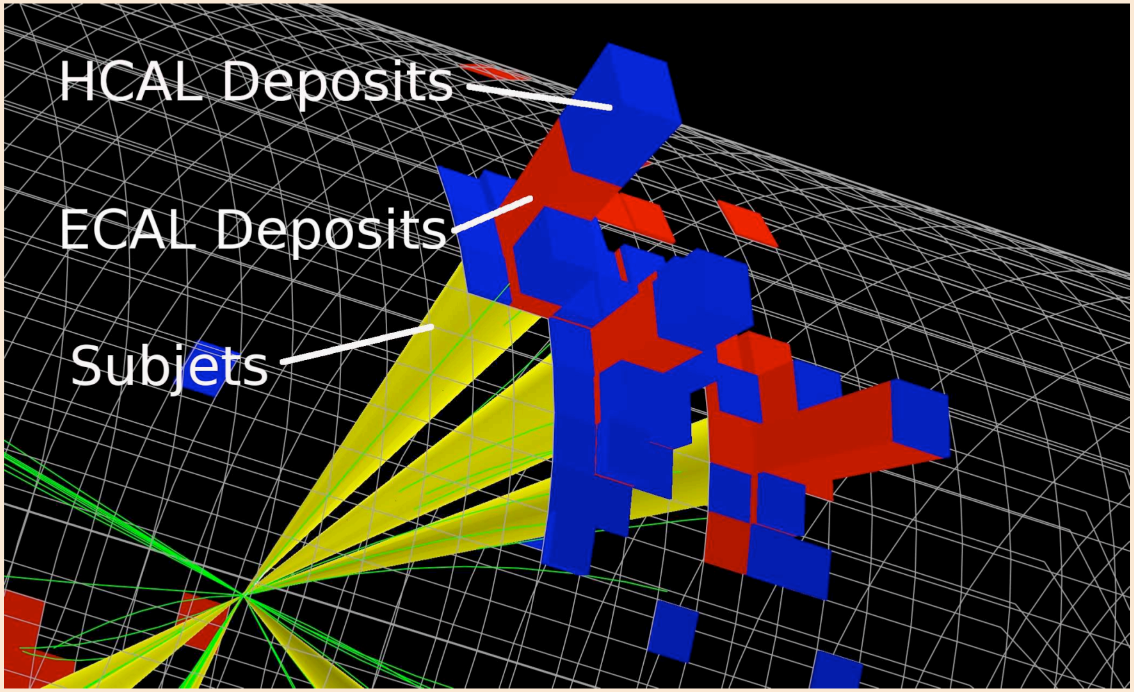


A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging

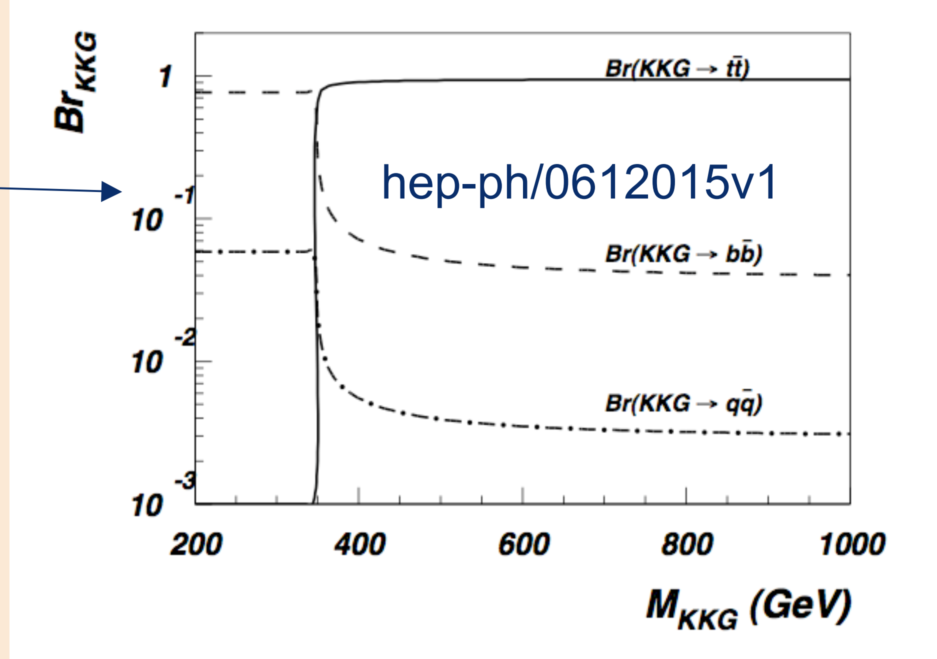
CMS PAS JME-09-001

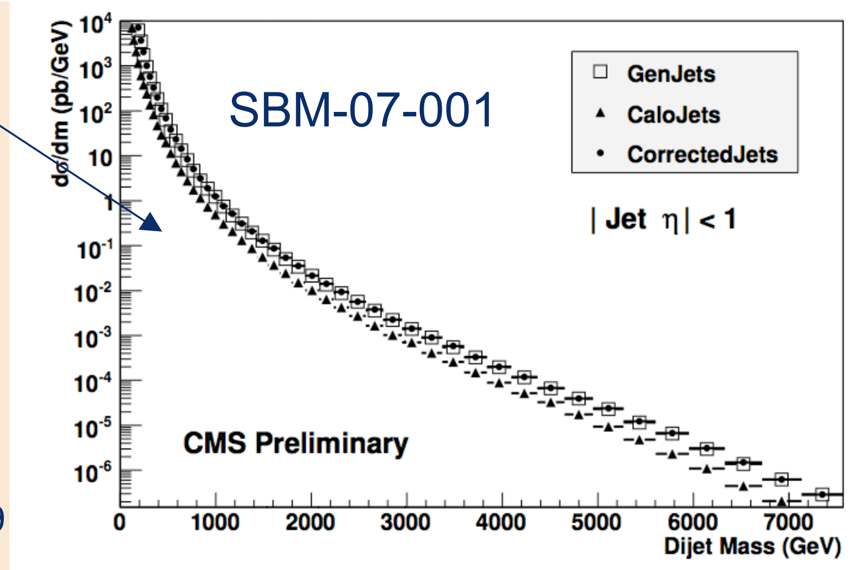
Salvatore Rappoccio

For the CMS Collaboration

Motivation

- New physics scenarios often involve ttbar resonances
- Large BR in all-hadronic channel (46%)
- Are we sensitive to this?
- Can we suppress the huge dijet background?





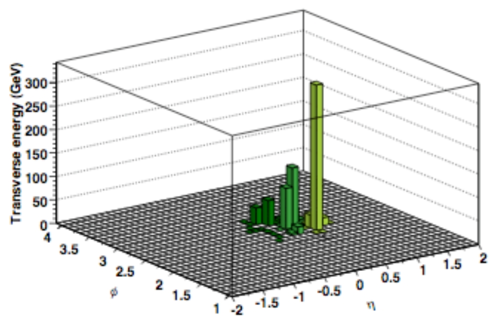
Motivation

Top-tagging: A Method for Identifying Boosted Hadronic Tops

David E. Kaplan, Keith Rehermann, Matthew D. Schwartz and Brock Tweedie
Department of Physics and Astronomy Johns Hopkins University Baltimore, MD 21218, U.S.A.

