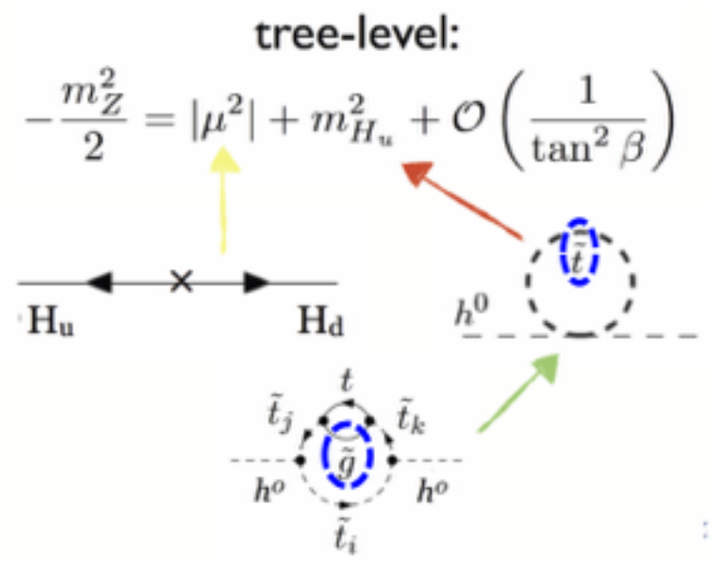
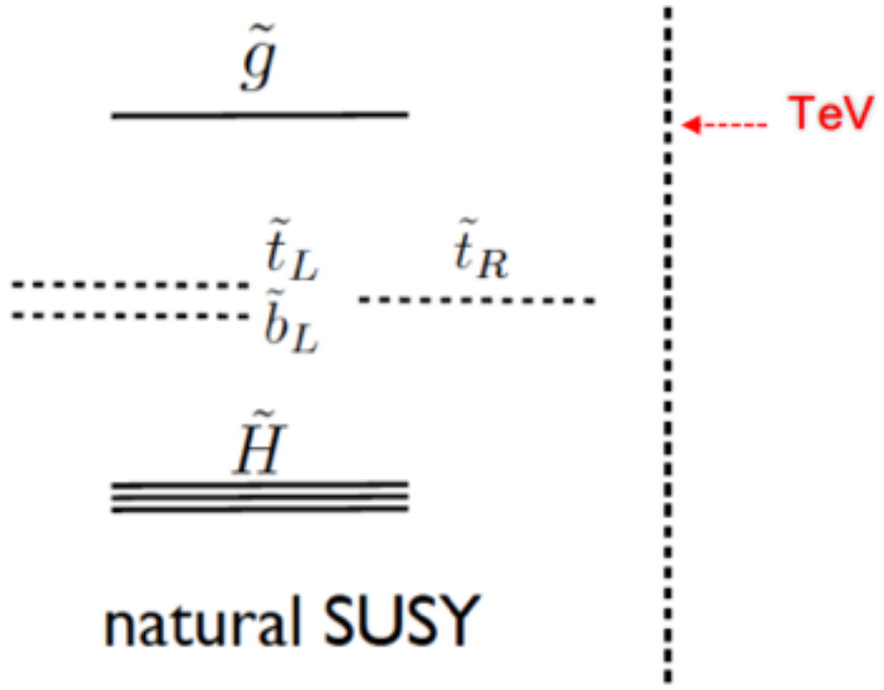
Search for SUSY with Hadronic Top Quark Tagging

By:
Scarlet Norberg
University of Puerto Rico

Motivation

- We know that the Standard Model (SM) is an incomplete theory
- Supersymmetry (SUSY) is one of the popular theories that can explain for example the: “hierarchy problem” & could provide a dark matter candidate (R-conserved)
- In “natural” (not-so-fine-tuned) SUSY argument: light gluino and top squark (stop) are essential and should be accessible at the LHC energy



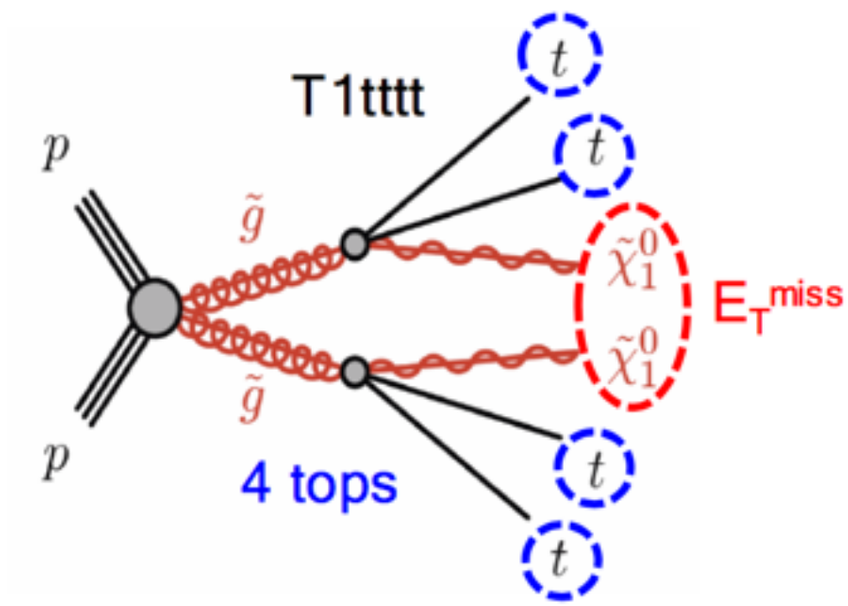


Papucci, Ruderman, Weiler, arXiv:1110.6926

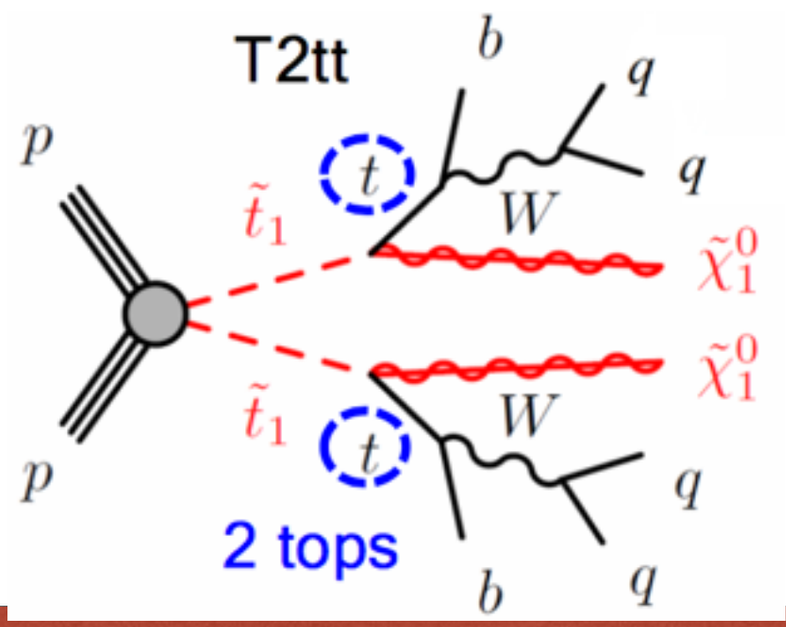
Stop search Strategy

- Two search scenarios(based on SUS-16-050 PAS)
  - Direct Stop production
  - Or gluino mediated stop production
- Top Tagging of a wide p_T range of tops is needed

