10 *b*-tagging performance

10.1 Introduction

Tagging the flavour of *b*-jets is useful for the selection of events containing the top quark, or Standard Model or supersymmetric Higgs bosons, which couple preferentially to heavy objects. This chapter describes the lifetime and leptonic decay methods that can be used for *b*-tagging and the performance that can be achieved, in terms of tagging efficiency and background rejection.

10.1.1 Detector layouts

The simulation of the events needed for the study of the *b*-tagging performance was done with the 97_6 and 98_2 geometries (see Section 2). They are identical as far as the Inner Detector is concerned; differences in the geometry of the cryostat and the calorimeters are taken into account by the reconstruction program and do not affect the performance. The muon identification efficiency was parametrised using an *ad hoc* simulation and then inserted into the global analysis program.

10.1.2 Jets used for *b*-tagging studies

Hadronic decays of Higgs bosons were used extensively as a means of testing pattern recognition within a jet of particles. Further, these decays were used for studies of the *b*-tagging capability of the Inner Detector which can be directly compared with physics requirements. In particular, comparison has been made between the decays $H \to b\bar{b}$ and backgrounds $H \to x\bar{x}$ where x is a *u*-, *d*-, *s*-, *c*-quark or gluon. While these background processes actually have negligible rates, the decays are considered representative of actual backgrounds which will be encountered at the LHC, and by generating them from Higgs decays, direct comparisons can be made with background jets having the same kinematics as the *b*-jets. Complete events, $pp \to WH + X$ and $pp \to ZH + X$ with $W \to \mu\nu$ and $Z \to \mu\mu$ were generated with PYTHIA [10-1].

Previous studies [10-2], [10-3] concentrated on *b*-tagging for a light Higgs ($m_H = 80\text{-}100$ GeV) since this is a plausible method for identifying a Standard Model Higgs in this mass range [10-4]. In these studies, at around 50% *b* efficiency the rejection of gluon jets was found to be limited by gluon splitting to heavy flavours, and no limitations arising from pattern recognition were observed. In the Inner Detector TDR [10-5] and the Pixel Detector TDR [10-6], the emphasis was put on $m_H = 400$ GeV, which provides a more stringent test of the two-track separation and pattern recognition capability of the Inner Detector. At this mass, the decay to *b*-jets would have a very small branching ratio. Nevertheless, this decay mode was used as a 'factory' for producing high-p_T *b*-jets. Such jets might be seen in the decays of light SUSY Higgs $h \to b\bar{b}$ which in turn may come from the decays of heavy supersymmetric particles. The current study uses both $m_H = 100$ and 400 GeV, in order to cover with sufficient statistics the p_T range between ~15 and ~400 GeV and the full pseudorapidity range $|\eta| < 2.5$. The results will be presented as a function of p_T and $|\eta|$. A comparison with the performance calculated using *b*-jets from $t\bar{t}$ events was made and the results found to be the same; this gives confidence in the extrapolation to other sources.

10.1.2.1 Properties of $H \rightarrow b\bar{b}$ events with $m_H = 100$ and 400 GeV

Figures 10-1 and 10-2 show the transverse momentum and the pseudorapidity distributions of the b-jets from Higgs decay, for $m_H = 100$ and 400 GeV. The distance ΔR of charged particles to the b-quark direction in $\eta{-}\phi$ is shown in Figure 10-3. For both $m_H = 100$ GeV and 400 GeV, 99% of charged particles produced from b-quark fragmentation are found in a cone of $\Delta R \leq 0.4$.
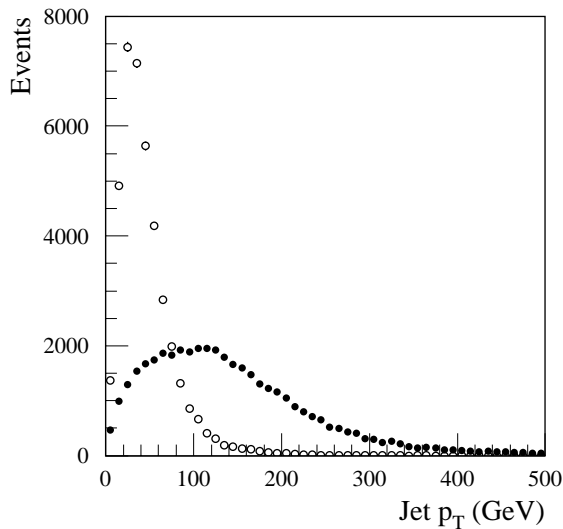
Figure 10-1 Transverse momentum distribution of b-jets from $H \rightarrow b\bar{b}$, $m_H = 100$ GeV (open circles) and 400 GeV (closed circles).
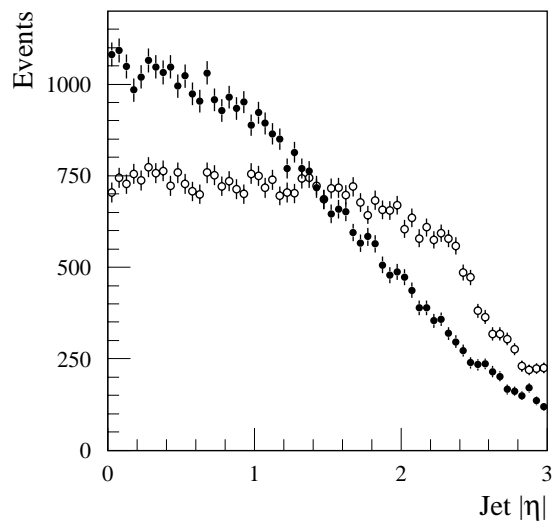
Figure 10-2 Pseudorapidity distribution of b-jets from $H \rightarrow b\bar{b}$, $m_H = 100$ GeV (open circles) and 400 GeV (closed circles).
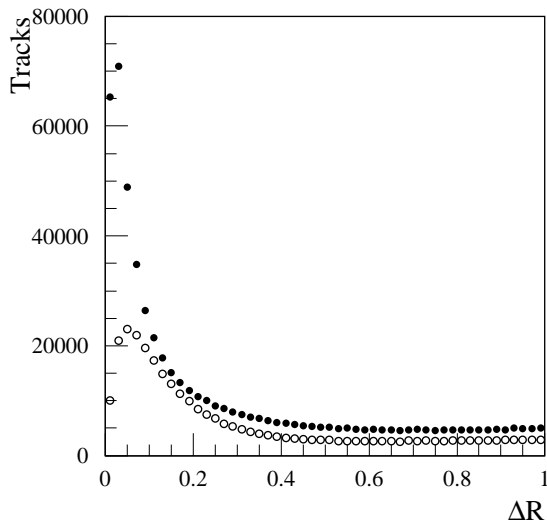
Figure 10-3 Distance ΔR of all charged particles with $p_T > 1$ GeV to the b-quark direction, for $m_H = 100$ GeV (open circles) and 400 GeV (closed circles).
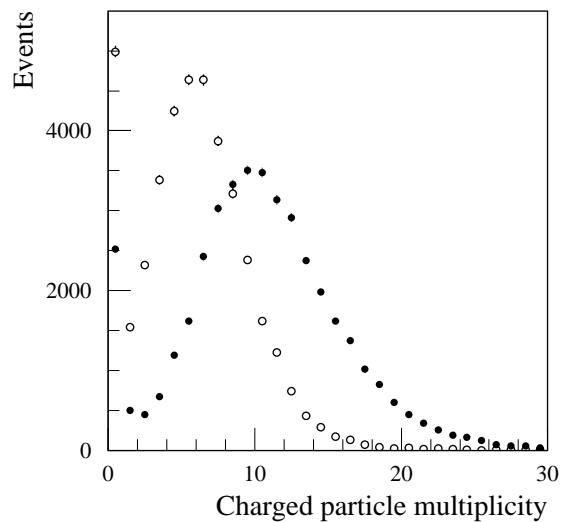
Figure 10-4 Multiplicity of charged particles with $p_T > 1$ GeV in $\Delta R < 0.4$ around a b-jet from $H \rightarrow b\bar{b}$, $m_H = 100$ GeV (open circles) and 400 GeV (closed circles).
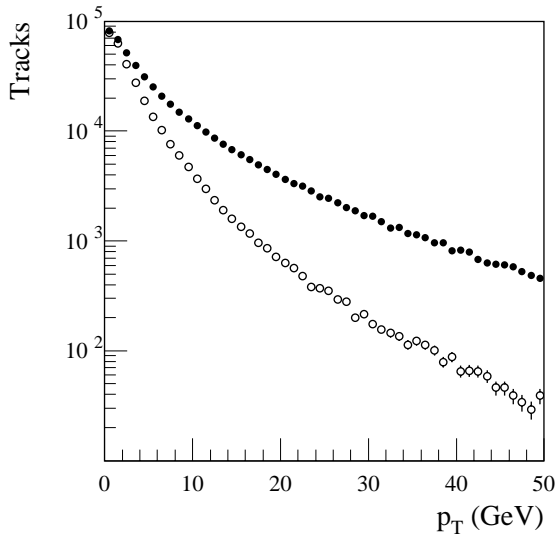
Figure 10-5 Transverse momentum of charged particles in a *b*-jet, $m_H = 100$ GeV (open circles) and 400 GeV (closed circles).
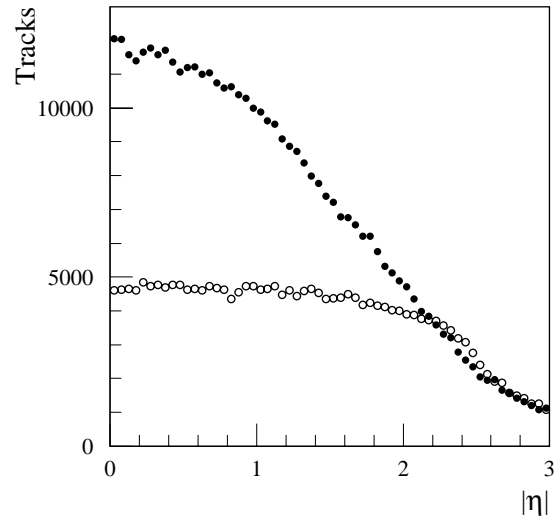
Figure 10-6 Pseudorapidity of charged particles in a *b*-jet, $m_H = 100$ GeV (open circles) and 400 GeV (closed circles).

The charged multiplicity of particles with $p_T > 1$ GeV, in a cone $\Delta R \le 0.4$ is shown in Figure 10-4. The charged multiplicity in the cone has a mean of 5.5 (10.0) for $m_H = 100$ GeV (400 GeV), 60% (40%) of which come from daughters of the *B*-hadron decay. In 12.5% (6.3%) of the jets, there are no charged particles with $p_T > 1$ GeV in the cone. In a cone $\Delta R \le 0.4$, the mean number of photons is 10 (18), of which 5 (9.5) have $p_T > 1$ GeV.

The transverse momentum and the pseudorapidity distribution of these particles are shown in Figures 10-5 and 10-6. The average p_T of particles with $p_T > 1$ GeV is 4.9 (8.7) GeV – 70% of particles have p_T less than this. The conclusion is that it is necessary to reconstruct accurately at least four tracks arising from *B*-hadron decays with $p_T < 10$ GeV in a region where multiple scattering dominates the impact parameter resolution.

The *b*-jet direction is not uniquely defined. *b*-quarks in the final state of an interaction or a decay can radiate gluons (Final State Radiation, FSR) and therefore change the direction. As, in practice, the reconstruction of the jet direction and energy is done using the energy deposited in the calorimeters, it is important to check the difference between the jet direction as measured by the calorimeters and the *b*-quark direction. Figure 10-7 shows the distribution of the angular distance ΔR between the *b*-jet axis as reconstructed by the calorimeter and the *b*-quark direction before and after FSR, and
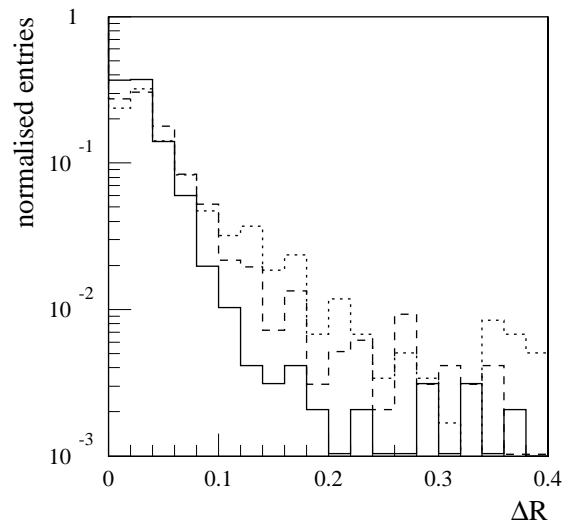

Figure 10-7 Angular distance ΔR between the *b*-jet axis as reconstructed by the calorimeter and the *b*-quark direction before Final State Radiation (dotted), the *b*-quark direction after FSR (dashed) and the *b*-jet axis as reconstructed from charged tracks (solid).

the *b*-jet axis as reconstructed from the centre of gravity of charged tracks. As the resolution of the *b*-jet direction is much better than the standard cone size $\Delta R \leq 0.4$, the use of a finite cone size does not cause significant loss of tracks.

10.1.2.2 Data sets and event reconstruction

The expected rejection factors for non-*b* jets are of the order of 100, at least for most of the allowed ranges of jet transverse momentum and pseudorapidity. In order to have a statistically significant number of events at the end of the selection and analysis procedure, 20 000 events were fully simulated for each of the channels $H \to b\bar{b}$, $H \to u\bar{u}$, $H \to c\bar{c}$, $H \to gg$ with $m_H = 100$ GeV, and at least 10 000 events for each of the same channels with $m_H = 400$ GeV. Events were selected at the generation level, so that both decay products of the Higgs have $p_T > 15$ GeV and at least one of them is within the pseudorapidity range $|\eta| < 2.5$, before final state radiation.

The reconstruction procedure has to be as realistic as possible, therefore no use was made of the generated event information. First, jets were identified as clusters in the combined electromagnetic and hadronic calorimeters, using a noise threshold of 2σ on each cell and a p_T threshold of 10 GeV for the reconstructed jets. Then the Inner Detector pattern recognition programs xKalman and iPatRec were run, using as seeds the reconstructed jet directions and a road half-width of 0.5; the p_T threshold for reconstructed tracks was set to 0.7 GeV. Finally the xConver algorithm was run in order to identify, and reject, electrons coming from photon conversions.

