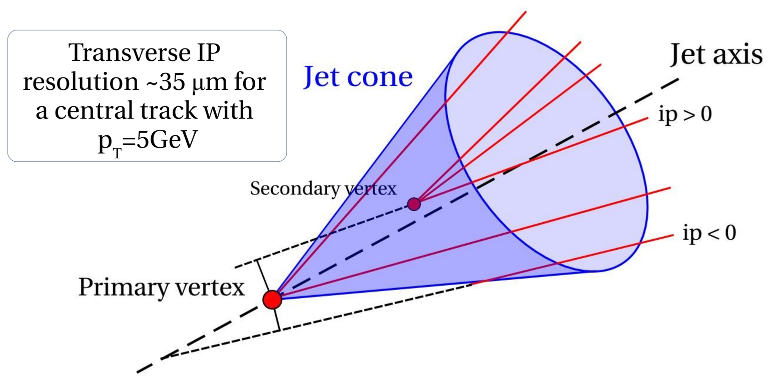
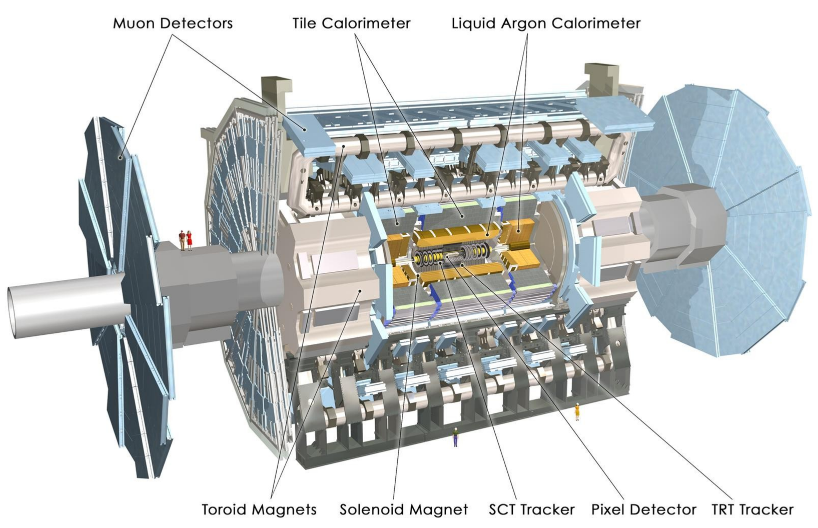
Tagging b-jets in ATLAS

Thomas Goepfert (Technische Universitaet Dresden) on behalf of the ATLAS collaboration

Tagging algorithms





Impact parameter (IP) based

TrackCounting: simple cut on the ordered IP significance of the n-th track (n=2, 3) in a jet

JetProb: calculates the probability that a track originates from the primary vertex by comparing the IP significance to a resolution function R for prompt tracks (obtained from tracks with IP<0)

$$P_{track} = \int_{-\infty}^{-|IP/\sigma(IP)|} R(x)dx$$

IPxD (x=1,2,3): Likelihood Ratio (LR) approach using the IP significance distribution S of tracks stemming from b and light (udsg) jets in x dim to calculate a jet weight

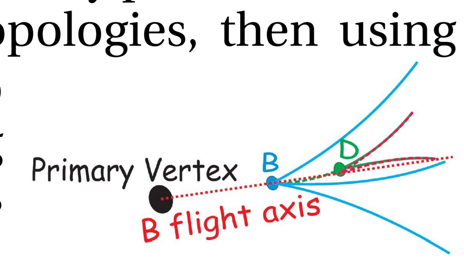
$$jet\ weight = \sum_{i=1}^{N_{tracks}} ln\frac{b(S_i)}{u(S_i)}$$

Performance gain by grouping the tracks into categories (e.g. p_T) with dedicated pdf's

