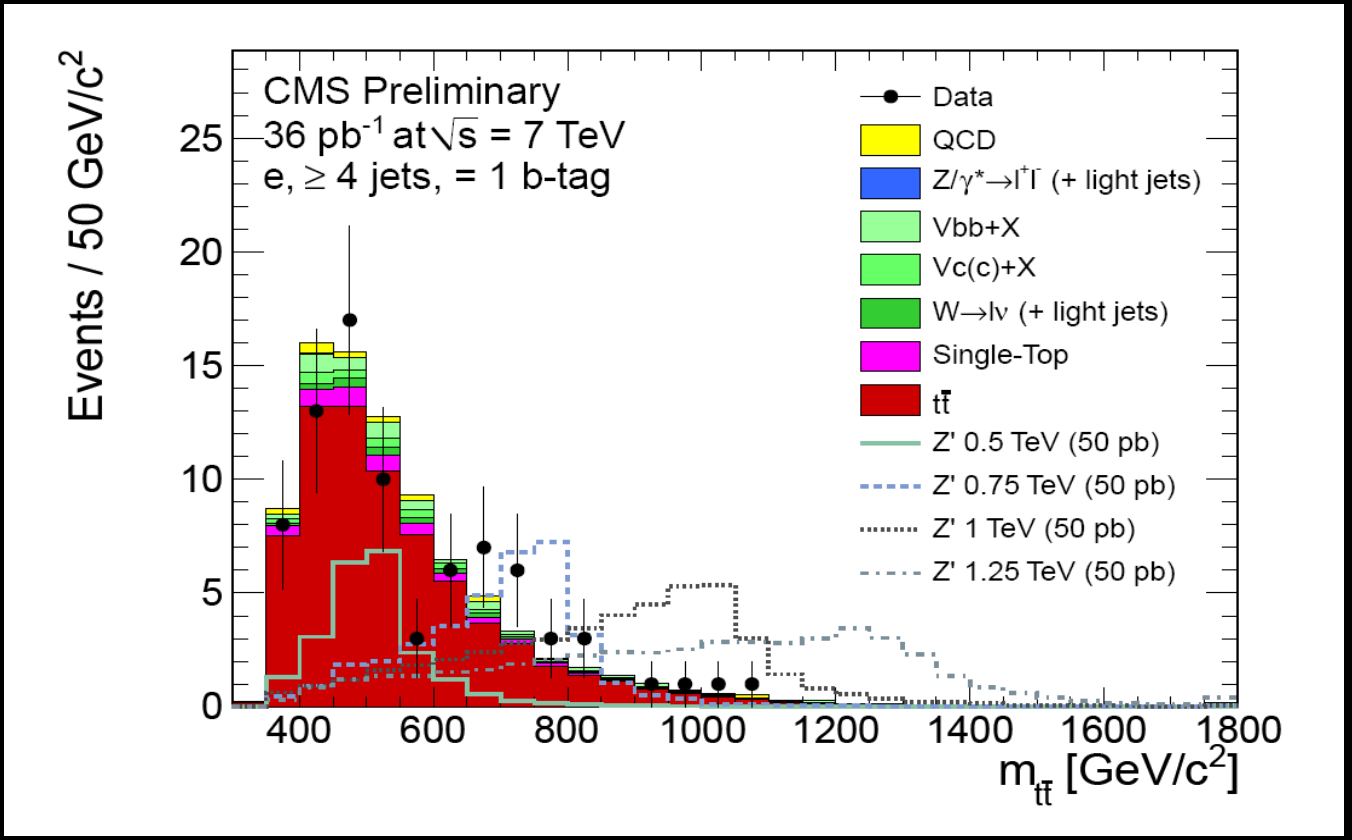
Top Tagging

Brock Tweedie
Boston University
20 March 2011

Boosted Tops are Here

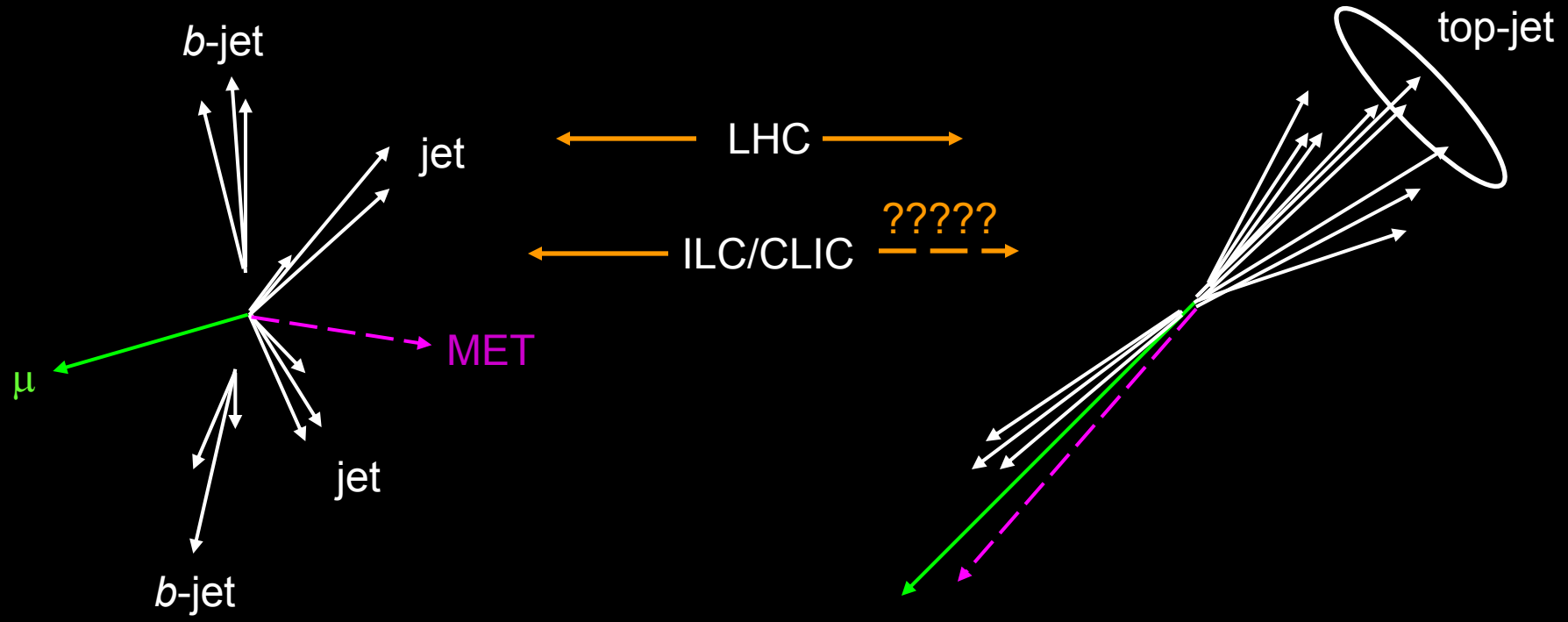


CMS PAS TOP-10-007

Facing the Inevitable: Top-Jets

$E_{CM} \sim 2m_t$

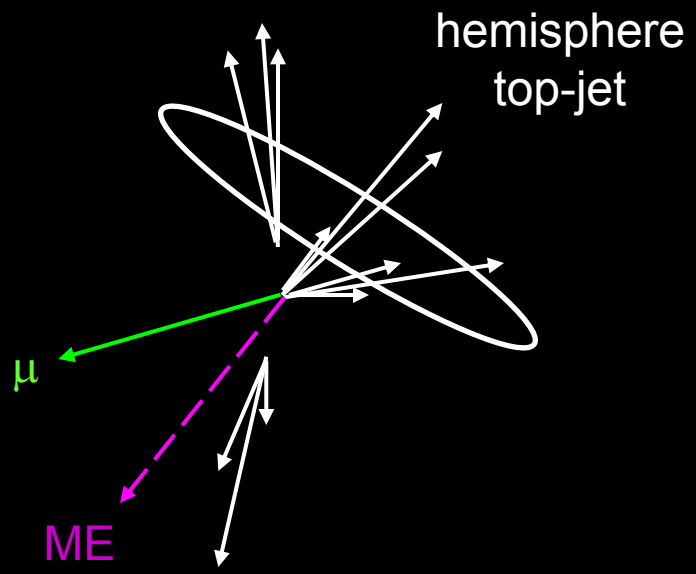
$E_{CM} >> 2m_t$



Sources of Top-Jets

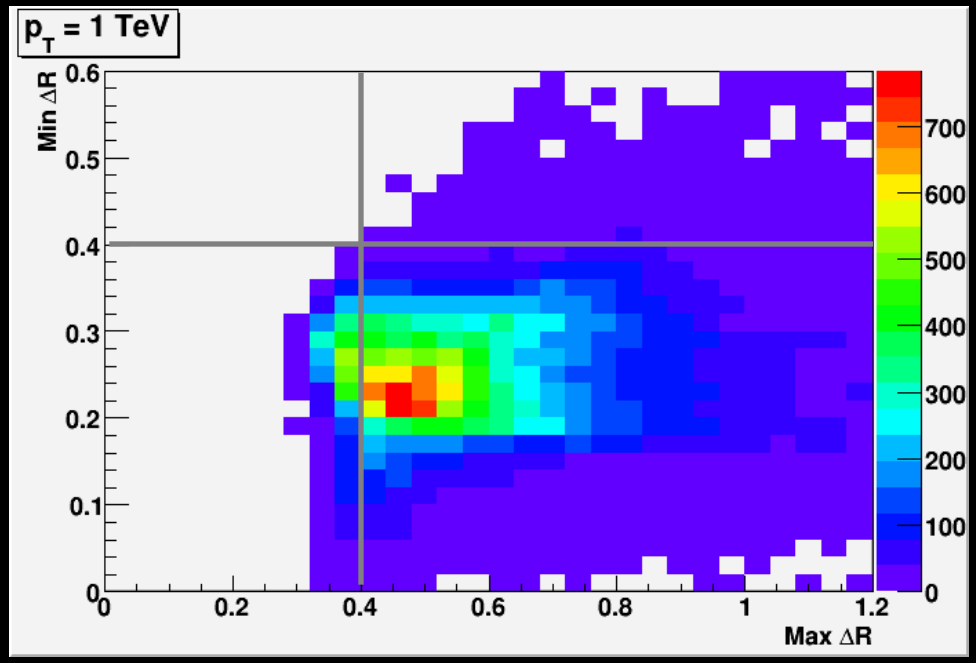
- The Standard Model
- Top resonances
  - Z' -> tt, g' -> tt, G' -> tt
- Top partners / SUSY
  - t' -> th / tZ
  - $\tilde{b}$' -> tW
  - t -> t $\tilde{\chi}^0$
- Top-Higgsstrahlung
  - high-energy tt->tth

