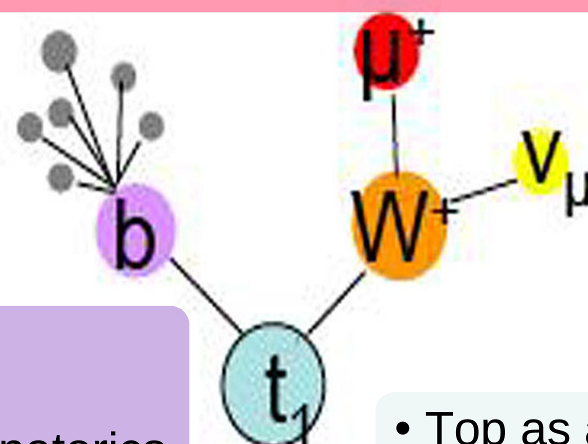
TOP2010: lepton identification, b-tagging and MET reconstruction at CMS after the first data

Joris Maes (jmmaes@cern.ch)
IIHE- Vrije Universiteit Brussel
On behalf of the CMS collaboration

Top quarks

Top quark physics involves reconstruction of objects using all subdetectors

- Leptons:
  - Used for triggering
  - Background suppression (opposite charge, isolation, identification)
  - High p_T, isolated leptons: from W decay
  - Soft, non isolated leptons: in b-jets



- MET
  - Top mass reconstruction
  - Background suppression
  - Sensitive to new physics

- b-tagging:
  - Identifying b jets
  - Resolving jet combinatorics
  - Background suppression

- Top as a calibration tool:
  - JES calibration
  - b-tag performance measurement

Reconstruction of jets: see Roger Wolf tomorrow

Data in this presentation

Cosmic muon data collected by the CMS detector

- 2008, Oct-Nov: Cosmic Run at 4 Tesla (CRAFT08):
  290M events with magnetic field of 3.8 Tesla. (23 Papers published)
- 2009, Summer: Cosmic run, very similar to CRAFT08

Collision data in CMS from 3 week running period in 2009 Nov-Dec:

- 900 GeV
- 2360 GeV

Collision data in CMS from last weeks/months:

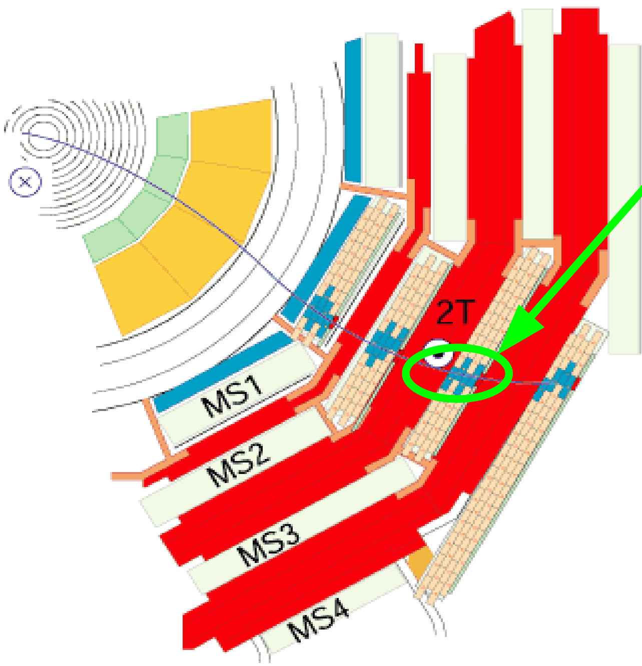
- 7 TeV data

Lepton section outline

- Muon
  - Reconstruction
  - Alignement using cosmic data
  - MC vs. data comparison
- Electron
  - Reconstruction
  - MC vs. collision data comparison
- Tag and Probe method

https://twiki.cern.ch/twiki/bin/view/CMS/PublicPhysicsResults

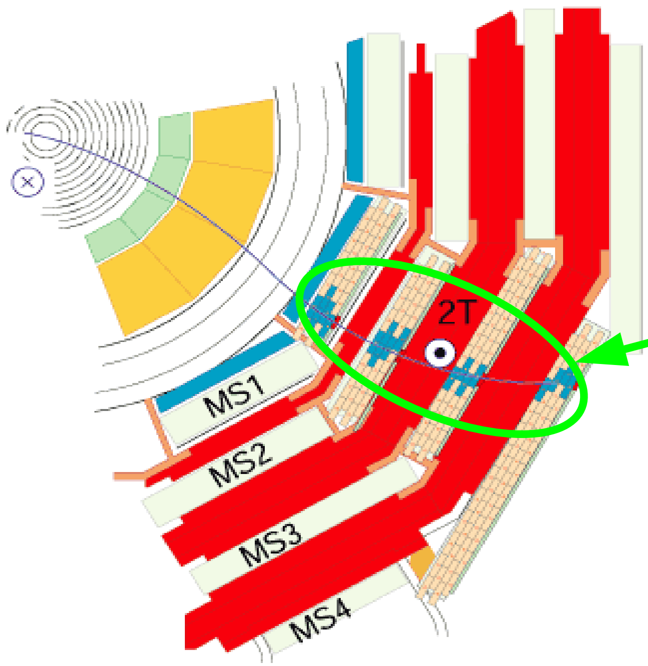
Muon - reconstruction



Muon reconstruction in 3 stages

- **Local muon** reconstruction
  - Hits are reconstructed in subdetectors (CSC, DT and RPC)
  - Make track segments from hits
- **Standalone muon** reconstruction
  - Combine track segments in Kalman Filter
  - Builds muon trajectory in muon system
- **Global muon** reconstruction
  - Combines silicon tracker information with muon system
  - Build global muon trajectory

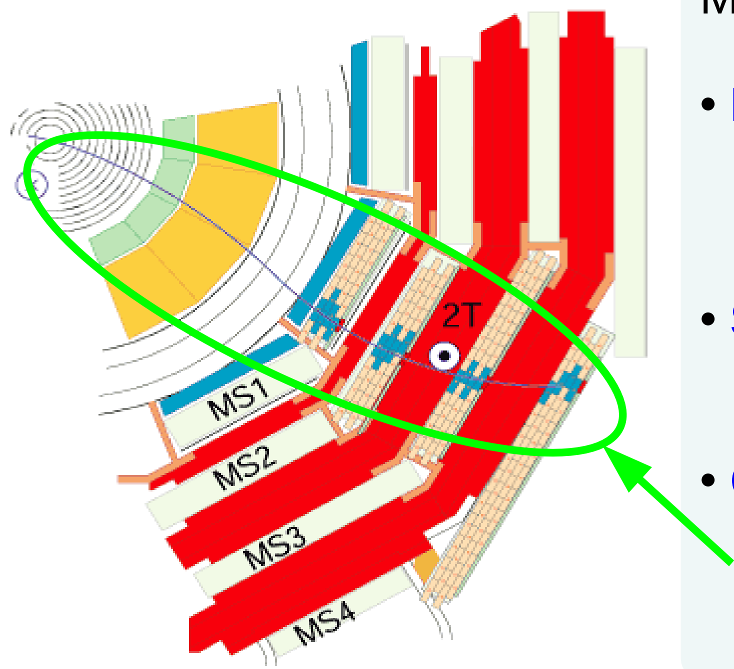
CRAFT data is used to **align the detector** components and study their **performance**



Muon reconstruction in 3 stages

- **Local muon** reconstruction
  - Hits are reconstructed in subdetectors (CSC, DT and RPC)
  - Make track segments from hits
- **Standalone muon** reconstruction
  - Combine track segments in Kalman Filter
  - Builds muon trajectory in muon system
- **Global muon** reconstruction
  - Combines silicon tracker information with muon system
  - Build global muon trajectory

CRAFT data is used to **align the detector** components and study their **performance**

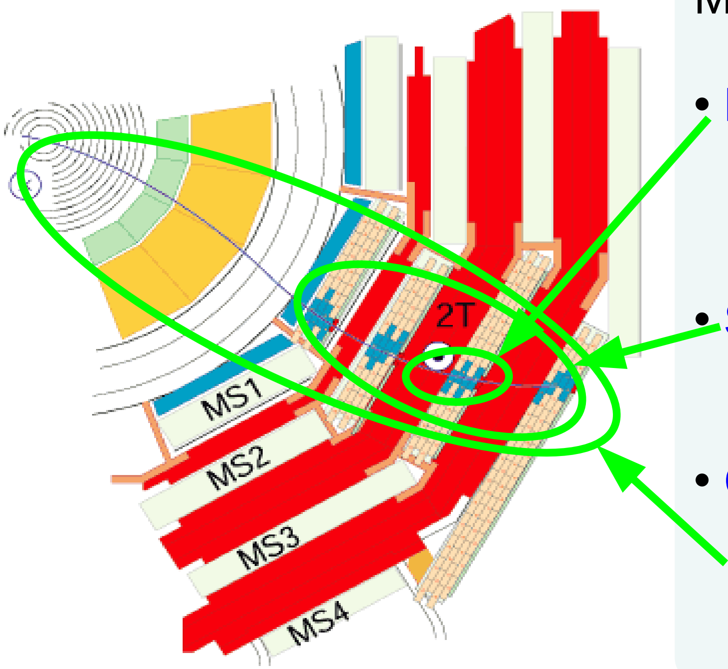


Muon reconstruction in 3 stages

- **Local muon** reconstruction
  - Hits are reconstructed in subdetectors (CSC, DT and RPC)
  - Make track segments from hits
- **Standalone muon** reconstruction
  - Combine track segments in Kalman Filter
  - Builds muon trajectory in muon system
- **Global muon** reconstruction
  - Combines silicon tracker information with muon system
  - Build global muon trajectory

CRAFT data is used to **align the detector** components and study their **performance**

Muon - reconstruction

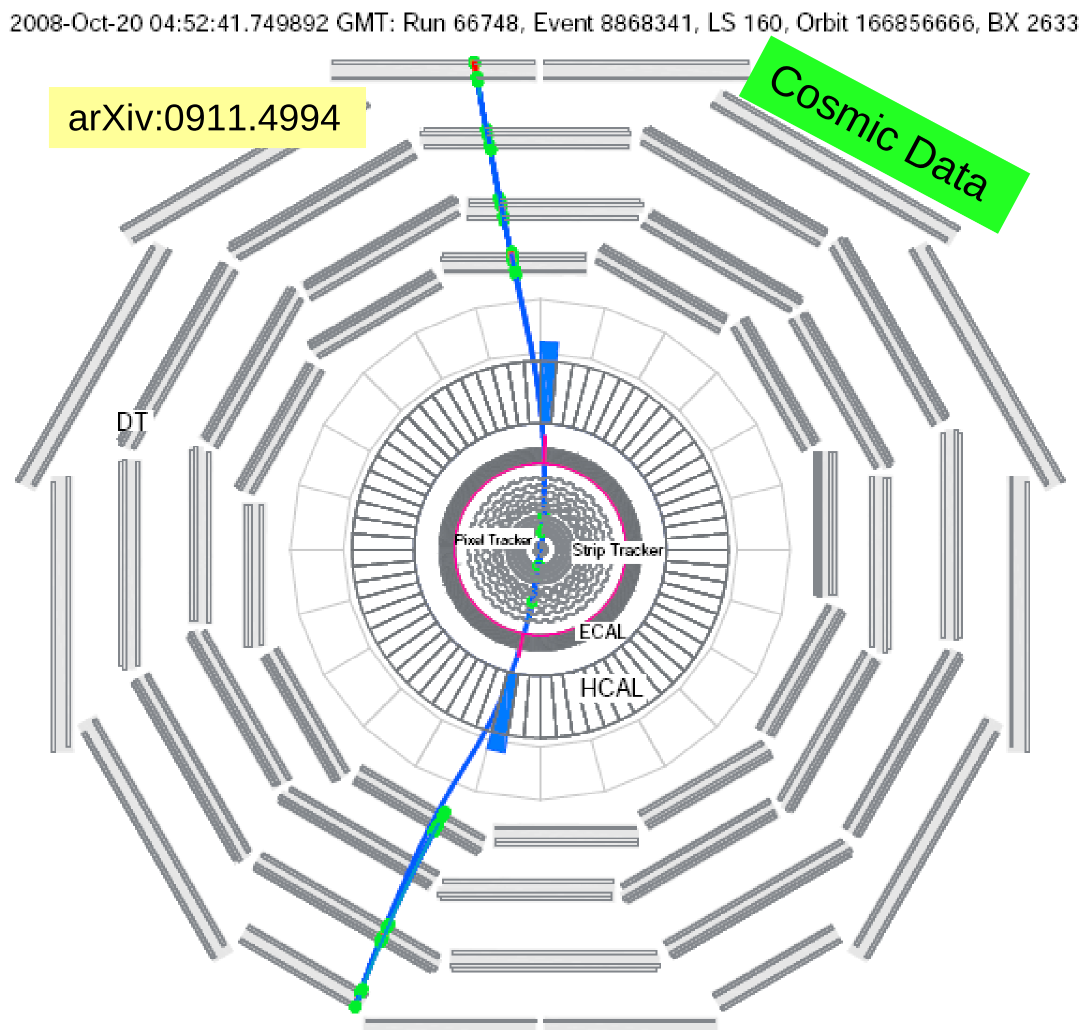


Muon reconstruction in 3 stages

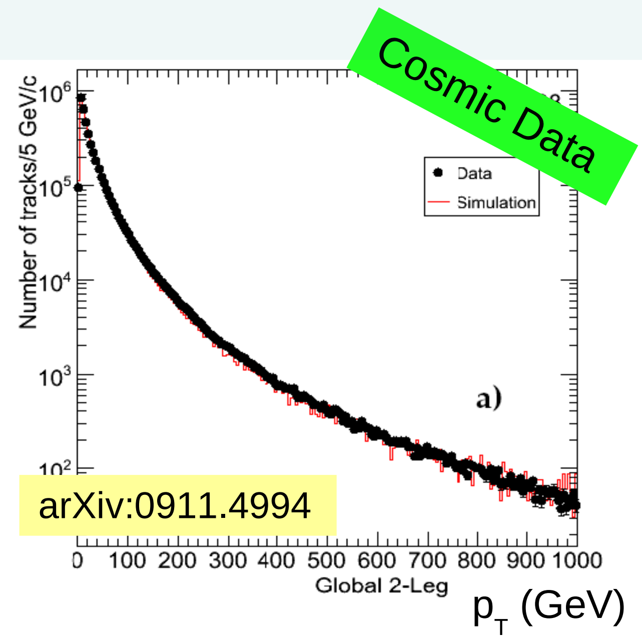
- **Local muon** reconstruction
  - Hits are reconstructed in subdetectors (CSC, DT and RPC)
  - Make track segments from hits
- **Standalone muon** reconstruction
  - Combine track segments in Kalman Filter
  - Builds muon trajectory in muon system
- **Global muon** reconstruction
  - Combines silicon tracker information with muon system
  - Build global muon trajectory

CRAFT data is used to **align the detector** components and study their **performance**

Event display of a cosmic muon track with magnetic field of 3.8 Tesla



- Comparison between dedicated MC simulation and data
- Muon required to pass closely by the nominal interaction point
- "2-leg" muon: reconstructed 2 tracks in the opposite hemispheres to mimic collision muons

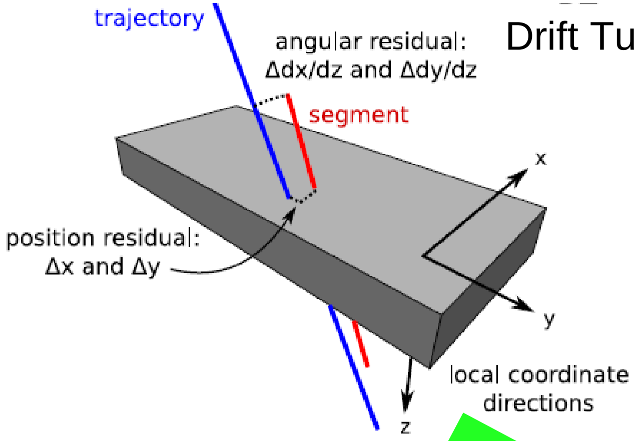


Muon system alignment

After the alignment of the individual modules with data, the **muon system is aligned w.r.t. to the tracker**.