Secondary vertex (SV) based

Inclusive SV: reconstruction of two-track vertices inside a jet, rejecting V0 and conversions and fitting to a common SV, cut on decay length significance or LR approach based on one or more dimensional distributions of vertex mass, energy fraction and number of two-track vertices

JetFitter: uses Kalman filter to find a common line of PV and B/D-hadron decay products which also includes *incomplete* topologies, then using LR approach similar to inclusive SV but taking different decay topologies into account



Combination of jet weights from IP and SV based taggers or multivariate techniques (BDT) for better performance

Soft lepton taggers

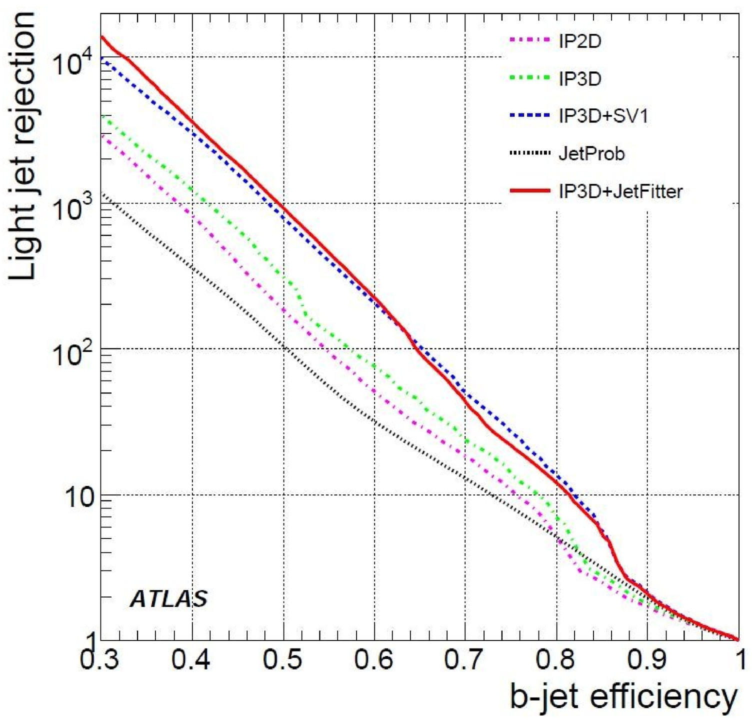
Soft muon: jet weight from 1 dim LR method using the muon relative transv. momentum p_T^{rel} wrt the jet axis

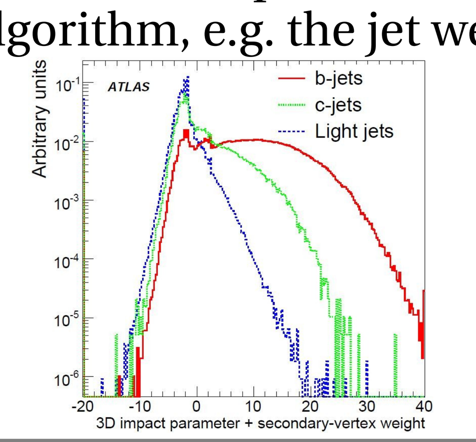
Soft electron: LR approach based on several variables from the inner detector, the calorimeter and both in combination

Exploiting B-hadron properties:
- relatively high mass (~5GeV)
- charged track multiplicity (~5)
- long lifetime (cτ ~ 450 µm) results in displaced decay vertex

Performance

The performance of b-tagging algorithms is described by the tagging efficiency and the rejection rate R of light/charm and tau jets, which is the inverse of the mis-tag rate. Those quantities are obtained from MC samples by cutting on the discriminating variable of the respective tagging algorithm, e.g. the jet weight.



