New analysis methods in CMS

Maria Mazza on behalf of the CMS Collaboration

Higgs Hunting 2025

July 16, 2025

CMS Compact Muon Solenoid

FSU FLORIDA STATE UNIVERSITY

Talk overview

Highlights of recent machine learning developments for CMS Higgs physics

- Jet tagging

  Status of algorithms for heavy flavor and boosted jet tagging

  A few online and offline applications in CMS

  Comprehensive use case: jet tagging in the search for $t\bar{t}H(c\bar{c})$

- Towards simulation based inference for EFT analyses

  $VH(b\bar{b})$ SMEFT interpretation

- Not an exhaustive list!

Jet tagging

Jet tagging in CMS

Jet tagging via particle clouds (H. Qu and L. Gouskos, 2020)
Particle Transformer for jet tagging (H. Qu, C. Li and S. Qian, 2022)

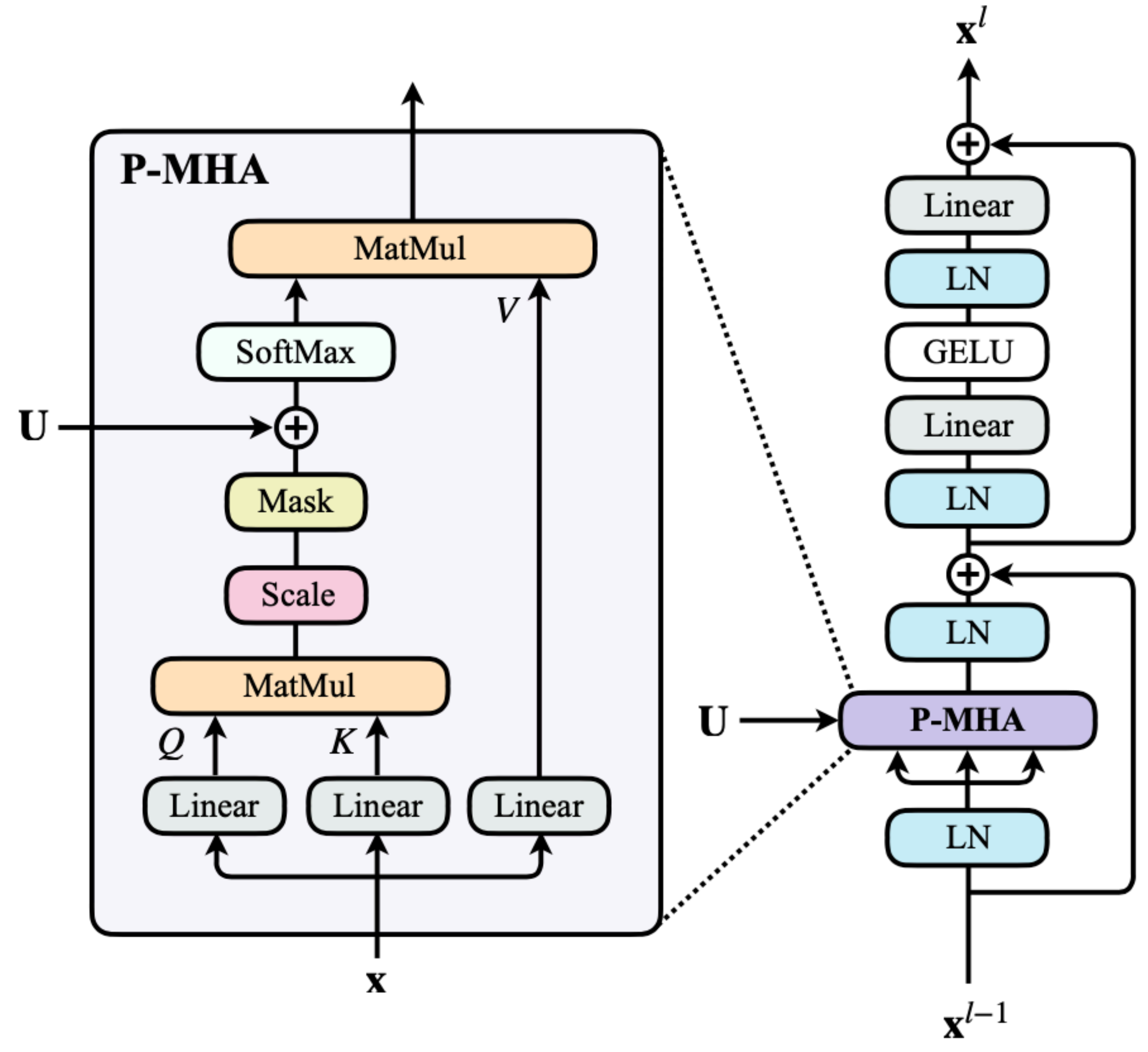
Major improvements in jet tagging performance leveraging new developments in AI community:

ParticleNet: A Dynamic Graph Convolutional Neural Network (DGCNN).

- Treat jets as unordered sets of particles ("particle clouds") for jet tagging (see backup)

ParticleTransformer:

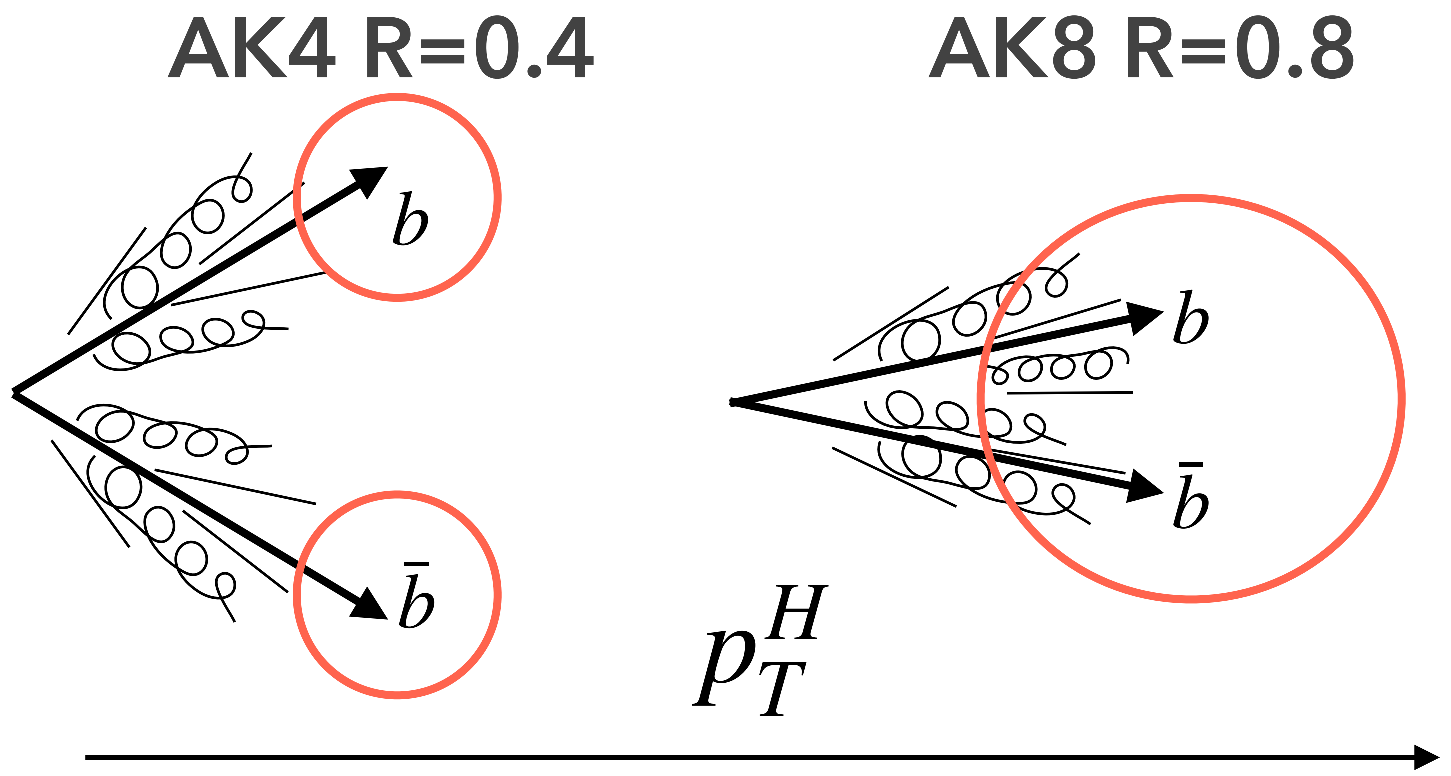
- Introduces particle interaction biases in multi-head attention, improving sensitivity to jet substructure (see backup)
- New tagger developments shifting in this direction



(b) Particle Attention Block

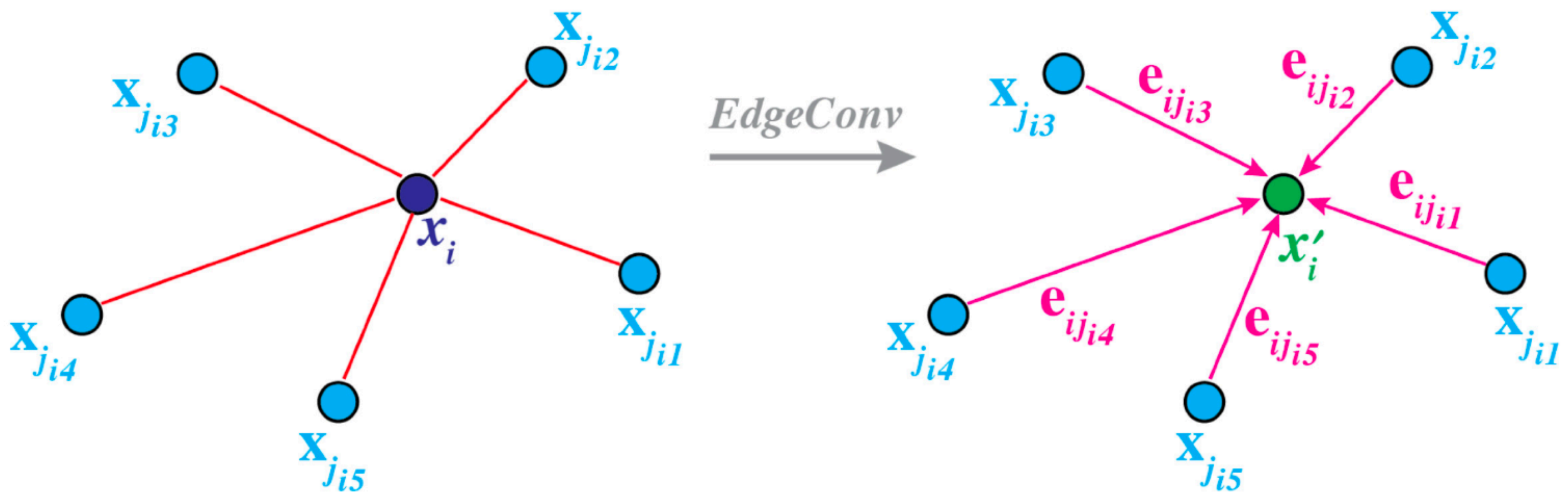
AK4 and AK8 jet identification techniques have evolved in parallel following shared trend of moving towards a more unified strategy:

- Multi-task: jet identification, mass or p_T regression, jet resolution regression
- Multi-class: e.g. τ_h and s-quark AK4 ID



Particle Net

Jet tagging via particle clouds (H. Qu and L. Gouskos, 2020)



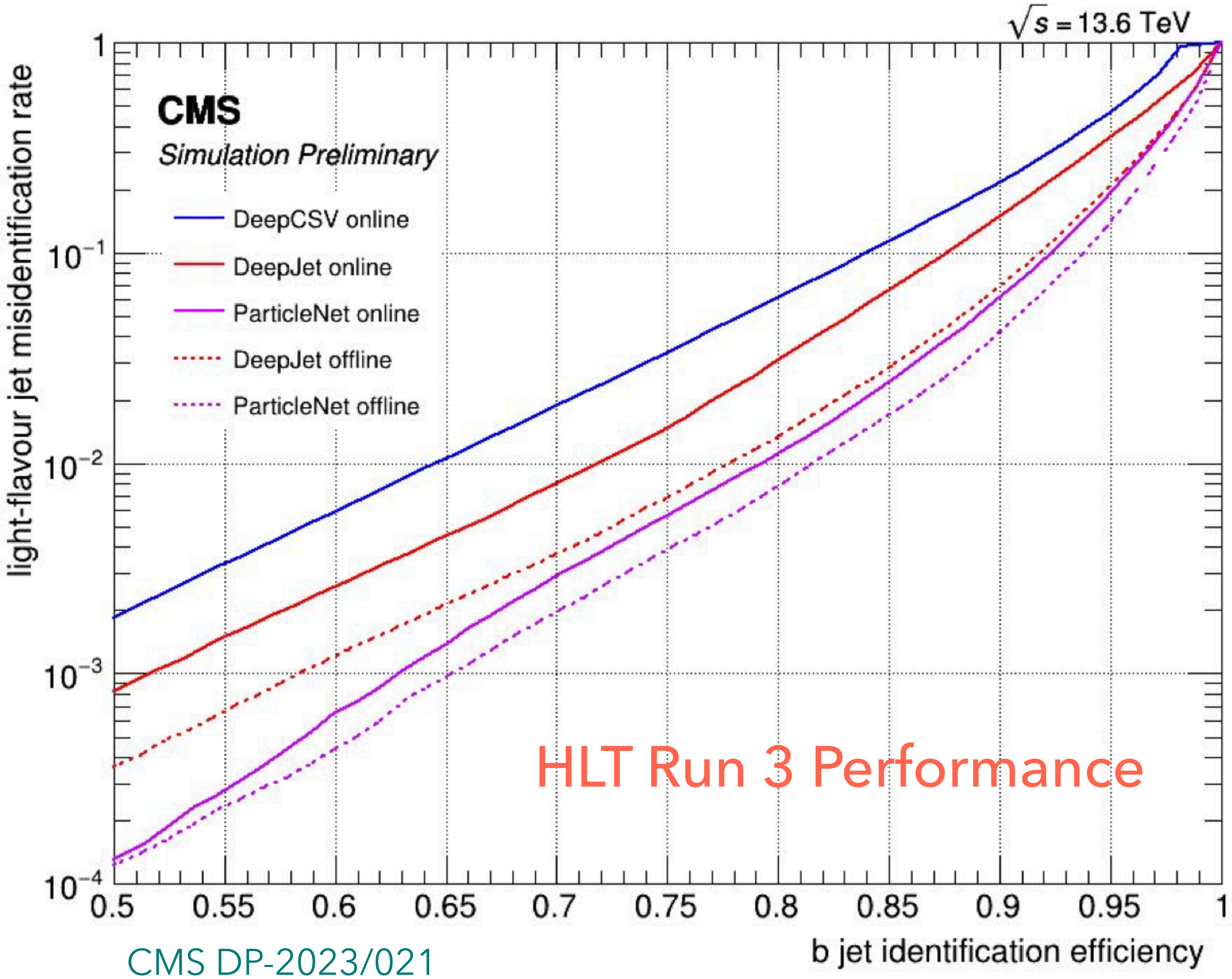
Developed at the end of Run 2, two independent trainings for AK4 and AK8 jet tagging.

A few extensions in terms of capabilities since then:

- Simultaneously provide an estimate of the truth-level jet p_T (mass) for AK4 (AK8) jets and an estimate of the jet resolution
- New jet classes for hadronic tau decays

Deployed at HLT for AK4 and AK8 b-tagging since Run 3.

