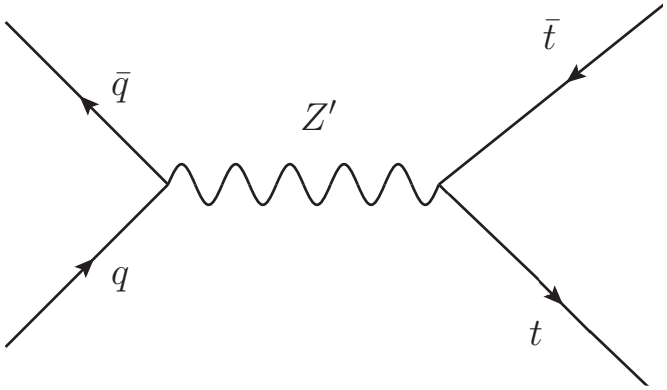
Searches for New Physics with Boosted Top Quarks at CMS

Christine McLean, University at Buffalo,
on behalf of the CMS Collaboration

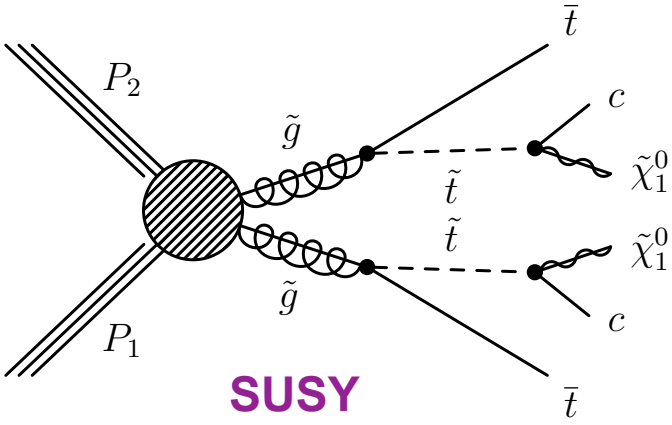
BOOST 2018
July 18, 2018

CMS Boosted Top Searches

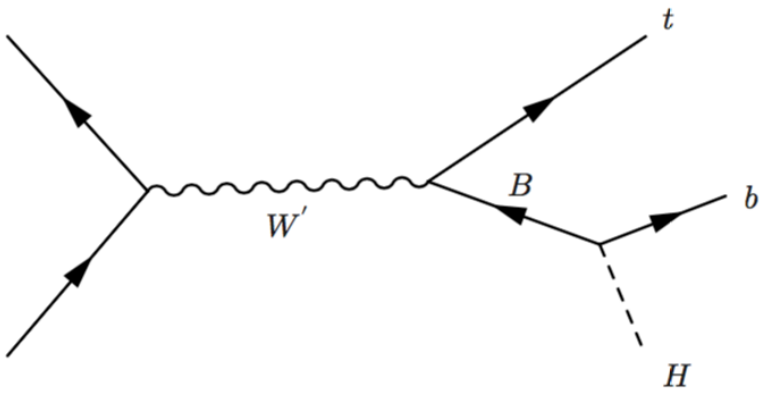
- Many theories predict new particles decaying to top quarks



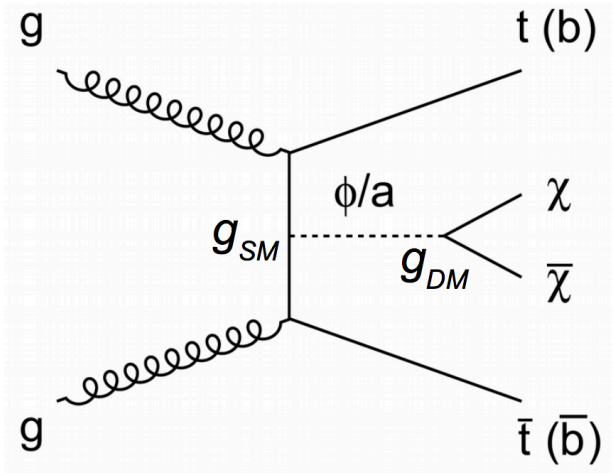
$t\bar{t}$ resonances: Z', RS model of extra dimensions



SUSY



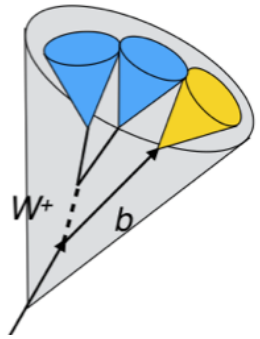
Vector-like Quarks (VLQs): composite and little Higgs models



DM + $t\bar{t}$: simplified models with new mediator and DM particle

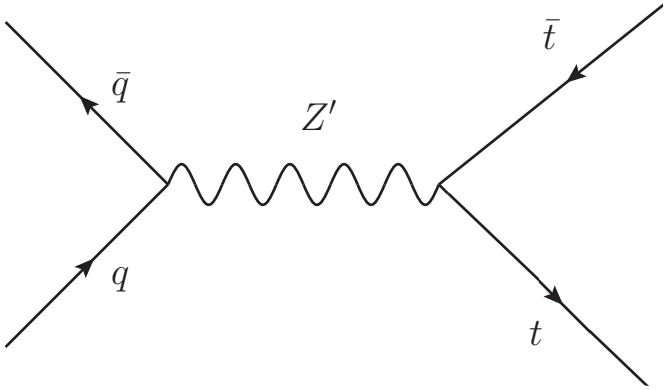
CMS top-tagging:

- mass grooming (soft drop)
- N-subjettiness
- subjet b-tagging
- MVA taggers

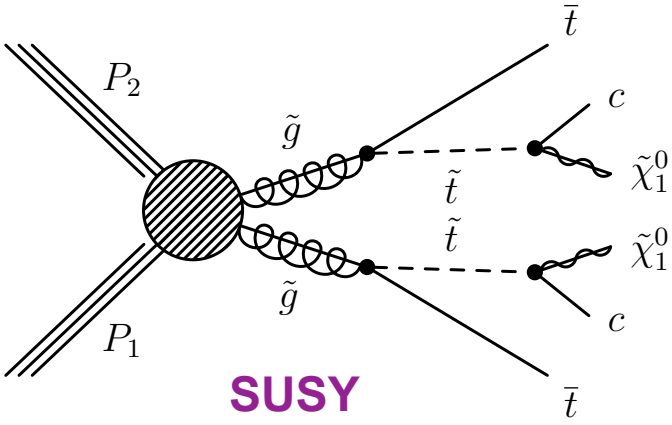


CMS Boosted Top Searches

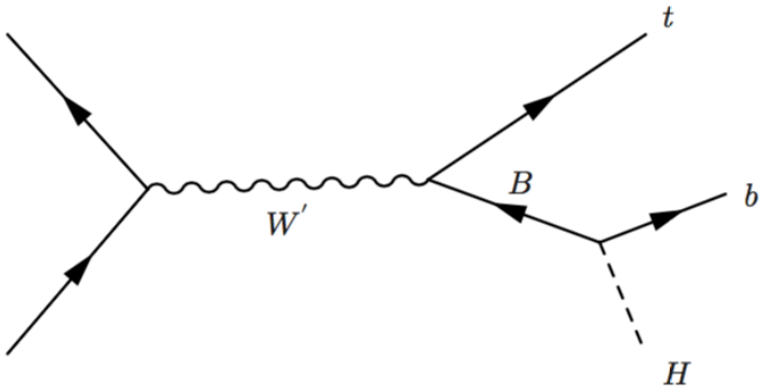
- Many theories predict new particles decaying to top quarks



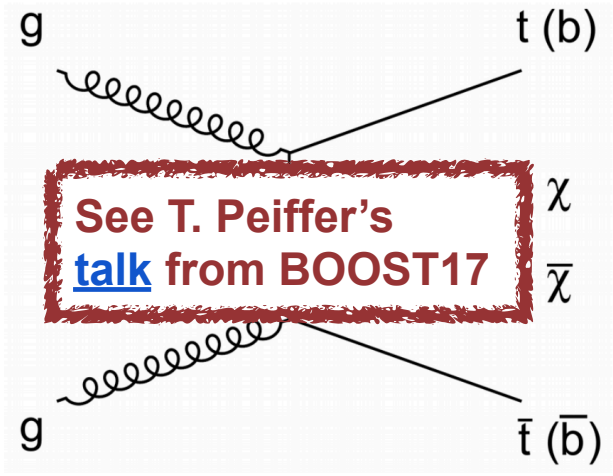
$t\bar{t}$ resonances: Z', RS model of extra dimensions



SUSY



Vector-like Quarks (VLQs): composite and little Higgs models

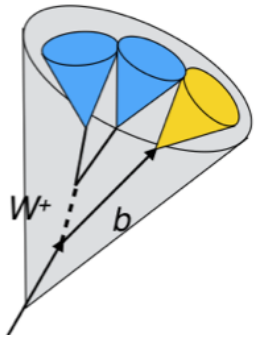


See T. Peiffer's talk from BOOST17

DM + $t\bar{t}$: simplified models with new mediator and DM particle

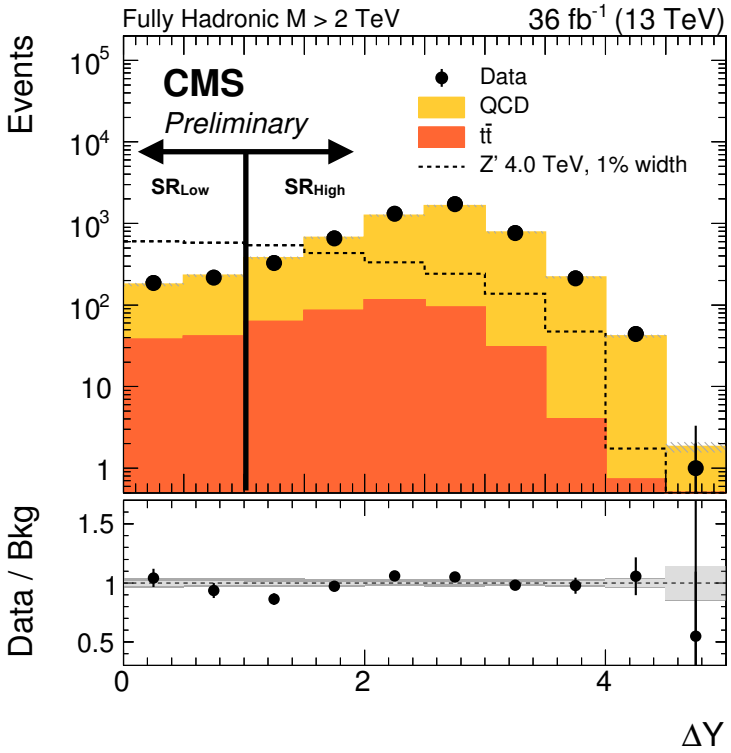
CMS top-tagging:

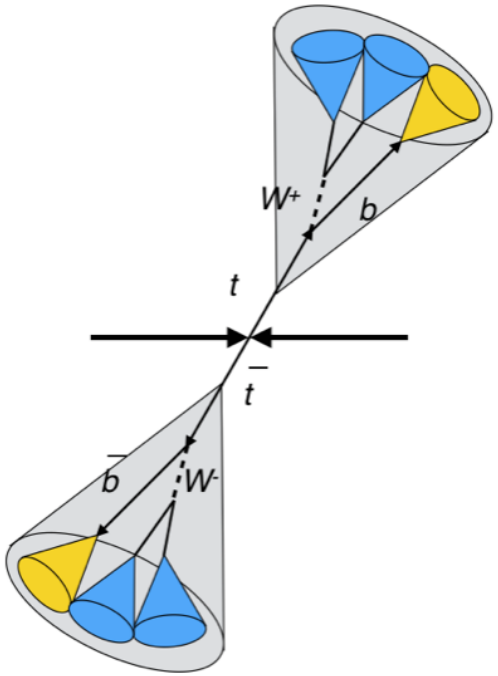
- mass grooming (soft drop)
- N-subjettiness
- subjet b-tagging
- MVA taggers



Heavy Resonances to $t\bar{t}$

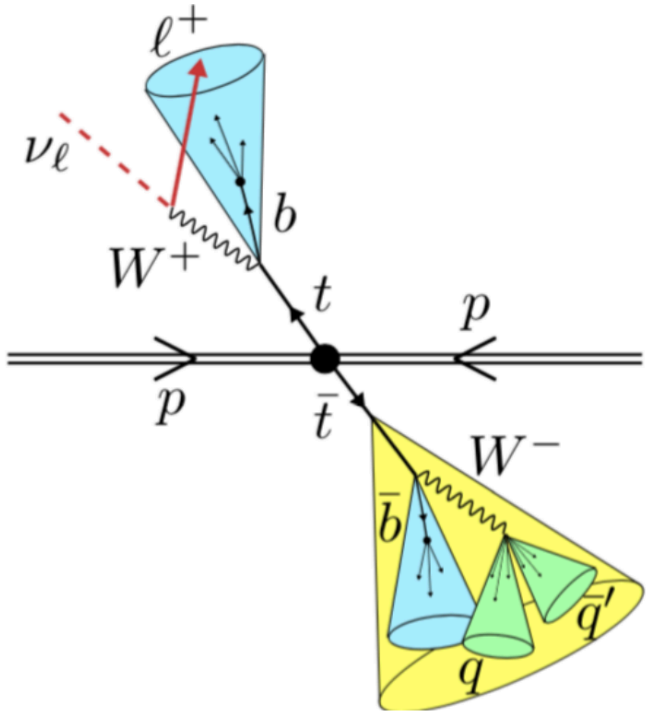
- Search for heavy spin-1 **RS KK gluon** & **Z'**(1%, 10%, 30% widths) decaying to $t\bar{t}$
  - **All-hadronic channel** - two back-to-back top-tagged **PUPPI** jets
    - Categorized by number of subjet b-tags (0,1,2) & $|\Delta y|_{tt}$
    - QCD background estimated from data, $t\bar{t}$ from MC





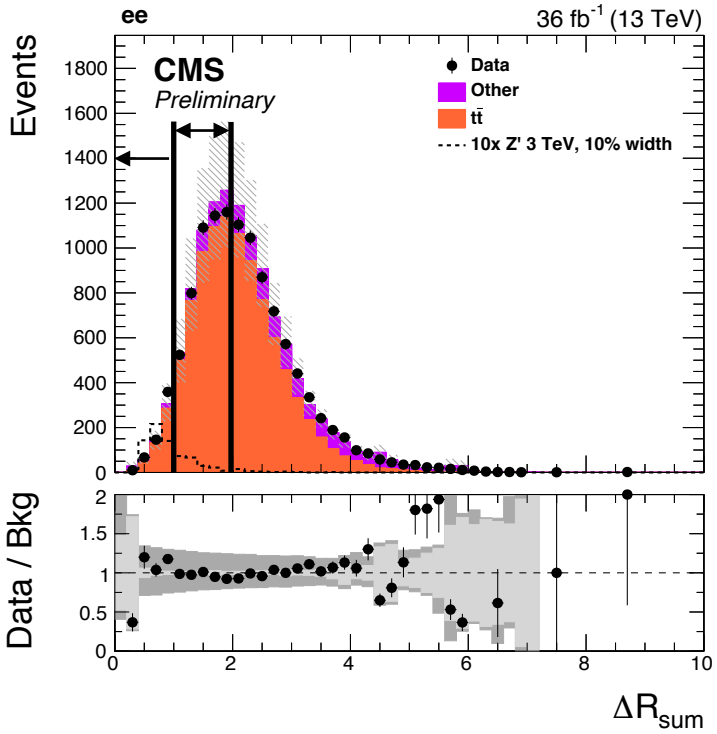
Heavy Resonances to $t\bar{t}$

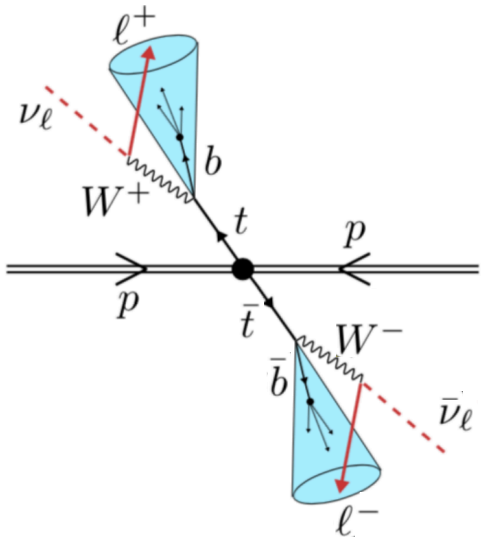
- Search for heavy spin-1 **RS KK gluon** & **Z'**(1%, 10%, 30% widths) decaying to $t\bar{t}$
  - All-hadronic channel
  - **Semileptonic channel:**
    - e+jets, μ+jets channels
    - Lepton 2D isolation cut
      - Suppresses QCD
    - X^2-based reconstruction of top system
    - Categorized by number of top tags (0,1)
    - Background estimated from Monte Carlo
      - **W + jets suppressed by BDT**



Heavy Resonances to $t\bar{t}$

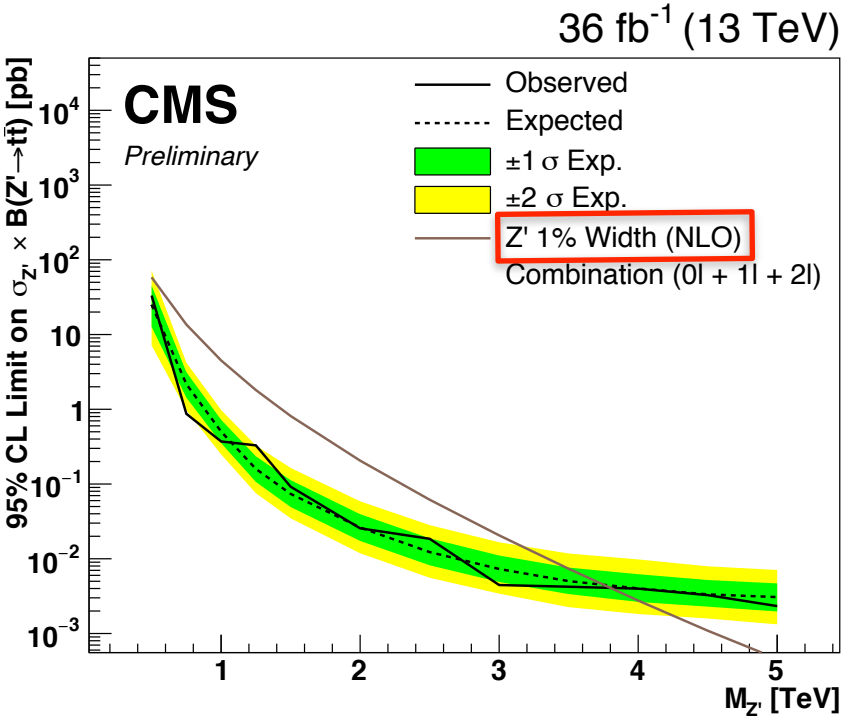
- Search for heavy spin-1 **RS KK gluon** & **Z'**(1%, 10%, 30% widths) decaying to $t\bar{t}$
  - All-hadronic channel
  - Semileptonic channel
- **Dileptonic channel**
  - ee, μμ channels
  - Lepton 2D isolation cut
  - At least one b-tagged jet
  - Characterize signal region by lepton/jet separation
  - Background: $t\bar{t}$ (MC)

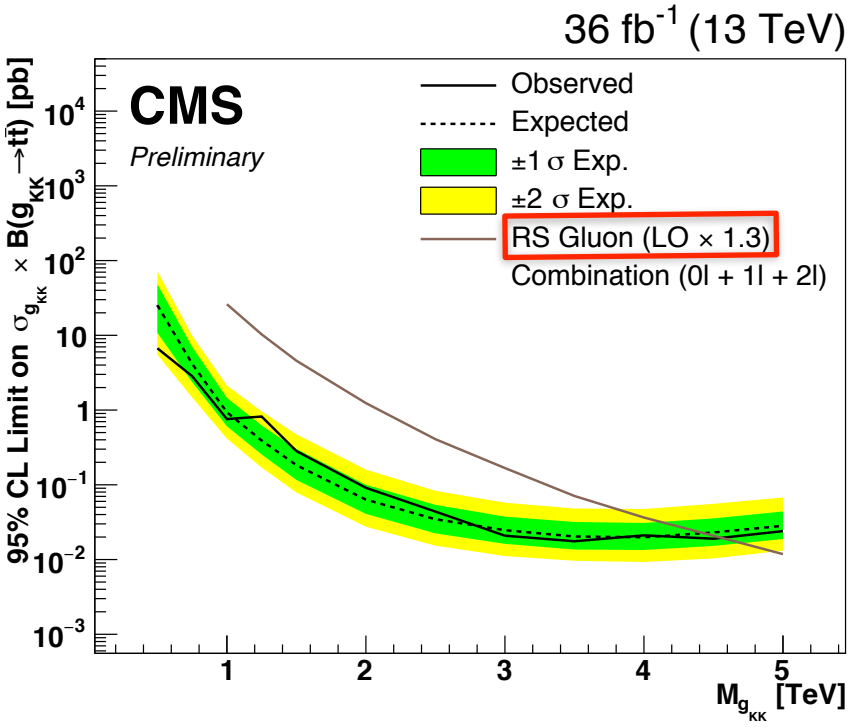




Heavy Resonances to $t\bar{t}$: Results

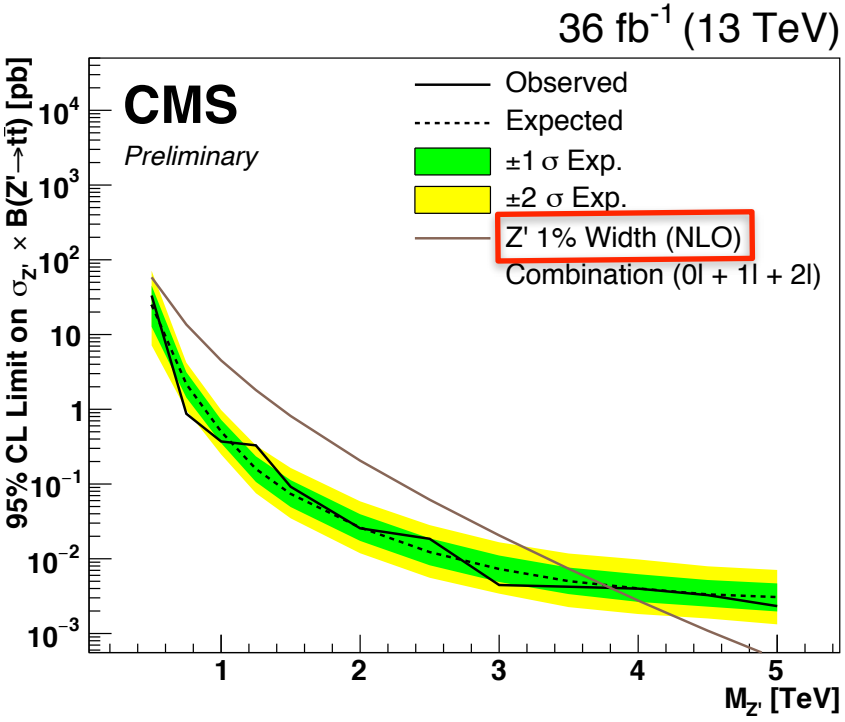
- Discriminating variables:
  - $M_{t\bar{t}}$ (all-hadronic, semileptonic), S_T (dileptonic)

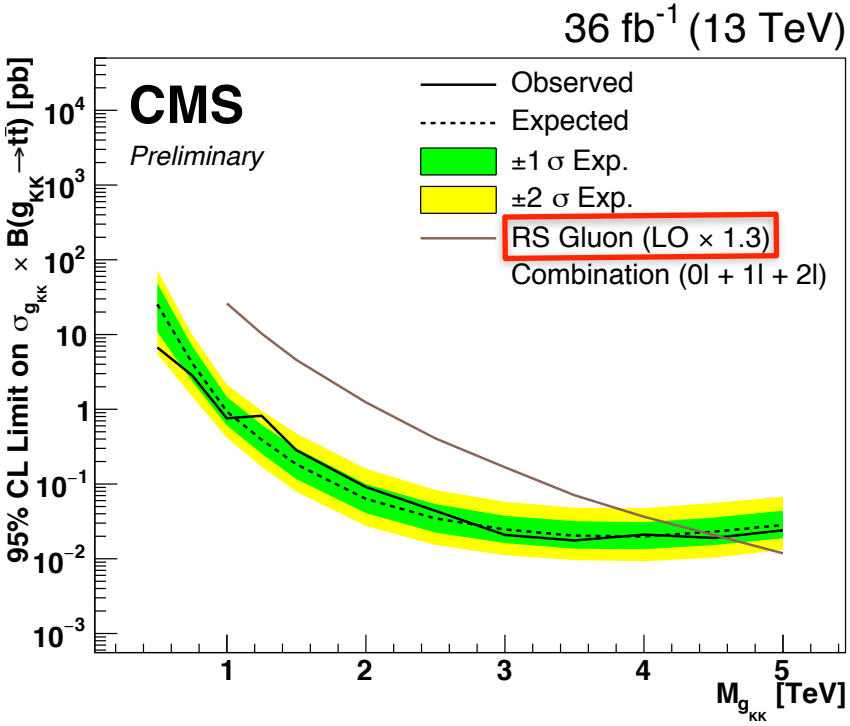




Heavy Resonances to $t\bar{t}$: Results

- Discriminating variables:
  - $M_{t\bar{t}}$ (all-hadronic, semileptonic), S_T (dileptonic)
- **Most recent CMS search: large gains in sensitivity!**





