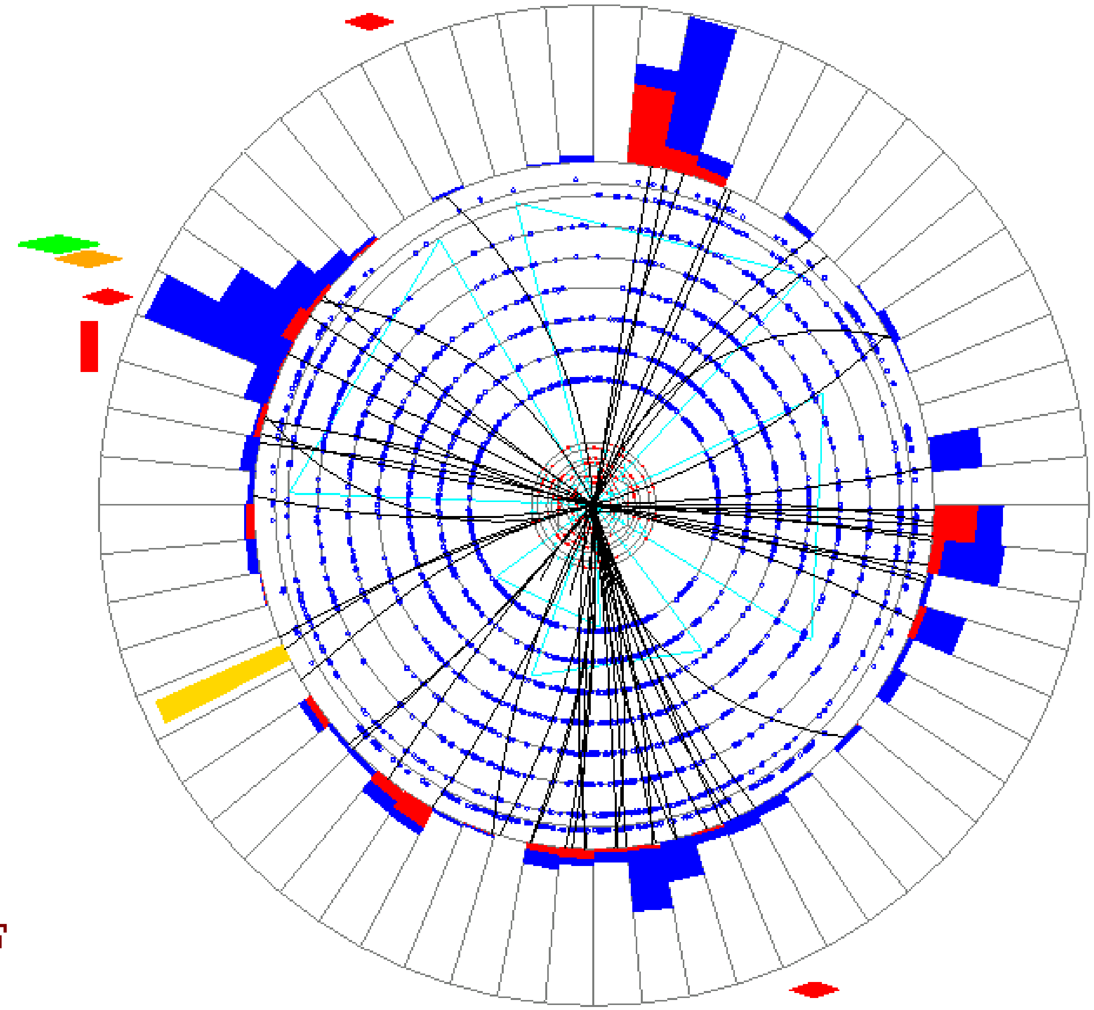
Jets and Missing E_T at the Tevatron

Andy Haas
Columbia University
DØ/ATLAS

US-ATLAS Beyond the SM Workshop

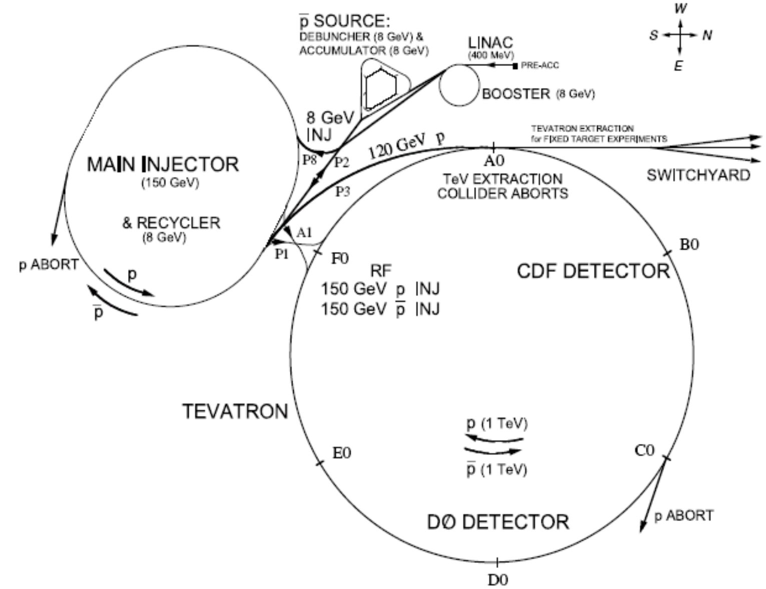
LBL
Jan. 10-13, 2006



Thanks to Daniel Whiteson
and Florencia Canelli from CDF

Fermilab Tevatron Accelerator

- 1 km radius p-pbar super-syncrotron
  - Energy: sqrt(s) = 1.96 TeV
  - Peak inst. luminosity: 1.7e32 $/cm^2/s$
  - Integrated luminosity: 1/fb
  - 36 bunches of p and pbar
  - 396 ns crossing period
  - Two 25 cm RMS interaction regions
- CDF and DØ: multi-purpose detectors
  - ~600 physicists each
  - LAr / Scintillator, hermetic calorimetry
    - Nearly compensating / non-compensating
  - Inner trackers w/ silicon vertex
  - Full simulation: Pythia/Herwig, GEANT3
- Basically an LHC 10% test!
  - Plenty of jets to study, up to pT=~500 GeV
  - Missing ET, up to ~200 GeV
- So, what has been learned about jets and mET?



Outline

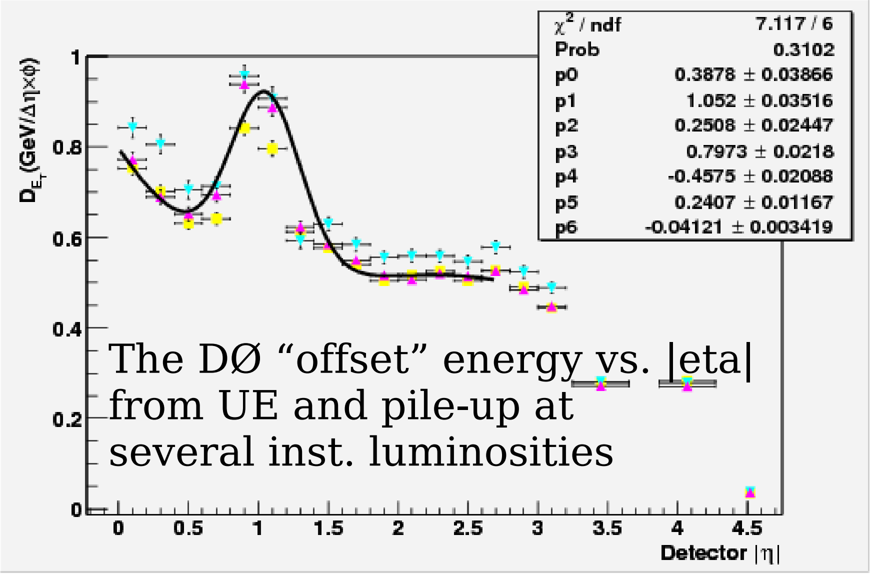
- How well does the MC resemble the data?
- How do you *make* the MC resemble the data?
  - MC tuning / scaling
  - Data quality issues
  - Jet energy scale and resolution
  - Trigger effects
- A look at some data/MC comparisons
  - Perturbative QCD data/MC comparisons
  - LO parton generators: multijets, heavy-flavor
- Some fancier jet corrections
  - Hadronic calorimeter calibration
  - Using tracks to improve jet energy resolution
- How are jets and mET used in physics analyses?
  - Different algorithm and correction choices
  - When are “QCD” backgrounds simulated or taken from data?
  - A look at a few DØ jets + mET analyses
  - MET significance : a per-event mET correction

