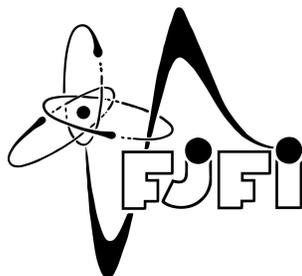


CZECH TECHNICAL UNIVERSITY IN PRAGUE
FACULTY OF NUCLEAR SCIENCE AND PHYSICAL ENGINEERING



RESEARCH REPORT

IMPACT OF MISALIGNMENT ON B-TAGGING AT ATLAS

Monika Panušková

Supervisor: Václav Vrba, CSc.

Acknowledgements

I am grateful to my supervisor Václav Vrba for introducing me to CERN and ATLAS collaboration in particular. Without his continuous support this research project would never come into being. I thank very much for all the help and advices I got from him during this entire work and also for all the support on missions to CERN and to several workshops outside CERN.

I would also like to thank all my colleagues in ATLAS collaboration. My special gratitude belongs to Grant Gorfine and Valentin Sipica for cooperation on the research project and on writing the CSC note. Grant Gorfine has lead the research project in CERN and did troubleshooting in case of any problems.

My big thanks belong to Alexandre Rozanov for motivation to study impact of misalignment in b -tagging, countless advices and also for all the people he introduced me to.

I am grateful also to Laurent Vacavant for leading the ATLAS b -tagging group and exceptionally good cooperation and support from him.

I am deeply indebted to Pavel Jež for countless consultations in which he introduced me to all pitfalls of ATLAS software, particularly to Athena and ROOT. He also shared with me his complete experience and helped with numerous technical and software problems and I cannot miss his contribution to b -tagging as a whole.

I am obliged to thank to the Institute of Physics of the Academy of Sciences of the Czech Republic and particularly to Jiří Popule for the material and organizational support.

My final thank is to Vladislav Šimák for his interest in my work and support on several missions in CERN.

This report was written with kind support of research programmes MŠMT VZ 3407391, 1P04LA212 and LA08032 by the Ministry of Education, Youth and Sports of the Czech Republic.

Declaration

I declare that I wrote my research report independently and exclusively with the use of cited bibliography.

Praha, 4.6.2008

Monika Panušková

Abstract: In the first chapter, there is a brief overview of the principles of the Standard Model. Attention is paid to its experimental validation, especially to the search for the Higgs boson. Most important discovery channels are reviewed there. Second chapter is dedicated to the LHC accelerator and experiments that operate on it. Special focus is on the ATLAS apparatus. The rest of the chapter is devoted to the computing on ATLAS. It includes description of the Athena - the ATLAS official software framework. The third chapter contains introduction to statistics in particle physics. The method for identifying jets from b quarks called b tagging is introduced and explained in detail. Fourth chapter covers author's own research activity: the effect of misalignment on b -tagging performance. Several alignment scenarios are compared in this study.

Key words: Standard Model, ATLAS, b-tagging, misalignment, Athena.

Contents

1	Description of the Elementary Particles	7
1.1	Overview of the Standard Model	7
1.2	Verification of the Standard Model	10
1.3	Summary	16
2	Experimental Instruments in High-Energy Physics	17
2.1	Large Hadron Collider	17
2.2	Experiment ATLAS	21
2.2.1	ATLAS Trigger	24
2.3	ATLAS computing	25
2.3.1	Introduction	25
2.3.2	Athena	27
2.3.3	ROOT	28
3	Introduction to B-tagging	29
3.1	Statistical Methods in Particle Physics	29
3.2	Introduction to B -tagging	33
3.2.1	Overview of B -tagging	33
3.2.2	B -tagging input	34
3.2.3	B -tagging performance estimation	35
3.3	b -tagging algorithms	36
3.3.1	Impact parameter based b -tagging	36
3.3.2	Secondary vertex based b -tagging	38
3.3.3	Other spatial b -tagging algorithms	40
3.3.4	Soft lepton algorithms	41
3.3.5	Combination of tagging algorithms	42
3.3.6	Calibration of b -tagging	42
3.4	B -tagging performance	42

4	Impact of Misalignment on b-tagging	47
4.1	Misalignment	47
4.2	Error Scaling	50
4.3	Impact on Tracking and Vertexing	52
4.4	Impact on B-tagging	54
4.4.1	Light Jet Rejections	54
4.4.2	Effects of Error Scaling	56
4.4.3	Effect of Purification	58
4.4.4	Comparison of the Different Taggers	58
4.5	Recalibration	62
4.6	Summary and Outlook	62
5	Conclusions	65

Chapter 1

Description of the Elementary Particles

1.1 Overview of the Standard Model

Nearly all observed phenomena in the subatomic world can be explained by the Standard Model (SM) - a theory of elementary particles which was formulated in late 1960s (electroweak part) and first half of 1970s (Quantum Chromodynamics (QCD): a theory of strong interaction). This chapter will provide brief overview of the Standard Model, more interested reader may study [1] for the Electroweak theory and [2] for QCD.

Standard Model postulates that there are twelve elementary fermions, falling into three structurally identical families. Each family contains doublet of leptons (one charged - electron, muon or tau - and one neutral: neutrino) and doublet of quarks (up type quark and down type quark). The interactions between fermions are mediated by twelve intermediating bosons. The electromagnetic interaction is mediated by photon, weak interaction is mediated by three massive weak bosons and strong interaction is mediated by eight massless gluons.

These all particles has been found and their properties match these predicted by SM with the extraordinary precision [3]. Apart from those twenty four particles, SM predicts also the existence of a scalar boson which is usually called Higgs boson. This particle should mediate the Spontaneous symmetry breaking and so giving masses to the elementary particles via Higgs Mechanism. The overview of all SM particles is given in tables 1.1 and 1.2.

The Standard Model is a gauge theory. This means, it has some internal symmetry on top of "normal" spacetime symmetry. In case of SM the internal symmetry is $SU(3) \otimes SU(2) \otimes U(1)$. The first term is internal symmetry of QCD, second is the symmetry of weak interaction and $U(1)$ is the symmetry of electromagnetic interaction.

The dynamics of Standard Model is contained in the demand that all gauge symmetries are local. This in turn produces another derivative term which can be added together with normal

		First generation	Second generation	Third generation	Electrical charge
fermions	quarks	u -up	c -charm	t -top	+2/3
		d -down	s -strange	b -bottom	-1/3
	leptons	e -electron	μ -muon	τ -tauon	-1
		ν_e - e -neutrino	ν_μ - μ -neutrino	ν_τ - τ -neutrino	0

Table 1.1: Elementary fermions.

	Interaction mediated	Name	Spin	Electrical charge	Mass	Count
bosons	electroweak	W^\pm	1	± 1	80.4 GeV	2
		Z	1	0	91.2 GeV	1
		γ	1	0	0	1
	strong	g	1	0	0	8
	Higgs	H	0	0	≥ 114.4 GeV	1

Table 1.2: Elementary bosons. Masses from [3].

derivative, thus constructing so called covariant derivative. The added term in covariant derivative is usually interpreted as describing the interaction between elementary fermions. Other interaction terms are originating in the noncommutativity of gauge groups $SU(2)$ and $SU(3)$, resulting in existence of weak boson and gluon selfinteraction.

All interaction vertices predicted by the SM has been found and no others. Their interaction strength was in a good agreement with the experiment.

However, the demand of exact inner symmetry brings serious phenomenological problem. Because the weak interaction violates parity, the left handed and right handed particles transform differently under $SU(2)$. This leads to a fact that a (Dirac) mass terms cannot be present in the theory: they are made from both right-and left-handed particles so insertion of such terms into SM Lagrangian would spoil its symmetry.

The cure to this problem is so-called Higgs Mechanism. It supposes the existence of another scalar field which can give masses to the elementary particles. The main idea is that this scalar field has all symmetries required by the theory, but its (nonzero) ground value does not. This is called "spontaneous symmetry breaking". There is also a theorem saying that for each broken symmetry one inserted scalar particle becomes massless and unphysical: it is so-called Goldstone boson. And at least one scalar field always survives the spontaneous symmetry breaking: the Higgs boson.

In case of SM, we want to break the $SU(2)$ symmetry, because this particular one prevents

the existence of mass terms. It has three generators, so that we need to introduce at least four scalar particles (three would become massless and one will become Higgs boson). After the symmetry breaking we are left with the ordinary gauge $U(1)$ symmetry which allows us to choose a gauge in which all massless Goldstone bosons disappear identically, so we are left with only one new particle.

Due to the original requirement of local gauge symmetry, the Higgs field is automatically coupled to all weakly interacting particles (quarks, leptons and weak bosons) and in this way they can obtain a mass. Important fact is that the strength of the coupling is proportional to the particle mass. Noteworthy exception from this scheme are neutrinos which interacts weakly and are massive, but they cannot obtain mass via Higgs mechanism. If so, we would be able to observe left-handed and right-handed neutrinos in the same amount, but this is not the case.

The fact that only left handed neutrinos were experimentally observed lead the founders of SM to postulate that it is massless because only particles moving with the speed of light can have fixed helicity. Nevertheless, the experiments in 1990s (notable SuperK [4]) showed that neutrinos are moving fast but not with the speed of light, indicating they are massive.

This could be explained in several ways but by far the most widely accepted proposal is the existence of super massive yet-to-be-observed right-handed Majorana neutrino (i.e. fulfilling the Majorana equation [5]) which is giving mass to the observed left-handed (Majorana) neutrino via Seesaw mechanism [7]. In principle, it says that there is a tiny mixing between right-handed and left-handed neutrinos. It has the property that the more massive right-handed neutrino, the less massive left-handed one. That is why it is named seesaw.

This idea is fully compatible with observed neutrino oscillations and in fact, the seesaw can be validated via neutrino oscillations. Nevertheless, the main input to the experimental verification of seesaw mechanism will be coming from collider experiments like LHC or future ILC, because the neutrino mass term is possible remnant of some higher symmetry between the interactions [8].

The unification of all interactions has been long term dream of physicists and the first models of so-called Grand Unification Theory which unifies electromagnetic, weak and strong interaction, was proposed soon after the SM. It exploited the fact that the symmetry group $SU(3) \otimes SU(2)$ could be nicely fitted into $SU(5)$ and that the strength of the interaction is dependent on the energy scale ("running coupling constant" - property of all renormalized quantum field theories [5,6]). The strong interaction becomes weaker with the energy (asymptotic freedom), while the weak and electromagnetic interaction becomes stronger.

The initial calculation suggested unification of couplings at about 10^{15} GeV. It also predicted interaction between leptons and quarks ultimately leading to statement that proton is unstable. However, precise calculation revealed that the coupling strengths does not unify and measurements showed that proton's lifetime is much larger than the initial GUT suggested.

Nevertheless, the introduction of Supersymmetry (which was designed to solve hierarchy problem) cured the GUT. With SUSY, the interactions unify at about 10^{16} GeV which leads also to a higher prediction for proton lifetime which is no longer in disagreement with the

experiment.

As was noticed, another large issue of SM is so-called Hierarchy problem. Its main aspect is that according to SM, we would need corrections of order 10^{15} to have Higgs mass stable in the perturbation theory. This corrections would have to be different in every order of perturbation theory which would mean that there is some highly improbable "fine tuning".

Because this idea is highly uncomfortable, yet-another symmetry was proposed: the symmetry between bosons and fermions and it was called Supersymmetry. It introduces whole new particle and interaction spectrum, but on the other hand, it brings relations between some SM parameters which are considered free (like Higgs mass). SUSY has not been observed, meaning that it must be broken. However if the supersymmetric particles are not too heavy (of order 1 TeV at most), it can still solve the problems it was designed for.

That is why many particle physicist are interested in SUSY these days and a lot of hope is given into the LHC experiments which should be able to observe SUSY events.

1.2 Verification of the Standard Model

As was outlined in the previous section, the Standard Model is emerging from few basic principles, among them Lorentz invariance, local gauge symmetry and spontaneous symmetry breaking. Therefore, its predictions are principled and unambiguous, so it is hard to tweak SM "a little" to fit experimental results.

Nevertheless, this was not necessary. Since its formulation, SM has withstood all the experimental tests with extraordinary precision - the LEP experiments in 1990s validated the key SM parameters with the precision of few per mille. Notable exception to this is neutrino sector which was found to behave differently from SM predictions (see previous section).

Also, one very important prediction of the SM has not been validated yet - it is the existence of neutral scalar boson: the Higgs boson. This particle was intensively looked for at LEP and searches continues also at Tevatron.

None of these machines was able to produce and identify Higgs boson, the only result is lower limit on Higgs mass (1.1). It is quite probable, that the Higgs has mass close to this limit, because there were several Higgs-like events on LEP just before its closure, but the statistical significance was not high enough to claim a discovery. Also, the theory suggests, that the Higgs mass is rather small.

The Higgs search will be one of the first physics campaigns on LHC. It can be started very early and the Higgs discovery can be claimed even with the 10 fb^{-1} of data (about 1 % of planned amount of LHC data) [11].

Although Higgs couples to all massive particles, the direct production of it in the proton-proton collision is not viable, as the proton mass is too small. The production via some heavy particle is much more efficient. The main production mechanisms of Higgs boson at hadron colliders are summarized in Fig. 1.2.

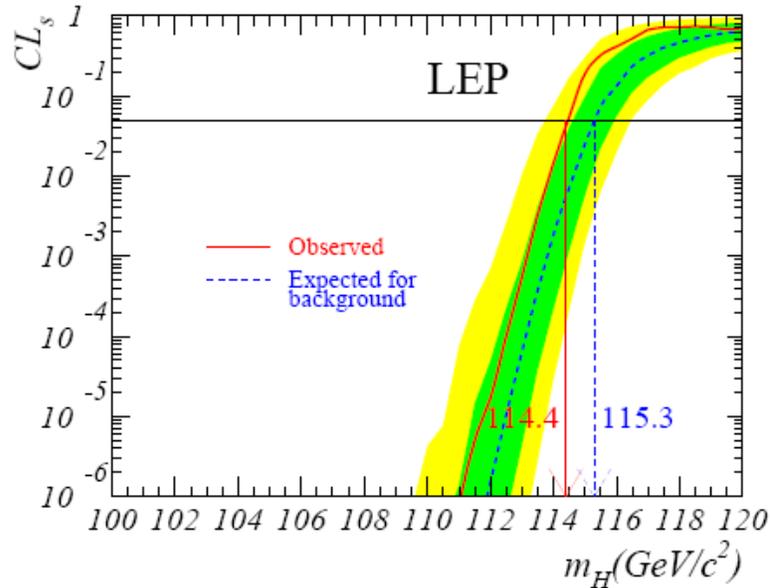


Figure 1.1: The lower limit on Higgs mass from LEP data (all experiments combined). The plot show confidence level (i.e. the probability that we have a Higgs in data, but we did not see it), horizontal line shows 5 %. Solid line is the theory prediction and the dashed line shows actual experimental results, indicating an excess of event in the region of 112 -114 GeV. The green and yellow band are 1σ and 2σ uncertainties, respectively. Figure from [9].

The first process (denoted a in the figure) is called gluon fusion and Higgs is produced via virtual top quark loop. Also other quarks contribute to this process, but because of its mass, the top contribution is by far the highest. The gluon fusion should have the largest cross section of all Higgs production channels and is therefore very important. Tens or hundreds of thousands Higgs bosons (depending on its mass) should be produced in this way over the LHC lifetime.

The second process (denoted b) is called weak boson fusion. The protons will scatter W or Z boson that will fuse and thus create a Higgs boson. This process has rather complicated topology (two very forward jets from the scattered quarks and the Higgs decay products in the center of the detector) making it very distinctive against background and difficult to mimic. Also this channel has only ten times smaller cross-section than the gluon fusion so it will play extremely important role in the LHC Higgs search, possibly becoming the discovery channel.

The third process (c) is called Higgsstrahlung and it was the only possible way of producing Higgs at LEP and it is major production channel at Tevatron. In this process, the Higgs is

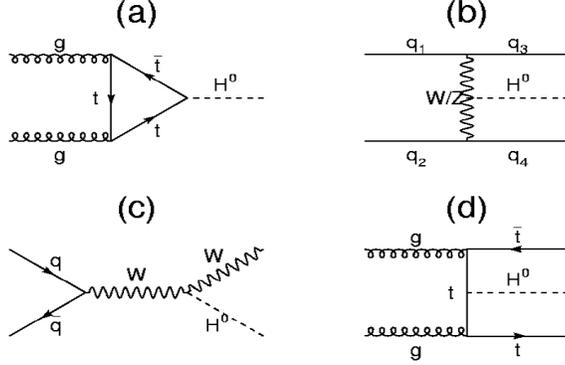


Figure 1.2: Example of four Higgs production channels on hadron collider. It depicts gluon fusion (a), weak boson fusion (b), Higgsstrahlung (c) and top quark associated production (d). Details are given in the text. Figure taken from [12].

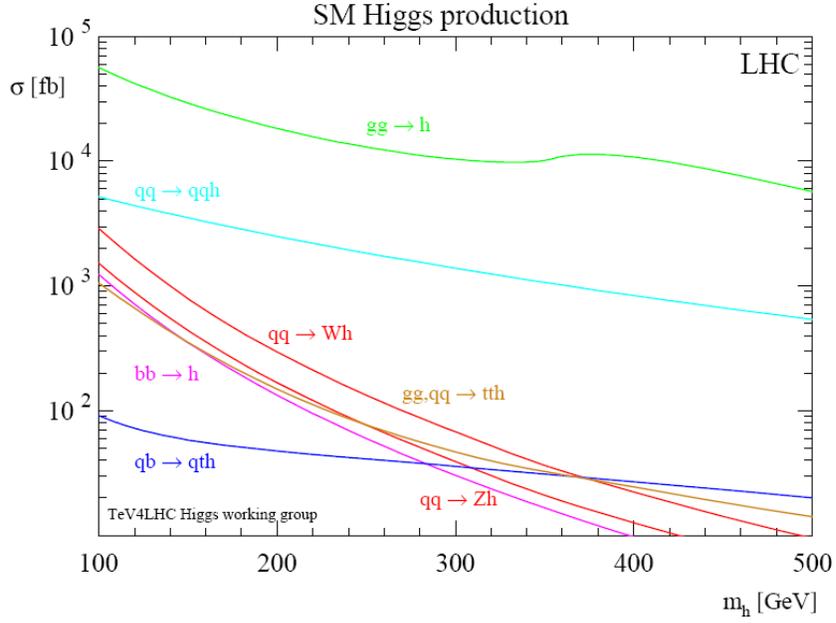
scattered from highly energetic W , formally resembling the Bremsstrahlung from QED. However, the cross-section for this channel is in LHC steeply falling with the Higgs mass. For Higgs over 200 GeV this production is marginal.

