Jet substructure and top tagging at CMS

Nhan Tran
Fermi National Accelerator Laboratory

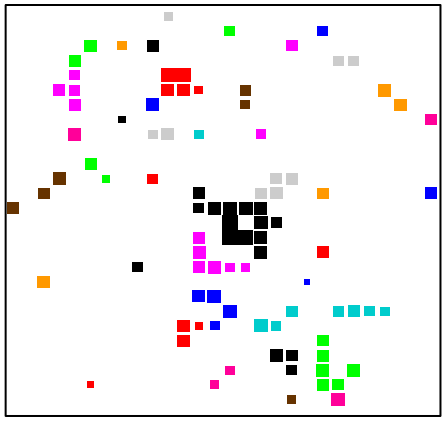
on behalf of the CMS collaboration

Chicago 2012 Workshop on LHC Physics
November 12, 2012

introduction

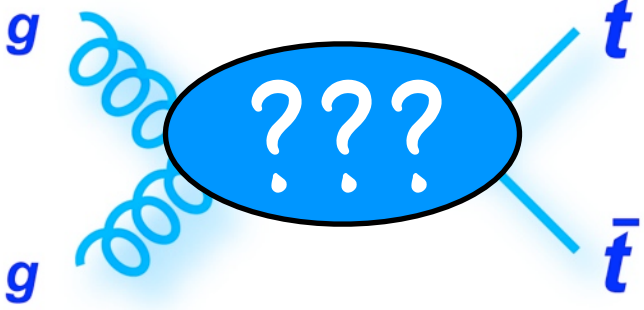
- Many recent phenomenological works highlight the potential of jet substructure for new physics searches
- Improved sensitivity in top analyses, boosted vector bosons and Higgs searches
- Requires an understanding of QCD radiation in jets

grooming

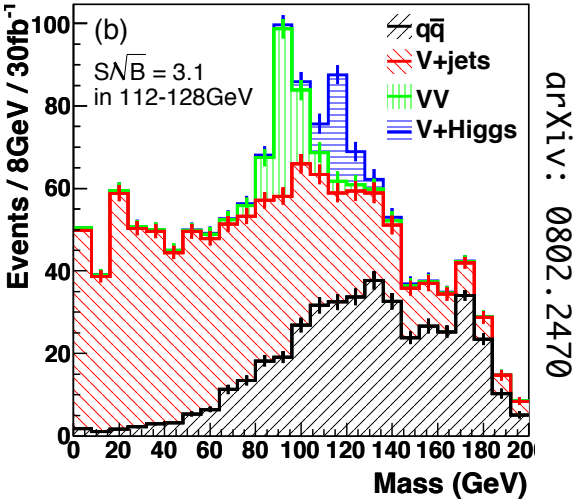


arxiv:0912.1342

boosted tops



higgs



arXiv: 0802.2470

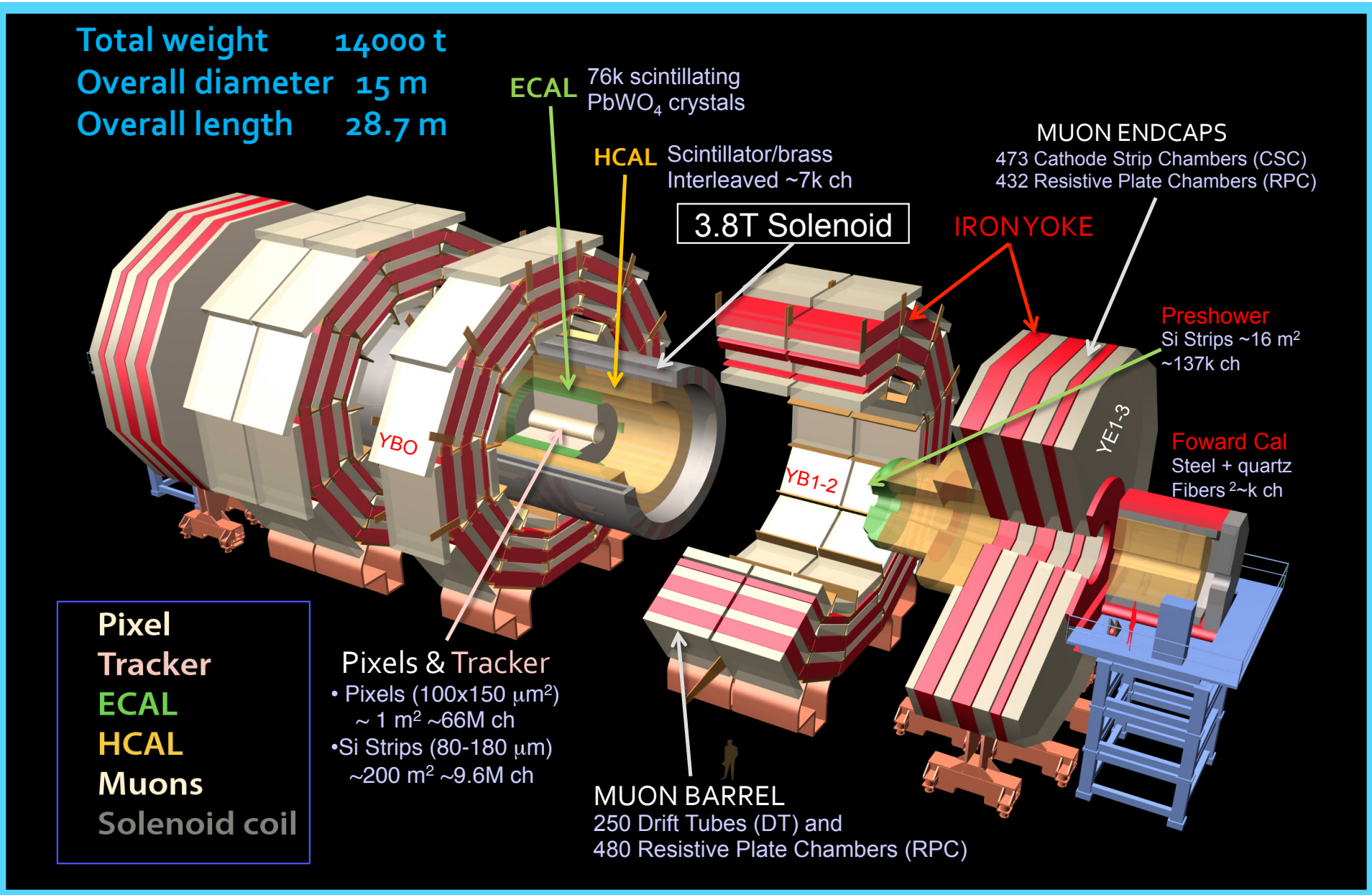
table of contents

- Jet mass in di-jet and V+jet events [1]
- Algorithms in new physics searches
- Searches with boosted vector bosons
  - $X \rightarrow ZV \rightarrow 2l2q$ [2], $X \rightarrow ZZ \rightarrow 2\nu2q$ [3]
  - Dijets: $X \rightarrow qV/VV$ [4]
- Searches with boosted tops
  - $X \rightarrow tt$ in all-hadronic final state [5-6]

[1] https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSMP12019
[2] CMS PAS EXO-11-061
[3] CMS PAS EXO-11-081
[4] CMS PAS EXO-11-095
[5] CMS PAS JME-10-013
[6] CMS PAS EXO-11-006, arXiv:1204.2488

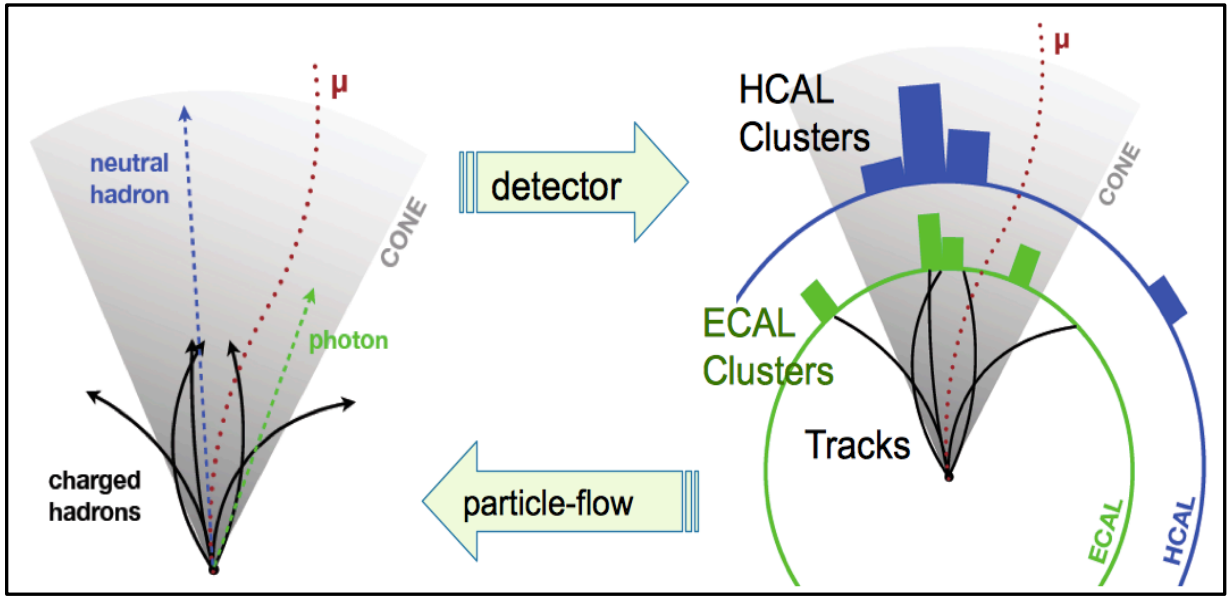
N.B. All searches on CMS 2011 data

the CMS experiment



global event description

CMS employs a particle flow algorithm translating detector measurements into physics objects