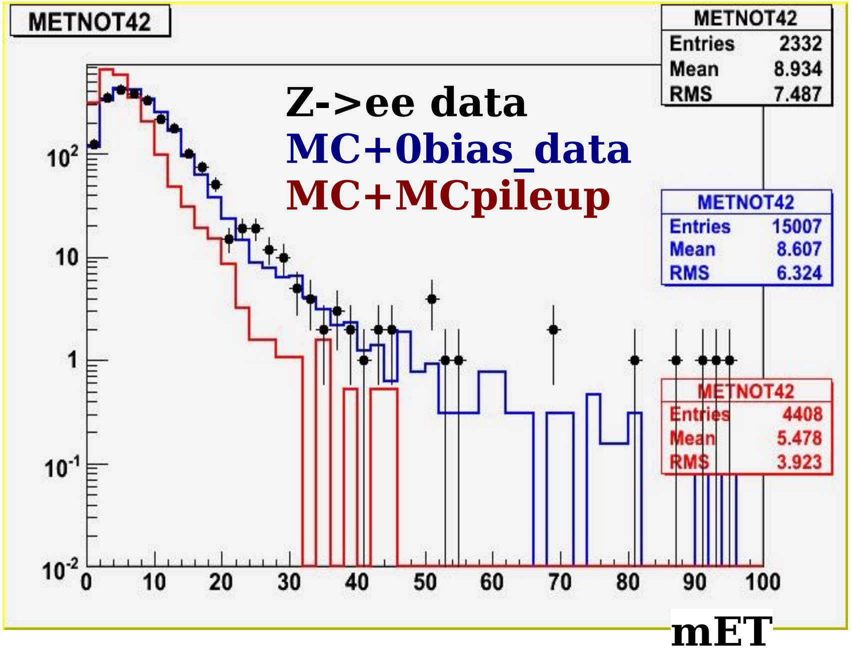
Simulations of Jets / mET Compared to Data

- Every MC simulation is already tuned to some experimental data
  - String-theory just isn't that advanced yet!
  - Even when it is, input will still be needed to determine our particular ground-state :)
- The Tevatron cheated, it had Run I
  - Even when used “out of the box” now,
    PDFs / generators / fragmentation / decay, are partially pre-tuned to Tevatron data
- *How different were the initial simulations from the first Run I data?*
  - *I don't know! I was 12 years old. :)*
- Will initial LHC simulations do a similarly good job of modeling LHC data?
  I think it's hard to compare the situation in the early '90s to 2007...
  - Better theoretical QCD techniques
  - Better simulations and software
- Beyond this basic tuning, both CDF and DØ have moved forward and performed more detailed tuning and corrections
  - Underlying event / pile-up / noise
  - Single-particle response (EM and Had)
  - Scaling / smearing, where necessary (yuck!)
- *The first year of LHC running, at low luminosity, will be a critical time for collecting data needed to tune the simulation!*

Underlying Event and Pile-up

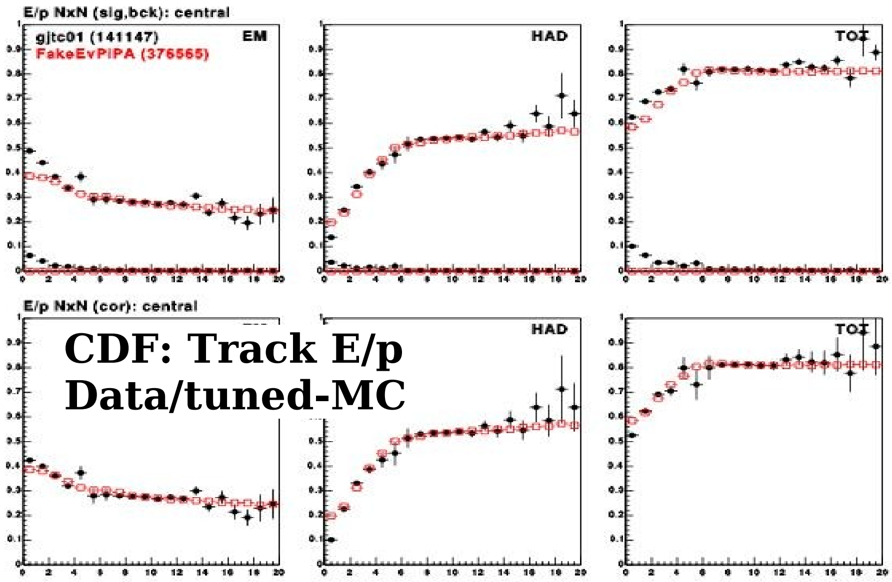
- UE / Pile-up is very important to tune
  - Affects jet energy and mET resolutions (as well as lepton isolation and other effects we won't discuss in this talk...)
- Must adjust MC parameters such that the simulated UE resembles data as closely as possible
  - Tevatron uses Pythia "tune A" - by CDF
- Can do the same for pile-up (minimum-bias events) - CDF
  or:
  Add zero-bias events from data – DØ recently switched to this approach
  - Data is a better idea in principle, but harder, due to data quality and technical issues
- *Very important to match the amount of pile-up in MC to the luminosity profile of the data!*
  - Critical for getting good data/MC agreement in mET resolution, etc.
  - This is easier to do with zero-bias data
- Detector noise is also tuned to data
  - And corrected for in jet energy scale

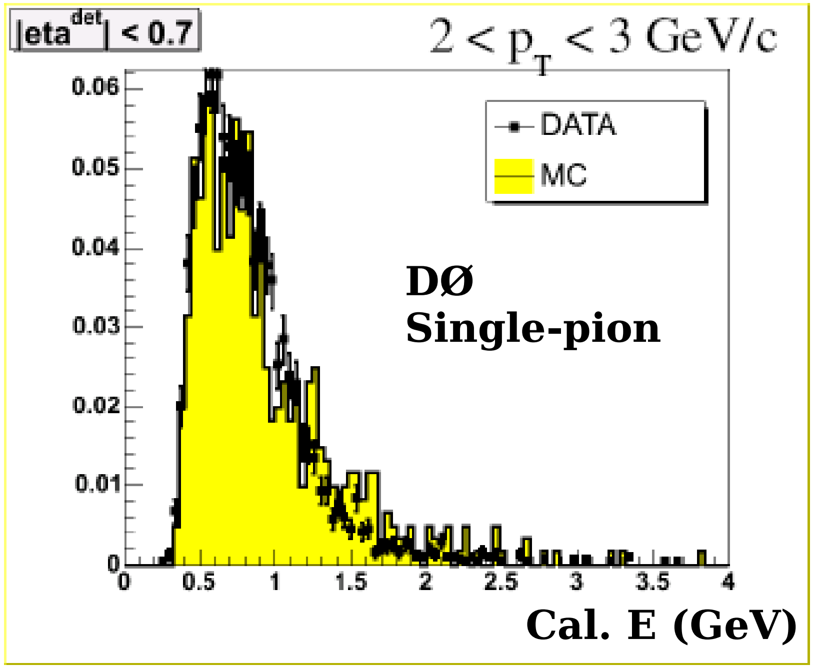




Single-particle Tuning

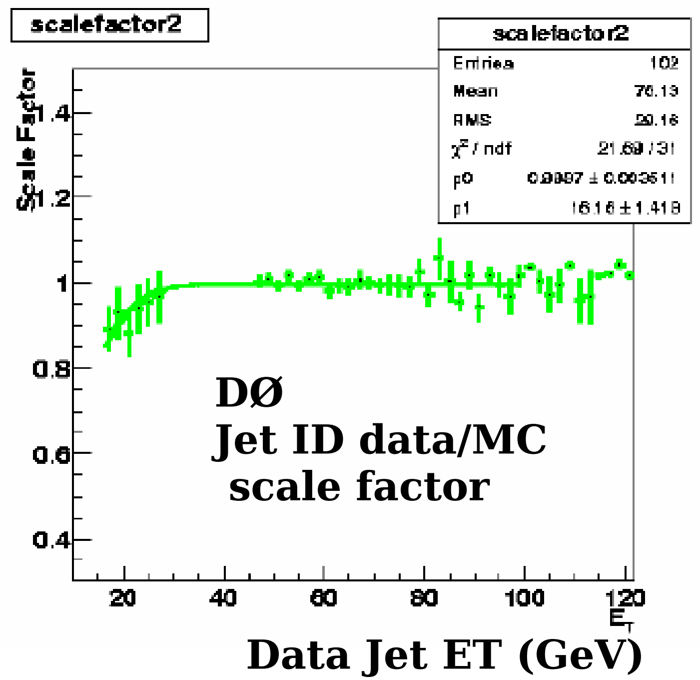
- For the MC to represent the data even better, the response of the calorimeter to single particles (pions) can be measured in data and then corrected for in the MC
  - *This was made more difficult by the lack of a dedicated test-beam for the DØ Run II calorimeter electronics!*
- CDF has fully incorporated these corrections into their MC simulation
  - A dedicated trigger regularly takes single-pion data
  - Critical for their small jet energy scale uncertainty
- DØ uses single-particle response data to correct jet energies on a jet-by-jet basis, based on track information
  - The trigger has not yet been optimized
- Difficult to obtain a clean single-pion data sample with large statistics and understand it
  - Study both gamma+jet events with an isolated track (in data and MC) and also single-pion MC
  - Require track and calorimeter isolation
  - Cuts on EM fraction





