# TAGGING *b*-JETS USING LOW- $p_{T}$ ELECTRONS\*

Stanisław Jagielski

Faculty of Physics and Nuclear Techniques, University of Mining and Metallurgy 30-055 Kraków, al. Mickiewicza 30, Poland

ANNA KACZMARSKA AND MARCIN WOLTER

Institute of Nuclear Physics 30-059 Kraków, Kawiory 26a, Poland

(Received December 18, 1999)

This note presents a study of *b*-jet tagging in ATLAS using soft electrons from semileptonic *b*-quark decays. Both Inner Detector and EM Calorimeter information is used by the *b*-tagging algorithm to identify signal electrons inside jets. The performance of this algorithm in terms of expected *b*-tagging efficiency and non-*b* jet rejection is discussed in detail over the small central rapidity range and for low luminosity operation. For an algorithm efficiency of 50% (and an overall efficiency of 7.2% taking into account the branching ratio of  $b \rightarrow e$  decays), rejections of 554 against *u*-jets, 192 against gluon jets and of 47 against *c*-jets can be achieved. Such a softelectron *b*-tagging algorithm can be combined with the higher-efficiency vertexing algorithm to improve the *b*-jet tagging performance.

PACS numbers: 07.05.Kf

### 1. Introduction

The capability for efficient identification of high- $p_{\rm T}$  jets originating from *b*-quark fragmentation plays a key role in the analysis of various physical processes expected to be observed in the ATLAS experiment. Many interesting physics channels have final states involving *b*-quarks:

— Higgs boson decays,  $H \to b\overline{b}$ , from associated WH and  $t\overline{t}H$  production [1]. This channel is important for SM Higgs searches and also for searches of Higgs bosons in the MSSM [2],

<sup>\*</sup> Work supported in part by Polish Government grants 115/E-343/SPUB/P03/004/97, 115/E-343/SPUB/P03/157/98, 2P03B00212 and KBN/S2000/IFJ/009/1998.

— cascade decays of SUSY particles [3, 4],

— top-quark decays  $(t \to Wb \text{ and possibly } t \to WbZ)$  [5].

In all of the above processes jets originating from lighter quarks and from gluons are sources of a potentially large background. Therefore the observability of signals in multi-*b*-jet final states generally requires a high rejection of non-*b* jets (typically  $\approx 10$ ) together with the highest possible efficiency for *b*-jet identification.

It was already shown by the CDF [6] and LEP [7] experiments, that *b*-tagging can be achieved by two independent methods. Vertexing algorithms make use of the relatively long *B* lifetime, while lepton tagging is based on the identification of the soft leptons from *B* decay. While the efficiency of the vertexing algorithms is only limited by the detector performance, the expected efficiency of the soft lepton tagging method is limited also by the fraction of *B*-hadron decays containing leptons ( $\approx 17\%$  per lepton family).

So far, ATLAS evaluations of the performance of the soft-lepton tagging have been restricted to the case of electrons [9-11]; however, similar studies using muons are foreseen [12]. It is also foreseen that both methods, vertexing [8] and soft-lepton tags, will be combined [13] to achieve the highest possible *b*-tagging performance.

As in [9-11], the present note also only discusses the *b*-tagging performance based on the identification of the soft electrons over a small central rapidity range. However the previous analyses have been completed and extended through:

- the analysis of a large variety of background samples using much higher statistics;
- the application of a  $\gamma$ -conversion and Dalitz decay finding algorithm;
- the reconstruction of the *b*-jet direction using fast simulation;
- the use of a combined discriminating function, built from several EM Calorimeter and Inner Detector specific variables and used to optimise the rejection power versus the *b*-tagging efficiency.

The present paper is organised as follows. In Sections 2 and 3, the analysed samples and the kinematical properties of signal and background events are discussed. Section 4 presents in detail the event reconstruction after full detector simulation, and Section 5 specifies the variables and the algorithm used by the *b*-tagging procedure. Section 6 discusses the results of the standard algorithm and also, in some detail, the impact of the use of some

of the variables on the final performance. Conclusions and an outlook are presented in Section 7. Some more technical details are given in Appendices A to C.

### 2. Samples

### 2.1. Generated samples

The analysis is performed using events from Higgs-boson associated production, WH with  $W \to \mu\nu$  and  $m_{\rm H} = 100$  GeV. Several event samples are considered: the signal sample of  $H \to b\bar{b}$  events and background samples of  $H \to gg$ ,  $c\bar{c}$ ,  $u\bar{u}$  events. The samples were generated using PYTHIA 5.702 and JETSET 7.408. The following preselection was applied on the generated events before performing full simulation of the ATLAS detector:

- for signal events:
  - both *b*-quarks (before final state radiation, FSR) are required to have  $p_{T}^{b} > 15$  GeV and  $|\eta_{b}| < 0.3$ ;
  - at least one electron with  $p_{\rm T}^e > 1$  GeV and  $|\eta_e| < 0.6$  is required to be found in a cone  $\Delta R = \sqrt{[\Delta \eta(e, b)]^2 + [\Delta \phi(e, b)]^2} < 0.4$ around each *b*-quark axis (before FSR);
- and for background events:
  - both partons (g,c,u) are required to have  $p_{\mathrm{T}}^{g,c,u} > 15~\mathrm{GeV}$  and  $|\eta_{g,c,u}| < 0.3$  .

A signal sample was filtered for both hard partons and for electrons, while a background sample only for hard partons. For signal events, this preselection produces a sample of *b*-jets which are expected to contain electrons coming from cascade decays within the jet cone<sup>1</sup>, to have high jet reconstruction efficiency and to have fragmentation products limited to the small rapidity range under study here ( $|\eta| < 0.6$ ). For the background events, the filter selects events with the expected high jet reconstruction efficiency and wuth fragmentation products limited to the small rapidity range.

<sup>&</sup>lt;sup>1</sup> In the first step of the analysis, jets are reconstructed with a fast algorithm and the *b*-quark after FSR is searched for in a cone around the reconstructed jet axis. Jets containing *b*-quarks with  $p_{\rm T}^{b-{\rm quark}} > 5\,{\rm GeV}$ ,  $\Delta R < 0.2$  are labelled as *b*-jets candidates. As explained later, the axis of the reconstructed jets is slightly displaced from the flight direction of the *b*-quark from  $H \rightarrow b\bar{b}$  decay. Nevertheless, the filter applied above is very efficient and, for 96% of the reconstructed jets labelled as *b*-jets, an electron of  $p_{\rm T}^e > 1$  GeV and  $\Delta R < 0.4$  is found after the jet reconstruction.

After filtering about 2100 signal  $H \to b\bar{b}$  events, 7000 events of  $H \to u\bar{u}$ , 4000 events of  $H \to c\bar{c}$  and 7200 events of  $H \to gg$  were processed through the ATLAS detector simulation using the DICE 96\_9 package [14] with activated simulation of the Inner Detector and the Electromagnetic Calorimeter.

### 2.2. Electrons in jets

True soft electrons are present both in the signal and the background jets. These electrons originate from the decays of particles from the hadronic cascade or from the interaction of particles with the detector material. Since the aim of the algorithm presented here is to identify the presence of a b-quark inside a reconstructed jet (the so called b-jet) by using as a tag soft electrons from B-hadron semileptonic decays, electrons will be classified as irreducible (or signal) and reducible (or background) throughout this note.

The signal (irreducible) electrons inside b-jets<sup>2</sup> came from following processes [15]:

- direct semileptonic decays of B mhadrons with an electron in the final state  $(b \rightarrow e)$ ;
- cascade decays of *B*-hadrons to charm hadrons, which later decay semileptonically with an electron in the final state  $(b \rightarrow c \rightarrow e)$ ;
- leptonic decays of  $J/\Psi$  coming from b decays  $(b \to J/\Psi \to e)$ ;
- decays of *B*-hadrons to  $\tau$ -leptons with susequent decays into electrons  $(b \to \tau \to e)$ .

In the background (non-b) jets such electrons are also expected. For example, inside the gluon jets, because of their larger heavy quark content, electrons from  $b \to e, b \to c \to e$  occur quite frequently and electrons from  $c \to e$  also obviously occur inside *c*-jets. The *u*-jet sample is almost free<sup>3</sup> such electrons. The presence of irreducible electrons in the background jets will of course limit the *b*-tagging performance at some level.

The background (reducible) electrons common to all types of jets arise from the following processes:

 $<sup>^2</sup>$  Jets from  $H \rightarrow b\bar{b}$  decays are labelled as *b*-jets if they contain a hard *b*-quark (after FSR) within a small cone  $\Delta R < 0.2$ . A similar definition is applied to *c*-jets, *u*-jets and *g*-jets, where the respective hard parton (before FSR) from H-boson decay is used to label the jet. One should note that, in general, it is not really possible in a parton-shower MC package such as PYTHIA/JETSET to identify the parent parton of a jet reconstructed using the final-statestable particles.

<sup>&</sup>lt;sup>3</sup> The contamination of the  $H \to u\bar{u}$  sample with heavy-flavour quarks from initialstate-radiation gluon splitting is below 2% for *c*-quarks and below 0.3% for *b*-quarks. In fact, in less than 0.2% of reconstructed *u*-jets, an electron from *D*-meson decay is identified at the particle level.

- semileptonic decays in the hadronic cascade;
- $\pi^0$  Dalitz decays  $(\pi^0 \to e^+ e^- \gamma);$
- $\gamma$ -conversions which occur in Inner Detector material.

While the signal electrons have very similar kinematic features in the signal and background jets (since in both cases they originate from the decays of heavy quarks), the kinematics of the background electrons is very different.

#### TABLE I

Fraction of jets containing electrons of a given origin (only electrons with  $p_{\rm T} \geq 2.0 \,\text{GeV}$  ( $p_{\rm T} \geq 1.0 \,\text{GeV}$ ) and within  $\Delta R \leq 0.4$  from the jet axis are considered). The numbers are given after applying filters described in the previous section and after full detector simulation.

