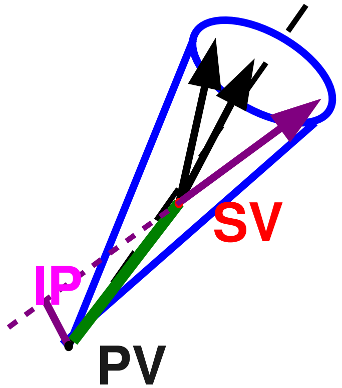
Higgs at LHC:
An overview of b and tau tagging

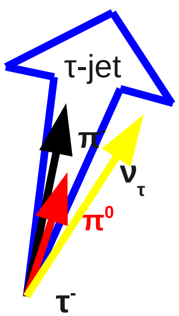
Michał Bluj
CNRS/LLR-École Polytechnique
on behalf of ATLAS and CMS collaborations

Outline

- Introduction
- B-tagging
  - Techniques & commissioning results



- Tau identification
  - Techniques & commissioning results



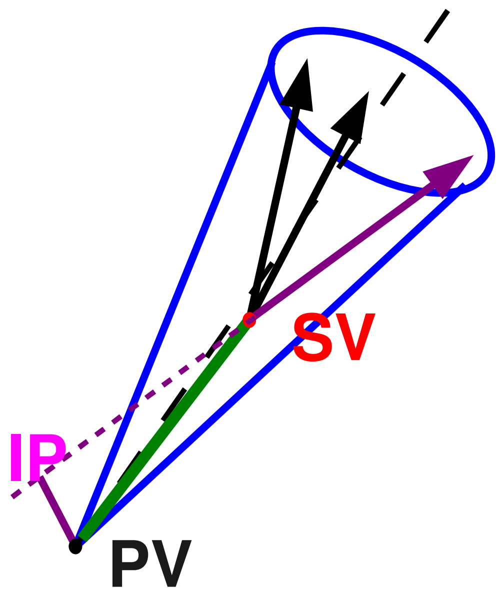
Higgs boson and b-quark & tau

- SM light Higgs bosons (m_H< 150 GeV)
  - **H→bb**: Dominant mode, but overwhelmed by QCD background. Possible in associated mode tt(H→bb)→**4b**2l or VH→V**bb** at hight p_T
  - **H→ττ**: Exploitable when produced in VBF – two forward tagging jets
- In MSSM with tanβ > ~3: significant enchencment of H to down fermions couplings :
  - (bb)H→ττ important also for high m_H
  - $H^{\pm}$→τν replacing $W^{\pm}$ in tt decays
- (Anti)b-tagging tool to suppress background with b: tt, Zbb,...

Understanding of b and tau tagging is very important for (light) Higgs searches at LHC.

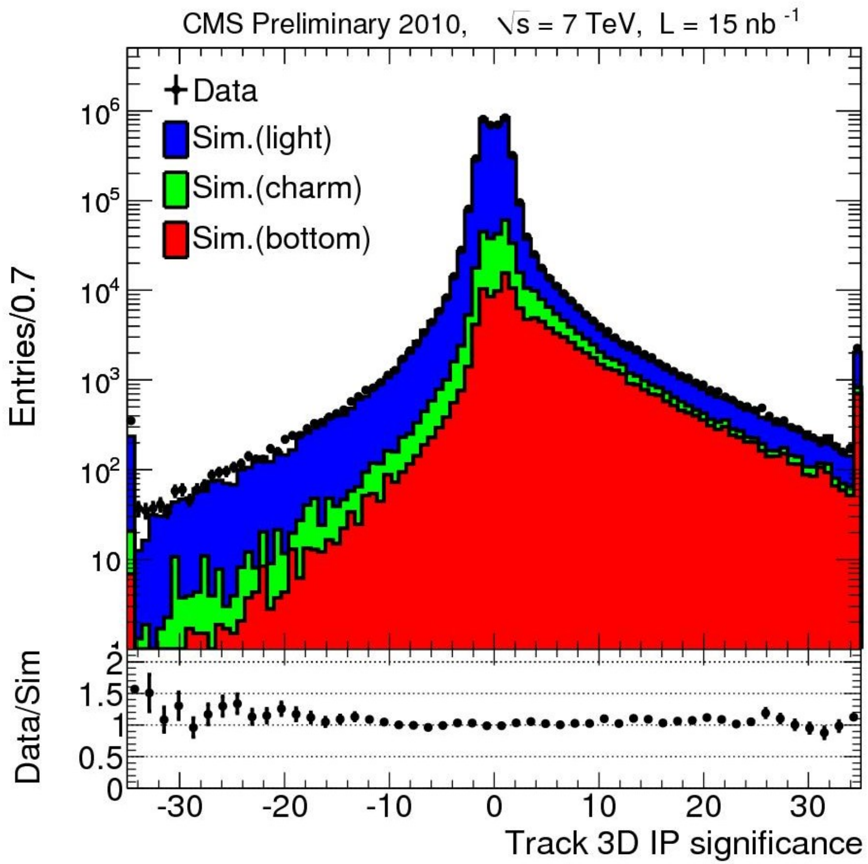
b-tagging, introduction

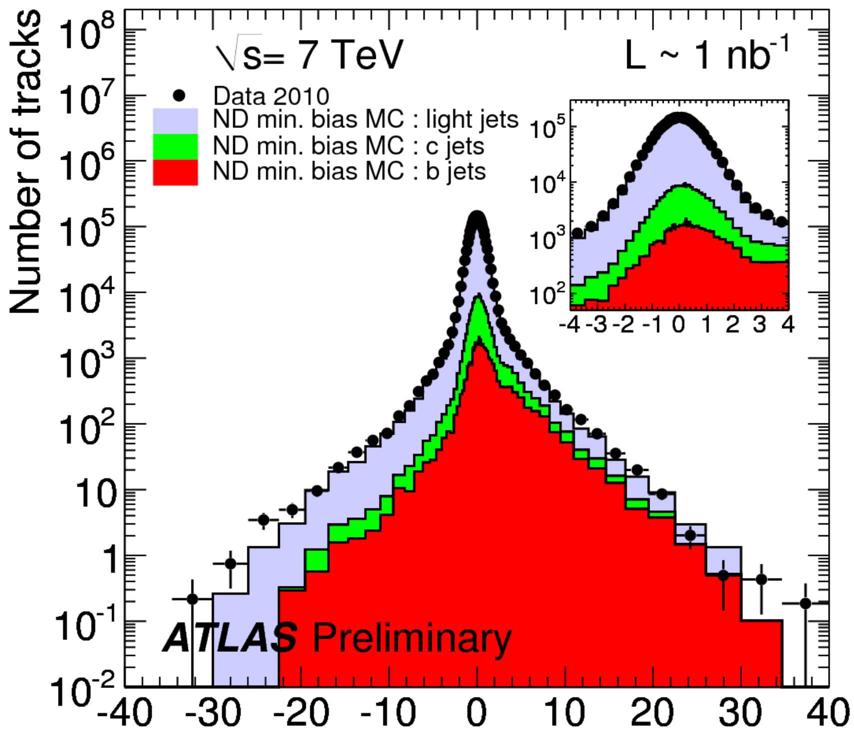
- There exist different b-tagging algorithms
  - Exploiting the long life time and high mass of B hadron ($c\tau_B$~0.5 mm, m_B~5 GeV)
    - **Track Counting (TC)** calculates and orders signed impact parameter significance of good tracks. B-tagger is significance of the N'th track.
    - **Jet Probability (JetProb)** calculates probability that tracks in jet are compatible with primary vertex (PV). Density of probability is obtained with negative IPs.
    - **Secondary Vertex (SV)** reconstructs the B decay vertex using good tracks in jet. Variables such as decay length significance and mass of tracks from SV used to calculate b-tag discriminator.
  - Looking for leptons from semileptonic b decay
    - Relative transverse momentum of lepton and/or its signed impact parameter significance
- In future all of the b-taggers will be combined into one b-tagging variable (with likelihood, MVA,...)

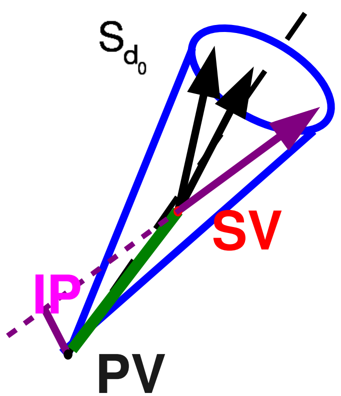


Some “control” variables with 7TeV data

b-tagging, IP significance





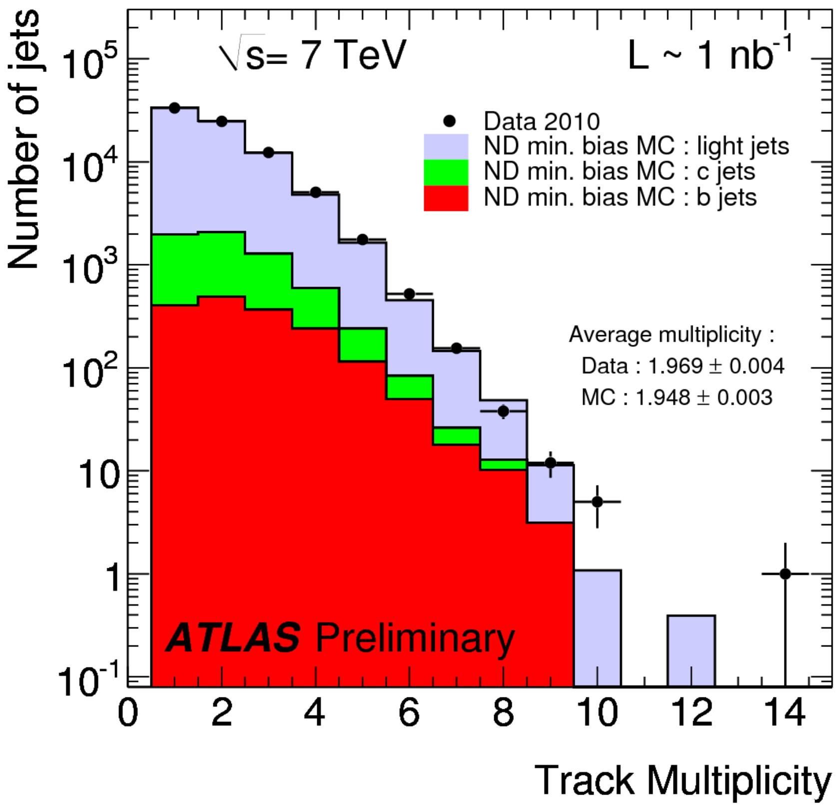


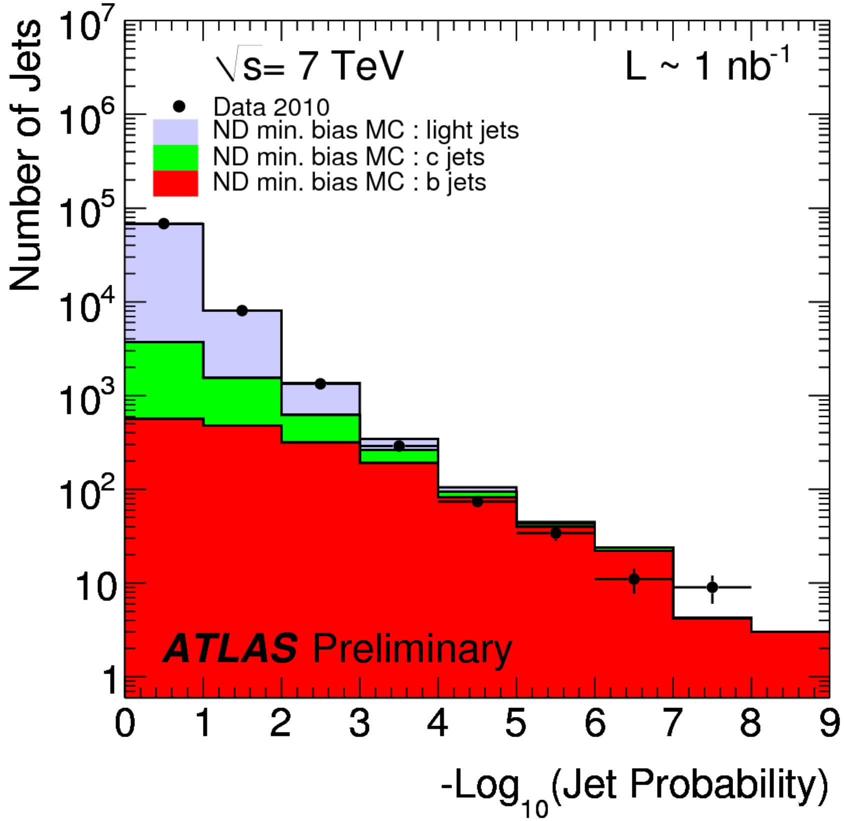
IP significance for all selected tracks associated to jet

