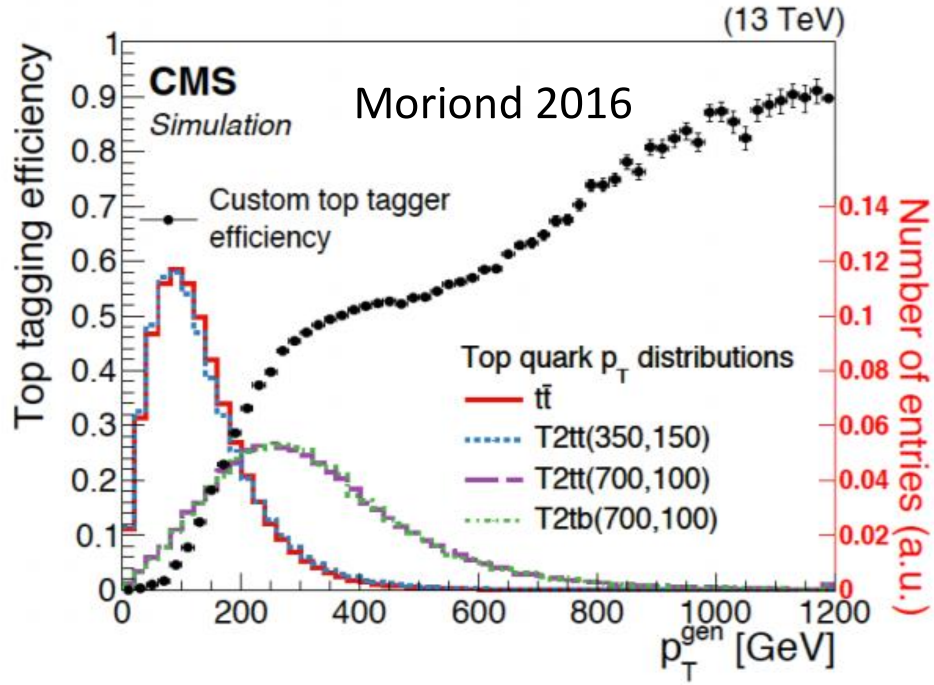
Search for Top Squarks Using Top Tagging at CMS

Joe Pastika
On behalf of the CMS Collaboration

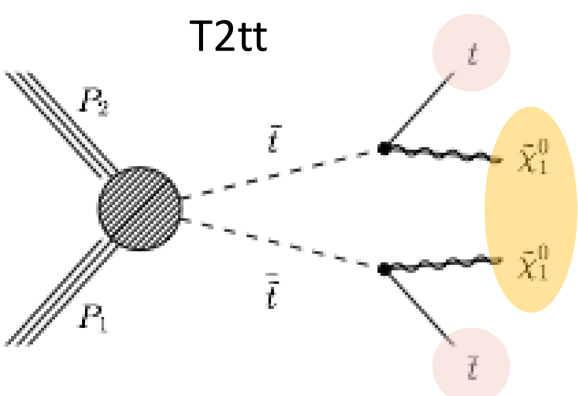
BAYLOR UNIVERSITY

Analysis strategy

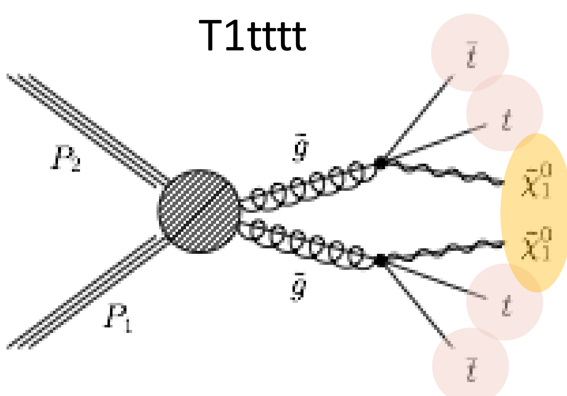
- All-hadronic analysis using top tagging
  - Custom tagging algorithm
  - Designed for good efficiency for large range of top p_T
- Search regions binned in N_t, N_b, p_T^{miss}, M_{T2}/H_T
- Interpret results using Simplified Model Spectrum (SMS)
- Link to full results