Jet type	Fraction of jets containing electrons [%]						
	$\gamma$ conversions	$\pi^0$ Dalitz	$D ext{-hadrons}$	$B ext{-hadrons}$	Other		
					sources		
b-jets	7.4(12.3)	6.2(7.5)	40.1 (41.2)	50.1 (50.4)	8.6(9.2)		
$g ext{-jets}$	$6.8\ (12.5)$	1.0(1.5)	$0.40\ (0.57)$	$0.10\ (0.13)$	4.4(7.3)		
$u ext{-jets}$	6.7 (11.7)	1.0(1.6)	- (-)	- (-)	4.5(5.1)		
$c ext{-jets}$	$6.6\ (12.2)$	1.0(1.8)	6.4(7.4)	- (-)	5.2(5.9)		

Jet type	Fraction of jets containing electrons [%]						
	Irreducible from $B$ and $D$	$\operatorname{Reducible}$					
<i>b</i> -jets <i>g</i> -jets <i>u</i> -jets <i>c</i> -jets	90.2 (91.6) 0.5 (0.68) - (-) 6.4 (7.4)	$\begin{array}{ccc} 16.0 & (24.3) \\ 12.0 & (20.2) \\ 11.6 & (17.4) \\ 12.2 & (18.6) \end{array}$					

The presence of any true electron in the background jets makes their rejection difficult. The fraction of reconstructed jets containing true electron tracks, KINE tracks, with  $p_{\rm T} \geq 2.0$  GeV ( $p_{\rm T} \geq 1.0$  GeV) and within  $\Delta R \leq 0.4$  from the jet axis is given in Table I for all jet type studied here. The results in Table I are given after the full detector simulation of the events<sup>4</sup>, so they take into account also electrons produced in the interaction with the material of the detector. The fraction of jets containing reducible electron tracks is at the level of 12% for the background jets and charged

<sup>&</sup>lt;sup>4</sup> For the signal sample (*b*-jets), the sum is greater than 100% because, after filtering, these jets sometimes contain more than one electron.

particles transverse momentum threshold  $p_{\rm T}^{\rm thr} = 2.0 \,{\rm GeV}$  and is dominated by electron tracks from  $\gamma$ -conversions (~ 60%). It increases almost by 50% if the threshold is lowered to  $p_{\rm T}^{\rm thr} = 1.0 \,{\rm GeV}$ . By lowering the threshold, the fraction of *b*-jets containing irreducible electron tracks increases by at most few percent.

It is obvious that there is a natural limitation to the rejection capability of jets not originating from b-quarks using electrons as tagging particles. As can be concluded from Table I, the best rejection factor will be achievable for the u-jets, an intermediate one for the g-jets, and the worst one for the c-jets. Very roughly, if electrons from B and D hadrons are treated as irreducible, for a 50% efficiency of the tagging algorithm, the best rejection that can be expected is about 30 against c-jets, and about 400 against g-jets. However, since not only signal electrons but also background electrons and misidentified hadrons (mostly pions) can tag jets as b-jets the above numbers indicate rather the ultimate performance achievable using soft electron identification. However, if the more global features of the tracks in the bjets can also be exploited by the b-jet tagging algorithm, the final rejection capability could be even higher than the numbers specified above.

### 3. Event kinematics

# 3.1. Jets

In the previous studies [10], jets were considered as bunches of tracks collected in a cone  $\Delta R = 0.4$  around the initial parton direction. The direction of the initial parton does not always coincide however with the bary centre of the cascade of stable particles originating from this parton. Figure 1 shows the  $\Delta \eta$  and  $\Delta \phi$  distributions between the initial parton and the axis<sup>5</sup> of the closest reconstructed jet (top) and between the bary centre of the stable particles and the axis of the closest reconstructed jet (bottom). The reconstructed jet axis reproduces the bary centre of all the particles with much better resolution and much less tails than the direction (after FSR) of the initial parton.

The procedure used for the jet reconstruction with the fast simulation algorithm [16] is as follows:

• the transverse energy of the stable particles, except neutrinos and muons, is deposited in calorimeter cells of  $0.1 \times 0.1$  granularity  $(\eta \times \phi)$ ;

<sup>&</sup>lt;sup>5</sup> The jet axis is defined as the bary centre of the transverse energy depositions in calorimeter cells, calculated for cells in a cone  $\Delta R = 0.4$  with respect to the cell with the highest transverse energy deposition.



Fig. 1. Distributions of  $\Delta \eta$  (left) and  $\Delta \phi$  (right) between initial parton and the axis of the closest reconstructed jet (top); and between the bary centre of the stable particles and the axis of the closest reconstructed jet (bottom) for WH,  $H \rightarrow b\bar{b}$  events and for  $m_{\rm H} = 100$  GeV.

- all calorimeter cells with transverse energies greater than 1.5 GeV are used as possible initiators of clusters to verify if the total cluster transverse energy in a cone  $\Delta R = 0.4$  is greater than the threshold value of 15 GeV;
- the  $E_{\rm T}$ -weighted bary centre of the cluster (centre of gravity of the cluster) is taken as the axis of the jet;
- the cluster is labelled as originating from a *b*-jet if a *b*-quark of  $p_{\rm T}^b > 5$  GeV (after FSR) is found in a cone  $\Delta R < 0.2$  around the cluster direction and  $|\eta^{\rm jet}| < 2.5$ ;
- for this analysis the labelling for c, u, g-jets is done in the same way;
- energy smearing as expected for the ATLAS calorimetry is applied to the reconstructed cluster energy.

Table II shows the efficiencies for the jet reconstruction, the mean jets transverse momentum,  $\langle p_{\rm T}^{\rm jet} \rangle$ , and the average charged particle multiplicities within the jet cone as a function of the transverse momentum threshold on the particles and of the jet type. The efficiency for the jet reconstruction is defined as the probability that a hard parton from the decay of the Higgs boson (b, g, c, u) with  $p_{\rm T}^{b,g,c,u} > 15$  GeV leads to a reconstructed jet with  $p_{\rm T}^{\rm jet} > 15$  GeV labelled as a b-, g-, c-, u-jet. As can be seen from Table II the multiplicity of charged particles within the jet cone varies rapidly as a function of the threshold set on the transverse momentum of the particles: decrease of  $p_{\rm T}^{\rm thr}$  from 2 GeV to 1 GeV results in an increase in the multiplicity by about 30-40%. The fraction of reconstructed jets with no charged tracks with  $p_{\rm T} > 0.5$  GeV within the jet cone is below 1%. The expected average charged particle multiplicity,  $\langle n^{\rm ch} \rangle$ , is higher for b-jets and g-jets,  $\langle n^{\rm ch} \rangle \approx 6.5$ , than for c-jets and u-jets,  $\langle n^{\rm ch} \rangle \approx 5.0$  (for  $p_{\rm T}^{\rm thr} > 1$  GeV). The jet reconstruction efficiency,  $\varepsilon^{\rm jet}$ , varies between 85% for u-jets and 72% for g-jets, an effect which essentially visualises the fraction of the jet energy collected within a cone of  $\Delta R = 0.4$ .

### TABLE II

Average values of the jet transverse momentum,  $\langle p_{\rm T} \rangle$ , jet reconstruction efficiencies,  $\varepsilon^{\rm jet}$ , and average values of the multiplicity of charged particles in a cone  $\Delta R = 0.4$  around the jet axis,  $\langle n^{\rm ch} \rangle$ , as a function of the transverse momentum threshold applied to the particles: (a)  $p_{\rm T}^{\rm thr} > 0.5$  GeV, (b)  $p_{\rm T}^{\rm thr} > 1$  GeV and (c)  $p_{\rm T}^{\rm thr} > 2$  GeV.

Jet type	$\langle p_{\mathrm{T}}^{\mathrm{jet}} \rangle [GeV]$	$\varepsilon^{ m jet}$ [%]		$\langle n^{\rm ch} \rangle$	
			a	b	с
b-jets $g$ -jets	$54.6\\46.1$	76 72	$\begin{array}{c} 7.5 \\ 7.4 \end{array}$	$\begin{array}{c} 6.5 \\ 6.3 \end{array}$	$\begin{array}{c} 4.9\\ 4.4 \end{array}$
$u ext{-jets}\ c ext{-jets}$	$\begin{array}{c} 51.2 \\ 55.0 \end{array}$	$\frac{85}{80}$	$\begin{array}{c} 5.8 \\ 6.4 \end{array}$	$5.1 \\ 5.7$	$\begin{array}{c} 3.9 \\ 4.4 \end{array}$

#### 3.2. Kinematics of charged particles

The particles inside reconstructed jets have rather low transverse momenta. Table III shows average values of the transverse momentum distributions for electrons and charged hadrons in various types of jets for thresholds of  $p_{\rm T}^{\rm thr} > 1$  GeV and  $dp_{\rm T}^{\rm thr} > 2$  GeV. Only particles inside a cone  $\Delta R = 0.4$  around the jet axis are considered and the numbers are given for the generated particles, *i.e.* neither the efficiency for track reconstruction nor electrons from  $\gamma$ -conversions are included. A cut  $|\eta| < 2.5$  was applied to all jets.

#### TABLE III

	$\langle p_{\mathrm{T}} \rangle ~(p_{\mathrm{T}}^{\mathrm{thr}})$	f > 1  GeV	$\langle p_{\mathrm{T}} \rangle  (p_{\mathrm{T}}^{\mathrm{th}})$	$^{\rm r} > 2 { m ~GeV}$
Jet type	electrons	hadrons	electrons	hadrons
b-jets g-jets u-jets c-jets	9.7 2.6 3.4 7.4	$5.1 \\ 4.6 \\ 6.2 \\ 6.0$	$11.2 \\ 4.0 \\ 5.2 \\ 9.0$	$6.5 \\ 6.0 \\ 7.8 \\ 7.4$

Average values of the transverse momenta,  $\langle p_{\rm T} \rangle$ , for electrons and charged hadrons within the jet cone for charged particle thresholds  $p_{\rm T}^{\rm thr} > 1 {\rm GeV}$  and  $p_{\rm T}^{\rm thr} > 2 {\rm GeV}$ .

The inclusive branching ratio  $BR^{\rm all}(b \to e)$  defined as the fraction of reconstructed *b*-jets with an electron inside a cone  $\Delta R = 0.4$  around the jet axis. The exclusive branching ratio for  $b \to e$  decay,  $BR^{\rm exclu}(b \to e)$ , includes only the direct semileptonic decay of *B* hadrons  $(b \to e)$  and the cascade decays  $(b \to c \to e)$  with semileptonic decays of charm hadrons, normalised to the total number of reconstructed b-labelled jets. In both cases, only *b*-jets with  $p_{\rm T}^{\rm b-jet} > 15 \,{\rm GeV}$  and electrons inside the *b*-jets with  $|\eta^{\rm ele}| < 2.5$  are considered.

Obviously, these inclusive and exclusive branching ratios depend on the cone around the jet axis in which the electrons are collected (two cones  $\Delta R = 0.4$  and  $\Delta R = 0.7$  have been compared) and on the threshold applied to the transverse momentum of the electron ( $p_T^{\rm ele} > 1$  GeV and  $p_T^{\rm ele} > 2$  GeV have been compared), as shown in Table IV. Both the inclusive and exclusive branching ratios are higher when a larger cone is used and the threshold on the transverse momentum is lower.

