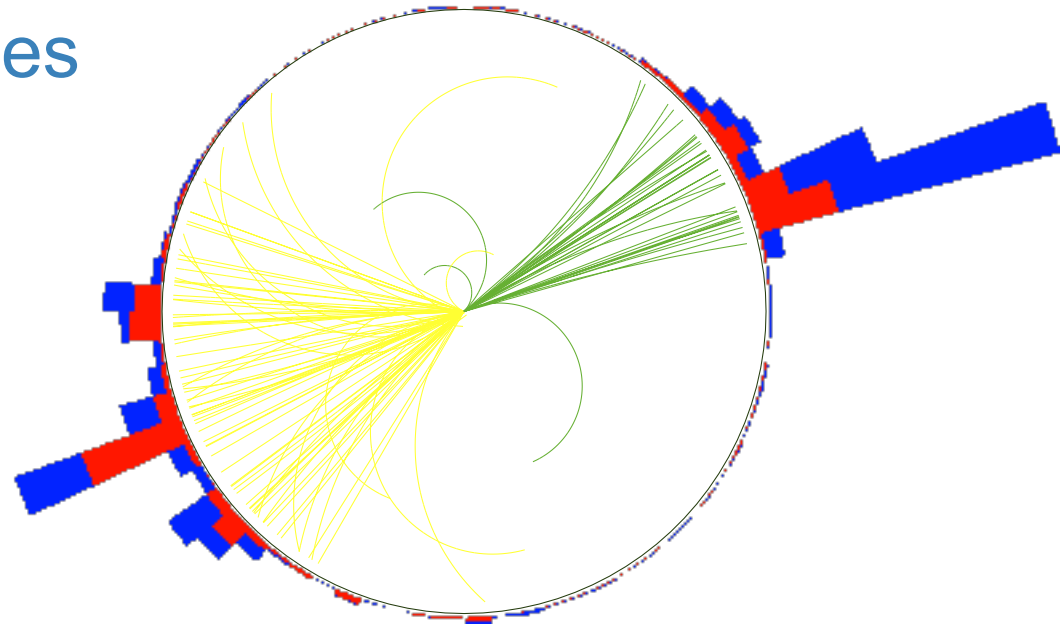
The Search for High-Mass Top Quark Resonances at CMS

Christine McLean, UC Davis

NorCal HEP-EXchange
December 2, 2017

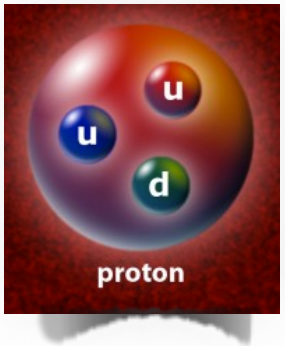
Outline

- The LHC and CMS
  - Particle detection
  - Jet reconstruction
- Z' to $t\bar{t}$, all-hadronic channel
  - Motivation
  - Analysis techniques
  - Results
- Conclusions

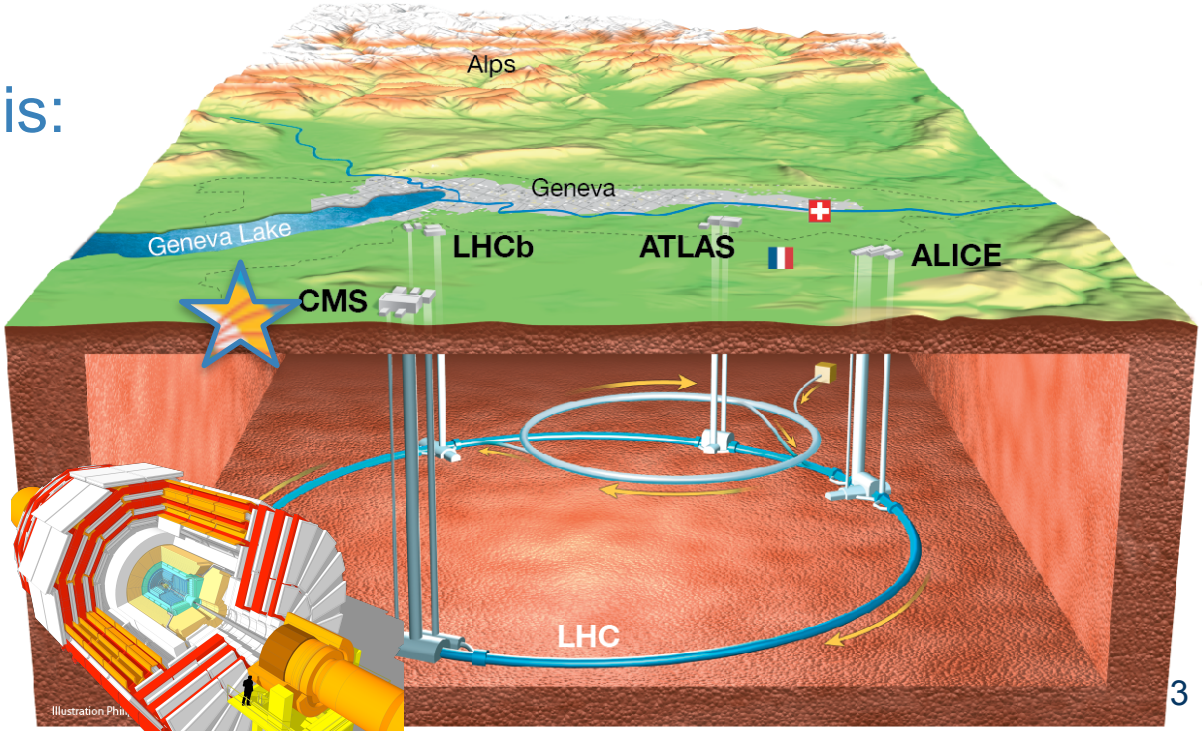


Large Hadron Collider

- Large Hadron Collider (LHC)
  - Circular proton-proton collider, 27 km around
    - Proton bunches collide every 25 ns
    - Two beams, each with E = 6.5 TeV
      - Collision energy: 13 TeV



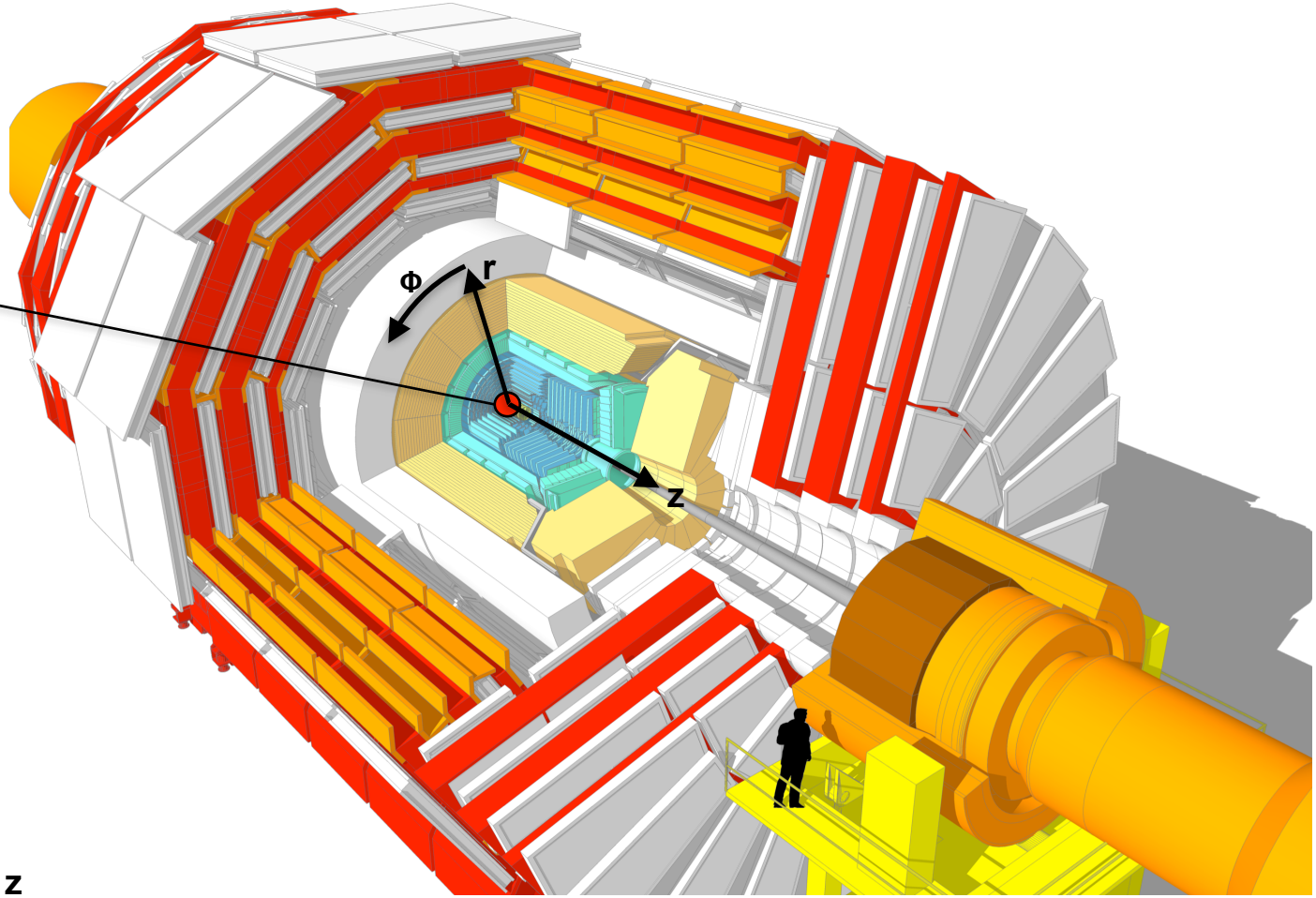
- Compact Muon Solenoid (CMS): one of the four main LHC experiments
  - My thesis analysis: uses 36 fb^{-1} of 2016 data

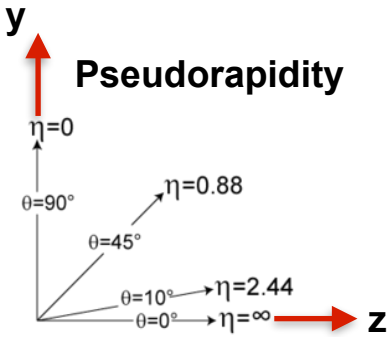


Compact Muon Solenoid

- Multipurpose particle detector
  - Four main subsystems: tracker, ECAL, HCAL, **muon detectors**

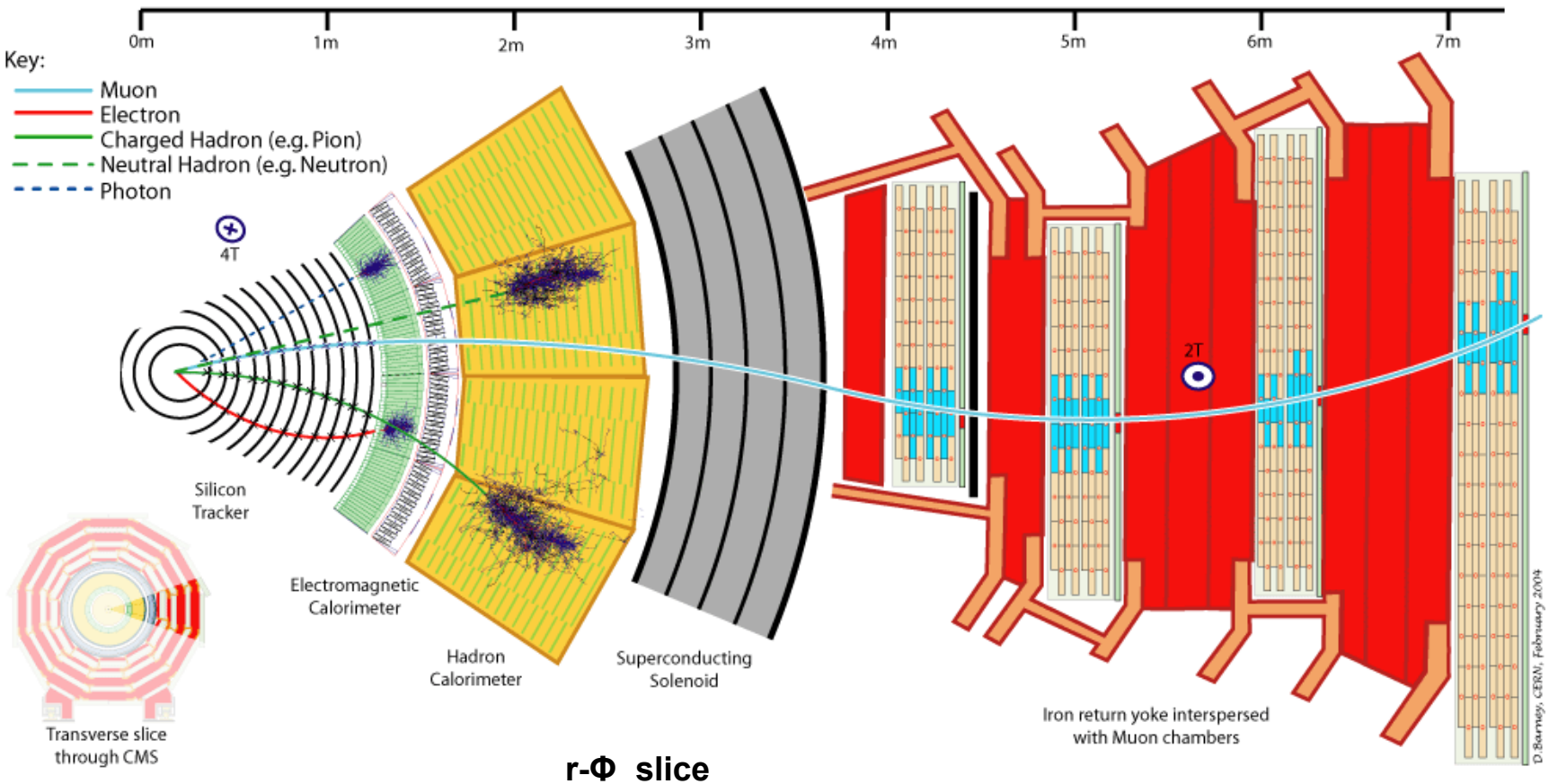
Interaction Point (IP): proton-proton collision point, (0,0)





Particle Detection in CMS

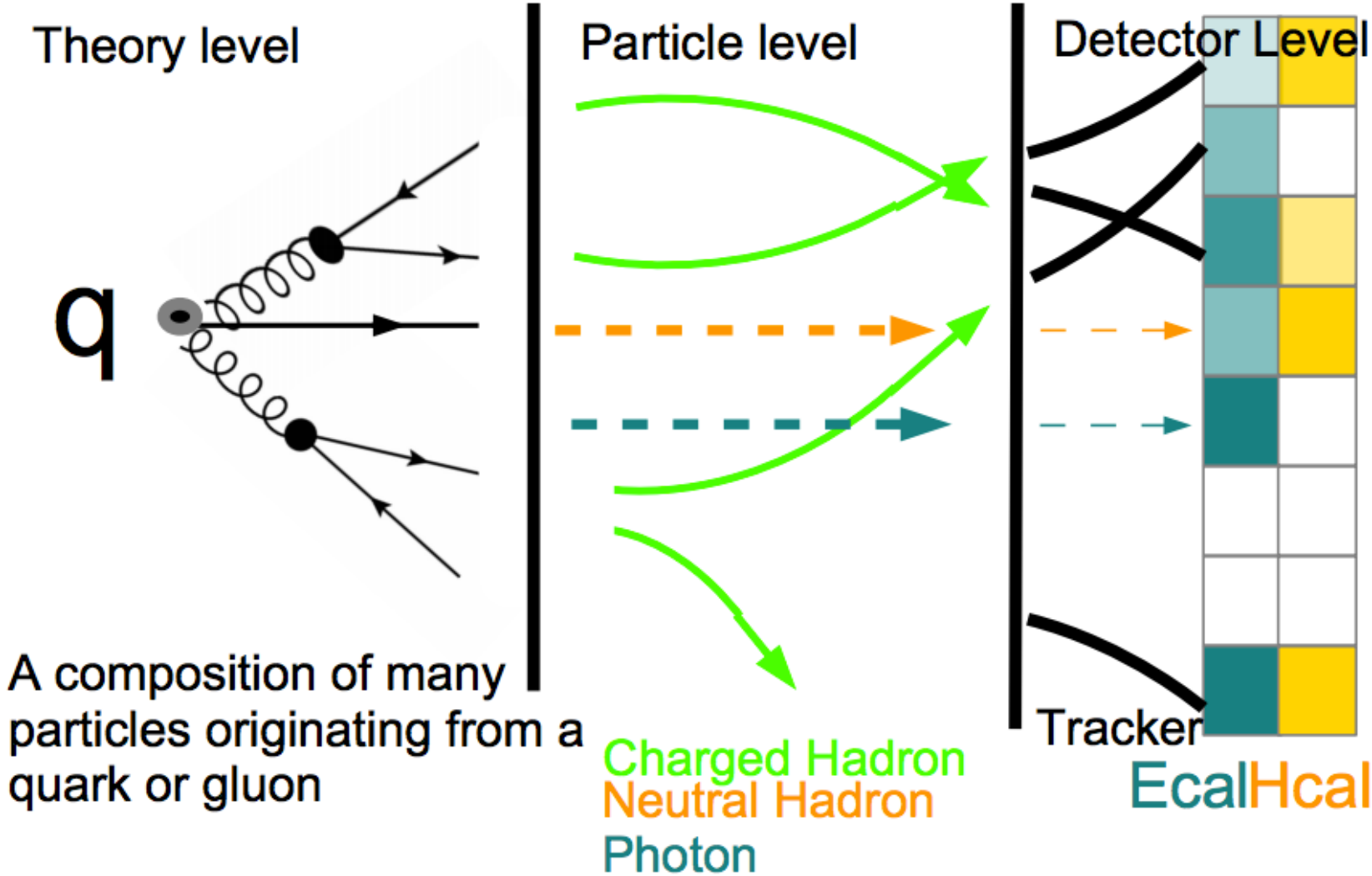
- Information from different subdetectors is combined to identify particles



r-Φ slice

Jets

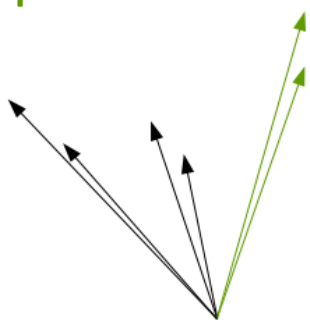
- **Jet**: a spray of particles from a quark or gluon
  - Due to QCD confinement