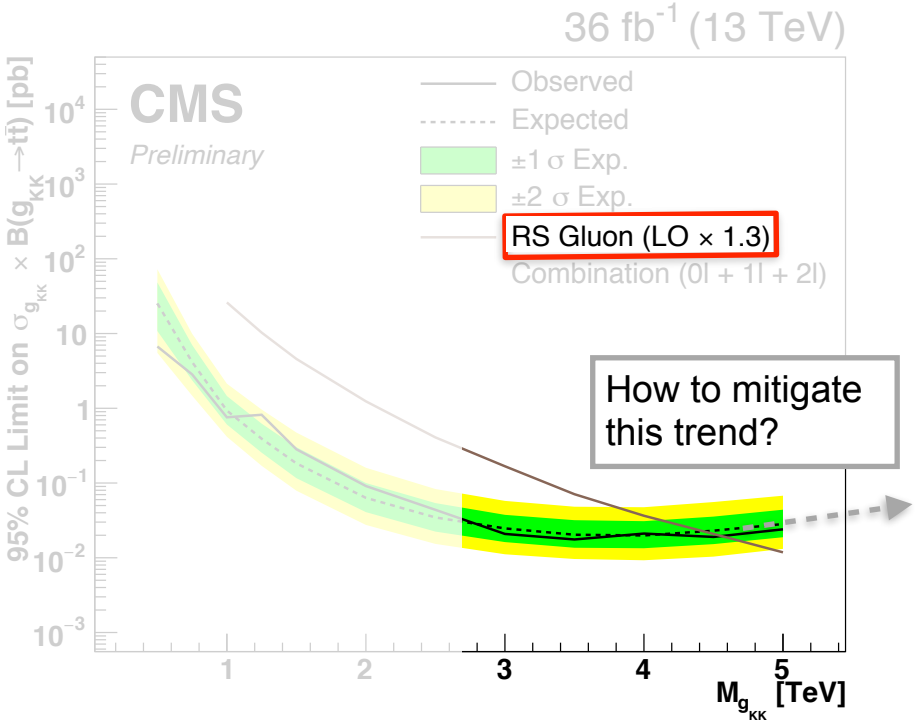
CMS Improvement (Combination)

1804.10823 ATLAS CMS Semileptonic Sensitivity ATLAS CMS

CMS Dileptonic, All-Hadronic Sensitivity

Heavy Resonances to $t\bar{t}$: Results

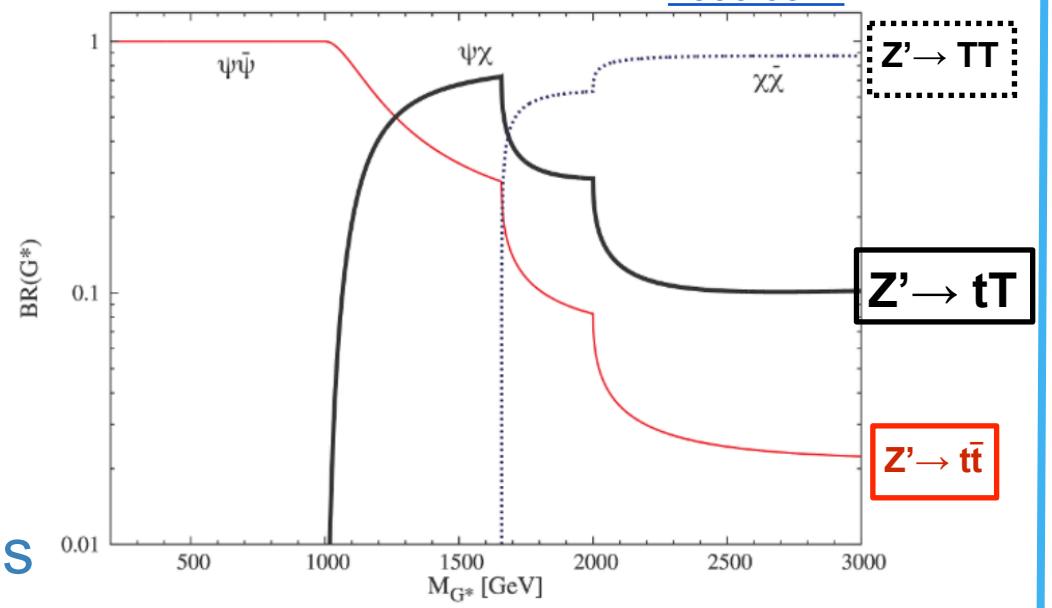
- Discriminating variables:
  - **$M_{t\bar{t}}$** (all-hadronic, semileptonic), **S_T** (dileptonic)
- Most recent CMS search: large gains in sensitivity!
- **Going forward: Analysis is luminosity limited**



Z' → tT (Lepton + jets)

- BSM models motivate existence of both Z' and vector-like quarks (VLQ)
  - e.g. composite Higgs, extra dimensions
  - **VLQ:** heavy spin-1/2 "quark"; left- and right-handed coupling to CCs

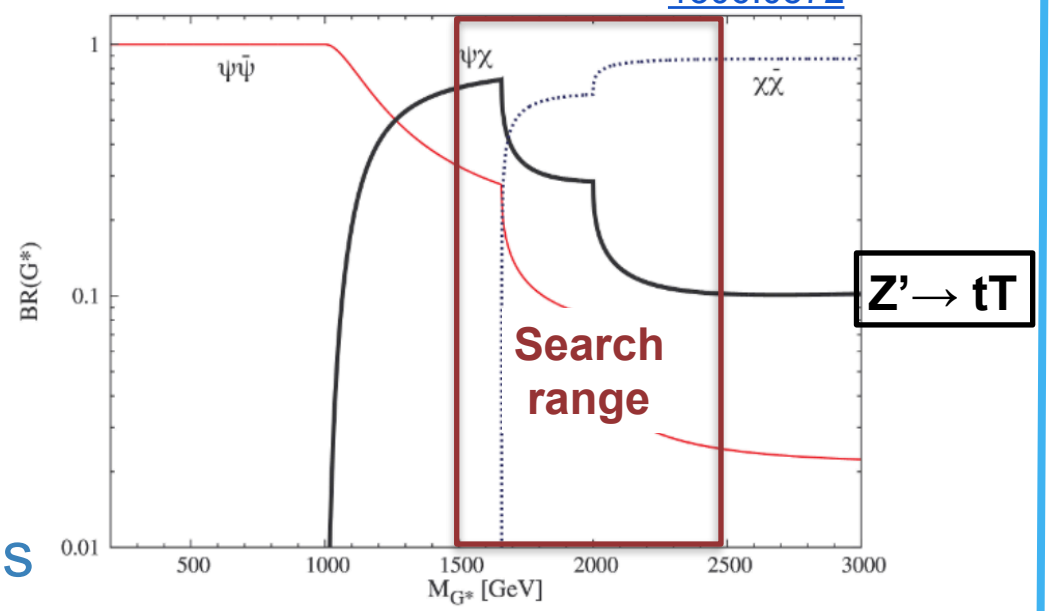
1306.0572



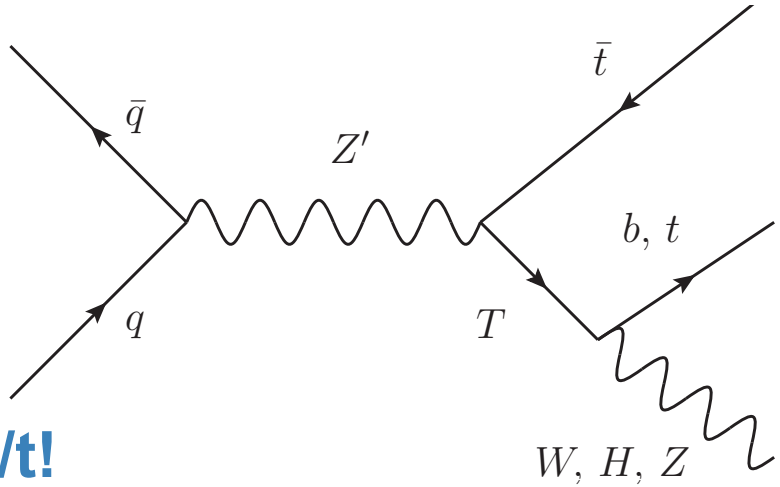
Z' → tT (Lepton + jets)

- BSM models motivate existence of both Z' and vector-like quarks (VLQ)
  - e.g. composite Higgs, extra dimensions
  - **VLQ:** heavy spin-1/2 "quark"; left- and right-handed coupling to CCs

1306.0572

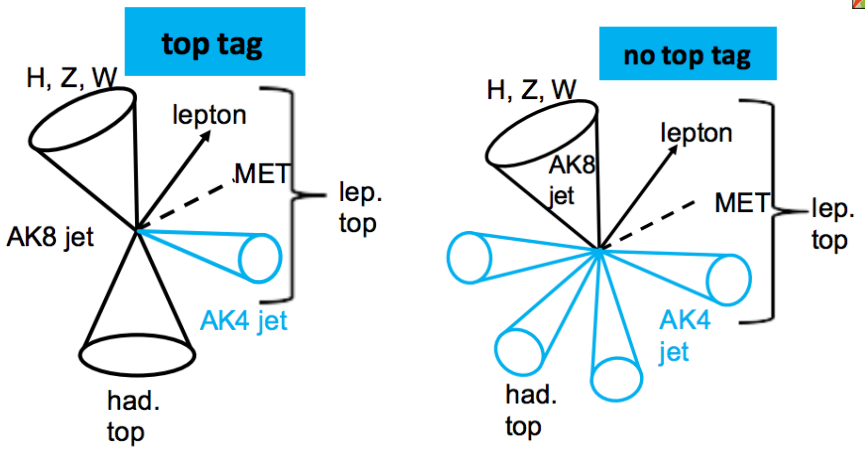


- This search:
  - Considers all three T decay modes
    - Optimized for Ht/Zt
  - Semileptonic
  - **Looks for boosted H/Z/W/t!**



Z' → tT: Event Selection

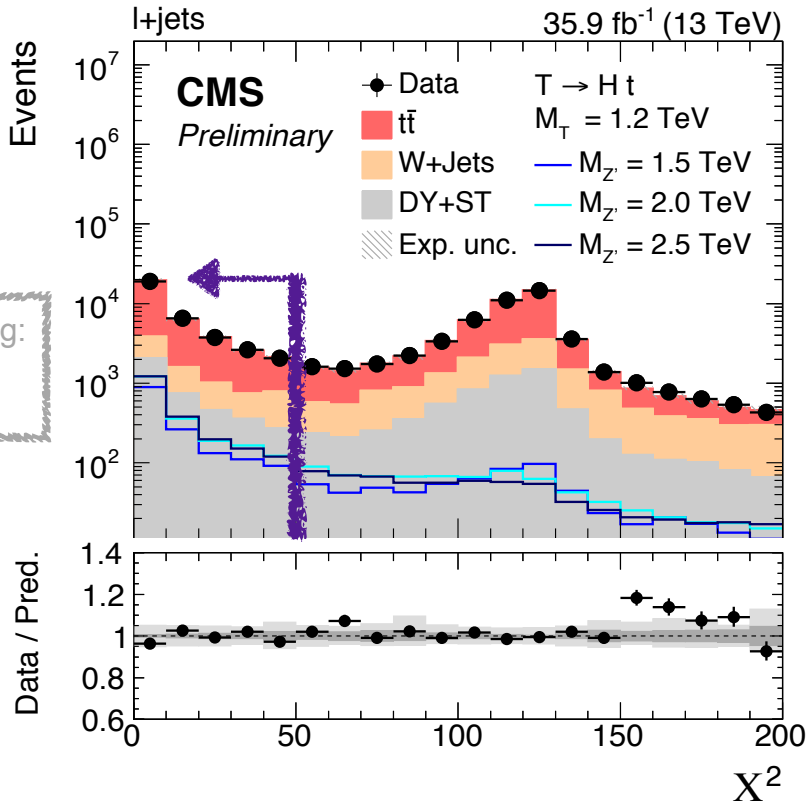
- Preselection:
  - exactly one e or μ
  - ≥ 1 AK4, ≥ 1 AK8 jet
  - MET
  - Non-isolated lepton



- Signal region:
  - Categorize based on number of boosted jet (Z/W/H/t) tags
    - Softdrop mass
    - τ_{32} (t), τ_{21} (Z/W)
    - subjet b-tags (H)

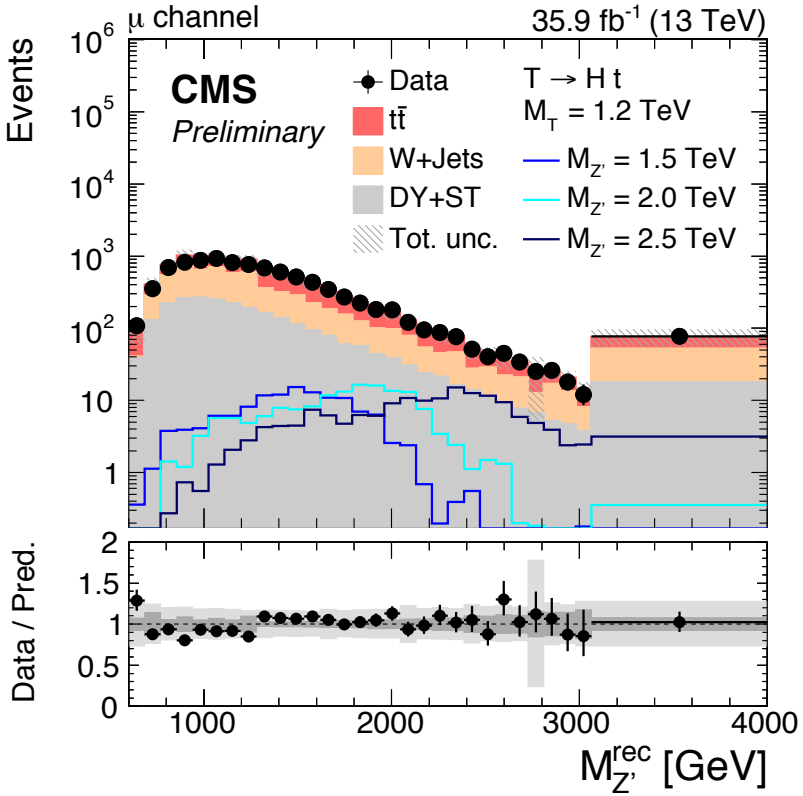
CMS V-tagging:
C. Suarez
T. Aarrestad

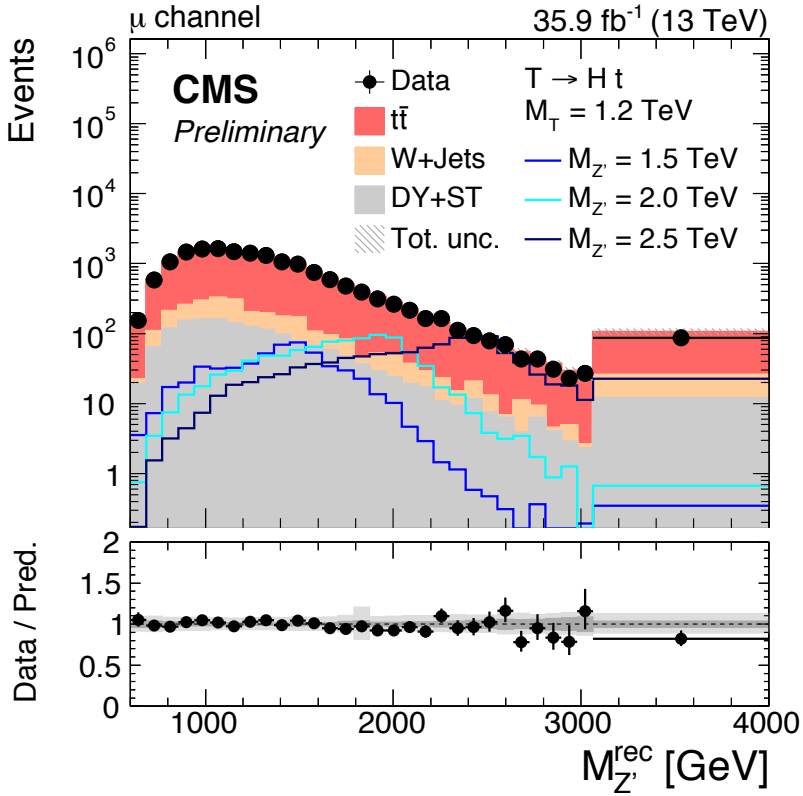
- Perform X^2 reconstruction of the system



Z' → tT: Backgrounds

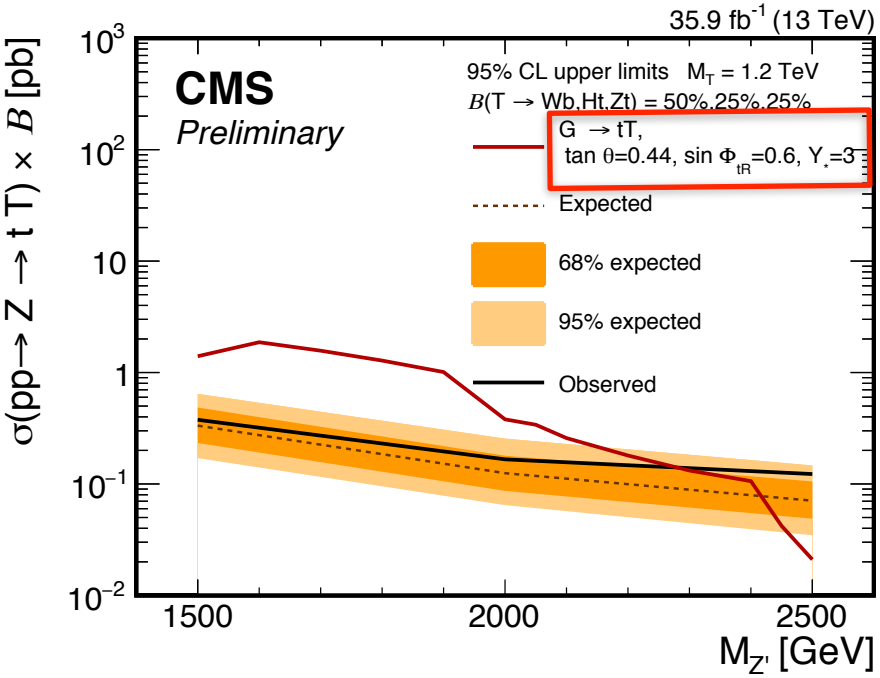
- Take background from simulation
- Fit control regions to simultaneously constrain backgrounds
  - Mass sideband: M_{SD} < 60 GeV & 150 GeV > M_{SD}
  - X^2 < 50
  - 0 b-tags: W+jets; 1 b-tag: $t\bar{t}$

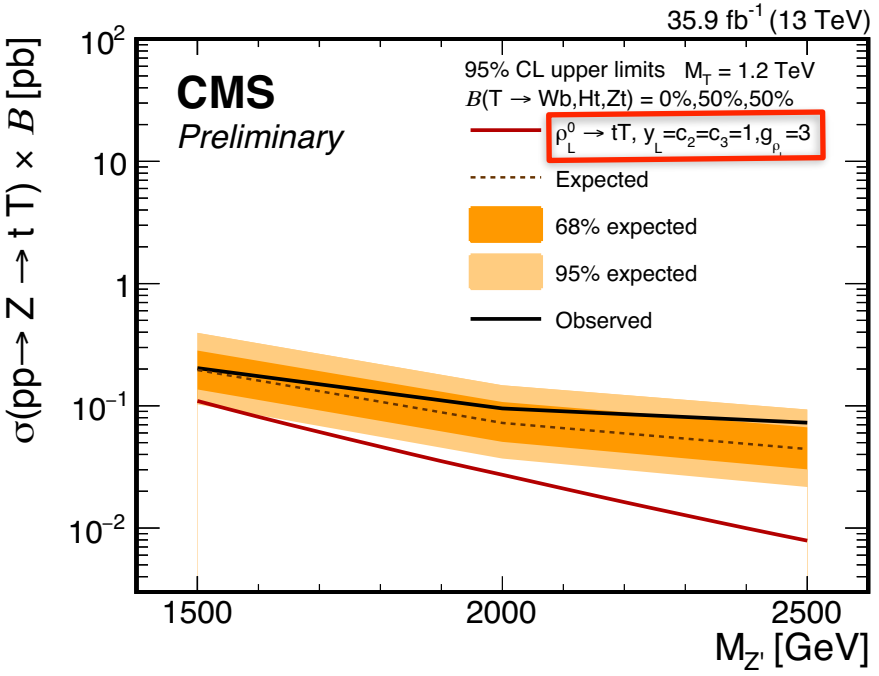




Z' → tT: Results

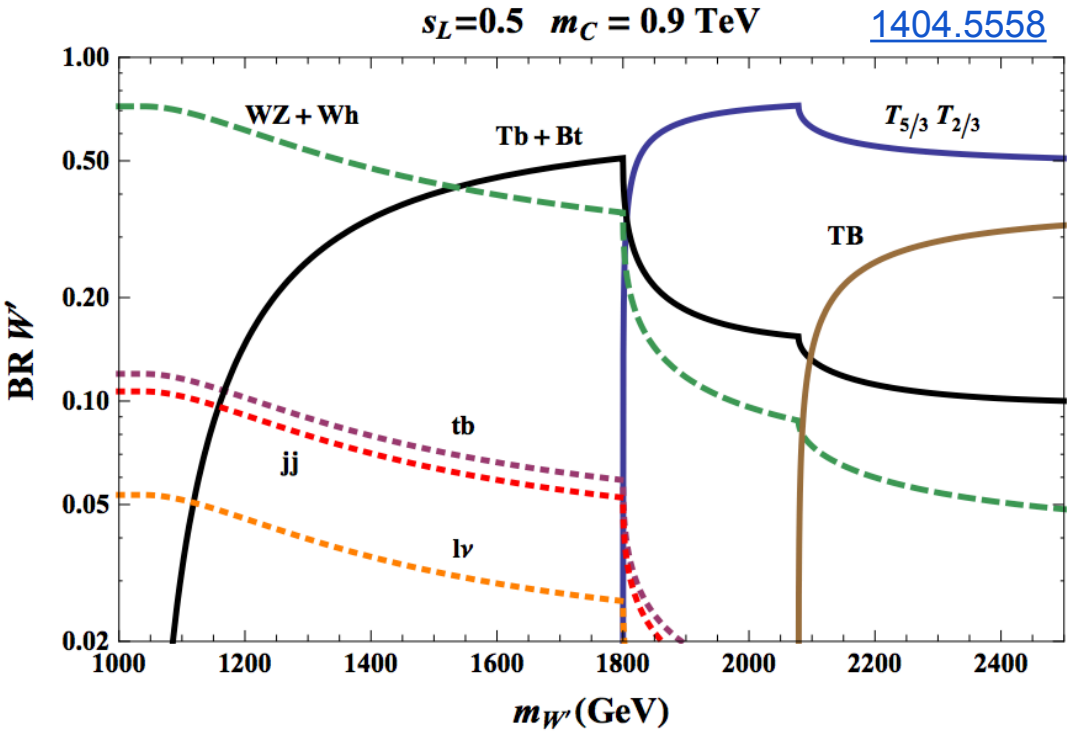
- Limits set as a function of Z' mass
  - **First results for Ht channel!**
  - Exceeds sensitivity of previous searches





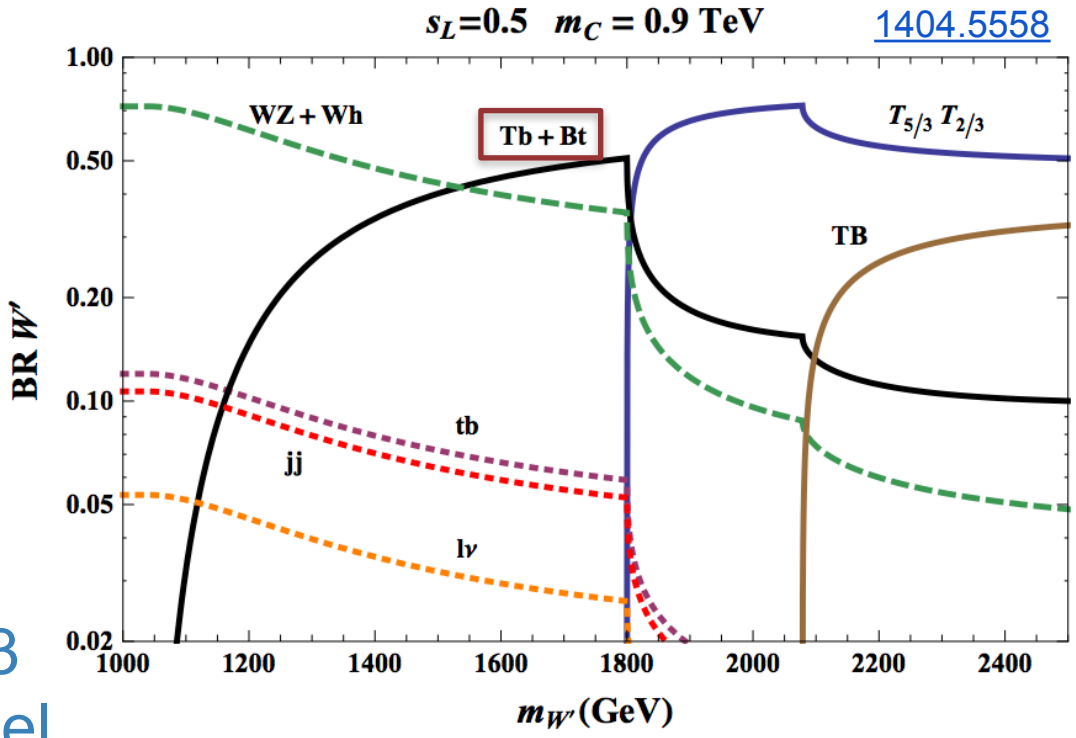
W' → Tb/tB (all-hadronic)