- Muon tracks are reconstructed with tracker only
- Chamber residuals like Δx, Δy, Δdx/dz, Δdy/dz are the differences between the predicted track and the observed track
- Alignment procedure performed with tracks that pass through both tracker and muon system with 100 GeV/c < p_T < 200 GeV/c

Drift Tubes

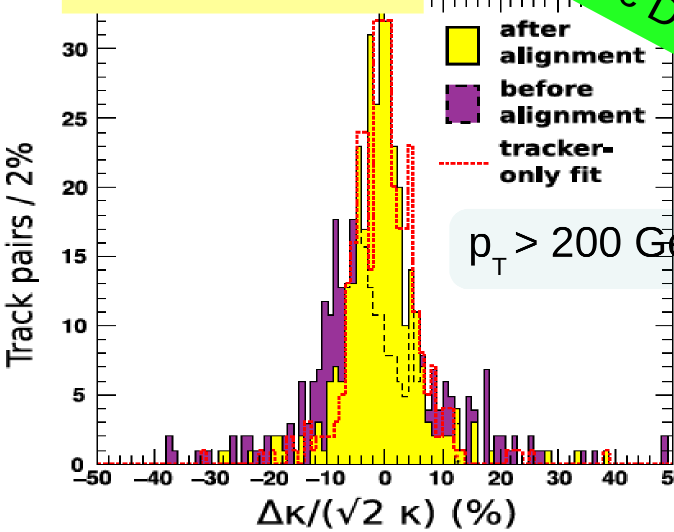


- Alignment performance for momentum shown
- The top half and bottom half of the cosmic ray trajectory are reconstructed separately ("2-leg")
- Difference between tracks is purely instrumental
- The fractional difference in curvature $\kappa = q/p_T$ is plotted before and after alignment

Difference in reconstructed p_T of the muons is much smaller after alignement

arXiv:0911.4022

Cosmic Data



p_T > 200 GeV/c

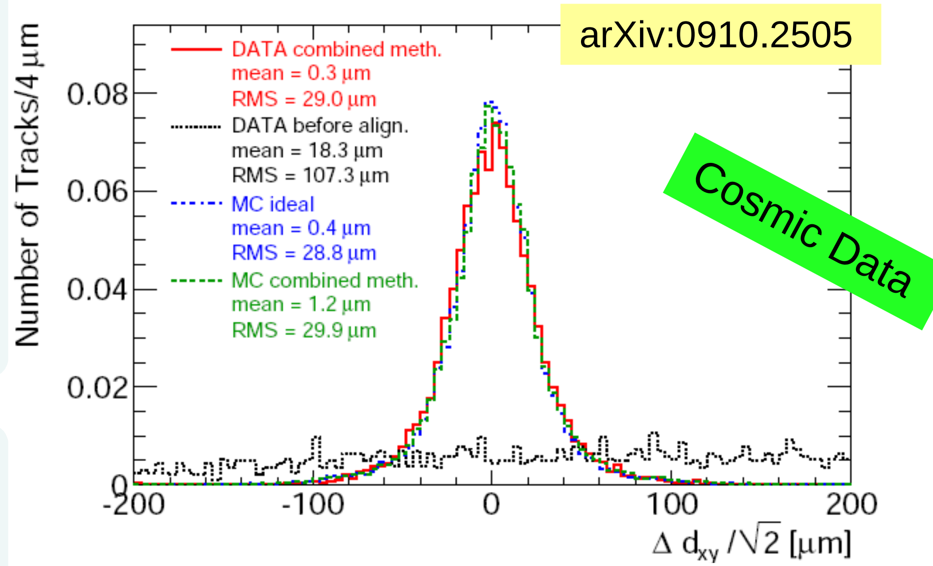
Silicon tracker alignment

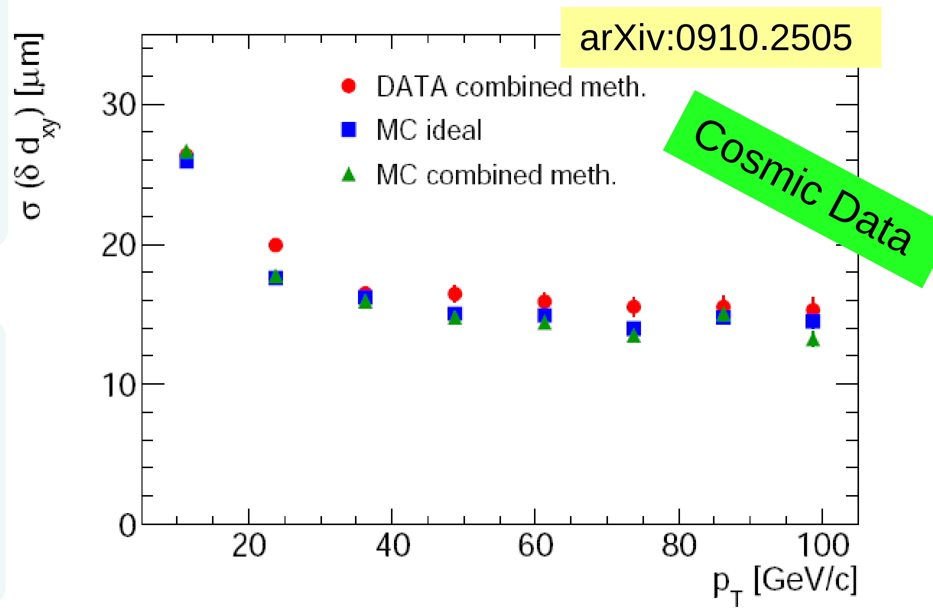
- Track reconstructed with "combinatorial track finder" algorithm
- Helix trajectory described by five parameters
  - Longitudinal/transverse direction
  - Azimuthal/polar angle
  - Transverse momentum

- Determine the silicon module positions with **track based alignment**
- Alignment is performed by minimizing the residuals between the hit position and the trajectory impact point:
  - "Hits and Impact Points"-algorithm
  - "Millepede II"-algorithm

- Track parameter resolutions validated with an independent reconstruction of upper and lower legs of cosmic ray tracks

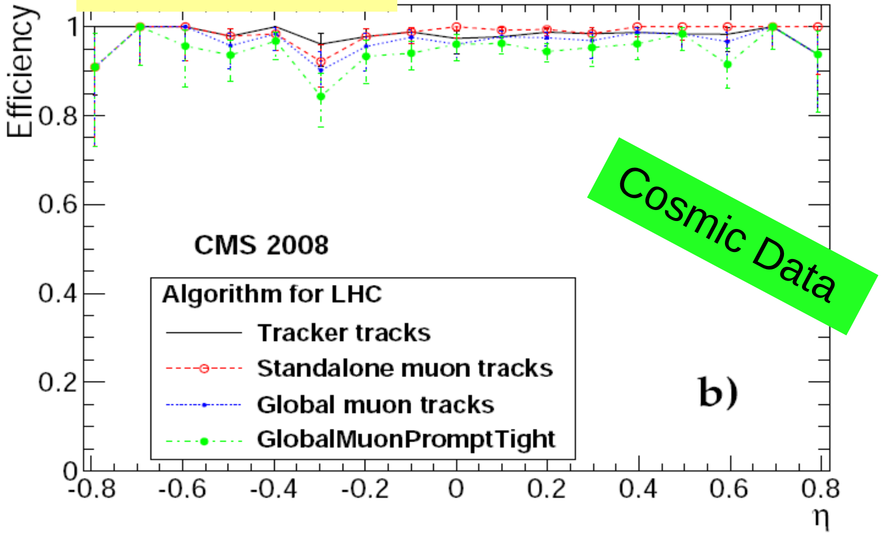
Performance already very close to ideal performance from MC simulations



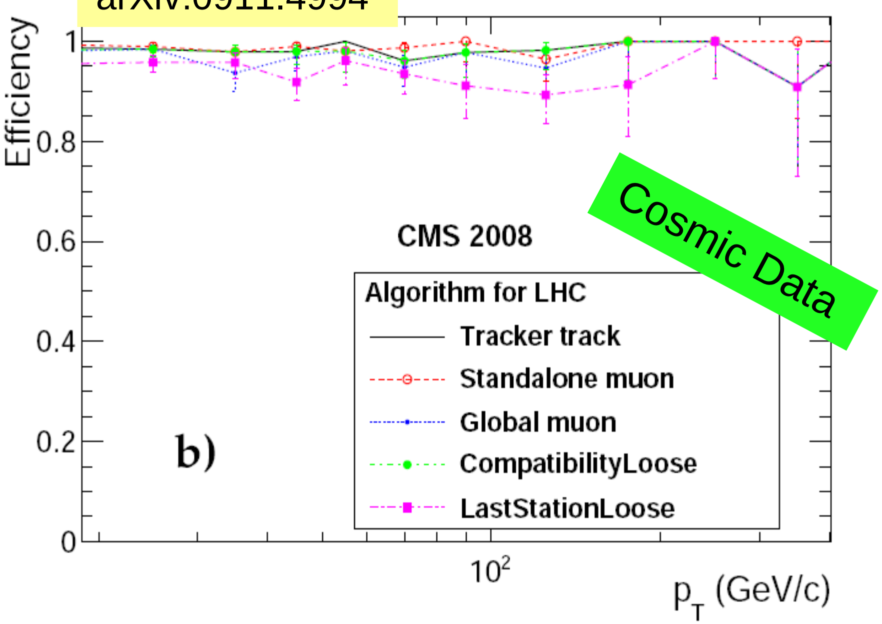


Muon reconstruction efficiency

arXiv:0911.4994



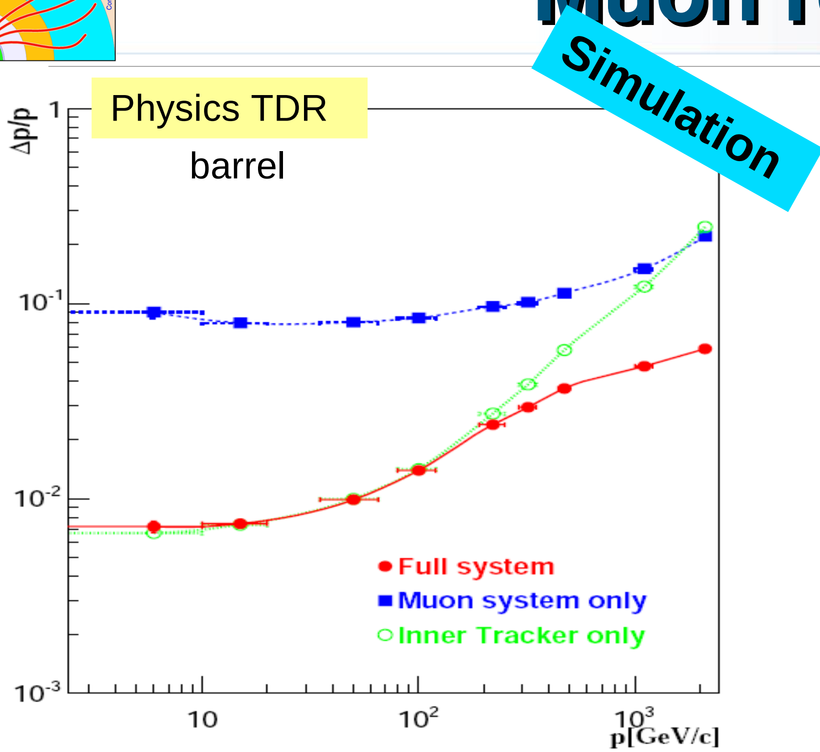
arXiv:0911.4994

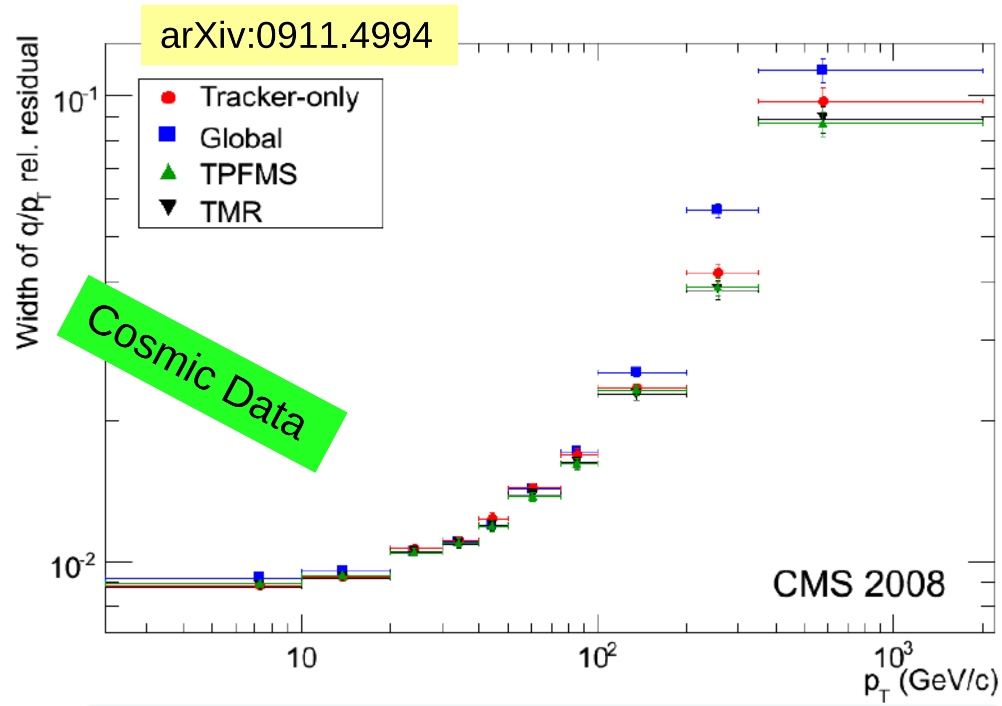


- Select a good quality muon in one of the two hemispheres
- Efficiency is obtained by searching for a matching muon in the opposite hemisphere with $|\Delta\varphi|<0.3$ and $|\Delta\eta|<0.3$ w.r.t direction of global muon track
- p_T >10 GeV/c to traverse whole detector

- **The efficiency for global muons in the barrel region is (97.1 +- 0.6)%**
- No η- and p_T-dependency is found

Muon resolution





"2-leg" muons with

- > 8 tracker hits
- > 1 pixel hit
- exactly 2 tracks

→ collision like muons

At p_T < 200 GeV/c inclusion of tracker hits in muon reconstruction improves the resolution significantly

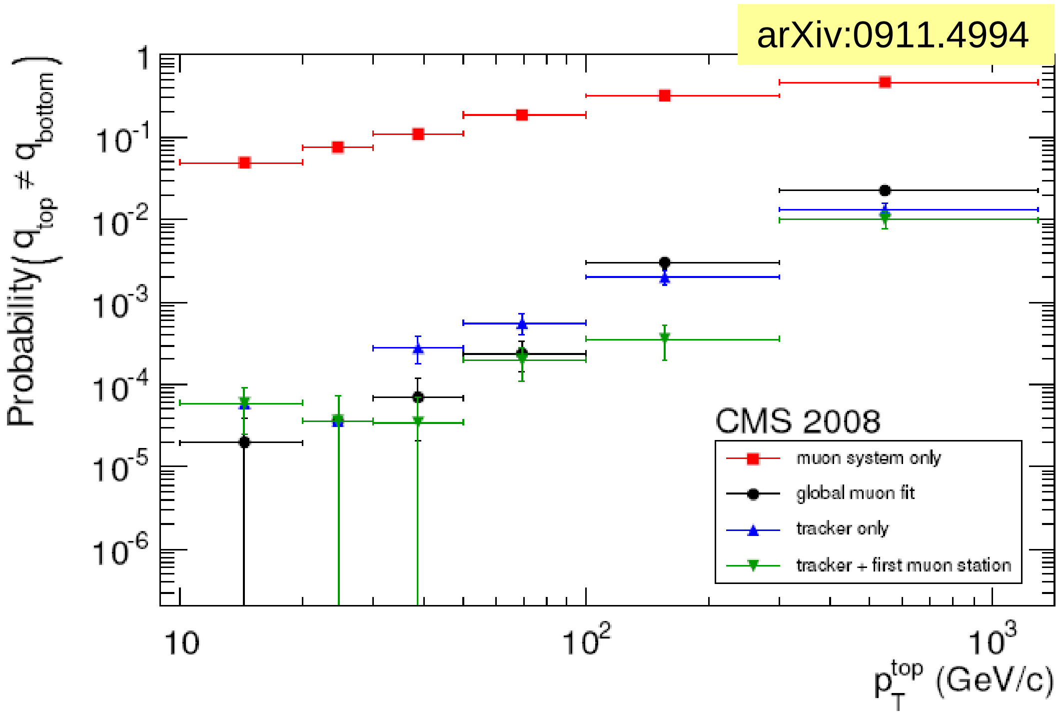
Relative momentum resolution for muons crossing the barrel after tracker and muon system alignment

- **<1% at 10 GeV/c**
- **~8% at 500 GeV/c**

Only a factor of two worse than expected with ideal alignment conditions

Reconstructed charge mis-assignment: probability that the charge of the same muon in top and bottom hemisphere disagree