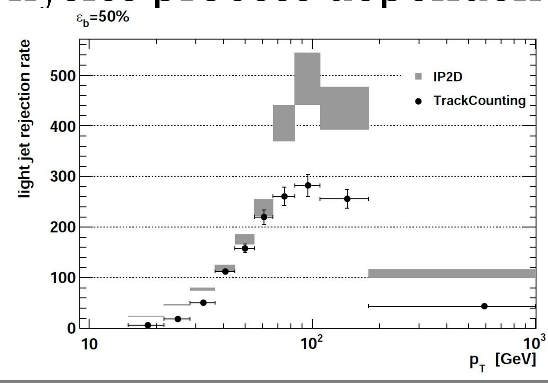


The expected rejection rate at 60% b-tag efficiency is ~30 for the simplest taggers (JetProb or TrackCounting) which are suitable for start-up and about 200 for high-end taggers like JetFitter.

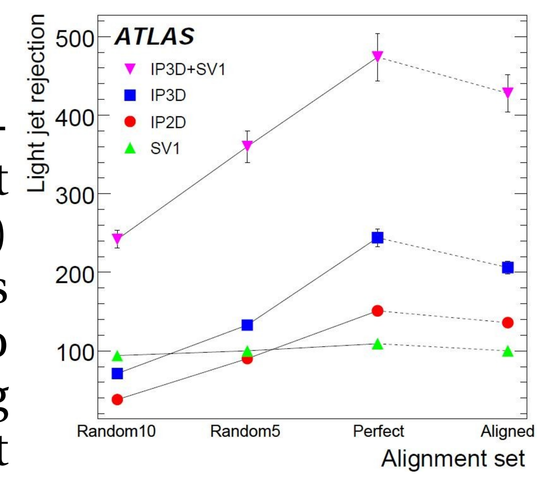
Further rejection rates

- Soft muon: R~300 at 10% (including BR of b → µ X)
- Soft electron: R~100 at 7%
- High level trigger (IP3D): R~10 at 70%

At a given b-tag efficiency the performance of the algorithms depends on the jet transverse momentum and rapidity and thus is physics process dependent.



Some critical ingredients

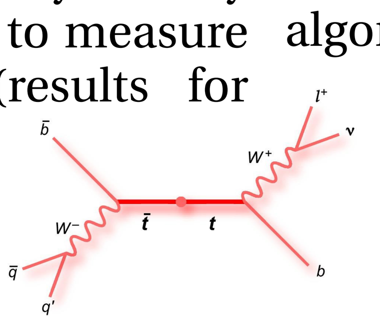


Residual misalignment: The impact on b-tag performance is studied with different misalignment sets with shifted (10-30 µm) and rotated (~0.3 mrad) Pixel modules and with a more realistic case (similar to the misalignment we see after aligning with cosmics) of actual realignment including systematic errors leading to a degradation of about 30%.

Primary Vertex reconstruction: At LHC the beamspot size is σ_{xy}=15 µm and σ_z=5.6 cm. With and without pile-up (~5 min bias events at L=2x10^{33} cm^{-2}s^{-1}) the PV resolution is about the same (σ_{xy}=12 µm, σ_z=40 µm) but up to 5% are misidentified in the presence of additional vertices. This leads to artificially big longitudinal IP and to a degradation of the performance of ~30% for IP3D and ~40% for IP3D+SV1 whereas b-tagging algorithms using only transverse IP are nearly not effected.

Calibration

Strategies to measure the tagging and mis-tagging efficiency directly in data. Two main approaches to measure b-tag efficiency at ATLAS (results for 14 TeV MC).

