CMS Searches With Boosted Top Quarks

Reyer Band
University of California, Davis
On Behalf of the CMS Collaboration

***11th* BOOST 2019**

25 July 2019

CMS Searches With Boosted Top Quarks…
Using Deep Neural Networks?

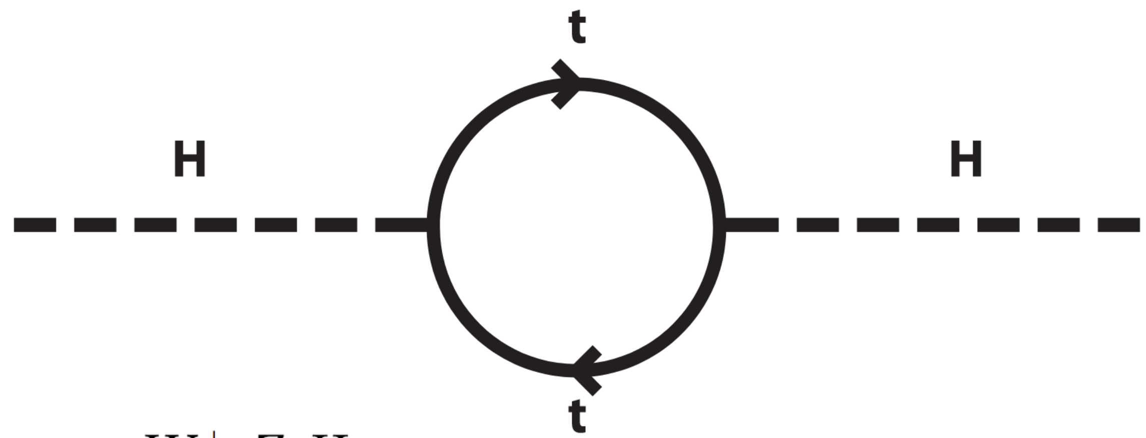
Reyer Band
University of California, Davis
On Behalf of the CMS Collaboration

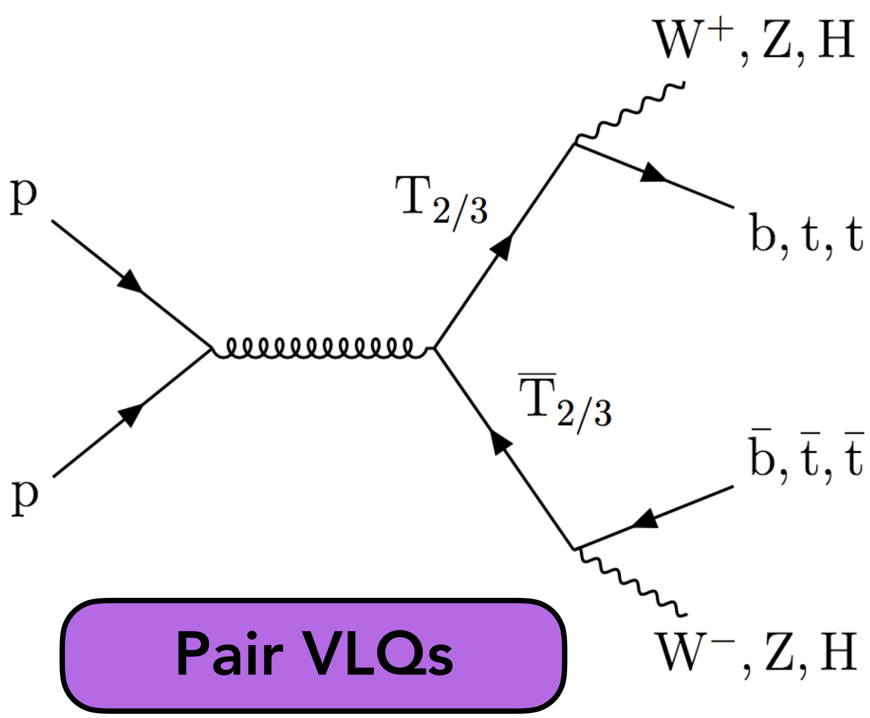
11th BOOST 2019

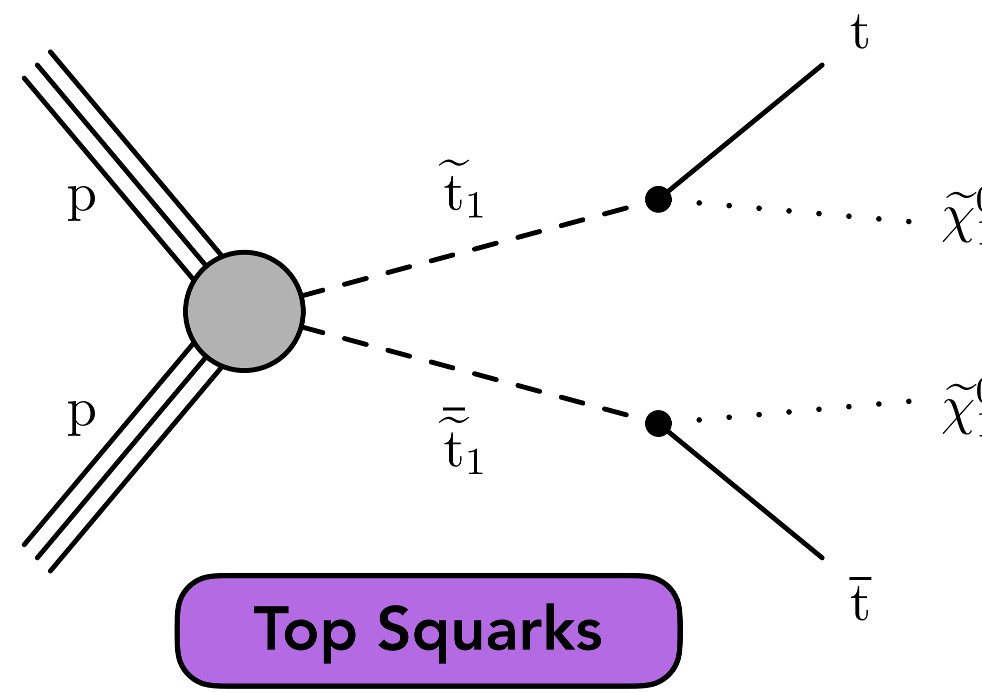
25 July 2019

Boosted Tops At CMS

Many theories of new physics contain heavy particles which couple to top quarks







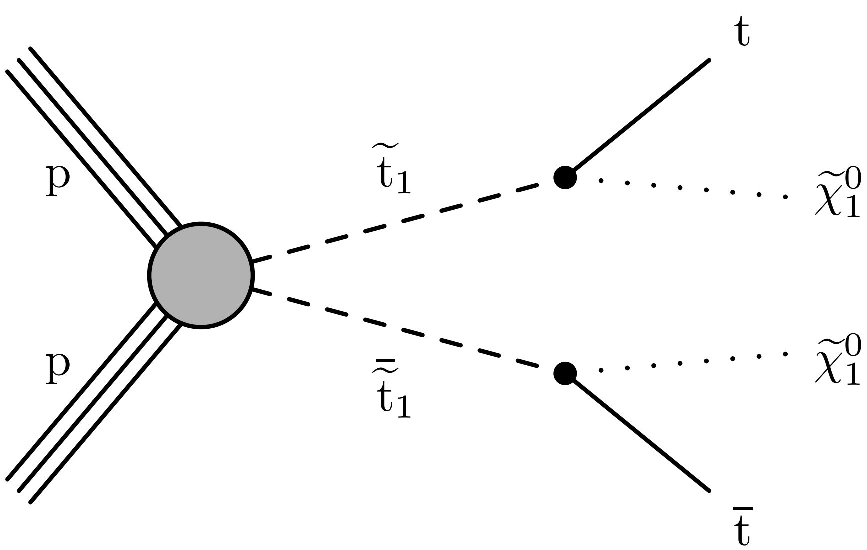
- A few recent results from CMS:
  - **Vector-Like Quark Pair Production**
  - **A Search For Top Superpartners**

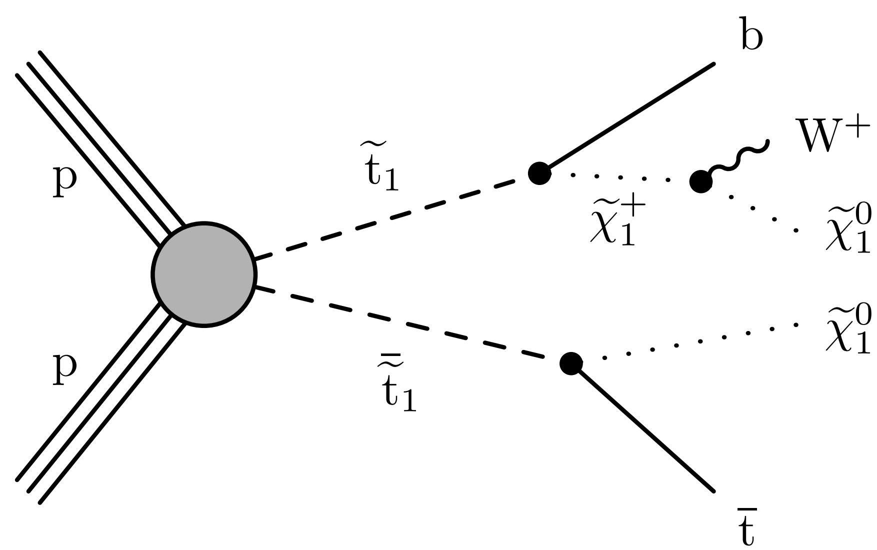
For other CMS results, see Christine's talk from BOOST18

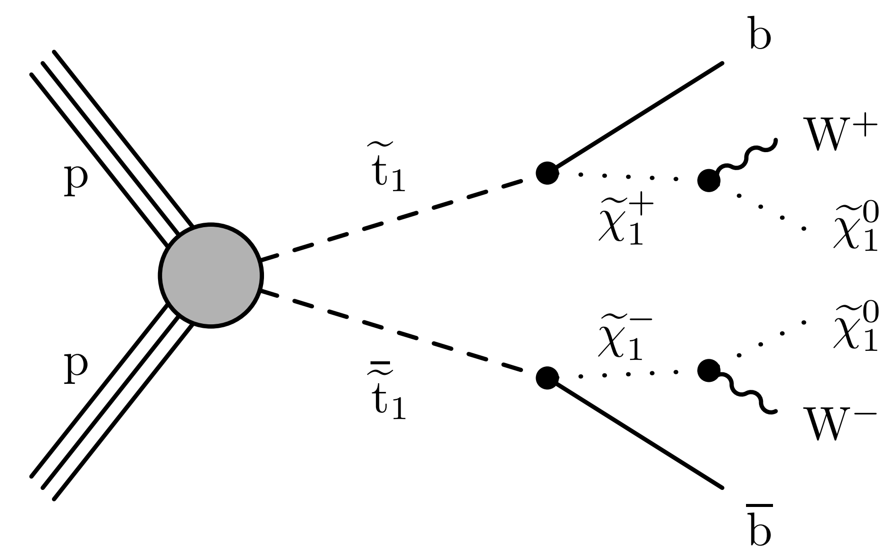
$\tilde{t}\bar{\tilde{t}}$ - Signal

SUS-19-009

- SUSY top super partner $\tilde{t}$ may decay to a top quark
- Look for events with a single lepton from a W, itself either from a top or chargino
- MET expected from neutralino and leptonic W



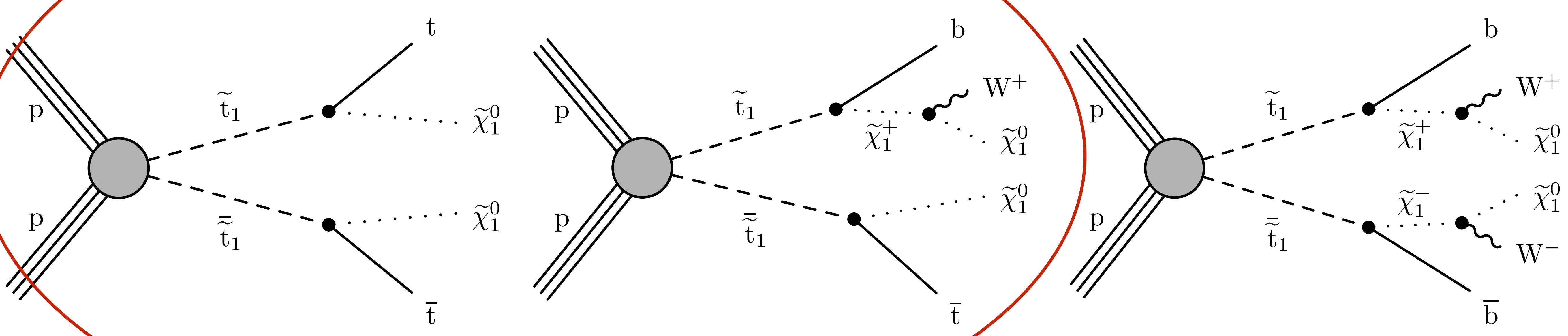




$\tilde{t}\bar{\tilde{t}}$ - Signal

SUS-19-009

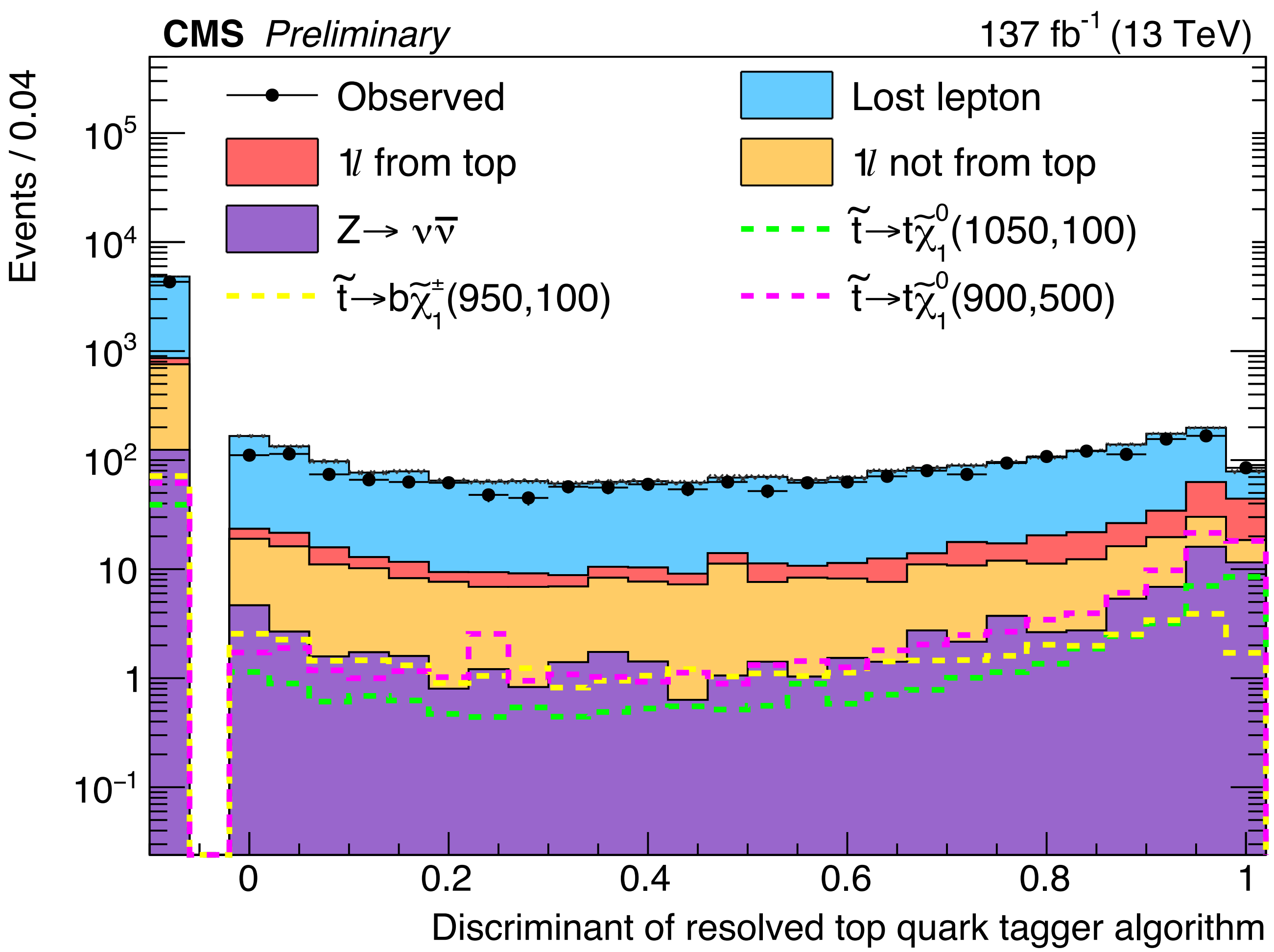
- SUSY top super partner $\tilde{t}$ may decay to a top quark
- Look for events with a single lepton from a W, itself either from a top or chargino
- MET expected from neutralino and leptonic W



- Hadronic decay of the top a key signature in two decay modes
  - If $m_{\tilde{t}} >> m_{\tilde{\chi}} + m_t$ ("uncompressed"), top quarks will be boosted
  - If not boosted, identify the resolved jets.
- Search performed in the **full Run 2** dataset (137 fb^{-1})

$\tilde{t}\bar{\tilde{t}}$ - Top Tagging

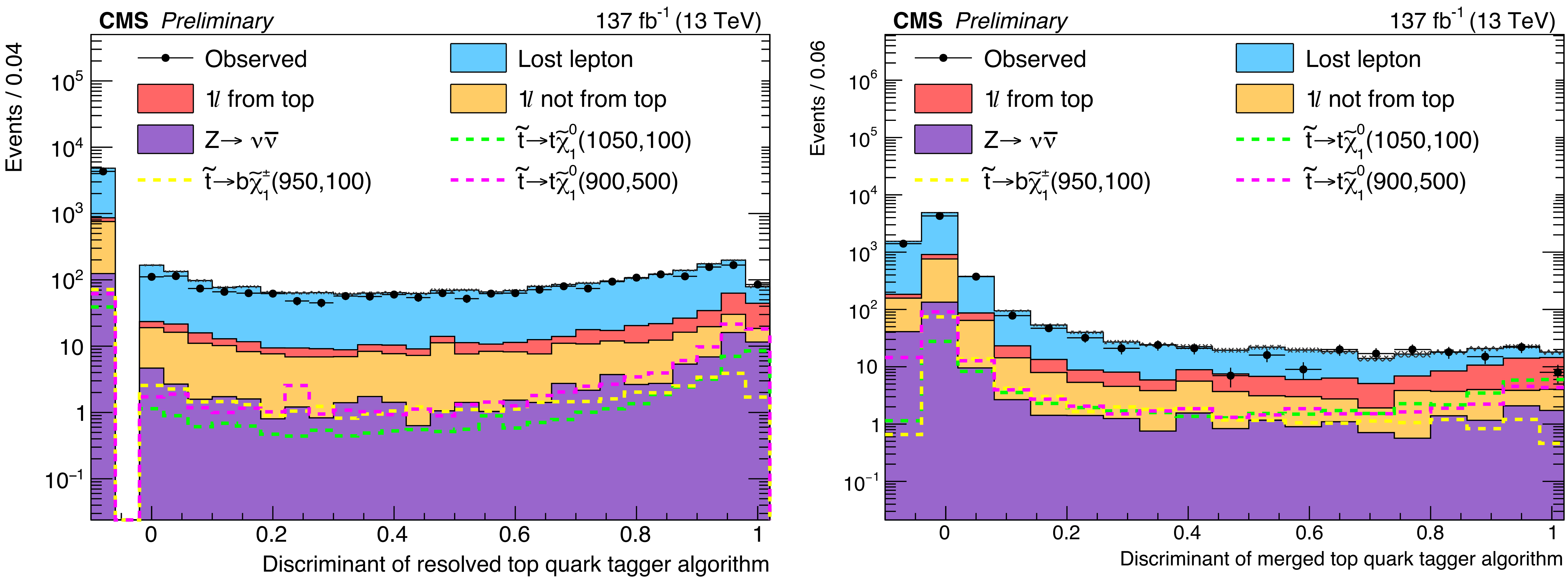
SUS-19-009



- Tag tops as either **merged** or **resolved**, boost depends on mass splitting
- **Resolved**:
  - Must be at least 3 (AK4) jets
  - Tag with **DeepResolved** network - trained on jet quantities and trijet variables

$\tilde{t}\bar{\tilde{t}}$ - Top Tagging

SUS-19-009



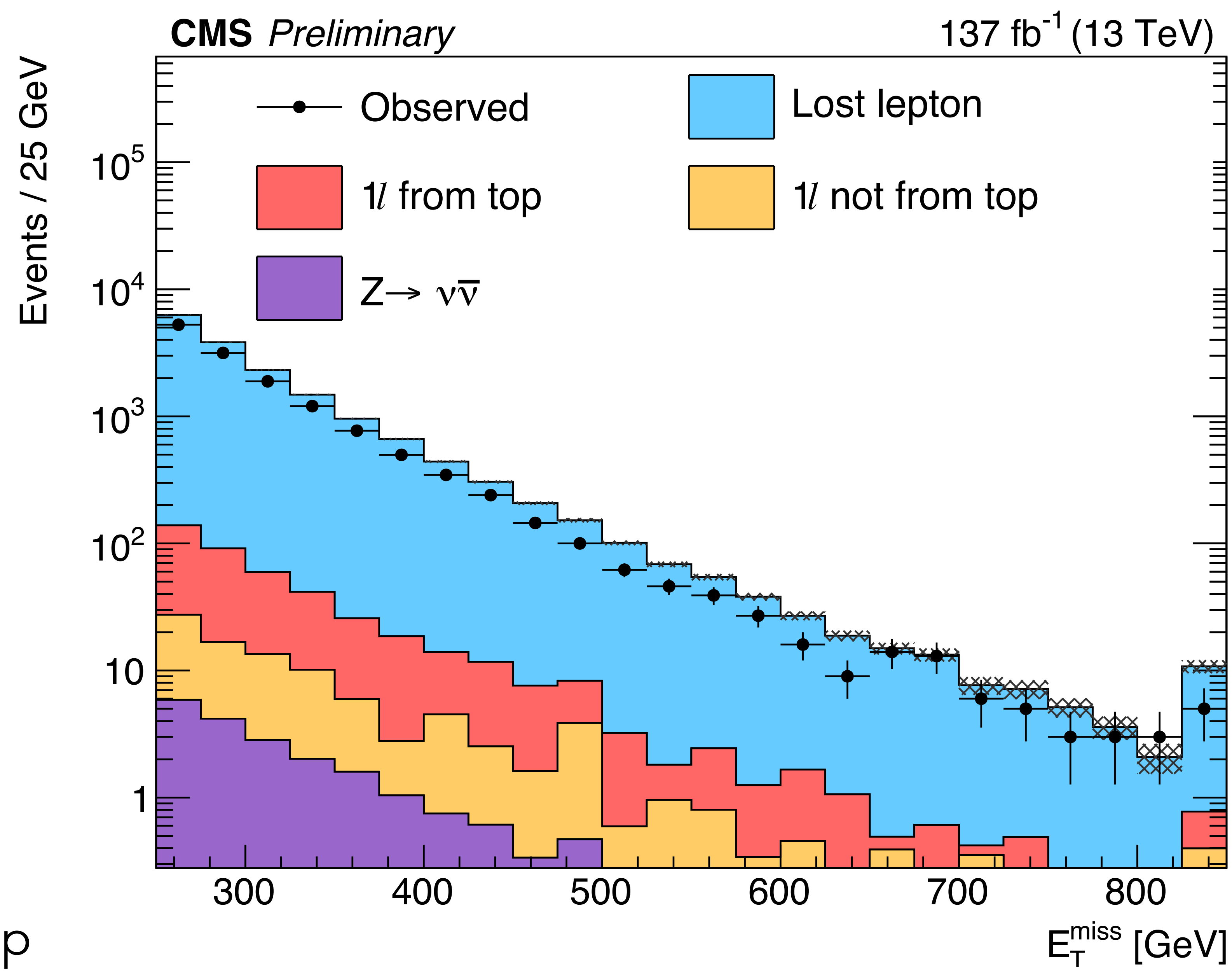
- Tag tops as either **merged** or **resolved**, boost depends on mass splitting
- **Resolved**:
  - Must be at least 3 (AK4) jets
  - Tag with **DeepResolved** network - trained on jet quantities and trijet variables
- **Merged**:
  - Heavily boosted top quark identified as one AK8 jet
  - Tag with **DeepAK8** network. See Meenakshi Narain's talk for all the details
- First search to make use of these neural networks