Gluino Mediated Stop Production

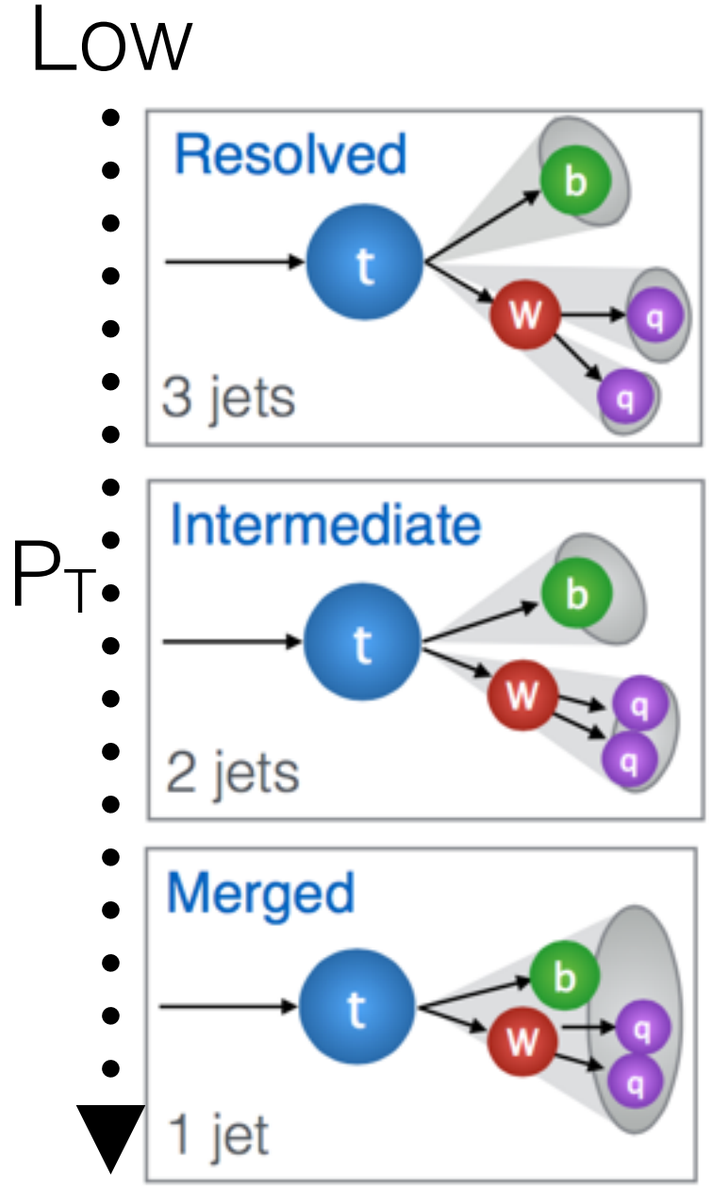


Direct Stop Production

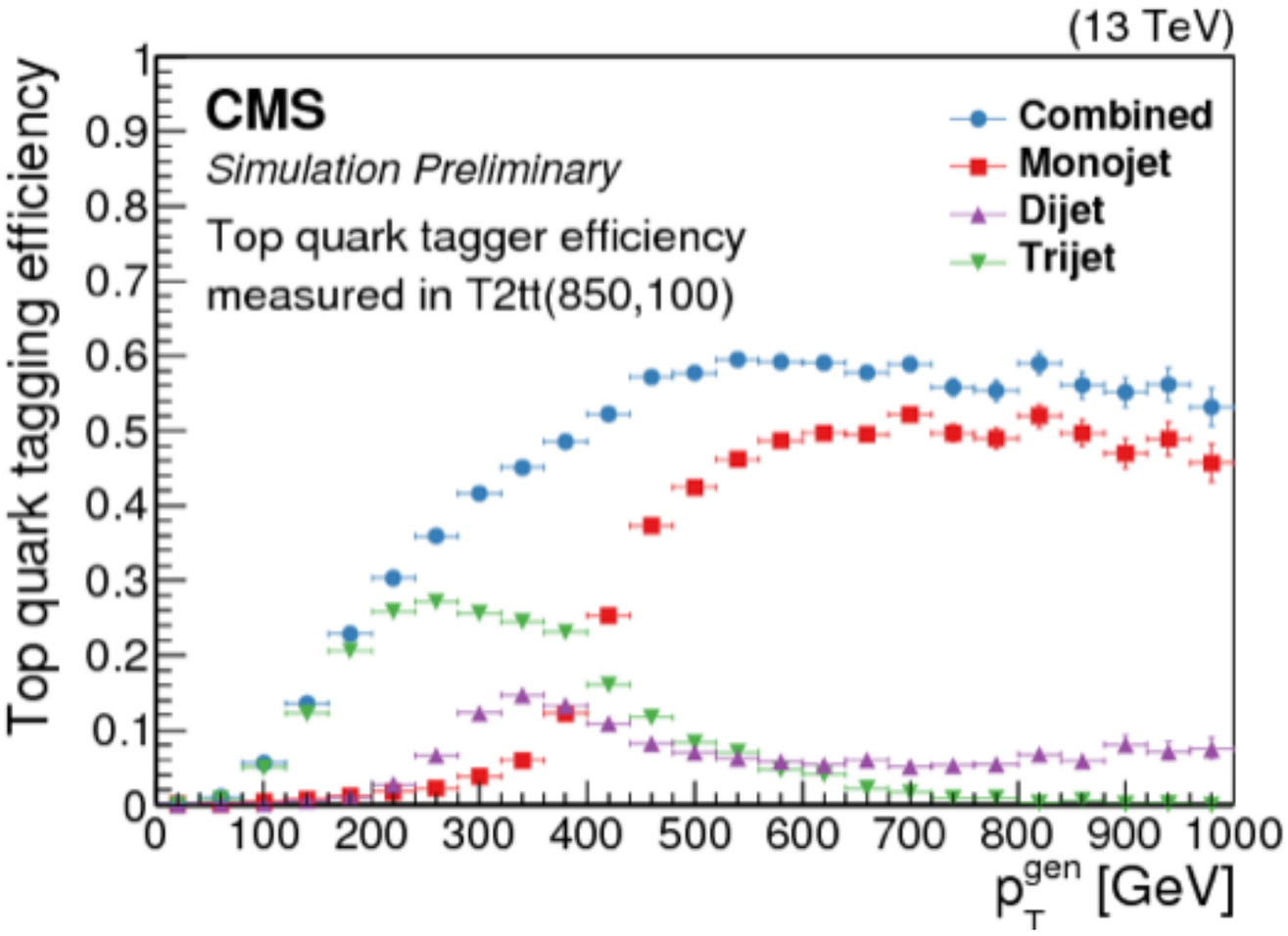


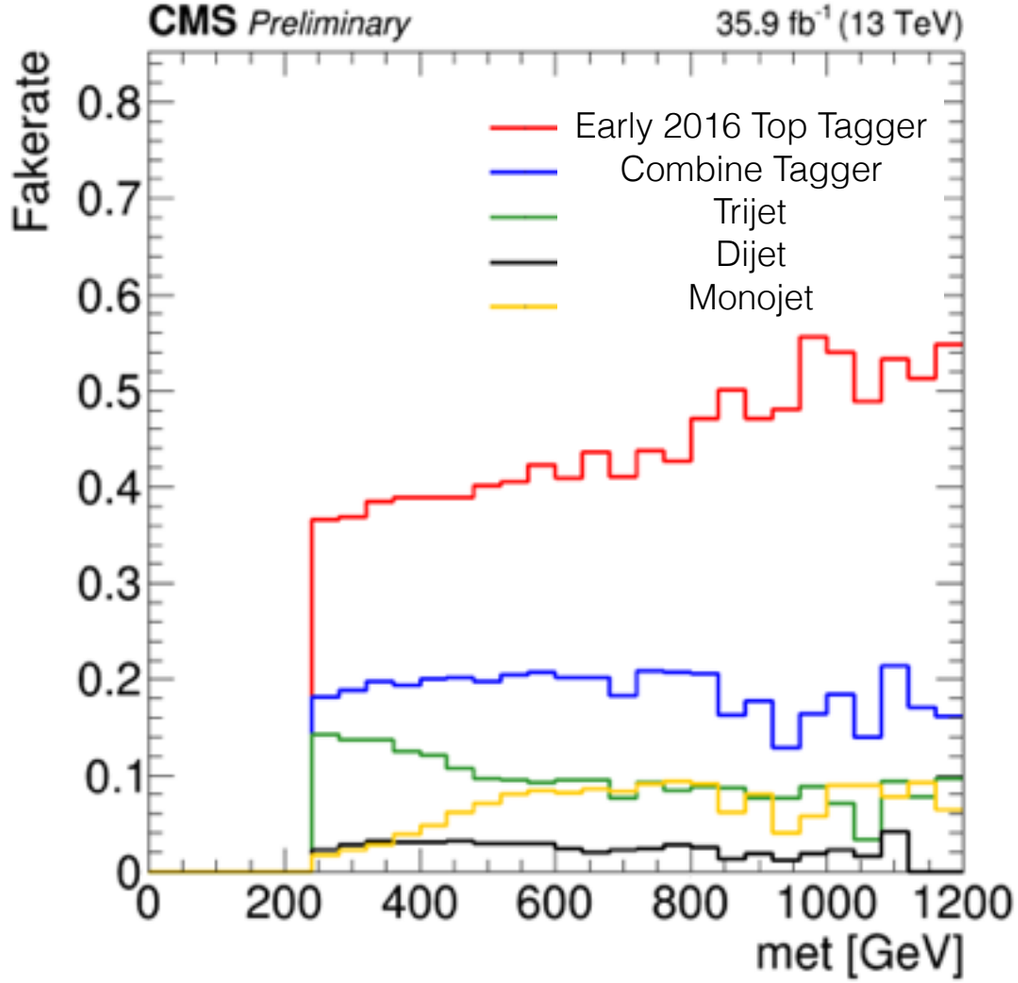
Top Tagger

- Consider only hadronic top quark decays (t→bW, W→qq)
- Resolved category selects tops reconstructed from combinations of 3 anti-kt R=0.4(AK4) jets. Signal is discriminated from background using a Random Forest decision tree using basic jet properties
- Intermediate category looks for anti-kt R=0.8 (AK8) tagged boosted W's with another AK4 jet
- Merged category looks at AK8 jets with pt>400 GeV using PUPPI. Softdrop grooming and N-subjettiness variables are used reject background jets
- Avoid overlaps between AK4 and AK8 jets by removing AK4 jets that are within ΔR=0.4 of a softdrop subjet of the tagged W or top AK8 jet.



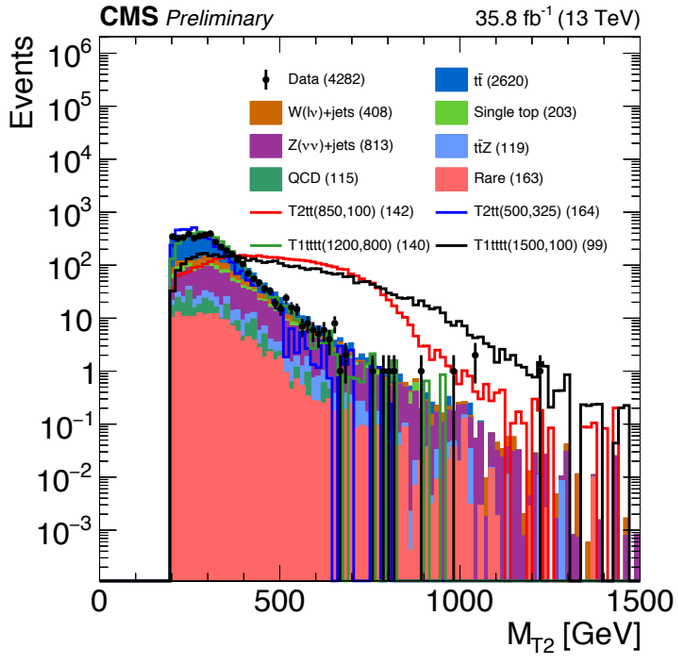
Top Tagger: Performance

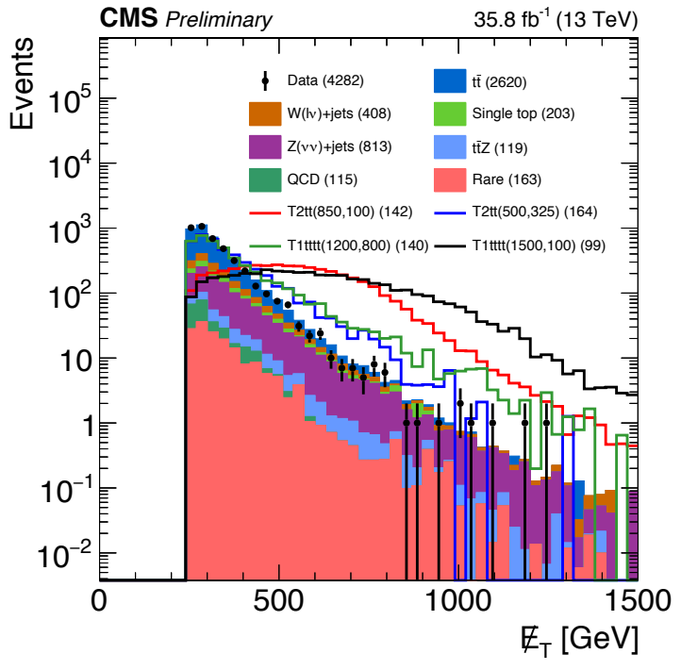


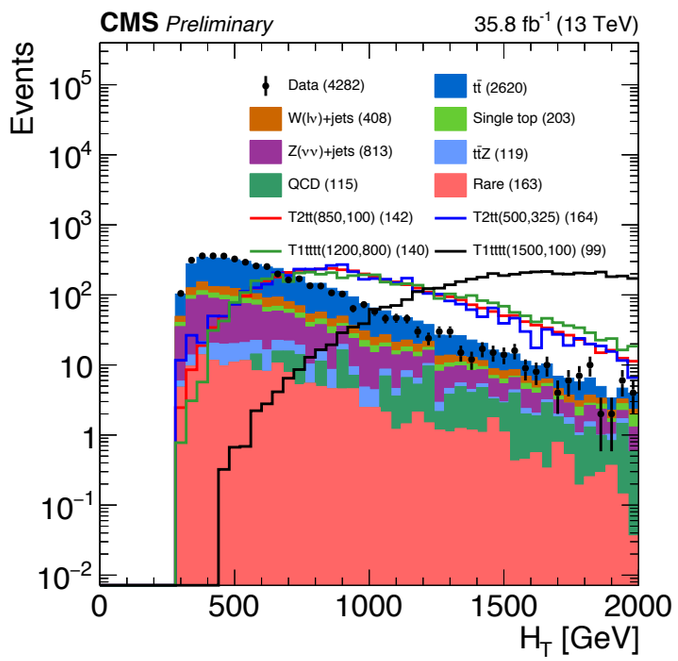


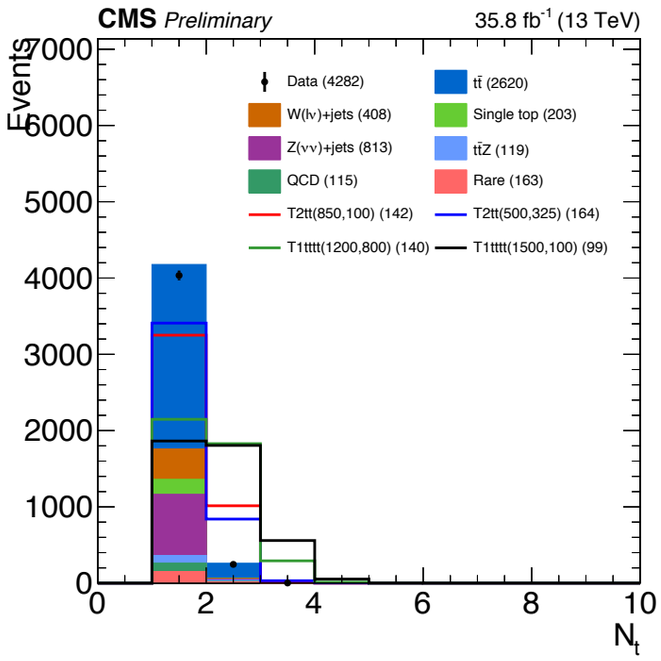
- Performance of the Top tagger used for the analysis
- Similar efficiencies compared with other Top Tagger, but much lower Mistag rate

