Boosted object tagging in Higgs, exotics, and other hadronic final state searches

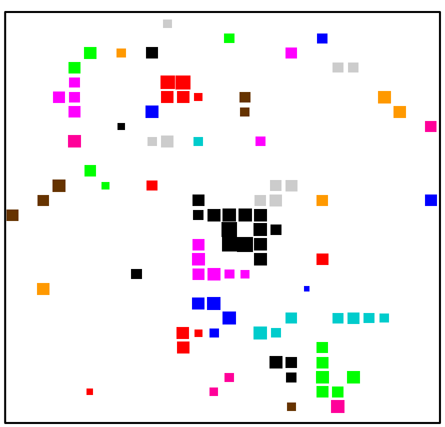
Nhan Tran
Fermi National Accelerator Laboratory

US ATLAS Hadronic Final State Analysis Forum:
Joint Theory/Experiment Open Session

December 03, 2012

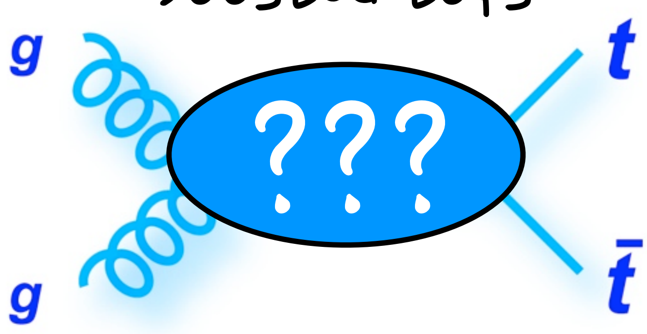
- Many recent phenomenological works highlight the potential of jet substructure for new physics searches
- Improved sensitivity in top analyses, boosted vector bosons and Higgs searches
- Requires an understanding of QCD radiation in jets

grooming

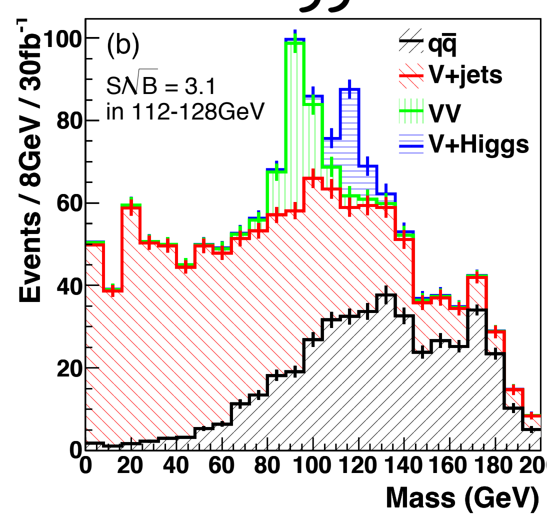


arxiv:0912.1342

boosted tops



higgs

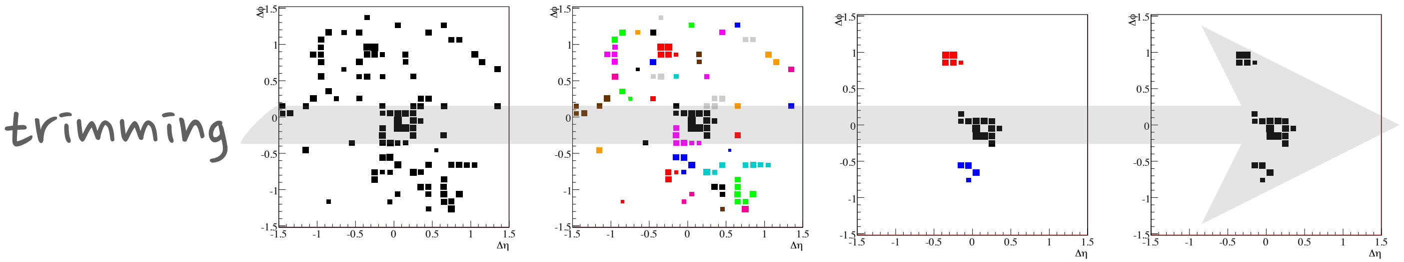


arXiv: 0802.2470

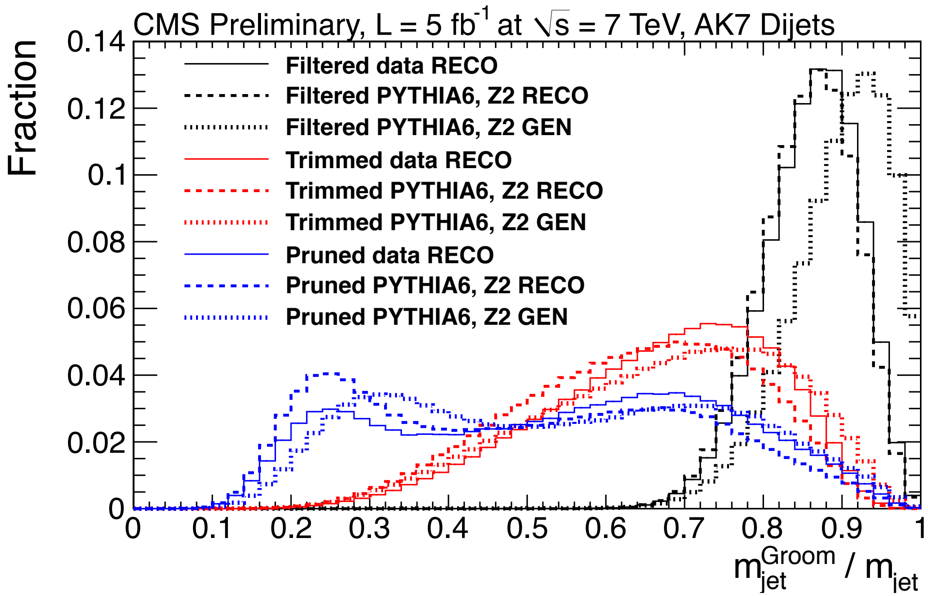
- A survey of the activities in CMS and ATLAS on tagging boosted objects to give a flavor of what has been done from the experimental side
- Experiments are scratching the surface of available phenomenological tools
- Outline:
  - Grooming algorithms
  - Boosted object taggers
  - b-tagging in boosted objects
- Focus on the algorithms and methods and experimental implementation

grooming algorithms

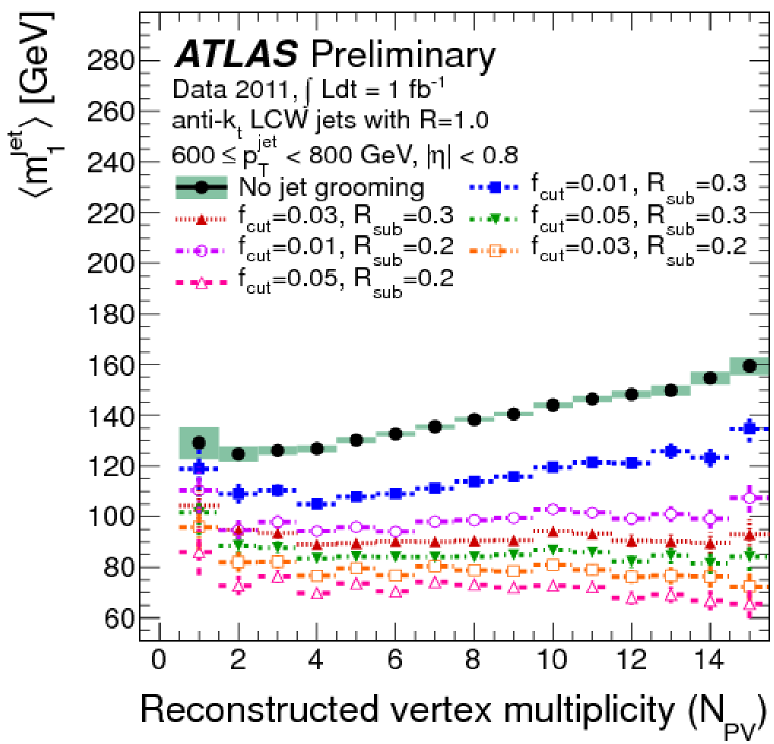
- Jet mass is one of the most important substructure variables for searches
- Jet grooming: removes soft components of jet due to UE and pileup
  - For searches, grooming can be used to further suppress background by pushing to lower jet mass
- Refer to the talks by Kalanand Mishra and David Miller earlier today for more details on inclusive properties of groomed jets
- Experiments' preferred grooming algorithms
  - CMS: pruned jets, ATLAS: trimmed jets



grooming algorithms



ATLAS scan of trimming parameters: Grooming algorithms are tunable to different levels of "aggressiveness".



CMS comparison of grooming algorithms using default parameters from reference

Pruning: $z_{cut} = 0.1$, $r_{cut} = 0.5$

Trimming: $r_{filt} = 0.2$, $pT_{frac,min} = 0.03$

Filtering: $r_{filt} = 0.3$, $n_{filt} = 3$

Pruning, found the be most aggressive algorithm, suitable for searches

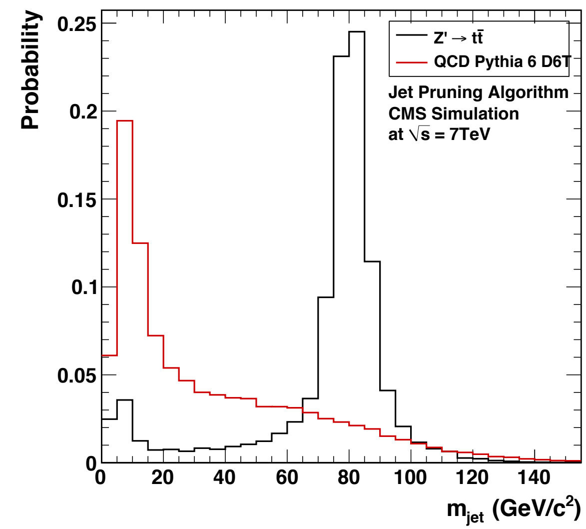
tagging algorithms

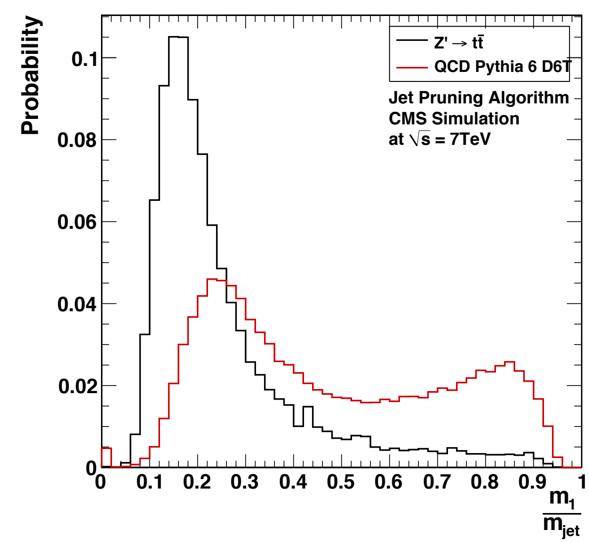
boosted W/Z tagging [CMS]

