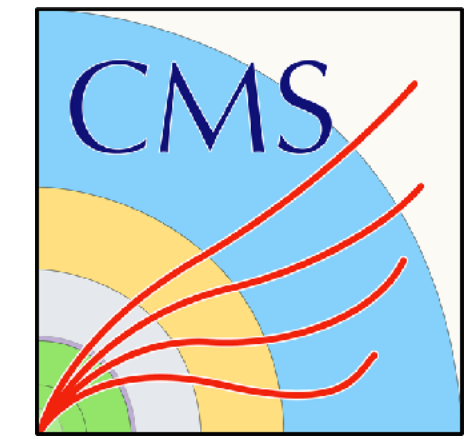




Squeezing the Standard Model: ML-driven top & Higgs physics at the LHC

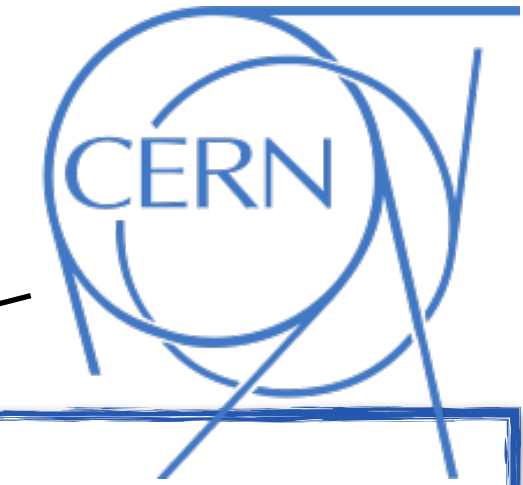
Melissa Quinnan

Color Meets Flavor Symposium

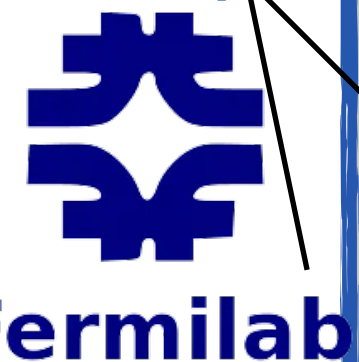
July 15, 2026

Initial Interactions...

- My research: **Innovate AI algorithms** at all stages of HEP experimental pipeline in order to **drive sensitivity in top and Higgs measurements**



- CERN Fellow (Experimental Physics) *January 2027*
- Postdoc at University of California San Diego (P.I. Javier Duarte) (2022-present)
- Fermilab LHC Physics Center Distinguished Researcher Junior Fellow (2026)
- RCSA Postdoctoral Fellow (2025-present)
- Schmidt AI in Science Postdoctoral Fellow (2023-2025)
- Accelerated AI Algorithms for Data-Driven Discovery (A3D3) NSF Institute Emerging Scientist Leadership Award (2023)
- CMS Leadership: Current L1 Trigger Offline Software Convener (senior "L2" manager), former Machine Learning Knowledge Subgroup Convener (junior "L3" manager)
- Ph.D. University of California Santa Barbara (P.I. Joseph Incandela) (2022)
- NSF Graduate Research Fellowship Program Fellow
- 2015 CERN Summer Student Program (ATLAS)



What to Expect

Setting the stage

- Why we want to study the **top quark** and the **Higgs boson**, why the LHC is a good place to do it, and why **machine learning (ML)** tools are vital to this endeavor

Higgs boson case study: boosted $H \rightarrow WW^*$

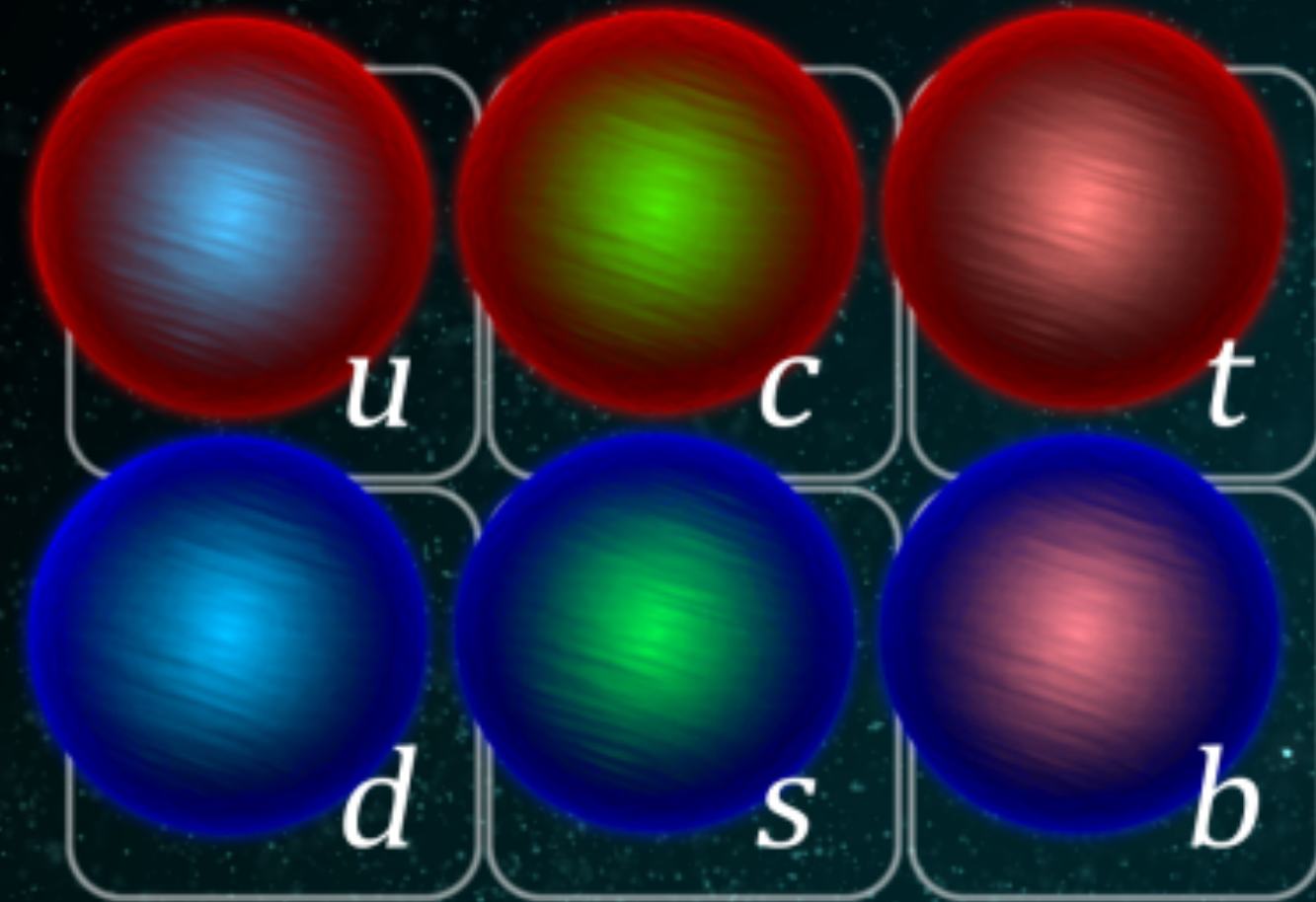
- Measurement of **boosted $H \rightarrow WW^*$** , facilitated by **transformer-based jet tagging** and **fine-tuning** techniques

Top quark case study: all-hadronic $t\bar{t}t\bar{t}$

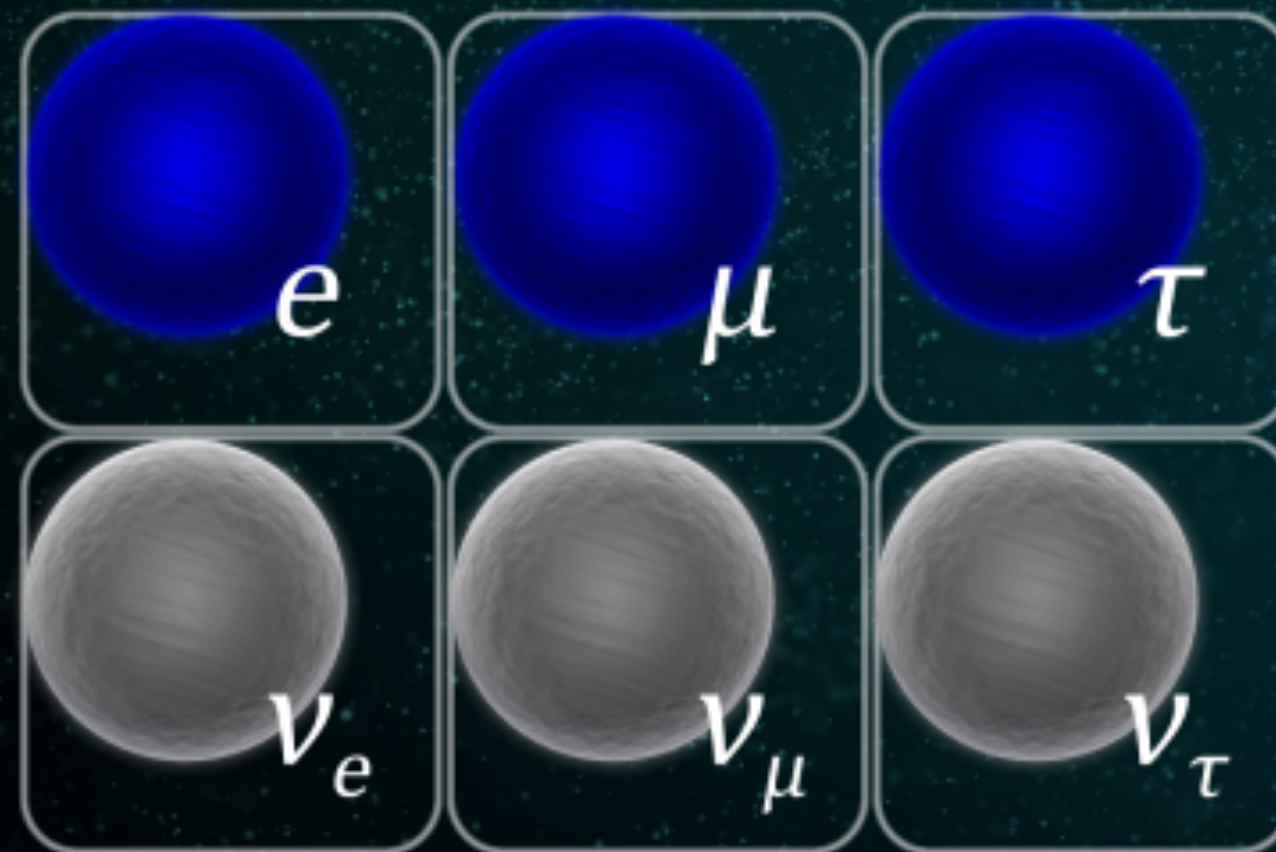
- Measurement of all-hadronic four-top production, facilitated by **ML-based data-driven background** estimation

Looking to the future

- Current and future developments of ML-based low-level **trigger algorithms** and **jet tagging** methods, and some concluding thoughts



Quarks



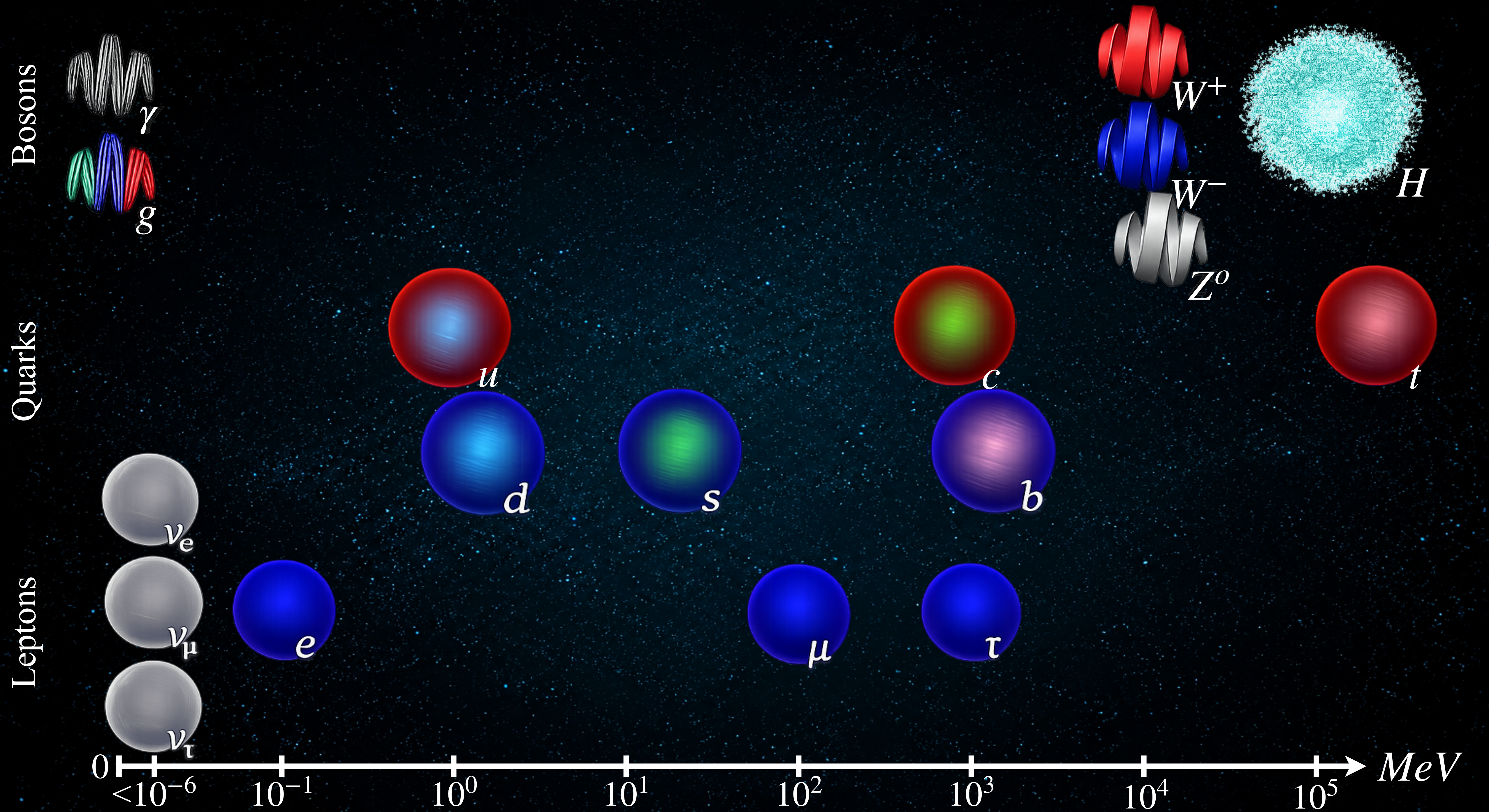
Leptons



Higgs boson



Forces



Big Questions...

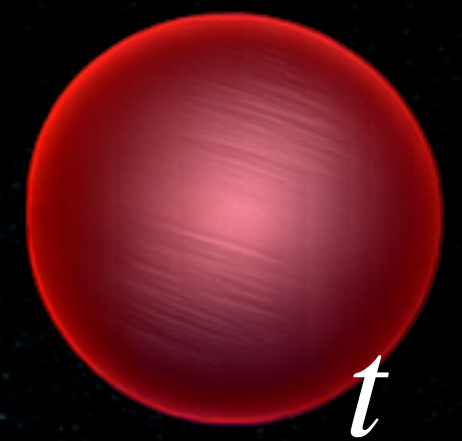
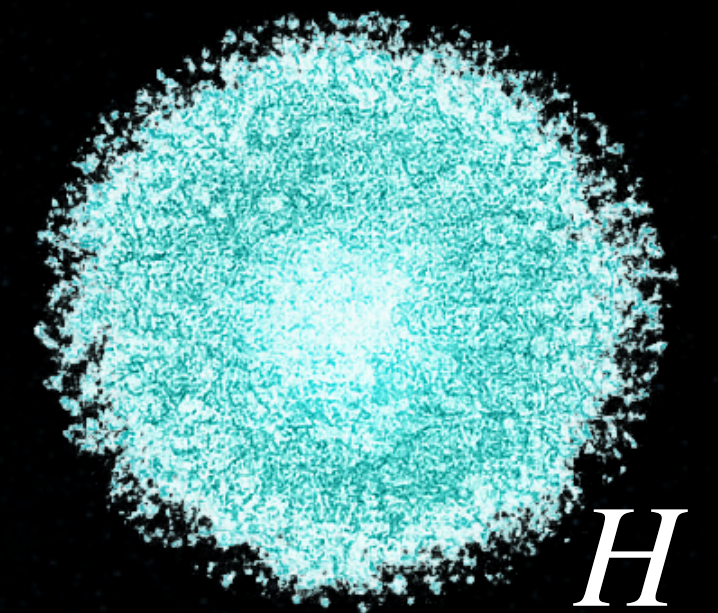
- **Naturalness/heirarchy problem:**

Why is the Higgs mass so small and the top mass so large, requiring an enormously fine-tuned theory?

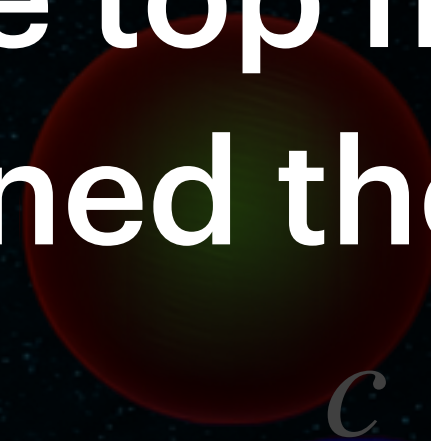
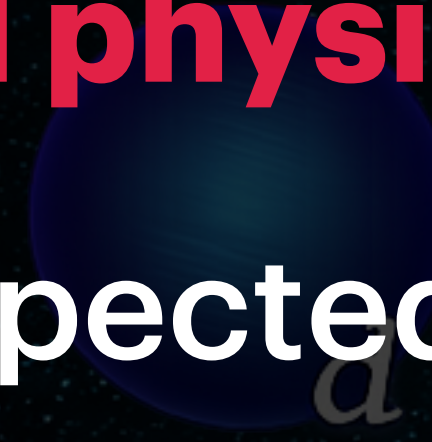
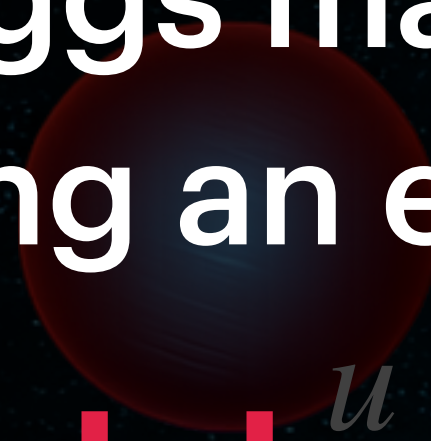
- **Undiscovered physics?**

Are there unexpected new particles to discover?

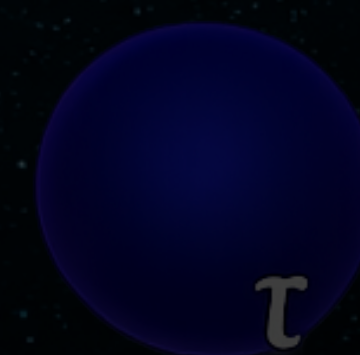
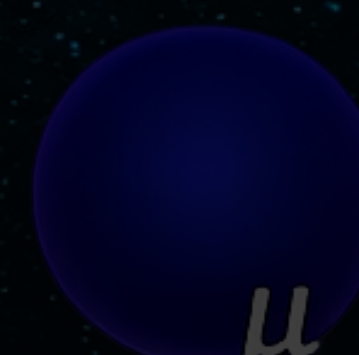
Bosons



Quarks



Leptons



Big Questions...

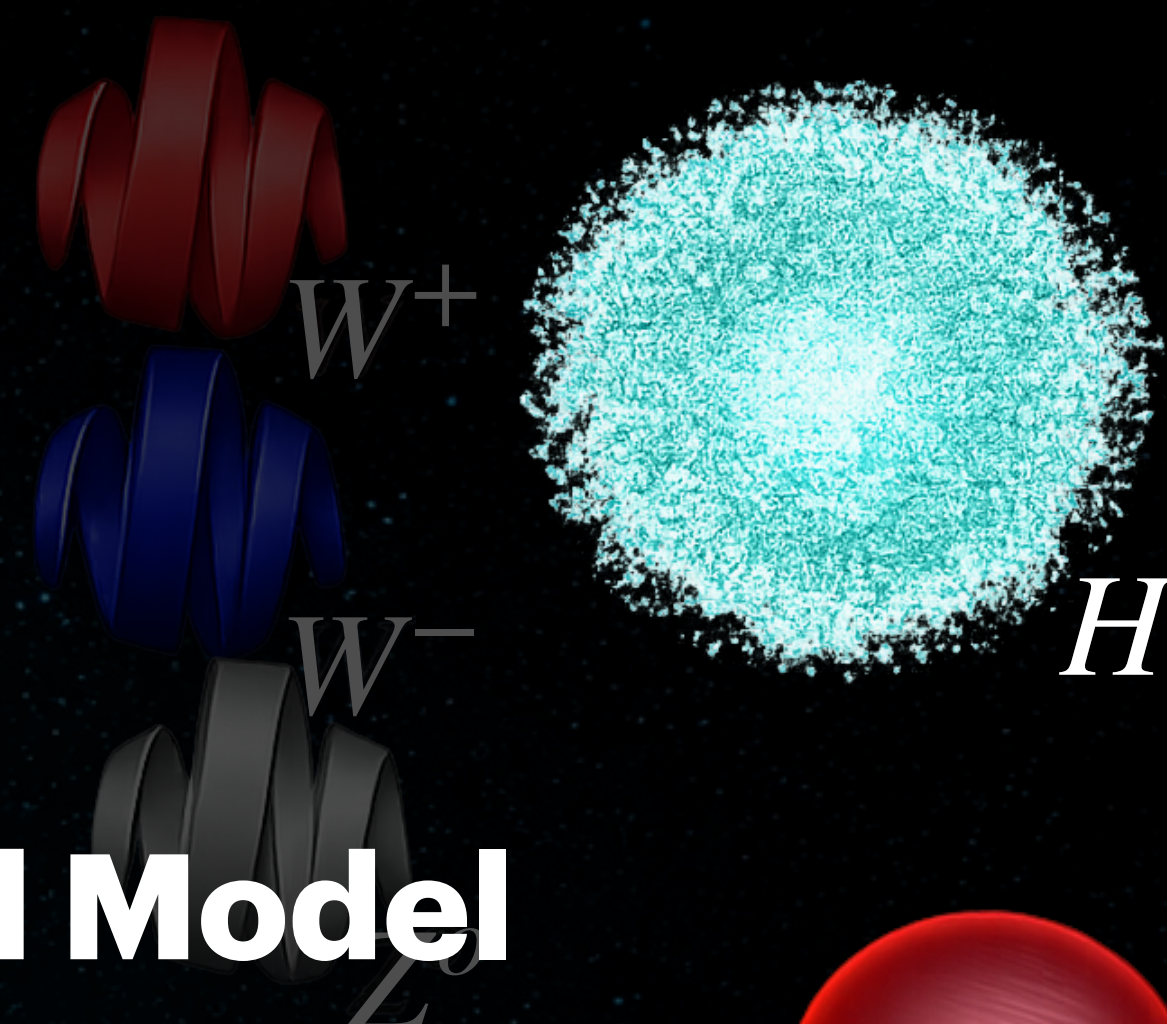
Require big (little) particles!

→ **Critical to understanding the Standard Model**

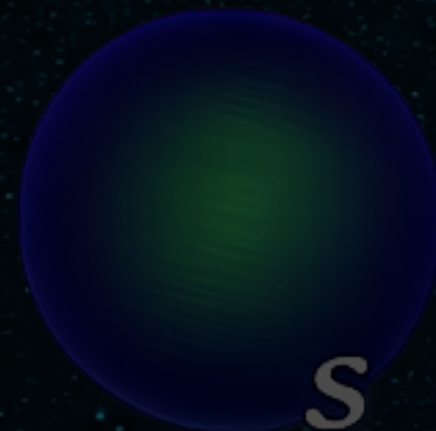
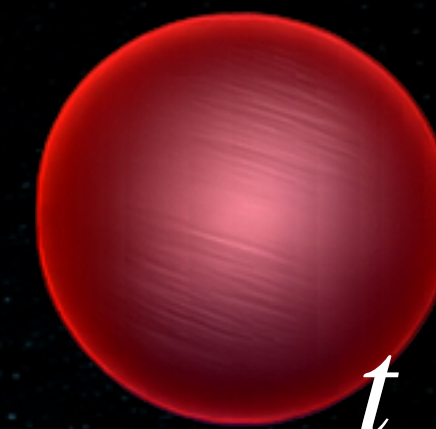
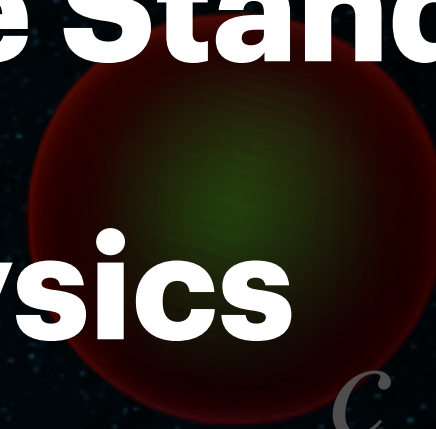
→ **Powerful probes for new physics**

Let's see some examples!

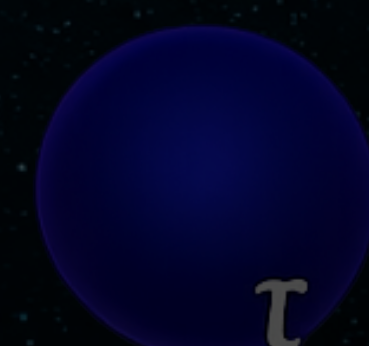
Bosons



Quarks

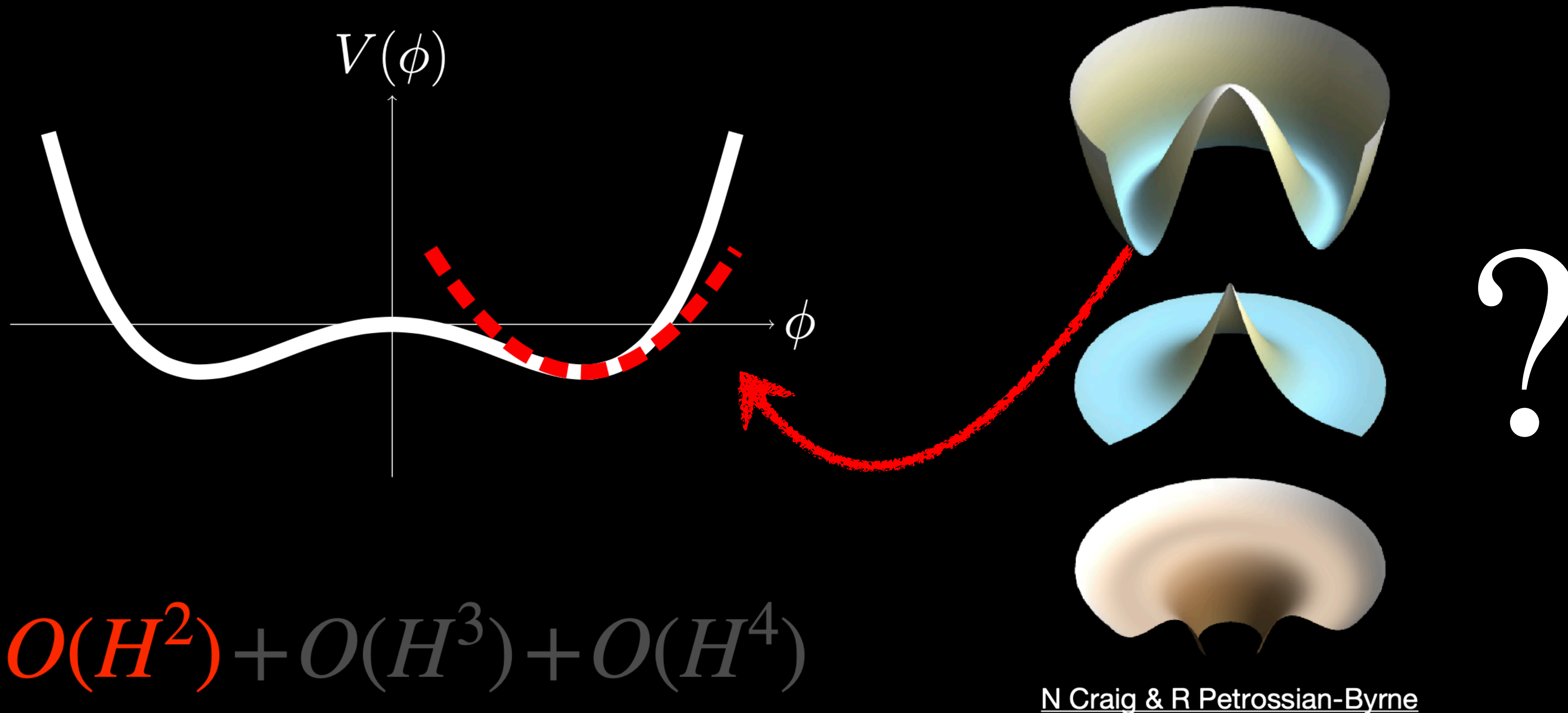


Leptons



Is the Higgs Potential “Standard?”

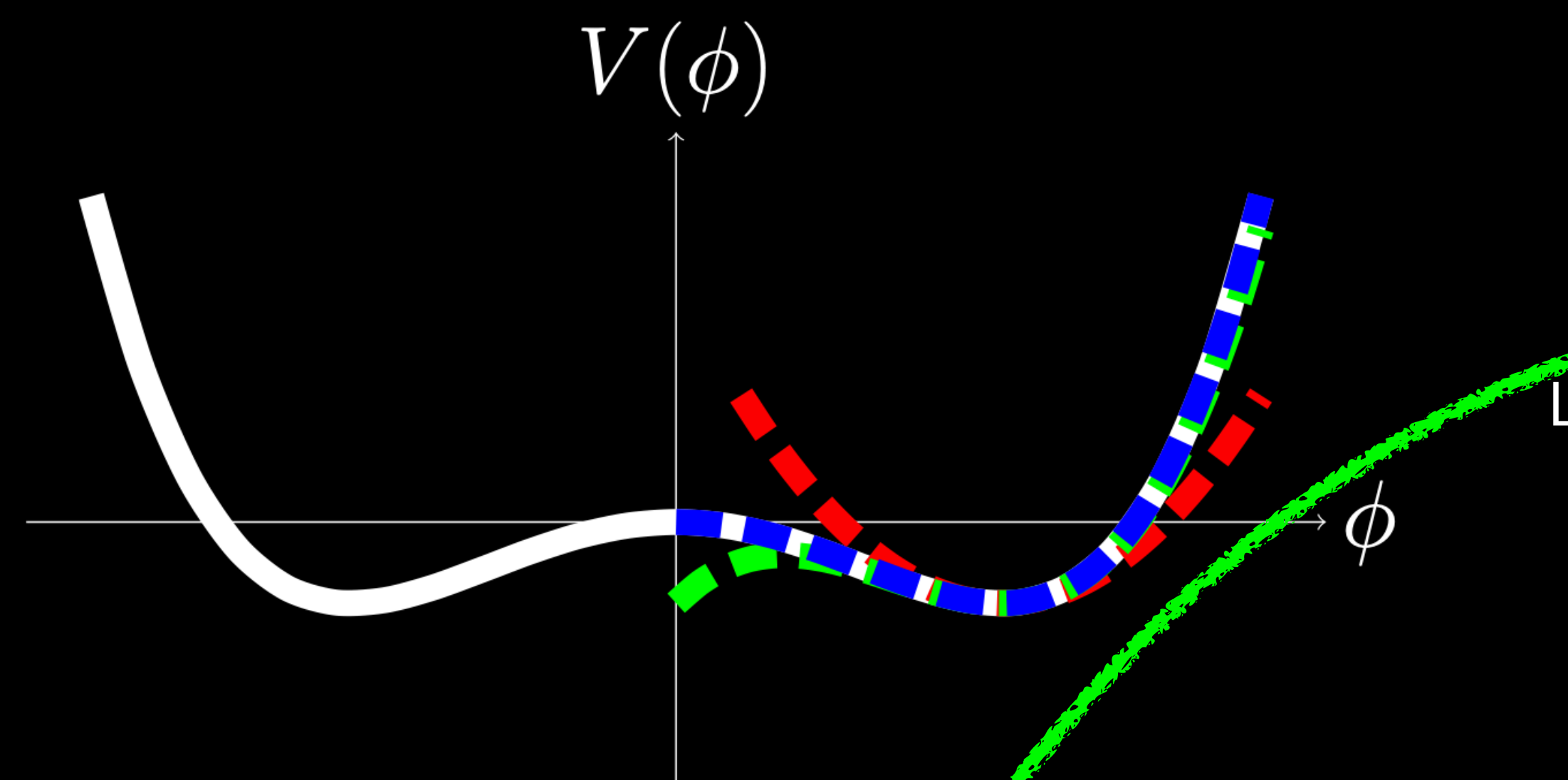
- Example: We need multi-Higgs production to understand the shape of the Higgs potential!



Figures by Larry Lee

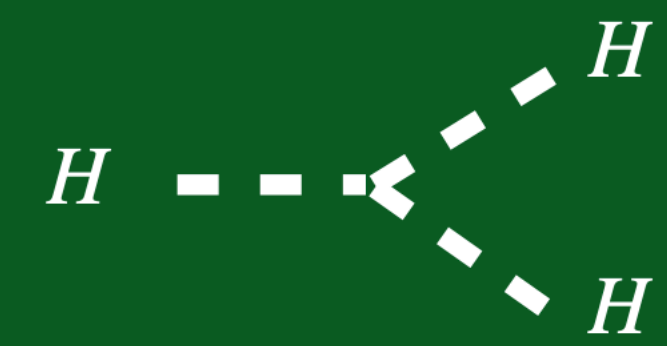
Is the Higgs Potential “Standard?”

- Example: We need multi-Higgs production to understand the shape of the Higgs potential!

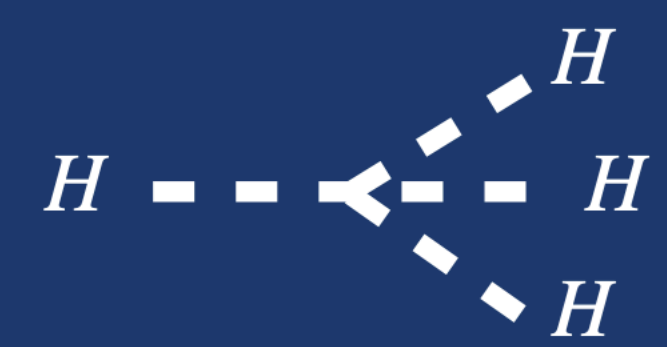


$$O(H^2) + O(H^3) + O(H^4)$$

Trilinear coupling
→ HH production*



Quartic coupling
→ HHH production*

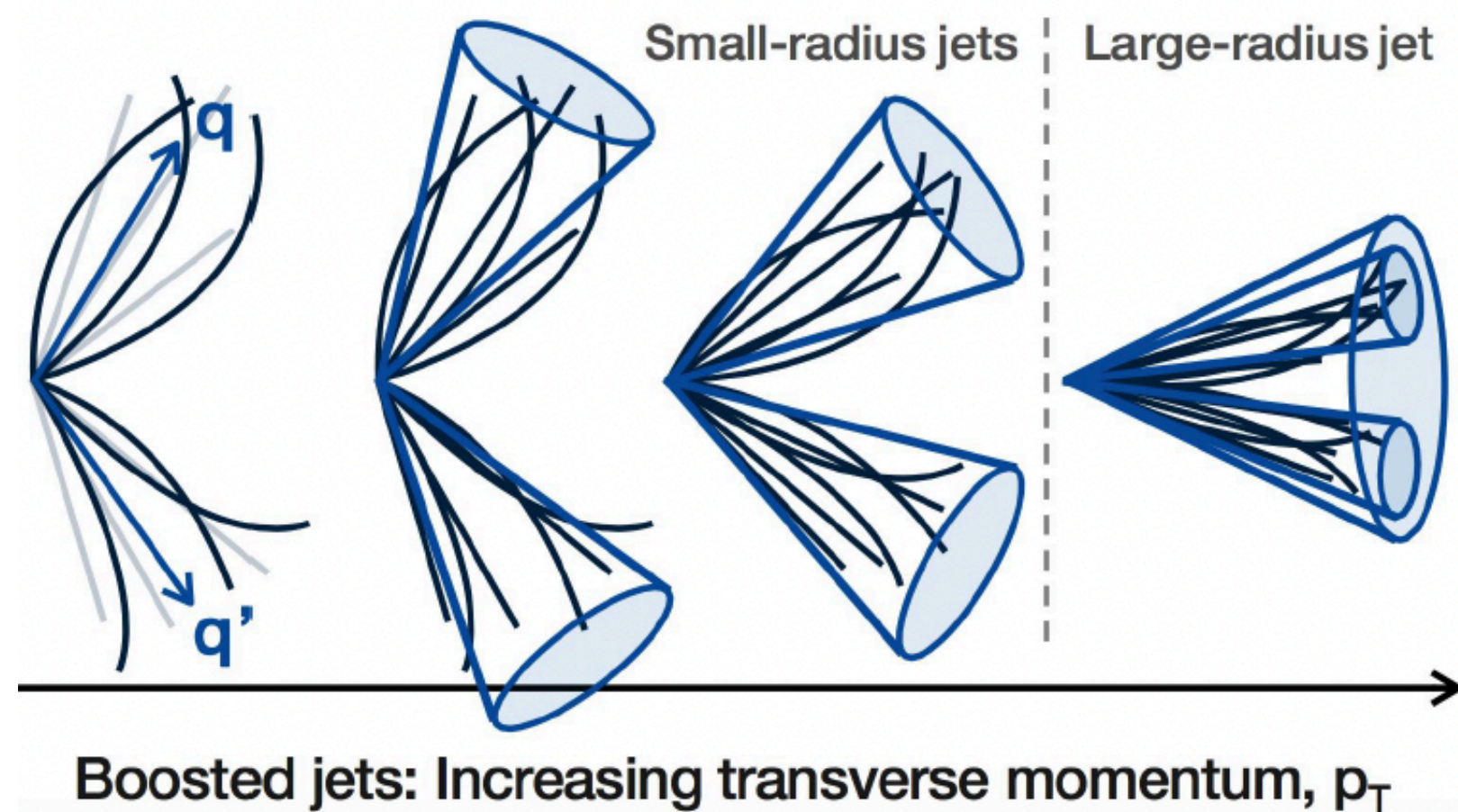


*work 
underway-
later in the
talk!

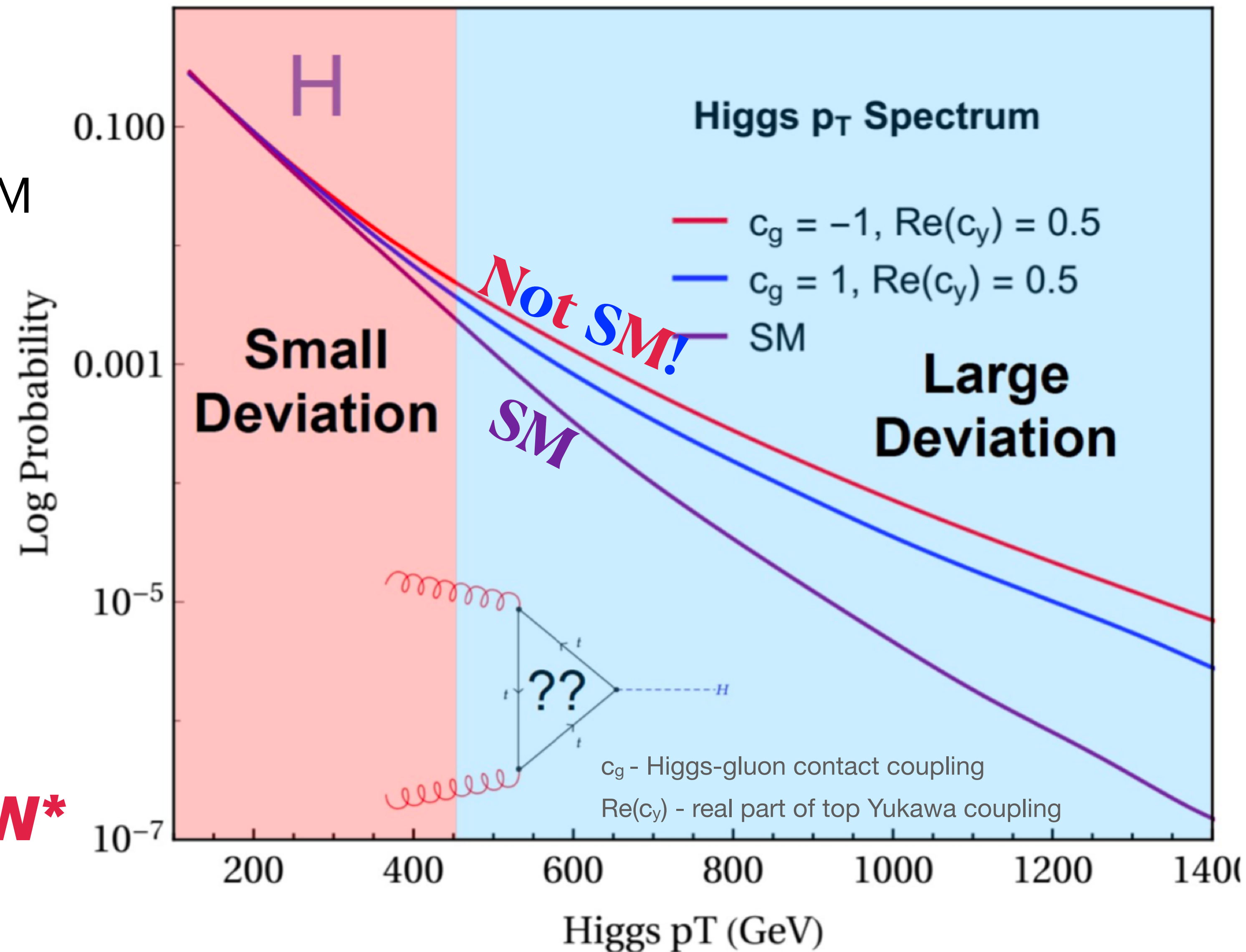
Figures by Larry Lee

Is new physics hiding at higher energies?

