Lepton and Photon ID at the Tevatron

P.Murat(FNAL) for the CDF and D0 collaborations

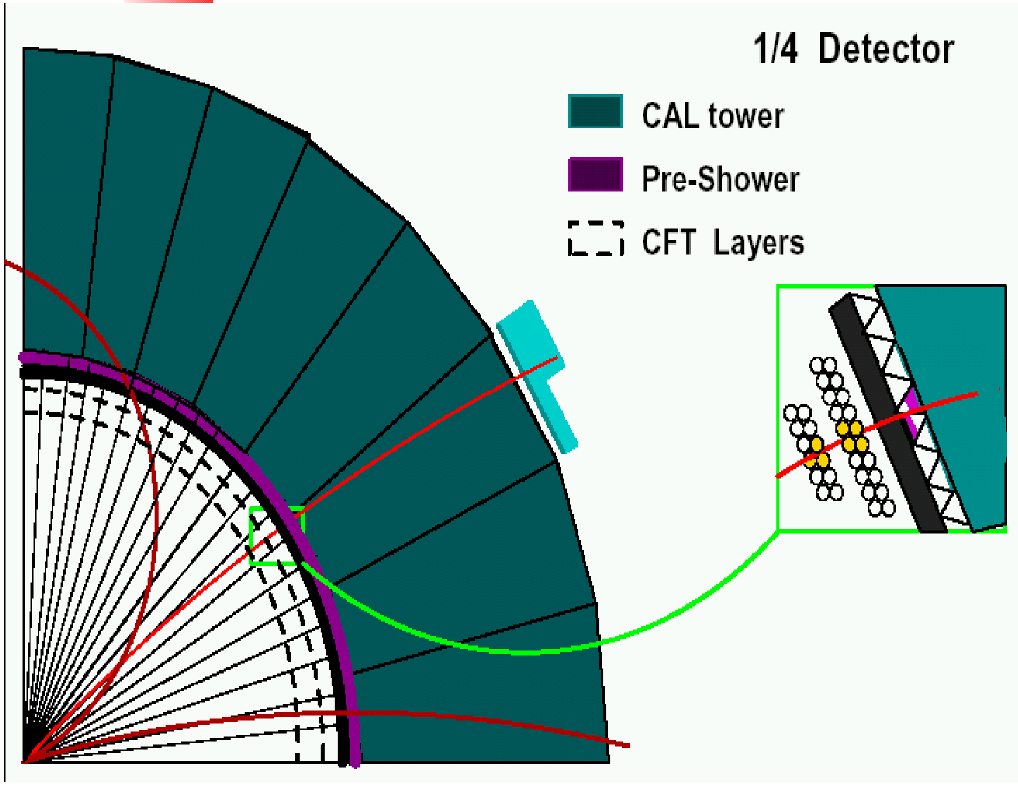
Introduction:

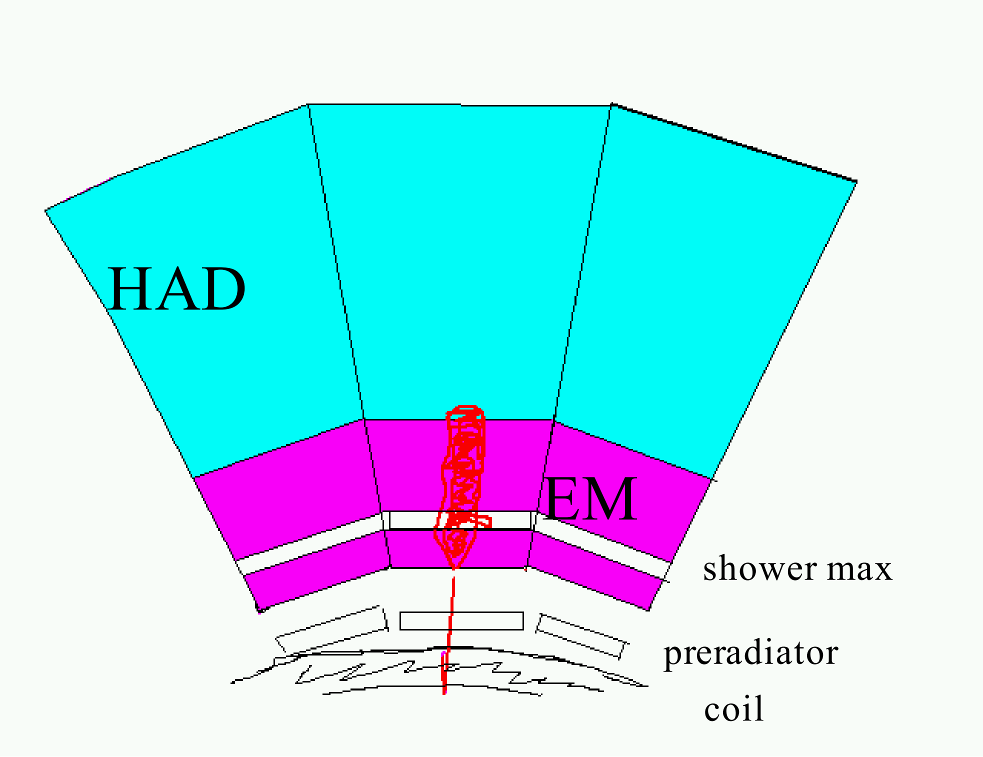
- High Pt leptons and photons are very important objects - can trigger on them
- Particle/object ID requirements driven by the physics
  - Isolated leptons and photons (W/Z, high-Pt searches – Z', SUSY...)
  - Non-isolated e/μ - tagging of the heavy quark jets
- Quantifying performance of ID techniques: **efficiencies**, **probabilities of misidentification**
- approaches: cut-based ("box"), likelihoods , neural networks

- Use - subject of a different talk

many thanks to D.Denisov, Y.Gershtein(D0), D.Waters(CDF)

Central calorimeters : face to face

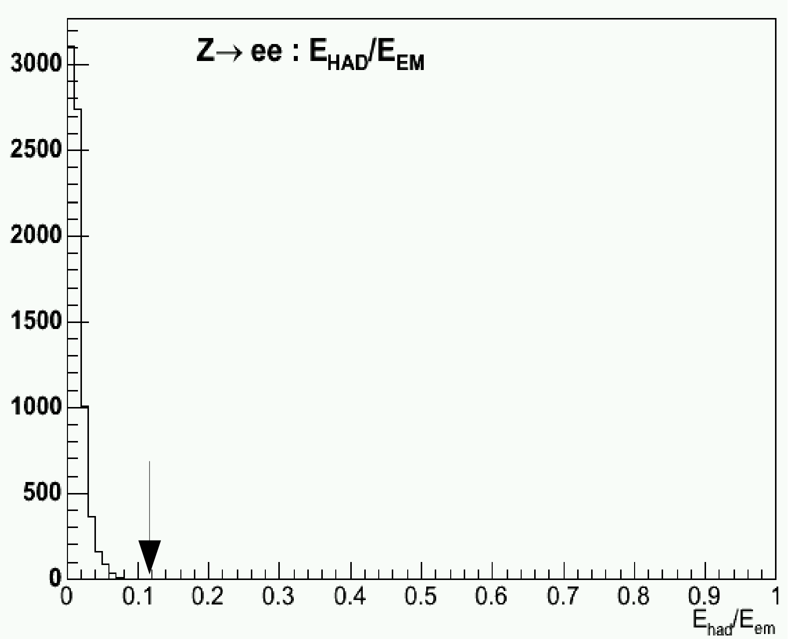


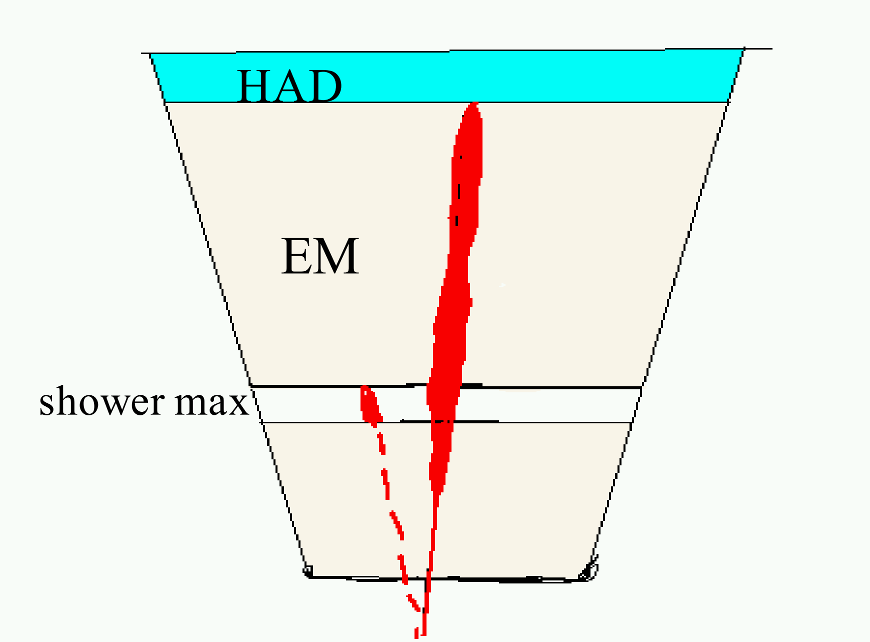


	CDF	D0
Technology	Sandwich (lead-sc/steel-sc)	LAr / Ur
Eta-phi segmentation	0.1 x 0.25	0.1 x 01
Long. Segmentation	2 (EM / HAD)	9/8 layers (first 4 – EM)
Preshower	MWPC => scint pads	Sc strips
Shower max	MWPC (pitch 1.5-2cm)	Layer 3 (0.05 x 0.05)
Total material	~5-7 interaction lengths	~7-9 interaction lengths

Electrons

- Electron clustering "box" ~ 0.3 x 0.3
- EM fraction > 0.9 (D0) , tighten later
- Shower shape consistent with that of EM shower
- track – shower max/preshower match ($\sigma \sim 2$-3mm)
- Consistency of the energy and momentum measurement: E/P < 2 (CDF)
- Isolated (calorimeter, sometimes – tracker), typical isolation cone size ~0.4
- Conversion removal
- both CDF and D0 reconstruct subclusters within the electron clustering cone
- Correlation between the ID variables