A method is introduced for distinguishing top jets – boosted, hadronically decaying top quarks – from standard model backgrounds using jet substructure. The procedure involves parsing the jet cluster to resolve features such as three light quark subjets, and then imposing angular and kinematic constraints. This method is much more efficient than simple invariant mass cuts or jet clustering with fixed angular size. With top-tagging, high p_T dijets can be rejected with an efficiency of around 99% while retaining 20-40% of the tops. This allows us to reach into the all-hadronic channel for new-physics signals, such as new heavy $t\bar{t}$ resonances, which ordinarily would be overwhelmed by the enormous dijet background. In addition, it will improve the reach for cases when one of the tops decays semi-leptonically, and may also have applications to single-top searches and studies of b-tagging efficiency at high p_T.

The Large Hadron Collider (LHC) is a top factory. The millions of $t\bar{t}$ pairs it produces, assuming they can be found, will provide profound insights into the standard model and its possible extensions. Most of the tops will be produced near threshold, and can be identified using the same kinds of techniques applied at the Tevatron – looking for the presence of a bottom quark through b-tagging, identifying the W boson, or finding three jets whose invariant mass is near m_t. However, some of the top quarks produced at the LHC will be boosted. In particular many of the interesting ones, such as those which come from $t\bar{t}$ resonances above 1 TeV, can have $E/m \gtrsim 5$, and these may be more difficult to identify. A boosted top will usually decay to a collimated collection of standard-model hadrons which looks remarkably simi-



- http://arxiv.org/abs/0806.0848
- Phys.Rev.Lett.101:142001,2008
- Based on discriminating top from non-top based on subjets of high-pt jets
- Assume crude deta-dphi segmentation (0.1x0.1) in paper

Achieved 99% background rejection, 40% top tagging efficiency

Cambridge-Aachen

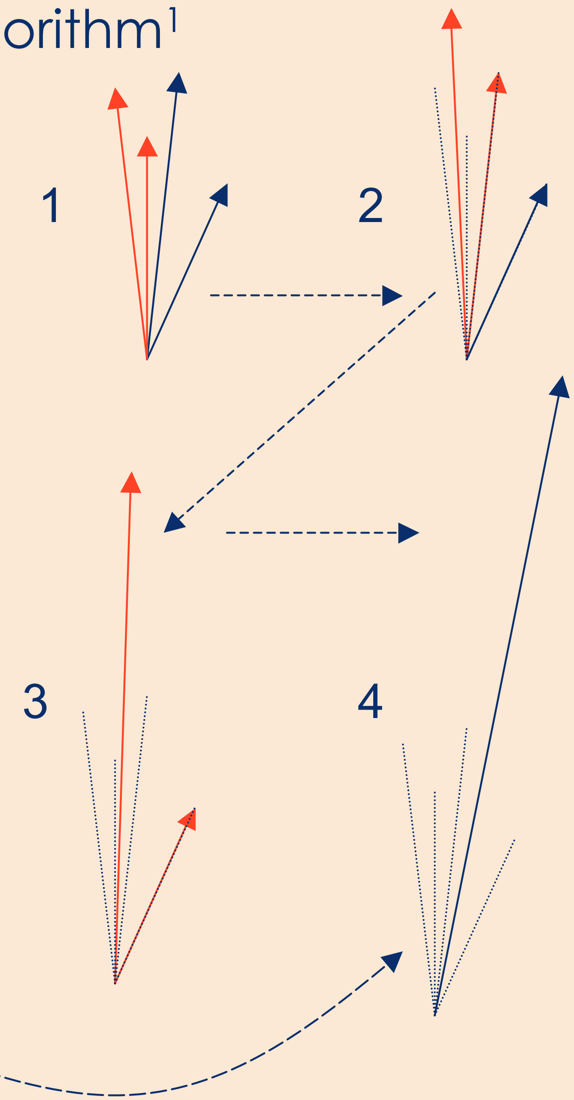
- C-A is a sequential recombination type algorithm[1]
- Pairwise examination of input 4-vectors
- Calculate d_{ij}

$$d_{ij} = min(k_{ti}^n, k_{tj}^n)\Delta R_{ij}^2 / R^2$$

  - N = 2: k_T
  - N = 0: Cambridge Aachen
  - N = -2: anti-k_T
- Also find the "beam distance"

$$d_{iB} = k_{T,i}^n$$

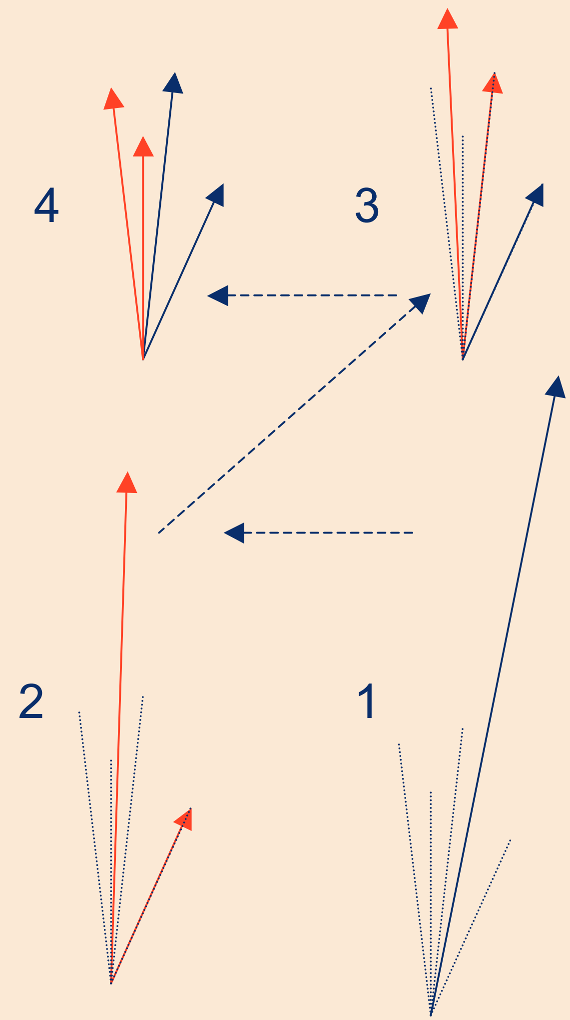
- Find min of all d_{ij} and d_{iB}
- If min is a d_{ij}, merge and iterate
- If min is a d_{iB}, classify as a final jet
- Continue until list is exhausted



[1]: arXiv:hep-ph/9707323

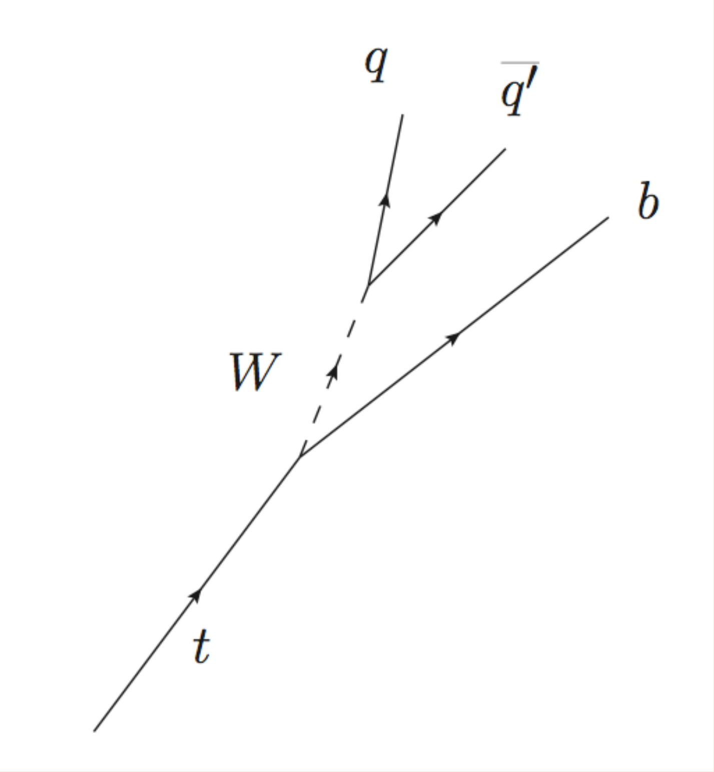
Top Tagging

- **"Hard jets":** Cluster jets with C-A
  - R = 0.8
  - p_T > 250 GeV
  - |y| < 2.5
- Reverse cluster sequence
  - Throw out soft clusters
    - Fraction of hard jet pt < 0.05
- Repeat on clusters until either
  - Have 3 or 4 hard clusters **(PASS)**
  - There are all soft clusters **(FAIL)**
- These are called **"subjets"**