A jet in the $H \to b\bar{b}$ events was labelled as a *b*-jet if a *b*-quark from the original hard process pointed, after final state radiation, along the jet within an angular distance of $\Delta R < 0.2$. Similarly, jets were labelled as for *u*-, *c*- and gluon jets for the other data samples.

10.1.3 Minimum bias events

Minimum bias events account for the vast majority of interactions which will result from beam collisions in ATLAS; by implication they do not contain hard-scattering processes. These events are of little interest *per se*, but at the high luminosities at LHC, multiple collisions within one beam-crossing will be inevitable, causing signal events to have several minimum bias events superimposed. The pile-up of these events on top of single particles is essential for realistic studies of pattern recognition.

Minimum bias events were generated individually using PYTHIA 5.7. The processes of interest for tracking studies are the inelastic, non-diffractive *pp* interactions, labelled in PYTHIA as 'QCD high-p_T processes' (the switches set in PYTHIA are: MSEL=1, MSTP(2)=2, MSTP(33)=3, MSTP(81)=1, MSTP(82)=4). The pseudorapidity distribution of charged particles in single minimum bias events is shown in
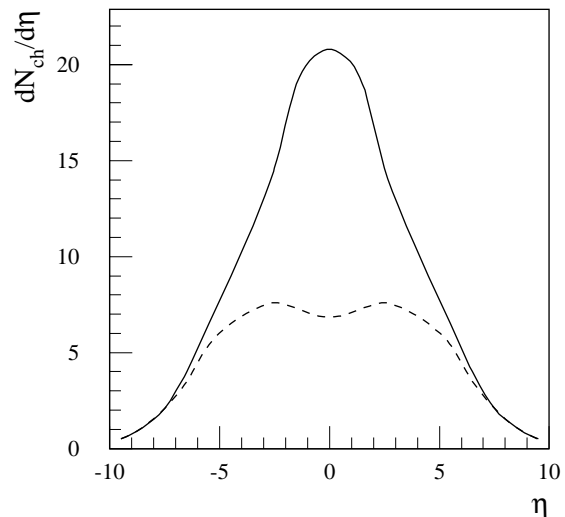

Figure 10-8 Pseudorapidity distribution of charged particles (no p_T cut). The solid line corresponds to all particles in events containing the decay of a 400 GeV Higgs to *b*-jets, the dashed line to all particles in minimum bias events.

Figure 10-8. Comparison is made with the distribution from events containing a hard scatter, namely $H \rightarrow b\bar{b}$ ($m_H = 400$ GeV). The latter is averaged over many events, but is quite different in character due to the high density of particles to be found in the jets and the additional initial state radiation.

The p_T distributions are shown in Figure 10-9, while the integrals of the p_T distributions are shown in Figure 10-10. It can be seen that the average charged particle multiplicity per unit of pseudorapidity, $dN/d\eta$, for a single minimum bias event is 7.5 (no p_T cut), falling to 0.64 for $p_T \geq 1$ GeV and 0.006 for $p_T \geq 5$ GeV. The mean $dN/d\eta$ for neutrals is 9.1, 90% of which are photons, and the mean E_T is 235 MeV [10-7]. Consequently, the number of charged particles with $p_T \geq 1$ GeV from the pile-up of minimum bias events is 1.2 within a cone $\Delta R \leq 0.4$ around a b-quark at a luminosity of 10^{34} cm^{-2}s^{-1}.
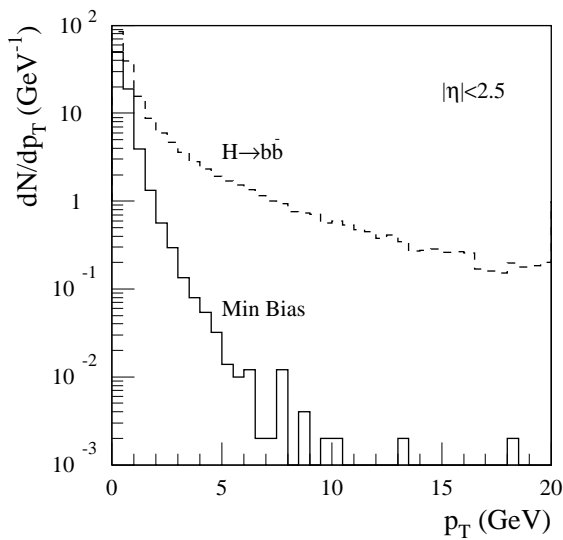

Figure 10-9 p_T distribution of charged particles in minimum bias and $H \rightarrow b\bar{b}$ events.
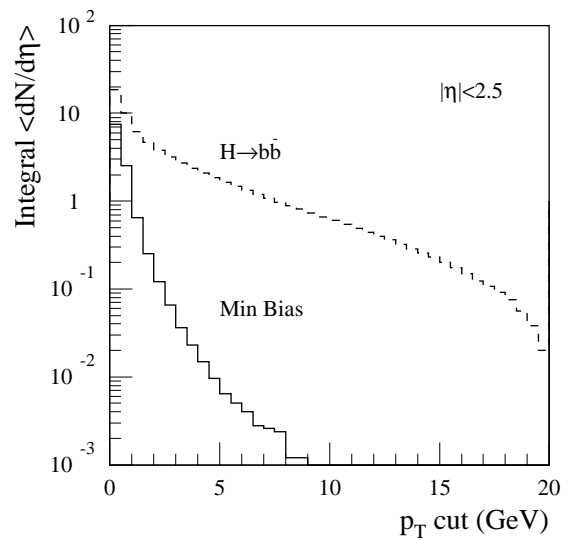
Figure 10-10 Average multiplicity per unit of pseudorapidity ($|\eta| \leq 2.5$) as a function of p_T cut (corresponding to left-hand edge of bin).

10.2 Vertexing algorithm

In this section, the tagging of b-quarks from the decay $H \rightarrow b\bar{b}$ ($m_H = 100$ and 400 GeV) is discussed and compared with the rejection of lighter-quark and gluon decays of the Higgs. The relatively long lifetimes of b-hadrons ($c\tau \approx 470$ µm) give rise to displaced vertices which may be tagged by either explicitly reconstructing the vertex or by examining the impact parameters of the daughters. In the work described in this section, the latter method was used since high b-tagging efficiency (rather than high purity) is required. The rejection of non b-jets is dependent on the fact that for light quarks, most of the stable particles which can be reconstructed in the Inner Detector come from the decays of short-lived objects and hence appear to come from the primary vertex. The extent to which this is true is determined by the impact parameter[1] resolution $\sigma(d_0)$ of the detector, by the rate of conversions and nuclear interactions in the detector ma-

---

1. Unless stated otherwise, impact parameter will refer to the transverse impact parameter d_0, since this is the most relevant one for b-tagging.

terial, and by the frequency of pattern recognition errors. Figure 10-11 shows the impact parameter resolution for primary pions in jets. The rise with $|\eta|$ is mainly due to the increase in material in the forward direction.

What is of interest for *b*-tagging is the signed impact parameter ($sign \times |d_0|$), where the *sign* is positive if the track appears to originate from in front of the primary vertex (*i.e.* the track crosses the jet axis in front of the primary vertex) and negative if it appears to originate from behind. The jet axis is determined accurately from the calorimeter; as explained above, it can also be determined by the charged tracks measured in the ID.

In principle, the primary vertex can be reconstructed from prompt tracks in the event, as described in Section 3.6.1. In practice, since the gain is small, this may not be done in *x* and *y* (there is a danger of introducing systematics), and here the impact parameter has been determined with respect to the nominal beam position ($x = 0$, $y = 0$ in this simulation), which is anticipated to be measured with high accuracy as a function of time. Therefore the uncertainty on the impact parameter is the combination of the measurement error $\sigma(d_0)$ and the spread of the beam-spot, 15 µm, taken in quadrature. The effect of this is to increase $\sigma(d_0)$ by about 4 µm.
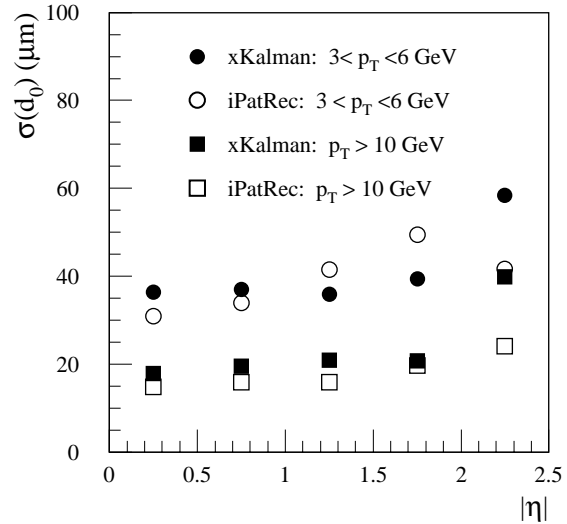

Figure 10-11 Impact parameter *rms.* resolution for tracks found in *b*-jets from $H \rightarrow b\bar{b}$ with $m_H = 100$ GeV.

The *z* coordinate of the primary interaction point can be reconstructed simply by taking the truncated weighted average of the z_0 coordinates of all well-reconstructed tracks in the event which have a transverse impact parameter $d_0 < 2\sigma(d_0)$. This simple and fast method achieves a resolution ~35 µm for events with at least four reconstructed tracks, which is always the case for Higgs events (see Figure 10-12).

The significance of the impact parameter is defined as the ratio of the signed impact parameter to its total error. This is shown in Figure 10-13 for tracks from *b* and *u*-jets. Both distributions have significant 'cores' which represent correctly reconstructed tracks coming from the primary vertex. These cores can be described by Gaussians of width close to one. The *b*-jets contain tracks with large positive significance, corresponding to genuine lifetime content. By contrast, light quarks and gluon jets have only a small excess of tracks which appear to contain lifetime arising from:

- interactions with material, *e.g.* photons converting to e^+e^- pairs or pions having nuclear interactions;

- daughters of V^0's;

- daughters of heavy quarks formed in the fragmentation (relevant only for gluon jets).

There are tails in the impact parameter distributions which are apparent on the negative side of the significance distribution. For the *b*-jets, these come mainly from an incorrect determination of the sign (corresponding to the uncertainty in determining the *b*-hadron direction) and from decays of charmed states. For the *u*-jets, the tail is dominated by secondaries from interactions.
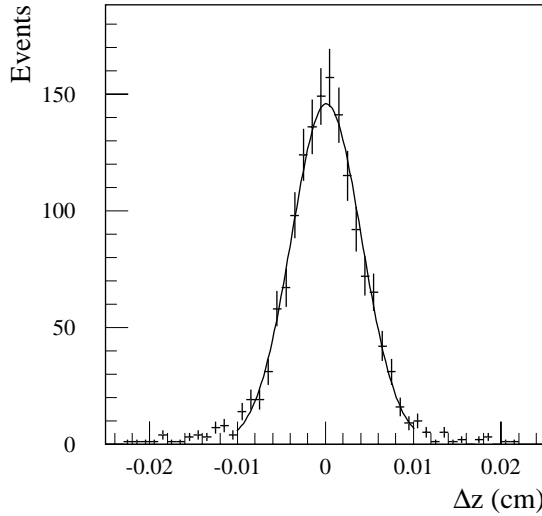
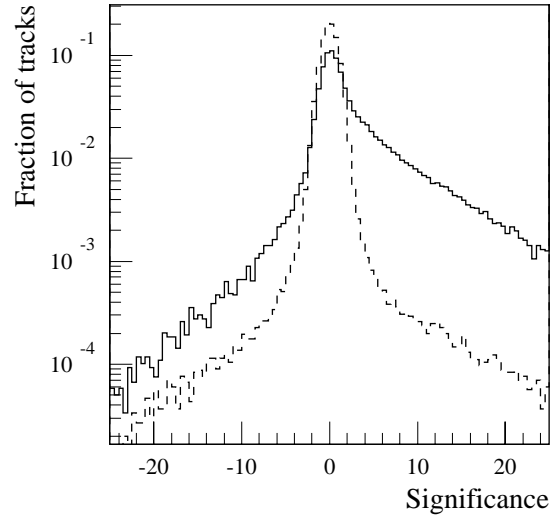**Figure 10-12** *z*-vertex resolution for $H \to b\bar{b}$ events.

Figure 10-13 Significance distribution: signed impact parameter divided by its error. Curves for *b*-jets (solid) and *u*-jets (dashed) are normalised to the same area.

10.2.1 *b*-tagging methodology

Various *b*-tagging methods have been investigated in the past. The likelihood ratio method appears to offer the best performance, therefore it is the method used for the studies presented here. The emphasis of these studies is on individual jets, rather than complete events. Throughout, the rejection R_j for different background jets is compared with the efficiency ε_b for keeping *b*-jets. The rejection is simply the reciprocal of the efficiency. If there is no discrimination at all, R will behave like $1/\varepsilon_b$.

The method was as follows:

1. For each selected track *i* in a jet, the significance S_i was calculated.

2. The ratio of the values of the significance probability distribution functions for *b*-jets and *u*-jets was computed: $r_i = f_b(S_i)/f_u(S_i)$.

3. A jet weight was constructed from the sum of the logarithms of the ratios: $W = \Sigma \log r_i$.

4. By keeping jets above some value of *W* (a value which can be varied), the efficiency for different jet samples can be obtained.

By using the significance distribution $f_u(S)$ for *u*-jets, the method was optimised for the rejection of *u*-jets. In the case of real data, since the jet type will not be *a priori* known, the rejection will have to be optimised for each specific background under study.

10.2.2 Track selection

Track finding was performed in restricted cones using both xKalman and iPatRec. Only tracks with $p_T > 1$ GeV and in a cone of $\Delta R < 0.4$ around the jet direction were considered. The efficiency for finding any primary charged track in a *b*-jet was 93.9% for xKalman and 92.8% for iPatRec. The following ('standard') cuts were then made:

- number of precision hits ≥ 9 (no explicit TRT requirement was made, although the valid drift-time hits were used for the global track fit);

- number of pixel hits ≥ 2;

- there must be a *B*-layer hit;
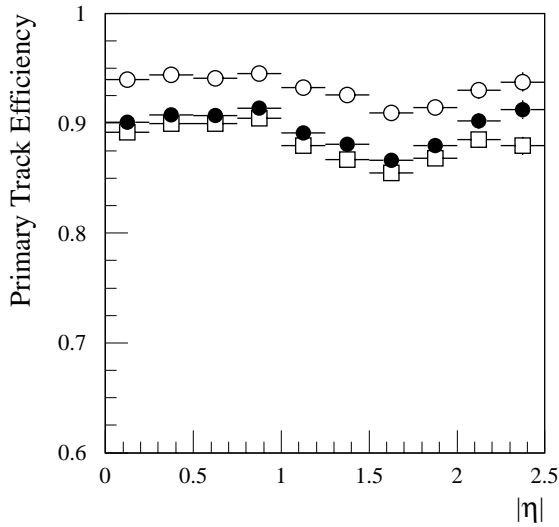
- $|d_0| < 1$ mm.