- Example: Boosted (high transverse momentum) Higgs production
- Seemingly small deviations from SM amplified at higher energies!

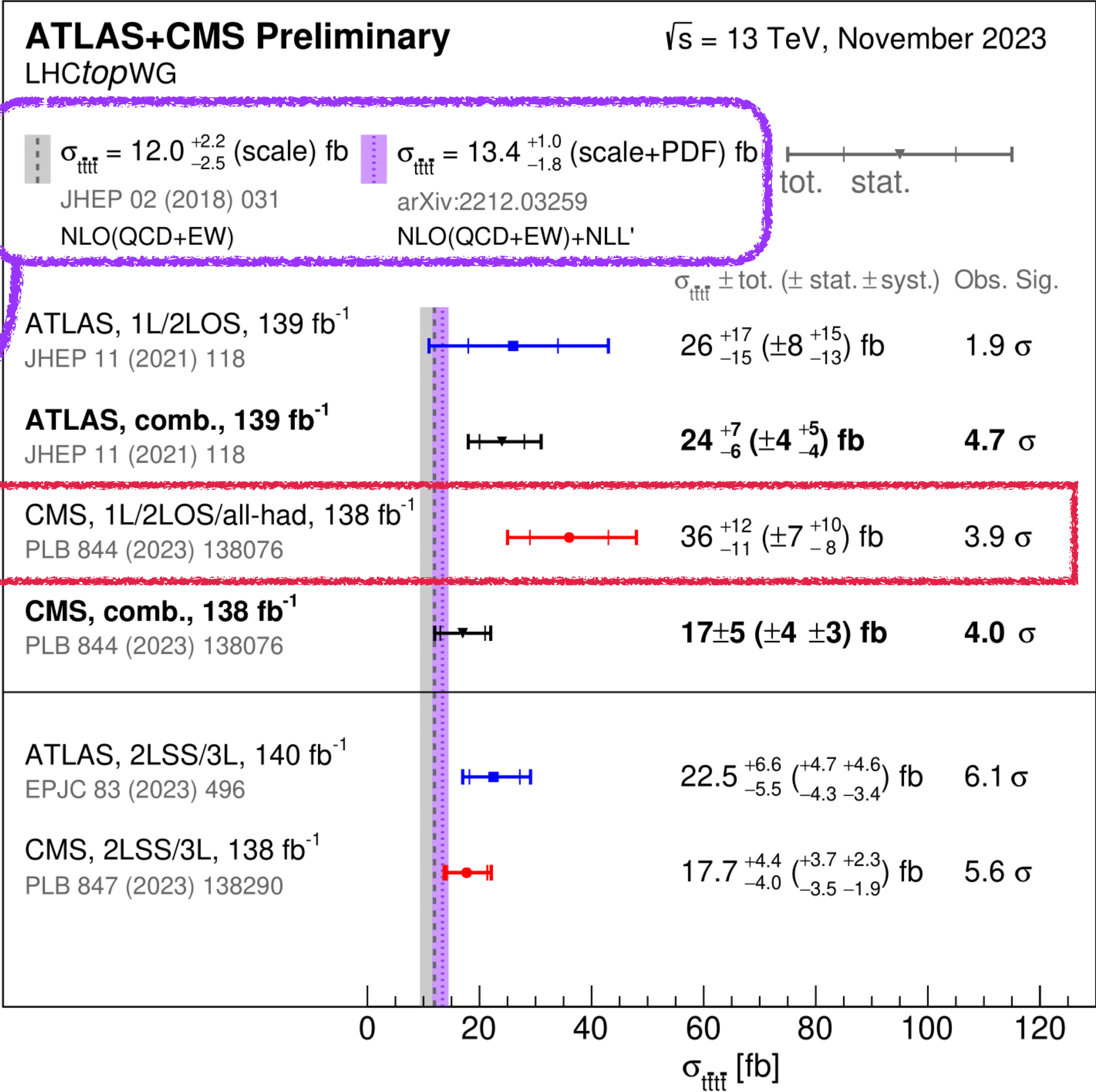
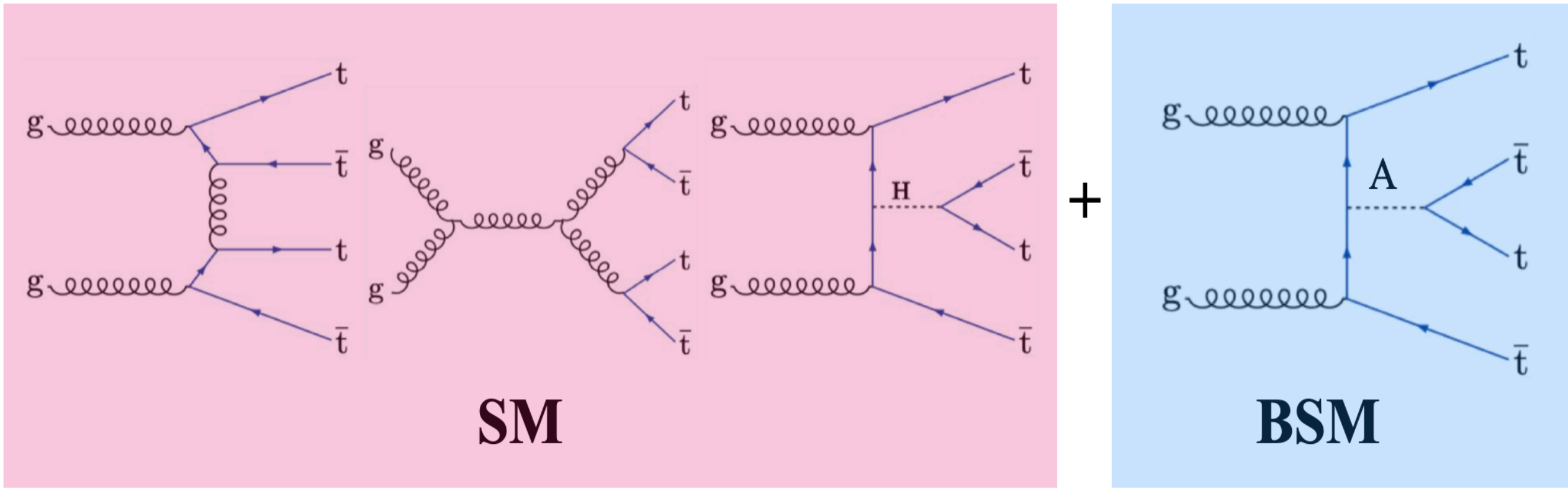


This talk: boosted $H \rightarrow WW^*$



The top quark as a probe for new physics?

- Example: four-top quark production
- Direct probe of top-Higgs Yukawa coupling
- Enhancements of cross section evidence of BSM physics
- Precise theory predictions are also vital!



This talk

Squeezing the Standard Model

Theory

Higher-order
calculations

Improved PDFs

New
computational
techniques

SM

BSM?

More powerful
colliders

Improved detection
technology

Cutting-edge
analysis
techniques

Experiment



Squeezing the Standard Model

Theory

Higher-order
calculations

Improved PDFs

New
computational
techniques

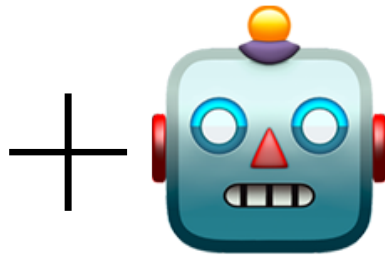
SM

More powerful
colliders

Improved detection
technology

Cutting-edge
analysis
techniques

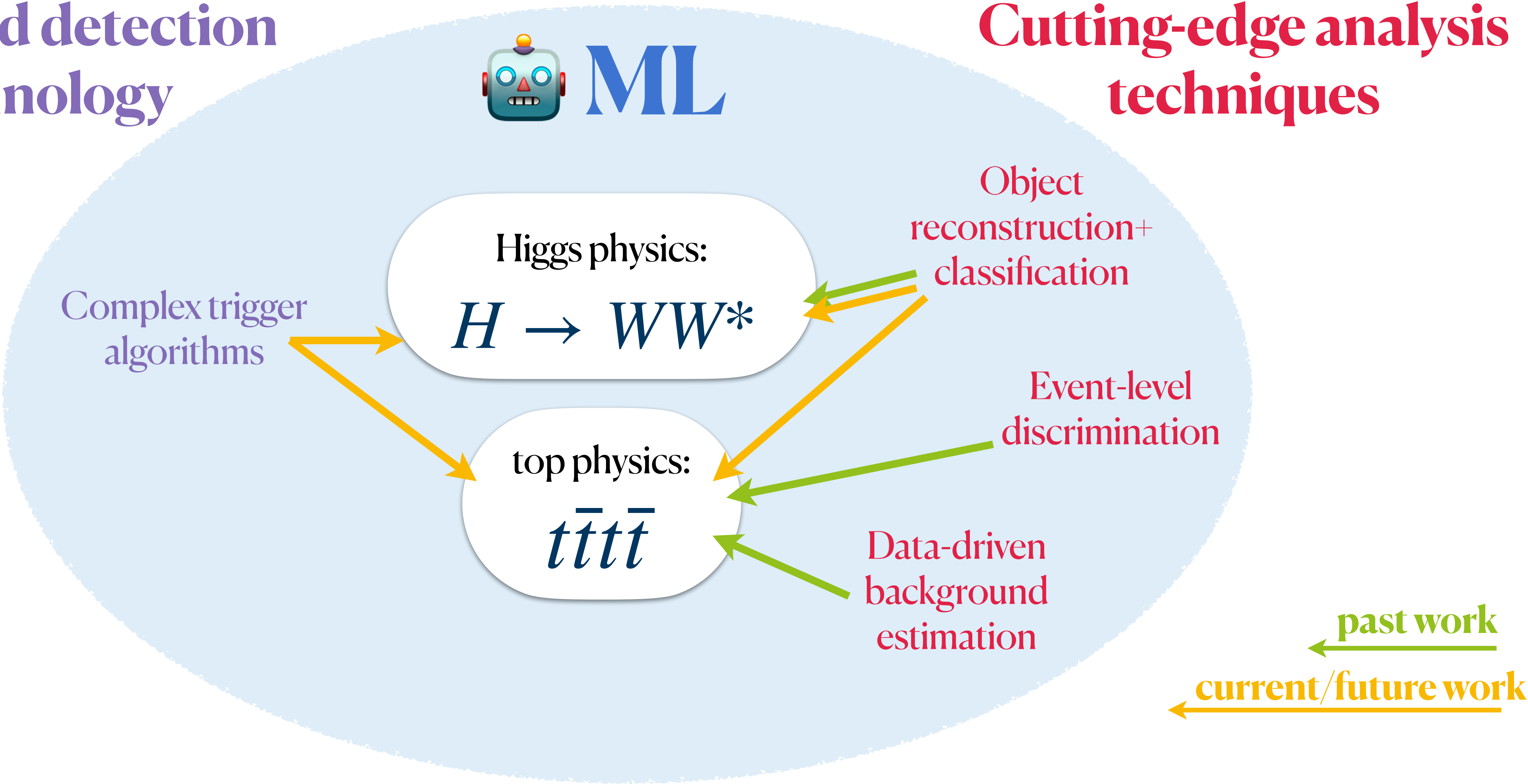
Experiment



Case Studies & their ML tools

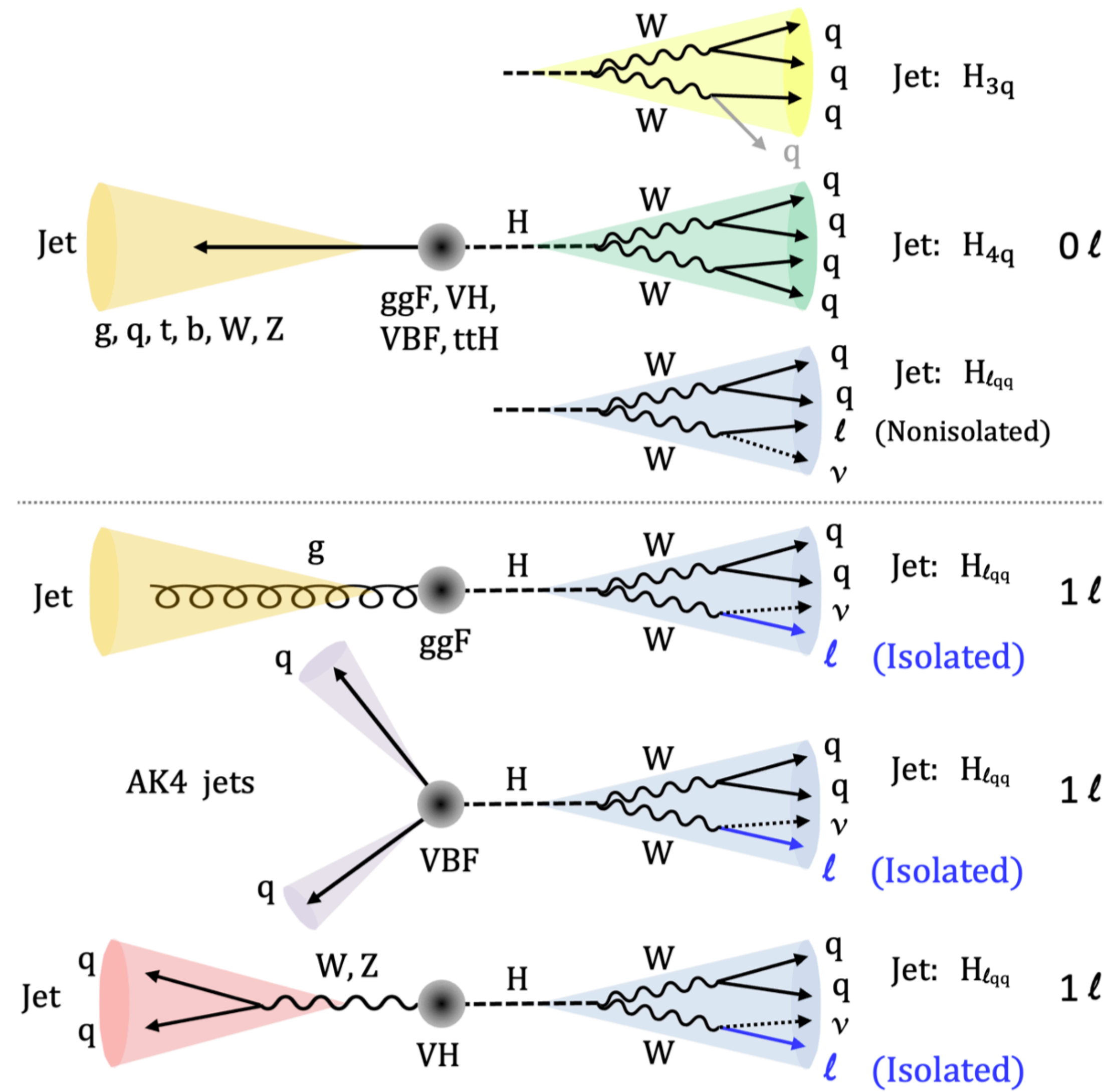
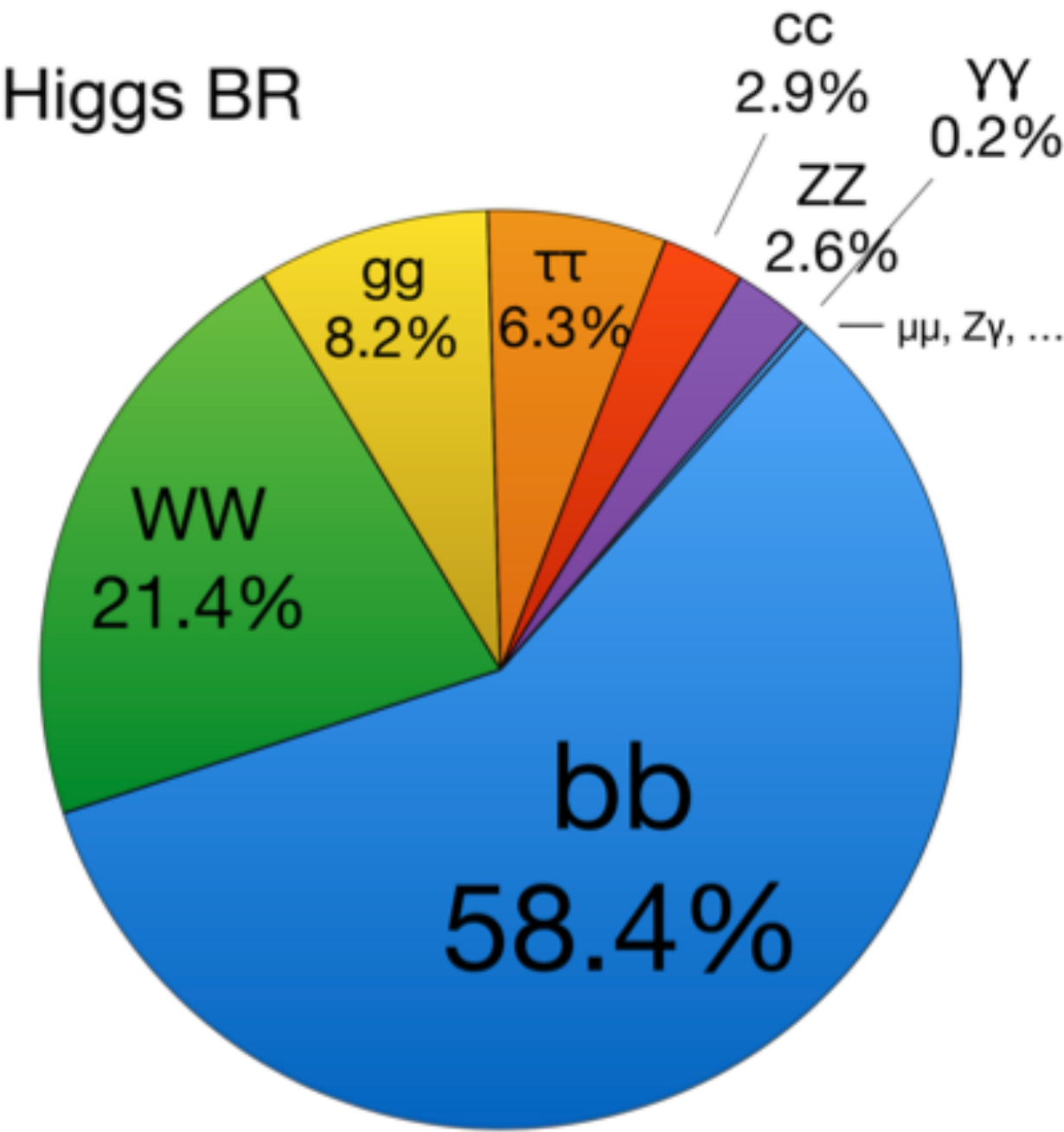
Improved detection
technology

Cutting-edge analysis
techniques



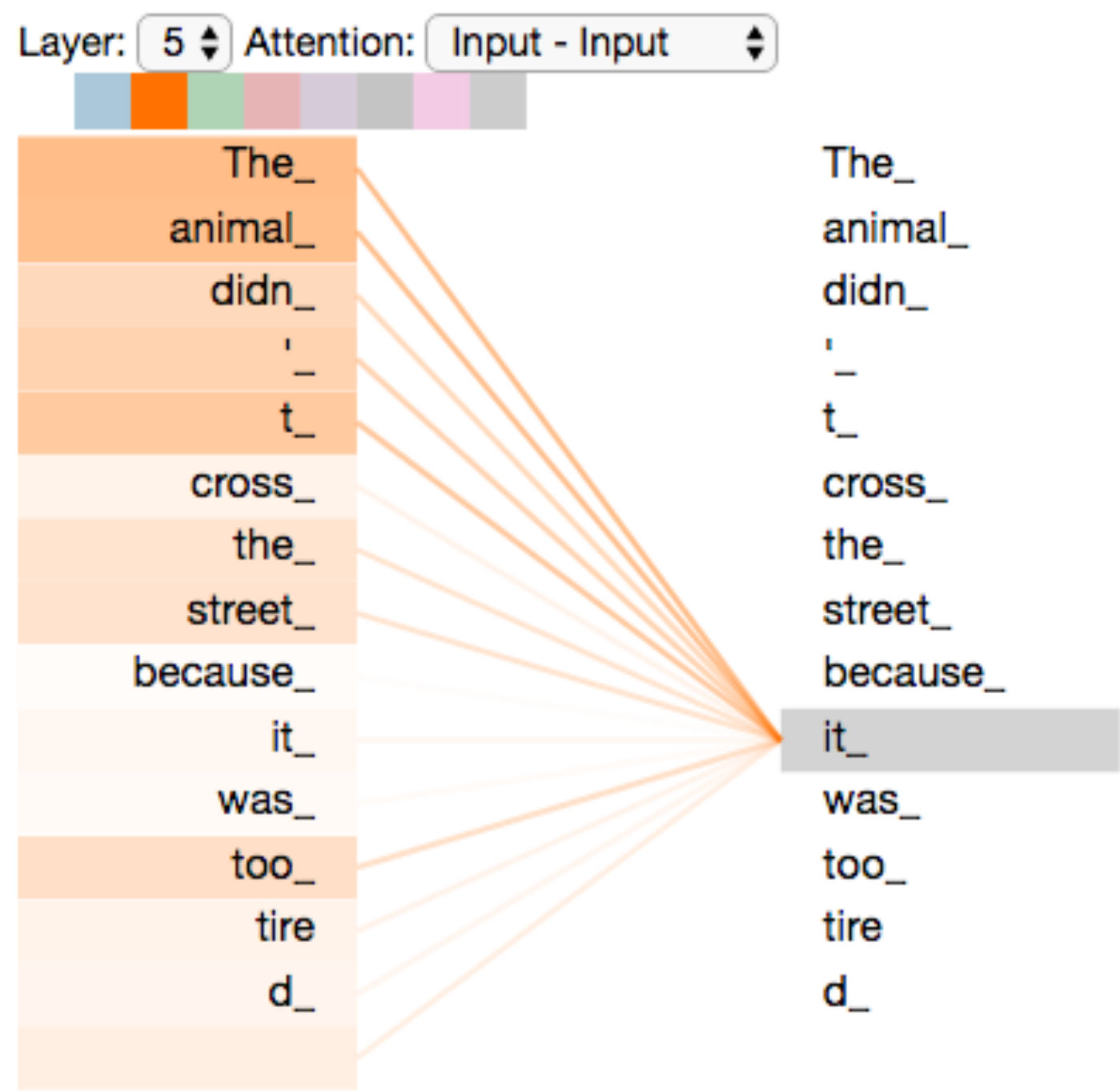
Search for Boosted $H \rightarrow WW^*$

- $H \rightarrow WW$ is 2nd largest BR ~21%
- First study of H production decaying to WW with at least 1 hadronic W (decaying to qq)
- Facilitated by NN-based jet tagging: "**ParT**" Particle Transformer



Transformers

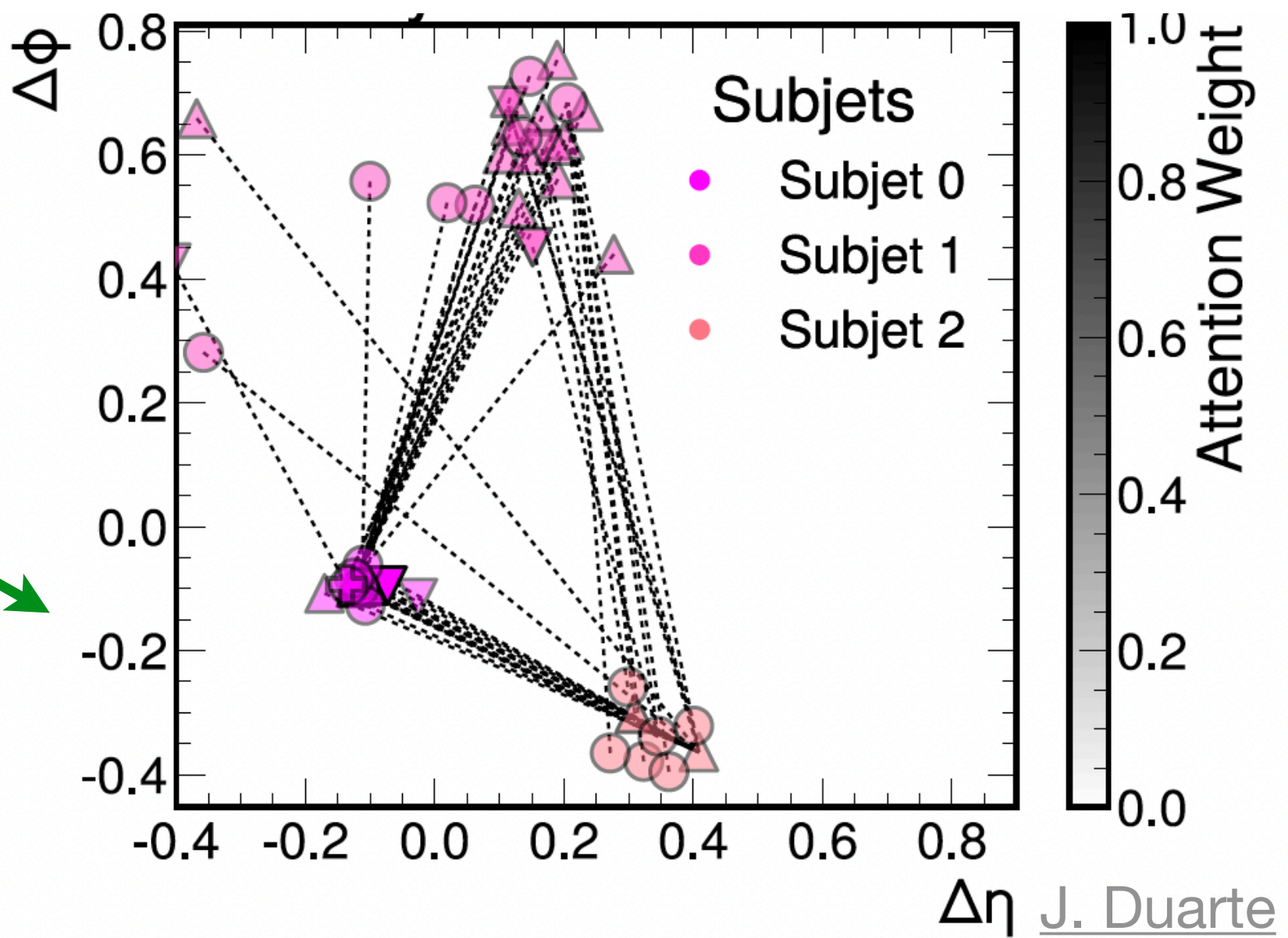
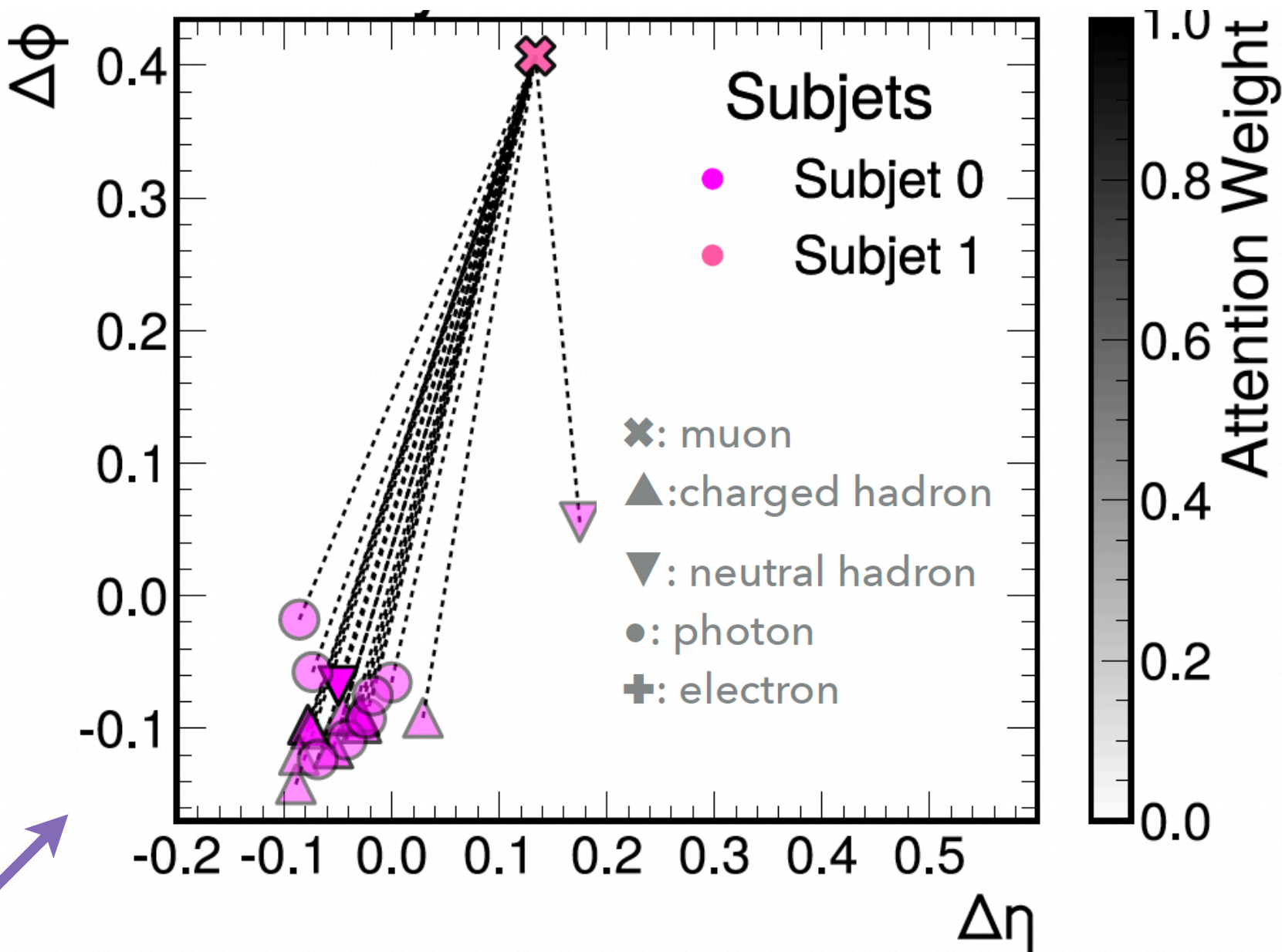
- Powerful neural networks at the core of LLMs based on the idea of attention
- Process a set of **items** (words, **jet constituents**...) and let each one gather **context** from all the others
- Each item computes its output from every other item in parallel → Learns efficient, complex relations



Leptonic decay:
attention focused
on lepton

Hadronic decay:
attention split
across subjects

→ Can form “foundational
models” for diverse
downstream tasks!



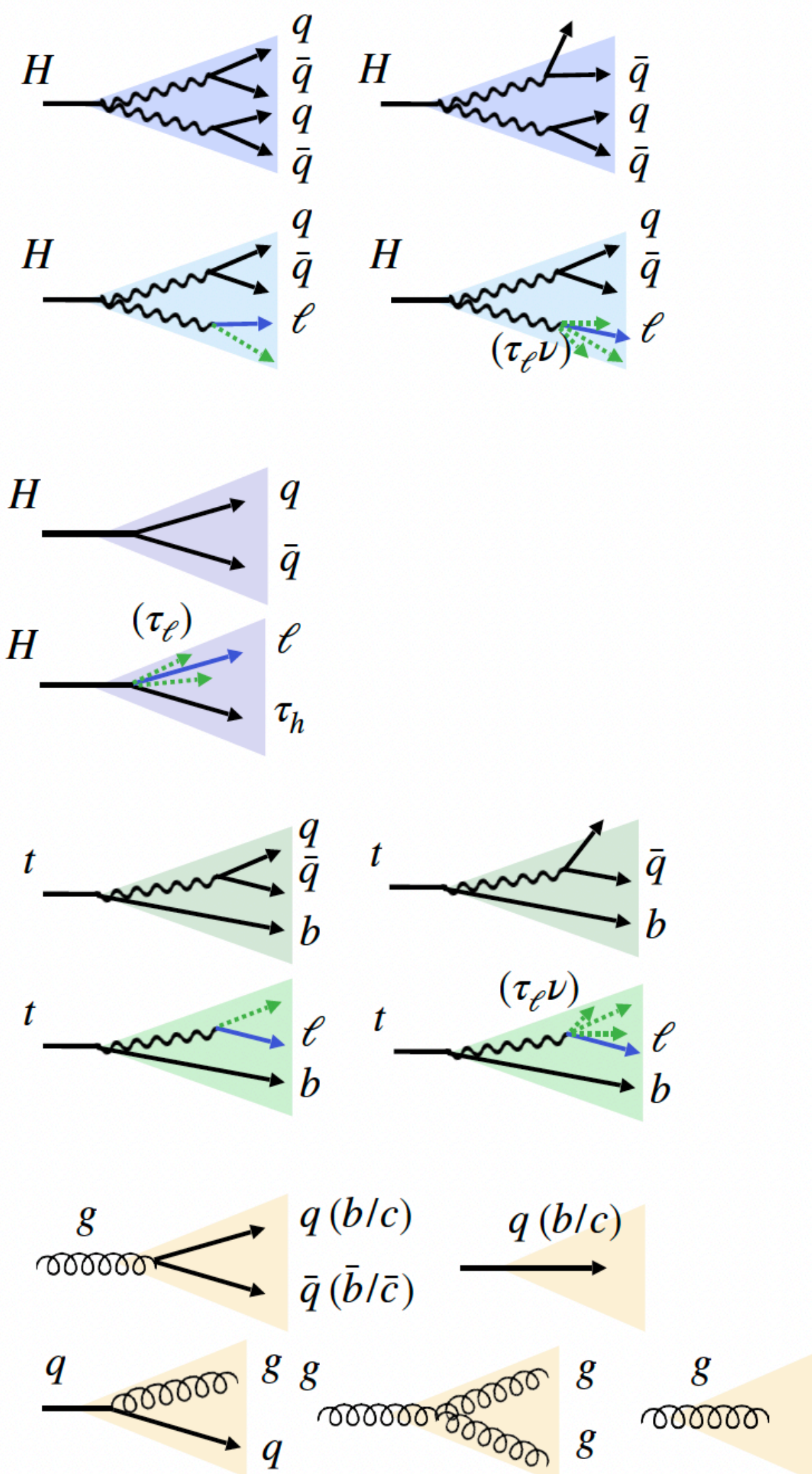
J. Duarte

Particle Transformer

- **ParT** model: Transformer architecture
 - Uses learned “attention” (and pairwise features) to give more weight or disregard certain particles in order to infer the origin of jets
- → expanded to GloParT

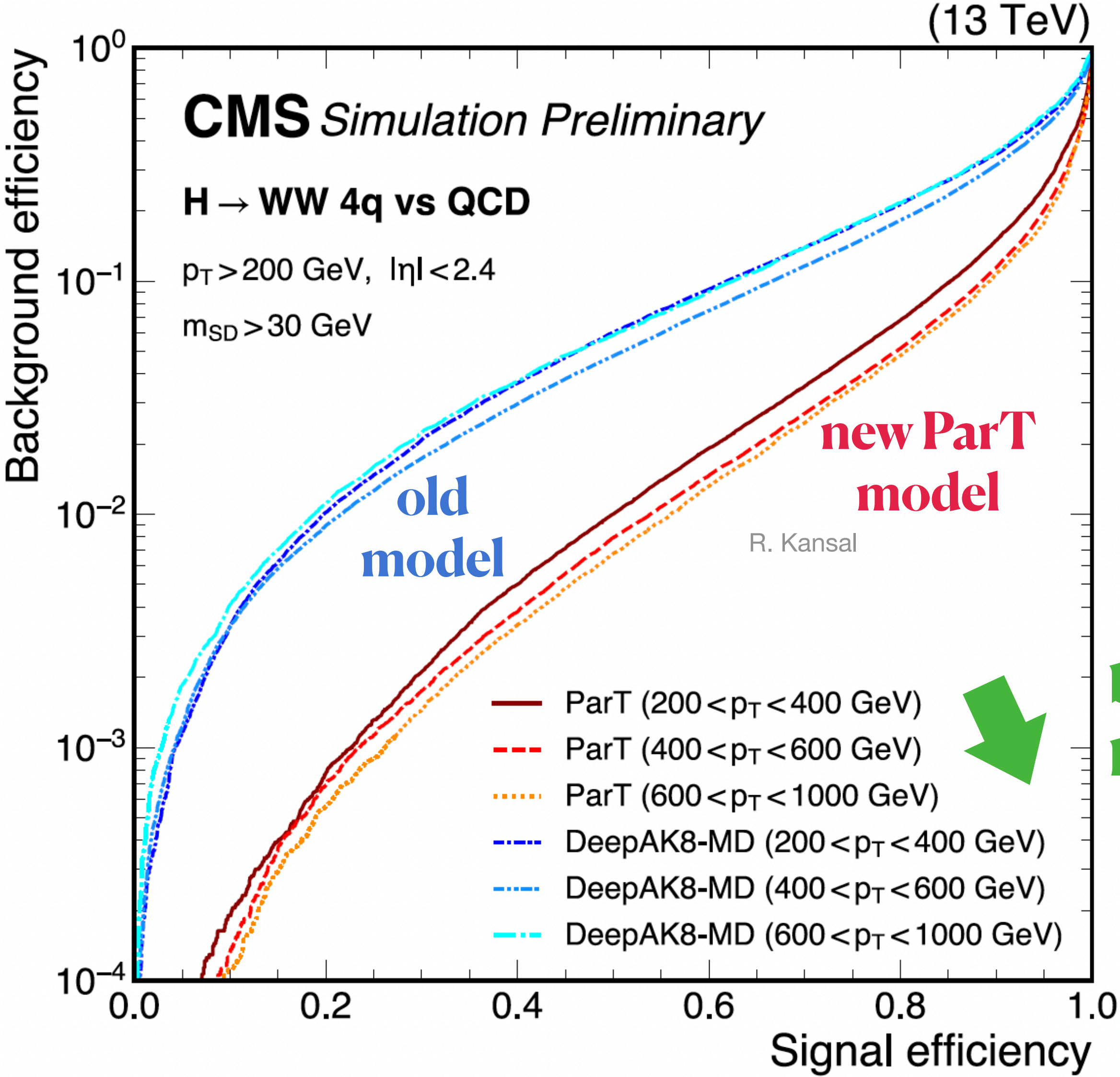
- GNNs for jet tagging
 - **CMS**: ParticleNet, **ATLAS**: GN1
- Transformers for jet tagging
 - **CMS**: ParT, UParT GloParT, **ATLAS**: GN2, GN2X, GN3, GN3X

Process	Final state/ prongness	heavy flavour	# of classes
H→VV (full-hadronic)	qqqq	0c/1c/2c	3
	qqq		3
H→WW (semi-leptonic)	eVqq	0c/1c	2
	μVqq		2
	τ _e Vqq		2
	τ _μ Vqq		2
	τ _h Vqq		2
H→qq		bb	1
		cc	1
		ss	1
		qq (q=u/d)	1
H→ττ	τ _e τ _h		1
	τ _μ τ _h		1
	τ _h τ _h		1
t→bW (hadronic)	bqq	1b + 0c/1c	2
	bq		2
t→bW (leptonic)	b _e V	1b	1
	b _μ V		1
	bτ _e V		1
	bτ _μ V		1
	bτ _h V		1
QCD		b	1
		bb	1
		c	1
		cc	1
		others (light)	1



Particle Transformer

- ParT significantly improves on previous state-of-the-art method for $H \rightarrow WW \rightarrow 4q$ tagging



