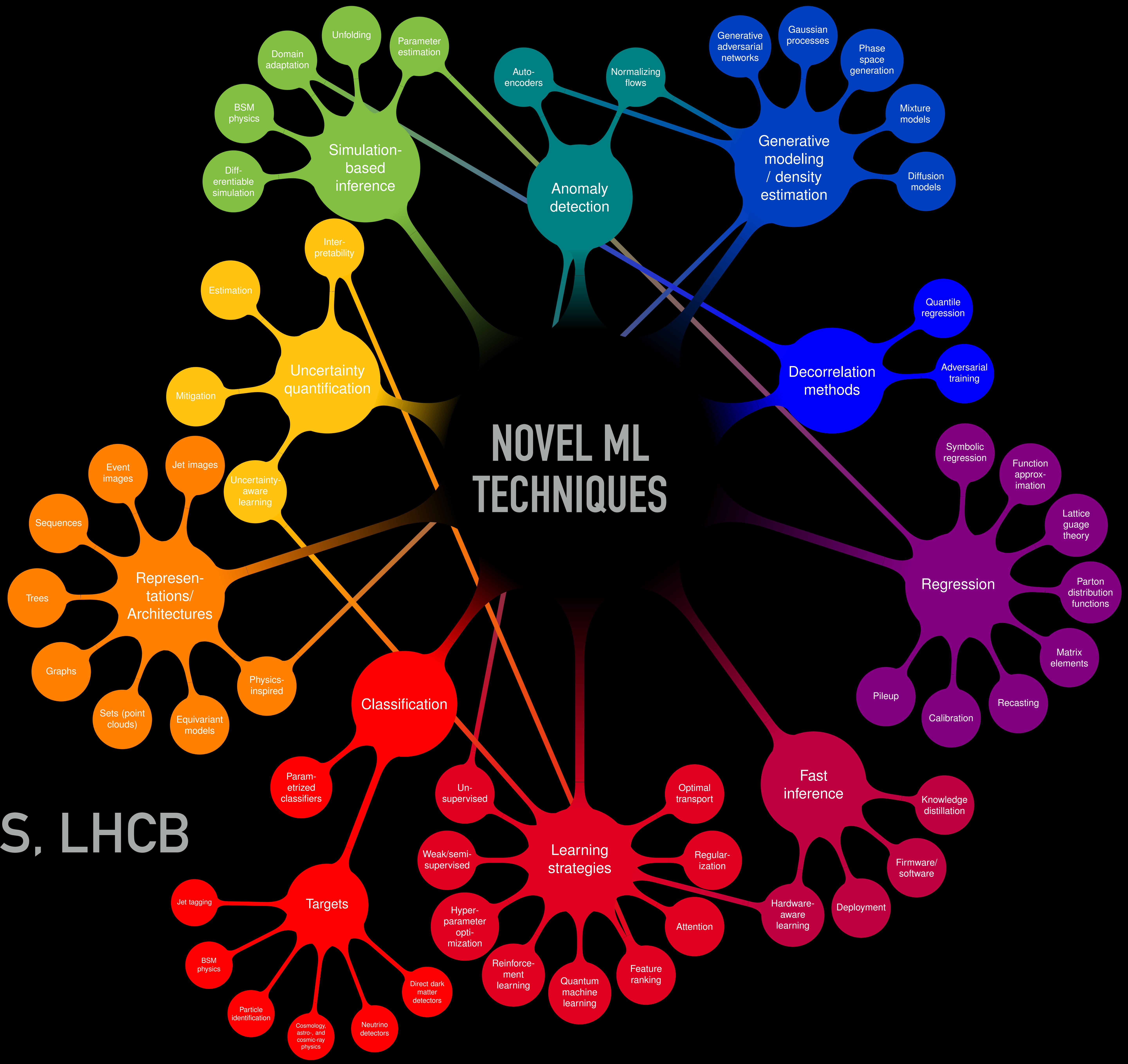
NOVEL ML TECHNIQUES

JAVIER DUARTE

ON BEHALF OF
ALICE, ATLAS, CMS, LHCB

ICHEP 2024

JULY 23, 2024



JAVIER DUARTE

ON BEHALF OF ALICE, ATLAS, CMS, LHCB

ICHEP 2024

JULY 23, 2024

PARTICLE PHYSICS + MACHINE LEARNING

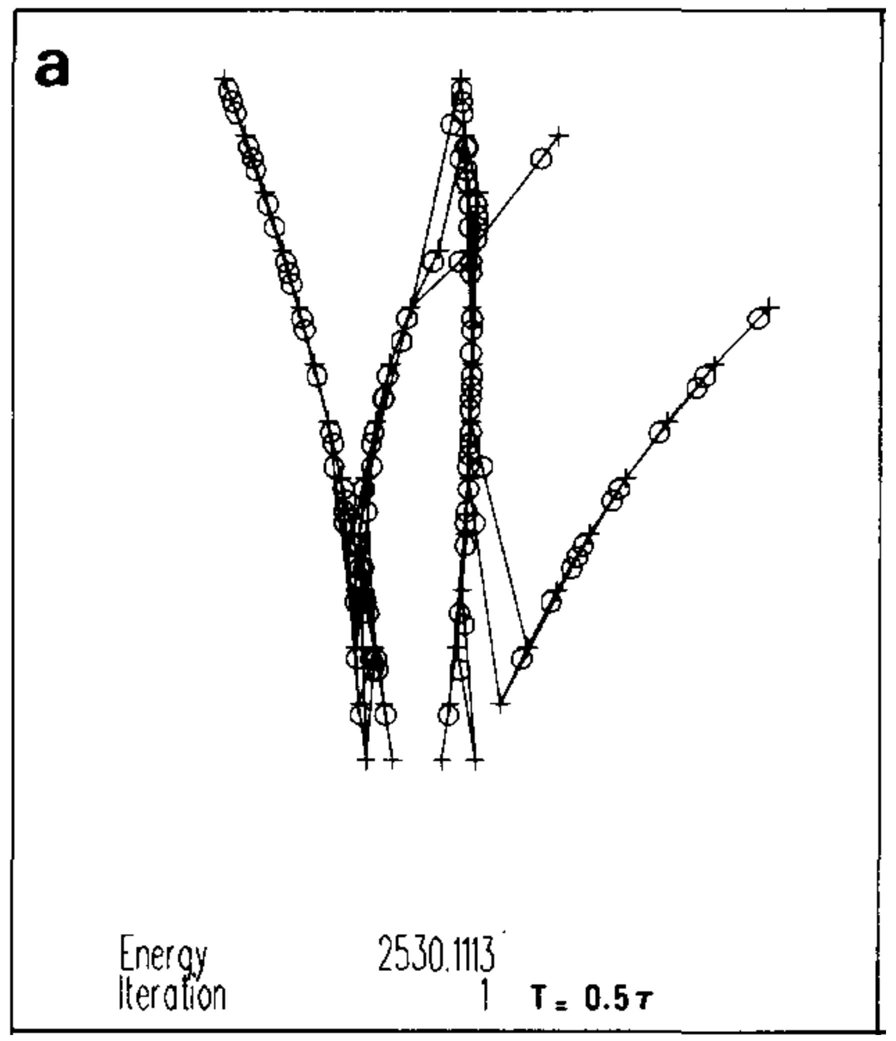
Comp. Phys. Comm. 49, 429 (1988)
Nature 560, 41 (2018)

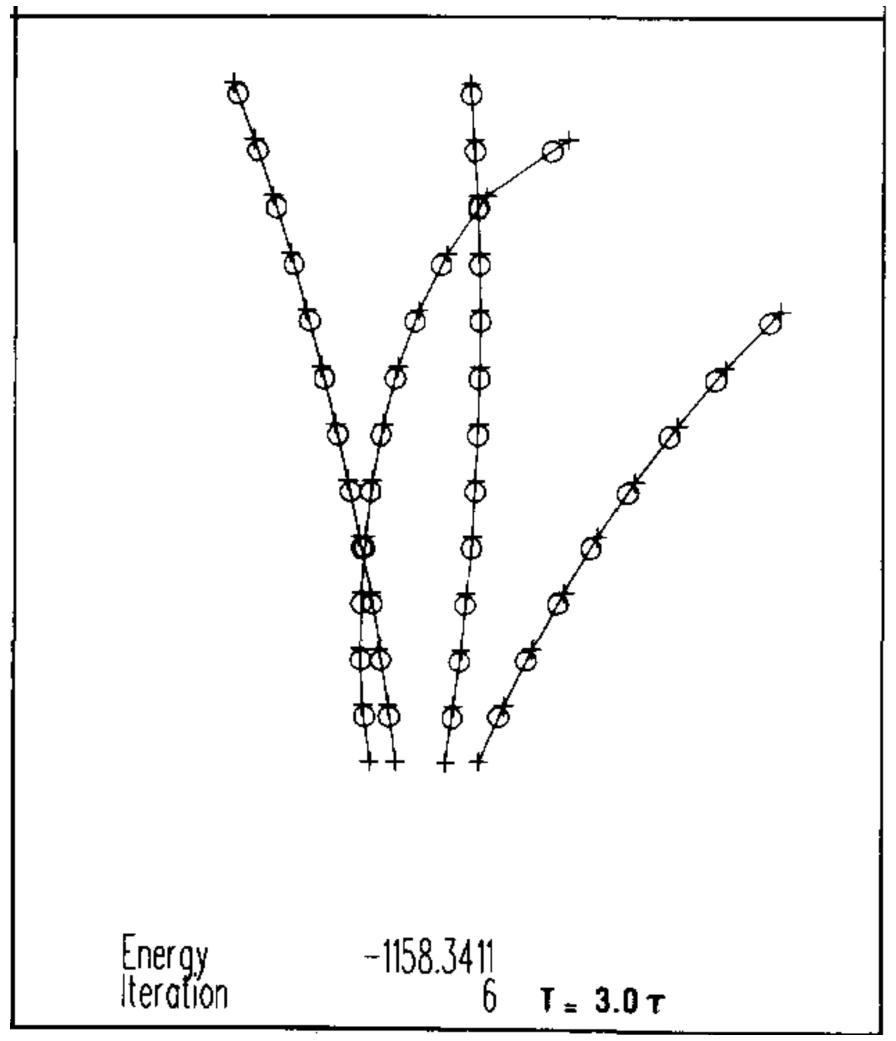
- Particle physics has a long history of applying machine learning

PARTICLE PHYSICS + MACHINE LEARNING

[Comp. Phys. Comm. 49, 429 (1988)](#)
[Nature 560, 41 (2018)](#)

- Particle physics has a long history of applying machine learning
  - From the early days in the 1980s applying neural networks to tracking

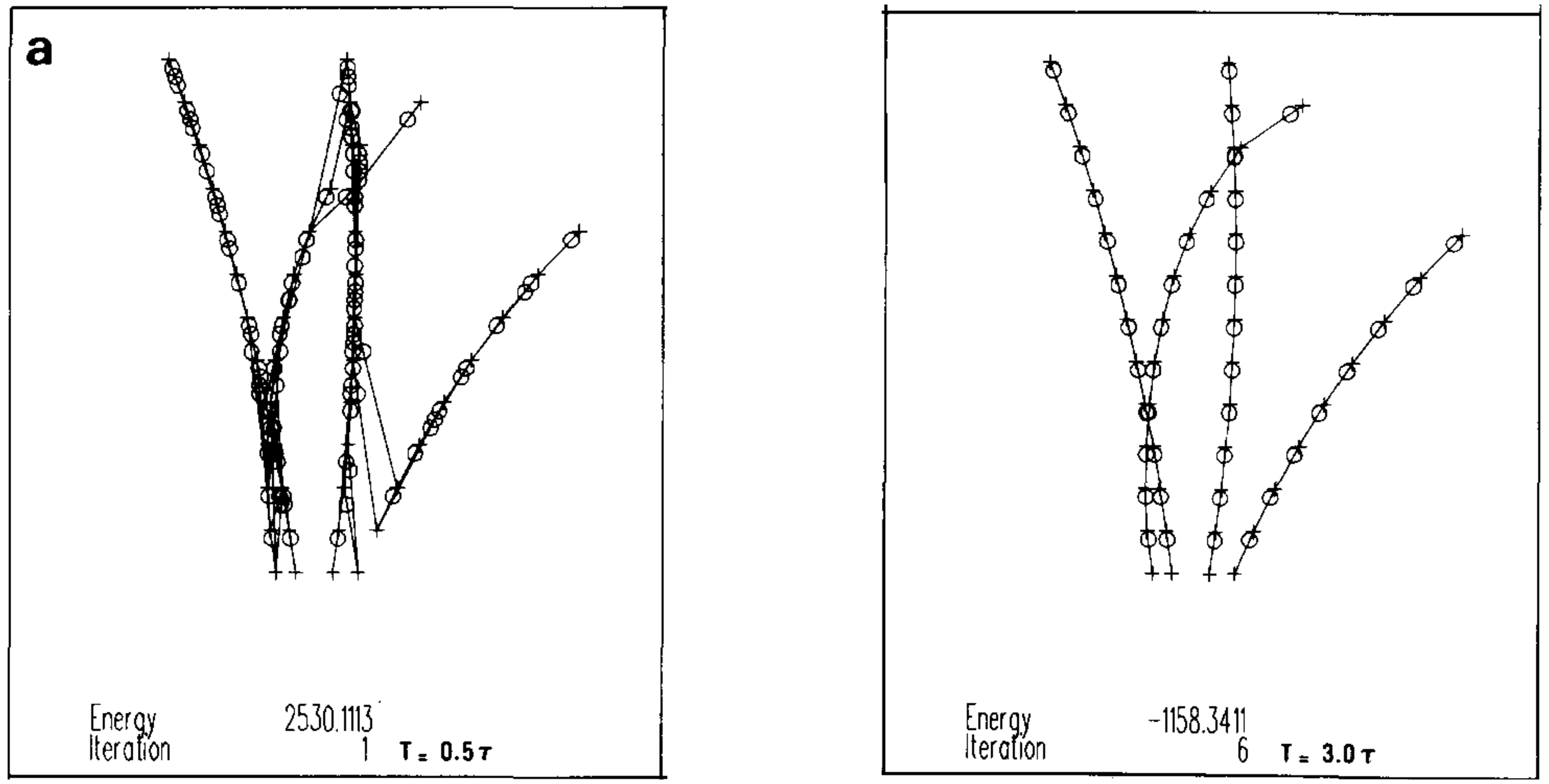


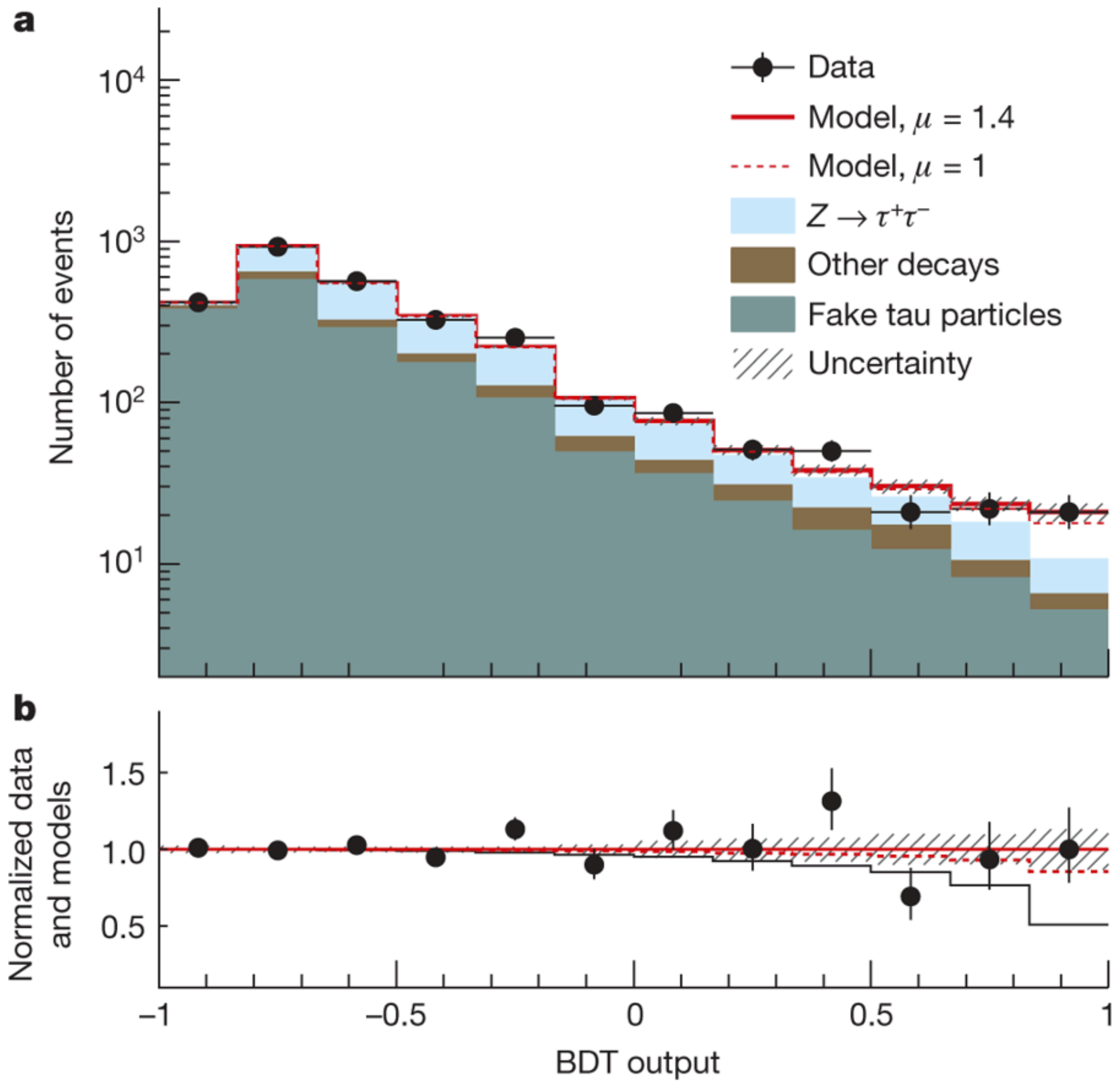


PARTICLE PHYSICS + MACHINE LEARNING

Comp. Phys. Comm. 49, 429 (1988)
Nature 560, 41 (2018)

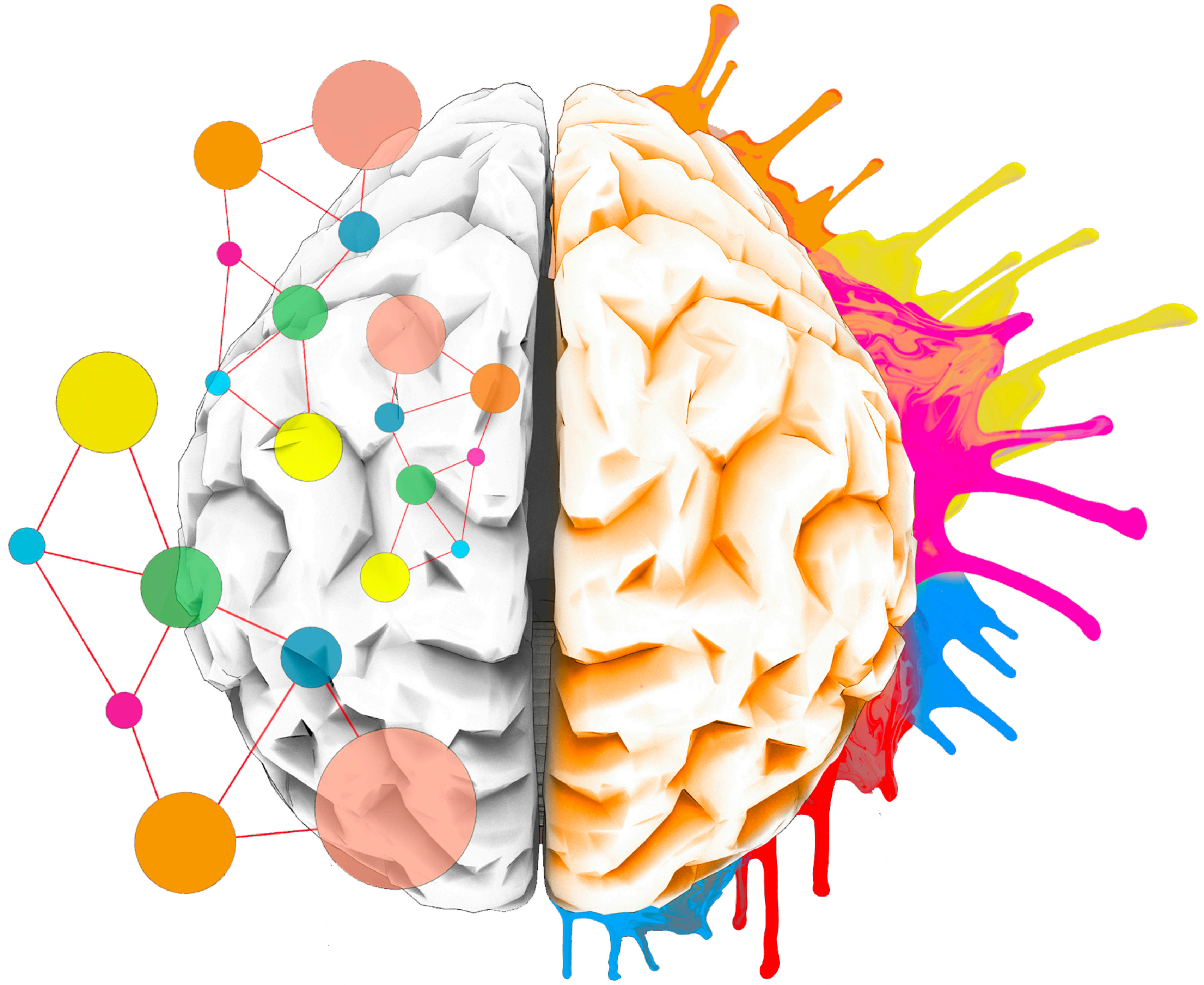
- Particle physics has a long history of applying machine learning
  - From the early days in the 1980s applying neural networks to tracking
  - To the Higgs boson discovery in 2012 applying boosted decision trees to identify $H \to \tau\tau$





INTRODUCTION & OUTLINE

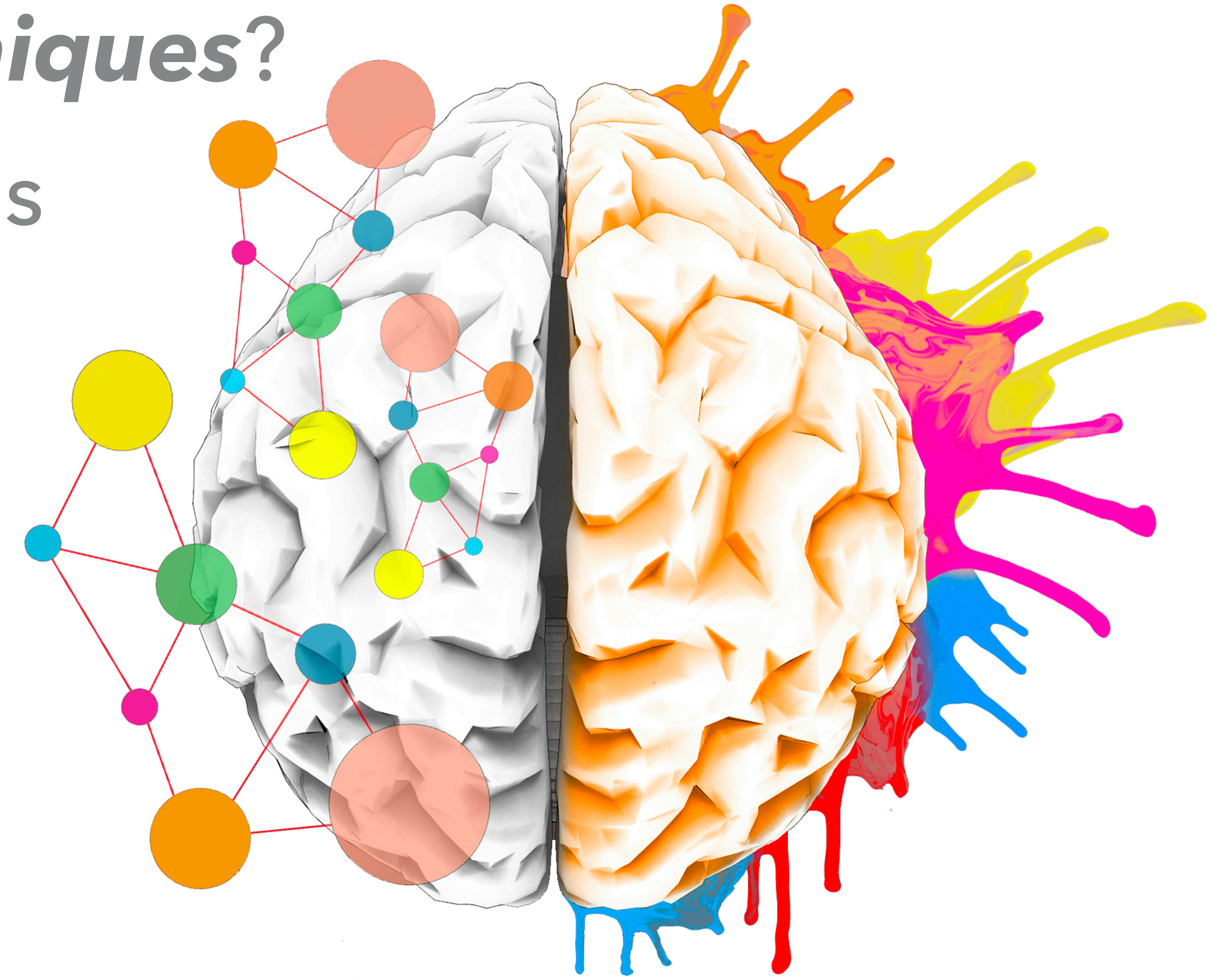
- ML has changed the way we do physics searches and measurements
    - It is an essential and versatile tool that we use to improve existing approaches
    - It enables fundamentally new approaches



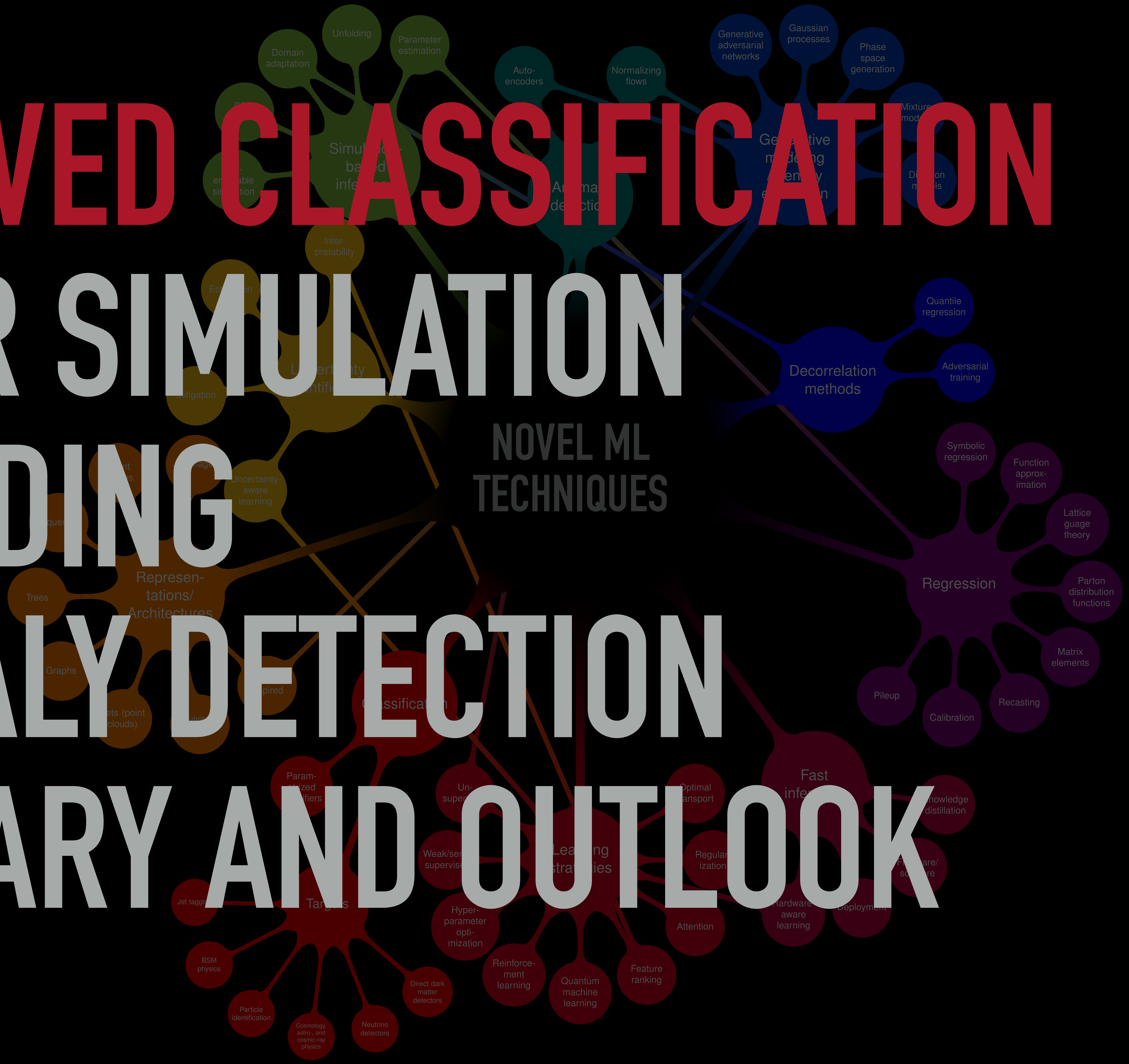
INTRODUCTION & OUTLINE

- ML has changed the way we do physics searches and measurements
  - It is an essential and versatile tool that we use to improve existing approaches
  - It enables fundamentally new approaches

- How are LHC experiments applying ***novel ML techniques***?
  - Improved classification, calibration, & uncertainties
  - Faster simulation
  - Unfolding
  - Anomaly detection
  - Summary and outlook

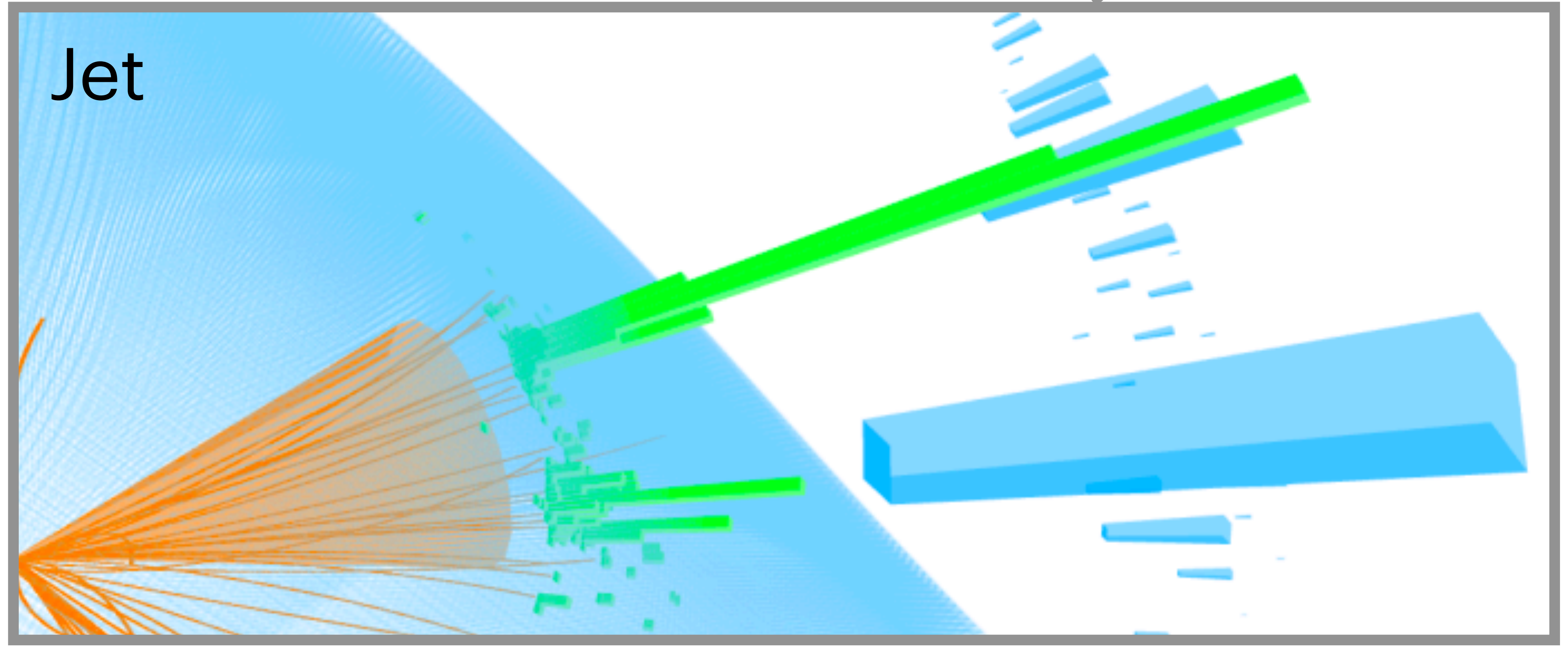


IMPROVED CLASSIFICATION
FASTER SIMULATION
UNFOLDING
ANOMALY DETECTION
SUMMARY AND OUTLOOK



FROM BDTS TO GNNS AND TRANSFORMERS

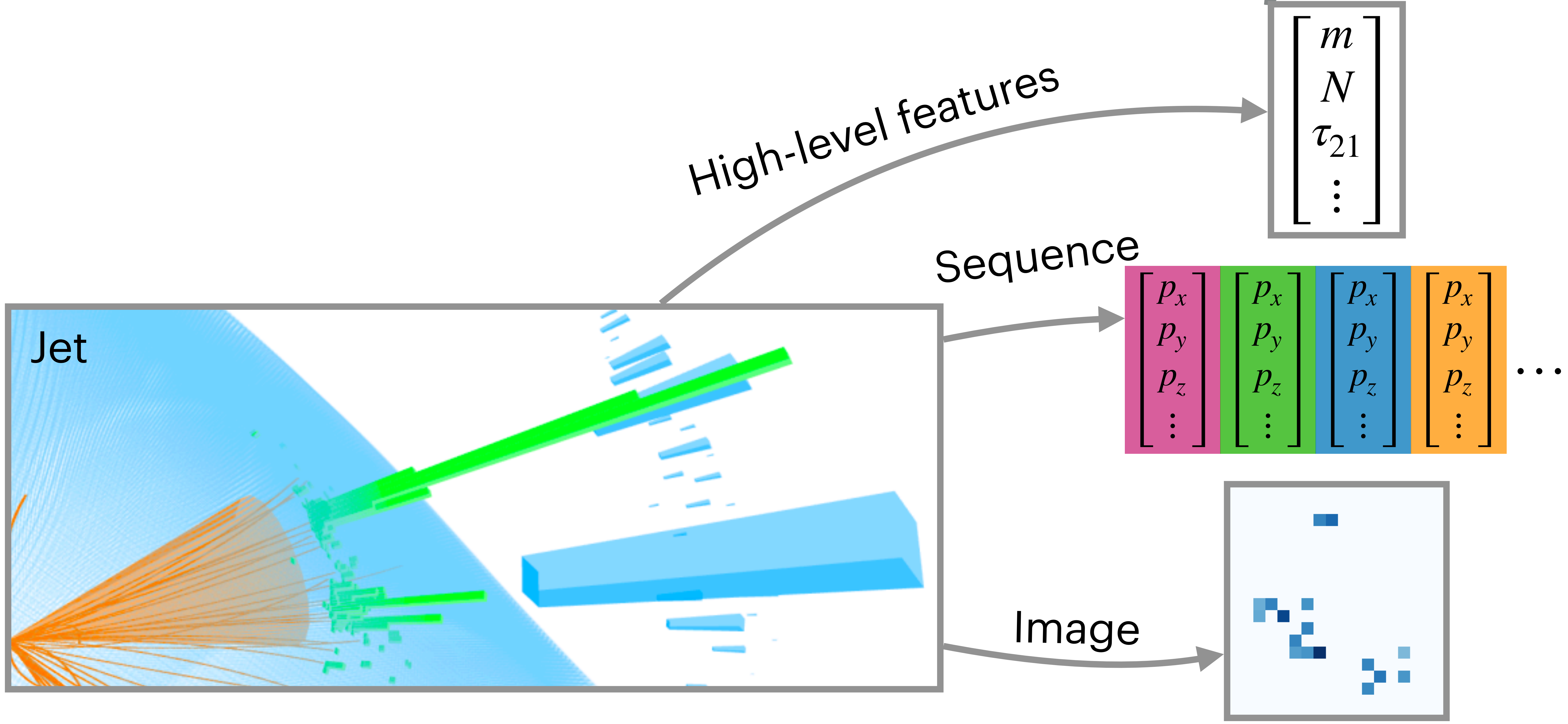
- Great strides made to leverage **rich low-level information** with **graph neural networks** and **transformers** for a variety of tasks including **jet classification**



High-level features

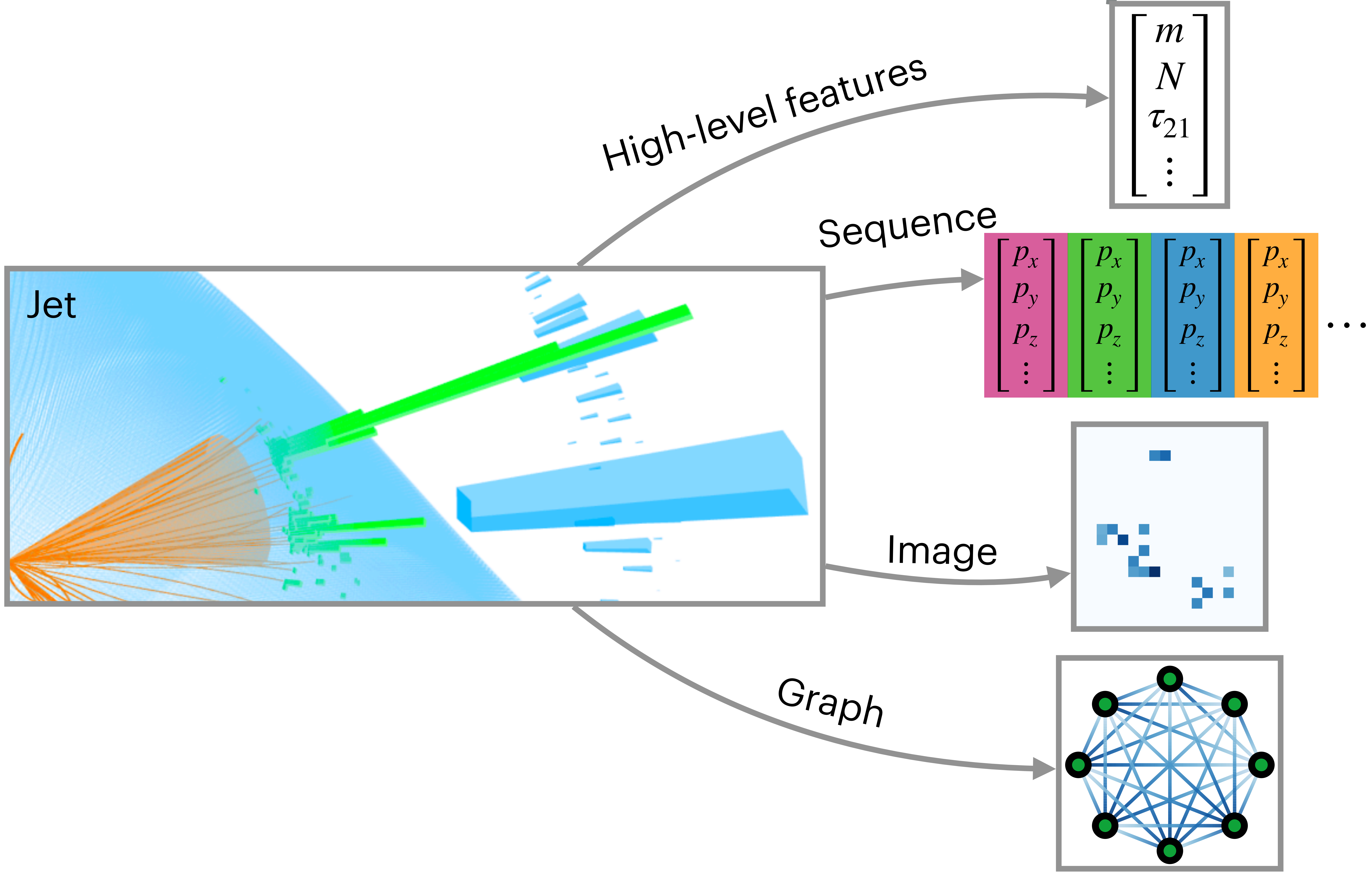
$$\begin{bmatrix} m \\ N \\ \tau_{21} \\ \vdots \end{bmatrix}$$

- Great strides made to leverage **rich low-level information** with **graph neural networks** and **transformers** for a variety of tasks including **jet classification**



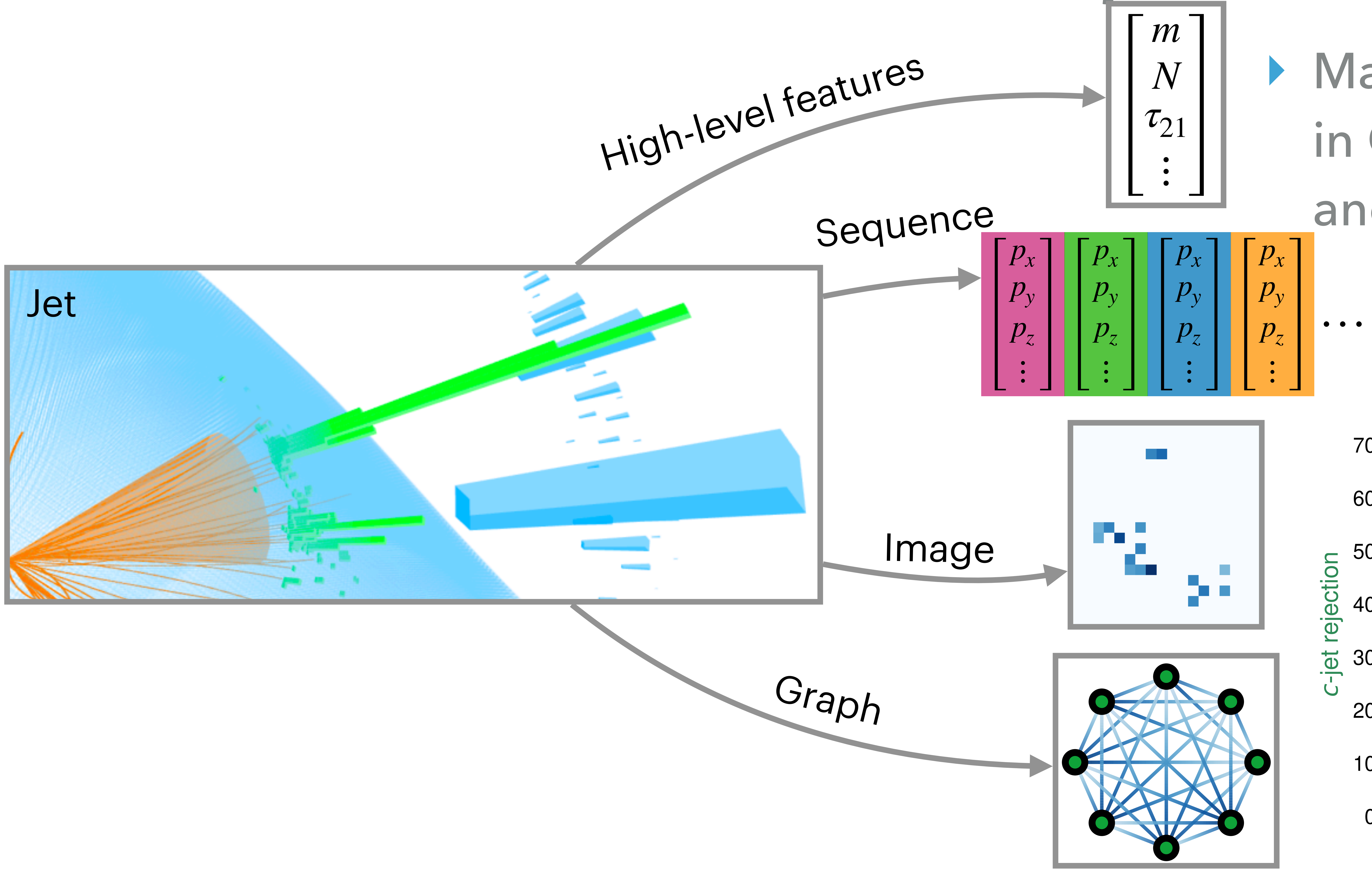
FROM BDTS TO GNNS AND TRANSFORMERS

- Great strides made to leverage **rich low-level information** with **graph neural networks** and **transformers** for a variety of tasks including **jet classification**

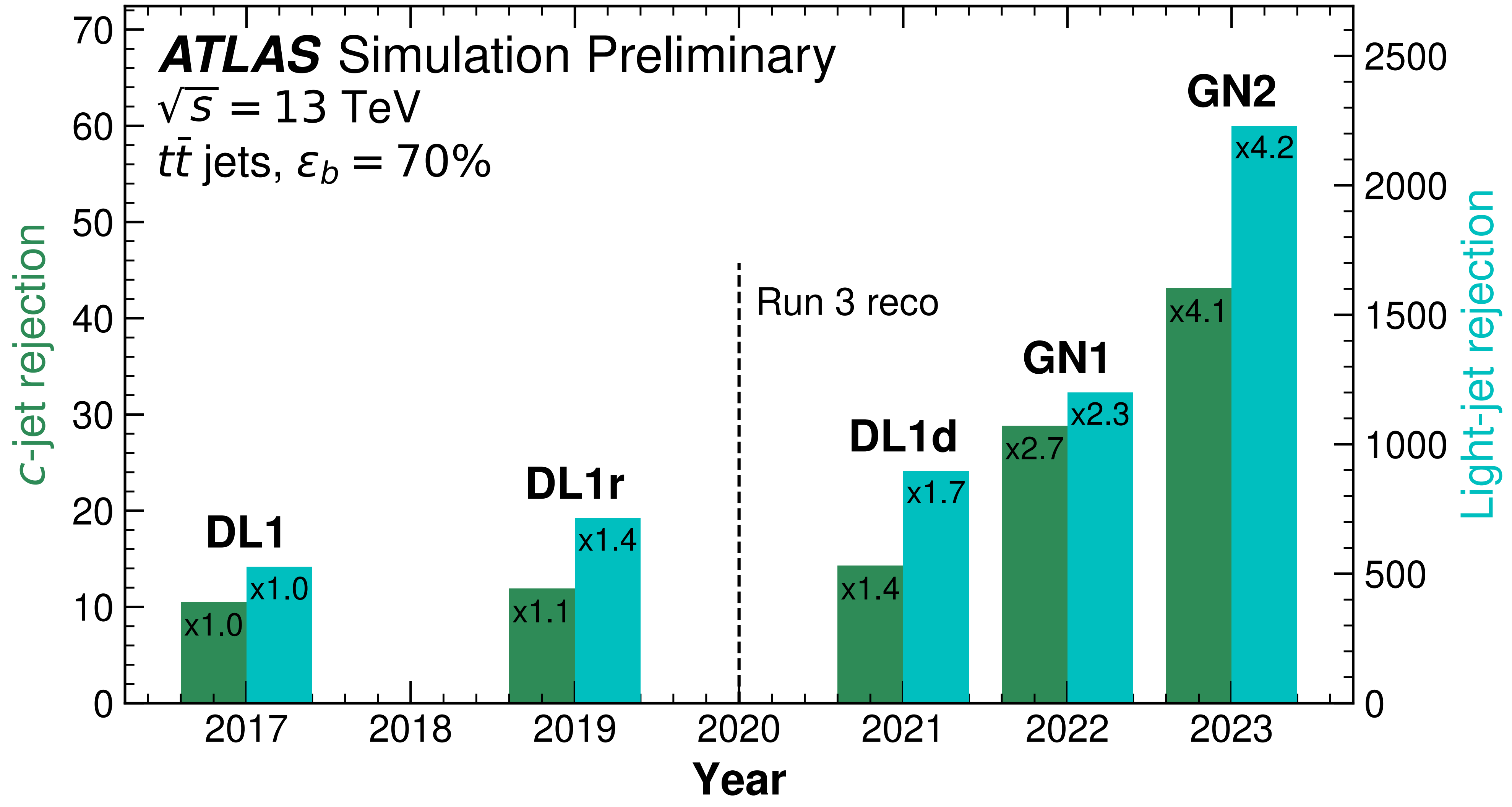


FROM BDTS TO GNNS AND TRANSFORMERS

- Great strides made to leverage **rich low-level information** with **graph neural networks** and **transformers** for a variety of tasks including **jet classification**