### TABLE IV

Inclusive  $BR^{\rm all}$  and exclusive  $BR^{\rm exclu}$  branching ratios of  $b \rightarrow e$  (direct and cascade decays) for  $p_{\rm T}^{\rm thr} > 1$  GeV and 2 GeV and for  $\Delta R = 0.4$  and 0.7.

$p_{\mathrm{T}}^{\mathrm{thr}}$	$\frac{BR^{\text{all}}}{\text{for }\Delta R = 0.4}$	$BR^{\rm exclu}(b \to e)$ for $\Delta R = 0.4$	$BR^{\text{ex clu}}(b \to e)$ for $\Delta R = 0.7$
> 1  GeV > 2  GeV	$16.0\%\ 14.5\%$	$15.5\%\ 13.8\%$	$16.4\%\ 14.2\%$

Tables V, VI and VII show respectively the fractions of all electrons<sup>6</sup>, of electrons from B and D hadron decays a nd hadrons inside jets for various  $p_{\rm T}$  bins. Only 11% of electrons from B and D and around 20% of charged hadrons have a transverse momentum in the range 1–2 GeV. These fractions are calculated for two jet cones  $\Delta R = 0.4$  and  $\Delta R = 0.7$  a nd the same  $p_{\rm T}$  bins will be used used for the optimisation of the discriminating function (see Section 5.2.4 and Section 6.1).

### TABLE V

Fraction of all electrons within a cone of given  $\Delta R$  from reconstructed jet axis for various  $p_{\rm T}^{\rm ele}$  bins.

	Fraction of all electrons in specified $p_{\rm T}^{\rm ele}$ bins in a cone $\Delta R = 0.4$ ( $\Delta R = 0.7$ ) [%]							
$p_{\mathrm{T}}^{\mathrm{ele}}$ bin	b-j	ets	g-	jets	<i>c</i> -	jets	u-	jets
1-2 GeV 2-5 GeV 5-8 GeV >8 GeV	$15.6 \\ 29.2 \\ 14.7 \\ 40.5$	(17.6) (30.0) (14.2) (38.2)	$51.7 \\ 39.2 \\ 7.6 \\ 1.5$	$(52.1) \\ (37.4) \\ (7.5) \\ (3.0)$	$20.8 \\ 29.9 \\ 18.8 \\ 30.5$	$\begin{array}{c}(22.0)\\(29.6)\\(18.4)\\(30.0)\end{array}$	$\begin{array}{r} 46.0 \\ 37.1 \\ 10.4 \\ 6.5 \end{array}$	$\begin{array}{c}(47.5)\\(36.3)\\(9.9)\\(6.3)\end{array}$

### TABLE VI

Fraction of signal electrons from  $(b \to e)$  within a cone of given  $\Delta R$  from the reconstructed jet axis for various  $p_T^{\text{ele}}$  bins.

	Fraction of electrons from $b \to e$ in specified $p_{\rm T}^{\rm ele}$ bins in a cone $\Delta R = 0.4$ ( $\Delta R = 0.7$ ) [%]					
$p_{\mathrm{T}}^{\mathrm{ele}}$ bin	$b ext{-jets}$	$g ext{-jets}$	$c ext{-jets}$			
1-2 GeV 2-5 GeV 5-8 GeV >8 GeV	$\begin{array}{rrrr} 11.0 & (13.1) \\ 28.2 & (29.3) \\ 15.2 & (14.8) \\ 45.6 & (42.8) \end{array}$	$\begin{array}{rrrr} 30.8 & (28.1) \\ 38.5 & (37.5) \\ 30.7 & (21.9) \\ 0.0 & (12.5) \end{array}$	$\begin{array}{rrrr} 11.8 & (12.3) \\ 28.6 & (28.7) \\ 21.8 & (21.6) \\ 37.8 & (37.4) \end{array}$			

<sup>&</sup>lt;sup>6</sup> Here all electrons denotes electrons from decays in the hadronic cascade but not from photon conversions in the detector material.

TABLE VII

941

#### Fraction of charged hadrons in specified $p_{\rm T}^{\rm had}$ bins in cone $\Delta R < 0.4$ $(\Delta R < 0.7)$ [%] $p_{\rm T}^{\rm had}$ bin b-jets g-jets c-jets *u*-jets 1-2 GeV26.324.123.128.5(31.0)(30.2)(28.4)(34.4)2-5 GeV40.537.637.341.6(39.2)(36.4)(36.4)(39.7) $5-8 \,\,\mathrm{GeV}$ 16.015.816.715.2(14.5)(14.0)(15.1)(13.3)> 8 GeV17.222.422.914.7(15.3)(12.6)(19.5)(20.1)

Fraction of charged hadrons within a cone of given  $\Delta R$  from the reconstructed jet axis for various  $p_{\rm T}^{\rm had}$  bins.

### 4. Event reconstruction

The basic reconstruction of the fully simulated events is done using standard reconstruction tools, *i.e.* the ATLSIM program with the XKALMAN, XCONVER and EMRECO packages. Jets are reconstructed with the fast algorithm supplied in the ATLFAST package. Results obtained with this fast algorithm for jet reconstruction are in good agreement with the jet reconstruction for the fully simulated events [17]. Events are reconstructed in the whole barrel part of the ATLAS detector, but only the information from tracks, e.m. clusters *etc.* which are closer than  $\Delta R = 0.4$  from the axis of the reconstructed jets is used in the analysis presented here.

### 4.1. Reconstruction in the Inner Detector

For XKALMAN all data cards are set to the default values and the internal bremsstrahlung fit is enabled. It is important because the electrons of interest have rather low momenta, so the inclusion of the corrections for bremsstrahlung effects plays a significant role in the proper calculation of the track parameters.

The fraction of converted photons and electrons from Dalitz decays, both in the signal and in the background samples, is rather high (see Table I). To reject electrons coming from the above sources the XCONVER package is used. It is executed in a reconstruction chain after XKALMAN, and to each track it assigns a value which quantifies the probability that the given track comes from a  $\gamma$ -conversion or a Dalitz decay. The use of XCONVER is described in more detail in Section 5.2.2 and in Appendix B. The track parameters obtained from the Inner Detector the reconstruction are used to initiate reconstruction procedure in the Electromagnetic Calorimeter.

The charged track reconstruction efficiency is about 93% for electrons as well as for hadrons. The average track multiplicity in *b*-jets is found to be ~ 5 for tracks with  $p_{\rm T} \geq 2$  GeV. The efficiency is in agreement with the XKALMAN reconstruction efficiency of 96.6% ±0.3% for electrons<sup>7</sup> and with the overall Inner Detector efficiency [18], since the XKALMAN efficiency was calculated only for "good quality" tracks.

#### 4.2. Reconstruction in the EM Calorimeter

For this step, the information stored in the databanks filled by the EM-RECO package is used. Since the EMRECO package is not optimised to reconstruct clusters in the EM calorimeter for low energy particles, only the information concerning the energy deposition inside individual cells is used by the algorithm presented below.

The track parameters calculated in the Inner Detector allow to localise the point in the vicinity of which the energy depositions in the EM Calorimeter should be searched for. The coordinates of this point are corrected taking into account the displacement of the primary vertex from its nominal position. The cell with the maximum energy deposition near this point is found. The coordinates of this cell define the central point for the calculation of the energy dependent variables. These calculations are performed in a  $3 \times 3$  window around the point defined by the track parameters to avoid misidentification of the energy depositions from various tracks which belong the same jet. This procedure is repeated separately for each longitudinal compartment of the EM Calorimeter. After this step, for each track, three arrays (one per each longitudinal compartment of the EM Calorimeter) are filled with the information necessary for calculating the energy dependent variables used by the soft-electron tag algorithm.

### 5. *b*-jet identification

To perform the identification of b-jets, a set of variables is constructed which constitute the b-jet signature. The variables used by the algorithm for b-jet tagging with electrons explore the characteristic features of soft electron tracks as reconstructed by the ATLAS detector. They were chosen to be efficient to reject other tracks and also electron tracks not originating from b-decays. An additional set of variables characterises more global features of the b-jets themselves and therefore is efficient to improve the rejection

<sup>&</sup>lt;sup>7</sup> Inner Detector TDR, vol. 1 p.152

of jets from non-b cascade decays (g, c, u), even if true electron tracks are present.

The identification variables are constructed by using the information from the Inner Detector and from the EM Calorimeter.

### 5.1. Identification variables

Most of the variables which are defined in this section have been already used in the algorithm described in [10], but at least a few of them have been slightly redefined. The complete set of the variables is presented below.

Inner Detector information only:

- $N_{\rm B}$  number of hits in the pixel detector B-layer on track (could be 0 or 1),
- $N_{\rm PIX}$  number of hits in the pixel detector on track,
- $N_{\rm Si}$  number of precision hits on the track (hits in silicon tracker, both in strip and pixel detectors),
- $N_{\rm RTR}$  number of high-energy hits (TR hits) in the TRT detector on the track,
- $A_0$  impact parameter of the track in the transverse plane,
- $\chi^2_{\rm CONV} \chi^2$  of the conversion fit.

EM Calorimeter information (based on track impact):

- $E_1/E$  fraction of the energy deposited in the first longitudinal compartment of the EM Calorimeter,
- $E_3/E$  fraction of the energy deposited in the third longitudinal compartment of the EM Calorimeter,
- ISO shower isolation in the EM Calorimeter,
- W width of the shower measured in units of distance between the strips,
- AS asymmetry of the shower measured in units of distance between the strips.

Combined information from the Inner Detector and the EM Calorimeter:

- $E_{\rm T}/p_{\rm T}$  ratio of the transverse energy deposited in the EM Calorimeter (for a 3×3 cell window) to the transverse momentum reconstructed in the Inner Detector,
- POS difference between the track and shower positions measured in units of distance between the strips,
- $P_{\rm TJ}$  transverse momentum of the track determined relative to the jet axis.

Variables constructed from the Inner Detector information are provided by the XKALMAN and XCONVER packages as described in Appendix C. For the EM Calorimeter, the exact algorithm for extracting these variables is also described in Appendix C.

## 5.2. Electron identification procedure

To efficiently identify reconstructed tracks as signal electrons and to reject other tracks, the set of variables described above is used. For some of them simple thresholds are applied and the others are used to determine the weight given to every track.

### 5.2.1. Identification of electrons

Most of the tracks reconstructed inside jets are not electrons, therefore the aim of the tagging algorithm is to reject these tracks and to select the electron track candidates with high efficiency. A subset of variables from the set described above was found to be the most efficient for such an algorithm:  $N_{\text{RTR}}$ ,  $E/p_{\text{T}}$ ,  $E_3/E$ ,  $E_1/E$ , ISO, W, AS, POS.

The choice of these variables is dictated by the nature of electromagnetic interactions of particles with matter, as described e.g. in [19]. Two main phenomena are explored: the transition radiation effect, and the shape of the shower energy deposition in the electromagnetic calorimeter.

