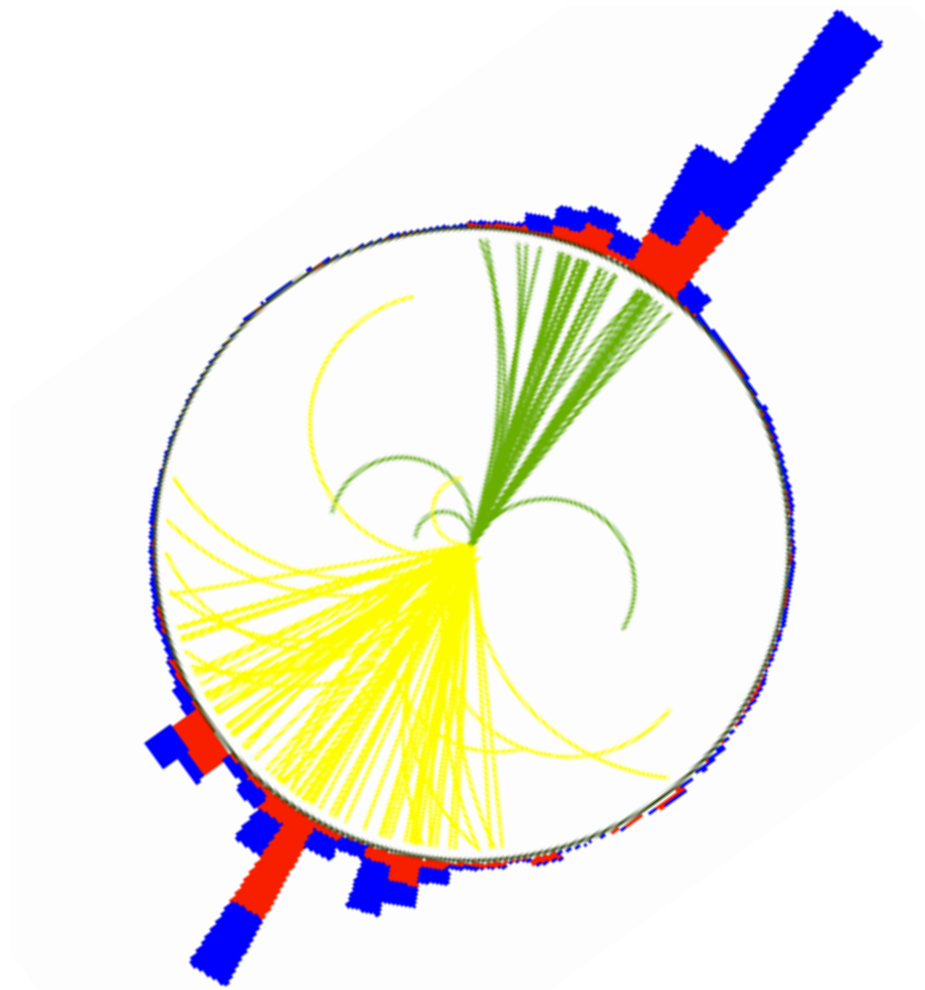
Boosted Top Quark Analyses at CMS

Justin Pilot University of California, Davis
on behalf of the CMS Collaboration

Top Physics Workshop
4 May 2012

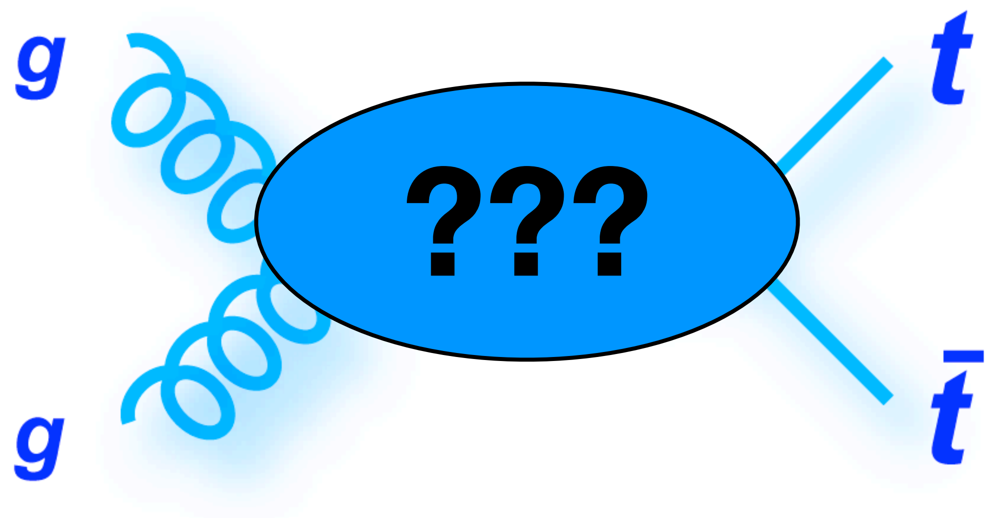
Outline

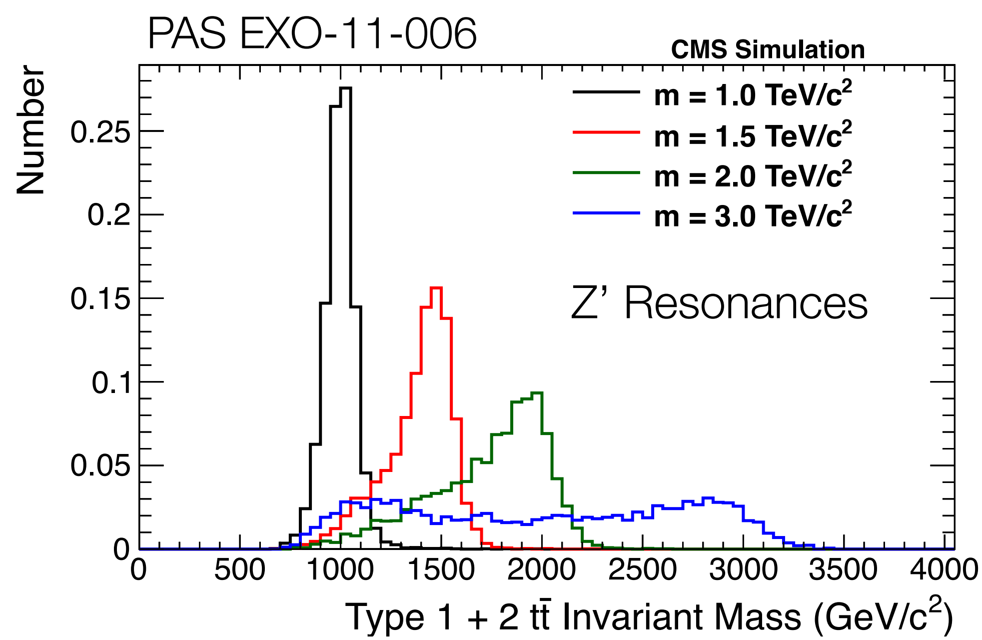
- Introduction to boosted top quarks
  - Why they are studied
  - How do we search for them
- Algorithms for boosted top reconstruction
  - Jet substructure
  - Jet grooming techniques
  - Event reconstruction
- Analyses using boosted tops at CMS
  - Lepton + jets search
  - All-hadronic search
- Plans for future improvements
  - Algorithm optimization



Introduction

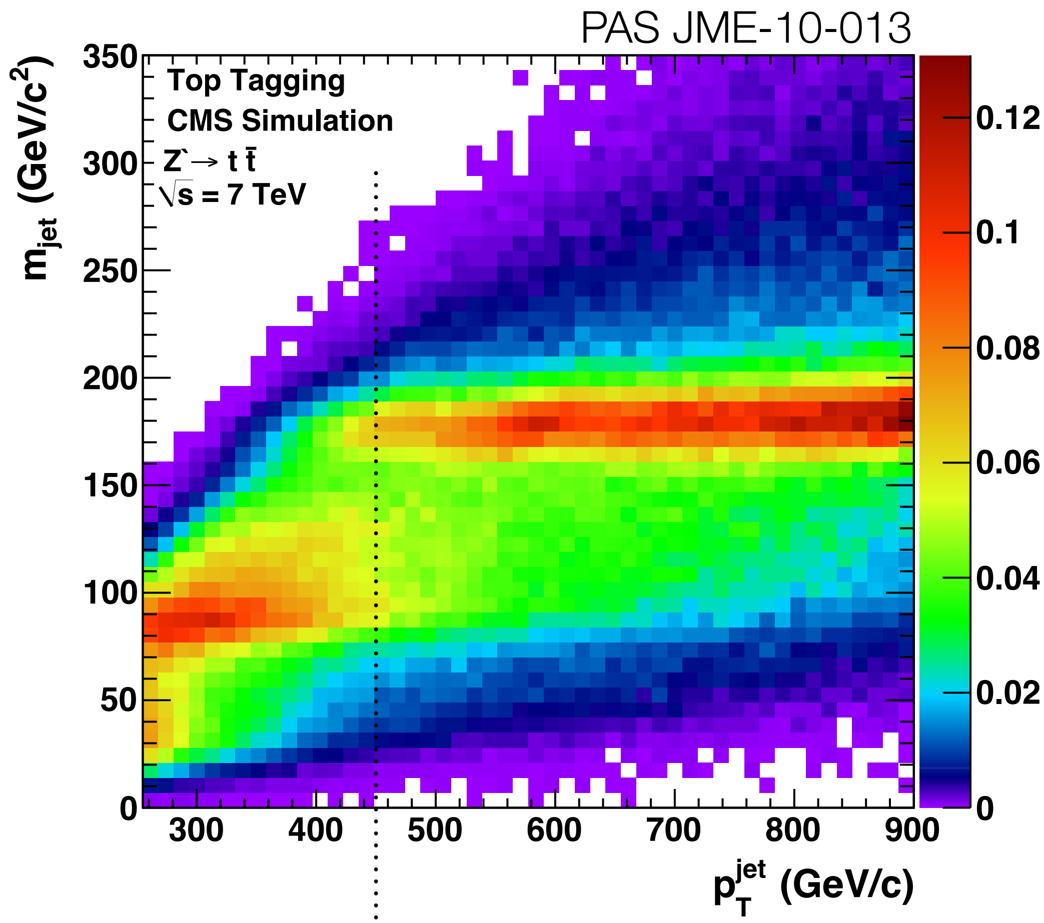
- The study of the top quark may give first insights into physics beyond the Standard Model
- Many models of new physics have large couplings to the top quark
  - Prefer the 3rd generation
- As we probe higher and higher mass scales, the phenomenology of the top quarks produced in collision events changes
  - Boosted regime
- We focus on model-independent searches
  - Narrow resonances
  - Overall enhancement in m_{tt} spectrum
- The LHC is a top factory -- a great place to carry out these searches!

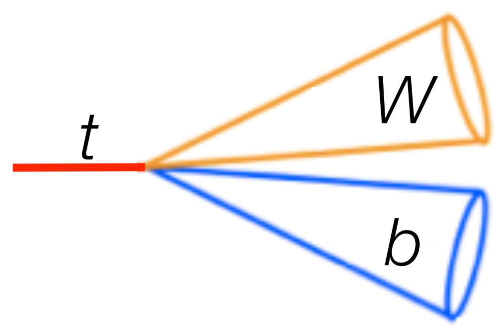




The Boosted Regime

- Searches covering different mass ranges call for different strategies
  - $X \rightarrow tt$ for example
- Low-mass searches (< ~1 TeV)
  - Decay products well-separated
  - Standard top quark methods used
- High-mass searches (> ~2 TeV)
  - Top quarks become boosted
  - Decay products collimated
  - Special reconstruction algorithms required
    - Jet substructure
- Intermediate mass range
  - Partially merged decay products
    - Mix of techniques





Boosted Top Algorithms

W Tagging Algorithm

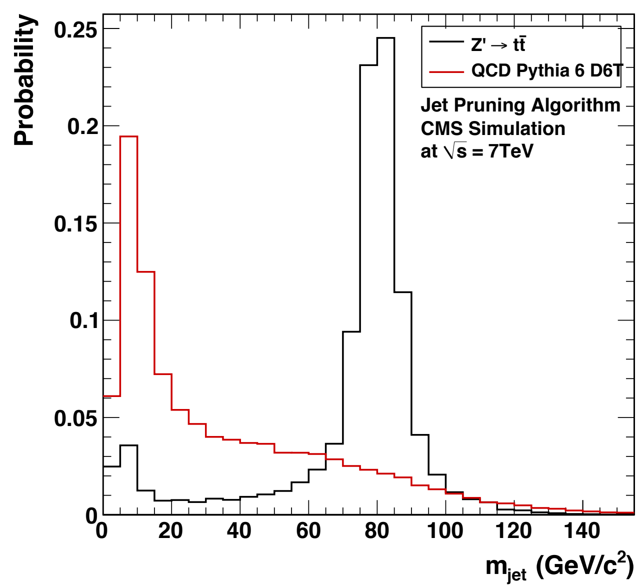
- The *W* tagging algorithm uses a jet pruning technique
  - J. Butterworth, A. Davison, M. Rubin, G. Salam -- PRL 100/242001 (2008)
  - S. Ellis, C. Vermilion, J. Walsh -- PRD 81/094023 (2010)
- Removes soft and large angle constituents of a jet
  - We use R = 0.8 jets
- In the clustering sequence, when comparing two constituents, the softer one is removed IF:

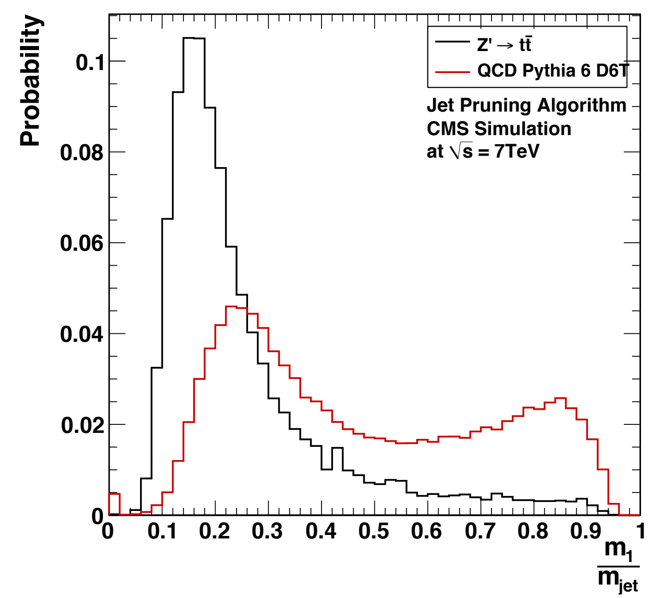
$$\Delta R_{ij} > R_{cut} \frac{m_j}{p_j}$$

AND: $$\frac{p_T}{p_T^{i+j}} < z_{cut}$$

- Default values are R_{cut} = 0.2, z_{cut} = 2%
- The *W* tagging algorithm requires
  - Jet mass in [60,100] GeV
  - Exactly 2 subjets
  - Mass drop μ < 0.4
    - Ratio of hardest subjet mass to total jet mass