Figure 10-14 Primary track efficiency for xKalman as a function of pseudorapidity without cuts (open circles), after applying standard cuts (closed circles) and after applying quality cuts (open squares).

Figure 10-15 Primary track efficiency for iPatRec as a function of pseudorapidity without cuts (open circles), after applying standard cuts (closed circles) and after applying quality cuts (open squares).

After all these cuts, xKalman and iPatRec yielded similar track finding efficiencies of 88.8% and 88.2% respectively for tracks from the primary vertex. The main limitation to the *b*-tagging performance arises from secondary tracks, *i.e.* tracks originating from interactions of primary tracks with detector material. Secondary tracks can only be rejected by applying appropriate cuts on track quality. The additional ('quality') cuts were:

- impact parameter in the *R-z* plane relative to the reconstructed primary interaction point $< 1.5 / \sin(\theta)$ mm;

- track fit $\chi^2/\mathrm{dof} < 3$;

- at most one hit in the Pixels shared with another reconstructed track;

- at most two hits in the precision detectors (pixel and SCT) shared with another reconstructed track.

Secondaries were now the most important component of the background; 37% of these secondaries were electrons from conversions and these could be removed by direct tagging. The program xConver was used to identify electrons from conversions and remove them from the track sample, as described in Section 7.5.1. Tracks were rejected if they had at least 5% high-threshold hits in the TRT (or no TRT hits at all, which happens in case of hard bremsstrahlung) and formed a good conversion candidate ($\chi^2 < 50$ for the conversion fit and conversion radius > 2 cm) with another track in the same jet. This cut removed 15% of the secondary tracks without affecting the track-finding efficiency for primary tracks.

Figure 10-16 Fraction of secondary tracks for xKalman as a function of pseudorapidity without cuts (open circles), after applying standard cuts (closed circles) and after applying quality cuts (open squares).
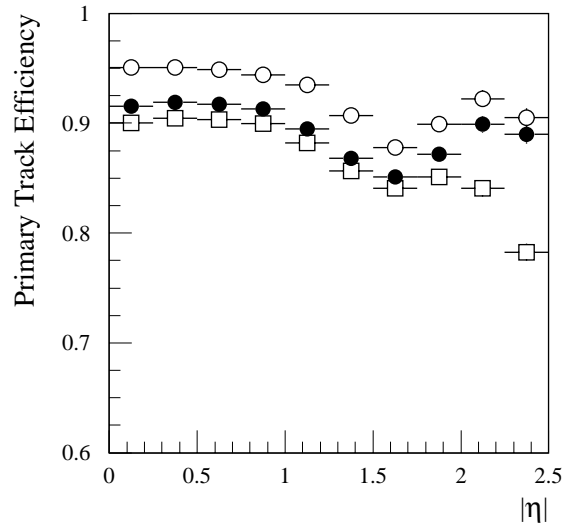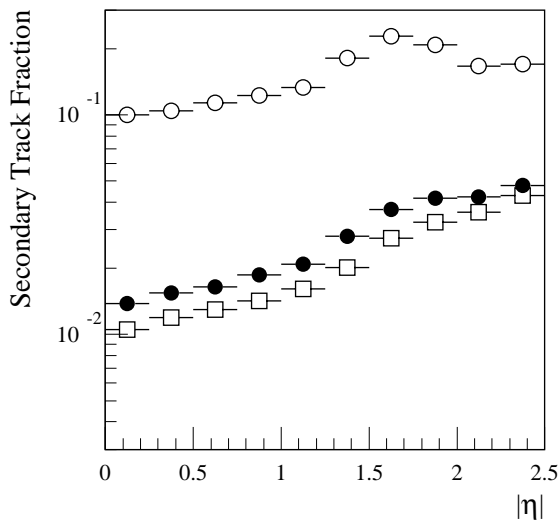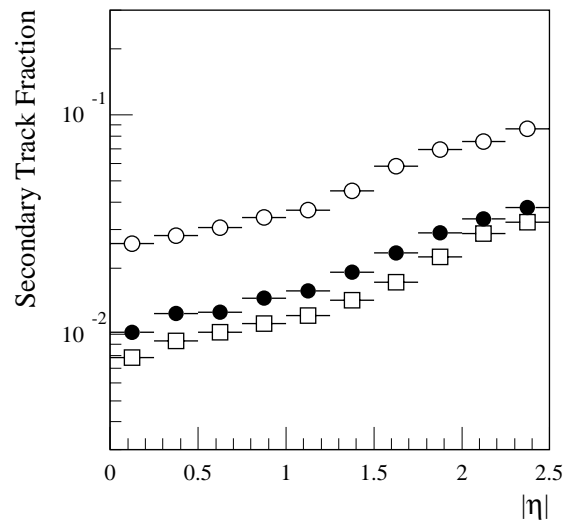

Figure 10-17 Fraction of secondary for iPatRec tracks as a function of pseudorapidity without cuts (open circles), after applying standard cuts (closed circles) and after applying quality cuts (open squares).
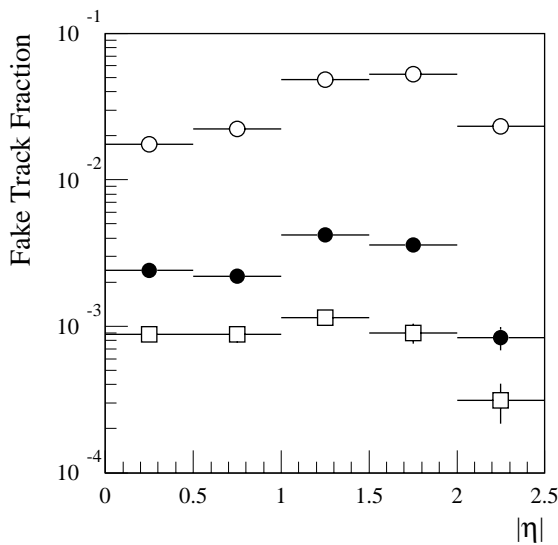
Figure 10-18 Fraction of fake tracks for xKalman as a function of pseudorapidity without cuts (open circles), after applying standard cuts (closed circles) and after applying quality cuts (open squares).
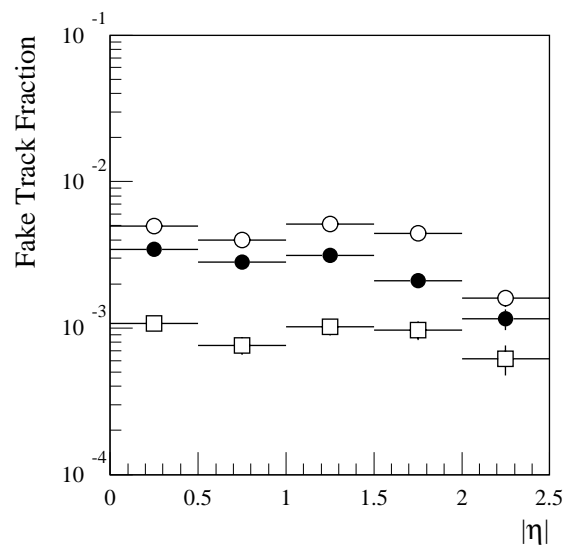
Figure 10-19 Fraction of fake tracks for iPatRec as a function of pseudorapidity without cuts (open circles), after applying standard cuts (closed circles) and after applying quality cuts (open squares).

As an example of the reconstruction performance using these cuts, Figures 10-14 to 10-19 show the primary track finding efficiency and the fraction of secondary and fake tracks in the final track sample.

The quality cuts effectively select a smaller but better-measured sample of tracks. Their net effect is quite different on jets reconstructed with xKalman or iPatRec:

- Even after the standard cuts, xKalman produces a larger fraction of secondary, duplicated and fake tracks, which are efficiently removed by the quality cuts. This is largely due to the sometimes conflicting requirements imposed on xKalman, which is the most widely

used pattern recognition program. These requirements have led to the removal of all or almost all internal quality cuts, *e.g.* to optimise the efficiency to find secondary tracks from long-lived particles. As a consequence, the quality of the reconstructed tracks is entirely determined by the selection cuts described above.

- In contrast, similar cuts to those described above are in some cases already implemented internally in iPatRec, leading to a better track quality after reconstruction.

In the case of iPatRec, the quality cuts described above tend to reduce the track sample, especially for $|\eta| > 2$ (see Figure 10-15), without degrading substantially the *b*-tagging performance, which, as described in Section 10.2.3, is limited much more by the amount of secondaries than that of fakes. Therefore, the results are quoted from now on, for xKalman with quality cuts and for iPatRec with standard cuts. The removal of conversions found by xConver is applied in all cases.

The pattern recognition performance presented in this section is somewhat better than at the time of the ID TDR: after the standard cuts the fraction of secondaries is lower by up to 40% depending on $|\eta|$ and the fraction of fakes is lower by almost a factor of two, for both pattern recognition programs. This can be explained by the continuous improvements and tuning of the pattern recognition and track-following algorithms in both programs.

10.2.3 Basic performance

The jet weights calculated with the likelihood ratio method are shown for *u* and *b*-jets in Figure 10-20. As all the relevant distributions are compatible in the p_T range where the two data sets overlap, in the following analysis the jets from the decay of 100 and 400 GeV Higgs are used together.

The following background compositions can be deduced for xKalman (iPatRec) respectively:

- 31% (21%) of tracks arise from interaction in the material of the detector (53% (32%) of these tracks are electrons from photon conversions, the other particles are products of nuclear interactions).

- 23% (21%) of tracks are produced in the decays of hadrons with significant lifetime (mainly K_s^0); hadrons containing heavy quarks are produced in 1.7% of *u*-jets.

Figure 10-20 Jet weights from likelihood ratio: *u*-jets with standard cuts (dotted line) and with quality cuts and conversion removal (solid line), *b*-jets with standard cuts (dash-dotted line) and with quality cuts and conversion removal (dotted line) (xKalman).

- 46% (58%) of tracks are produced at the primary vertex, with large deflections from multiple scattering or possibly pattern recognition problems.
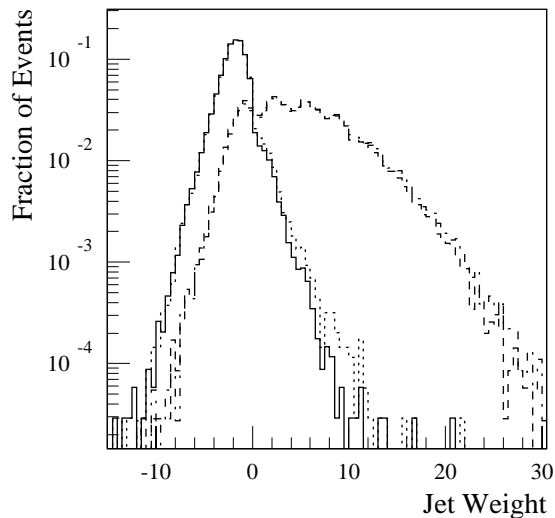
By varying the cut on the jet weights, it is possible to derive the curves of jet rejections as functions of b-jet efficiency; Figures 10-21 and 10-22 show the rejections obtained for different jet types. The rejections for $\varepsilon_b = 50\%$ and 60% are shown in Table 10-1. The rejection factors presented here are substantially better than those published in the ID TDR [10-5]. The improvement can be explained by:

- the improvement of the quality of tracks reconstructed by the pattern recognition programs (after the standard cuts only 1.2% (1.7%) of the tracks in jets from 100 (400) GeV Higgs are in the tails of the impact parameter distribution);

- the choice of jet axis: jets produced by very low-p_T quarks (after final state radiation), which are now not reconstructed by the calorimeter, would contain low-momentum particles, whose impact parameter resolution is dominated by multiple scattering.

The rejection of c-jets, R_c, is limited by the lifetime of charmed hadrons: for $D^{\pm}$, $c\tau = 317$ μm; for D^0, $c\tau = 124$ μm. The rejection of gluon jets, R_g, is limited by gluon splitting: $BR(g \rightarrow c\bar{c}) = 6\%$ and $BR(g \rightarrow b\bar{b}) = 4\%$ for $m_H = 400$ GeV.
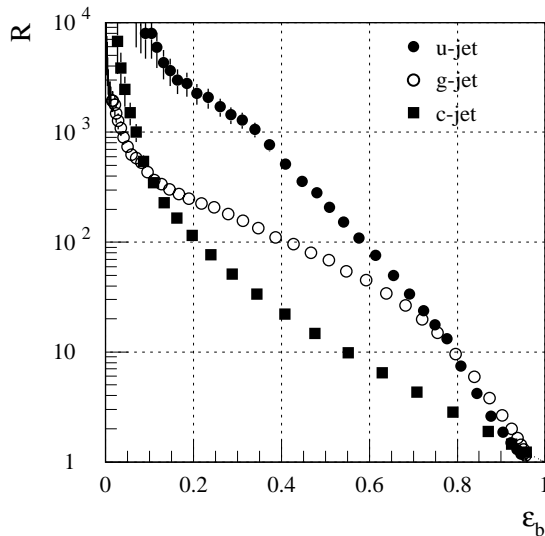
Figure 10-21 Background rejections as a function of b-jet efficiency obtained with xKalman, applying quality cuts and conversion removal.
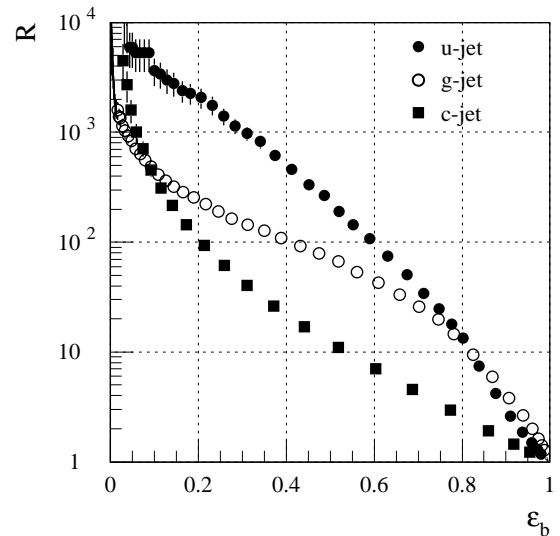
Figure 10-22 Background rejections as a function of b-jet efficiency obtained with iPatRec, applying standard cuts and conversion removal.