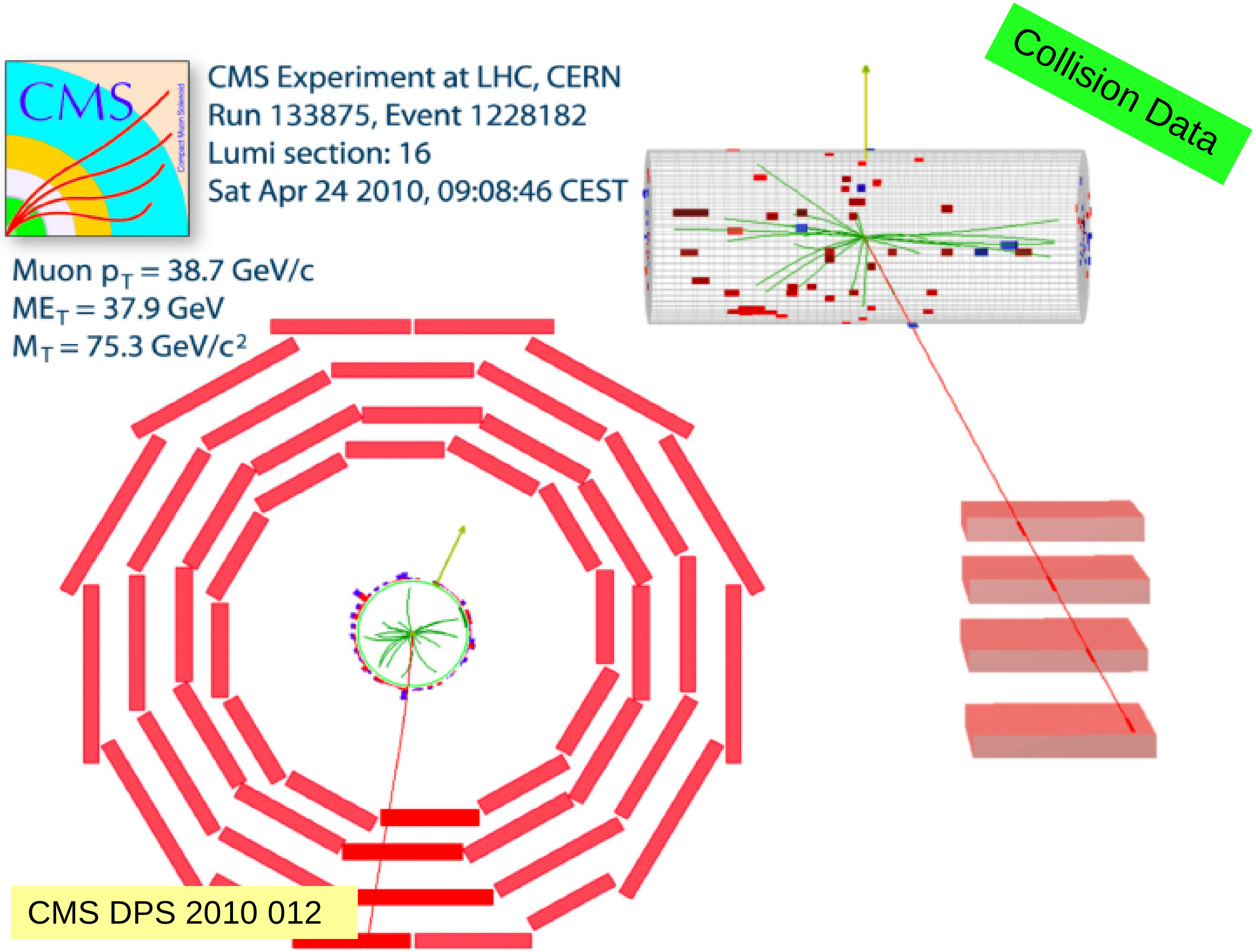
- Tag: muon track in the upper hemisphere
- Probe: charge of the other track in the bottom hemisphere
- Only barrel muon system and central tracker used in plot
- Reconstruction of charge is much improved for global muons



Charge mis-assignment:
- < 0.01% at 100 GeV/c
- ~1% at 500 GeV/c

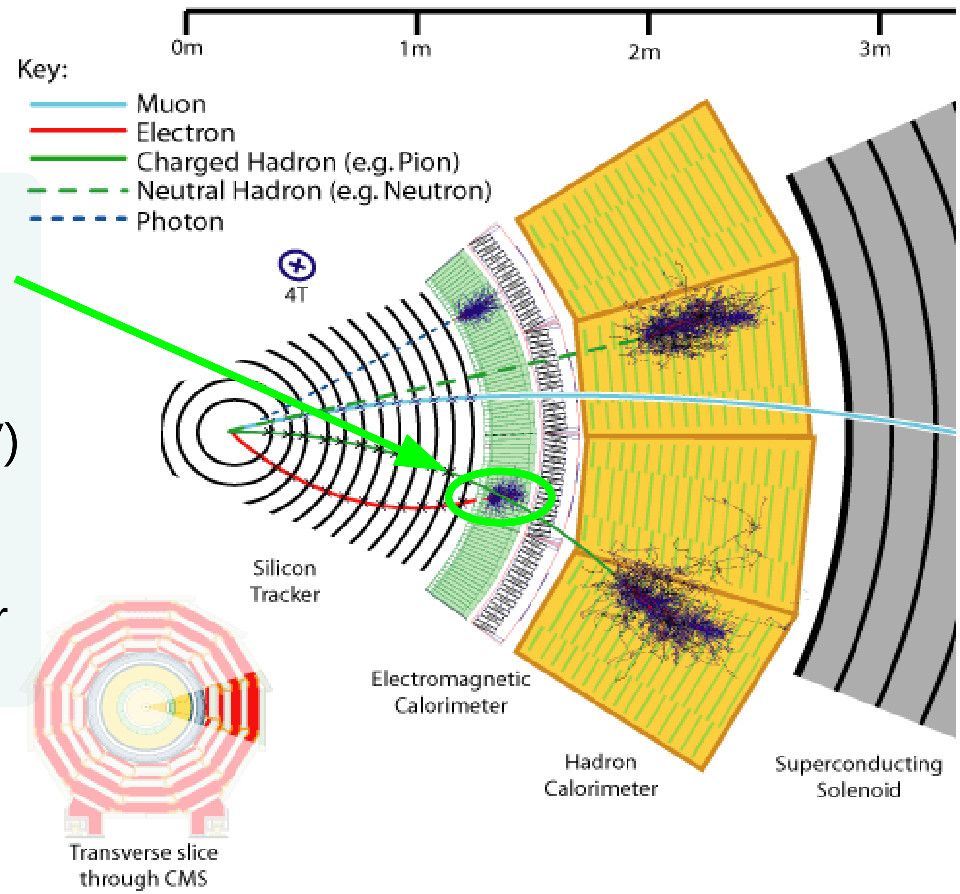
Very good charge assignment observed

Event display of a $W^{+} \rightarrow \mu^{+}\nu_{\mu}$ candidate in 7 TeV collision data



Electron reconstruction:

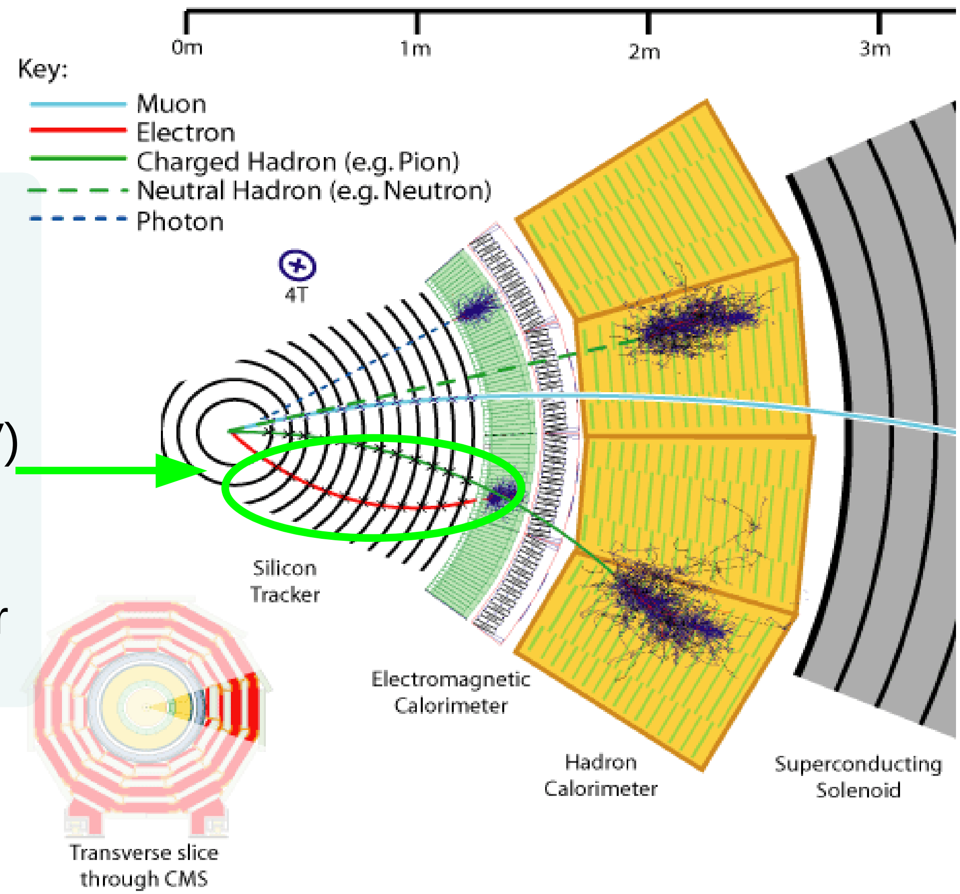
- ECAL driven, starting from ecal super clusters E_T>4 GeV, optimized for high p_T (>5GeV) isolated electrons from W/Z
- Tracker driven, low p_T electrons (<5 GeV) and non-isolated electrons inside jets
- Tracks are reconstructed using a Gaussian-Sum filter (GSF) to account for Bremsstrahlung energy losses



Electron reconstruction

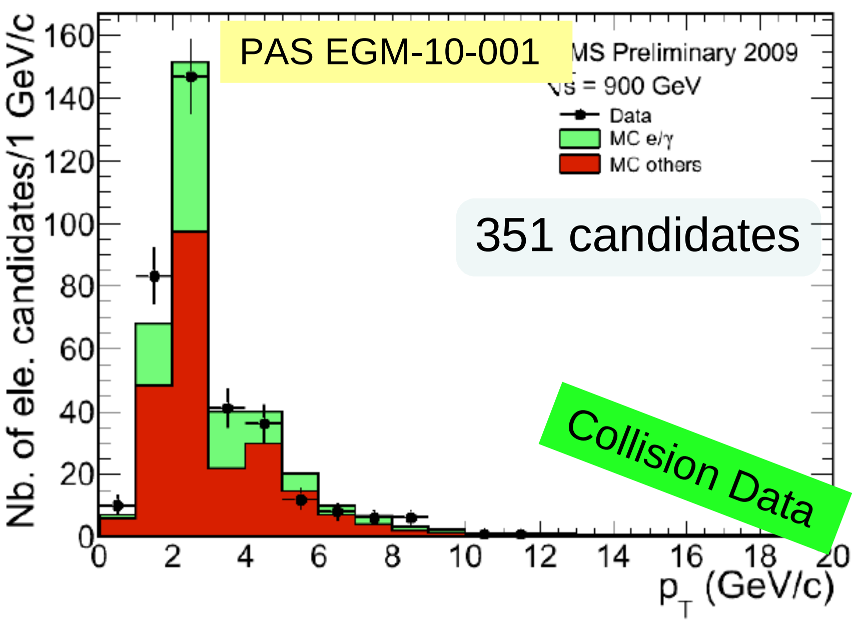
Electron reconstruction:

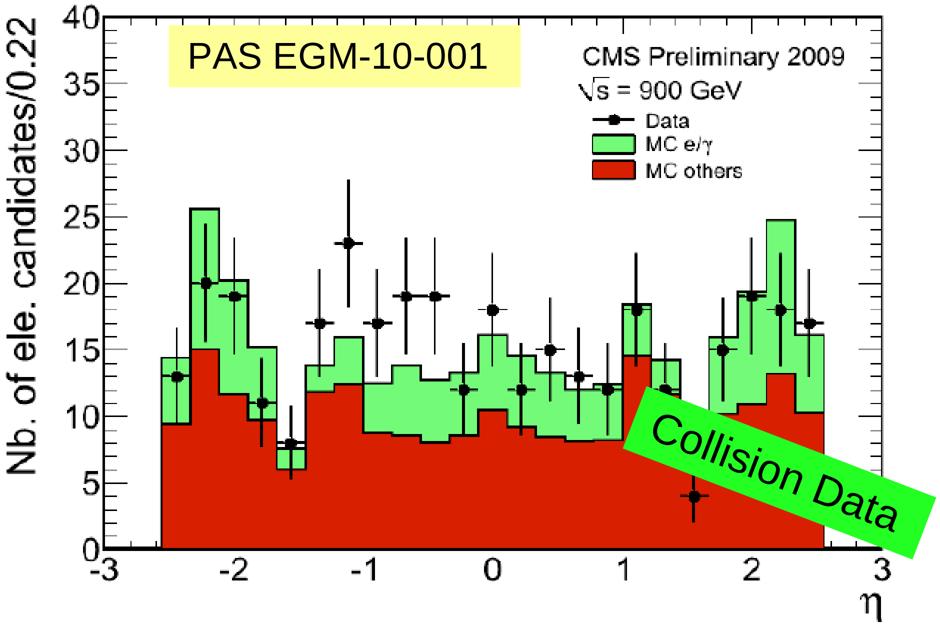
- ECAL driven, starting from ecal super clusters E_T>4 GeV, optimized for high p_T (>5GeV) isolated electrons from W/Z
- Tracker driven, low p_T electrons (<5 GeV) and non-isolated electrons inside jets
- Tracks are reconstructed using a Gaussian-Sum filter (GSF) to account for Bremsstrahlung energy losses



Electrons in collision data

- Electron candidates reconstructed in **900 GeV proton collision data**
- Both ECAL and tracker driven electron candidates are included
- Good agreement found between η and p_T of the electron candidates and MC
- Most of the electron and photons candidates are due to fakes
- Real electrons are non-isolated electrons from photon conversions





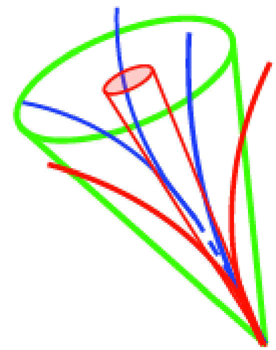
Electron - isolation

Isolation is quantified by the activity surrounding the electron

ECAL isolation:

• transverse energy sum in $\Delta R = 0.4$

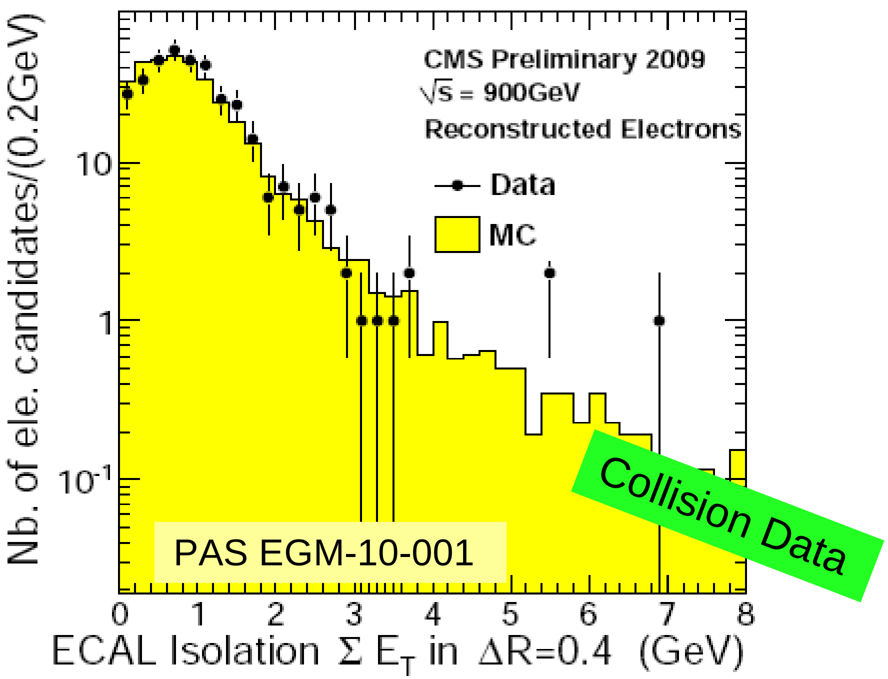
• inner veto cone radius corresponds to size of 3 ECAL crystals (ΔR=0.05 in barrel)