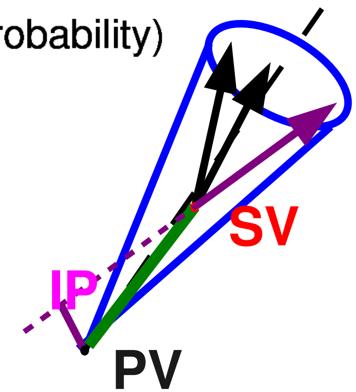
- Left, CMS: p_T(jet) > 30 GeV and |η| < 2.4
- Right, ATLAS: p_T(jet) > 20 GeV and |η| < 1.8

Remarkable good agreement between collision data and MC expectations

b-tagging, jet probability







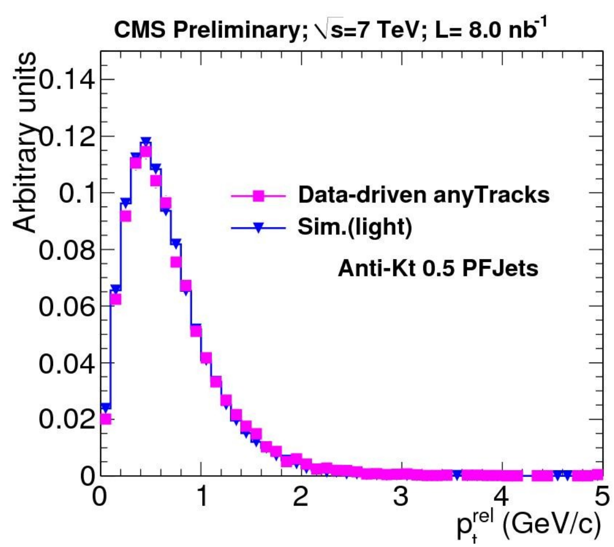
- Left, ATLAS: number of tracks with positive IP in jet
- Right, ATLAS: probability that jet is compatible with a light flavour hypothesis

Colours indicate flavour composition of MC sample.

Remarkable good agreement between collision data and MC expectations

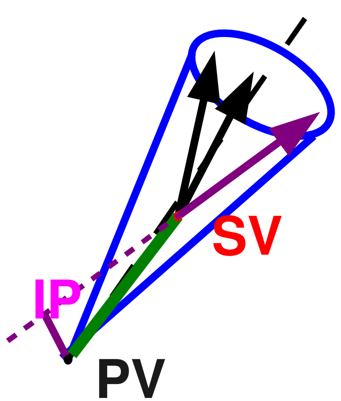
b-tagging, performance from data

- Measurement of b-tag performance from data is an important issue
  - flavour content of samples not perfectly known,
  - modeling with MC sensitive to many systematic uncertainties, ...
- Several methods possible
  - Use sample with jets with muons (p_T^{rel}, system8 methods)
  - Use samples with well defined jet flavour composition tt, Z+jets,...
- p_T^{rel} method
  - Muon p_T^{rel} from heavy flavours (HF) harder than from light flavours (LF)
  - Fit mu p_T^{rel} spectrum in data with template distributions for HF (from MC) and LF (data-driven) to obtain flavour content of jets



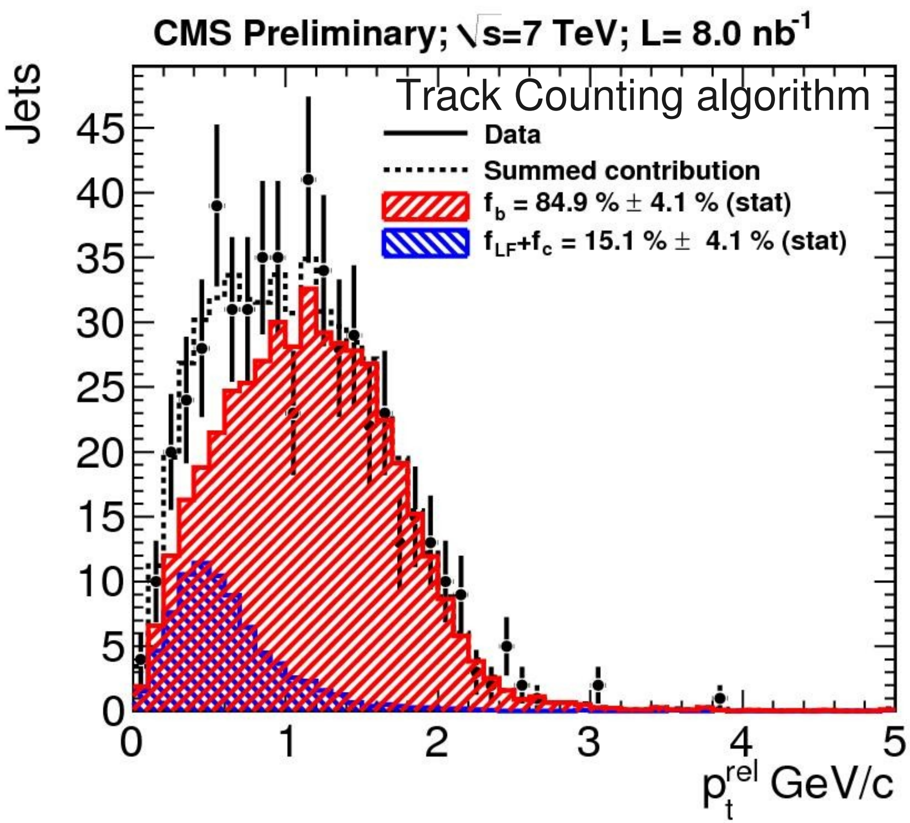
Data-driven p_T^{rel} spectrum for muons in LF jets (any track) compared with simulated spectrum.

Good agreement observed.



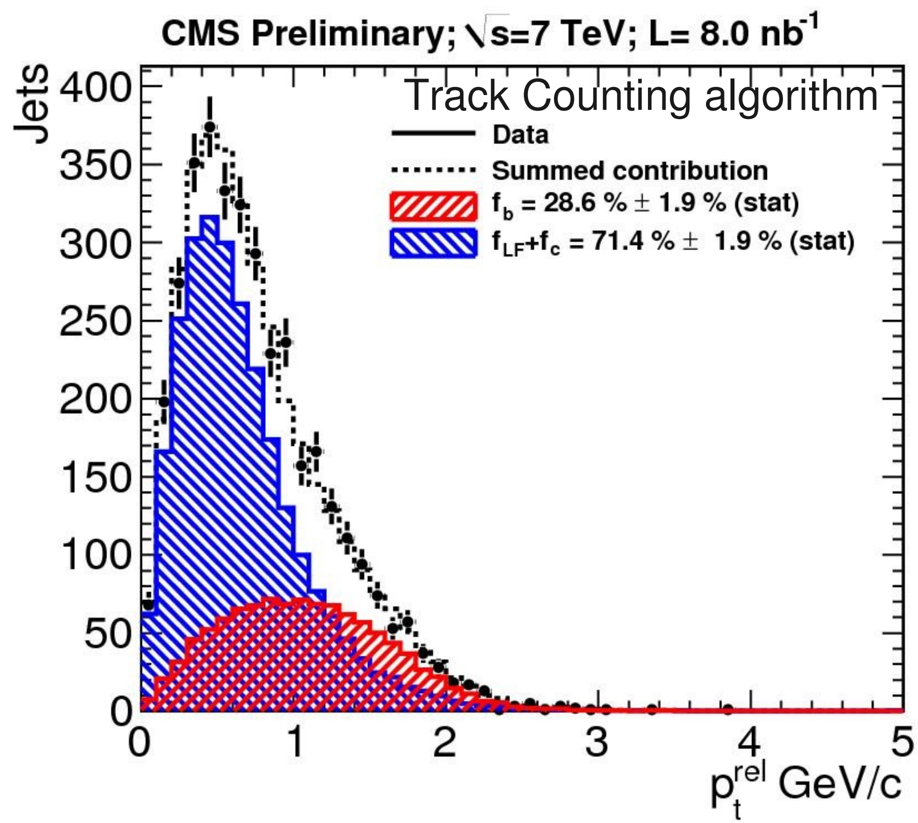
b-tagging, performance from data

◎ Measurement of b-tag efficiency form data: p_T^{rel} method



Template fit for b-tagged jets

⇨ $f_b^{tag} = 85\%$, $f_{LF+c}^{tag} = 15\%$



Template fit for anti b-tagged jets

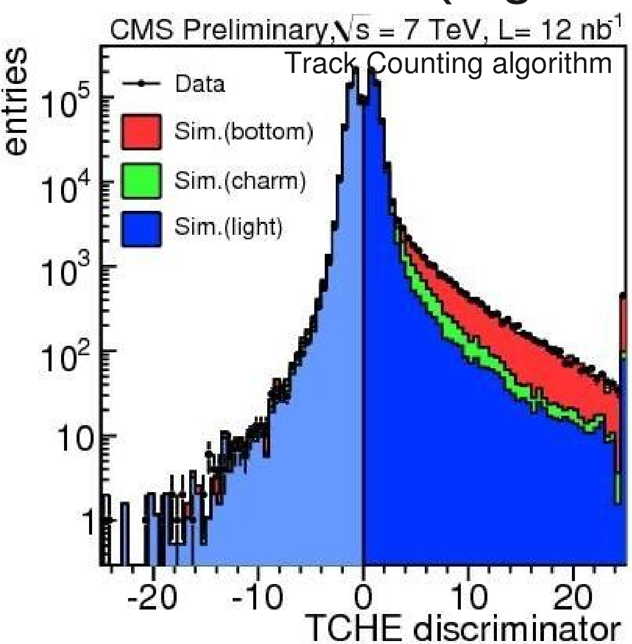
⇨ $f_b^{untag} = 29\%$, $f_{LF+c}^{untag} = 71\%$

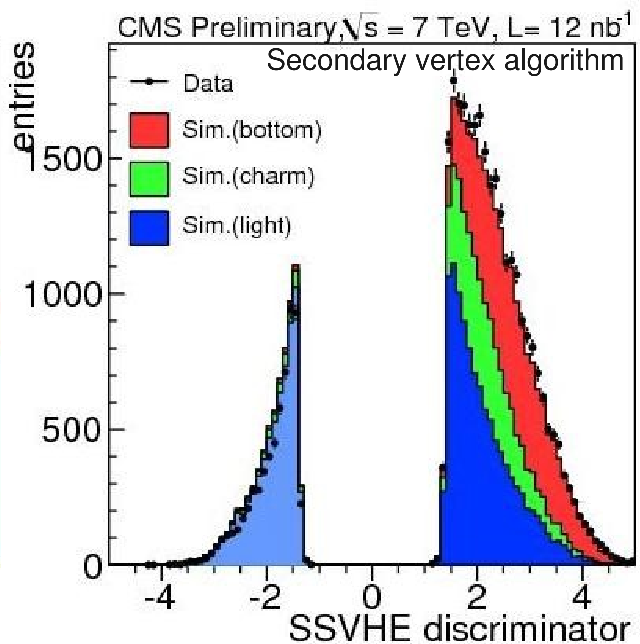
Combining the flavour content and number of tagged/untagged jets one can compute efficiency (or mistag rate) of given b-tagger:

- $\varepsilon_b^{tag} = f_b^{tag} \cdot N^{tag} / (f_b^{tag} \cdot N^{tag} + f_b^{untag} \cdot N^{untag})$

b-tagging, mistag from data

- Mistag rate can be evaluated from data using negative impact parameters or negative decay lengths
  - Negative impact parameters (or decay lengths) reflect an instrumental precision – they should be same as positive ones for track from primary vertex (light flavours, w/o V^0's)
- Method:
  - Compute rate of jets tagged with “b-taggers” obtained for tracks with the negative impact parameters (negative decay lengths) - “negative tag rate”
  - Correct the rate for ratio of light flavour mistags to negative tags from simulation (e.g. V^0's)



