

# b-tagging with DC1 data

S. Corr ard, V. Kostioukhine, J. Lev eque, A. Rozanov, J.B. de Vivie

25 November 2003

## Abstract

This note describes the analysis of the b-tagging performances of the ATLAS Inner Detector expected in WH and  $t\bar{t}H/t\bar{t}$  events. The simulation, reconstruction and analysis were done in the framework of the ATLAS Data Challenge. The effects of the changes in the inner detector (increase of the b-layer radius, more realistic material estimate, staging of the intermediate pixel layer, disk and TRT C-wheels, increase of the pixel pitch size in the b-layer from 300  $\mu\text{m}$  to 400  $\mu\text{m}$ ) were studied. The introduction of pile-up, detector inefficiencies and full vertex reconstruction brings the simulation results closer to the reality. New b-tagging methods show considerable improvements in the performances.



# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Detector layouts and data samples</b>	<b>2</b>
2.1	Detector description . . . . .	2
2.2	Simulation . . . . .	6
2.3	Data samples . . . . .	7
2.4	Reconstruction . . . . .	7
2.5	Performances of track reconstruction . . . . .	9
<b>3</b>	<b>b-tagging algorithms</b>	<b>10</b>
3.1	The transverse impact parameter b-tagging algorithm . . . . .	11
3.2	The 3D b-tagging algorithm . . . . .	13
3.3	b-tagging with secondary vertices . . . . .	13
3.4	Statistical errors of b-tagging results . . . . .	14
3.5	Perspectives for other methods and improvements . . . . .	14
<b>4</b>	<b>b-tagging with WH events</b>	<b>16</b>
4.1	DC1 versus Physics TDR layouts . . . . .	16
4.2	Staging of the intermediate pixel layer . . . . .	17
4.3	Impact of inefficiencies . . . . .	17
4.4	Impact of 300/400 $\mu\text{m}$ pixel pitch in the b-layer . . . . .	17
4.5	Impact of pile-up at low luminosity . . . . .	18
4.6	Secondary vertex b-tagging and rejection of interacting tracks . . . . .	19
4.7	Impact of primary vertex reconstruction . . . . .	20
4.8	Comparison between ATLFAST and Full Simulation jets . . . . .	21
4.9	Comparison between DC0 and DC1 data . . . . .	23
4.10	Realistic b-tagging and staging . . . . .	24
<b>5</b>	<b>b-tagging with <math>t\bar{t}H</math> and <math>t\bar{t}</math> events</b>	<b>25</b>
5.1	Calibration and overlapping jets . . . . .	25
5.2	Jet isolation . . . . .	26
5.3	Performance as a function of the b-jet origin . . . . .	27
5.4	Rejection as a function of the jet flavour . . . . .	28
5.5	Performances with the initial detector . . . . .	31
<b>6</b>	<b>Comparison between the <math>t\bar{t}H</math> / <math>t\bar{t}</math> and WH channels</b>	<b>34</b>
6.1	Global performances . . . . .	34
6.2	Differential performances with $p_T$ and $\eta$ . . . . .	34
6.3	Performance at high luminosity . . . . .	37
6.4	Parametrisation for ATLFAST . . . . .	38
<b>7</b>	<b>Effect of the Silicon Tracker misalignment</b>	<b>39</b>
<b>8</b>	<b>Conclusions</b>	<b>41</b>

# 1 Introduction

b-tagging is an important tool for many physics analyses at LHC, such as precision measurements in the top quark sector or Higgs boson and new physics searches. Since the Physics TDR publication [1], the Inner Detector (ID) geometry and the reconstruction software have been significantly modified. In particular, the pixel b-layer radius was increased and a realistic material estimate was introduced leading to an important degradation of the b-tagging performances. Some of these aspects were studied in [2]. Moreover, the effects of the potential pixel detector staging (one layer and two disks could be missing at start of data taking), and the modification of the pixel pitch size in the b-layer ( $400\ \mu\text{m}$  instead of  $300\ \mu\text{m}$ ) have to be considered.

This note aims at providing a realistic evaluation of the b-tagging performances expected for the initial and full versions of the ATLAS detector. The simulations include pile-up and detection inefficiencies, and the effect of  $300/400\ \mu\text{m}$  pixels is studied. Section 2 details the data samples and the detector versions used for the different analyses. Section 3 presents the different b-tagging methods used in this work. The results are given in Section 4 for the “reference” channel WH, with the standard 2D algorithm and for the improved methods including secondary vertices. Section 5 presents the b-tagging performances expected in the more realistic  $t\bar{t}H$  and  $t\bar{t}$  processes. Section 6 gives a comparison between the WH and  $t\bar{t}H/t\bar{t}$  channels and the parametrisation of the b-tagging performances for the fast simulation with the ATLFAST [3] package. Expected performances at high luminosity are also given in this section. Section 7 presents the effects of the alignment and commissioning on b-tagging.

## 2 Detector layouts and data samples

The baseline of the ATLAS Inner Detector is described in the Inner Detector TDR [4] and Pixel Detector TDR [5]. The reference for the tracking and b-tagging performances is the Physics TDR [1], which was followed by several studies in the framework of the ATLAS Data Challenges (DC) [6].

### 2.1 Detector description

Since the Physics TDR publication most of the changes occurred in the Pixel Detector:

- The b-layer radius was increased from 4.3 to 5.0 cm to accommodate a larger beam-pipe.
- Fully Insertable Layout on Pixel Support Tube (PST) was introduced.
- The Pixel system is integrated and installed with the beam pipe.
- Installation and upgrades of the pixel system are possible independently of other items of the Inner Detector.
- Pixel services are fully located inside the Silicon Strip Tracker (SCT) endcap and outside the coverage of the inner detector ( $|\eta| < 2.5$ ).
- Pixel material at  $\eta = 0$  was increased by a factor 1.5 due to thicker sensors and the evolution of module design and services.
- Ganged and long pixels were introduced in the inter-chip areas of the pixel modules.

- Most of the geometrical parameters like azimuthal tilt angles, polar inclination angles, module overlaps, radial positions were modified.
- Most of the services and mechanic elements were changed.

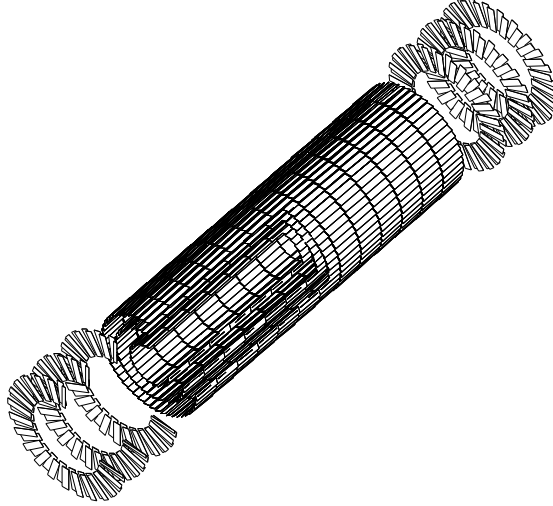


Figure 1: Stereo view of the full Pixel detector.

The stereo view of the latest pixel subsystem layout is shown on Fig. 1. The evolution of Pixel Layouts since Physics TDR is given in the Table 1.

	Physics TDR	Insertable DC0-DC1	Initial
b-layer R (cm)	4.0	5.05	5.05
layer-1 R (cm)	11.0	8.85	staged
layer-2 R (cm)	14.2	12.25	12.25
Tilt angle (deg)	-9.5	-20.0	-20.0
Tilt angle b-layer (deg)	-10.5	-20.0	-20.0
dZ/dR angle (deg)	0.0	1.1	1.1
Pitch $R\phi$ ( $\mu\text{m}$ )	50	50	50
Pitch $Z$ ( $\mu\text{m}$ )	300	400	400
Pitch $Z$ b-layer ( $\mu\text{m}$ )	300	300	400
Big pixels ( $\mu\text{m}$ )	no	600	600
Big pixels b-layer ( $\mu\text{m}$ )	no	500	600
Ganged pixels	no	yes	yes
Sensor thickness ( $\mu\text{m}$ )	150	250	250
Sensor thick. b-layer ( $\mu\text{m}$ )	150	200	250
Number of disks	4	3	2

Table 1: Evolution of Pixel Layouts since Physics TDR.

Many small changes in the SCT and Transition Radiation Tracker (TRT) were introduced due to design changes or more detailed descriptions. In particular the following modifications

were made:

- The SCT tilt angle sign was reversed.
- TRT straw modular geometry was introduced.
- More realistic service, connector, thermal enclosure descriptions were included.

The azimuthal and longitudinal cuts of the Full ATLAS inner detector in DC1 layout are shown on Fig. 2 and 3 respectively.

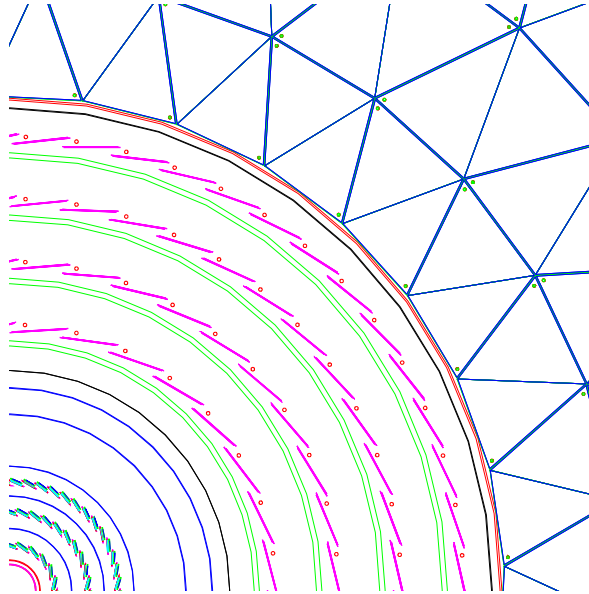


Figure 2: Transverse view of the full Inner Detector.

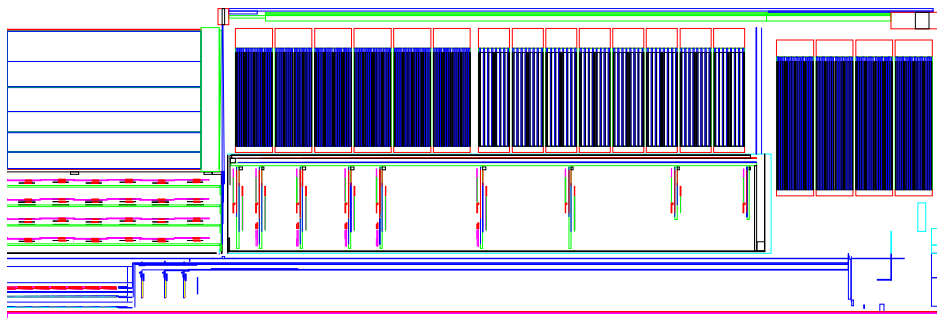


Figure 3: Longitudinal view of the full Inner detector.

Several financial and schedule constraints could eventually result in the staging of some ID components. In the present scenario the most important items proposed for staging in the Initial ATLAS are:

- Pixel Barrel Layer-1  $R = 9$  cm,

- Pixel End-cap Disk-2 at  $z = 58$  cm,
- Transition Radiation Tracker C-Wheels.

The present projection of how the Initial ATLAS ID could be in April 2007 at the beginning of data taking is presented on Fig. 4.

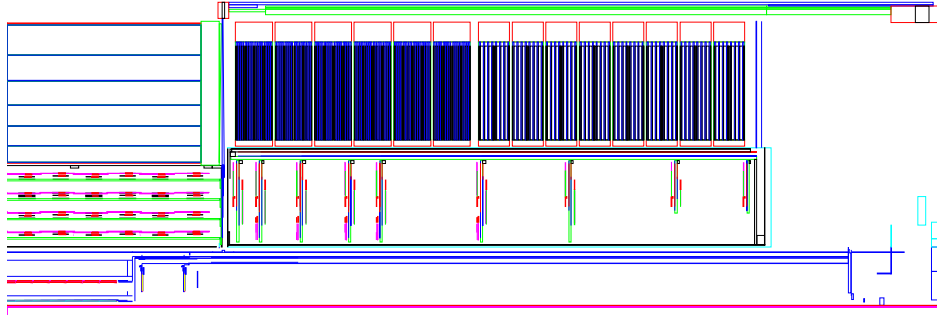


Figure 4: Longitudinal view of the initial Inner detector.

The accurate description of the passive material of the ID is very important for all kinds of realistic simulations, but in particular for b-tagging studies. The latest engineering data were used for the settings of the DC1 simulation layout. However one should expect more updates to be done, as detectors enter into production stage. Produced parts will be weighted and confronted with the expected values. At last more precise data are needed for the final material used in patch panels and moderators. Material changes between the Physics TDR

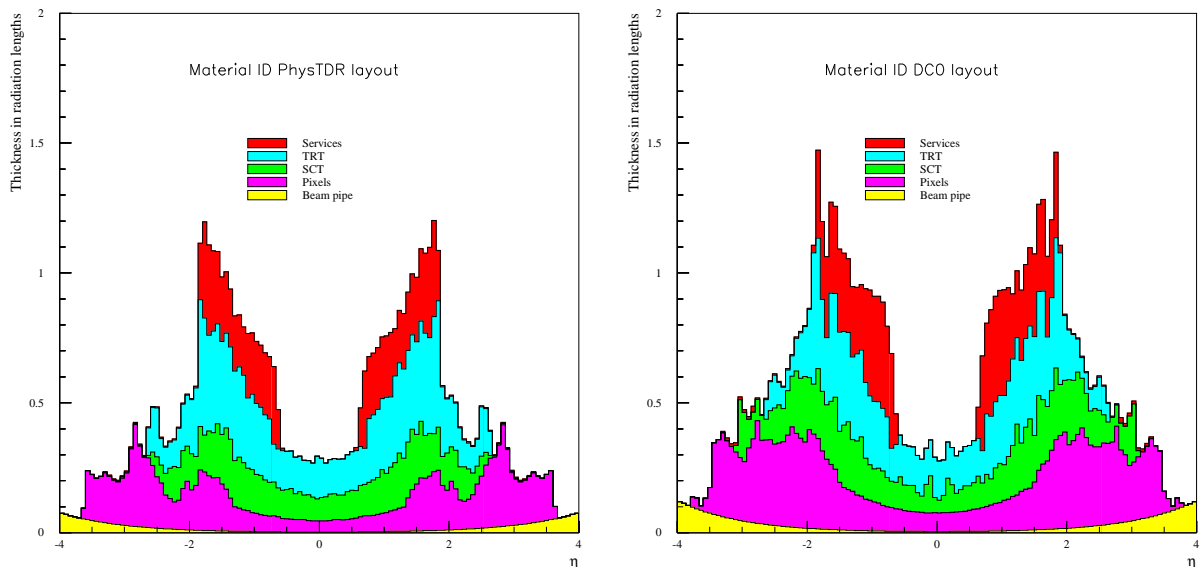


Figure 5: Material in % of radiation length as a function of pseudo-rapidity for Physics TDR (left) and DC1 (right) layouts.

and DC1 layouts can be seen by comparing the total material thickness in units of radiation length as a function of pseudo-rapidity as illustrated on Fig. 5. The service material increase

at high radius or  $z$ -coordinate has little impact on ID performances and b-tagging as it is outside the ID coverage. However a small increase of the rates in TRT wheels and a higher radioactive activation of the detector can be expected.

The main effects of the changes between the Physics TDR and DC1 layouts for b-tagging performance are due to:

- increase of the b-layer radius
- increase of the b-layer material

They result in:

- At low  $p_T \sim 1 \text{ GeV}/c$ :
  - degradation of the momentum resolution by a factor 1.2
  - degradation of the impact parameter resolution by a factor 1.6
- At high momentum the performances are almost identical, but small degradations due to an increase of secondary interactions are expected.

## 2.2 Simulation

DC1 is a new step toward more realistic simulations. Several additional effects were taken into account after the Physics TDR:

- Latest versions of the geometry, material budgets and digitisation of the detectors were used,
- Simulations were done with a non-uniform solenoid field of 2T, reduced to 1T at  $z = 2.7 \text{ m}$ ,
- $K_s^0, \Lambda, \Sigma, \Xi, \Omega^-$  were allowed to fly in the detector during GEANT3 (G3) simulation, to interact with the material and to give hits in sensitive detectors,
- Systematic studies with new low luminosity (4.6 events per bunch crossing on average) and high luminosity pile-up were done,
- New pile-up scheme in the calorimeters and muon system was introduced. In this scheme the pile-up is mixed with the main event at the level of G3 hits, therefore full correlation between ID and calorimeter pile-up is assured,
- Inefficiencies of modules and chips were introduced,
- Alignment effects and commissioning scenario were studied,
- New neutron background pile-up in the muon system was introduced after DC1 simulation. It was not used in this note, but should be introduced in the next generation of the full simulation b-tagging studies.

The primary Monte-Carlo events were generated for DC1 by PYTHIA 6.203 under the ATHENA framework and stored in ROOT format. CTEQ5L structure functions were used for parton distributions. The underlying events were treated in PYTHIA by the multiple parton interaction model tuned to CDF data. Full Monte Carlo events were simulated in the ATLAS detector with the G3 based ATLSIM/DICE program (version 3.2.1) with as input the event kinematics from ROOT-format files and as output the detector hits and digits in ZEBRA-format files. A part of the data was simulated with the ID only.

## 2.3 Data samples

Associated WH production processes with light ( $m_H = 120 \text{ GeV}/c^2$ ) or heavy Higgs ( $m_H = 400 \text{ GeV}/c^2$ ) boson hereafter labelled WH(120) and WH(400) are traditionally used as benchmarks for b-tagging studies at LHC. The typical data samples for the WH process contain 20K events with  $b\bar{b}$  Higgs boson decays for the b-jet signal and 50K or 100K events with  $u\bar{u}$  Higgs boson decays for light jet background. For the  $t\bar{t}H$  production ( $m_H = 120 \text{ GeV}/c^2$ ) the typical samples contain 20K events, while for the  $t\bar{t}$  background, the sample size is 100K events.

The exact number of simulated events in each data sample and the detector configuration are given in the Appendix in the Tables 28, 29, 30. The reason to have some data samples simulated with the Inner Detector only is to have high statistics with relatively small CPU consumption. It takes about 43 seconds CPU to simulate one  $t\bar{t}H$  event in the ID only with a PIII-1Ghz processor, while the full simulation with the calorimeters and the muon system takes 1010 seconds, *i.e.* 23 times longer. Most of the datasets were produced with the Full ATLAS setup with or without the intermediate pixel barrel layer. Some of them were also simulated for the Initial ATLAS detector.

