Massive High P_T Jets: Updates from CDF

Outline

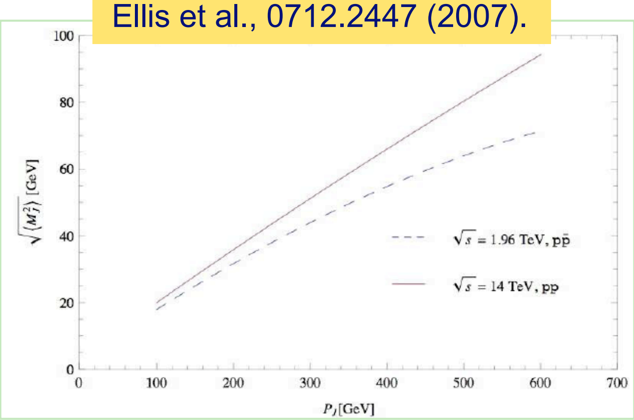
1. Introduction and Motivation
2. Data Selection
3. Calibrating and Correcting Jet Observables
4. Systematic Uncertainties
5. Algorithm Dependencies
6. Next Steps

Representing CDF Collaboration

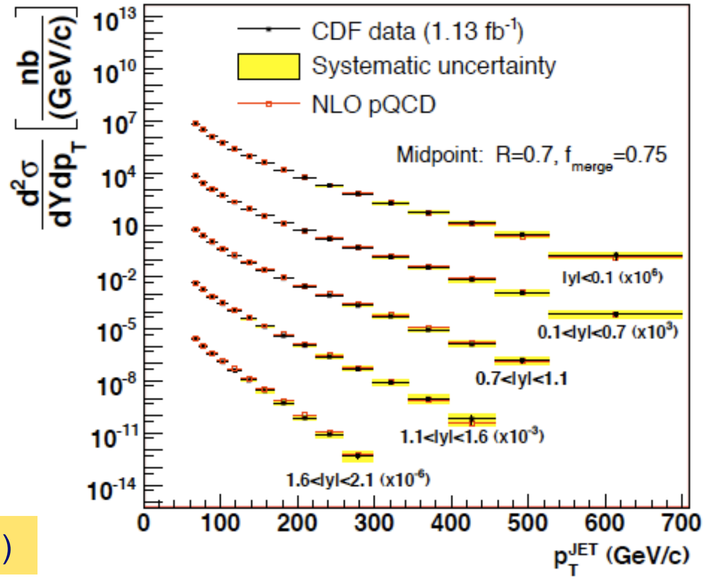
Raz Alon, Gilad Perez & Ehud Duchovni
Weizmann Institute of Science
&
Pekka K. Sinervo, FRSC
University of Toronto

Study Motivation

- **Mass of high-p_T jets important property – but mostly theory studies**
  - **High mass: QCD at NLO predicts jet mass (eg., Ellis et al, 0712.2447, Alemeida, et al. 0810.0934)**
  - **Such jets form significant background to new physics signals**
    - Examples: high p_T tops, Higgs, neutralino …
- **Focus on jets with $p_T > 400$ GeV/c**
  - **CDF II has collected ~9 fb^{-1}**
  - **Have several thousand jet candidates**
  - **Opportunity for 1stsystematic study of jet mass, other substructure observables**



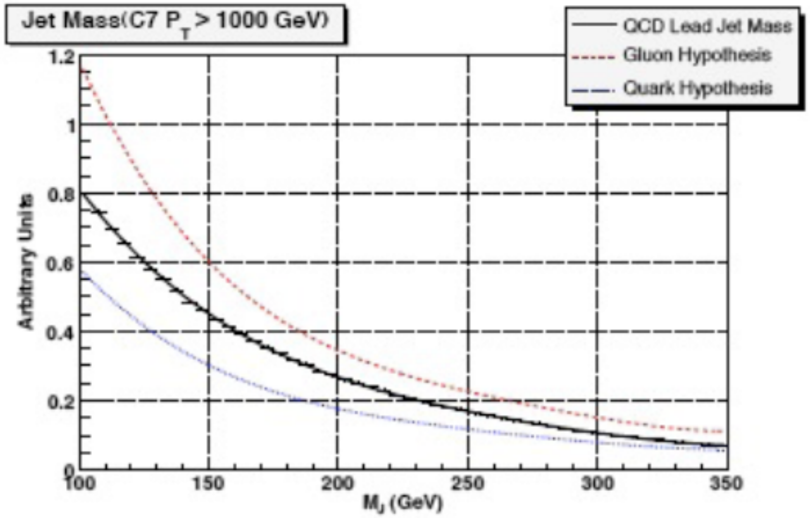
Ellis et al., 0712.2447 (2007).



CDF Collaboration, PRD 78, 052006 (2008)

NLO QCD Predicts Jet Function

- **Expect high mass jets arise primarily from 1-gluon radiation**



$$\frac{\sigma}{dp_T\, dm^{jet}} \quad J^{q,g}(m^{jet}, p_T, R)\frac{d\hat{\sigma}}{dp_T}$$

$$J^{q,g}(m^{jet}, p_T, R) \cong \alpha_s\left(m^{jet}\right)\frac{4C^{q,g}}{\pi m^{jet}}\log\left(Rp_T / m^{jet}\right)$$

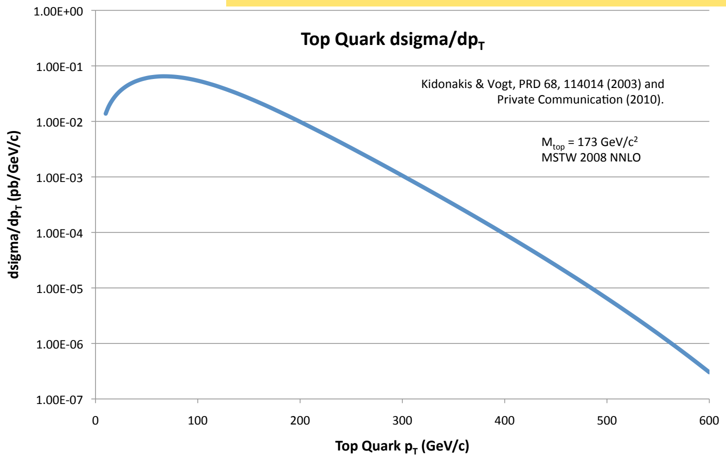
Alemeida, et al. 0810.0934

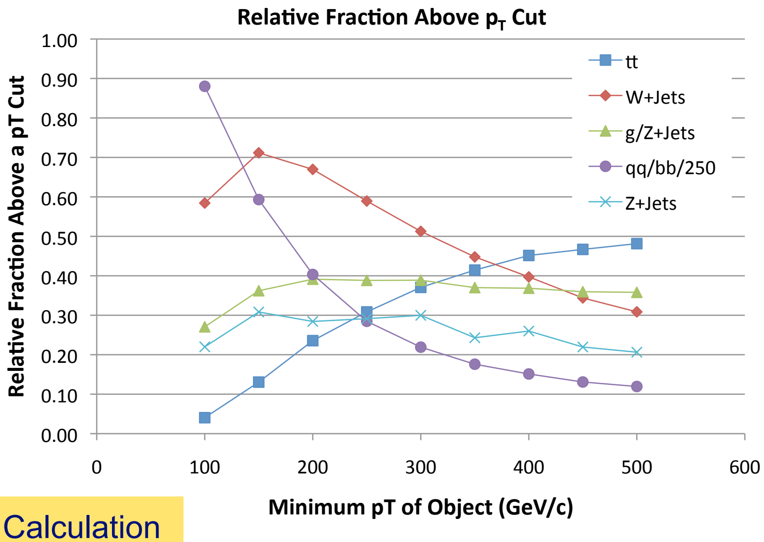
  - **Robust NLO prediction for**
    - Shape of high mass tail (and quark/gluon difference)
    - Relative rate of high mass QCD jets