The last mentioned process (d) is top quark associated production. Unlike in gluon fusion, the top quarks have enough energy to come real and are produced together with the Higgs boson. Thanks to this the process has also very distinctive signature (three very heavy particles), nevertheless due to extremely high LHC collision energy and luminosity it is quite easy to mimic with process $pp \rightarrow t\bar{t}b\bar{b}$ for light Higgs which decay almost exclusively to $b\bar{b}$. Therefore for light Higgs this channel can never acquire sufficient statistical significance to discover Higgs boson. It may be however useful for heavy Higgs decaying to W pair.

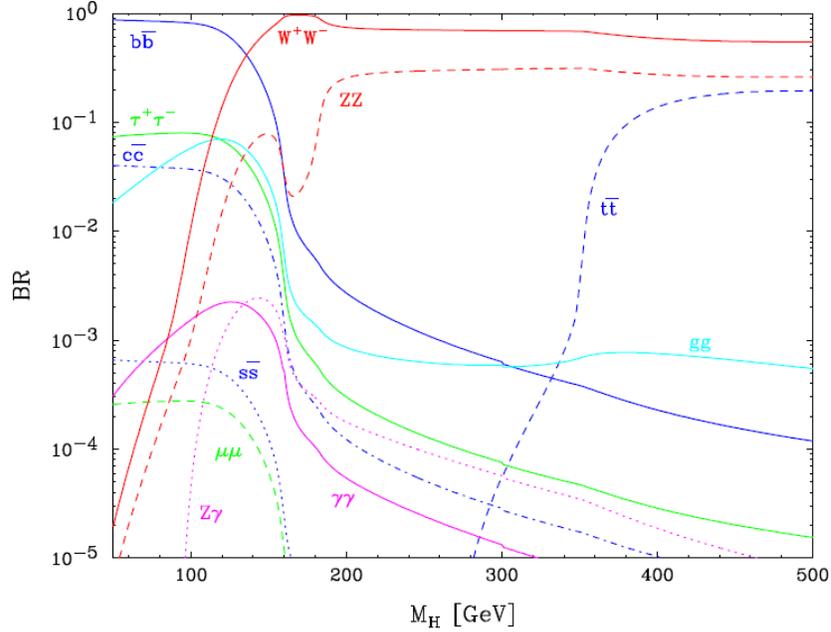
The cross section of all this processes as a function of Higgs mass is on figure 1.3 (a).

Maybe even more important that various Higgs production channels are the way of its decay. Because Higgs is highly unstable particle it can be identified only via its decay products. The figure 1.3 (b) shows the branching ratios of various decay channels. The branching ratio is the proportion of the time the Higgs boson decays in some particular way. We observe, that Higgs couples to the heavies particles available. The light Higgs couples to $b\bar{b}$ and to $\tau^+\tau^-$, while the heavier Higgs boson prefers W or Z pair or even $t\bar{t}$ for the very heavy Higgs.

The most common decay is also least possible to identify. For example, in channel $gg \rightarrow H \rightarrow b\bar{b}$ is the background nine orders of magnitude larger than signal. The similar statements are true also for the other channels. The main reason is that b quarks produced via weak processes from Higgs boson (itself being produced weakly) cannot compete with the direct QCD b quark production as the latter has much larger coupling (and therefore cross section).



(a) Higgs Production



(b) Higgs Decay

Figure 1.3: Cross section of main Higgs production channels at LHC(a). Branching ratio of various Higgs decay channels (b). Both plots are function of a Higgs mass. Figure from [12].

Thus, for light Higgs the only hope is the decay to tau, because leptons are much easier to identify at hadron collider than hadrons. The tau reconstruction is a bit tricky, because it decays quite quickly in either leptonic way (and so we obtain two neutrinos that smear the precision of reconstruction) or in the hadronic way via pions and we obtain very large background.

Rather complicated cuts are employed to make proper selection of tau events and experience shows that the finer criteria the better signal over background ratio. The proper b-tagging is of particular importance here as it can enhance the signal/background several times without losing much efficiency.

For heavier Higgs the situation is much easier as decay to W pair has only electroweak background which has the same order of magnitude as Higgs production. Therefore the heavy Higgs can be identified quite easily, especially in the case of weak boson fusion (Fig. 1.4).

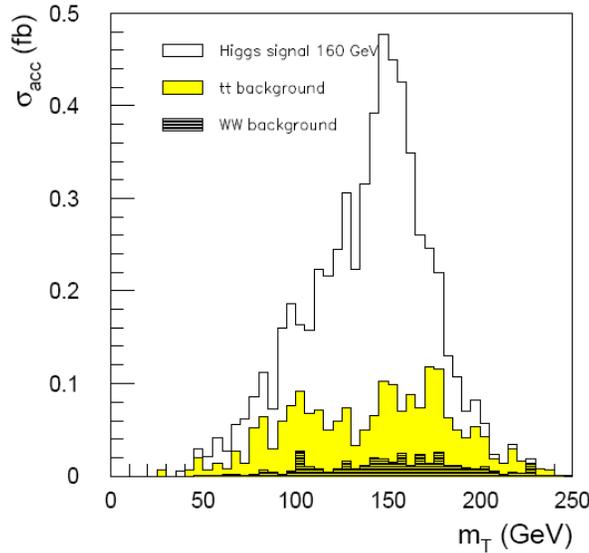


Figure 1.4: Distribution of the signal and background for Higgs production via Weak boson fusion and decay to 2 W bosons. Signal is white, various backgrounds yellow and black. We can see clear separation of signal. Figure taken from [12].

The overview of the statistical significance of Higgs discovery channels for early data at ATLAS is shown in Fig. 1.5. We observe that we would be able to claim a discovery over very large region of Higgs masses.

The mass and charge of the newly discovered particle will be known immediately after the discovery. Nevertheless, to claim that this new particle is really Higgs boson, we must measure also its spin and couplings. The reason is that we must prove that the new particle is scalar

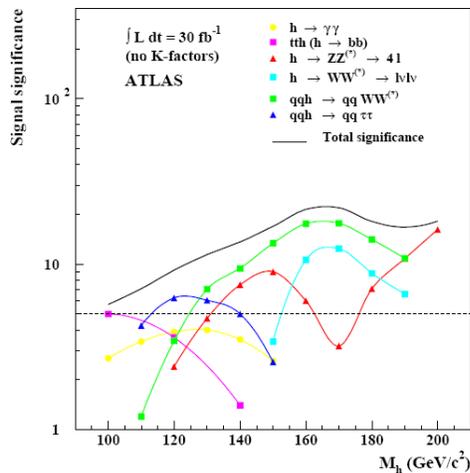


Figure 1.5: The statistical significance for various Higgs channels as a function of its mass. Plot describes the situation for 30 fb^{-1} which is about a tenth of planned LHC data. This particular figure is valid for ATLAS. Figure taken from [12].

boson whose couplings to the other particles are proportional to their mass. This requires very high precision and more data, but dedicated studies show [12] that these measurements will be possible at LHC.

But the Higgs discovery would not be complete without proving the existence of spontaneous symmetry breaking potential. This is equivalent to measuring couplings of 3-Higgs and 4-Higgs interaction. The first can be studied at LHC and although it may not measure the coupling exactly it may quite well exclude the possibility that 3-Higgs coupling is not existing. The coupling of four Higgs bosons is much more difficult to prove, as the observables are much more sensitive to the fluctuation of value of 3-Higgs coupling than to the 4-Higgs coupling. It can very well happen that the difference between some Higgs 4-coupling and no Higgs 4-coupling will be within the error on measurement of Higgs 3-coupling.

The Higgs sector would be also significantly influenced by the SUSY. The minimal SUSY extension of SM (MSSM) predicts that the number of Higgs bosons rises to five (two complex Higgs doublets = 8 fields - 3 Goldstone bosons). Two of them should be SM-like (neutral scalars), one of them should be neutral pseudoscalar and two should be charged. In case of discovering this pattern of particles we can obtain an indirect proof of SUSY.

1.3 Summary

This chapter showed that Standard Model is remarkable simple and economic description of (nearly) all observable particle phenomenology. It gives excellently precise predictions, most of them were experimentally confirmed [9].

Nevertheless one key property of SM has not been proven yet. It is the existence of the Higgs boson which should be responsible for the masses of elementary particles. If the Standard Model is right, the Higgs boson should be observable at LHC.

There are four main production channels which together should yield $10^5 - 10^6$ Higgs bosons during the LHC run, turning it into true Higgs factory. However, not all of them would be identifiable against the background (some decay modes would not be identifiable at all), but still, it is possible to see Higgs boson with necessary statistical significance even in the first 1 % of data.

With the complete data, LHC should determine the Higgs mass with the precision better than 1 % and also measure its other properties which will confirm (or exclude) that it is indeed neutral scalar boson. LHC should also partly prove the spontaneous symmetry breaking. This measurements are however very subtle and require careful analysis with input from sophisticated analysis routines like b-tagging.

Standard model has also several weak spots. The most profound are neutrino masses, the unification of interaction and hierarchy problem. The first of this problems is conventionally solved by the seesaw mechanism, which however is untrivial to prove experimentally. The unification of interactions should happen at a very large energy scale (twelve orders of magnitude higher than accessible at LHC) but one of its effects on a accessible energy scale is already mentioned neutrino mass [8].

Very promising extension of the SM is the Supersymmetry which can solve hierarchy problem and also fix the ordinary GUT by giving protons proper lifetime and exactly unifying the interaction strength. If SUSY exists, there is a good chance that it will be observable at LHC - both directly and indirectly. The main effect of SUSY on SM is in the Higgs sector where the SUSY predicts five physical particles instead of one. At least one of them will be always observable at LHC and if the conditions are good, all five can be observed.

Chapter 2

Experimental Instruments in High-Energy Physics

2.1 Large Hadron Collider

At present¹ the LHC is being finished in CERN, Switzerland. It will be the most powerful hadron collider ever built. It is designed to collide protons with the center-of-mass energy of 14 TeV. It will be also capable of colliding heavy ions - mainly the nuclei of lead (Pb^{82+}). It is the last part of CERN accelerator chain (see Fig. 2.1), so it hugely benefits from the laboratory's practice of linking accelerators together. Protons are created and initially accelerated in older CERN accelerators like PS and SPS, so that they are injected to LHC with quite high energy (450 GeV). LHC then does the rest of acceleration.

The machine is stored in 27 km long underground tunnel which was previously used for LEP. Because it collides particles with the same charge, it has to have two separate pipes, one for each direction, with differently oriented magnetic field. There are 8 regions (called Insertion regions) where there are no magnets and particles run straight sharing the same pipe (See Fig.2.1). At this places protons can collide.

To achieve this high luminosity, the proton beams have to be particularly dense. In fact, there will be 2835 bunches of 10^{11} particles, separated by 25 ns. On the other hand, the higher density causes quicker degradation of a beam by deflection of protons. So that, there is some "critical" density which should not be crossed to maintain high luminosity. Altogether, at LHC this would mean some twenty collisions for every bunch crossing at the full luminosity.

This all requires extremely strong magnetic field which in return requires huge amount of energy. The only way to create such strong fields is using superconductors. Thanks to this, LHC operating temperature is lower than 2K.

¹Summer 2008

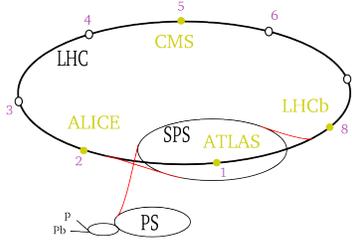


Figure 2.1: The scheme of the CERN accelerator chain together with location of LHC experiments. Figure from [14]

Table 2.1: LHC Statistics

particles used	protons and heavy ions (Pb^{82+})
circumference	26.659 m
injected beam energy	450 GeV (protons)
beam energy at collision	7 TeV
magnetic field at 7 TeV	8.33 T
operating temperature	1.9 K
revolution frequency	11.2455 kHz
power consumption	120 MW

Previous paragraphs outlined that LHC is really a technical and scientific challenge. It has very high research potential and it should open a completely new era in particle physics. Its vital parameters are summarized in Table 2.1.

The LHC could (at least theoretically) produce particles with mass as high as 14 TeV. But the production rate of such a particles will be extremely low and thus unobservable. In general, it is agreed, that the particles with mass up to 1 TeV could be spotted at LHC with reasonable statistics. Even this value is quite impressive - the heaviest known particle has 172 GeV.

The LHC has the capability to reliably explore the region of energies where the Standard Model is distinguishable from the "Beyond the Standard Model Theories" like Supersymmetry. LHC will be also the second machine in the world able to produce top quark. But it will produce much more of them than Tevatron (it is estimated about 10^7 $t\bar{t}$ pairs a year at the luminosity of 10^{33} $\text{cm}^{-2}\text{s}^{-1}$). Thus, this higher statistics will allow more precise measurements of top quark mass and its couplings.

There are many other particles LHC can produce apart from the mentioned ones, but these are the most important. Important aspect of a high-energy physics experiment is the detector.

There are four major experiment at LHC designed to analyze output of collisions:

ATLAS (A Toroidal LHC ApparatuS) is a general purpose detector designed to exploit the full LHC potential. It is being built at Point 1 (see fig. 2.1), directly opposite the CERN main entrance. The view of the whole detector is at Figure 2.2 The project involves collaboration of more than 1800 people from 34 countries. Because of the high collision energy and production rate, the ATLAS is the most complex detector ever built. The main lines of the ATLAS research are:

- The search for the Higgs boson or any other mechanism of the electroweak symmetry breaking
- The investigation of CP violation in B-decays
- The precise measurement of mass of heavy particles like top quark or W boson
- The search for supersymmetric particles or any other new models of physics
- The studies of compositeness of fundamental fermions

To fulfil these goals the ATLAS consists of several components which together provide the full information about the collision. These subdetectors will be described later.

CMS (Compact Muon Solenoid) is also a general purpose detector. The name "compact" means that it is somewhat smaller than ATLAS (about 8 times in volume), but has about twice its weight. It is being built at Point 5 (cf. fig. 2.1) - unlike ATLAS it is being assembled on the surface and lowered to the experimental cavern afterwards. The name also signalizes that CMS is optimized for tracking muons and its magnet will be the largest solenoid ever built, producing a magnetic field of the strength of 4 Tesla. The CMS collaboration involves about 2000 scientists and engineers from 36 countries. The scientific goals of the CMS are similar to that of ATLAS, namely

- The search for origin of the spontaneous symmetry breaking (Higgs boson)
- The search for physics beyond the SM - for example supersymmetric particles
- The study of heavy ion collisions and of the formation of the quark-gluon plasma, thus emulating the very first moments after the Big Bang

Although the construction of two similar detectors may seem as a waste of time and money, it fulfils the natural requirement on experimental physics - that any result should be independently confirmed. Also, combined results from both detectors have reduced systematic as well as random errors.

ALICE (A Large Ion Collider Experiment) is a detector specially designed to study the collisions of heavy ions. Experiments in the CERN in 1990's and in the Brookhaven National

Laboratory, USA, in 2000's showed that at very high temperatures the quarks are free in a state which was called the quark gluon plasma.

The LHC should create the QGP by colliding nuclei of lead with an energy of 5.5 TeV per nucleon. The QGP will be then identified thanks to the specific signatures of leaving particles - for example the production of strange particles and the suppression (compared to what is expected from ordinary theory) of the production of J/ψ mesons (made from charm and anticharm pair of quarks), because the turmoil of QGP prevents forming of heavy quark pairs. This was seen both in CERN and in BNAL, and it is supposed, that on higher energy also the production of bottomium (bound state of b and \bar{b}) will be suppressed from the same reason.

ALICE is constructed at Point 2 and its collaboration involves more than 1000 people from 28 countries.

LHCb (Large Hadron Collider beauty) is an experiment devoted to the measurement of CP violation. It is expected that it could be most clearly seen in the difference between the decay of B_d meson ($d\bar{b}$) to J/ψ ($c\bar{c}$) and K^0 ($d\bar{s}$) and the decay of anti- B_d meson to respective antiparticles. By studying the difference in the decay times, we would be able to determine the complex phase of CKM matrix [1].

This type of experiment has been already tried (among others) at the LEP, SPS or Tevatron. Nevertheless, none of these machines produced enough b quarks to observe such a subtle effect like CP-violation. The LHC is able to produce much more b quarks than previous accelerators, thus hopefully making the observation of CP-violation possible.

The LHCb is located at Point 8. This experiment has nearly 900 participants from 13 countries.

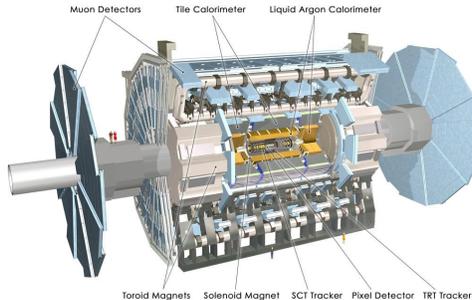


Figure 2.2: The ATLAS detector. The picture was taken from [19].

2.2 Experiment ATLAS

As was outlined in previous chapter, the physics goals at LHC are really ambitious. And its main tool for accomplishing them is the ATLAS detector. It is huge in size and more complex than any other detector. ATLAS is roughly cylindrical in shape and could be divided into three main layers:

- Inner tracking - done by Inner Detector which measures the tracks of all charged particles coming out of the collision.
- Calorimetry - several types of calorimeters measure the energy of particles which passes through them - most of them will be stopped by the calorimeters so we just measure the deponed energy. Note that neutrinos and muons are able to penetrate the calorimeter system without any problem
- Muon spectrometer - whole detector is surrounded by large muon spectrometer (inner radius 5 m, outer radius 10 m) which helps to identify and track muons, which escape from the detector.

