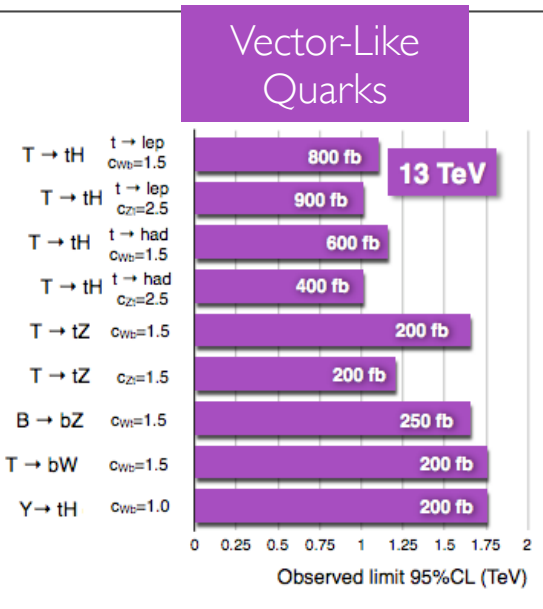
Jet Substructure at CMS: Algorithms, Searches, and Outlook

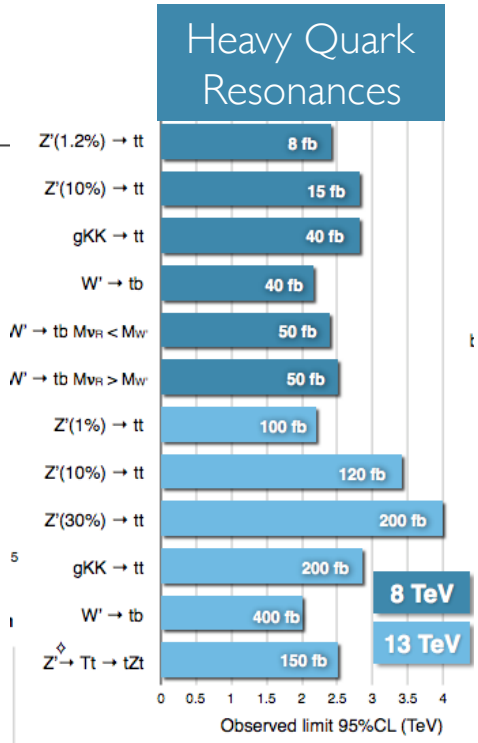
Justin Pilot, *UC Davis*
CMS Collaboration

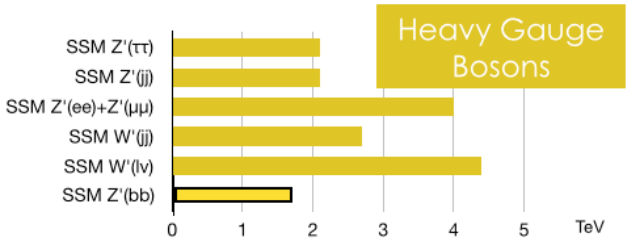
High Energy Physics Seminar
University of Virginia
Charlottesville, VA
16 November 2016

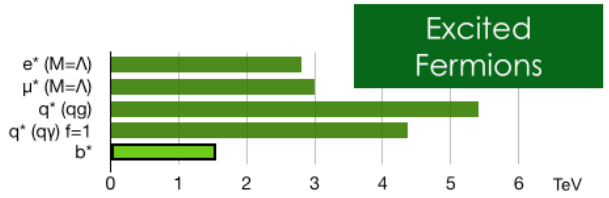
Introduction

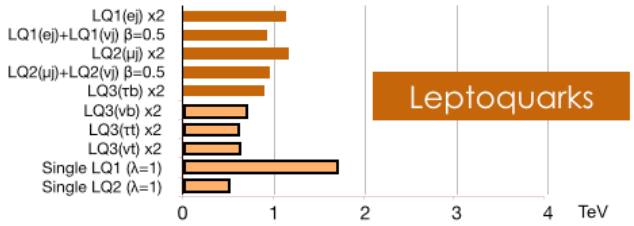
- Many searches for new physics underway at LHC experiments
  - No discoveries yet!
  - Mass exclusion limits for new particles reaching the ~several TeV level
- Probing higher mass scales means decay products have higher p_T
  - Hard to resolve decay products
- Running conditions at the LHC continue to become more and more challenging
  - High pileup
- Is there any hope??
  - YES!!

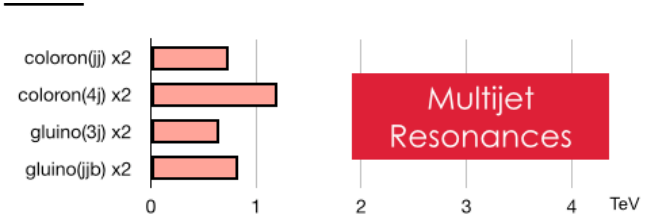








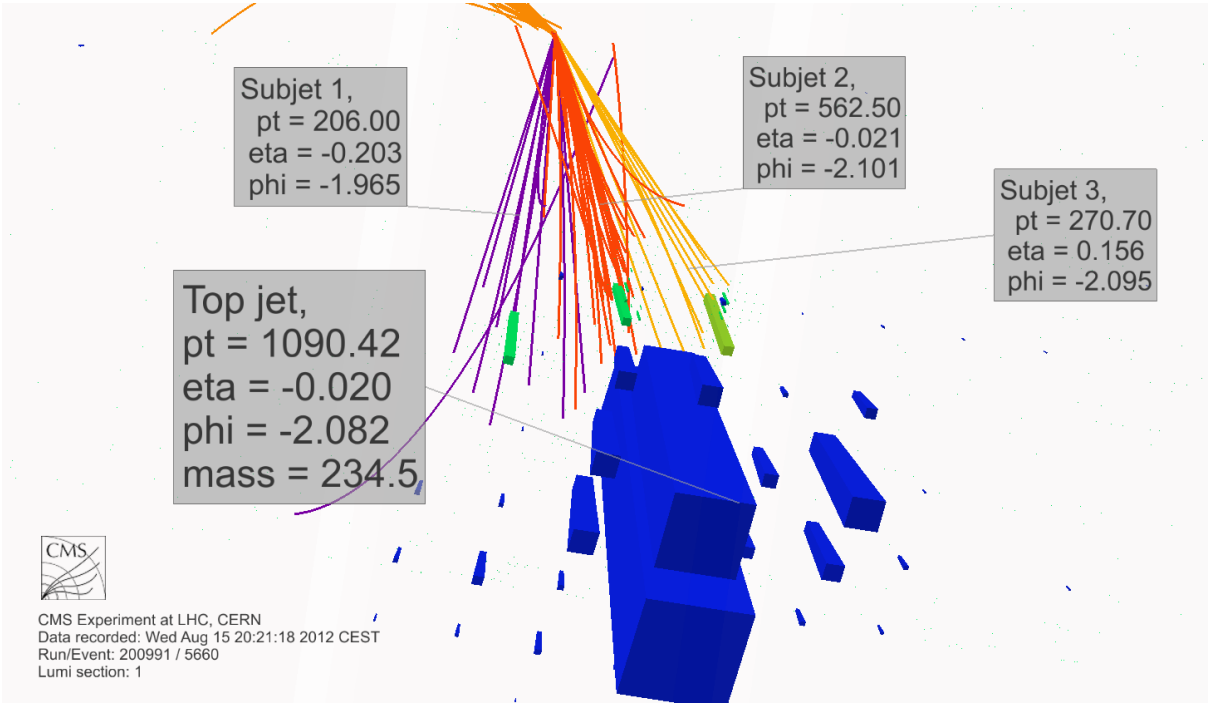




Introduction

- **Jet substructure** is the solution to all of these problems
  - Look inside of a reconstructed jet
- Can resolve individual decay products at high p_T
- Can mitigate degradation effects due to high pileup activity
- Enhances search power for very high-mass particles

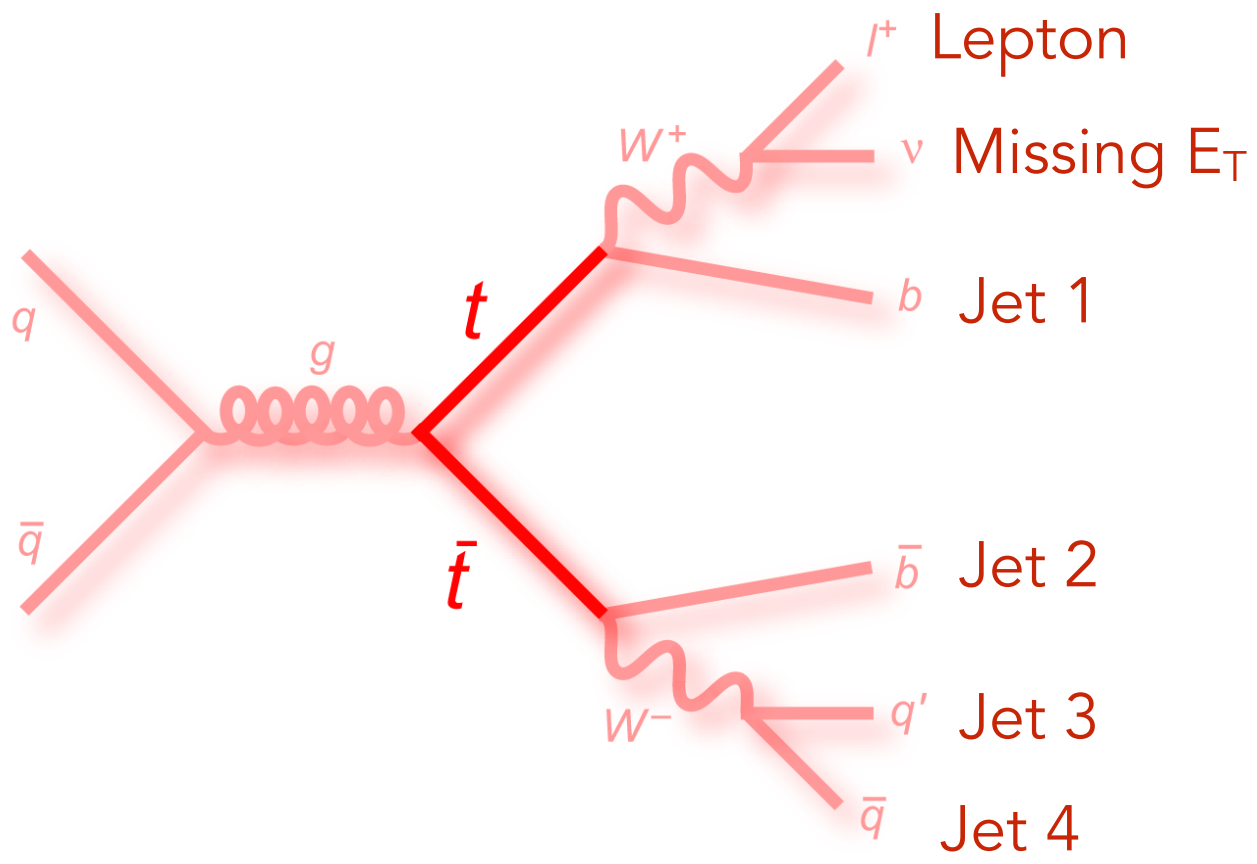
- Today I will discuss development of these algorithms and their use in CMS search analyses
  - Historical Perspective
  - Jet substructure algorithms
  - Search results
  - Future Developments



Jet Substructure Development

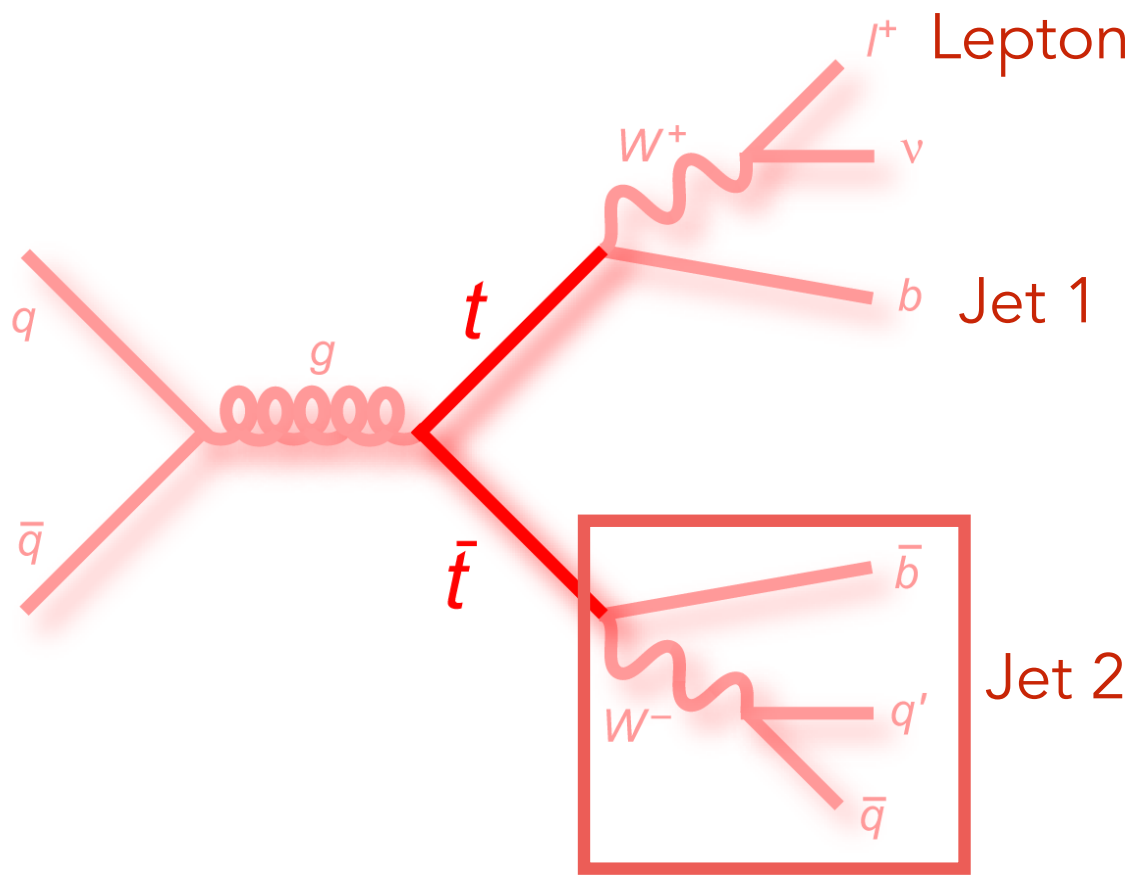
Historical Perspective

- Identifying SM ttbar events historically done by associating one object to each final state decay product
  - Combine objects to reconstruct each top quark
- Combinatorics can become wieldy
  - 6+ jets in all-hadronic decay mode!



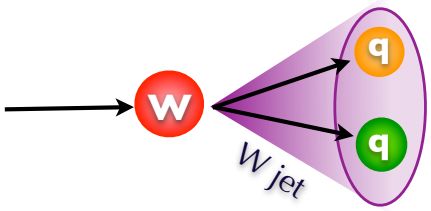
Historical Perspective

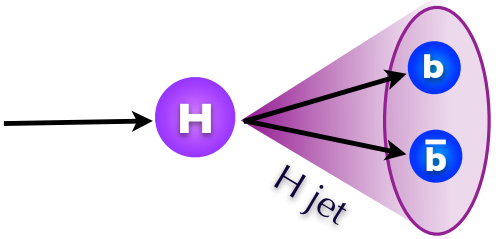
- If the top quarks are boosted, what happens to the number of objects?
  - Hadronic decay products reconstructed in the same final-state object!
- Hadronic final states now become accessible with a dijet final state (in this case)

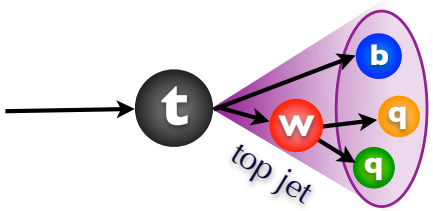


Historical Perspective

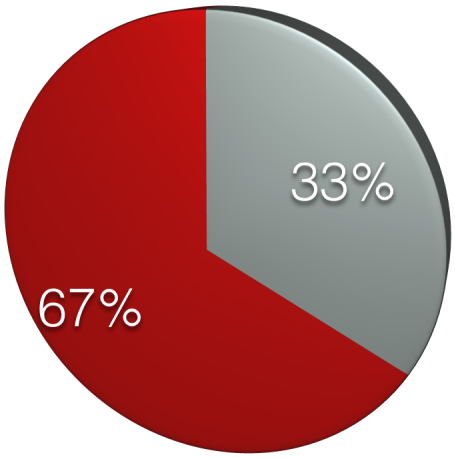
- These merged decays can be used in other cases as well
  - W, Z, Higgs bosons
- Large amount of acceptance can be gained from hadronic channels!
- How to identify these objects?



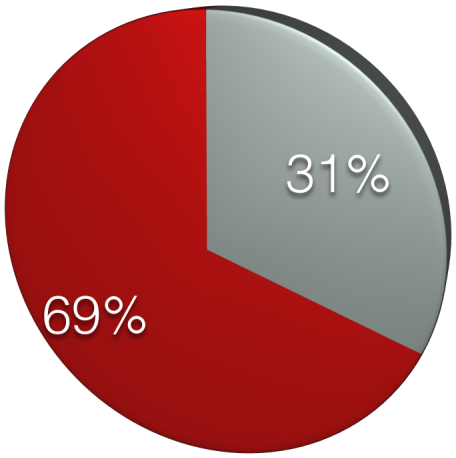




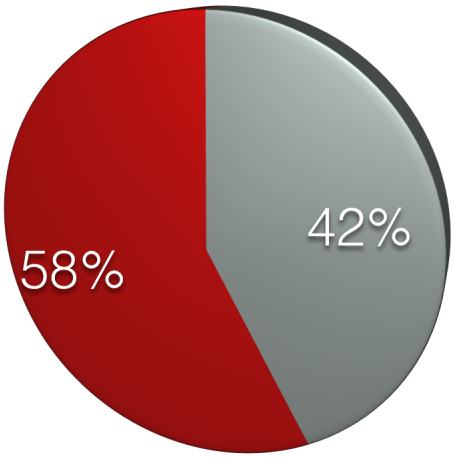
W Boson Decays



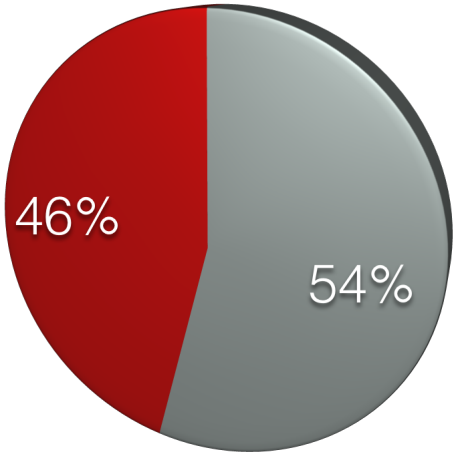
Z Boson Decays



H Boson Decays



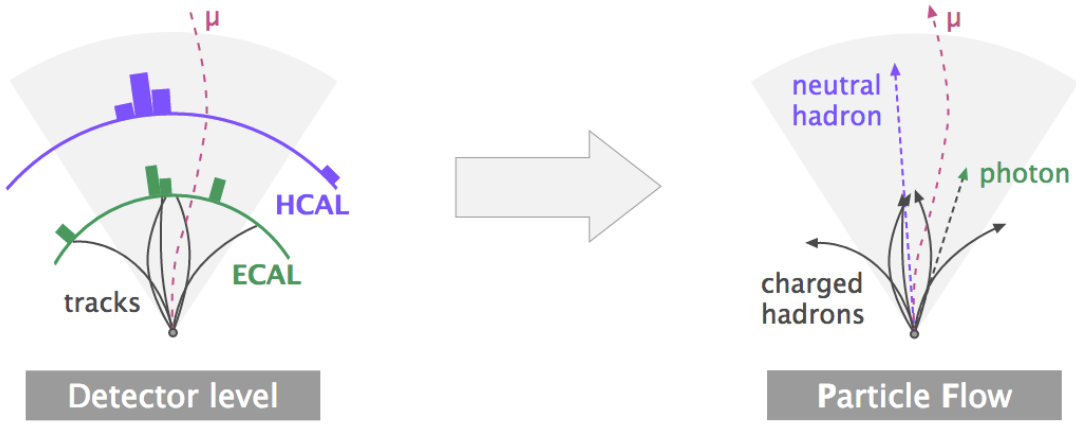
Top Pair Decays



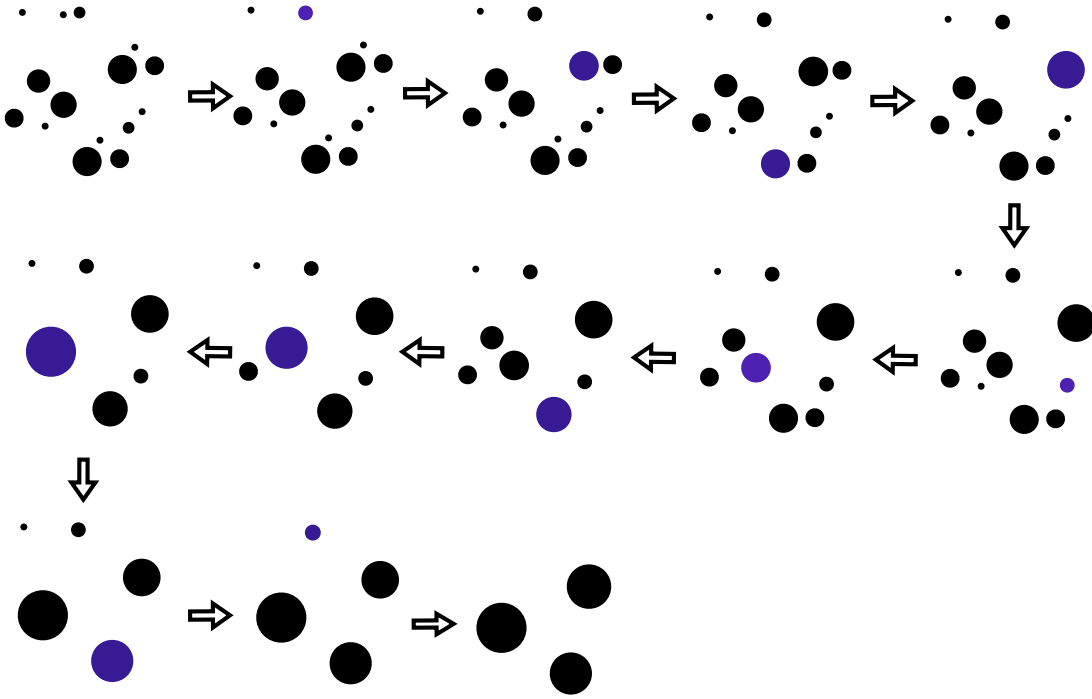
Hadronic Other

Jet Reconstruction

- CMS uses the Particle Flow algorithm to reconstruct all particles produced in a collision event
  - Use information and signals from all detector components