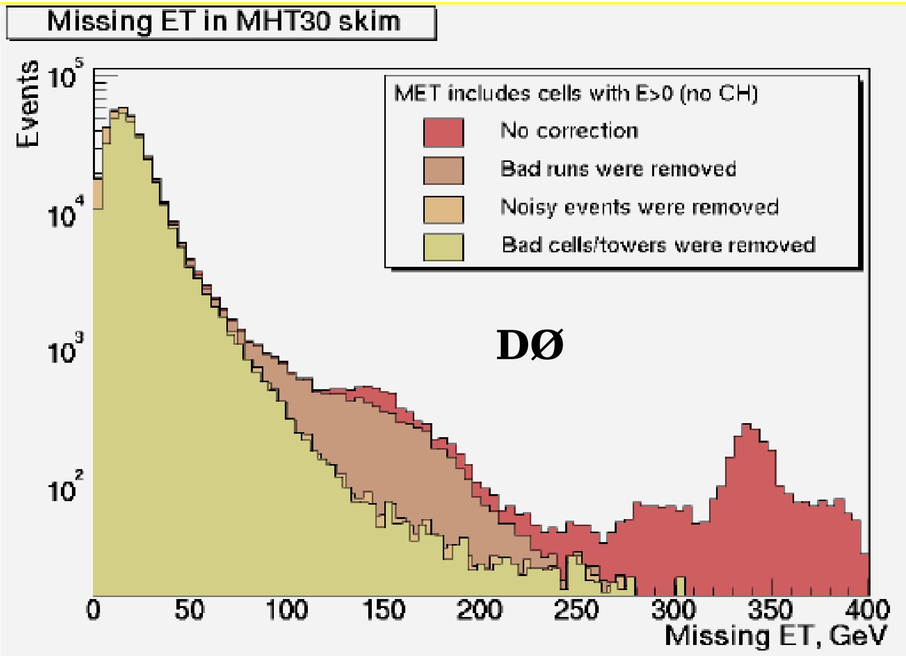
Scaling and Smearing

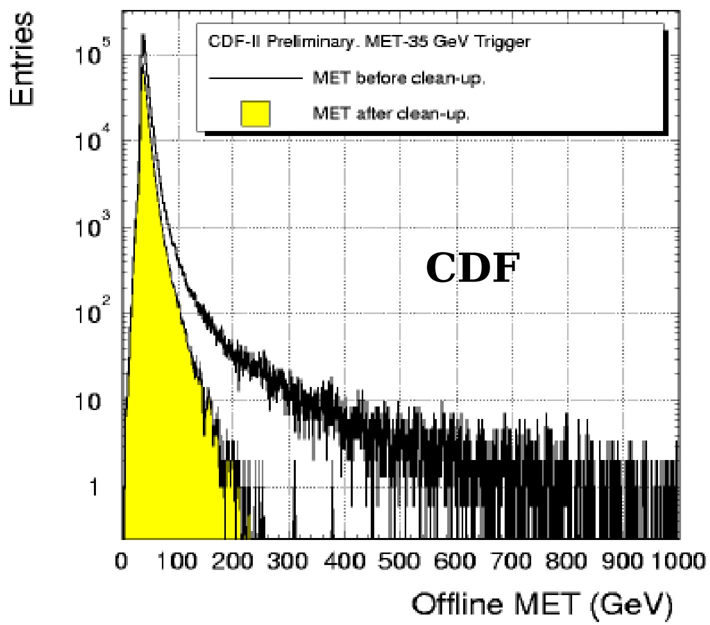
- An experiment never wants to do this... :)
  - But sometimes it has to!
  - MC must be smeared and scaled until the simulations themselves can be tuned properly
- *It's likely that ATLAS will have to make these kinds of rough corrections in the early days of data taking, while MC is being tuned*
- DØ corrects for:
  - Jet-id efficiency at low-pT
  - Jet energy resolution
- mET in good agreement now at DØ after adding real zero-bias for pile-up in MC
- It is partially understood why these are not modeled well in simulations
  - Hadronic response is too high in MC, because it doesn't simulate the shortened charge-collection time (420ns) in Run II electronics correctly
- CDF does not have to make any object-level jet or mET corrections!



Data Quality

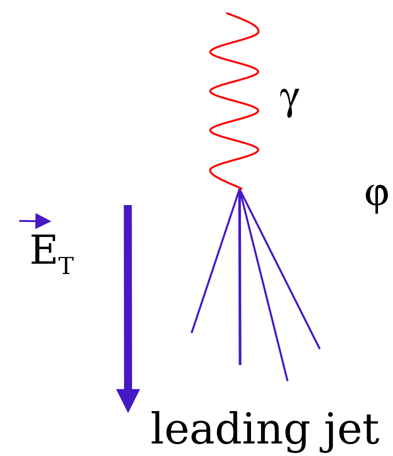
- Missing ET is monitored online and studied offline
  - By run
  - By 1-minute blocks
- Fake mET comes from:
  - Detector problems
    - External RF noise / grounding problems
    - Hardware failures (can be detected online but must still be flagged offline)
    - Replaced / repaired hardware
    - Improper online calibrations
  - "Physics" effects
    - Beam losses
    - Halo muons
    - Cosmics
- Problems are detected offline and corrected for, or events flagged as "bad"
  - The fraction of "physics events" which are flagged bad is also measured in data – and then corrected for in MC

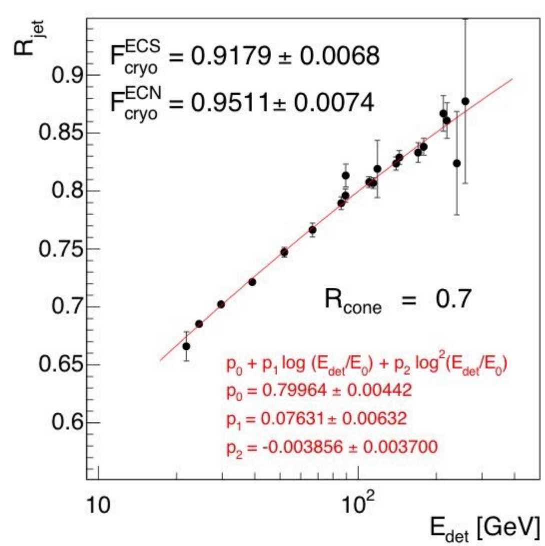




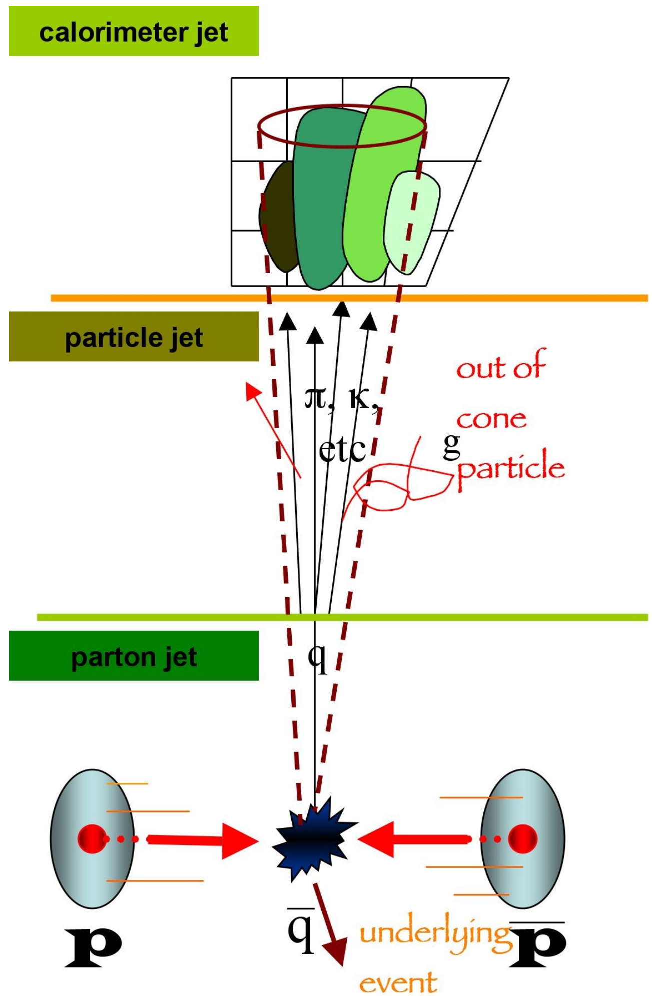
Jet Energy Scale

- Jet energies are corrected for:
  - overlap from the underlying event
    - depends on tower |eta|
  - overlap from other ppbar interactions
    - depends on |eta|, nPV
  - out-of-cone showering
    - depends on ET
  - calorimeter response
    - depends on ET, |eta|
    - mET is corrected for this!
  - muons in the jet, for b-jets
- Use data gamma+jet events, assuming ET balance
- *CDF focuses more on the relative differences between JES in data compared to MC*
- DØ has developed absolute JES for data and MC separately, but will soon also have a relative scale



