Review of Preparations for V+Jets in CMS

Piergiulio Lenzi
on behalf of the CMS collaboration

MC4LHC Workshop
01-04-2010

Outline

- Motivations for V+jets studies
- Observables
  - What we will study in the beginning
  - Precision studies
- MC generators for V+jets in CMS
  - What we use
  - What we would like
- Status of analysis in CMS
  - Analysis strategy for early data
  - Long term plans
- Plans

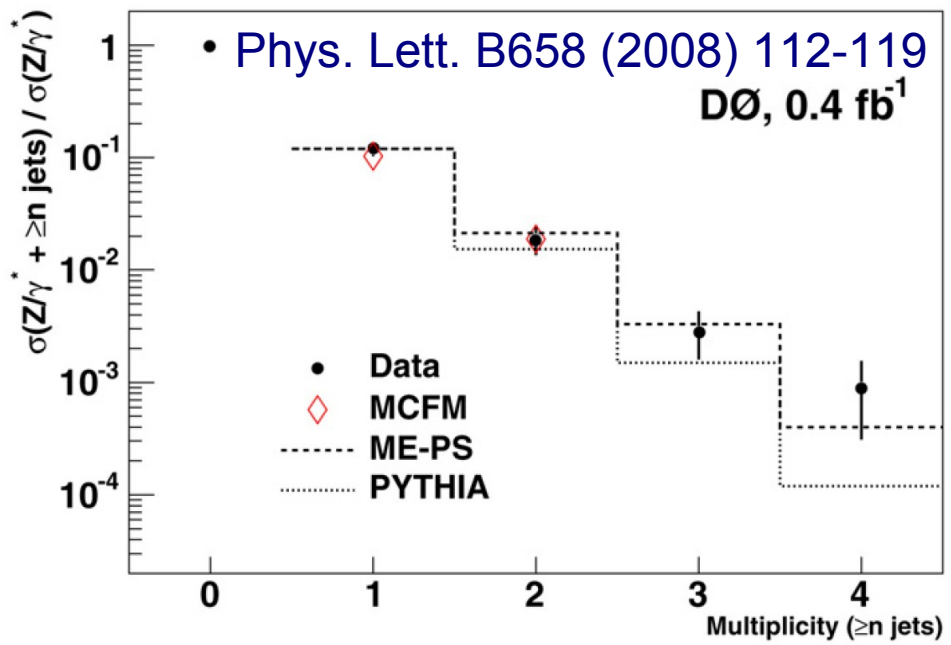
Motivations

- Important test of perturbative QCD
  - Compare rates to NLO predictions (MCFM, BlackHat, Rocket)
  - Compare shapes and relative rates to matrix element + parton shower calculations
- Final states with a vector boson plus jets are background to many new physics processes
- Important for detector commissioning
  - Jet energy scale calibration
  - MET commissioning in W events using Z

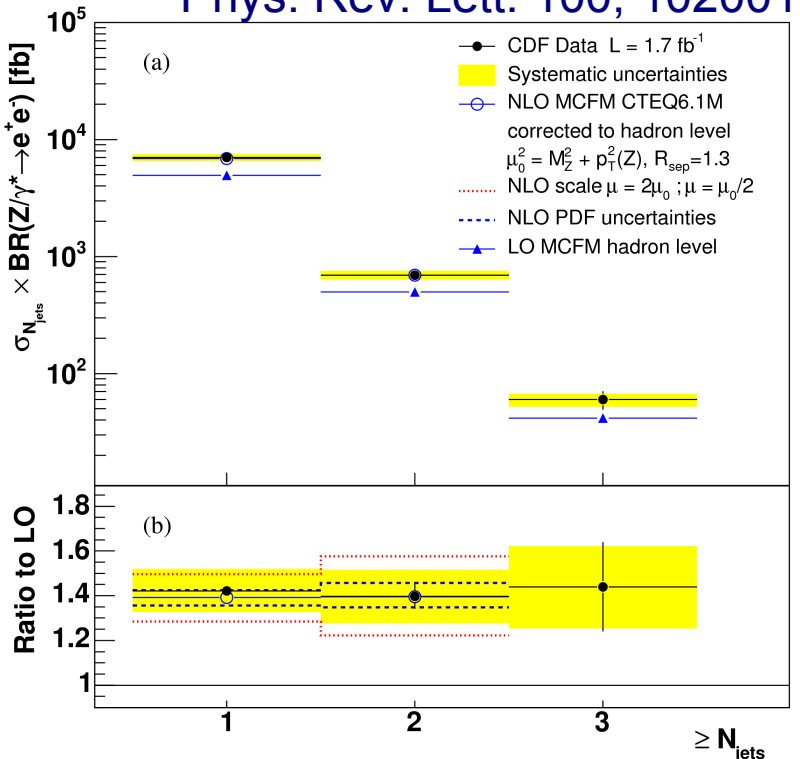
Observables

Short term plan, O(50 pb^{-1}):

- Measure the rate of jet production in association with a weak boson
  - Inclusive jet counting
  - $(1/\sigma_0)\ d\sigma/dNj$, $d\sigma/dNj$, $\sigma[V+Nj]/\sigma[V+(N+1)j]$
  - Comparison of rates in association with W and Z, $\sigma(W+Nj)/\sigma(Z+Nj)$



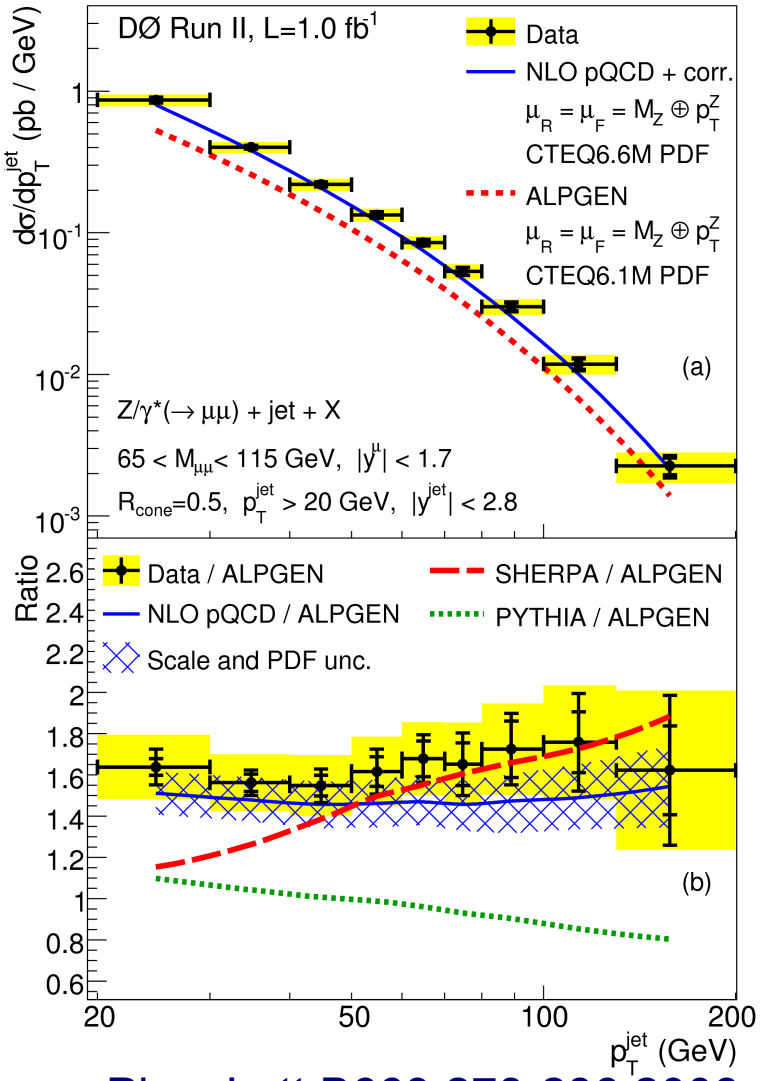
Phys. Rev. Lett. 100, 102001



Observables

Longer term plan:

- Characterize V + jets in greater detail
  - $d\sigma/dE_T$ for each jet, $d\sigma/dR_{jj}$, $d\sigma/dM_{jj}$, $d\sigma/d\Delta y_{jj}$
- Events shapes
  - Differential jet rates
- We would also like to look into variables related to rapidity gaps in V+2 jets
  - Angular decorrelation
  - Average number of jets VS rapidity gap