- From this list of particles one can form **jets**, an object to reconstruct the shower of particles produced from a quark or gluon
  - Clustering algorithm used with size parameter *R* ("jet cone size")
  - If $d_{ij} < d_{ii}$, combine constituents

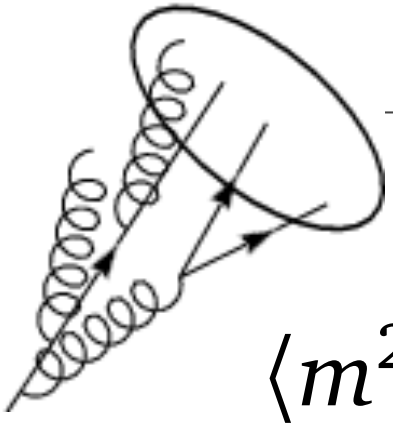


$$d_{ii} = p_{T,i}^{2p}$$

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2}$$

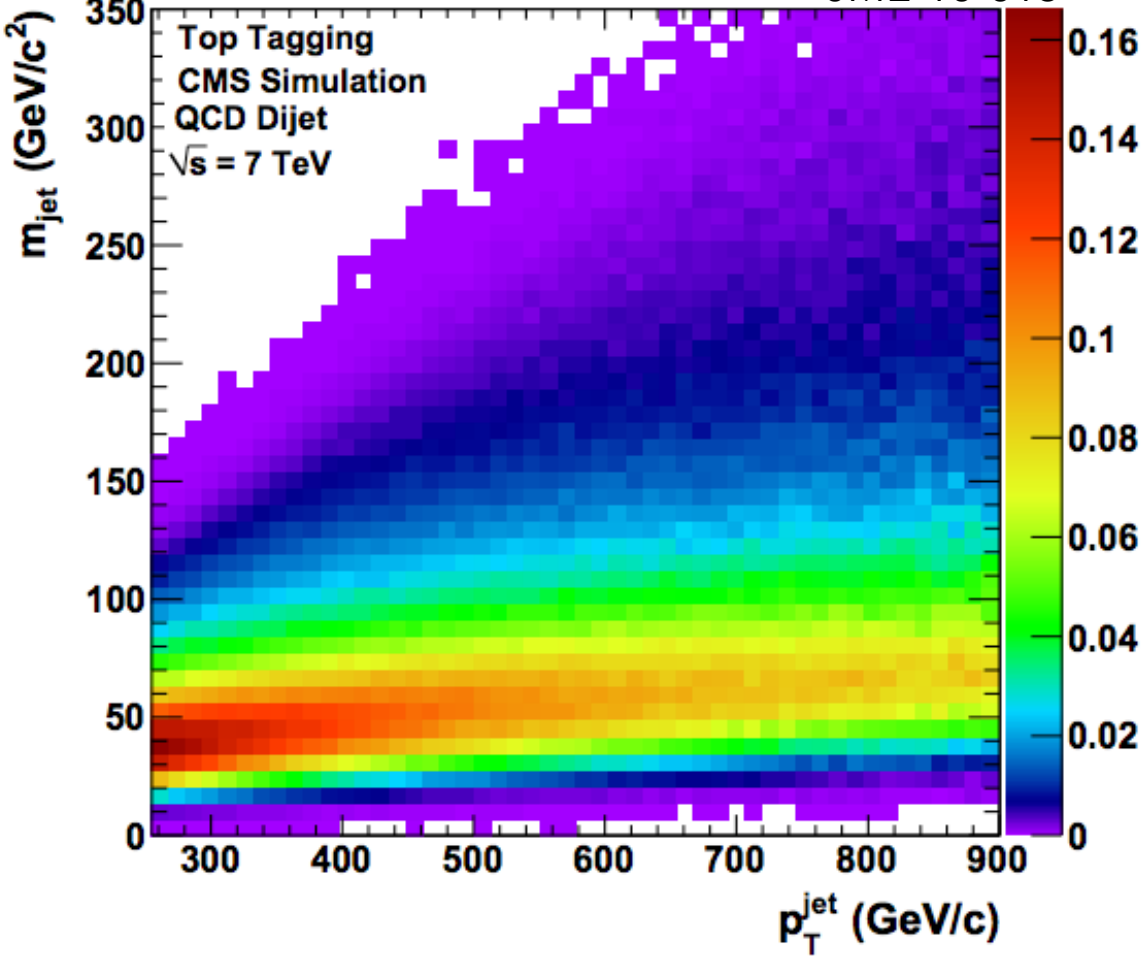
Jet Mass

- One important quantity to identify these jets is the jet mass
- Computed by adding up constituent particle 4-vectors and computing the mass
- For QCD jets, this depends on:
  - The momentum (p_T)
  - The size of the jet used (R)



$$\langle m^2 \rangle \sim \frac{C\alpha_S}{\pi} p_T^2 R^2$$

JME-10-013

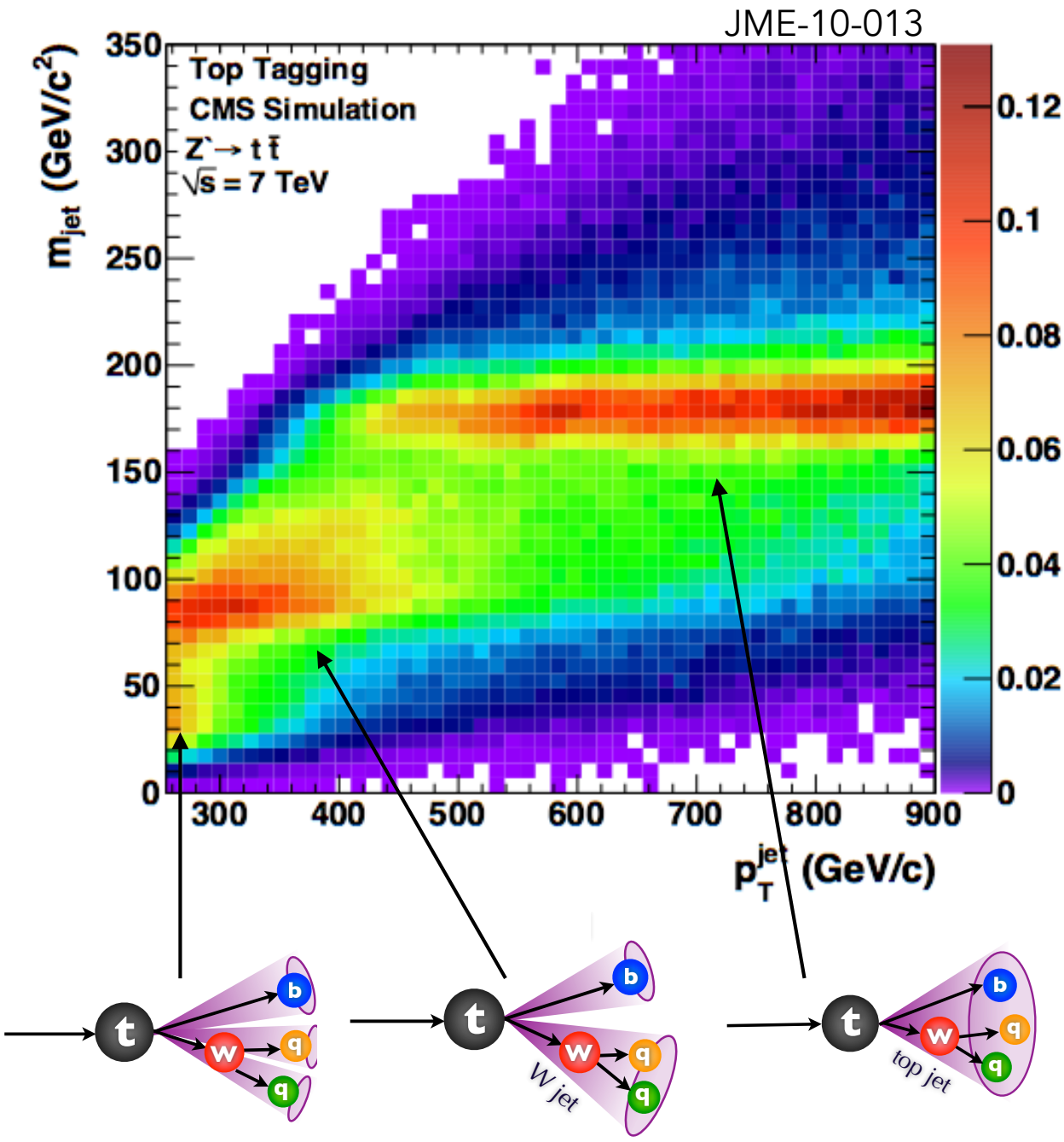


Jet Mass

- For heavy objects, the picture is different
- The jet mass is relatively stable at the heavy particle mass
- Choose correct cone size to reconstruct all decay products in jet:

$$\Delta R \sim \frac{2 \cdot m_{\text{particle}}}{p_T}$$

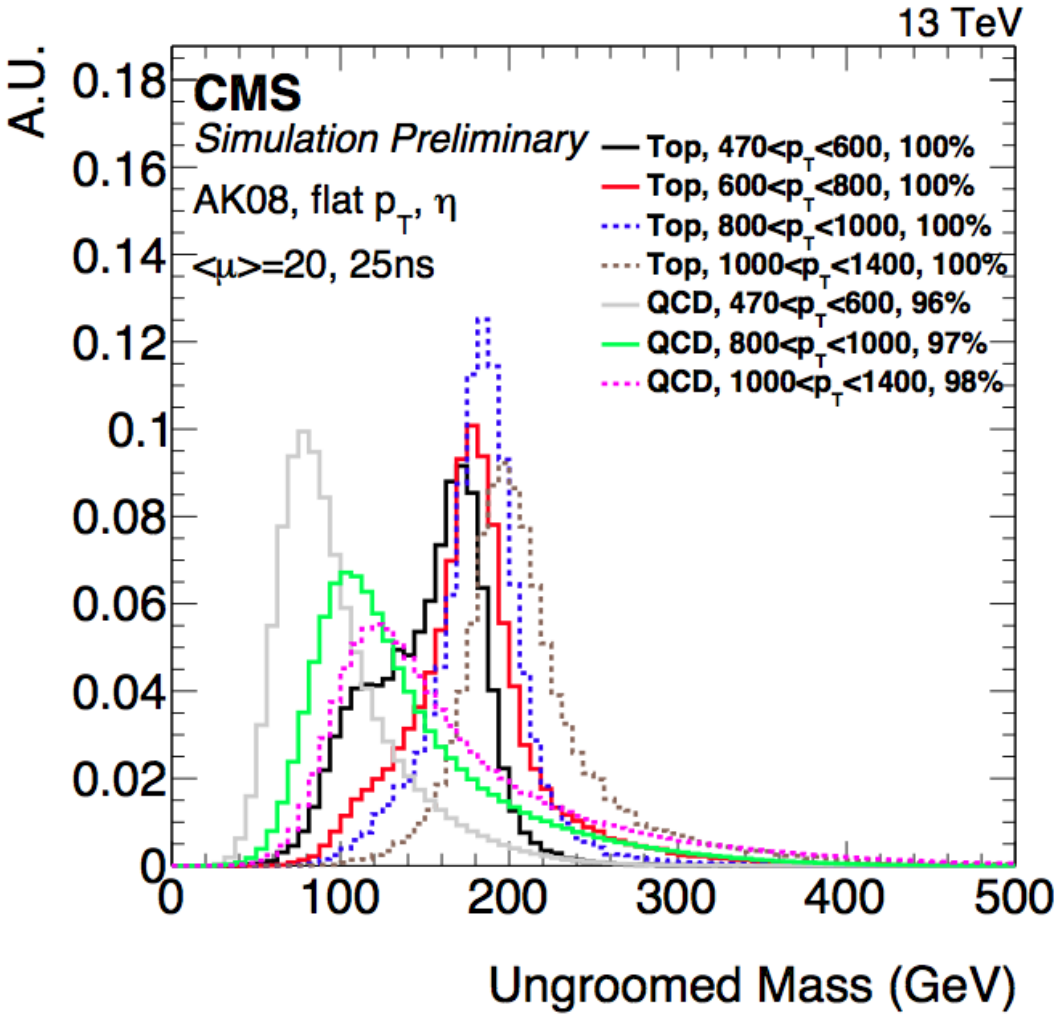
- CMS uses R = 0.8 for heavy object reconstruction
  - Merged W/Z at ~200 GeV
  - Merged top at ~400 GeV



Jet Mass

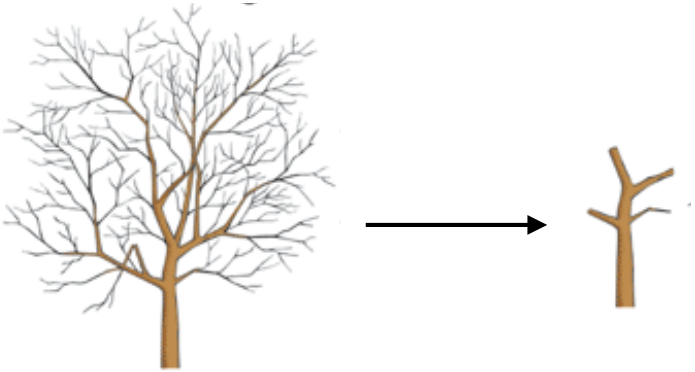
JME-15-002

- Some discrimination obtained when using this ‘raw’ jet mass
- We can do better by looking inside the jet at the individual constituents
- Using **jet grooming** algorithms can improve the discrimination between QCD and top quark jets

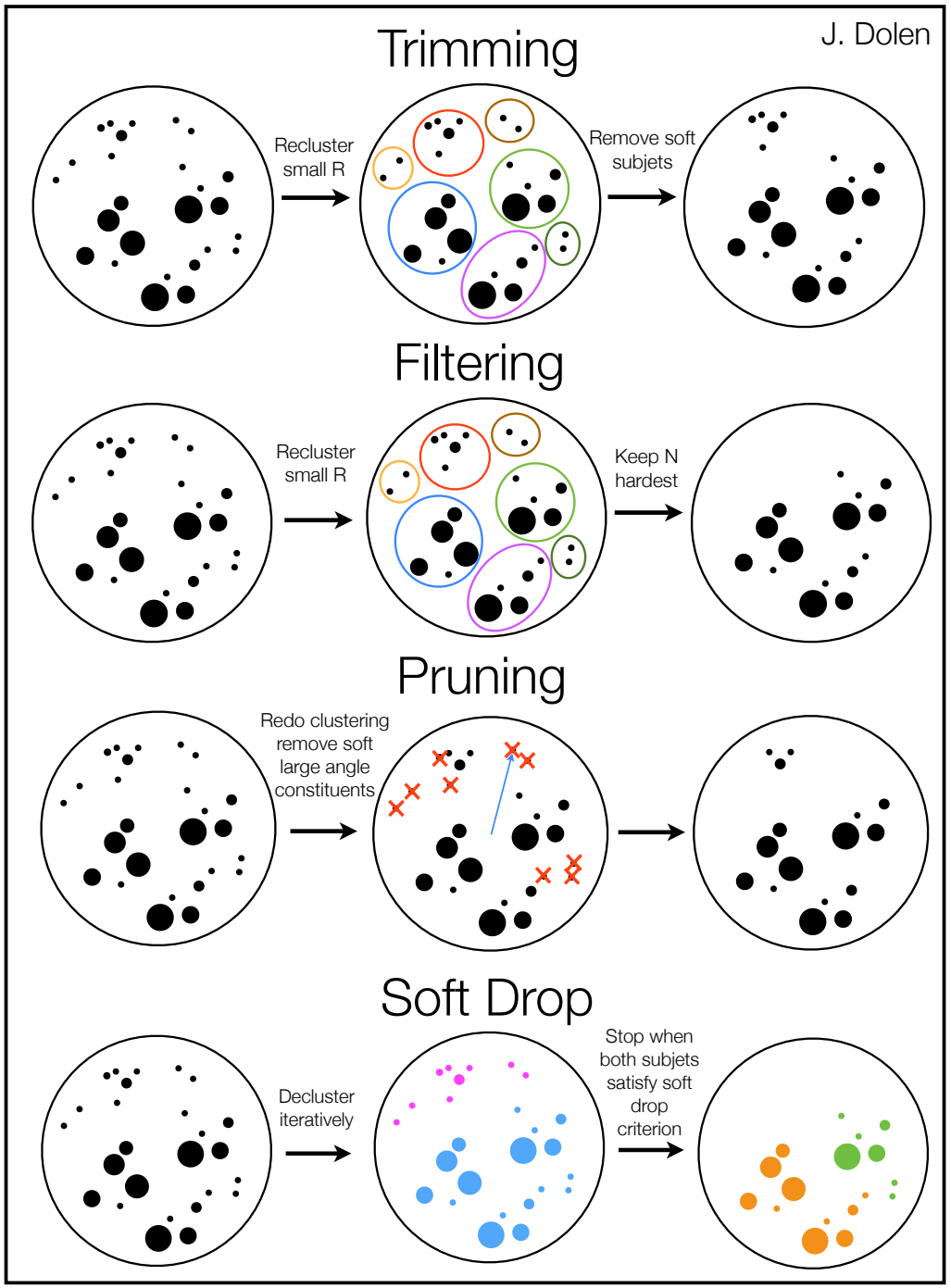


Jet Grooming Algorithms

- Many different approaches
- Standard idea: remove **soft and wide-angle radiation** from within the jet
  - Recluster with smaller R
  - Remove subjets
  - Remove constituents during clustering



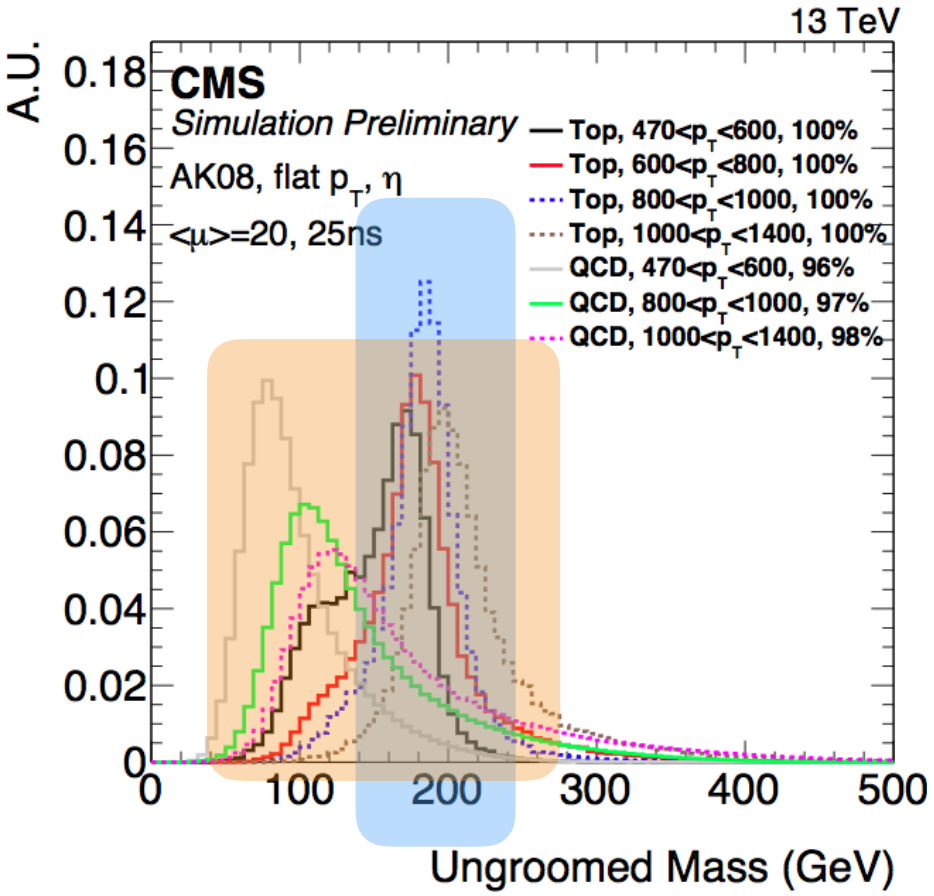
- Soft drop algorithm chosen for latest analyses

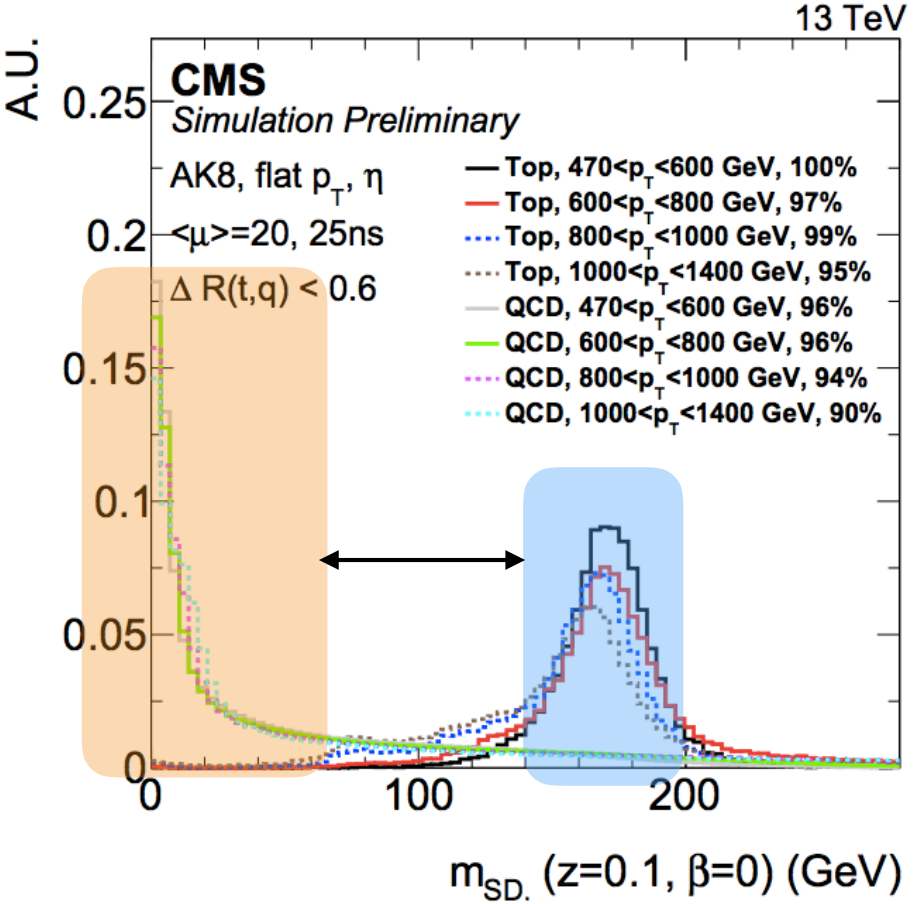


Jet Grooming Algorithms

JME-15-002

- Dramatically improves the separation of QCD and top quark jets
- Merged top quarks can be identified with a window around the top quark mass