Boosted Objects at Tevatron

- SM sources for high-p_T objects calculable
  - Dominated by light quarks & gluons
- Expect other contributions
  - Fraction of top quarks ~1.5% for $p_T > 400$ GeV/c
    - Total rate 4.45±0.5 fb (Kidonakis & Vogt)
    - PYTHIA 6.216 rate is 6.4 fb (scaling total cross section to measured world average)
  - Expect W/Z production of similar order

Kidonakis & Vogt, PRD 68, 114014 (2003)

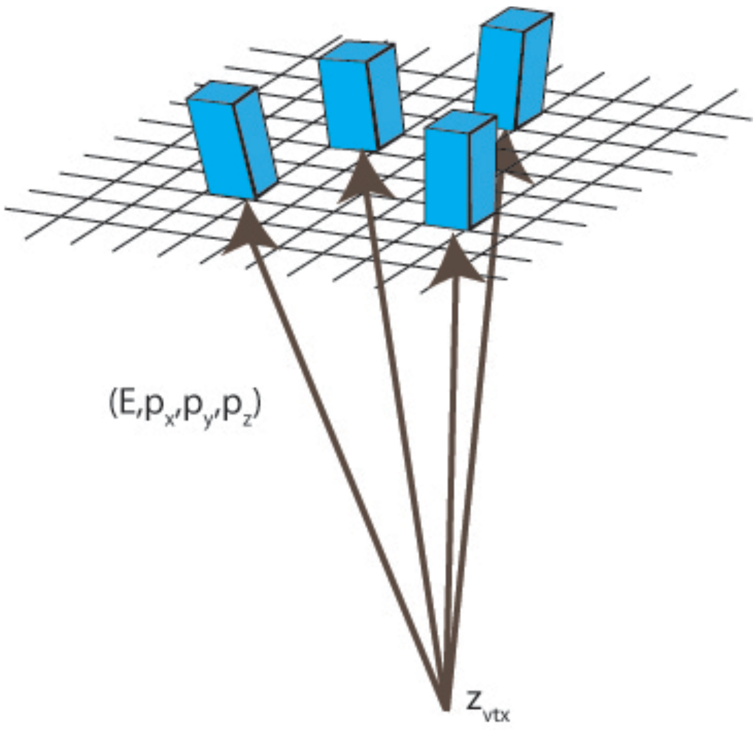




PYTHIA 6.4 Calculation

Strategy for Analysis

- **Select high p_T jets in CDF central calorimeter**
  - **Use tower segmentation to measure jet mass**
    - Confirm with tracking information
  - **Employ standard "e-scheme" for mass calculation**
    - 4-vector sum over towers in jet
    - Each tower is a particle with m = 0
    - Four vector sum gives (E, p_x, p_y, p_z)



- **Employ Midpoint cone jets**
  - **Best understood in CDF II context**
  - **Compare results with anti-k_T and Midpoint with "search cones" (Midpoint/SC)**

N.B. CDF central towers are $\Delta\eta \times \Delta\phi \sim 0.11 \times 0.26$

Jet Algorithms

- Cone algorithms used for most Tevatron studies
  - Long history – quite separate from e^+e^- work
  - JetClu was CDF reference
    - Required "seed" to initiate
    - Significant IRC sensitivity
- Midpoint developed to reduce IRC sensitivity
  - Use seeds, but then recluster with seeds "midway" between all jets

Use Fastjet Framework!

M. Cacciari, G.P. Salam and G. Soyez,
Phys. Lett. B641, 57 (2006) [hep-ph/0512210].

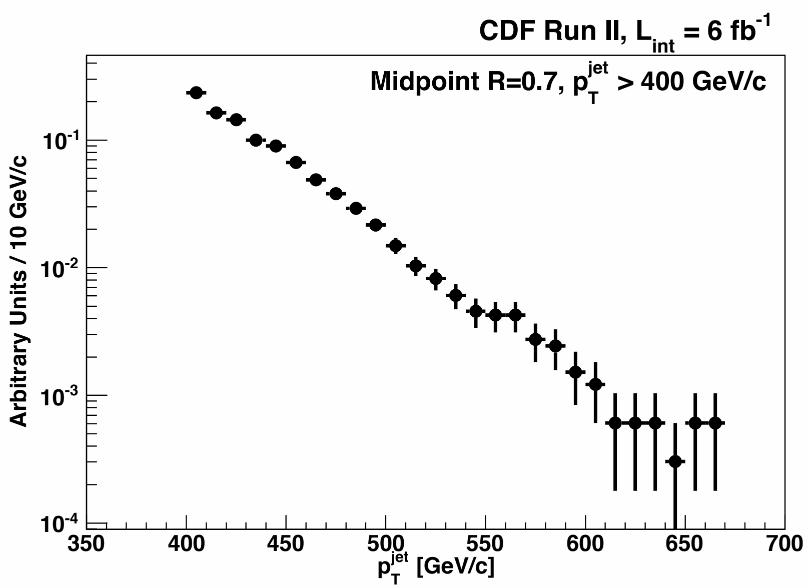
- Cone algorithms had "dark tower" problem
  - Unclustered energy due to split/merge/iteration procedure
  - Proposed solution: Midpoint with "search cones"
    - Find jets with cone size R/2
    - Fix jet direction, cluster with size R
  - Midpoint/SC was used for various studies 2006-2008
- Anti-k_T algorithm developed
  - No IR sensitivity
  - Still retained many of the benefits of a "cone" algorithm

Data Selection

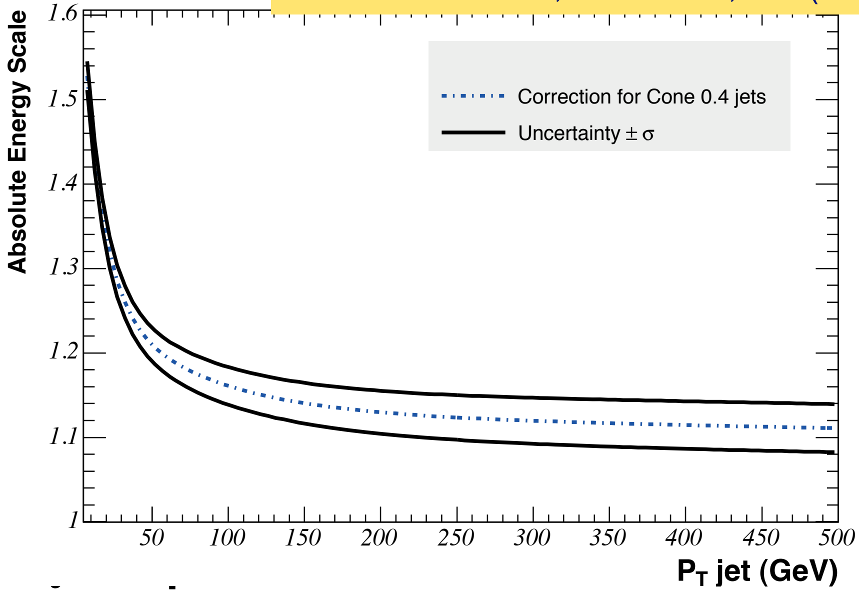
- **Analyzed inclusive jet sample**
  - **Trigger requires $E_T > 100$ GeV**
  - **Analyzed 5.95 fb^{-1} sample**
- **Selected data with focus on high p_T objects**
  - **Kept any event with**
    - Jet with $p_T > 300$ GeV/c and $|\eta| < 0.7$
    - Used cones of R=0.4, 0.7 and 1.0
- **Processed 76M events**
  - **Selected subsample with**
    - $p_T > 400$ GeV/c
    - $|\eta| \in (0.1, 0.7)$
- **Performed cleaning cuts**

$$S_{MET} \equiv E_T^{MISS} \Big/ \sqrt{\sum_{i\ towers} E_T^i}$$

  - **Event vertex, jet quality and loose S_{MET} (< 14)**
- **Resulted in 2700 events using jets with R=0.7**