The new low luminosity  $\mathcal{L} = 2 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$  and the “standard” high-luminosity  $\mathcal{L} = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  were used in these simulations. The main repository for all b-tagging data samples is on the HPSS system at the Computing Center of IN2P3 at Lyon in the directories: `/hpss/in2p3.fr/group/atlas/dc1` and `/hpss/in2p3.fr/group/atlas/cppm/dc1`.

The data samples simulated at Lyon have the simulation conditions stored in the ATLAS VDC database. All dataset properties are stored in the ATLAS bookkeeping database AMI and in the GRID oriented MAGDA database. All three databases are using MySQL.

## 2.4 Reconstruction

The track reconstruction was done with the xKalman++ pattern recognition program [7] (version xKalmanppAtrecon-00-00-74) under the ATLSIM/ATRECON framework. A minimum transverse momentum of  $0.7 \text{ GeV}/c$  and at least seven precision hits were required during the track search at low luminosity. For the reconstruction of the high luminosity data, the cuts were increased to  $0.9 \text{ GeV}/c$  and 8 hits. The selectivity of the track finding was set to 2, which means a reasonably high number of combination of track candidates. The internal xKalman magnetic field interpolation was used. An increased pixel wafer buffer size of 4096 digits (axxxrec-00-05-11, starting from release 6.1.0) was used, which eliminates the effects of buffer overflows observed in preliminary studies.

### Tracks

Tracks are of course the main objects to perform b-tagging. In order to suppress the contribution from badly reconstructed, fake or pile-up tracks, four levels of cuts need to be applied:

1. Transverse momentum cut  $p_T \geq 1 \text{ GeV}/c$
2. Standard quality cuts
  - transverse impact parameter  $|a_0| \leq 1 \text{ mm}$  ;
  - at least seven (eight at high luminosity) precision hits (pixels + SCT) ;
  - at least two pixel hits, one of which must be in the b-layer ;



### 3. Additional quality cuts

- track fit quality  $\chi_{fit}^2 \leq 3$  ;
- $|z_0 - z_{\text{vertex}}| \sin \theta \leq 0.15$  cm. This cut is used against tracks from pile-up ;
- no shared hit in the pixel detector, no more than one in the SCT ;

### 4. No ambiguity in the first pixel wafer.

Tracks satisfying all these requirements are referred to as *good* tracks hereafter.

## Primary vertex

The primary vertex is needed for the “additional quality cuts” and the 3D and Secondary Vertex (SV) algorithms described later. Several primary vertex estimations were used in the b-tagging studies:

- Beam position in transverse direction and Monte Carlo Truth in  $z$ -coordinate
- Monte Carlo Truth vertex
- Primary vertex reconstructed by the xKalmanpp package
- Primary vertex reconstructed by the VKalVrt package [9]
- Beam position in transverse direction and  $z$ -coordinate reconstructed by the VKalVrt package

By default, unless otherwise specified, we use in this note the beam position for the transverse primary vertex and the  $z$ -coordinate from VKalVrt for the longitudinal primary vertex.

## Jets

The jet direction is needed to sign the impact parameter and attribute tracks to the jet. The following jet definitions were used in different studies:

- Jets made from all stable particles (GENZ seeds). This jet definition was used in the Physics TDR and various studies afterward.
- ATLFAST jets.
- FullSim - jets reconstructed in ATRECON from the Full simulation of calorimeters.

Both for ATLFAST and FullSim jets the “crude” cone algorithm was used with a cone size  $\Delta R = 0.4$ . A cut on the transverse momentum of the jet  $p_T > 15$  GeV/ $c$  was applied. In this note, unless otherwise specified, ATLFAST jets are used.

## 2.5 Performances of track reconstruction

As will be explained in Section 3, the track impact parameters are among the most important variables for b-tagging. Fig. 6 (upper plots) illustrates the resolution on the transverse impact parameter  $a_0$  for all good primary tracks in  $W(H \rightarrow u\bar{u})$  (120) DC1 events as a function of  $|\eta|$ . The left plot shows the resolution for the Perfect DC1 complete detector, without any module/chip inefficiencies at 0 luminosity. The middle plot corresponds to the staged Initial detector (cf. Section 4.2). The right one is given for a *realistic* DC1 detector, labelled Real DC1, (which corresponds to a possible upgrade of ATLAS detector) with three pixel layers, a z pitch of  $300 \mu\text{m}$  in the b-layer, module/chips inefficiencies of 1/2%, at  $\mathcal{L} = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  (cf. Section 6.3). No big degradation is expected for the resolution on  $a_0$  when only the initial detector is available since although the pattern recognition is more difficult, the material budget is decreased. The lower plots in Fig. 6 give the corresponding resolution for the z coordinate of the track at closest approach to the beam line. In this case, the impact of the increased z pitch for the b-layer pixels (initial detector) is visible, *e.g.* in high  $p_T$  tracks at  $|\eta| = 0$ , with a 25% degradation.

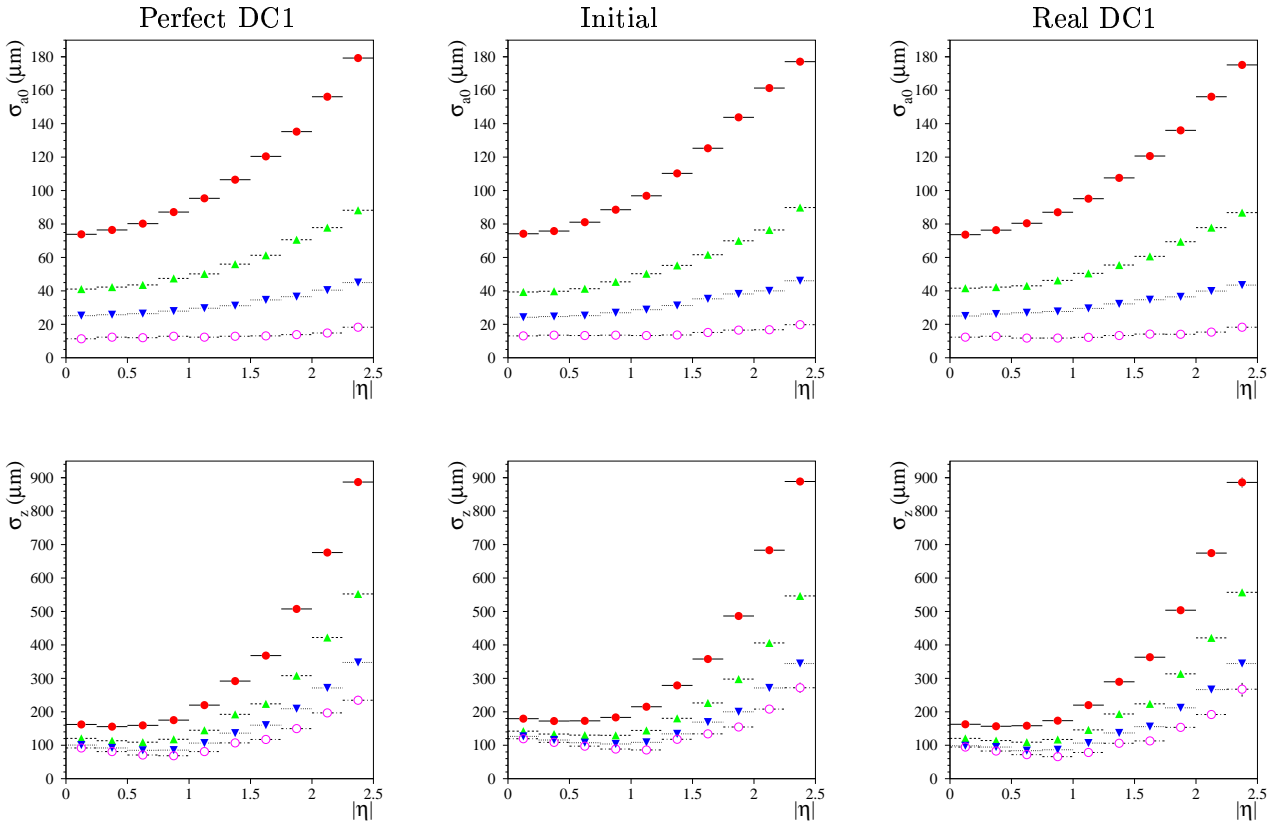


Figure 6: As a function of  $|\eta|$ , resolution on the transverse impact parameter (upper plots) and the z coordinate of tracks at closest approach to the beam line (lower plots) for the Perfect DC1 (left), Initial (middle) and Real DC1 (right) detectors. The different symbols correspond to different  $p_T$ : 1-2, 2-5, 5-10,  $\geq 50$  GeV/c from the upper to the lower curves.

The primary track reconstruction efficiency is shown in Fig. 7 as a function of  $|\eta|$  for different levels of quality cuts. Only tracks which may give at least five hits in the SCT are used for the normalisation (*i.e.* tracks giving nuclear interactions in the pixel detector or before the third SCT layer are not considered). The average efficiency is  $\sim 91\%$  for the

perfect detector, and  $\sim 79\%$  for the realistic initial detector. The degradation is mainly due to the requirement of at least two hits in the pixel detector. With only two layers, any further pixel, chip or module inefficiency implies a loss of the track. A tuning of the track selection criteria should be done for the initial detector. The impact of the “no ambiguity” cut is also higher in the initial detector case, since the hit multiplicity is higher when pile-up is added and the probability to pick a wrong hit increases due to the missing information from the intermediate staged pixel layer. For the Real DC1 detector, the average efficiency is  $\sim 83\%$ . The main losses are due to the increased module/chip inefficiencies and to the cut on the longitudinal impact parameter, when a wrong vertex from pile-up is chosen instead of the hard event one. The lower plots in Fig. 7 illustrate the fake track fraction for the same quality cuts. A track is considered as fake if less than 50% of precision hits originate from a single MC track. Fake good tracks are very rare (with a rate below the 0.02% level) over all the  $|\eta|$  range. When no quality cut is applied, a relatively high rate of fake tracks is observed. This however is expected since a loose cut on the number of precision hits has been chosen.

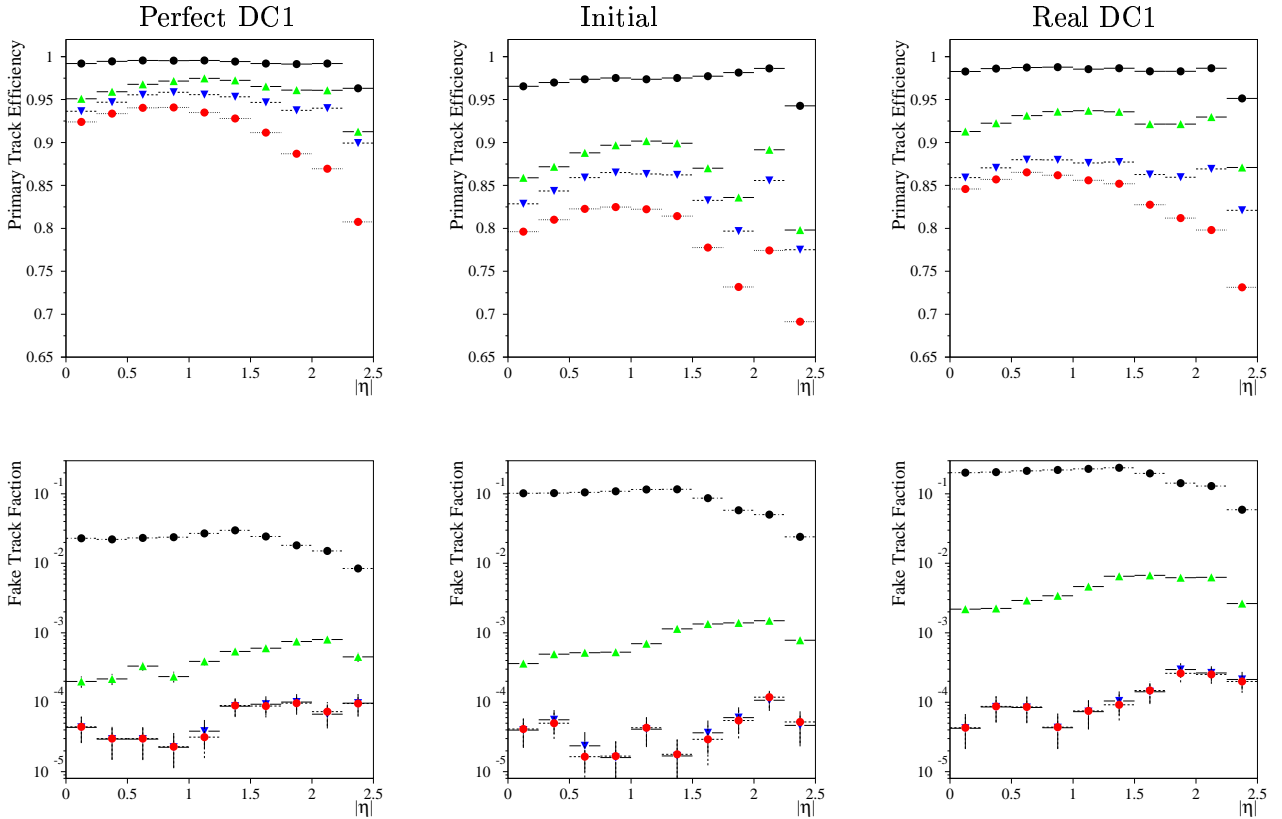


Figure 7: As a function of  $|\eta|$ , primary track reconstruction efficiency (upper plots) and fake track fraction (lower plots) for different levels of quality cuts for the Perfect (left), Initial (middle) and Real DC1 (right) detectors.

### 3 b-tagging algorithms

All space b-tagging algorithms exploit the long lifetime of B-hadrons. Inside a b-quark jet, a B-hadron may decay far from the primary vertex. The charged particles coming from that decay may have a non-negligible impact parameter and form displaced secondary vertices. In this note, only space b-tagging is considered, leaving apart soft lepton tagging methods.

The general approach to the space b-tagging is the following:

- The first b-tagging algorithm uses the transverse impact parameter  $a_0$ , defined as the closest distance between the primary vertex and track helix in the transverse plane. The impact parameter is signed according to the jet direction. The lifetime of B-hadrons is reflected in Fig. 8a as a long tail at positive values in the track transverse impact parameter distribution. On the contrary tracks from light jets give a narrow and symmetrical distribution.
- The longitudinal impact parameter  $z_0$  can be also used for b-tagging, but is less discriminating due to a worse resolution. It is usually defined at the trajectory point with smallest transverse impact parameter. Despite larger absolute values (Fig. 8b), the discriminating power of this variable is smaller than the one of the transverse impact parameter.
- The distributions of the impact parameters are used to construct likelihood functions for tracks coming from the primary interactions (small  $|a_0|$  and  $|z_0|$ ) or from a B-particle decay (large  $|a_0|$  and  $|z_0|$ ). The “likelihoods” of all good tracks inside the jet are combined into a “probability” for the jet to be a b-jet or a light jet.
- If there are two or more tracks in a jet with significant impact parameters, one can search for a secondary vertex displaced from the primary vertex. Some discriminating variables of the secondary vertex can then be combined into the jet “probability”.

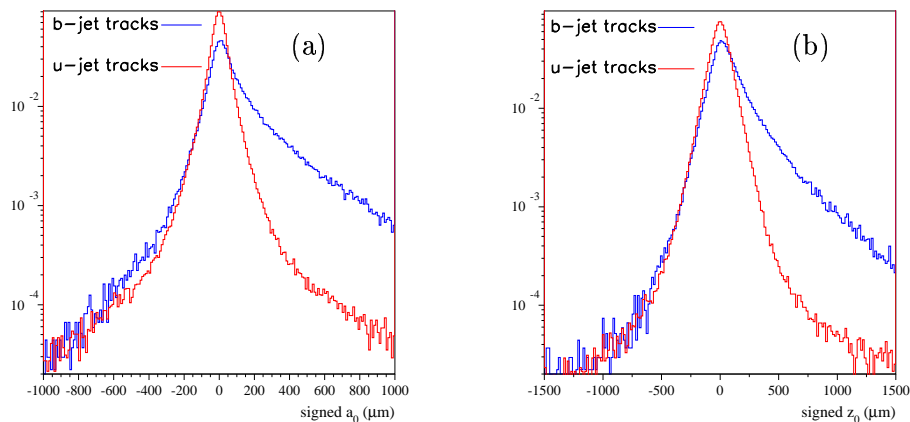


Figure 8: Transverse (a) and longitudinal (b) impact parameter distributions for all good tracks (WH(120) events).

In the following sections the details of how the likelihood functions are constructed and used in different b-tagging algorithms are given.

### 3.1 The transverse impact parameter b-tagging algorithm

Tracks from B-hadron decays are expected to have on average a large and positive transverse impact parameter  $a_0$ . The sign of the impact parameter is positive if the track crosses the jet axis in front of the primary vertex and negative otherwise (see Fig. 9). To give more weight to precisely measured tracks, the significance  $S$  of the impact parameter is used to

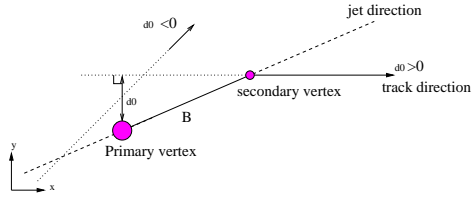


Figure 9: Definition of the transverse impact parameter  $|a_0|$  and its sign.

disentangle b-jets from light jets:

$$S_{a_0} = \frac{a_0}{\sigma_{a_0}} \quad \text{where } \sigma_{a_0} \text{ is the error on } a_0.$$

The distributions of the track significance are shown in Fig. 10a for tracks from b and u jets. The large tails for tracks from b-jets reflect the long lifetime of B-hadrons. After smoothing these histograms are used to construct the likelihood functions  $P_b$  and  $P_u$  for a track to originate from a b-jet or a u-jet. A track weight  $w_t$  is then computed as the ratio of the probabilities:

$$w_t = P_b(S_{a_0})/P_u(S_{a_0})$$

The ratio of the probability density functions represents the test statistics with which one

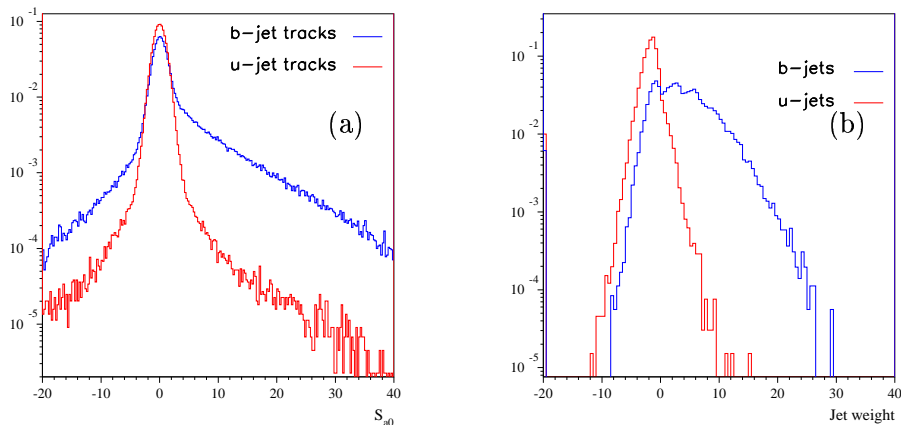


Figure 10: Distributions of the transverse impact parameter significance for tracks in light (udsg) and b jets (a) and of the jet weight for light (udsg) and heavy (b) flavour jets (b).

may obtain the highest purity sample for a given signal efficiency according to Neyman-Pearson lemma [8]. At last, a jet weight  $w_{jet}$  is defined as the sum of the logarithms of the track weights for all *good* tracks in the jet:

$$w_{jet} = \sum_{i \in jet} \ln w_t^i$$

The distributions of these weights are illustrated in Fig. 10b for b and u jets. By applying different cuts on this variable, a rejection factor  $R_j$  is obtained for light jets (u, d, s quark or gluon) for a given b-jet selection efficiency  $\epsilon_b$ . The same procedure, with the same likelihood

functions is applied to determine the c-jet rejection  $R_c$ . This algorithm (called the 2D algorithm) is simple, robust and since the beam spot at LHC is small ( $\sim 15 \mu\text{m}$ ), it is efficient even without the primary vertex reconstruction. However with the real experimental data, the knowledge of the position of the beam spot is needed from a special calibration. In the DC1 data the beam spot position was taken from the Monte Carlo to be (0.,0.), assuming that the error on the mean transverse beam position is negligible. The 2D algorithm can be used for all b-jets with at least one charged track with significant impact parameter.