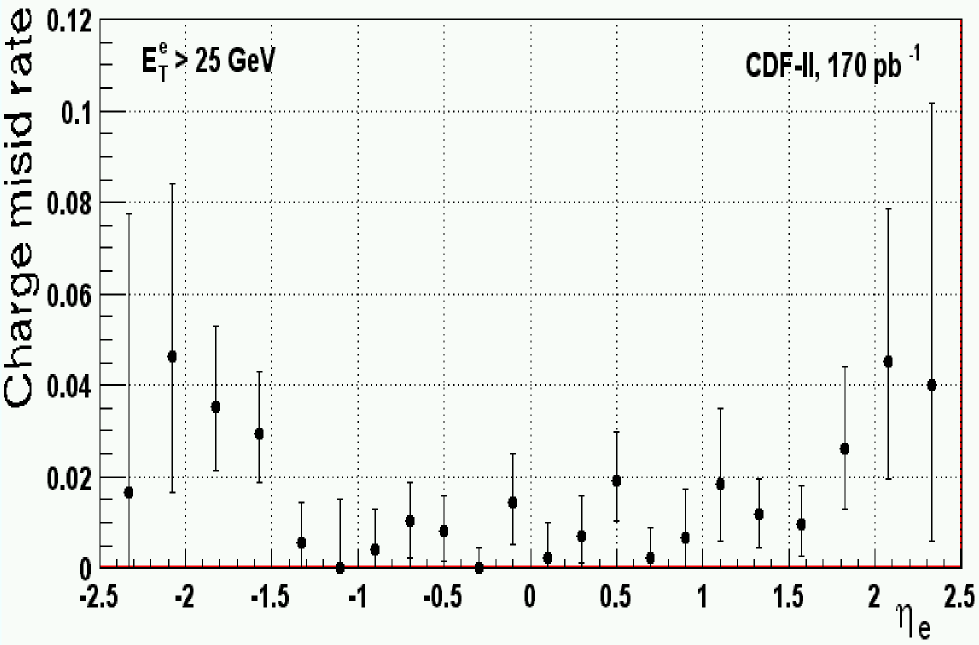




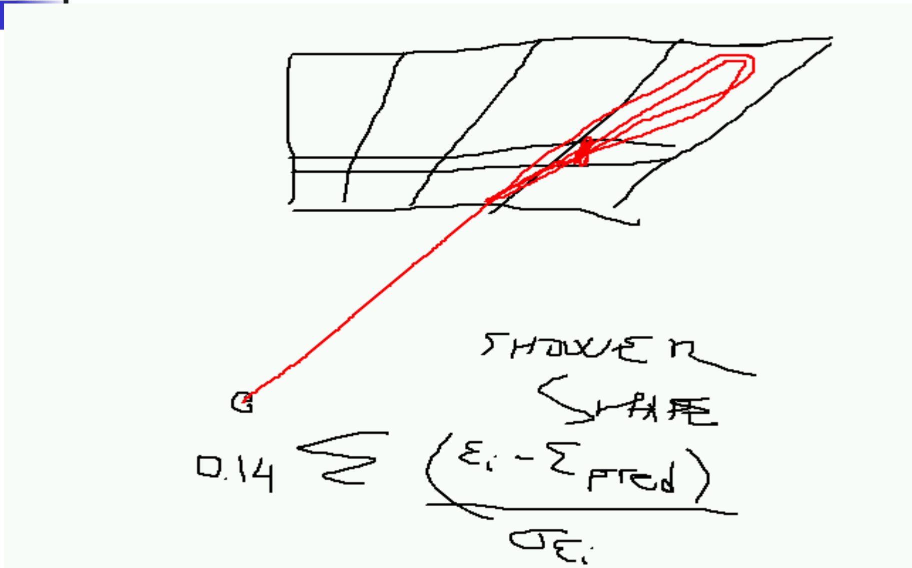
Electron ID: efficiencies

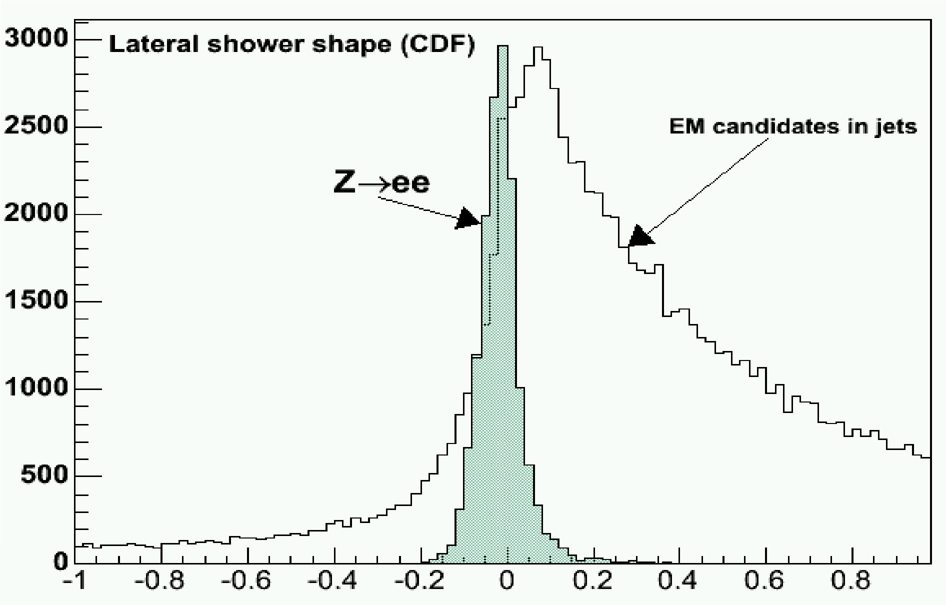
- Measure ID efficiencies for high-Pt electrons using Z->ee decays
- Several quality classes (tight, loose)
- Efficiency: 85-95%
- data-to-MC scale factors : <5%, uncertainties <1%
- Backgrounds for ID efficiency measurement small (same-sign events under the Z peak) , which makes the efficiency measurements very robust
- SUSY (multileptons): isolated electrons above ~5 GeV
  - Calibrations sample: low mass Drell-Yan e+e- events
  - Background : same-sign e+e-candidates
  - Efficiencies/scale factors - similar to above
- Charge misID in the forward region: ~4% at |η| ~ 2