$\tilde{t}\bar{\tilde{t}}$ - Background

SUS-19-009

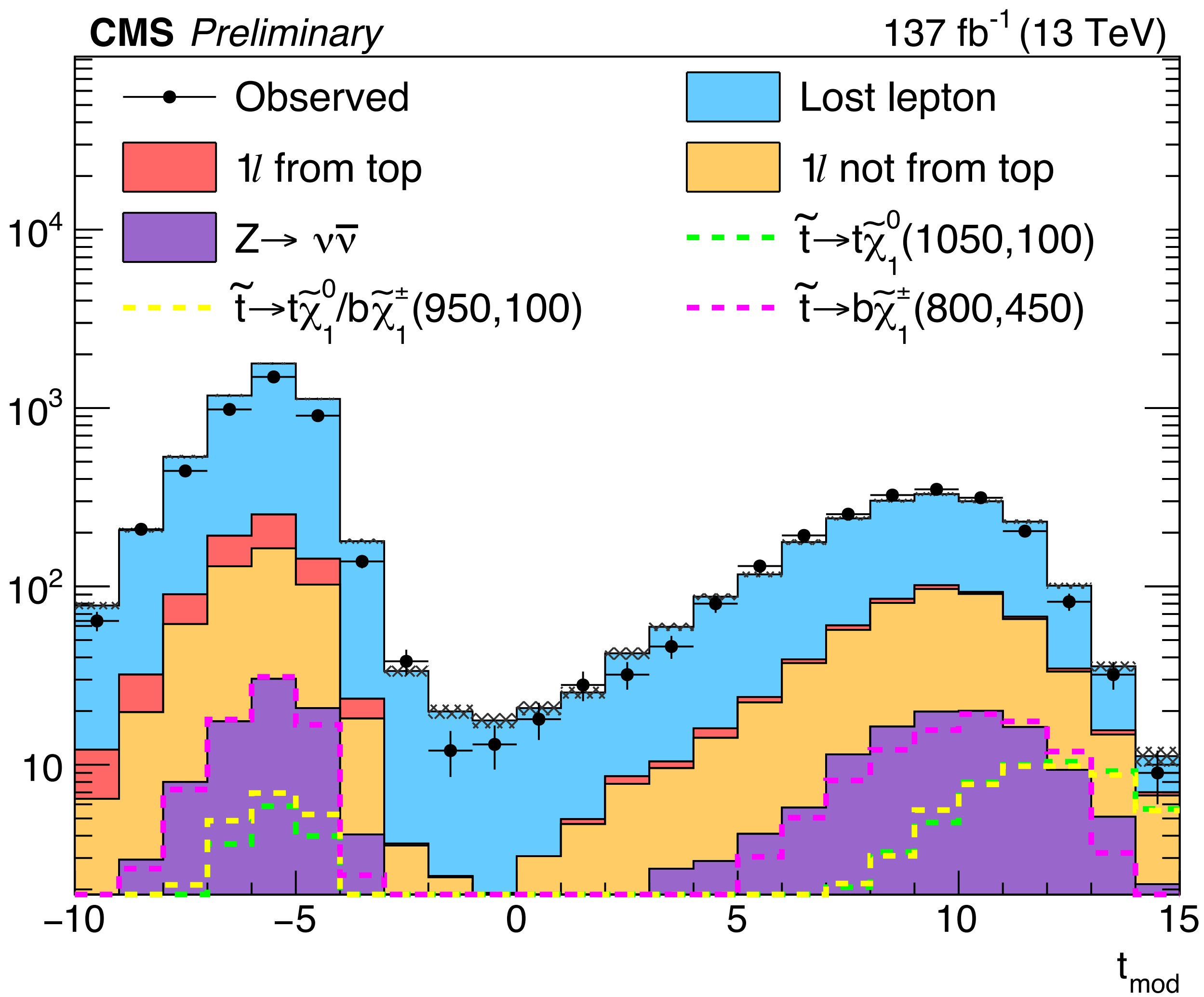
- **Lost lepton** from dilepton event
  - Mostly from $t\bar{t}$->2l
  - Measure transfer factor in MC by inverting lepton veto (right)
  - Reduced by t_{mod}
- **W lepton** (besides top)
  - W+Jets, WW with 1 lepton
  - Genuine MET from neutrino
- **$t\bar{t}$ semi-leptonic**
  - High M_T requirement kills this background
  - Predicted from simulation
- **Z->MET**
  - Produced with W or $t\bar{t}$
  - May have actual hadronic top

$$N^{SR}_{lost-\ell} = N^{CR}_{\ell\ell} \times \frac{M^{SR}_{lost-\ell}}{M^{CR}_{\ell\ell}}$$



$\tilde{t}\bar{\tilde{t}}$ - t_{mod}: Modified topness

SUS-19-009

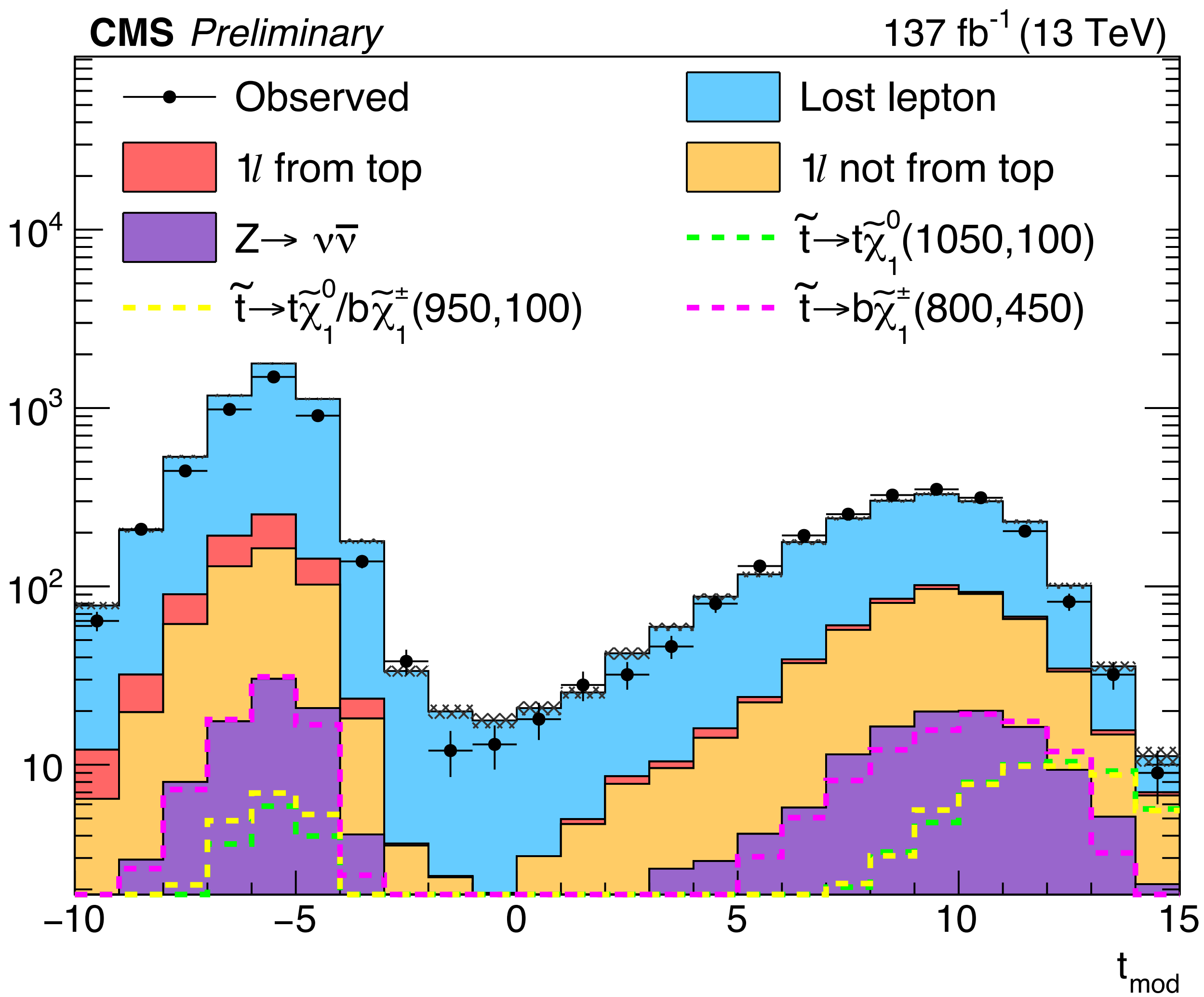


- χ^2-like variable
- $t\bar{t}$->2l system with missing lepton
  - Missing neutrino and a W
  - System under-constrained
- Expect low S (or t_{mod}) from $t\bar{t}$ and high from signal

$$t_{\text{mod}} = \ln(\min S)$$

$\tilde{t}\bar{\tilde{t}}$ - t_{mod}: Modified topness

SUS-19-009



- χ^2-like variable
- $t\bar{t}$->2l system with missing lepton
  - Missing neutrino and a W
  - System under-constrained
- Expect low S (or t_{mod}) from $t\bar{t}$ and high from signal

$$t_{\text{mod}} = \ln(\min S)$$

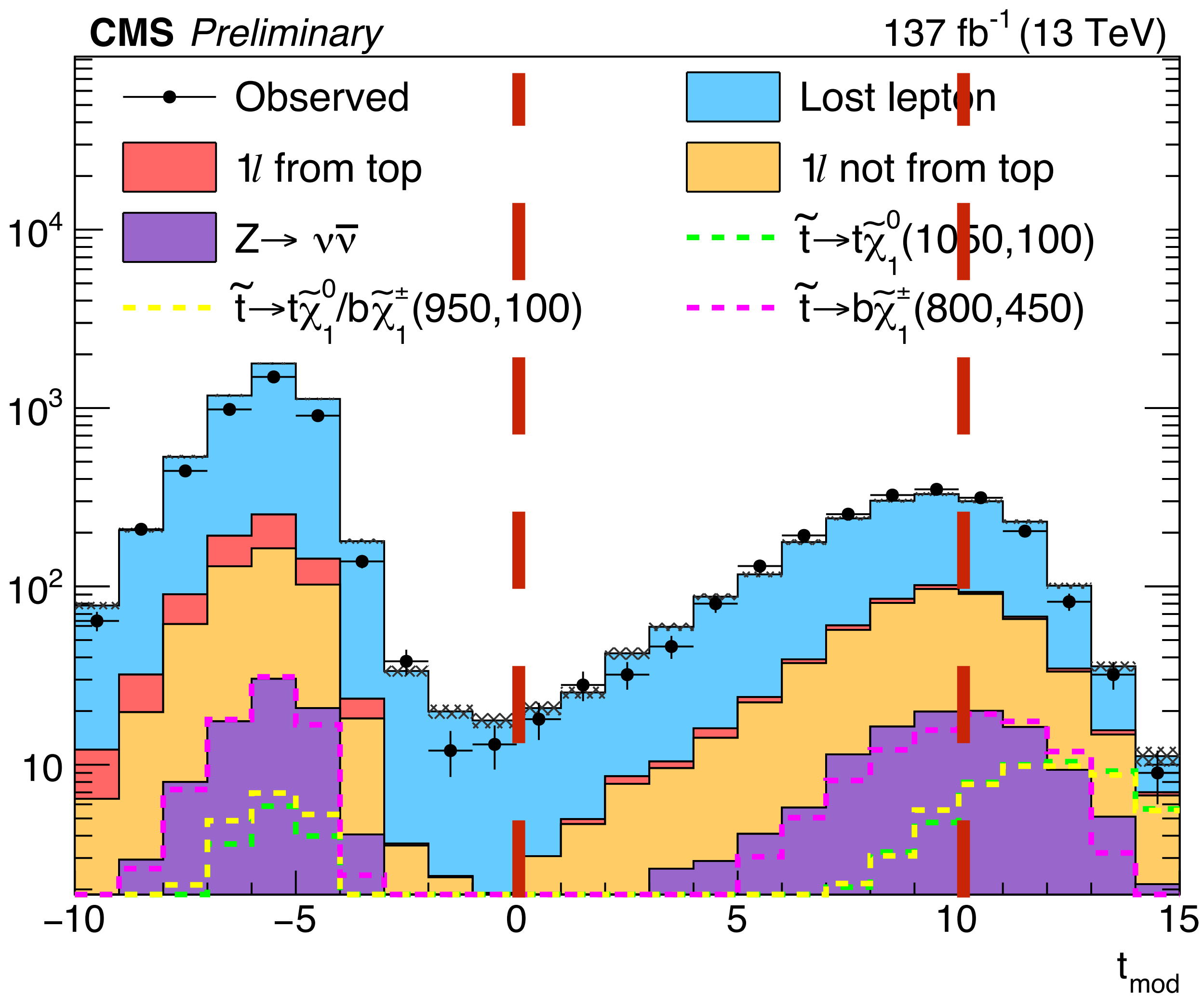
$$S = \frac{\left(m_{\text{W}}^2 - (p_\nu + p_\ell)^2\right)^2}{a_{\text{W}}^4} + \frac{\left(m_{\text{t}}^2 - (p_{\text{b}} + p_{\text{W}})^2\right)^2}{a_{\text{t}}^4}$$

W mass constraint from visible lepton

top mass constraint from missing lepton

$\tilde{t}\tilde{\bar{t}}$ - t_{mod}: Modified topness

SUS-19-009



- χ^2-like variable
- $t\bar{t}$->2l system with missing lepton
  - Missing neutrino and a W
  - System under-constrained
- Expect low S (or t_{mod}) from $t\bar{t}$ and high from signal

$$t_{\text{mod}} = \ln(\min S)$$

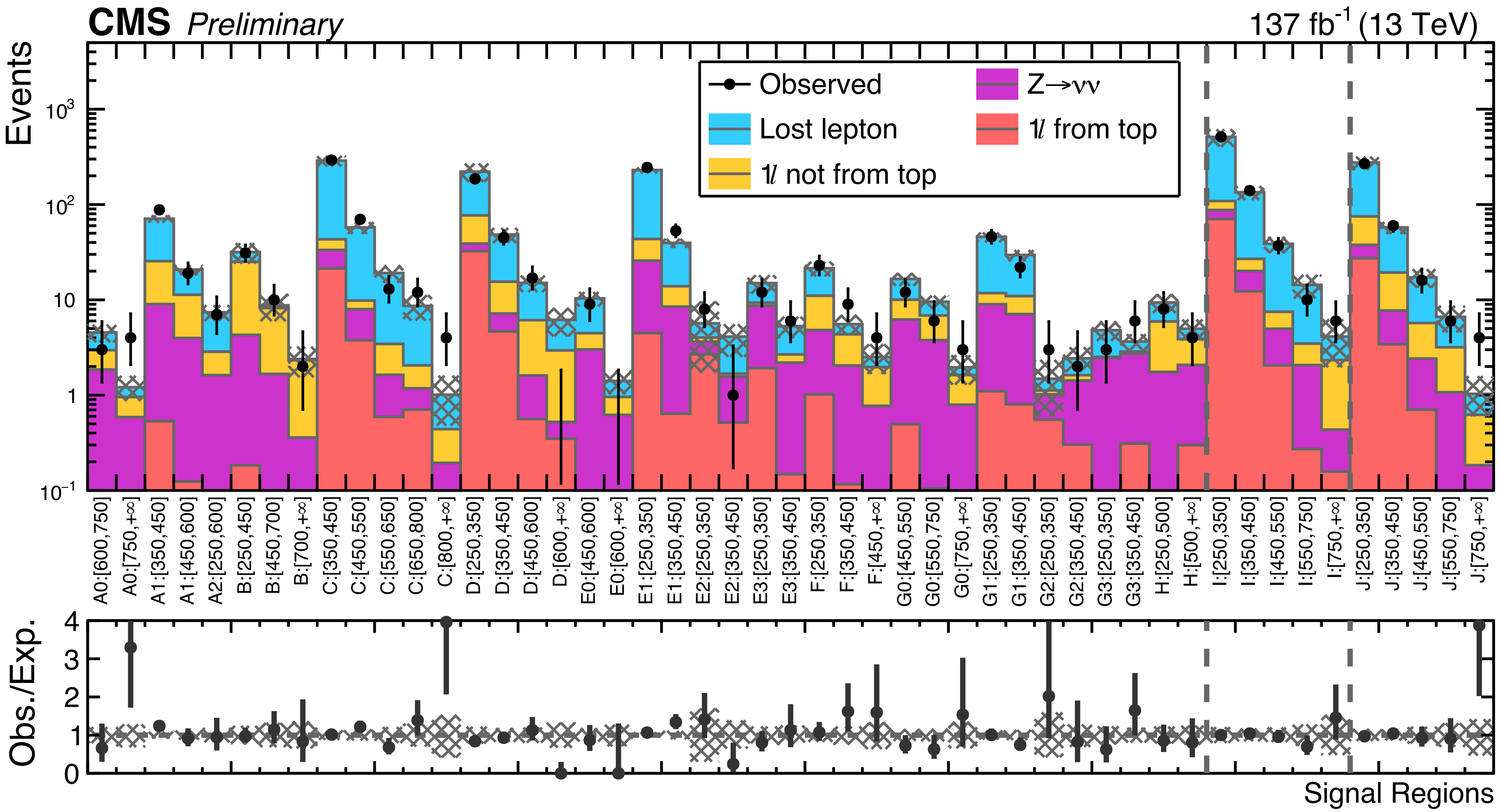
$$S = \frac{\left(m_{\text{W}}^2 - (p_\nu + p_\ell)^2\right)^2}{a_{\text{W}}^4} + \frac{\left(m_{\text{t}}^2 - (p_{\text{b}} + p_{\text{W}})^2\right)^2}{a_{\text{t}}^4}$$

W mass constraint from visible lepton

top mass constraint from missing lepton

$\tilde{t}\bar{\tilde{t}}$ - Results

SUS-19-009



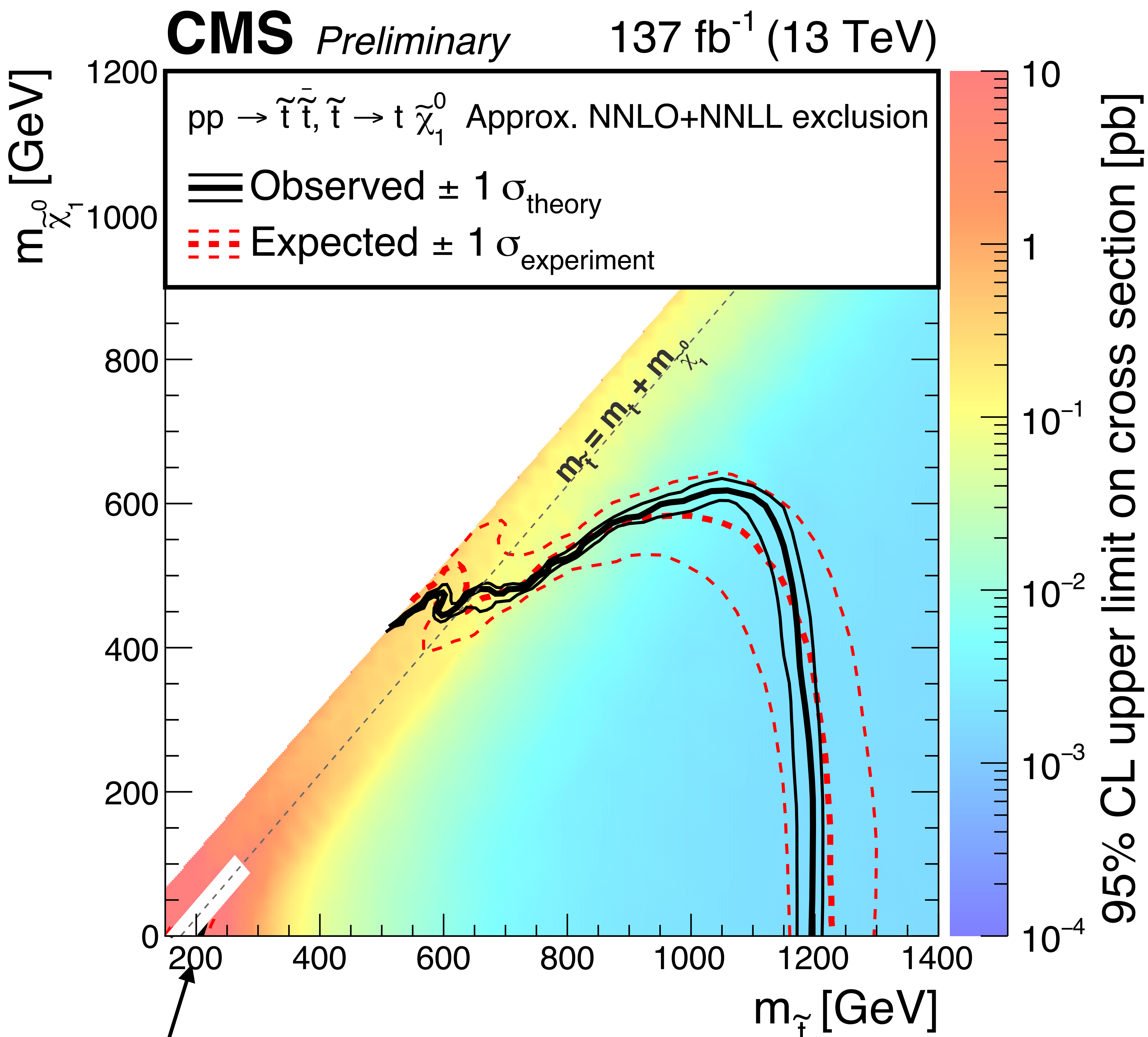
	N_J	t_{mod}	M_{lb} [GeV]
A	2–3	> 10	≤ 175
B	2–3	> 10	> 175
C	≥ 4	≤ 0	≤ 175
D	≥ 4	≤ 0	> 175
E	≥ 4	0–10	≤ 175
F	≥ 4	0–10	> 175
G	≥ 4	> 10	≤ 175
H	≥ 4	> 10	> 175

X0: Inclusive
X1: Untagged
X2: Boosted top
X3: Resolved top

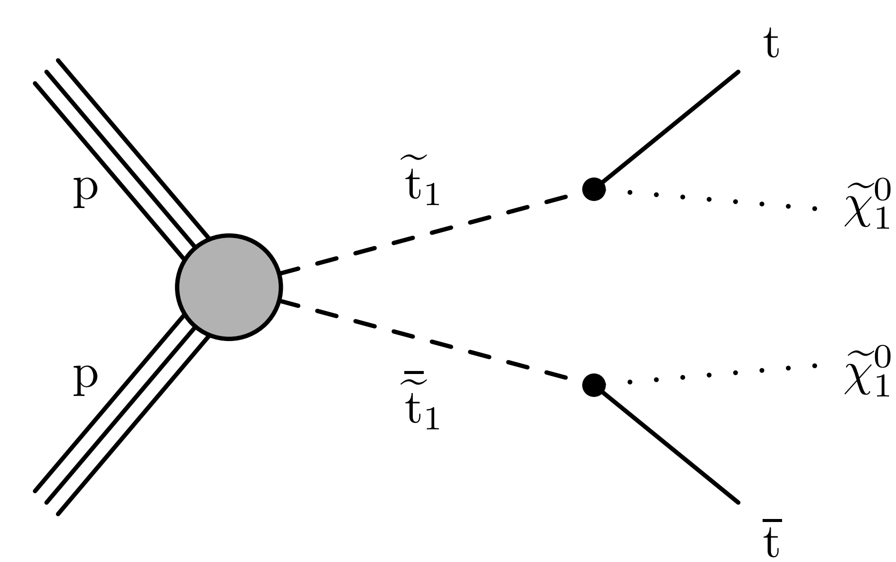
I: $N_J \geq 5$, $N_{b,med} \geq 1$
J: $N_J \geq 3$, $N_{b,soft} \geq 1$