Phys.Lett.B669:278-286,2008

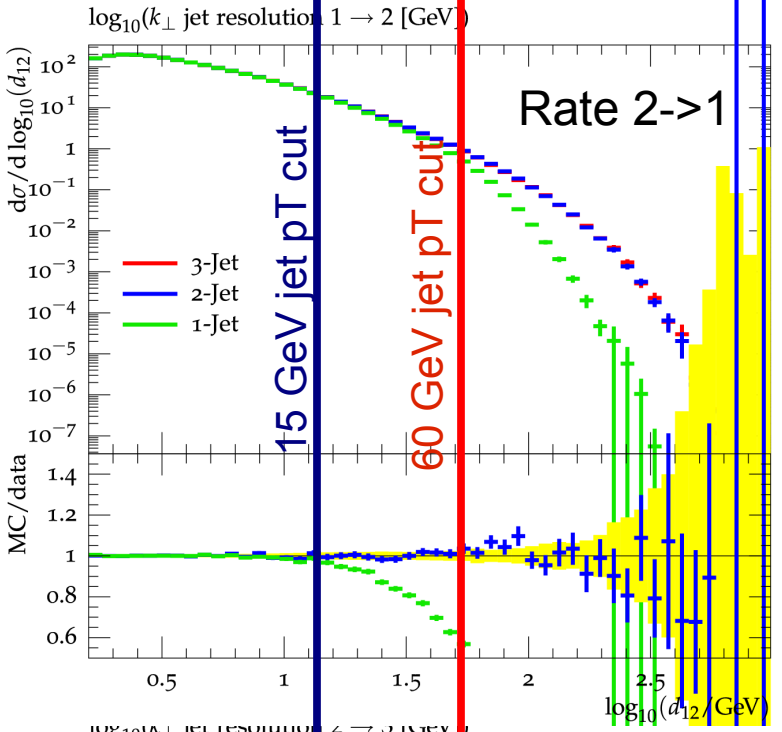
Observables

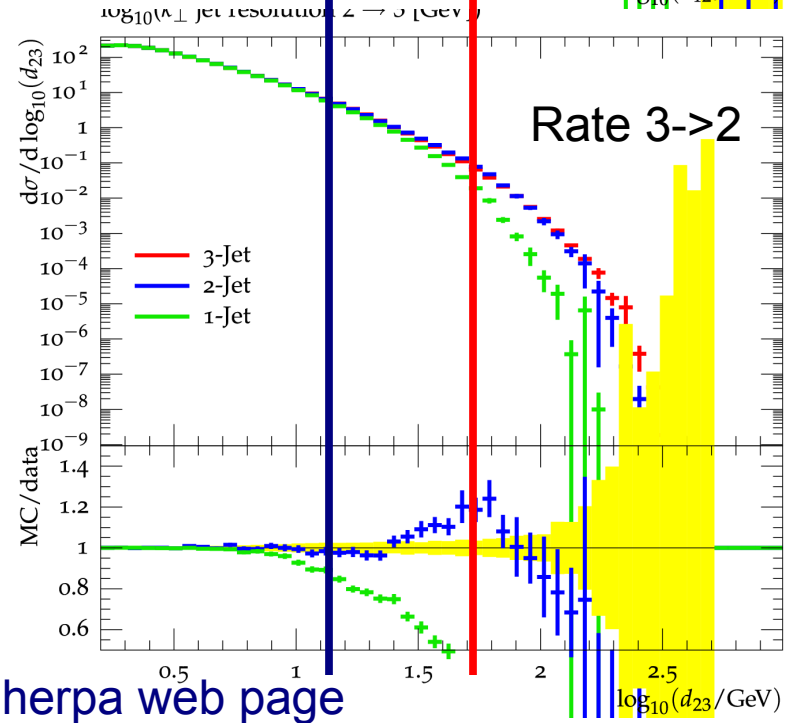
Jet algorithms

- Only infrared-collinear safe algorithms used in CMS
  - Anti-kt is our default
- Where to put the pt cut?
  - Low p_T cut increases the statistics (especially for early data)
  - High p_T cut discriminates between models
  - Effect of underlying event

Differential jet rates:

- Describe the QCD radiation pattern accompanying V production
- Very challenging measurement



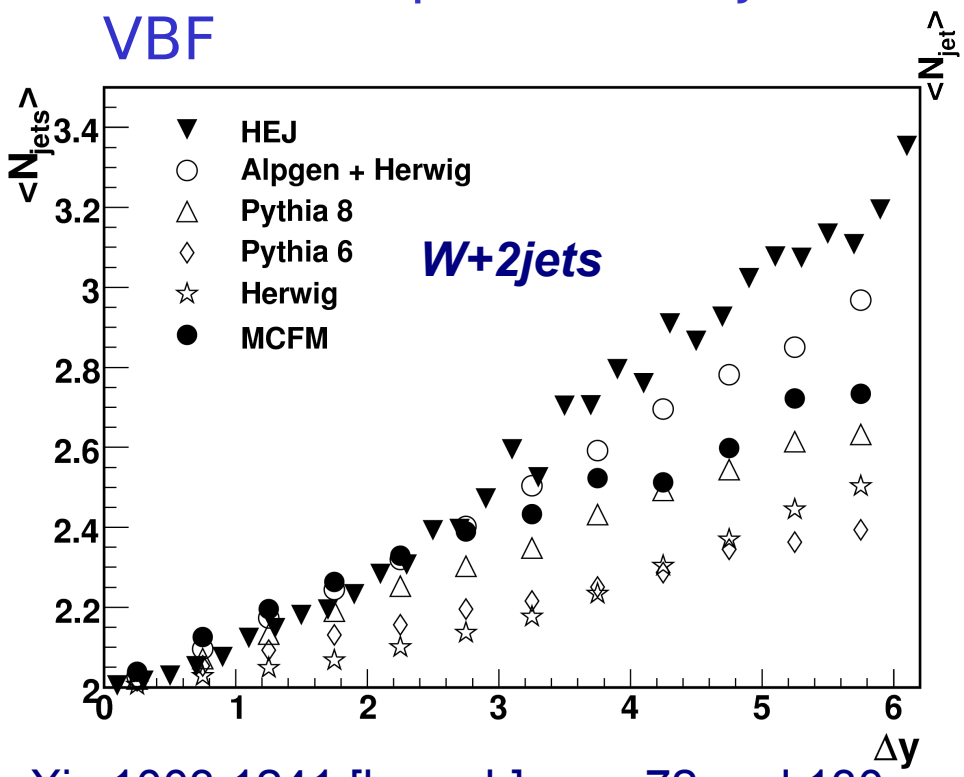


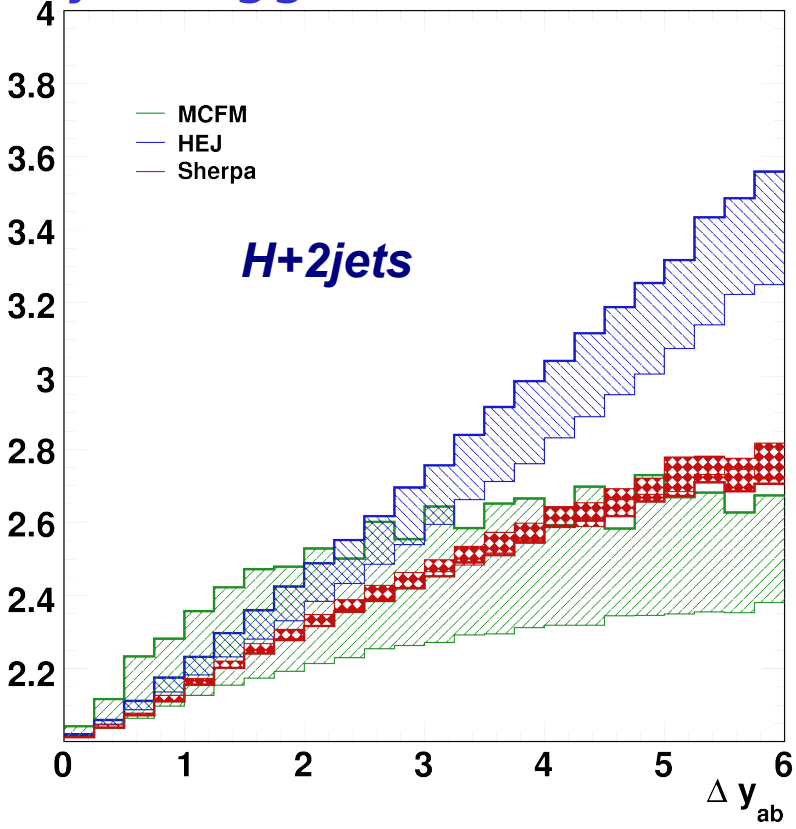
plots from Sherpa web page

Observables

Rapidity gaps:

- Average number of jets VS rapidity gap between the forward and the backward jets
- Similar pattern in W+di-jets and in gluon fusion Higgs
- We can exploit W+di-jet to study the gg contamination to VBF





arXiv:1003.1241 [hep-ph], pag 72 and 130

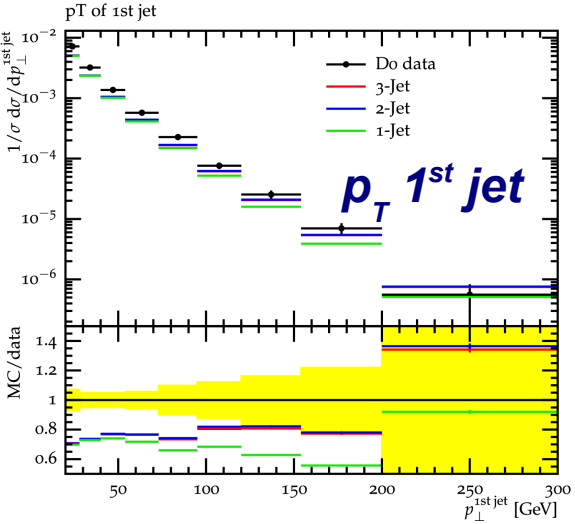
Monte Carlos for V+jets

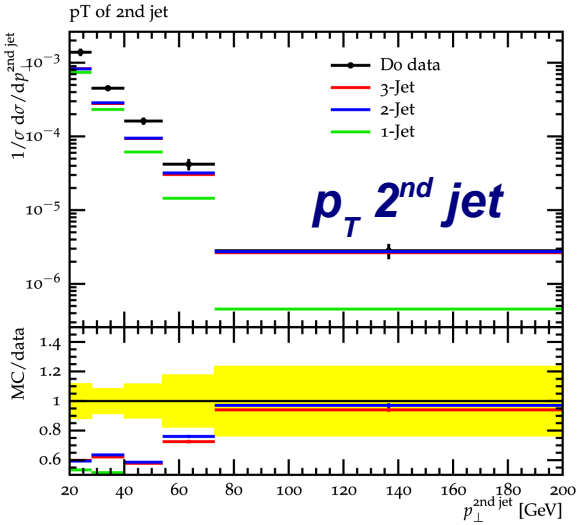
What are the desirable features?