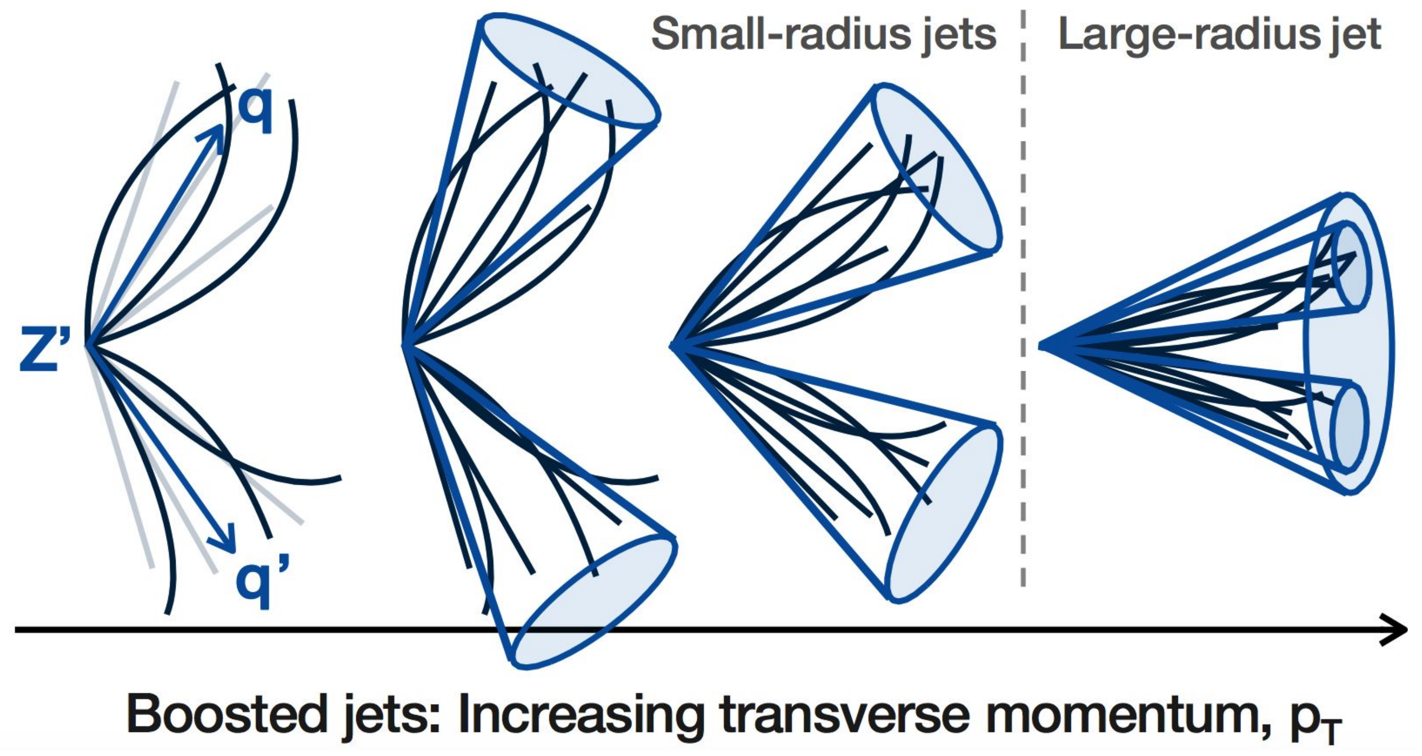


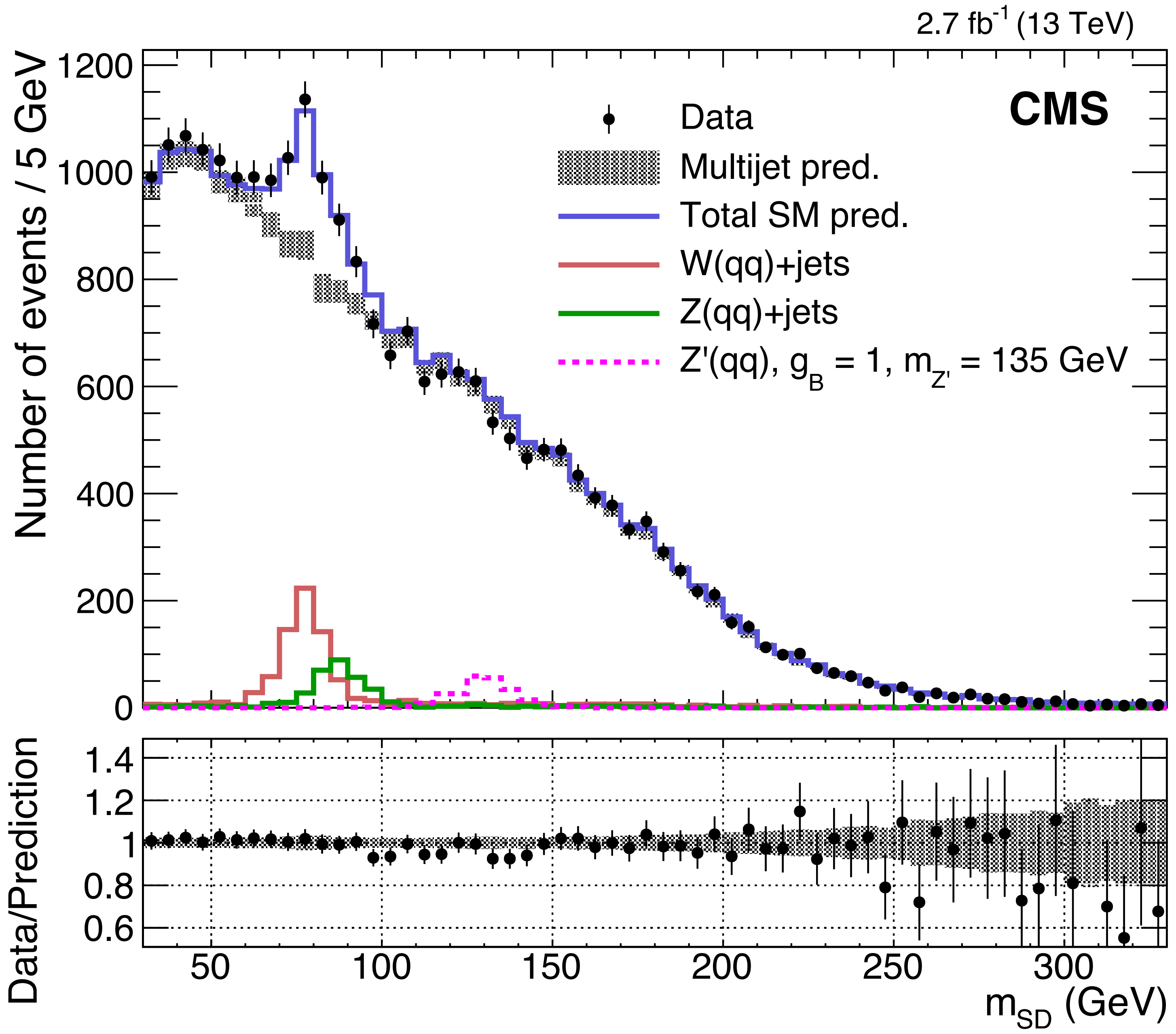
- Major improvements demonstrated in CMS (ParticleNet/Transformer) and **ATLAS (GN1/GN2)**



IMPACT OF GNN TAGGING

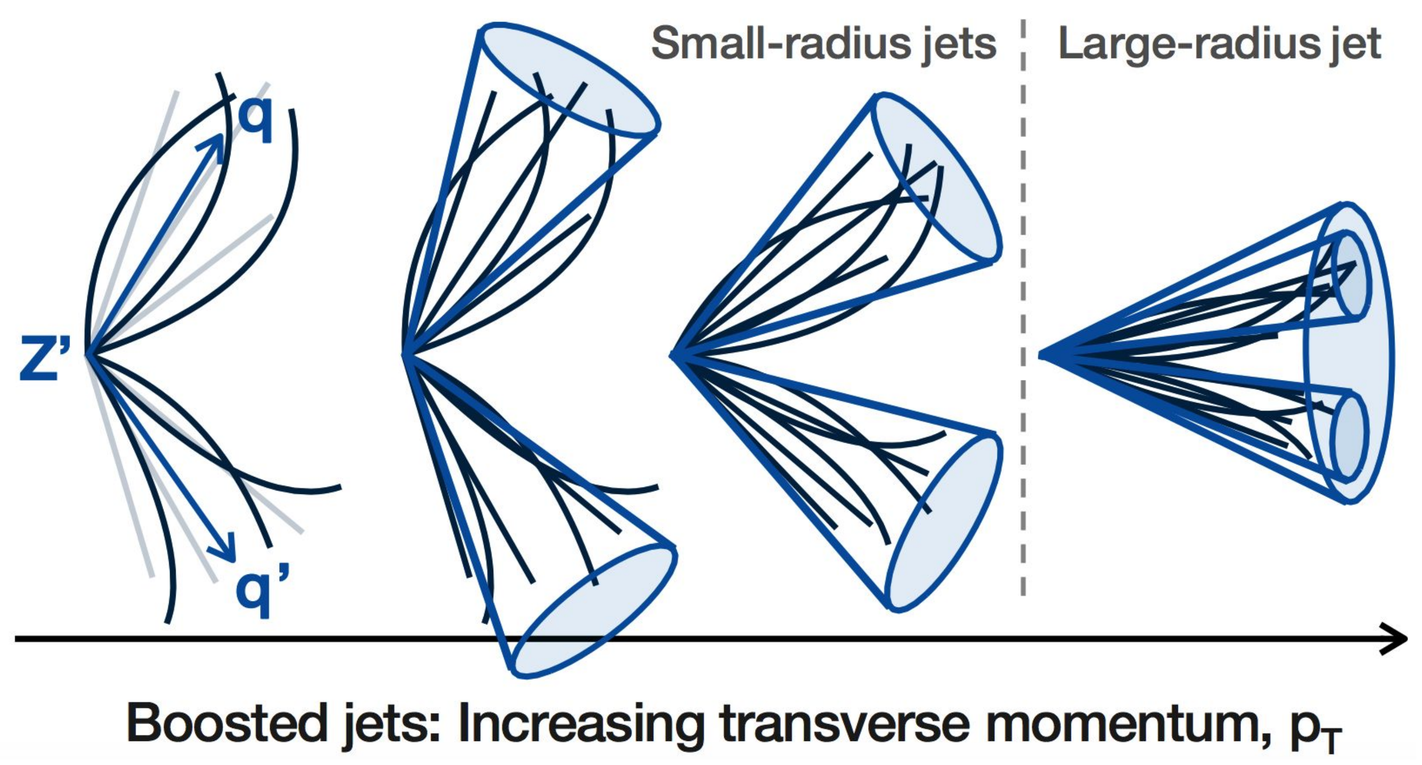
- Previously search for boosted $Z' \to qq$ resonances reconstructed as large-radius jets with substructure

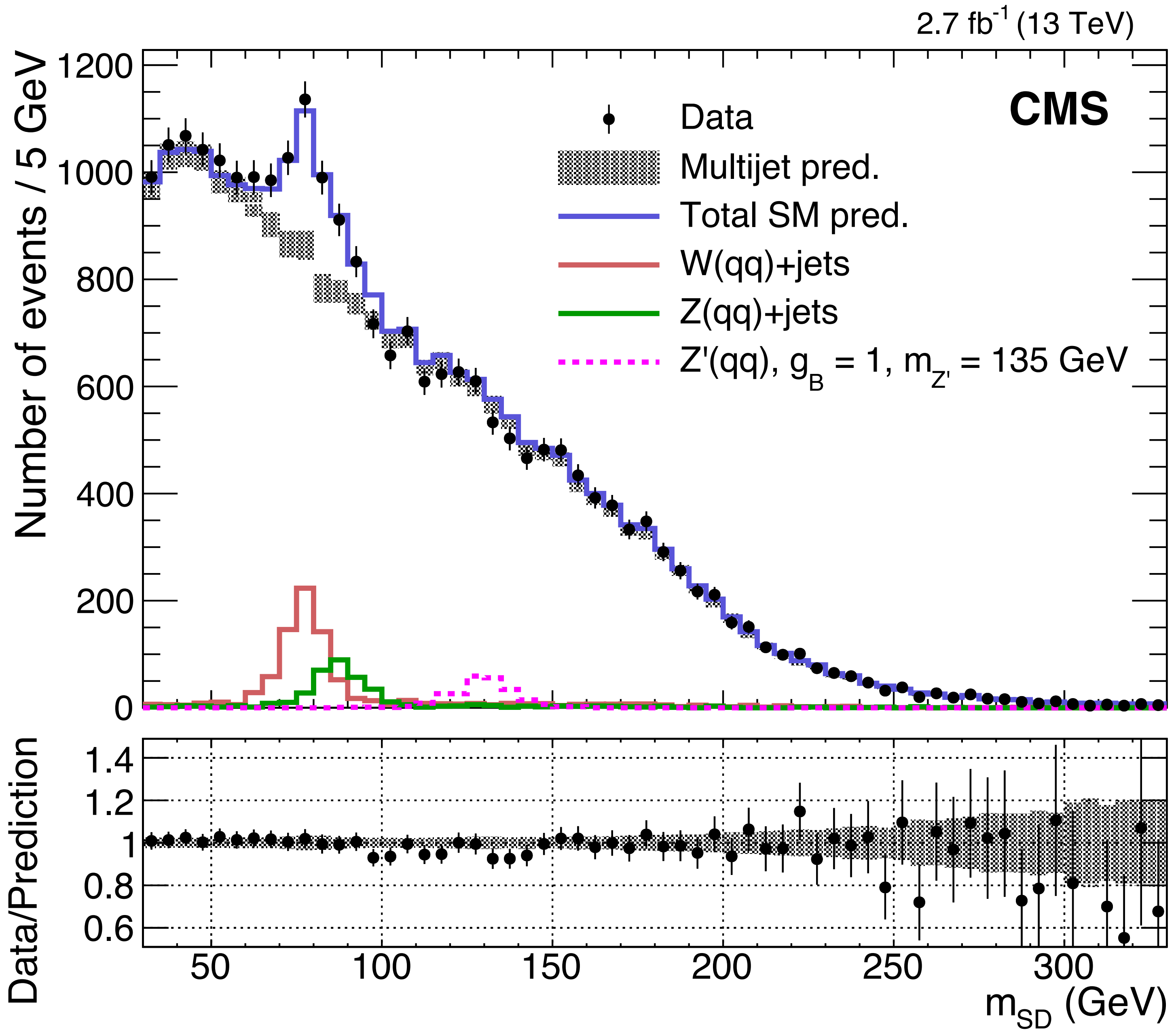


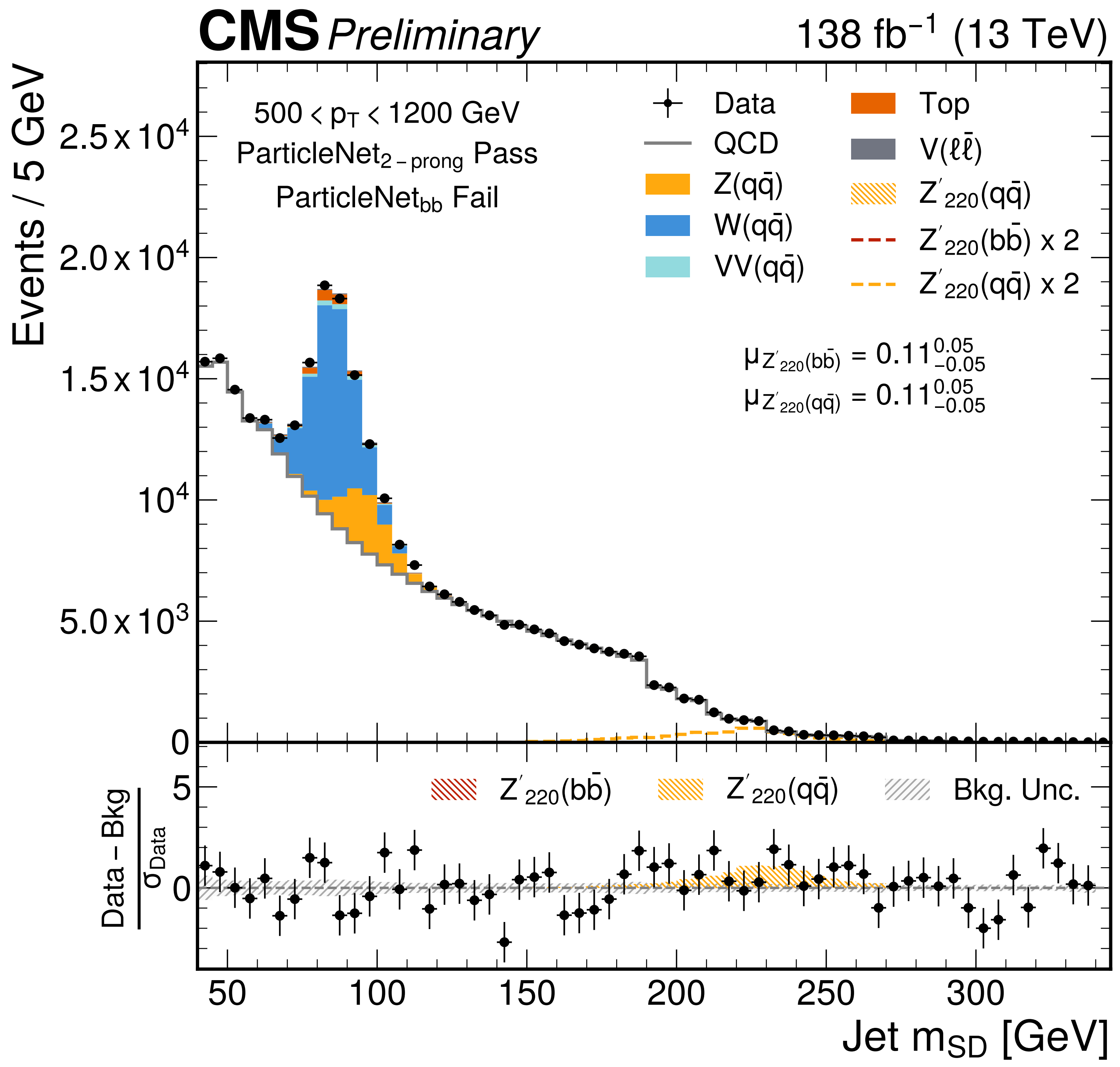


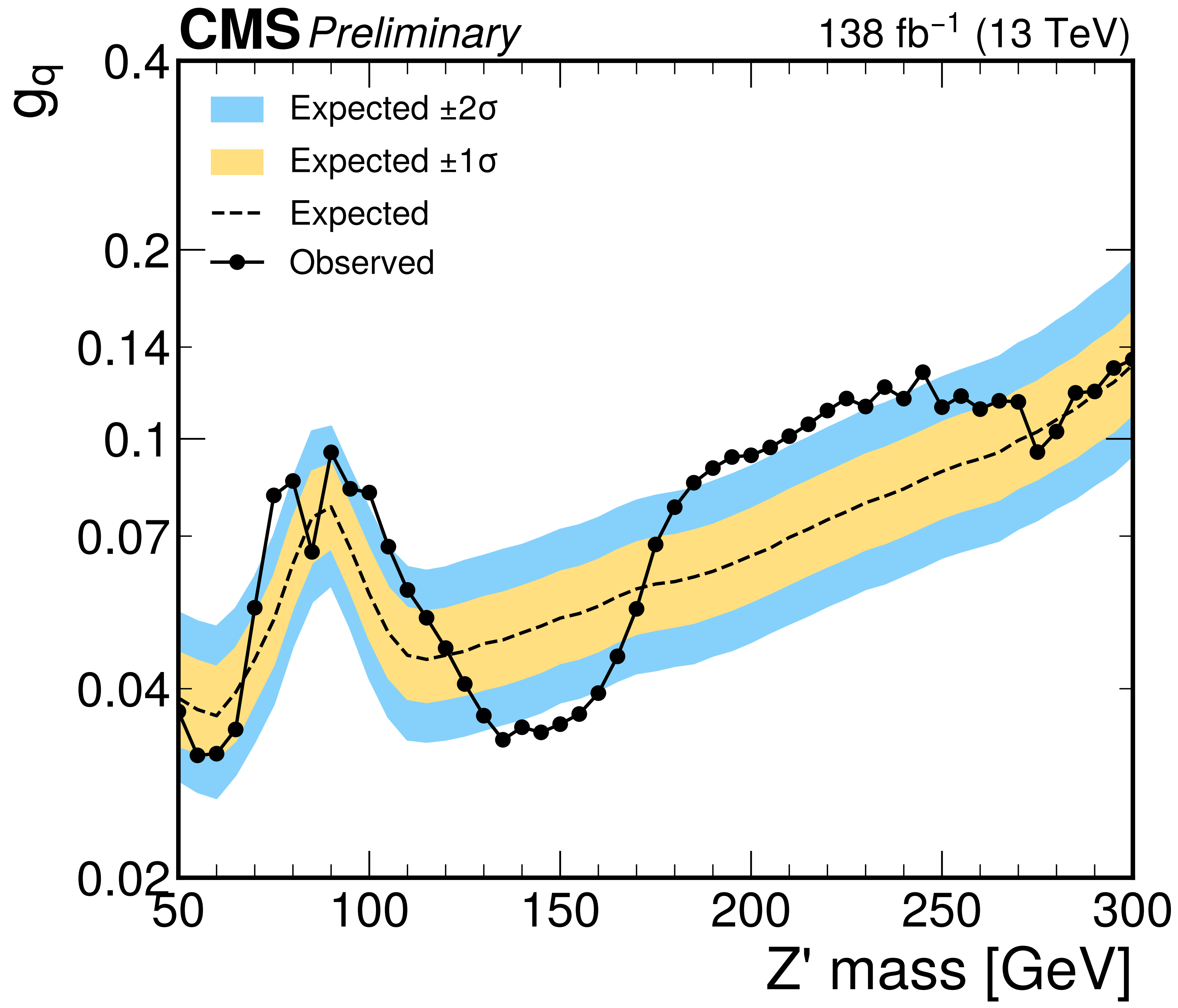
IMPACT OF GNN TAGGING

- Previously search for boosted $Z' \rightarrow qq$ resonances reconstructed as large-radius jets with substructure
- Now signal distinguished from the backgrounds using **ParticleNet GNN** discriminants
  - Stringent limits on universal g_q coupling





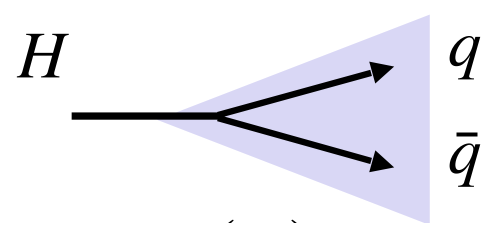


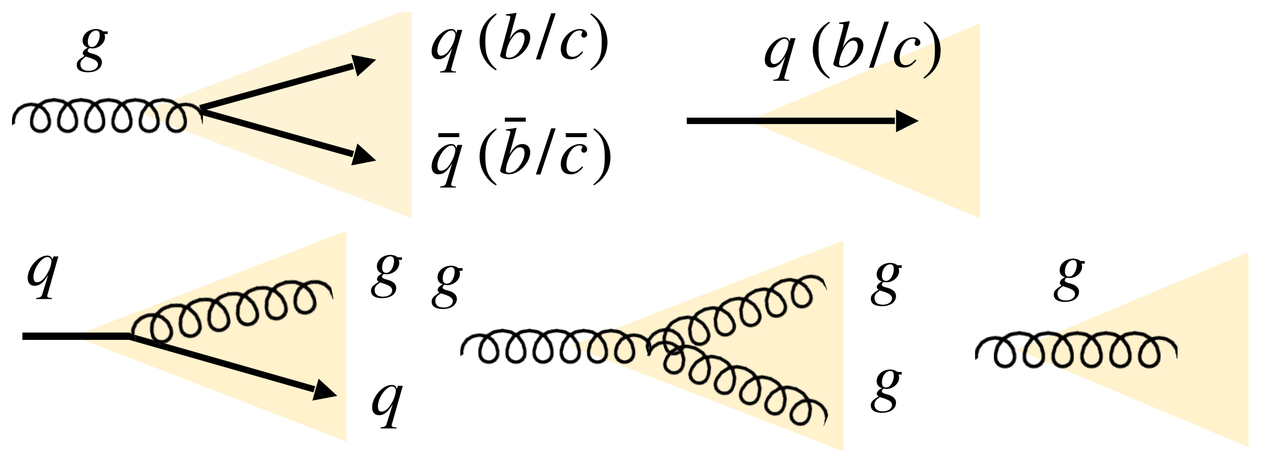


LANDSCAPE OF BOOSTED HIGGS BOSON FINAL STATES

- CMS ParticleNet has been successfully deployed to identify H→qq vs. QCD large-radius jets in mass-agnostic way using a highly granular multiclassifier

Process	Final state/ prongness	heavy flavour	# of classes
H→qq		bb	1
		cc	1
		qq (q=u/d)	1
QCD		b	1
		bb	1
		c	1
		cc	1
		others (light)	1

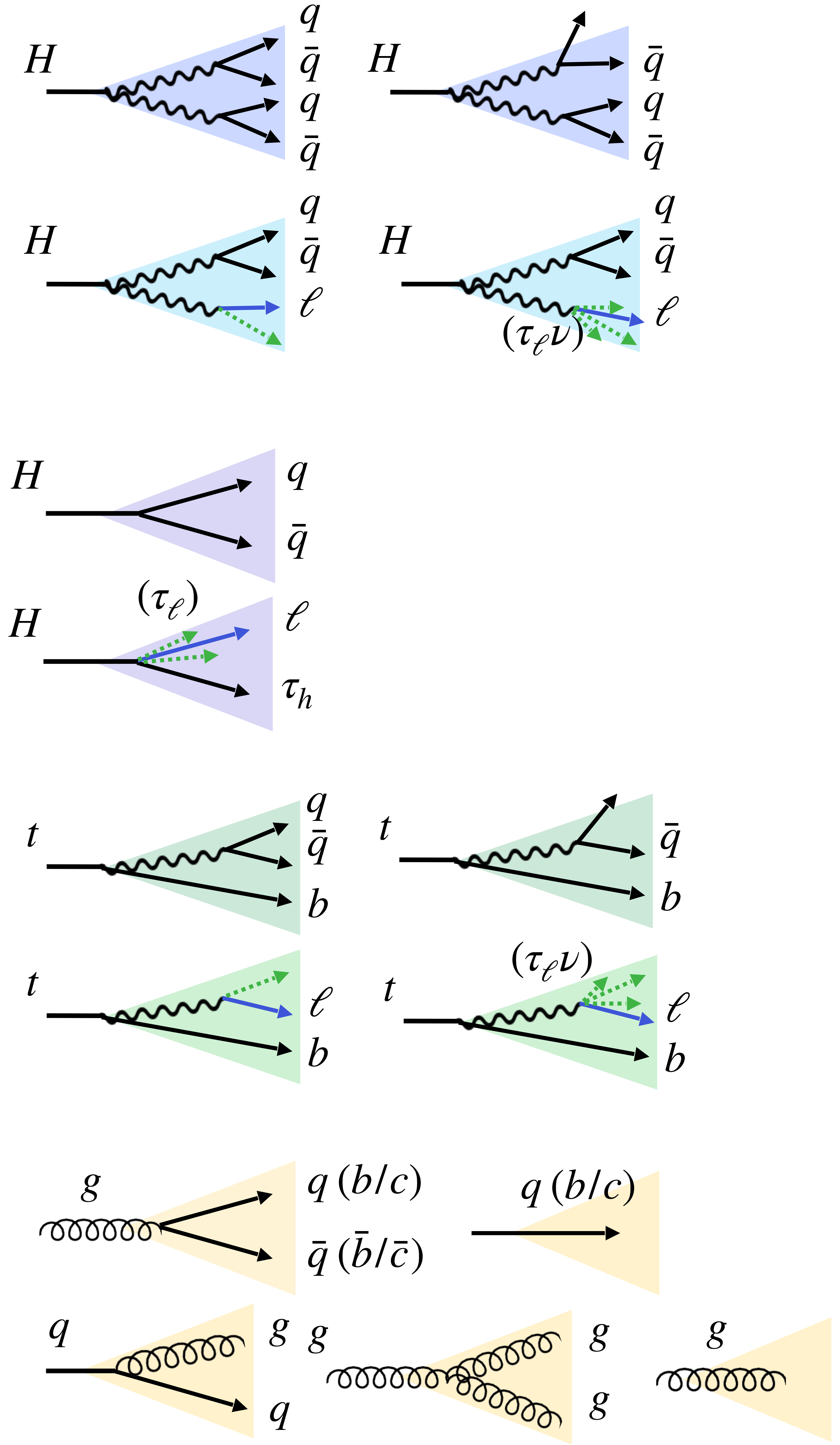




LANDSCAPE OF BOOSTED HIGGS BOSON FINAL STATES

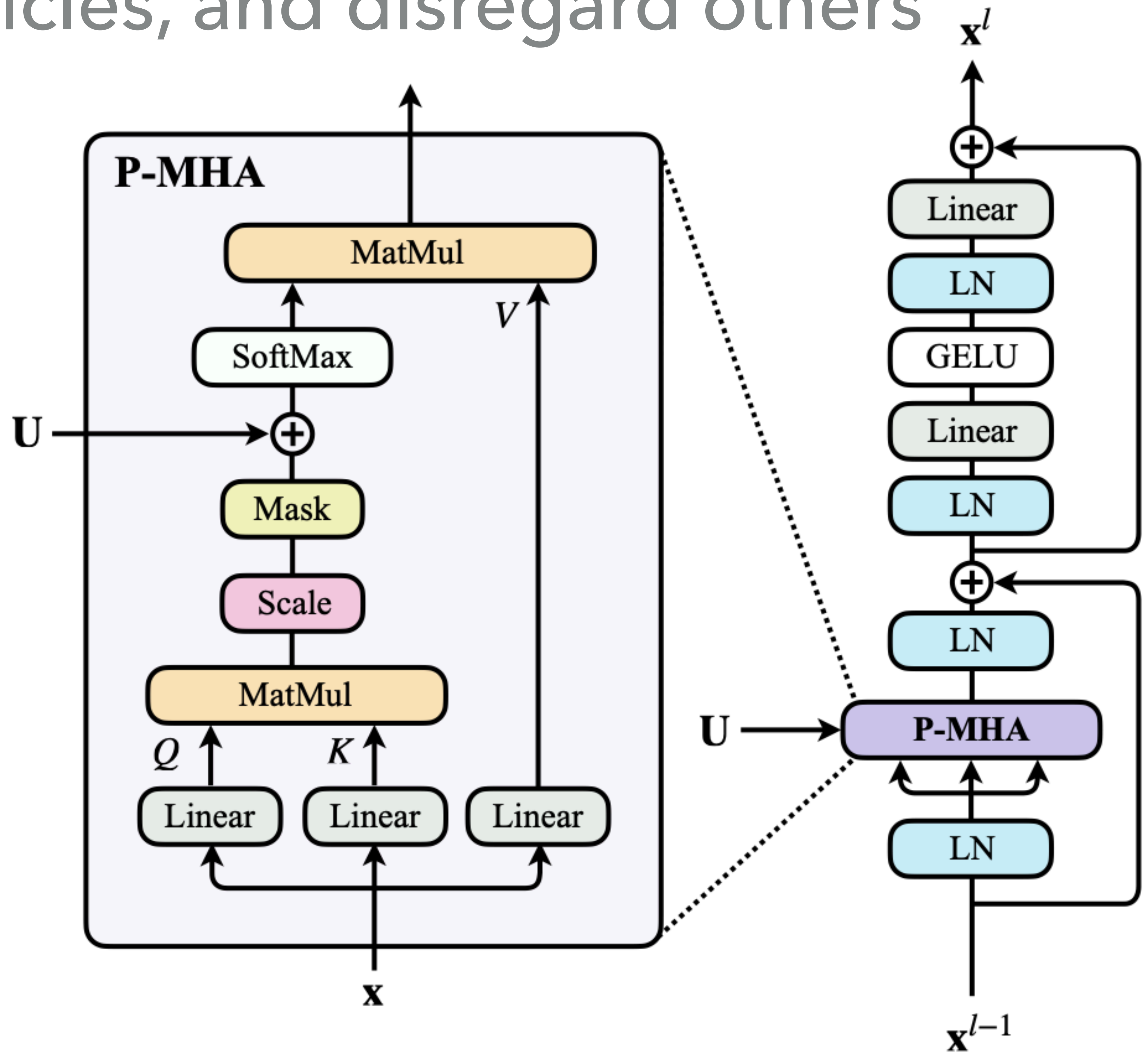
- CMS ParticleNet has been successfully deployed to identify H→qq vs. QCD large-radius jets in mass-agnostic way using a highly granular multiclassifier
- Extend this same approach to a large array of final states, including H→VV, all-hadronic (3- or 4-prong), and semi-leptonic modes

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
	τ_eνqq		2
	τ_μνqq		2
	τ_hνqq		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
	bτ_eν		1
	bτ_μν		1
	bτ_hν		1
QCD		b	1
		bb	1
		c	1
		cc	1
		others (light)	1



GLOBAL PARTICLE TRANSFORMER

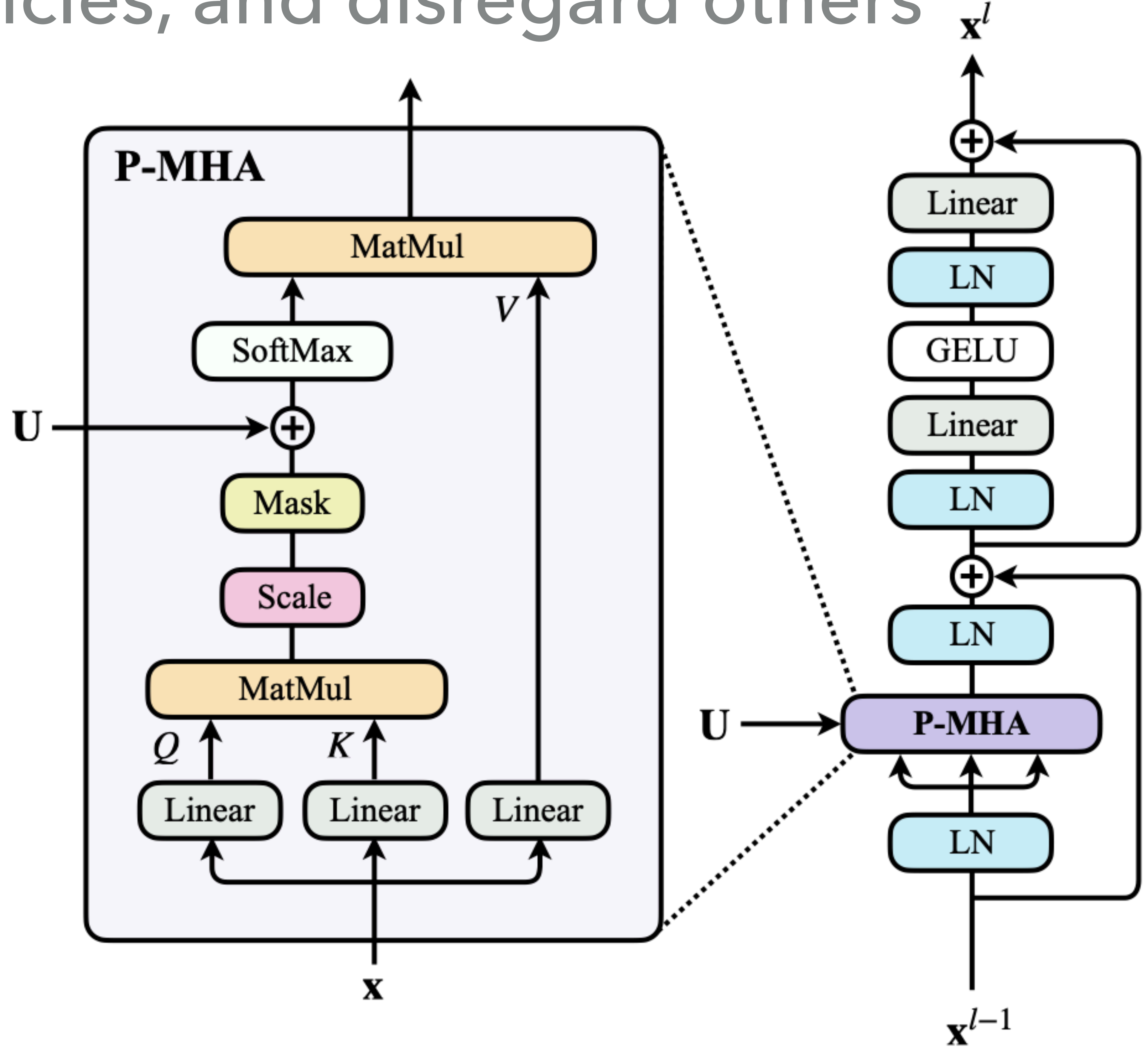
- Global Particle Transformer algorithm uses learned "attention" (and pairwise features) to give more weight to certain particles, and disregard others in order to infer the origin of jets



$$\mathrm{softmax}\left(\frac{QK^\intercal}{\sqrt{d_k}} + U\right)V$$

GLOBAL PARTICLE TRANSFORMER

- Global Particle Transformer algorithm uses learned "attention" (and pairwise features) to give more weight to certain particles, and disregard others in order to infer the origin of jets
- Challenge to calibrate these taggers when there is no SM analogue for the signal we can isolate
  - New technique uses Lund jet plane for calibration



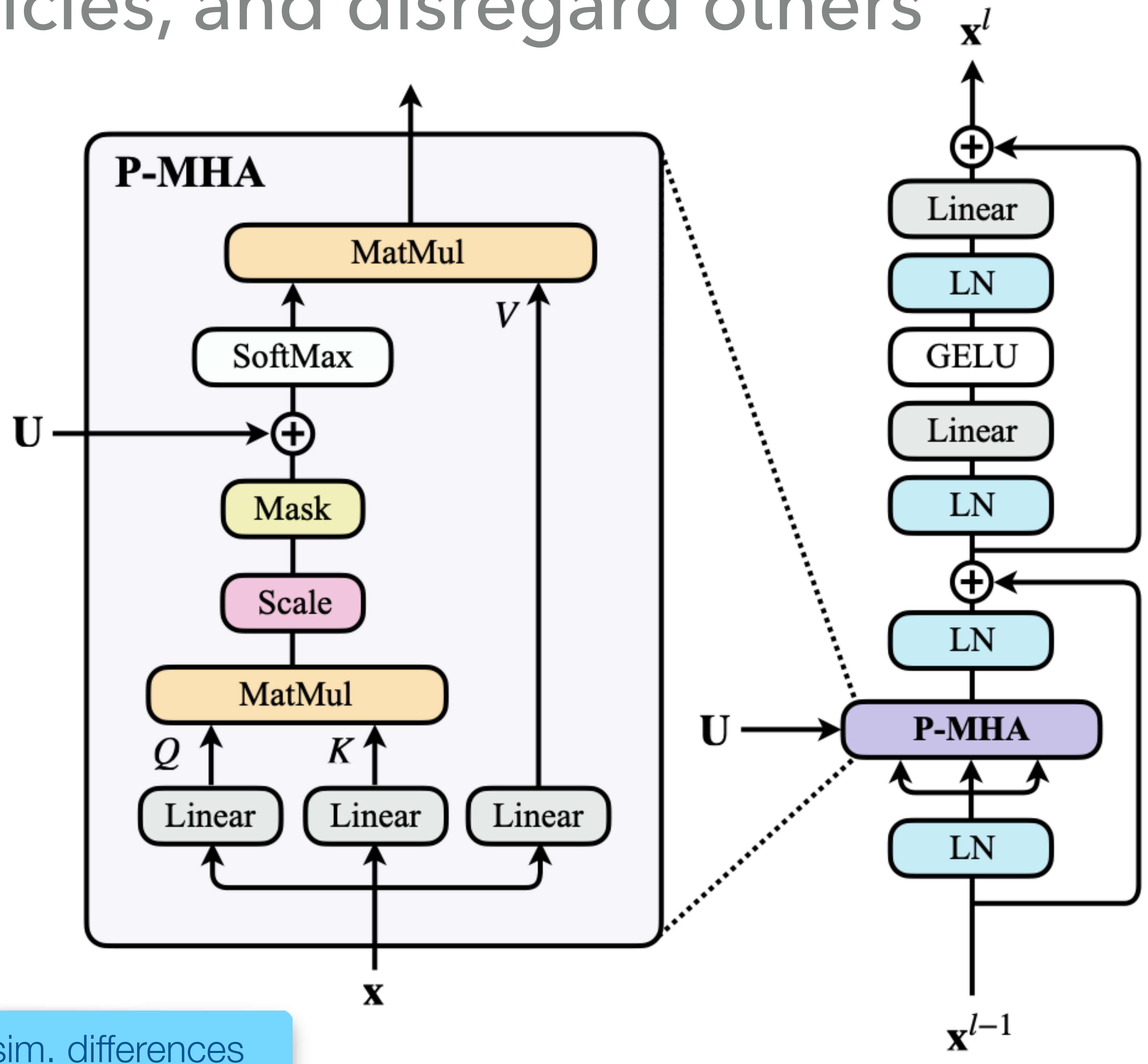
$$\mathrm{softmax}\left(\frac{QK^\intercal}{\sqrt{d_k}} + U\right)V$$

GLOBAL PARTICLE TRANSFORMER

[arXiv:2202.03772](https://arxiv.org/abs/2202.03772)
CMS-DP-2023-046

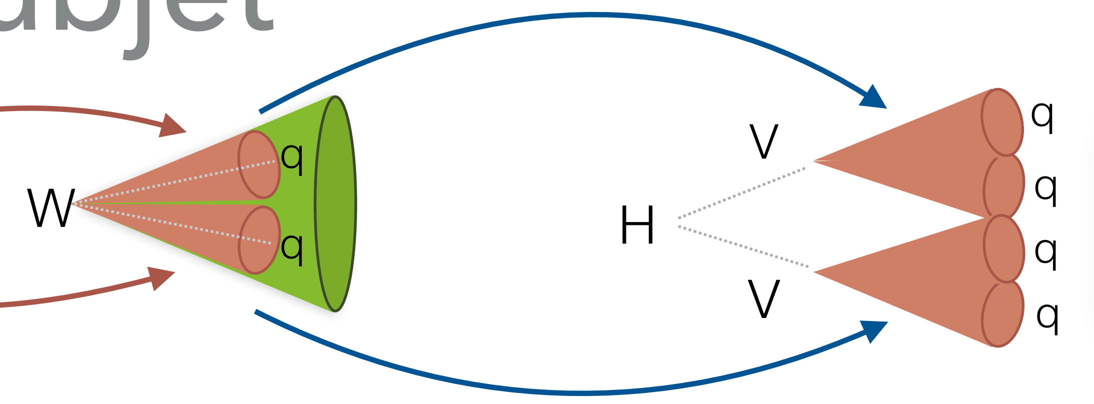
- Global Particle Transformer algorithm uses learned “attention” (and pairwise features) to give more weight to certain particles, and disregard others in order to infer the origin of jets

- Challenge to calibrate these taggers when there is no SM analogue for the signal we can isolate
  - New technique uses Lund jet plane for calibration
  - Effectively **measure scale factors** per quark subjet



$$\mathrm{softmax}\left(\frac{QK^{\intercal}}{\sqrt{d_k}} + U\right)V$$

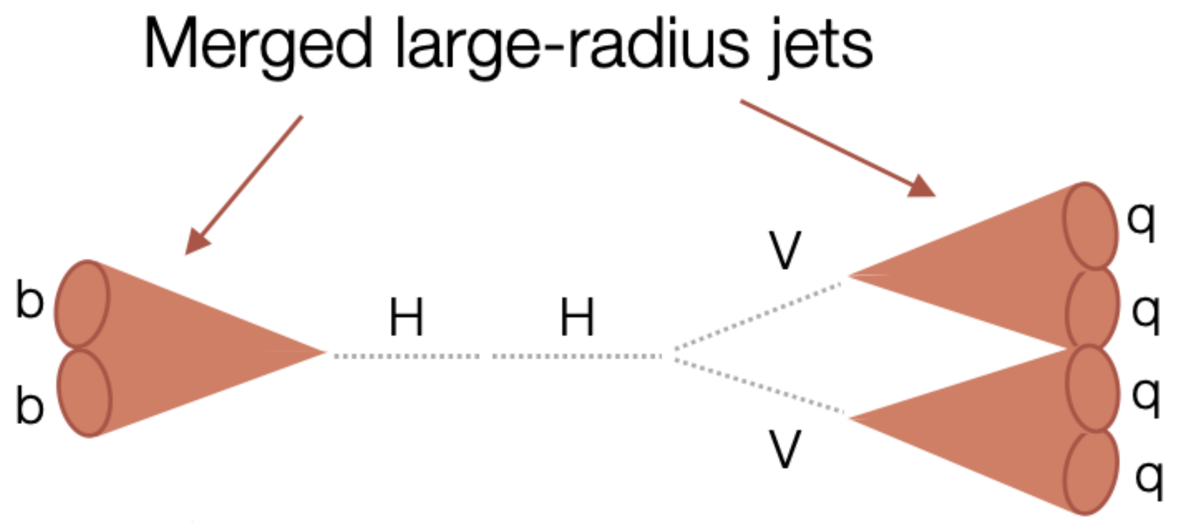
Measure tagger efficiency per-quark

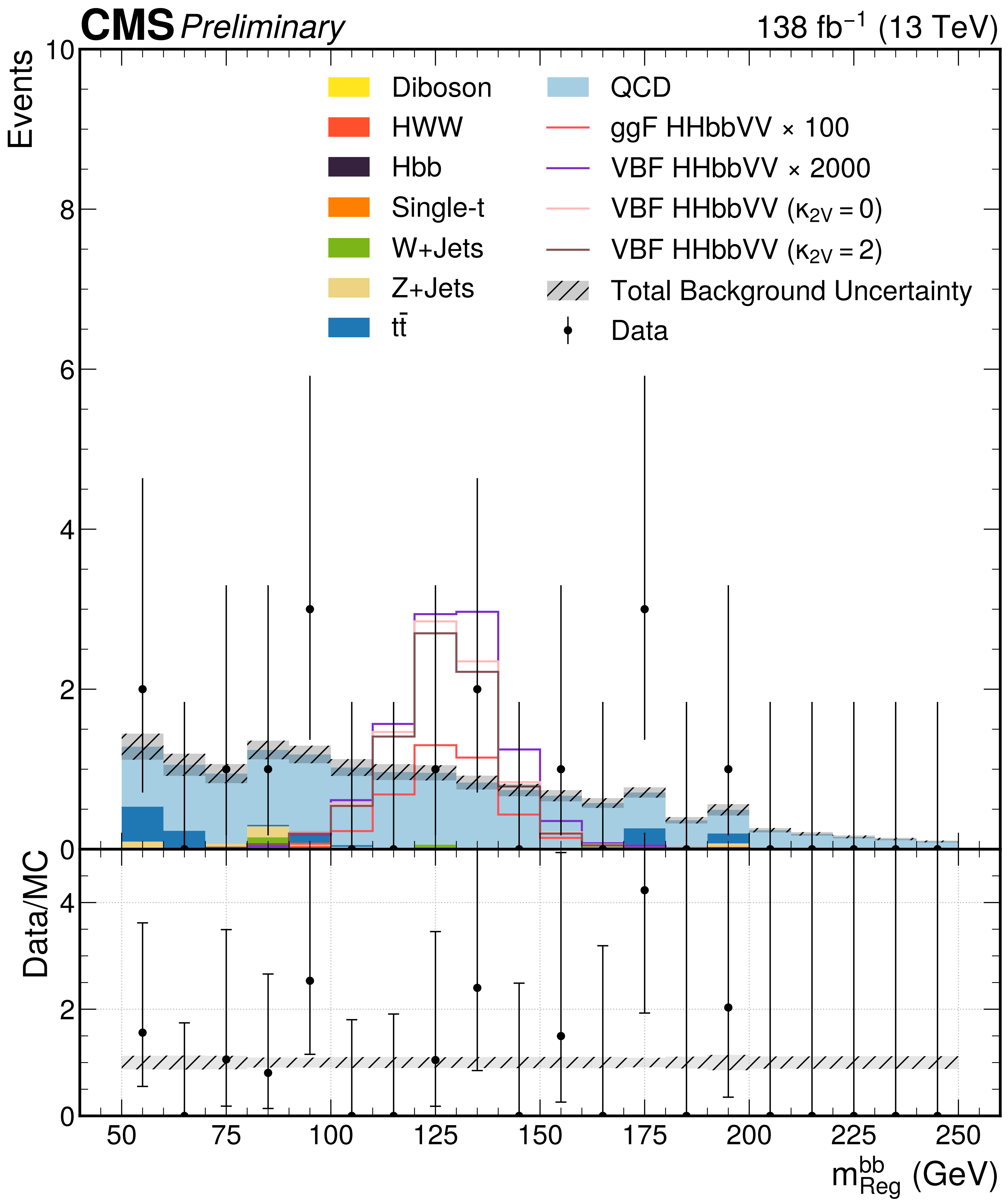


Product of per-quark data / sim. differences
⇒ SF + uncertainty on HVV(4q) tagging