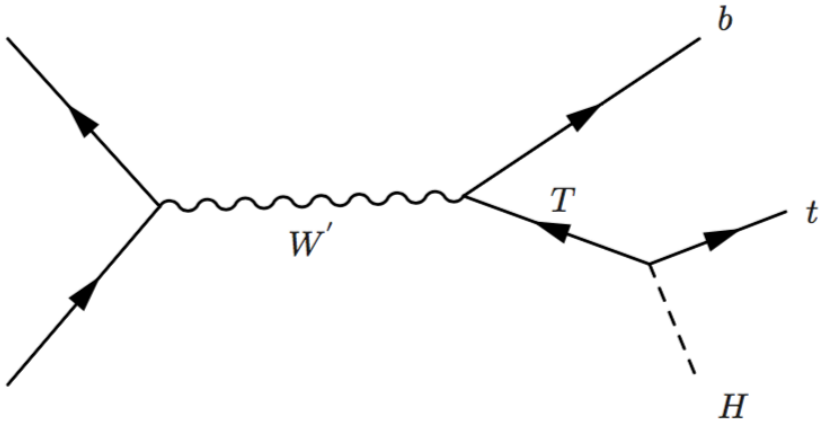
- **W':** also predicted by e.g. little Higgs and extradimensional models
  - Many possible decay modes

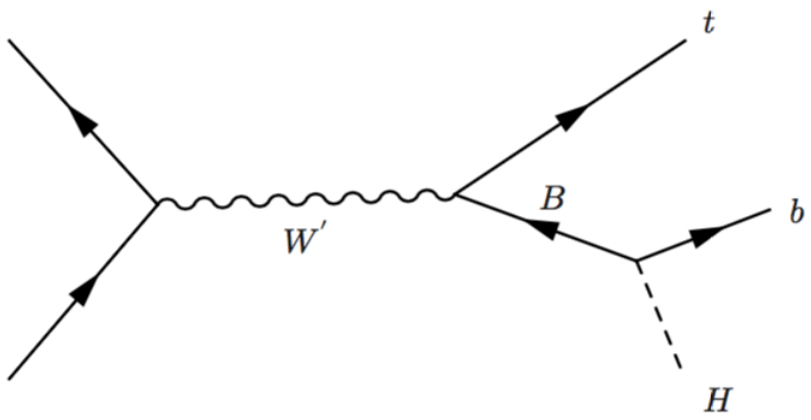


W' → Tb/tB (all-hadronic)

- **W':** also predicted by e.g. little Higgs and extradimensional models
  - Many possible decay modes
- **W' → Tb+tB model**
  - T → tH, B → bH
  - s_L = 0.5, cot(θ_2) = 3
  - All-hadronic channel







W' → Tb/tB: Signal Selection

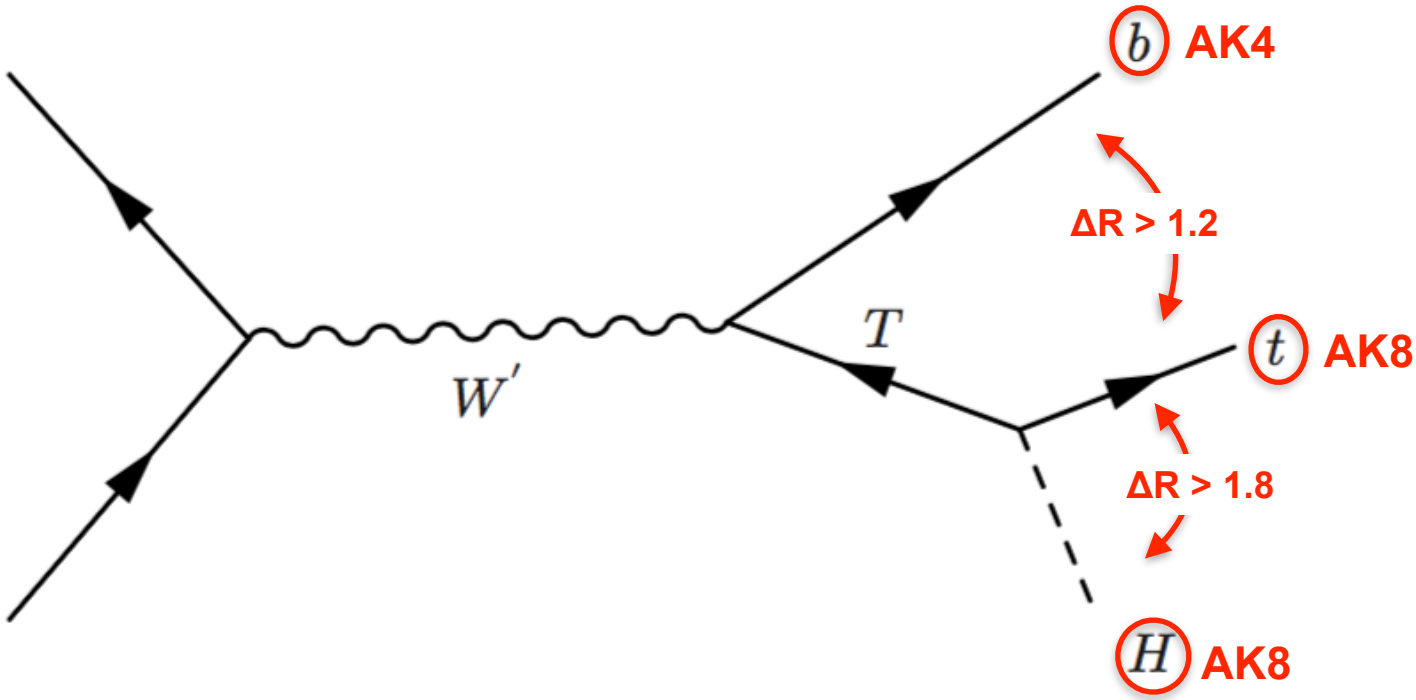
1. Higgs jet tag
   - Double b-tagger

T. Aarrestad

2. Top jet tag
   - with subjet b-tag

3. b jet tag

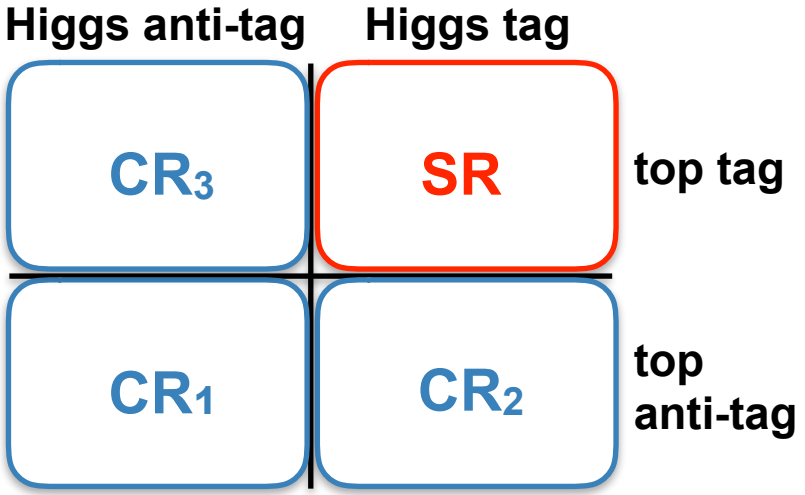
4. Combine the three jets' invariant mass to get the W'

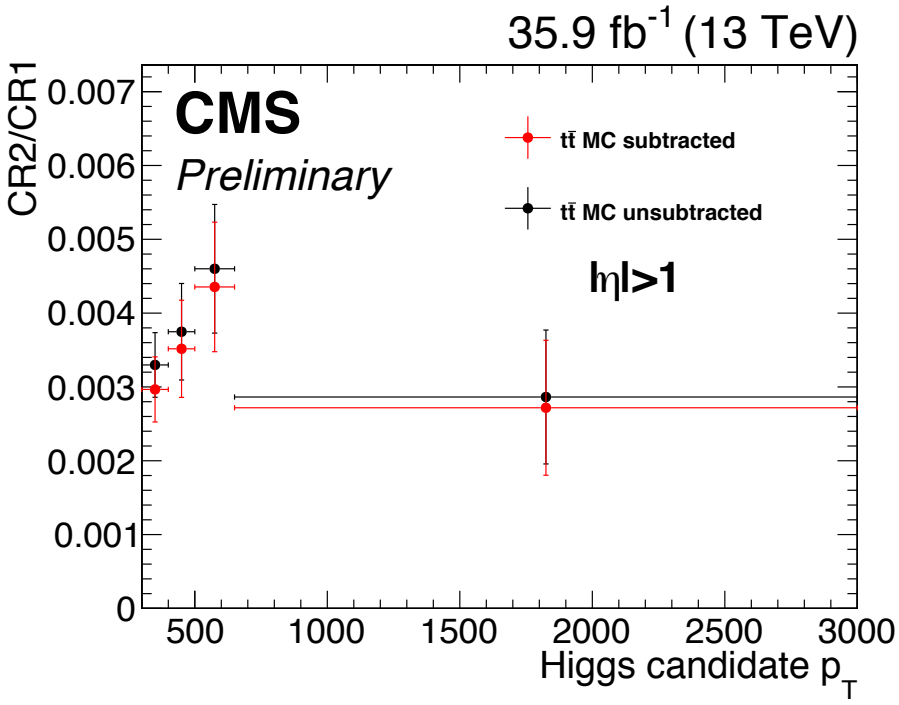


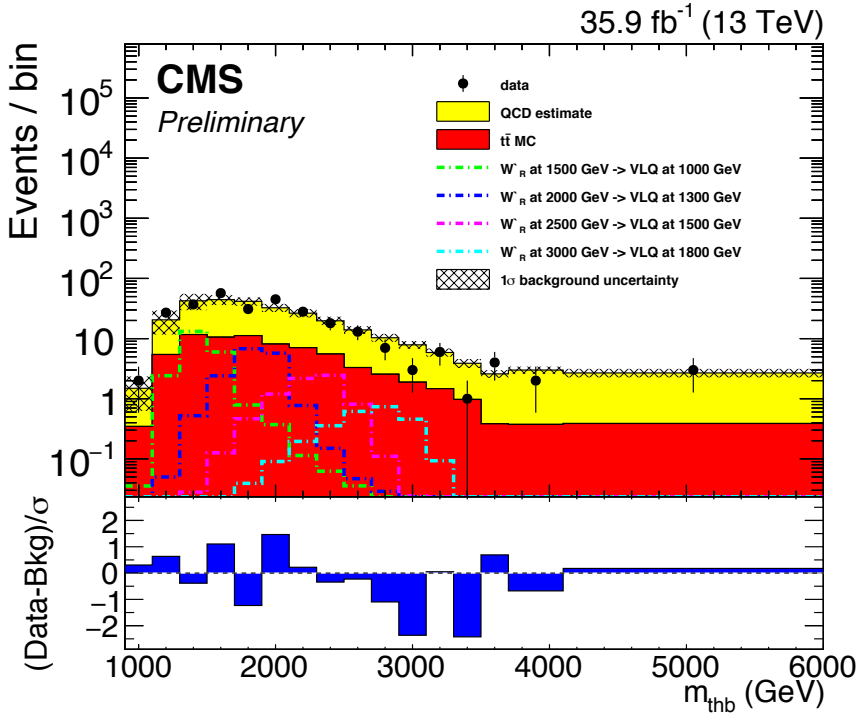
W' → Tb/tB: Background Estimation

- **$t\bar{t}$:** estimated from MC
- **QCD:** estimate from data using transfer function
  - Define Higgs, top anti-tag regions

$$SR = CR_3 \times \frac{CR_2}{CR_1}$$

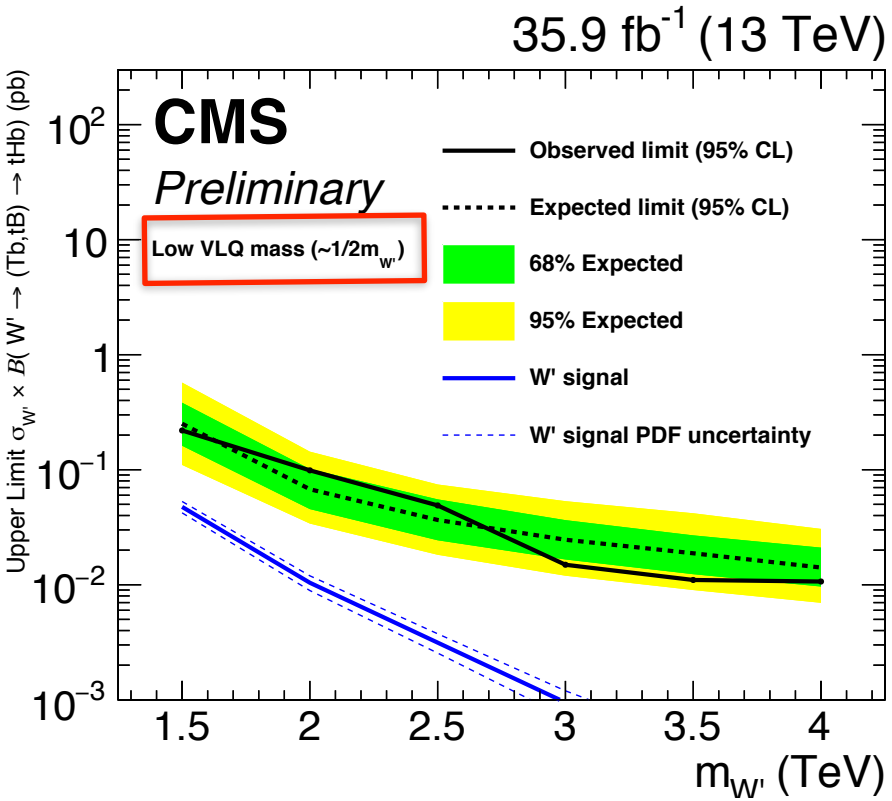


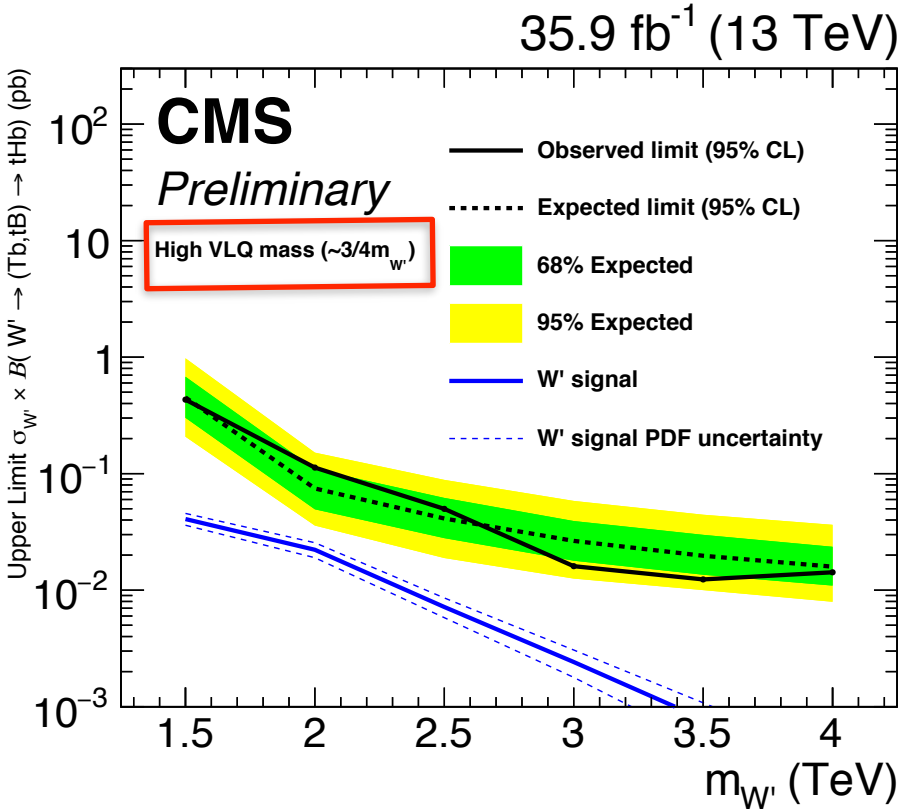




W' → Tb/tB: Results

- Limits set for several VLQ mass ranges
- **New for BOOST!**
- No mass exclusion limits set yet
  - Could benefit from new, advanced taggers



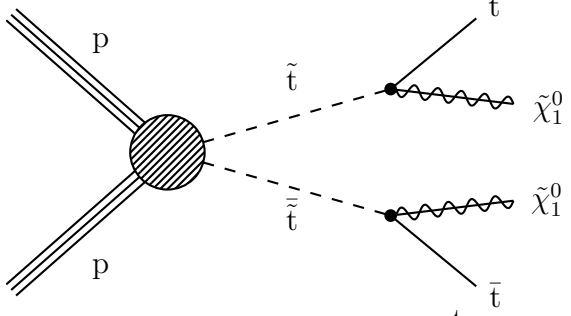


10.1103/PhysRevD.97.012007

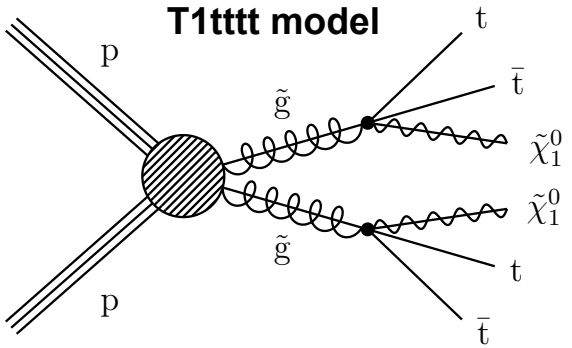
SUSY with Top Tagging

- SUSY production can lead to top quarks in the final state
- This analysis: consider five SUSY models of direct or gluino-mediated top squark production
  - Look for hadronically decaying tops over large p_T range

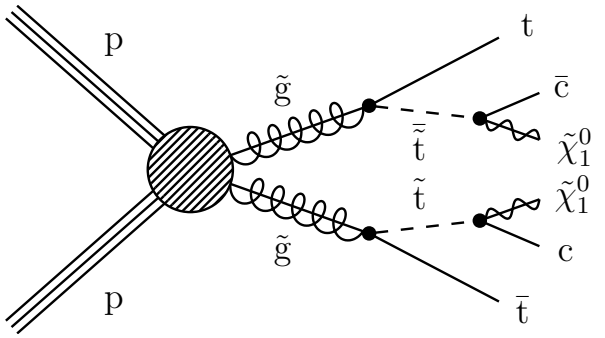
T2tt model



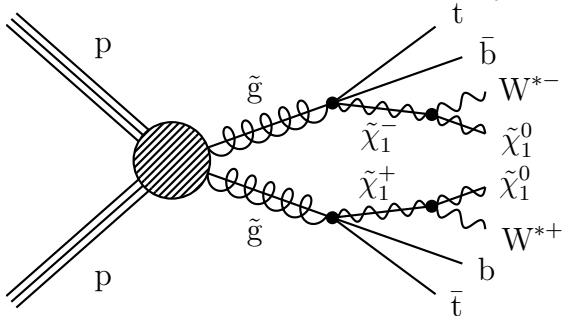
T1tttt model



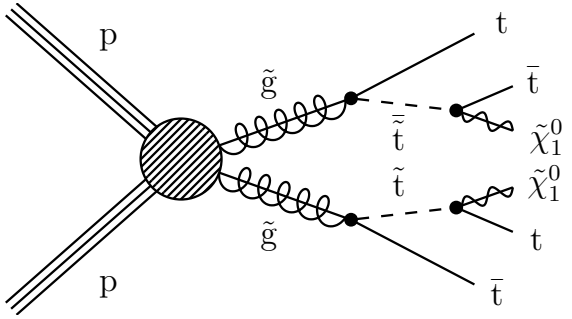
T5ttcc model



T1ttbb model



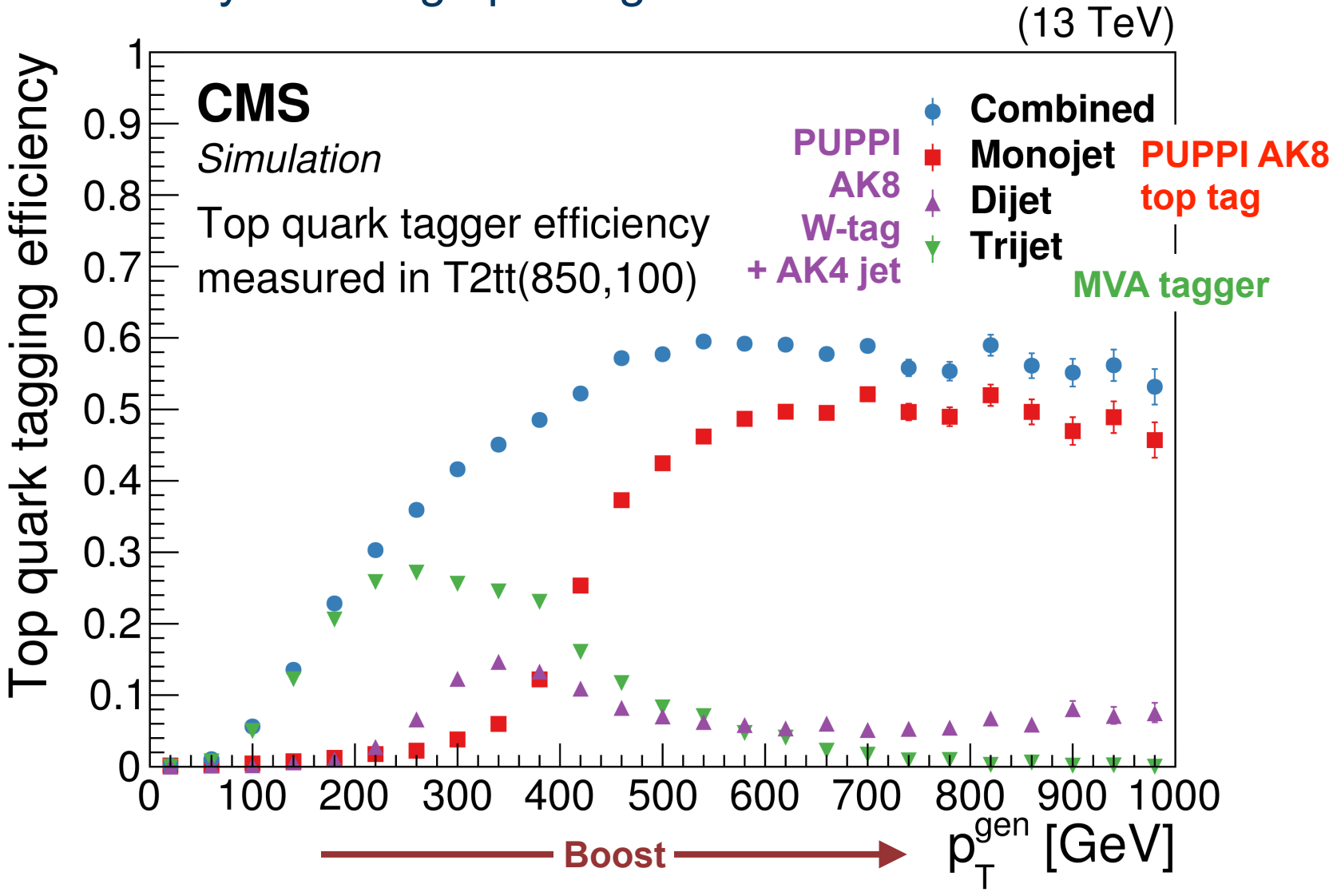
T5tttt model



10.1103/PhysRevD.97.012007

SUSY Top Tagging Method

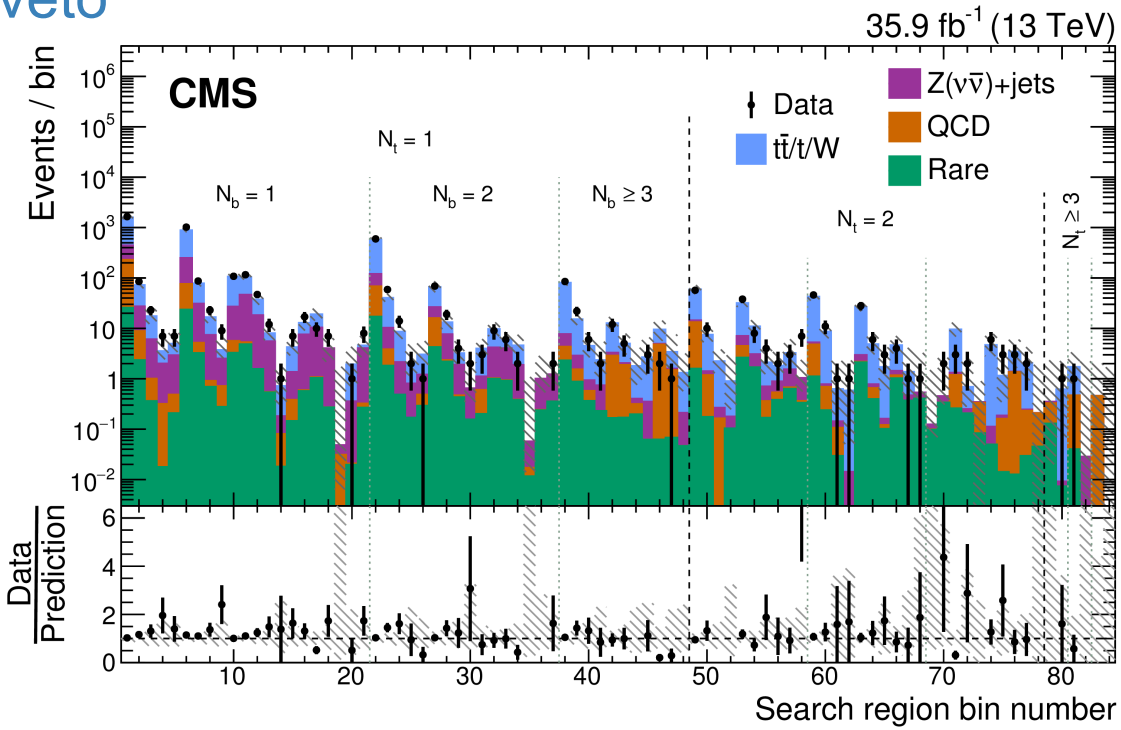
Combine fully merged, partially merged and resolved taggers to gain efficiency over large p_T range