PAS JME-10-013

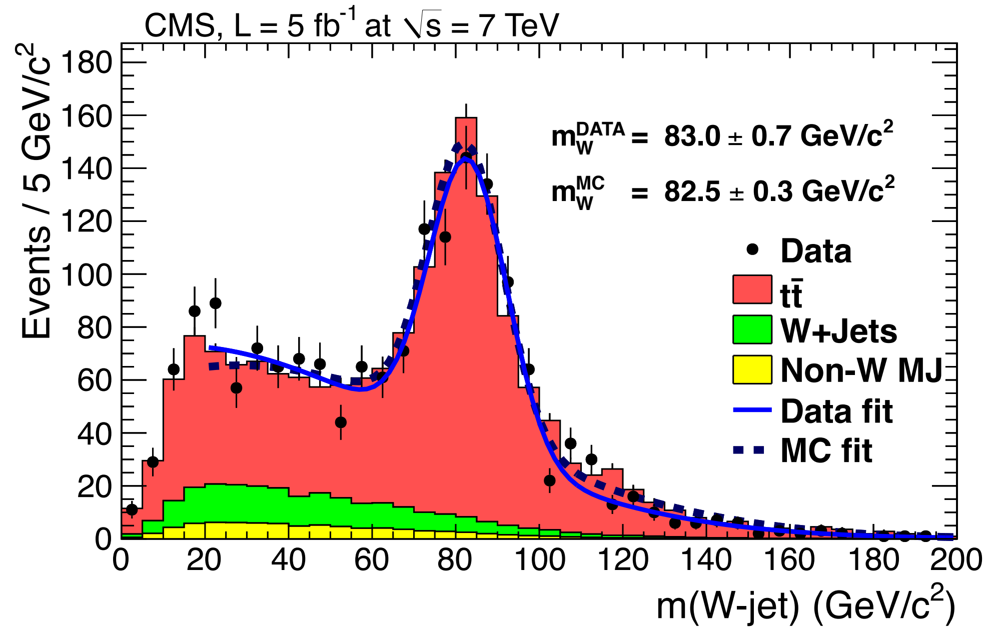




W Tagging Validation

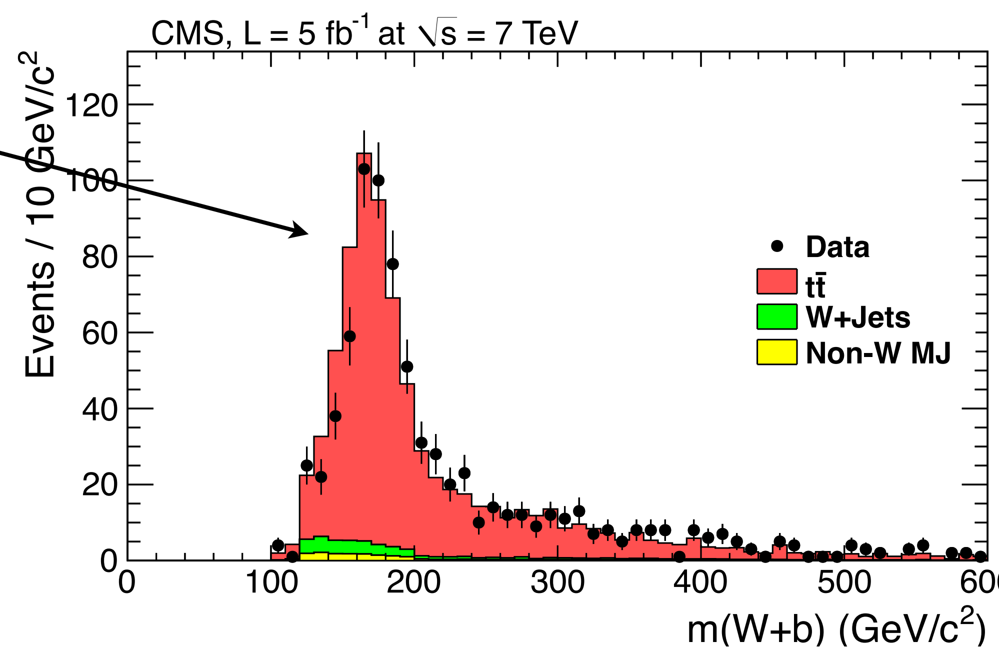
arXiv:1204.2488

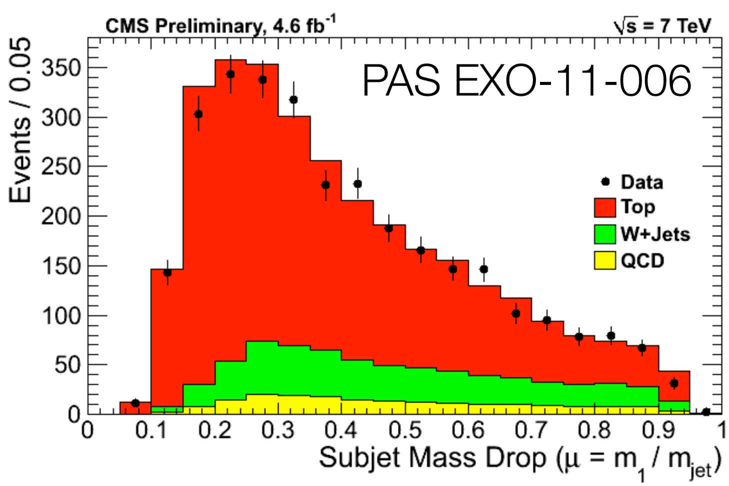
- In the Z' all-hadronic search, the *W* tagging algorithm is validated in data using a muon+jets selection
  - Used to determine efficiencies and scale factors for signal region



- Data agrees quite well with the Monte Carlo expectation

- These *W*-tagged jets are combined with an additional jet to form top candidates (Type 2)

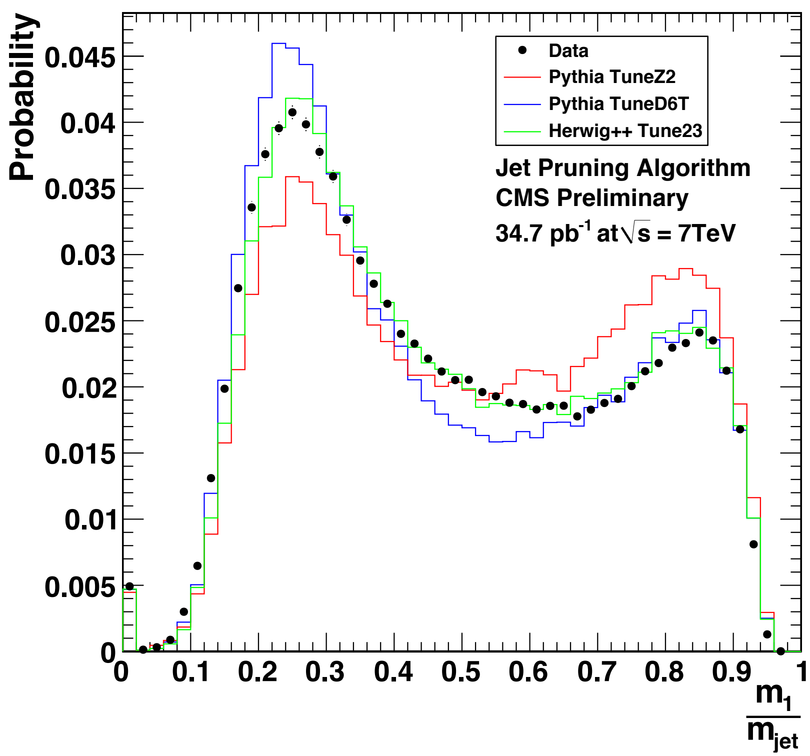


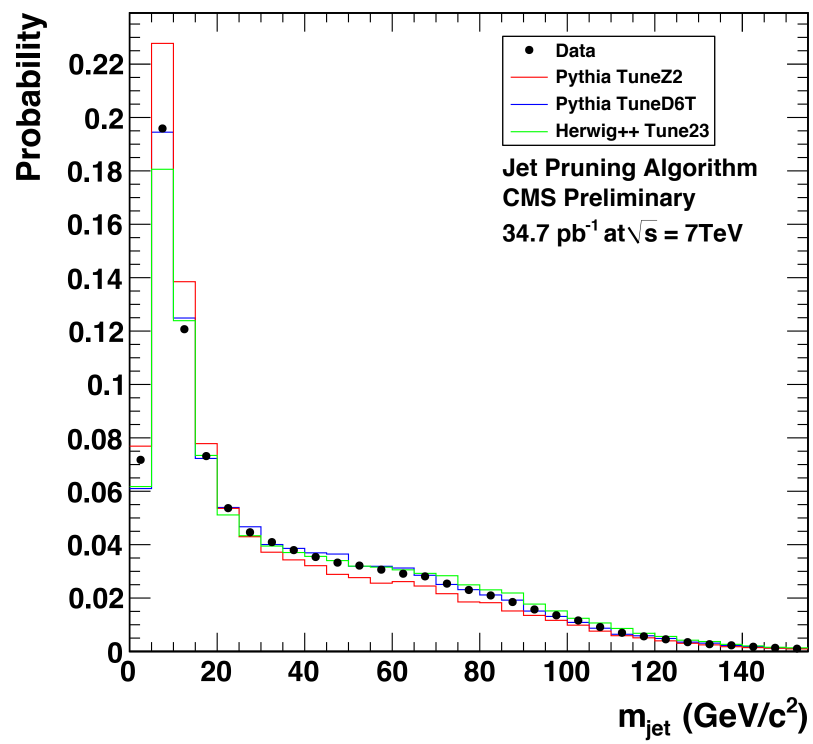


W Tagging Validation

PAS JME-10-013

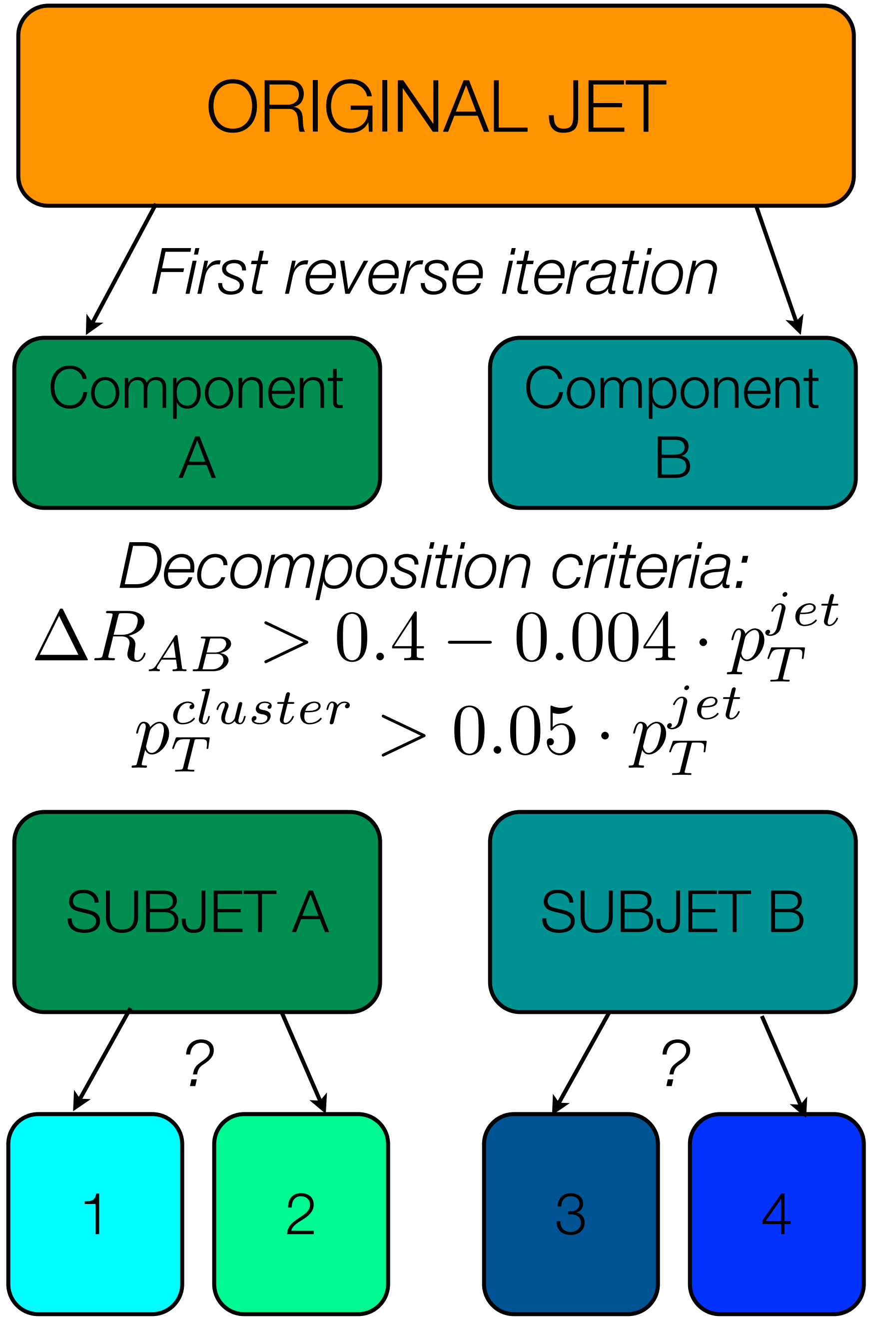
- The substructure quantities agree well in shape with the data distributions
  - Shown here for the leading p_T jet
  - All curves normalized to unit area
- Different generators plotted here
  - Herwig++ describes the mass drop variable well





Top Tagging Algorithm

- Based on JHU top tagging algorithm
  - Kaplan, Rehermann, Schwartz, Tweedie, PRL 101/142001 (2008)
- The algorithm uses jets with distance parameter *R* = 0.8, clustered with Cambridge-Aachen
- Uses cuts based on jet substructure information
  - Acquired by reversing the jet clustering algorithm
  - Step back in the pairwise sequence to find substructure
  - Can find a maximum of 4 subjets if all decomposition criteria are met
    - Optimized in simulation

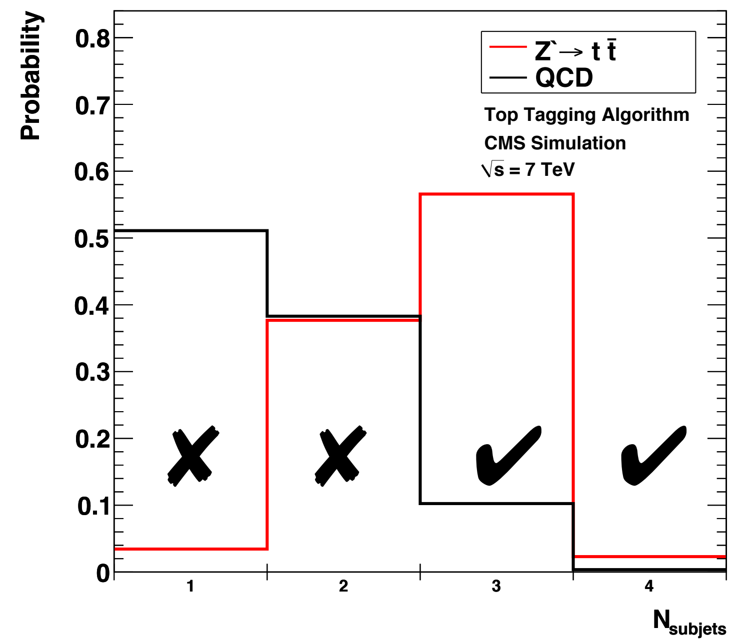


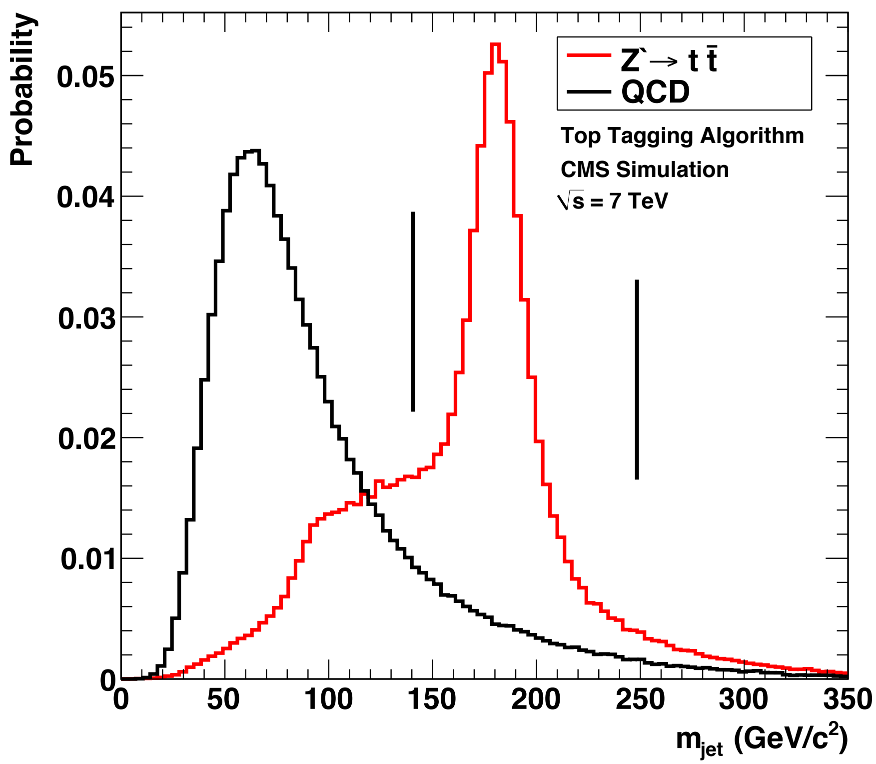
$$\Delta R_{AB} > 0.4 - 0.004 \cdot p_T^{jet}$$
$$p_T^{cluster} > 0.05 \cdot p_T^{jet}$$