Identifying boosted W/Z bosons

not uniform across existing analyses but discuss most aggressive treatment

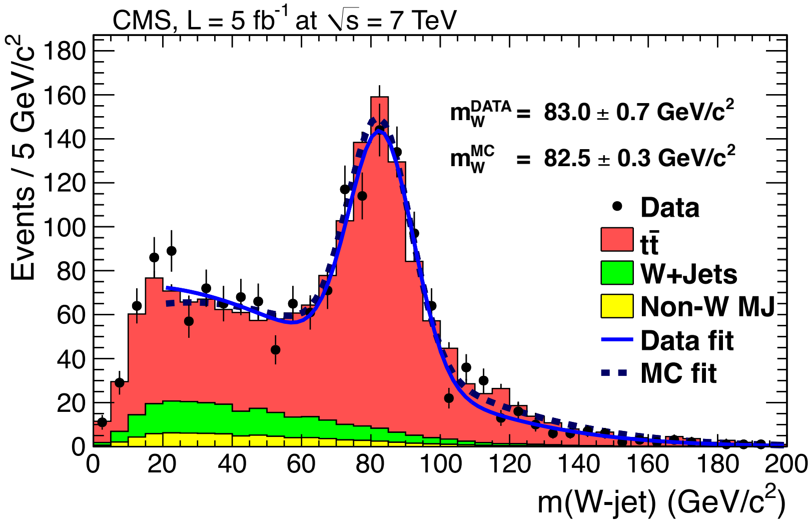
- Jet-finding with a large radius algorithm, typically CA8, AK7
- Prune jets with reference default parameters, removes soft and large angle constituents
- Cut on the mass drop, $\mu = m_1/m$
  - m_1 is mass of highest pT subjet
  - subjets defined by un-clustering last step

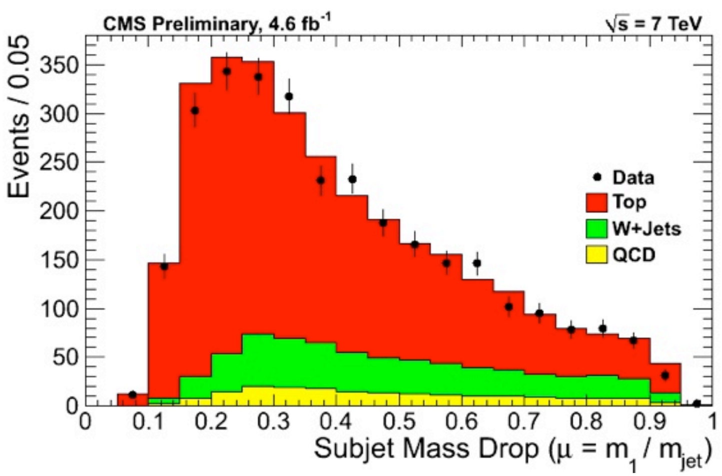


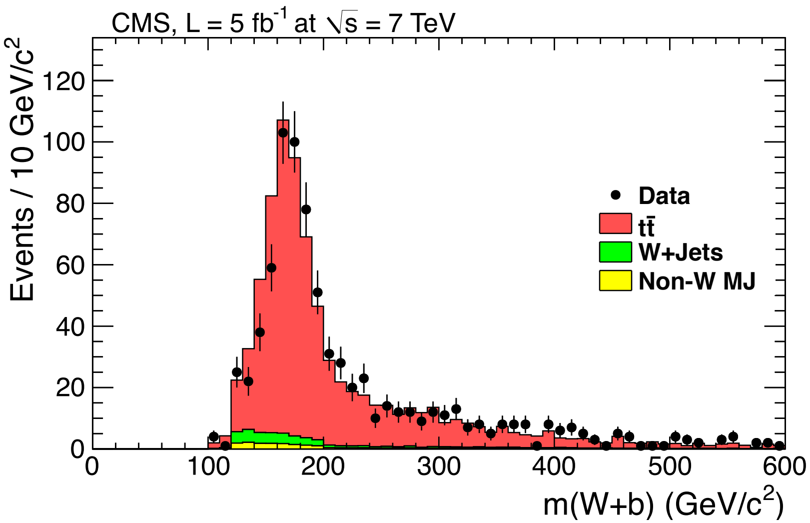


boosted W validation [CMS]

- Validation using merged W bosons in semi-leptonic $t\bar{t}$ sample
- Require a b-tagged jet in addition to high pT jet passing boosted W requirements
- Clear observation of merged W's and valuable sample for understanding mass scale and efficiencies



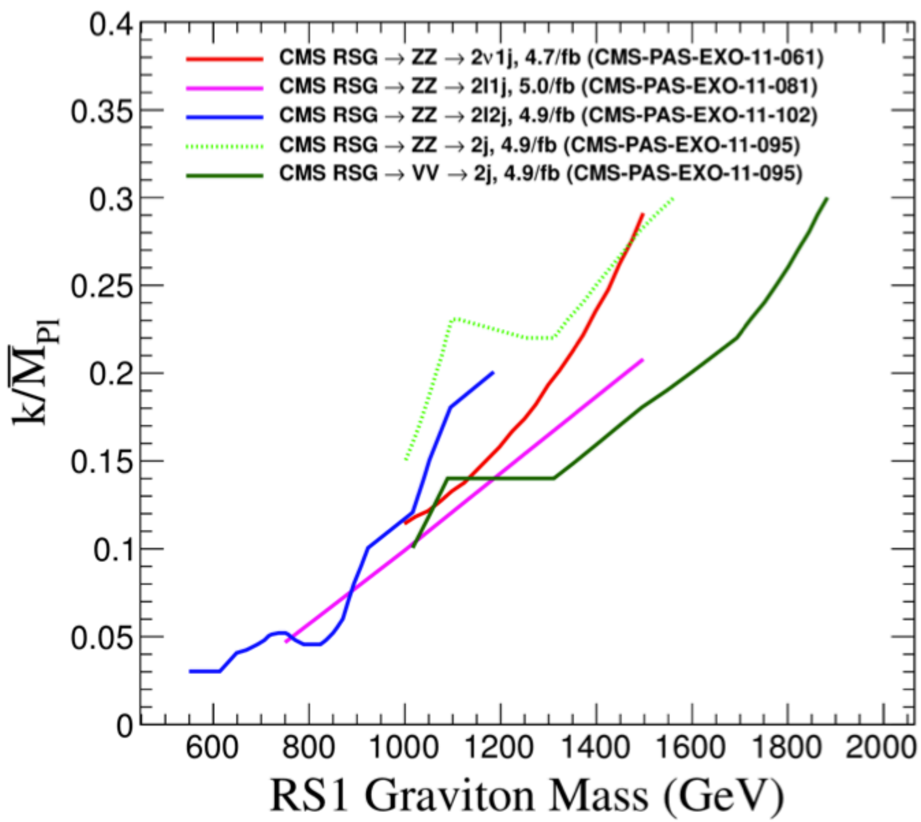




application

exotic graviton searches [CMS]

Combination of CMS searches
for RS1 Graviton in di-boson final states



$X \to ZZ \to 2\nu+1j$:

AK7 jet mass [CMS-PAS-EXO-11-061]

$X \to ZZ \to 2l+1j$:

AK7 jet mass [CMS-PAS-EXO-11-081]

$X \to VV \to 2j$:

CA8 pruned jet mass + mass drop cut [CMS-PAS-EXO-11-095]

4 of 5 analyses using jet substructure observables

W/Z tagging [ATLAS]

- Require a highly boosted AK4 jet
- Selection: Jet mass requirement

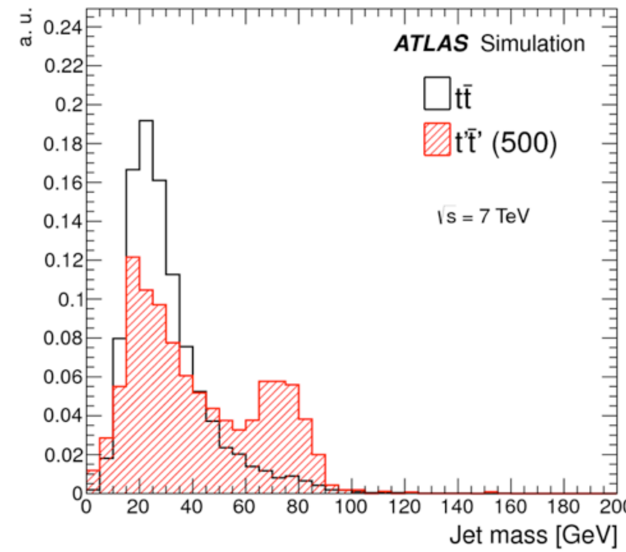
application

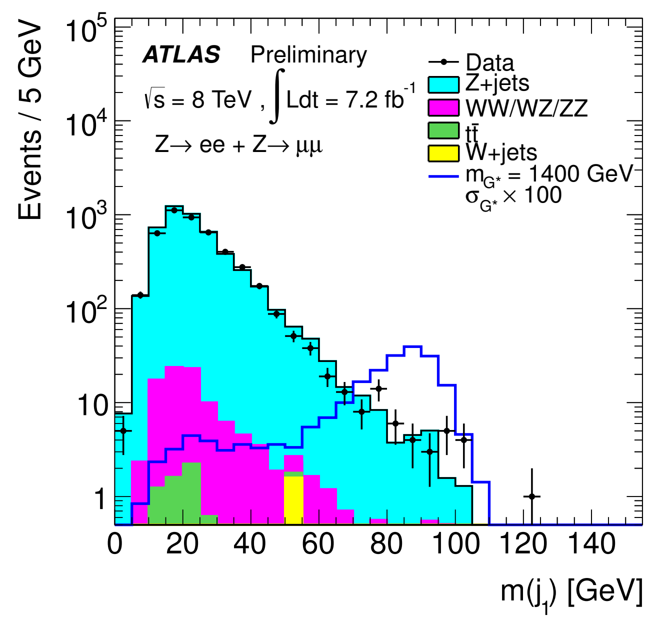
Pair produced t’: $pT_J > 250$ GeV and m_J = [60-110] GeV

application

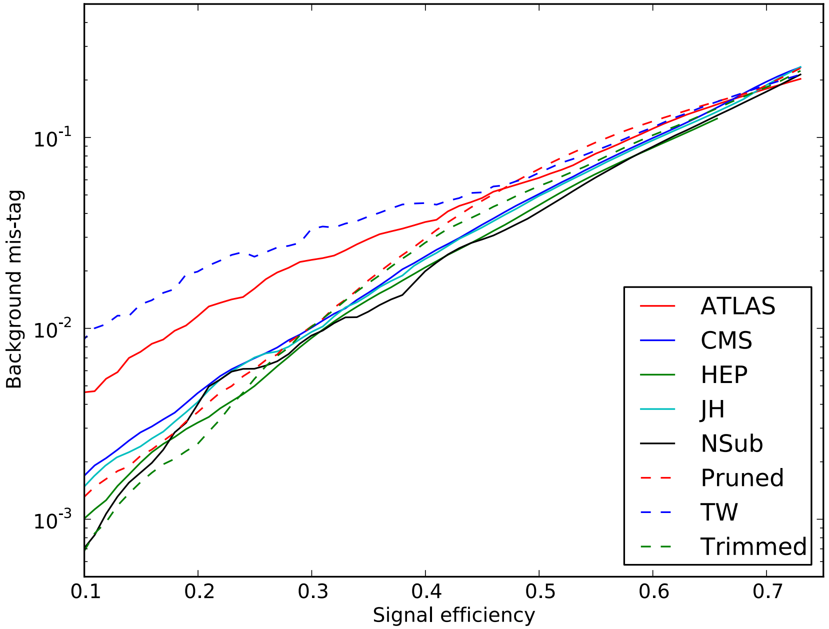
$X \to ZZ \to llqq$: $m_J > 40$ GeV and pT_{ll}, $pT_J > 200$ GeV