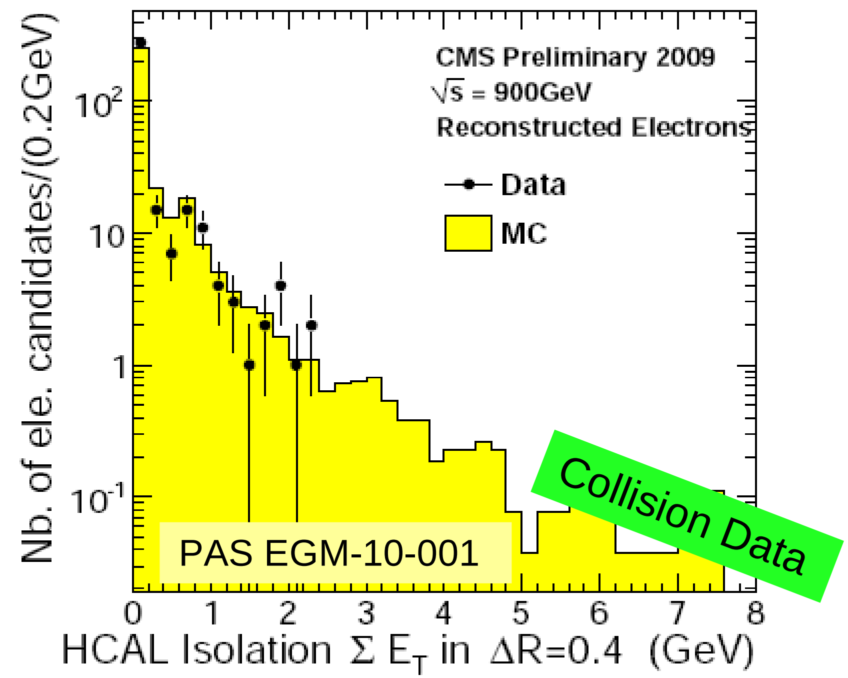


HCAL isolation:

• transverse energy sum $\Delta R = 0.4$, veto cone $\Delta R = 0.15$

•The energy is summed of towers which have an energy greater than 0.7 GeV in the barrel



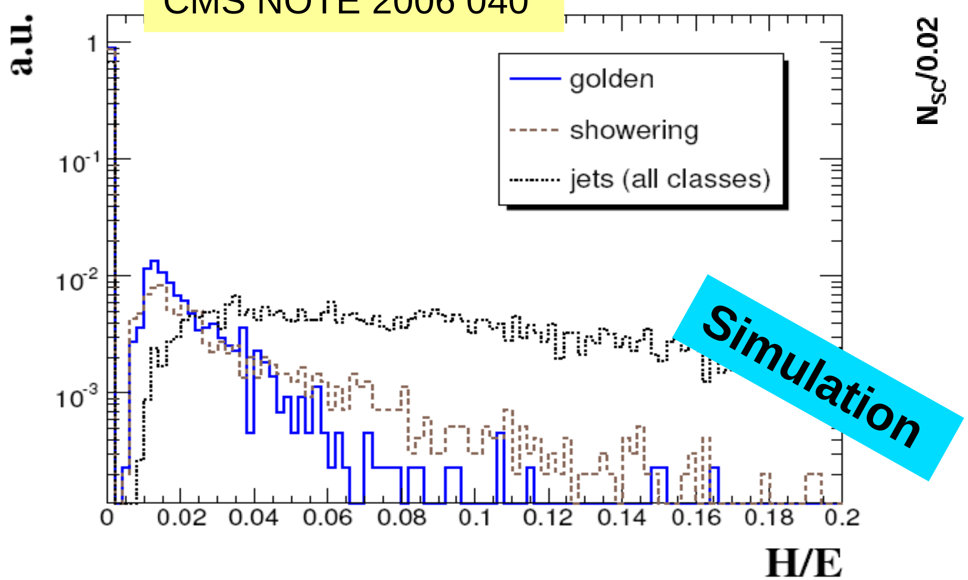


PAS TOP-09-004

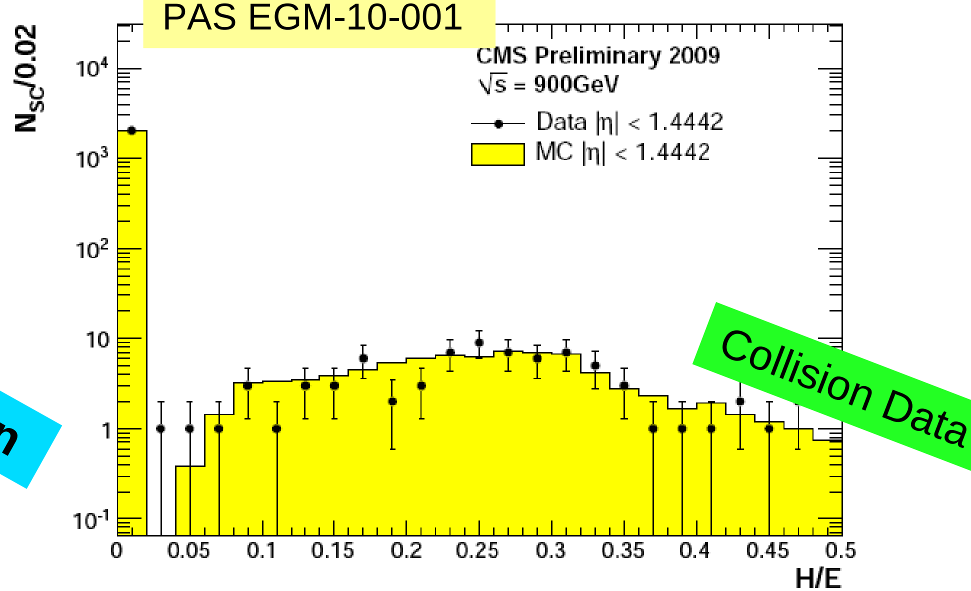
Electron identification (ID) variables used in startup study of tt cross section

- The differences between φ and η of the supercluster and the GSF track
- Ratio of energy measured in the HCAL behind the ECAL shower: H/E
- Shower shape in η
- Variables are insensitive to tracker misalignment

CMS NOTE 2006 040



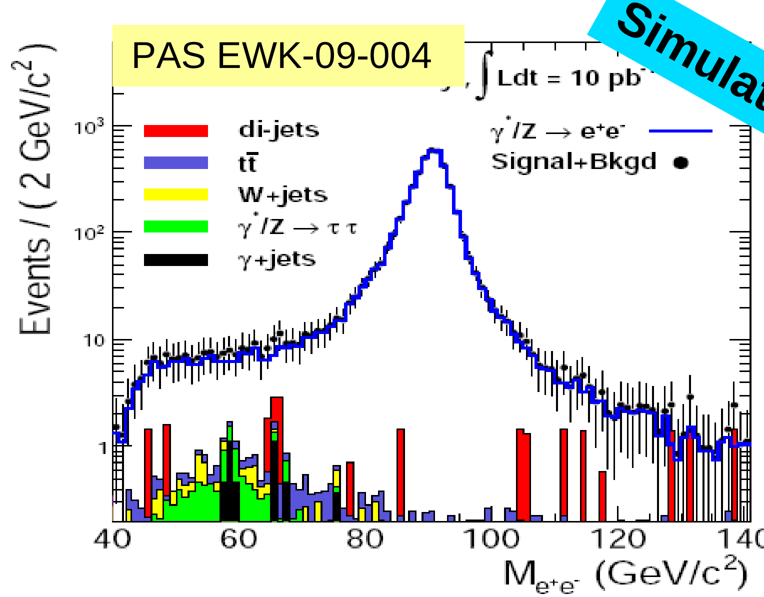
PAS EGM-10-001

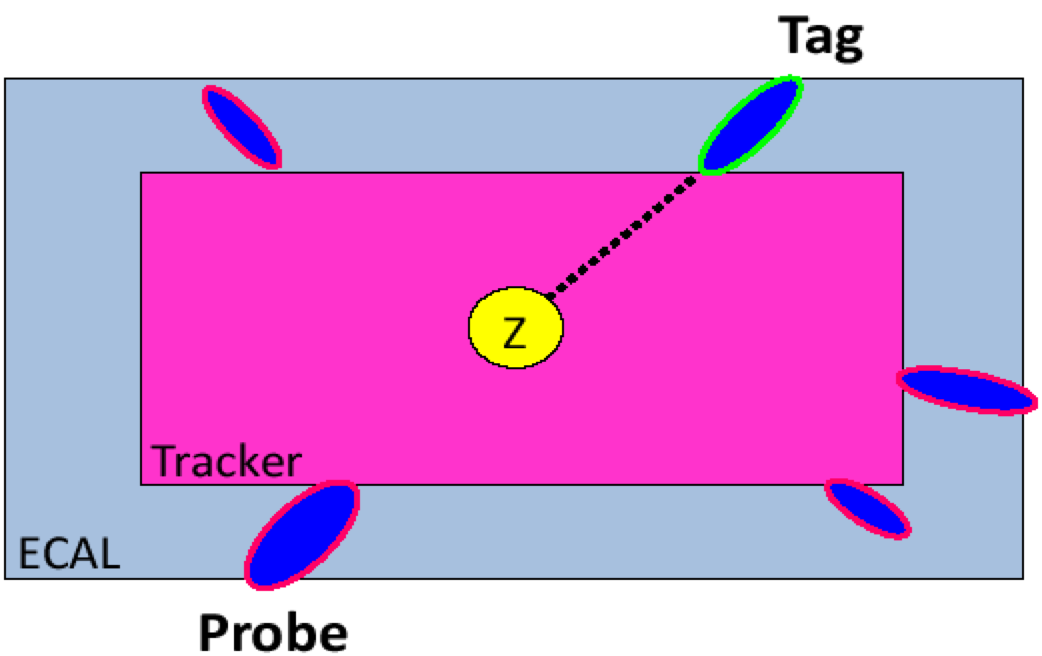


Tag and probe methods

Method to obtain **trigger and reconstruction efficiency** of leptons
$Z \rightarrow e^+e^-$ (shown)
$Z \rightarrow \mu^+\mu^-$

- Electron-pair must be in Z-mass window
- High electron purity can be reached with thight ID cut on the tagged electron

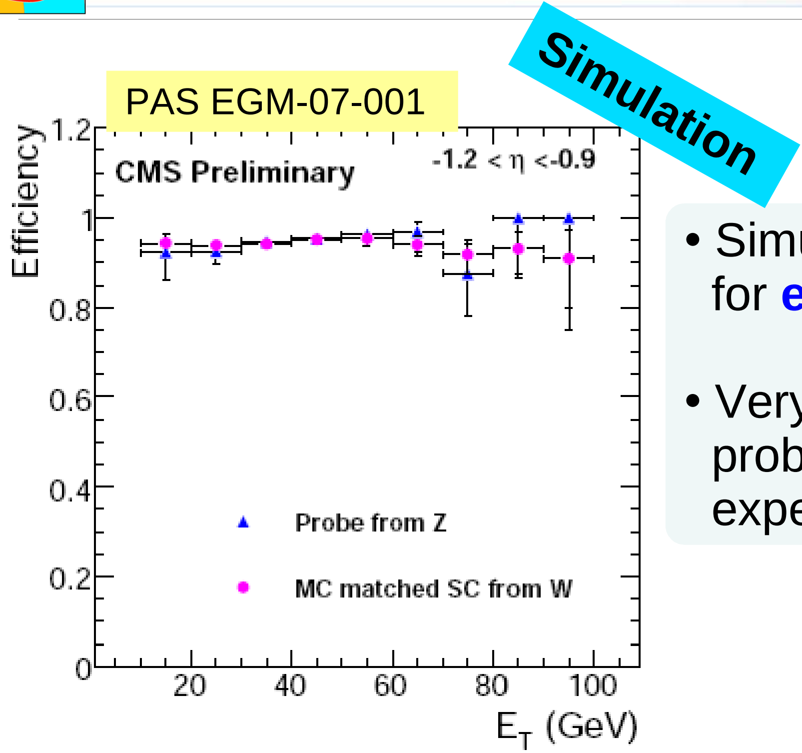




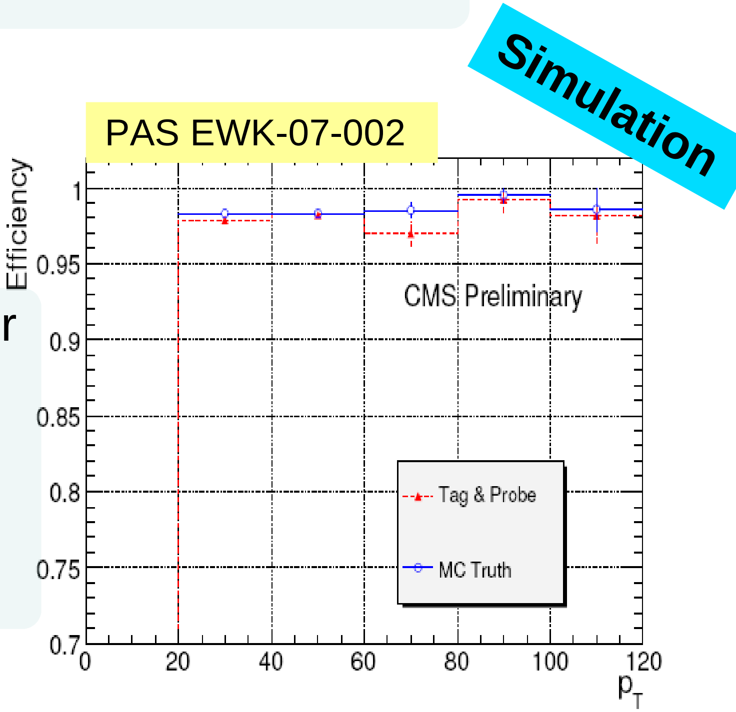
Tag:
Passes stringent electron identification criteria

Probe:
Passes a set of identification criteria depending on the efficiency under study

Tag and probe methods



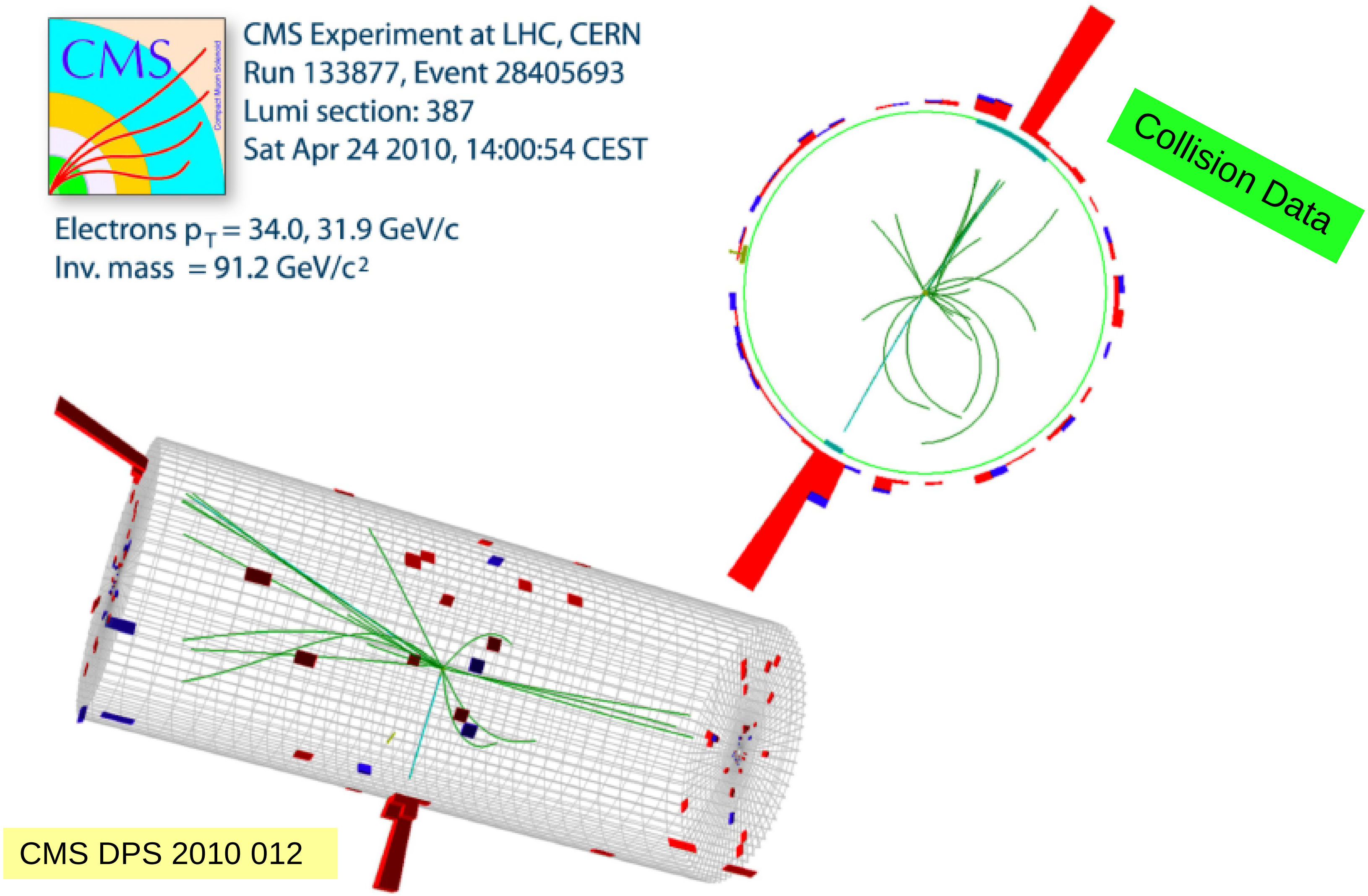
- Simulation of track matching efficiency for **electrons** for early data 10/pb
- Very good agreement between 'tag and probe' method (using MC) and MC expectations