Jet Mass Corrections

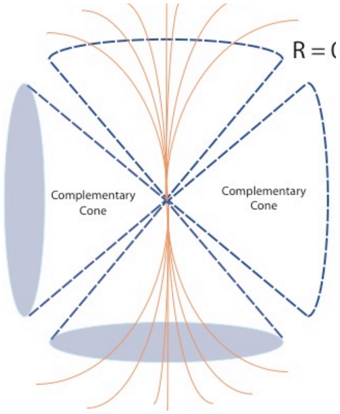


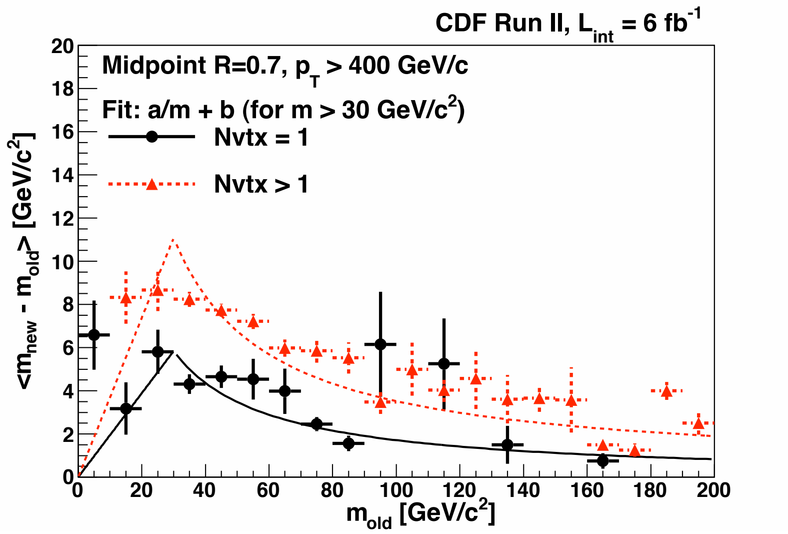
CDF Collaboration, NIM A 566, 375 (2006)

- **Corrected jet mass using standard jet corrections**
  - **Further correction needed for multiple interactions (MI)**
  - **Use N_{vtx}=1 and N_{vtx}>1 events to determine MI effect**
- **Investigated other effects:**
  - **Effect of calorimeter inhomogen**
    - Varied pseudorapidity window – no significant changes in mass
  - **Calorimeter segmentation and jet recombination**
    - Varied position of towers (especially azimuth) and corrections for geometry
  - **Calorimeter response across face of jet**
    - Detailed study of tracking/calorimeter response in data and MC/detector simulation
  - **Jet energy scale vs algorithm (Midpoint, Midpoint/SC, anti-k_T)**
    - Saw < 1 % difference

Effects of MI and UE

- **Additional contribution from**
  - **Underlying Event (UE)**
  - **Multiple Interactions (MI)**
    - Average # interactions ~3/crossing
- **Looked at purely dijet events**
  - **Defined cones (same size as jet) at 90° in azimuth (same η)**
  - **Took towers in cones, and added to leading jet in event**
    - Mass shift, on average, is same shift coming from UE and MI
- **Separately measure N_{vtx}=1 events**
  - **Gives UE correction separately**

