b*-tagged category of Fig. 6.7b. The jet pairs that can erroneously pass the $1b$ selection *because of the misidentification of a light-flavour jet as b-jet* are the lc and ll pairs. Let $\#ll_0$ and $\#lc_0$ be the number of *light-light* and *light-c* jet pairs before the $1b$ selection, and $P_{l\to b}$ the probability of misinterpreting the flavour of a *light*-flavour jet as being a *b*-jet. The total number of *light* jets that could wrongly pass the selection is thus $\#l_0 = 2\#ll_0 + \#lc_0$.

The number of *light*-flavour jets that pass the $1b$ requirement is given by:

$$\#l_1^{old} = (2\#ll_0 + \#lc_0)P_{l\to b}^{old}; \tag{6.1}$$

where *old* means that the cluster information has not been added yet. Since the *light*-flavour jet rejection is defined as $1/P_{l\to b}$, the ratio *rejratio* between the new *light*-flavour jet rejection after the addition of the cluster information and the old one (Fig. 5.24) implies

$$rejratio = \frac{P_{l\to b}^{old}}{P_{l\to b}^{new}} \implies P_{l\to b}^{new} = \frac{P_{l\to b}^{old}}{rejratio}. \tag{6.2}$$

Thus, using Eq. 6.2,

$$\#l_1^{new} = \frac{\#l_0 P_{l\to b}^{old}}{rejratio} \tag{6.3}$$

The scaling on the number of *light*-flavour jets passing the $1b$ selection determines the decrease of the total number of dijets in the $1b$ category.

$$\#j_1^{new} = (\#j_1^{old} - \#l_1^{old}) + \#l_1^{new}. \tag{6.4}$$

The first term in brackets of the right hand side of Eq. 6.4 represents the portion of the sample that does not contain *light*-flavour jets. Substituting the expressions of l_1^{new} and l_1^{old} in Eq. 6.4 and further dividing by $\#j_1^{old}$, one obtains that, since $\#j_1 = 2\#jj_1$, the total number of dijets in the $1b$ categories scales with

$$s = \frac{\#jj_1^{new}}{\#jj_1^{old}} = (1 - f_{ll} - 0.5f_{lc}) + \frac{f_{ll} + 0.5f_{lc}}{rejratio}, \tag{6.5}$$

where f_{ll} and f_{lc} are the fractions of dijet pairs of the subscripted type that pass the $1b$ selection (Fig. 6.7b); when the improved *b*-tagging algorithm is applied.

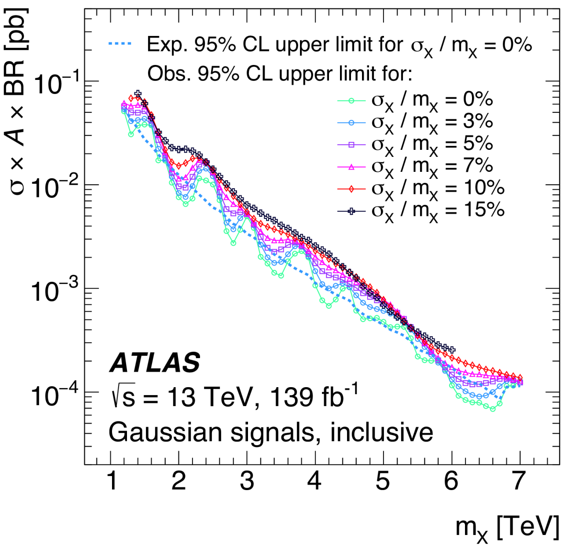
Fig. 6.9 shows the expected upper limit on the number of signal events for Gaussian shaped resonances with standard deviation $\sigma = 200$ GeV and masses ranging from 1.85 TeV to 2.2 TeV at the exclusion CL of 95 %, before and after the scaling of the background determined by the addition of the cluster information under the assumptions made in this estimation. Since *rejratio* is not constant along the energies of jets that originate from the decay of dijet resonances of the considered masses, for each m_{jj} value a *rejratio* value is computed by averaging on all the bins of Fig. 6.9 with p_T values smaller than m_{jj}. This choice was made by considering that the p_T of each of the two jets can assume any value between 0 and m_{jj}, the only constrain being on the two impulses summing up to m_{jj}. Furthermore, for the sake of simplicity, $f_{ll} + 0.5f_{lc}$ was assumed to be equal to 0.4 for all the considered energies, motivated by observing Fig. 6.7b. Improvements of up to 3% of the upper limits for an exclusion CL of 95% on Gaussian shaped signals with the above mentioned standard deviation would be reached in the $1b$ category of the 2019 ATLAS search for new resonances, if improvements of the *light*-flavour jet rejection performance of the scale of those we observe by adding the cluster information to the BDT trained with the IP3D output were made on DL1r. Since the formula that links the number of observed events to the cross section (σ) of the production of the observed signal times the acceptance (A) of the detector times the efficiency (ϵ) of *b*-tagging times the Branching Ratio (BR) for the resonance to decay into a dijet belonging to the $1b$ category is

$$N_{obs} = \sigma \times \mathrm{L}_{int} \times \mathrm{A} \times \epsilon \times \mathrm{BR}, \tag{6.6}$$

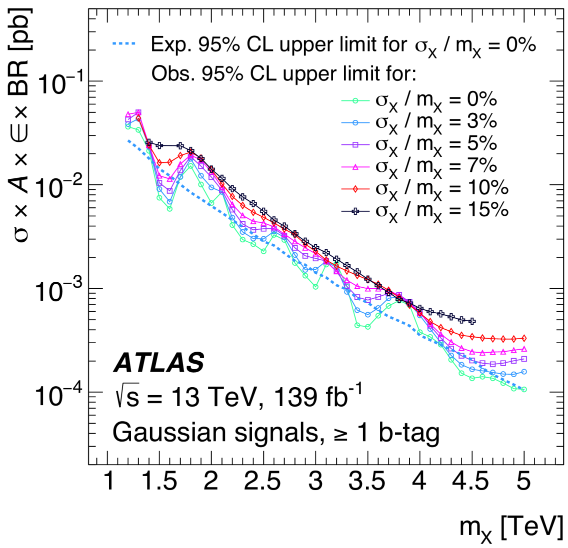
information encapsulated in the clusters will be preserved when tested in conjunction with DL1r.

a improvement of 3% of the upper limits on the number of observed signal events at an exclusion CL of 95% simply translates in a improvement of the same amount of the upper limits on the combined cross section times acceptance times efficiency of b-tagging time Branching ratio that are shown in Fig. 6.6.