- Simulation of reconstruction efficiency for **muons** for of early data (10/pb)
- Very good agreement between 'tag and probe' method results (on MC) and MC expectations

Electron pair in collision data

Event display of a Z→ e^+e^- candidate in 7 TeV collision data



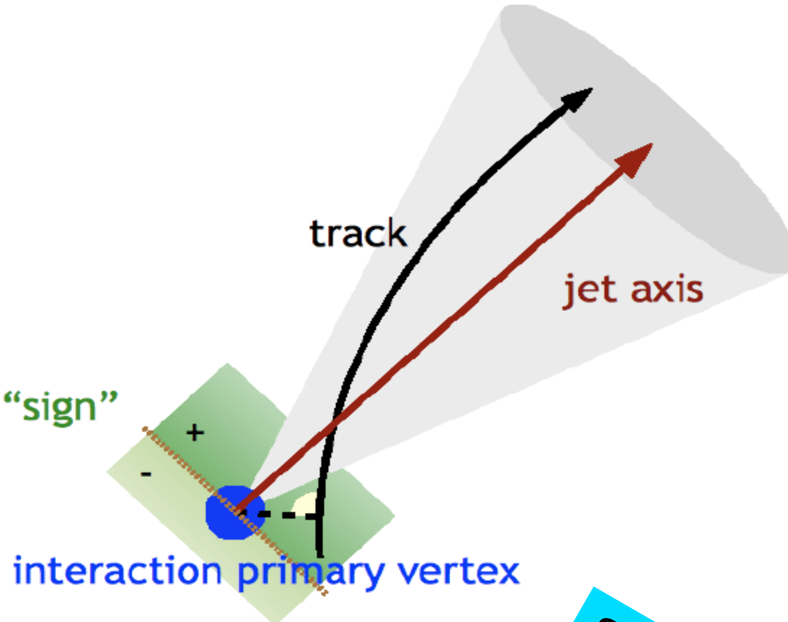
b-tagging section outline

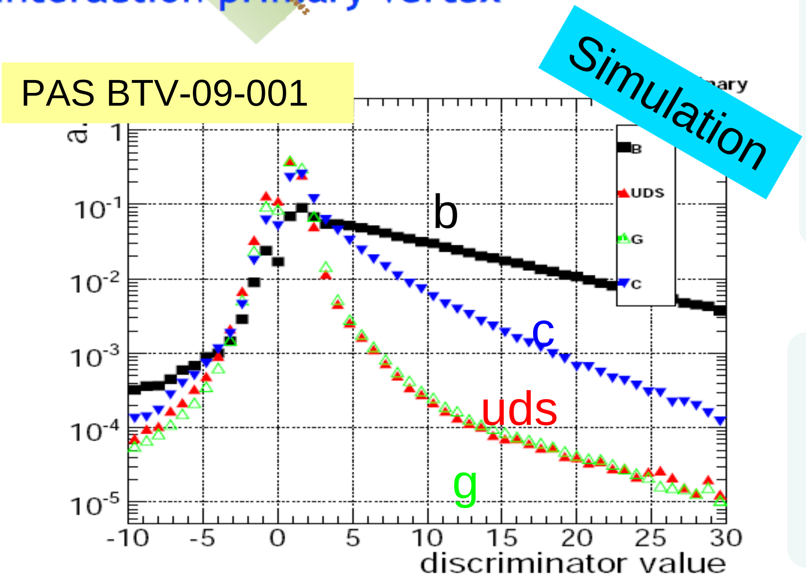
- Lifetime based taggers
  - Impact parameter significance (data vs. MC)
  - 2nd vertex (data vs. MC)
- Soft lepton based taggers
- Performance of b-tag algorithms
- Effect of misalignment on performance
- Performance measurement methods

https://twiki.cern.ch/twiki/bin/view/CMS/PublicPhysicsResults

Lifetime based tagger

- Hadronization process of a b quark contains long-lived B-hadrons (1.5 ps)
- They travel ~O(mm) before decaying
- Gives rise to **displaced vertex and tracks**
- Tracks are associated to jets if they are within a cone around the jet direction
- Tracks are selected with some basic quality cuts p_T, #hits, χ^2 of the fit...

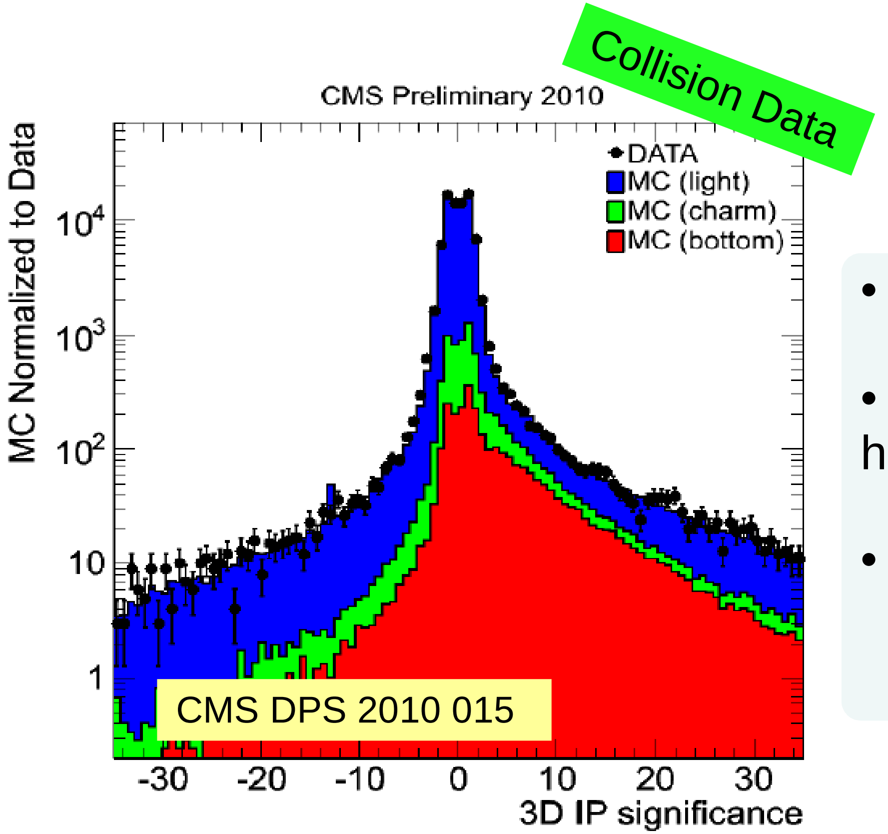




- Closest distance of a track w.r.t. the primary vertex is the **Impact Parameter**
- Impact Parameter (IP) is signed
- Calculated in 2D (xy) and 3D
- Impact Parameter significance (=IP/σ_{IP})

- Track counting algorithm:
  - IP significance of 2nd (shown) or 3rd track
- Jet probability algorithm:
  - Combines IP of all associated tracks

Impact parameter significance



- 7 TeV collision data (0.919 nb^{-1})
- Changes made in the reconstruction chain to have a sufficiently large set of data
- Jet:
  - Anti-kt algorithm with a cone of 0.5
  - p_T>40 GeV & |η|<1.5

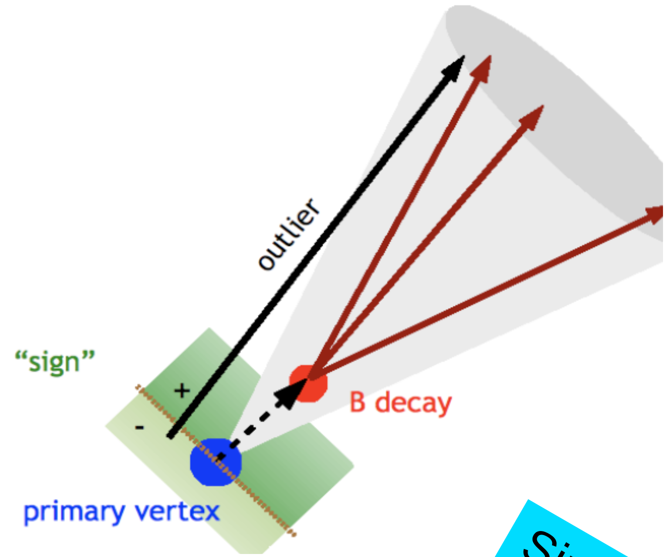
Remarkable good agreement between track Impact Parameter significance from collision data and MC expectations

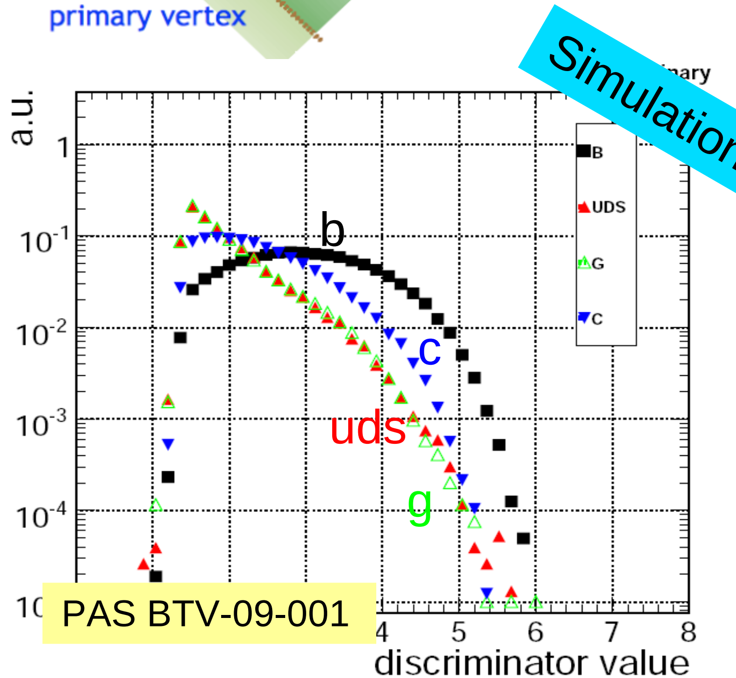
Secondary vertex based tagger

- Secondary vertex reconstructed with Adaptive Vertex Fitter
- Selection criteria to select secondary vertex
  - Mass constraints to remove K_s
  - Transverse distance with primary vertex >100 µm and <2.5 cm
  - Invariant mass < 6.5 GeV/c^2

- Variables connected to 2^{nd} vertex:
  - Invariant mass of charged particles
  - Multiplicity of charged particles
  - Flight significance: significance of distance between the primary vertex and the secondary vertex in the transverse plane
  - Energy of charged particles associated to the secondary vertex divided by the energy of all charged particles associated to the jet
  - impact parameter significance of tracks

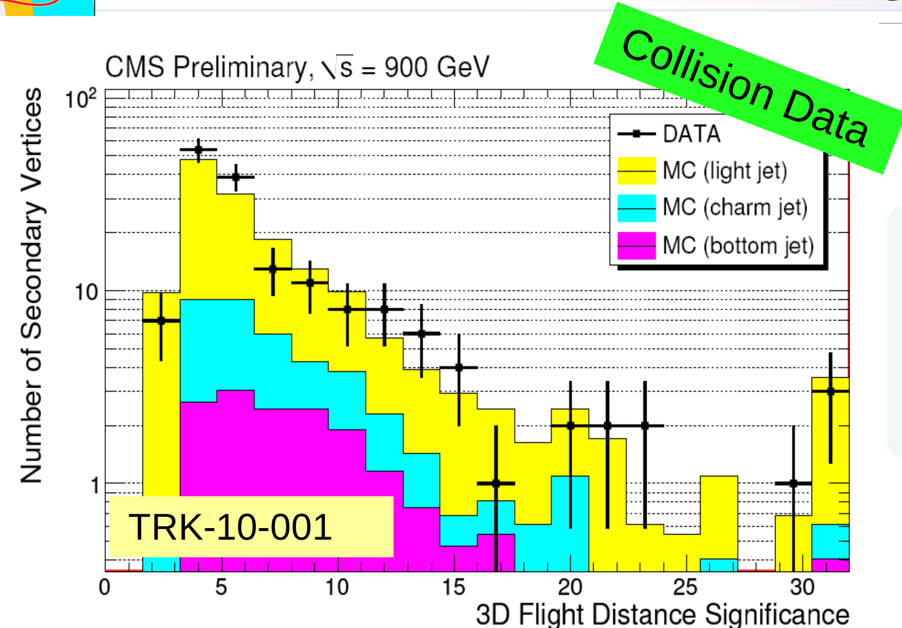
- **Simple 2^{nd} vtx algorithm: flight significance**
- Combined 2^{nd} vtx algorithm: Likelihood ratio combination of variables