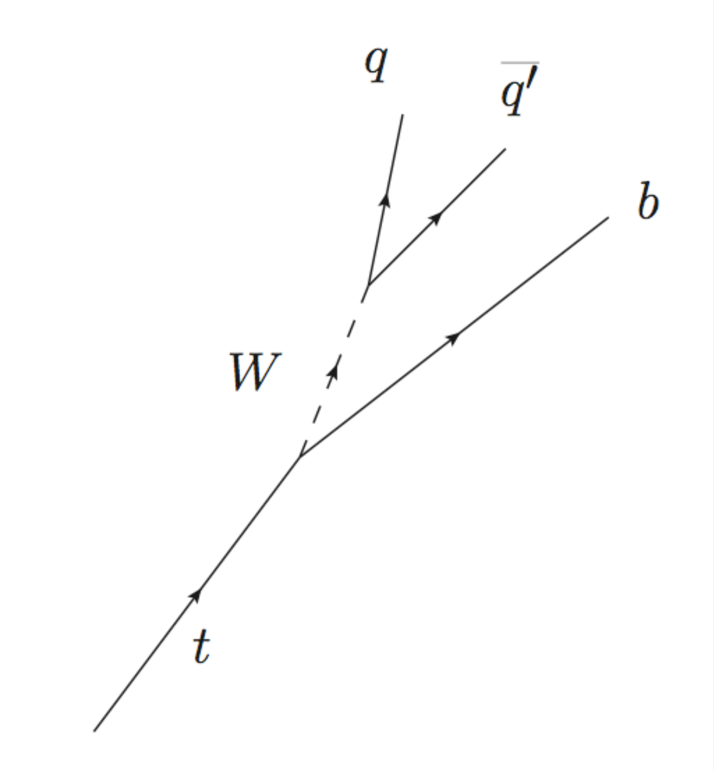
Top Tagging

- Discriminate **top jets** against **non-top jets**



Top Tagging

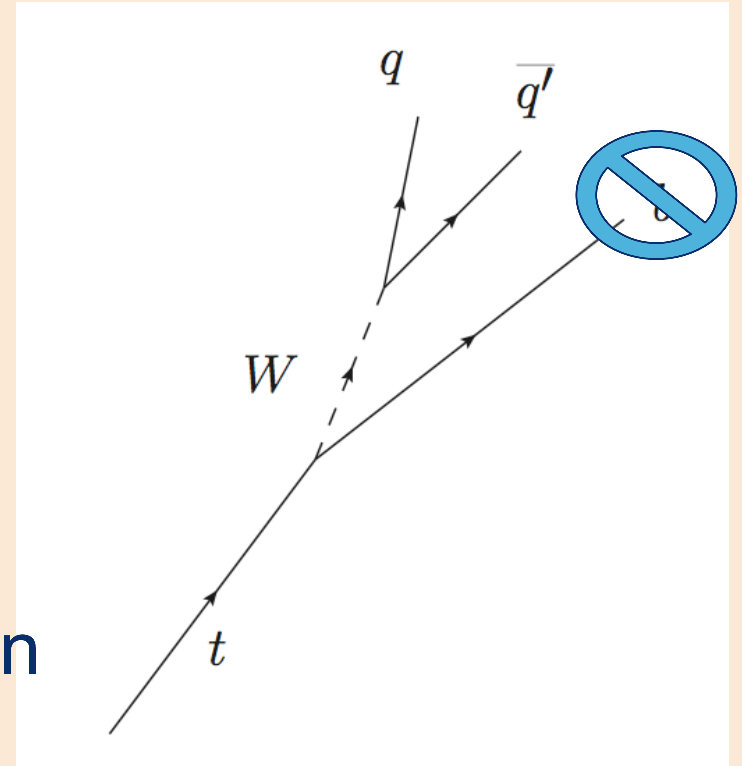
- Discriminate **top jets** against **non-top jets**
  - Top mass
  - W mass
  - b-tagging



Top Tagging

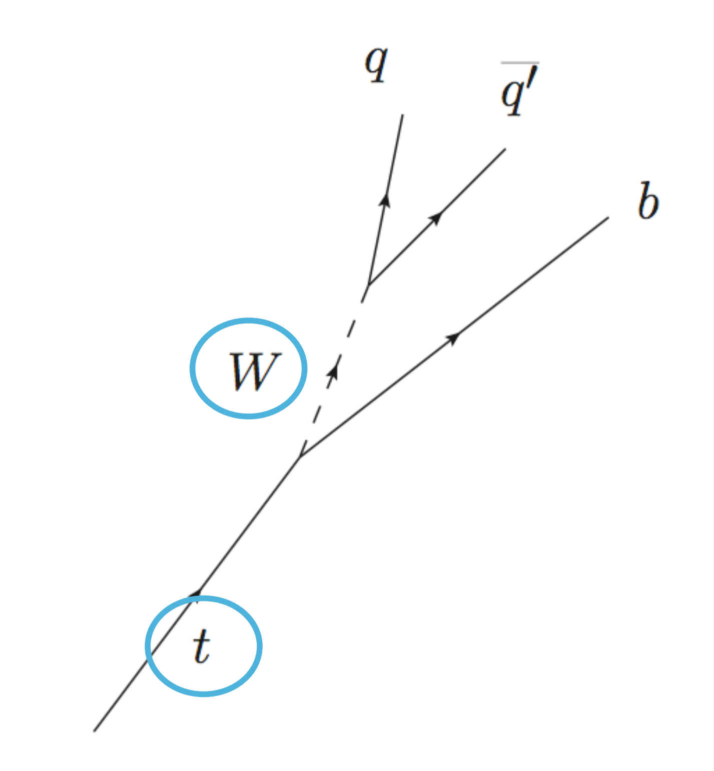
- Discriminate **top jets** against **non-top jets**
  - Top mass
  - W mass
  - ~~b tagging~~