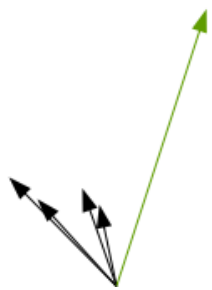
Jet Clustering: Anti-k_t

- Take the smallest $\Delta R \min(p_T^i, p_T^j)^\alpha$ and combine
  - $\Delta R^2 = \Phi^2 + \eta^2$
  - i,j: jet indices
  - $\alpha = 1$: k_t, $\alpha = 0$: Cambridge Aachen, $\alpha = -1$: Anti-k_t

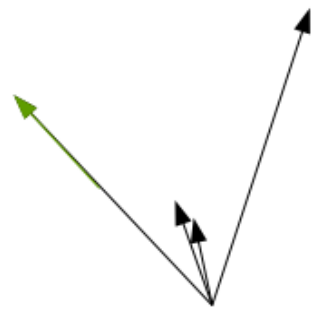
1.) Start with high p_T particles



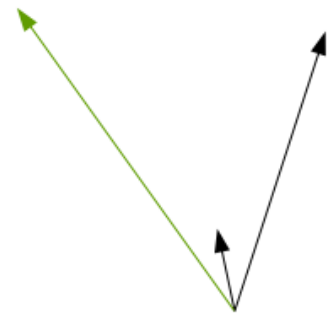
2.) Merge



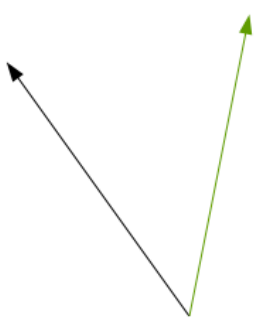
3.) Merge next pair



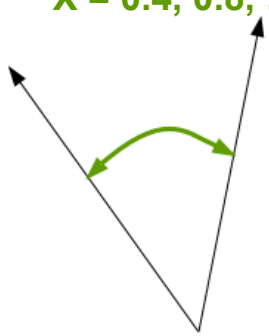
4.) Merge next pair



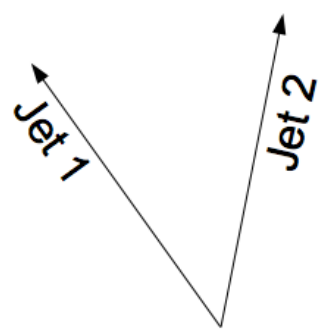
5.) Merge next pair

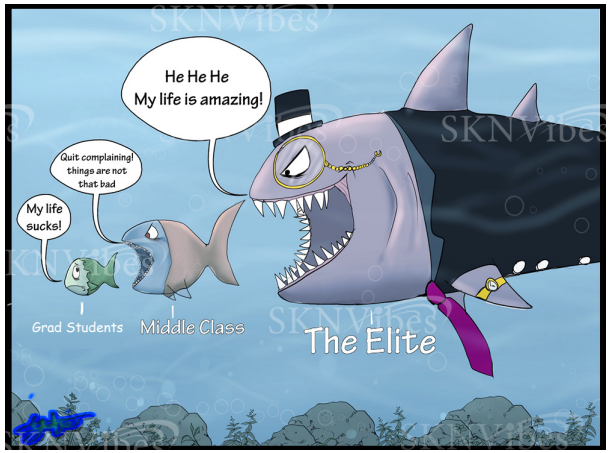


6.) Stop! $\Delta R > X$
$X = 0.4, 0.8, \ldots$



7.) Done!

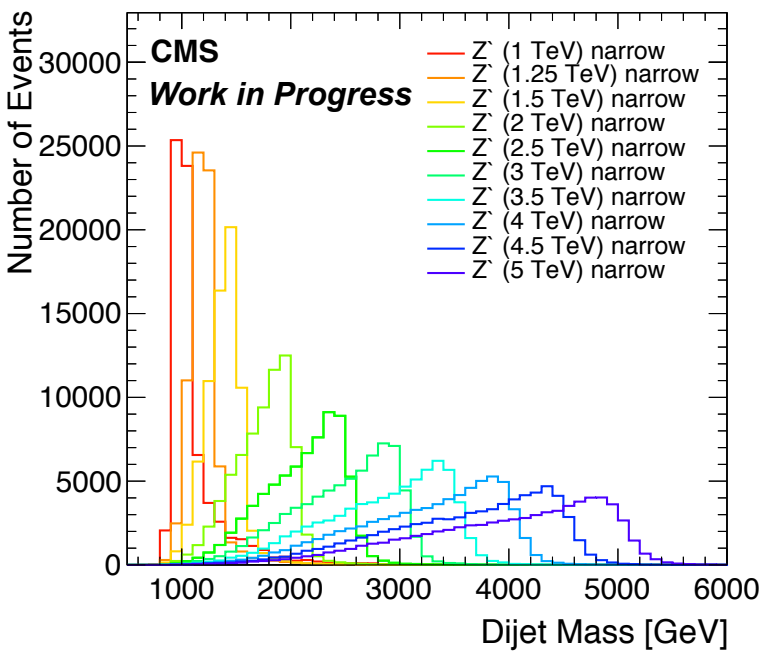


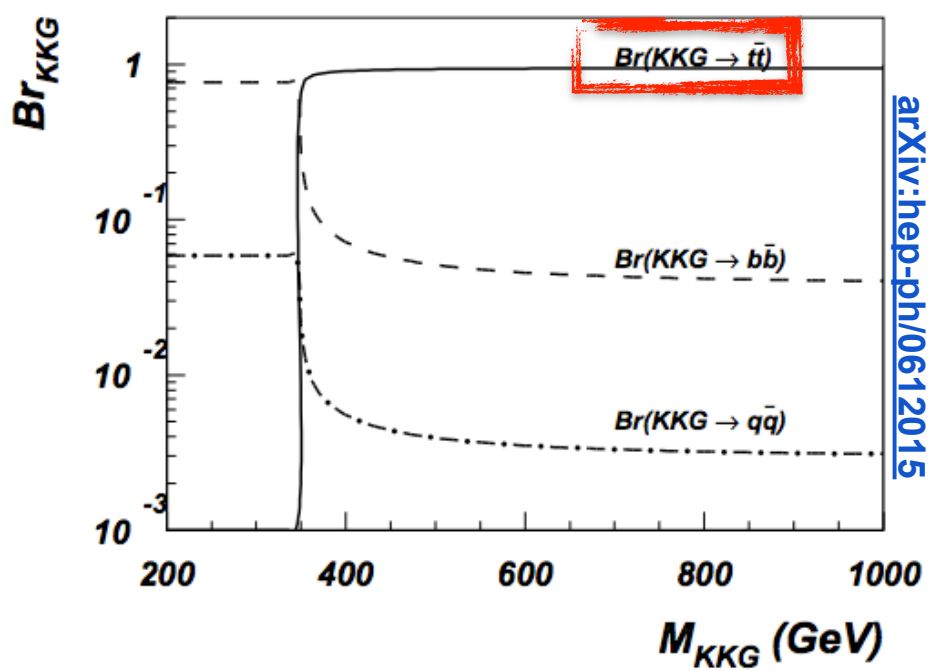


Analysis

Heavy Resonances Decaying to Top Pairs

- The SM is quite successful, but leaves unanswered questions, e.g. why gravity is so weak
- Randall Sundrum model of extra dimensions predicts heavy gluon (RS KK gluon)
  - Decays almost exclusively to $t\bar{t}$ pairs
- Also search for generic Z′ boson decaying to $t\bar{t}$
  - Narrow, wide, and extra wide models (Γ/M = 1%,10%, 30%)

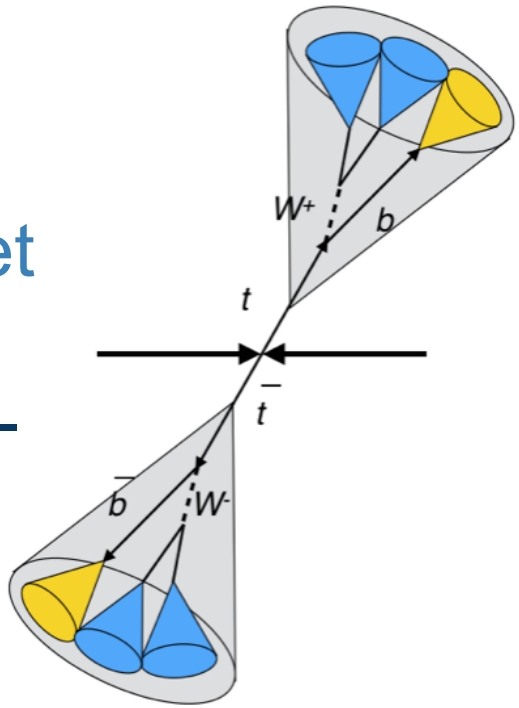




arXiv:hep-ph/0612015

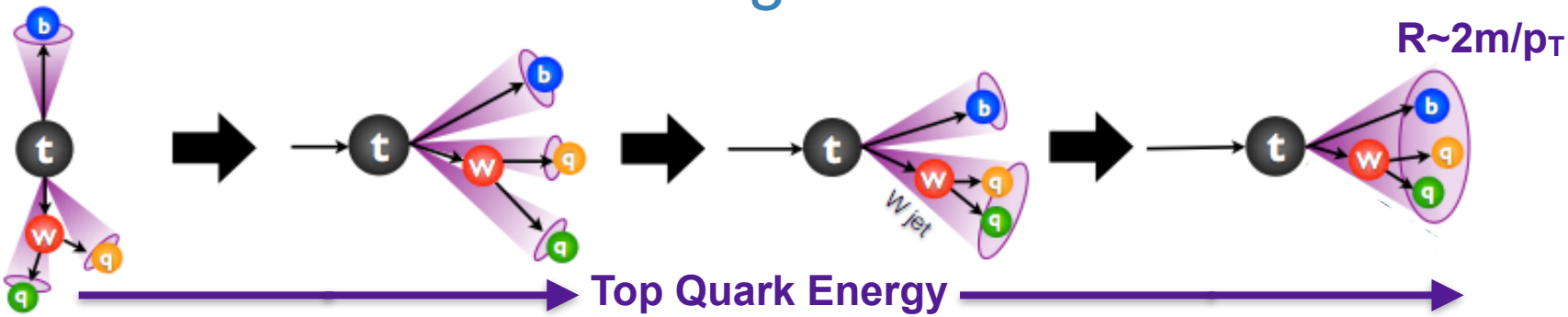
Z′ to $t\bar{t}$

- Final state: two hadronic tops
  - Massive resonance → tops produced with high p_T → decay products collimated into a single jet



- Signature: two large (R=0.8) back-to-back jets: $|\Delta\Phi| > 2.1$, $p_T > 400$ GeV

- Background: SM $t\bar{t}$ and QCD events
  - QCD estimated using data

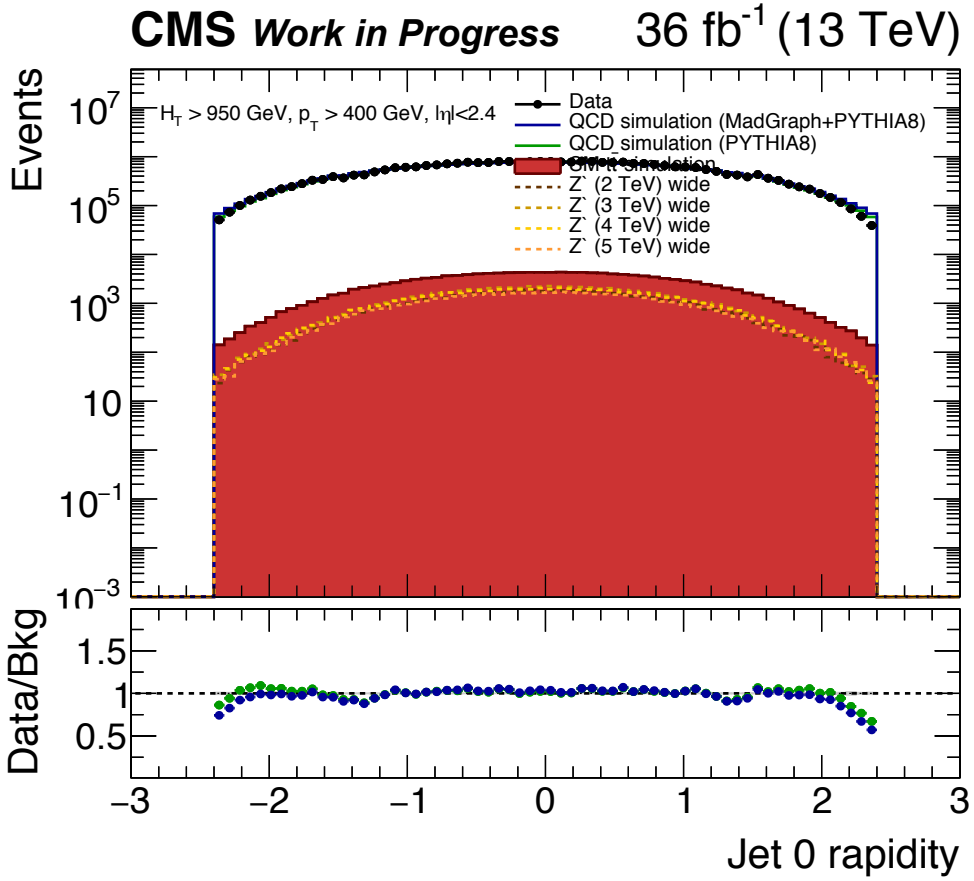


Background Reduction: Top-tagging

Why Top-tagging?