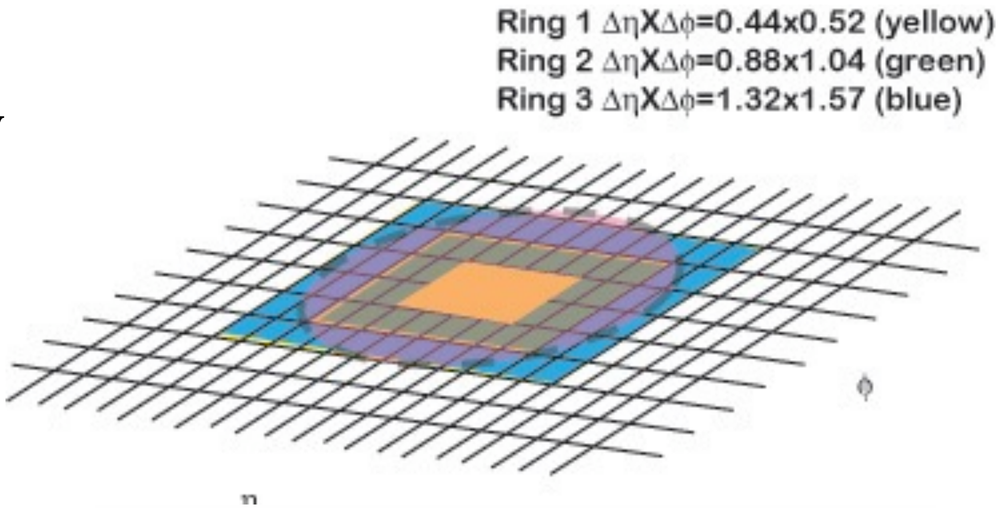


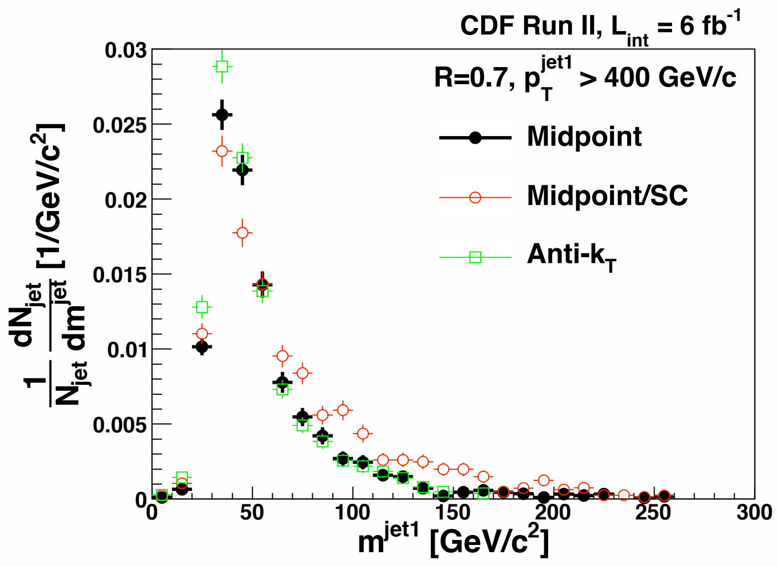


Correction scales as R^4

Inter-Jet Energy Calibration

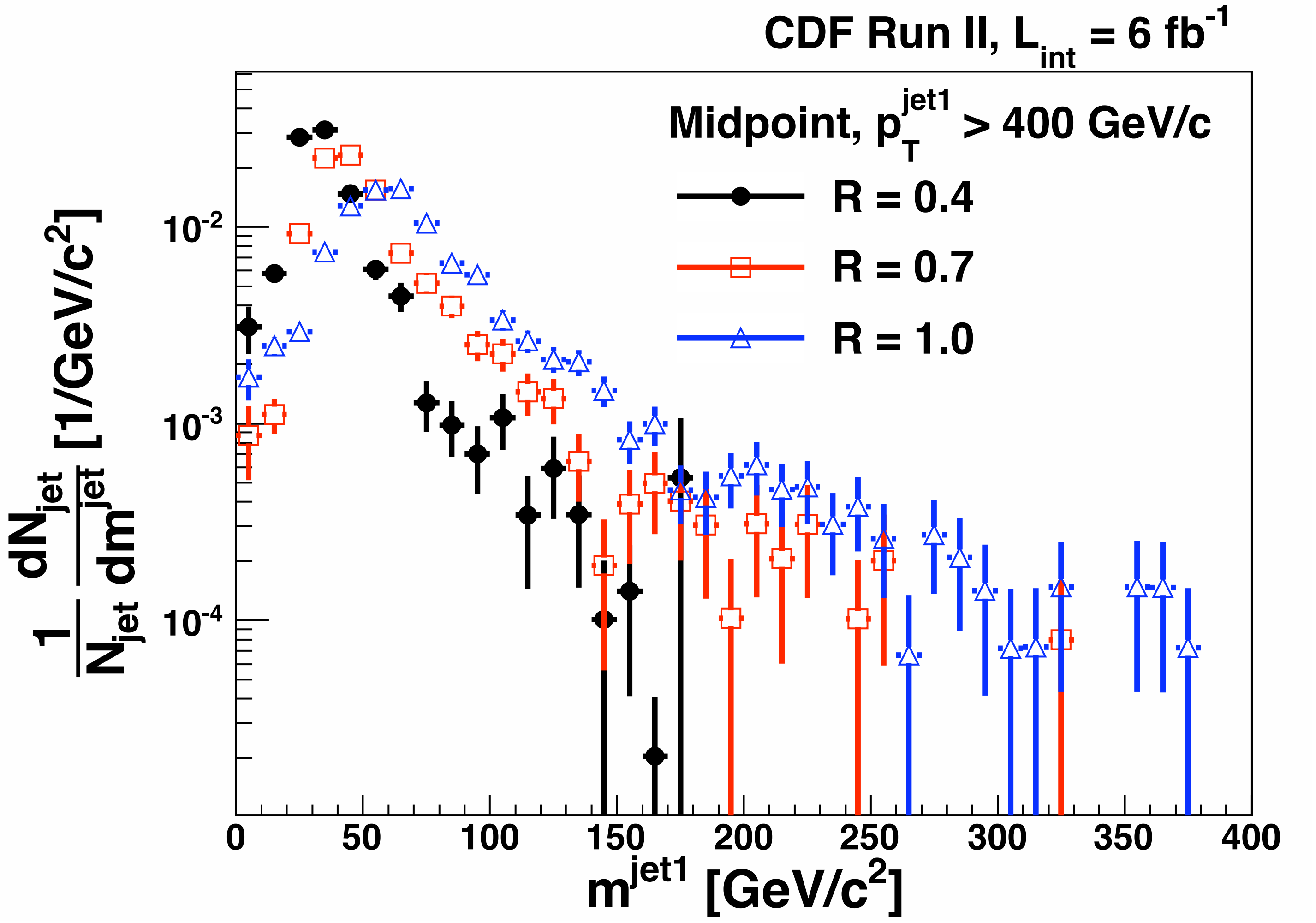
- **Jet mass arises from deposition of varying energy per tower**
  - **Performed study to compare momentum flow vs calorimeter energy internal to jet**
    - Defined 3 rings and compared observed p_T/E_T with simulation
- **Resulted in constraints on calorimeter relative response**
  - **At m^{jet}=60 GeV/c^2, σ_m=1 GeV/c^2**
  - **At m^{jet}=120 GeV/c^2, σ_m=10 GeV/c^2**
- **Largest source of systematic uncertainty**





Comparison with Cone Size

- **Compare**
  - ○ R=0.4
  - ○ R=0.7
  - ○ R=1.0



Systematics on m^{jet}

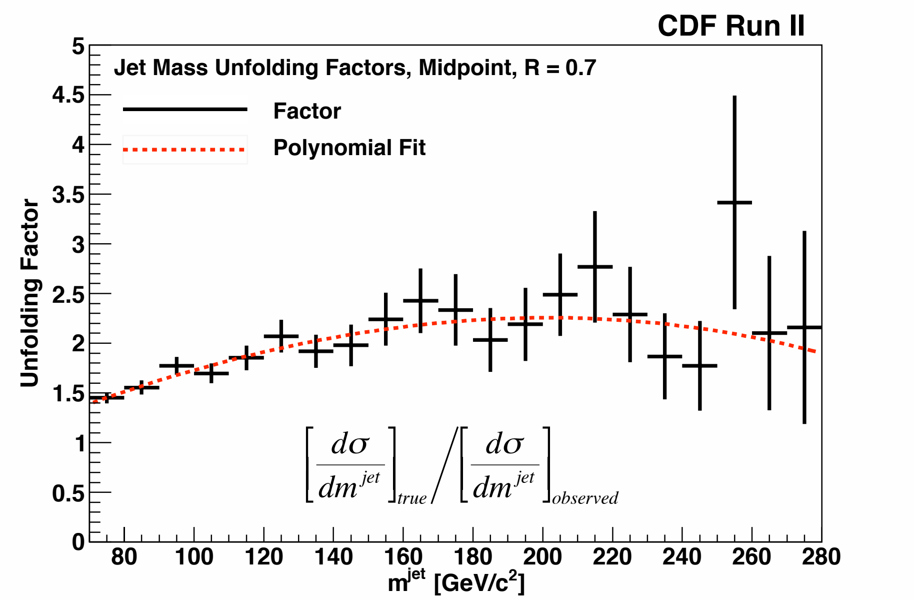
- **Sources of systematics:**
  - **Calorimeter energy scale**
    - Varies from 1 to 10 GeV/c^2 for 65 to 120 GeV/c^2 mass jets
  - **UE and MI modelling**
    - Estimate 2 GeV/c^2 based on uncertainty in high mass correction
  - **PDF Uncertainties**
    - Used standard 20 eigenvector decomposition to assess MC uncertainties
    - Shown when direct comparison made with PYTHIA 6.216
- **Uncertainties are uncorrelated**
  - **Combined in quadrature, gives total jet mass uncertainty of**
    - 3.4 GeV/c^2 for m^{jet} = 60 GeV/c^2
    - 10.2 GeV/c^2 for $m^{jet} >$ 100 GeV/c^2
- **Effects jet mass distributions arising from bin-to-bin migration**
  - **Small systematic shifts in other substructure variables**
  - **Determined using 90° cone approach (see G. Perez's talk)**

Determining Jet Function

- Key prediction is “jet function”

$$J^{q,g}(m_{jet},p_T,R) \cong \alpha_s\left(m_{jet}\right)\frac{4C^{q,g}}{\pi m_{jet}}\log\left(Rp_T/m_{jet}\right)$$

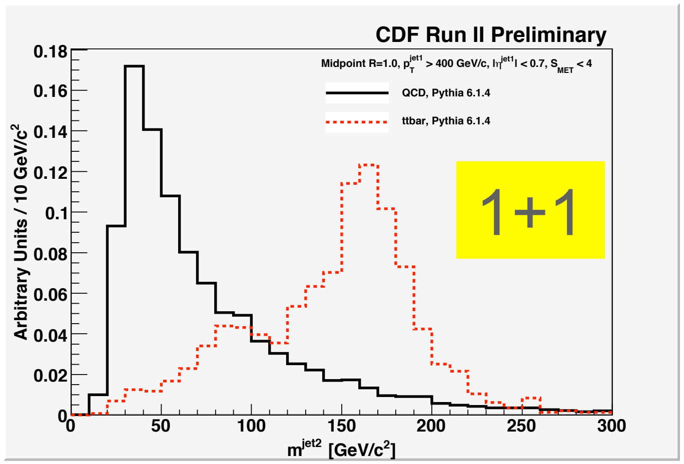
  - Just m^{jet} distribution?
- However, large correction comes from jet p_T cut
  - p_T of low mass jets has ~10% broader resolution than high mass jets
  - More events in sample with true $p_T < 400$ GeV/c at low m_{jet} vs high m_{jet}
    - Aggravated by much larger rate at low jet mass

