PROCEEDINGS OF SCIENCE

B-Tagging at LHC: Expected Performance and its Calibration in Data

Giacinto Piacquadio* **on behalf of the ATLAS Collaboration**
Physikalisches Institut
Albert-Ludwigs-Universität Freiburg
Hermann-Herder-Str. 3
79104 Freiburg
E-mail: giacinto.piacquadio@physik.uni-freiburg.de

The ability to identify jets containing *b*-hadrons is important for the high-p_T physics program of a general-purpose experiment at the LHC such as ATLAS. This is in particular useful to select very pure top samples, to search and/or study Standard Model or supersymmetric Higgs bosons which couple preferably to heavy objects or are produced in association with heavy quarks. After a review of the algorithms used to identify *b*-jets, their anticipated performance is discussed as well as the impact of various critical ingredients such as the residual misalignments in the tracker. The prospects to measure the *b*-tagging performance in the first few hundreds pb^{-1} of data with di-jet events and $t\bar{t}$ events are then also discussed.

Prospects for Charged Higgs Discovery at Colliders
September 16-19 2008
Uppsala, Sweden

*Speaker.

PoS(CHARGED2008)033

1. Introduction

The identification of *b*-quark jets (*b*-jets) is an important task at a multi-purpose experiment like ATLAS, where many interesting processes giving rise to *b*-jets in the final state will be produced in the collision of the 7 TeV energy proton beams delivered by the LHC accelerator. An efficient and accurate identification of the *b*-quark jets will help to differentiate between the interesting physics and the backgrounds, the latter often being dominated by *u*-, *d*- and *s*-quark jets. Examples of such interesting processes are: the production of $t\bar{t}$ pairs, Higgs boson decays into a pair of *b*-quarks (e.g. $pp \to t\bar{t}H$ with $H \to b\bar{b}$), Higgs associated production in supersymmetric models ($pp \to bbH/A$ with $H,A \to \tau^+\tau^-$), processes involving charged Higgs bosons ($H^+ \to t\bar{b}$ or $t \to H^+b$) and some exotic scenarios, like the decay of a heavy new particle into $b\bar{b}$. Two main signatures are available to separate *b*-jets from jets generated by the fragmentation of lighter quarks (u,d,s): a spatial signature, relying on the lifetime of *b*-hadrons ($c\tau \approx 0.5$ mm), whose decay products form a secondary vertex which is typically displaced with respect to the interaction point by several mm, and a lepton based signature, relying on the semileptonic decay of a *b*-hadron or subsequent *c*-hadron into a muon or electron.

The spatial signature is exploited by the following two classes of algorithms: **Impact Parameter** based algorithms, relying on the (in)compatibility of the individual tracks in a jet with the primary interaction vertex of the event and **Secondary vertex** based algorithms, explicitly requiring the determination of a *b*-hadron decay vertex.

Based on the jet direction, the reconstructed charged particles tracks are associated with the jet to be tagged if $\Delta R(Track, Jet) < 0.4$. The track hit resolution in the innermost pixel layers – in the barrel region $\approx 10\mu m$ in $r\phi$ and $\approx 115\mu m$ in z – determines the impact parameter resolution for high p_T tracks. A degradation is expected for lower p_T tracks due to effect of multiple scattering in the detector material. The primary vertex resolution is $\approx 15\mu m$ in the transverse plane and $\approx 50\mu m$ in the longitudinal plane, depending on the topology of the event [1]. The identification of the signal vertex in the presence of additional overlaid interactions due to pile-up events is crucial for *b*-tagging algorithms.

2. Impact parameter based *B*-tagging algorithms

The impact parameter significance of all tracks, $S_{r\phi} = \frac{d_0}{\sigma(d_0)}$ (in the $r\phi$ plane) and $S_z = \frac{z_0}{\sigma(z_0)}$ (in the *z*-coordinate), are used as input. For each track a sign for the impact parameter is determined, according to whether the track is compatible with having its origin in a vertex in front or behind the primary vertex (with respect to the jet direction).