CMS BOOSTED HADRONIC HHBBVV SEARCH

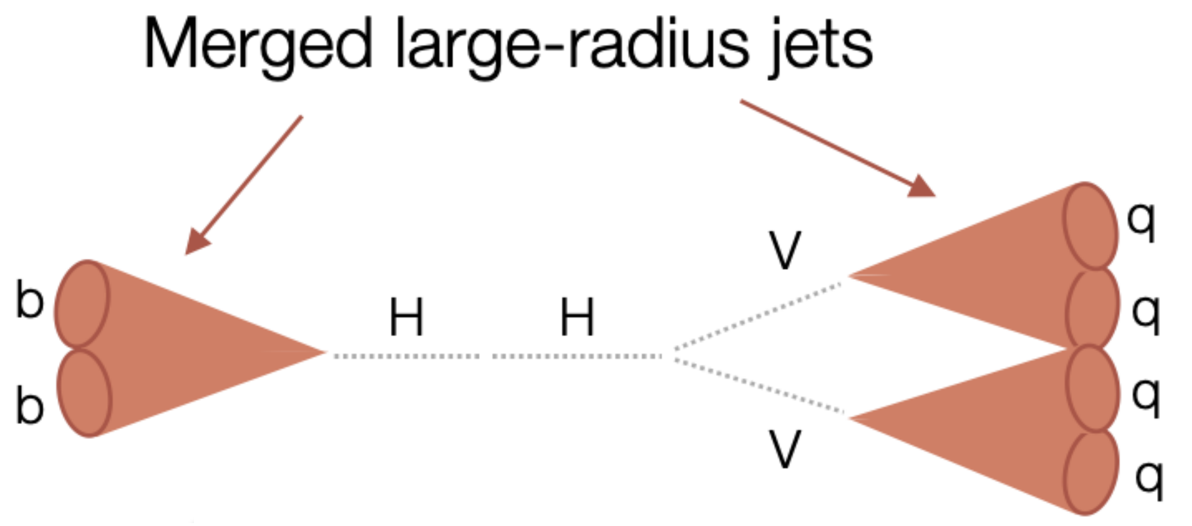
- Enables a new search for boosted $HH \rightarrow bbVV \rightarrow bb4q$

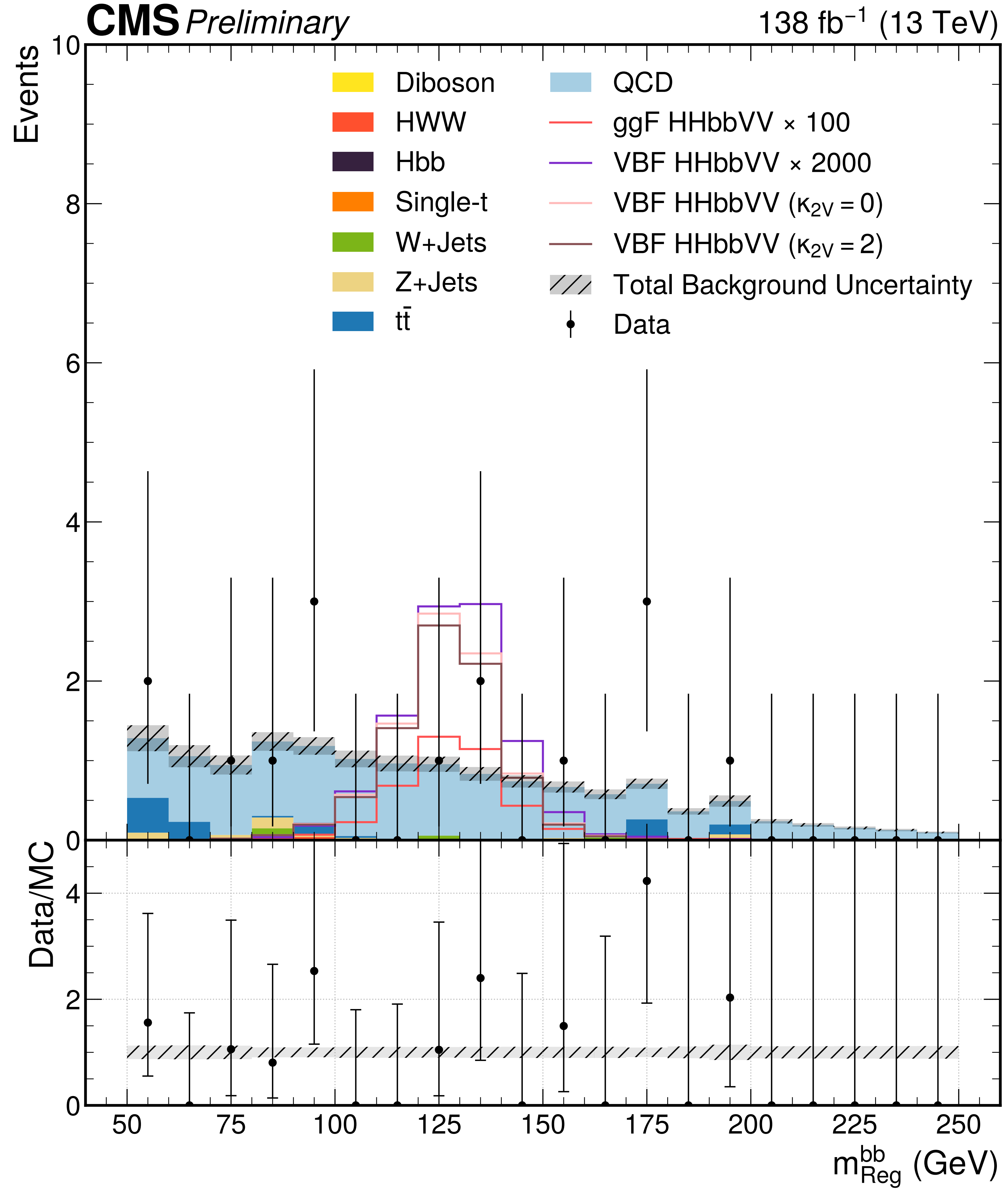


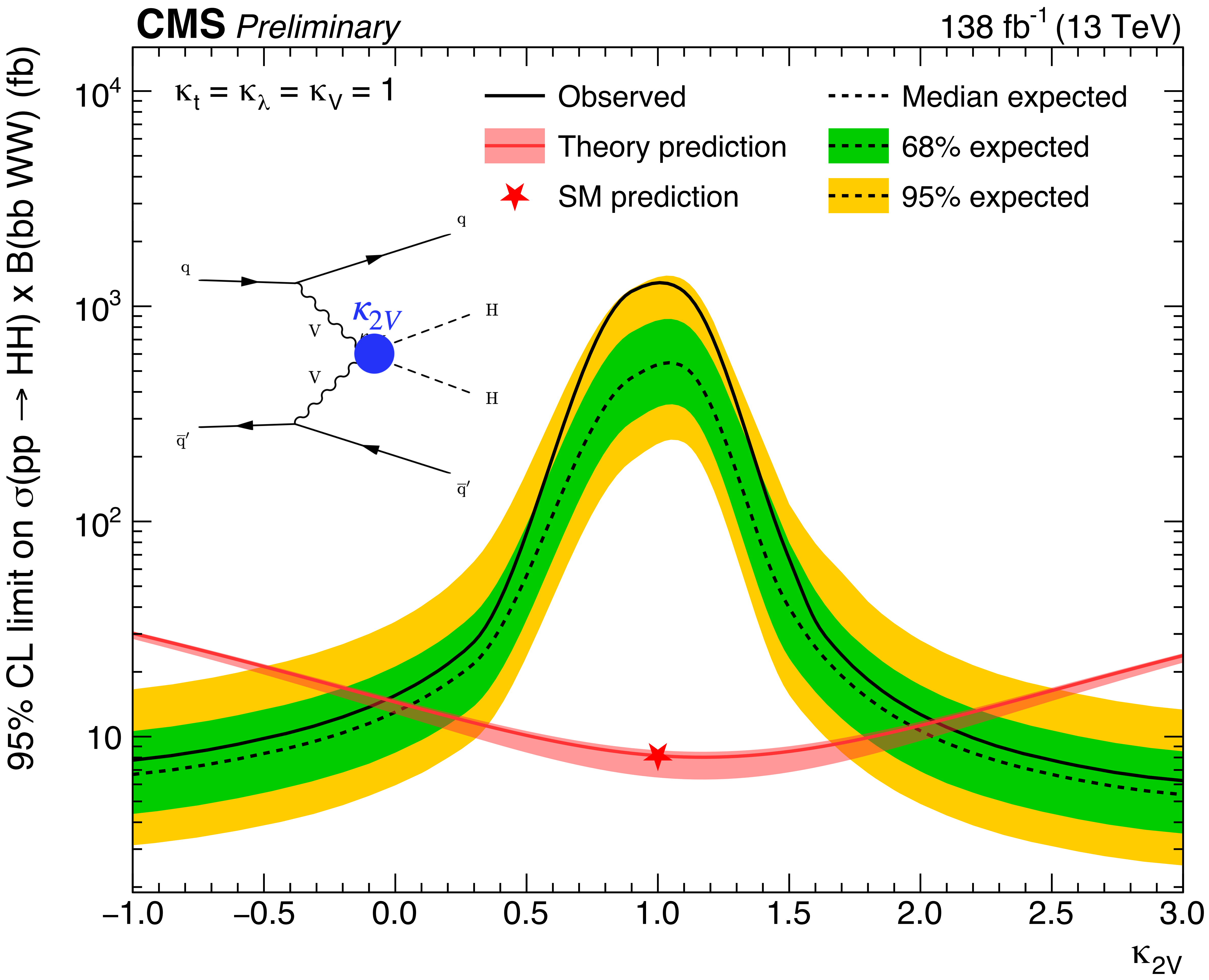


CMS BOOSTED HADRONIC HHBBVV SEARCH

- Enables a new search for boosted $HH \rightarrow bbVV \rightarrow bb4q$
- Provides second-best constraint on $HHVV$ coupling κ_{2V}

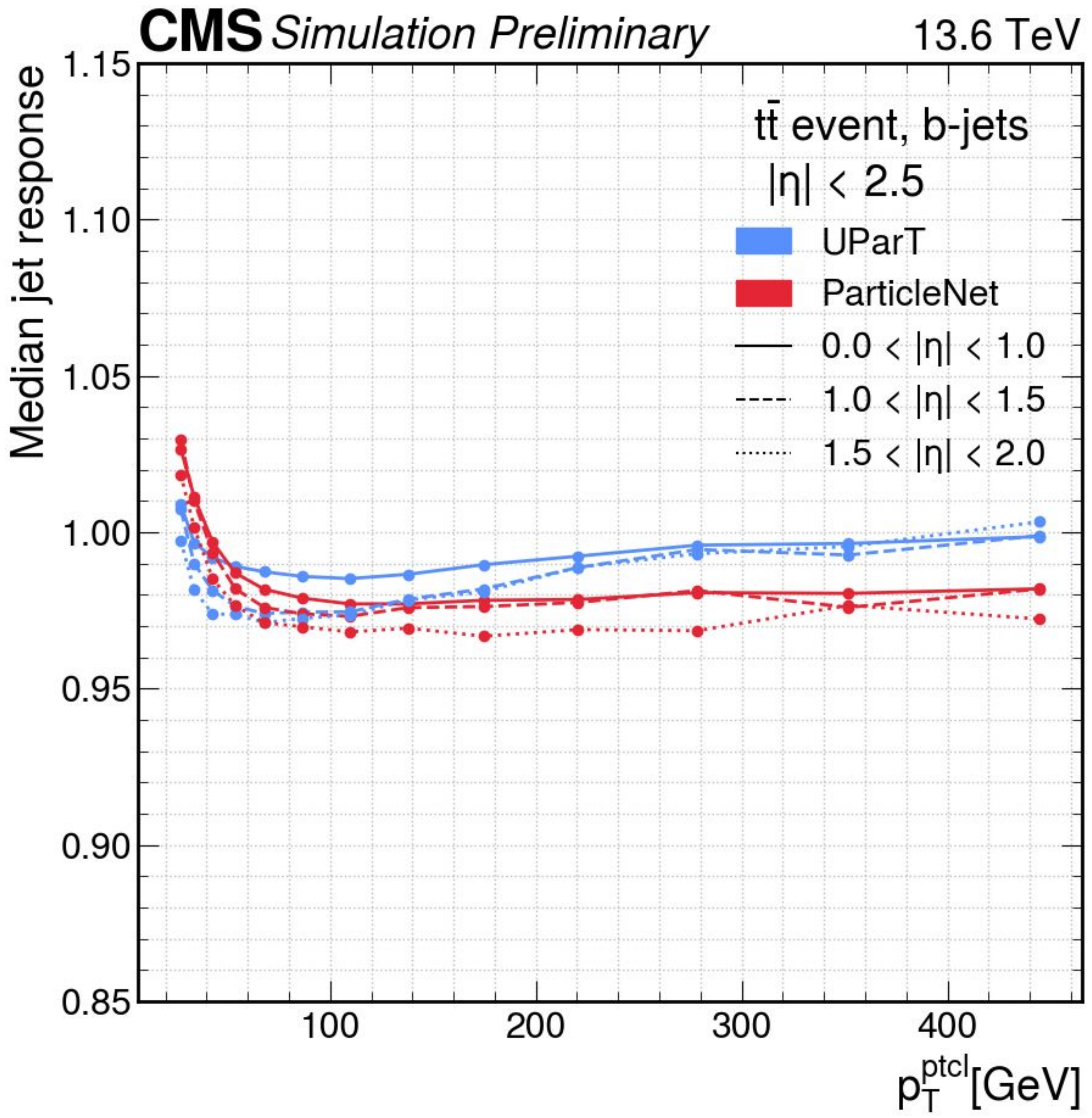






UNIFIED PARTICLE TRANSFORMER SMALL-RADIUS JETS

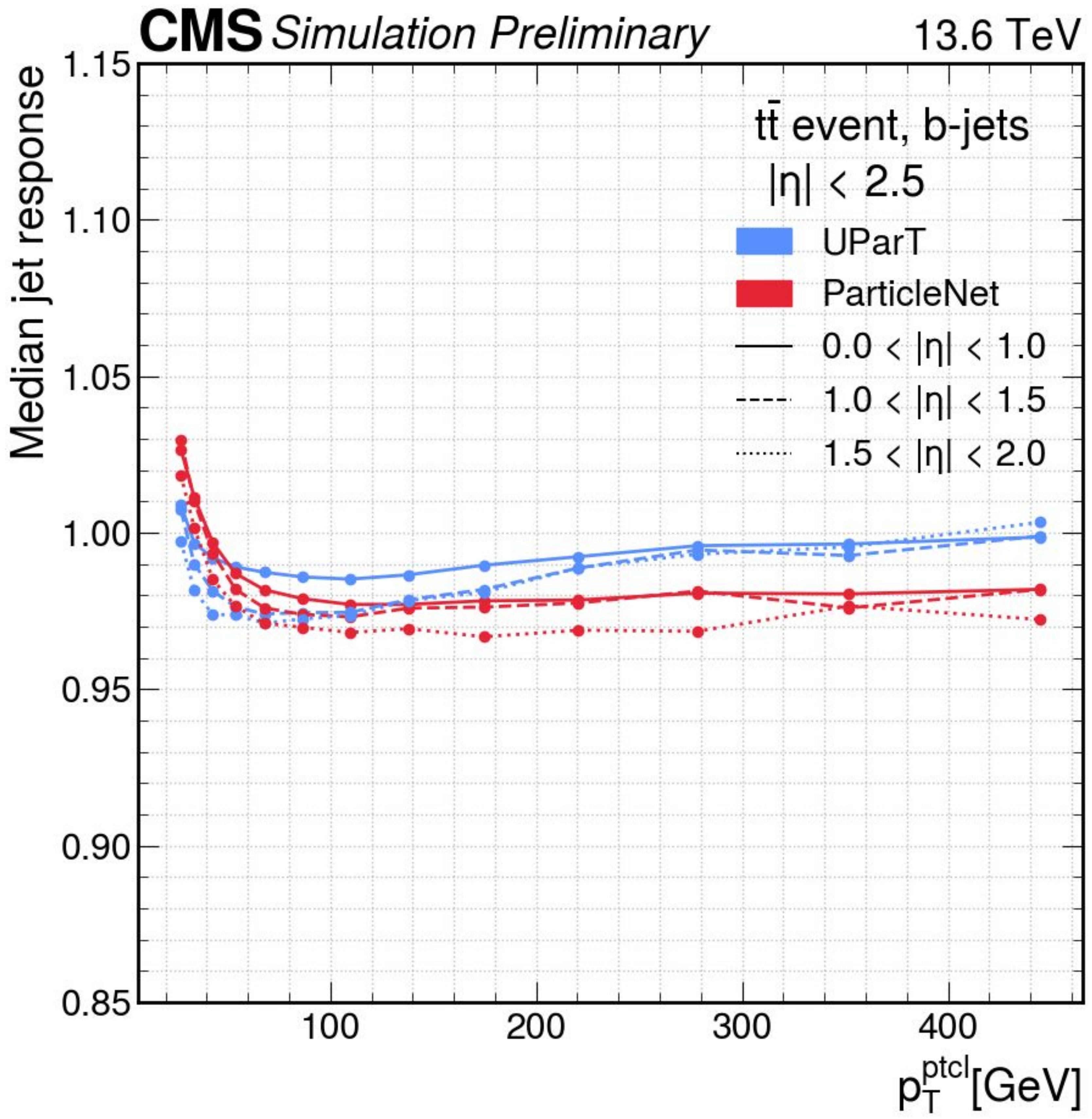
- Unified approach using same architecture to tag heavy flavor, tag hadronic taus, **regress jet energy**, and estimate jet energy resolution

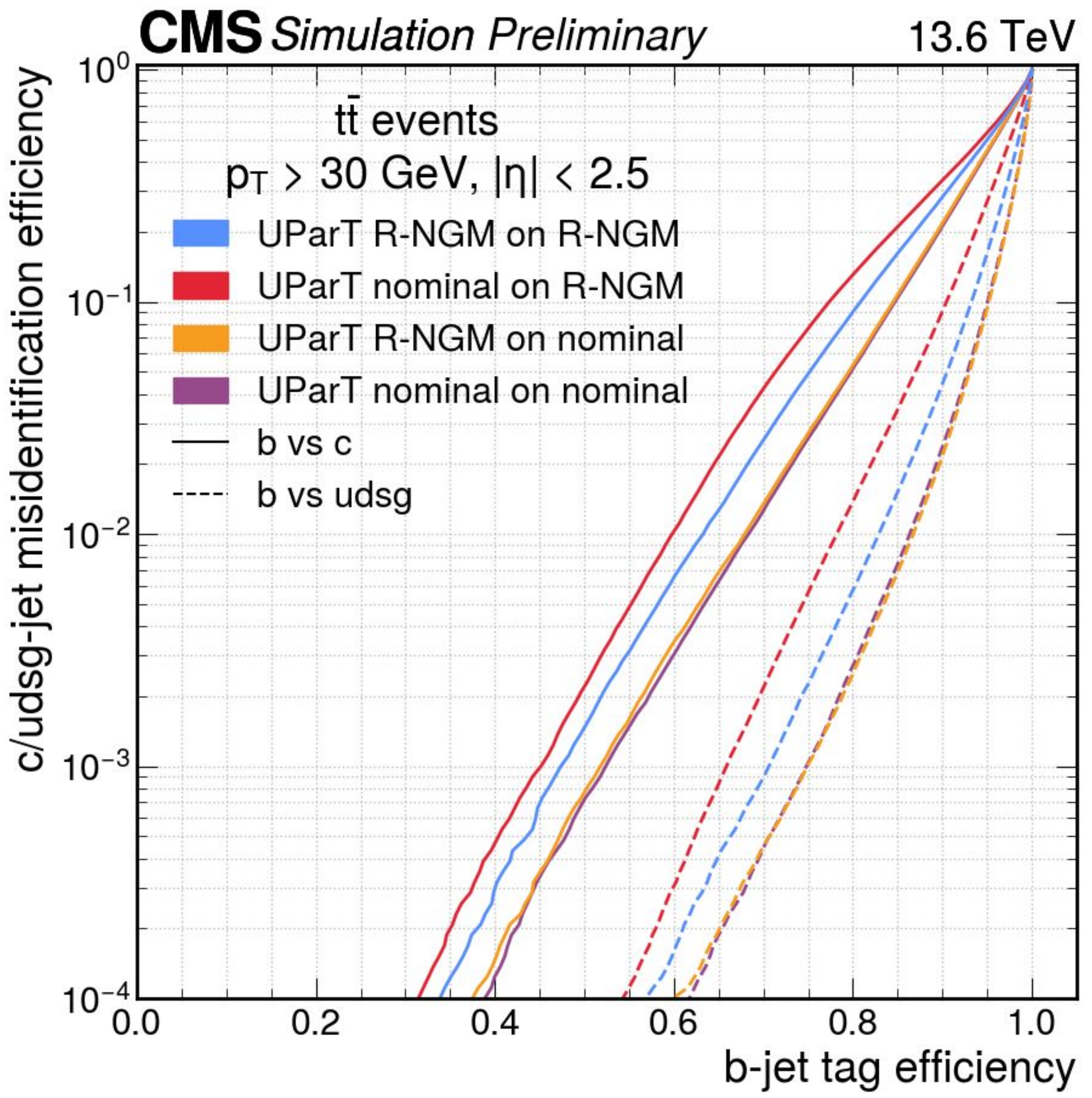


UNIFIED PARTICLE TRANSFORMER SMALL-RADIUS JETS

CMS-DP-2024-066
CMS-DP-2024-064

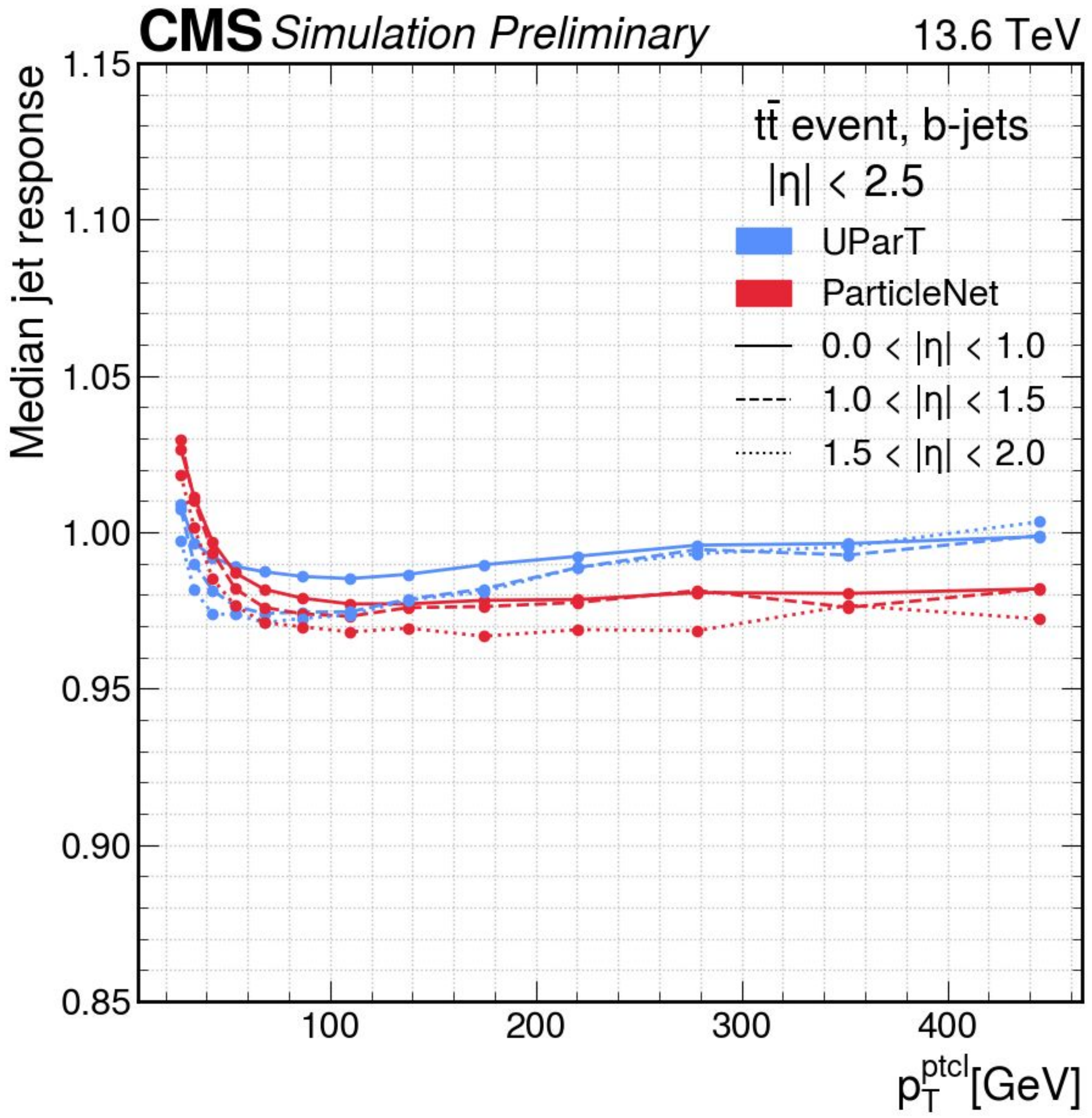
- Unified approach using same architecture to tag heavy flavor, tag hadronic taus, **regress jet energy**, and estimate jet energy resolution
- Rectified normed gradient method (R-NGM) adversarial training used improve model robustness even when facing perturbed or mismodeled inputs



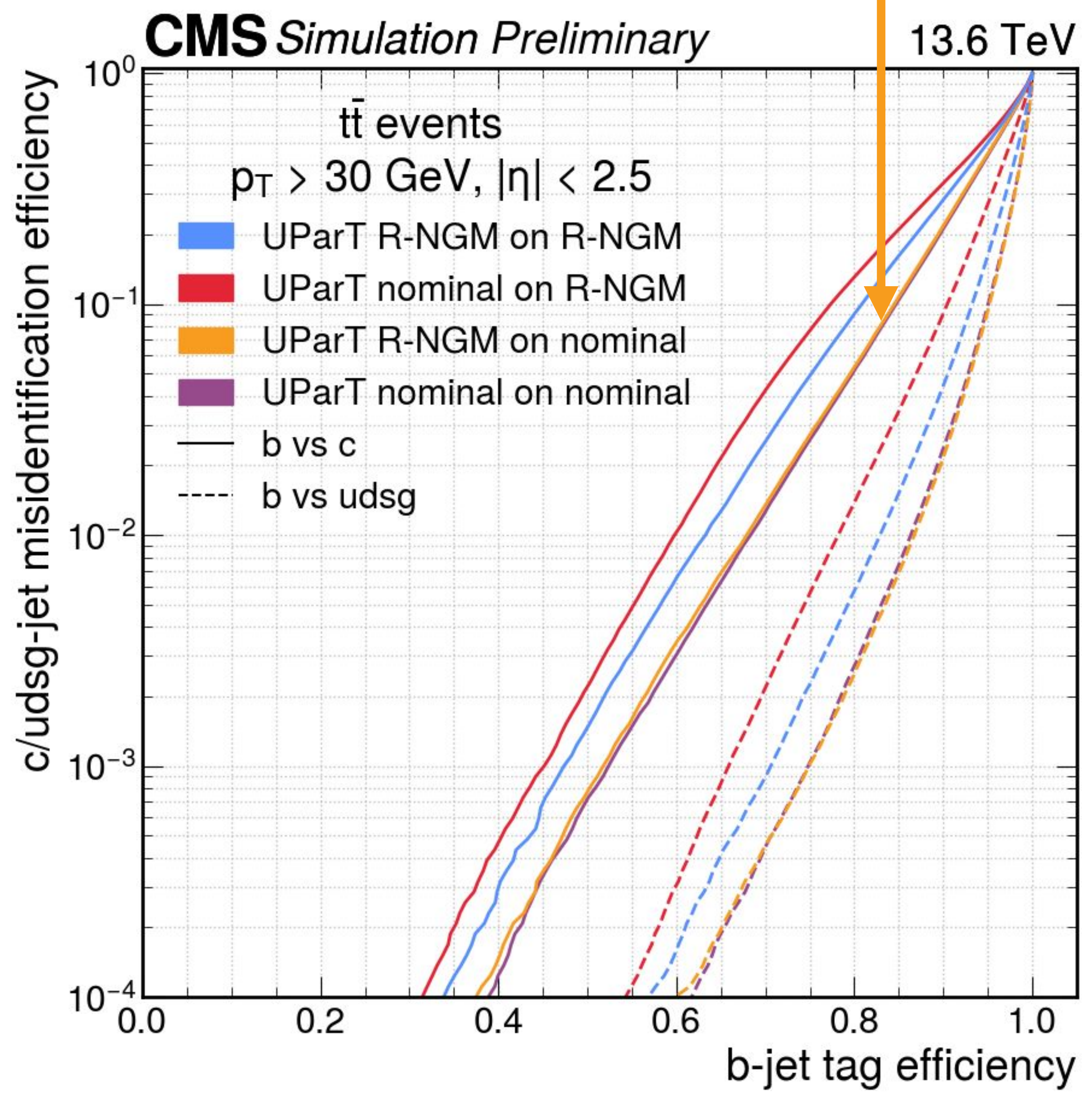


UNIFIED PARTICLE TRANSFORMER SMALL-RADIUS JETS

- Unified approach using same architecture to tag heavy flavor, tag hadronic taus, **regress jet energy**, and estimate jet energy resolution
- Rectified normed gradient method (R-NGM) adversarial training used improve model robustness even when facing perturbed or mismodeled inputs

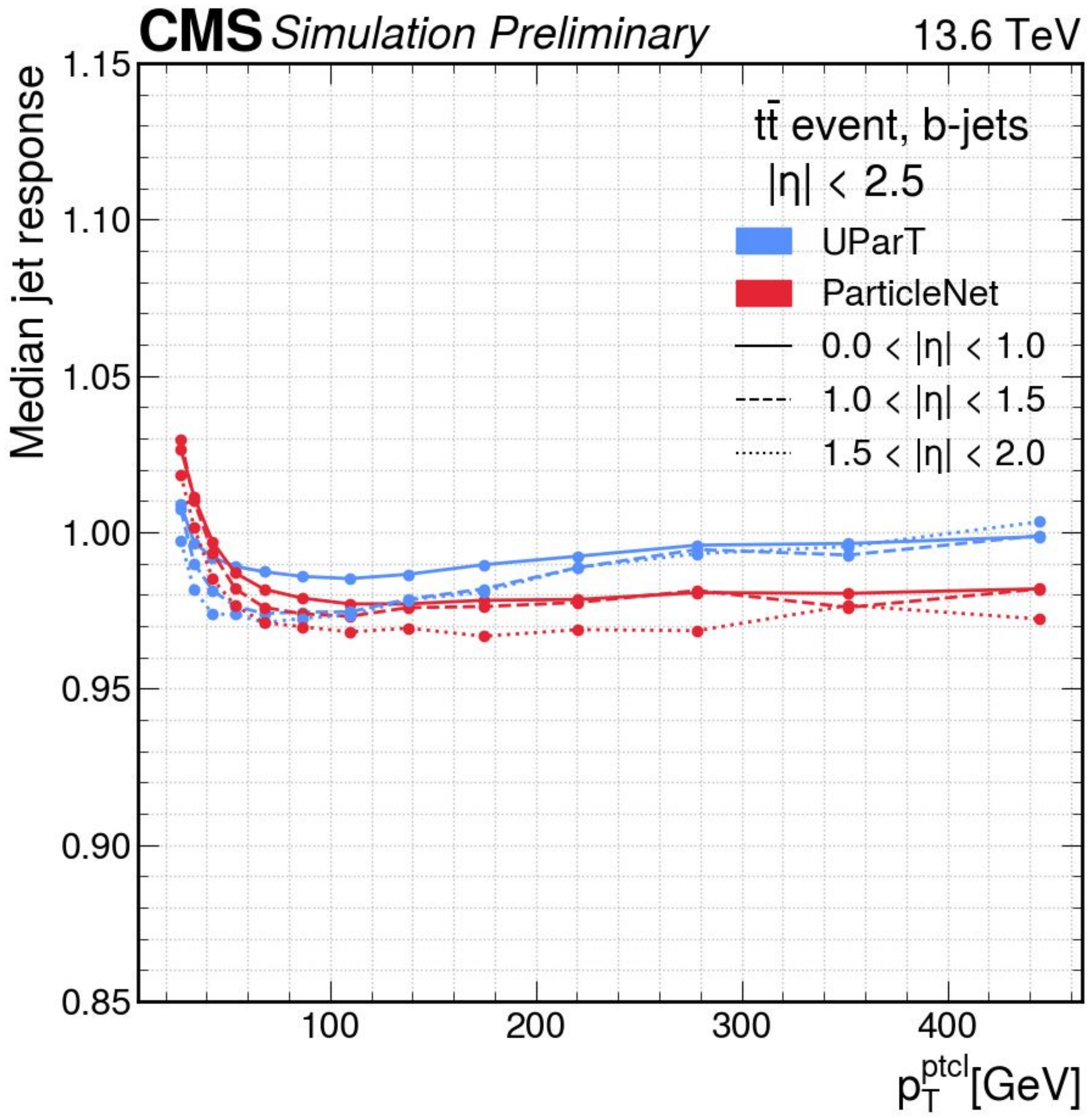


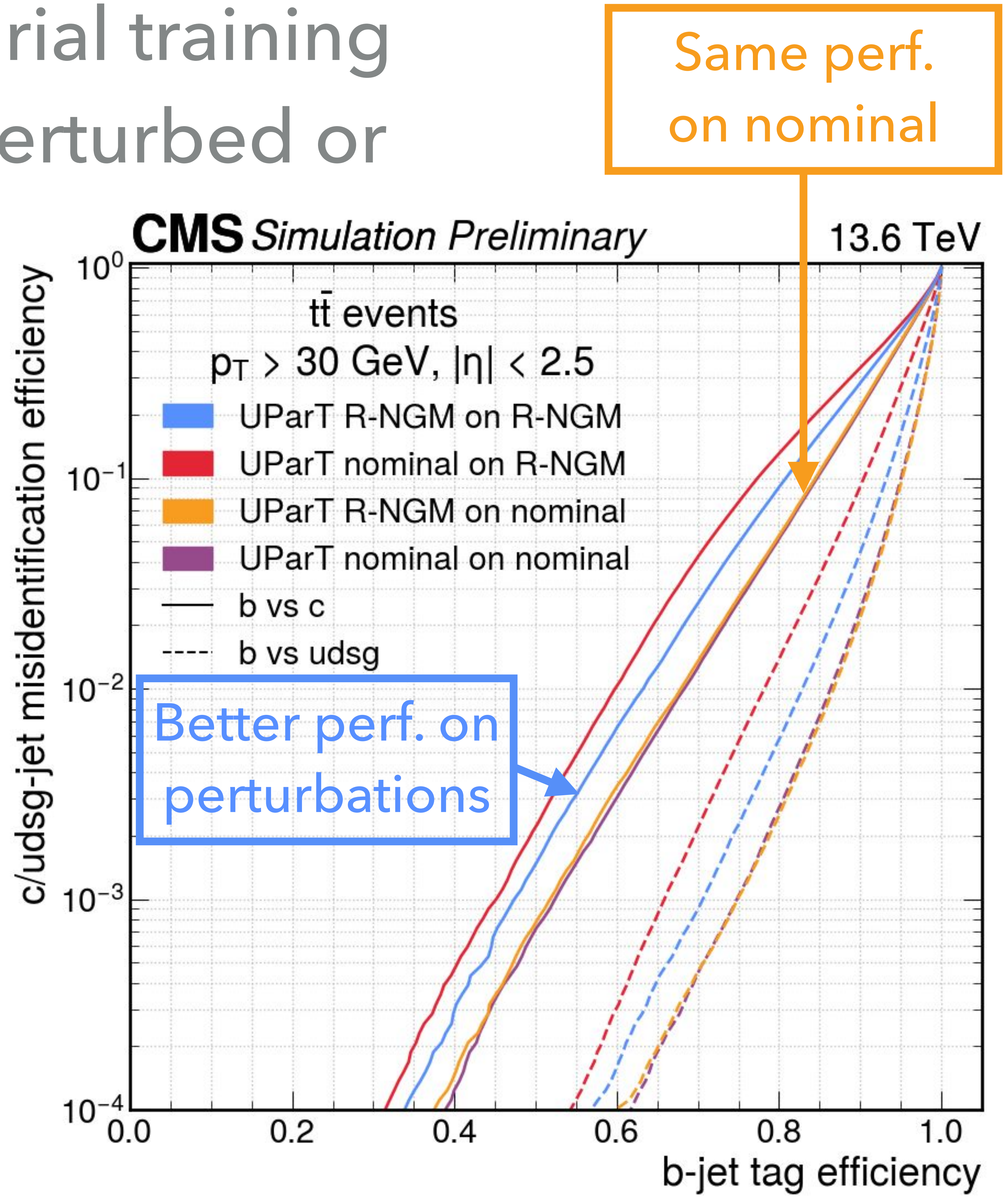
Same perf. on nominal



UNIFIED PARTICLE TRANSFORMER SMALL-RADIUS JETS

- Unified approach using same architecture to tag heavy flavor, tag hadronic taus, **regress jet energy**, and estimate jet energy resolution
- Rectified normed gradient method (R-NGM) adversarial training used improve model robustness even when facing perturbed or mismodeled inputs

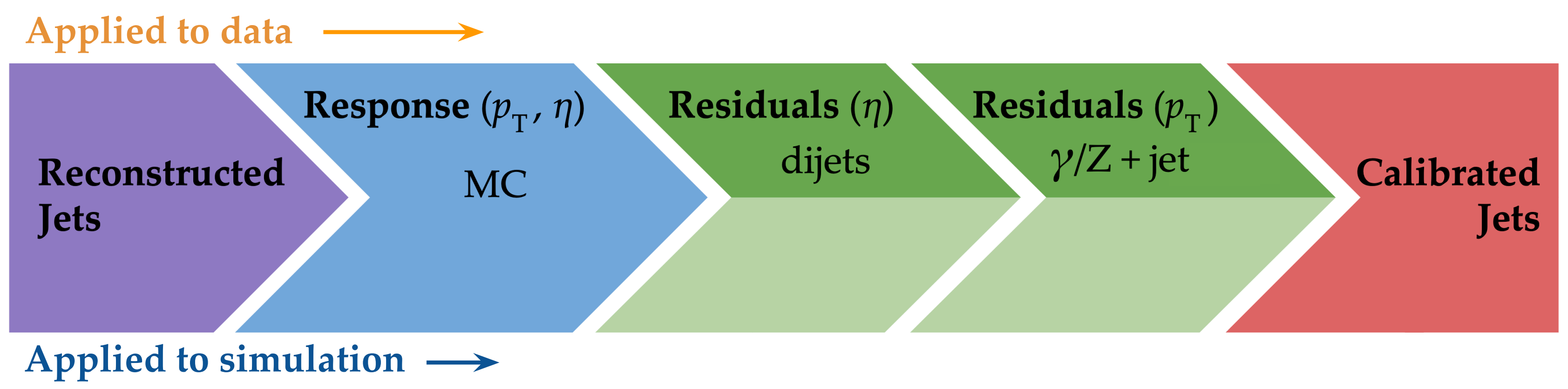


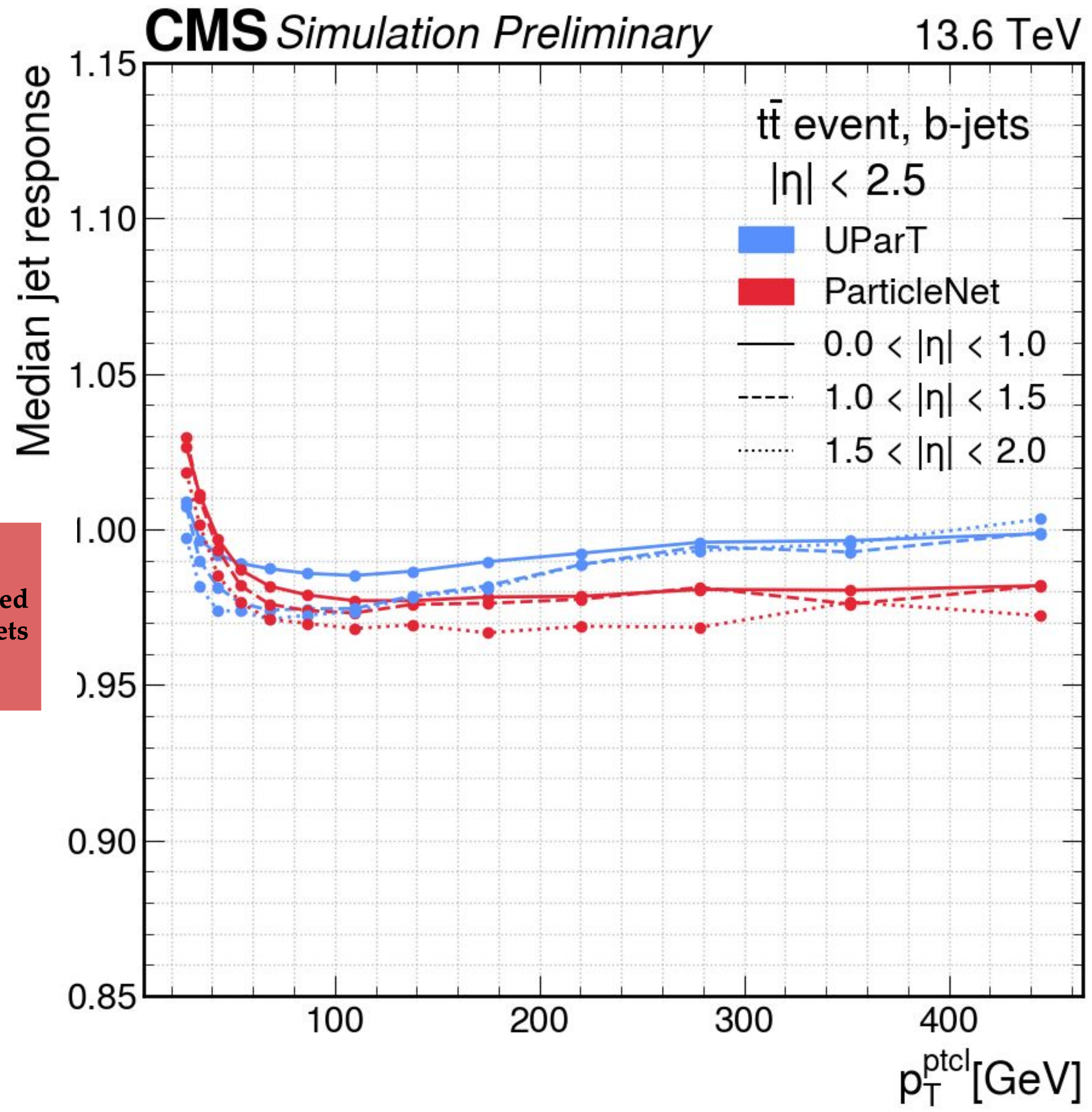


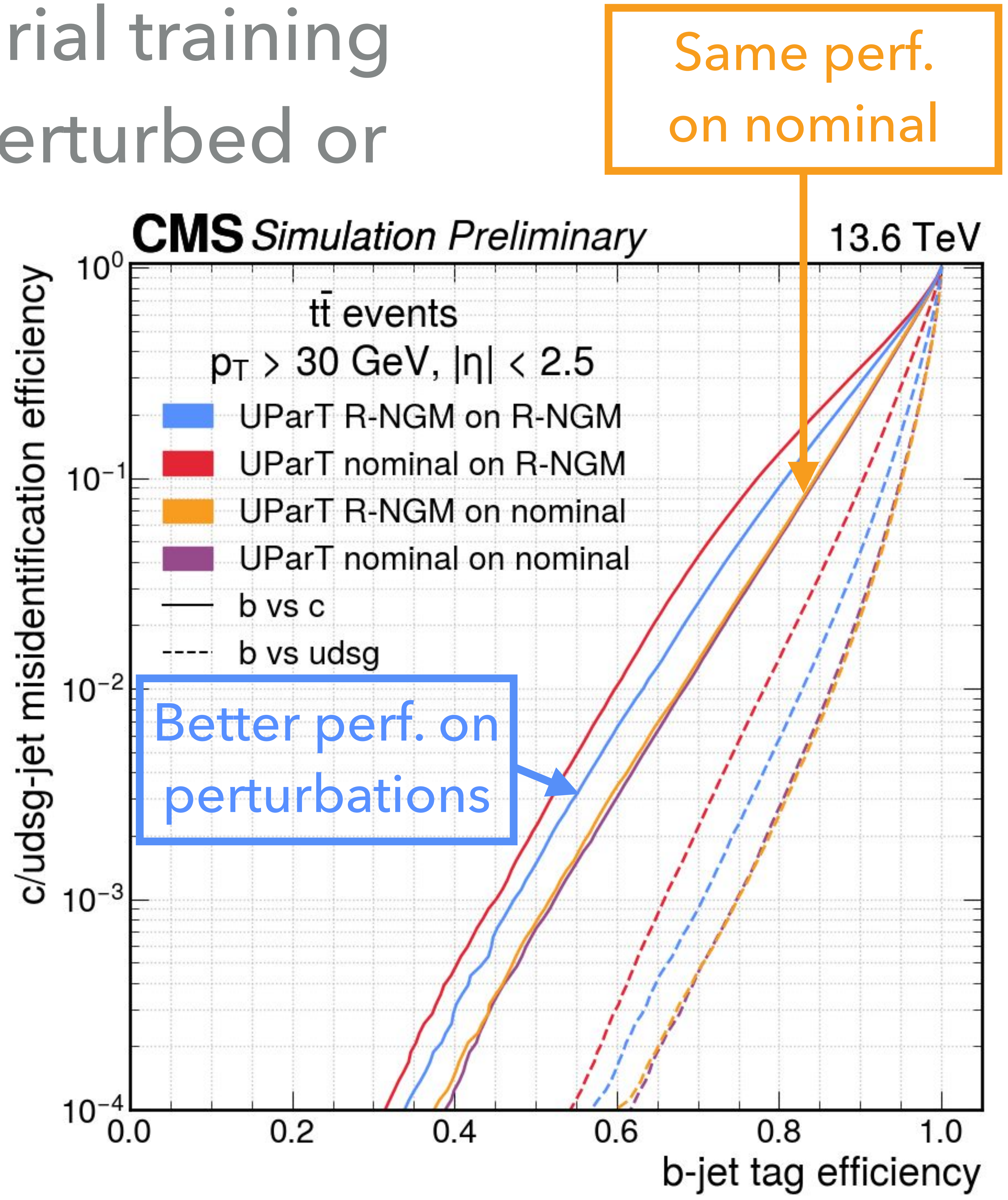
UNIFIED PARTICLE TRANSFORMER SMALL-RADIUS JETS

CMS-DP-2024-066
CMS-DP-2024-064

- Unified approach using same architecture to tag heavy flavor, tag hadronic taus, **regress jet energy**, and estimate jet energy resolution
- Rectified normed gradient method (R-NGM) adversarial training used improve model robustness even when facing perturbed or mismodeled inputs
- Calibration in progress for 2022+2023 data