10.1103/PhysRevD.97.012007

SUSY with Top Tagging: Search Method

- Preselection:
  - ≥ 4 AK4 jets
  - MET > 250 GeV
  - H_T > 300 GeV
  - $\Delta\Phi(j_{1,2,3},\text{MET})$
  - $N_b \geq 1$, $N_t \geq 1$
  - M_{T2} > 200 GeV
  - lepton veto
- Search binned in H_T,M_{T2} and MET
  - Categorized by N_b and N_t
  - 84 total categories
- Backgrounds
  - tt/W+jets, Z to invisibles, QCD estimated using data
  - Others use MC only

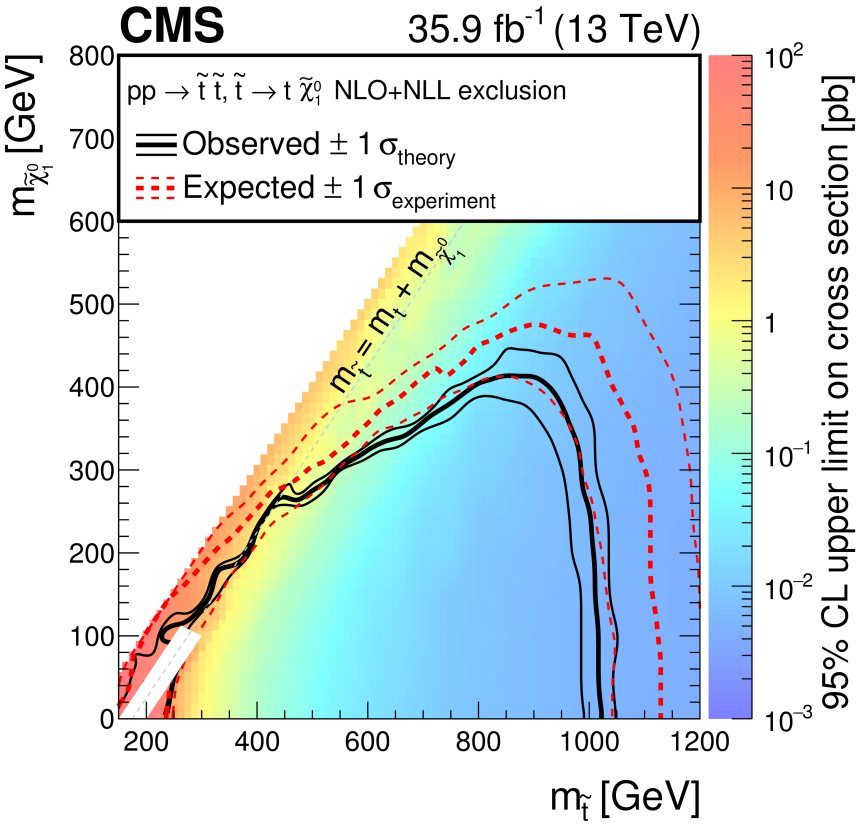


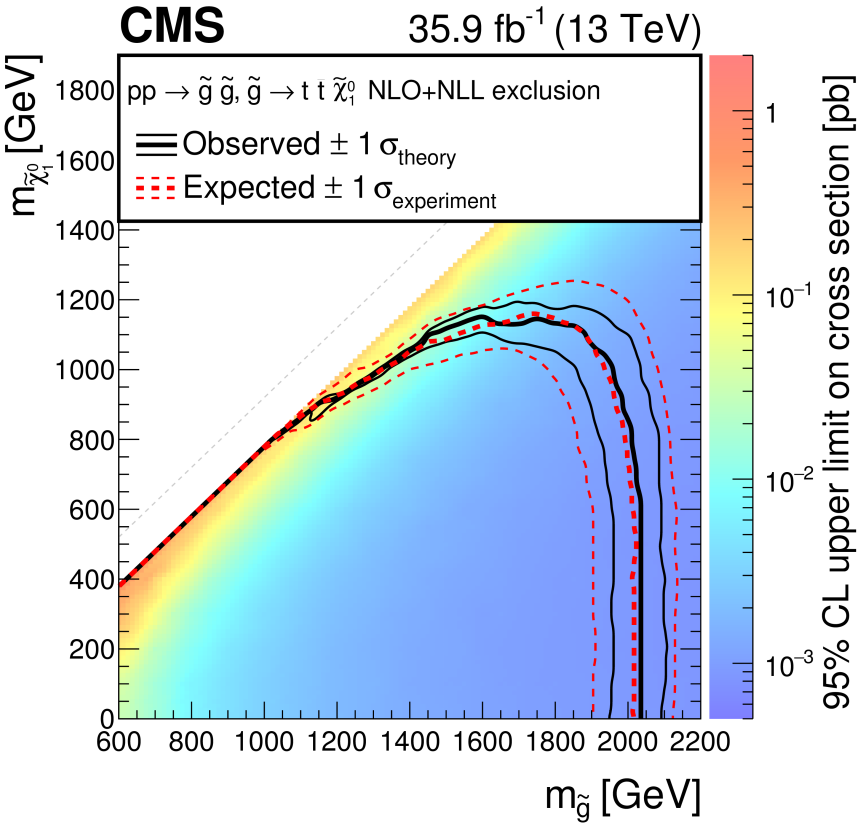
10.1103/PhysRevD.97.012007

SUSY with Top Tagging: Limits

- Evaluated limits for several SUSY models
  - Direct and gluino-mediated top squark production

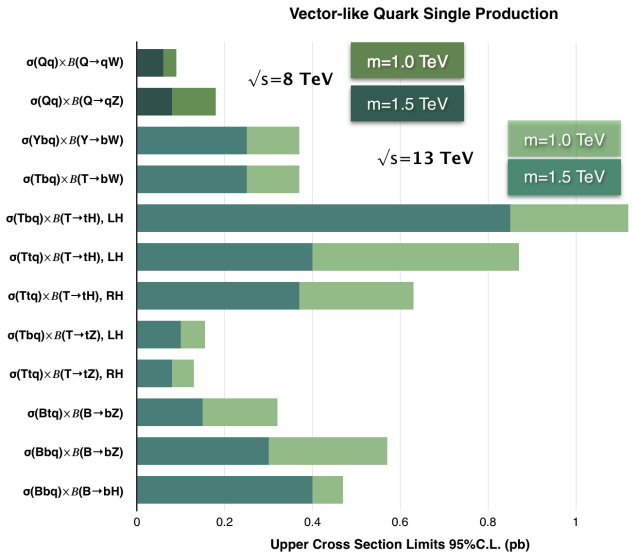
- Exclude gluinos, neutralinos, and top squarks up to 2040, 1150 and 1020 GeV

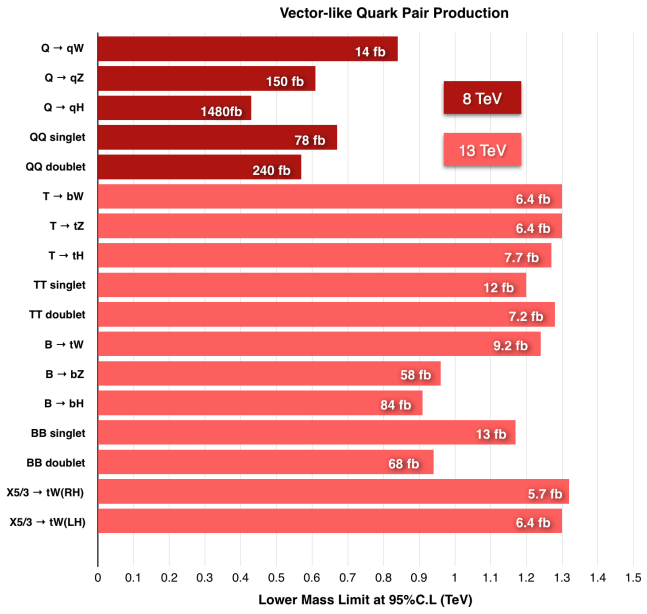


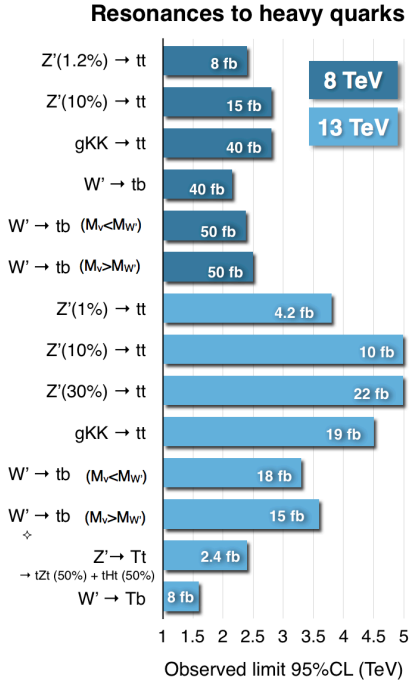


Summary

- Presented several new CMS searches with boosted tops
  - Rely upon jet substructure techniques
  - Path forward: luminosity limited, will need new techniques to tackle QCD background
- More on top tagging: see talks by R. Kogler, S. Lee, A. Sogaard, C. Suarez, and M.Guzzi
- Hope to see new developments in top tagging being used at next year's BOOST!



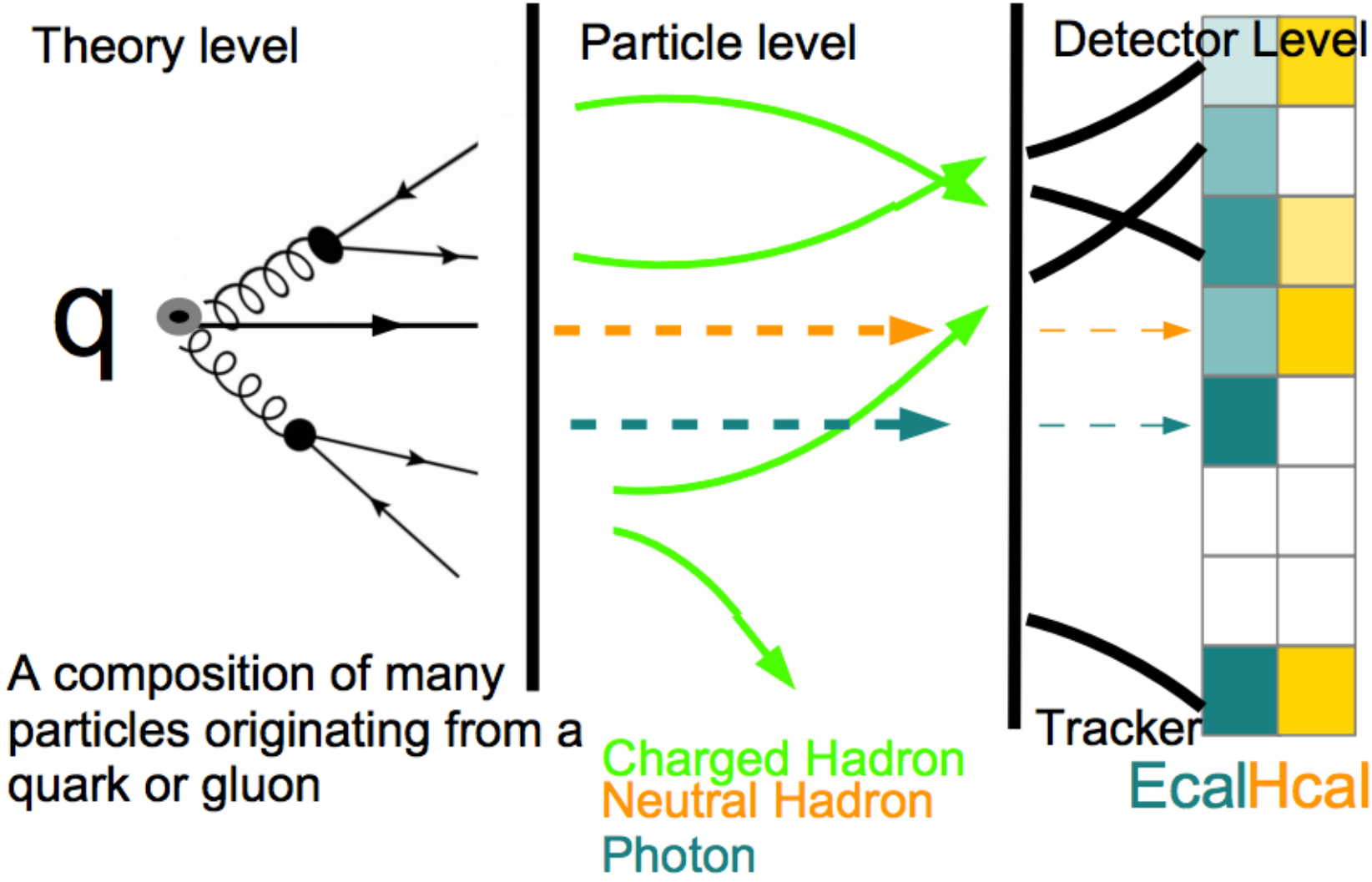




Additional Material

Jets

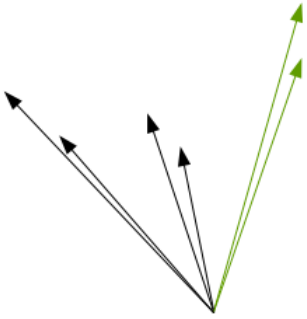
- **Jet**: a spray of particles from a quark or gluon
  - Due to QCD confinement



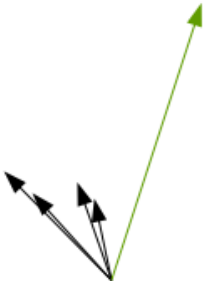
Jet Clustering: Anti-k_t

- 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

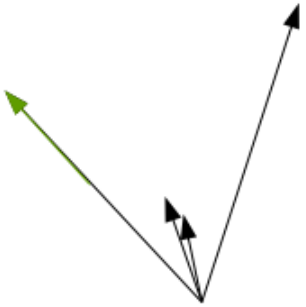
1.) Start with high p_T particles



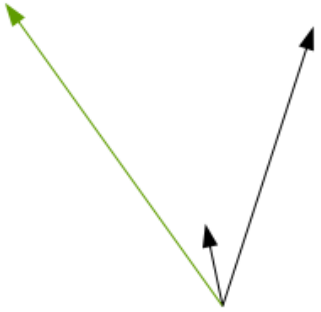
2.) Merge



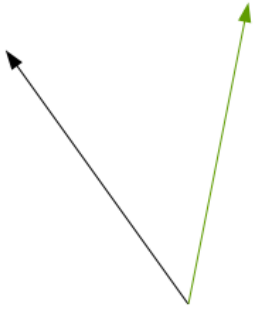
3.) Merge next pair



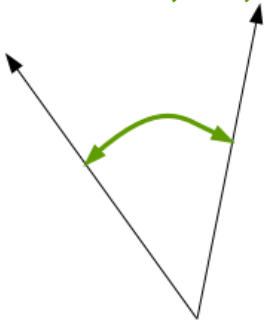
4.) Merge next pair



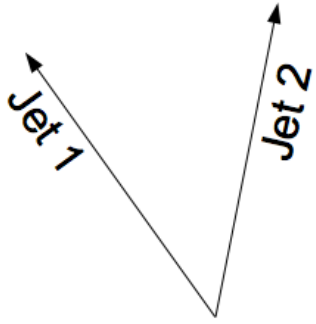
5.) Merge next pair

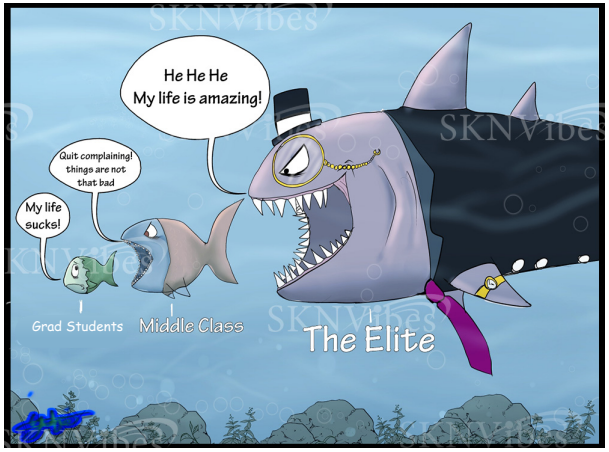


6.) Stop! $\Delta R > X$ X = 0.4, 0.8, …



7.) Done!



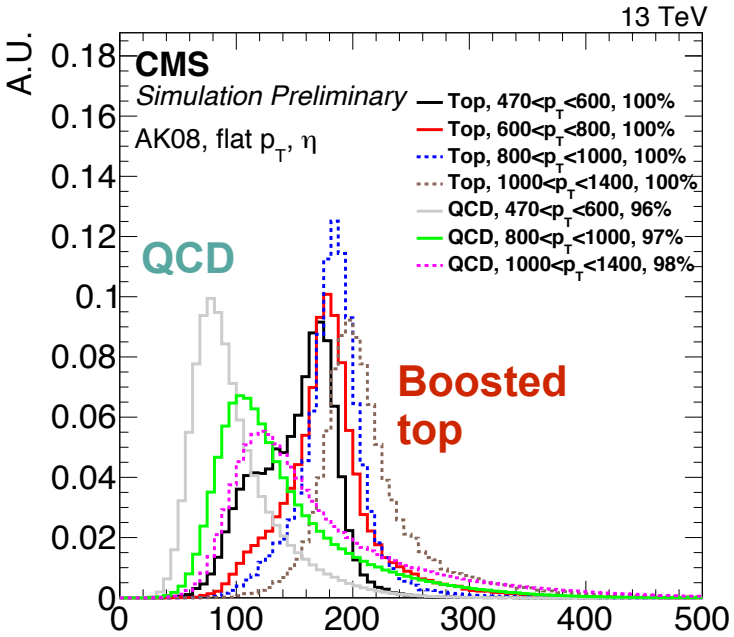


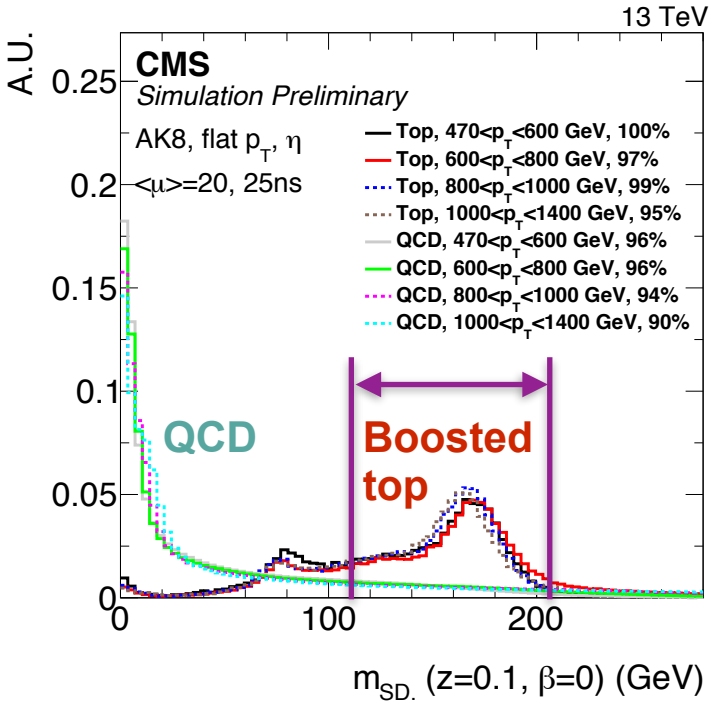
Boosted Top Quark ID: Jet Mass

- Signal jet mass should be ~ top quark mass

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

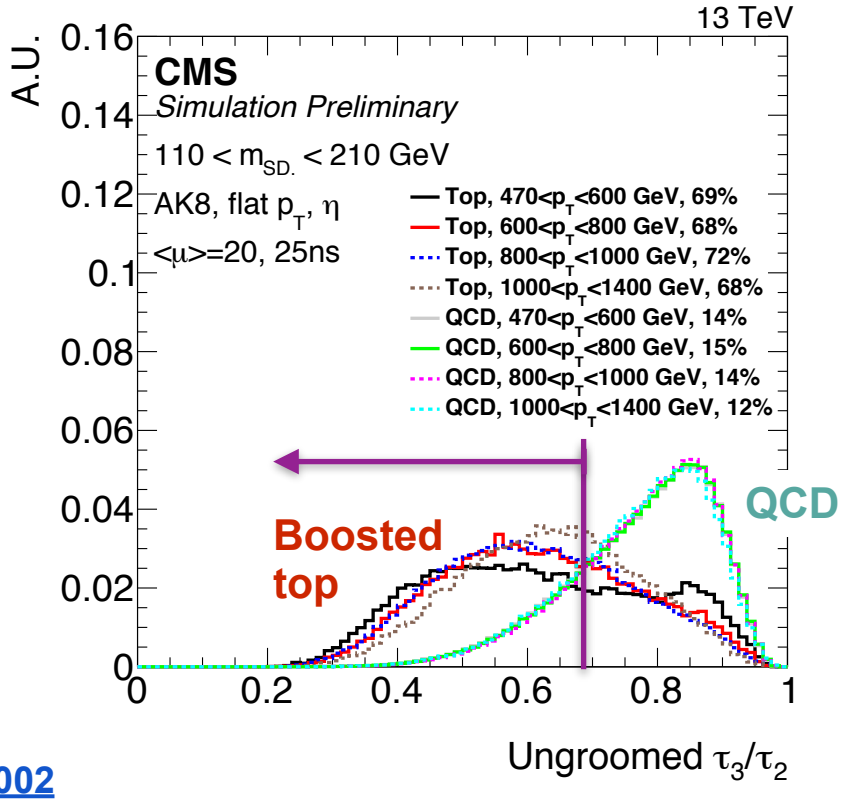
- N-subjettiness: Jet shape variable to measure consistency of jet to have N or fewer subjets