- Enabled development of new trigger strategy targeting HH and HHH signatures CMS DP-2023/021
- e.g. increased absolute trigger efficiency for HH(4b) from 52% to 68(82)% in 2022 (2023)



CMS DP-2023/021

Unified Particle Transformer (UParT)

[A unified approach for jet tagging in Run 3 (CMS-DP-2024-066)](#)

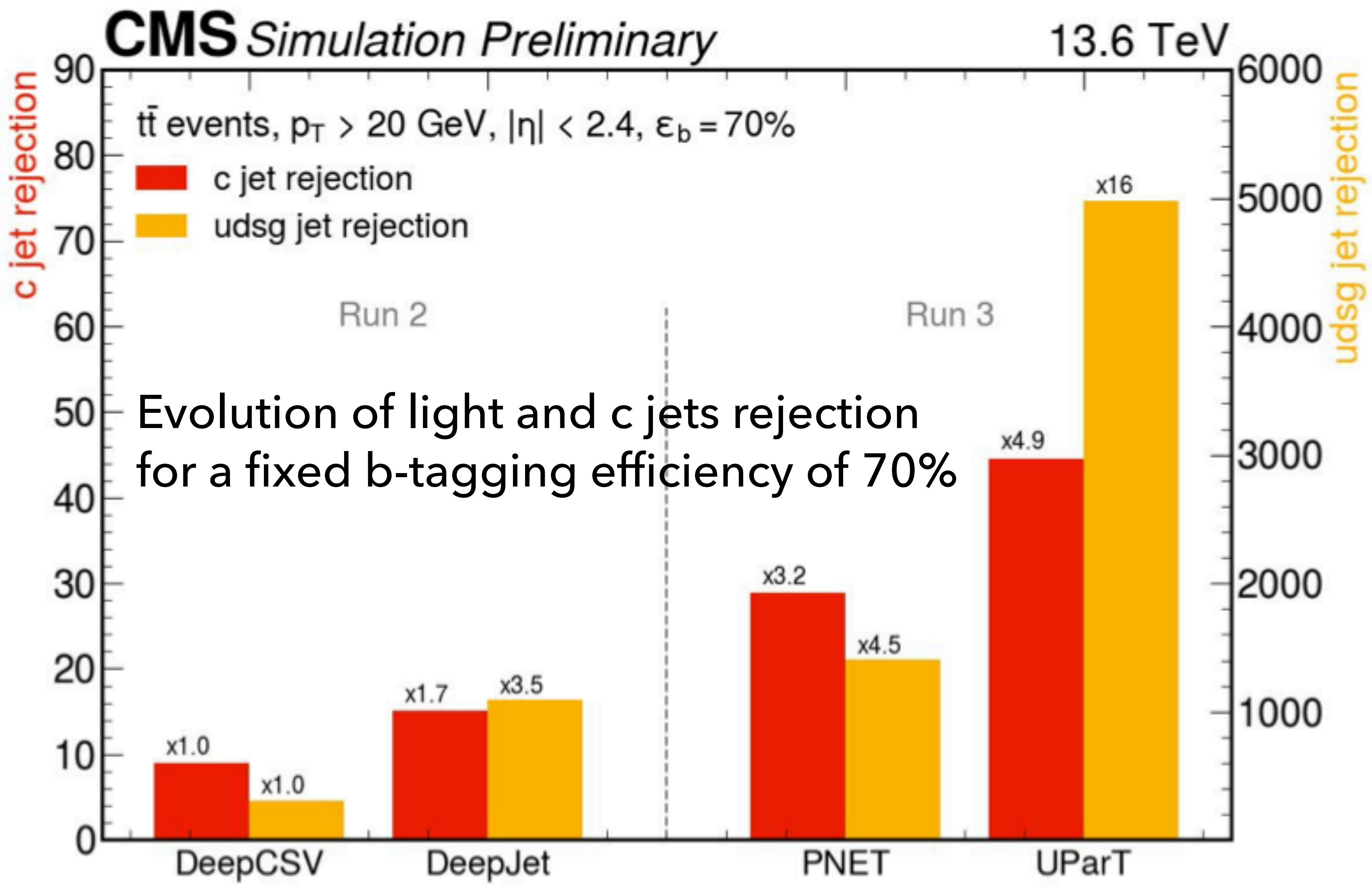
AK4 Jet heavy flavor tagging

A modified ParticleTransformer, with the inclusion of Adversarial Training to enhance robustness against MC mismodeling.

- Perform a distortion of the input features, allowing the model to learn to classify the jet flavor in a region around the jet input features

Additional outputs and augmented loss function:

- τ_h + first time identification of jets originating from s-quarks in CMS.
- Regression tasks for flavor-aware jet energy regression and resolution estimation



UParT shows state-of-the-art heavy flavor jet tagging performance

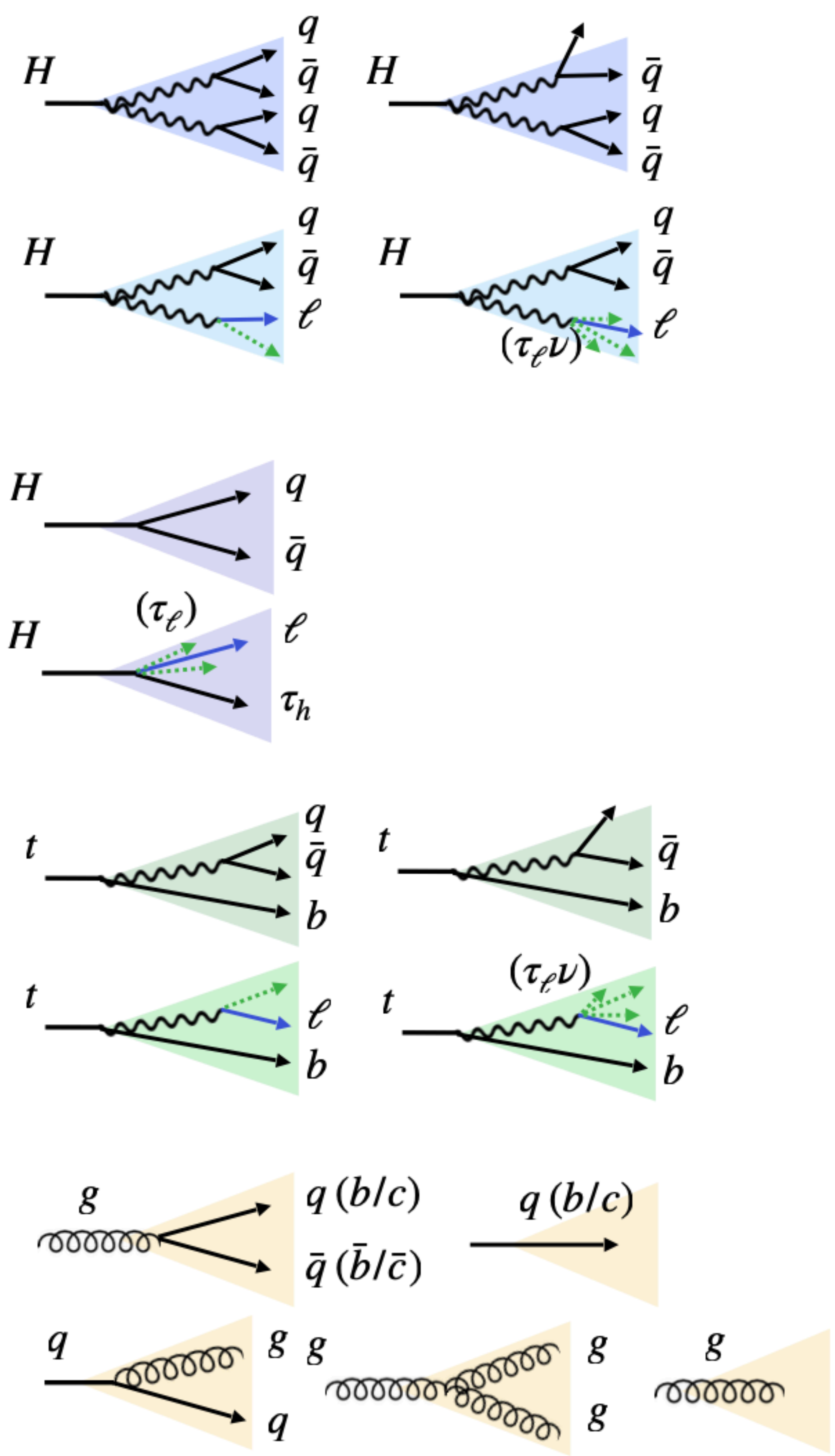
Classification	Regression	Quantile regression (resolution estimation)

$$L = CatEntropy(x, x_{truth}) + \gamma_{regr} \cdot log(cosh(y - y_{truth})) + \gamma_{quantile} \cdot [p_{0.16}(z - z_{truth}) + p_{0.84}(z - z_{truth})]$$

Global Particle Transformer

Boosted AK8 jet tagging

Process	Final state/ prongness	heavy flavour	# of classes
H→VV (full-hadronic)	qqqq	0c/1c/2c	3
	qqq		3
H→WW (semi-leptonic)	eνqq	0c/1c	2
	μνqq		2
	$\tau_e \nu qq$		2
	$\tau_\mu \nu qq$		2
	$\tau_h \nu qq$		2
H→qq		bb	1
		cc	1
		ss	1
		qq (q=u/d)	1
H→ττ	$\tau_e \tau_h$		1
	$\tau_\mu \tau_h$		1
	$\tau_h \tau_h$		1
t→bW (hadronic)	bqq	1b + 0c/1c	2
	bq		2
t→bW (leptonic)	beν	1b	1
	bμν		1
	$b\tau_e \nu$		1
	$b\tau_\mu \nu$		1
	$b\tau_h \nu$		1
QCD		b	1
		bb	1
		c	1
		cc	1
		others (light)	1



Trained to classify between QCD background and various all-hadronic and semileptonic Higgs and top quark decays

- Inputs: AK8 jets with up to 128 Particle Flow candidates and 7 secondary vertices
- 37 classification scores
- Mass decorrelation: training with Higgs- and top quark-like resonances with flat mass spectrum between [15, 250] GeV

Enabled boosted searches for $HH \rightarrow bbWW \rightarrow bb4q$ (CMS PAS HIG-23-012) and $H \rightarrow WW^*$ (CMS PAS HIG-24-008)

Jet tagging in the search for $t\bar{t}H(c\bar{c})$

Search for $t\bar{t}H(c\bar{c})$

Overview

PAS-HIG-24-018 (2025)

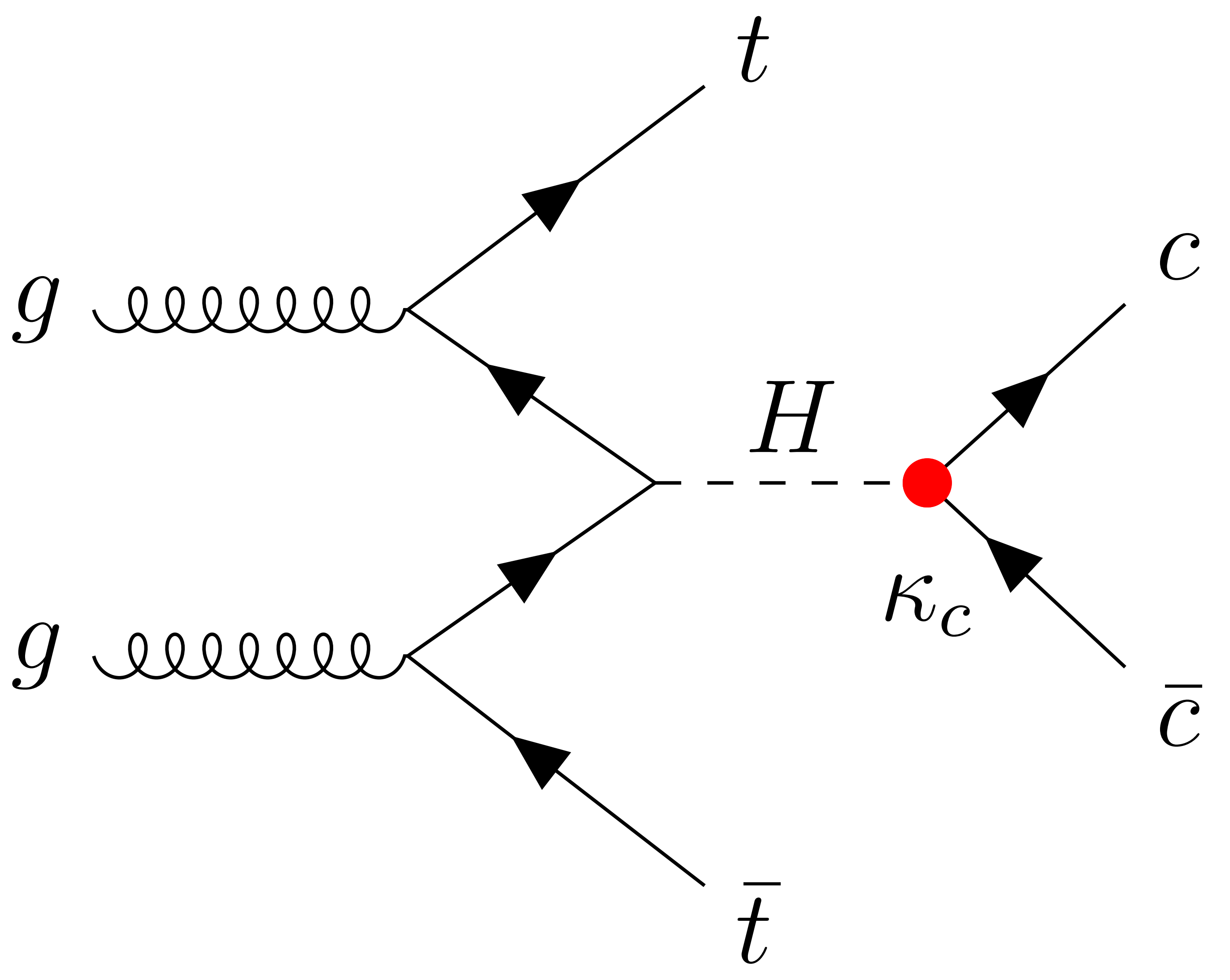
Search for $H \to c\bar{c}$ via associated production of a Higgs boson with a $t\bar{t}$ pair, plus simultaneous measurement of $t\bar{t}H(b\bar{b})$.

Two key challenges:

1. $H \to c\bar{c}$ vs. $H \to b\bar{b}$ discrimination
   - Requires ability to distinguish between jets initiated by b or c quarks
2. $t\bar{t}H$ signal vs. $t\bar{t}$+jets background discrimination
   - Jet multiplicity makes jet-parton assignment task difficult

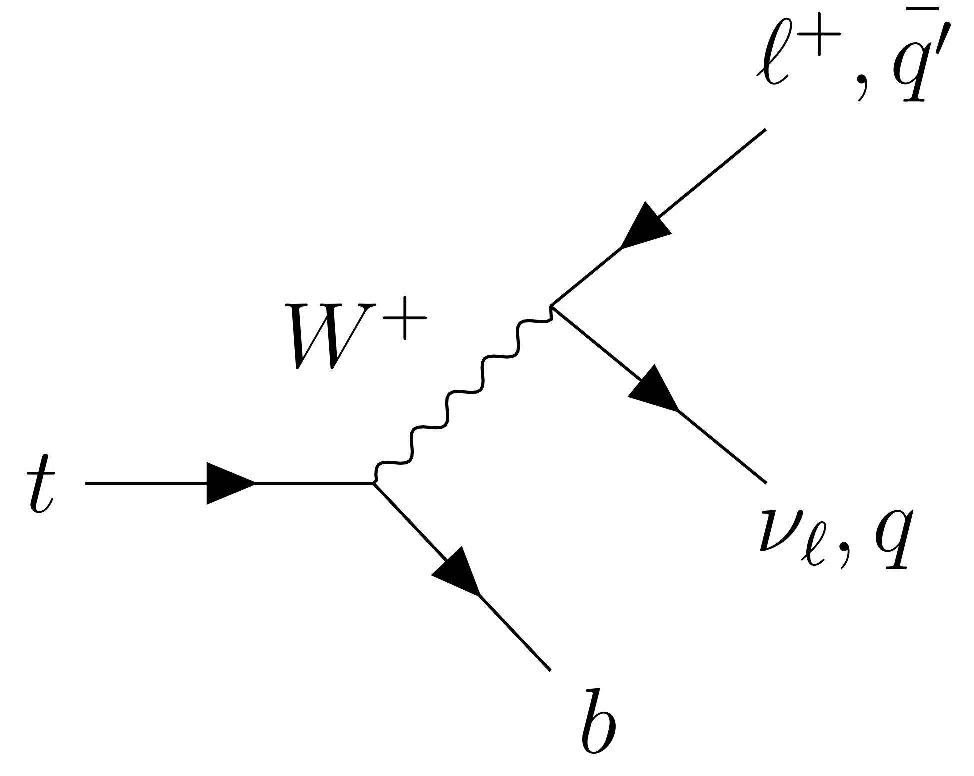
New ML algorithms allow to address these challenges:

- ParticleNet for jet flavor identification
- Bypass reconstruction of parent parton with Particle Transformer for signal vs. background event classification



Decay mode of the top quarks define three categories:

0, 1, and 2 lepton



Search for $t\bar{t}H(c\bar{c})$

PAS-HIG-24-018 (2025)

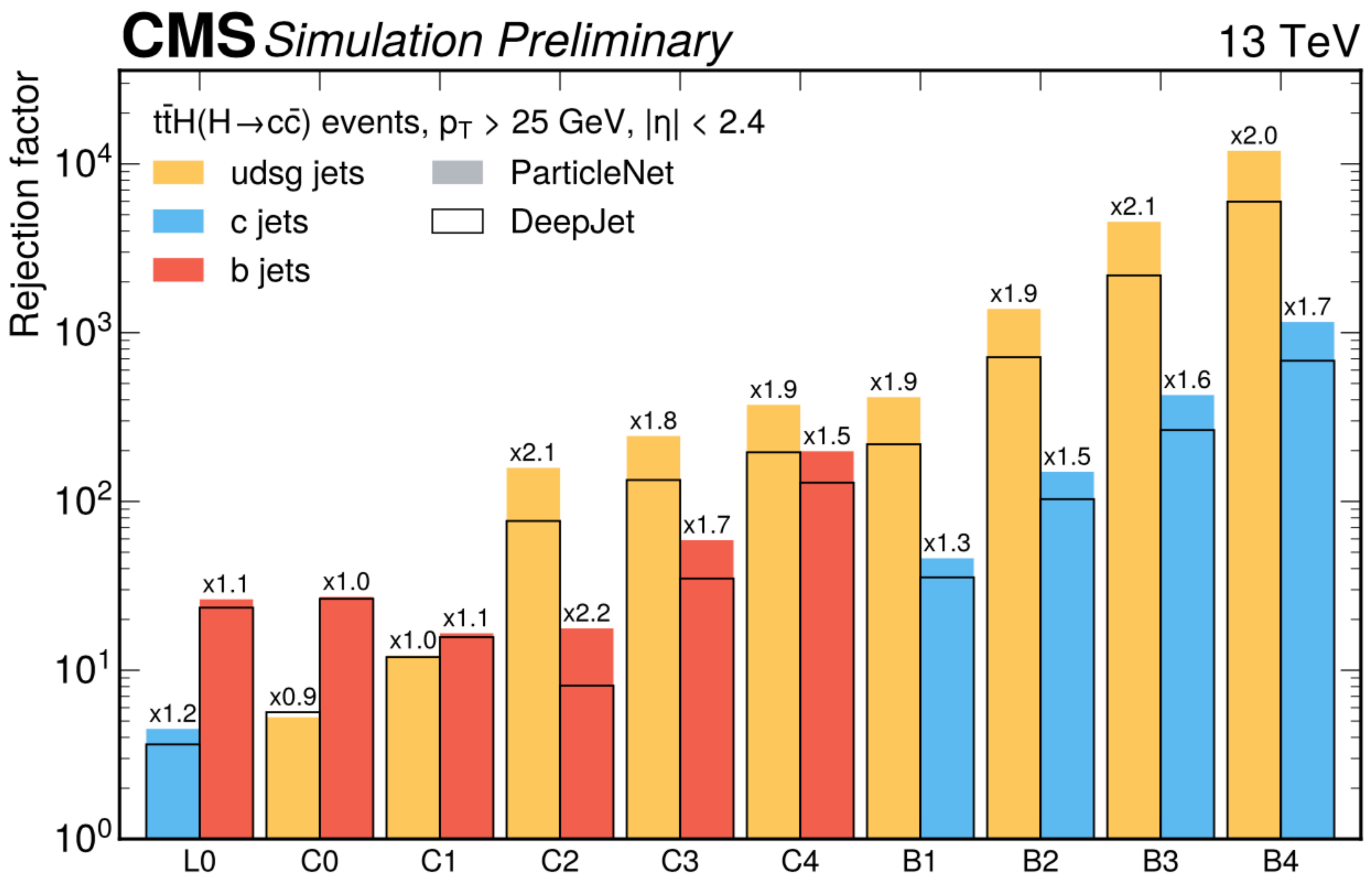
ParticleNet for jet flavor identification

Two discriminants are constructed from the ParticleNet output:

$$p_{B+C} = \frac{p_b + p_{bb} + p_c + p_{cc}}{p_b + p_{bb} + p_c + p_{cc} + p_{uds} + p_g}$$

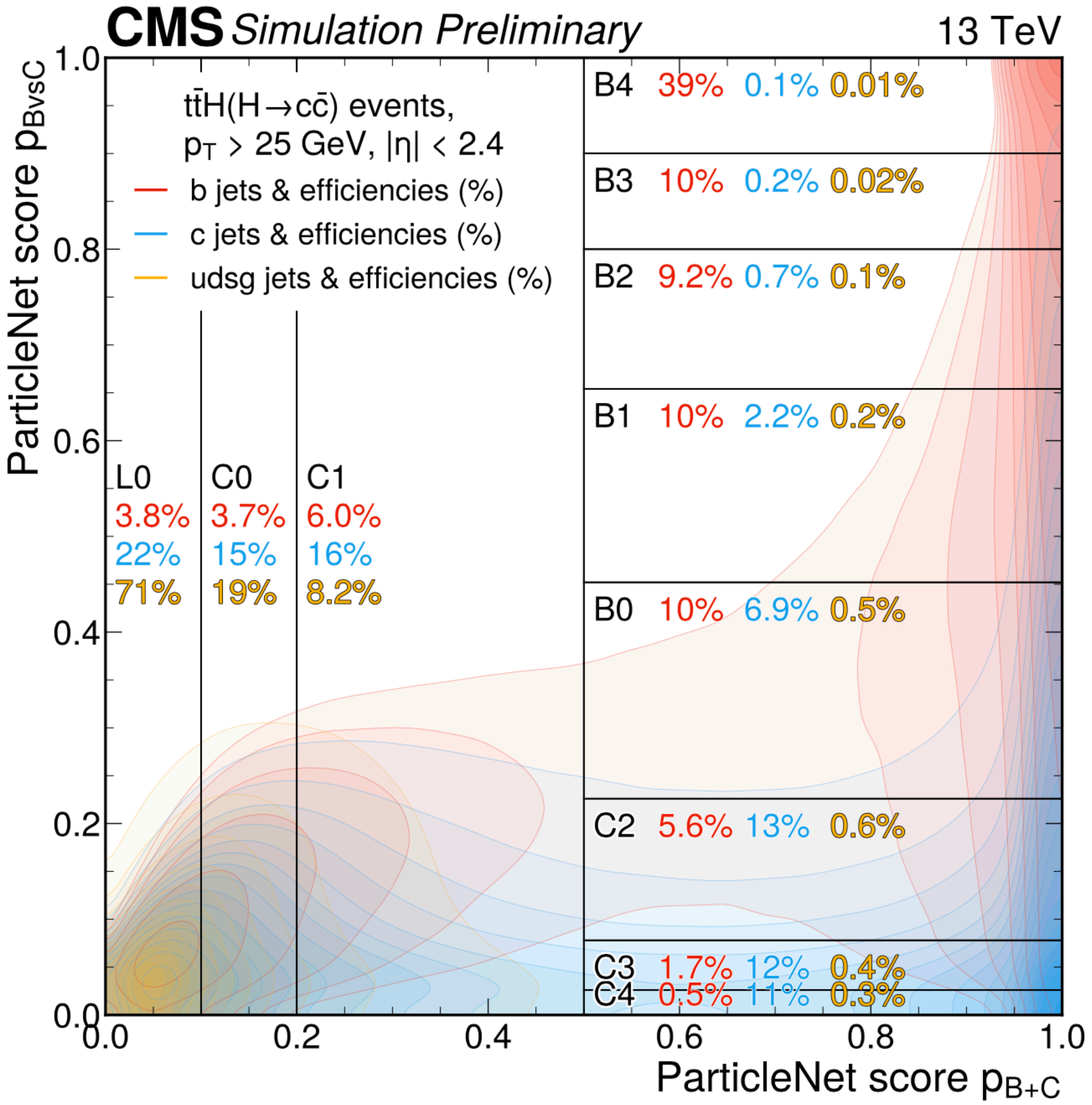
$$p_{BvsC} = \frac{p_b + p_{bb}}{p_b + p_{bb} + p_c + p_{cc}}$$

p_{BvsC} and p_{B+C} are used to define 11 mutually exclusive tagging categories.



~x2 background rejection improvement over early Run 2 DeepJet at the same signal jet efficiency.

Higher p(b) vs.p(c)



Higher p(b and c) vs. p(u, d, s, and g)

Search for $t\bar{t}H(c\bar{c})$

PAS-HIG-24-018 (2025)

ParT for event classification

A multi-class event classifier based on ParT is developed to separate signal vs. background events directly from final state objects, avoiding explicit reconstruction of partons.

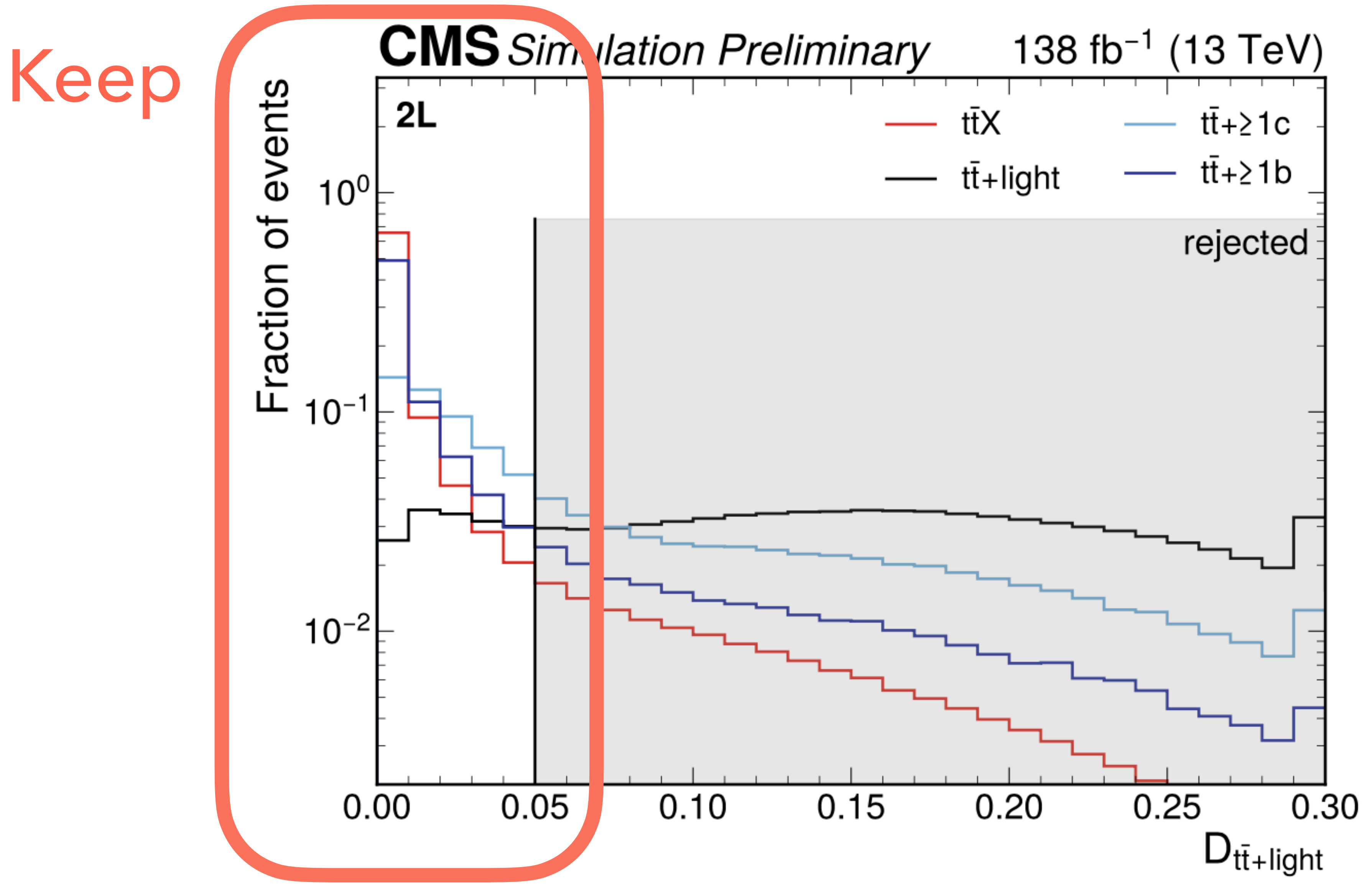
- Kinematics and tagging information with pairwise features for all object pairs

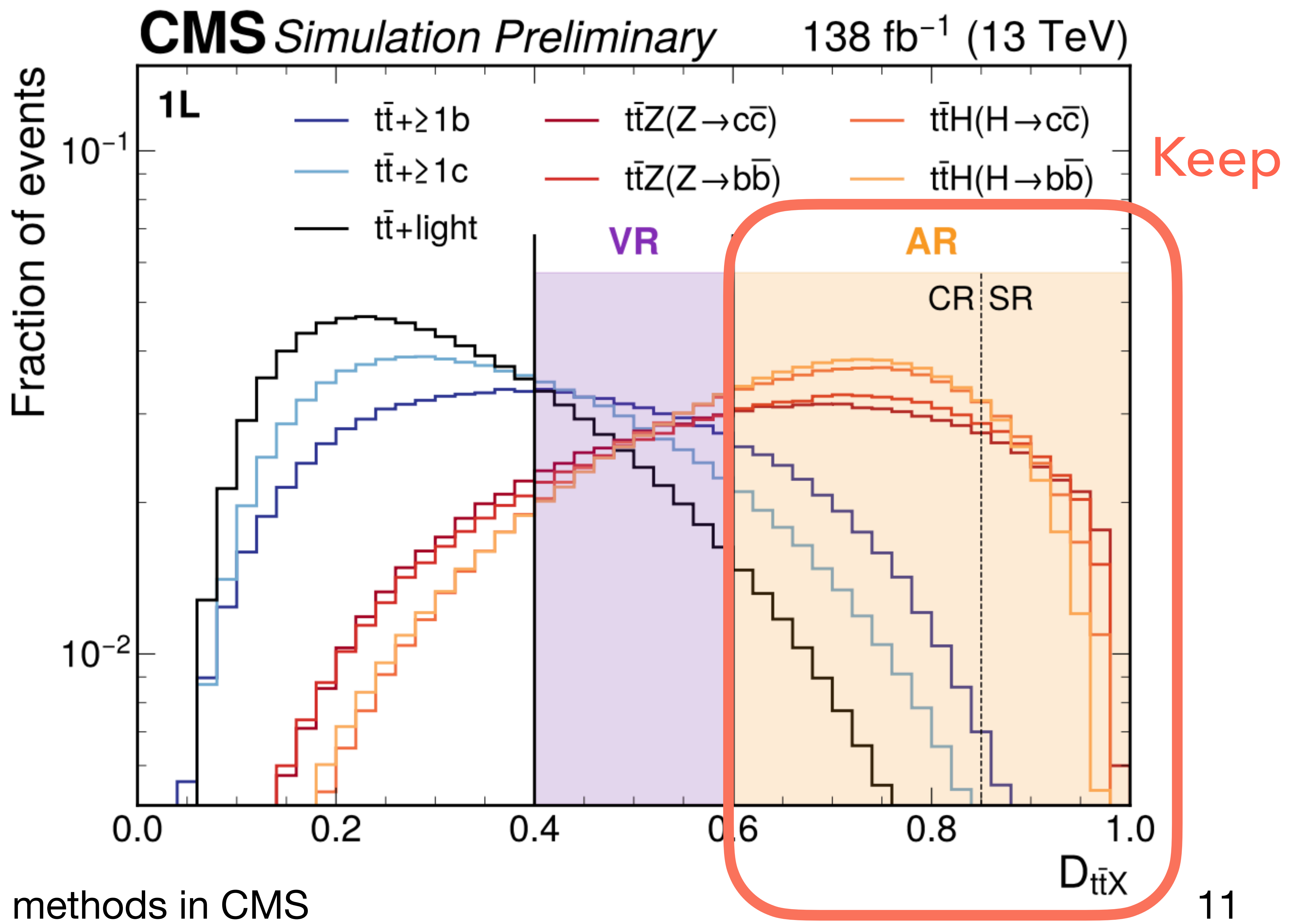
Output classes

- Signal like: $t\bar{t}H(c\bar{c})$, $t\bar{t}H(b\bar{b})$, $t\bar{t}Z(c\bar{c})$, $t\bar{t}Z(b\bar{b})$
- $t\bar{t}$+jets: $t\bar{t}$+light, $t\bar{t}+b$, $t\bar{t}+\geq 2b$, $t\bar{t}+c$, $t\bar{t}+\geq 2c$
- QCD (0Lep only)

Cuts on the output scores allows to reject background and select regions of higher signal purity.

$$D_{t\bar{t}X} = D_{t\bar{t}H(cc)} + D_{t\bar{t}H(bb)} + D_{t\bar{t}Z(cc)} + D_{t\bar{t}Z(bb)}$$





Search for $t\bar{t}H(c\bar{c})$

ParT output as fit discriminant

Binned profile likelihood fit of the ParT event classifier discriminant.

