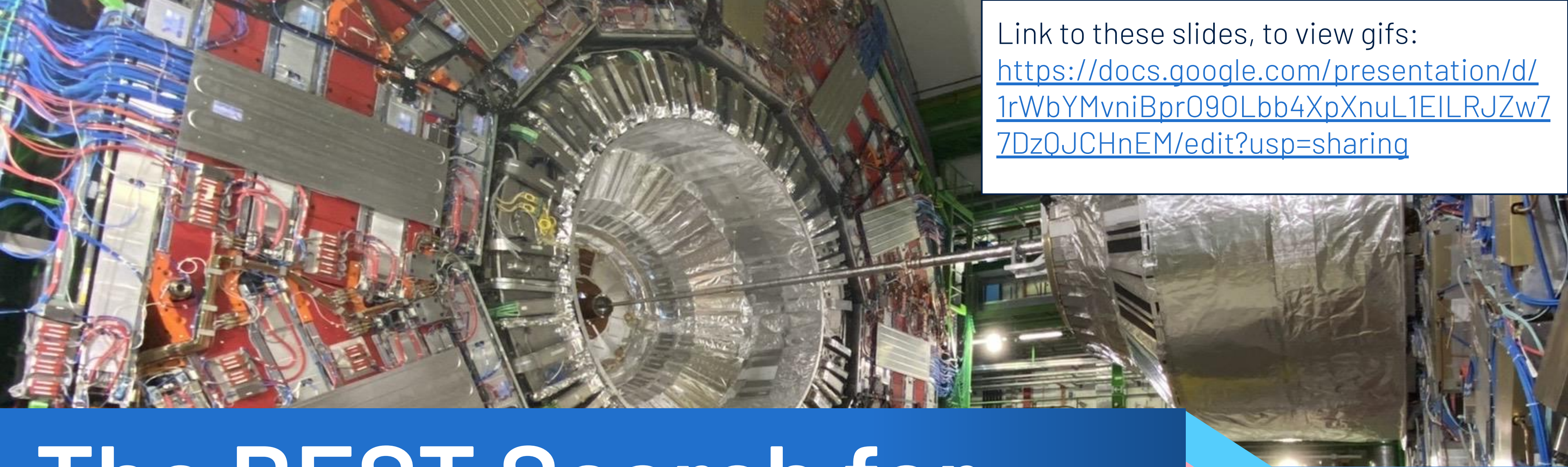


Link to these slides, to view gifs:
https://docs.google.com/presentation/d/1rWbYMvniBprO9OLbb4XpXnuL1EILRJZw77DzQJCHnEM/edit?usp=sharing

The BEST Search for Vector-Like Quarks

Samantha Abbott (She/Her)

UCDAVIS

US LHC Users Association Annual Meeting 2024
Stanford, CA

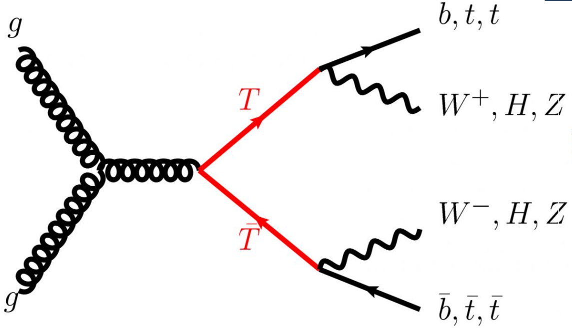
Dec 18, 2024

CMS

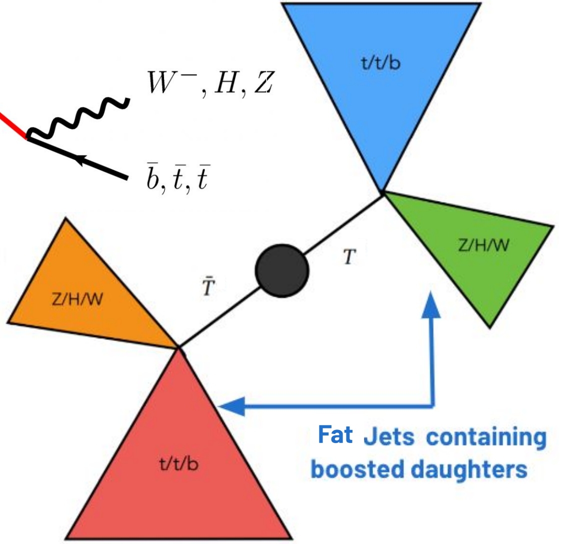
Quirky Quarks: Vector-Like Quarks (VLQs)

$T \to W^+b$	$B \to W^-t$
$T \to Zt$	$B \to Zb$
$T \to Ht$	$B \to Hb$

- Search for new physics: **Vector-Like Quarks (VLQs)**
  - Commonly predicted by SM extensions
    - e.g. Composite & Little Higgs models
  - Top-like (**T**) and bottom-like (**B**)
  - Non-chiral -> **vector-like**
  - **Very massive (>1 TeV)**

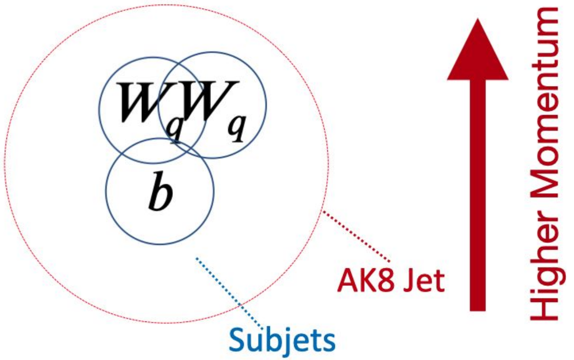


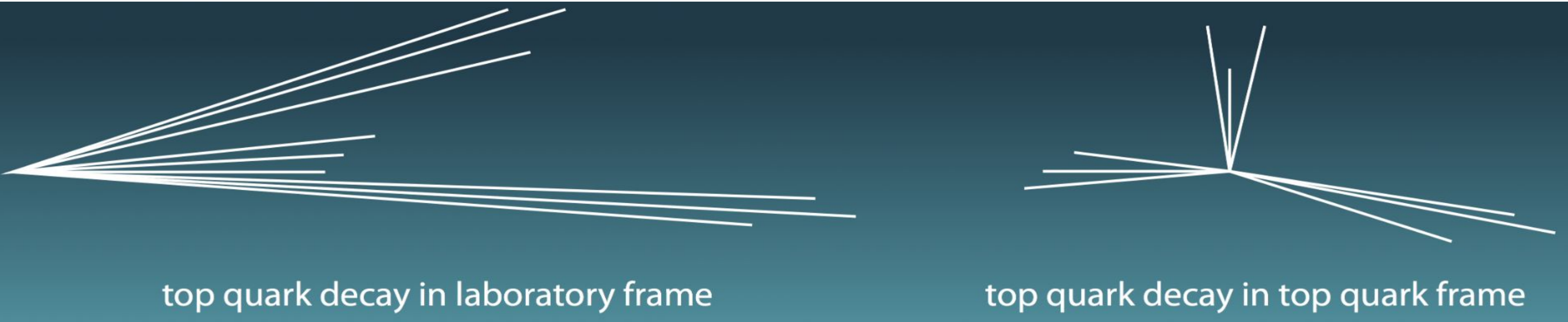
- We study **pair-produced VLQs**
  - Final State is **diverse** and **boosted**:
    - **Fully hadronic**
    - **4 high - p_T AK8 jets**
    - **W, Z, Higgs, *t*, *b***



Momentum Brings Us Closer

- **Heavy** Particles -> **Light** Decays -> **High Momentum** Decays
  - VLQ mass is on the **TeV** scale
  - W, Z, Higgs, *t*, *b* are on the **GeV** scale
- **AK8 Jets** contain several constituents
  - These '**daughters**' are
    - **boosted (high momentum)**
    - **highly collimated**

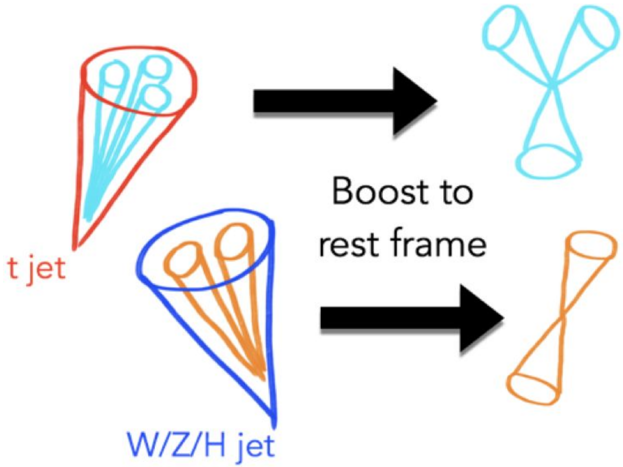




The BEST Method

- The **B**oosted **E**vent **S**hape **T**agger (**BEST**) steps:
  - **Boost** AK8 jet into different **rest frames**
  - Calculate **BES** variables in each frame
    - Include some **frame invariant** variables
  - Feed vars into simple Neural Net
  - Classify!
    - **Multi-Object** classifier: W, Z, Higgs, top, bottom, QCD
    - Generalized

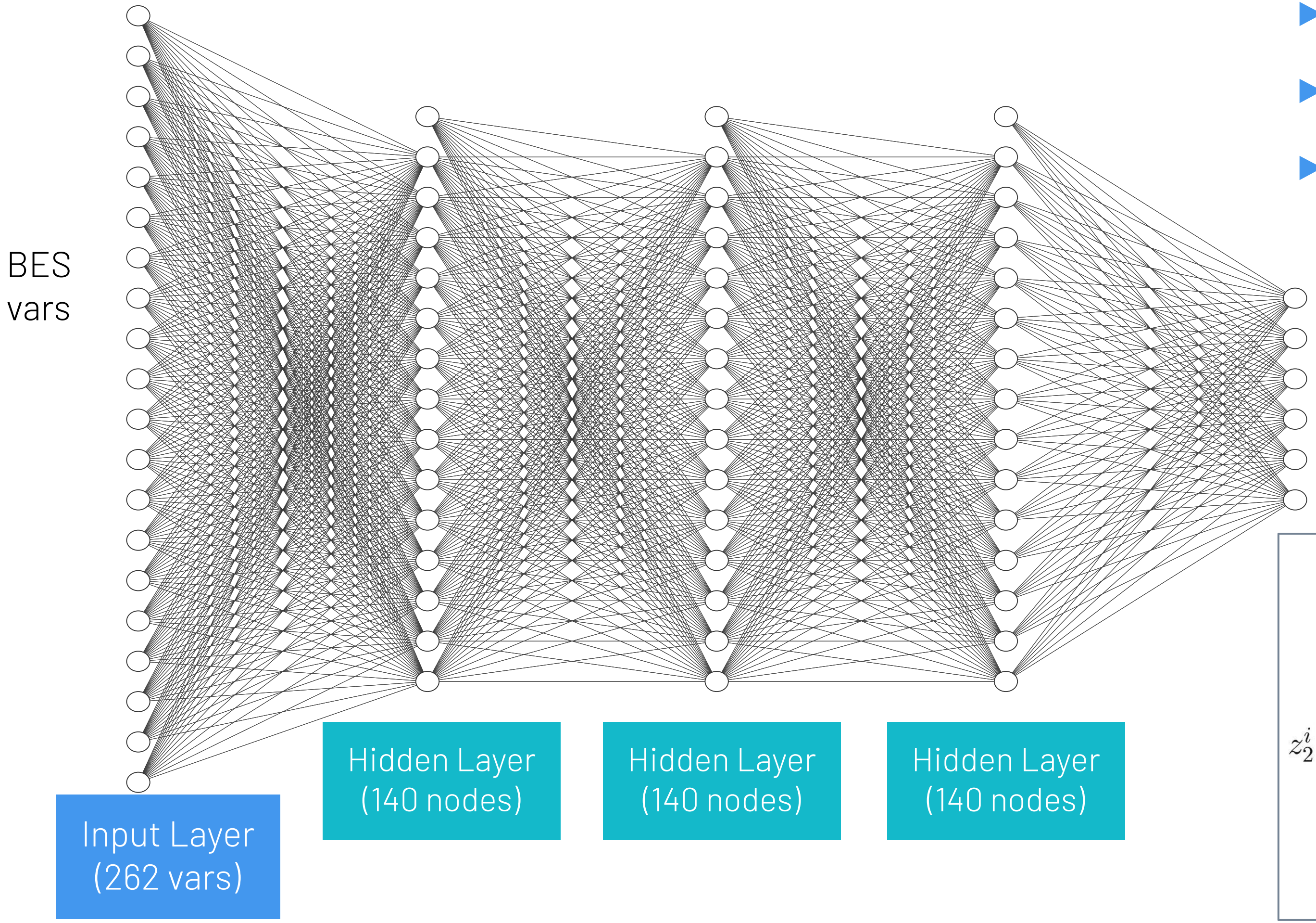
7 Boost Frames:
m_{jet}, m_{SD},
W, H, top,
300, 400 GeV