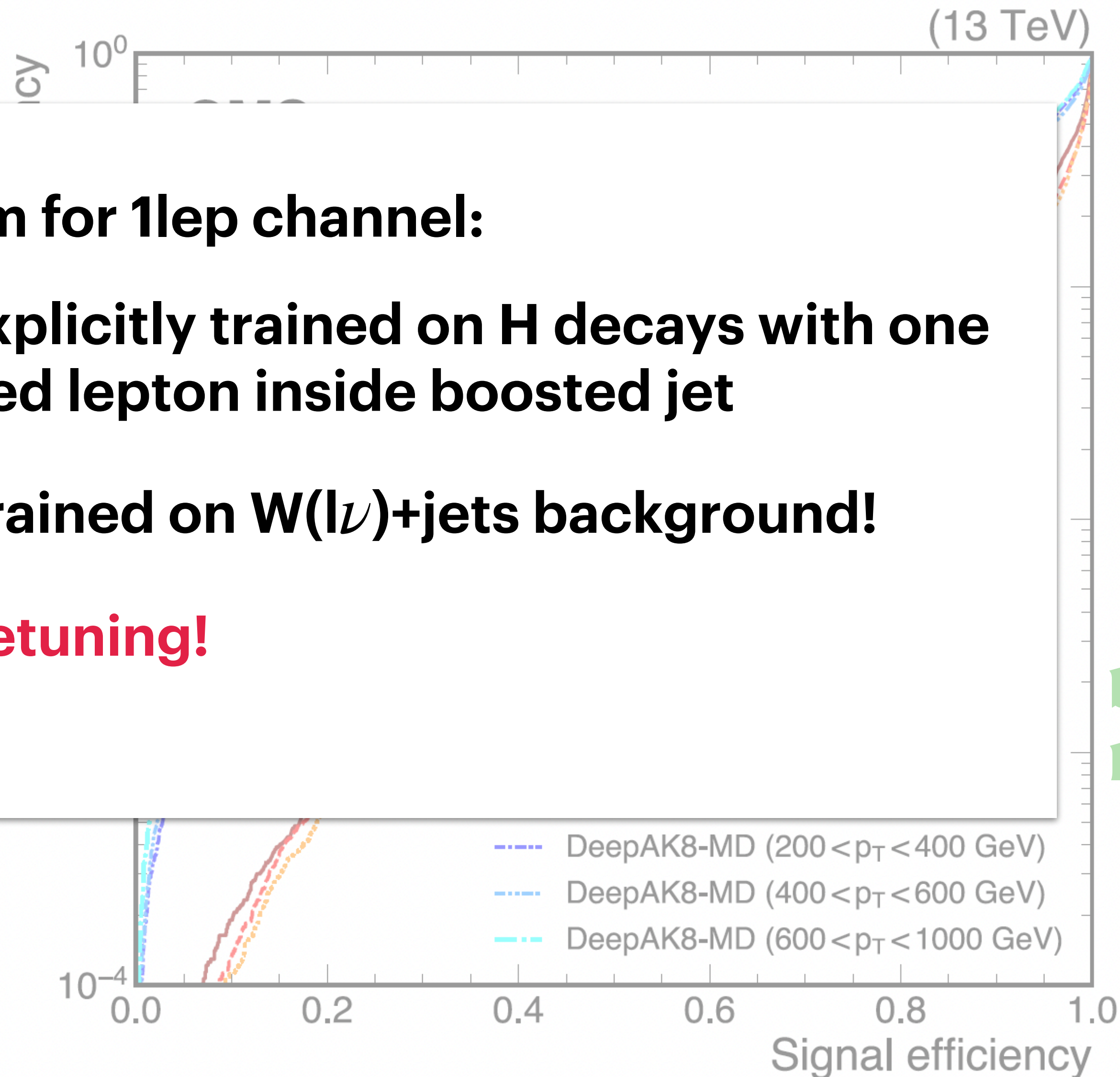
**better
performance**

Particle Transformer

- ParT significantly improved on previous state-of-the-art method for $H \rightarrow WW$ tagging

Problem for 1lep channel:

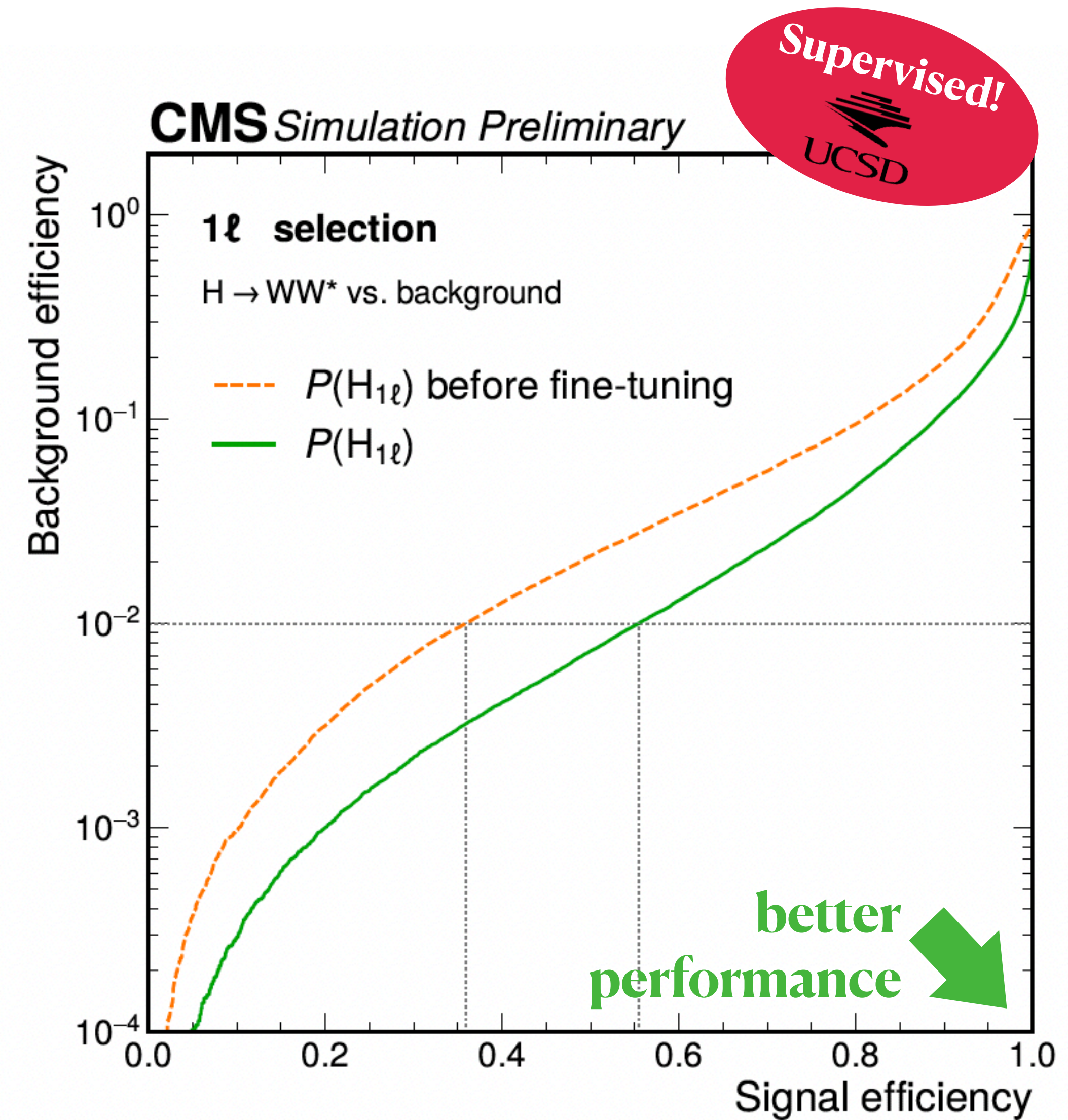
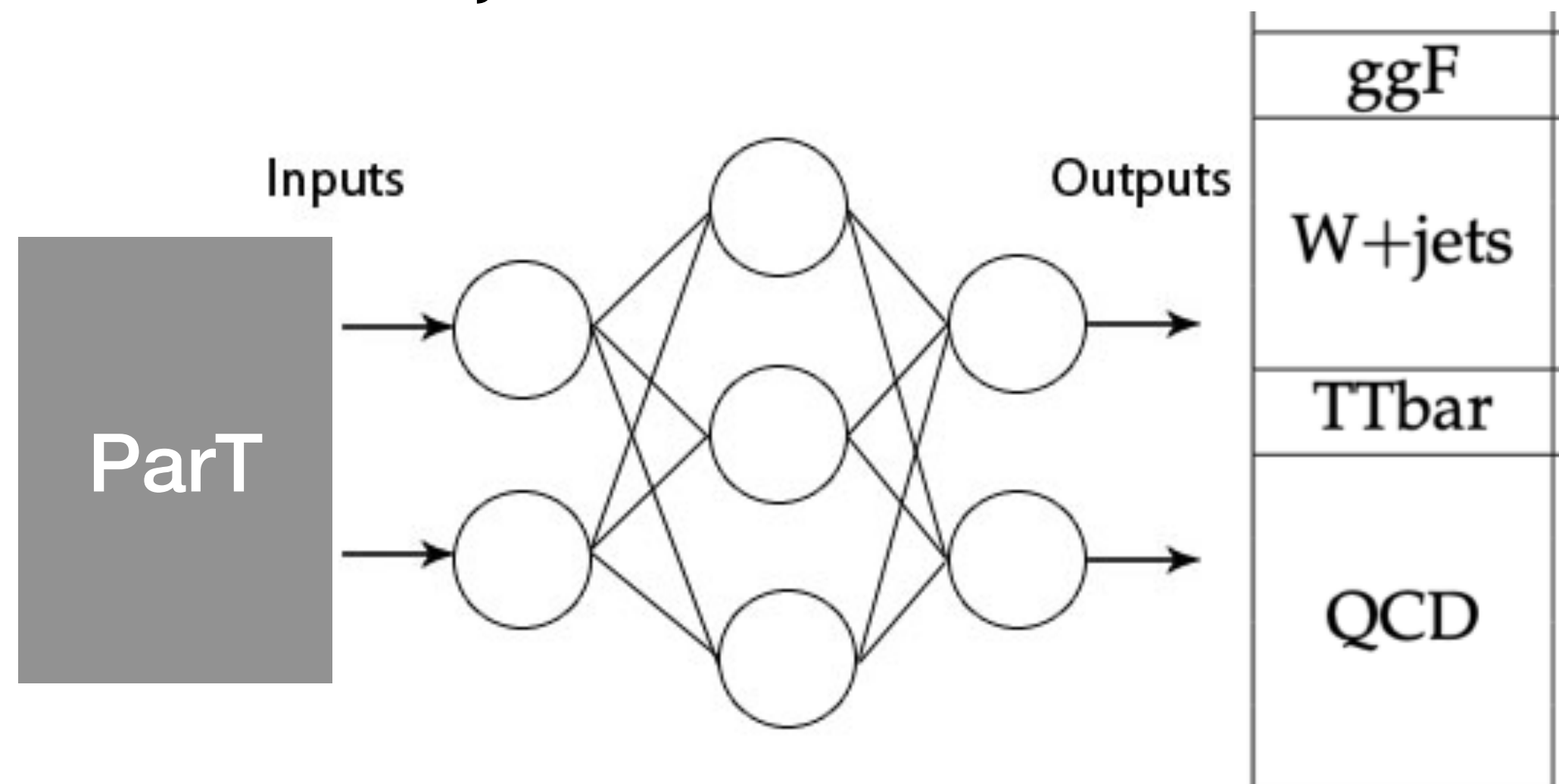
- Not explicitly trained on H decays with one isolated lepton inside boosted jet
- Not trained on $W(l\nu)$ +jets background!
- → **Finetuning!**



Search for boosted $H \rightarrow WW^*$

→ Train smaller specialized model starting from output of broad "foundation" model

- Inputs 128 neurons from last hidden layer of ParT
- Outputs probability of 1lep $H \rightarrow WW$, QCD, $t\bar{t}$, or W+jets



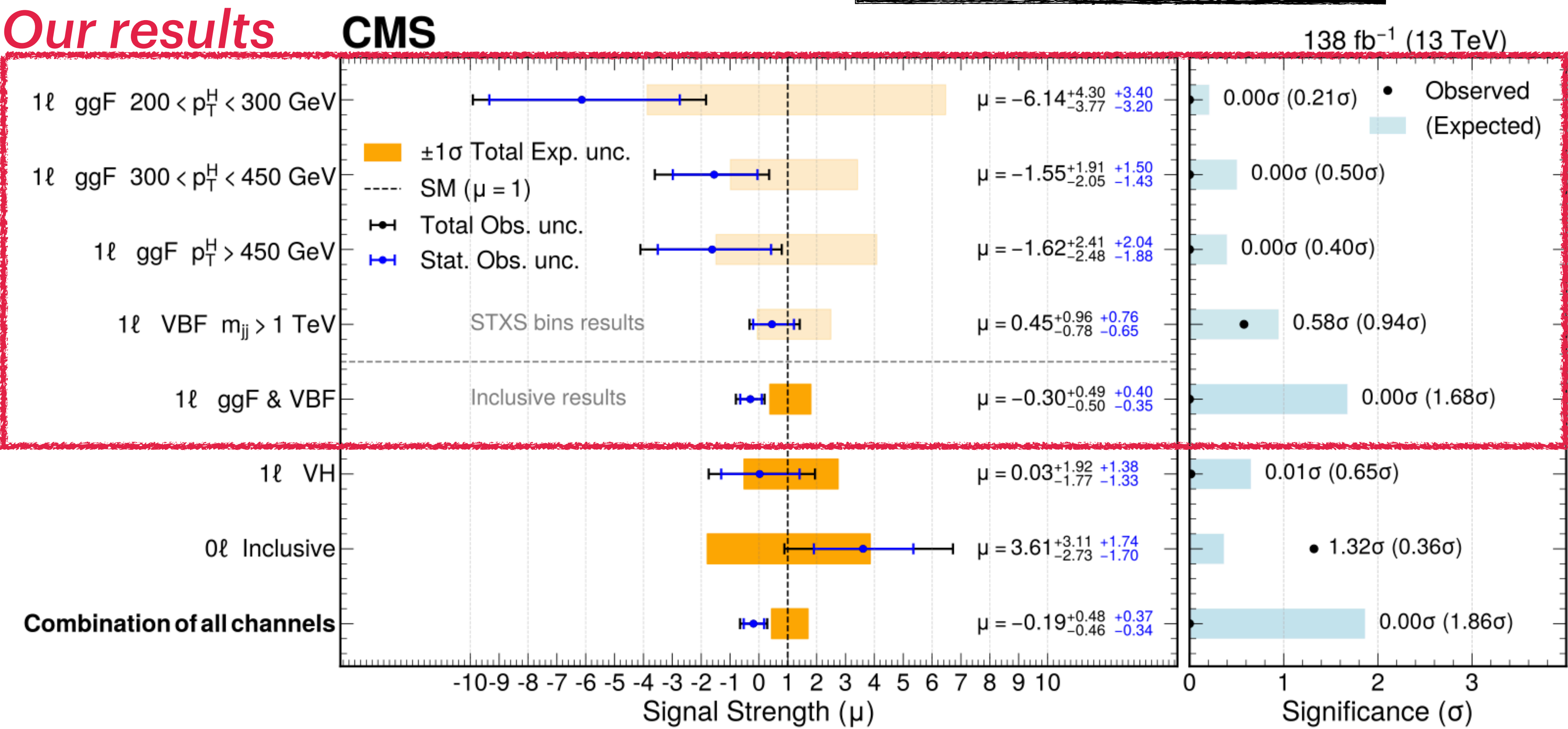
20% more signal efficiency for the same background rejection

Search for boosted $H \rightarrow WW^*$



CMS-HIG-24-008

- First study of boosted $H \rightarrow WW$ production with at least 1 hadronic W
- Facilitated by **ParT** boosted jet tagger + **finetuning** techniques!
- **→ ML-driven jet reconstruction+ classification**

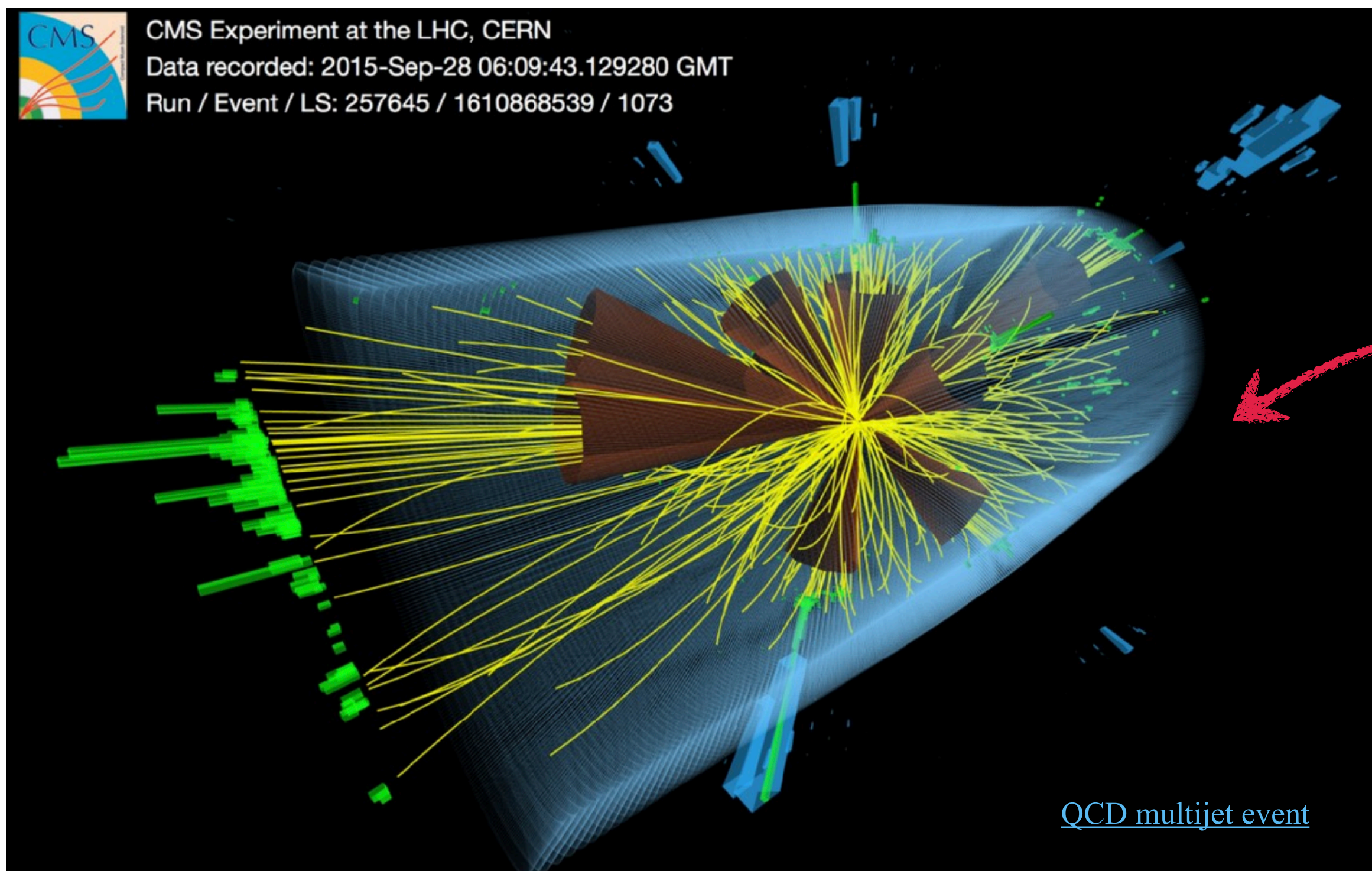


All-Hadronic Four-Top Production

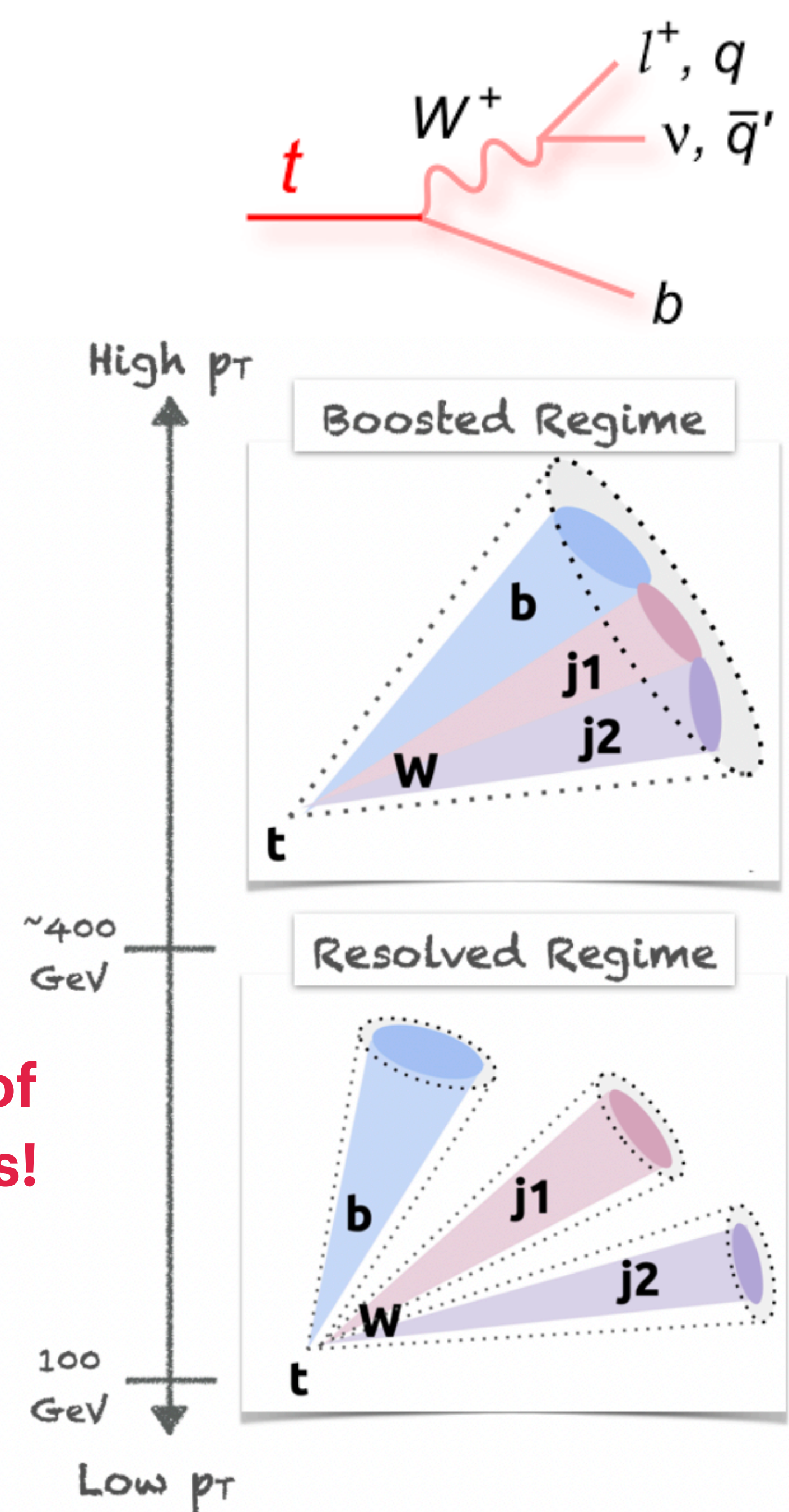
- First all-hadronic $t\bar{t}t\bar{t}$ analysis (published 2023)
- Very challenging final state due to large QCD-multijet backgrounds

PhD Thesis

UCSB

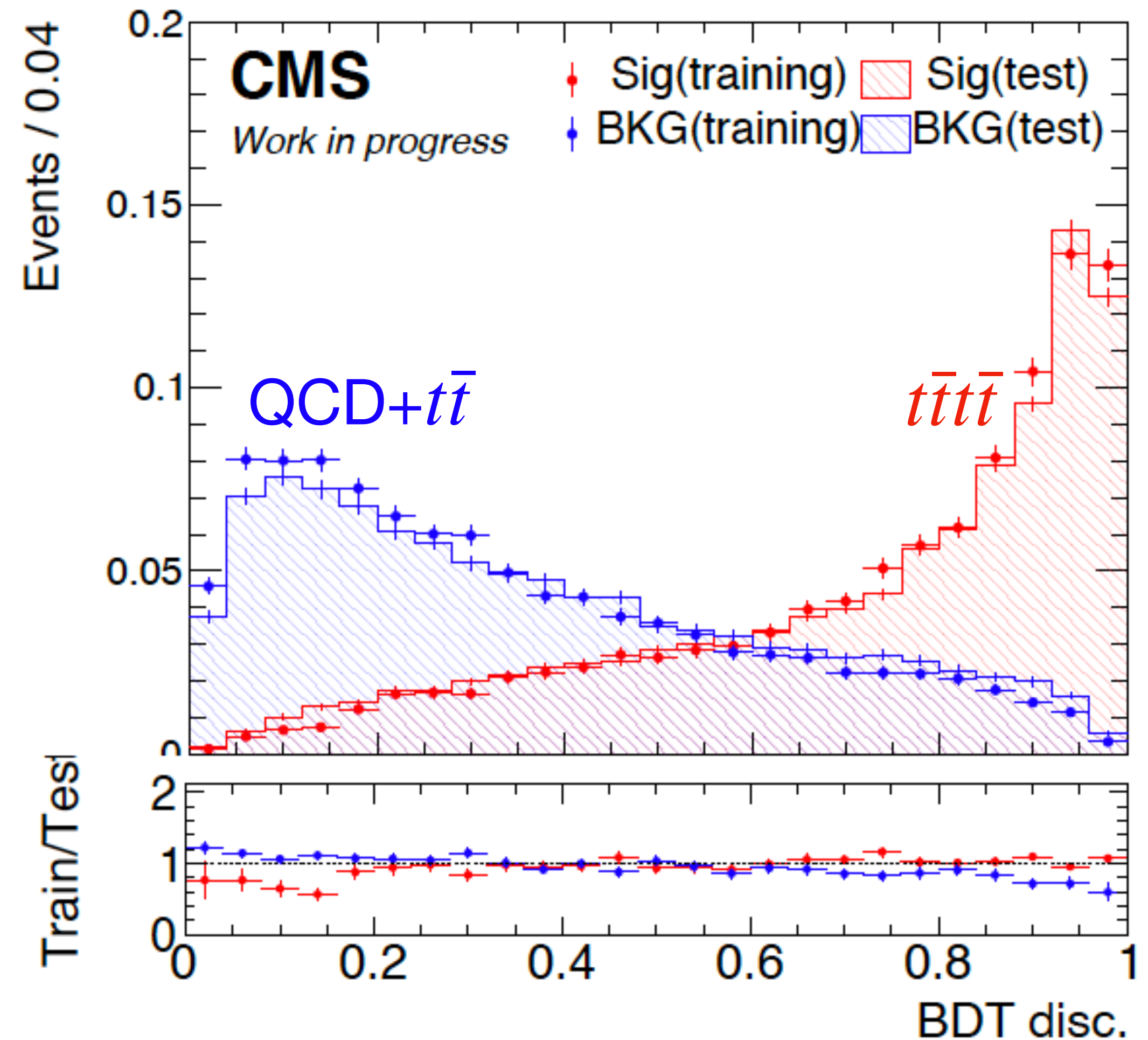


ML-based data-driven estimation of major backgrounds!



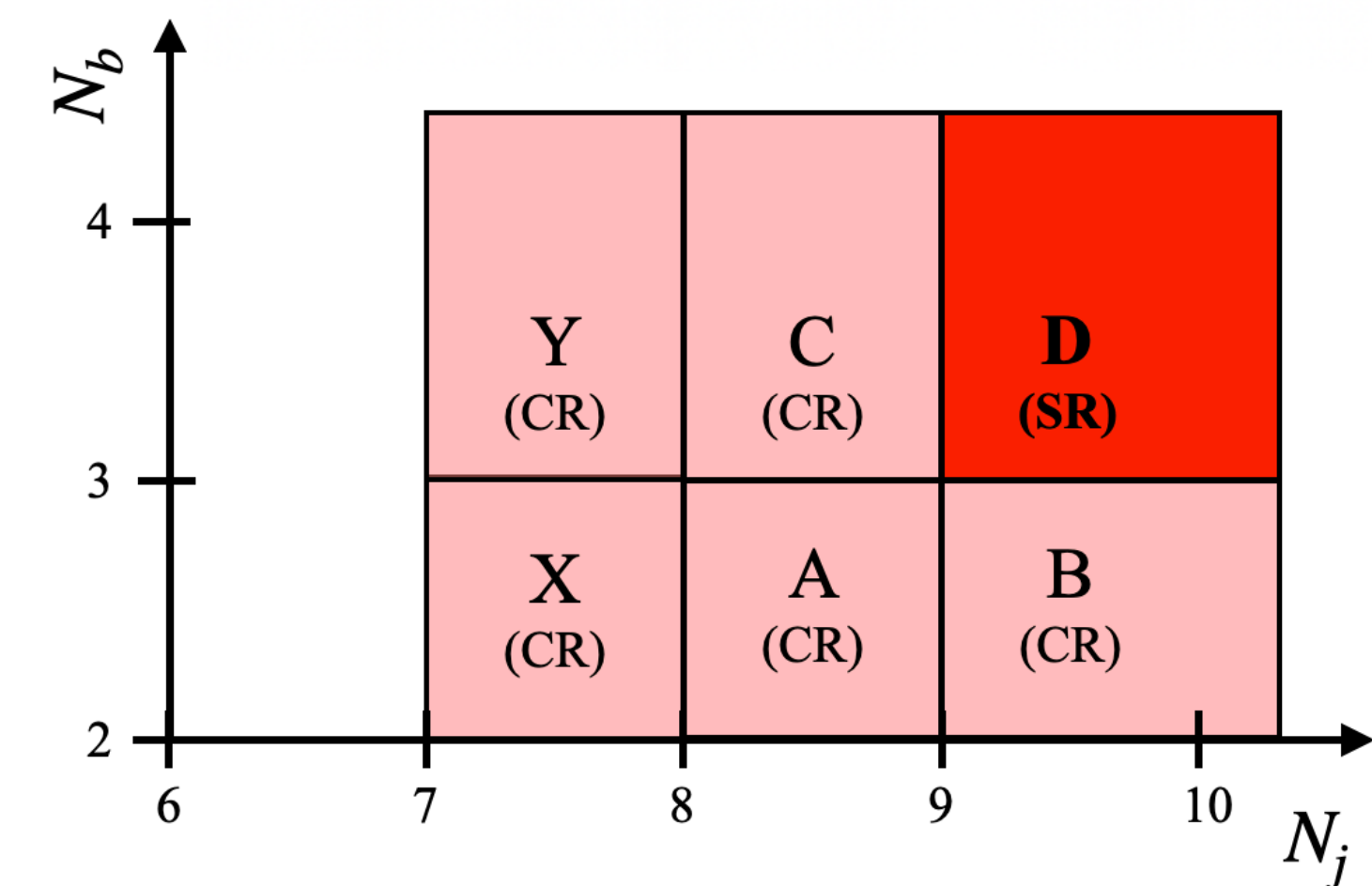
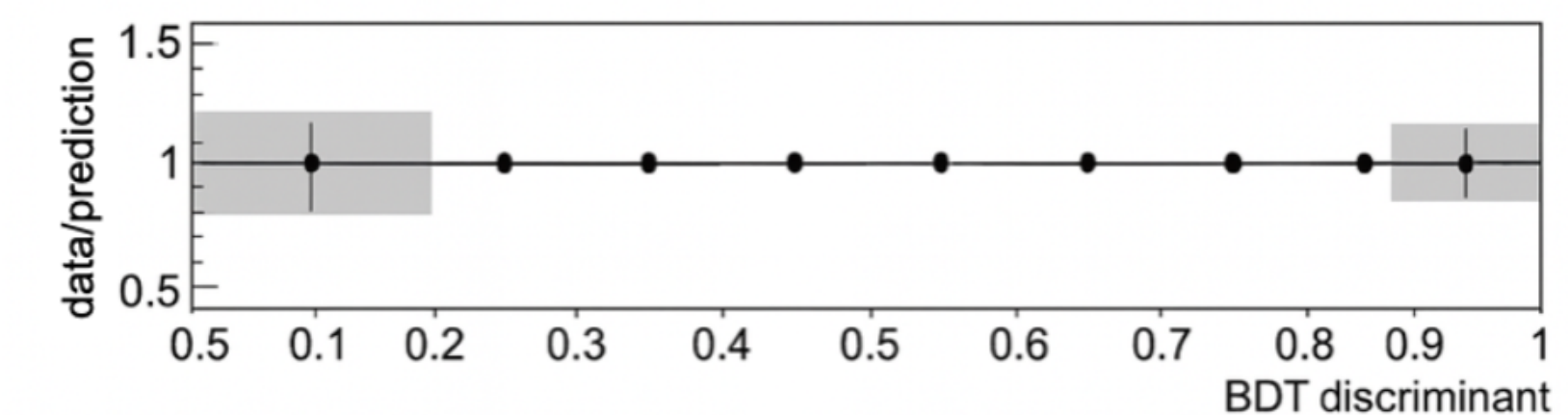
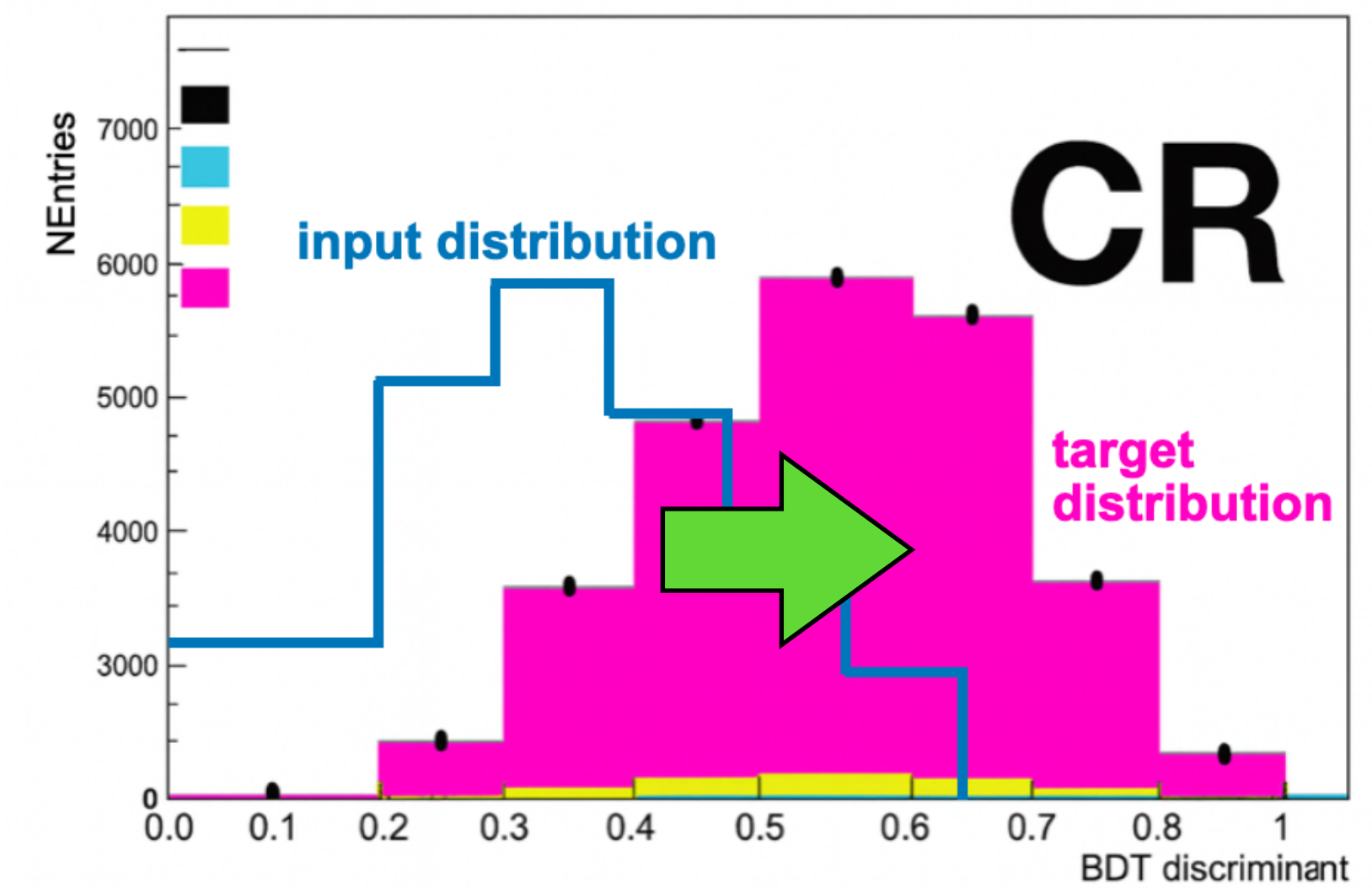
Event-Level Classifier

- **BDT event-level classifier:** boosted decision tree trained to discriminate $t\bar{t}t\bar{t}$ vs. QCD multijet and $t\bar{t}$ backgrounds
- **Problem:** Could not model QCD multijet background in simulation
- Need way of estimating BDT discriminant **shape** of major backgrounds from data



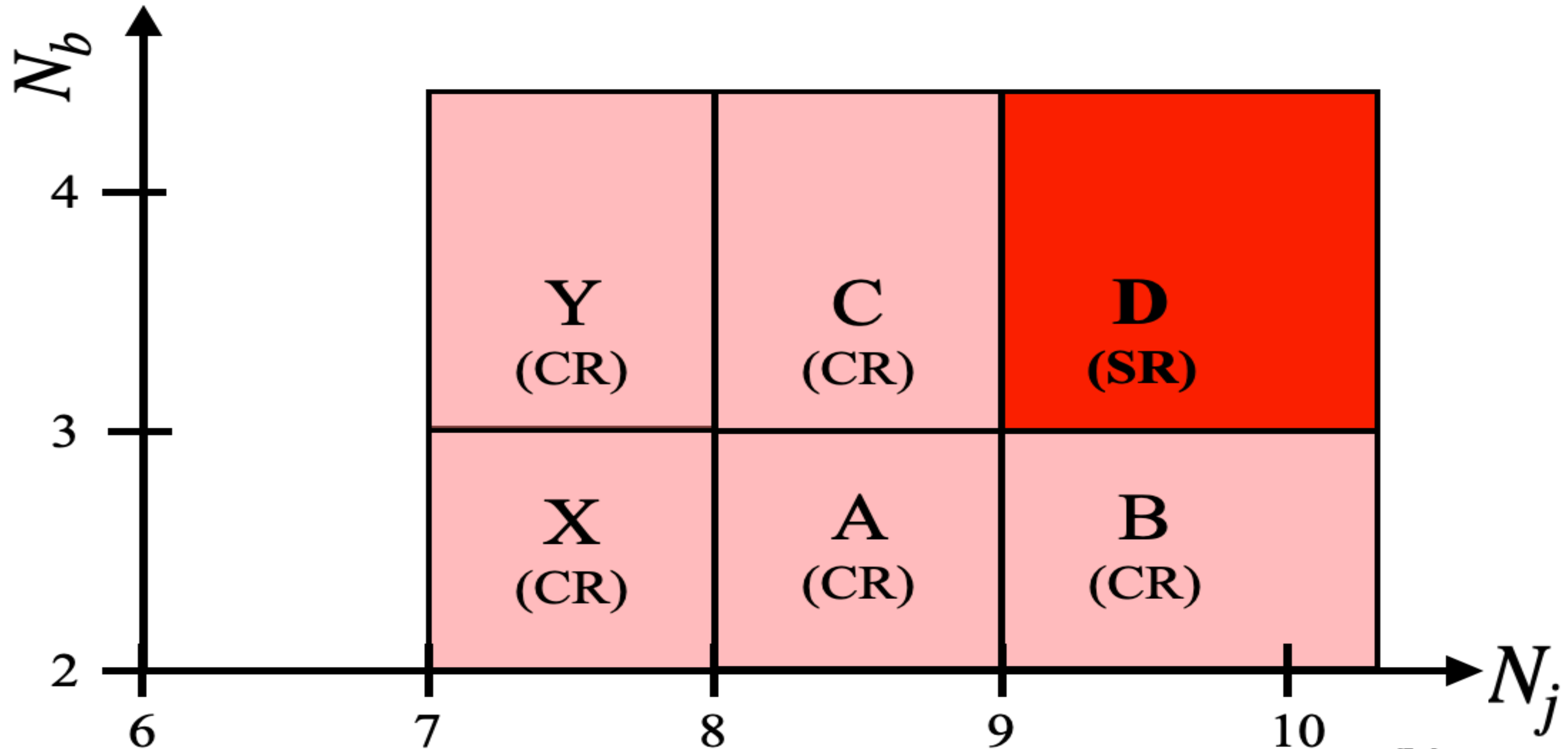
NN Background Estimation

- **Goal:** predict BDT shape of $\text{QCD}+t\bar{t}$ backgrounds from data
- Neural autoregressive flows (NAFs) are a form of normalizing flow (NF)
 - Learn **transformation** τ between **input** ($t\bar{t}$) and **target** (data) distributions in order to predict **feature variables** (BDT discriminant)
 - τ modified autoregressively by current and previous **conditional variables** (control regions)
 - **→ Learns how transformation varies across control regions in order to apply to signal region**