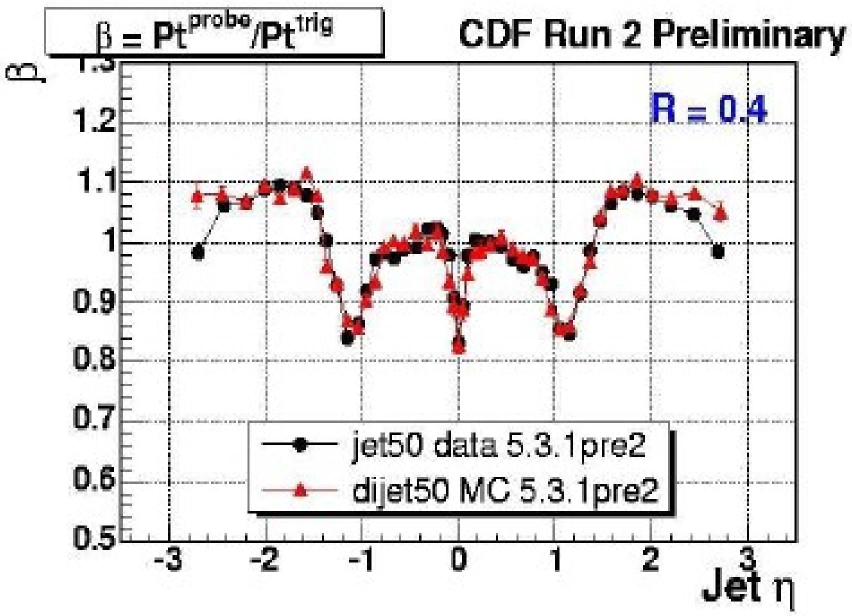


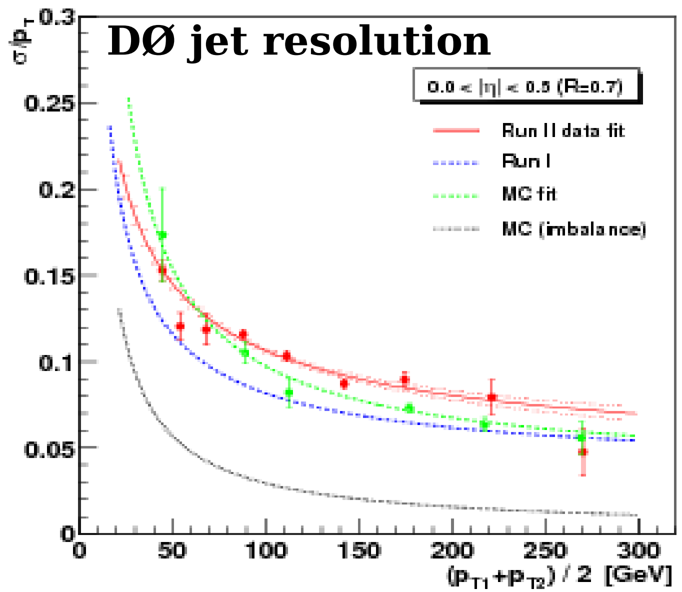
DØ jet response vs. E



Jet Energy Scale and Resolution

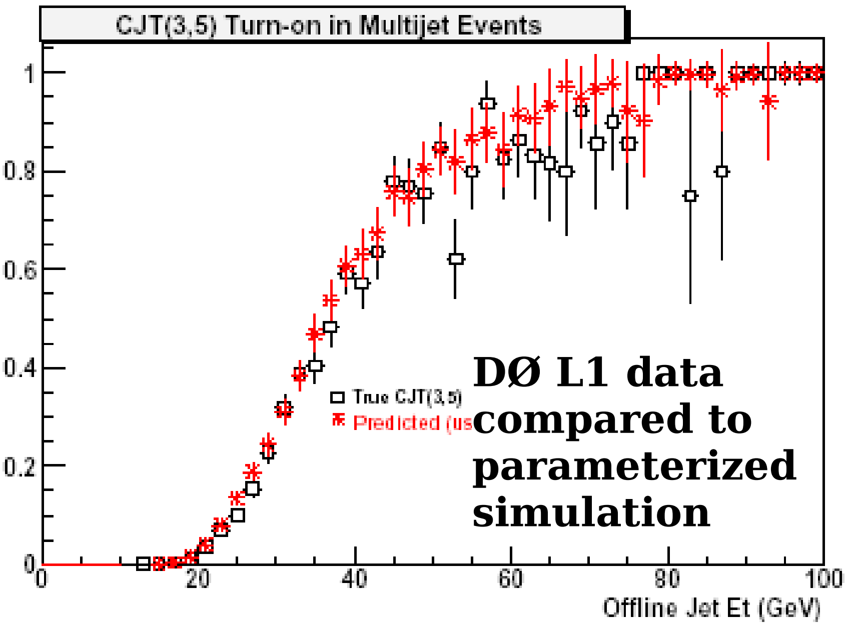
- Systematic uncertainty is estimated by varying many analysis parameters
  - photon ID cuts
  - event topology (dphi)
  - PV selection
  - Seeing where data and MC disagree (CDF)
- *Systematics are also added for the fact that gamma+jet does not agree with Z+jet !*
- Now have made a preliminary JES measurement using ttbar->W(lnu)+W(jets)
  - Can also calibrate the b-jet energy scale
- Resolution is studied in gamma+jet and di-jet events (to extend to higher ET) and compared to MC
  - DØ Run II jet resolution is worse than Run I: more material in front of the calorimeter
- Z->bb would be a great cross-check of jet energy scale and resolution for b-jets
  - Very difficult analysis!
  - But some preliminary results exist