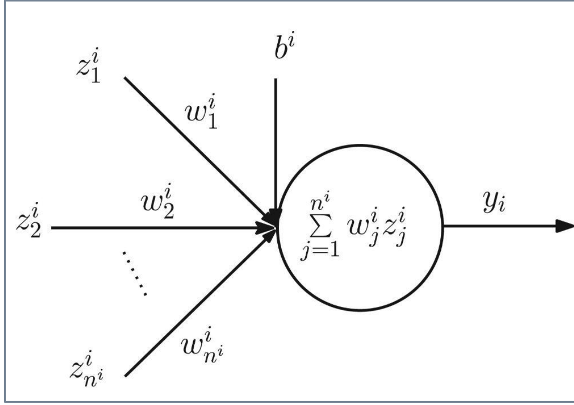


- **Perfect for our all-hadronic search for VLQs!**

Neural Network: The BEST Artificial Brain

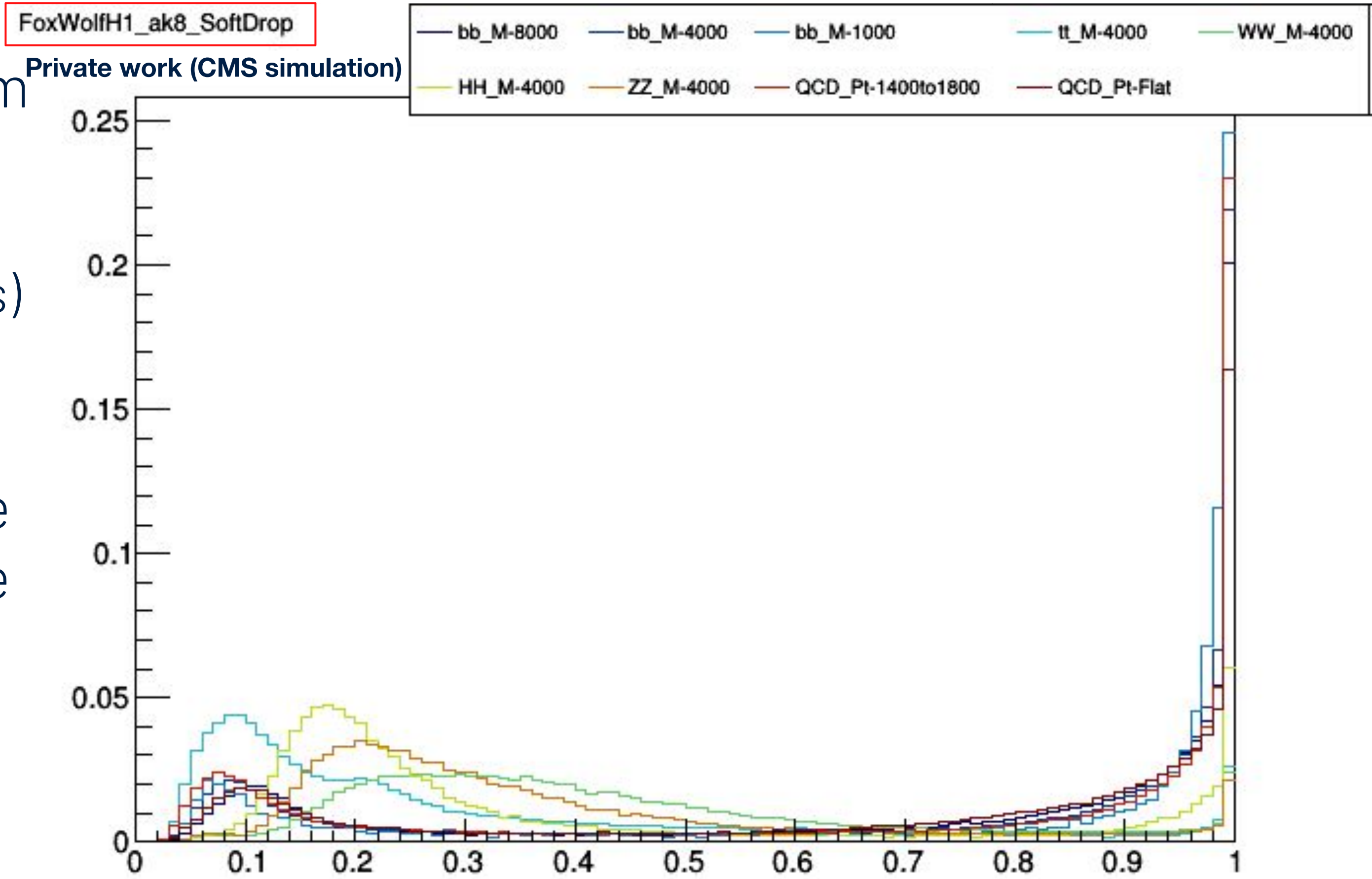
- Feedforward
- Fully Connected
- Dense Layers





Scanning Through Boosts

- Our NN learns from **non-linear transformations** (e.g. Lorentz Boosts)
  - *What Lorentz boosts are BEST?*
- Want BES variable distributions to be **distinguishable**
  - Naked eye can pick some out; **artificial eye is even better**

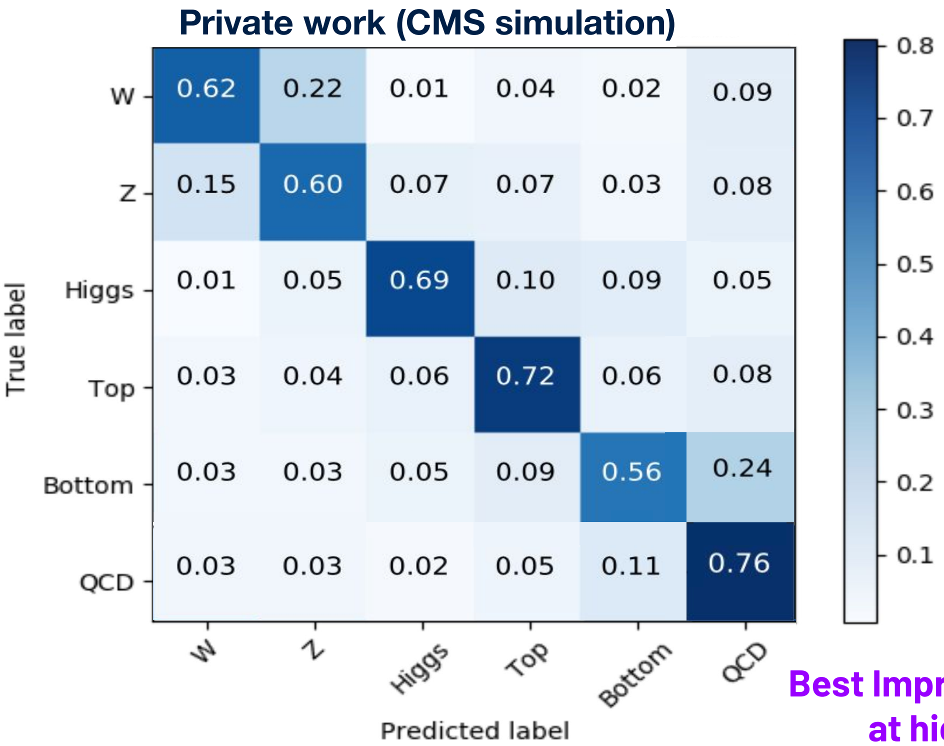


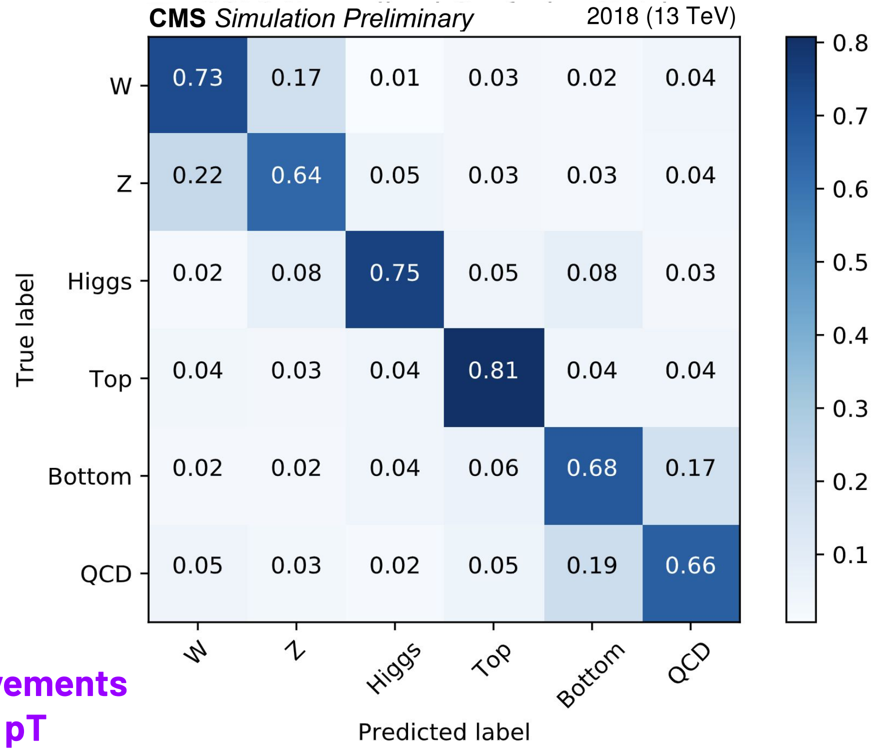
Improving on the Classics

Classic BEST (2016)