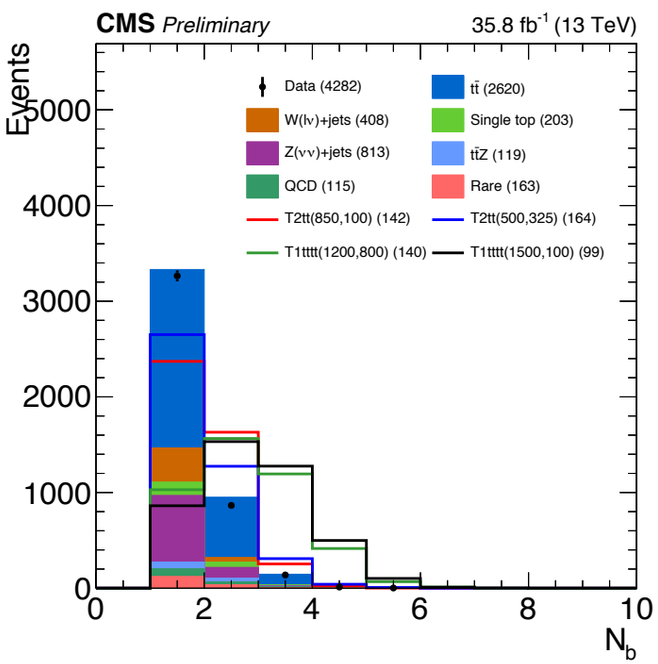
Kinematic Variables Used







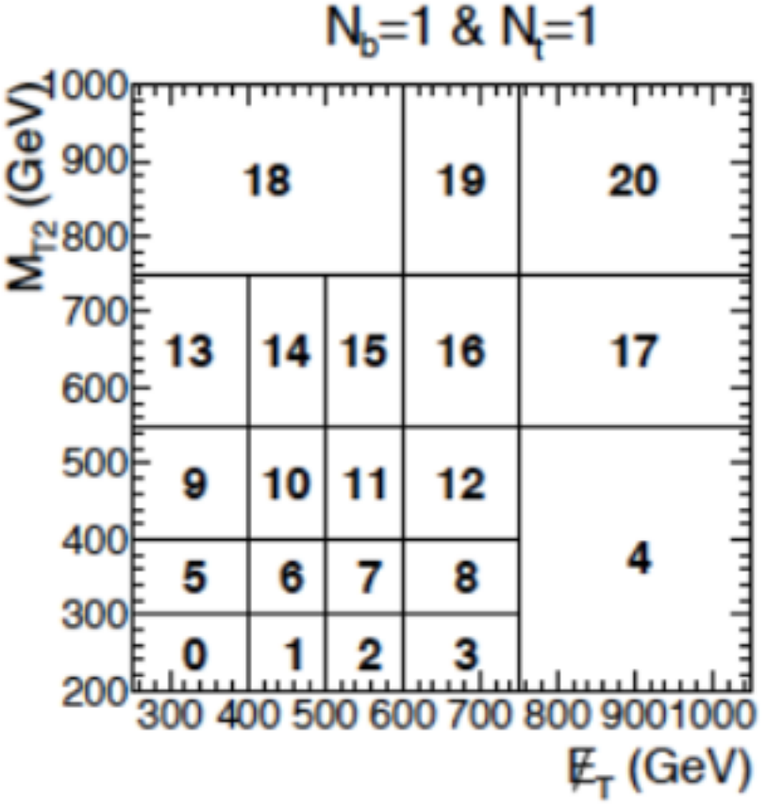


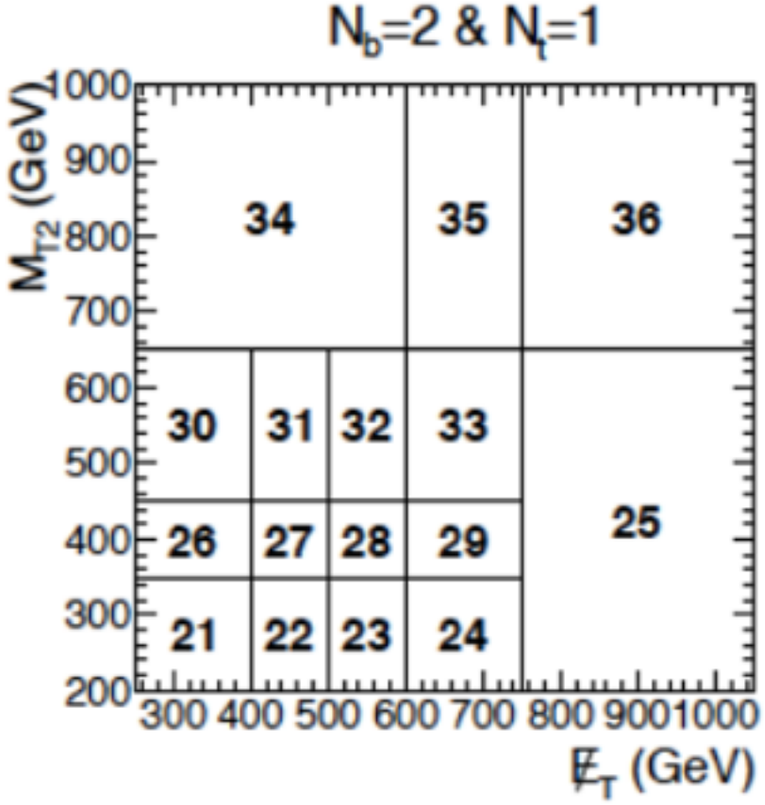


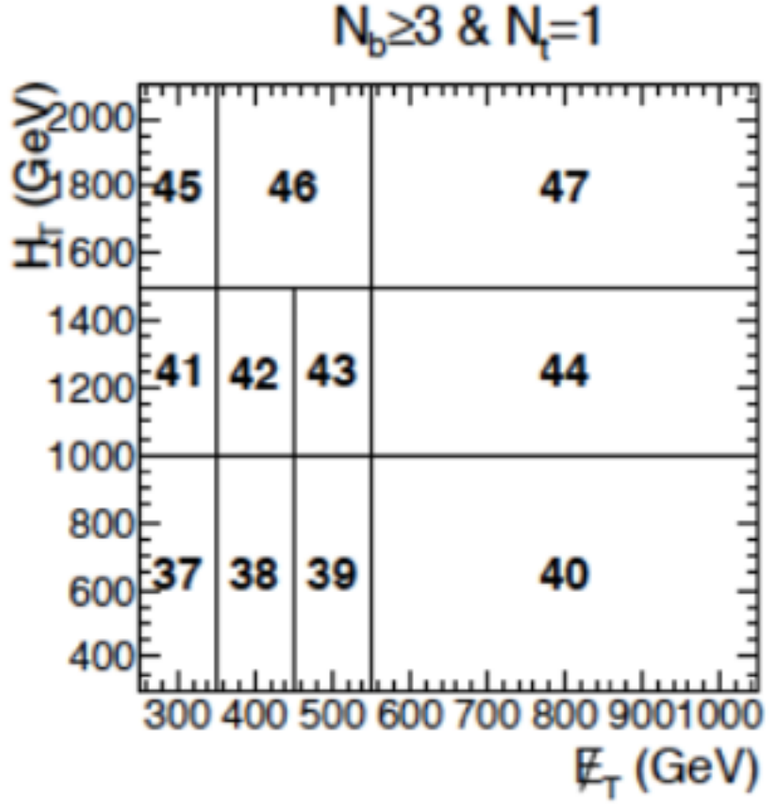
- Variables after baseline selection
- All variables have good discrimination power
- Baseline selection:
  - lep veto, Nj≥4, N_b≥1, N_t ≥1, ΔΦ, p_T^{miss}>250 GeV, H_T>300 GeV, M_{T2}>200 GeV

Search Bins

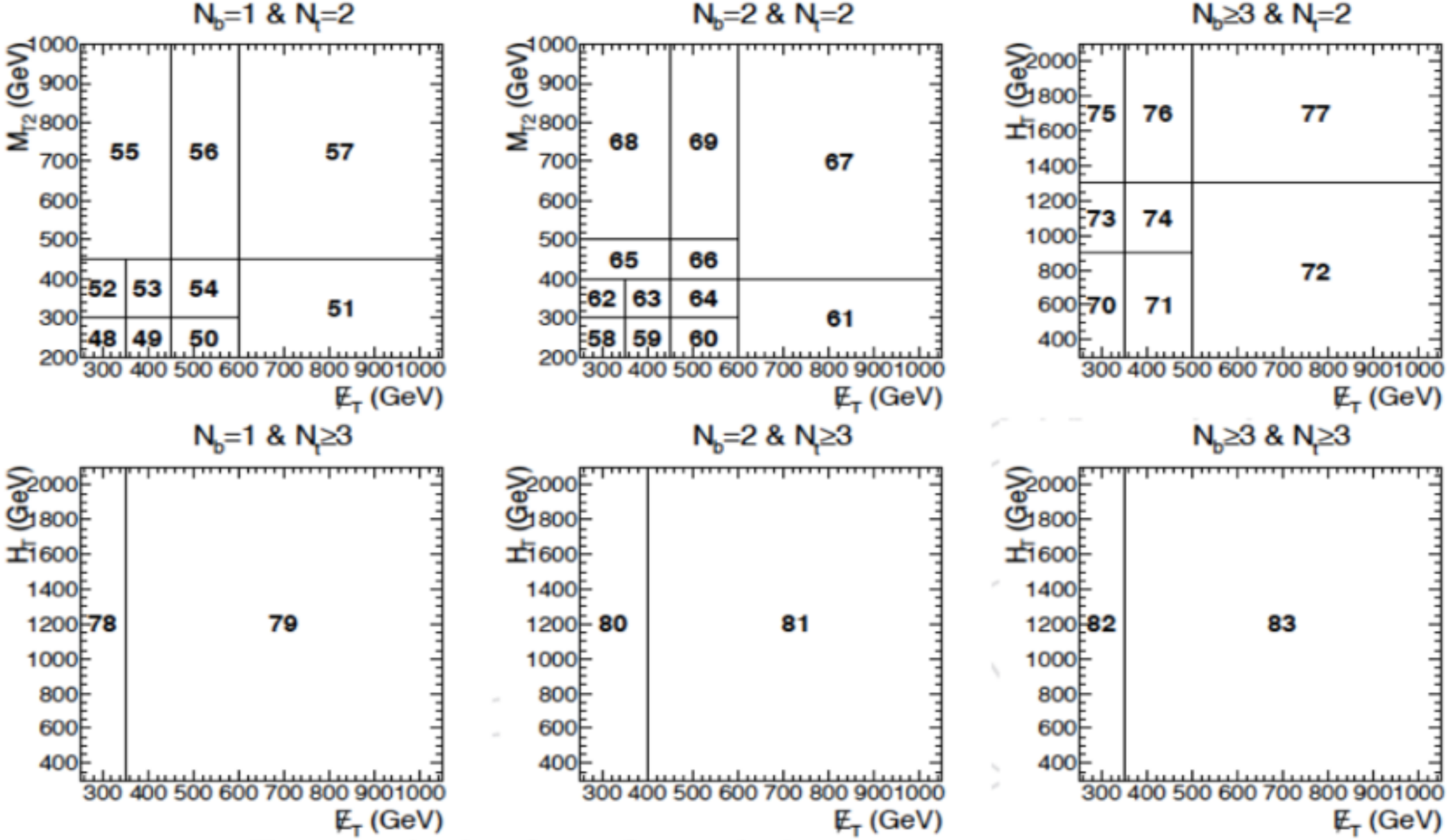
- 84 search bins in N_b, N_t, p_T^{miss}, M_{T2}/H_T
  - N_b and N_t = 1,2 :Targeting T2tt, each block is binned in p_T^{miss}, M_{T2}
  - N_b or $N_t \geq 3$: Targeting T1tttt, each block is binned in p_T^{miss}, H_T only







Search Bins Continued



- Dominant backgrounds are ttbar/W/single top (80%)
- Zinv ~10%, QCD ~9% and ttZ <2%