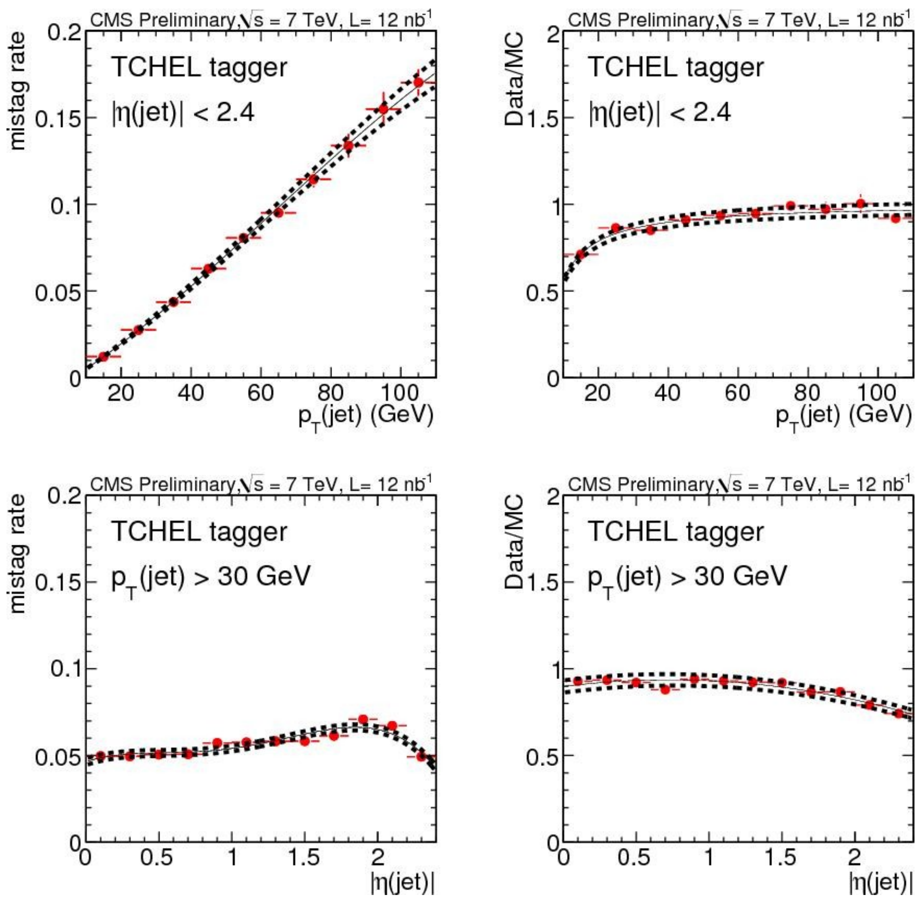


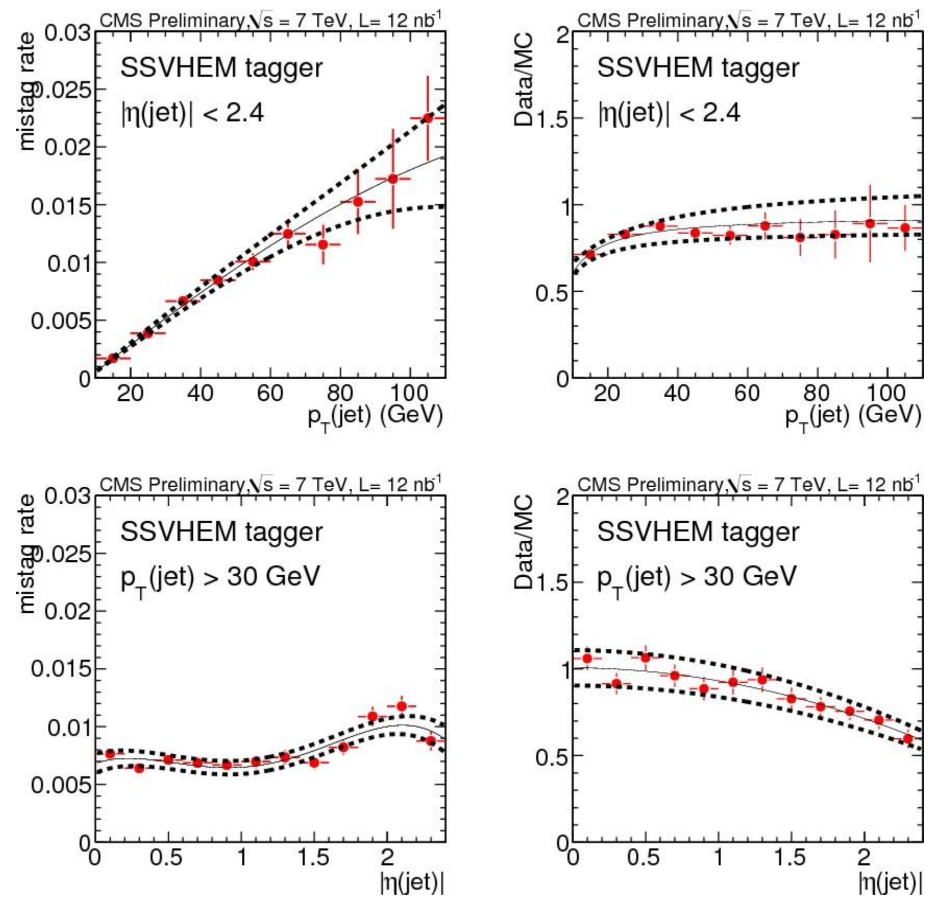
Positive and negative b-tag discriminators

Negative (positive) discriminators in simulation indicated by light (dark) shading

Good agreement observed

b-tagging, mistag from data





Mistag rate for Track Counting tagger

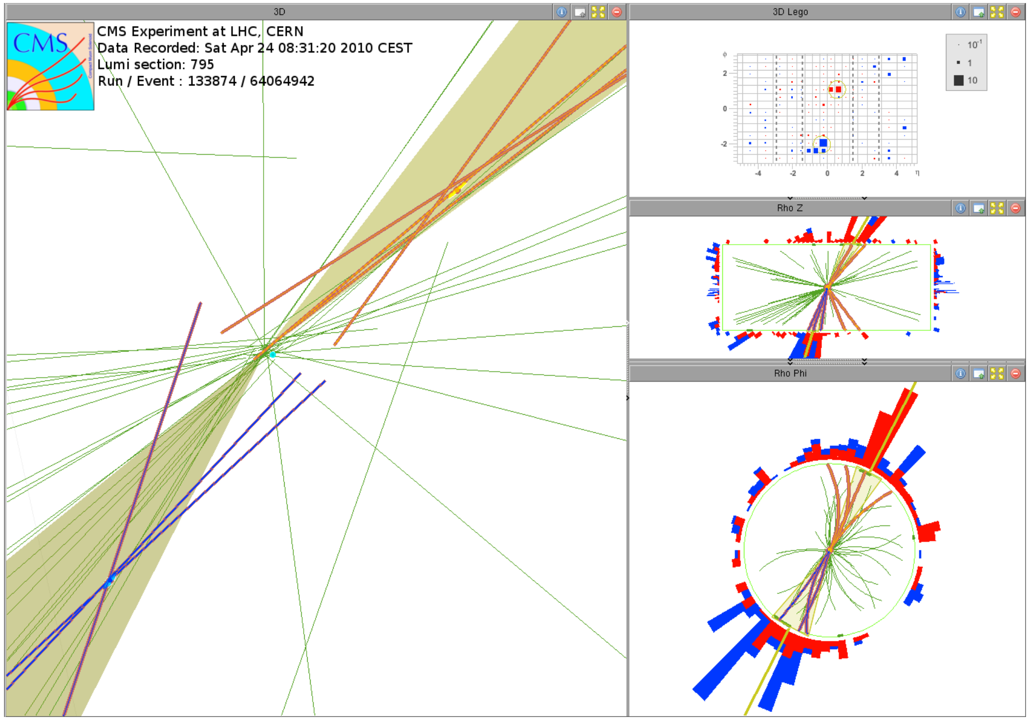
- Right column: data-driven measurement
- Left column: data/MC scale factor

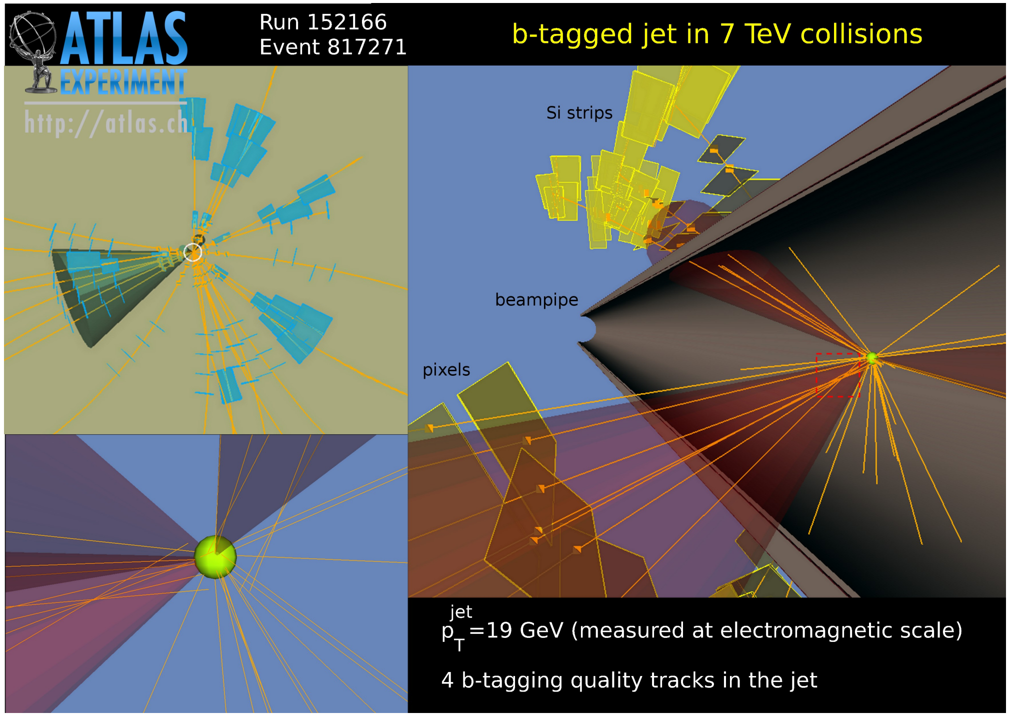
Mistag rate for Secondary Vertex tagger

- Right column: data-driven measurement
- Left column: data/MC scale factor

Solid curve is a result of fit of polynomial to data points, dashed ones represents overall statistical and systematic uncertainties of the measurement.

b-jet candidates





2 b-jet candidates

- p_T=43.7 / 40.3 GeV
- Good secondary vertices

b-jet candidate [bottom left]

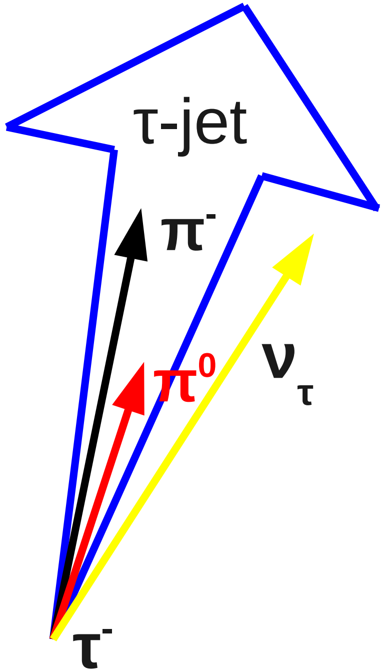
- p_T=19 GeV
- Good secondary vertex

Three other jets with p_T>15 GeV

Good b-jet candidates are being found by both experiments.