### 3.2 The 3D b-tagging algorithm

Although the resolution on the longitudinal impact parameter  $z_0$  is much worse than the one on the transverse impact parameter, an important improvement of the performances can be expected by combining the *longitudinal significance* with the *transverse* one (the 2D+1D algorithm, labelled 3D hereafter). The jet weight is now a 2-dimensional function:

$$w_t = P_b(S_{a_0}, S_{z_0})/P_u(S_{a_0}, S_{z_0})$$

For the time being, in order to have enough statistics to build the likelihood functions, a binning of the plane  $(S_{a_0}, S_{z_0})$  has been adopted, as shown in Fig 11. In the future a smooth function should be implemented.

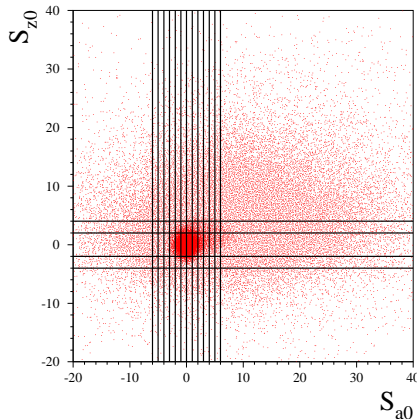


Figure 11: Definition of the binning of the  $(S_{d_0}, S_{z_0})$  plane used to compute the likelihood functions in the 3D method.

### 3.3 b-tagging with secondary vertices

The power of the 3D b-tagging method can be significantly improved by using the additional information on the presence or absence of a secondary vertex in the jet. This method starts with a primary vertex reconstruction, whose quality is crucial in the presence of many pile-up vertices. In this study, the VKalVrt package [9] is used for the primary vertex reconstruction. This package is also used for the search of the secondary vertices. There is no assumption on the exact topology and multiplicity of the secondary vertices inside the jet, therefore a method of inclusive vertex search was developed. The search starts with a selection of all track pairs which form good two-track vertices (see Ref. [10] for details). At a first stage these two-track vertices are used to reject “bad” tracks coming from secondary interactions in the volumes with dense material (beam pipe, b-layer, etc) or coming from  $K^0$  and  $\Lambda$  decays. At

the second step all tracks in two-track vertices are combined into a single effective secondary vertex. Due to the cascade B to D decays, this effective secondary vertex corresponds to some mean position between the B-hadron and D-hadron decay points.

In order to combine secondary vertex information with track impact parameter significances one needs some discriminating variables, which do not have too strong correlations with the impact parameters. In these studies the following variables were used:

- The fraction of the jet energy in the secondary vertex, Fig. 12a
- The invariant mass of all particles in the secondary vertex, Fig. 12b
- The number of two track secondary vertices, Fig. 12c

This method contributes only if there are at least two charged tracks reconstructed from the same B/D-hadron decay. Used in combination with 3D b-tagging, the SV method improves the performance by a factor 2-3.

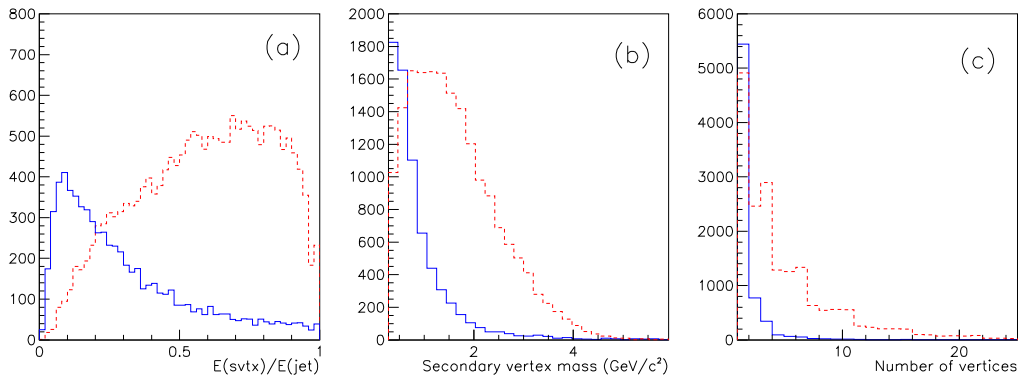


Figure 12: (a) Fraction of the jet energy coming from the secondary vertex. (b) Mass of the secondary vertex in the jet. (c) Number of two track secondary vertices in the jet. All distributions are given for b-jets (dashed line) and light jets (solid line).

However this method is very sophisticated and sensitive to the quality of the software for track and vertex reconstruction as many correlated effects enter in the final results. Therefore both the results of the much simpler methods with transverse and longitudinal impact parameters and the results with secondary vertices are presented in this note.

### 3.4 Statistical errors of b-tagging results

Many results in these studies are presented as ratios of the light jet rejections for different setups or conditions. In this case the same events are used, leading to strong statistical correlations between results. Due to the presence of these correlations and the complexity of the computational procedure, the error on the ratios of rejections can not be easily estimated. In this note we present the mean values of the ratios and avoid to quote their uncertainties.

### 3.5 Perspectives for other methods and improvements

Further progress in the b-tagging performances are possible either by improving the described methods or by combination with others:

- Charge will be available in the Pixel system in form of a Time Over Threshold measurement (TOT). This will improve the spatial resolution up to  $7 \mu\text{m}$ . This analog information will be introduced in the next generation of clustering. Nevertheless the expected improvement will be probably counterbalanced by the introduction of alignment errors.
- The impact parameters of  $V^0$  vertices ( $K^0$ ,  $\Lambda$ , etc) could bring additional discrimination.
- Many options in the SV method were not yet explored. In particular, one could try to use directly multiple secondary vertices in the likelihood functions instead of the “average” secondary vertex.
- All current methods rely completely on the Monte Carlo simulations for the resolution functions, efficiencies, fake track fraction, track multiplicities, fragmentation functions etc. In order to reduce this MC dependence, an experimental calibration of b-tagging will be needed. The first idea is to use  $t\bar{t}$  events in which one b-jet is found by tagging and the other through kinematical constraints. This second b-jet will be used to measure the b-tagging efficiencies or to correct the calibration functions.
- It is possible to tag b-jets with soft leptons from semileptonic B/D hadrons decays  $b \rightarrow l$ ,  $b \rightarrow c \rightarrow l$ . One of the difficulties in this method is its strong correlation with space tagging. In previous studies the simple combination of the two algorithms improved only marginally the final performance [1]. However it was shown that a more elaborated combination with neural networks of the space b-tagging with low- $p_T$  electron can improve by 16% the u-jet rejection at 50% b-jet efficiency (see Ref.[11]) for the simple 2D method. Soft muons can also contribute to b-tagging in the same way. More studies are needed to evaluate quantitatively the improvements given by the combination with the latest space methods.



## 4 b-tagging with WH events

### 4.1 DC1 versus Physics TDR layouts

There are two main sources of differences between Physics TDR and DC1 b-tagging results. The first one is the evolution of the reconstruction software and b-tagging analysis. The second one is the the evolution of the detector description. Since in the Physics TDR only 2D b-tagging was used we can trace the full evolution only with the 2D method. Table 2 presents both the software and hardware driven changes.

2D b-tagging	Phys TDR xKalman	Phys TDR with DC1 soft.	$R_{new/old}$	DC1	$R_{DC1/TDR}$
$m_H=100, 120 \text{ GeV}/c^2$					
$\epsilon_b = 50 \%$	$326 \pm 37$	$566 \pm 66$	1.74	$187 \pm 8$	0.33
$\epsilon_b = 60 \%$	$124 \pm 9$	$203 \pm 14$	1.64	$59 \pm 1$	0.29
$m_H=400 \text{ GeV}/c^2$					
$\epsilon_b = 50 \%$	$126 \pm 9$	$280 \pm 21$	2.22	$163 \pm 5$	0.58
$\epsilon_b = 60 \%$	$65 \pm 3$	$91 \pm 4$	1.40	$59 \pm 1$	0.65

Table 2: Light jet rejections in DC1 WH events for 2D b-tagging, 300  $\mu\text{m}$  pitch in the b-layer, no pile-up, no inefficiency, z-coordinate of the primary vertex from MC.

The new software improves b-tagging by a factor  $\sim 1.4 - 2.2$ . In fact the new version of the xKalmanpp pattern recognition program uses much more elaborated track finding methods. In particular the track finding starts from the highly granular pixel detector, while the Physics TDR version of xKalman started from the TRT hits. The change in the detector layout decreases the performance by a factor  $\sim 0.3$  for light Higgs boson events ( $m_H = 100/120 \text{ GeV}/c^2$ ). The main reasons are the increase of multiple scattering and secondary interactions due to the increase of the material budget and b-layer radius. The loss of performance is smaller (factor  $\sim 0.6$ ) for heavy Higgs boson events ( $m_H = 400 \text{ GeV}/c^2$ ), where multiple scattering effects are less important due to the high momentum of tracks in these events; the bigger radius of the b-layer is also favorable for pattern recognition inside dense jets.

In the case of the 3D algorithm only the hardware evolution can be determined (Table 3). The decrease factor of the performance is  $\sim 0.30$  for light Higgs boson events ( $m_H = 100/120 \text{ GeV}/c^2$ ) and  $\sim 0.58$  for heavy Higgs boson ( $m_H = 400 \text{ GeV}/c^2$ ) events.

3D b-tagging	Phys TDR with DC1 soft.	DC1	$R_{DC1/TDR}$
$m_H=100,120 \text{ GeV}/c^2$			
$\epsilon_b = 50 \%$	$1406 \pm 259$	$421 \pm 26$	0.30
$\epsilon_b = 60 \%$	$399 \pm 39$	$115 \pm 4$	0.29
$m_H=400 \text{ GeV}/c^2$			
$\epsilon_b = 50 \%$	$460 \pm 44$	$260 \pm 10$	0.57
$\epsilon_b = 60 \%$	$144 \pm 8$	$86 \pm 2$	0.60

Table 3: Light jet rejections in DC1 WH events for 3D b-tagging, 300  $\mu\text{m}$  pitch in the b-layer, no pile-up, no inefficiency, z-coordinate of the primary vertex from MC.

## 4.2 Staging of the intermediate pixel layer

The staging of the intermediate pixel layer will affect the quality of the track reconstruction and b-tagging. The negative effects of the absence of the intermediate pixel layer come from more difficult pattern recognition, associations of wrong hits in the b-layer and strong dependencies on the inefficiencies. However the smaller amount of material due to the absence of this layer has also positive effects in reducing the multiple scattering for low energy tracks. The net result is therefore not trivial and depends on the physical process. In order to get the best possible results in case of two pixel layers the high selectivity option (DCKILL=2) in the xKalmanpp program is needed, which corresponds to considering a high number of combinations of the space points for the track candidates. The drawback is the significant increase of the CPU time consumed by the pattern recognition program. That increase can be avoided with the three layer system.

Even for the ideal situation with 300  $\mu\text{m}$  b-layer pitch, absence of pile-up and z-vertex from MC truth, a significant degradation occurs as illustrated in Table 4.

	2 layers	3 layers	$R_{2/3}$
$m_H=120 \text{ GeV}/c^2$			
2D $\epsilon_b = 50 \%$	$149 \pm 4$	$187 \pm 8$	0.80
$\epsilon_b = 60 \%$	$50 \pm 1$	$59 \pm 1$	0.85
3D $\epsilon_b = 50 \%$	$336 \pm 14$	$421 \pm 26$	0.80
$\epsilon_b = 60 \%$	$94 \pm 2$	$115 \pm 4$	0.82
$m_H=400 \text{ GeV}/c^2$			
2D $\epsilon_b = 50 \%$	$129 \pm 4$	$163 \pm 5$	0.79
$\epsilon_b = 60 \%$	$47 \pm 1$	$59 \pm 1$	0.80
3D $\epsilon_b = 50 \%$	$200 \pm 7$	$260 \pm 10$	0.77
$\epsilon_b = 60 \%$	$66 \pm 1$	$86 \pm 2$	0.77

Table 4: Light jet rejections in DC1 WH events for 2D/3D b-tagging, no pile-up, 300  $\mu\text{m}$  pitch in the b-layer, no inefficiency, z-coordinate of the primary vertex from MC.

## 4.3 Impact of inefficiencies

Inefficiencies in the pixel layers reduce the number of good tracks used in b-tagging and degrade the impact parameter resolutions. Most critical are the inefficiencies in the b-layer, which is crucial for the selection of good tracks. On top of the default inefficiency of 3% assumed for individual pixels, SCT strips and TRT straws, the effect of the dead pixel chips and modules has been estimated. As can be seen in Table 5 the performances are reduced by  $\sim 18\%$  by introducing inefficiencies of 1% in modules and 2% in chips. In the more pessimistic case where the inefficiencies are 2% and 4% for modules and chips respectively the performances are reduced by  $\sim 30\%$ .

In the next generation of studies special algorithms for the treatment of zones with dead active elements could be used to reduce the sensitivity to inefficiencies.

## 4.4 Impact of 300/400 $\mu\text{m}$ pixel pitch in the b-layer

The default DC0/DC1 layout of the ATLAS pixel detector has a pixel size of 50  $\mu\text{m}$  x 400  $\mu\text{m}$  with 250  $\mu\text{m}$  sensor thickness in the external barrel layers (so called layer-1 and layer-2)

Module ineff. chip ineff.	no no	1.0% 2.0%	$R_{inef/noinef}$	2.0% 4.0%	$R_{inef/noinef}$
$m_H=120 \text{ GeV}/c^2$					
2D $\epsilon_b = 50 \%$	$149 \pm 4$	$120 \pm 3$	0.81	$98 \pm 2$	0.66
$\epsilon_b = 60 \%$	$50 \pm 1$	$43 \pm 1$	0.86	$37 \pm 1$	0.74
3D $\epsilon_b = 50 \%$	$336 \pm 14$	$254 \pm 9$	0.76	$210 \pm 7$	0.63
$\epsilon_b = 60 \%$	$94 \pm 2$	$76 \pm 1$	0.81	$64 \pm 1$	0.68
$m_H=400 \text{ GeV}/c^2$					
2D $\epsilon_b = 50 \%$	$129 \pm 4$	$107 \pm 3$	0.83	$88 \pm 2$	0.68
$\epsilon_b = 60 \%$	$47 \pm 1$	$40 \pm 1$	0.85	$34 \pm 1$	0.72
3D $\epsilon_b = 50 \%$	$200 \pm 7$	$159 \pm 5$	0.80	$130 \pm 4$	0.65
$\epsilon_b = 60 \%$	$66 \pm 1$	$53 \pm 1$	0.80	$46 \pm 1$	0.70

Table 5: Light jet rejections in DC1 WH events for 2D/3D b-tagging, no pile-up, 300  $\mu\text{m}$  pitch in the b-layer, 2 pixel layers, z-coordinate of the primary vertex from MC, as a function of the inefficiencies (x in modules, 2 x in chips, where x = 0, 1 or 2 %).

and end-cap disks. In the b-layer the pixel size is assumed to be 50  $\mu\text{m}$  x 300  $\mu\text{m}$  with 200  $\mu\text{m}$  sensor thickness. In this case b-layer modules are different from other pixel modules. The design of the electronics with 300  $\mu\text{m}$  pitch would require significant changes or conversion to a new rad-hard technology (e.g 0.13  $\mu\text{m}$  DSM). The unification of the pixel modules to one type of 400  $\mu\text{m}$  can simplify the production and accelerate the installation schedule for the Initial ATLAS Detector. In this section we show the influence of the increase of the pixel longitudinal pitch from 300  $\mu\text{m}$  to 400  $\mu\text{m}$  in the b-layer. The corresponding Engineering Change Request [12] was submitted for the ATLAS approval. The small effect of the increase of the thickness from 200  $\mu\text{m}$  to 250  $\mu\text{m}$  was neglected, since it is believed to be very small and would require new G3 simulation of the data samples.

The effects of the pixel longitudinal pitch increase were reported in details in the ECR backup document [13], only the summary and conclusions are given here.

- 2D b-tagging is almost insensitive to the increased pitch with 4% average worsening.
- 3D b-tagging degrades on average by 10 %.
- 3D + SV b-tagging is also sensitive with  $\sim 8\%$  worsening.
- The effect of the increased pixel pitch can be almost completely compensated by choosing the best modules in the b-layer. In this case the average loss factor is  $R \sim 0.97$  (Table 6) instead of  $R \sim 0.90$ .

The next generation of studies should take into account the remaining small effects: 250  $\mu\text{m}$  b-layer thickness, latest test beam results on the pixel performance with new FE-I electronics, detailed description of the chip and MCC buffers, better estimations of the contributions of noisy pixels and early events to the buffer occupancy.