Charge misID for Z->ee

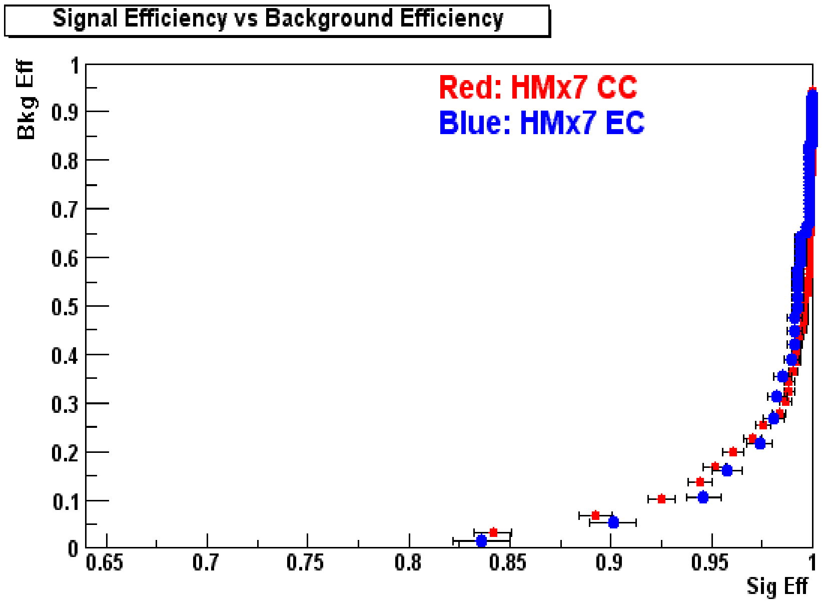


exploiting shape of EM showers



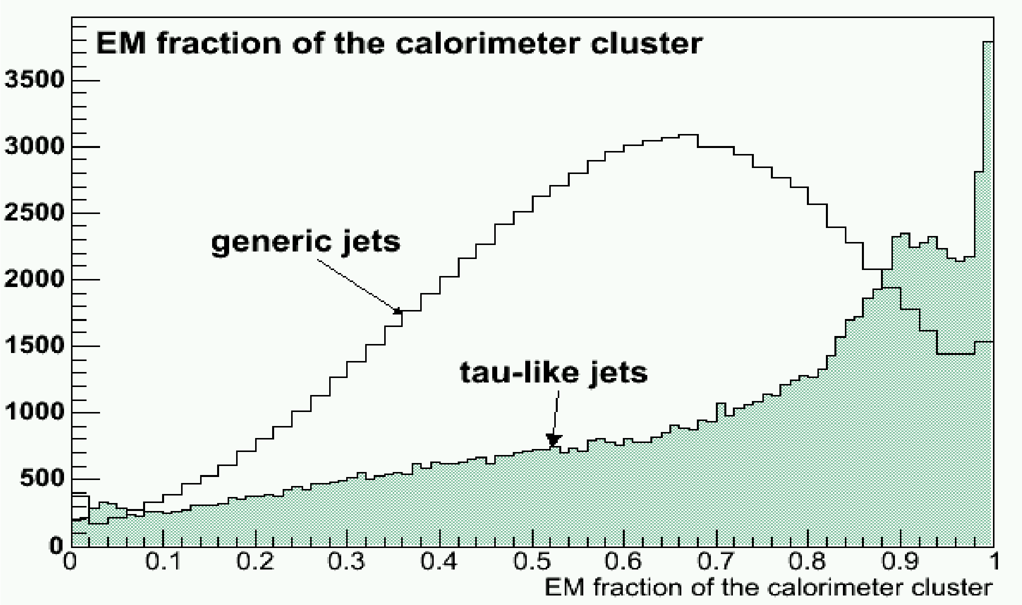


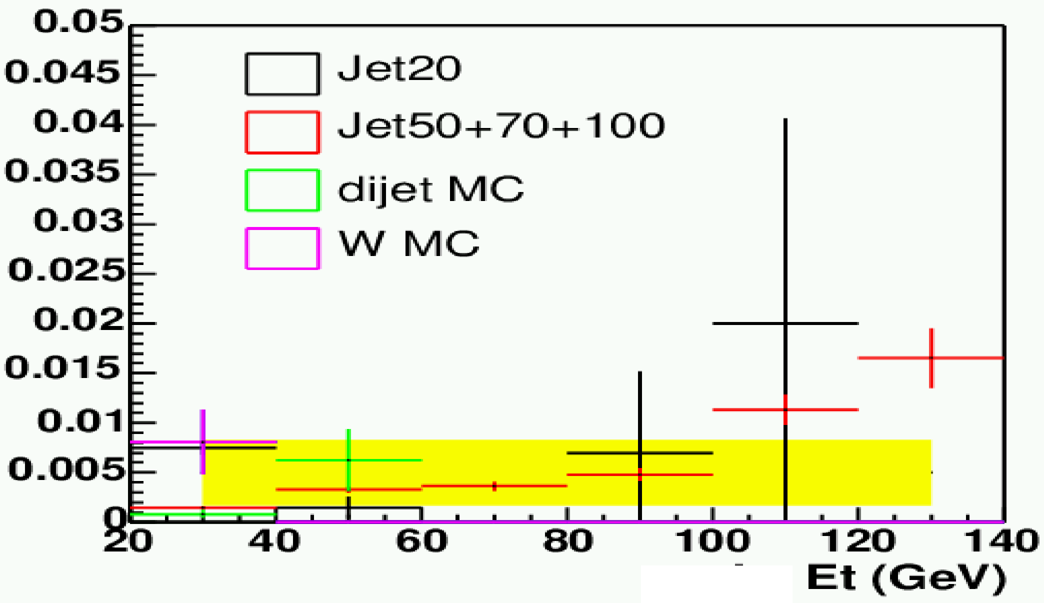
- Both experiments use cluster shape variables in electron ID
  - CDF : calculate lateral shower shape analytically
- EM cluster and a track
  - Matching: track-cluster – shower max
- D0 - "H-matrix" - measurements in 9 layers(5x5 matrix) calculate chi2 of the shower using 7 or 8 variables
  - Account for the shower energy



Electrons: misidentification rates

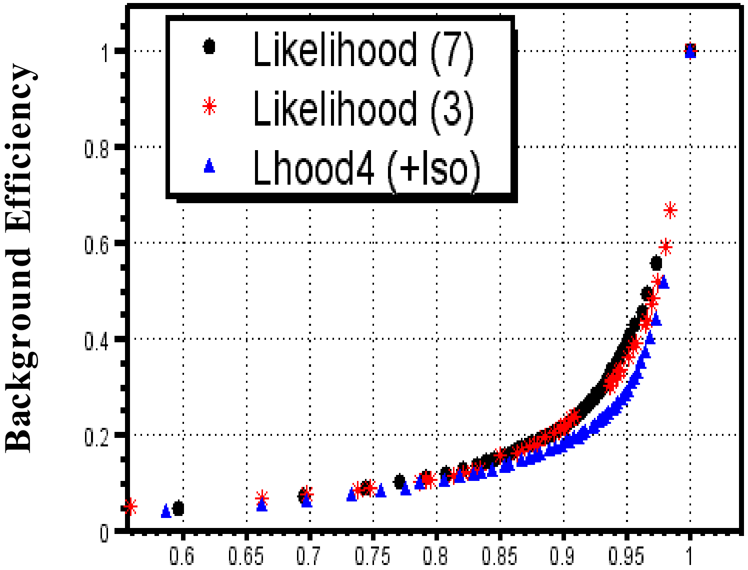
- Typical QCD patterns:
  - Converted leading photon
  - Tails of the fragmentation: leading pi0 overlaps with pi+
- Normalization of the fake probabilities:
  - "Per jet" (more traditional)
    - energies of a jet and a fake electron are different
    - Expect sample dependence (top plot)
  - "Per EM object "– use the same variables
- MisID Probabilities are low
  - D0: (0.6+/-0.1) % per jet w/o preshower
  - CDF : per jet is about 10^{-4}
- Sample dependence: ~ 30-50%
- In many cases QCD backgrounds are small, large uncertainty is more matter of principle

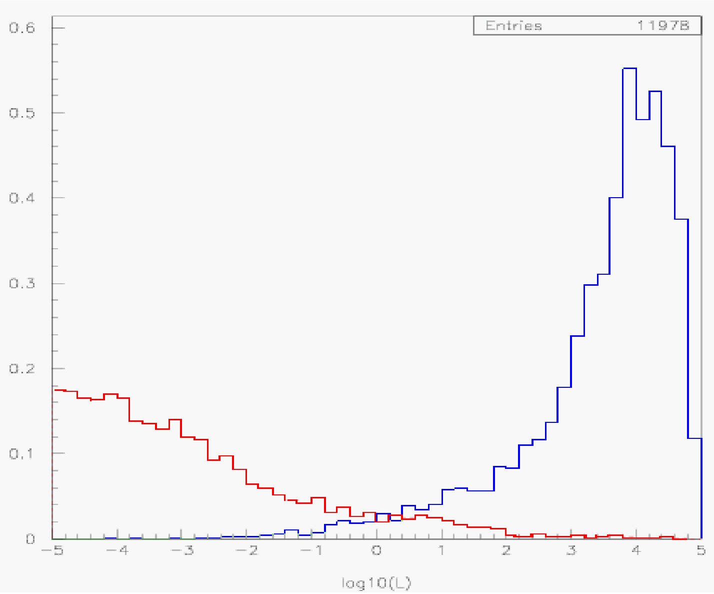




Electrons: likelihoods

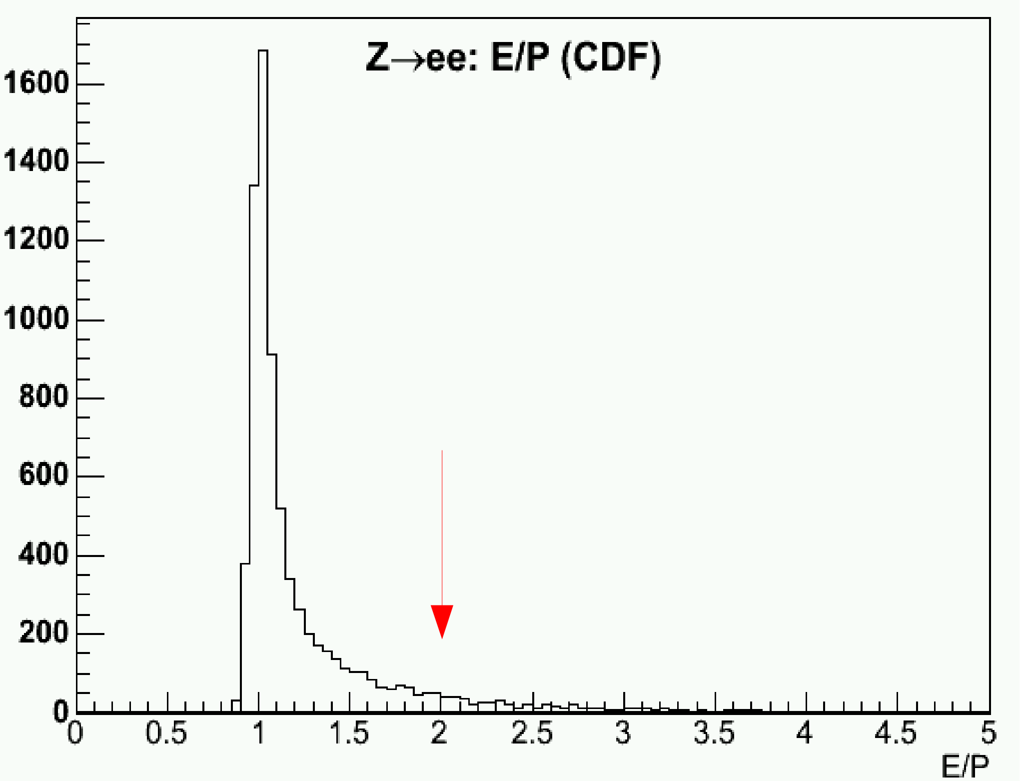
- D0: “H-Matrix”
  - Takes 7 or 8 longitudinal and lateral shower distribution variables and calculates a χ^2 discriminant
    - Layer energy fractions
    - Lateral shower widths
  - Currently tuned with full Monte-Carlo simulations
- CDF:
  - +5% efficiency
  - 40% better QCD background rejection
- Decorrelation
- Stability wrt the definition of the likelihood