Tracking unreliable
in boosted top due to collimation
of particles



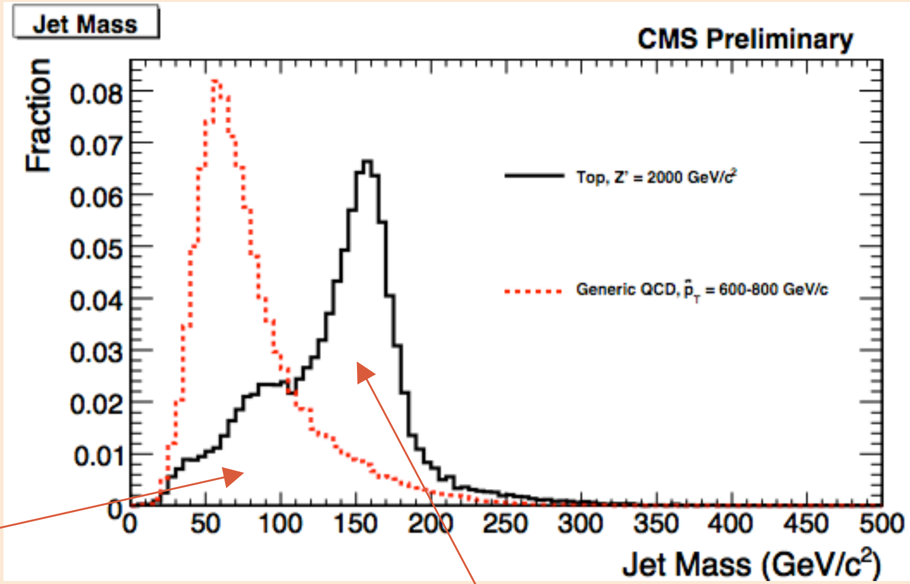
Top Tagging

- Discriminate **top jets** against **non-top jets**
  - Top mass
  - W mass
  - b-tagging



Top Tagging

- Discriminate **top jets** against **non-top jets**
  - **Top mass ~ jet mass**
  - W mass



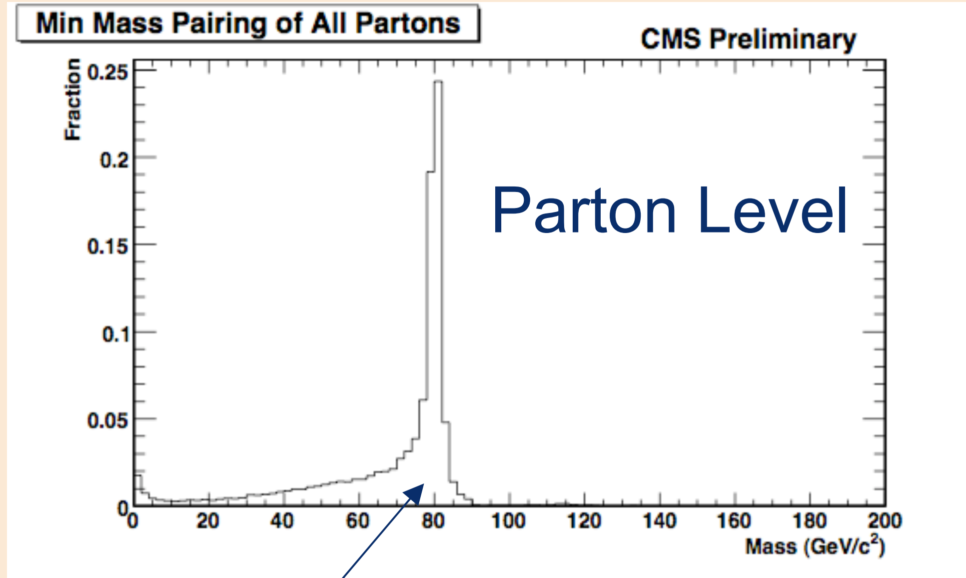
Partially merged jets

Fully Merged jets

Jet mass good approximation for top mass

Top Tagging

- Discriminate **top jets** against **non-top jets**
  - Top mass
  - **W mass ~ min di-subjet mass**



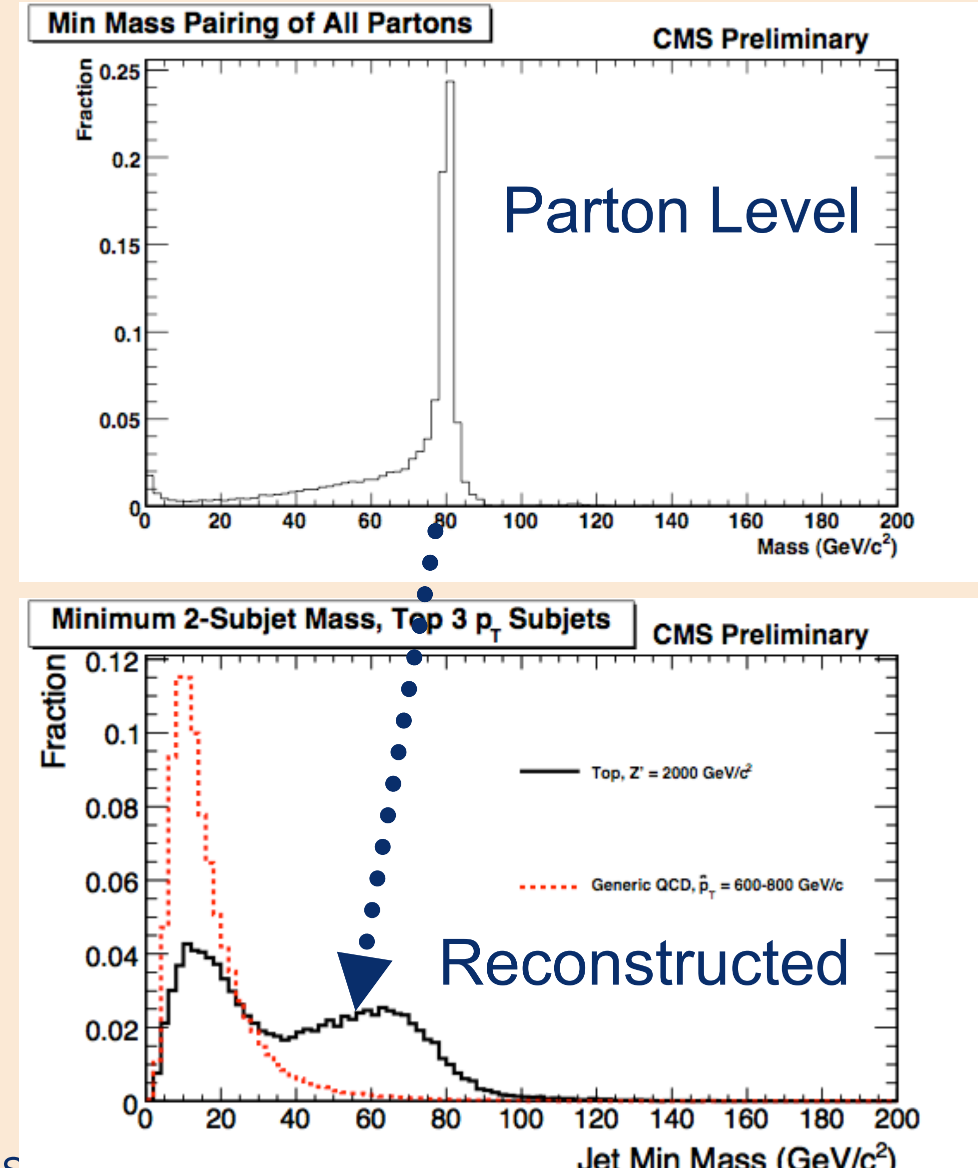
bq
bq'
qq'

Take minimum mass pairing

Top Tagging

- Discriminate **top jets** against **non-top jets**
  - Top mass
  - **W mass ~ min di-subjet mass**