- Algorithm uses information from all subsystems, returns a list of muons, electrons, photons, and neutral and charged hadrons
- List is used as the building blocks for final physics objects in analyses:
  - Jets and b-jets, taus, and missing transverse energy (MET)

foundations: jet mass measurement

Inclusive analysis of 2011 CMS data for di-jet and V+jet events

- Complementary probe of gluon-jet and quark-jet enriched environments
- Systematically examine grooming algorithms using algorithm parameters from original references
  - Trimming: http://arxiv.org/abs/0912.1342
  - Filtering: http://arxiv.org/abs/0802.2470
  - Pruning: http://arxiv.org/abs/0903.5081
- Probe several different jet radii and finding algorithms:
  - Anti-KT [AK] (R = 0.5,0.7, 0.8), Cambridge-Aachen [CA] (R = 0.8, 1.2)
- Jet details, more details about jet corrections in backup
  - Particle flow jets with charged hadrons subtracted, neutrals corrected via ρA Fastjet correction
  - AK5 and AK7 jets have dedicated jet energy corrections

jet mass measurements

- Comparison of data to simulation as a validation of parton showering models
- Contrast herwig++ and Pythia 6 samples, di-jet samples are also compared against Pythia 8
- Provide detector-unfolded jet mass distributions as inputs to the theoretical community

observable: Jet mass distributions in bins of pT

$$PDF(m_J) = \frac{1}{\frac{d\sigma}{dp_T}} \times \frac{d^2\sigma}{dp_T dm_J}.$$

Di-jet presents the average jet pT and jet mass for the two leading jets in the event

V+jet presents the leading jet pT and jet mass in the event

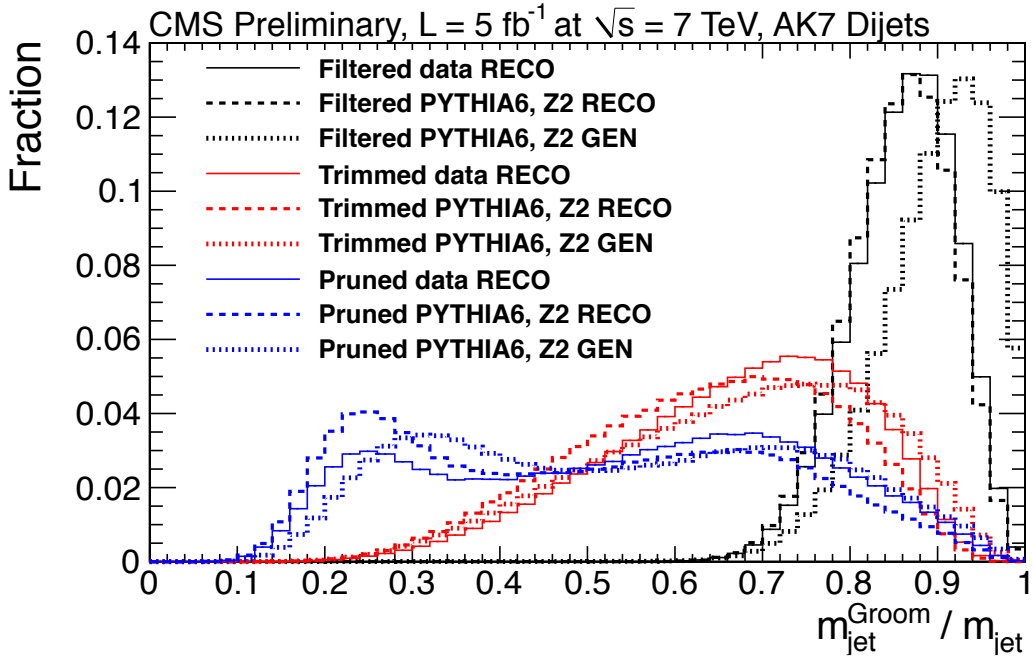
pT binning

bin	‹pT› for di-jet pT for V+jet
1	125-150
2	150-220
3	220-300
4	300-450
5	450-500
6	500-600
7	600-800
8	800-1000
9	1000-1500

V+jets only
Di-jet only
Di-jet and V+jet

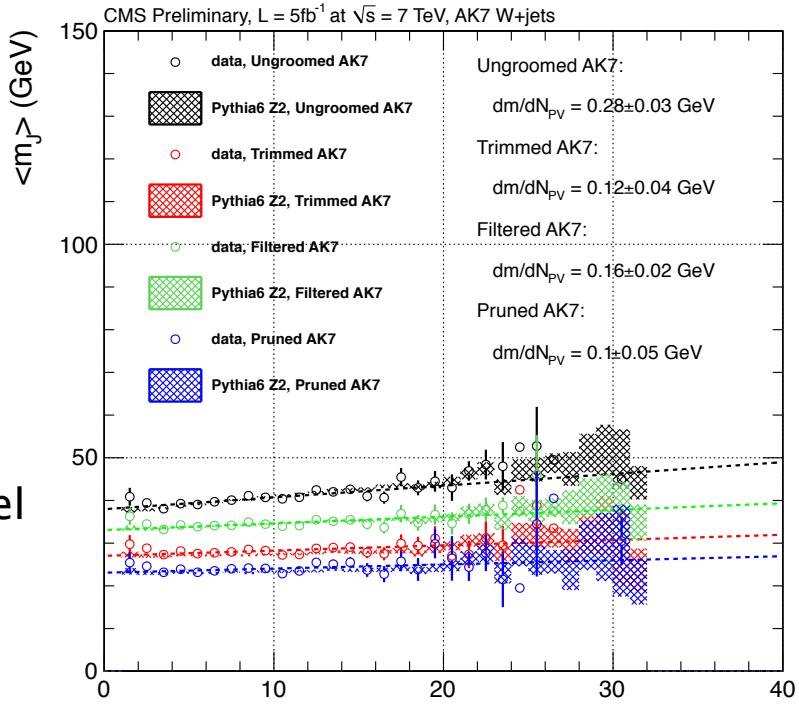
grooming algorithm performance

Ratio of the groomed to ungroomed jet mass, encapsulates behavior of each grooming algorithm



Comparison of grooming algorithms at particle level (GEN), reconstructed simulation (RECO) and data

Pruning is the most aggressive, filtering is the least aggressive

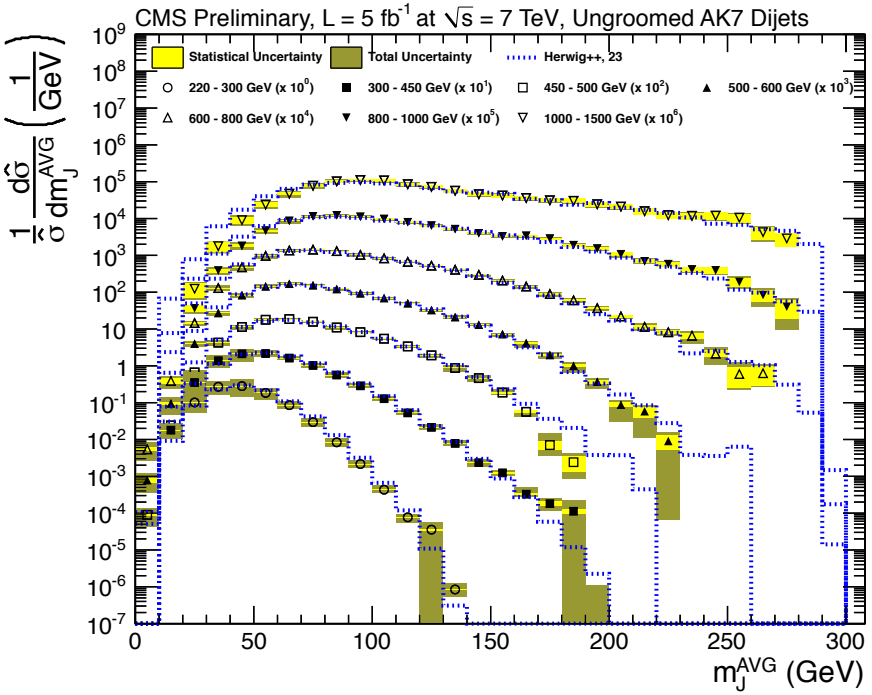


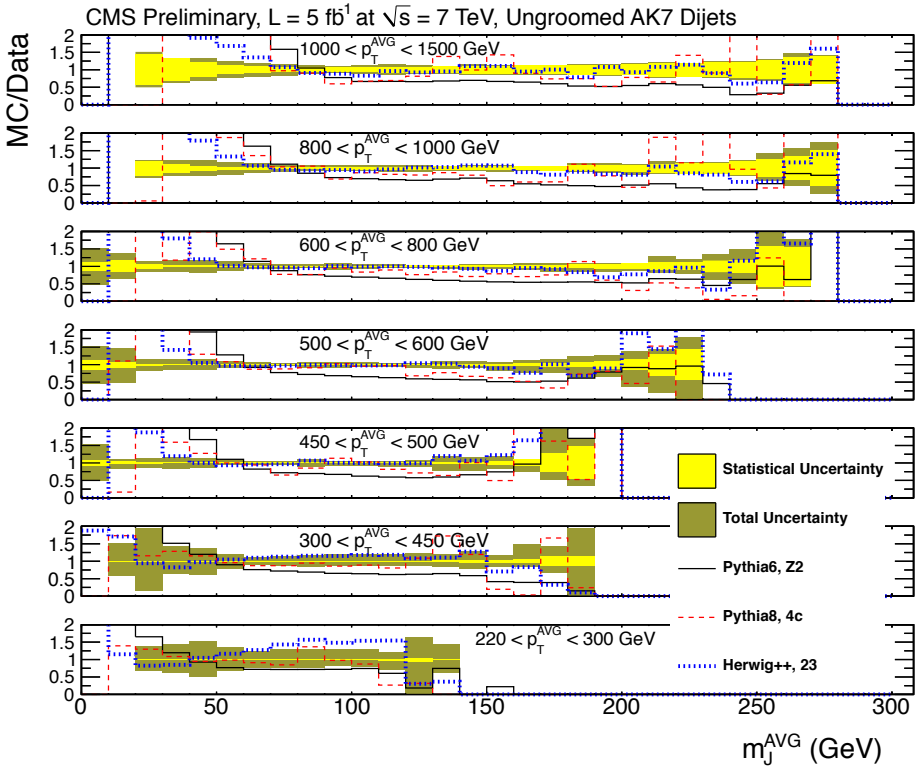
Average jet mass versus number of primary vertices