(only 8 TeV analysis presented today!)

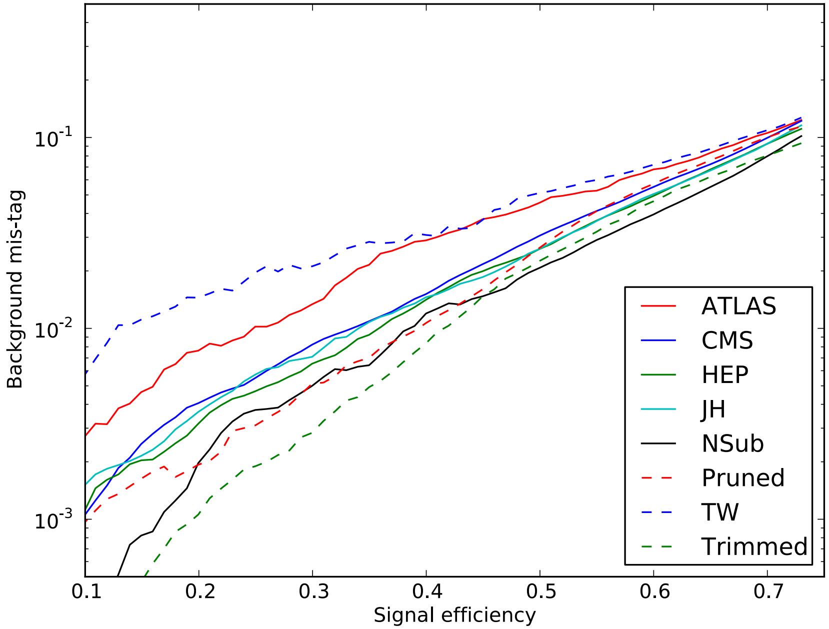




top tagging



(a) all p_T, optimised

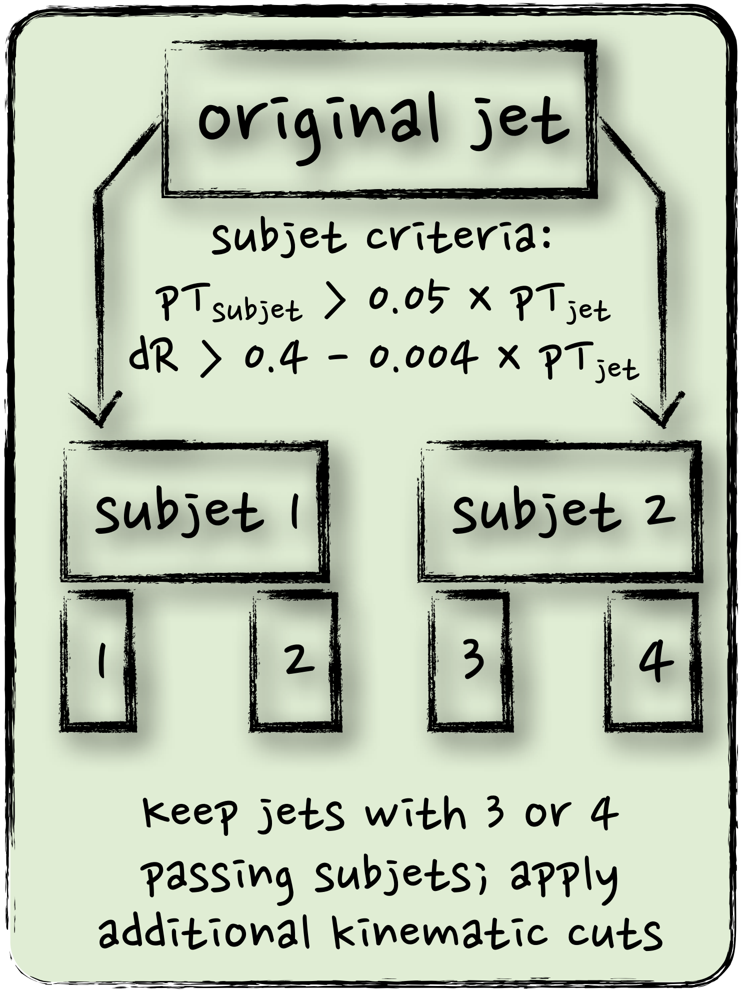


(b) p_T 500–600 GeV, optimised

BOOST 2011 Report

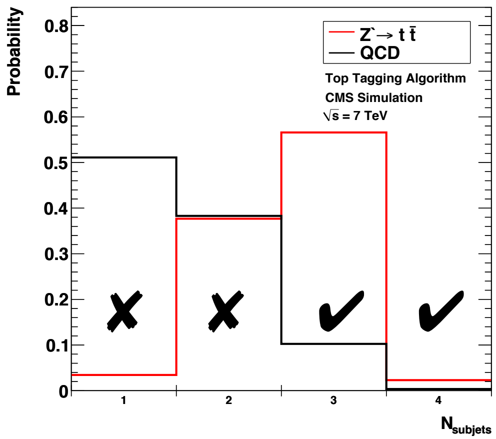
CMS top tagger [CMS]

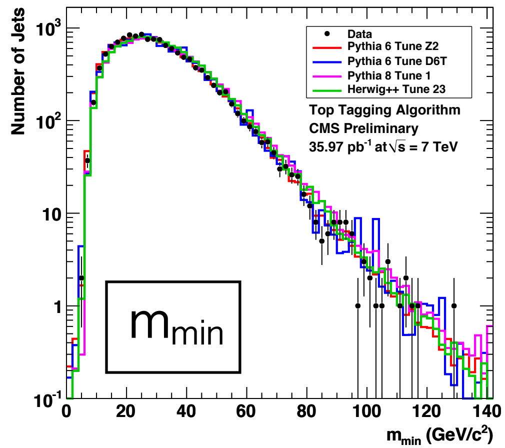
- Based on the JHU top tagger:
  PRL 101/142001 (2008)
  Kaplan et al.
- Cluster jets with CA8 algorithm
- Reverse clustering algorithm to find subjets, keep subjets passing following criteria
  - $pT_{subjet} > 0.05 \times pT_{jet}$
  - $dR > 0.4 - 0.004 \times pT_{jet}$
- Keep original jets with 3 or 4 passing subjets
  - Jet mass is [100-250] GeV
  - Minimum pairwise mass of hardest 3 subjets, $m_{min} > 50$ GeV

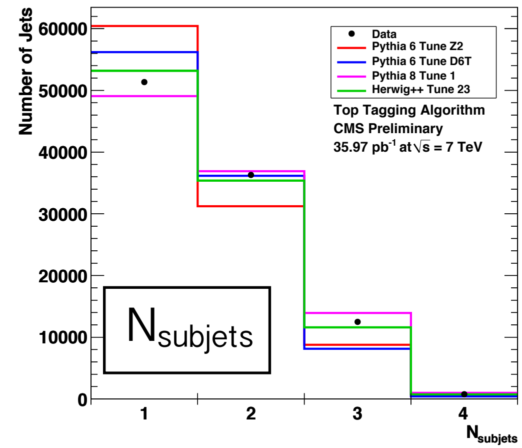


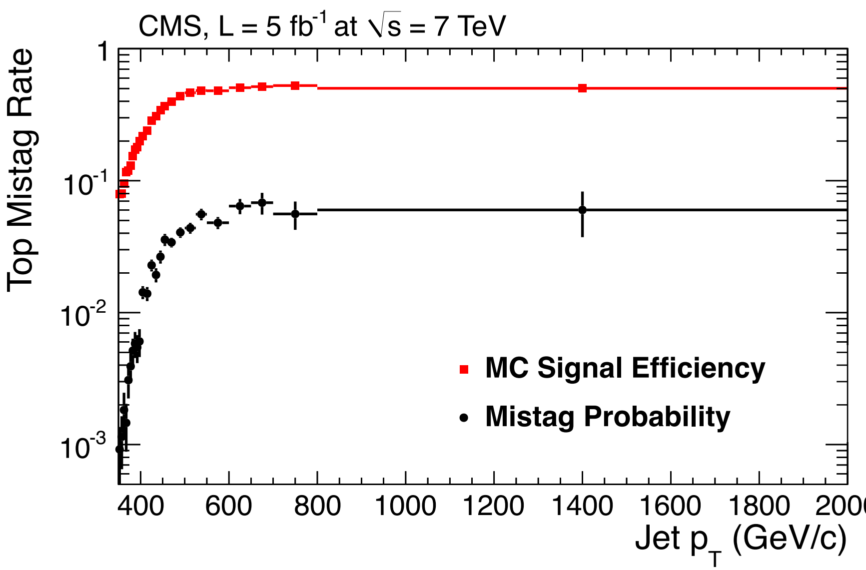
top tagging validation [CMS]

Good agreement between data and MC for algorithm observables, particularly with Herwig++