WINNER TAKE ALL

Current BEST

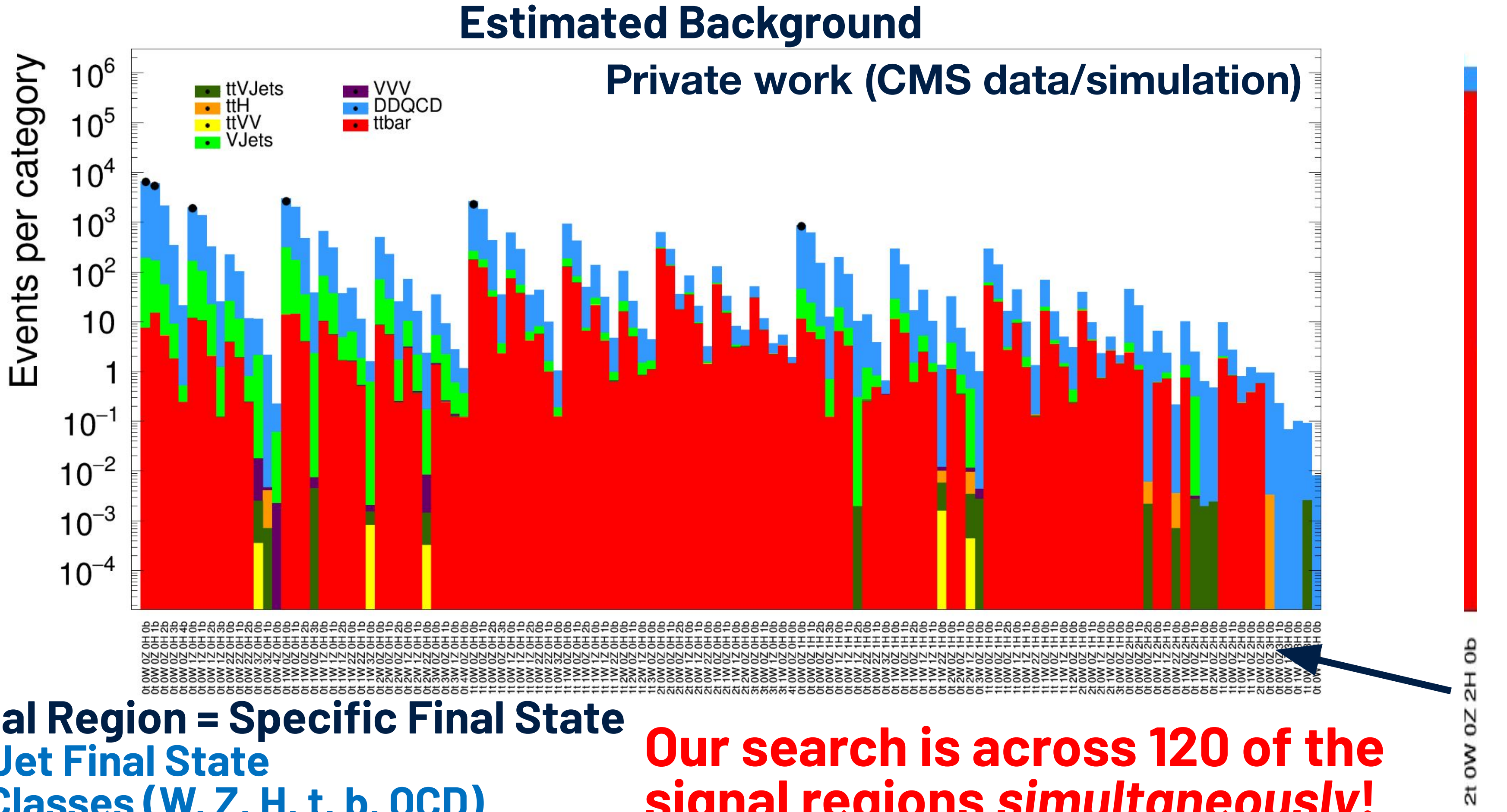




Best Improvements at high pT

Oops, All Signal Regions!

UCDAVIS



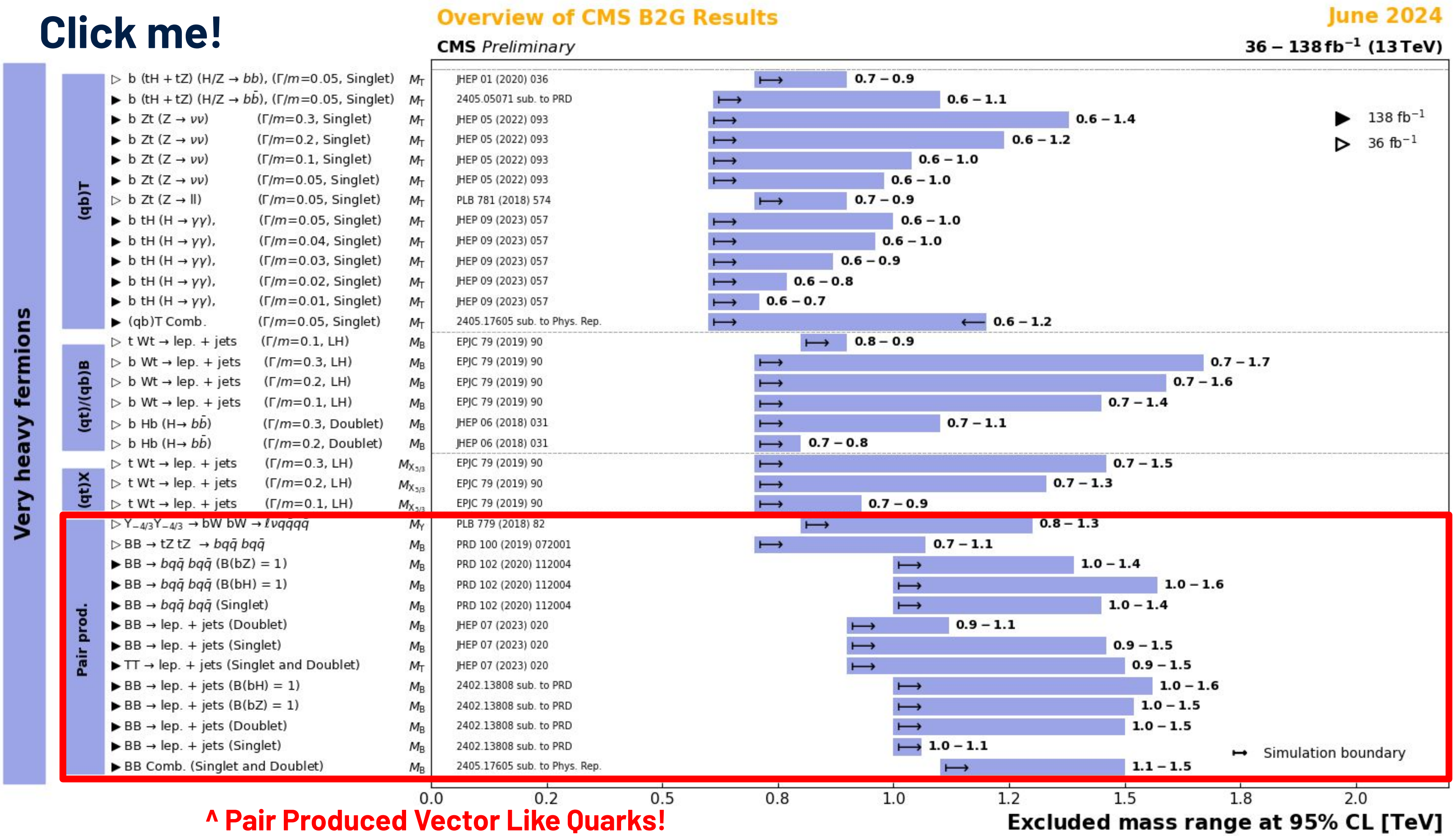
- **Signal Region = Specific Final State**
  - **4 Jet Final State**
  - **6 Classes (W, Z, H, t, b, QCD)**
    - **126 possible combinations!**

Our search is across 120 of the signal regions *simultaneously*!

Soon to be BEST

UC DAVIS

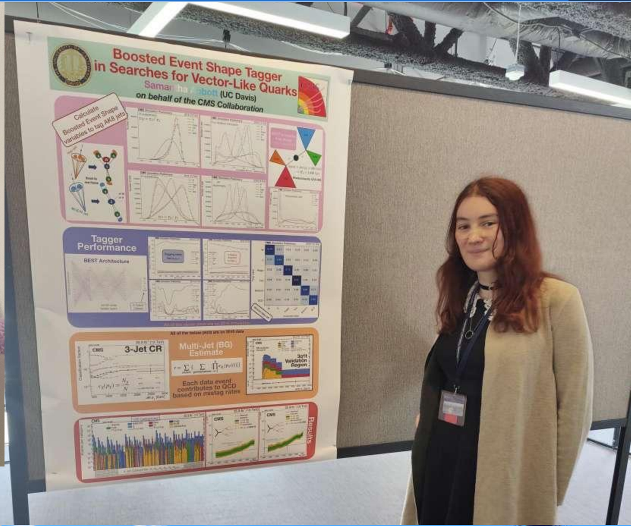
Click me!



^ Pair Produced Vector Like Quarks!

Thank You!





Boosted Event Shape Tagger
in Searches for Vector-Like Quarks
(UC Davis)
on behalf of the CMS Collaboration
Tagger Performance
3-Jet CR
Multi-Jet (BG) Estimate
Results

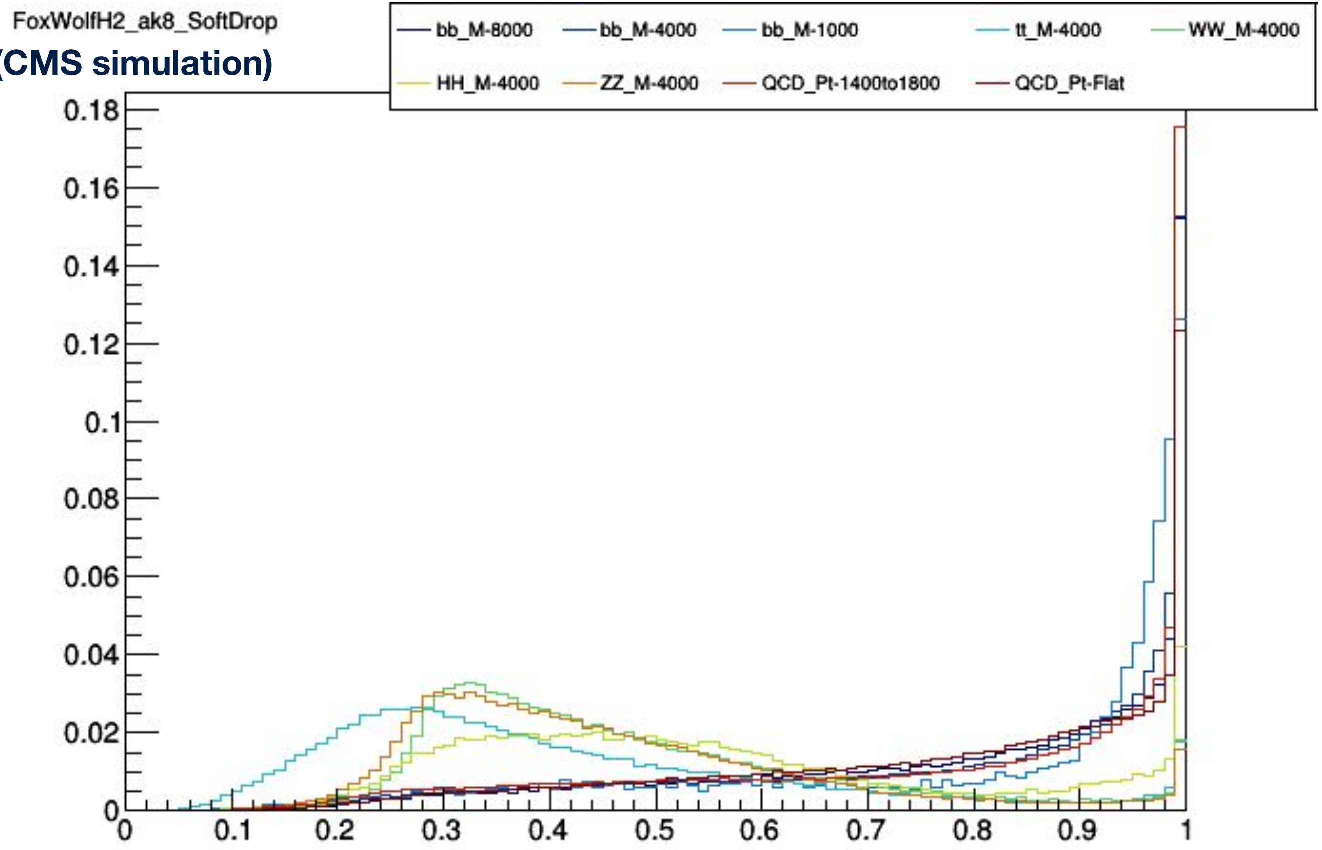

More Gifs

Link to these slides, to view gifs: https://docs.google.com/presentation/d/1rWbYMvniBpr090Lbb4XpXnuL1EILRJZw77DzQJCHnEM/edit?usp=sharing