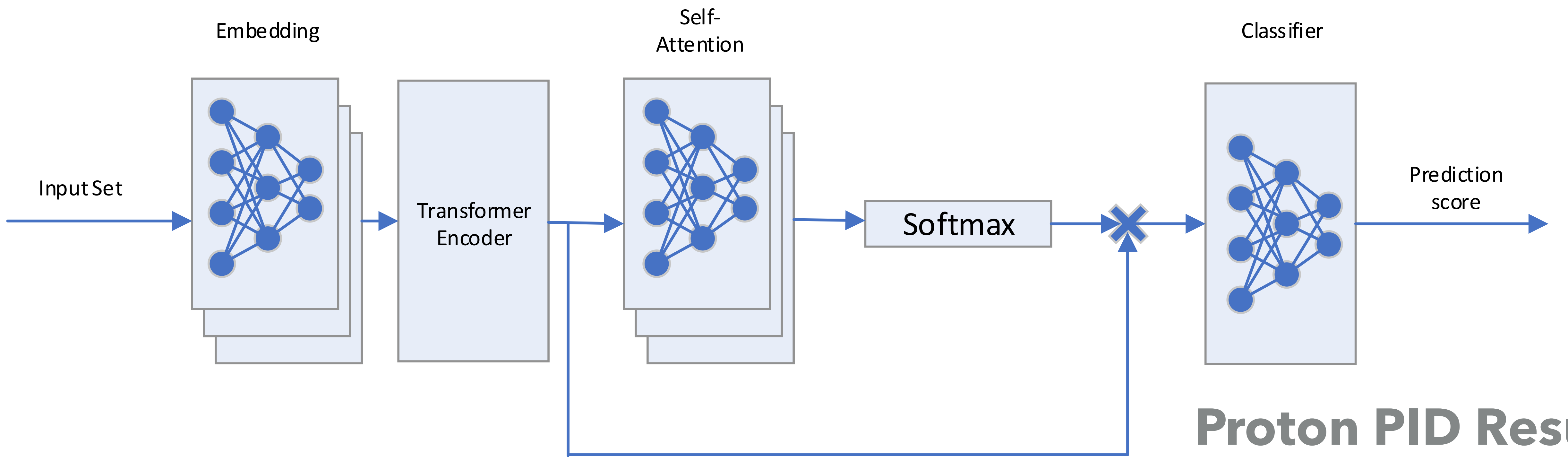


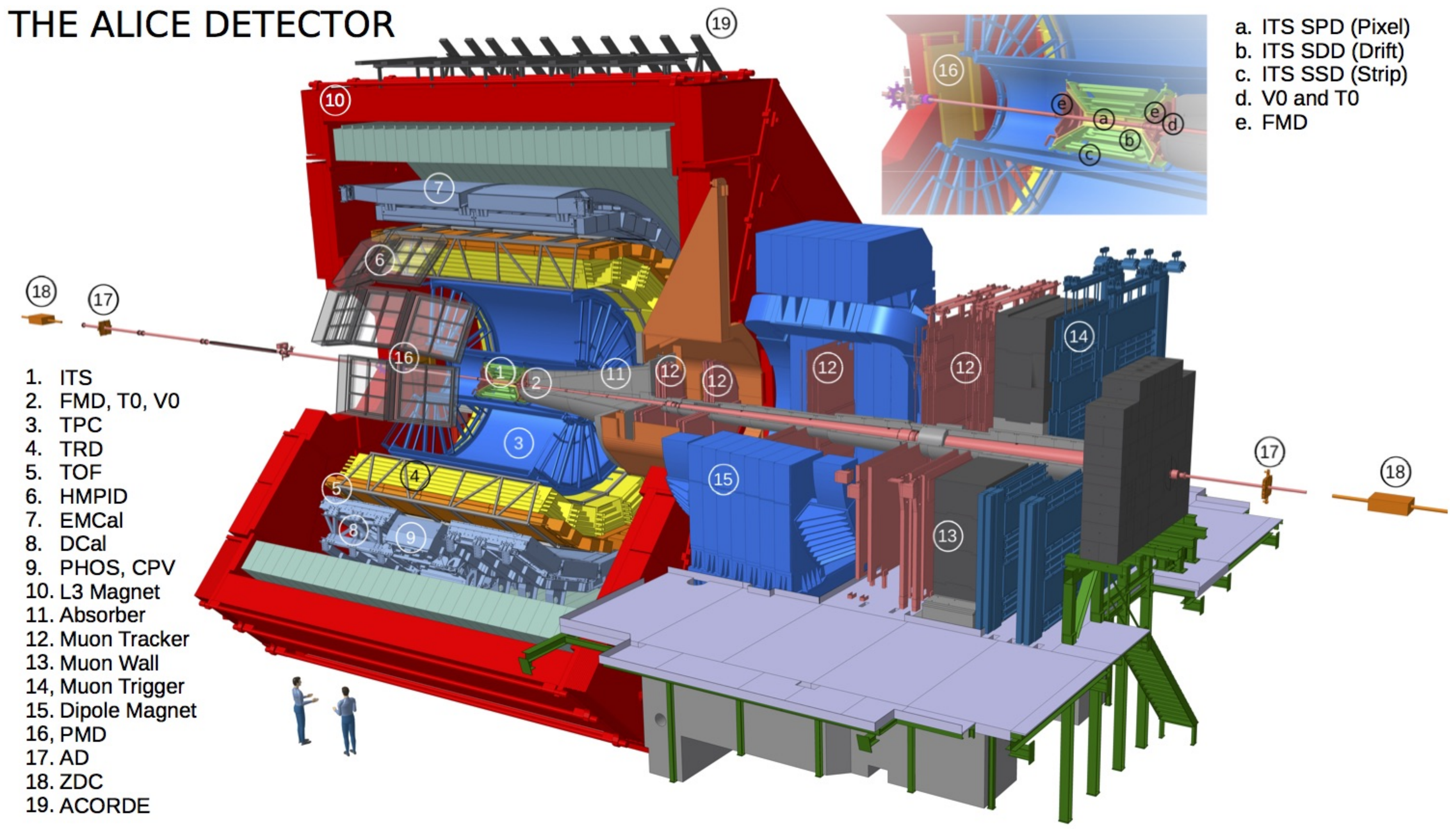


PARTICLE ID WITH TRANSFORMER

[arXiv:2403.17436](https://arxiv.org/abs/2403.17436)
[arXiv:2401.01905](https://arxiv.org/abs/2401.01905)

- **Transformer** for particle ID in ALICE can result in higher purity and efficiency than standard methods





Proton PID Results

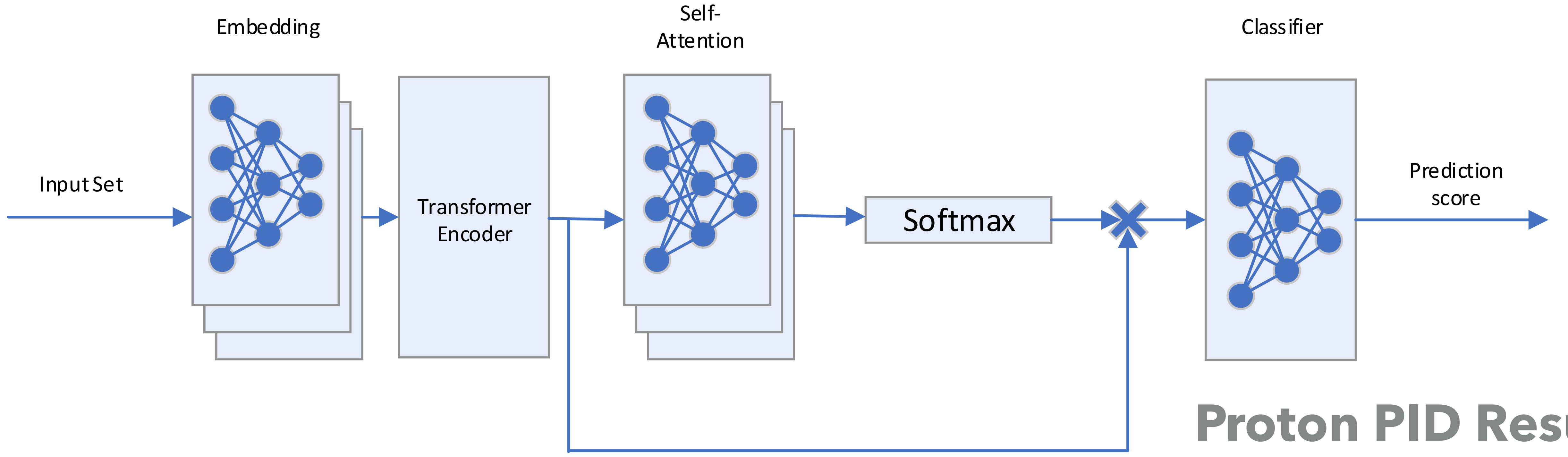
Model	Precision	Recall	F_1
Standard	**99.40 ± 0.01**	59.72 ± 0.03	74.61 ± 1.88
Ensemble	97.16 ± 0.46	93.74 ± 0.30	95.42 ± 0.12
Mean	97.85 ± 0.41	93.34 ± 0.32	95.54 ± 0.06
Proposed	97.80 ± 0.44	**93.86 ± 0.27**	**95.79 ± 0.07**
Regression	97.38 ± 0.40	93.67 ± 0.38	95.49 ± 0.15

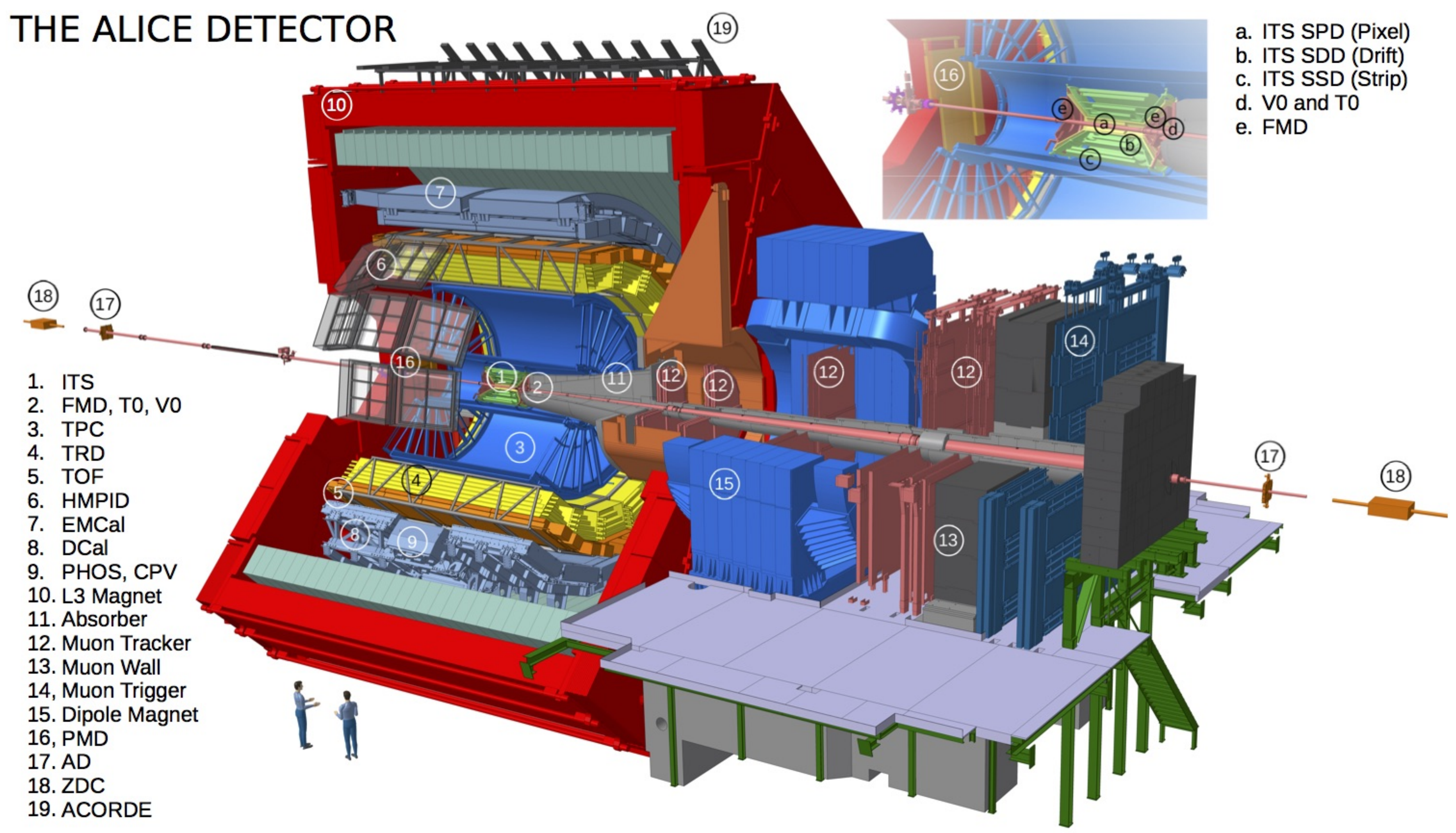
Transformer

PARTICLE ID WITH TRANSFORMER

arXiv:2403.17436
arXiv:2401.01905

- **Transformer** for particle ID in ALICE can result in higher purity and efficiency than standard methods
- Use **domain adversarial neural networks** to mitigate data-simulation differences

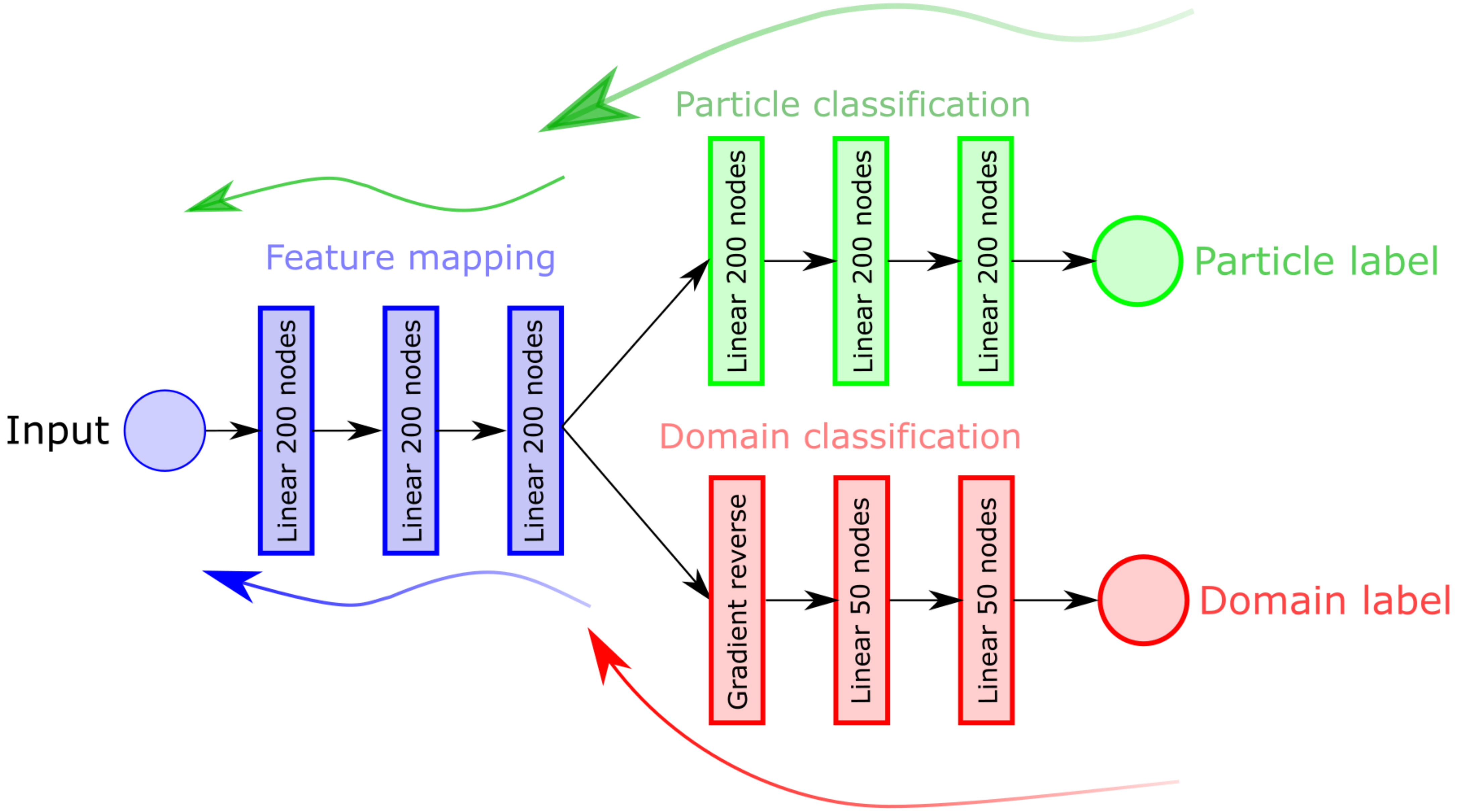




Proton PID Results

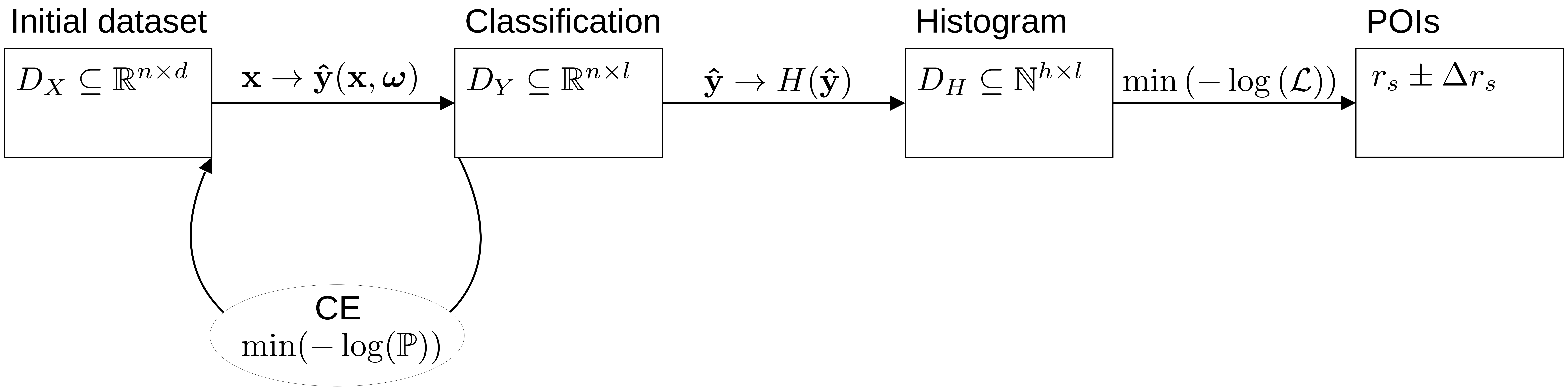
Model	Precision	Recall	F_1
Standard	**99.40 ± 0.01**	59.72 ± 0.03	74.61 ± 1.88
Ensemble	97.16 ± 0.46	93.74 ± 0.30	95.42 ± 0.12
Mean	97.85 ± 0.41	93.34 ± 0.32	95.54 ± 0.06
Proposed	97.80 ± 0.44	**93.86 ± 0.27**	**95.79 ± 0.07**
Regression	97.38 ± 0.40	93.67 ± 0.38	95.49 ± 0.15

Transformer



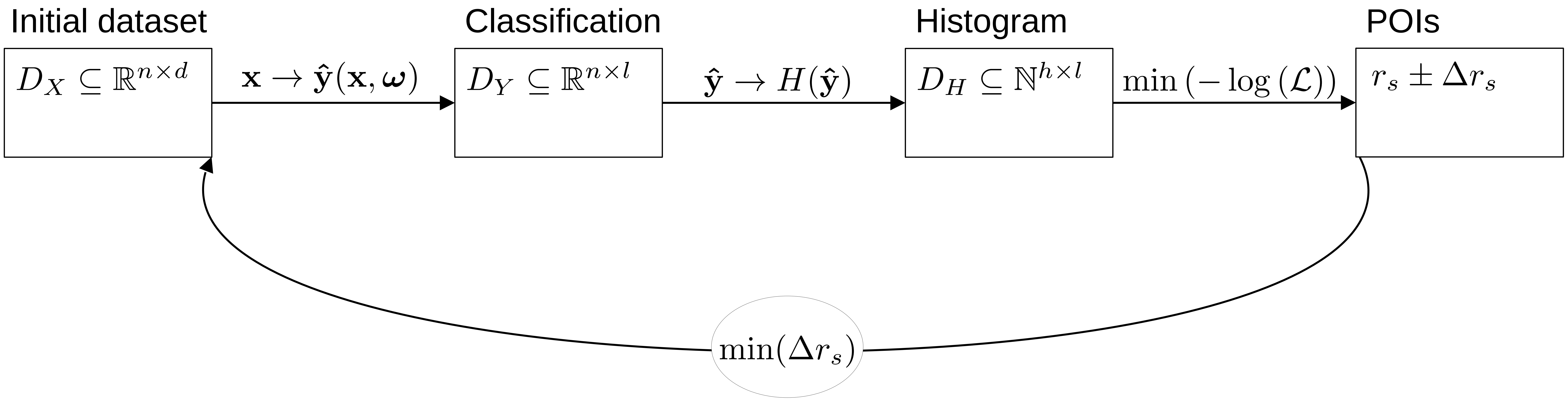
SYSTEMATIC-AWARE LEARNING

- Standard classifier training (CENNT) optimizes for signal vs. background discrimination without considering systematics and other effects that affect the ultimate figure: uncertainty Δr_s on a physics parameter r_s



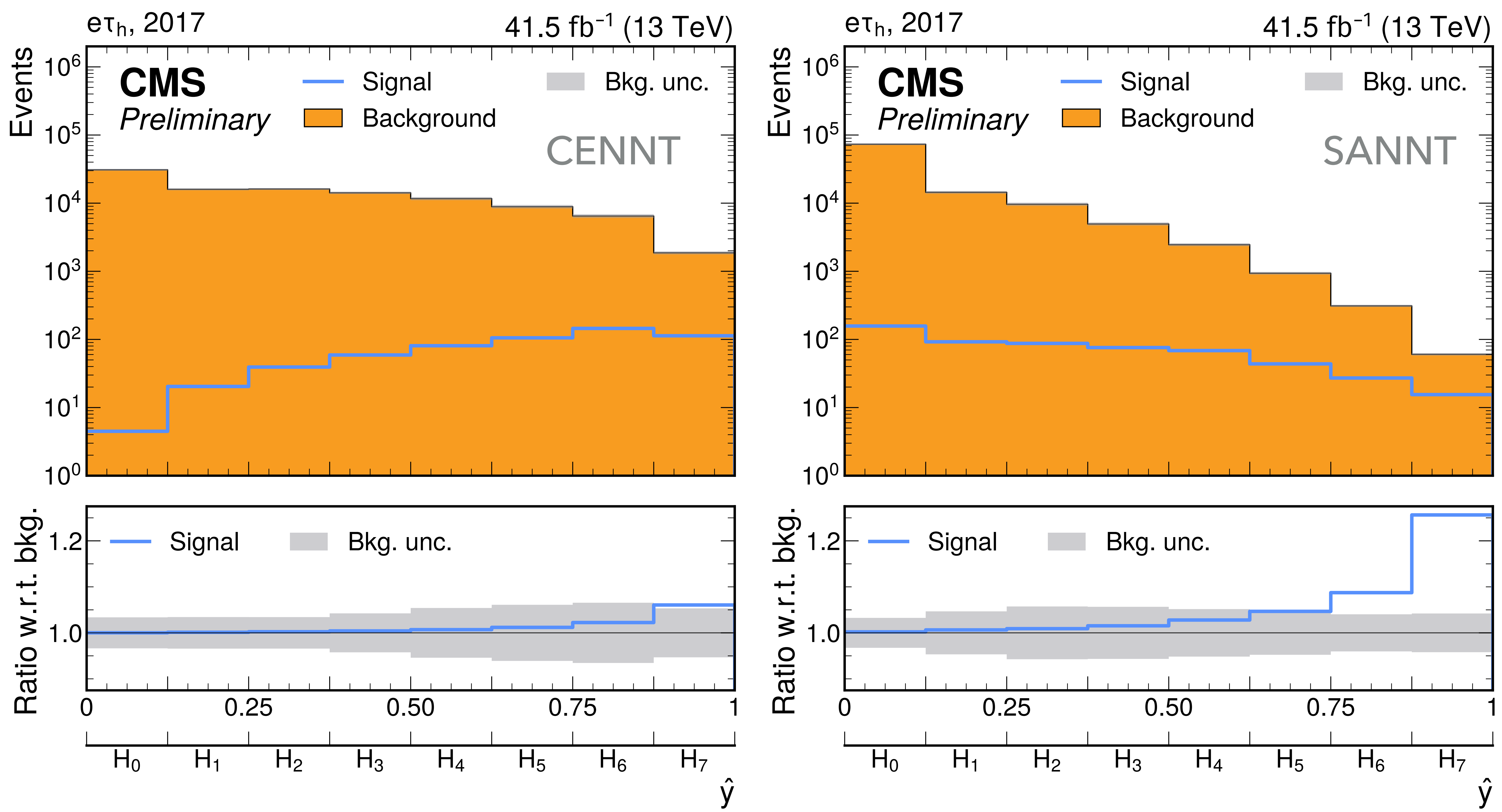
SYSTEMATIC-AWARE LEARNING

- Standard classifier training (CENNT) optimizes for signal vs. background discrimination without considering systematics and other effects that affect the ultimate figure: uncertainty Δr_s on a physics parameter r_s
- By implementing the analysis chain (including systematics) in a **differentiable way**, we can directly optimize for min. Δr_s in the neural network training!
  - Key choosing gradients for histogram operation $\hat{\mathbf{y}} \rightarrow H(\hat{\mathbf{y}})$



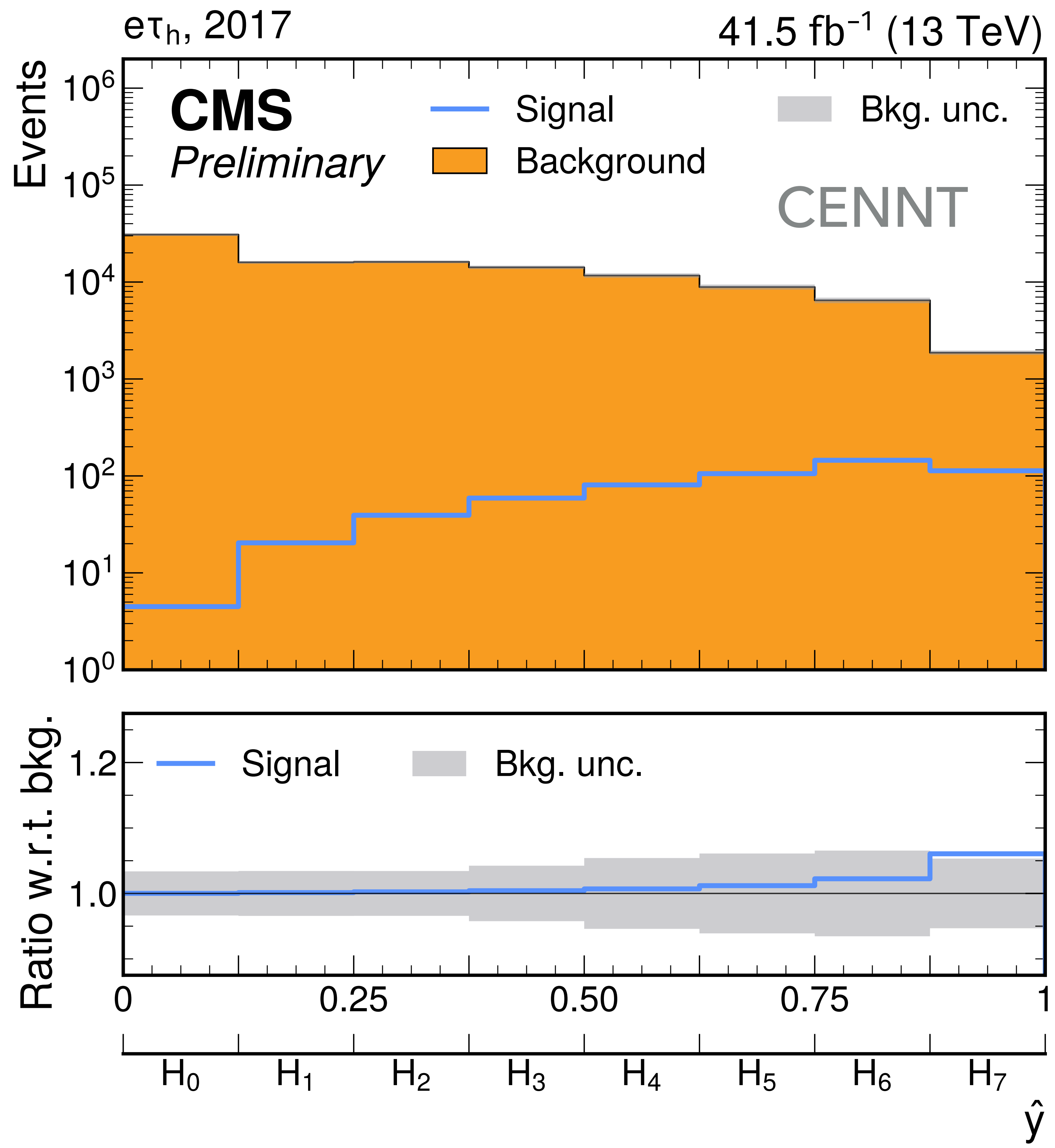
SYSTEMATIC-AWARE LEARNING

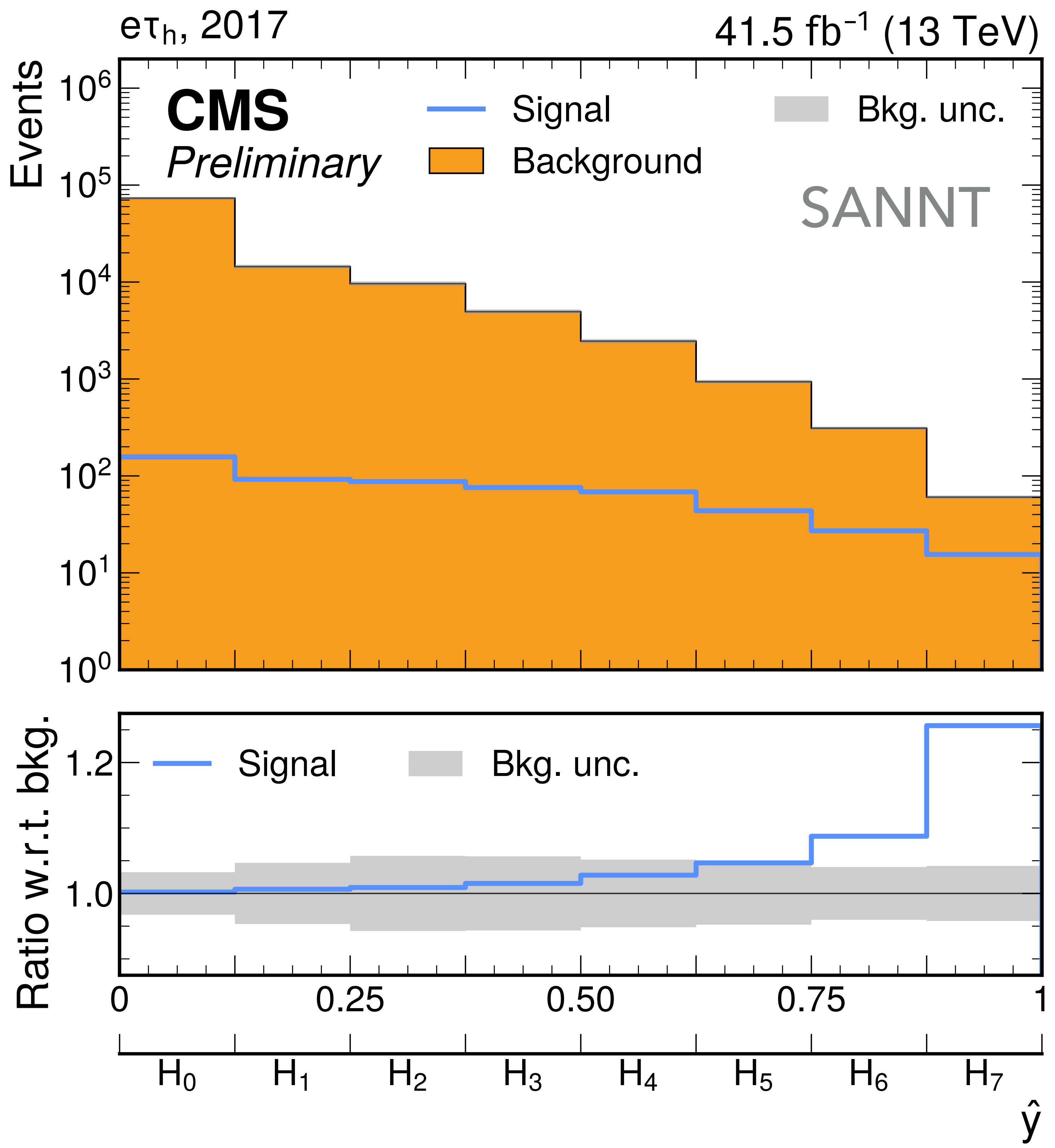
- CENNT optimizes for separation, while SANNT (Δr_s) concentrates signal in bins with smaller background uncertainty

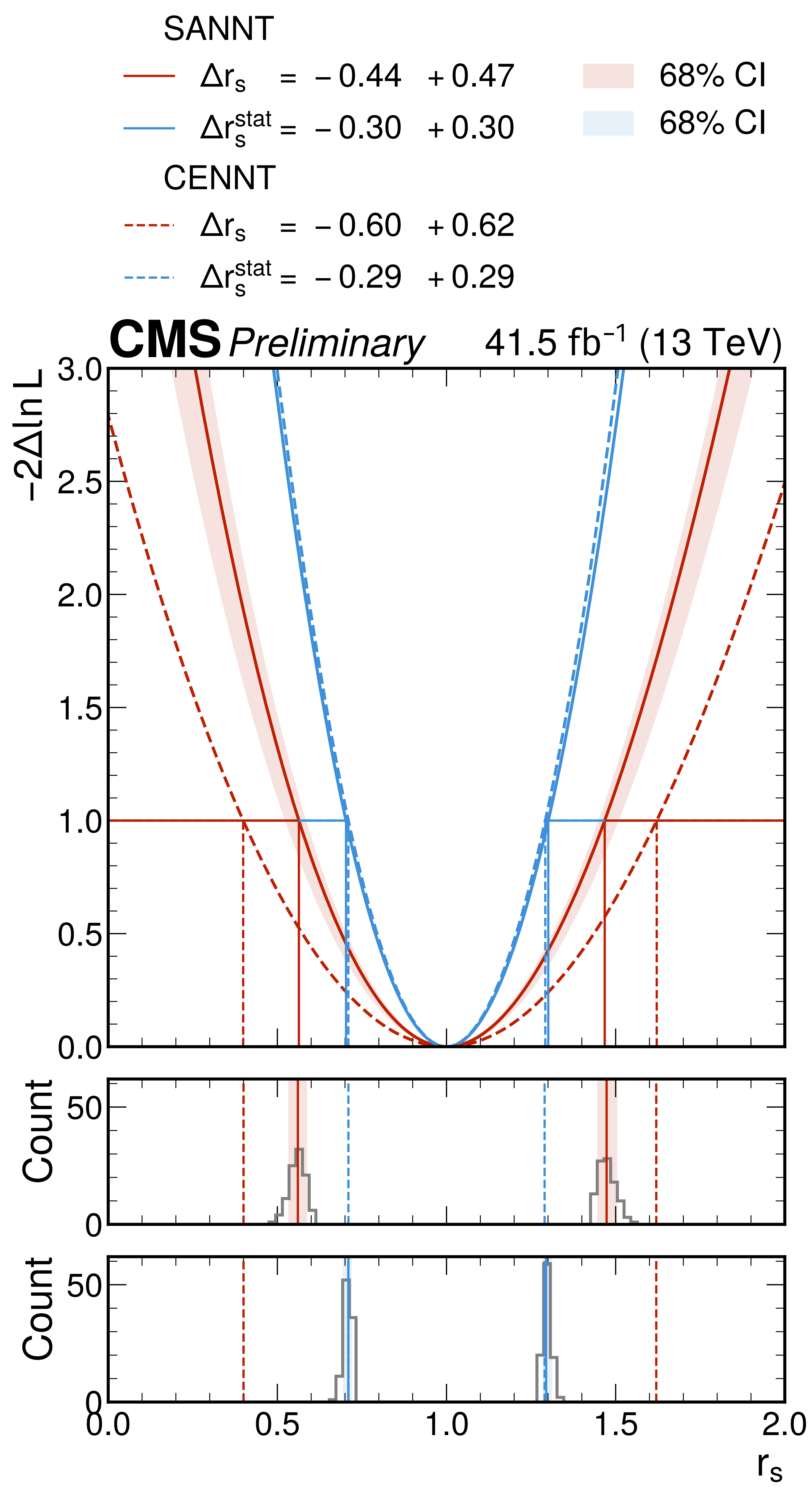


SYSTEMATIC-AWARE LEARNING

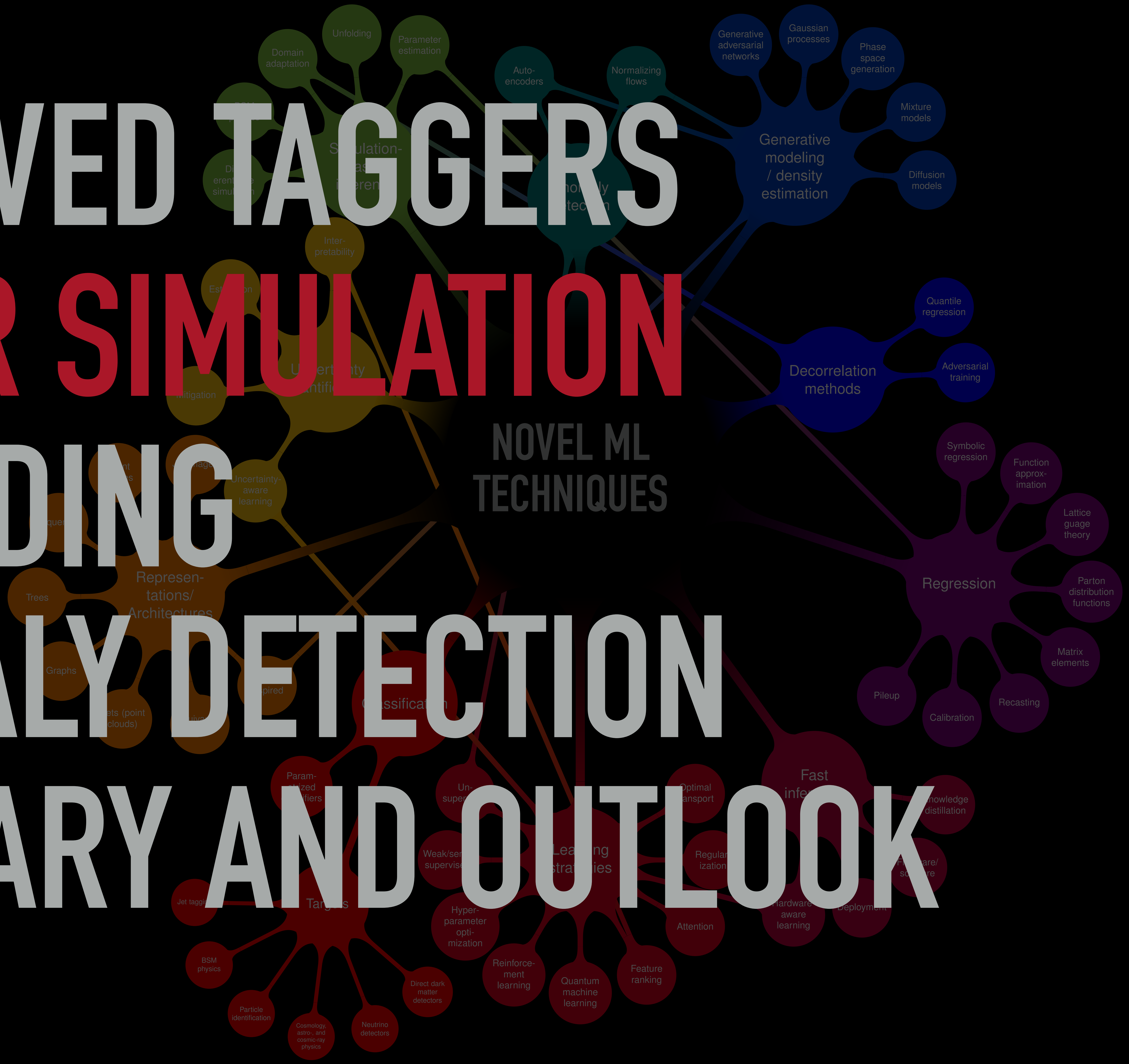
- CENNT optimizes for separation, while SANNT (Δr_s) concentrates signal in bins with smaller background uncertainty
- Total uncertainty improves by 25%

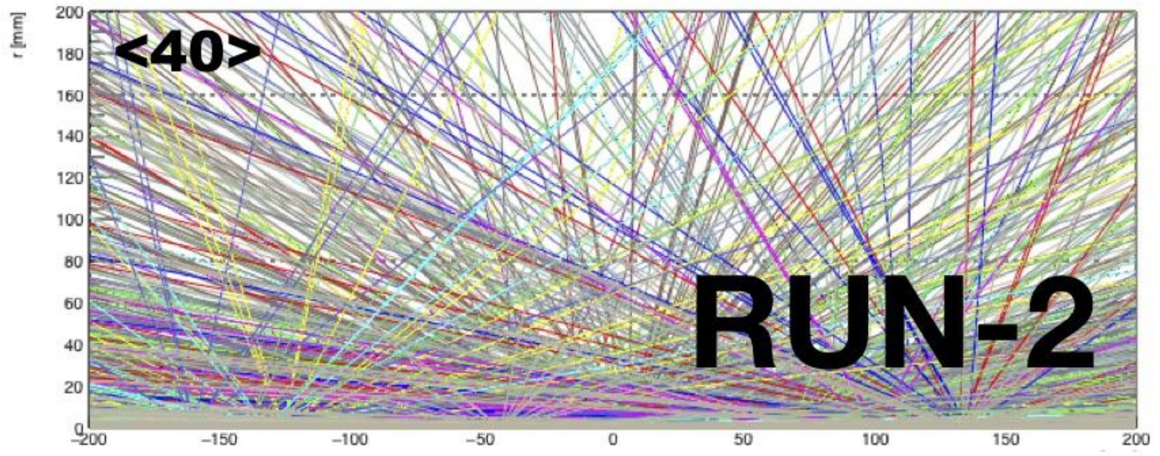


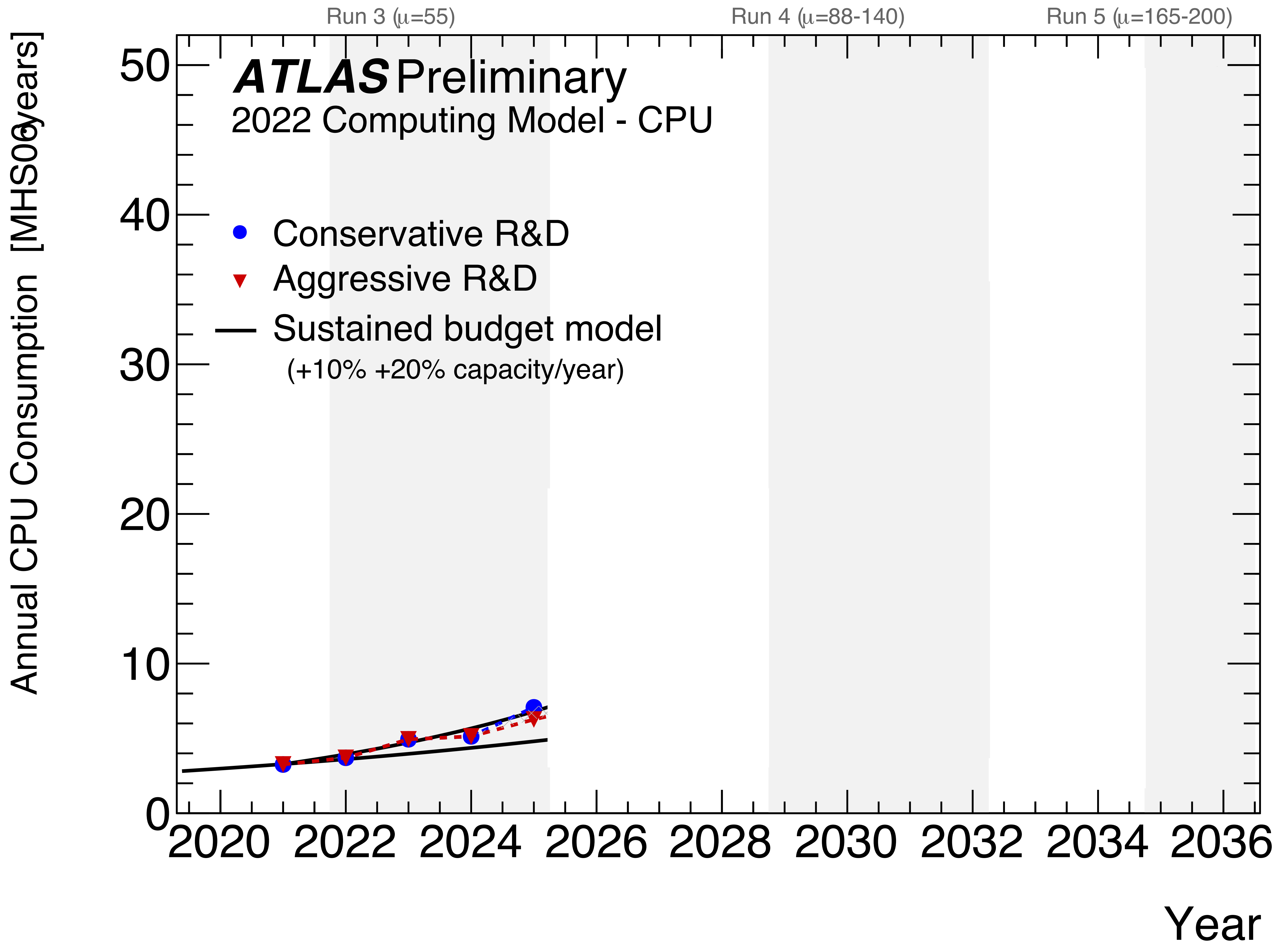




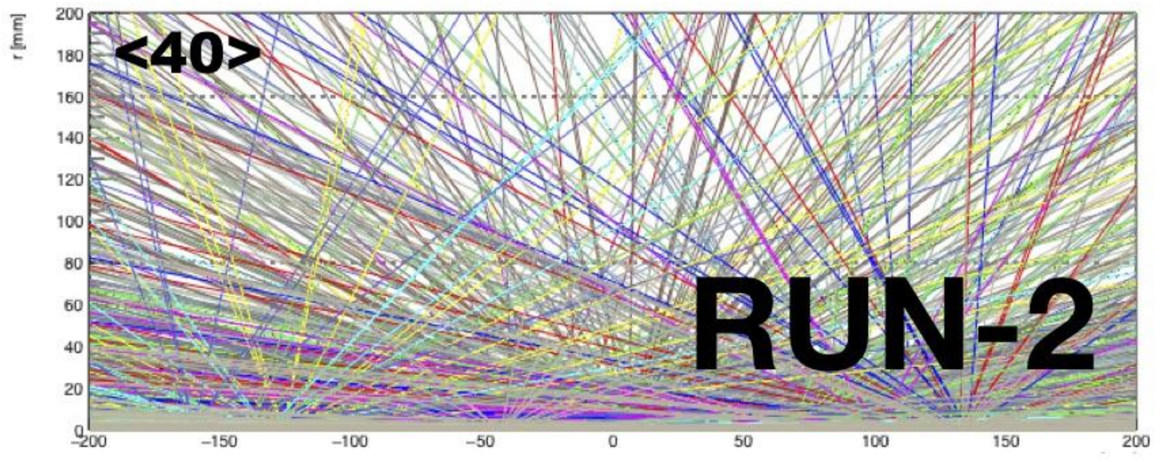
IMPROVED TAGGERS
FASTER SIMULATION
UNFOLDING
ANOMALY DETECTION
SUMMARY AND OUTLOOK