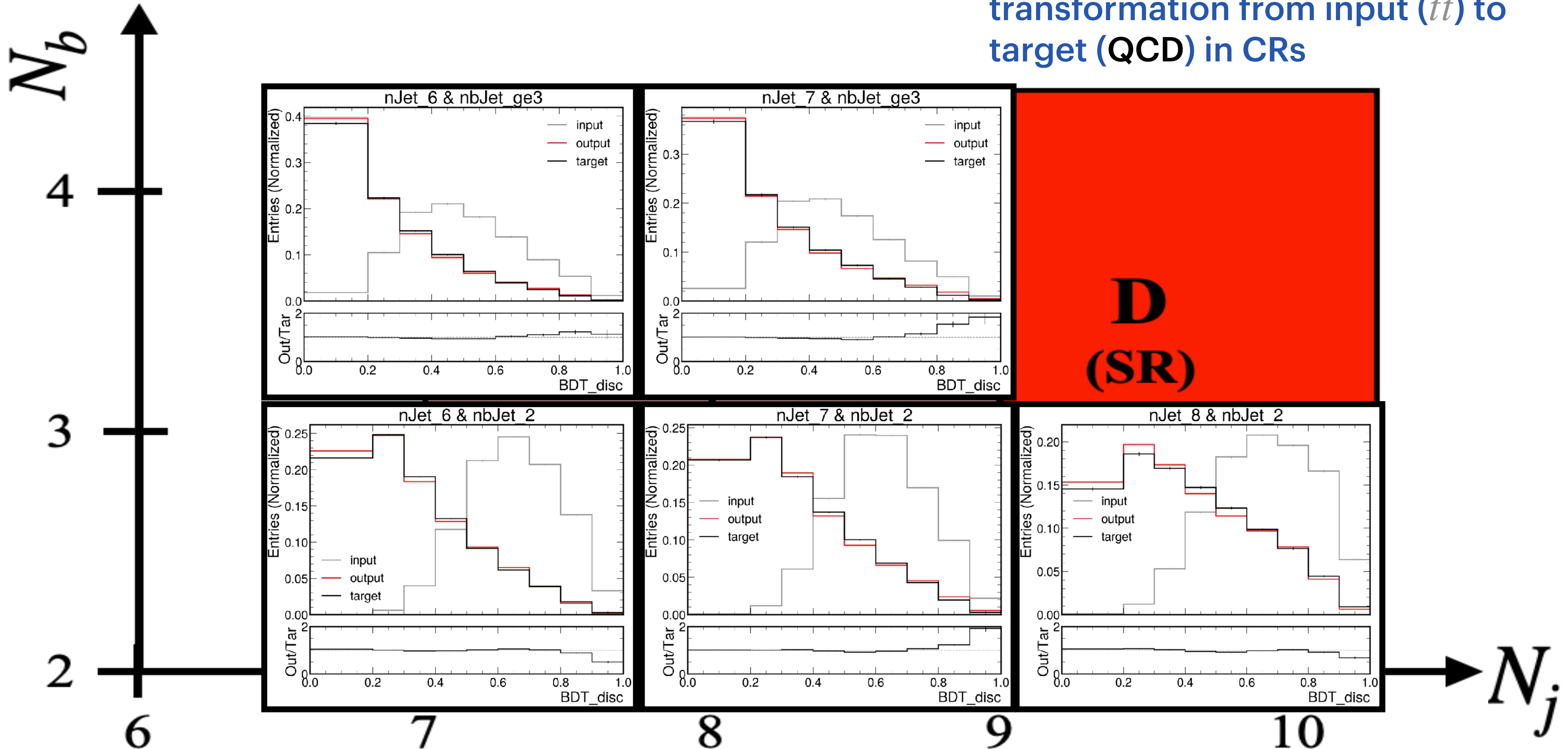
SR & associated CRs

NN Background Estimation



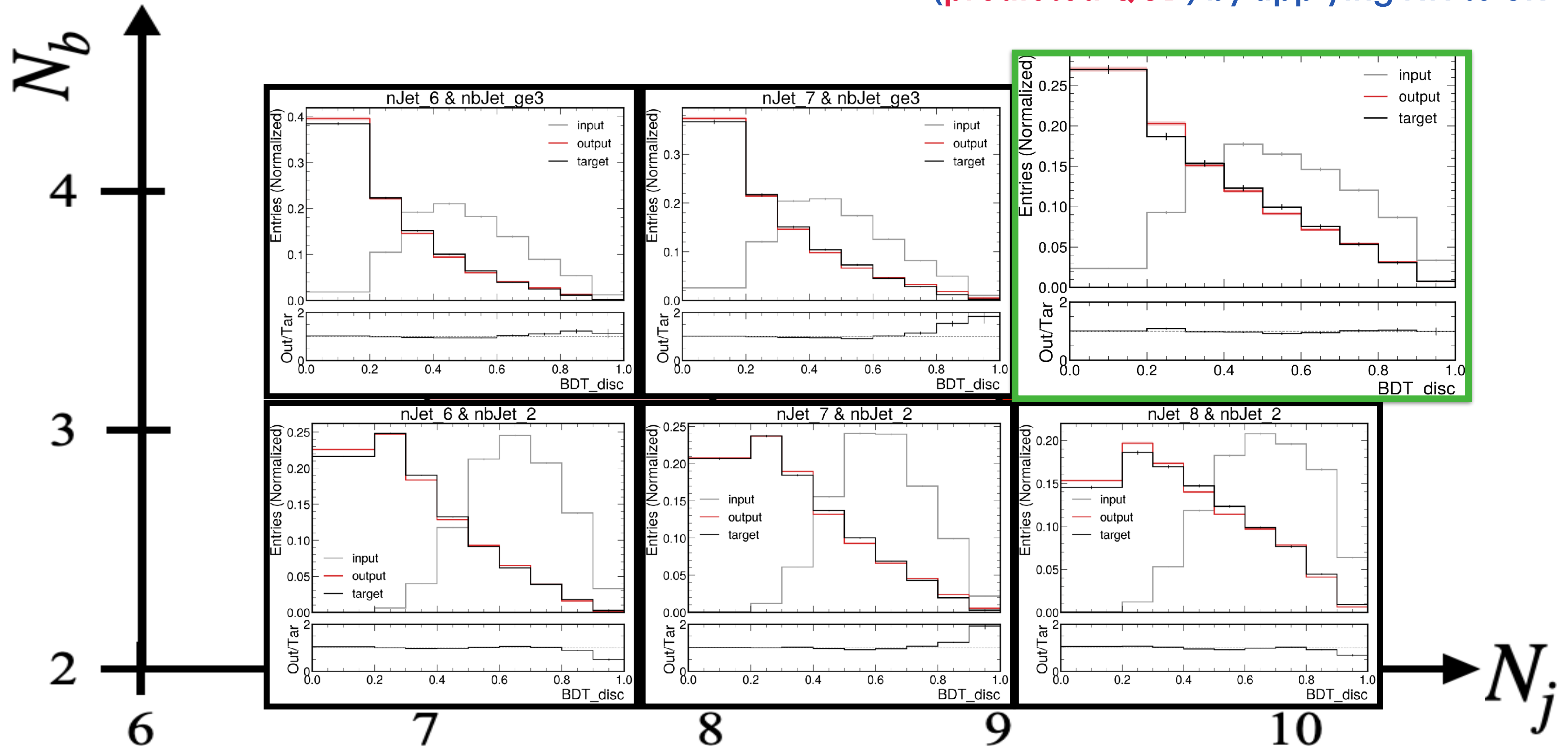
NN Background Estimation

1) Train NN to predict output
(predicted QCD) by learning
transformation from input ($t\bar{t}$) to
target (QCD) in CRs



NN Background Estimation

2) Infer final output prediction
(predicted QCD) by applying NN to SR



All-Hadronic Four-Top Production

- Analysis facilitated by this normalizing flow based data-driven background estimation technique
- Developed for [first all-hadronic four-top analysis](#)
- Also in upcoming Run 3 version of this analysis
- Potentially applicable to other hadronic signatures
- ML paper and corresponding public git repository coming soon!

PhD Thesis
UCSB

CMS-TOP-21-005

Supervised!

UCSD

CMS-MLG-23-004

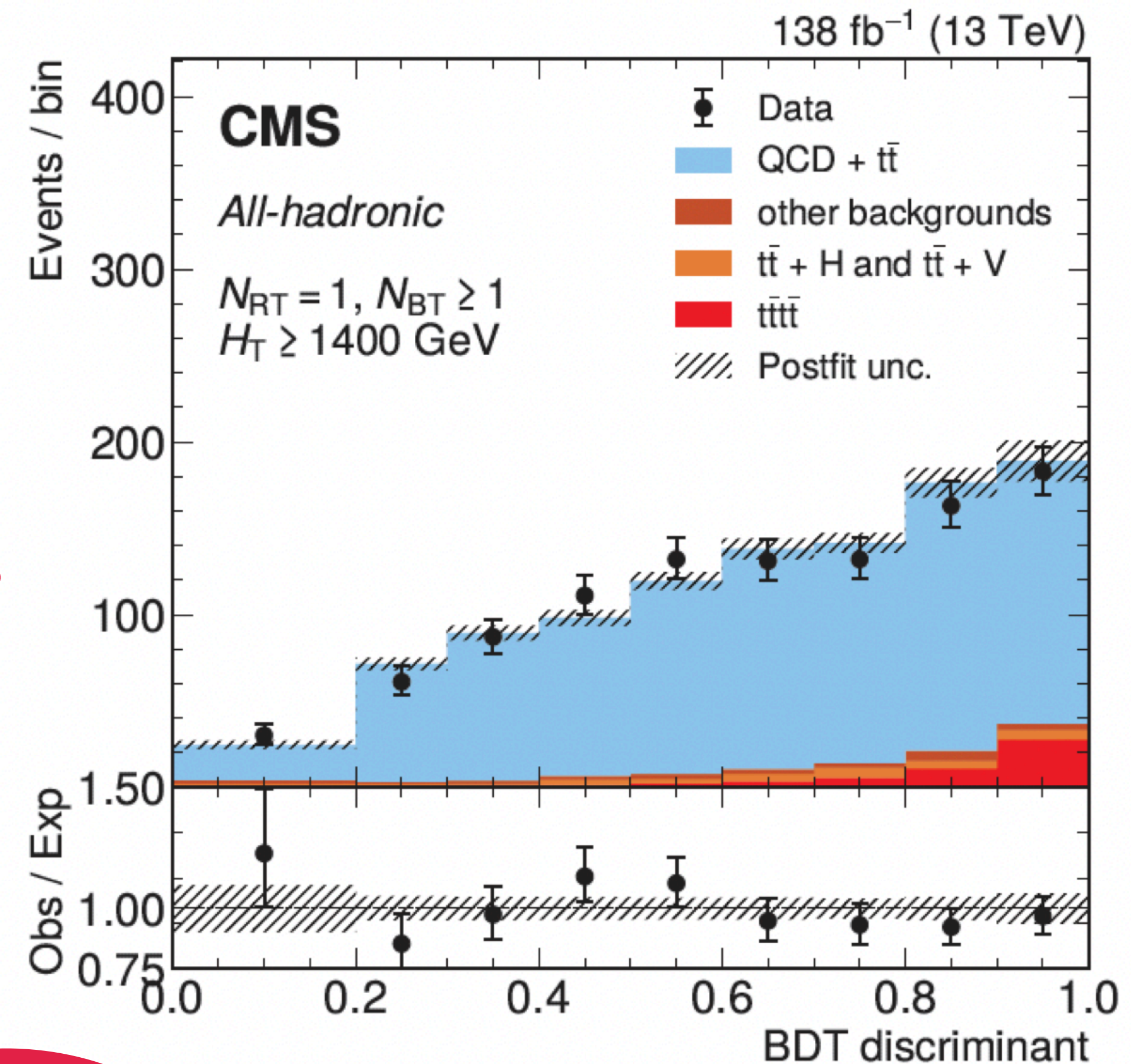
Postdoc work

UCSD

Supervised!



CMS-TOP-25-006



All-Hadronic Four-Top Production

- Analysis facilitated by this normalizing flow based data-driven background estimation technique

- Development

- Also

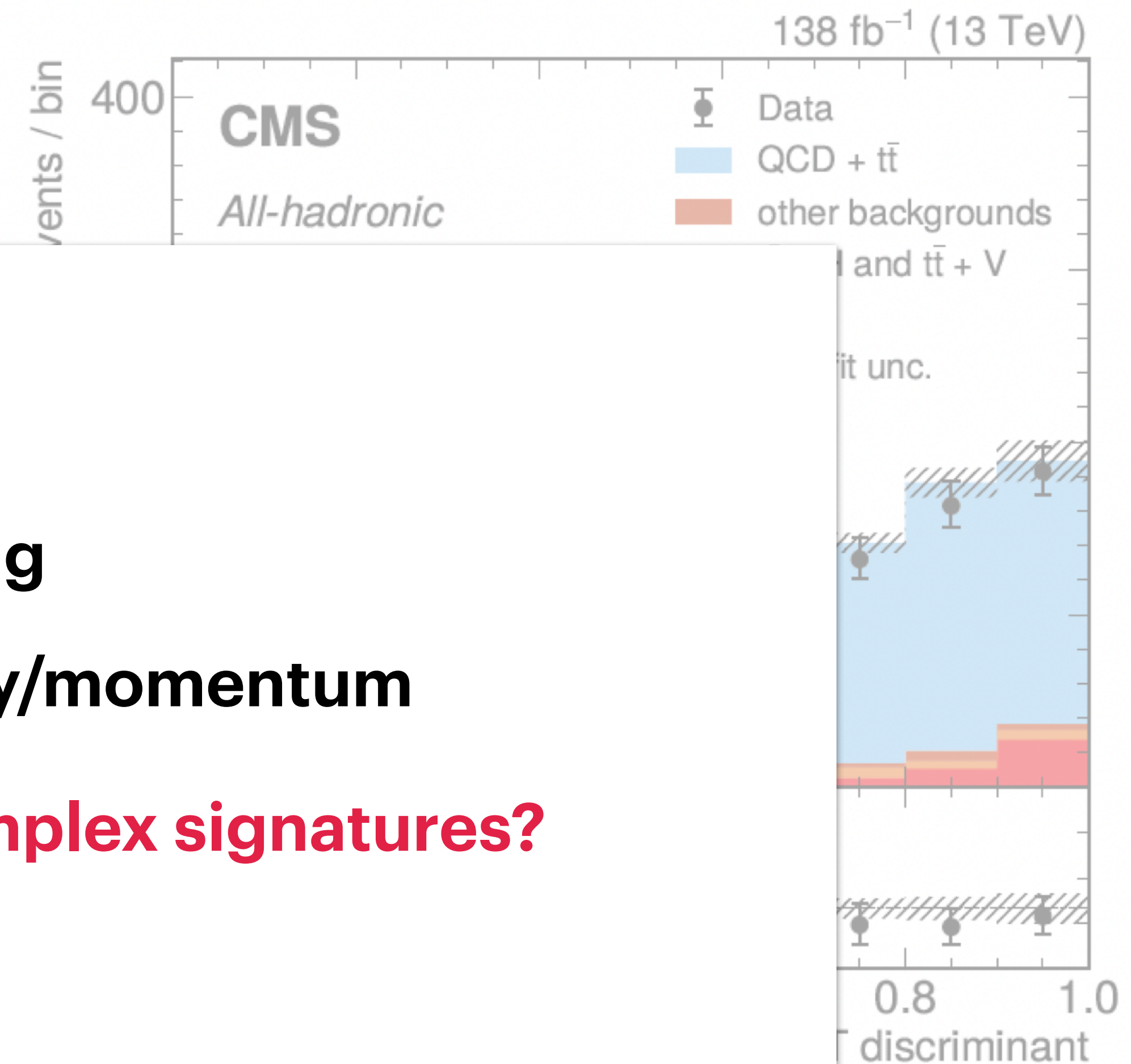
- Potential

- ML p

comi

Potential improvements:

- Could benefit from improved top tagging
- Limited by triggers at low jet multiplicity/momentum
 - → Improved trigger algorithms for complex signatures?



Supervised!

UCSD

Postdoc work

UCSD

CMS-MLG-23-004

Supervised!



CMS-TOP-25-006

All-Hadronic Four-Top Production

- Analysis facilitated by this normalizing flow based data-driven background estimation technique

- Development

- Also

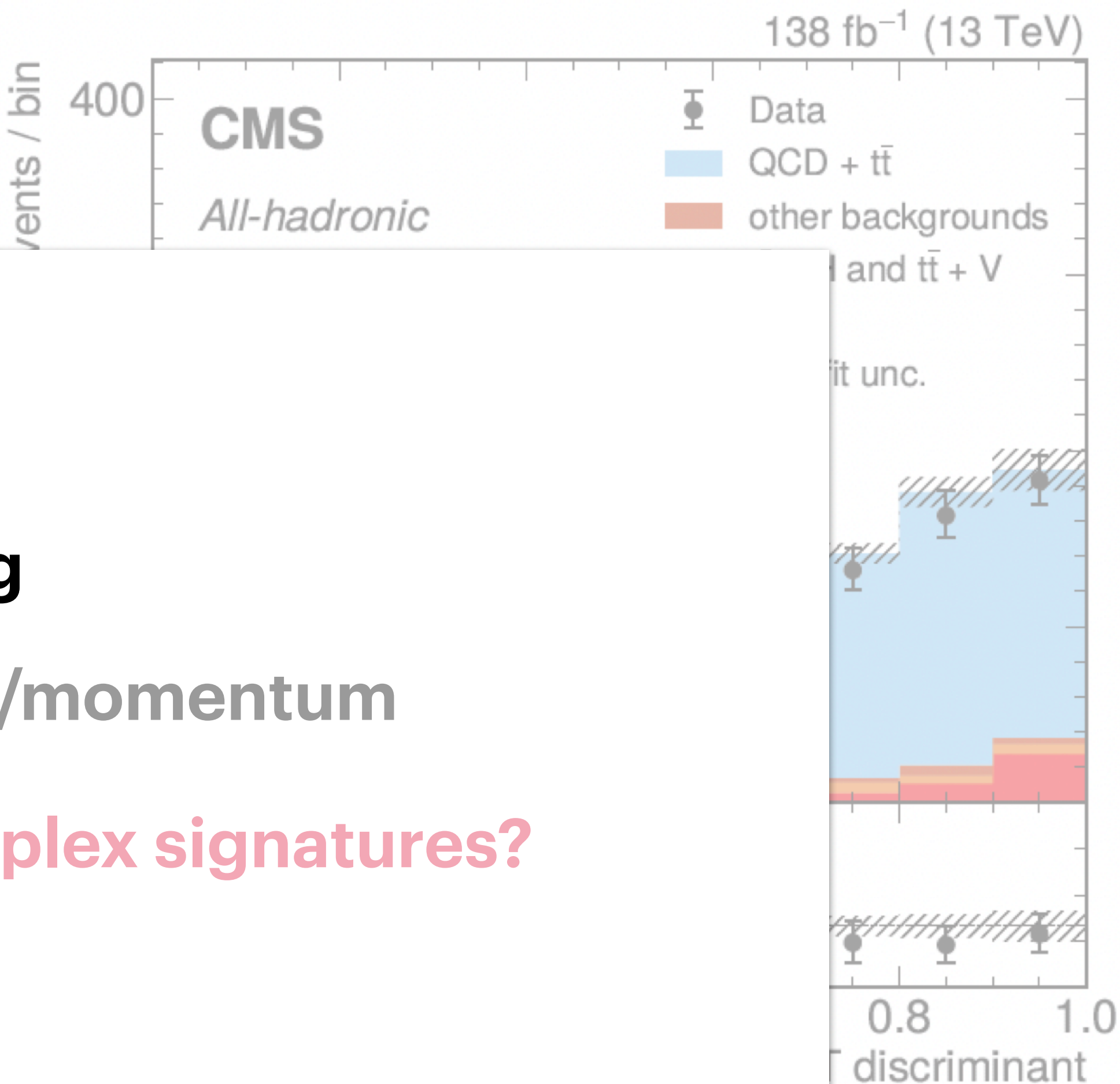
- Potential

- ML p

comi

Potential improvements:

- Could benefit from improved top tagging
- Limited by triggers at low jet multiplicity/momentum
 - → Improved trigger algorithms for complex signatures?



Supervised!

UCSD

Postdoc work

UCSD

CMS-MLG-23-004

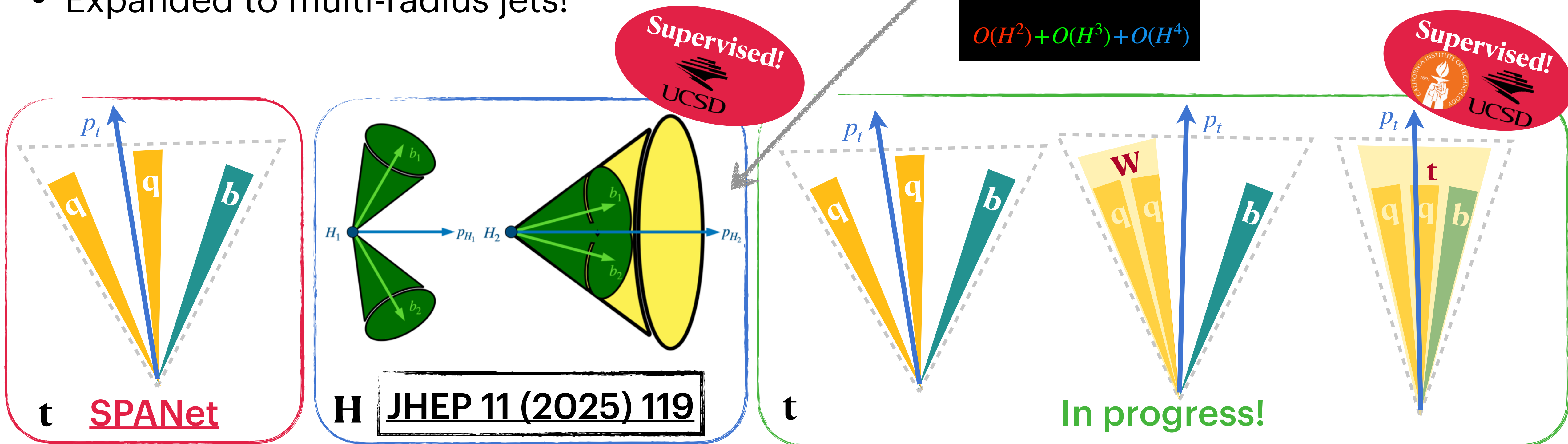
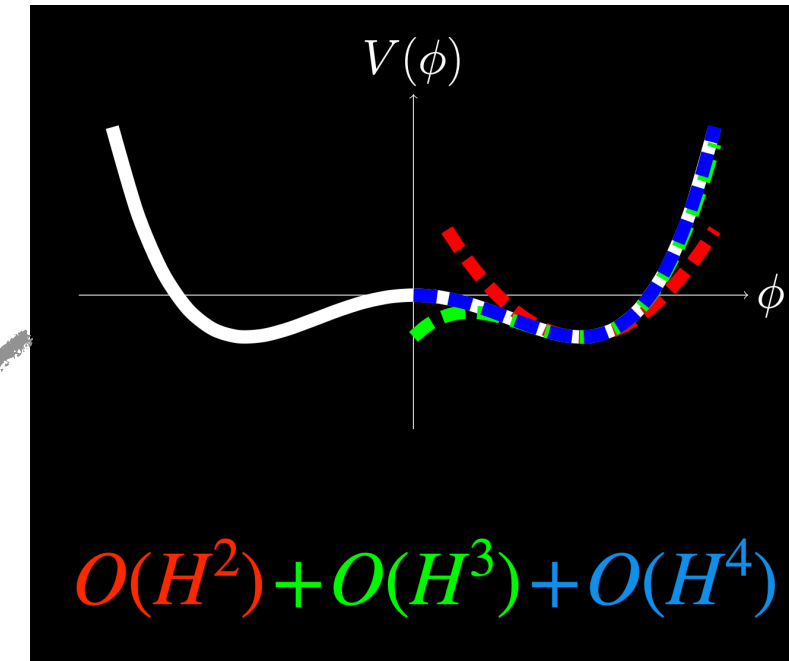
Supervised!



CMS-TOP-25-006

SPANet for topology-inclusive Jet Tagging

- Traditional jet assignment: try every possible assignment, pick the best by χ^2
 - Scales combinatorially, intractable in high-multiplicity final states
 - **SPANet**: “Symmetry Preserving Attention Network”
 - Permutationless, efficient, and variable set assignment
 - Expanded to multi-radius jets!
- 

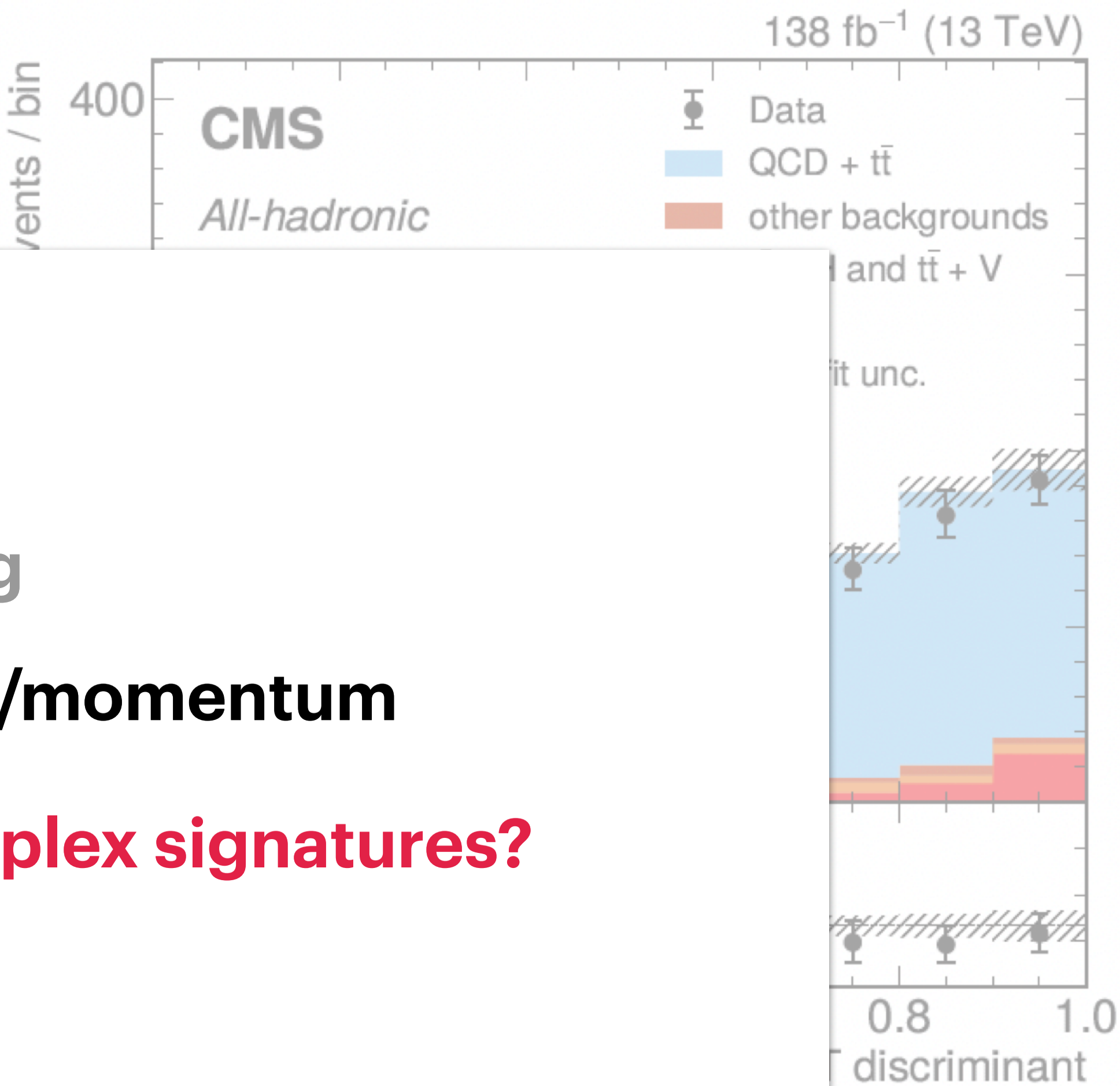


All-Hadronic Four-Top Production

- Analysis facilitated by this normalizing flow based data-driven background estimation technique

Potential improvements:

- Could benefit from improved top tagging
- Limited by triggers at low jet multiplicity/momentum
 - → Improved trigger algorithms for complex signatures?

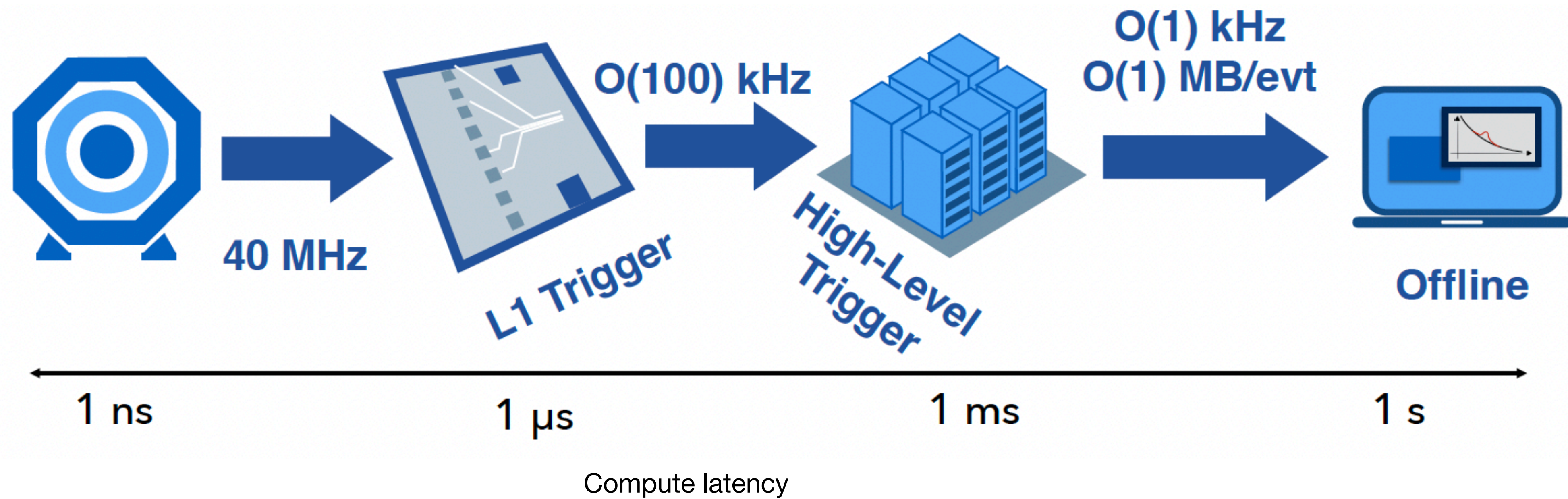


CMS-MLG-23-004



CMS-TOP-25-006

An analysis is only as good as its data!



Challenges:

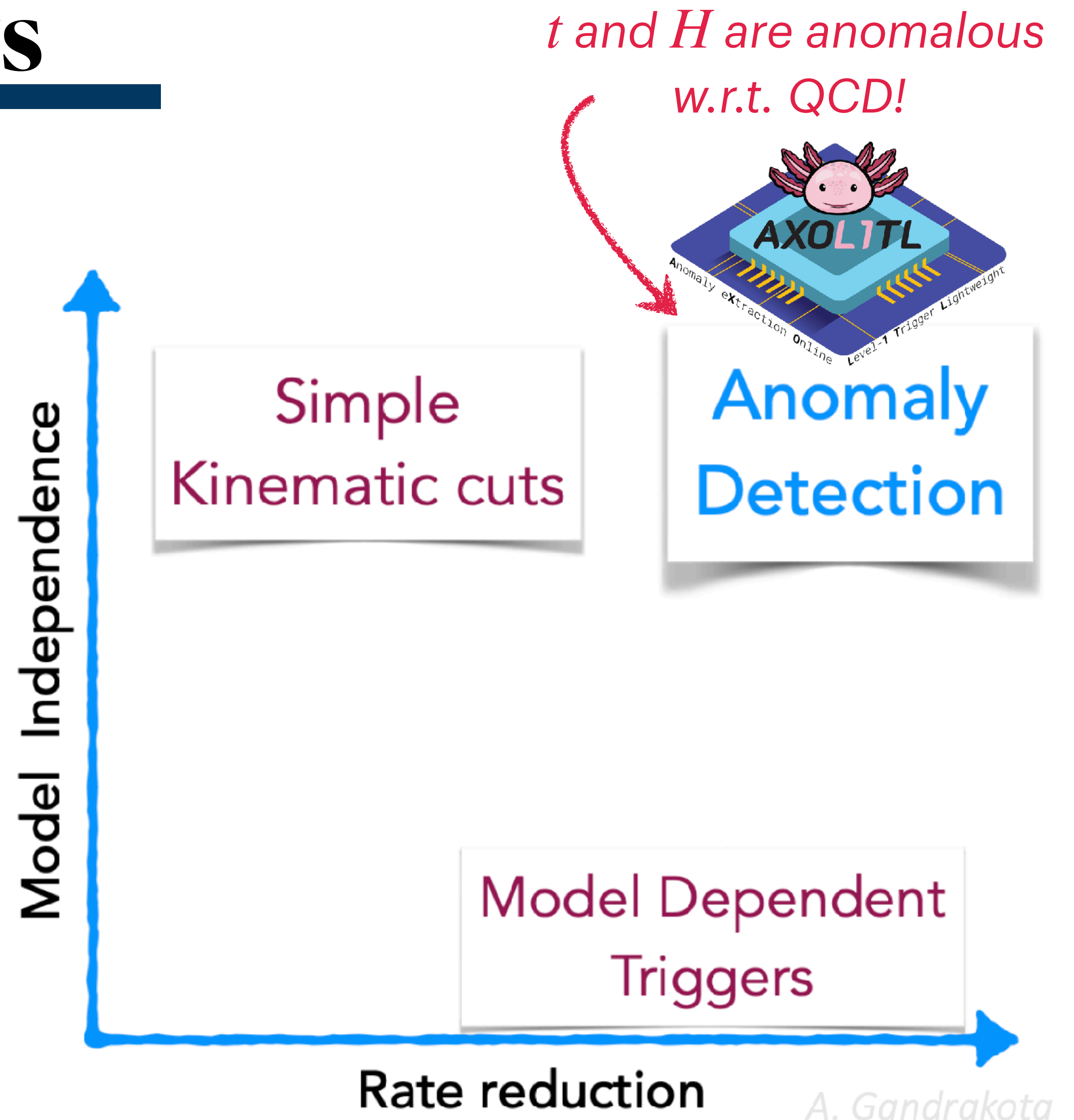
- Each collision produces $O(10^3)$ particles
- Detector has $O(10^8)$ sensors
- Extreme data rates $O(100 \text{ TB/s})$
- Exabyte scale datasets

**High rates, low
latency, low resource
utilization**

**Very large, complex
datasets**

ML for Hardware Triggers

- ML facilitates complex algorithms at the trigger level
- L1 Trigger rejects >98% events
- What if we are missing physics because we did not design a suitable trigger?
- Application: **Anomaly Detection**
- Detect new physics?
- Sensitive to a broad spectrum of SM/BSM signatures without redundancy or cut-based inefficiencies



Autoencoders

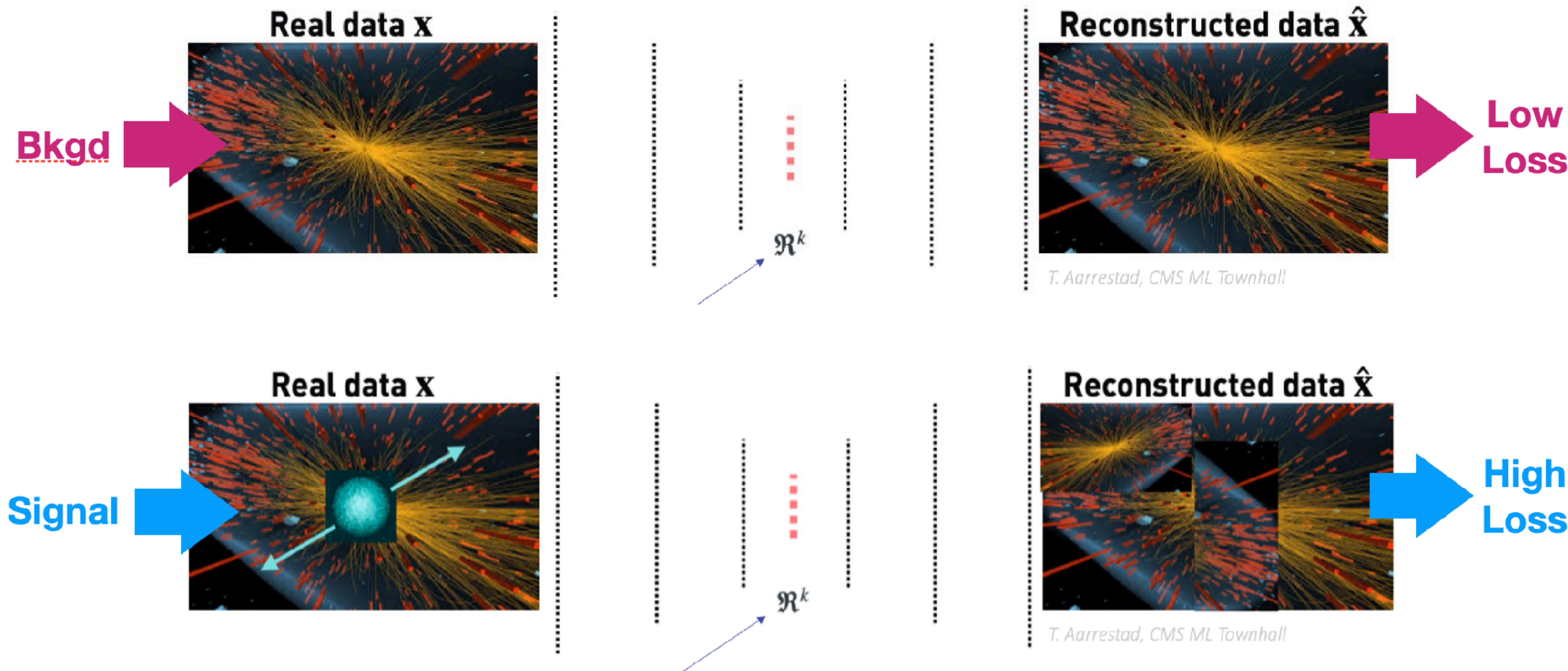
$$\text{Loss} = ||x - \hat{x}||^2$$

Reconstruction loss

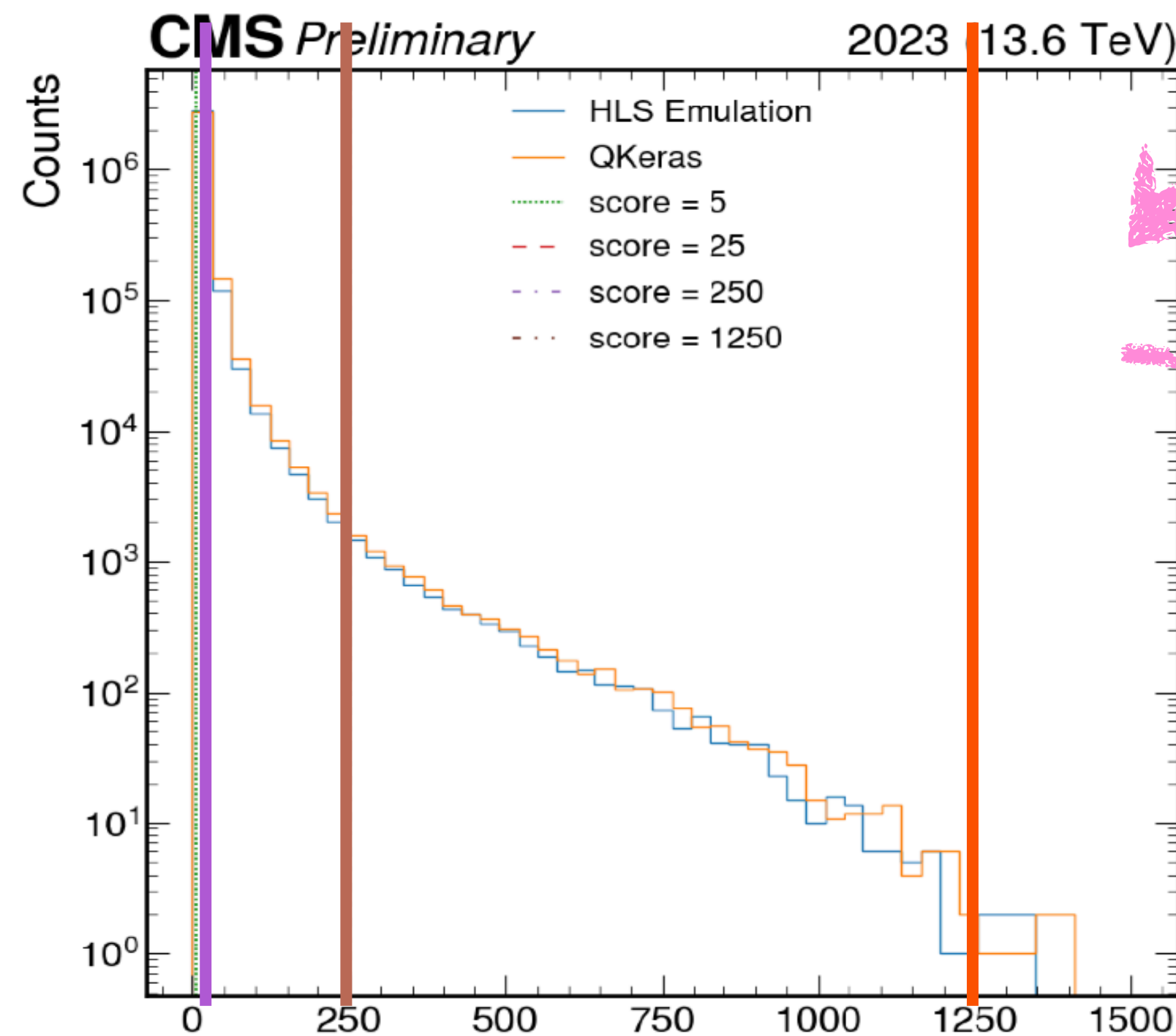
Train on ZeroBias LHC data

Bottleneck: autoencoder learns to compress high dimensional inputs into low dimensional latent space