Together with detectors, ATLAS also have a magnet system consisting of central solenoid and several air-core toroid magnets on the perimeter. They produce magnetic field with the strength of about 2 T (central solenoid) and 4 T (toroid magnets) which helps to measure momentum of charged particles. Now let's look at the ATLAS subsystems in detail.

Inner Detector

Inner Detector (ID) is the innermost part of the ATLAS detector. It is a cylinder with the length of 7 m and radius of 1.15 m. ID is a very complex device, consisting of tree subsystems: Pixels, SCT (Semiconductor Tracker) and TRT (Transition Radiation Tracker). The first two systems are in principle high-precision semiconductor detectors, the last one is less precise, but, in the other hand gives large number of hits.

In the central area the detector looks like a series of concentric cylinders (so called "barrel" area), further from center (or, equivalently, in the area, where particles have small angle with respect to proton beam) are detectors glued on a disks perpendicular to the beam axis (these disks are on both end of the detectors and are called endcaps). The total pseudorapidity² range covered by the ID is $|\eta| \lesssim 2.5$.

Inner detector subsystems:

²Pseudorapidity is defined as

$$\eta = -\ln \operatorname{tg} \frac{\theta}{2},$$

where θ is azimuthal angle. It is the ultrarelativistic approximation of rapidity.

Pixel detector In the core of the whole ATLAS detector is the Pixel detector. It consists of three layers (at radii 5 cm, 9 cm and 12 cm) in barrel and three disks in the endcaps. The rectangular pixel modules are mounted on a barrel staves so that their longer axis is parallel with the beam axis. On the endcap disks, the modules are arranged like rays coming from the center of the disk. It is good to note that the number of layers cannot be too high - even if we do not care for the price, each additional layer brings more material into the detector volume and so causes more energy to dissipate, influencing the experiment outcome more than it is necessary.

High granularity is absolutely essential for the whole experiment: it is necessary for high impact parameter resolution and for measuring the lifetimes of the heavy particles like B-hadrons. The quality of b-tagging relies mostly on the quality of the data from the pixel detector. Only it can precisely determine the positions of primary and secondary vertices and measure their impact parameters.

To achieve high precision, over 46 000 tiny pixels are on each module. Their size is $50 \times 400 \mu\text{m}$ and their resolution is $14 \times 115 \mu\text{m}$. However, the practical precision is somewhat lower. The reason is, that it is technically impossible to place all modules on the right places, and also, the whole detector degrade in time due to mechanical fatigue. This leads to so-called misalignment, i.e. modules being aligned out of their rightful places. That is why we must do an alignment before any serious physical measurement.

Pixel detector is constructed independently from the other parts of the ID and will be put in place only after the SCT and TRT.

SCT The SCT is constructed similarly to the pixel detector - it has 4 layers in the barrel and 18 disks in the endcaps. The detection of particles is done by silicon microstrip detectors - they are similar in principle to the pixels, but they are longer - meaning that they have worse resolution in longitudinal direction.

The SCT also contributes to precision measurements of tracks and particle momenta and suffer from similar alignment problems as pixels. However, their effect on tracking resolution is not such big as in the case of pixels.

TRT Transition Radiation Tracker is different in construction from the Pixels or SCT. It is based on use of a straw detectors which are aligned parallel with the beam axis. Therefore it has quite good resolution in the plane perpendicular to the beam ($r - \phi$ plane) - about $170 \mu\text{m}$, but much worse in the direction of beam. However, there are quite a lot of this straws, so that TRT produces more than 30 hits for each track passing through its volume.

It is composed from 420 000 straw detectors (100 000 in barrel and 320 000 in the endcaps). Electron identification capability is enhanced by Xenon gas which is set between and inside the straws. Each particle create a different number of transition radiation photons, so that the TRT can distinguish between them.

Table 2.2: Inner Detector subsystems parameters

System	Element size	Resolution	η coverage
Pixels	$50 \times 400 \mu\text{m}$	$\sigma_{R\phi} = 14 \mu\text{m}$ $\sigma_z = 87 \mu\text{m}$ $\sigma_R = 87 \mu\text{m}$	± 2.5
SCT	75 or $112.5 \mu\text{m} \times 12 \text{ cm}$	$\sigma_{R\phi} = 15 \mu\text{m}$ $\sigma_z = 770 \mu\text{m}$	± 2.5
TRT	4 mm diameter 150 cm long	$\sigma_{R\phi} = 170 \mu\text{m}$ per straw	± 2.5

Important facts about ID subsystem are summarized in Table 2.2.

ATLAS calorimeters

After leaving the Inner Detector, particles come into the calorimeters. This device is kind of opposite of the ID - its purpose is to cause particles to dissipate as much energy as possible and do not leave the calorimeter. The dissipated energy is converted into electrical current and then measured. This information is essential in determining the event outcome.

ATLAS calorimeter is composed from several calorimeters - Liquid Argon Calorimeter and Tile Calorimeter.

Liquid Argon Calorimeter is next to the Inner Detector. This calorimeter consists from both electromagnetic and hadronic calorimeters. It is made from absorber plates (made from steel or lead) and filled with liquid Argon. The electrons and photons lose their energy on that plates and decay into showers of less energetic particles. These ionize the molecules of liquid Argon and the resulting electric potential is measured. The same principle is also applied to the measurement of the energy of hadrons. The calorimeter itself has also some degree of granularity, and so it can give us the information not only about the energy but also about the position and direction of the particle.

Tile Calorimeter is the outer one of ATLAS calorimeters. Barrel part is made from 64 wedges, each 5.6 m high and 20 tonnes in mass. Each endcap is also made from 64 wedges (which are however a bit smaller than in the barrel). Tile calorimeter is made from plastic scintillator plates which emit light when a particle passes through. This light is converted to electric signal and measured.

Muon Spectrometer

The purpose of muon spectrometer is to identify and measure muons. As was said before, they are the only measurable particles able to escape from the detector. Muon Spectrometer allows measurement of their tracks and momenta with high precision. This is very important for ATLAS physics, because usually the most popular events are these with leptonic signature - they are cleaner as there is no QCD background.

In principle it is quite similar to the straw detector in ID - the TRT. The only difference is that the straws are slightly larger.

2.2.1 ATLAS Trigger

The task of ATLAS trigger is to select bunch crossing containing interesting physics and then record them for the analysis. This is very challenging task because the LHC has very high bunch crossing frequency (one bunch every 25 ns - this is shorter time than it takes photon to travel to the edge of the detector) and because every bunch crossing contains approx. 20 interactions.

The trigger has also to be very selective - the data come with the frequency of about 10^9 Hz, while recording to permanent data storage is done at the frequency of 100 Hz. And, also, it has to be able "to find a needle in a haystack" - for example, the Higgs decay to 4 leptons happens only in one out of 10^{13} events.

To achieve this, ATLAS trigger is divided into 3 levels. Level 1 (LVL1) uses the data only from muon spectrometer and calorimeters. It has to decide very quickly whether some event interesting or not. LVL1 defines Regions of Interest (RoI), which are simply the areas of detector where something interesting could have happened (Fig. 2.3). The candidates selected by the LVL1 can be high- p_T leptons, photons, hadrons or jets. The information in RoI usually contains direction of the particle (polar angle ϕ and pseudorapidity η), transverse momentum and the energy sums (i.e. also the information about missing energy, which are usually neutrinos. LVL1 reduces the data flow from some 40 MHz to 10-100 kHz.

Level 2 (LVL2) trigger uses the data from all detectors lying in the Regions of Interest to investigate the event in more detail and decide whether the data from the event will be further processed. It reduces data rate to 0.1 - 1 kHz. And finally the third level finally uses the data from the full detector to finally decide if the event will be chosen for permanent storage and further analysis. This is done with the frequency of 10-100 Hz.

I showed that the trigger has crucial meaning in the ATLAS experiment. It really determines what physics we will actually see, so the trigger has to be set as carefully as possible not to miss anything important and the other hand not to accept a lot of uninteresting background events.

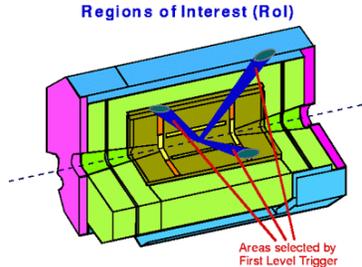


Figure 2.3: Regions of interest in the ATLAS detector. Description of them is in the text. The picture was taken from [19]

2.3 ATLAS computing

2.3.1 Introduction

Average LHC event would be made from several hundreds of particles. And these events will be produced at rate 40 MHz at the full LHC luminosity. It is clear that without computers it would be impossible to do any physics at the LHC. The first part of the computing system was already outlined in the previous chapter - I mean the ATLAS trigger which chooses every second 10-100 interesting events out of 40 millions. It is closely related to the Data Acquisition (DAQ), a system which is responsible that all data will be extracted from the detectors and stored on a long-term storage device (typically a tape). Trigger and Data Acquisition together forms ATLAS online computing (together with DCS: detector control system and other online services). And because it has to be quick, all computers are placed underground, as near the interaction point as possible.

But there is also ATLAS offline computing - it is something which is done in surface laboratories and research institutes which can be half a planet away from the ATLAS. This is that part of the experiment which produces some publishable results. ATLAS offline computing has rich inner structure, so let's examine it in more detail.

Although one may think that there is no work for the offline software when no data are available, the contrary is true. There is lively computing activity at every phase of experiment, including the planning and aftermath. The software can be divided into several groups:

Generators are programs which more or less accurately try to generate some event. In principle, these generators take input from the physical theory which tells them what should happen in the terms of probability. Then they employ some Monte Carlo generator which will produce events with the theoretically predicted probability. So the quality of generator is highly dependent on the quality of our theory.

Simulators are programs which take care of the detector and the surrounding area. They cooperate closely with the generators. In principle, they just add the material of detector into the particle tracks. This influences both the particle, which loses some or all its energy when it meets with the body of detector, and also the detector, which has a "hit". Because of the detector complexity, the simulation is very time-consuming process even on the fastest computer.

Digitization software What we get from the detector in reality is basically some electrical current converted to zeros and ones. So digitization takes the hits from the simulation and turns them into a digital signal. It has to be aware of imperfectness of the detectors (i.e. final resolution, noise, defects, etc.) so that the outcome of the digitization is very similar to what will be produced by ATLAS online software. And here we can see one of the main tasks of the offline computing during the construction of the detector - it prepares our computers to what they can see in the actual experiment. This process is called Computing System Commissioning (CSC) and one of its purpose is to find out whether our present calculational and storage capacity will be enough for the ATLAS experimental data.

Reconstruction From the experiment or digitization we get some form of detector response to a physical event. The purpose of reconstruction is to determine what happened from the detector response the event had induced. So that the algorithm collects the information about the hits in the tracker and about the energy deposited in calorimeters. From the hits it tries to reconstruct track - by some form of interpolation, and it tries to match the track with the energy in calorimeter.

The output of reconstruction tells us something like: "In event number N there was a particle flying with polar angle ϕ and azimuthal angle θ , with an energy E and momentum p . It had charge Q , spin S and mass m ." The amount of stored information of course depends on situation and also on the reconstruction software.

Analysis Analysis is closely related to the reconstruction and sometimes is hard to tell where ends the reconstruction and where starts the analysis. The basic type of physical analysis is (and always has been) the testing of physical hypotheses, i.e. we have some prediction coming from the physical theory and the experiment tests it.

The actual analysis is almost exclusively done by statistics - we extract from the reconstructed data only the important observables and then we do a statistical analysis on them, determining the mean as well as the deviation of measurement.

Now two things may happen - the results are in accord with the theory - this was the case of LEP (the results agreed with the Standard Model up to few per mille) - or there is a disagreement with the prediction. This could mean two things - either we are bad

experimenters (that happened a lot of times during past century), or the theory was wrong - this way an atomic nucleus was discovered.

Where lies the border between measurement error and the new discovery? This is not always clear, but generally accepted is that 5σ statistical deviation from the theory expectation ("background") means new discovery.

Visualization So far we have seen the software which takes a bunch of numbers, do something with them and return another bunch of numbers. Because this is not an ideal input for human mind, there is software to convert data about events to something more convenient, typically some picture.

Basically there are two types of this software. The first one are event viewers, which visualize the event in the detector (for experienced physicist, this is the fastest way how to decide whether something interesting happened or not). This is done for example by Atlantis.

The second way of visualization is to plot the data to make some histograms. They usually have the familiar structure of some flat background with gaussian peaks. This way the physicist loses the contact with the individual events, but on the other hand he acquires a comprehensive view on all interesting data.

2.3.2 Athena

The practically all offline software in ATLAS is implemented in a framework called Athena. It is based on C++ and among others brings common user interface to all offline software. It provides common tool for installing and managing the software called CMT (Configuration Management Tool), it has shared libraries of important objects and tools, so that all collaboration is using the same. Last but not least, it provides common data format, making the analysis of results easier and more transparent. All this encourages common approach and reusability, making it easier for people to join as well as to go from one physics group to another if they like.

Athena consists of the software (i.e. algorithms and libraries) and of the tools which manage this software. The ATLAS offline software is organized in the forms of small packages, each doing some specific task. These packages can be browsed by CVS local repository ([22]) or by LXR [20] - this tool is extremely useful, because it contains very powerful search engine.

Software can be obtained Athena's management tool CMT to "check out" that package - it means downloading its contents from the central repository to the local machine and making links with the other software which the package uses (for example the description of detector) and usually is common for a lot of other packages. Thanks to Athena, all pieces of software are installed in the same way.

Then the user can change the main algorithm of the package to better suits his needs. Because everything is written in C++ using the same objects and tools, he does not have to

learn much new when wanting to work with a new package. When he is satisfied with the code, he compiles it (using the CMT) and runs it - the commands to run a package are also the same for all of them.

However, the modification of the main algorithm is usually not necessary. The code contains dozens of switches which are set independently in so-called Job Options file. This allows the physicist to modify the algorithm to his needs without even knowing the C++, because Job Options are written in much more simple Python language. And because every package contains a lot of already written Job Option files, the user can just pick one of them and change the values. So that even knowledge of Python is not necessary to work with ATLAS software.

Because of the common data format, the output of any package could be used as an input for other packages. There are some restrictions coming from the organization of the software - the output of simulation could be used as input for digitization, but not the other way round.

2.3.3 ROOT

In previous section I mentioned ROOT as an important tool for physics analysis. ROOT, in fact, is a software framework on its own. The reason is that every analysis is different - every physicist needs his own analysis routine.

ROOT framework provide common objects and services (histograms, fitting...) which are used in a statistical analysis. Physicists then use them to make the analysis macro they need. And because of common functionality it could be easily taken by some other user and modified to his needs.

ROOT is based on C++ and could be used in two ways: interactively or in a batch mode. In the first case, the user writes commands in simplified C++ and ROOT interprets them. If the user is not very familiar with C++, he or she can turn on the graphical user interface and do almost everything by clicking on the menus.

For complex tasks it is more convenient to write a short script, rather than repeat the commands in the interactive mode again and again. When user have a script he can run it as many times as he like in a batch mode. The script can even be compiled. After that they are about ten times faster than the interpreted scripts.

ROOT is not a part of Athena, however it is connected with it. For one thing, the Athena uses root files as its input and output data files. Secondly, it is possible to call ROOT directly from Athena and so quickly produce histograms. Also it is possible to use ROOT advanced mathematical tools (in addition to statistics it knows mathematical analysis and linear algebra) in Athena.

Here I showed only some of the ROOT functions. For example it can also draw the Feynman diagrams or make 3D models of ATLAS detector. Comprehensive documentation on all functions (as well as many tutorials and examples) are on ROOT home page [21].

Chapter 3

Introduction to B-tagging

3.1 Statistical Methods in Particle Physics

When we do some measurement in particle physics, we can never obtain one sharp value, but the result is rather a set of numbers following some statistical distribution. Reason for this is that our detector is imperfect (finite resolution, dead regions, faulty electronics), the data taking and reconstruction software has *always* some bugs and that the particle physics has probability character.

Therefore, to properly analyze the output of measurement, we must use mathematical statistics. This section will do a brief overview of statistical methods used in a particle physics. For comprehensive review of statistical methods the reader is referred to [3].

The basic object in all statistical tools is the *probability*. Consider the set S called sample space (this can be for example the set of all possible outcomes of experiment). The probability is a real valued function on S fulfilling following axioms:

1. For every subset A in S , $P(A) \geq 0$.
2. For disjoint subsets A and B ($A \cap B = \emptyset$), $P(A \cup B) = P(A) + P(B)$.
3. $P(S) = 1$.

These axioms ensures that the probability behaves in a natural way, for example the probability of no result is equal to zero and the probability of any result is less or equal to one.

We can also define conditional probability $P(A|B)$, i.e. the probability of getting result A given the result B :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Example is for example the probability of getting 6 in a dice roll if we know that the result is even. The probability of result being 6 *and* even is equal to $1/6$. The probability of getting even result is equal to $1/2$. So the probability of getting 6 given the result is even is $1/6 : 1/2 = 1/3$.

In practice, we do not work with the element of sample space but rather with the random variables which are numerical characteristics assigned to them. It can be interpreted as an outcome of repeatable experiment.

Consider a random variable x and suppose it can take on any value from a continuous range. Then we can define $f(x; \theta)dx$ as a probability that x would have value between x and $x + dx$. Function $f(x; \theta)$ is called probability density function (often abbreviated p.d.f.) and can depend on several parameters θ . In case x can take on only discrete values, $f(x; \theta)$ is directly the probability of obtaining result x .

Probability density functions are always normalized to unit area (unit sum in case of discrete distribution) and can be used to construct cumulative distribution functions $F(a)$ that tells what is the probability that the result will be less or equal to a :

$$F(a) = \int_{-\infty}^a f(x)dx$$

The parameters θ of distribution functions are often unknown and their determination is subject of statistics.

Any function of random variable is also a random variable, but with (in general) different p.d.f. Nevertheless the original p.d.f. can be used to calculate *expectation value* $E[u(x)]$ (or mean) of the function:

$$E[u(x)] = \int_{-\infty}^{\infty} u(x)f(x)dx$$

In a similar way we can calculate the *variance* σ^2 of the function which is the measure of the square of its width (or in other words the measure of uncertainty):

$$\sigma^2 = E[(u(x) - E[u(x)])^2] = \int_{-\infty}^{\infty} (u(x) - E[u(x)])^2 f(x)dx$$

We can also define joint p.d.f when we say that $f(x, y)dx dy$ is the probability that random variable x will be between x and $x + dx$ and similarly for y . We can also define conditional p.d.f as $f(x|y) = f(x, y)/f(y)$.