The distribution of signed impact parameters in the transverse plane is shown in Fig. 1 separately for *b*-, *c*- and light-jets. These distributions define the probability density functions (*PDFs*) for single tracks. These are then combined into a single discriminator, the *jet weight*, using the likelihood ratio formalism: $W_{JET}^{IP} = \sum_{tracks} log\left(\frac{PDF_b(S_{r\phi})}{PDF_{light}(S_{r\phi})}\right)$. To combine the transverse and longitudinal impact parameters information two-dimentional *PDF*s ($PDF(S_z, S_{r\phi})$) are used, taking at the same time their correlations correctly into account (the corresponding jet weight is shown in Fig. 2).

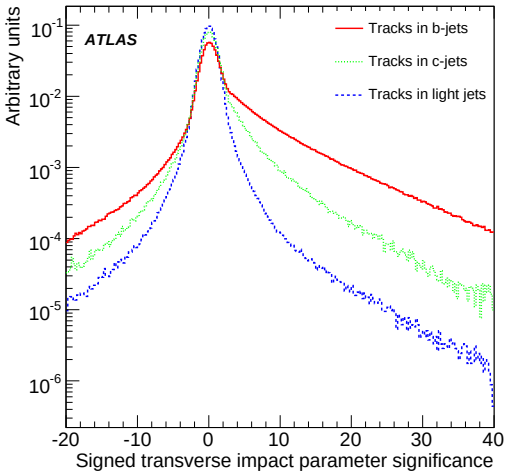


Figure 1: Signed transverse impact parameter significance distribution.

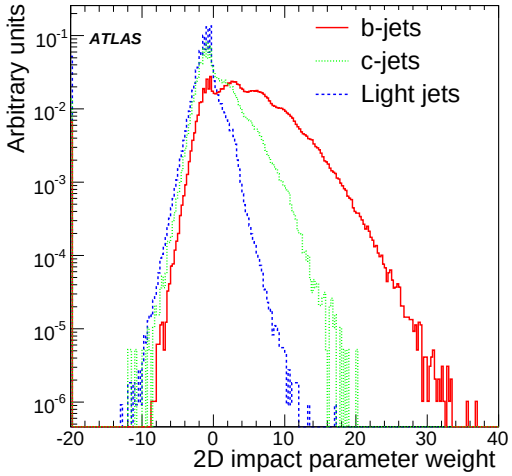


Figure 2: Jet weight distribution based on both transverse and longitudinal impact parameters.

Simpler algorithms, not relying on any assumption of the impact parameter significance distribution in *b*-jets are also available, such as simply counting the number of tracks with impact parameter significance above a predefined threshold or evaluating the probability for the tracks in a jet to originate from a light-jet. These will be particularly important for the commissioning of *b*-tagging algorithms during the early data taking phase of ATLAS.

3. Secondary vertex based *B*-tagging algorithms

Two strategies to detect a secondary decay vertex in *b*-jets are implemented in ATLAS: a fully *inclusive* approach relying on the reconstruction of a single displaced vertex and a *topological* approach which attempts to explicitly reconstruct the $PV \rightarrow B \rightarrow D$ decay chain.

3.1 Inclusive secondary vertex reconstruction

First the displaced tracks to be used in the secondary vertex fit are identified, by considering all possible displaced two-track vertices. Tracks corresponding to vertices compatible with K_s, Λ decays, conversions or hadronic interactions are removed. The tracks surviving this selection are used as an input for the vertex fit (as implemented in the *VKalVrt* package [4]): incompatible tracks are iteratively removed from the fit. A likelihood function is defined for *b*- and light-jets, based on the fraction of jets with a reconstructed secondary vertex, the invariant mass of the charged particles assigned to the secondary vertex, the fraction of charged particles energy in the reconstructed secondary vertex with respect to all charged particles in the jet and the number of good two-track vertices. A *jet weight* is then defined using the likelihood ratio of the *b*- and light-jet hypotheses, $W_{JET}^{SV} = log\left(\frac{L_b}{L_l}\right)$. A combined secondary vertex and impact parameter based *b*-tagging algorithm is then easily formed by adding their respective jet weights: $W_{JET} = W_{JET}^{IP} + W_{JET}^{SV}$.