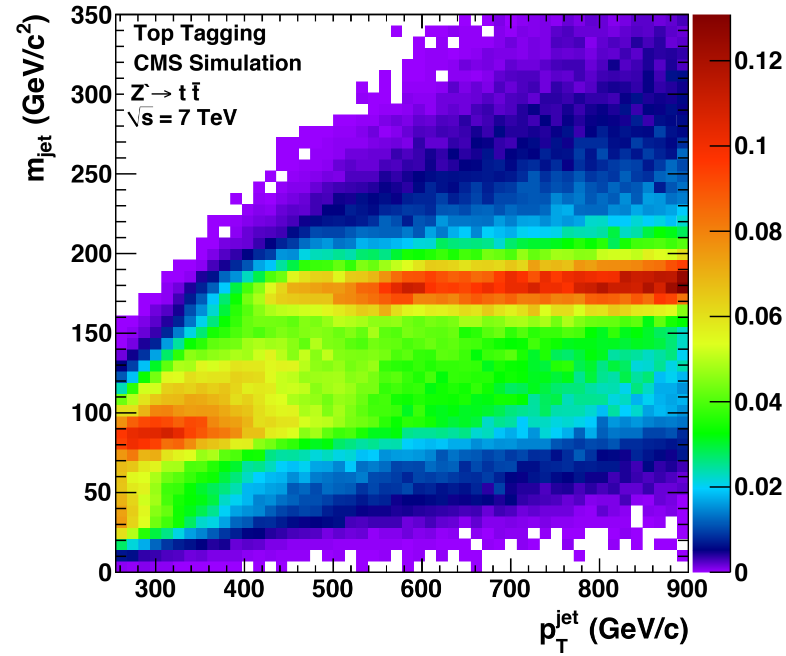
"Anti-tag and probe" method

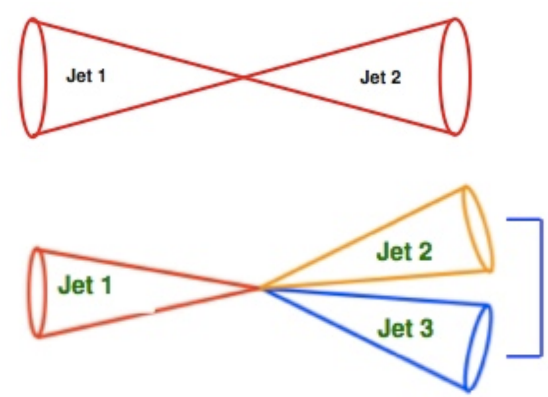
Invert the substructure requirements on tag side and determine how often jet is top-tagged on probe side

application

$X \rightarrow t\bar{t}$, all-hadronic channel [CMS]

- Two types of top reconstructions
- Type I: top fully merged in one jet
    - Jet pT > 350 GeV, apply the CMS top tagging criteria
- Type II: W merged into one jet, b jet reconstructed separately
    - Lead jet pT > 200 GeV
      m_J = [60-100] GeV, μ < 0.4;
      second jet pT > 30 GeV (no b-tag)
- Event classes, define hemispheres based on Type I candidate
    - Type I + Type I
    - Type I + Type II



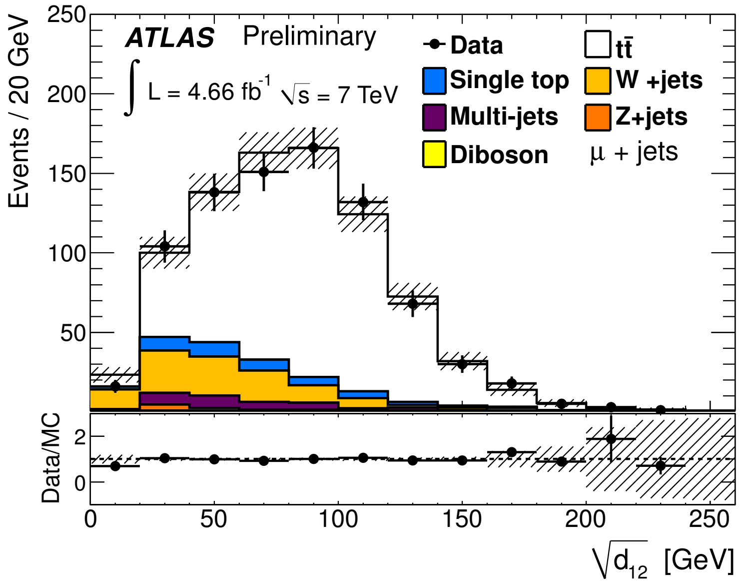


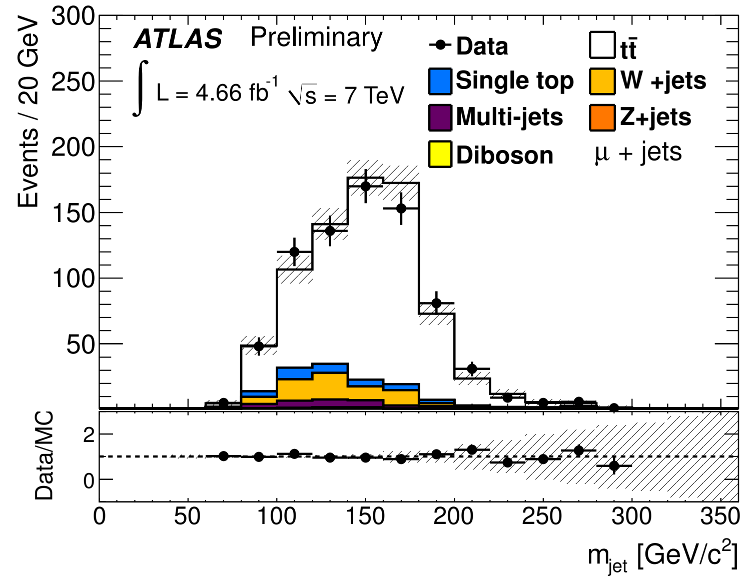
top tagging [ATLAS]

- Search for AK10 jet with pT >350 GeV, mass > 100 GeV
- Recluster with the k_t algorithm to determine the first k_t splitting scale $\sqrt{d_{12}} = \min(pT1, pT2) \times \Delta R_{12} > 40$ GeV

application

Semi-leptonic ttbar resonance search at 7 TeV

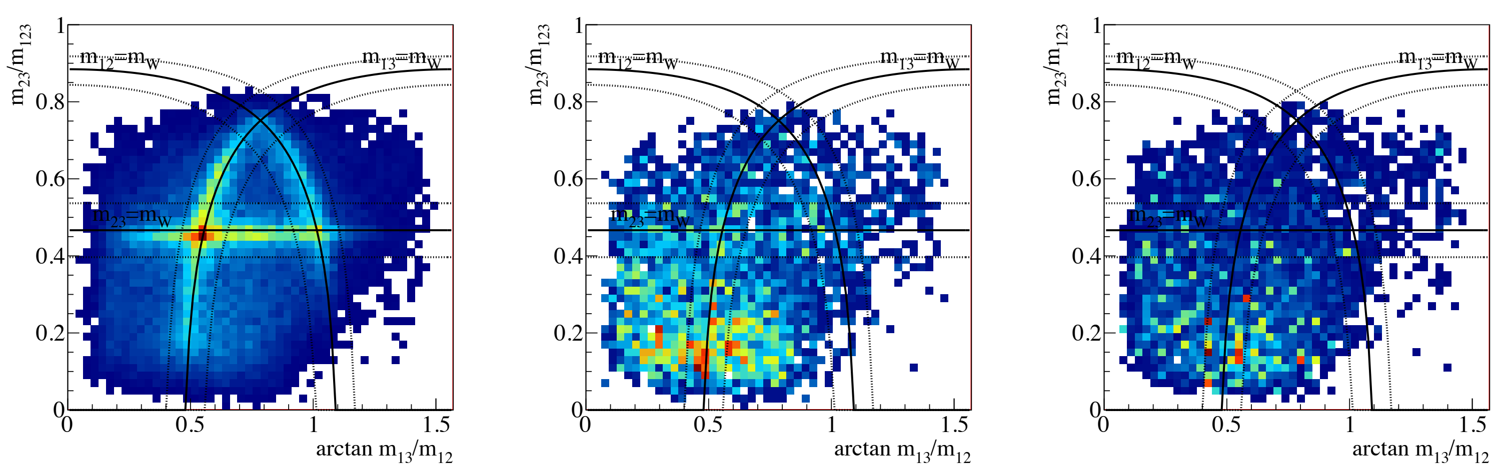




arXiv:1006.2833
Plehn et al.

HEP top tagger [ATLAS]

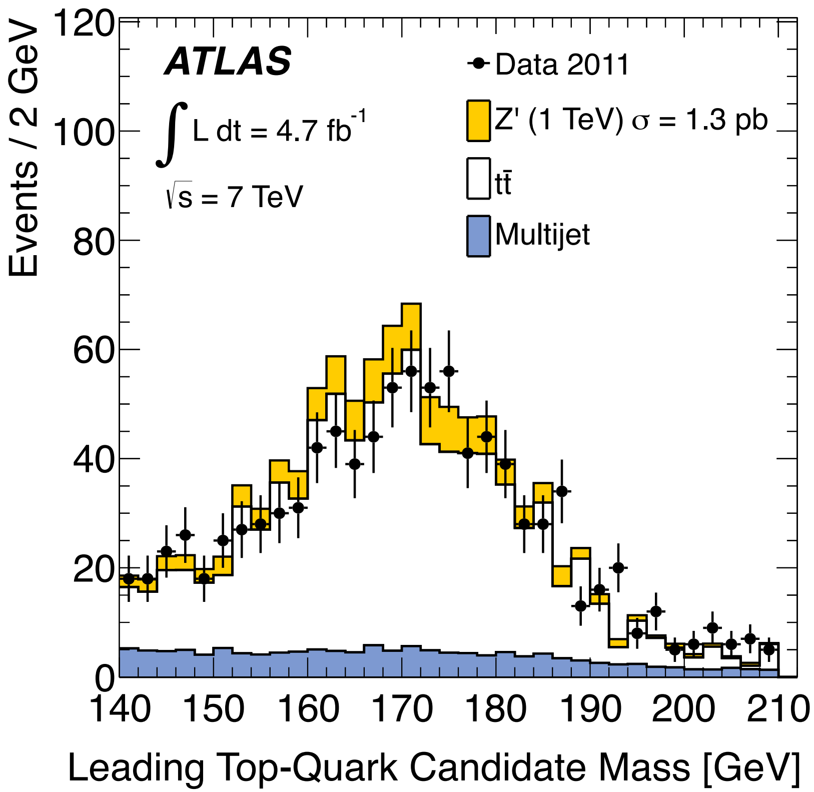
- For identifying moderately boosted tops, CA fat jets (R = 1.5, 1.8) with pT > 200 GeV
- Decluster jet keeping subjets that pass the mass drop criterion, $m_{j1} > m_{j2}$ and $m_{j1} < 0.8 \times m_j$ until each subjet each subjet has $m_{j,i} <$ 30 GeV
- Filter all combinations of triplets of subjets to remove UE/PU contributions, keeping 5 hardest filtered consituents to compute the jet mass; keep triplet with jet mass closest to m_t
- Apply kinematic constraints on all mass pairings: $\{m_{12}, m_{23}, m_{13}\}$