- QCD rate >> expected signal rate
  - Use top-tagging to reduce QCD background

- Top-tagging - uses jet substructure tools
  - Jet mass
  - Subjet multiplicity
  - Subjet b-tagging
  - Pileup removal

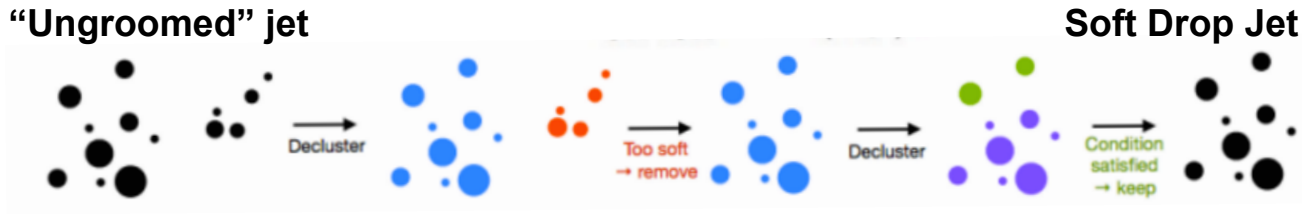


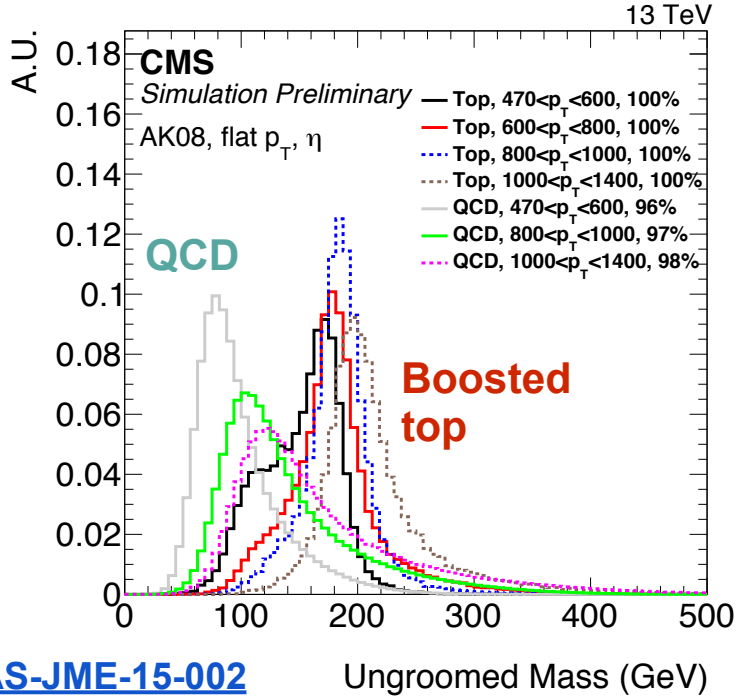
Boosted Top Quark ID: Jet Mass

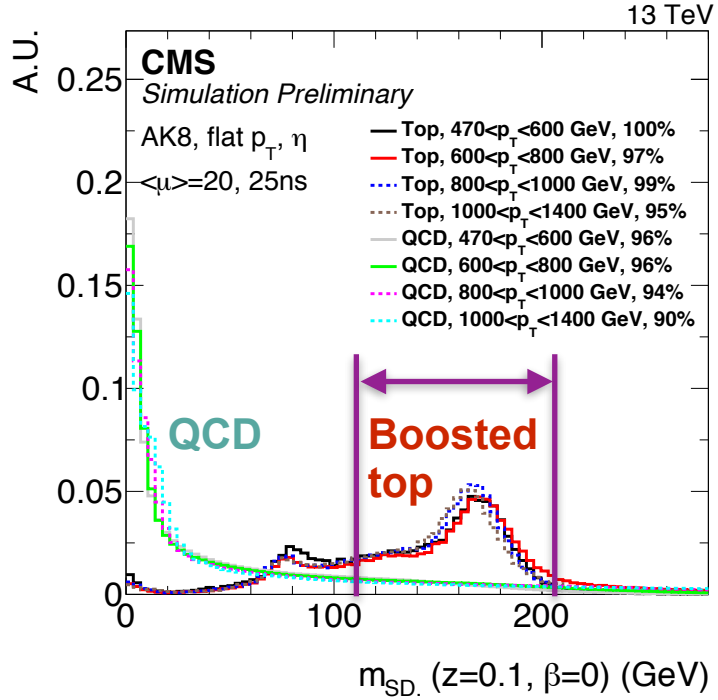
- Signal jet mass should be ~ top quark mass

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\rm cut} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$ tunable variables

- Measure groomed jet mass
  - **Soft drop** algorithm removes soft, wide-angle radiation from a jet







CMS-PAS-JME-15-002

Boosted Top Quark ID: N-subjettiness

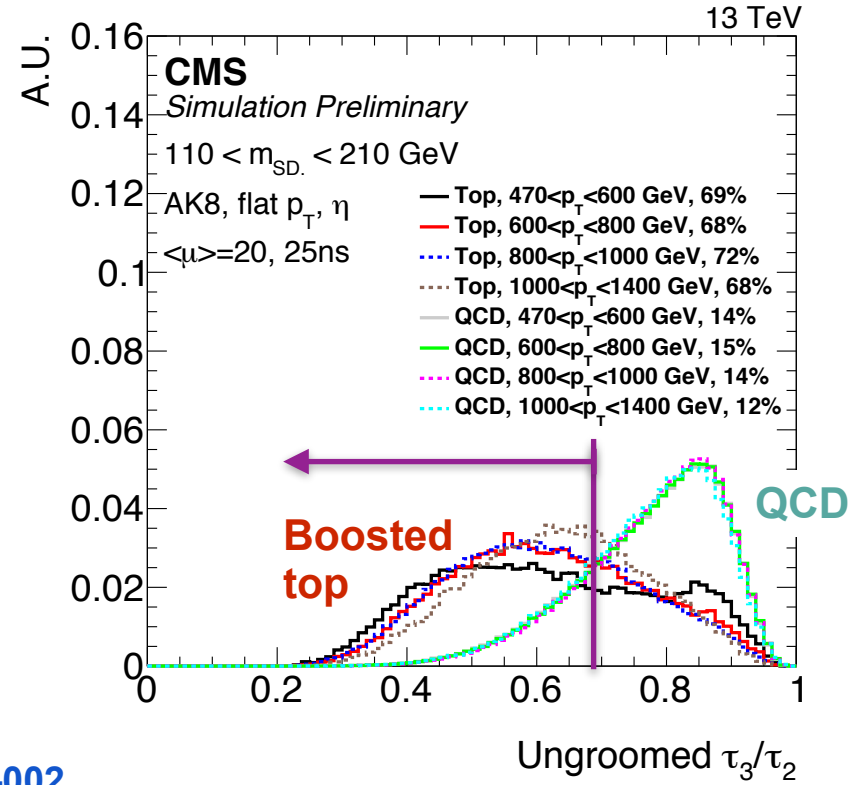
- N-subjettiness: Jet shape variable to measure consistency of jet to have N or fewer subjets

$$\tau_N = \frac{1}{d_0}\sum_k p_{Tk} \times \Delta R_k^{min} \quad \text{with} \quad d_0 \equiv \sum_k p_{Tk} \times R$$

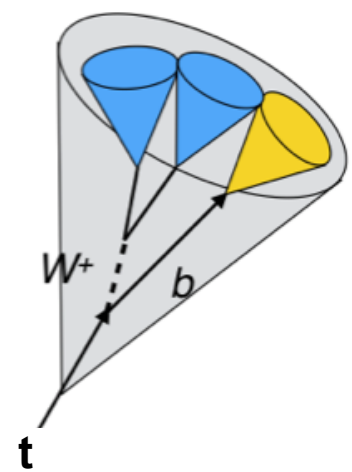
p_T of constituent k

distance between constituent k & axis of closest subjet

large-R jet distance parameter



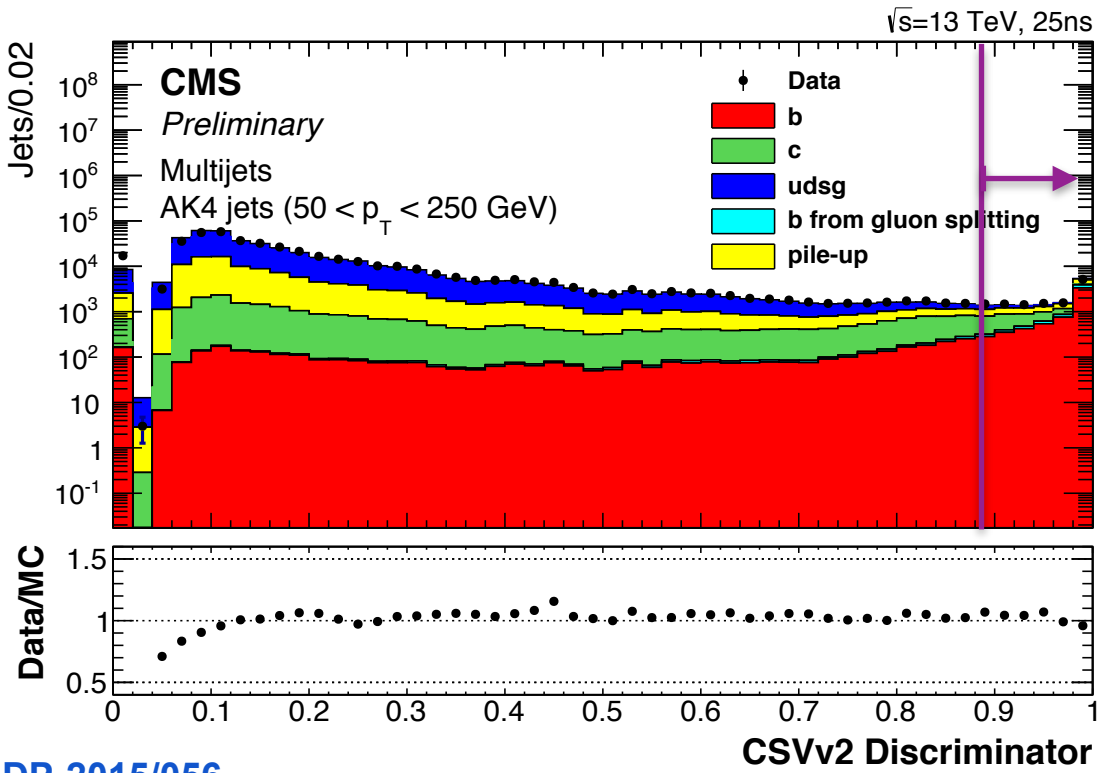
Boosted top: 3 subjets

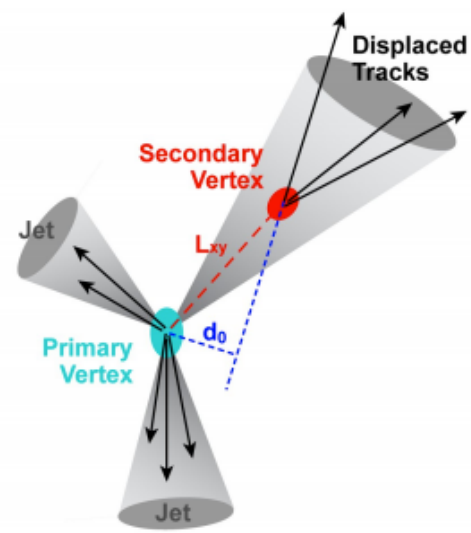


CMS-PAS-JME-15-002

Boosted Top Quark ID: Subjet b-tagging

- b quarks are massive and have a relatively long lifetimes
  - b jets decay from secondary vertex
- Combined Secondary Vertex Version 2 (CSVv2) Algorithm
  - Uses **displaced track information** and **secondary vertex information** to determine probability that a (sub)jet comes from a b quark

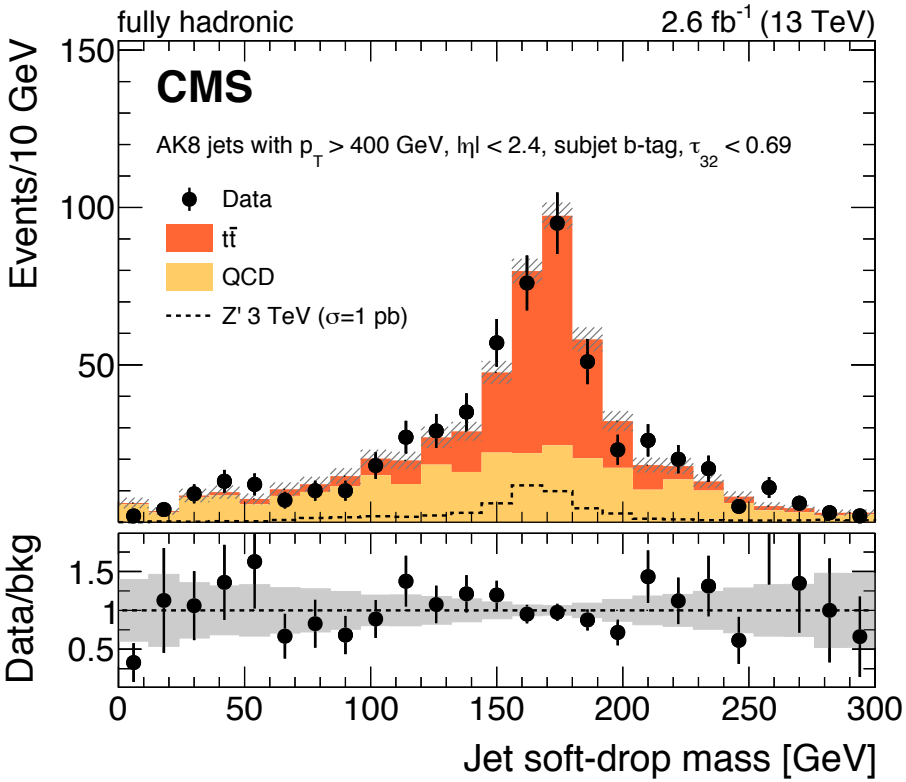


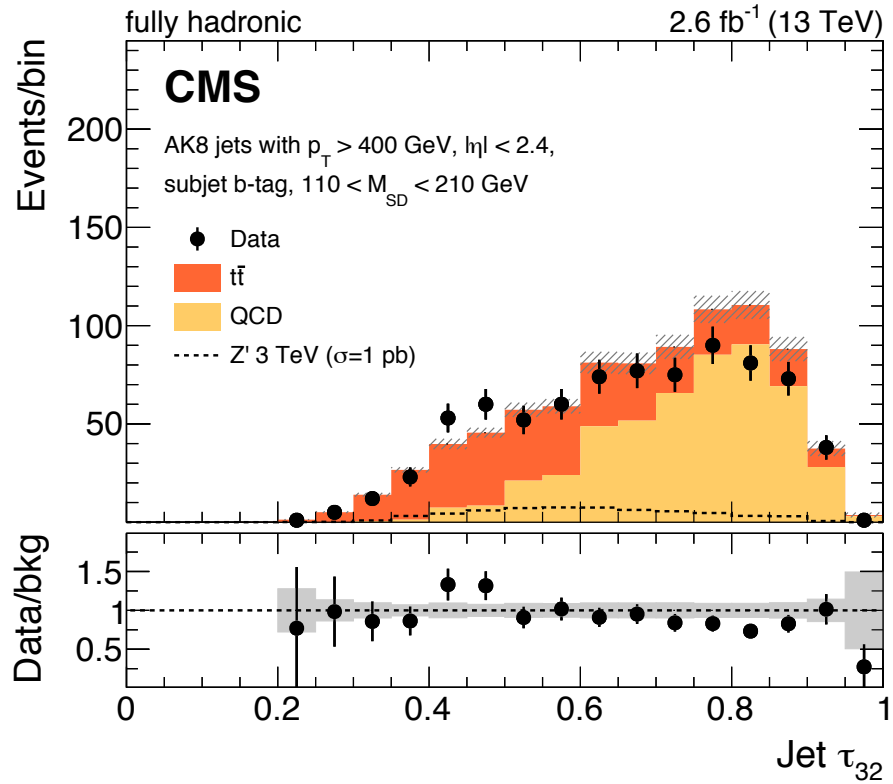


CMS DP-2015/056

Combine Jet ID variables: CMS Top Tagger Version 2

- CMS top tagger for Run II
- CMSTT V2 variables show good data/MC agreement

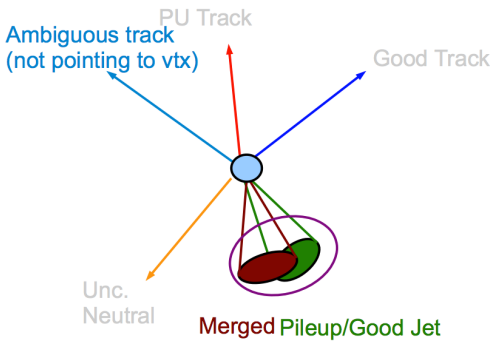




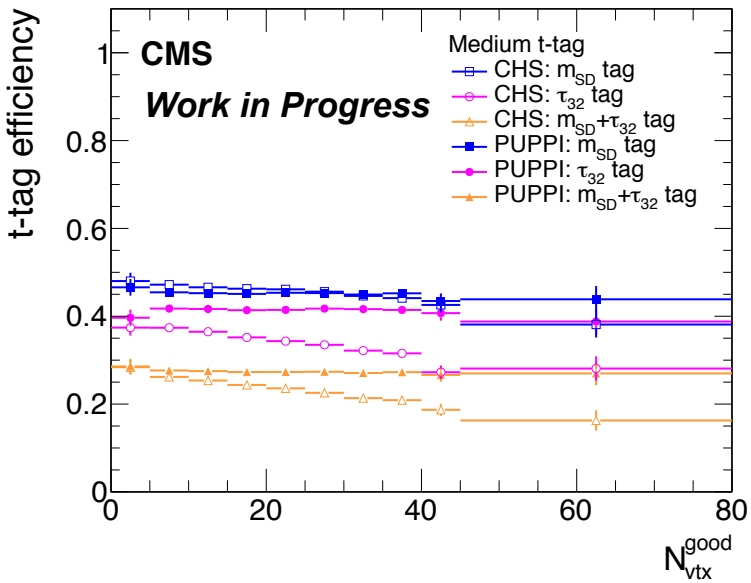
arXiv:1704.03366

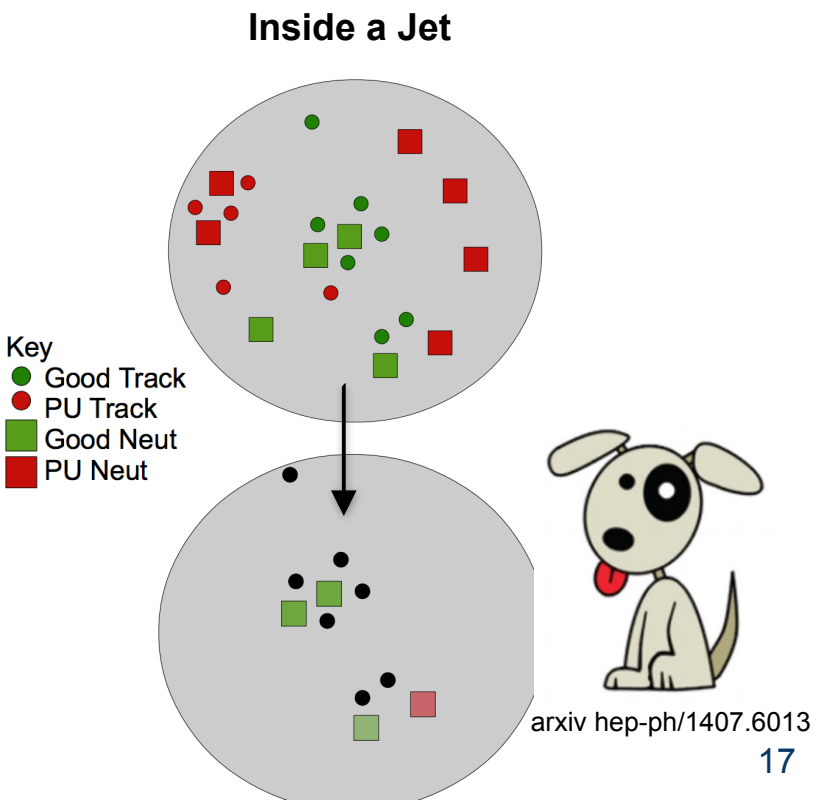
New for the 2016: PUPPI

- **PUPPI - PileUp Per Particle Id**
  - weights particles based on their probability to have originated from a pileup vertex