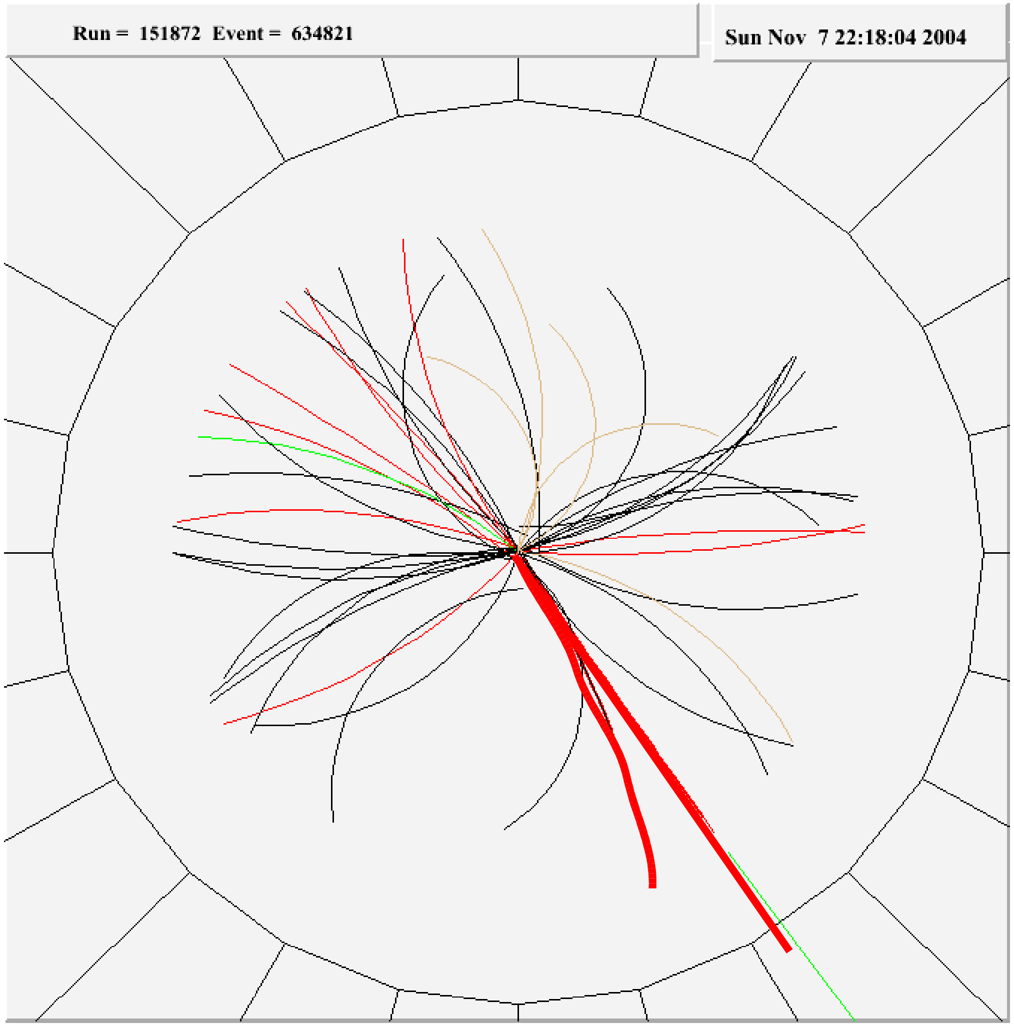




Learning from the likelihoods

- Why likelihood-based approach performs better? - no E/P cut
- What does this cut remove ?
  - asymmetric conversion pairs
- CDF Si tracker – 15% of rad length in average
- CMS – up to 1.5 X_0 … 35%… + pileup



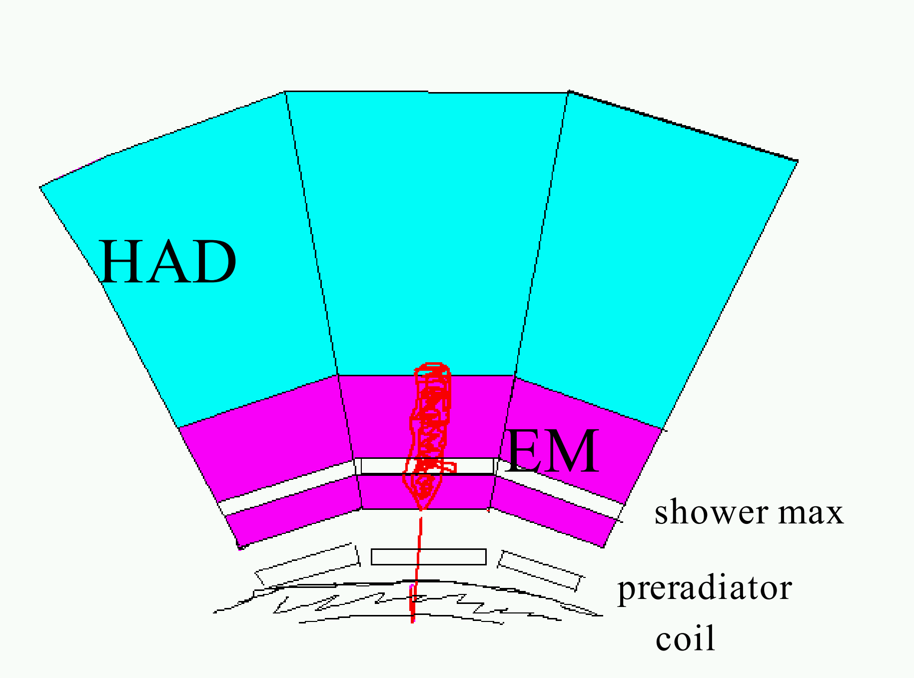


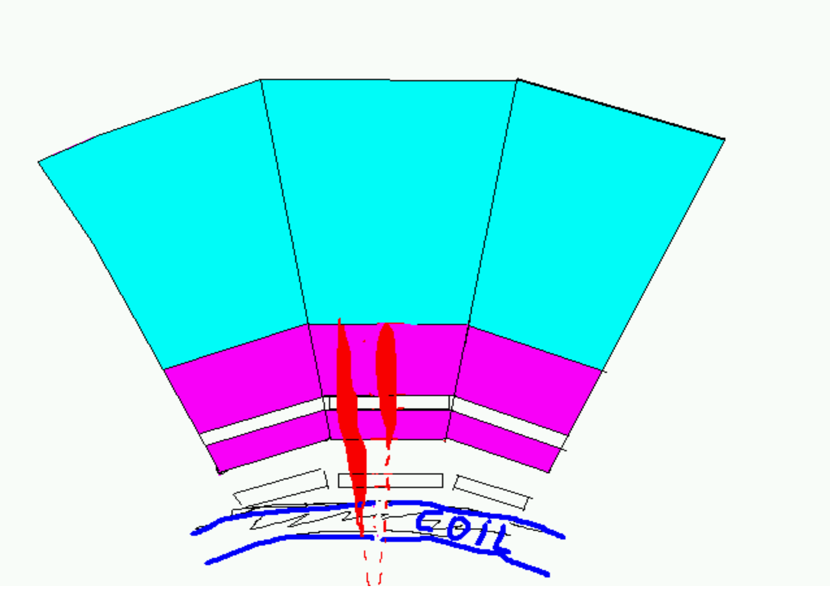
Electrons: summary

- Run II : improve and extend identification techniques first developed in Run I
- ID efficiencies for (Pt> 20GeV) ~ 85-95% up to |eta| ~ 2
- MisID probabilities: “per jet” vs “per EM object”
  - Low, calorimeter alone (D0): 6*10-3 , preradiator commissioned
  - Using shower max information (CDF): ~ (1-2)e^{-4}
  - Forward region (CDF) ~ (5-6)*10^{-4}
  - Shower shape important
  - [jet] sample dependence: ~30-50%
- Conversion removal: important at the Tevatron, even more at LHC (material, pileup)
- Likelihood-based approaches:
  - Typically better S/B than the box-type cuts
  - Useful for estimating the backgrounds
  - Breakdown of improvemnts: x*(better technique) +(1-x)* smarter people
  - X = ?

Photons

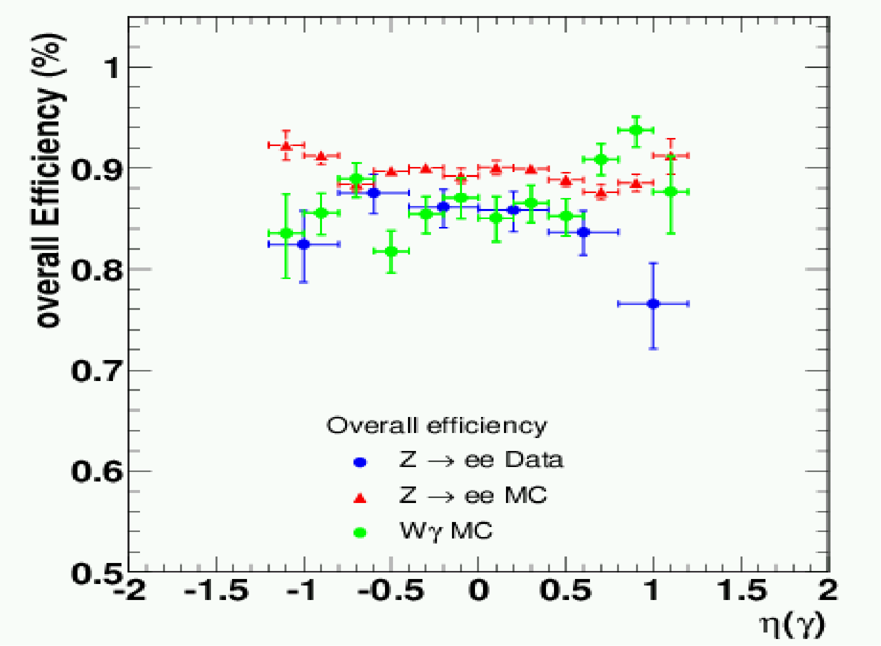
- ***A* photon:**
  - Narrow EM shower in the calorimeter, no tracks
  - Cluster in shower max detector, consistent with EM profile
  - No other clusters in shower max detector
  - always "an isolated photon"
- Major background: jets fragments into leading pi0 with 2 photons merged
- CDF shower max ~ 1.8m from the interaction point, symmetric pi0 decay:
  - $\Delta(R\varphi) \sim 50$cm/Et, cluster width ~ 2cm
  - Can't resolve 2 EM showers above ~50 GeV
- Preradiator: $\sim 1 X_0$, probability to have a shower started in the coil: **$P(\pi^0) \sim 2{*}P(\gamma)$ independent on how close the 2 photons are**
- Extend Et range above 50 GeV

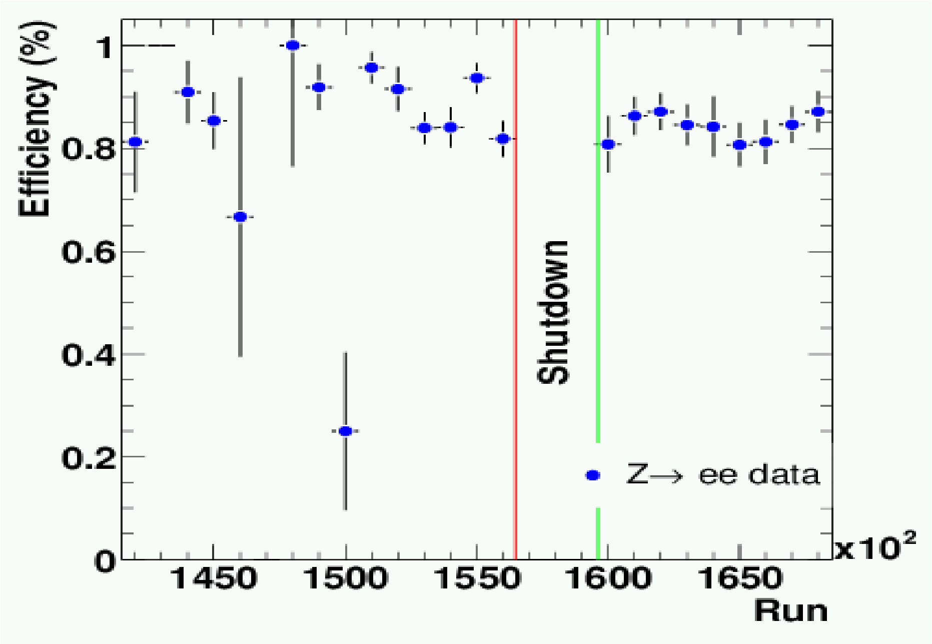




Photons: ID efficiency

- ID efficiencies: no tagged photons
- Use Z->ee data, remove electron tracks
- suppress bremsstralung: 0.9 < E/P < 1.1
- Compare to Z->ee MC , determine scale factor
- **ε_{ID} ~ 90% (Et > 20 GeV, |η|<1)**
- **Scale factor stable: ~ (94±2.3)%**