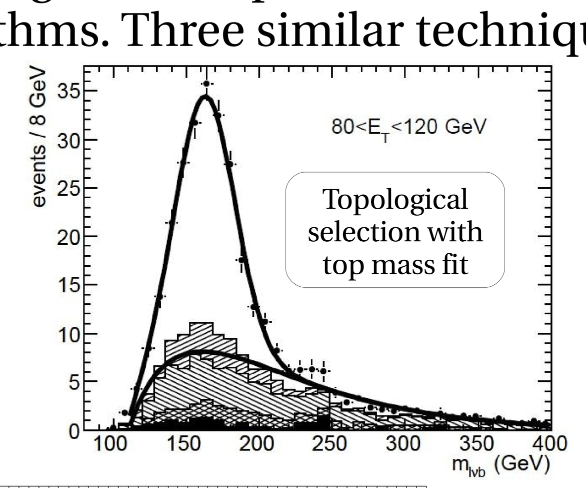


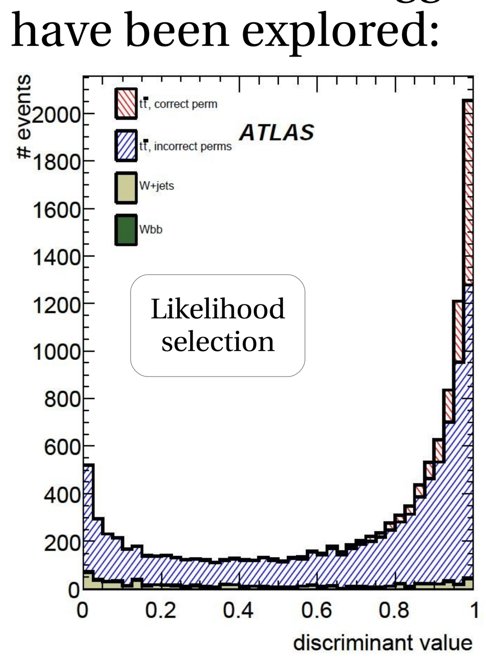
$t\bar{t}$ events

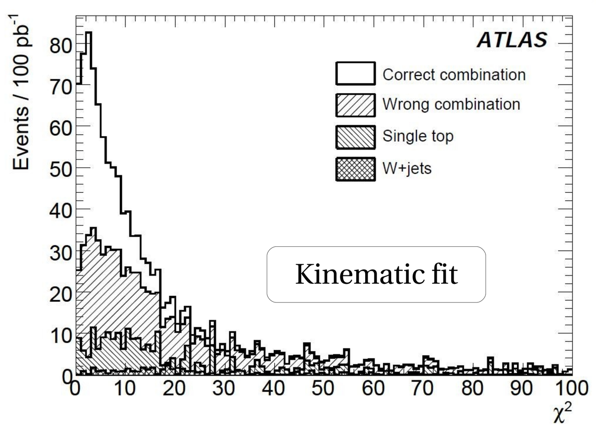
Event counting: count number of events with 1,2,3 b-tags (lepton+jets) or 1,2 b-tags (dilepton channel), likelihood fit gives b-tag efficiency and $t\bar{t}$ cross section at the same time as well as the c-tag efficiency in the lepton+jets channel

With 100 pb^{-1} the relative precision at 60% b-tag efficiency is ±2.7(stat.)±3.4(syst.)% in the lepton+jets channel (slightly worse in the dilepton channel).

Event reconstruction: obtaining a pure sample of identified b-jets by full reconstruction of a semileptonic $t\bar{t}$ decay chain, studying the output distribution of the relevant tagging algorithms. Three similar techniques have been explored:







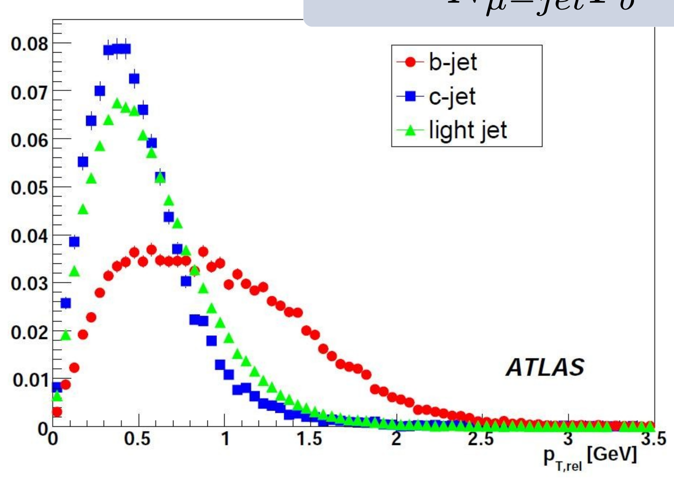
For the topological method, 200 pb^{-1} of data are needed and allow a relative precision of ±2.7(stat.)±3.4(syst.)%.