- Signal strengths for $t\bar{t}Z(c\bar{c})$ and $t\bar{t}Z(b\bar{b})$ production are measured to validate analysis strategy and agree with SM.
- 27 final regions (plot shows 2-lepton channel)

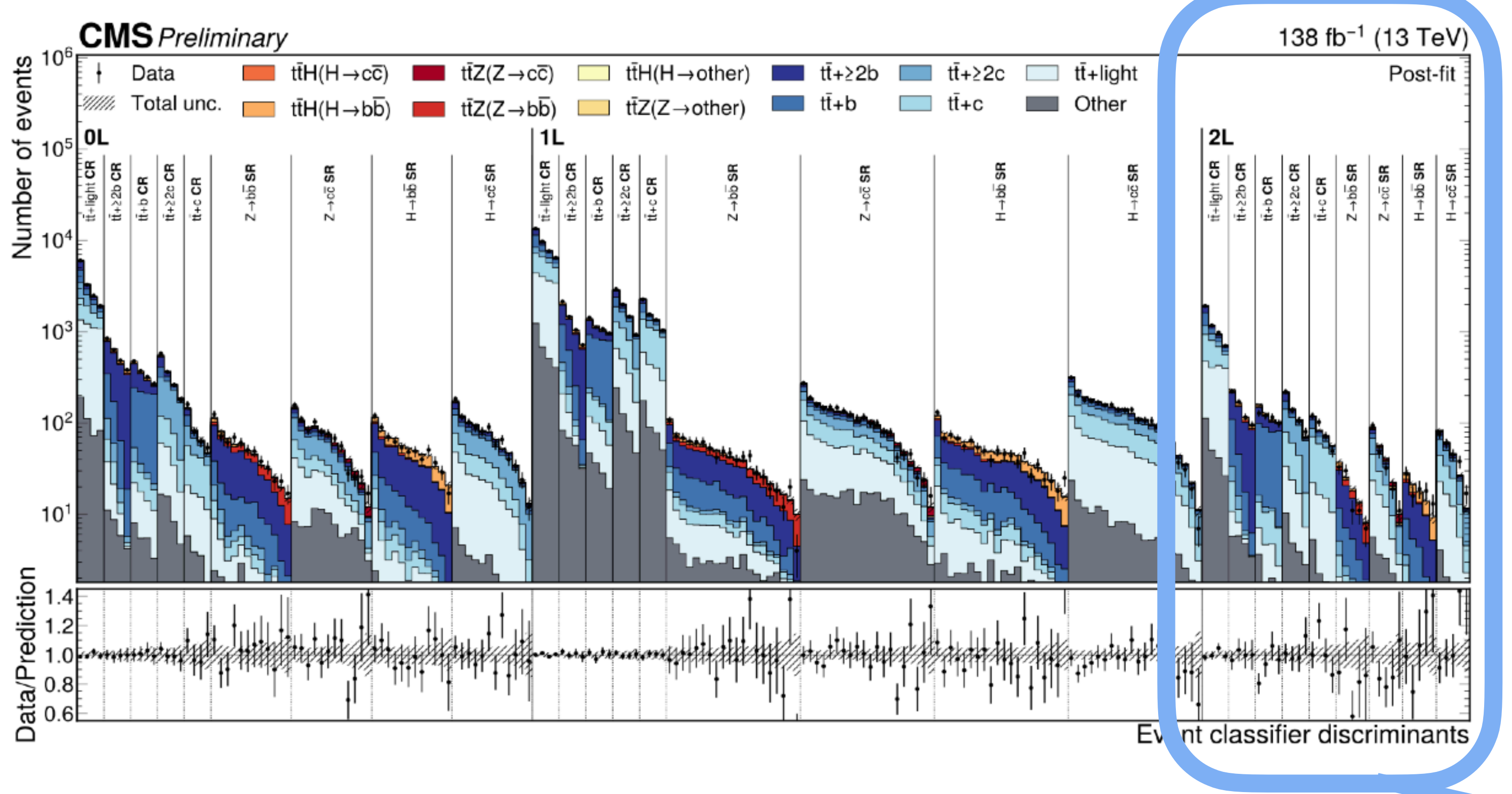
Category	Discriminant
$H\to c\bar{c}$ SR	$\mathcal{D}_{t\bar{t}H(H\to c\bar{c})}$
$H\to b\bar{b}$ SR	$\mathcal{D}_{t\bar{t}H(H\to b\bar{b})}$
$Z\to c\bar{c}$ SR	$\mathcal{D}_{t\bar{t}Z(Z\to c\bar{c})}$
$Z\to b\bar{b}$ SR	$\mathcal{D}_{t\bar{t}Z(Z\to b\bar{b})}$
$t\bar{t}$+light CR	$\mathcal{D}_{t\bar{t}+\text{light}}/\mathcal{D}_{t\bar{t}+\text{jets}}$
$t\bar{t}$+c CR	$\mathcal{D}_{t\bar{t}+c}/\mathcal{D}_{t\bar{t}+\text{jets}}$
$t\bar{t}+\geq 2c$ CR	$\mathcal{D}_{t\bar{t}+\geq 2c}/\mathcal{D}_{t\bar{t}+\text{jets}}$
$t\bar{t}$+b CR	$\mathcal{D}_{t\bar{t}+b}/\mathcal{D}_{t\bar{t}+\text{jets}}$
$t\bar{t}+\geq 2b$ CR	$\mathcal{D}_{t\bar{t}+\geq 2b}/\mathcal{D}_{t\bar{t}+\text{jets}}$

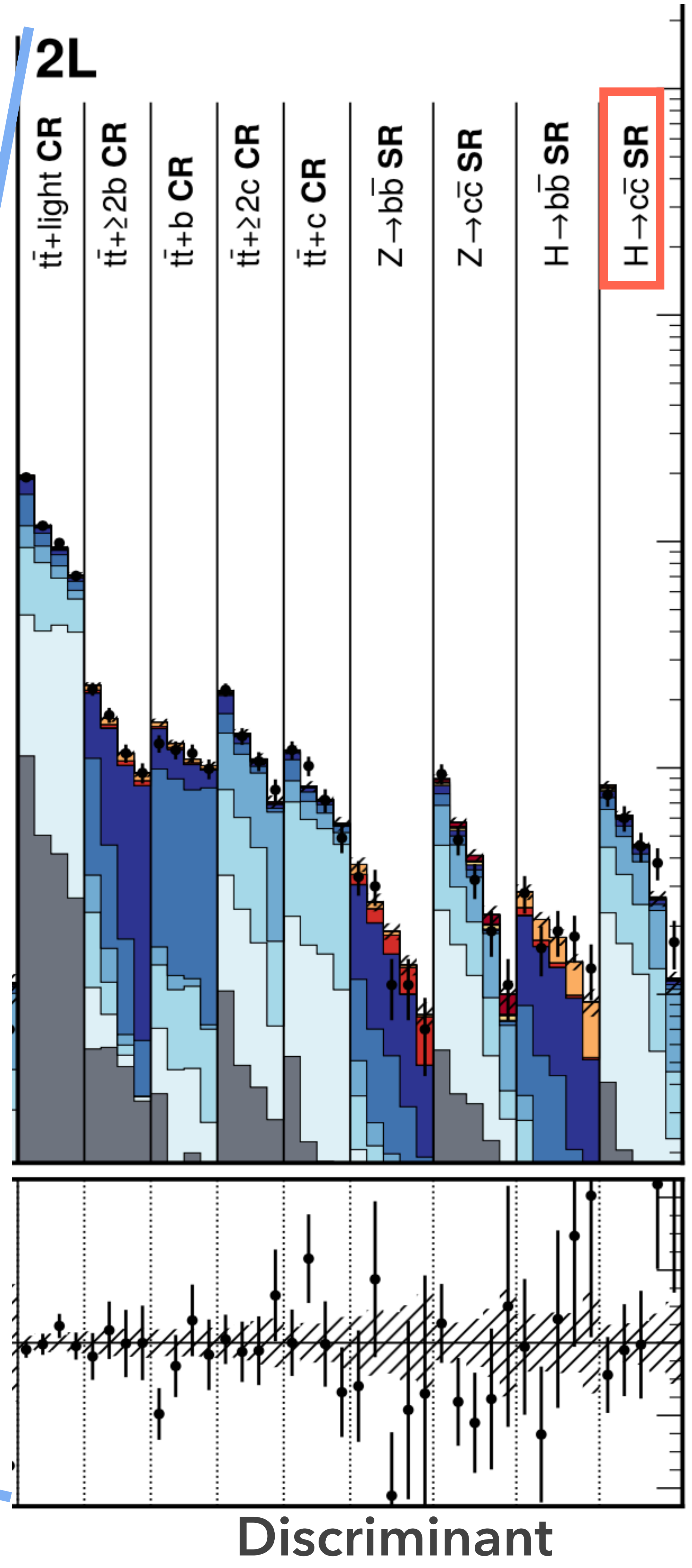
Results

First limit on $ttH(cc)$ production
Obs. (Exp.) 95% CL upper limit on $\sigma(t\bar{t}H)B(H\to c\bar{c})$ of 0.11(0.13) pb

Most stringent individual limit on κ_c
Obs. (Exp.) $\kappa_c < 3$ (3.5) at $\kappa_b = 1$

Highest individual sensitivity to $ttH(bb)$ production
Observed with 4.4σ significance



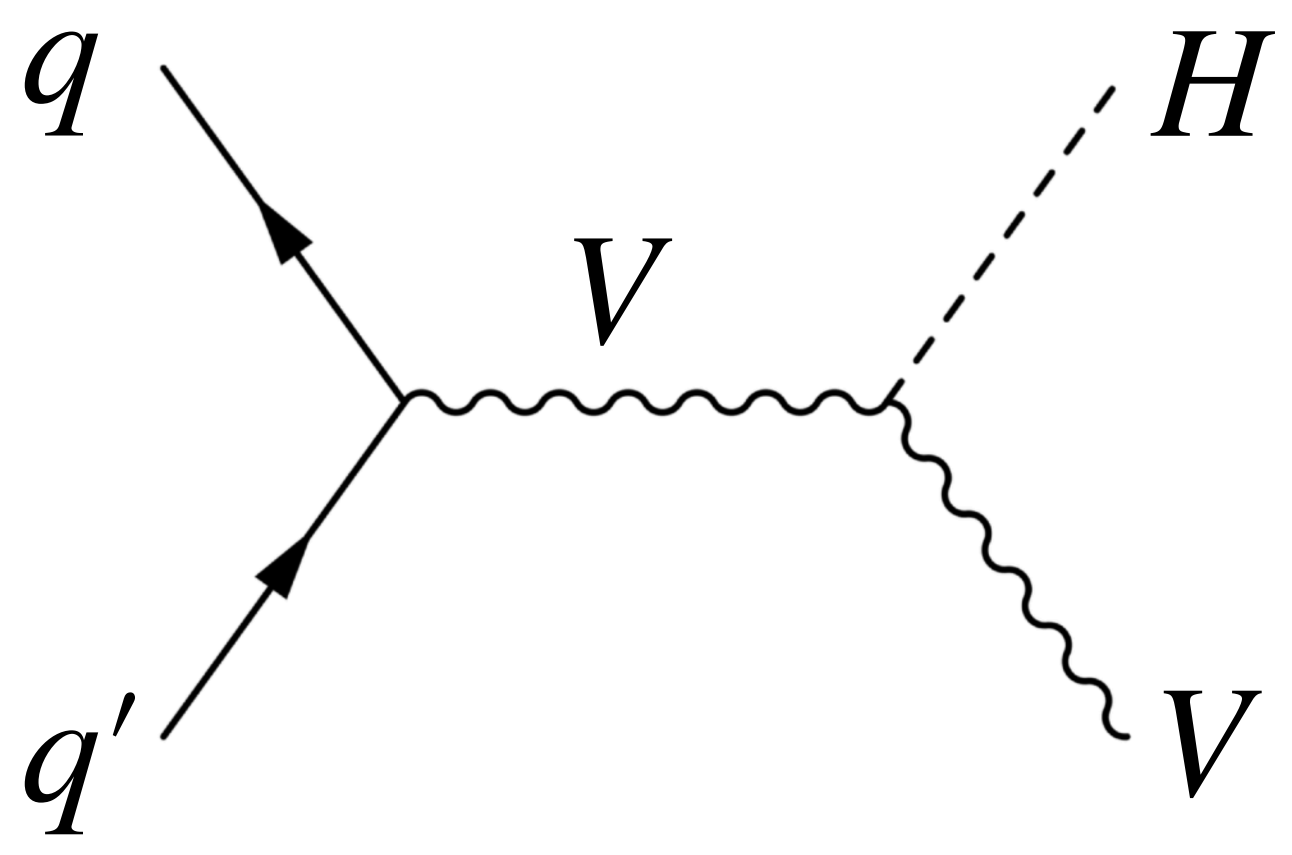


$VH(b\bar{b})$ SMEFT interpretation

$VH(b\bar{b})$ SMEFT interpretation

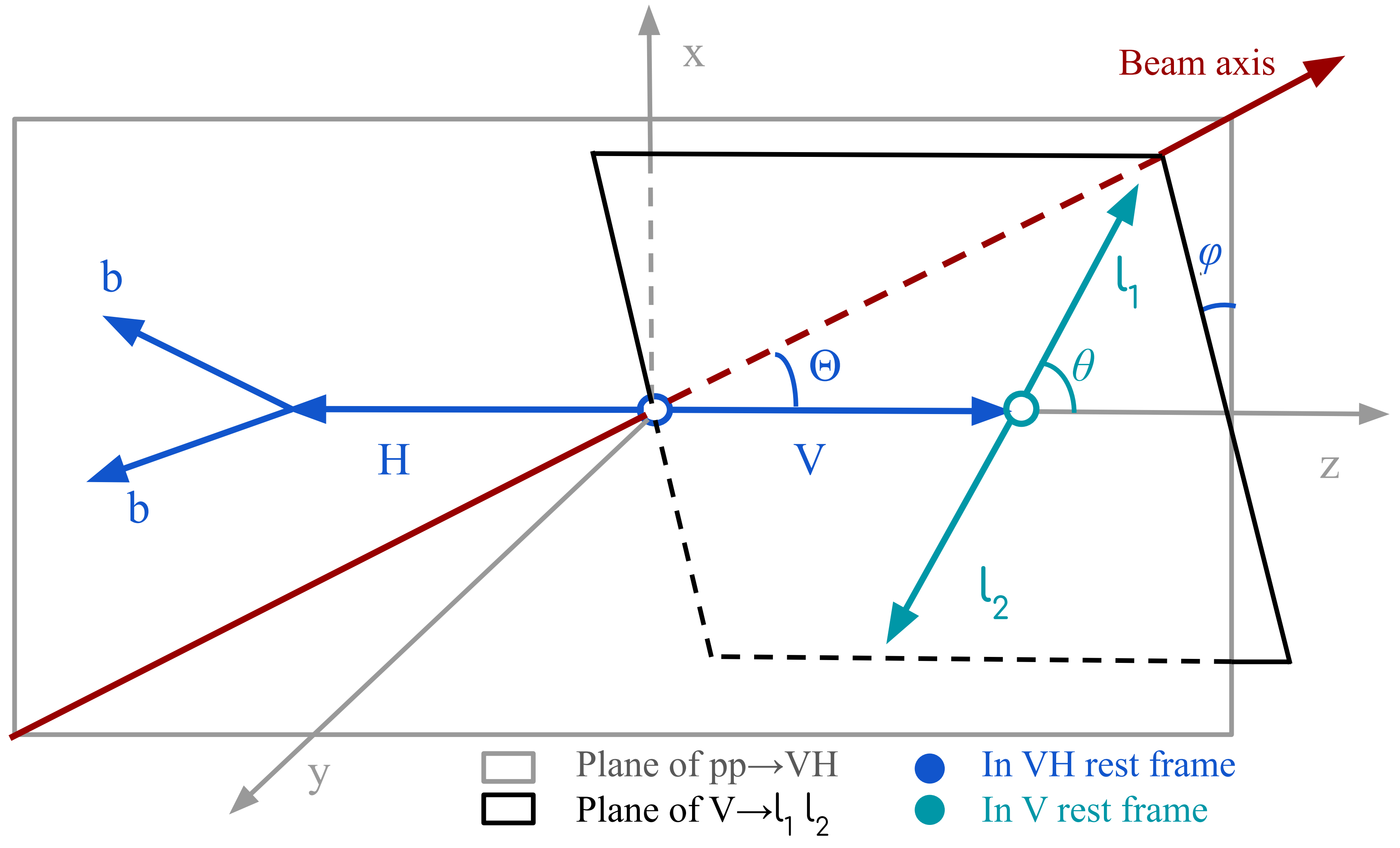
JHEP 03 (2025) 114

Overview



Measurement to probe dimension-six SMEFT operator coefficients contributing to $H \rightarrow b\bar{b}$ production in association with a vector boson V (V= W, Z).