Grooming techniques show less sensitivity to pileup

unfolded jet mass distributions

AK7

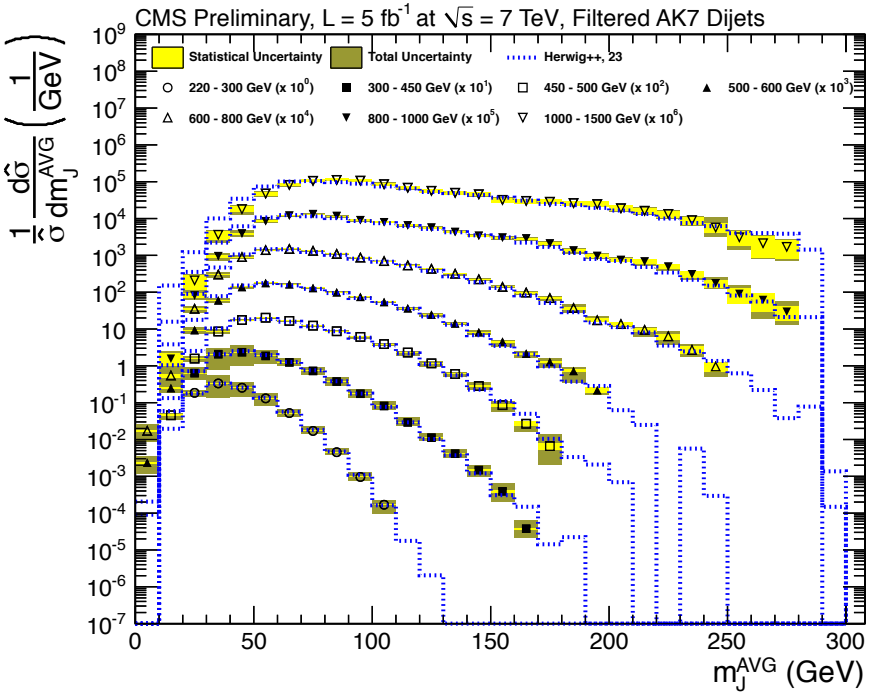


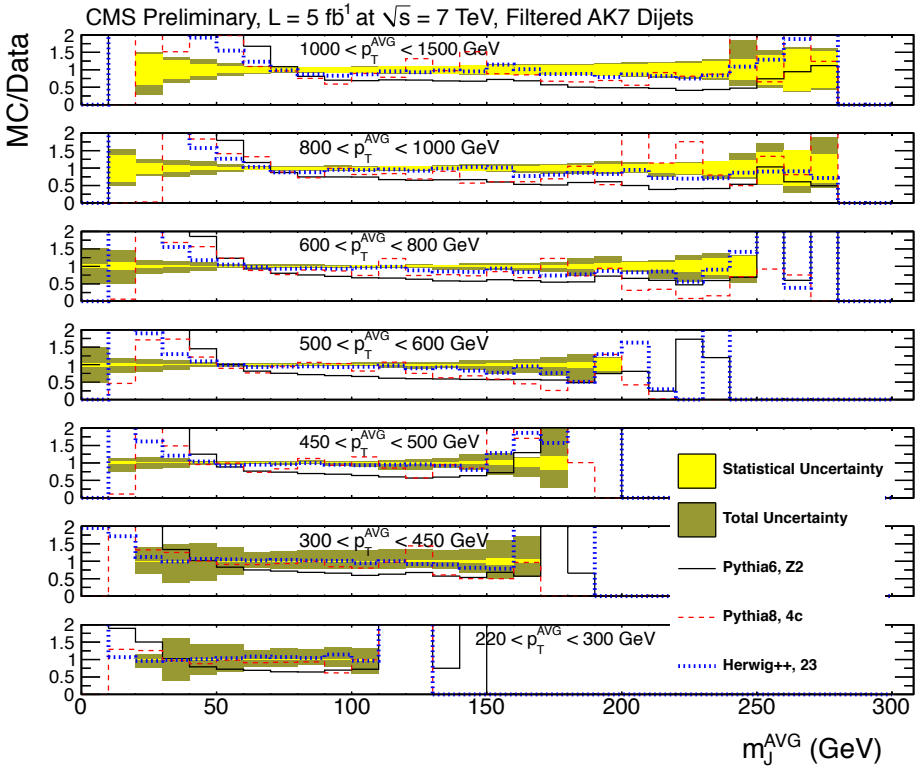


For unfolding, use iterative Bayesian method to unfold to the true distributions.
Combined electrons and muons in the case of V+jets using inclusive acceptance