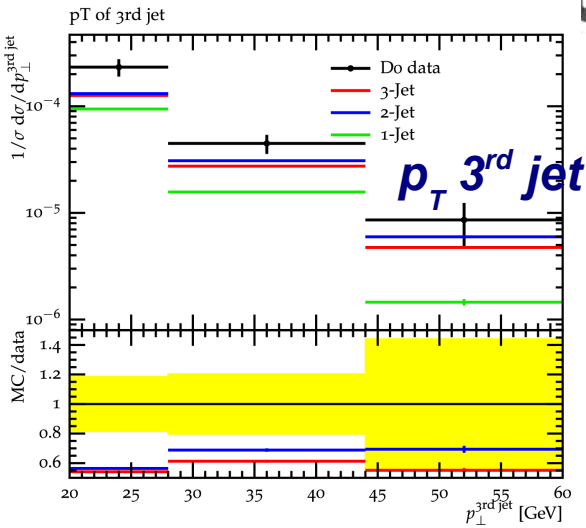
- Describe multi jet topologies
  - ME + PS techniques
- Describe angular separation between jets and leptons
  - Our selections rely on isolation
- Refined treatment of QED fsr
  - FSR photons can fake jets
- Light and heavy flavor jets
  - We have both in the data
- It has to be tuned to data
  - Extremely important for a good understanding of isolation efficiencies

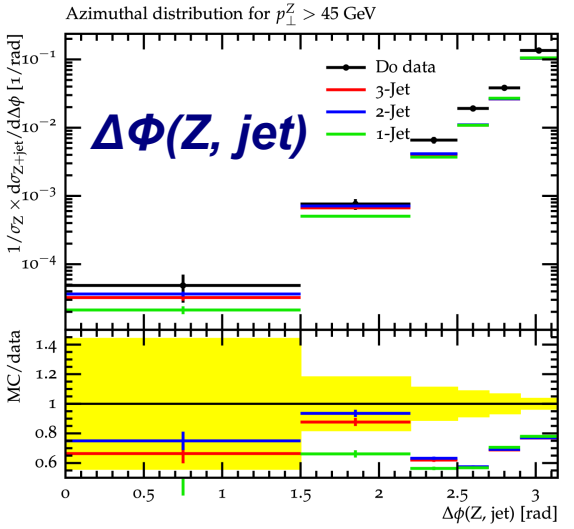
In the beginning we will not care too much about absolute rates, but we need tools that are good at shapes

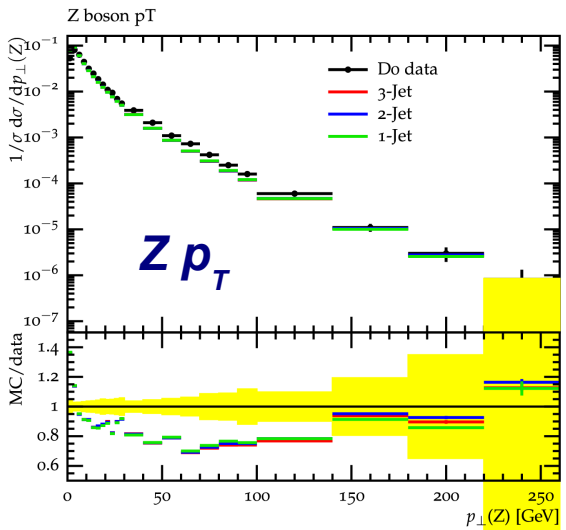
Comparisons to Tevatron data











- Need for a clear picture of how different generators compare to Tevatron
- Example: Comparison to Z + jets @D0

plots from Sherpa web page

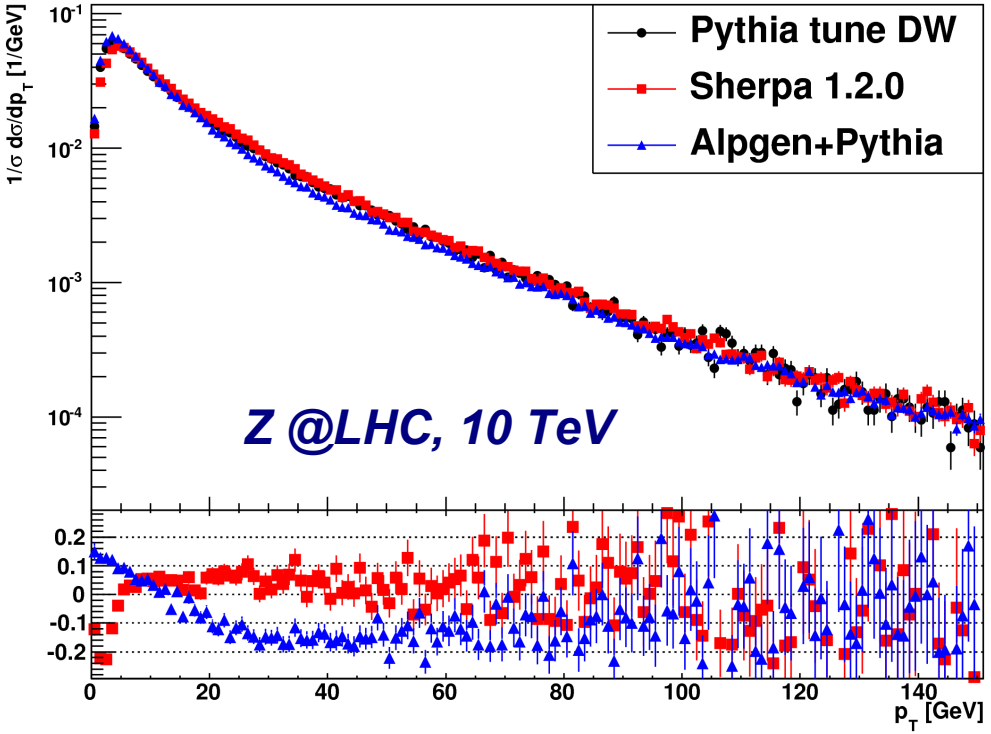
MC for V+jets @ CMS

We are using all the most popular MC tools available:

- Alpgen (+Pythia)
- MadGraph (+Pythia)
- Sherpa

The tuning is one of our major concerns

- How does the matching interact with/spoil Pythia tunes?
- Is there an interplay between the tuning of α_s and the matching?
- What is the tuning status of Sherpa?



arXiv:1003.1643[hep-ph]

Parton level calculations

- MCFM
  - To compute overall k-factors
- BlackHat and Rocket
  - We are in contact with the authors to get ntuples for W+1,2,3 jets at NLO 7TeV
  - It would be nice to have samples produced with at least two scale choices
  - We would like to have parton level ntuples, not with already formed jets, so that we can recluster partons with different algorithms/cuts

Analysis strategy for early data @ CMS

- Cut based event selection
  - High efficiency to maximize signal yield and to allow background modeling
  - Data driven extraction of signal yields
- Different jet reconstruction algorithms
  - To be robust and to allow checking of our findings using different sub-detectors
    - Traditional calorimeter jets, track jets, Particle Flow(*) jets

(*) Particle flow reconstruction gathers information from all CMS subdetectors and reconstructs all stable particles, classified as electrons, muons, taus, photons, charged hadrons, neutral hadrons

Analysis road-map

- Start measuring ratios
  - Z/W + n jets absolute cross section suffers from experimental uncertainties (luminosity, jet energy scale, acceptance...)
  - Inclusive jet counting (avoid UE promoting n to n+1 jets)
  - We can check the Berends-Giele scaling in Z/W +n jets over Z/W+(n+1) jets ratio
    - Many systematics cancel out
  - Measure the double ratio $\frac{(W+n)/(W+n+1)}{(Z+n)/(Z+n+1)}$
    - Check consistency with 1
    - Keeping W and Z selections in sync allows an almost complete cancellation of reconstruction efficiency
- Proceed with full unfolding of detector effects
  - Deliver data directly comparable with particle level predictions

Event selection strategy

- Single, non isolated, muon and electron triggers
- Electron/muon reconstruction and identification
- P_T > 20 GeV on the leading leg
- Isolation (Tracker+ECAL+HCAL)
- Primary vertex compatibility

Z Specific:
- At least 2 leptons
- 60 GeV < M_{ll} < 110 GeV

CMS-PAS-EWK-08-006

W Specific:
- At least 1 lepton
- Veto on Z like events
- MET > 15 GeV
- MT > 20 GeV (μ), 30 GeV (e)