3.2 Topological reconstruction of the $PV \rightarrow B \rightarrow D$ decay chain

The fragmentation of a *b*-quark results in a decay chain composed of a secondary vertex from the weakly decaying *b*-hadron and typically one or more tertiary vertices from *c*-hadron decays. These vertices are very difficult to separate efficiently. In this algorithm, the assumption is made that the transverse momentum of the *c*-hadron with respect to the *b*-hadron flight direction is negligible with respect to the overall *c*-hadron momentum, so that the primary vertex and the decay vertices of the *b*- and *c*-hadrons in the decay chain lie on the same line. A vertexing algorithm based on this principle (first adopted by SLD in the *ghost track algorithm*[2]), has been implemented in *JetFitter* [5] using an original extension of the Kalman Filter formalism commonly used for vertexing [6]: an arbitrary number of vertices with one or more tracks can be fitted efficiently, constraining them to lie on a common *b*-hadron flight axis whose origin is the primary vertex.

A first fit is performed initializing the flight axis with the jet direction and considering all tracks in the jet to form single-track vertices along this axis. A clustering procedure is then performed, merging the vertices two by two in decreasing order of probability, until a stable configuration is reached. Considering *b*-jets simulated in $t\bar{t}$ and $t\bar{t}jj$ Monte Carlo events the topological approach obtains a **single multi-track vertex** in $\approx 50\%$ of *b*-jets, a **single multi-track vertex plus an additional single-track vertex** from a second vertex in $\approx 16\%$ of cases, **two multi-track vertices** in $\approx 5\%$ of cases and **two single-track vertices** in $\approx 5\%$ of cases.

A likelihood function and a jet weight is then defined analogously to the fully *inclusive approach*, but including the additional information about the identified decay chain topology, and again combined with the impact parameter only based discriminator.

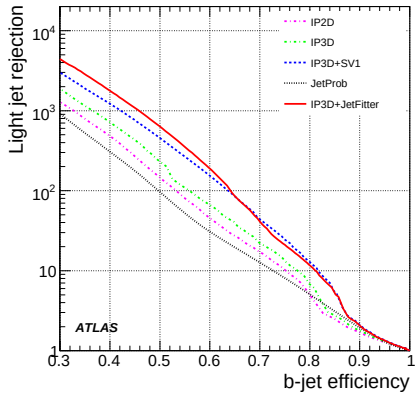
4. *B*-tagging performance of spatial algorithms

The *b*-tagging performance was tested on a sample of fully simulated $t\bar{t}$ and $t\bar{t}jj$ Monte Carlo events. The *b*-tagging efficiency is defined as the fraction of identified *b*-jets, while the *b*-tagging rejections (r_u and r_c) are defined as the inverse of the fraction of light- or charm-quark jets mistagged as *b*-jets. The light-quark rejection as a function of the *b*-tagging efficiency is shown in Fig. 3 for various *b*-tagging algorithms, together with a table showing the light-quark and charm-quark rejections at a fixed *b*-tagging efficiency of 50% and 60% for the three most important algorithms.

An ideal geometry and a pixel single hit inefficiency of 5% were assumed in the simulation used for this study. A degradation of the *b*-tagging performance is expected due to residual misalignment. Recent studies show that, after a realistic alignment procedure, starting from a randomly misaligned detector with misalignments of the order of $10-100$ μm as expected from fabrication precision and survey measurements, a degradation of less than 25% in the light-quark rejection is expected. Further studies are ongoing, in particular to control global deformations which are only weakly constrained by the alignment procedure.