$\tilde{t}\bar{\tilde{t}}$ - Results

SUS-19-009

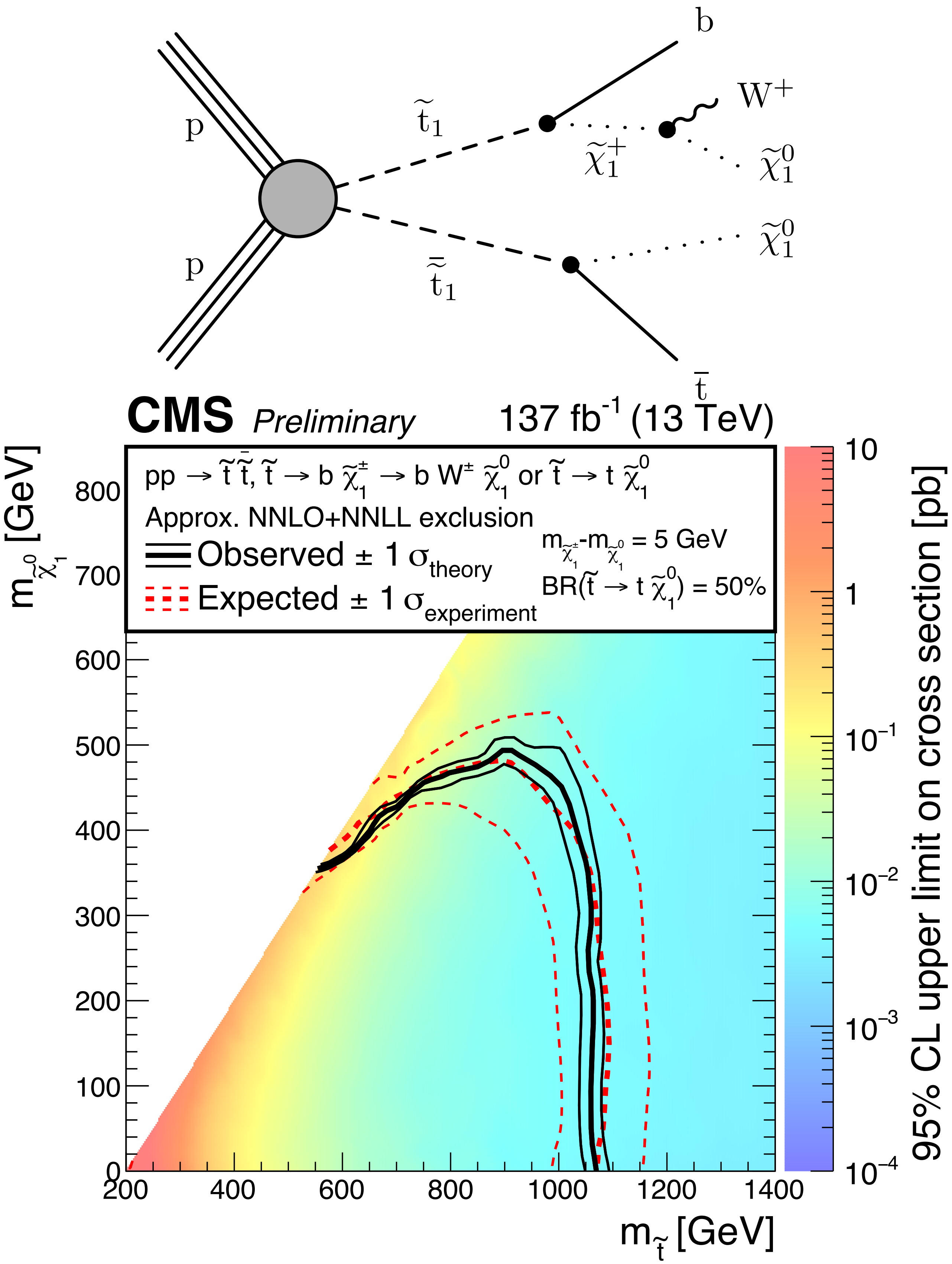


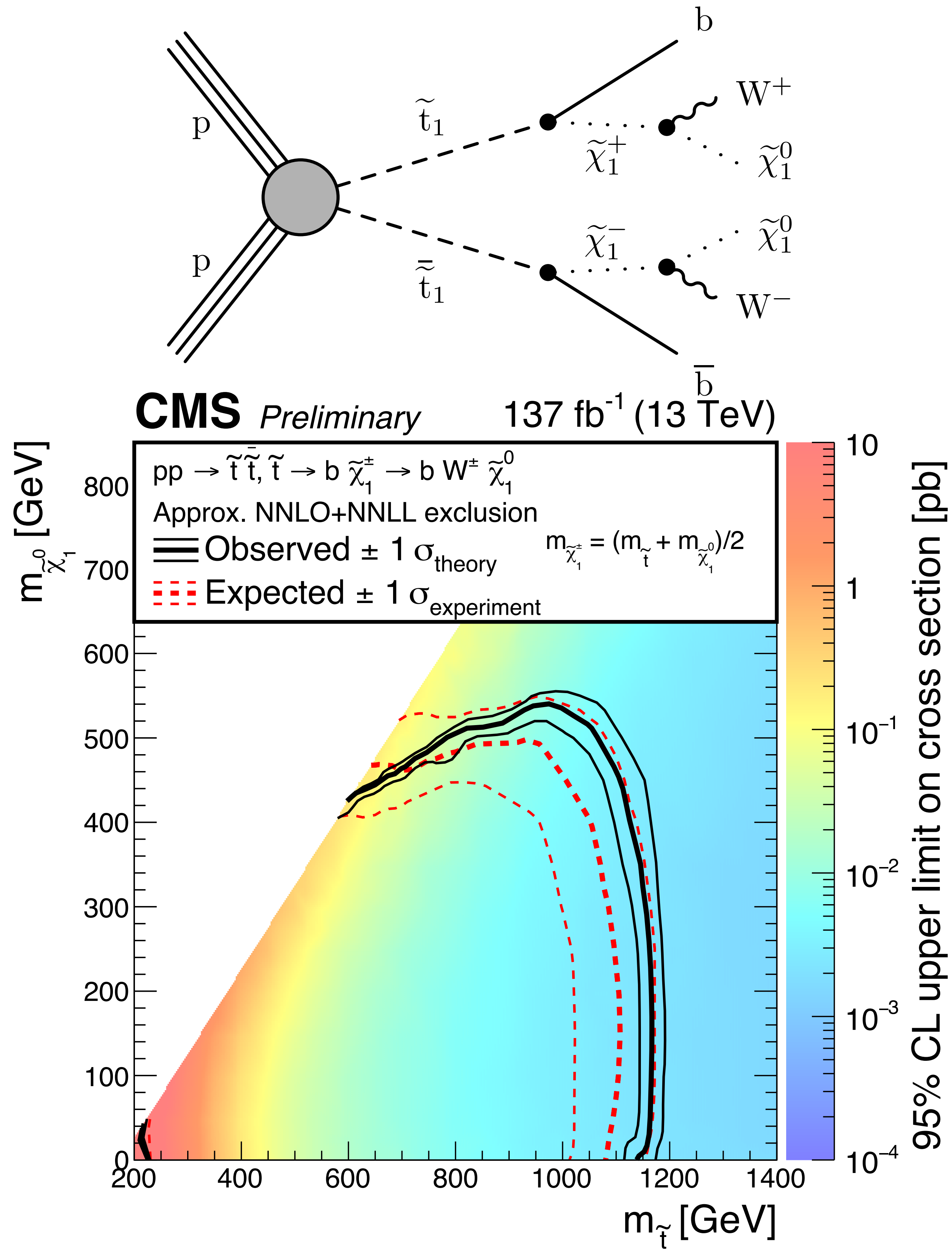
$m_{\tilde{t}}$ - m_t - m_χ < 25 GeV



$\tilde{t}\bar{\tilde{t}}$ - Results

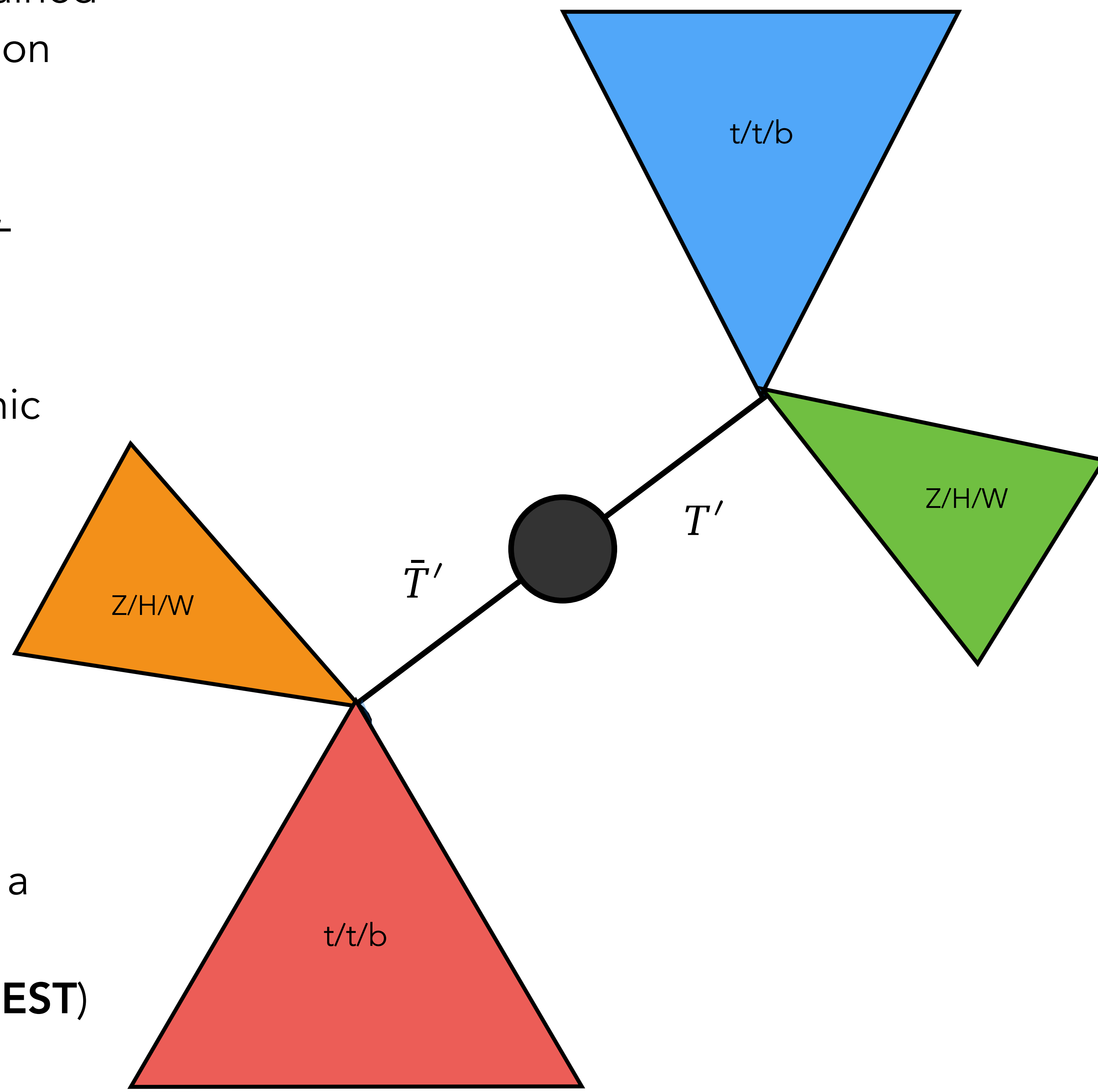
SUS-19-009





VLQ - Strategy

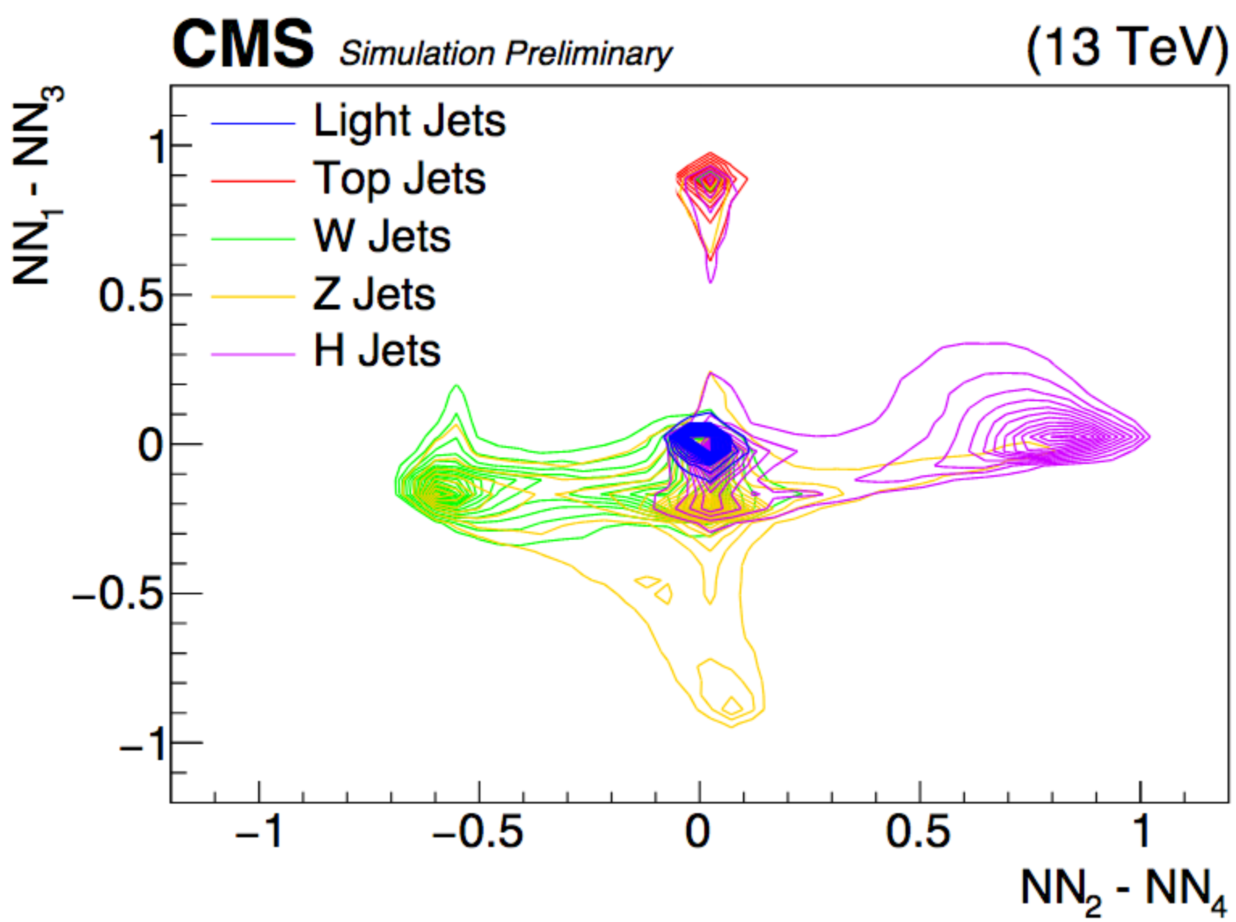
- **Vector-like quarks (VLQs)** are not constrained by observed Higgs production cross section
    - Left- and right-handed components transform equally under SU(2) ('vector-like')
    - Consider pair production in all-hadronic final states
- Two searches performed with different tagging approaches:
    - One targeting **T->bW** decay
    - One targeting **all modes,** utilizing a novel neural net based tagger: "Boosted Event Shape Tagger" (**BEST**)
- Perform these searches in the 2016 dataset (35.9 fb^{-1})

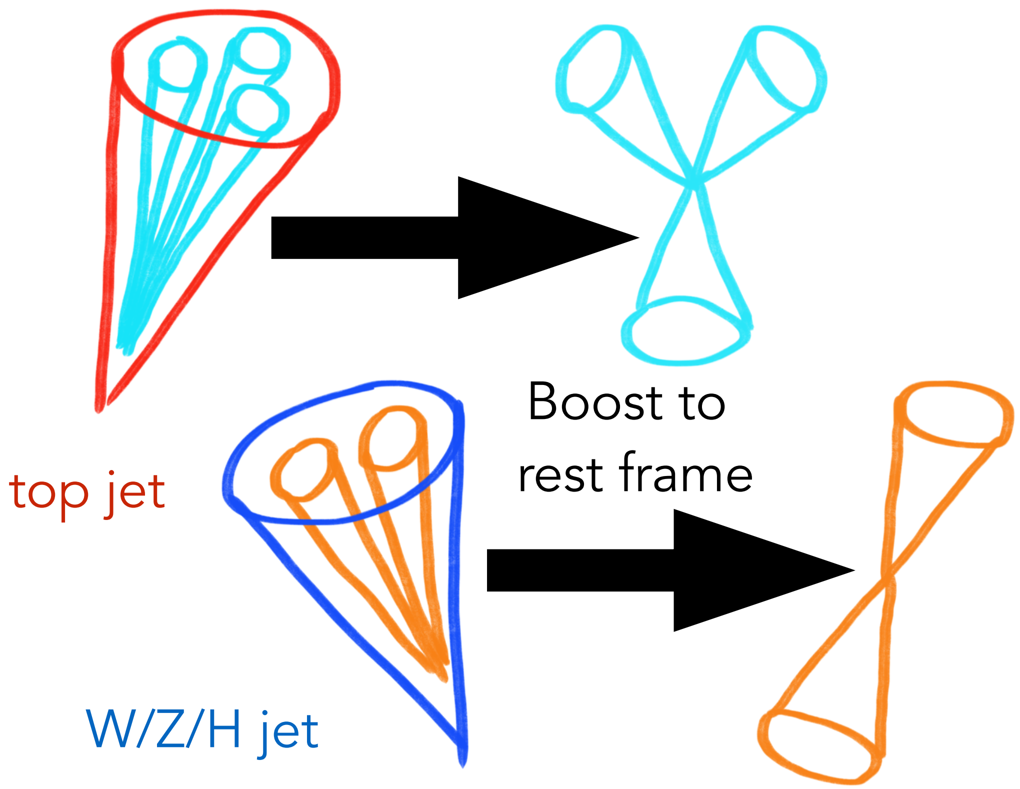


VLQ - BEST Procedure

B2G-18-005

- Hypothesize different particles as origin of jet: top, H, Z, W
  - Jet constituents should be isotropic in the rest frame of that particle
  - Use particle masses to apply different boosts to jet constituents
  - In each reference frame, calculate event shape variables



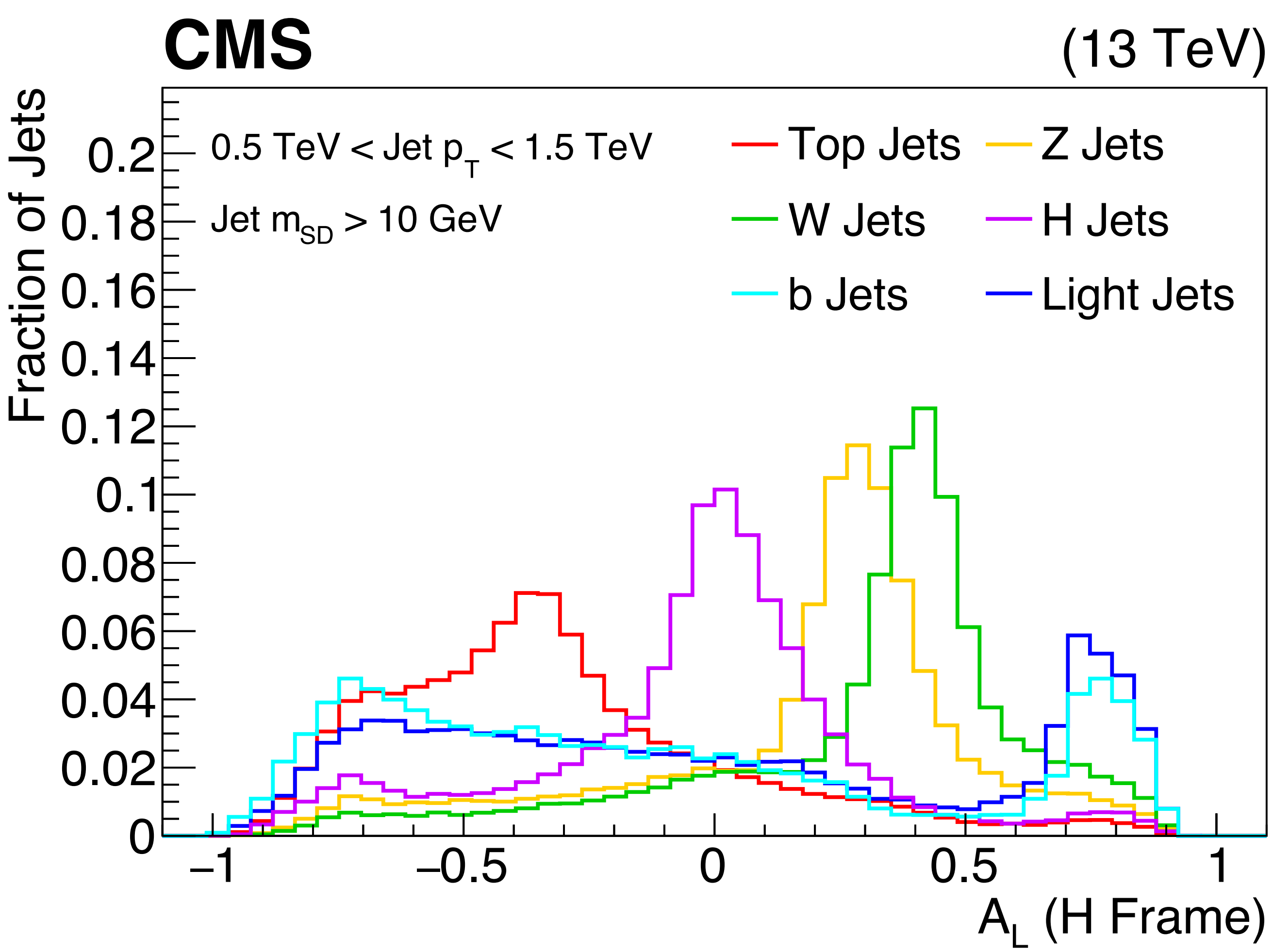


- Also use some lab frame inputs, such as soft drop mass and CSV
- Six possible classifications:
  - t, W, Z, H, b, light [u/d/s/c/g]

VLQ - BEST Inputs

B2G-18-005

- **Longitudinal Asymmetry** - momentum balance along jet axis in rest frame
  - Balanced momentum (A_L = 0) more likely in correct frame
  - Calculate value in each frame