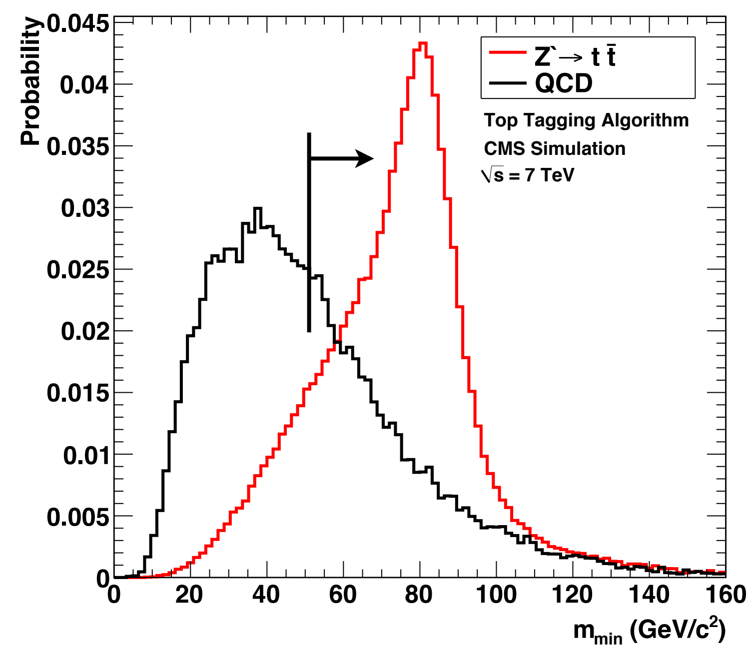
Top Tagging Algorithm

PAS JME-10-013

- Top-tagged jets are required to have
  - Jet mass in [140, 250] GeV
    - Consistent with top
  - 3 or 4 subjets
  - Minimum di-Subjet mass > 50 GeV
    - Consistent with *W*



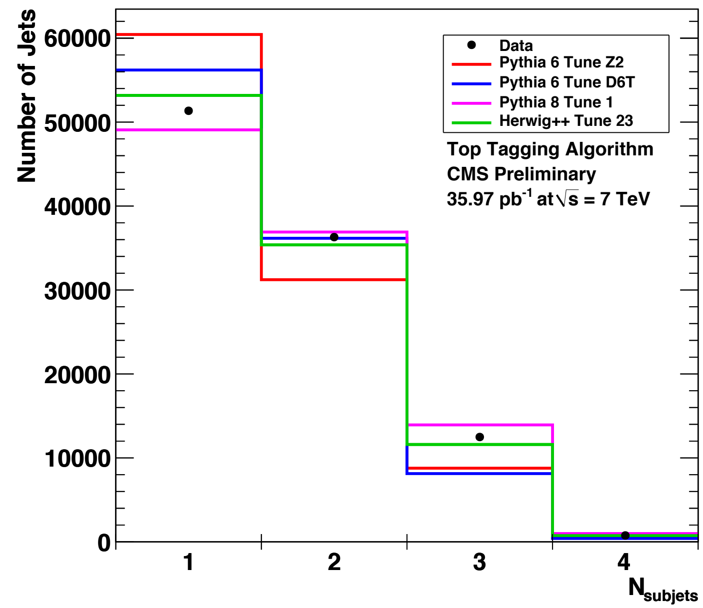


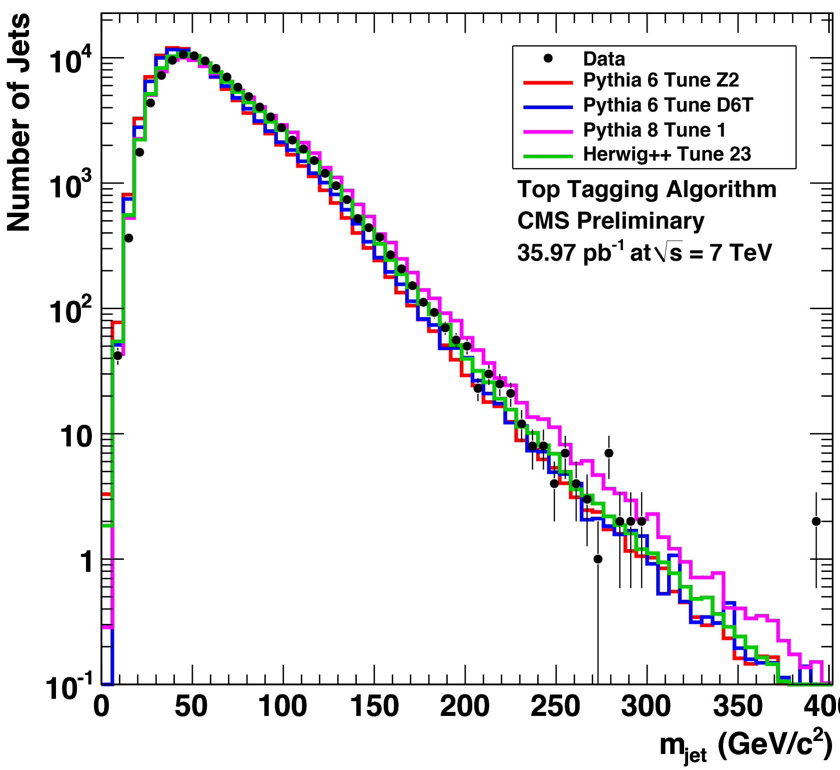


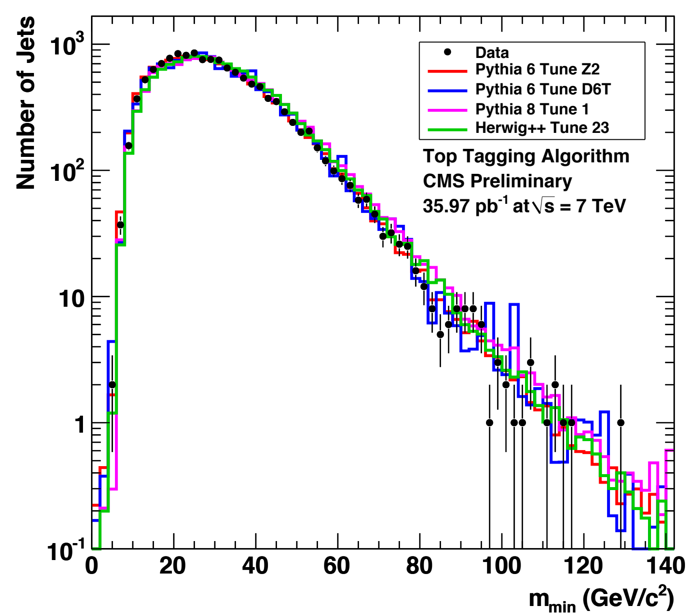
Top Tagging Validation

PAS JME-10-013

- Again the shapes of substructure quantities agree between data and MC
  - MC samples scaled to the number of observed data events

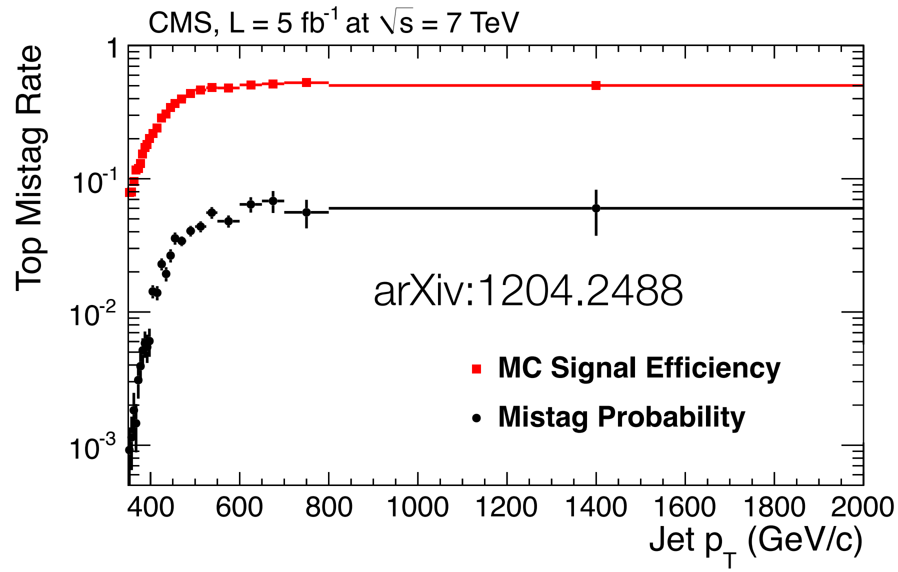


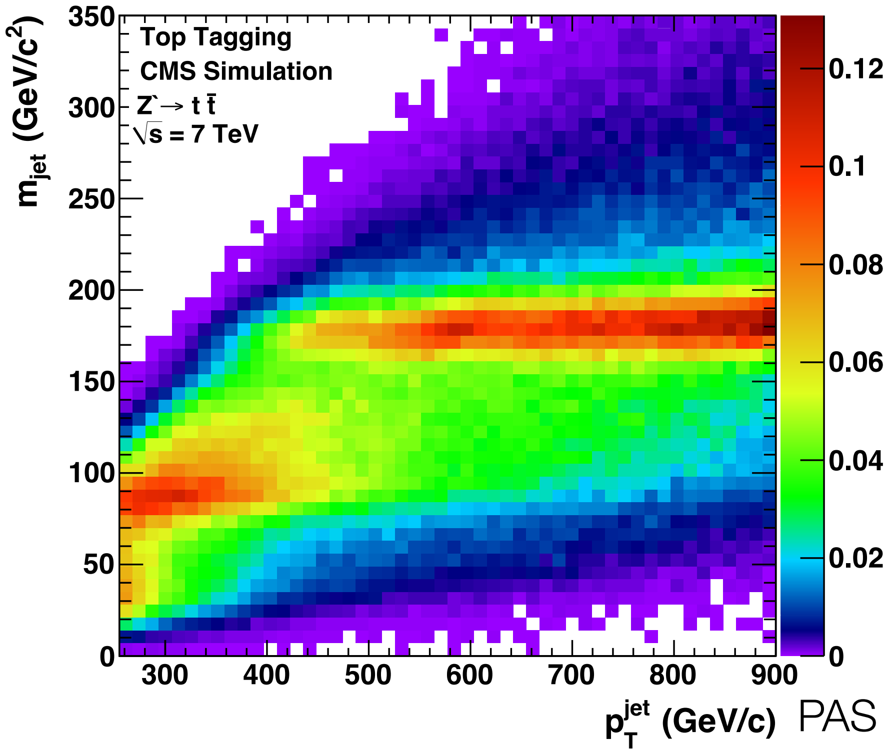




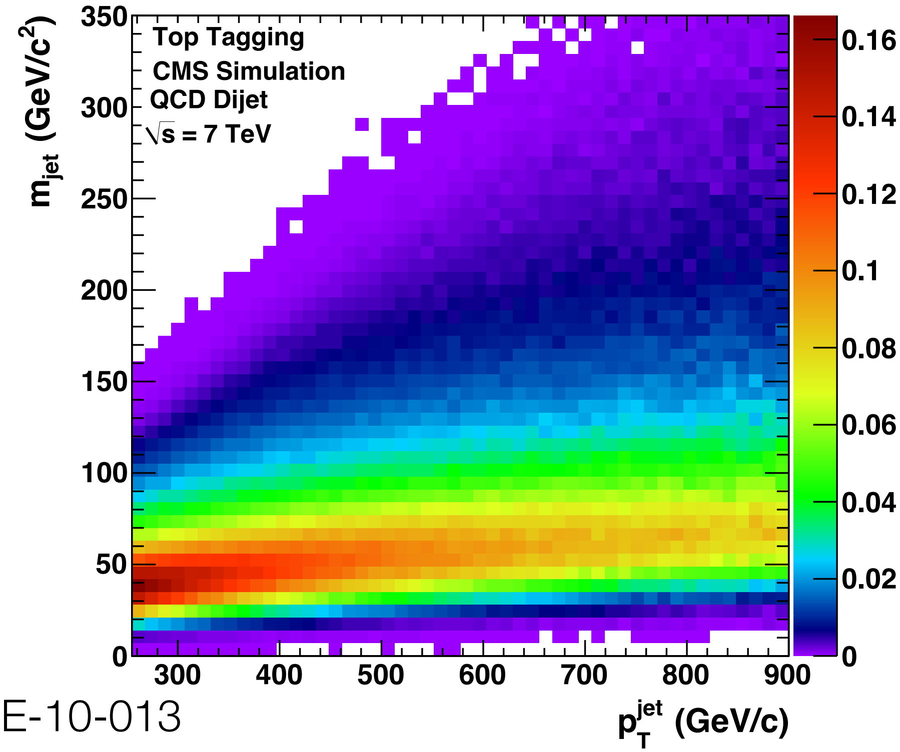
Top Tagging Algorithm

- Mistag rate can be determined by inverting the substructure requirements
  - “Anti-tag” and probe method
- QCD mistag rate is ~few percent or smaller across entire jet p_T range





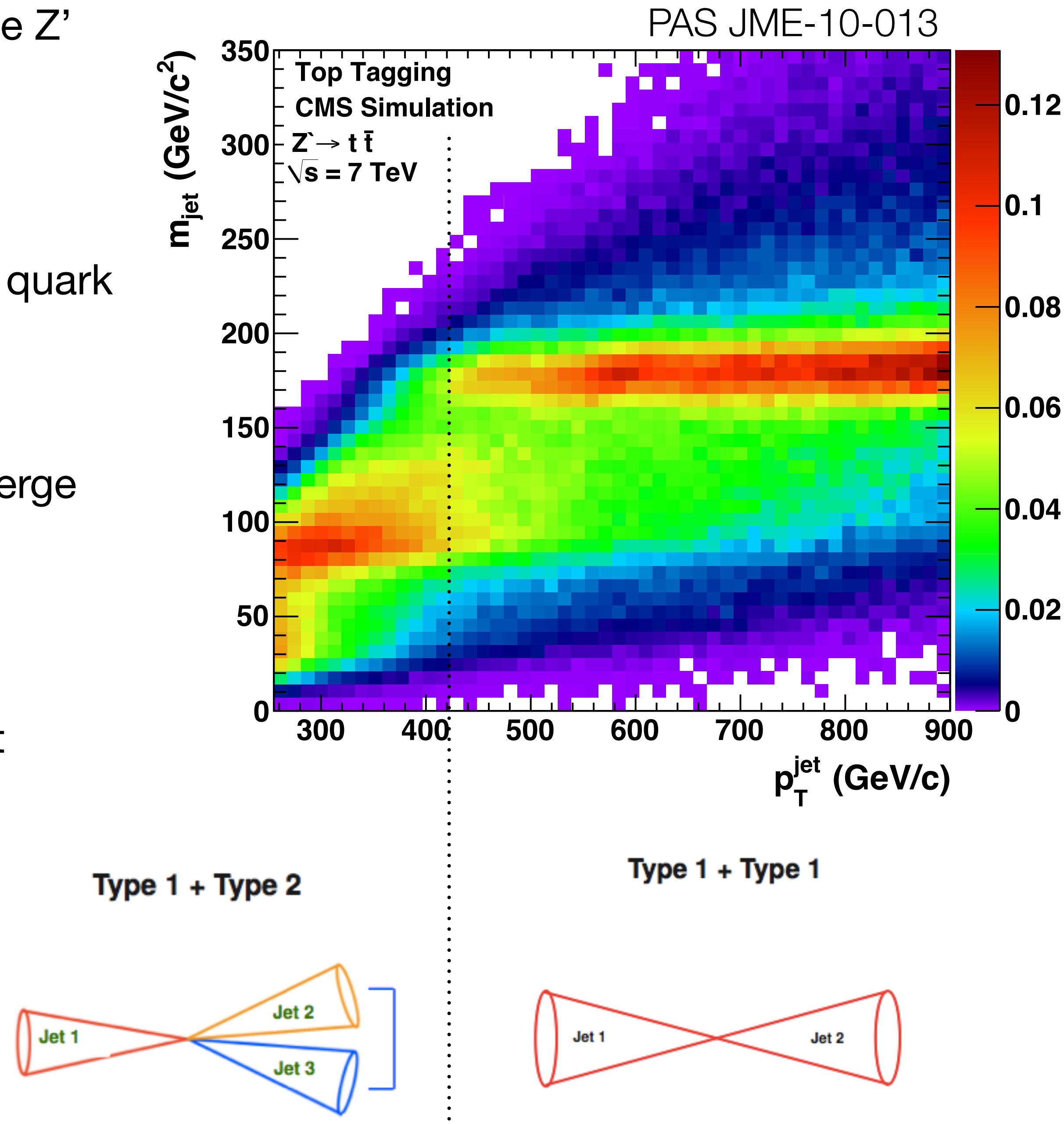
PAS JME-10-013



CMS Analyses and Results

All-Hadronic Z' Search

- Both top quarks produced by the Z' decay hadronically
  - Two top candidate types:
    - "TYPE 1"
    - All three jets from the top quark merge into a single jet
    - "TYPE 2"
    - The *W* decay products merge into a single jet
    - The *b* jet is reconstructed separately
  - Analysis considers two event topologies
    - Type 1 + 1
    - Type 1 + 2