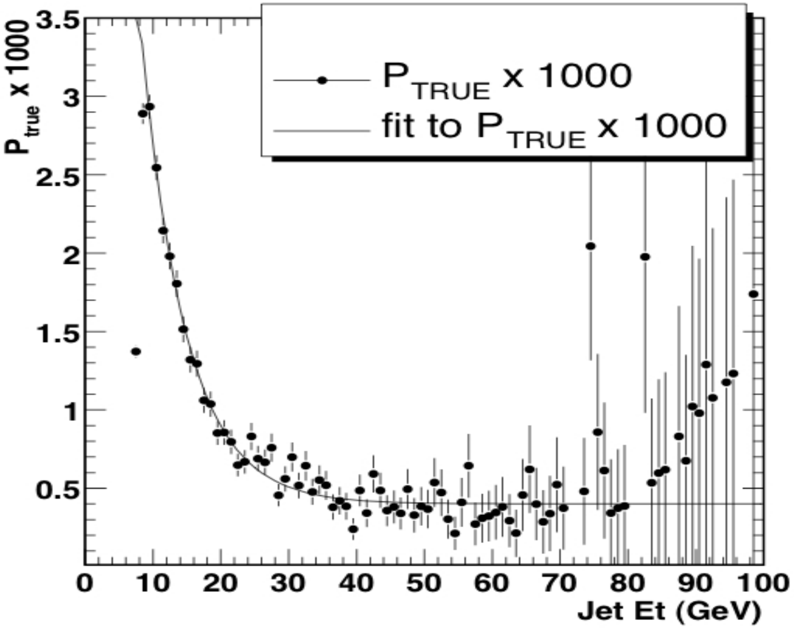


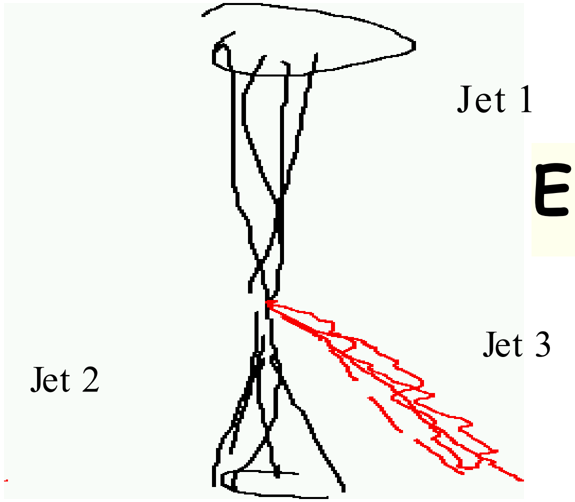


P. Murat, TeV4LHC, 2005/02/05, BNL

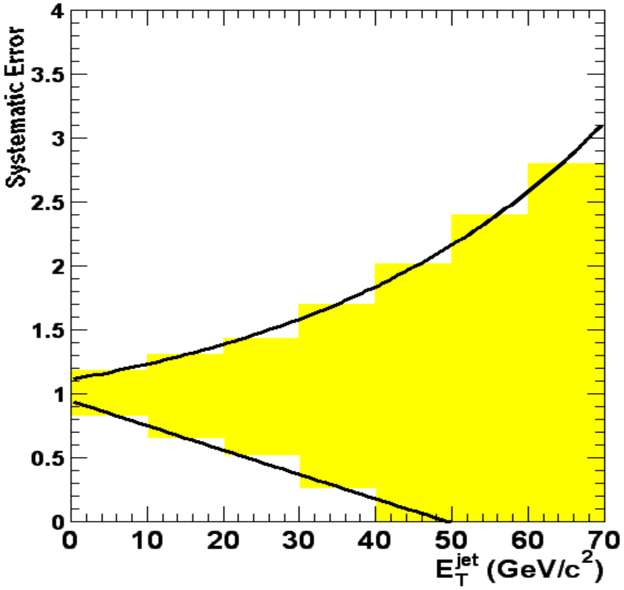
Photon ID: misidentification rates

- Fake rate:
  - probability for a pi0 to be reconstructed as a photon
- Measure in the data (jet samples with different trigger thresholds 20, 50, 70 and 100 GeV)
- Order jets in Et, ignore the 1^{st} one (trigger bias)
- use 3rd, 4th, 5th …, highest E_T jets
- **Jet#2: measure lower mis-ID rate than jet#3**
  - **Lower fraction of prompt photons**