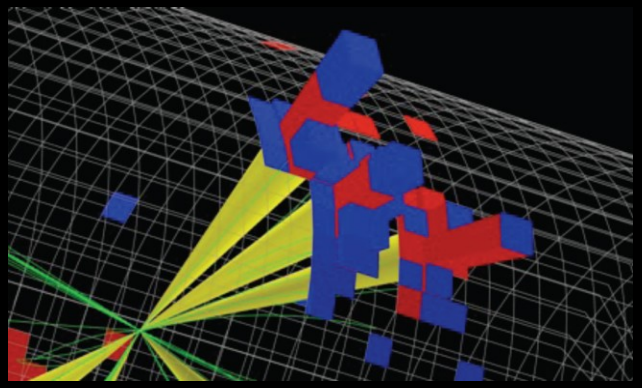
Intermediate-Boost Tops at ILC?



may even be relevant for organizing event information closer to top pair threshold…treat the entire event as one “ditop-jet”

Angular Scales in Top Decay





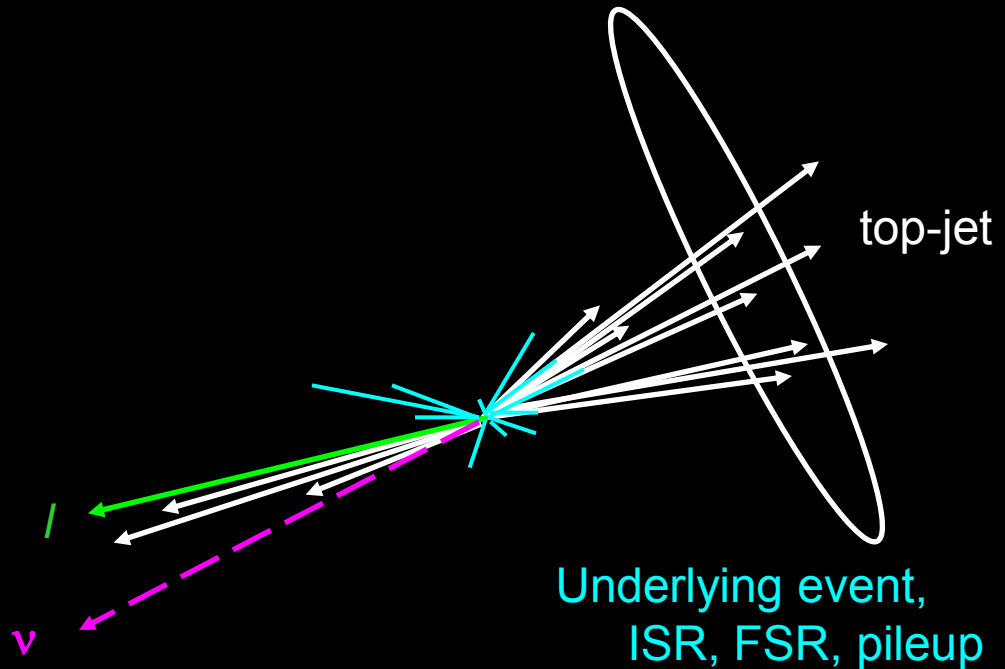
- Typical LHC jet size
  - ΔR ~ 0.4 - 0.7
- HCAL cells
  - ΔR ~ 0.1
- ECAL cells
  - ΔR ~ 0.02
- Tracker
  - ΔR ~ 0.001

So What?

- QCD is our enemy
  - collinear splittings => easy to get several "hard" partons emitted into a small solid angle
  - jet mass ~ sqrt(α_s)*p_T*R ~ 100 GeV

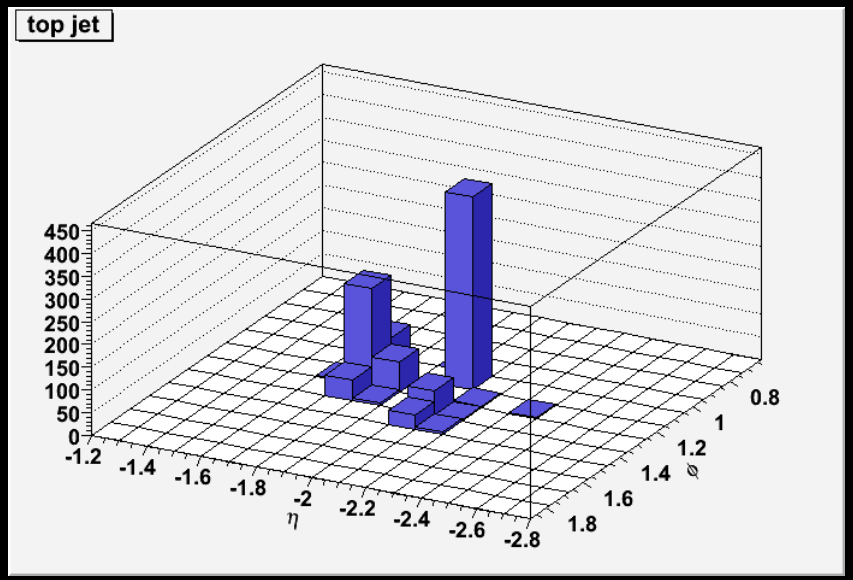
So What?

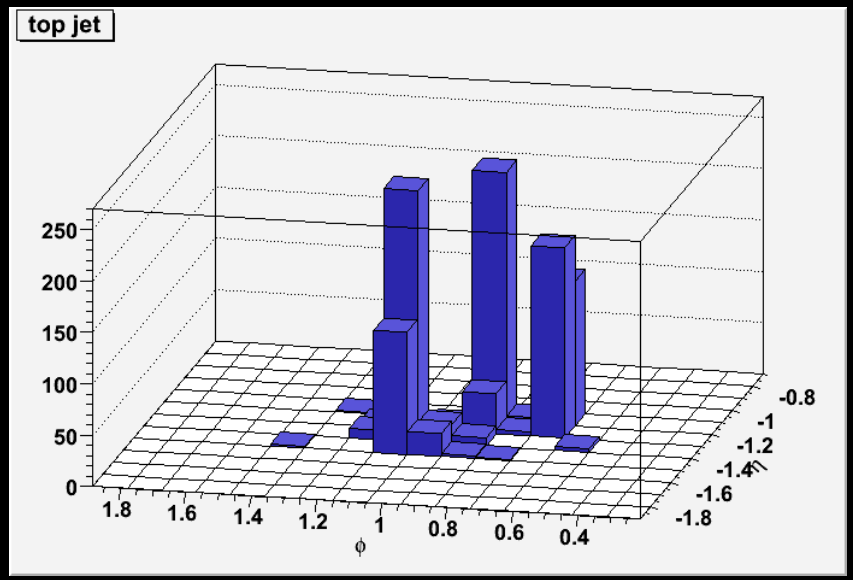
- QCD is our enemy
  - All jets are full of junk!
  - Size too large compared to ΔR => messed-up jet-mass measurements