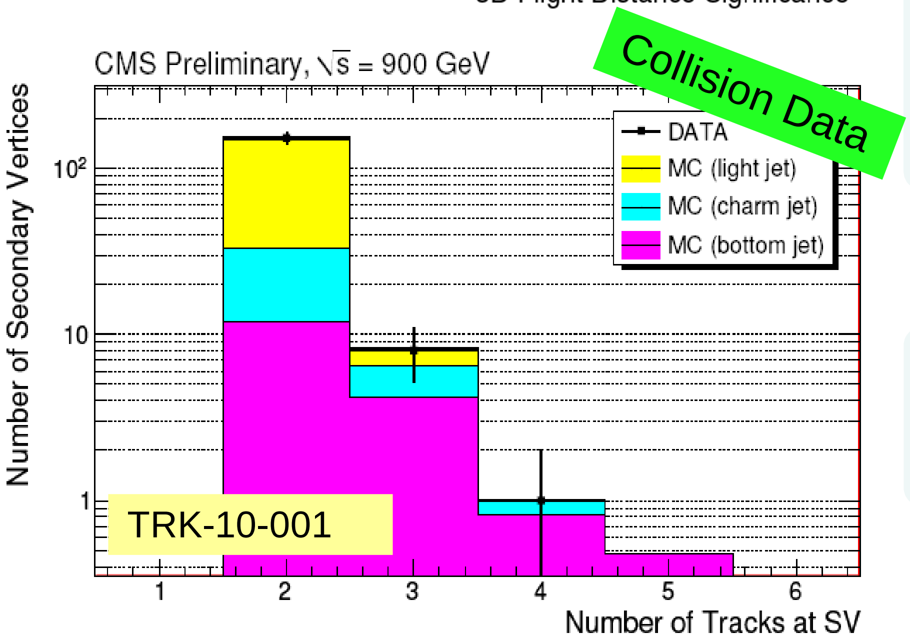




PAS BTV-09-001

Secondary vertex



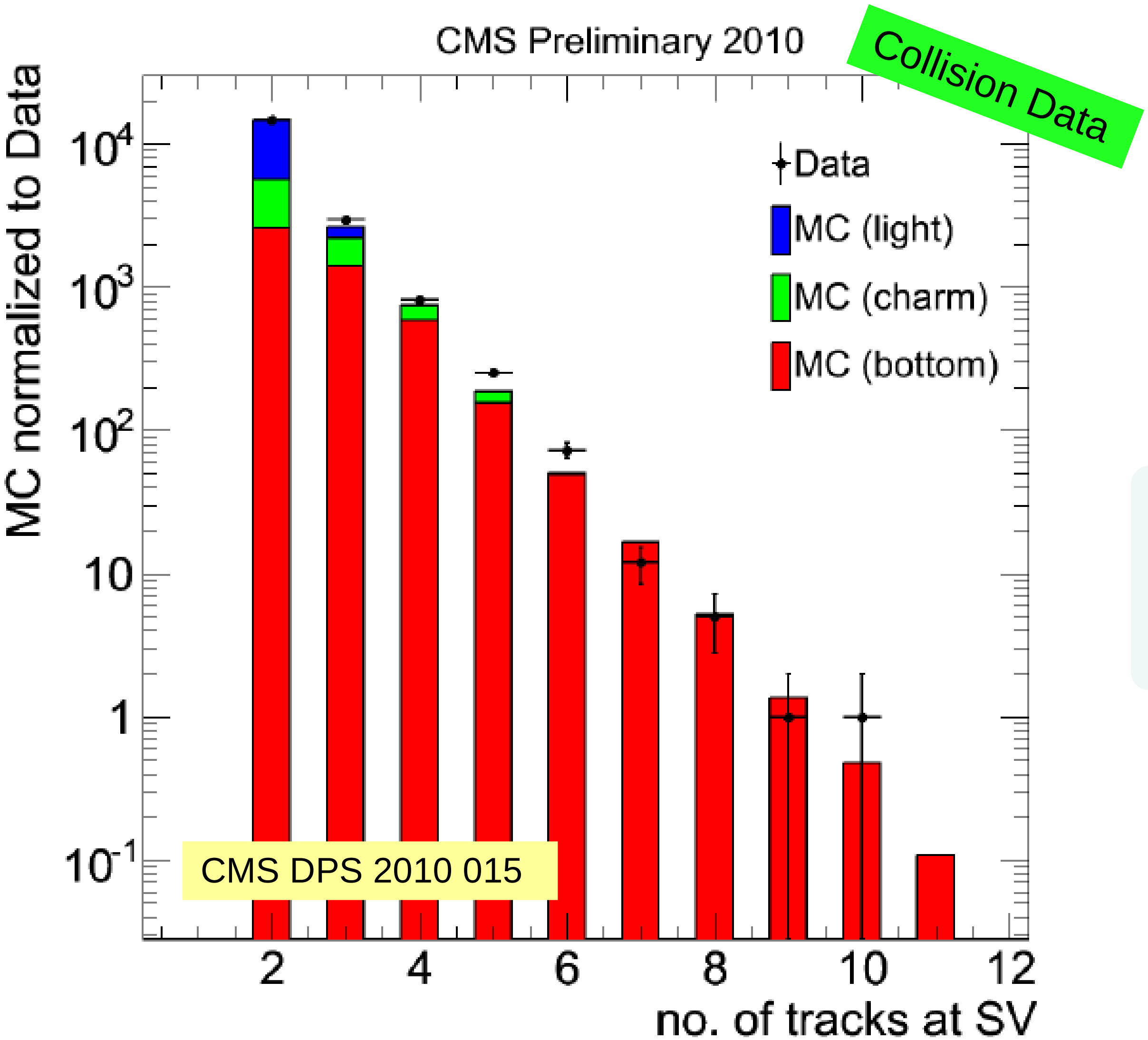


- 2^{nd} vertex related variables are reconstructed in 900 GeV data
- Relaxed requirements to obtain enough events with 2^{nd} vertex

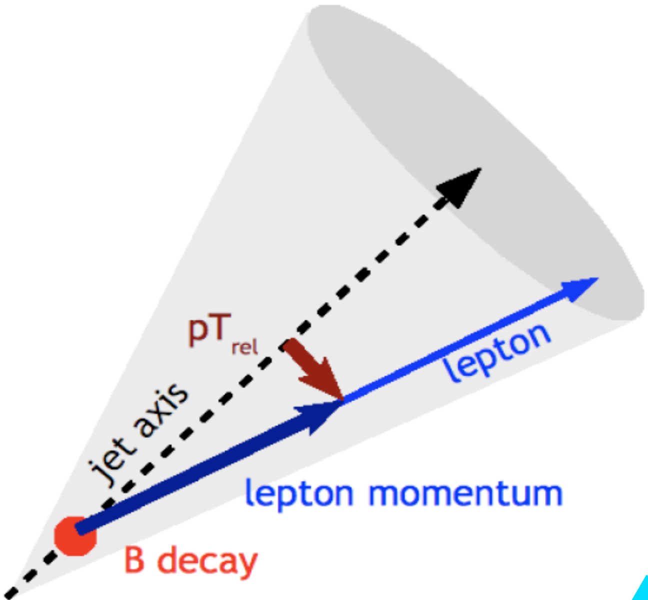
- 3D significance of flight distance (upper plot)
- Number of tracks associated to the 2^{nd} vertex (lower plot)

Very good agreement between distributions from min. bias MC and collision data

Secondary vertex

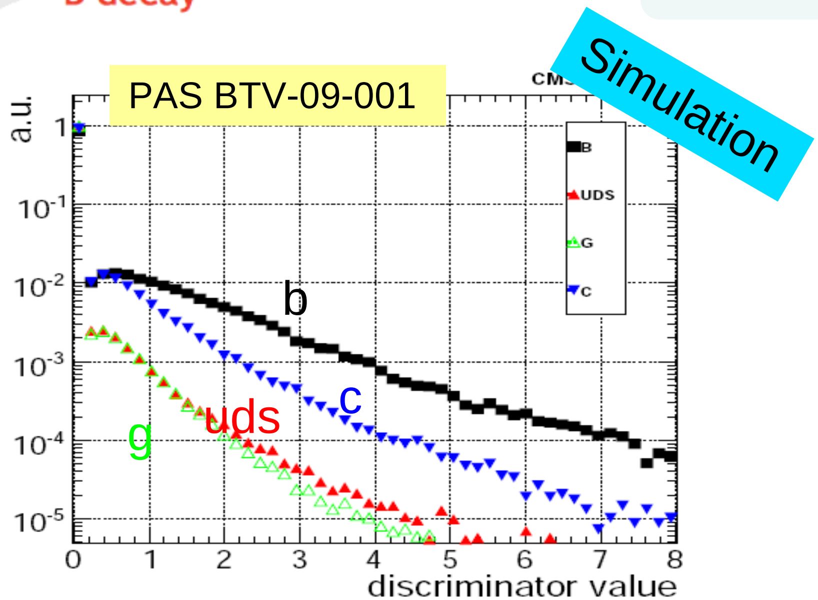


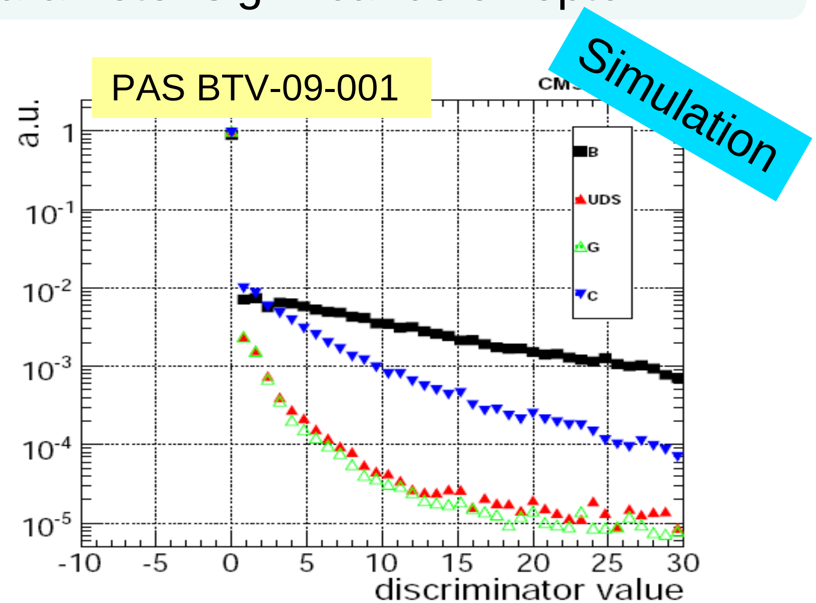
- Number of tracks associated to the 2nd vertex
- 7 TeV data



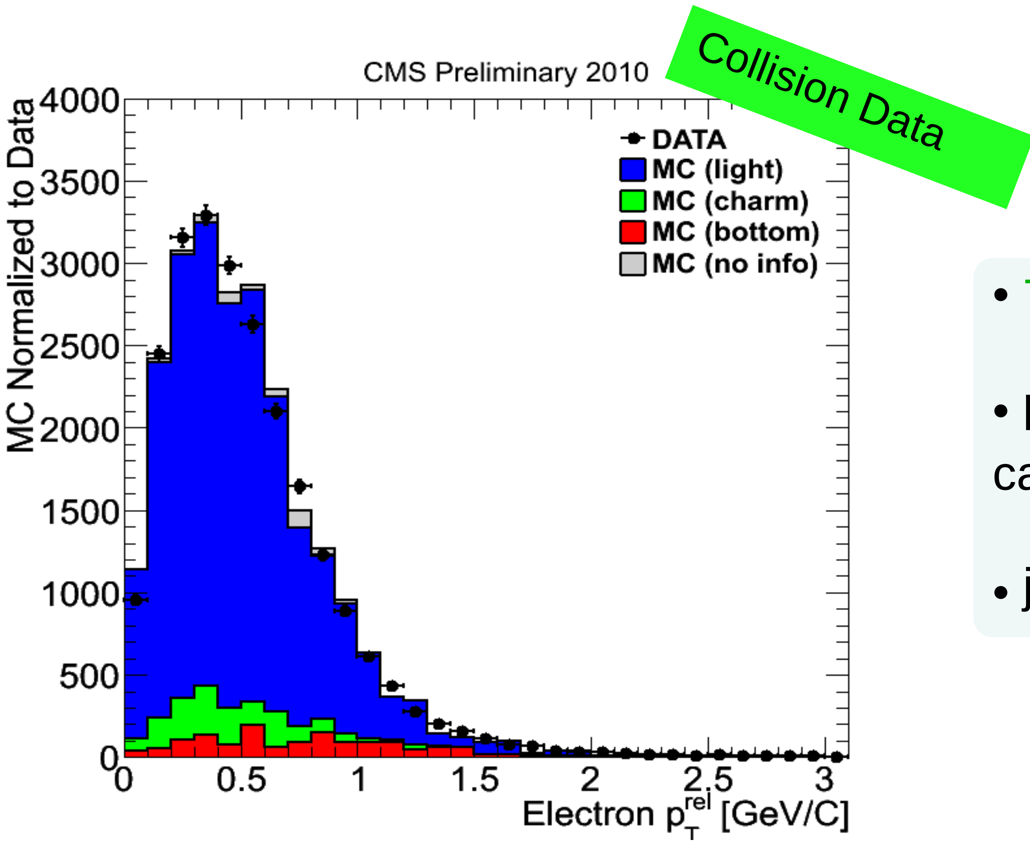
- 20% of all b jets have a **muon or electron in the jet** because of the decay of the B hadron
- Lepton information used in soft-lepton taggers

- Soft *muon* by p_T^{rel} algorithm (left)
  - Relative transverse momentum p_T^{rel} of muon w.r.t. axis of the jet
- Soft *muon* by IP significance algorithm (right)
  - impact parameter significance of lepton





Soft lepton in b jets

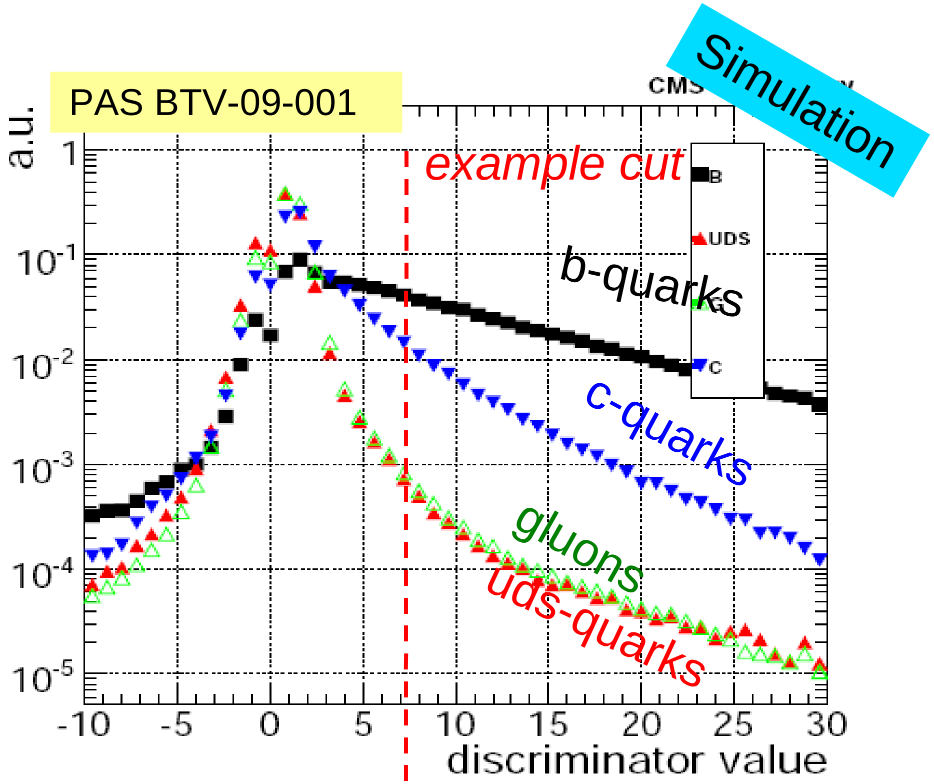


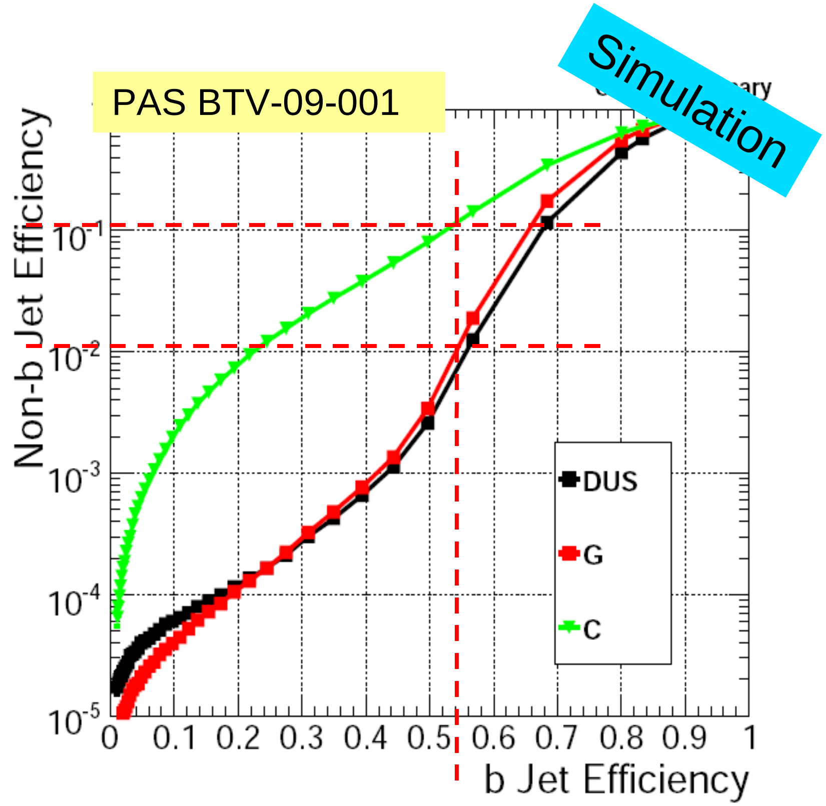
CMS DPS 2010 015

- 7 TeV collision data (0.919 nb^{-1})
- p_T relative to the jet axis for electron candidates
- jets: p_T > 10 GeV and $|\eta|$ < 2.4