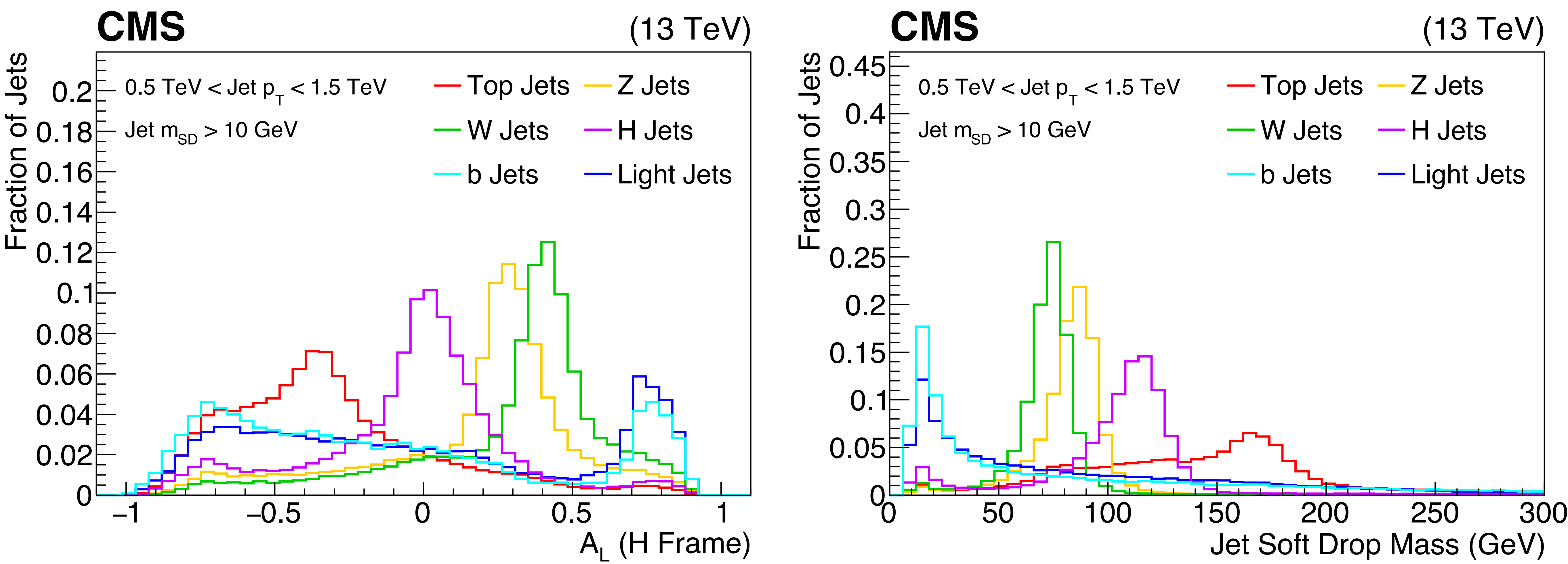


$$A_L = \frac{\sum_{jet} p_L^{jet}}{\sum_{jet} p^{jet}}$$

VLQ - BEST Inputs

B2G-18-005

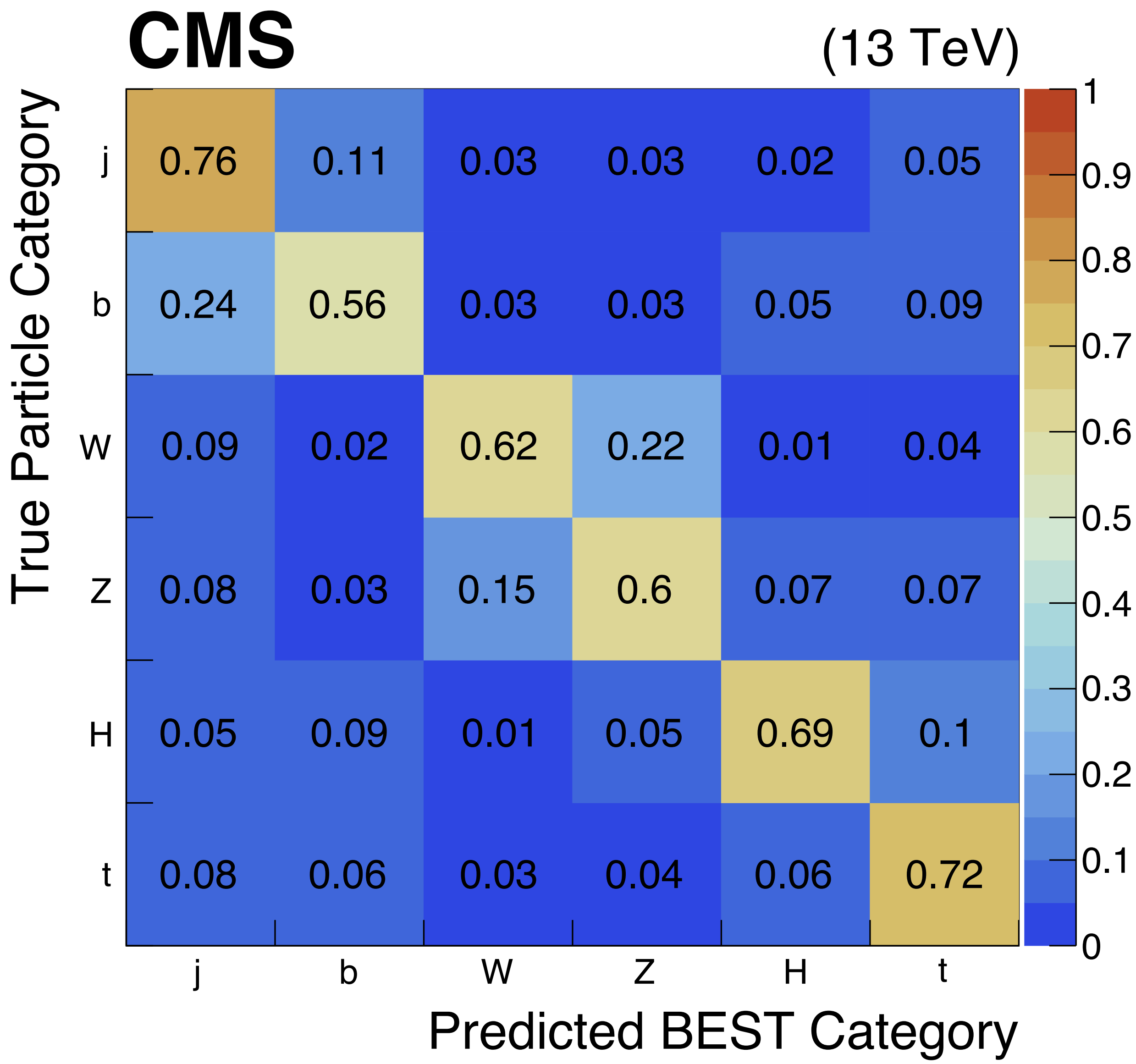
- **Longitudinal Asymmetry** - momentum balance along jet axis in rest frame
  - Balanced momentum (A_L = 0) more likely in correct frame
  - Calculate value in each frame

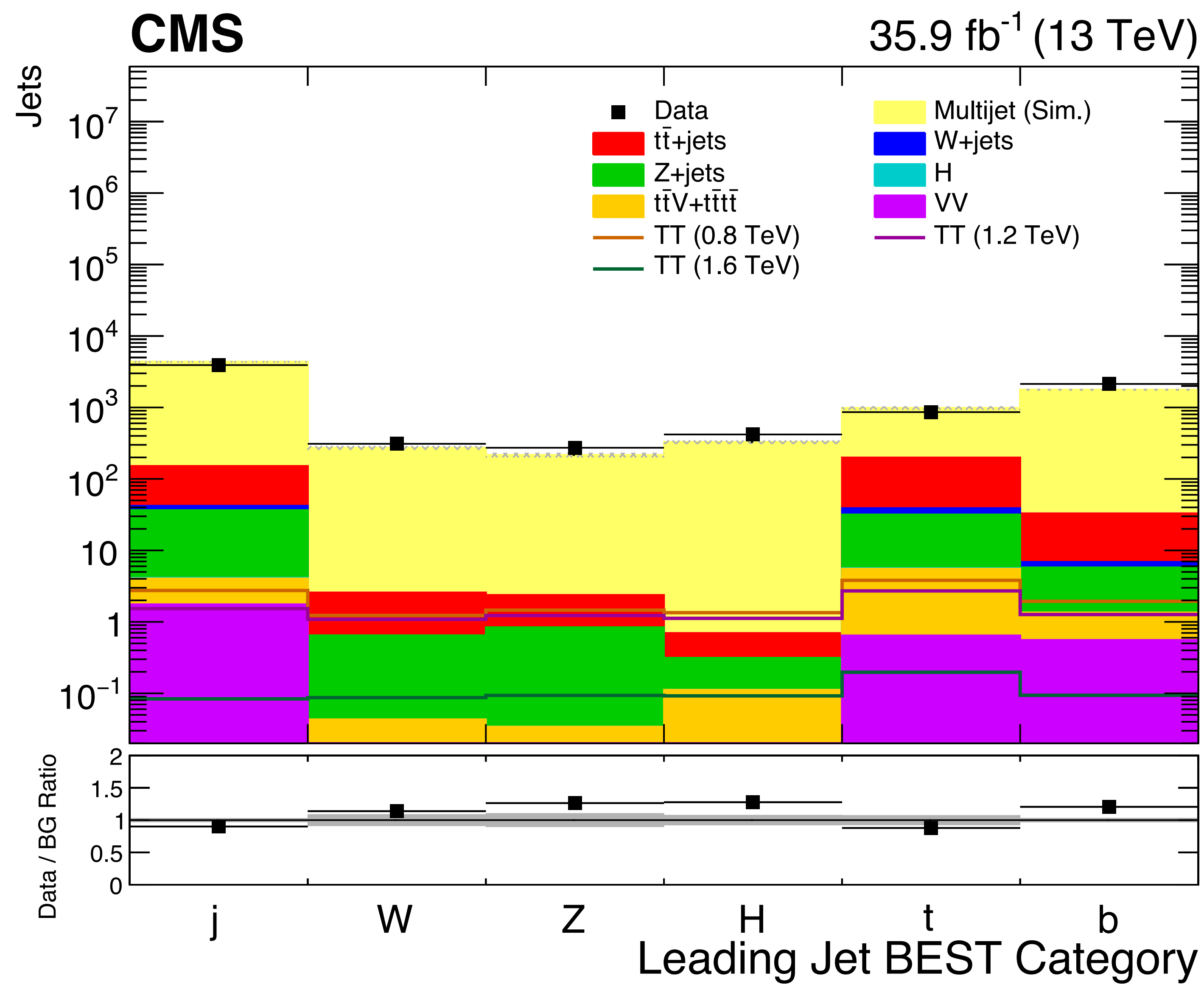


- **Soft Drop Mass** - Well known variable for identifying jets from heavy objects
  - Small cut on m_{SD} > 10 GeV to reduce light background

VLQ - BEST Performance

B2G-18-005



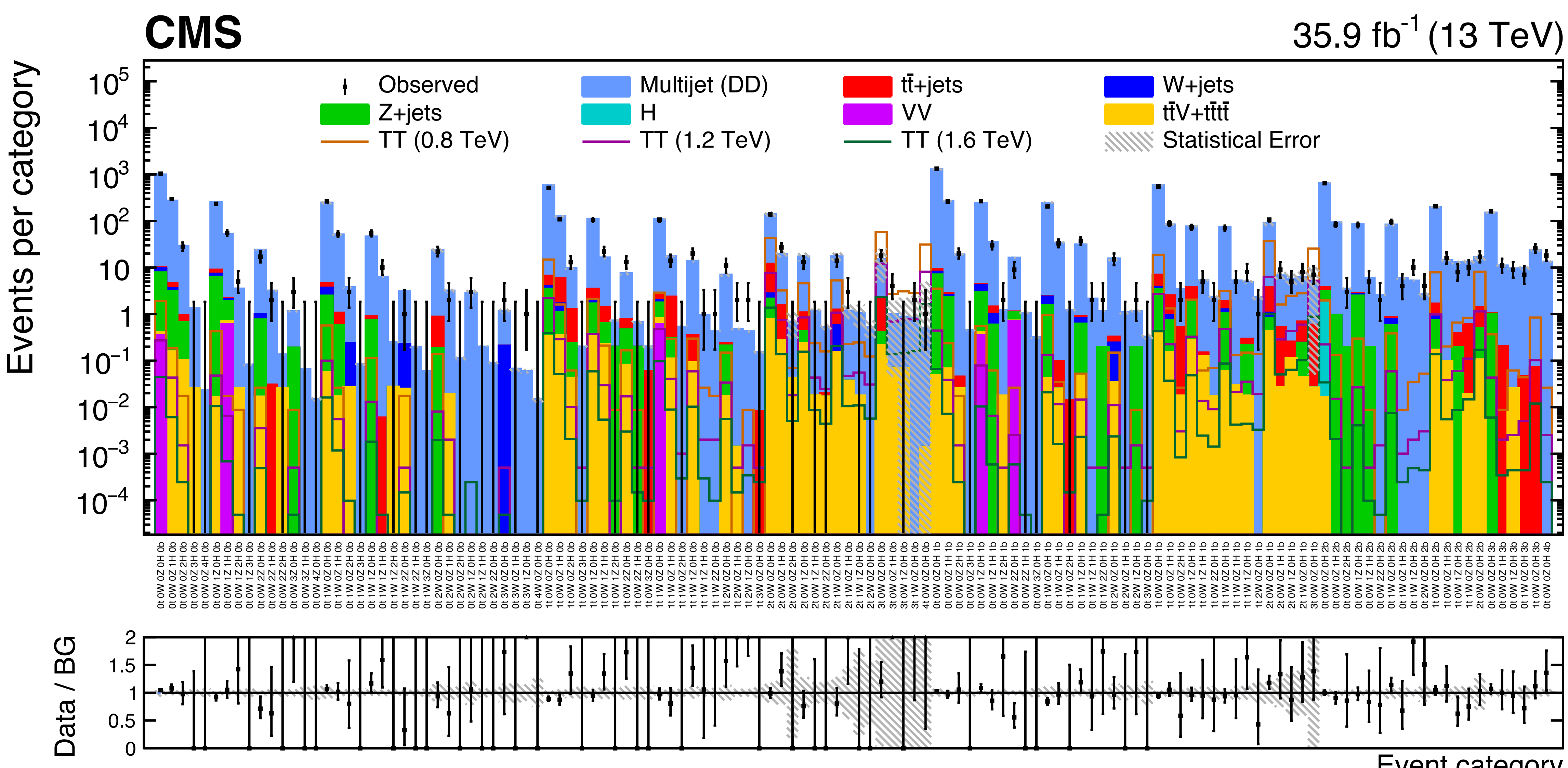


- Train fully connected deep neural network on 500K simulated jets
  - Require jets with p_T > 500 GeV, to ensure decay products are fully merged
- Achieve good discrimination between particle types
- Tagging rates found to be close to simulation

VLQ - Event Selection

B2G-18-005

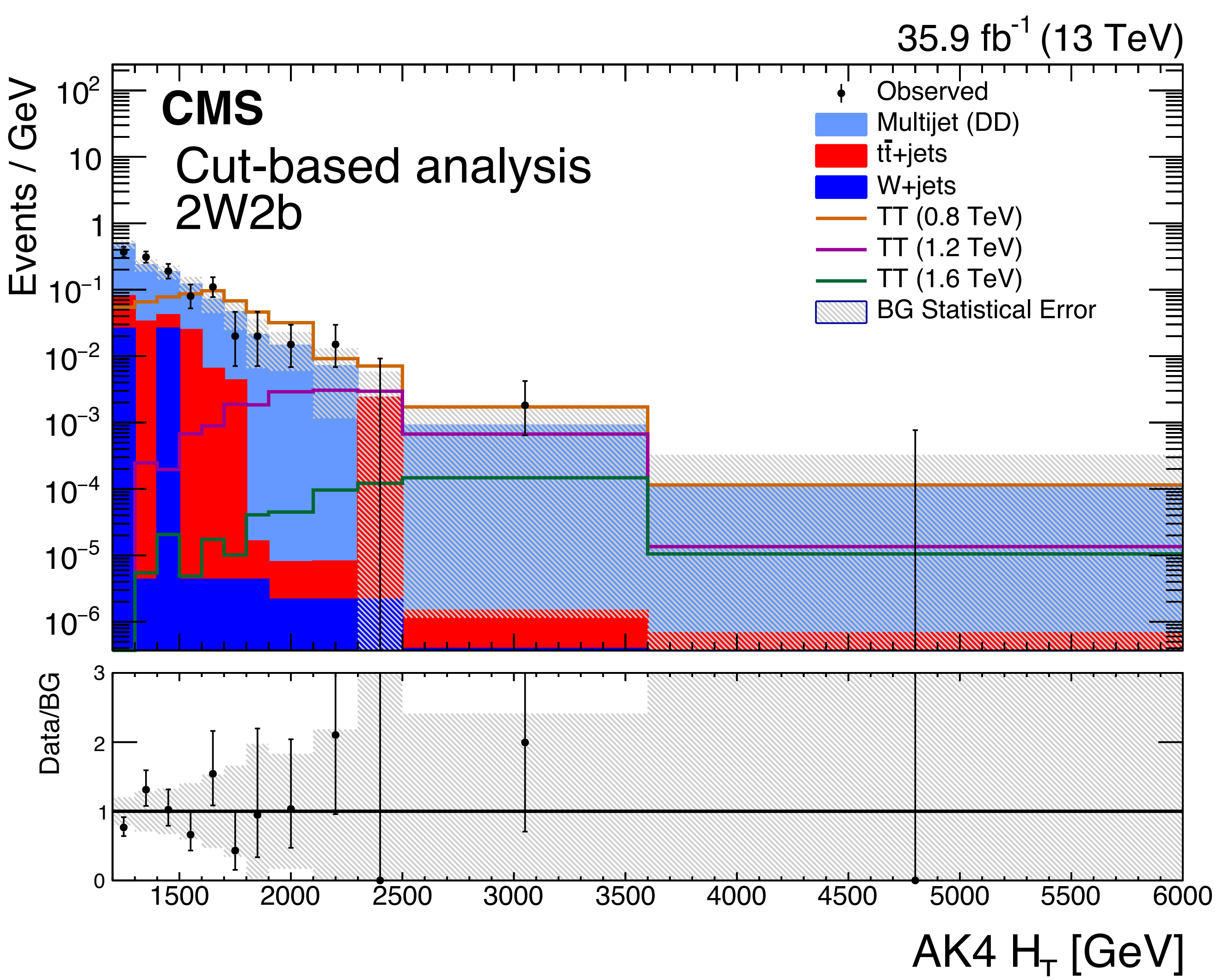
- **Neural Net**-based analysis signal regions include exactly 4 jets
  - Count multiplicities of objects according to BEST classification
    - Unique set of (N_t, N_H, N_W, N_Z, N_b, N_j), sum of the N_i = 4
  - **126 independent signal regions**



VLQ - Event Selection

B2G-18-005

- **Neural Net**-based analysis signal regions include exactly 4 jets
  - Count multiplicities of objects according to BEST classification
    - Unique set of (N_t, N_H, N_W, N_Z, N_b, N_j), sum of the N_i = 4
  - **126 independent signal regions**



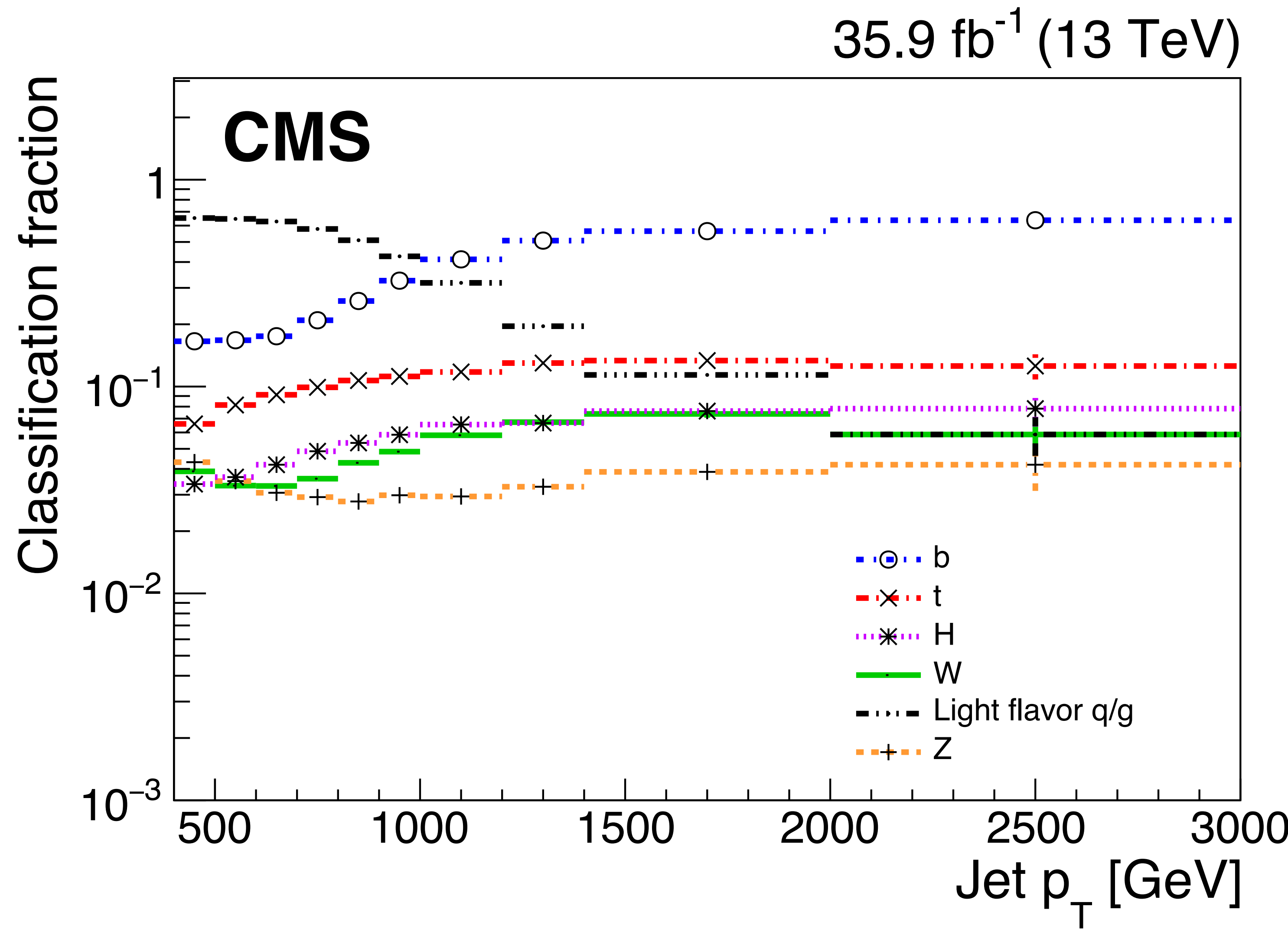
- **Cut-based** analysis counts multiplicities of W, b tags, requiring > 1 of each tag
- Tag W jets via cuts on Soft Drop Mass and τ_{21}
- Tag b jets with cuts on CSV discriminant
- **4 independent signal regions**
- In both analyses, use H_T distribution to discriminate signal from background

VLQ - Background Estimate

B2G-18-005

- Dominant background is QCD
  - Estimate this from observed data
- **Cut-based** analysis uses ABCD method
  - H_T, VLQ mass difference sideband
- **NN-based** analysis measures a tagging rate in independent 3 jet sample
  - 3-jet events entirely QCD
  - Apply rates to whole 4-jet sample to get QCD distribution per category
  - Other backgrounds taken from simulation: W+jets, $t\bar{t}$

$$\epsilon_X(p_T) = \frac{\text{Number of jets with BEST class X}}{\text{Number of jets}}$$

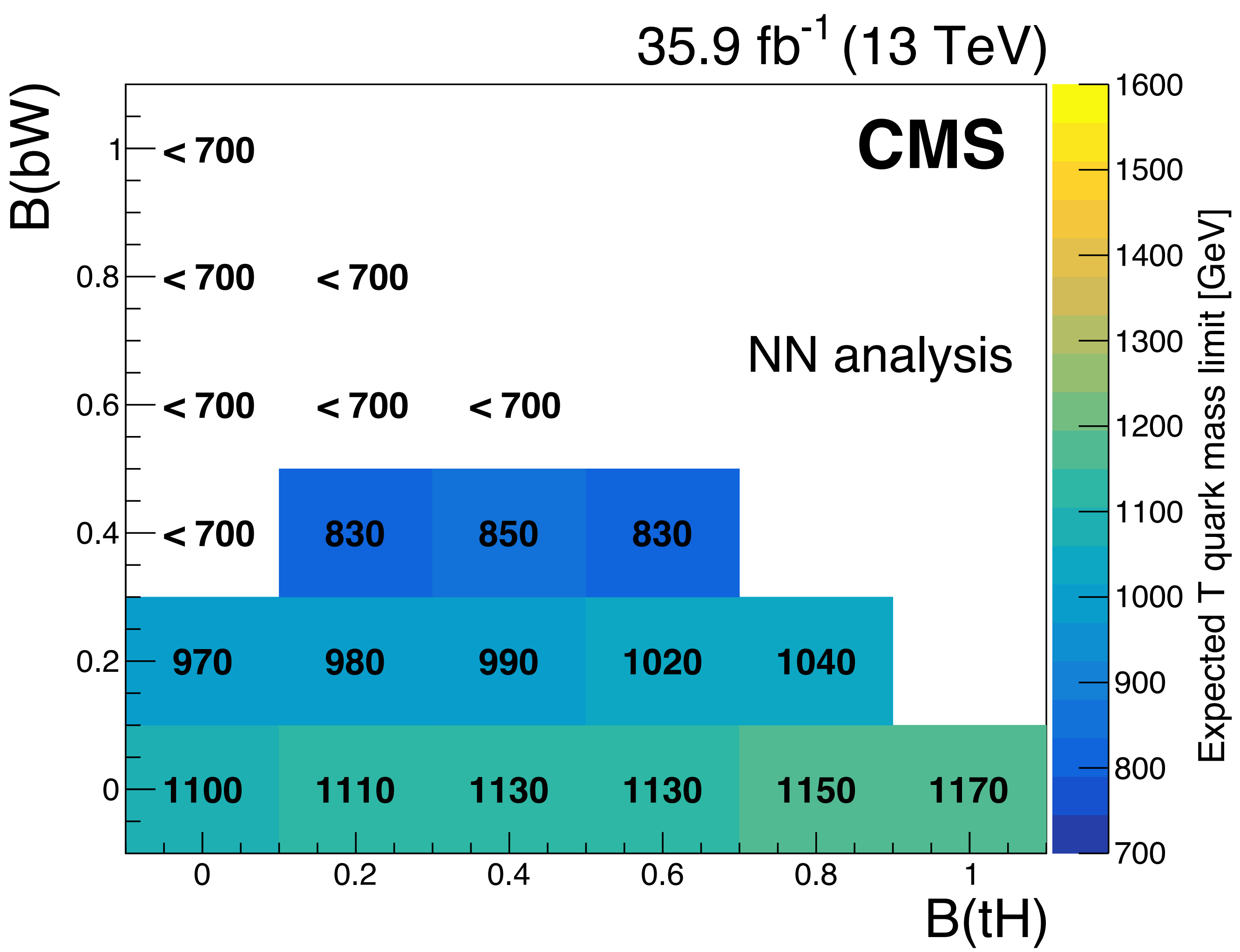


$$R = \sum_{events} [r] = \sum_{events} [\sum_{perms} (\prod_{i=1}^{4} \epsilon_{X_i}(p_T(i)))]$$