$x - \hat{x}$ represents degree of abnormality



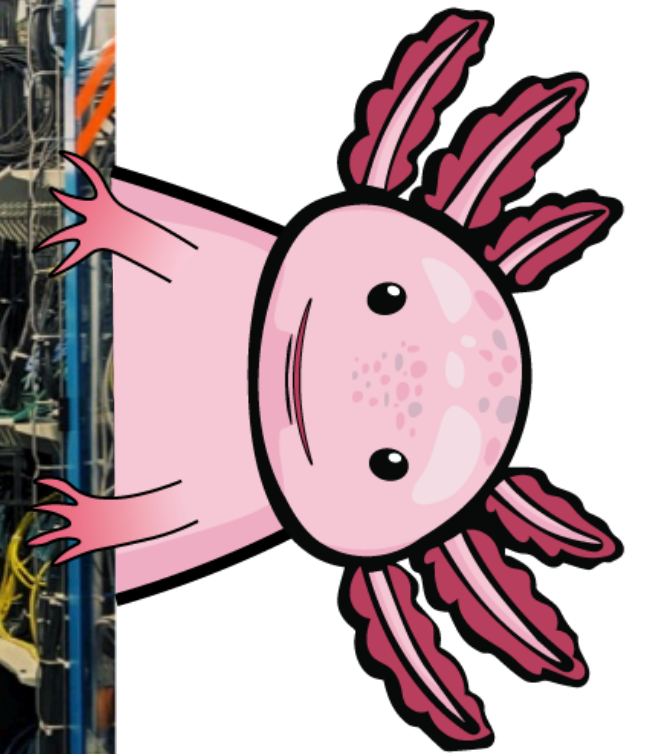
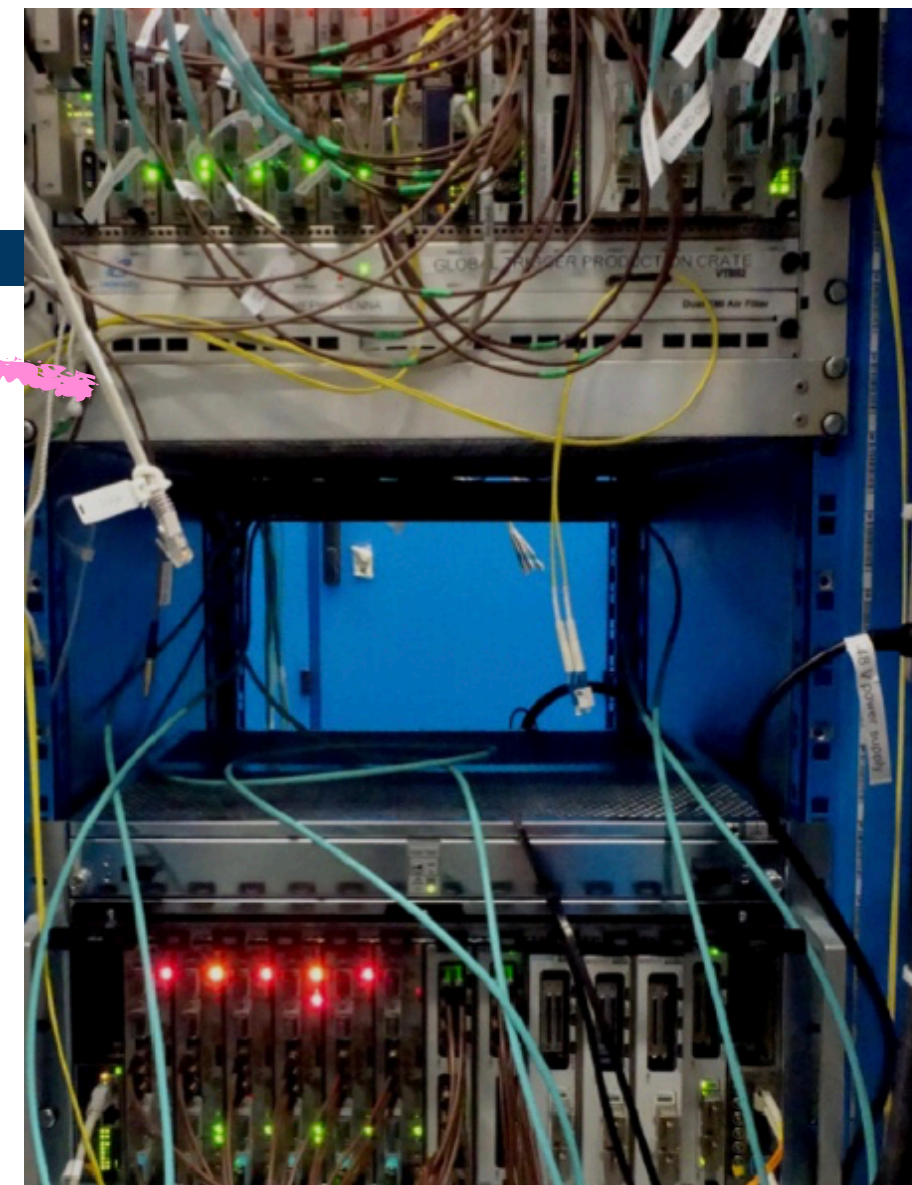
Deployment



$$\text{Loss} = ||x - \hat{x}||^2$$

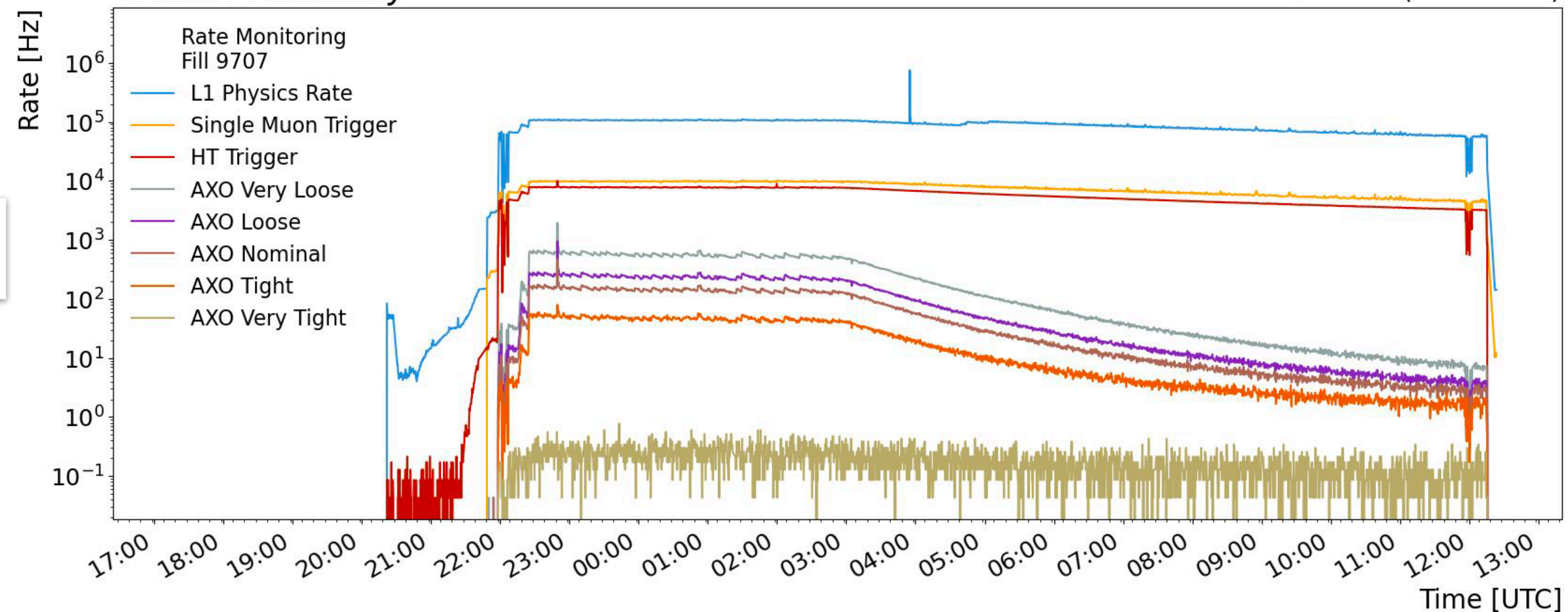
Reconstruction loss

- Thresholds target different trigger rates
- Rates stable relative to other L1 triggers

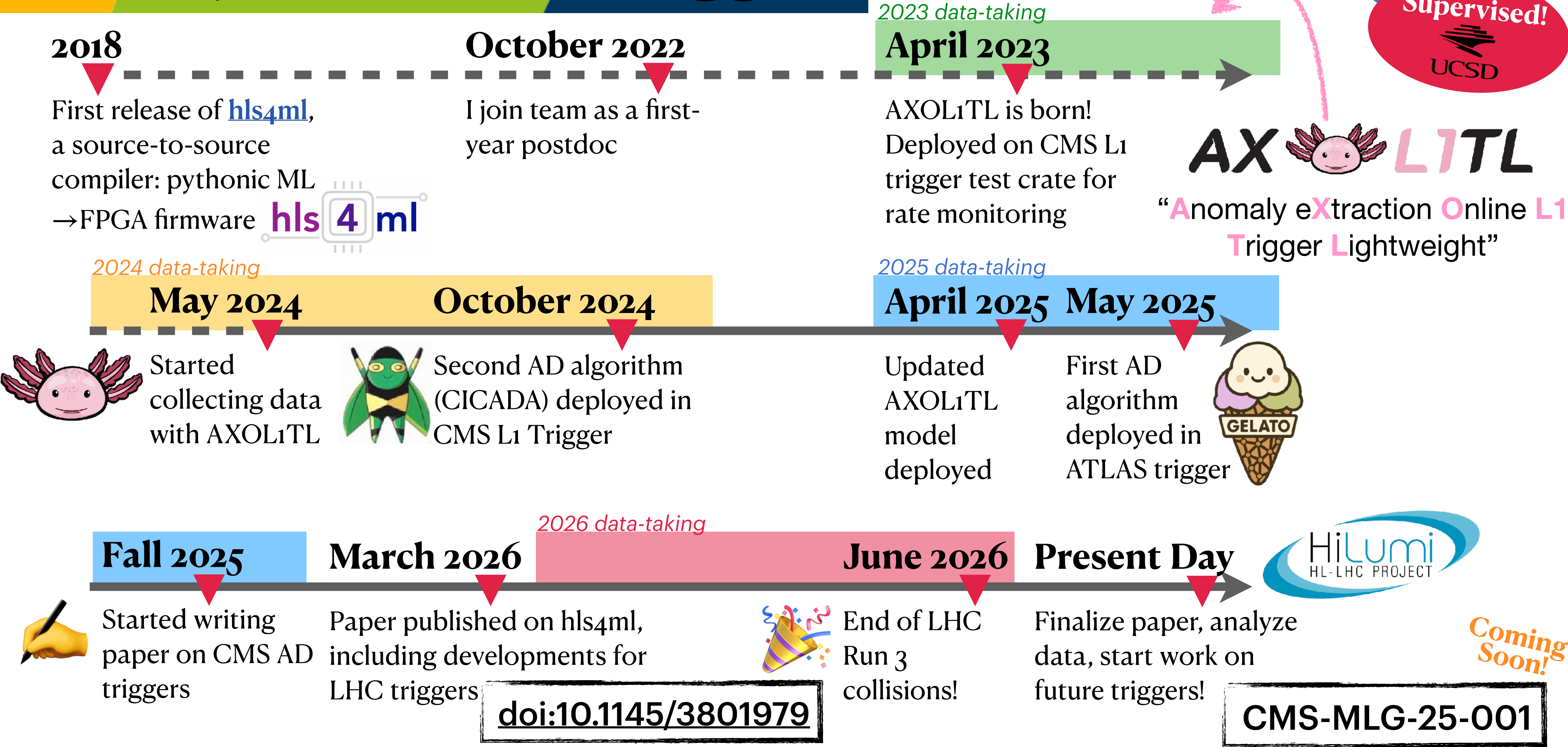


CMS Preliminary

0.767 fb⁻¹, 2024 (13.6 TeV)

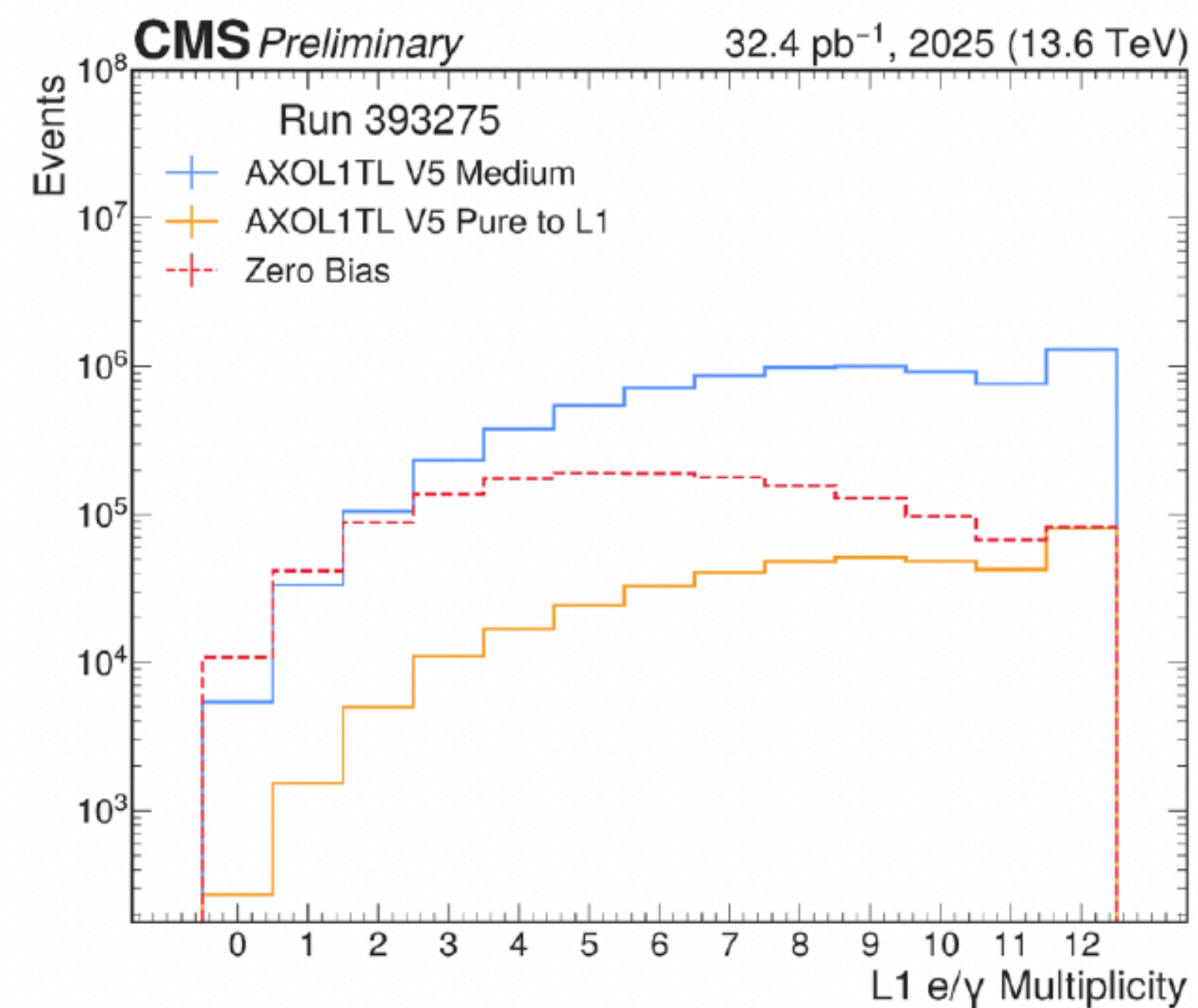
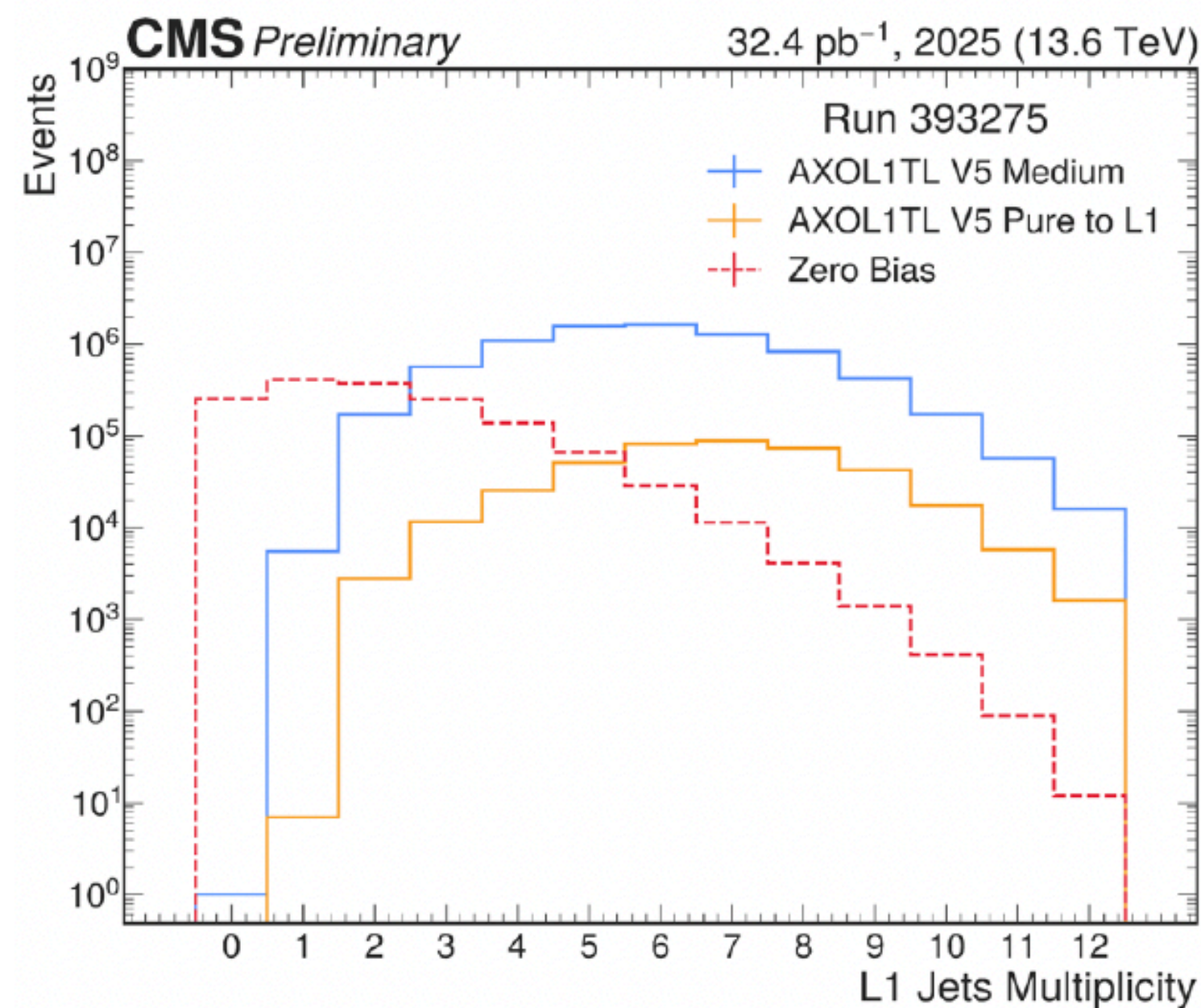
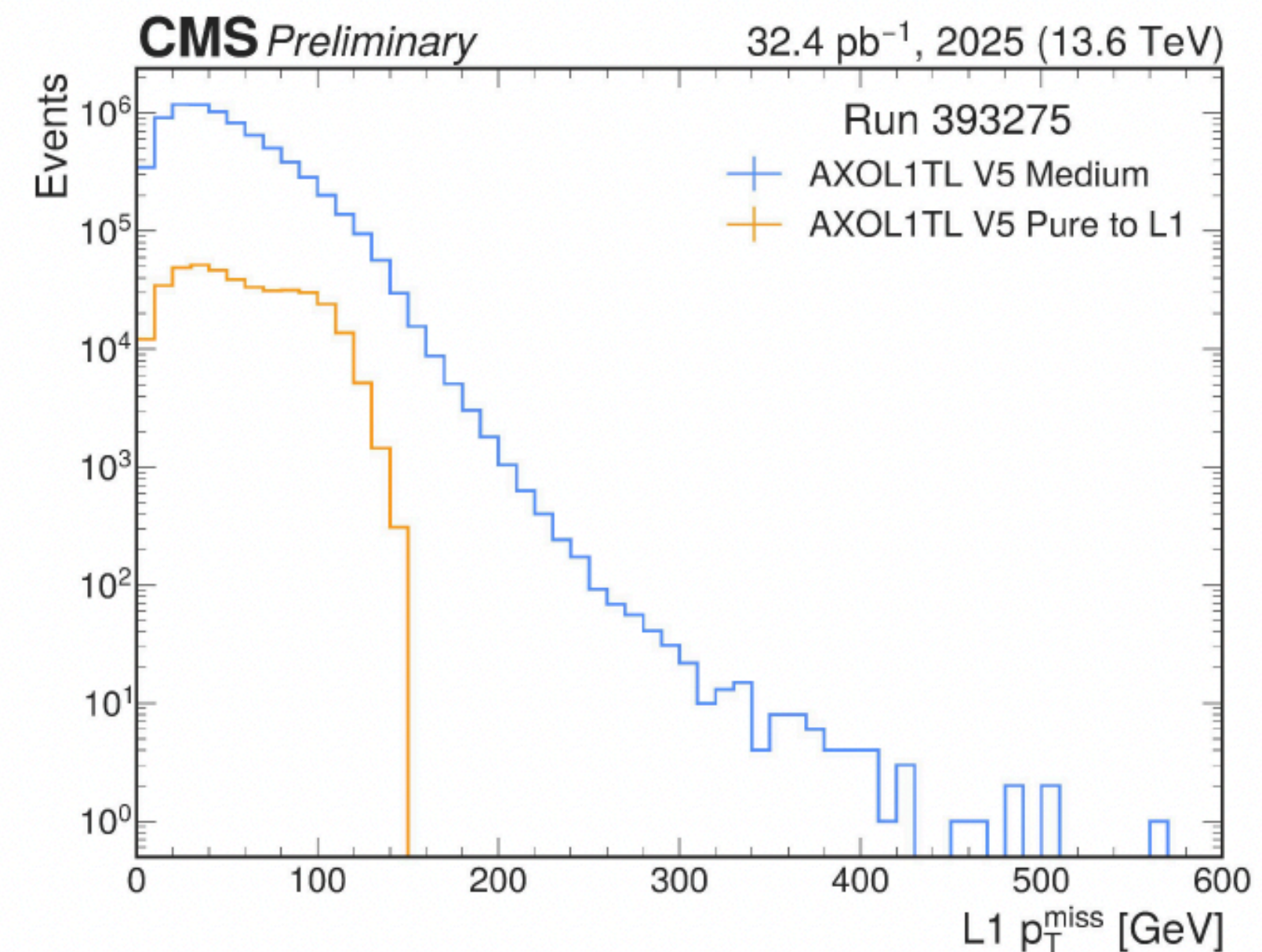
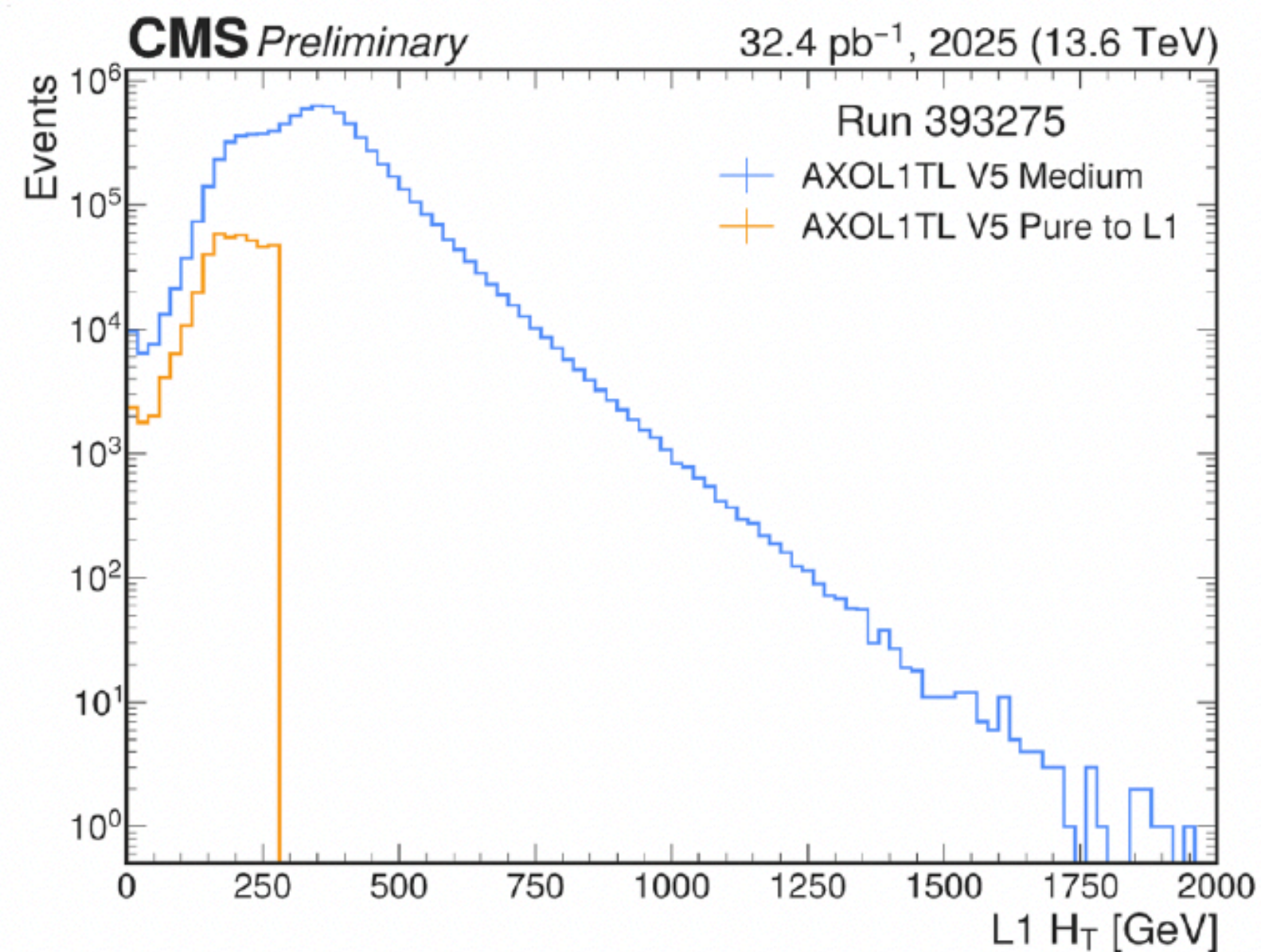
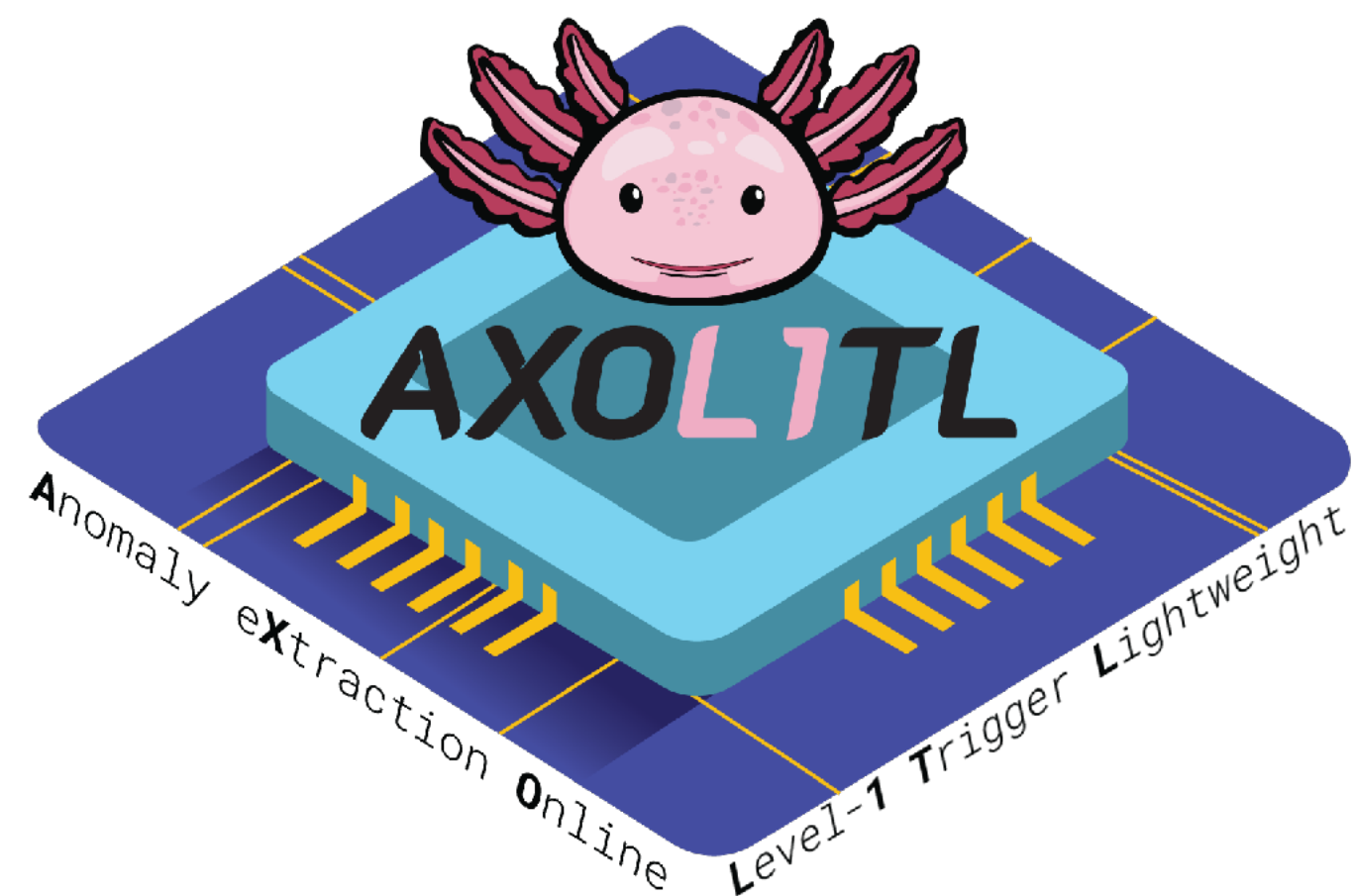


Anomaly Detection Trigger Timeline



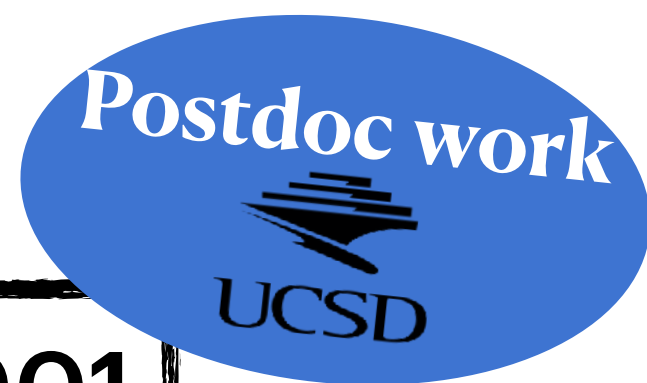
Performance

- Enhanced sensitivity to lower L1 H_T and p_T^{miss} values for pure events
- Preference for higher object multiplicity especially jets and e/γ

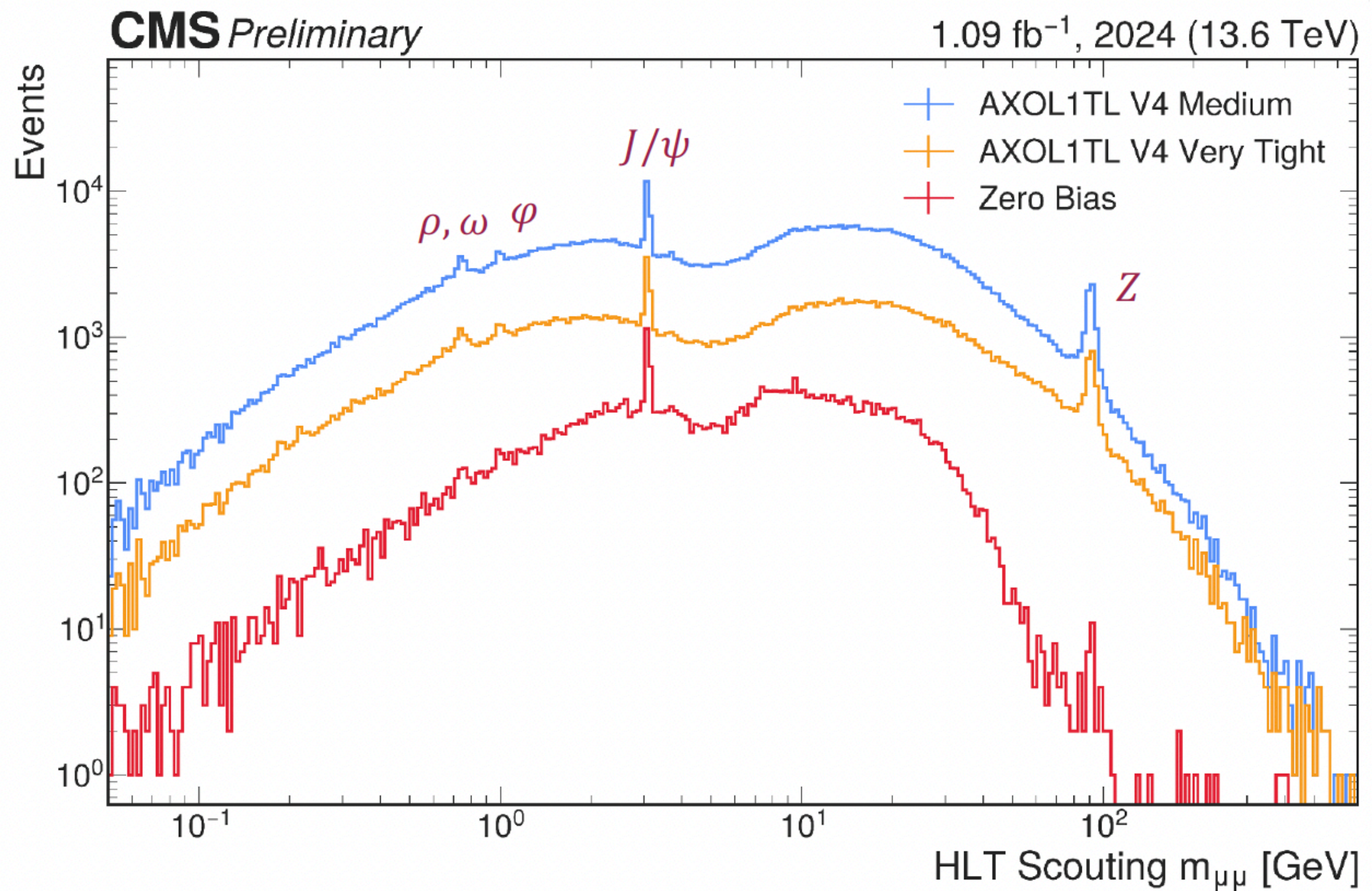
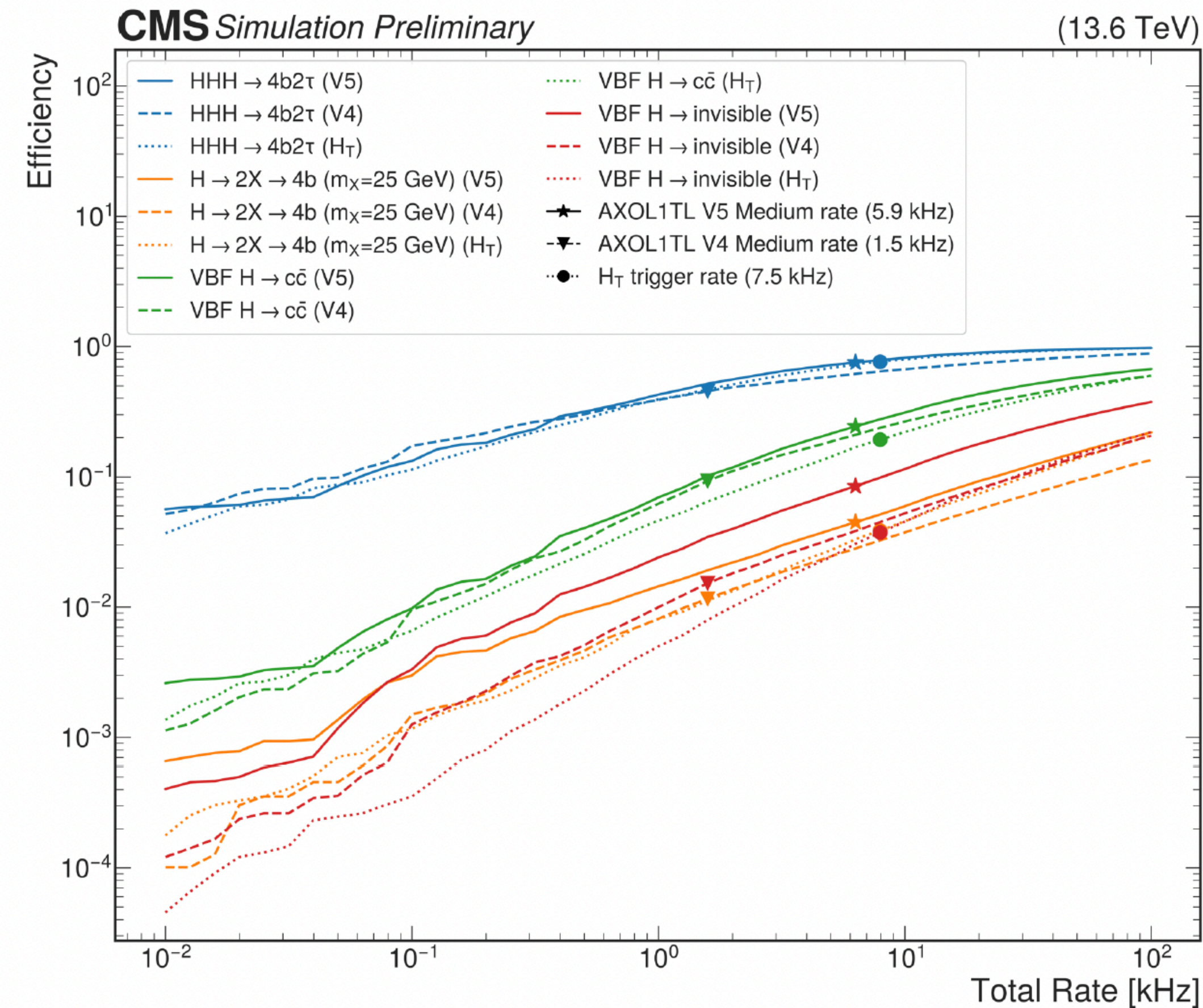


Physics Sensitivity

Coming Soon!