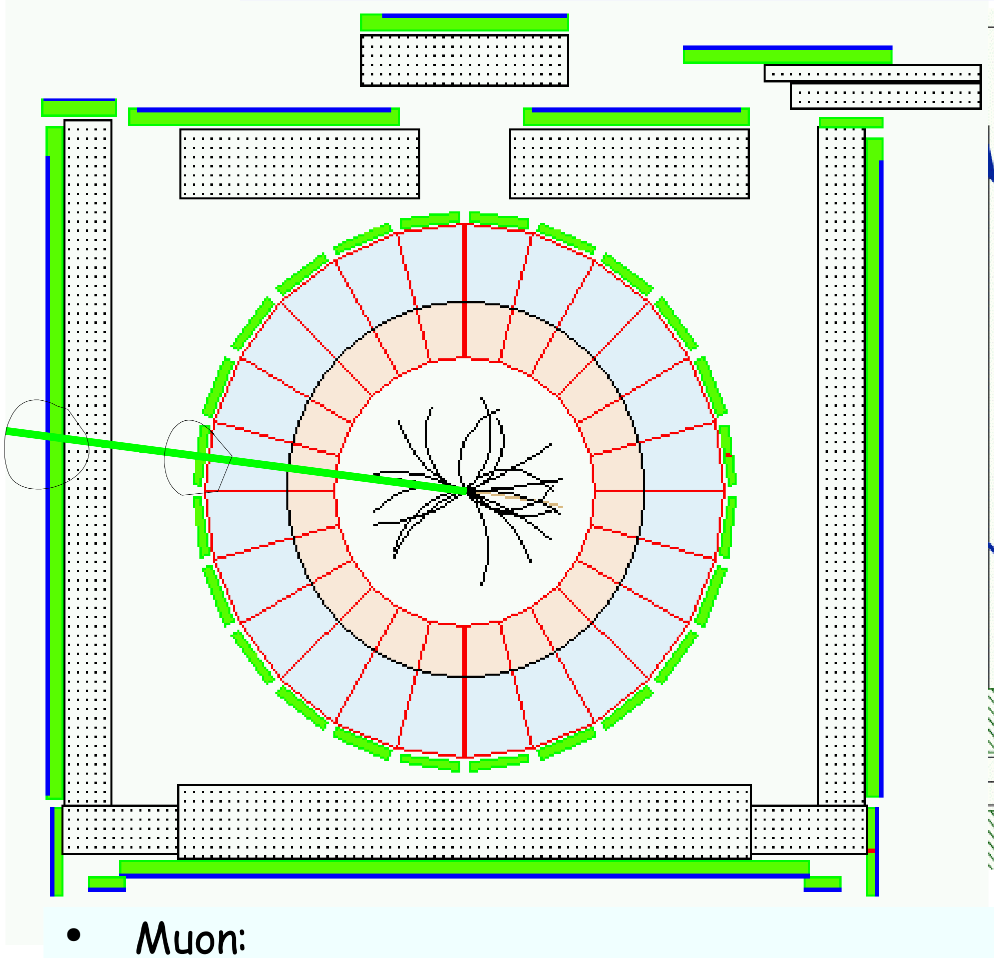


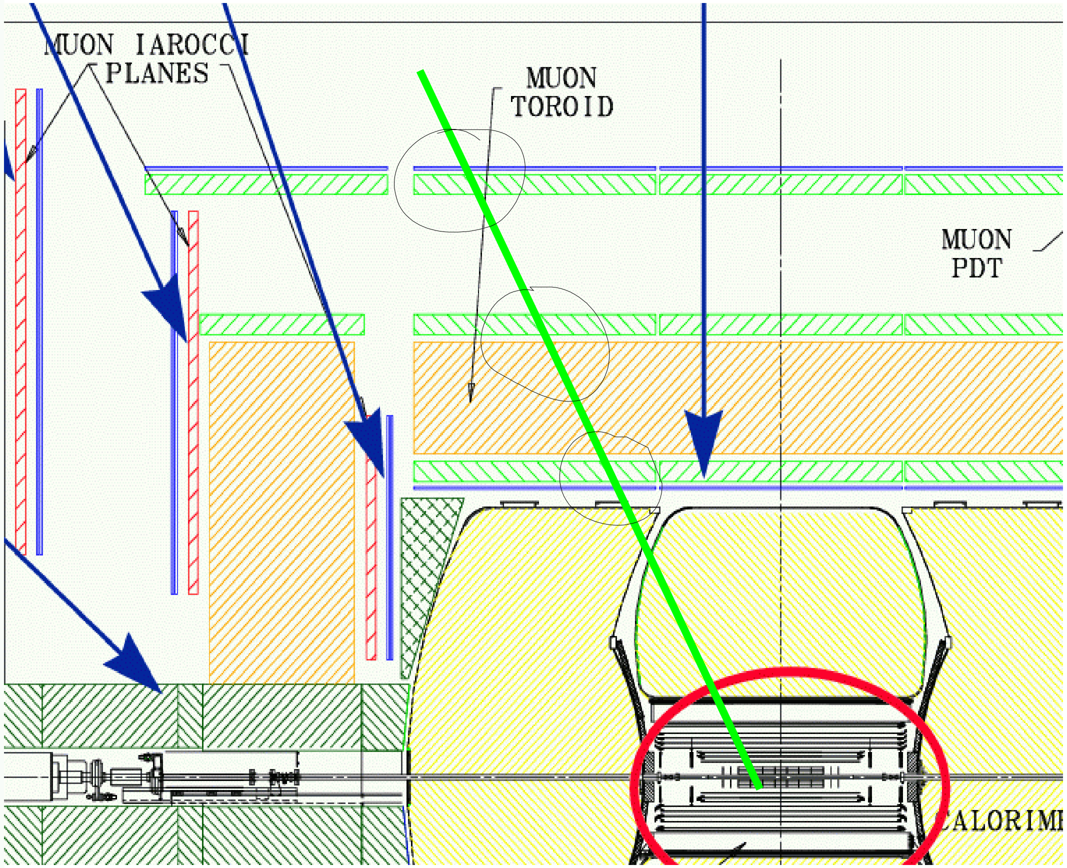


$E_{fake} \sim (5+/-5)10^{-4}$ @ 50 GeV



Muons

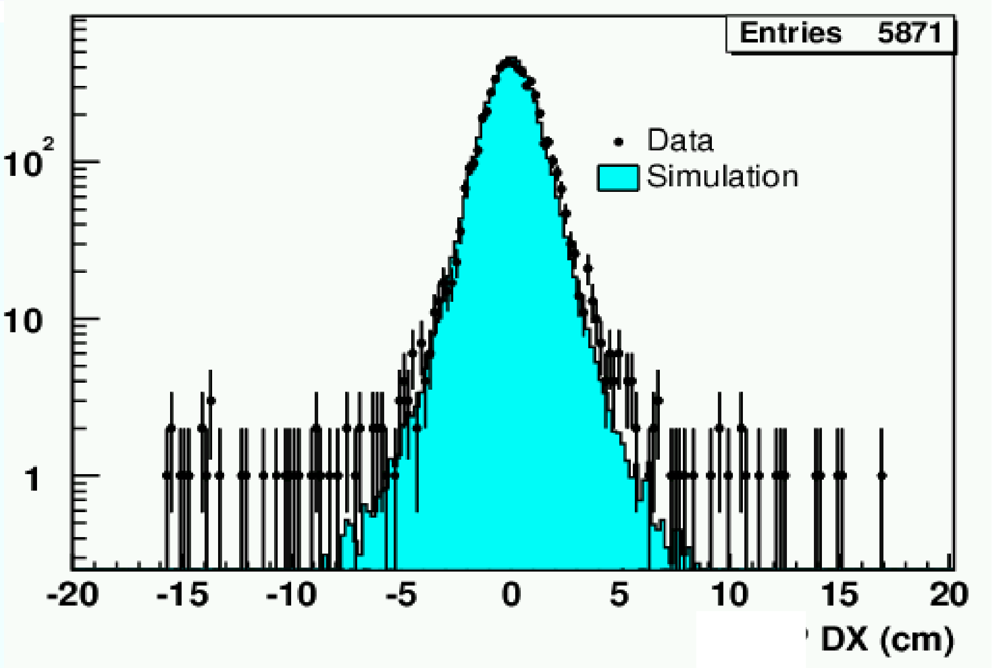


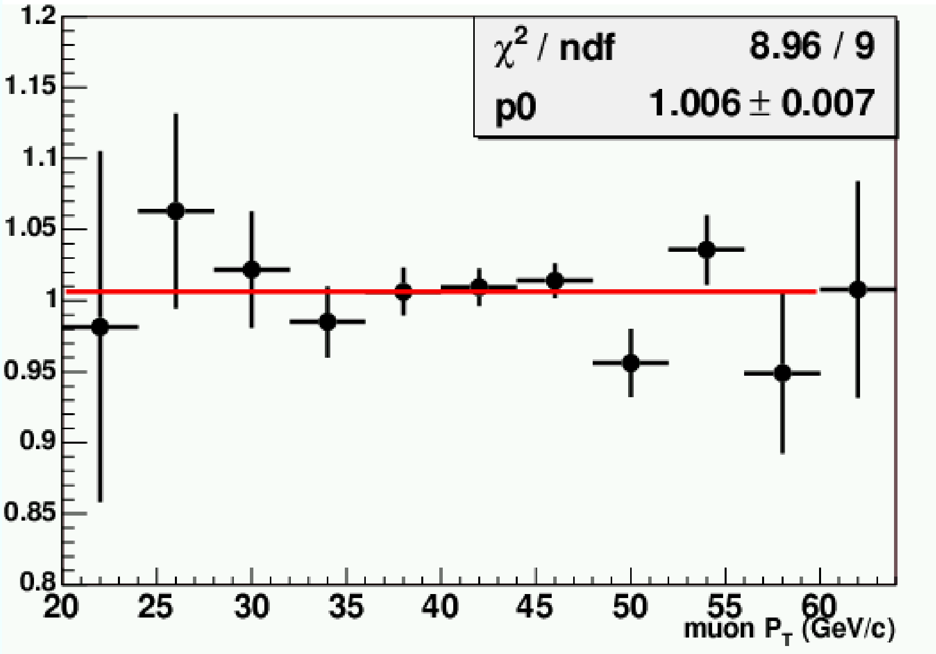


- Muon:
  - track pointing to 1 or more stubs in the muon chambers
    - D0 : 3 stubs , 2-3 layers of sc. Counters, standalone measurement of muon momentum
    - CDF: 2 stubs, 1 layer of scintillators not in the trigger
  - Consistent with MIP energy deposition in the calorimeter (D0 – cross check only)
  - Isolated in the calorimeter, less often - in the tracker
  - Timing: muon scintillators/ CDF calorimeter scintillators

Muons: ID efficiency

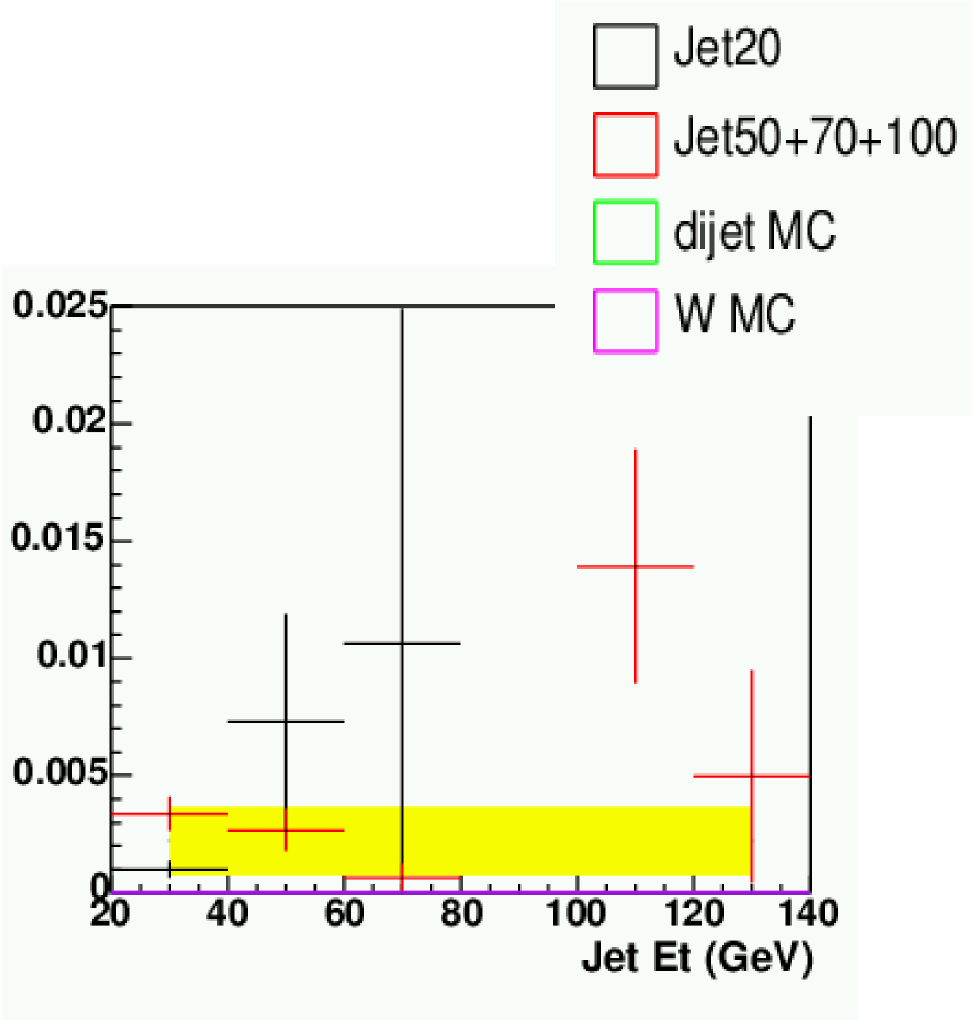
- Need to deal with
  - Muon stub reconstruction efficiency
  - Track extrapolation
    - fringed magnetic fields
    - Multiple scattering
    - alignment
  - Fake muons - hadrons punching through
  - Real muons from pi/K decays in flight – suppress by requiring good quality of the track fit
  - Cosmic muons
- Several categories of muons – from tight to loose
- CDF(E_{ID}) ~ 90% for the "best quality muons"
- **Data-to-mc scale factor consistent with 1**