5. Lepton based *b*-jet identification

The identification of *b*-jets based on a muon or electron from the semileptonic decay of the $b-$ or *c*-hadron from the $b \rightarrow c$ decay chain is limited by the semileptonic branching fraction ($BR(b \rightarrow lX) \approx 11\%$ and $BR(b \rightarrow c \rightarrow lX) \approx 10\%$ for both $l =$muon or electron). The lepton based



	IP3D	IP3D+SV1	IP3D+JetFitter
light-jet rejection			
$\varepsilon_b = 50\%$	232±2	456±4	635±7
$\varepsilon_b = 60\%$	67±0	154±1	189±1
c-jet rejection			
$\varepsilon_b = 50\%$	10.6±0.0	12.4±0.1	12.3±0.1
$\varepsilon_b = 60\%$	6.5±0.0	7.4±0.0	7.4±0.0

Figure 3: *b*-tagging performance as a function of *b*-tagging efficiency (left) and for fixed values of 50 and 60% *b*-tagging efficiency (right), for various algorithms: *JetProb* (based only on the resolution function for prompt tracks), IP2D (based on transverse impact parameters), IP3D (based on transverse + longitudinal impact parameters), IP3D+SV1 (based on one single inclusive secondary vertex, combined with IP3D) and JetFitter+IP3D (based on the reconstruction of the $PV \rightarrow b \rightarrow c$ decay chain, combined with IP3D).

signature is however nearly uncorrelated with the spatial signature, which makes the lepton based *b*-tagging algorithms particularly useful to calibrate the spatial algorithms directly on data. After a lepton candidate in a jet has been identified as muon or electron, the transverse momentum of the lepton with respect to the jet flight axis ($p_{T,rel}$) is used in order to further separate the signal leptons from fakes or from real leptons in light-quark jets. While the muon signature is very clean, so that a light-quark rejection of ≈ 300 can be reached at a *b*-tagging efficiency of 10%, it is much more challenging to identify electrons in a dense jet environment, so that a rejection of ≈ 100 can be reached at a *b*-tagging efficiency of 7%[1].

6. Measurement of *B*-tagging performance on data

A lot of effort has been put into trying to ensure that the ATLAS Monte Carlo simulations will resemble the behaviour of the ATLAS Detector with real data as closely as possible. However the real *b*-tagging performance will need to be measured on data itself. Two main strategies have been set up for this. The first is based on the use of two uncorrelated *b*-tagging algorithms, relying one on the lepton, the other on the spatial signature and on the selection of a sample of QCD dijet events where the $b\bar{b}$ component is enriched through the online selection of a jet with a contained muon. The second method relies on the kinematic selection of $t\bar{t}$ events in order to obtain a very pure *b*-jet sample.

6.1 Measuring performance in QCD dijet events

The so called *System 8* method, first used at the DØ experiment, uses two samples with different flavour composition: the first is dijet events with a *Jet* + *Muon* signature, the second requires an additional jet on the opposite side to be identified as *b*-jet by an impact parameter based algorithm.

[1]The ATLAS electron based tagging algorithm includes the lepton impact parameter information, while the muon based one doesn't.

Two uncorrelated *b*-tagging algorithms are used, corresponding to four possible combinations: no *tag*, μ *tag*, spatial *tag* and spatial+μ *tag*. Applying these four combinations to the two samples already mentioned, a system of eight equations with eight unknowns can be formulated. The solution of the system yields the flavour composition of the two samples and the tagging efficiencies for the different flavours. An alternative method, based on the determination of the flavour sample composition of the selected jets in dijet events applying templates of $p_{T,rel}$ for *b*-, charm- and light-jets before and after *spatial b*-tagging is applied, has also been studied [3].

Both the *System 8* and the $p_{T,rel}$ methods are expected to be dominated by systematic uncertainties already after 50 pb^{-1} of collected data with a dedicated trigger. A p_T and η dependent measurement of the *b*-tagging efficiency with a precision of 6% up to a *b*-jet transverse momentum of 80 GeV seems to be feasible, with larger errors up to 150 GeV. Further studies are ongoing to evaluate the systematic uncertainty on the Monte Carlo based correction which is needed to account for the bias introduced by requiring the $Jet + Muon$ signature and extend the method to higher jet energies with more data.