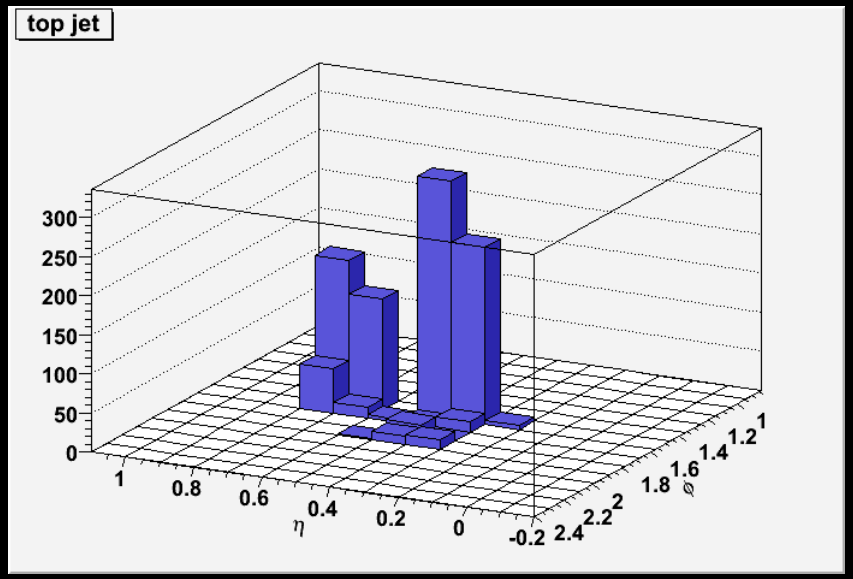


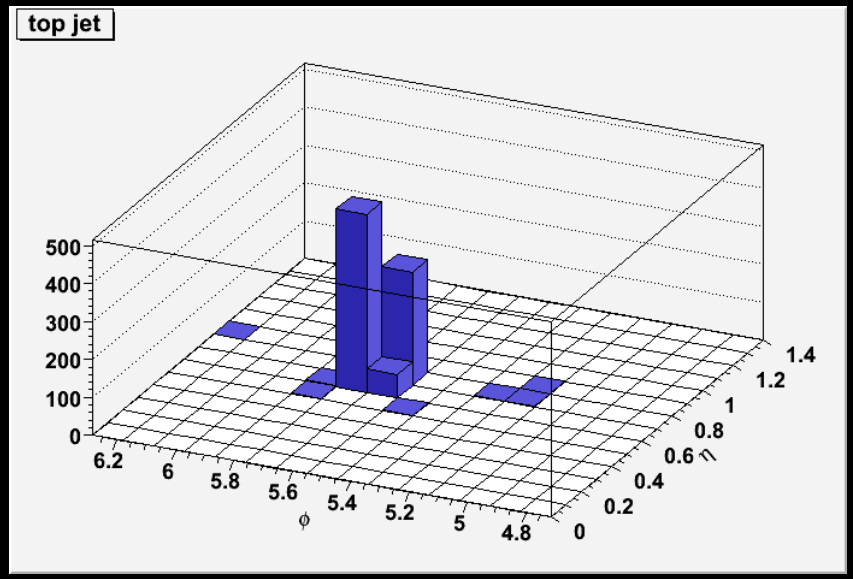
* idealized 0.1 x 0.1 calorimeter

1 TeV Top-Jet Gallery

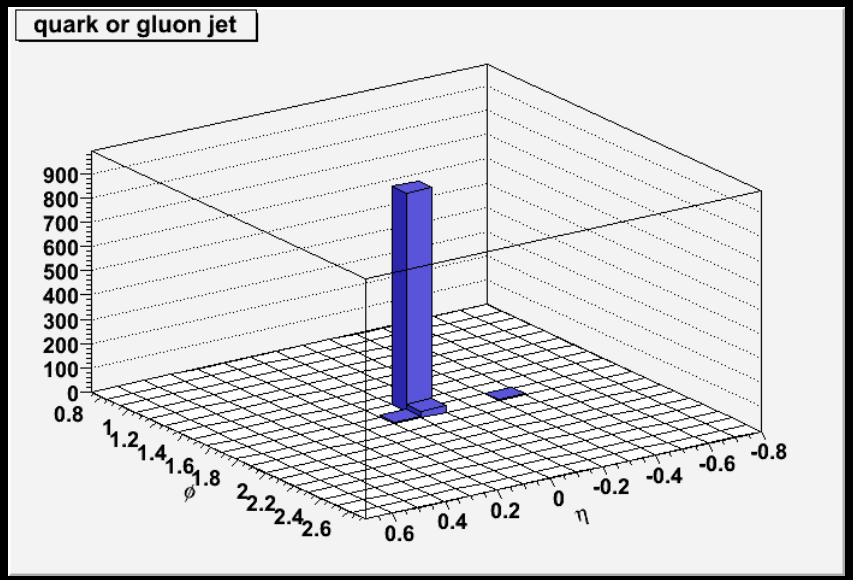


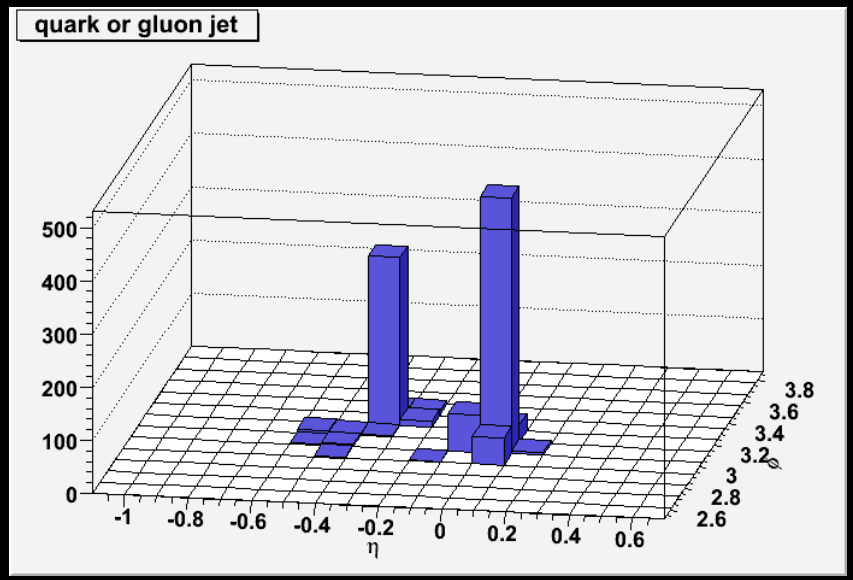


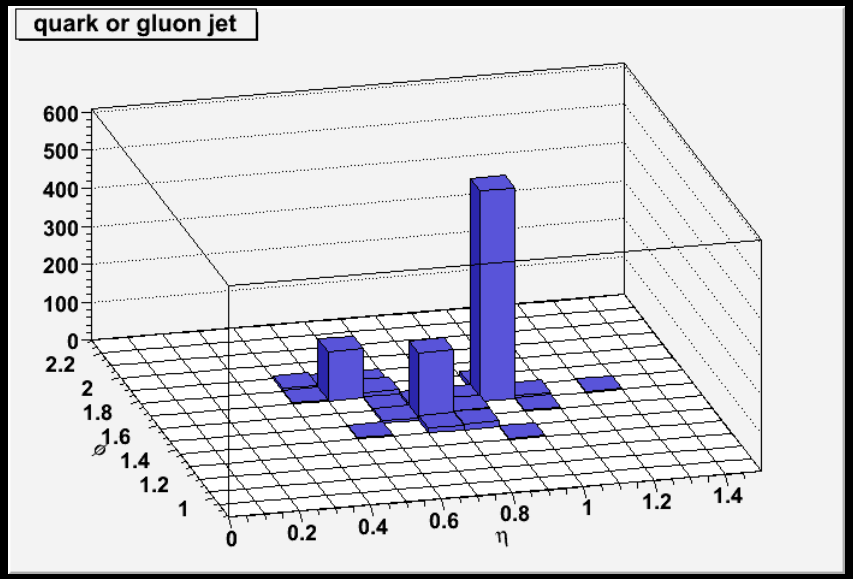


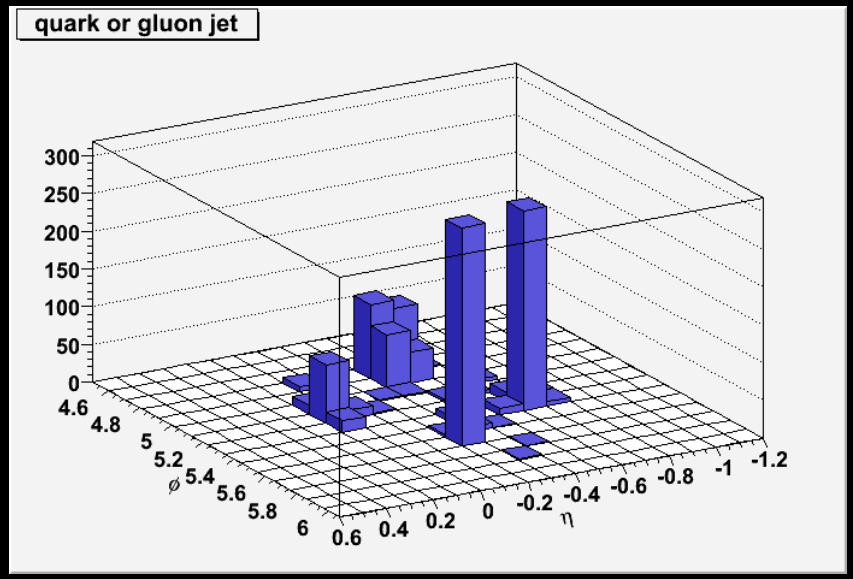


1 TeV QCD-Jet Gallery









Top-Tag Tactics

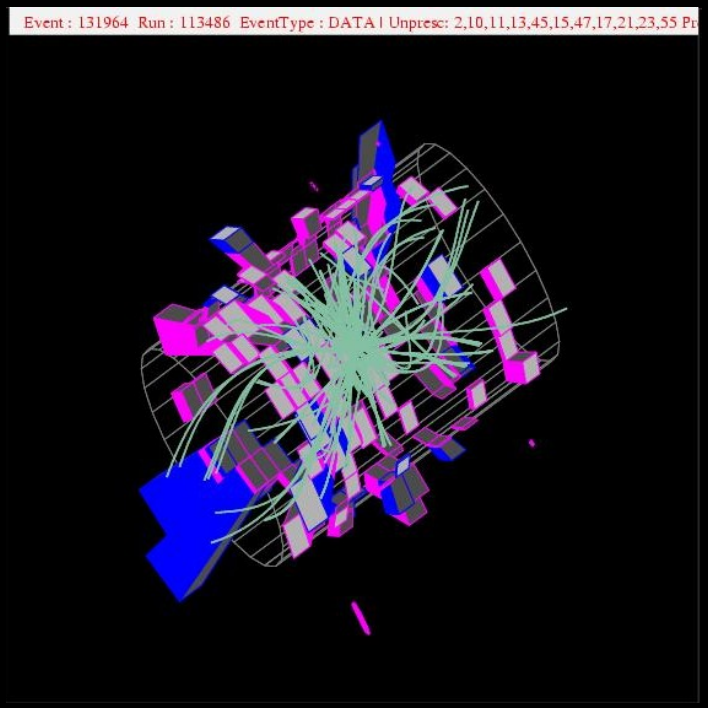
- Y-Splitter: Brooijmans
  - cluster jets with k_T algorithm, decluster 2 or 3 stages and study (dimensionful) splitting scales
  - evolved into the ATLAS top-tagger
- Thaler & Wang
  - cluster jets with anti-k_T algorithm, exclusively recluster with k_T into 3 "subjets" and apply multibody kinematic cuts
- Hopkins: Kaplan, Rehermann, Schwartz, Tweedie
  - cluster jets with Cambridge/Aachen algorithm, decluster recursively until 3 or 4 subjets are found and apply multibody kinematic cuts
  - evolved into the CMS top-tagger
- Jet Shapes: Almeida, Lee, Perez, Sterman, Sung, Virzi
  - angularities, planar flow, etc
- Pruning: Ellis, Vermilion, Walsh