- Six Wilson coefficients probed by the analysis: $c^{(1)}_{Hq}$, $c^{(3)}_{Hq}$, c_{Hu}, c_{Hd}, g_2^{ZZ}, g_4^{ZZ} (see backup) See Suman's talk for EFT discussion
- Targeting leptonic decays of V bosons: $Z \rightarrow \nu\nu$, $W \rightarrow l\nu$, $Z \rightarrow ll$
- Resolved and boosted categories according to reconstruction mode of the Higgs



Towards the ultimate differential SMEFT analysis (S. Banerjee et al., 2020)

First time the complete event information (including angular observables) is exploited to build optimal discriminant.

- Train a ML model to regress the likelihood ratio from simulation
- Handle many observables without loss of information

$VH(b\bar{b})$ SMEFT interpretation

JHEP 03 (2025) 114

Simulation based inference

The likelihood ratio q_θ is the optimal test statistic to discriminate between two EFT hypothesis θ_0 and θ (Neyman-Pearson).

Dependence of the observables on the theory parameters factorizes into:

- Parton-level process (contains the full dependence on the EFT coefficients)
- Part describing parton shower, hadronization, and detector effects

By minimizing the cross-entropy loss w.r.t. $g(x|\theta)$ to regress the joint likelihood ratio $\hat{r}(\mathbf{x},\mathbf{z}|\theta,\theta_0)$, one can recover $r(\mathbf{x}|\theta,\theta_0)$ from the minimizer $g^*(x|\theta)$.

The minimization is performed using Boosted Information Tree (BIT).

- Node splitting criterion maximizes the Fisher information

Tree boosting for learning EFT parameters (S. Chatterjee et al., 2022)

x: set of all observables after shower, detector, and reconstruction
z: parton level four-momenta
θ: theory parameters (e.g. Wilson coefficients)

$$r(\mathbf{x}|\theta,\theta_0) = \frac{\mathbf{p}(\mathbf{x}|\theta)}{\mathbf{p}(\mathbf{x}|\theta_0)}$$

Intractable

$$p(x|\theta) = \int dz\; p(x,z|\theta) = \int dz p(x|z) p(z|\theta)$$

$$L[g(x|\theta)] = -\left(\int dxdz \frac{\sigma(\theta)}{\sigma(\theta_0)} \frac{p(x,z|\theta)}{p(x,z|\theta_0)} \log(1-g(x|\theta)) + \log(g(x|\theta))\right)$$

Known from simulation

$$\hat{r}(\mathbf{x},\mathbf{z}|\theta,\theta_0) = \frac{\mathbf{p}(\mathbf{x},\mathbf{z}|\theta)}{\mathbf{p}(\mathbf{x},\mathbf{z}|\theta_0)} = \frac{\mathbf{p}(\mathbf{z}|\theta)}{\mathbf{p}(\mathbf{z}|\theta_0)}$$

$$g^*(x|\theta) = \frac{1}{1 + r(x|\theta,\theta_0)}$$

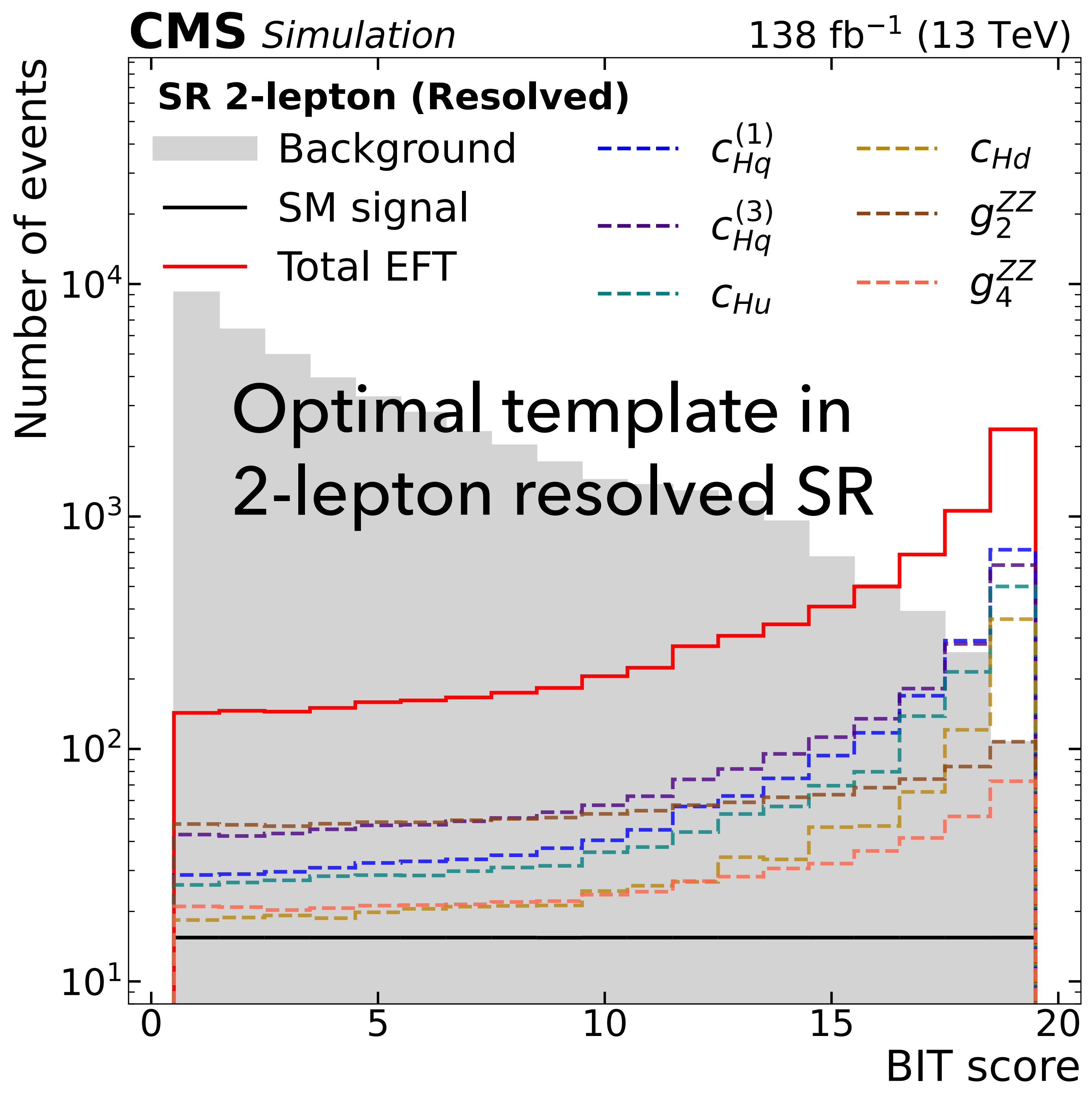
$VH(b\bar{b})$ SMEFT interpretation

BIT template

JHEP 03 (2025) 114

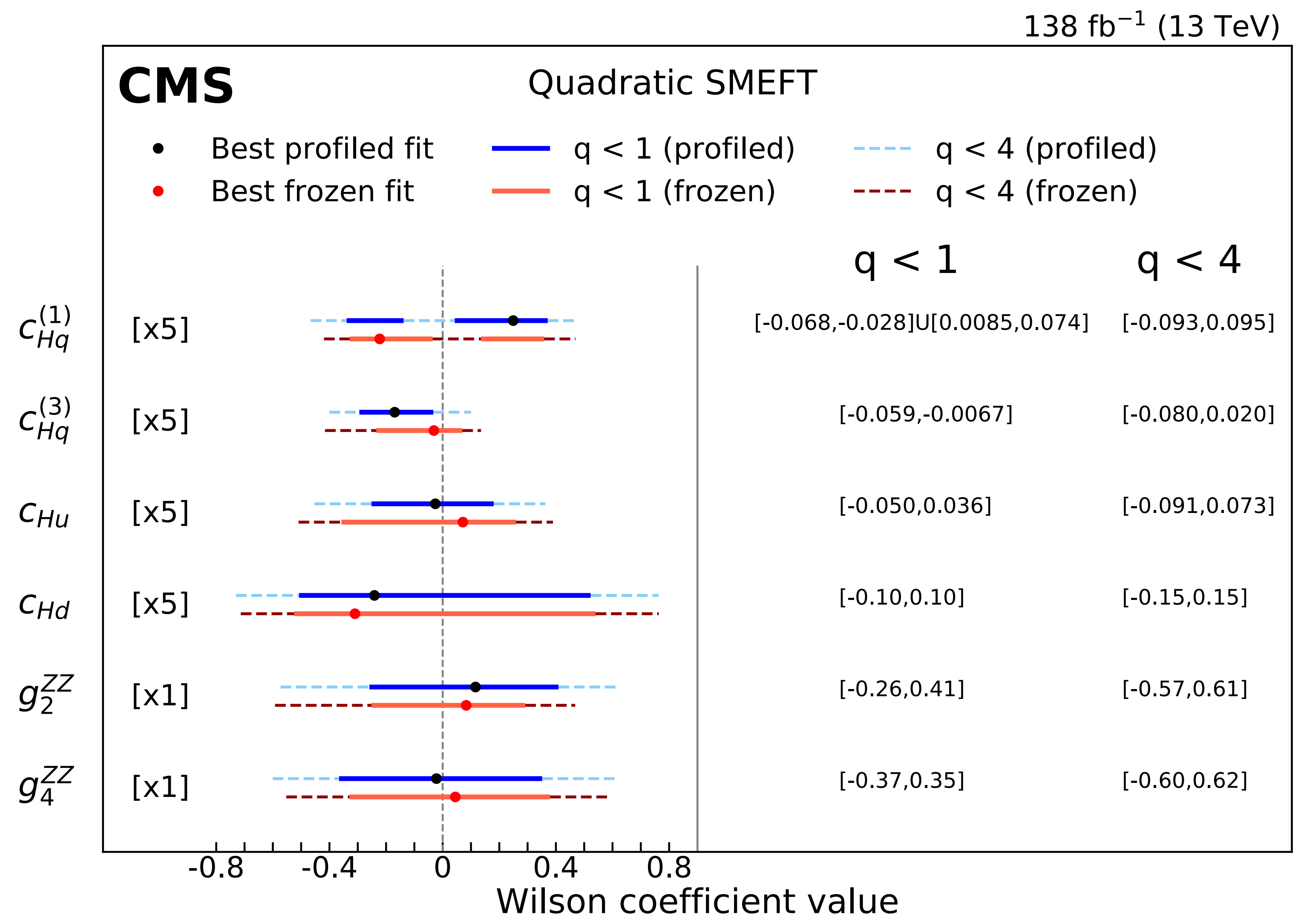
The ratio $r(\mathbf{x}\,|\,\theta, \theta_0)$ will not be the optimal observable for another hypothesis θ'.

- To retain optimal separation power, a different BIT training would be required at each point in WC space
- Optimal template $r(\mathbf{x}\,|\,\theta^*, \theta_0)$ found via Bayesian optimization to maximize the expected sensitivity to the WCs



Binned BIT template used for signal extraction in the SRs, while separate CRs are used to constrain backgrounds.

- Background peaks at low BIT score → separates bkg from signal
- SMEFT effects peak at high BIT score → correctly identifies SMEFT-sensitive events



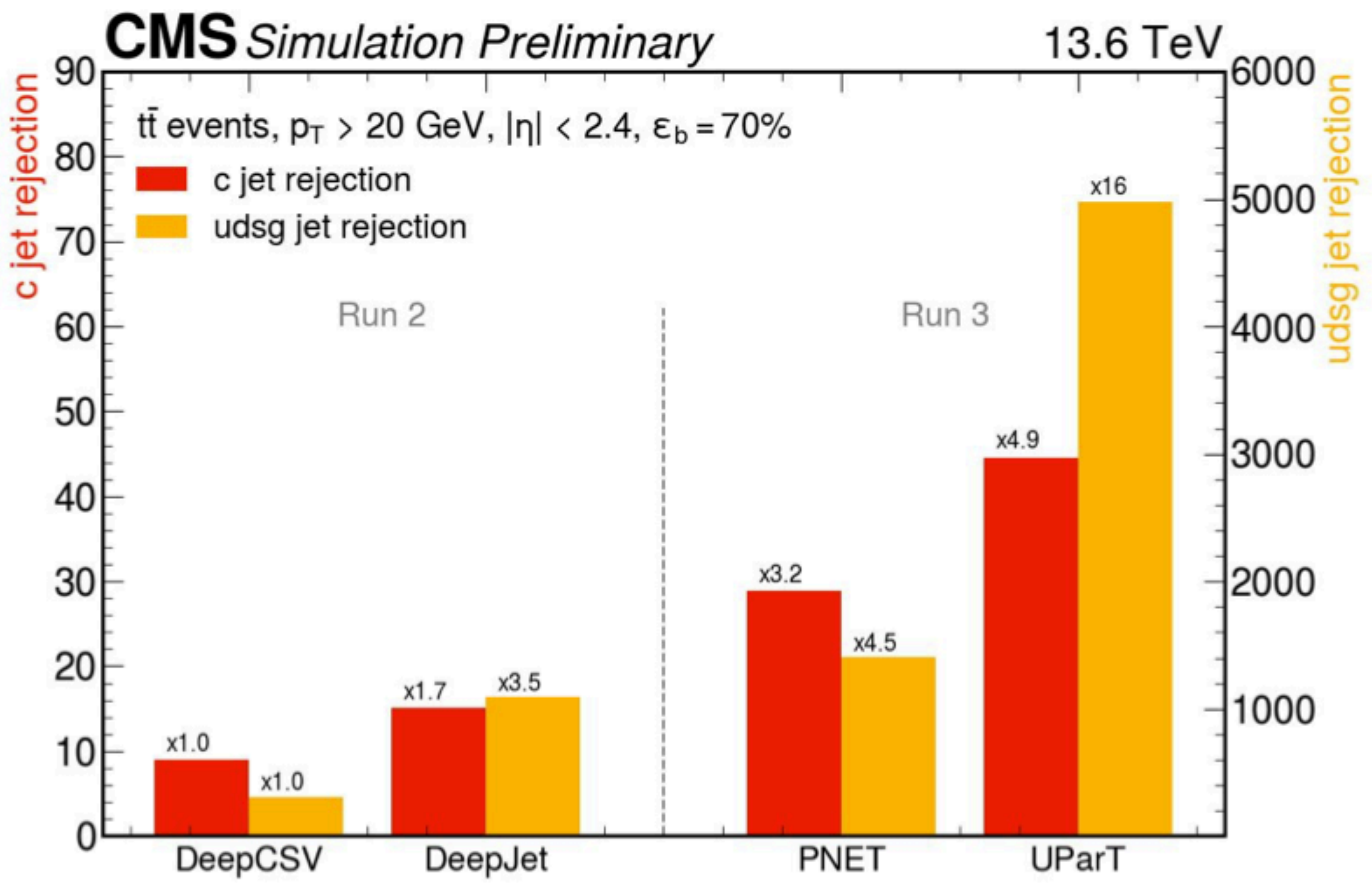
Most comprehensive SMEFT analysis in this channel!