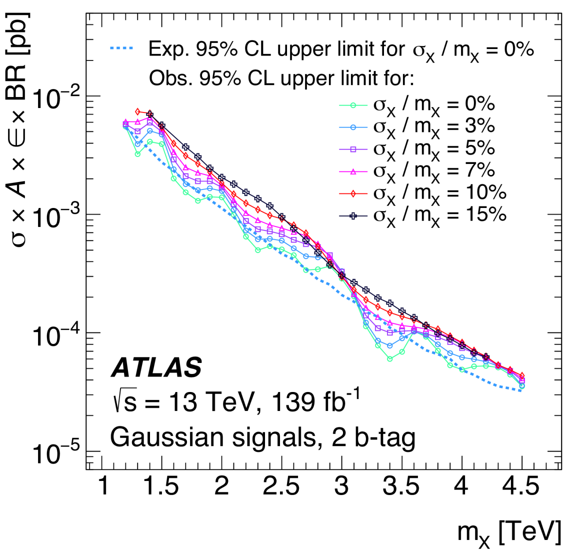
Fig. 6.10 shows the upper limits obtained for various widths of the Gaussian-shaped signal.



(a) Inclusive

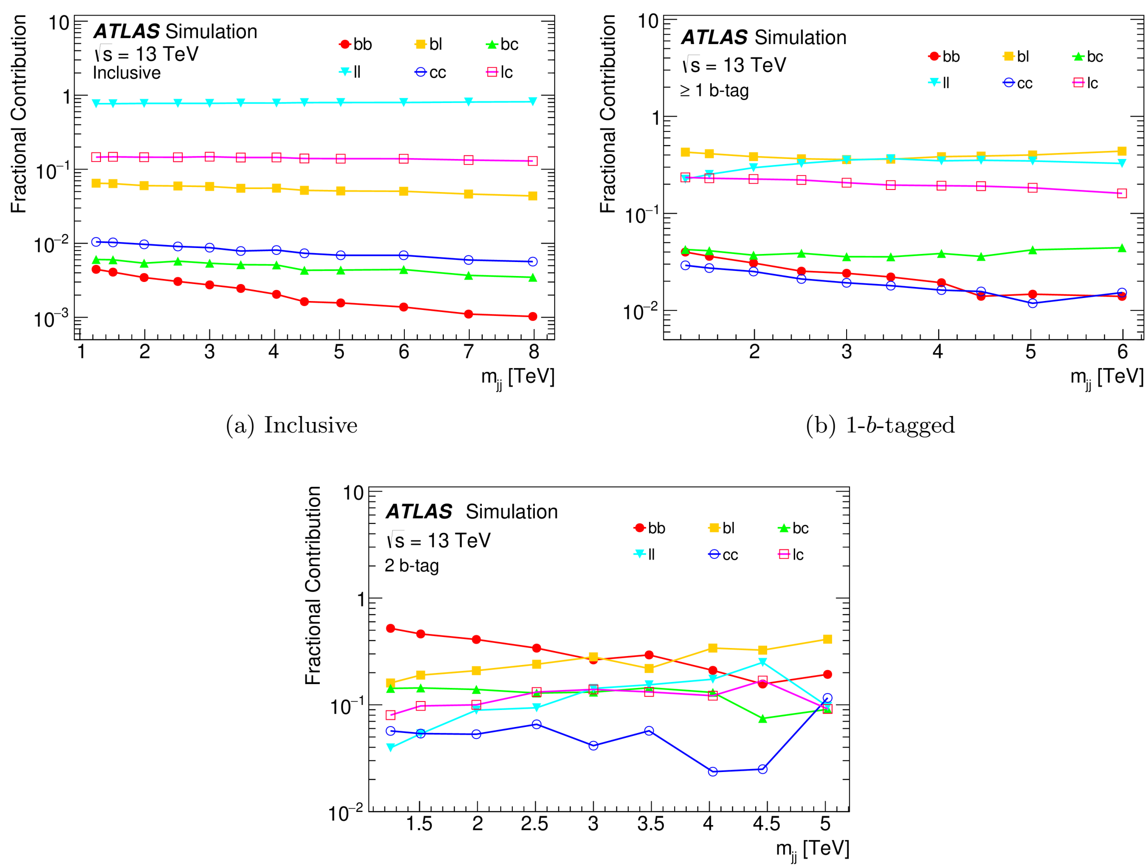


(b) 1-b-tagged



(c) 2-b-tagged

Figure 6.6: The 95 % upper limits on the cross-section times kinematic acceptance times branching ratio for resonances with a generic Gaussian shape, as a function of the Gaussian mean mass m_x in the different categories. For the 1b and 2b categories, also the b-tagging efficiencies are included. Gaussian shapes with different widths are considered.



(a) Inclusive

(b) 1-b-tagged

(c) 2-b-tagged

Figure 6.7: Fractional contributions of the different flavours of simulated jet pairs in the MC background sample in the different categories, as a function of the dijet invariant mass. Looking at (b) and (c), it can be seen that a non null fraction of pairs that should not pass the b-tagged requirement is still present in the samples.