Lost Lepton and Hadronic Tau TF Method Background

Nt=0, Nb=2

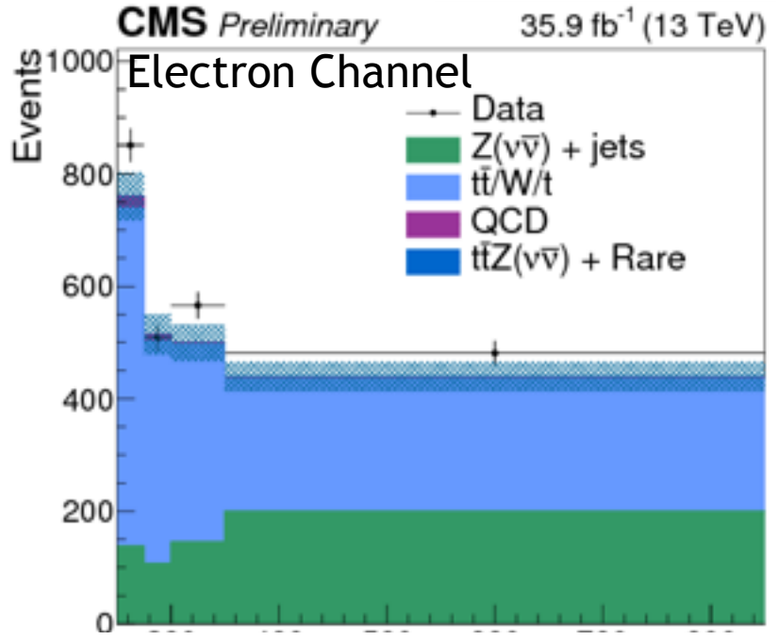
- Measure Translation Factor (TF) TF = N(search region)/N(control region)

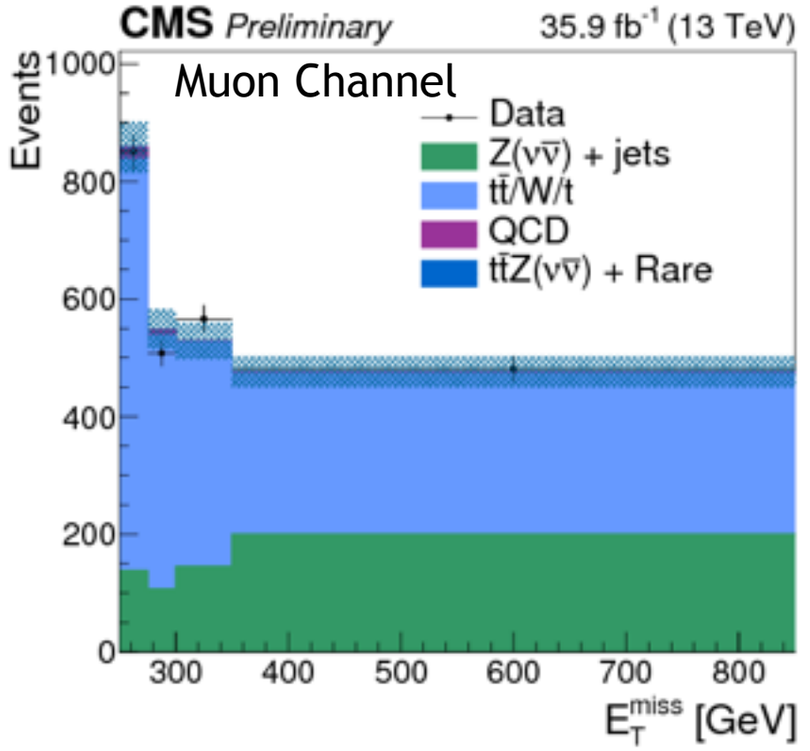
- Calculated for every search bin in TTJets, WJets, Single Top MC

- Control Region(CR): Signle Muon/ Electron with $M_{TW} < 100$ GeV

- TF measured for lost e/mu and tau respectively using MC truth information

- 2 sets of TF one for muon and electron channels

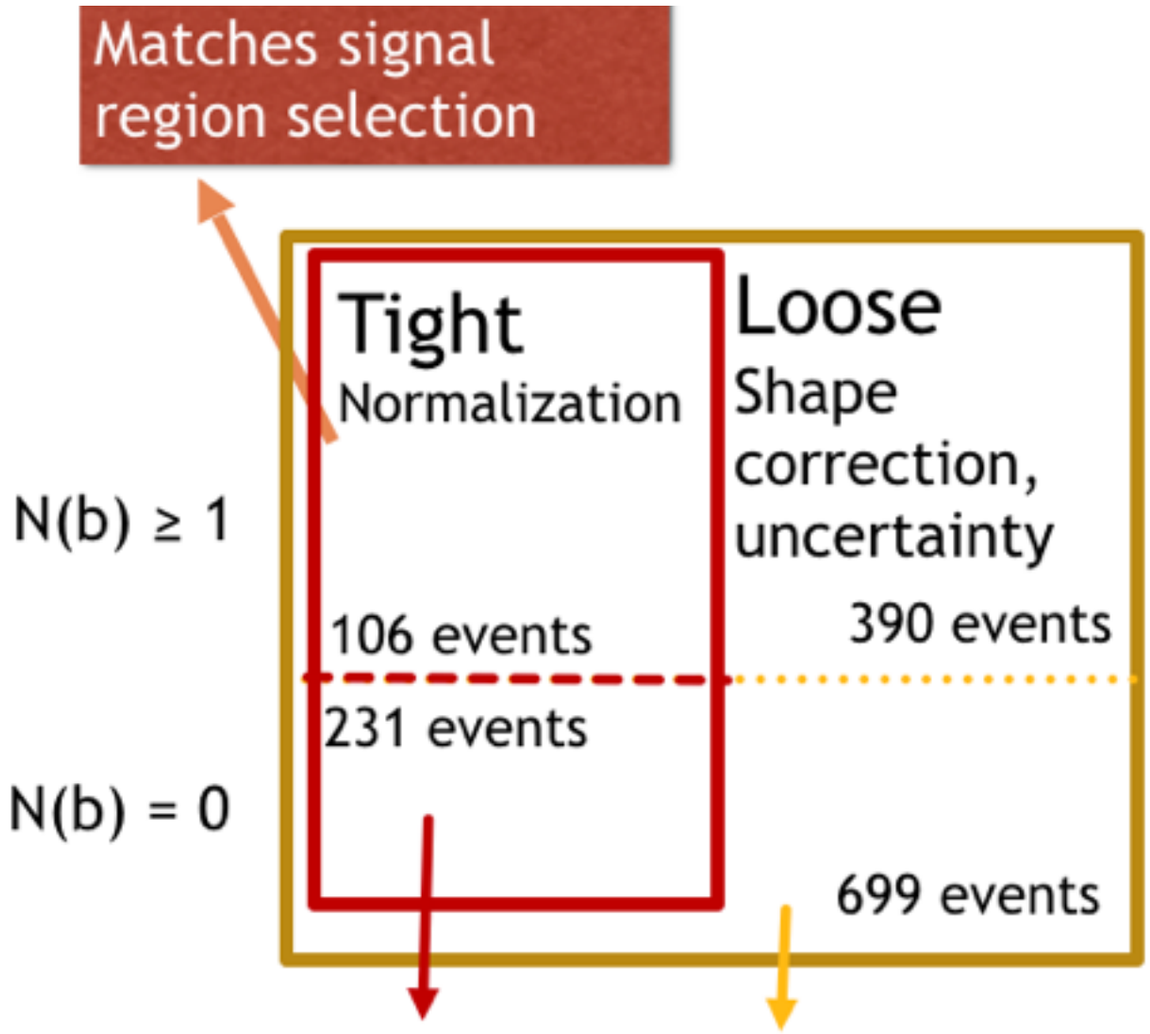




Z Invisible Background

- N-jet Reweighting
  - For ttbar ISR reweighing is applied
  - Derive a scale factor (SDY(Njet)) for Drell Yan going mumu in the loose control region, and apply this to the z->nunu MC in the signal region
  - Corrects for both shape and normalization differences and is done separately for 0Bjet and >= 1bjet regions
- Normalization
  - Calculated in the tight region
  - Done after all the reweighting is applied to the MC