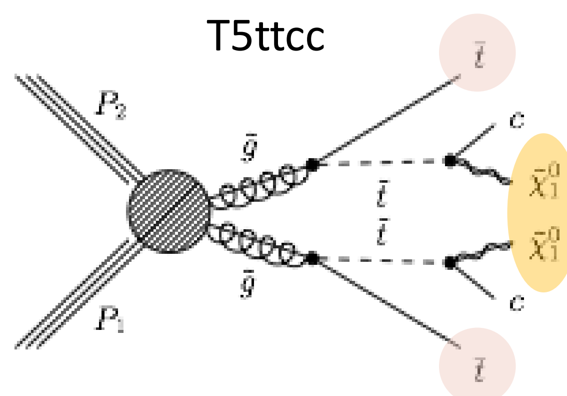
T2tt



T1tttt



T5ttcc

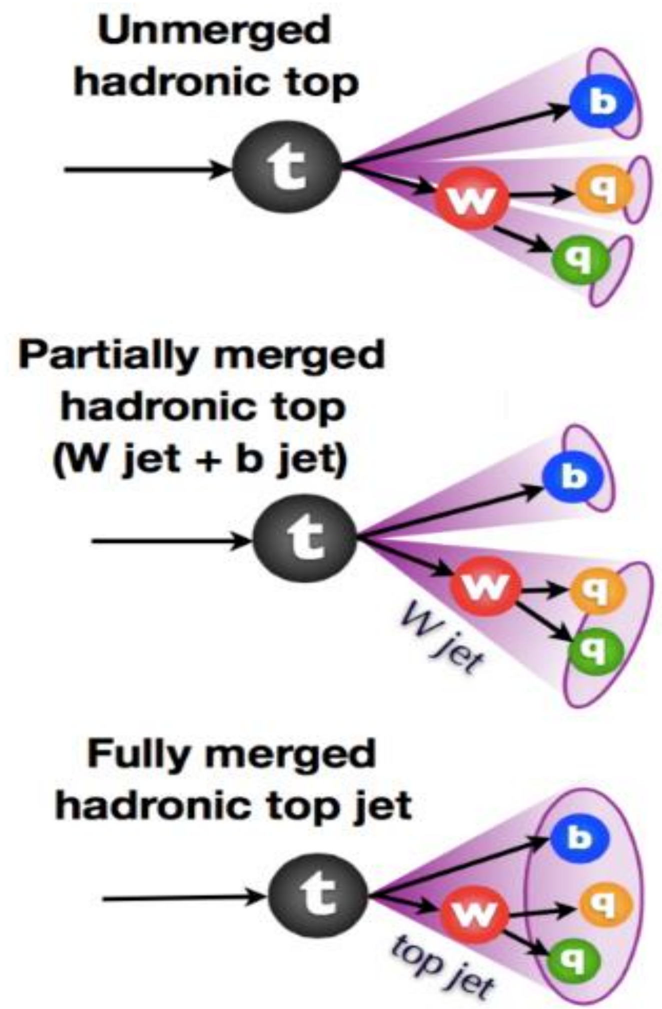


Top Tagger: Overview

Top tagging algorithm

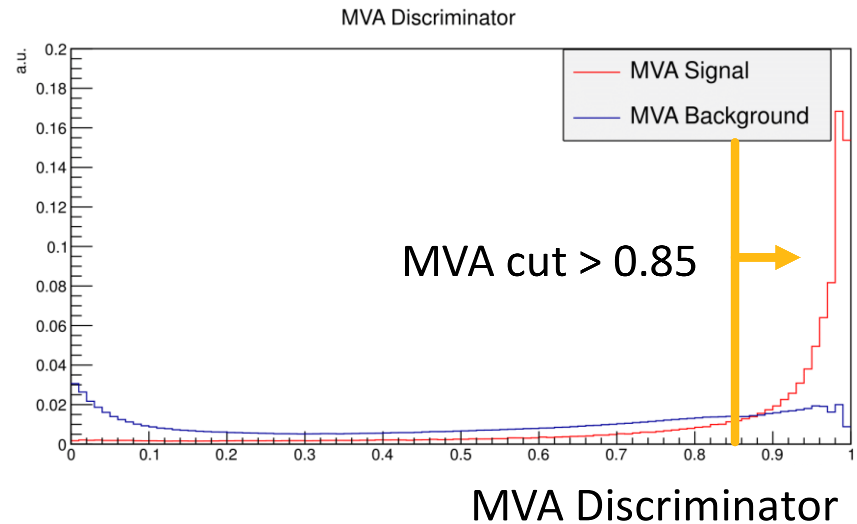
- First tag fully merged tops
  - Using AK8 boosted top tagging
  - Discriminates using soft-drop mass and sub-jettiness variables
- Tag medium p_T tops with W+jet category
  - AK8 as a boosted W (loose working point)
  - Combine with a nearby AK4 jet
  - Require combined mass consistent with top and AK8 mass to combined mass ratio to be consistent with M_W/M_t
- Tag low p_T tops with combinations of 3 resolved AK4 jets with MVA

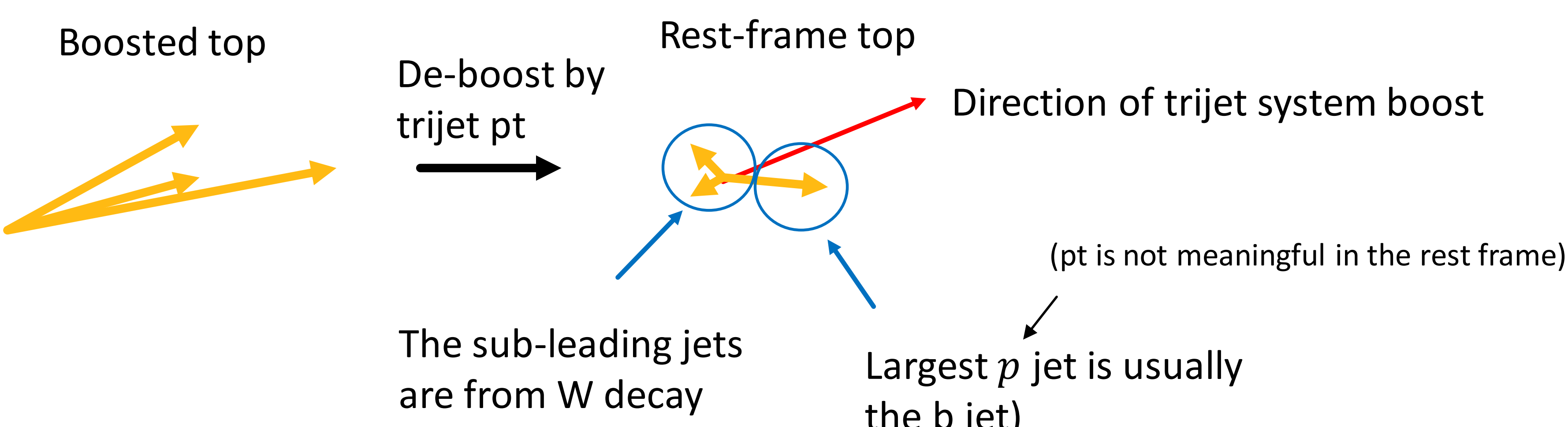
- These three categories are combined
  - Fully merged are tagged first, then W+jet, then resolved, removing jets from consideration form subsequent steps



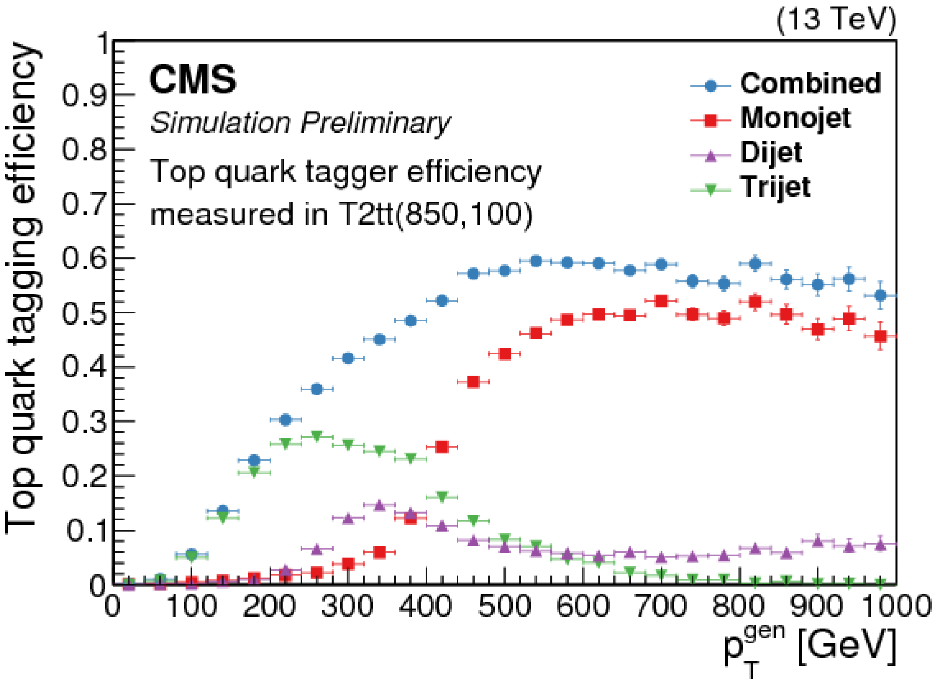
Top Tagger: Resolved Top

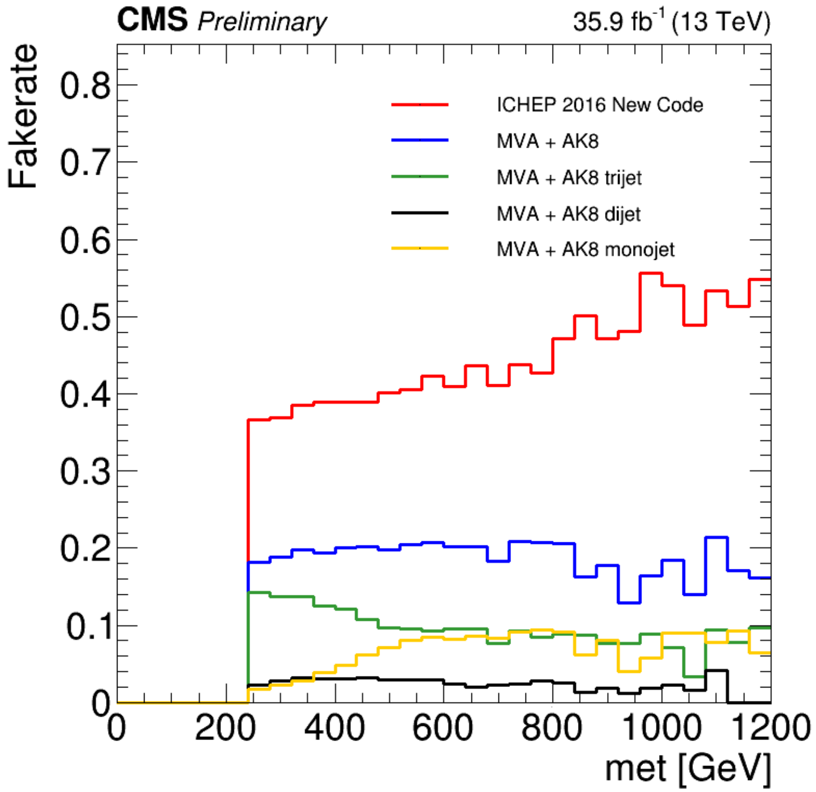
- To avoid overlap : AK4 jets that are $\Delta R < 0.4$ matched to an AK8 jet subjet are removed as input of resolved top tagger
- Resolved Top – All combinations of 3 AK4 jets
  - $p_T > 30$ GeV and $|\eta| < 5$)
  - $\Delta R < 1.5$
  - Trijet mass between 110 and 250 GeV
- MVA training on jet properties (Input variables : tri-jet mass, restframe angular seperations, p, b-tagging discriminator, quark-gluon discriminator)