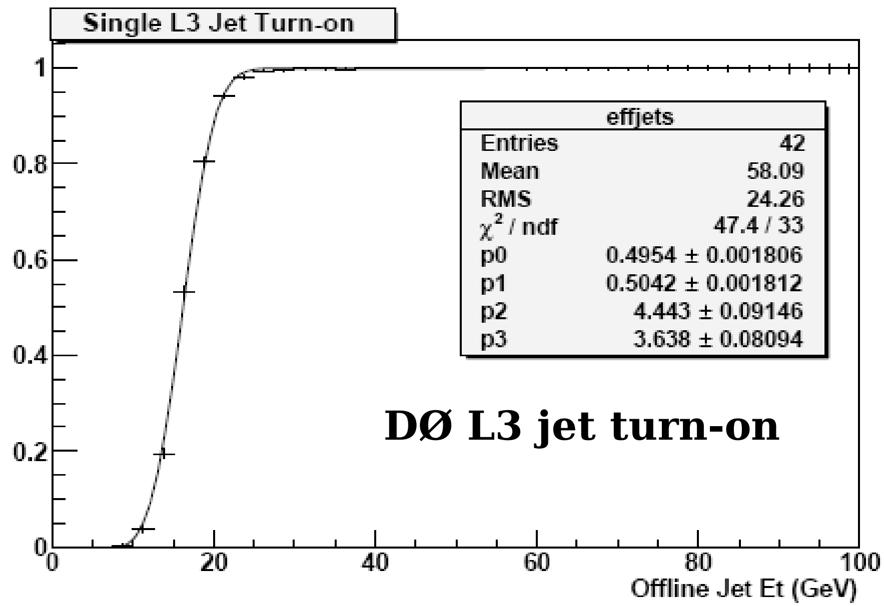




Trigger Simulation

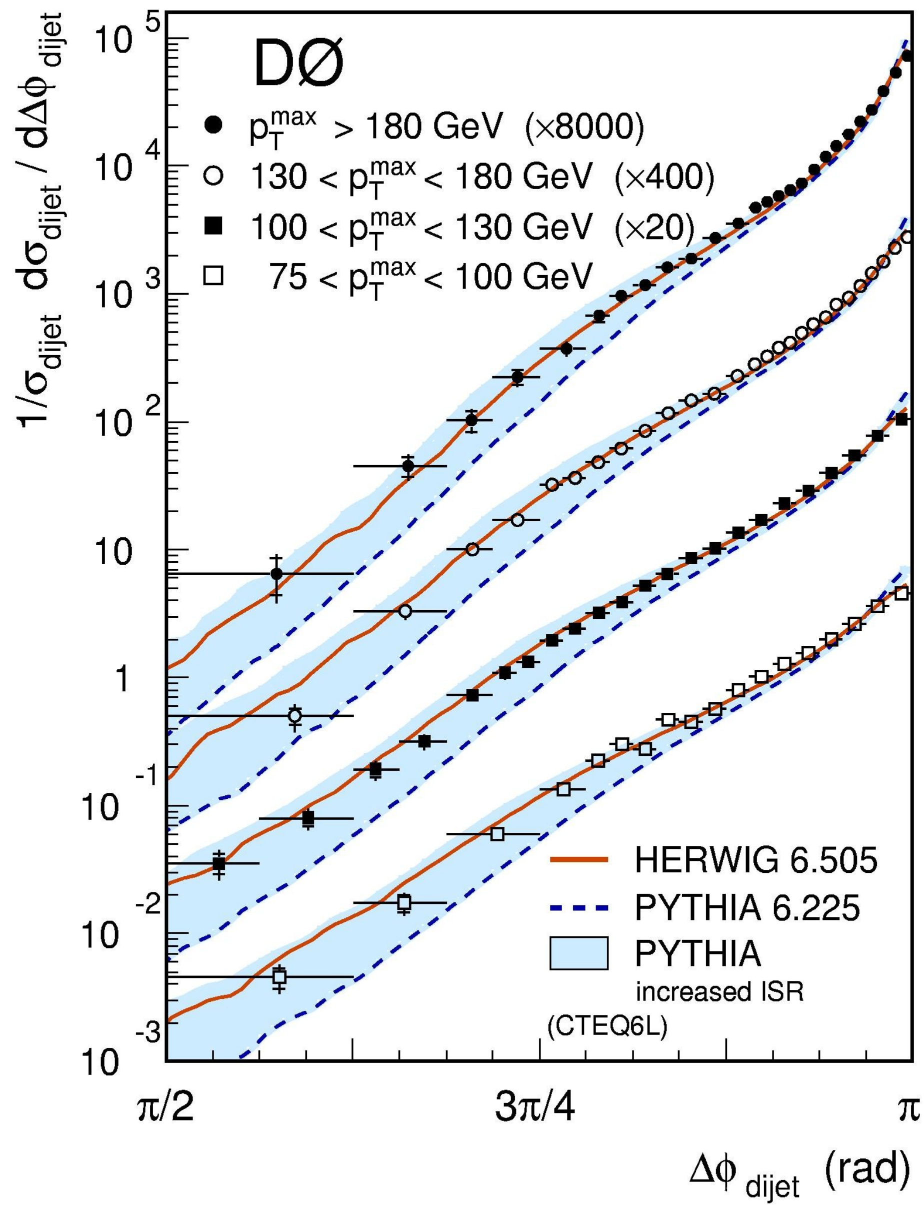
- Needed for studying hypothetical new triggers
- Needed for optimizing trigger lists for best efficiency at a given rate
- Needed for analysis: model the effect on efficiency and kinematics of simulated MC signals and backgrounds
  - *Almost all analyses at CDF and DØ measure "turn-on" curves in data, and apply them to MC jets and mET objects*
  - Sometimes several turn-ons are combined in a parameterization to model a single term
  - More accurate than the trigger simulations
- Care has not been taken to calibrate the trigger-level quantities in the simulation
  - It's a difficult job!
  - The simulations of the algorithms often have bugs or do not reproduce data for some reason
- Trigger conditions are often different for various sub-sets of a data-set
  - A proper simulation would require an extensive database infrastructure
  - Much easier to deal with this "in situ"





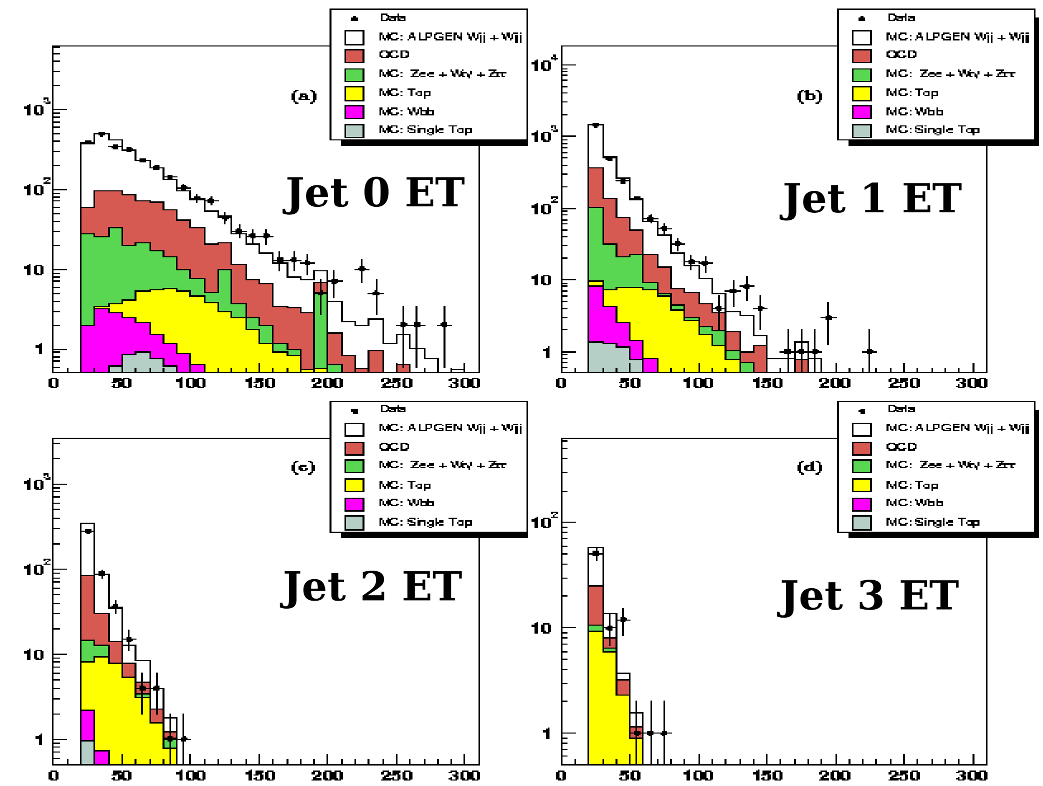
Perturbative QCD Data/MC Comparisons

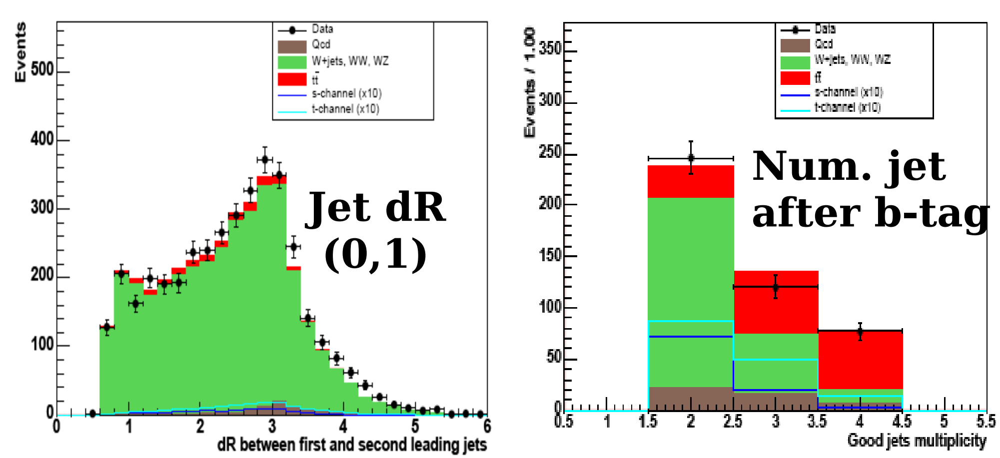
- How well have we done? Does MC now look like our data?
- Look at the dphi between the two leading jets ("angular decorrelation")
  - A measure of the jet "radiation"
  - Also a good test of PDFs
- Use a cone jet with R=0.7
- Apply JES corrections
- Restrict to the central |eta| region, <0.5
- Compare data to:
  - LO pQCD (does not describe low and high dphi)
  - NLO pQCD (does a good job)
  - Pythia and Herwig showering algorithms
    - Herwig is good "out of the box"
    - Pythia needs a little more ISR... but is within tuning uncertainties