unfolded jet mass distributions

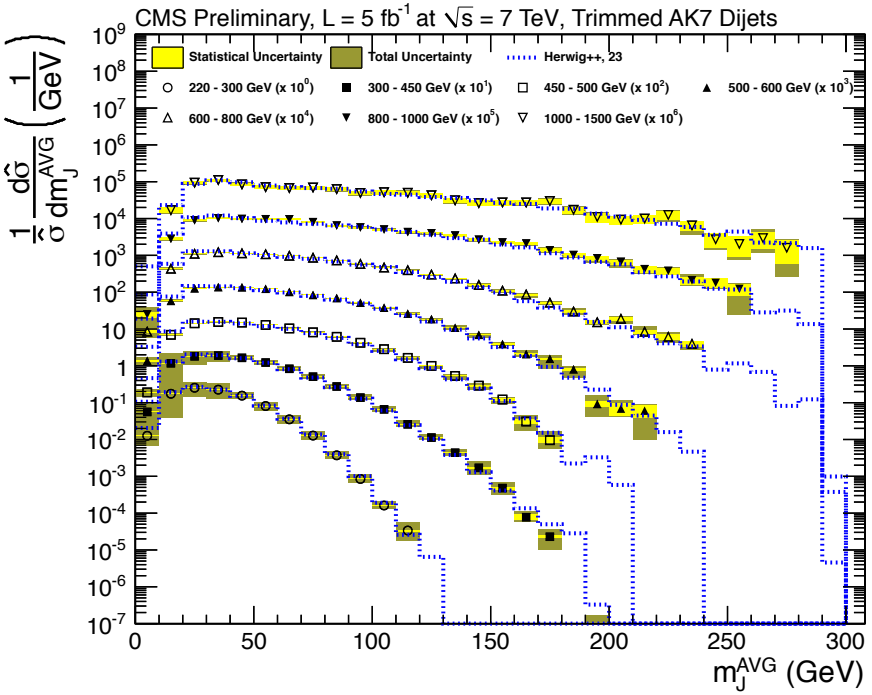
AK7, Filtered

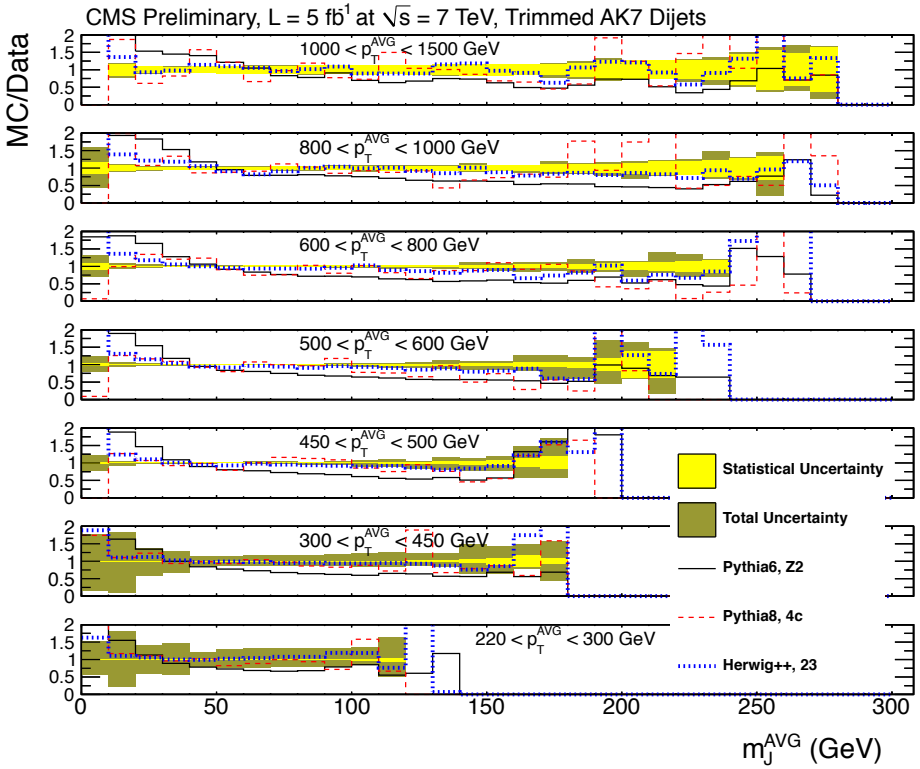




unfolded jet mass distributions

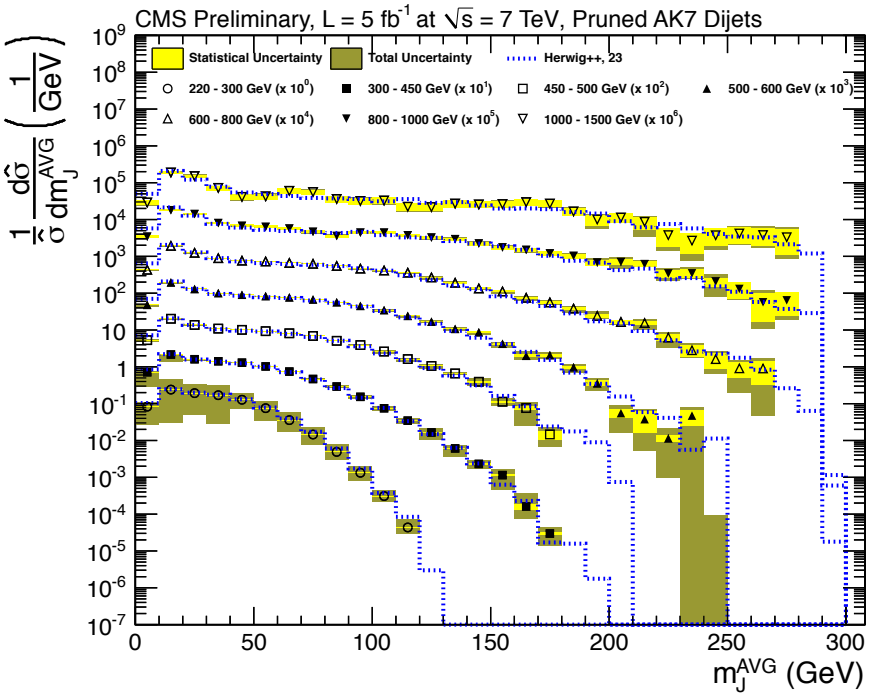
AK7, Trimmed

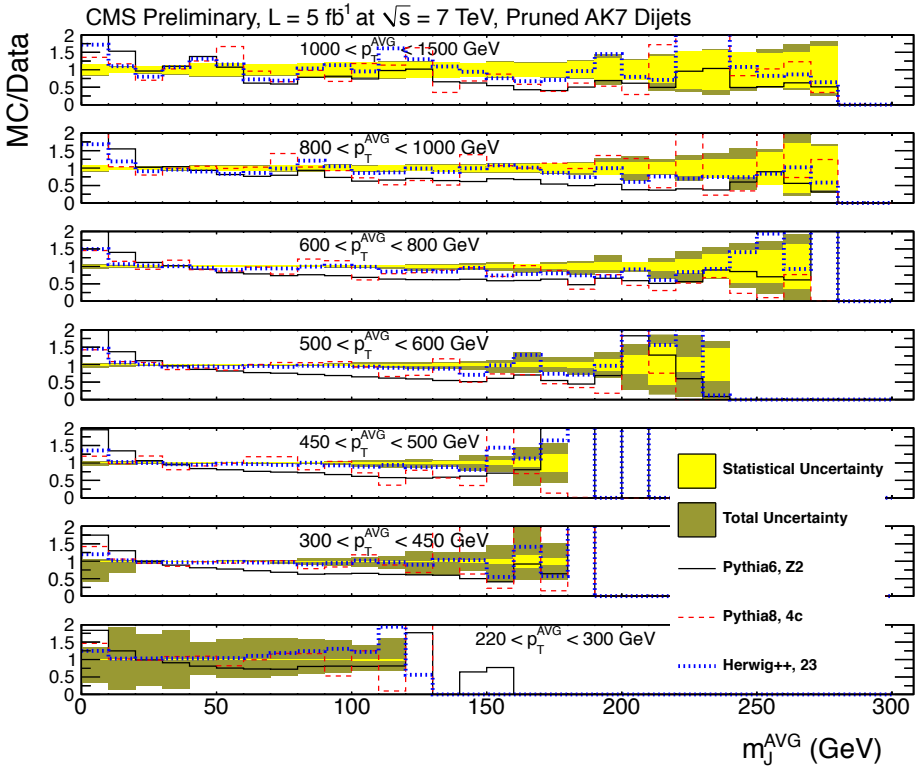




unfolded jet mass distributions

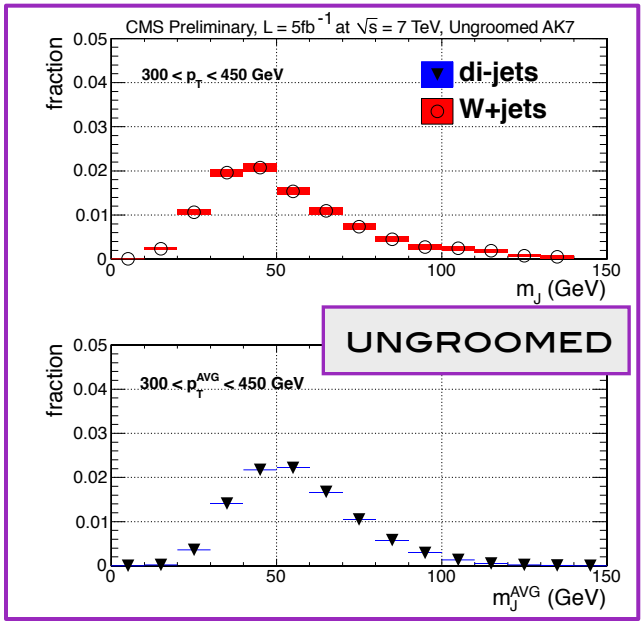
AK7, Pruned

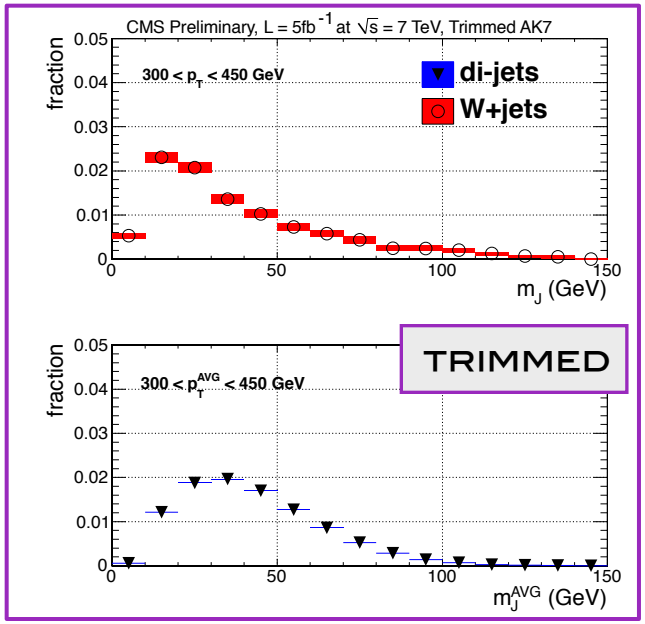




For unfolding, use iterative Bayesian method to unfold to the true distributions.
Combined electrons and muons in the case of V+jets using inclusive acceptance

jet mass measurement summary





- Analysis is an important benchmark in understanding various algorithms for searches for new physics and providing input for theoretical calculations
- Leading systematics coming from parton shower modeling, jet energy resolution and pileup
- Comparison with Pythia Z2 and Herwig++ simulation gives reasonable agreement where Herwig++ appears to agree better
- Detector-unfolded distributions are provided

For visualization:
N.B. stat. err. only

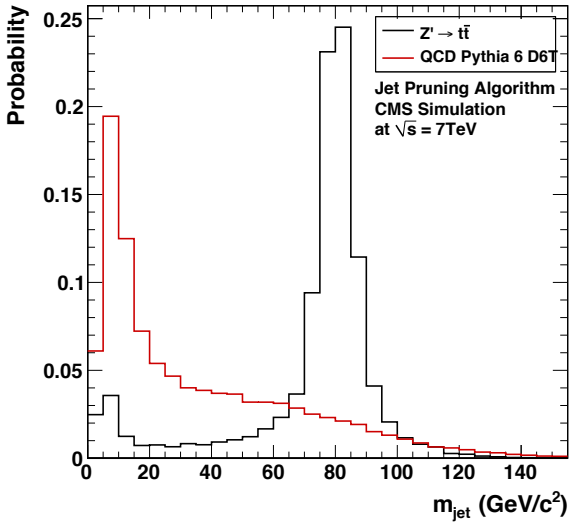
search algorithms

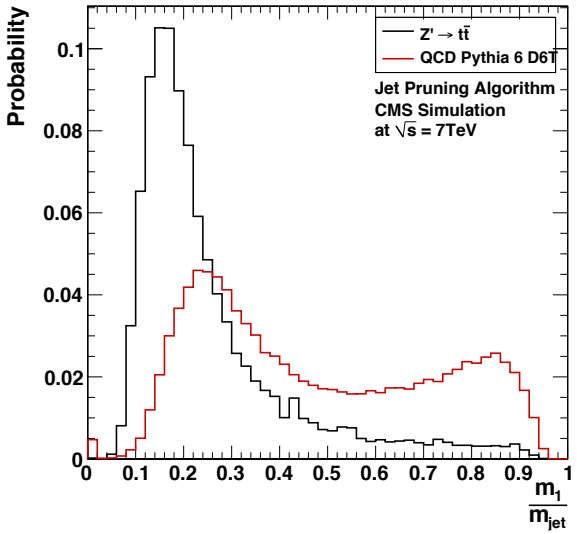
search channel algorithms, boosted W/Z

Identifying boosted W/Z bosons

not uniform across existing analyses but discuss most aggressive treatment, more details given per analysis

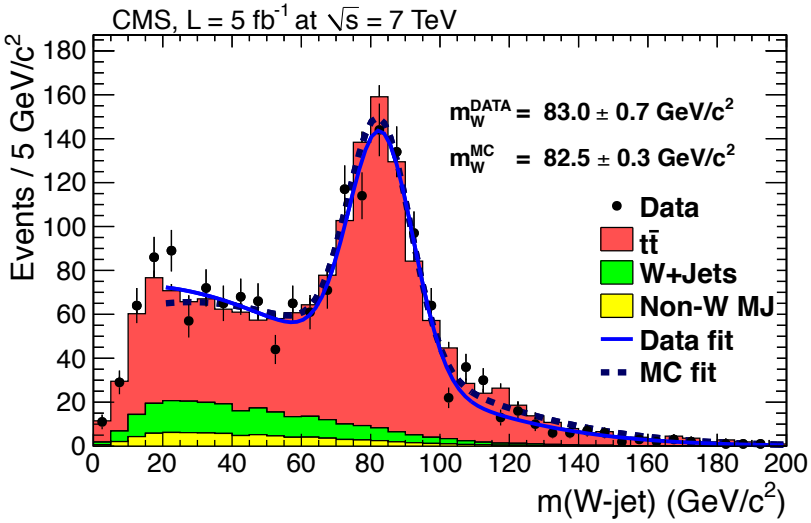
- Jet-finding with a large radius algorithm, typically CA8, AK7
- Prune jets with reference default parameters, removes soft and large angle constituents
- Cut on the mass drop, $\mu = m_1/m$
  - m_1 is mass of highest pT subjet
  - subjets defined by un-clustering last step