dijet events

System 8 method: uses two samples of jets with different b-quark content and two uncorrelated tagging algorithms (e.g. soft muon and spacial tagger) to form a system of 8 equations and 8 unknowns, one of which is the b-tagging efficiency

p_T^{rel} method: uses jets with an associated muon, fitting templates of muon p_T relative to the jet+muon axis for b/c- and light jets before and after b-tagging

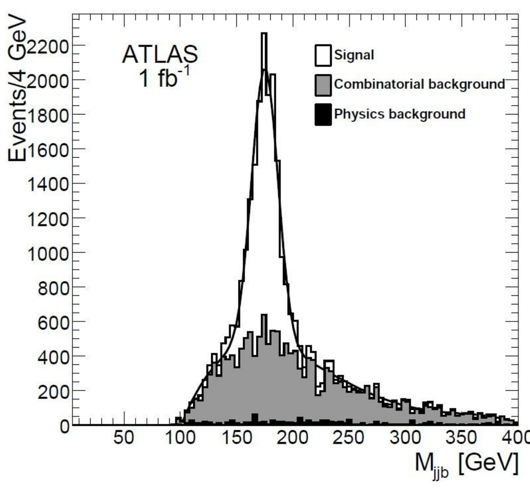
$$\epsilon_b = \frac{N_{\mu-jet}^{tag} F_b^{tag}}{N_{\mu-jet} F_b}$$



Both methods give results binned in p_T and η. With 50 pb^{-1} a relative precision of about 6% can be reached.

Use cases for b-tagging

Top mass measurement

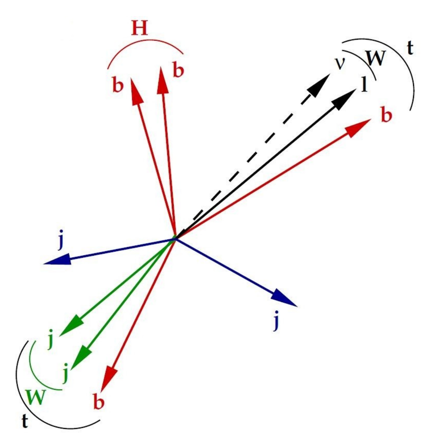


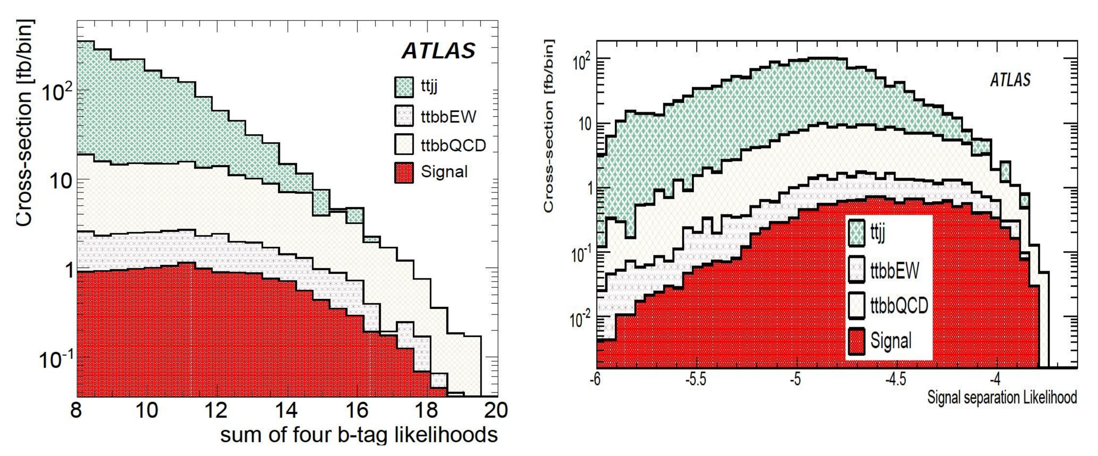
The best top quark mass determination is achieved with two b-tagged semi-leptonic $t\bar{t}$ events and a top mass estimator taken as the invariant mass of the three jets from hadronically decaying top quark. The uncertainty on the top quark mass will be dominated by systematics and relies mainly on the jet energy scale uncertainty. A precision of 1(3.5) GeV is reachable with 1 fb^{-1}, assuming a jet energy scale uncertainty of 1(5)%.