Parton Generators: Data/MC

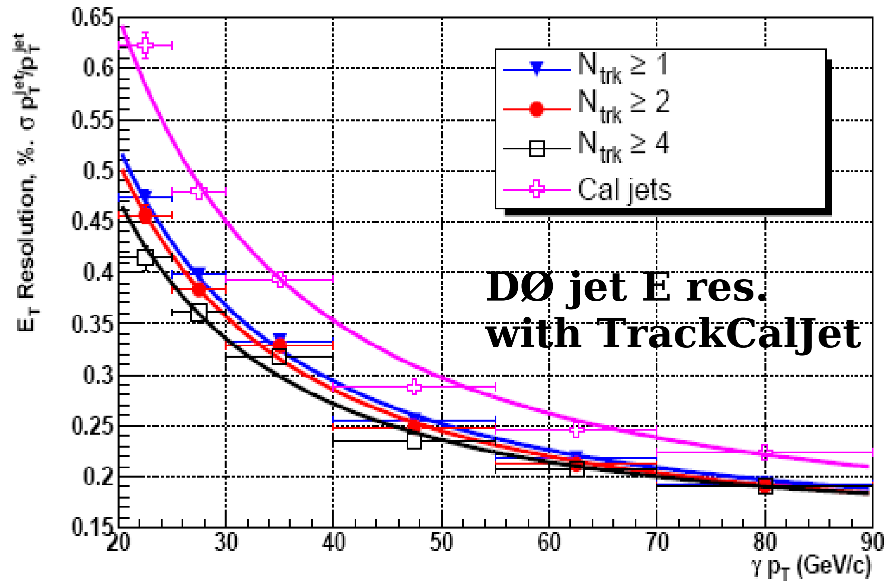
- Jets and mET in analyses often suffer from higher-order processes:
  - Multi-jet events, W/Z+jets
  - Heavy-flavor production / fragmentation
- *Modern generators do a very good job*
  - ALPGEN, etc.
  - Jet parton-shower matching algorithms
  - Pythia fragmentation, gluon splitting
- Can get correct: nJets, pT spectra, eta distributions / correlations, heavy-flavor fractions
- But! Use data where possible
  - "QCD" here is from "matrix method" - inverting lepton id cuts in data
  - A huge amount of MC QCD would be required to model this background
- Always tune / scale processes to data
  - ~+20% are common "k-factors" - just one is enough for W+jets...
  - Heavy-flavor fraction is well-modeled!

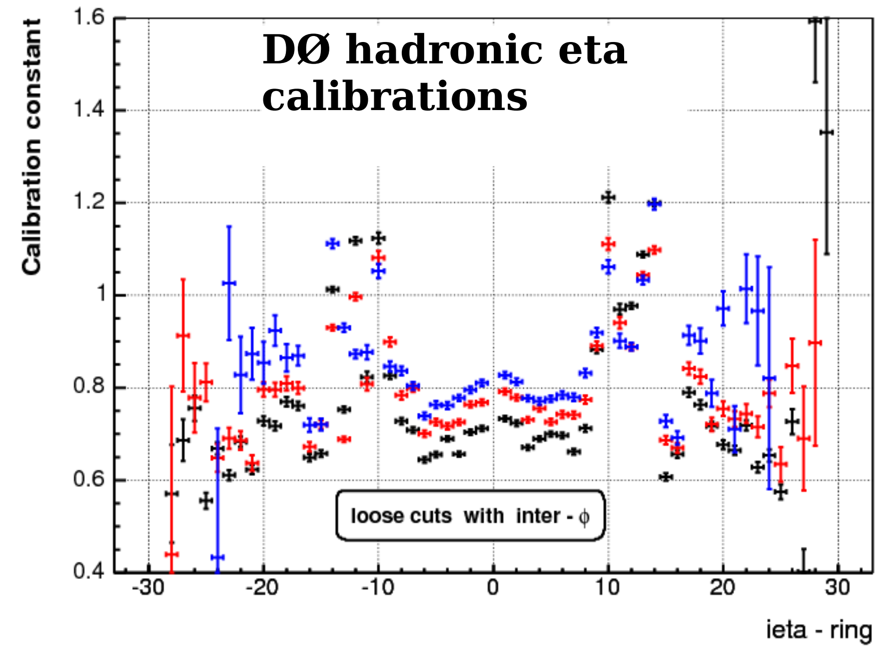




Other Calibrations

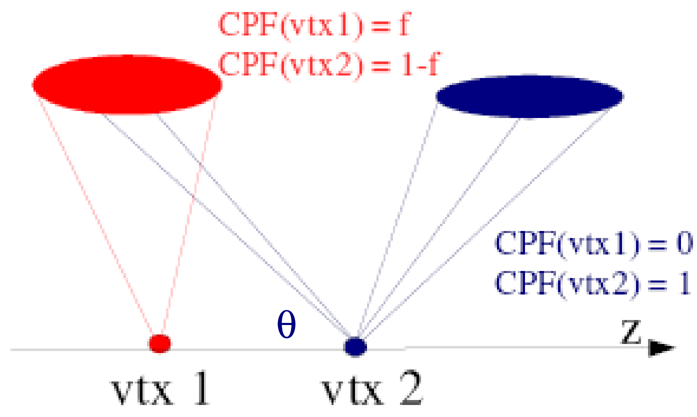
- Beyond the basic cone-jet algorithm, several methods have been tested for improving jet energy resolution
  - *Both experiments have implemented the kT algorithm – no major gains seen… not used in any analysis I know of…*
  - DØ: TrackCalJet – 10-15% improvement
    - Use the single-pion response functions measured in data and MC to improve the jet response – a kind of "energy flow" algorithm
- Both experiments have also devoted significant effort to calibrating the calorimeters in situ
  - For EM layers, use Z->ee
  - For Hadronic layers:
    - Use single particles for inter-phi calibrations (take advantage of azimuthal symmetry of beam / detector)
    - DØ: Use di-jet events to calibrate inter-eta / inter-layer -> 5% improvement in jet energy resolution
  - CDF: H1-type calibrations
    - How big is the improvement?
    - Were there any surprises seen or interesting lessons learned?



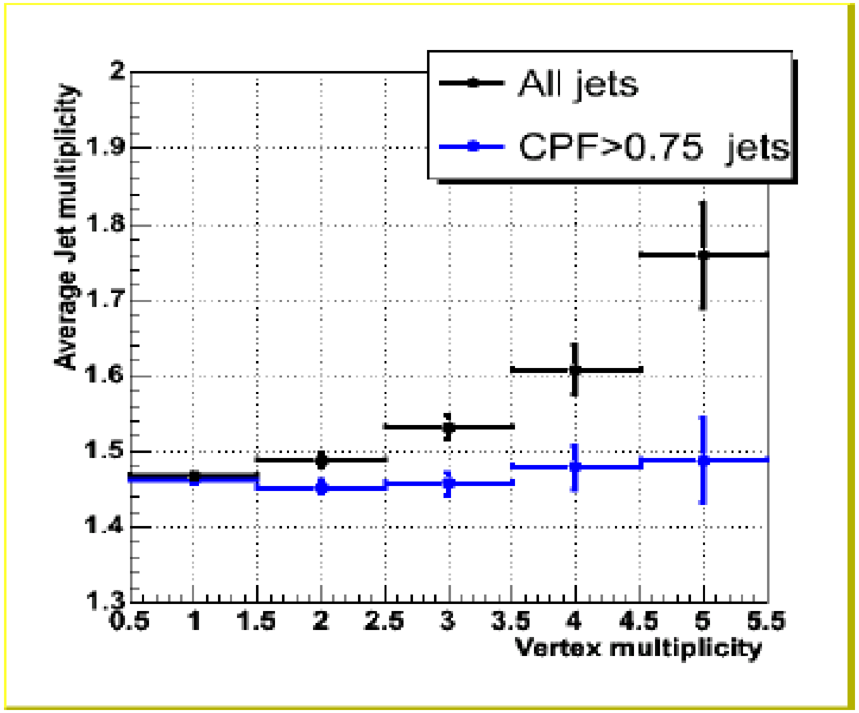


Analysis-specific Algorithm Choices

- Jet cone sizes are sometimes different between analyses, but are mainly historic choices
  - *Real differences in terms of resolution or systematics are minimal*
  - Top / SUSY / Higgs tends to use smaller cone (.5 at DØ, .4 at CDF), because of busier event
  - QCD uses 0.7 cone
- Require jets to come from the same primary vertex (charged-particle-fraction cut) ?
  - Also has an impact on the mET (E sinθ)
  - Seems sensible, but can induce larger systematics due to min-bias modeling
- Apply track-based jet energy corrections?
  - This has better resolution, but possibly will have larger systematic uncertainty on scale
- There are also differences in the mET algorithms / corrections used
  - Do you reject events with possibly mismeasured leptons (in cracks)?
  - Reject mET aligned with jets and/or leptons?