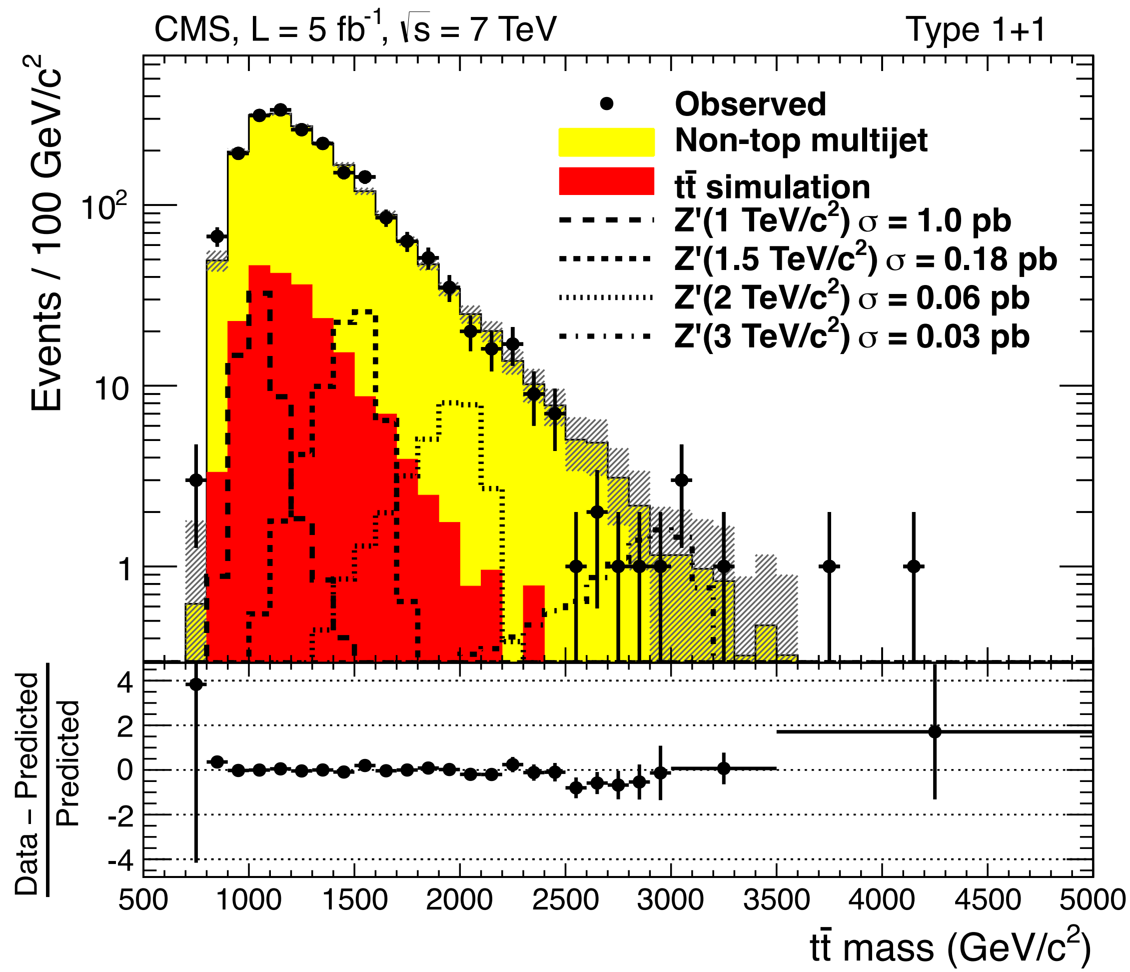


All-Hadronic Event Selection

arXiv:1204.2488

- Split event into two hemispheres based on Type 1 top candidate
  - Only utilize Type 1+1 and 1+2 events
- Type 1 selection
  - Jet p_T > 350 GeV, $|\eta|$ < 2.5
  - Passes Top Tagging criteria
- Type 2 selection (two jets)
  - Lead jet p_T > 200 GeV
  - Second jet p_T > 30 GeV
    - No *b*-tagging
  - Pass *W* Tagging criteria
- Main background is QCD
  - Determined by data-derived method, inverting selection criteria
  - Anti-tag and probe

- m_{tt} spectrum used to set limits

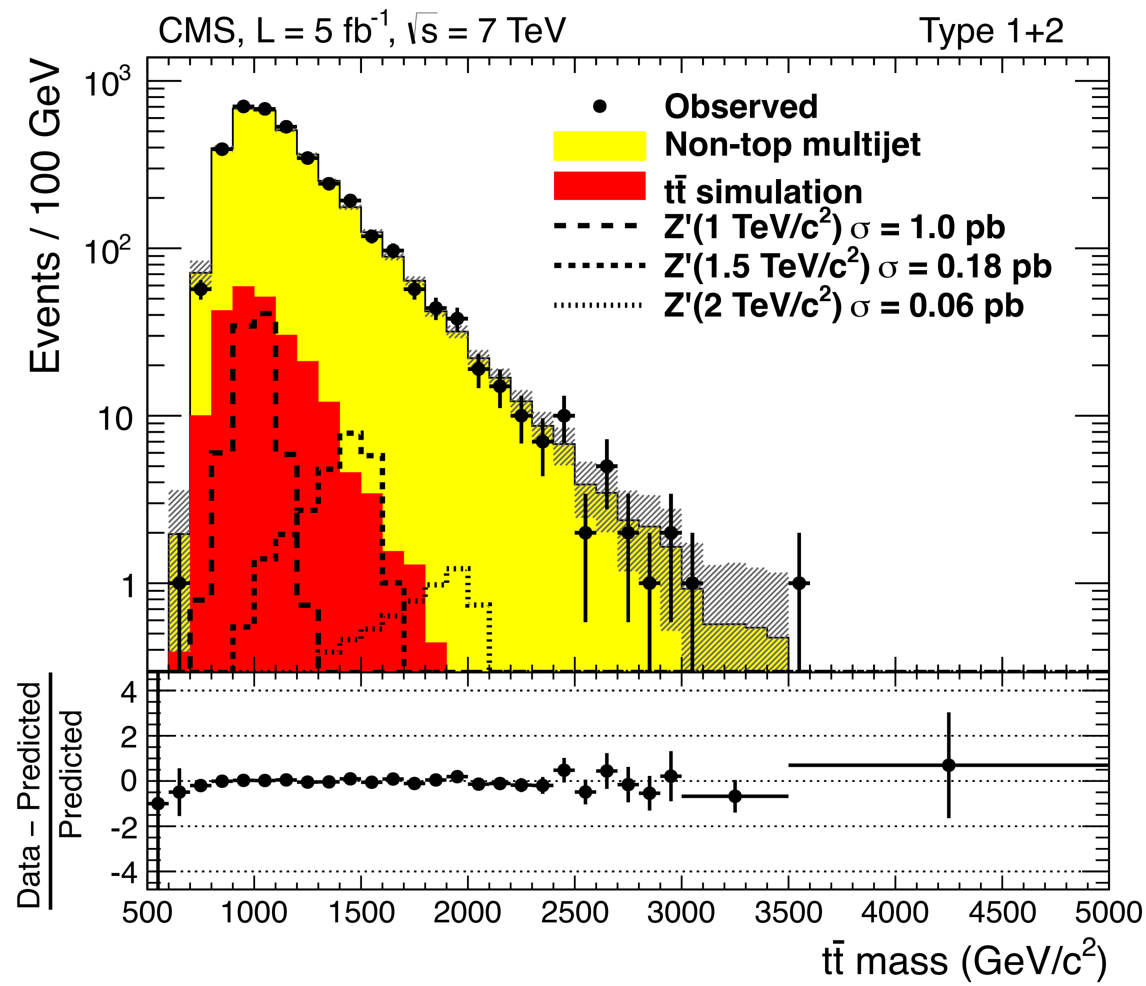


All-Hadronic Event Selection

arXiv:1204.2488

- Split event into two hemispheres based on Type 1 top candidate
  - Only utilize Type 1+1 and 1+2 events
- Type 1 selection
  - Jet p_T > 350 GeV, $|\eta|$ < 2.5
  - Passes Top Tagging criteria
- Type 2 selection (two jets)
  - Lead jet p_T > 200 GeV
  - Second jet p_T > 30 GeV
    - No *b*-tagging
  - Pass *W* Tagging criteria
- Main background is QCD
  - Determined by data-derived method, inverting selection criteria
  - Anti-tag and probe

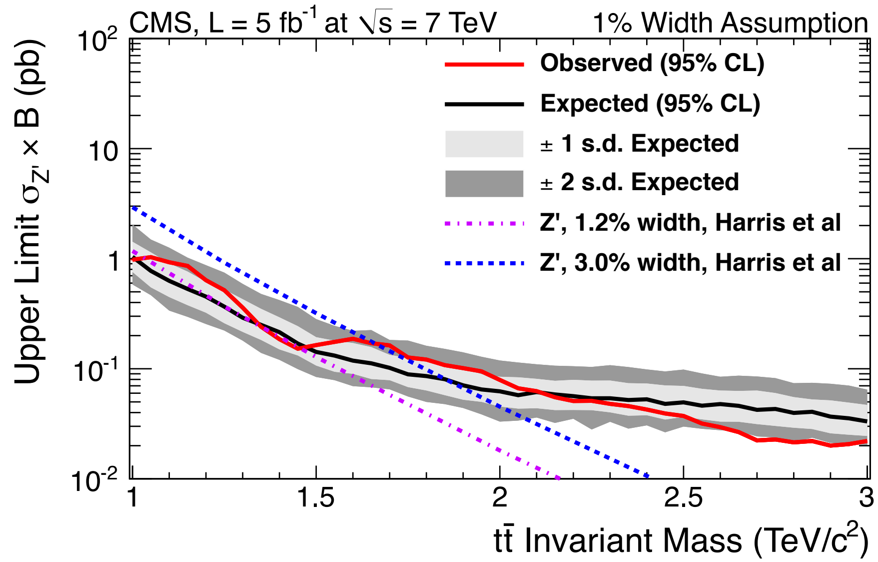
- m_{tt} spectrum used to set limits

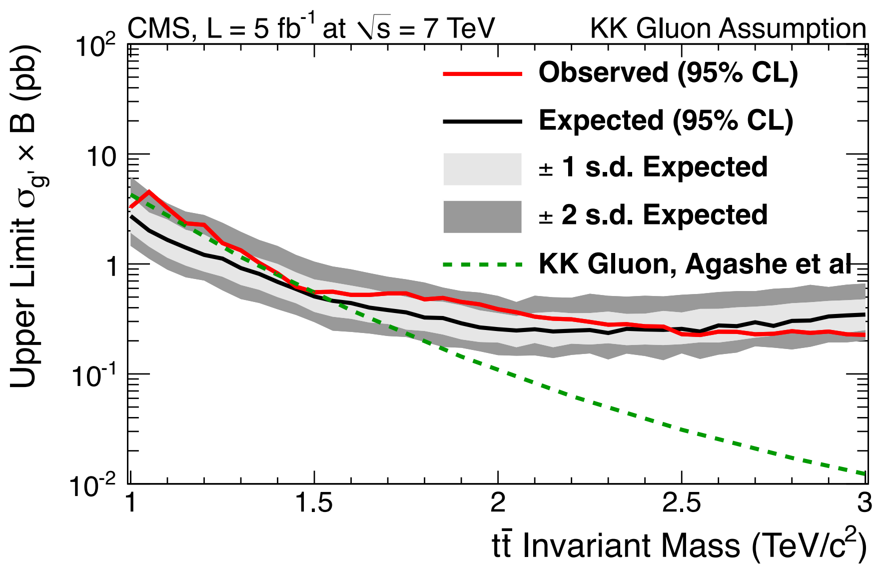


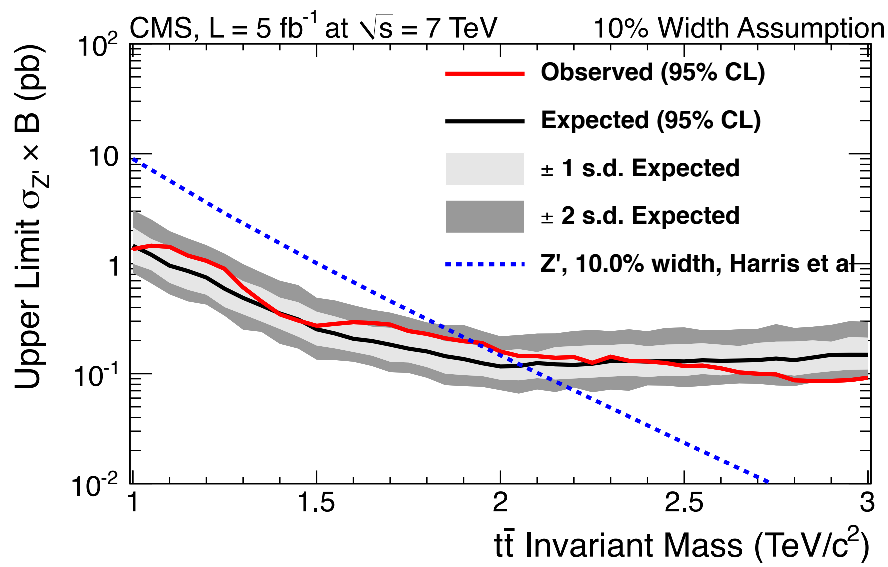
All-Hadronic Results

arXiv:1204.2488

- Shape analysis performed using several signal models
  - Z' narrow and wide resonances
  - KK Gluon production
- Mass ranges up to ~2 TeV are excluded
- Recently submitted for publication in JHEP
  - arXiv:1204.2488