- PUPPI variables improve performance w.r.t. number of primary vertices





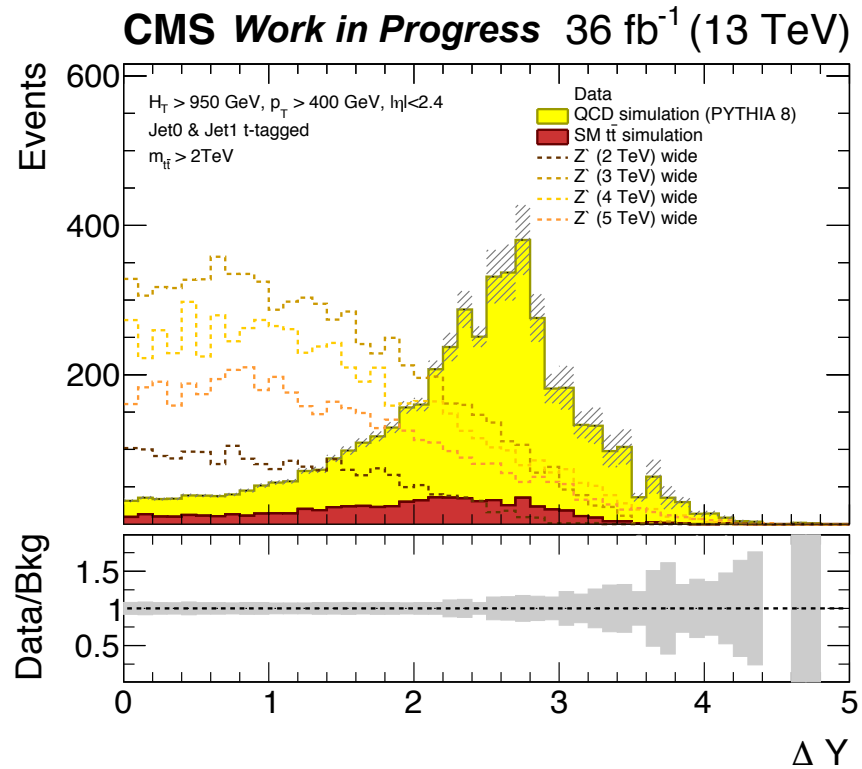
arxiv hep-ph/1407.6013

Back to Z′!

Z′ Event Selection

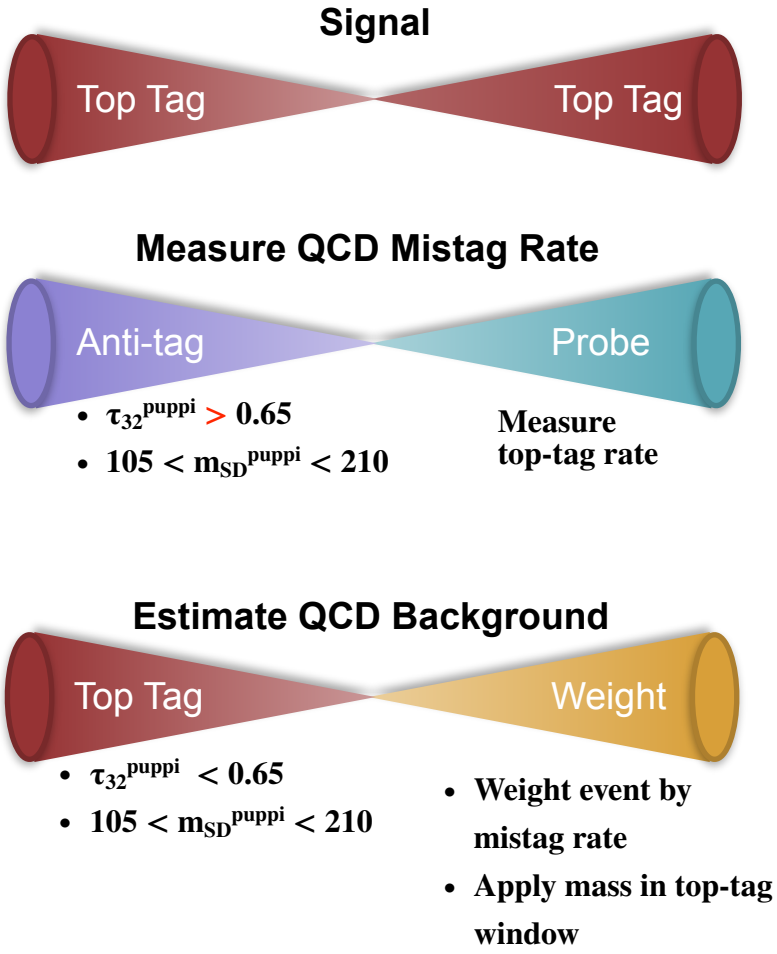
- Preselection: 2 AK8 jets
  - p_T > 400 GeV, |ΔΦ| > 2.1
  - H_T > 950 GeV
- Signal Region: 2 PUPPI-top-tagged jets
- Six event categories
  - Number of jets with a PUPPI subjet b-tag (0, 1, 2)
  - Low (ΔY < 1) or high (ΔY > 1) rapidity difference
    - Low ΔY region more sensitive
- Sensitive variable: $t\bar{t}$ invariant mass

	PUPPI top tag
Soft drop mass	105-210
N-subjettiness	$\tau_{32}^{puppi} < 0.65$



Backgrounds

- $t\bar{t}$: estimated from simulation
- QCD: estimated from data
  - Use "anti-tag and probe method"
    - Measure mistag rate as a function of jet momentum
  - Estimate QCD background in a singly top-tagged region
    - Use previously measured mistag rate
  - Subtract $t\bar{t}$ contamination using simulation

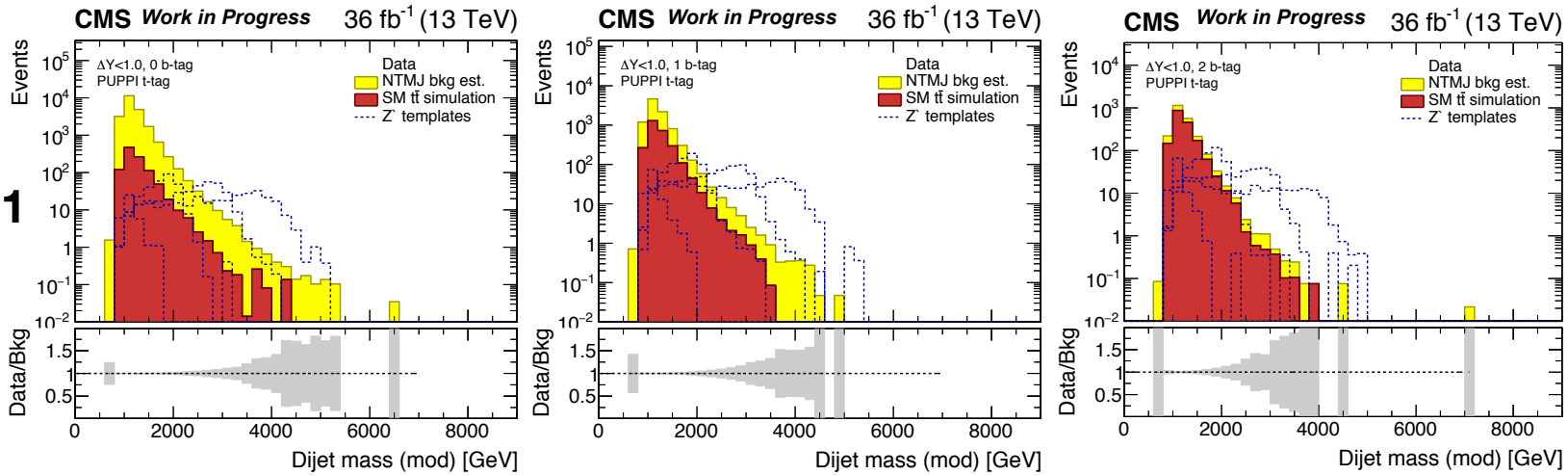


Dijet Mass Templates (Blinded)

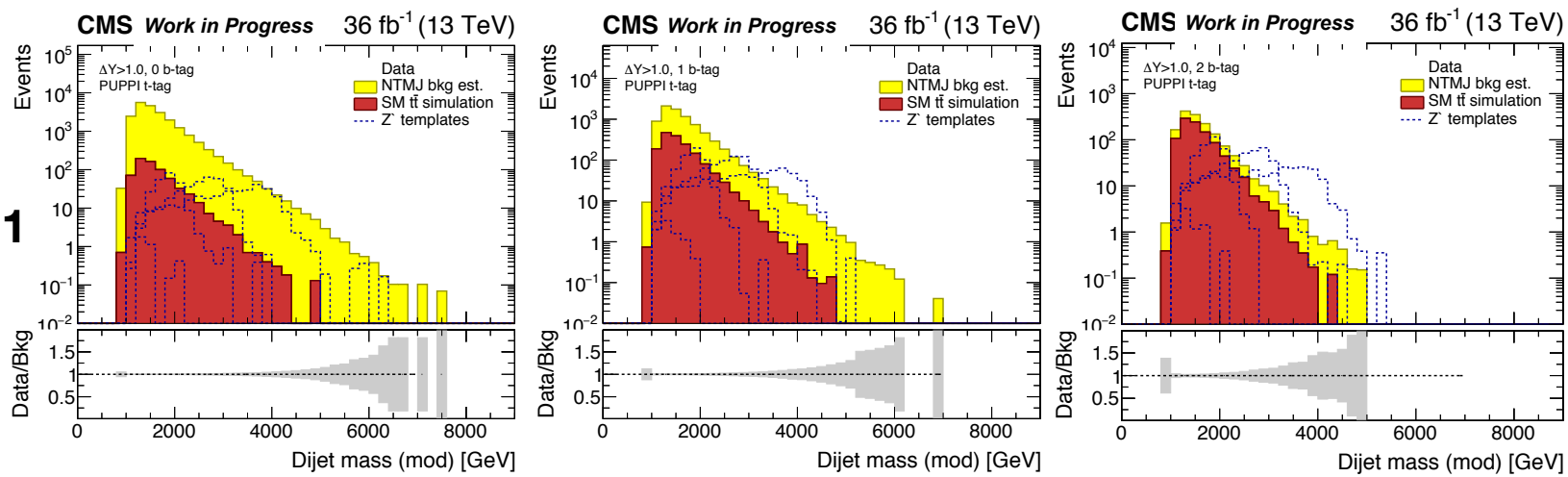
- Two b-tag, low delta rapidity region most sensitive
- Combine signal region templates to set limits

0 b-tags | **1 b-tag** | **2 b-tags**

|Δy| < 1

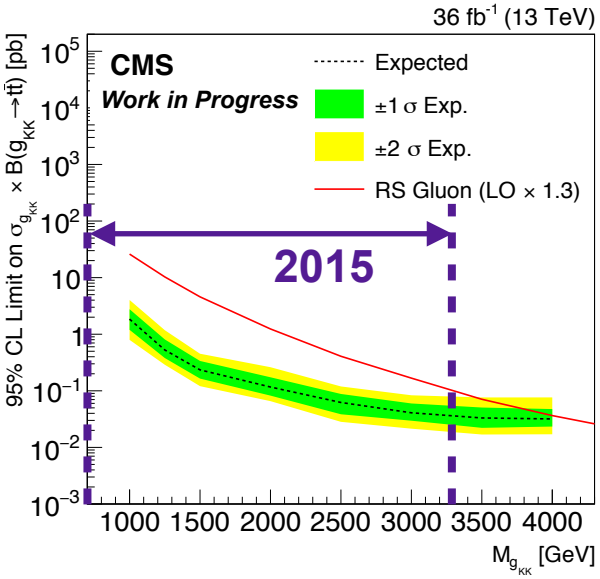


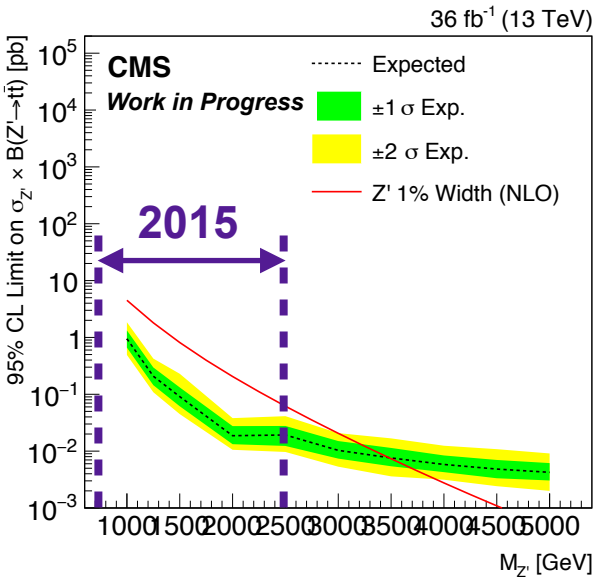
|Δy| > 1

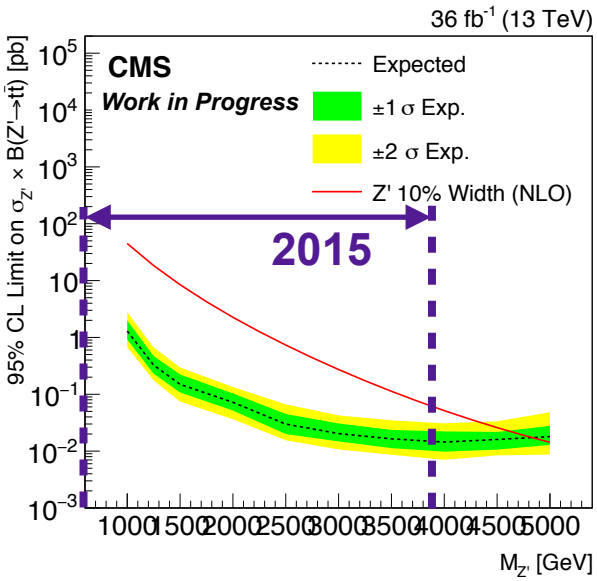


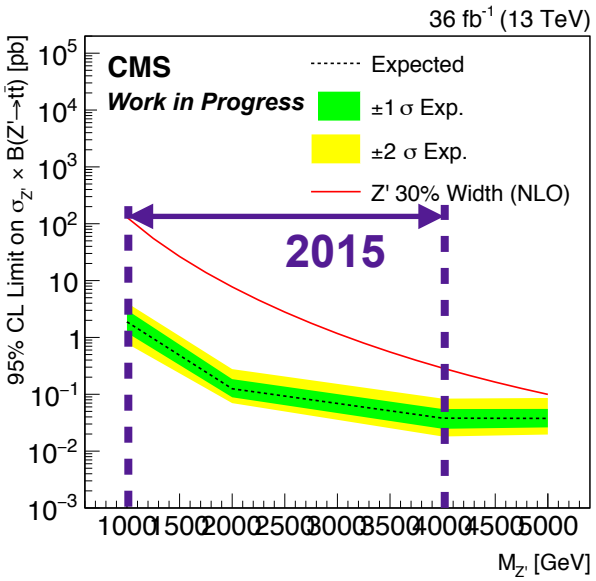
Expected Limits

- Apply full set of systematics
- Expected limits: improvement on **2015 combined limits**
  - Combination of all-hadronic and semi-leptonic final states published in JHEP earlier this year