  - selective jet clustering removes junk and self-organizes substructure simultaneously
- HEP Tagger: Plehn, Spannowsky, Takeuchi, Zerwas
  - decluster into arbitrary # subjets, sophisticated kinematic discrimination
  - works with for large top-jets with additional activity inside
- Template Overlap: Almeida, Lee, Perez, Sterman, Sung
  - calorimeter cell pattern -> multidimensional vector
  - check dot products with ensembles of template top-jets and QCD-jets
- N-Subjettiness: Thaler and Van Tilberg
  - continuous scores assigned for mono-subjet-like, di-subjet-like, tri-subjet-like, etc
- Dipolarity: Hook, Jankowiak, Wacker
  - improved discrimination using observables sensitive to color connections

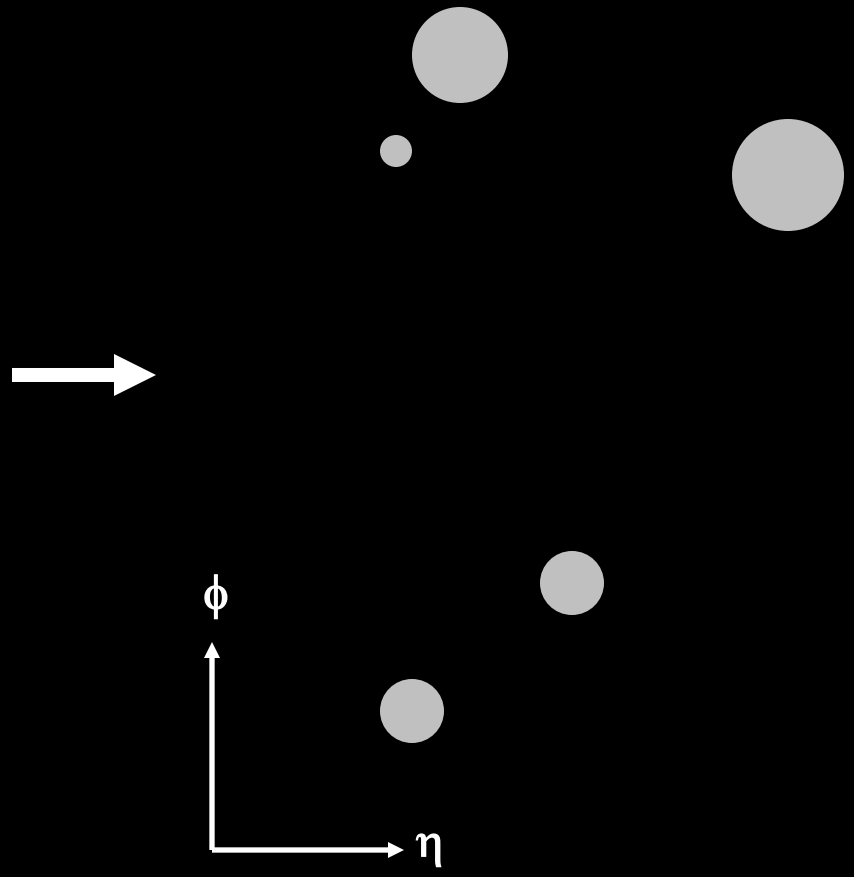
Cambridge/Aachen Jet Clustering

- Sequentially sum up nearest-neighbor 4-vectors in the $\eta-\phi$ plane
- *Stop* when all 4-vectors are distanced by more than a prespecified R

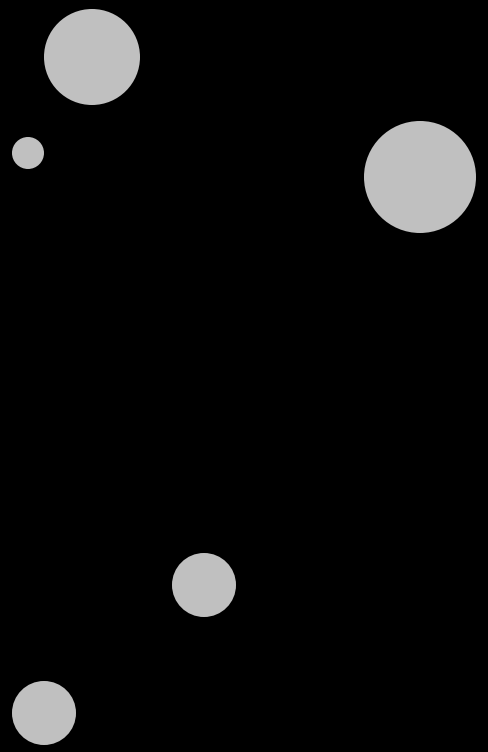
nice code for running this:
http://www.lpthe.jussieu.fr/~salam/fastjet/