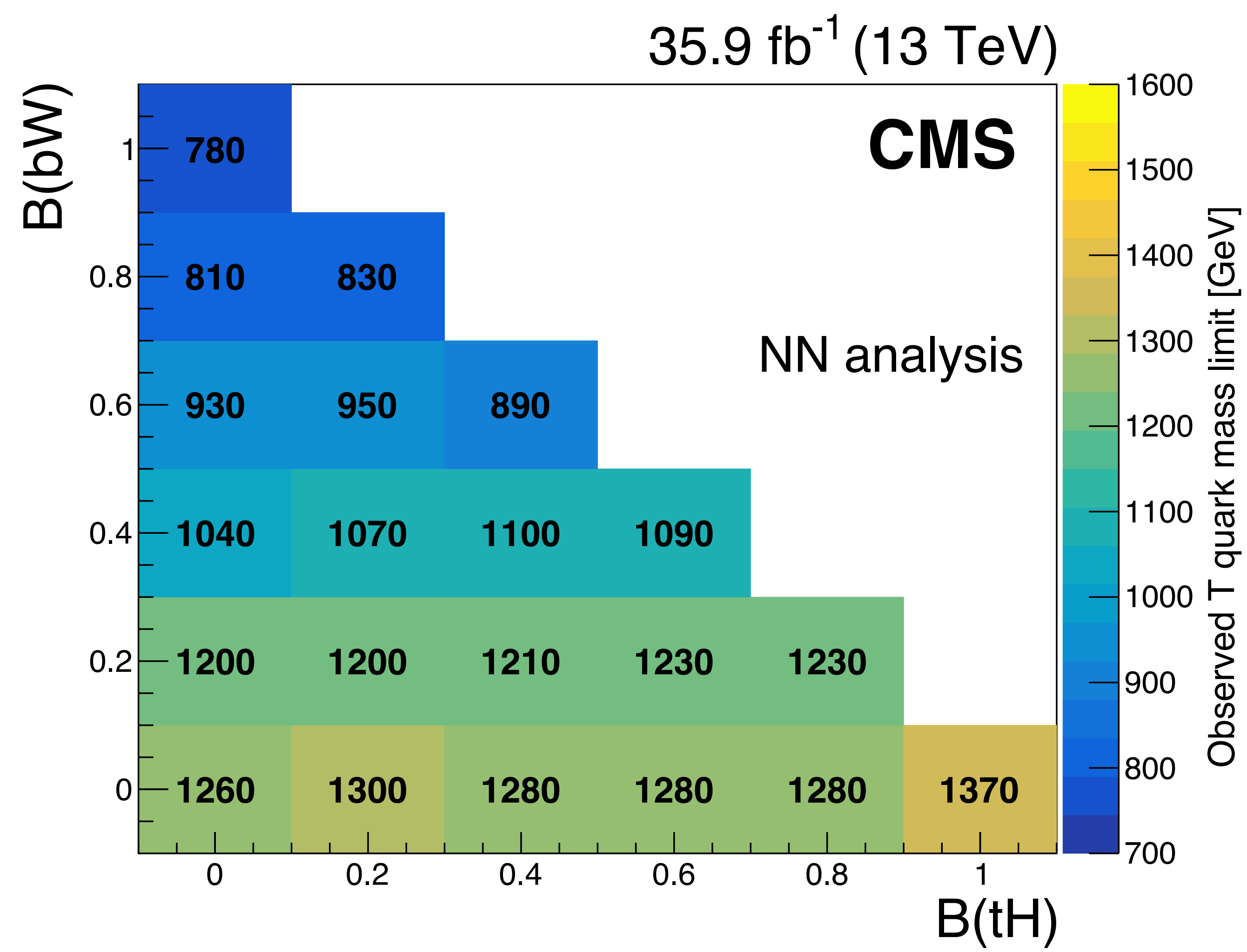
VLQ Results — TT Limits By Branching Fraction

B2G-18-005

- Scan limits over the branching fractions of the VLQ
  - BEST sensitive to T->tH and T->tZ



Expected

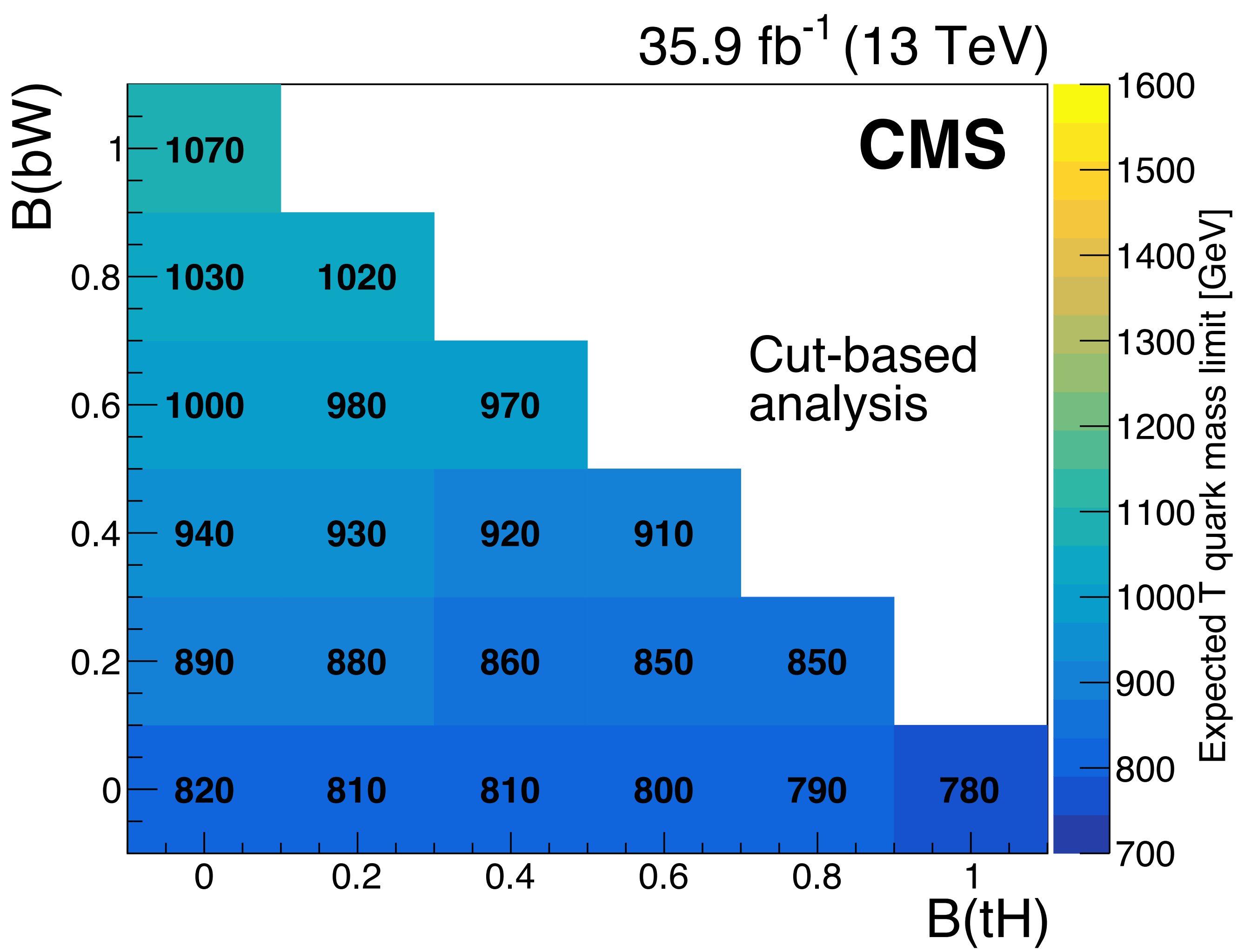


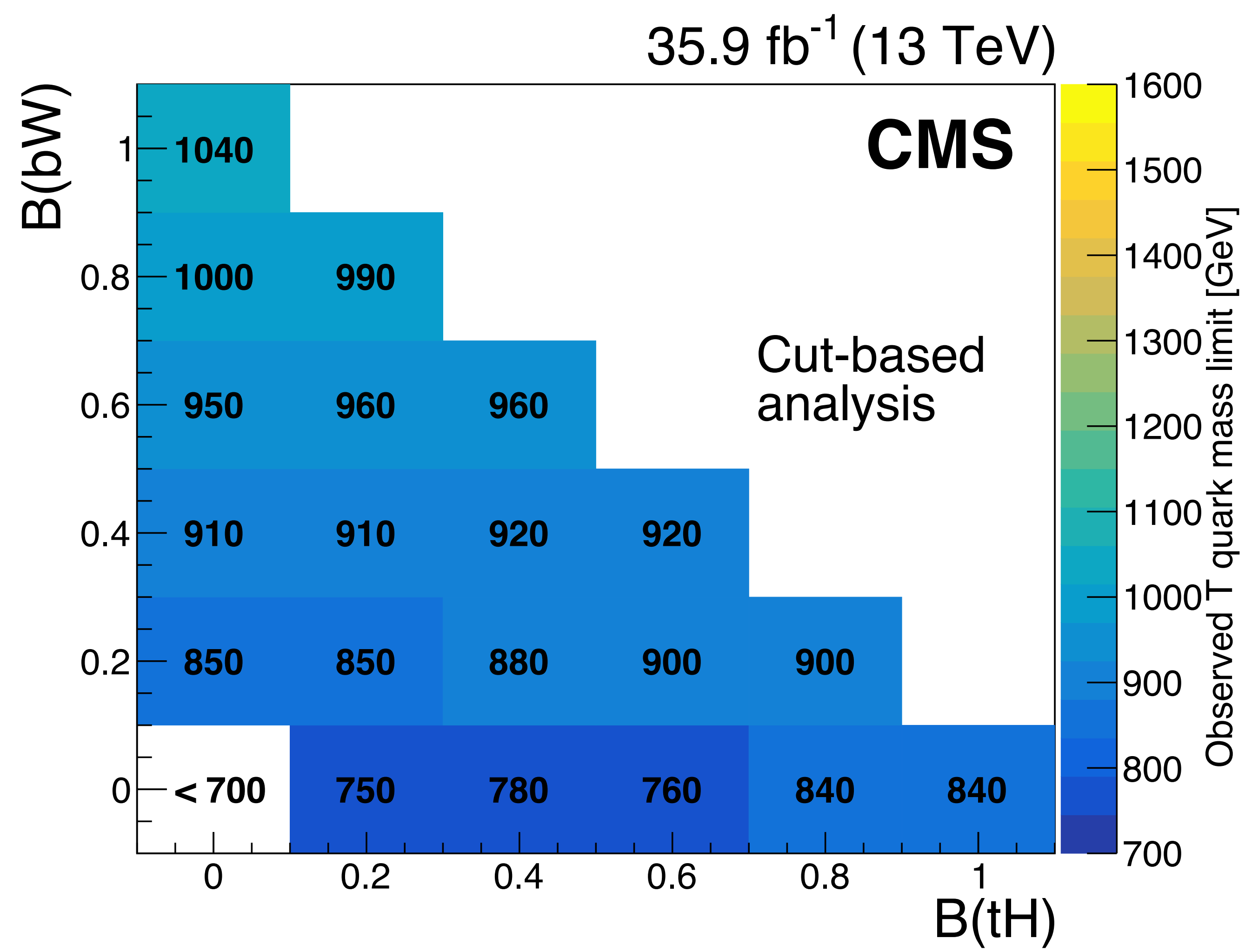
Observed

VLQ Results — TT Limits By Branching Fraction

B2G-18-005

- Scan limits over the branching fractions of the VLQ
  - Cut-based analysis performs better in T->bW corner



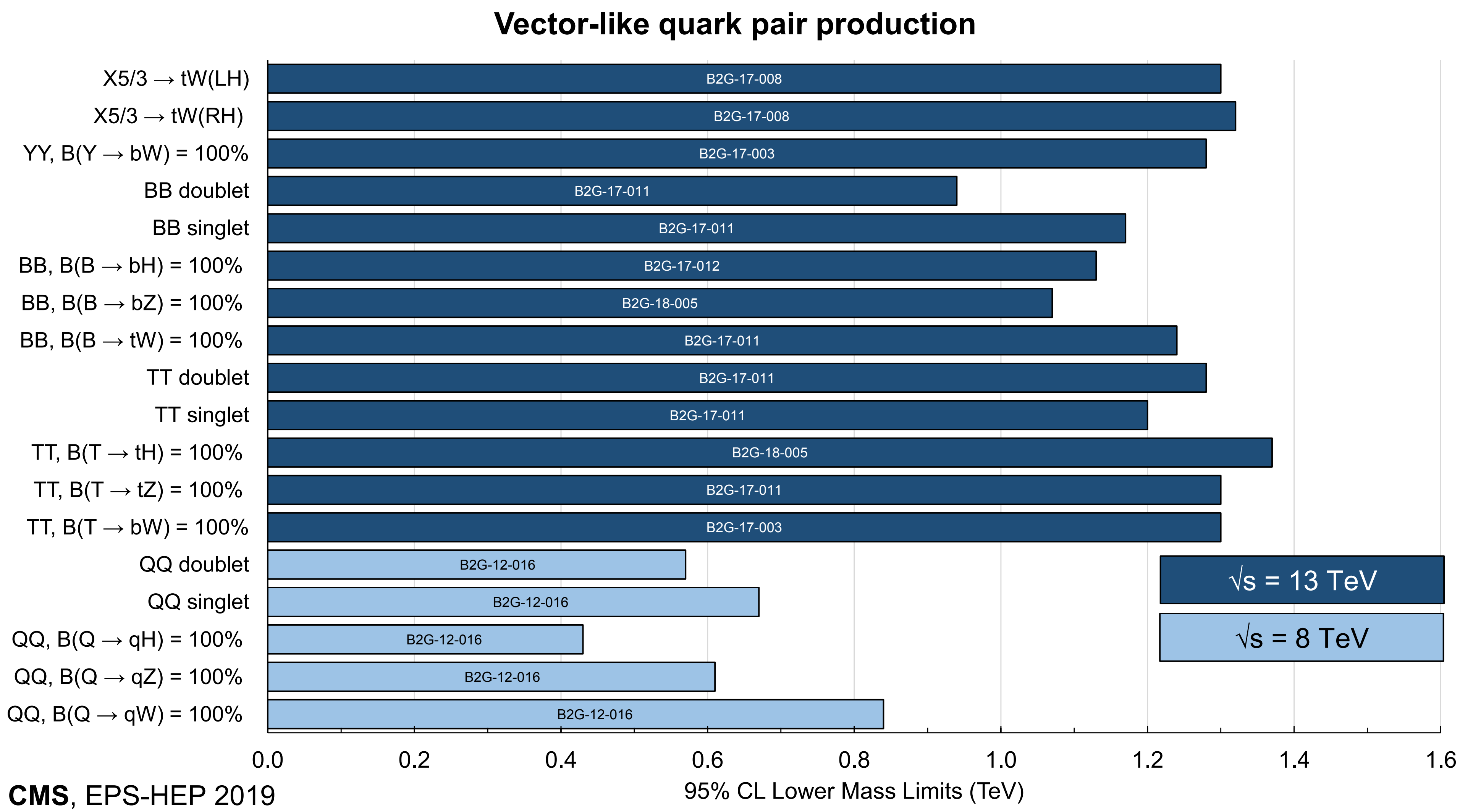


Expected

Observed

Summary

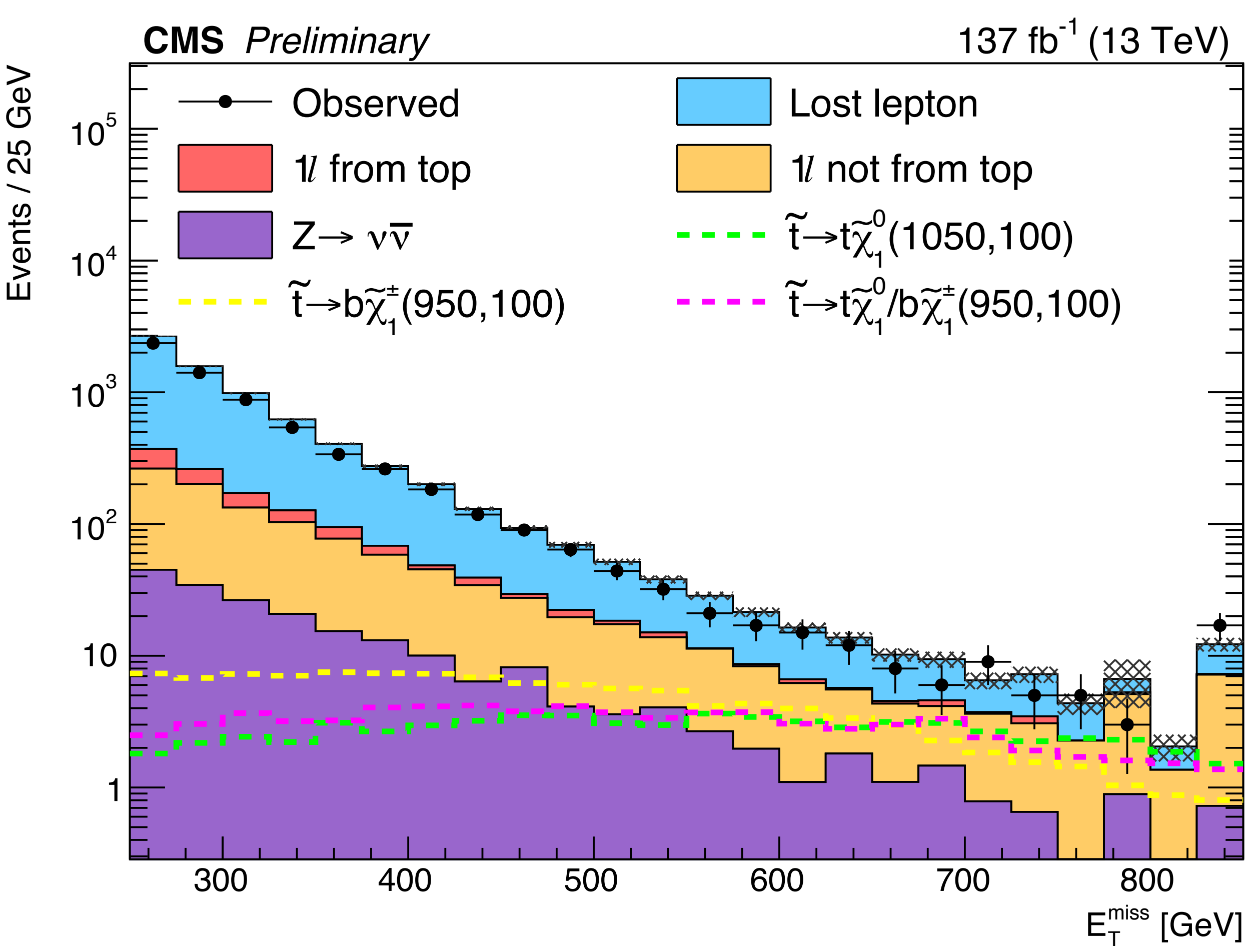
- Several new searches by CMS using boosted tops have been presented
    - Both make use of new tagging techniques based on deep neural nets
    - More advances in this field being made, look forward to even more results at the next BOOST conference!
- See Meenakshi Narain's talk from this conference for more information on CMS tagging techniques featuring top quarks

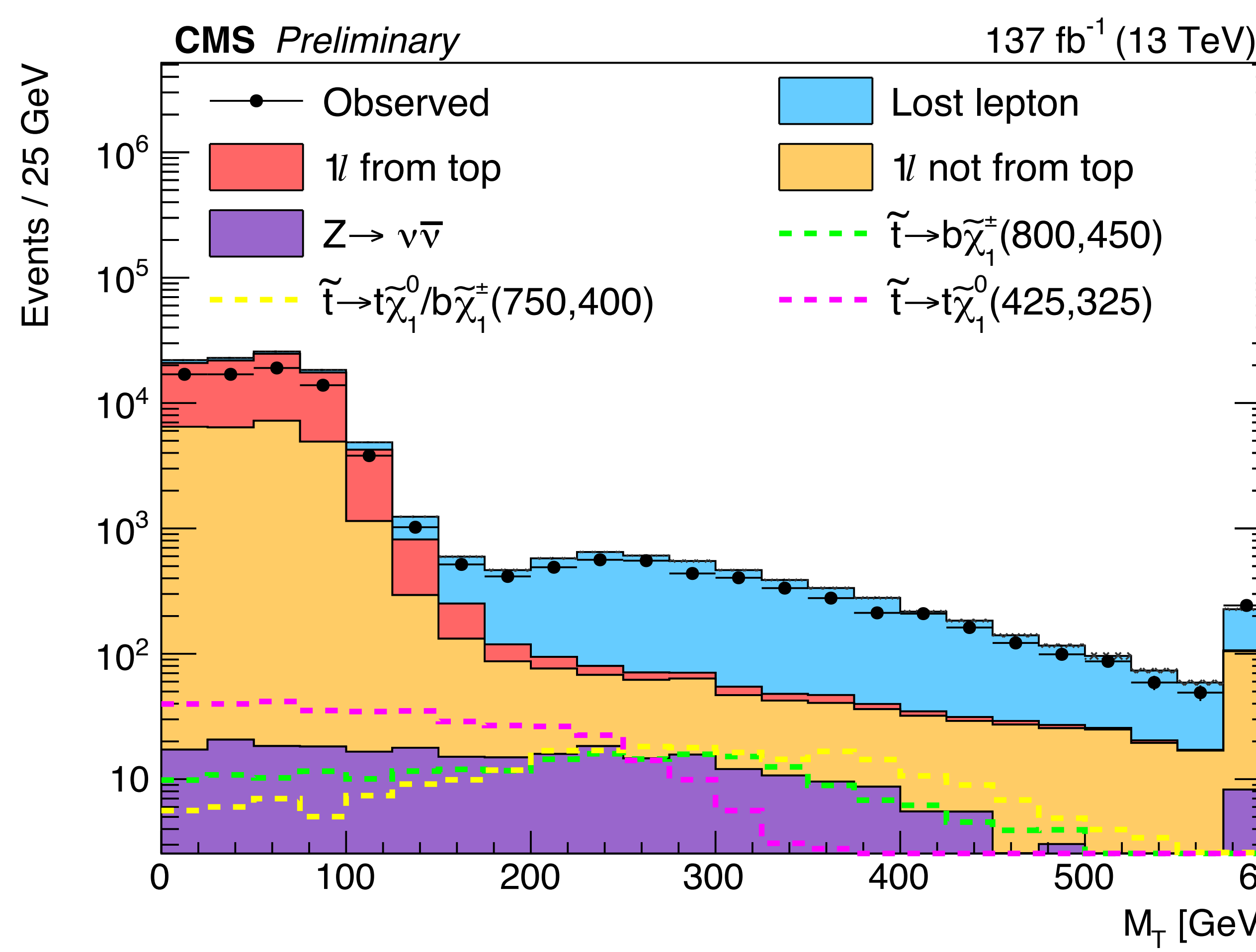


Backup Material

SUS - Pre-selection

- Require exactly one isolated lepton
- Veto extra isolated leptons with $p_T > 5$ GeV
- Veto tracks (potentially from taus) with $p_T > 10$ GeV
- Medium b-tag requirement (special soft >20 GeV tag for compressed spectra)
- $M_T > 150$
- MET > 250, with a separation from leading jets $\Delta\Phi > 0.8$ (0.5 for compressed)
- At least 2 AK4 jets