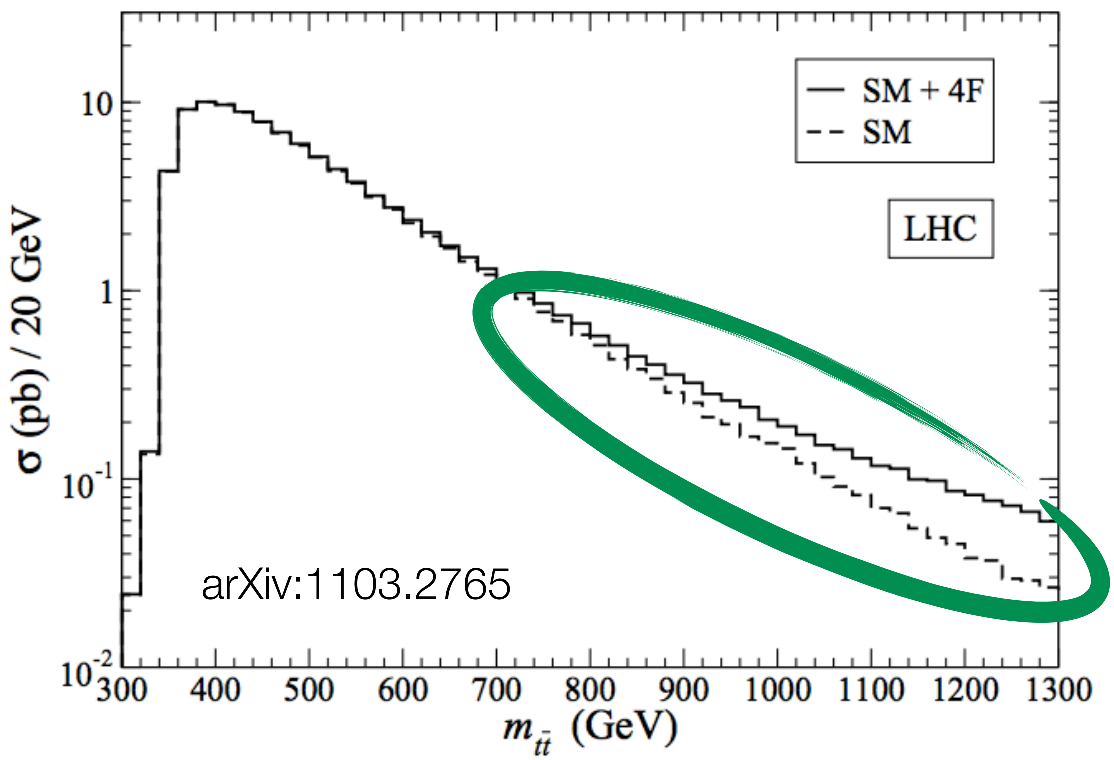






m_{tt} Enhancement Analysis

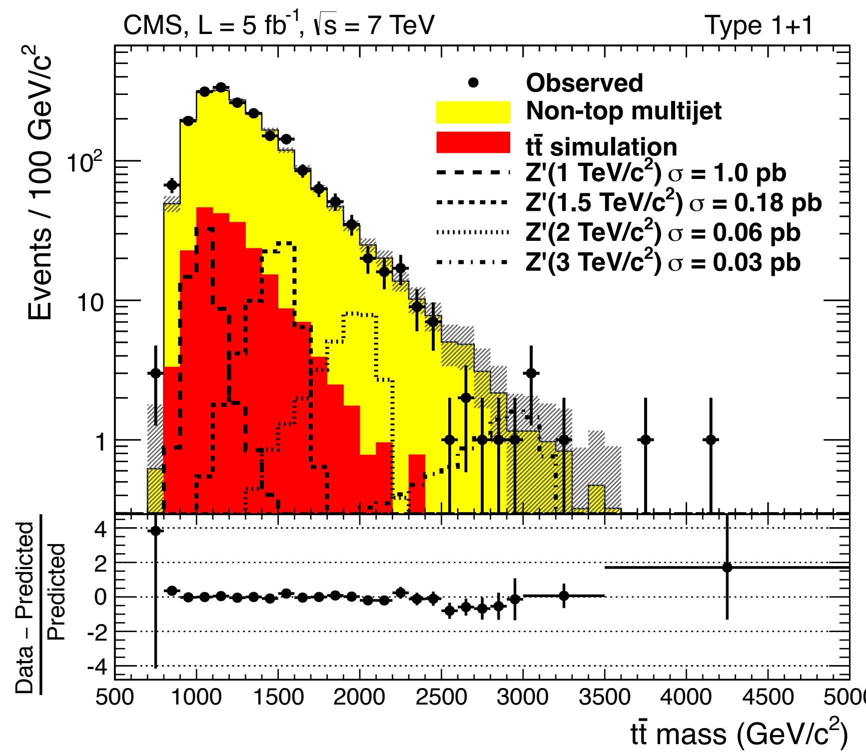
- Many new physics scenarios which explain the top forward-backward asymmetry also predict a general enhancement in the m_{tt} distribution
  - Non-resonant production due to interference terms with SM
- J. A. Aguilar-Saavedra, M. Perez-Victoria -- JHEP 1105/034 (2011)
- C. Delaunay, O. Gedalia, Y. Hochberg, G. Perez, Y. Soreq -- JHEP 1108/031 (2011)
- The all-hadronic result includes an analysis to set limits on the presence of such an enhancement



m_{tt} Enhancement Analysis

arXiv:1204.2488

- Use the same Type 1+1 and Type 1+2 events as in the Z' search
- Search for an enhancement above threshold mass
  - 1 TeV used here
- Need to use "true" top pair invariant mass to compare to theory predictions
  - Must correct the reconstructed mass to obtain true number of events above threshold
- Assume efficiency for new physics signal is the same as that for SM top pair production



	1+1	1+2
Expected SM $t\bar{t}$ events	194 ± 106	129 ± 80
Expected non-top multijet events	1546 ± 45	2271 ± 130
Total expected events	1740 ± 115	2400 ± 153
Observed events	1738	2423
$t\bar{t}$ efficiency	$(2.5 \pm 1.3) \times 10^{-4}$	$(1.6 \pm 1.0) \times 10^{-4}$

m_{tt} Enhancement Results

arXiv:1204.2488

- A counting experiment is performed for all events with m_{tt} > 1 TeV
- Limit set on *S* -- the ratio of number of events above 1 TeV with the enhancement to the SM contribution
- Observed *S* < 2.6 at 95% CL
  - Expected 1σ range [2.0, 3.5]
  - Expected 2σ range [1.7, 5.5]
  - Expected *S* = 2.5
- Models predicting enhancements of more than 2.6x the SM top pair production rate are excluded

$$\mathcal{S} = \frac{\int_{m_{t\bar{t}}>1\,\mathrm{TeV}/c^2} \frac{\mathrm{d}\sigma_{SM+NP}}{\mathrm{d}m_{t\bar{t}}} \mathrm{d}m_{t\bar{t}}}{\int_{m_{t\bar{t}}>1\,\mathrm{TeV}/c^2} \frac{\mathrm{d}\sigma_{SM}}{\mathrm{d}m_{t\bar{t}}} \mathrm{d}m_{t\bar{t}}}$$

< 2.6 at 95% CL

Z' Searches in Lepton + Jets

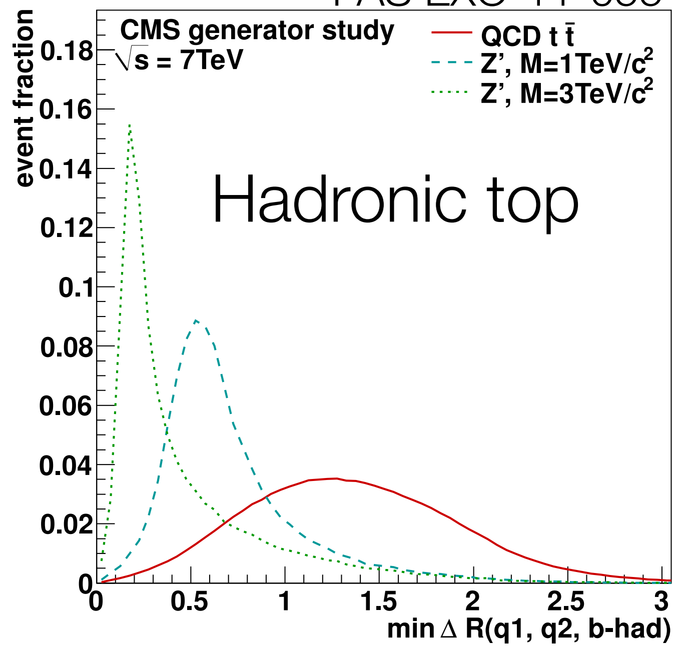
- Search for Z'→tt
  - One hadronic top
  - One leptonic top

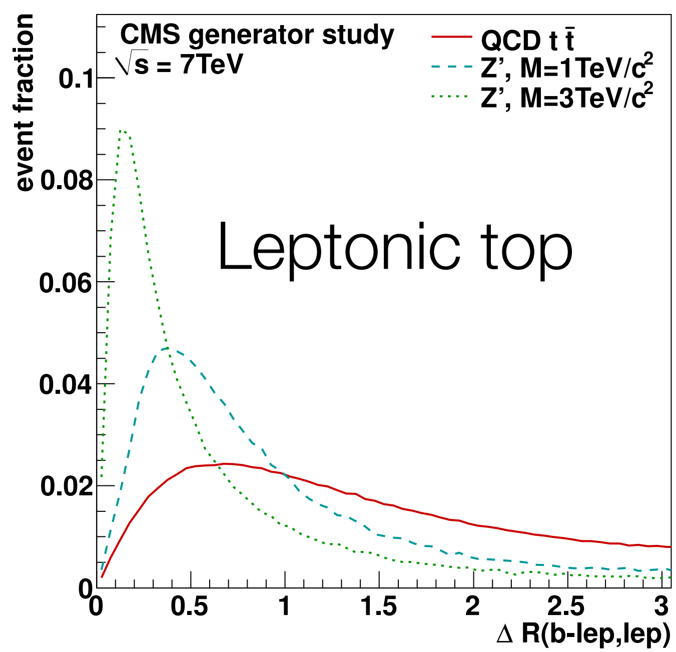
- Basic selection
  - Exactly one high-quality lepton
    - Electron $p_T > 70$ GeV, $|\eta| < 2.5$
    - Muon $p_T > 35$ GeV, $|\eta| < 2.1$
    - Reduce di-leptonic top pair and Z backgrounds
    - Special isolation requirements

  - Two or more jets
    - Leading jet $p_T > 250$ (150) GeV
    - Other jets $p_T > 50$ GeV

  - In the boosted regime, the lepton becomes closer to a jet, and the jets may merge
    - Special techniques to reduce QCD contribution
    - Still allows for high selection efficiency of boosted objects

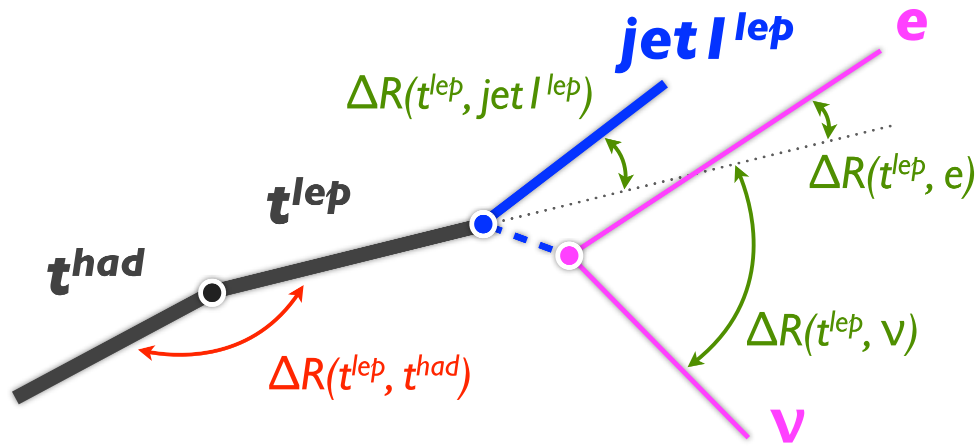
PAS EXO-11-055





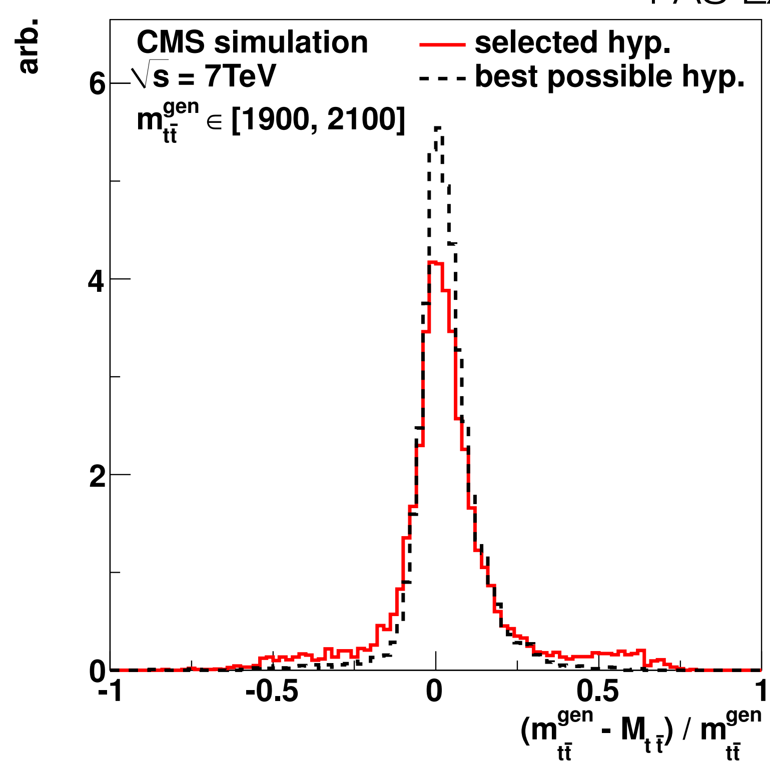
Event Reconstruction

- With one lepton and several jets, how to find the right combinations to form top candidates?
- Solve for the missing component of the neutrino momentum
- Form hypotheses by assigning jets to either the leptonic or hadronic top quark candidate
  - Choose the hypothesis with the minimal sum ΔR between leptonic decay products and the leptonic top candidate
  - If more than one, maximize ΔR between top candidates
- Comparison with simulation shows this method performs well