In general the p.d.f. is different for every random variable and for every experiment, nevertheless we can see that the distribution of results follows limited number of patterns. The most common distribution functions are

Binomial distribution Suppose we have a random process that can have exactly two outcomes ("success" or "failure"). If a probability of success is p than the probability of obtaining

exactly r successes after N trials is

$$f(r; N, p) = \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r}$$

The mean of this distribution is Np and variance is $Np(1-p)$

Poisson distribution It gives the probability of finding exactly n events in a given interval of x (for example time), when the events occur independently on one another and the average rate per given interval is ν :

$$f(n; \nu) = \frac{\nu^n e^{-\nu}}{n!}$$

This distribution can be viewed as a limit case of binomial distribution for $N \rightarrow \infty$ $p \rightarrow 0$ $Np = \nu$. For example, in accelerator experiment we have very large number of events but only a tiny probability p that some particular process will occur. Therefore Poisson distribution well model the number of events we can get in some time interval. The mean and variance of this distribution are the same and equal to ν .

Normal distribution Also called Gaussian distribution describes many physical processes. Its importance is coming from the central limit theorem which says that if we have several independent random variables y_1, y_2, \dots, y_n behaving according any p.d.f.s, their sum $x = \sum_i y_i$ will have a p.d.f that approaches the Gaussian for large n . Therefore the sum of a large number of fluctuations (typically measurement error) will be distributed as Gaussian although the particular fluctuations have different (and often unknown) p.d.f.s. The Gaussian probability distribution is given by

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-(x-\mu)^2/2\sigma^2)$$

The mean of this distribution is μ and variance is σ^2 .

The experimental results often follow normal distribution so that it is used to set the criteria for the new discovery. The typical is 5σ criterium which demands that the probability that the result is due to statistical fluctuation is less then probability of getting $x \geq \mu + 5\sigma$ or $x \leq \mu - 5\sigma$ in Gaussian distribution.

χ^2 distribution If x_1, x_2, \dots, x_n are independent random variables following the Gaussian distribution, then the sum $z = \sum_i (x_i - \mu_i)^2 / \sigma_i^2$ follows χ^2 distribution with n degrees of freedom. The distribution function is

$$f(z; n) = \frac{z^{n/2-1} e^{-z/2}}{2^{n/2} \Gamma(n/2)}$$

For large n χ^2 distribution approaches a Gaussian distribution with mean $\mu = n$ and variance $\sigma^2 = 2n$.

Exponential distribution If some process generates events as a function of x (e.g. time) according to Poisson distribution (e.g. radioactive decay), the distance in x from arbitrary starting point (e.g. probability of decay as a function of time) follows exponential distribution. Its distribution function is

$$f(x; \lambda) = \lambda e^{-\lambda x}$$

where λ is Poisson parameter ν per unit x . The mean of this distribution is $1/\lambda$ and variance is $1/\lambda^2$.

To validate probabilistic model or to find a parameters of p.d.f.s from data we use statistics. Common approach in particle physics is called frequentist statistics which interprets the probability as a frequency of an outcome of repeatable experiment. Other approach, called Bayesian which is using subjective probability called degree of belief. This approach will not be discussed here, but it can be found in ref. [3].

One of the main tasks of the statistics is to estimate value of parameters of probability distribution from data. The function of data which gives the probable value of distribution parameters is called estimator and is written with hat: $\hat{\theta}$. The most common are estimators for mean and variance. The mean estimator is simple the arithmetic average:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$$

The variance estimator is

$$\widehat{\sigma^2} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu})^2$$

These are consistent (for $N \rightarrow \infty$ we obtain $\hat{\mu} \rightarrow \mu$ and $\widehat{\sigma^2} \rightarrow \sigma^2$) and unbiased ($E[\hat{\mu}] - \mu = 0$) estimators of μ and σ^2 .

Another method for estimating parameters is the maximal likelihood. Suppose, we have N measured values $\vec{x} = (x_1, x_2, \dots, x_N)$ which are described by joint p.d.f. $f(x, \theta)$ where θ is vector of parameters $\vec{\theta} = (\theta_1, \dots, \theta_n)$. If the measurements are independent and each of them follows the p.d.f. $f(x, \theta)$ then the likelihood function L is

$$L(\vec{\theta}) = \prod_{i=1}^N f(x_i; \theta)$$

The likelihood function is viewed as a function of the parameters, evaluated on data, and we try to find such parameters that maximize the likelihood function. From technical reasons

is easier to work with the logarithm of this function, because it is maximized for the same set of parameters. We obtain set of equations

$$\frac{\partial \ln L}{\partial \theta_i} = 0$$

The solution of these equations are the demanded parameters of the p.d.f.

There are also other useful methods for estimating the distribution parameters from data (e.g. the least squares method), they are discussed in [3].

Very important use of statistics are the hypothesis tests. These are rules for accepting or rejecting a hypothesis depending on a data which are represented by a vector \vec{x} . The hypothesis is statement about the distribution of \vec{x} , i.e. the data. Usually we have two hypotheses: H_0 (background or null hypothesis) and H_1 (signal) which are exactly complementary (i.e. either one or the second will take place).

The statistical test then says for which values of \vec{x} we should accept the signal hypothesis. These values form so-called acceptance region in \vec{x} space.

In particle physics, the probability to accept signal hypothesis H_1 is called selection efficiency. Similarly, the probability to accept H_0 is called background efficiency and one minus background efficiency is called purity of the sample.

It is often inconvenient to work with a full data vector \vec{x} , so we often define test statistics t , which can be either single number or a vector with fewer components than the original data vector. Hypotheses about the distribution of \vec{x} will determine a distribution of t and the acceptance region in \vec{x} will define an interval in t (in case it is a number).

Question is how to construct a test which maximizes the purity (or minimizes background efficiency) for a given signal efficiency. The Neyman-Pearson lemma states that optimal test statistics is

$$\lambda(x) = \frac{f(x|H_1)}{f(x|H_0)}$$

or any monotonous function of it. Signal is accepted when this ratio is larger than given constant which determines the selection efficiency.

In case the functions $f(x|H_1)$ and $f(x|H_0)$ are not known, there is a lot of various multivariate classifiers which can efficiently filter signal events from background. These include for example neural networks or decision trees. For details please refer to [3].

3.2 Introduction to B -tagging

3.2.1 Overview of B -tagging

After the general overview of the HEP statistics, this section will present a particular example of a statistical test: selection of jets stemming from the hadronization of b -quark called b -tagging.

The ability to identify jets containing b -hadrons is important for the high- p_T physics program of a general-purpose experiment at the LHC such as ATLAS. This is in particular useful to select very pure top samples, to search and/or study Standard Model or supersymmetric (SUSY) Higgs bosons which couple preferably to heavy objects or are produced in association with heavy quarks, to veto the large dominant $t\bar{t}$ background for several physics channels and finally to search for new physics: SUSY decay chains, heavy gauge bosons, etc.

The identification of b -jets takes advantage of several of their properties which allow us to distinguish them from jets which contain only lighter quarks. First the b -hadron retains about 70% of the original b quark momentum and also the mass of b -hadrons is relatively high (> 5 GeV). Thus, their decay products may have a large transverse momentum with respect to the jet axis. Another and most important property is the relatively long lifetime of hadrons containing a b quark, of the order of 1.5 ps ($c\tau \approx 450\mu\text{m}$). A b -hadron in a jet with $p_T = 50$ GeV will therefore have a significant flight path length $\langle l \rangle = \beta\gamma c\tau$, traveling on average about 3 mm in the transverse plane before decaying. Such displaced vertices can first be identified inclusively by measuring the impact parameters of the tracks from the b -hadron decay products.

The transverse impact parameter, d_0 , is the distance of closest approach of the track to the primary vertex point, in the $r - \varphi$ projection. The longitudinal impact parameter, z_0 , is the z coordinate of the track at the point of closest approach in $r - \varphi$. The tracks from b -hadron decay products tend to have rather large impact parameters so they can be distinguished from tracks stemming from the primary vertex. The other more demanding option is to reconstruct explicitly the displaced vertices. These two approaches are called spatial b -tagging.

Finally, the semi-leptonic decays of b -hadrons can be used by tagging the lepton in the jet. In addition, thanks to the high mass of b -hadrons, the lepton will have a relatively large transverse momentum and also a large momentum relative to the jet axis. This is the so-called soft lepton tagging (the lepton being soft compared to high- p_T leptons from W or Z decays).

3.2.2 B -tagging input

The main ingredient for b -tagging are the tracks of charged particles reconstructed in the Inner Detector. On average, reconstructed track has 3 pixel hits, 4 SCT hits and about 36 hits in TRT. The resolution of various parts of the Inner Detector was quoted in the previous chapter.

ID has ability to reconstruct tracks with $|\eta| < 2.5$ and with momentum larger than 500 MeV. For typical b -tagging track (central, with transverse momentum 5 GeV) it has impact parameter resolution about $35 \mu\text{m}$ and transverse momentum resolution about 1.5 % [23].

Before using tracks for b -tagging a baseline selection is applied which rejects badly measured and fake tracks as well as tracks coming from long-lived particles like K and material interaction. Two levels of selection are used. The standard quality cut requires that track has at least seven precision hits (i.e. from Pixel/SCT) and transverse and longitudinal impact parameters must fulfil $|d_0| < 2$ mm and $|z_0 - z_{pv}| \sin \theta < 10$ mm respectively, where z_{pv} is longitudinal coordinate of the primary vertex. The b -tagging quality cut further requires that we have at least two

hits in Pixel detector, one of them being in the innermost layer as well as $|d_0| < 1$ mm and $|z_0 - z_{pv}| \sin \theta < 1.5$ mm.

Another key ingredient for b -tagging is the primary vertex of the event. The impact parameters of tracks are recomputed with respect to its position and tracks compatible with the primary vertex are excluded from the secondary vertex searches. At LHC the beam-spot size will be $\sigma_{xy} = 15$ μm and $\sigma_z = 5.6$ cm: therefore the primary vertex is especially important for the z direction, while in the transverse plane only the beam-line could be used. The strategies to find the primary vertex and their performance are explained in [24]. The efficiency to find the primary vertex is very high in the high- p_T events of interest, and the resolution on its position is around 12 μm in each transverse direction and 50 μm along z .

Other input for b -tagging are jets which are reconstructed with standard Athena tools. For b -tagging purposes, only the jet direction is relevant. In the first place, this direction is used to define which tracks should be associated with the jets. The actual tagging is done on this subset of tracks. Currently all tracks within a distance $\Delta R < 0.4$ of the jet axis in the $\Delta\eta \times \Delta\phi$ space are associated to the jet. The jet direction is also used to sign the impact parameters of the tracks in the jet as explained later.

Last input are the leptons from b -hadrons semi-leptonic decays. Muons are identified in muon spectrometer and their track is matched to a track in the ID. Muons satisfying some basic requirements ($p_T > 3$ GeV, $|d_0| < 4$ mm) are associated to the closest jet provided that $\Delta R < 0.5$. Finally, the kinematic properties of the jet-muon system are used in order to reject the background caused by punch-throughs and decays-in-flight in light jets.

Reconstructing soft electrons in the calorimeter inside a jet is more difficult. This is achieved by matching an inner detector track to an electromagnetic cluster. For a given track, only the energy contained in a small window around the track extrapolation is used. The contribution of neighbouring hadronic showers is therefore reduced. The identification procedure takes full advantage of the tracking capabilities of the inner detector as well as of the granularity of the electromagnetic calorimeter: a likelihood ratio combines inner detector information such as transition radiation hits with shower shape variables from the calorimeter. The performance is however highly dependent on the track density in jets as well as the quantity of matter in front of the electromagnetic calorimeter.

3.2.3 B -tagging performance estimation

To define b -tagging performance, the Monte-Carlo event history is used to know the type of parton from which a jet originates. This *labeling* procedure is however not unambiguous. For this report a quark labeling has been used: a jet is labeled as a b -jet if a b quark with $p_T > 5$ GeV is found in a cone of size $\Delta R = 0.3$ around the jet direction. The various labeling hypotheses are tried in this order: b quark, c quark and τ lepton. When no heavy flavour quark nor τ -lepton satisfies these requirements, the jet is labeled as a light-jet. No attempt is made to distinguish between u , d , s quarks and gluon since such a label is even more ambiguous.

For performance studies, only jets fulfilling $p_T > 15$ GeV and $|\eta| < 2.5$ are considered and referred to as taggable jets. In the following, jets for which no track passed the b -tagging quality cuts are still counted in the performance estimators. However events where the primary vertex could not be reconstructed are ignored.

The tagging efficiency is naturally defined as the fraction of jets labeled as b -jets (see previous section) which are actually tagged as b -jets by the tagging algorithm under study. The mis-tagging rate is the fraction of jets not labelled as b which are actually tagged as b -jets. For historical reasons the jet rejection is used instead: this is simply the inverse of the mis-tagging rate.

A difficulty arises as soon as the jet multiplicity is high and various jet flavours are present in a single event: a jet with $\Delta R(\text{jet} - b) = 0.31$ is labelled as a light jet, although tracks from b -hadron decay with high lifetime content are likely to be associated to it.

This leads to a decrease of the estimated performance, not related to the b -tagging algorithm itself but to the labelling procedure which strongly depends on the activity of the event. In order to obtain a more reliable estimation of b -tagging performance, a purification procedure has been devised: light jets for which a b quark, a c quark or a τ lepton is found within a cone of size $\Delta R = 0.8$ around the jet direction are not used to compute the rejection.

The performance estimated after purification represents the intrinsic power of the b -tagging algorithms and should be similar for different kinds of hard event, whereas results obtained for the complete light jet sample are more dependent on the event type. On the other hand, the latter is more representative of the actual b -tagging power for a given physics analysis. This is illustrated in Figure 3.1: the light jet rejection in simple WH events is similar without or with purification (left plot), while for busier $t\bar{t}$ events (right plot) the two curves differ in the region where lifetime content as opposed to resolution effects dominates (*i.e.* for $\epsilon_b < 80\%$). In the following, jets fulfilling the purification procedure will be referred to as purified or pure jets, while all the jets will be called raw jets or non-pure jets.

3.3 b -tagging algorithms

In this section several b -tagging algorithms will be introduced. The most powerful ones are spatial algorithms, built on tracks and subsequently vertices. Most of them are using likelihood approach described in the beginning of this chapter, but there also exist simple and more robust algorithms.

3.3.1 Impact parameter based b -tagging

All tracks in the jet fulfilling the b -tagging quality cuts described previously are considered for the tagging algorithms based on displaced tracks or vertices. In typical $t\bar{t}$ events, the average number of those tracks per light (b -) jet is 3.7 (5.5) and their average p_T is 6.6 (6.3) GeV, respectively.

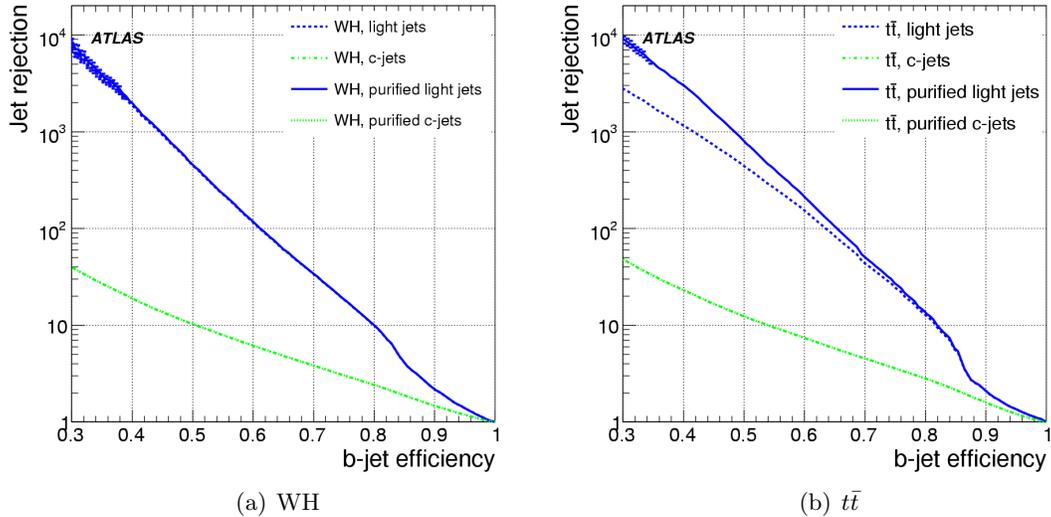


Figure 3.1: Rejection of light jets and c -jets with and without purification versus b -jet efficiency for WH ($m_H = 120$ GeV) and $t\bar{t}$ events, using the tagging algorithm based on 3D impact parameter and secondary vertex. Plots from [23].

The impact parameters of tracks are computed with respect to the primary vertex. The impact parameter is signed using the jet direction as measured by the calorimeters. Tracks crossing the jet axis behind the primary vertex have a negative impact parameter. For illustration see Fig. 3.2.

The distribution of the signed transverse impact parameter d_0 is shown on figure 3.3, left plot, for tracks coming from b -jets, c -jets and light-jets. The right plot shows the significance distribution d_0/σ_{d_0} which gives more weight to precisely measured tracks. Combining the impact parameter significances of all the tracks in the jet is the basis of the first method to tag b -jets. Three tagging algorithms are defined in this way: IP1D relies on the longitudinal impact parameter, IP2D on the transverse impact parameter and finally IP3D which uses two-dimensional histograms of the longitudinal versus transverse impact parameters.