Tau-Id, introduction

- Goal: Reconstruct tau decaying into hadrons (τ-jet)
  - Leptonically decaying taus are indistinguishable from prompt leptons (due to small time life of tau cτ~90μm)
  - BR(τ→hadrons) ≈ 64%
- Tau jet properties
  - Narrowness
  - Leading (charged) particle, low multiplicity
  - Visible decay products of taus are soft due to escaping neutrinos – requires low thresholds to preserve acceptance
- Neutrinos from tau decays contribute to MET
- Main background: quark/gluon jets
  - Fake rate at O(1%) level - one order of magnitude higher than for electrons and muons

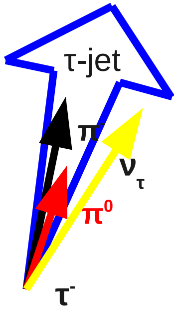


Tau-jet, identification (ATLAS)

- Each calorimetric jet (antiKt R=0.4) with $p_T > 10$ GeV or track with $p_T > 6$ GeV is a seed for hadronic tau candidate reconstruction

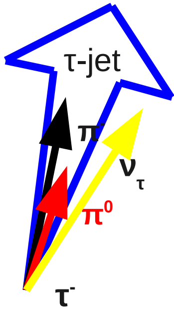
  ~70% candidates is double seeded (calorimetric and double seeded candidates considered further in this early data study)

- Tracks in cone R=0.2 around seed axis are associated to it (signal tracks).
- EM clusters associated to seed isolated with respect to tracks are interpreted as π^0 candidates
- For such candidates a set of topological variables is computed. The most robust ones are analysed with early collision data:
  - Mass of calorimetric topclusters ($m_{clusters}$)
  - Mass of associated tracks (m_{tracks})
  - Track radius (R_{track}): p_T of tracks weighted by its distance to tau candidate axis
  - EM radius (R_{EM}): transverse EM energy weighted by distance to tau candidate axis
  - Core energy fraction (f_{core}): ratio of E_T in core of tau candidate (R<0.1) to its total energy

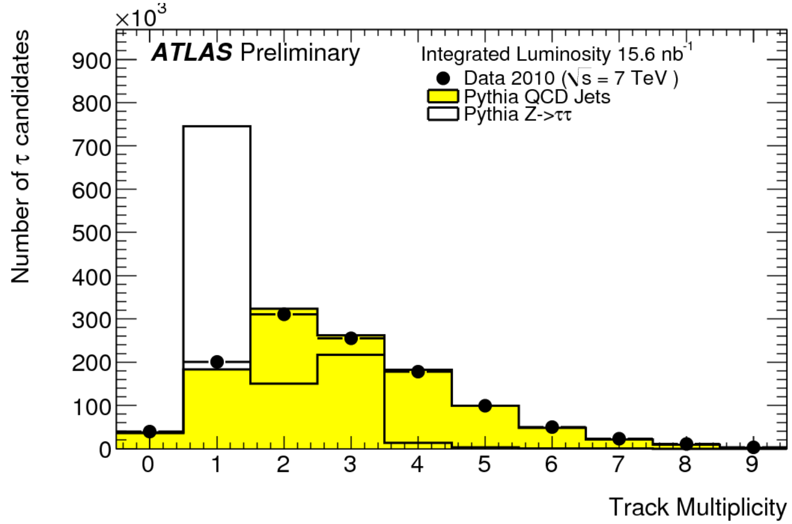


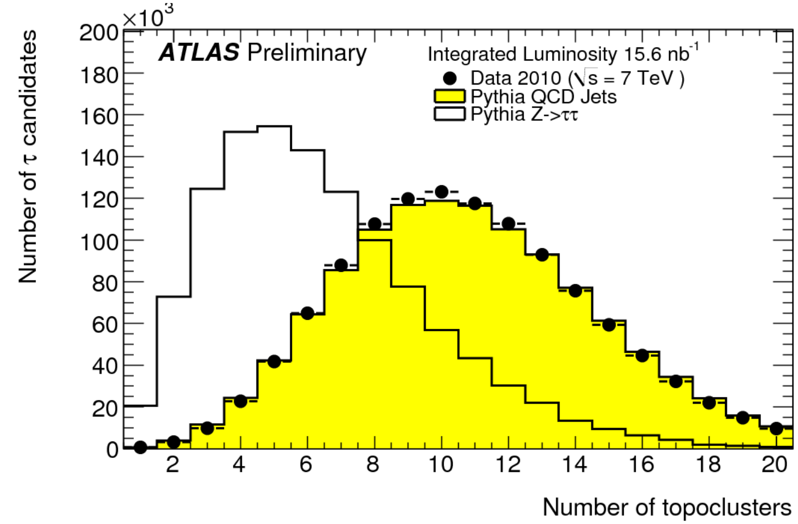
Tau-jet, identification (CMS)

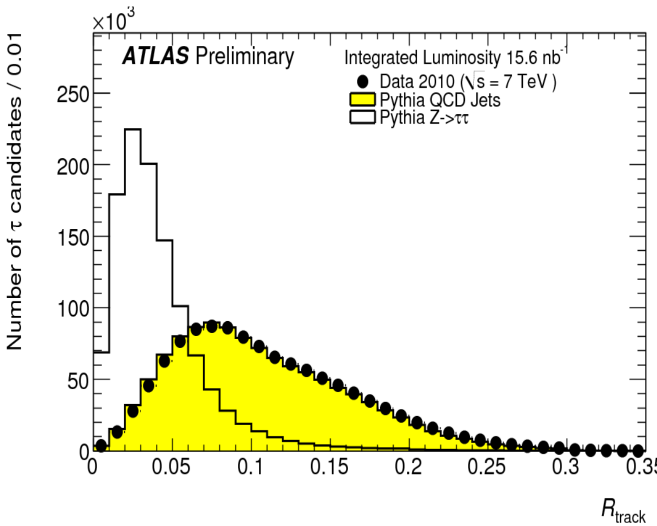
- Each reconstructed jet is a seed for hadronic tau reconstruction
  - Both calorimetric and particle flow jets can be used (antiKt R=0.5)
- Simple robust cone algorithms (ShrinkingConeTau and TCTau)
  - Two cones around a leading particle close to jet axis (charged or π^0 candidate)
    - Narrow signal cone with all tau decay products ($R=5/E_T$ or $R=0.07$)
    - Isolation cone around signal one without any charged or EM activity above some threshold
- More complex algorithms analyse jet constituents to identify specific tau decay mode (using information provided by particle flow technique)
  - Hadron Plus Strip (HPS) algorithm – cut on multiplicity and invariant mass of charged particles and EM strips in $R=2.8/E_T$ cone combined with requirement of no other activity in the jet above some threshold
  - Tau Neural Classifier (TaNC) – reconstruct decay mode by combining charged and π^0 candidates. Then set of NN (one for each decay mode) combines tau-jet candidate properties and returns discriminant variable

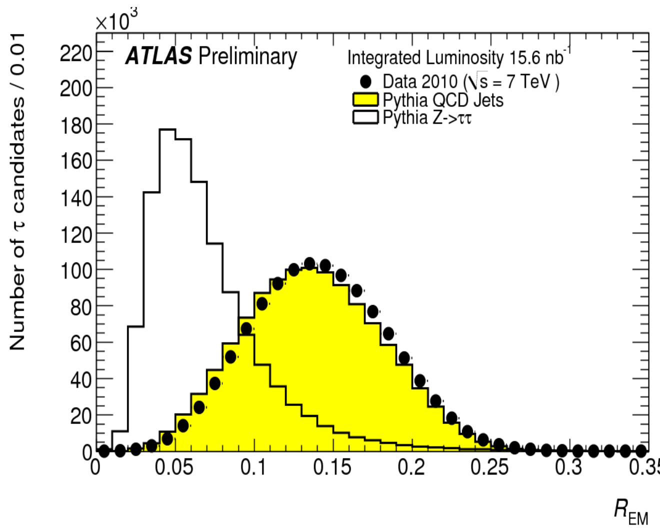


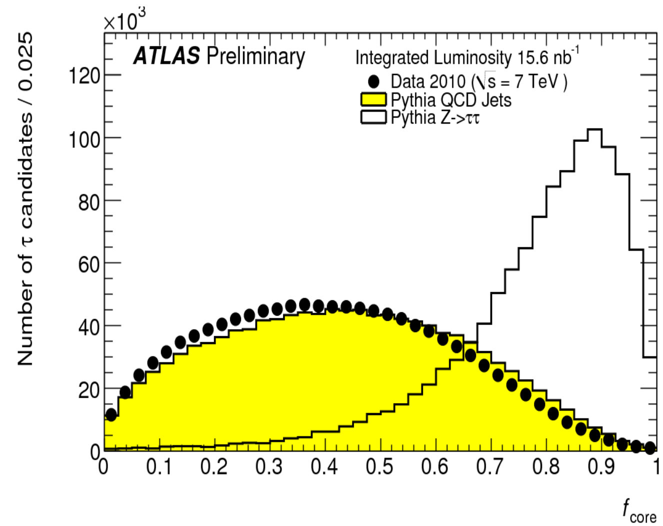
Tau-jet, basic variables (ATLAS)











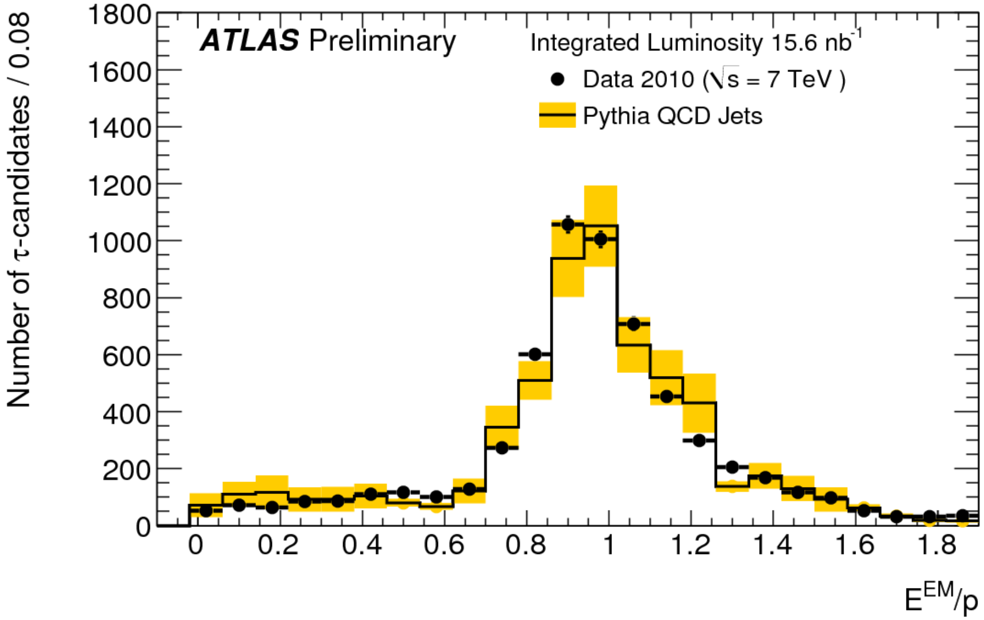
- Basic variables to distinguish hadronic taus and q/g jets tested with data.
- Signal distribution (simulated $Z\rightarrow\tau\tau$) superimposed for comparison.

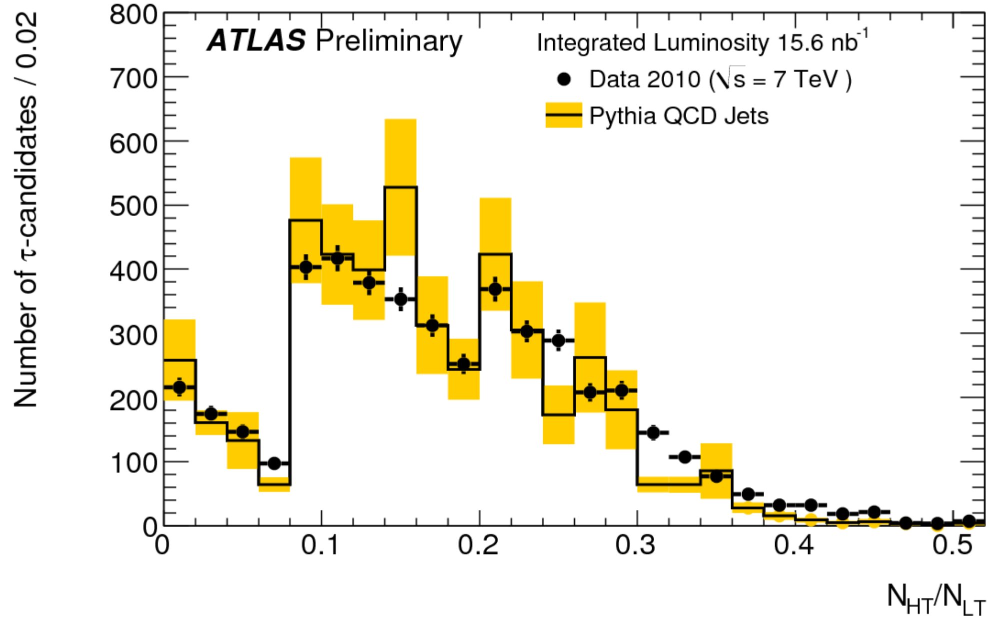
Good agreement between collision data and MC expectations