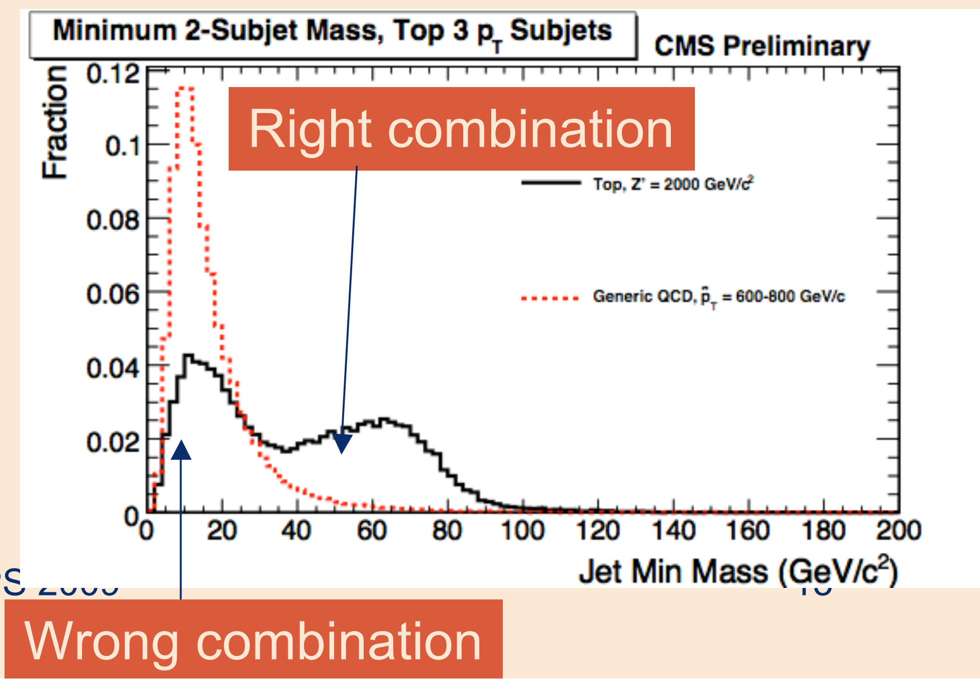
bq → j1+j2
bq' → j1+j3
qq' → j1+j3



Top Tagging

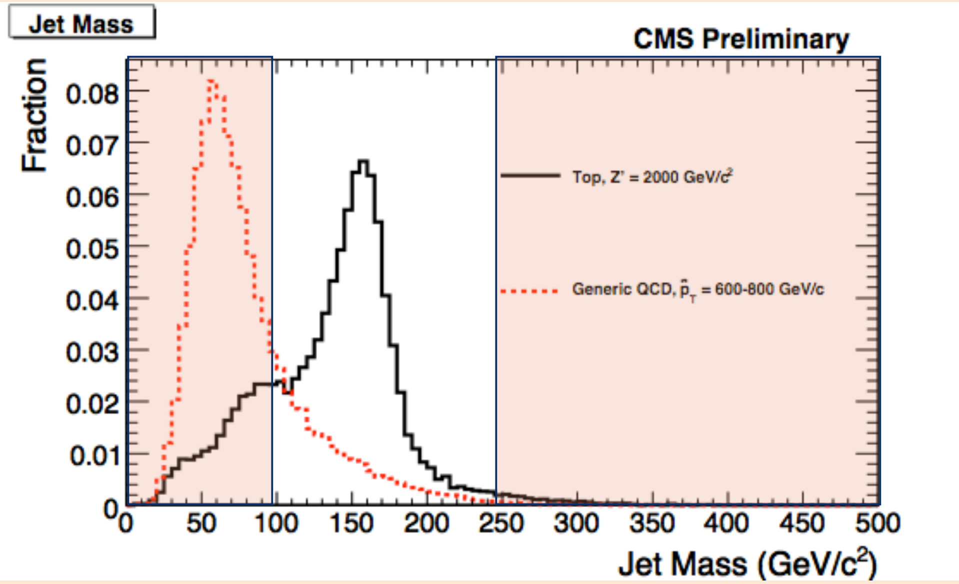
- Discriminate **top jets** against **non-top jets**
  - Top mass
  - **W mass ~ min di-subjet mass**

After reconstruction, still often gets W mass



Top Tagging

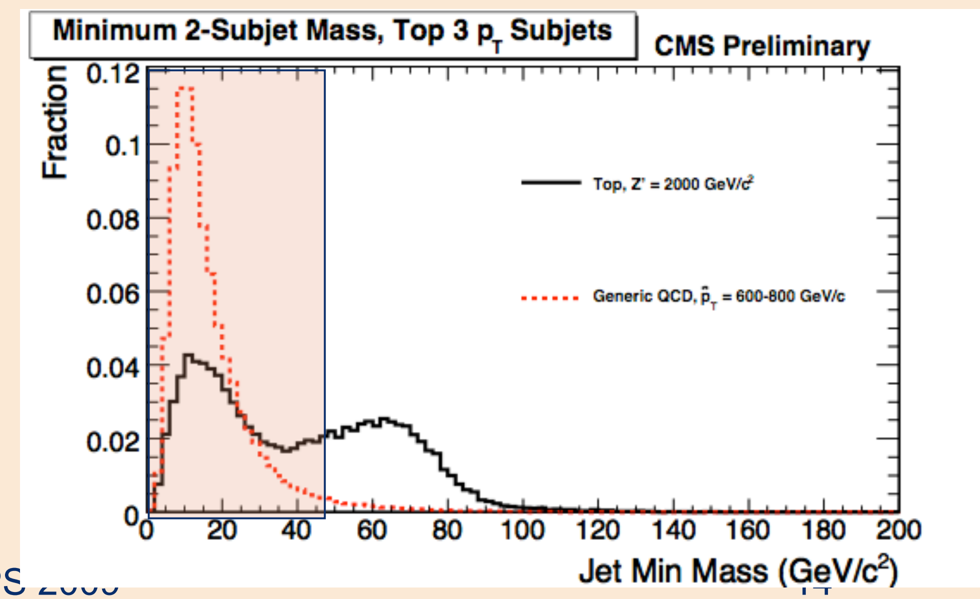
- Discriminate **top jets** against **non-top jets**
  - Top mass
  - W mass



Require

$100 < m_{jet} < 250$ GeV/c^2

50 GeV/c^2 $< m_{min}$



Efficiency Estimate

- Efficiency derived from MC
- Systematic uncertainties will include:
  - Theoretical uncertainties
  - Smearing detector-based resolutions