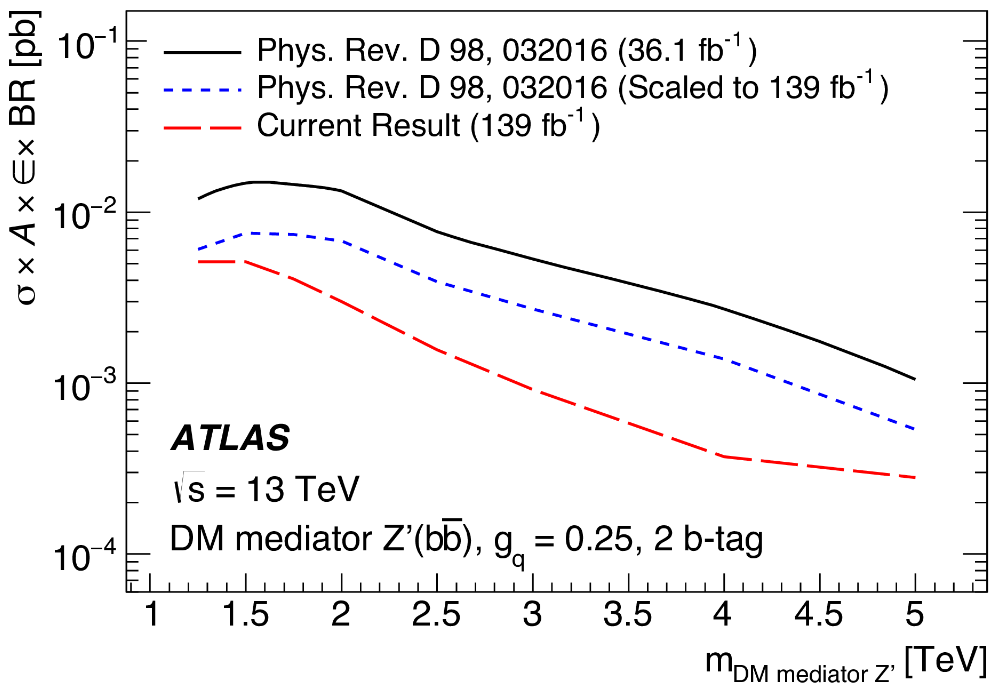


Figure 6.8: The expected 95 % CL upper limits on the cross-section times acceptance times b-tagging efficiency times branching ratio as a function of the DM mediator Z' mass for the current (red line) analysis and the previous one (black line). The upper limit of the previous result is also shown scaled to the 139 fb^{-1} integrated luminosity of the current analysis to show the impact of an improved b-tagging (blue line).

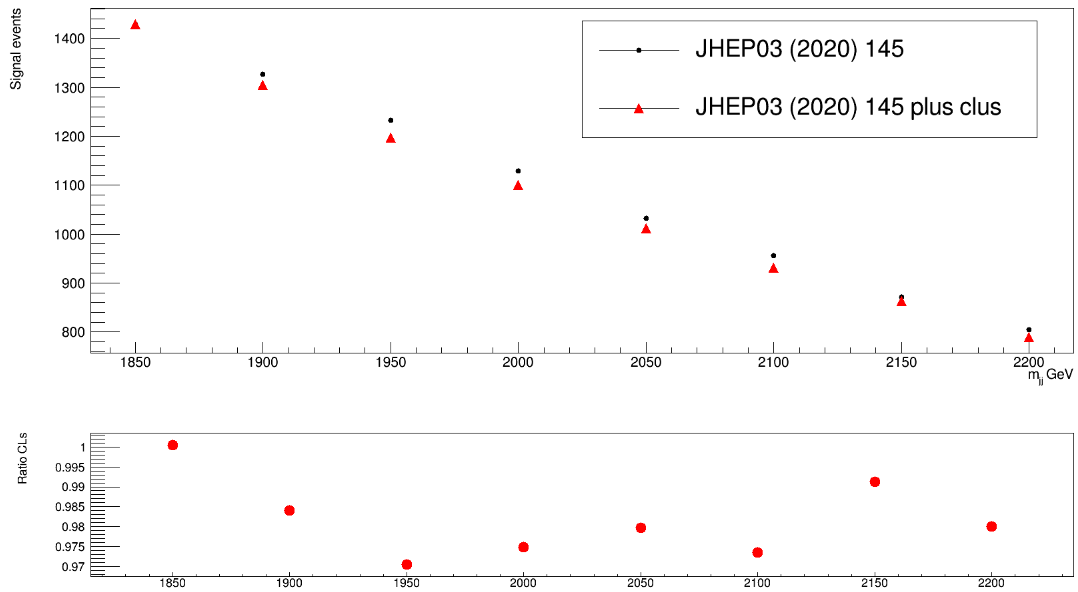


Figure 6.9: Expected upper limits on the number of jet pairs coming from Gaussian shaped signals with a σ of 200 GeV decaying to dijets with at least $1b$ before (JHEP03 (2020) 145) and after (JHEP03 (2020) 145 plus clus) the scaling determined by the improvement in the *light*-flavour jet rejection after the addition of the cluster information.

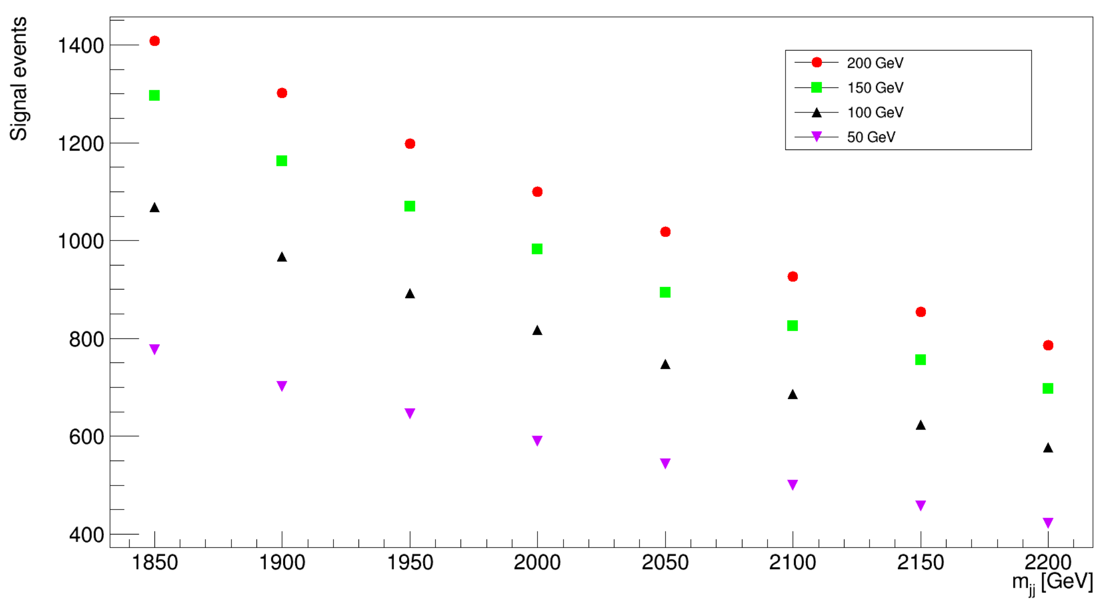


Figure 6.10: Upper limits on the number of jet pairs coming from Gaussian shaped signals in the $1b$ category when different Gaussian widths are considered.

Chapter 7

Further studies

The study presented in this research is the first step in the path of establishing a novel category of algorithms based on the cluster information to be integrated in the ATLAS b-tagging algorithms. While the issues on the deterioration of the b-tagging performance at high transverse momentum were illustrated in Chapter 5 as they are observed in very recent public ATLAS analyses; the improvements given by the addition of the cluster information in Sec. 5.5 were gained on a simplified algorithm that do not exploit the state-of-the-art track and vertex information to their full discriminating power. Therefore a natural further development of the work presented in this thesis is the study of the impact of the cluster information on the performance of the DL1r algorithm; to convince ourselves and the ATLAS community that the improvements gained are worth the effort.