The transition radiation effect is used first to select electron tracks, which due to their high velocity are emitting X-ray photons while passing through the TRT detector. These photons are detected in the straw tubes as high-amplitude signals called "high energy hits". On average, for 2 GeV electrons with  $|\eta| < 0.8$ , 6.6 high energy hits are expected compared to 1.5 hits expected for pions of the same energy <sup>8</sup>.

The development of the shower in the electromagnetic calorimeter is different for electrons (electromagnetic cascades) and for hadrons (hadronic cascades). For electromagnetic cascades, most of the energy is deposited in

<sup>&</sup>lt;sup>8</sup> Inner Detector TDR, vol. 1, p.118

the first longitudinal sampling and only a small fraction in the third one<sup>9</sup>. The EM shower in the calorimeter is also expected to be isolated, that is the energy leaking outside the  $3 \times 3$  tower cluster is expected to be small<sup>10</sup>.

For each of the variables listed above, probability density function that a given track is produced by an electron from *b*-decay or by any other particle in non-*b* cascades is derived from the analysed samples. Figure 2 shows the normalised distributions of these variables for electron tracks in *b*-jets coming from *B* and *D* decays, and for non-electron tracks (mostly pions) in non-*b* jets. Only these two extreme cases are shown, since the first one corresponds to more than 90% of true electron tracks in *b*-jets and the second one to 95 % of all tracks in *u*-, *c*- or *g*-jets. However most of the variables shown are correlated and the rejection power of the whole set is not the product of the rejection capabilities of the single variables.

#### 5.2.2. Suppression of $\gamma$ -conversions and Dalitz decays

A significant source of low  $p_{\rm T}$  electron and positron tracks in jets are photon conversions and Dalitz decays of neutral pions  $\pi^o \to e^+e^-\gamma$  (see Appendices A.2 and A.1). Such tracks might be identified by the *b*-tagging algorithm as signal electron tracks. To suppress this type of background, the following four discriminating variables are used by the algorithm:  $N_{\rm Si}$ ,  $N_{\rm B}$ ,  $N_{\rm PIX}$ , and  $\chi^2_{\rm conv}$ .

The first three describe simple features of the reconstructed track. The probability density functions for these variables are shown in figure 3 for signal electrons in *b*-jets and for background electrons from photon conversions and Dalitz decays. The signal electrons originate from a region close to the pp interaction vertex, therefore they are expected to pass through all layers of the silicon tracker. However,  $\gamma$ -conversions can occur at any radius. That is why it is required that the signal track candidate have hits in all three pixel layers. This also implies, that there should also be a hit in the first pixel detector layer, the *B*-layer. Finally, the track should have at least 10 hits in the semiconductor tracker. All the above discrete criteria in fact are used as *good-quality* track reconstruction requirements by most of the standard analyses performed in ATLAS [20].

In the case of  $\gamma$ -conversions and Dalitz decays  $e^+e^-$  pairs with small invariant mass are observed in the detector. A common analysis tool, the xCONVER package [21], is used to identify electrons from the above sources. The algorithm searches for pairs of oppositely charged particles and performs the  $\chi^2$  fit assuming a common vertex and a zero opening angle between the tracks (see Appendix B). If, for the electron track candidate, xCONVER finds

<sup>&</sup>lt;sup>9</sup> Calorimeter TDR, p.140

<sup>&</sup>lt;sup>10</sup> TDR Calorimeter p.137



Fig. 2. Probability density functions of the variables used to identify signal electron track for signal electron tracks in b-jets (open histogram) and for non-electron tracks in non-b jets (shaded histogram).

a partner with  $\chi^2_{\rm conv}$  below a given threshold the track is rejected. In figure 3 the  $\chi^2_{\rm conv}$  distributions for signal electrons from *B* and *D* mesons in *b*-jets and background electrons from Dalitz decays and  $\gamma$ -conversions are compared. A strong peak near  $\chi^2_{\rm conv} \approx 0$  can be seen for the background distribution and indicates that  $e^+e^-$  pairs with small invariant mass exist in this sample and are correctly reconstructed. For the signal electron tracks, since no matching with any oppositely charged track is expected this distribution is flat. All tracks with  $\chi^2_{\rm conv} < 200$  are rejected by the tagging algorithm.



Fig. 3. probability density functions for the variables used to reject electrons from Dalitz decays and from  $\gamma$ -conversions, for signal electrons in *b*-jets (open histogram) and for background electrons from Dalitz decays and  $\gamma$ -conversions in non-*b* jets (shaded histogram).

As discussed in [21], the above method allows to reconstruct 75% of photons having a transverse momentum around 50 GeV and converting at a radius below 10 cm. However, photons and neutral pions in b-jets have typically much lower momenta, on average 2.6 GeV for photons and 3.8 GeV for  $\pi^{o}$ 's (for a threshold  $p_{\rm T} > 0.5$  GeV). As described in Appendices A.2 and A.1, in general, the photon momentum is not equally divided between the electron and the positron. and one of them takes most of the available momentum. The lower energetic partner may have a transverse momentum below 0.5 GeV and therefore will not be reconstructed by the XKALMAN track reconstruction package. Due to the high track density in the jet also hits wrongly associated to the tracks decrease the efficiency of the conversion finding algorithm. These two effects lead to a lower efficiency for identifying  $\gamma$ -conversions and Dalitz decays by the XCONVER package than that expected from the sample of photons with  $p_{\rm T} = 50$  GeV.

For the analysed sample of electrons from photon conversions and Dalitz decays inside *b*-jets, the identification efficiency for the  $e^+(e^-)$  using the  $\chi^2$ 

variable is around 50% if no other selection criterion except requiring hits in all three pixel layers is applied. The obtained efficiency is shown in figure 4 as a function of the transverse momentum of the electron (positron) track candidate. The efficiency is flat over the relevant transverse momentum spectrum.



Fig. 4. Efficiency for identifying  $e^+(e^-)$  originating from photon conversions or Dalitz decays as a function of the  $e^+(e^-)$  transverse momentum. The positron (electron) candidate is requested to have three hits in the pixel layers. The results shown for tracks reconstructed in *b*-jets.

It should be mentioned that  $\chi^2_{\rm conv}$  is to some extent correlated with other variables used in the analysis. Therefore, a 50% reduction of background electrons originating from photon conversions and Dalitz decays, obtained without applying any initial selection cuts, does not necessarily imply an additional 50% reduction in the sample remaining after applying the selection criteria described in Section 5.2.1.

#### 5.2.3. Charged tracks inside b-jets

In the two previous Sections, variables efficient for rejecting non-electron tracks or electron tracks from  $\gamma$ -conversions and Dalitz decays have been defined. In this Section two additional variables which characterise some of the global features of any track coming from the *b*-jet are discussed. Since a *b*-jet is in fact reconstructed from the hadronic cascade initiated by the heavy (~5GeV) and long-lifetime *B*-hadron, one expects that many of its charged tracks will have significant impact parameter  $A_0$  (long lifetime) and also large  $p_{\rm T}$  in respect to the jet axis  $P_{\rm TJ}$  (decay of a heavy object).

The  $A_0$  and  $P_{\text{TJ}}$  distributions are shown in figure 5 for signal electron tracks from B and D mesons in b-jets, and for other charged tracks in non-bjets. Similarly to the variables described in Section 5.2.1, for each value of  $A_0$  and  $P_{\text{TJ}}$ , probability density functions that the given track is produced by a signal electron in b-jet, or by a hadron in a non-b jet, are derived from the analysed sample.



Fig. 5. Probability density functions for the variables  $A_0$  and  $P_{TJ}$  used to identify tracks in *b*-jets for signal electron tracks in *b*-jets (opened histogram) and for all tracks in non-*b* jets (shaded histogram).

The rejection power of the  $A_0$  variable is much higher than that of  $P_{\rm TJ}$ (notice the logarithmic scale for  $A_0$ ). The average value of  $A_0$  for tracks inside *b*-jet is expected to be ~ 468  $\mu m$  [22], so the ATLAS transverse impact parameter resolution of 12  $\mu m$  for  $p_{\rm T} = 200$  GeV and  $60\mu m$  for  $p_{\rm T} = 1$  GeV<sup>11</sup> obviously provides a large rejection of tracks not originating from *b*-jets.

One should be aware however that the  $A_0$  variable is strongly correlated with other ones used by the vertexing algorithm [8].

#### 5.2.4. Discriminating function

The discriminating function  $D_{\text{track}}$  is defined for each track as:

$$D_{\text{track}} = \log \frac{\prod_i P_e(x_i)}{\prod_i P_h(x_i)}$$

where  $x_i$  denotes the value of the *i*-th variable for the given track,  $P_e(x_i)$  is the probability that the track originates from a signal electron,  $P_h(x_i)$  is the probability that the track originates from non-signal hadron and *i* runs through all the variables used by the algorithm as described in Sections

<sup>&</sup>lt;sup>11</sup> Inner Detector TDR, vol. 1, p.112

5.2.1 and 5.2.3. For continuous variables  $(E_1/E, E_3/E, \text{ISO}, W, \text{AS}, \text{POS}, E_{\text{T}}/p_{\text{T}}, P_{\text{TJ}} \text{ and } A_0)$ , the probability density functions are smoothed using PAW smoothing procedure [23]. The only exception is  $N_{\text{RTR}}$  for which a look-up table is prepared instead. For some of the variables  $(N_B, N_{\text{PIX}}, N_{\text{Si}} \text{ and } \chi^2_{\text{CONV}})$ , thresholds are defined and if they are not satisfied, a discrete value of the discriminating function  $D_{\text{track}}, D_{\text{track}} = -40$ , is assigned to the track. If the value of  $D_{\text{track}}$  is not well defined<sup>12</sup> then it is set to -40 too. Track with a low value of  $D_{\text{track}}$  is always rejected by the cut on the  $D_{\text{track}}$  value.

A higher rejection power can be achieved by the algorithm if, for tracks with transverse momentum below 5 GeV, the value of  $D_{\text{track}}$  is rescaled as follows:

$$D_{\rm track} \rightarrow D_{\rm track} + 1.5 \times p_{\rm T} - 7.5$$
,

where  $p_{\rm T}$  is the transverse momentum of the reconstructed track given in GeV. Even if there is no physical explanation for this rescaling, it is effective in rejecting the low- $p_{\rm T}$  background electron tracks from Dalitz decays and  $\gamma$ -conversions.

The probability distributions of some of the identifying variables (for example  $N_{\rm RTR}$ ) show a significant dependence on the transverse momentum of the track. The algorithm used by the *b*-tagging procedure therefore calculates the probability density distributions and the look-up-tables for three  $p_{\rm T}$  bins, a procedure which significantly improves the discrimination power for low- $p_{\rm T}$  tracks.