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

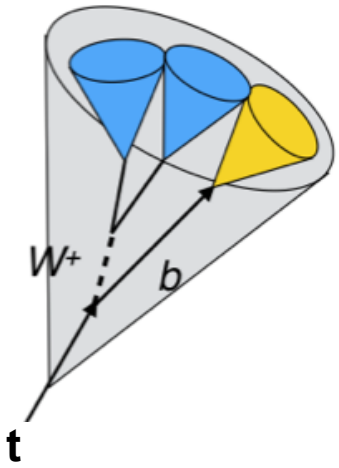
p_{Tk}: p_T of constituent k

$\Delta R_k^{\min}$: distance between constituent k & axis of closest subjet

R: large-R jet distance parameter



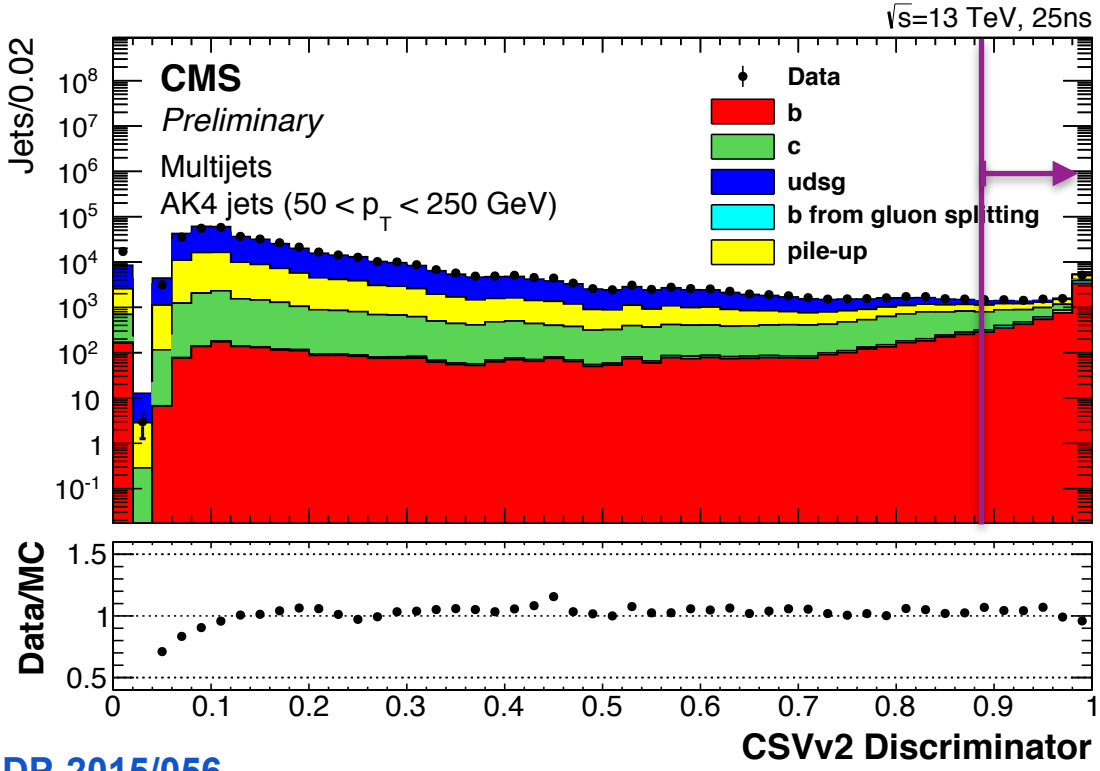
Boosted top: 3 subjets

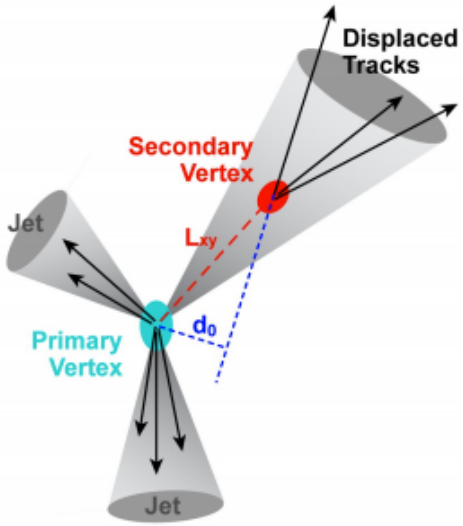


CMS-PAS-JME-15-002

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

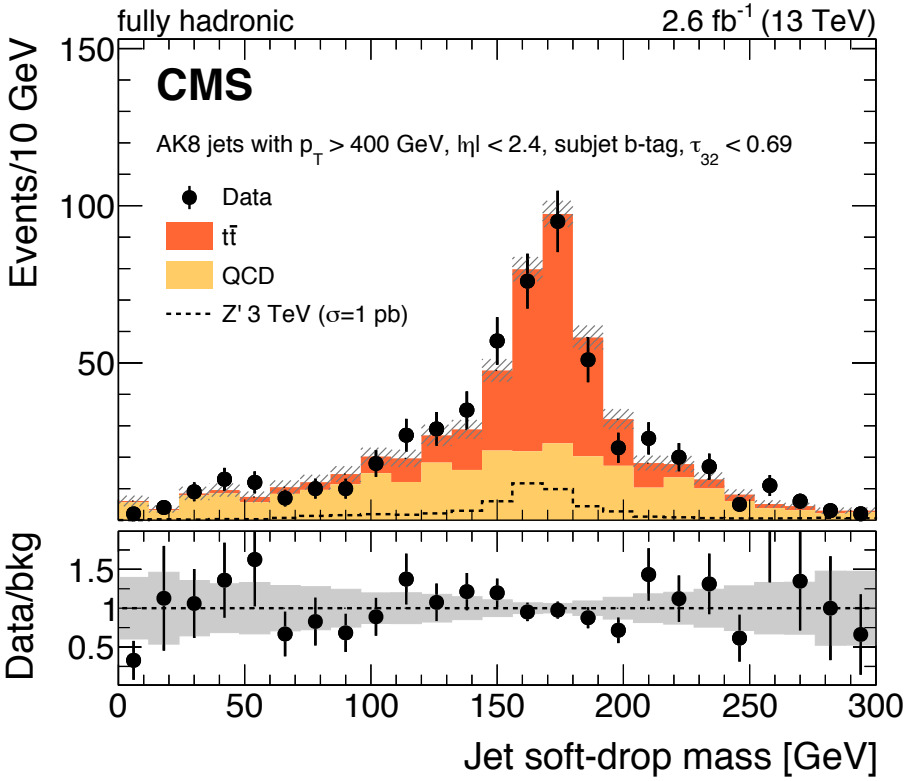


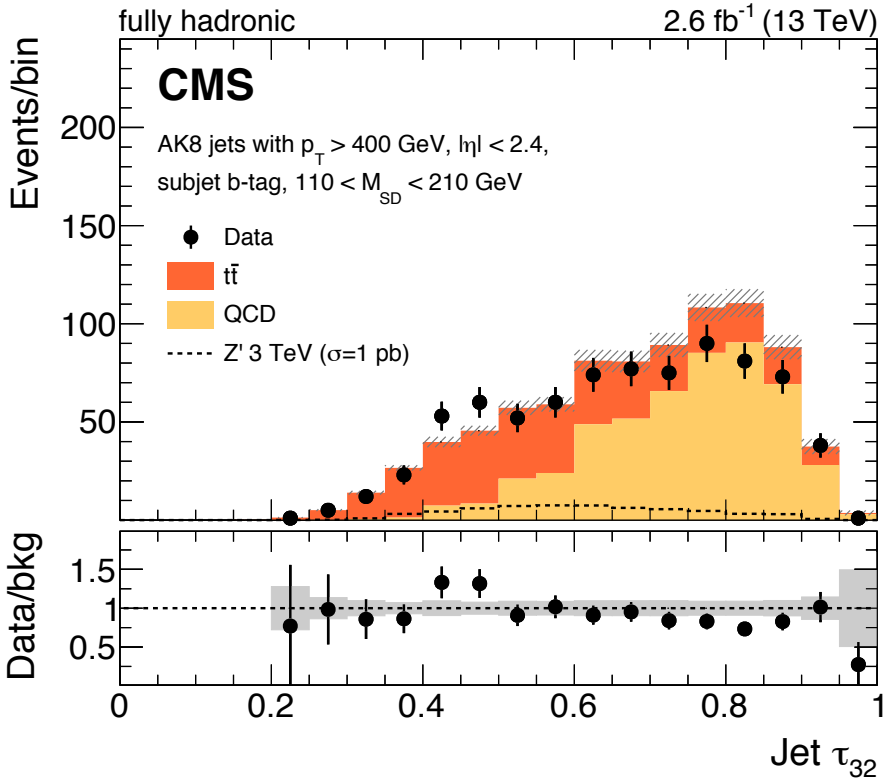


[CMS DP-2015/056]

Combine Jet ID variables: CMS Top Tagger Version 2

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

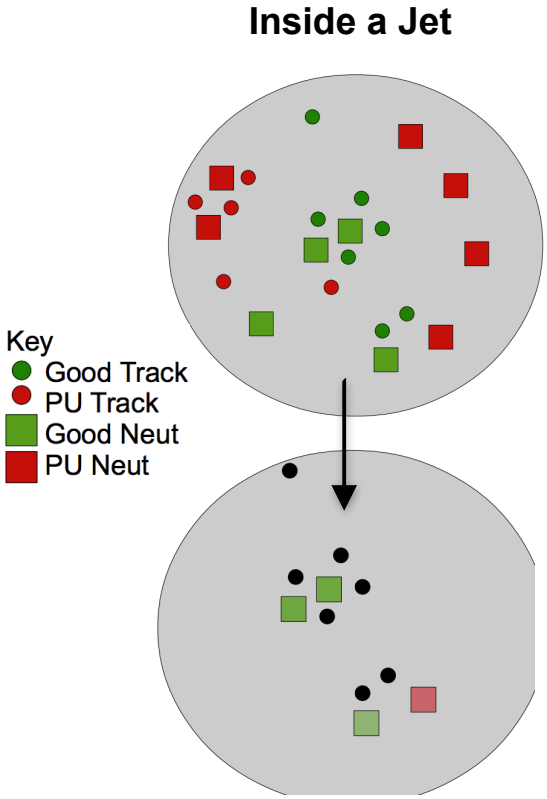


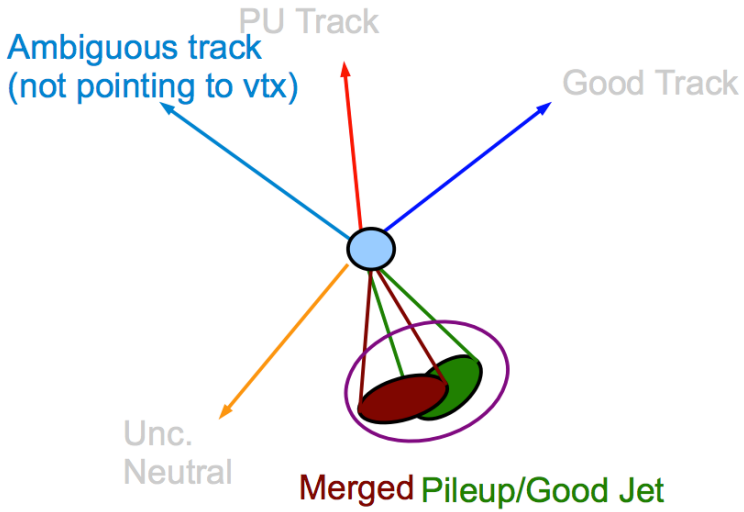


arXiv:1704.03366

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

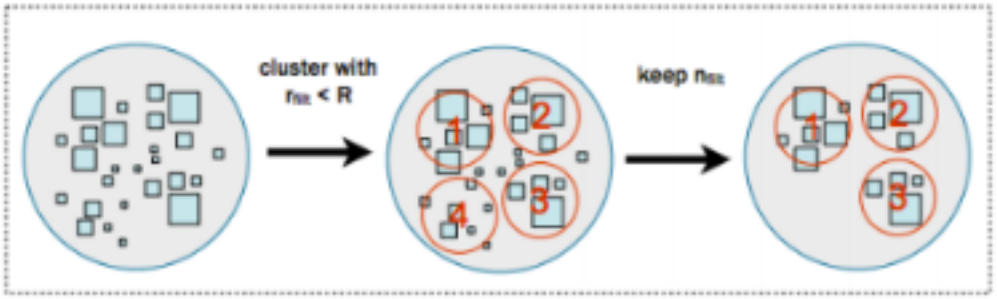




arxiv hep-ph/1407.6013

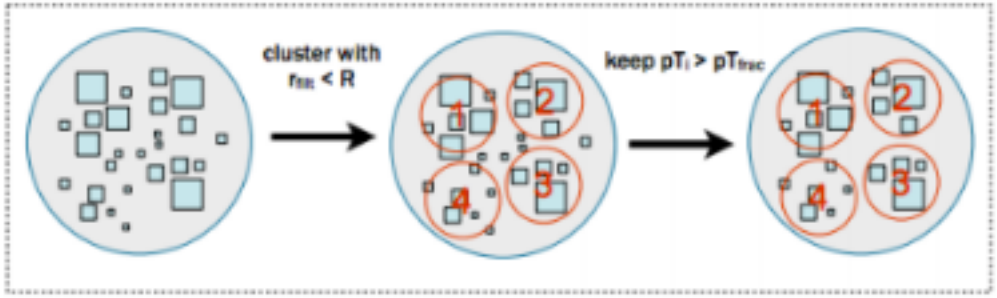
Mass Groomers

filtering



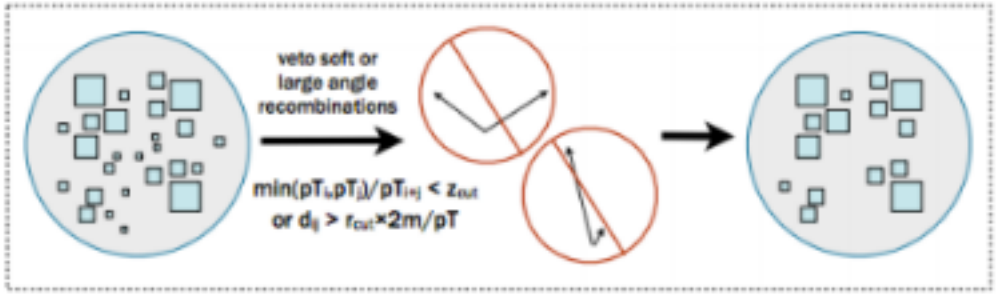
default:
$r_{filt} = 0.3$
$n_{filt} = 3$

trimming



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

pruning



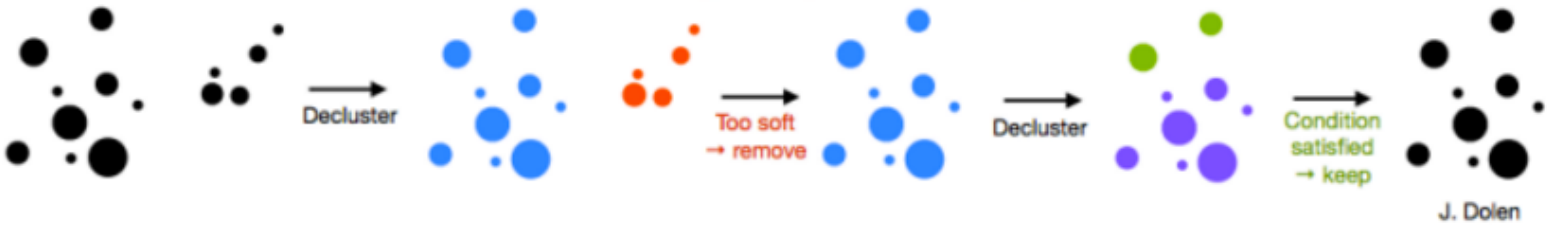
default:
$z_{cut} = 0.1$
$r_{cut} = 0.5$

Jim Dolen, Justin Pilot, Joshua Swanson, Nhan Tran

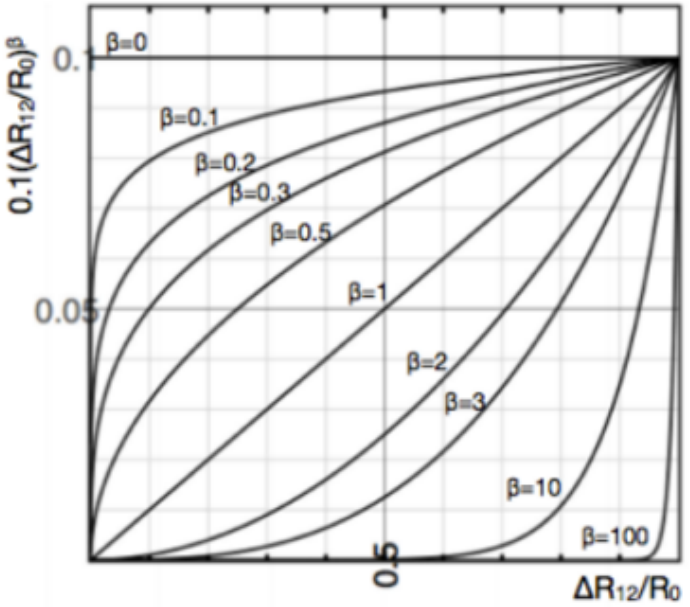
Soft Drop Algorithm

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

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



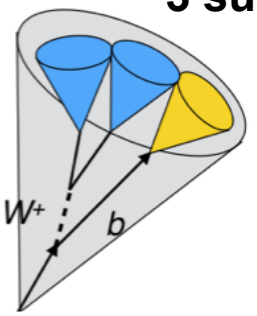
- Soft wide angle radiation fails the condition
  - As z_{cut} ↑ ⇒ more aggressive grooming
  - As β ↓ ⇒ more aggressive grooming
- Example ($z_{cut} = 0.1$) :
  - If β = 0, remove softer subjet if p_T fraction < 0.1 (~equivalent to MMDT)
  - If β > 0, remove softer subjet if p_T fraction < x, where 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)

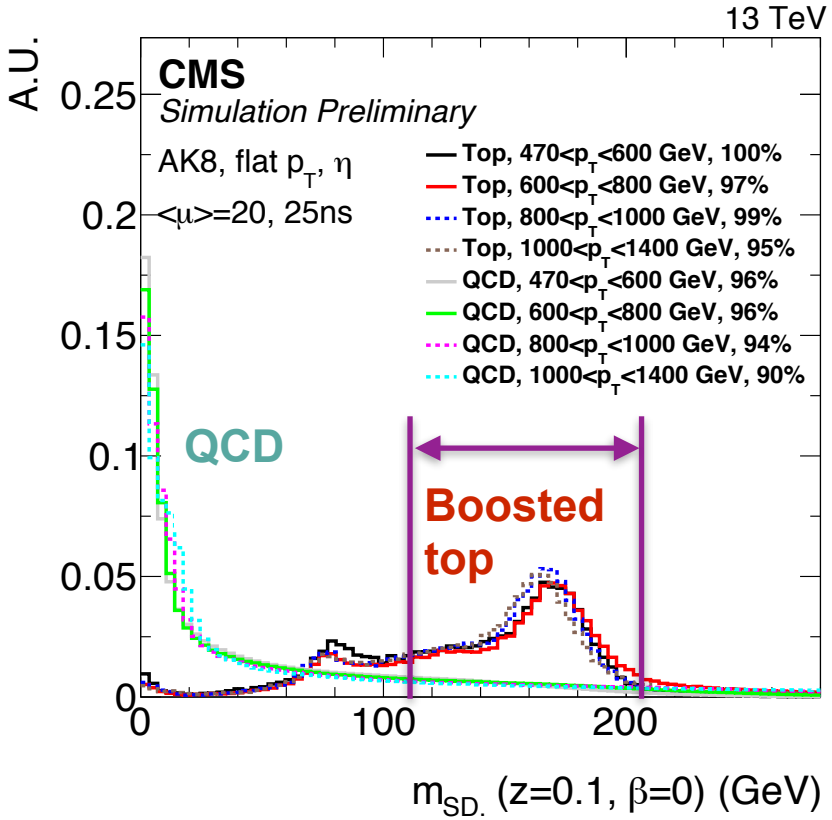


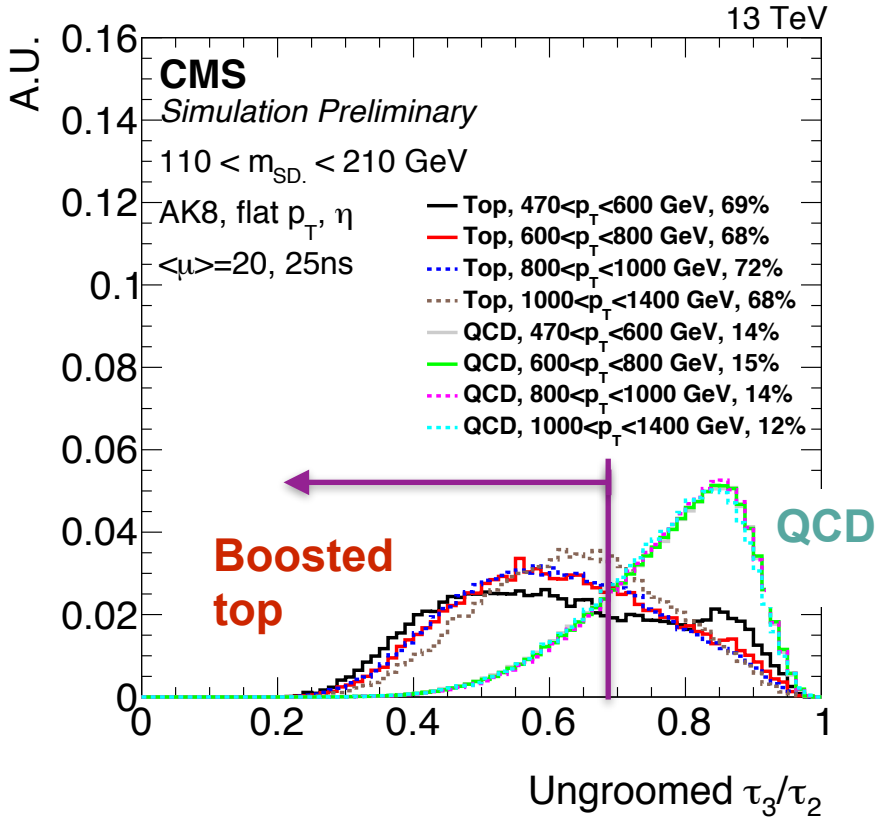
Techniques for Boosted Top ID

- **Groomed jet mass** - CMS uses soft drop mass
- **N-subjettiness** - used to distinguish a jet with N or fewer subjets
- **Subjet b-tagging** - identify secondary vertex of b jet decays

Boosted top: 3 subjets

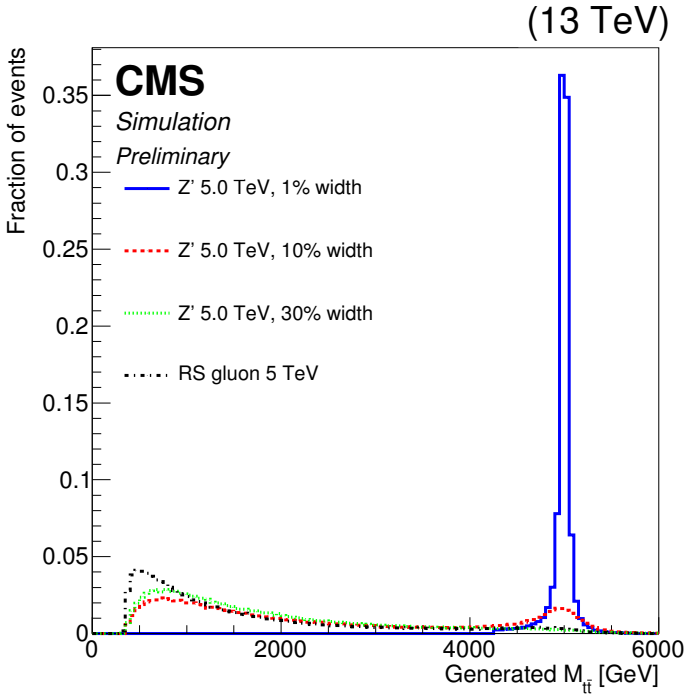


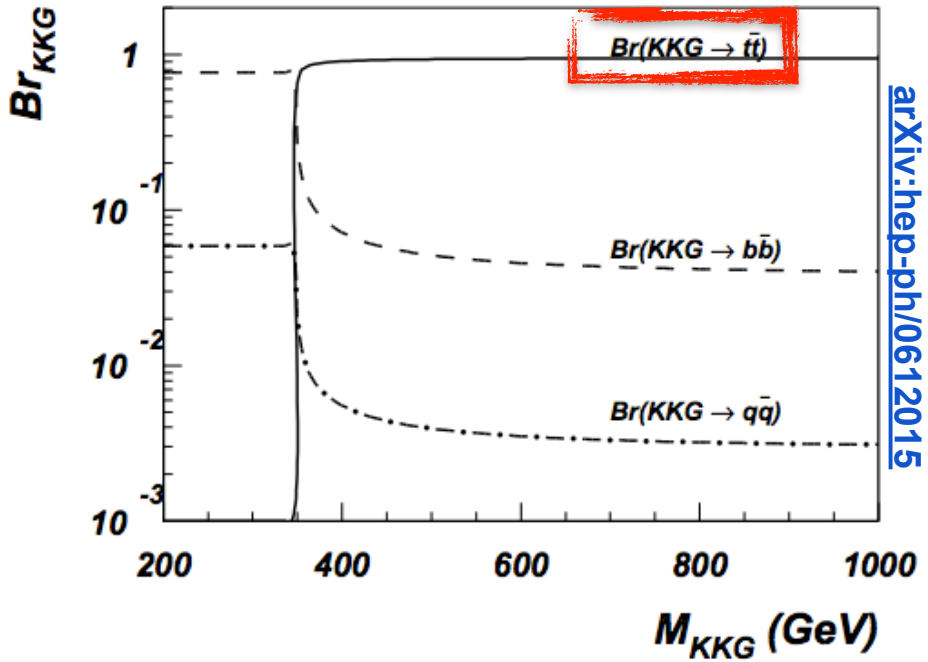




Heavy Resonances Decaying to Top Pairs

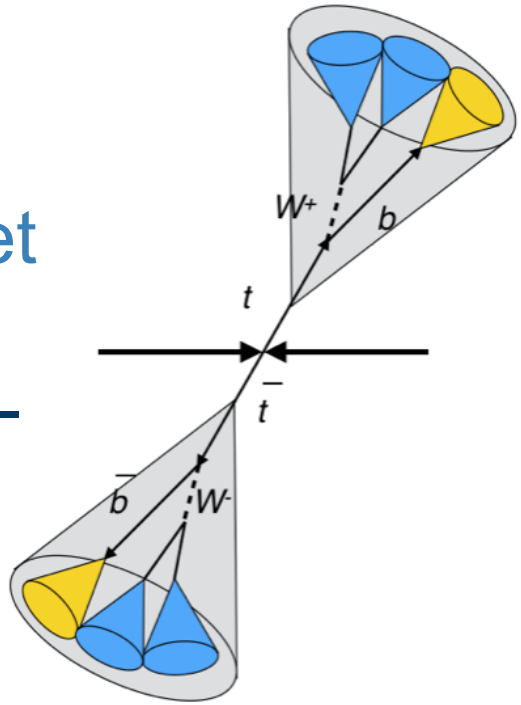
- The SM is quite successful, but leaves unanswered questions, e.g. why gravity is so weak
- Randall Sundrum model of extra dimensions predicts heavy gluon (RS KK gluon)
  - Decays almost exclusively to $t\bar{t}$ pairs
- Also search for generic Z′ boson decaying to $t\bar{t}$
  - Narrow, wide, and extra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