Table 10-1 Rejections of various types of background jets.

	$\varepsilon_b = 50\%$				$\varepsilon_b = 60\%$			
	$m_H = 100$ GeV		$m_H = 400$ GeV		$m_H = 100$ GeV		$m_H = 400$ GeV	
	xKalman	iPatRec	xKalman	iPatRec	xKalman	iPatRec	xKalman	iPatRec
R_u	326 ± 37	391 ± 49	126 ± 9	183 ± 17	124 ± 9	148 ± 11	65 ± 3	80 ± 5
R_g	135 ± 12	148 ± 14	59 ± 3	58 ± 3	72 ± 4	74 ± 5	38 ± 1	37 ± 1
R_c	13.6 ± 0.4	11.7 ± 0.3	13.3 ± 0.4	13.4 ± 0.4	7.7 ± 0.2	6.9 ± 0.1	8.0 ± 0.2	7.3 ± 0.2

10.2.4 Jet p_T and pseudorapidity dependence

The rejections as a function of $|\eta|$ are shown in Figure 10-23. The drop in rejection observed for $|\eta| > 1.5$ (compared to $\eta \sim 0$) is related to the increase of material in this region.

The dependence on the jet p_T is shown in Figure 10-24. Both xKalman and iPatRec show the same trends: a sharp fall in the rejection at low p_T as p_T tends to 0, and a slow degradation at high p_T as p_T increases.

The sharp fall at low p_T is due to:

- The decrease of charged particle multiplicity in jets. For jet $p_T < 30$ GeV, the charged particle multiplicity after acceptance ($p_T > 1$ GeV) and selection cuts is 2.3 (to be compared with an average of 4.3 for $p_T < 100$ GeV) and 30% of the b-jets have less than two selected tracks (compared with 10% for $p_T < 100$ GeV).

- The worse impact parameter resolution arising from increased multiple scattering of tracks coming from a softer p_T spectrum.
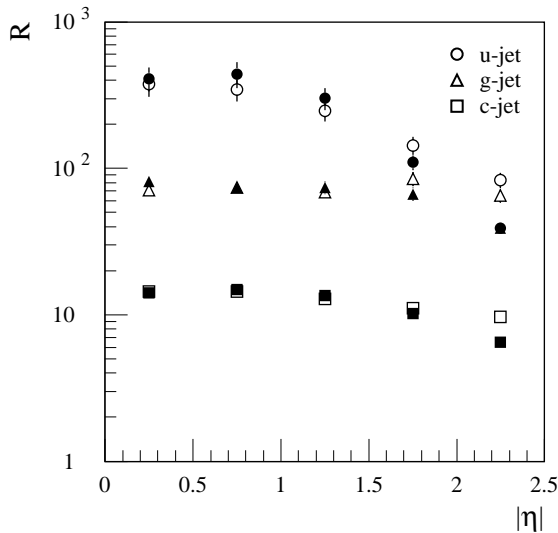
Figure 10-23 Background rejections as a function of jet $|\eta|$ for $\varepsilon_b = 50\%$. xKalman: full symbols; iPatRec: open symbols.
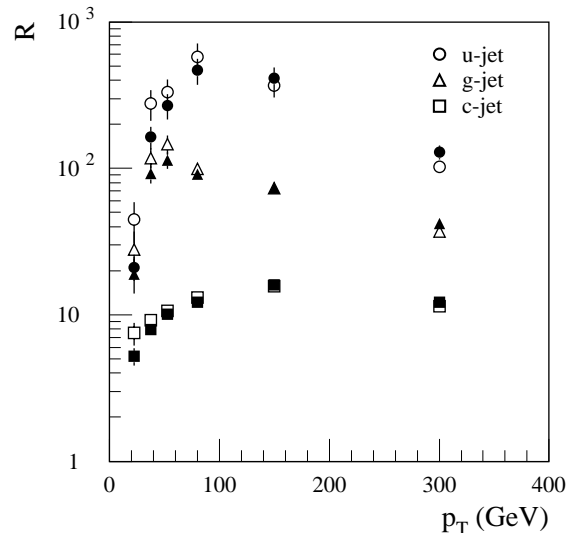
Figure 10-24 Background rejections as a function of jet p_T for $\varepsilon_b = 50\%$. xKalman: full symbols; iPatRec: open symbols.

The slow degradation at high p_T is the consequence of several factors:

- The increase in multiplicity of charged particles from the primary interaction vertex due to the increase of the hadronisation component of the jet. Consequently, the discrimination between b- and u-jets is reduced. For jets with $p_T < 100$ GeV, the number of tracks (after selection cuts) is 4.3 on the average, whereas it is 10.4 for $p_T > 270$ GeV.

- The increase in the fraction of reconstructed secondaries contained in the jet, due to the higher momentum of the interacting particles: 1.3% for jet $p_T < 100$ GeV and 2.2% for $p_T > 270$ GeV.

- The difficulties of pattern recognition in more dense environment at larger jet p_T values result in more tails in the impact parameter distributions. The number of prompt tracks with $|d_0| > 3\sigma(d_0)$ is 1.1% for $p_T < 100$ GeV, and 2.3% for $p_T > 270$ GeV.

In order to study the correlation between the p_T and the pseudorapidity dependence of the non-b jet rejection factor, one can consider the normalised rejection factor R_j^n, *i.e.* the ratio between $R_j(p_T,\eta)$ and the mean value $R_j(p_T)$. R_j^n does not depend much on the actual value of the b-jet tagging efficiency ε_b taken as reference. Figures 10-25 and 10-26 show respectively the normalised u- and c-jet rejection factors as functions of the jet p_T and pseudorapidity. As can be seen, there is little correlation between the p_T and pseudorapidity dependence of R_u, whereas R_c, which is limited by the physics of charm decays, is almost constant. Table 10-2 shows the rejection factor R_u in several bins of $|\eta|$ and p_T.
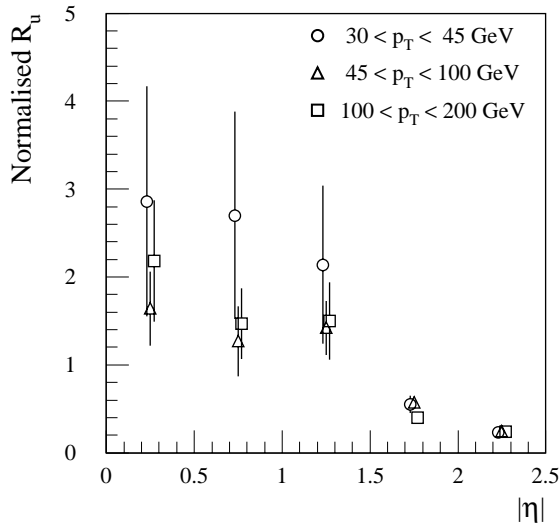

Figure 10-25 *u*-jet rejection factors, normalised to their mean values, as function of pseudorapidity for jets in different p_T ranges.
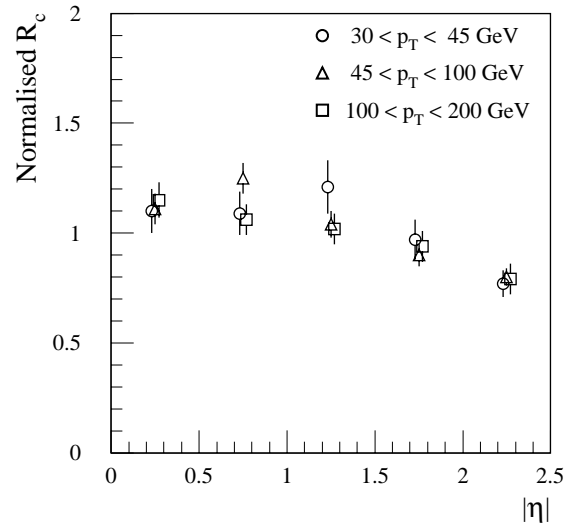

Figure 10-26 *c*-jet rejection factors, normalised to their mean values, as function of pseudorapidity for jets in different p_T ranges.

Table 10-2 *u*-jet rejection for 50% *b*-jet tagging efficiency, for different p_T and $|\eta|$ intervals. In each cell, the first number refers to xKalman and the second to iPatRec. Errors are statistical only.

	all η	$0 < |\eta| < 0.5$	$0.5 < |\eta| < 1$	$1 < |\eta| < 1.5$	$1.5 < |\eta| < 2$	$2 < |\eta| < 2.5$
all p_T	231 ± 16	411 ± 76	442 ± 88	301 ± 52	110 ± 13	39 ± 3
	237 ± 17	375 ± 66	346 ± 60	246 ± 38	144 ± 20	83 ± 10
$15 < p_T < 30$	21 ± 4	-	-	-	-	-
	45 ± 14	-	-	-	-	-
$30 < p_T < 45$	164 ± 29	267 ± 137	192 ± 81	298 ± 134	153 ± 60	14 ± 2
	277 ± 66	392 ± 200	301 ± 156	433 ± 210	140 ± 47	54 ± 13
$45 < p_T < 60$	269 ± 54	369 ± 184	821 ± 385	334 ± 150	153 ± 51	20 ± 3
	331 ± 75	671 ± 384	519 ± 290	372 ± 186	143 ± 50	110 ± 36
$60 < p_T < 100$	467 ± 95	646 ± 311	554 ± 257	568 ± 277	311 ± 124	60 ± 11
	581 ± 132	610 ± 295	432 ± 176	720 ± 358	468 ± 230	398 ± 185
$100 < p_T < 200$	413 ± 74	889 ± 444	714 ± 318	482 ± 200	184 ± 54	101 ± 26
	368 ± 62	699 ± 310	633 ± 284	546 ± 243	168 ± 47	111 ± 31
$p_T > 200$	129 ± 14	188 ± 44	196 ± 51	116 ± 26	61 ± 14	94 ± 36
	103 ± 10	160 ± 34	127 ± 27	76 ± 14	73 ± 18	80 ± 29

The jet rejections $R_j(p_T, \eta)$ are functions of p_T and $|\eta|$. The same rejections in the relevant bins of p_T and $|\eta|$ were obtained using fully simulated $t\bar{t}$ events. The extrapolation to other processes is therefore justified.

10.3 Soft electrons

The most powerful way to tag b-jets is to look for tracks with significant impact parameter – this is discussed in the previous section. However, the tagging of soft leptons (electrons and muons) will provide a valuable complement to this. In this section, the potential to tag b-jets using electrons from semileptonic decays of b-quarks while rejecting gluon and light-quark jet background is evaluated. Only jets from decays of Higgs bosons with $m_H = 100$ GeV were considered for this study. Only xKalman was used for the pattern recognition in the Inner Detector.

10.3.1 Event characteristics

The branching ratio for the $b \rightarrow e$ decay, $BR(b \rightarrow e)$, is calculated taking into account the direct semileptonic decay of B mesons ($b \rightarrow e$) and the cascade decay of B to D mesons and D semileptonically to e ($b \rightarrow c \rightarrow e$) and normalised to the number of reconstructible b-labelled jets. Only electrons inside b-jets with $|\eta^e| < 2.5$ and b-jets with $|\eta^b| < 2.5$, $p_T{}^b > 15$ GeV were taken into account. For 100 GeV Higgs events, taking a cone around the b-quark direction of half-width $\Delta R < 0.4$, this branching ratio is 15.5% for electrons with $p_T{}^e > 1$ GeV and 13.8% for electrons with $p_T{}^e > 2$ GeV.

The signal (irreducible) electrons in b-jets come from the following processes:

- the direct semileptonic decay of B mesons with an electron in the final state ($b \rightarrow e$);

- the cascade decay of b-hadrons to c-hadrons which later decay semi-leptonically with an electron in the final state ($b \rightarrow c \rightarrow e$);

- the leptonic decays of J/ψ's coming from b decays;

- the decays of b-hadrons to τ leptons which later decay into electrons.

These electrons are also expected in the background jets, for example in gluon jets, because of their possible content of heavy quarks. The contamination of the $H \rightarrow u\bar{u}$ sample with heavy flavour quarks from initial state radiation gluon splitting is below 2% for c-quarks and below 0.3% for b-quarks.

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

- photon conversions which occur in the detector;

- π^0 Dalitz decays;

- semileptonic decays in the hadronic cascade.

While the signal electrons have very similar kinematic features in the signal and background jets (originating in both cases from the decays of heavy quarks), the kinematics of the background electrons is very different. Another important source of background are hadrons misidentified as electrons.

The fraction of reconstructed jets containing true electron tracks with $p_T > 2$ GeV and within $\Delta R < 0.4$ from the jet axis is given in Table 10-3. Results in the Table are given after the full detector simulation of the events, so they take into account also electrons produced in the interac-

Table 10-3 Fraction of jets containing electrons of a given origin (only electrons with $p_T > 2$ GeV and within $\Delta R < 0.4$ from the jet axis are considered). The numbers are given after the full detector simulation.

Jet type	Fraction of jets containing electrons (%)						
	from γ	from π^0	from B	from D	from other sources	irred. from B and D	reducible
b-jets	3.8	1.8	15.3	10.9	1.1	26.2	6.7
g-jets	3.8	0.7	0.1	0.4	0.1	0.5	4.6
c-jets	3.8	0.9	-	6.1	0.1	6.1	4.8
u-jets	3.8	1.0	-	-	0.1	-	4.9

tions with the detector material. The fraction of jets containing reducible electron tracks is of the order of 5% for the background jets and is dominated by electron tracks from photon conversion (~79%). It increases almost by 50% if the p_T threshold is lowered to 1 GeV. Lowering the threshold, the fraction of b-jets containing irreducible electron tracks increases by at most few percent.

10.3.2 Electron identification

Tracks reconstructed in the Inner Detector were selected applying the same cuts ('standard cuts') as for the vertexing algorithm (see Section 10.2.2), including the removal of electrons from reconstructed conversions. Variables for electron identification were then constructed using information from the Inner Detector and from the EM Calorimeter, as described in Section 7.3 and, in more detail, in Ref. [10-8] and [10-9]. Two additional variables which characterise global features of any track coming from a b-jet were considered, d_0 (the transverse impact parameter) and p_{TJ} (the transverse momentum relative to the jet axis). One expects that the charged tracks from B decays would have significant impact parameter (long lifetime) and also large p_T with respect to the jet axis (decay of a heavy object).