Conclusions and Outlook

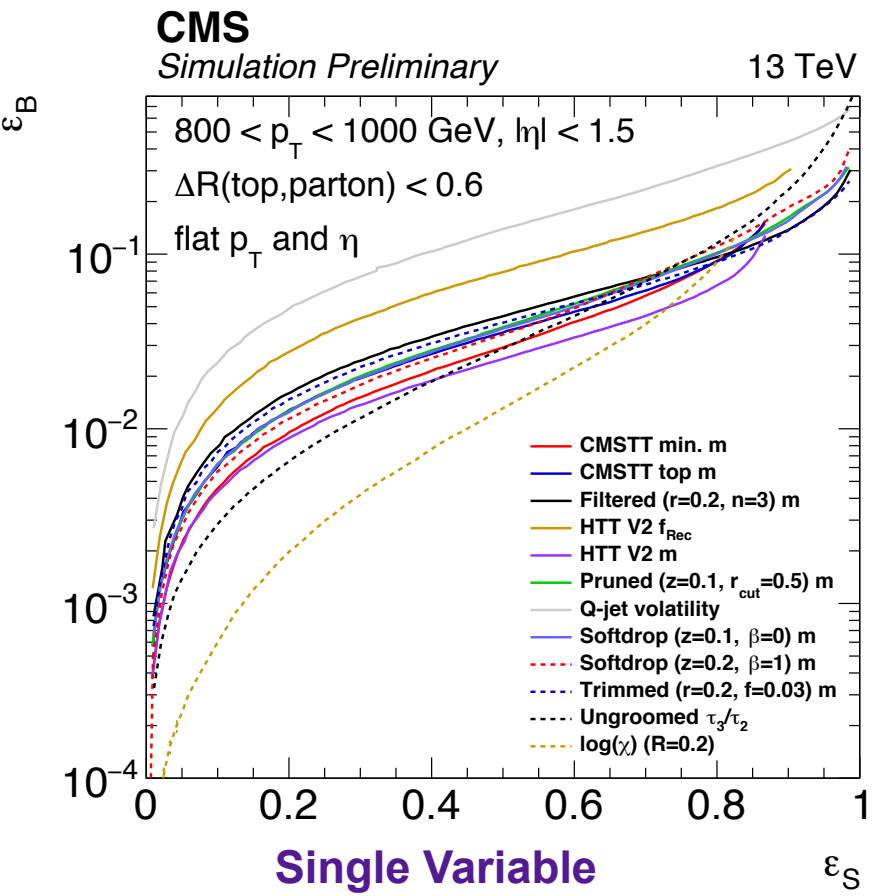
- My thesis analysis: Z' to $t\bar{t}$
  - All-hadronic channel
  - Developed Run II CMS top tagger
  - Introduced PUPPI variables for 2016 analysis
  - Employed QCD background estimate techniques
  - Extended limits to 5 TeV
    - 2016 analysis is looking to be the most sensitive search yet!

- Future plans
  - February 2018: starting CMS postdoc position at SUNY Buffalo
  - Will work on developing more tools to apply to boosted CMS searches and measurements

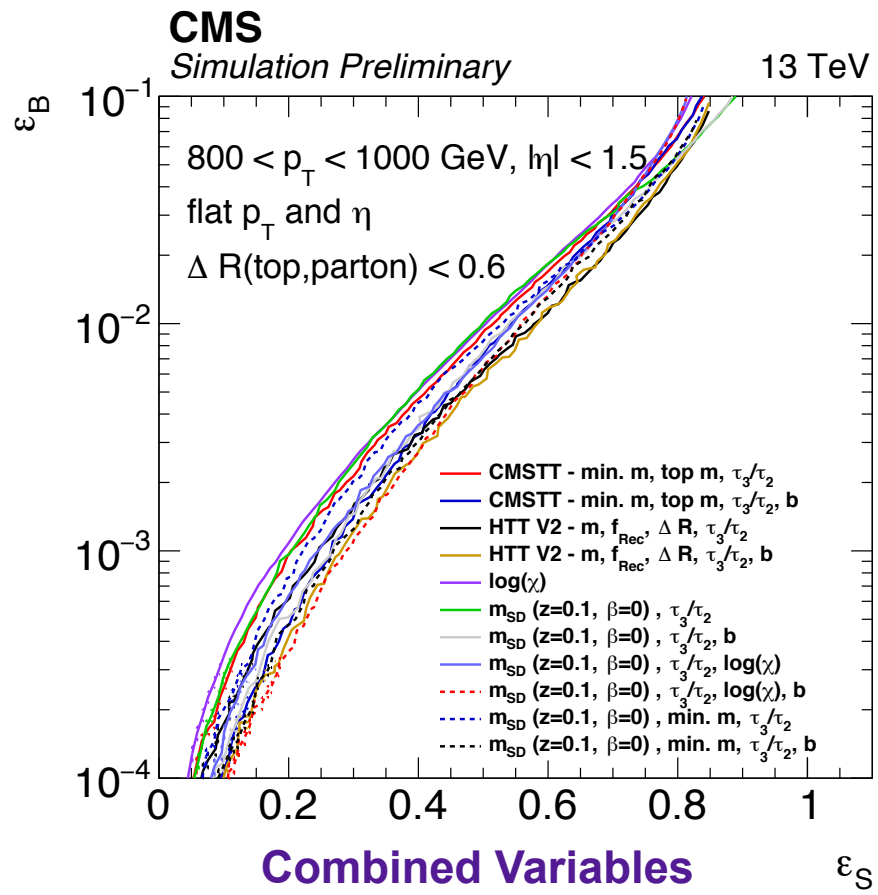
Additional Material

CMS Top Tagger V2 (CMSTT V2) Development

- I investigated optimizing top tagging for Run II
  - First, compared performance of different variables in MC
    - QCD mistag rate vs boosted top-tagging efficiency



Single Variable

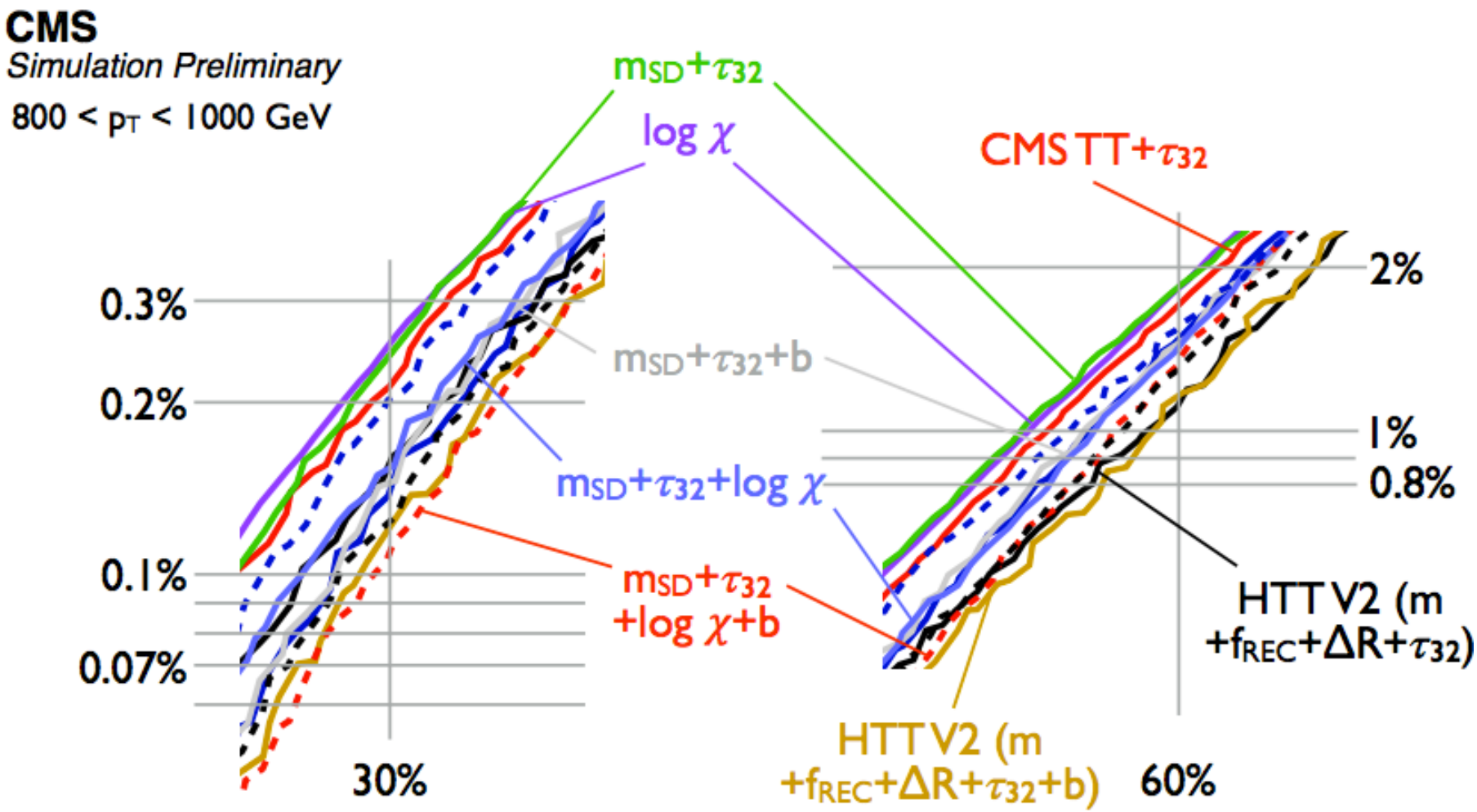


Combined Variables

Improved Performance →

Combined Variable ROC Curves: Zoom In

- Combined variables similarly performant
  - Consider behavior as a function of kinematic variables

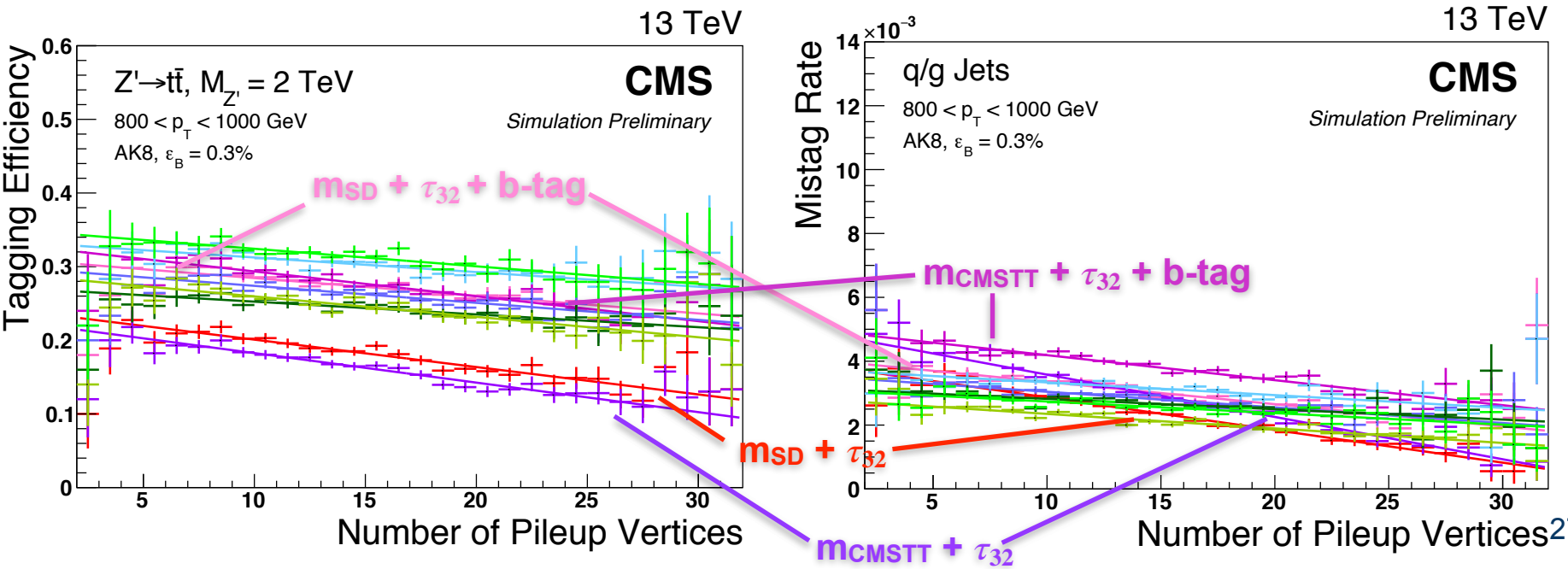


Developed Working Points

- Recommended for Run II: Soft drop + τ_{32} (CMSTT V2)
  - With added b-tagging, improvement in background rejection w.r.t. CMSTT

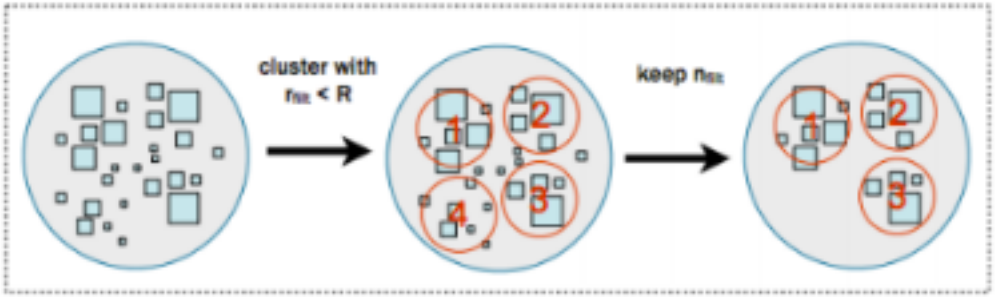
Substructure Algorithms Used for High p_T Working Points

	$110 < M_{SD}(z=0.1, \beta=0) < 210$ GeV	CMSTT V2		$\tau_3/\tau_2 < 0.5$	
	$110 < M_{SD}(z=0.1, \beta=0) < 210$ GeV			$\tau_3/\tau_2 < 0.6$	b tag > 0.8
	$140 <$ CMSTT Mass < 250 GeV			$\tau_3/\tau_2 < 0.5$	
	$140 <$ CMSTT Mass < 250 GeV			$\tau_3/\tau_2 < 0.6$	b tag > 0.8
	$120 <$ HTT V2 Mass < 180 GeV	$f_{Rec} < 0.2$	$-0.1 < \Delta R_{opt} < 0.1$	$\tau_3/\tau_2 < 0.6$	
	$90 <$ HTT V2 Mass < 170 GeV	$f_{Rec} < 0.2$	$-0.2 < \Delta R_{opt} < 0.2$	$\tau_3/\tau_2 < 0.7$	b tag > 0.8
		SD $\log(\chi_1) > 5.4$			
	$90 < M_{SD}(z=0.1, \beta=0) < 180$ GeV	SD $\log(\chi_1) > 4.6$		$\tau_3/\tau_2 < 0.7$	
	$70 < M_{SD}(z=0.1, \beta=0) < 190$ GeV	SD $\log(\chi_1) > 3.9$		$\tau_3/\tau_2 < 0.7$	b tag > 0.6



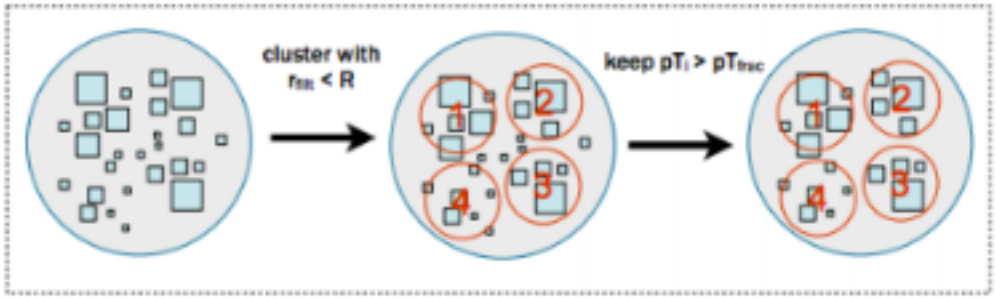
Soft Drop Algorithm

filtering



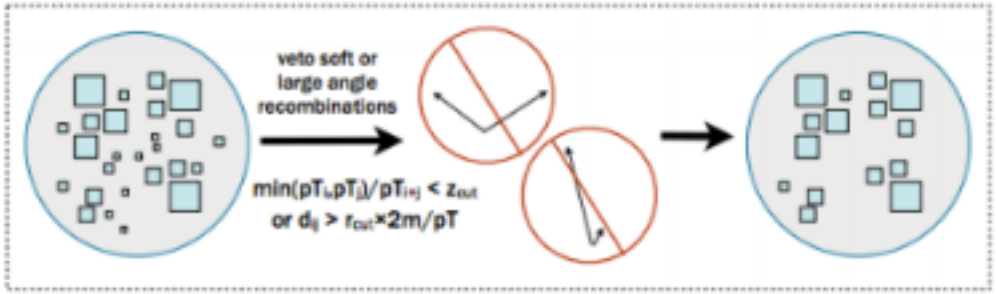
default:
$r_{filt} = 0.3$
$n_{filt} = 3$

trimming



default:
$r_{filt} = 0.2$
$pT_{frac} = 0.03$

pruning



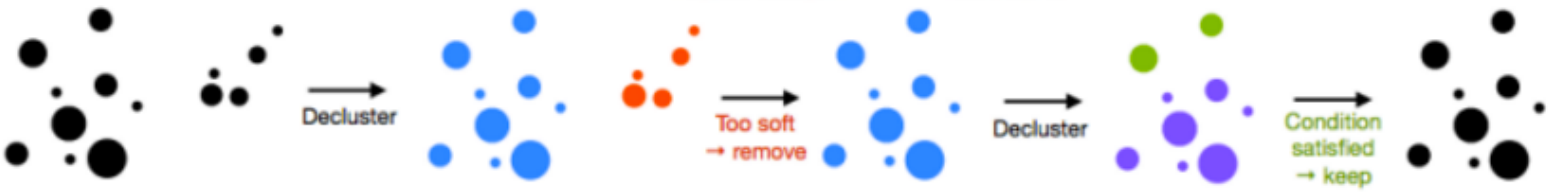
default:
$z_{cut} = 0.1$
$r_{cut} = 0.5$