Efficiency Estimate

Theory

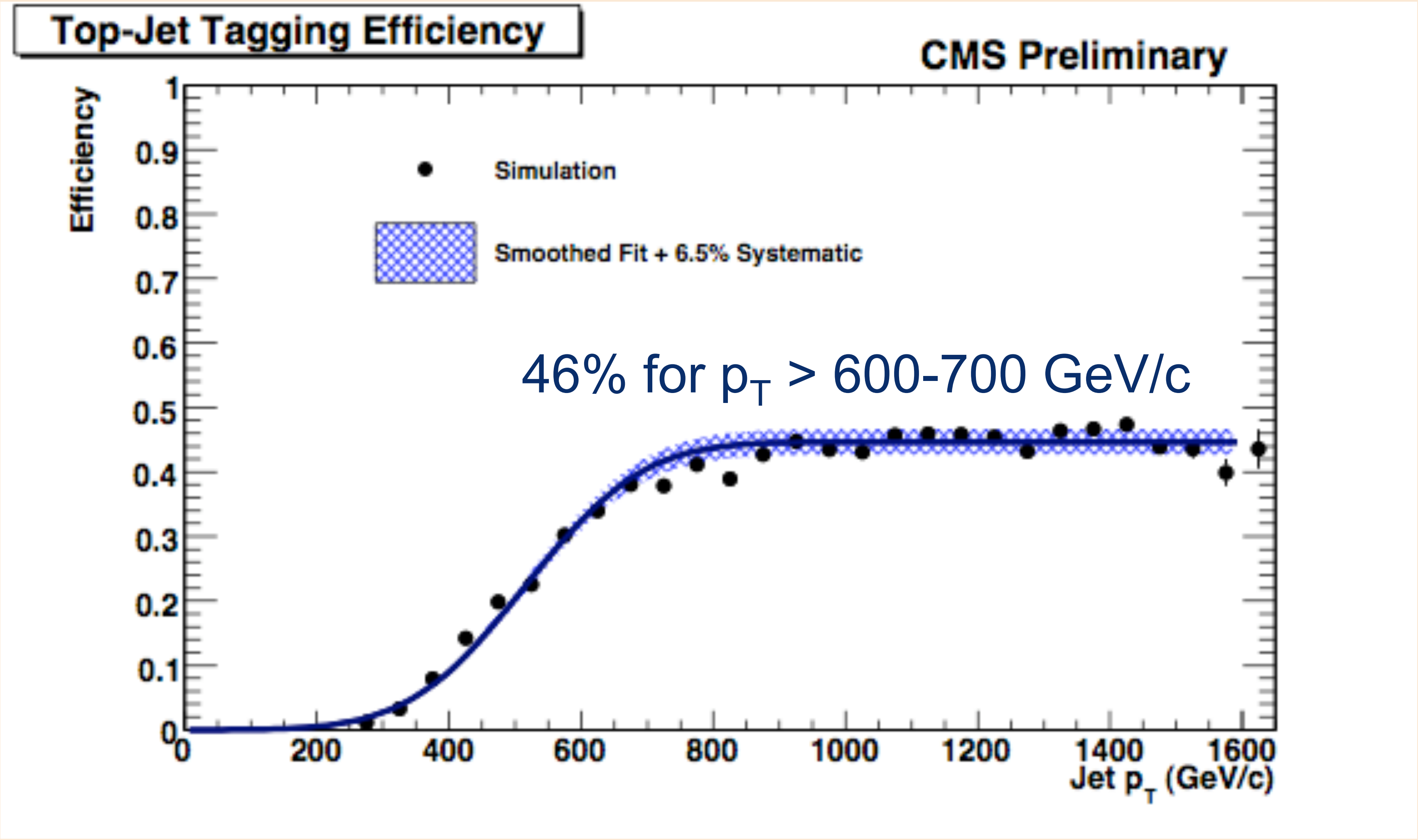
- Heavy quark fragmentation
- Light quark fragmentation
- ISR/FSR
- Λ_{QCD}

Effect	Systematic Uncertainty (%)
Initial State Radiation	1
Final State Radiation	2
Renormalization Scale	3
Light Quark Fragmentation	< 1
Heavy Quark Fragmentation	< 1
Theoretical Uncertainty	3.8
Momentum Smearing + 10%	3.3
Azimuthal Smearing + 50%	2.9
Rapidity Smearing + 50%	2.9
Detector-Based Uncertainty	5.3
Total Systematic Uncertainty	6.5

Detector

- Select partons from t->Wb decay
  - b, q, q'
- Compare true value to response in subjet
- Parameterize resolutions with parton p_T
- Smear p_T by 10%
- Smear Y, phi by 50%

Efficiency Estimate

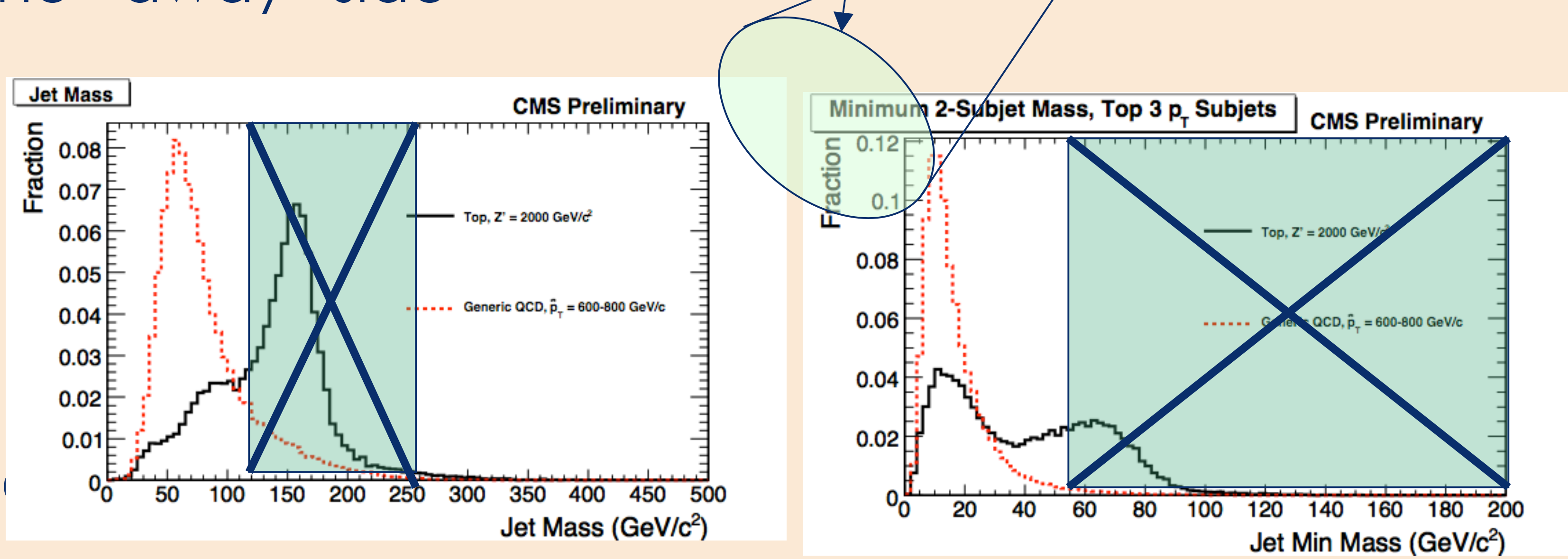


Mistag Parameterization

- **"Anti-tag-and-probe"**
  - Look at "anti-tagged" sample collected from dijet triggers
  - Have a signal-depleted sample on the "away" side