Top Tagger: Performance





Similar efficiencies compare with 2016 ICHEP Tagger, but much lower mistag rate

Baseline Selection

- Jets and p_T^{miss}:
  - $N_{jets}(p_T > 50) \geq 2$ (AK4 jets)
  - $N_{jets}(p_T > 30) \geq 4$ (AK4 jets)
  - p_T^{miss}> 250 GeV
  - H_T> 300 GeV
  - $\Delta\phi\left(j_{1,2,3}, p_T^{miss}\right) > 0.5, 0.3, 0.3$
  - $N_b \geq 1$
- Veto e, μ, and tracks consistent with τ
- Top reconstruction:
  - $N_t \geq 1$ with top tagger
  - $M_{T2} > 200$ GeV

Binning variables

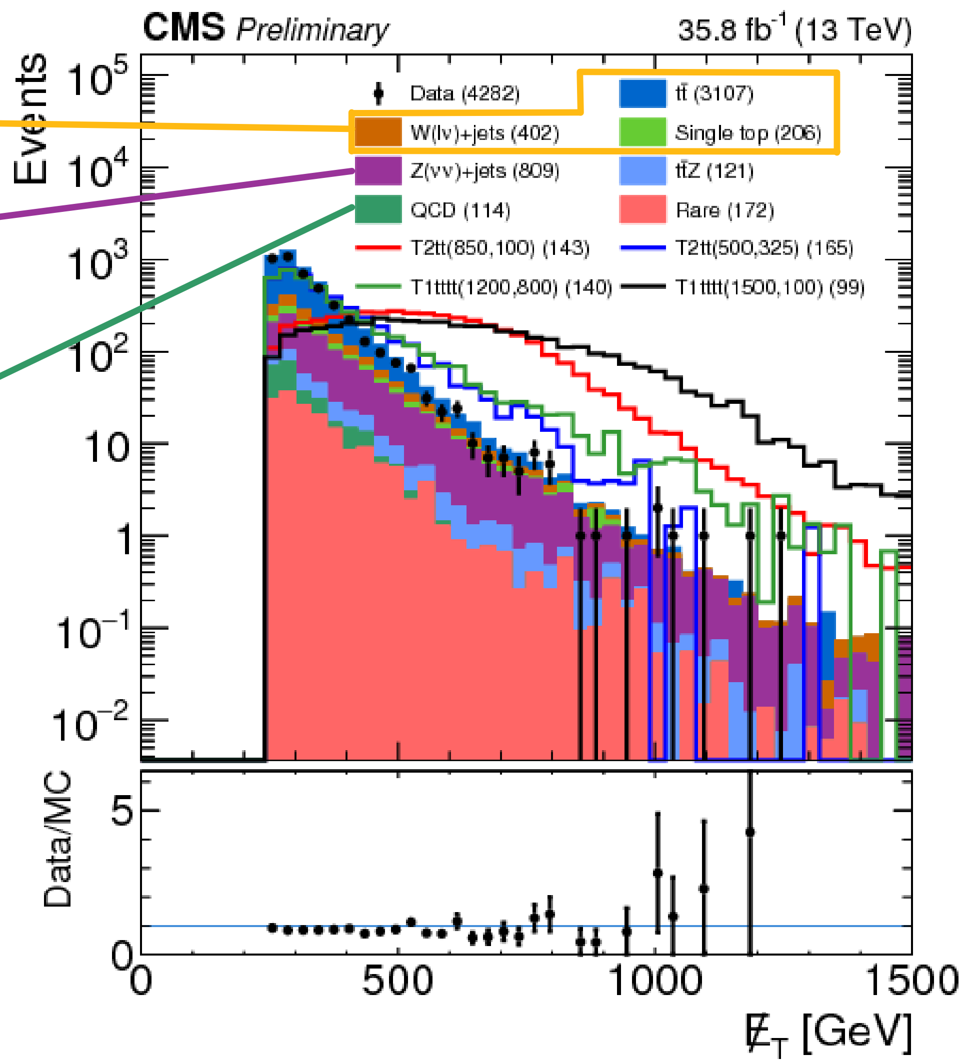
N_t \ N_b	**1**	**2**	**≥3**
1	p_T^{miss}, M_{T2}	p_T^{miss}, M_{T2}	p_T^{miss}, H_T
2	p_T^{miss}, M_{T2}	p_T^{miss}, M_{T2}	p_T^{miss}, H_T
≥3	p_T^{miss}, H_T	p_T^{miss}, H_T	p_T^{miss}, H_T

$$M_{T2} = \min_{\vec{p}_T^{miss(1)} + \vec{p}_T^{miss(2)} = \vec{p}_T^{miss}} \left[\max\left(M_T^{(1)}, M_T^{(2)}\right)\right]$$

Background Estimations

- Main Backgrounds
  - $t\bar{t}/t/W + jets$
    - Estimated by data driven approach using translation factors
  - $Z \rightarrow \nu\nu$
    - Estimated by data corrected MC
  - QCD
    - Estimated in data sideband with data-normalized translation factors derived from MC

- Sub-dominant backgrounds
  - $t\bar{t}Z$
  - Diboson
  - Triboson
  - Estimated from MC



$t\bar{t}/W$ Background Estimate

- Estimated using single lepton (e & μ) data side band
  - Matches baseline selection but with lepton veto replaced with explicit selection of one e or μ
  - $m_T(p_T^{miss}, \ell) < 100$ GeV
- Translation factors are derived per search bin
  - Derived using MC
  - Move estimate from single lepton control samples to signal sample
- The estimate is derived separately using the e and μ control regions and both are used to constrain the final estimate
- The dominant systematic uncertainty comes from the statistics of the translation factor