PAS EXO-11-092

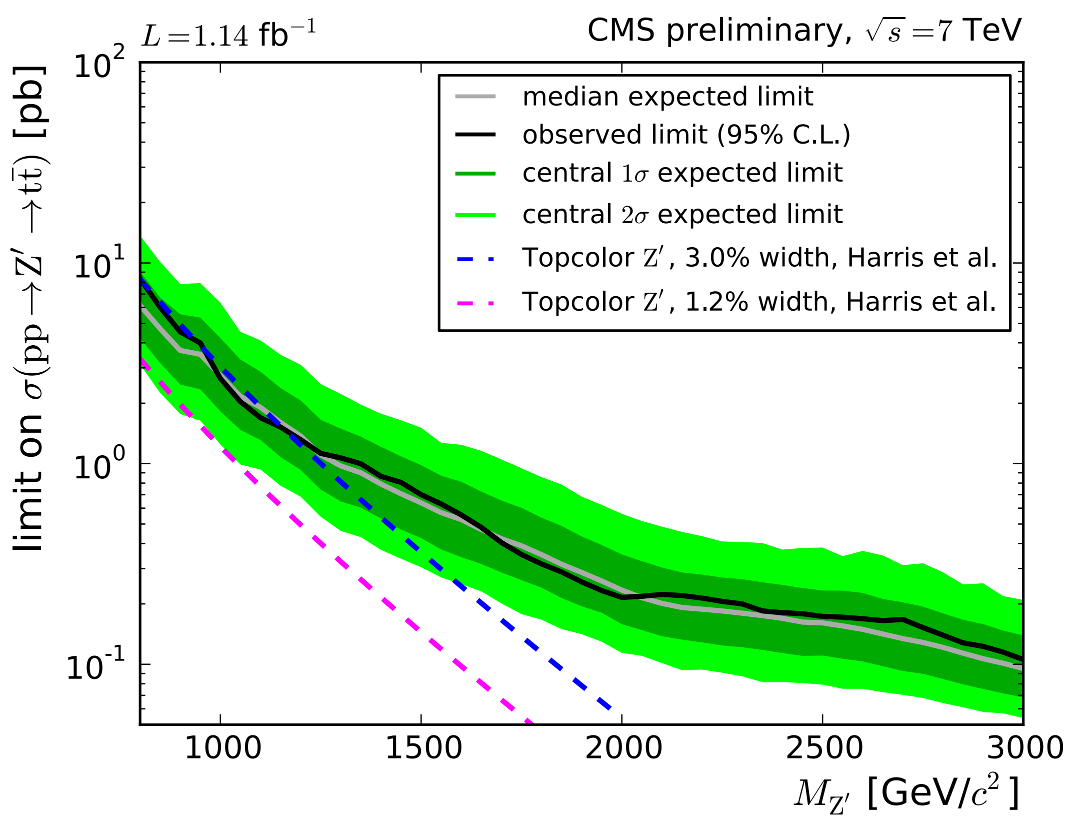
PAS EXO-11-055

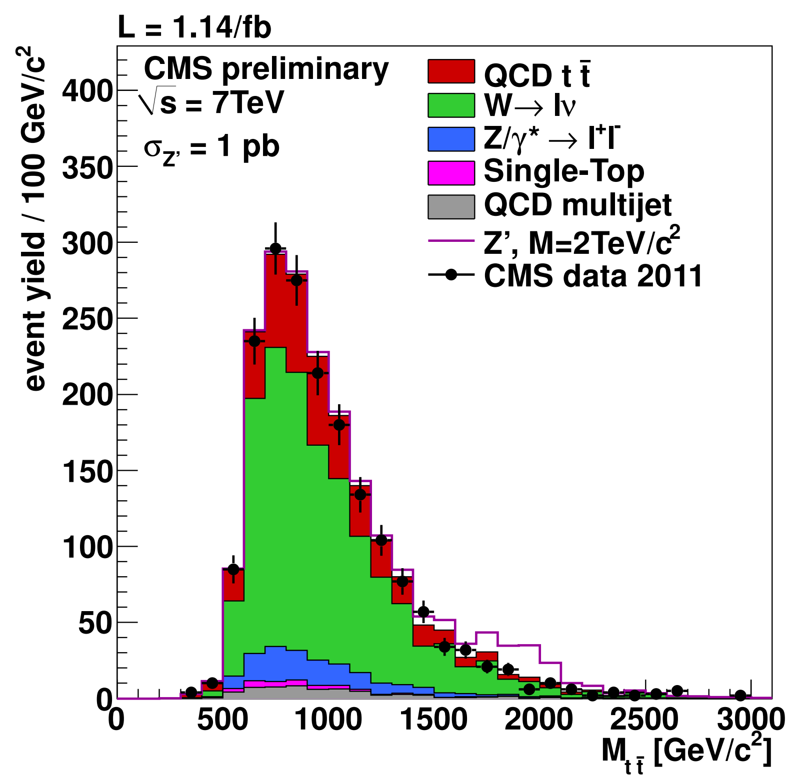


Muon + Jets Results

PAS EXO-11-055

- Signal region defined by cut
  - H_T^{lep} = muon p_T + MET > 150
- Simultaneous fit to low H_T^{lep} and m_{tt}
- Bayesian limit-setting procedure to obtain 95% CL limits
- Limits exclude cross sections of < 1 pb above $m_{Z'}$ ~ 1.3 TeV
- Update coming soon with full dataset!

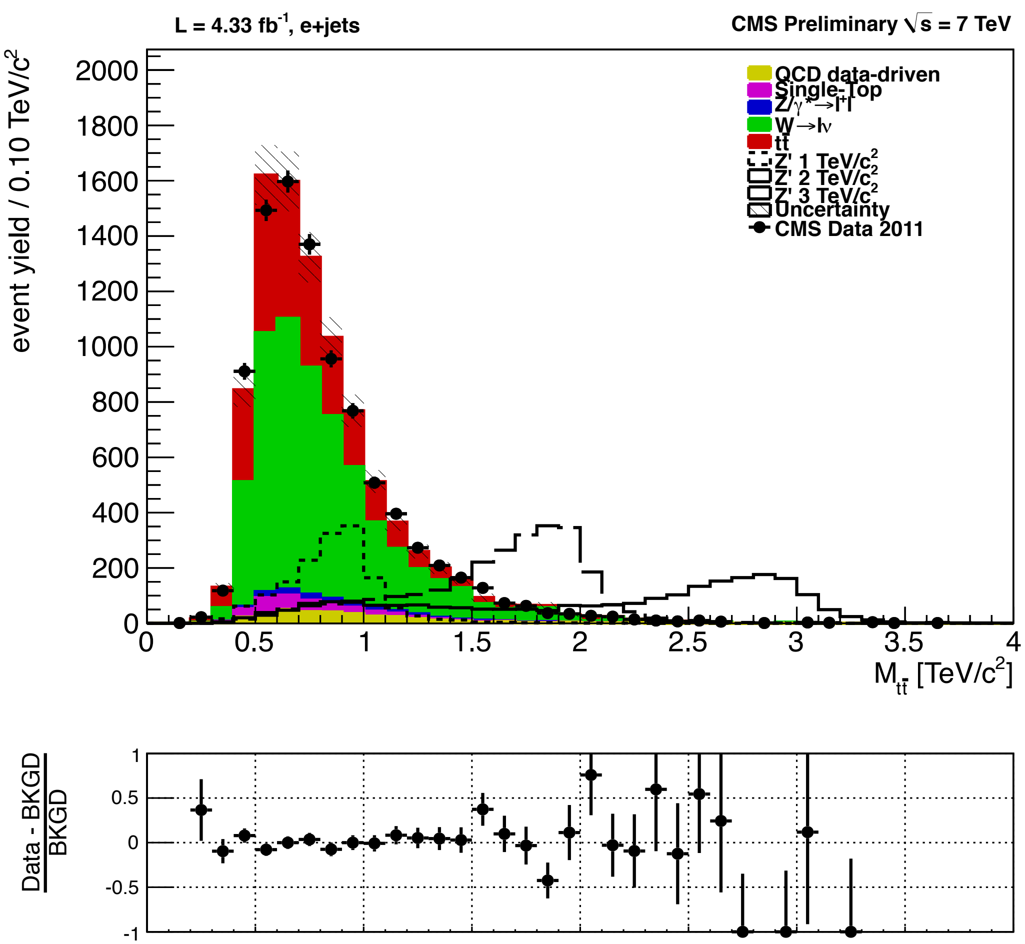


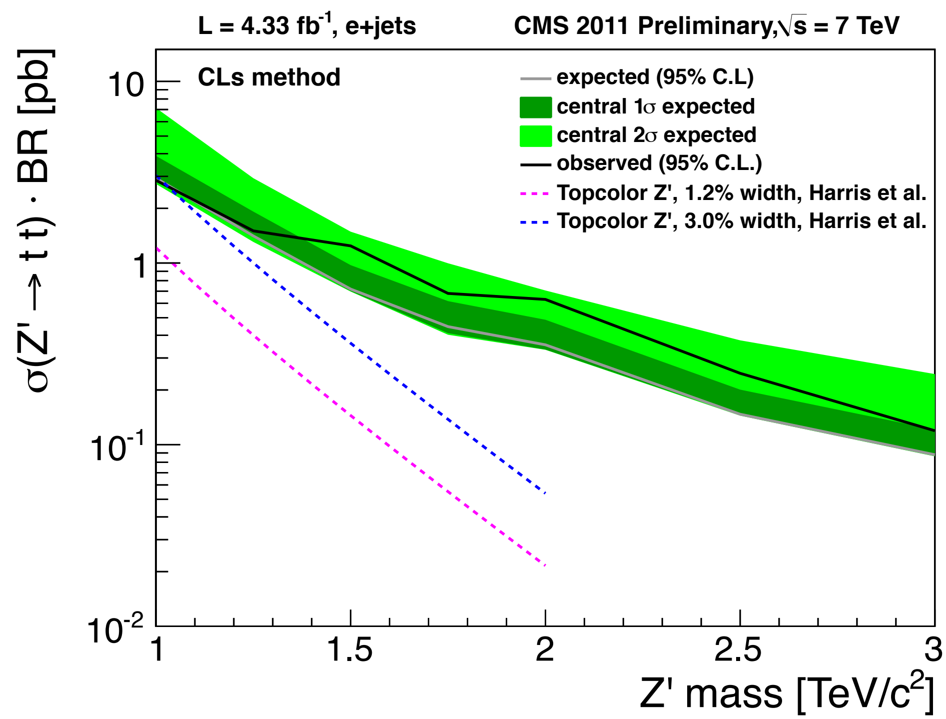


Electron + Jets Results

PAS EXO-11-092

- QCD normalization determined from fit to MET distribution
- m_{tt} spectrum used for limit setting
- CLs method used to obtain 95% CL limits on Z' production cross section
- Again cross sections < 1 pb are excluded for high masses

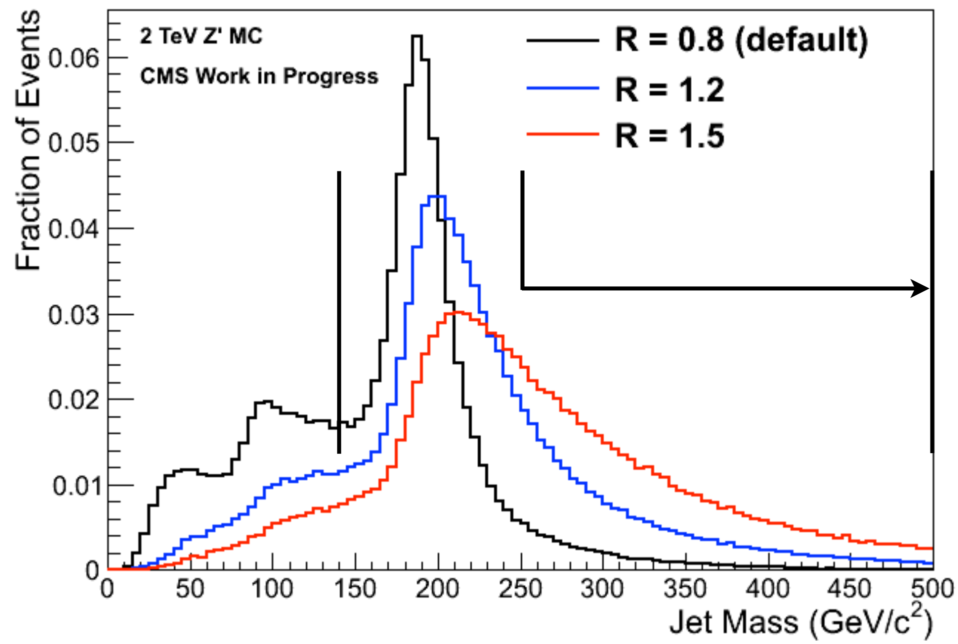




Future Plans

Top Tagging Improvements

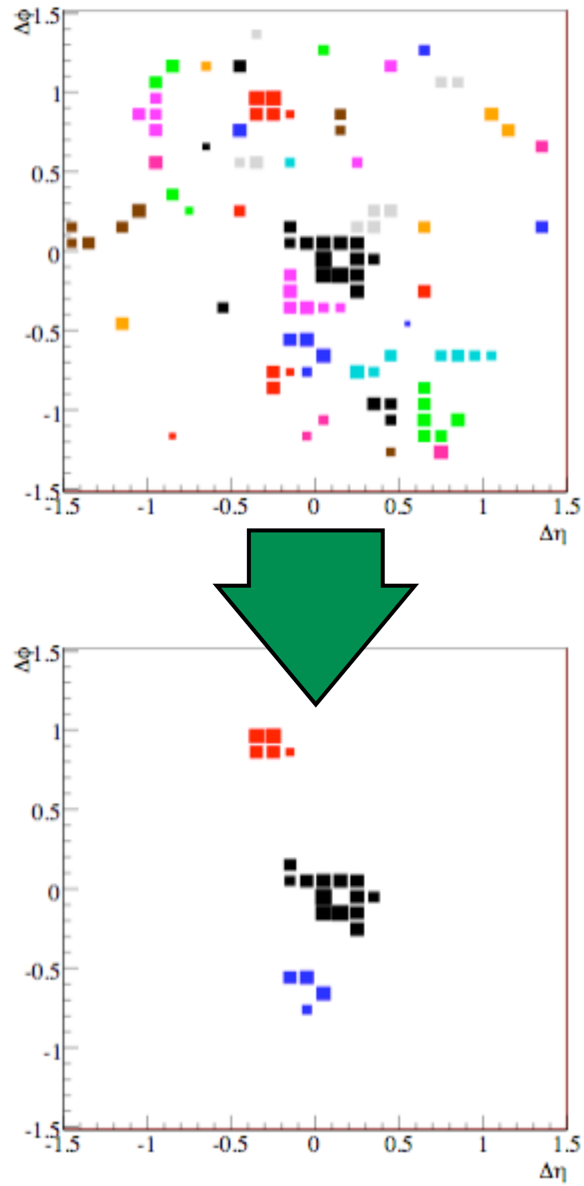
- Investigating increasing the distance parameter of jets used for the algorithm
  - Captures additional top quarks that may not be as boosted
- Modification of the algorithm necessary
  - Extending the bounds of the jet mass window due to observed behavior
  - Can be more efficient for boosted tops
- Possible efficiency gain for Z' signal!
- However, QCD rate needs to be kept under control

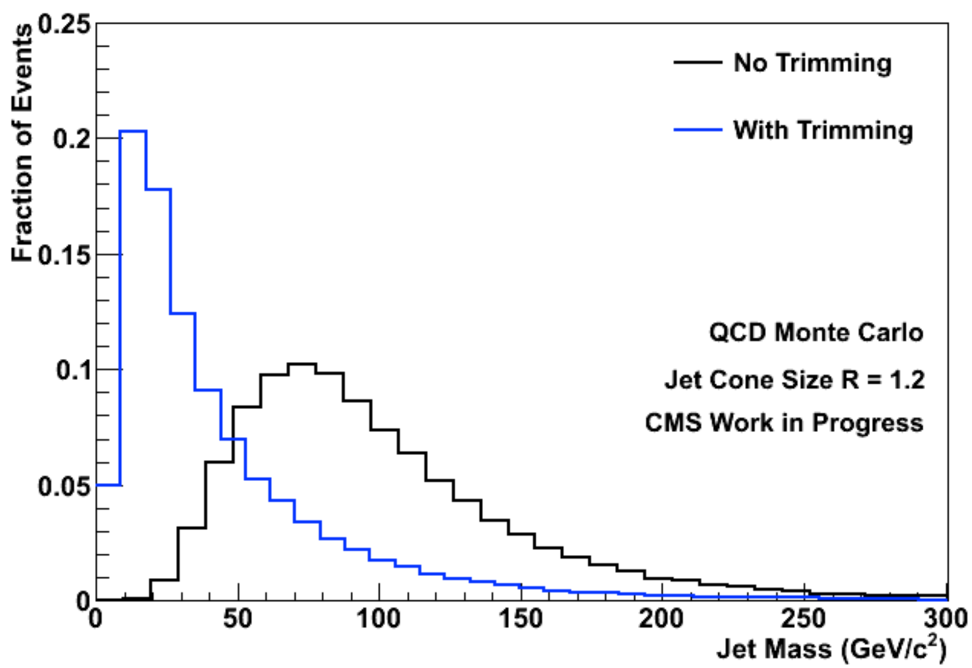


QCD Rejection

- Studying Jet Trimming
  - Krohn, Thaler, Wang -- JHEP 1002/084 (2010)
  - Clusters jets into subjets with a small *R* parameter (0.2 is default)
  - Removes those subjets with less than 3% of the total jet p_T
- Signal efficiency similar, but mistag rate reduced by a large amount
  - Reduced jet mass