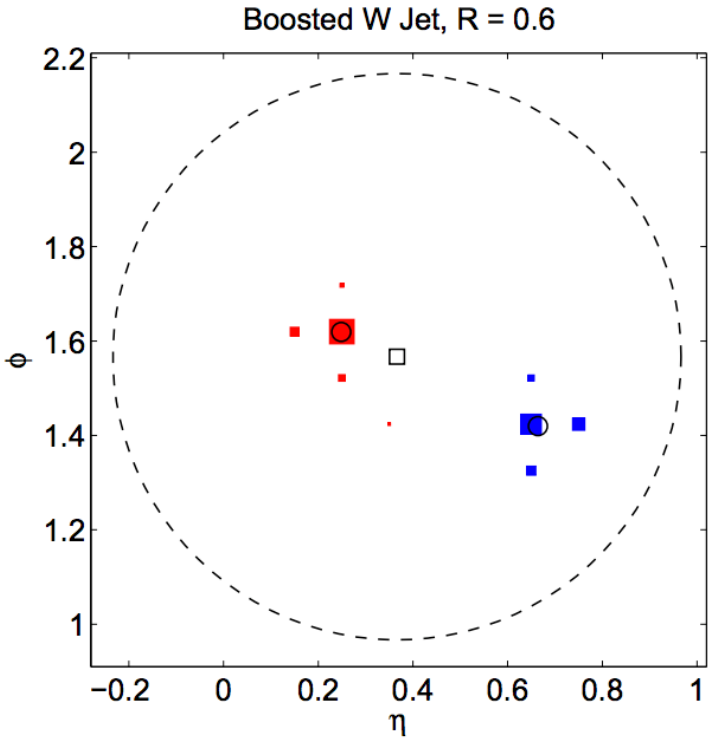


Topological Algorithms

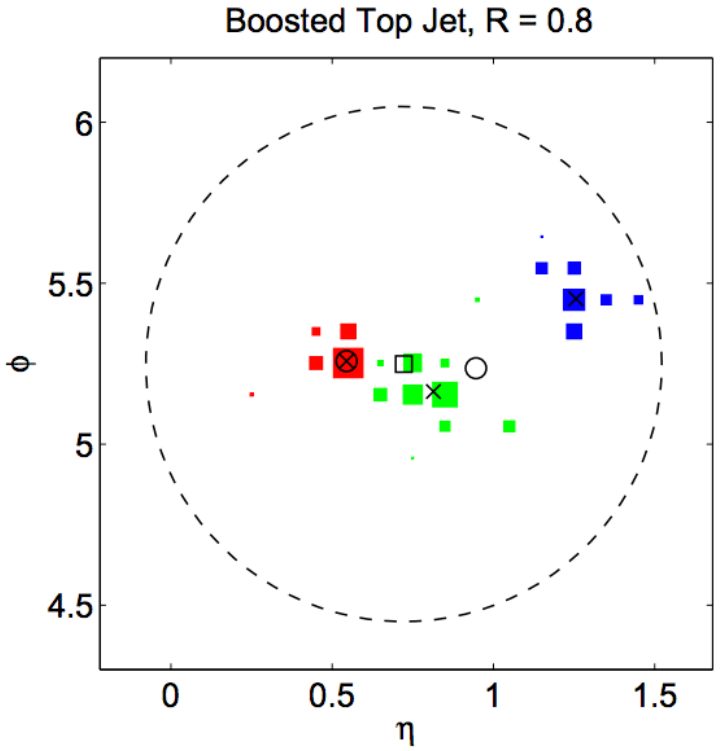
arXiv:1011.2268

- We know how many final state objects to expect from these decays
  - Can look inside the jet for the expected substructure
    - Top decays → 3 subjets
    - W/Z/H decays → 2 subjets
- A quantity called N-subjettiness is a measure of how consistent a jet is with a hypothesized number of subjets
  - Low τ_N → consistent with N (or fewer) subjets
  - Ratios used for additional discrimination

$$\tau_N = \frac{1}{\sum_i p_{T,i} \cdot R} \sum_i p_{T,i} \cdot \min(\Delta R_{1,i}, \Delta R_{2,i}, \ldots, \Delta R_{N,i})$$



□ 1 subjet hyp.
○ 2 subjet hyp.
× 3 subjet hyp.



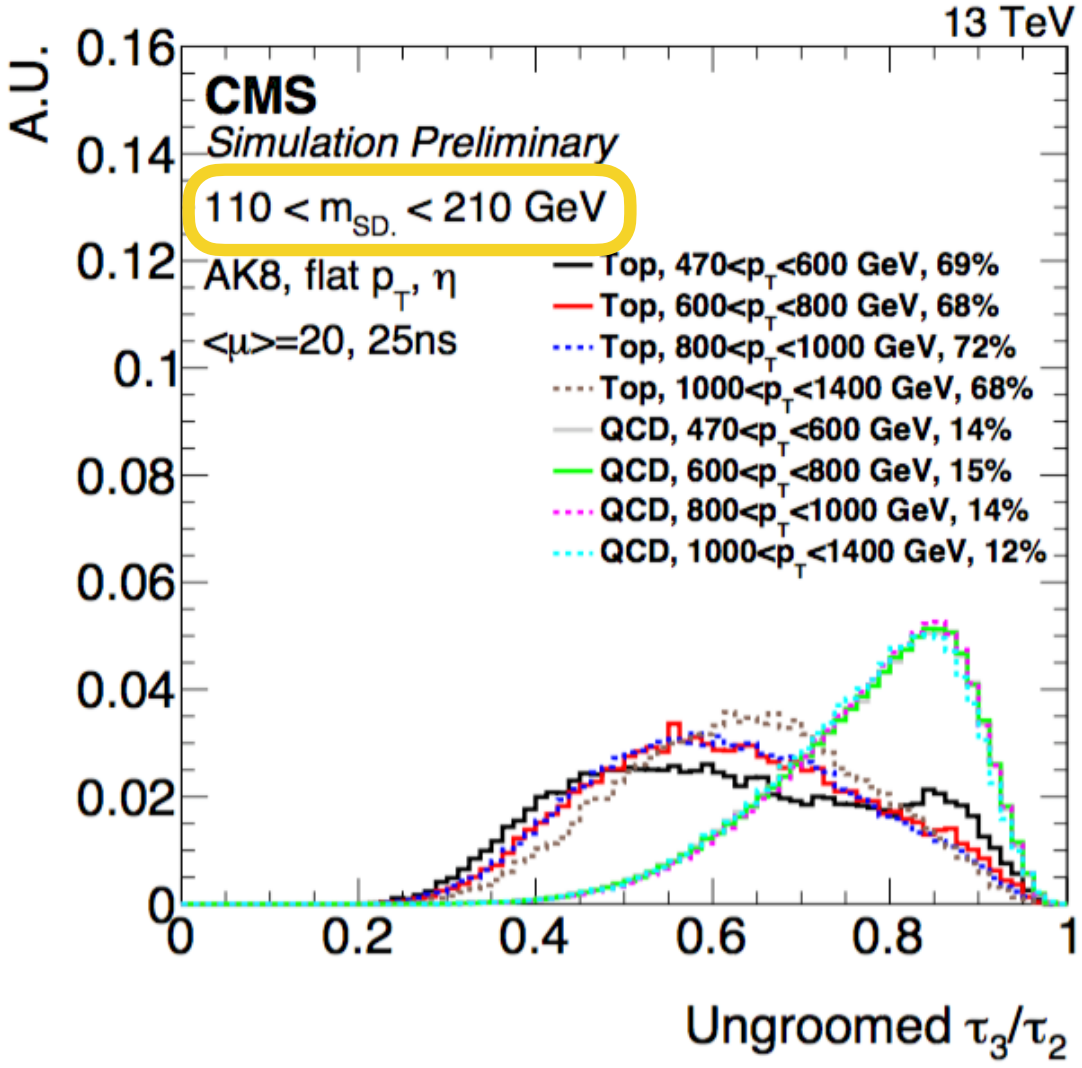
N-Subjettiness

JME-15-002

- Ratios used for discrimination
  - τ_3 / τ_2 for top quark jets
  - τ_2 / τ_1 for W/Z/H jets

- Provides additional power when used in conjunction with the groomed jet mass

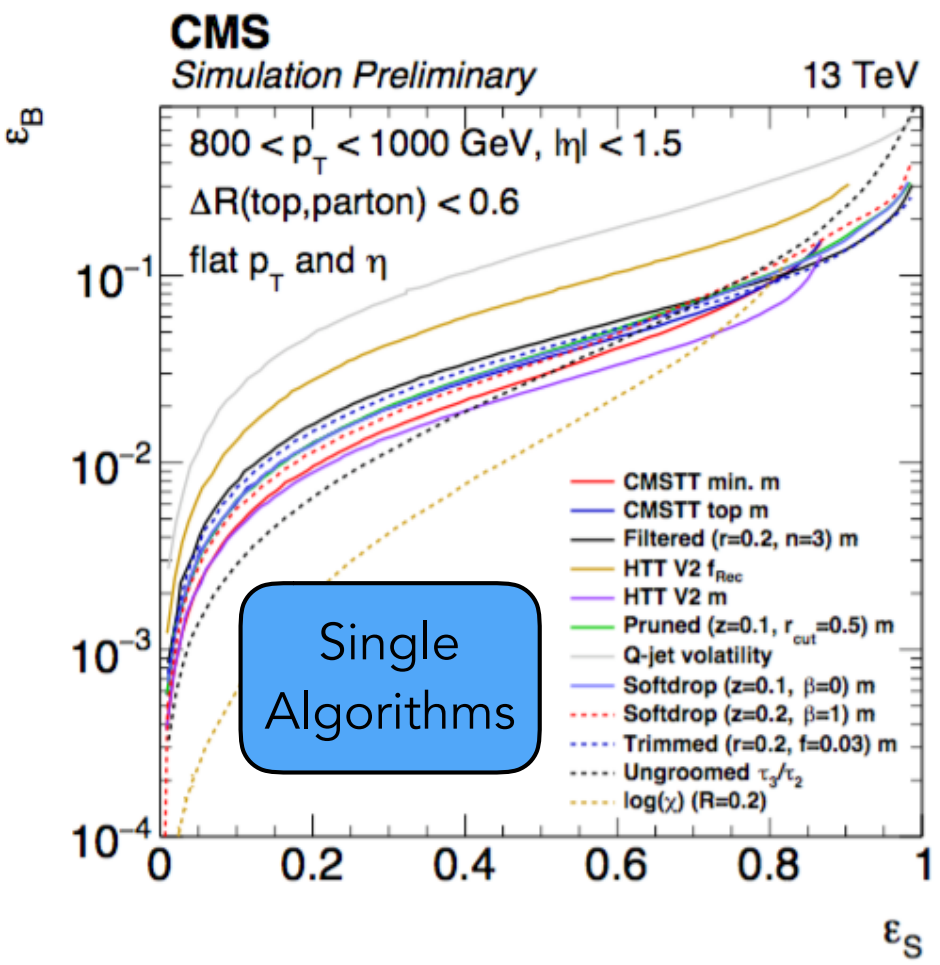
- Can we do better?

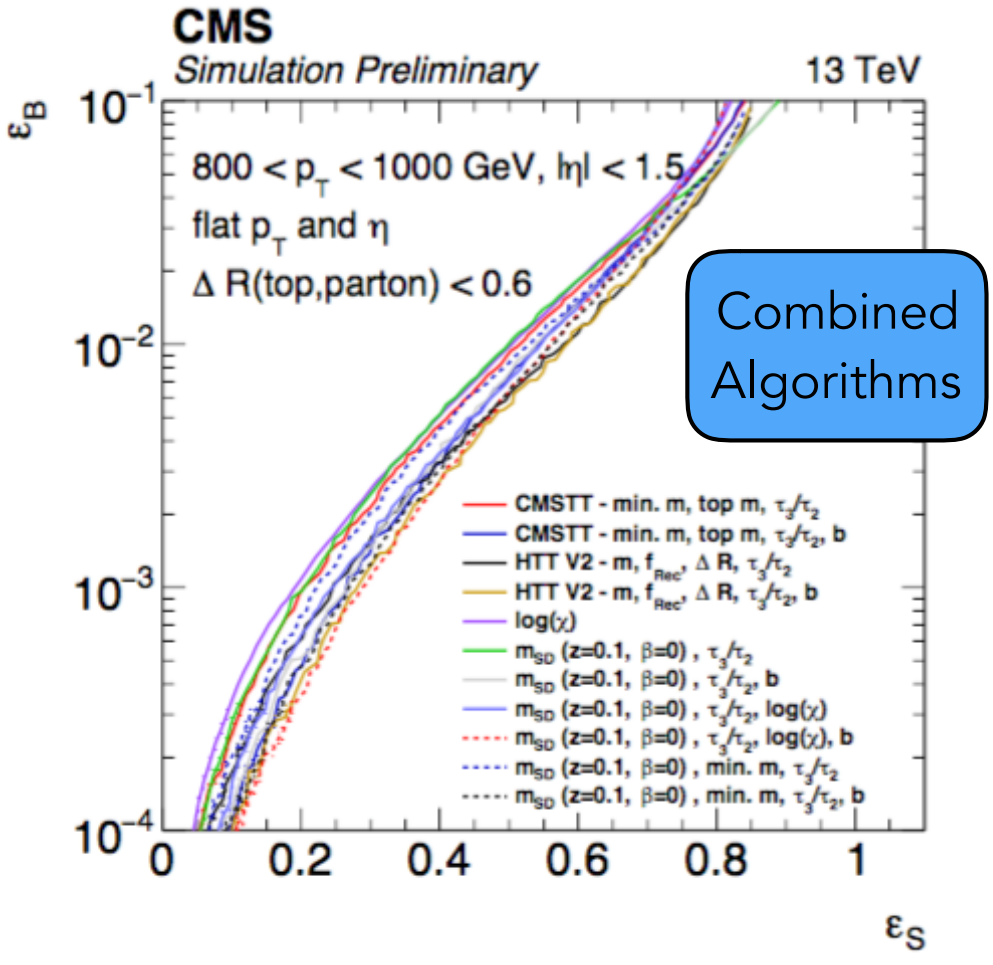


Combining Algorithms

JME-15-002

- The algorithms provide mutual information that increases performance
- Best combination — used for top quark identification in 13 TeV analyses
  - **Soft-drop mass + N-subjettiness (+ b-tagging)**





Working Points

JME-15-002

- Choose cut values to ensure stability across both p_T and pileup activity
  - m_{SD} in [110, 210] GeV
  - $\tau_{32} < 0.69$
  - Slight degradation in efficiency as a function of pileup

$M_{SD}(z = 0.1, \beta = 0) + \tau_3/\tau_2$

$M_{SD}(z = 0.1, \beta = 0) + \tau_3/\tau_2$ + b tag

CMSTT Mass + τ_3/τ_2

CMSTT Mass + τ_3/τ_2 + b tag

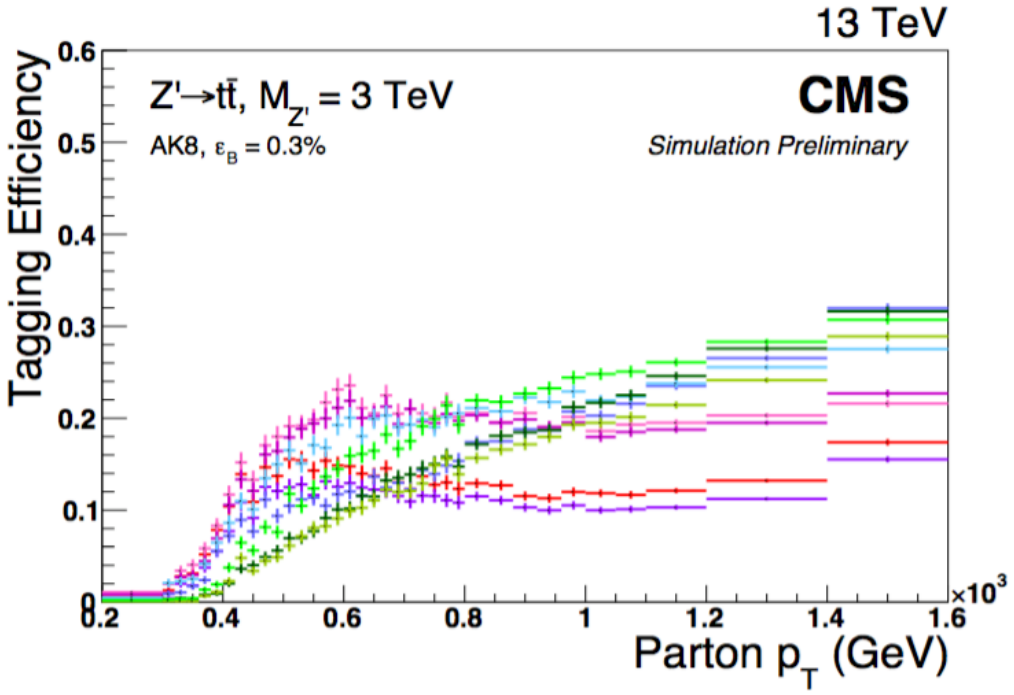
HTT V2 Mass + $f_{Rec} + \tau_3/\tau_2 + \Delta R$

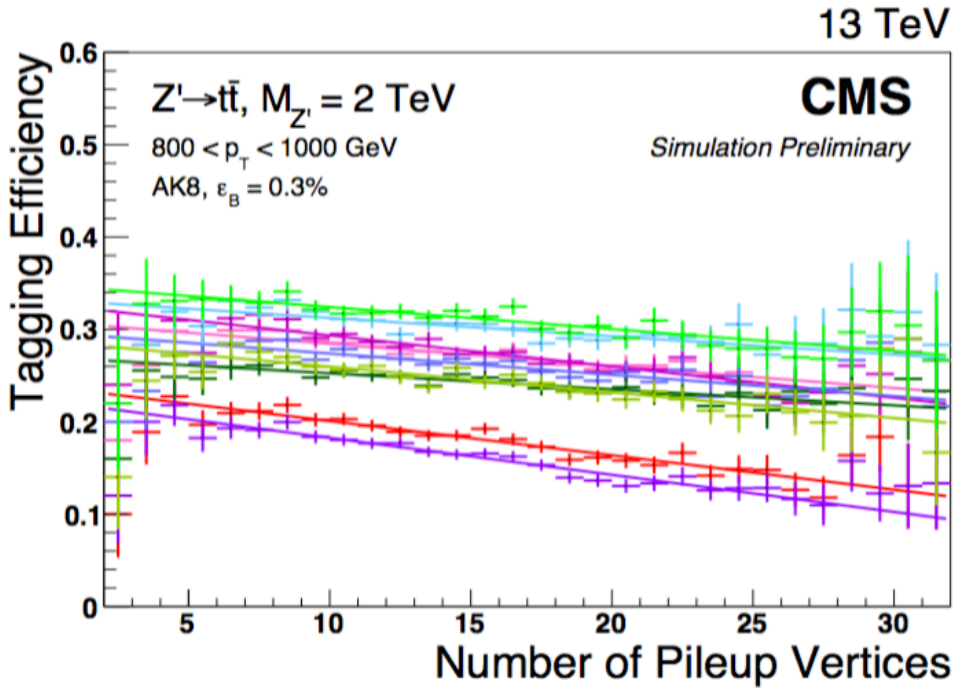
HTT V2 Mass + $f_{Rec} + \tau_3/\tau_2 + \Delta R$ + b tag

log(χ)