$$R_{norm} = 1.06 \pm 0.08$$



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

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

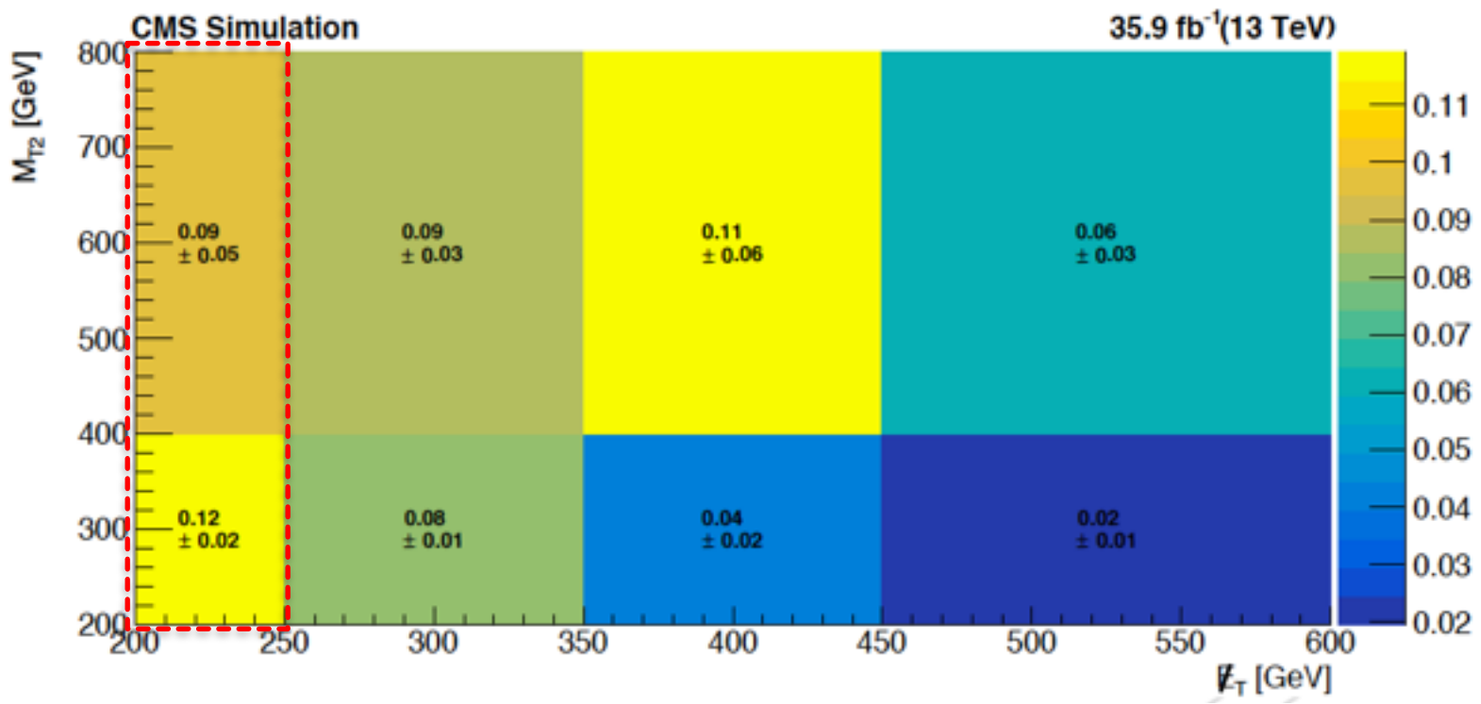
QCD Background

- QCD a T-Factor method is used

$$N_{QCD}^{SR} = N_{QCD}^{\Delta\bar{\phi}} \times T_{QCD}$$

$$T_{QCD} = \frac{N_{QCD}^{\Delta\phi}}{N_{QCD}^{\Delta\bar{\phi}}}$$

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

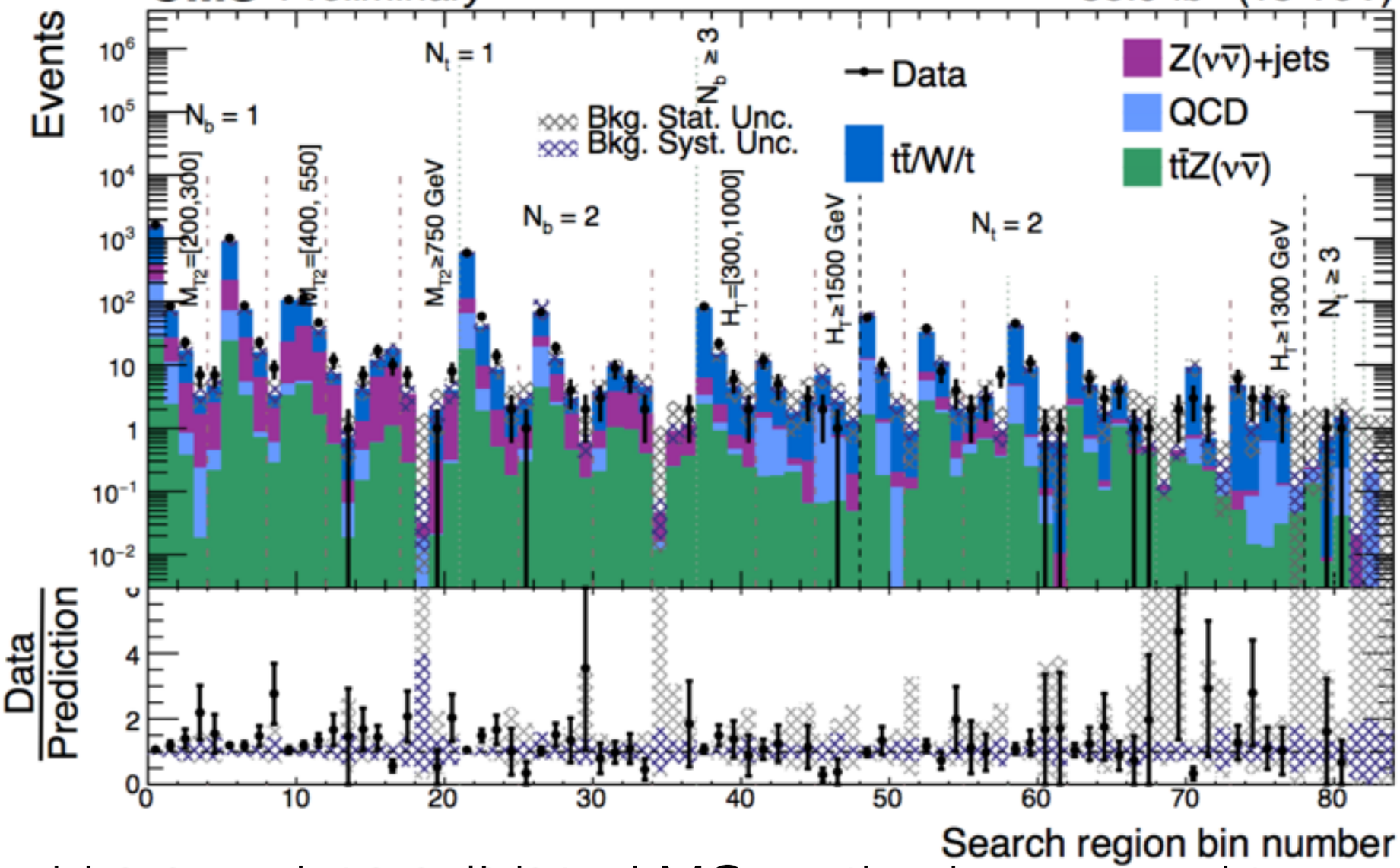


- Use inverted Δϕ control sample to predict the number of QCD events, using a translation factor (T-factor)

$$N_{QCD}^{\Delta\bar{\phi}} = N_{Data}^{\Delta\bar{\phi}} - N_{LL}^{\Delta\bar{\phi}} - N_{\tau_h}^{\Delta\bar{\phi}} - N_{Z\to\nu\nu}^{\Delta\bar{\phi}}$$

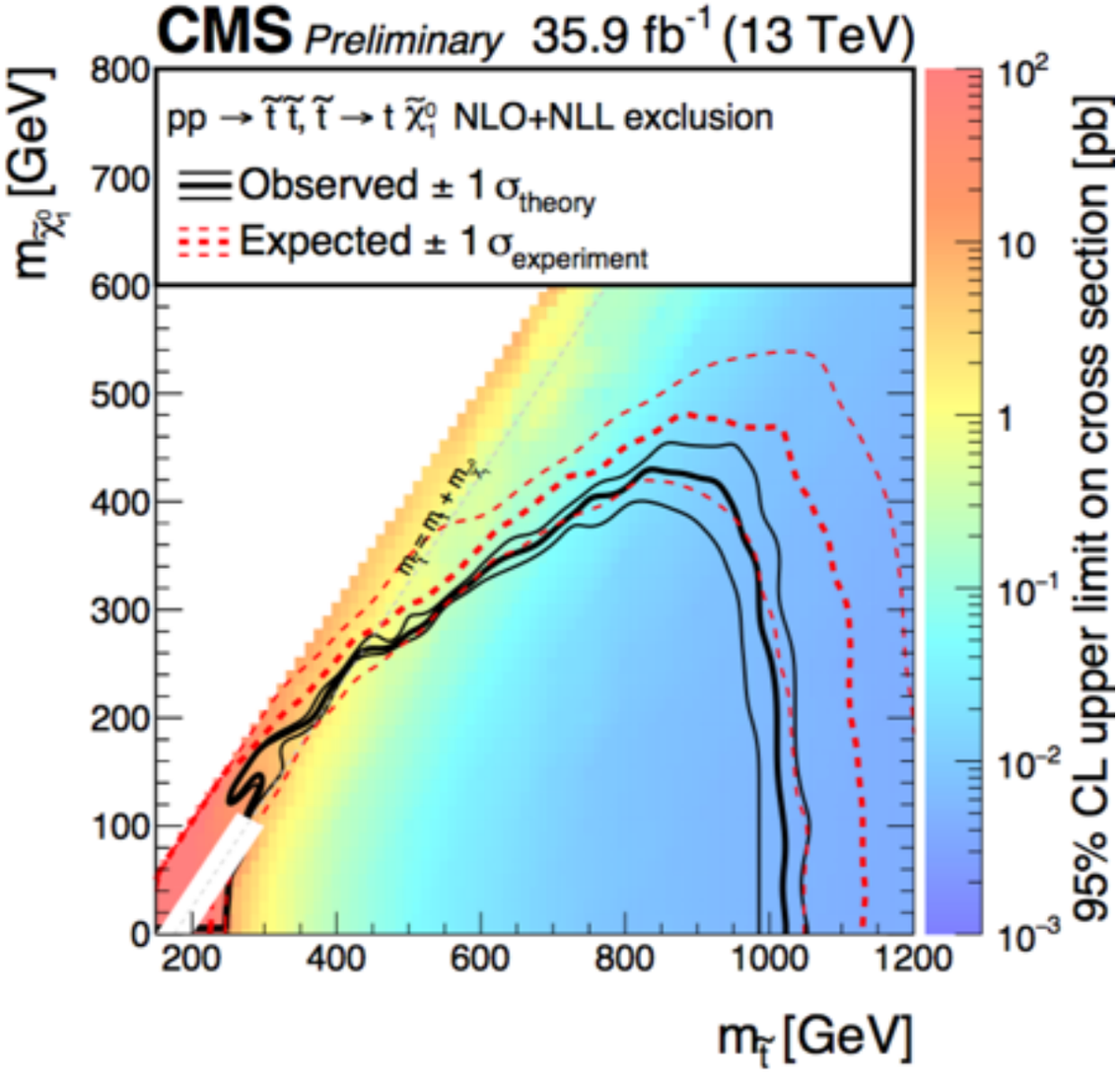
- The T-factor is measured in data in a sideband region: ($200 < P_T^{miss} < 250$ GeV), and extrapolate in higher P_T^{miss} region using MC results

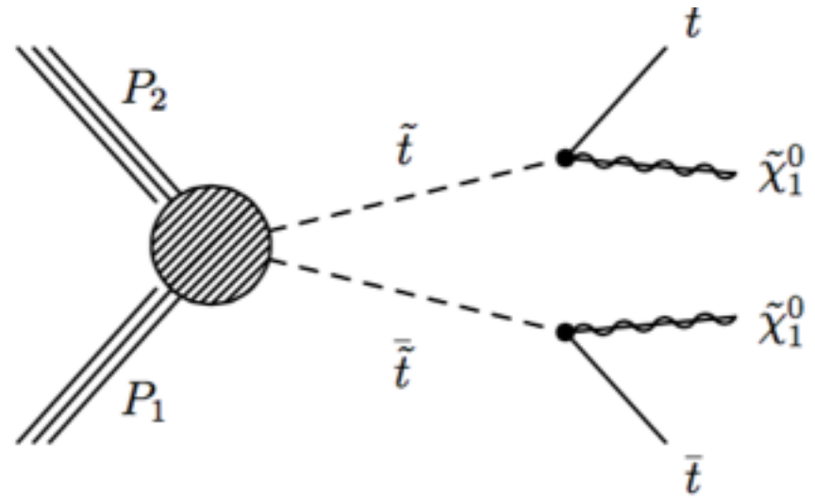
Results



- Data driven or data validated MC methods are used to predict all backgrounds
- No significant excess seen