Moreover, in this research we focused on studying the cluster information in the pixel layers only, whereas its benefits should also extend to the SCT. b-hadrons with transverse momenta greater than 2 TeV can penetrate all the pixel layers and decay outside the Pixel detector, causing greater difficulties in track reconstruction and further decrease of the b-tagging performance. In such cases, the jump in the number of clusters is *not* located between two of the pixel layers as explained in Sec. 5.3, but it is expected to be found in the SCT. A similar scenario would be the one in which quarks decaying before the fourth pixel layer were so impulsive that they products would be too collimated to be resolved by the pixels: eventually, such products would separate, and the jump would be located again in the SCTs. Hence, feeding the BDT with the information on the SCT cluster as well is another attempt yet to be made to chek whether the addition could push the improvements gained by the cluster information on the BDT's b-tagging performance to jets of even more extreme momenta.

Lastly, before attempting the difficult task of significantly improving the existing ATLAS algorithms, more complex ways of extracting discriminant information from the clusters could be studied in order to extract more discriminating power from it. We think that the geometrical distribution of the clusters in each jet could contain some of this power still latent. In the next section, I will describe a *topological approach* to refine the cluster information that we are convinced is worth investigating.

7.1 Recipe for a new topological approach

Particles belonging to the hadronisation process of the one which generates the jet are contained in a cone whose typical angular opening depends on the parent particle's mass and transverse momentum through [12]

$$\theta_J \sim \frac{2m_J}{p_T}. \tag{7.1}$$

In very impulsive jets, thus, the particles coming from the "interesting" physical process[1] tend to be very much collimated to the flight direction of the parent, as can be seen in Fig. 5.4. Furthermore,

[1] I.e. the hadronisation shower of the parent particle.

due to the features of the anti-k_T algorithm (Sec. 3.2), the direction of the axis of each jet is more influenced by the direction of flight of the most impulsive particles they contain. For example, since boosted b-hadrons reach energies up to 80% of their total b-jets', their trajectories are very influential in the determination of the jet axis directions. The combination of these two factors is reflected in Fig. 5.9, in which the clusters left by the particles coming from the b-hadron decay are shown to be mainly located in a narrow cone around the jet axis. In the same figure, the angular distribution of the clusters *not coming from the b-hadron* is shown to be evenly distributed in the jet. This fact is not a surprise, since spurious pile up signals, hardware noises and unwanted hits coming from the near-by fragmentation tracks randomly access the jet. The same reasoning can be applied to high-p_T *light*-flavour jets, since the *light* quarks produced in the hard scattering of the partons of the colliding protons are likely to have the highest fraction of the total jet p_T.

Unfortunately, as it was the case in Sec. 5.2 (in which, thanks to the Monte Carlo *truth* information, we were able to explain why the track-based algorithms do not maintain their performance at high-p_T), the possibility to observe the exact locations of the clusters coming from the parent particle decay in the simulation jets (Fig. 7.1) is not present when studying real data. Nevertheless, the MC *truth* can be used to check whether it is possible to identify some features of the real clusters that helps separating those belonging to the physical hadronisation shower (let them be called *signal* clusters from now on) from the spurious noise. Because of the previous reasoning, trying to separate the signal clusters from the noise clusters exploiting their expected greater collimation seems to be worth trying.

With this aim, we study the distribution of the angular separation of each cluster in the layer $n+1$ from each cluster in the layer n for each jet, and we call it dR_{cc}^{subs} (Fig. 7.2a-7.2c), where *subs* stands for subsequent. We look at the dR_{cc}^{same} distribution as well, defined by the angular separation of each cluster in each layer from each cluster in the same layer (Fig. 7.2b-7.2d). We reckon the figures to be promising: they show that the dR_{cc}^{subs} distribution has a peak at low values that the dR_{cc}^{same} has not. Since the geometrical correlations between signal clusters in subsequent layers caused by the characteristic collimation of the hadronisation process are not present when computing dR_{cc} in the same layers, the resulting dR_{cc}^{same} distributions *show only the casual combinatorial distribution that arise from the evenly distributed, uncorrelated clusters*, whereas the correlations clearly stand out in the left peak of dR_{cc}^{subs}.

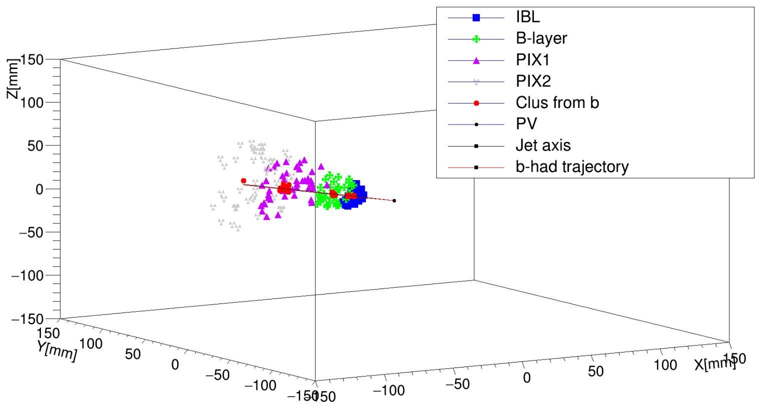
If we plot the dR_{cc}^{subs} distribution of b-jets only for the clusters coming from the b-hadron's shower, though, we realize that the separation between the two peaks is not perfect: a long tail is observed that reaches the right peak's region (Fig. 7.3).

Nevertheless, if we count the number of clusters coming from the b-hadron over the total number of clusters in Fig. 7.2a, we see that the fraction goes from 0.19 when all the clusters are considered, to 0.23 when we consider only the clusters that enter the $dR_{cc} < 0.05$ region, to 0.37 when we look at the clusters that enter the $dR_{cc} < 0.015$[2]. This observation indicates that cuts in dR_{cc} can in principle increase the density of the signal clusters in the *cleaned* sample of those that pass them. Hence, a n-clus-*clean* variable could be studied by using it in a BDT similar to that of Sec. 5.5: since the noise clusters are expected to be similarly distributed in *light*- and b-jets, but the signal clusters resent of the differences in the hadronisations of b- and *light*-quarks, the discrimination power of the cluster information should be enhanced by the cleaning procedure.