Fox Wolfram Moments: H2

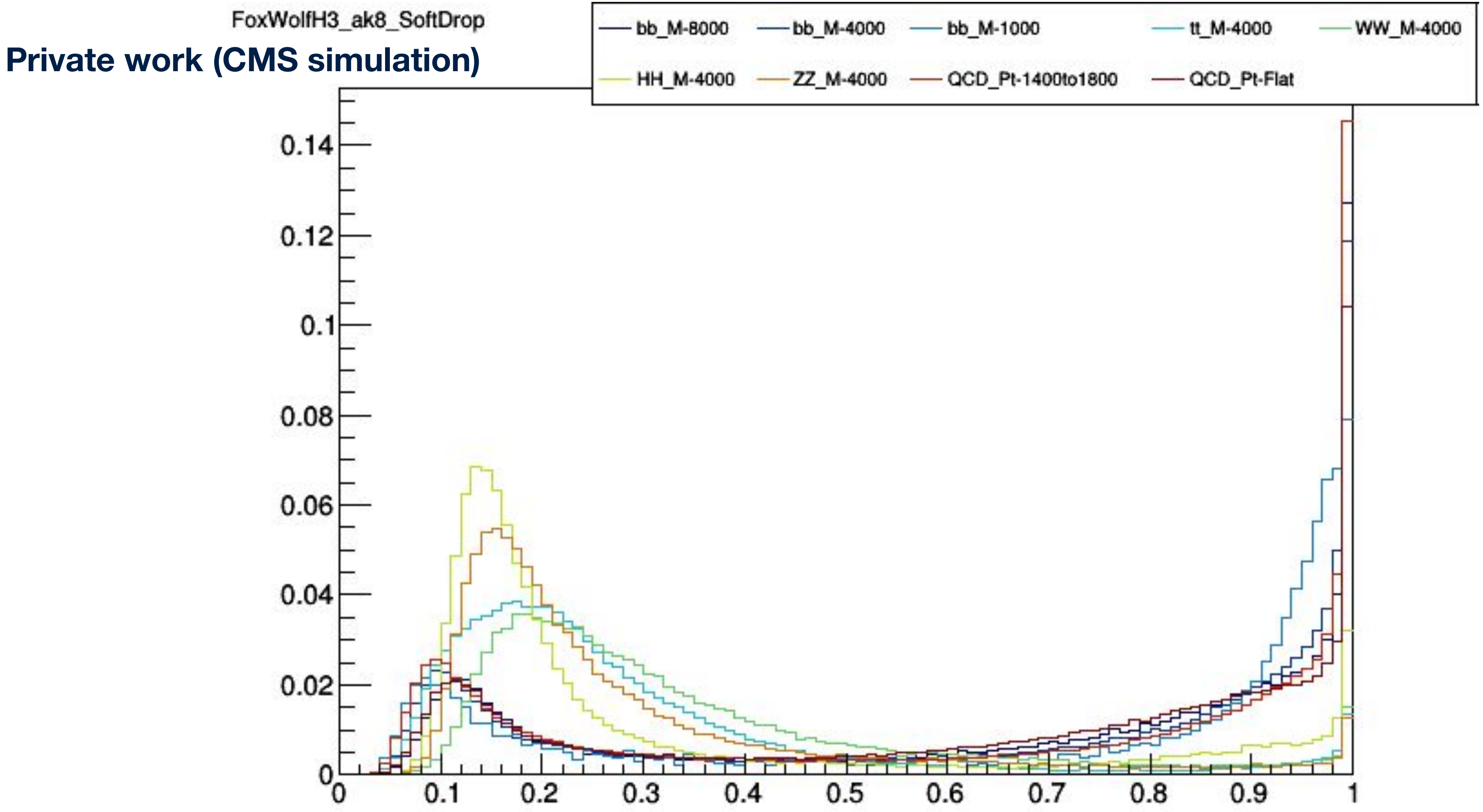
UCDAVIS

Private work (CMS simulation)