Single lepton control sample
- 1 e or μ
- $m_T(p_T^{miss}, \ell) < 100$ GeV

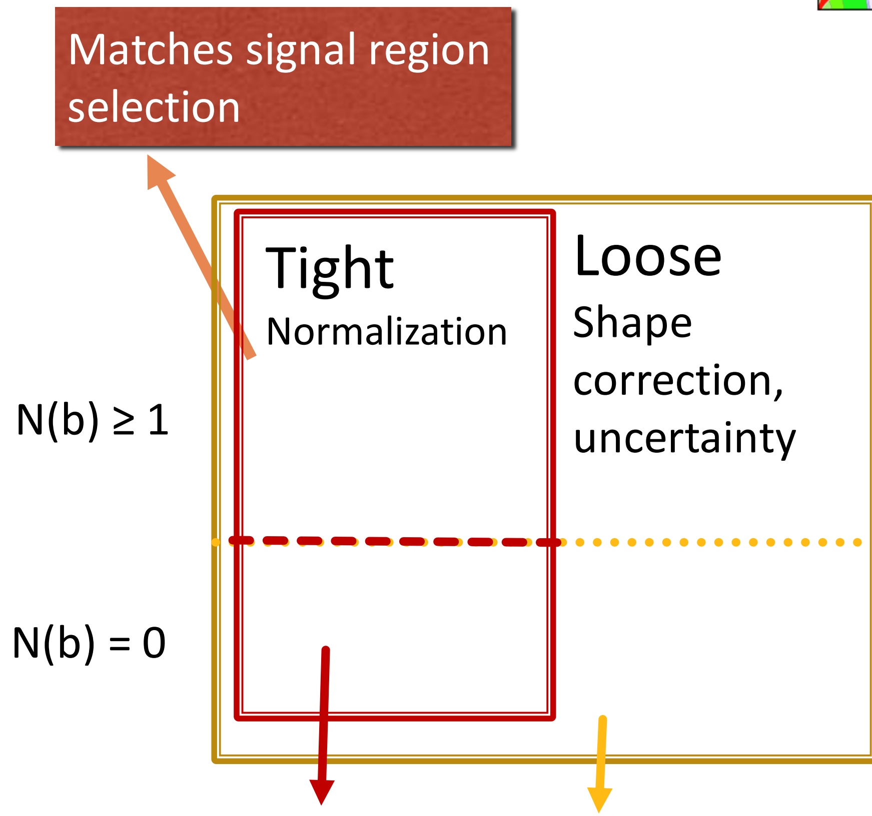
Signal Sample
- Baseline selection

$$\hat{N}_{SS}^{B} = T^{B}\hat{N}_{CS}^{B}$$

$$T^{B} = \frac{\hat{N}_{SS}^{MC:B}}{\hat{N}_{CS}^{MC:B}}$$

$Z \rightarrow \nu\nu$ Background Estimate

- MC base estimate, validated in $Z \rightarrow \mu\mu$ control sample
  - Require 2 opposite sigh muons
  - $81 < M_{\mu\mu} < 101$ GeV
- DY loose region:
  - $N_j \geq 4$, elec veto, $\Delta\phi$,
  - $p_T^{miss} > 100$ GeV
  - $H_T > 300$ GeV,
  - $N_T \geq 1$
- Tight region adds:
  - $p_T^{miss} > 250$ GeV
  - $M_{T2} > 200$ GeV
- Dominant uncertainty from residual Data/MC disagreement in Loose region

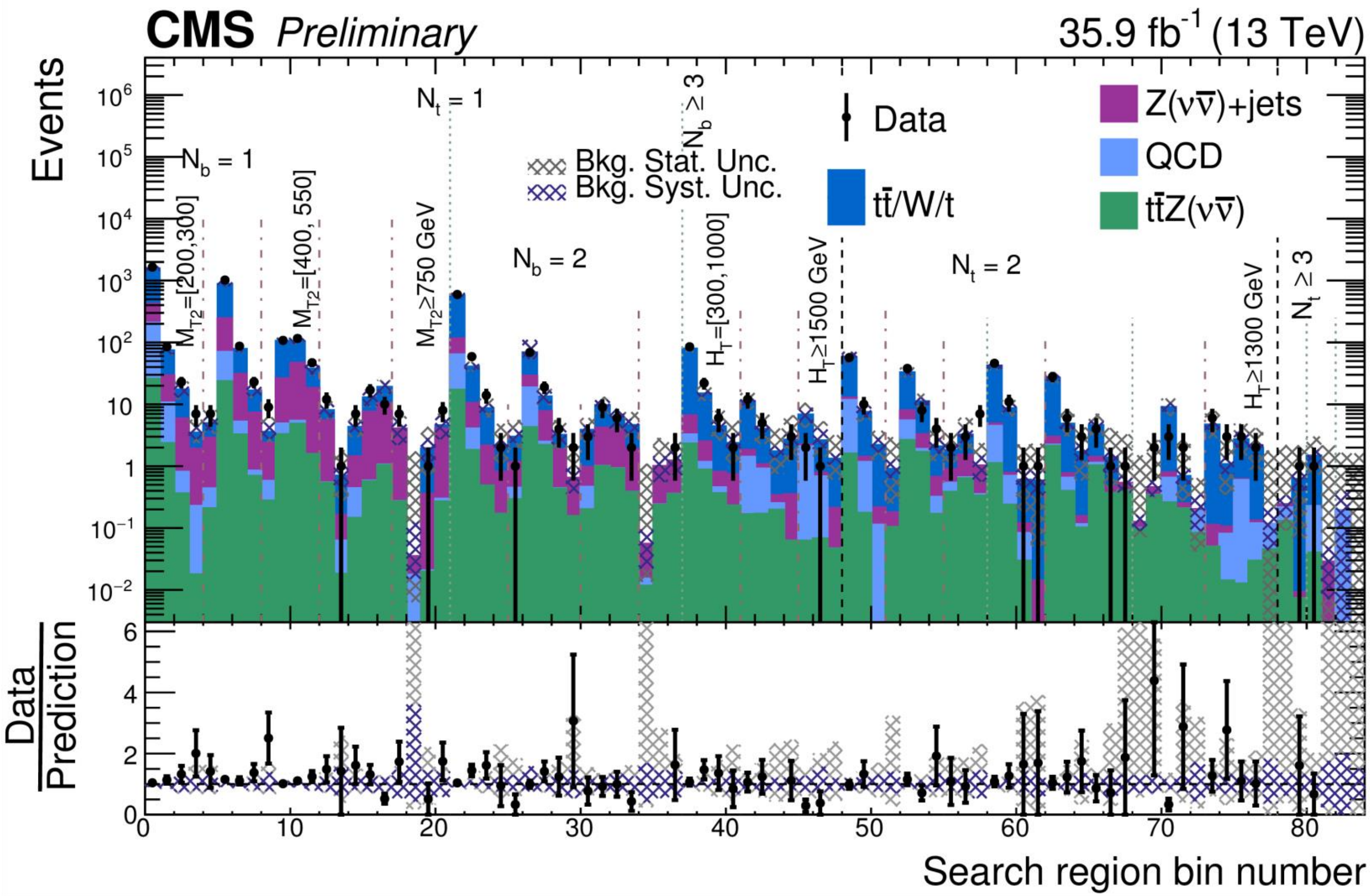


$$\hat{N}_B = \boxed{R_{norm}} \sum_{events \in B} \boxed{S_{DY}(N_{jet})} w_{MC}$$

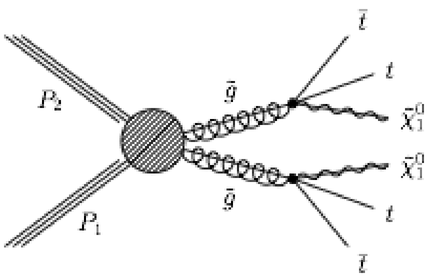
$$w_{MC} = (\sigma \times \mathcal{L}) \epsilon_{trig} w_{btag} w_{pileup}$$