CMS-PAS-EWK-09-006

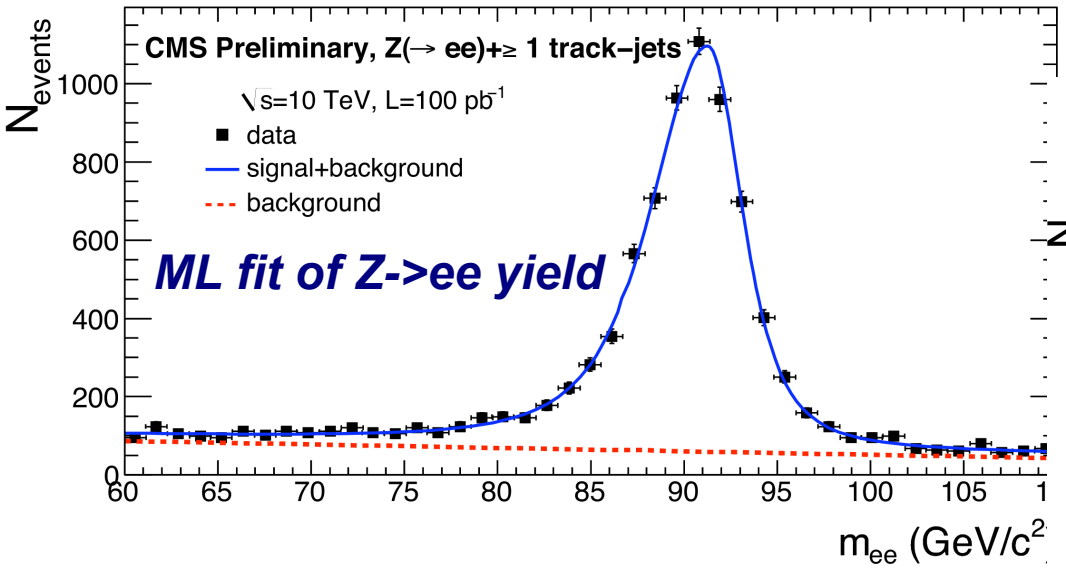
Data driven extraction of Z yield

- For each jet multiplicity bin we performe an unbinned extended ML fit of the di-lepton invariant mass

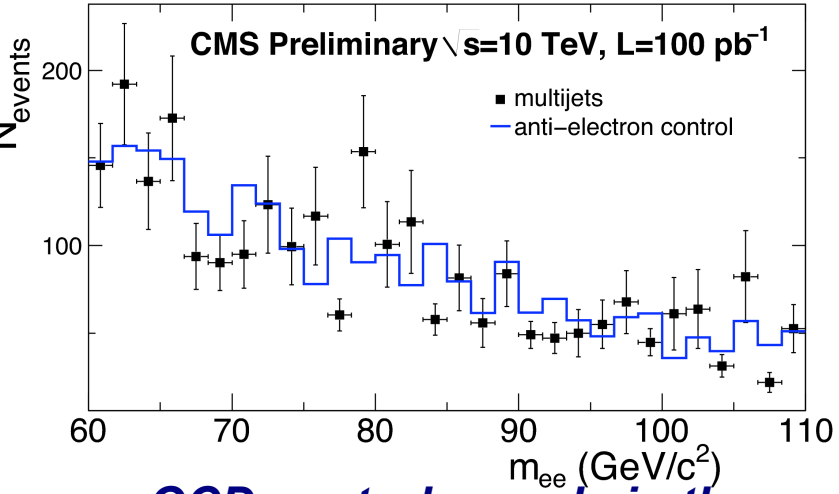
$$L = \frac{e^{-(N_S+N_B)}}{(N_S+N_B)!} \prod_i [N_S \cdot P_S(m(ll)_i) + N_B \cdot P_B(m(ll)_i)]$$

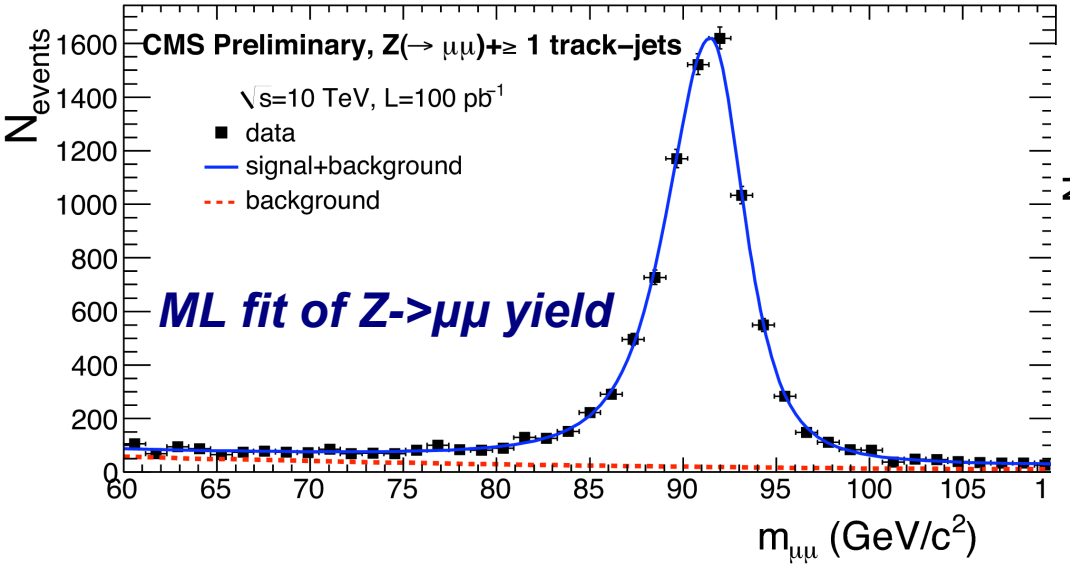
  - N_S (N_B) is the number of signal (background) events
  - P_S (P_B) is the probability density functions for signal (background) of the invariant mass
- We take the shape for the signal from inclusive Z analyses (no dependency of shape on #jets)
- We check the functional form of background on a control sample obtained inverting isolation requirement
- Signal and background abundances are extracted from the fit