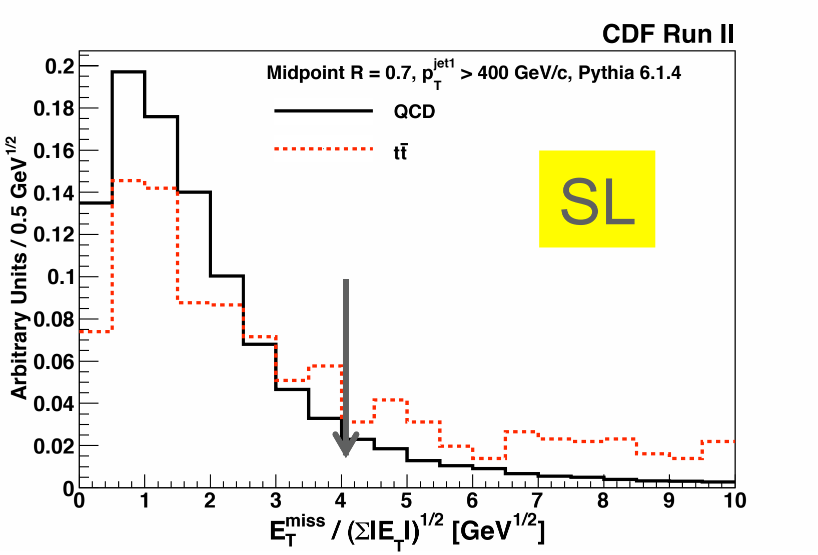


- Verified by studies of recoil jet
  - No intrinsic p_T bias
- Calculated correction with MC
  - Hadronization uncertainty 10%

Reducing Top Contamination

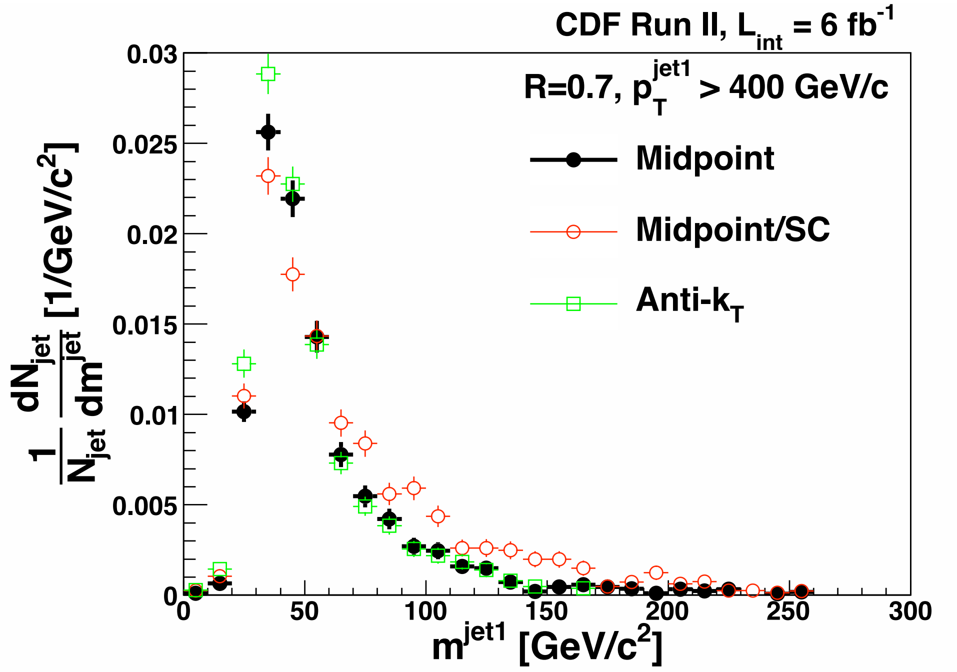
- **Expect about 1.6 fb of high p_T jets from top in sample**
  - **Eliminate by rejecting events with**
    - $m^{jet2} > 100$ GeV/c^2
    - Missing E_T Significance (S_{MET}) > 4
  - **Use jet cone of R=1.0 for improving top jet tagging**
  - **Lose 28% of jet candidates**
    - 2576 events using R=0.7 jets
    - 145 events with jet with $p_T > 500$ GeV/c
- **After top-rejection, expect ~0.3 fb of top jets**
  - **Comparable rates for W/Z jets**





Properties of QCD Jet Sample

- **After top rejection**
  - **Left with sample dominated by light quarks and gluon**
  - **Compare high mass region with QCD theory**
  - **Algorithm dependence?**
    - Midpoint and anti-k_T very similar
    - Midpoint/SC quite different

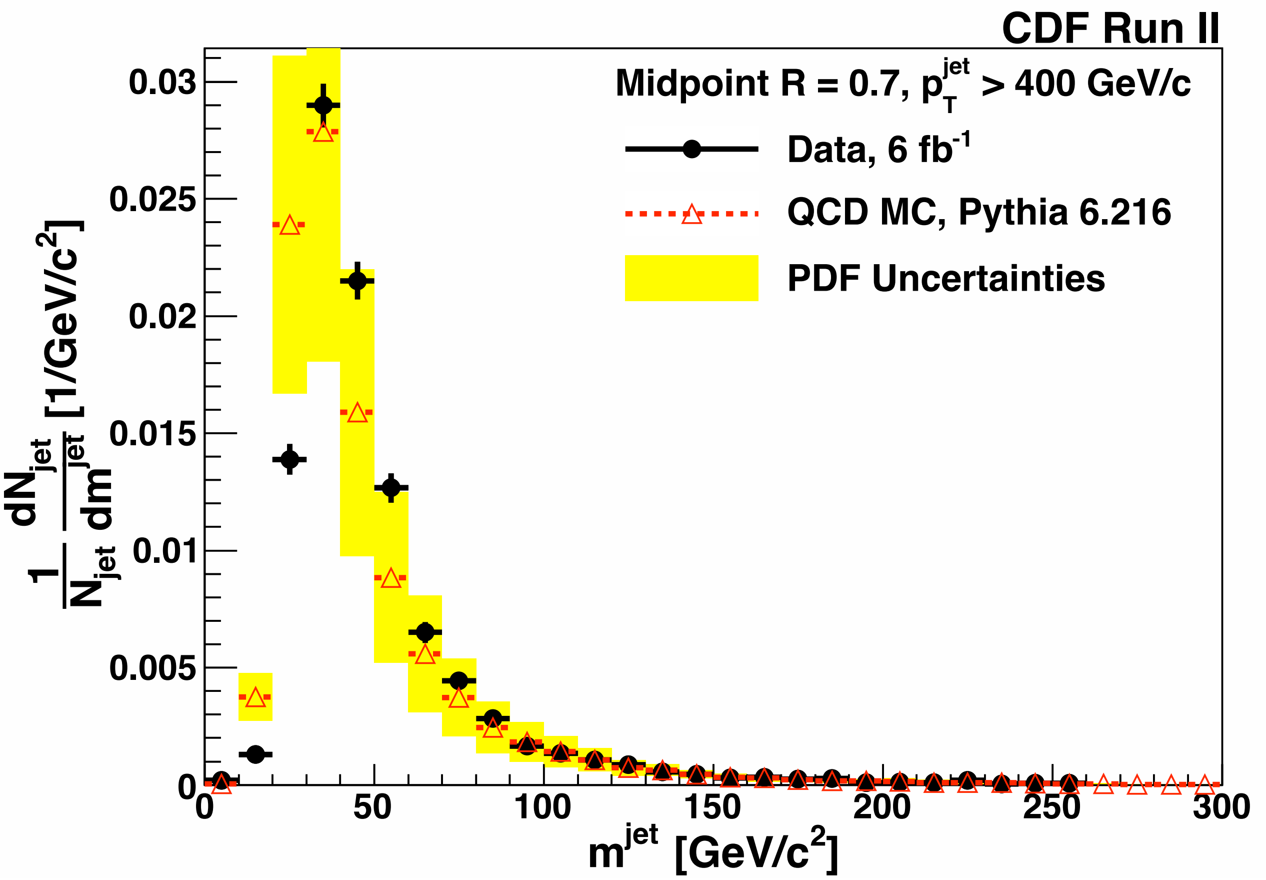


Cut Flow		
All Data, 5.95 fb^{-1}	75,764,270 events	
	R = 0.4	**R = 0.7**
At least one jet with $p_T > 400$ GeV/c, $\|\eta\|$ in (0.1, 0.7), and event quality cuts	2153 events	2700 events
$m^{jet2} < 100$ GeV/c^2 and $S_{MET} < 4$ (with $p_T^{jet2} > 100$ GeV/c and MI corrections)	1837 events	2108 events