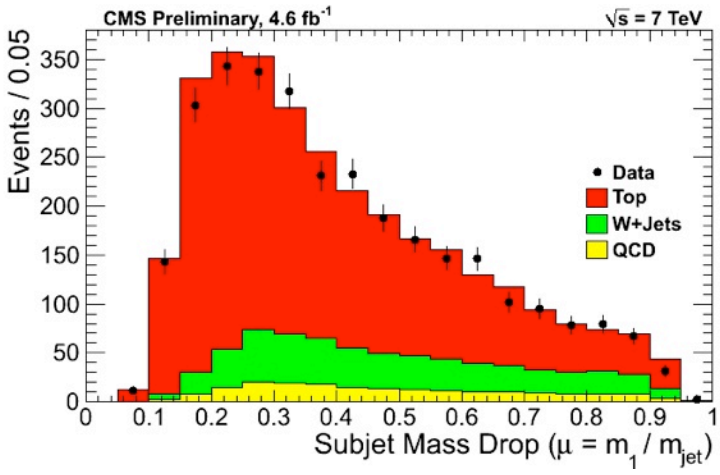


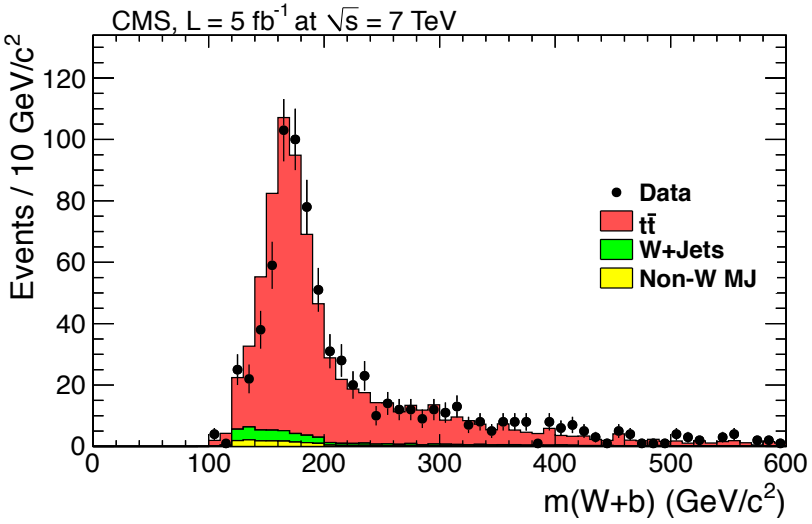


boosted W validation

- 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

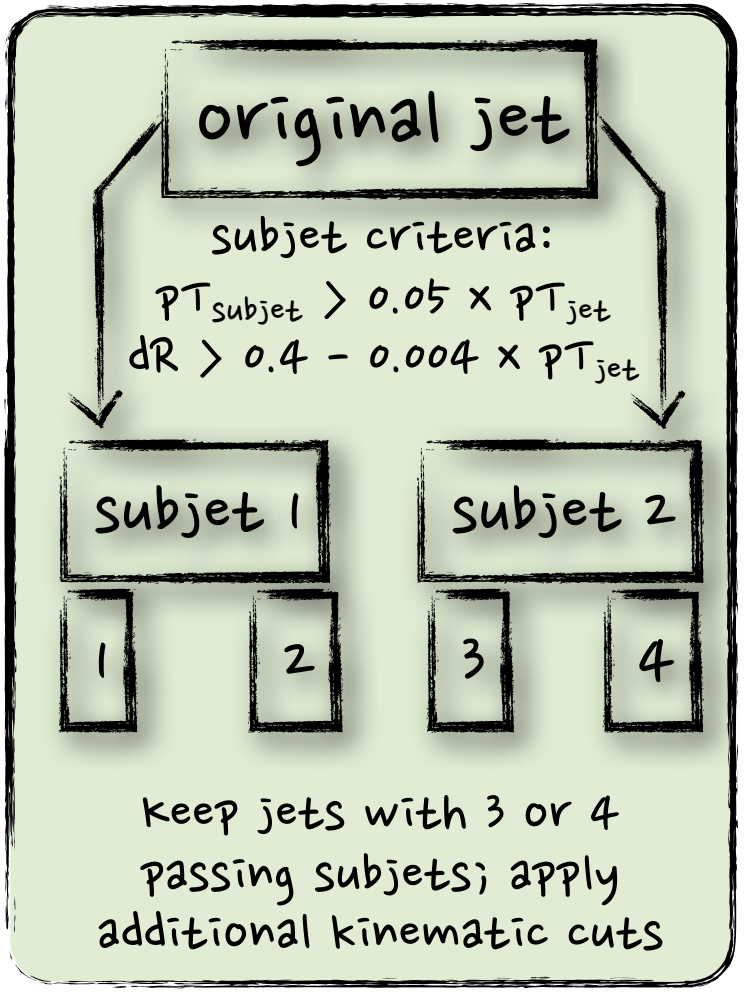






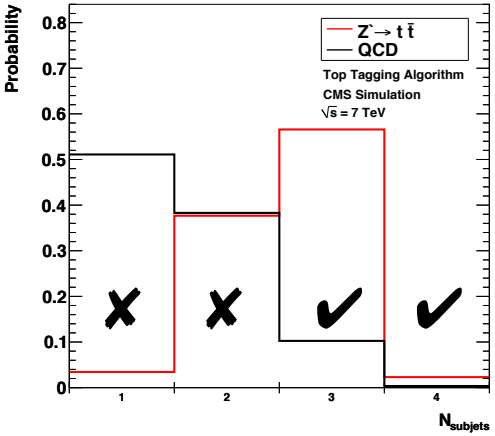
identifying boosted tops, CMS top tagger

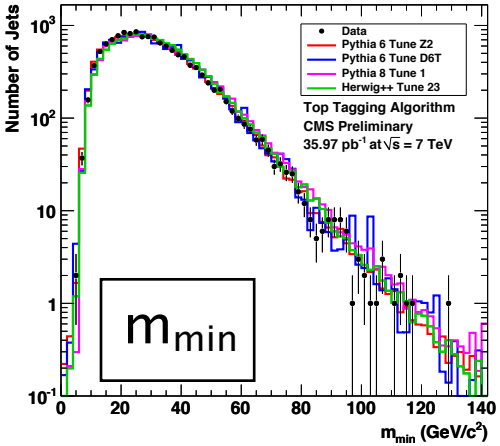
- Based on the JHU top tagger: PRL 101/142001 (2008)
- Cluster jets with CA8 algorithm
- Reverse clustering algorithm to find subjets, keep subjets passing following criteria
  - $pT_{subjet} > 0.05 \times pT_{jet}$
  - $dR > 0.4 - 0.004 \times pT_{jet}$
- Keep original jets with 3 or 4 passing subjets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

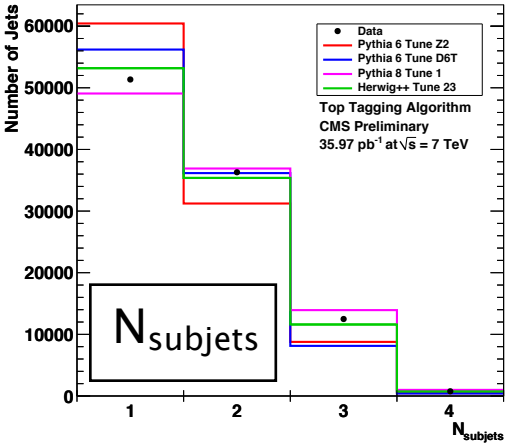


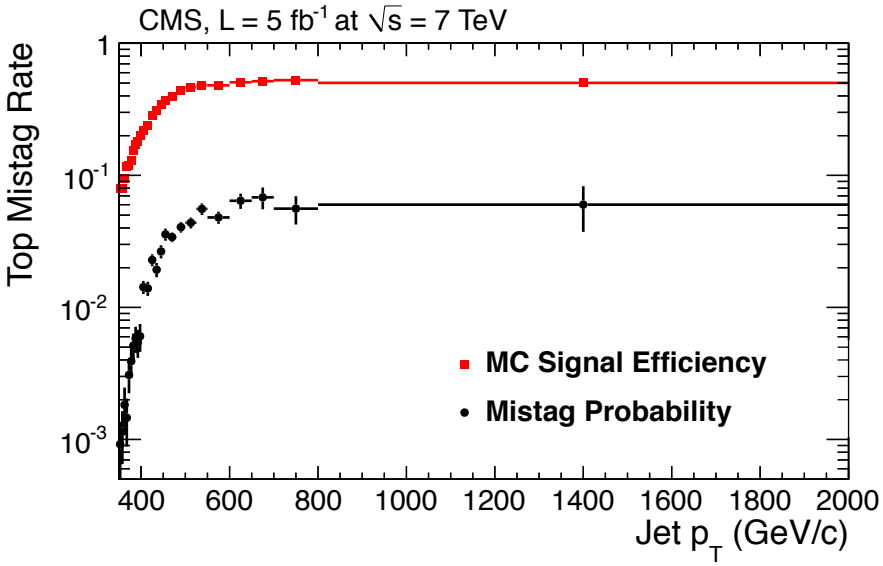
top tagging validation

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

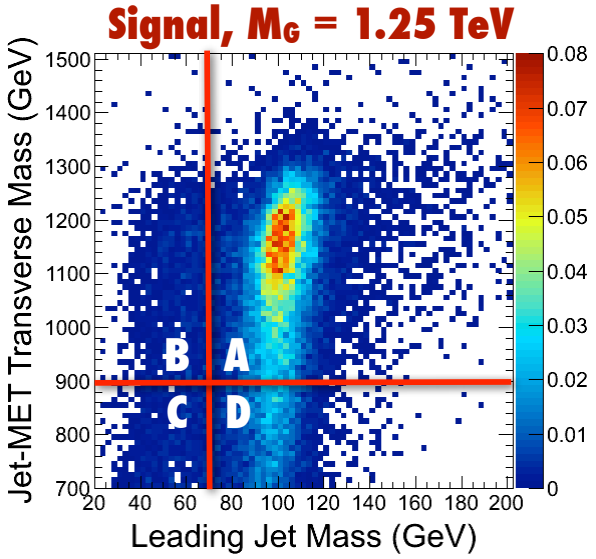
searches

$X \to ZV \to 2l2q$, $X \to ZZ \to 2\nu2q$

- $X \to ZV \to 2l2q$
  - AK7 Jet (merged) and two isolated leptons final state
  - $m_Z = [70,100]$ GeV; $pT_Z > 150$ GeV; $pT_{jet} > 250$ GeV, $m_{jet} = [65,120]$ GeV
  - Background shape constrained from sidebands
- $X \to ZZ \to 2\nu2q$
  - AK7 Jet (merged) and MET final state
  - Jet pT > 300 GeV, MET > 300 GeV
  - Discriminating observables:
    transverse mass M_T(Jet–MET) > 900 GeV
    and AK7 jet mass > 70 GeV
  - Background estimation technique:

$$B_{est} = N_D \cdot \frac{N_B}{N_C} \cdot \frac{1}{\rho}$$

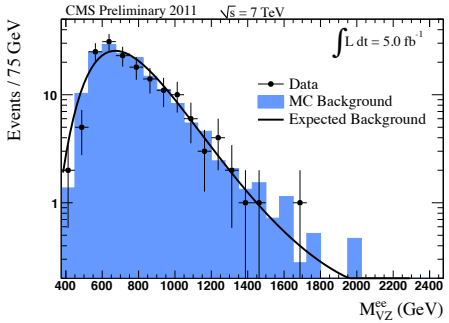
← 2D correlation factor

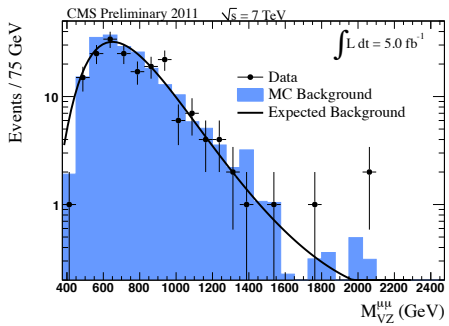


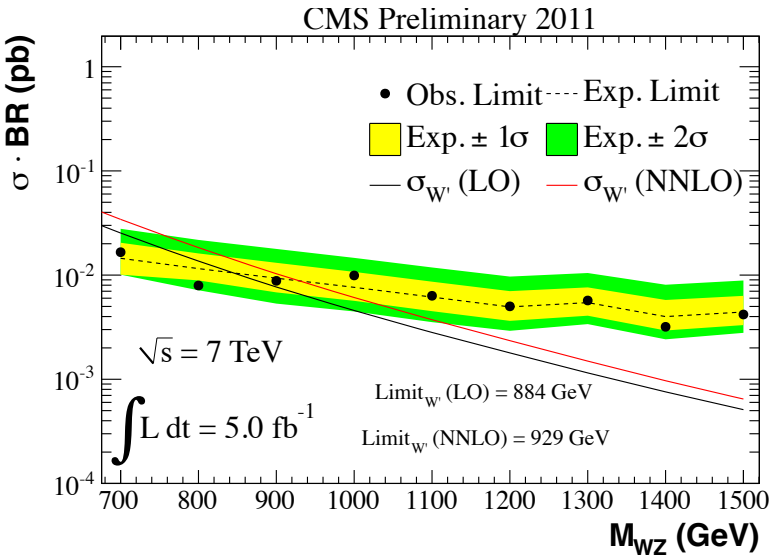
X → ZV → 2l2q, X → ZZ → 2ν2q

No significant excesses observed

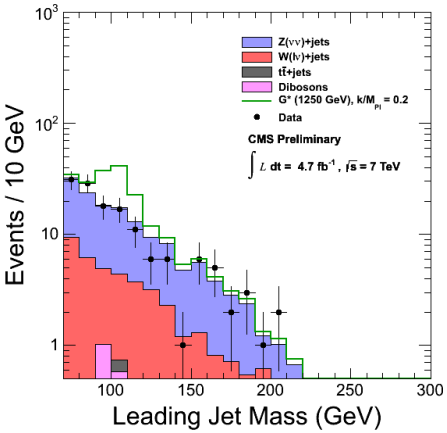
X → ZV → 2l2q

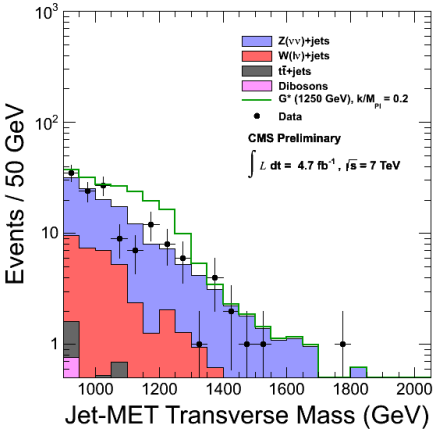


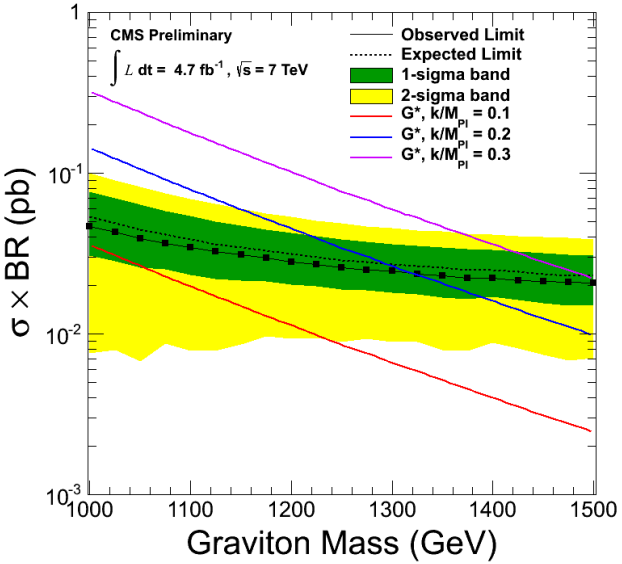




X → ZZ → 2ν2q

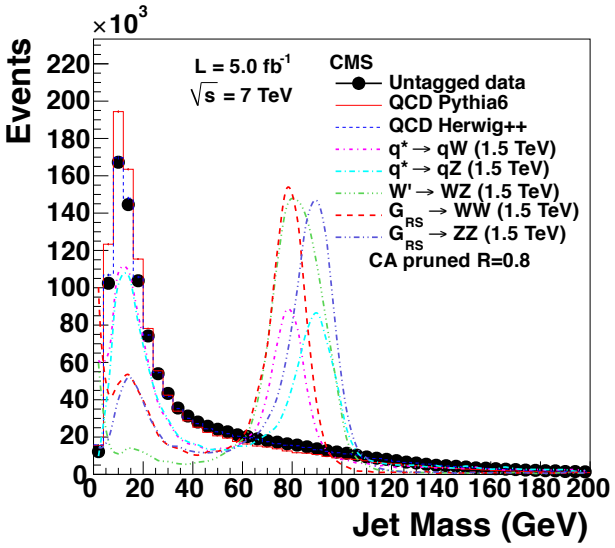


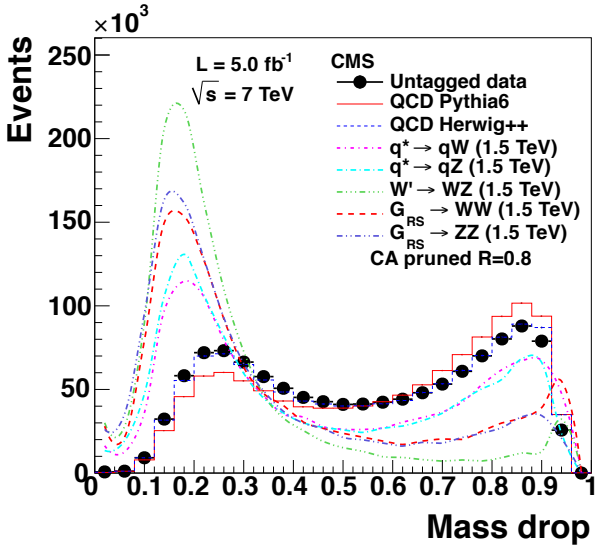


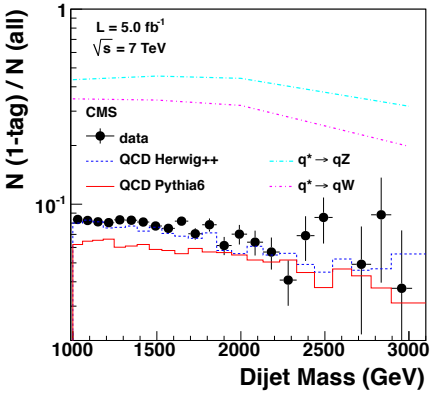


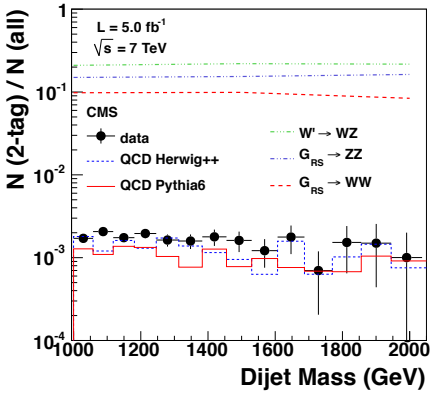
X → qV/VV

- 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 $\mu < 0.25$
  - Identify 1-tag and 2-tag topologies depending on final state
    - Tag rate from semileptonic tt



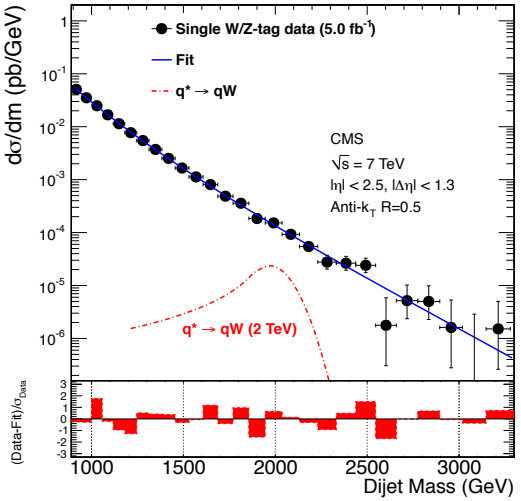


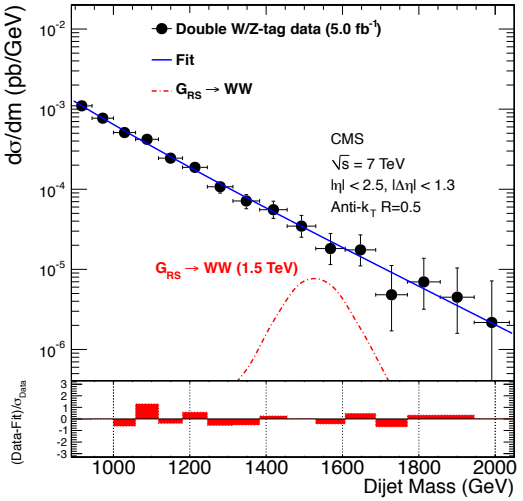




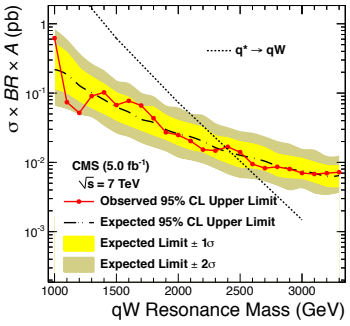
X → qV/VV

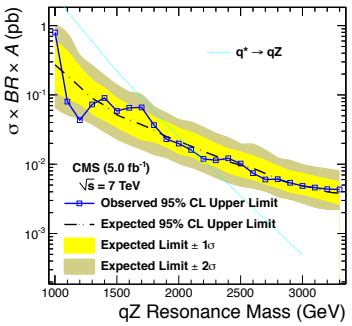
No significant excesses in 1-tag or 2-tag final states