To select b -jets a likelihood ratio method is used: the measured value S_i of a discriminating variable (in this case signed impact parameter significance) is compared to pre-defined smoothed and normalized distributions for both the b - and light jet hypotheses, $b(S_i)$ and $u(S_i)$. These distributions are shown in Fig. 3.3. Two-dimensional probability density functions are used as well for IP3D. The ratio of the probabilities $b(S_i)/u(S_i)$ defines the track weight, which can be combined into a jet weight W_{Jet} as the sum of the logarithms of the N_T individual track weights

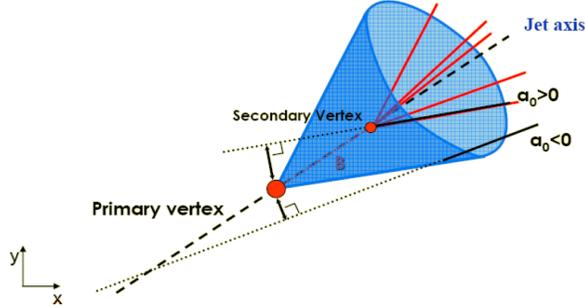


Figure 3.2: The explanation of meaning of impact parameter. Dashed line is the jet axis, red solid lines are tracks associated to a given jet which is shown as blue cone. Primary and secondary vertex are denoted by red dots. Whole picture is projection to transverse plane.

W_i in accordance with Neyman-Pearson lemma:

$$W_{Jet} = \sum_{i=1}^{N_T} \ln W_i = \sum_{i=1}^{N_T} \ln \frac{b(S_i)}{u(S_i)} \quad (3.1)$$

The distribution of such a weight is shown in Figure 3.4 for b -, c - and light jets for tagging algorithm IP2D. To select b -jets, a cut value on W_{Jet} must be chosen, corresponding to a given efficiency. The relation between the cut value and the efficiency depends on the jet transverse momentum and rapidity, and therefore is different for different samples.

3.3.2 Secondary vertex based b -tagging

For further discrimination between b -jets and light jets we can look for the inclusive vertex formed by the decay products of the bottom hadron, including the products of the eventual subsequent charm hadron decay. This of course require the reconstruction of secondary vertex which is much more complicated than just looking for the impact parameters of single tracks. Therefore b -tagging using secondary vertex is much more delicate than the impact parameter based one.

The search starts by building all two-track pairs that form a good vertex, using only tracks far enough from the primary vertex. Vertices compatible with a primary vertex or material interaction are rejected. All tracks from the remaining two-track vertices are combined into a single inclusive vertex, using an iterative procedure to remove the worst track until the χ^2 of the vertex fit is good.

Three of the vertex properties are exploited: the invariant mass of all tracks associated to the vertex, the ratio of the sum of the energies of the tracks participating to the vertex to the sum

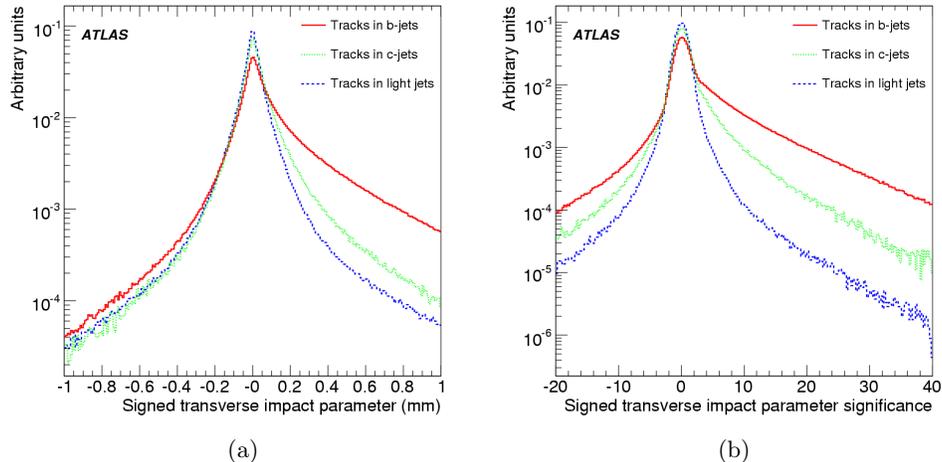


Figure 3.3: Signed transverse impact parameter d_0 distribution (a) and signed transverse impact parameter significance d_0/σ_{d_0} distribution (b) for b -jets, c -jets and light jets. Figure from [23].

of the energies of all tracks in the jet and the number of two-track vertices. These properties are illustrated in Figure 3.5 for b -jets and light jets.

The SV tagging algorithms make different use of these properties: SV1 relies on a 2D-distribution of the two first variables and a 1D-distribution of the number of two-track vertices, while SV2 is based on a 3D-histogram of the three properties which requires quite some statistics. The secondary vertex search efficiency depends in particular on the event topology, but the typical efficiency is higher than 60%.

The discrimination between b -jets and light jets are done using the same likelihood method as described in the previous section. The only difference is that for SV algorithms the discriminating variables are the three introduced above. Due to the same formalism, the jet weights from SV algorithms can be easily combined with jet weights from impact parameter based algorithms by just summing the weights. This is shown on Fig. 3.4.

A completely new algorithm, JetFitter, is also available, which exploits the topological structure of weak b - and c -hadron decays inside the jet. A Kalman filter is used to find a common line on which the primary vertex and the beauty and charm vertices lie, as well as their position on this line approximating the b -hadron flight path. With this approach, the b - and c -hadron vertices are not merged, even when only a single track is attached to each of them. The discrimination between b -, c - and light jets is based on a likelihood using similar variables as the SV tagging algorithm above, and additional variables such as the flight length significances of the vertices. This algorithm and its performance are also described in detail in Ref. [24].

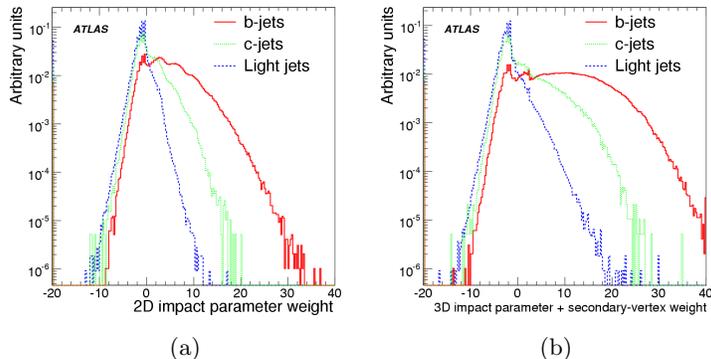


Figure 3.4: Jet b -tagging weight distribution for b -jets, c -jets and purified light jets for the IP2D tagging algorithm (a) and for combined weights in IP3D+SV1 tagging algorithm (b). Figure from [23].

3.3.3 Other spatial b -tagging algorithms

Impact parameter and secondary vertex based b -tagging require an a-priori knowledge of properties of b -jets and light jets. There are methods how to obtain them from the data, but this would require at least 100 pb^{-1} of collected data. So there are few other spatial tagging algorithms, that are less powerful and have less reliance on Monte Carlo. They are expected to be easier and faster to commission with the first real data.

The simplest approach that could be used, at least at the beginning, is the counting of tracks with large impact parameter or large impact parameter significance. Requiring a few of these tracks provides a sample enriched in b -jets. Such a simple tagger may also be very useful at the trigger level.

Another approach is to combine the impact parameter of all the tracks in the jet. JetProb is an implementation of the ALEPH tagging algorithm [29], used extensively at LEP and later at the Tevatron.

The signed impact parameter significance d_0/σ_{d_0} of each selected track in the jet is compared to a resolution function \mathcal{R} for prompt tracks, in order to measure the probability that the track i originates from the primary vertex (Figure 3.6(a)):

$$\mathcal{P}_i = \int_{-\infty}^{-|d_0^i/\sigma_{d_0}^i|} \mathcal{R}(x) dx \quad (3.2)$$

The resolution function can be measured in data and the individual probability of each of the N tracks associated to the jet are then combined to obtain a jet probability \mathcal{P}_{jet} which

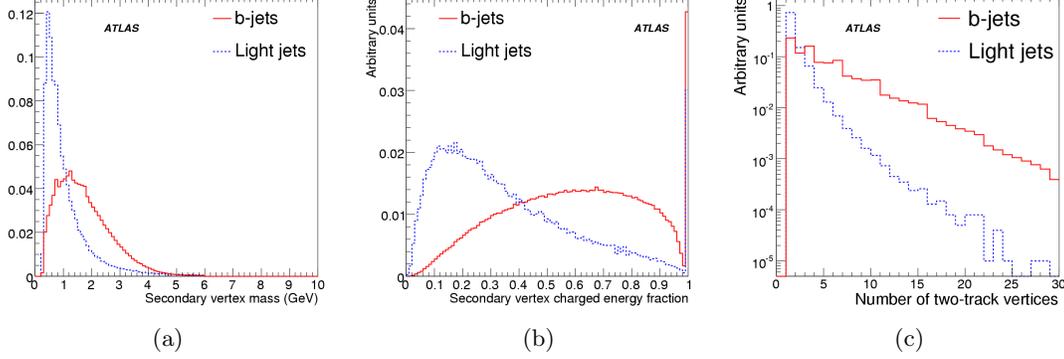


Figure 3.5: Secondary vertex variables: invariant mass of all tracks in vertex (a), energy fraction vertex/jet (b) and number of two-track vertices (c) for b -jets and light jets. Figure from [23].

discriminates b -jets against light jets (Figure 3.6(b)):

$$\mathcal{P}_{jet} = \mathcal{P}_0 \sum_{j=0}^{N-1} \frac{(-\ln \mathcal{P}_0)^j}{j!} \quad (3.3)$$

where

$$\mathcal{P}_0 = \prod_{i=1}^N \mathcal{P}'_i \quad \text{and} \quad \begin{cases} \mathcal{P}'_i = \frac{\mathcal{P}_i}{2} & \text{if } d_0^i > 0 \\ \mathcal{P}'_i = \left(1 - \frac{\mathcal{P}_i}{2}\right) & \text{if } d_0^i < 0 \end{cases} \quad (3.4)$$

3.3.4 Soft lepton algorithms

Soft lepton tagging relies on the semi-leptonic decays of bottom and charm hadrons. Therefore it is intrinsically limited by the branching ratios to leptons: at most 21% [3] of b -jets will contain a soft lepton of a given flavour, including cascade decays of bottom to charm hadrons. However, tagging algorithms based on soft leptons exhibit very high purity and low correlations with the track-based tagging algorithms, which is very important for checking and cross-calibrating performance in data.

The association of muon to jet was briefly described in previous section. The selection of b -jets is again done using the likelihood methods. The tagging algorithm and its performance is described in detail in [26]. Soft muon tagging can reach rejections of about 300 for b -jet selection efficiency 10 %.

Similar methods are applied in case of soft electron tagging. Again, we would not go to much detail as this tagging algorithm is described in [27]. Soft electron tagging can reach 100 for efficiency about 7 %.

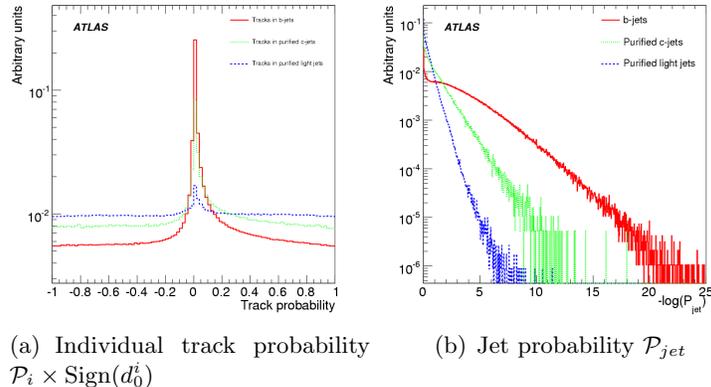


Figure 3.6: Distributions of the probability of compatibility with the primary vertex for individual tracks (a) and for all tracks in the jet (b) as defined for JetProb. The cases of b -jets (red plain), c -jets (green dashed) and light jets (blue dotted line) are shown. Figure from [23].

3.3.5 Combination of tagging algorithms

As was showed in the previous sections, we can readily combine jet weights calculated by impact parameter based taggers and by SV taggers, because they use the same formalism. Similarly we can combine IP taggers and JetFitter. Although no other combination was tried, recently some studies appeared which try to combine all taggers (including soft lepton ones) using multivariate analysis (boosted decision trees). Details are in Ref. [23].

3.3.6 Calibration of b -tagging

We have seen that likelihood based tagging algorithms require the knowledge of p.d.f.s of discriminating variable for both b -jet and light jet hypotheses. These can be obtained either from Monte Carlo or from data, when available. The process of acquiring them is called *calibration* of tagging algorithms and the resulting plots are called reference histograms. Proper calibration is very important because in principle the distribution of discriminating variables is different in every geometry configuration, leading to improper tagging of b -jets (see Chapter 4).

3.4 B -tagging performance

There are basically two types of events which are used as a benchmark for b -tagging performance. The first one are events with low jet multiplicities when Higgs boson is produced together with W boson via Higgsstrahlung. It is denoted “WH”. As was said in the first chapter, this process

has little physical relevance, but from historical reasons is used in performance studies.

W boson decays in this sample leptonically and Higgs boson decays to $b\bar{b}$ pair. For the determination of rejection of charmed and light jets the Higgs boson is forced to decay to $c\bar{c}$ or to the unlikely $u\bar{u}$ channel respectively. Higgs boson mass was set to 120 GeV.

To study more energetic jets, similar physics processes have been considered for a different Higgs boson mass, $m_H=400$ GeV (again such a choice is unphysical since a 400 GeV Higgs boson would not decay to $b\bar{b}$ but is useful for these studies). The light jet rejection for WH events are shown in Tab 3.1.

Besides these events with simple topology, b -tagging performance is also studied in the multi-jet channels, mostly with $t\bar{t}$ pair. In this case the jet multiplicity is much higher. Every top decays to W and b in this sample and we accept only these events where at least one W decays leptonically - so-called semileptonic channel. In case both W decay to leptons we have dileptonic channel.

For the dominant one lepton+jets channel, there are usually at least four jets from the hard process and extra jets from radiation. Several flavours of jets are present at the same time in the event: two b -jets from the top quarks, light jet(s) and often a c -jet from the W decaying hadronically. This increases the likelihood of having light jets contaminated with heavy flavour and also makes the labeling of jets even more ambiguous as discussed previously.

The jets of the various flavours were taken from the same sample in this case, unlike for events of the WH channels. Table 3.1 shows the light jet rejection achieved in $t\bar{t}$ events and in $t\bar{t}jj$ events. The latter events are $t\bar{t}$ events which were filtered in order to have at least six jets, of which four are taggable (see section 3.2.3 for explanation). Because the performance on both of them is similar, they have been merged.

Table 3.1 shows also the light jet rejection achieved in even more complex topologies with at least six jets. Those channels are relevant for the Higgs discovery channel $t\bar{t}H$ which requires a high b -tagging efficiency since four jets are b -tagged and the cross section is low: therefore the more typical working points of ϵ_b around 60-70% are shown. The performance for dominant background channel for $t\bar{t}H$ which is $t\bar{t}b\bar{b}$ is also shown.

The light jet rejections as a function of b -tagging efficiency for various tagging algorithms is shown in Fig. 3.7.

The spatial b -tagging performance depends also strongly on the jet momentum and rapidity. The p_T and η dependencies of the b -tagging efficiency and light jet rejection for a given cut on the b -tagging weight are shown in Figure 3.8. At high p_T or at high $|\eta|$, the b -jet tagging performance is poor, regardless of which tagging algorithm is used. At low p_T , maintaining a reasonable b -jet efficiency is possible only by loosening the cut on the weight, at the price of a very low rejection of light jets. Because of these strong p_T and η dependencies, and since various samples have very different spectra, it is not straightforward to compare between channels the integrated rejection numbers shown in the following.

Table 3.1: Integrated rejection of light jets (with and without purification when it applies), for various event types and for several tagging algorithms. For each case, the cut on the b -tagging weight is chosen to lead to the quoted average b -tagging efficiency ϵ_b over the sample considered. Numbers from [23].

	JetProb	IP2D	IP3D	IP3D+SV1	IP3D+JetFitter
WH ($m_H = 120$ GeV) events					
$\epsilon_b = 50\%$	83±1	116±2	190±3	458±13	555±17
$\epsilon_b = 60\%$	30±0	42±0	59±1	117±2	134±2
WH ($m_H = 400$ GeV) events					
$\epsilon_b = 50\%$	73±1	163±3	179±3	298±7	396±11
$\epsilon_b = 60\%$	27±0	56±1	58±1	96±1	123±2
$t\bar{t}$ and $t\bar{t}jj$ events					
Raw, $\epsilon_b = 50\%$	91±0	146±1	232±2	456±4	635±7
Purified, $\epsilon_b = 50\%$	97±0	186±1	310±3	789±10	924±13
Raw, $\epsilon_b = 60\%$	28±0	46±0	67±0	154±1	189±1
Purified, $\epsilon_b = 60\%$	28±0	51±0	76±0	206±1	224±2
$t\bar{t}H$ events					
Raw, $\epsilon_b = 60\%$	23±0	35±0	49±1	90±2	113±2
Purified, $\epsilon_b = 60\%$	25±0	48±1	72±1	188±5	188±5
Raw, $\epsilon_b = 70\%$	10±0	14±0	18±0	32±0	31±0
Purified, $\epsilon_b = 70\%$	11±0	17±0	22±0	46±1	37±1
$t\bar{t}b\bar{b}$ events					
Raw, $\epsilon_b = 60\%$	23±0	34±0	50±1	100±2	123±2
Purified, $\epsilon_b = 60\%$	24±0	41±0	64±1	156±4	166±4
Raw, $\epsilon_b = 70\%$	10±0	13±0	18±0	32±0	28±0
Purified, $\epsilon_b = 70\%$	10±0	15±0	20±0	40±0	31±0

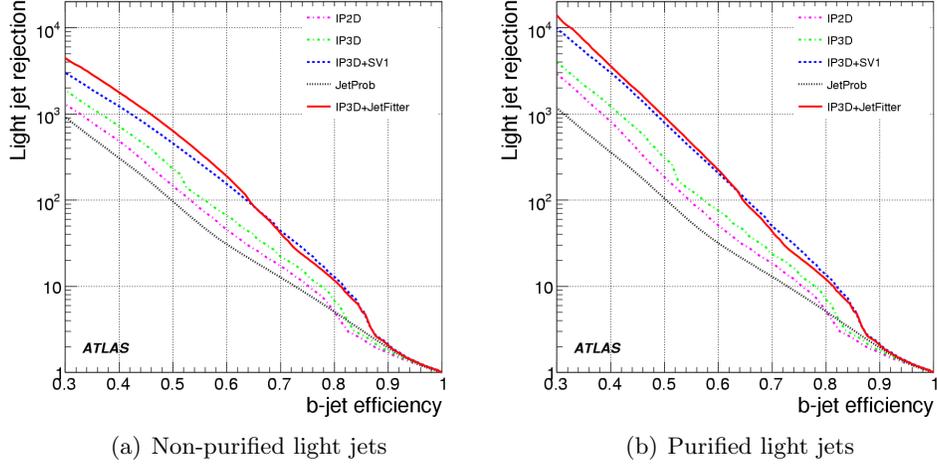


Figure 3.7: Rejection of light jets versus b -jet efficiency for $t\bar{t}$ and $t\bar{t}jj$ events and for all tagging algorithms: JetProb, IP2D, IP3D, IP3D+SV1, IP3D+JetFitter. Figure from [23].