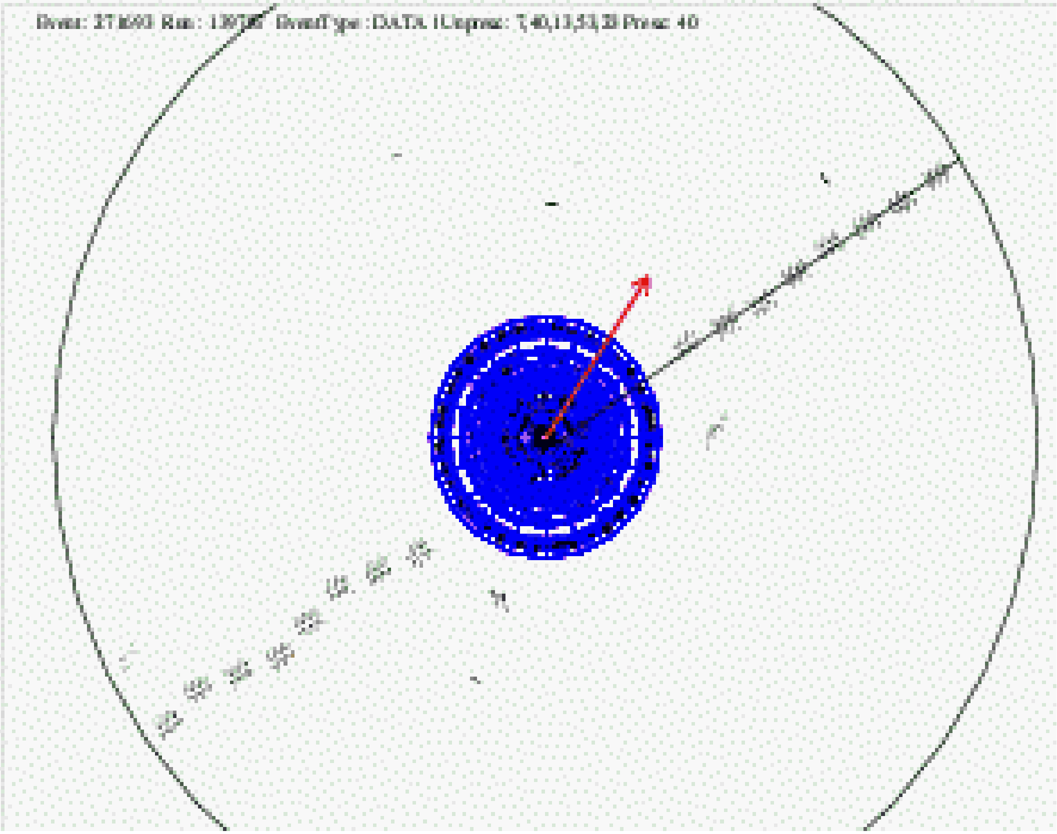
Muons: misidentification probability

- Mis-ID probability (CDF):
  - probability for a high-Pt track to be reconstructed as a muon
- Of the order of 1% (CDF), significantly lower for D0
- Use generic jet samples, accuracy severely limited statistically
- backgrounds due to the fake central muons are small
- might be useful to implement special backup triggers aimed in misID rate measurements



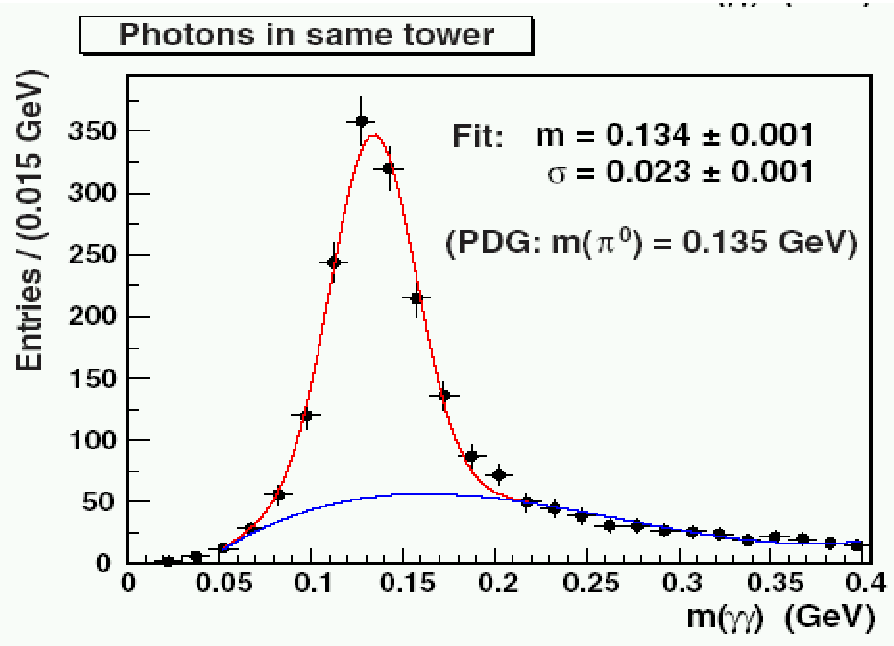
Rejection of the cosmic muons

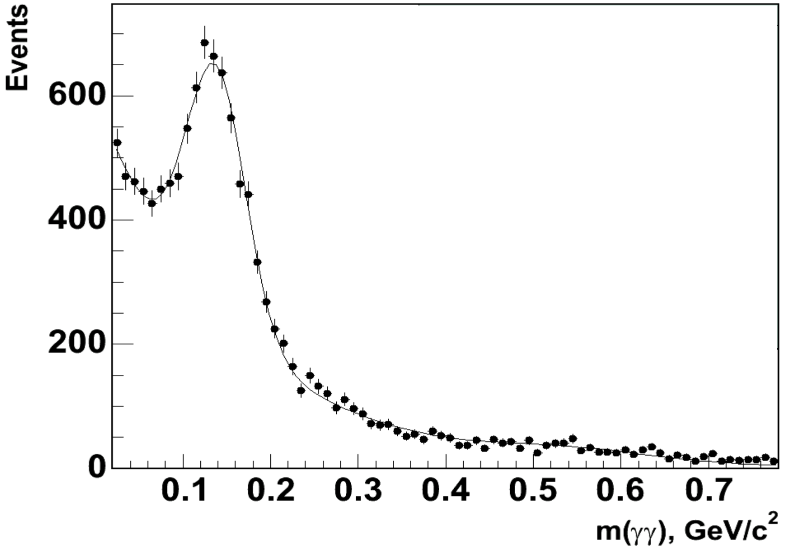
- D0 uses robustness and redundancy of the muon system
  - Timing from the muon scintillators + track impact parameter
- CDF relies on the efficiency and resolution of the drift chamber
  - For each muon candidate try to reconstruct its 2nd leg
  - If found, test if 2 legs correspond to:
    - 2 particles
    - 1 particle
  - If the best chi2 corresponds to 1 particle may call it a cosmic muon
- for inclusive W and Z cross section measurements background from cosmic muons is negligible



Tau ID: reconstructing pi0's

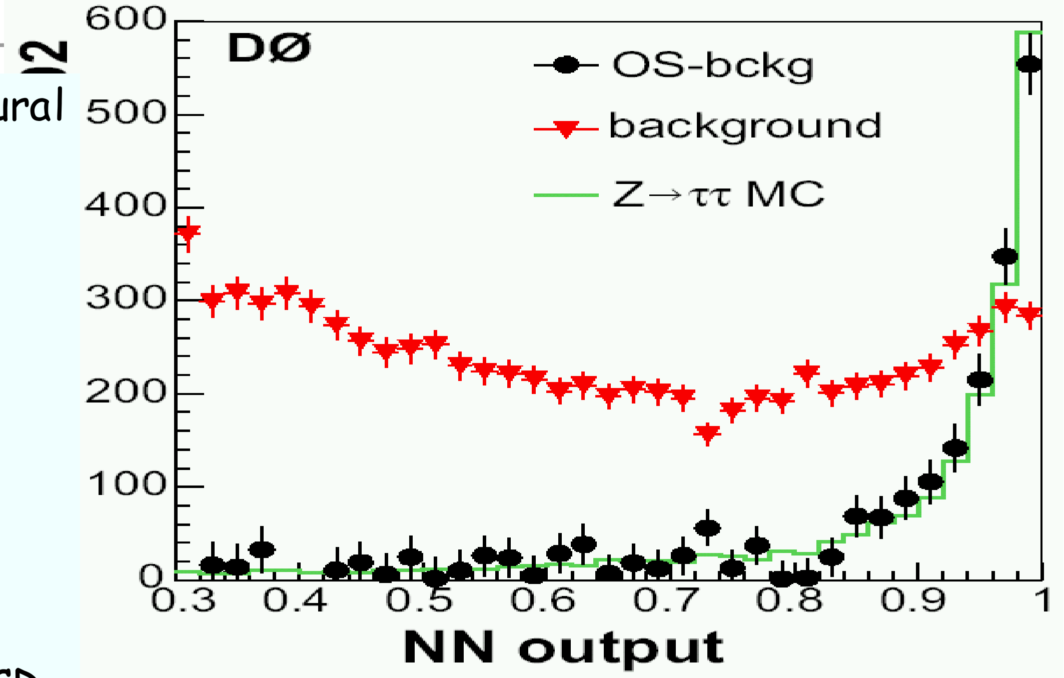
- Tau decays result in many pi0's/photons in the final state
- Coordinate resolution of the calorimeter alone not enough to resolve them
- Both Tevatron experiments demonstrated their ability to reconstruct pi0 using
  - shower max detector (CDF) with coordinate resolution about 2-3 mm
  - preshower (D0), similar resolution

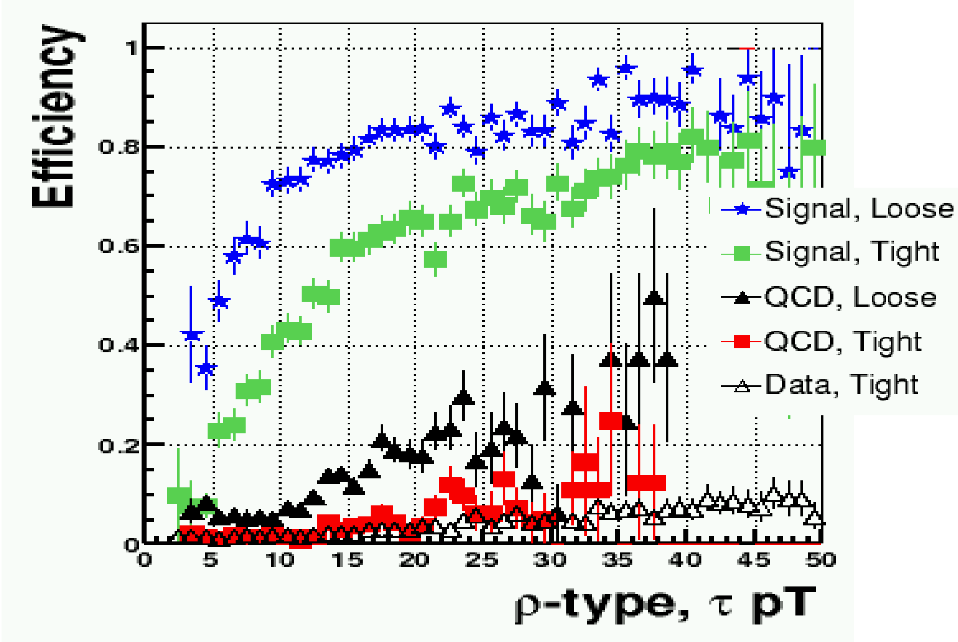




Tau ID: neural networks (D0)