CMS-MLG-25-001



- Efficiency of simulated SM & BSM signals
- 2025 model more sensitive than 2024
- Competitive with standard H_T trigger

→ **Sensitivity to broad range of physics scenarios**

Future Triggers

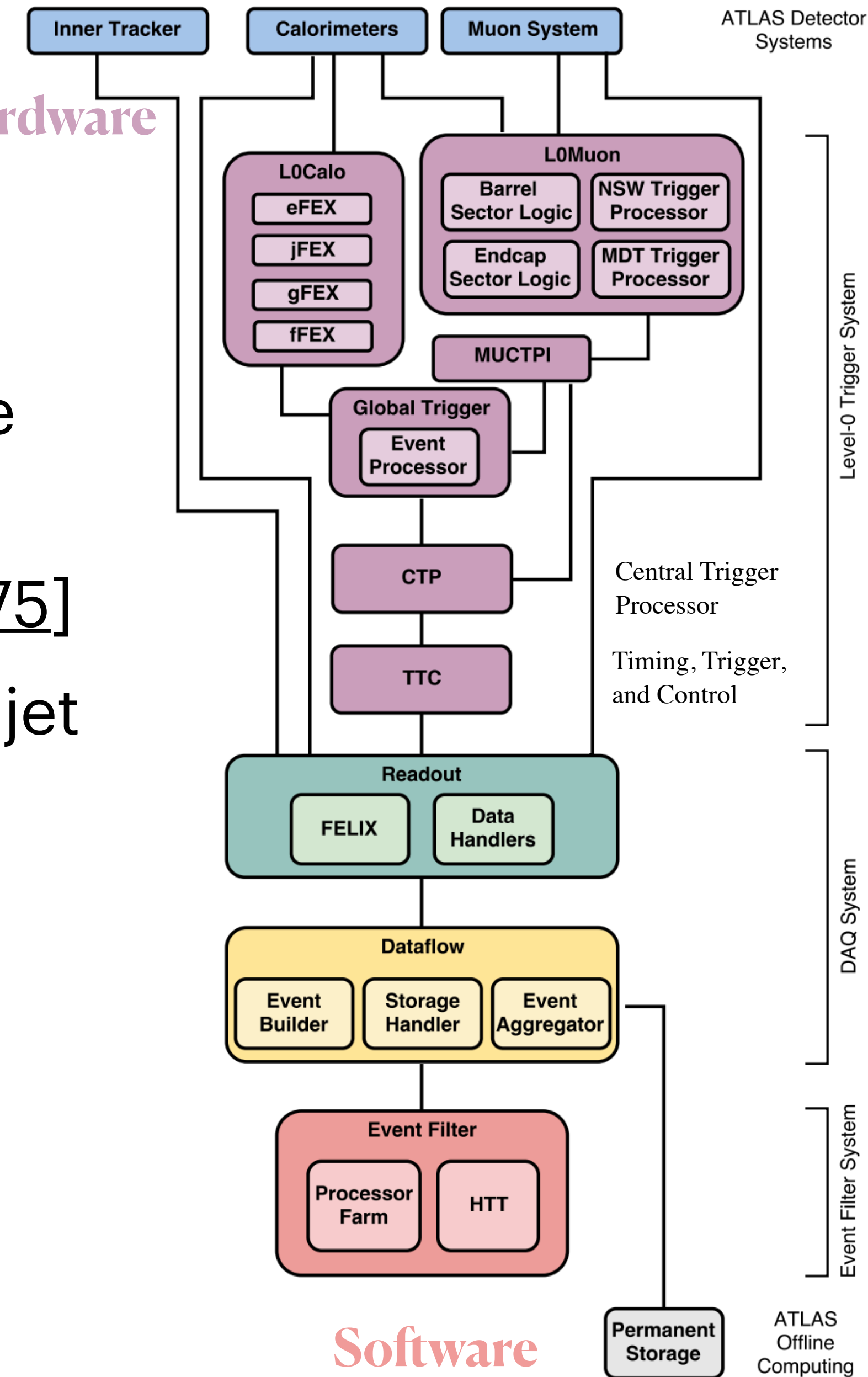
ATLAS Phase II TDAQ TDR

Phase-2 ATLAS TDAQ system

- **ML algorithms increasingly essential for Phase 2 hardware triggers in CMS & ATLAS**
- ATLAS Phase 2 Global Trigger: Time-multiplexing architecture makes full-event, offline-quality reconstruction feasible at L0
- hls4ml/ML based algorithms targeting GEP FPGAs [2406.12875]
 - b-tagging, VBF classification, MET regression, quark-gluon jet tagging...
- Both broad + targeted algorithms for improved efficiency + detection capabilities

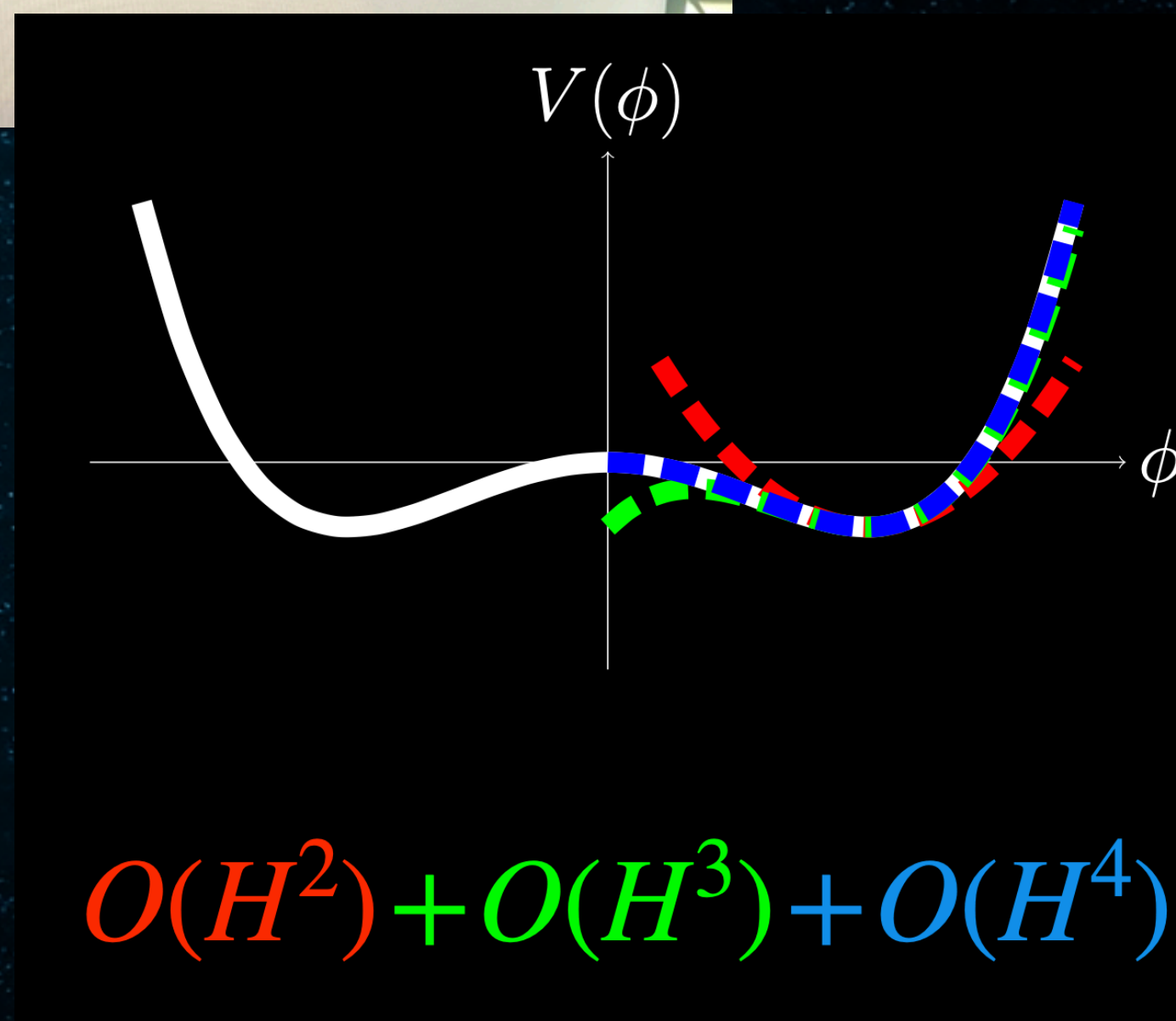
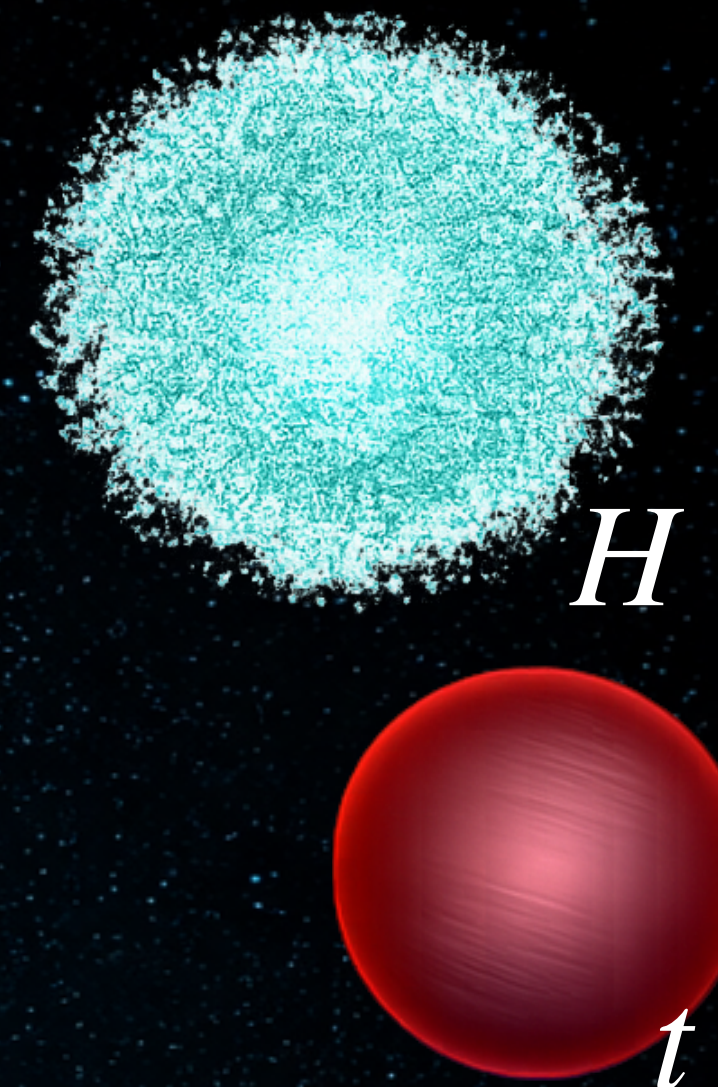
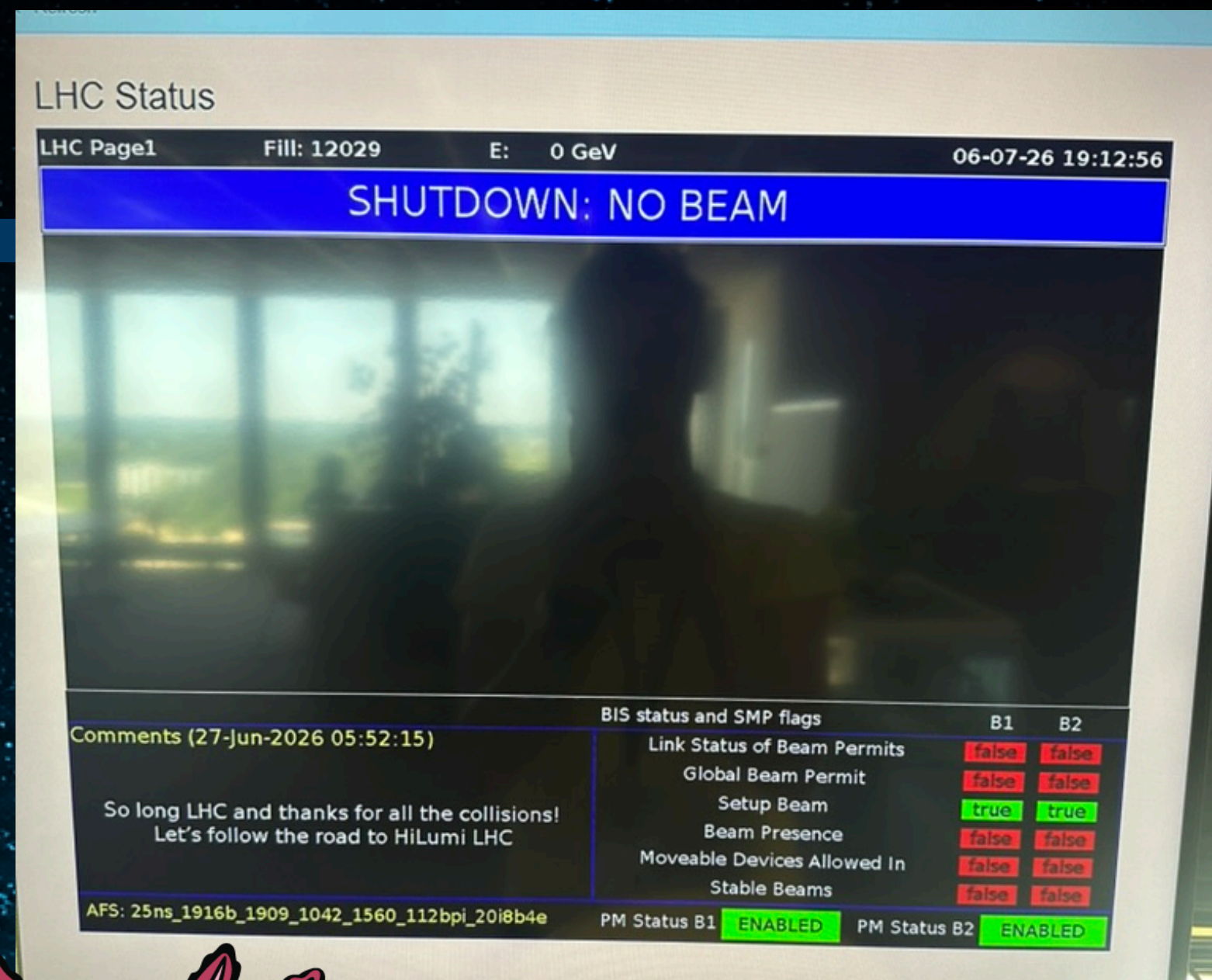


Hardware



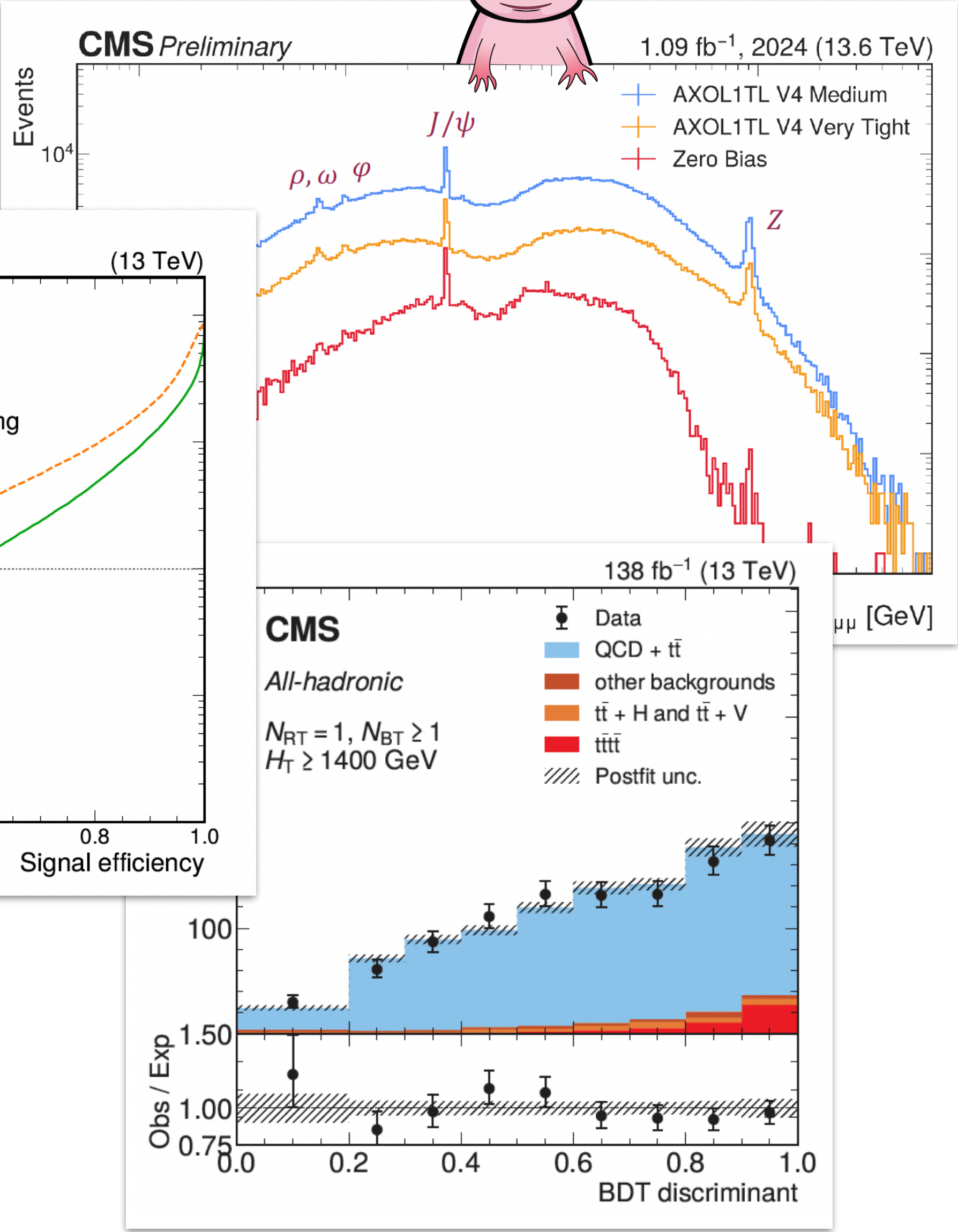
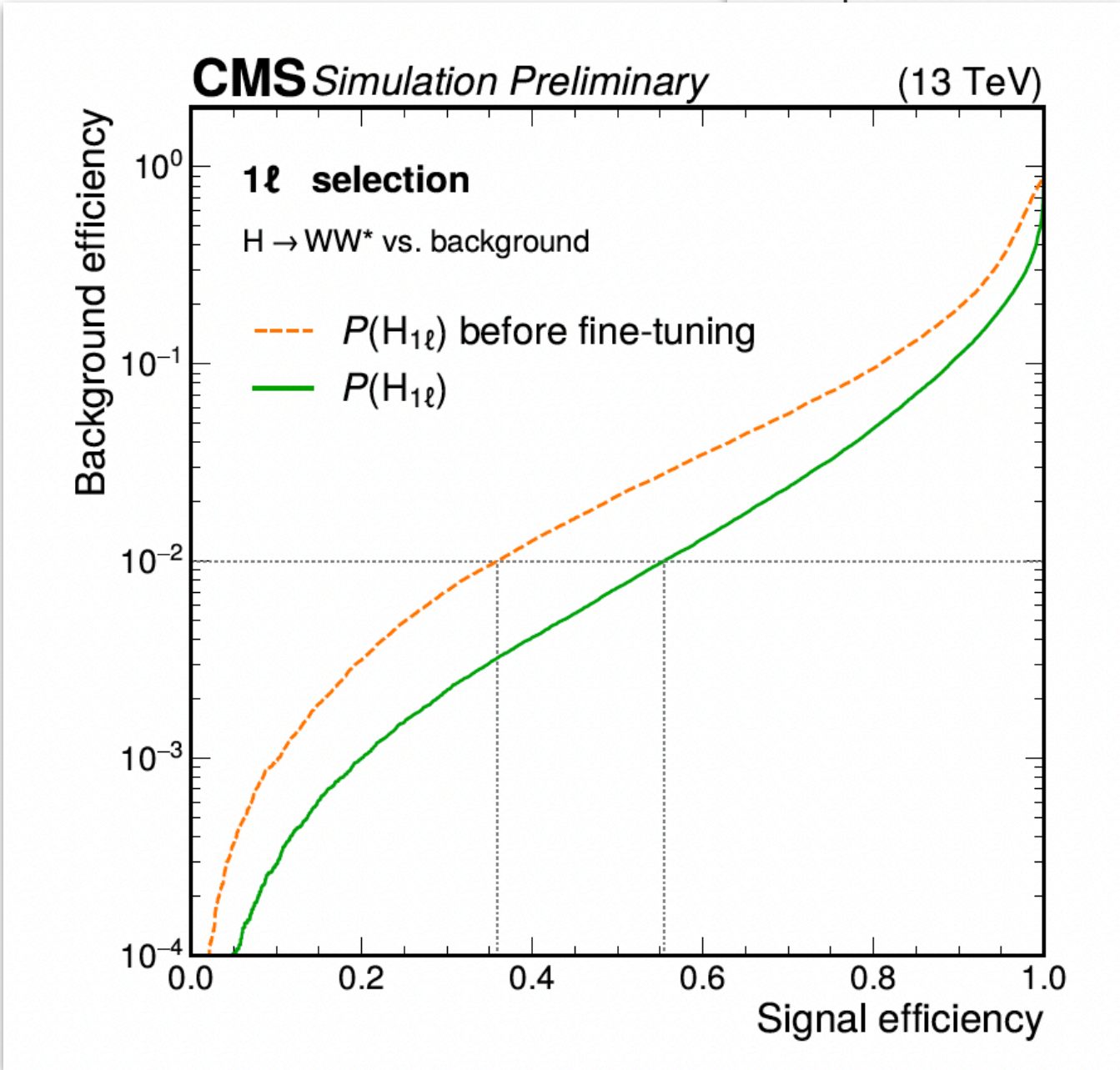
What's next?

- Analyze current LHC data utilizing cutting-edge ML to push boundaries of sensitivity
 - $H \rightarrow WW..., t\bar{t}t\bar{t}, (HH, t\bar{t}H?)$
- Build upon analysis+ML+trigger experience to inform High-Luminosity LHC strategy
 - More robust trigger algorithms
 - Bigger, better (foundation?) models for jet tagging



Final Thoughts

- Precise experimental measurements of the **top quark** and **Higgs boson** are vital to **understanding the Standard Model** + revealing unexpected physics beyond it
- **Cutting-edge ML algorithms** at all levels of the experimental pipeline are **driving sensitivity and discovery**:
 - Improving trigger algorithms + jet reconstruction
 - Facilitating complex measurements
 - Addressing challenges of HL-LHC & beyond
 - *Relevant to other scientific challenges!*



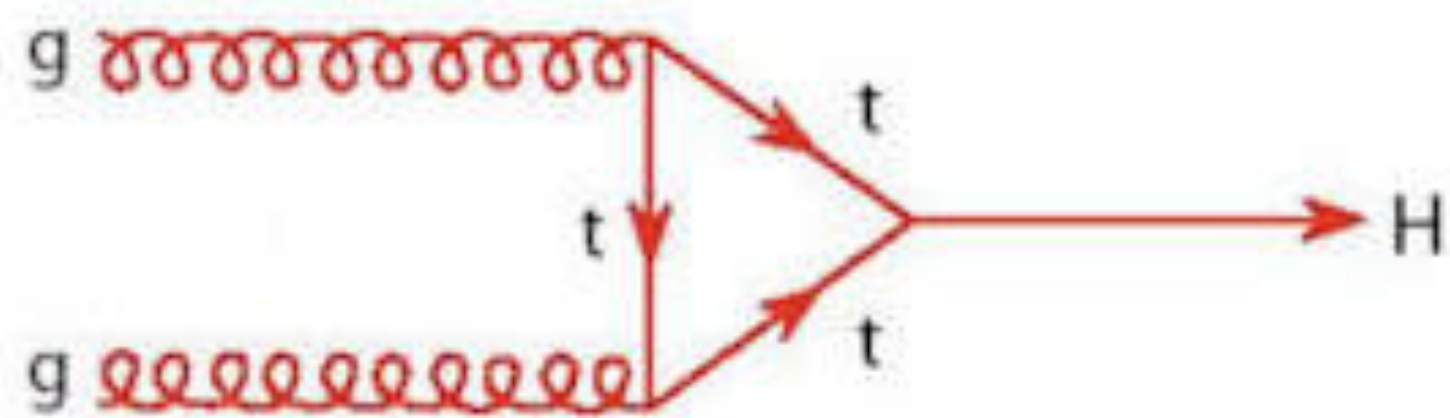

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ + i\bar{\psi}D\not{\psi} - m\bar{\psi}\psi \\ + |D_{\mu}\phi|^2 - V(\phi)$$

Thank you!

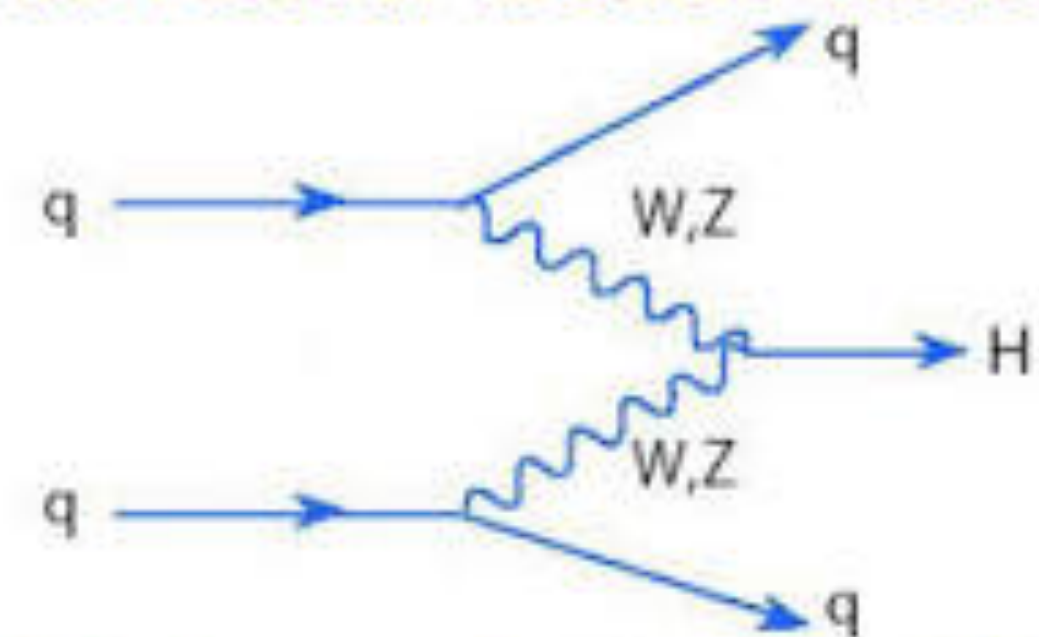
Backup

Higgs Decay Modes

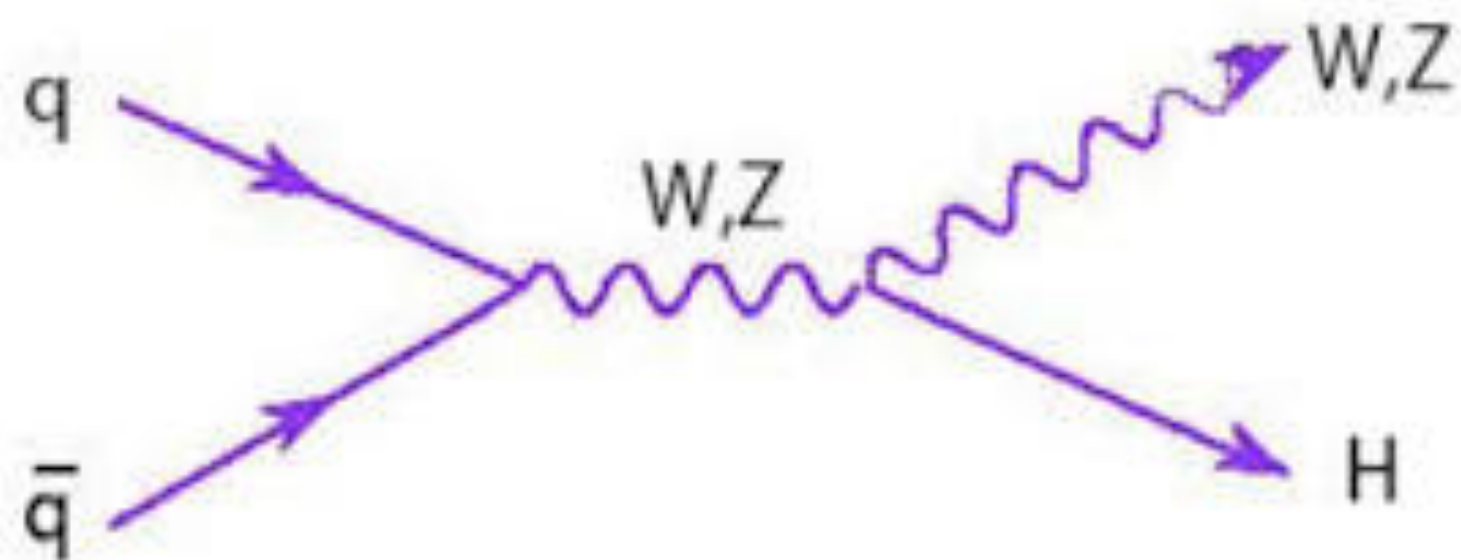
~88% Gluon fusion (ggF)



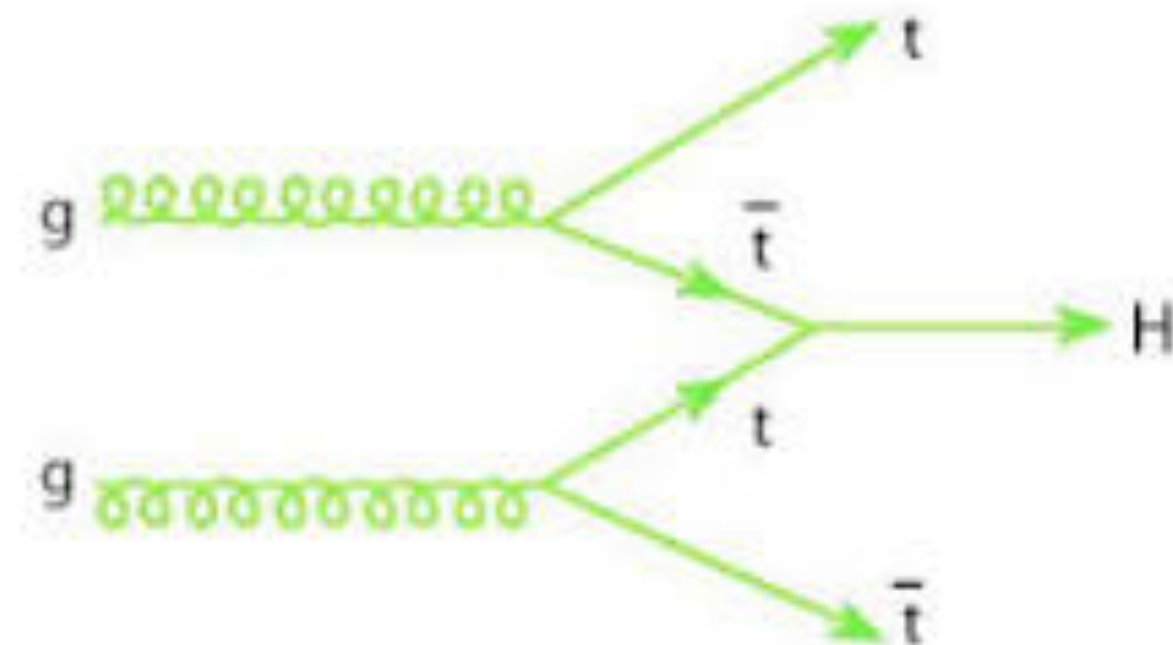
~7% Vector-boson-fusion (VBF)



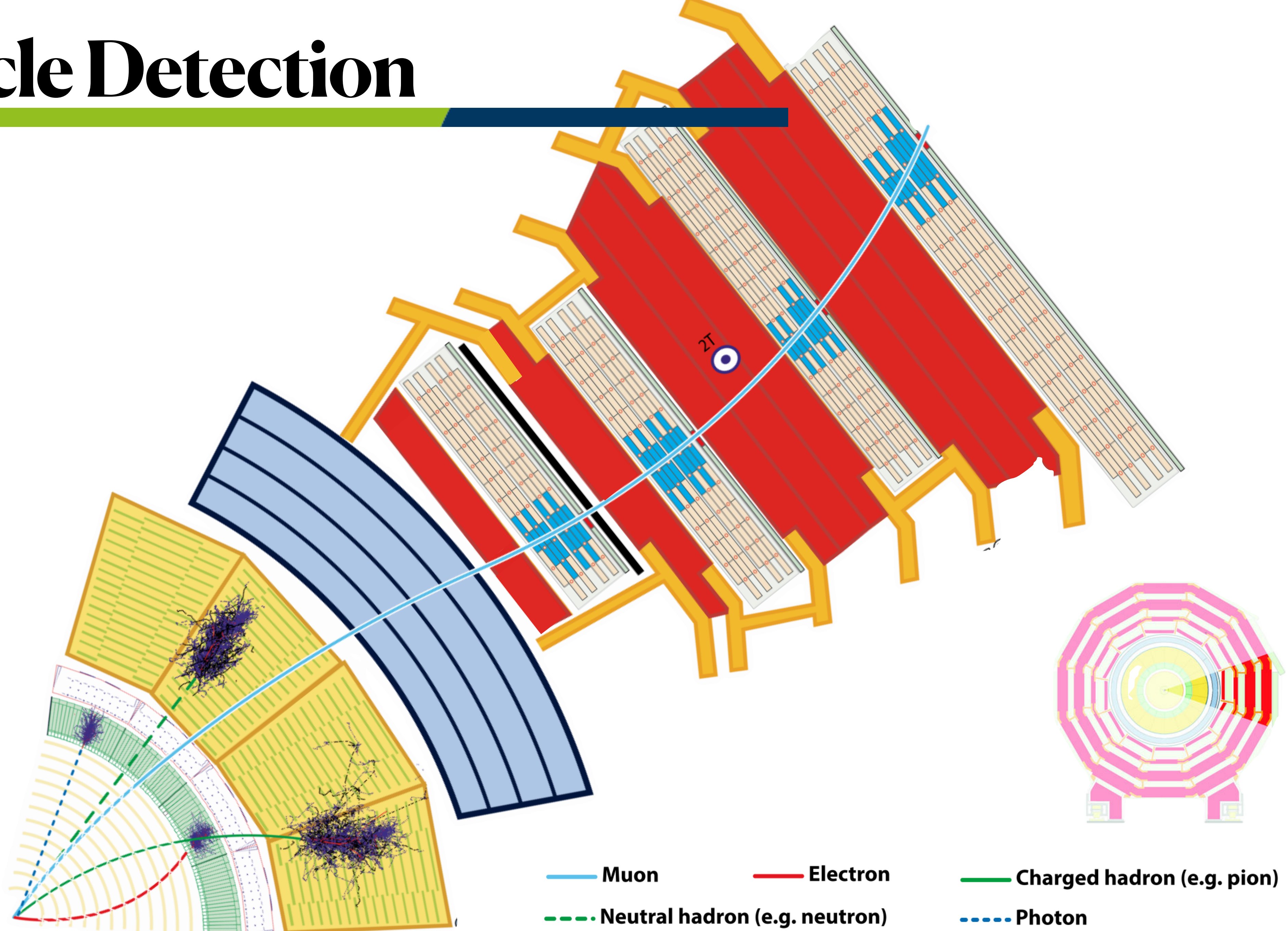
~4% Higgs-Strahlung (VH)



~1% Association with $t\bar{t}$ (ttH)

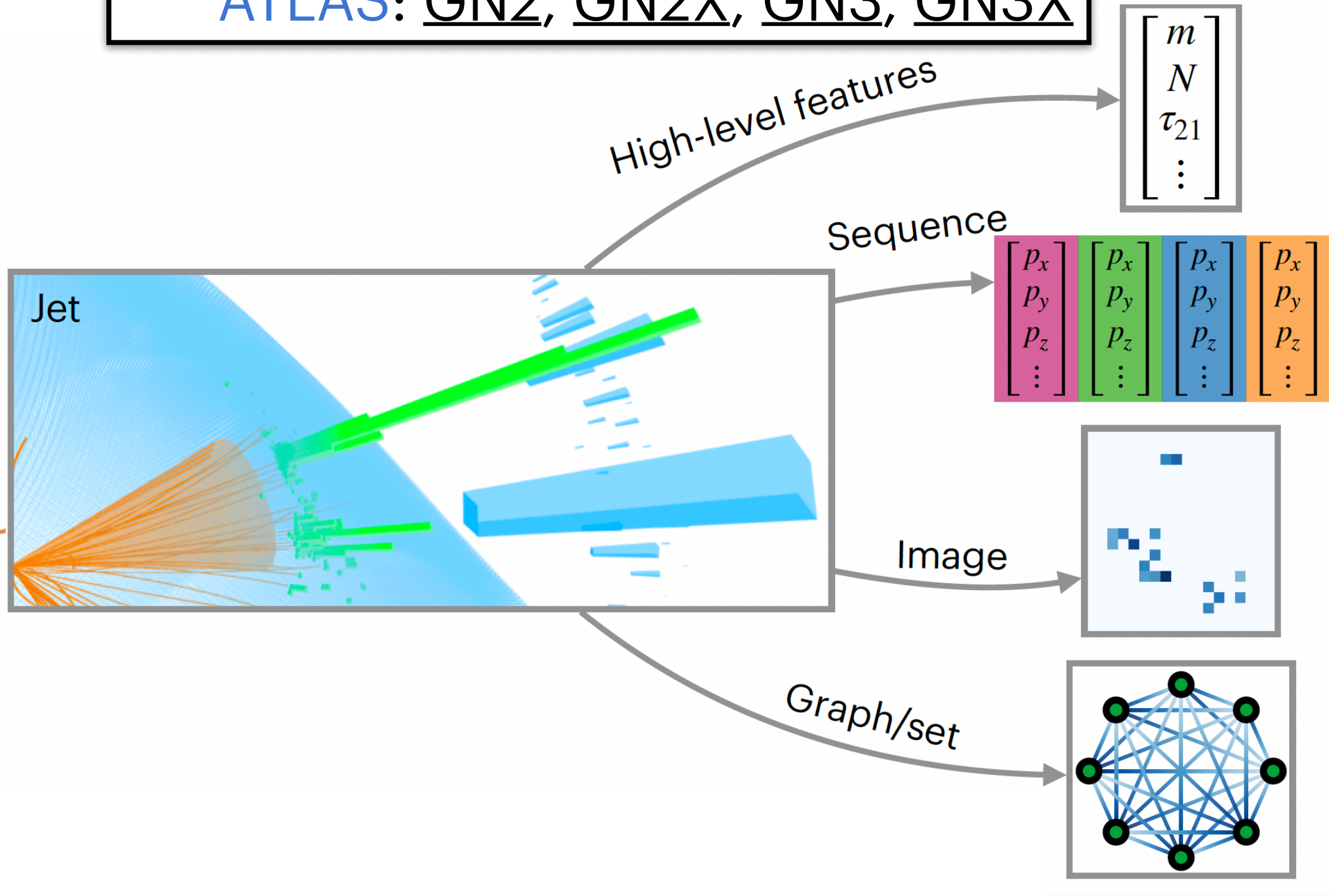
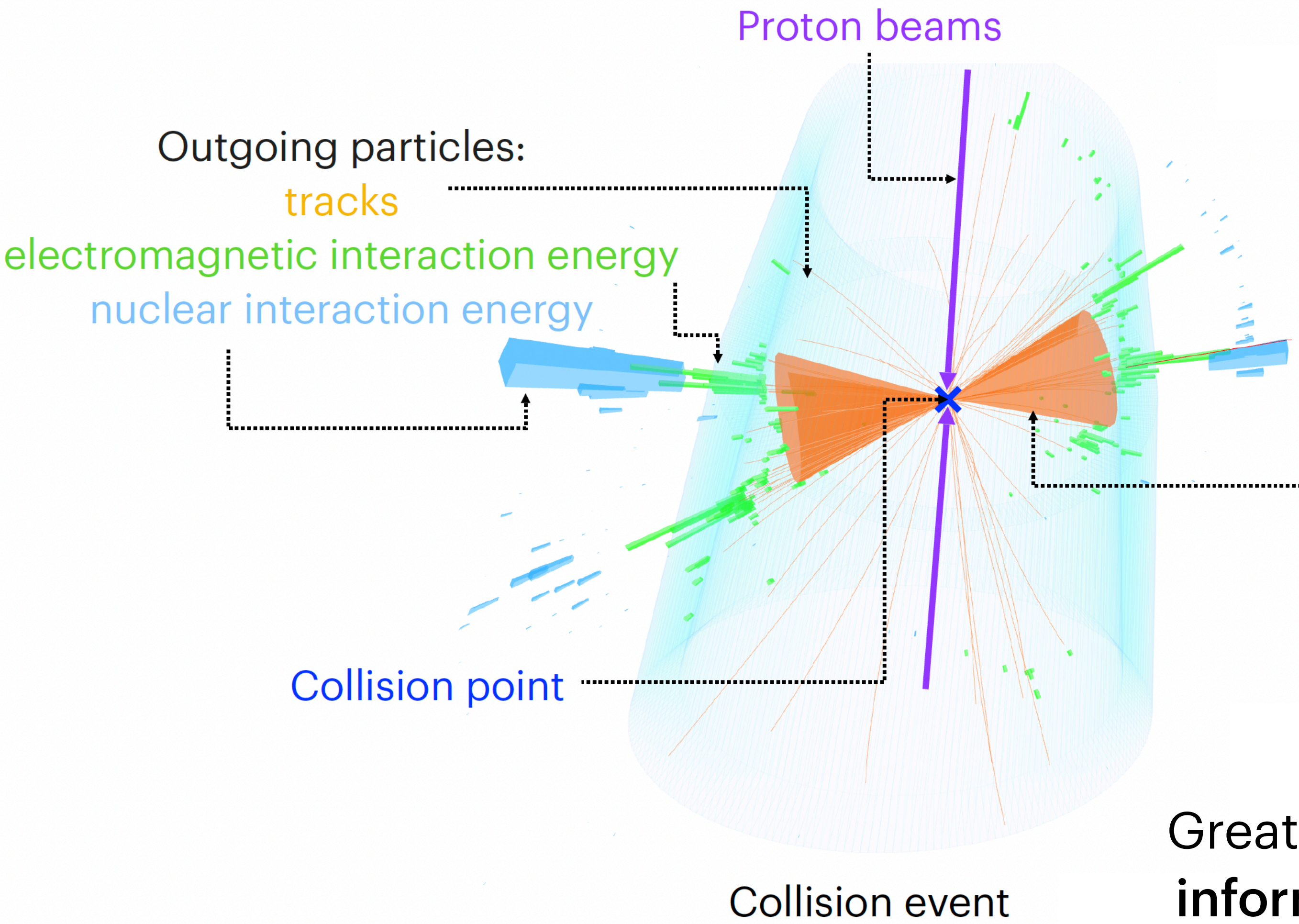


Particle Detection



Jet Tagging

- GNNs:
 - CMS: ParticleNet, ATLAS: GN1
- Transformers:
 - CMS: ParT, UParT GloParT,
ATLAS: GN2, GN2X, GN3, GN3X

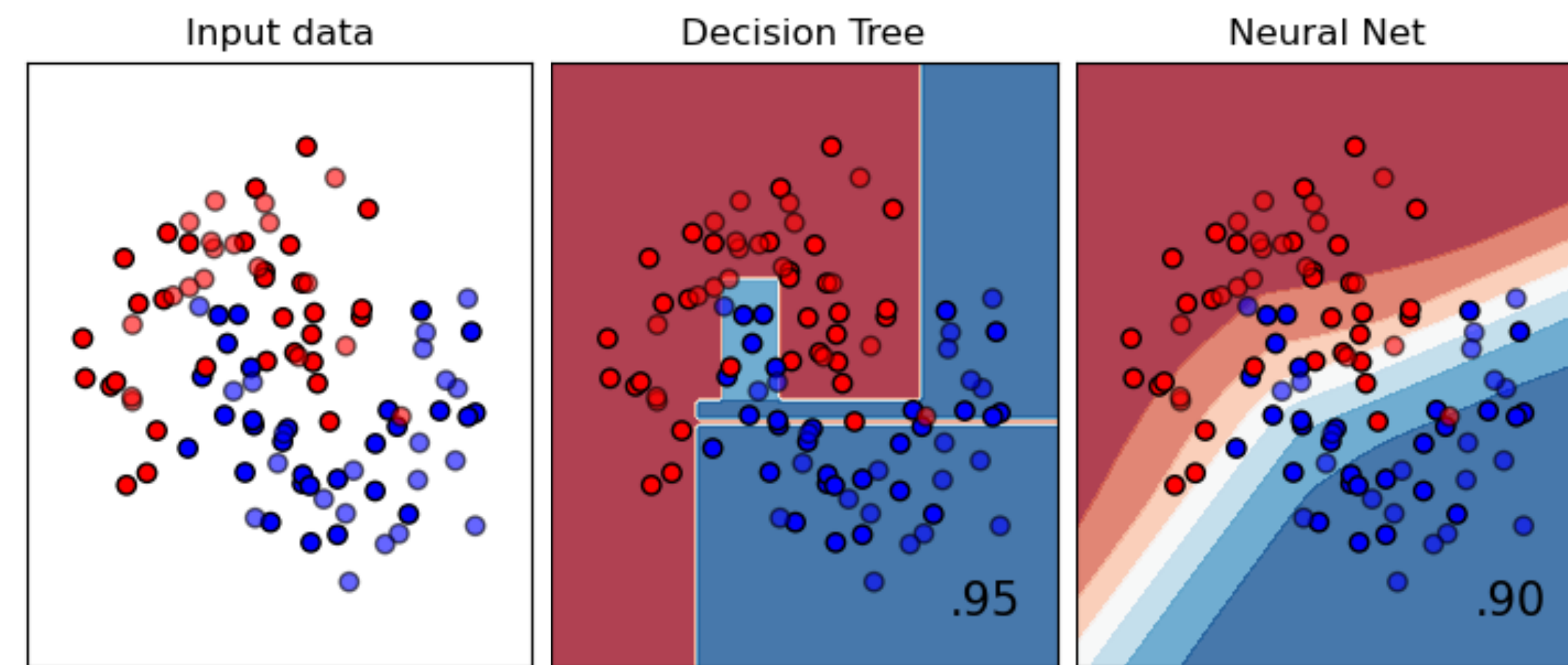
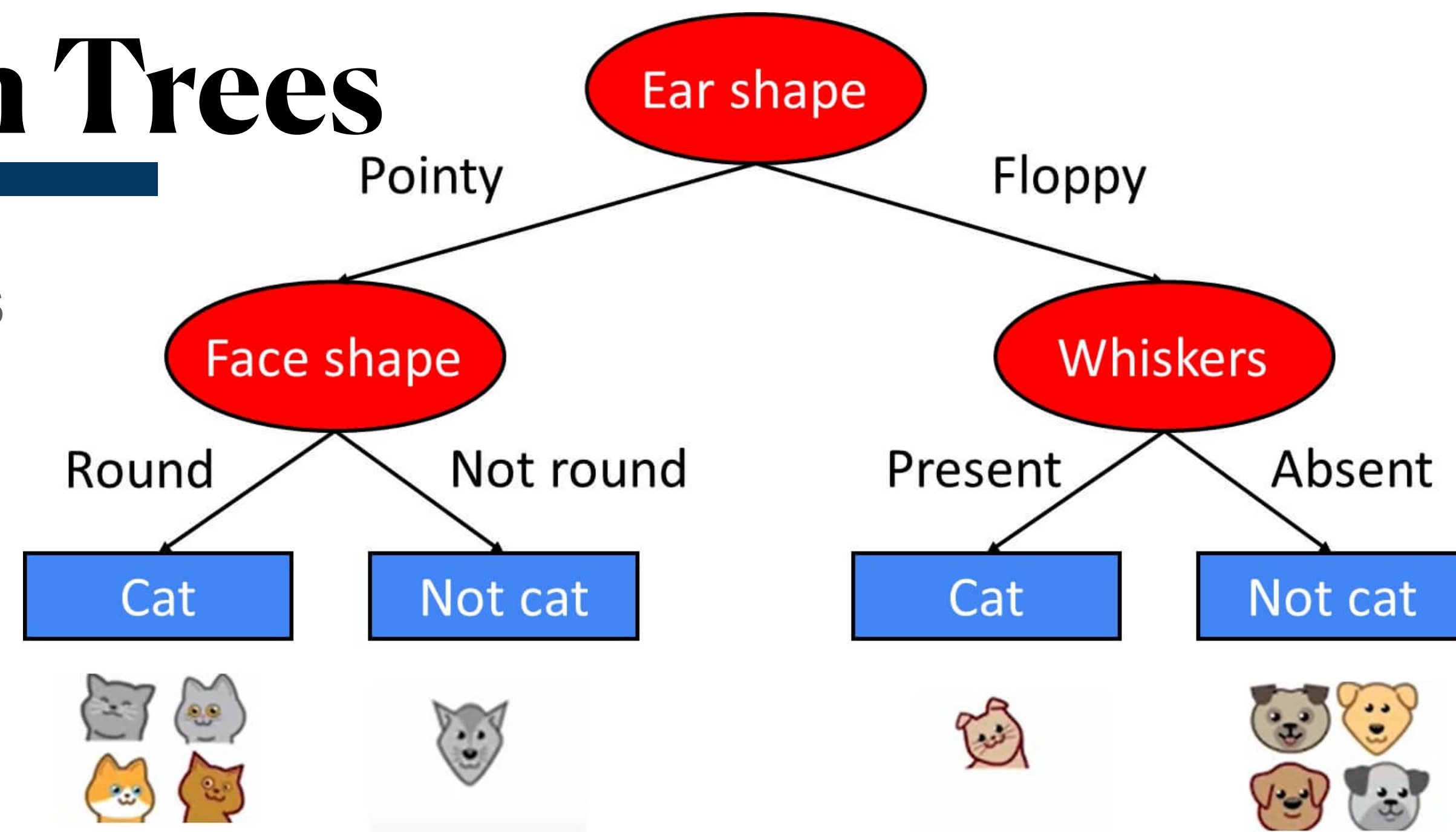


Great strides made to leverage rich low-level information with graph neural networks and transformers for jet classification!