- **Low-mass peak arises from non-perturbative QCD effects**
  - **Opportunity to study the properties of the high mass jets**
  - **Gilad will say more…**

Comparison with PYTHIA

- **PYTHIA 6.216**
  - Standard CDF II QCD sample
  - PDF uncertainties based on eigenvector decomposition
- **Agreement is reasonable**
  - Low-mass peak few GeV/c^2 lower
  - Large PDF uncertainties at low mass

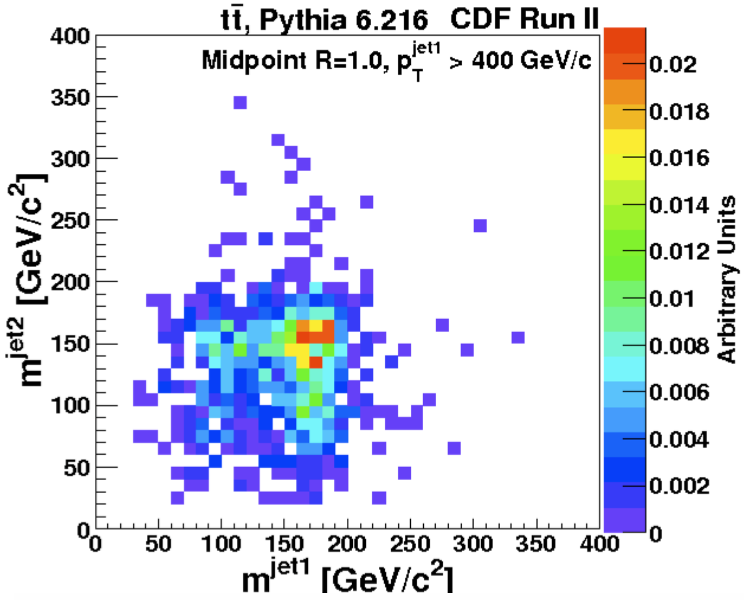


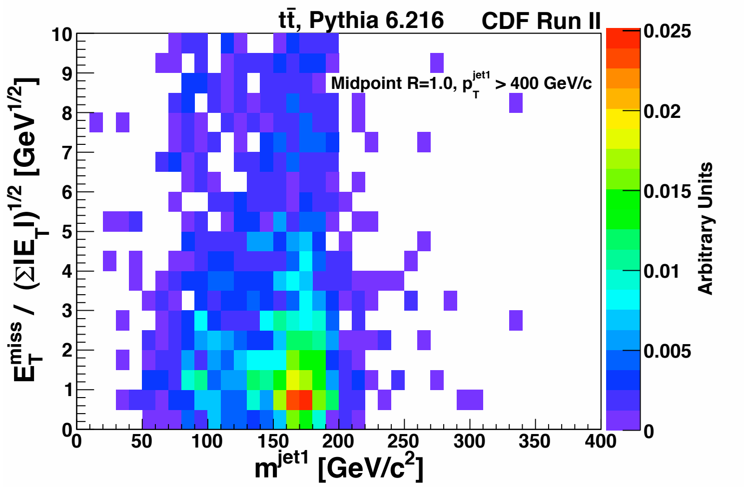
Search for Boosted Top

- **Analysis suggests sensitivity to boosted top**

$\gamma \sim 2.5$

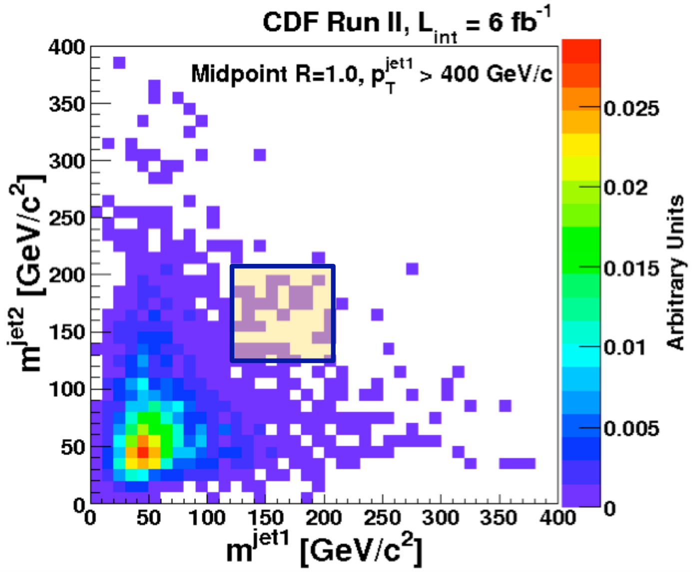
- **Two topologies:**
  1. **All hadronic ("1+1")**
     - Two massive jets recoiling ($\varepsilon \sim 11\%$)
  2. **Semi-leptonic decay ("SL")**
     - Require $S_{MET} > 4$ ($\varepsilon \sim 7\%$)
- **MC predicts ~0.8 fb**
  - ○ **Divided about 60:40 between topologies**
    - Highest efficiency channel for top (~18%)
  - ○ **Important handles for background:**
    - masses of QCD di-jets not correlated
    - Jet mass and S_{MET} not correlated