SUS - Signal Regions

SUS-19-009

- Uncompressed spectra:
  - $m_{\tilde{t}} > m_t + m_\chi$
  - Split categories by:
    - Number of jets
    - t_{mod}
    - M_{lb}
    - MET
  - **39 categories**
- Compressed spectra:
  - $m_{\tilde{t}} \sim m_t + m_\chi$ or $\sim m_W + m_b + m_\chi$
  - Softer b jets
  - Leading jet from ISR
  - Veto high p_T lepton
  - Split by MET bin
  - Number of jets: 3 or 5
  - **10 categories**

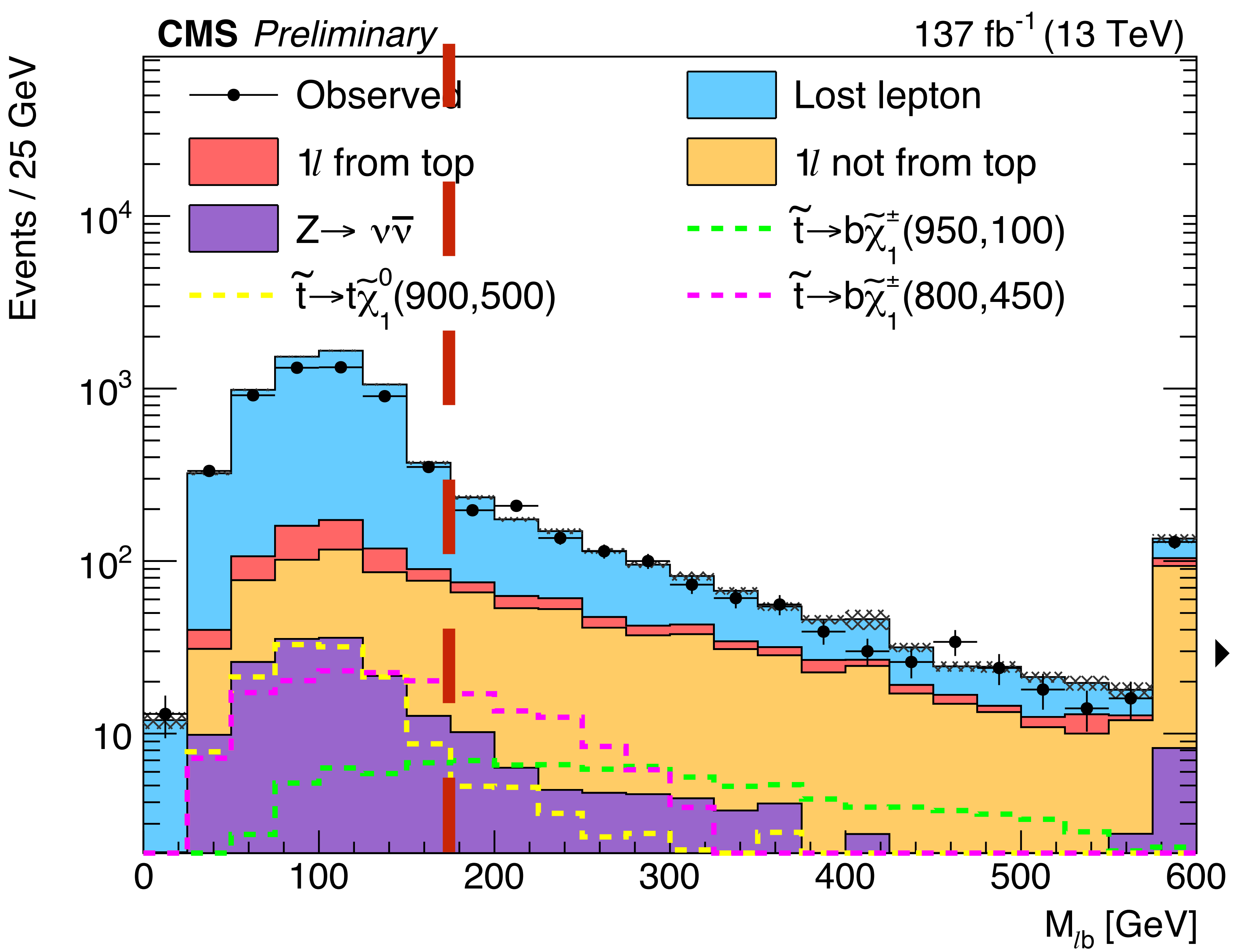
Uncompressed

Label	N_J	t_{mod}	$M_{\ell b}$ [GeV]	top tagging category	E_T^{miss} bins [GeV]
A0	2–3	≥ 10	< 175	–	[600, 750, $+\infty$]
A1				U	[350, 450, 600]
A2				M	[250, 600]
B			≥ 175	–	[250, 450, 700, $+\infty$]
C	≥ 4	< 0	< 175	–	[350, 450, 550, 650, 800, $+\infty$]
D			≥ 175	–	[250, 350, 450, 600, $+\infty$]
E0		0–10	< 175	–	[450, 600, $+\infty$]
E1				U	[250, 350, 450]
E2				M	[250, 350, 450]
E3				R	[250, 350, 450]
F			≥ 175	–	[250, 350, 450, $+\infty$]
G0		≥ 10	< 175	–	[450, 550, 750, $+\infty$]
G1				U	[250, 350, 450]
G2				M	[250, 350, 450]
G3				R	[250, 350, 450]
H			≥ 175	–	[250, 500, $+\infty$]

Compressed

Compressed spectra with $\Delta m(\tilde{t}, \tilde{\chi}_1^0) \sim m_t$ (label: I)	
Selection criteria	$N_J \geq 5$, leading-p_T jet not b-tagged, $p_T^\ell < \max\left(50,\ 250 - 100 \times \Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^\ell)\right)$ GeV,
E_T^{miss} bins [GeV]	250–350, 350–450, 450–550, 550–750, > 750
Compressed spectra with $\Delta m(\tilde{t}, \tilde{\chi}_1^0) \sim m_W$ (label: J)	
Selection criteria	$N_J \geq 3$, leading-p_T jet not b-tagged, $p_T^\ell < \max\left(50,\ 250 - 100 \times \Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^\ell)\right)$ GeV,
E_T^{miss} bins [GeV]	250–350, 350–450, 450–550, 550–750, > 750

SUS - M_{lb}



- Invariant mass of lepton+b quark in a tt system is bounded

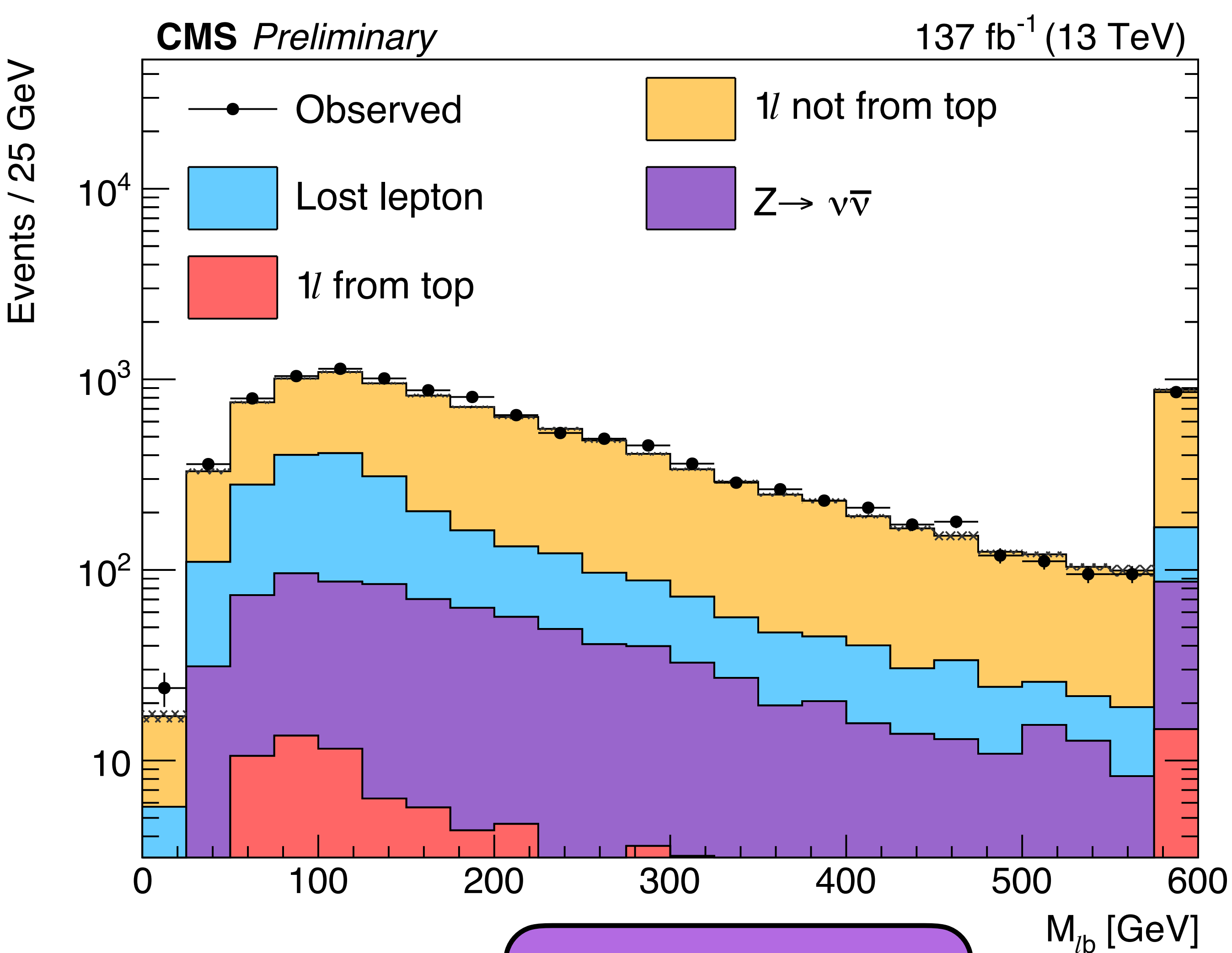
$$M_{\ell b} \leq M_t \sqrt{1 - \frac{M_W^2}{M_t^2}}$$

- Mass constraint applies to signal as well, in T2tt channel

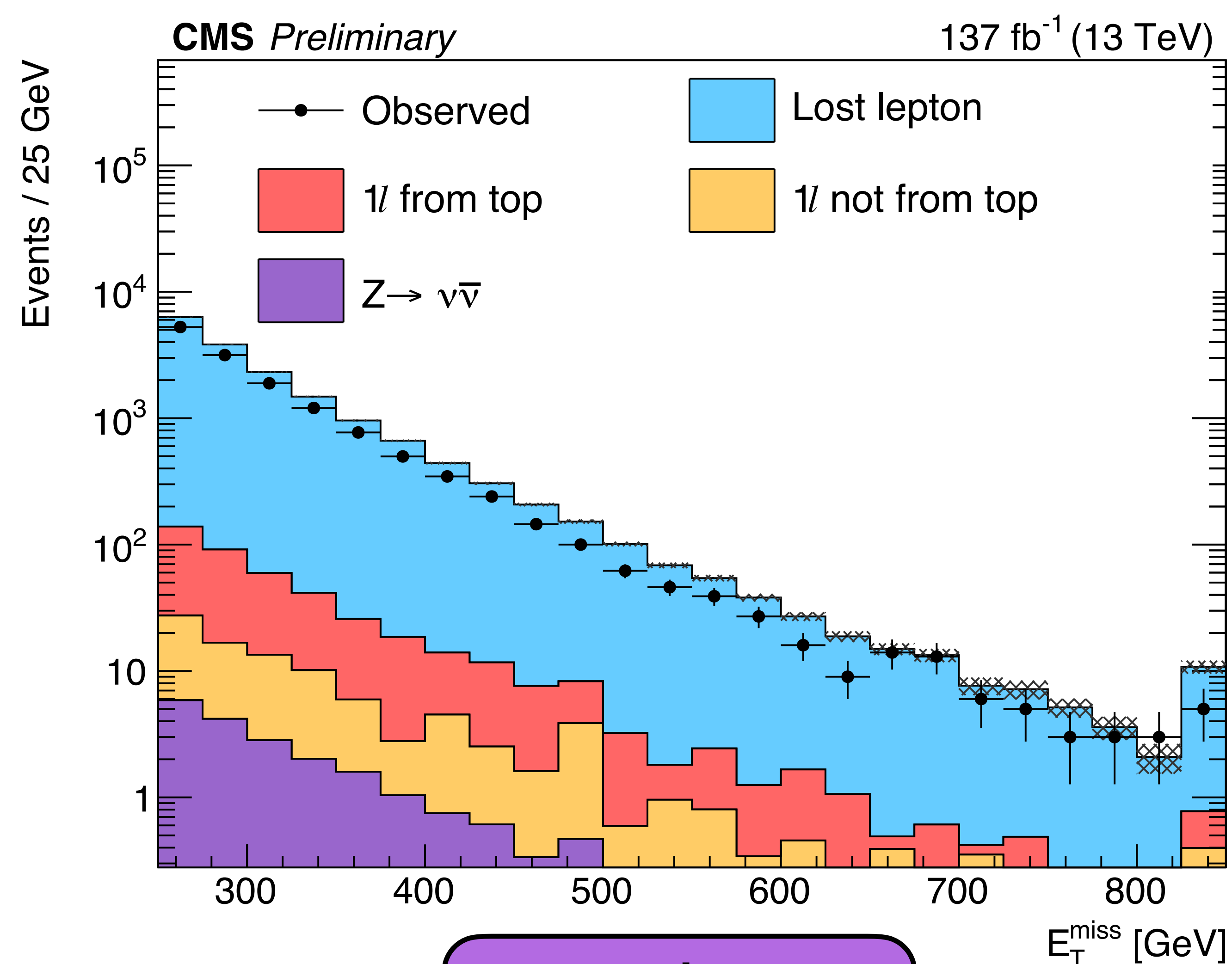
- Violated by backgrounds from other processes, like W+jets, and other signal channels

SUS - Control Regions

- Control region samples used for the background estimation
- MET distribution used for the dilepton transfer factor
  - Require presence of a second isolated lepton with p_T > 10 GeV
  - Add trailing lepton p_T to the MET
- W+jets estimated from the 0b control region
  - The "b" jet used for M_{lb} distribution is jet with highest deepCSV value



2 Leptons

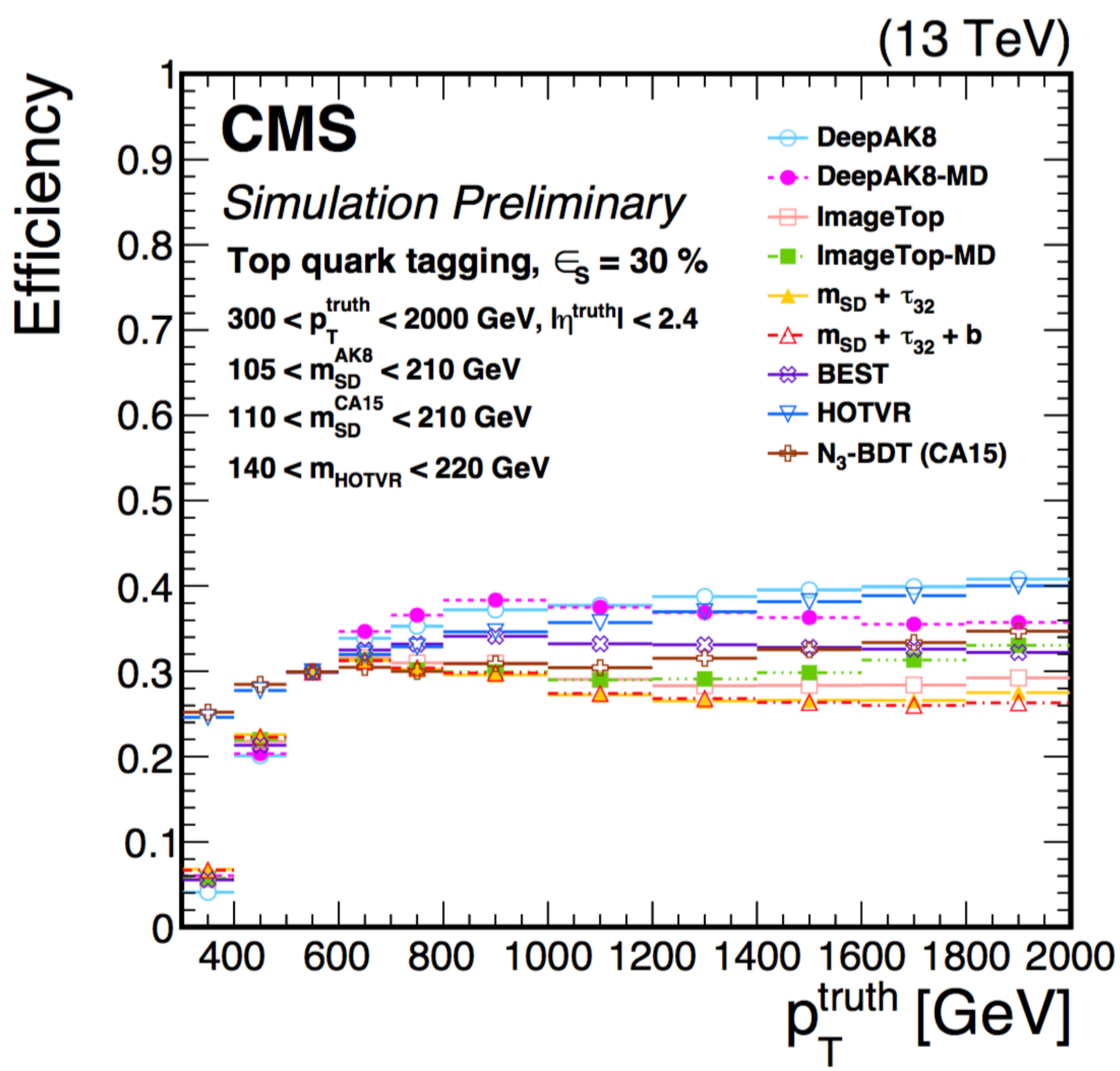


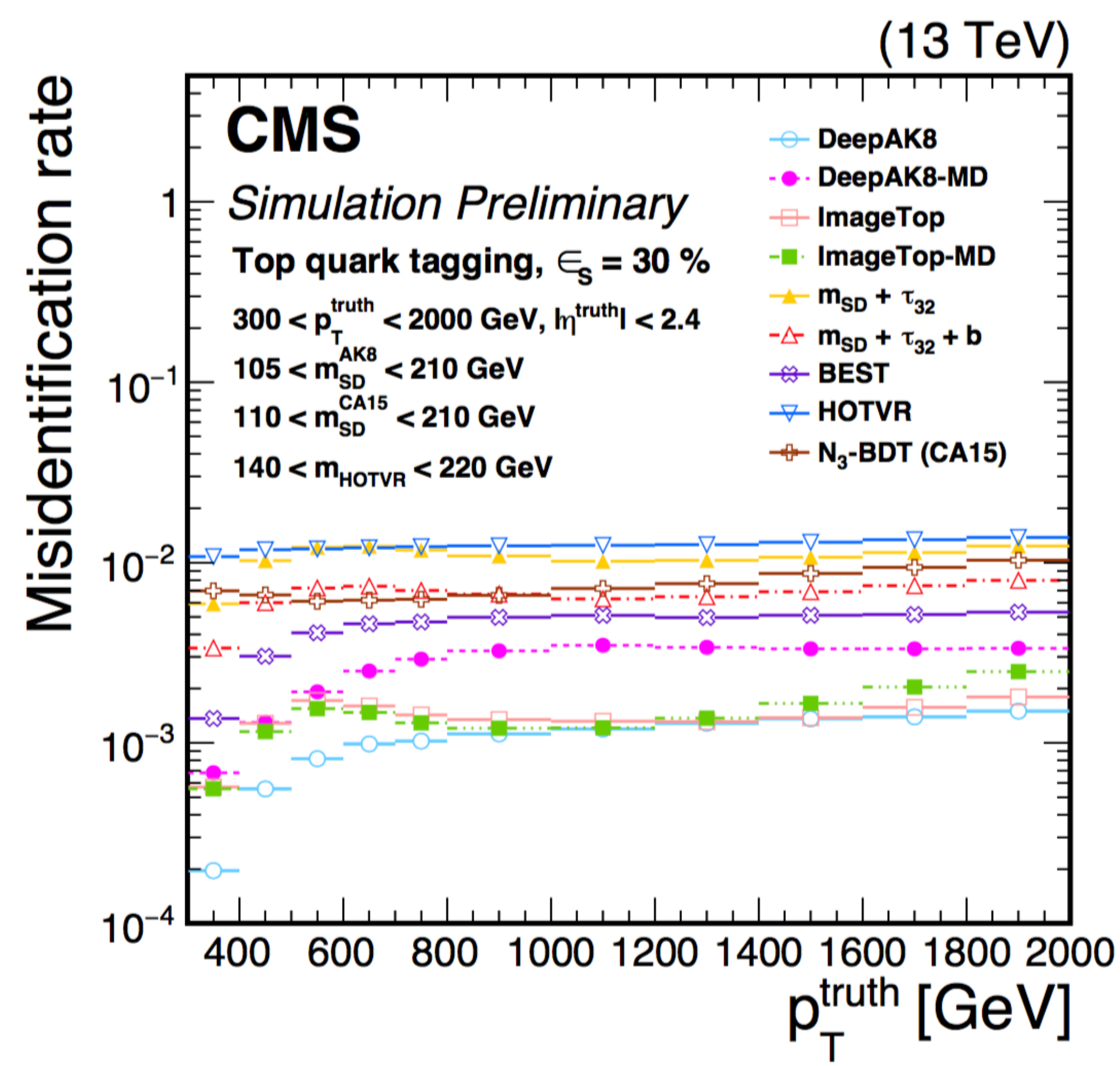
0b

SUS - Systematics

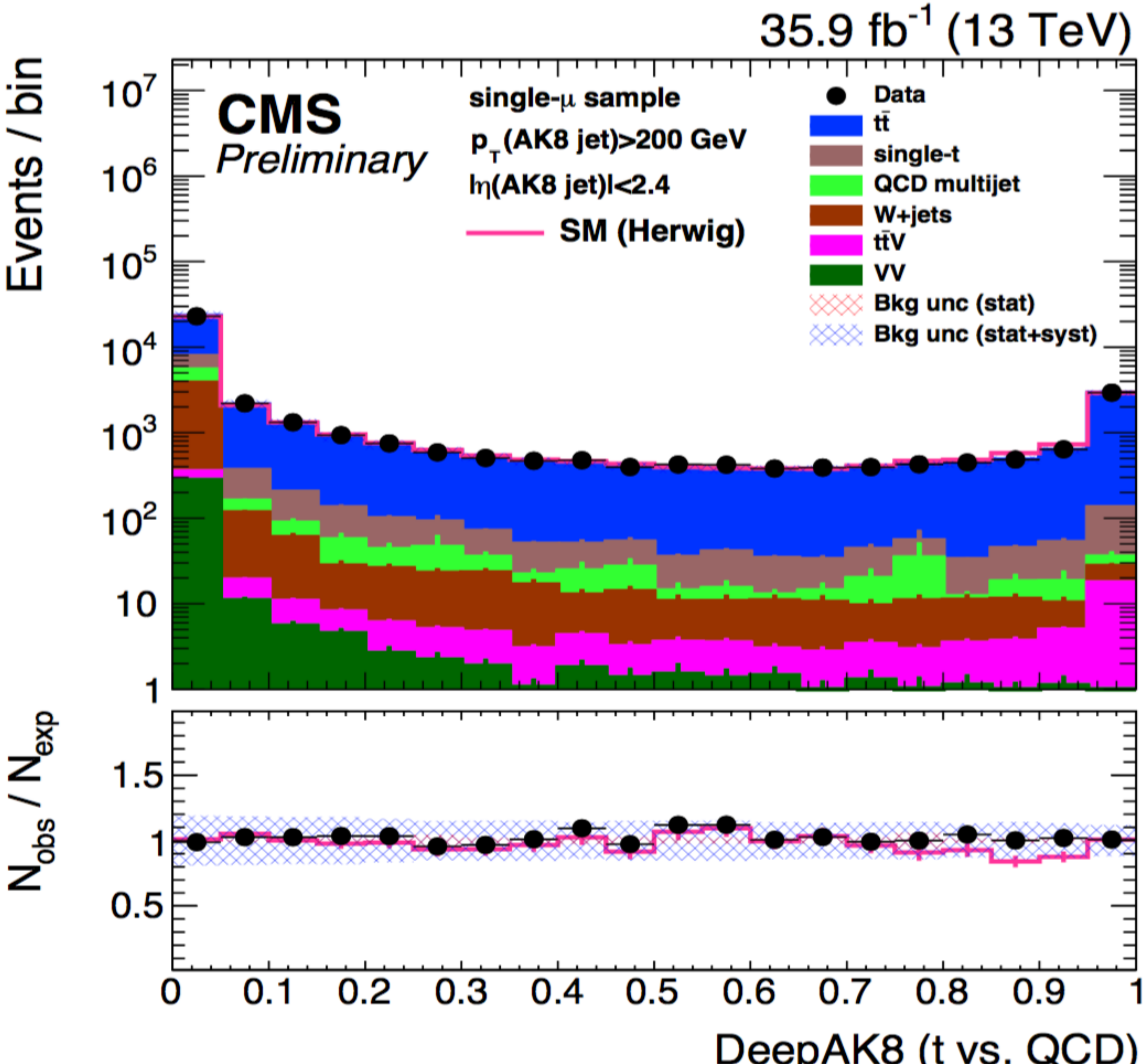
Source	Signal	Lost lepton	1ℓ background	$Z \to \nu\bar{\nu}$
Data statistical uncertainty	—	5–50%	4–30%	—
Simulation statistical uncertainty	6–36%	3–68%	5–70%	4–41%
$t\bar{t}$ E_T^{miss} modeling	—	3–50%	—	—
QCD scales	1–5%	0–3%	2–5%	1–40%
Parton distribution	—	0–4%	1–8%	1–12%
Pileup	1–5%	1–8%	0–5%	0–7%
Luminosity	2.3–2.5%	—	—	2.3–2.5%
$W + b$ cross section	—	—	20–40%	—
$Z \to \nu\nu$ estimate	—	—	—	5–10%
System recoil (ISR)	1–13%	0–3%	—	—
Jet energy scale	2–24%	1–16%	1–34%	1–28%
E_T^{miss} resolution	—	1–10%	1–5%	—
Trigger	2–3%	1–3%	—	2–3%
Lepton efficiency	3–4%	2–12%	—	1–2%
Merged top tagging efficiency	3–6%	—	—	5–10%
Resolved top tagging efficiency	5–6%	—	—	3–5%
b tagging efficiency	0–2%	0–1%	1–7%	1–10%
Soft b tagging efficiency	2–3%	0–1%	0–1%	0–5%