#### 4.5 Impact of pile-up at low luminosity

The presence of pile-up events makes the pattern recognition more difficult and time consuming. It also increases the probability of selecting fake tracks and the fraction of tracks

	300 $\mu\text{m}$ b-layer	400 $\mu\text{m}$ b-layer	$R_{400/300}$
b-layer inefficiency	1% and 2%	0.5% and 1%	
general inefficiency	1% and 2%	1% and 2%	
$m_{\text{H}}=120 \text{ GeV}/c^2$			
$\epsilon_b = 50 \%$	$166 \pm 5$	$172 \pm 5$	1.04
$\epsilon_b = 60 \%$	$58 \pm 1$	$57 \pm 1$	0.98
$m_{\text{H}}=400 \text{ GeV}/c^2$			
$\epsilon_b = 50 \%$	$139 \pm 4$	$133 \pm 4$	0.96
$\epsilon_b = 60 \%$	$50 \pm 1$	$48 \pm 1$	0.96

Table 6: Light jet rejections in DC1 WH events for 3D b-tagging, inefficiencies and selection of the best modules in the b-layer. Results are given for two pixel layers at low luminosity.

with shared and ambiguous hits. Pile-up events also have wide  $z$ -vertex distribution and may overlap with the main event giving sometimes tracks with artificial longitudinal impact parameter or degrading the primary vertex resolution. Another negative effect is that one of the pile-up vertices can be taken as the primary vertex of the main event leading also to artificial impact parameters and vertices.

This effect results in a loss of performance by 1-2% and 2-5% for the 2D and 3D algorithms respectively, as shown in Table 7. This result confirms that the quality of the present version of the pattern recognition is good enough to deal with low luminosity pile-up.

	zero lumi.	low lumi.	$R_{zero/low}$
$m_{\text{H}}=120 \text{ GeV}/c^2$			
2D $\epsilon_b = 50 \%$	$149 \pm 4$	$147 \pm 4$	0.99
$\epsilon_b = 60 \%$	$50 \pm 1$	$50 \pm 1$	1.00
3D $\epsilon_b = 50 \%$	$336 \pm 14$	$320 \pm 13$	0.95
$\epsilon_b = 60 \%$	$94 \pm 2$	$90 \pm 2$	0.96
$m_{\text{H}}=400 \text{ GeV}/c^2$			
2D $\epsilon_b = 50 \%$	$129 \pm 4$	$127 \pm 4$	0.98
$\epsilon_b = 60 \%$	$47 \pm 1$	$46 \pm 1$	0.98
3D $\epsilon_b = 50 \%$	$200 \pm 7$	$189 \pm 7$	0.95
$\epsilon_b = 60 \%$	$66 \pm 1$	$65 \pm 1$	0.98

Table 7: Light jet rejections in DC1 WH events for 2D/3D b-tagging, 300  $\mu\text{m}$  pitch in the b-layer, for 2 pixel layers, no inefficiencies,  $z$ -coordinate of the primary vertex from MC.

#### 4.6 Secondary vertex b-tagging and rejection of interacting tracks

Since the Physics TDR publication the b-tagging algorithm evolved from the simple 2D method, using only the transverse impact parameter and track quality cuts toward the 3D method using the two-variable distribution of transverse and longitudinal impact parameters. The performance with the simple 2D method is shown in the second column of the Table 8. The beam spot position (0,0) was used in this table as the transverse primary vertex and

	2D	3D	3D corr	3D +track. rej.	3D + track. rej. +sec. vertex
$m_H=120 \text{ GeV}/c^2$					
$\epsilon_b = 50 \%$	$115 \pm 3$	$181 \pm 5$	$179 \pm 5$	$204 \pm 6$	$428 \pm 19$
$\epsilon_b = 60 \%$	$42 \pm 1$	$60 \pm 1$	$60 \pm 1$	$64 \pm 1$	$130 \pm 3$
$m_H=400 \text{ GeV}/c^2$					
$\epsilon_b = 50 \%$	$96 \pm 2$	$138 \pm 4$	$138 \pm 4$	$155 \pm 5$	$461 \pm 25$
$\epsilon_b = 60 \%$	$36 \pm 1$	$49 \pm 1$	$49 \pm 2$	$52 \pm 1$	$167 \pm 5$

Table 8: Light jet rejections in DC1 WH events for different b-tagging methods at low luminosity with 2 pixel layers, 1%(2%) of module(chip) inefficiencies, (0.5%(1%) in the b-layer) and 400  $\mu\text{m}$  pitch in the b-layer.

the reconstructed z-coordinate from VKalVrt package <sup>1</sup>. The 3D method results in the third column are significantly better. The estimation of the longitudinal impact parameter significances can be slightly improved by taking into account all vertex covariance matrix terms in the point of closest approach of the track helix to the vertex. The results after these corrections with the 3D method are shown in the fourth column and they do not exhibit a significant difference. The next improvement consists in rejecting *bad* tracks from the jet weight computation. A track is identified as *bad* during vertex reconstruction if its vertex is located in the region of dense material concentration (beam-pipe, b-layer etc) or if it is part of a good  $K^0$  or  $\Lambda$  candidate. The results in the fifth column show the expected improvement due to this selection. The last method combines the 3D algorithm together with the *bad* track rejection and the secondary vertex based discriminating variables. That method gives significantly better result, as shown in the last column.

#### 4.7 Impact of primary vertex reconstruction

The quality of the primary vertex reconstruction enters into the resolution on the impact parameters. Due to the small transversal beam size at LHC ( $\sigma_{beam} = 15 \mu\text{m}$ ), the use of the transverse beam crossing position is already a very good approximation for the primary transverse vertex. However the longitudinal primary vertex should be reconstructed on an event by event basis. The b-tagging results in Table 9 show that if the z-coordinate is reconstructed by VKalVrt the loss of performance is small for the 2D case ( $\sim 4\%$ ), but important for the 3D algorithm especially for low  $p_T$  jets coming from a light Higgs boson.

In the case of more sophisticated secondary vertex b-tagging (see Table 10) the performance is more sensitive to the quality of the primary vertex reconstruction. For low  $p_T$  jets an important degradation by 35-42 % is observed. This is partly due to the primary vertex resolution, but mostly due to rare cases where a vertex from pile-up was attributed to the primary vertex, instead of the hard-event vertex. Such situations are equivalent to non-gaussian tails in the vertex resolution. It should be possible to improve this in the future by a better tuning of the criteria for choosing the primary vertex and by the use of additional information *e.g.* the pointing of a lepton to the primary vertex. Other criteria based on particular physics process could be used.

<sup>1</sup>The improved version of VKalVrt used in this table and everywhere for the SV algorithm results in a slight 1-5% increase of performance.

	MC vertex (0,0, $z_{mc}$ )	VKalVrt vertex (0,0, $z_{vkal}$ )	$R_{vkal/MC}$
$m_H=120 \text{ GeV}/c^2$			
2D $\epsilon_b = 50 \%$	$121 \pm 3$	$114 \pm 3$	0.94
$\epsilon_b = 60 \%$	$43 \pm 1$	$41 \pm 1$	0.95
3D $\epsilon_b = 50 \%$	$231 \pm 8$	$172 \pm 5$	0.74
$\epsilon_b = 60 \%$	$71 \pm 1$	$57 \pm 1$	0.80
$m_H=400 \text{ GeV}/c^2$			
2D $\epsilon_b = 50 \%$	$105 \pm 3$	$102 \pm 3$	0.97
$\epsilon_b = 60 \%$	$38 \pm 1$	$37 \pm 1$	0.97
3D $\epsilon_b = 50 \%$	$149 \pm 5$	$133 \pm 4$	0.89
$\epsilon_b = 60 \%$	$52 \pm 1$	$48 \pm 1$	0.92

Table 9: Light jet rejections in DC1 WH events for 2D/3D b-tagging, low luminosity,  $400 \mu\text{m}$  pitch in the b-layer, 1.0/2.0 % inefficiencies and selection of the best modules in the b-layer with 0.5/1.0 % inefficiencies, 2 pixel layers.

	MC vertex (0,0, $z_{mc}$ )	VKalVrt vertex ( $x_{vkal}, y_{vkal}, z_{vkal}$ )	$R_{vkal/MC}$
$m_H=120 \text{ GeV}/c^2$			
SV $\epsilon_b = 50 \%$	$900 \pm 82$	$519 \pm 36$	0.58
$\epsilon_b = 60 \%$	$220 \pm 10$	$164 \pm 6$	0.75
$m_H=400 \text{ GeV}/c^2$			
SV $\epsilon_b = 50 \%$	$759 \pm 52$	$662 \pm 42$	0.87
$\epsilon_b = 60 \%$	$267 \pm 11$	$257 \pm 10$	0.96

Table 10: Light jet rejections in DC1 WH events for SV b-tagging, low luminosity,  $400 \mu\text{m}$  pitch in the b-layer, 1.0/2.0 % inefficiencies and selection of the best modules in the b-layer with 0.5/1.0 % inefficiencies, 3 pixel layers.

#### 4.8 Comparison between ATLFAST and Full Simulation jets

Up to now, only ATLFAST jets have been used to estimate the performances. They are very practical since they only require the simulation of the inner detector, which is much faster than the calorimeter simulation and therefore allow to use large data samples. Nonetheless, it is important to check whether the detector effects (beyond those parametrized in ATLFAST) in reconstructing jet can lead to significant changes in the results.

The crude cone algorithm has been used to reconstruct jets from fully simulated events. The jets have been re-calibrated with the *cal $\theta$ vjet* routine [15] to correct for dead material, non-compensation and leakage. This is needed to estimate the light jet rejection as a function of the  $p_T$  of the jet. At last, the jet pseudo-rapidity has not been corrected for the longitudinal size of the LHC beams. The staged initial detector has been considered for this comparison. The global light jet rejections are presented in Table 11 (column Fast and Full).

The results are in a remarkable agreement. However, the fraction of jets with no good track is much higher in the light fully simulated jets. This is mainly due to pile-up jets at low  $p_T$ , not present in the ATLFAST sample. Although these jets cannot be tagged as b-jets, they are used in the normalisation to compute the efficiencies. This could bias the comparison,

	WH (120)			WH (400)		
	Fast	Full	Full + match	Fast	Full	Full + match
2D $\epsilon_b = 50\%$	114±3	121±3	116±3	102±3	123±3	109±3
$\epsilon_b = 60\%$	41±1	45±1	42±1	37±1	44±1	39±1
3D $\epsilon_b = 50\%$	172±5	169±4	164±4	133±4	140±4	130±4
$\epsilon_b = 60\%$	57±1	59±1	56±1	48±1	52±1	48±1

Table 11: Comparison of light jet rejections for “Fast” and “Full” simulation jets. The column “Full + match” gives the results for full simulation jets that lie at a distance  $\Delta R$  less than 0.4 from an ATLFAST jet.

leading *by chance* to comparable rejections in both samples. Since these 0-track jets carry information on the track reconstruction efficiency, they cannot be simply discarded. A way to get rid of most of these pile-up jets in the full simulation is to use only jets which match an ATLFAST jet (*i.e.* an ATLFAST jet is found at a distance  $\Delta R$  less than 0.4 around the full simulation jet direction). The results with this matching are shown in Table 11 (column Full + match). The evolution of the rejection at  $\epsilon_b = 60\%$  as a function of  $|\eta|$  and  $p_T$  for the simple 2D algorithm are presented in Fig. 13. The upper plots show the ATLFAST results while the lower ones show the ratio of rejection between Full+match and Fast jets. Given

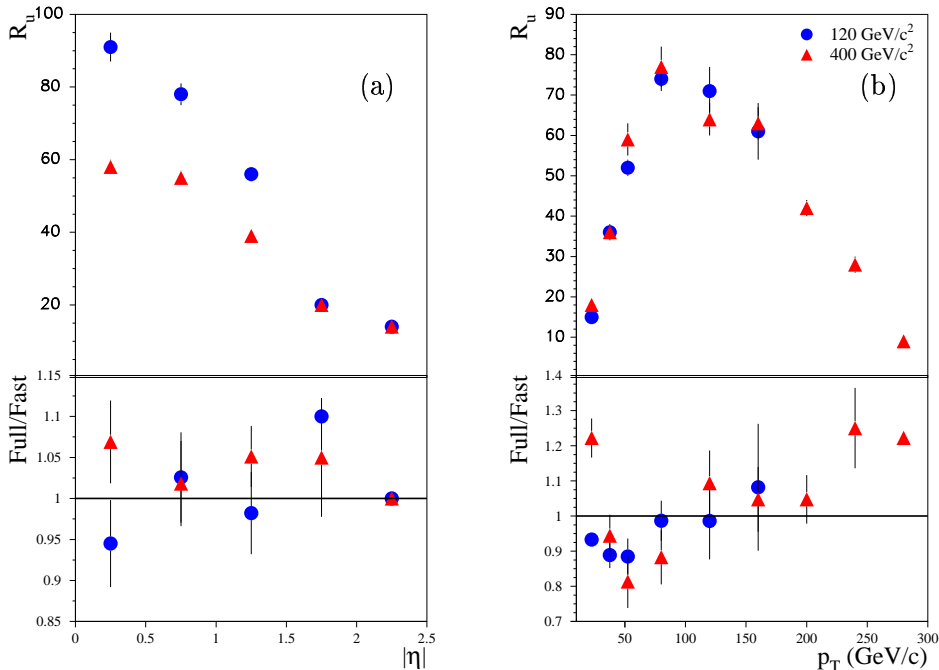


Figure 13: As a function of  $|\eta|$  (a) and  $p_T$  (b), light jet rejection for ATLFAST jets (upper plots) at 60% b-tagging efficiency and for the simple 2D algorithm. The lower plots show the ratio of rejection for Full + match (see text) and ATLFAST jets.

that the calibration constants (in *cal6vjet*) have been determined with old simulations, some differences between the two jet samples are expected. The agreement visible in Fig. 13 is however satisfactory over all the  $|\eta|$  and  $p_T$  ranges. ATLFAST jets can therefore be safely used to estimate the performances.

## 4.9 Comparison between DC0 and DC1 data

At the early stage of these studies, DC1 data were not available. Only events simulated with the DC0 layout from old GENZ tapes could be used. The DC0 and DC1 detector layouts are very similar, but the physics events do differ, since they have been generated with different PYTHIA versions. In particular, parameters in the B-hadron sector and for the underlying event are different. The comparison between DC0 and DC1 results is therefore interesting and could be considered as a first look at systematic uncertainties. For this comparison, only the perfect complete ATLAS detector has been considered. From the better treatment of long-lived particles (especially hyperons and  $K_S^0$ ) and the increase of low  $p_T$  track and jet multiplicities from the underlying event in DC1 data, a slight degradation of the performances can be expected. The global light jet rejections for DC0 data are shown in Table 12. These numbers should be compared with those in Tables 2, 3. The loss of performance between DC0 and DC1 is around 16% (22%) for a light (heavy) Higgs boson. Although the light Higgs boson samples correspond to two different masses (100  $\text{GeV}/c^2$  in DC0 and 120  $\text{GeV}/c^2$  in DC1 events), the event kinematics are not expected to change a lot.

	WH (100)	WH (400)
2D $\epsilon_b = 50\%$	$218 \pm 16$	$212 \pm 12$
$\epsilon_b = 60\%$	$71 \pm 3$	$78 \pm 3$
3D $\epsilon_b = 50\%$	$497 \pm 54$	$316 \pm 22$
$\epsilon_b = 60\%$	$143 \pm 8$	$110 \pm 5$

Table 12: Light jet rejections for DC0 data, 2D/3D b-tagging, complete ATLAS ID, no pile-up, no module/chip inefficiency and z-coordinate of the primary vertex from MC.

The evolutions of the rejection as a function of  $|\eta|$  and  $p_T$ , at 60% b-tagging efficiency and for the simple 2D algorithm, are shown in Fig. 14 for WH events and a heavy Higgs boson. Although some differences between both datasets have been identified, no quantification of

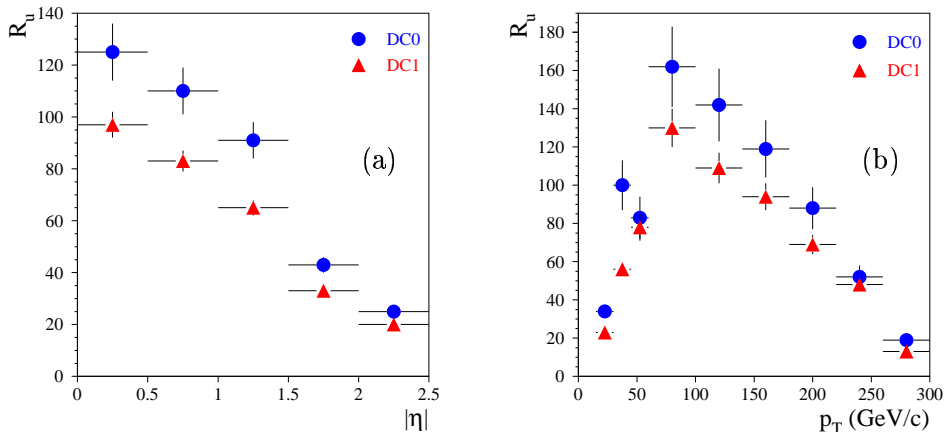


Figure 14: As a function of  $|\eta|$  (a) and  $p_T$  (b), light jet rejection in DC0 (dots) and DC1 (triangles) WH(400) events, at 60% b-tagging efficiency and for the simple 2D algorithm.

their impacts on the results has been performed for the time being. This is an important issue and will require further studies. It should be also stressed that much higher discrepancies have been observed with HERWIG data (although in very different events with a higher track activity). Further investigations are needed to understand these behaviours.



## 4.10 Realistic b-tagging and staging

The impact of the the staging of the intermediate pixel layer should be evaluated taking into account all realistic effects which can influence the final performances.

As shown in Section 4.2 the performances are degraded after staging even in case of a perfect detector without pile-up. This degradation becomes stronger after including the realistic effects of 400  $\mu\text{m}$  pitch in the b-layer, low luminosity pile-up and z-coordinate of the primary vertex reconstructed by VKalVrt as presented in Table 13. This worsening is

	2 layers	3 layers	$R_{2/3}$
$m_{\text{H}}=120 \text{ GeV}/c^2$			
2D $\epsilon_b = 50 \%$	$114 \pm 3$	$141 \pm 5$	0.81
$\epsilon_b = 60 \%$	$41 \pm 1$	$50 \pm 1$	0.82
3D $\epsilon_b = 50 \%$	$172 \pm 5$	$229 \pm 11$	0.75
$\epsilon_b = 60 \%$	$57 \pm 1$	$76 \pm 2$	0.75
$m_{\text{H}}=400 \text{ GeV}/c^2$			
2D $\epsilon_b = 50 \%$	$102 \pm 2$	$136 \pm 4$	0.75
$\epsilon_b = 60 \%$	$37 \pm 1$	$49 \pm 1$	0.76
3D $\epsilon_b = 50 \%$	$133 \pm 4$	$188 \pm 4$	0.71
$\epsilon_b = 60 \%$	$48 \pm 1$	$66 \pm 1$	0.73

Table 13: Light jet rejections in DC1 WH events for 2D/3D b-tagging, low luminosity, 400  $\mu\text{m}$  pitch in the b-layer, 1.0/2.0 % inefficiencies and selection of best modules in b-layer with 0.5/1.0 % inefficiencies.

observed with SV b-tagging too, as can be seen in Table 14.