Results

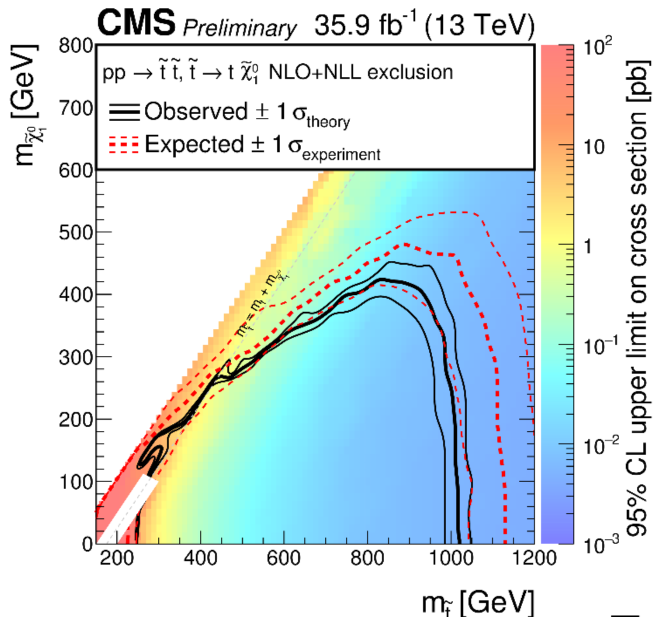
- Background predictions compared with observed data
- No Significant excess observed