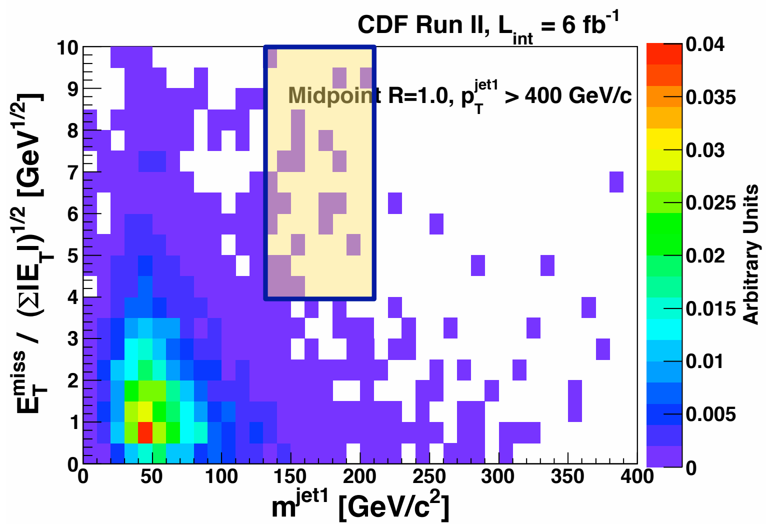




Strategy for Detecting Top

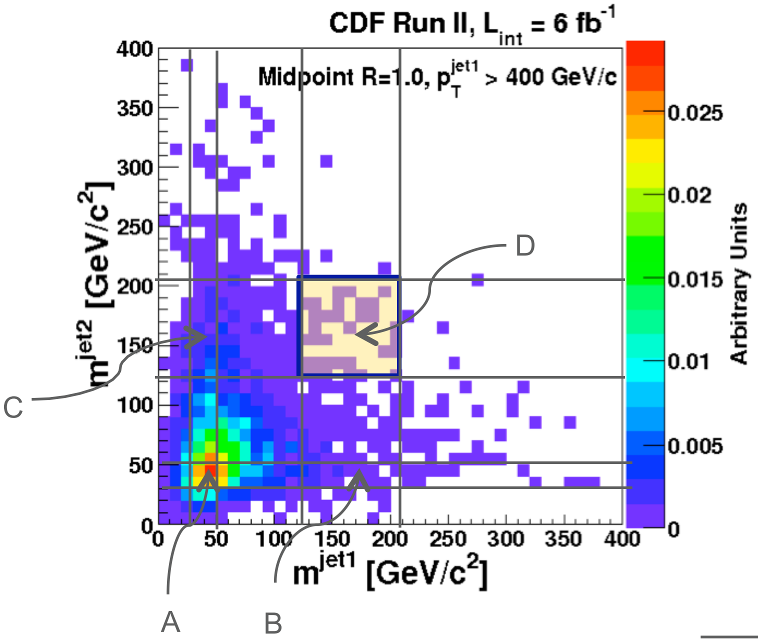
- **Keep selection simple**
    - Focus on two separate channels
- **All Hadronic Top (1+1)**
    - Require 2 jets with $130 < m^{jet} < 210$ GeV/c^2
    - Require $S_{MET} < 4$
    - Estimate background using "ABCD" technique
- **Semi-leptonic top**
    - Require $4 > S_{MET} > 10$
    - Require 1 jet with $130 < m^{jet} < 210$ GeV/c^2
    - Estimate background using "ABCD" technique





Best “Simple” Counting of 1+1

- With R=1.0 cones, m^{jet1} and m^{jet2} are equally powerful
  - Use jet mass (130,210) GeV/c^2 to define ttbar candidates
  - Expect 3.0±0.8 top quark events to populate this region

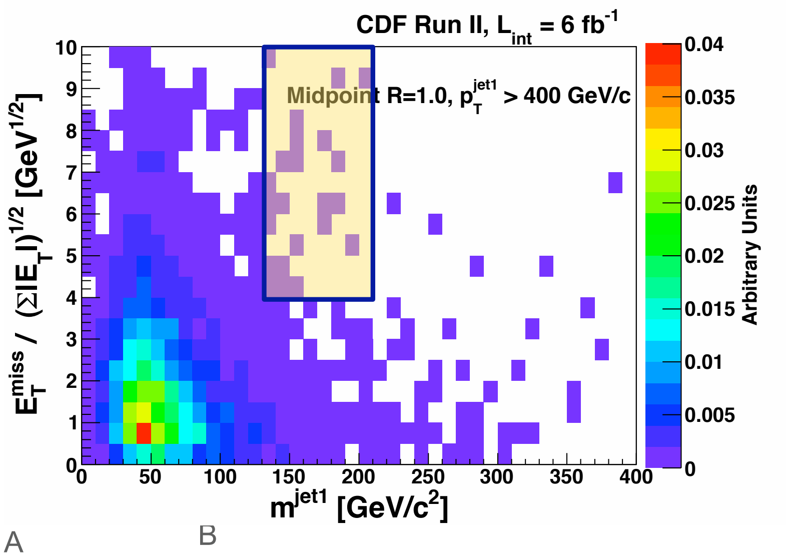


- Employ data to estimate backgrounds
  - Define mass windows
    $m^{jet} \in (130,210)$ GeV/c^2
    $m^{jet} \in (30,50)$ GeV/c^2
    - Use fact that m^{jet} distributions uncorrelated for background
    - Signal is region D
    - In “1+1” sample, predict 13±2.4 (stat) bkgd events
      - Observe N_D=32 events

$$N_D^{Pred} = N_C \left[\frac{N_B}{N_A} \right]$$

Best "Simple" Counting for SL

- In case of recoil semileptonic top, use m^{jet1} and S_{MET}
  - Assumption is the S_{MET} and m^{jet1} are uncorrelated
  - Expect 1.9±0.5 top quark events to populate this region



- Employ data to estimate backgrounds
  - Use regions $m^{jet1} \in (30,50)$ & (130,210) GeV/c^2
  - $S_{MET} \in (2,3)$ & $S_{MET} \in (4,10)$
    - In "SL" sample, predict 31±8 (stat) bkgd events
      - Observe N_D=26 events
      - About a -0.6 σ deficit!

Region	m^{jet1} (GeV/c^2)	S_{MET} ($\sqrt{GeV/c^2}$)	Data (Events)	MC (Events)
A	(30, 50)	(2, 3)	256	0.01
B	(130, 210)	(2, 3)	42	1.07
C	(30, 50)	(4, 10)	191	0.03
D (signal)	(130, 210)	(4, 10)	26	1.90
Predicted QCD in D			31.3 ± 8.1	

Uncertainties

- Background uncertainty (±10.2 GeV/c^2 jet mass scale)
  - ±30% uncertainty
- Uncertainties on top efficiency (SM production)
  - Primarily jet energy scale of ±3% on pT -> ±25% on σ
- Background statistics
  - ±11% from counting
- Luminosity
  - ±6% on integrated luminosity
- MC m^{top} (±2 GeV/c^2)
  - ±0.3%
- Overall uncertainties added in quadrature
  - ±41% overall
- Incorporated into upper limit calculation
- Use a CL_s frequentist method
  - Marginalize nuisance parameters
  - Same as used in Higgs and single top searches

Top Quark Cross Section Limit

- **Assume we observe signal + background**
  - **Set upper limit on SM production σ for top quark with $p_T > 400$ GeV/c**
- **Observe 58 events with 44+/-8 background**
  - **Calculate 95% CL upper limit using CL_s method**
    - Systematic uncertainties incorporated a la CDF 8128 (T. Junk)
    - $N_{LIM} = 43.3$ events
  - **Efficiency from MC**
    - 452 & 283 ttbar expected in 2 channels (out of 4041 MC events)
    - Efficiency = 0.182
- **Upper limit on cross section for $p_T > 400$ GeV/c**

$$\sigma_{95\%} = \frac{N_{LIM}}{\int L\,dt\;\varepsilon}$$

$$= \frac{43.3}{(5.95)(0.182)} = 40 \text{ fb}$$

- **Can also set limit on 1+1 only**
  - **Assume massive (m ~ mtop) object, pair-produced, decaying hadronically**
  - **Include SM top as background**

$$\sigma_{95\%} = \frac{N_{LIM}}{\int L\,dt\;\varepsilon}$$

$$= \frac{30.2}{(5.95)(0.254)} = 20 \text{ fb}$$

Conclusions

- **First measurement of jet mass (and substructure) for high p_T jets**
  - **Being confronted by data forces one to understand systematics**
    - Multiple interaction corrections
    - Calibration of mass scale
  - **Allows for test of QCD predictions**
  - **Algorithm dependence**
    - Anti-k_T and Midpoint very similar
    - Midpoint/SC produces "fatter" jets
- **Next talk will show substructure results**
- **Search for boosted top possible**
  - **Achieve**

$$S/\sqrt{B} \approx 0.75$$

  - **Set $\sigma < 40$ fb at 95% CL**
  - **Limited by statistics!**
- **Real task is to observe at LHC**
  - **Tevatron program will end 2011**
  - **ATLAS and CMS already have comparable sized samples with 50 pb^{-1}**
  - **Much higher p_T jets already!**