  - Preliminary studies indicates that observed disagreement depends on QCD tune

Tau-jet, electron rejection

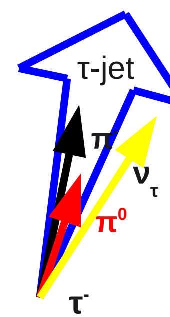
- Electrons are an important source of fake taus
  - They are isolated calorimetric objects





Variables to distinguish hadronic taus and electrons for tau candidates matched to well reconstructed electrons

- Left: Ratio of EM energy around the impact cell to the momentum of the leading track of tau candidate
- Right: Ratio of number of high threshold to low threshold hits in the Transition Radiation Tracker of tau candidate

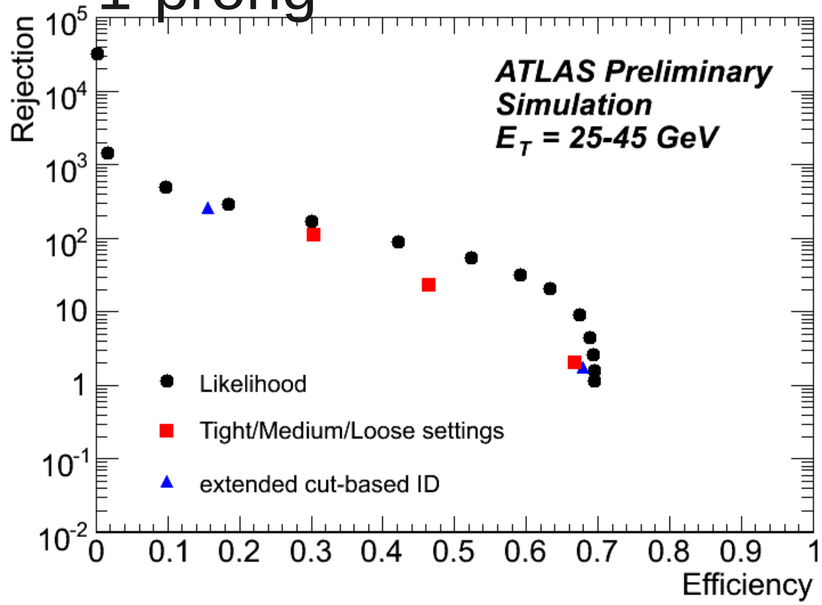


Very good agreement between collision data and simulation

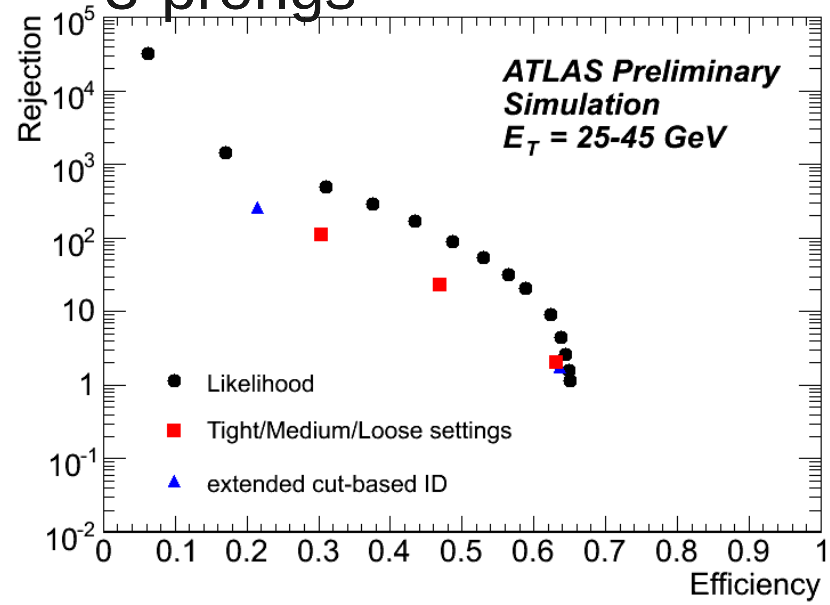
Tau-jet, expected performance

ATLAS

1-prong



3-prongs



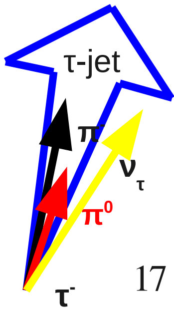
Efficiency of tau identification vs quark/gluon jet rejection obtained with simulation ($Z \rightarrow \tau\tau$ and QCD di-jets samples)

- Optimized separately for 1- and 3-prongs in p_T bins

- **Expected performance** (medium working point, p_T=25-45 GeV):

  efficiency $\varepsilon_{sig} \approx$ 45-50% for rejection r ≈ 23 (“fake rate”: $\varepsilon_{bkg} \approx$ 4%)

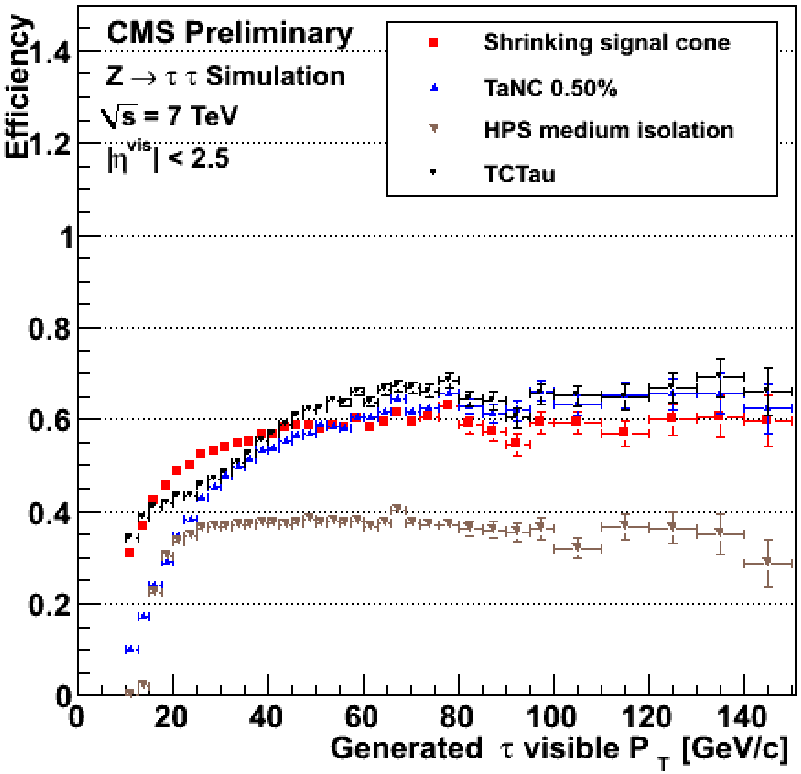
  - Rejection: $r = 1/\varepsilon_{bkg} - 1$



Tau-jet, expected performance

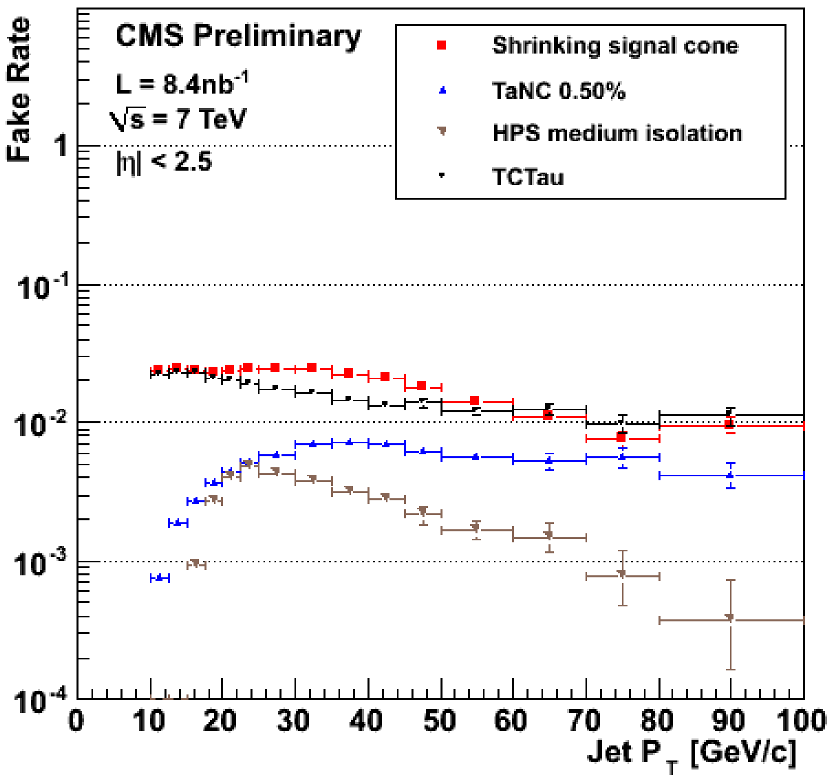
CMS

Simulation



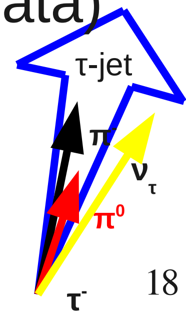
Efficiency of tau-jet identification for 4 algorithms (simulated $Z \rightarrow \tau\tau$)

Data

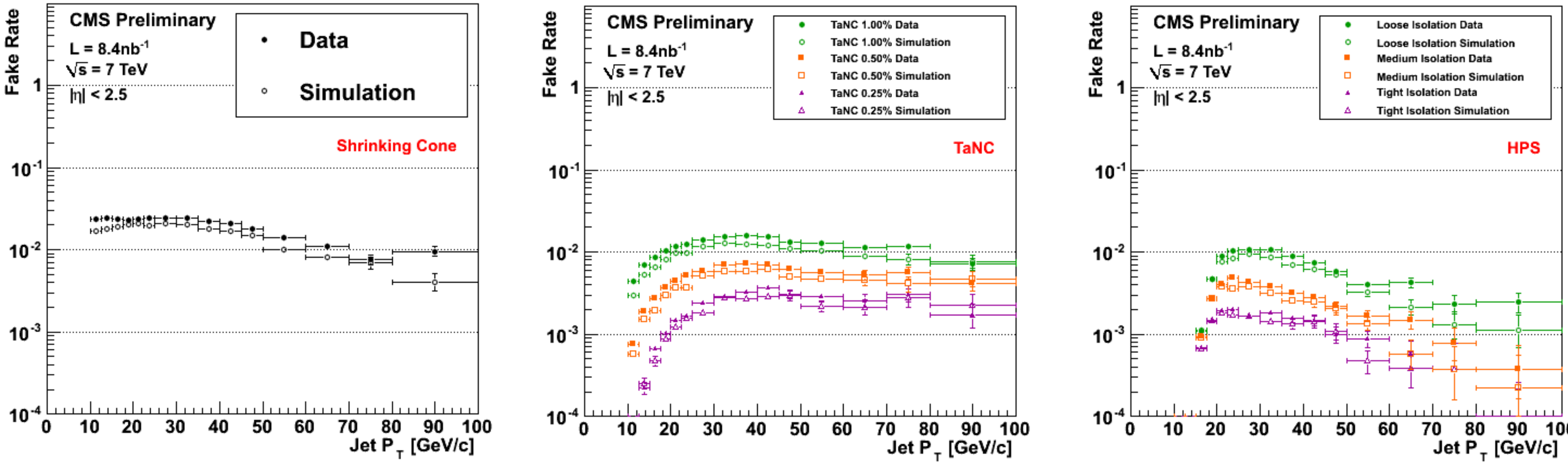


"Fake rate": probability that quark/gluon jet passes tau-jet identification (7TeV collision data)

Expected average efficiency $\varepsilon_{sig} \approx$ 30-45% for "fake-rate" $\varepsilon_{bkg} \approx$ 0.5-2%

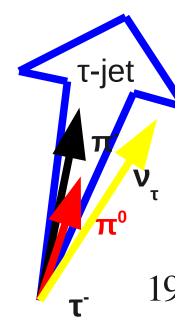


Tau-jet, fake rate from data

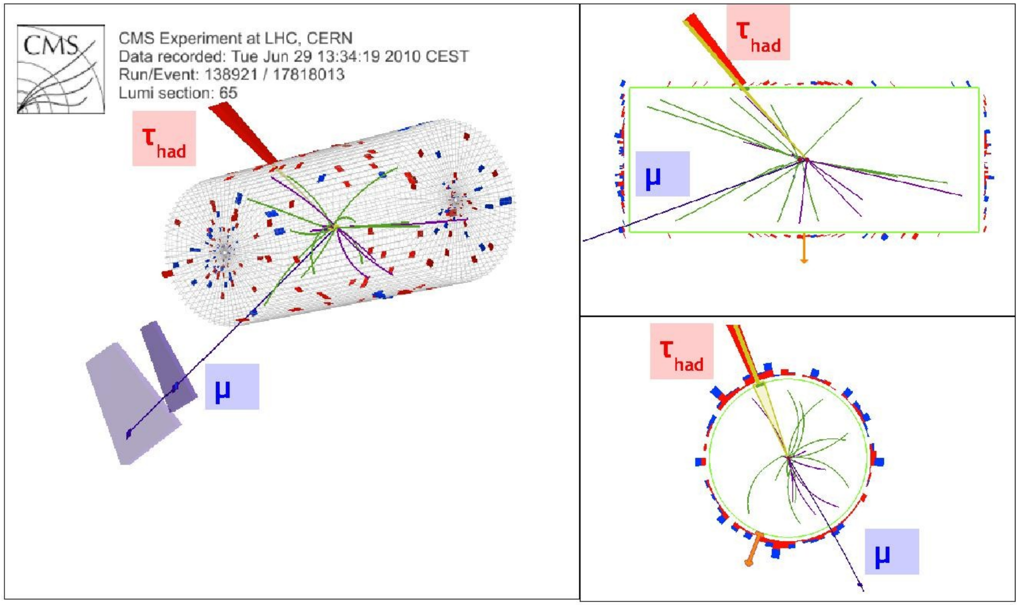


"Fake rate" measured with jets from 7TeV collision data (solid symbols) compared to MC simulation prediction (open symbols) for 3 algorithms.