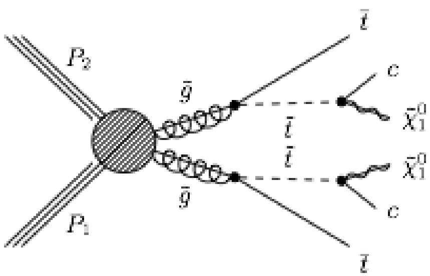


Final Limits

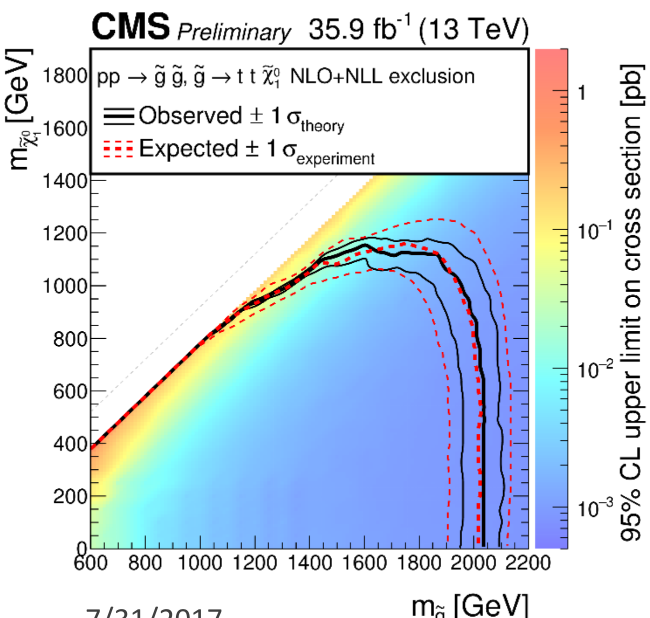


T1tttt

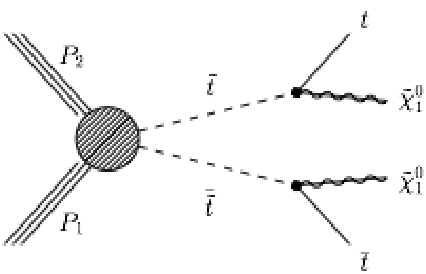


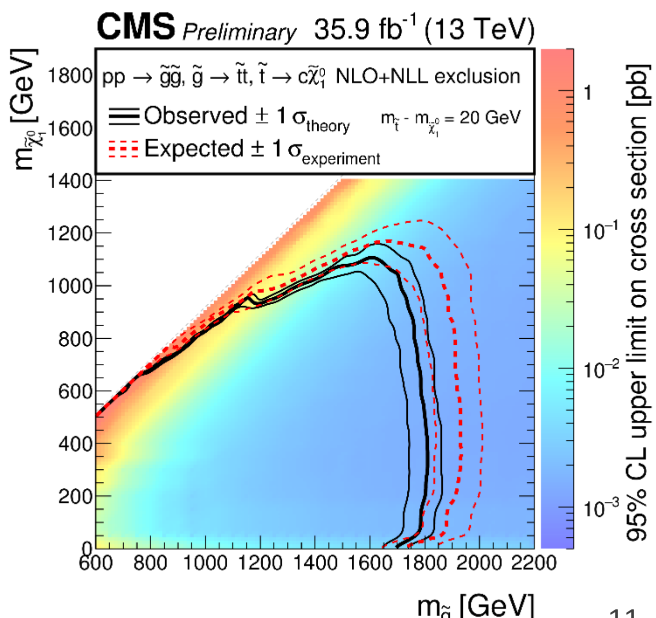


T5ttcc



T2tt

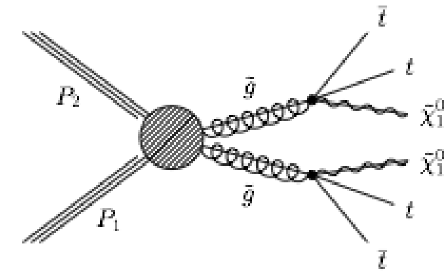




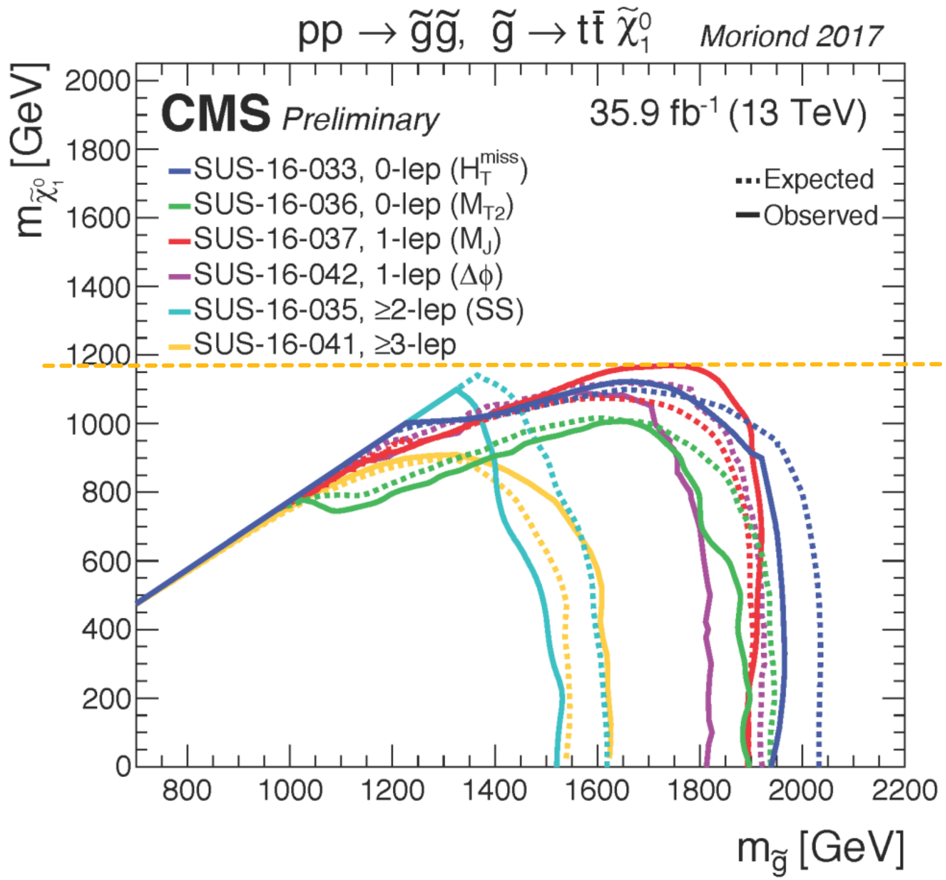
Final Limits

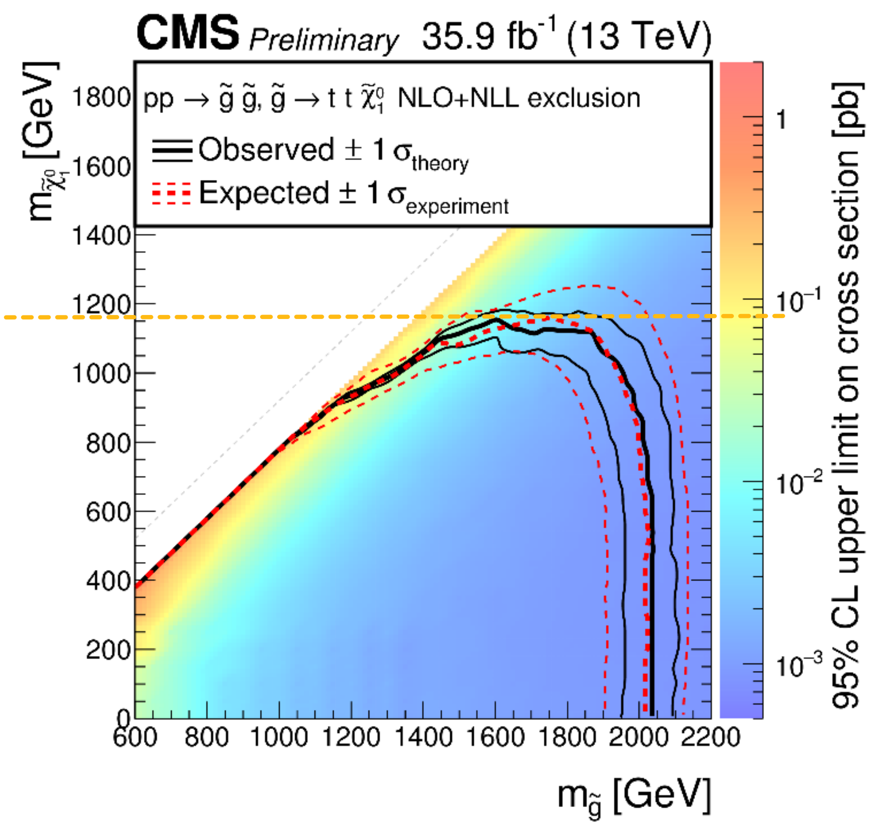
Top tagging particularly helps at high M_χ

T1tttt



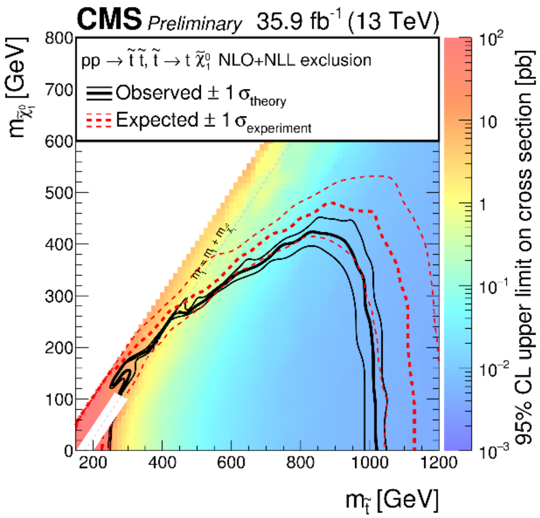
More CMS limits



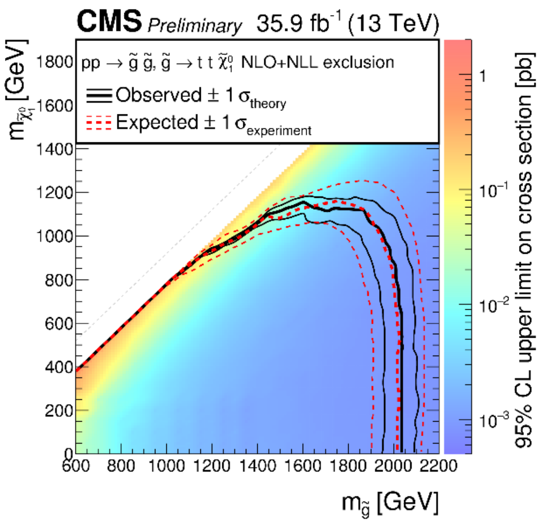


Conclusions

- We have designed an updated top tagging algorithm
  - Similar efficiency as past tagger but ~50% of its mistag rate
  - Tagger code released as standalone package



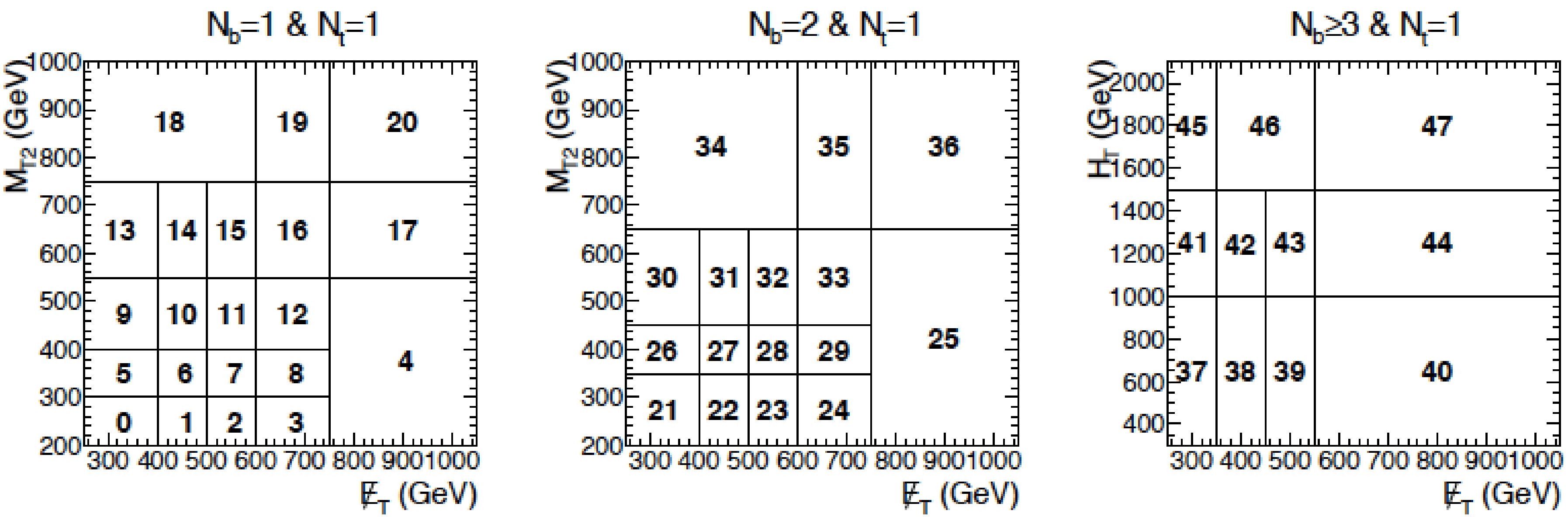
- We have excluded T2tt models up to ~1000 GeV in stop mass and ~400 GeV in LPS mass
- We have excluded T1tttt models up to ~2000 GeV in stop mass and ~1100 GeV in LPS mass