Performance

The efficiency and mistag rate are evaluated at each discriminant cut

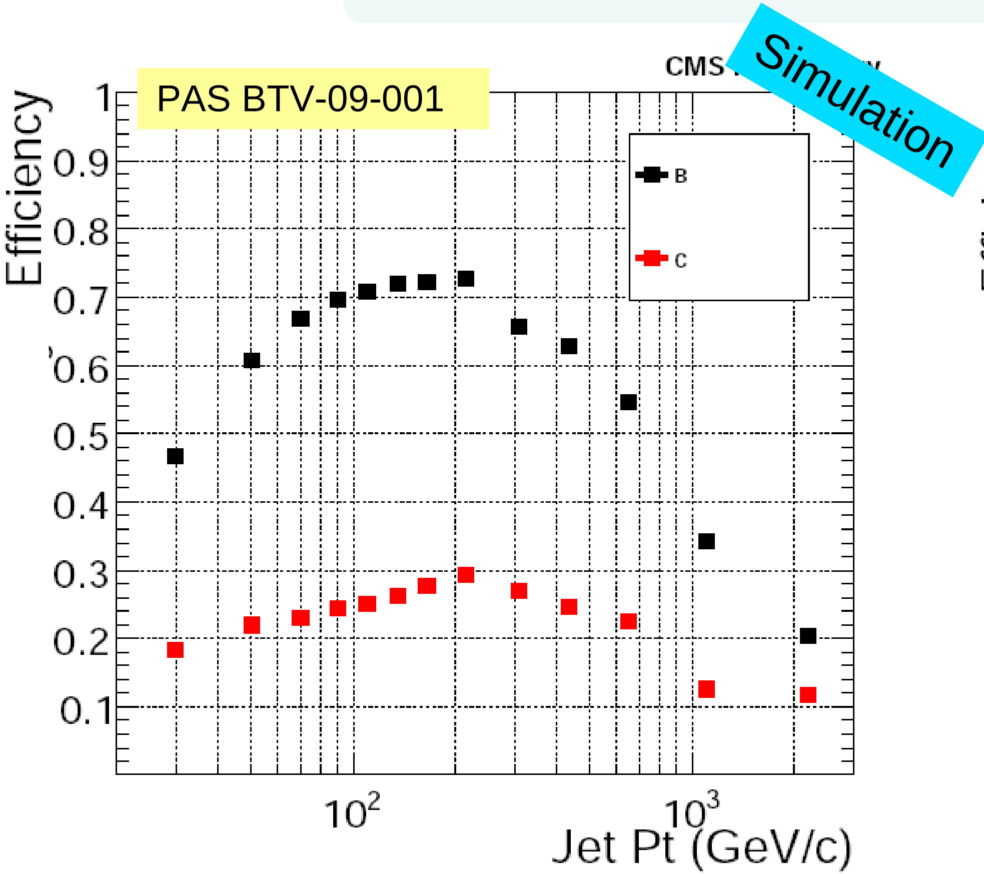


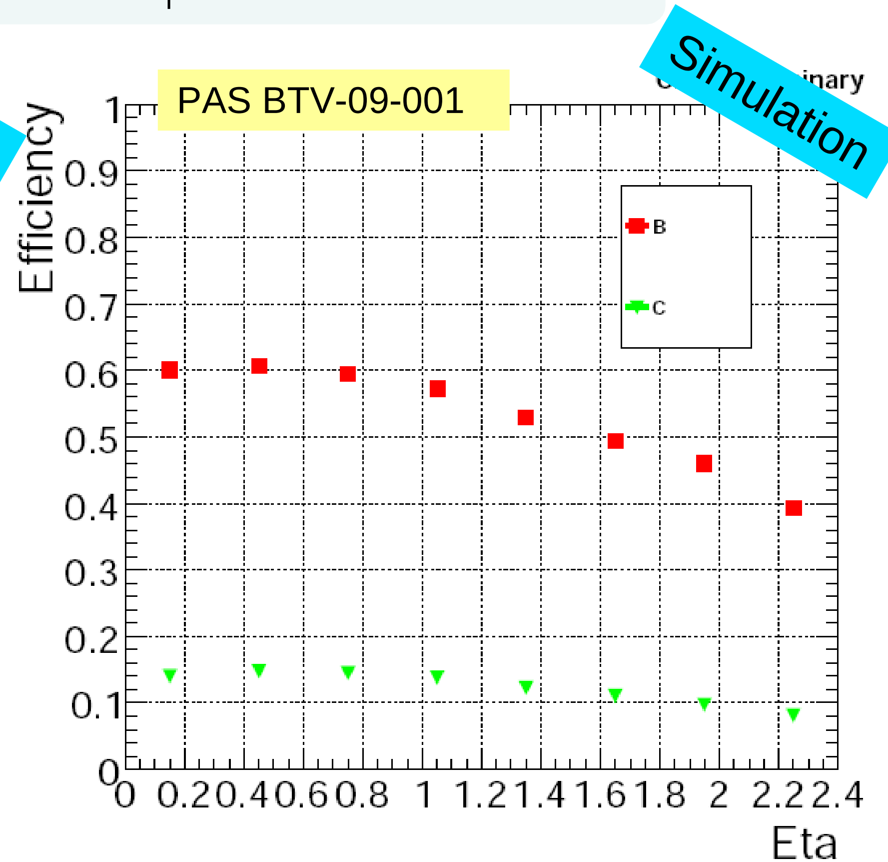


Typical working points are defined by the uds-quark jet mistag rate:
for trackCounting tagger: $\varepsilon_{\text{uds-mistag}}$ = 1%, ε_b = 55%, $\varepsilon_{\text{c-mistag}}$ = 10%

Performance

The performance is dependent on p_T and η of the jets

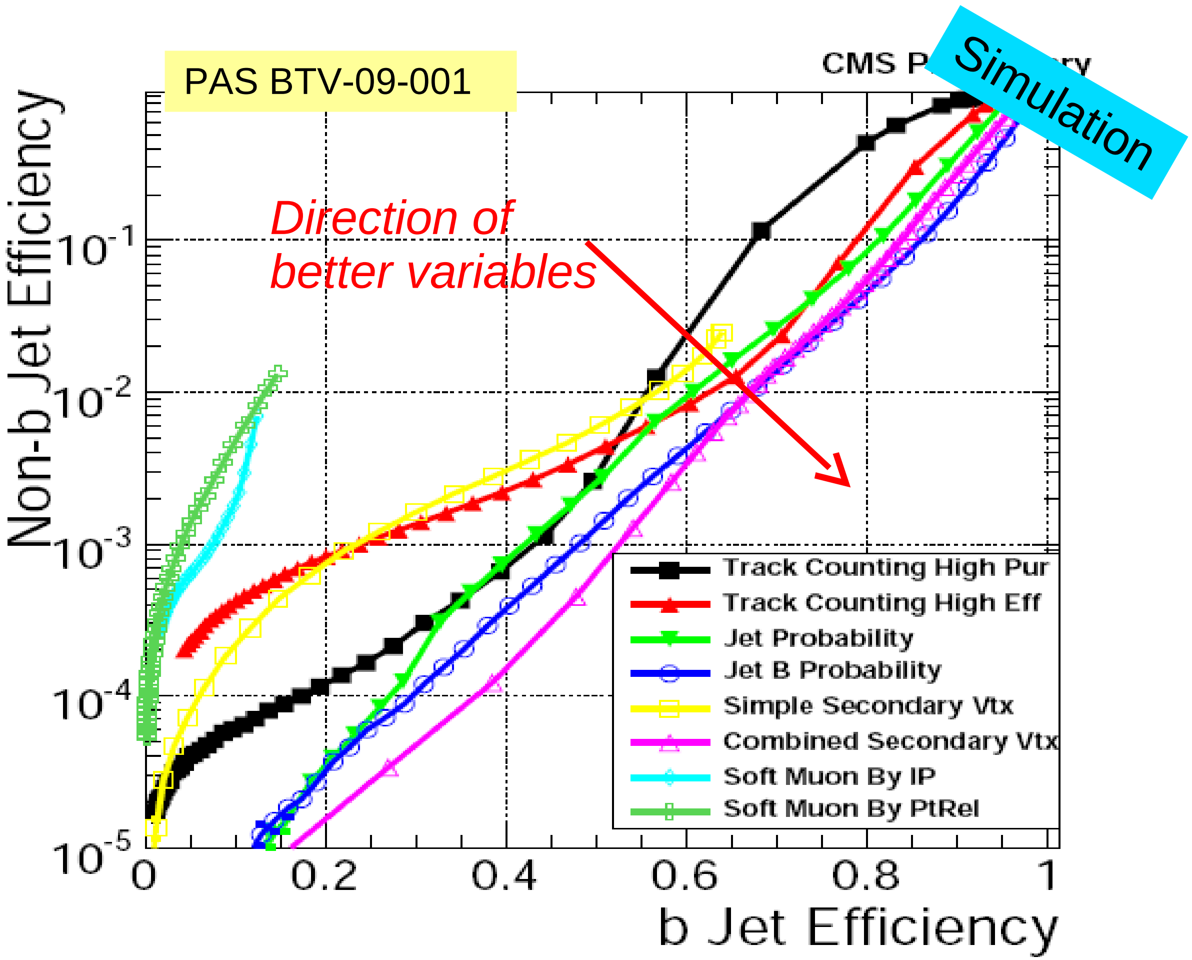




b-tag efficiency and c-mistag rate for a fixed uds-mistag rate of 5% for the track counting b-tagger (2nd track IP)

b-tag efficiency and c-mistag rate for a fixed uds-mistag rate of 1% for the track counting b-tagger (2nd track IP)

Comparison of performance of all b-taggers in CMS



The best variable is the combined 2nd vertex, but it is the most complex one

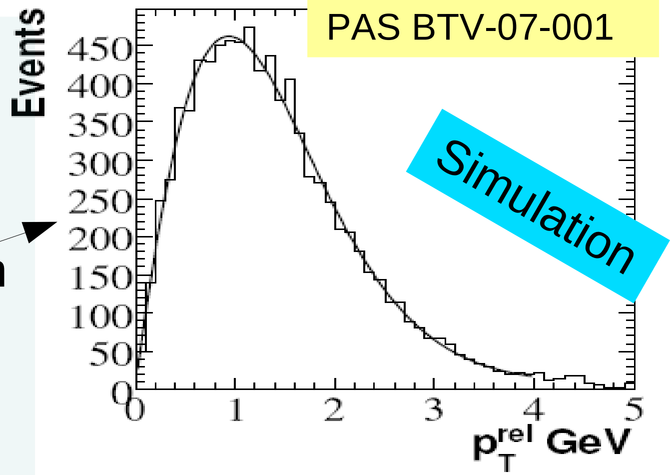
Soft-lepton taggers can only go up to ~20% eff. This is due to the lepton branching ratio in b-jets

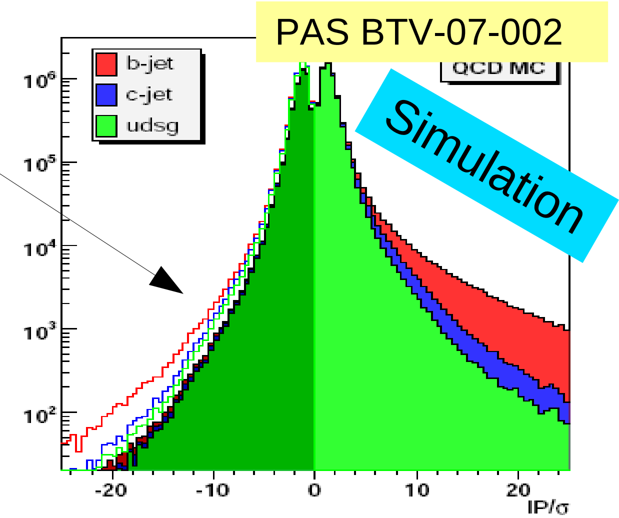
simple secondary vertex and **track counting (2nd track)** are the most robust in the start-up phase (ref. back-up)

Measurement of performance in data

- System 8 method: solve a system of eight equations constructed from the total number of events in two samples with different b jet content, before and after tagging with two b-tagging algorithms.
- p_T^{rel} method: relies directly on a fit to the p_T^{rel} distribution of the muon before and after tagging the muon-jet
- Counting method: based as well on p_T^{rel}

- Measurement of mistag rate from negative tags





- Using the constraint $|v_{tb}|=1$ in top quark pair events can be used as well to measure b-tag efficiency

CMS NOTE 2006/013

- Different methods based on different jet topologies needed since b-tag performance depends on p_T, η, number of tracks
- All methods expect roughly a **relative uncertainty of 15% for 10/pb and down to 5% for 1000/pb**
- Top quark based measurements only feasible > 100/pb

ME_T section outline

- ME_T reconstruction algorithms
  - ME_T performance on data
  - ME_T resolution from data

https://twiki.cern.ch/twiki/bin/view/CMS/PublicPhysicsResults

MET

CMS DPS 2010 014

Collision Data

MET is reconstructed from the transverse vector sum of energy deposits in the calotowers

$$\vec{\not{E}}_T = -\sum_{i=1}^{\text{towers}} \vec{E}_T^{\,i}$$

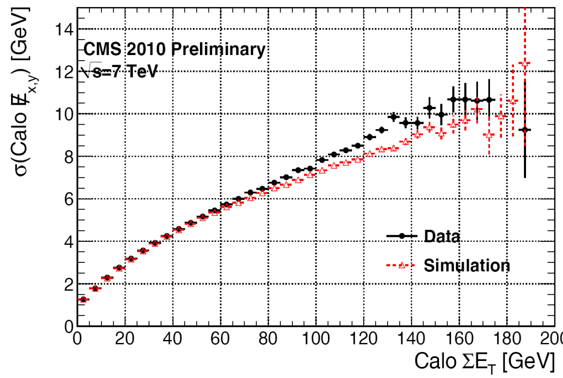
Correction for the muons by subtracting the p_T of the muon and adding the calorimeter energy deposit and additional corrections are made for other tracks in the events

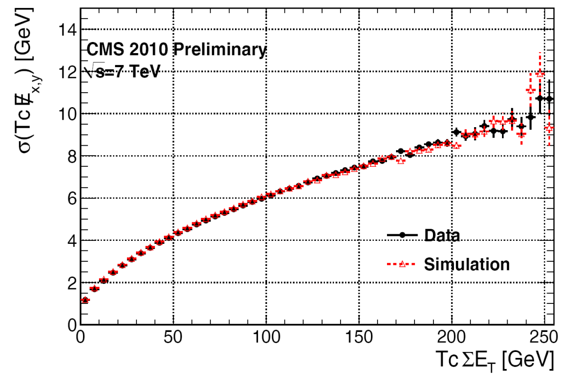
JME-09-010

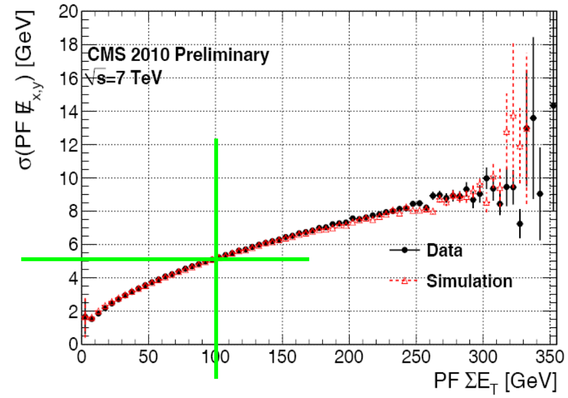
$$\not{E}_T^{\text{tc}} = -\sum_{\text{towers}} \vec{E}_T - \sum_{\substack{\text{good}\\\text{muons}}} \vec{p}_T + \sum_{\substack{\text{good}\\\text{muons}}} \vec{E}_T^{\text{MIP}} + \sum_{\substack{\text{good}\\\text{tracks}}} \langle \vec{E}_T \rangle - \sum_{\substack{\text{good}\\\text{tracks}}} \vec{p}_T$$

Particle flow missing energy is reconstructed by performing the transverse momentum-sum of all particles in the event reconstructed from all subdetectors combined

PFT-09-001





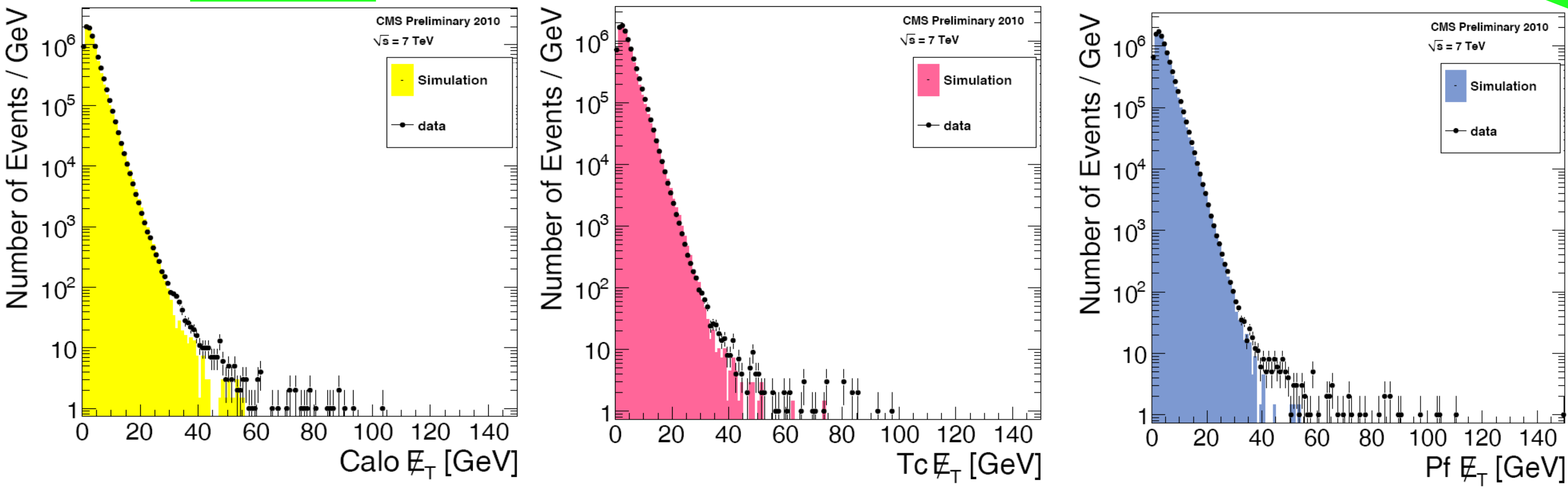


MET

CMS DPS 2010 014

Collision Data

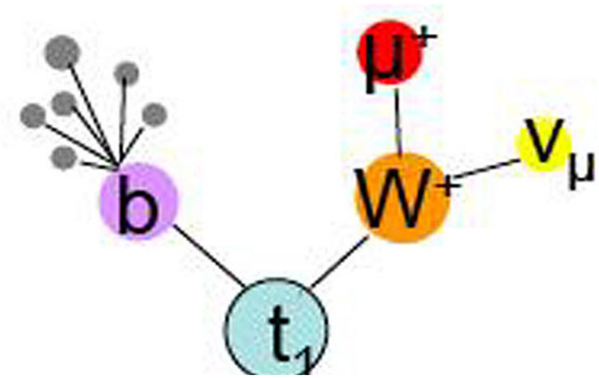
7 TeV



- Core of MET is well described
- Noise cleaning is applied, will be further improved in the future