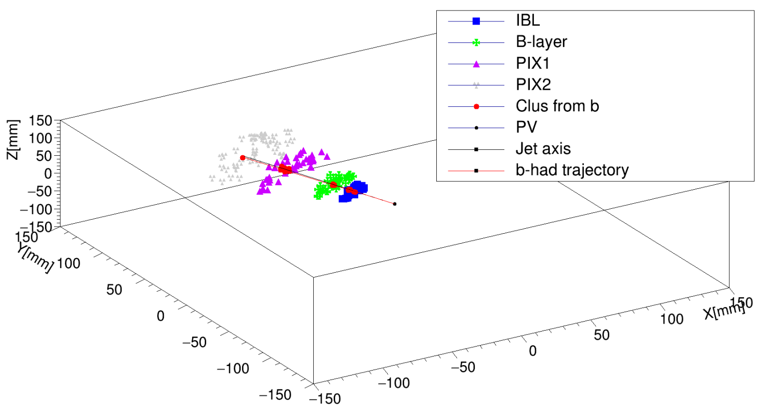
Looking even further, the discriminating power hidden in the geometrical distribution of the clusters in jets of different flavours could be further inspected. The recipe I have sketched in this section ultimately reduces to *counting* the number of clusters in each layer after the new cleaning procedure is applied. The geometry of the cluster distribution could also be *directly* used by feeding a Neural Network with the full set of clusters, each one with its position, letting the machine learn the geometrical correlations by itself. Similar approaches are successfully implemented for the track information by IPRNN and DIPS (Sec. 4.3.2). It is difficult to tell beforehand whether explicitly feeding the network with some dR_{cc}-like information would be beneficial, or if it would learn the correlation independently

[2] To compute this fraction, we clean any cluster that does not enter the dR_{cc} region below the cut in any of all the possibles couples that it forms with the clusters of the preceding layer

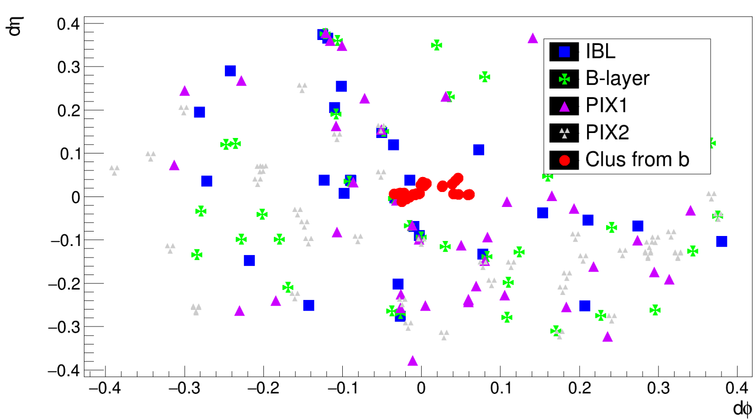
from the cluster positions: the answer to this question will be another small step in the path towards the future cluster-based b-tagging in ATLAS.



(a)

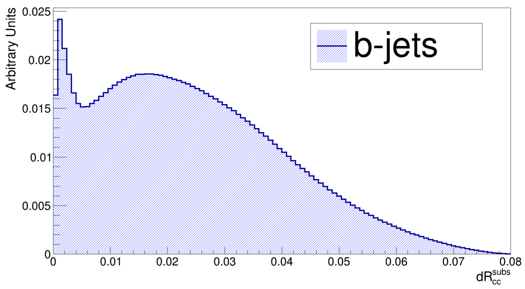


(b)

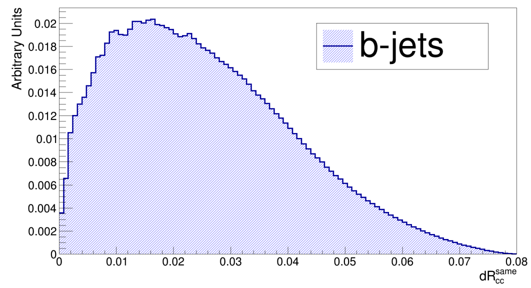


(c)

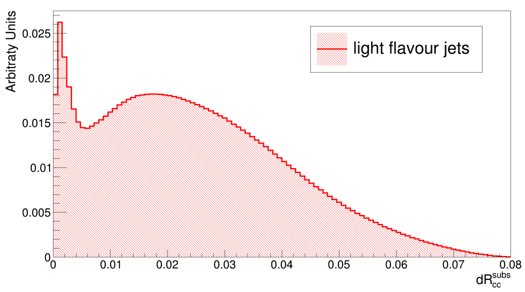
Figure 7.1: Visualisation of the spatial distribution of the clusters of a simulated b-jet. (a) and (b) show the global coordinates of each cluster: those coming from the b-hadron are close to the jet axis; and the trajectory of the b-hadron is very collimated to the jet axis as well. (c) shows the angular separation of the clusters on the different layers from the jet axis: the clusters coming from the b-hadron agglomerate around 0.



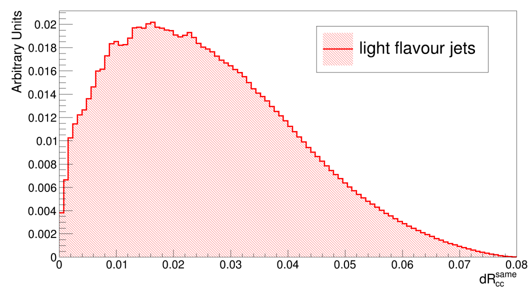
(a)



(b)

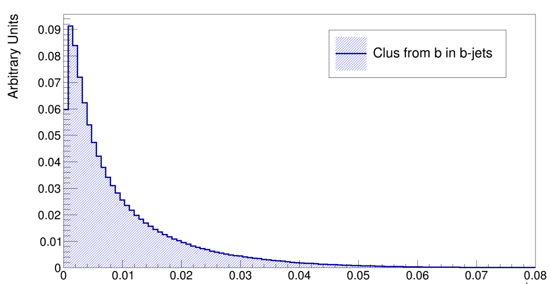


(c)

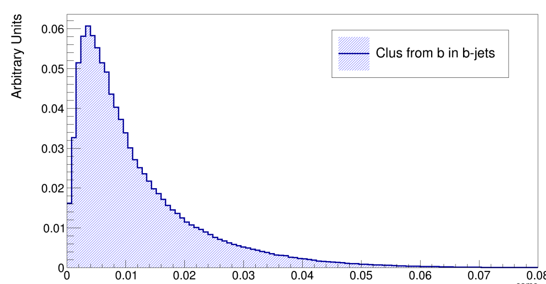


(d)

Figure 7.2: dR_{cc}^{subs} (a) and dR_{cc}^{same} (b) distributions for b-jets (c) and *light*-flavour jets (d).