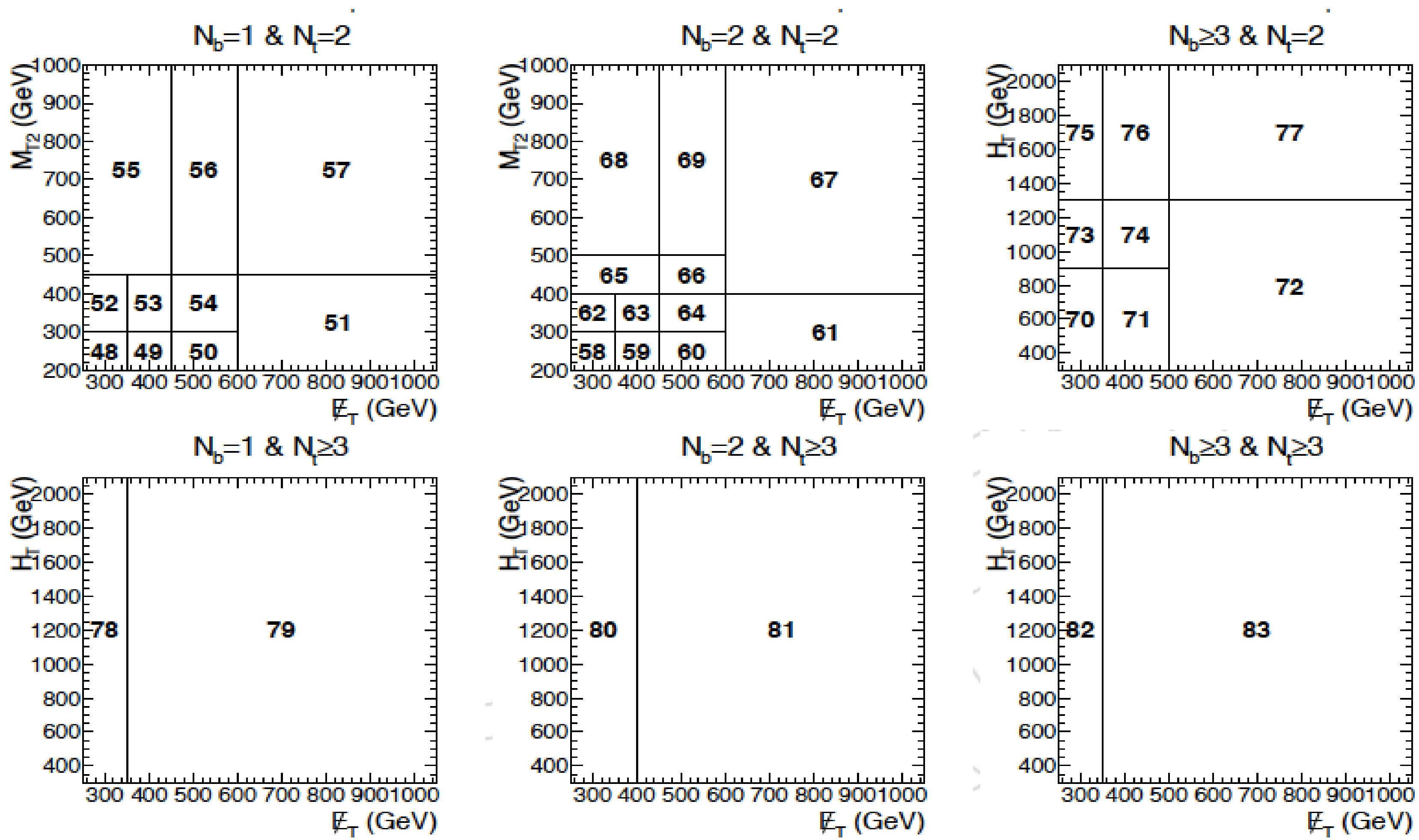
Backup Slides

Search Bins

- 84 search bins in N_b, N_t, p_T^{miss}, M_{T2}/H_T
- $N_b, N_t = 1{,}2$: Targeting T2tt, each block is binned in p_T^{miss}, M_{T2}
- N_b or $N_t \geq 3$: Targeting T1tttt, each block is binned in p_T^{miss}, H_T only



Search Bins

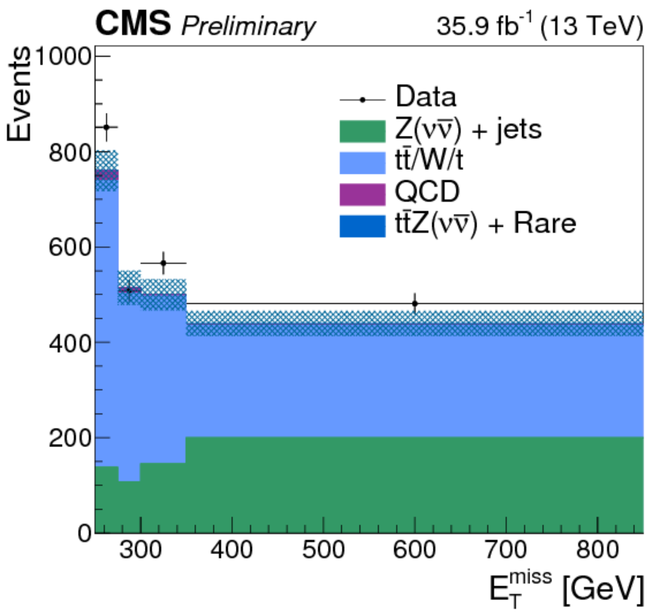


N_b=1 & N_t=2
M_T2 (GeV)
55
56
57
52
53
54
51
48
49
50
E_T (GeV)
N_b=2 & N_t=2
M_T2 (GeV)
68
69
67
65
66
62
63
64
61
58
59
60
E_T (GeV)
N_b≥3 & N_t=2
H_T (GeV)
75
76
77
73
74
72
70
71
E_T (GeV)
N_b=1 & N_t≥3
H_T (GeV)
78
79
E_T (GeV)
N_b=2 & N_t≥3
H_T (GeV)
80
81
E_T (GeV)
N_b≥3 & N_t≥3
H_T (GeV)
82
83
E_T (GeV)

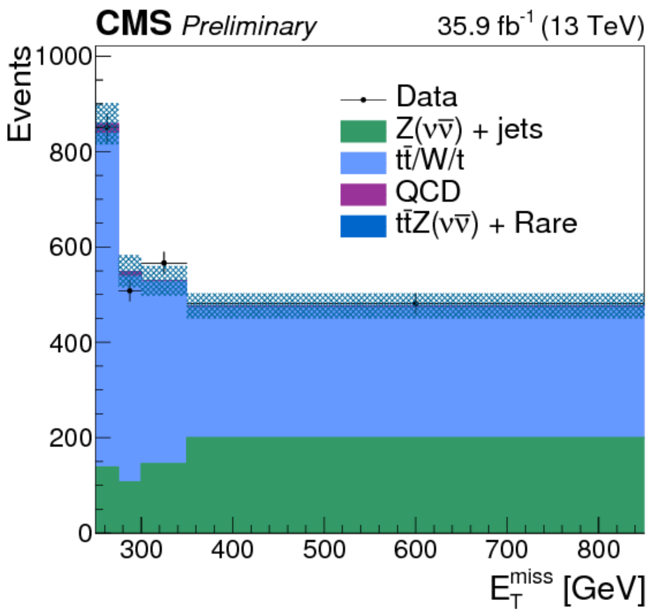
TF Method: Data Validation

- Data sideband selected using full baseline selection except $N_b \geq 2$, $N_t = 0$, and a stricter cut to reduce QCD of $\Delta\phi\left(\vec{p}_T^{miss}, j_{1,2,3,4}\right) > 0.5$
- Cross-check shows good agreement in both electron and muon channel