Sidenote: Boosted Decision Trees

- ~Flowchart of binary cuts on multiple variables
- Boosting: combine many trees:
 - When a tree misclassifies events, the next tree is trained with those misclassified events given higher weight
 - The final score is a weighted vote across all trees.
- Interpretable, easy to train
- Excellent results for appropriate problems!

→ ***Pick the model that suits your task***



[BDT tutorial](#)

scikit-learn.org

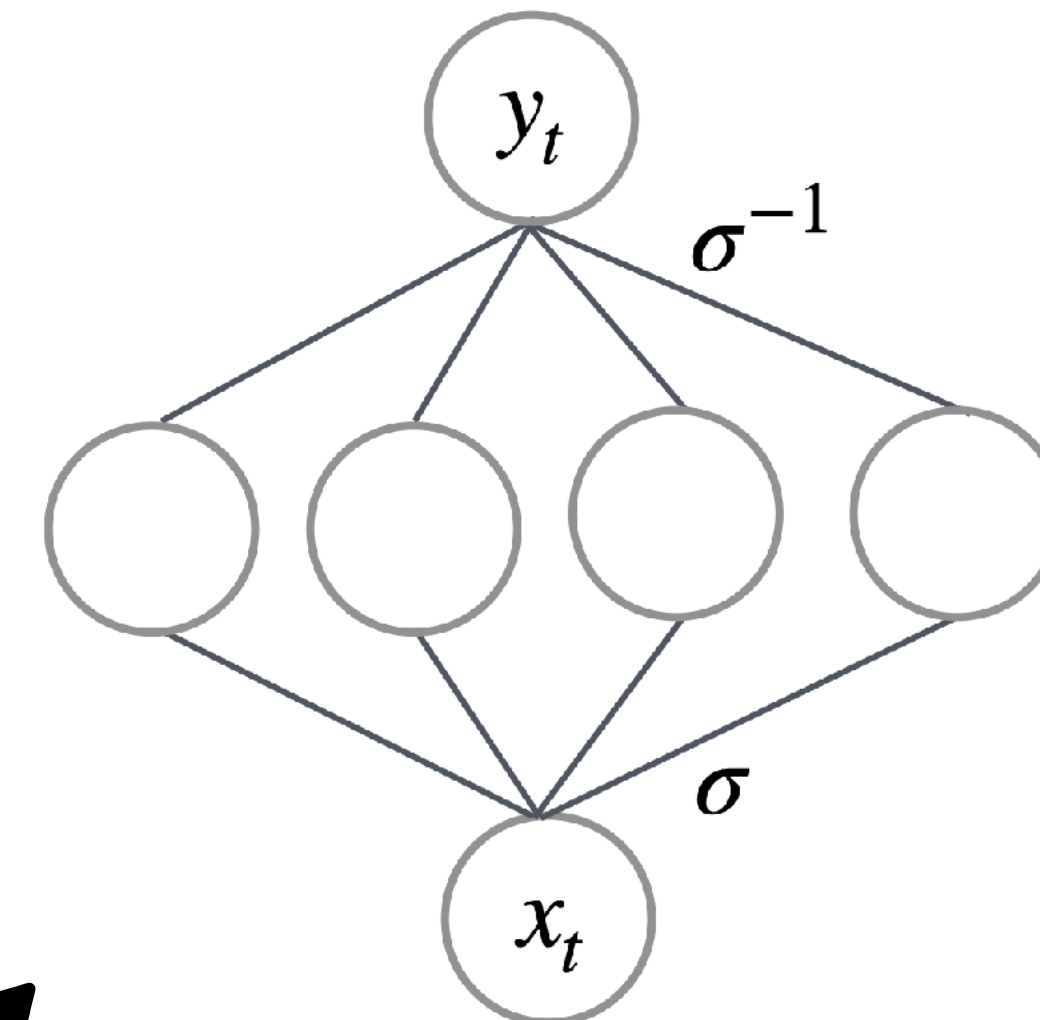
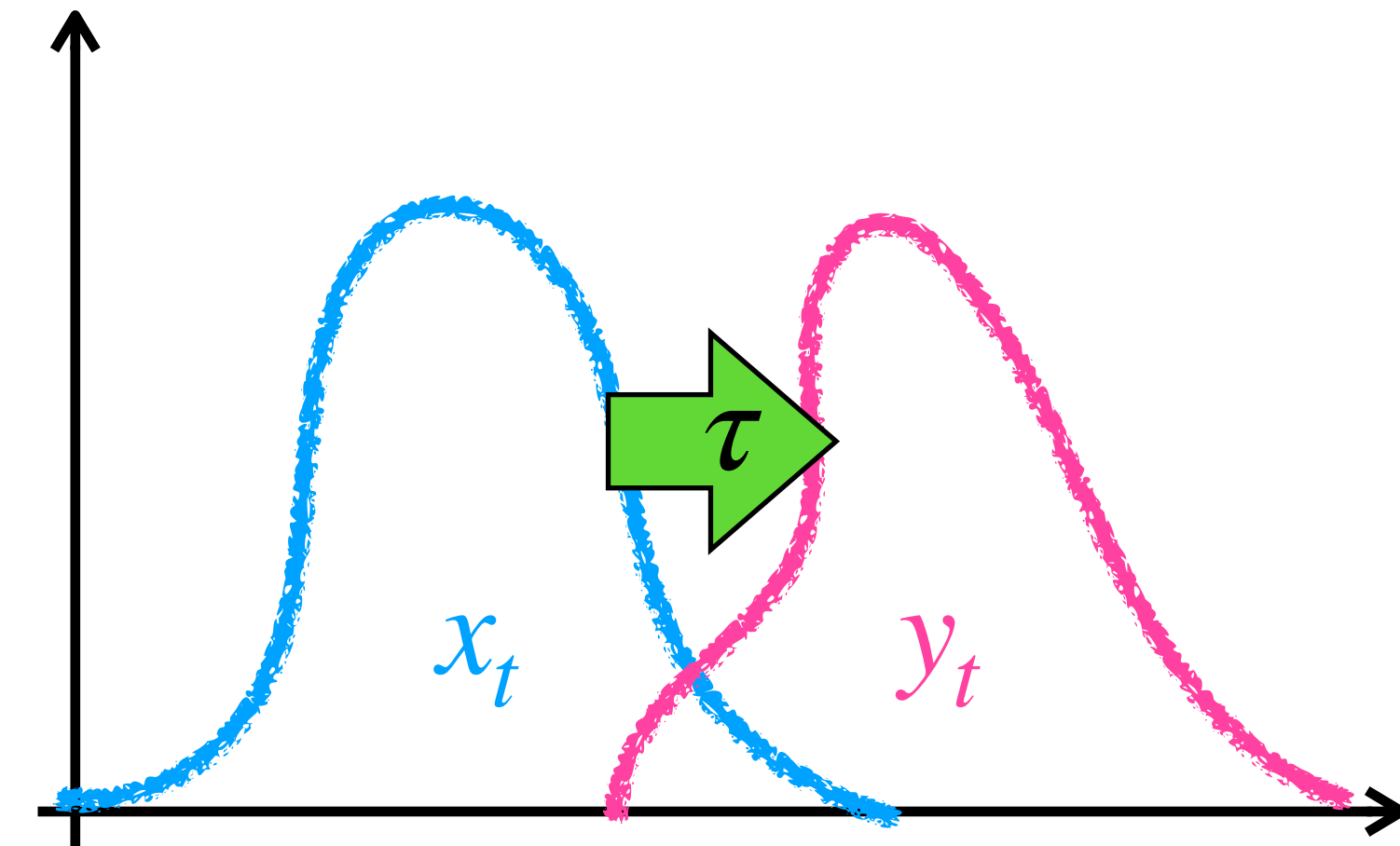
Shape Estimation: NAFs

- Neural autoregressive flows (NAFs) are a form of normalizing flow (NF)
- Latent variables y_t from a target distribution $p_Y(y_t)$ **expressed as an invertible transformation τ** of observable (“feature”) variables x_t from an input distribution $p_X(x_t)$
- τ is a transformer that depends on observables x_t via a **conditioner function c** that modifies transformer based on current and previous components (“conditional variables”) $c(x_{1:t-1})$:

$$y_t = \tau(c(x_{1:t-1}), x_t) = \tau_c(x_t)$$

- Recommended transformer (τ): **deep sigmoidal flow** (DSF) unit where w , a , and b are parameters of a model adjusted during training:

$$y_t = \tau(x_t) = \sigma^{-1}(w^T \cdot \sigma(a \cdot x_t + b))$$



$$y_t = \tau_{DSF}(x_t)$$

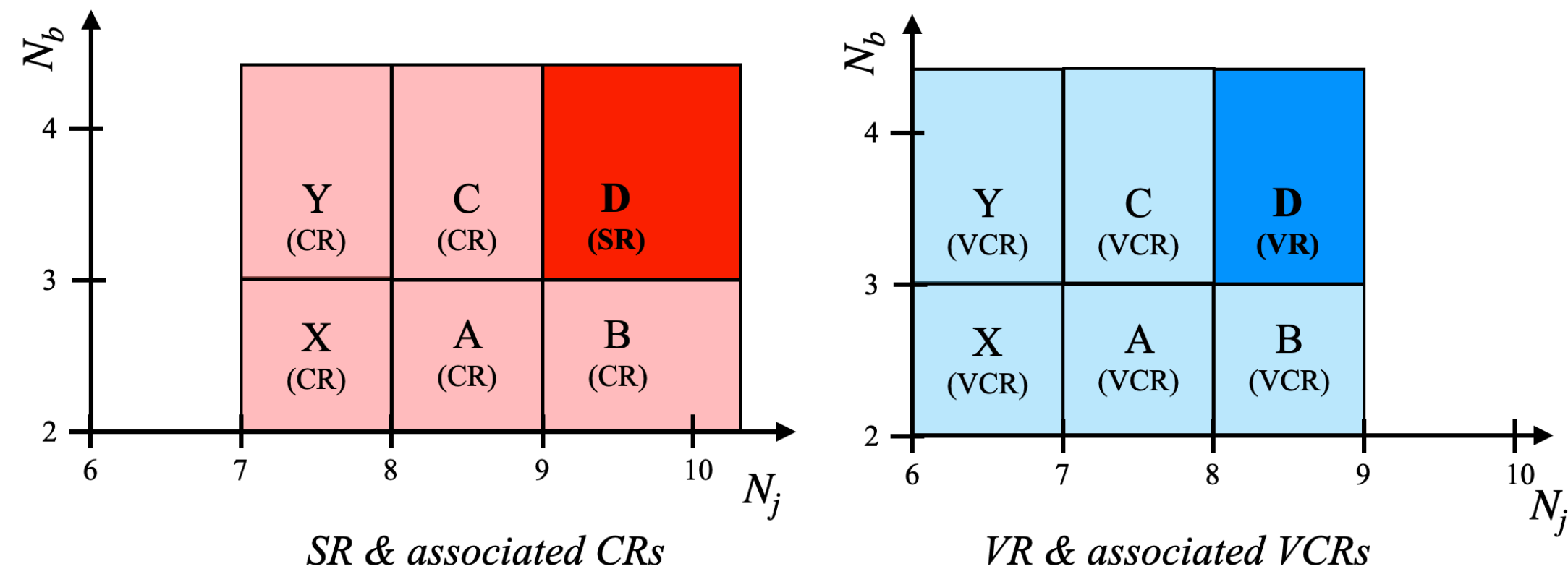
DSF unit with sigmoid activation, 1 hidden layer and an inverse sigmoid at the output layer. Sigmoid activation function σ is defined as $\sigma_z = 1/(1 + e^{-z})$

Shape Estimation: ABCDNN

- We define an “ABCDNN” NAF:

$$y_t = \tau(c(x_{1:t-1}), x_t) = \tau_c(x_t)$$

- Target distribution:** the data-driven backgrounds to be predicted
- In this case major QCD multijet and $t\bar{t}$ backgrounds (data-minor MC backgrounds)
- Input distribution:** MC simulation sample with sufficient statistics
 - In this case $t\bar{t}$ is used



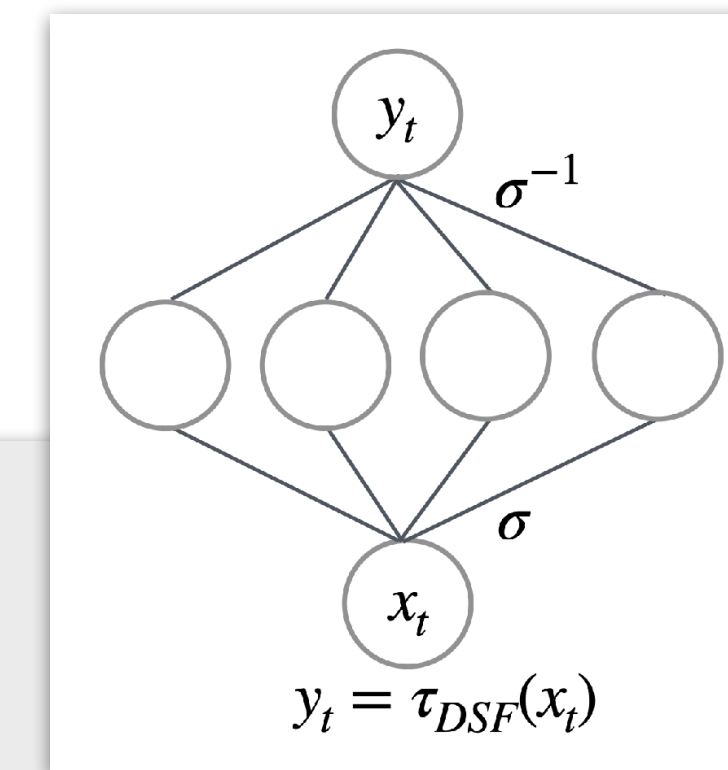
- Feature variables** that the network transforms: BDT discriminant distributions and H_T (in order to bin in H_T categories)
- Conditional variables** are one-hot encoded labels reflecting different CR/VCR regions
- Loss function:** maximum mean discrepancy (MMD)
- Network learns to transform feature variables to match target distribution by generalizing over control regions → **predicts BDT discriminant distribution of QCD multijet and $t\bar{t}$ backgrounds in SR (VR) from data in CRs (VRs)**

Training Procedure

- **Training Loop (per mini-batch):**

Forward Propagation (repeat for a given network depth)

$$y_t = \sigma^{-1}(\mathbf{w}^T \cdot \sigma(\mathbf{a} \cdot x_t + \mathbf{b}))$$



1. **Input permutation**
shuffle features and concat conditionals to include outputs of previous mini-batches
2. **CCN* 1**
Compute weights $\mathbf{a}$ and $\mathbf{b}$ from conditionals
3. **Sigmoid**
Sigmoid σ transformation of DSF unit
4. **CCN* 2**
Compute weight $\mathbf{w}$ from conditionals
5. **Inverse Sigmoid**
Inverse sigmoid σ^{-1} transformation of DSF unit

Learning Rate Adjustment and Early Stopping
Sawtooth learning rate scheduler, early stopping if loss does not improve over number of training steps

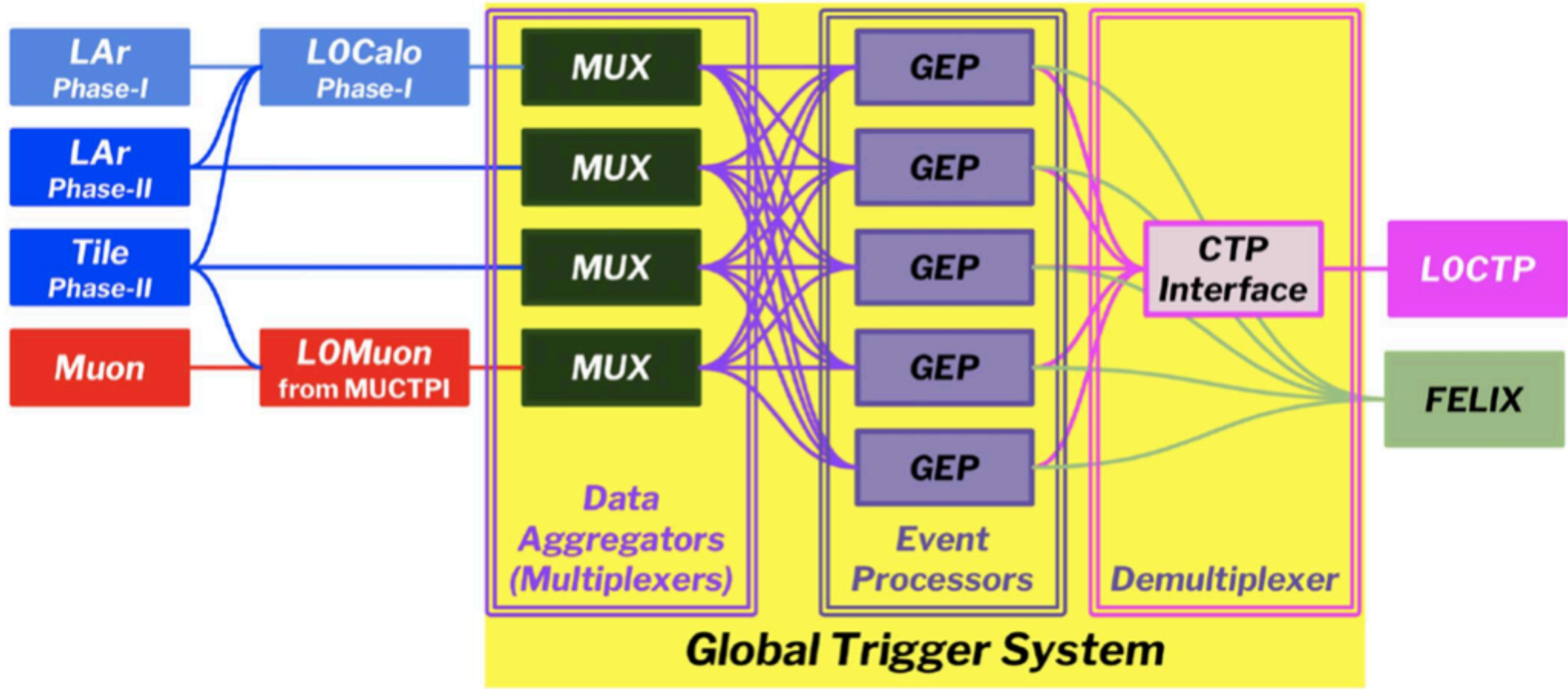
Backpropagation
ADAM optimizer used to update model weights

Compute Loss
maximum mean discrepancy (MMD) loss

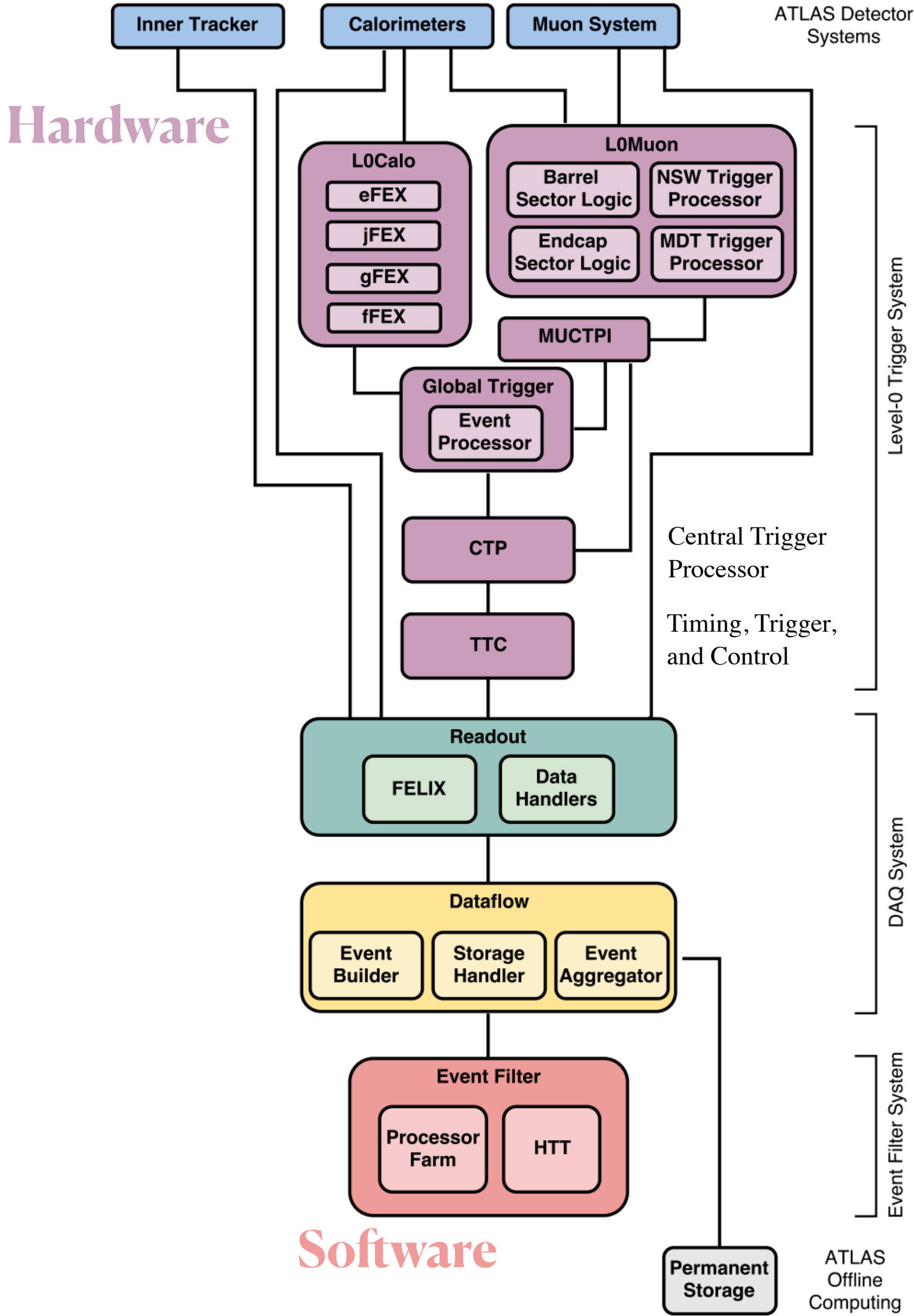
*CCN: Conditional coupling network (CCN) with 2 fully connected layers and swish activation function $\text{Swish}(z) = z\sigma(z)$

- LO: $\sim 10 \mu\text{s}$ latency, advanced FPGAs+optical modules
- Global Trigger: full-granularity calorimeter processing and topological algorithms in hardware
 - Time-multiplexing architecture makes full-event, offline-quality reconstruction feasible at LO
 - Full event processing on GEP FPGAs
- hls4ml/ML based algorithms proposed for GEP FPGAs [2406.12875]
 - B-tagging, VBF classification, MET regression, quark-gluon jet tagging...
 - Improved performance within $1.2 \mu\text{s}$ latency, $<5\%$ resource utilization

Phase-2 ATLAS Global Trigger



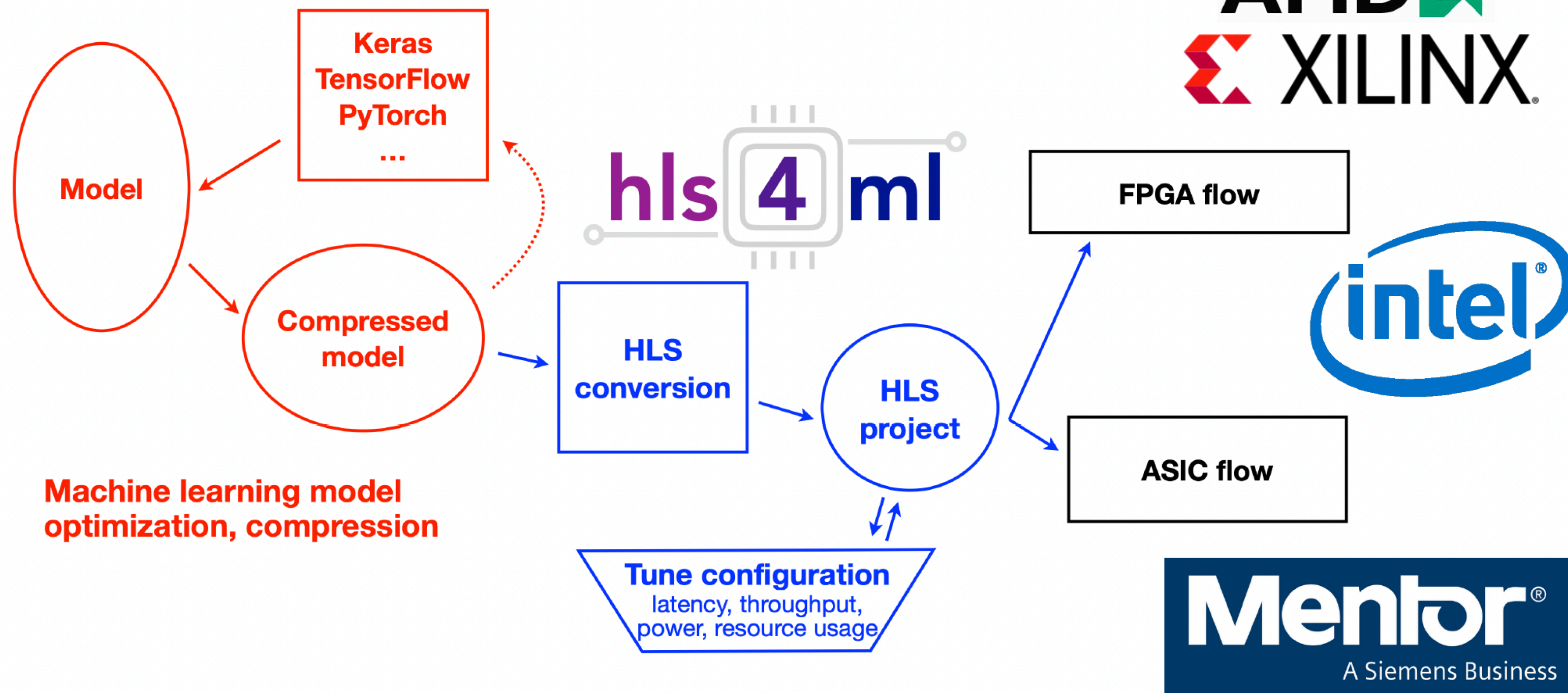
Phase-2 ATLAS TDAQ system



How to put ML on FPGAs?

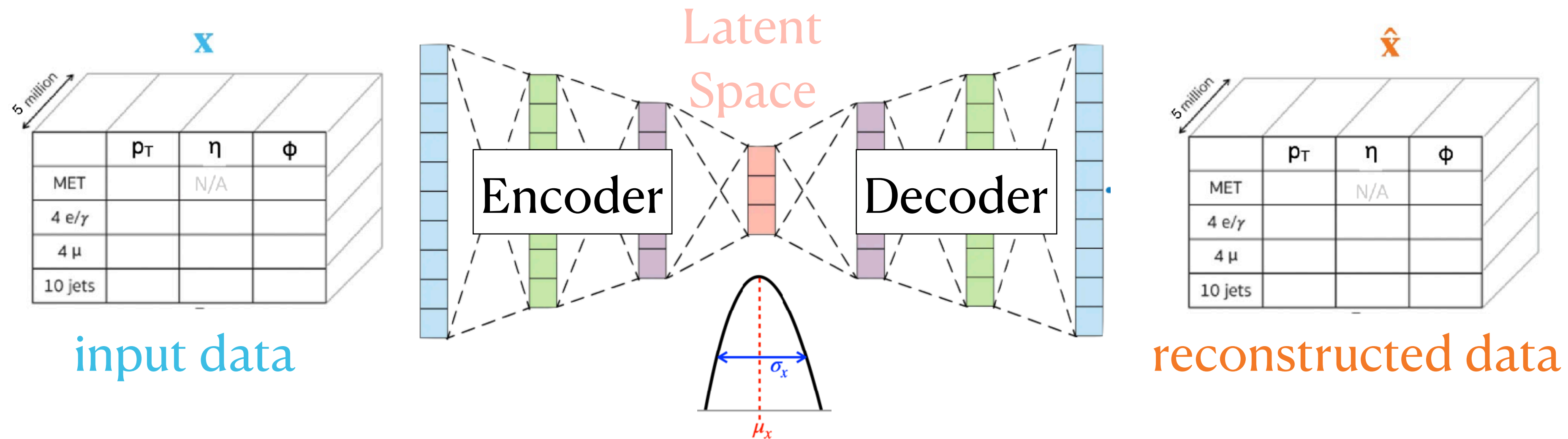
ML for data collection

- Codesign: intrinsic development loop between ML design, training, and implementation
- [hls4ml](#) - a source-to-source compiler for translating ML algorithms written in python into FPGA firmware



AXOL1TL Design: 2024

- **Inputs L1 trigger objects:** (p_T , η , ϕ) of MET, 4 electron/photons, 4 muons, 10 jets
- **Train on ZeroBias data** collected by CMS in 2023/2024
- **Variational autoencoder (VAE):** Additional KL-divergence loss term regularizes latent space to be normal gaussian ($N(\mu_x, \sigma_x) \rightarrow N(0,1)$) to prevent overfitting



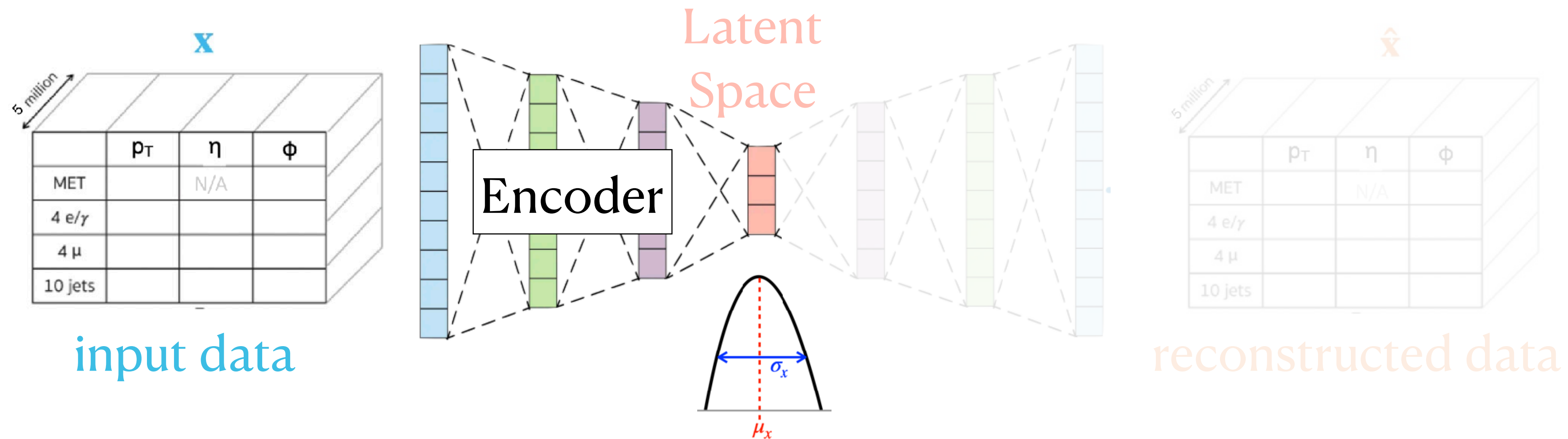
$$\text{Loss} = ||x - \hat{x}||^2 + \text{KL}[N(\mu_x, \sigma_x) \rightarrow N(0,1)]$$

Reconstruction loss

*Full regularization term
(Kullback–Leibler divergence)*

AXOL1TL Design: 2024

- **Only deploy encoder half of the network**, compute degree of abnormality from latent space directly → **Halves the network size and latency!**
- Small, fully connected network architecture
- **Satisfies strict L1 Trigger requirements**: latency of 4 μ s, low resource utilization



$$\text{Loss} = \left[\frac{1}{2} \left(\sum_{i=1}^N \mu_i^2 + \sum_{i=1}^N \sigma_i^2 \right) + \frac{1}{2} \log \left(\prod_{i=1}^N \sigma_i^2 \right) \right] + \text{KL term}$$

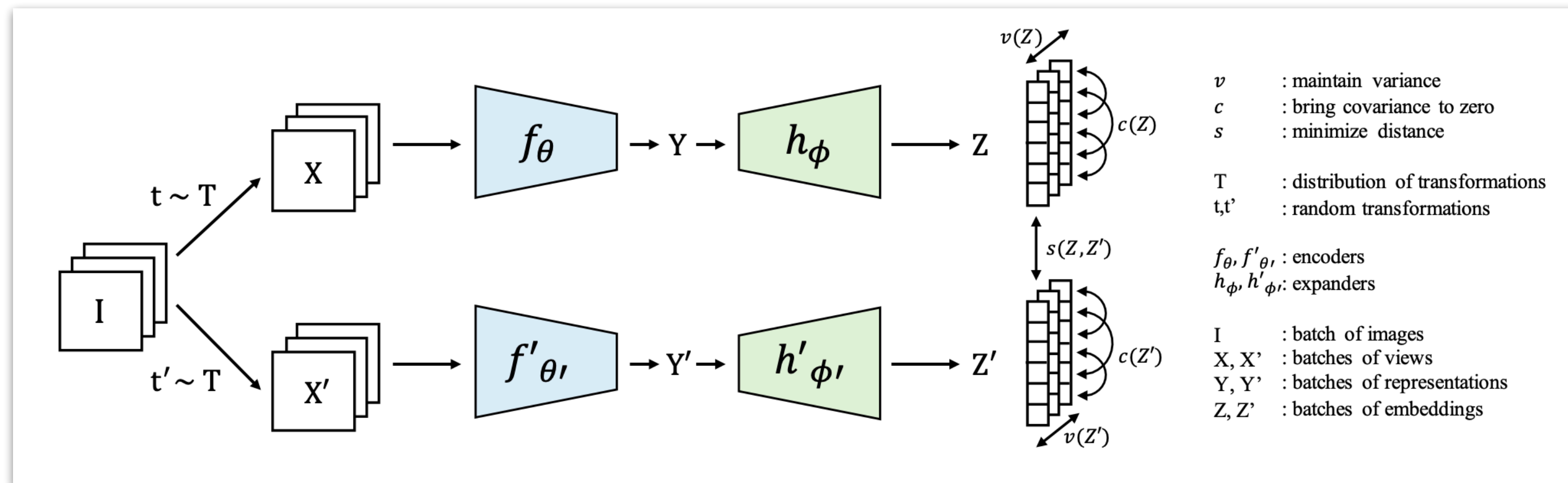
Anomaly Score $\sim \sum \mu_x^2$
(Kullback–Leibler divergence)

AXOL1TL Design: 2025

Self-supervised learning: Learn mapping function $f: X \rightarrow Y'$, from input space X (observations) to Y' where Y' are pseudo-labels generated from X

→ Encode the underlying structure of the input space through supervised-style training to learn meaningful/distinguishing features

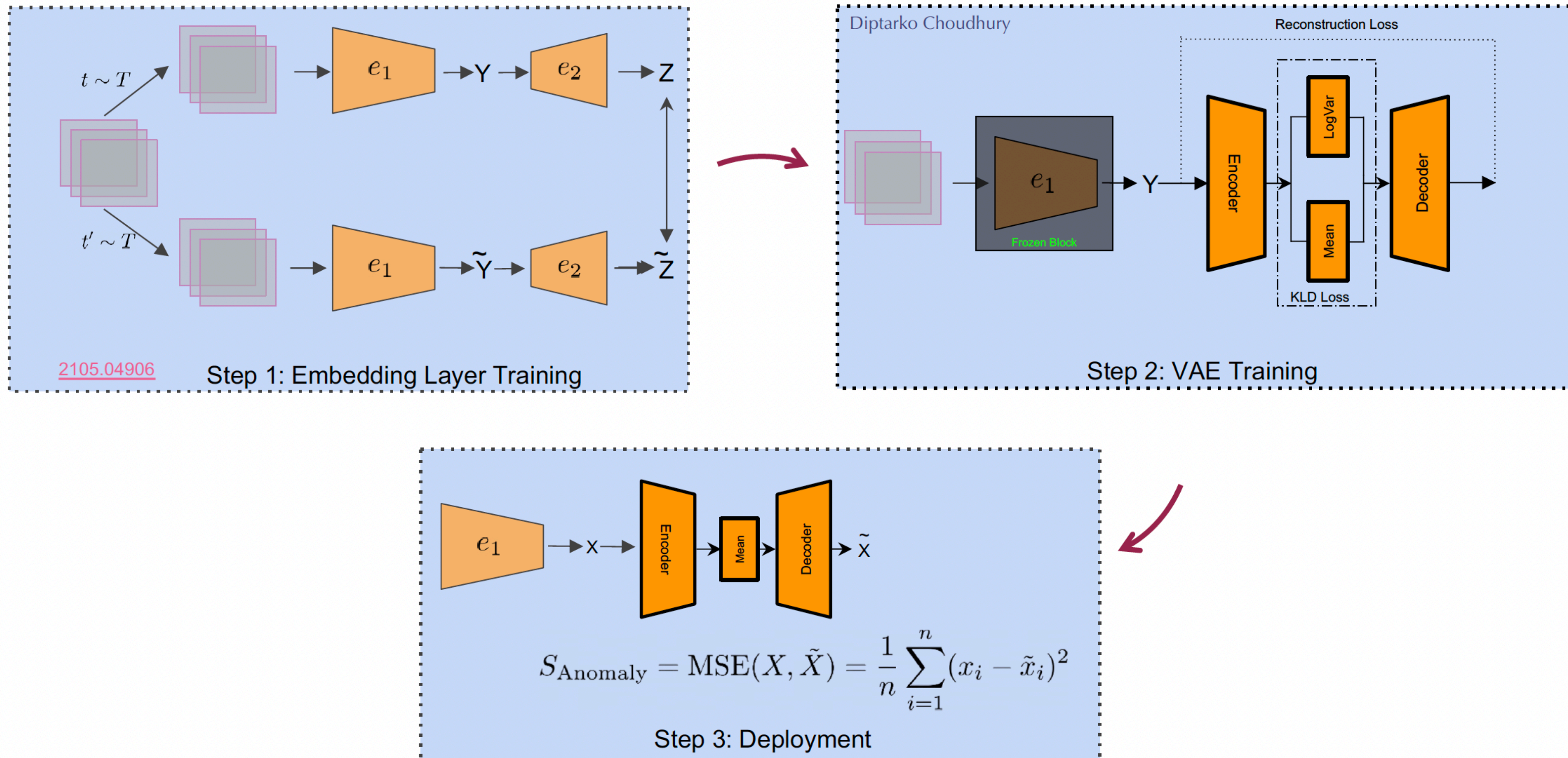
→ [VICReg](#) “Variance-Invariance-Covariance Regularization”



$$s(Z, Z') = ||Z - Z'||^2$$

AXOL1TL Design: 2025

- Train VICReg to learn embedding of data given L1 trigger kinematics
- Learned **embedding input to VAE**
- **Anomaly score is MSE loss** between embedding/VAE output

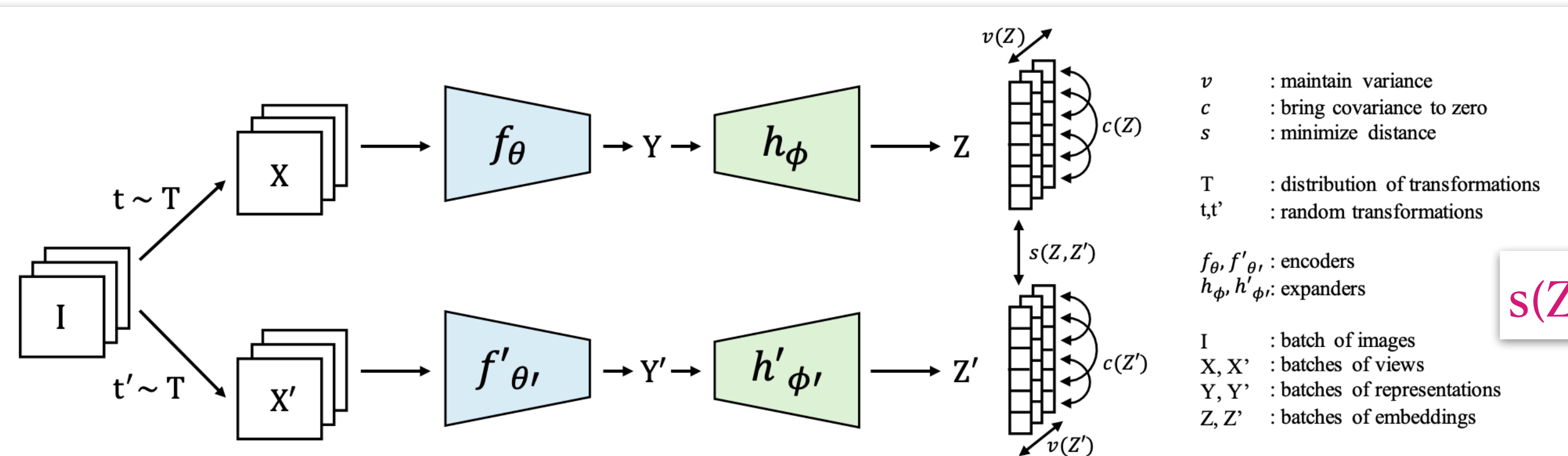


AXOL1TL Design: 2025

Self-supervised learning: Learn mapping function $f: X \rightarrow Y'$, from input space X (observations) to Y' where Y' are pseudo-labels generated from X

→ Encode the underlying structure of the input space through supervised-style training to learn meaningful/distinguishing features

→ [VICReg](#) “Variance-Invariance-Covariance Regularization”



$$s(Z, Z') = ||Z - Z'||^2$$

Figure 1: **VICReg: joint embedding architecture with variance, invariance and covariance regularization.** Given a batch of images I , two batches of different views X and X' are produced and are then encoded into representations Y and Y' . The representations are fed to an expander producing the embeddings Z and Z' . The distance between two embeddings from the same image is minimized, the variance of each embedding variable over a batch is maintained above a threshold, and the covariance between pairs of embedding variables over a batch are attracted to zero, decorrelating the variables from each other.