$M_{SD}(z = 0.1, \beta = 0) + \log(\chi) + \tau_3/\tau_2$

$M_{SD}(z = 0.1, \beta = 0) + \log(\chi) + \tau_3/\tau_2$ + b tag

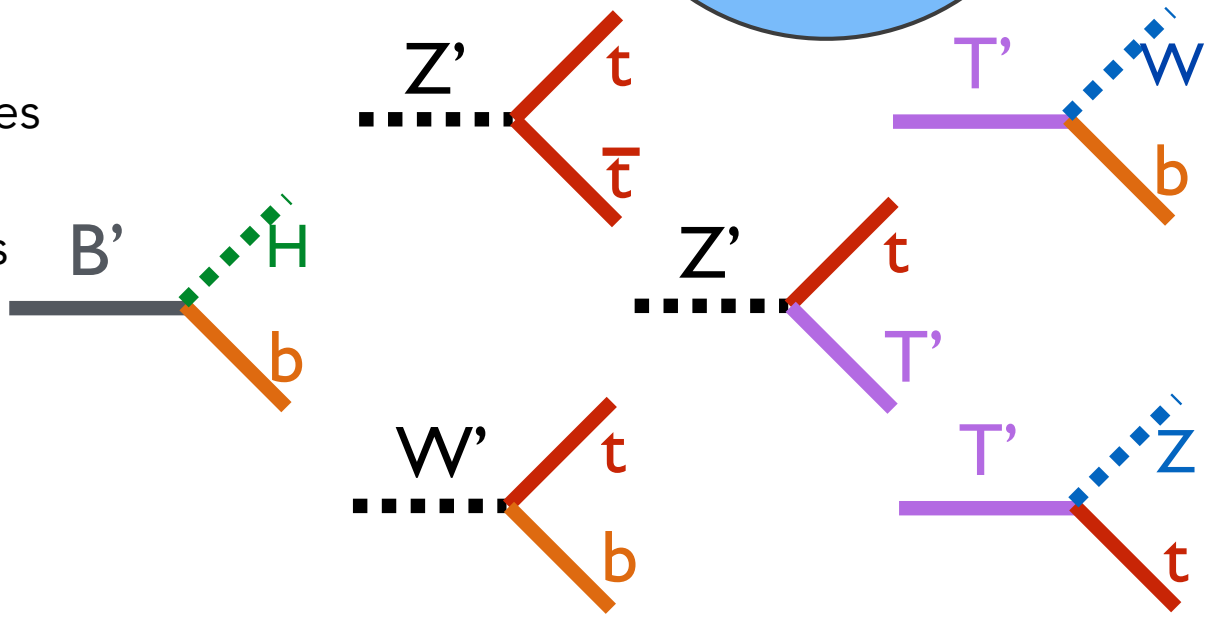




Searches with Jet Substructure

CMS B2G Group

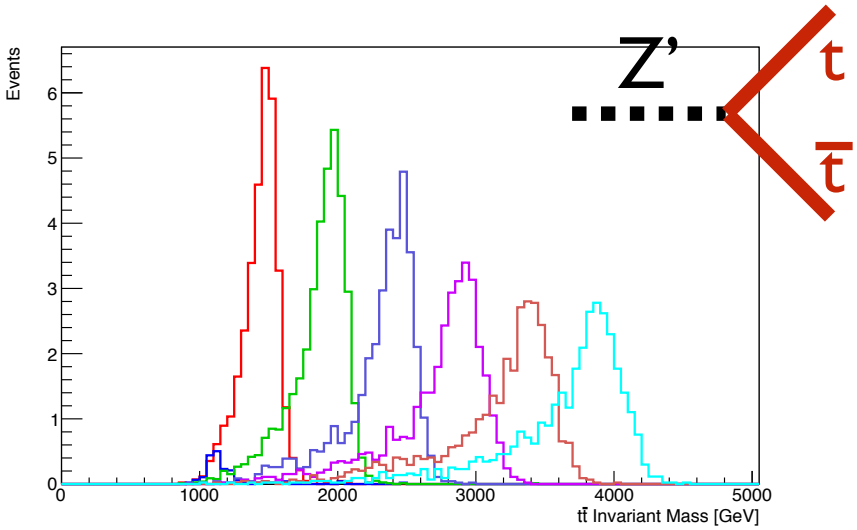
- The 'Beyond 2 Generations' group in CMS is focus on searches which decay to heavy objects (t/W/Z/H)
- **Our specialty is jet substructure and heavy object tagging!**
- There are many searches which take advantage of these signatures
  - Top pair resonances
  - Top + bottom resonances
  - Diboson resonances
  - Heavy vector-like quarks
  - Hybrid decay modes



Top Quark Pair Resonances

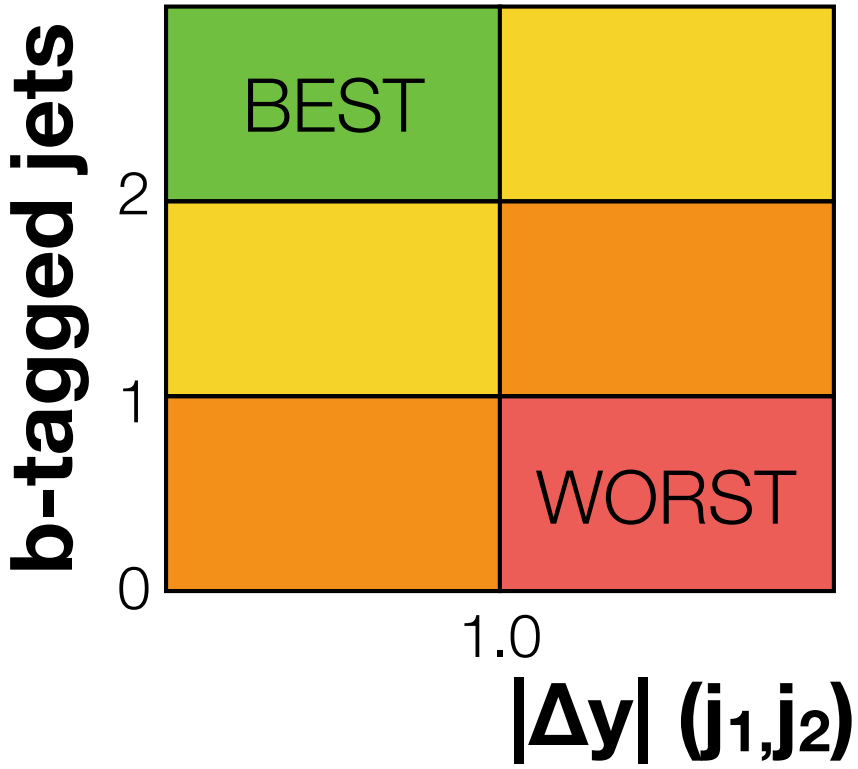
B2G-15-003

- Hadronic channel — dijet event topology



- Search for events with 2 back-to-back top-tagged jets
  - p_T > 400 GeV
  - m_{SD} in [110, 210] GeV
  - τ_{32} < 0.69

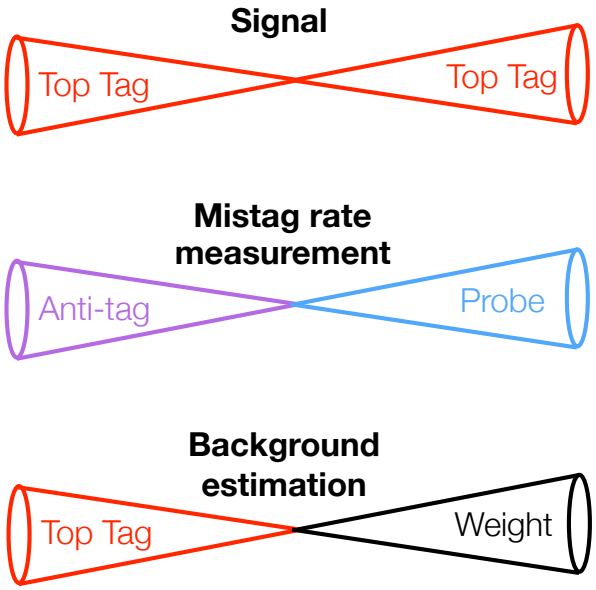
- Divide events into categories based on number of subjet-b-tagged jets and rapidity separation
- Use top pair invariant mass to discriminate signal



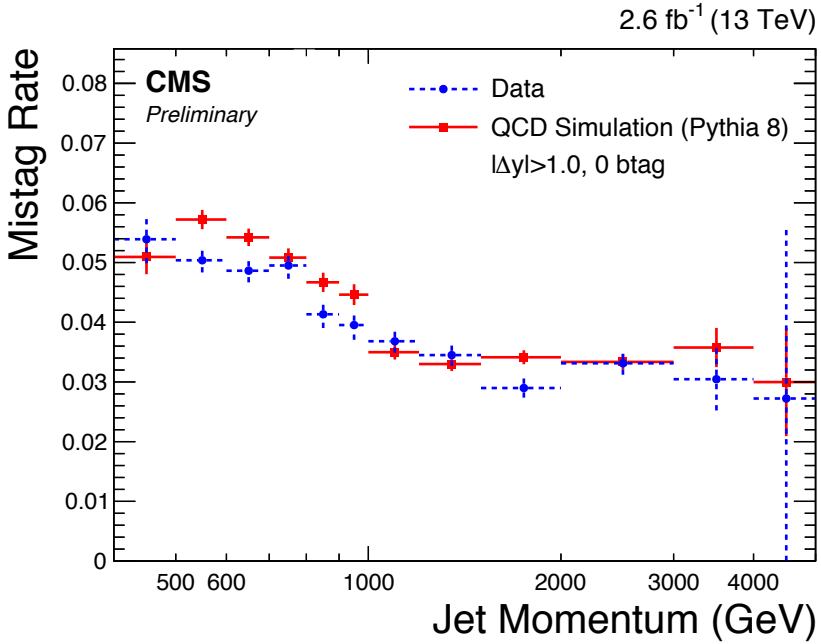
Top Pair Resonances

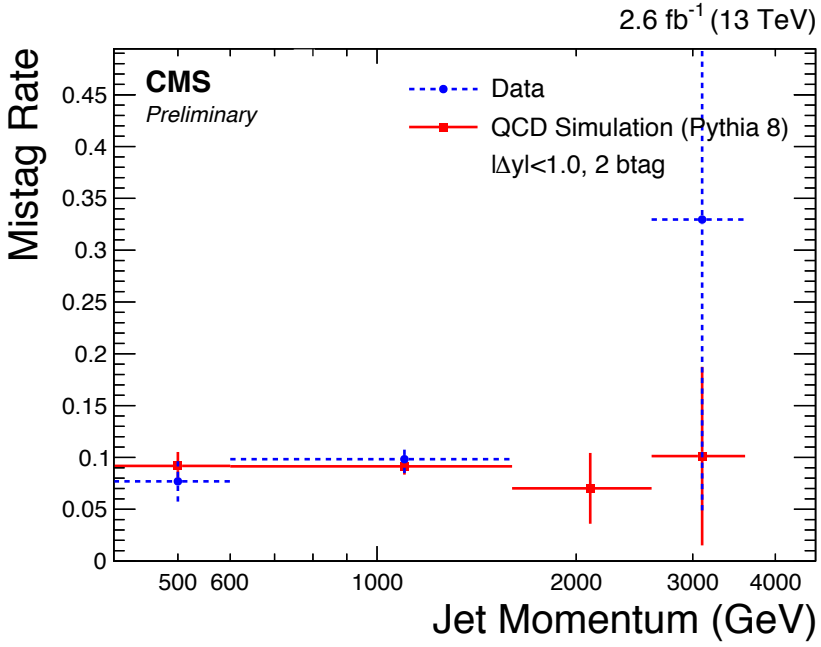
B2G-15-003

- Mistag rate measured in an anti-tag and probe sample
  - Function of jet p (not p_T) to account for eta-dependence



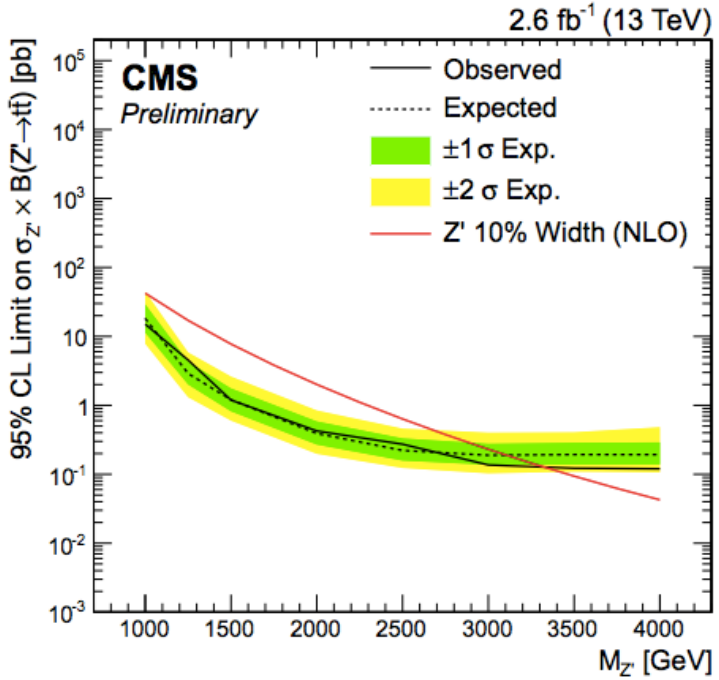
- One rate for each category of the analysis
  - Applied to sample of single-top-tagged events





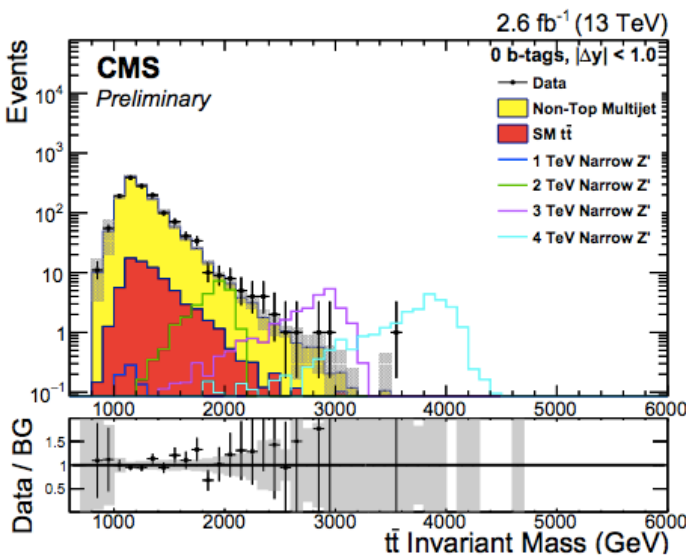
Top Pair Resonances

- Good agreement seen in discriminating distributions
- Set limits ranging from 1.6 - 3.8 TeV depending on the signal model considered

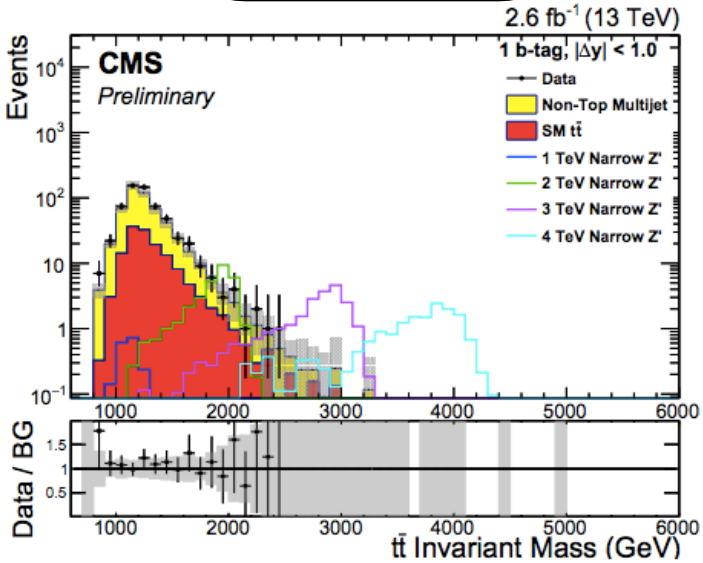


B2G-15-003

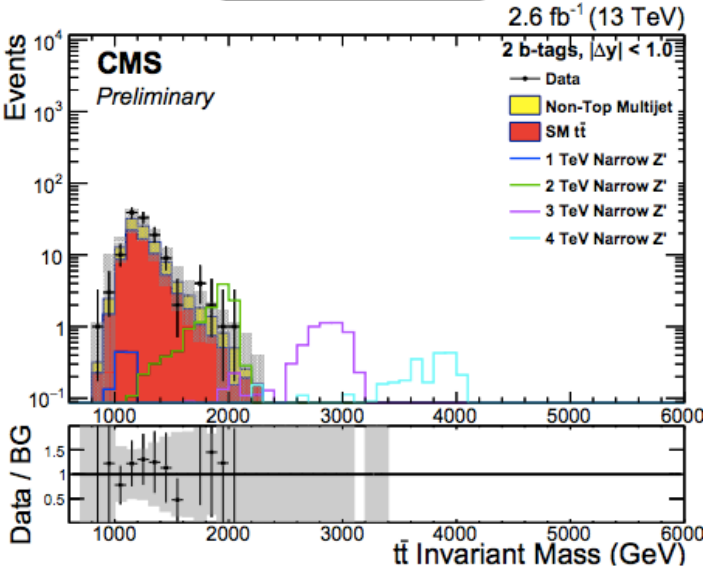
0 b-tag



1 b-tag



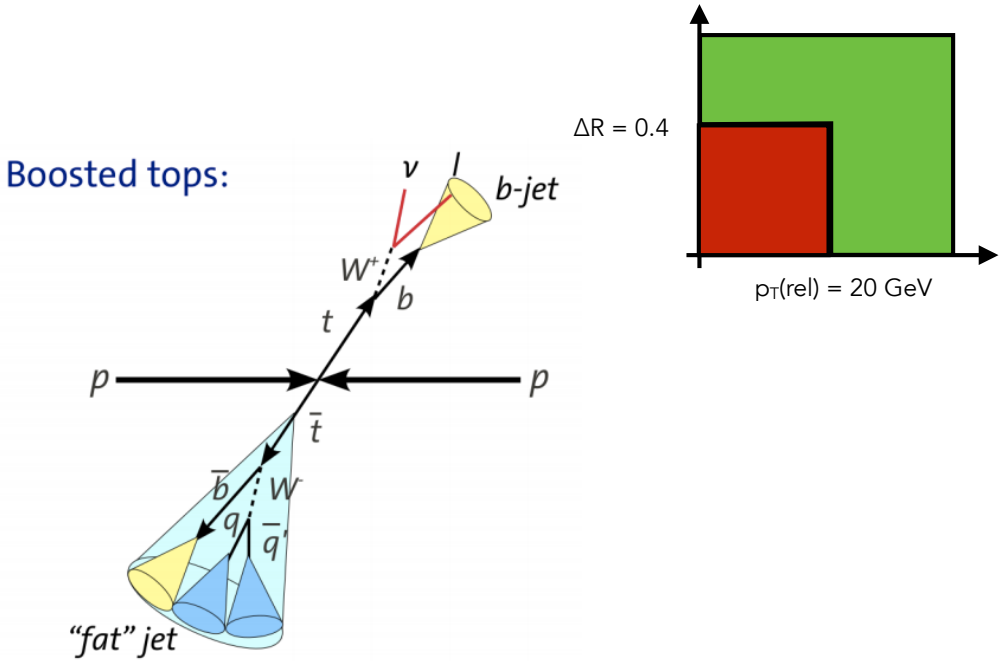
2 b-tag

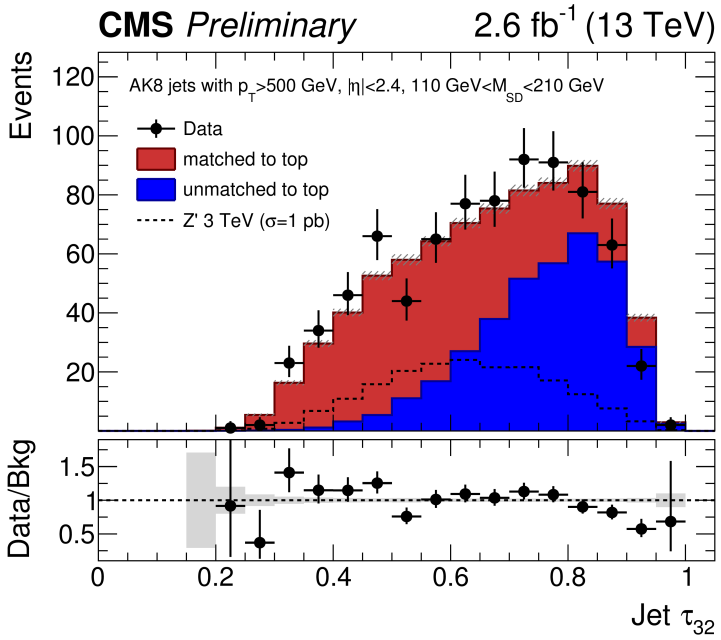


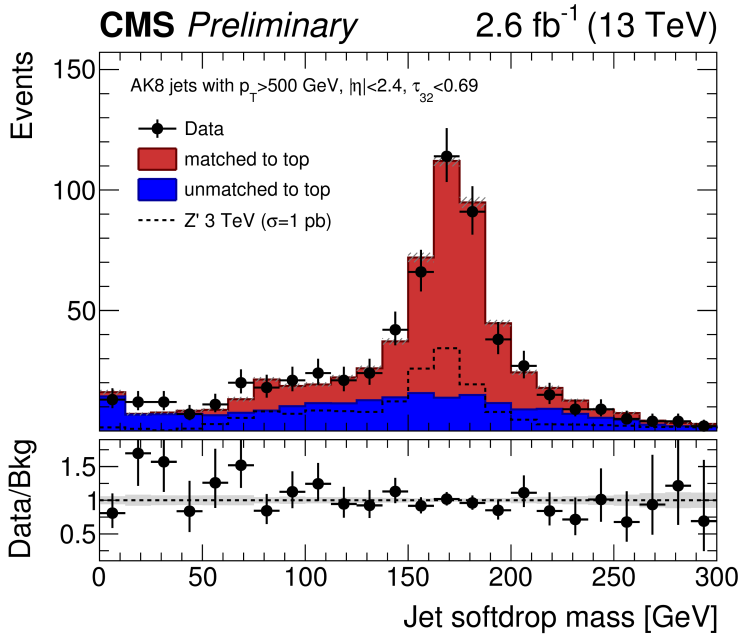
Top Pair Resonances

B2G-15-002

- Also consider the lepton+jets decay mode in this search
  - At high p_T, lepton may overlap with jet → special reconstruction needed to reject QCD
    - $\Delta R > 0.4$ or $p_T(\text{rel}) > 20$ GeV
- Same top-tagging algorithm used