(a)



(b)

Figure 7.3: dR_{cc}^{subs} (a) and dR_{cc}^{same} (b) distributions of the clusters coming from the b-hadron shower.

Conclusion

The problem of the deterioration of the track information in highly-impulsive jets, due to the increasing difficulties of the tracking algorithms to deal with dense jet environments and late decays of b-hadrons, strongly affects the performance of the state-of-the-art b-tagging algorithms used by the ATLAS collaboration.

To face this issue, new techniques are being studied to enhance the track reconstruction capabilities and to use the available track information in a more efficient way.

Another way of attacking the problem is to design new algorithms that make use of other types of inputs to perform the tagging task. In this research work, we have shown how the information on the number of clusters per pixel layer in the jet contains discriminating power at high impulse, allowing to write a new chapter in the development of flavour tagging algorithms.

Our results extend and consolidate the first observations performed in previous studies ([69], [70]). In particular we have demonstrated that this detector-level information can add discriminant power to that already provided by taggers based on track impact parameters (IP3D). To do so, we built a Boosted Decision Tree fed with the IP3D output and the cluster information and we tested its performance over a Monte Carlo sample that loyally mimics the ATLAS Inner Detector; showing significant improvements over a range of jet transverse momenta reaching 2.5 TeV. This is an important step forward with respect to previous studies on the impact of the cluster information, which were performed on simplified models of the detector.

We also tried to feed fed the BDT with the additional information on the *jumps* of the number of clusters per layer provoked by the displaced decays of the boosted b-hadrons, a variable known in literature to own discriminating power; and showed that the algorithm does not significantly benefit from this addition. We were pleased by this observation, which proves that our BDT is able to *independently* learn correlations in the data, since the information of the jump on the numbers of clusters in subsequent layers is already intrinsic to the numbers of clusters themselves.

Enhancing the b-tagging performance is not a matter of secondary importance. Many of the most exciting analyses and searches for new physics in the ATLAS detector strongly rely on the capability of identifying b-jets: significant b-tagging improvements would directly translate in an enhanced confidence on the statements we make on the presence of absence of new physics in the observed data, and in more precision in the study of the latest SM processes that have been discovered.

We have estimated improvements of about 3 % on the expected upper limits on the cross section of a gaussian shaped resonance coupling to b-quarks, gained on the results of [50] by the information on the number of cluster per layer in each jet.

These gains have to be considered preliminary, since many steps in the exploitation of cluster information lay ahead of us. We already identified promising procedures to clean the spurious clusters that blur the differences between b and *light*-flavour jets. Following this path will also pave the way to more complex methods of treating the individual clusters to extract the full information hidden in their geometrical distributions which are still to be tried.

If the necessary efforts will be made by the ATLAS collaboration to walk this path, I am firmly convinced that, *per aspera ad astra*, the day will come in which the cluster information will be integrated in the future b-tagging algorithms; helping to push forward the boundaries of our present day knowledge of the fundamental interactions at particle colliders. Shall that day come soon!

Bibliography

[1] Daniel Abercrombie et al. "Dark Matter benchmark models for early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum". In: *Physics of the Dark Universe* 27 (Jan. 2020), p. 100371. ISSN: 2212-6864. DOI: 10.1016/j.dark.2019.100371. URL: http://dx.doi.org/10.1016/j.dark.2019.100371.

[2] R.J. Adler, B. Casey, and O.C. Jacob. "Vacuum catastrophe: An Elementary exposition of the cosmological constant problem". In: *Am. J. Phys.* 63 (1995), pp. 620–626. DOI: 10.1119/1.17850.

[3] P.A. Zyla et al. *b-quark mass*. URL: https://pdglive.lbl.gov/DataBlock.action?node=Q005M.

[4] M Aleksa and M Diemoz. *Discussion on the electromagnetic calorimeters of ATLAS and CMS*. Tech. rep. ATL-LARG-PROC-2013-002. Geneva: CERN, May 2013. URL: https://cds.cern.ch/record/1547314.

[5] ALICE. "ALICE: Physics performance report, volume I". In: *J. Phys. G* 30 (2004). Ed. by F Carminati et al., pp. 1517–1763. DOI: 10.1088/0954-3899/30/11/001.

[6] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *Journal of High Energy Physics* 2014.7 (July 2014). ISSN: 1029-8479. DOI: 10.1007/jhep07(2014)079. URL: http://dx.doi.org/10.1007/JHEP07(2014)079.

[7] Johan Alwall, Philip C. Schuster, and Natalia Toro. "Simplified models for a first characterization of new physics at the LHC". In: *Physical Review D* 79.7 (Apr. 2009). ISSN: 1550-2368. DOI: 10.1103/physrevd.79.075020. URL: http://dx.doi.org/10.1103/PhysRevD.79.075020.

[8] Ryan Atkin. "Review of jet reconstruction algorithms". In: *Journal of Physics: Conference Series* 645 (Oct. 2015), p. 012008. DOI: 10.1088/1742-6596/645/1/012008. URL: https://doi.org/10.1088%2F1742-6596%2F645%2F1%2F012008.

[9] "ATLAS pixel detector electronics and sensors". In: *JINST* 3 (2008), P07007. DOI: 10.1088/1748-0221/3/07/P07007. URL: http://cds.cern.ch/record/1119279.

[10] J. Katharina Behr et al. "Boosting Higgs pair production in the $b\bar{b}b\bar{b}$ final state with multivariate techniques". In: *The European Physical Journal C* 76.7 (July 2016). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-4215-5. URL: http://dx.doi.org/10.1140/epjc/s10052-016-4215-5.

[11] Fady Bishara, Roberto Contino, and Juan Rojo. "Higgs pair production in vector-boson fusion at the LHCand beyond." In: *Eur. Phys. J. C* 77.arXiv:1611.03860. 7 (Nov. 2016). Comments: 41 pages, 17 figures, 10 tables, 481. 42 p. DOI: 10.1140/epjc/s10052-017-5037-9. URL: http://cds.cern.ch/record/2232873.

[12] Shikma Bressler et al. "Hadronic calorimeter shower size: Challenges and opportunities for jet substructure in the superboosted regime". In: *Physics Letters B* 756 (2016), pp. 137–141. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2016.02.068. URL: http://www.sciencedirect.com/science/article/pii/S0370269316001647.