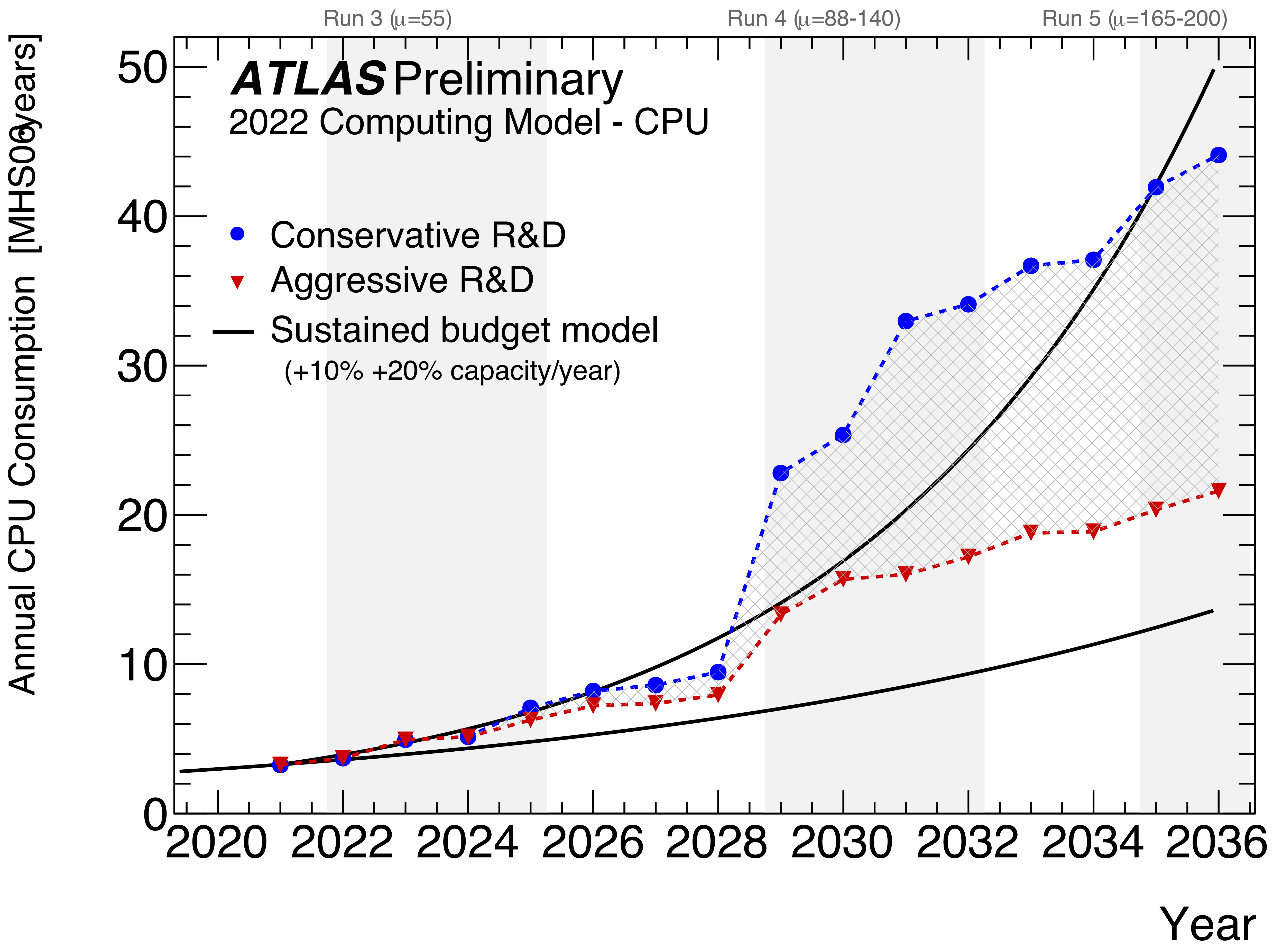


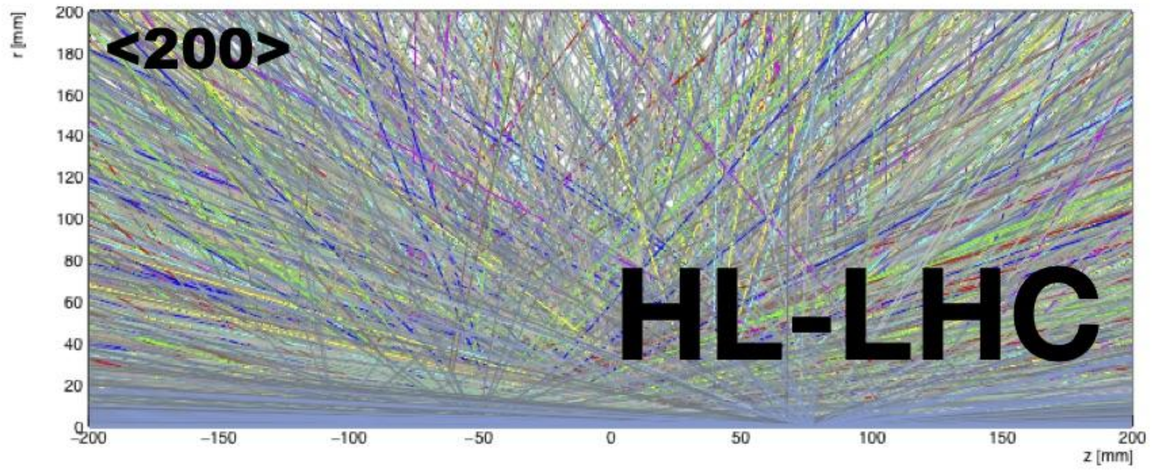

<40>
RUN-2


Run 3 (μ=55)
Run 4 (μ=88-140)
Run 5 (μ=165-200)
ATLAS Preliminary
2022 Computing Model - CPU
Conservative R&D
Aggressive R&D
Sustained budget model
(+10% +20% capacity/year)
Annual CPU Consumption [MHS06years]
Year

TOWARD HL-LHC

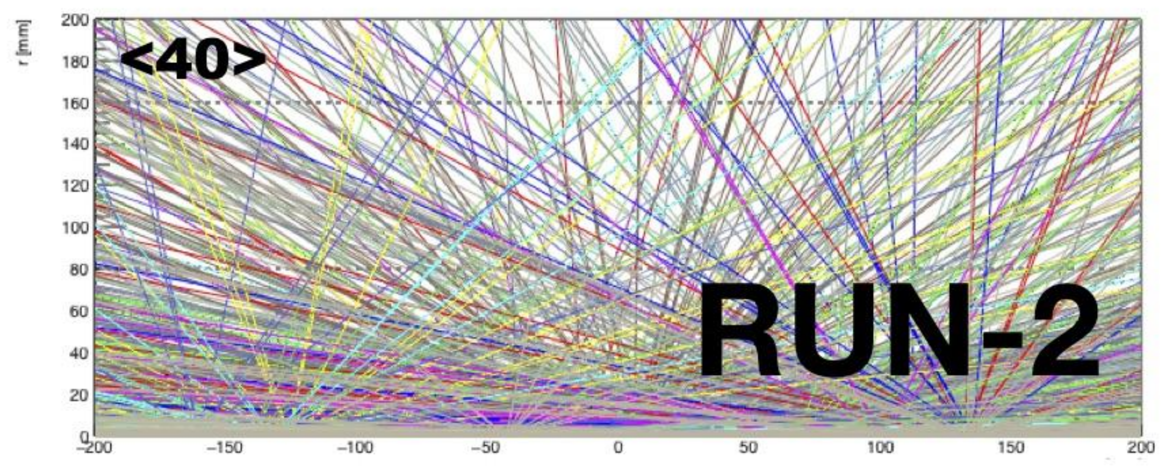


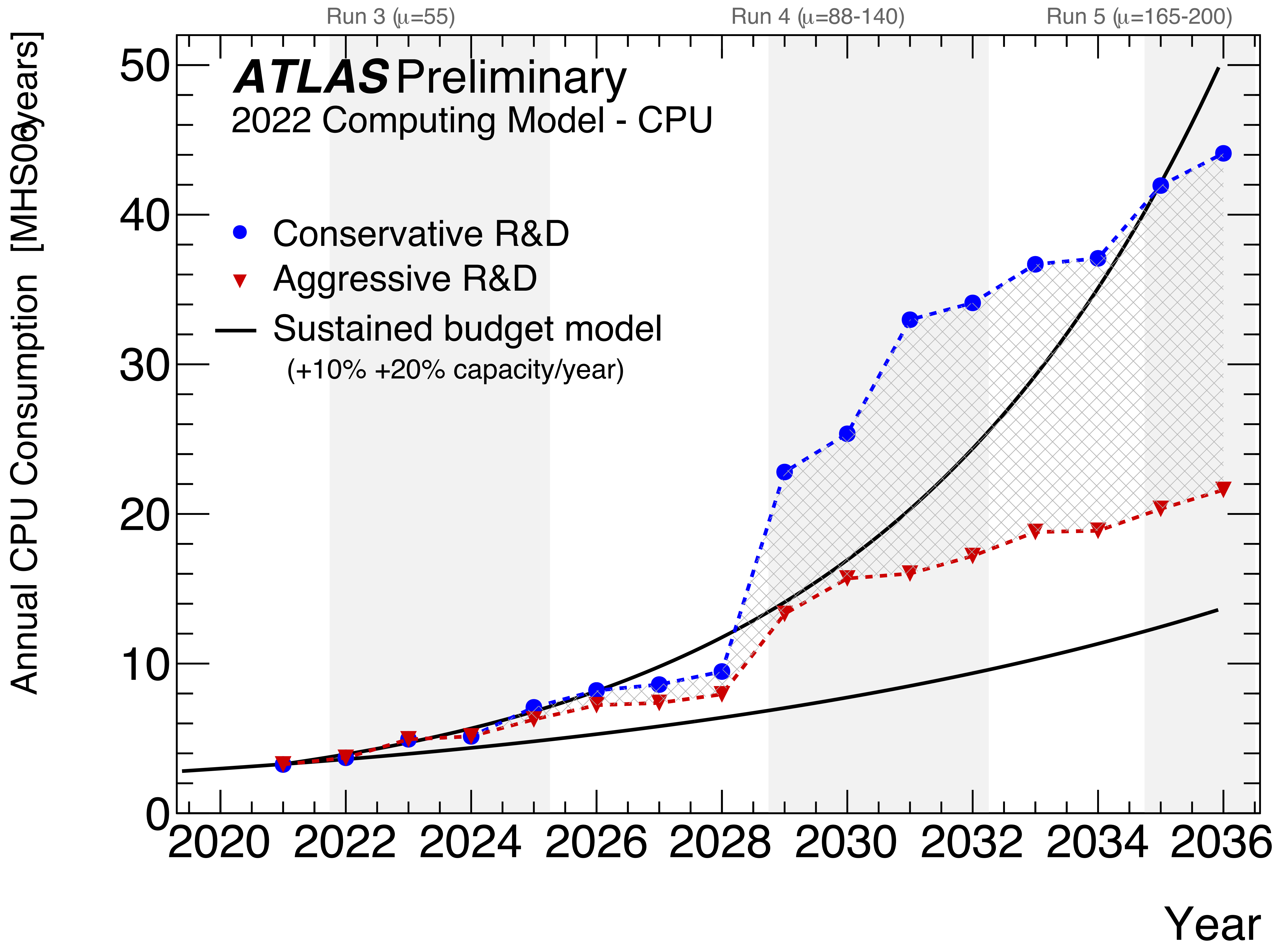


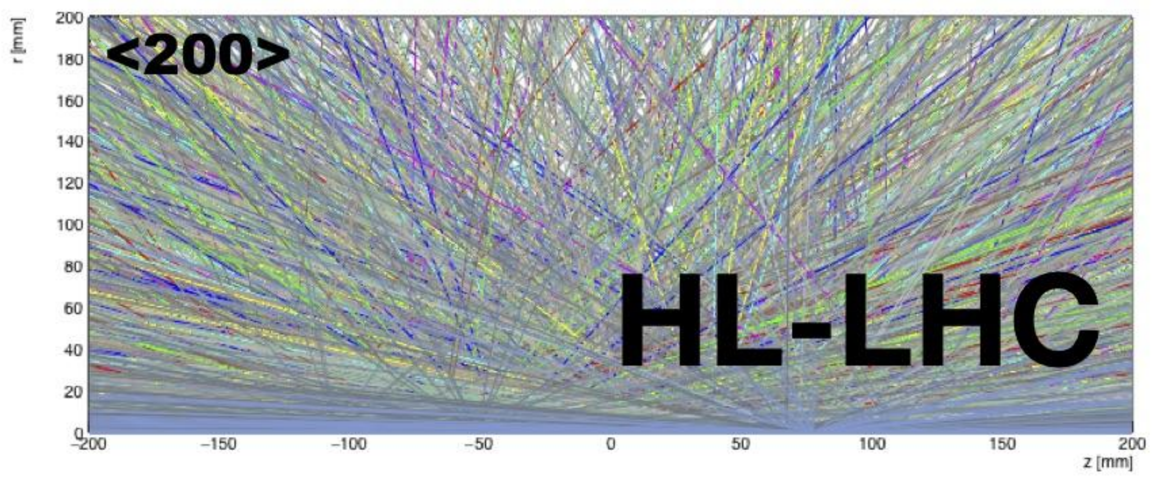


- Simulation is a key driver of CPU needs for the HL-LHC

TOWARD HL-LHC



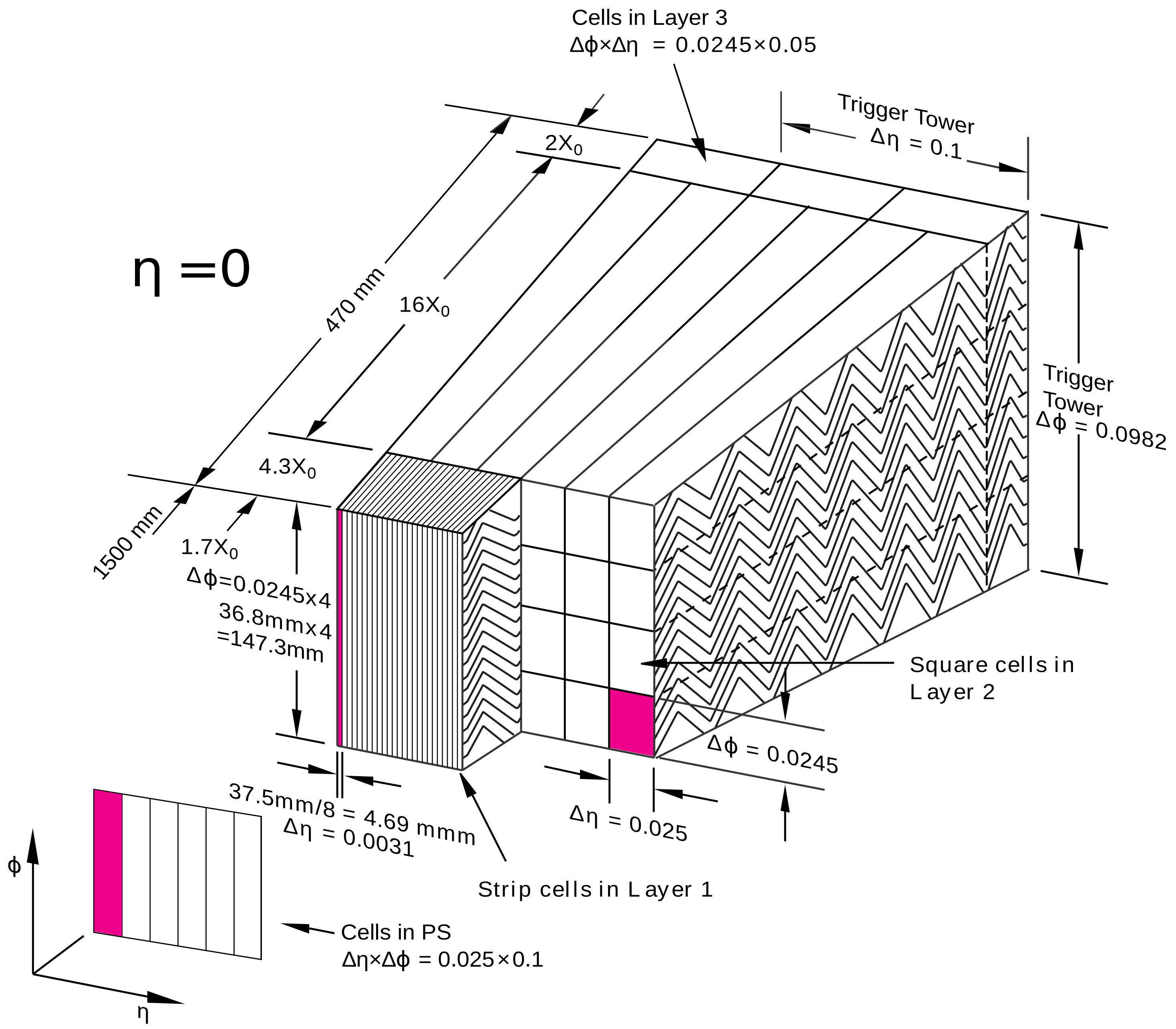




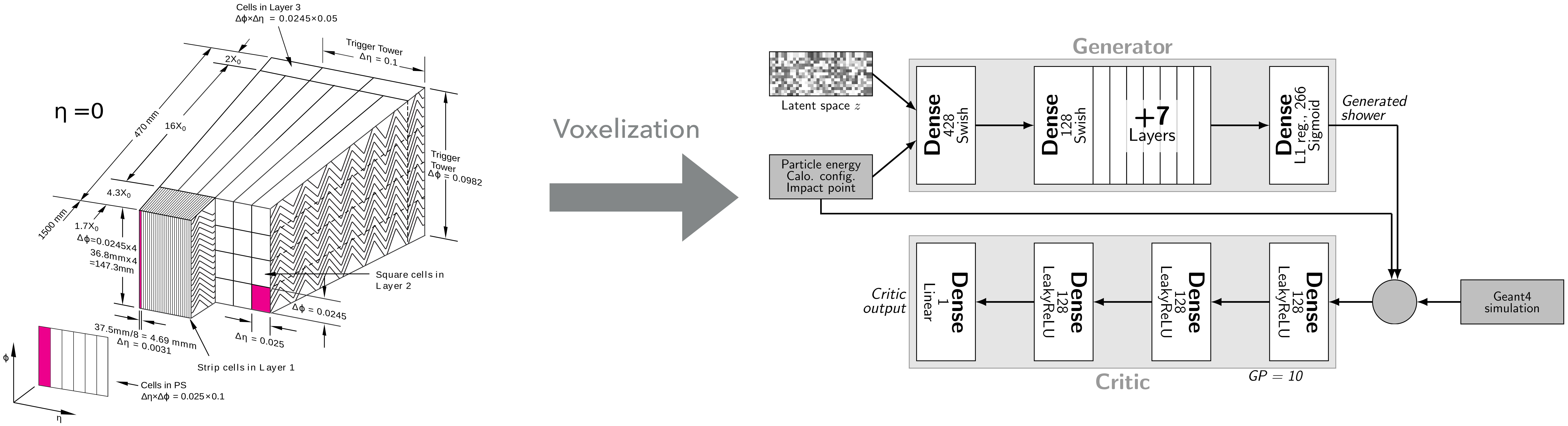
- Simulation is a key driver of CPU needs for the HL-LHC
- ML can be used to “short cut” simulation needs, e.g. additional samples for evaluating systematic uncertainties or higher order corrections

FASTCALOGAN

CSBS 8, 7 (2024)



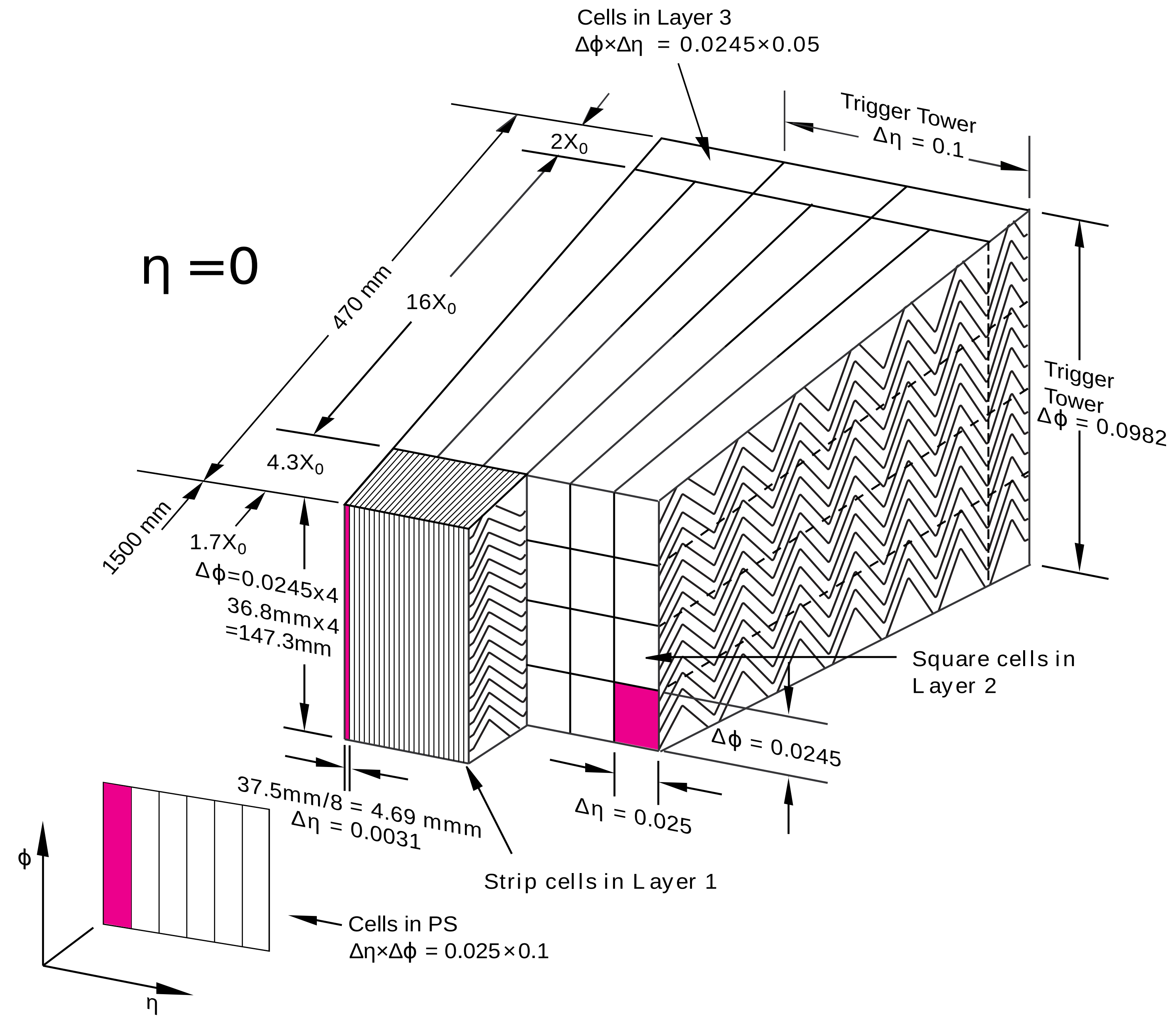
- Can we replace slow Geant4-based simulation of ATLAS calorimeter with fast generative ML?



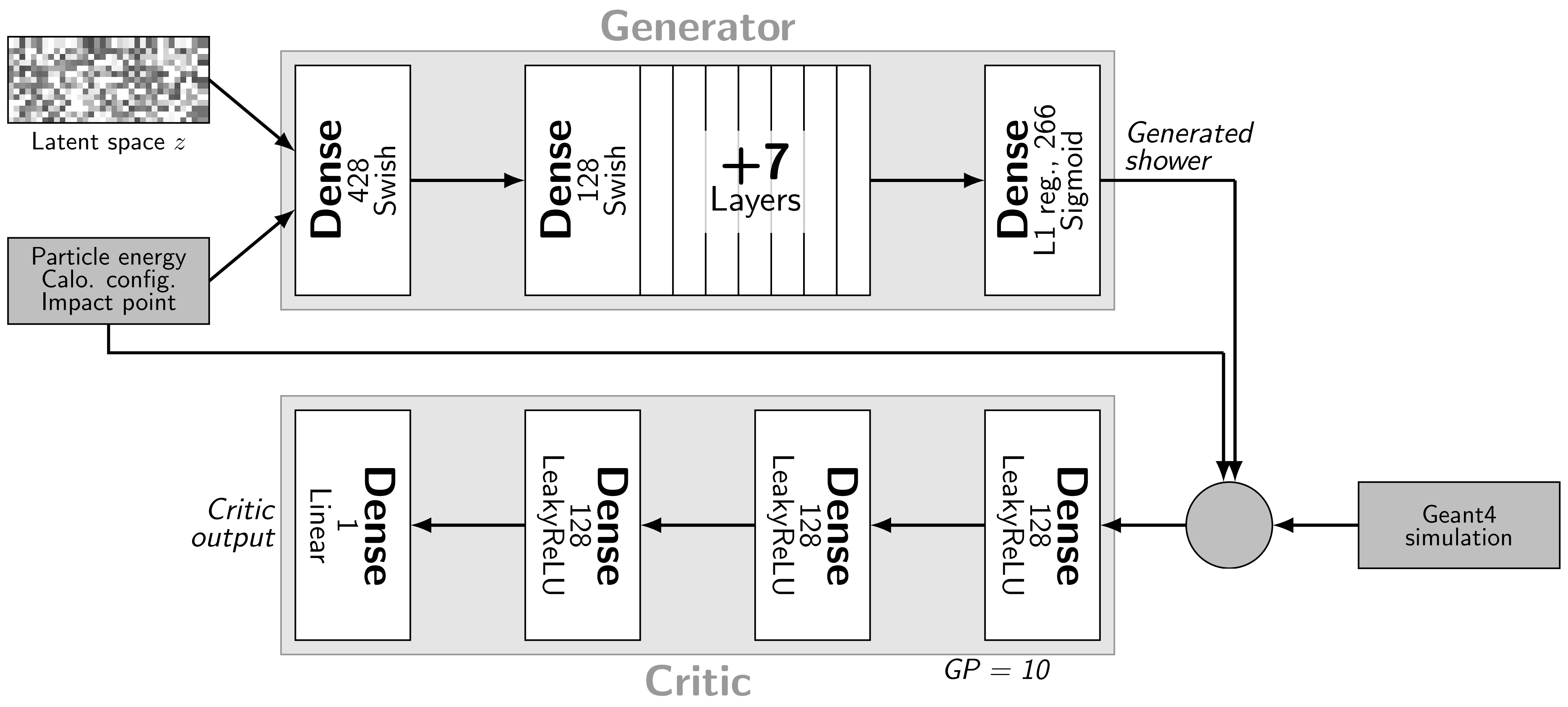
- Can we replace slow Geant4-based simulation of ATLAS calorimeter with fast generative ML?
- GAN/VAEs trained to parametrize the detector response to photons, electrons, and pions

FASTCALOGAN

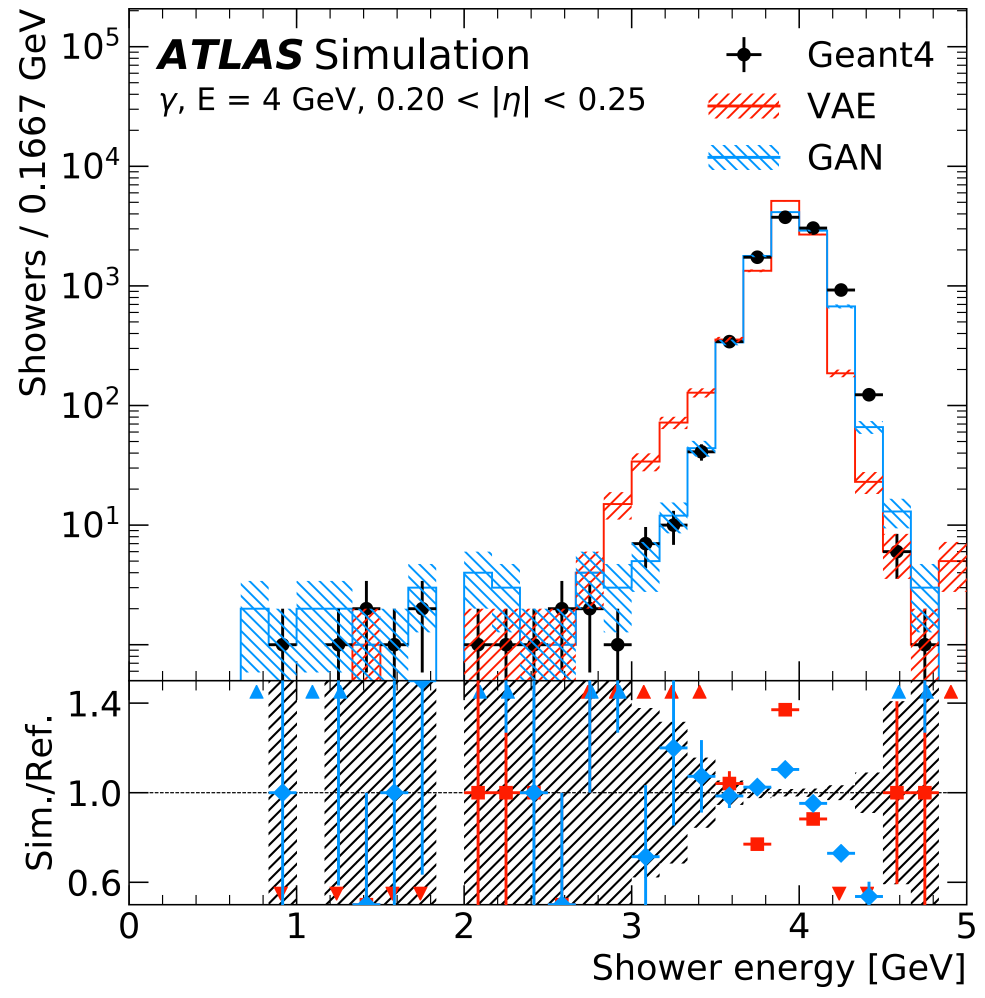
[CSBS 8, 7 (2024)](#)

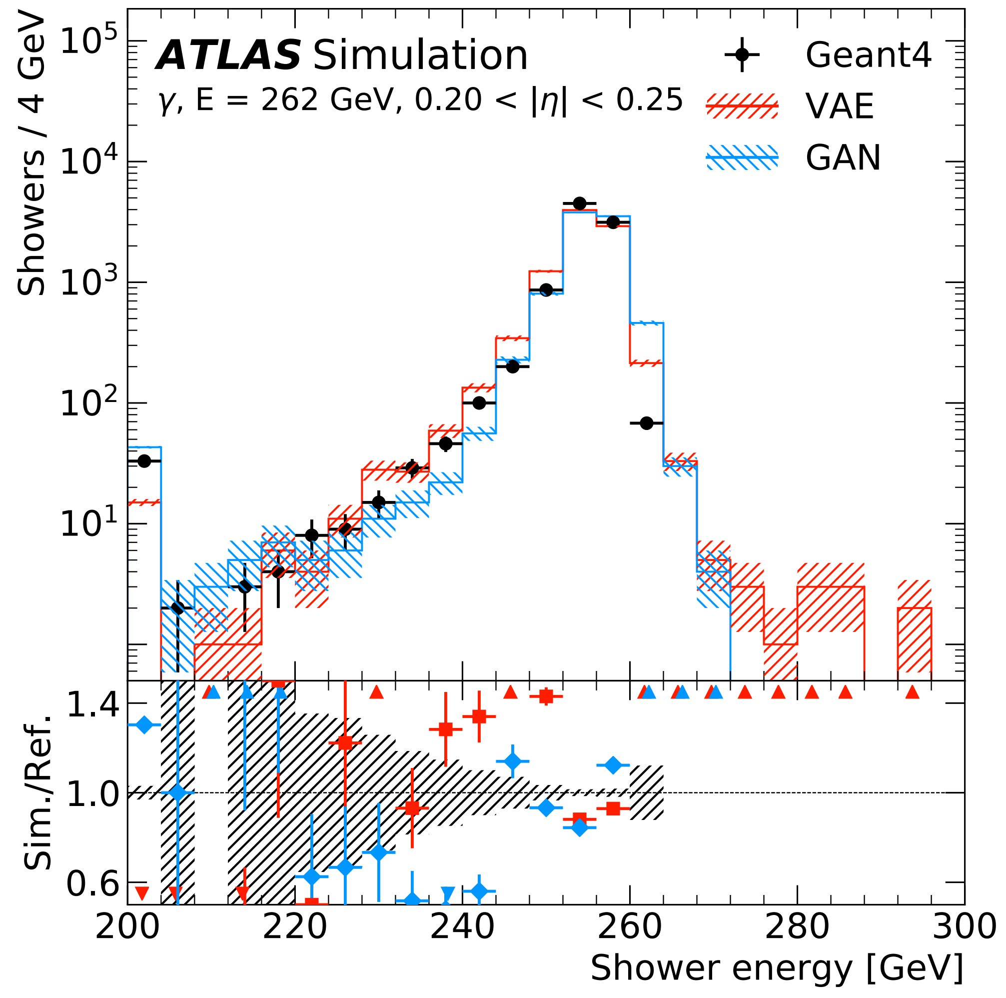


Voxelization



- Can we replace slow Geant4-based simulation of ATLAS calorimeter with fast generative ML?
- GAN/VAEs trained to parametrize the detector response to photons, electrons, and pions
- Good agreement between GAN/VAE and Geant4 for showers of different energies

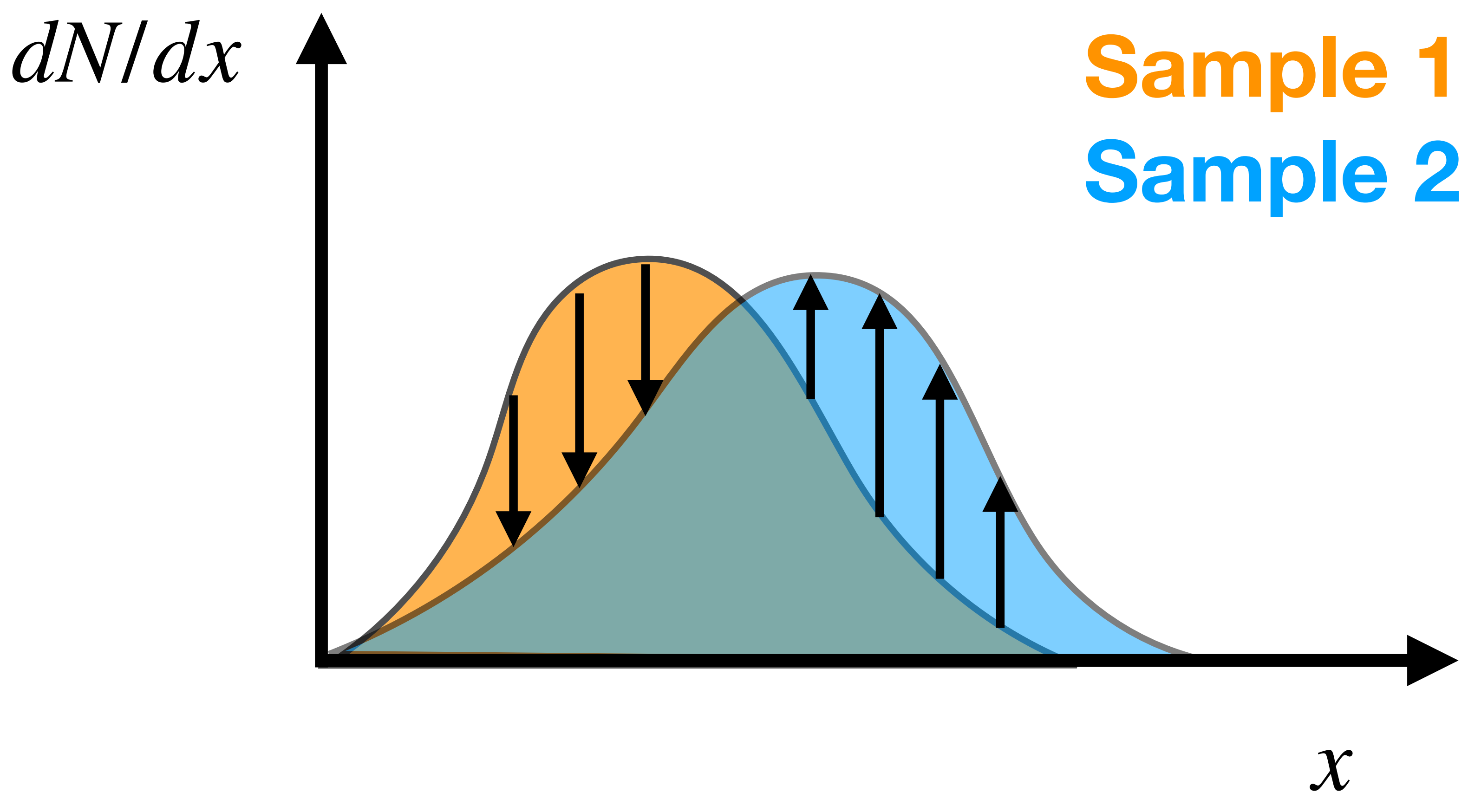




NEURAL NETWORK REWEIGHTING

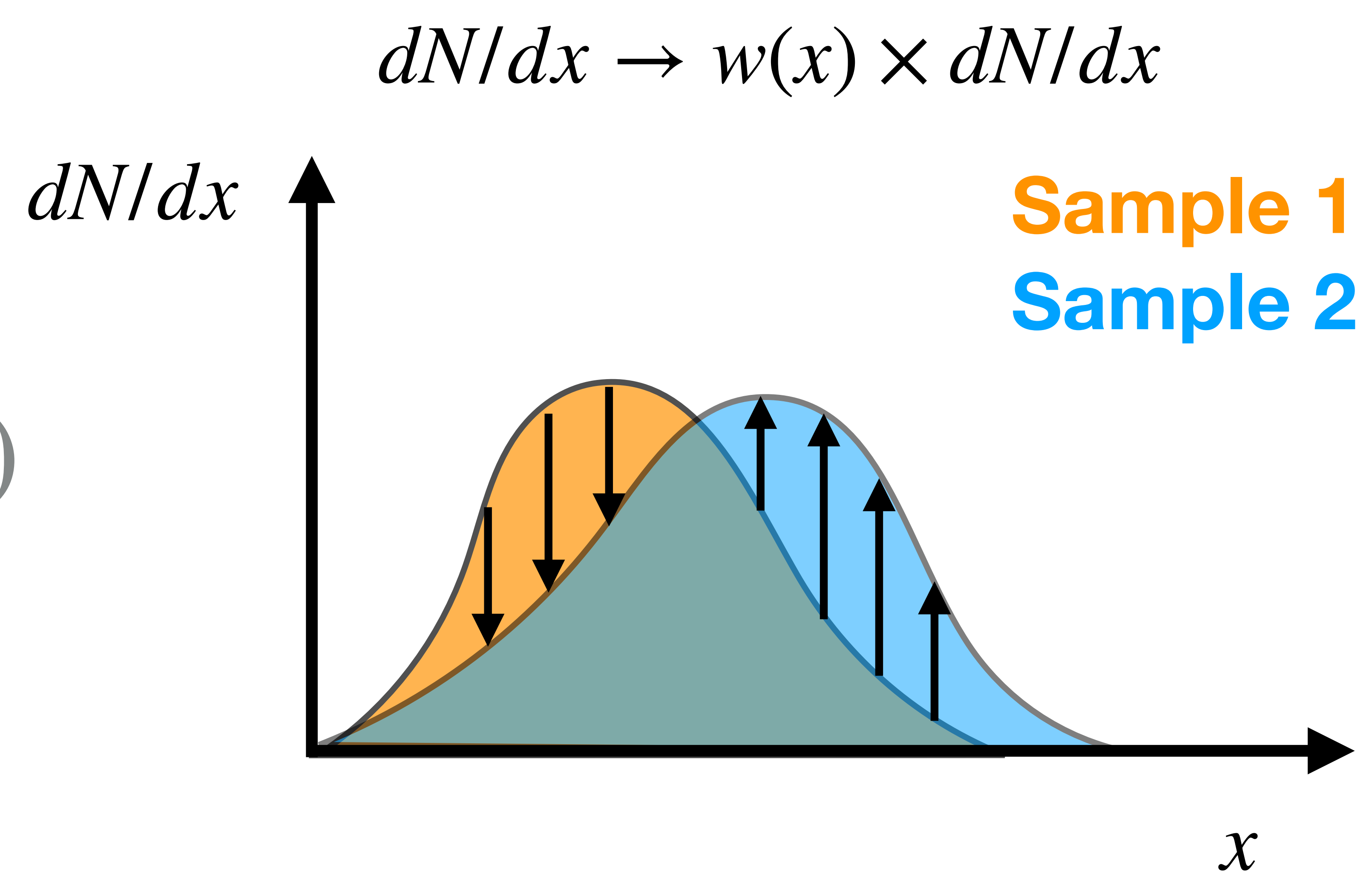
- To reweight sample from $p_1(x)$ to sample from $p_0(x)$ need $w(x) = p_0(x)/p_1(x)$
- **Likelihood ratio trick**: a neural network classifier $f(x)$ trained to distinguish the two samples approximates this ratio: $f(x)/(1-f(x)) \approx p_0(x)/p_1(x)$

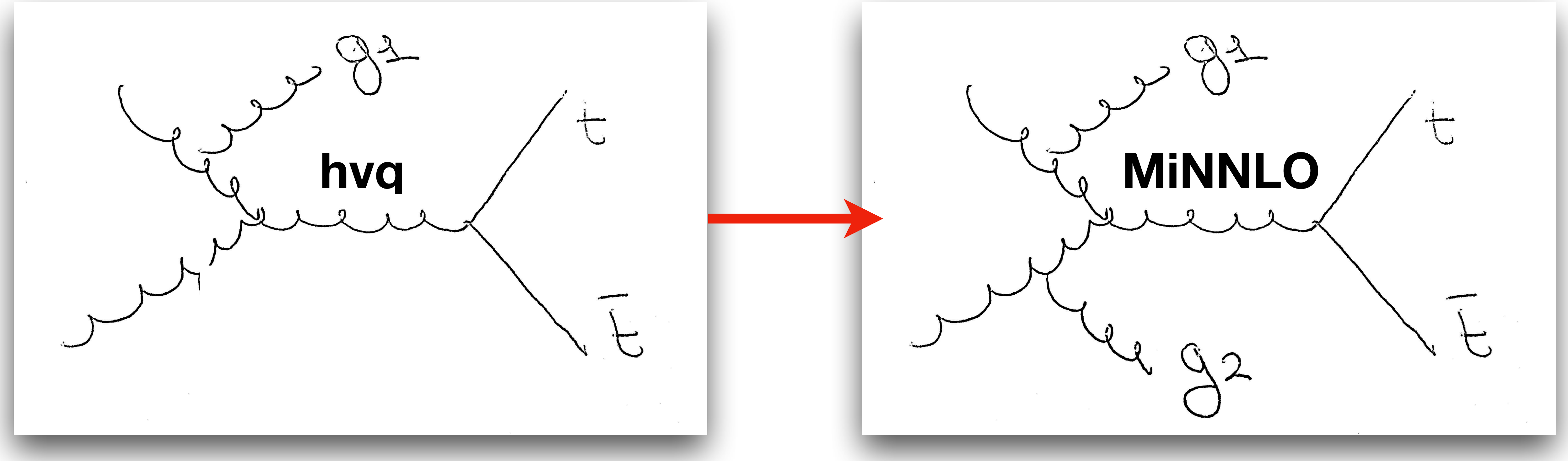
$dN/dx \rightarrow w(x) \times dN/dx$



NEURAL NETWORK REWEIGHTING

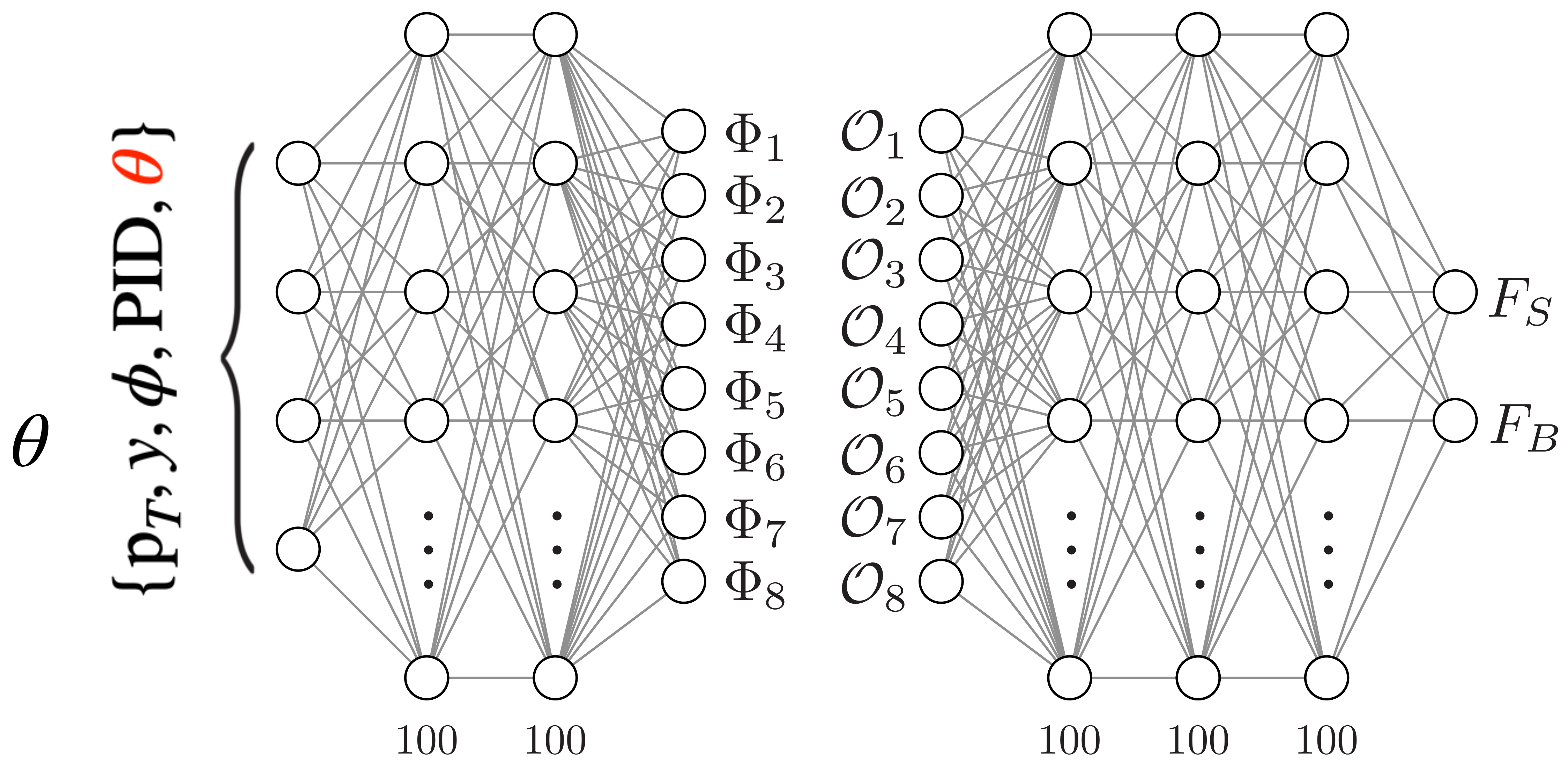
- To reweight sample from $p_1(x)$ to sample from $p_0(x)$ need $w(x) = p_0(x)/p_1(x)$
- **Likelihood ratio trick**: a neural network classifier $f(x)$ trained to distinguish the two samples approximates this ratio: $f(x)/(1-f(x)) \approx p_0(x)/p_1(x)$
- CMS use cases for top quark physics
  - Reweight MC for evaluation of systematic uncertainties
  - **Reweight MC to higher-order theory predictions**