DeepAK8+BEST Performance In Simulation

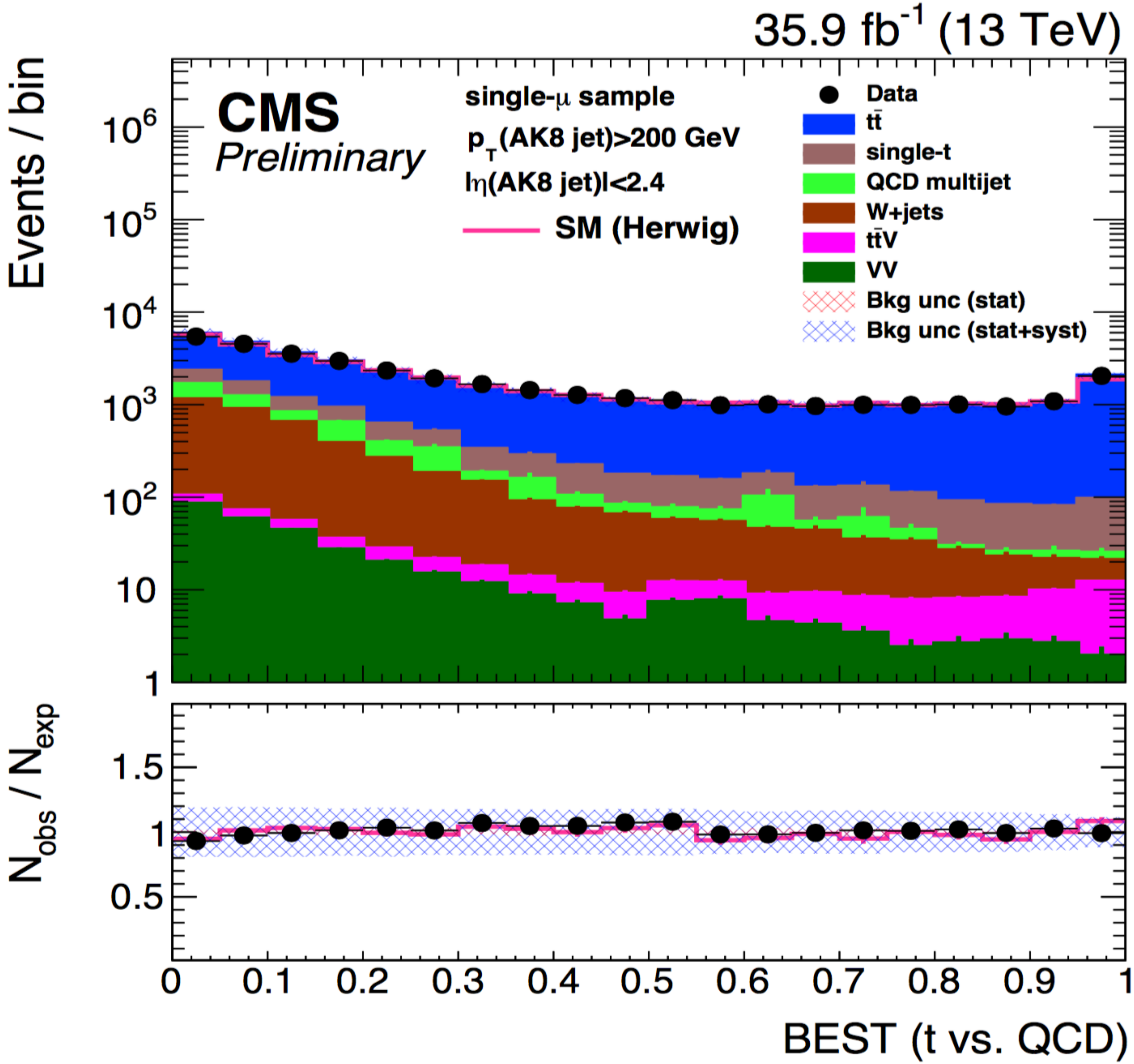




DeepAK8 Top Tagging Validation



BEST Top Tagging Validation



VLQ - Event Selection - Cut Based

- Use tagging multiplicities of W's and b's to define signal regions
- Suppress background by requiring multiple W/b tags and high hadronic activity:
  - H_T > 1200 GeV, where H_T is the scalar sum of all jet's transverse momenta
  - Signal requires > 0 of each tag.
- Signal-poor areas used for **validating** multijet QCD background estimation
- **Sideband** used to extract a normalization scale factor

2W0b	2W1b	2W2b
1W0b	1W1b	1W2b
0W0b	0W1b	0W2b

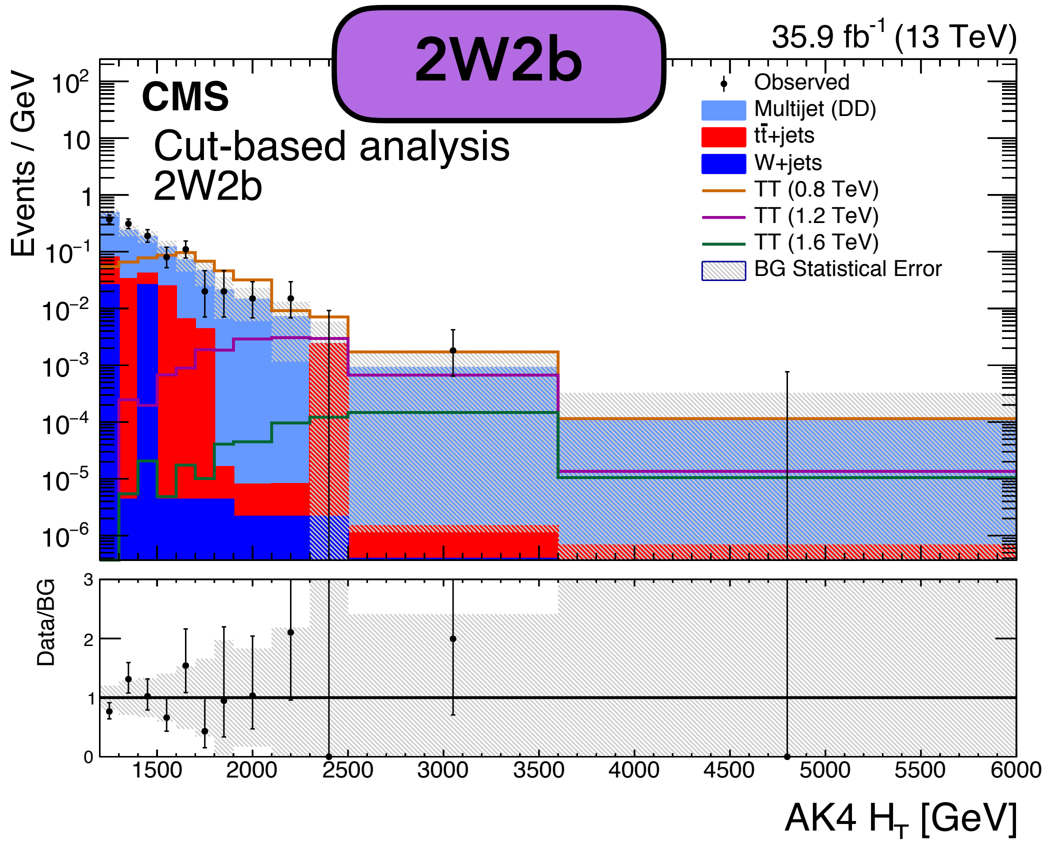
= 0 W tags, any b tags: Validation Regions

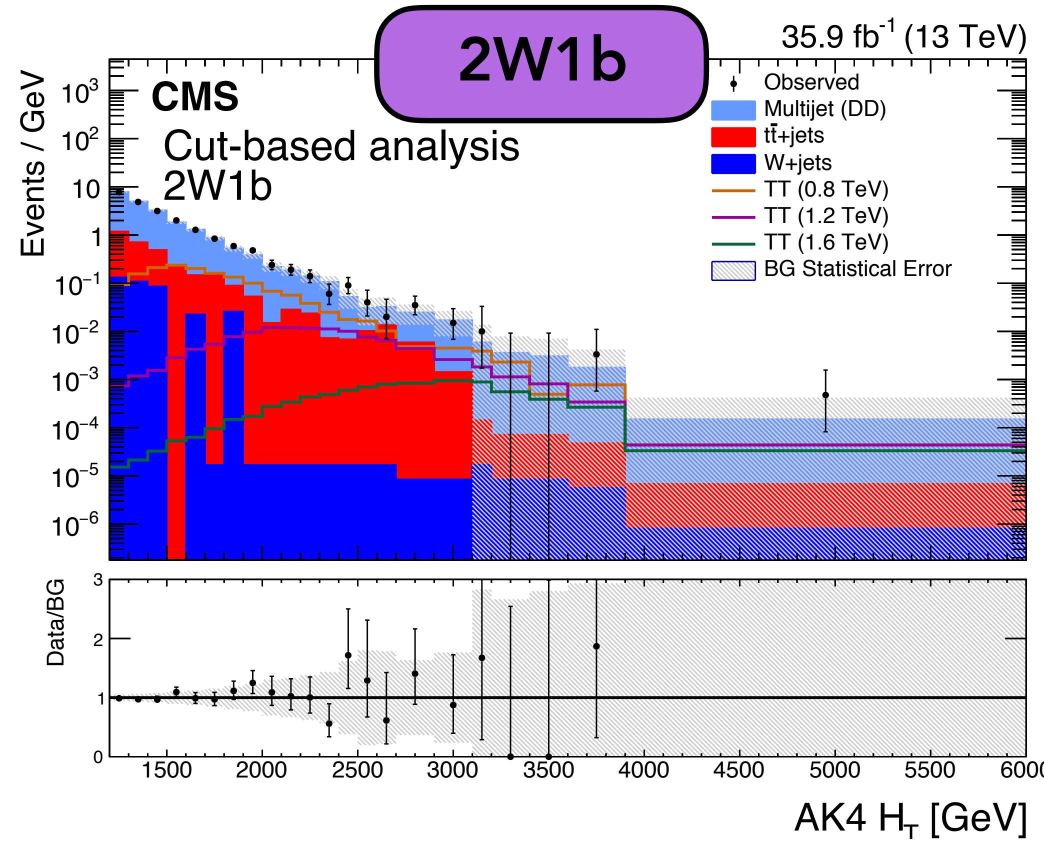
> 0 W tags, = 0 b tags: Sideband Regions

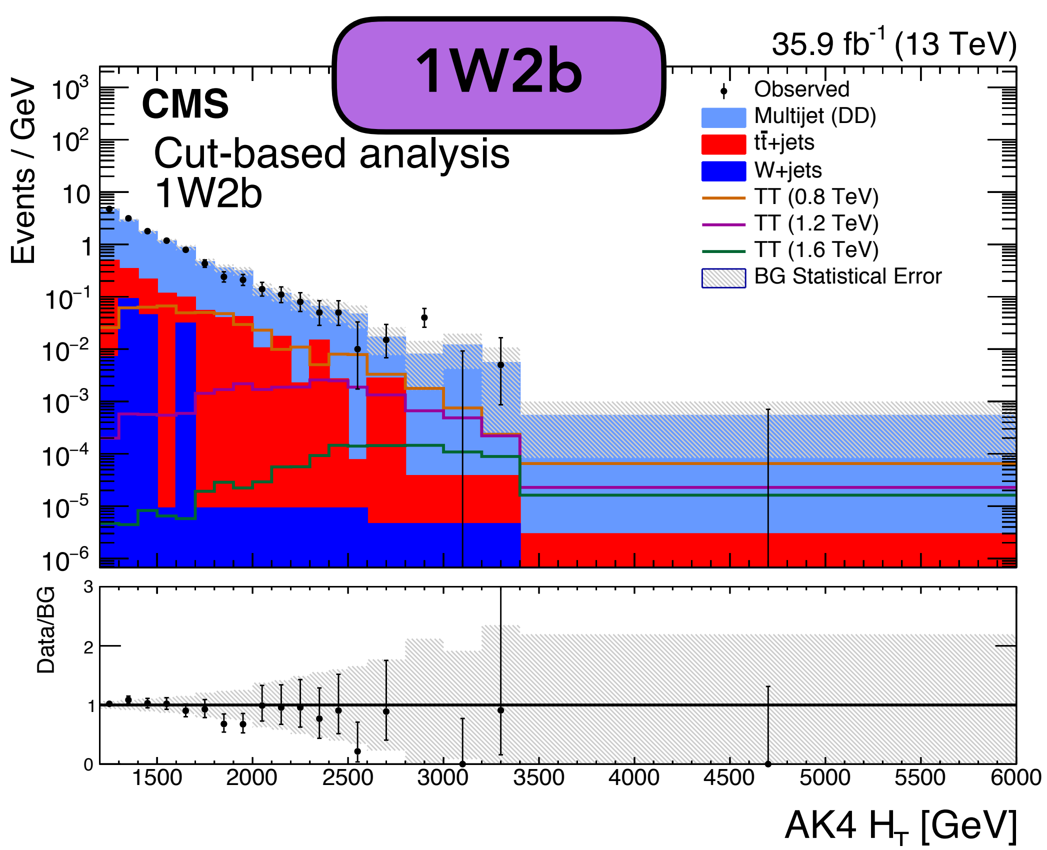
> 0 W tags, > 0 b tags: Signal Regions.

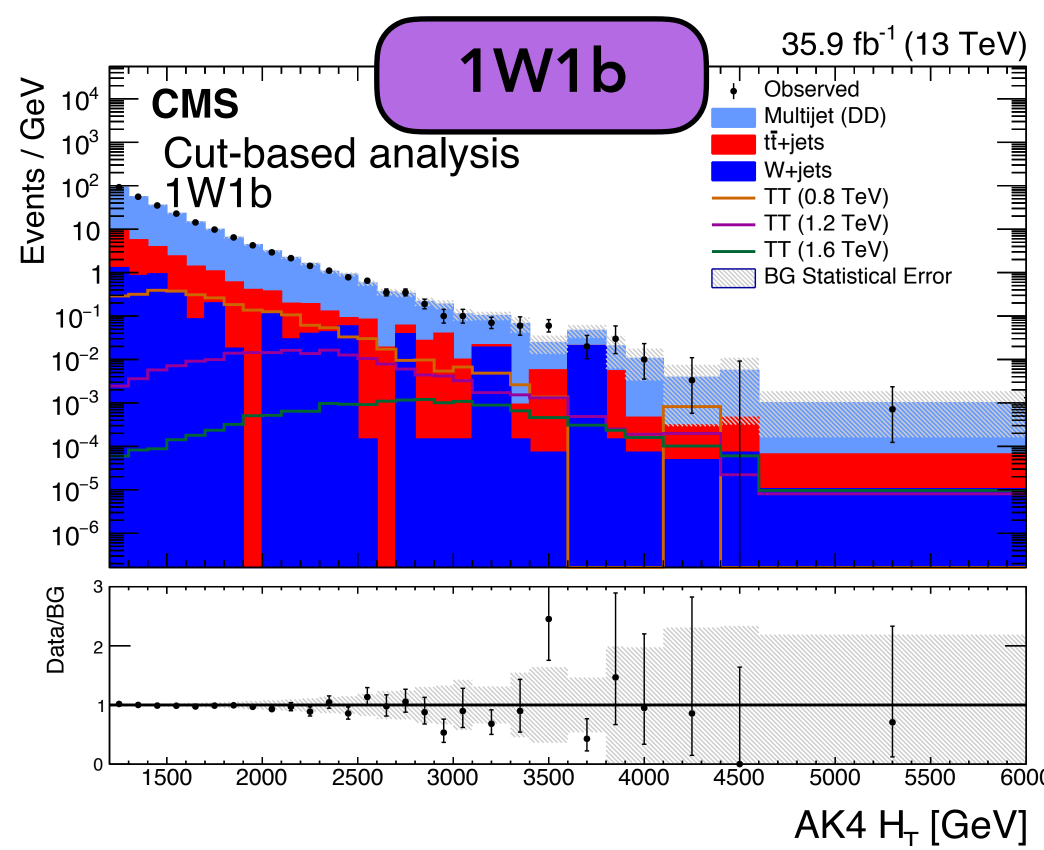
VLQ - Cut-Based Final Distributions

- Final signal regions for cut-based analysis (post-fit).



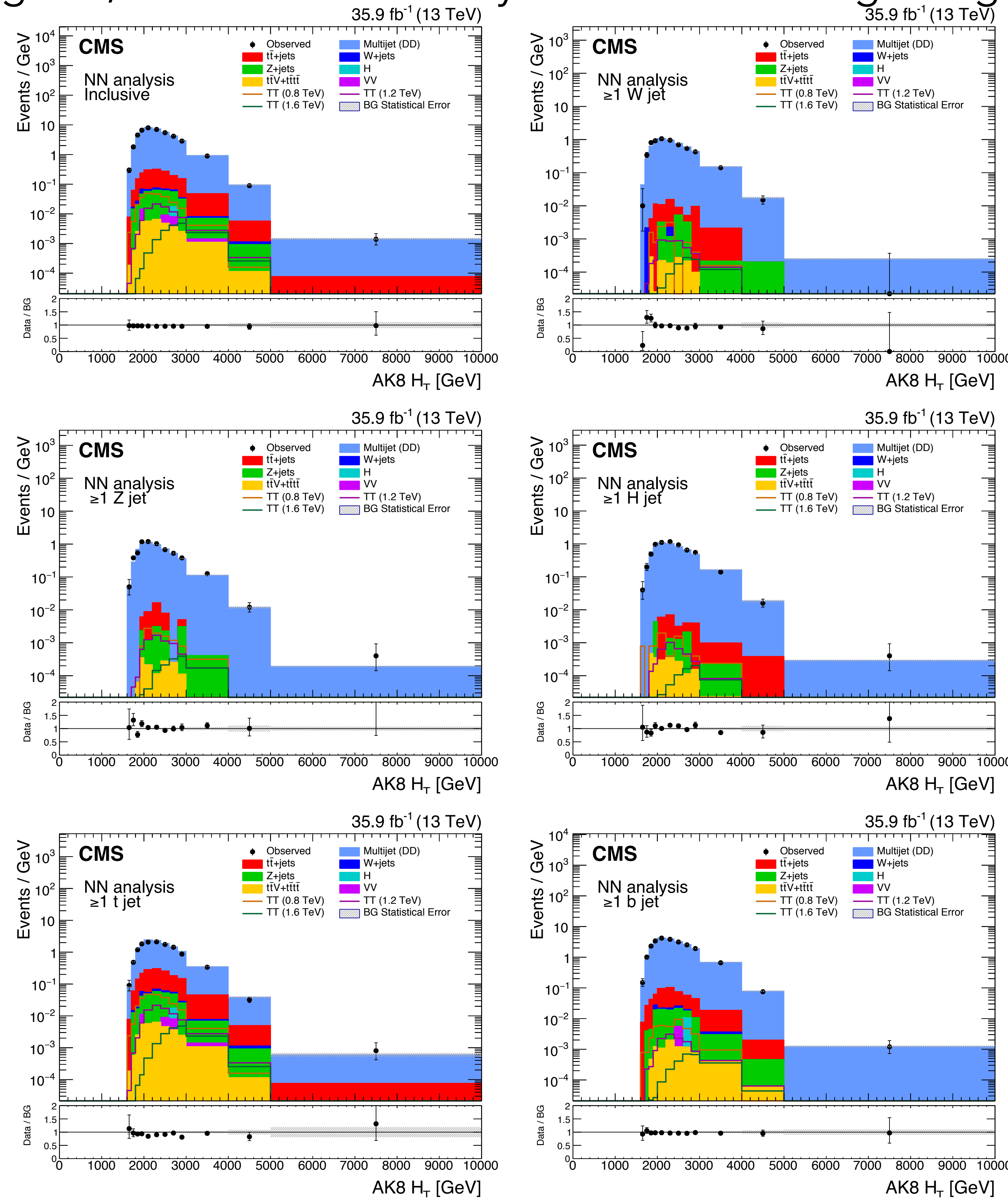






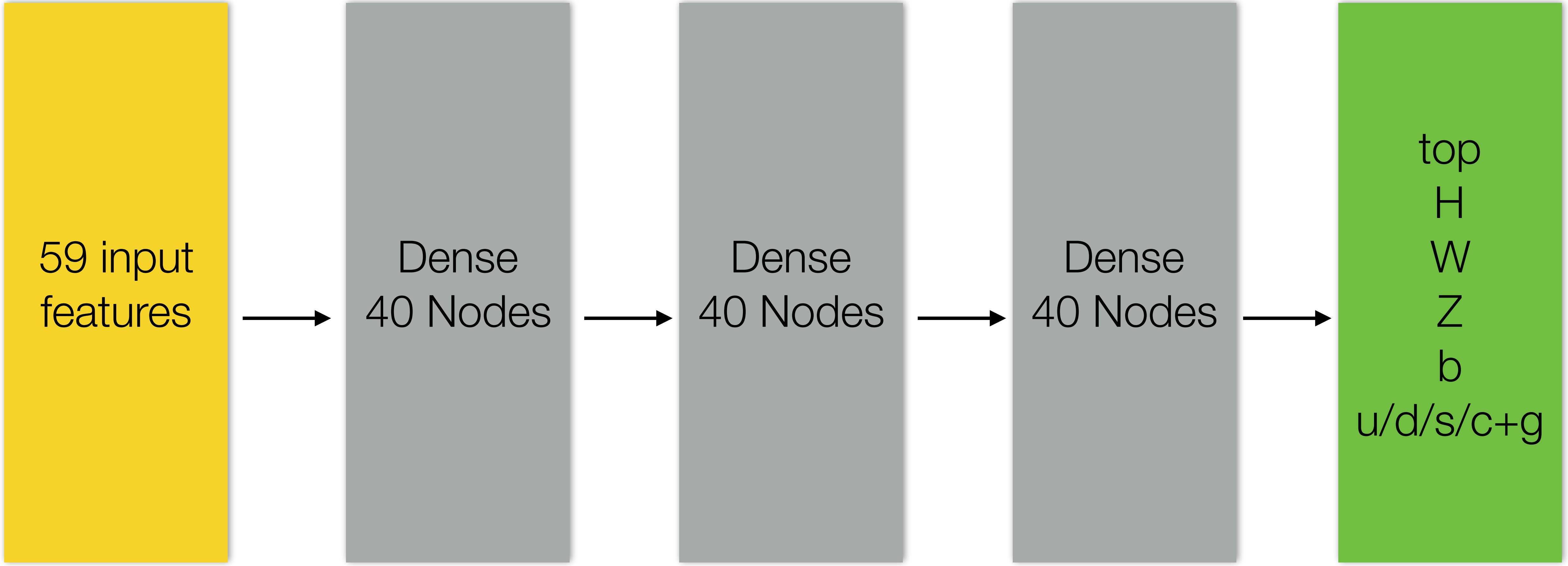
VLQ - BEST - Signal Region Summary

- 126 signal regions, summarized here by inclusive BEST tag categories.