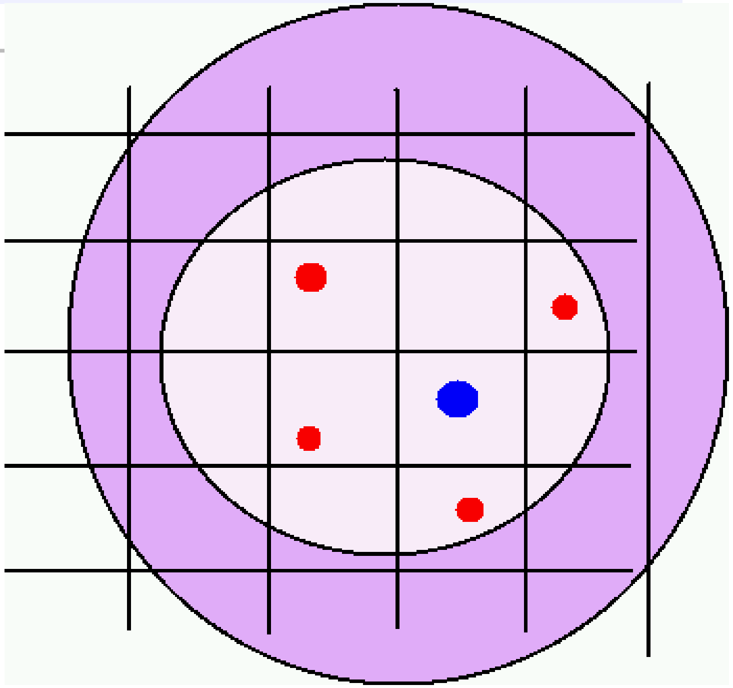
- D0 measured Z->tau tau cross section using neural network-based approach
- Trigger on muons, identify Z->$\tau(\mu)$ $\tau(e/h)$ events
- 3 classes of events, 3 separate NN's
  - τ^{-} -> π^{-}/K^{-} ν
  - τ^{-} -> π n(π^{0}/ γ) nu / e $\nu\nu$(rho-type)
  - τ^{-} -> π- π- π+ ν (K..)
- Training: use single τ +minbias events MC and QCD data
- ε_{ID} **platoes at ~ 80%, jet misID rate ~ 5-10%** (caveat: normalized to "type2-looking jet", x5)
- Flattish distribution in X_{NN} for background useful for cross checks

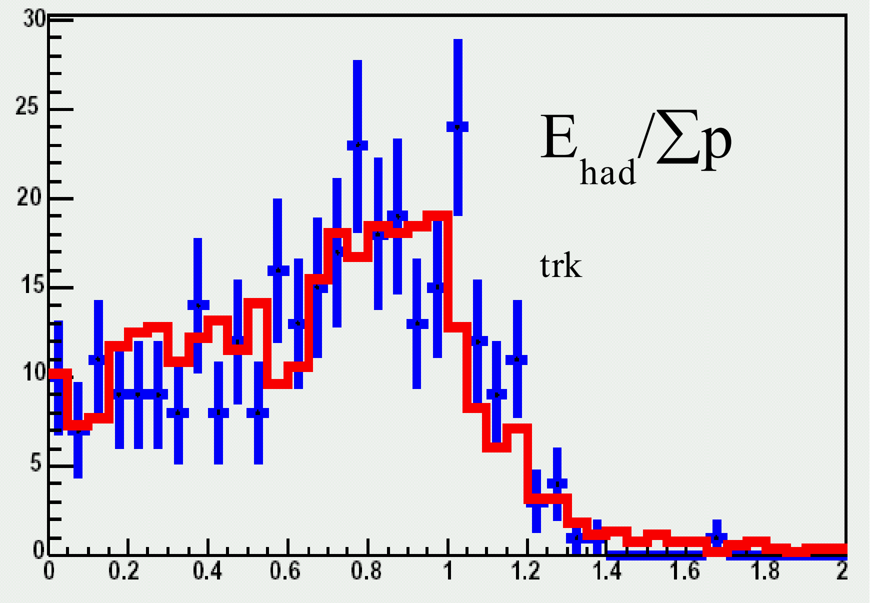




Tau ID: cut-based approach (CDF)

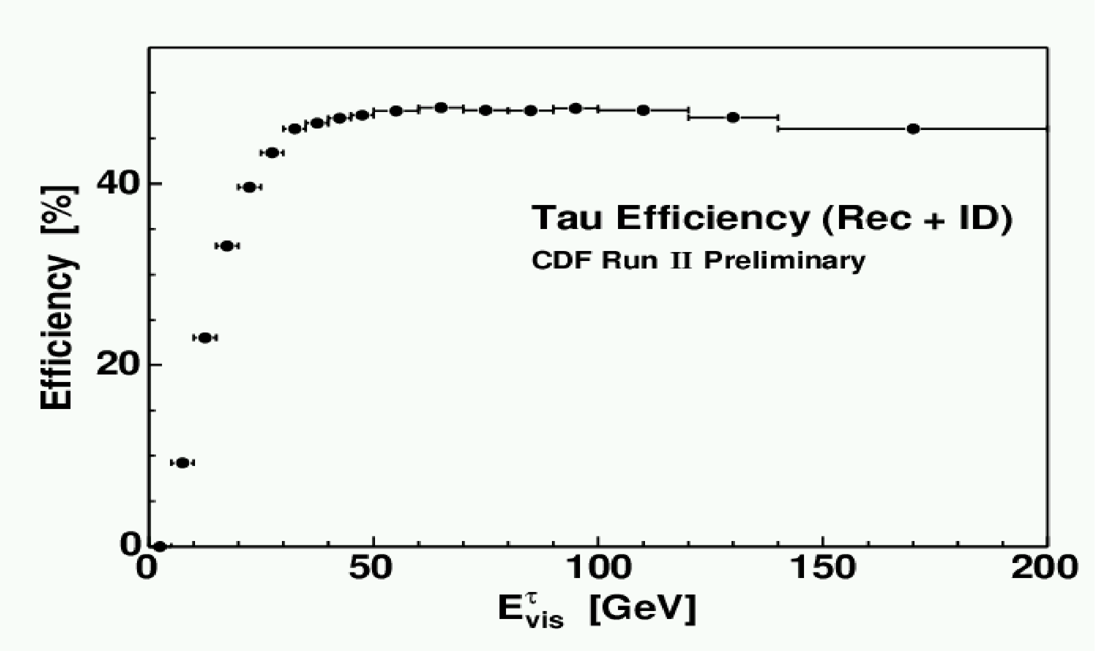
- Target on hadronic decays of taus
- Signature: narrow jet with low track multilicity
- Find narrow clusters in the calorimeter, count **tracks pointing to a cluster**
- Use shower max detector to reconstruct **pi0's / photons** down to ~0.5 GeV (resolution in pi0 energy about 25-30%)
- M(tracks+pi0) < Mtau
- Require tau candidate to be isolated
- **Diference with NN-based approach: understand effect of each cut separately**
- **Electron removal**: require energy deposition in the hadron compartment not to be small compated to the sum of track momenta (typical cut values used ~0.1-0.2)

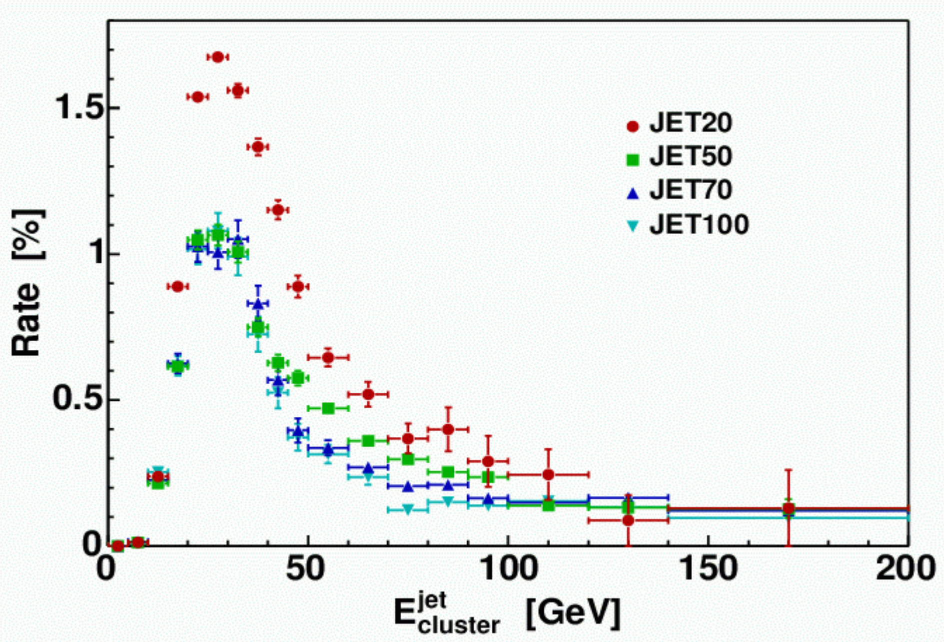




Tau ID:

- efficiency platoes at about 50%.
- Fake probability falls with energy (mass cut)
  - ~1% at 20 GeV, 0.2% at 100GeV
- Need to understand different from NN dependence on energy
- Sample dependence: ~50%
- 2D parametrization: E_{jet}, $\gamma = E_{jet}/M_{jet}$ reduces sample dependence to ~20%
- Jet and tau reconstruction algorithms calculate parameters of the same object differently - **determine fake probability per “very loose tau candidate” (~3 times higher)**
- Still art, more intellectual effort needed





Summary

- Lepton/photon ID techniques at the Tevatron are well established
- Reliable calibration sources (W/Z, J/psi's, upsilons) exist in Pt range ~ few-50 GeV
- Electrons/muons/photons (Pt > 20 GeV): ε_{ID} are 90% and above, fake probabilities low (10^{-3}-10^{-4})
- Taus – ID efficiencies in the range (50-80)% demonstrated, fake rates vary significantly depending on the approach (NN vs "box") (0.2-2% at 100 GeV)
- QCD mis-ID probabilities: sample-dependent at the level of 30-50%, still art
- understanding misidentification needs a lot of thought put into design of the calibration/backup triggers

Hymalayan Crystal Salt –
the purest form of Natural Salt

Backup triggers with prescales above 10 often not enough