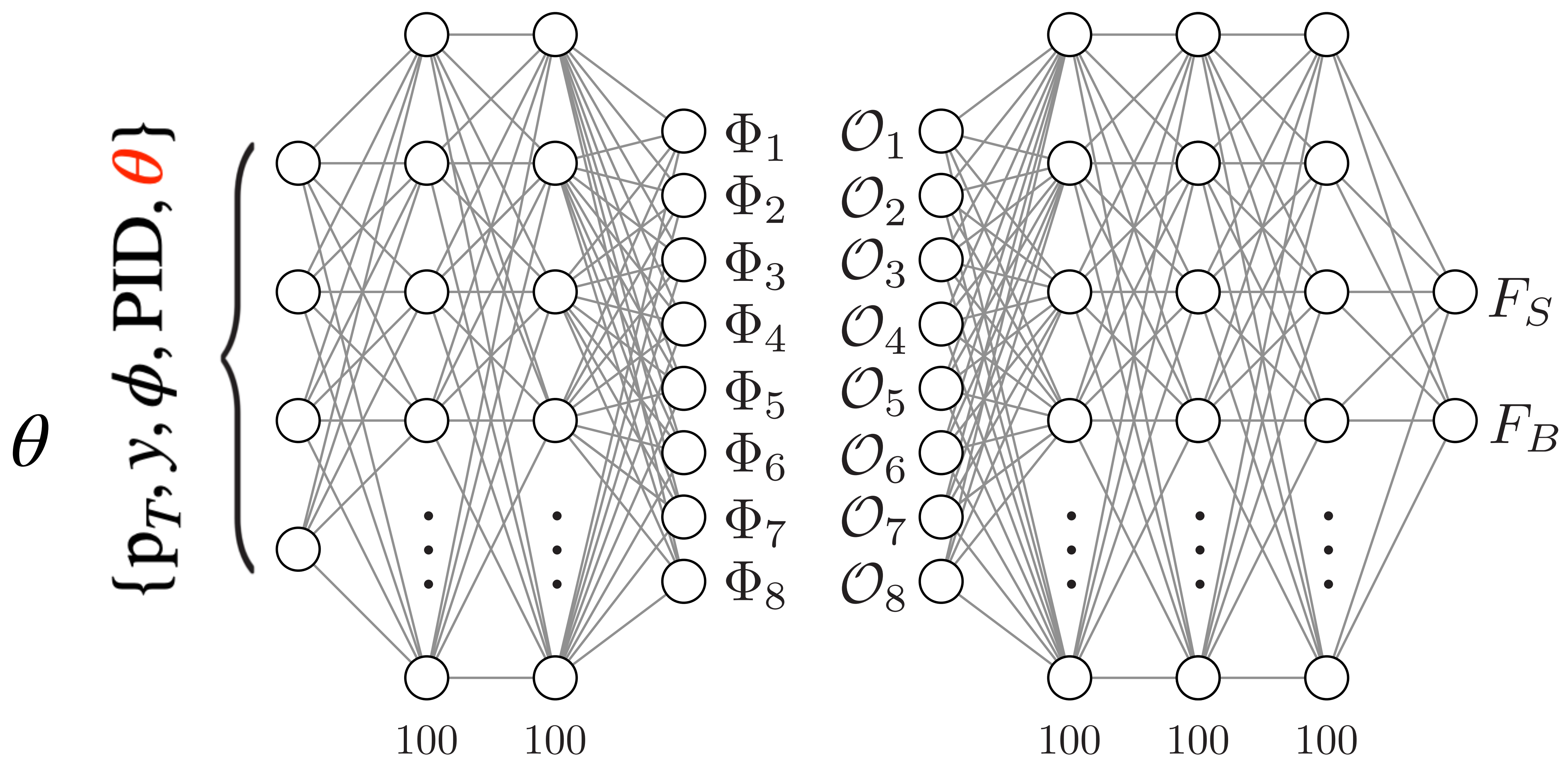
HIGHER ORDER REWEIGHTING RESULTS

- Parton level information input to the Particle Flow Network:
  - 4-vector $(p_{\mathrm{T}}, \eta, \phi, m)$ and PID of $[t, \bar{t}, t\bar{t}]$ system of the showered events

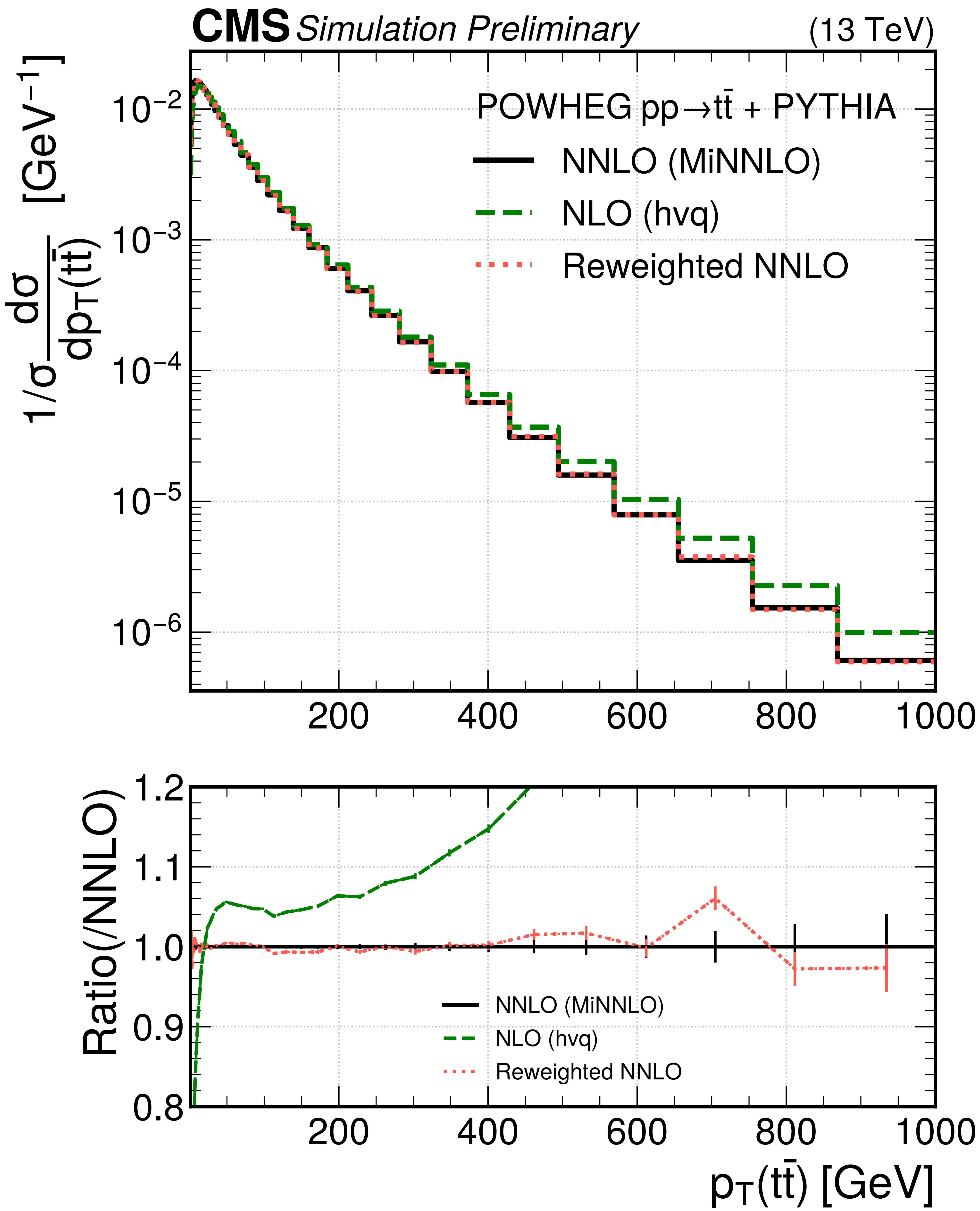


HIGHER ORDER REWEIGHTING RESULTS

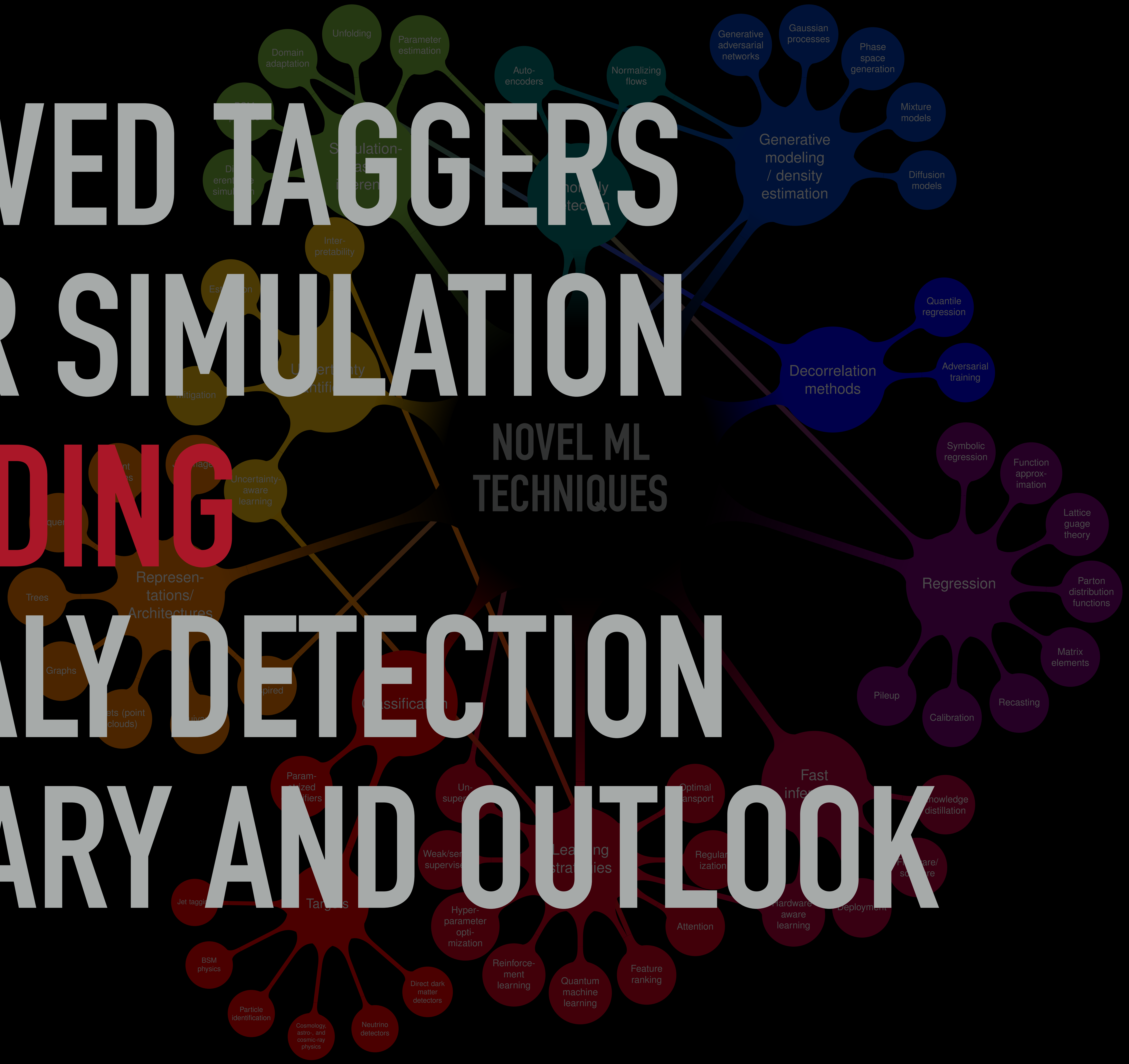
- Parton level information input to the Particle Flow Network:
  - 4-vector $(p_{\mathrm{T}}, \eta, \phi, m)$ and PID of $[t, \bar{t}, t\bar{t}]$ system of the showered events



- Method closes within ~2% of target NNLO distribution

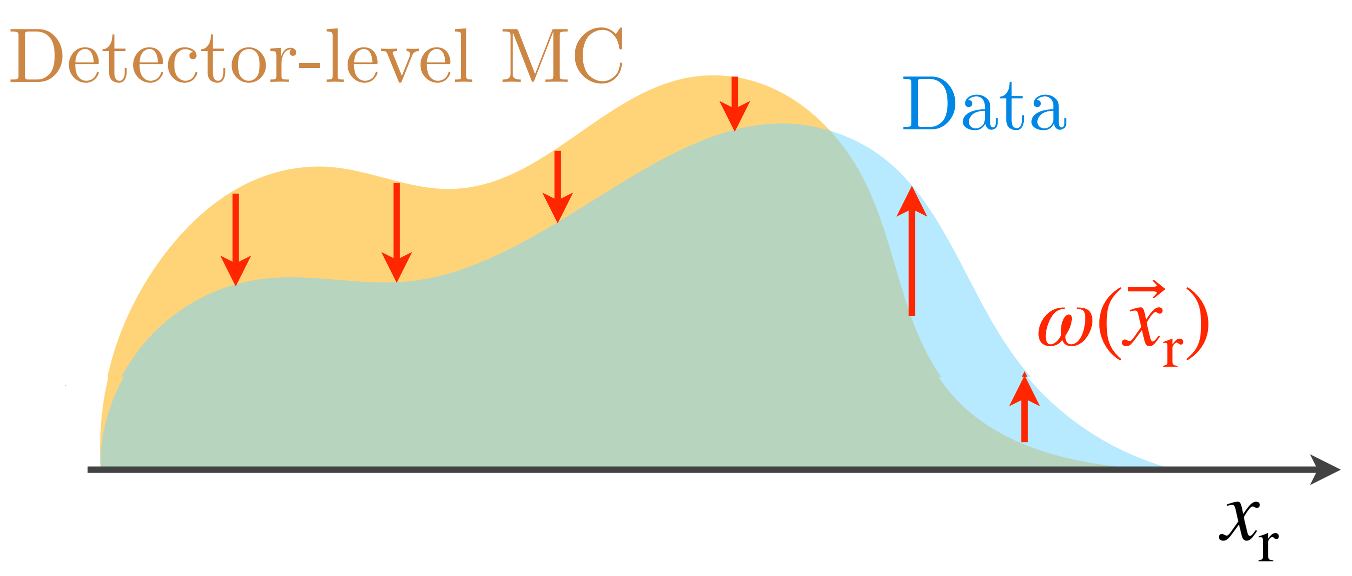


IMPROVED TAGGERS
FASTER SIMULATION
UNFOLDING
ANOMALY DETECTION
SUMMARY AND OUTLOOK



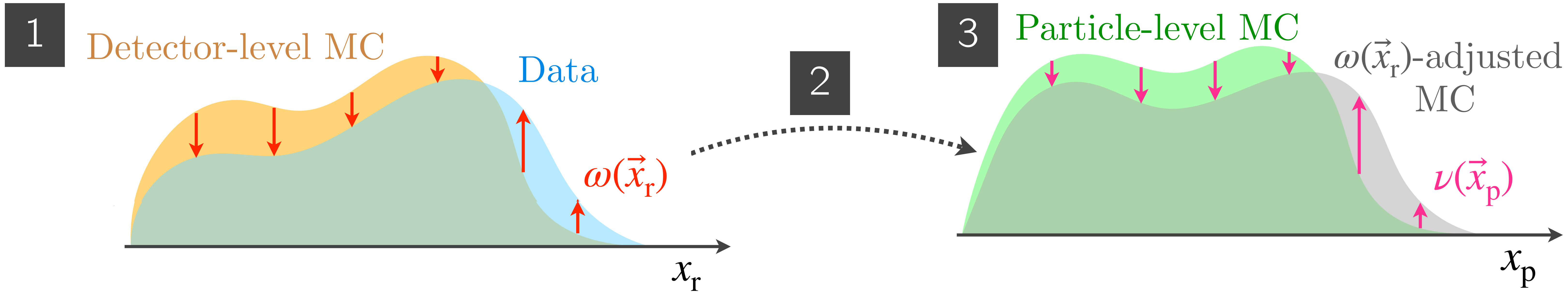
- Omnifold used to unfold 24 Z+jets kinematic observables in ATLAS
  - Detector-level MC corrected by $\omega(\vec{x}_r)$ to match data

1



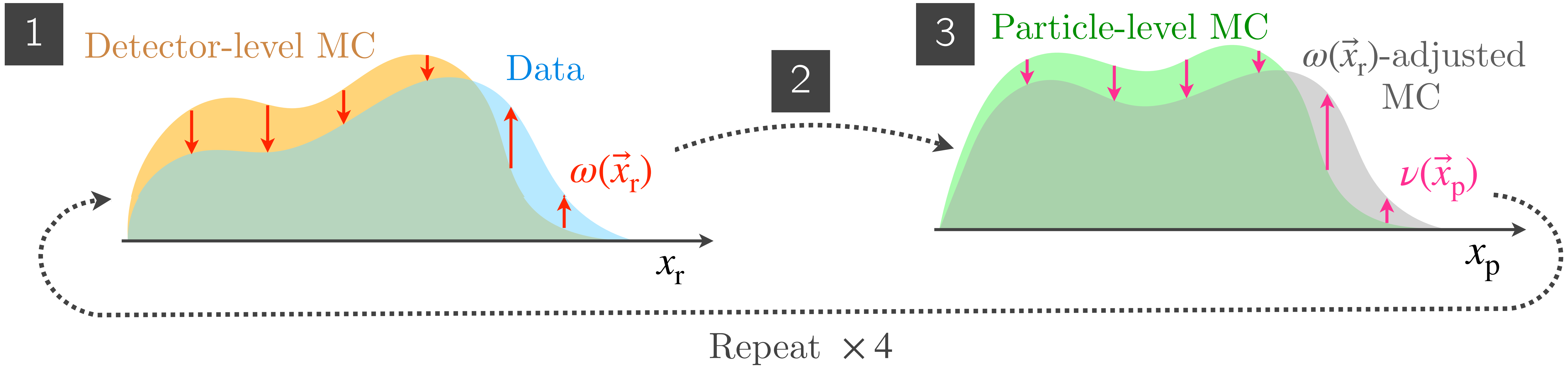
MEASUREMENTS & UNFOLDING WITH OMNIFOLD

- Omnifold used to unfold 24 Z+jets kinematic observables in ATLAS
  - Detector-level MC corrected by $\omega(\vec{x}_r)$ to match data
  - Particle-level MC corrected by $\nu(\vec{x}_p)$ to match $\omega(\vec{x}_r)$-adjusted MC



MEASUREMENTS & UNFOLDING WITH OMNIFOLD

- Omnifold used to unfold 24 Z+jets kinematic observables in ATLAS
  - Detector-level MC corrected by $\omega(\vec{x}_r)$ to match data
  - Particle-level MC corrected by $\nu(\vec{x}_p)$ to match $\omega(\vec{x}_r)$-adjusted MC
  - Method repeated four more times



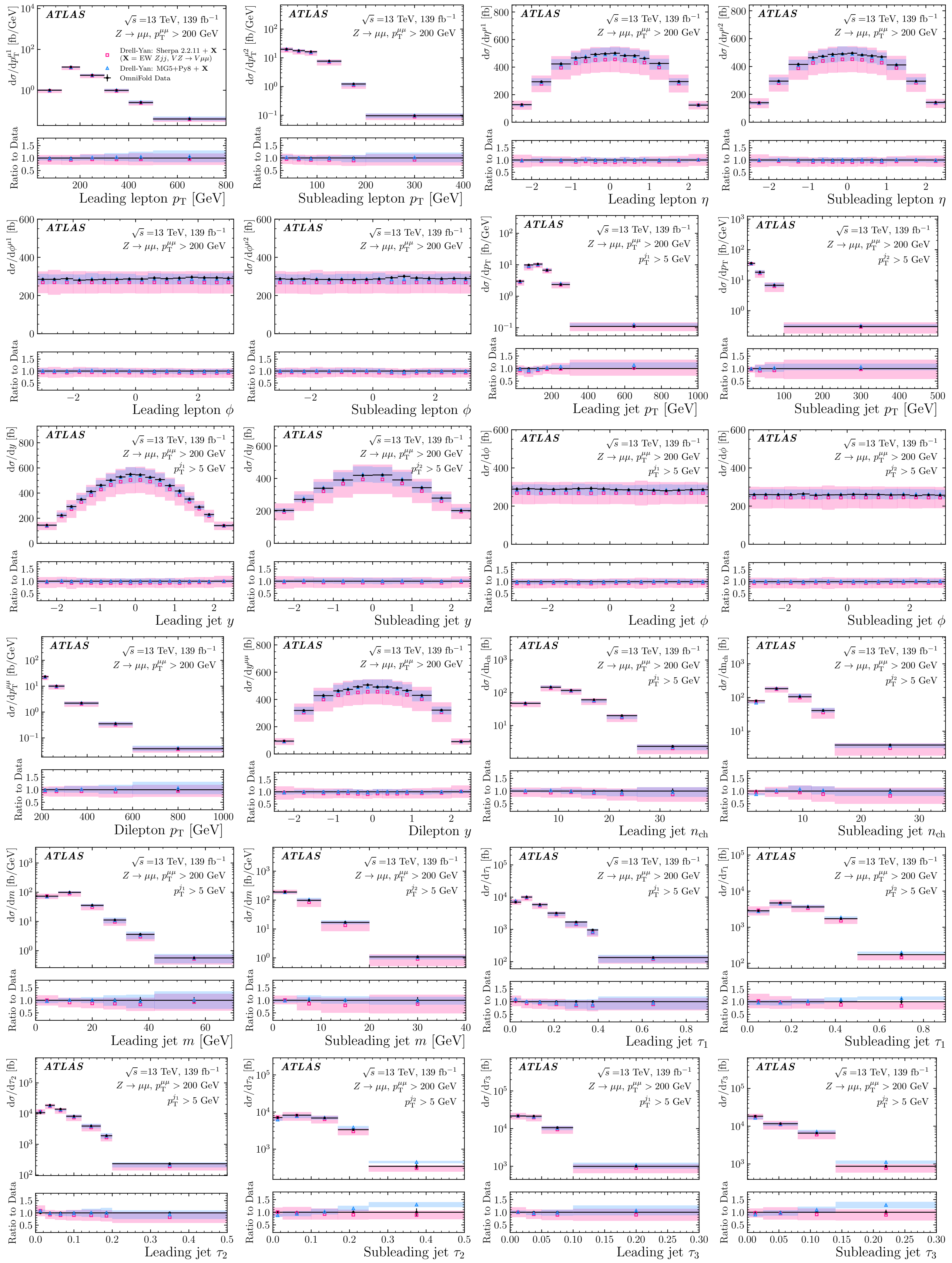
MEASUREMENTS & UNFOLDING WITH OMNIFOLD

- Omnifold used to unfold 24 Z+jets kinematic observables in ATLAS
  - Detector-level MC corrected by $\omega(\vec{x}_r)$ to match data
  - Particle-level MC corrected by $\nu(\vec{x}_p)$ to match $\omega(\vec{x}_r)$-adjusted MC
  - Method repeated four more times
  - Final measurement is $\nu(\vec{x}_p)$-weighted MC events

MEASUREMENTS & UNFOLDING WITH OMNIFOLD

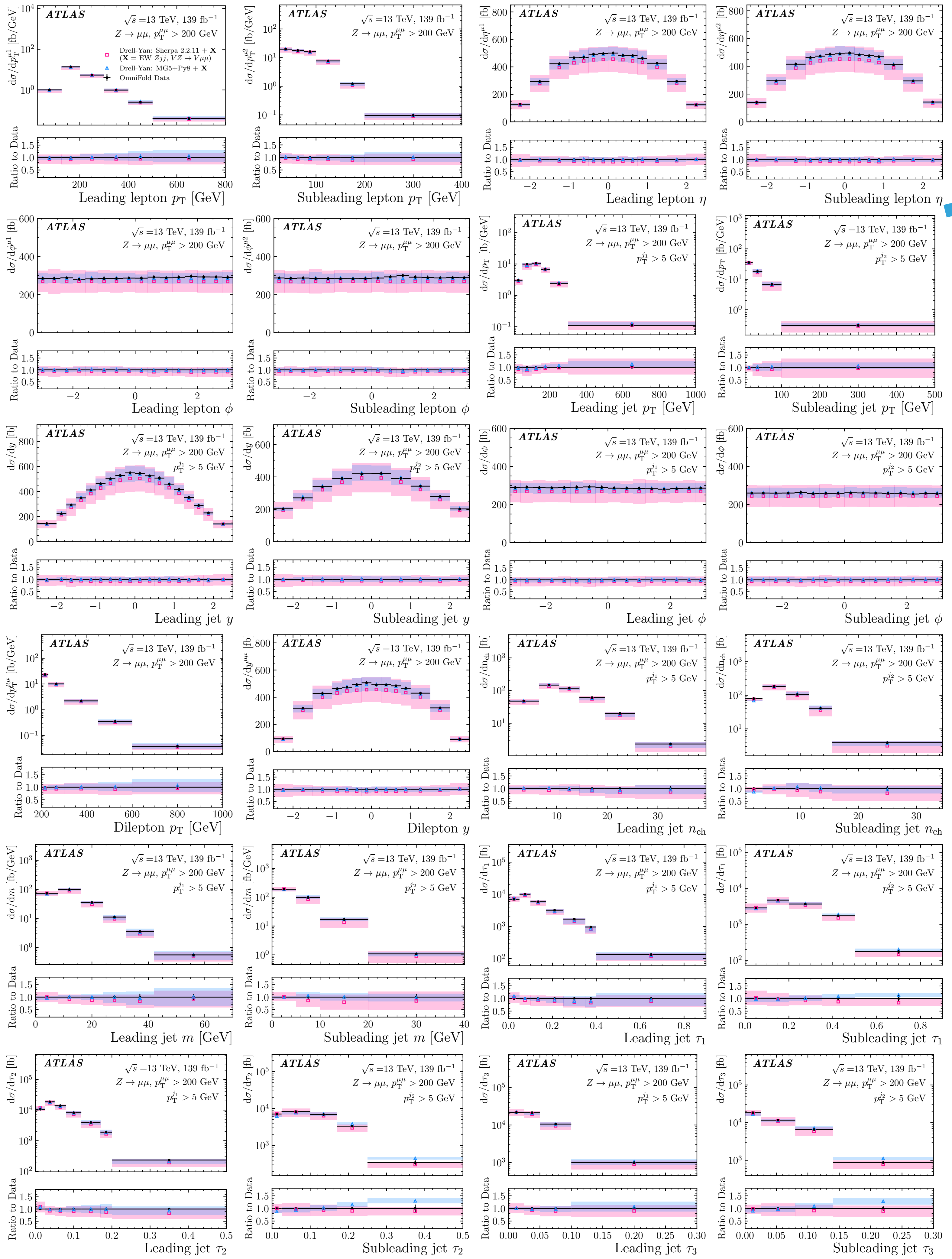
arXiv:2405.20041

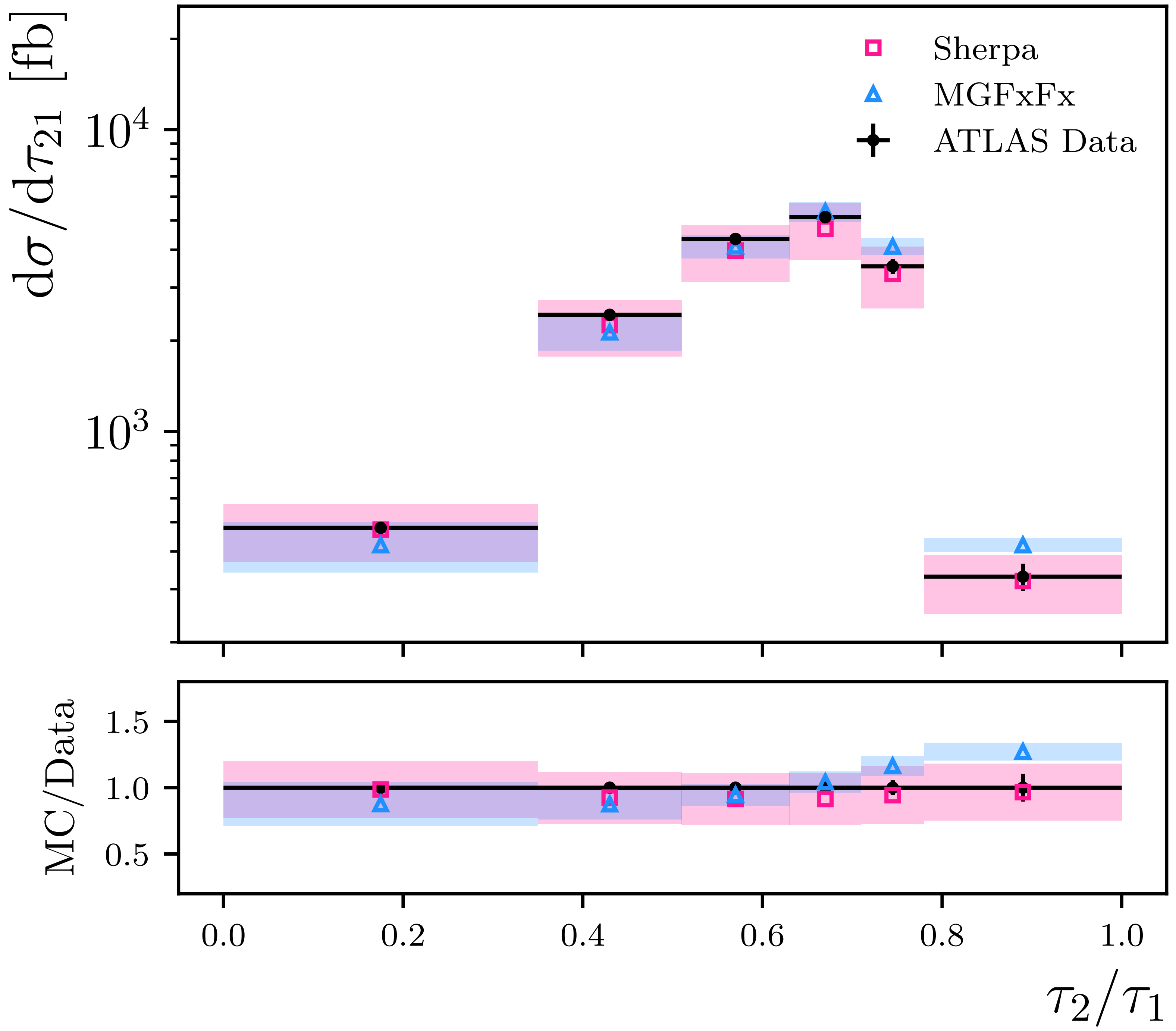
- Unbinned results provided as 24-dim. dataset on Zenodo and code on GitLab

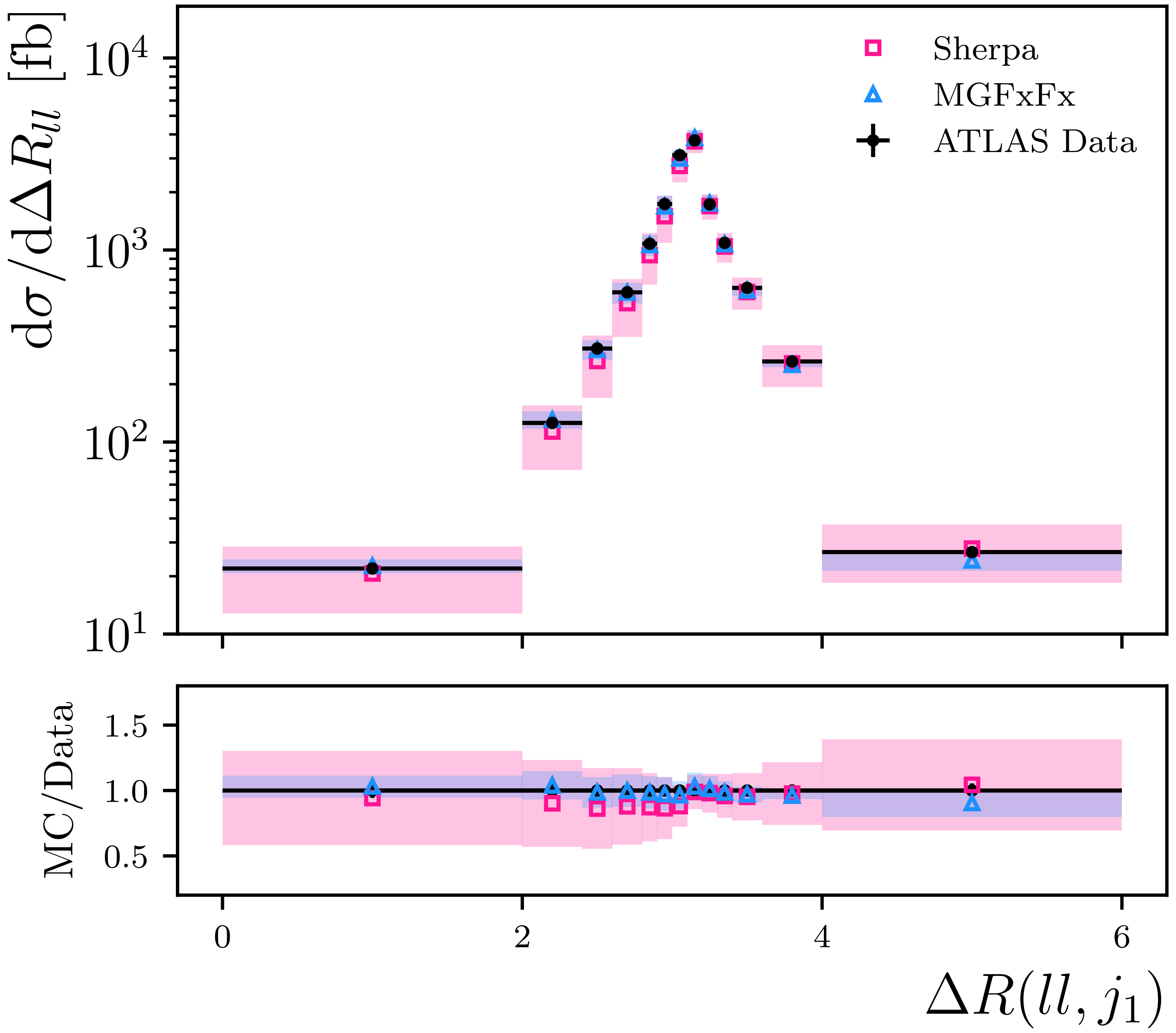


MEASUREMENTS & UNFOLDING WITH OMNIFOLD

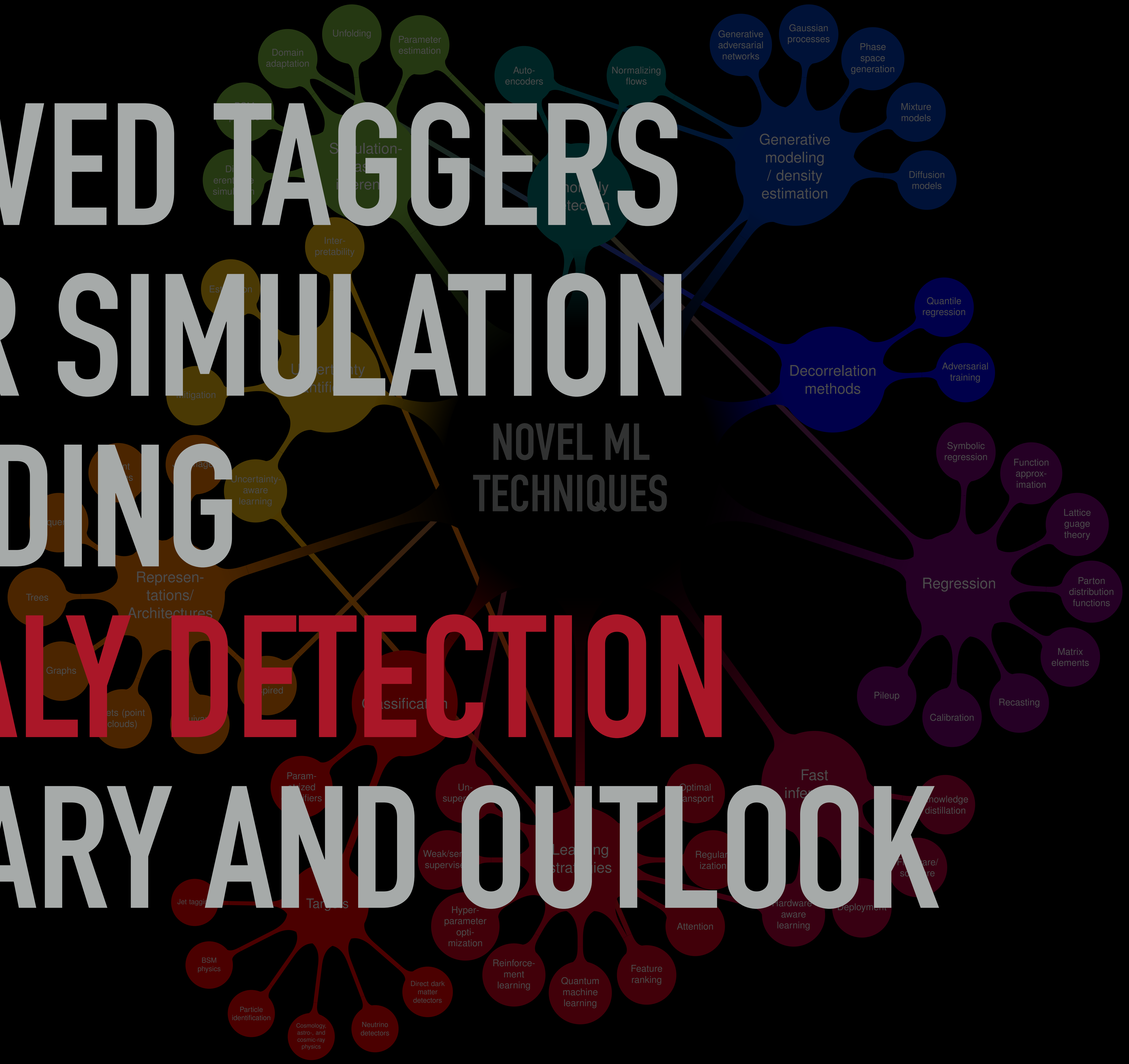
- Unbinned results provided as 24-dim. dataset on Zenodo and code on GitLab
- Enables measurement of new observables not presented in original paper!





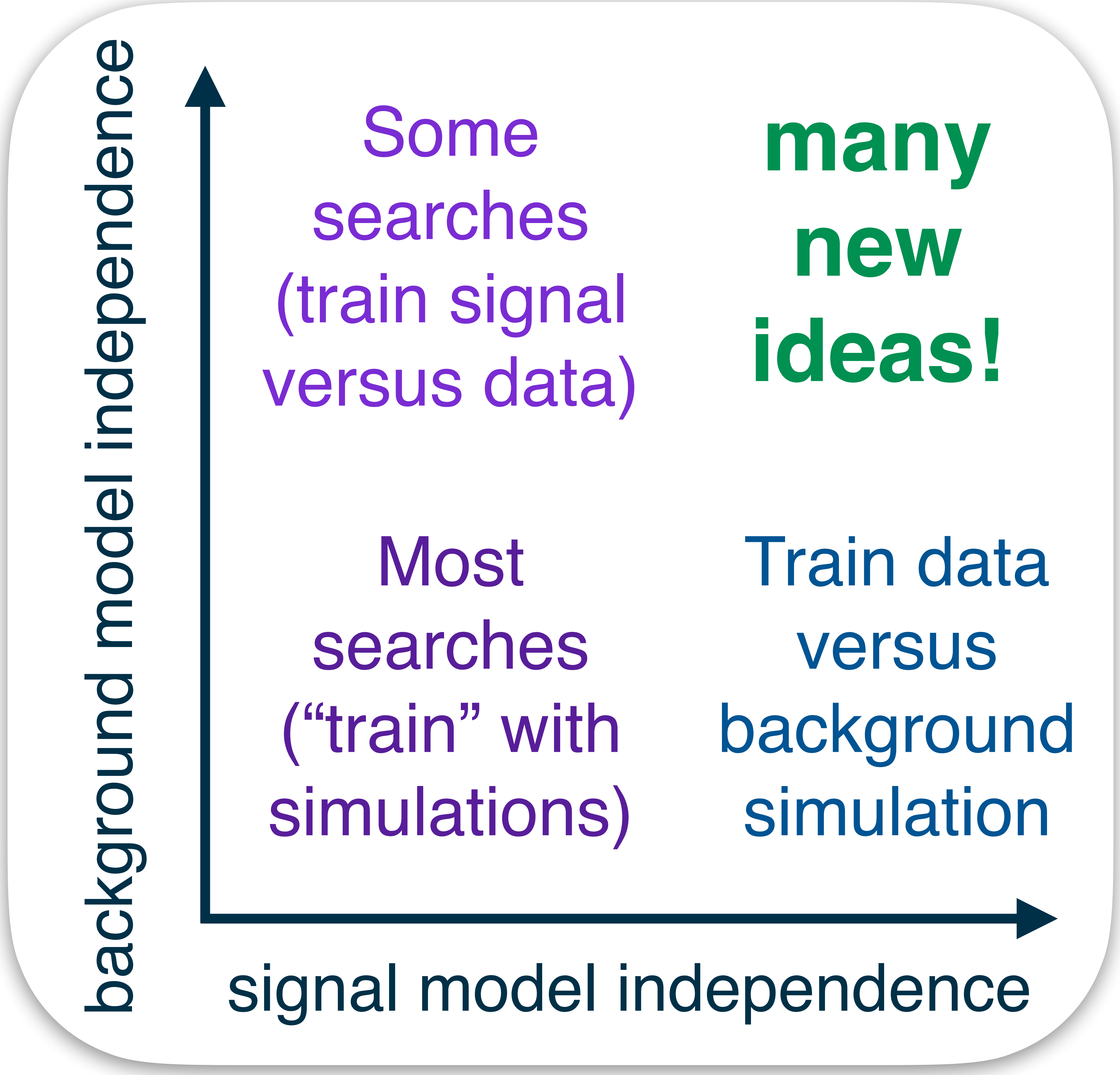


IMPROVED TAGGERS
FASTER SIMULATION
UNFOLDING
ANOMALY DETECTION
SUMMARY AND OUTLOOK

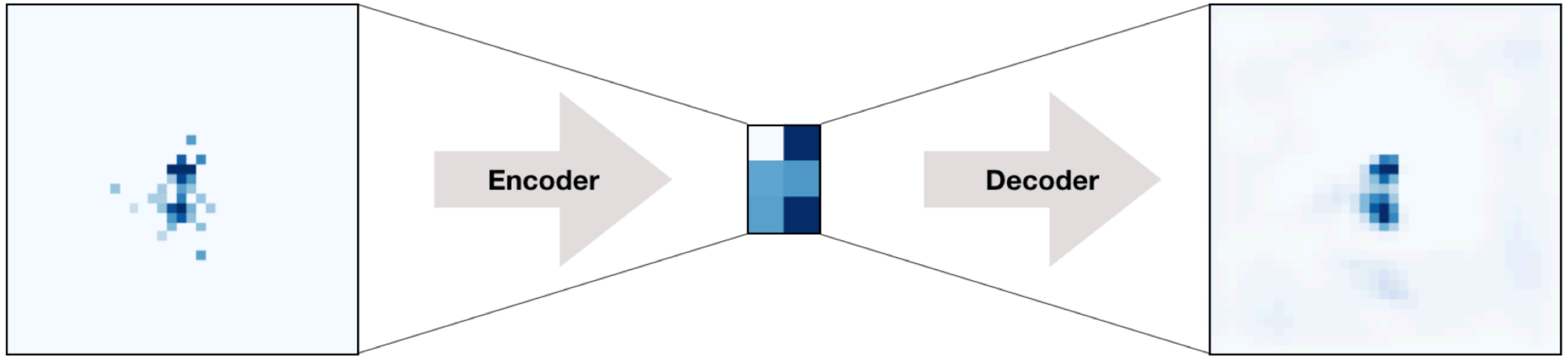


LANDSCAPE OF BSM SEARCHES

Credit: B. Nachman
https://indico.cern.ch/event/1188153/



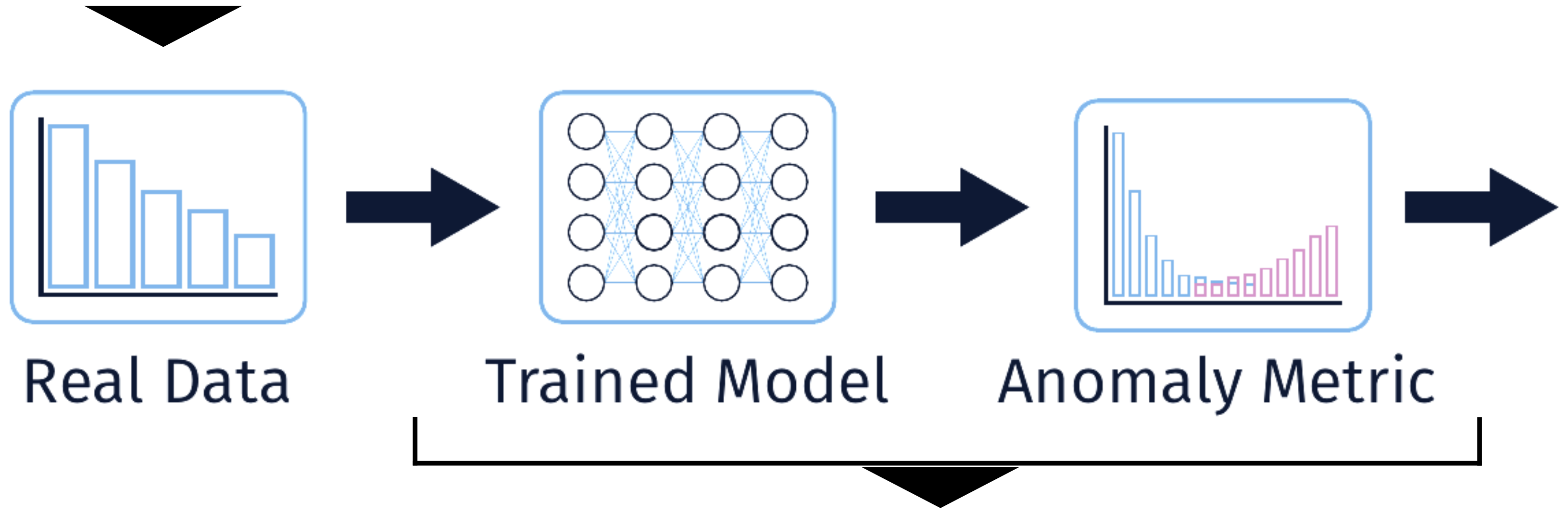
- Supervised = full label information
- Semi-supervised = partial labels
- Weakly-supervised = noisy labels
- Unsupervised = no labels
  - Example: autoencoders compress data and then uncompress it
  - Assumption: if x is far from $\mathrm{Decoder}(\mathrm{Encoder}(x))$, then x has low $p_{\mathrm{bkgd}}(x)$