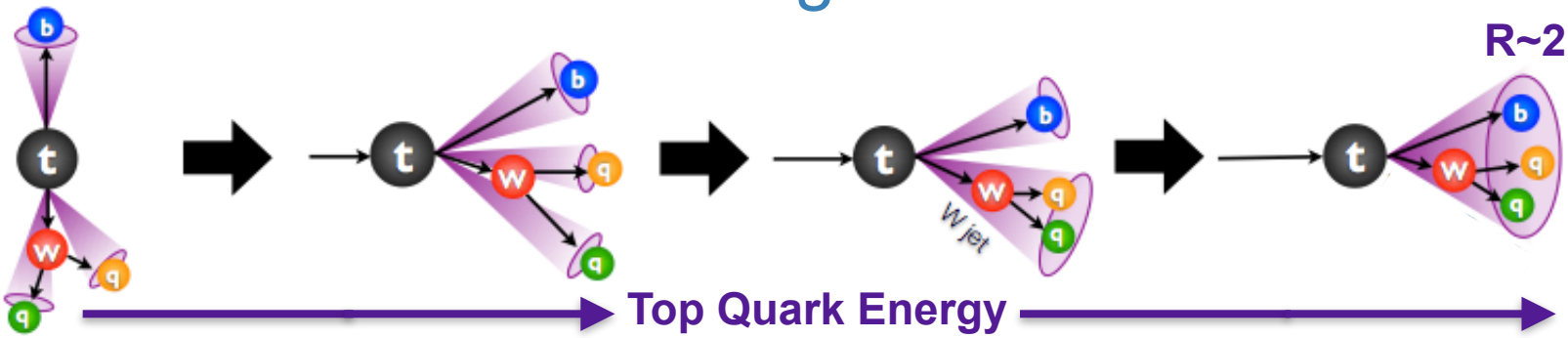
Z′ to $t\bar{t}$

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



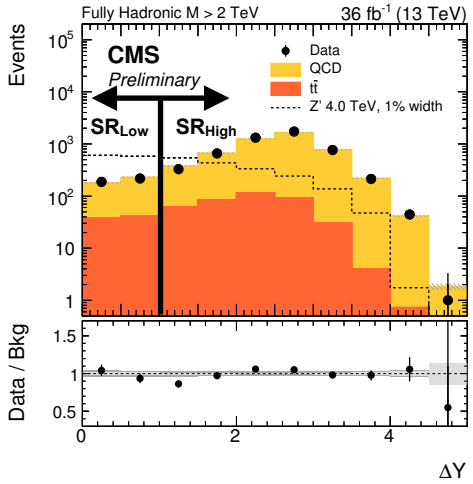
- Signature: two large (R=0.8) back-to-back jets: $|\Delta\Phi| > 2.1$, $p_T > 400$ GeV

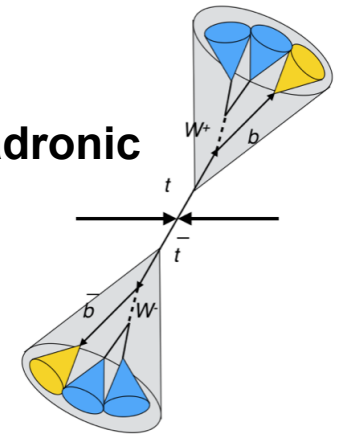
- Background: SM $t\bar{t}$ and QCD events
  - QCD estimated using data

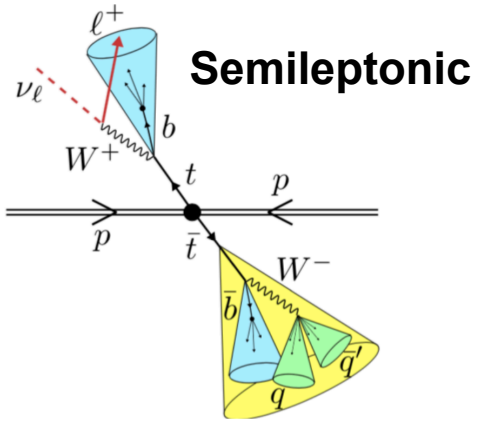


Heavy Resonances to $t\bar{t}$: Signal Regions

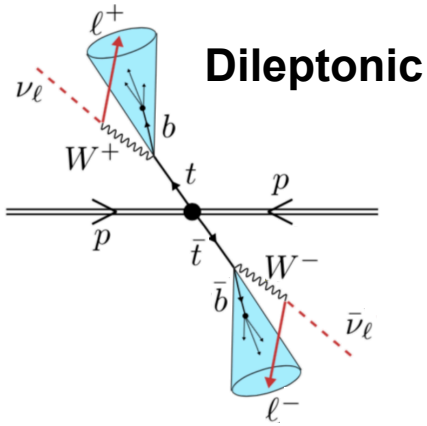
- All-hadronic: 2 back-to-back PUPPI top-tagged jets
  - Categorized by number of subjet b-tags (0,1,2) and $|\Delta y|_{tt}$







- Semileptonic: e+jets, μ+jets channels
  - Lepton 2D isolation cut
  - X^2-based reconstruction of top system
  - Categorized by number of top tags (0,1)
- **Dileptonic: ee, μμ channels**
  - **Lepton 2D isolation cut**
  - **At least one b-tagged jet**
  - **Characterize signal region by lepton/jet separation**



Heavy Resonances to $t\bar{t}$: Main Backgrounds

- All-hadronic: $t\bar{t}$ (MC), QCD (estimated from data)

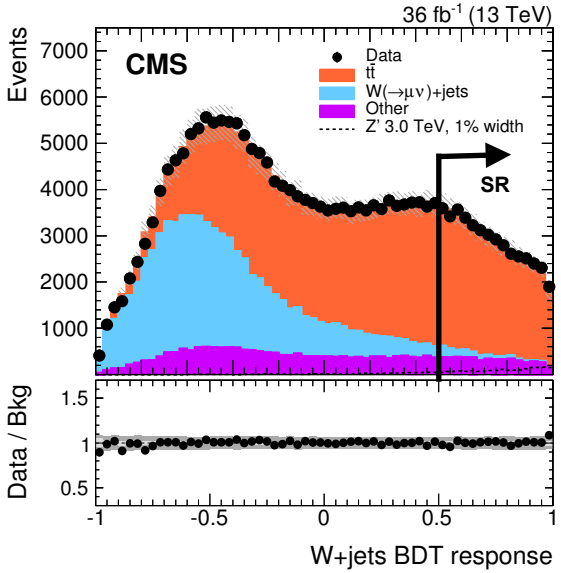
Measure QCD Mistag Rate

- $\tau_{32}^{puppi} > 0.65$
- $105 < m_{SD}^{puppi} < 210$

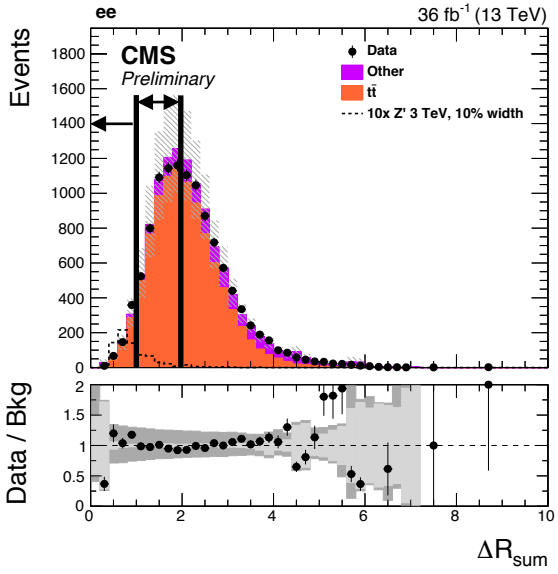
Measure top-tag rate

Estimate QCD Background

- $\tau_{32}^{puppi} < 0.65$
- $105 < m_{SD}^{puppi} < 210$
- Weight event by mistag rate
- Apply mass in top-tag window



- Semileptonic: $t\bar{t}$ (MC), W+jets
  - W+jets suppressed by BDT



- Dileptonic: $t\bar{t}$ (MC)
  - Reduced by choice of SR
    - $\Delta R_{sum} < 1$, $1 < \Delta R_{sum} < 2$
    - $\Delta R = \Delta R_{min}(l_1,j) + \Delta R_{min}(l_2,j)$

Heavy Resonances to tt: Semileptonic

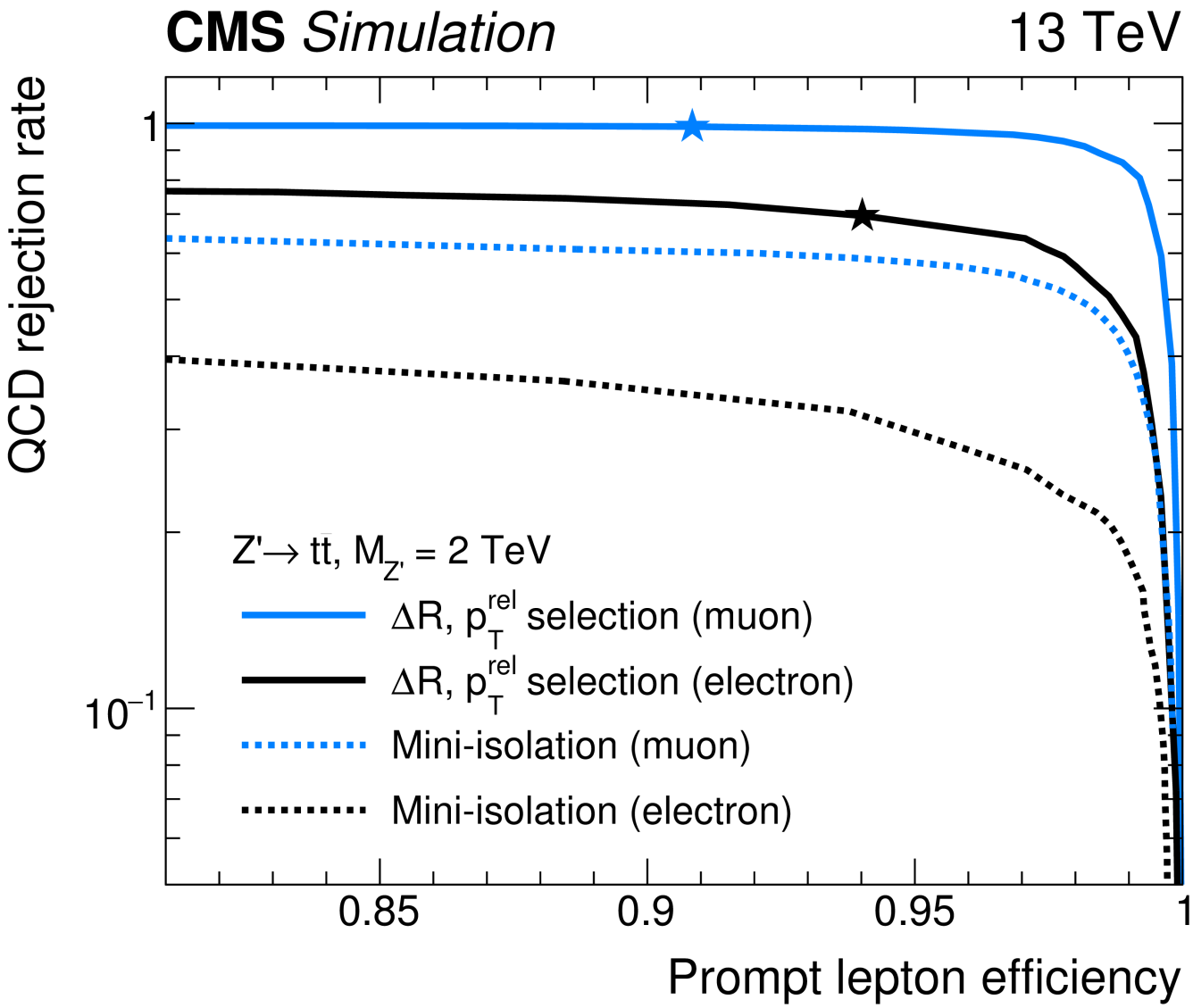
- W+jets BDT:

  - $\Delta R_{min}(\ell, j)$,the separation between the lepton and its closest jet
  - The CSV score of the subleading and the leading AK4 jets
  - The number of jets
  - $p_{T,rel}(\ell, j)$
  - The reconstructed mass of the leading AK4 jet
  - $\Delta R_{min}(\ell, j) \times p_T(j)$
  - The reconstructed mass of the subleading AK4 jet
  - The shape variable S_{33} of the sphericity tensor $S^{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |p_i|^2}$ where α, β correspond to the x, y and z components of the momentum vectors of the jets
  - $S_T \equiv H_T + H_T^{lep}$, where H_T is the scalar sum of the p_T of the jets

$$\chi^2 < 30$$

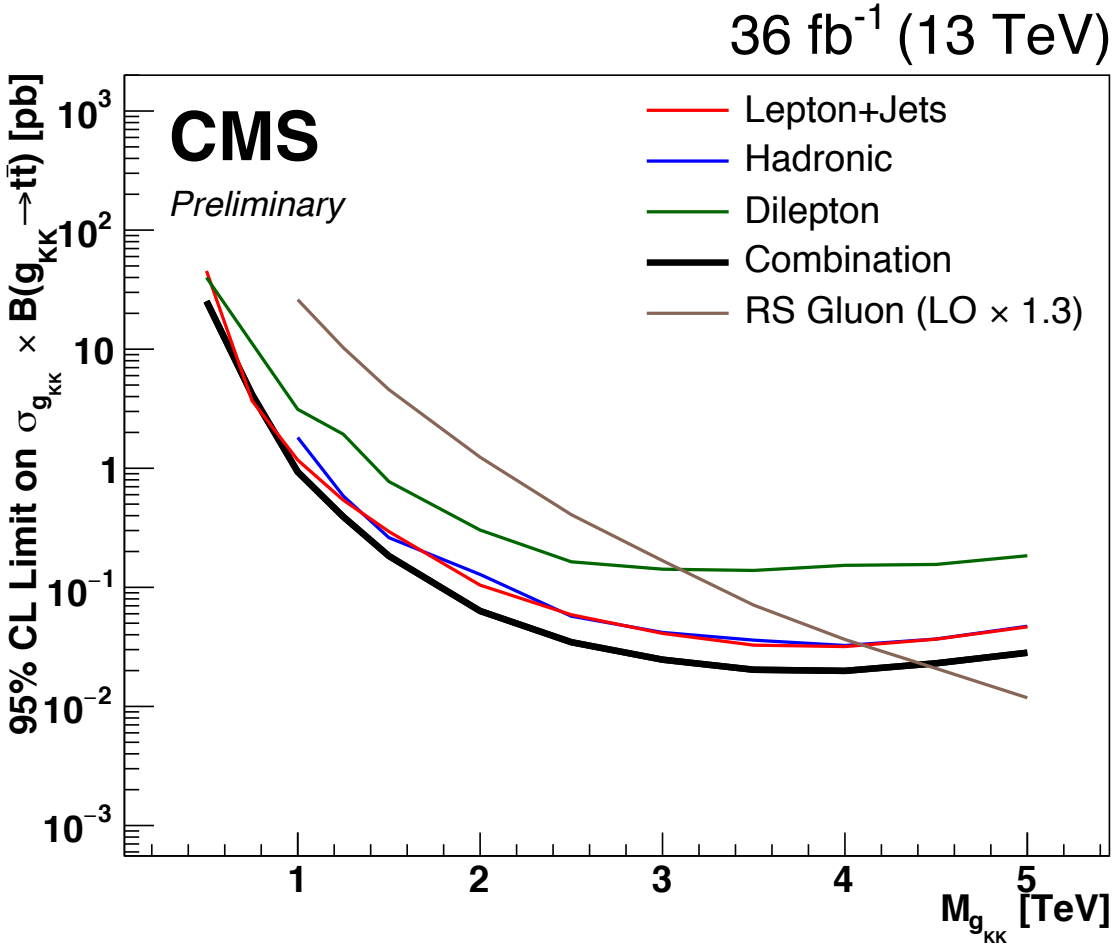
$$\chi^2 = \chi^2_{lep} + \chi^2_{had} = \left[\frac{M_{lep} - \bar{M}_{lep}}{\sigma_{M_{lep}}}\right]^2 + \left[\frac{M_{had} - \bar{M}_{had}}{\sigma_{M_{had}}}\right]^2$$

Heavy Resonances to tt: Semileptonic



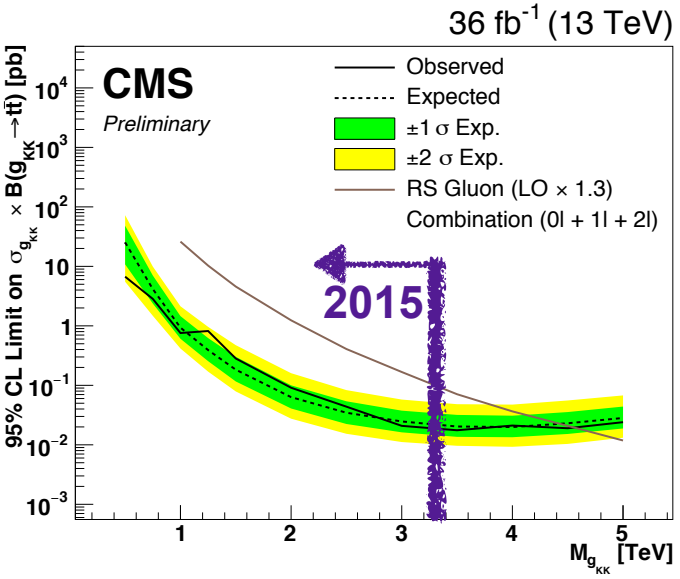
Heavy Resonances to $t\bar{t}$: Results

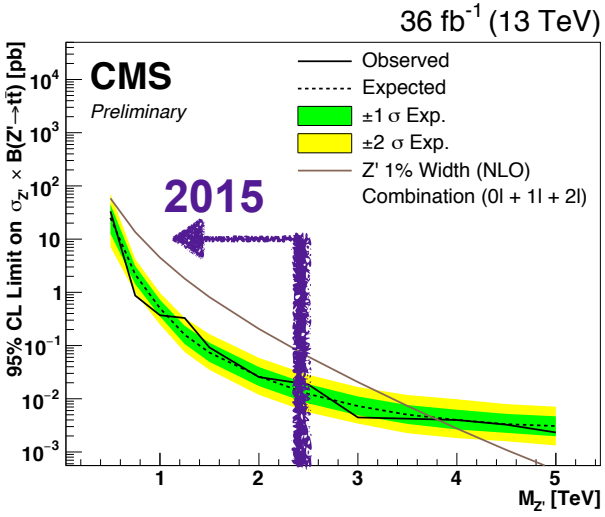
- Discriminating variables
  - $M_{t\bar{t}}$ - all-hadronic and semileptonic channels
  - S_T - dileptonic channel

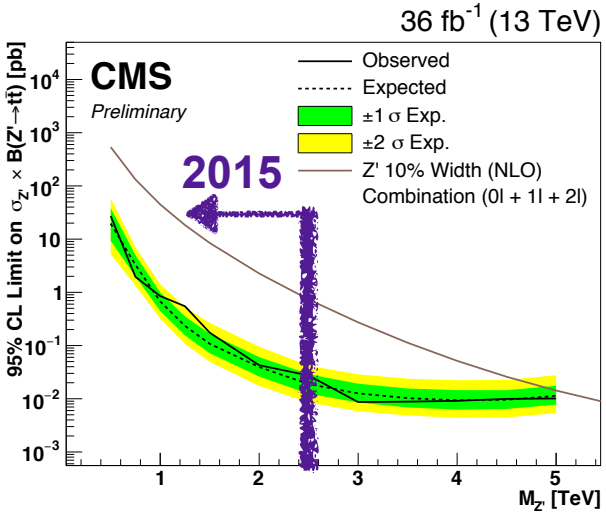


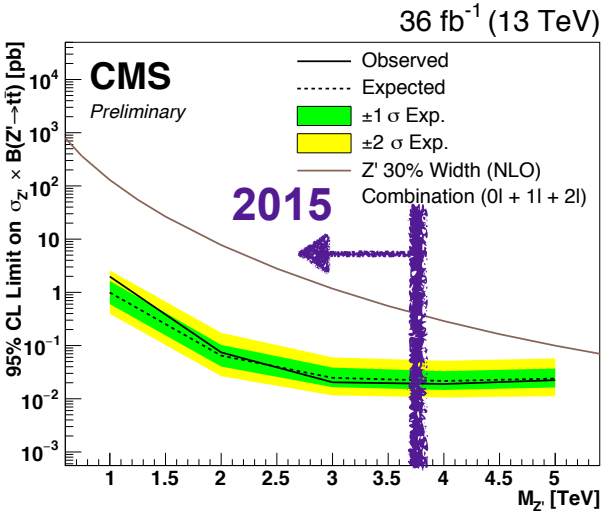
Heavy Resonances to $t\bar{t}$: Results

- Discriminating variables
  - $M_{t\bar{t}}$ - all-hadronic and semileptonic channels
  - S_T - dileptonic channel
- Significant improvement over 2015 results!