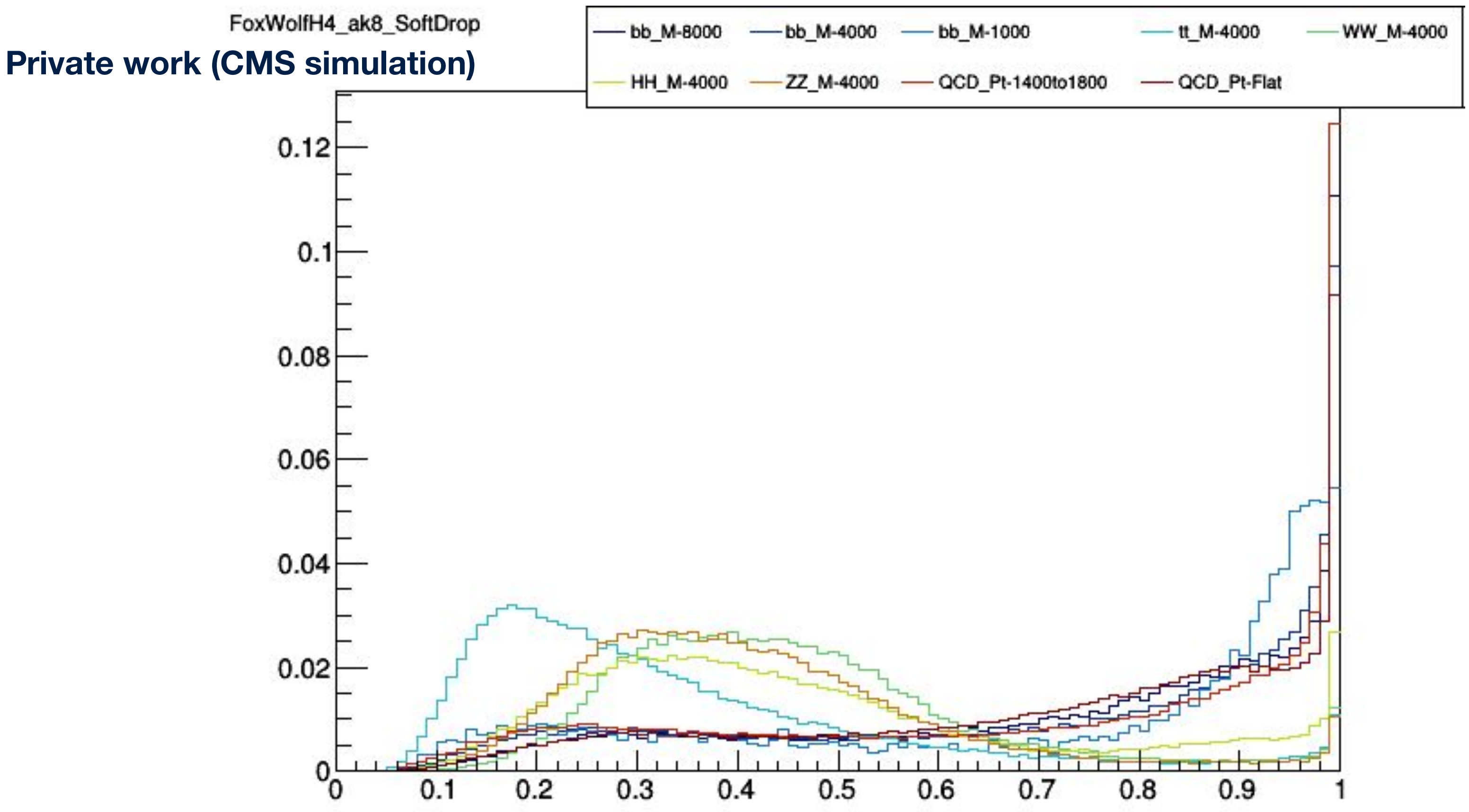


Fox Wolfram Moments: H3

UCDAVIS

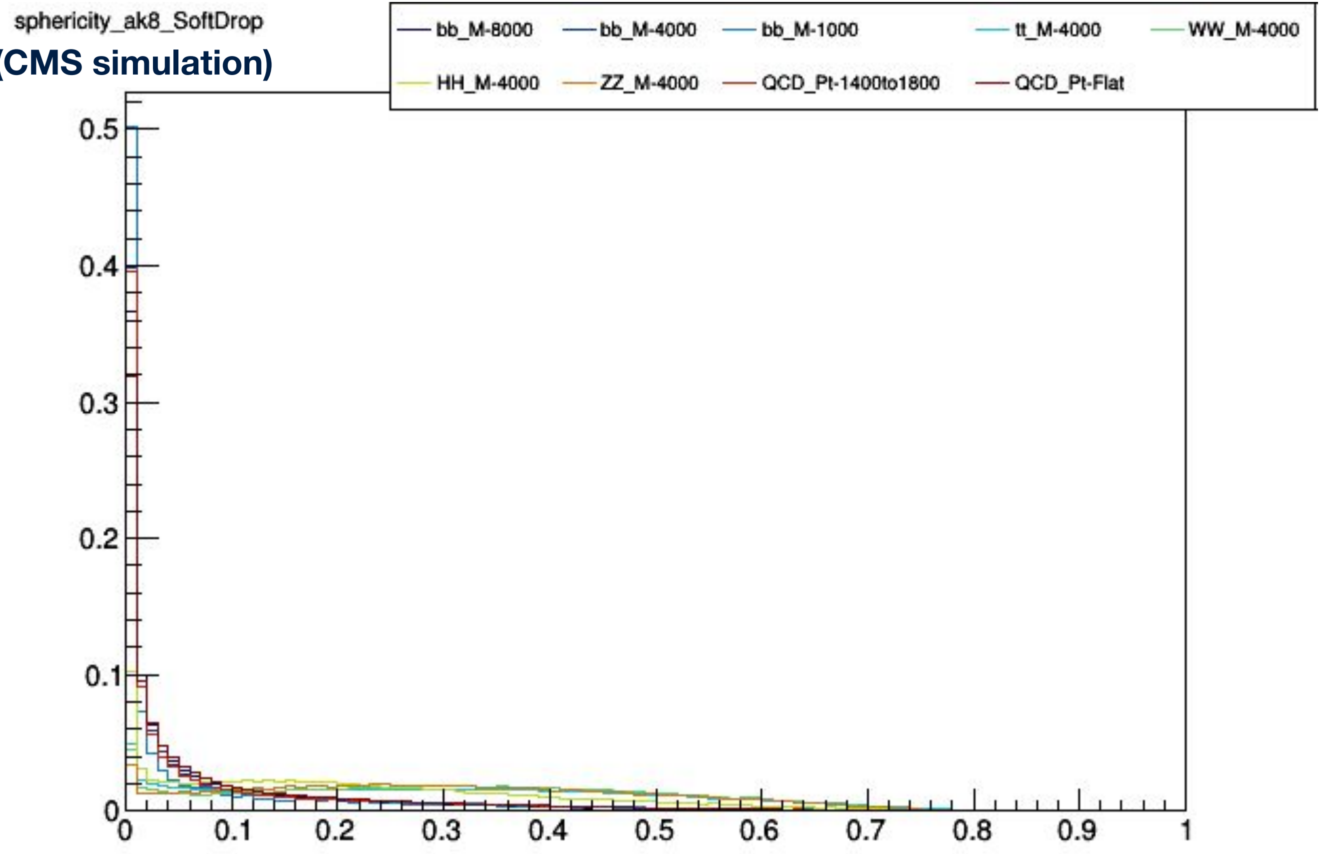


Fox Wolfram Moments: H4



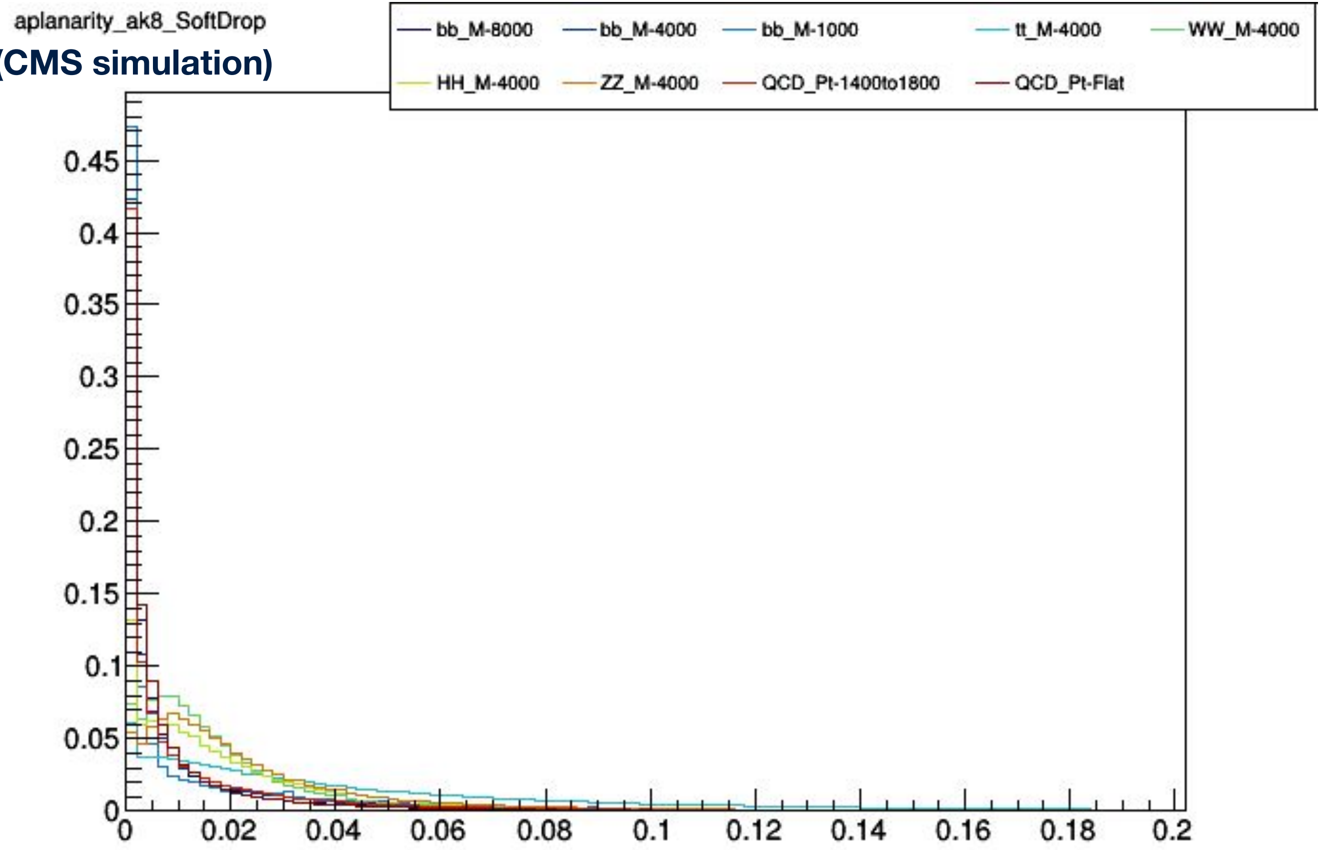
Sphericity Tensor: Sphericity

Private work (CMS simulation)



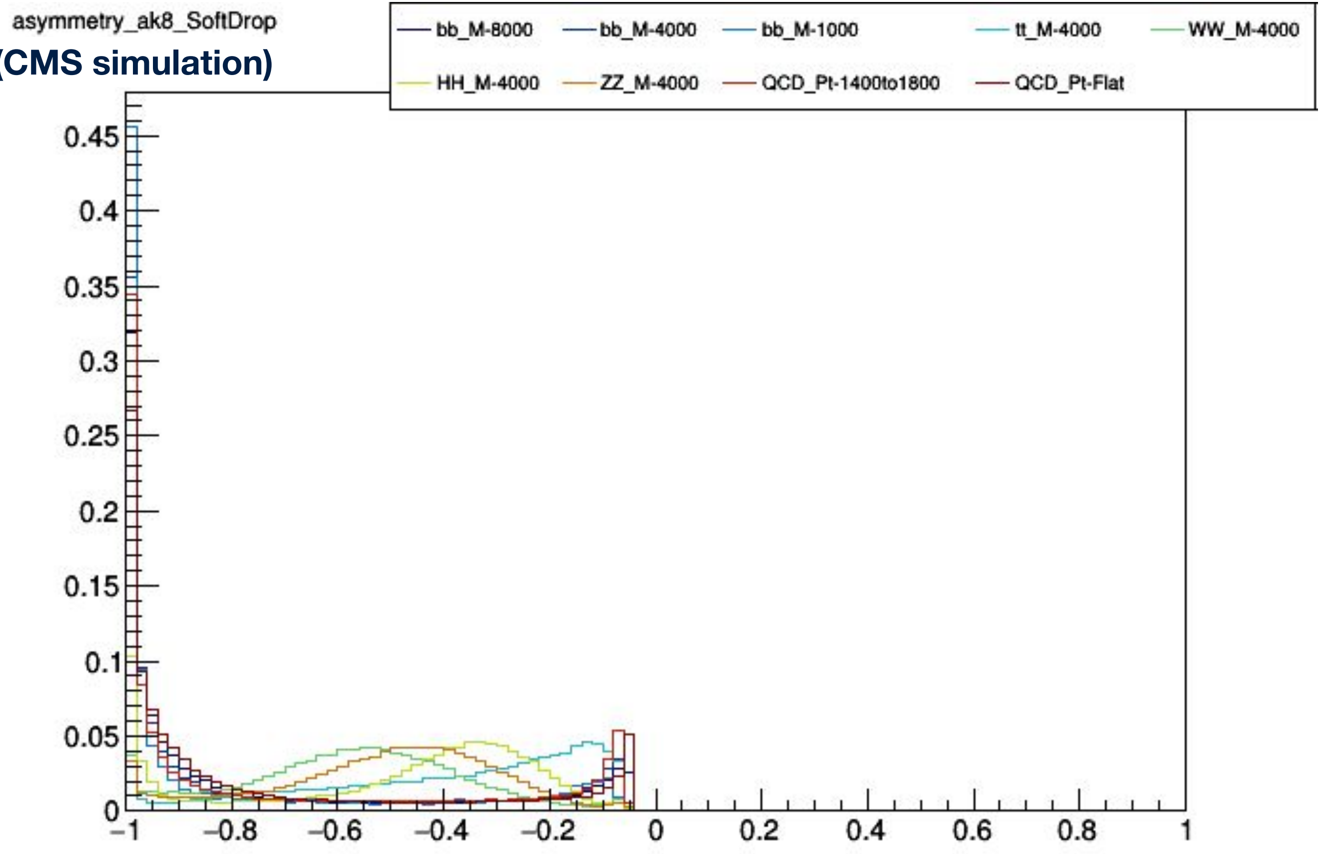
Sphericity Tensor: Aplanarity

Private work (CMS simulation)



Sphericity Tensor: Asymmetry

Private work (CMS simulation)



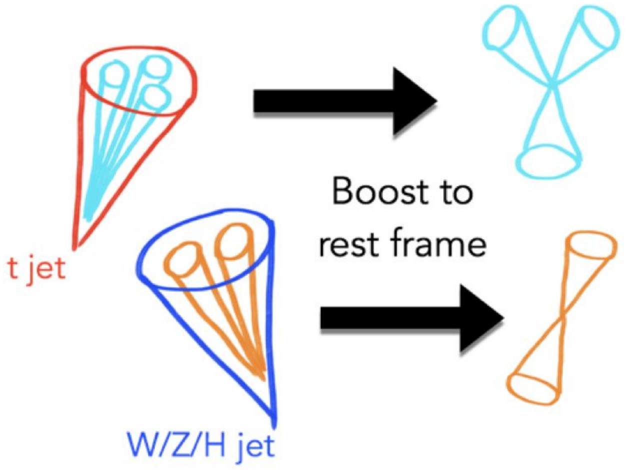
Backup

Input Features: Brain Food

- Boosted Event Shape (**BES**) variables *(262 total)*:
  - Frame Invariant *(17 total)*:
    - AK8 Jet position, charge, & mass
    - Isotropy
    - Secondary Vertices
    - b-tagging scores
    - Subjettiness
  - Frame Dependent *(35 per frame, 245 total)*:
    - Fox-Wolfram moments
    - Sphericity tensor
    - Thrust
    - Re-clustered jet energies, momenta, & angles

7 Boost Frames:

m_{jet}, m_{SD},
W, H, top,
300, 400 GeV



Input Features

Sphericity	Aplanarity	Thrust	Longitudinal Asymmetry
FWM H_1/H_0	FWM H_2/H_0	FWM H_3/H_0	FWM H_4/H_0
p_{x_1}	p_{x_2}	p_{x_3}	p_{x_4}
p_{y_1}	p_{y_2}	p_{y_3}	p_{y_4}
p_{z_1}	p_{z_2}	p_{z_3}	p_{z_4}
E_1	E_2	E_3	E_4
m_{12}	m_{13}	m_{23}	m_{1234}
$cos_{12}(\theta)$	$cos_{13}(\theta)$	$cos_{23}(\theta)$	$cos_{1234}(\theta)$
$cos_{12}(\Delta\theta)$	$cos_{13}(\Delta\theta)$	$cos_{23}(\Delta\theta)$	

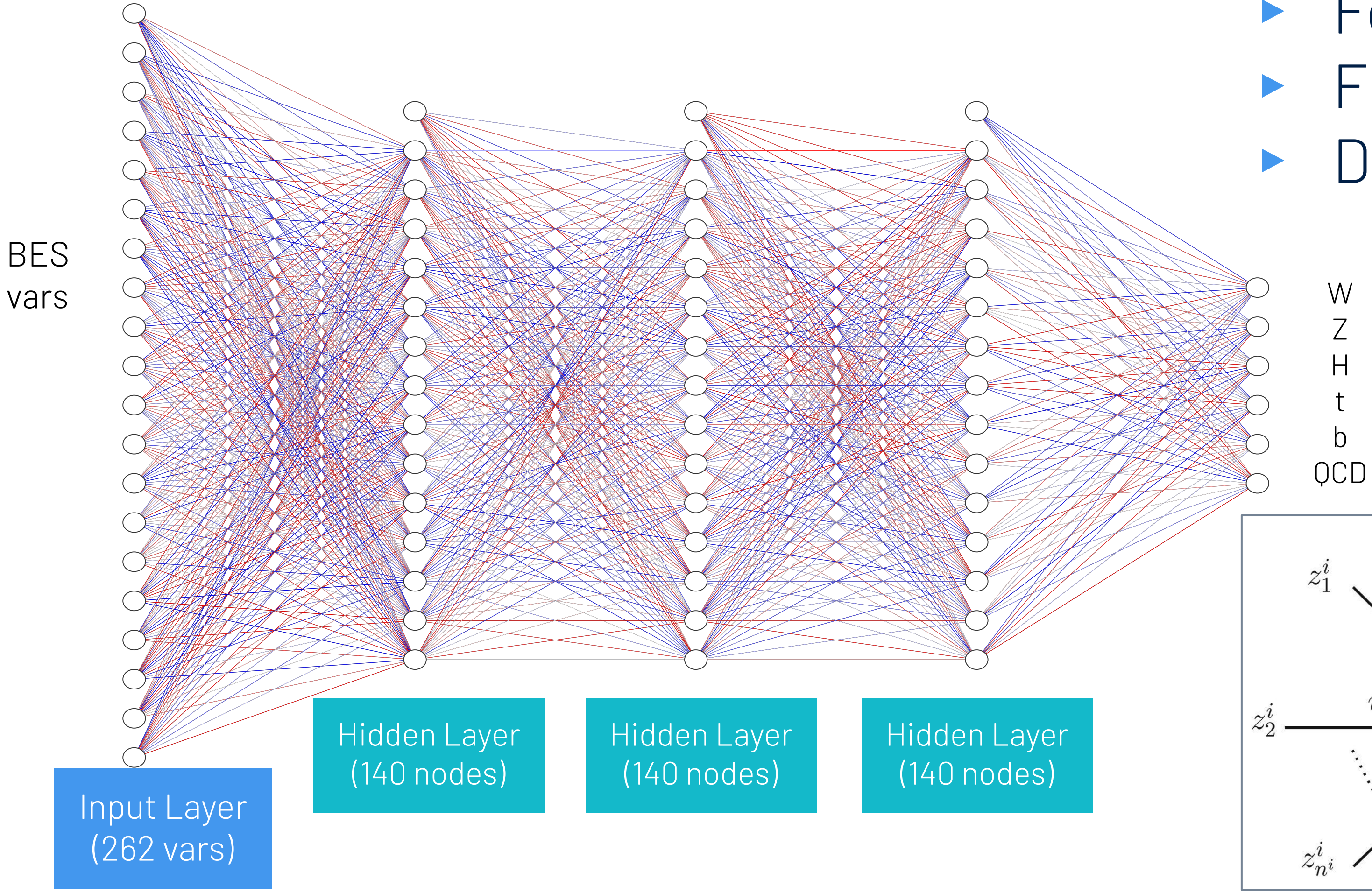
Table 8: List of frame-dependent input features for training the Boosted Event Shape Tagger. For each frame, features such as the normalized Fox-Wolfram Moment (FWM) are examined, along with the four leading-energy reclustered jets.