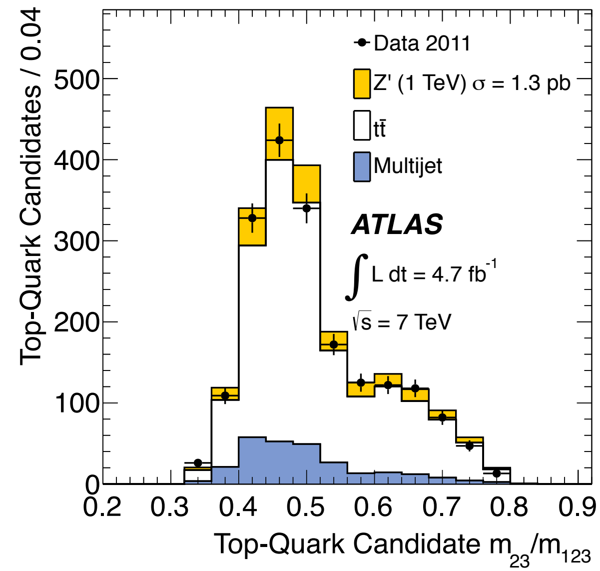


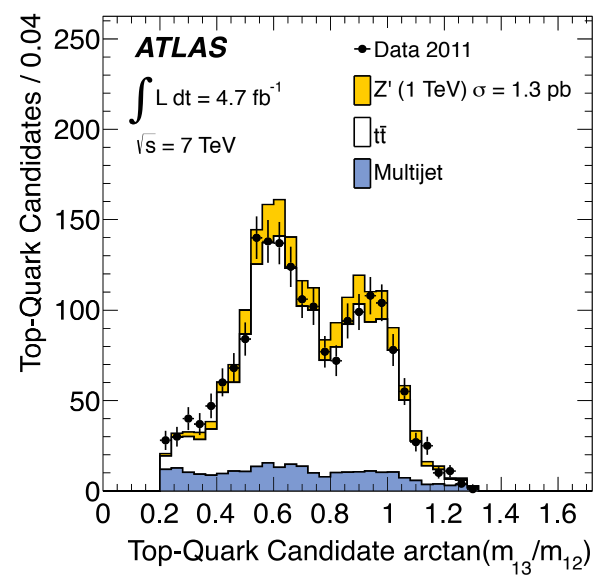
application

HEP top tagger [ATLAS]

- Fully hadronic ttbar resonance analysis
- HEP top tagger observables show good agreement with data







arXiv:1006.2035
Almeida et al.

Top Template tagger [ATLAS]

- Selection: AK10 jets with pT_1 (pT_2) > 500 (450) GeV
- Energy flow inside a jet compatibility with top quark decay
- Given a library of ~300k templates, encode the overlap into a single observable OV_3 (from 0-1)
- Libraries in bins of 100 GeV starting from 450 GeV

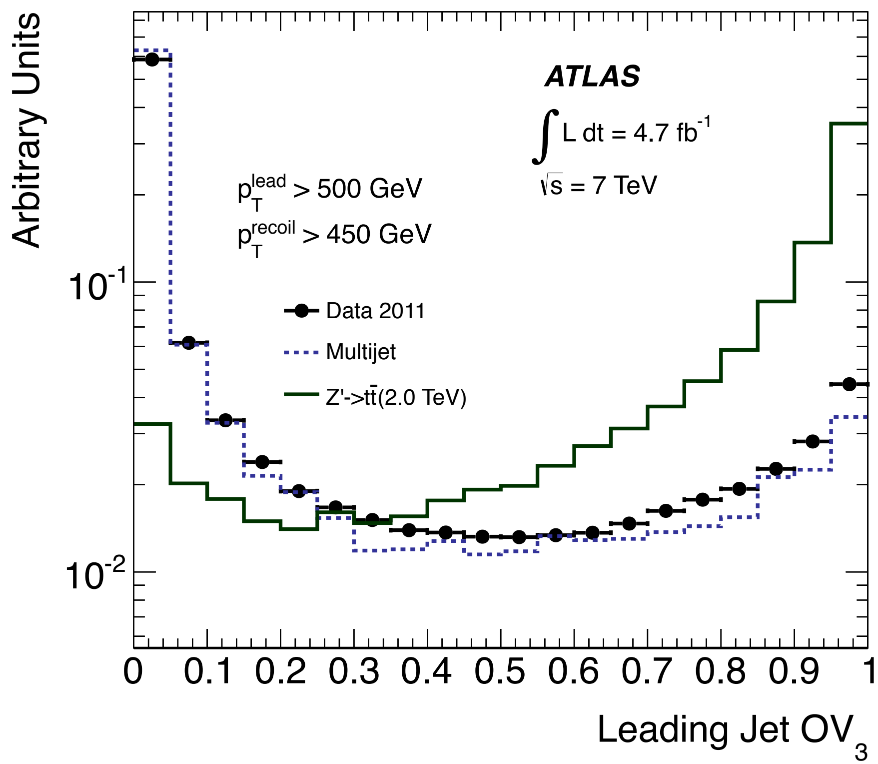
$$OV_3 = \max_{\{\tau_n\}} \exp\left[-\sum_{i=1}^{3} \frac{1}{2\sigma_i^2}\left(E_i - \sum_{\substack{\Delta R(\text{topo},i)\\<0.2}} E_{\text{topo}}\right)^2\right]$$

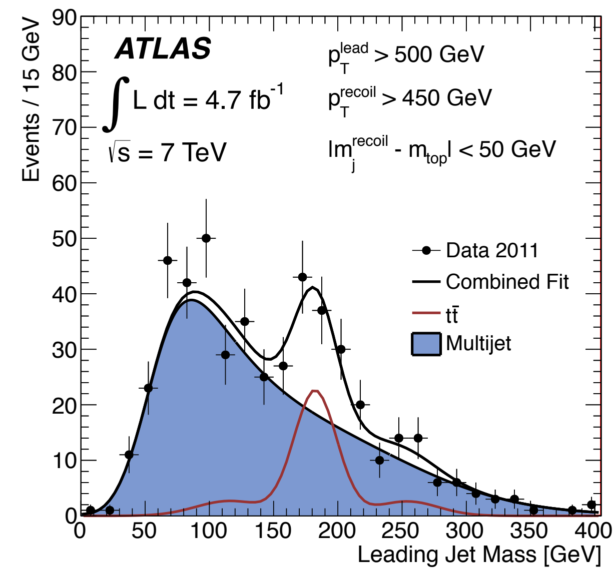
  - τ_n is set of templates
  - i sums over top-quark decay daughters, $\sigma_i = E_i/3$ is weight factor, Etopo is energy of topocluster required to be within $\Delta R < 0.2$
- Selection, make a cut on $OV_3 > 0.7$

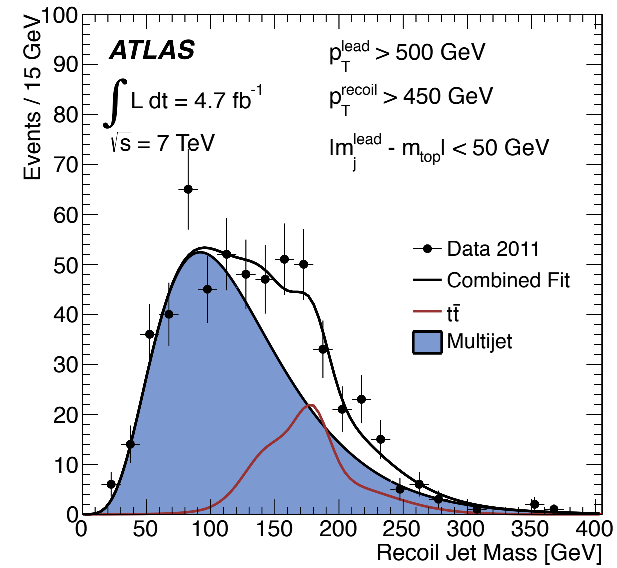
application

Top Template tagger [ATLAS]

- Fully hadronic ttbar resonance analysis
- Complimentary phase space w.r.t. the HEP top tagger, higher boost





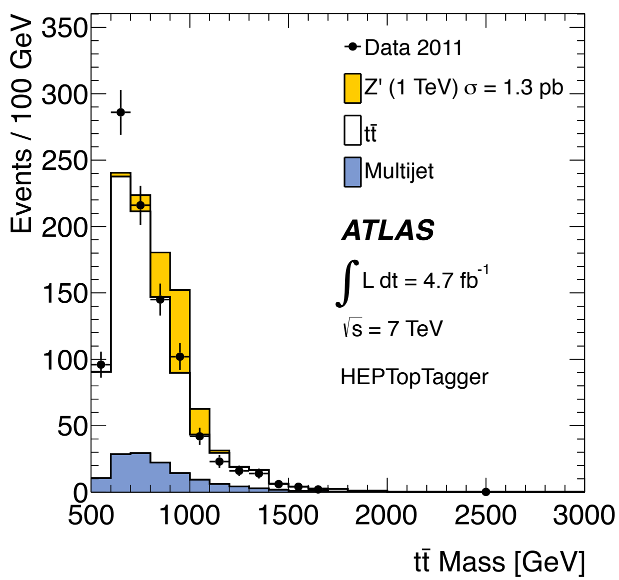


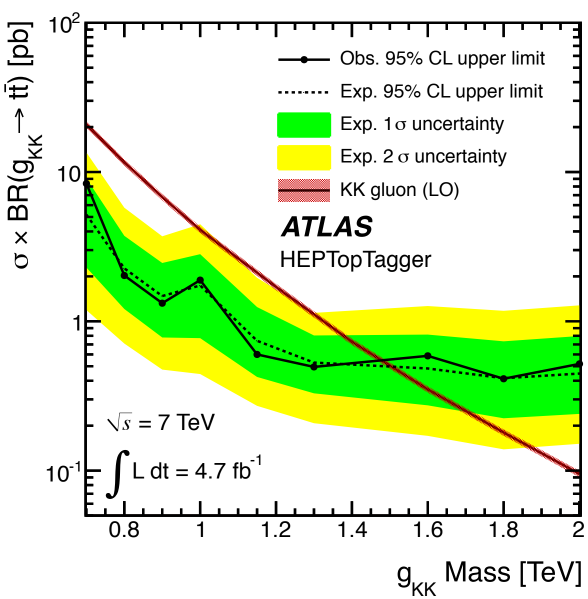
application

arXiv:1211.2202
submitted to JHEP

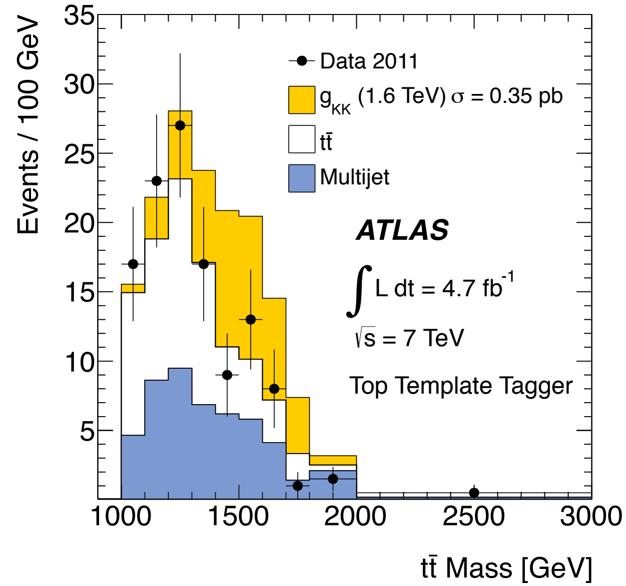
full hadronic ttbar [ATLAS]