	2 layers	3 layers	$R_{2/3}$
$m_{\text{H}}=120 \text{ GeV}/c^2$			
SV $\epsilon_b = 50 \%$	$433 \pm 19$	$519 \pm 36$	0.84
$\epsilon_b = 60 \%$	$131 \pm 3$	$164 \pm 6$	0.80
$m_{\text{H}}=400 \text{ GeV}/c^2$			
SV $\epsilon_b = 50 \%$	$486 \pm 26$	$662 \pm 42$	0.73
$\epsilon_b = 60 \%$	$170 \pm 5$	$257 \pm 10$	0.66

Table 14: Light jet rejections in DC1 WH events for SV b-tagging, 400  $\mu\text{m}$  pitch in the b-layer, 1.0/2.0 % inefficiencies and selection of the best modules in b-layer with 0.5/1.0 % inefficiencies, x,y,z-coordinates of the primary vertex reconstructed by VKalVrt.

The degradation is larger ( $\sim 27 - 34\%$ ) for jets at high  $p_T$  originating from a heavy Higgs boson, where multiple scattering effects are less important. The absence of the intermediate layer has a negative impact on the pattern recognition, but has a positive effect too, thanks to the decrease of the amount of material near the vertex region. The net impact on b-tagging comes therefore from the balance between these negative and positive effects. This explains why the loss of performance is less visible in the light Higgs boson sample, where multiple scattering has the largest influence on performances.

## 5 b-tagging with $t\bar{t}H$ and $t\bar{t}$ events

Up to now, b-tagging studies have essentially been performed in the “clean” WH environment. These results can be considered as good benchmark performances if the assumption that b-jet identification does not depend on the event type is made. An interesting issue is the check of this hypothesis in a much more realistic channel, with a higher jet multiplicity.

The  $t\bar{t}H$  channel is one of the most promising for the discovery of a Higgs boson with mass below  $120 \text{ GeV}/c^2$ . Its relevance strongly depends on the main reducible background  $t\bar{t}jj$  rate [14], related to the performances of b-jet identification. The b-tagging with  $t\bar{t}H$  and  $t\bar{t}$  events was studied in details in [16] and we report here the main results with the latest software upgrades. These events are quite busy, since they contain at least six jets, among which four should be b-jets. The data samples used for the analysis are the following:

- 20K  $t\bar{t}H$  events, with  $t\bar{t} \rightarrow \ell\nu b jjb$ ,  $H \rightarrow b\bar{b}$
- 100K  $t\bar{t}$  events, filtered with ATLFAST at the generation level with the following criteria:
  - ≥ 1 electron (muon) with  $p_T > 15(4) \text{ GeV}/c$  and  $|\eta| < 2.8$
  - ≥ 6 jets with  $p_T > 10 \text{ GeV}/c$  and  $|\eta| < 5$
  - ≥ 4 of these jets should be in the pixel detector acceptance  $|\eta| < 2.5$

### 5.1 Calibration and overlapping jets

For the first part of the analysis, all b-jets and light jets labelled by ATLFAST in  $t\bar{t}H$  and  $t\bar{t}$  events are used to construct the likelihood functions. The complete detector version without pile-up or inefficiencies is considered.

The first estimation of the light jet rejection (including u,d,s quark and gluon jets) gives very bad results compared to the reference channel WH as shown in the second column Table 15.

	First estimation	Cleaning of udsg jet sample	Track-jet association	Reference channel WH(120)
$\epsilon_b = 50\%$	$110 \pm 2$	$225 \pm 6$	$300 \pm 10$	$187 \pm 8$
$\epsilon_b = 60\%$	$47 \pm 1$	$74 \pm 1$	$83 \pm 1$	$59 \pm 1$

Table 15: Light jet rejections in  $t\bar{t}H/t\bar{t}$  events, compared to the reference channel WH(120), with the complete detector and 2D b-tagging.

The first difficulty arises from the likelihood distribution construction. In WH events, the distributions for light jets and b-jets are obtained from independent data samples. In  $t\bar{t}H$  and  $t\bar{t}$ , the different jet flavours are extracted from the same events, according to the ATLFAST labelling. During this labelling procedure, a heavy flavour jet (b or c) is identified if an heavy quark with  $p_T > 5 \text{ GeV}/c$  is found in a cone of size 0.2 around the jet direction. The b and c jet samples obtained from this procedure are very pure, but the light jet sample is contaminated with misidentified b or c jets, leading to biased track significance and jet weight distributions.

To “clean” the light jet sample, additional criteria were applied: light jets are used only in events where all the heavy flavours have been labelled (*i.e.* the number of heavy jets found

in the event is the same as the number of heavy quarks produced). The jet sample obtained with this criterium is very pure, but the selection is too tight: only 48% of the light jets are selected. To increase the statistics, light jets with the closest heavy quark outside a cone of size 0.8 around its direction are added. The selection efficiency increase up to 87%. After this “cleaning” procedure, the rejections are significantly improved (column 3 of Table 15).

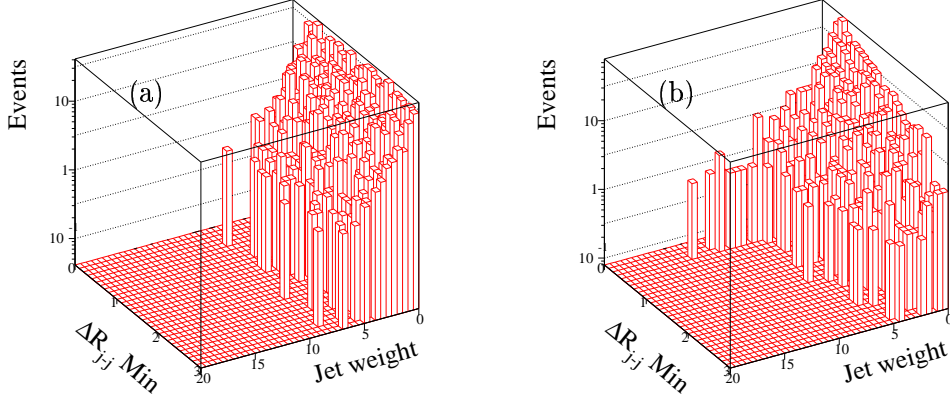


Figure 15: Light jet weight distributions in WH(120) (a) and  $t\bar{t}H/t\bar{t}$  (b) events, as a function of the distance to the closest jet.

Fig. 15 shows the light jet weight distribution for WH(120) (a) and  $t\bar{t}H/t\bar{t}$  (b) events, after the cleaning procedure. Tails at large weights remain in the  $t\bar{t}H/t\bar{t}$  sample, and are always found for jets very close to another jet. Jets within a distance of 0.8 may have common tracks. With the default b-tagging algorithm, those tracks in the cone intersection are used for both jets. To solve this problem, each track is associated to the closest jet and used only once. With this simple track-jet association, the rejections are increased by a factor of 33% at  $\epsilon_b=50\%$ .

## 5.2 Jet isolation

Table 16 compares the rejections obtained for *isolated* and *non-isolated* jets. A jet is defined as *isolated* when the distance to the closest jet is greater than 0.8. The global rejections expected for the non-isolated jets are degraded by 35% with respect to isolated jets.

	All jets	Isolated jets	Non-isolated jets
$\epsilon_b = 50\%$	$300 \pm 10$	$338 \pm 14$	$220 \pm 12$
$\epsilon_b = 60\%$	$83 \pm 1$	$90 \pm 2$	$66 \pm 2$

Table 16: Light jet rejections in  $t\bar{t}H/t\bar{t}$  events for all, isolated and non-isolated jets (2D b-tagging).

Fig. 16a shows the rejection dependence as a function of the closest jet distance. The difference between isolated and non-isolated jets is expected to come from the  $p_T$  spectrum of the jets. The non-isolated jets are often gluon jets emitted by the main quarks of the event, and are less energetic than isolated jets (Fig. 16b). Fig. 17a compares the rejections as a function of the jet transverse momentum for both types of jets. Including the  $p_T$  dependence, the results become comparable.

The remaining difference is explained by Fig. 17b, for non-isolated jets, some of the tracks may be associated to the wrong jet, and the sign of their impact parameter is reversed. The

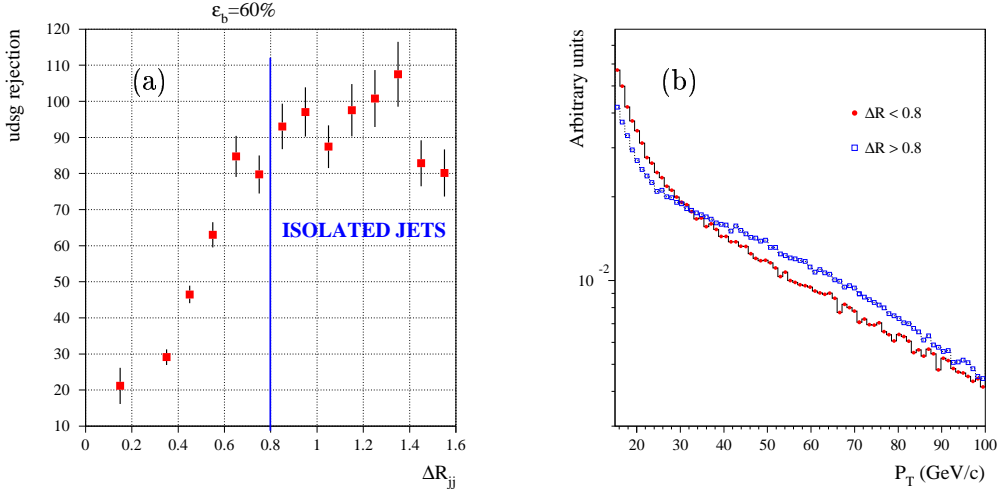


Figure 16: (a) Light jet rejection as a function of the closest jet distance for  $\epsilon_b=60\%$ . (b) Comparison of  $p_T$  distribution for isolated (empty squares) and non-isolated (dots) jets.

weight of tracks which would have had a large positive impact parameter is reduced during this wrong association, leading to a small shift both in b-jet and light jet weight distributions (Fig. 18), which tends to reduce the discrimination.

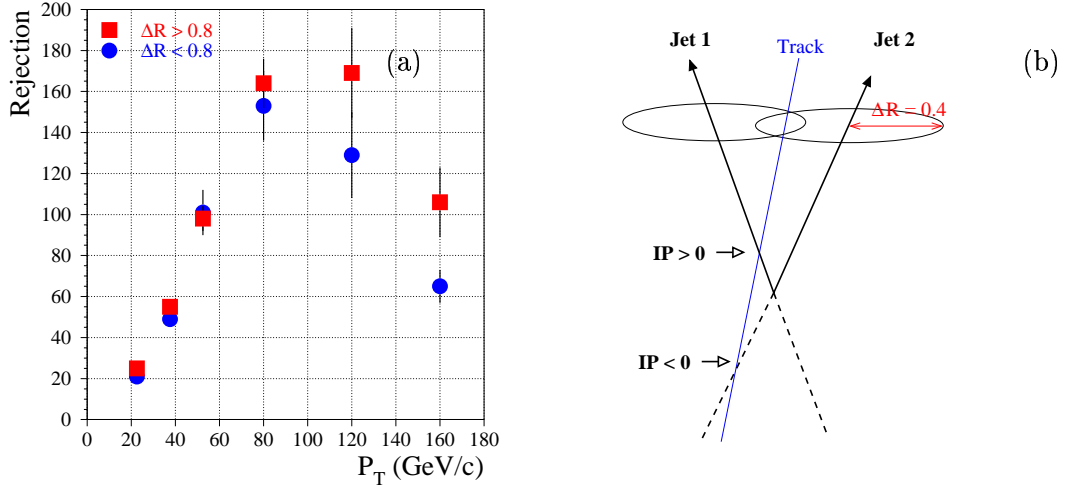


Figure 17: (a) Comparison of the rejection for isolated (squares) and non-isolated (dots) jets as a function of  $p_T$ . (b) Change of the impact parameter sign from a wrong track-jet association.

### 5.3 Performance as a function of the b-jet origin

The b-jet identification performances have always been assumed to be independent on the b production process. A check of this hypothesis is of interest since the quark colour flows are very different for a b-jet coming from the colour singlet Higgs decay and for a jet created from the top quark decay. These different processes could lead to different fragmentation and jet properties, and  $t\bar{t}H$  and  $t\bar{t}$  events allow to study this effect.

Table 17 shows the global rejections for three different b-jets samples, depending on their

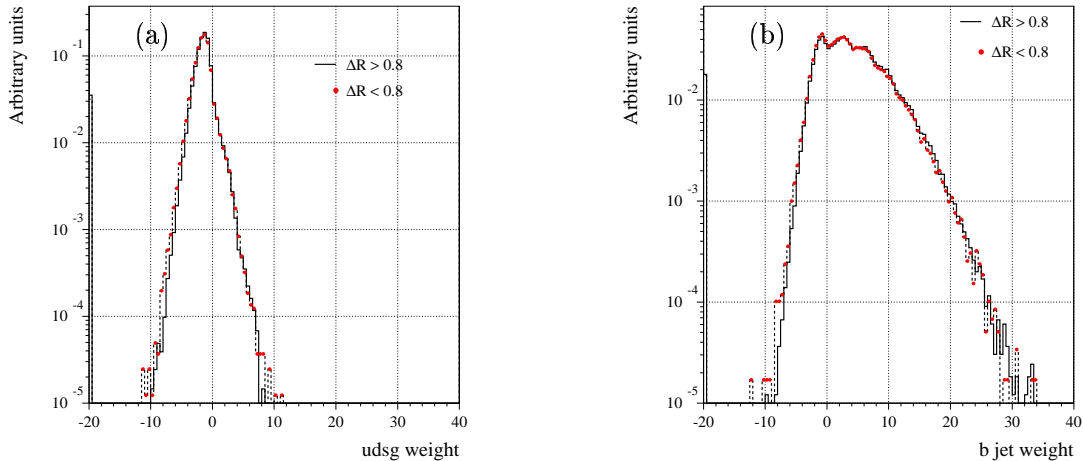


Figure 18: Distributions of light jet (a) and b-jet (b) weights for isolated (solid line) and non-isolated (dashed line) jets.

$\epsilon_b = 60\%$	$t\bar{t}H$		$t\bar{t}$
	top decay	Higgs decay	top decay
$R_{udsq}$	$86 \pm 1$	$79 \pm 1$	$84 \pm 1$
$\langle p_T \rangle_b$ (GeV/c)	<b>83.8</b>	75.3	70.6
$\langle \Delta R \rangle_{jb}$	1.1	1.0	<b>1.4</b>
$\langle N \rangle_b$	6.7	6.5	6.3

Table 17: Comparison of the light jet rejection obtained with different b-jet samples, for  $\epsilon_b=60\%$  (2D b-tagging). The jet mean  $p_T$ , distance to the closest jet and track multiplicity are also shown.

origin: b-jet from top quark decay in  $t\bar{t}H$  or  $t\bar{t}$  events, and b-jets from Higgs boson decay. The global rejections are better for the b-jet sample from top decay. In  $t\bar{t}H$  events, this result is explained by greater mean  $p_T$  value, which indicates a smaller fraction of very low  $p_T$  jets, difficult to identify (see Fig. 17a). In  $t\bar{t}$  events, where the jet multiplicity is smaller than for  $t\bar{t}H$  (although the  $t\bar{t}$  sample was biased at the generation level toward a high jet multiplicity), the better results are explained by smaller fraction of non-isolated jets.

The comparison of the rejections for the three samples, as a function of  $p_T$  and for the isolated jets only is presented on Fig. 19. The previous differences observed for global rejections are greatly reduced. With the PYTHIA generator, the b-jets coming from top quark or Higgs boson have the same properties, and their identification do not depend on the event type, if the momentum and isolation differences are taken into account.

#### 5.4 Rejection as a function of the jet flavour

The  $t\bar{t}H$  and  $t\bar{t}$  events provide the possibility to study the rejection for different light jet type. The interesting issue is the study of the differences between “quark jets” coming from W boson decay and “gluon jets” from radiation. For this study, all lights jets (u, d or s quark, gluon) and b-jets from both type of events are used. c-jet rejections are also given.

Fig. 20 compares the pseudo-rapidity, momentum and distance to the closest jet for gluon jet and u, d or s quark jet. From these distributions, we can expect a worse global rejection

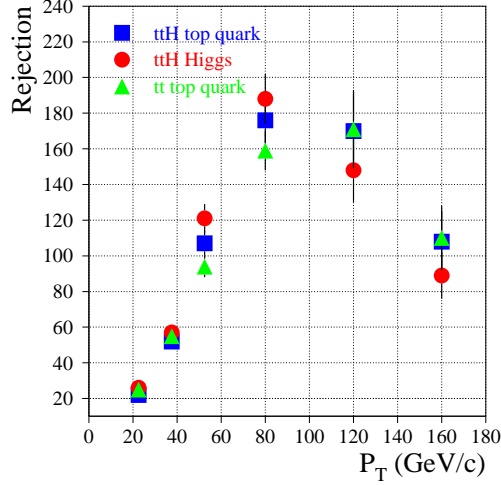


Figure 19: Comparison of the rejections as a function of the jet momentum, for the three different b-jet samples: b-jets from top quark decay in  $t\bar{t}H$  (squares) or  $t\bar{t}$  (triangles) events, and b-jets from Higgs boson decay (dots) in  $t\bar{t}H$  events. Only the isolated jets have been used in this comparison.

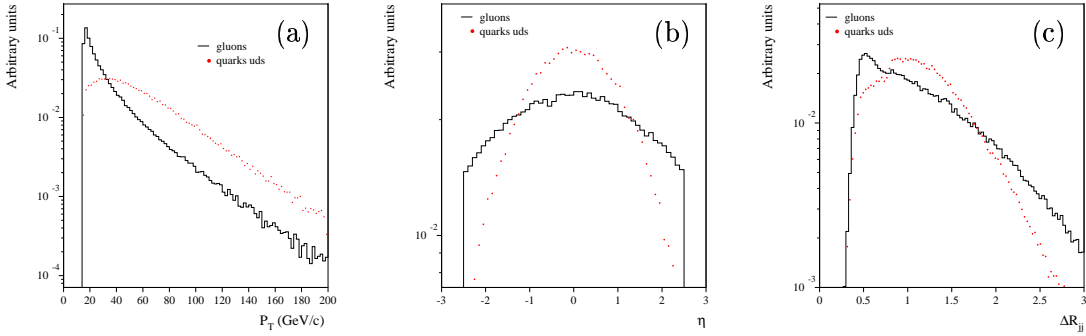


Figure 20: Comparison of the  $p_T$  (a),  $\eta$  (b) and minimal distance between jets (c) distributions, for u,d or s quark jets (solid line) and gluon jets (dots).

for gluon jets, which have a larger pseudo-rapidity, smaller  $p_T$  and are less often isolated than quark jets.