Electron Channel

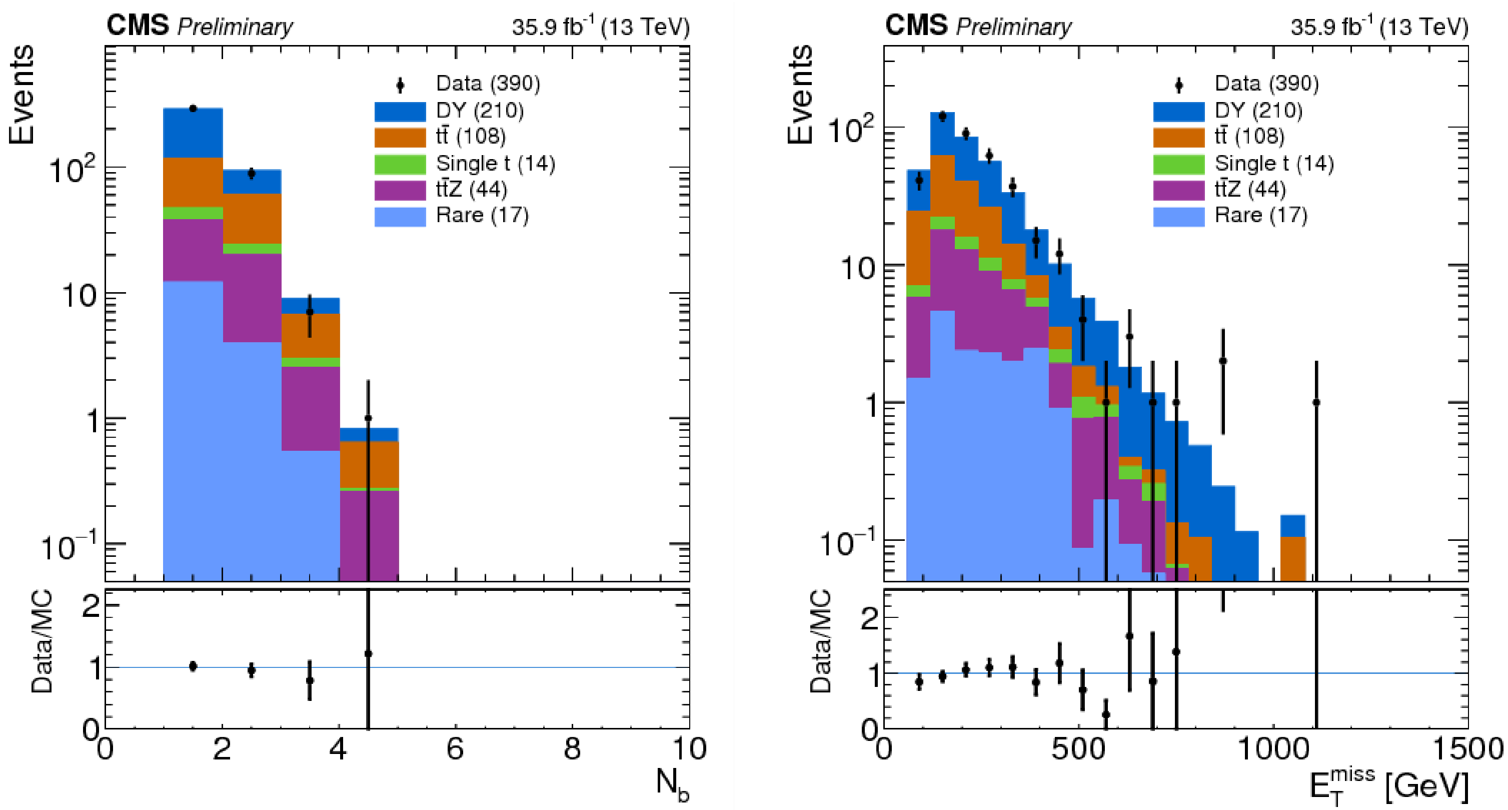


Muon Channel



Z control region after reweighting

The loose DY control region with $N_b \geq 1$ after all weights are applied

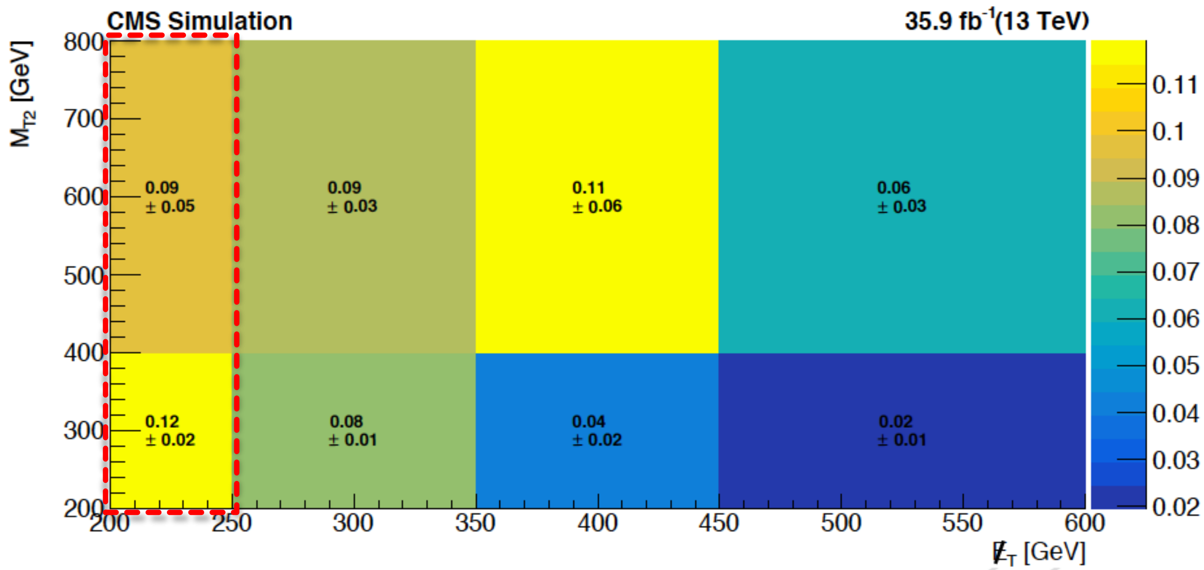


QCD Background

- Use **inverted Δϕ control sample to predict the number of QCD events**, using a translation factor (T-factor) $\boxed{N_{QCD}^{SR} = N_{QCD}^{\Delta\bar{\phi}} \times T_{QCD}}$ $T_{QCD} = \frac{N_{QCD}^{\Delta\phi}}{N_{QCD}^{\Delta\bar{\phi}}}$
  - The T-factor is measured in data in a sideband region ($200 < E_T^{miss} < 250$ GeV), and extrapolate in higher E_T^{miss} region using MC results
  - MC gives the shape and systematics and data sideband gives the normalization

- $N_t \leq 2$ and $N_b \leq 2$: 2D in (M_{T2}, E_T^{miss})
- $N_t \geq 3$ or $N_b \geq 3$: 1D in E_T^{miss}

p_T^{miss} [200,250]	p_T^{miss} [250,Inf]
0.113	0.095



- Red circled numbers are compared to measurement directly from data → determine normalization to the T-factor numbers for the higher E_T^{miss} bins