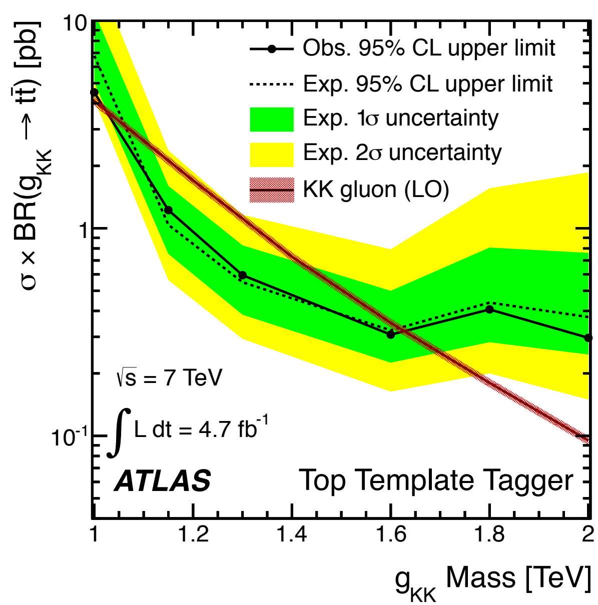
HEP top tagger





Top Template tagger





RPV gluinos w/N-subjettiness [ATLAS]

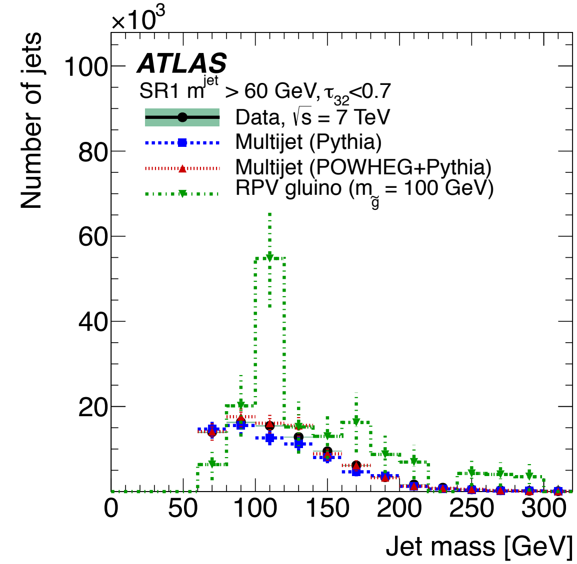
application

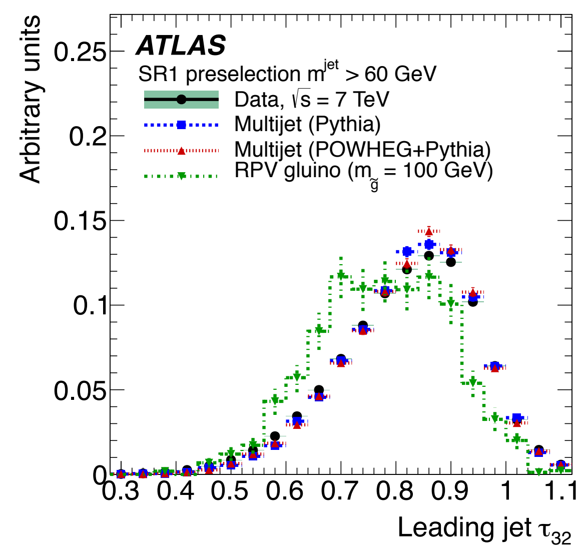
arXiv:1210.4813, submitted to JHEP

- Exotic resonance search: $\tilde{g}\to q\tilde{q}\to qq$, pair produced RPV gluinos
- For light gluinos, decaying quarks can be highly collimated
- AK10 jets with pT > 350 (or 200) GeV
- N-subjettness, τ_3/τ_2, variable used to identify jets with 3 subjets
  - Require $\tau_3/\tau_2 < 0.7$
  - Exclusive k_t axes

$$\tau_N = \frac{1}{d_0}\sum_k p_{\mathrm{T}k} \times \min(\delta R_{1k}, \delta R_{2k}, ..., \delta R_{Nk})$$

$$d_0 \equiv \sum_k p_{\mathrm{T}k} \times R$$

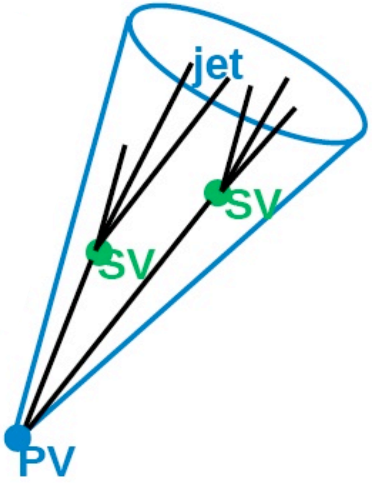




b-tagging and substructure

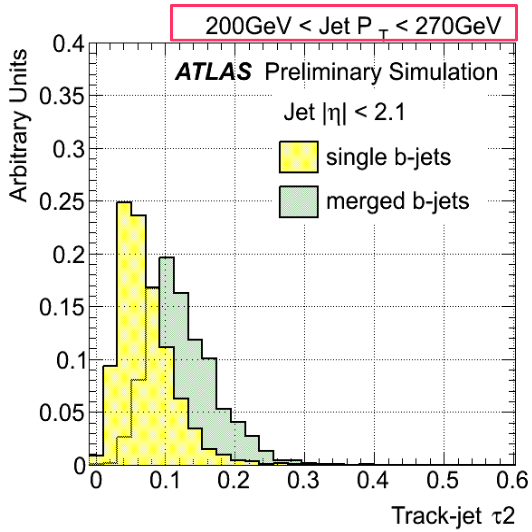
b-tagging in the boosted regime

- What happens to the performance of b-tagging in the highly boosted regime?
- Example, ATLAS fully hadronic ttbar: require b-tagged AK4 jet within given ΔR of fat jet
  - b-tagging for high pT top quark: ε = 50-70% decreasing with pT due to highly collimated decays
  - mistag rate is 3.5% (7%) for pT = 200 GeV (1 TeV) jets
- What about double b-tags? Hbb?
  - Searching for Hbb with substructure
  - Exotic production of boosted Higgs, i.e. t'→tH

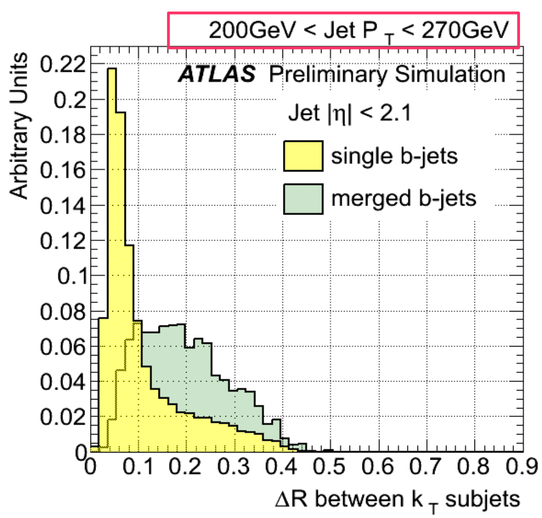


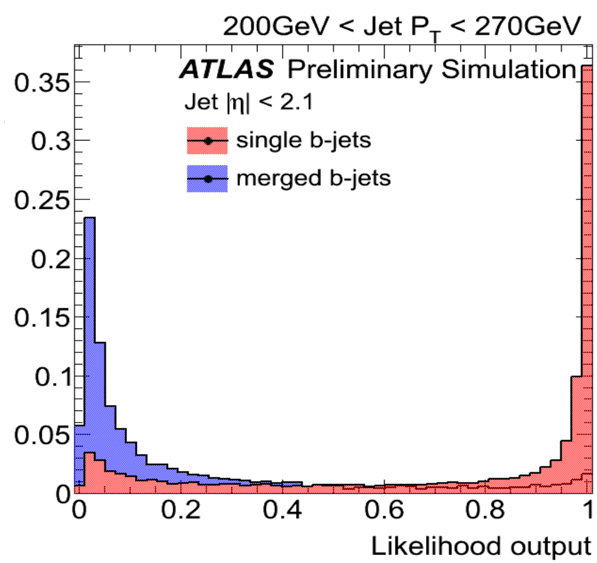
See M.L. Gonzalez-Silva's talk at BOOST 2012

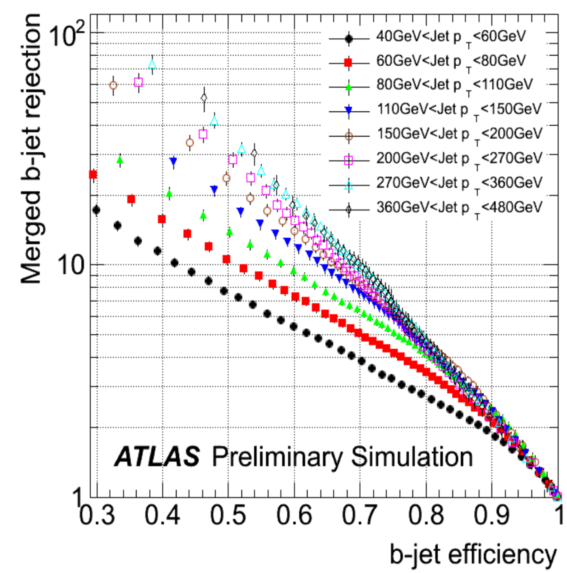
double b-hadron tagging [ATLAS]