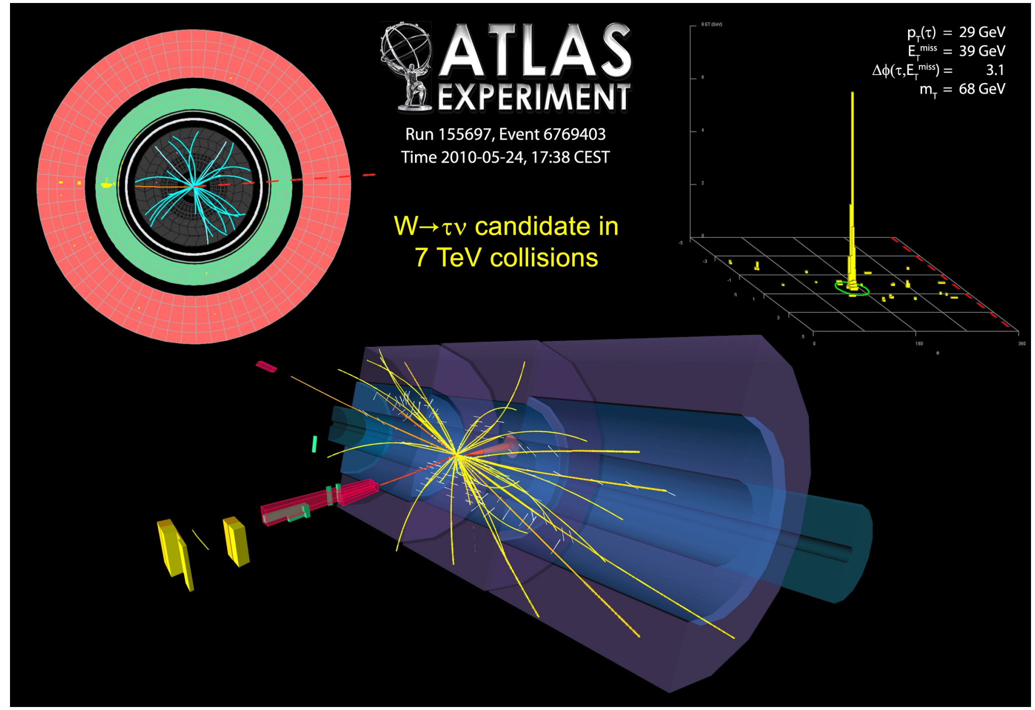
- The "fake rate" should be controlled with data – exact modeling of hadronisation for atypical collimated, low particle quark/gluon jets is difficult
- Shapes of measured and expected fake rate distributions are in good agreement
  - Preliminary studies indicates that observed disagreement depends on used QCD tune



$Z\rightarrow\tau\tau/W\rightarrow\tau\nu$ candidates



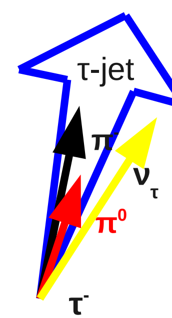
μ p_T = 22.8 GeV/c	T_{had} E_T = 32.9 GeV	Vis. Mass = 60.8 GeV/c^2 M_T(μ,MET) = 10.1 GeV



$Z\rightarrow\tau\tau\rightarrow\mu+\tau$-jet candidate passing all selection criteria observed in dataset corresponding to 70/nb

$W\rightarrow\tau\nu$ candidate

- $p_T(\tau)$ = 29 GeV
- E_T^{miss} = 39 GeV
- m_T = 68 GeV



First $Z\rightarrow\tau\tau/W\rightarrow\tau\nu$ candidates are being found by both experiments.

Conclusions

- An overview of b and tau tagging techniques developed by ATLAS and CMS experiments was shown
- The ATLAS and CMS communities are analysing data being collected this year
  - Detectors performs very well
  - Commissioning of b and tau tagging algorithms with early 7TeV collisions data shows good agreement between data and simulation expectation
- Preparation to analyse physics phenomena with b-quarks and tau-leptons ongoing:
  - First candidates for $tt \rightarrow (bW)(bW)$ and $Z \rightarrow \tau\tau/W \rightarrow \tau\nu$ events are being found
- Candidates for Higgs-like final states approach

References

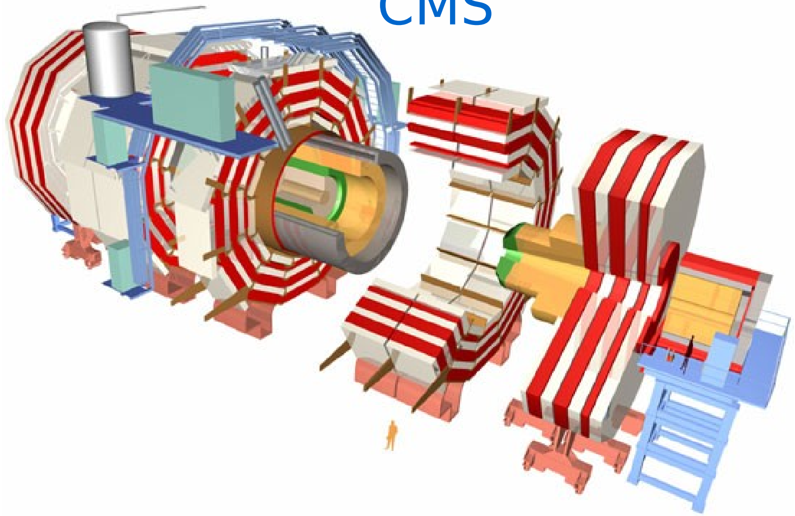
- B-tagging
  - ATLAS
    - ATLAS-CONF-2010-040 (tracking for b-tagging)
    - ATLAS-CONF-2010-041 (impact parameter b-tagging)
    - ATLAS-CONF-2010-042 (secondary vertex b-tagging)
  - CMS
    - CMS PAS BTV-10-001 (tracking for b-tagging, impact parameter and secondary vertex b-tagging)
- Tau identification
  - ATLAS
    - ATLAS-CONF-2010-001 (cut-based tau-Id)
    - ATLAS-CONF-2010-059 (tau-Id with 7TeV collision data)
  - CMS
    - CMS PAS PFT-10-004 (tau-Id fake-rate with 7TeV collision data)

Backup

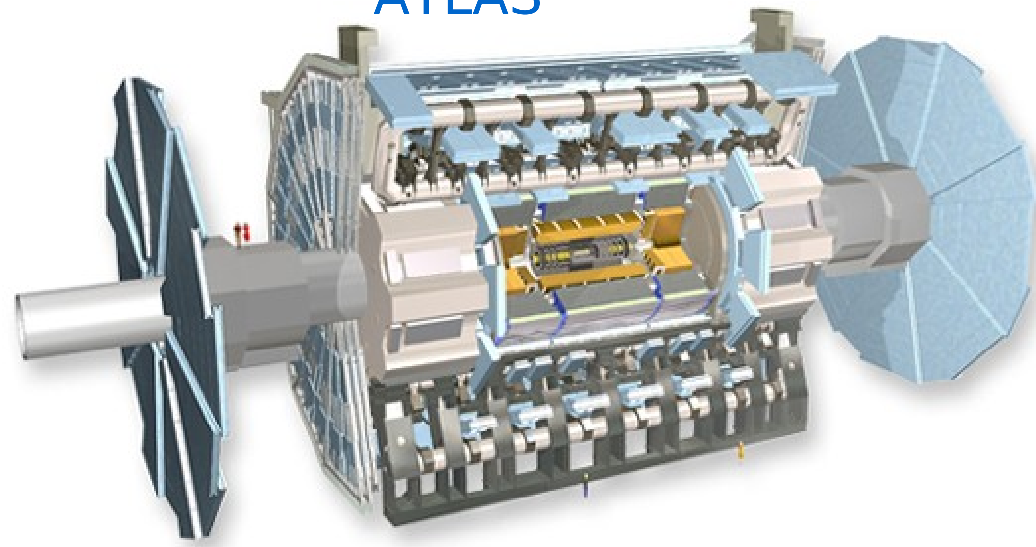
The LHC Experiments

Polytechnique

CMS



ATLAS



Tracking $|\eta|<2.5$, B=3.8T

- Si pixels and strips
  - $\sigma/p_T \approx 1.5 \cdot 10^{-4} p_T \oplus 0.005$

Calorimetry $|\eta|^{EM}<3$, $|\eta|^{HAD}<5$

- EM: homogeneous $PbWO_4$ crystals
  - $\sigma/E \approx 2.8\%/\sqrt{E} \oplus 12\%/E \oplus 0.3\%$
- HAD: Cu-Zn/scint. + Fe/Quartz (5.8λ)
  - $\sigma/E \approx 100\%/\sqrt{E} \oplus 0.05$

Muon Spectrometer $|\eta|<2.4$

- Solenoid return yoke instrumented
- DT/RPC + CSC/RPC

Tracking $|\eta|<2.5$, B=2T

- Si pixels and strips
- Transition radiation detector
  - $\sigma/p_T \approx 5 \cdot 10^{-4} p_T \oplus 0.01$