Jim Dolen, Justin Pilot, Joshua Swanson, Nhan Tran

Soft Drop Algorithm

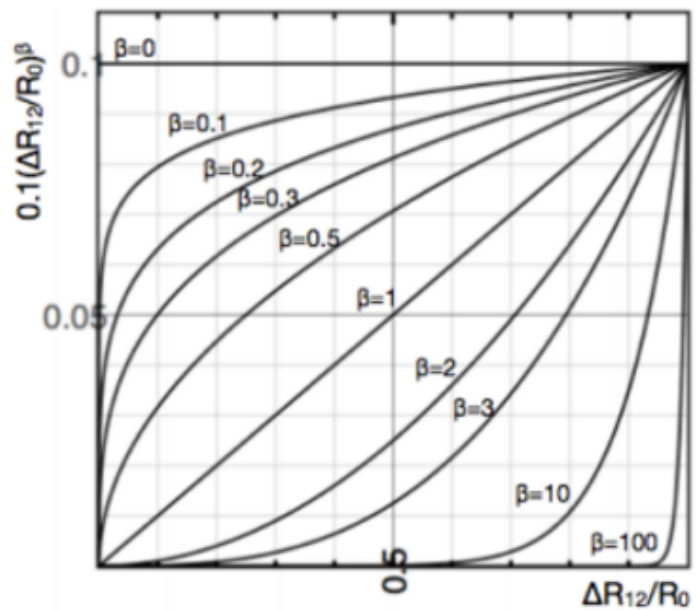
- Recursively decluster jet. Remove the softer component unless the soft drop condition is satisfied.

$$\text{Soft Drop Condition:} \quad \frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta}$$



J. Dolen

- Soft wide angle radiation fails the condition
  - As z_{cut} ↑ ⇒ more aggressive grooming
  - As β ↓ ⇒ more aggressive grooming
- Example (z_{cut} = 0.1) :
  - If β = 0, remove softer subjet if p_T fraction < 0.1 (~equivalent to MMDT)
  - If β > 0, remove softer subjet if p_T fraction < x, where x increases with ΔR and has maximum value 0.1
  - β → ∞ no grooming
  - β <0 soft drop becomes a tagger instead of a groomer (finds jets with hard, large angle subjets)



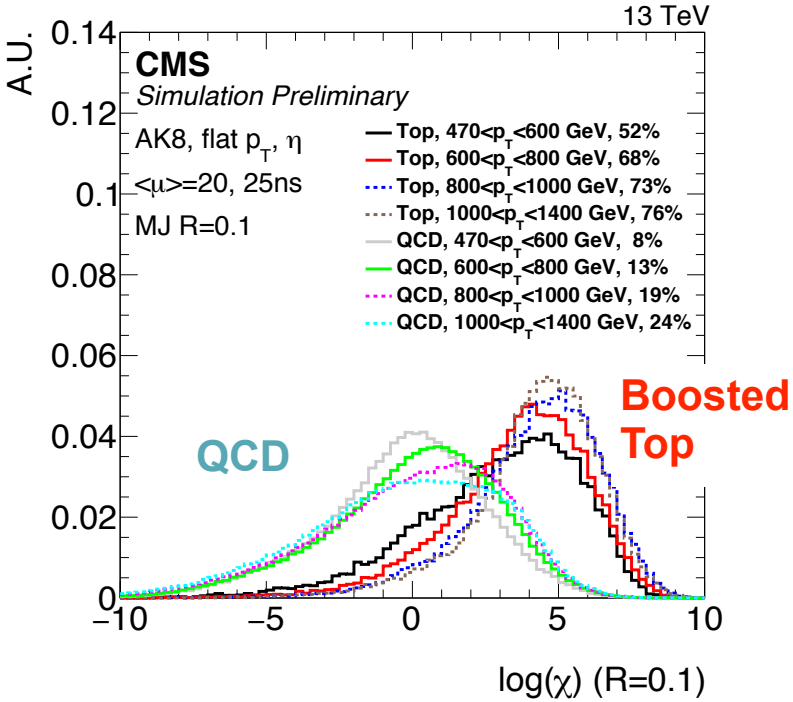
Shower Deconstruction

CMS-PAS-JME-15-002

- Likelihood that large-R jet is signal (hadronic top decay) vs background process (QCD)
  - Input: R=0.1 kt "microjets"
  - Assume each microjet comes from certain source of radiation (top decay, ISR, ...)
  - Calculate probability that microjet configuration comes from particular decay chain
  - Discriminating variable:

$$\chi = \frac{\text{sum of signal probabilities}}{\text{sum of bkg probabilities}}$$

- Powerful but more complicated variable



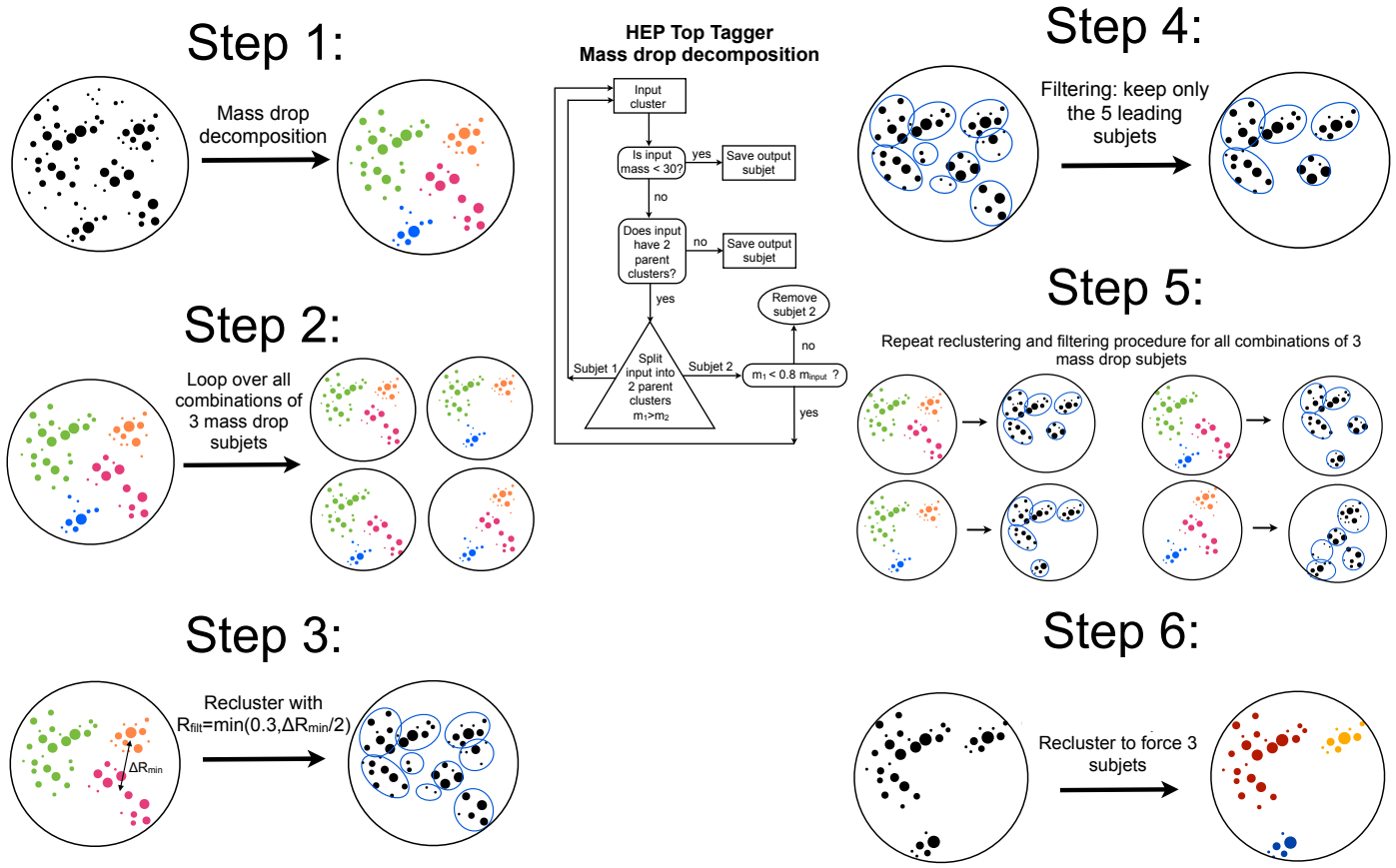
HEP Top Tagger Version 2

CMS-PAS-JME-13-007

- Input: R=1.5 Cambridge-Aachen jets (p_T > 200 GeV)
- Decision variables:
  - Top mass & f_{REC} (normalized ratio of W mass to top mass)

$$f_{REC} = min_{ij} \left| \frac{\frac{m_{ij}}{m_{123}}}{\frac{m_W}{m_t}} - 1 \right|$$

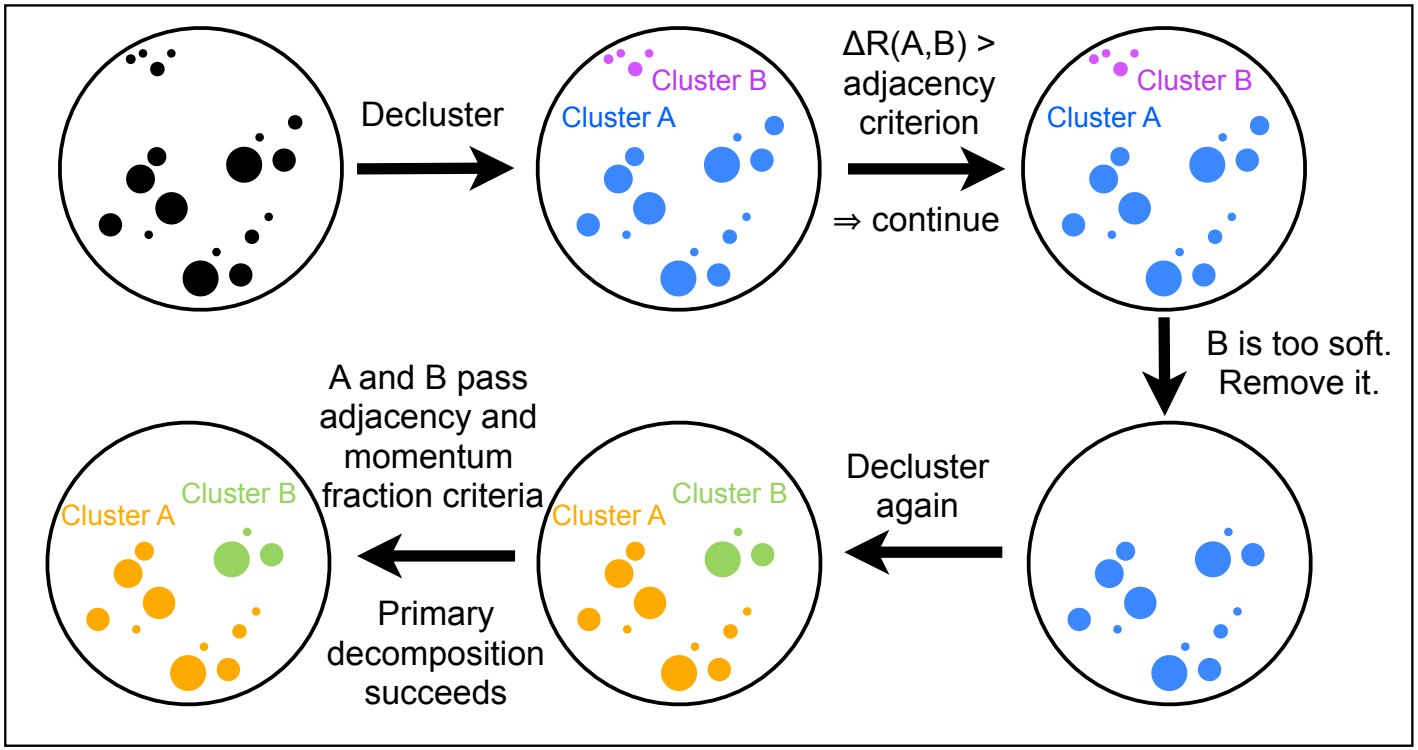
- New! Determines optimal R; $\Delta R = R_{opt} - R_{exp}$



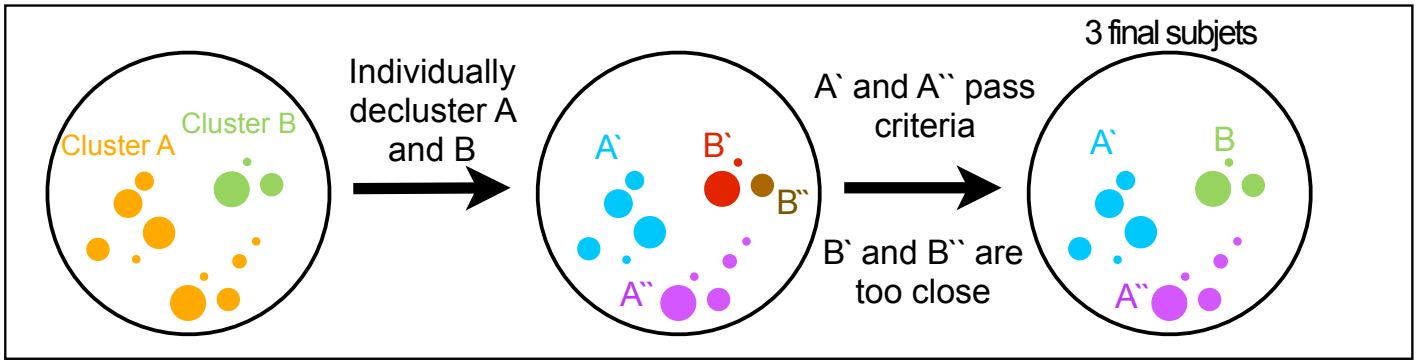
CMS Top Tagger

CMS-PAS-JME-13-007

Primary decomposition

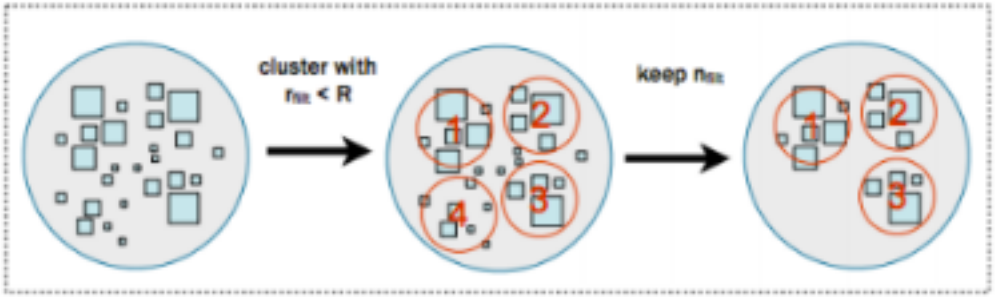


Secondary decomposition



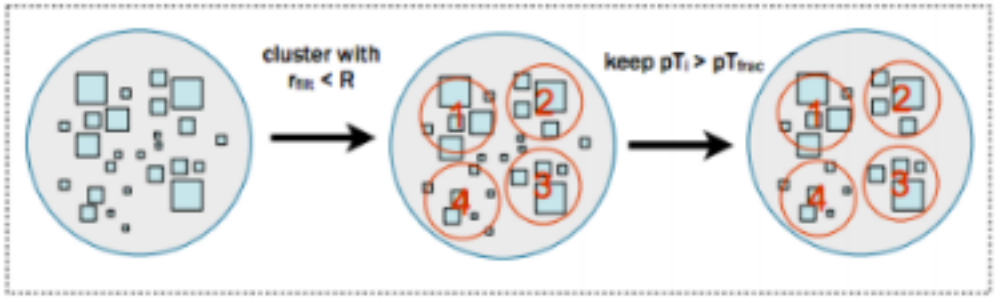
Soft Drop Algorithm

filtering



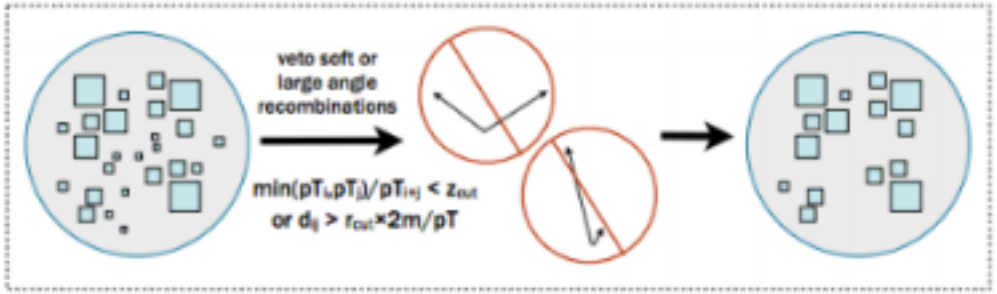
default:
$r_{filt} = 0.3$
$n_{filt} = 3$