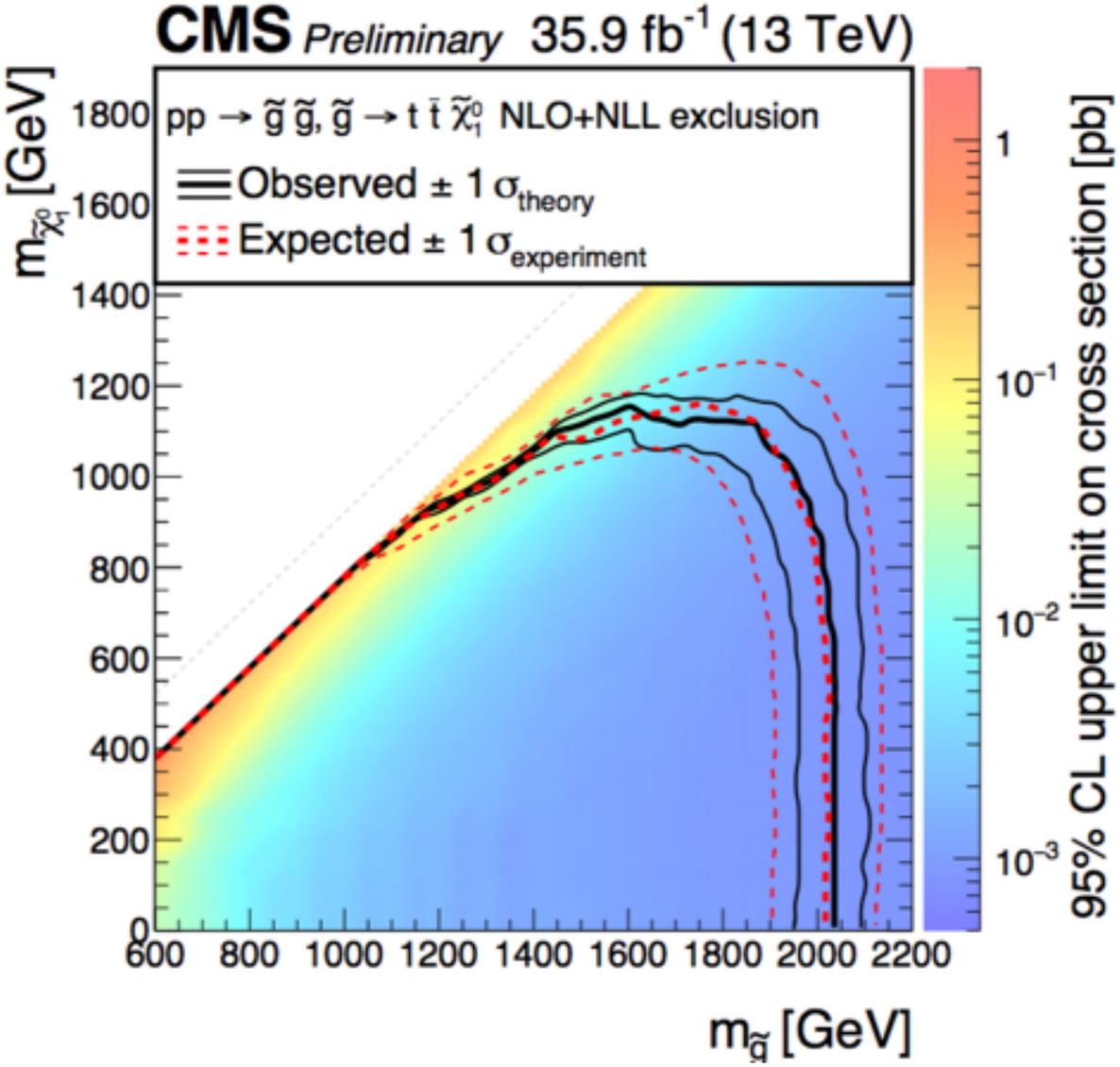
Interpretation T2tt

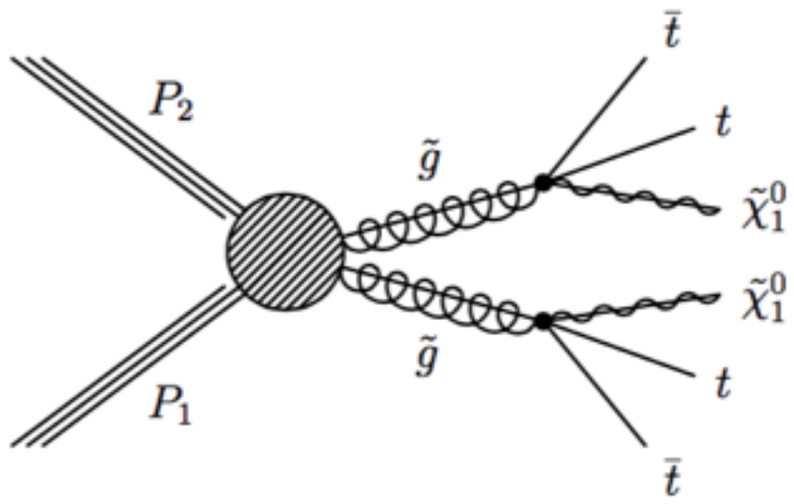




- ~100 GeV in stop mass better than the previous results
- Exclude stop mass (obs) up to 1020 GeV and LSP mass up to 430 GeV

Interpretation T1tttt





- ~200 GeV for gluino mass better than was seen previously
- Exclude gluino mass (obs) up to 2040 GeV and LSP mass up to 1150 GeV for T1tttt

Conclusion

- Direct Stop: Exclude stop mass (obs) up to 1020 GeV and LSP mass up to 430 GeV
- Mediated stop: Exclude gluino mass (obs) up to 2040 GeV and LSP mass up to 1150 GeV for T1tttt
- Simplified Top tagger will be available for theorist soon.
- Future plans is to have more coverage of SUSY hadronic models target top/W SUSY signatures

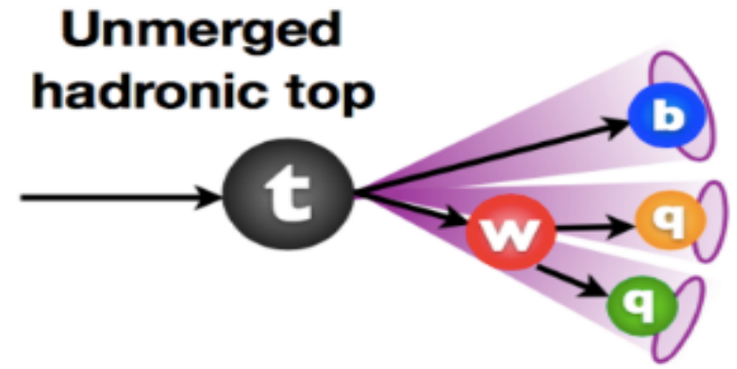
Backup Slides

Top Tagger: Overview

- Hadronically decayed tops have a wide range of p_T
- Hadronic top tagger needs to be
  - High efficiencies in both low and high P_T
  - Low mistag rate

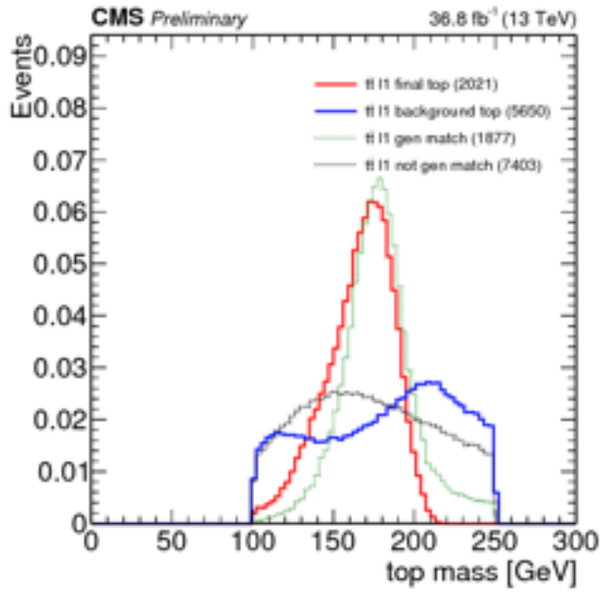
Top Tagging Algorithm

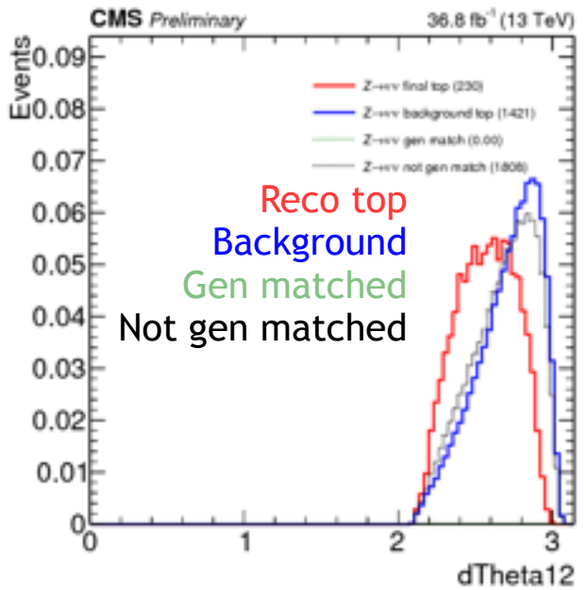
- First tag fully merged tops
  - Using standard PUPPI AK8 boosted top (loose working point)
- Tag medium p_T tops with W+jet category
  - PUPPI AK8 as a boosted W (loose working point)
  - Combine with a nearby AK4 jet
  - Require combined mass consistent with top and AK8 mass to combined mass ratio to be consistent with M_W/M_t
- Tag low p_T tops with combinations of 3 resolved AK4 jets with MVA
- The three categories are combined
  - Fully merged are tagged first, then W+jet, then resolved, removing jets from consideration form subsequent steps

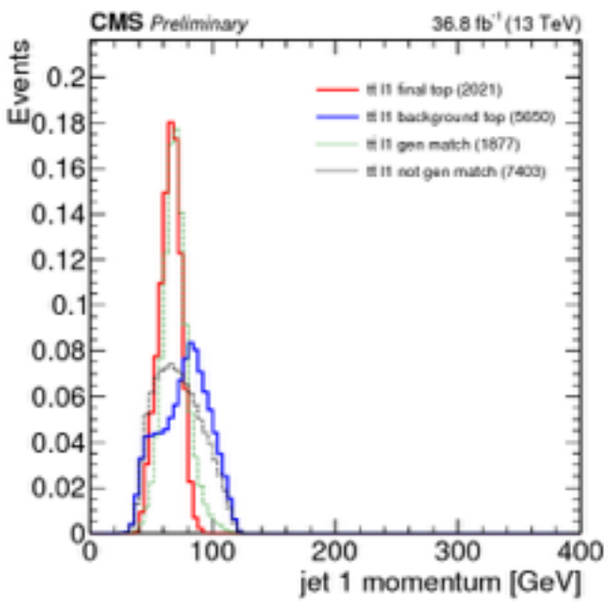


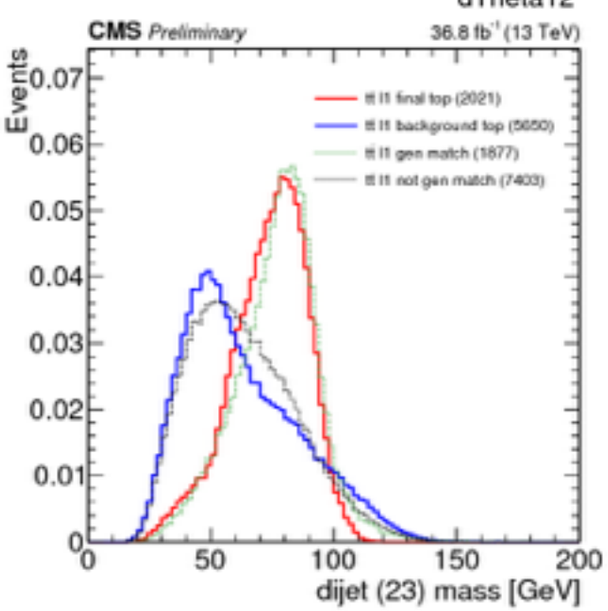
Top Tagger: Resolved Case

- 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 $|\eta| < 5$), within $\Delta R < 1.5$ and combined mass between 110 and 250 GeV
- MVA training on jet properties (Input variables : Jet p, Jet CSV, Jet QGL, etc)
- Cut on MVA discriminator to select Top candidate
- Additional requirement : No more than 1 b-jet in a top candidate
  - Reduces the mistag rate in high b-jet multiplicity
  - small penalty of efficiency