Maximum likelihood fit of Z yield

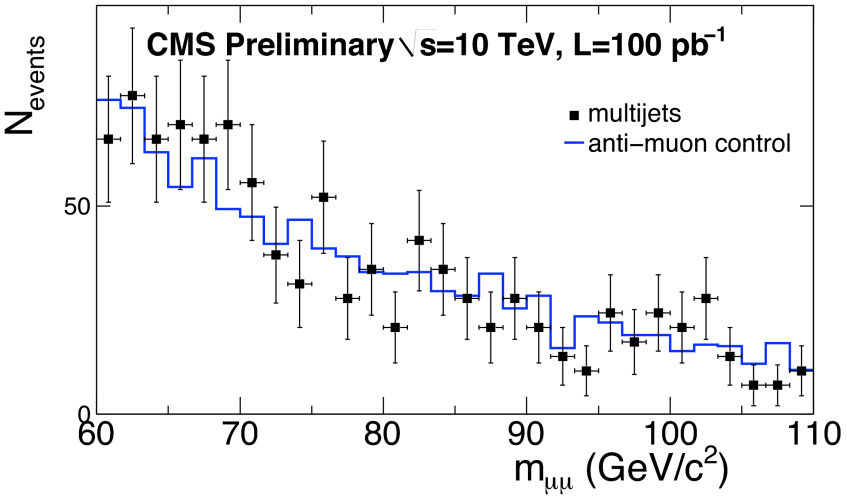


QCD control sample in the electron channel





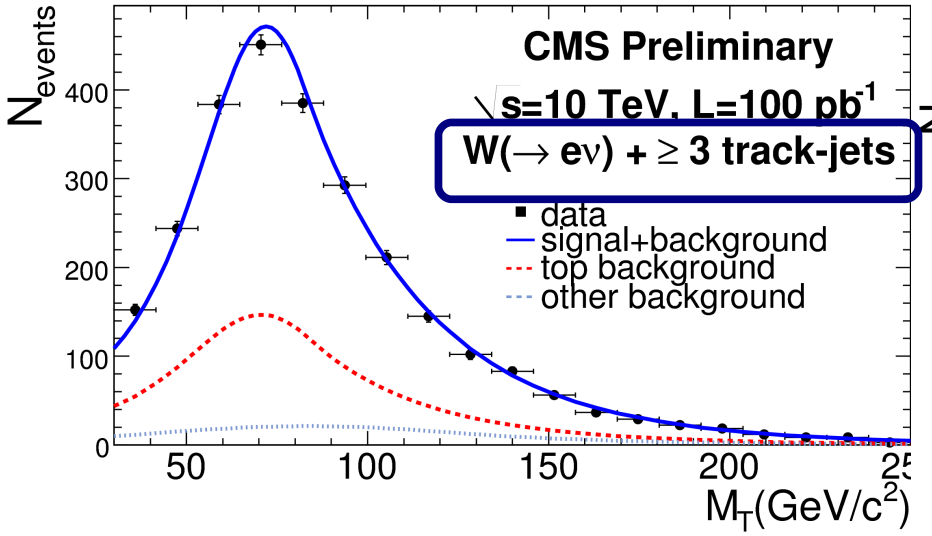
QCD control sample in the muon channel

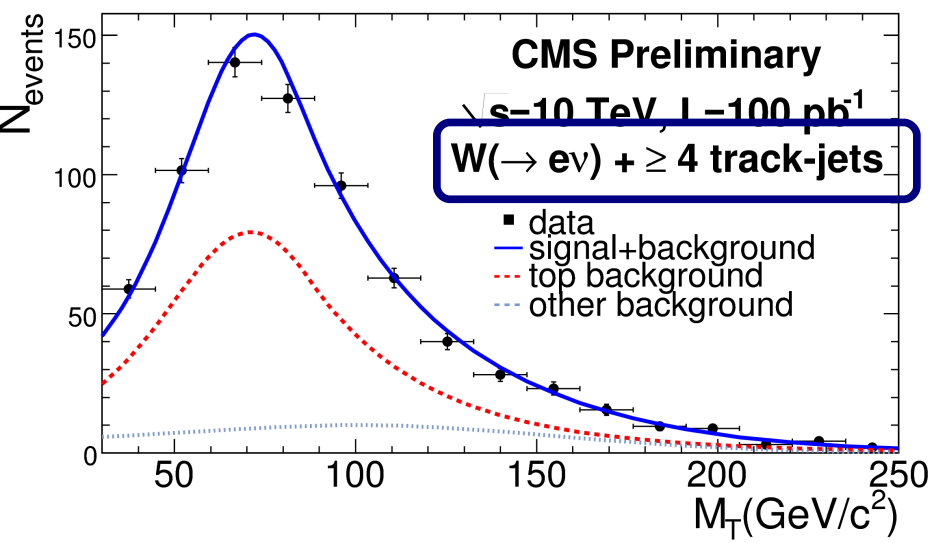


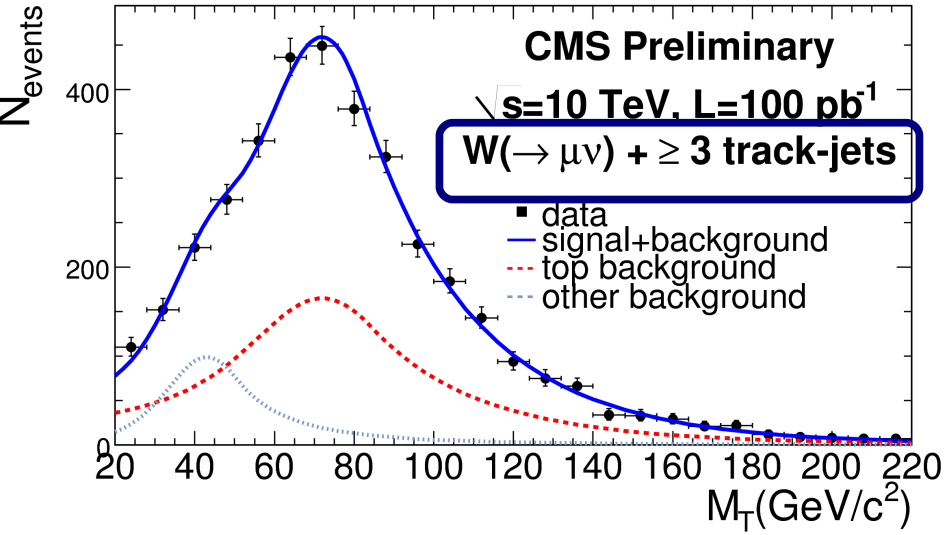
ML fit of W yield

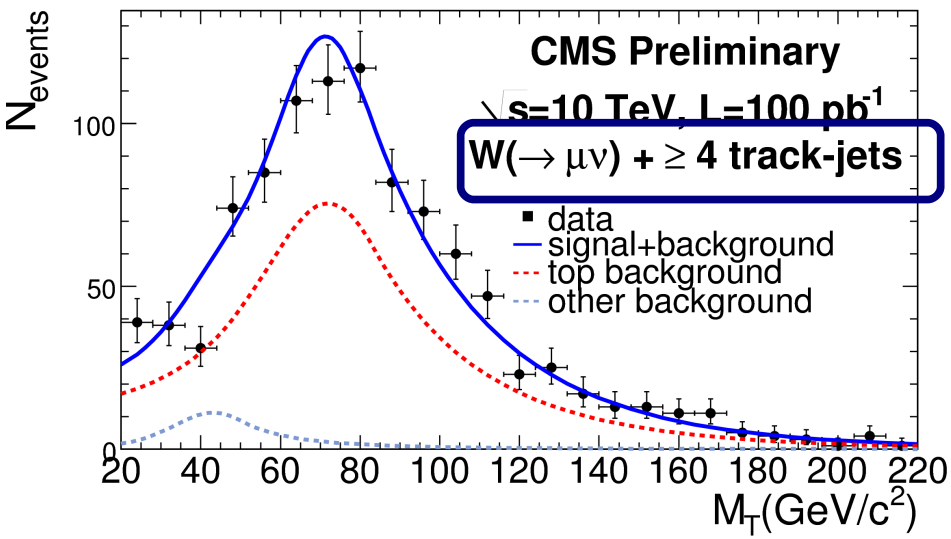
- W is much more difficult than Z
  - MET enters the game
  - High background from real W in ttbar + >= 2 jets
  - Transverse mass discriminates between events with a real W (signal and top) and events with fake-W (QCD)
- To improve the rejection of top in multi jet events:
  - We split the data sample in an heavy flavor enriched (top like) and hf-depleted (signal like) sample
    - We use an average event impact parameter to classify events (b-jets are present in top events)
- We fit simultaneously the two data samples and extract the signal yield from the fit
  - The probability for signal (background) events to be in the hf-depleted sample is taken from the MC and can be validated against a control sample in data

Fit results of W yield



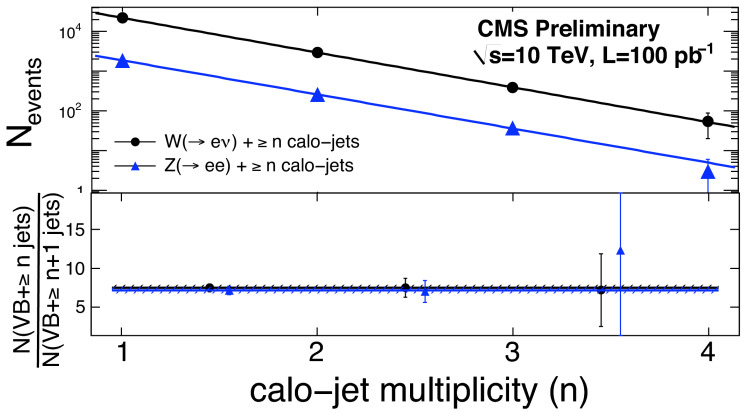


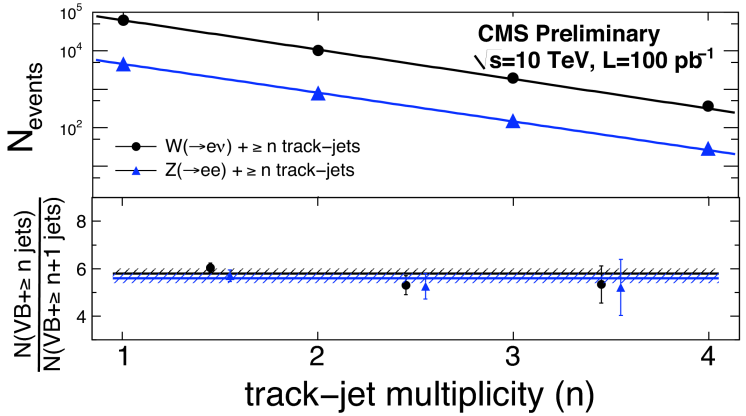


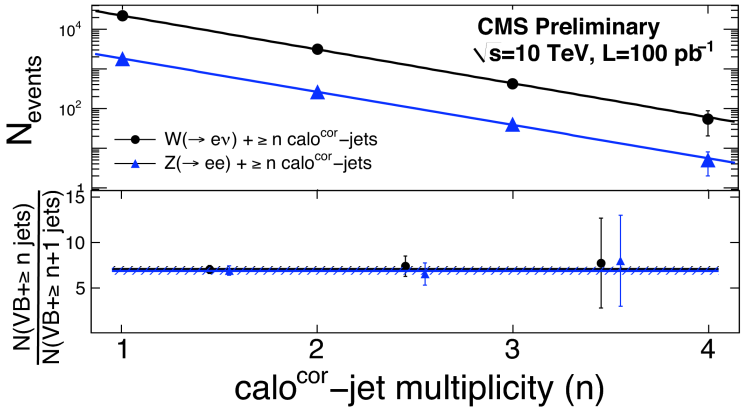


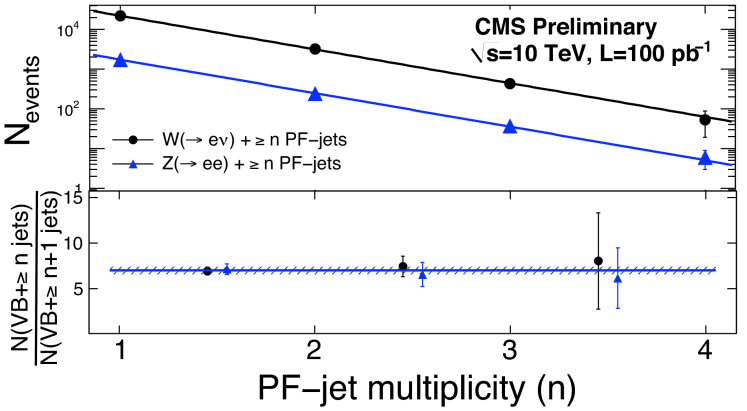
Check of Berends scaling

- The signal efficiencies do not depend significantly on the number of jets
  - We can check the Berends-Giele scaling directly from the extracted yields
- We performed the study for 4 different types of jets