BACKUP SLIDES

MI/UE Corrections

- Looked at how to make MI correction in a variety of ways
  - Looked at mass corrections event-by-event
  - But statistical fluctuations large, event-to-event
  - Chose to develop a parametrized correction
- Note that:

$$\delta m^{jet} \simeq \frac{E_{tower} E_{jet} \Delta R}{m^{jet}}$$

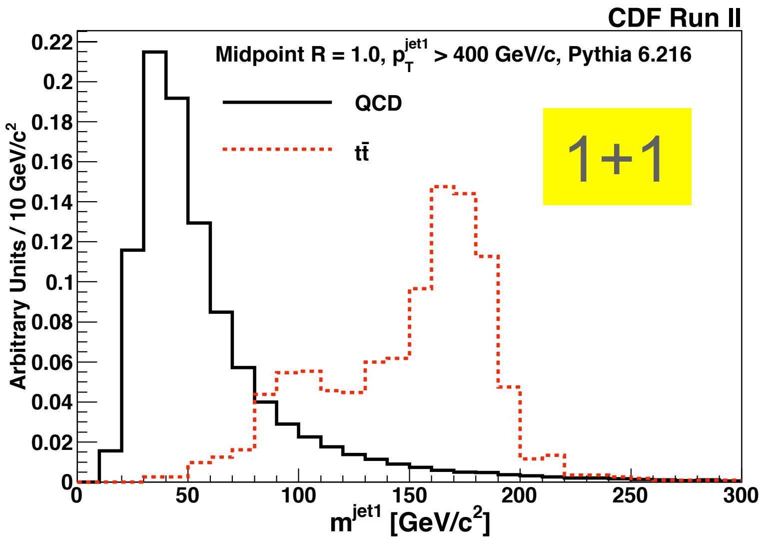
- Expect MI correction to scale with R^4:
  - Exactly what we see when comparing R=0.4 and R=0.7
- PYTHIA UE agrees well with data – same UE mass correction
- Use that to scale corrections for R=1.0
  - Method doesn't work with larger cone because of overlap

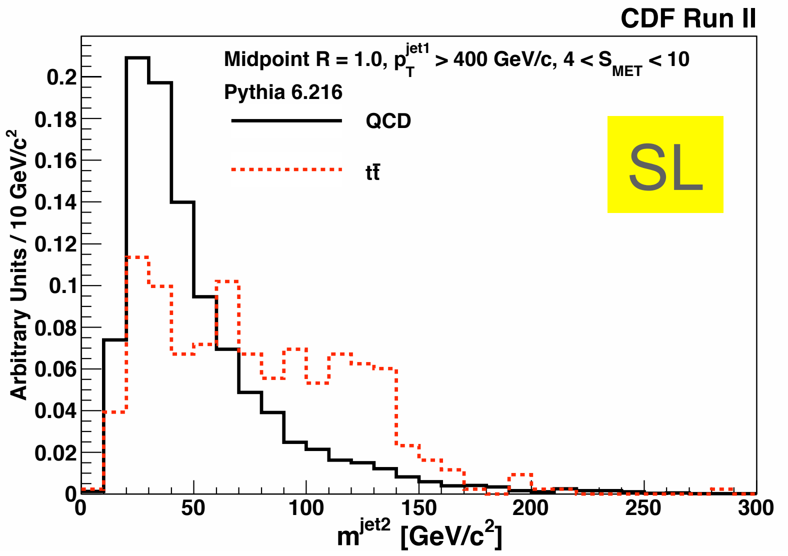
Internal Jet Energy Scale

- Overall jet energy scale known to 3%
  - The relative energy scale between rings known to 10-20%, depending on ring
  - Use this to constrain how far energy scale can shift
- Do first for $m^{jet} \sim 60$ GeV/c^2 – use average jet profile
  - Extract from that a limit on how much “Ring 1” energy scale can be off - ± 6%
  - Then do the same for mjet ~ 120 GeV/c^2
- Resulting systematic uncertainty is 9.6 GeV/c^2
  - Conservative estimate – used a very broad energy profile
    - No localized substructure assumed
- Take this as systematic uncertainty
  - Could constrain it better using single particle response
  - Note that fixed cone size is an advantage here

Reconstruction of Top

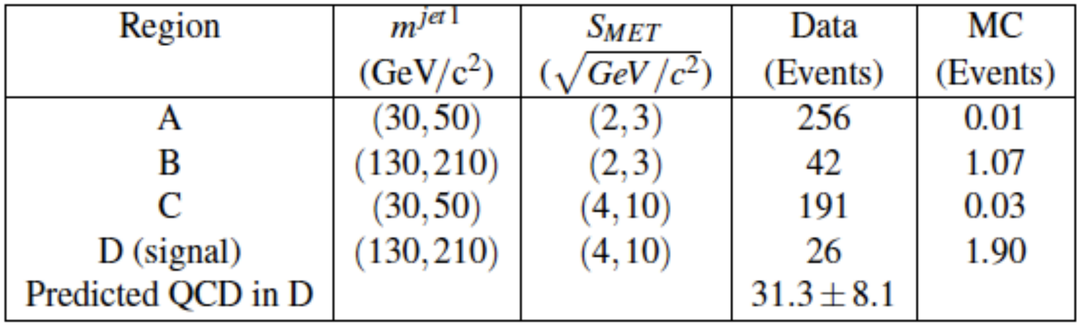
- **Leading jet in ttbar events has clear top mass peak**
  - **All events between 70 and 210 GeV/c^2 for R=1.0**
  - **See evidence of W peak**
    - B quark jet presumably nearby in those cases
  - **Clear that higher mass cut gives greater QCD rejection**
    - But also start to lose efficiency
  - **S_{MET} cut effectively identifies semi-leptonic decays (8%)**
- **B tagging not used**
  - **Can estimate mis-tags using data -> ~0.05%/jet**
  - **But large uncertainty in tagging efficiency in high pT jets**





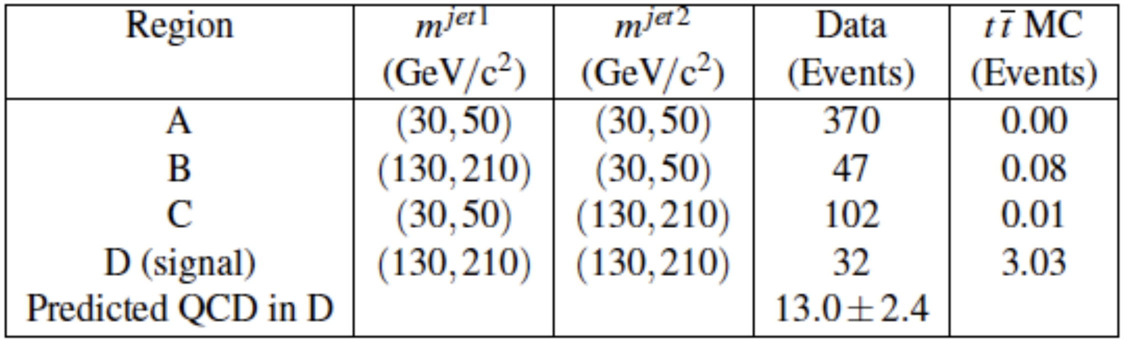
Background Calculations

- **Background calculations used "ABCD" technique**
- **SL**



Region	m^{jet1} (GeV/c^2)	S_{MET} ($\sqrt{GeV}/c^2$)	Data (Events)	MC (Events)
A	(30,50)	(2,3)	256	0.01
B	(130,210)	(2,3)	42	1.07
C	(30,50)	(4,10)	191	0.03
D (signal)	(130,210)	(4,10)	26	1.90
Predicted QCD in D			31.3 ± 8.1	

- **1+1**



Region	m^{jet1} (GeV/c^2)	m^{jet2} (GeV/c^2)	Data (Events)	$t\bar{t}$ MC (Events)
A	(30,50)	(30,50)	370	0.00
B	(130,210)	(30,50)	47	0.08
C	(30,50)	(130,210)	102	0.01
D (signal)	(130,210)	(130,210)	32	3.03
Predicted QCD in D			13.0 ± 2.4	