trimming



default:
$r_{filt} = 0.2$
$pT_{frac} = 0.03$

pruning



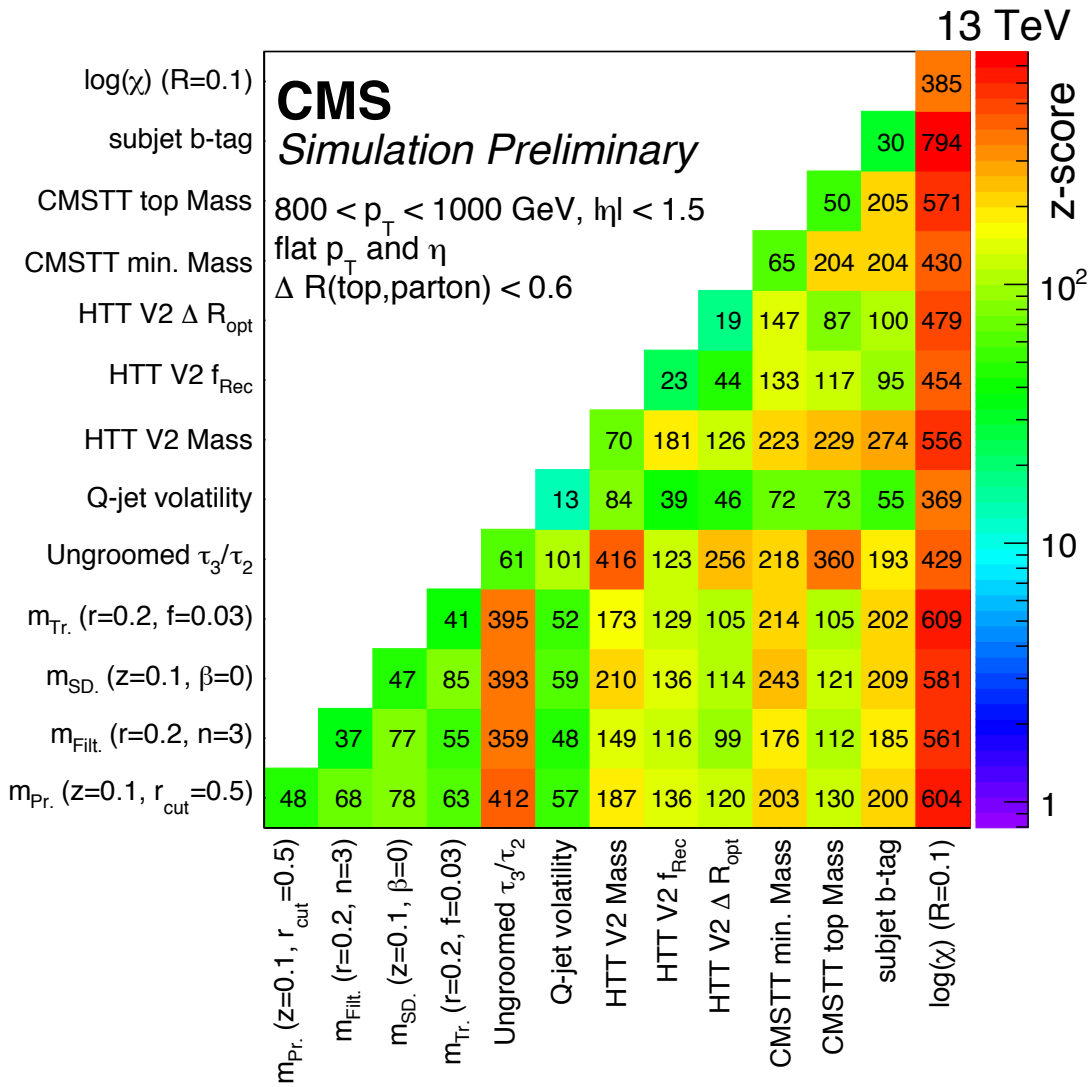
default:
$z_{cut} = 0.1$
$r_{cut} = 0.5$

Jim Dolen, Justin Pilot, Joshua Swanson, Nhan Tran

Combining Variables

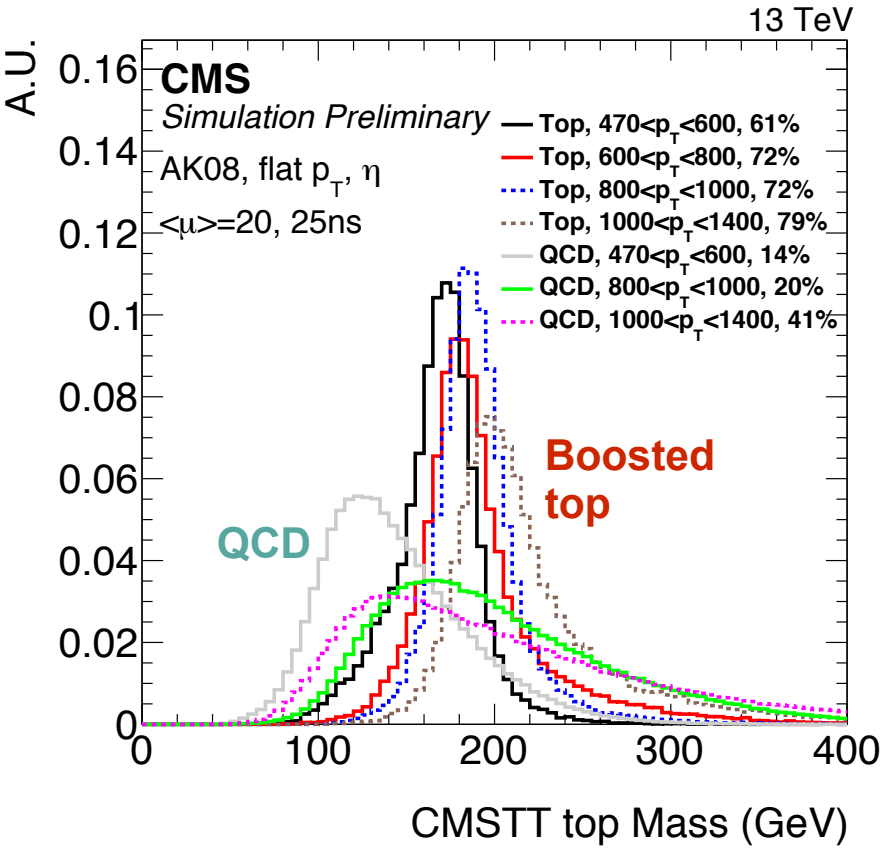
CMS-PAS-JME-15-002

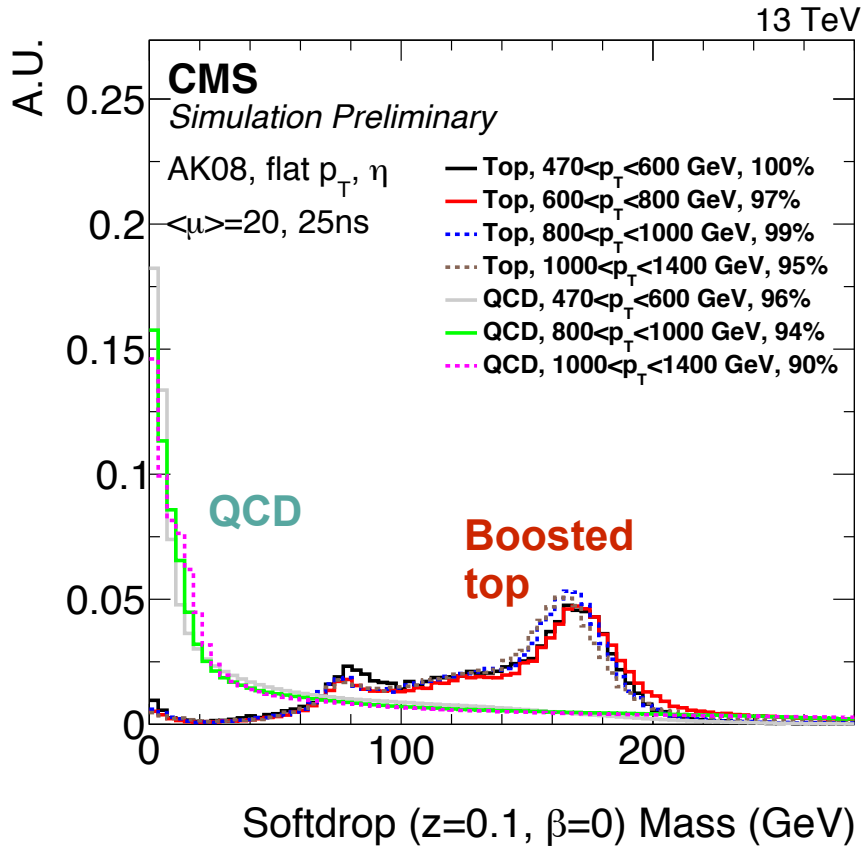
- z-score:
  - **$1/\varepsilon_B$**, for $\varepsilon_B = 0.3$
  - Determined using BDTs (or likelihood ratios for diagonals)
- Best combinations
  - **Shower deconstruction** still very promising
  - **Mass + τ_{32}** also potentially powerful



CMSTT Mass vs Soft Drop Mass

- Soft drop mass is more stable across p_T bins





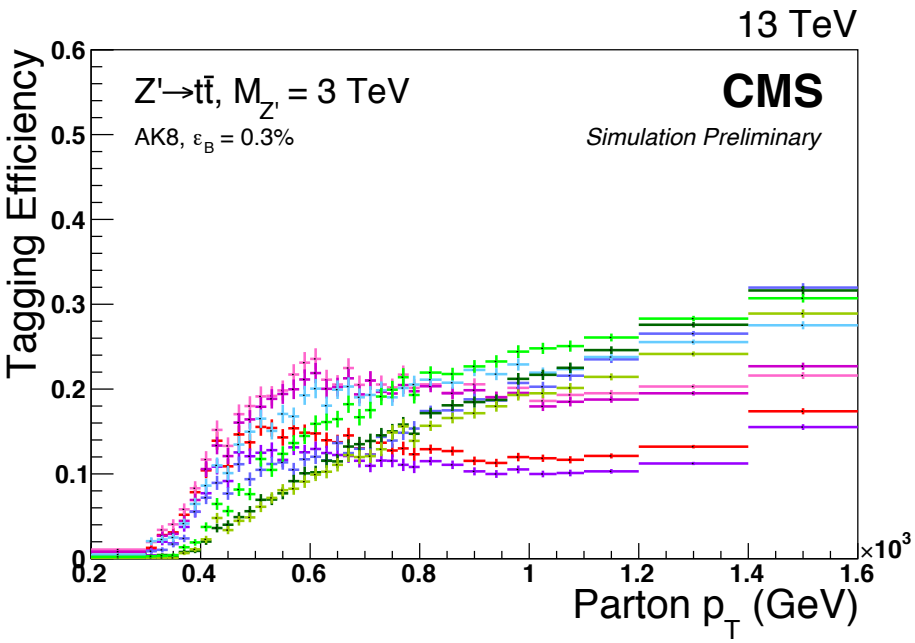
Performance vs p_T

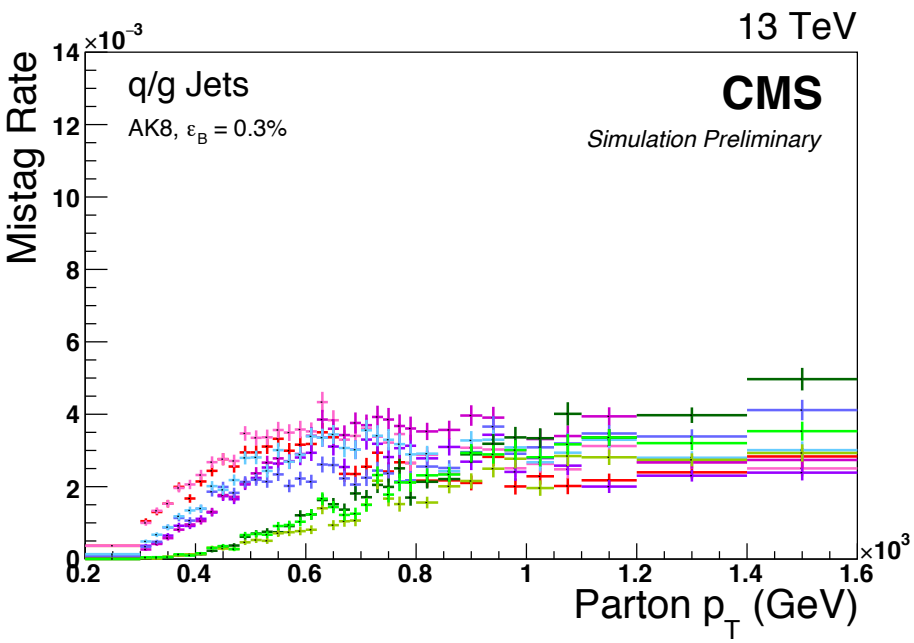
CMS-PAS-JME-15-002

$\varepsilon_B = 0.3\%$

Substructure Algorithms Used for High p_T Working Points

$110 < M_{SD}(z=0.1, \beta=0) < 210$ GeV			$\tau_3/\tau_2 < 0.5$	
$110 < M_{SD}(z=0.1, \beta=0) < 210$ GeV			$\tau_3/\tau_2 < 0.6$	b tag > 0.8
$140 <$ CMSTT Mass < 250 GeV			$\tau_3/\tau_2 < 0.5$	
$140 <$ CMSTT Mass < 250 GeV			$\tau_3/\tau_2 < 0.6$	b tag > 0.8
$120 <$ HTT V2 Mass < 180 GeV	$f_{Rec} < 0.2$	$-0.1 < \Delta R_{opt} < 0.1$	$\tau_3/\tau_2 < 0.6$	
$90 <$ HTT V2 Mass < 170 GeV	$f_{Rec} < 0.2$	$-0.2 < \Delta R_{opt} < 0.2$	$\tau_3/\tau_2 < 0.7$	b tag > 0.8
	SD $\log(\chi_1) > 5.4$			
$90 < M_{SD}(z=0.1, \beta=0) < 180$ GeV	SD $\log(\chi_1) > 4.6$		$\tau_3/\tau_2 < 0.7$	
$70 < M_{SD}(z=0.1, \beta=0) < 190$ GeV	SD $\log(\chi_1) > 3.9$		$\tau_3/\tau_2 < 0.7$	b tag > 0.6