The d_0 and p_{TJ} distributions are shown in Figure 10-27 for signal electron tracks from B and D mesons in b-jets, and for other charged tracks in non-b jets. The rejection power of the d_0 variable is much higher than that of p_{TJ} (notice the logarithmic scale for d_0). One should be aware however that the d_0 variable is effectively the same as the variable used by the vertexing algorithm.

10.3.2.1 Discriminating function

The discriminating function D_{track} is defined for each track as: $D_{track} = \log\dfrac{\Pi_i P_e(x_i)}{\Pi_i P_h(x_i)}$,
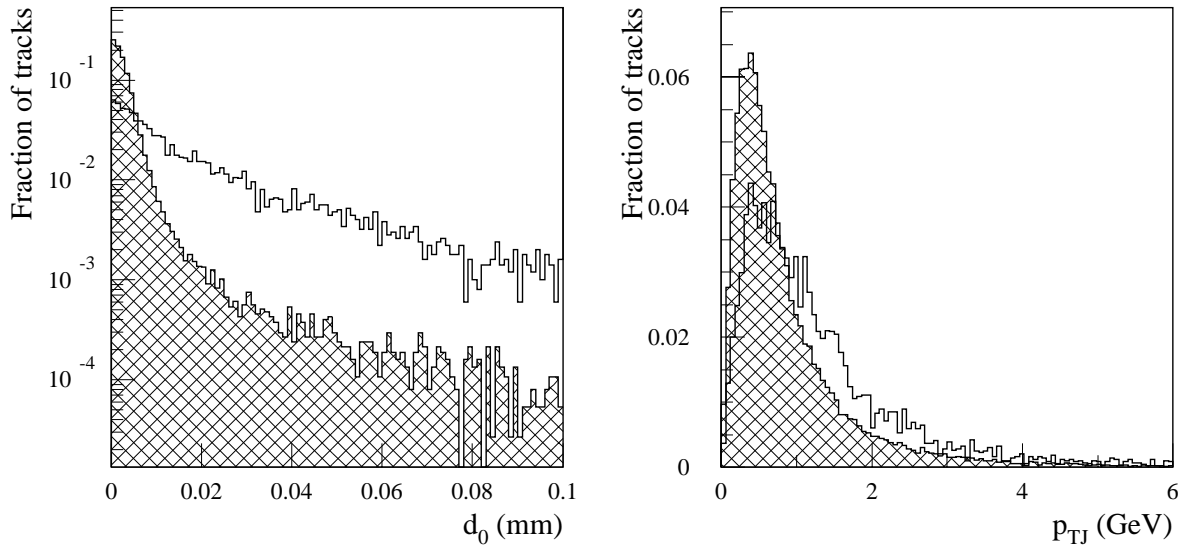
Figure 10-27 Probability density functions of the variables identifying tracks in *b*-jet. Open – signal electron tracks in *b*-jets, shaded – all tracks in non-b jets.

where x_i denotes the value of the *i*-th variable for this track, $P_e(x_i)$ is the probability that the track is a signal electron. $P_h(x_i)$ is the probability that the track is a non-signal hadron. Tracks that fail the track-quality cuts are assigned the value D_{track} = -40.

Higher rejection power can be achieved by the algorithm if, for tracks with transverse momentum below 5 GeV, the value of the discriminating function is rescaled as follows: $D_{track} \rightarrow D_{track} + 1.5 \times p_T$ - 7.5, where p_T is the transverse momentum of the track reconstructed in the Inner Detector. This p_T weighting is effective in rejecting the low-p_T background electron tracks from Dalitz decays and from γ-conversions.

Figure 10-28 shows the distribution of the discriminating function D_{track} for various types of tracks. Applying a threshold on the value of the D_{track} function a good separation of signal electron tracks from *b*-jets and other tracks can be achieved.

Figure 10-28 Distributions of the discriminating function D_{track} for signal electron tracks in *b*-jets from *B* and *D*-meson decays (solid histogram), compared to background electron tracks from γ-conversions or Dalitz decays in non-*b* jets (dotted) and pion tracks from non-*b* jets (dashed).

The identification of a track as a signal electron track is based on the value of D_{track} assigned to a given track. Those for which D_{track} is below threshold (typically D_{track}^{thr} ~ 5-7) are rejected as signal electron track candidates. To quantify the performance of the identification procedure, the following efficiency and rejection factors are defined: the electron identification efficiency ε_e, the pion rejection R_π and the conversion rejection R_{conv}.
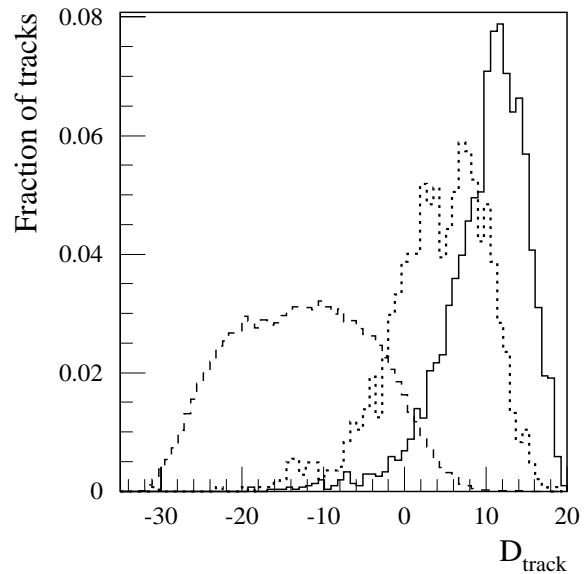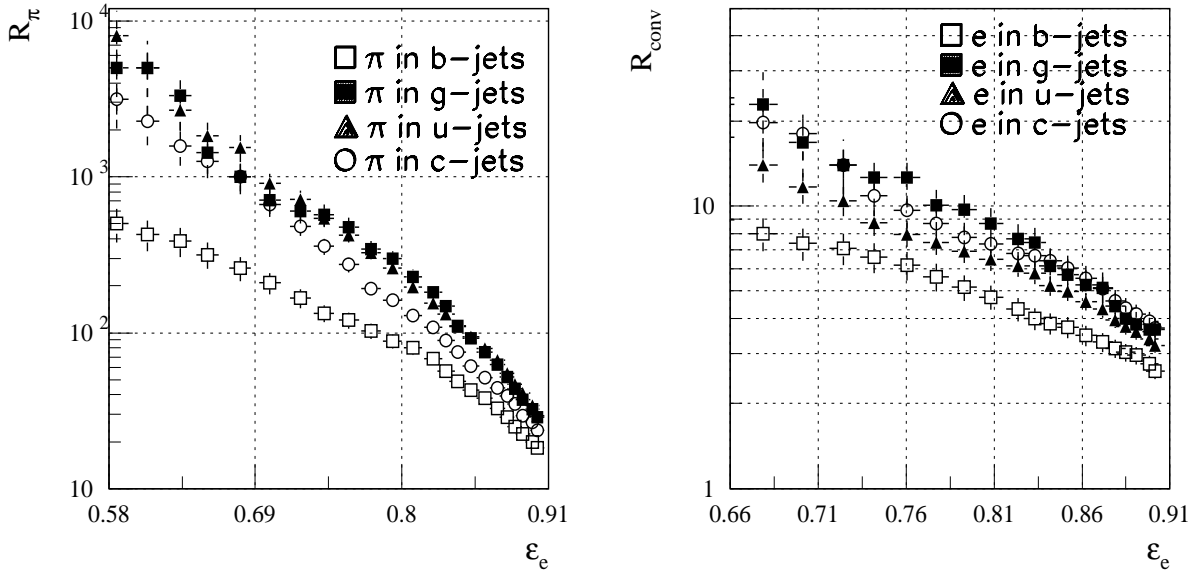
Figure 10-29 R_π (left), and R_{conv} (right) rejection factors for tracks in various jet types as a function of the efficiency of identifying signal electron tracks from B and D decays in b-jets. 'e in b-jets', 'e in g-jets', *etc.* denotes electron tracks from γ-conversions or Dalitz decays in b-, g-, u- or c-jets. Only statistical errors are shown.

Figure 10-29 shows the rejections R_π and R_{conv} as functions of the signal electron identification efficiency. As expected the worst rejection factors are obtained for tracks originating from b-jets, both for pions and also for electrons from γ-conversions and from Dalitz decays. While the two rejection factors are almost the same for b-jets, they increase significantly for background jets. This can be explained by the fact that the d_0 variable has no discriminating power in this case as several tracks are in fact originating from B-meson decays.

10.3.3 Jet tagging procedure

The jet tagging procedure is based on the electron track identification procedure. The algorithm is constructed as follows: for each track in the jet the value of the discriminating function D_{track} is calculated, then the track with the highest value of D_{track} is chosen and its value of the discriminating function is assigned as the value of the discriminating function for the jet D_{jet}; jets with $D_{jet} > D_{jet}^{thr}$ are tagged as b-jets.

Figure 10-30 shows the distributions of D_{jet} for signal jets and for background jets: u-, c- and g-jets. The D_{jet} distribution for b-jets is shifted with respect to the distribution for u-, c-, and g-jets. Some fraction of events for each jet type has jets with $D_{jet} = -40$ which corresponds to jets for which none of the tracks passed the quality selection criteria.

10.3.4 *b*-tagging results

In order to improve the rejection power, the probability density distribution functions were calculated separately for several p_T and $|\eta|$ bins. Figure 10-31 shows the jet rejection factors for various types of jets as a function of the efficiency of the b-tagging algorithm. As expected, the highest rejection is obtained for u-jets, intermediate for gluon jets, and the lowest for c-jets.
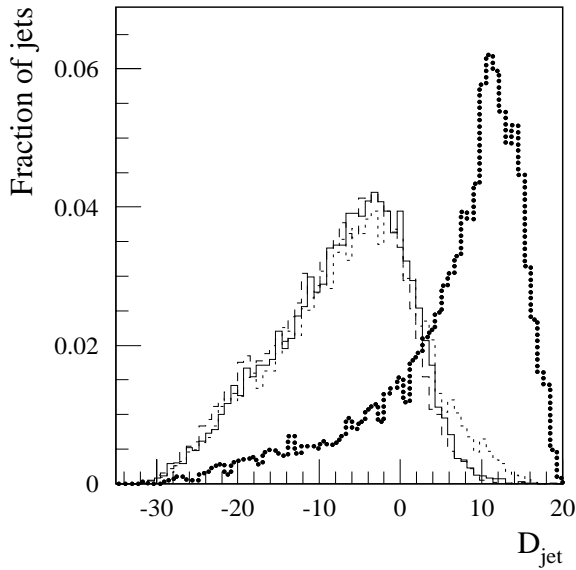
Figure 10-30 Probability density functions of the values of D_{jet} for *b*-jets (bold-dotted), compared to gluon jets (solid), *u*-jets (dashed), *c*-jets (dotted).
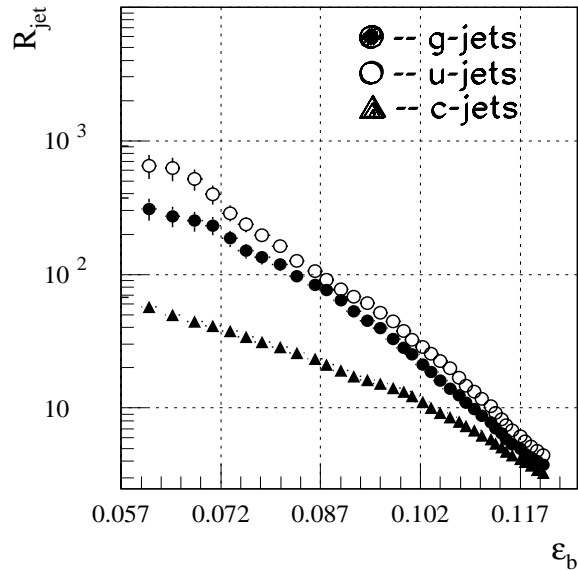

Figure 10-31 Jet rejection factor R_{jet} as a function of the efficiency of the *b*-tagging algorithm for various jet types.

Table 10-4 shows the jet rejection factors for a *b*-tagging efficiency of 7.2%, which corresponds to ~50% efficiency for *b*-jets with an electron. The reason for the lower rejection factors with respect to previous studies [10-8] is mainly the extension of this study to the full rapidity range. As particles at larger $|\eta|$ cross more material, the electron-pion separation power decreases rather rapidly: Figure 10-32 shows the pseudorapidity dependence of the rejection factors. Decreasing the threshold on p_T causes the jet rejection to decrease, as not only the charged track multiplicity, but also the fraction of background jets with true electron tracks in the jet is significantly higher.
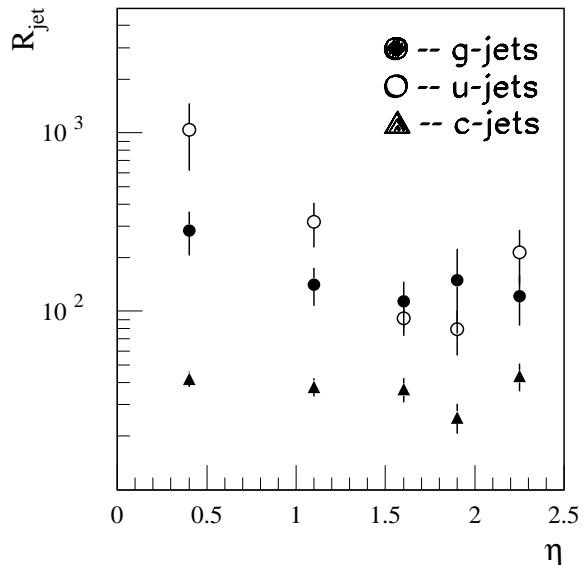
Table 10-5 shows the fraction of jets tagged by a given track for ε_b = 7.2%, which corresponds to a 50% tagging efficiency for *b*-jets with a true electron.

Most of the jets are tagged by a true electron track, independently of the jet type. This indicates that the electron identification procedure works with high efficiency and good purity. Most of the electron tracks which tag the jets, except for *u*-jets, are signal electron tracks (they come from *B* or *D* decays). For the *b*-jet sample this type of track accounts for 94% of all tagged jets. Noting that in the original sig-

Figure 10-32 Jet rejection factors as a function of pseudorapidity for a fixed (7.2%) *b*-tagging efficiency.