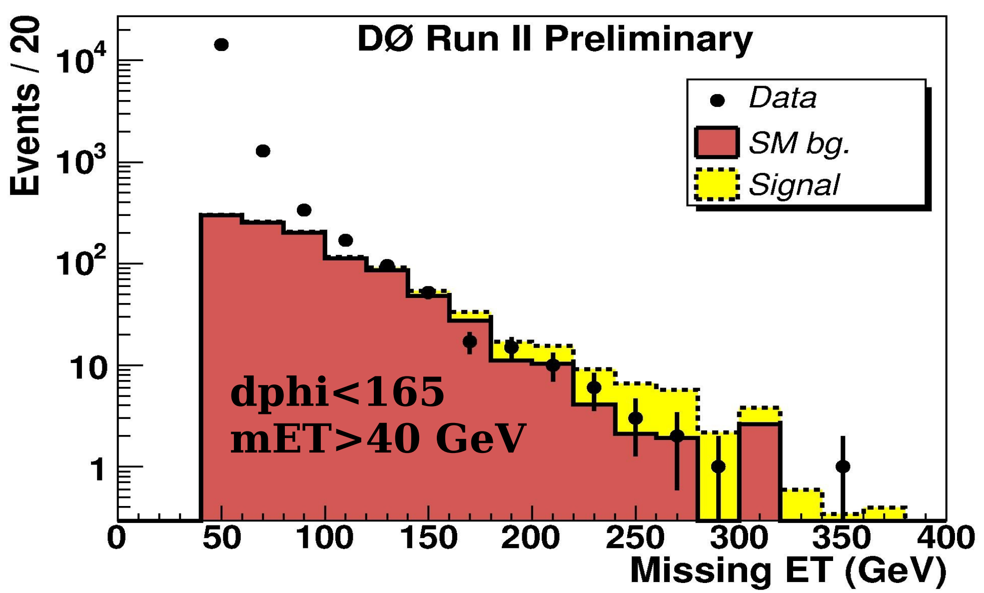


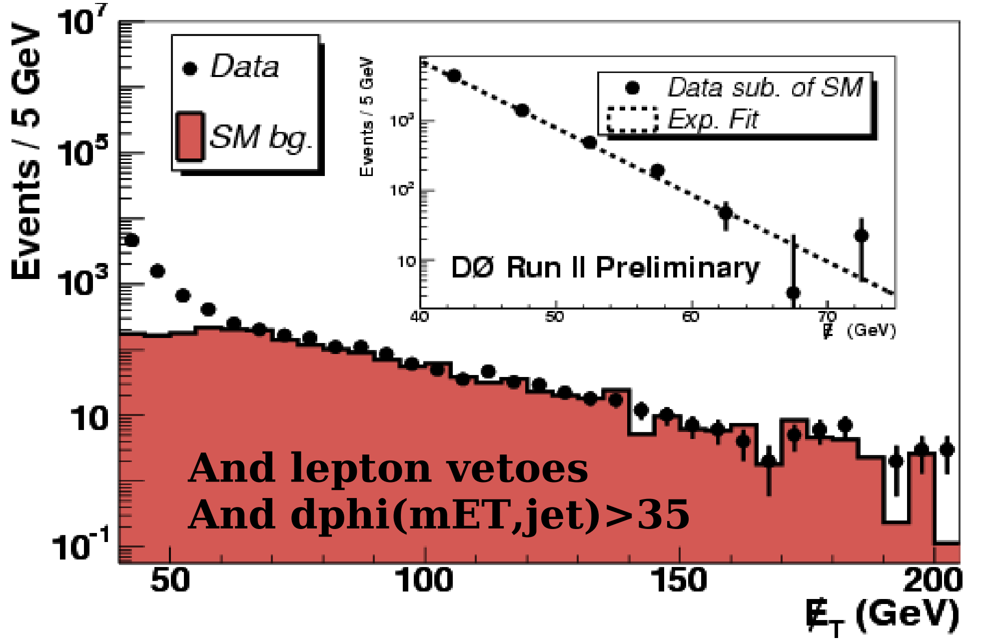
Cutting on CPF removes jets from pile-up



An Example Jets+mET Analysis

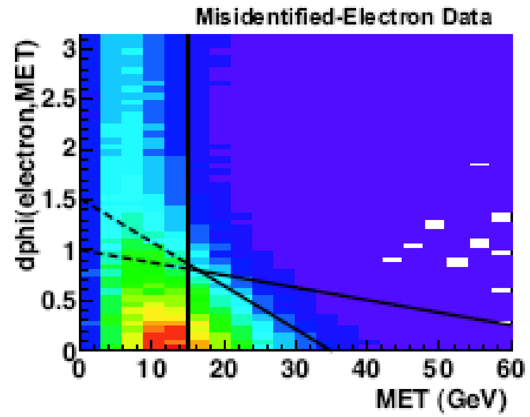
- The search for squarks and gluinos looks for excessive mET in events with at least two jets
  - Fake mET is suppressed in di-jet events by requiring them to not be back-to-back, dphi<165
- *Note that "QCD" is not simulated: It's measured in data and extrapolated under the signal region, where it's determined to be negligible*
- The shape of the "SM" W+jets, and Z(nunu)+jets, etc. reproduces the data
- Extra selections are added for the sbottom search, which later adds b-tagging
  - Lepton vetoes and requiring mET isolation from jets and leptons reduce the QCD dramatically, and also the W+jets
- Mono-jet search is challenging!
  - Very sensitive to JES, at high jet ET
  - Sensitive to cosmic and beam backgrounds
  - No approved plots yet, sorry!

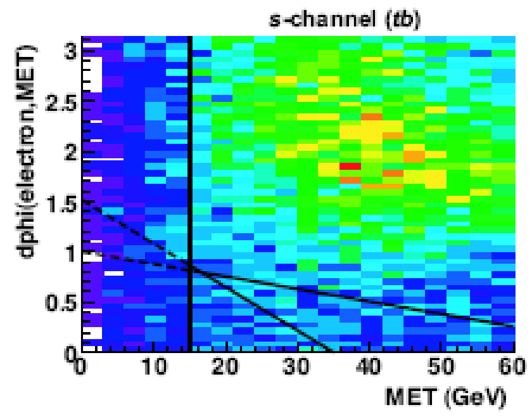


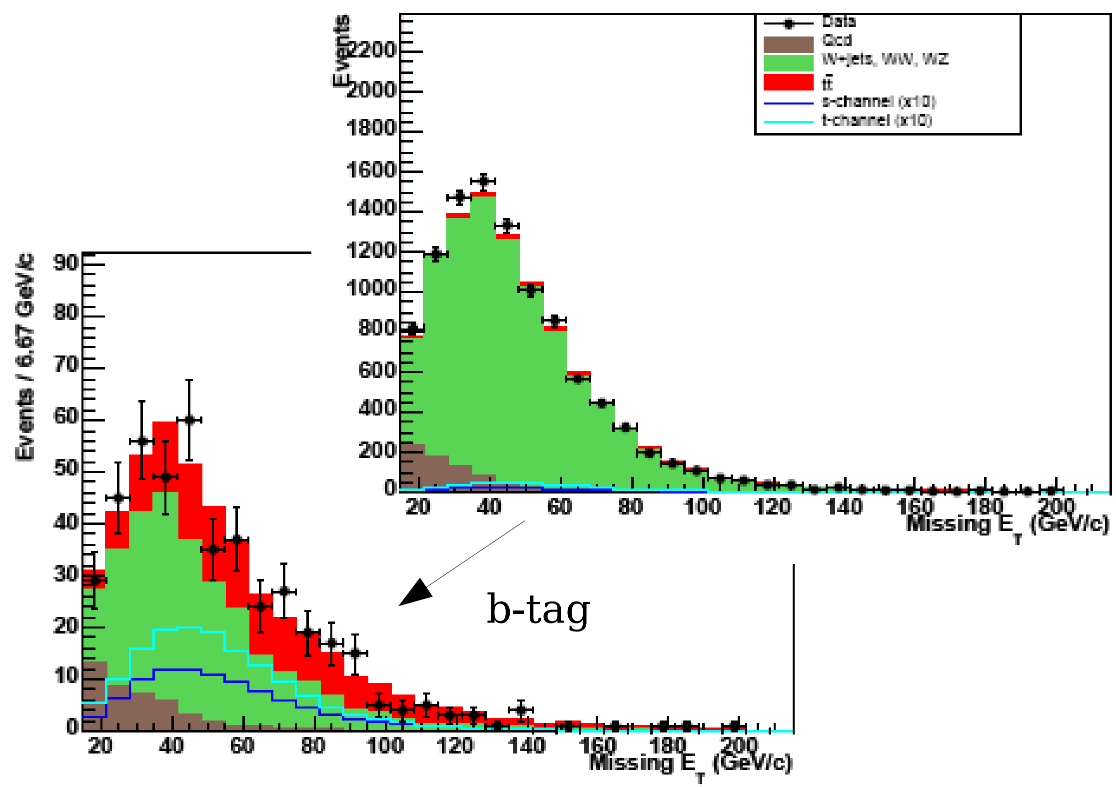


Search for Single-Top in Leptons+Jets+mET

- The search for single-top production looks at the kinematics of W(lnu)+b+jet events and compares to MC simulations
- So called “triangle cuts” in planes of:
  - mET vs. dphi(mET,jet[0,1])
  - mET vs. dphi(mET,lepton)
- *Triangle cuts allow for a more efficient selection of signal, while rejecting fake mET background*
- Good agreement in mET both before and after b-tagging!
- Note that again, “QCD” is taken from data, but W/Z+jets and ttbar is from ALPGEN+Pythia MC simulation