Table 18 gives the rejection for the three types of light jets (gluon, ud, s) and c-jets, with the 2D algorithm. The slight discrepancy observed at low b-tagging efficiency between gluon and ud jets is under investigation. The  $p_T$ ,  $|\eta|$  and track multiplicity spectra of these two types of jets are quite different and more statistics would be needed to really understand the correlated effects of these variables on the rejections. Moreover, due to the intricate nature of these  $t\bar{t}H/t\bar{t}$  events, the labelling procedure is not obvious and may sometimes lead to wrong labels. This discrepancy remains when only isolated jets are considered, as illustrated in Table 18. On the contrary, the 20 % degradation observed for s-jets is expected: jets produced from a s quark contain some hadrons with a non negligible life time like  $\Lambda_0$  or hyperons.

Fig. 21 shows the evolution of the rejections for each jet flavour, as a function of the b-jet selection efficiency  $\epsilon_b$ . This figure is comparable to the reference one given in the Physics TDR ([1] Fig. 10.21, p 327), obtained with WH events and an old detector version. Since the TDR publication, the detector performances have been degraded [2]. The main difference

	$t\bar{t}H/t\bar{t}$			
	gluon	u,d quark	s quark	c quark
	all jets			
$\epsilon_b = 50\%$	$329 \pm 15$	$286 \pm 17$	$237 \pm 19$	$11.9 \pm 0.2$
$\epsilon_b = 60\%$	$85 \pm 2$	$87 \pm 3$	$71 \pm 3$	$7.0 \pm 0.1$
	isolated jets			
$\epsilon_b = 50\%$	$387 \pm 23$	$319 \pm 23$	$250 \pm 23$	$12.2 \pm 0.2$
$\epsilon_b = 60\%$	$95 \pm 2$	$95 \pm 4$	$70 \pm 3$	$7.2 \pm 0.1$

Table 18: Comparison of the rejections for different jet flavour for all jets and for isolated jets (2D b-tagging).

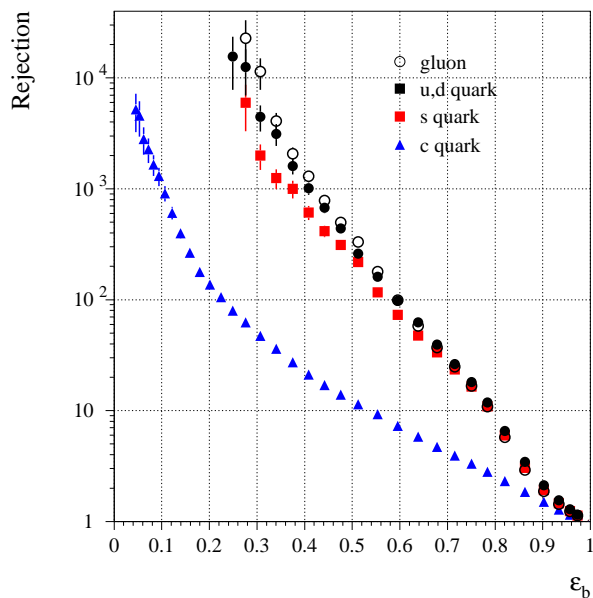


Figure 21: Comparison of the rejections for the different jet flavours, as a function of the b-jet identification efficiency  $\epsilon_b$ . Only isolated jets are considered to produce these curves.

between Fig. 21 and the TDR curve is found for the gluon jet: in the TDR, gluon conversions in heavy quark pairs ( $g \rightarrow c\bar{c}$  or  $b\bar{b}$ ) were included as gluon jets.

The rejection factor for light jets (u, d quarks or gluons) for a b-tagging efficiency of 60% is close to the canonical requirement  $R_{udg}=100$ . This factor decreases by 20% for jets coming from s quarks. The impact parameter method is limited for the c quark jets, which contain D-hadrons with lifetimes smaller than but comparable to those of B-hadrons.

## 5.5 Performances with the initial detector

At the beginning of data taking, one layer of the pixel barrel detector and two disks will be missing. One of the TRT wheels (C end cap) will also be staged. The second/third columns of Table 19 compare the rejections for the complete detector and the initial detector.

$\epsilon_b$	Complete	Initial				
		perfect	+pile-up	+400 $\mu\text{m}$	+Ineff 1/2%	+Ineff 0.5/1%
2D 50%	300 $\pm$ 10	204 $\pm$ 6	203 $\pm$ 5	200 $\pm$ 5	156 $\pm$ 4	164 $\pm$ 4
	60%	83 $\pm$ 1	62 $\pm$ 1	60 $\pm$ 1	58 $\pm$ 1	49 $\pm$ 1
3D 50%	650 $\pm$ 31	401 $\pm$ 15	387 $\pm$ 14	346 $\pm$ 12	261 $\pm$ 8	278 $\pm$ 9
	60%	151 $\pm$ 3	112 $\pm$ 2	109 $\pm$ 2	97 $\pm$ 2	79 $\pm$ 1

Table 19: Comparison of the rejections obtained with different detector layouts, for the standard 2D and improved 3D algorithms (z-coordinate of the primary vertex from MC).

The detector is considered as “perfect”, since no pile-up nor inefficiency are added. The degradation due to the detector staging is between 25% and 30% for the 2D algorithm, and between 30% and 35% for the 3D one. It is due to the lower reconstruction efficiency, from the requirement on the number of pixel hits.

The degradations are essentially visible in the regions where the track reconstruction is normally the most efficient as illustrated on Fig. 22, at small pseudo-rapidity and for  $p_T$  in

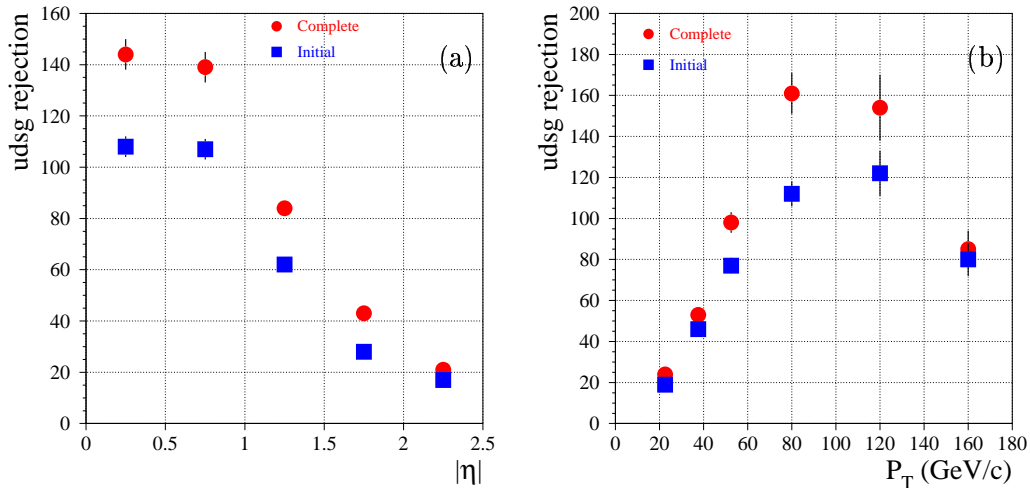


Figure 22: Comparison of the rejections for a perfect detector, in two different configurations: complete (dots) and initial (squares), as a function of jet pseudo-rapidity (a) and momentum (b).

the range [50-140] GeV/c.



The consequences of these degradations have been studied for the  $t\bar{t}H$ ,  $H \rightarrow b\bar{b}$  channel in [17], which relevance strongly depends on the b-tagging efficiency. This study was performed using a parametrisation in ATLFast as a function of a jet  $p_T$  from the full simulation, for the 2D method, no pile-up and no pixel inefficiency. The change from the full ATLAS to the initial detector increases the reducible background  $t\bar{t}jj$  by 20 %, and decreases the significance by 4% for a Higgs boson mass of 120 GeV/ $c^2$ . The effect of pile-up at low luminosity (Table 19, third column) results in an additional degradation at a level of 2-3%. More studies are needed including all realistic effect and the latest 3D and SV b-tagging results.

Table 20 shows the comparison of the rejections for all jets and for isolated jets. The

	Complete perfect			Initial perfect		
	all	$\Delta R > 0.8$	$R_{all/isol}$	all	$\Delta R > 0.8$	$R_{all/isol}$
2D $\epsilon_b = 50\%$	$300 \pm 10$	$338 \pm 14$	0.89	$204 \pm 6$	$229 \pm 8$	0.90
$\epsilon_b = 60\%$	$83 \pm 1$	$90 \pm 2$	0.92	$62 \pm 1$	$68 \pm 1$	0.96
3D $\epsilon_b = 50\%$	$650 \pm 31$	$772 \pm 48$	0.84	$401 \pm 15$	$465 \pm 22$	0.85
$\epsilon_b = 60\%$	$151 \pm 3$	$179 \pm 5$	0.84	$112 \pm 2$	$127 \pm 3$	0.89

Table 20: Comparison of the rejections obtained for jets with and without the isolation cut  $\Delta R > 0.8$ , for the standard (2D) and improved (3D) algorithms.

introduction of the isolation cut improves the rejection by 4-16 % both for 2D and 3D algorithms.

The fourth column of Table 19 gives the rejection when the  $z$ -pitch of the b-layer pixels is increased from 300 to 400  $\mu\text{m}$ . The degradation is expected to be very small for the standard 2D algorithm. The effect is more visible for the 3D algorithm, which uses the information from the longitudinal impact parameter: the rejection decreases by 11%.

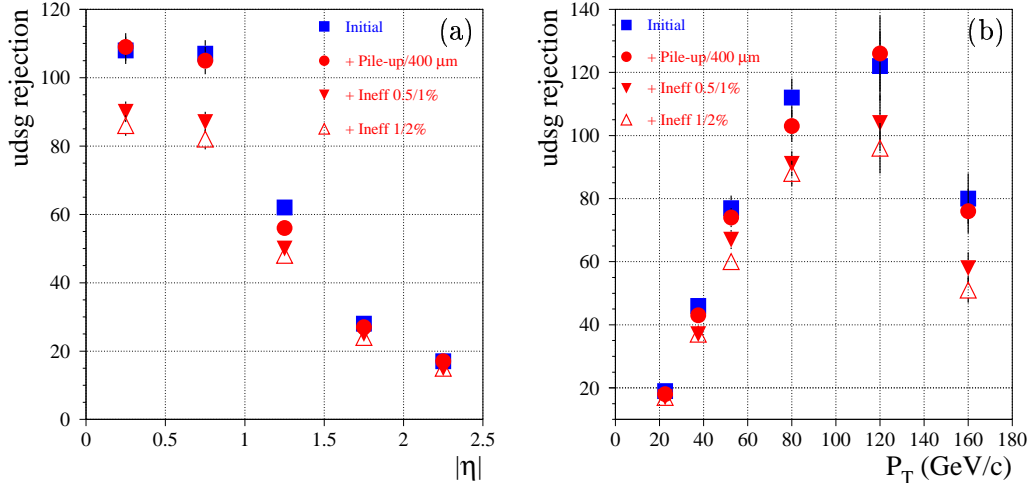


Figure 23: Light jet rejections with the “perfect” initial detector version (squares), with pile-up added and 400  $\mu\text{m}$  pitch in the b-layer (dots), with 1/2% inefficiencies in all layers (full triangles) and with 0.5/1% inefficiencies in the b-layer (empty triangles). The curves are given for the standard 2D algorithm, as a function of pseudo-rapidity (a) and  $p_T$  (b).

Finally, the detection inefficiencies due to dead channels in modules or chips have been included in the simulations. A rate of 1/2% inefficiencies for the modules/chips respectively

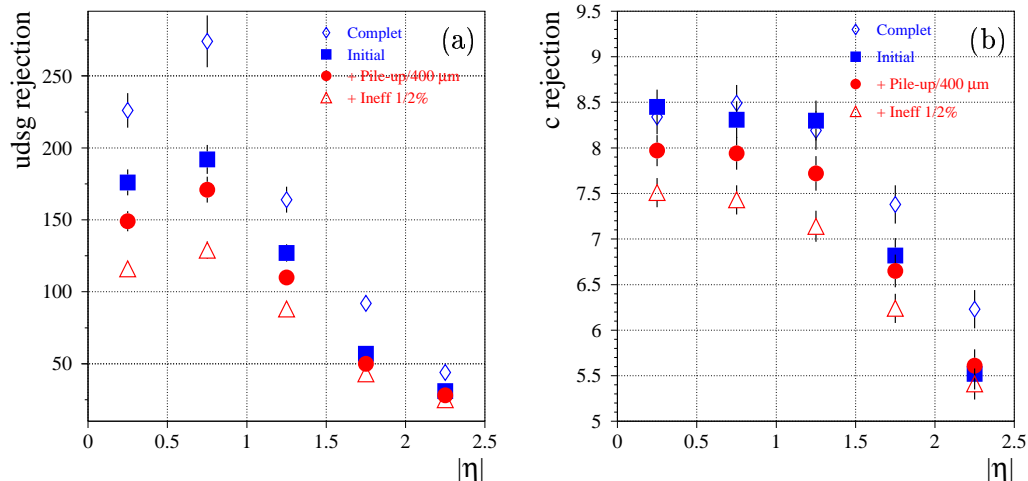


Figure 24: Light (a) and c (b) jet rejections as a function of jet pseudo-rapidity, with the complete and perfect detector (diamonds), with the initial detector (squares), with pile-up added and  $400\ \mu\text{m}$  pitch in the b-layer (dots), with 1/2% inefficiencies in all layers (triangles), for the 3D algorithm.

and for all layers has been adopted, leading to a degradation between 20 and 25% for  $\epsilon_b=50\%$ . If the pixel modules are identical in the b-layer and in other layers, the inefficiencies could be reduced by a factor 2 in the b-layer, by choosing the most efficient modules. In this case the rejections increase by 5% compared to the previous inefficiencies set.

Fig. 23 shows the rejections obtained for light jets with the different simulations, as a function of jet momentum and pseudo-rapidity. The canonical factor  $R_{\text{udsg}} = 100$  assumed in the simplified ATLFAST simulations can not be reached with the standard 2D algorithm and the most realistic initial detector version, including pile-up and inefficiencies. With the 3D algorithm (Fig. 24), the factor 100 is reached in the small pseudo-rapidity region. The rejection of 10 assumed for c-jets can not be reached, even with the improved 3D algorithm and the complete detector layout. The relative degradation of the c-jet rejection when going to more realistic detectors is less important than for the light jet rejection. This is expected since the physics (long lifetime of D-hadrons) and not the detector effects is the dominant factor in c-jet rejection.

A special study was performed to understand the importance of the intermediate barrel layer alone. In this study, the 3 pixel disks and TRT C-wheels were assumed to be present. The realistic conditions of low luminosity with  $400\ \mu\text{m}$  z-pitch in the b-layer, 1/2% general module/chip inefficiencies and 0.5/1% b-layer module/chip inefficiencies were used.

	2 layers	3 layers	$R_{2/3}$
2D $\epsilon_b = 50\%$	$158 \pm 4$	$253 \pm 8$	0.62
$\epsilon_b = 60\%$	$50 \pm 1$	$73 \pm 1$	0.68
3D $\epsilon_b = 50\%$	$254 \pm 8$	$435 \pm 17$	0.58
$\epsilon_b = 60\%$	$81 \pm 1$	$119 \pm 2$	0.68

Table 21: Light jet rejections in  $t\bar{t}H/t\bar{t}$  events for 2D/3D b-tagging, low luminosity,  $400\ \mu\text{m}$  pitch in the b-layer, 1.0/2.0 % inefficiencies and selection of the best modules in the b-layer with 0.5/1.0 % inefficiencies, 2 or 3 pixel layers (with 3 disks and TRT C-wheels).

The comparison of the layouts with 2 or 3 pixel barrel layers in Table 21 shows a degradation of the performances by 32-42 %.

## 6 Comparison between the $t\bar{t}H/t\bar{t}$ and WH channels

### 6.1 Global performances

Table 22 compares the global rejections obtained in WH events with two different Higgs boson masses, and in  $t\bar{t}H/t\bar{t}$  events. Although the multiplicities of tracks and jets are higher in  $t\bar{t}H$  and  $t\bar{t}$  events, the b-tagging performances are better than those observed in the WH channels. This result can be understood from the pseudo-rapidity and momentum distributions of the jets in both types of events and will be discussed in the next section. The rejections of c-jets, which originates from  $H \rightarrow c\bar{c}$  or  $W \rightarrow cs$  decays, are essentially the same in these three processes.

	$t\bar{t}H/t\bar{t}$		WH (120)		WH (400)	
	udsg jet	c jet	udsg jet	c jet	udsg jet	c jet
udsg $\epsilon_b = 50\%$	$300 \pm 10$	$11.9 \pm 0.2$	$187 \pm 8$	$9.8 \pm 0.2$	$163 \pm 5$	$10.8 \pm 0.2$
$\epsilon_b = 60\%$	$83 \pm 1$	$7.0 \pm 0.1$	$59 \pm 1$	$5.9 \pm 0.1$	$59 \pm 1$	$6.3 \pm 0.1$

Table 22: Light and c-jet rejections in  $t\bar{t}H/t\bar{t}$  events, compared to the reference channels WH(120) and WH(400), with the complete detector and the standard 2D algorithm.

### 6.2 Differential performances with $p_T$ and $\eta$

The  $p_T$  and  $\eta$  of the jets, the track multiplicity and the  $p_T$  of the tracks are shown on Fig. 25 and Fig. 26, for the b-jets and light jets respectively and for each type of event. The specific property of WH(400) events is the very high  $p_T$  of the jets (Fig. 25a and Fig. 26a). The tagging of very high  $p_T$  jets is not very efficient, since they contain a lot of tracks (Fig. 25c and Fig. 26c) in a small opening angle, leading to a difficult pattern recognition for the track reconstruction.

In WH(120), the pseudo-rapidity distribution of the jets (Fig. 25b and Fig. 26b) is wider than for the other types of events. This results from the production mode of the WH events which is a quark-antiquark annihilation. The quark carries more energy than the antiquark in the proton, leading to a boost of the WH system in the quark direction, and a larger pseudo-rapidity for the Higgs decay products. At large pseudo-rapidity, the particles cross more material and suffer more from multiple scattering, leading to a bad tagging efficiency. This effect is not visible in WH(400) events, in which the heavy Higgs boson is produced at rest. Furthermore the mean  $p_T$  value of b-jets is lower than in  $t\bar{t}H$  and  $t\bar{t}$  events. Jets with low  $p_T$  contain low  $p_T$  tracks, more sensitive to multiple scattering.