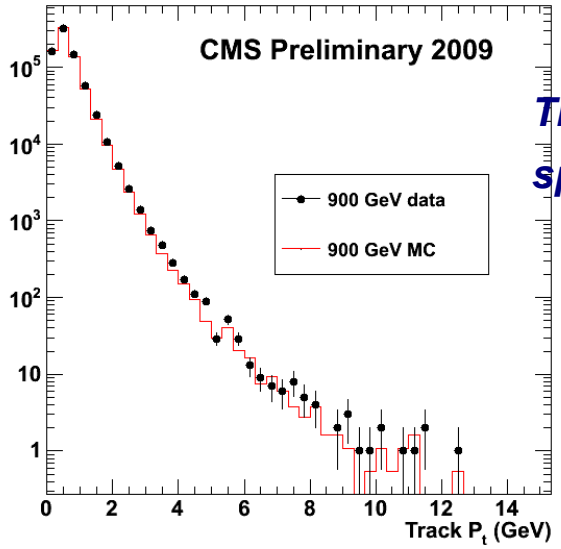




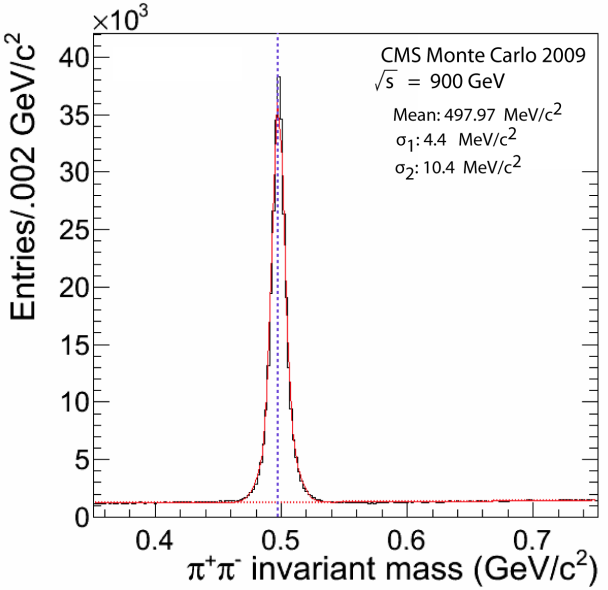




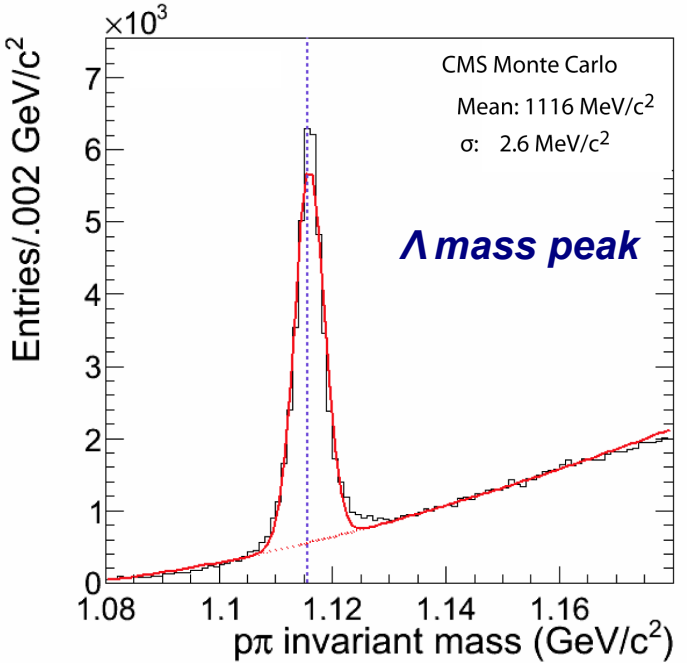
Are we ready? Tracking @ 900 GeV



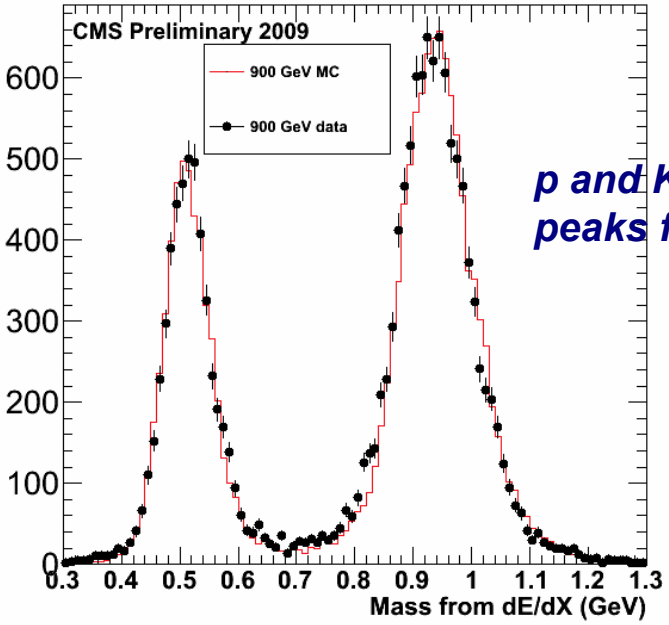
Track p_T spectrum



K_s mass peak



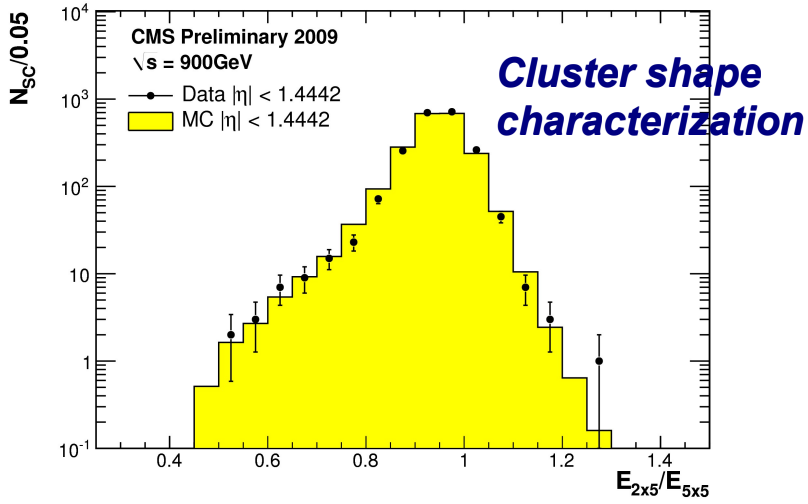
Λ mass peak



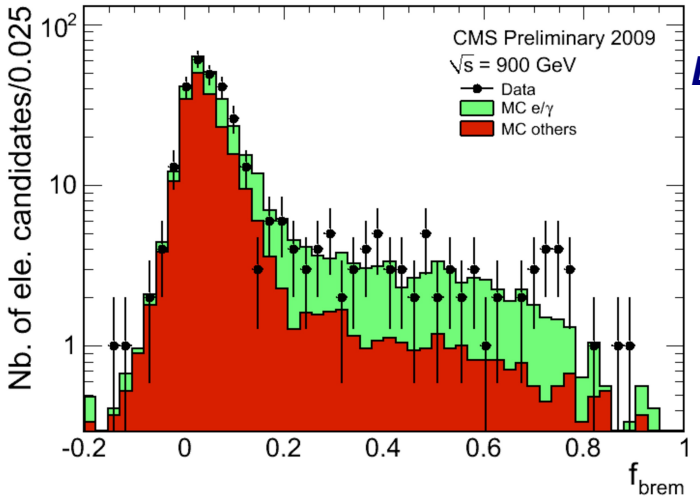
p and K mass peaks from dE/dx

Electrons @ 900 GeV

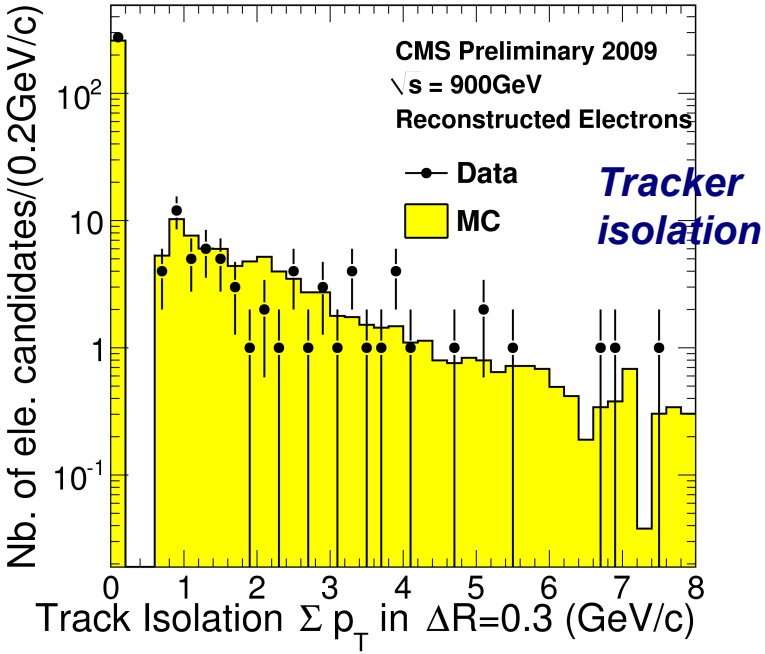
CMS-PAS-EGM-10-001



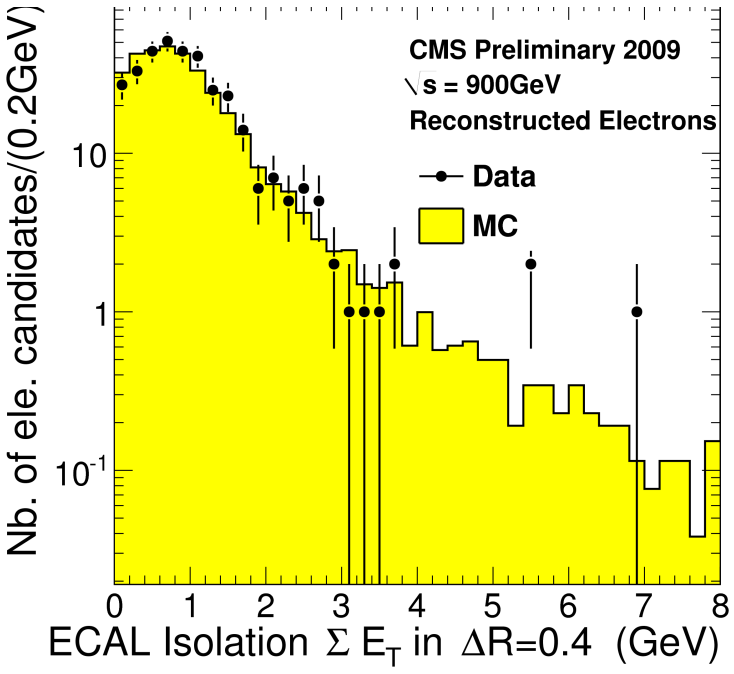
Cluster shape characterization



Bremsstrahlung

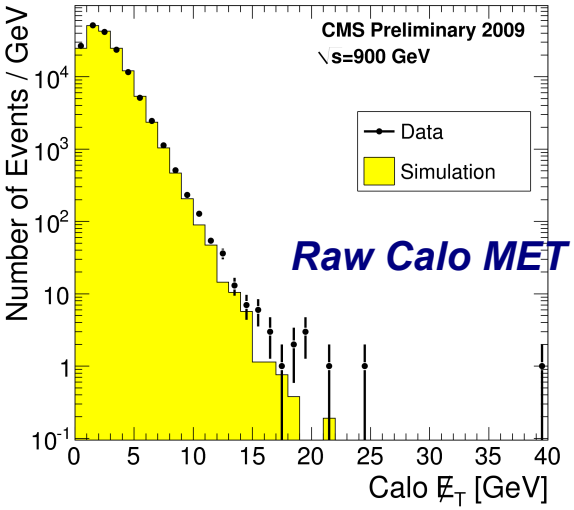


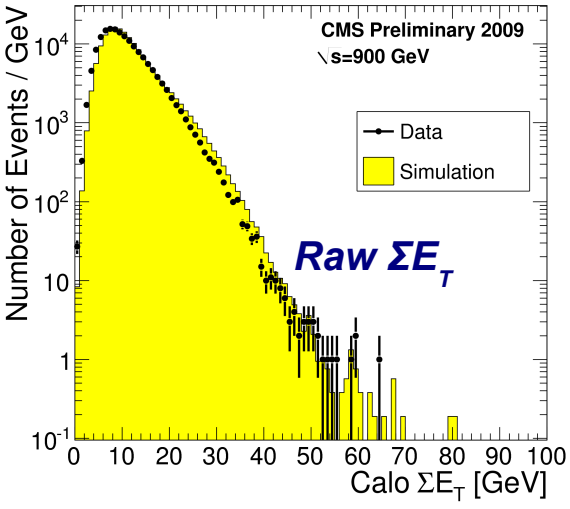
Tracker isolation

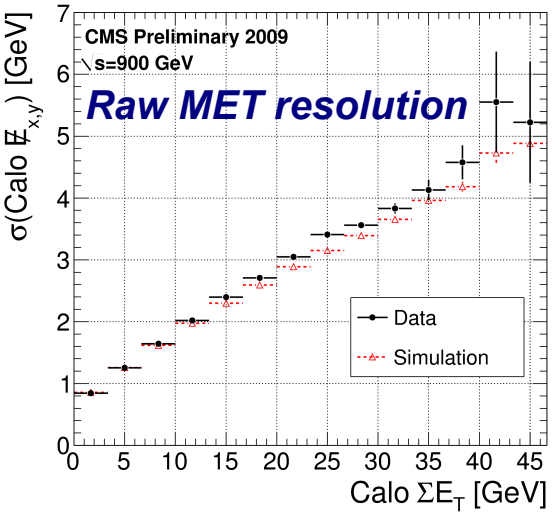


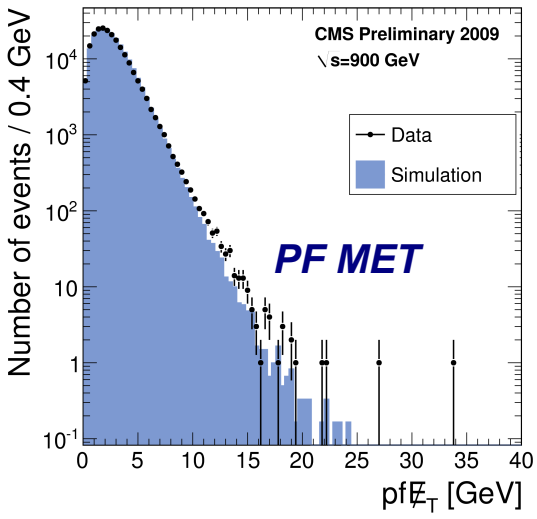
ECAL isolation