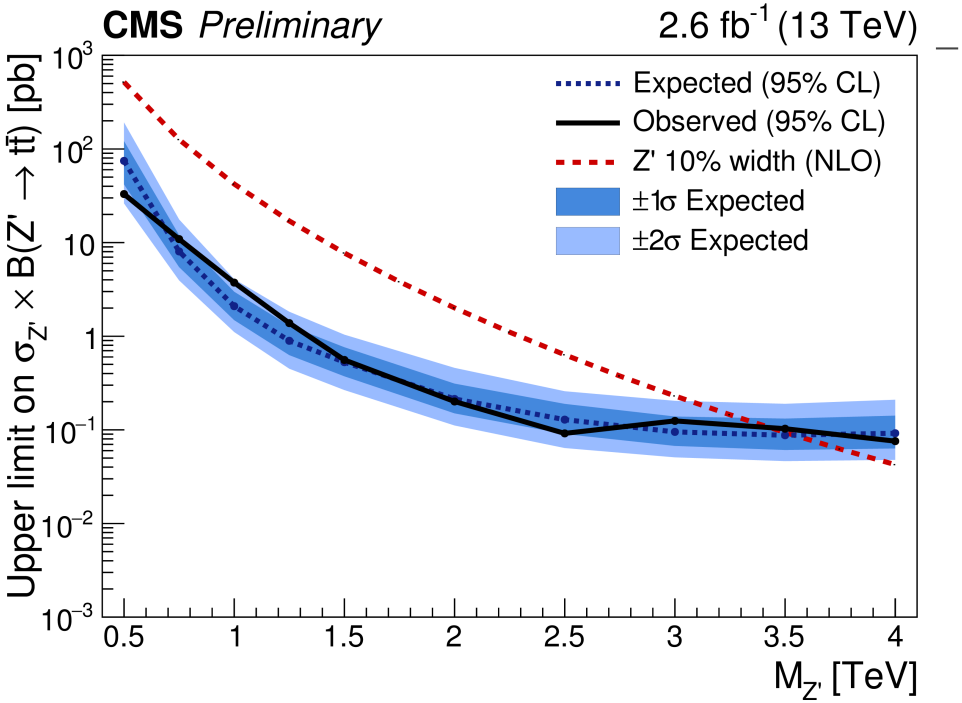


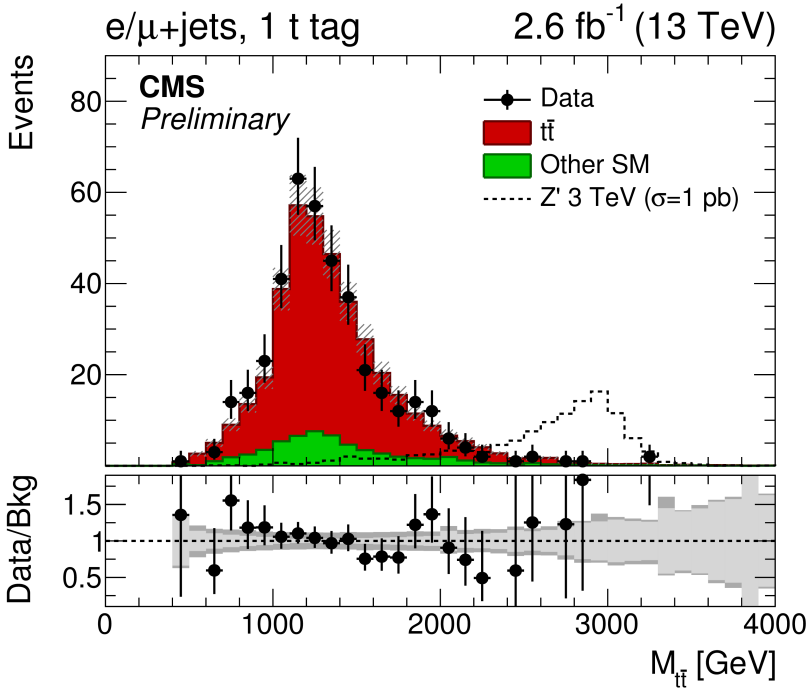


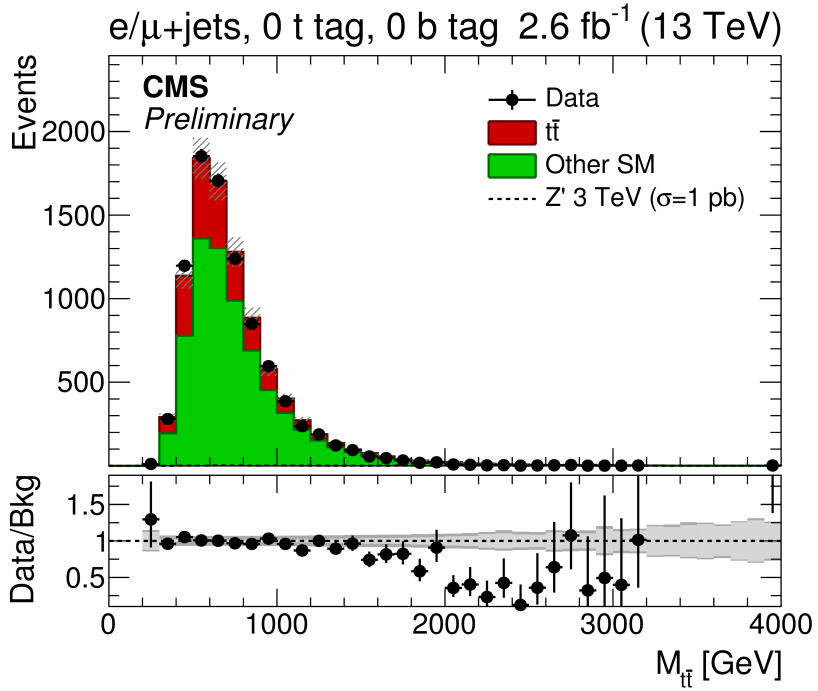


Top Pair Resonances

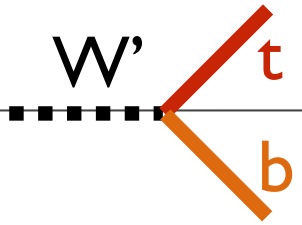
- Events are divided by lepton flavor and number of b/t-tagged jets
  - Top pair invariant mass used for discrimination
- Limits range from 2.3 - 4.0 TeV depending on the physics model





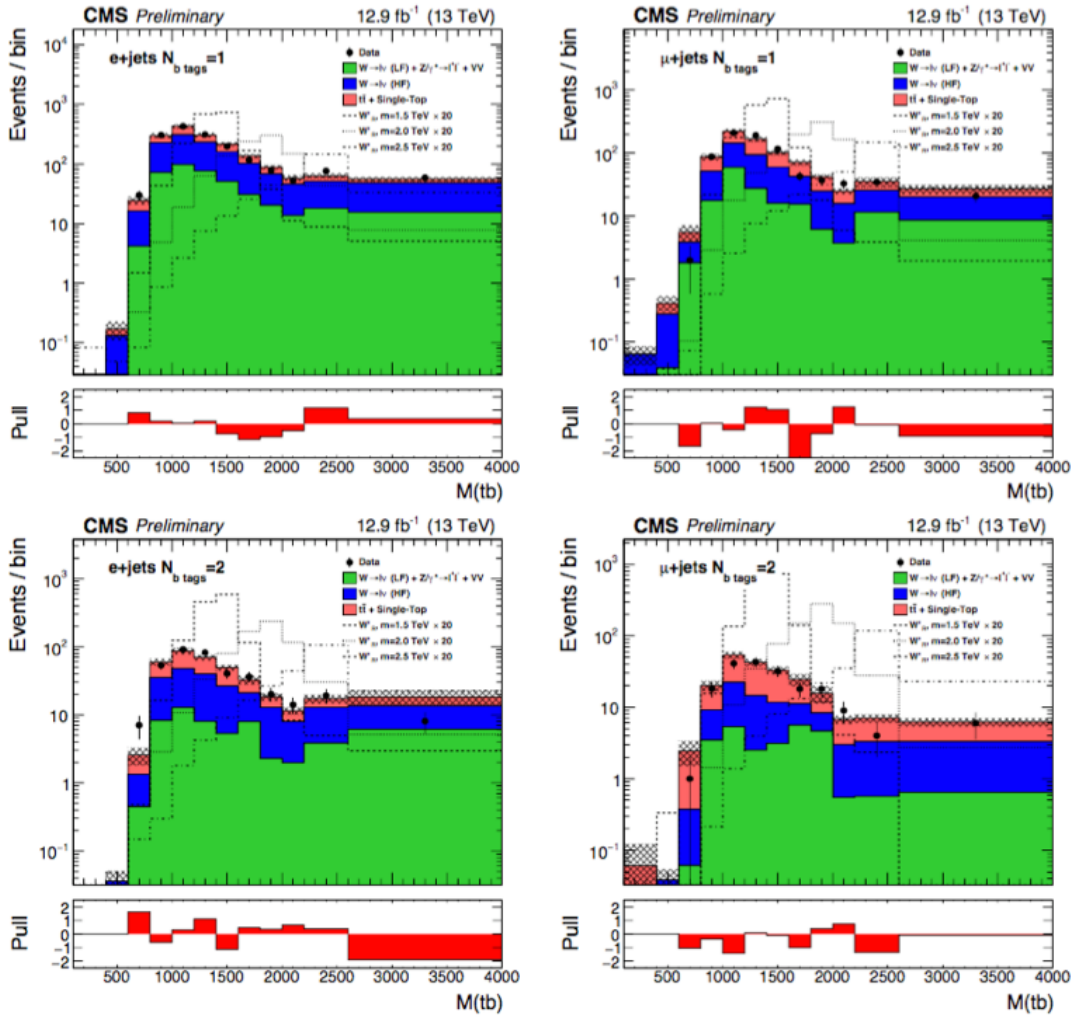


t+b Resonances



B2G-16-017

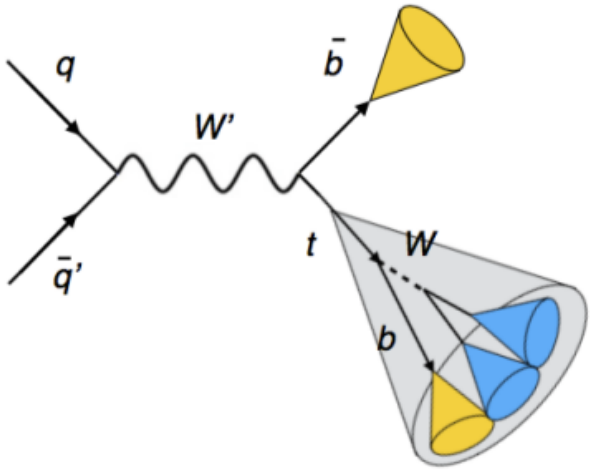
- **W′ → tb → lepton + jets**
  - Single non-isolated muon/electron
  - Leading jet pT > 350 GeV (e), 450 GeV (μ)
  - MET > 50 GeV (μ), 120 GeV (e)
- Kinematic reconstruction of top quark and W′ mass
- Main backgrounds ttbar, W+jets taken from simulation with checks in dedicated control regions
- Invariant mass of t+b used for signal discrimination
  - Categories based on lepton flavor and number of b-tags

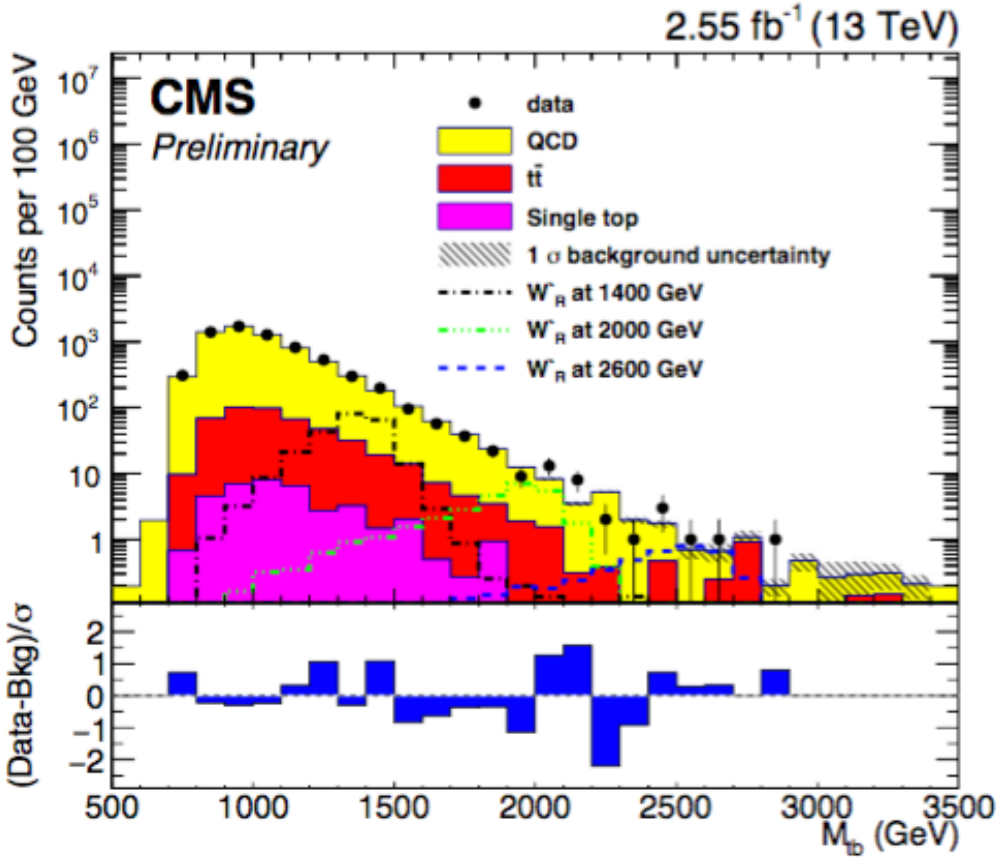


t+b Resonances

B2G-16-009

- The hadronic channel takes advantage of top-tagging algorithms
- Dijet topology
  - Top-tagged jet opposite from a high-pT b-jet
  - Each jet p_T > 350 GeV
  - b-jet mass < 70 GeV to remove ttbar background
- QCD background estimated using data-driven mistag rate
- Dijet (t+b) mass used for signal discrimination





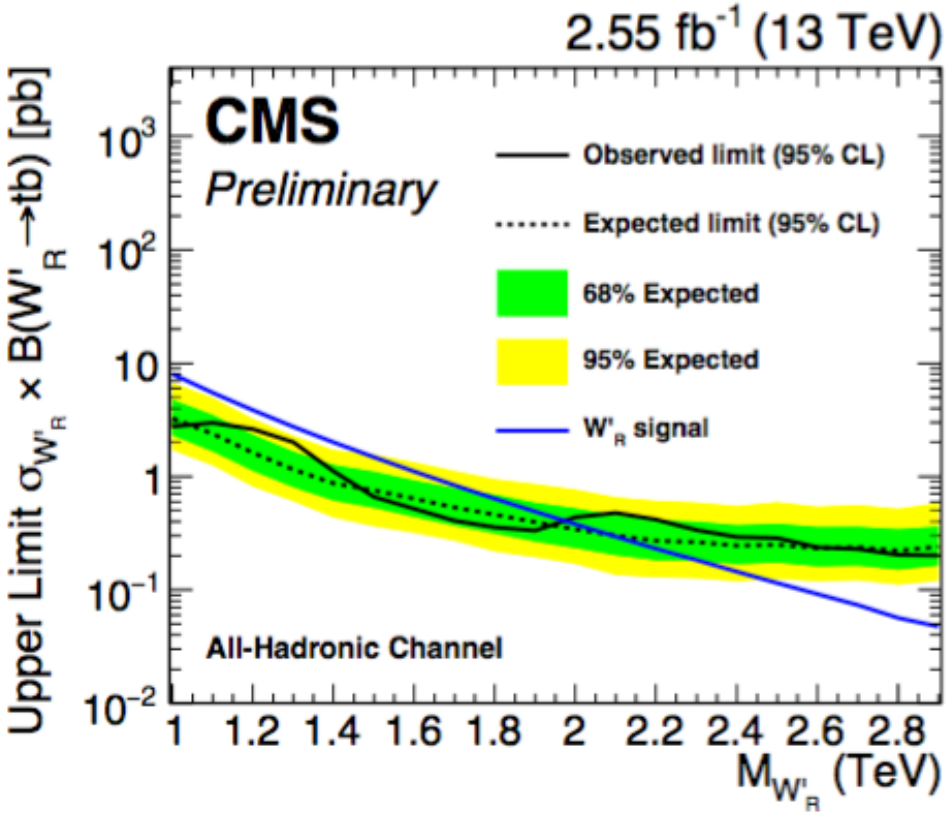
W′ Search Results

B2G-16-009
B2G-16-016

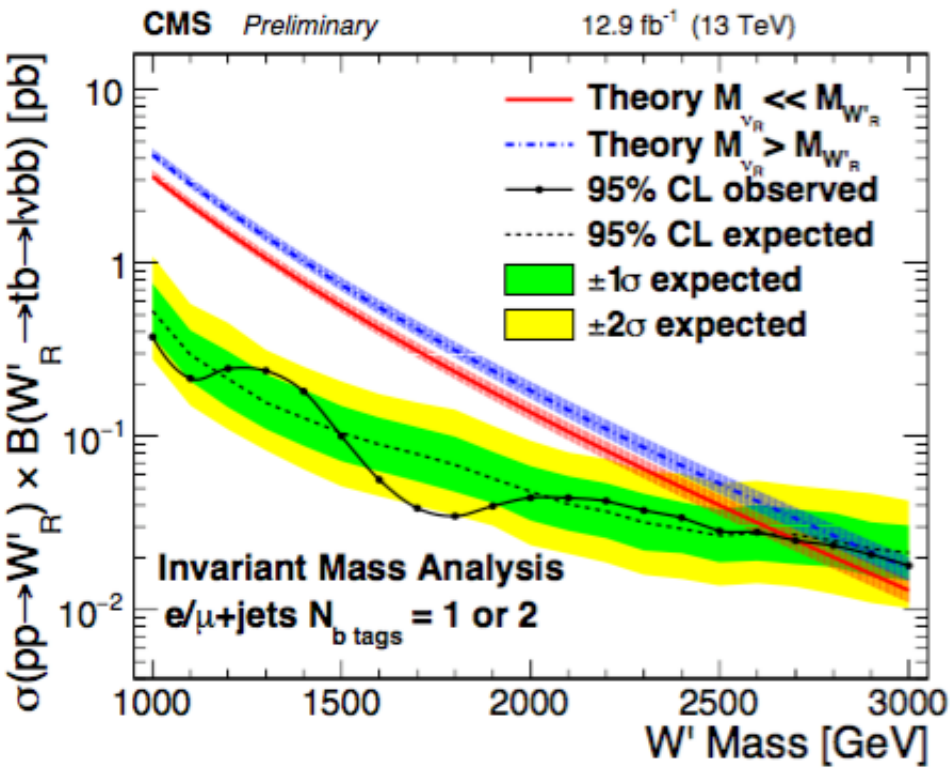
- Mass exclusions up to 2.0 TeV (hadronic) and 2.7 TeV (lepton+jets)

- Combination currently in progress with updated datasets

Hadronic

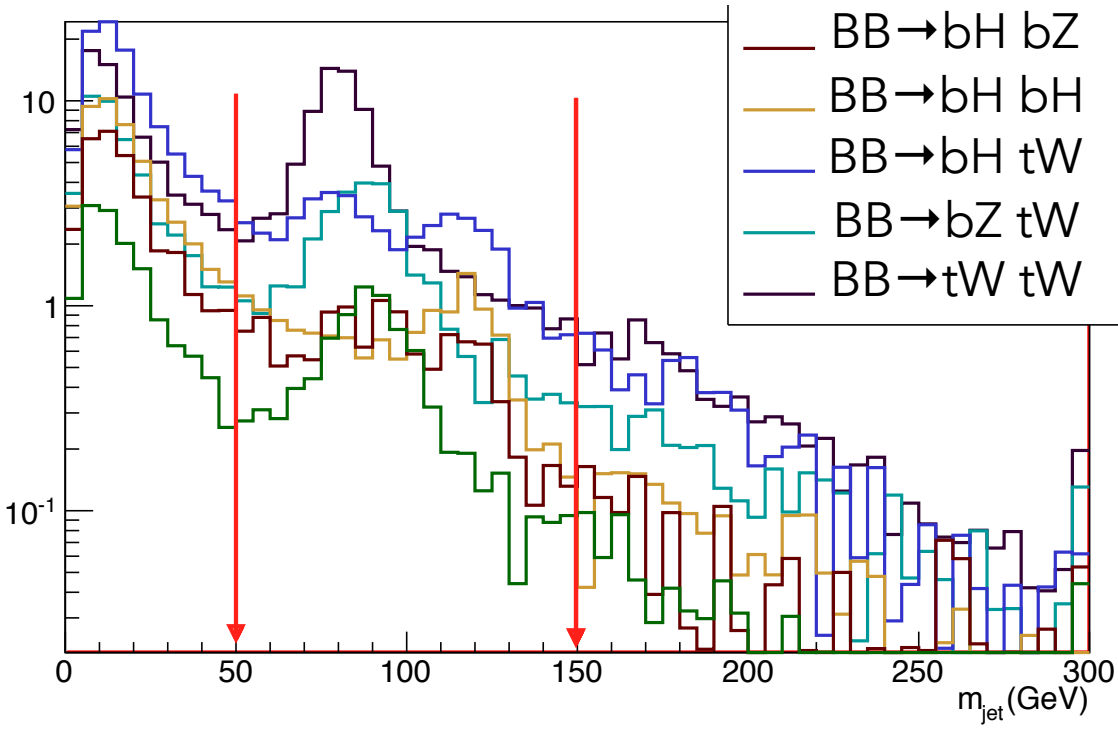


Lepton+Jets



Vector-Like Quarks

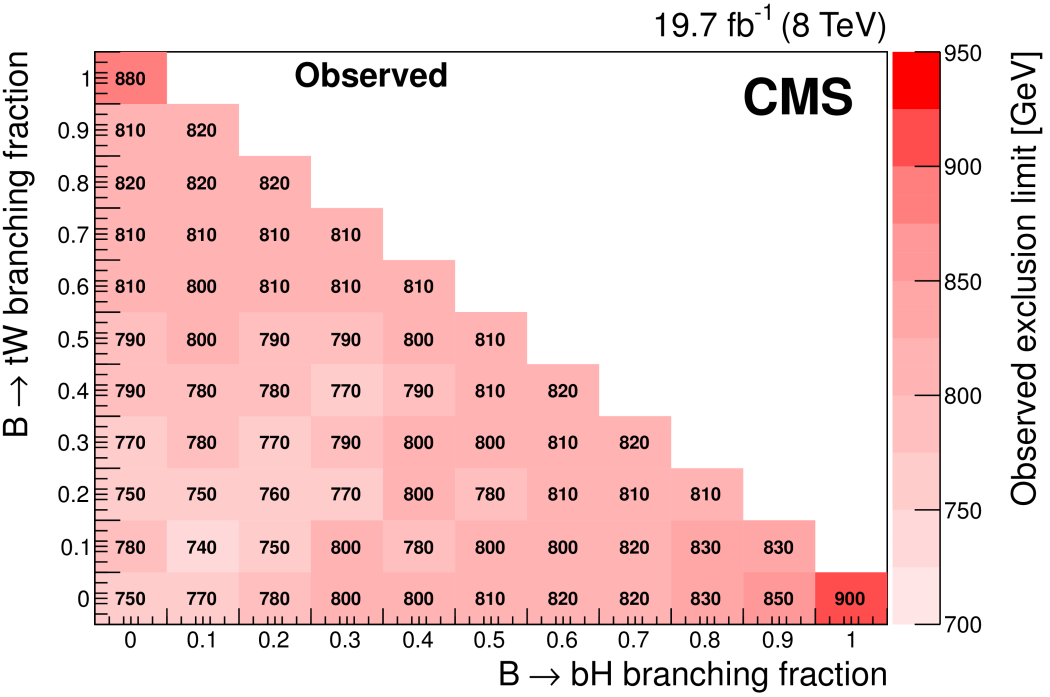
- A special type of 4th generation particle
- After Higgs discovery, constraints have become much more stringent
  - 'Generic' 4th generation t', b' would have enhanced Higgs production cross section by ~5x
- "Vector-like" quarks can escape these constraints and are predicted by many models
  - Triplets under SU(3), L and R components have same quantum numbers
  - Do not obtain mass through Higgs Yukawa coupling
- Different decay modes possible
  - B → tW, bZ, bH
  - T → bW, tZ, tH
- Diverse final states possible when considering pair production
  - Several heavy taggable objects