#### 5.2.5. Identification efficiency for electron tracks

Figure 6 shows the distribution of the discriminating function  $D_{\text{track}}$  for various types of tracks. By applying a cut to the value of  $D_{\text{track}}$ , a good separation between signal electron tracks from *b*-jets and other tracks can be achieved.

Figure 6(a) shows the  $D_{\text{track}}$  distribution for the signal electron tracks in *b*-jets. The distribution is peaked around  $D_{\text{track}} = 10$  and extends below -10 only for 8.6% of the analysed tracks. The singular bin at  $D_{\text{track}} = -40$ corresponds to signal electron tracks for which one of the discrete variables did not satisfy the threshold. Figure 6b shows the same distribution (open histogram) overlayed with the distribution for all tracks in non-*b* jets (shaded histogram). The distribution of the background tracks is shifted towards lower values with a peak position around  $D_{\text{track}} = -10$ . Again, the singular bin at  $D_{\text{track}} = -40$  corresponds to tracks for which at least one of the discrete variables did not satisfy the threshold. The separate cases of

<sup>&</sup>lt;sup>12</sup> The logarithm of an argument close to zero has a singular behaviour.



Fig. 6. Probability density distributions of the discriminating function  $D_{\text{track}}$  for: (a) signal electron tracks in *b*-jets from *B* and *D* hadrons, (b) signal electron tracks in *b*-jets from *B* and *D* hadrons (open histogram) and all tracks in non-*b* jets (shaded histogram), (c) signal electron tracks in *b*-jets from *B* and *D* hadrons (open histogram) and background electron tracks from  $\gamma$ -conversions or Dalitz decays in non-*b* jets (shaded histogram), (d) signal electron tracks in *b*-jets from *B* and *D* hadrons (open histogram) and pion tracks from non-*b* jets (shaded histogram). The distributions are plotted only for tracks with a reconstructed transverse momentum  $p_{\rm T} \geq 2$  GeV.

background electron tracks and charged pion tracks from the non-b jets are shown respectively in figures 6(c) and 6(d).

The identification of a candidate track as originating from a signal electron track is based on the value of  $D_{\text{track}}$  assigned to the given track. Those tracks for which  $D_{\text{track}}$  is below a given threshold (typically  $D_{\text{track}}^{\text{thr}} = 5$  to 7) are rejected. The identification efficiency versus the rejection power of the algorithm is obtained by varying the value of  $D_{\text{track}}^{\text{thr}}$ .

To quantify the performance of the identification procedure the following efficiency and rejection factors are defined:

• electron identification efficiency:

$$\varepsilon_e = \frac{N_e^t}{N_e} \,,$$

where  $N_e^t$  is the number of reconstructed signal electron tracks in the  $H \rightarrow b\bar{b}$  sample (*i.e.* electrons from B or D hadrons) identified as a

signal electron track, *i.e.* for which  $D_{\text{track}} > D_{\text{track}}^{\text{thr}}$ , and  $N_e$  is the number of reconstructed signal electron tracks from that sample,

• charged pion rejection:

$$R_{\pi} = \frac{N_{\pi}}{N_{\pi}^t} \,,$$

where  $N_{\pi}$  is the number of reconstructed charged pion tracks, and  $N_{\pi}^t$  is the number of reconstructed pion tracks misidentified as a signal electron track, *i.e.* for which  $D_{\text{track}} > D_{\text{track}}^{\text{thr}}$ ,

• conversion rejection:

$$R_{\rm conv} = \frac{N_{\rm conv}}{N_{\rm conv}^t} \,,$$

where  $N_{\rm conv}$  is the number of reconstructed tracks coming from  $\gamma$ -conversions and Dalitz decays,  $N_{\rm conv}^{\rm t}$  is the number of reconstructed tracks coming from  $\gamma$ -conversions and Dalitz decays misidentified as a signal electron track, *i.e.* for which  $D_{\rm track} > D_{\rm track}^{\rm thr}$ .

The use of the impact parameter  $A_o$  in the discriminating function is a potential source of correlation of the soft-electron and vertexing *b*-tagging methods. Figure 7 shows the expected rejections  $R_{\pi}$  and  $R_{\text{conv}}$  as a function of the signal electron identification efficiency without (top) and with (bottom) the  $A_0$  variable included in the discriminating function. As expected, the worst rejection factors are obtained for tracks inside *b*-jets, both for pions and also for electrons from  $\gamma$ -conversions and Dalitz decays. While the rejection factors  $R_{\pi}$  and  $R_{\text{conv}}$  as a function of the identification efficiency remain almost the same for *b*-jets when including the  $A_o$  variable in the algorithm they increase significantly in the case of the background jets. This can be explained by the fact that the  $A_0$  variable has no discriminating power in the case of *b*-jets where tracks originate in fact from *B* or *D* hadron decays.

#### 5.3. Jet tagging procedure

The jet tagging procedure is based on the electron track identification procedure described above. The algorithm is constructed as follows:

- for each track in the jet, the value of the discriminating function  $D_{\text{track}}$  is calculated;
- for each jet, the track with the highest value of  $D_{\text{track}}$  is chosen and its value of the discriminating function is assigned as the value of the discriminating function for the jet,  $D_{\text{jet}} = \max(D_{\text{track}})$ ;
- for a given threshold  $D_{\text{jet}}^{\text{thr}}$  a jet with  $D_{\text{jet}} \ge D_{\text{jet}}^{\text{thr}}$  is tagged as a *b*-jet.



Fig. 7. Rejection factors  $R_{\pi}$  (left) and  $R_{\text{conv}}$  (right) for tracks in various jet types as a function of the efficiency for identifying signal electron tracks from B and Dhadrons in *b*-jets. Only tracks with a reconstructed transverse momentum  $p_{\text{T}} \geq$ 2 GeV are considered. The labels "e in *b*-jets", "e in *g*-jets" *etc.* denote electron tracks from  $\gamma$ -conversions or Dalitz decays in *b*-, *g*-, *u*- or *c*-jets. The two upper plots show the rejections without using the impact parameter  $A_0$  and the bottom ones after including  $A_0$  in the set of selection variables. The error bars denote only the statistical uncertainties.

Figure 8 shows the distributions of  $D_{\rm jet}$  for signal jets, *i.e. b*-jets, and for background jets: *g*-, *u*- and *c*- jets. The  $D_{\rm jet}$  distribution for *b*-jets is shifted with respect to the distributions for *g*-, *u*-, and *c*-jets, peaking at values  $D_{\rm jet}^{\rm peak} \approx 10$  (for *b*-jets) and  $D_{\rm jet}^{\rm peak} \approx -10$  (for other jets). A fraction of jets for each jet-type have a discrete value,  $D_{\rm jet} = -40$ , which corresponds to jets for which none of the tracks passed all the discrete selection criteria.

The flexibility of the procedure allows to vary continuously the required b-tagging efficiency as a function of the rejection of other jets by changing the value  $D_{jet}^{thr}$ . In addition, it is possible to combine the discriminating function  $D_{jet}$  with the corresponding discriminating function constructed for the vertexing b-tagging algorithm. This would allow to obtain a combined discriminating function for the overall b-tagging algorithm, including vertexing and soft lepton information.



Fig. 8. Probability density distributions of the discriminating function  $D_{\text{jet}}$  for various types of jets: (a) *b*-jets, (b) *b*-jets (open histogram) and gluon jets (shaded), (c) *b*-jets (open histogram) and *u*-jets (shaded histogram), (d) *b*-jets (white) and *c*-jets (grey). Only tracks with a reconstructed transverse momentum  $p_{\text{T}} \geq 2$  GeV are used for the *b*-tagging algorithm.

### 6. Results

In this Section, the rejection power against non-*b* jets obtained by the algorithm described above is discussed as a function of the *b*-tagging efficiency. The reference algorithm is compared to modified algorithms, without the  $\gamma$ conversion finding, without the use of the  $A_o$  and  $P_{\rm TJ}$  variables, and with a lower threshold applied to the electron  $p_{\rm T}$ .

The following notation is used:

• efficiency of *b*-tagging algorithm:

$$\varepsilon_b^{\mathrm{alg}} = \frac{N_b^{\mathrm{t}}}{N_b}$$

where  $N_b^{\rm t}$  is the number of reconstructed *b*-jets in the  $H \to b\bar{b}$  sample, which has been filtered to contain at least one electron per *b*-jet as described in Section 2.1, tagged as *b*-jets, and  $N_b$  is the number of reconstructed *b*-jets in the sample, • soft electron *b*-tagging efficiency:

$$\varepsilon_b^{\text{soft}} = \frac{N_b^{\text{soft}}}{N_b},$$

where  $N_b^{\text{soft}}$  is the number of reconstructed *b*-jets tagged by the softelectron *b*-tagging procedure, and  $N_b$  is the number of reconstructed *b*-jets in a sample of  $H \to b\bar{b}$  decays not filtered to contain electrons,

• jet rejection factor:

$$R_{\rm jet} = \frac{N_j}{N_j^{\rm t}},$$

where j is a given jet type (g,u,c),  $N_j$  is the number of reconstructed jets of j type in the  $H \to j\bar{j}$  sample, and  $N_j^t$  is the number of reconstructed j type jets in this sample tagged as b-jets.

The soft-electron *b*-tagging efficiency,  $\varepsilon_b^{\text{soft}}$ , is obtained by multiplying the efficiency of the *b*-tagging algorithm  $\varepsilon_b^{\text{alg}}$  by the inclusive branching ratio  $BR^{\text{all}}$  for all electrons inside *b*-labelled jets, calculated for the chosen threshold of the electron  $p_{\text{T}}$  (*e.g.*  $BR^{\text{all}} = 14.5\%$  for  $p_{\text{T}}^{\text{thr}} > 2$  GeV, as shown in Table IV).

#### 6.1. Standard algorithm

The results for the standard algorithm are obtained with the following assumptions:

- the detector is assumed to work with the nominal performance (in particular for all layers of the pixel detector);
- the procedure for identifying γ-conversions and Dalitz decays is applied with the following thresholds on the identifying variables required for the signal electron tracks:
  - number of hits in the silicon tracker  $N_{\rm Si} \ge 9$ ;
  - number of hits in the pixel detector  $N_{\text{PIX}} \ge 2$ ;
  - a hit in the pixel detector B-layer is required, hence  $N_{\rm B} = 1$ ;
  - $-\chi^2$  of the conversion fit  $\chi^2_{\rm conv} = 0$  or  $\chi^2_{\rm conv} \ge 200;$
- the variables identifying *b*-jet tracks,  $A_0$  and  $P_{TJ}$ , are used;
- only tracks with transverse momentum  $p_{\rm T} \ge 2$  GeV are considered;

- the probability density distribution functions and look-up-tables are calculated separately for three  $p_{\rm T}$  bins:  $2 \leq p_{\rm T} \leq 5$ ,  $5 < p_{\rm T} \leq 8$ , and  $p_{\rm T} > 8$  GeV;
- all reconstructed charged tracks are considered by the tagging algorithm, so any type of particle (e.g.  $\pi^{\pm}$ ,  $\gamma \rightarrow e^+e^-$ ) can tag a jet as a *b*-jet.