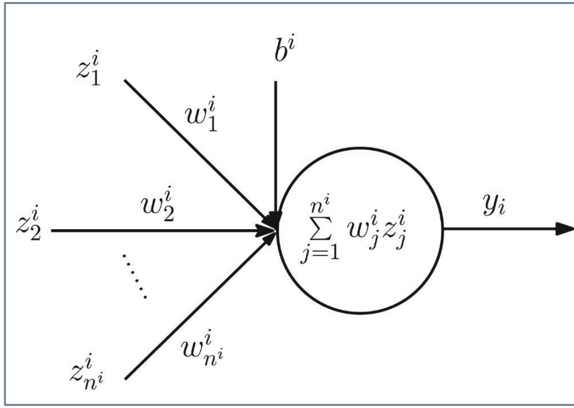
Jet τ_1	Jet τ_2	Jet τ_{21}
Jet τ_3	Jet τ_4	Jet τ_{32}
Jet η	Jet ϕ	Jet Charge
Jet Mass	Jet Soft-Drop Mass	Isotropy
Subjet 1 b-tag	Number of Secondary Vertices	Subjet 1 double b-tag
Subjet 2 b-tag		Subjet 2 double b-tag

Table 9: List of frame-invariant input features for training the Boosted Event Shape Tagger. b-tagging values provided by DeepJet.

Neural Network: The BEST Artificial Brain

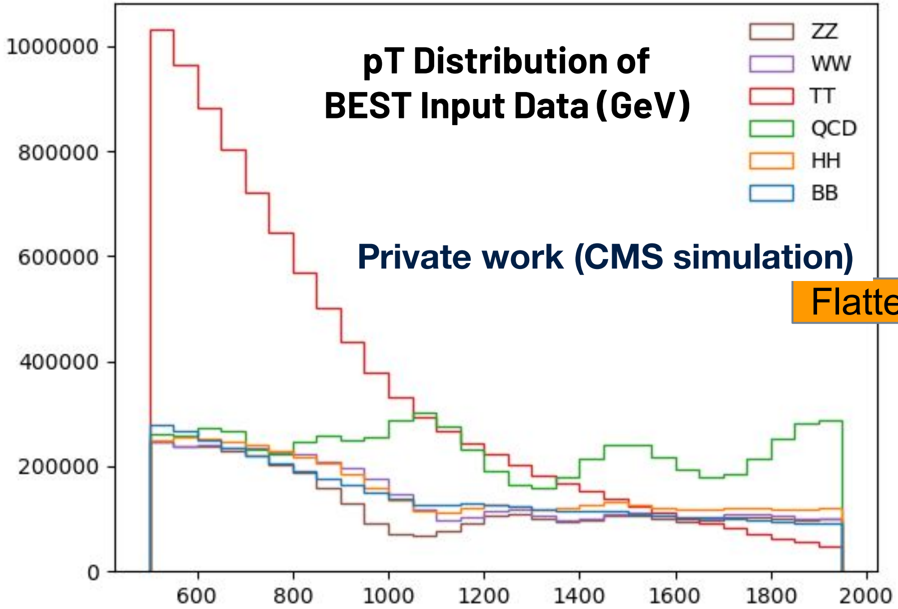
- Feedforward
- Fully Connected
- Dense Layers



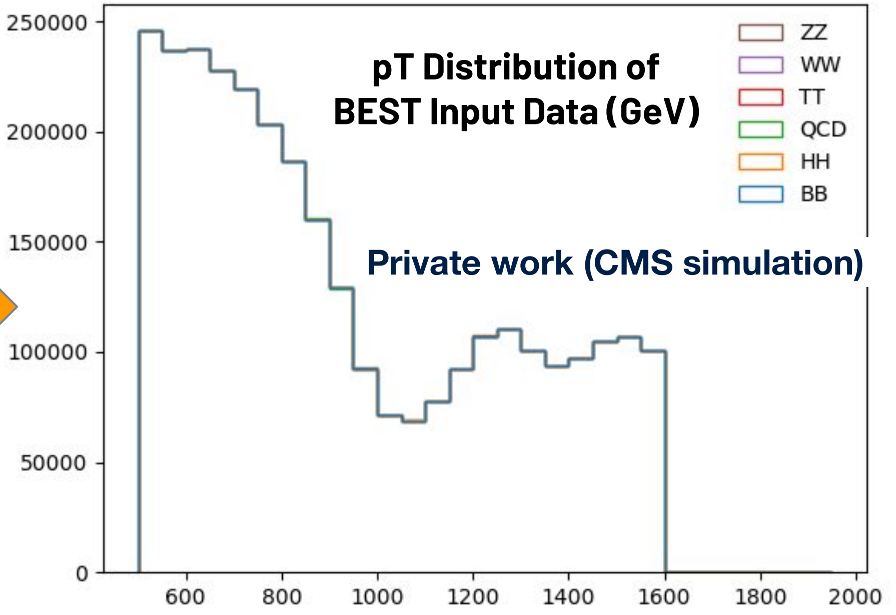


How to Flatten

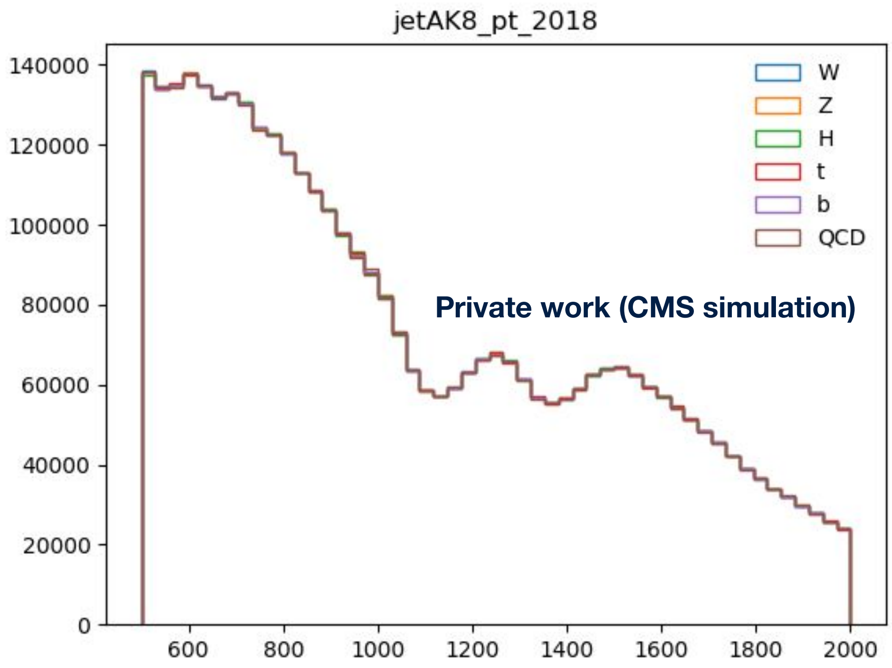
- For each **pT bin**:
  - Identify **process** with the **least events**
  - For the other **5 processes**, begin **randomly dropping events**
    - Continue until they match the process with the **least events**
- Now, the pT distribution for all 6 processes will have **identical shape**