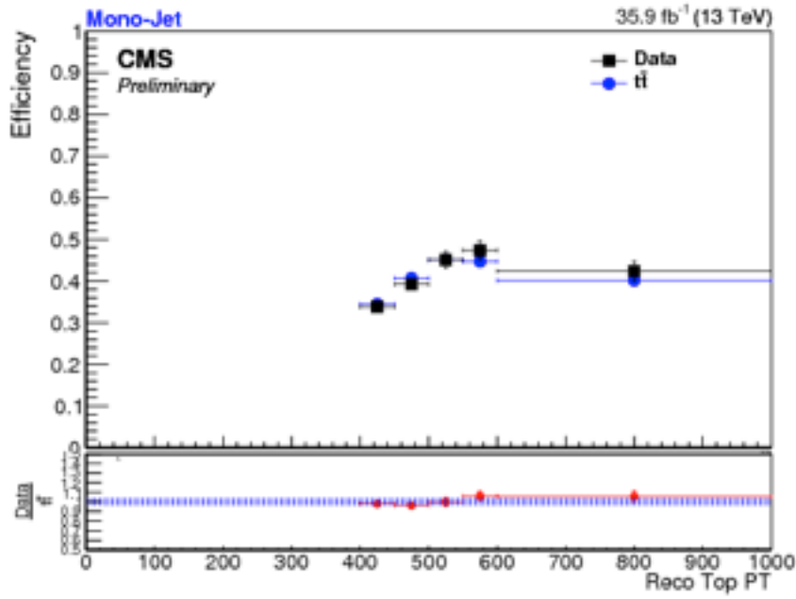


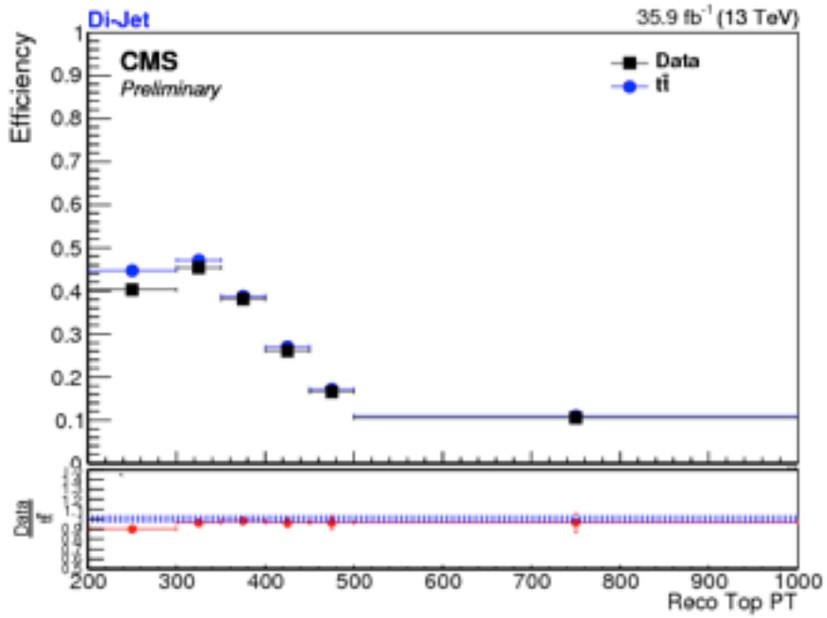


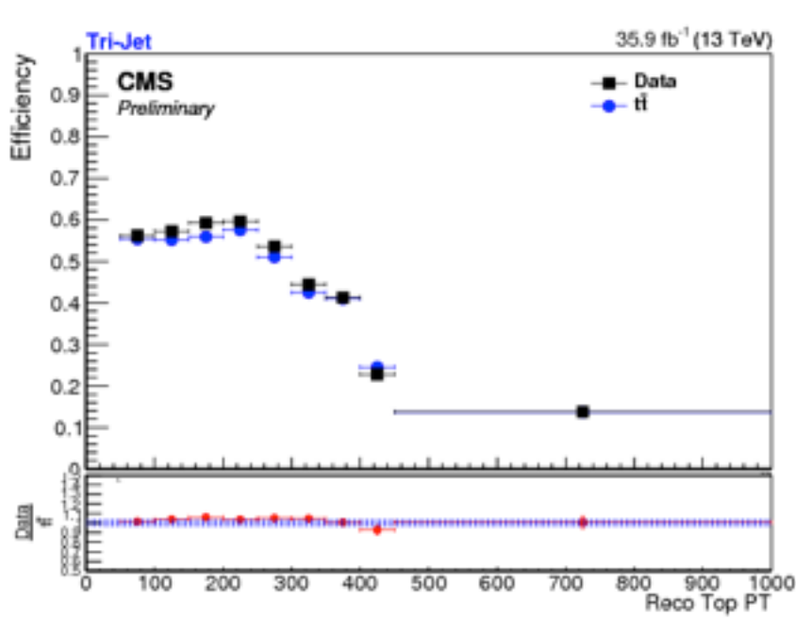


Top Tagger: Data/MC

- Top tagger efficiencies scale factor in data:
- Select tt-enriched region: HLT_Mu24 || HLT_Mu50, Pass Filters, At least 3 jets with p_T > 30 GeV, At least one muon with p_T > 45 GeV, |η|<2.1
- Scale factor in Reco-Top P_T
- Applied in analysis as a scale factor, mono-Jet, Di-Jet and Tri-Jet respectively





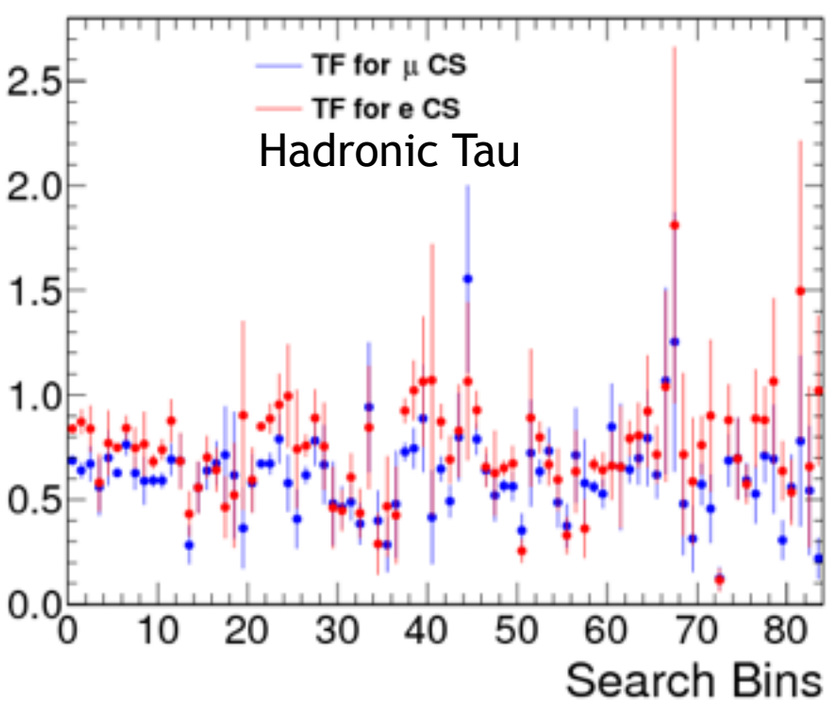


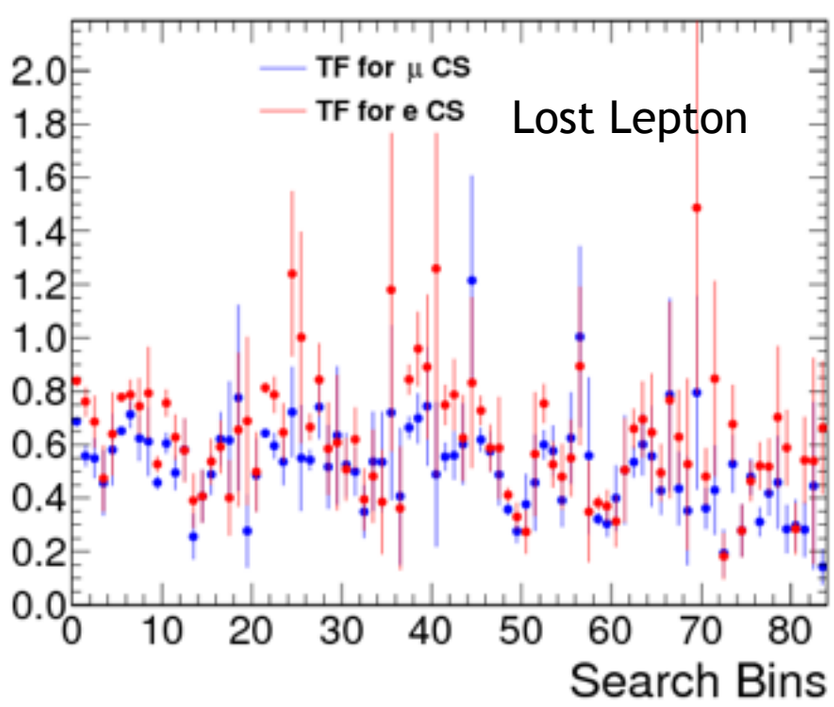
Baseline

- Noise cleaning: HBHENoiseFilter, HBHENoiseIsoFilter, EcalDeadCellTriggerPrimitiveFilter, GoodVertices, eeBadScFilter, CSCTightHalo2016Filter, badPFMuonFilter, badChargedHadronFilter, LooseJetID + PFMET/CaloMET<5
- Jets and p_T^{miss}:
  - AK4PF jets with CHS: $N_{jets}(p_T > 50) \geq 2$ and $N_{jets}(p_T > 30) \geq 2$
  - p_T^{miss}> 250 GeV
  - H_T> 300 GeV
  - $\Delta\phi\left(j_{1,2,3}, p_T^{miss}\right) > 0.5, 0.3, 0.3$
  - $N_b \geq 1$ (CSVM)
- Lepton/track veto:
  - μ veto: $p_T > 10$ GeV, medium ID && miniISO
  - e veto: $p_T > 10$ GeV, veto ID && miniISO
  - IsoTrack: e/μ track (relIso < 0.2 && $p_T > 5$ GeV) or π track (relIso < 0.1 && $p_T > 10$ GeV), $M_{tW} < 100$ GeV
- Top reconstruction:
  - $N_t \geq 1$ with top tagger
  - $M_{T2} > 200$ 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 $\Delta\phi\left(\vec{p}_T^{\,miss}, j_{1,2,3,4}\right) > 0.5$
- Yields a W enriched, signal-depleted region
- Cross-check shows good agreement in both electron and muon channel

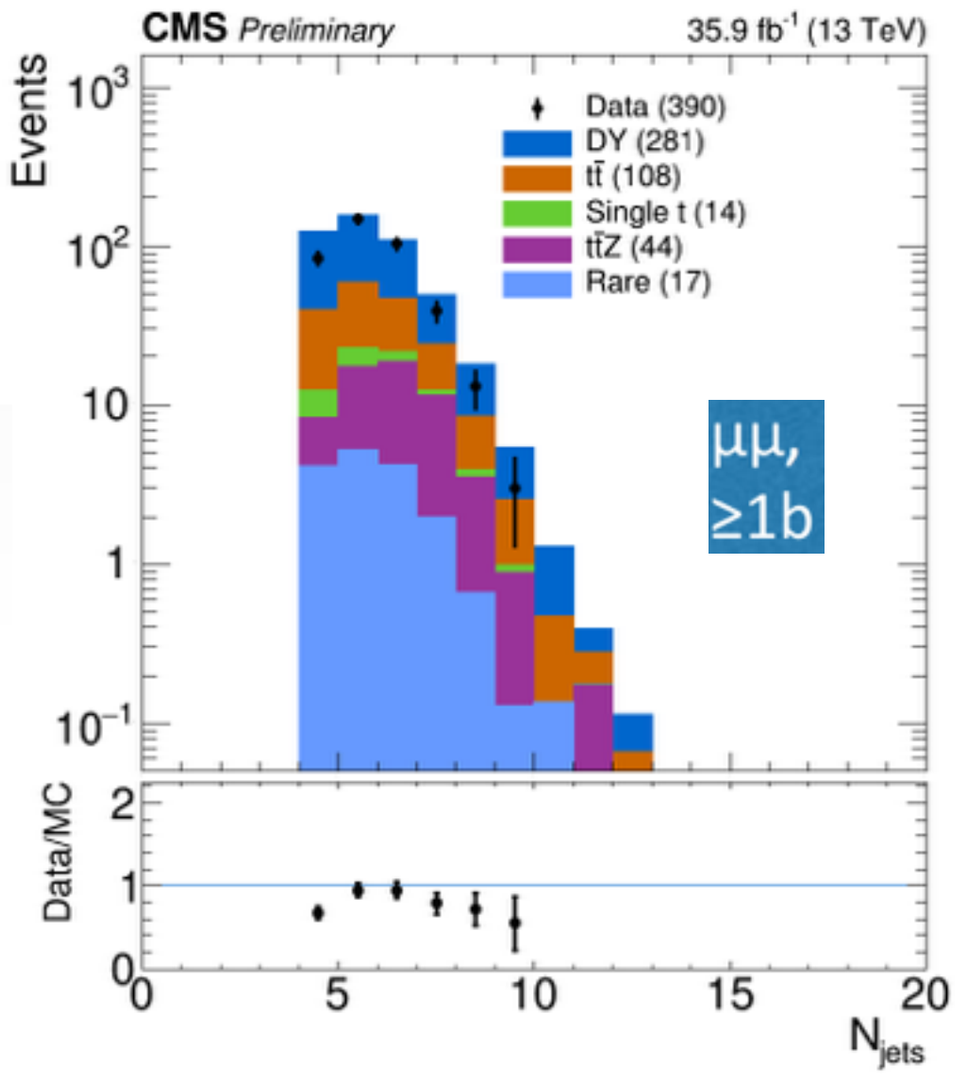




Lose and Tight Z Invisible

- MC base estimate, validated in $Z \to \mu\mu$ control sample
  - Trigger: HLT_Mu50_v*
  - Require 2 medium ID opposite sigh muons
  - $81 < M_{\mu\mu} < 101$ GeV