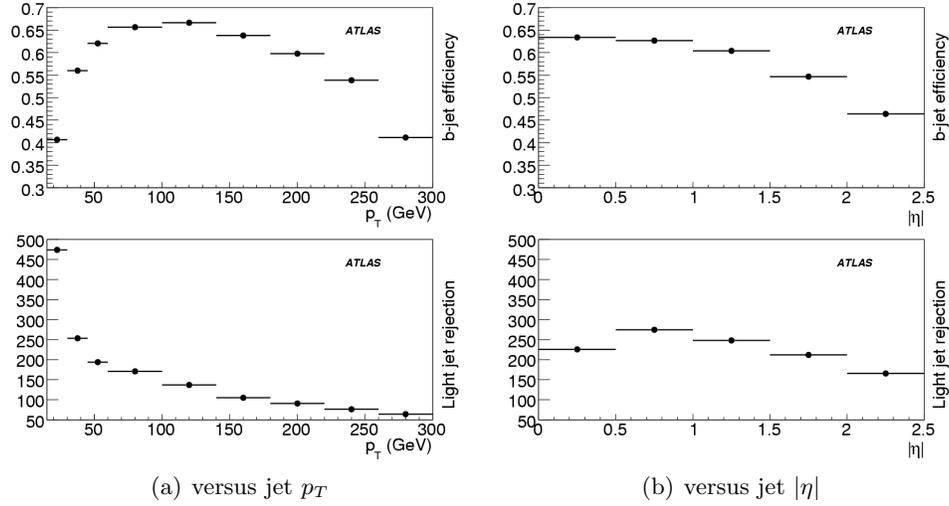


Figure 3.8: b -tagging efficiency and purified light jet rejection obtained with the IP3D+SV1 tagging algorithm operating at a fixed cut of 4 on the b -tagging weight, for $t\bar{t}$ events.

Chapter 4

Impact of Misalignment on b -tagging

4.1 Misalignment

During the assembly, it was not possible to place all parts of ATLAS detector exactly on places where they should be according to plan. Sometimes also an already installed subsystem moved due to change of surrounding conditions or due to some accident in assembly.

Because of this, our knowledge about the current detector part positions is imperfect. This state is called misalignment, meaning that ATLAS is aligned in a different way than we want and we do not know how is aligned. Misalignment are not only shifts from nominal positions, but also deformations of the material the detecting modules have been made from. The effect of misalignment is shown on Fig. 4.1.

We can see that misalignment leads to improper tracking which subsequently affects all services which use tracks as in input, particularly b -tagging. Therefore a study was done to quantify the effect of misalignment on b -tagging. The result of this study are summarized in [25].

The effects of misalignment on b -tagging has been studied using a number of different alignments sets. These include the perfectly aligned detector, two hand made sets with known levels of random misalignment and an alignment set produced with the actual alignment algorithms to be used in ATLAS. The performance of b -tagging for each of these sets has been investigated with both $t\bar{t}$ and $WH(m_H = 120 \text{ GeV})$ samples. Because the misalignments effectively add an additional error to the measurements, the errors assigned to the clusters need to be adjusted depending on the level of misalignment. The error tuning procedure and resulting error scalings will be presented in next sections. All samples were investigated with and without this error scaling.

Four different alignment sets were considered in this study. The misalignment was introduced in the level of Monte Carlo simulation. The level of misplacement we used is representative of the amount of misalignment expected before any attempt to align the detector. The misalignments

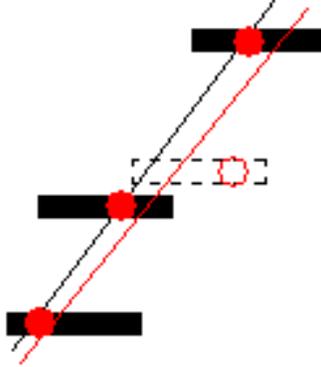


Figure 4.1: Illustration of effect of misalignment. The full black rectangles are physical positions of the detecting modules. The middle is however misaligned: its nominal position is designed by dashed shape of module. Solid black line is the real particle track and full red circles are the places of hit. The empty red circle shows where the software thinks the second hit was made. Finally, the red line shows the fitted track and we can observe the difference to reality due to misalignment of central module.

are of the order of 10–100 μm at the level of individual modules and assembly structures such as layers and disks and misalignments of the order of a few mm at the whole subsystem level.

The misalignment introduced are too large to allow for reasonable reconstruction. What is desired is to reconstruct the resulting data sets with alignment corrections that are typical of what is expected in the real detector, after which only small misalignments should remain. In addition to the perfectly aligned case, three different alignments sets were considered. These alignment sets correct most of the misalignment but do not correct it perfectly leaving some *residual misalignment*.

- **Perfect:** This is the ideal case where the same set of alignments used in the simulation are used in the reconstruction and so one does not see any misalignment.
- **Aligned:** This uses an alignment set produced using the actual track based alignment algorithms developed for the ATLAS detector. It is expected to include any systematic deformations that the alignment procedure itself causes. While some systematic effects were included in the misalignments introduced in the simulation, such as clocking effects where each subsequent layer was rotated by increasing amounts, it does not contain several systematic deformations which are expected. In particular large scale structures such as layers and discs were treated as rigid objects without any internal deformations such as a twist. Also pixel stave bows which are known to occur were not introduced. So it is possible that this set is still optimistic. This set is a first attempt at the full scale alignment of the

inner detector and so should not be considered the final answer of what will be seen in the real detector, however, it is considered to be the most realistic case studied here.

- **Random10:** This is a hand made alignment set that takes the misalignment set used in simulation and randomly shifts the module positions by small amounts such that it corrects most of the misalignments but leaves some residual misalignment. These residual misalignments were introduced at different levels in the hierarchy. Random shifts and rotations were made to individual modules, and whole layers and disks. A small shift and rotation was also made to the whole pixel structure. Only pixel residual misalignments were put in since the degradation of the b -tagging performance is expected to be dominated by the alignment of the pixel system. The SCT and TRT were corrected perfectly as in the perfect alignment case. Due to movements of higher level structures in this set some systematics effects may exist. The levels of misalignment are given in Table 4.1. The axis definitions for the module level uses a local frame where x and y are the $r\phi$ and η measurement directions respectively and z is out of the plane. For higher levels they correspond to the global frame with z -axis along the beam direction. $RotX$, $RotY$, $RotZ$ are rotations around the corresponding axes. The module level shifts in the $r\phi$ measurement direction are around $10 \mu\text{m}$. The set is a guess at the level of misalignments expected during the early running period. It is not well known what levels of misalignments are expected after certain running periods so this is just an indication rather than being a firm prediction of what is expected at start up. Comparison with the real alignments (“Aligned” set) shows this to be a rather pessimistic scenario.
- **Random5:** As with “Random10”, but with levels of misalignment better by about a factor of 1.5 to 2. This is a guess at what might be expected after several years of running. Like “Random10”, this set introduces misalignments at the three levels of hierarchy with levels of misalignment given in Table 4.2.

Table 4.1: Residual misalignment for “Random10”. Random misalignments were generated with a Gaussian distribution with σ as tabulated. Shifts are in μm and rotations are in mrad.

Level	x	y	z	$RotX$	$RotY$	$RotZ$
Module	10	30	30	0.3	0.5	0.2
layer	10	10	15	0.05	0.05	0.1
disk	10	10	30	0.2	0.2	0.1
Whole Pixel	10	10	15	0.1	0.1	0.1

Table 4.2: Residual misalignment for “Random5”. Random misalignments were generated with a Gaussian distribution with σ as tabulated. Shifts are in μm and rotations are in mrad.

Level	x	y	z	$RotX$	$RotY$	$RotZ$
Module	5	15	15	0.15	0.3	0.1
layer	7	7	10	0.02	0.02	0.05
disk	7	7	20	0.1	0.1	0.05
Whole Pixel	7	7	10	0.05	0.05	0.05

4.2 Error Scaling

The intrinsic error of a hit will depend on a number of factors. These factors are taken into account when calculating the intrinsic error of the hit. In the case of a perfectly aligned detector, if these intrinsic errors are properly determined one expects the pull distribution of the hit residuals (i.e. the distribution of difference between the hit position and the fitted track in the units of measurement error) to be close to one. The differences between the real positions of individual hits and those recorded by a misaligned detector lead to an additional error term that must be added in quadrature to the intrinsic error of the hits.

The errors on the hits directly affect whether a hit is associated to a track, the track propagation and track parameter errors and other objects that use tracks as input, such as vertexing. Of particular importance to b -tagging is the precision of the impact parameter and vertexing. It is therefore necessary to have accurately assigned hit errors. In order to correct the hit errors, the size of the error was modified using two parameters a and c :

$$\sigma'^2 = a^2 \cdot \sigma^2 + c^2 \quad (4.1)$$

where:

- σ is the original error assigned to the hit. This should normally be close to the intrinsic resolution if properly determined,
- a is a factor to the error, which is meant to compensate for inaccuracies in the intrinsic error determination,
- c is a constant added in quadrature to the error. This is meant to correct effects attributed purely to residual misalignments.

Since each detector component can have significantly different behaviour, the granularity of each detector component has to be taken into account, and therefore different sets of (a, c) have

to be computed separately for the barrel and endcap regions for each detector technology, as well as for the different $r\phi$ and η measurement directions in the case of the pixel detector.

For the derivation of the (a, c) pairs, the distributions of hit residuals and their pull distributions are analyzed, and in particular the deviations of the pull widths from the ideal value of 1 are investigated.

Since the scale factor a is intended to correct the intrinsic resolutions, this is most easily obtained with a perfectly aligned geometry. Naturally, this is not possible with real data, where more in depth studies would be needed to determine if the assigned intrinsic errors are appropriate. Currently the factor a is needed as the reconstruction does not match the simulation. It is assumed, however, that the best knowledge from testbeam and simulation will be put into the determination of the intrinsic error such that a will be close to 1 and any remaining differences would be absorbed into the parameter c .

The widths of the resulting pull distributions can be used directly as the scaling factors a . This is iterated a few times, applying the correction, rerunning reconstruction and then determining new values of a . The iterations are necessary due to correlations between detector components. The factor of c is set to zero when determining the a factor.

The resulting factors of a are then kept constant when used for the misaligned detector. Several iterations (apply (a, c) factors, reconstruct sample, analyze pulls) are performed using a misaligned detector, in order to determine the c factor. It is computed using the formula:

$$c_i^2 = (p_{obs}^2 - 1)a^2\sigma_0^2 + p_{obs}^2c_{i-1}^2 \quad (4.2)$$

where c_i and c_{i-1} are the values of the c factor obtained in the iteration i and $i - 1$, respectively, p_{obs} is the hit residual pull width observed at step i , and σ_0 is the average intrinsic detector resolution. The detector resolution was presented in Table 2.2.

Table 4.3: Error scaling parameters for the different alignment scenarios. The parameter a was tuned using the ‘‘Perfect’’ case and used for all alignment scenarios.

	All a	Perfect $c(\mu\text{m})$	Random10 $c(\mu\text{m})$	Random5 $c(\mu\text{m})$	Aligned $c(\mu\text{m})$
Pixel Barrel Phi	1.03	0	31	13	3
Pixel Barrel Eta	0.97	0	71	34	13
Pixel Endcap Phi	1.05	0	30	14	3
Pixel Endcap Eta	1.08	0	43	11	15
SCT Barrel	0.78	0	0	2	7
SCT Endcap	0.86	0	6	5	8
TRT Barrel	0.82	0	11	3	37
TRT Endcap	0.77	0	11	10	19

Table 4.4: Efficiency of track reconstruction.

Setup	efficiency (%)
Perfect	97.09 ± 0.02
Random10	95.50 ± 0.03
Random10 + error scaling	97.22 ± 0.02
Aligned	97.07 ± 0.02
Aligned + error scaling	97.04 ± 0.02

The performance of b -tagging was investigated with $t\bar{t}$ and $WH(m_H = 120 \text{ GeV})$ samples which are standard samples used in b -tagging performance studies in ATLAS as show in the last chapter. The $t\bar{t}$ sample includes semi-leptonic and di-lepton channels and the one sample is used for measuring both b -jet an light jet efficiencies. The $WH(120)$ sample contains two sub samples. $WH(120) \rightarrow \mu\nu b\bar{b}$ is used to measure b -jet efficiencies and $WH(120) \rightarrow \mu\nu u\bar{u}$ for light jet efficiencies.

4.3 Impact on Tracking and Vertexing

The tracking and vertexing performance was studied with the $WH(120) \rightarrow \mu\nu b\bar{b}$ sample, although the other sample could equally have been chosen. A sample size of 27,000 events was used. This sample was reconstructed using three of the alignment sets described previously: “Perfect”, “Random10” and “Aligned”. All other settings were kept the same. The samples were investigated with and without error scaling.

The track reconstruction efficiency was computed for each scenario by comparing truth tracks (i.e. from Monte Carlo) to corresponding reconstructed tracks. Tracks with $p_T > 1 \text{ GeV}$ and $|\eta| < 2.5$ were selected. A truth track was matched a reconstructed track and the efficiency was computed as the ratio of the number of matched tracks to the number of all truth tracks. The results for the efficiency for each of the three scenarios are shown in Table 4.4. The presence of residual misalignment in the “Random10” set causes a loss of about 2% in the efficiency, while the introduction of error scaling completely recovers the loss of performance. The “Aligned” set shows no significant change with respect to the “Perfect” case, with or without error scaling.

The number of fake tracks was also investigated in a similar manner to the track reconstruction efficiency calculation. A track was labeled as “fake” if it could not be matched to any track from Monte Carlo. The percentage of fake tracks from the total accepted tracks is shown in Table 4.5 for the different alignment scenarios. The misalignments result in more fakes and as with the efficiency this is recovered with the introduction of error scaling.

The performance of the primary vertex finding algorithm was also investigated. The efficiency for the primary vertex finding was computed as the ratio between the total number

Table 4.5: Ratio of fake tracks from the total accepted tracks.

Setup	fake tracks (%)
Perfect	2.33 ± 0.02
Random10	2.46 ± 0.02
Random10 + error scaling	2.29 ± 0.02
Aligned	2.34 ± 0.02
Aligned + error scaling	2.27 ± 0.02

Table 4.6: Primary vertex resolution.

Setup	res. in x (μm)	shift in x (μm)	res. in z (μm)	shift in z (μm)
Perfect	11.4 ± 0.1	-0.13 ± 0.07	51.1 ± 0.4	-8.2 ± 0.3
Random10	15.1 ± 0.1	4.2 ± 0.1	63.0 ± 0.4	1.4 ± 0.4
Random10 + error scaling	13.2 ± 0.1	2.6 ± 0.1	56.6 ± 0.4	2.3 ± 0.4
Aligned	13.9 ± 0.1	-0.18 ± 0.09	53.7 ± 0.4	-91.5 ± 0.4
Aligned + error scaling	13.8 ± 0.1	-0.15 ± 0.09	55.4 ± 0.4	-91.6 ± 0.4

of reconstructed vertices to the total number of true vertices. It was found that this remains constant, at a value of $99.7 \pm 0.4\%$, irrespective of the misalignment scenario considered.

The primary vertex resolution was evaluated by looking at the difference between the reconstructed and the true vertex position. The resulting distribution for x and z directions is displayed in Fig. 4.2. The results for the y direction were similar to that in the x direction. The introduction of residual misalignment causes the distributions to become wider as would be expected with a degradation of the hit resolutions. The shift for the hand-made sets is consistent with the shift of the entire pixel detector that was introduced. For the ‘‘Aligned’’ set a shift of about $90 \mu\text{m}$ is apparent. The reason is that no constrain was applied on the six degrees of freedom of the whole detector during the alignment procedure and so it can easily result in such a shift when comparing with truth information.

The values for the resolution are computed as the width of a Gaussian fit to the distributions in Fig. 4.2 and are shown in Table 4.6. The resolution is degraded by residual misalignment, for both x and z directions. Error scaling helps to partially recover the loss of performance for the ‘‘Random10’’ scenario.

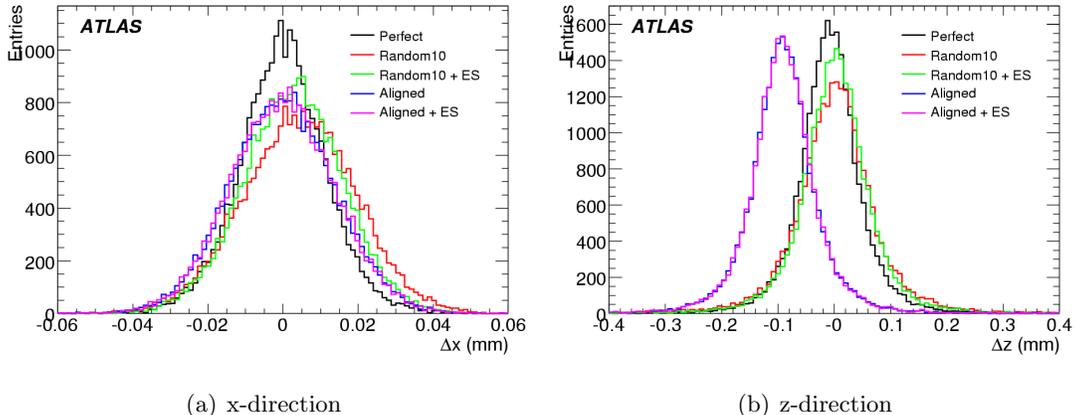


Figure 4.2: Primary vertex resolution, shown for the x direction (a) and z direction (b) for the various misalignment scenarios.