Table VIII summarises the variables used by the algorithm and specifies whether their contribution to the discriminating function D is calculated from a probability density function or is defined by a discrete threshold<sup>13</sup>. Figure 9 shows the jet rejection factors for various types of jet as a function of the efficiency of the *b*-tagging algorithm. As expected, the highest rejection is obtained for *u*-jets, an intermediate one for gluon jets, and the lowest one for *c*-jets.

### TABLE VIII

Variables used by the standard algorithm. In the case of variables not satisfying a certain discrete threshold the value of the discriminating function D is set to D = -40.

Variable	Discriminating feature			
	Inner Detector			
$N_{\rm B}$	threshold $N_{\rm B} = 1$			
$N_{\rm PIX}$	threshold $N_{\rm PIX} \ge 2$			
$N_{\rm Si}$	threshold $N_{\rm Si} \ge 9$			
$N_{\rm RTR}$	continuous probability density			
$A_0$	continuous probability density			
$\chi^2_{ m conv}$	$\chi^2_{\rm conv} = 0$ (no conversion partner found)			
	or threshold $\chi^2_{\rm conv} \ge 200$			
EM Calorimeter				
$E_1/E$	continuous probability density			
$E_3/E$	continuous probability density			
ISÓ	continuous probability density			
W	continuous probability density			
$\mathbf{AS}$	continuous probability density			
$\mathbf{POS}$	continuous probability density			
	Combined ID+EM Cal.			
$E/p_{\rm T}$	continuous probability density			
$P_{\mathrm{TJ}}^{\prime 11}$	continuous probability density			

<sup>&</sup>lt;sup>13</sup> As can be seen from Table VIII the cuts used for the precision tracker variables are the same as used by the b-tagging vertexing algorithm.



Fig. 9. Jet rejection factor  $R_{jet}$  as a function of the efficiency of the *b*-tagging algorithm,  $\varepsilon_{b}^{alg}$ , for various jet types.

To compare these results with the ones which will be given for the modified algorithms, a nominal working point corresponding to a 50% efficiency of the *b*-tagging algorithm was chosen. For  $\varepsilon_b^{\text{alg}} = 50 \pm 1\%$ , the nominal rejection factors against non *b*-jets are:  $R_g = 192 \pm 27$ ,  $R_u = 554 \pm 113$  and  $R_c = 47 \pm 4$ . The quoted errors are purely statistical. These results are in general agreement with the conclusions given in section 2.2. The somewhat higher rejection obtained here, as compared to the limit given there, can be explained by the additional rejection power of the discrimination criteria which explore the properties of the *b*-jets themselves ( $A_0$  and  $P_{\text{TJ}}$ ).

Table IX shows the fraction of jets tagged by a given type of track for  $\varepsilon_b^{\text{alg}} = 50 \%$ . Most of the jets are tagged by true electron tracks, independently of the jet type. It indicates that the electron identification procedure works with a high efficiency and good purity. Most of the electron tracks which tag the jets, except for *u*-jets, are signal electron tracks (from *B*- or *D*-hadron decays). For the *b*-jet sample, these tracks tag almost 96% of all tagged jets. If one takes into account that, in the original signal sample, the percentage of jets containing such electrons is ~ 90% (see Table I), the conclusion is that, after the tagging procedure, the *b*-tagged sample is enriched with this type of electrons. This effect is even stronger for the *g*- and *c*-jets, where the electrons from *B* and *D* hadrons constitute 4% (*g*-jets) and 35% (*c*-jets) of the jets containing an electron, but where these fractions increase respectively to 52% and ~ 90% for the *b*-tagged jets of these types. The sample of *u*-jets does not contain electrons from *B* and *D*, but these jets

are also sometimes tagged, mainly (~ 83%) by true electron tracks from  $\gamma$ -conversions and Dalitz decays or from other sources. An improvement of the rejection power against g- and u-jets is still possible. It would require further improvement of the  $\gamma$ -conversion and Dalitz decay finding procedure and also an improvement of the pion track rejection. For the c-jets, a better rejection would be hard to achieve.

### TABLE IX

Fractions of jets (in %) tagged by a given type of track for  $\varepsilon_b^{\text{alg}} = 50\%$ . Only tracks with  $p_{\text{T}} \geq 2$  GeV are used. The label "*e from*  $\gamma$ " denotes electron tracks from  $\gamma$ -conversions and Dalitz decays.

		Frac	tion of jets	tagged by a	ı given typ	pe of tr	ack [%]
Jet type	all $e$	$e \text{ from } \mathbf{B}$	$e \text{ from } \mathbf{D}$	$e \text{ from } \gamma$	other $e$	π	others
b	99.8	64.0	31.8	0.07	3.93	0.14	0.06
g	96.0	14.0	38.0	28.0	16.0	4.0	0.0
u	83.3	0	0	33.3	50.0	13.7	0.0
c	95.6	0	90.1	1.2	4.3	3.7	0.7

Table X gives the efficiencies of the *b*-tagging algorithm  $\varepsilon_b^{\text{alg}}$  and  $\varepsilon_b^{\text{soft}}$  versus the jet rejections  $R_g$ ,  $R_u$ ,  $R_c$  for a few values of  $D_{\text{jet}}^{\text{thr}}$  and for the tagging algorithm applied to tracks with  $p_{\text{T}} \geq 2$  GeV. The inclusive branching ratio value  $BR^{\text{all}} = 14.5\%$  from Table IV is used to recalculate  $\varepsilon_b^{\text{soft}}$  from  $\varepsilon_b^{\text{alg14}}$ .

### 6.2. Algorithm without $\gamma$ -conversions and Dalitz decays finding

As already explained above electrons from  $\gamma$ -conversions and Dalitz decays can degrade the rejection power of the procedure for soft-electron tagging. As shown in Table I, for g-, u- and c-jets such electron tracks are present in about 8% of all jets, if only tracks with  $p_{\rm T} \geq 2$  GeV are taken into account. To evaluate the improvement of the rejection power due to the  $\gamma$ -conversion and Dalitz decay finding procedure, the results with and without this procedure are compared.

Figure 10 shows a comparison of the jet rejection factors obtained for the standard algorithm and for the same algorithm without using the  $\gamma$ -conversion and Dalitz decay finding procedure variables defined in Section 5.2.2. The rejection factors are plotted as a function of the efficiency of

<sup>&</sup>lt;sup>14</sup> This inclusive branching ratio does not include electrons from  $\gamma$ -conversions and Dalitz decays as it is calculated at particle level. However, since such electrons tag less than 0.1% of all *b*-tagged jets from  $H \to b\bar{b}$  events (see Table IX), it is nevertheless correct to use  $BR^{\rm all}$  to recalculate  $\varepsilon_b^{\rm soft}$  from  $\varepsilon_b^{\rm alg}$ .

TABLE X

Efficiencies of the *b*-tagging algorithm,  $\varepsilon_b^{\text{alg}}$ , of the overall soft-electron tagging,  $\varepsilon_b^{\text{soft}}$ , and jet rejection factors  $R_g$ ,  $R_u$ ,  $R_c$  for a few values of  $D_{\text{jet}}^{\text{thr}}$  (only tracks with  $p_{\text{T}} \geq 2$  GeV are used to tag jets and  $\varepsilon_b^{\text{soft}} = \varepsilon_b^{\text{alg}} \cdot BR^{\text{all}}$ ).

$\begin{array}{c} \varepsilon^{\mathrm{alg}}_{b}  [\%] \\ \varepsilon^{\mathrm{soft}}_{b}  [\%] \end{array}$	$35 \pm 1 \\ 5.1 \pm 0.1$	$50 \pm 1$ $7.2 \pm 0.1$	$65 \pm 1 \\ 9.1 \pm 0.1$	$80 \pm 0.7 \\ 11.6 \pm 0.1$
$\begin{array}{c} R_g \\ R_u \\ R_c \end{array}$	$456 \pm 100$ $1478 \pm 492$ $87 \pm 9$	$192 \pm 27 \\ 554 \pm 113 \\ 47 \pm 4$	$110 \pm 12 \\ 205 \pm 25 \\ 26 \pm 1$	$19 \pm 1$ $29 \pm 1$ $9.1 \pm 0.3$
$D_{ m jet}^{ m thr}$	8.65	6.88	4.88	1.1

the *b*-tagging algorithm  $\varepsilon_b^{\text{alg}}$ . For all types of jets, the  $(R_g, R_u, R_c)$  curves show similar behaviour. For a very high efficiency of the *b*-tagging algorithm ( $\varepsilon_b^{\text{alg}} > 75\%$ ), the rejections are slightly higher without applying the conversion finding procedure, while they are lower for lower efficiency.



Fig. 10. Rejection factors against *g*-jets  $(R_g)$ , *u*-jets  $(R_u)$  and *c*-jets  $(R_c)$  as a function of the efficiency of the *b*-tagging algorithm,  $\varepsilon_b^{\text{alg}}$ . The results are shown for the standard algorithm described in the previous Section *(stand)* and for the same algorithm without the use of the  $\gamma$ -conversion and Dalitz decay finding procedure *(no-conv)*.

This behaviour can be explained by the fact that, when using the standard algorithm, additional few percent of the *b*-jets are tagged by tracks otherwise identified as background electron tracks from  $\gamma$ -conversions or Dalitz decays, and hence rejected by the  $D_{\text{jet}}^{\text{thr}}$  requirement (see figure 8). These jets are therefore *b*-tagged if the conversion finding procedure is not applied. In that case, other tracks (*e.g.* pion tracks) in the background jets still have a low value of the  $D_{\text{track}}$  discrimination function, and are therefore rejected.

A most significant improvement when using the standard algorithm, which includes the conversion finding, is observed for all three types of background jets, but with a different magnitude for each jet type. The biggest improvement is observed for the *u*-jets, an intermediate one for the *g*- jets, and only a small improvement can be seen in the case of *c*-jets. These results can be explained by two effects: first of all, the fraction of jets with electrons from  $\gamma$ -conversions and Dalitz decays with respect to jets with any electron is the highest for the *u*-jets, secondly the fraction of jets with a signal electron is the highest one for the *c*-jets (see Table IX).