[13] R. Brun and F. Rademakers. "ROOT: An object oriented data analysis framework". In: *Nucl. Instrum. Meth. A* 389 (1997). Ed. by M. Werlen and D. Perret-Gallix, pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.

[14] Oliver Brüning and Paul Collier. "Building a behemoth". In: (2007). DOI: 10.1038/nature06077. URL: https://doi.org/10.1038/nature06077.

[15] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[16] CDF and D0 Collaborations. "Evidence for a Particle Produced in Association with Weak Bosons and Decaying to a Bottom-Antibottom Quark Pair in Higgs Boson Searches at the Tevatron". In: *Physical Review Letters* 109.7 (Aug. 2012). ISSN: 1079-7114. DOI: 10.1103/physrevlett.109.071804. URL: http://dx.doi.org/10.1103/PhysRevLett.109.071804.

[17] CERN. *ATLAS*. URL: https://home.cern/science/experiments/atlas.

[18] CERN. *Calorimeter*. URL: https://atlas.cern/discover/detector/calorimeter.

[19] CERN. *CERN releases analysis of LHC incident*. URL: https://home.cern/news/press-release/cern/cern-releases-analysis-lhc-incident.

[20] CERN. *DarkMatter*. URL: https://home.cern/science/physics/dark-matter.

[21] CERN. *Linear Accelerator 2*. URL: https://home.cern/science/accelerators/linear-accelerator-2.

[22] CERN. *Linear Accelerator 4*. URL: https://home.cern/science/accelerators/linear-accelerator-4.

[23] CERN. *Proton Synchrotron*. URL: https://home.cern/science/accelerators/proton-synchrotron.

[24] CERN. *Proton Synchrotron Booster*. URL: https://home.cern/science/accelerators/proton-synchrotron-booster.

[25] CERN. *Super Proton Synchrotron*. URL: https://home.cern/science/accelerators/super-proton-synchrotron.

[26] CERN. *The Large Hadron Collider*. URL: https://home.cern/science/accelerators/large-hadron-collider.

[27] Georgios Choudalakis. *On hypothesis testing, trials factor, hypertests and the BumpHunter*. 2011. arXiv: 1101.0390 [physics.data-an].

[28] ATLAS Collaboration. "A neural network clustering algorithm for the ATLAS silicon pixel detector". In: *Journal of Instrumentation* 9 (June 2014). DOI: 10.1088/1748-0221/9/09/P09009.

[29] ATLAS Collaboration. *ATLAS calorimeter performance: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1996. URL: http://cds.cern.ch/record/331059.

[30] ATLAS Collaboration. *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331063.

[31] ATLAS Collaboration. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19. Sept. 2010. URL: http://cds.cern.ch/record/1291633.

[32] ATLAS Collaboration. *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1996. URL: https://cds.cern.ch/record/331061.

[33] ATLAS Collaboration. *ATLAS magnet system: Technical Design Report, 1*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/338080.

[34] ATLAS Collaboration. *ATLAS tile calorimeter: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1996. URL: https://cds.cern.ch/record/331062.

[35] ATLAS Collaboration. “ATLAS: letter of intent for a general-purpose pp experiment at the large hadron collider at CERN”. In: (1992). URL: http://cds.cern.ch/record/291061.

[36] ATLAS Collaboration. *Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS*. Tech. rep. ATL-PHYS-PUB-2020-014. Geneva: CERN, May 2020. URL: http://cds.cern.ch/record/2718948.

[37] ATLAS Collaboration. *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*. Tech. rep. ATL-PHYS-PUB-2017-003. Geneva: CERN, Mar. 2017. URL: https://cds.cern.ch/record/2255226.

[38] ATLAS Collaboration. *Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector*. 2020. arXiv: 2007.02873 [hep-ex].

[39] ATLAS Collaboration. “Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector”. In: *Physics Letters B* 786 (Nov. 2018), pp. 59–86. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2018.09.013. URL: http://dx.doi.org/10.1016/j.physletb.2018.09.013.

[40] ATLAS Collaboration. *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. Tech. rep. ATL-PHYS-PUB-2017-013. Geneva: CERN, July 2017. URL: https://cds.cern.ch/record/2273281.

[41] ATLAS Collaboration. *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*. Tech. rep. ATL-PHYS-PUB-2016-012. Geneva: CERN, June 2016. URL: https://cds.cern.ch/record/2160731.

[42] ATLAS Collaboration. *Secondary vertex finding for jet flavour identification with the ATLAS detector*. Tech. rep. ATL-PHYS-PUB-2017-011. Geneva: CERN, June 2017. URL: https://cds.cern.ch/record/2270366.

[43] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08003. 437 p. DOI: 10.1088/1748-0221/3/08/S08003. URL: https://cds.cern.ch/record/1129811.

[44] ATLAS Collaboration. *The Silicon Microstrip Sensors of the ATLAS SemiConductor Tracker*. Tech. rep. ATL-INDET-PUB-2007-007. ATL-COM-INDET-2007-008. CERN-ATL-COM-INDET-2007-008. 1. Geneva: CERN, Mar. 2007. DOI: 10.1016/j.nima.2007.04.157. URL: https://cds.cern.ch/record/1019885.

[45] ATLAS Collaboration. *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC*. Tech. rep. ATL-PHYS-PUB-2018-025. Geneva: CERN, Oct. 2018. URL: https://cds.cern.ch/record/2645405.

[46] ATLAS collaboration. “ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 79.11 (2019), p. 970. DOI: 10.1140/epjc/s10052-019-7450-8. arXiv: 1907.05120 [hep-ex].

[47] ATLAS collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[48] ATLAS collaboration. “Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC”. In: *Eur. Phys. J. C* 77.5 (2017), p. 332. DOI: 10.1140/epjc/s10052-017-4887-5. arXiv: 1611.10235 [physics.ins-det].

[49] ATLAS collaboration. “Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s}$=13 TeV with the ATLAS detector”. In: *Physical Review D* 96.5 (Sept. 2017). ISSN: 2470-0029. DOI: 10.1103/physrevd.96.052004. URL: http://dx.doi.org/10.1103/PhysRevD.96.052004.

[50] ATLAS collaboration. “Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *JHEP* 03 (2020), p. 145. DOI: 10.1007/JHEP03(2020)145. arXiv: 1910.08447 [hep-ex].

[51] ATLAS collaboration. "Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector." In: *HEPData* (). DOI: 10.17182/hepdata.91126.