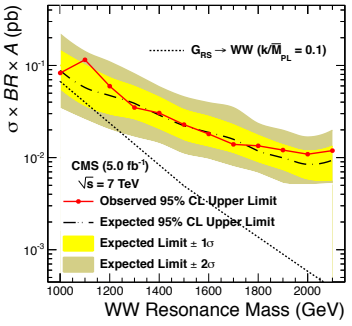


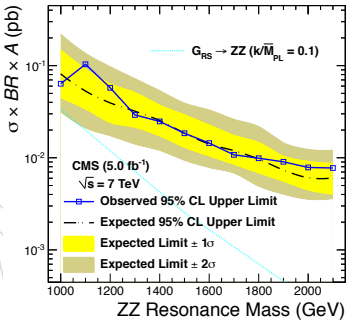


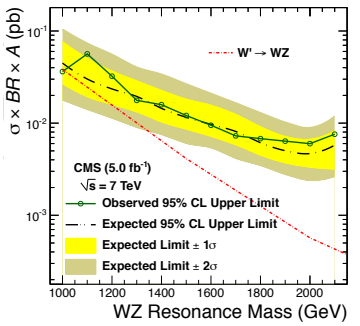
Results interpreted in many final states
qW, qZ, WW, ZZ, WZ





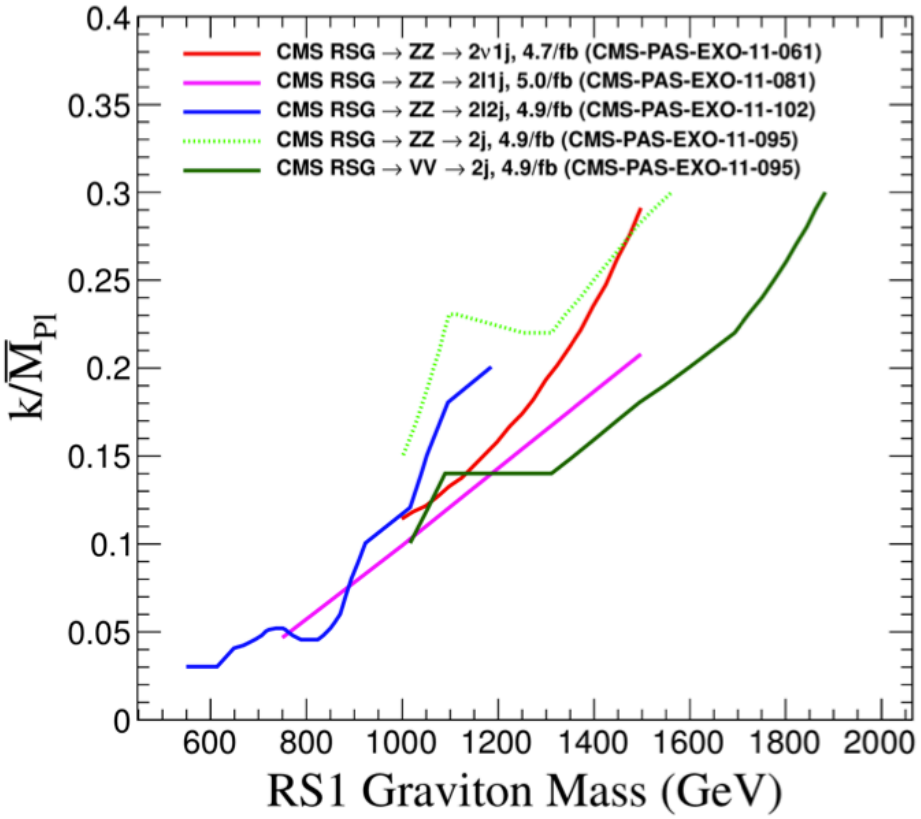






overview of exotic graviton searches

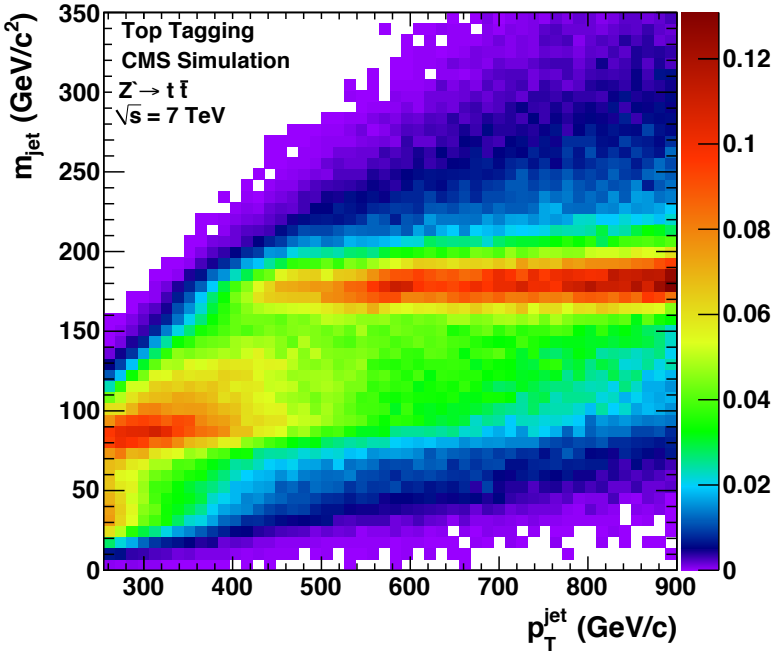
Combination of CMS searches
for RS1 Graviton in di-boson final states