Cambridge/Aachen Example

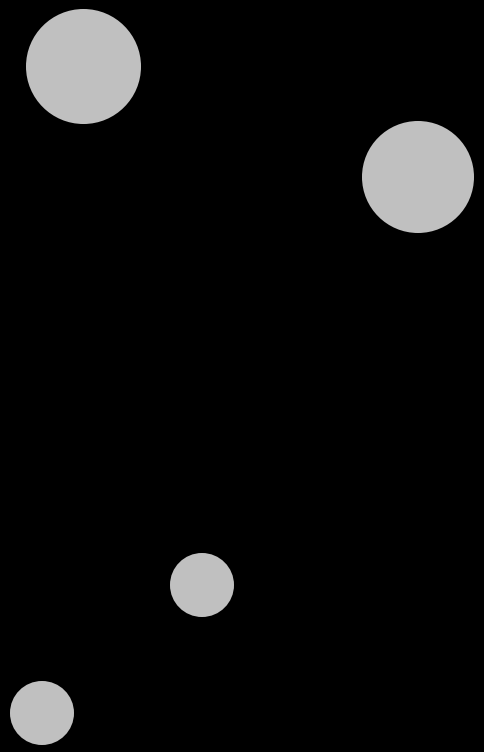




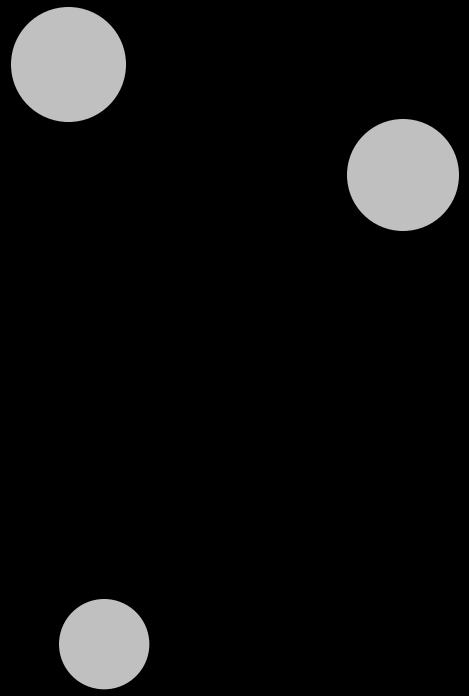
Cambridge/Aachen Example



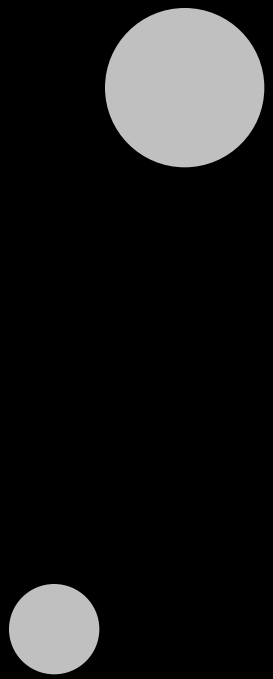
Cambridge/Aachen Example



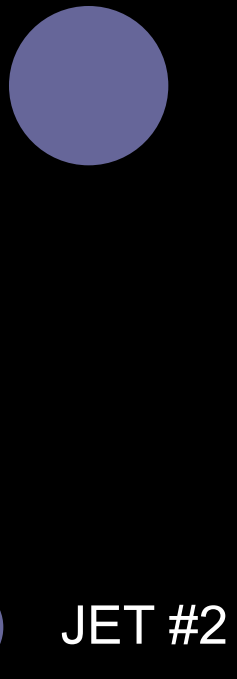
Cambridge/Aachen Example



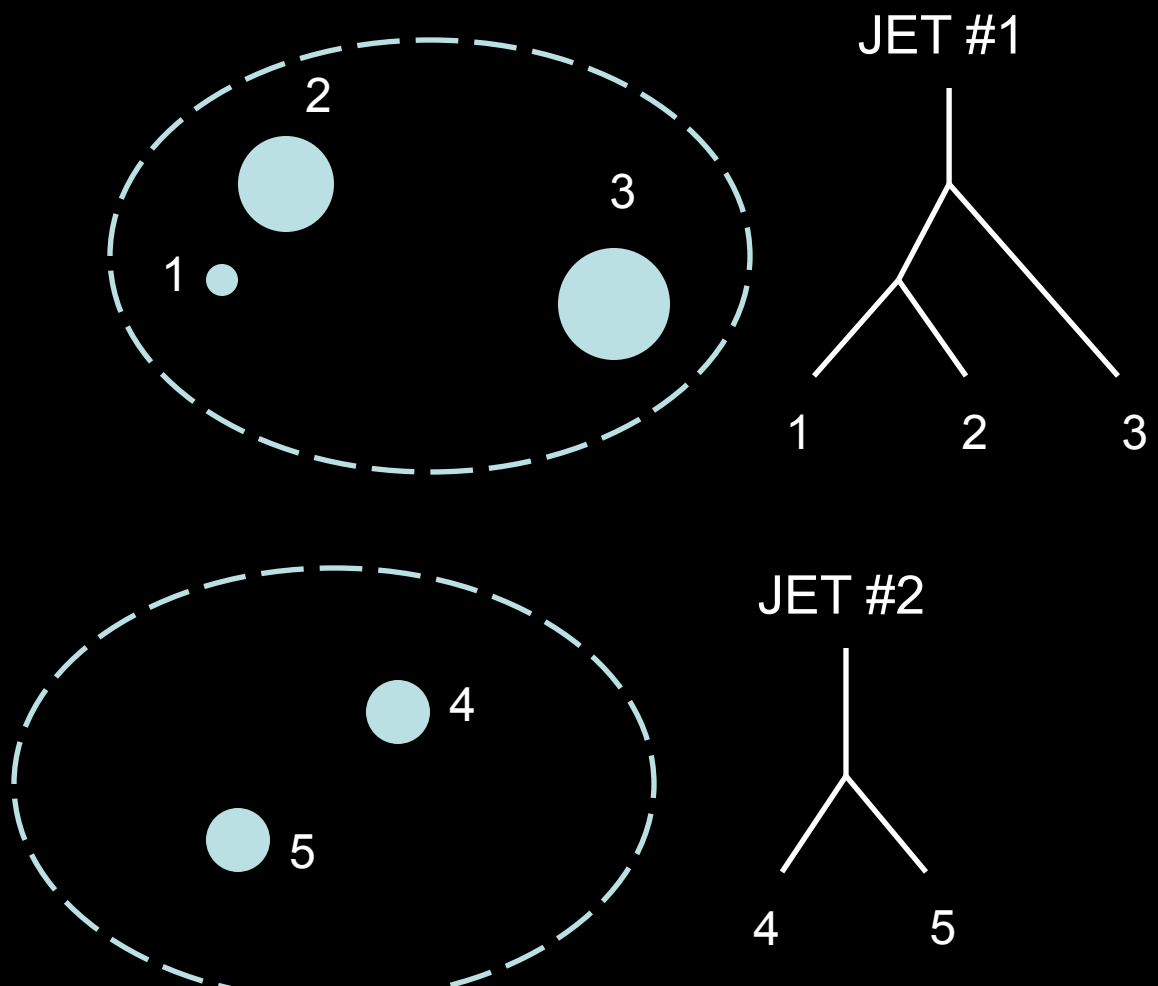
Cambridge/Aachen Example



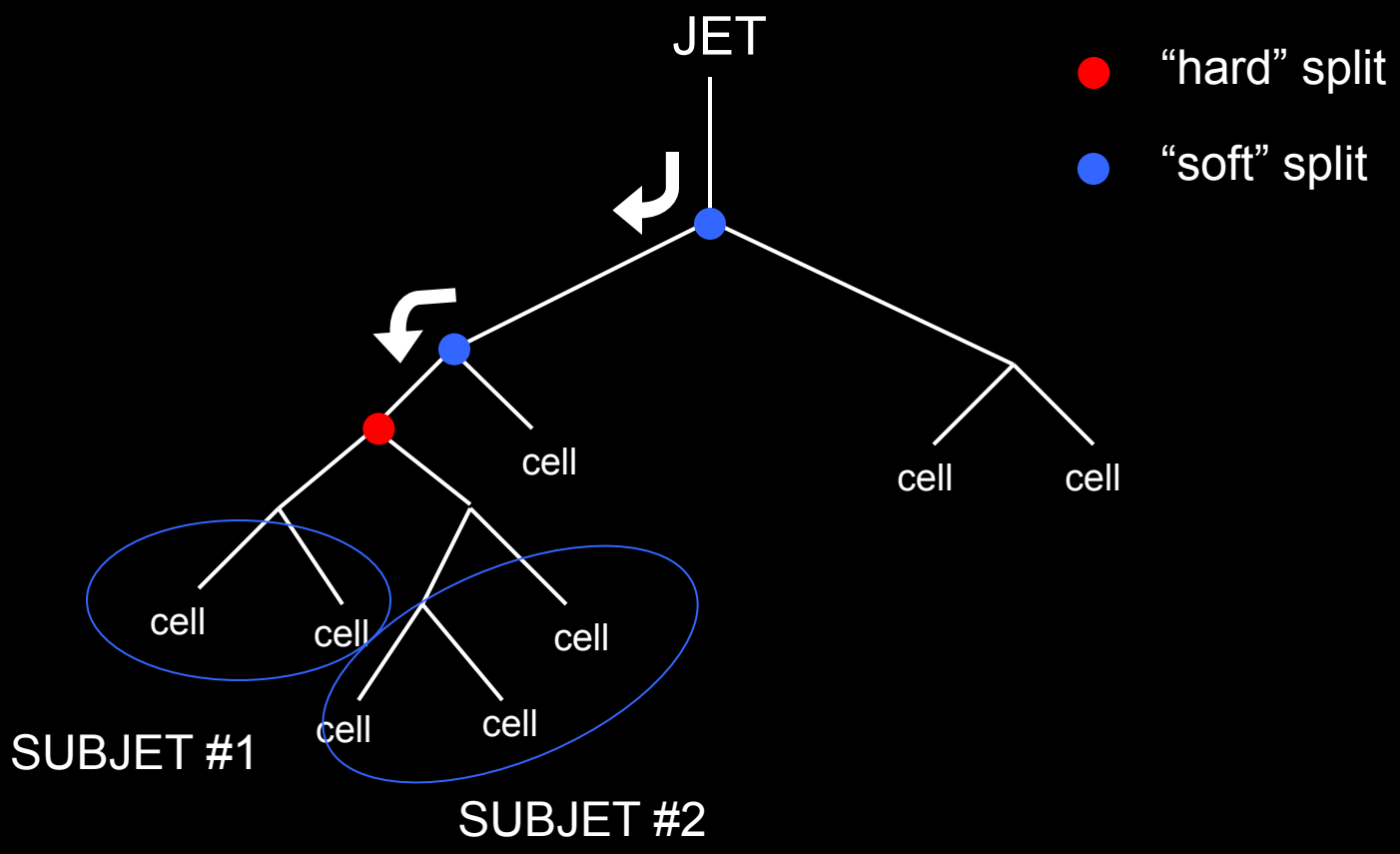
Cambridge/Aachen Example



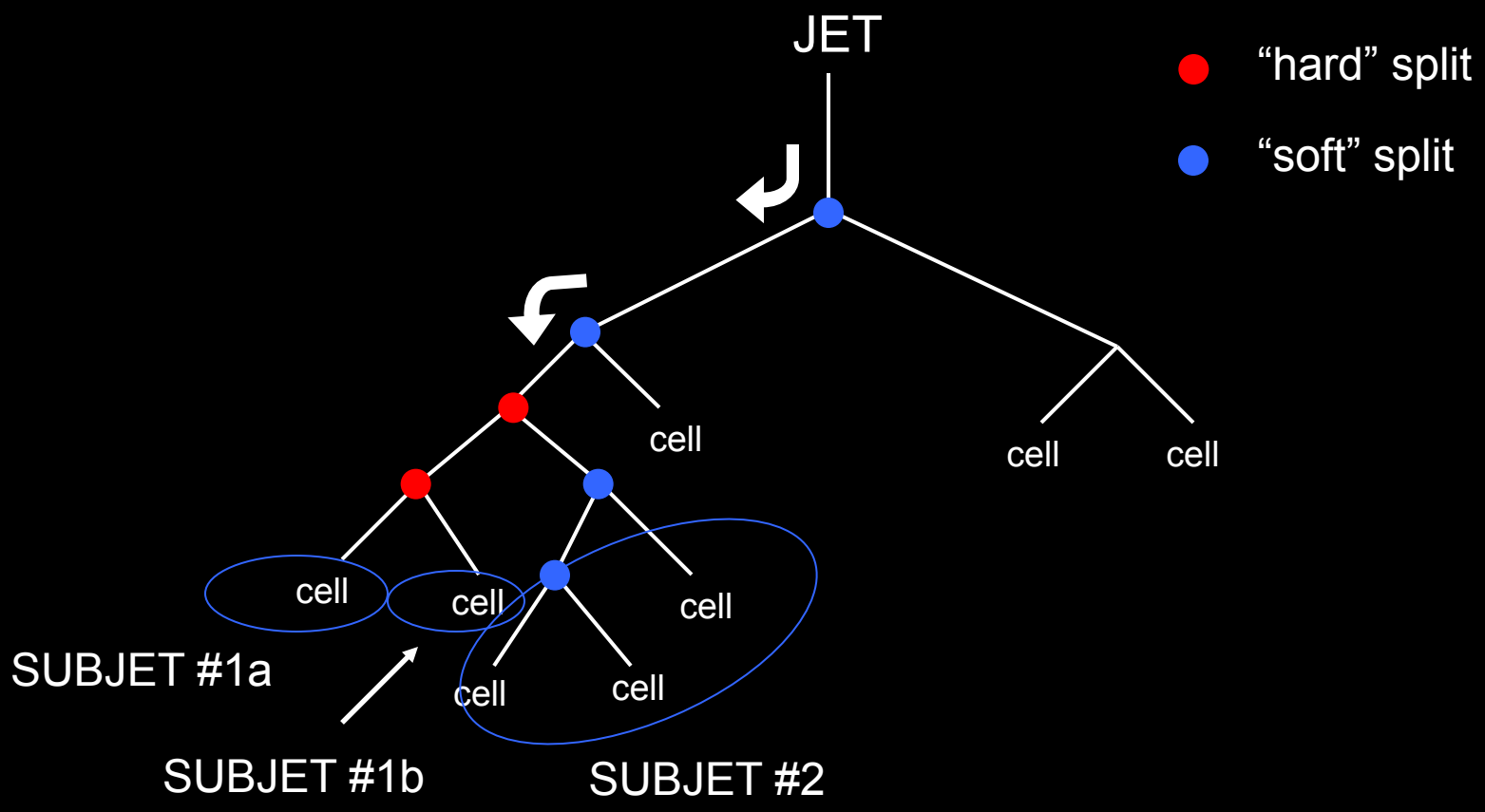
Sequential Clustering History



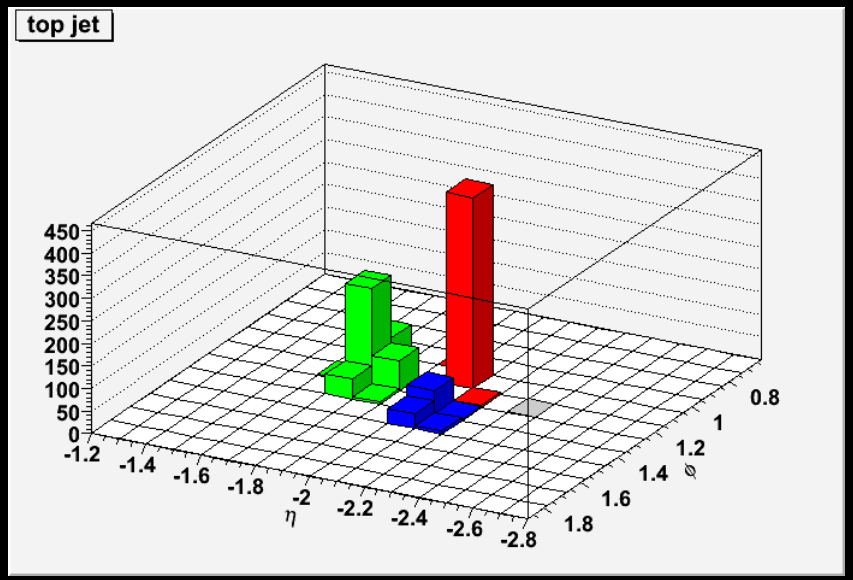
Jet Declustering

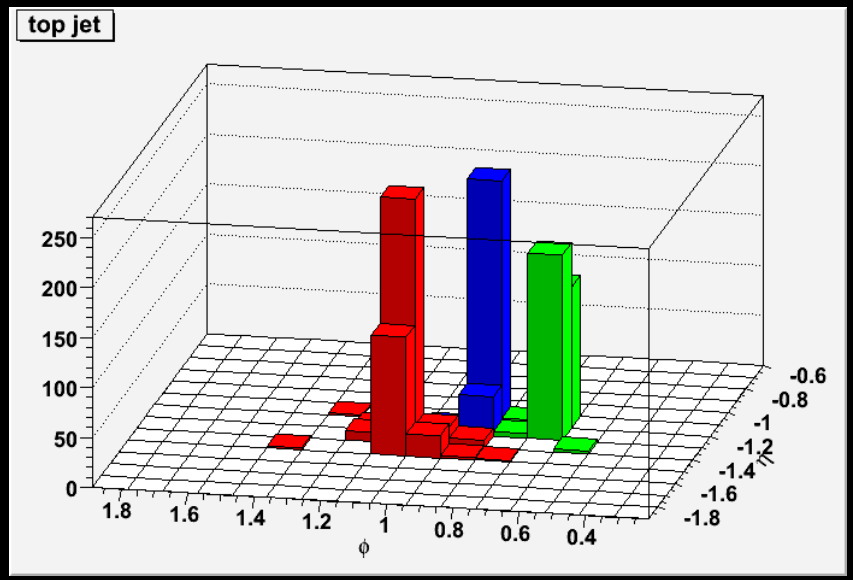


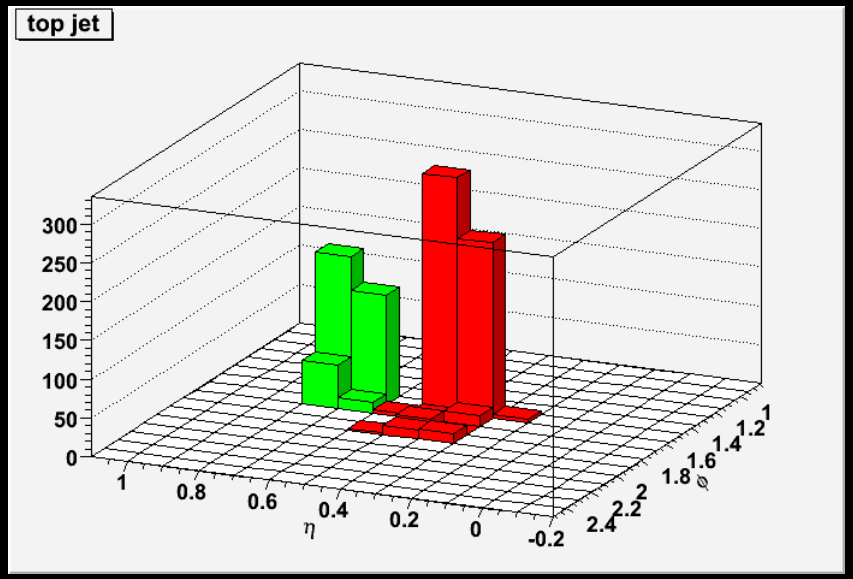
Jet Declustering