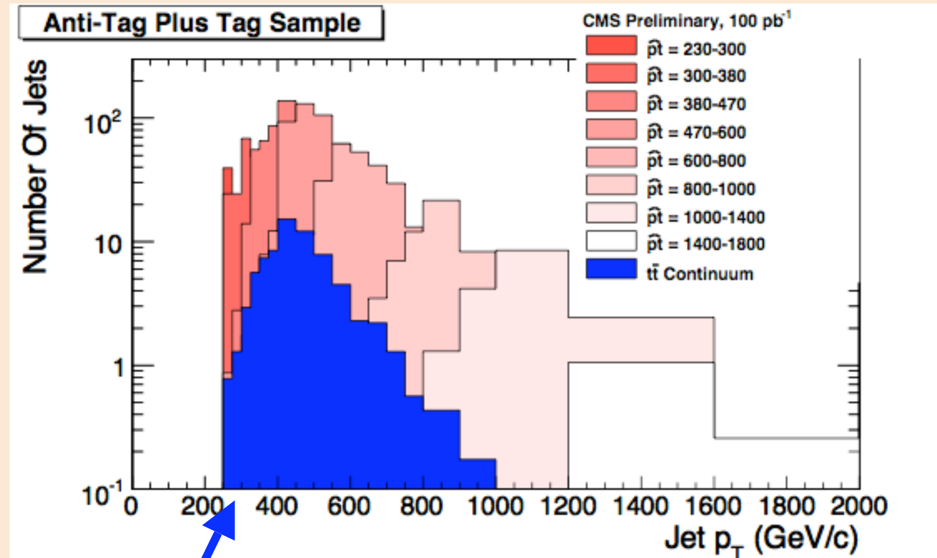
Signal depleted

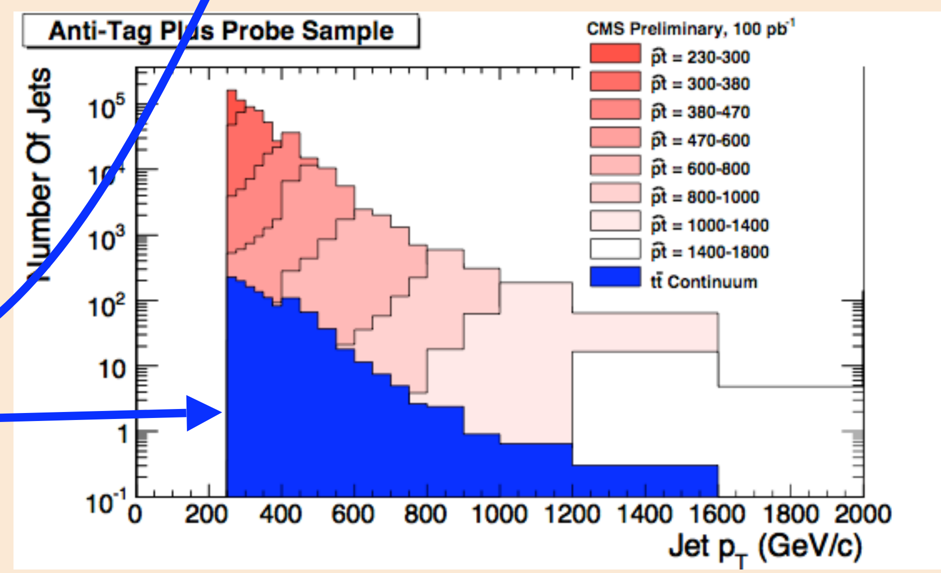
Require veto (anti-tag)



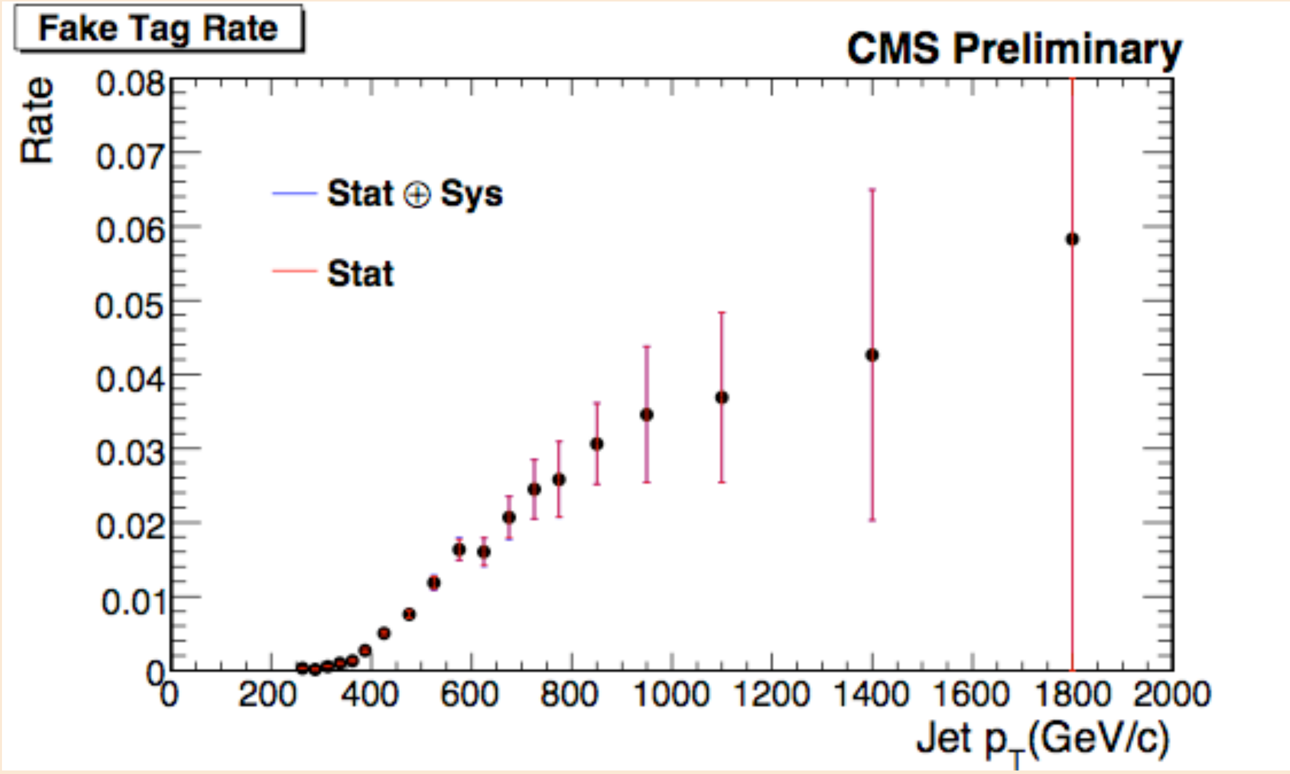
Mistag Parameterization

- **Data-based background estimate**
- Parameterize the background rate with jet p_T
- Numerator: Anti-tag Plus Tag
- Denominator: Anti-tag Plus Probe
- For simulation: scale to 100 pb^{-1}
  - Remove ttbar