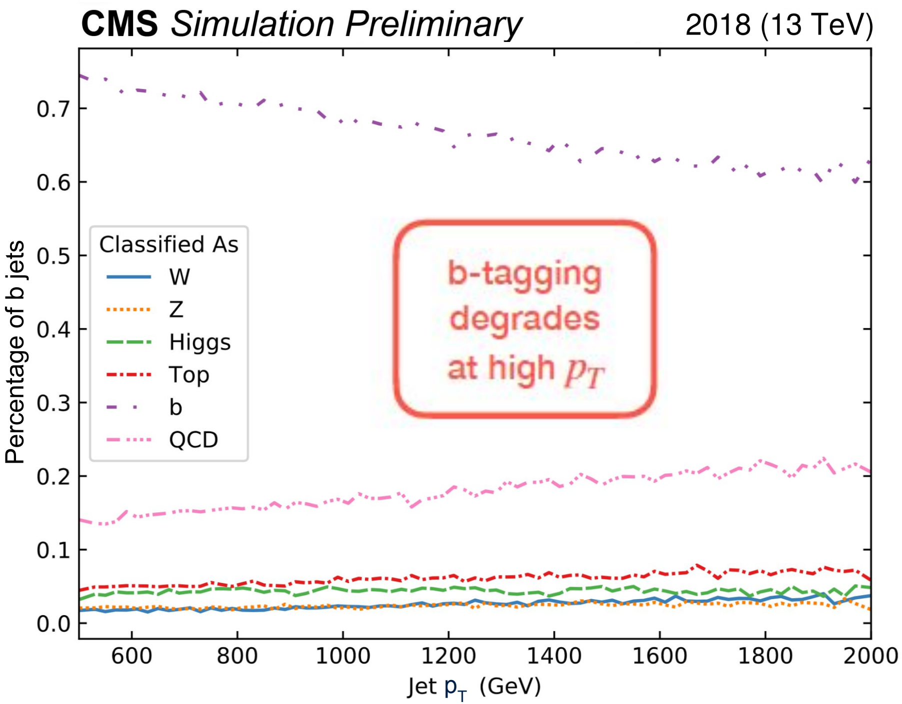


Flattening

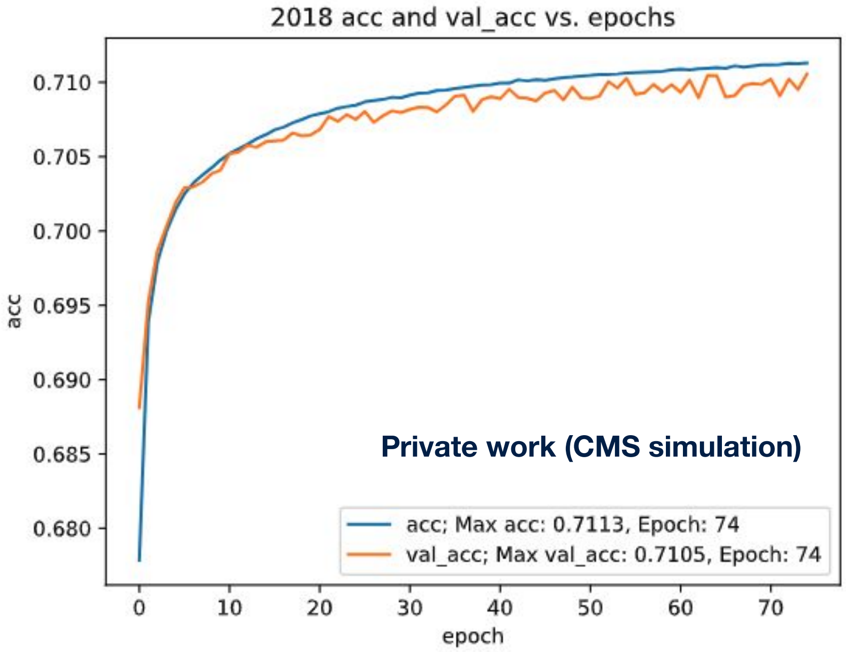


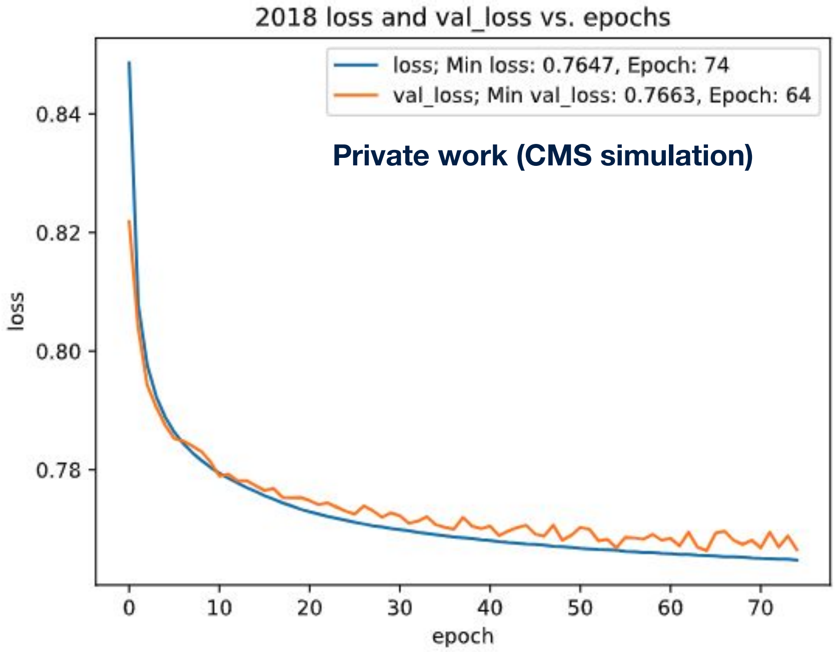
Tagging Behavior pT Dependence





Overtraining





7.2 Statistical Methods

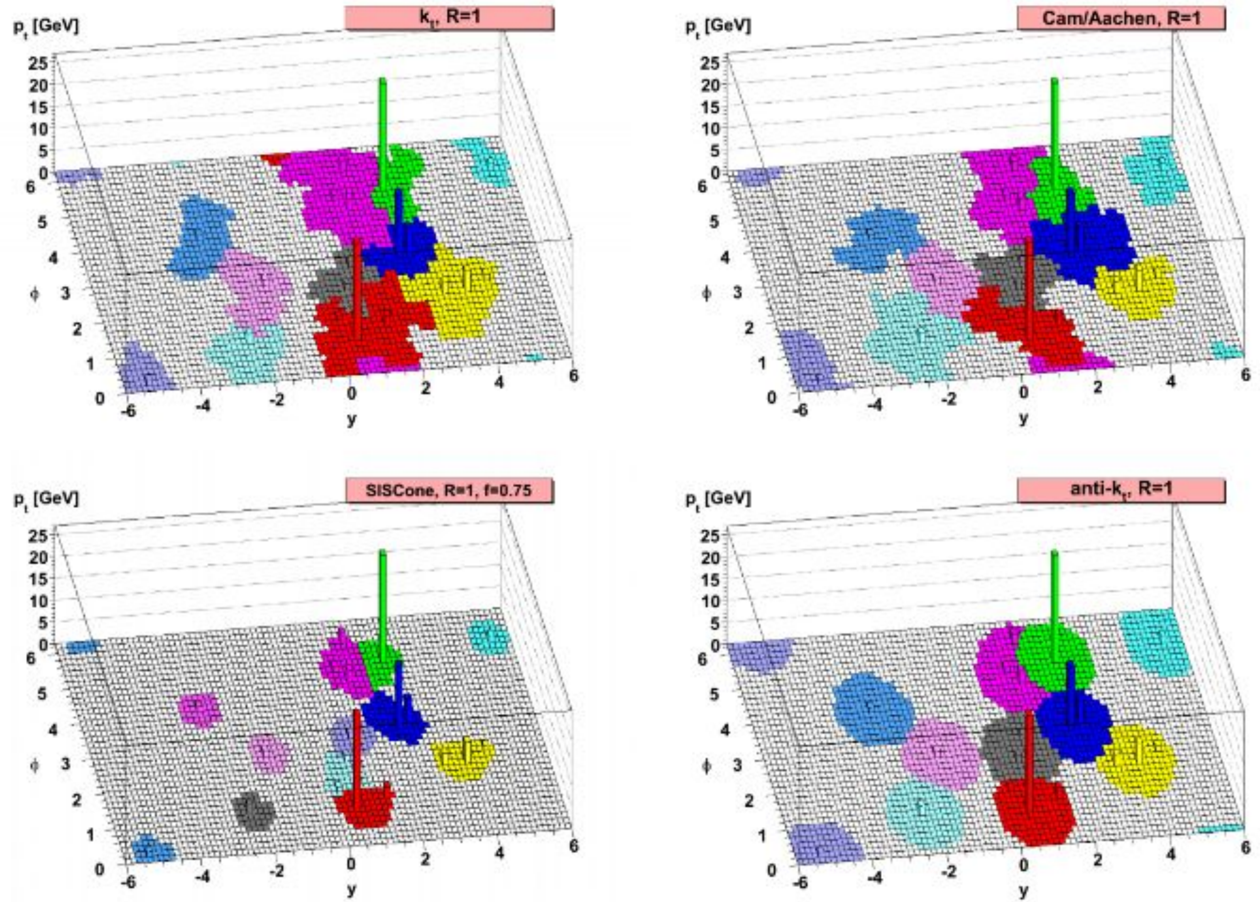
This search uses a binned likelihood to test for the presence of any discrepancy from the predicted background. A shape-based analysis is performed on the H_T distributions in all 120 signal regions. Then, these orthogonal regions are statistically combined together and exclusion limits are placed at 95% confidence.

In this analysis, the exclusions limits are set with the CLS method in the Higgs Combine tool. As the analysis evolves there will be comparisons between the CLS results and Bayesian results in order to check the health of the statistical methods. The starting point is to construct a binned likelihood function. This is done using a Poisson likelihood defined as

$$L(d|r,\vec{\theta}) = \prod_{i=1}^{N} \frac{f_i^{d_i} e^{-f_i}}{d_i!} \tag{19}$$

Where f_i is the model prediction in bin i, and d_i is the data observed in bin i. f_i is actually $f(r,\vec{\theta})$ where r is the parameter of interest and $\vec{\theta}$ are the nuisance parameters of the model [26].

Anti-kT



BEST Confusion Matrix

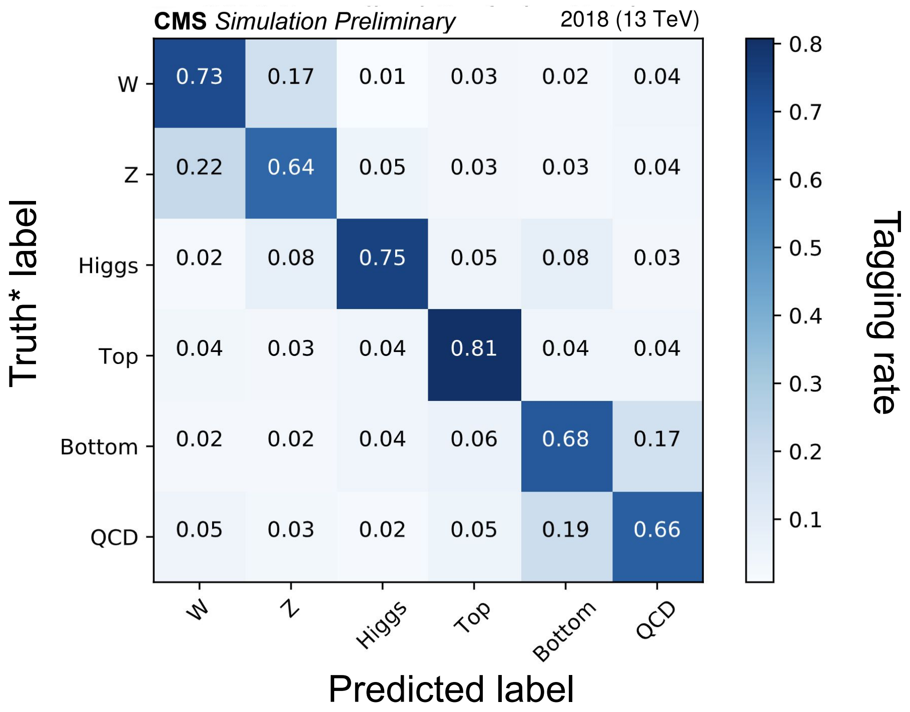


Figure 6: A Confusion Matrix for BEST using 2018 Monte Carlo Simulated data. The x-axis is the classification predicted by BEST, and the y-axis is the truth* label (the actual simulated jet). Along the diagonal are the correct identification rates, and off diagonal entries are mis-identification rates. Each horizontal row is normalized individually, by the total number of truth jets for that category. So, the horizontal rows will sum to 1, but the vertical columns will not.

BEST is performs well at categorizing *H* and *t* jets (with respect to the other categories). BEST shows confusion when trying to discriminate between *W* vs. *Z* jets, but is fairly good at categorizing a jet as *W* ***or*** *Z* (with respect to the other categories). Similarly, BEST confuses *b* vs. *QCD* jets, which makes it difficult to separate *b* jets from background.

BEST: Fraction of Truth Jets Tagged as Higgs Jets by p_T for 2018

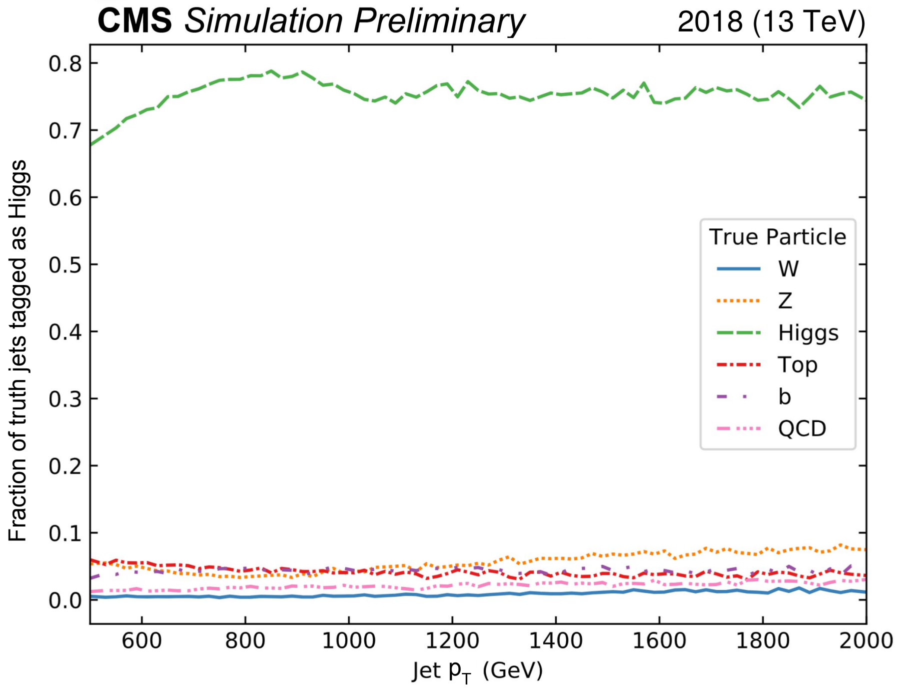


Figure 7: Fraction of truth jets tagged as Higgs jets as a function of AK8 Jet p_T. Notably, the BEST tagging rates are flat in p_T, which is by design. BEST correctly classifies a truth level Higgs jet ~75% of the time.