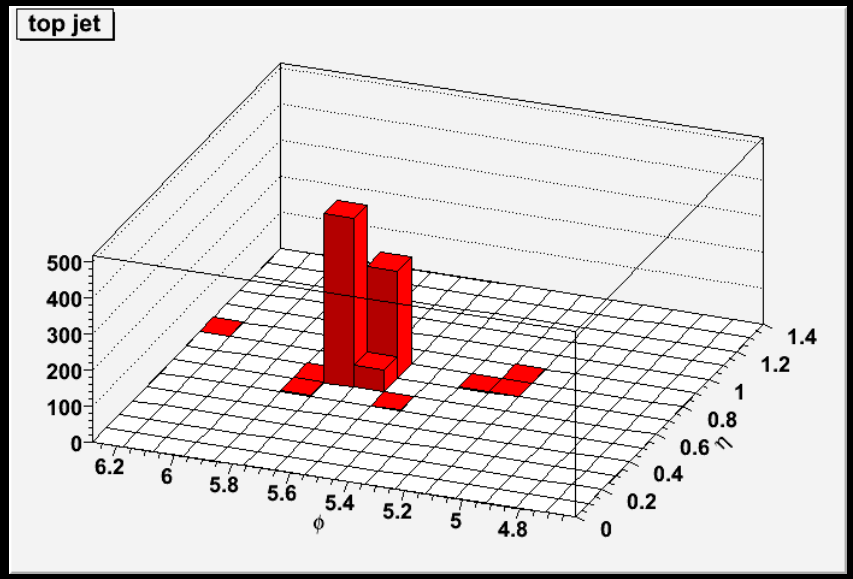


1 TeV Top-Jet Gallery

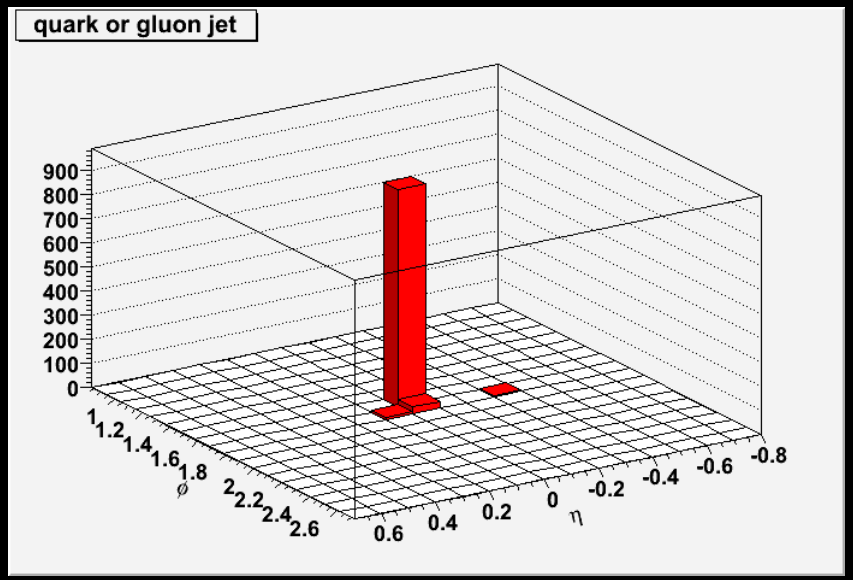


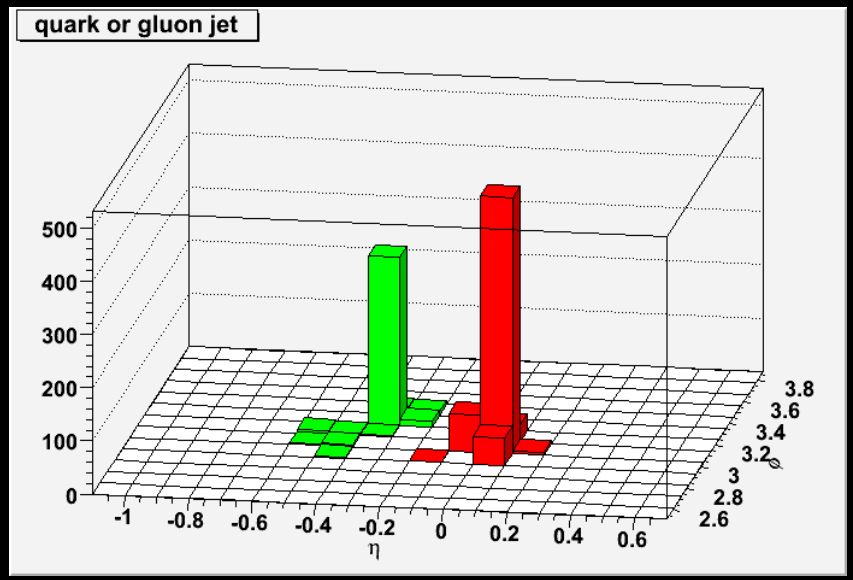


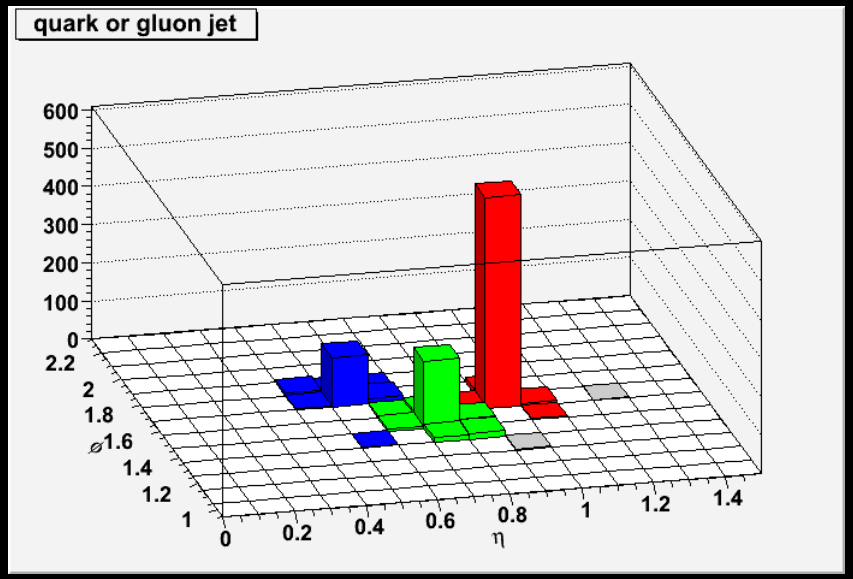


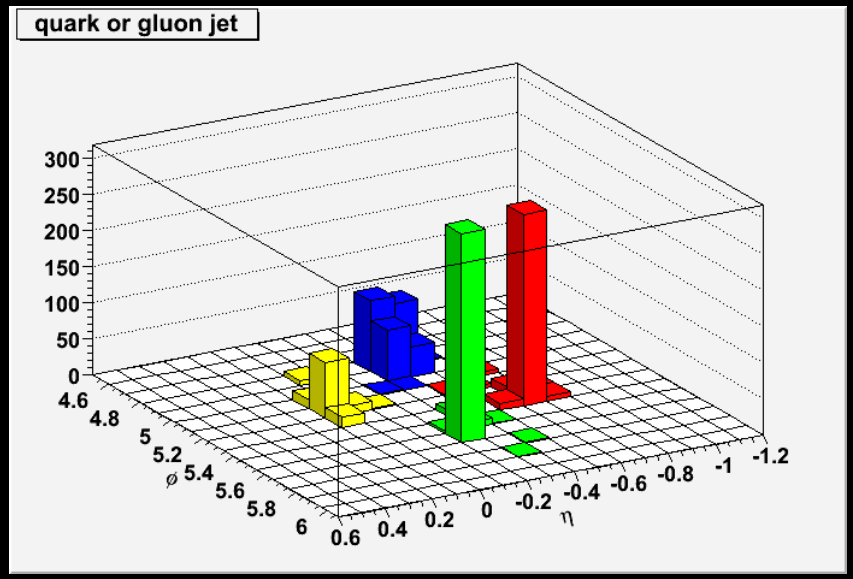


1 TeV QCD-Jet Gallery

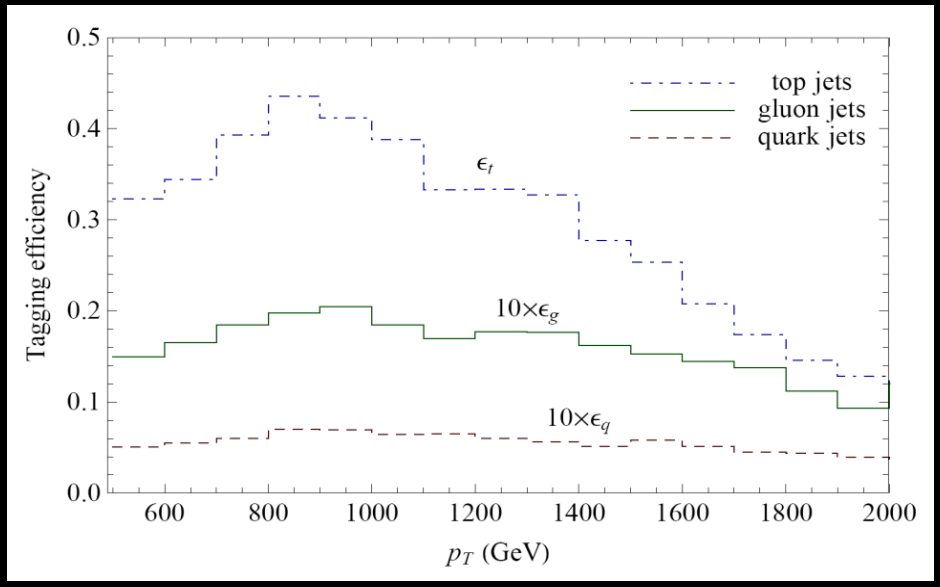




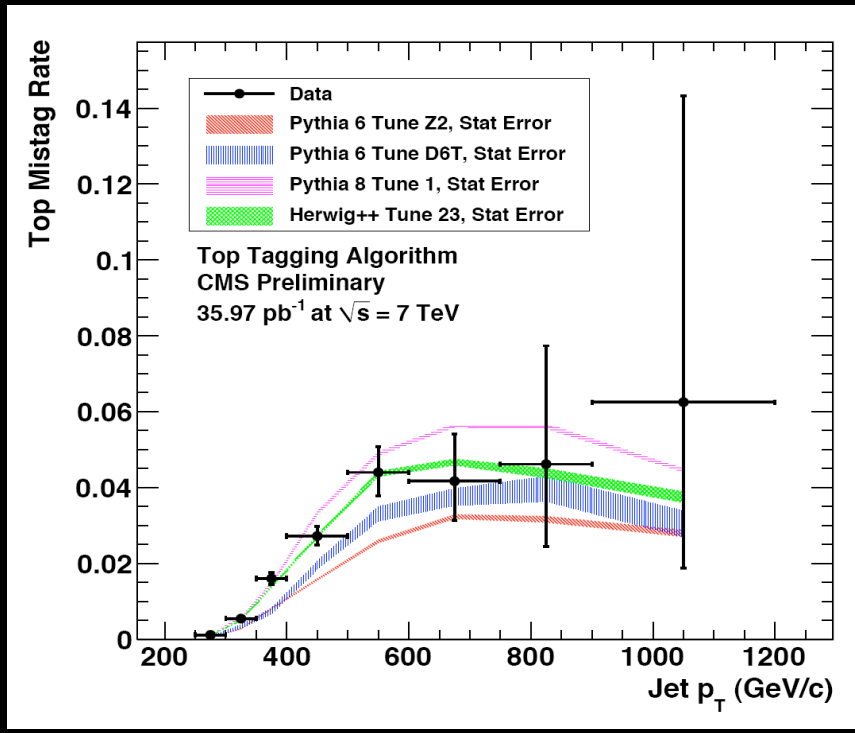




Tag/Mistag Rates

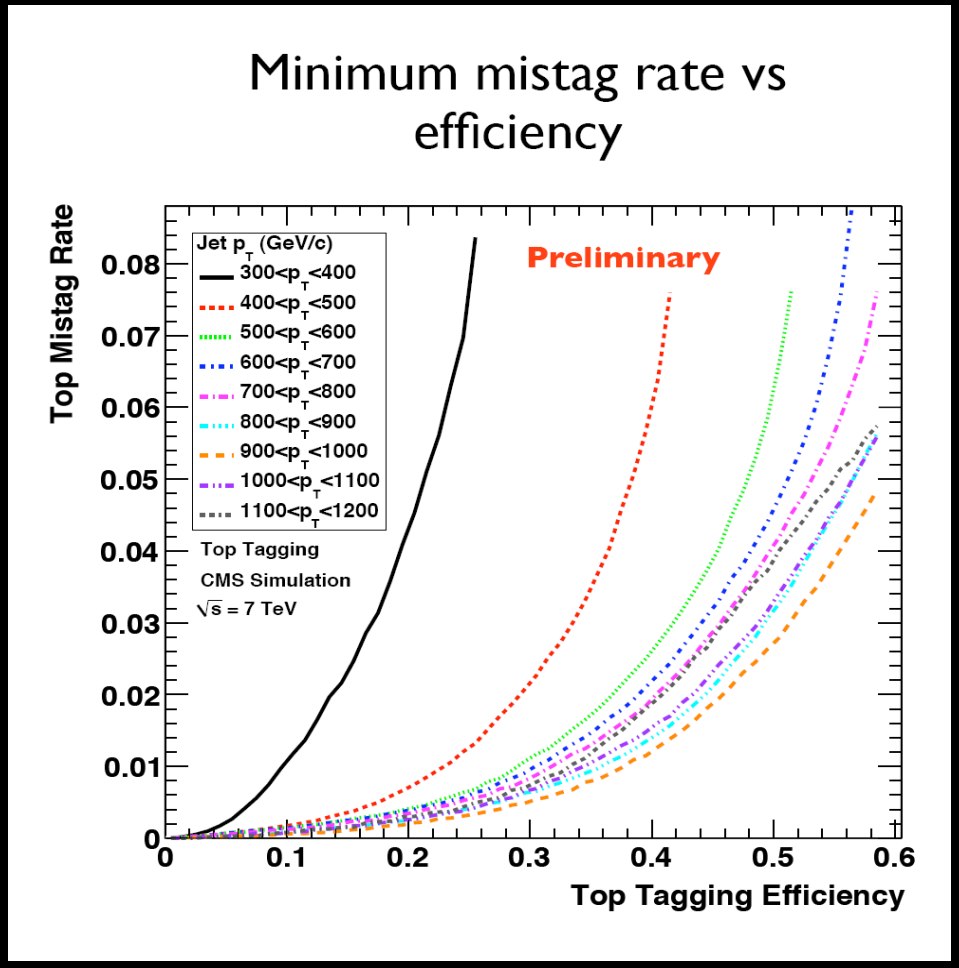


Hopkins top-tagger on our simple simulation



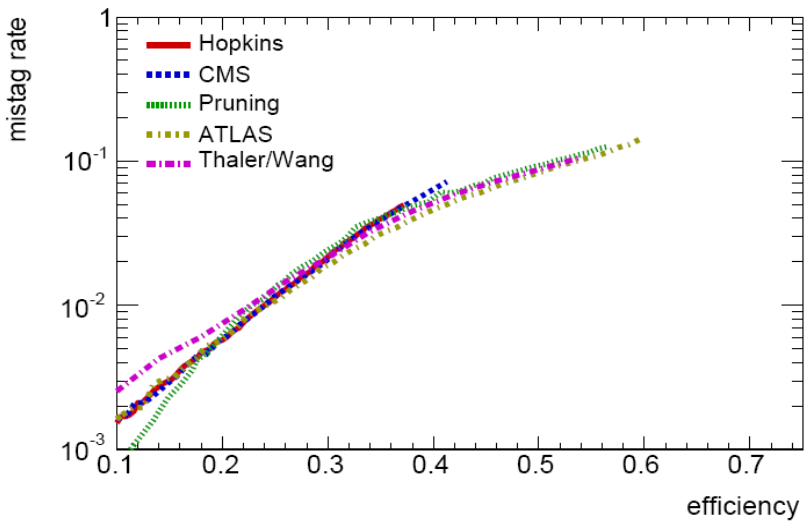
CMS top-tagger on data

CMS Tag vs Mistag

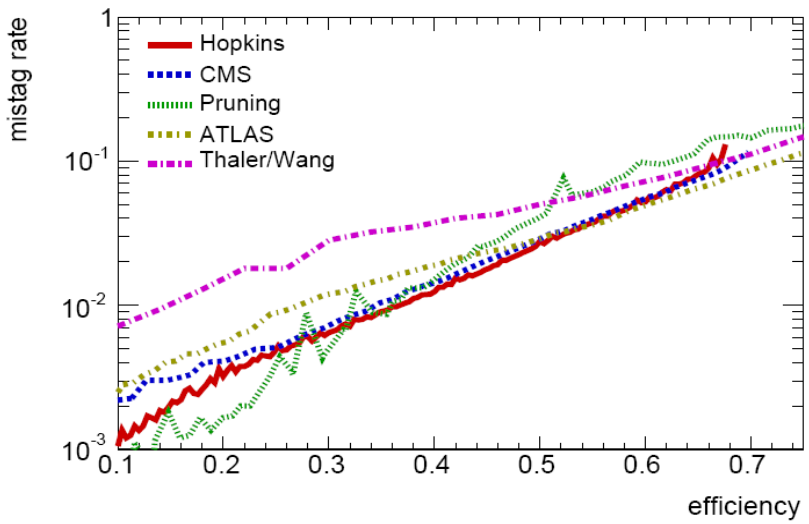


Full simulation study

Performance Comparison



(c) 300–400 GeV

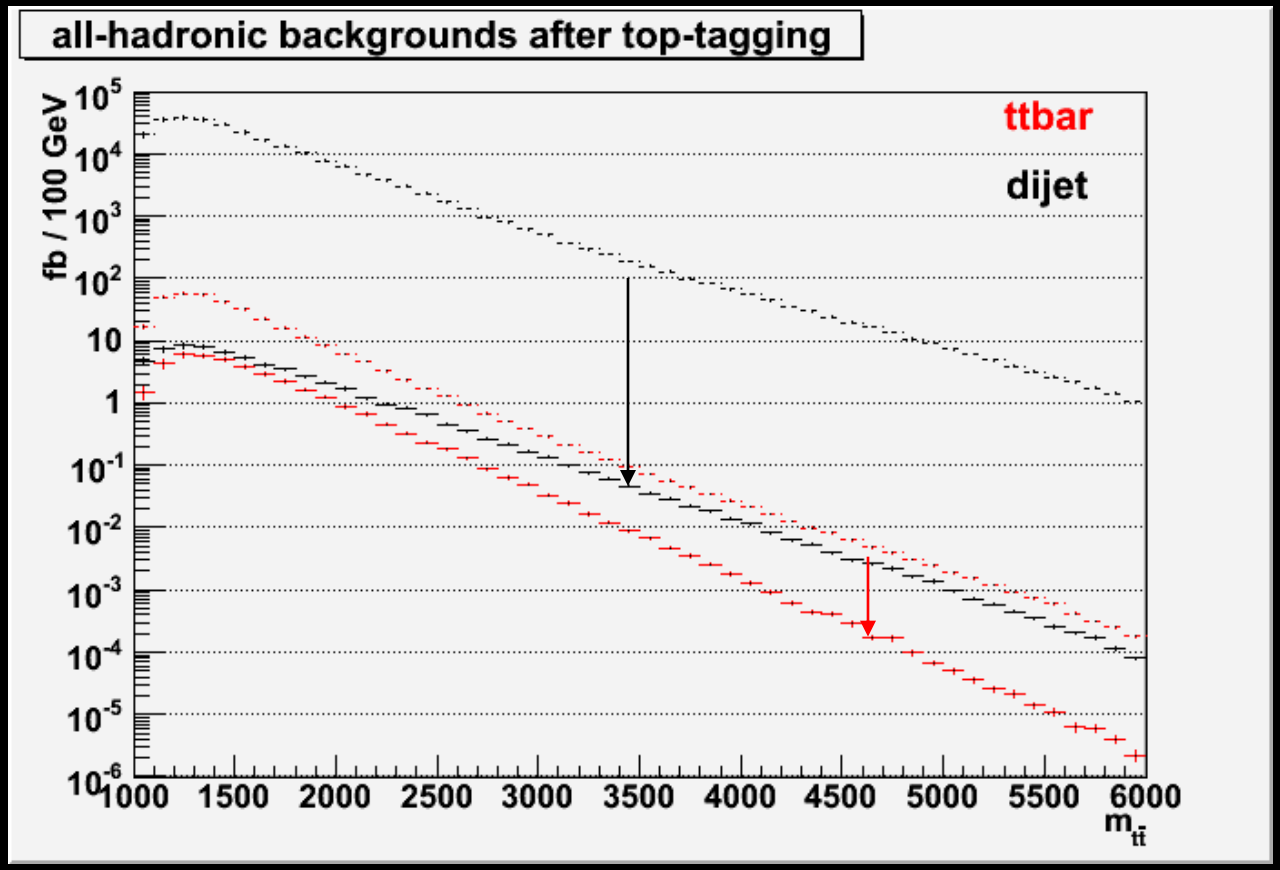