Fig. 27b compares the rejections for the  $t\bar{t}H/t\bar{t}$ , WH(120) and WH(400) events, as a function of  $p_T$ . As a function of  $p_T$  rejections are closer for all three types of events. The rejection in WH(120) events with  $p_T$  lower than 100 GeV/c is slightly worse than in other samples. This can be explained by wider pseudo-rapidity distribution of WH(120) events. For higher  $p_T$  jets, the statistics is too small in the  $t\bar{t}H$ ,  $t\bar{t}$  and WH(120) samples to conclude on their difference. As a function of  $|\eta|$  (Fig. 27a), the difference between WH(120) and  $t\bar{t}H/t\bar{t}$  events is also reduced. Rejections are lower in WH(400) events, where the jets have a

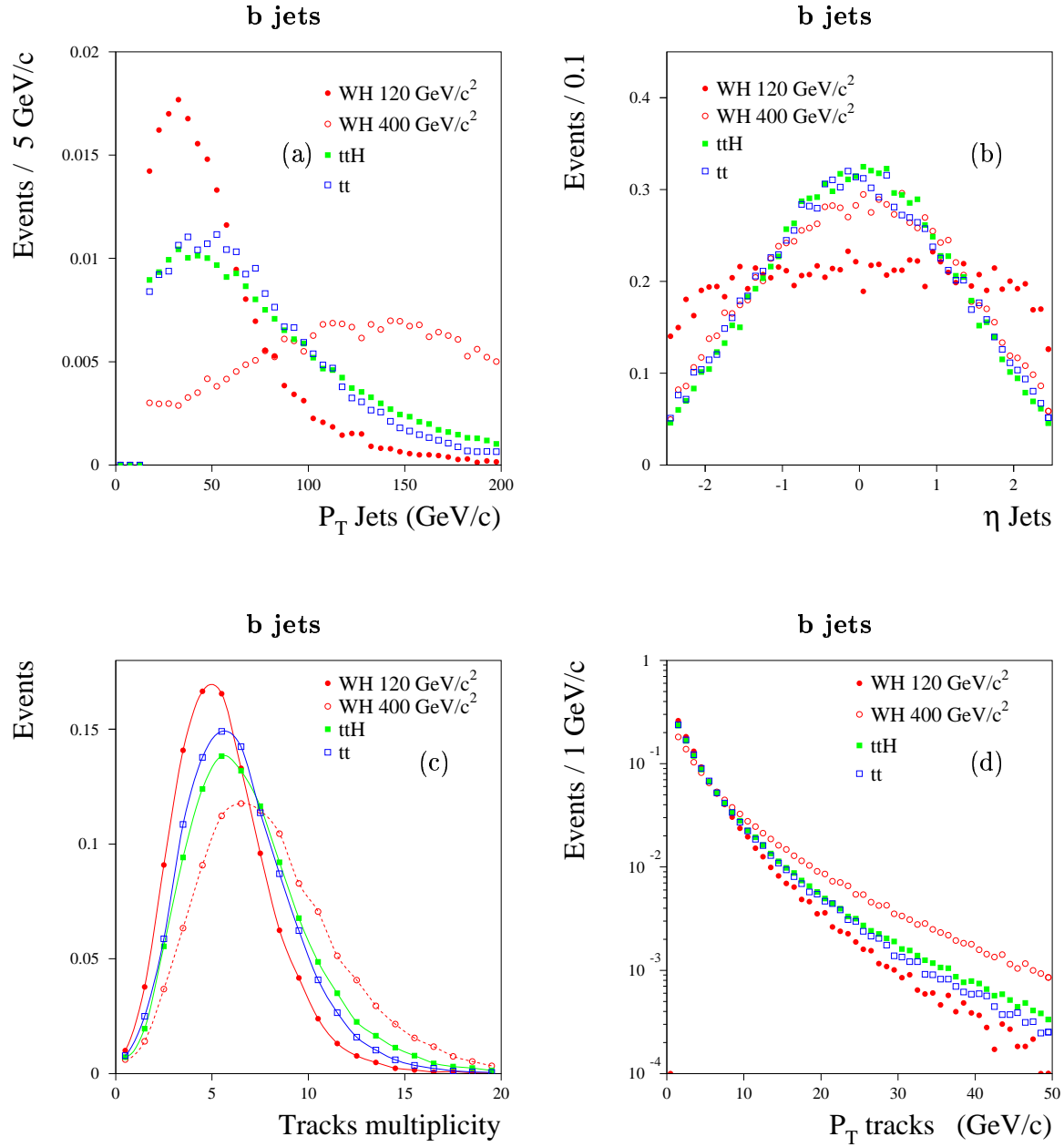


Figure 25:  $p_T$  (a) and pseudo-rapidity (b) distributions of b jets, track multiplicity (c) and  $p_T$  track distributions (d) in b-jets for WH(120), WH(400),  $t\bar{t}H$  and  $t\bar{t}$  events.

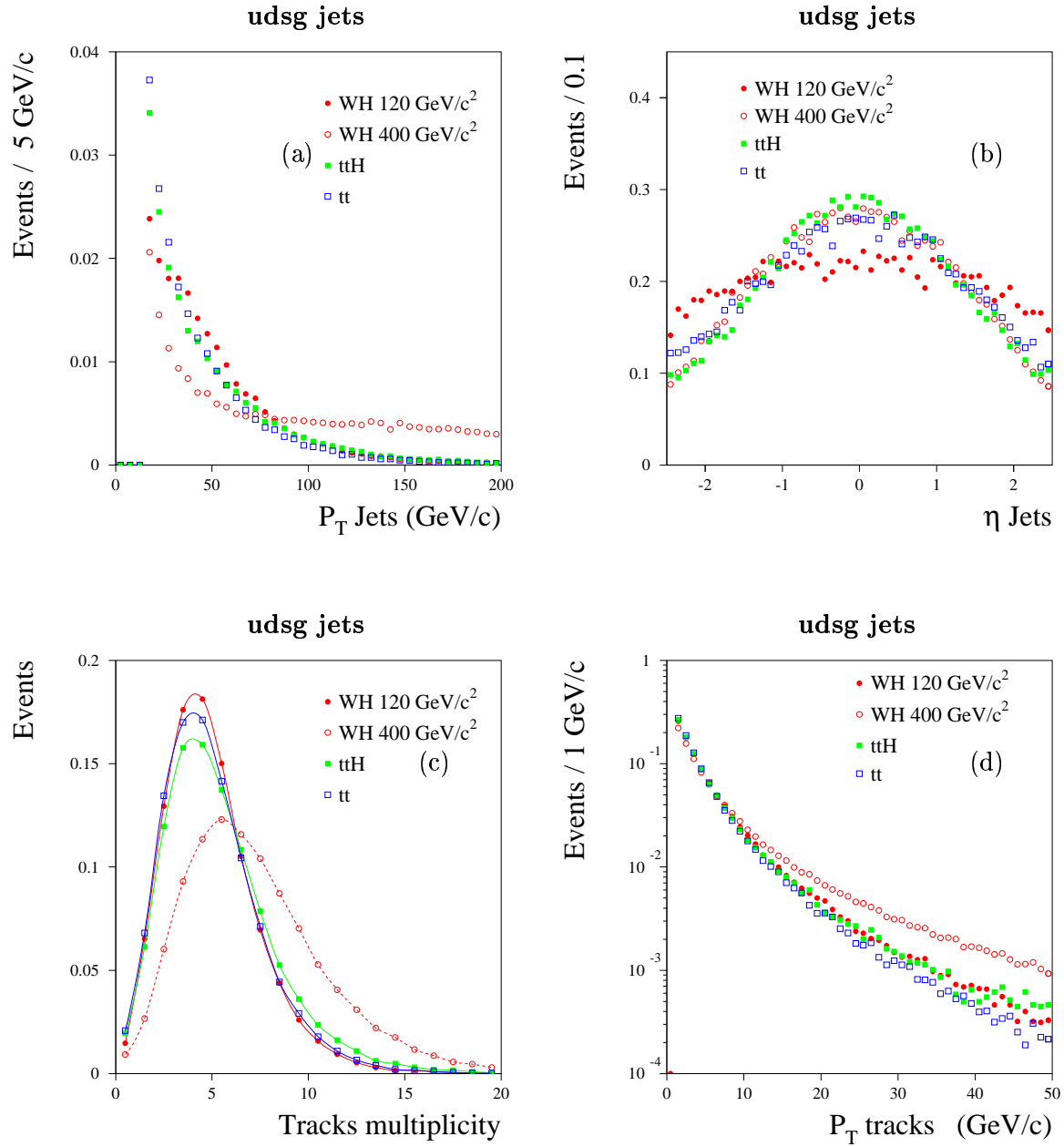


Figure 26:  $p_T$  (a) and pseudo-rapidity (b) distributions of light jets, track multiplicity (c) and  $p_T$  track distributions (d) in light jets for WH(120), WH(400),  $t\bar{t}H$  and  $t\bar{t}$  events.

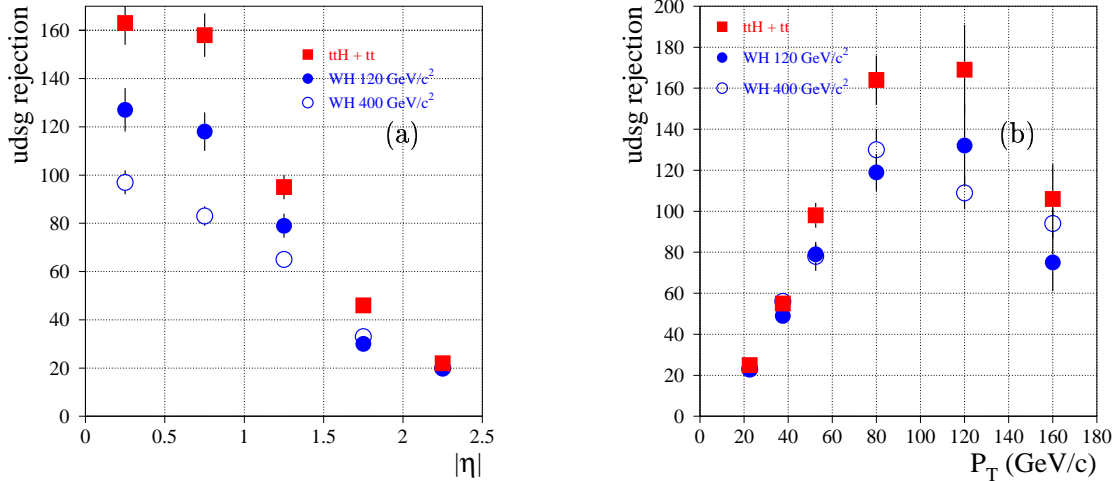


Figure 27: Comparison of the rejection factor for light jets in  $t\bar{t}H+t\bar{t}$ , WH(120 GeV/c<sup>2</sup>) and WH(400 GeV/c<sup>2</sup>) events, as a function of  $|\eta|$  (a) and  $p_T$  (b). Only the isolated jets in  $t\bar{t}H+t\bar{t}$  events are used in these distributions.

very high  $p_T$  on average. For WH(120) events, the lower performances come from the lower mean  $p_T$  jet value: jets with  $p_T$  below 50 GeV/c are more difficult to identify.

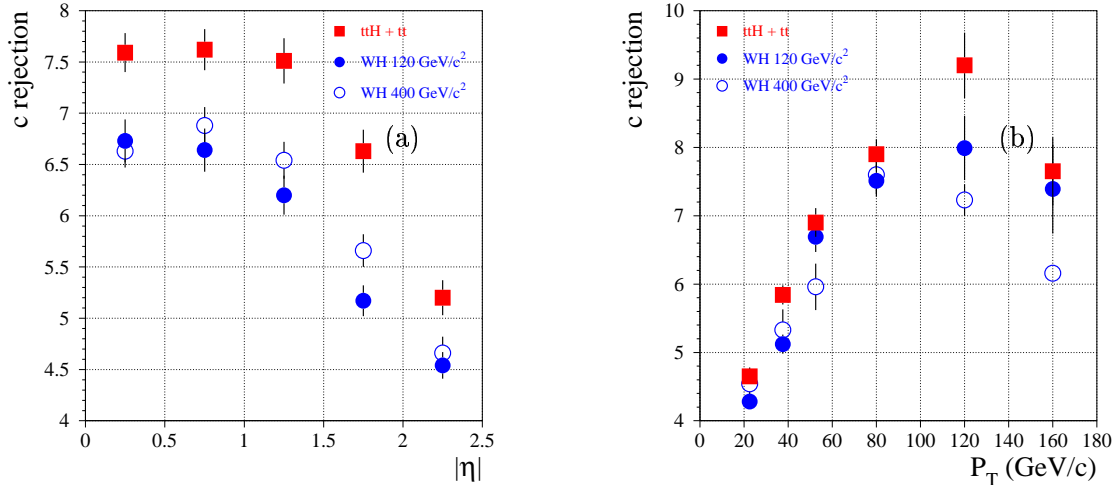


Figure 28: Comparison of the rejection factor for c-jets in  $t\bar{t}H+t\bar{t}$ , WH(120) and WH(400) events, as a function of  $|\eta|$  (a) and  $p_T$  (b). Only isolated jets in  $t\bar{t}H+t\bar{t}$  events are used in these distributions.

Fig. 28a and Fig. 28b give the corresponding comparisons for c-jet rejection.

### 6.3 Performance at high luminosity

The performances of the simple 2D and 3D algorithms have also been estimated in the very busy environment of LHC running at high luminosity  $\mathcal{L} = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . In these studies, ATLFAST jets have been used, with the *low luminosity* smearing. Cuts on the track  $p_T$  at 0.9 GeV/c and on the number of precision hits at 8 have been applied already at the reconstruction level. At such a high luminosity, the track multiplicity is very high and pattern recognition is *a priori* difficult. However, thanks to the high granularity of the pixel

detector and the excellent skill of xKalman for track reconstruction in dense environment, the degradation is small. This is illustrated in Table 23, which gives the rejections of light

	$t\bar{t}H/t\bar{t}$	WH (120)	WH (400)
2D $\epsilon_b = 50\%$	$274 \pm 8$	$177 \pm 10$	$156 \pm 5$
$\epsilon_b = 60\%$	$79 \pm 1$	$54 \pm 2$	$55 \pm 1$
3D $\epsilon_b = 50\%$	$524 \pm 23$	$357 \pm 29$	$241 \pm 9$
$\epsilon_b = 60\%$	$133 \pm 3$	$104 \pm 5$	$79 \pm 2$

Table 23: Light jet rejections for 2D/3D b-tagging, complete ATLAS ID, high luminosity, no module/chip inefficiency and z-coordinate of the primary vertex from the MC.

jets obtained with a perfect detector and the longitudinal coordinate of the primary vertex taken from the Monte Carlo. The numbers should be compared to those in Tables 2,3 for WH events and Table 16 for  $t\bar{t}H/t\bar{t}$  events. The degradations are below 10% and 20% for the 2D and 3D algorithms respectively.

However, one of the crucial aspects is the identification of the correct interaction point from the hard event among the many pile-up vertices. Table 24 shows the results for the most realistic detector proposed at high luminosity to date: the complete ATLAS detector, a z-pitch of 300  $\mu\text{m}$  in the b-layer, and 1% (2%) inefficiencies for the modules (chips) in all pixel layers.

	$t\bar{t}H/t\bar{t}$		WH (120)		WH (400)	
	udsg jet	c jet	udsg jet	c jet	udsg jet	c jet
2D $\epsilon_b = 50\%$	$228 \pm 7$	$11.4 \pm 0.2$	$132 \pm 7$	$8.9 \pm 0.2$	$133 \pm 4$	$10.2 \pm 0.2$
$\epsilon_b = 60\%$	$67 \pm 1$	$6.6 \pm 0.1$	$42 \pm 1$	$5.3 \pm 0.1$	$48 \pm 1$	$5.9 \pm 0.1$
3D $\epsilon_b = 50\%$	$354 \pm 12$	$13.0 \pm 0.2$	$157 \pm 9$	$10.2 \pm 0.2$	$151 \pm 5$	$11.3 \pm 0.2$
$\epsilon_b = 60\%$	$100 \pm 2$	$7.5 \pm 0.1$	$50 \pm 1$	$5.8 \pm 0.1$	$58 \pm 1$	$6.6 \pm 0.1$

Table 24: Light and c-jet rejections for 2D/3D b-tagging, complete ATLAS ID, high luminosity and 1/2% module/chip inefficiencies.

The degradation induced by the use of the reconstructed primary vertex is huge, especially for the 3D algorithm, and for hard events with low track multiplicity. These results should however be considered as very conservative, since the search of the primary vertex has been tuned on events without pile-up. As stated above, other information should be used to increase the probability of finding the correct hard interaction point. The influence of the pile-up on the c-jet rejection is small (see Tables 22 and 24).

## 6.4 Parametrisation for ATLFAST

Most of the physics studies in ATLAS are done with fast simulations using the ATLFAST program. The results of the full simulation should be parametrised in order to be used in a simple way in these fast simulation studies. As demonstrated in the previous sections the b-tagging performance depends strongly on the pseudo-rapidity and transverse momentum of jets. This motivates us to give a b-tagging parametrization as function of these two variables. Since the dependence on the process type is rather weak or not visible, events from different datasets (WH (120), WH (400),  $t\bar{t}H$ ,  $t\bar{t}$  and  $t\bar{t} b\bar{b}$ ) are combined together in order to have the highest statistics.

As an example the light jet rejection is given in Table 25 as a function of  $p_T$  and  $|\eta|$  at 60% b-jet efficiency with the 3D method at low luminosity in the Initial ATLAS layout.

$ \eta $ $p_T$ (GeV/c)	0.0-0.5	0.5-1.0	1.0-1.5	1.5-2.0	2.0-2.5
15-30	$40 \pm 1$	$39 \pm 1$	$27 \pm 1$	$15 \pm 0$	$9 \pm 0$
30-45	$101 \pm 6$	$112 \pm 7$	$83 \pm 5$	$35 \pm 1$	$20 \pm 1$
45-60	$169 \pm 15$	$192 \pm 19$	$133 \pm 12$	$57 \pm 4$	$29 \pm 1$
60-100	$250 \pm 23$	$244 \pm 23$	$183 \pm 16$	$69 \pm 4$	$39 \pm 2$
100-140	$198 \pm 24$	$278 \pm 43$	$171 \pm 22$	$71 \pm 7$	$39 \pm 3$
140-180	$138 \pm 19$	$227 \pm 41$	$127 \pm 19$	$50 \pm 5$	$35 \pm 4$
180-220	$106 \pm 16$	$119 \pm 20$	$92 \pm 14$	$44 \pm 5$	$21 \pm 2$
220-260	$85 \pm 14$	$69 \pm 11$	$44 \pm 6$	$23 \pm 3$	$13 \pm 2$
> 260	$19 \pm 1$	$17 \pm 1$	$16 \pm 1$	$8 \pm 0$	$9 \pm 1$

Table 25: Light jet rejections for the 3D algorithm as a function of  $p_T$  and  $|\eta|$  at 60% b-tagging efficiency, for the Initial ATLAS with two pixel layers,  $400 \mu\text{m}$  z-pitch in the b-layer, at low luminosity and 1/2% module/chip inefficiencies in all layers and 0.5/1% module/chip inefficiencies in b-layer.