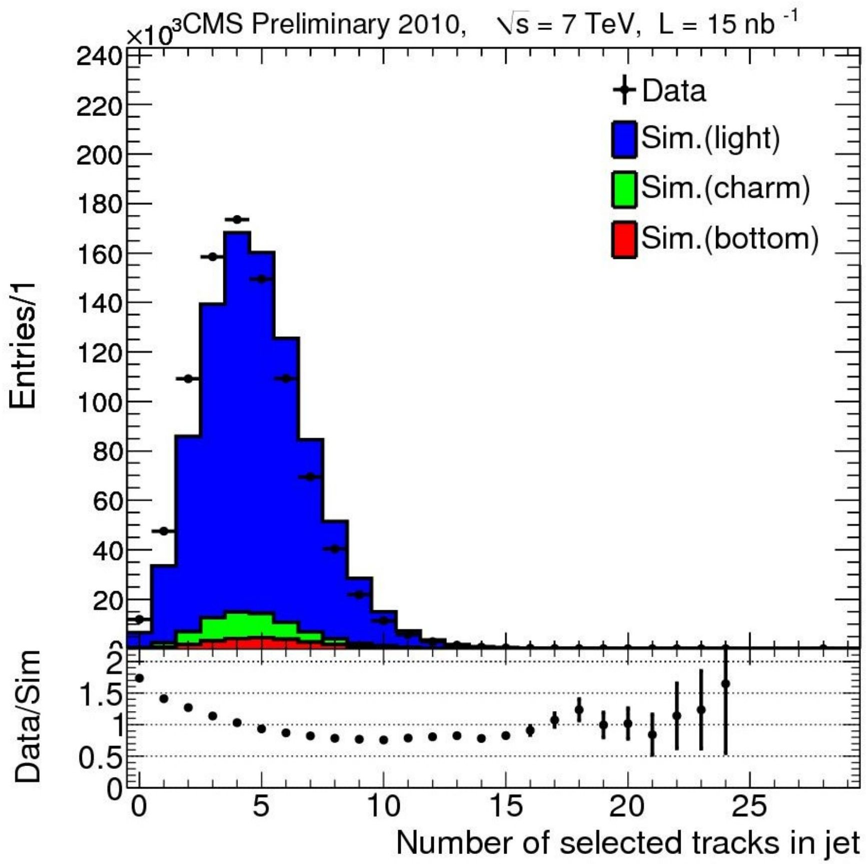
Calorimetry $|\eta|<5$

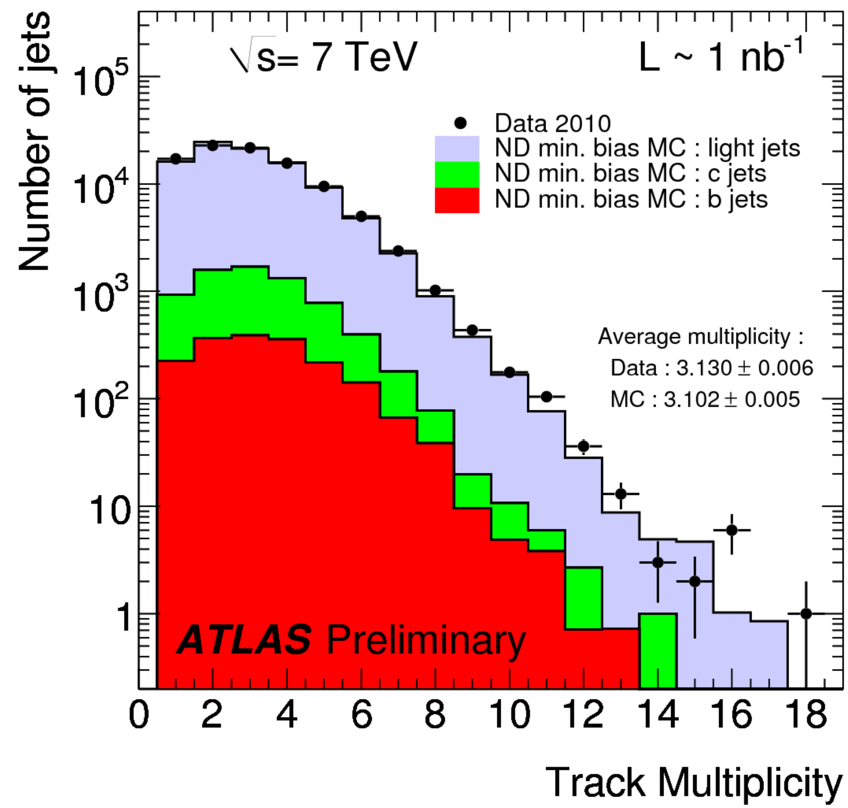
- EM: sampling; Pb/LAr accordeon
  - $\sigma/E \approx 10\%/\sqrt{E} \oplus 0.007$
- HAD: Sampling Fe/scint.+Cu-W/LAr (10λ)
  - $\sigma/E \approx 50\%/\sqrt{E} \oplus 0.03$

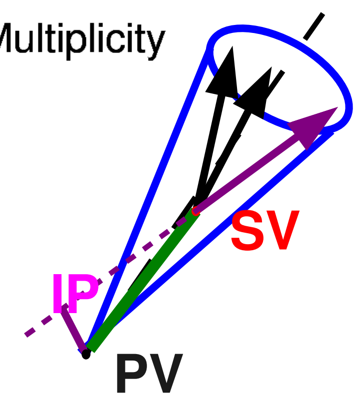
Muon Spectrometer $|\eta|<2.7$

- Air-core toroids with muon chambers

b-tagging, track multiplicity



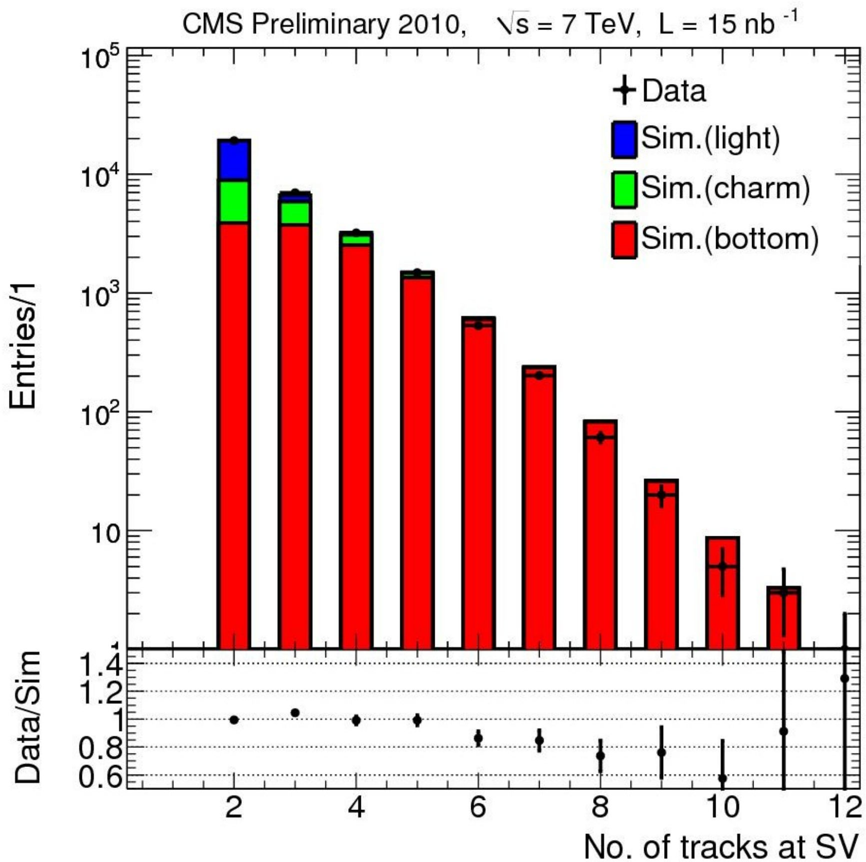


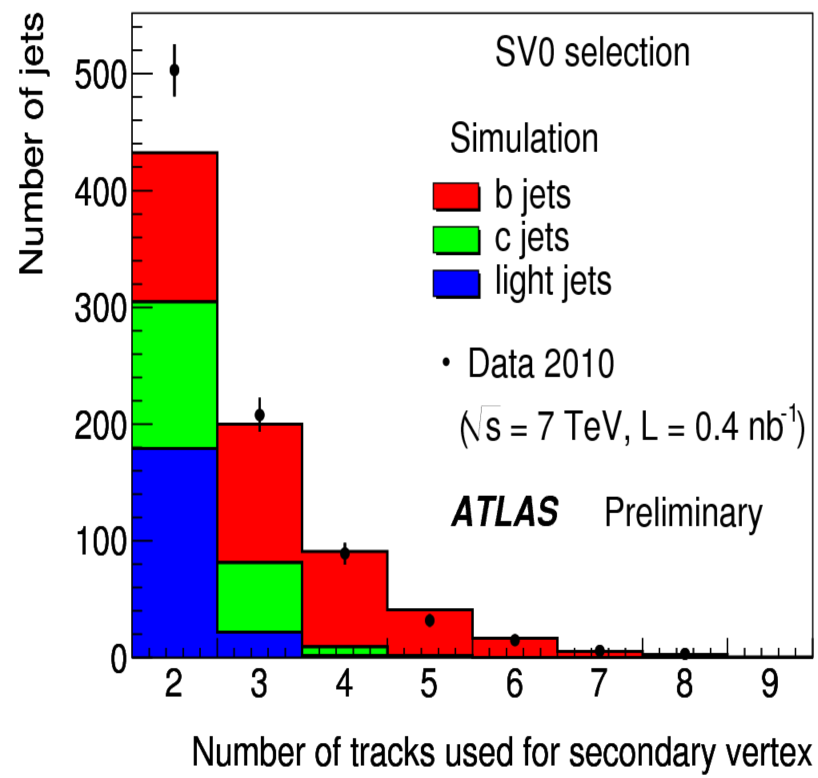


Number of selected tracks associated to jet

- Left, CMS: p_T(jet) > 30 GeV and |η| < 2.4
- Right, ATLAS: p_T(jet) > 20 GeV and |η| < 1.8

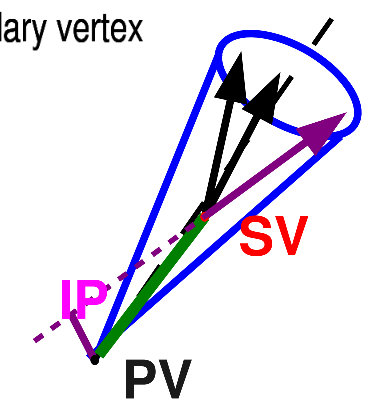
b-tagging, secondary vertex





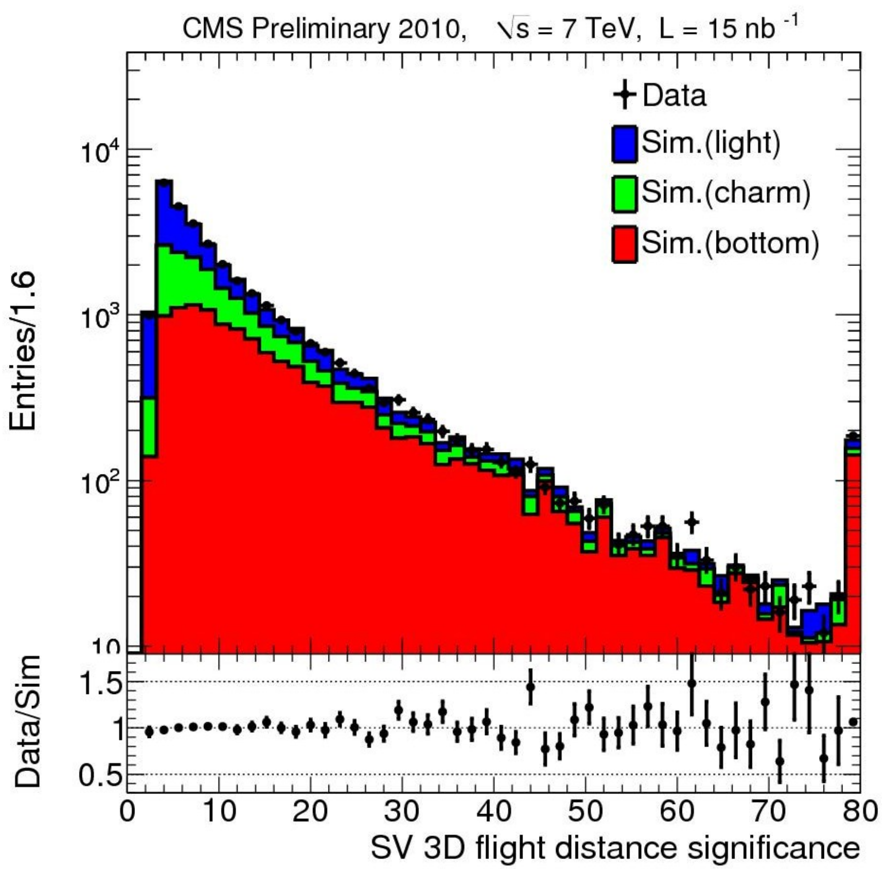
No. of tracks in secondary vertex (SV)

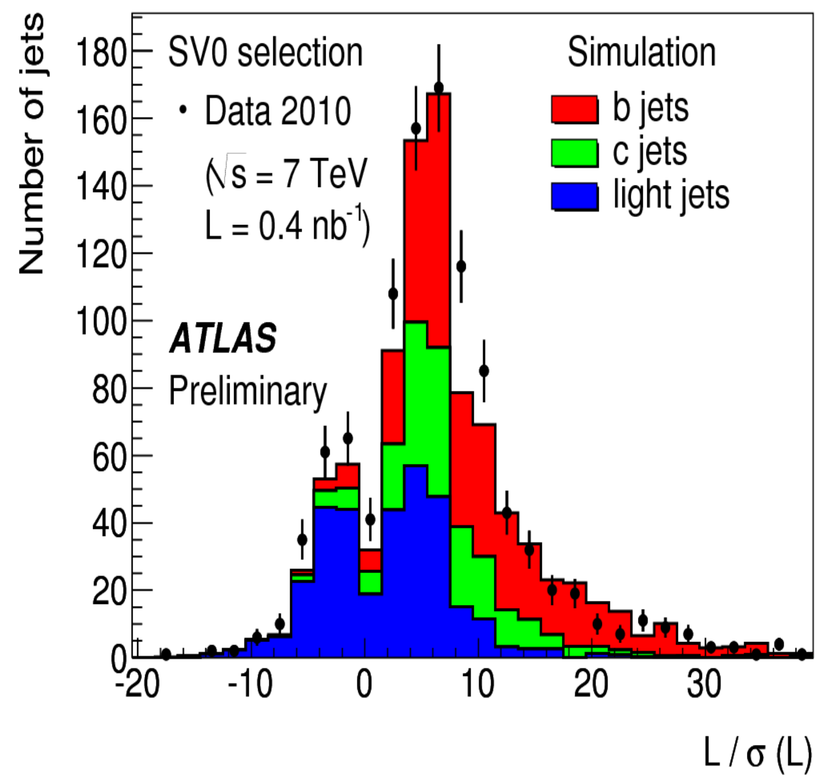
- Left, CMS: p_T(jet) > 30 GeV and |η| < 2.4
- Right, ATLAS: p_T(jet) > 20 GeV and |η| < 1.8