Since the aim of the algorithm is to give the highest possible rejection power, the standard algorithm is more efficient in this respect. For  $\varepsilon_b^{\text{alg}} =$ 50% the procedure without  $\gamma$ -conversion and Dalitz decay finding yields the rejection factors:  $R_g = 83 \pm 8$ ,  $R_c = 38 \pm 3$  and  $R_u = 96 \pm 8$ , while the standard procedure (nominal results) gives  $R_g = 192 \pm 27$ ,  $R_c = 47 \pm 4$  and  $R_u = 554 \pm 113$ . In both cases, the errors are purely statistical.

### 6.3. Algorithm without the $A_0$ and $P_{TJ}$ variables

The standard algorithm does not use only variables which identify signal electron tracks, but also explores features which characterise most of the tracks inside b-jets. These are quantified by the two identifying variables:  $A_0$ , the transverse impact parameter of the track, and  $P_{\rm TJ}$ , the transverse momentum of the track determined with respect to the jet axis. Figure 11 shows the jet rejection factors for the standard algorithm, for the algorithm without the use of the  $A_0$  variable and for the algorithm without use of the  $A_0$  nor  $P_{\rm TJ}$  variables. The rejection factors are plotted as a function of the efficiency of the b-tagging algorithm. For all jet types, the results obtained with the standard algorithm are better (higher rejection), and the biggest improvement, as expected, is observed for the u-jets, the smallest for the c-jets. Comparing the results for  $\varepsilon_b^{\text{alg}} = 50\%$ , one finds that the described algorithm (without the use of the  $A_0$  nor  $P_{\rm TJ}$  variables) gives  $R_q = 126 \pm 14, R_c = 33 \pm 2$  and  $R_u = 229 \pm 30$  while the nominal results are  $R_g = 192 \pm 27$ ,  $R_c = 47 \pm 4$  and  $R_u = 554 \pm 113$ . In both cases the errors are purely statistical. At 50% efficiency, the use of the  $P_{\rm TJ}$  variable (but not  $A_0$ ) improves the rejection only up to  $R_q = 145 \pm 18$ ,  $R_c = 37 \pm 2$ and  $R_u = 283 \pm 41$  .



Fig. 11. Rejection factors against *g*-jets  $(R_g)$ , *u*-jets  $(R_u)$ , and *c*-jets  $(R_c)$  as a function of the efficiency of the *b*-tagging algorithm,  $\varepsilon_b^{\text{alg}}$ . The results are shown for the standard algorithm (*stand*), for the algorithm without the use of the  $A_0$  variable  $(no-A_0)$ , and for the algorithm without the use of the  $A_0$  nor  $P_{\text{TJ}}$  variables  $(no-A_0 \text{ and } no-P_{\text{TJ}})$ .

The systematic improvement observed for the rejection factors  $R_g$ ,  $R_u$ ,  $R_c$  by including the  $A_0$  and  $P_{\rm TJ}$  variables in the algorithm can be easily explained. A significant fraction of both g-jets (52.0%) and c-jets (90.1%) is tagged by the signal electron tracks from B- or D-hadron decays. Only the remaining fraction can be rejected using  $A_0$  and  $P_{\rm TJ}$ . For the u-jets, however, which are mostly tagged by background electron tracks or by other tracks, one does not expect these tracks to have significant impact parameter or transverse intrinsic momentum. The  $A_0$  and  $P_{\rm TJ}$  variables improve the rejection significantly in this case.

The impact parameter  $A_0$  is used as well by the vertexing *b*-tagging algorithm [8]. This fact introduces a potential strong correlation between the two methods unless  $A_0$  is dropped from the soft-electron tagging algorithm. A careful evaluation of this correlation is necessary before combining results from both methods to obtain the overall *b*-tagging performance.

# 6.4. The impact of $p_{\rm T}^{\rm thr}$

As already discussed in Section 3.2, the inclusive branching ratio  $BR^{\rm all}$ of the process  $b \to e$  depends on the threshold on the transverse momentum of the tagging track,  $p_{\rm T}^{\rm thr}$ . The value of  $BR^{\rm all}$  is higher when this threshold is lower and therefore for the same value of the efficiency of the b-tagging algorithm  $\varepsilon_b^{\rm alg}$ , the soft-electron tagging efficiency,  $\varepsilon_b^{\rm soft}$ , is higher. However, with decreasing  $p_{\rm T}^{\rm thr}$ , the jet rejection decreases for a fixed

However, with decreasing  $p_{\rm T}^{\rm thr}$ , the jet rejection decreases for a fixed value of  $\varepsilon_b^{\rm alg}$ . The first reason for this effect is that the average charged track multiplicity rises as  $p_{\rm T}^{\rm thr}$  decreases (see Table II). Assuming that the rejection power,  $R_p$ , against the background tracks is constant as a function of  $p_{\rm T}$ , the jet rejection factor  $R_{\rm jet}$  can be calculated as:

$$R_{\rm jet}^{-1} = 1 - (1 - R_p^{-1})^m$$
,

where m is the average charged track multiplicity of the jet. Therefore, to keep the value of  $R_{jet}$  at the same level as in the standard algorithm for larger values of m,  $R_p$  should increase as well.

The second reason is that not only the charged track multiplicity is higher for a lower value of  $p_{\rm T}^{\rm thr}$ , but also the fraction of background jets with true electron tracks inside the jets is significantly higher. These two observations indicate that for the same value of the efficiency of the *b*-tagging algorithm,  $\varepsilon_b^{\rm alg}$ , the rejection capability would probably decrease if one decreases  $p_{\rm T}^{\rm thr}$ from 2 GeV to 1 GeV.

Table XI shows the results obtained with the standard algorithm for two values of  $p_{\rm T}^{\rm thr}$ : 1 GeV and 2 GeV. Taking into account that the corresponding values of  $BR^{\rm all}(b \to e)$  are respectively 16% and 14.5%, different values of  $\varepsilon_b^{\rm alg}$  correspond in the end to the same values of the overall efficiency of the soft-electron tagging,  $\varepsilon_b^{\rm soft}$ . A higher jet rejection appears to be achievable for  $p_{\rm T}^{\rm thr} = 2$  GeV than for  $p_{\rm T}^{\rm thr} = 1$  GeV<sup>15</sup>. The largest improvement in the rejection power with rising the  $p_{\rm T}^{\rm thr}$  is observed for *u*-jets, and it is due to the fact that, with lower  $p_{\rm T}^{\rm thr}$ , the increase in the number of background electrons from  $\gamma$ -conversions and from Dalitz decays is the highest for these jets.

In the  $p_{\rm T}$  bin 1 – 2 GeV for any threshold applied on the discriminating function the *b*-tagging algorithm performance is worse than for higher  $p_{\rm T}$  values. Therefore the overall background rejection for  $\varepsilon_b^{\rm alg} \approx 50\%$  becomes worse after lowering  $p_{\rm T}^{\rm thr}$  down to 1 GeV.

<sup>&</sup>lt;sup>15</sup> The rejections are in fact comparable within statistical errors, but, as the sample used for the higher momentum threshold is a subsample of the  $p_{\rm T}^{\rm thr} = 1$  GeV sample, values obtained for both  $p_{\rm T}^{\rm thr}$  are strongly correlated and the central values should be compared.

### TABLE XI

Efficiencies of the *b*-tagging algorithm,  $\varepsilon_b^{\text{alg}}$ , and of the soft-electron tagging efficiency,  $\varepsilon_b^{\text{soft}}$ , and jet rejection factors  $R_g, R_u, R_c$  obtained for two values of the threshold on the transverse momentum of the tagging track for the same value of the soft electron tagging efficiency,  $\varepsilon_b^{\text{soft}} = 7.2\%$ . The values of  $D_{\text{jet}}^{\text{thr}}$  used to obtain these results are also given.

$p_{\mathrm{T}}^{\mathrm{thr}}$	$arepsilon_b^{\mathrm{alg}}~[\%]$	$\varepsilon_b^{\mathrm{soft}}$ [%]	$R_{g}$	$R_u$	$R_c$	$D_{ m jet}^{ m thr}$
1 GeV 2 GeV	$45.3 \pm 0.8 \\ 50 \pm 1$	$7.2 \pm 0.1$ $7.2 \pm 0.1$	$\begin{array}{c} 171 \pm 23 \\ 192 \pm 27 \end{array}$	$403 \pm 70 \\ 554 \pm 113$	$\begin{array}{c} 43\pm 3\\ 47\pm 4\end{array}$	$\begin{array}{c} 6.92 \\ 6.88 \end{array}$

### 7. Conclusions

In the studies presented in this note the soft-electron b-tagging procedure is discussed. Table XII summarises results obtained for the jet rejection for variations around the standard algorithm and for the same value of the efficiency of the soft-electron b-tagging. For a nominal efficiency of the soft electron b-tagging of ~ 7.2% the rejection against gluon jets is ~ 200, the rejection against c-jets is ~ 45 and the rejection against u-jets ~ 600. This standard algorithm can be combined with the vertexing b-tagging algorithm to improve the overall ATLAS b-tagging performance.

### TABLE XII

For a fixed value of  $\varepsilon_b^{\text{soft}} \sim 7.2\%$ , expected efficiencies of the *b*-tagging algorithm,  $\varepsilon_b^{\text{alg}}$ , and of the jet rejection factors  $R_g, R_u, R_c$  for the standard and modified algorithms.

	fixed $\varepsilon_b^{\text{soft}} \sim 7.2\%$			
Procedure	$arepsilon^{\mathrm{alg}}_{b}$ $[\%]$	$R_{g}$	$R_u$	$R_c$
standard	$50 \pm 1$	$192\pm27$	$554 \pm 113$	$47 \pm 4$
$p_{\rm T}^{\rm thr} > 2 {\rm GeV}$ no conv. finding	$50 \pm 1$	$83 \pm 8$	$96 \pm 8$	$38 \pm 3$
$p_{\rm T}^{\rm thr} > 2 \text{ GeV}$ no $A_0$ $p^{\rm thr} > 2 \text{ GeV}$	$50 \pm 1$	$145 \pm 18$	$283 \pm 41$	$37 \pm 2$
$p_{\rm T} > 2 \text{ GeV}$ no $A_0$ and no $P_{Tj}$	$50 \pm 1$	$126 \pm 14$	$229\pm30$	$33\pm2$
$p_{\rm T}^{\rm pr} > 2 \text{ GeV}$ standard $p_{\rm T}^{\rm thr} > 1 \text{ GeV}$	$45.3\pm0.8$	$171 \pm 23$	$403\pm70$	$43 \pm 3$

The best results are achieved for the standard algorithm, and the worst for algorithms not using the  $\gamma$ -conversion and Dalitz decay finding procedure. Since a significant fraction of the background jets identified as *b*-jets is tagged through background electrons from  $\gamma$ -conversions and Dalitz decays (~ 28% of *g*-jets and ~ 33% of *u*-jets), an improved overall background rejection could be hoped for, if a better performance of the conversion finding procedure would be achievable.