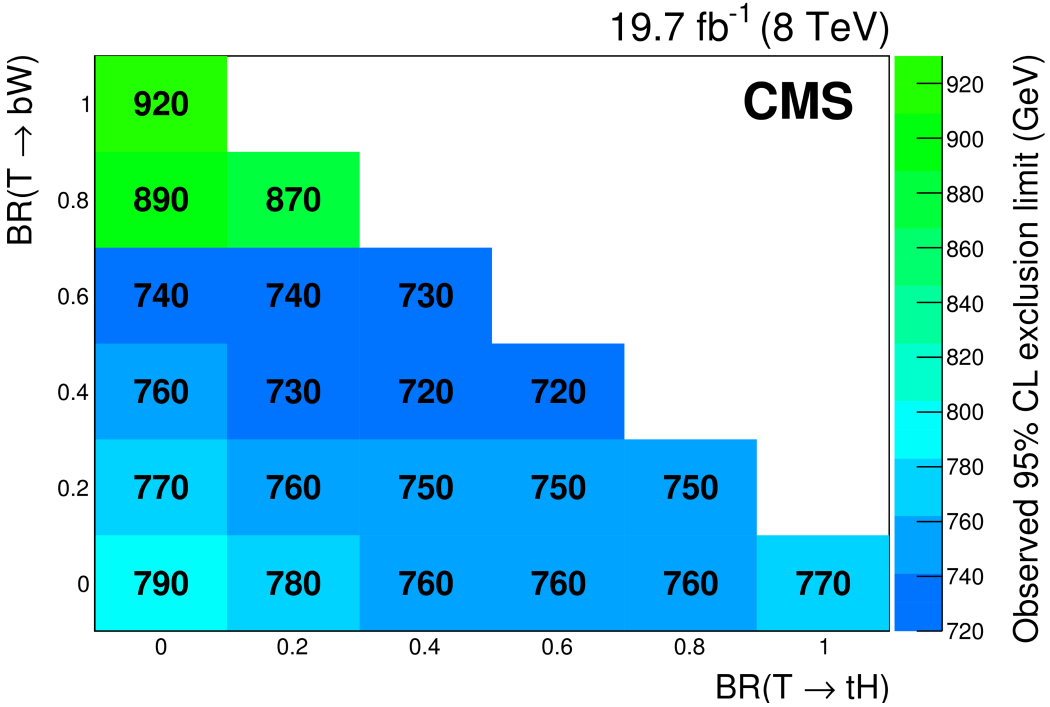


8 TeV Status

B2G-13-005/6

- Combinations of T and B searches published in PRD
  - Exclusions between ~700 and ~900 GeV depending on BR combination
- ~10 individual analyses represented here
  - Reliant on substructure techniques such as W/Z-tagging, top-tagging, H→bb tagging

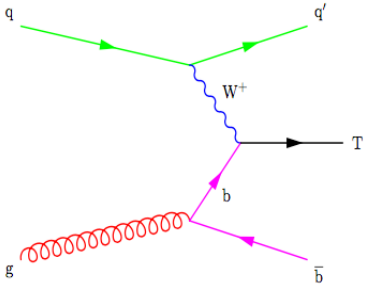


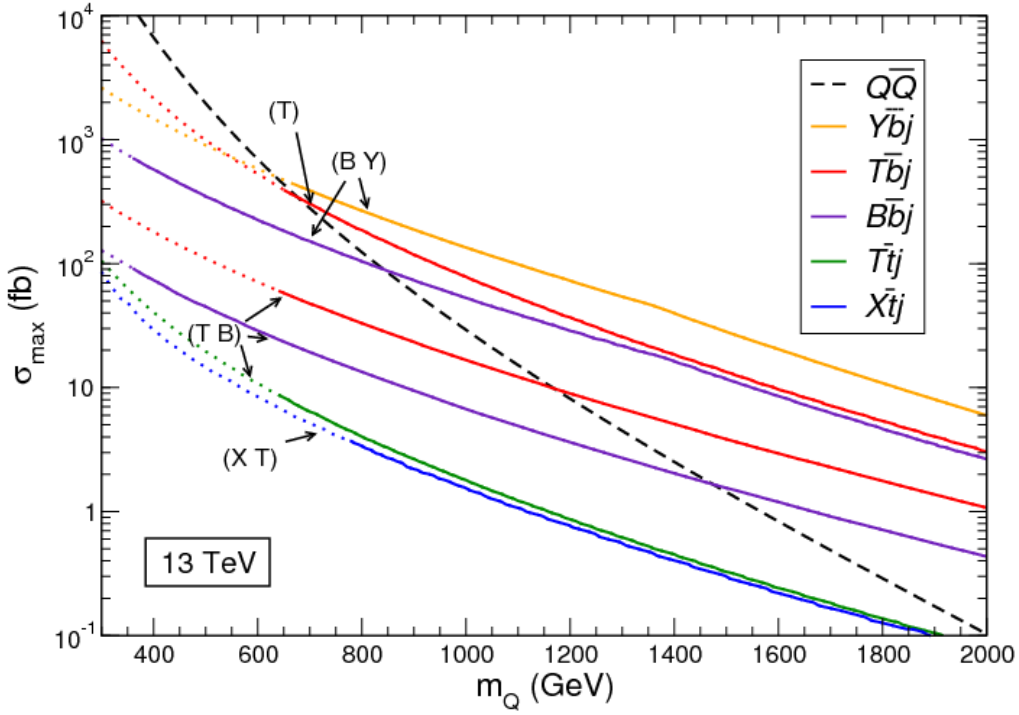


New VLQ Search Results

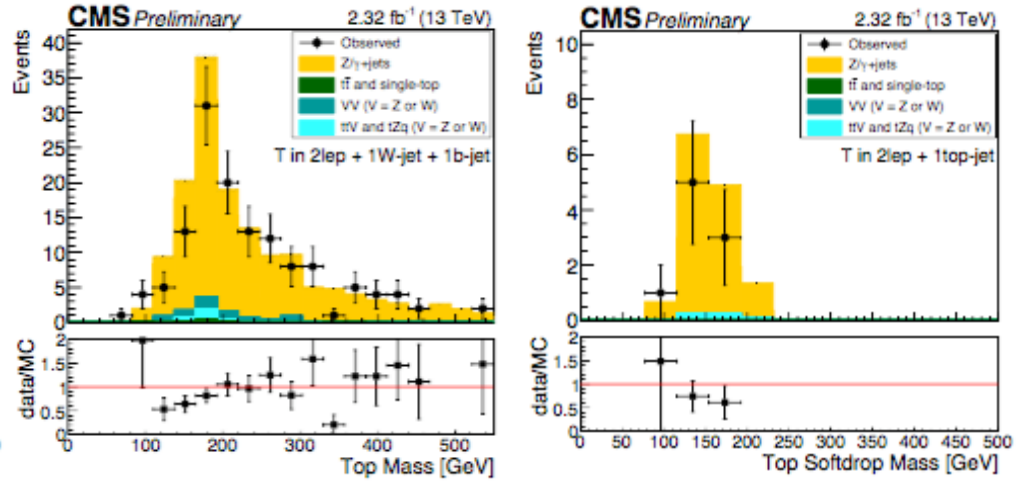
B2G-16-001

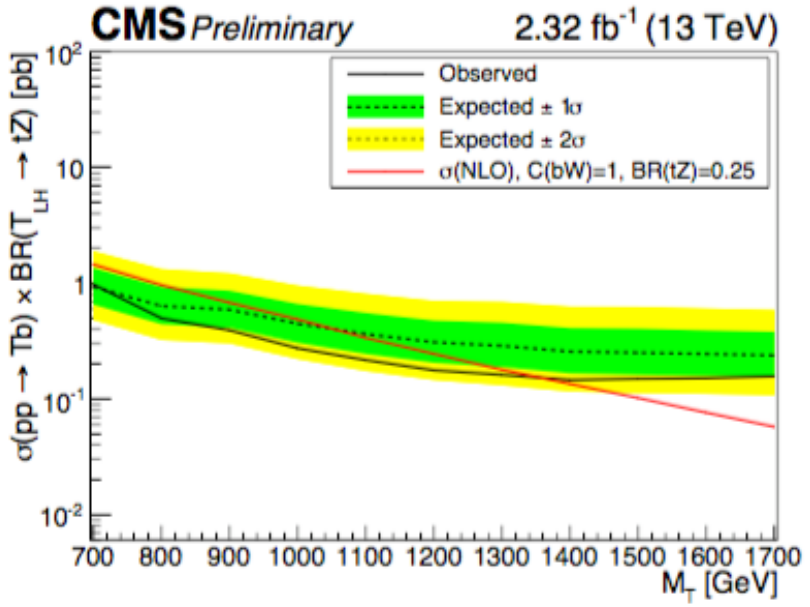
- As we move to 13 TeV, single B or T production becomes dominant for high-mass VLQs
  - Produced in association with forward heavy quark





- One recent analysis: T → tZ in the dilepton channel

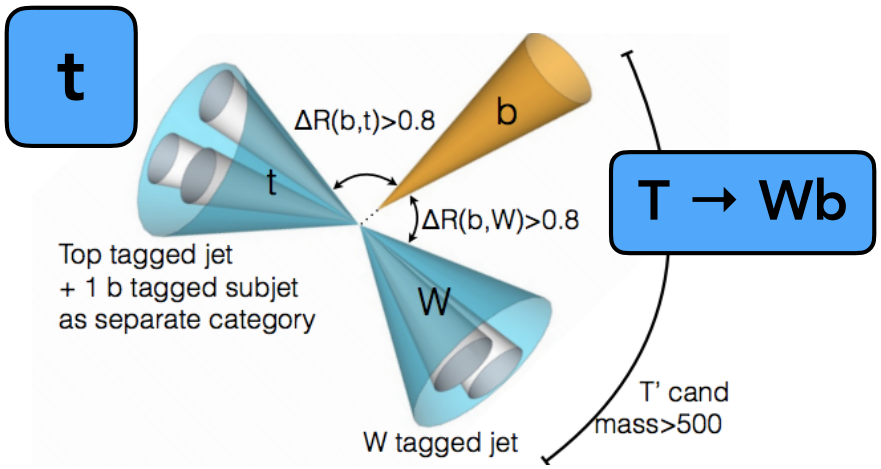




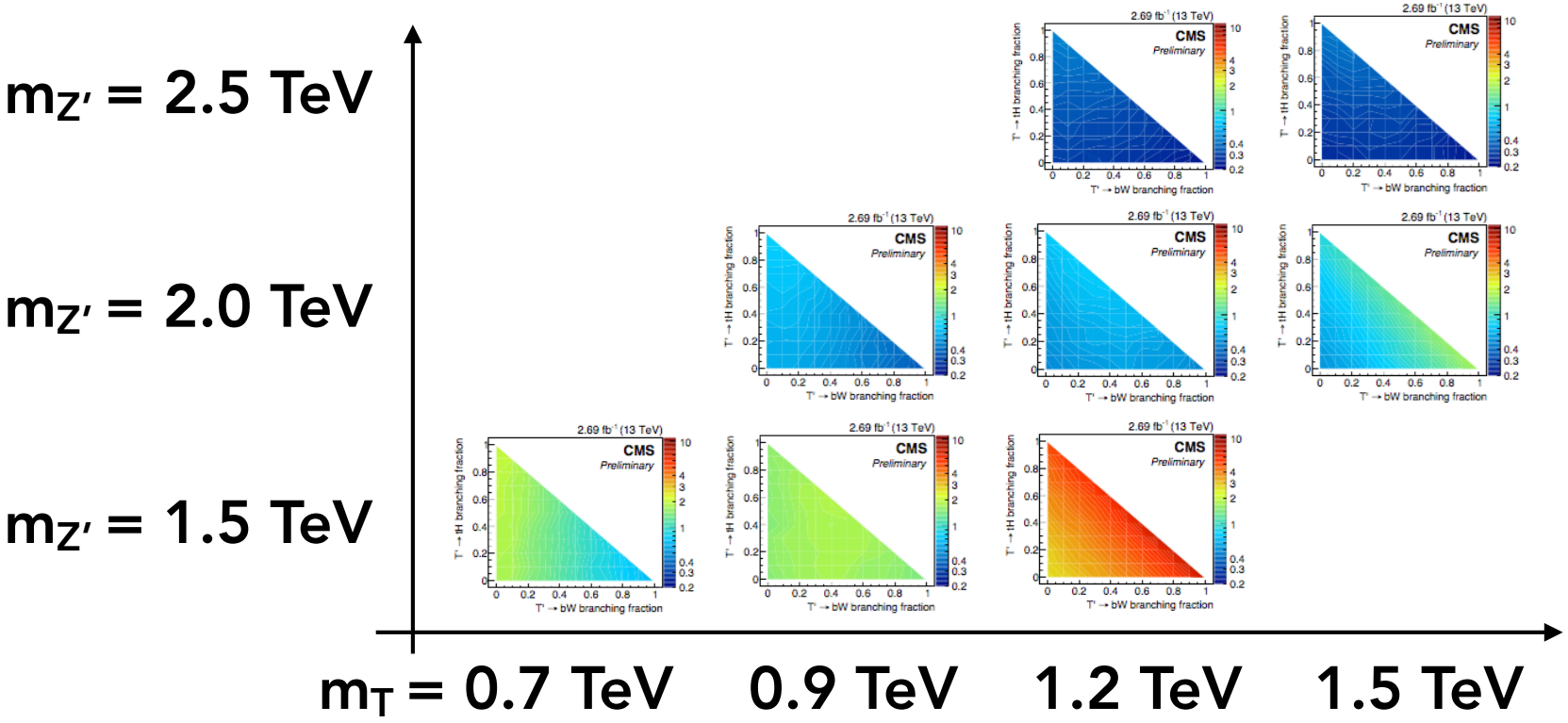
Hybrid Analyses

B2G-16-013

- There are models which allow the new heavy gauge bosons to decay to the VLQs
  - Z′ → tT, bB
  - W′ → bT, tB
  - Masses of Z′, T/B can vary
    - "Triangles within triangles"



- First public result is search for Z′→ tT

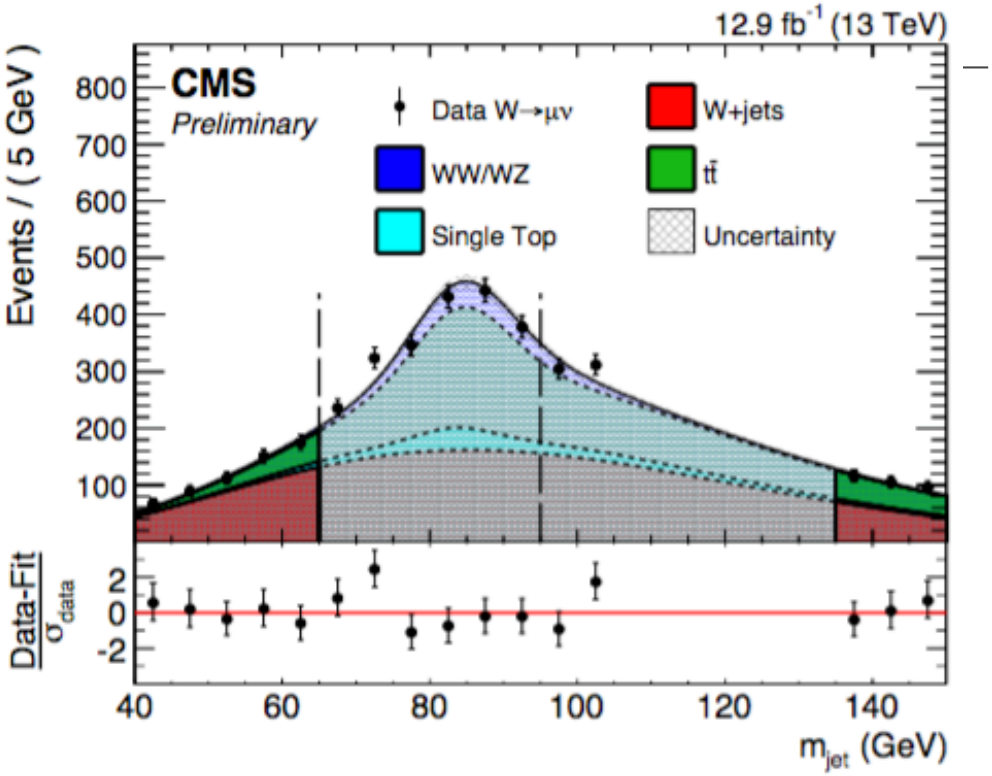


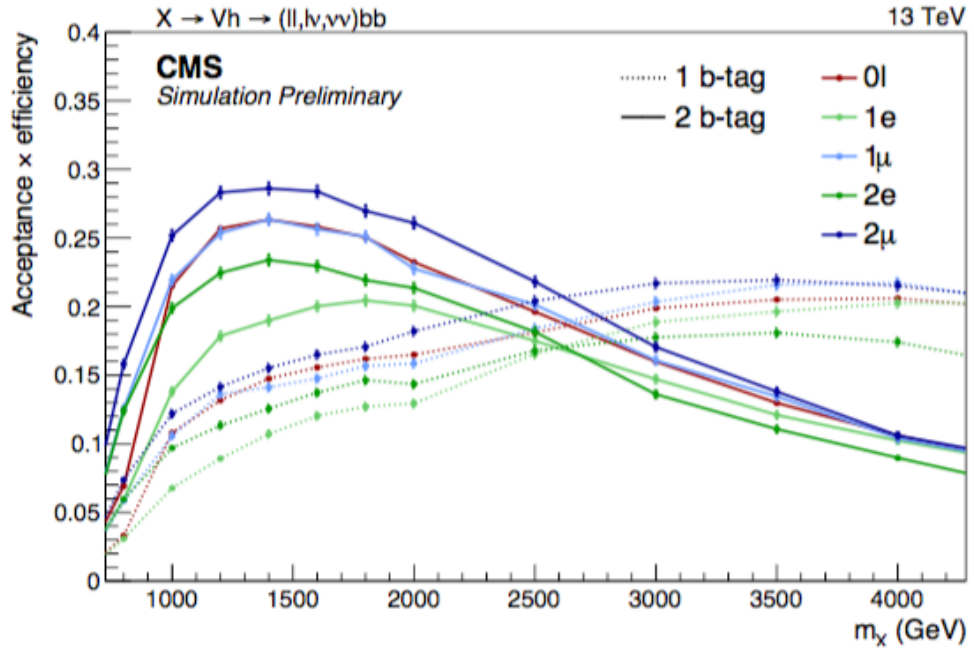
Diboson Resonances

- B2G group also includes searches for heavy resonances decaying to dibosons (WW/WZ/ZZ/HH)

- Some channels use groomed jet mass to identify merged W/Z jets

- Boosted Higgs(→bb) bosons better identified at high-mass with a single b-tag due to subjet merging!