Table 10-4 Jet rejection factors for a fixed (7.2%) *b*-tagging efficiency using XKalman. The p_T threshold is 2 GeV.

R_u	338 ± 50
R_g	202 ± 30
R_c	40 ± 2

nal sample the percentage of jets with such electrons is ~90% (see Table 10-3), then one concludes that the tagging enriches the sample. This effect is even stronger for the *g*- and *c*-jets, where the fraction of jets containing electrons from *B* and *D* mesons is greatly increased by the *b*-tagging algorithm. The sample of *u*-jets does not contain electrons from *B* and *D*, but these jets are also tagged mainly (60%) by the electron tracks from γ-conversions and Dalitz decays or from other sources. In the higher pseudorapidity region ($|\eta| > 1$), pions become a major source of background.

Table 10-5 Fraction of jets tagged by a specified track for $\varepsilon_b = 7.2\%$. Only tracks with $p_T > 2$ GeV are used. The '*e* from γ' column denotes electron tracks from γ-conversions and Dalitz decays.

Jet type	Fraction of jets tagged by a specified track (%)						
	all *e*'s	*e* from *B*	*e* from *D*	*e* from γ	other *e*	π	others
b	99.4	68.7	26.0	1.4	3.3	0.4	0.2
g	82.2	17.8	24.4	37.8	2.2	15.6	2.2
c	91.9	0	86.9	4.6	0.4	5.8	2.3
u	65.2	0	0	65.2	0	26.1	8.7

10.4 Soft muons

In principle, muons are just as good as electrons for the identification of the semileptonic decays of heavy quarks. In practice, although the efficiency for detecting muons in the Inner Detector is very high even at very low p_T, they can be identified as such only if they have enough momentum to cross the whole thickness of the calorimeters. The mean energy loss at central rapidity is ~3 GeV, therefore this is the effective threshold for muon identification in the barrel part of the Muon System. Muons with $p_T > 6$ GeV cross the middle muon chamber station and can be found as tracks in the muon detector; this effective threshold decreases to 2 GeV for muons crossing the end-cap toroids. The energy deposit in the last segment of the Hadronic Tile Calorimeter can be used to discriminate muons with $p_T > 2$ GeV in the barrel rapidity range $|\eta| < 1.7$.

10.4.1 Event characteristics

The branching ratio for the $b \to \mu$ decay, $BR(b \to \mu)$, was calculated taking into account the direct semileptonic decay of *B* mesons ($b \to \mu$) and the cascade decay of *B* to *D* mesons and *D* semileptonically to μ ($b \to c \to \mu$) and normalised to the number of reconstructible *b*-labelled jets. In both cases only muons inside *b*-jets with $|\eta^\mu| < 2.5$ and *b*-jets with $|\eta^b| < 2.5$, $p_T^b > 15$ GeV were taken into account.

Table 10-6 specifies the branching ratios for two values of the p_T^μ threshold ($p_T^\mu > 3$ GeV and $p_T^\mu > 6$ GeV). As shown in Table 10-6, if muons below the threshold for muon identification in the Muon System can be successfully tagged (using the information from the last segment of the Hadronic Tile Calorimeter), they can add a considerable fraction to the *b*-tagging efficiency for lower-p_T jets.

Table 10-6 The $b \to \mu$ (direct and the cascade decay) branching ratios for $p_T > 3$ and 6 GeV and for $m_H = 100$ and 400 GeV.

	m_H = 100 GeV	m_H = 400 GeV
$p_T^\mu > 3$ GeV	10.0%	15.7%
$p_T^\mu > 6$ GeV	6.5%	13.3%

Signal muons have clearly the same properties as the signal electrons used for the soft electron tagging studies in the previous section. The background consists of:

- muons from π and *K* decays;

- misidentified particles in jets containing real muons, as a result of incorrect matching between the tracks reconstructed in the Inner Detector and the Muon System;

- (for low-p_T muons) particles extrapolated to a region of the Hadronic Tile Calorimeter with an energy deposit compatible with a muon's.

10.4.2 Muon identification

The Muon System performance was parametrised and applied to events which were fully simulated and reconstructed in the Inner Detector and calorimeters, since a full simulation was not possible due to the delayed availability of the Muon System geometry.

The procedure to reconstruct muons in the Muon System and to match them to Inner Detector tracks is described in detail in Section 8.1. Applying this procedure to a sample of *b*-jets where the *B* mesons were forced to decay semileptonically, the muon reconstruction efficiency and hadron misidentification rates were parametrised as functions of $|\eta|$ and p_T as shown in Section 8.4 and in Figures 8-22 and 8-23. Then a muon identification probability was assigned to each track reconstructed in the Inner Detector, depending on the track type and its parameters.

Muons with $p_T > 2$ GeV in the range $|\eta| < 1.7$ can be identified through their energy deposit in the last segment of the Hadronic Tile Calorimeter. If no muon was found in the Muon System and matched to an Inner Detector track for a given jet, the method described in Section 8.4 was used to assign a 'quality factor' to particles in the range $2 < p_T < 6$ GeV which corresponded to an energy deposit in the last Hadronic Tile Calorimeter segment comprised in the band shown in Figures 8-24 and 8-25.

10.4.3 Jet tagging procedure

The jet tagging procedure selected for each jet the muon which had the highest probability of coming from a semileptonic *B* decay. The impact parameter significance, the transverse momentum relative to the jet axis (p_{TJ}) and the muon energy fraction relative to the jet energy (E_{fr}) are all good discriminating variables, and muons from semileptonic *B* decays tend to have larger values of these variables compared to background muons (see the distributions of p_{TJ} and E_{fr} for muons in *b*- and *u*-jets in Figures 10-33 and 10-34). The logarithm of the product of these
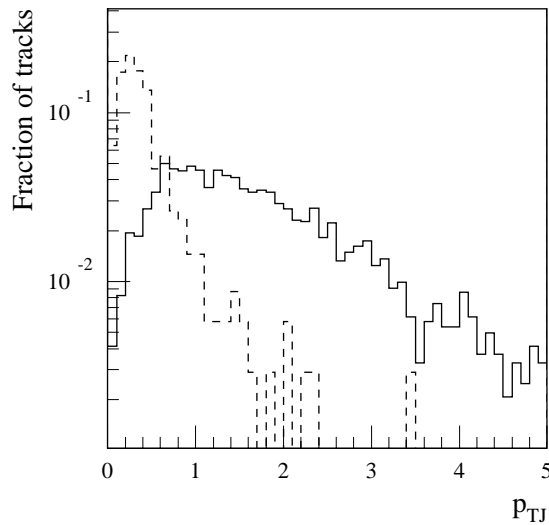
Figure 10-33 Transverse momentum relative to the jet axis for muons from *B* and *D* decays in *b*-jets (solid line) and any muon in *u*-jets (dashed line).
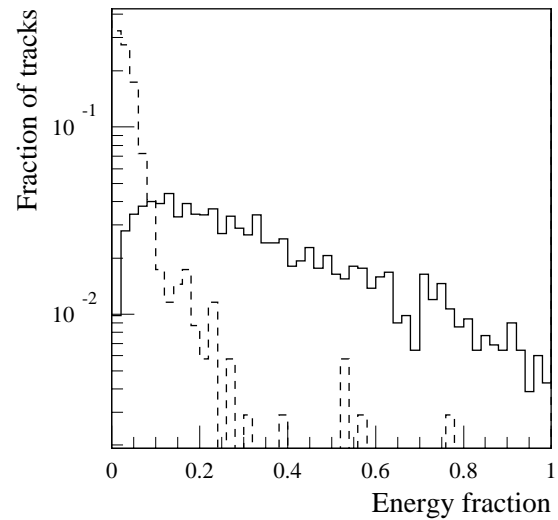
Figure 10-34 Energy fraction relative to the jet energy (as reconstructed by the calorimeters) for muons from *B* and *D* decays in *b*-jets (solid line) and any muon in *u*-jets (dashed line).

three variables was used as the discriminating function D_{track}. The largest value of D_{track} for all tracks in a jet was then used to calculate the jet weights and rejection factors for *u*-, *g*- and *c*-jets. Figure 10-35 shows the jet weights for *b*-jets compared to background jets.

10.4.4 *b*-tagging results

The jet rejection curves as function of the *b*-jet efficiency are shown in Figure 10-36. As can be immediately noted, the much better signal-to-background ratio of the muons in the Muon System sample partially compensates the loss of efficiency for low-p_T central jets; the contribution of the muons recognised in the Hadronic Tile Calorimeter is significant only in the low-p_T range. For the sake of comparison with the soft electron analysis, Table 10-7 shows the jet rejection factors for a fixed (7.2%) *b*-jet efficiency, separately for jets produced in the decays of 100 and 400 GeV Higgs.

Figure 10-37 shows the pseudorapidity dependence of the rejection factors for fixed *b*-jet efficiency. As expected, the lower p_T threshold for muons recognised in the Hadronic Tile Calorimeter contributes to improve the rejection power in the central pseudorapidity region.
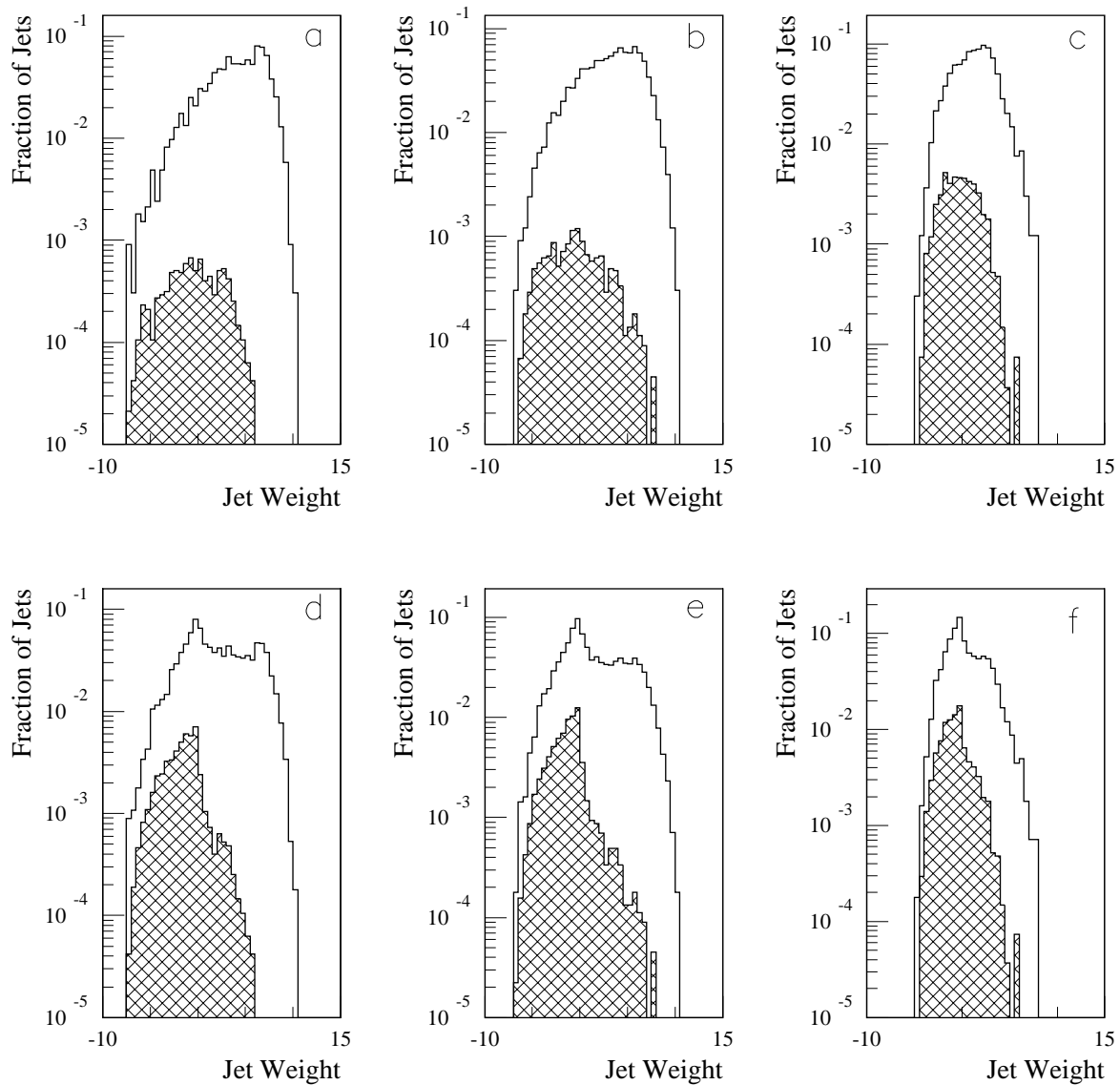
Figure 10-35 Probability distribution functions of jet weights for muon *b*-tagging. Top: muons recognised in the Muon System only, bottom: muons recognised in the Muon System or the Hadronic Tile Calorimeter. Open histograms: *b*-jets, shaded histograms: background jets. a) and d): *b*-jets compared to *u*-jets. b) and e): *b*-jets compared to gluon jets. c) and f): *b*-jets compared to *c*-jets. Jets without muons are taken into account for the normalisation of the histograms.