- Study AK4 jets, b-tagged with ATLAS MV1 algorithm
- Perform multivariate analysis using substructure variables to distinguish between single and double b-hadron cases
    - track multiplicity, jet width, ΔR subjets, τ_2
- Input observables validated in data with good agreement with MC





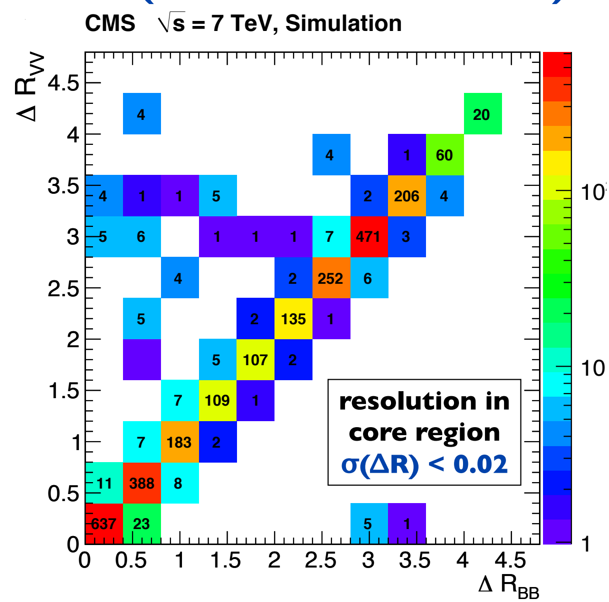


See A. Schmidt's talk at BOOST 2012

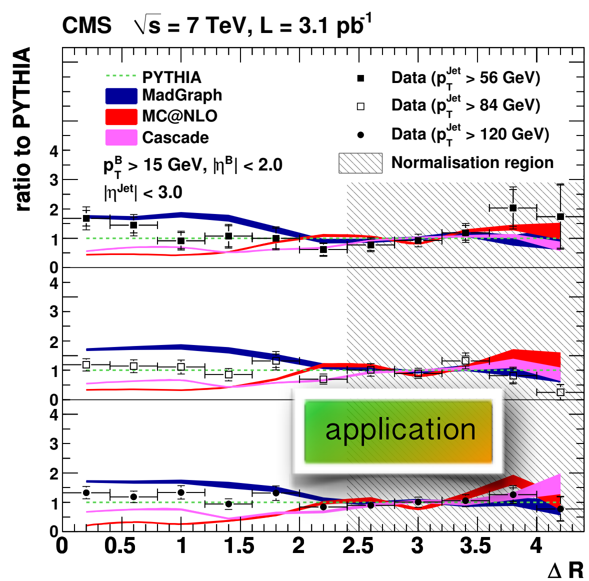
iterative vertex finder [CMS]

- New technology for bbar at small angles, uses only secondary vertices, no jets required (still can use if wanted)
- Seed from tracks with high impact parameter [technical details in backup]
- Resolution scale, $\sigma(\Delta R)$ ~ 0.02 in the core region

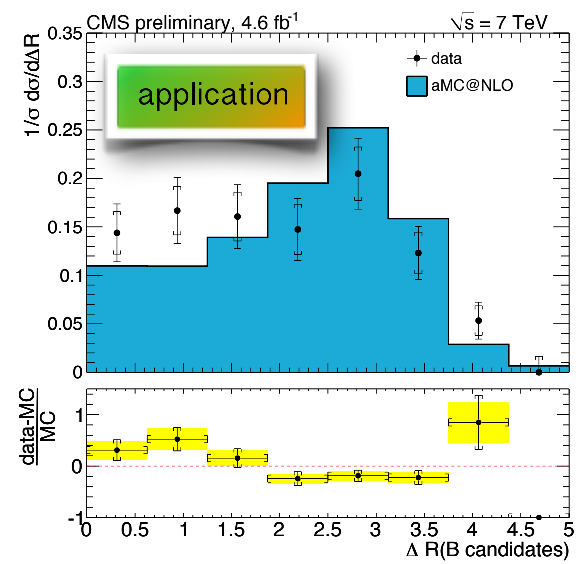
ΔR (true vs. reconstructed)



bb angular correlations
JHEP 03 (2011) 136



Zbb angular correlations
CMS PAS EWK-11-015



summary and outlook

- A summary of boosted object taggers is presented in both CMS and ATLAS
  - Grooming algorithms used to suppress backgrounds
  - W/Z “taggers” using jet mass and mass drop
  - Top tagging with CMS top tagger, kt splitting scales, HEP top tagger, Template top tagger
  - RPV gluinos with τ_3/τ_2
- Results for b-tagging in the boosted regime and methods presented for identifying bbar pairs at small angles

summary and outlook

Tuhin Roy

[Johns Hopkins] Top Tagging
(Kaplan, Rehermann, Schwartz, Tweedie; 0806.0848)

New jet shapes
(Almeida, et al.; 0807.0234, 0810.0934)

3-body kinematic variables
(Thaler, Wang; 0806.0023)

Pruning
(Ellis, Vermilion, Walsh; 0903.5081)

CMS Top Tagging
(0909.4894)

HEP Top Tagger
(Plehn, Salam, Spannowsky; 0910.5472)

Jet Trimming
(Krohn, Thaler, Wang; 0912.1342)

More jet shapes
(Chekanov, Proudfoot, Levy, Yoshida; 1002.3982, 1009.2749)

HEP Top Tagger++
(Plehn, Spannowsky, Takeuchi, Zerwas; 1006.2833...)

Jet Pull
(Gallicchio, Schwartz; 1001.5027)

New physics multi-tagging
(Kribs, Martin, Roy, Spannowsky; 0912.4731, 1006.1656)

Template overlap
(Almeida et al.; 1006.2035)

ISR tagging
(Krohn, Randall, Wang 1101.0810)

N-Subjettiness
(Thaler, Van Tilburg 1011.2268; Kim 1011.1493)

"Jet substructure without trees"
(Jankowiak, Larkoski; 1104.1646)

Shower deconstruction
(Spannowsky, Soper 1102.3480)

Dipolarity
(Hook, Jankowiak, Wacker; 1102.1012)

Multivariate quark/gluon discrimination
(Gallicchio, Schwartz; 1106.3076)

too much creativity to fit in one page

summary and outlook

Tuhin Roy

A lot of new ideas out there to be tried!

W/Z/Higgs search algorithms,
boosted tops improvements,
multivariate W-tagging techniques [1012.2077],
Qjets [1201.1914],
jet charge [1209.2421],
PU corrections to jet shape variables [1211.2811],
etc...

[Johns Hopkins] Top Tagging
(Kaplan, Rehermann, Schwartz, Tweedie; 0806.0848)

New jet shapes
(Almeida, et al.; 0807.0234, 0810.0934)

Pruning
(Ellis, Vermilion, Walsh; 0903.5081)

CMS Top Tagging
(0909.4894)

HEP Top Tagger
(Plehn, Salam, Spannowsky; 0910.5472)

Jet Trimming

HEP Top Tagger++

Jet Pull

New physics multi-tagging
(Kribs, Martin, Roy, Spannowsky; 0912.4731, 1006.1656)

Template overlap
(Almeida et al.; 1006.2035)

ISR tagging
(Krohn, Randall, Wang 1101.0810)

N-Subjettiness
(Thaler, Van Tilburg 1011.2268; Kim 1011.1493)

"Jet substructure without trees"
(Jankowiak, Larkoski; 1104.1646)

Shower deconstruction
(Spannowsky, Soper 1102.3480)

Dipolarity
(Hook, Jankowiak, Wacker; 1102.1012)

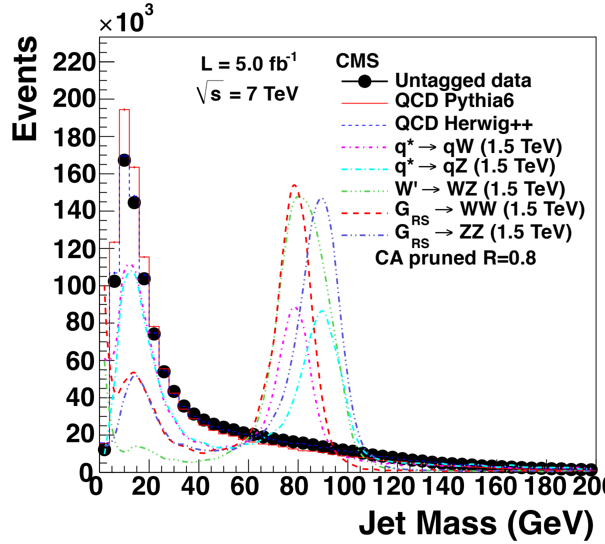
Multivariate quark/gluon discrimination
(Gallicchio, Schwartz; 1106.3076)

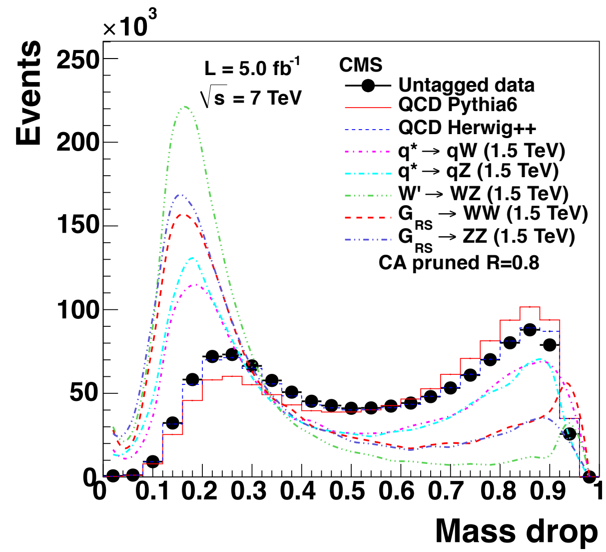
too much creativity to fit in one page

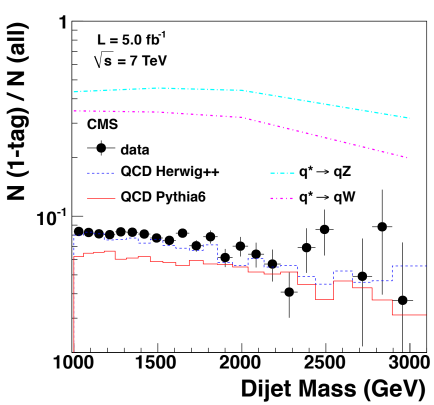
backup

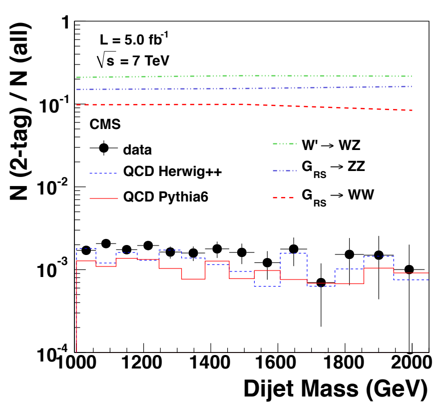
X → qV/VV

- Search for dijet final states:
  - qW, qZ, WW, ZZ, WZ
  - model interpreation: excited quark q*, RS gravitons and W’
- Analysis strategy
  - Employs V-tagging: CA8 pruned jets + mass drop cut; m_J = 70-100 GeV and μ < 0.25
  - Identify 1-tag and 2-tag topologies depending on final state
    - Tag rate from semileptonic tt