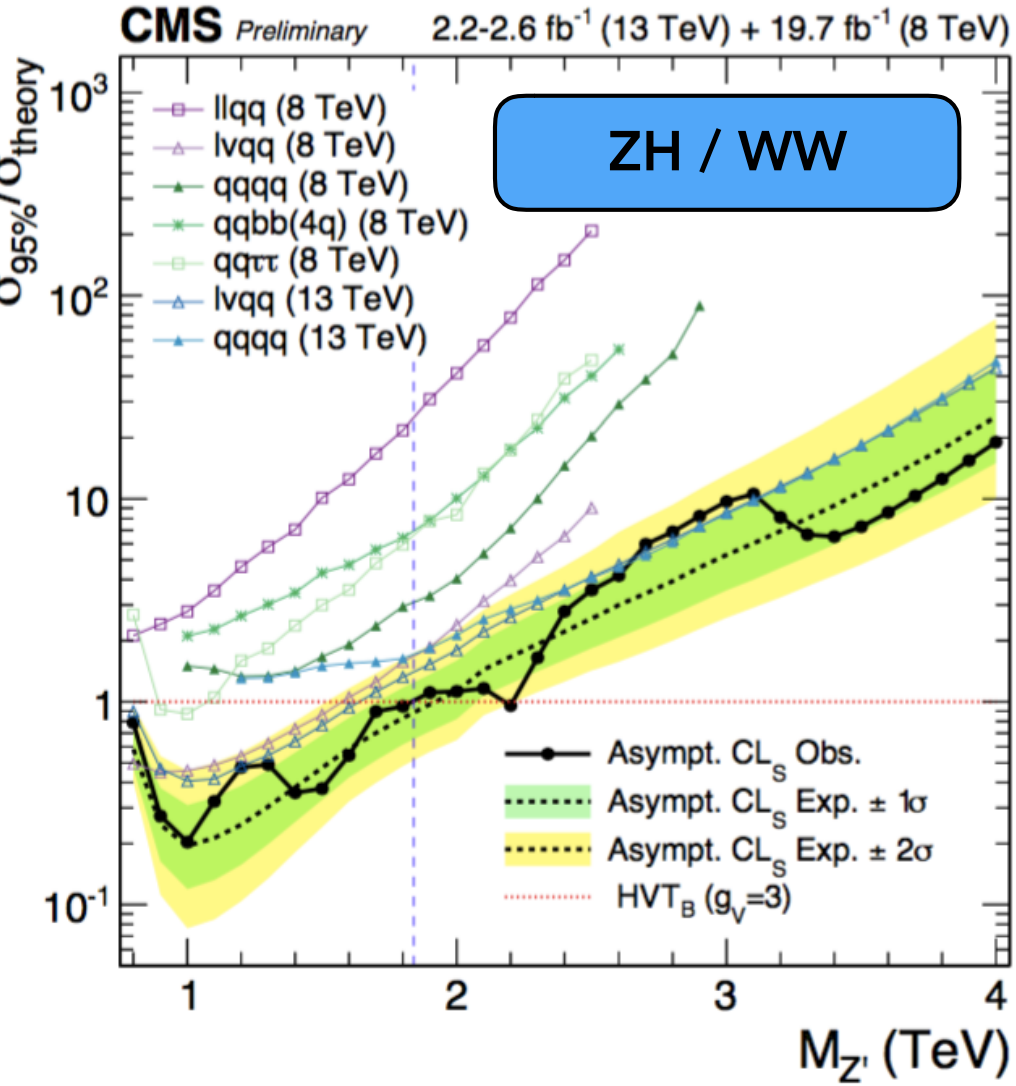


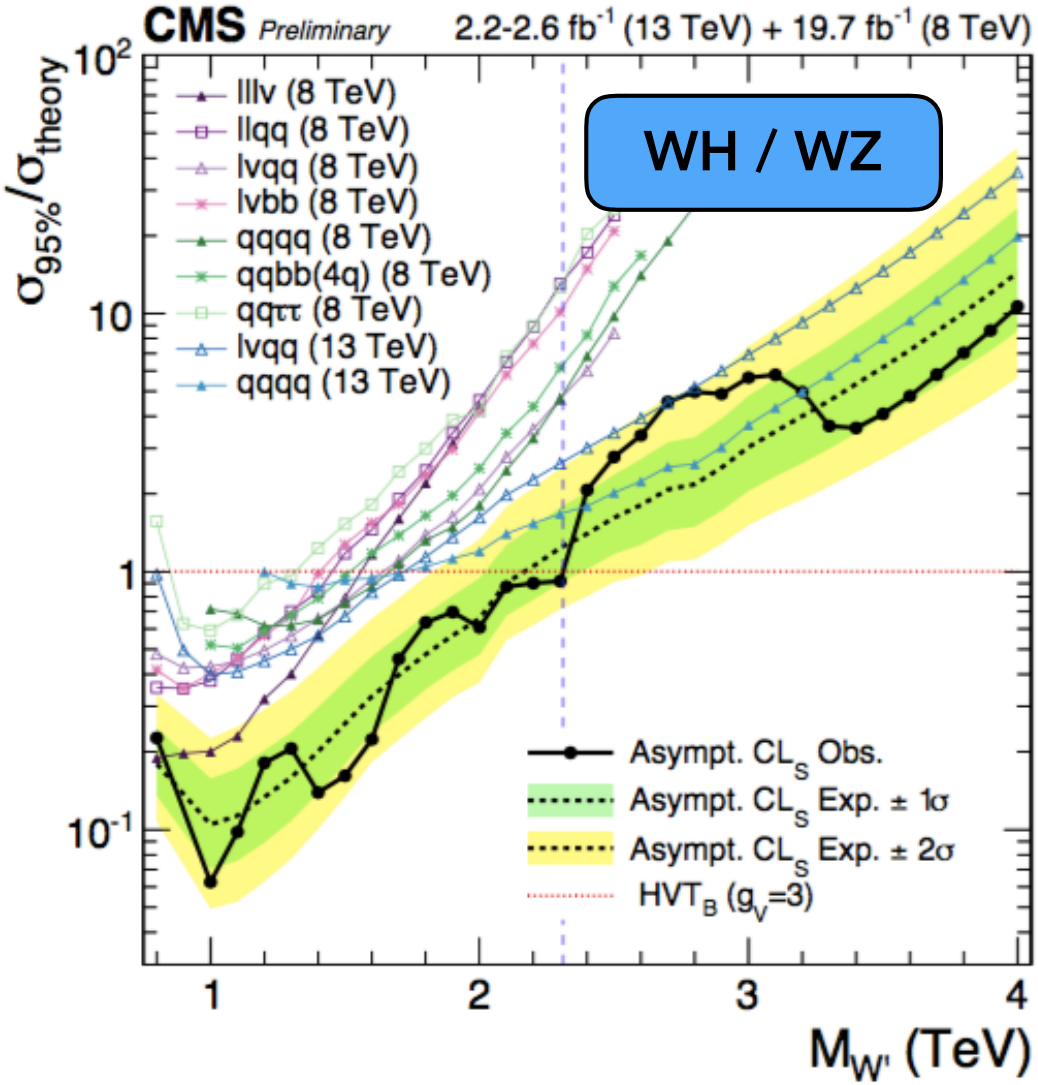


Diboson Combination

B2G-16-007

- Some new results from ICHEP not yet included here

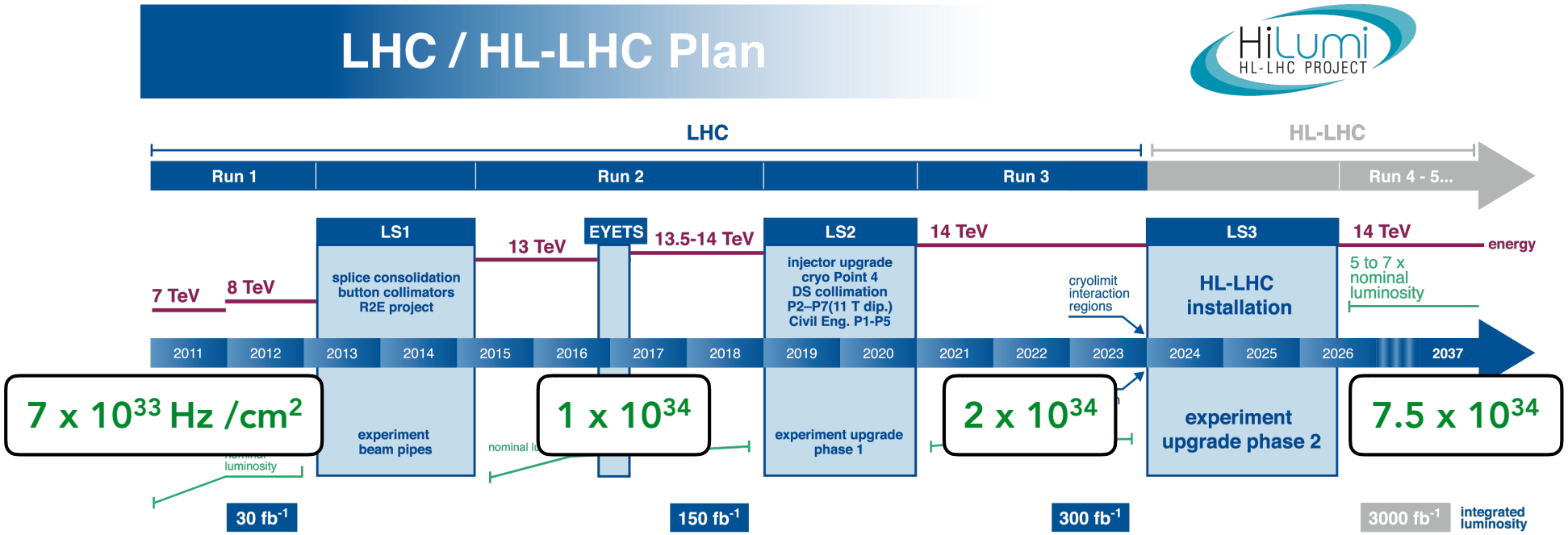




Looking to the Future

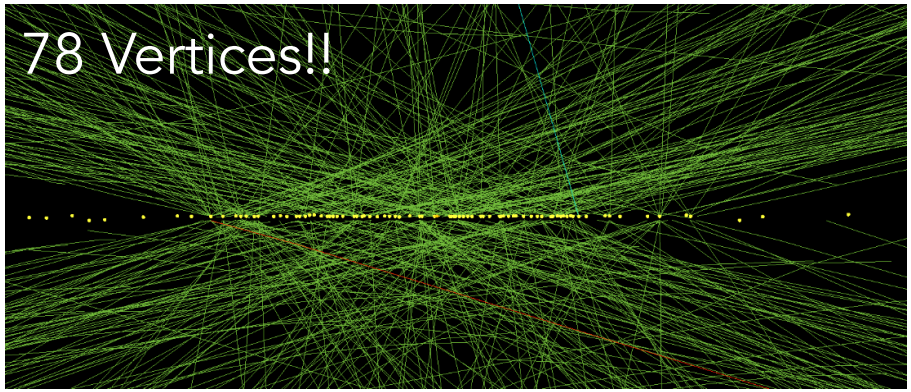
Where to go from here?

- Many search results have dramatically increased reach compared to 13 TeV
  - Probing very high masses
- What's going to happen next?
  - More data, of course
  - Higher instantaneous luminosities
  - Higher pileup
  - Higher energy (?)
- We are already testing ideas to ensure that we can remain efficient in identifying these objects with jet substructure



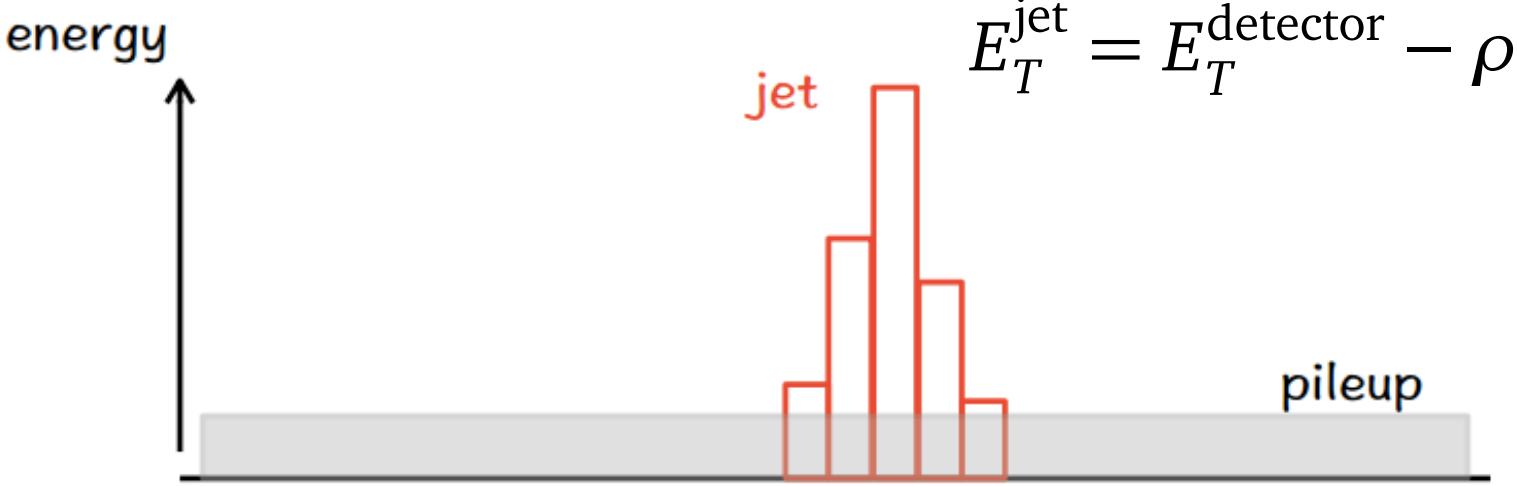
Pileup

- In the future running, we can expect up to 200 pileup events in each bunch crossing of the LHC!



- Can easily determine which charged particles are from pileup vertices
  - Reconstructed tracks point back
  - "Charged Hadron Subtraction" (CHS)
  - What about neutrals??

- Assume uniform energy distribution of pileup energy ρ

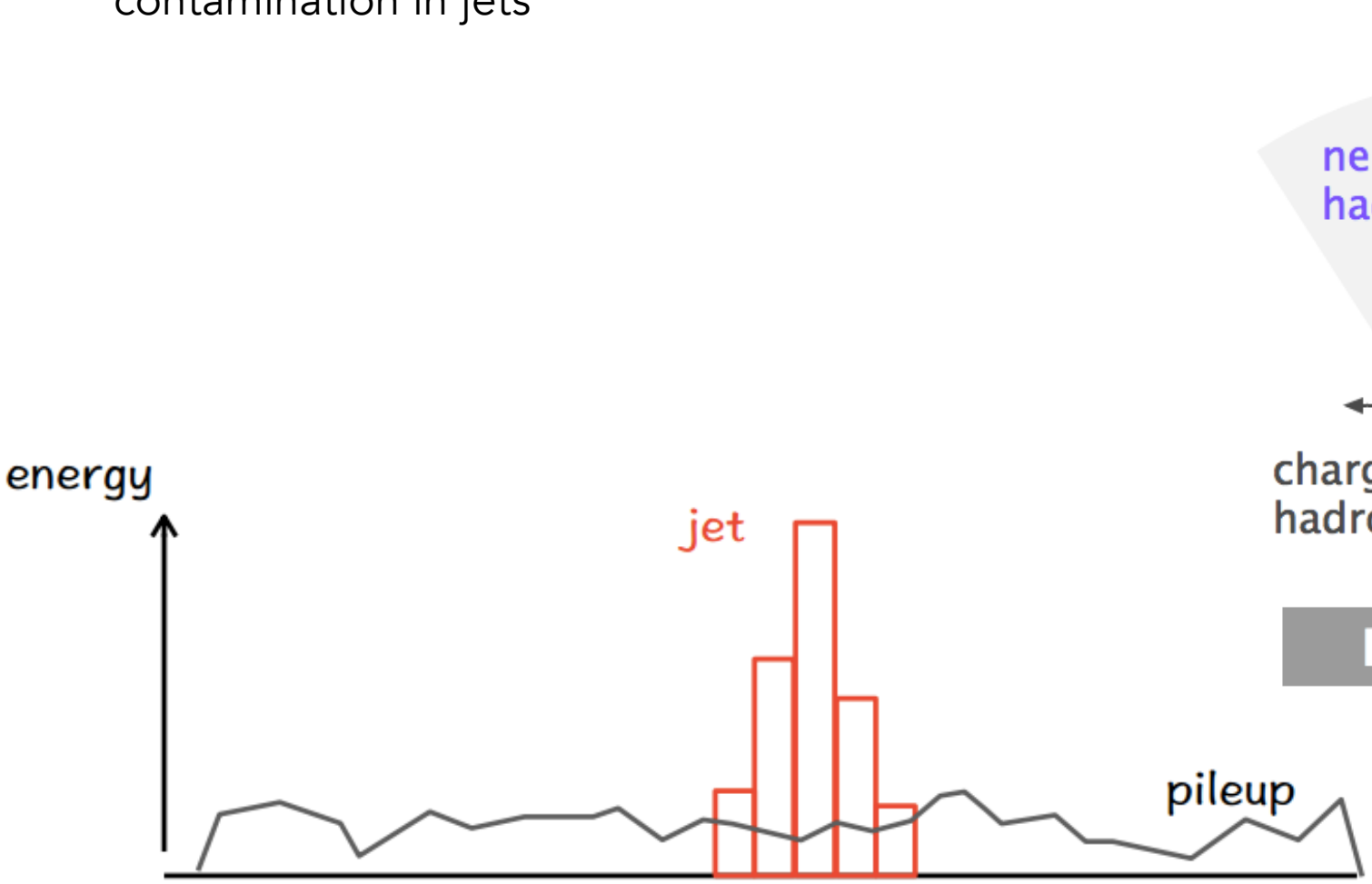


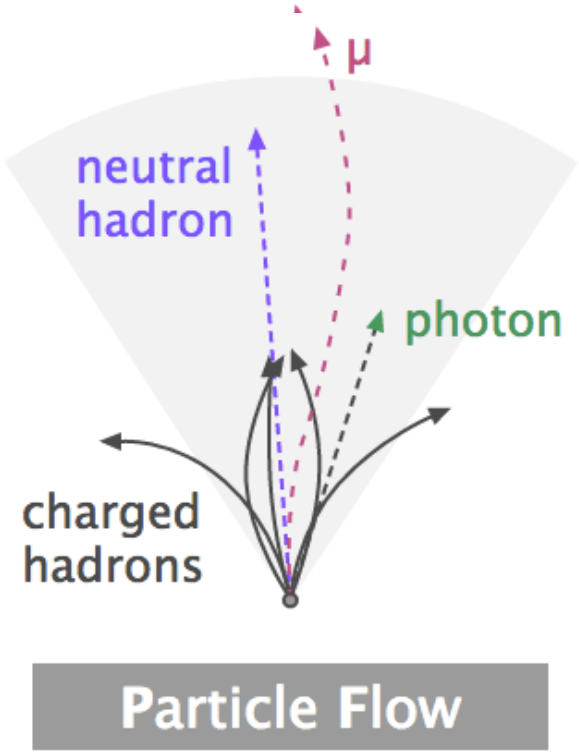
$$E_T^{\text{jet}} = E_T^{\text{detector}} - \rho \cdot A_{\text{jet}}$$

PUPPI arXiv:1407.6013

- The **p**ile**u**p **p**er-**p**article **i**dentification algorithm is designed to use the particle information in the event to improve mitigation of pileup contamination in jets

- We can use tracking information to decide probabilistically if a particle originated from pileup or not



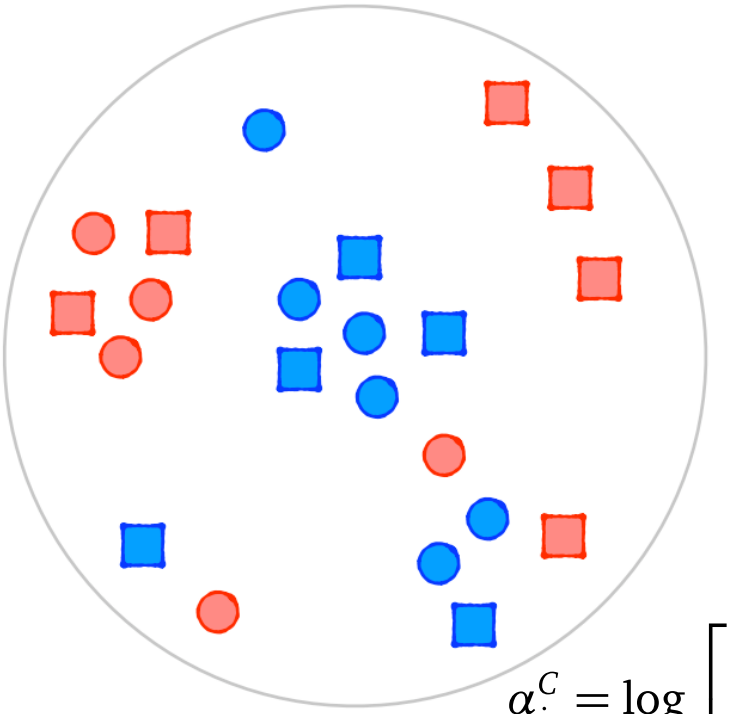


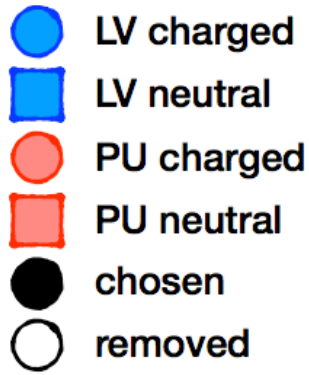
PUPPI

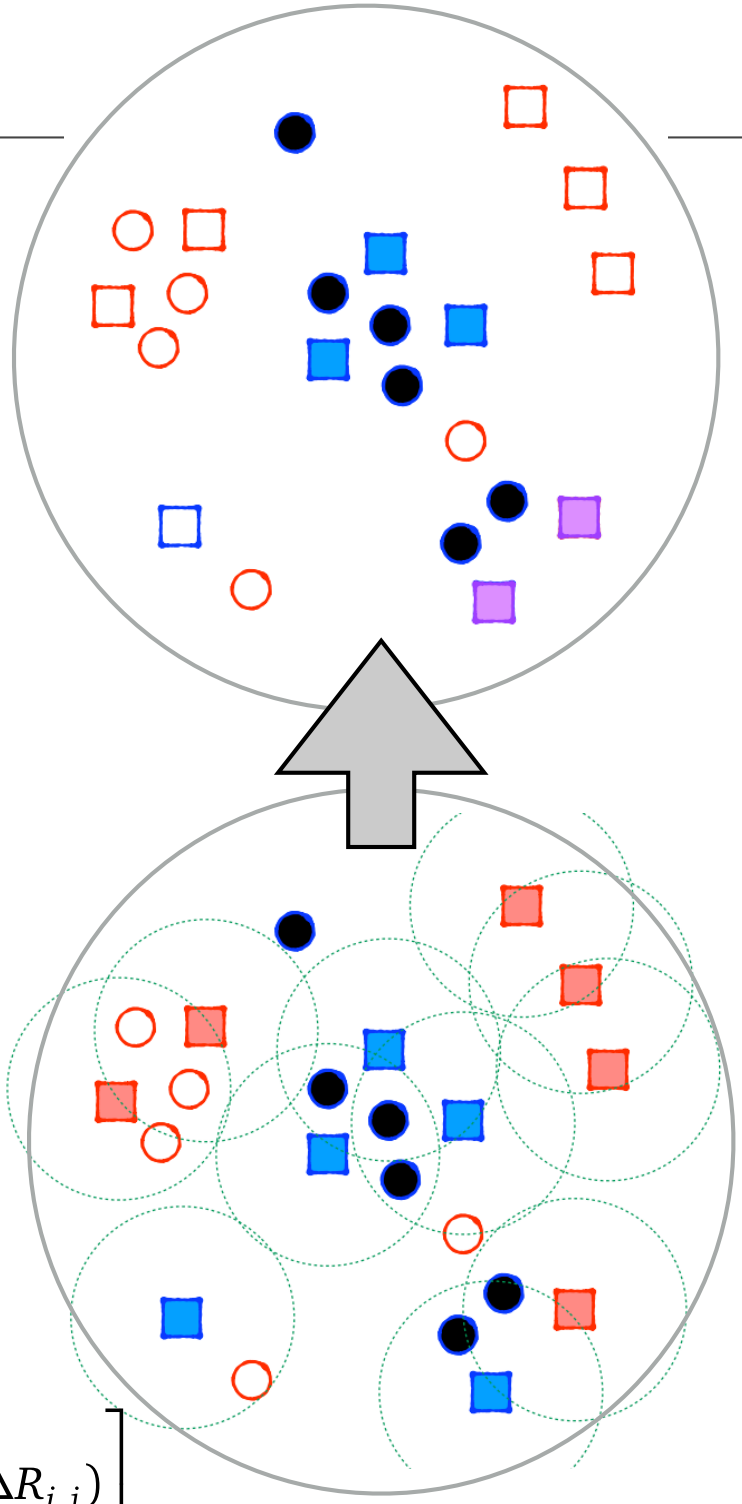
arXiv:1407.6013

- Easy to use tracking information to find which particles originated from the primary vertex