10.5 Robustness of performance

10.5.1 Effects of high-luminosity pile-up

A large fraction of the Higgs events used for the *b*-tagging studies were processed with the addition of minimum bias events to simulate the high luminosity environment. A random number (on average 24) of minimum bias events were added to all detectors; in addition, on average
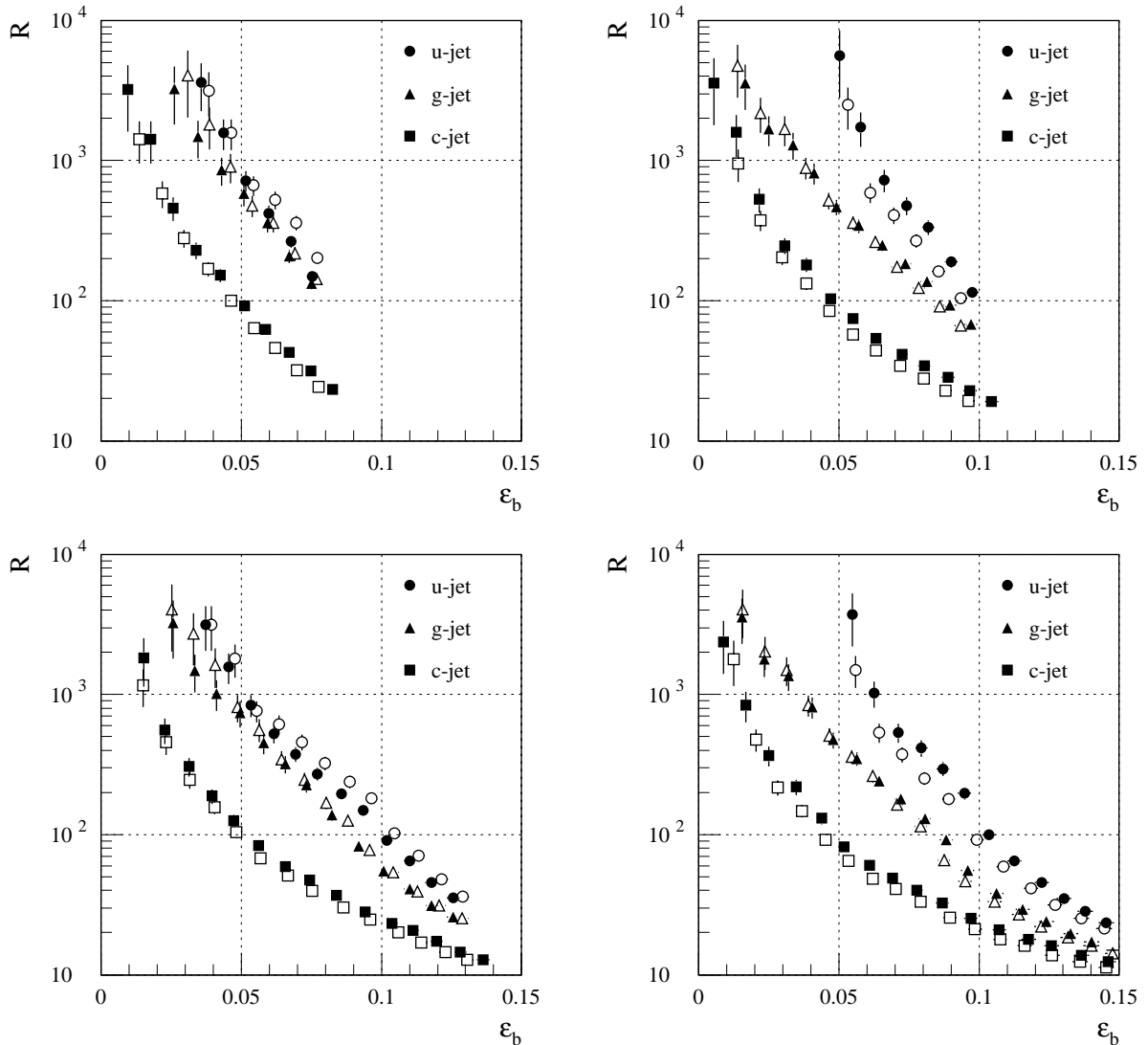
Figure 10-36 Rejection factors as a function of *b*-jet efficiency for the muon tags. Top: only muons recognised in the Muon System. Bottom: all muons recognised in the Muon System or the Hadronic Tile Calorimeter. Left: $m_H = 100$ GeV. Right: $m_H = 400$ GeV. Full symbols: xKalman; open symbols: iPatRec.

Table 10-7 Jet rejection factors for a fixed (7.2%) *b*-tagging efficiency.

	Muon System only				Muon System and Hadronic Tile Calorimeter			
	$m_H = 100$ GeV		$m_H = 400$ GeV		$m_H = 100$ GeV		$m_H = 400$ GeV	
	xKalman	iPatRec	xKalman	iPatRec	xKalman	iPatRec	xKalman	iPatRec
R_u	189 ± 16	299 ± 33	510 ± 77	347 ± 43	334 ± 38	455 ± 61	513 ± 77	379 ± 49
R_g	169 ± 17	182 ± 19	198 ± 16	169 ± 13	250 ± 31	255 ± 32	181 ± 14	159 ± 12
R_c	35 ± 2	30 ± 1	43 ± 2	34 ± 2	50 ± 3	44 ± 3	46 ± 3	40 ± 2

eight events were added to the TRT to simulate its long sensitive time. The pile-up and pulse shaping in the calorimeters was simulated as described in Section 2.3. No pile-up simulation was performed for the Muon System elements as the simulation code was not available in time.
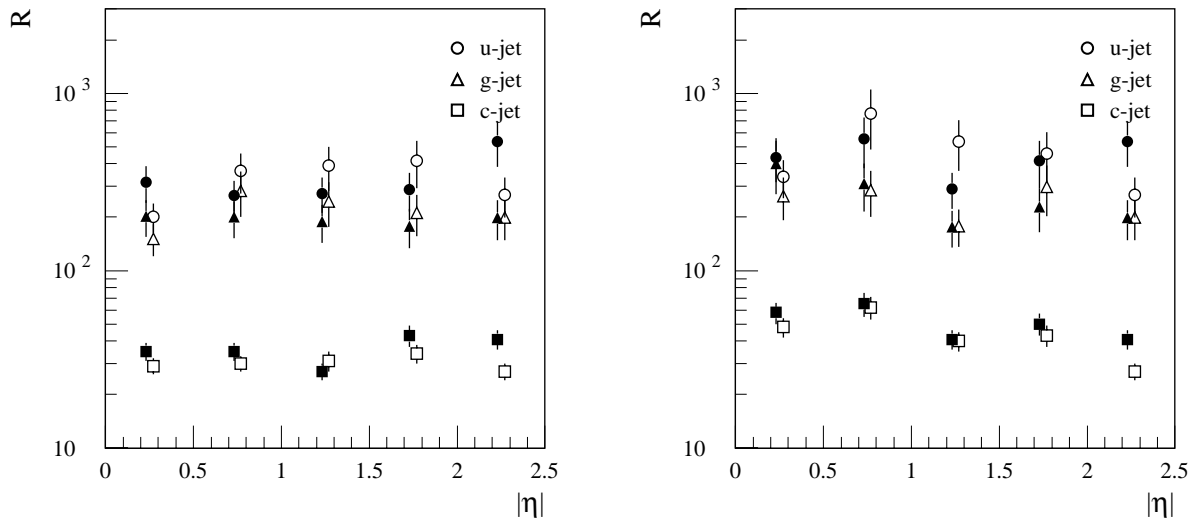
Figure 10-37 Rejection factors as a function of pseudorapidity for 7.2% *b*-jet efficiency for the muon tags (m_H = 100 GeV). Left: only muons recognised in the Muon System. Right: all muons recognised in the Muon System or the Hadronic Tile Calorimeter. Full symbols: xKalman; open symbols: iPatRec.

The same reconstruction and analysis software was run on these events, as for events without pile-up. In order to save running time, and as it was shown by all analyses presented so far that the performance of xKalman and iPatRec for this type of studies are equivalent, only the faster iPatRec algorithm was used. Figures 10-38 to 10-40 show the primary track efficiency and the rates of secondary tracks and fakes for pile-up events.
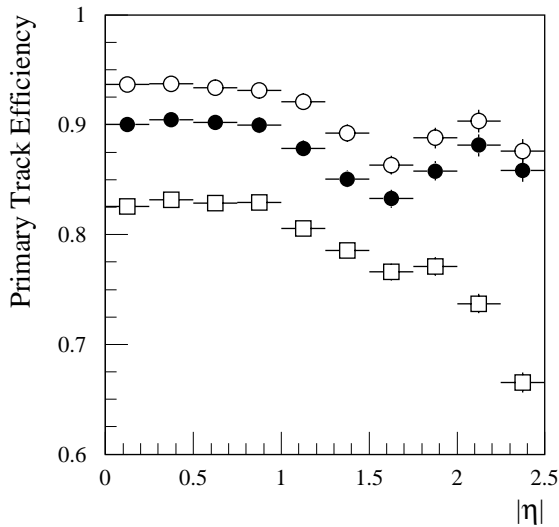

Figure 10-38 Primary track efficiency for iPatRec as a function of pseudorapidity without cuts (open circles), after applying standard cuts (closed circles) and after applying quality cuts (open squares).
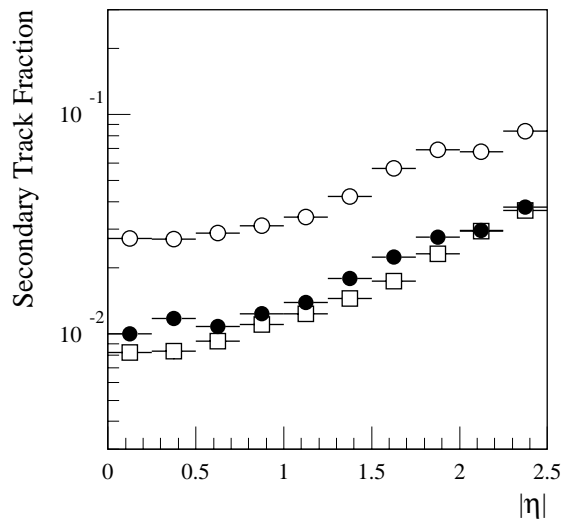
Figure 10-39   Fraction of secondary tracks for iPatRec as a function of pseudorapidity without cuts (open circles), after applying standard cuts (closed circles) and after applying quality cuts (open squares).

It can be noted that for standard cuts the pattern recognition performance is comparable to the performance without pile-up. The quality cuts (especially the cut on the number of shared hits) affect adversely the track-finding efficiency but remove a large fractions of tracks which did not originate from the main Higgs event.

Figure 10-40 Fraction of fake tracks as a function of pseudorapidity without cuts (open circles), after applying standard cuts (closed circles) and after applying quality cuts (open squares) (iPatRec).
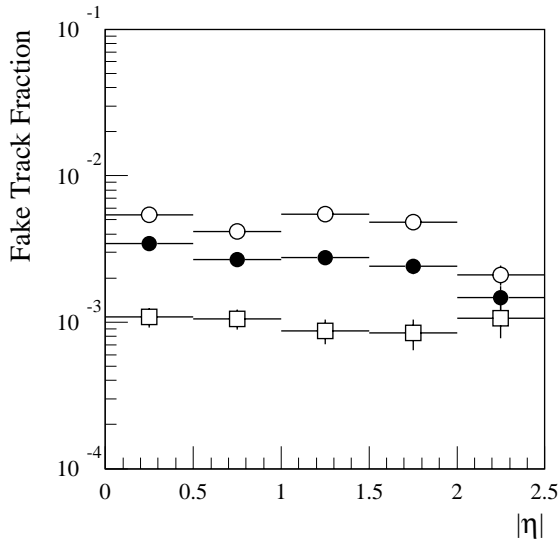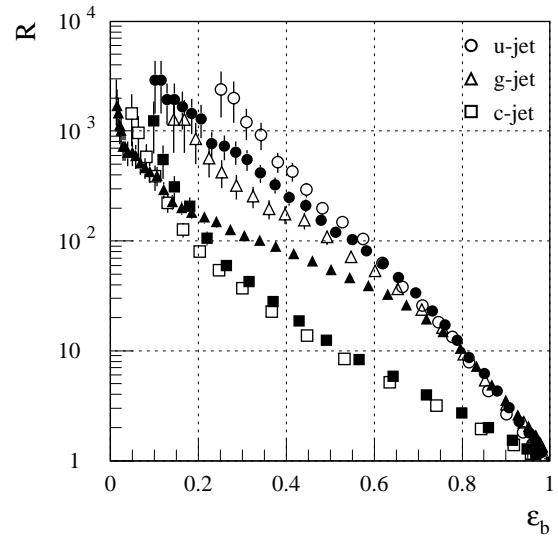
Figure 10-41 Jet rejection factors for the vertex b-tagging method, with high-luminosity pile-up. Open symbols: $m_H = 100$ GeV, full symbols: $m_H = 400$ GeV.

The z position of the primary vertex can be reconstructed with the same resolution (~35 μm) but the fraction of events in the tails of the distribution increases from 9% to 14%.

Table 10-8 Jet rejection factors for different b-tagging methods, with high-luminosity pile-up. Rejections are calculated for $\varepsilon_b = 50\%$ for the vertex method and 7.2% for the lepton methods.

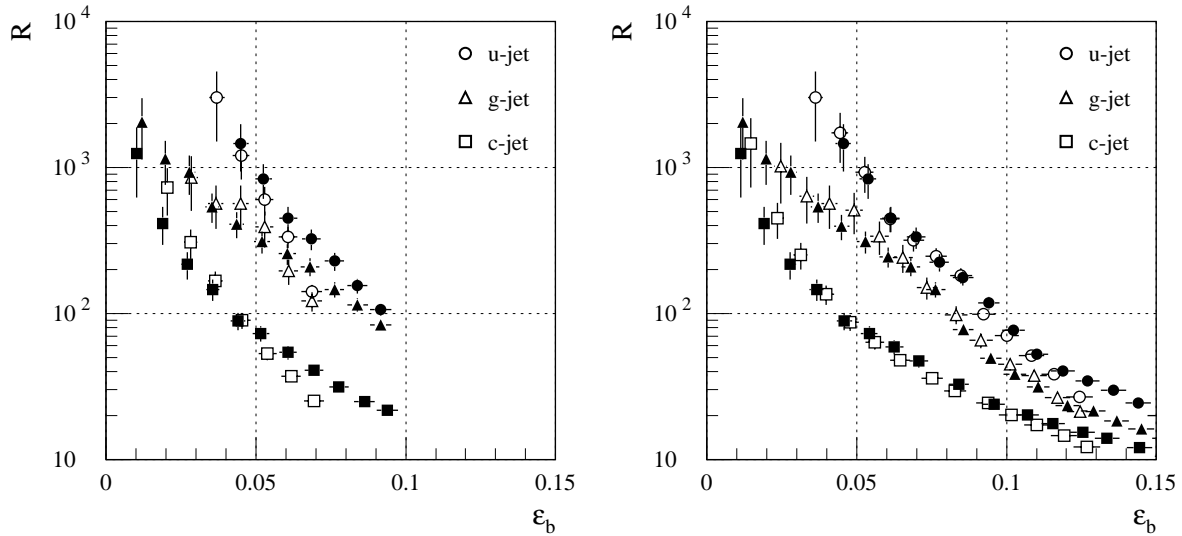|	R_u		R_g		R_c	
	$m_H = 100$	$m_H = 400$	$m_H = 100$	$m_H = 400$	$m_H = 100$	$m_H = 400$
Vertex	178 ± 22	133 ± 14	104 ± 15	55 ± 4	10.5 ± 0.4	12.1 ± 0.6
Electron	118 ± 11	74 ± 6	115 ± 17	135 ± 16	34 ± 3	31 ± 2
Muon	223 ± 30	284 ± 44	170 ± 31	168 ± 22	32 ± 2	37 ± 3
Muon + Tile	289 ± 45	282 ± 44	157 ± 27	183 ± 24	40 ± 3	46 ± 4

Figure 10-42 Rejection factors as function of b-jet efficiency for the muon tags, with high-luminosity pile-up. Left: only muons recognised in the Muon System are used; right: all muons recognised in the Muon System or in the Hadronic Tile Calorimeter are used. Open symbols: $m_H = 100$ GeV, full symbols: $m_H = 400$ GeV.