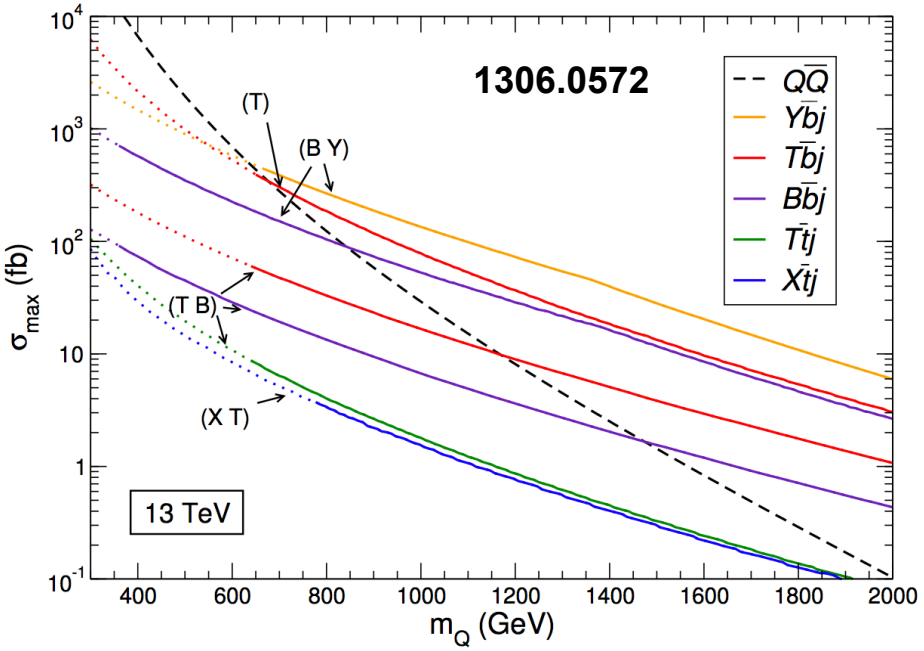


Vector-Like Quarks

- Non-traditional quarks
  - **The standard:** spin-1/2, colored, charged particles
  - **The unusual:** both left- and right-handed coupling to CCs
    - Mass not acquired through Higgs coupling - **don't affect Higgs cross section or width!**
- Predicted by composite Higgs and little Higgs models
- Both single- and pair-production
  - Single production dominates at high masses (above ~ 2 TeV)

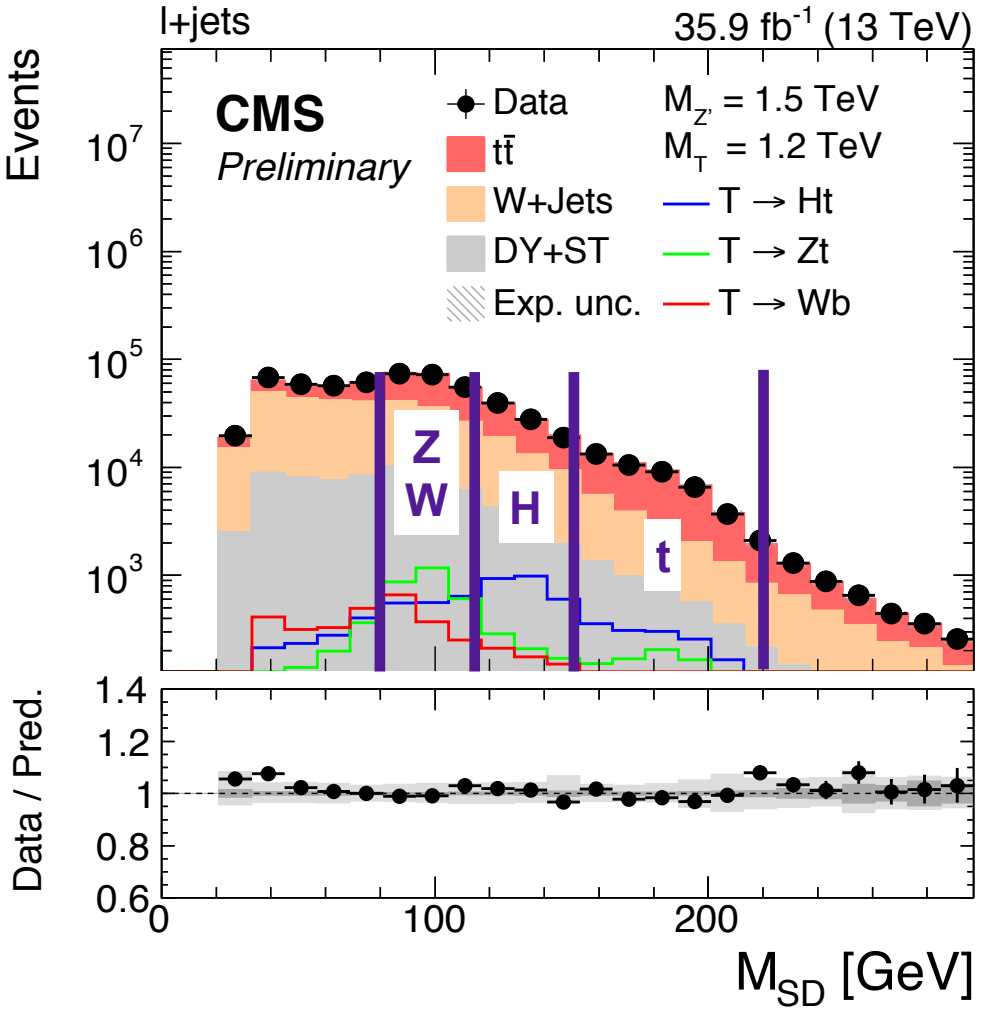
Weak multiplets (isospin$_{\text{hypercharge}}$)

Singlets	Doublets	Triplets
$\mathbf{1}_{2/3} = T$	$\mathbf{2}_{1/6} = \begin{pmatrix} T \\ B \end{pmatrix}$	$\mathbf{3}_{2/3} = \begin{pmatrix} X \\ T \\ B \end{pmatrix}$
	$\mathbf{2}_{7/6} = \begin{pmatrix} X \\ T \end{pmatrix}$	
$\mathbf{1}_{-1/3} = B$	$\mathbf{2}_{-5/6} = \begin{pmatrix} B \\ Y \end{pmatrix}$	$\mathbf{3}_{-1/3} = \begin{pmatrix} T \\ B \\ Y \end{pmatrix}$



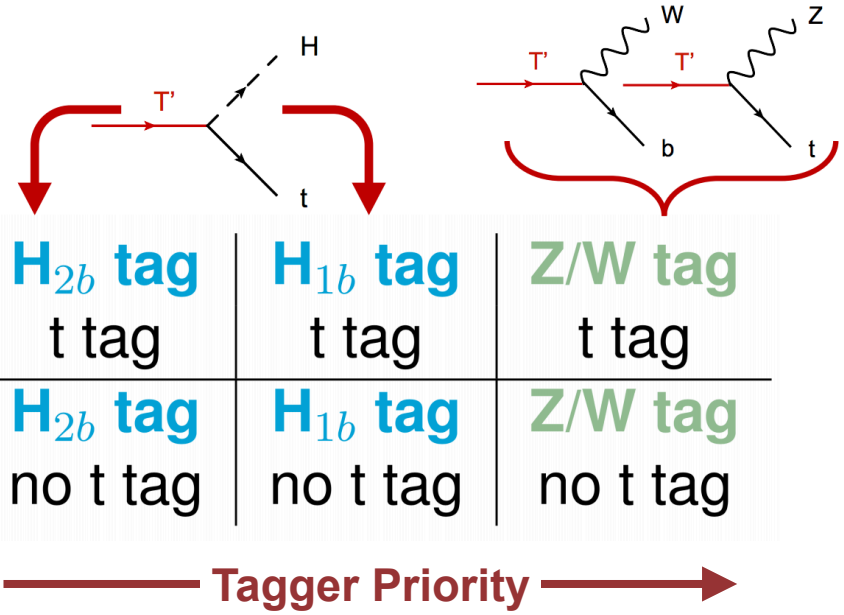
Z' → tT: Event Selection

- Preselection:
  - exactly one e or μ
  - ≥ 1 AK4 jet
  - ≥ 1 AK8 jet
  - MET
  - Non-isolated lepton:
    - $p_{T,rel}$ > 40 GeV || ΔR > 0.4
- Signal region:
  - Boosted jet tagging:
    - Softdrop mass
    - τ_{32} (t), τ_{21} (Z/W)
    - subjet b-tags (H)

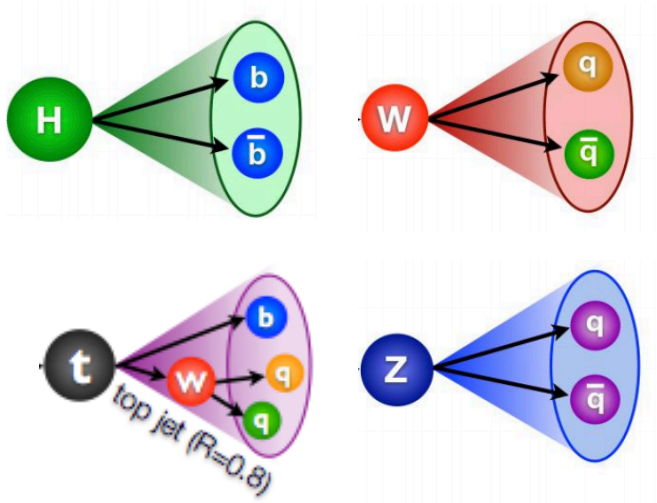


Z' → tT: Event Selection

- Preselection:
  - exactly one e or μ
  - ≥ 1 AK4 jet
  - ≥ 1 AK8 jet
  - MET
  - Non-isolated lepton:
    - $p_{T,rel}$ > 40 GeV || ΔR > 0.4

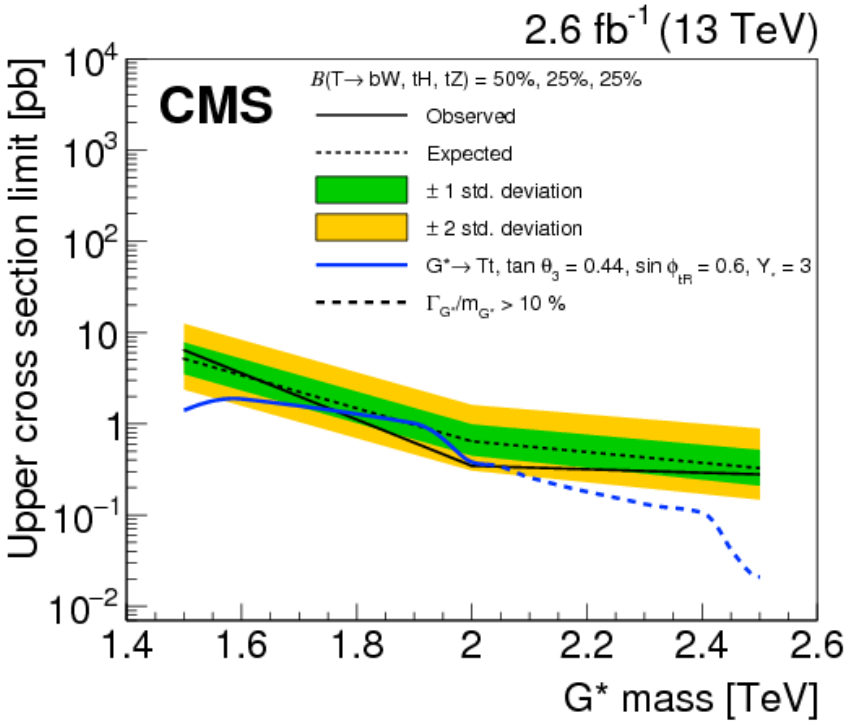


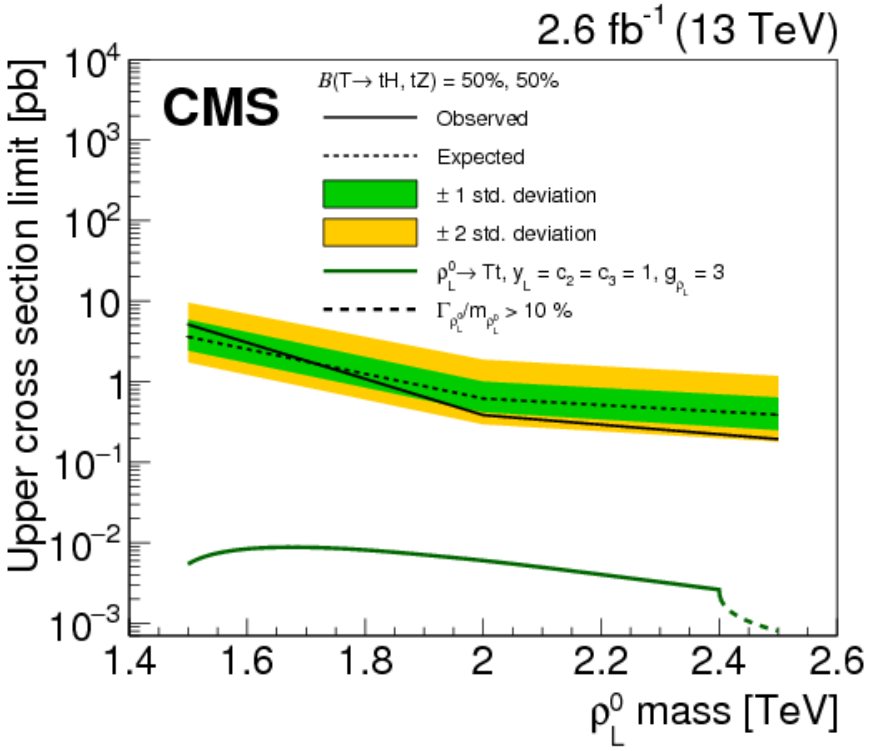
- Signal region:
  - Boosted jet tagging:
    - Softdrop mass
    - **τ_{32} (t), τ_{21} (Z/W)**
    - **subjet b-tags (H)**
  - **Categorize based on number of boosted jet (Z/W/H/t) tags**



Z' → tT: Previous Results

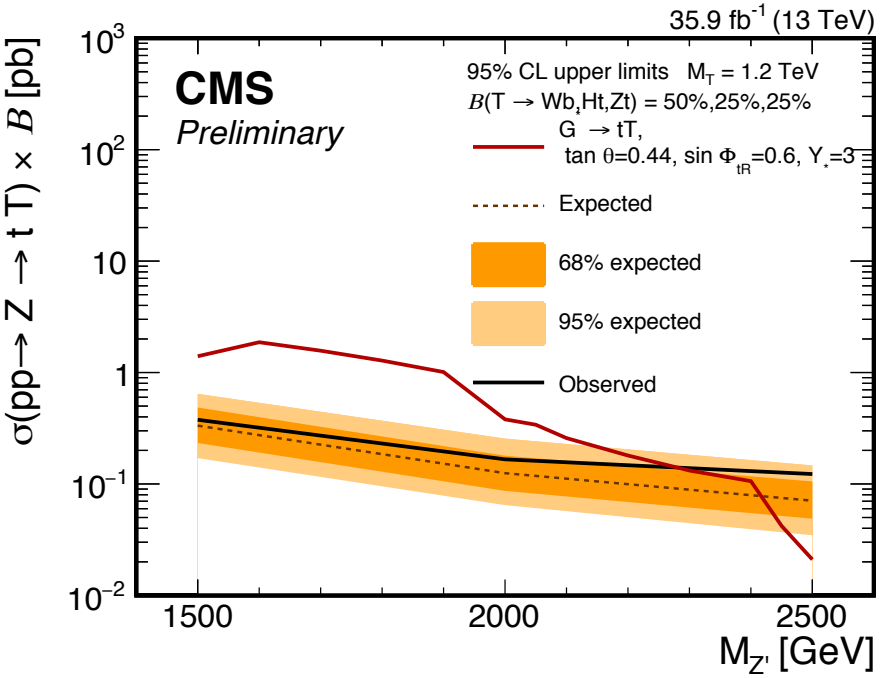
- Limits set as a function of Z' mass
  - **Optimized for bW channel**

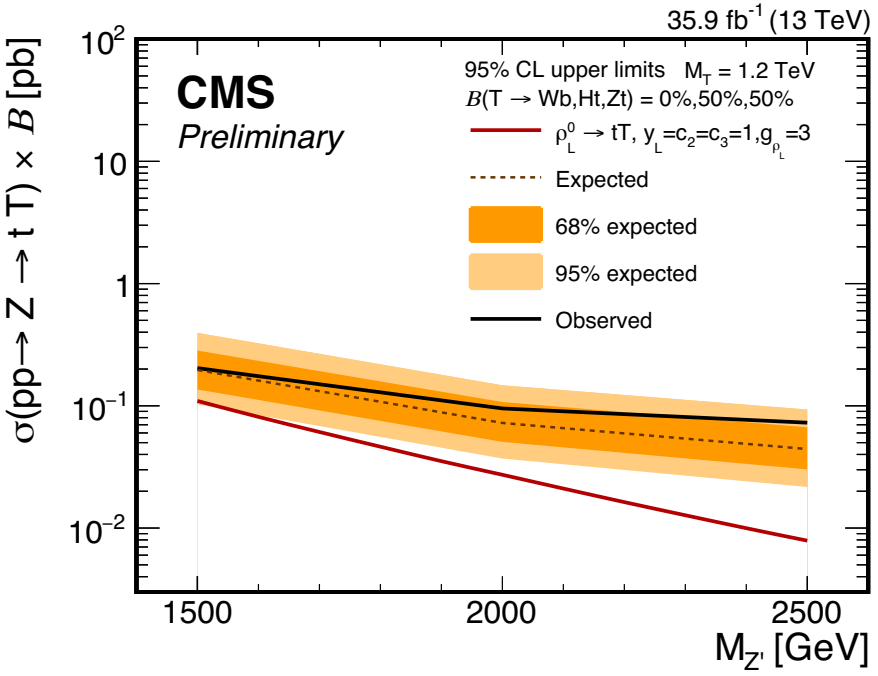




Z' → tT: Results

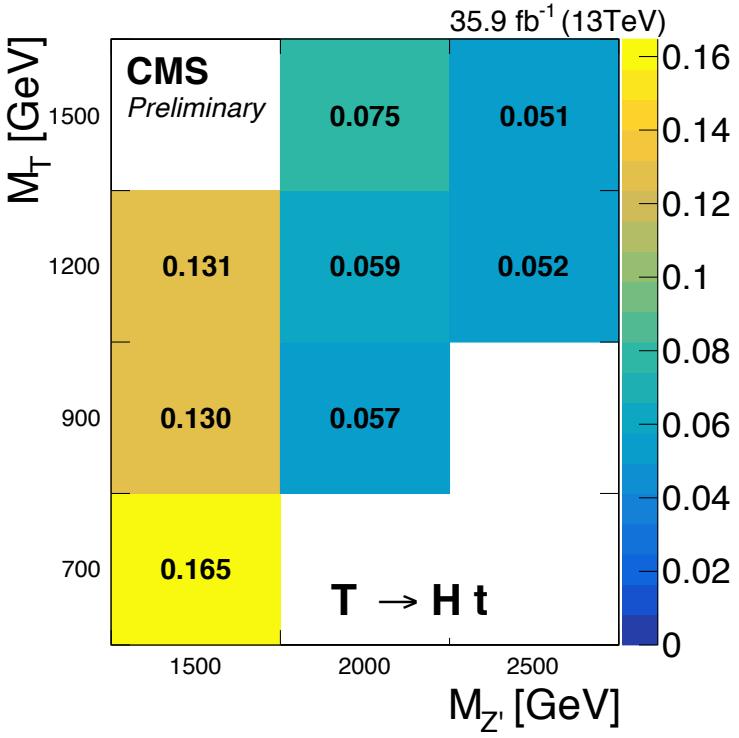
- Limits set as a function of Z' mass
  - **First results for Ht channel!**
  - Exceeds sensitivity of previous searches

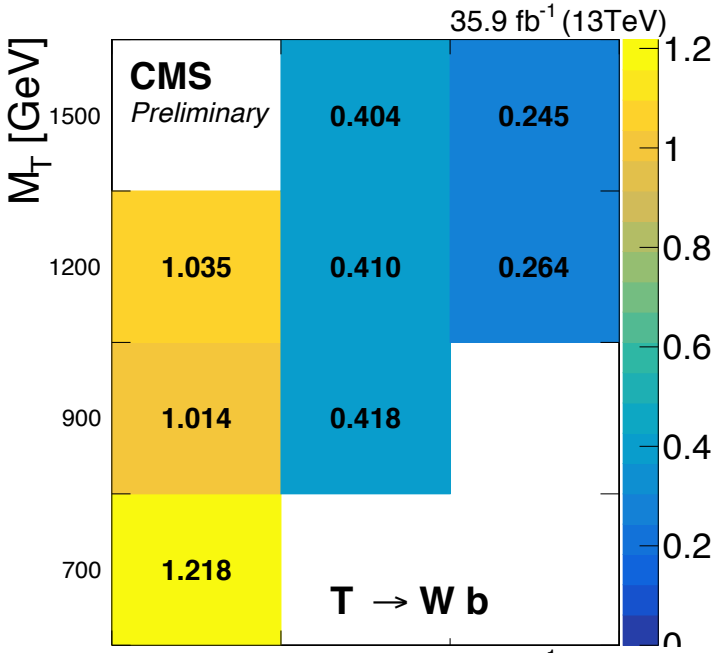


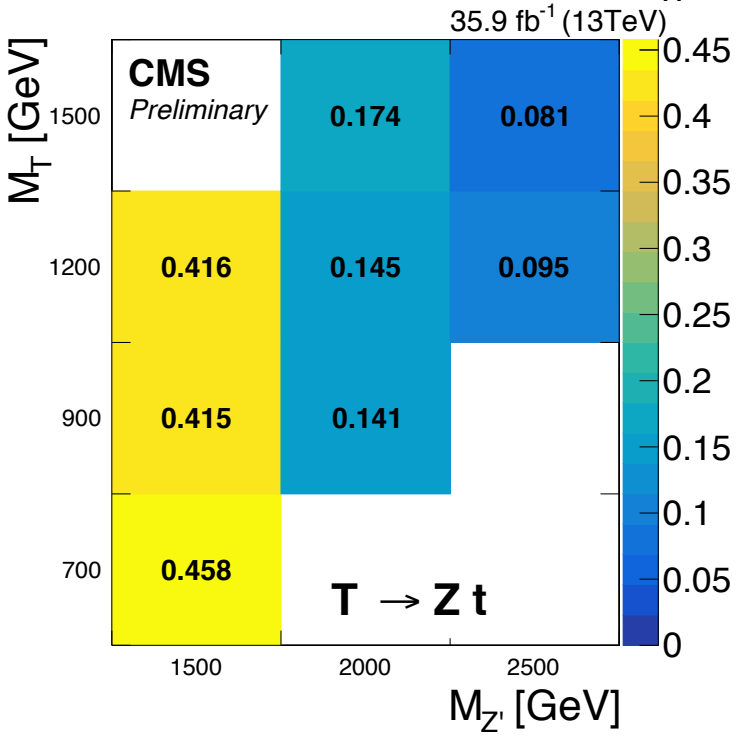


Z' → tT: Results

- Limits set as a function of Z' mass and T mass
  - **First results for Ht channel!**
  - Assume 100% BR for given T decay mode