SPANet

- Traditional approach (try every possible jet-to-parton assignment, pick the best by χ^2) scales combinatorially and becomes intractable in high-multiplicity final states
- SPANet: “Symmetry Preserving Attention Network”
 - Based on a novel attention mechanism with embedded symmetries
 - Attention architectures are permutation invariant because rearranging the order of the elements in the input sequence induces the same rearrangement in the attention weights -> **Permutationless many-jet event reconstruction**
 - Performs effective and efficient set assignment tasks
 - Traditional permutation approach: list every possible combination of candidates into positions and scores each one at a time
 - SPANet: Simultaneously compute compatibility score between every candidate and every position, solve for best overall matching, and if two positions are identical do not bother to distinguish them (jet A->parton 1 and jet B->parton 2 is the same as jet B->parton 2 and jet A->parton 2)
 - Originally for resolved top tagging in hadronic [ttbar](#), later [expanded to variable-size set of observed aK4 jets](#)

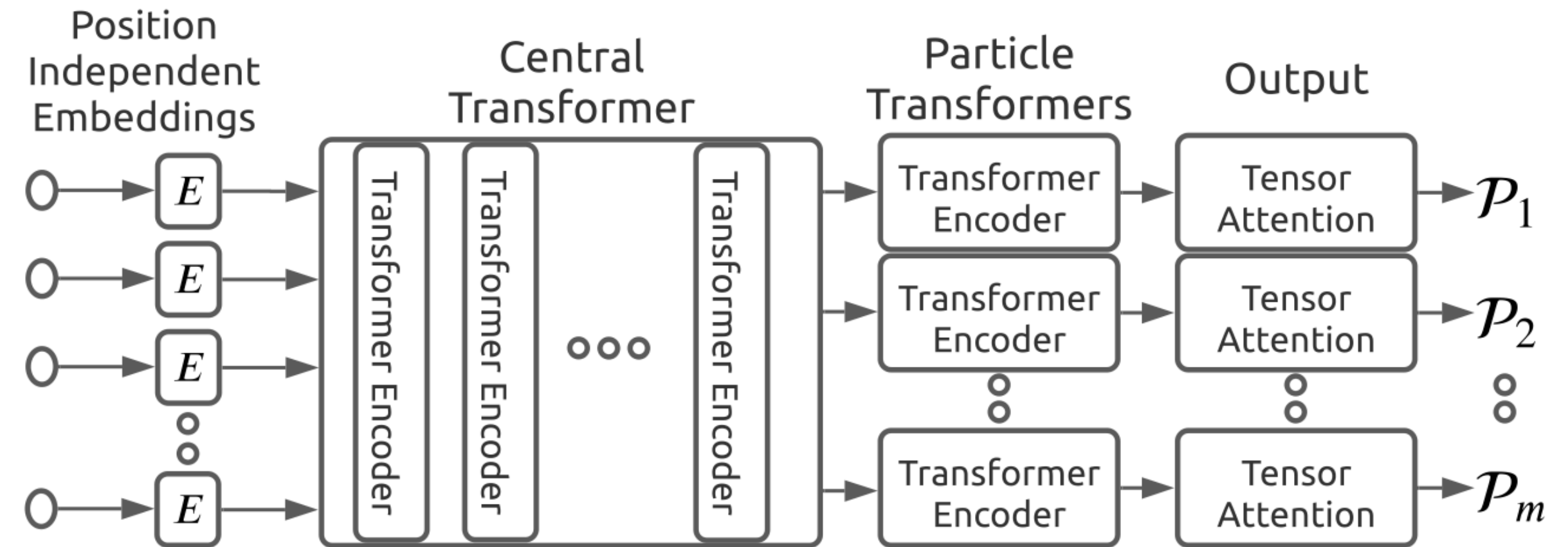
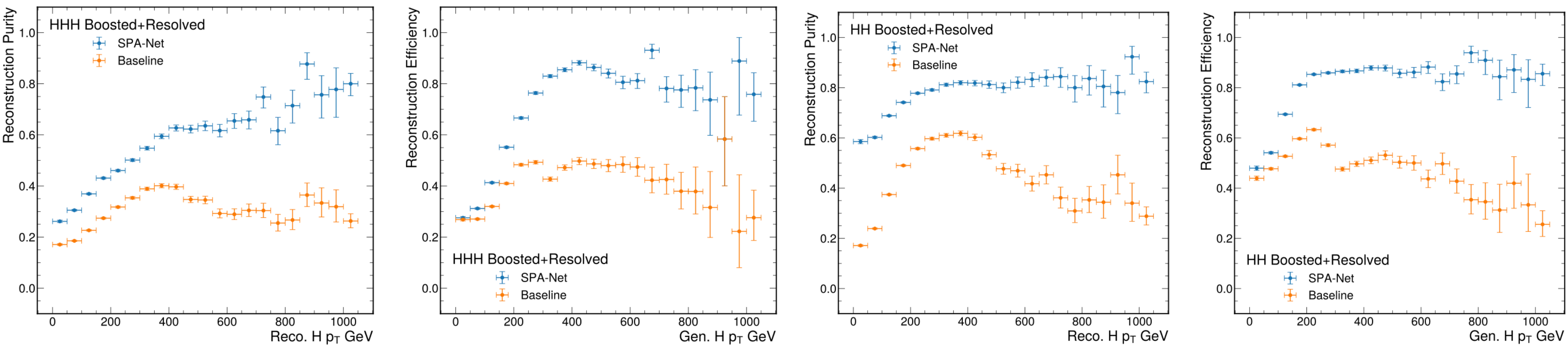
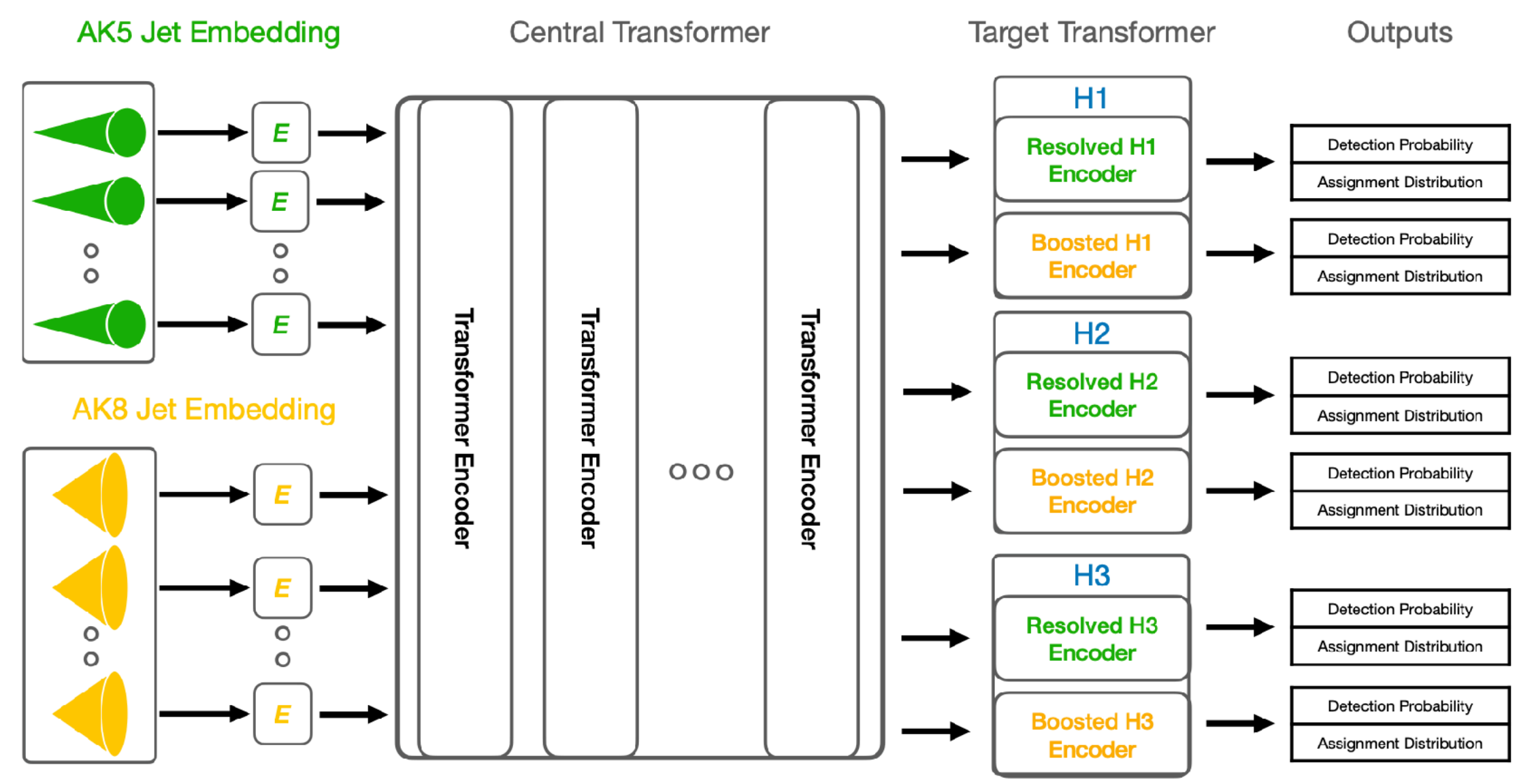


Figure 2: A visualization of the high level structure of SPA-NET.

significantly outperforms existing state-of-the-art methods, correctly assigning all jets in 80.7% of 6-jet, 66.8% of 7-jet, and 52.3% of ≥ 8 -jet events respectively

SPANet for boosted+resolved hadronic HH/HHH

- Retrain SPANet for simultaneous boosted+resolved H->bb tagging for up to 3 Higgs
 - Increases the Higgs boson reconstruction purity by 56–80% and the efficiency by 37–38% compared to baseline depending on final state
- <https://arxiv.org/abs/2412.03819>



CMS/ATLAS Tagging NNs

- **ATLAS**

- GN1: graph based resolved jet tagging
- GN2: transformer based resolved jet+flavor tagging
 - GN2: transformer based resolved jet+flavor tagging, + more classes
 - GN2X: transformer based boosted jet tagging (+flavor information)
 - GN3X: transformer based boosted jet tagging (+flavor information), + more classes

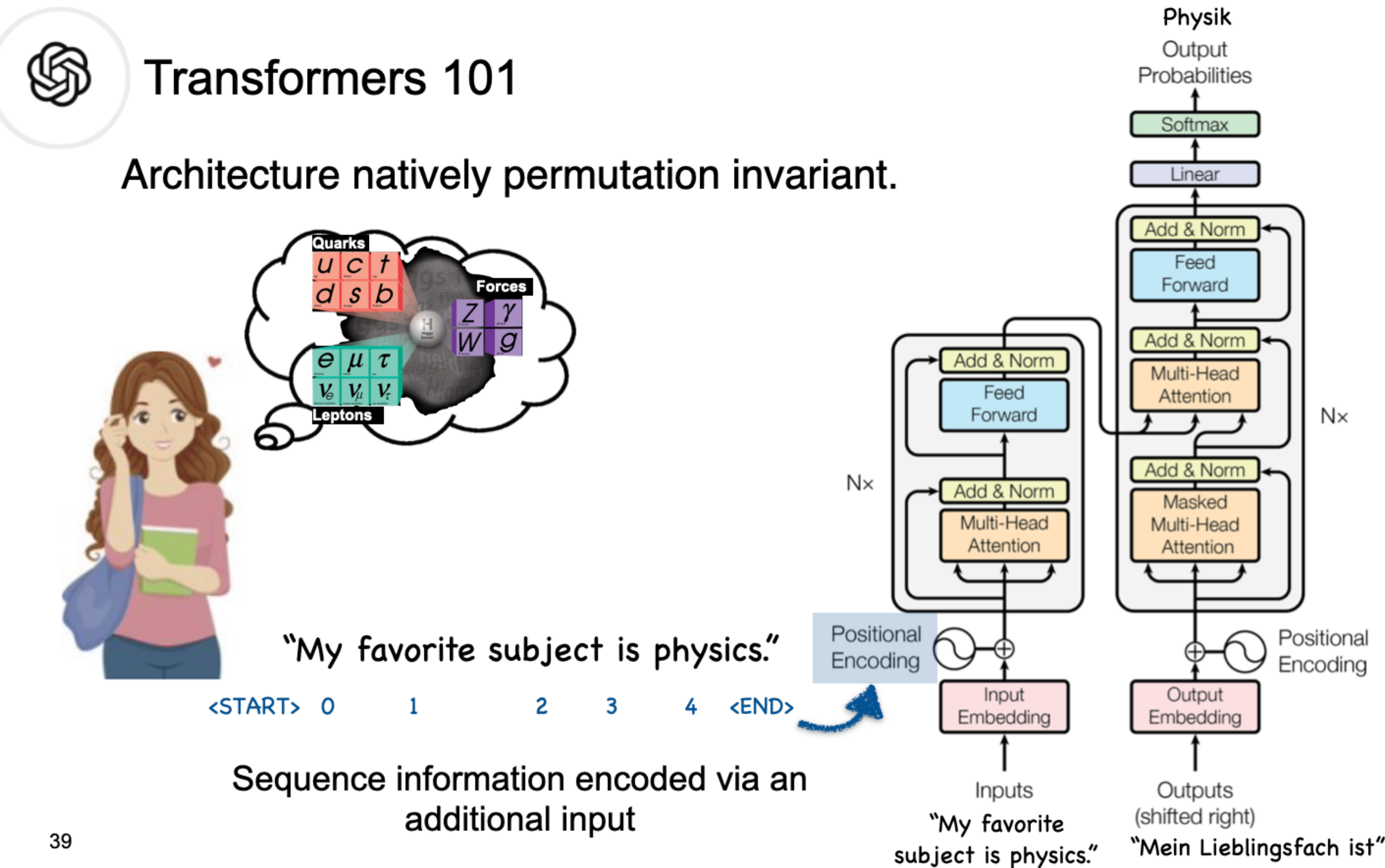
- **CMS**

- ParticleNet: graph based boosted (later resolved in ParticleNetAK4) jet tagging
- ParT: transformer based boosted jet tagging
 - GloParT: architecture evolved and scaled up to ~37 classes (+flavor information), includes mass regression
- UParT: transformer based resolved jet+flavor tagging

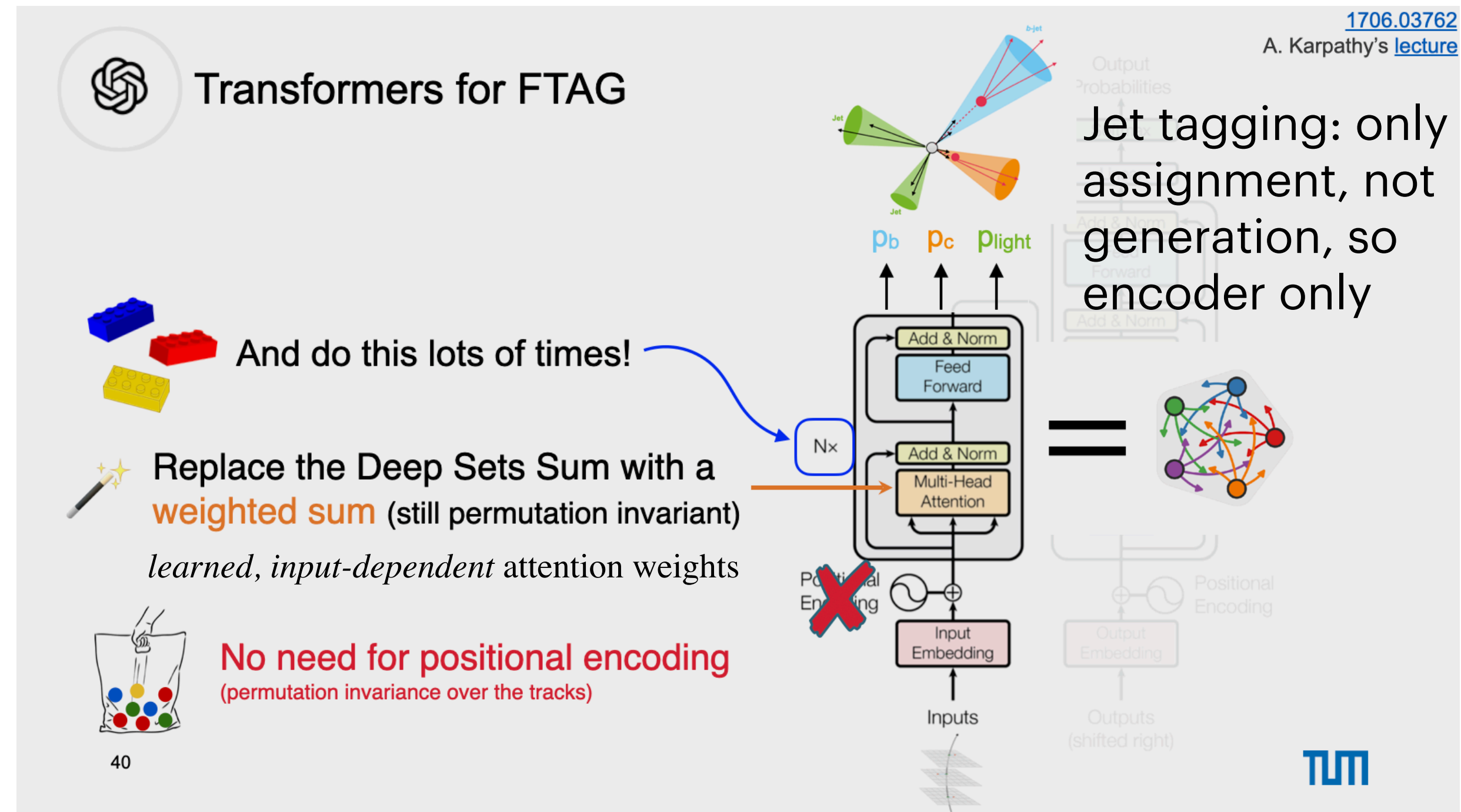
- GNNs:
 - **CMS**: ParticleNet, **ATLAS**: GN1
- Transformers:
 - **CMS**: ParT, UParT GloParT,
ATLAS: GN2, GN2X, GN3, GN3X

Transformers

Images by Nicole Hartmann



39



1706.03762
A. Karpathy's [lecture](#)

Input embedding: raw inputs (tokens, or here, tracks) mapped to continuous feature vectors

Positional encoding: added to embeddings so the model knows sequence order — *only needed when order is physically meaningful*

Multi-head self-attention: every item computes a weighted combination of every other item, with several parallel heads each learning a different notion of relevance

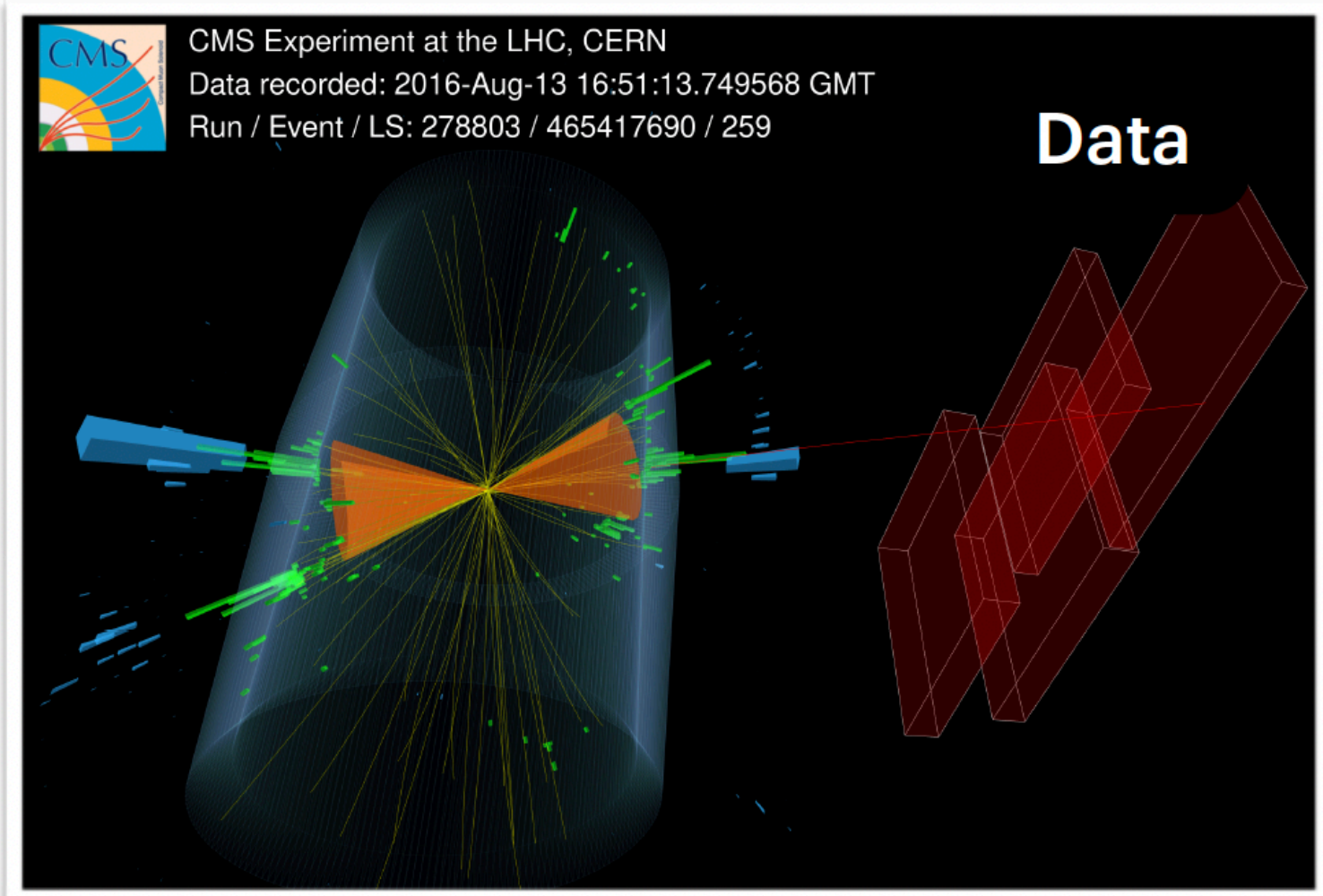
Add & Norm: residual connection + layer normalization, for stable gradient flow through deep stacks

Feed-forward: per-item MLP, applied identically to each position — where per-item transformation happens (attention handles mixing, this handles processing)

Nx: the whole block (attention → add&norm → FF → add&norm) repeated N times, building progressively richer contextualized representations

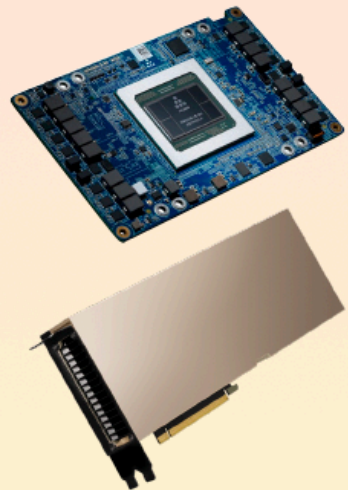
No decoder needed for classification/tagging tasks — just the encoder stack, followed by a task-specific head (here: linear + softmax → class probabilities)

A HEP Foundation Model?



- Tracks
- EM Clusters
- Had. Clusters
- Muon Tracks

Training



Adaptation

**HEP
Foundation
Model**

Tasks

Jet Tagging

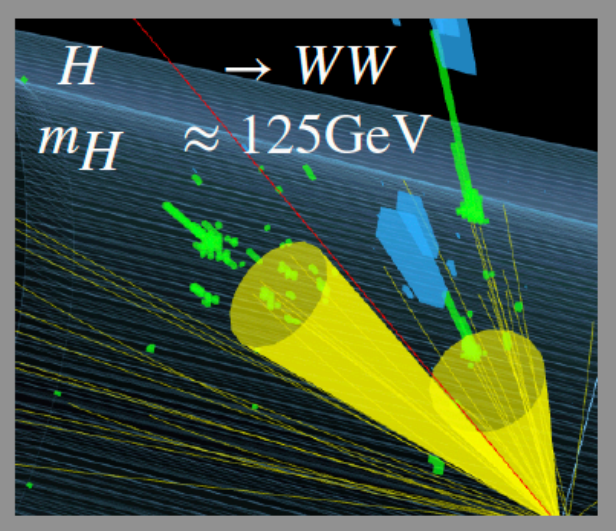
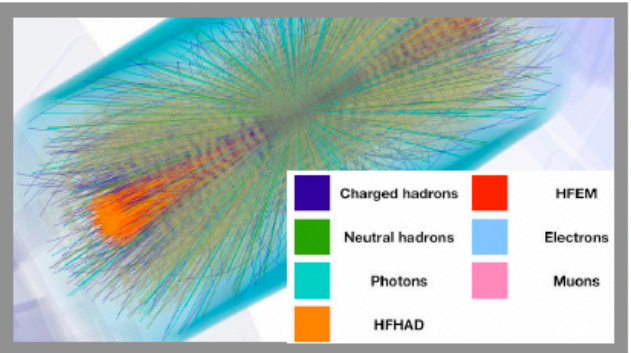
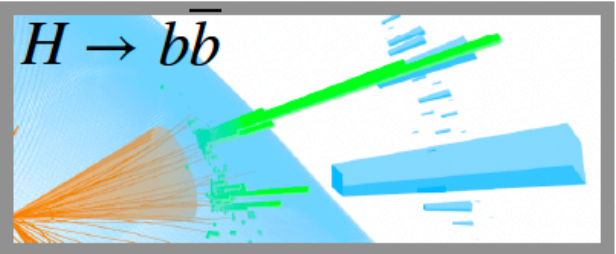
Jet Energy
Correction

Particle-Flow
Reconstruction

Pileup
Mitigation

Event
Analysis

Anomaly
Detection



Naturalness background

- Radiative correction of Higgs mass from any loop of a particle it couples to comes in quadratically in the Planck scale Λ
 - For the observed, tiny Higgs mass to emerge, you'd need this enormous quantum correction to cancel to a conspiratorial **~34 decimal places**
 - = a fine-tuned theory where the low-energy physics is exquisitely, coincidentally sensitive to whatever happens at arbitrarily high, unknown scales.
 - if nature is "natural," there should be new particles near the TeV scale that cancel the dominant (top-loop) quadratic divergence (but none so far!)- any of these partners, if they exist and couple to the top sector, would shift the effective top-Higgs coupling away from its SM value
 - A measured self-coupling that disagrees with the SM prediction is direct, model-independent evidence the assumed Higgs potential isn't the minimal SM one
 - new CP-violating operators in the top-Higgs sector are exactly what you'd expect if naturalness is solved by new top-coupled states (a stop, a top partner, or a composite resonance can carry its own CP phase). So CP violation is a diagnostic signature to look for as a consequence of a naturalness-motivated model

$$\delta m_H^2 \sim \frac{y_t^2}{16\pi^2} \Lambda^2$$

~ 1 $\sim 10^9 \text{ GeV}$

$$= 10^{34} \times (125 \text{ GeV})^2$$

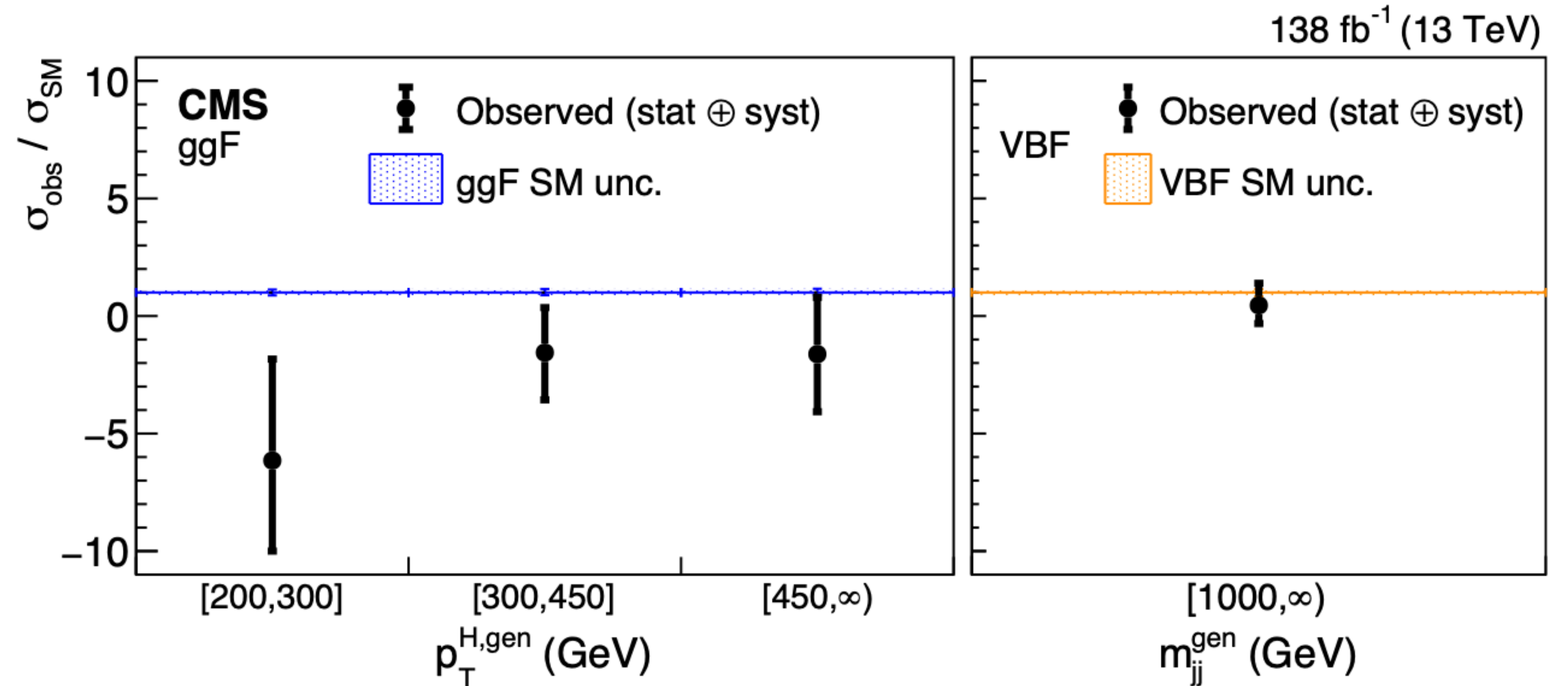
Unfolding background

Unfolding

- Detector measurements are **smeared**: finite resolution, reconstruction inefficiencies, and limited acceptance distort the true kinematic distribution
- **Unfolding** = statistically inverting that detector response to recover the best estimate of the true, particle-level distribution
- Uses a response matrix (built from simulation) mapping true $\rightarrow$ reconstructed bins, inverted with regularization (naive inversion amplifies noise)
- **Why bother**: an unfolded result is detector-independent — reusable by any theorist or any other experiment, without needing to rerun your detector simulation

STXS (Simplified Template Cross Sections)

- Standardized binning scheme: split cross sections by **production mode** (ggF , VBF , VH , $t\bar{t}H$...) and **kinematics** (e.g. Higgs p_T ranges, jet multiplicity)
- A middle ground between:
 - **Too inclusive** (one total rate) $\rightarrow$ loses sensitivity, blends regions with very different BSM sensitivity
 - **Too exclusive** (analysis-specific categories) $\rightarrow$ not reusable/comparable across experiments or reinterpretable years later
- Bins chosen to (a) isolate regions where different BSM effects show up distinctly, (b) minimize theory uncertainty within each bin



Four top theory

<https://arxiv.org/pdf/1504.04630>
<https://arxiv.org/pdf/1602.01934>

Why $t\bar{t}H$ and $t\bar{t}t\bar{t}$ matter

- $y_t \sim 1$ is the largest SM coupling $\rightarrow$ dominant driver of the Higgs mass's quadratic radiative correction $\rightarrow$ **the** direct handle on naturalness
- $t\bar{t}H$: only **tree-level, direct** access to y_t — unlike loop-level probes ($ggF, H \rightarrow \gamma\gamma$) which can't distinguish a shifted y_t from a new colored particle in the loop
- But $t\bar{t}H$ alone can't isolate y_t : signal strength $\propto \kappa_t^2/\Gamma_H$ — coupling and width are **degenerate**
- $t\bar{t}t\bar{t}$ breaks the degeneracy: Higgs-mediated diagram $\propto y_t^4$, **independent of Higgs decay** (no on-shell Higgs, no branching-ratio dependence)
- $\rightarrow$ Combined strategy: pin down y_t from $t\bar{t}t\bar{t}$, then extract Γ_H (or other couplings) from $t\bar{t}H$
- $t\bar{t}t\bar{t}$ also has **enhanced EFT sensitivity** — multiple four-fermion operator insertions + high partonic threshold energy $\rightarrow$ constrains $qqtt$ operators competitively with $t\bar{t}$, despite far lower precision

Why the theory prediction keeps increasing at each order

- **LO $\rightarrow$ NLO QCD**: new real-emission channel opens ($qg \rightarrow t\bar{t}t\bar{t}q$), boosted by the large gluon PDF
- **NLO QCD $\rightarrow$ complete NLO (QCD+EW)**: large, *individually* sizeable EW terms ($O(\alpha_s^3\alpha), O(\alpha_s^2\alpha^2)$) from genuine $t\bar{t} \rightarrow t\bar{t}$ EW-scattering diagrams — mostly cancel inclusively, but the cancellation is "accidental," and effects are large near threshold / high- p_T
- **Fixed order $\rightarrow$ threshold resummation (NLL')**: $t\bar{t}t\bar{t}$ sits close to its ~ 700 GeV kinematic threshold $\rightarrow$ soft-gluon exchange is enhanced and **attractive** in the dominant color channels $\rightarrow$ resumming these logs gives a genuinely positive, one-directional +26–27% shift, and **shrinks the scale uncertainty by $\sim 2\times$**
- Common root cause: unusually **high partonic energy threshold** + unusually **many colored final-state particles** (6 colored objects) $\rightarrow$ far more new channels/topologies/color-exchange configurations than a typical $2\rightarrow 2$ process, so each new order isn't just "shrinking corrections," it's uncovering genuinely new physical contributions

EFT background

- If new physics lives at a high scale Λ , you don't need the full UV theory to describe LHC-scale data
- Integrate out heavy states $\rightarrow$ encode their effects as higher-dimension operators added to the SM Lagrangian:

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i}{\Lambda^{d-4}} O_i^{(d)}$$

- **Model-independent:** parametrizes *any* possible new physics as coefficient deviations, rather than committing to one specific UV model (2HDM, SUSY, composite Higgs...)
- **SMEFT** = the standard realization (SM gauge symmetry + particle content below Λ); dimension-6 "Warsaw basis" = 59 independent operators
- Global fits (e.g. SMEFiT) combine many processes to simultaneously constrain dozens of Wilson coefficients, breaking degeneracies no single measurement can resolve alone

Why $t\bar{t}t\bar{t}$ is especially useful for EFT constraints

- Sensitive to $qqtt$ -type four-fermion contact operators — exactly what a heavy colored resonance leaves behind once integrated out
- **Enhanced EFT sensitivity** (Zhang 2018): constraining power on these operators is already competitive with $t\bar{t}$ measurements, despite $t\bar{t}$ being $>100\times$ more precise
- Two compounding reasons:
 - **Multiple operator insertions** in the squared amplitude $\rightarrow$ leading BSM term can scale as $(C/\Lambda^2)^2$, not just linear interference
 - **High partonic threshold energy** ($\sqrt{\hat{s}} \gtrsim 700$ GeV) $\rightarrow$ operator contributions grow with energy, and $t\bar{t}t\bar{t}$ sits unusually deep in that regime
- Rare, small-rate process — but that same kinematic extremity is exactly what gives it outsized leverage per event