4.4 Impact on B-tagging

4.4.1 Light Jet Rejections

For the study of the impact of the residual misalignment on the b -tagging performance, several data sets were produced with $WH(m_H = 120 \text{ GeV})$ and $t\bar{t}$ events. Eight cases were considered corresponding to each specific scenario of residual misalignment (“Perfect”, “Aligned”, “Random10” and “Random5”) with and without error scaling. The performance of the b -tagging has been assessed by looking at the rejection rate of light quarks at b -jet efficiencies of 50% and 60% using various tagging algorithms: IP2D, IP3D, SV1 and the combined tagger IP3D+SV1.

For each of the scenarios using $WH(120)$ samples, 45,000 signal events ($WH(120) \rightarrow b\bar{b}$) and 175,000 background events ($WH(120) \rightarrow u\bar{u}$) were used. The $t\bar{t}$ samples contained 50,000 events each with the exception of the “Perfect” scenario without error scaling which had 570,000 events and the “Aligned” scenario with error scaling which had 500,000 events.

Figure 4.3 shows the b -jet efficiency versus light jet rejection for $t\bar{t}$ and $WH(120)$ samples for the four misalignment sets with and without error scaling. Rejections for the IP3D and IP3D+SV1 tagger at b -tag efficiencies of 50% and 60% are tabulated in Table 4.7 for $WH(120)$ and in Table 4.8 for $t\bar{t}$.

The results for the IP3D+SV1 are also summarized in Fig. 4.4. As expected, the larger the misalignment, the greater the degradation of the b -tagging performance. In the case of “Random10”, which represents the highest amount of misalignment (shifts of 10 μm in the pixel $r\phi$ measurement direction), there is almost a factor 2 drop in performance. For “Random5”, where the level of misalignment is lower (shifts of the order of 5 μm), the decrease of the light jet

Table 4.7: Light jet rejections rates computed for b -jet efficiencies of 50% and 60% for the IP3D and IP3D+SV1 taggers for the various misalignment scenarios with and without error scaling (ES) for $WH(120)$ events.

Setup	Rejection rate			
	IP3D (50%)	IP3D (60%)	IP3D+SV1 (50%)	IP3D+SV1 (60%)
Perfect	211 ± 4	67 ± 1	399 ± 11	104 ± 2
Perfect + ES	215 ± 5	67 ± 1	372 ± 11	98 ± 2
Random10	51 ± 1	23 ± 1	49 ± 1	21 ± 1
Random10 + ES	80 ± 1	29 ± 1	166 ± 3	49 ± 1
Random5	144 ± 3	49 ± 1	165 ± 3	53 ± 1
Random5 + ES	182 ± 7	53 ± 1	311 ± 16	80 ± 2
Aligned	193 ± 4	62 ± 1	300 ± 8	84 ± 1
Aligned + ES	190 ± 4	62 ± 1	306 ± 8	87 ± 1

Table 4.8: Standard light jet rejections rates computed for b -jet efficiencies of 50% and 60% for the IP3D and IP3D+SV1 taggers for the various misalignment scenarios with and without error scaling (ES) for $t\bar{t}$ events.

Setup	Rejection rate			
	IP3D (50%)	IP3D (60%)	IP3D+SV1 (50%)	IP3D+SV1 (60%)
Perfect	238 ± 11	68 ± 2	480 ± 30	166 ± 6
Perfect + ES	244 ± 11	70 ± 2	474 ± 30	161 ± 6
Random10	86 ± 2	32 ± 1	95 ± 3	38 ± 1
Random10 + ES	71 ± 2	25 ± 0	242 ± 11	77 ± 2
Random5	192 ± 7	56 ± 1	290 ± 14	95 ± 3
Random5 + ES	133 ± 4	46 ± 1	360 ± 20	116 ± 4
Aligned	234 ± 10	67 ± 2	442 ± 27	143 ± 5
Aligned + ES	206 ± 8	62 ± 1	428 ± 24	138 ± 5

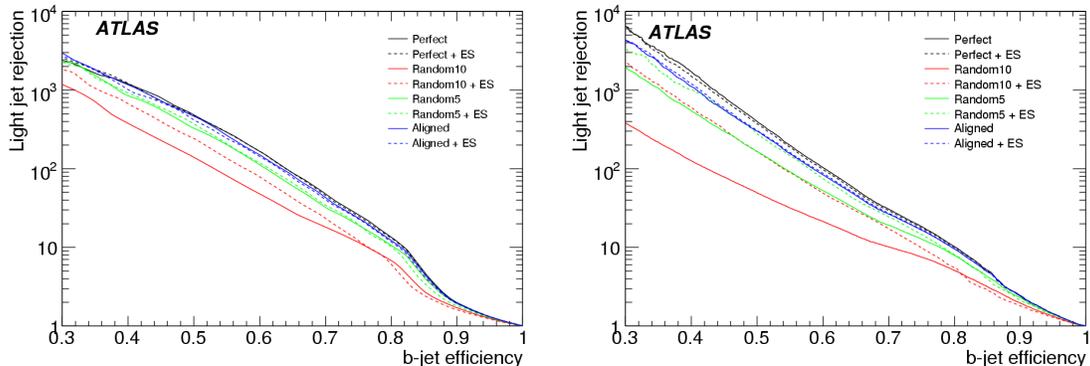


Figure 4.3: Light jet rejection versus b -tagging efficiency for the four different alignment sets for IP3D+SV1 for $t\bar{t}$ (left) and $WH(120)$ (right).

rejection rates is lower than in the previous case at around 30% degradation. For the “Aligned” set the loss in performance is around 10 to 20% and lies somewhere between the “Perfect” alignment and the “Random5” set. This is consistent with the level of misalignment suggested by the parameter c in the error tuning which is $3 \mu\text{m}$ in the pixel $r\phi$ measurement direction.

Figure 4.5 shows the b -tag weight for IP3D+SV1 tagger for the different alignment scenarios. The differences between the “Aligned” and “Perfect” sets are difficult to see in such plots but for the larger misalignments (“Random10” and “Random5”) it is seen that the light jets have slightly larger weights while the b -jets have lower weights resulting in the loss of discrimination.

4.4.2 Effects of Error Scaling

It is observed in Fig. 4.4 that for the larger misalignment scenarios (“Random10” and “Random5”) the error scaling gives a significant improvement, while for the “Aligned” and “Perfect” case the impact of error scaling is small.

Figure 4.6 compares the b -tag weights with and without error scaling. Only the “Random10” results are shown. For the other scenarios the difference were less pronounced. The $t\bar{t}$ events shows some differences in behaviour for the error scaling as compared with the WH events. For $t\bar{t}$, the error scaling results in only a small difference for the light jets while for the b -jets the difference are more evident. This is in contrast with the WH events where the light jets show more differences and the b -jets are less affected.

The differences are thought to be associated with the p_T spectra of light jets and b -jets which are different from each other and different for the two event types. The effectiveness of the error scaling is expected to have some p_T dependence since for lower p_T , the multiple scattering will dominate and differences in hit errors will be less important. There was insufficient statistics in

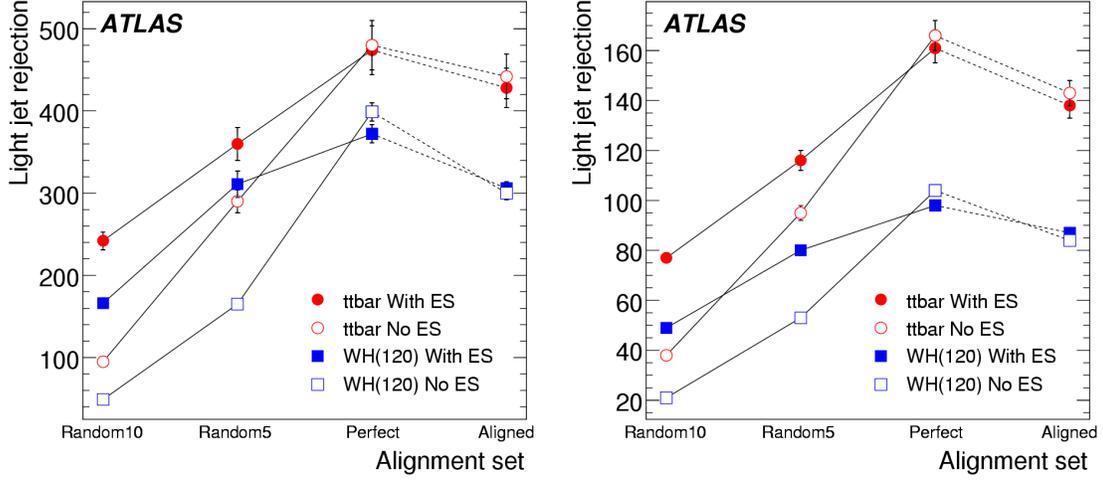


Figure 4.4: Light jet rejections using IP3D+SV1 tagger for the four misalignment scenarios at b -tagging efficiency working points of 50% (left) and 60% (right). Results are shown before and after error scaling (ES).

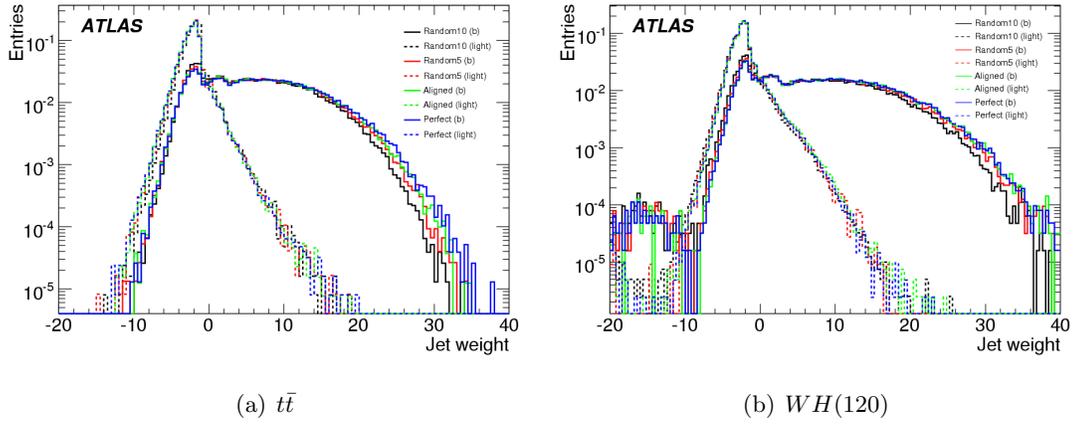


Figure 4.5: Jet weight distributions for the IP3D+SV1 tagger for the difference alignment scenarios with error scaling for $t\bar{t}$ (a) and $WH(120)$ (b).

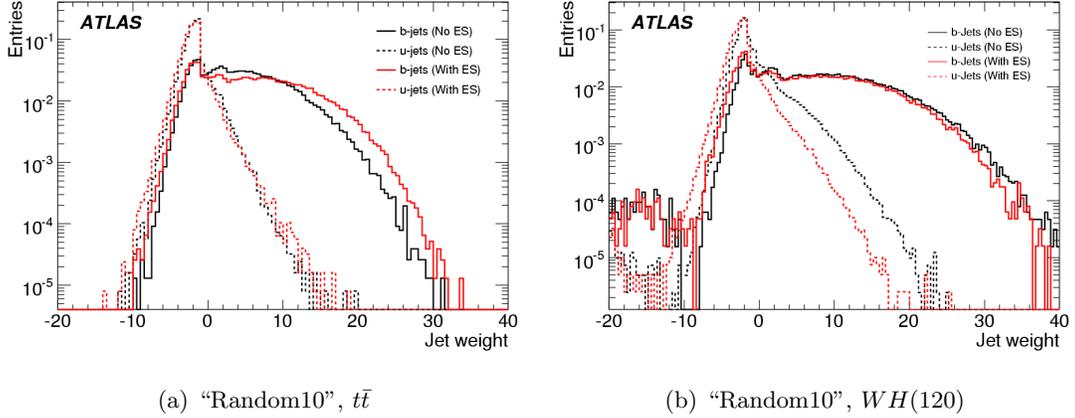


Figure 4.6: Jet weight distributions for the IP3D+SV1 tagger comparing with and without error scaling for “Random10” scenario for $t\bar{t}$ (a) and $WH(120)$ (b).

the samples however to verify if this was indeed the case.

4.4.3 Effect of Purification

For comparison with other studies, also the purified jets (see last chapter) were investigated. Table 4.9 shows the results for purified jets for $t\bar{t}$. For WH events only standard jets were considered as other studies [23] show similar results with and without purification. Figure 4.7 compares standard and purified jets for $t\bar{t}$ events and it can be seen that the rejections are higher for purified jets but the trends are similar for the different alignment scenarios. The degradation of the “Aligned” case with respect to the “Perfect” alignment is more pronounced after purification (19 – 23% degradation) than for the standard jets (10 – 14% degradation). The effects of error scaling showed similar behaviour for both standard and purified jets.

4.4.4 Comparison of the Different Taggers

Figure 4.8 shows the rejections for different taggers for standard jets. The impact parameter based taggers are the most affected by misalignment with the “Aligned” set showing 10 – 20% lower rejections than the “Perfect” case and up to a factor 3 degradation for the largest misalignment. After error scaling, the SV1 tagger shows rather uniform performance for all scenarios considered, with the “Aligned” set giving 10% degraded performance with respect to the “Perfect” case.

Without error scaling (see Fig. 4.9) the SV1 tagger shows significant differences for the different alignment scenarios. The ratio between rejections with error scaling to those without

Table 4.9: Purified light jet rejections rates computed for b -jet efficiencies of 50% and 60% for the IP3D and IP3D+SV1 taggers for the various misalignment scenarios with and without error scaling (ES) for $t\bar{t}$ events.

Setup	Rejection rate			
	IP3D (50%)	IP3D (60%)	IP3D+SV1 (50%)	IP3D+SV1 (60%)
Perfect	331 ± 19	80 ± 2	914 ± 86	243 ± 12
Perfect + ES	332 ± 19	80 ± 2	872 ± 79	234 ± 11
Random10	97 ± 3	34 ± 0	106 ± 3	41 ± 1
Random10 + ES	76 ± 2	26 ± 0	316 ± 17	89 ± 3
Random5	250 ± 12	62 ± 2	387 ± 23	113 ± 4
Random5 + ES	154 ± 6	50 ± 1	558 ± 41	148 ± 6
Aligned	321 ± 18	77 ± 2	714 ± 59	190 ± 8
Aligned + ES	273 ± 14	70 ± 2	706 ± 56	180 ± 7

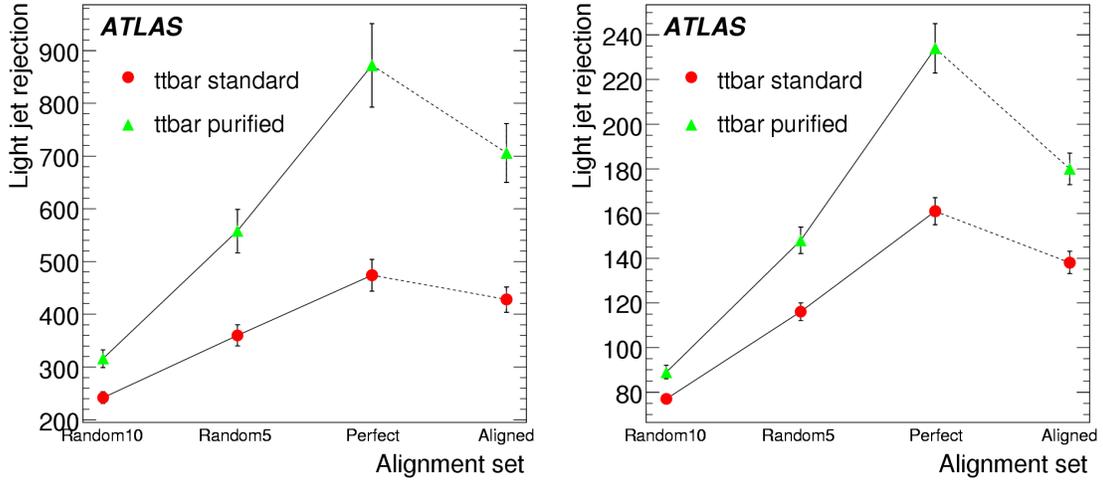


Figure 4.7: Comparison of light jet rejections for standard and purified jets using $t\bar{t}$ events. Left plot: 50% b -tag efficiency. Right plot: 60% b -tag efficiency.

error scaling is shown in Fig. 4.10. It can be seen that the error scaling has the most beneficial affect with the larger misalignments (“Random10” and “Random5”) for the SV1 performance. For the “Aligned” scenario the error scaling has little affect while for the “Perfect” case it actually degrades the performance slightly. For the impact parameter based taggers the error scaling has a smaller effect on the b -tagging performance and even degrades the performance in $t\bar{t}$ events.

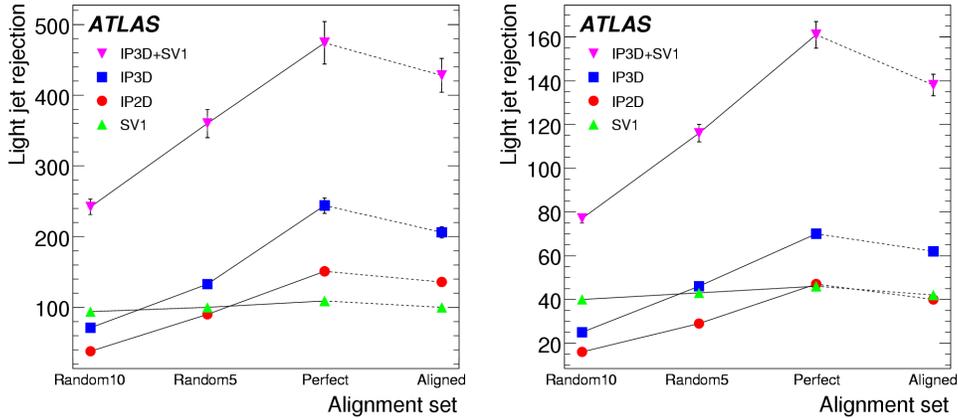


Figure 4.8: Comparison of the light jet rejections for the different taggers, IP2D, IP3D, SV1 and the combined tagger IP3D+SV1. Left plot: 50% b -tag efficiency. Right plot: 60% b -tag efficiency. Results are with error scaling using $t\bar{t}$ events.