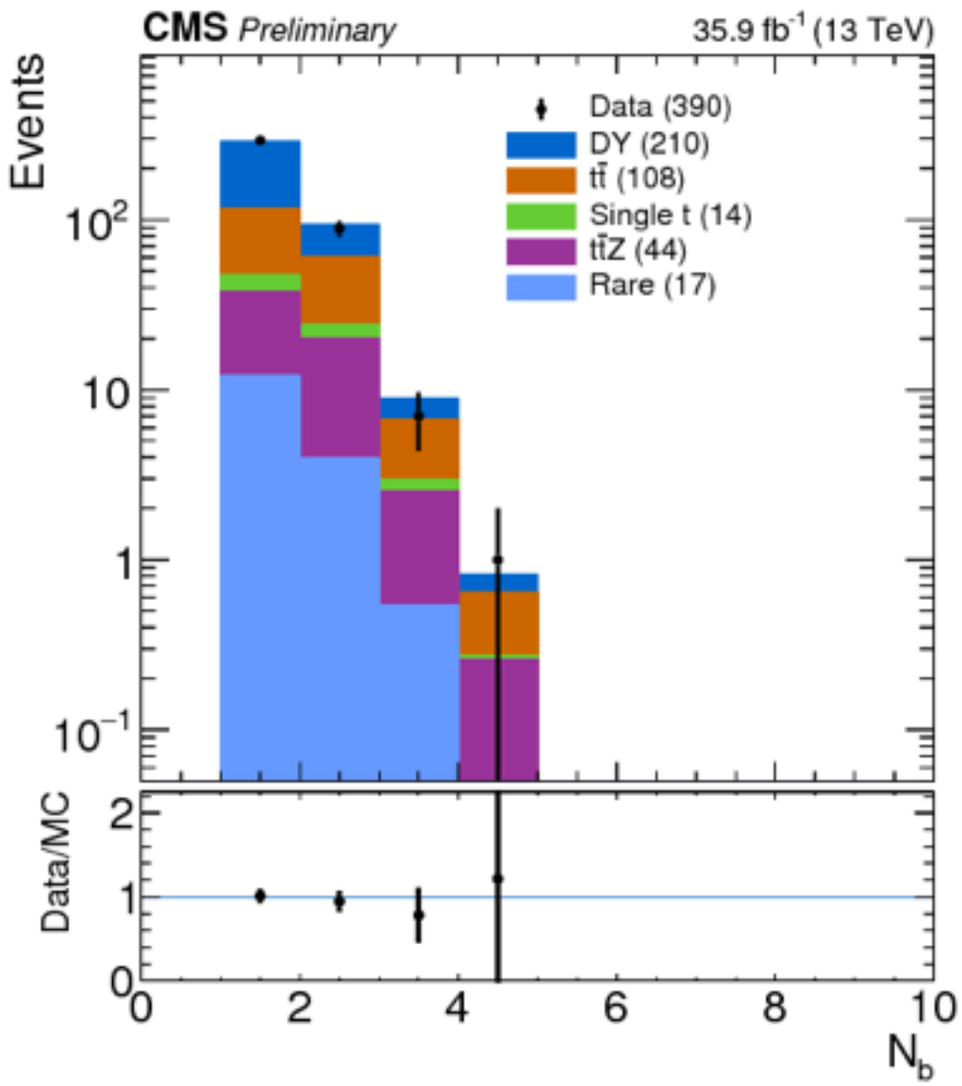
- DY loose region:
  - $N_j \geq 4$, elec veto, $\Delta\phi$,
  - $p_T^{miss} > 100$ GeV
  - $H_T > 300$ GeV,
  - $N_T \geq 1$
- Tight region adds:
  - $p_T^{miss} > 250$ GeV
  - $M_{T2} > 200$ GeV

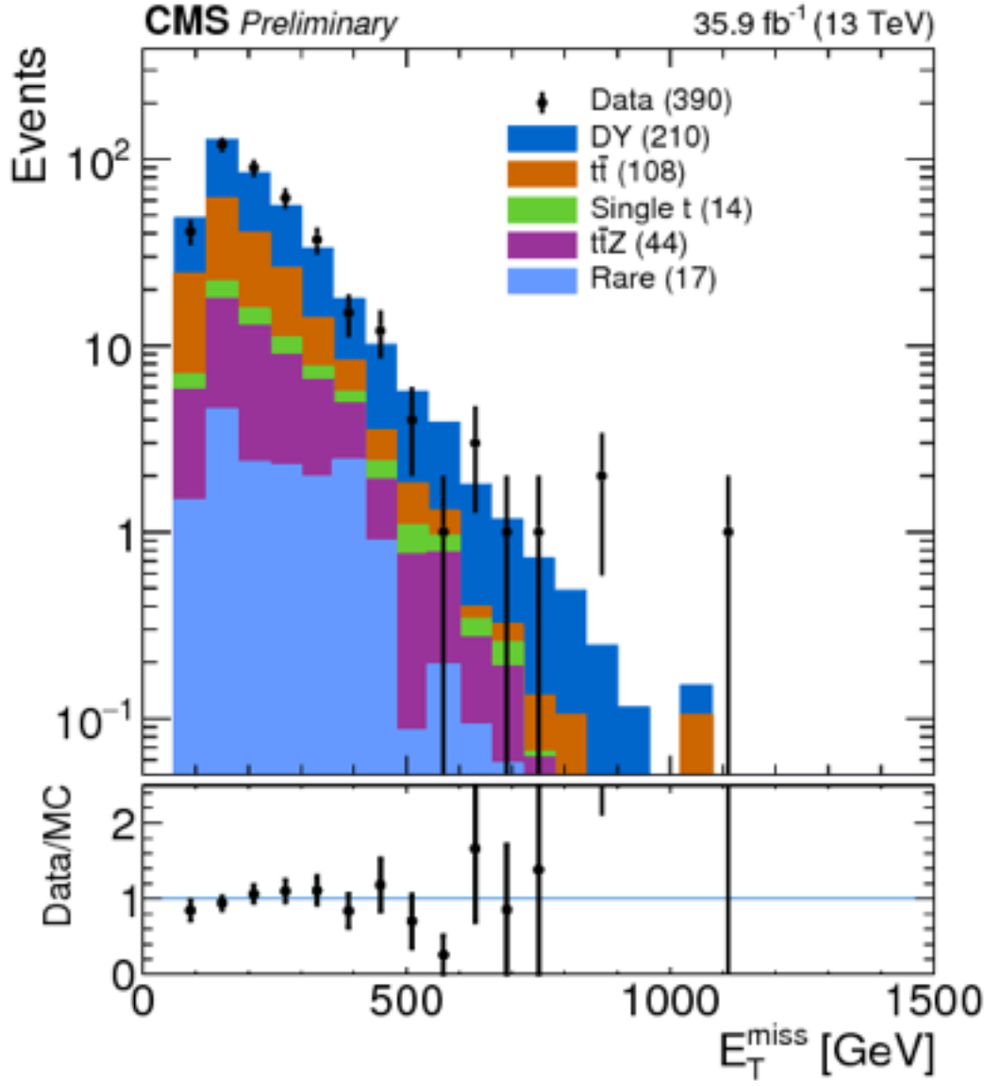


- Shape weights applied

Z Control Region after Reweighing

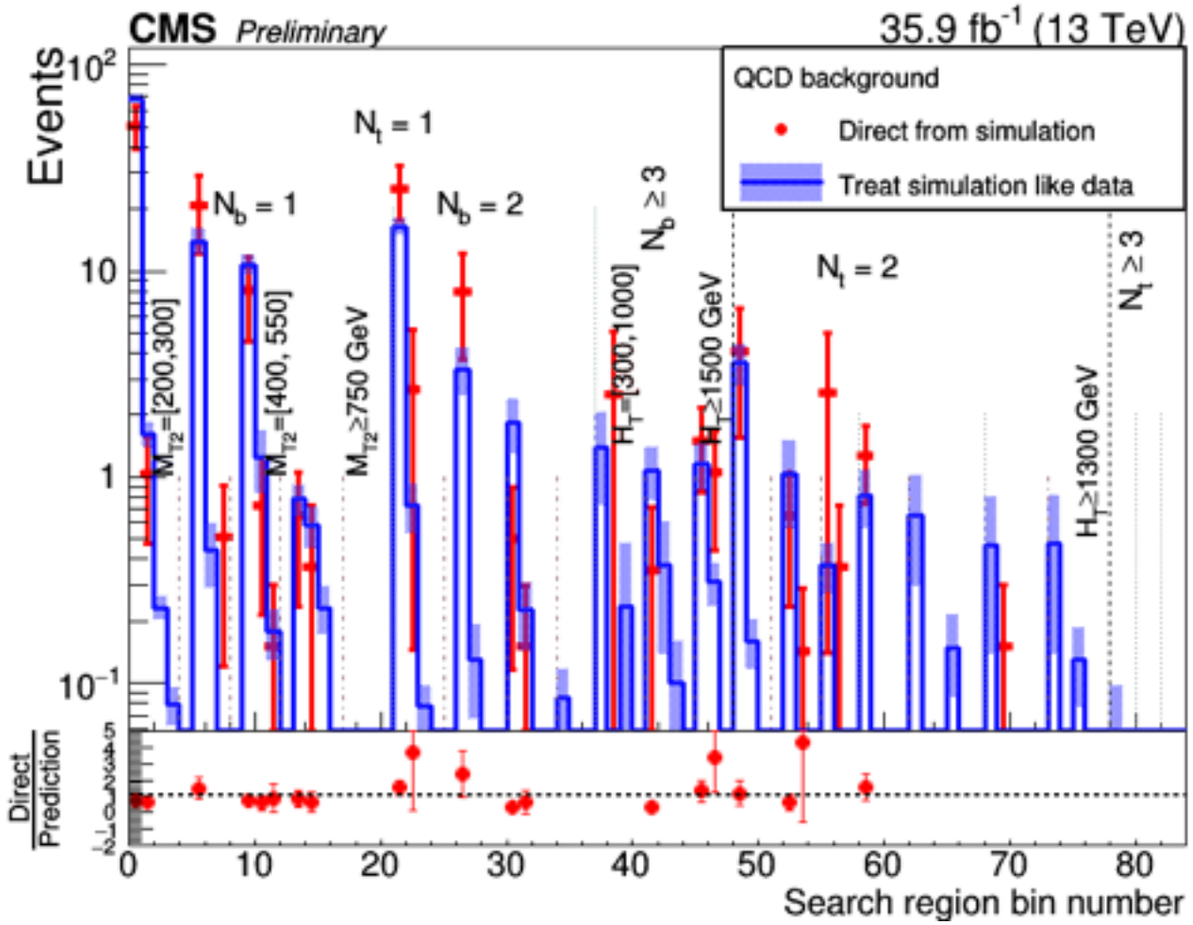
The loose DY control region with $N_b \geq 1$ after all weights are applied





QCD Test and Prediction

- 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}



- Test the T-factor paramterization method works for the QCD predictions in our search bins
- (As shown) direct closure test for search bins if having enough statistics
- When statistics is low or no MC yields, 2D closure tests of (E_T^{miss}, M_{T2})/(E_T^{miss}, H_T) or quadratic combination of four 1D closure uncertainties are considered

QCD systematics

- Closure is the major systematic uncertainty

Table 27: Contributions from different sources of systematic uncertainty to the QCD background prediction.

Process	Source	Effect on QCD Prediction in %
T_{QCD}^{Scale} factors	Statistical uncertainty on T_{QCD}^{MC} and T_{QCD}^{Data}	30 to 330
Closure	Non-closure and statistical precision of the closure	30 to 500
Contamination from other backgrounds	Uncertainties from other background input	2 to 50