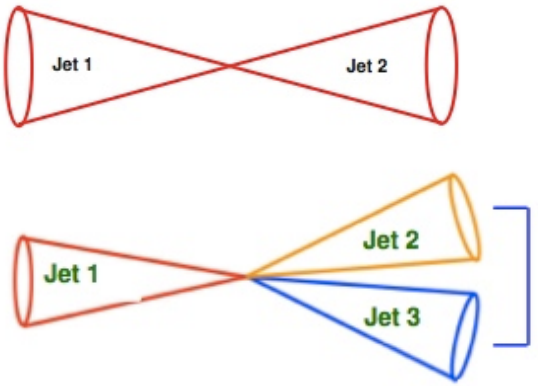


4 of 5 analyses using jet substructure observables

Z' → tt, all-hadronic channel

- 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

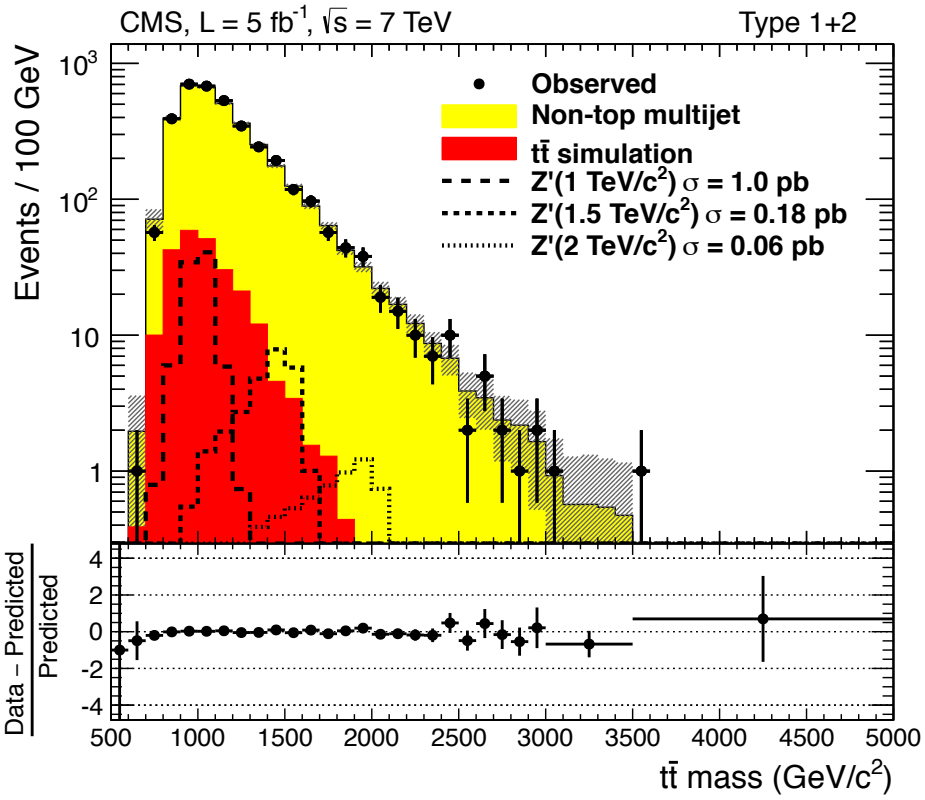


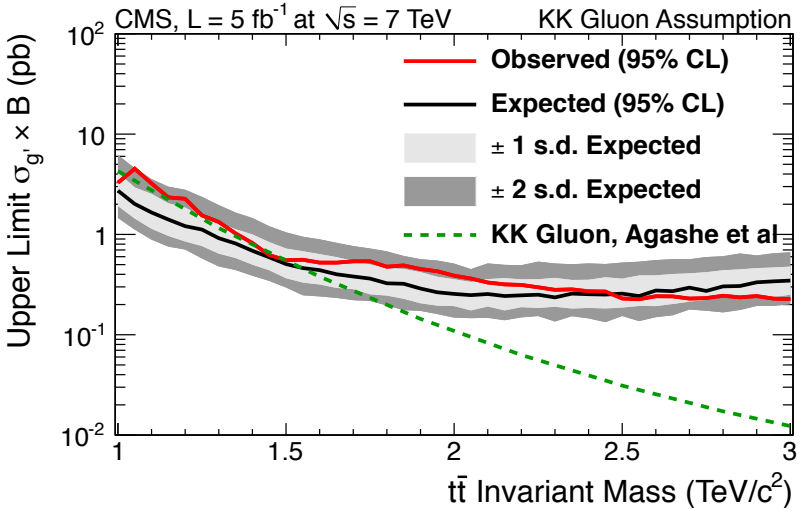


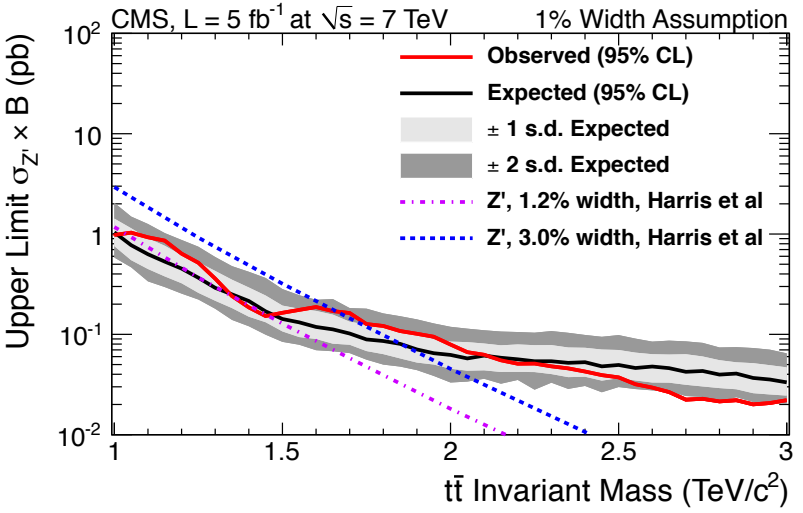
Z' → tt, all-hadronic channel

No significant excess in tt mass

Interpret searches in terms of KK gluons and Z' of various widths





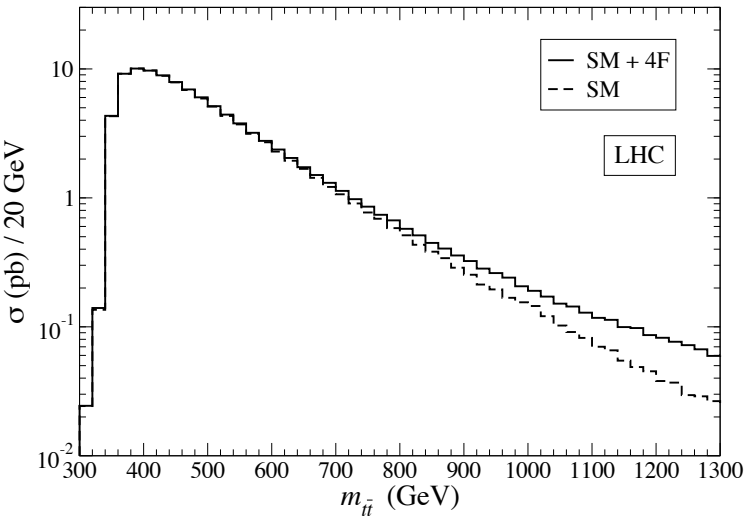


m_{tt} enhancement

- New physics models explaining $t\bar{t}$ asymmetry predict enhancement in the $t\bar{t}$ production rate at high $\sqrt{s}$
- Search for excess of $t\bar{t}$ events above 1 TeV
- Assume same efficiency of SM $t\bar{t}$
- Set limit on enhancement factor, S, w.r.t SM production expectation
- Excludes $S > 2.6 \cdot SM$ at 95% CL

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)



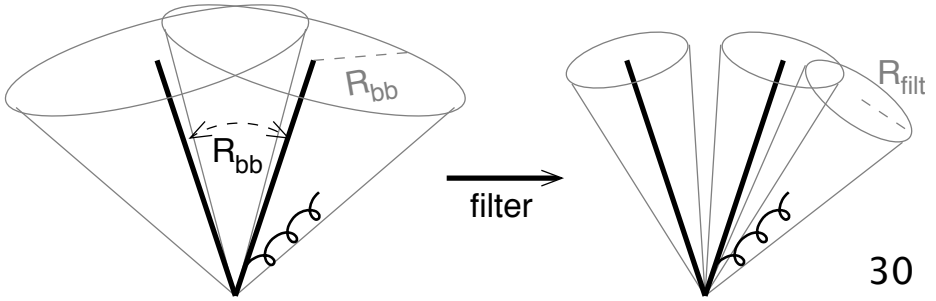
summary and prospects

- Substructure methods on display at CMS
- Inclusive measurements of jet mass in di-jet and V+jet events with various grooming algorithms
- Several searches are employing substructure observables
  - Merged vector bosons in many final states
  - Boosted tops with CMS top tagger
- Future prospects:
  - Plan for several 2012 analyses employing substructure techniques
  - More mature methods, more powerful phenomenological tools exist

backup

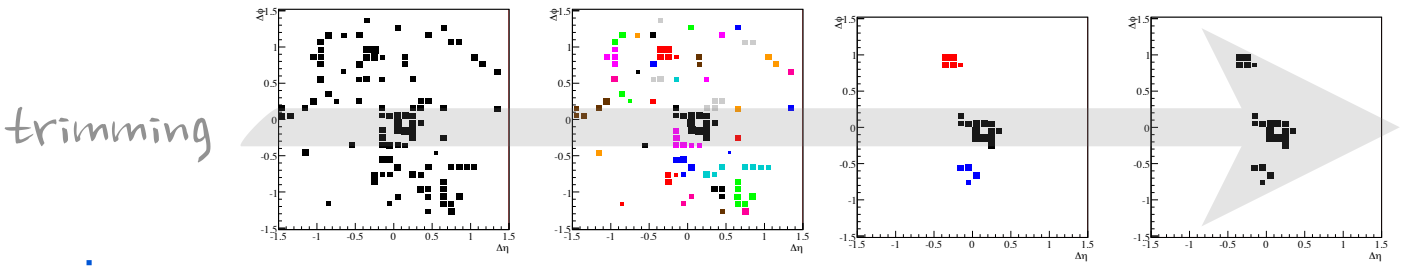
jet substructure algorithms

- 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



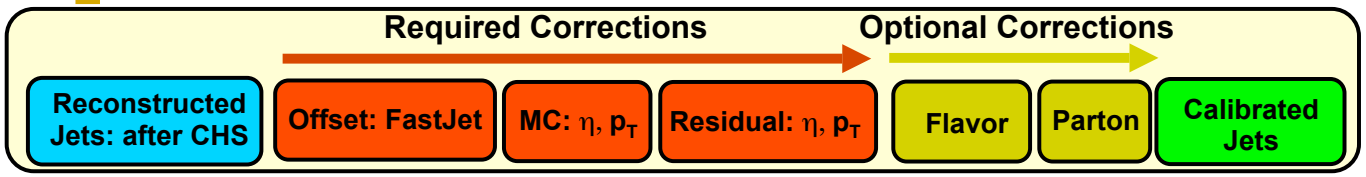
jet substructure algorithms

- 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

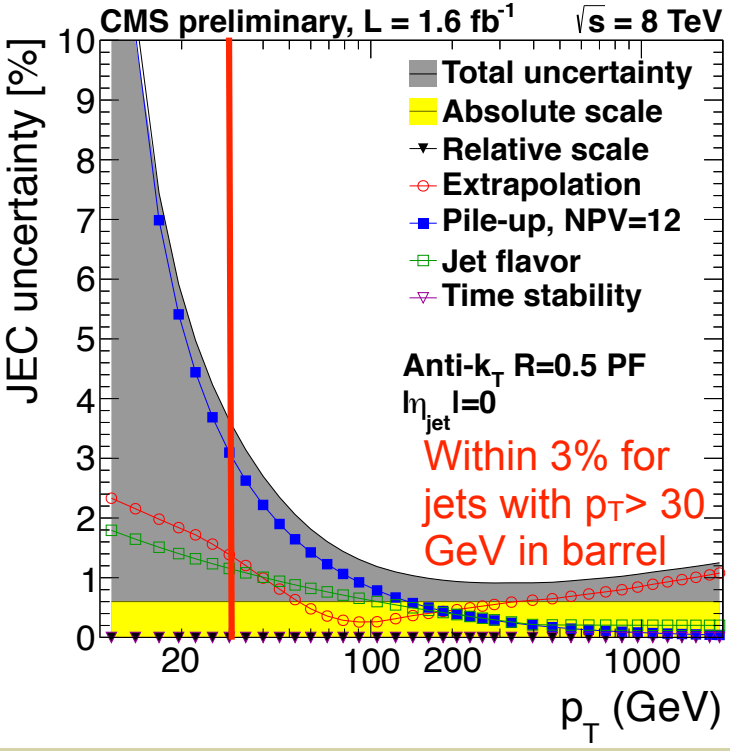
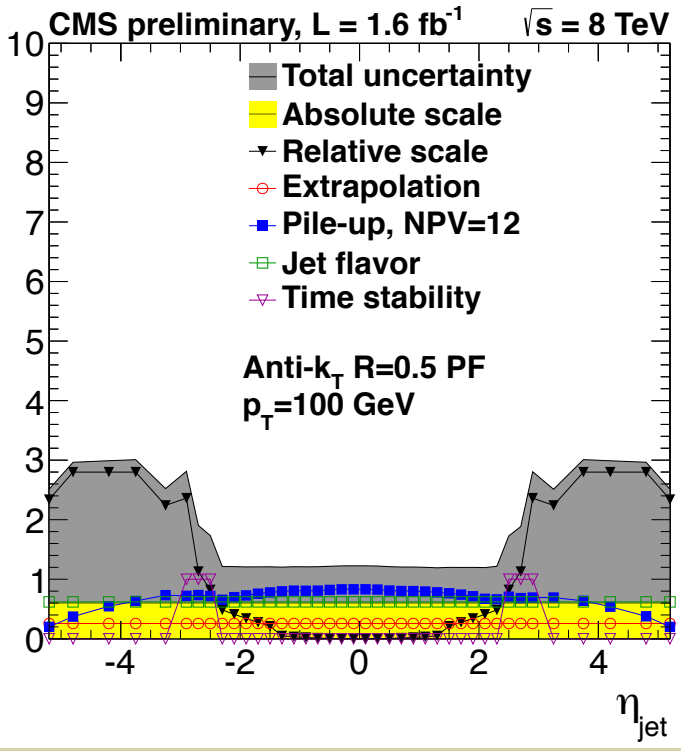
Jet energy calibration: overview



- 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.