See Suman's talk for discussion of EFT results

Conclusion & outlook

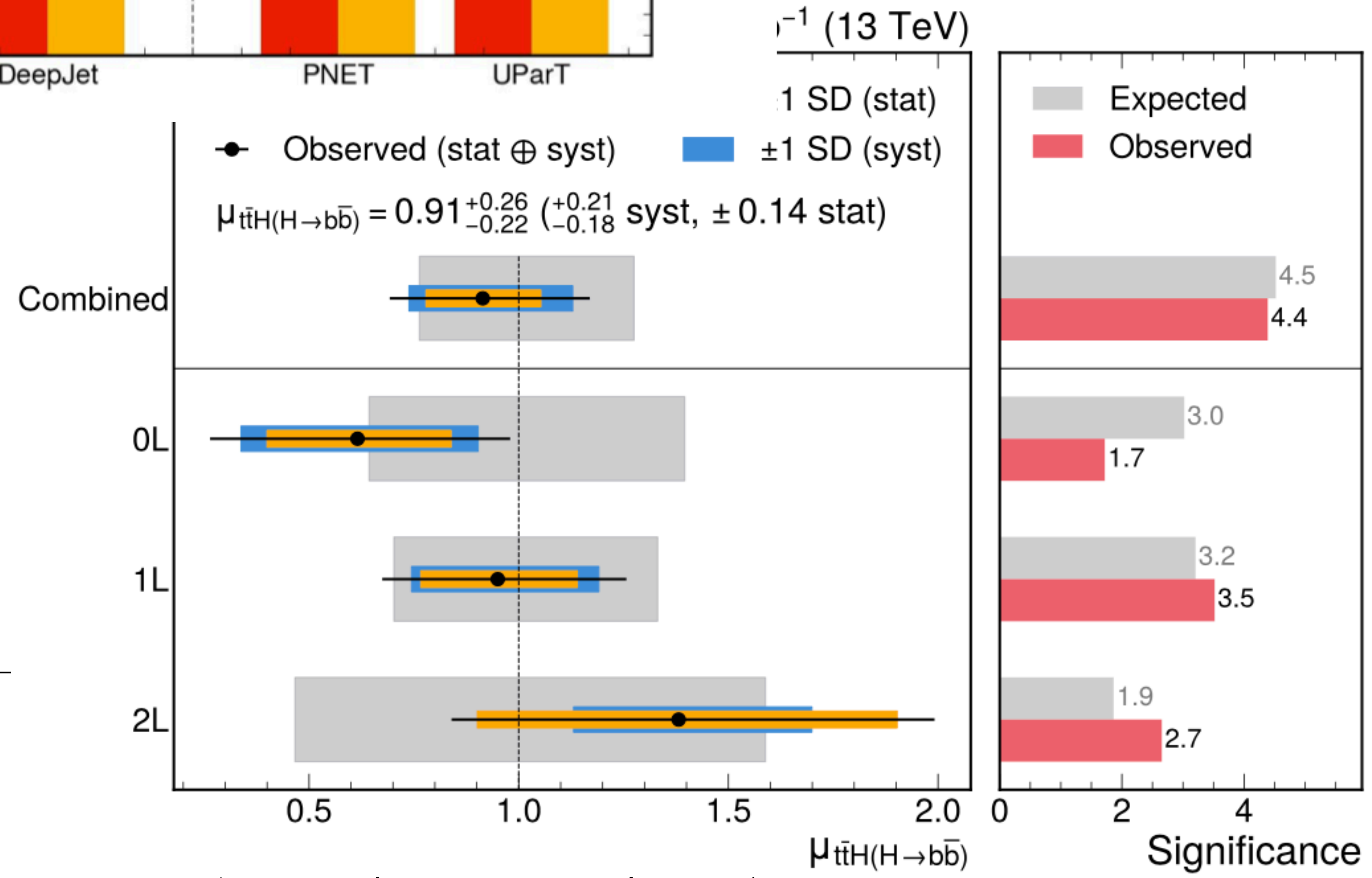
PNet and Transformer based algorithms show state-of-the-art heavy flavor and boosted jet tagging performance

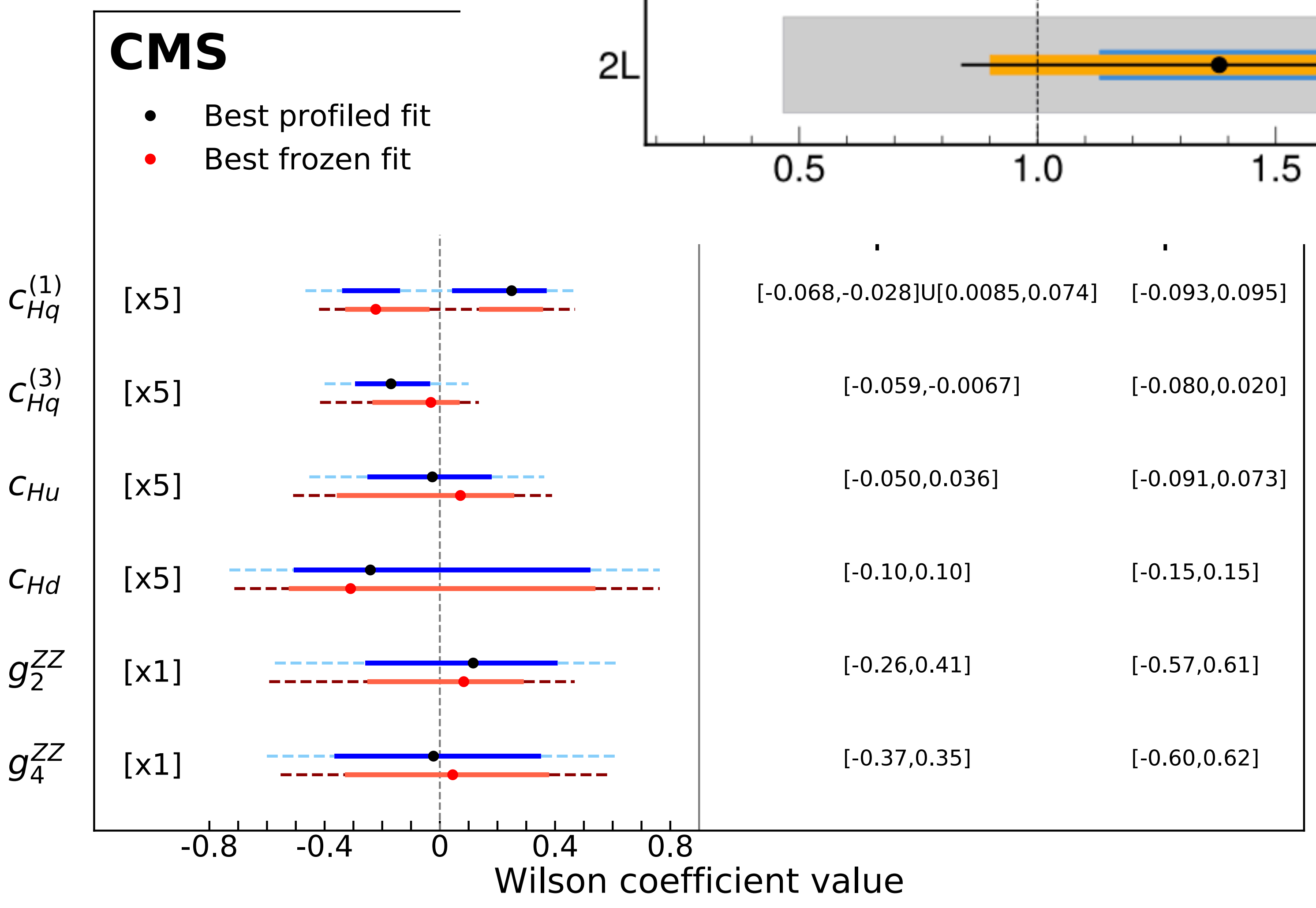
- Towards a unified tagger strategy: multiple tasks and comprehensive classification categories
- Stay tuned for WIP Run 3 analyses using these methods!



Beyond jet tagging:

- Towards simulation based inference
- Adversarial DNN for signal agnostic event classification (see Benedetta's talk), systematics aware NN (backup), ABCDNet, …





Machine learning is now ubiquitous in CMS analyses, allowing us to perform analyses not possible before and directly impacting our physics reach!

Backup

Systematic aware NN training (SANNT)

CMS-MLG-23-005 (2025)

Proof of principle study showcases new method for sig. vs. bkg discrimination that takes into account systematic uncertainties

- Evaluated on measurement of $H \rightarrow \tau\tau$ produced via ggF and VBF

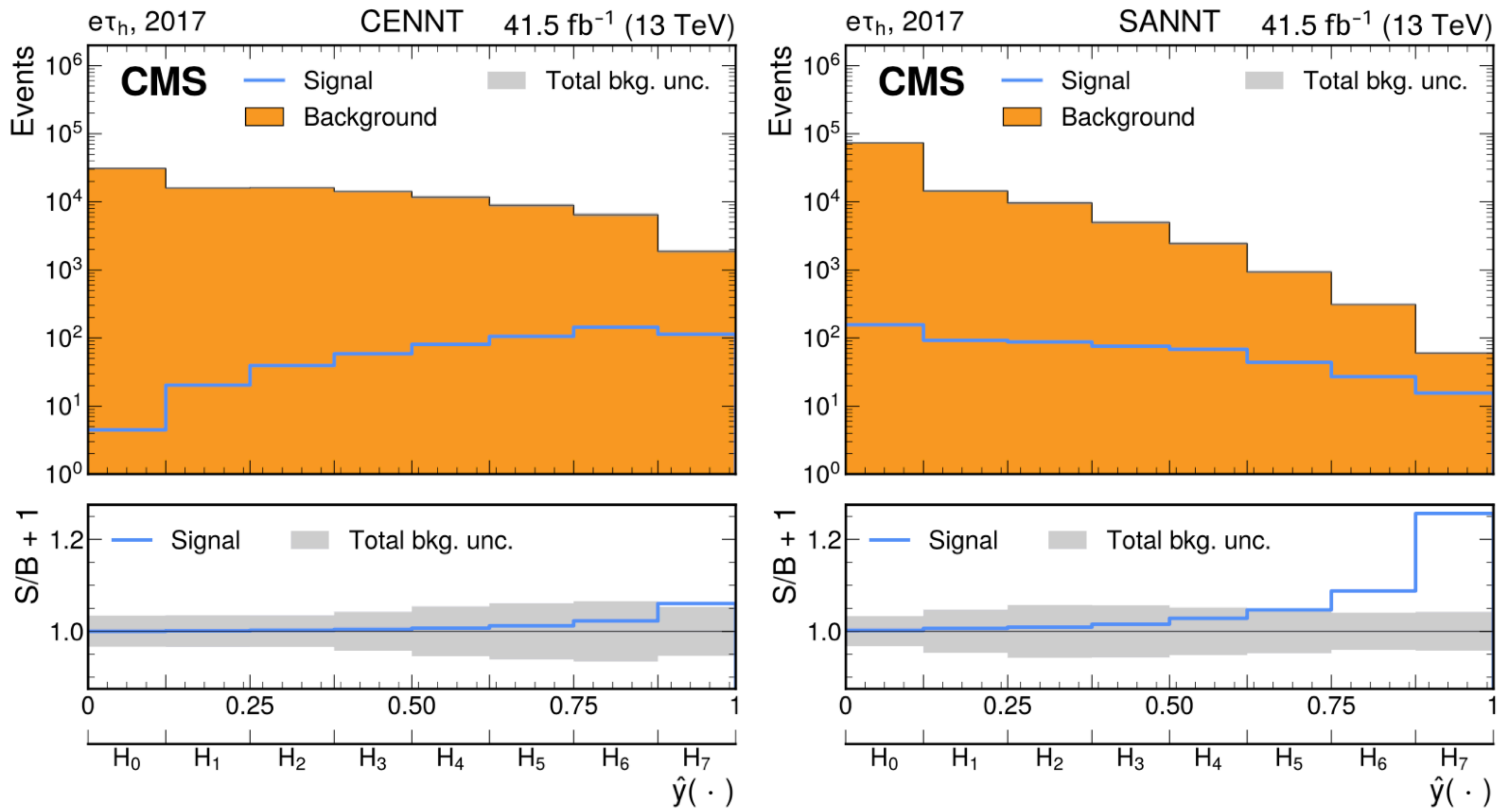
Compared to CENNT: Standard NN method based on categorical cross entropy loss (equivalent to negative log-likelihood of multinomial distribution)

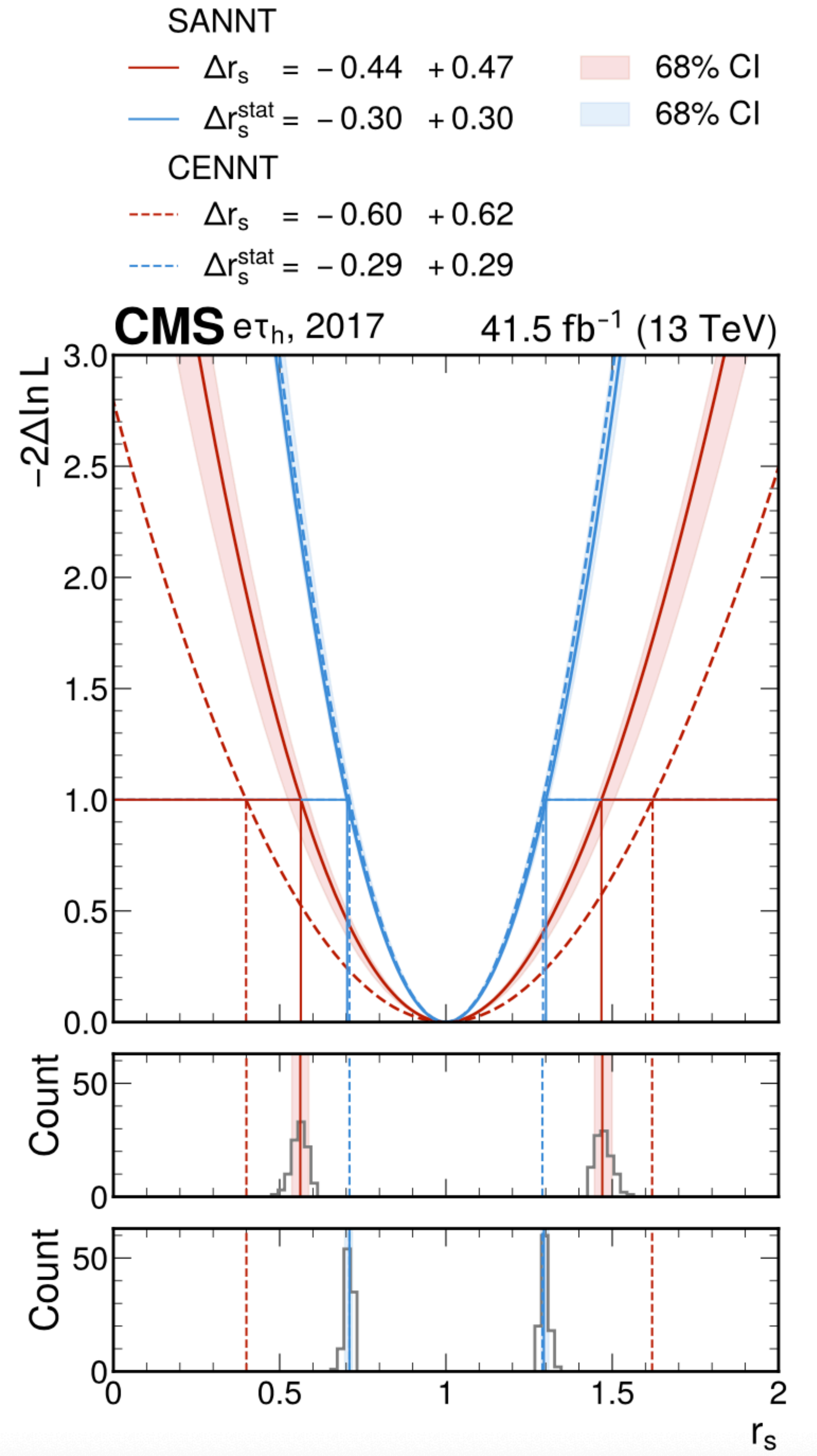
- Optimal discriminator when only statistical uncertainties are present