ANOMALY DETECTION: GENERAL STRATEGY

Start from **data**

Anti-k_T jets with R = 0.8
Basic selection criteria



5 different methods

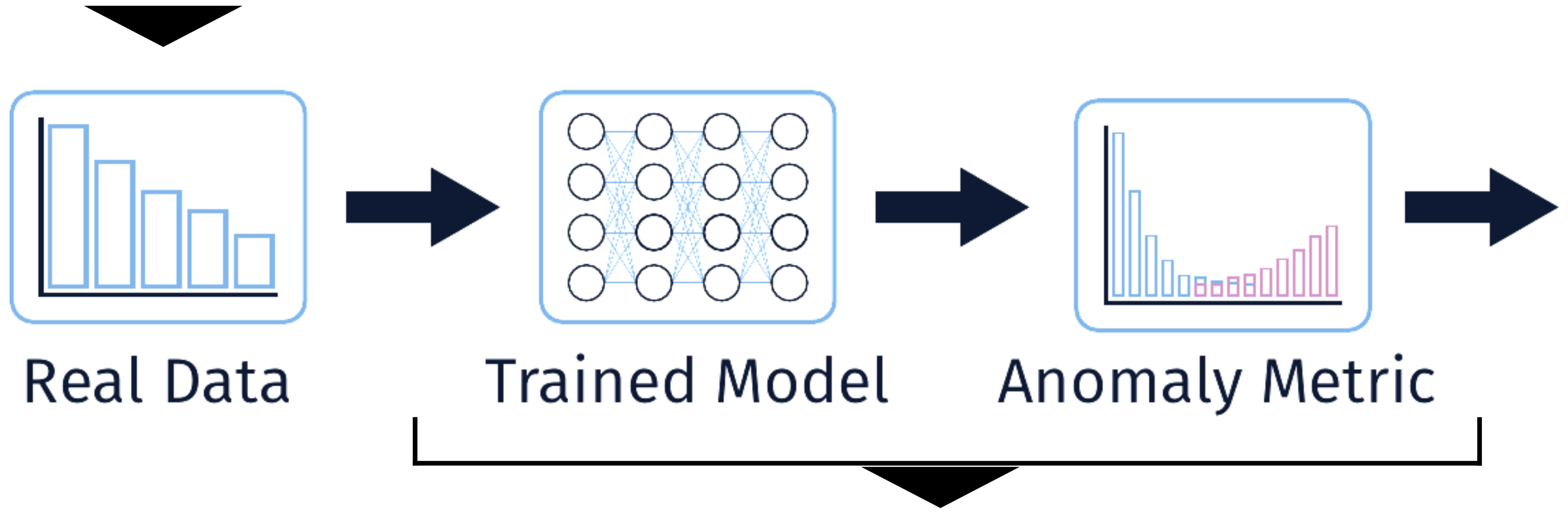
CWoLa Hunting / TNT /
CATHODE(-b) / VAE / QUAK

ANOMALY DETECTION: GENERAL STRATEGY

Start from **data**

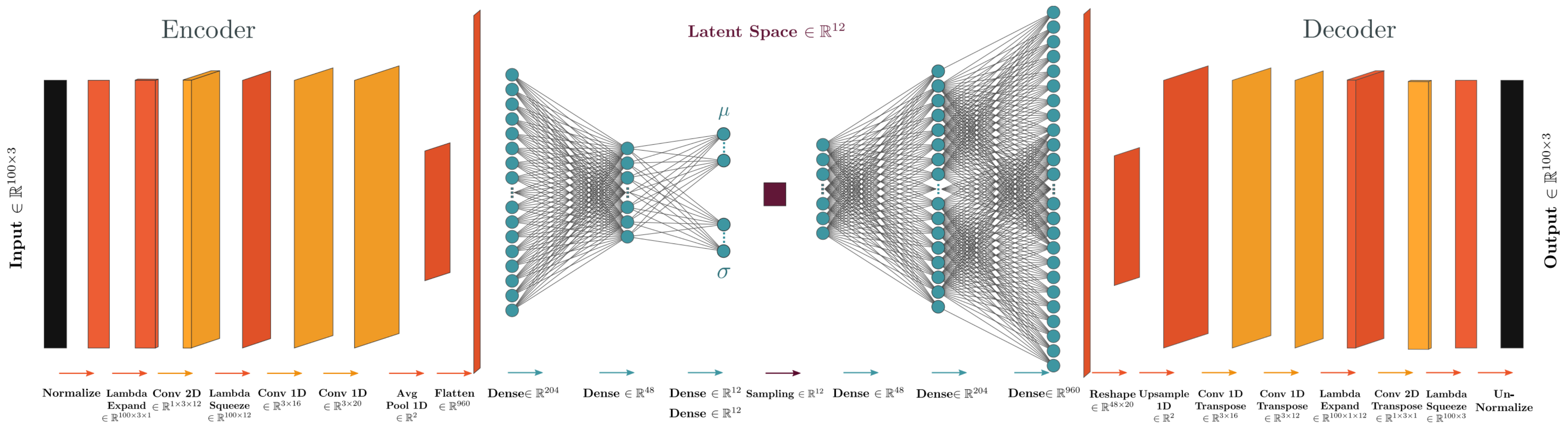
Anti-k_T jets with R = 0.8

Basic selection criteria



5 different methods

CWoLa Hunting / TNT /
CATHODE(-b) / VAE / QUAK



ANOMALY DETECTION: GENERAL STRATEGY

Start from **data**

Anti-k_T jets with R = 0.8
Basic selection criteria

Real Data → Trained Model → Anomaly Metric → Cut → SIG like / BG like

SIG like → Bump Hunt

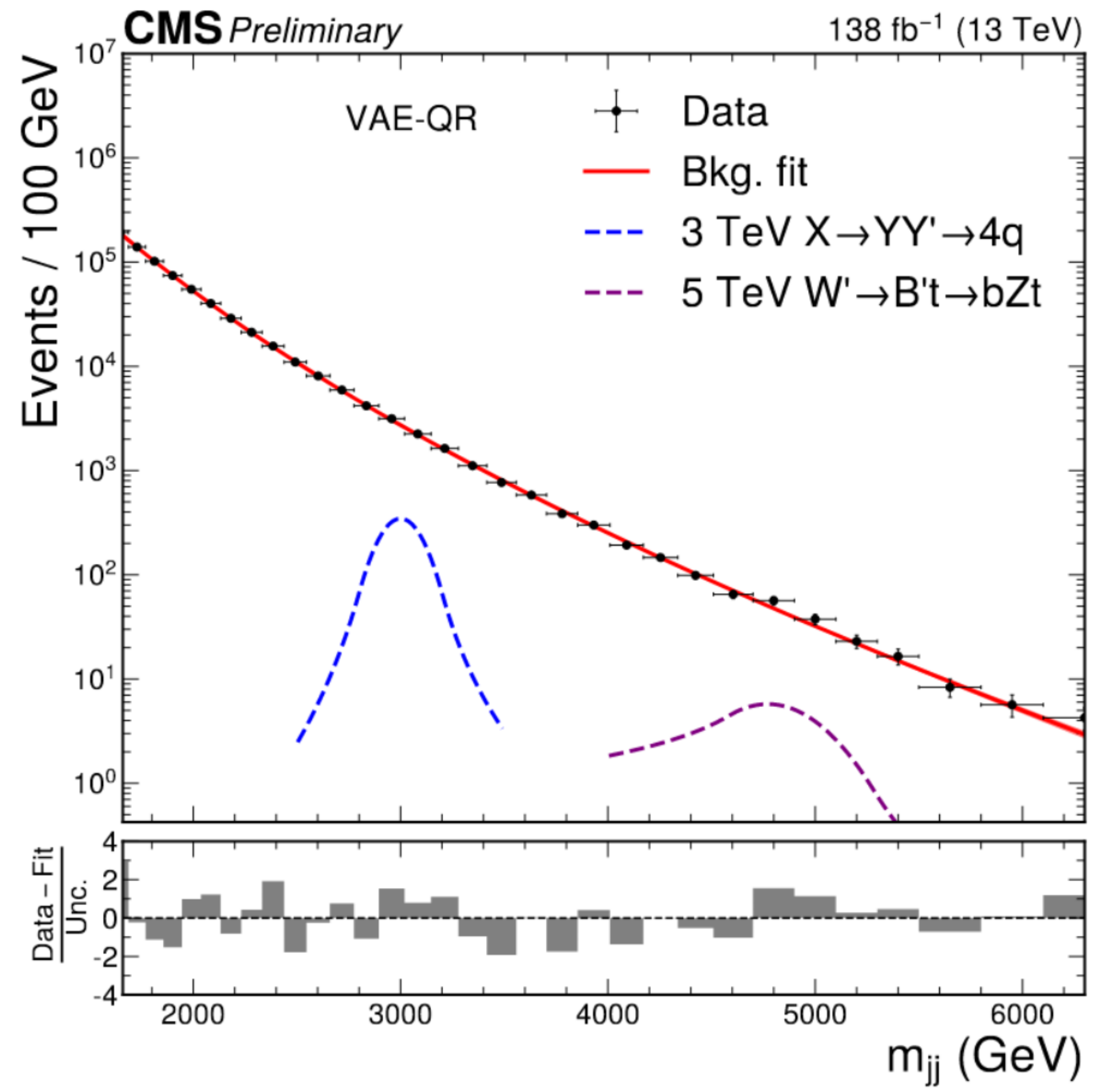
Keep ~**1%** most anomalous events

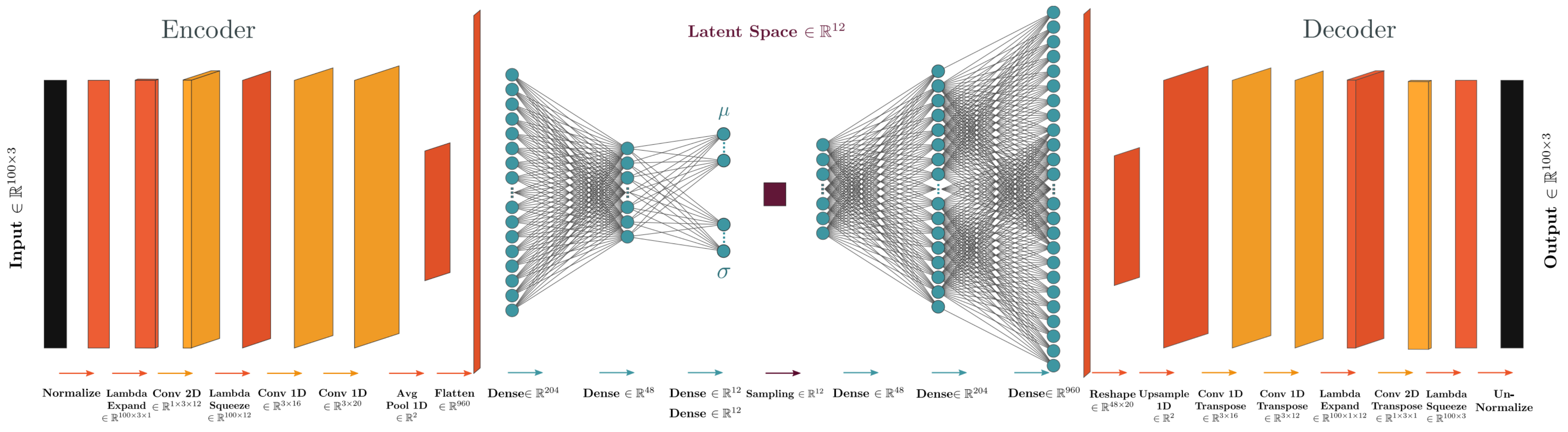
5 different methods

CWoLa Hunting / TNT / CATHODE(-b) / VAE / QUAK

Resonance?

Fit m_{jj} spectrum and obtain significance



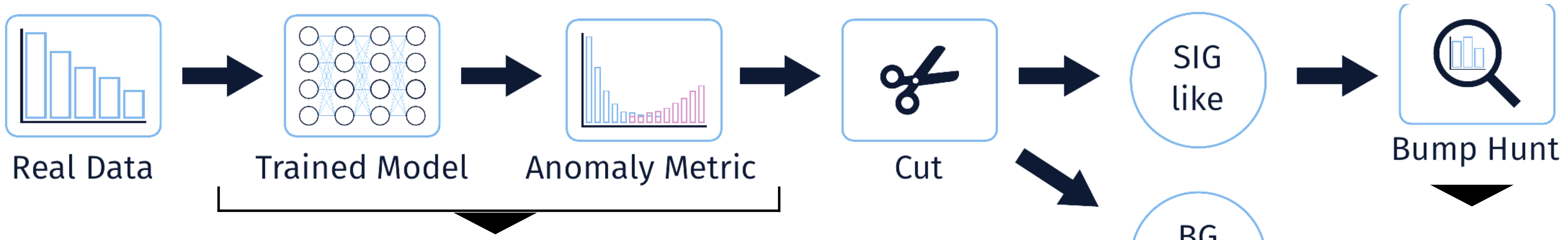


ANOMALY DETECTION: GENERAL STRATEGY

Start from **data**

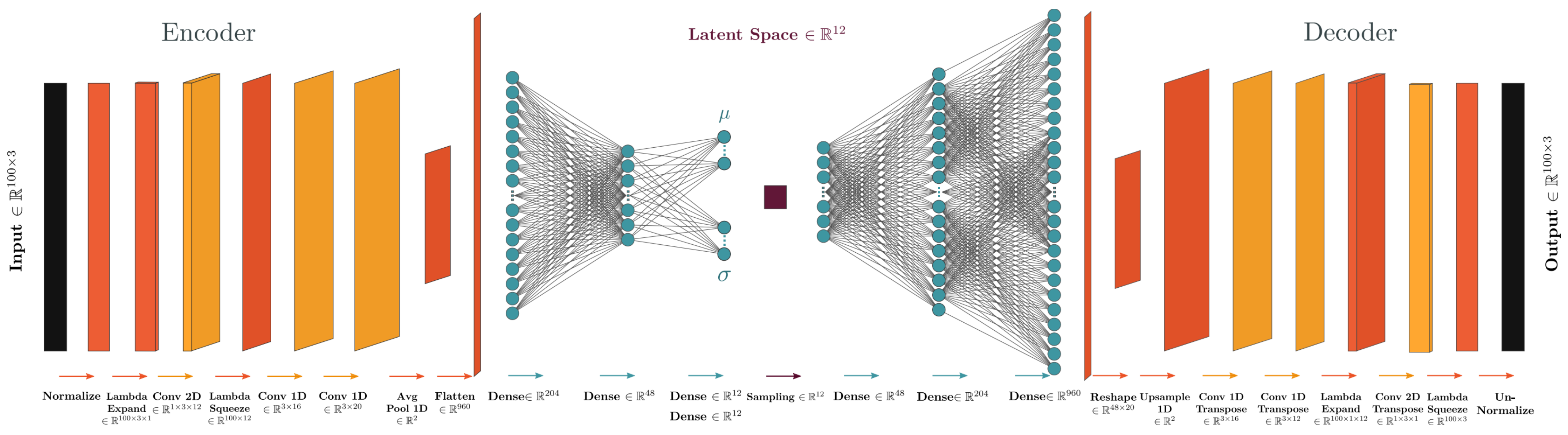
Anti-k_T jets with R = 0.8
Basic selection criteria

Keep ~**1%** most anomalous events



5 different methods

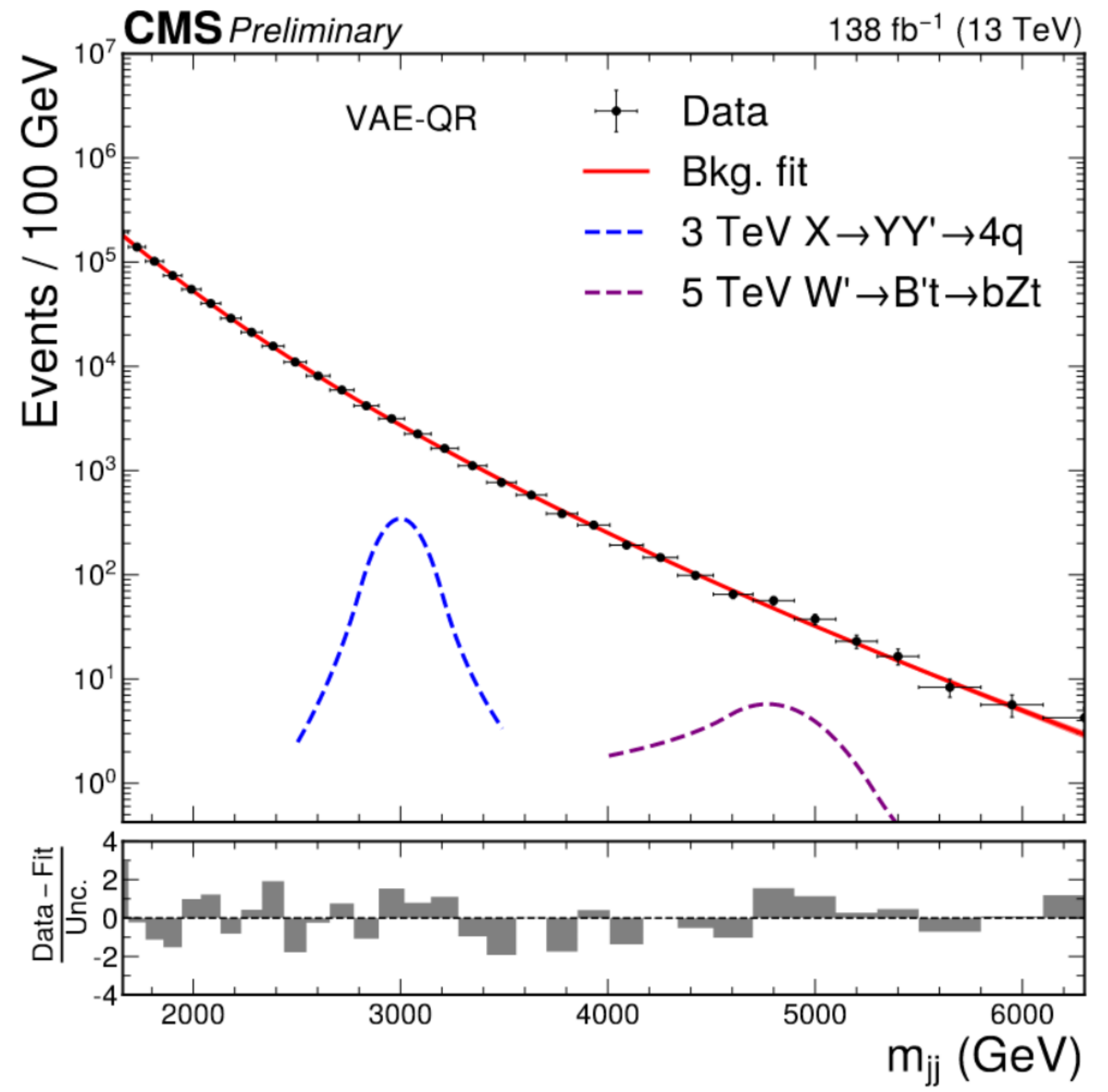
CWoLa Hunting / TNT / CATHODE(-b) / VAE / QUAK

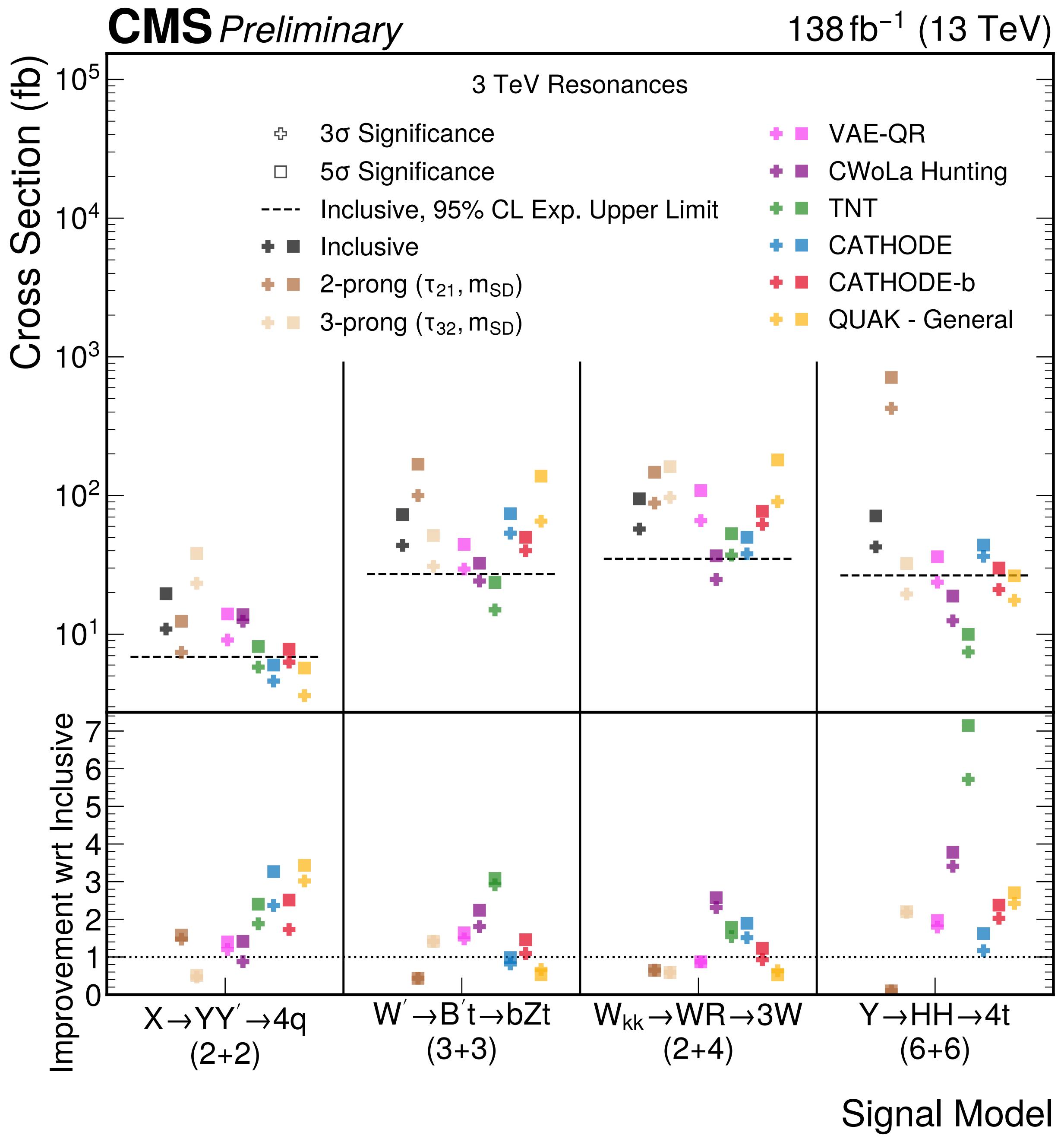


Resonance?

Fit m_{jj} spectrum and obtain significance

- Improvement in cross section needed for 5 σ discovery compared to inclusive dijet search by up to factor of 7

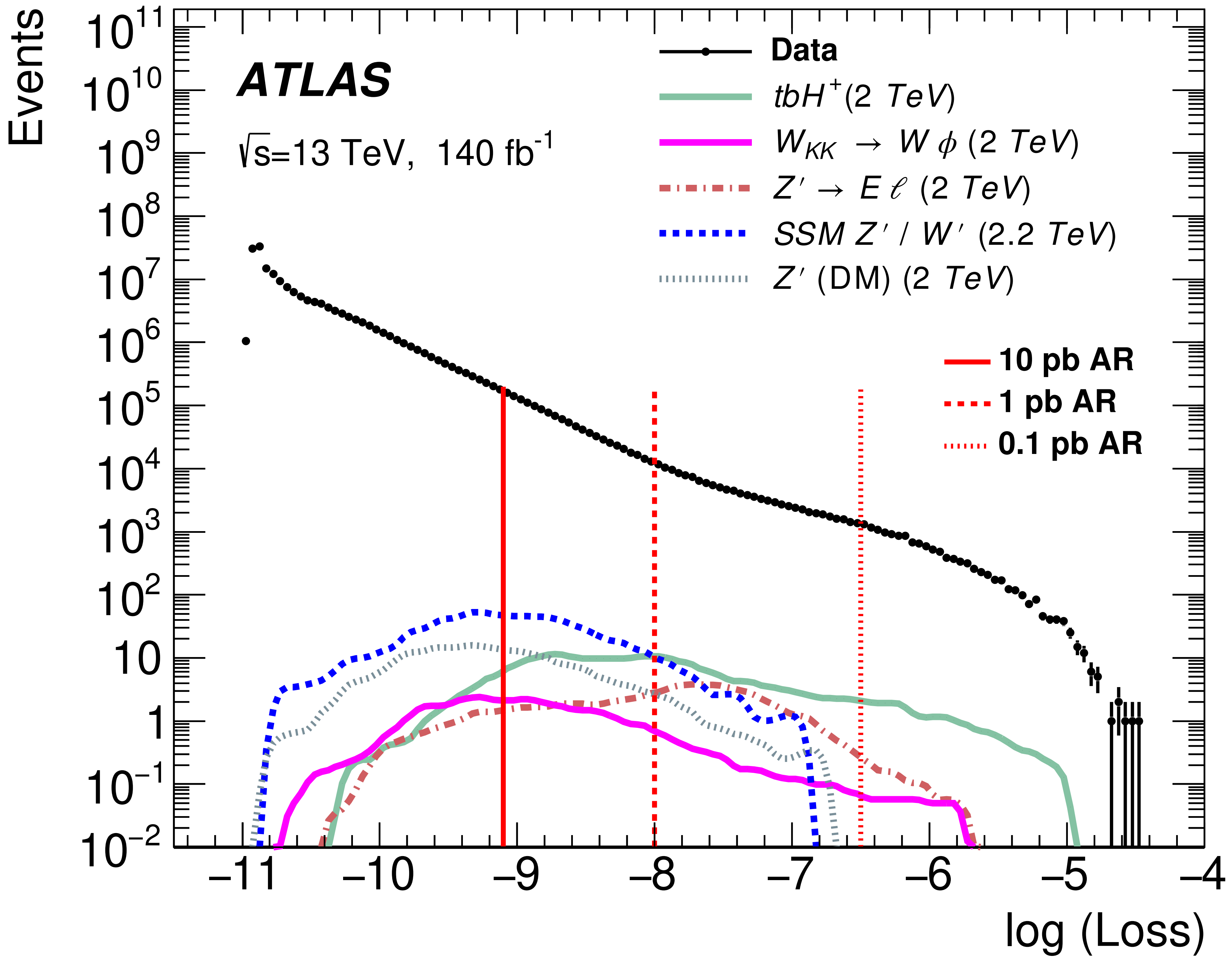




ANOMALY DETECTION: OBJECT PAIRS

- Autoencoder applied to physics-informed representation (rapidity-mass matrix)

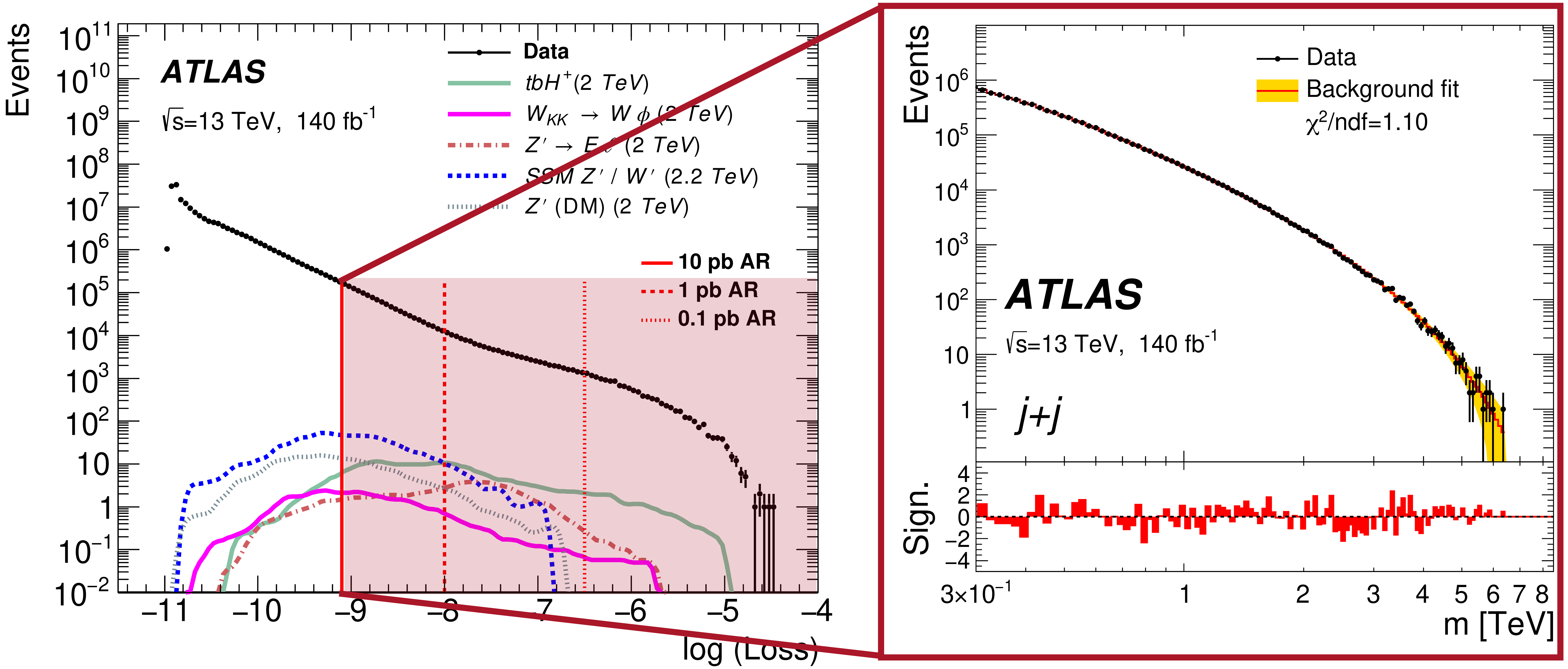
$$\begin{pmatrix}
e_T^{miss} & m_T(j_1) & m_T(j_2) & \ldots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \ldots m_T(\mu_N) \\
h_L(j_1) & e_T(j_1) & m(j_1,j_2) & \ldots m(j_1,j_N) & m(j_1,\mu_1) & m(j_1,\mu_2) & \ldots m(j_1,\mu_N) \\
h_L(j_2) & h(j_1,j_2) & \delta e_T(j_2) & \ldots m(j_2,j_N) & m(j_2,\mu_1) & m(j_2,\mu_2) & \ldots m(j_2,\mu_N) \\
\ldots & \ldots & \ldots & \ldots, & \ldots & \ldots & \ldots \\
h_L(j_N) & h(j_1,j_N) & \ldots & \ldots \delta e_T(j_N) & m(j_N,\mu_1) & m(j_N,\mu_2) & \ldots m(j_N,\mu_N) \\
h_L(\mu_1) & h(\mu_1,j_1) & h(\mu_1,j_2) & \ldots h(\mu_1,j_N) & e_T(\mu_1) & m(\mu_1,\mu_2) & m(\mu_1,\mu_N) \\
h_L(\mu_2) & h(\mu_2,j_1) & h(\mu_1,j_2) & \ldots h(\mu_2,j_N) & h(\mu_1,\mu_2) & \delta e_T(\mu_2) & m(\mu_2,\mu_N) \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\
h_L(\mu_N) & h(\mu_N,j_1) & h(\mu_N,j_2) & \ldots h(\mu_N,j_N) & h(\mu_N,\mu_1) & h(\mu_N,\mu_2) & \delta e_T(\mu_N)
\end{pmatrix}$$



ANOMALY DETECTION: OBJECT PAIRS

- Autoencoder applied to physics-informed representation (rapidity-mass matrix)
- Anomaly score selection corresponding to 10 pb

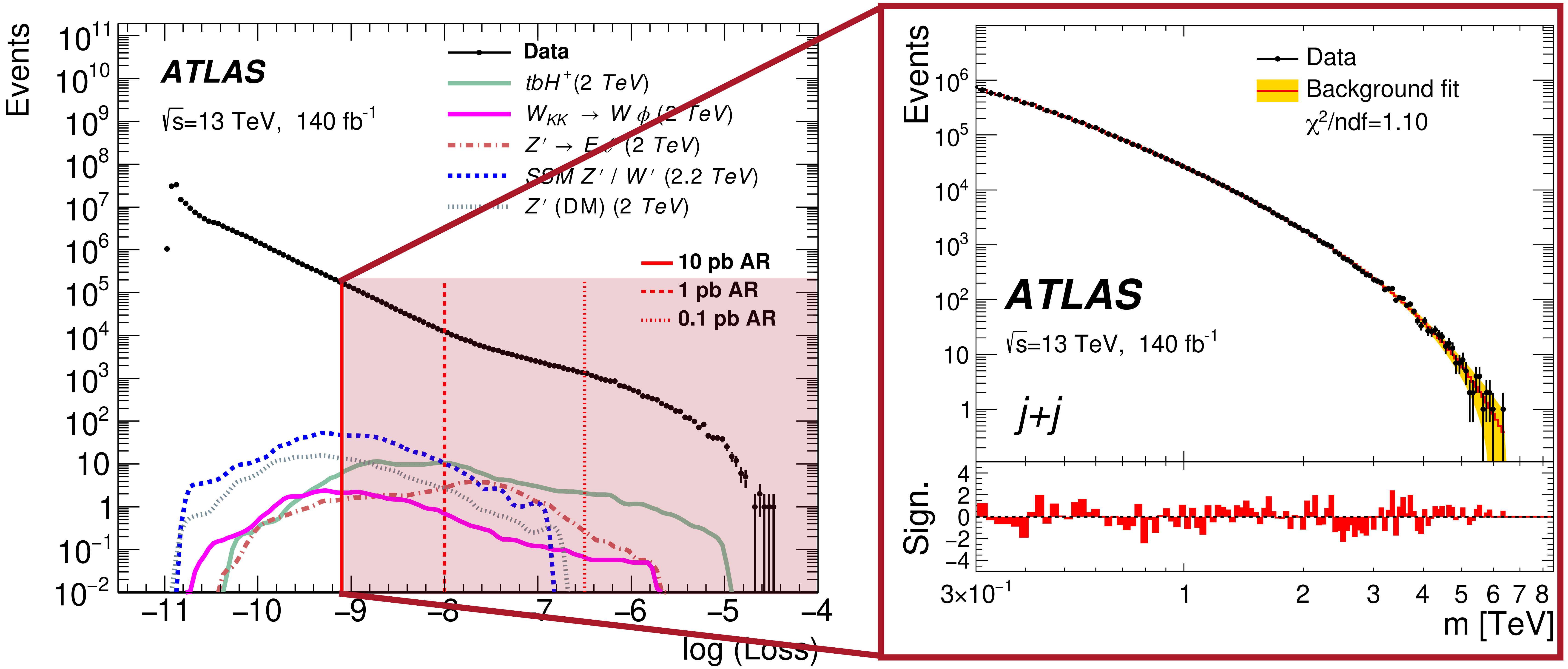
$$\begin{pmatrix}
e_T^{miss} & m_T(j_1) & m_T(j_2) & \dots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \dots m_T(\mu_N) \\
h_L(j_1) & e_T(j_1) & m(j_1,j_2) & \dots m(j_1,j_N) & m(j_1,\mu_1) & m(j_1,\mu_2) & \dots m(j_1,\mu_N) \\
h_L(j_2) & h(j_1,j_2) & \delta e_T(j_2) & \dots m(j_2,j_N) & m(j_2,\mu_1) & m(j_2,\mu_2) & \dots m(j_2,\mu_N) \\
\dots & \dots & \dots & \dots, & \dots & \dots & \dots \\
h_L(j_N) & h(j_1,j_N) & \dots & \dots \delta e_T(j_N) & m(j_N,\mu_1) & m(j_N,\mu_2) & \dots m(j_N,\mu_N) \\
h_L(\mu_1) & h(\mu_1,j_1) & h(\mu_1,j_2) & \dots h(\mu_1,j_N) & e_T(\mu_1) & m(\mu_1,\mu_2) & m(\mu_1,\mu_N) \\
h_L(\mu_2) & h(\mu_2,j_1) & h(\mu_1,j_2) & \dots h(\mu_2,j_N) & h(\mu_1,\mu_2) & \delta e_T(\mu_2) & m(\mu_2,\mu_N) \\
\dots & \dots & \dots & \dots & \dots & \dots & \\
h_L(\mu_N) & h(\mu_N,j_1) & h(\mu_N,j_2) & \dots h(\mu_N,j_N) & h(\mu_N,\mu_1) & h(\mu_N,\mu_2) & \delta e_T(\mu_N)
\end{pmatrix}$$

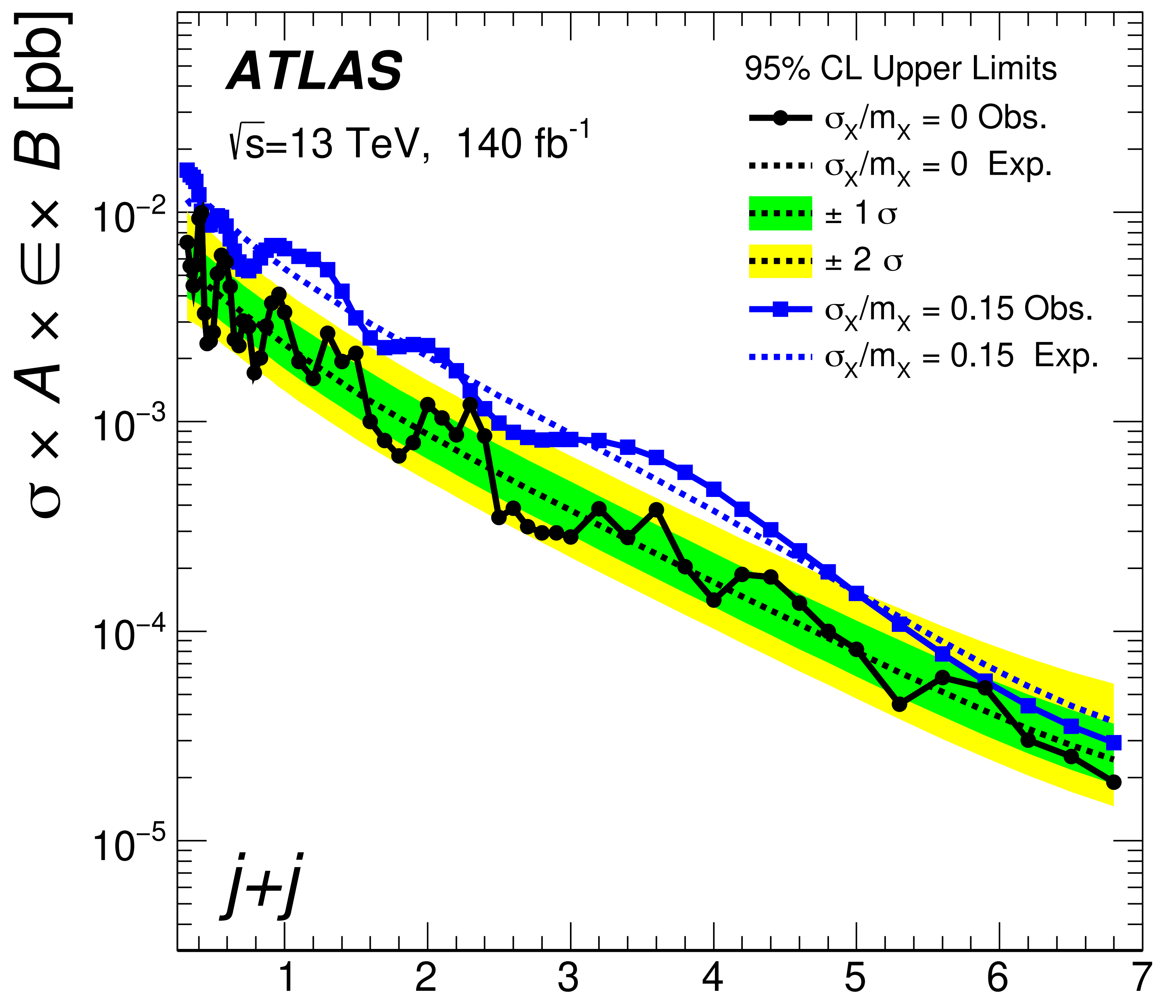


ANOMALY DETECTION: OBJECT PAIRS

- Autoencoder applied to physics-informed representation (rapidity-mass matrix)
- Anomaly score selection corresponding to 10 pb
- Model-independent limits placed in 9 final states: **j+j**, j+b, b+b, j+e, b+e, j+γ, j+μ, b+μ, and b+γ

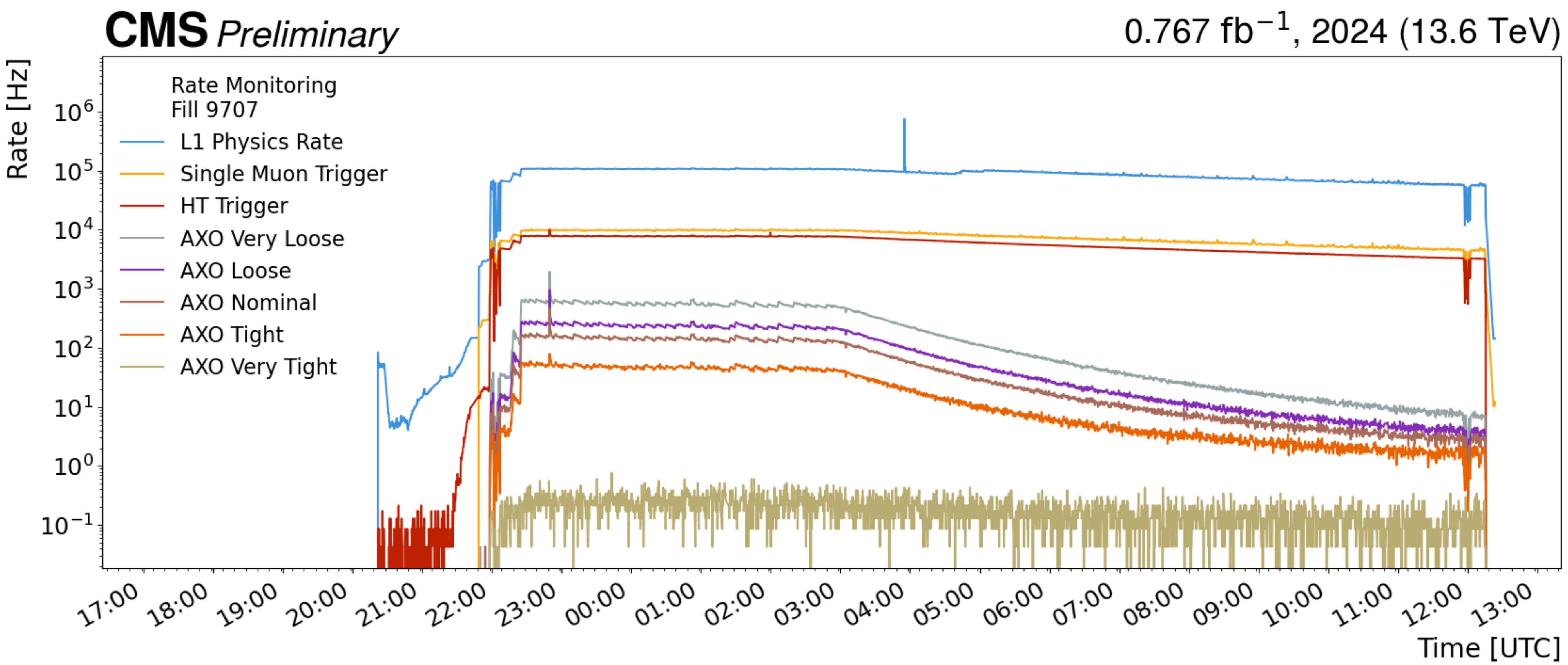
$$
\begin{pmatrix}
e_T^{miss} & m_T(j_1) & m_T(j_2) & \dots m_T(j_N) & m_T(\mu_1) & m_T(\mu_2) & \dots m_T(\mu_N) \\
h_L(j_1) & e_T(j_1) & m(j_1,j_2) & \dots m(j_1,j_N) & m(j_1,\mu_1) & m(j_1,\mu_2) & \dots m(j_1,\mu_N) \\
h_L(j_2) & h(j_1,j_2) & \delta e_T(j_2) & \dots m(j_2,j_N) & m(j_2,\mu_1) & m(j_2,\mu_2) & \dots m(j_2,\mu_N) \\
\dots & \dots & \dots & \dots, & \dots & \dots & \dots \\
h_L(j_N) & h(j_1,j_N) & \dots & \dots \delta e_T(j_N) & m(j_N,\mu_1) & m(j_N,\mu_2) & \dots m(j_N,\mu_N) \\
h_L(\mu_1) & h(\mu_1,j_1) & h(\mu_1,j_2) & \dots h(\mu_1,j_N) & e_T(\mu_1) & m(\mu_1,\mu_2) & m(\mu_1,\mu_N) \\
h_L(\mu_2) & h(\mu_2,j_1) & h(\mu_1,j_2) & \dots h(\mu_2,j_N) & h(\mu_1,\mu_2) & \delta e_T(\mu_2) & m(\mu_2,\mu_N) \\
\dots & \dots & \dots & \dots & \dots & \dots & \\
h_L(\mu_N) & h(\mu_N,j_1) & h(\mu_N,j_2) & \dots h(\mu_N,j_N) & h(\mu_N,\mu_1) & h(\mu_N,\mu_2) & \delta e_T(\mu_N)
\end{pmatrix}
$$