6.2 Measuring performance in $t\bar{t}$ events

Three methods have been proposed to measure the *b*-tagging efficiency in $t\bar{t}$ events [3]. The method based on the topological selection of one leptonic and one hadronic top will be briefly described here. After a basic preselection, the *b*-jet stemming from the hadronic top is required to pass a *b*-jet identification cut, while the leptonic one is left unbiased. Based on the leptonic and hadronic top mass, a signal region is defined. The background turns out to be almost purely combinatorial, originating from $t\bar{t}$ itself: a signal free region is defined using the mass sidebands and requiring the *b*-jet from the leptonic top not to pass a very loose *b*-jet identification cut. The distribution of the *b*-jet weight (i.e. of the discriminating variable) for the *b*-jet from the leptonic top is then obtained after subtracting the background on a statistical basis. The knowledge of the *b*-jet weight can be translated directly into a *b*-tagging efficiency for an arbitrary cut value.

This method permits to obtain a measurement of the *b*-jet efficiency in bins of jet p_T. With 200 pb^{-1} of data and for jets with $p_T > 40$ GeV, a relative precision of $\pm$7.7% (stat.) and $\pm$3.2% (syst.) can be achieved.

7. Outlook

A lot of effort is being spent to further optimize the performance of the *b*-tagging algorithms. An area which has lately attracted more attention is the specific optimization of the charm-quark rejection. Charm-quark jets are more difficult to reject, because they fragment into *c*-hadrons, which also have a detectable lifetime. However, since a $PV \rightarrow B \rightarrow D$ decay chain is expected out of the hadronization of a *b*-quark and typically only one decay vertex out of a *c*-quark, using the topological reconstruction of the $PV \rightarrow B \rightarrow D$ decay chain and using dedicated *PDF*s for charm-jets it is possible to enhance the charm-quark rejection, at the cost of a decreased light-quark rejection. Preliminary results show that at 50% *b*-tagging efficiency an increase of up to $\approx$ 60% in charm-quark rejection is achievable, at the cost of a decrease in light-quark rejection of up to $\approx$ 65%. The optimal working point will clearly depend on the flavour composition of the background of the specific physics analysis of interest.

8. Conclusions

The methods developed to identify *b*-jets in ATLAS have been described. In terms of light-quark rejection, at 60% *b*-tagging efficiency the performance achievable by various algorithms, in order of expected commissioning, is ≈ 30, relying only on the resolution function of prompt tracks, then up to ≈ 70, using the transverse + longitudinal impact parameter based algorithm, and finally $\approx 150-190$, using the most sophisticated secondary vertex based algorithms. Preliminary studies show that the residual misalignement of the Inner Detector should degrade these numbers by less than 25%. Methods have been established to measure the *b*-tagging efficiency on data to about 6% accuracy with 100 pb^{-1} of data. Studies are ongoing to determine the mistagging rate on data, but around 10% precision is expected from the Tevatron experience.

References

[1] G. Piacquadio, K. Prokofiev and A. Wildauer, *Primary Vertex Reconstruction In The ATLAS Experiment At Lhc*, *J. Phys. Conf. Ser.* **119** (2008) 032033.

[2] K. Abe *et al.* [SLD Collaboration], *Time dependent B_s^0 $\bar{B}_s^0$ mixing using inclusive and semileptonic B decays at SLD*, in proceedings of the *19th Intl. Symp. on Photon and Lepton Interactions at High Energy LP99* ed. J.A. Jaros and M.E. Peskin,

[3] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment, Detector, Trigger and Physics*, *CERN-OPEN-2008-020*, Geneva, 2008, to appear

[4] V. Kostyukhin, *VKalVrt - A package for vertex reconstruction in ATLAS*, *ATL-PHYS-2003-031* (2003).

[5] G. Piacquadio and C. Weiser, *A New Inclusive Secondary Vertex Algorithm For B-Jet Tagging In ATLAS*, *J. Phys. Conf. Ser.* **119** (2008) 032032.

[6] R. Frühwirth, *Nucl. Instrum. Meth.* A 262 (1987) 444