Signal strength is extracted from the output of the NN

- Largest background uncertainty contributions move away from signal-enriched region

Improvements of 12% (16%) in the uncertainty in the signal strengths for ggF (VBF) compared to CENNT





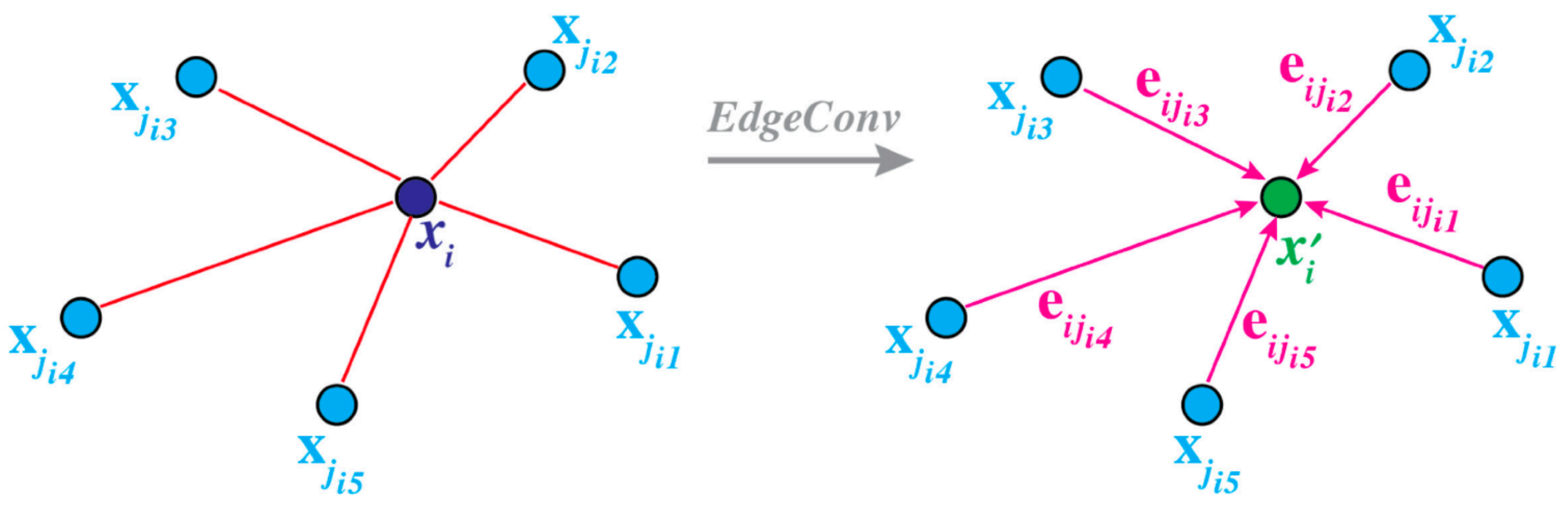
Backup - ParticleNet

Dynamic graph CNN for learning on point clouds. (Wang et al., 2019)

Jet tagging via particle clouds (H. Qu and L. Gouskos, 2020)

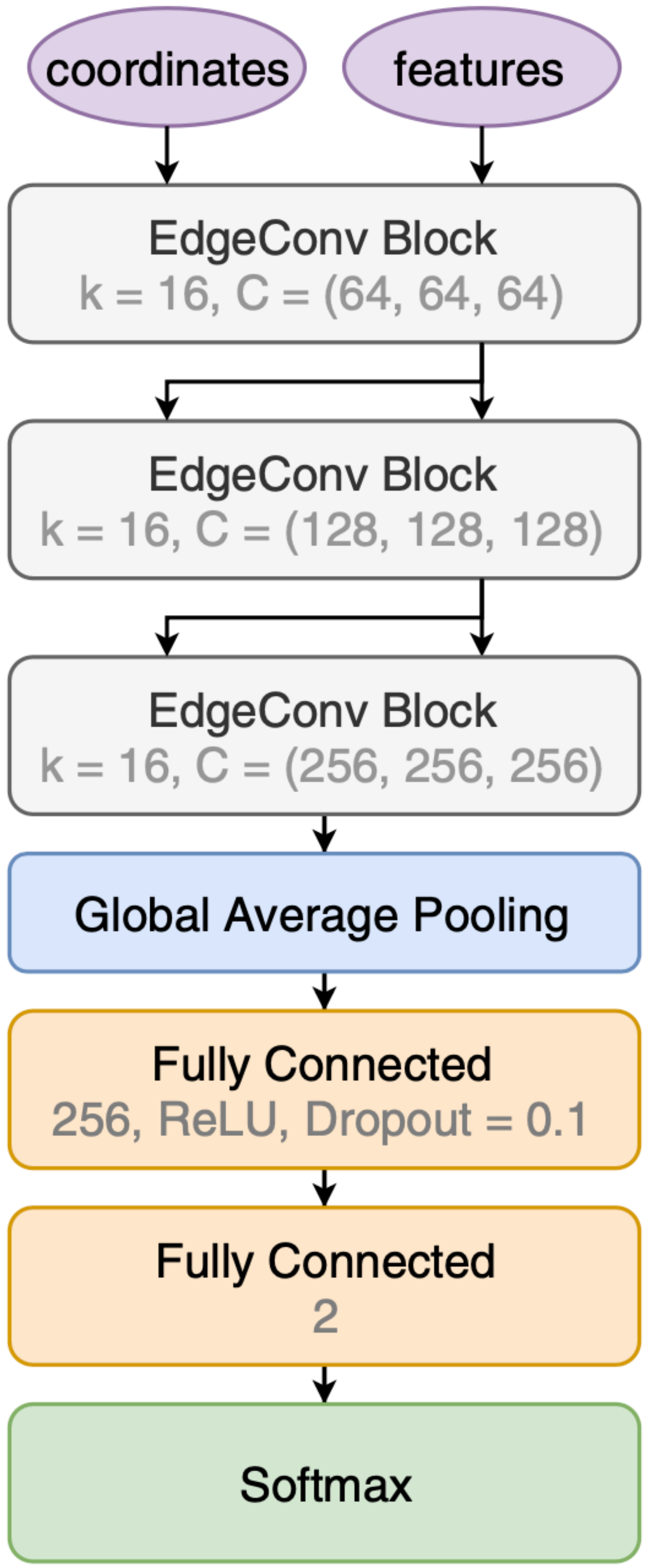
ParticleNet is a Dynamic Graph Convolutional Neural Network (DGCNN) used for jet tagging.

- Treat jet as unordered sets of particles ("particle clouds")



Based on edge convolution (EdgeConv) blocks:

- Construct a graph by representing the points as vertices, and defining the edges as the distance to its k nearest neighbors
- Convolution operation is applied on this local patch
- The graph is dynamic: built per-layer, adapting to the feature space
  - Only the first EdgeConv block uses the spatial coordinates to compute the distances, while the subsequent blocks use the learned feature vectors



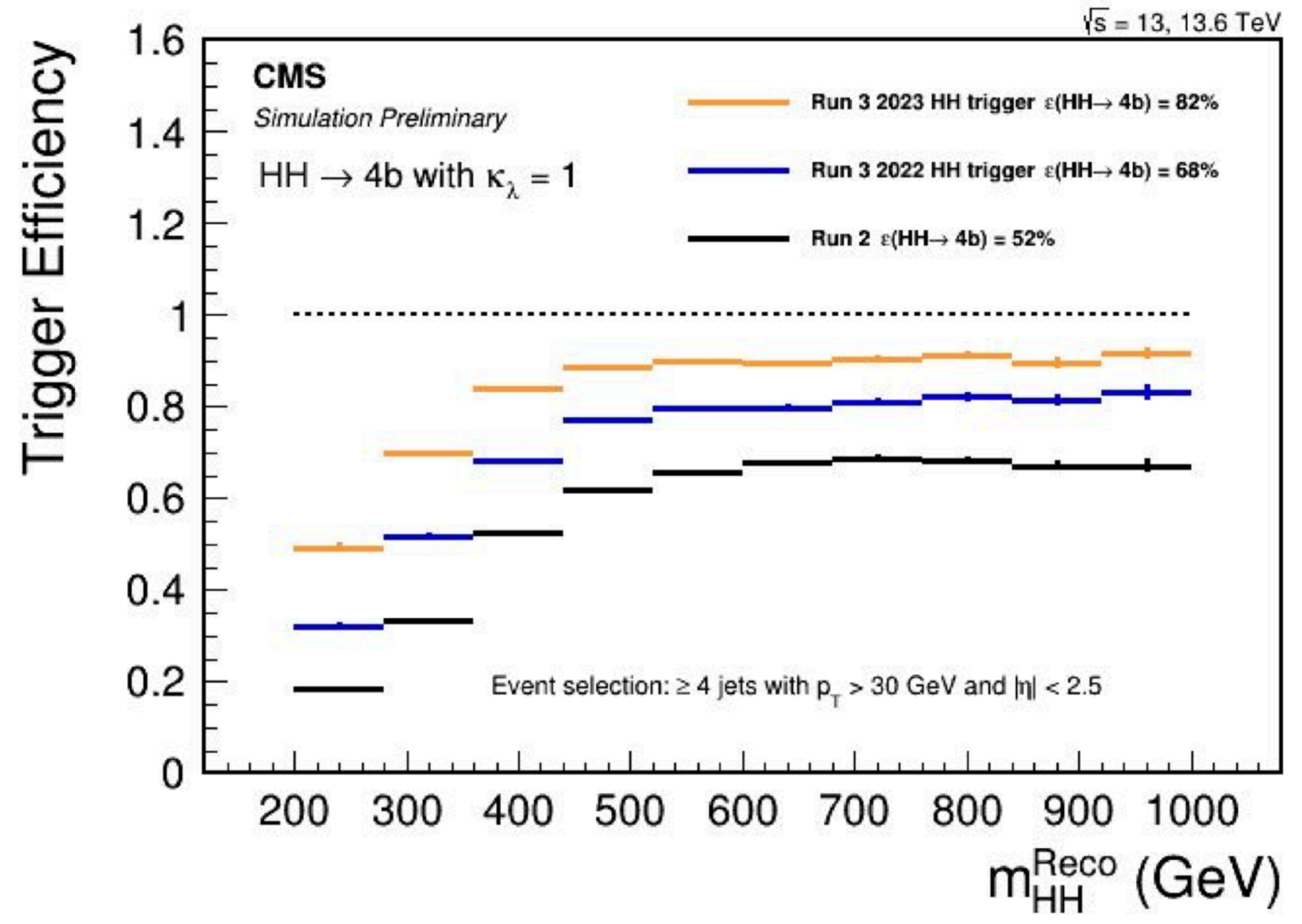
ParticleNet at the HLT

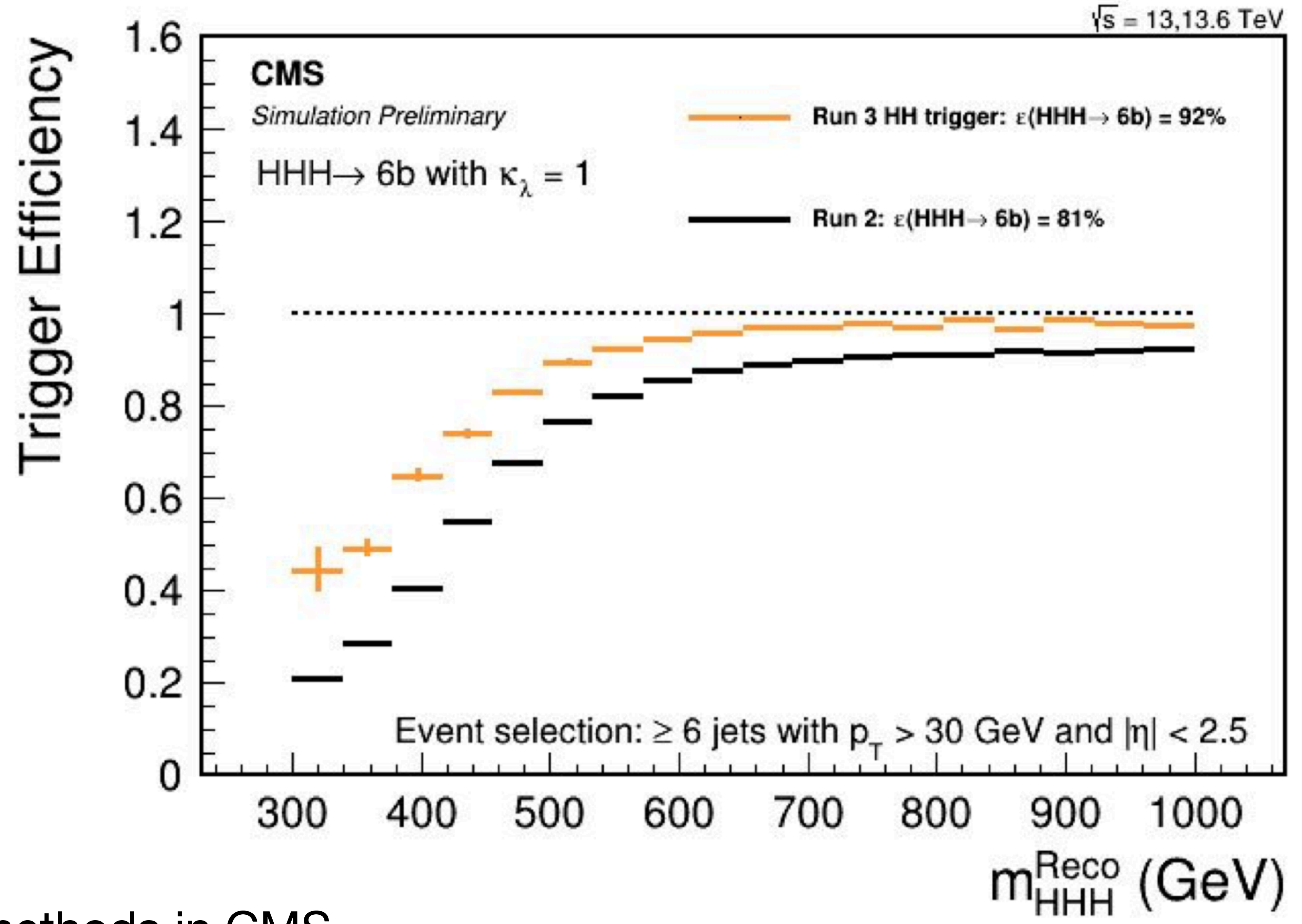
Run 3 HH and HHH triggers

Trigger	**Requirement**
2023 HH trigger	HT > 280 GeV, 4 jets with pT > 30 GeV, PNet@AK4(mean 2 highest b-tag score) > 0.55
2022 HH trigger	4 jets pT > 70, 50, 40, 35 GeV, PNet@AK4(mean 2 highest b-tag score) > 0.65

Similar improvements in $HH \to 2b2\tau$ and $HHH \to 4b2\tau$.

CMS DP-2023/021





Backup - Particle Transformer

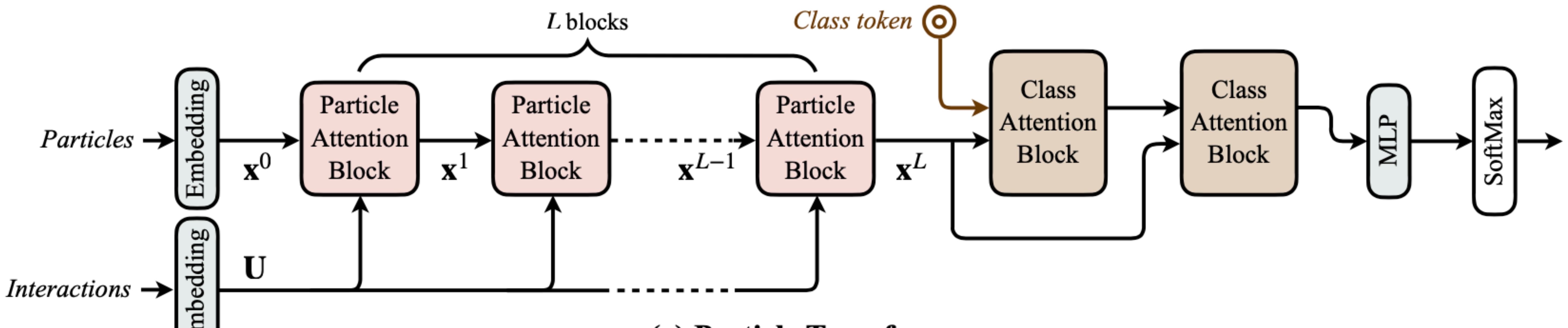
Particle Transformer for jet tagging (H. Qu, C. Li and S. Qian, 2022)

A Transformer-based model for jet tagging, incorporating pairwise particle interactions and a class token to enhance classification performance.