A complementary approach relying on b-tagging infers the top quark mass from the mean transv. decay length of B-hadrons coming from the top decays, here the uncertainty due to the jet energy scale is negligible.

Search for $t\bar{t}H$ ($H \to b\bar{b}$) in the SM

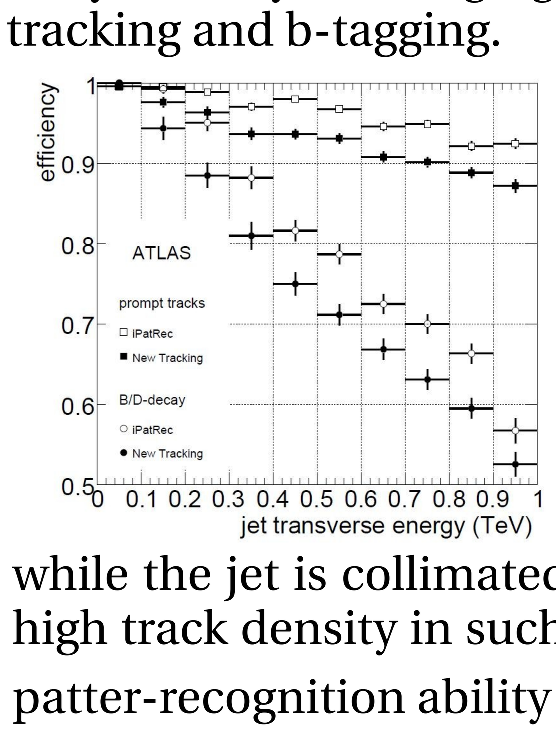
B-tagging helps to reduce or even to eliminate large backgrounds like $t\bar{t}$+jets, W+jets, tW. The b-tagging weight is used as a cut and as input variable to a likelihood selection for jet assignment and signal to background separation.

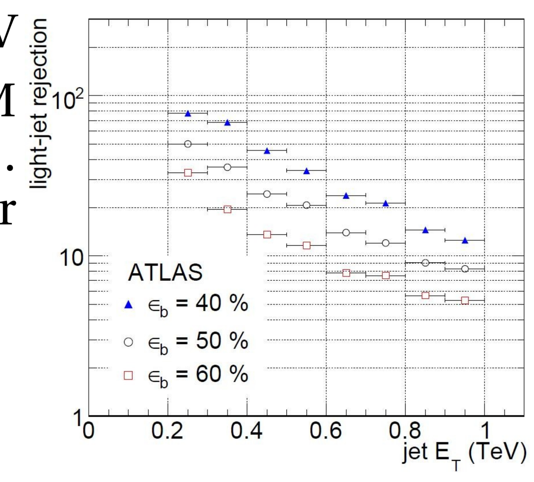




High p_T b-tagging for BSM/Exotic

Jets with p_T above 500 GeV appear in many BSM models, e.g. Z'→Zh ($h \to b\bar{b}$). They are very challenging for tracking and b-tagging.





Performance is worse for such jets due to several effects. The fraction of fragmentation tracks increases with parton p_T, while the jet is collimated into a narrower cone. The high track density in such high p_T jets challenges the patter-recognition ability of the tracking software.

XXIV International Symposium on Lepton Photon Interactions at High Energies