[52] ATLAS collaboration. "The ATLAS Inner Detector commissioning and calibration". In: *Eur. Phys. J. C* 70 (2010), pp. 787–821. DOI: 10.1140/epjc/s10052-010-1366-7. arXiv: 1004.5293 [physics.ins-det].

[53] ATLAS TRT Collaboration. "The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance". In: *JINST* 3 (2008), P02013. DOI: 10.1088/1748-0221/3/02/P02013. URL: http://cds.cern.ch/record/1094549.

[54] CMS Collaboration. "CMS Physics Technical Design Report, Volume II: Physics Performance". In: *Journal of Physics G Nuclear and Particle Physics* 34 (Jan. 2007), p. 995.

[55] CMS Collaboration. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[56] Geant4 Collaboration. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: http://www.sciencedirect.com/science/article/pii/S0168900203013688.

[57] LHCb collaboration. "Observation of $J/\psi\phi$ structures consistent with exotic states from amplitude analysis of $B^+ \to J/\psi\phi K^+$ decays". In: *Phys. Rev. Lett.* 118.2 (2017), p. 022003. DOI: 10.1103/PhysRevLett.118.022003. arXiv: 1606.07895 [hep-ex].

[58] LHCb collaboration. "Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$". In: *Phys. Rev. Lett.* 122.22 (2019), p. 222001. DOI: 10.1103/PhysRevLett.122.222001. arXiv: 1904.03947 [hep-ex].

[59] LHCb collaboration. "Observation of CP violation in charm decays at LHCb". In: *54th Rencontres de Moriond on Electroweak Interactions and Unified Theories*. 2019, pp. 263–270. arXiv: 1905.05428 [hep-ex].

[60] LHCb collaboration. "Observation of structure in the J/ψ-pair mass spectrum". In: (June 2020). arXiv: 2006.16957 [hep-ex].

[61] LHCb collaboration. "Observation of the resonant character of the $Z(4430)^-$ state". In: *Phys. Rev. Lett.* 112.22 (2014), p. 222002. DOI: 10.1103/PhysRevLett.112.222002. arXiv: 1404.1903 [hep-ex].

[62] LHCb collaboration. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[63] Fermilab. *Fermilab Experiment Discovers New Particle Upsilon*. URL: https://history.fnal.gov/botqrk.html.

[64] Yoav Freund and Robert E. Schapire. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *J. Comput. Syst. Sci.* 55.1 (1997), pp. 119–139. DOI: 10.1006/jcss.1997.1504.

[65] Maxime Gouzevitch et al. "Scale-invariant resonance tagging in multijet events and new physics in Higgs pair production". In: *Journal of High Energy Physics* 2013.7 (July 2013). ISSN: 1029-8479. DOI: 10.1007/jhep07(2013)148. URL: http://dx.doi.org/10.1007/JHEP07(2013)148.

[66] Trevor Hastie. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Jan. 2009. ISBN: 9780387848570. DOI: 10.1007/978-0-387-84858-7.

[67] S. W. Herb et al. "Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions". In: *Phys. Rev. Lett.* 39 (5 Aug. 1977), pp. 252–255. DOI: 10.1103/PhysRevLett.39.252. URL: https://link.aps.org/doi/10.1103/PhysRevLett.39.252.

[68] Andreas Hocker et al. “TMVA - Toolkit for Multivariate Data Analysis”. In: (Mar. 2007). arXiv: physics/0703039.

[69] B Todd Huffman, Charles Jackson, and Jeff Tseng. “Tagging b-quarks at extreme energies without tracks”. In: *Journal of Physics G: Nuclear and Particle Physics* 43.8 (July 2016), p. 085001. DOI: 10.1088/0954-3899/43/8/085001. URL: https://doi.org/10.1088%2F0954-3899%2F43%2F8%2F085001.

[70] B. Todd Huffman, Thomas Russell, and Jeff Tseng. “Tagging b quarks without tracks using an Artificial Neural Network algorithm”. In: (Jan. 2017). arXiv: 1701.06832 [hep-ex].

[71] Pier Paolo Ippolito. *Hyperparameters Optimization.* URL: https://towardsdatascience.com/hyperparameters-optimization-526348bb8e2d.

[72] Makoto Kobayashi and Toshihide Maskawa. “CP Violation in the Renormalizable Theory of Weak Interaction”. In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[73] Paul Langacker. “The Physics of Heavy Z' Gauge Bosons”. In: *Rev. Mod. Phys.* 81 (2009), pp. 1199–1228. DOI: 10.1103/RevModPhys.81.1199. arXiv: 0801.1345 [hep-ph].

[74] D.J. Lange. “The EvtGen particle decay simulation package”. In: *Nucl. Instrum. Meth. A* 462 (2001). Ed. by S. Erhan, P. Schlein, and Y. Rozen, pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.

[75] Sandro Palestini. “The muon spectrometer of the ATLAS experiment”. In: *Nuclear Physics B - Proceedings Supplements* 125 (2003). Innovative Particle and Radiation Detectors, pp. 337–345. ISSN: 0920-5632. DOI: https://doi.org/10.1016/S0920-5632(03)91013-9. URL: http://www.sciencedirect.com/science/article/pii/S0920563203910139.

[76] A L Read. “Presentation of search results: theCLstechnique”. In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (Sept. 2002), pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313. URL: https://doi.org/10.1088/2F0954-3899.

[77] Andreas Salzburger. “Track Simulation and Reconstruction in the ATLAS experiment”. PhD thesis. Innsbruck U., 2008.

[78] Ethan Siegel. *Five Years After The Higgs, What Else Has The LHC Found?* URL: https://www.forbes.com/sites/startswithabang/2018/04/11/five-years-after-the-higgs-what-else-has-the-lhc-found/#1d2dcf9f552c.

[79] A.M. Sirunyan et al. “Observation of Higgs Boson Decay to Bottom Quarks”. In: *Physical Review Letters* 121.12 (Sept. 2018). ISSN: 1079-7114. DOI: 10.1103/physrevlett.121.121801. URL: http://dx.doi.org/10.1103/PhysRevLett.121.121801.

[80] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. “A Brief Introduction to PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[81] *Tagging and suppression of pileup jets with the ATLAS detector.* Tech. rep. ATLAS-CONF-2014-018. Geneva: CERN, May 2014. URL: https://cds.cern.ch/record/1700870.

[82] Mark Thomson. *Modern particle physics.* New York: Cambridge University Press, 2013. ISBN: 978-1-107-03426-6.

[83] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[84] Victor Zhou. *Machine learning for Beginners: an Introduction to Neural Networks.* URL: https://victorzhou.com/blog/intro-to-neural-networks/.