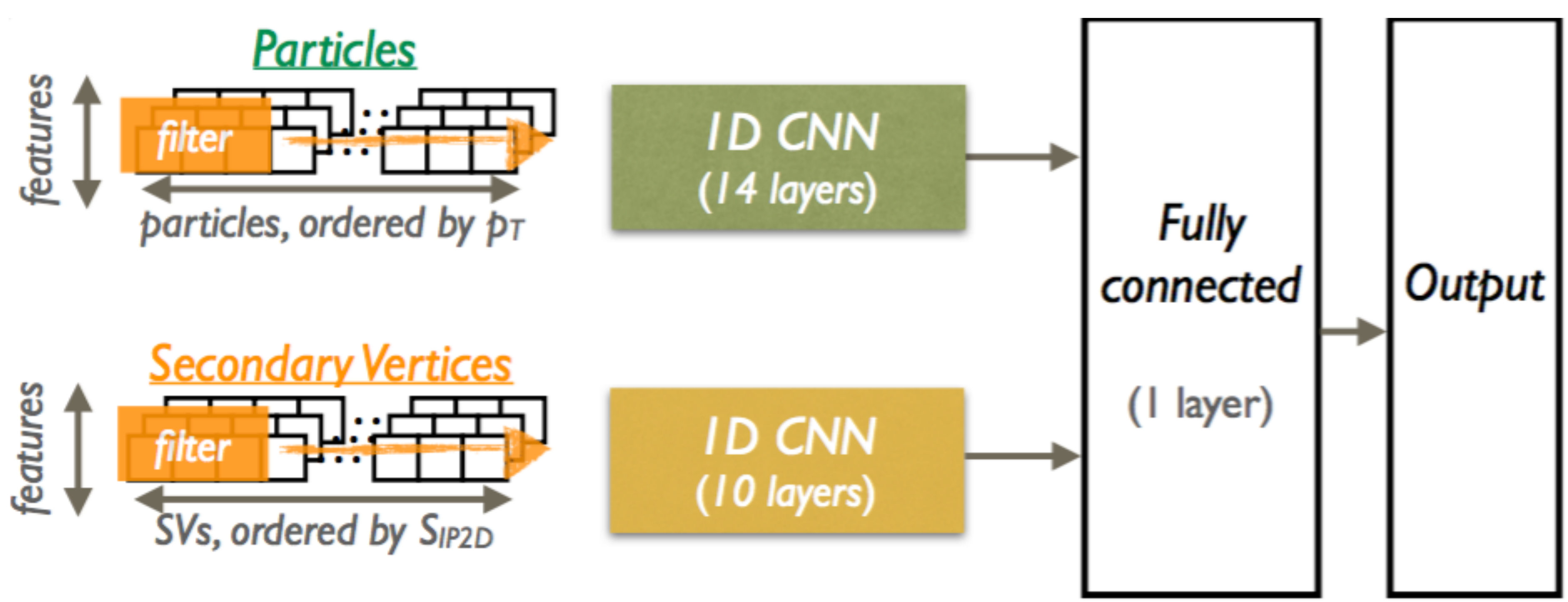
VLQ - BEST Architecture

- Feed-forward deep neural network with 3 hidden layers, 40 nodes each.
- Multi classification algorithm with six outputs.
- Node with highest output determines the jet label.
- 59 input features, 500,000 jets used to train the network.
- Use samples of high-pT boosted SM particles from high-mass resonance decays for training
  - Z'→tt, RSG->HH, etc.



DeepAK8 (Merged Tagger) Structure

- Trained on 40 million jets
- First 100 reconstructed particles in a jet used as inputs
  - 42 features per particle, such as basic kinematics, tracking info
- Up to 7 secondary vertices fed into parallel CNN
  - 15 features per SV, including displacement and quality
- Both CNNs are fed into one fully connected layer
- Discriminate between decay modes of heavy objects



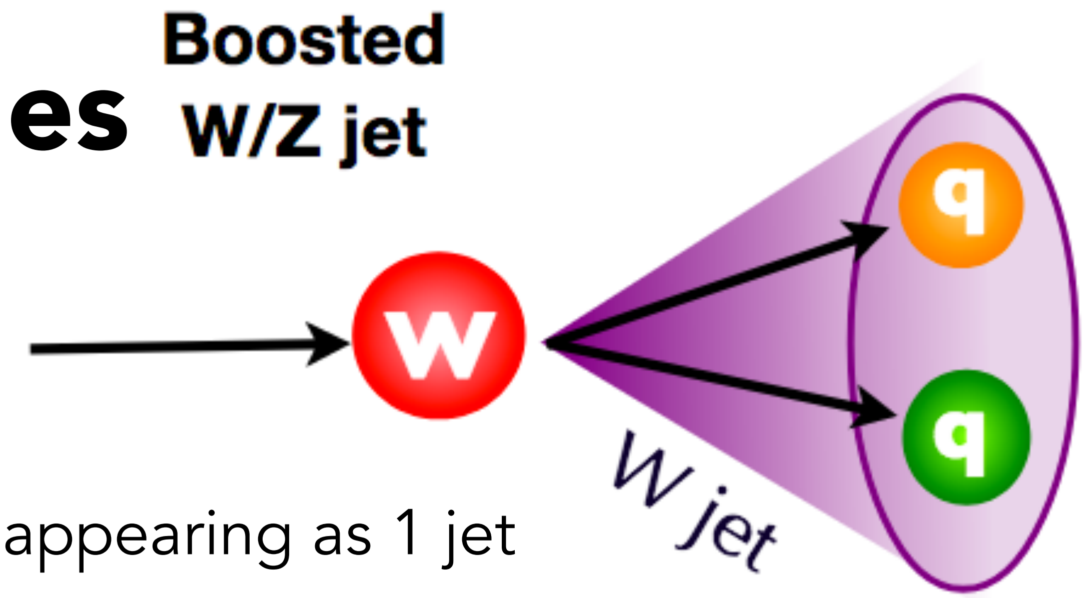
VLQ - BEST Variable List

NN Input Quantities	
Sphericity (t, W, Z, H)	Jet Soft-Drop Mass
Isotropy (t, W, Z, H)	Jet η
Aplanarity (t, W, Z, H)	Jet τ_{21}
Thrust (t, W, Z, H)	Jet τ_{32}
Jet Asymmetry A_L (t, W, Z, H)	Jet Charge
Fox-Wolfram H_1/H_0 (t, W, Z, H)	Maximum Subjet CSV Value
Fox-Wolfram H_2/H_0 (t, W, Z, H)	Subjet 1 CSV Value
Fox-Wolfram H_3/H_0 (t, W, Z, H)	Subjet 2 CSV Value
Fox-Wolfram H_4/H_0 (t, W, Z, H)	m_{13} (t,W,Z,H)
m_{12} (t,W,Z,H)	m_{1234} (t,W,Z,H)
m_{23} (t,W,Z,H)	

VLQ - Systematic Uncertainties

Uncertainty		Contribution to:		
Source	Uncertainty	Cut-based	NN	Applies to samples:
Diboson cross section	50%		✓	VV only
Rare top quark process cross sections	50%		✓	$t\bar{t}V$, $t\bar{t}t\bar{t}$
Higgs boson cross section	50%		✓	H only
W+jets cross section	15%	✓	✓	W+jets only
Z+jets cross section	15%		✓	Z+jets only
Integrated luminosity measurement	2.5%	✓	✓	All simulation
Pileup reweighting	$\pm1\sigma$	✓	✓	All simulation
Jet energy scale	$\pm1\sigma(p_{\mathrm{T}}, \eta)$	✓	✓	All simulation
Jet energy resolution	$\pm1\sigma(\eta)$	✓	✓	All simulation
Parton distribution functions	$\pm1\sigma$	✓	✓	$t\bar{t}$, VLQ
Renormalization and factorization scales	$\pm1\sigma$	✓	✓	$t\bar{t}$, VLQ
CSVv2 discriminant reshaping	δ(wgt., unwgt.)		✓	All simulation
BEST classification fractions	$\pm1\sigma(p_{\mathrm{T}})$		✓	QCD multijet
BEST classification scale factor	5%		✓	All simulation
BEST misclassification scale factor	5%		✓	All simulation
Trigger	2%	✓		All simulation
W tag scale factor	$\pm1\sigma$	✓		All simulation
Soft drop jet mass scale	$\pm1\sigma$	✓		All simulation
Soft drop jet mass resolution	$\pm1\sigma$	✓		All simulation
b tag scale factor	$\pm1\sigma$	✓		All simulation
Extrapolation fit	$\pm1\sigma$	✓		Background from data
Normalization of 1Wbackground prediction	1.9%	✓		Background from data
Normalization of 2Wbackground prediction	1.1%	✓		Background from data

VLQ - Substructure Variables

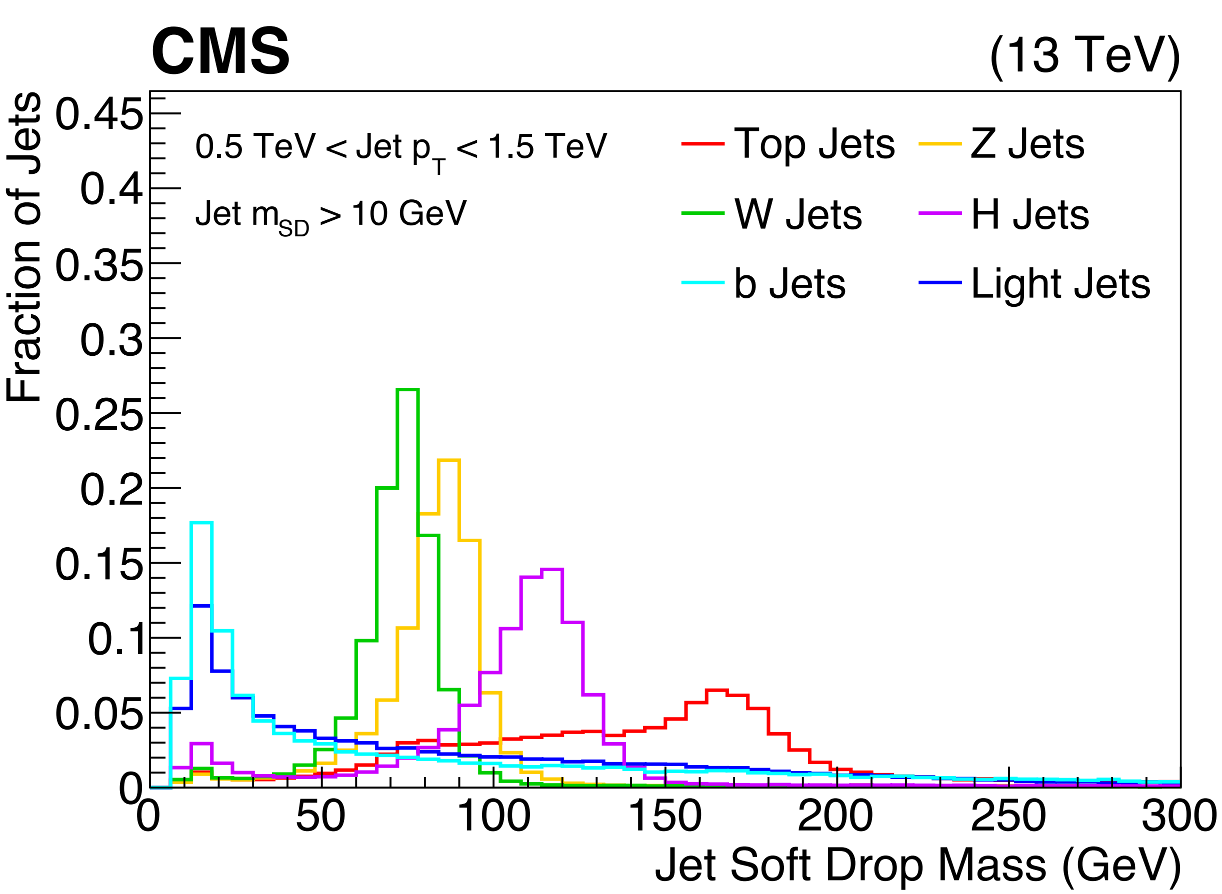


Tagging Boosted Jets

Heavy object (e.g. W boson) decays to two quarks

If W is sufficiently boosted, resulting jets will overlap, appearing as 1 jet

Can identify decays of heavy objects from typical light jets by mass and composite structure



"Soft Drop" Mass

Remove contribution of soft radiation to jet mass.

Resulting dist. is centered on original particle mass.

Cut-based: Require mass in the range [65, 105] GeV.

One of the most sensitive variables in the NN as well

N-subjettiness: "τ_N"

A measure of consistency with N subjets.

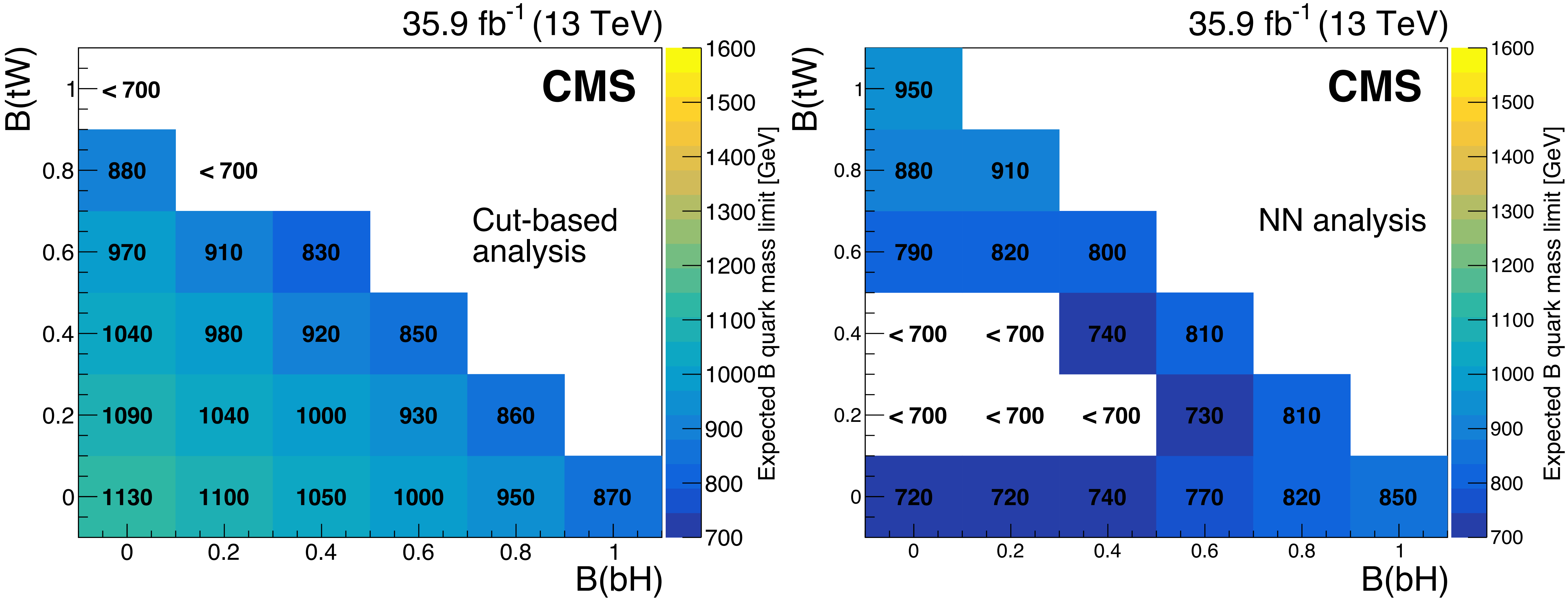
Cut on the ratio of τ_N to τ_{N-1}

Cut-based: For W jet tagging, $\tau_{21} < 0.55$

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min \{\Delta R_{1,k}, \Delta R_{2,k}, \cdots, \Delta R_{N,k}\}$$

VLQ Results — BB Limits Expected

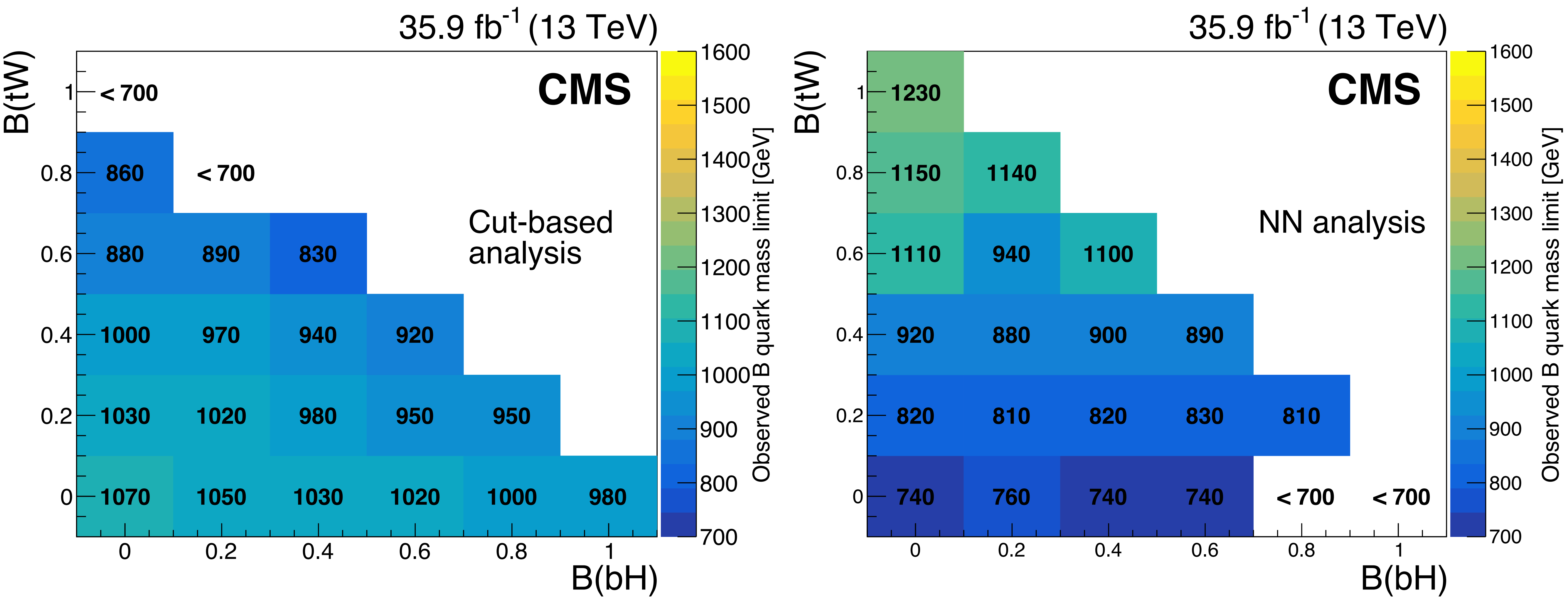
B2G-18-005



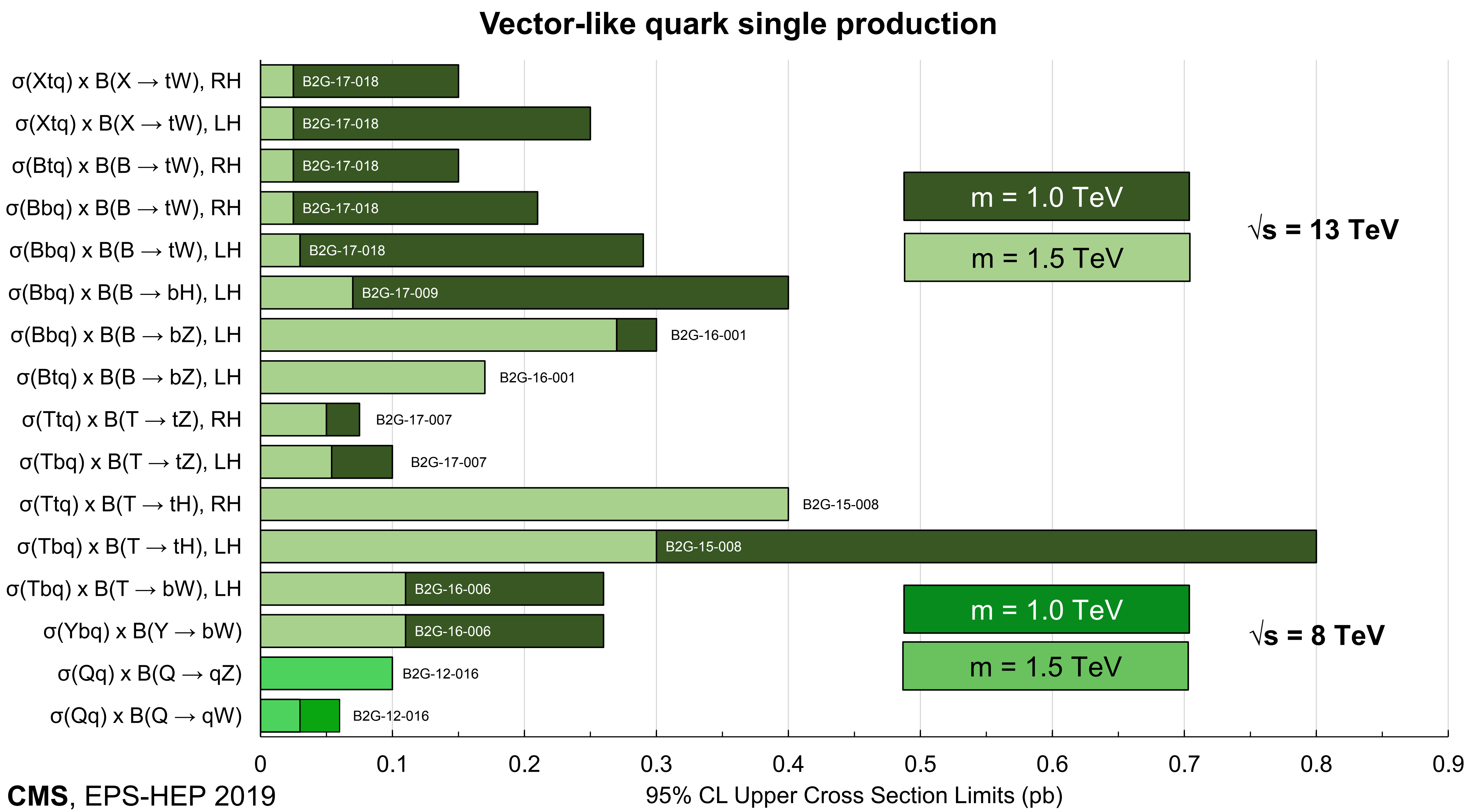
VLQ Results — BB Limits By Branching Fraction

B2G-18-005

- Scan limits over the branching fractions of the VLQ
  - BEST sensitive to B->tW, B->bZ and B->bH covered by other analysis



Summary Of Limits on Production of Single VLQs



CMS, EPS-HEP 2019