For the combined tagger (IP3D+SV1) due to the beneficial effect of the error scaling on the secondary vertexing, the overall performance is improved with error scaling for the “Random10” and “Random5” sets. The error scaling parameter c is zero for the “Perfect” alignment and small for the “Aligned” set and consequently for both cases the effect of error scaling on the performance of b -tagging for the combined tagger is also small. For the “Aligned” case, while the relative difference is smaller than the hand made sets, the error scaling degrades the performance slightly.

The reason for loss of performance with error scaling for some cases can be explained by the following. Since the error scaling will generally increase the errors, it will reduce the impact parameter significance. This is desirable for light jets as it will make them more compatible with zero impact parameter. For b -jets, however, it also reduces the significance and so will reduce the b -tagging efficiency for a given weight cut or in other words one needs a lower weight cut to obtain the same efficiency and hence results in lower rejection for light jets for a given b -jet efficiency. The overall effect depends on these two competing effects and so can potentially lead to a loss in performance. While a decrease in performance with error scaling was observed in

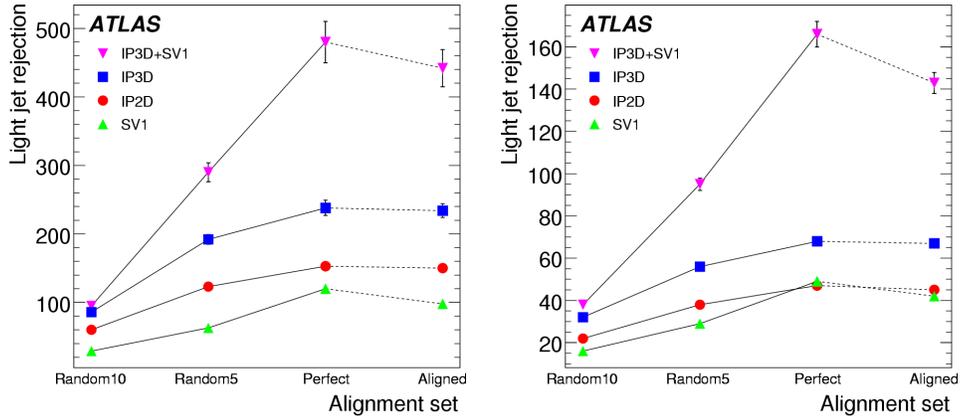


Figure 4.9: As in Fig. 4.8 but without error scaling.

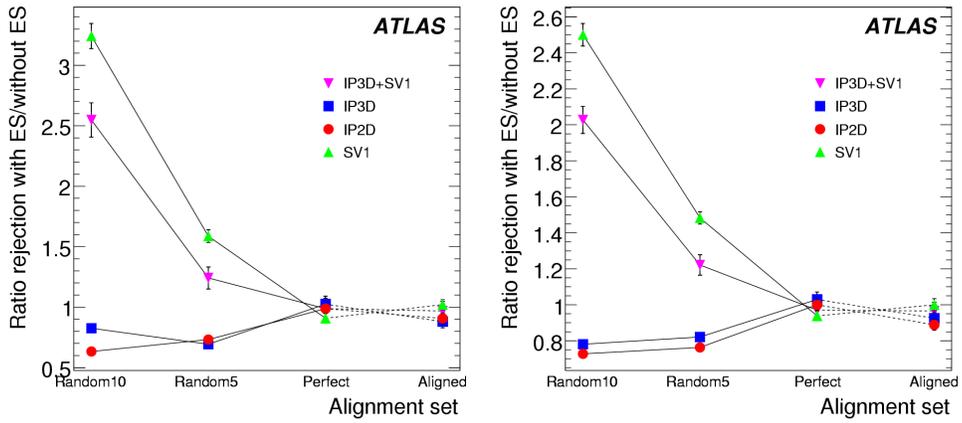


Figure 4.10: Ratio of rejections with error scaling to rejections without error scaling. Left plot: 50% b -tag efficiency. Right plot: 60% b -tag efficiency. Compares the various taggers.

$t\bar{t}$ events for the IP2D and IP3D taggers, the opposite was observed for WH events. As was already discussed above for the b -tag weight in Fig. 4.6, the error scaling affects differently the light jets and b -jets in the two physics sample and this is thought to lead to the differences in the error scaling behavior seen here.

The effect of error scaling on the secondary vertex tagger will be to recover some secondary vertices which would have otherwise failed quality cuts due to an underestimated error. One will lose some true secondary vertices that are close to the primary vertex as the larger error will make them compatible with the primary vertex, however for similar reasons it will result in less fake secondary vertices close to the primary vertex.

4.5 Recalibration

The taggers require prior probability distribution functions for light jets and b -jets as described before and the process of obtaining the set of these reference distributions is known as calibration. The results presented here use the same set of calibrations as used in Ref. [23]. Since misalignments will alter these distributions, it is possible that one can obtain some more discriminating power by recalibrating using the misaligned sample. Methods for obtaining these calibrations from real data are explored in Refs. [28].

To investigate whether recalibrating results in any gain in performance, a new set of reference distributions was obtained for each sample with and without error scaling. In practice one should use an independent samples to calibrate and to test the performance, however, here the reference distributions were obtained with the same or a subset of the sample used to measure the performance.

As seen in Fig. 4.11 recalibration gives better performance, although after error scaling the difference between the fixed calibration and recalibration is only marginal.

One might expect that recalibration may compensate any miscalculation of the errors in a similar way to the error scaling. This is only partially the case as can be seen in Fig. 4.12, where the relative improvement with error scaling is reduced when recalibrating.

4.6 Summary and Outlook

The effects of misalignment on the tracking performance as measured by tracking efficiency and fake rates was only small and error scaling recovered performance close to what was seen in the perfect alignment. The degradation of the primary vertex was more apparent with resolutions about $2.5 \mu\text{m}$ degraded and an increase in the number of outliers. The “Aligned” set showed similar performance to the “Random10” despite better b -tagging performance.

The performance of b -tagging was clearly degraded with misalignment and the amount of degradation was found to be roughly proportional to the amount of random misplacement of modules. Systematic effects are expected to also play a important role, however, the random

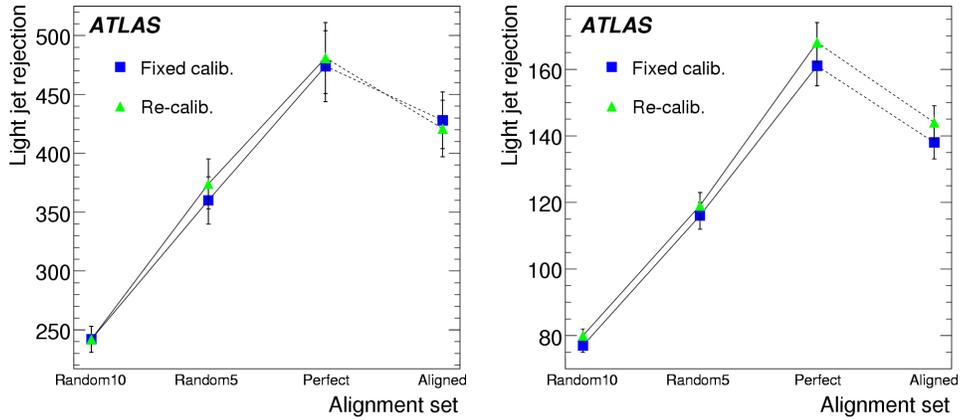


Figure 4.11: Comparison of using a fixed calibration or recalibrating for each separate sample. Left plot: 50% b -tag efficiency. Right plot: 60% b -tag efficiency.

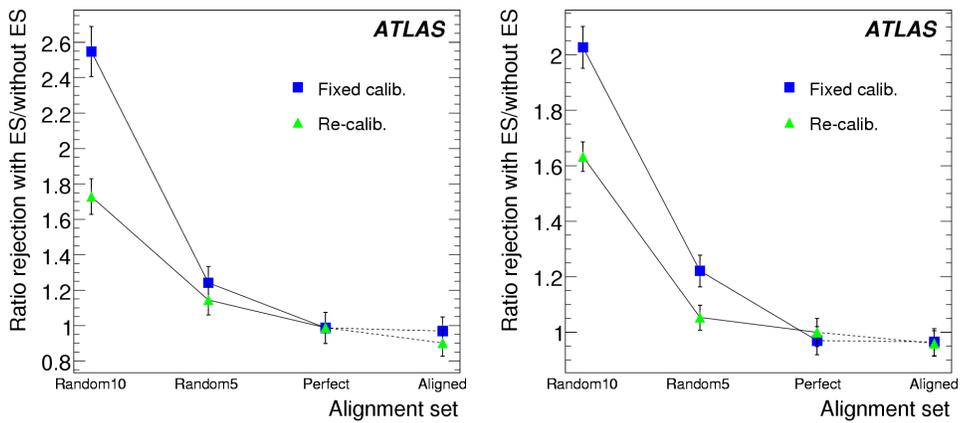


Figure 4.12: Ratio of rejections with error scaling to rejections without error scaling. Left plot: 50% b -tag efficiency. Right plot: 60% b -tag efficiency. Compares using a fixed calibrations and recalibrating for each separate sample.

misplacement was the only aspect that was quantified in this study by the parameter c obtained in the error scaling procedure. In order to disentangle the contributions from random and systematic effects it would be necessary to create dedicated residual misalignment sets with known systematic distortions, however, this was not explored in this study.

The impact parameter based taggers were observed to be the most affected by the misalignment and the introduction of error scaling had little or even negative impact on the b -tagging performance in the case of $t\bar{t}$ events. The error scaling was important for the performance of the secondary vertex finding, and without it, for larger misalignments the degradation for the secondary vertex based tagger was significant. With error scaling most of the degradation was recovered and the secondary vertex tagger showed uniform performance for all alignment scenarios consider. The behaviour of the combined tagger, IP3D+SV1, follows what one might conclude from the behaviour of the separate taggers, that is, showing a degradation with misalignment even after error scaling and also benefiting from using error scaling.

The “Aligned” set was the most realistic misalignment scenario looked at in these studies and the results were encouraging with rather moderate degradation to the b -tagging performance. In purified jets from $t\bar{t}$, the degradation was more evident with 19% degradation at 50% b -tagging efficiency. However, in a more realistic environment as seen by looking at standard jets, the amount of degradation was only 10%. At 60% b -tagging efficiency, the degradation was slightly higher with 23% degradation for purified jets and 14% degradation for standard jets. For WH events the degradation was similar with around 18% degradation at 50% b -tagging efficiency and 11% degradation at 60% b -tagging efficiency.

Chapter 5

Conclusions

Since its formulation, Standard Model has been our best description of the subatomic world. It is remarkable simple and principled, yet it has tremendous dynamical content and is able to explain nearly all observed phenomenology.

The crucial prediction of the Standard Model is the spontaneous symmetry breaking, i.e. the ground state does not possess all symmetry of the theory. The particle connected with this symmetry breaking is called Higgs boson and although large experimental effort has been made to find it, it has not been spotted so far.

The SM has also other weak points. For example it predicted that neutrino is exactly massless, while dedicated experiment found the opposite. Other strange thing about SM is rather haphazard assignment of charges to particles and non-simple inner symmetry group.

There are several proposals how to address this problems. In general they include some form of unifying the interactions at very high energy scale (GUT), or the prediction of new, superheavy particles (right-handed neutrino). However, it seems that these theories have effects even for larger energies and under some assumptions these effects can be distinguishable from the SM even at the reachable energy scale.

Promising theory is also Supersymmetry which predicts (broken) symmetry between fermions and bosons. This theory can address problem of scalar boson self-energy in SM and it can also, for free, modify the predictions of GUT to suit the latest experimental findings. Supersymmetry is also predicted by the superstring theory. Like in previous case, there is a good chance of seeing SUSY in reachable energy scales.

These are the main motivations for building the largest hadron collider in the world: LHC. It is being constructed in CERN and it should collide protons with center of mass energy equal to 14 TeV.

The outcome of the collisions will be measured by large detectors constructed around the collision points. The largest of them will be the experiment ATLAS. It has no particular focus, but rather it should measure outcome of as much collisions as possible as precisely as

possible. That is why it is composed of numerous subsystems which together are able to identify practically all SM particles.

The first physical program on ATLAS would be of course looking for the Higgs boson, but ATLAS should be also able to study many new physics effects like Supersymmetry, microscopic black holes or extra dimensions.

Because Higgs should be heavy particle and also most of the new models include high pT physics, the identification of jets containing b quark would be extremely important. This is called b -tagging and CSC studies in CERN showed that its usage significantly improves the quality of physical analysis.

B -tagging makes use of several b quark properties which make its unique among other quarks. There are several decision algorithms called “taggers”, which decide whether jet in question is b -jet. The most powerful are algorithms based on impact parameter of tracks in the jet because the B -hadrons decay outside primary vertex. Also the algorithms which try to reconstruct secondary vertex are quite successful. The most powerful is then combination of criteria based in impact parameter and secondary vertex properties. Such a choice is able to select b -jet sample with excellent purity both in few jet and multi jet environments.

B -tagging is very sensitive to the quality of track and vertex fitting. Because we know that ATLAS is assembled imperfectly, all parts of the detector suffer some level of misalignment. This is in general misplacement from the nominal positions as well as spatial deformations. However, we do not know the level of misalignment and so the tracking loses some performance. The subject of this study therefore was to find what is the effect of misalignment on b -tagging.

Because misalignment affects the determination of the tracking error, a procedure called error scaling was developed to prevent deterioration of tracking.

An effect of both misalignment and error scaling on b -tagging has been studied in two standard environments: Higgs boson produced together with W with only two jets coming from the Higgs decay and the multijet $t\bar{t}$ production. In the study several different levels of misalignment has been used.

The results showed that b -tagging performance decrease rapidly with the increasing level of misalignment, but this effect can be cured to a large extend using error scaling. The effect of error scaling was large for more misaligned environments. It also acted very differently on various taggers. The performance of impact parameter based taggers actually decreased after using error scaling, but the performance of combined tagger was increased, because error scaling had only positive effects on secondary vertex based taggers.

The most important were the results for the realistic alignment level measured recently by cosmic rays. The effect on b -tagging in this case was rather small, indicating that b -tagging could be used from the start of data taking.

The study was summarized in the CSC-BT2 note and was highly praised by the ATLAS community. It was also presented on workshops in Ringberg and Genova [30,31].

Bibliography

- [1] J. Hořejší, *Fundamentals of Electroweak Theory*, Karolinum Press, 2002
- [2] J. Chýla, *Quarks, partons and Quantum Chromodynamics*, lecture notes, <http://www-hep2.fzu.cz/Theory/notes/text.pdf>
- [3] W.-M. Yao et al., *Journal of Physics G* 33, 1 (2006)
- [4] Super-Kamiokande Collaboration, *Phys. Rev. Lett.* 81, 1562 (1998), hep-ex/9807003
- [5] M. E. Peskin, D. V. Schroeder, *Introduction to quantum field theory*, Westview Press, 1995
- [6] L. H. Ryder, *Quantum Field Theory*, 2nd edition, Cambridge University Press, 1996
- [7] R. T. Klauber, *The Seesaw Mechanism*, <http://www.quantumfieldtheory.info/TheSeesawMechanism.htm>
- [8] M. R. Buckley, H. Murayama, *How can we test seesaw experimentally?*, hep-ph/0606088v1
- [9] The LEP Collaborations, *A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model*, hep-ex/0612034v2
- [10] H. Georgi, S. Glashow, *Phys. Rev. Lett.* 32, 438 (1974)
- [11] K. Cranmer, *SM Higgs Combination*, Higgs WG meeting, 10 April 2008, CERN, <http://indico.cern.ch/conferenceDisplay.py?confId=31775>
- [12] D. Rainwater, *Searching for the Higgs boson*, hep-ph/0702124v1
- [13] S. Asai et al., *Eur. Phys. J. C* 32S2, 19 (2004).
- [14] http://commons.wikimedia.org/wiki/Image:LHC_octants.png
- [15] CMS Outreach, <http://cmsinfo.cern.ch/>
- [16] ALICE experiment: Panorama of ALICE, <http://aliceinfo.cern.ch/Public/panorama/>

- [17] LHCb public pages, <http://lhcb-new.web.cern.ch/>
- [18] LHC machine Outreach, <http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/>
- [19] ATLAS experiment public pages, <http://atlas.ch/>
- [20] The LXR Cross-Referencer, <http://alxr.usatlas.bnl.gov/lxr/source>
- [21] ROOT framework homepage, <http://root.cern.ch/>
- [22] Web interface to ATLAS offline software repository ViewCVS, <http://atlas-sw.cern.ch/cgi-bin/viewcvs-atlas.cgi/offline/>
- [23] L. Vacavant et al. *B-tagging performance*
- [24] V. Kostyukhin, C. Weiser et al., *Vertex reconstruction for b-tagging*, CSC Note BT1
- [25] G. Gorfine, M. Panušková, V. Sipica *b-tagging performance with ID misalignment*, CSC Note BT2
- [26] H. Bachacou et al., *Soft Muon Tagger: algorithm and performance*, CSC Note BT3
- [27] F. Derue, A. Kaczmarska et al., *Soft electron b-tagging*, CSC Note BT4
- [28] H. Bachacou, R. Hawkings et al., *b-tagging calibration with $t\bar{t}$ events*, CSC Note BT8
- [29] ALEPH Collaboration, *A precise measurement of $\Gamma_{Z\rightarrow b\bar{b}}/\Gamma_{Z\rightarrow hadrons}$* , Phys. Lett. **B313**, (1993) 535; D. Brown, M. Frank, *Tagging b-hadrons using track impact parameters*, ALEPH-92-135
- [30] G. Gorfine, *b-tagging - misalignment/error scaling*, 2nd Ringberg ID workshop, April 17, 2008, <http://indico.cern.ch/conferenceTimeTable.py?confId=23706>
- [31] M. Panušková, *Impact of alignment on b-tagging*, ATLAS b-tagging workshop, May 5, 2008, <http://indico.cern.ch/conferenceTimeTable.py?confId=27771>