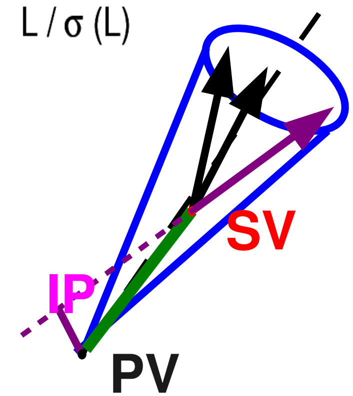


Good agreement between collision data and MC expectations

b-tagging, decay length





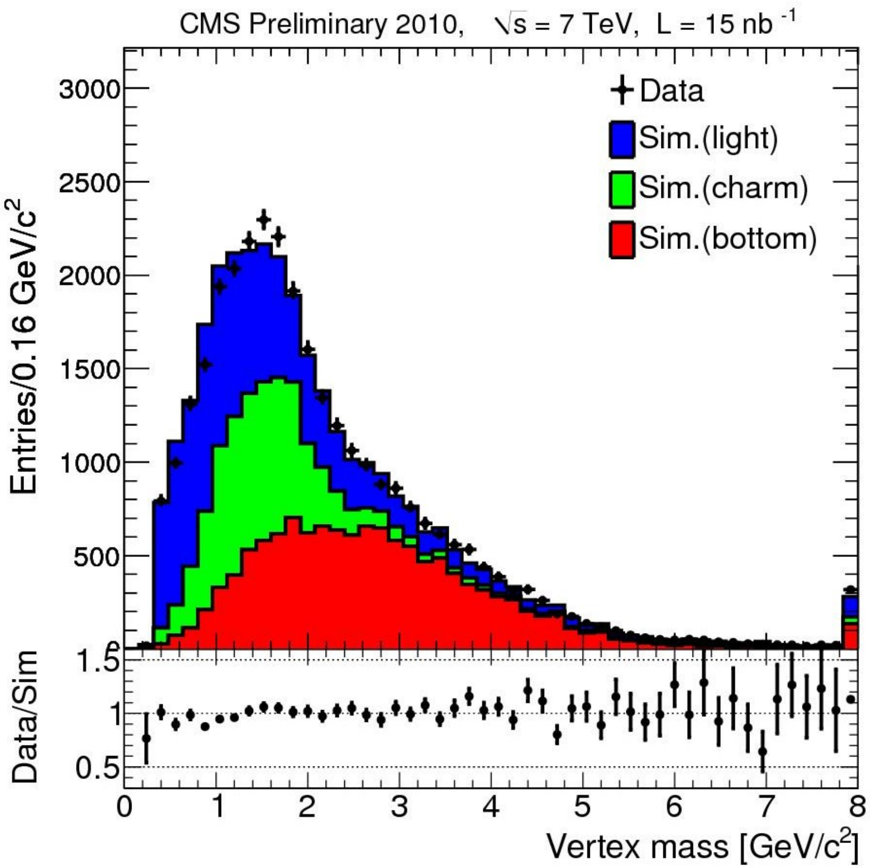


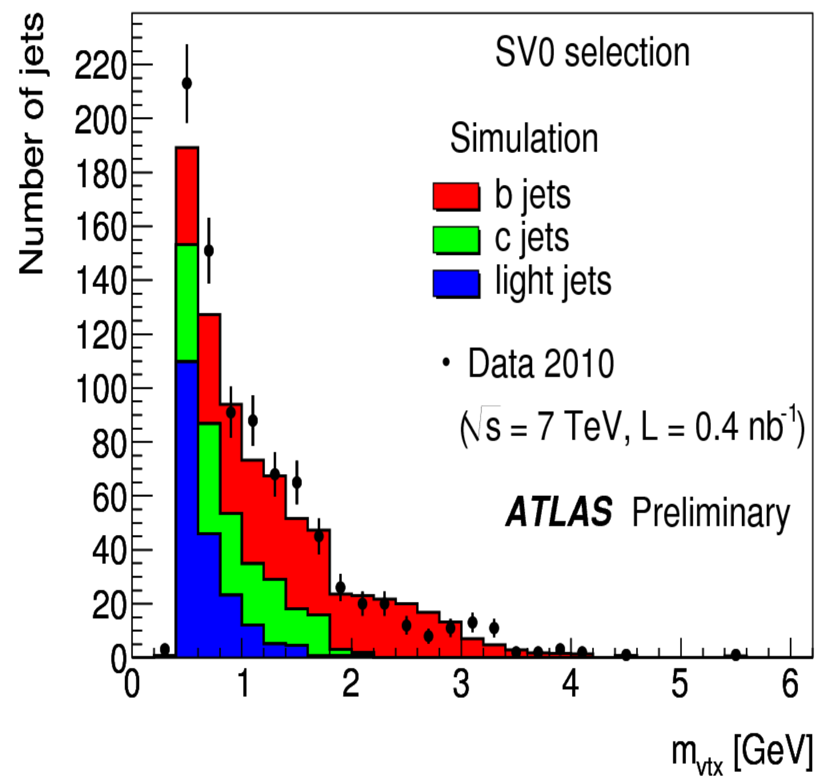
Decay length significance

- Left, CMS: p_T(jet) > 30 GeV and |η| < 2.4
- Right, ATLAS: p_T(jet) > 20 GeV and |η| < 1.8

Good agreement between collision data and MC expectations

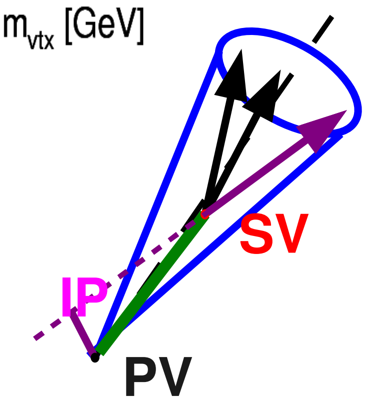
b-tagging, secondary vertex mass





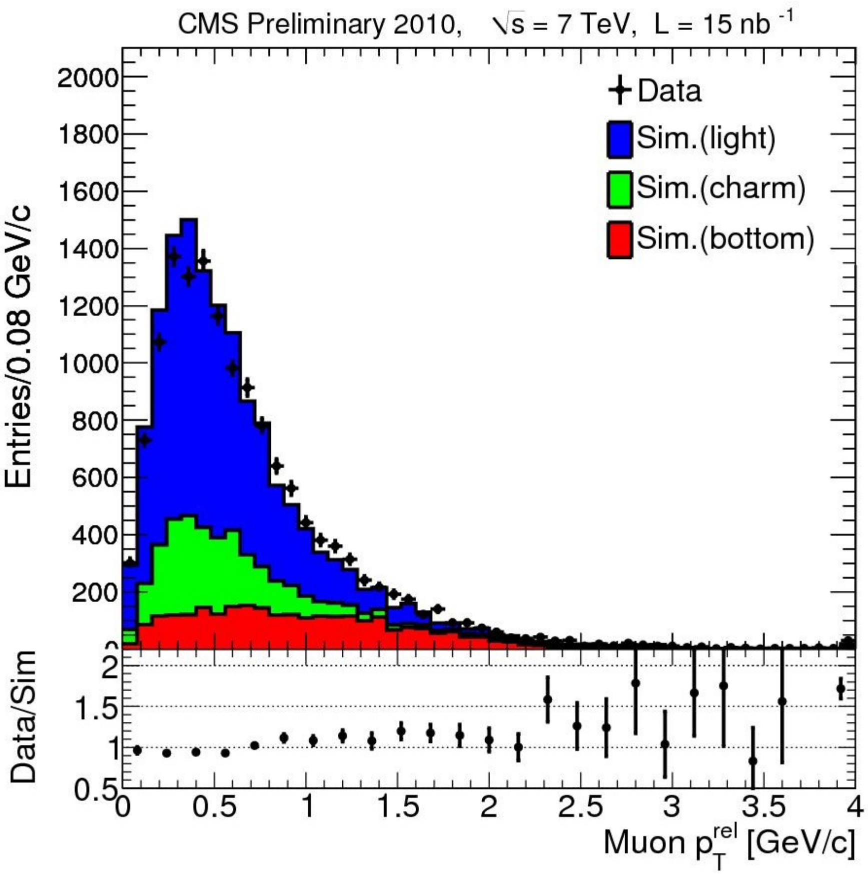
Mass of tracks from secondary vertex

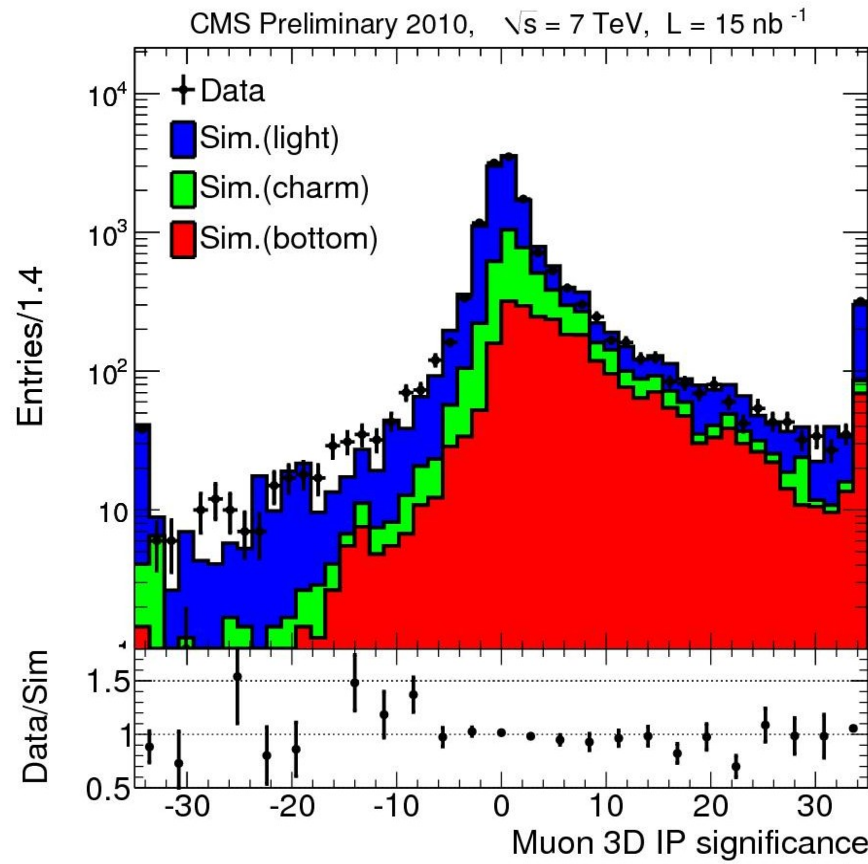
- Left, CMS: p_T(jet) > 30 GeV and |η| < 2.4
- Right, ATLAS: p_T(jet) > 20 GeV and |η| < 1.8



Good agreement between collision data and MC expectations

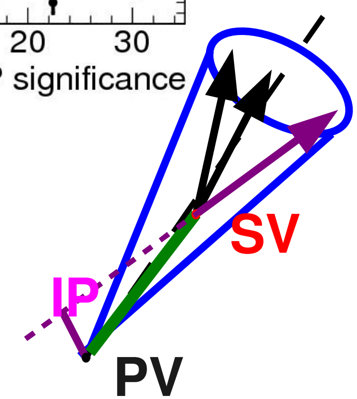
b-tagging, soft lepton





- Left (CMS): relative p_T(mu) w.r.t. jet axis
- Right (CMS): muon IP significance

Similar distributions for electrons



Good agreement between collision data and MC expectations