The rejection factors obtained in this study are much better than those presented previously in [9,10] and [11], and these results confirm the importance of this additional *b*-tagging technique to improve the overall *b*-tagging performance of the ATLAS detector. It should be stressed, however, that at the moment neither the electronic noise nor the pile-up effects were included in the reconstruction of the EM Calorimeter information. In fact, pile-up effects should be studied in the future for both the Inner Detector and the EM Calorimeter reconstruction in order to evaluate the performance of softelectron *b*-tagging at high luminosity. The pile-up will probably degrade the overall tagging performance, but also some improvement is expected while taking into account the spiralling of low  $p_{\rm T}$  electrons in the magnetic field<sup>16</sup>.

In addition the results presented in this note were obtained for the central region of the ATLAS detector only. It is planned to extend the present studies to the full rapidity range in the near future.

Even if the algorithms presented in this note have been studied with the exclusive aim of tagging *b*-jets from Higgs boson decay, they could obviously be used more generally for any physics involving soft electrons in jets.

The authors acknowledge the very helpful and fruitful collaboration with M. Seman in the part of the analysis concerning the EM Calorimeter reconstruction and several crucial comments from U. Egede concerning the procedure of the photon conversion finding. We are grateful to J. Chwastowski for generating MC events and for valuable comments and advice on many technical details. We acknowledge warmly the very constructive suggestions from D. Froidevaux, E. Ros and D. Barberis. Last but not least, we thank E. Richter-Was and P. Malecki for their attention and support for this work and for never being satisfied with the achievable 'rejection factor', and to all the Cracow ATLAS Group for many valuable comments.

<sup>&</sup>lt;sup>16</sup> Many thanks to Michal Seman for pointing up on the spiralling effect.

### Appendix A

#### Photon conversions and Dalitz decays

### A.1 Photon conversions

Pair production is the most important electromagnetic process at high energies by which photons interact with matter [19]. The intense electric field near the nucleus can cause a photon to convert into an electron and a positron. The nucleus must be there to satisfy conservation of momentum, but the energy transfer to the nucleus is typically small. Therefore the  $e^+e^$ opening angle is also small. The positrons and the electrons are likely to be produced at any allowed energy. In general, they do not have the same energy.

### A.2 Dalitz decays of neutral pions

The neutral pion decays into two photons with a probability of nearly 99%. However, with a probability of approximately 1.2% an  $e^+e^-$  pair and one photon are produced. The second decay mode, called Dalitz decay, is also a pure electromagnetic process. In this decay, one real and one virtual photon are produced. The latter undergoes an internal conversion and decays into an electron and positron pair [24]. Compared to the  $\pi^o$  decay into two photons, this process is suppressed by the electromagnetic coupling constant. Since the mass of the virtual photon is small (although larger than for  $\gamma$ -conversions) a small opening angle between the  $e^+$  and  $e^-$  is expected. In general, the positron and electron have different energies.

### Appendix B

### Description of the XCONVER algorithm

The XCONVER algorithm [21] is designed to reconstruct  $e^+e^-$  pairs originating from photon conversions or Dalitz decays of neutral pions. In both cases the signature of such a process is a pair of oppositely charged particles with a common production vertex and an invariant mass close to zero.

The algorithm makes use of the fact that both the electron and the positron originate from the same conversion vertex and have a zero opening angle. In addition, it is assumed that the primary photon has zero impact parameter, thus neglecting any beam spread in the transverse plane. XCON-VER finds all the pairs of tracks of oppositely charged particles. The distance in the transverse plane between the tracks must be smaller than a given cut-off (normally 5 cm), where the distance is computed at the radius of the hit

closest to the primary vertex on both tracks. After this fast preselection, a  $\chi^2$ -fit is performed for all preselected pairs of tracks using the MINUIT [25] package. The obtained value of  $\chi^2$  is a measure of the goodness of the fit and is used to reject  $e^+e^-$  pairs originating from photon conversions and Dalitz decays.

# Appendix C

# Variables

The identifying variables are calculated as follows:

• The following variables are taken, as described in [21] and in [26], directly from the banks of the XKALMAN and XCONVER reconstruction packages:

Identifying <sup>•</sup>	variable	Reconstruction	package	variable

$N_{ m Si}$	tfit_NSiHits
$N_{\rm PIX}$	from tfit_Pattern
$N_{ m B}$	from tfit_Pattern
$N_{ m RTR}$	$tfit_NTRHits$
$A_0$	tfit_A0Vert
$\chi^2_{ m CONV}$	$cfit\_Chi2$

- Using information from the matrices filled during reconstruction in the EM Calorimeter (see section 4.2), the following variables are calculated:
  - $-E_1$  energy deposited in the first longitudinal sampling of the calorimeter,
  - $-E_2$  energy deposited in the second longitudinal sampling of the calorimeter,
  - $-E_3$  energy deposited in the third longitudinal sampling of the calorimeter,
  - $-E = E_1 + E_2 + E_3$  energy deposited in the whole depth of the calorimeter,

The energy is calculated in a  $3 \times 3$  window size in units of the cell sizes in the second longitudinal sampling of the EM Calorimeter.

- isolation of the shower

$$ISO = 1 - \frac{E_{3\times 3}}{E_{3\times 7}},$$

where  $E_{3\times3}$  and  $E_{3\times7}$  is the energy deposited in the  $3\times3$  and  $3\times7$  window around the cell with maximum energy deposition summed over the three samplings of the EM Calorimeter,

 shower width measured in the strips (first longitudinal sampling of the EM Calorimeter)

W = 
$$\sqrt{\frac{\sum_{i=j-1}^{i=j+1} E_i \times (i-j)^2}{\sum_{i=j-1}^{i=j+1} E_i}},$$

where j is the cell with the maximum energy deposition and  $E_i$ is the energy deposited in the *i*-th cell in the  $\eta$ -direction for a constant  $\phi = \phi_{max}$ ,

- asymmetry measured in the strips

$$AS = \left| \frac{\sum_{i=j+1}^{i=j+3} E_i - \sum_{i=j-3}^{i=j-1} E_i}{\sum_{i=j+1}^{i=j+3} E_i + \sum_{i=j-3}^{i=j-1} E_i} \right|,$$

where j is the cell with the maximum energy deposition and  $E_i$  is the energy deposited in the *i*-th cell in the  $\eta$ -direction for a constant  $\phi = \phi_{\text{max}}$ .

During the simulation of the events with the DICE package, no electronic noise in the EM Calorimeter nor pile-up events were taken into account.

- From combined Inner Detector and EM Calorimeter:
  - energy and transverse momentum matching

$$E_{\rm T}/p_{\rm T} = \frac{E_{T1} + E_{T2} + E_{T3}}{p_{\rm T}},$$

where  $p_{\rm T}$  is the transverse momentum of the track taken from the XKALMAN reconstruction package:

$$p_{\rm T} = |1.0/tfit\_PTInvVert|$$
,

— position of the shower measured in the strips

POS = 
$$\frac{\sum_{i=j-7}^{i=j+7} E_i \times (i-j)}{\sum_{i=j-7}^{i=j+7} E_i}$$
,

where j is the impact cell for the track reconstructed in the Inner Detector and  $E_i$  is the energy deposited in the *i*-th cell in the  $\eta$ -direction for constant  $\phi$  given by the track parameters,  transverse momentum of the track determined relative to the jet axis

$$P_{\rm TJ} = p_{\rm T} / \sin(\theta) \times \sin(\Delta R),$$

where  $p_{\rm T}$  is as defined above,  $\theta$  is the track angle with respect to the z axis calculated from XKALMAN variables, and  $\Delta R$  is the distance between the jet axis taken from the fast jet reconstruction package and the track direction.

### REFERENCES

- [1] D. Froidevaux, E. Richter-Was, ATLAS Internal Note PHYS-No-043 (1994).
- [2] E. Richter-Was et al., ATLAS Internal Note PHYS-No-074 (1996).
- [3] E. Richter-Was, D. Froidevaux, J. Soderqvist, ATLAS Internal Note PHYS-No-108 (1996).
- [4] G. Polesello et al., ATLAS Internal Note PHYS-No-109 (1996).
- [5] M. Cobal, ATLAS Internal Note PHYS-No-093 (1996).
- [6] T. Dorigo *et al.*, Observation of Z Decays to b Quark Pairs at the Tevatron Collider, hep-ex/9806022.
- [7] P. Abreu *et al.*, Z. Phys. C70, 531 (1996).
- [8] I. Gavrilenko et al., ATLAS Internal Note INDET-No-115 (1995).
- [9] F. Gianotti, ATLAS Internal Note PHYS-No-049 (1995).
- [10] S. Jagielski, ATLAS Internal Note PHYS-No-087 (1996).
- [11] ATLAS Collaboration, Inner Detector performance TDR, pp.160–162.
- [12] D. Barberis, presentation on ATLAS Physics Workshop, Grenoble'98.
- [13] E. Ros, presentation on ATLAS Physics Workshop, Grenoble'98.
- [14] A. Artamonov, A. Dell'Acqua, D. Froidevaux, M. Nessi, P. Nevski, G. Poulard, ATLAS Internal Note SOFT-No-014 (1995).
- [15] OPAL Collaboration, Measurement of  $\Gamma(Z^0 \to b\bar{b})/\Gamma(Z^0 \to hadrons)$  using leptons. Z. Phys. C58, 523 (1993).
- [16] E. Richter-Was, ATLAS Internal Note PHYS-No-079 (1996).
- [17] D. Cavalli, S. Resconi, ATLAS Internal Note PHYS-No-100 (1997).
- [18] A.G. Clark, A. Pickford, A. Poppleton, G. Stavropoulos, M. Tadel, ATLAS Internal Note INDET-No-204 (1998).
- [19] For a description of photon interactions with matter, see for example: R.C. Fernov, *Introduction to Experimental Particle Physics*, Cambridge University Press 1986.
- [20] See for example P. Pralavorio, presentation on ATLAS Physics Workxshop, Grenoble'98.
- [21] U. Egede, LUNDF6/(NFFL-714X)1997.

- [22] Particle Data Group, The European Physical Journal, C3 (1998).
- [23] J.J. Bunn, An evaluation of PAW, Proceedings L. Taylor, C.E. Vandoni, CERN, Geneva, CERN-97-01 (123–125) 1997.
- [24] R.H. Dalitz, Proc. Phys. Soc. London, Sect. A 64, 667 (1951).
- [25] F. James MINUIT Function Minimisation and Error Analysis, CERN, March 1994.
- [26] U. Egede LUNDF6/(NFFL-7133)1996.