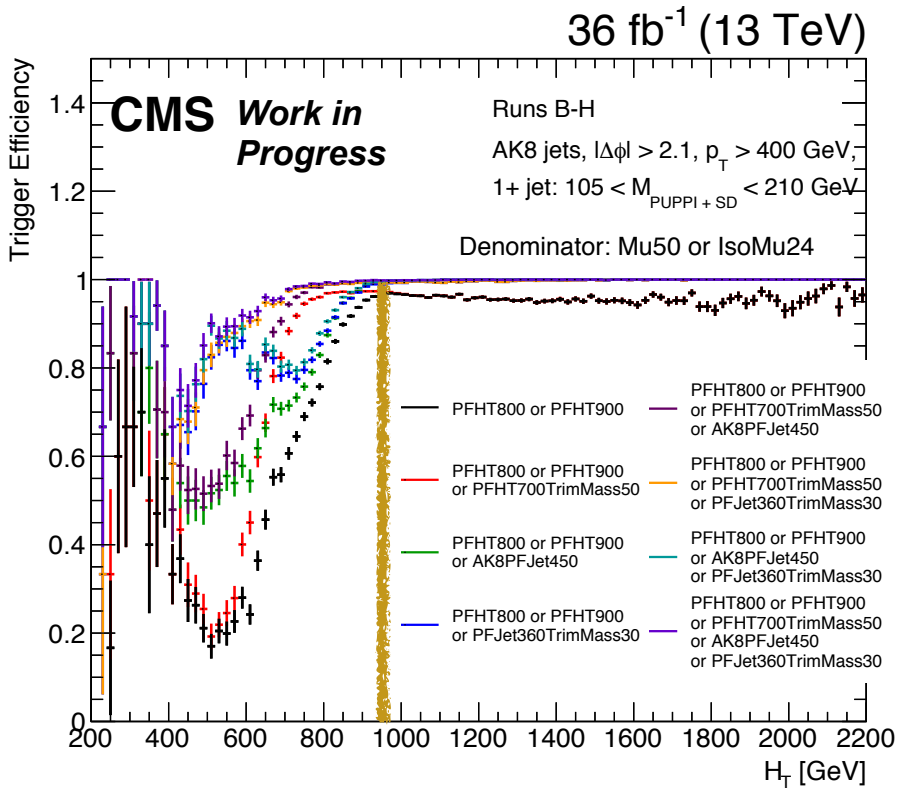




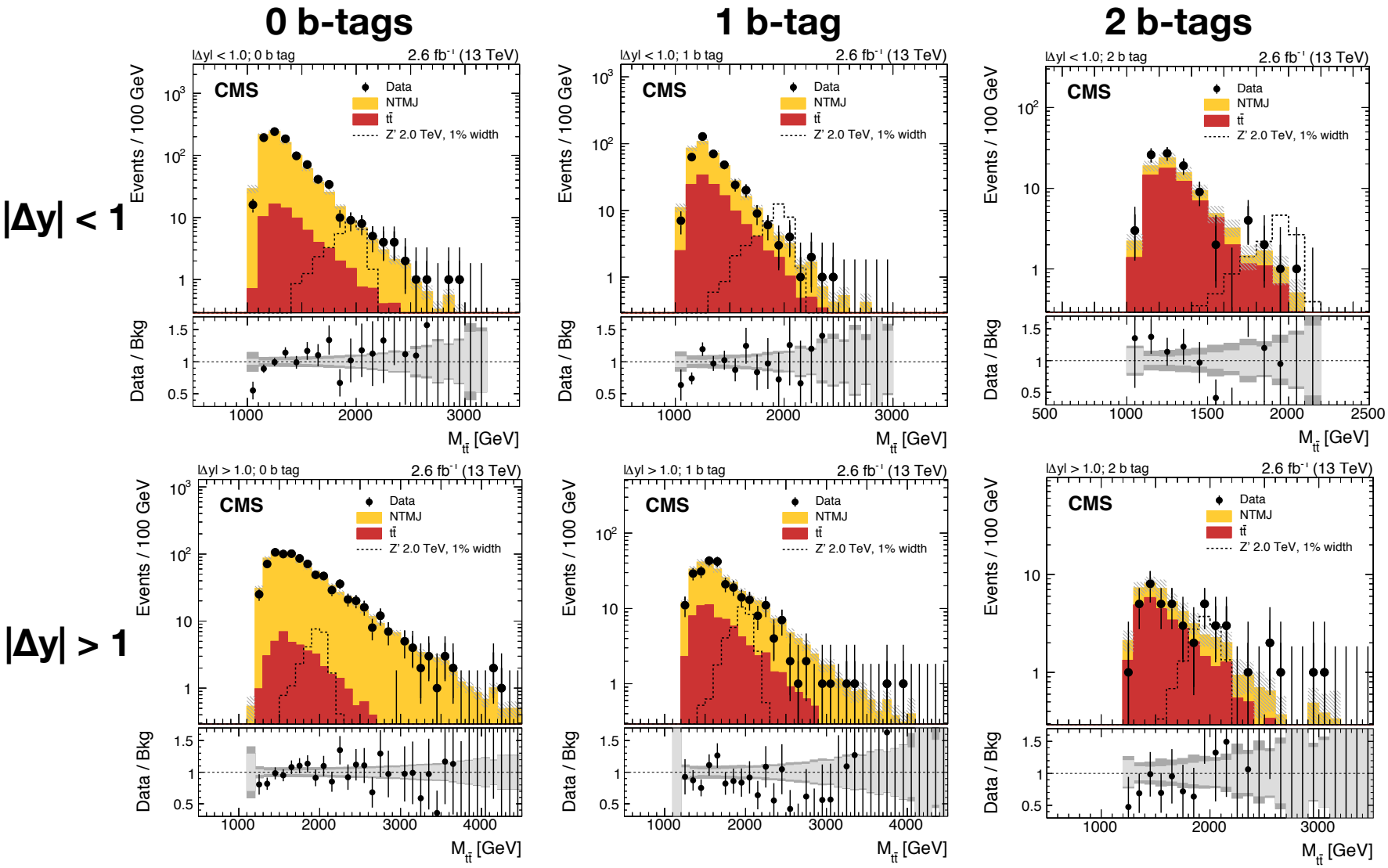
Signal Selection

- Trigger: OR of HT, single jet, and substructure triggers
  - Enables lower p_T, H_T cuts - improves low-mass sensitivity
- Preselection
  - $p_T > 400$ GeV, $|\Delta\Phi| > 2.1$
  - $H_T > 950$ GeV
- Signal Region: 2 top-tagged jets

	PUPPI top tag
Soft drop mass	105-210
N-subjettiness	$\tau_{32}^{puppi} < 0.65$

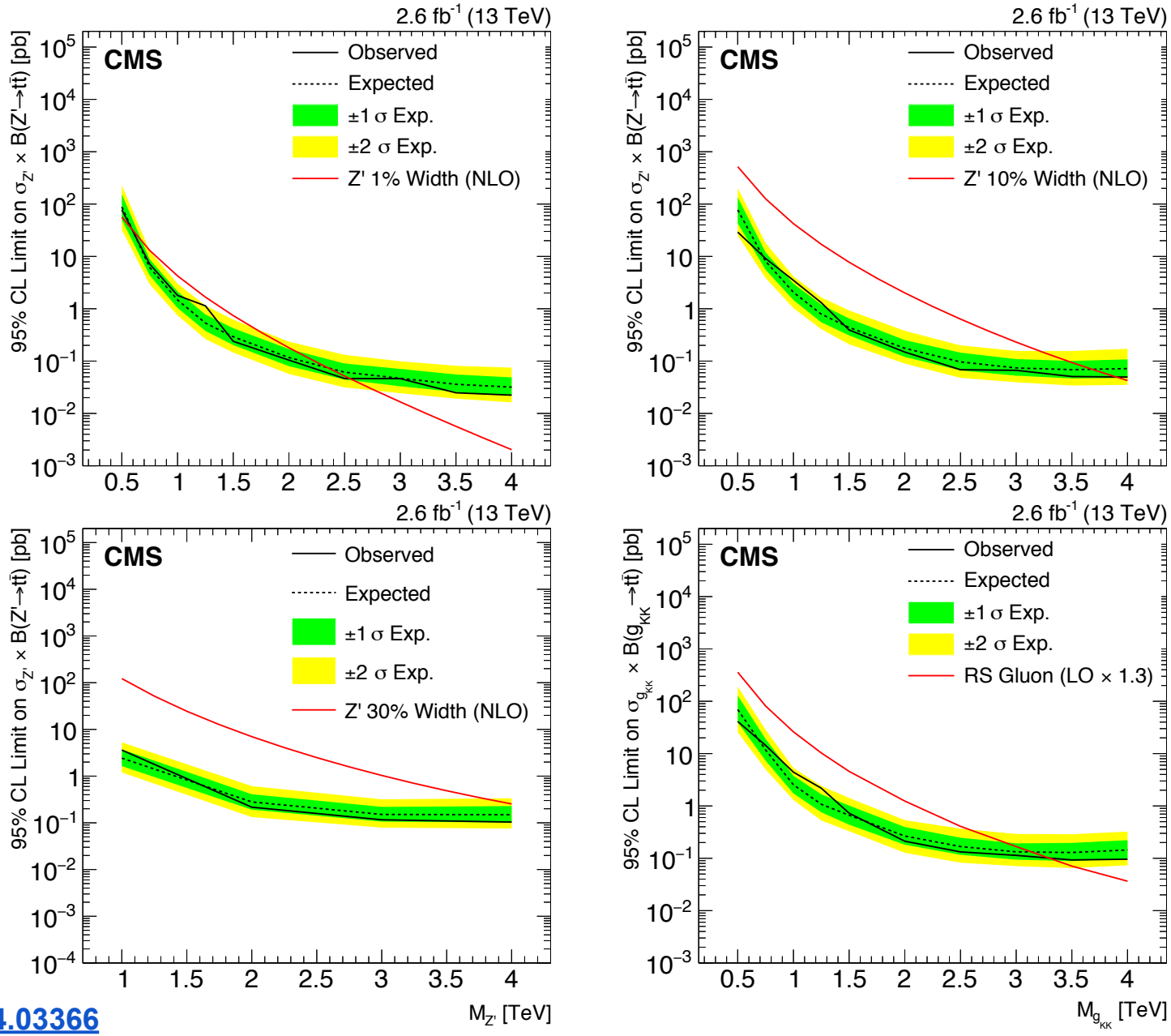


2015 Templates



arXiv:1704.03366

2015 Combined Limits

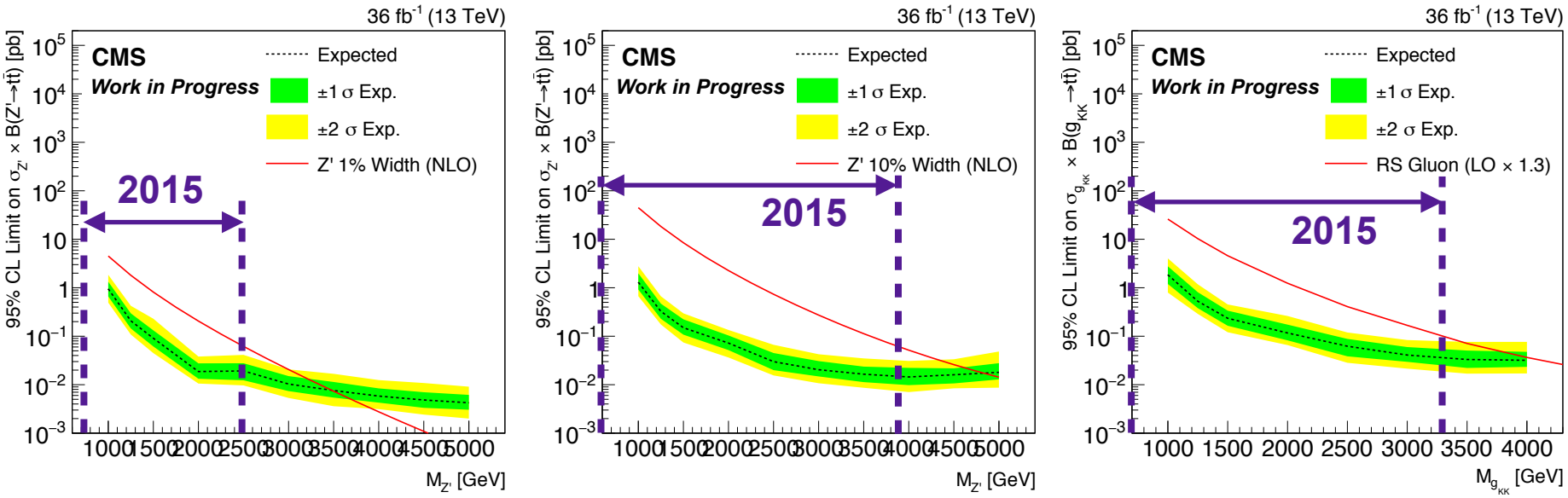


arXiv:1704.03366

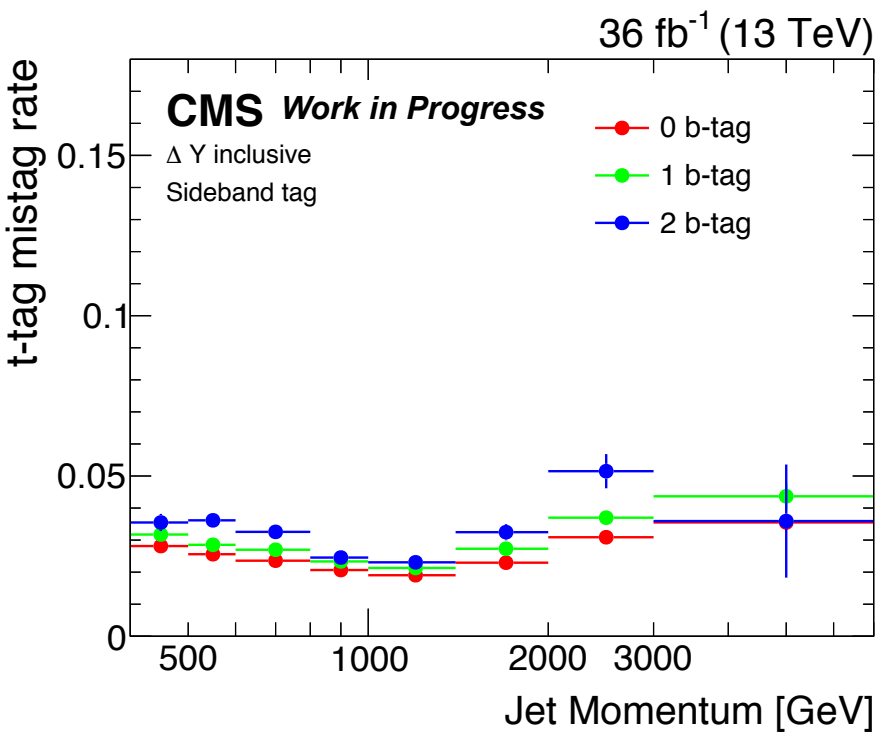
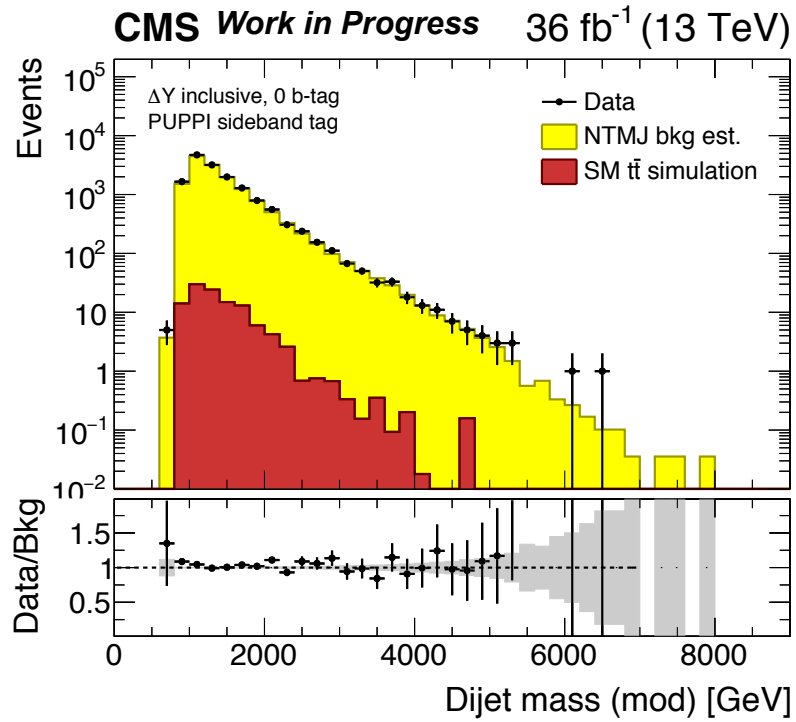
Systematics and Expected Limits

- Apply full set of systematics
- Expected limits: improvement on **2015 combined limits**
  - Combination of all-hadronic and semi-leptonic final states published in JHEP earlier this year

Systematic Uncertainty	Value	Type
b-Tag SF Uncertainty	$\pm 1\sigma$	Rate + Shape
CMS Top Tagger SF Uncertainty	+25% -9%	Rate
Jet Energy Corrections	$\pm 1\sigma$	Rate + Shape
Jet Energy Resolution	$\pm 1\sigma$	Rate + Shape
Luminosity	2.5%	Rate
PDF Uncertainty	$\pm 1\sigma$	Rate + Shape
Pileup Uncertainty	$\pm 1\sigma$	Rate + Shape
Top p_T Reweighting Uncertainty	$\pm 1\sigma$	Rate + Shape
$t\bar{t}$ Q^2 Uncertainty	$\pm 1\sigma$	Rate + Shape
$t\bar{t}$ Cross Section	8%	Rate
Mistag Rate Uncertainty	$\pm 1\sigma$	Rate + Shape
QCD Modified Mass Procedure	$\pm 1\sigma$	Rate + Shape



Background Estimate Closure Test

- Data closure using sideband
  - Define alternate tag definition
  - Perform full background estimate
  - Good agreement observed

	PUPPI top tag	Sideband tag (PUPPI)
Soft drop mass	105 - 210	40 -105
N-subjettiness	$\tau_{32}^{puppi} < 0.65$	$\tau_{32}^{puppi} < 0.65$