- Particle inputs: C features for every particle (N,C)
- Interaction input: C′ features for every pair (N,N,C′)

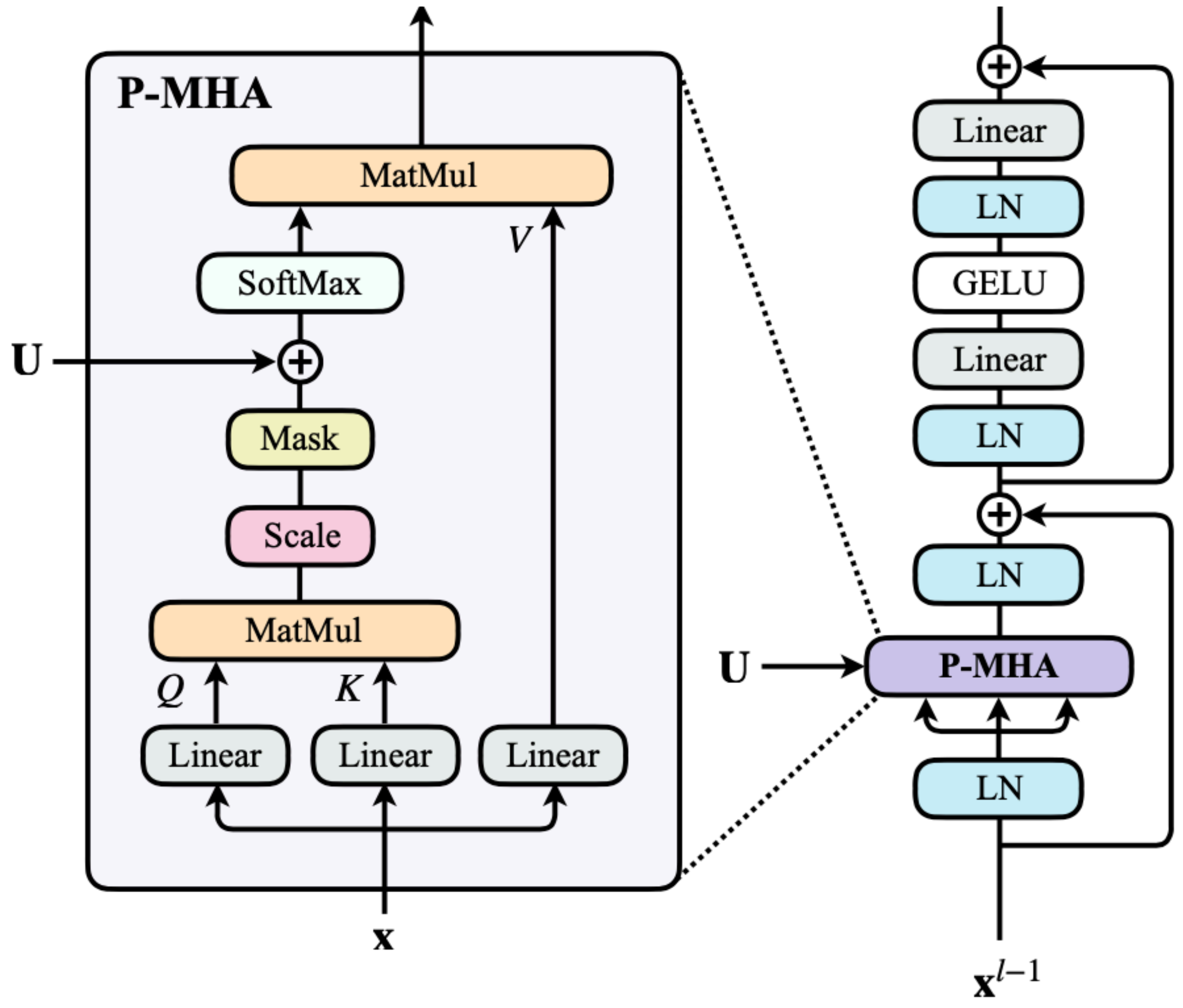
Pairwise Interactions: Introduces particle interaction biases in multi-head attention, improving sensitivity to jet substructure.

Class Token: Used as the query in dedicated Class Attention layers, attending to concatenated class and particle embeddings (keys/values) to produce an updated class token optimized for classification.



(a) Particle Transformer

$$\Delta = \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2},$$
$$k_\mathrm{T} = \min(p_{\mathrm{T},a}, p_{\mathrm{T},b})\Delta,$$
$$z = \min(p_{\mathrm{T},a}, p_{\mathrm{T},b})/(p_{\mathrm{T},a} + p_{\mathrm{T},b}),$$
$$m^2 = (E_a + E_b)^2 - \|\mathbf{p}_a + \mathbf{p}_b\|^2,$$



(b) Particle Attention Block

$$\text{P-MHA}(Q, K, V) = \text{SoftMax}(QK^T/\sqrt{d_k} + \mathbf{U})V,$$

Backup - $VH(b\bar{b})$

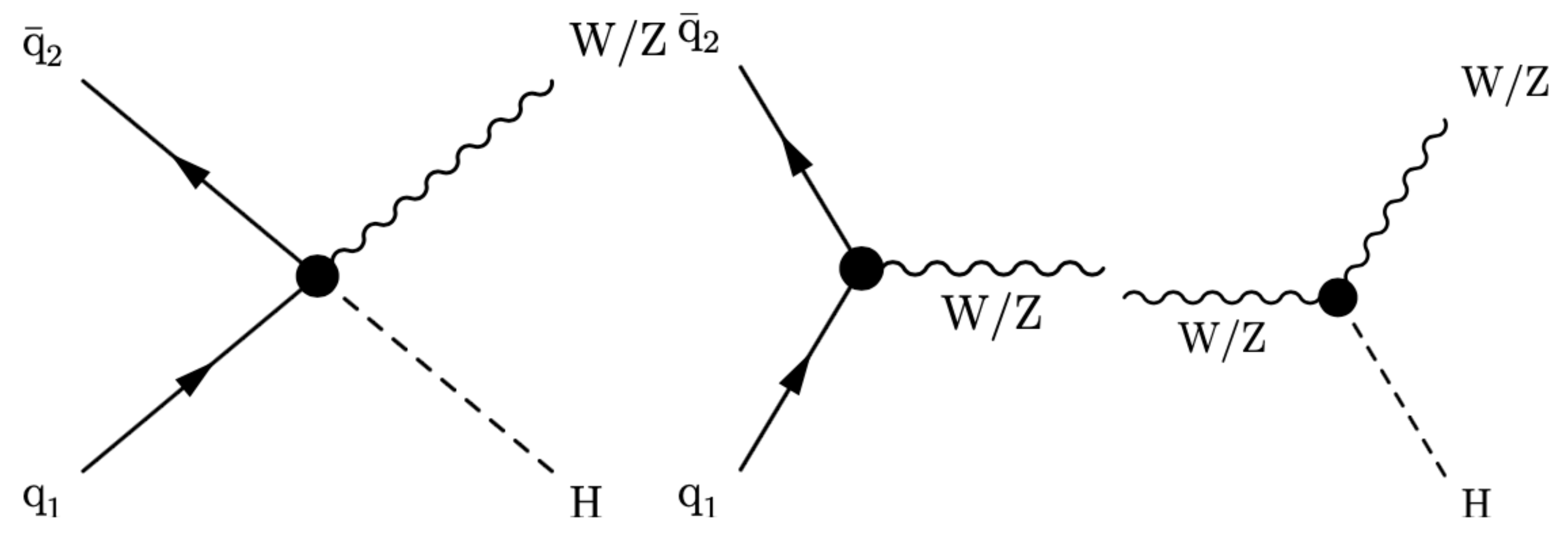
JHEP 03 (2025) 114

EFT parameters

Operator	Definition	Wilson coefficient	Operator	Definition	Wilson coefficient
$\mathcal{O}^{(1)}_{Hq}$	$iH^\dagger \overleftrightarrow{D}_\mu H \bar{q}_L \gamma^\mu q_L$	$c^{(1)}_{Hq}$	$\mathcal{O}_{HWB}$	$H^\dagger \sigma^a H W^a_{\mu\nu} B^{\mu\nu}$	c_{HWB}
$\mathcal{O}^{(3)}_{Hq}$	$iH^\dagger \sigma^a \overleftrightarrow{D}_\mu H \bar{q}_L \sigma^a \gamma^\mu q_L$	$c^{(3)}_{Hq}$	$\mathcal{O}_{H\widetilde{W}B}$	$H^\dagger \sigma^a H W^a_{\mu\nu} \widetilde{B}^{\mu\nu}$	$c_{H\widetilde{W}B}$
$\mathcal{O}_{Hu}$	$iH^\dagger \overleftrightarrow{D}_\mu H \bar{u}_R \gamma^\mu u_R$	c_{Hu}	$\mathcal{O}_{HW}$	$(H^\dagger H) W^a_{\mu\nu} W^{a\mu\nu}$	c_{HW}
$\mathcal{O}_{Hd}$	$iH^\dagger \overleftrightarrow{D}_\mu H \bar{d}_R \gamma^\mu d_R$	c_{Hd}	$\mathcal{O}_{H\widetilde{W}}$	$(H^\dagger H) W^a_{\mu\nu} \widetilde{W}^{a\mu\nu}$	$c_{H\widetilde{W}}$
$\mathcal{O}_{HD}$	$(H^\dagger D^\mu H)^* (H^\dagger D_\mu H)$	c_{HD}	$\mathcal{O}_{HB}$	$(H^\dagger H) B_{\mu\nu} B^{\mu\nu}$	c_{HB}
$\mathcal{O}_{H\Box}$	$(H^\dagger H)\Box(H^\dagger H)$	$c_{H\Box}$	$\mathcal{O}_{H\widetilde{B}}$	$(H^\dagger H) B_{\mu\nu} \widetilde{B}^{\mu\nu}$	$c_{H\widetilde{B}}$

Dimension-six operators in the Warsaw basis affecting VH production at LO:

- Current operators $\mathcal{O}^{(1)}_{Hq}$, $\mathcal{O}^{(3)}_{Hq}$, $\mathcal{O}_{Hu}$, $\mathcal{O}_{Hd}$ introduce four point interactions (left) and determine increase in production cross section
- $\mathcal{O}_{HD}$, $\mathcal{O}_{HWB}$: together with current operators, describe all relevant V-fermion coupling modifications (middle)
- Gauge coupling operators $\mathcal{O}_{HW}$, $\mathcal{O}_{HWB}$, $\mathcal{O}_{HB}$ (and their CP conjugates), together with $\mathcal{O}_{HD}$ and $\mathcal{O}_{H\Box}$, modify the H-V coupling (right)



After a rotation of the mass eigenstate basis, the coefficients c_{HW}, c_{HB}, and c_{HWB} can be expressed as:

$$g_2^{ZZ} = -2\frac{v^2}{\Lambda^2}\left(s_w^2 c_{HB} + c_w^2 c_{HW} + s_w c_w c_{HWB}\right),$$

$$g_2^{Z\gamma} = -2\frac{v^2}{\Lambda^2}\left(s_w c_w (c_{HW} - c_{HB}) + \frac{1}{2}(s_w^2 - c_w^2) c_{HWB}\right),$$

$$g_2^{\gamma\gamma} = -2\frac{v^2}{\Lambda^2}\left(c_w^2 c_{HB} + s_w^2 c_{HW} - s_w c_w c_{HWB}\right),$$

The corresponding degrees of freedom in the WH production are fully constrained in this change of basis, i.e. constraints g_2^{WW} and g_4^{WW} can be obtained from the ones on g_2^{ZZ} and g_4^{ZZ}.

The anomalous couplings $g_2^{Z\gamma}$ and $g_2^{\gamma\gamma}$ are not expected to be constrained by WH and ZH production, and are therefore not considered for the analysis.

The coefficients $\mathcal{O}_{HD}$ and $\mathcal{O}_{H\Box}$ are also dropped due to their small effect.

Backup - $VH(b\bar{b})$

EFT weights

$$p(z|\theta) = \frac{1}{\sigma(\theta}\frac{d\sigma(\theta)}{dz}$$

$$p(x|\theta) = \int dz p(x|z)p(z|\theta)$$

The cross section has at most a quadratic dependence on the Wilson coefficients (WCs):

$$d\sigma \sim |\mathcal{M}_{SM} + \sum_i \frac{\theta_i}{\Lambda^2}\mathcal{M}_{6,i}|^2 \sim |\mathcal{M}_{SM}|^2 + \sum_{i=1}^{n} \frac{\theta_i}{\Lambda^2} 2Re(\mathcal{M}^\dagger_{SM}\mathcal{M}_{6,i}) + \sum_{i=1}^{n} \frac{\theta_i^2}{\Lambda^4}|\mathcal{M}_{6,i}|^2 + \sum_{i=1}^{n}\sum_{j,j>i}^{n} \frac{\theta_i\theta_j}{\Lambda^4}\mathcal{M}_{6,i}\mathcal{M}_{6,j}$$

Given enough basis points, one can solve the system of equations to obtain the quadratic dependence of the differential cross section on the couplings at each point in the phase space.

- Each simulated event is augmented with a weight polynomial function $w_i(\theta)$:
- Augmented data set $= \{w_i(\theta), x_i, z_i\}_{i=1}^{N}$

$$\int_{\Delta z} p(z|\theta)dz \approx \frac{1}{N}\sum_{z_i \in \Delta z} w_i(\theta)$$

Because x and z are jointly available, the joint likelihood $p(x, z|\theta)$ is known from simulation.

Assuming no BSM effects in the strong interactions driving the parton showering, the factorization of polynomial dependency is valid also

$$p(x|\theta) = p_0(x) + \sum_{1\le i\le M} p_i^l(x)\frac{\theta_i}{\Lambda^2} + \sum_{1\le i\le M} p_i^q(x)\frac{\theta_i^2}{\Lambda^4} + \sum_{1\le i<j\le M} p_{i,j}^m(x)\frac{\theta_i\theta_j}{\Lambda^4}$$

Backup - $VH(b\bar{b})$

Training

Making use of the polynomial dependency in θ of the likelihood ratio

$$R(\mathbf{x}|\theta,\theta_{\mathbf{0}}) = \frac{\sigma(\theta)\mathbf{p}(\mathbf{x}|\theta)}{\sigma(\theta_{\mathbf{0}})\mathbf{p}(\mathbf{x}|\theta_{\mathbf{0}})} \longrightarrow R(x|\theta,\theta_0) = 1 + \sum_{1\le i\le M}(\theta_i - \theta_0)R_i(x) + \sum_{1\le i\le j\le M}\frac{1}{2}(\theta_i-\theta_0)(\theta_j-\theta_0)R_{i,j}(x)$$

$$R_i(x) = \frac{\partial_i(\sigma(\theta)p(x|\theta))|_{\theta,\theta_0}}{\sigma(\theta_0)p(x|\theta_0)} \qquad R_{i,j}(x) = \frac{\partial_i\partial_j(\sigma(\theta)p(x|\theta))|_{\theta,\theta_0}}{\sigma(\theta_0)p(x|\theta_0)}$$

Minimizing the loss using $\hat{R}(x,z)$ as the target function, will yield a minimizing function equal to $R(x)$.

$$\hat{R}_i(x,z) = \frac{\partial_i(\sigma(\theta)p(x,z|\theta))|_{\theta,\theta_0}}{\sigma(\theta_0)p(x,z|\theta_0)} \qquad \hat{R}_{i,j}(x,z) = \frac{\partial_i\partial_j(\sigma(\theta)p(x,z|\theta))|_{\theta,\theta_0}}{\sigma(\theta_0)p(x,z|\theta_0)}$$

Each component in this expansion corresponds to a BDT training with target function given by $\hat{R}_i(x,z)$ or $\hat{R}_{i,j}(x,z)$.
Given N Wilson coefficients, a total number of N_T trainings is performed to regress each component.

$$N_T = 2N + \frac{1}{2}N(N-1)$$