MET is in fairly well agreement with expectations from MC simulations

Conclusions



- Leptons: eg. Muon 50 GeV/c
  - Reconstruction efficiency > 97%
  - Resolution <2%
  - Charge mis-asignment prob. ~0.1%
- Data confirms in general the MC expectations for both muons and electrons

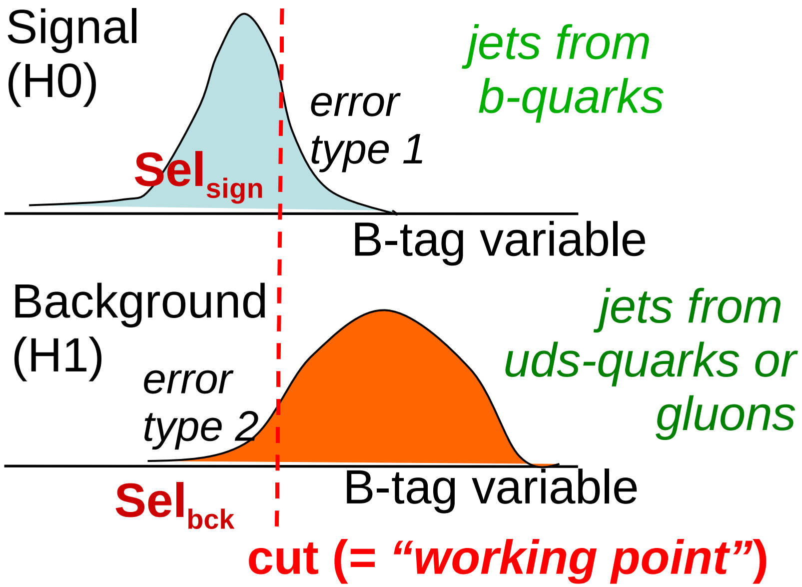
- MET
  - MET value and resolution agree fairly well with simulations

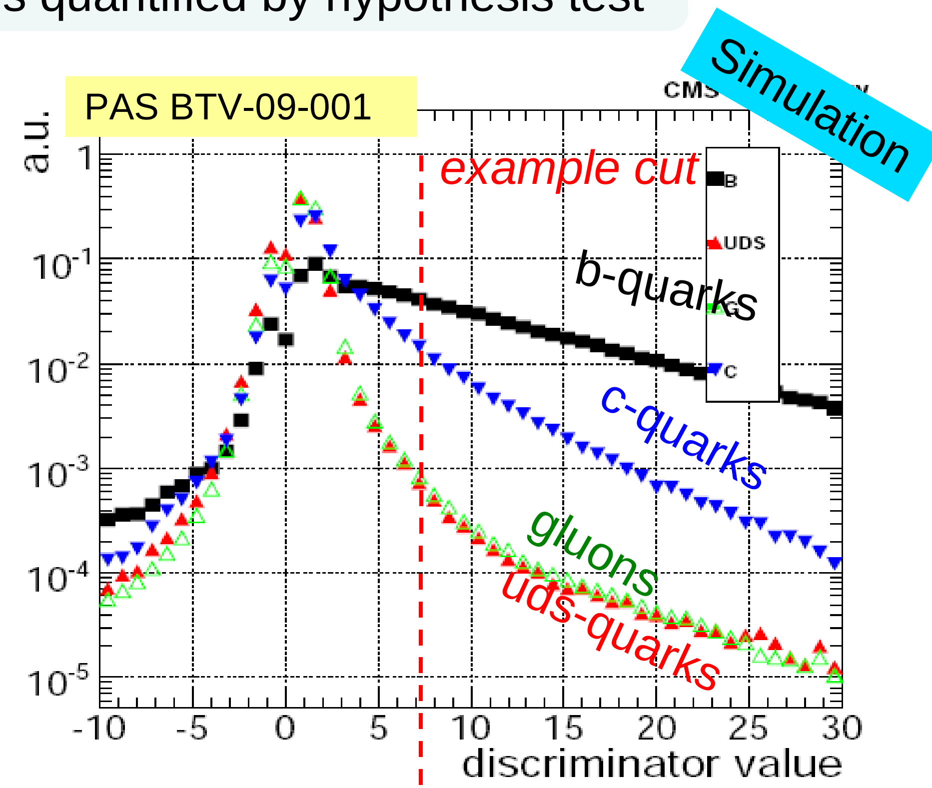
- b-tagging:
  - Key variables look very promising
  - Both soft lepton tagger and lifetime based taggers are in good shape
  - Techniques for data driven determination are ready

Back-up

Performance

The **performance of b-tag algorithms** is quantified by hypothesis test



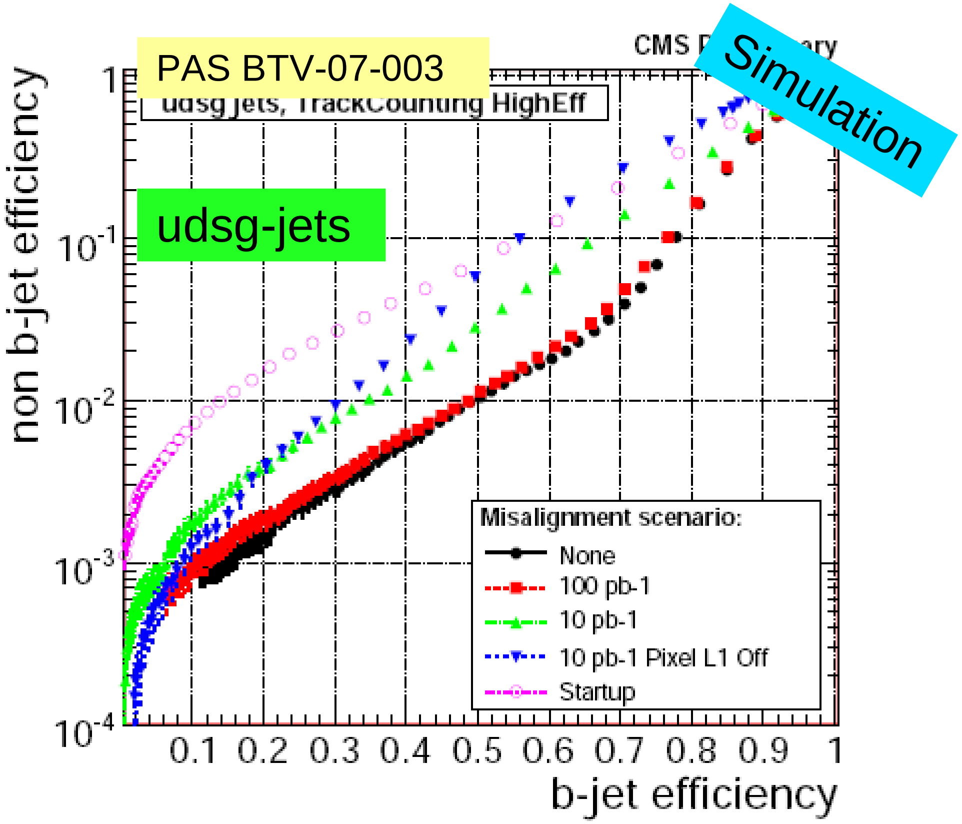


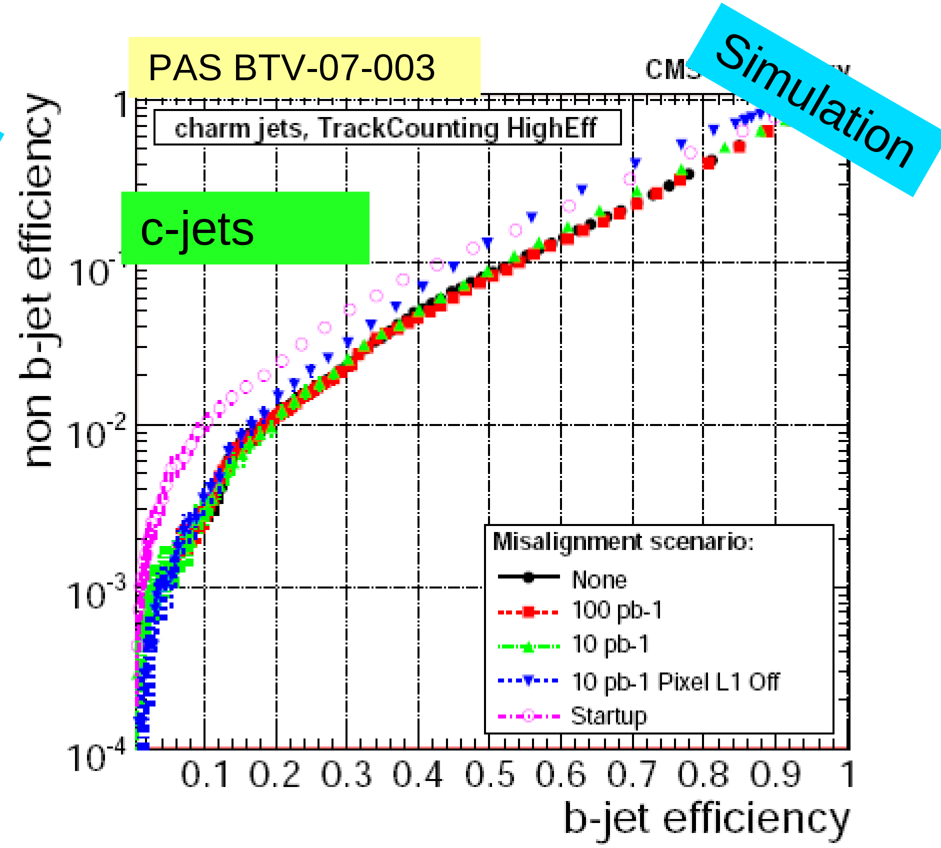
- Error of type 1:
  rejecting a real b-quark jet
- Error of type 2:
  accepting a uds-quark or gluon

- ε_b = b-tag efficiency =
  #Sel$_{sign}$ / #total b-quark jet
- $\varepsilon_{non\text{-}b}$ = mistag rate =
  #Sel$_{bck}$ / #total non-b-quark jets

Misalignment and performance

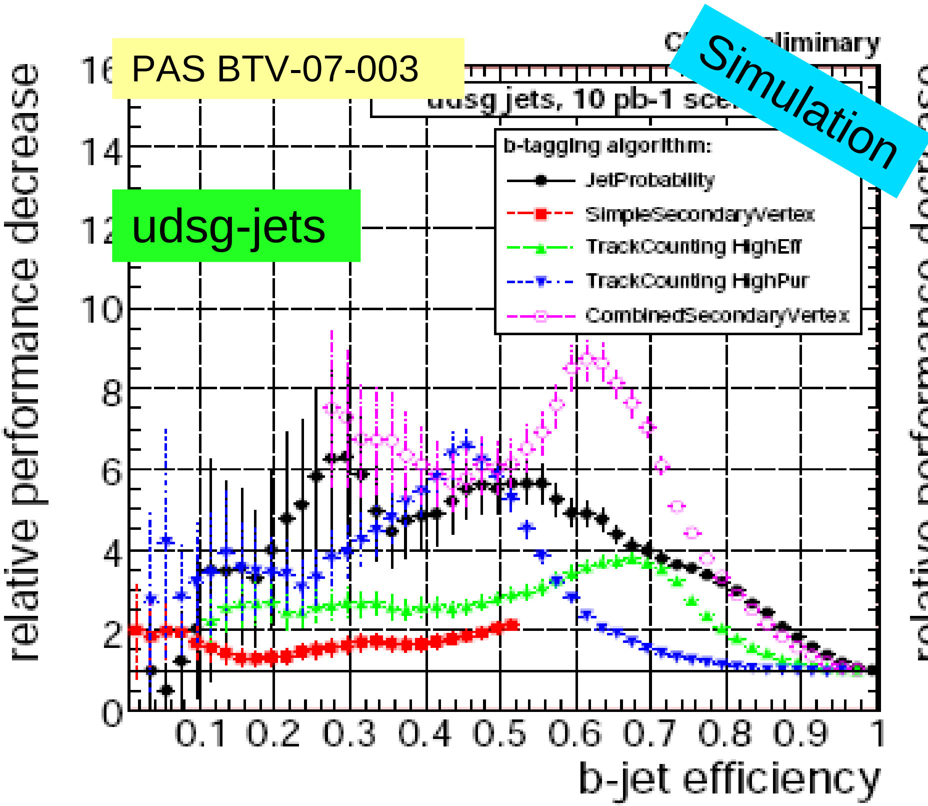
- Robustness of b-tag algorithms was studied for mis-alignment scenario's ('07)
- Plots show performance for various scenarios which have been found to be much too pessimistic, tracker is already very well aligned
- Start-up scenario shows strong degradation of the track-counting tagger

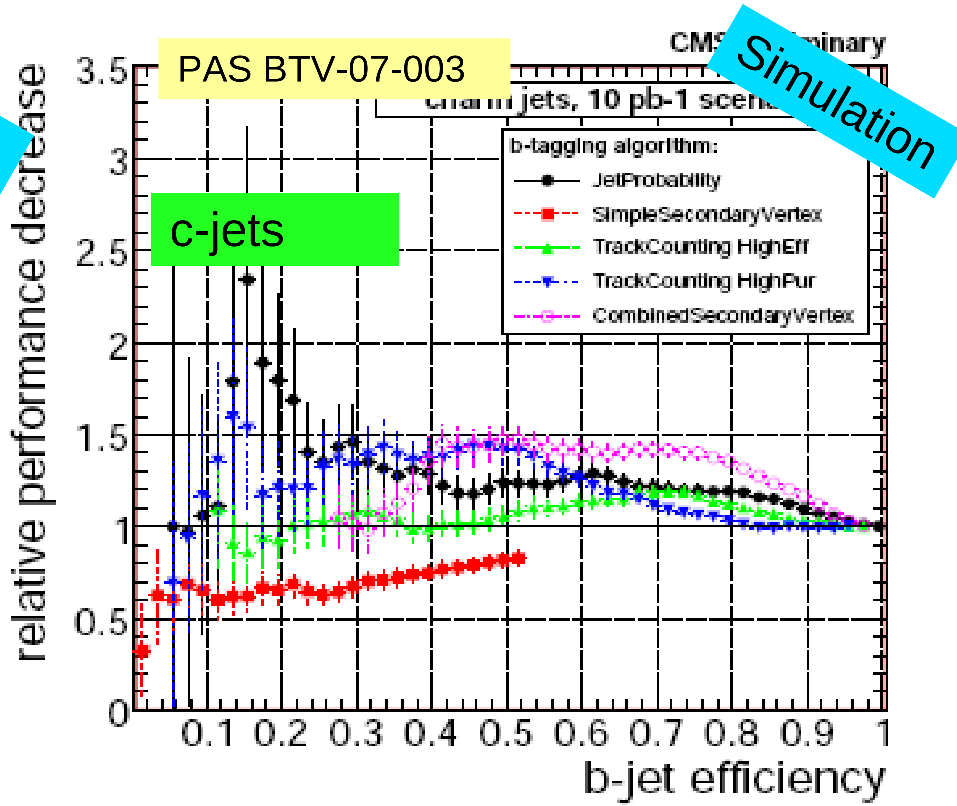




Misalignment and performance

Relative performance decrease is different for various taggers for 10/pb scenario





The performance decrease is much overestimated but **simple secondary vertex** and **track counting (2nd track)** are the most robust in the start-up phase