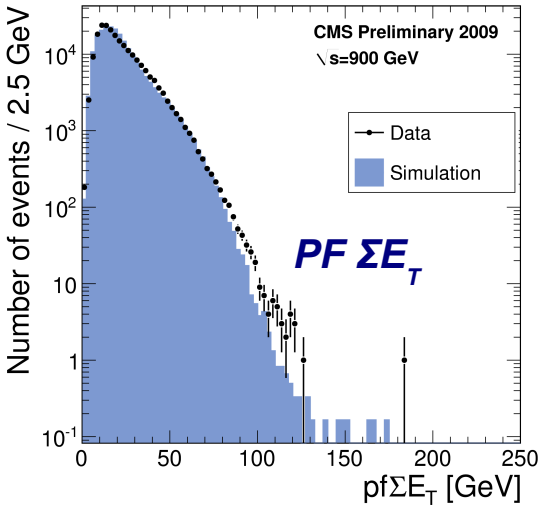
MET @ 900 GeV

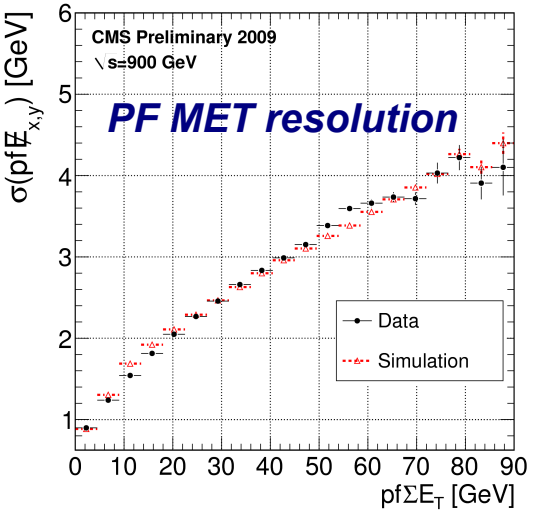






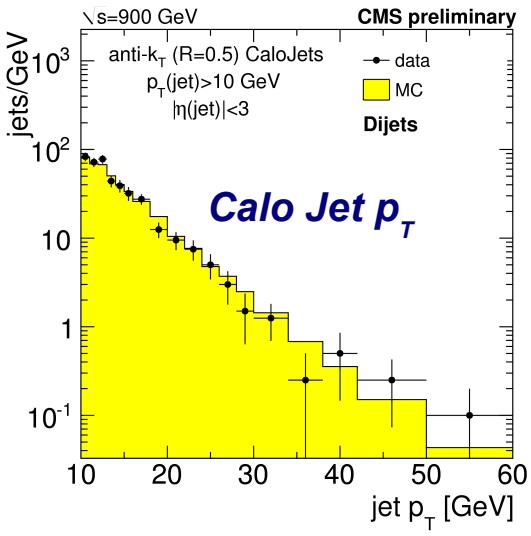


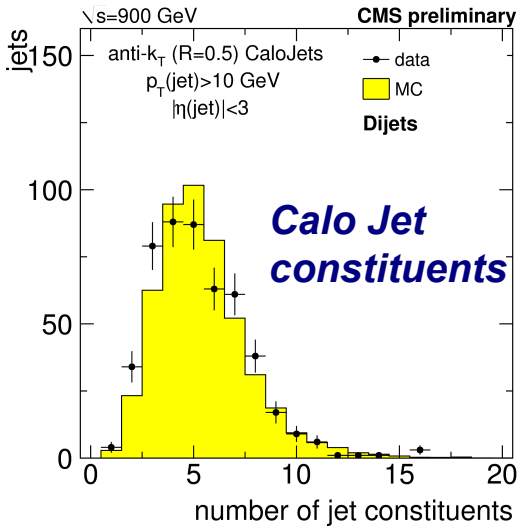


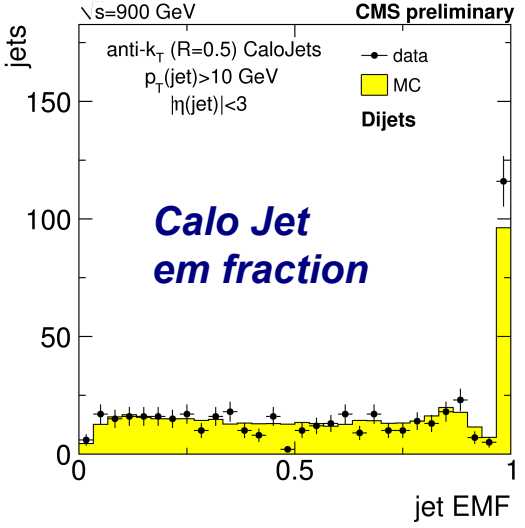


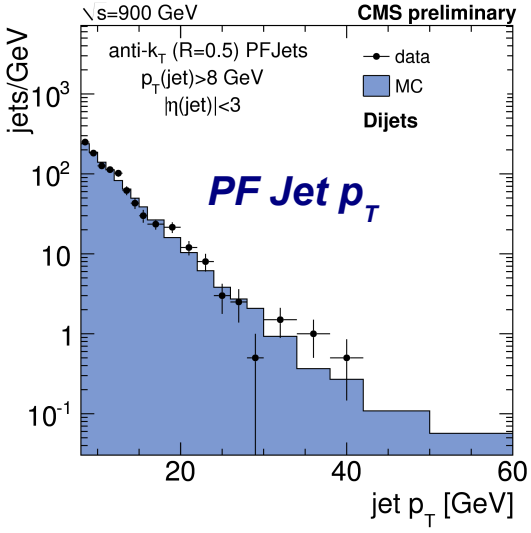
CMS-PAS-JME-10-002

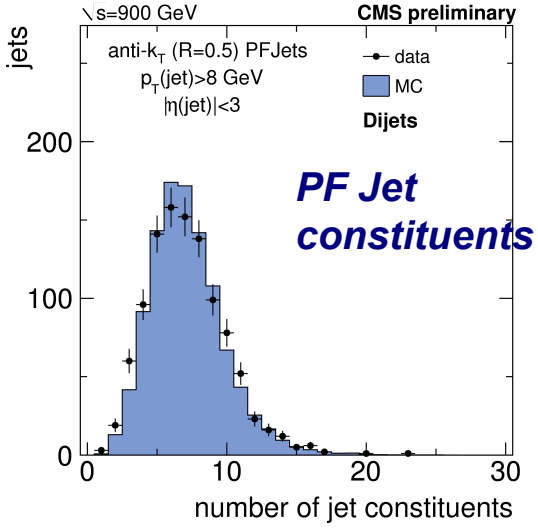
Jets @ 900 GeV

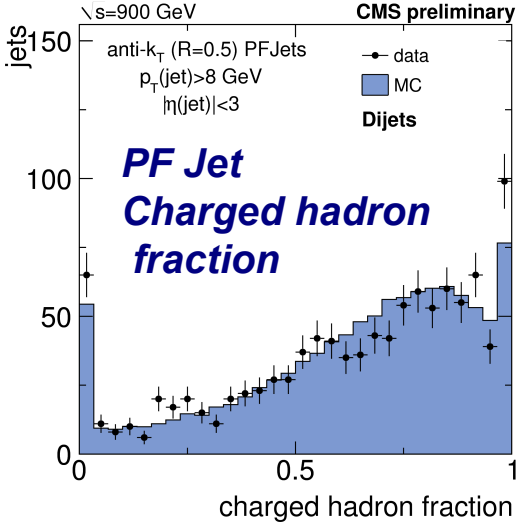












CMS-PAS-JME-10-001

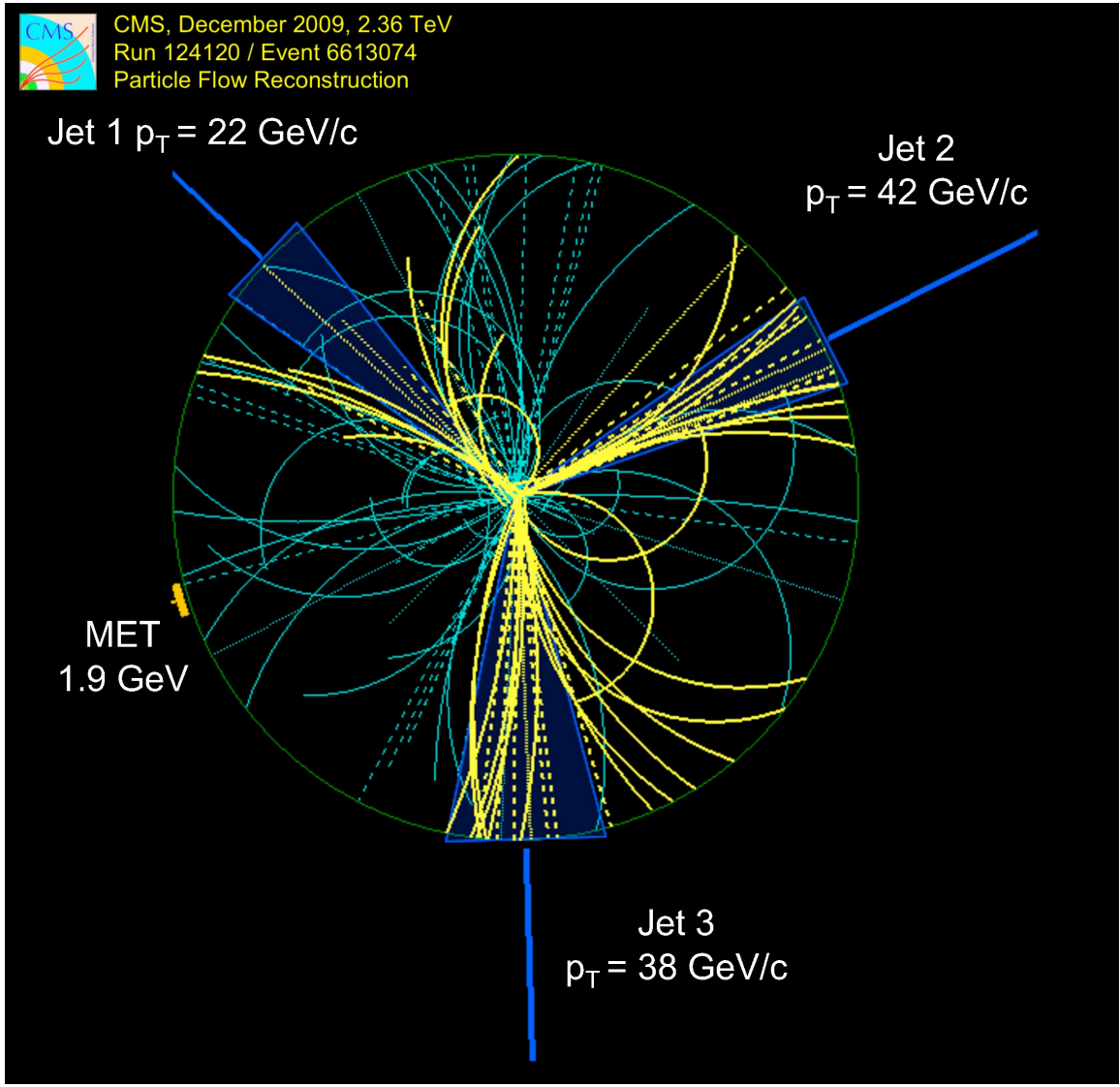
Conclusion

- Our plan for V+jets can be summarized as:
  - Short term: measure ratio of rates, rates
  - Longer term: measure differential distributions, event shapes
- We are using all the most popular particle level MC for our simulations
  - Status of tuning
  - Clear picture of how they compare to Tevatron