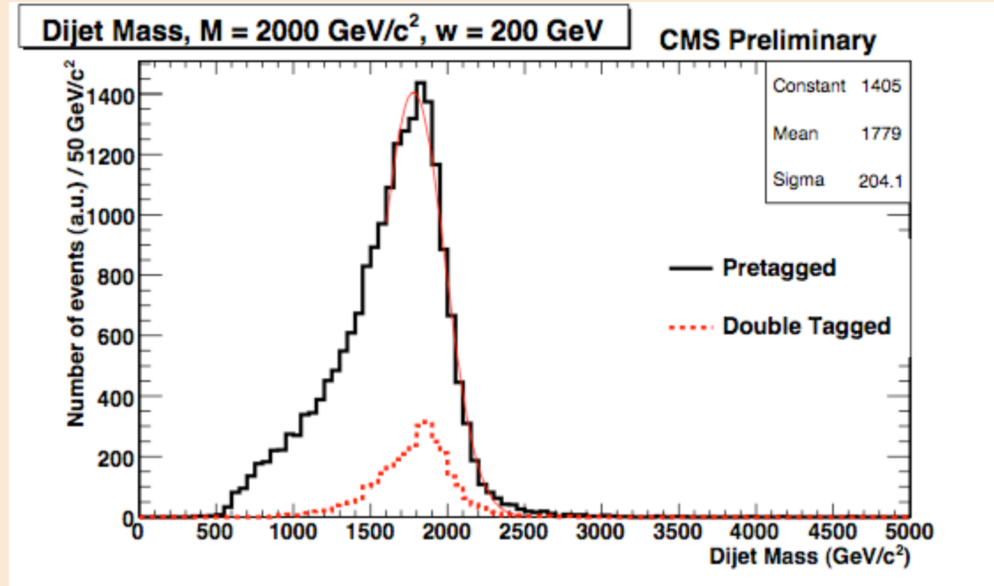
Mistag Parameterization

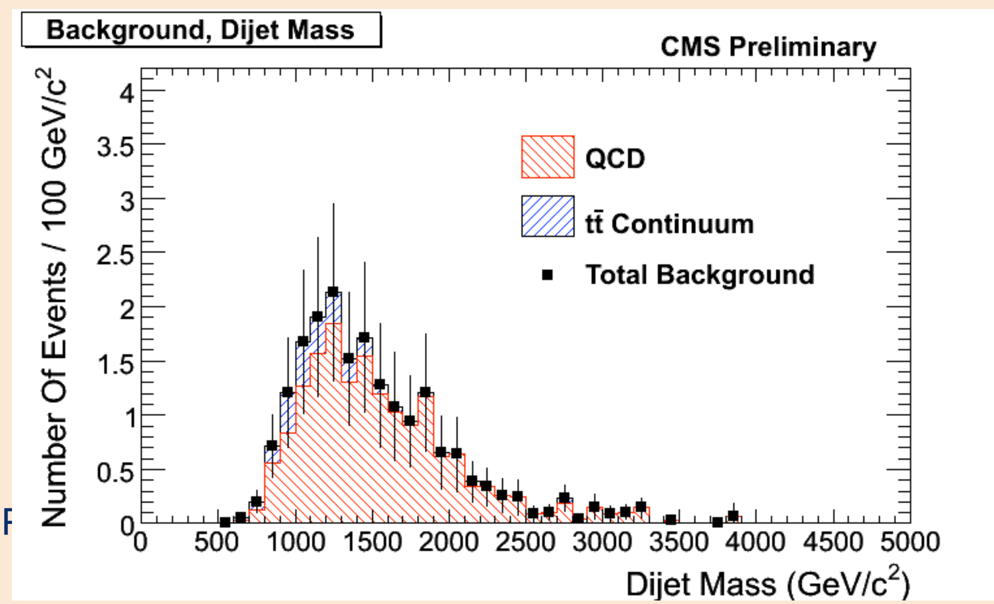


- Total Systematic: 100% of subtracted ttbar
- Statistically dominated
- 98% rejection for p_T = 600 GeV/c

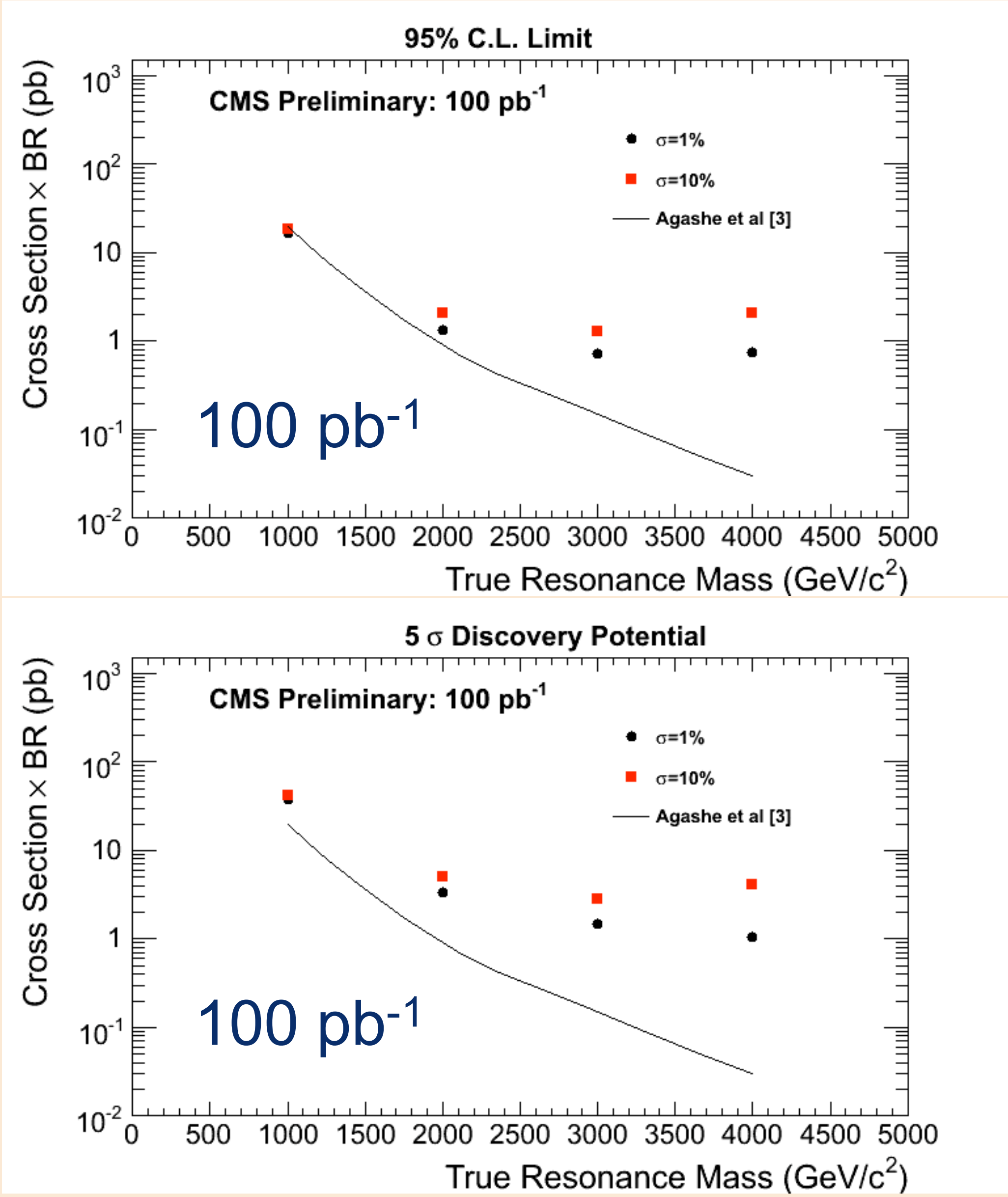
Application: Dijet Search

- CMS PAS EXO-09-002
- Examine dijet search for resonances decaying to ttbar in hadronic channel
- Simple bump-hunt
- Signal from MC
- Background:
  - QCD dijets **(red)** : data-driven.
  - Ttbar **(blue)** : from MC





Application: Dijet Search



With ~200 pb^{-1} can begin to probe realistic new physics scenarios giving boosted top

Conclusions

- Have presented the C-A based algorithm for tagging highly boosted top jets
  - Validated C-A
  - Validated the Top Tagging algorithm
- Presented data-driven fake-tag estimate
  - ~98% rejection of non-top jets at 600 GeV/c
- Presented MC-driven efficiency estimate until data-driven approach is possible
  - ~46% efficiency of top jets at 600 GeV/c
- Good sensitivity to popular new physics models