*This corresponds to the **Predicted Label: Higgs** vertical line of entries on the Confusion Matrix. Each **True Particle** line here matches the corresponding **True Label** box on that **Higgs** vertical line.*

BEST: Percentage of Truth Level Bottom Jets Classified as X by p_T for 2018

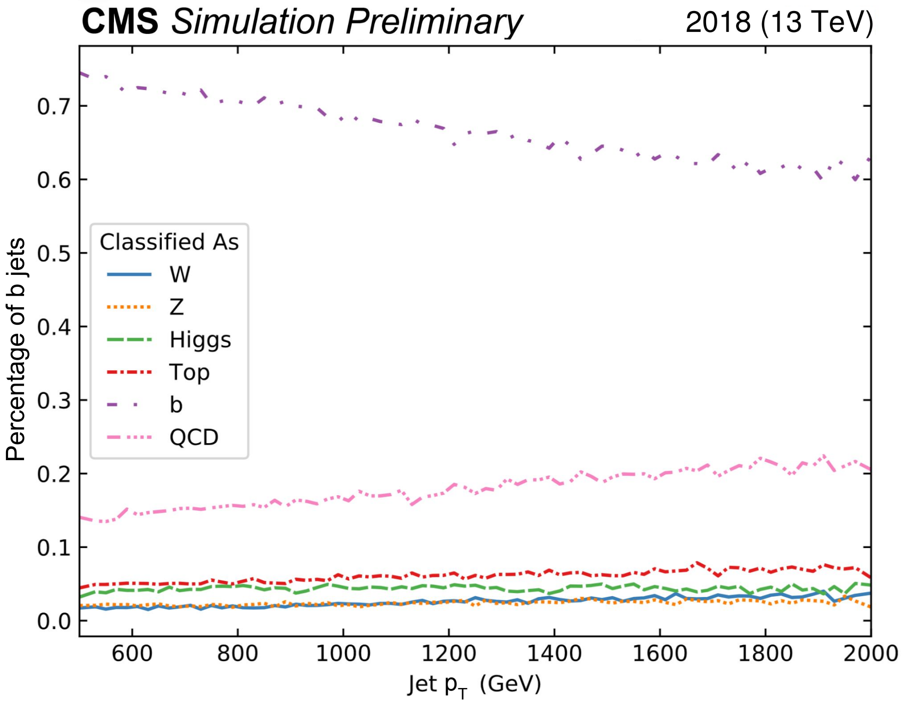
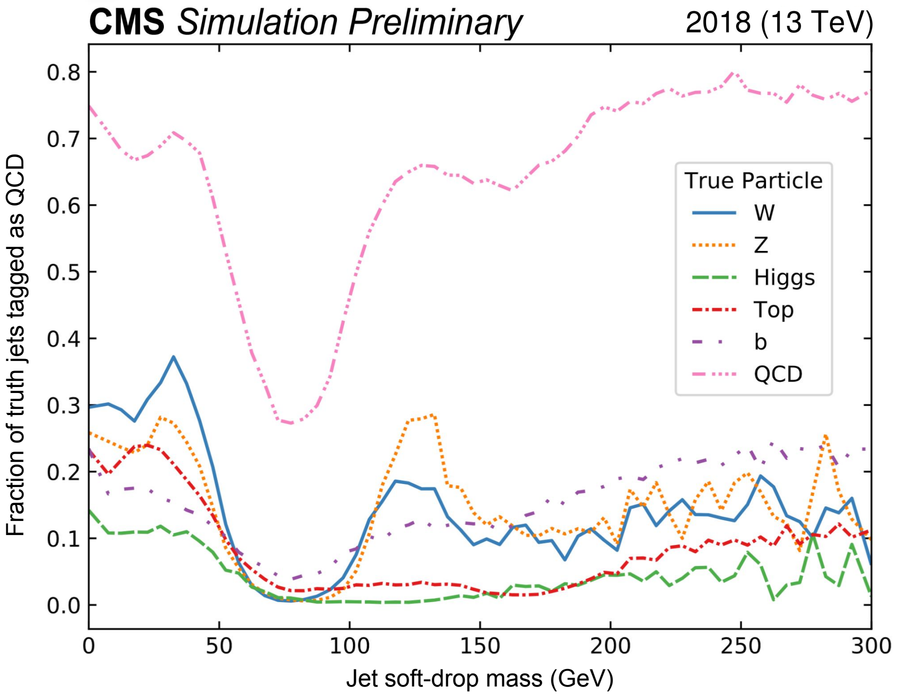


Figure 8: Each BEST category's classification rate for truth level bottom jets, by AK8 Jet p_T. Note that b-tagging degrades at high p_T. This illustrates some of BEST's confusion between b jets and QCD jets.

*This corresponds to the **True Label: Bottom** horizontal line of entries on the Confusion Matrix. Each **Classified As** line here matches the corresponding **Predicted Label** box on that **Bottom** horizontal line.*

BEST: Fraction of Truth Jets Tagged as QCD Jets by SoftDrop* Mass for 2018



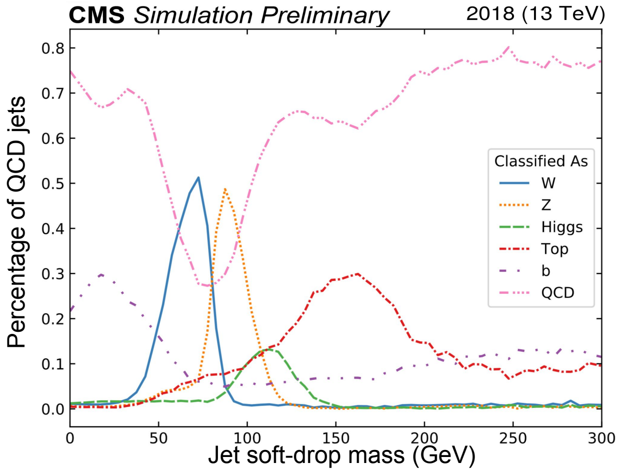


Figure 10: *Each BEST category's classification rate for truth level QCD jets, by AK8 Jet SoftDrop Mass. The QCD lines in both plots are identical. The above plot helps explain the dip in jets tagged as QCD near the W and Z mass.*

Figure 9: Fraction of truth jets tagged as QCD jets as a function of AK8 Jet SoftDrop Mass. BEST is *not* blind to mass like it is to p_T, so these rates are not flat. The dip in the plot near 75 GeV is due to most jets near that mass being categorized as *W* or *Z* jets, across all categories.

This corresponds to the ***Predicted Label: QCD*** *vertical line of entries on the Confusion Matrix. Each* ***True Particle*** *line here matches the corresponding* ***True Label*** *box on that* ***QCD*** *vertical line.*

BEST: Percentage of Truth Level Top Jets Classified as X by Soft-Drop Mass for 2018

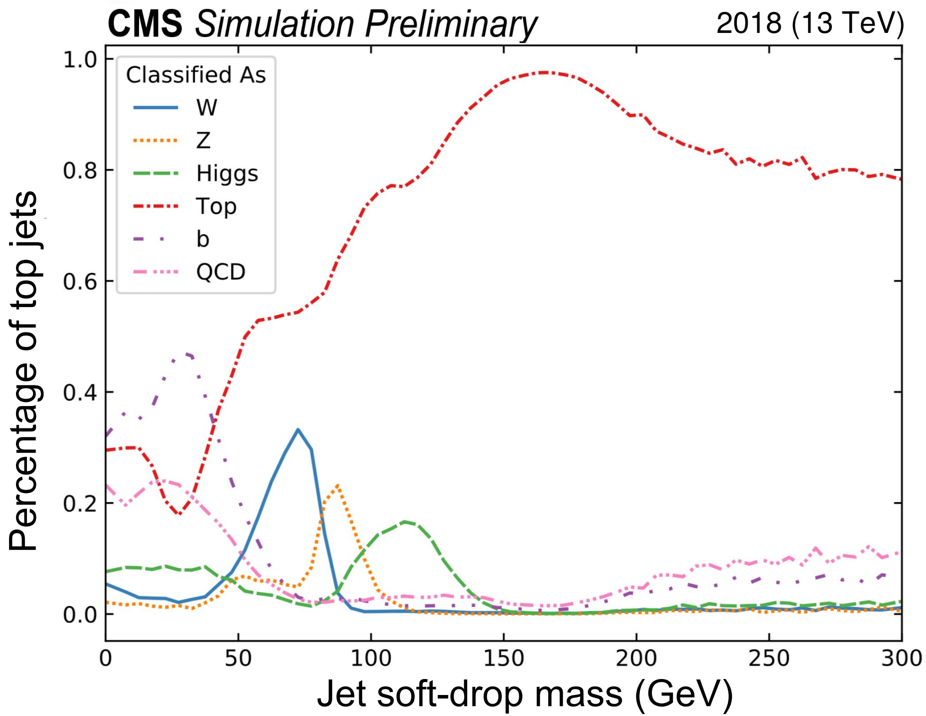


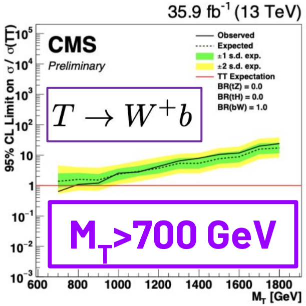
Figure 11: Each BEST category's classification rate for truth level top jets, by AK8 Jet Soft-Drop Mass. The higher the top jet's Soft-Drop Mass, the better BEST is at classifying it. BEST shows confusion when Soft-Drop mass is near another category's rest mass, as seen in the peaks in the misclassification rates for the top quark.

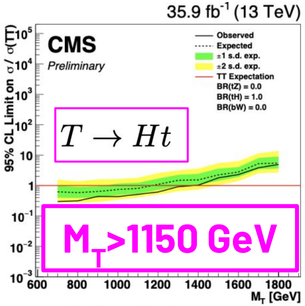
*This corresponds to the **True Label: Top** horizontal line of entries on the Confusion Matrix. Each **Classified As** line here matches the corresponding **Predicted Label** box on that **Top** horizontal line.*

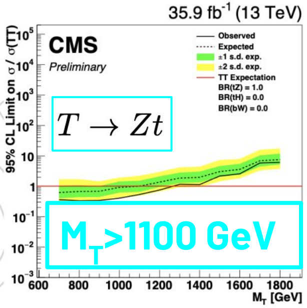
Exclusively the BEST

UCDAVIS CMS

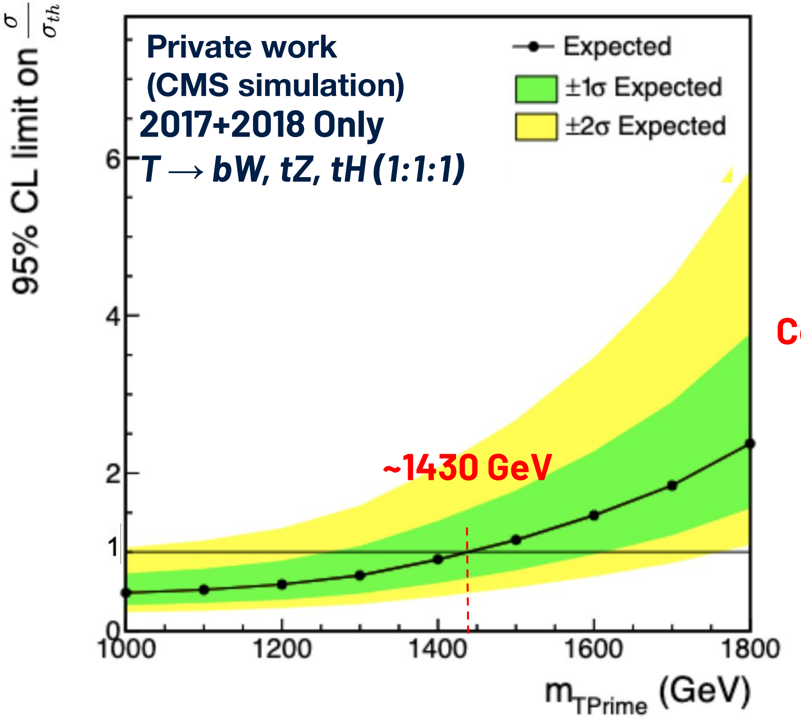
2016 BEST → 2024 BEST







- Mass exclusion limits
  - **Improved** from previous analysis!
    - More data, improved Tagger



- Have not **unblinded**!
  - This plot is a measurement of our **sensitivity**

Compare to world's best:

- **m_{TPrime} > ~1440 GeV**
  - **Full Run 2 (*all years*)**
  - **1:1:1 Branching Ratio**
  - **Leptonic Final States**