Figures from J. Thaler



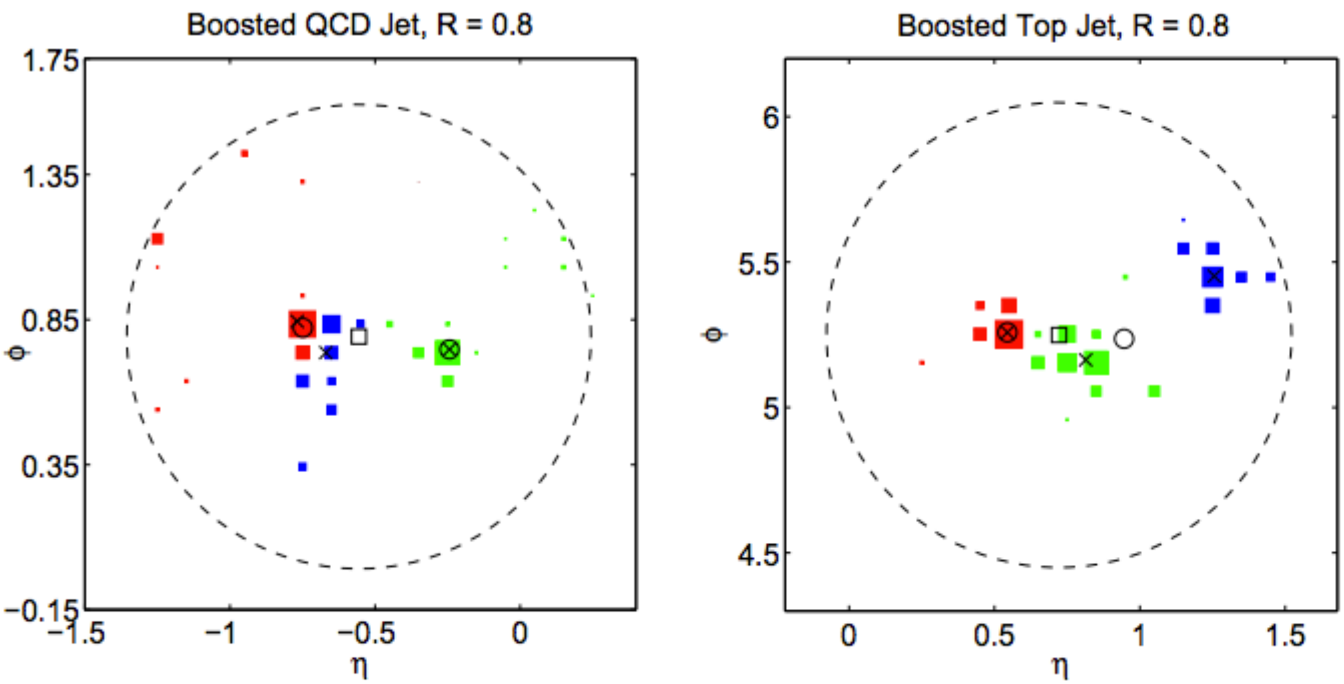


N-Subjettiness

- Additional signal discriminating power could be obtained from N-subjettiness
  - J. Thaler, K. Van Tilburg, arXiv:1011.2268
- Determines the consistency of the particles in the jet with N number of subjets
- Requires clustering the jet into N exclusive subjets

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \cdot \min(\Delta R_{1,k}, \cdots, \Delta R_{N,k})$$

Normalization factor (d_0) — Jet constituent index (k) — Subjet index (N)



Subjet positions after clustering into *N* subjets:
□ *N* = 1
○ *N* = 2
× *N* = 3

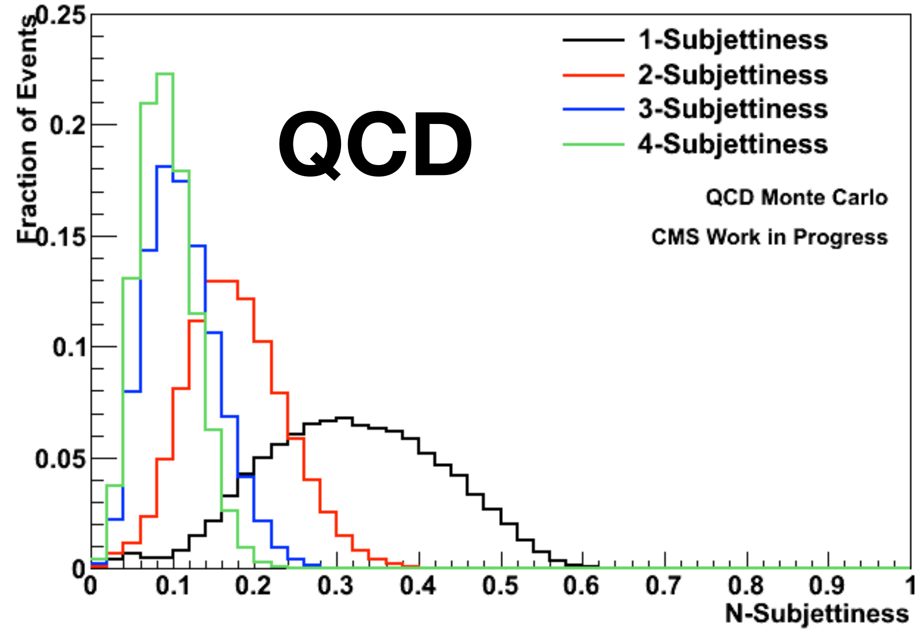
From arXiv:1011.2268

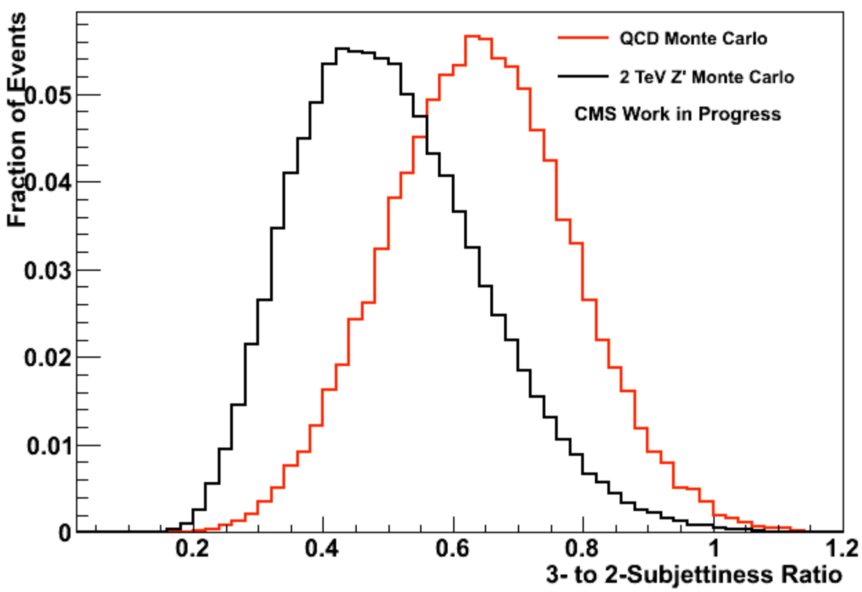
N-Subjettiness

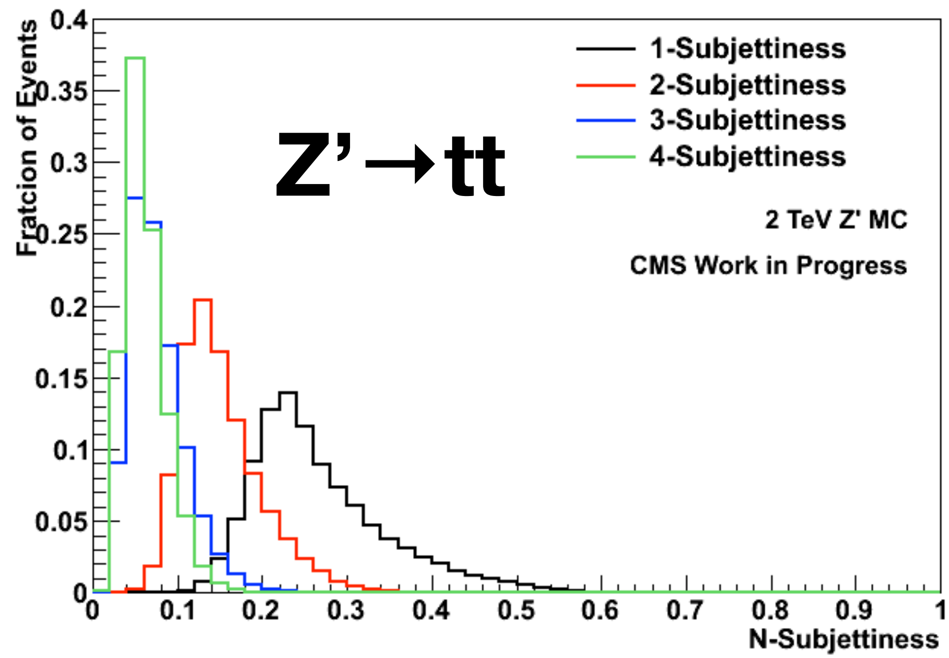
- First look shows promise
- Possibility to combine this with other quantities in an MVA method for enhanced discrimination

- Ratios more useful than individual values
  - τ_2 / τ_1 can distinguish *W* jets
  - τ_3 / τ_2 can distinguish top jets

- Jets here pass top tagging requirements

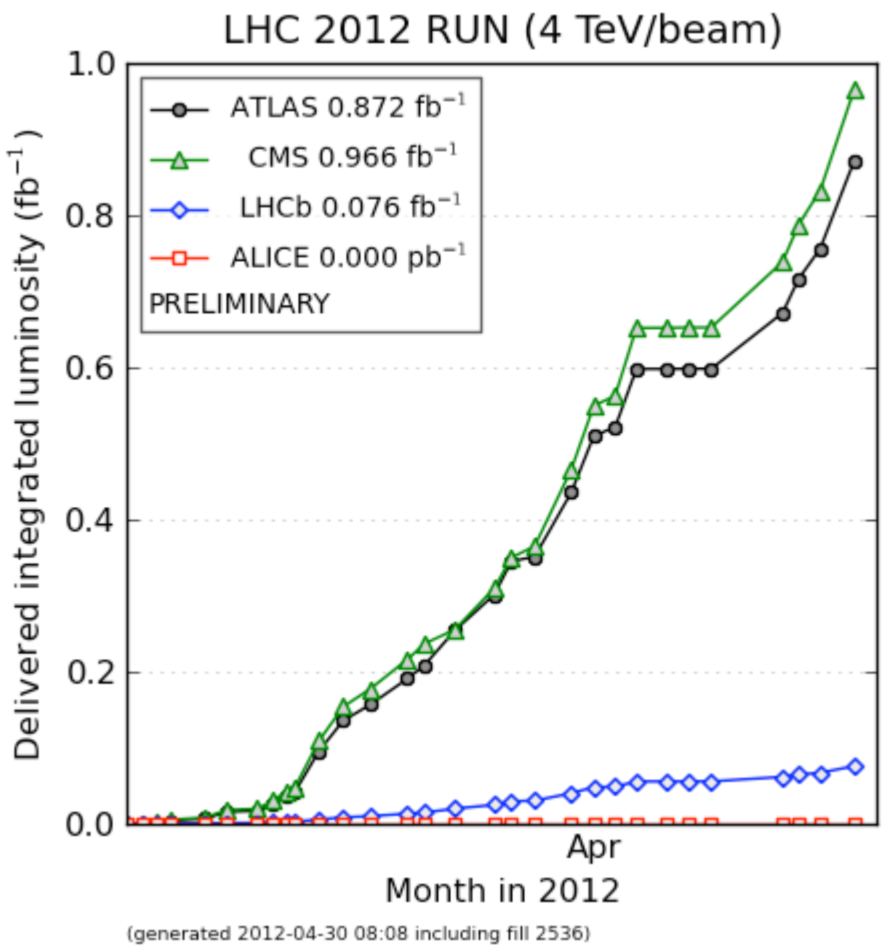






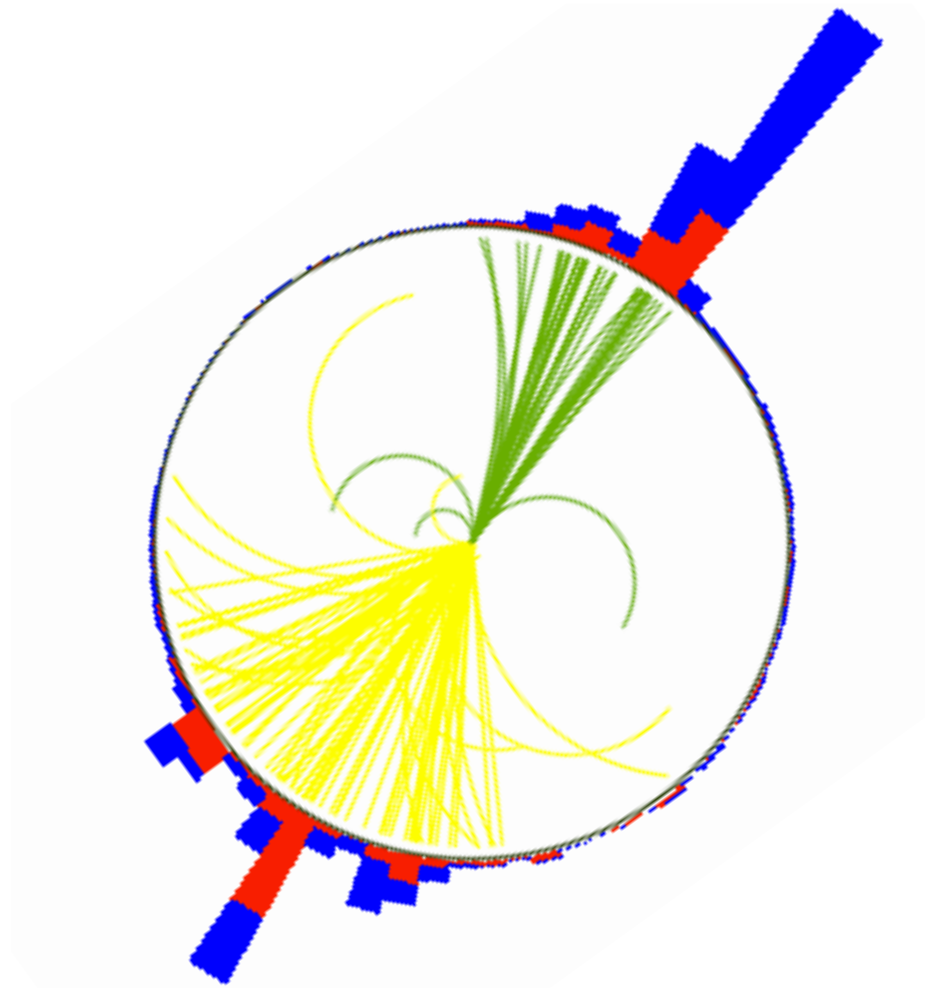
Planned Updates

- Boosted top analyses are planning updates for this coming summer
  - Addition of *b*-tagging to searches
  - Full 5 fb^{-1} dataset for μ+jets channel
  - Improved top mass reconstruction techniques
- Results using the new 8 TeV dataset
  - Already ~1 fb^{-1} delivered!
  - Quick updates with existing methods underway (for summer)
  - New analyses with planned improvements later this year
- It will be an exciting year to say the least!



Conclusions

- CMS has been active in analyses involving boosted top signatures
- Many new techniques have been developed, validated, and implemented into the searches
- We continue to utilize the latest ideas from the theory community to improve our sensitivity to new physics processes as much as possible
- Stay tuned for updates as we continue to collect and analyze more data!