The main trend in the two variables parametrization is the best performance in the  $p_T$  region of 60-140 GeV/c and rapid degradation at high pseudo-rapidity. One can see that the peak in the  $p_T$  dependence of the rejection becomes less pronounced in the region of high pseudo-rapidity. This table can be directly used for physics analyses. More results on the light jets and c-jets rejection parametrisations at 50 % and 60 % b-tagging efficiencies are given in [18]. It will be extended in the near future to the SV b-tagging methods.

## 7 Effect of the Silicon Tracker misalignment

The effects of the detector misalignment were studied with the  $t\bar{t}H$  and  $t\bar{t}$  events. The Initial layout with  $400 \mu\text{m}$  b-layer pixel pitch was used with low luminosity pile-up. The additional pixel chip and module inefficiencies were 2(1)% respectively. For the b-layer these inefficiencies were reduced to 1(0.5)% according to the scenario of selecting the best modules for the b-layer. The misalignment was introduced as a random Gaussian shift of the modules in the  $R\phi$  and  $z$  directions in the local coordinate system. No rotational misalignment was assumed in the simulation.

The usual requirements on the alignment errors from the Inner Detector TDR [4] (p.345, p.510) are:  $\sigma_{R\phi} = 5 \mu\text{m}$  and  $\sigma_z = 10 \mu\text{m}$  for the pixel barrel,  $\sigma_{R\phi} = 12 \mu\text{m}$  and  $\sigma_z = 50 \mu\text{m}$  for the SCT barrel. Studies of alignment with tracks [1] (p.94) show in some cases the possibility to achieve a better precision, up to  $1 \mu\text{m}$  both for pixel and SCT modules. However the systematic uncertainties on the alignment errors are not yet completely understood. Therefore in this study equal values of misalignment errors were assumed for pixel and SCT modules for simplicity. Higher values of the misalignment errors should be considered during the commissioning period of the ATLAS Inner Detector during the first year of operation. It is usually assumed that these errors will gradually decrease with the accumulation of statistics and improved understanding of the systematics during the first few months. Four sets of alignment errors have been studied:  $\sigma_{R\phi} = 0, 5, 10, 20 \mu\text{m}$  and  $\sigma_z = 0, 15, 30, 60 \mu\text{m}$  respectively. The results are presented in the Table 26 for the 2D method. In this case the



transverse misalignment errors are the most important. Even for the smallest misalignment, a 2-5 % degradation of the performance is observed. For a typical misalignment at the beginning of commissioning,  $\sigma_{R\phi} = 20 \mu\text{m}$ , the loss of performance is as high as 34 %.

	$R_0 = R_u$ perfect alignm.	$R_u$ $\sigma_{R\phi} = 5\mu\text{m}$ $\sigma_z = 15\mu\text{m}$	$R_u$ $\sigma_{R\phi} = 10\mu\text{m}$ $\sigma_z = 30\mu\text{m}$	$R_u$ $\sigma_{R\phi} = 20\mu\text{m}$ $\sigma_z = 60\mu\text{m}$
2D $\epsilon_b = 50\%$	$164 \pm 4$	$157 \pm 4$	$146 \pm 3$	$106 \pm 2$
$\epsilon_b = 60\%$	$53 \pm 1$	$52 \pm 1$	$49 \pm 1$	$39 \pm 1$
	$R_u/R_0$	$R_u/R_0$	$R_u/R_0$	$R_u/R_0$
2D $\epsilon_b = 50\%$	1.00	0.95	0.89	0.65
$\epsilon_b = 60\%$	1.00	0.98	0.91	0.74

Table 26: Comparison of the rejections obtained with different silicon detector misalignments for the standard (2D) b-tagging algorithm.

The results for the 3D method are presented in the Table 27. In this case both transverse and longitudinal alignment errors contribute. 3D b-tagging is slightly less sensitive to misalignment errors than 2D b-tagging.

	$R_0 = R_u$ perfect alignm.	$R_u$ $\sigma_{R\phi} = 5\mu\text{m}$ $\sigma_z = 15\mu\text{m}$	$R_u$ $\sigma_{R\phi} = 10\mu\text{m}$ $\sigma_z = 30\mu\text{m}$	$R_u$ $\sigma_{R\phi} = 20\mu\text{m}$ $\sigma_z = 60\mu\text{m}$
3D $\epsilon_b=50\%$	$262 \pm 8$	$259 \pm 8$	$237 \pm 7$	$175 \pm 4$
$\epsilon_b=60\%$	$81 \pm 1$	$79 \pm 1$	$74 \pm 1$	$57 \pm 1$
	$R_u/R_0$	$R_u/R_0$	$R_u/R_0$	$R_u/R_0$
3D $\epsilon_b=50\%$	1.00	0.99	0.91	0.67
$\epsilon_b=60\%$	1.00	0.97	0.92	0.71

Table 27: Comparison of the rejections obtained with different silicon detector misalignments for the 3D b-tagging algorithm.

In this study the correlation between the misalignment errors of different detectors were neglected. In the real case of alignment with tracks and with the primary vertex constraint the misalignment errors in a given azimuthal and polar sectors and different layers could be correlated. In this case the errors in the impact parameter due to the misalignment could be smaller than in the simplified estimation neglecting these correlations due to the vertex constraint in the alignment procedure. In future studies, in order to test the importance of this effect, the full misalignment error matrix with all correlation terms will be needed to generate the detector misplacements accordingly. For the time being such an error matrix is not available from alignment simulation studies and the present study should therefore be considered as a conservative estimate.

The main conclusion is that the alignment precision should be better than  $5 \mu\text{m}$  and  $15 \mu\text{m}$  in transverse and longitudinal coordinates respectively in order to have a negligible impact on the b-tagging quality.

## 8 Conclusions

In this note the b-tagging results obtained in the framework of the ATLAS DC1 were presented. With the introduction in the simulation of the latest detector layouts and many realistic running conditions the b-tagging performance is considerably worse than the previous Physics TDR results. In particular the following factors contributed to the degradation of the performances results:

- The latest detector layout reduced the performance by a factor  $\sim 0.5 \pm 0.2$ , essentially due to the increase of the material and radius of the b-layer.
- The eventual staging of the intermediate pixel layer reduced the performance by a factor  $\sim 0.7 \pm 0.1$
- The inclusion of the realistic effects of the pile-up, detector inefficiency and alignment reduced the performance by a factor of  $\sim 0.75 \pm 0.05$
- Switching to  $400 \mu\text{m}$  pixels in b-layer reduce the performance by  $\sim 10 \%$ , but can be almost completely compensated by the increase of the efficiency by choosing the best modules in the b-layer.

The net effect of all these degradations is a reduction of the expected light jet rejection to a level which is too low for most of the physics analyses at LHC.

Therefore, efforts were made to improve the reconstruction and analysis methods, leading to substantial progress. The following efforts contributed to this improvement:

- The developments of the xKalman pattern recognition and track fitting package increased the performance by a factor  $\sim 1.8 \pm 0.4$  with respect to the version used for the Physics TDR.
- 3D b-tagging impact parameter method gained a factor  $\sim 1.9 \pm 0.4$  with respect to the 2D method.
- The introduction of the Secondary Vertex method and suppression of the tracks from secondary interactions and strange particle decays improved further the performance by a factor  $\sim 2.8 \pm 0.6$ .

Taking into account all these improvements the light jet rejection becomes  $\sim 150 \pm 20$  at 60% b-tagging efficiency with all realistic effects taken into account. This result super-seed the traditional ATLAS default assumption of light jet rejection  $R_{uds} = 100$  at 60% b-tagging efficiency. Such a performance opens very important possibility to target higher b-tagging efficiency of  $\sim 65\%$  with reasonable rejection of the light quark jets, which would greatly improve the significance of the light Higgs signal and efficiency of SUSY searches. In the most optimistic situation (3 pixel layers, increased efficiencies of pixel detectors,  $300\mu\text{m}$  reduced z-pitch of the b-layer) one can put a goal to reach even 70% b-tagging efficiency keeping  $R_{uds} = 100$ .

The sensitivity of the b-tagging performance to several systematic effects, like inefficiencies, longitudinal pixel pitch, pile-up, primary vertex reconstruction, jet reconstruction, jet overlap, jet flavour and origin, jet  $p_T$  and pseudo-rapidity and misalignment were studied in this note.

Much more developments are subject of ingoing and future work:

- Parametrisation of SV performance for fast simulation.

- Implementation of the SV algorithm in the ATHENA framework.
- Combination of the latest methods of space b-tagging with soft lepton b-tagging.
- New methods using for b-tagging multiple secondary vertices and neutral strange particles.
- Development of vertex c-tagging methods.
- Development of experimental b-tagging calibration procedures.
- Optimisation of multiple jet b-tagging, in particular tagging events with four b-jets for Higgs search in  $t\bar{t}H$  process.

## Acknowledgments

We are grateful to I. Gavrilenko for the continuous support in the use of xKalman package, P. Nevski for the support in using ATLSIM-DICE for simulation and reconstruction, K. Bernardet for the work on the data processing and management, D. Barberis, E. Ros, F. Gianotti, L. Rossi and D. Rousseau for helpful discussions on the b-tagging and ID performance.

## Appendix

### List of b-tagging data samples

Only the samples marked with a star have been used for these studies. The samples labelled by  $t\bar{t}jj$  have been biased at the generation level with an ATLFast-based filter (see Section 5.1). All samples have been generated with PYTHIA, except the 2332 one, generated with AcerMC interfaced to PYTHIA and biased at the generation level too.

Process	$m_{H/A/h}$ (GeV/ $c^2$ )	Pixel detector	Dataset	Events
WH $\rightarrow \mu\nu b\bar{b}$	120		2055*	20K
WH $\rightarrow \mu\nu u\bar{u}$	120		2056*	50K
WH $\rightarrow \mu\nu c\bar{c}$	120		2057*	20K
WH $\rightarrow \mu\nu g\bar{g}$	120		2078	50K
WH $\rightarrow \mu\nu b\bar{b}$	120	2 layers	2062*	20K
WH $\rightarrow \mu\nu u\bar{u}$	120	2 layers	2063*	100K
WH $\rightarrow \mu\nu c\bar{c}$	120	2 layers	2065*	20K
WH $\rightarrow \mu\nu g\bar{g}$	120	2 layers	2079	50K
WH $\rightarrow \mu\nu b\bar{b}$	400		2323*	20K
WH $\rightarrow \mu\nu u\bar{u}$	400		2324*	50K
WH $\rightarrow \mu\nu c\bar{c}$	400		2325*	20K
WH $\rightarrow \mu\nu g\bar{g}$	400		2326	50K
WH $\rightarrow \mu\nu b\bar{b}$	400	2 layers	2317*	20K
WH $\rightarrow \mu\nu u\bar{u}$	400	2 layers	2318*	50K
WH $\rightarrow \mu\nu c\bar{c}$	400	2 layers	2319*	20K
WH $\rightarrow \mu\nu g\bar{g}$	400	2 layers	2320	50K
H $\rightarrow hh \rightarrow b\bar{b}b\bar{b}$	300(109)		2071	10K
H $\rightarrow hh \rightarrow b\bar{b}b\bar{b}$	300(109)	2 layers	2075	10K
$b\bar{b}$ A/H $\rightarrow b\bar{b}b\bar{b}$	400		2327	10K
$b\bar{b}$ A/H $\rightarrow b\bar{b}b\bar{b}$	400	2 layers	2321	10K
$b\bar{b}$ A/H $\rightarrow b\bar{b}u\bar{u}$	400		2328	50K
$b\bar{b}$ A/H $\rightarrow b\bar{b}u\bar{u}$	400	2 layers	2322	50K
$t\bar{t}$ A/H $\rightarrow b\bar{b}b\bar{b}$	400		2070	10K
$t\bar{t}$ A/H $\rightarrow b\bar{b}b\bar{b}$	400	2 layers	2072	10K
$t\bar{t}$ A/H $\rightarrow b\bar{b}u\bar{u}$	400		2074	50K
$t\bar{t}$ A/H $\rightarrow b\bar{b}u\bar{u}$	400	2 layers	2076	50K
$t\bar{t}H \rightarrow \ell\nu\ell\nu b\bar{b} b\bar{b}$	120		2058	20K
$t\bar{t}H \rightarrow \ell\nu\ell\nu b\bar{b} b\bar{b}$	120	2 layers	2066	20K
$t\bar{t} \rightarrow WW b\bar{b}$			2082	20K
$t\bar{t} \rightarrow WW b\bar{b}$		2 layers	2083	20K
$t\bar{t}H \rightarrow \ell\nu q\bar{q}' b\bar{b} b\bar{b}$	120		2311	20K
$t\bar{t}H \rightarrow \ell\nu q\bar{q}' b\bar{b} b\bar{b}$	120	2 layers	2310	20K
$t\bar{t} jj \rightarrow \ell\nu q\bar{q}' b\bar{b} jj$			2313	20K
$t\bar{t} jj \rightarrow \ell\nu q\bar{q}' b\bar{b} jj$		2 layers	2312	20K

Table 28: Data samples produced for DC1 b-tagging studies with the full ATLAS detector. The cut  $|\eta| < 3$  has been applied for all samples except the last four where the  $|\eta|$  upper limit has been increased to 7. The 2083, 2310 and 2312 samples have 2 pixel barrel layers but  $2 \times 3$  disks.

Process	$m_H$ (GeV/ $c^2$ )	Pixel detector	Dataset	Events
$t\bar{t}H \rightarrow l\nu l\nu b\bar{b}b\bar{b}$	120		2089	40K
$t\bar{t}H \rightarrow l\nu l\nu b\bar{b}b\bar{b}$	120	2 layers	2090	40K
$t\bar{t} \rightarrow WW b\bar{b}$			2073	120K
$t\bar{t} \rightarrow WW b\bar{b}$		2 layers	2077	120K
$t\bar{t}H \rightarrow l\nu q\bar{q}' b\bar{b} b\bar{b}$	120		2306*	20K
$t\bar{t}H \rightarrow l\nu q\bar{q}' b\bar{b} b\bar{b}$	120	2 layers	2303*	20K
$t\bar{t} jj \rightarrow l\nu q\bar{q}' b\bar{b} jj$			2307*	100K
$t\bar{t} jj \rightarrow l\nu q\bar{q}' b\bar{b} jj$		2 layers	2304*	100K

Table 29: Data samples produced for DC1 b-tagging studies (ATLAS ID only). In these samples, “2 layers” means 2 barrel layers and  $2 \times 3$  disks.

Process	$m_H$ (GeV/ $c^2$ )	Filter	Dataset	Events
$t\bar{t}H \rightarrow l\nu l\nu b\bar{b} b\bar{b}$	120	$ \eta  < 3$	2096	20K
$t\bar{t}H \rightarrow l\nu l\nu b\bar{b} u\bar{u}$	120	$ \eta  < 3$	2097	20K
$t\bar{t} \rightarrow WW b\bar{b}$		$ \eta  < 3$	2098	100K
$t\bar{t}H \rightarrow l\nu q\bar{q}' b\bar{b} b\bar{b}$	120	$ \eta  < 7$	2308*	20K
$t\bar{t}H \rightarrow l\nu q\bar{q}' b\bar{b} u\bar{u}$	120	$ \eta  < 7$	2305	20K
$t\bar{t} jj \rightarrow l\nu q\bar{q}' b\bar{b} jj$		$ \eta  < 7$	2309*	100K
$t\bar{t} b\bar{b} \rightarrow l\nu q\bar{q}' b\bar{b} b\bar{b}$		$ \eta  < 7$	2332*	20K

Table 30: Data samples produced for DC1 b-tagging studies with the Initial ATLAS detector.

## References

- [1] ATLAS Collaboration, *ATLAS detector and physics performance*, Technical Design Report, CERN/LHCC/99-15.
- [2] S. Gonzalez de la Hoz, E. Ros and M. Vos, *The b-tagging performance of the complete ATLAS DC1 layout using WH events*, ATL-COM-INDET-2003-017.
- [3] E. Richter-Was, D. Froidevaux and L. Poggioli, *ATLFAST 2.0 a fast simulation package for ATLAS*, ATL-PHYS-98-131.
- [4] ATLAS Collaboration, *Inner Detector Technical Design Report*, Technical Design Report, CERN/LHCC/97-16 and CERN/LHCC/97-17.
- [5] ATLAS Collaboration, *ATLAS Pixel detector*, Technical Design Report, CERN/LHCC/98-13.
- [6] ATLAS Collaboration, *ATLAS Data Challenge 1*, ATL-SOFT-2003-012.
- [7] I. Gavrilenko, *Description of Global Pattern Recognition Program (XKalman)*, ATL-INDET-97-165; ATL-I-PN-165.
- [8] Particle Data Group, Phys.Rev. 66 (2002), p.232
- [9] V. Kostioukhine, *VKalVrt - package for vertex reconstruction in ATLAS*, ATL-PHYS-2003-031.
- [10] V. Kostioukhine, *Secondary vertex based b-tagging*, ATL-PHYS-2003-033.
- [11] M. Wolter and A. Kaczmarska, *Combining b-tagging methods using a neural network approach*, ATL-INDET-2000-023.
- [12] *b-layer pixel length changes to 400 $\mu$ m*, ATLAS Engineering Change Request, ATL-IP-EC-0005, August 2003.
- [13] V. Kostioukhine, J. Levêque, A. Rozanov and J. B. de Vivie, *Physics impact of the change of the pixel b-layer longitudinal pitch from 300 to 400 microns*, ATL-INDET-2003-016.
- [14] J. Cammin and M. Schumacher, *The ATLAS discovery potential for the channel  $t\bar{t}H$ ,  $H \rightarrow b\bar{b}$* , ATL-PHYS-2003-024.
- [15] <http://ific.uv.es/~camarena/jet.html>
- [16] J. Levêque, *Recherche d'un boson de Higgs produit en association avec une paire de quarks top dans l'expérience ATLAS*, PHD thesis, Université de la Méditerranée, Aix-Marseille II, 30 June 2003.
- [17] J. Cammin, *The channel  $t\bar{t}H$ ,  $H \rightarrow b\bar{b}$* , talk given at the ATLAS Higgs Meeting, May 2003.  
  
<http://agenda.cern.ch/fullAgenda.php?ida=a03583>
- [18] <http://marpix1.in2p3.fr/Pixel/dice/btagdc1/btagdc1.html>