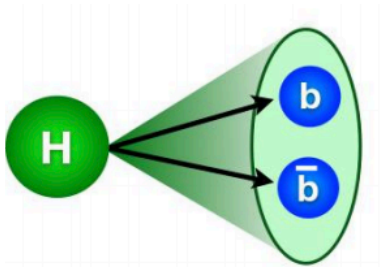


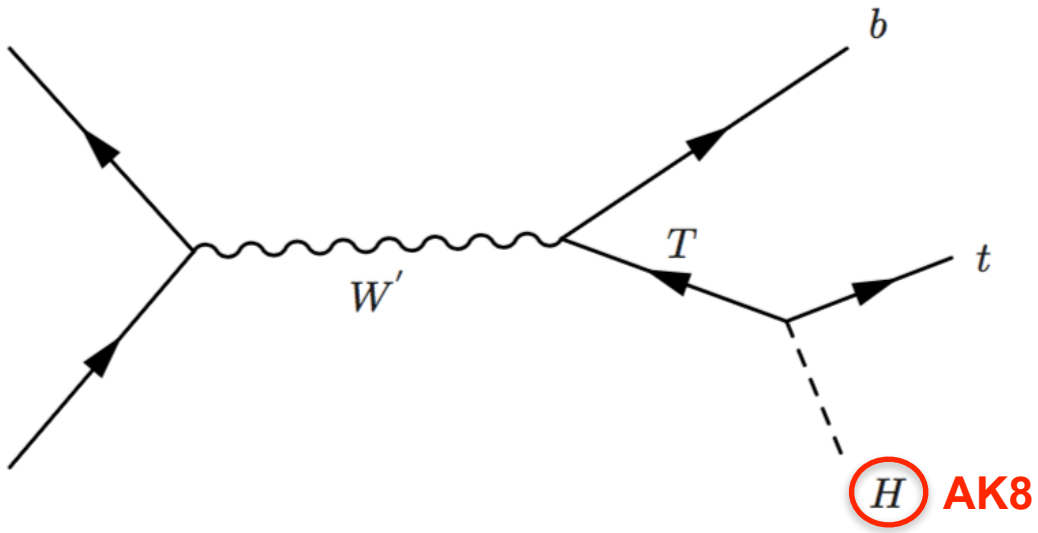


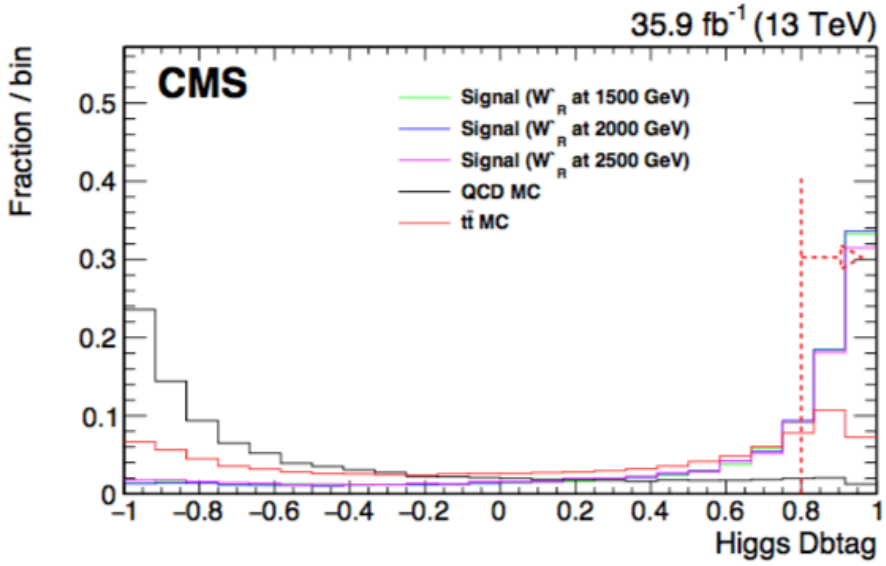


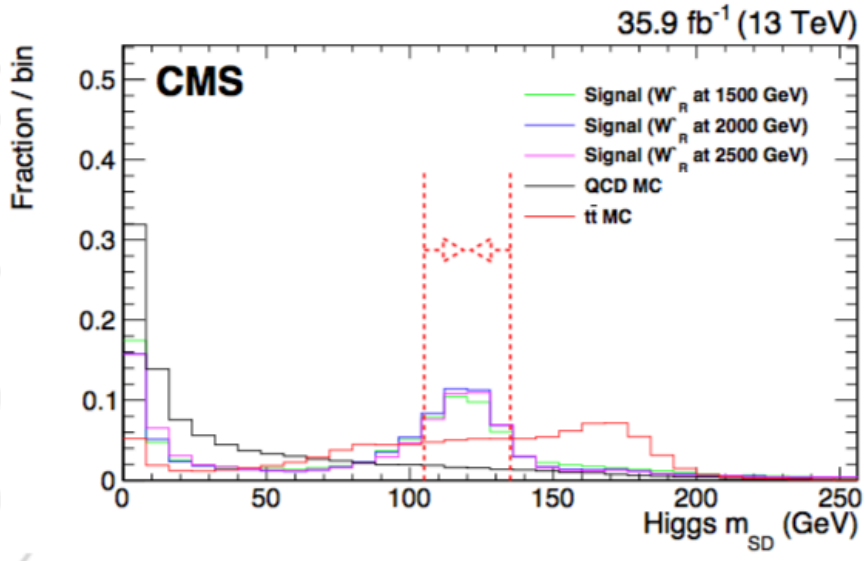
W' → Tb/tB: Signal Selection

1. Higgs jet
   - Double b-tagger



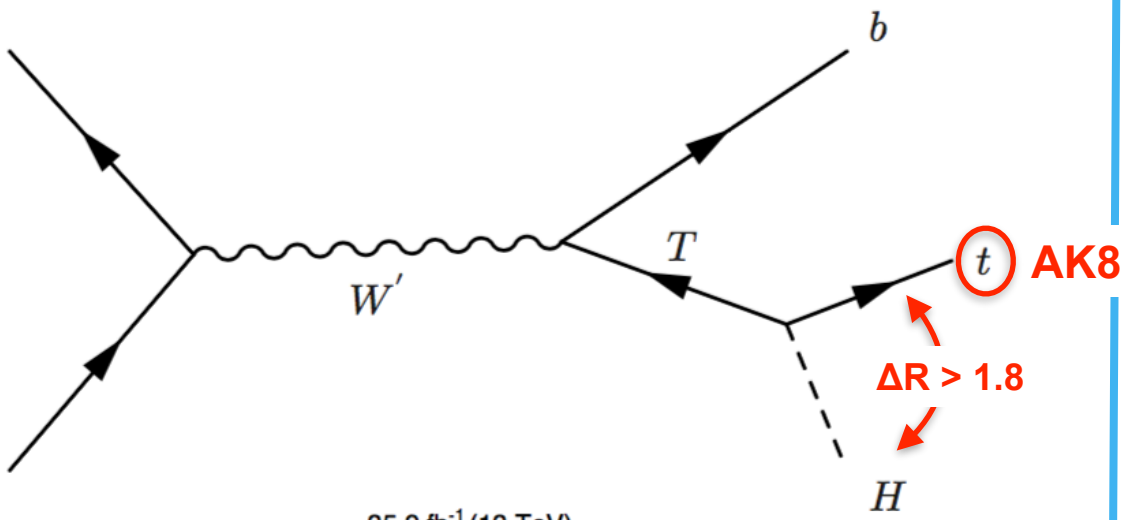


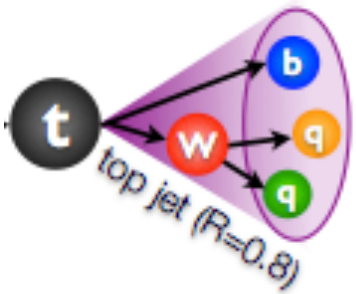


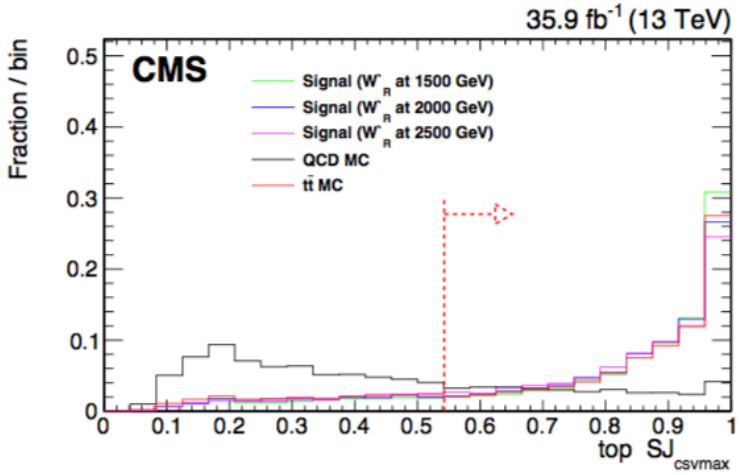


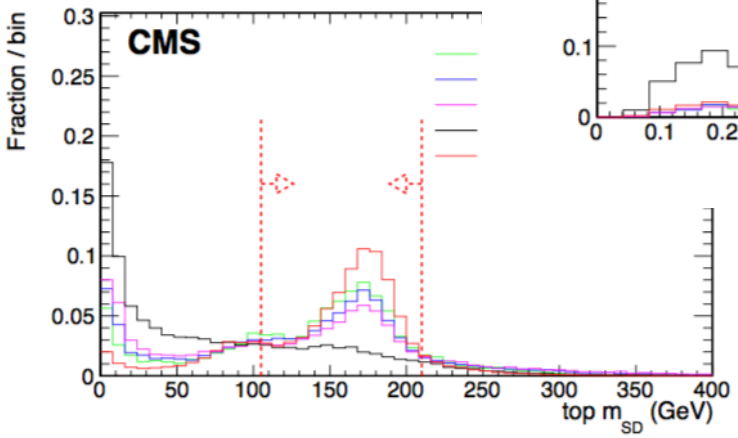
W' → Tb/tB: Signal Selection

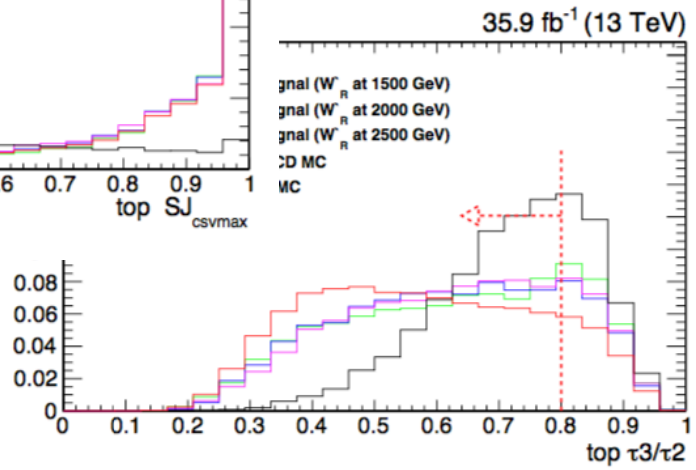
1. Higgs jet tag
   - Double b-tagger
2. Top jet tag
   - CMSTTv2 + b-tag





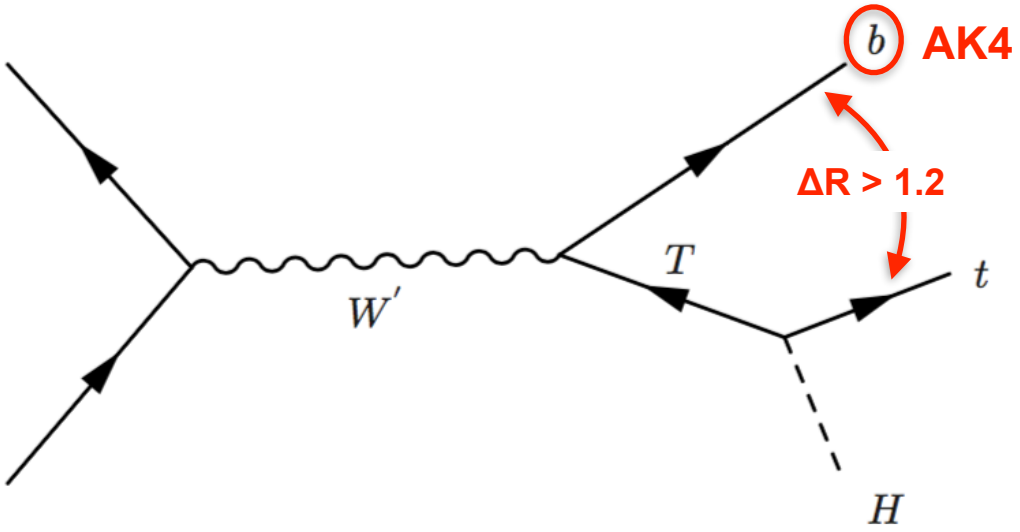


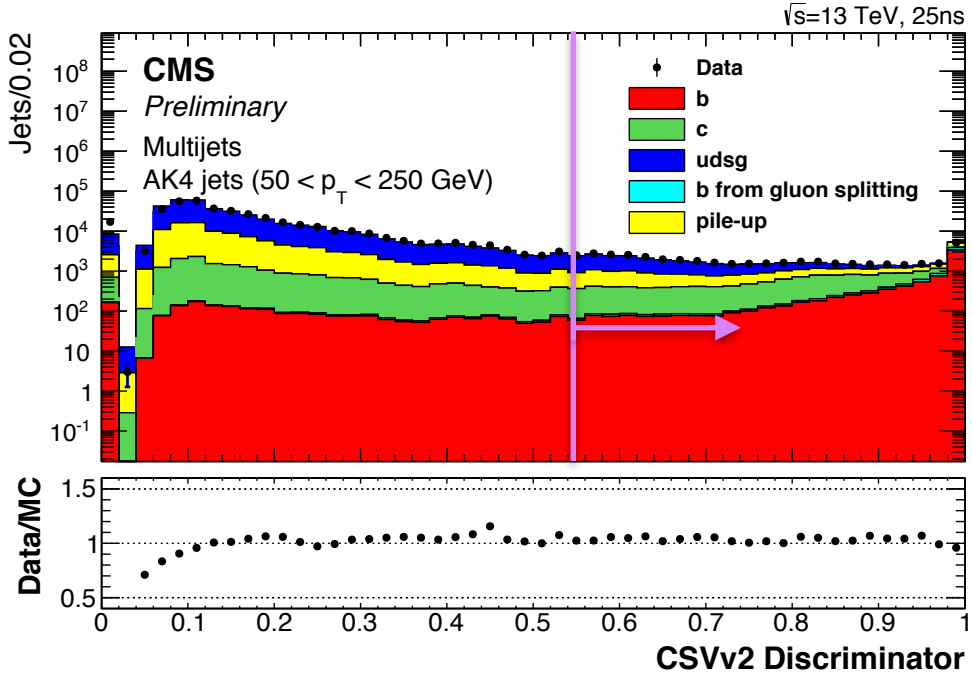


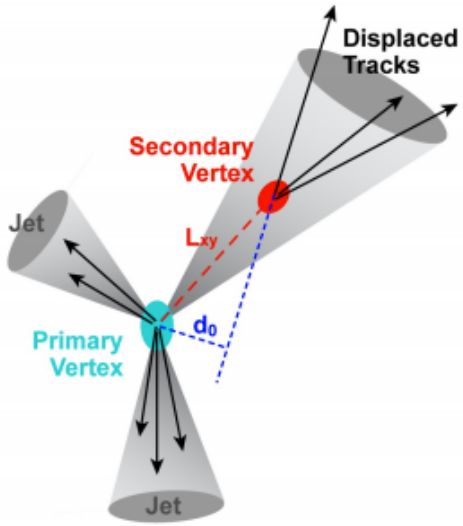


W' → Tb/tB: Signal Selection

1. Higgs jet tag
   - Double b-tagger
2. Top jet tag
   - CMSTTv2 + b-tag
3. b jet tag





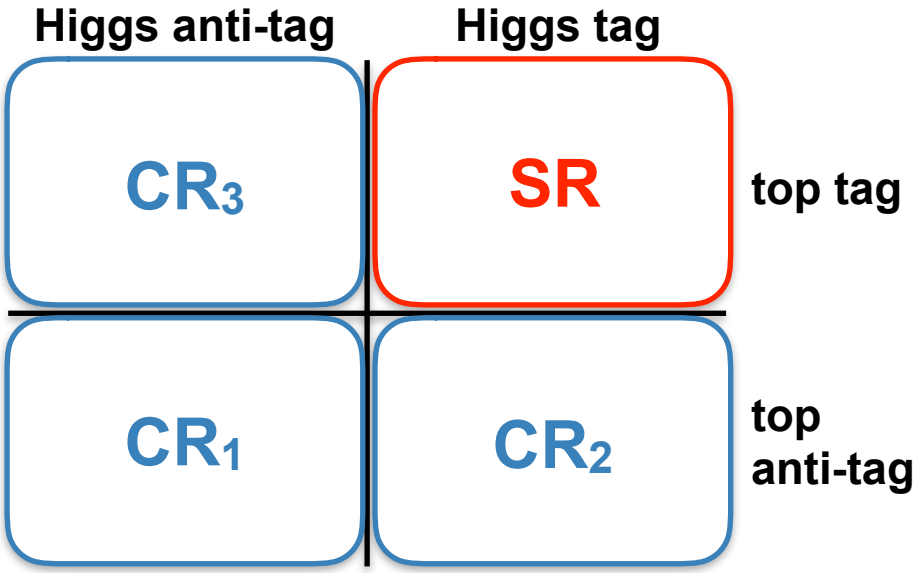


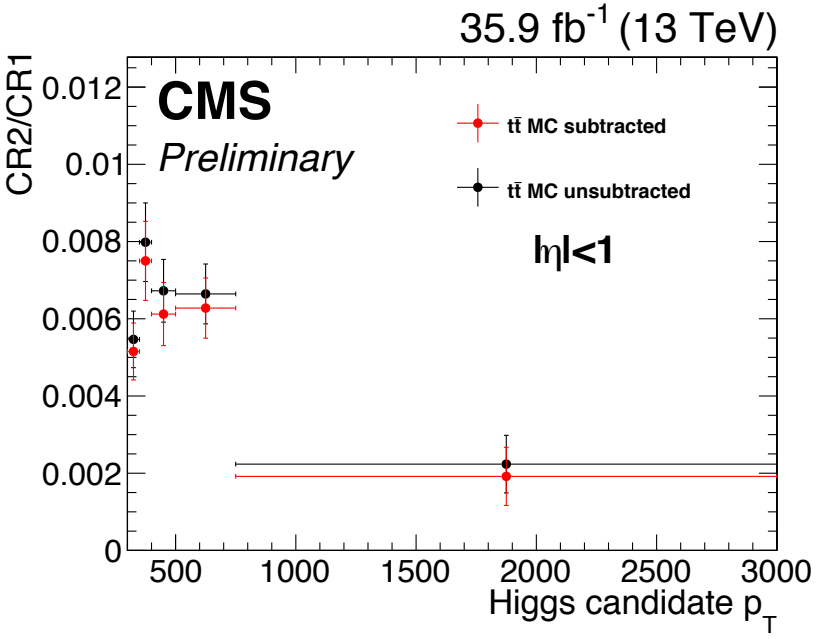
CMS DP-2015/056

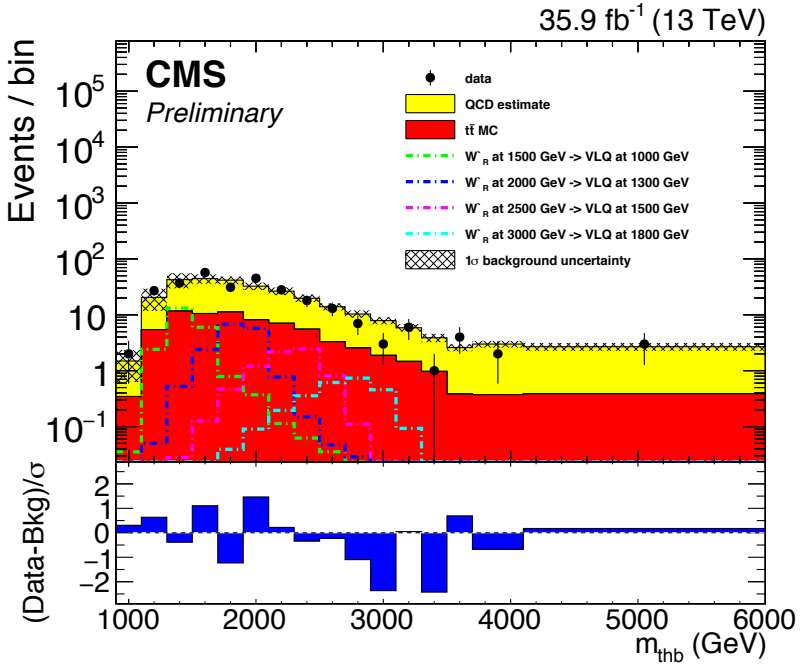
W' → Tb/tB: Background Estimation

- **$t\bar{t}$:** estimated from MC
- **QCD:** estimate from data using transfer function
  - Define Higgs and top anti-tag regions

$$SR = CR_3 \times \frac{CR_2}{CR_1}$$

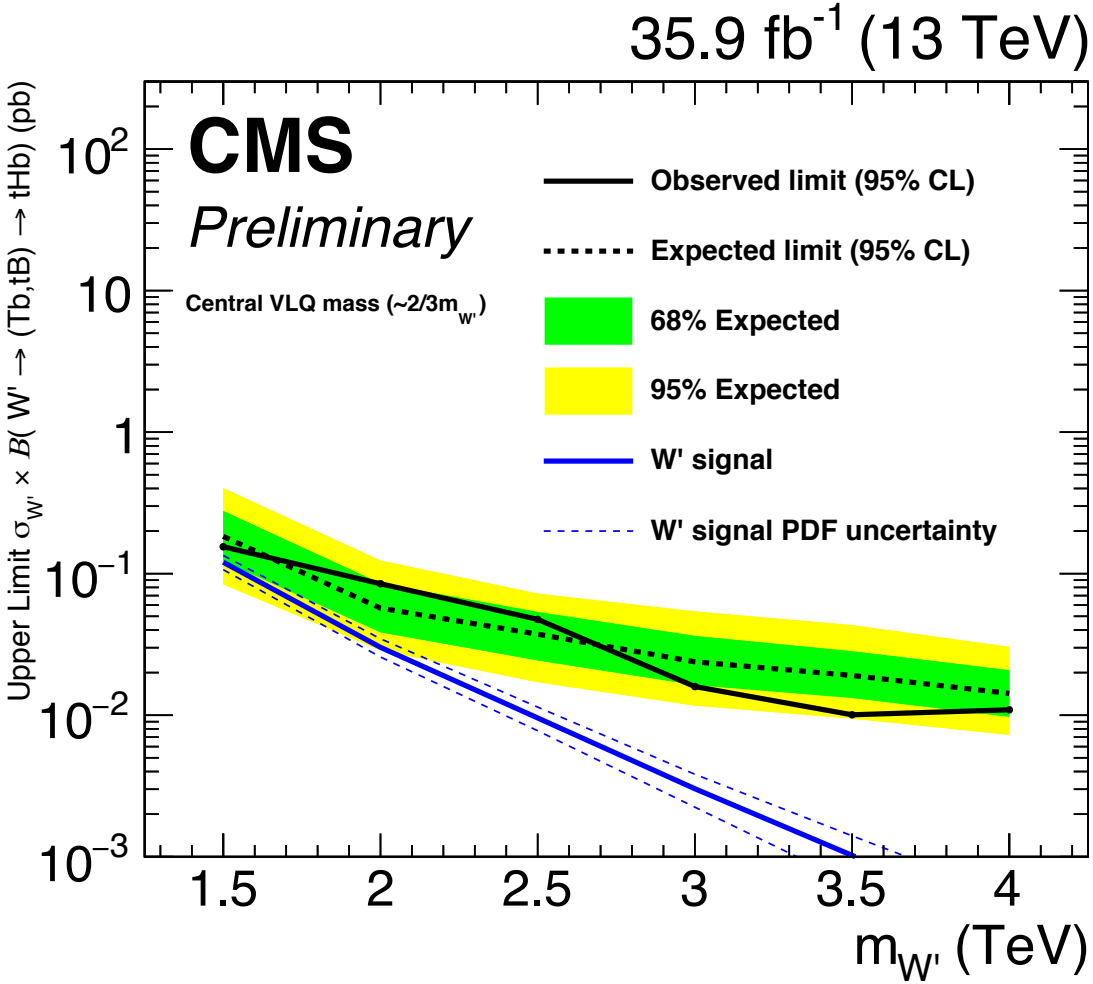






W' → Tb/tB: Central VLQ Mass Results

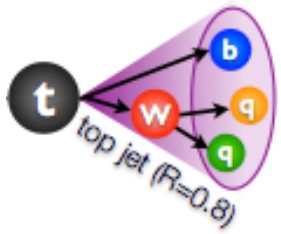
- New for BOOST!
- No limits set yet
  - Could benefit from new, advanced taggers



10.1103/PhysRevD.97.012007

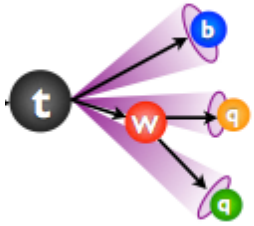
SUSY top tagging method

- Want to tag tops over a large p_T range
  1. First tag fully merged tops
     - Use standard PUPPI AK8 jet tagging (m_{SD} + τ_{32})



  2. Next, tag partially merged tops
     - PUPPI AK8 W-tag
     - Combine with nearby AK4
     - m_t (AK8 + AK4) and m_W/m_t (AK8/(AK4+AK8)) requirements

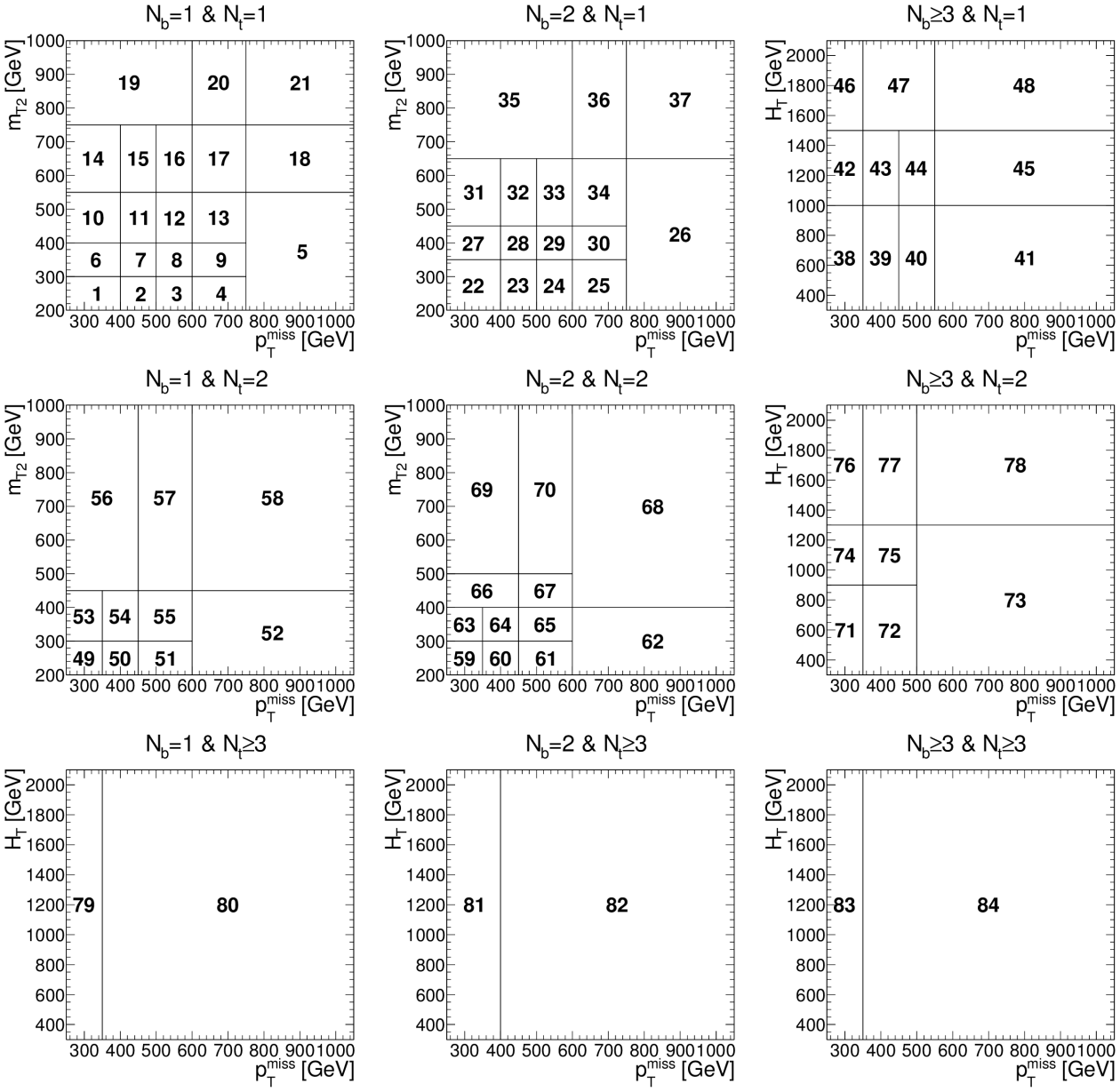
  3. Use MVA to tag 3 resolved AK4 jets
     - Input variables: Jet p, Jet CSV, Jet QGL, etc
     - ΔR and m_t requirements
     - One b-jet maximum



  4. Combine

10.1103/PhysRevD.97.012007

SUSY with Top Tagging: SR categories



10.1103/PhysRevD.97.012007

SUSY with Top Tagging

- Main Backgrounds
  - $t\bar{t}$/W + jets
    - Data-driven estimate using translation factor
  - $Z \rightarrow \nu\nu$
    - Estimated by data-corrected MC
  - QCD
    - Data-driven estimate using data-normalized translation factor derived from MC
- Sub-dominant backgrounds (MC)
  - $t\bar{t}Z$
  - Diboson
  - Single t
  - Triboson