Table 10-8 and Figures 10-41 to 10-43 present the rejection factors that can be obtained, using the vertex and lepton tags. The rejections are somewhat degraded with respect to the zero-luminosity case (see Tables 10-1, 10-4 and 10-7) but still acceptable and useful.

Figure 10-44 shows the background rejections, for 50% b-tagging efficiency, obtained using the vertex method only, as functions of $|\eta|$ and p_T. Comparing to the corresponding Figures 10-23 and 10-24, it can be noted that the loss in overall performance is effectively due to a slower increase of the performance for jets with $p_T < 100$ GeV; above this value the performance is not affected by the presence of pile-up.
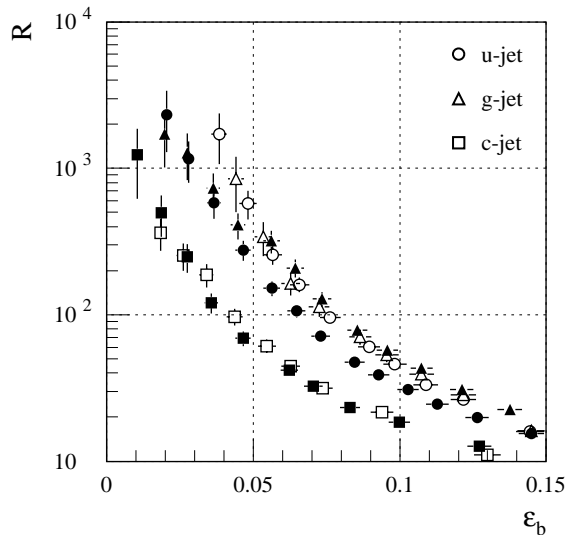

Figure 10-43 Rejection factors as a function of b-jet efficiency for the electron tags, with high-luminosity pile-up. Open symbols: $m_H = 100$ GeV, full symbols: $m_H = 400$ GeV.

10.5.2 Effects of reduced detector efficiency and missing detector layers

The effects of reduced detector efficiency and/or of missing detector layers were studied in detail in Ref. [10-10]. Those studies used the same Inner Detector geometry as used for the present study. The reconstruction and analysis software were less developed at that time, which produced a worse b-tagging performance than can be achieved now, but it is believed that the relative changes in performance as function of the detector efficiency still apply. Here only the main results are summarised.
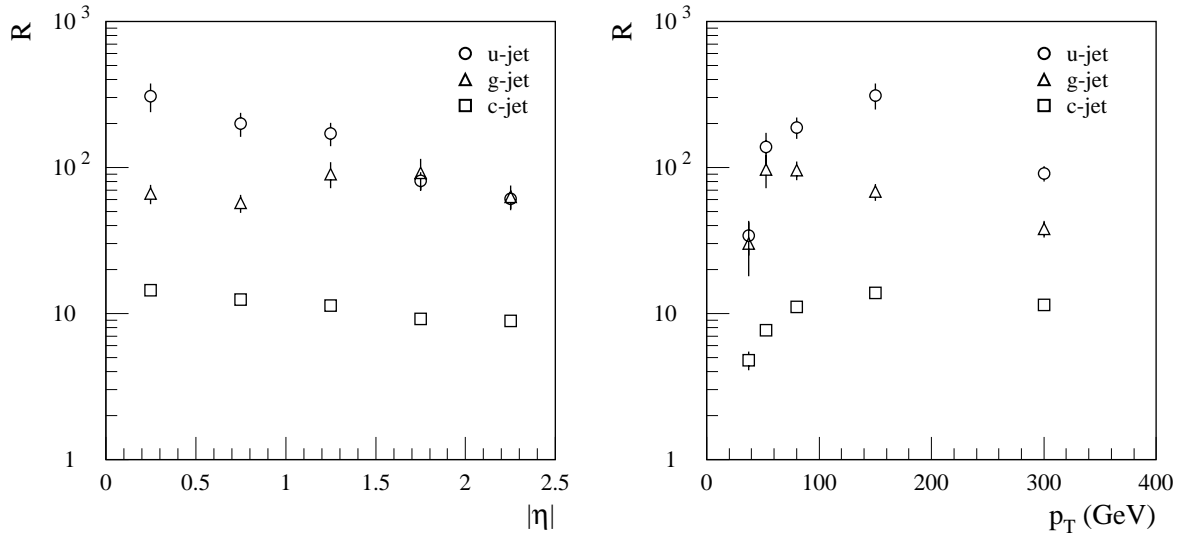
Figure 10-44 Background rejections as a function of jet $|\eta|$ and p_T for $\varepsilon_b = 50\%$, with high-luminosity pile-up. Vertex method only.

Reducing the efficiency of the Pixel and SCT detectors from the default random 97% to 90% reduced the u-jet rejection power to 76% of its original value, independent of the pattern recognition program and of the track selection cuts.

The loss, or removal, of the middle or outer Pixel layer reduced the u-jet rejection power to 70% of its original value, independent of the pattern recognition program and of the track selection cuts. Similarly, the loss, or removal, of an SCT layer reduced the u-jet rejection power to 90% of its original value.

The effect of the loss of the B-layer was studied in the Inner Detector TDR: in this case the u-jet rejection power is reduced to 40% of its original value.

10.5.3 Evolution of the Inner Detector layout

The current design of the Inner Detector layout differs in several details from the geometry used for the present studies (see Section 3.2.1). The main impact on the b-tagging performance is that of the changes in Pixel Detector geometry: slightly larger B-layer radius, slightly smaller outer layer radius, re-arranged disk regions. All these changes result in small degradations of the performance. The cumulative effect is to reduce the u-jet rejection power to 90% of its original value; if hardware constraints forced the adoption of 400 μm long pixels in the B-layer, there would be a further 10% performance loss.

10.6 Combined *b*-tagging performance

The algorithms described above for the vertex tagging and the soft lepton tagging all calculate a 'jet weight' which is used as a measure to discriminate b-jets from other jets. There is more than one way to combine these weights into a single combined variable. As the rejection power of the vertex method is superior to that of the lepton methods and the lepton tags are efficient only for a few percent of the b-jets, the following method was used to select the 'tagged' events: for each

value of the cut on the vertex weight, the cut on the lepton weight is chosen so as to optimise the global performance. The combined performance was then compared to the performance of the vertex method for the same *b*-tagging efficiency or for the same rejection.

Table 10-9 Jet rejection for different combinations of *b*-tagging methods, for 50% *b*-jet efficiency, compared to the vertex method. In each cell, the first column refers to xKalman and the second to iPatRec; the first (second) line is for m_H = 100 (400) GeV.

	R_u	R_g	R_c
Vertex	326 / 391	135 / 148	13.6 / 11.7
only	126 / 183	59 / 58	13.3 / 13.4
Vertex +	329 / 398	148 / 154	13.8 / 12.0
Electron	127 / 185	69 / 58	13.3 / 13.5
Vertex +	326 / 391	135 / 150	13.7 / 11.9
Muon	128 / 187	60 / 58	13.4 / 13.5
Vertex +	326 / 391	135 / 150	13.7 / 11.9
Muon + Tile	128 / 187	60 / 58	13.4 / 13.5

Table 10-9 shows the increase in jet rejection factors, for 50% *b*-tagging efficiency, for different combinations of *b*-tagging methods. For a constant jet rejection factor (corresponding to 50% *b*-tagging efficiency with the vertex method alone), the increase in efficiency is at most 1%. Given the very high rejection achieved with the vertex method, the combination of the vertex method with the lepton methods, which apply only to a small fraction of events, does not produce much improvement in the performance.

The contribution of the lepton methods can become important if the rejection of the vertex method is decreased with respect to the values shown above. If it would be reduced to ~40% of its value (which would be the case if the *B*-layer were not operational), then the lepton tags would add up to 12% to the rejection at 50% efficiency, or up to 2% efficiency for constant rejection. In addition, they make possible a cross-check of results in order to reduce any systematic errors.

10.7 Impact of *b*-tagging performance on the channel *WH*, $H \rightarrow b\bar{b}$

The detailed understanding of the *b*-tagging performance achieved through the studies described in this chapter allowed the implementation of a realistic parametrisation of the *b*-tagging performance into the fast simulation (see Section 2.5.6) used for the study of this signal and backgrounds. This parametrisation was implemented as described below.

The results for m_H = 100 GeV were adapted to account for the different methods used for labelling jet flavour in the full-simulation studies and in the fast simulation program (the latter has to treat many background sources with varying jet multiplicities and flavours). The *b*-tagging performance was parametrised for jets from *u, d, s, g*, which do not contain *c*-quarks nor *b*-quarks in the final parton shower process, and from charm, including the $|\eta|$ and p_T dependence.

Table 10-10 For an integrated luminosity of 30 fb^{-1} and $m_H = 100$ GeV, expected rates for the signal from Standard Model *WH* associated production followed by $H \to b\bar{b}$ decay and for the various background processes after all selection cuts as a function of the average *b*-tagging efficiency using vertexing; $|\eta|$ and p_T-dependent rejection factors of non-*b* jets are used. The signal-to-background ratio, the ratio R$_{red/irred}$ of reducible to irreducible background and the significance expressed as $S/\sqrt{B}$ are also indicated.

Process	$\varepsilon_b{}^{vert} = 50\%$ $\varepsilon_c{}^{vert} = 8.2\%$ $\varepsilon_j{}^{vert} = 0.43\%$	$\varepsilon_b{}^{vert} = 60\%$ $\varepsilon_c{}^{vert} = 14.4\%$ $\varepsilon_j{}^{vert} = 1.1\%$
$WH,\ H \to b\bar{b}$	294	423
$WZ,\ Z \to b\bar{b}$	422	625
$Wb\bar{b}$	2449	3603
$t\bar{t} \to WWb\bar{b}$	2885	4233
$t\bar{b} \to Wb\bar{b}$	558	755
Wbc	621	1350
$Wc\bar{c}$	698	2045
Wbj	332	1079
Wcj	1082	4796
Wjj	314	2093
$R_{red/irred}$	0.48	1.23
Total bgd (B)	9361	20579
S/B	3.1%	2.1%
$S/\sqrt{B}$	3.04	2.95

The *WH* signal and all the background processes were simulated and reconstructed using the fast simulation, modified to account for the *b*-tagging performance in a statistical way for each reconstructed jet. Table 10-10 shows the expected significance for two values taken from the *b*-tagging efficiency versus rejection curves for an integrated luminosity of 30 fb^{-1} and for $m_H = 100$ GeV. The background processes can be separated into irreducible background containing real *b*-jets and reducible background (predominantly from *W*+jet events), containing non-*b* jets wrongly tagged as *b*-jets. The ratio $R_{red/irred}$ of the total reducible background to the irreducible background increases from a value of 0.5 for a *b*-tagging efficiency of 50% to a larger value of 1.2 for a *b*-tagging efficiency of 60%. At the same time, the signal-to-background ratio decreases by about 30% from 3.1% to 2.1%. For consistency with previous estimates, the tighter trigger threshold was applied for isolated electrons (30 GeV instead of 20 GeV) in Table 10-10. If the lower threshold is applied, the sensitivity increases by 4% only [10-11].

Clearly this channel remains a difficult challenge at the LHC. The uncertainties on the *b*-tagging performance shown above in terms of possible improvements or degradations do not change the results shown in Table 10-10 by more than 10% for *b*-tagging efficiencies below or around 50%, since the dominant background in this case is the irreducible background. The extraction of the rather broad signal peaks expected for invariant masses involving *b*-jets with a signal-to-background ratio often significantly below unity will obviously require a careful monitoring of the *b*-tagging performance with real data. The most abundant source of events con-

taining *b*-jets which could be used to calibrate the efficiency of the *b*-tagging algorithms is $t\bar{t}$ production, since clean samples can be selected by requiring only one tagged *b*-jet. The second *b*-jet from top-quark decay can then be used for calibration.

10.8 References

10-1   T. Sjostrand, Computer Physics Commun. **82** (1994) 74.

10-2   I. Gavrilenko *et al.*, 'Present status of *b*-tagging studies in ATLAS', ATLAS Internal Note INDET-NO-115 (1995).

10-3   S. Haywood, '*b*-tagging with Atlas Inner Detector using Fast Simulation', ATLAS Internal Note INDET-NO-116 (1995).

10-4   D. Froidevaux and E. Richter-Was, Zeit. Phys. **C67** (1995) 213.

10-5   ATLAS Collaboration, Inner Detector Technical Design Report, Volume I, CERN/LHCC/97-16 (1997).

10-6   ATLAS Collaboration, Pixel Detector Technical Design Report, CERN/LHCC/98-13 (1998).

10-7   ATLAS Collaboration, Calorimeter Performance Technical Design Report, CERN/LHCC/96-40 (1996).

10-8   S. Jagielski, A. Kaczmarska and M. Wolter, 'Tagging *b*-jets using low-p_T electrons', ATLAS Internal Note ATL-PHYS-98-129 (1998), published in Acta Physica Polonica **B30** (1999) 4.

10-9   A. Kaczmarska and M. Seman, 'Identification of low-p_T electrons with the ATLAS detector', ATLAS Communication ATL-COM-PHYS-99-047 (1999).

10-10   D. Barberis *et al.*, 'A Comparative Study of Reduced Layouts of the ATLAS Inner Detector', ATLAS Internal Note INDET-NO-188 (1997).

10-11   B.J. Dick, 'Further work on *WH, H* $\rightarrow b\bar{b}$', ATLAS Communication ATL-COM-PHYS-99-019 (1999).