- If neutral particles are mostly surrounded by charged particles from pileup activity, they are probably pileup as well







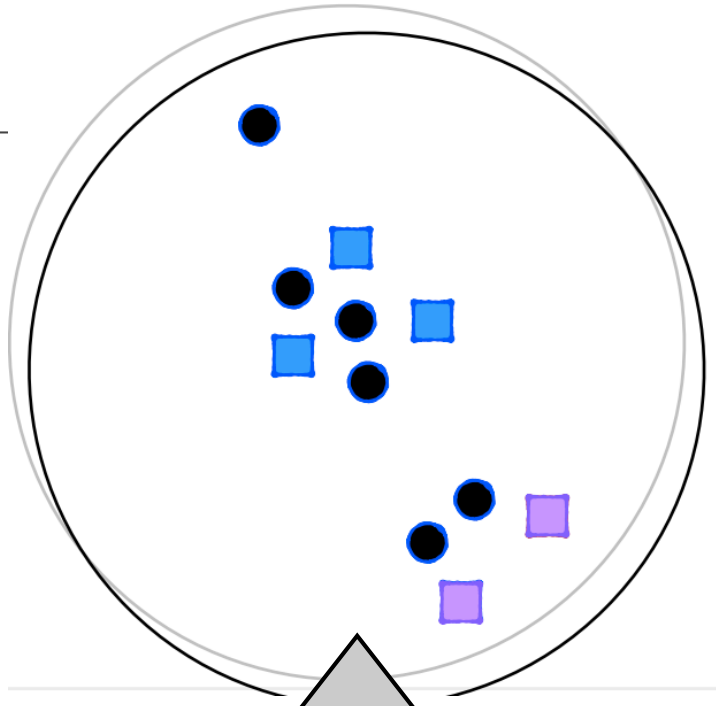
$$\alpha_i^C = \log\left[\sum_{j\in \mathrm{Ch,LV}} \frac{p_{T,j}}{\Delta R_{i,j}} \Theta(R_0 - \Delta R_{i,j})\right]$$

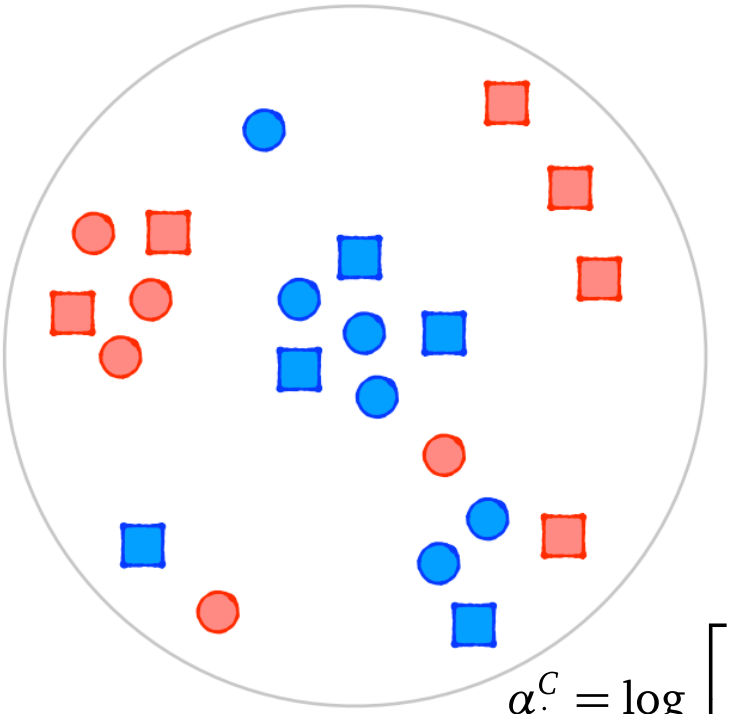
PUPPI

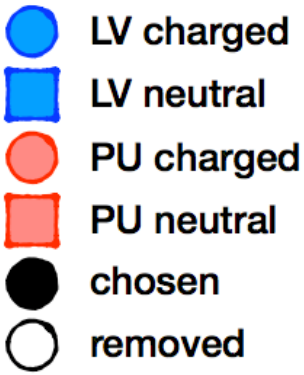
arXiv:1407.6013

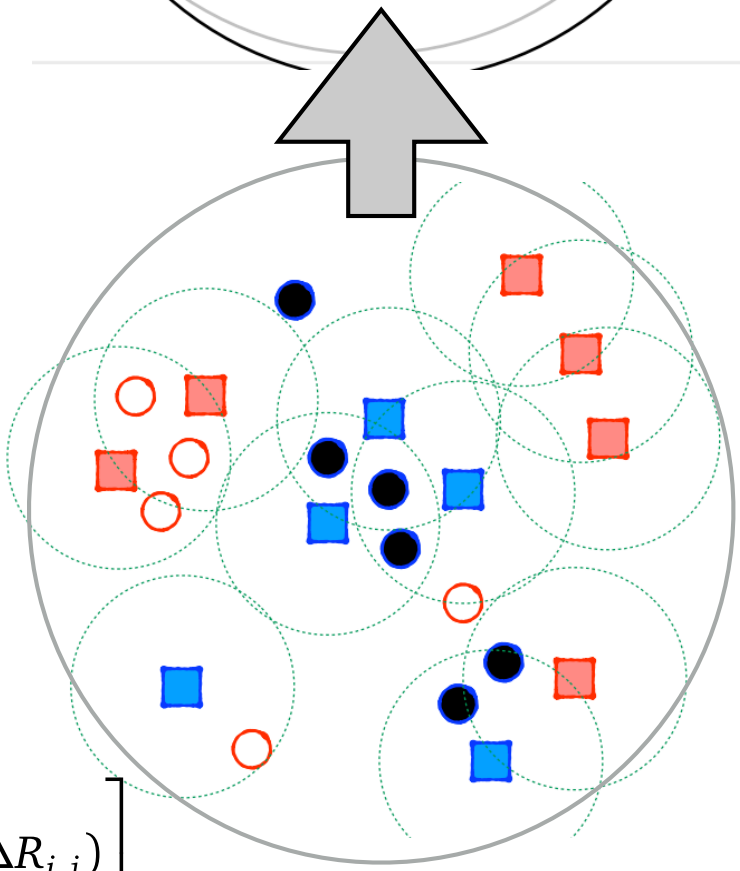
- Easy to use tracking information to find which particles originated from the primary vertex

- If neutral particles are mostly surrounded by charged particles from pileup activity, they are probably pileup as well





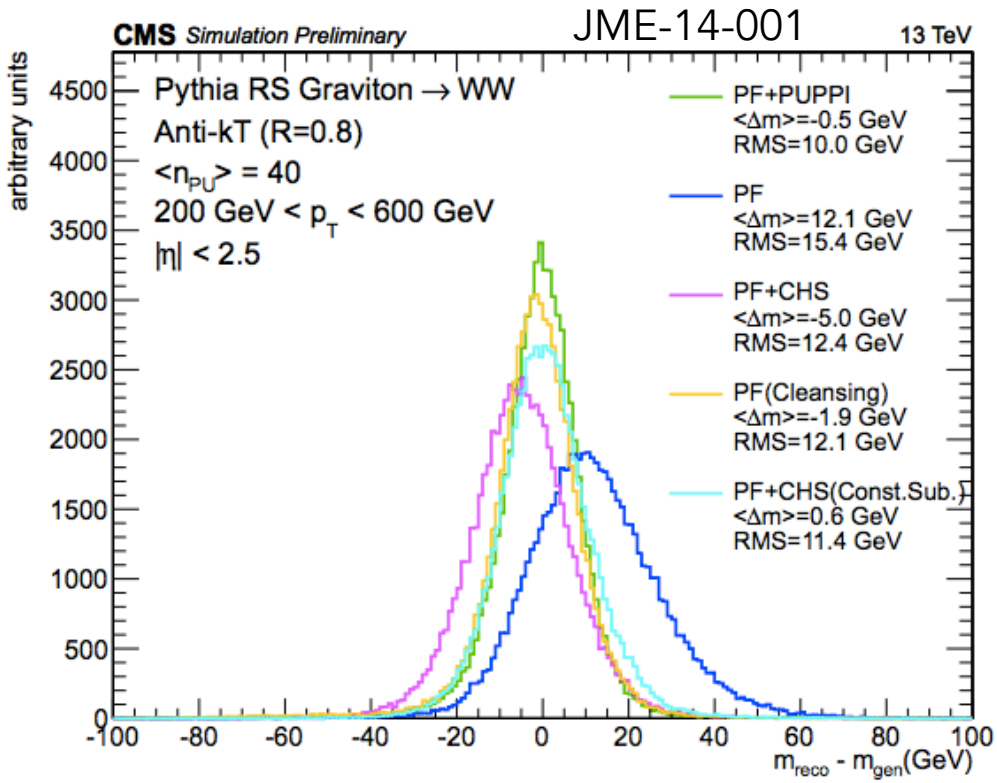


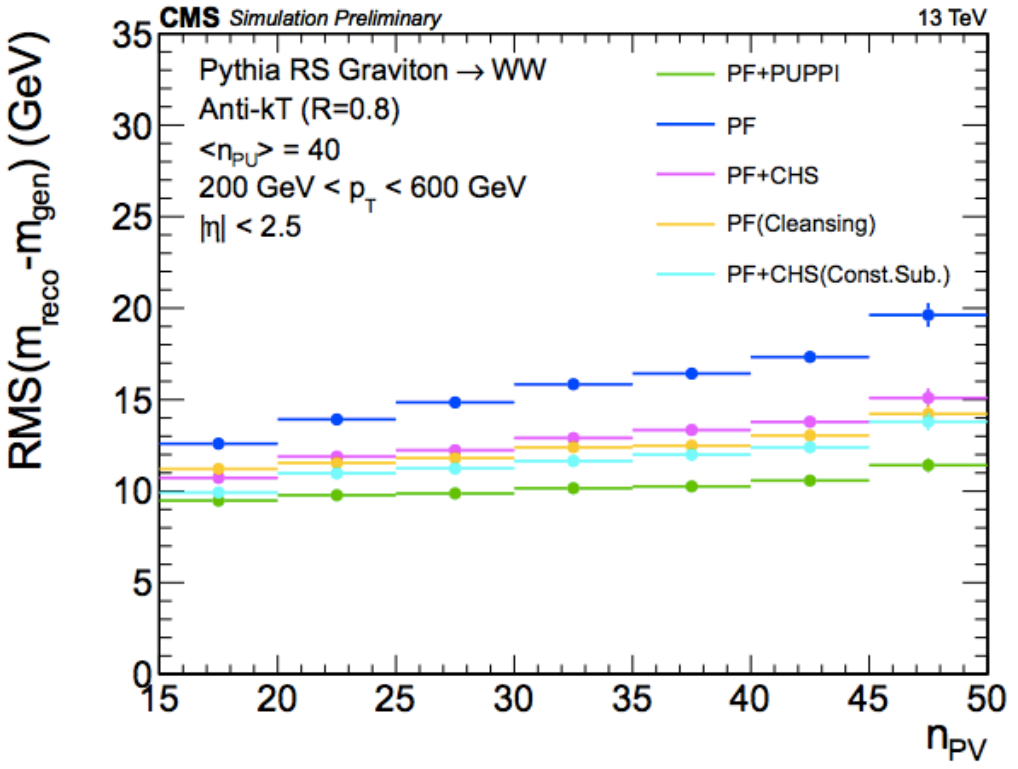


$$\alpha_i^C = \log\left[\sum_{j\in \mathrm{Ch,LV}} \frac{p_{T,j}}{\Delta R_{i,j}} \Theta(R_0 - \Delta R_{i,j})\right]$$

PUPPI Performance

- Improves jet mass resolution significantly
  - Stable as a function of pileup vertices
- Being applied to latest round of CMS searches now
- Also useful for quantities such as MET, lepton isolation

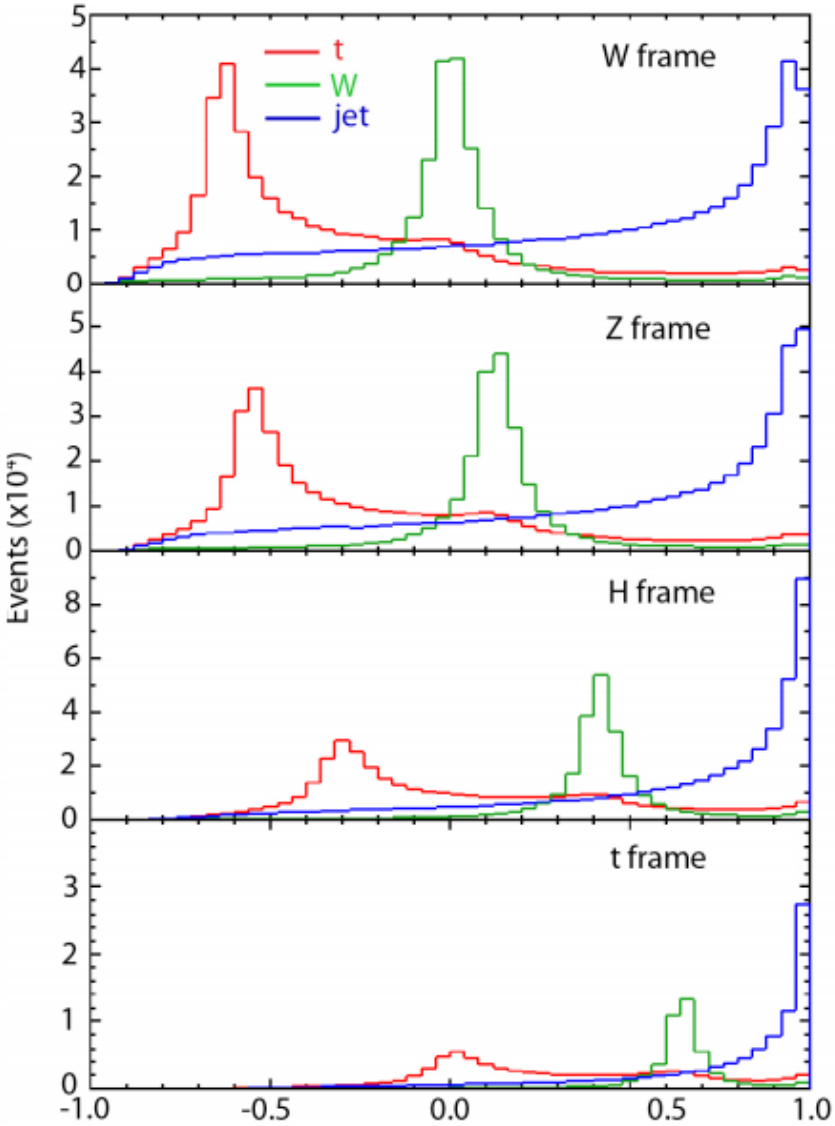


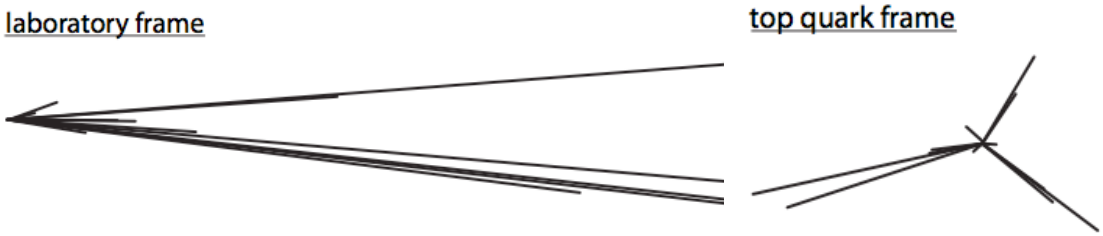


New Substructure Algorithms

arXiv:1606.06859

- Many new substructure algorithms also on the market
- A new method is being explored called the Boosted Event Shape Tagger ("BEST")
- What if we "guess" the particle origin of a jet
  - Extra information can be obtained with different guesses, i.e. W/Z/H/t
- Use boosted reference frames for each hypothesis, compare jet constituent distributions and quantities in each



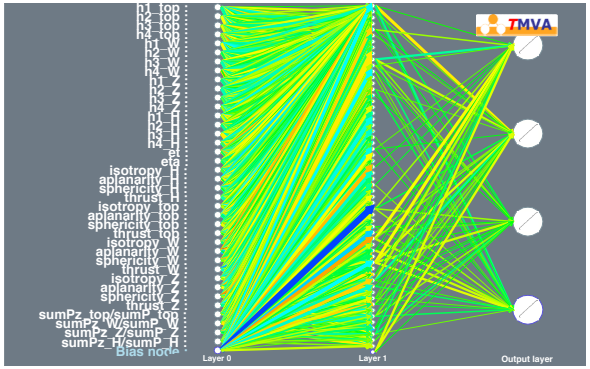


$$A_L = \frac{\sum_{jet} p_z^{jet}}{\sum_{jet} p^{jet}}$$

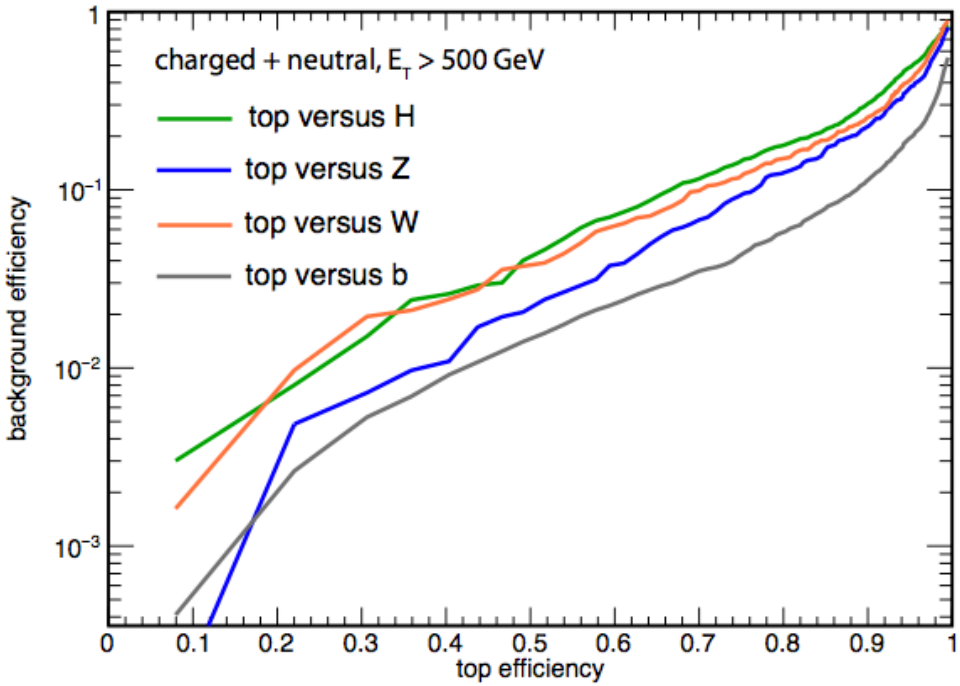
Boosted Event Shape Tagger

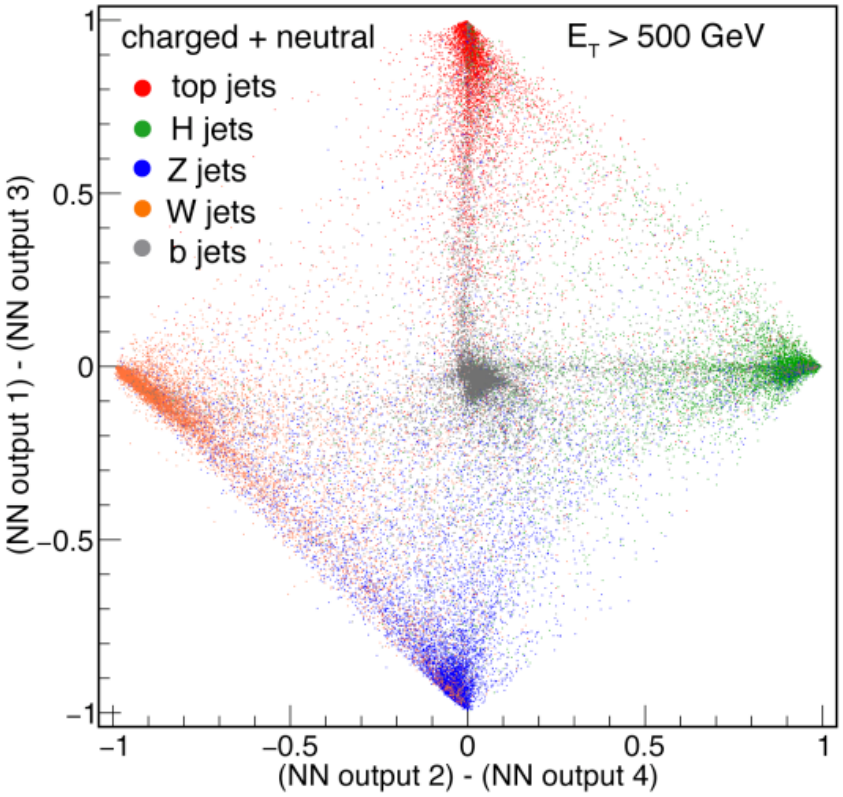
arXiv:1606.06859

- Use these angular quantities and distributions to train a NN to separate the different particle species
  - One output for each particle species



- Discrimination can be obtained for each of the particle species
  - ~70% signal efficiency with ~95% background rejection





- Currently being optimized with CMS data for use in analyses!

Conclusions

Conclusions

- CMS has a broad program of searches using hadronic final states that take advantage of jet substructure algorithms
  - Additional signal acceptance can be added in hadronic channels
- These channels will be critical to cross-check the leptonic channels in the case of a hint of signal events
  - Orthogonal channel, orthogonal methods
- We continue to work on optimization of algorithms for the changing running conditions
  - Will be extremely useful to maintain performance in the future!

- **Thank you for your attention!**
  - **Comments, questions?**