  - Need a good description of observables related to isolation and QED fsr
- We have a well established strategy for early analysis:
  - data driven techniques
  - unfolding of detector effects

Backup

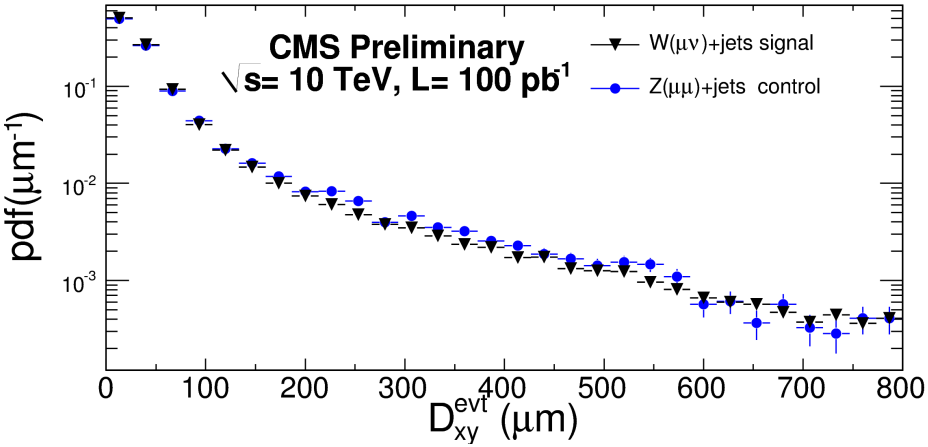
Particle Flow

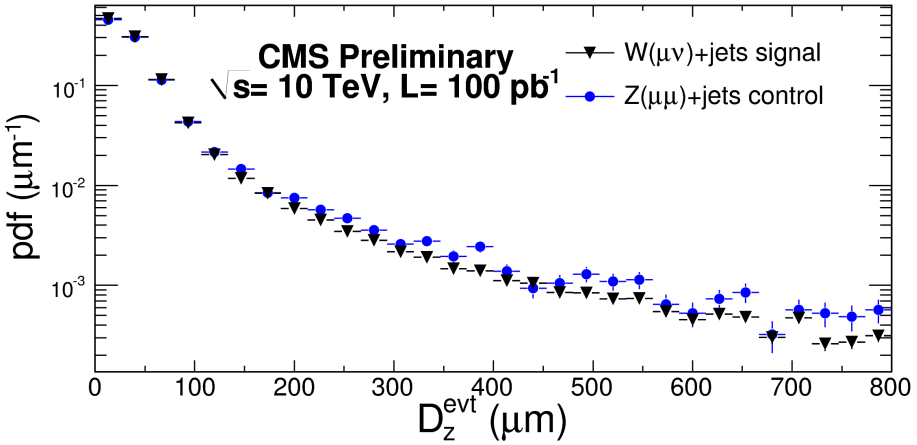
- The particle flow reconstruction gathers information from all the CMS subdetectors and reconstructs all stable particles in the events, classified as:
  - Electrons, muons, photons, charged and neutral hadrons



Control samples

- Control samples for the shapes of transverse and longitudinal event average IP are extracted from Z events





- Control samples for ttbar are obtained inverting the IP cut on the signal lepton and asking for at least 4 jets