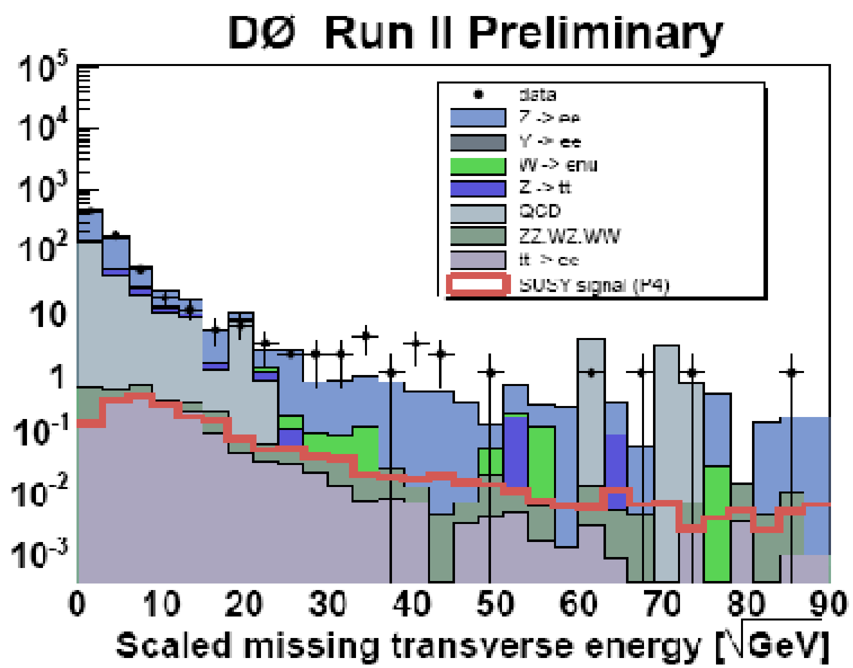






H->WW: mET Significance

- The search for SM Higgs->WW looks for an invariant mass bump in the WW spectrum
  - Very important to reduce non-WW background such as Z(->ee)+ jets
- Purify the mET measurement by:
  - dphi cuts w.r.t. jets and leptons
  - remove events with jets going into poorly instrumented regions (calorimeter cracks)
- Also use “mET significance”, a continuous quantity based on the mET resolution for each event, based on the objects in the event
  - Sum over degradation of the mET from jets, leptons, and unclustered energy
- *A mET significance cut has better S/sqrt(B) than a simple mET cut*
- Also useful for SUSY searches!



$$\mathrm{Sig}(\not{E}_T) = \frac{\not{E}_T}{\sqrt{\sum_{\mathrm{jets}} \sigma^2_{E_T^j \| \not{E}_T}}}$$

Conclusions

- Jets and mET are studied at the Tevatron in a somewhat LHC-like environment
  - *But, relatively low radiation / luminosity*
  - *Little attempt made at reconstruction of jets past |eta|>2.5*
- MC does a good job of reproducing Tevatron data...
  - But, only after careful tuning, corrections, and normalizations
  - And it's still preferable to use data for "QCD" backgrounds where possible
- *Building a calorimeter with excellent data quality and low noise will make the data agree with the MC more easily, as well as simplify all jet and mET calibrations!*
- We've learned how to measure the jet energy scale using gamma/Z+jet and have started testing fancier methods like Z->bb and ttbar
- We've applied several advanced calibration techniques: di-jet eta intercalibration, H1, and track-based methods for pile-up rejection and calorimeter response corrections
- Analyses use several different jet and mET corrections, depending on whether they are needed and don't induce unwanted systematic uncertainty
  - For mET: require acoplanar jets, reject "crack" events, require mET "isolation" from jets and leptons, use "triangle cuts", cut on mET significance

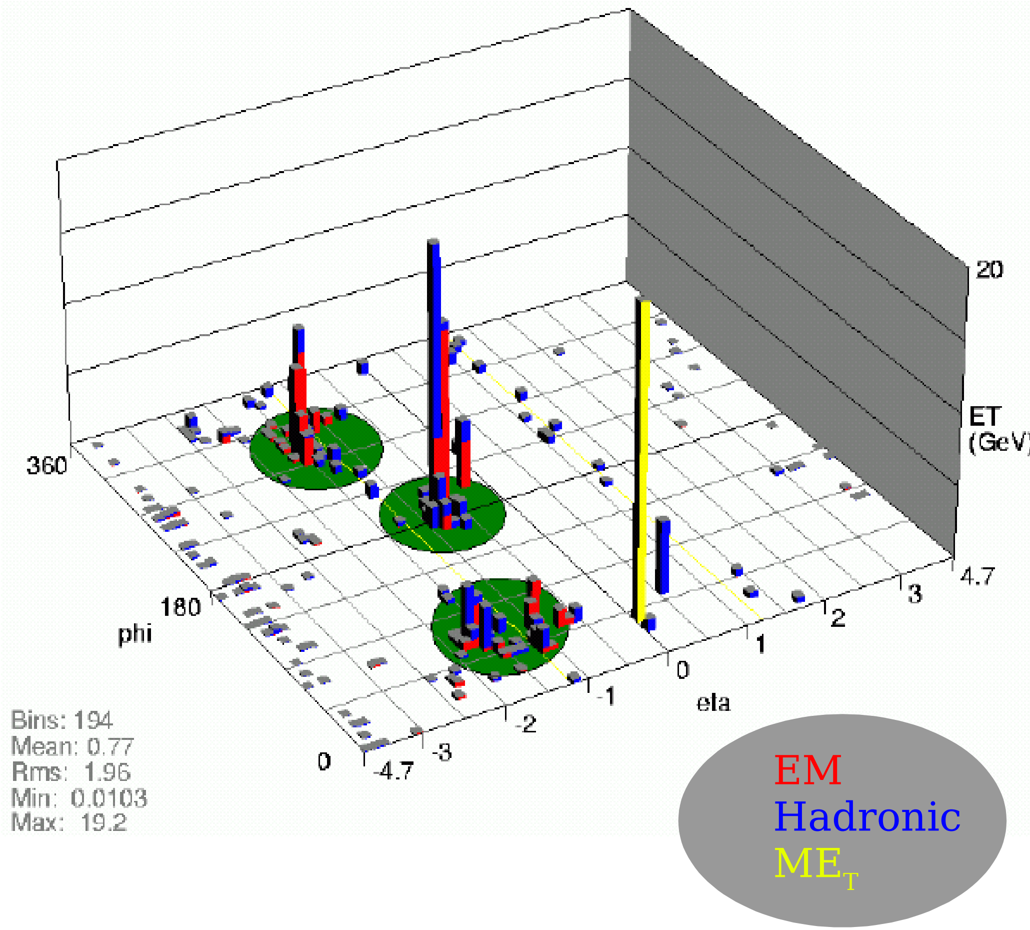
Finally, another interesting way the Tevatron could help... :)

- **Still hoping to establish a Neutralino signal at the Tevatron to provide a standard candle for E_T calibration at the LHC**

Backup slides...

Jet Reconstruction

- The E_T in (0.1x0.1)(η-ϕ) towers is summed, and seeds are found
- Energy is clustered in cones of $\Delta R = \mathrm{sqrt}(\Delta\eta^2 + \Delta\phi^2) = 0.5$ around the seeds
- The midpoints (in η-ϕ) between stable cones are also used as seeds
- The unique cones with $E_T > 8$ GeV are merged or split, depending on whether they share more or less than 50% of a jet's E_T
- Jets' E_T are calibrated, using the jet energy scale (JES)
  - accounts for out-of-cone showering and the underlying event, on average



$$E_{jet}^{corrected} = \frac{E_{jet}^{colorimeter} - E_{offset}}{R_{jet} \cdot R_{cone}}$$

A Quadruple b-tagged Event