(d) 500–600 GeV

simple simulations again

BOOST 2010, arXiv:1012.5412

Backgrounds to All-Hadronic Resonance Search



PYTHIA 6.4 continuum QCD and top pair (LO + parton shower)

Improving the Tag?

- All taggers (mostly) use the coarsest level of detector information, HCAL @ $\Delta R \sim 0.1$
- Folding in full ECAL & tracker info may allow further improvements at $p_T > 1$ TeV

Substructure at ILC or CLIC?

- No UE, no ISR
  - almost all radiation in the jet is meaningful
- Color singlet machine => “lepton+jets” backgrounds are of a very different quality
  - e.g., Wq, Wg absent
  - W+jets still there, but *much* lower rate
- Clearly useful for all-hadronic search channels (q-qbar backgrounds)
- Suggests intelligent strategies to recover “parton-level” top decay kinematics from final-state hadrons, and more detailed aspects of how tops decay with full QCD

Summary

- Lots of ideas to beat QCD and identify top jets at the LHC
  - ~40% tag rates vs few % mistags
  - hopefully improvable
- Forces us to consider novel ways to organize the activity in a region of a detector
  - circling interesting patches with fixed size -> adaptively clustering interesting hotspots
- So far mistag rates look sane in data
- Potential relevance for ILC or CLIC
  - intermediate-boost tops
  - precision top-jets?