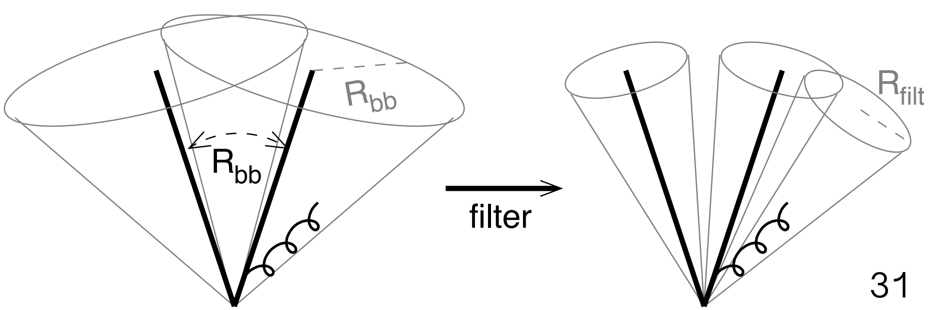






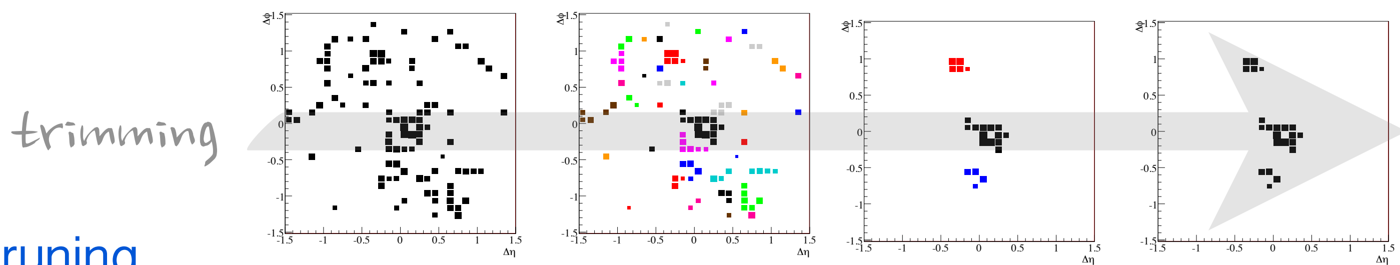
jet substructure algorithms

- Study jet mass properties under three jet grooming techniques
  - Trimming: http://arxiv.org/abs/0912.1342
  - Filtering: http://arxiv.org/abs/0802.2470
  - Pruning: http://arxiv.org/abs/0903.5081
  - This round of analysis uses default parameters from each of the references.
- Filtering
  - reclustering jet constituents with smaller radius, r_{filt}, keeping n_{filt} hardest sub-jets
  - default parameters: r_{filt} = 0.3, n_{filt} = 3
  - sub-jet clustering algorithm: Cambridge-Aachen



jet substructure algorithms

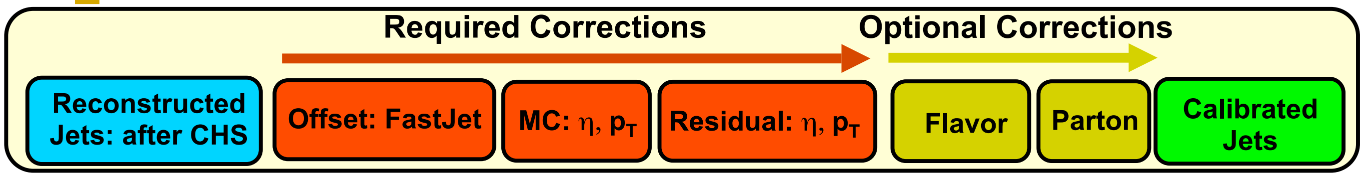
- Trimming
  - reclustering with smaller radius, r_{filt}, keeping sub-jets with a fraction, $pT_{frac,min}$, of original jet pT
  - default parameters: $r_{filt} = 0.2$, $pT_{frac,min} = 0.03$
  - sub-jet clustering algorithm: kT



- Pruning
  - reclustering with sequential recombination algorithm, veto soft and large-angle recombinations between pseudojets i and j
    - veto: $d_{ij} > r_{cut} \times 2m/pT$; $z = \min(pT_i, pT_j)/pT_{i+j} < z_{cut}$
  - default parameters: $z_{cut} = 0.1$, default $r_{cut} = 0.5$
  - subjet clustering algorithm: Cambridge-Aachen

jet corrections

Jet energy calibration: overview



✦ Factorization facilitates the use of data-driven corrections

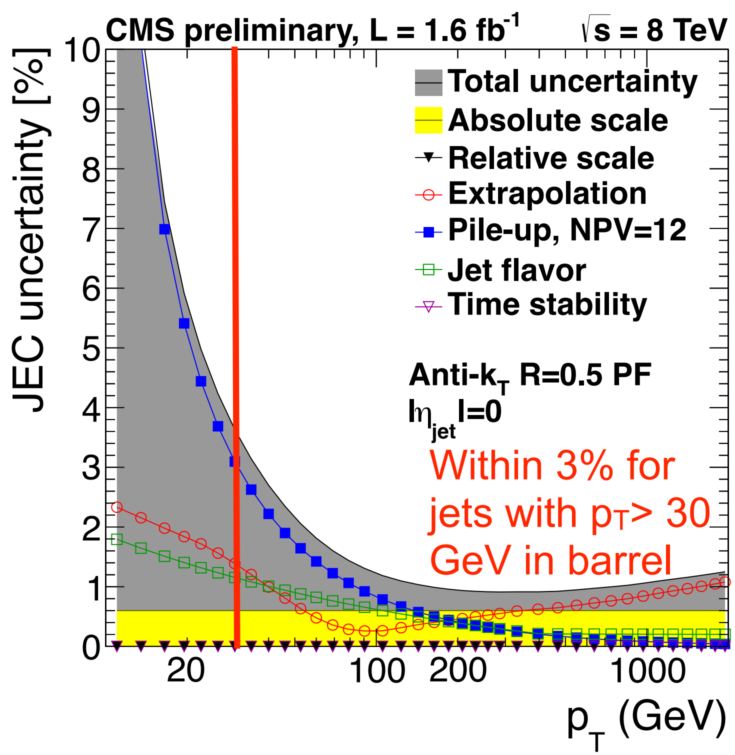
- Breaking the correction into pieces that are naturally measured in collider data:
    - **Offset**: pile-up and noise measured in zero-bias events.
    - **MC**: jet response vs. η, P_T using MC truth.
    - **Residual**: jet response vs. η, P_T using dijet balance and γ/Z+jet in data.

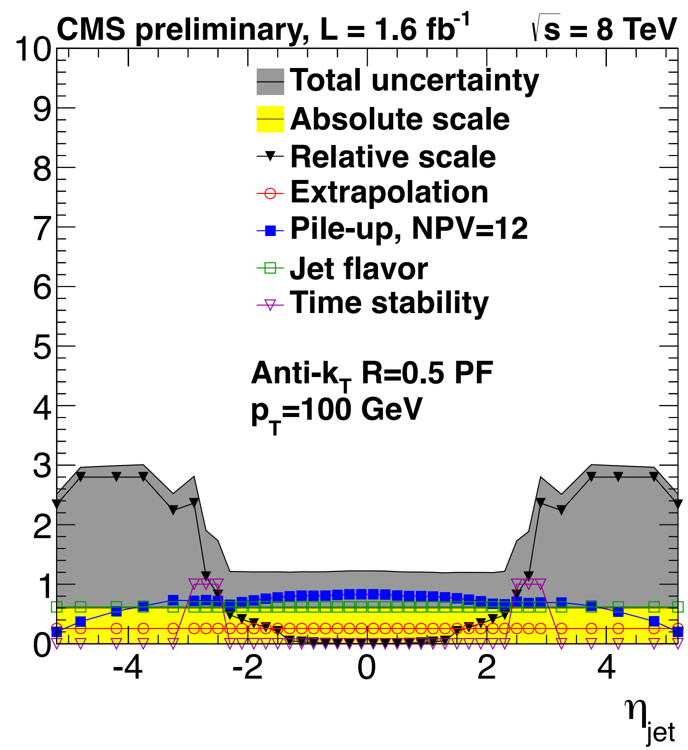
In CMS the most widely used jet is anti-k_T 0.5 (0.7 for QCD measurements). For pileup studies, consider anti-k_T 0.5, 0.7, 0.8 with various grooming techniques: filtering, trimming, pruning.

jet correction uncertainties

Correction uncertainties

- ✦Uncertainties in 2012 data comparable to 2010, 2011.
  - Pileup uncertainties increasing due to higher average pileup.





technical details

1. coarse pre-clustering of seeds based on
   - track distances, angles

2. vertex reconstruction/fitting with "adaptive fitter"
   - iterative procedure, outlier resistant (small weights for outliers)
   - χ^2/ndof < 10, 3D-significance > 0.5, 2D-significance>2.5, pointing angle cos(α)> 0.98

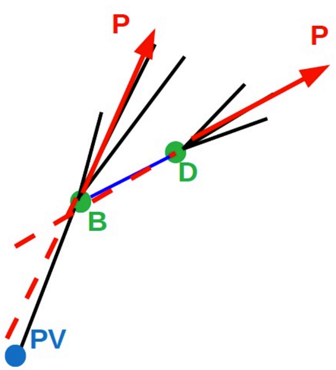
3. vertex merging
   - check all vertices for shared tracks
   - remove vertex if shared fraction > 0.2 (and dist. significance<2)

4. vertex arbitration
   - trade off tracks between PV and SV based on track distance (significance) to vertices
   - refit vertices with new track selection

5. vertex merging again (step 3.)

B candidates

- B decays can have two vertices ($B \rightarrow D$ cascade)
- merging of two-SV kinematics into one B candidate if
  - $\Delta R<0.5$,
    $m_{tot}<5.5$ GeV,
    pointing angle (BD, P): $\cos(\alpha)>0.99$



- B candidate selection:
  - $m>1.4$ GeV, $p_t>8$ GeV, N(tracks)>=3, flight significance >5, $|\eta|<2$

- retain events with two B candidates
  $m_1+m_2>4.5$ GeV