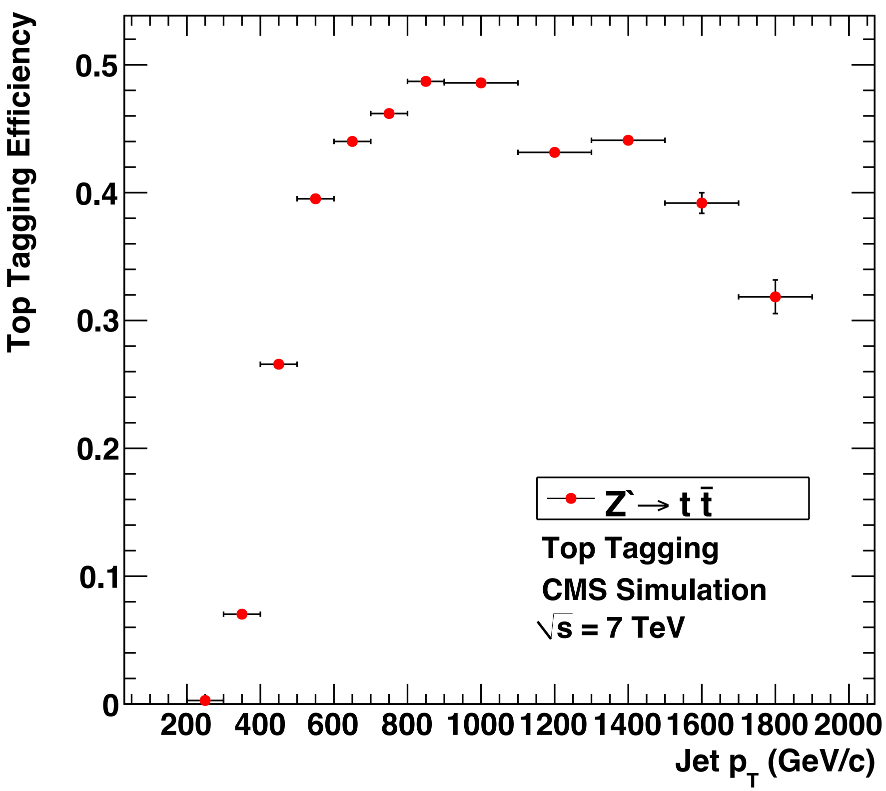
- Thanks very much for your attention!

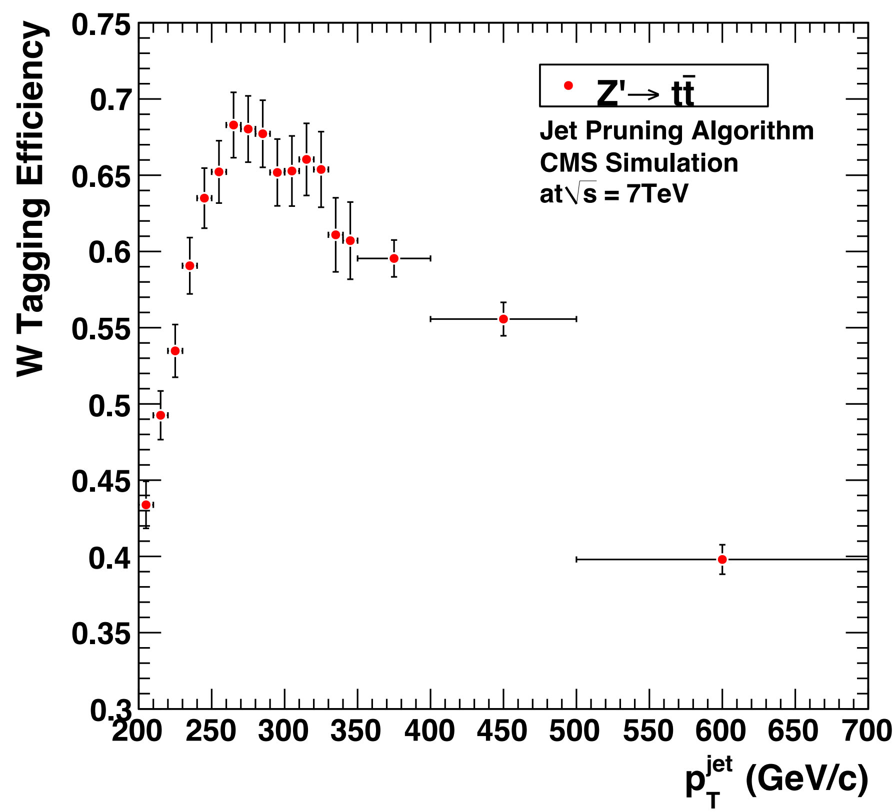
BACKUP MATERIAL

Substructure Algorithms

- Since the top decay products merge into a single jet, different reconstruction algorithms are needed
  - Top Tagging algorithm
  - W Tagging algorithm (pruning)
- Top tagging peak efficiency ~50%
- W tagging peak efficiency ~70%

PAS JME-10-013





Reducing QCD Contamination

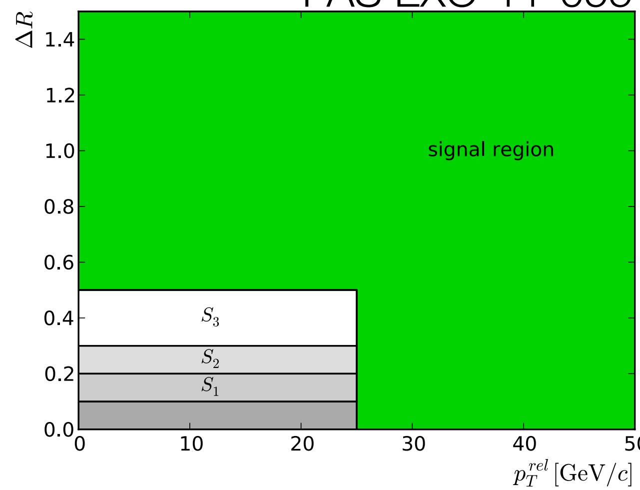
- The selected leptons do not have any isolation requirements applied
  - Additional cuts required
  - Lepton is required to pass '2D' cut
    - If within the jet cone, must have 25 GeV of momentum orthogonal to jet axis
  - Electron channel uses triangular cuts to reduce QCD
    - Uses angles between leading jet, electron and MET according to

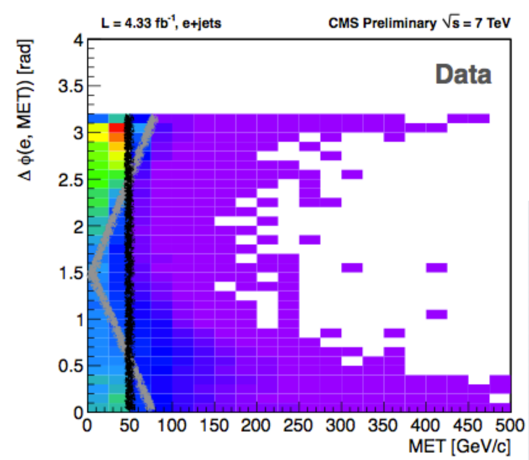
$$\Delta\phi(e, E_{\mathrm{T}}^{\mathrm{miss}}) < \frac{1.5}{75}E_{\mathrm{T}}^{\mathrm{miss}} + 1.5, \quad \Delta\phi(e, E_{\mathrm{T}}^{\mathrm{miss}}) > -\frac{1.5}{75}E_{\mathrm{T}}^{\mathrm{miss}} + 1.5,$$

$$\Delta\phi(j_1, E_{\mathrm{T}}^{\mathrm{miss}}) < \frac{1.5}{75}E_{\mathrm{T}}^{\mathrm{miss}} + 1.5, \quad \Delta\phi(j_1, E_{\mathrm{T}}^{\mathrm{miss}}) > -\frac{1.5}{75}E_{\mathrm{T}}^{\mathrm{miss}} + 1.5.$$

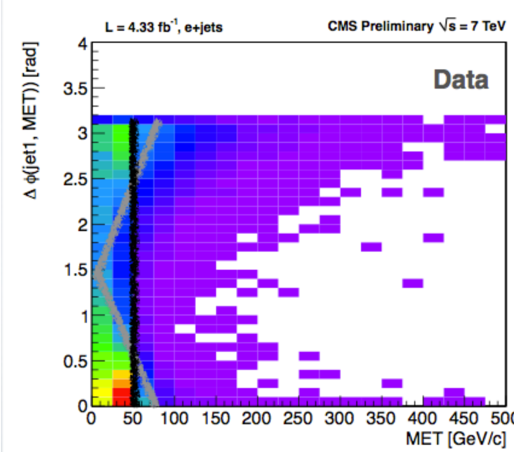
    - ~4% QCD contribution in final selection

PAS EXO-11-055





PAS EXO-11-092

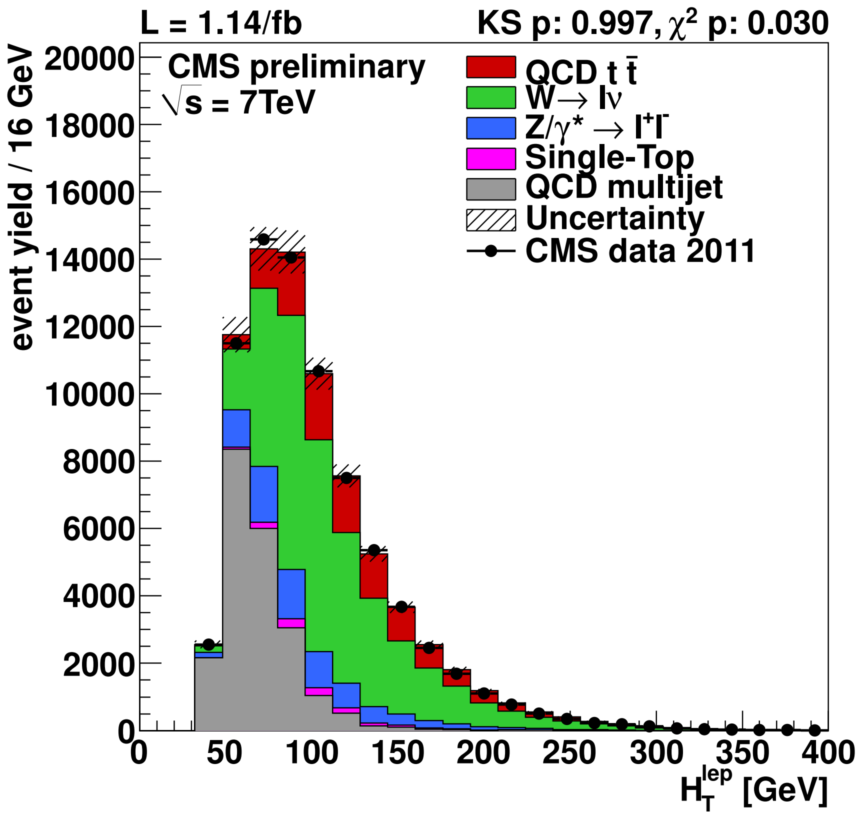


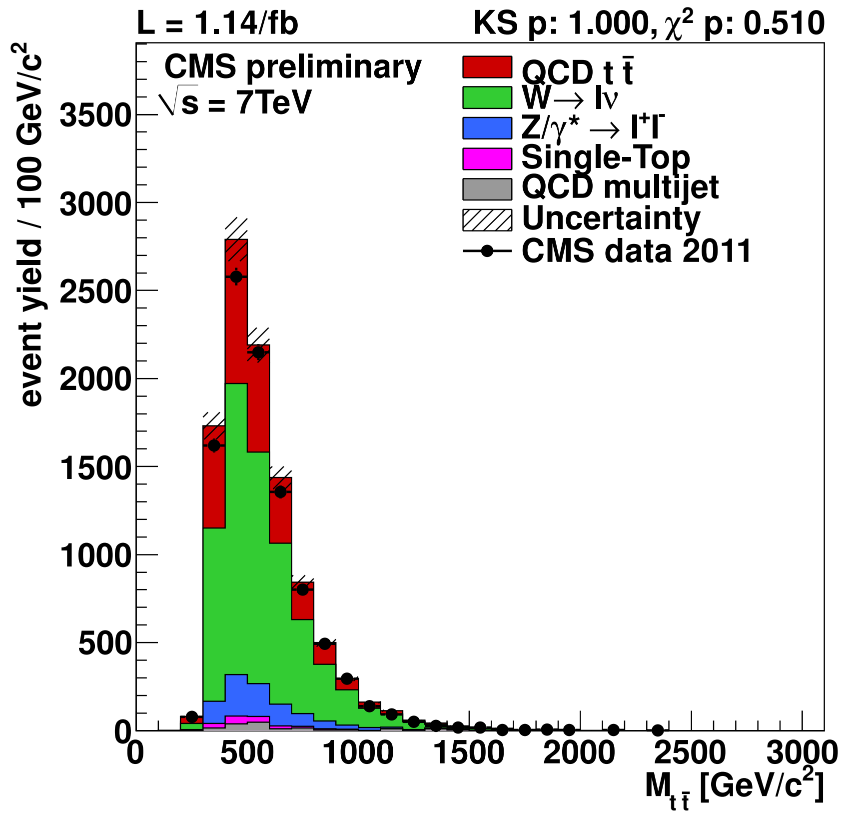
Muon + Jets Results

PAS EXO-11-055

- Final signal region defined by additional cut
  - H_T^{lep} = muon p_T + MET > 150 GeV
- Background model is validated in sideband region defined by inverting leading jet p_T cut

Process	N_{exp} Jet cut	N_{exp} Lepton cut	N_{exp} $H_{T,lep}$ cut	ϵ [%]
W+jets	2780.7	2157.4	1385.9 ± 93	3.9×10^{-3}
Z+jets	452.3	393.4	204.3 ± 13	5.9×10^{-3}
QCD $t\bar{t}$	1090.8	613.9	385.3 ± 60	0.21
Single-Top	73.7	40.0	26.0 ± 1.8	0.058
QCD multijet	—	—	57.2 ± 9	—
Total background	—	—	2058.7 ± 130	—
Data	—	—	1817	—
Z′ ($M = 1$ TeV/c^2)	125.7	58.7	46.5	4.1
Z′ ($M = 2$ TeV/c^2)	306.6	130.2	121.8	10.7
Z′ ($M = 3$ TeV/c^2)	294.6	119.7	113.6	9.9

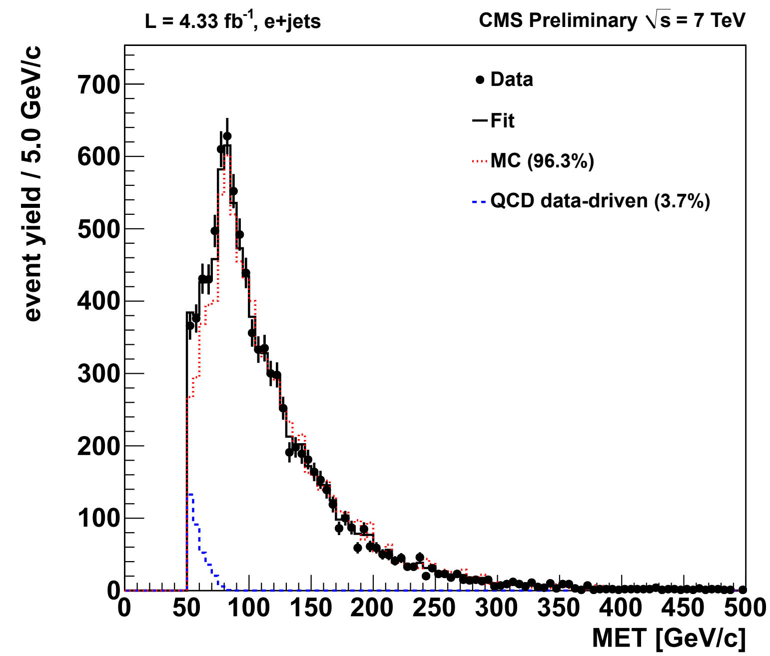




Electron + Jets Results

PAS EXO-11-092

- Fit performed to determine background composition using MET distribution
- Proceed to set limits using the m_{tt} spectrum after this fitting procedure



Sample	Event Yield
Z', $M = 1\,\mathrm{TeV}/c^2$	153 ± 2
Z', $M = 1.5\,\mathrm{TeV}/c^2$	244 ± 3
Z', $M = 2\,\mathrm{TeV}/c^2$	273 ± 3
Z', $M = 3\,\mathrm{TeV}/c^2$	229 ± 3
Single Top	285 ± 4
W+jets	5408 ± 105
Z+jets	152 ± 8
$t\bar{t}$	2707 ± 29
Total MC	8551 ± 109
QCD data-driven	375 ± 16
Total Background	8927 ± 111
Data 2011	9236