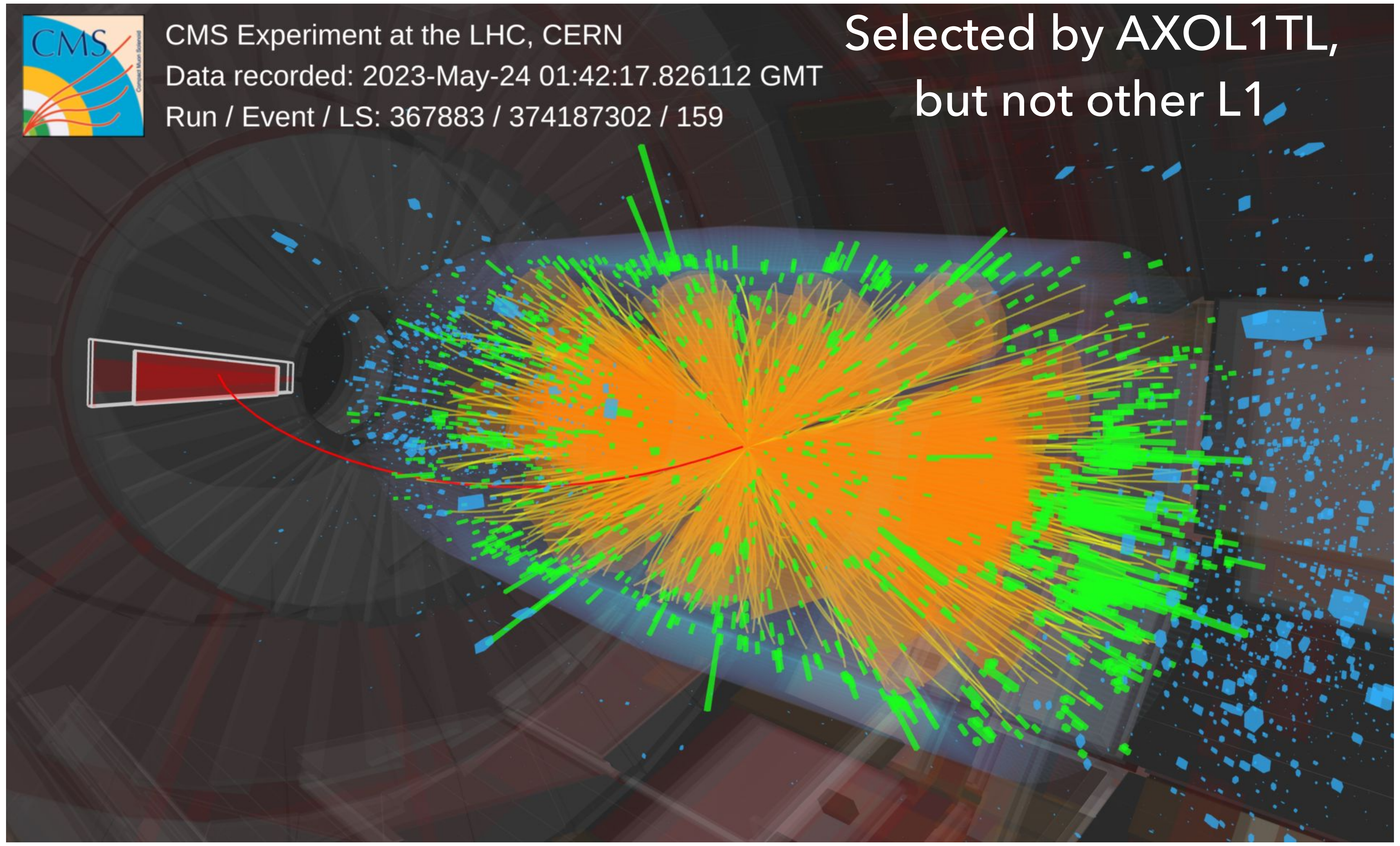
ANOMALY DETECTION @ LEVEL-1 TRIGGER

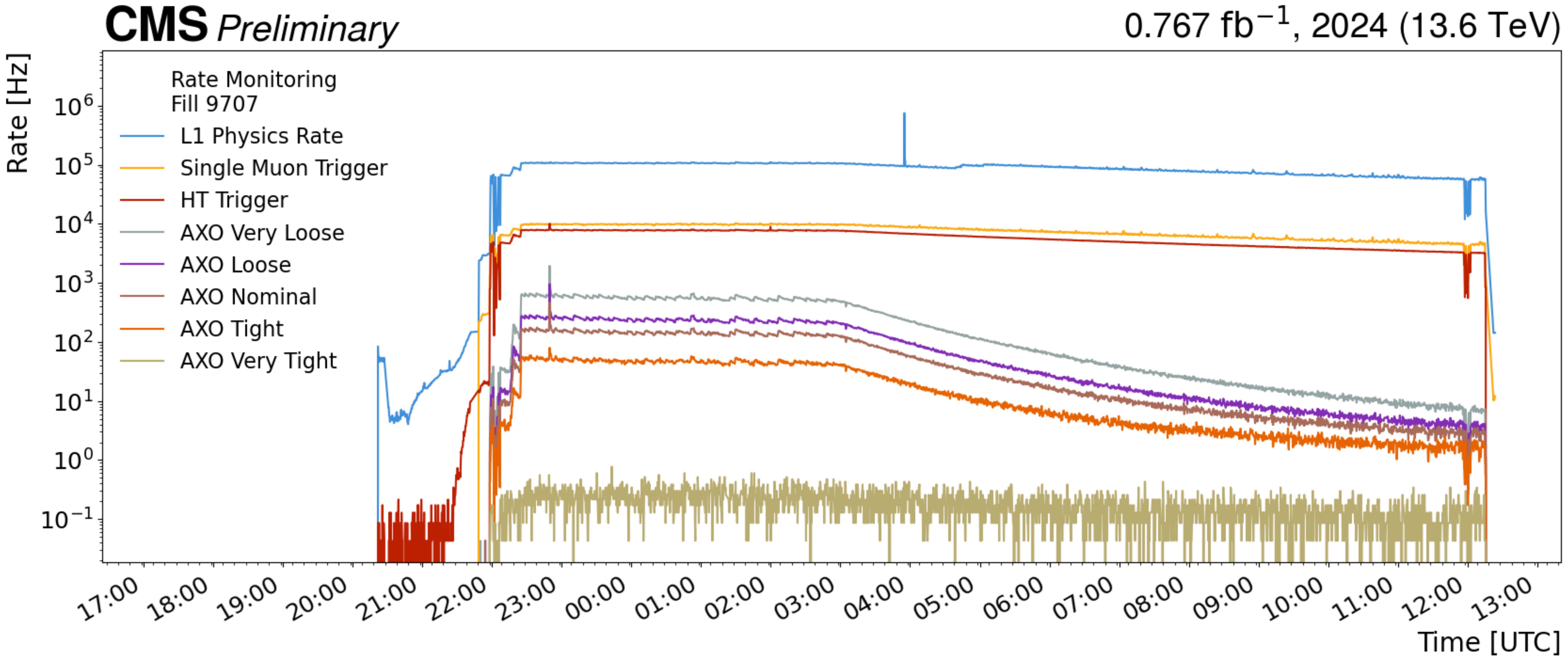
- AXOL1TL anomaly detection algorithms for the level-1 trigger based on a variational autoencoder
- AXOL1TL rates and score distributions for 2024 data taking

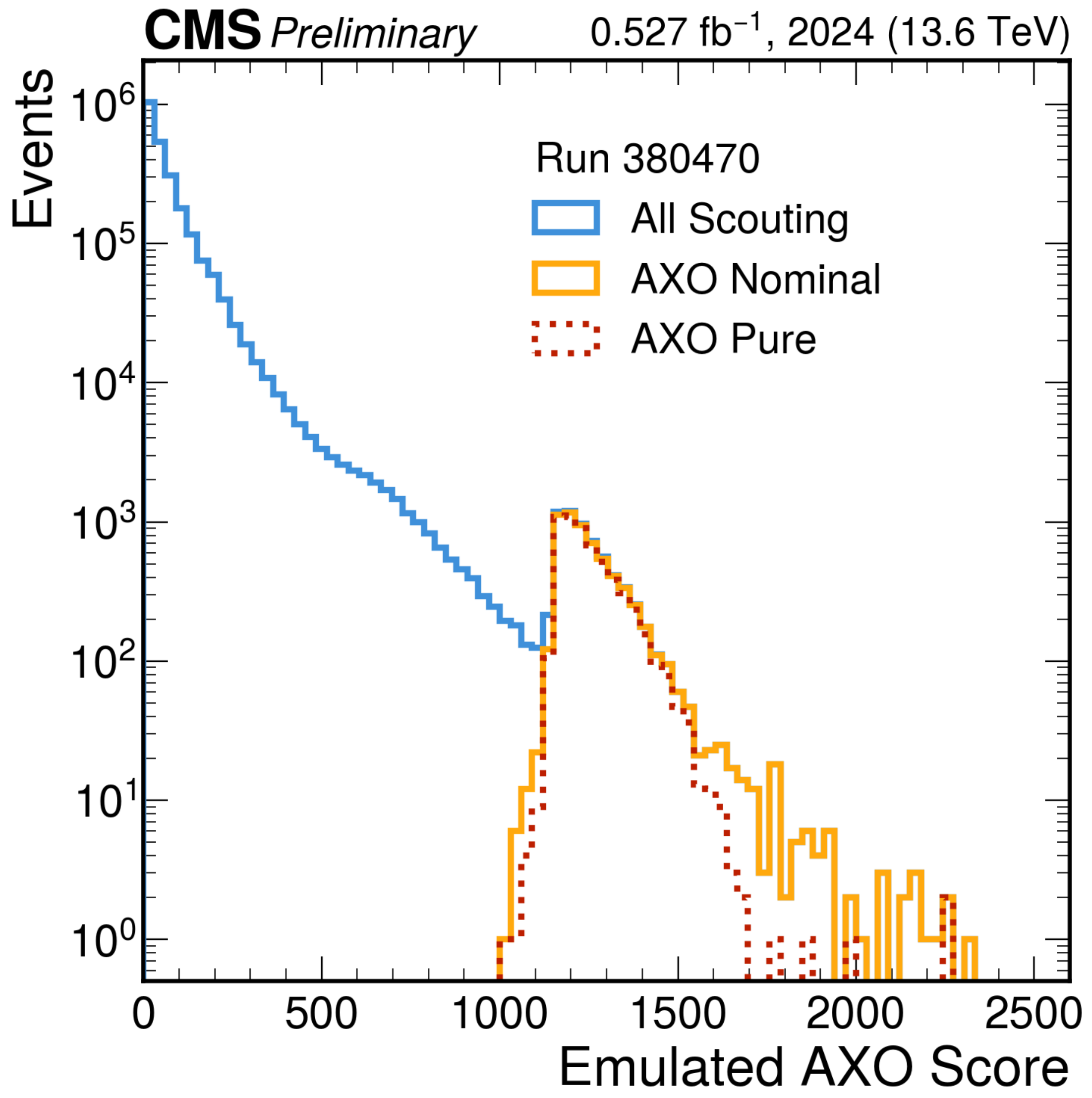


ANOMALY DETECTION @ LEVEL-1 TRIGGER

- AXOL1TL anomaly detection algorithms for the level-1 trigger based on a variational autoencoder
- AXOL1TL rates and score distributions for 2024 data taking
- Pure contribution shows AXOLT1TL selects unique events relative to existing level-1 trigger



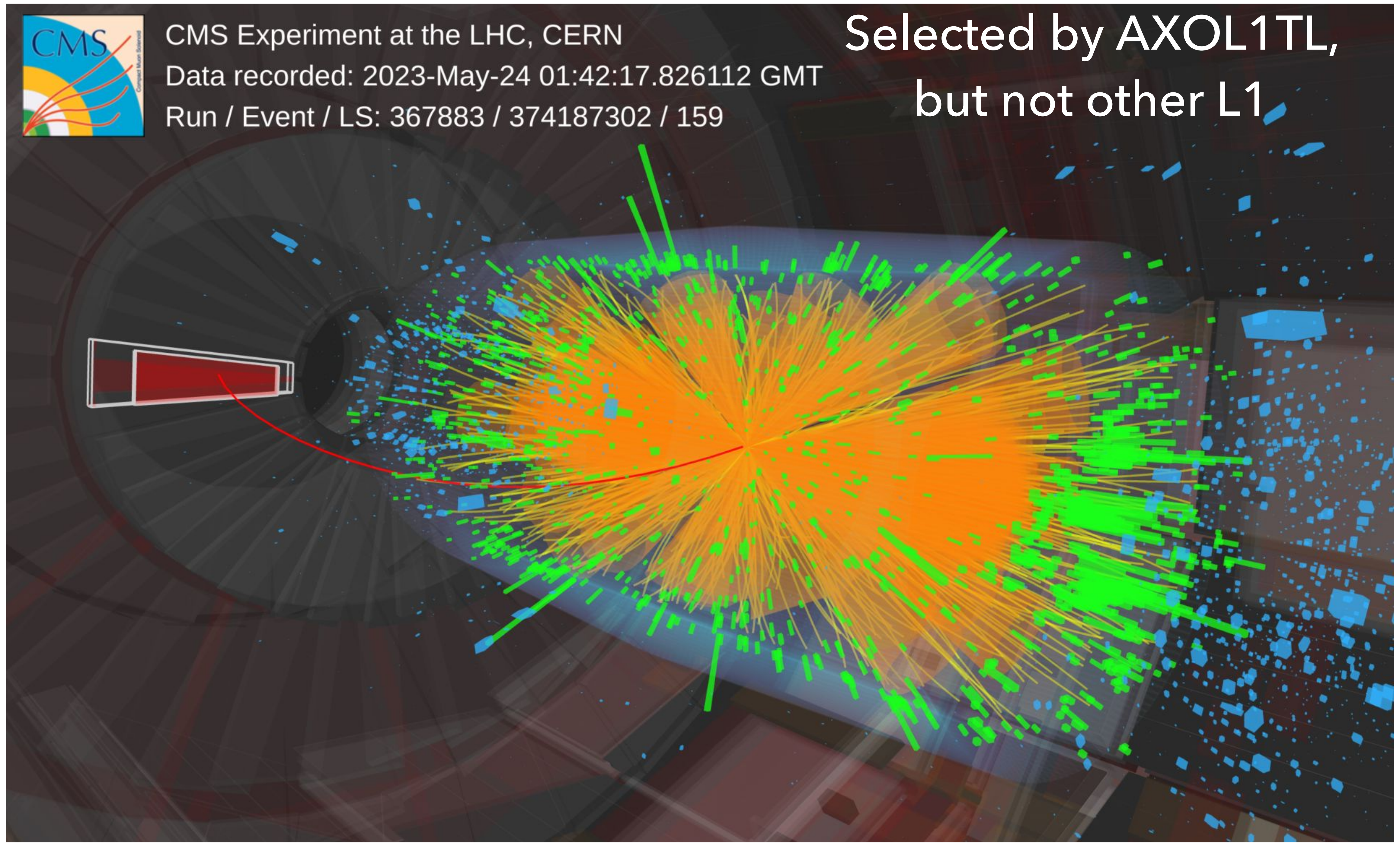


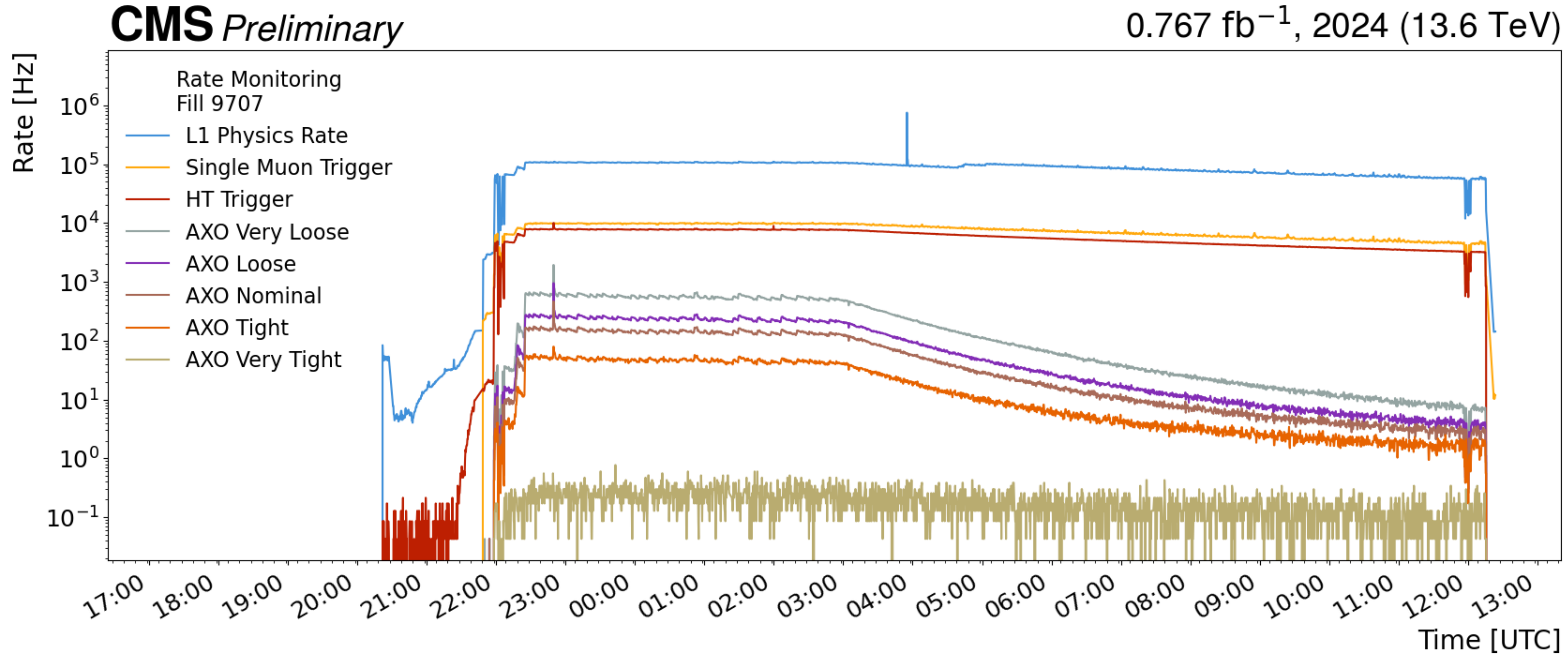


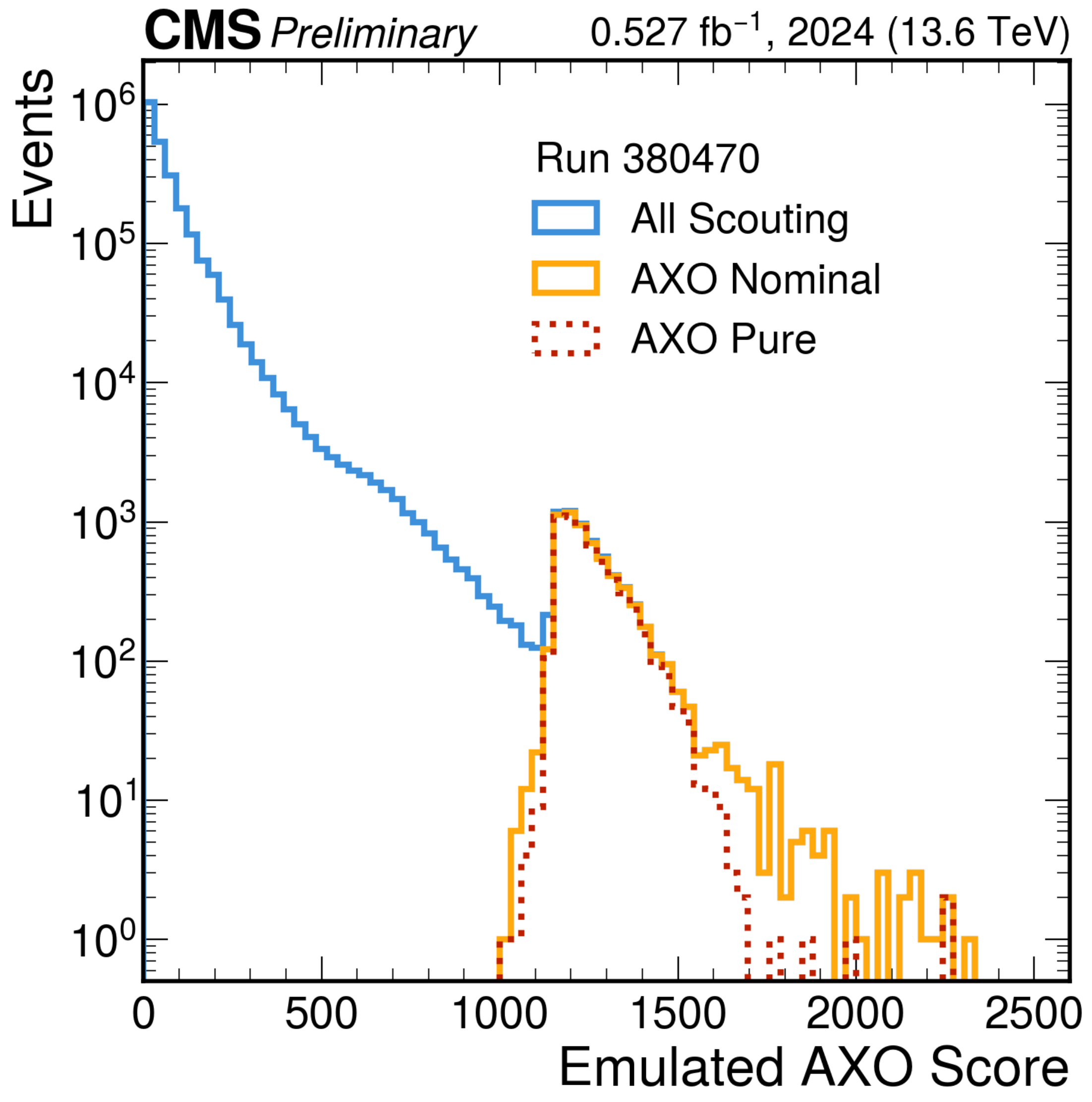
ANOMALY DETECTION @ LEVEL-1 TRIGGER

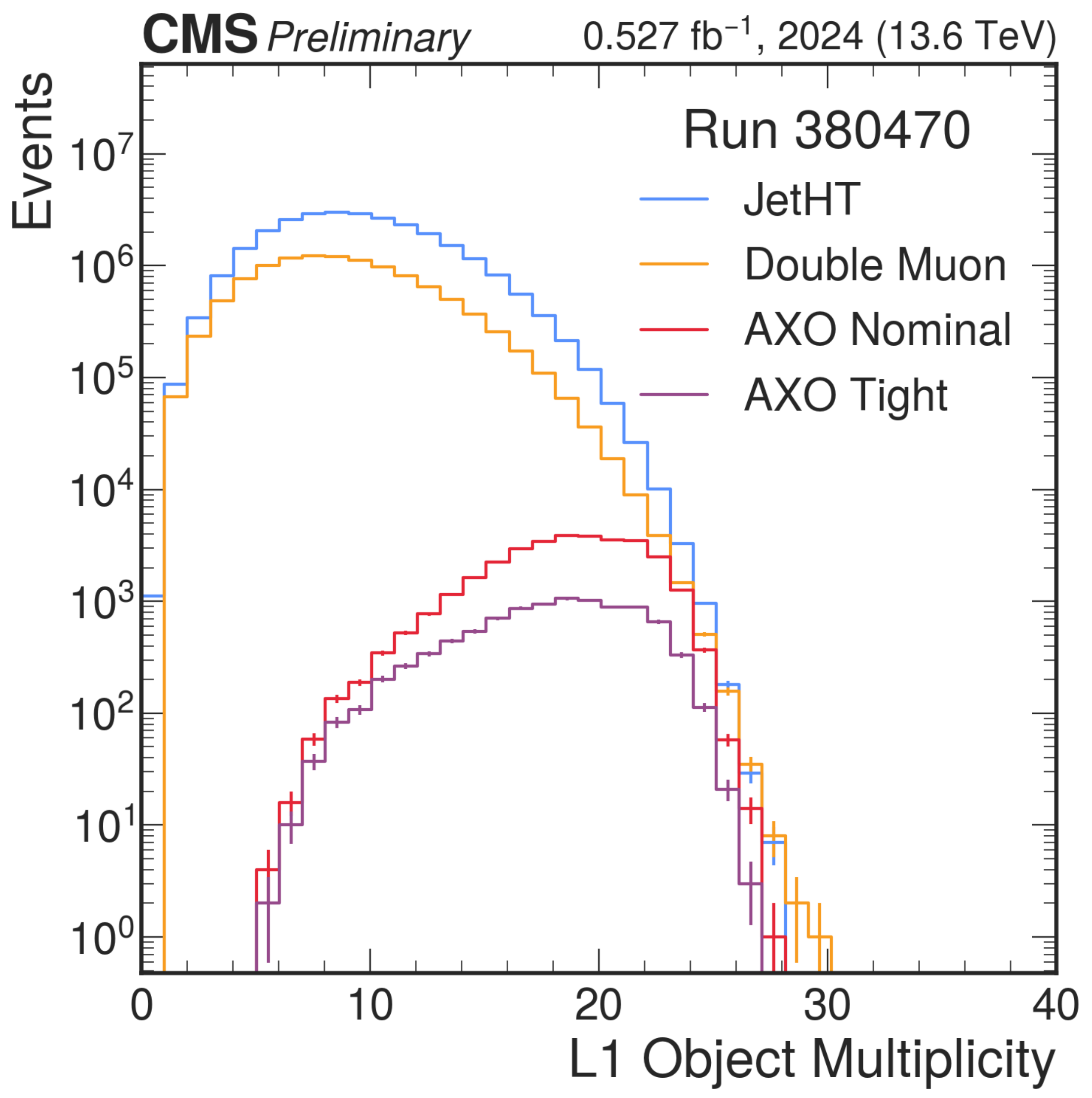
CMS-DP-2023-079
CMS-DP-2024-XXX

- AXOL1TL anomaly detection algorithms for the level-1 trigger based on a variational autoencoder
- AXOL1TL rates and score distributions for 2024 data taking
- Pure contribution shows AXOLT1TL selects unique events relative to existing level-1 trigger
- Preference for high multiplicity events



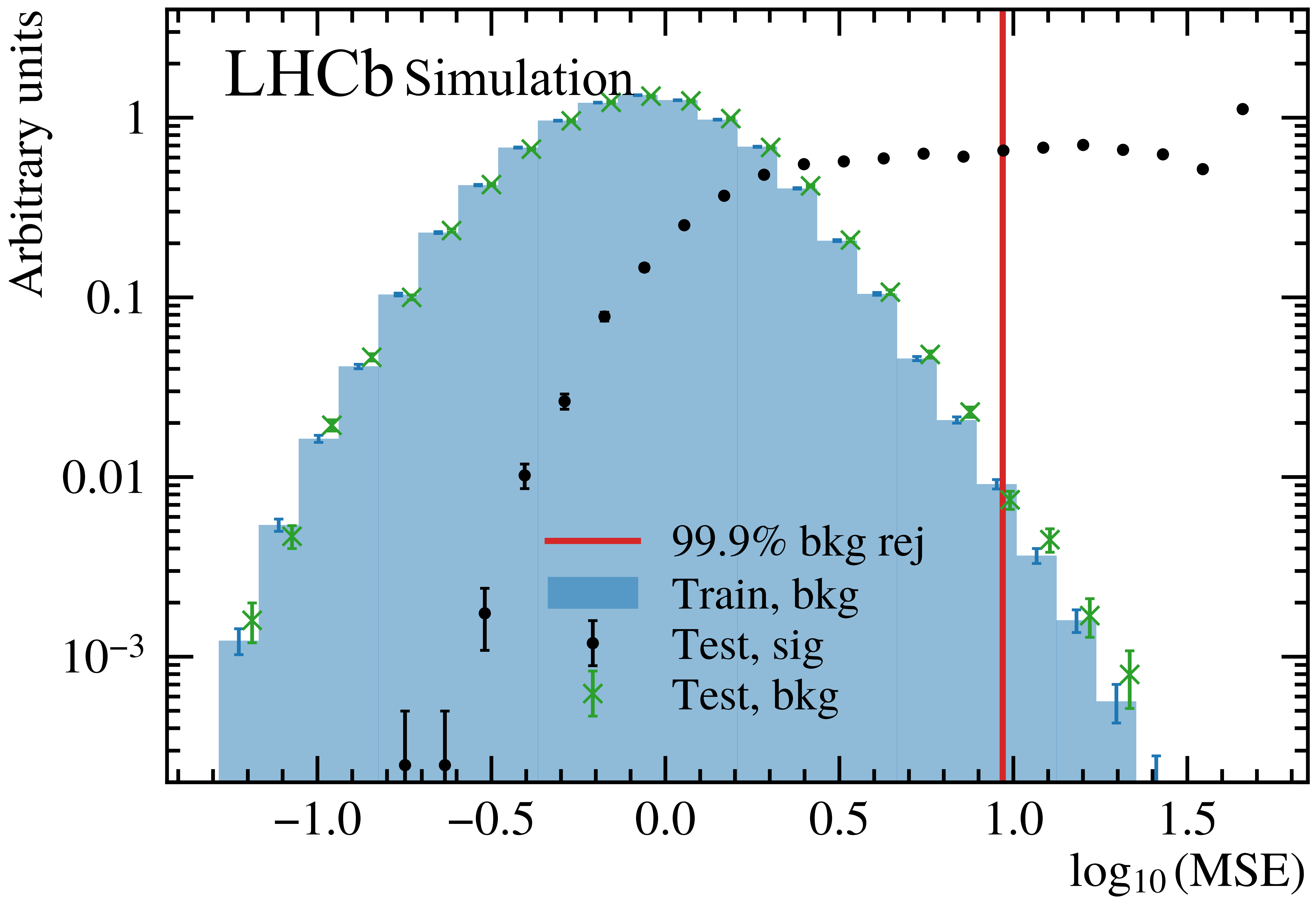






NORMALIZED AUTOENCODER

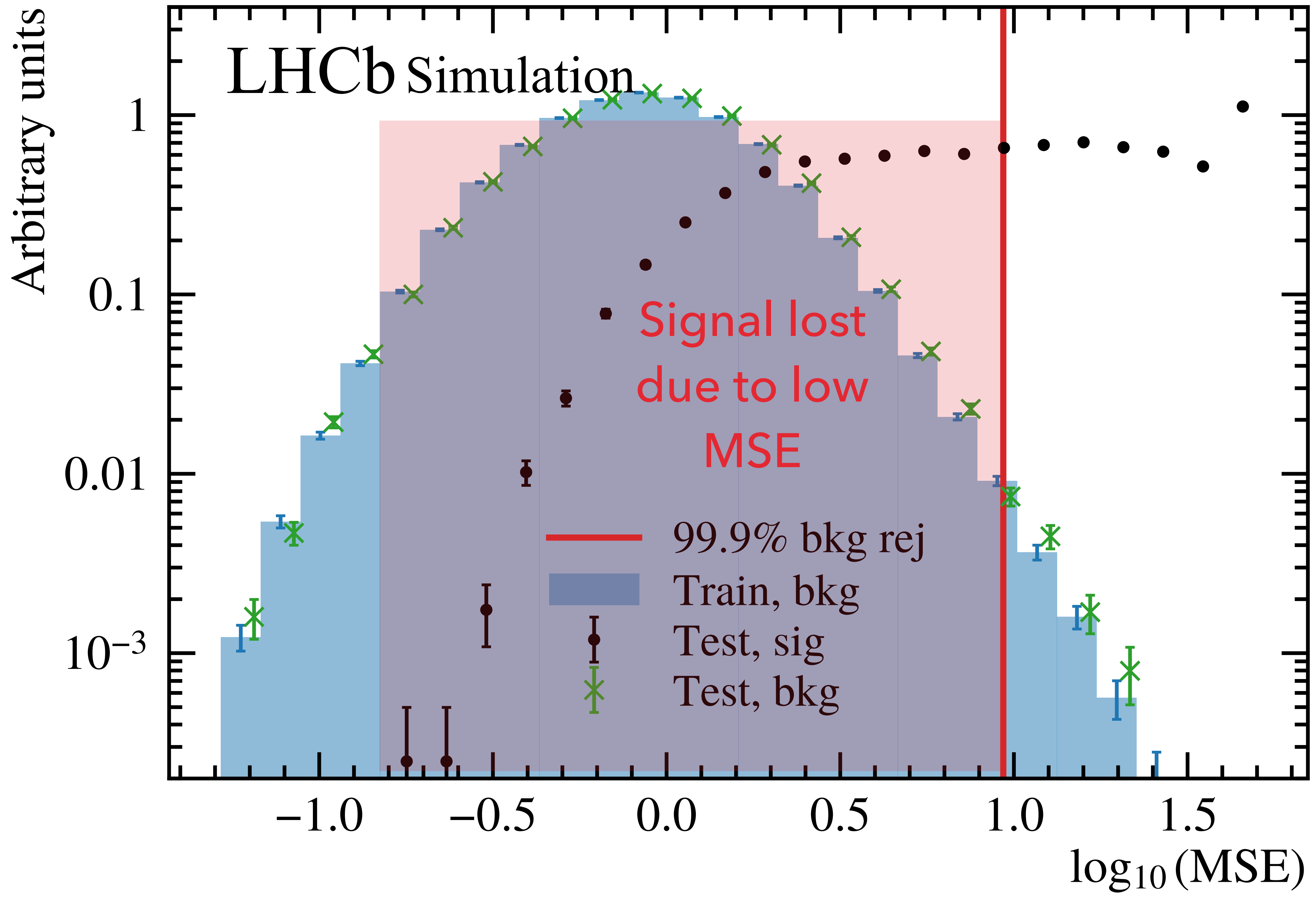
- Autoencoders can be too good at reconstructing signal, meaning signal is not flagged as anomalous (high reconstruction error)
- Normalized autoencoder approach designed to mitigate this issue—align low reconstruction error phase space with background phase space



NORMALIZED AUTOENCODER

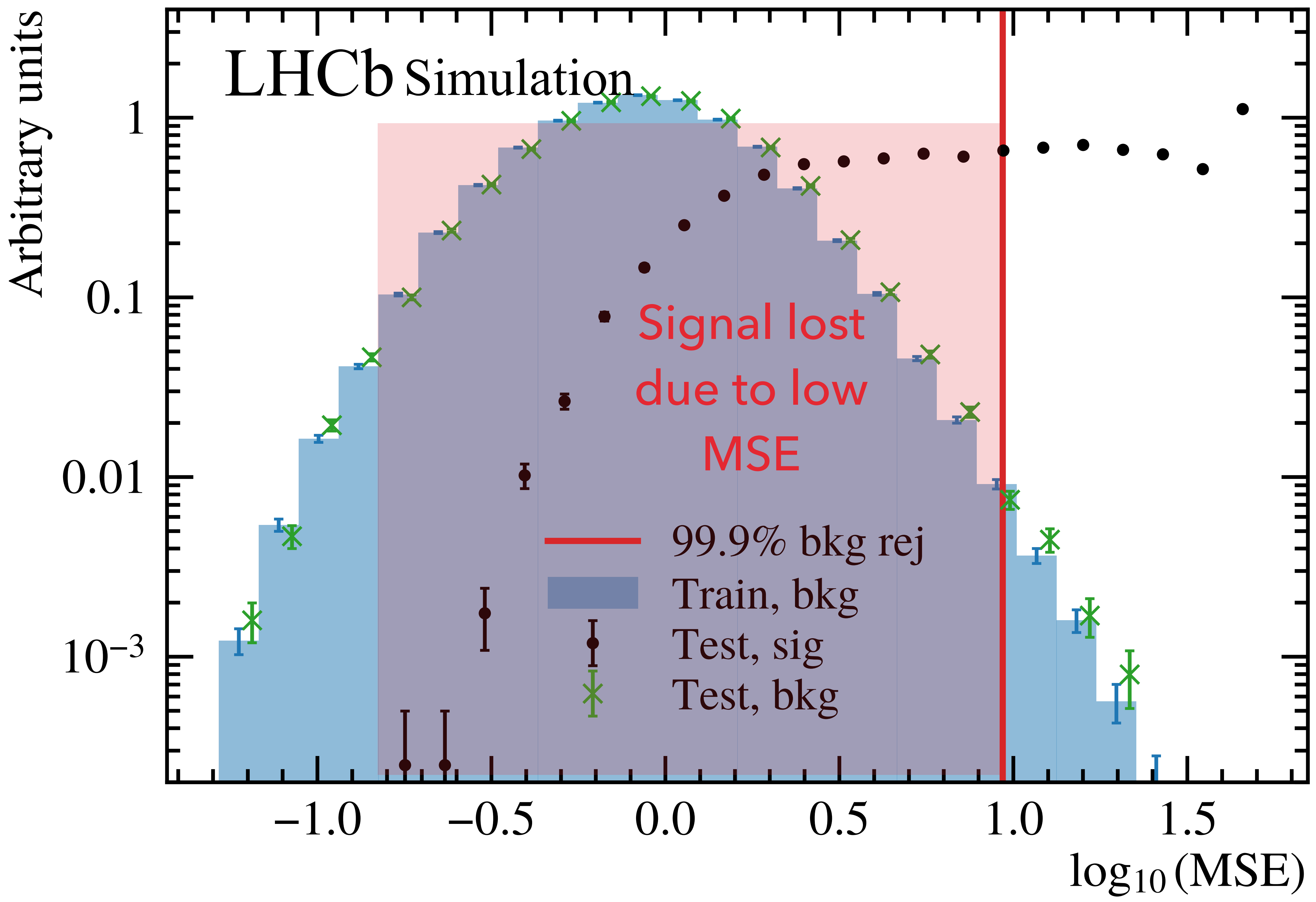
LHCB-FIGURE-2024-015

- Autoencoders can be too good at reconstructing signal, meaning signal is not flagged as anomalous (high reconstruction error)
- Normalized autoencoder approach designed to mitigate this issue—align low reconstruction error phase space with background phase space



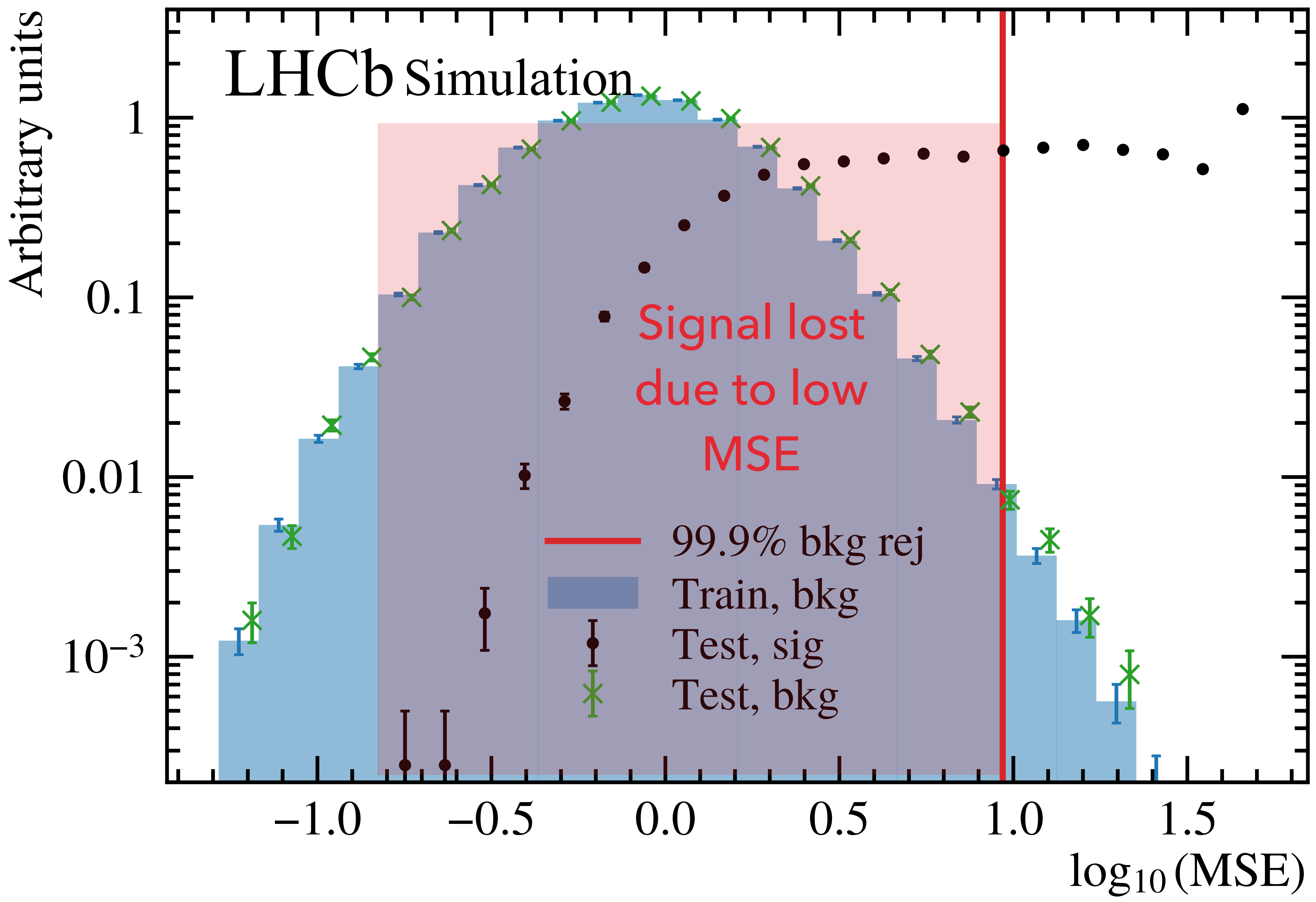
NORMALIZED AUTOENCODER

- Model data as $p_\theta(x) \propto \exp\left[-\mathrm{MSE}\left(x, \mathrm{AE}_\theta(x)\right)\right]$ and minimize $-\ln p_\theta(x)$



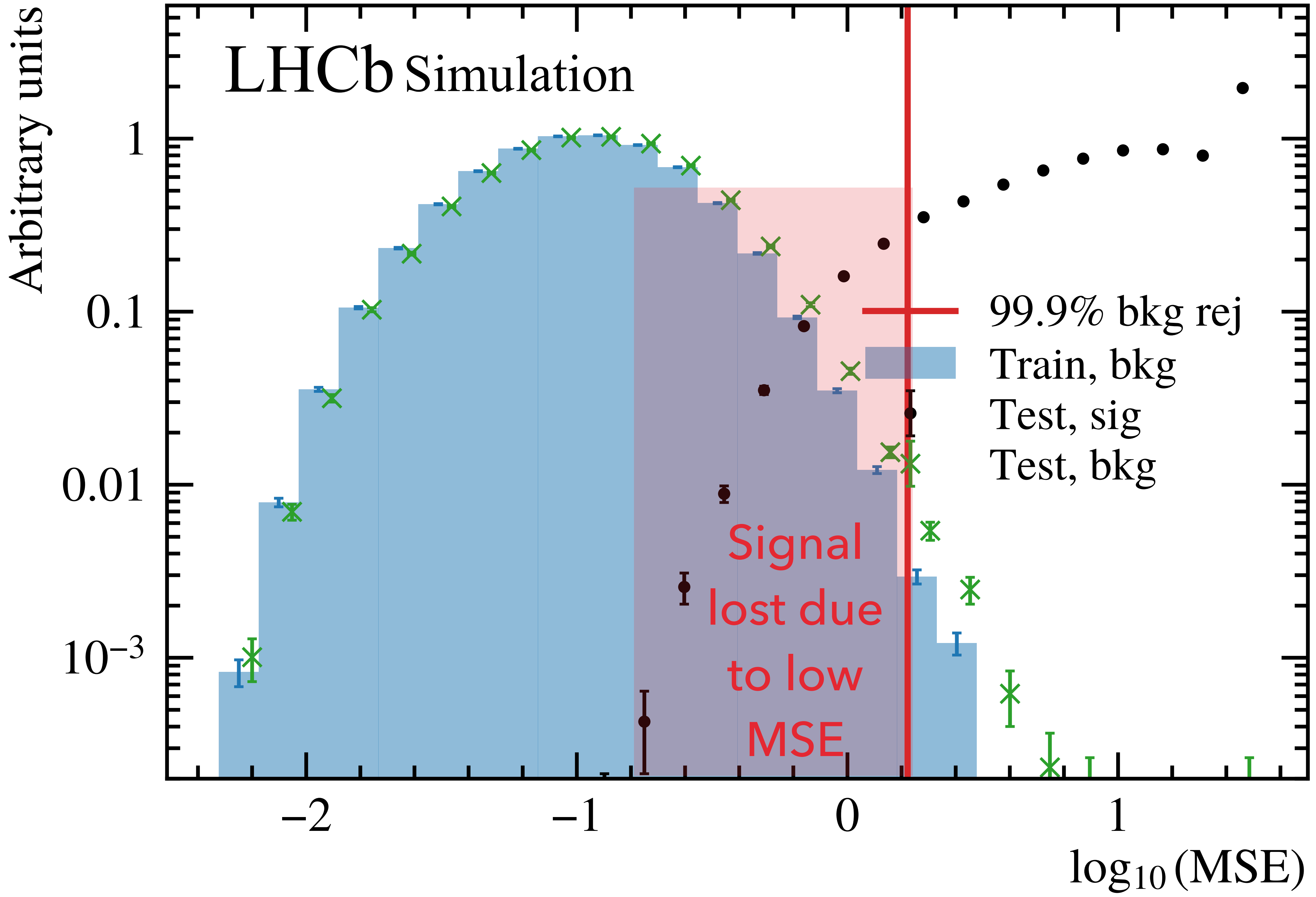
NORMALIZED AUTOENCODER

- Model data as $p_\theta(x) \propto \exp\left[-\mathrm{MSE}\left(x, \mathrm{AE}_\theta(x)\right)\right]$ and minimize $-\ln p_\theta(x)$
- In practice, approximately minimize this through clever reframing
  - Sample "negative" samples from $p_\theta(x)$ using MCMC
  - Sample "positive" samples from $p_{\mathrm{data}}(x)$
  - Minimize difference in expected MSE between negative and positive samples

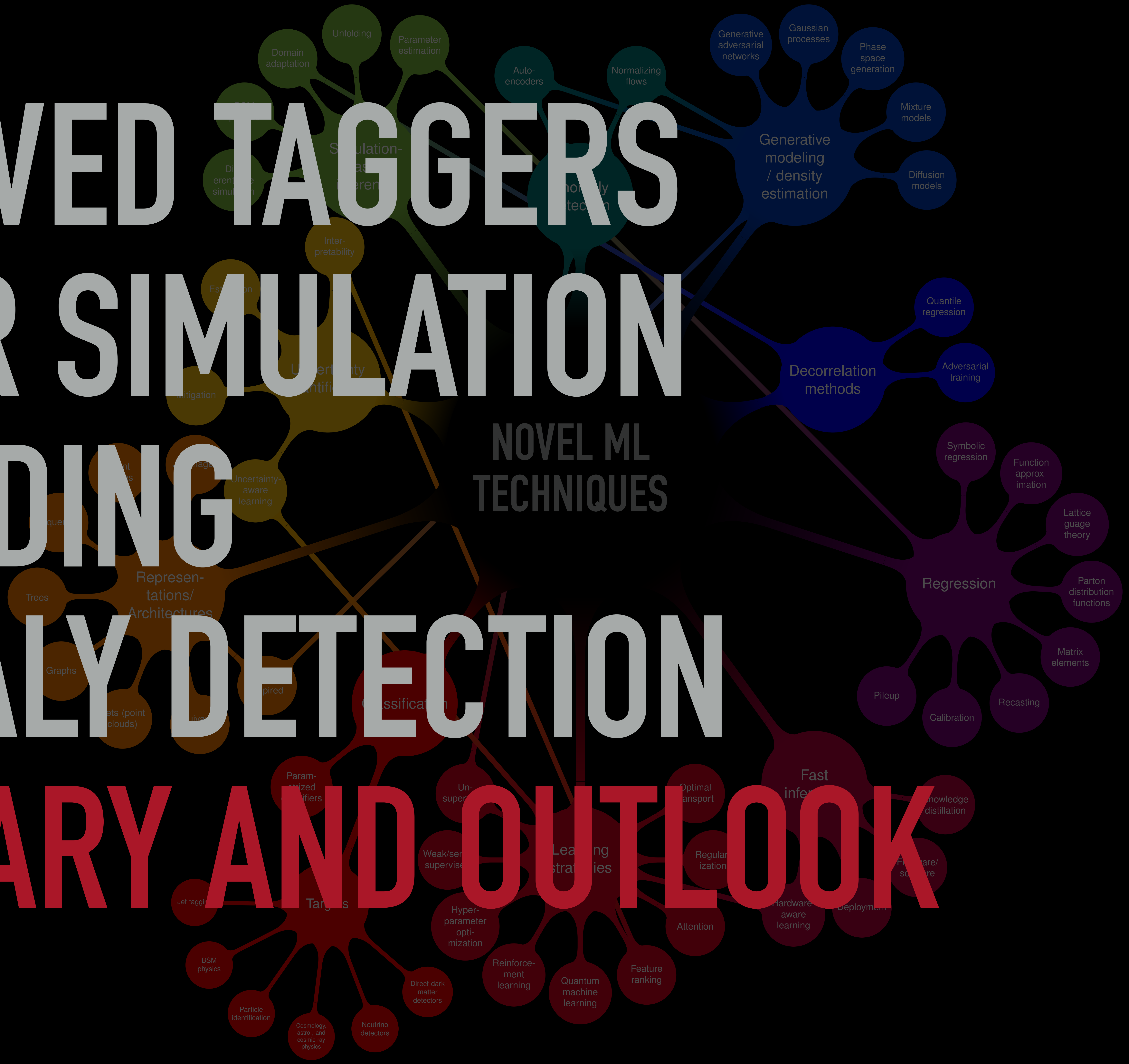


NORMALIZED AUTOENCODER

- Model data as $p_\theta(x) \propto \exp\left[-\mathrm{MSE}\left(x, \mathrm{AE}_\theta(x)\right)\right]$ and minimize $-\ln p_\theta(x)$
- In practice, approximately minimize this through clever reframing
  - Sample "negative" samples from $p_\theta(x)$ using MCMC
  - Sample "positive" samples from $p_{\mathrm{data}}(x)$
  - Minimize difference in expected MSE between negative and positive samples
- With this change in training, can capture more signal as anomalous!

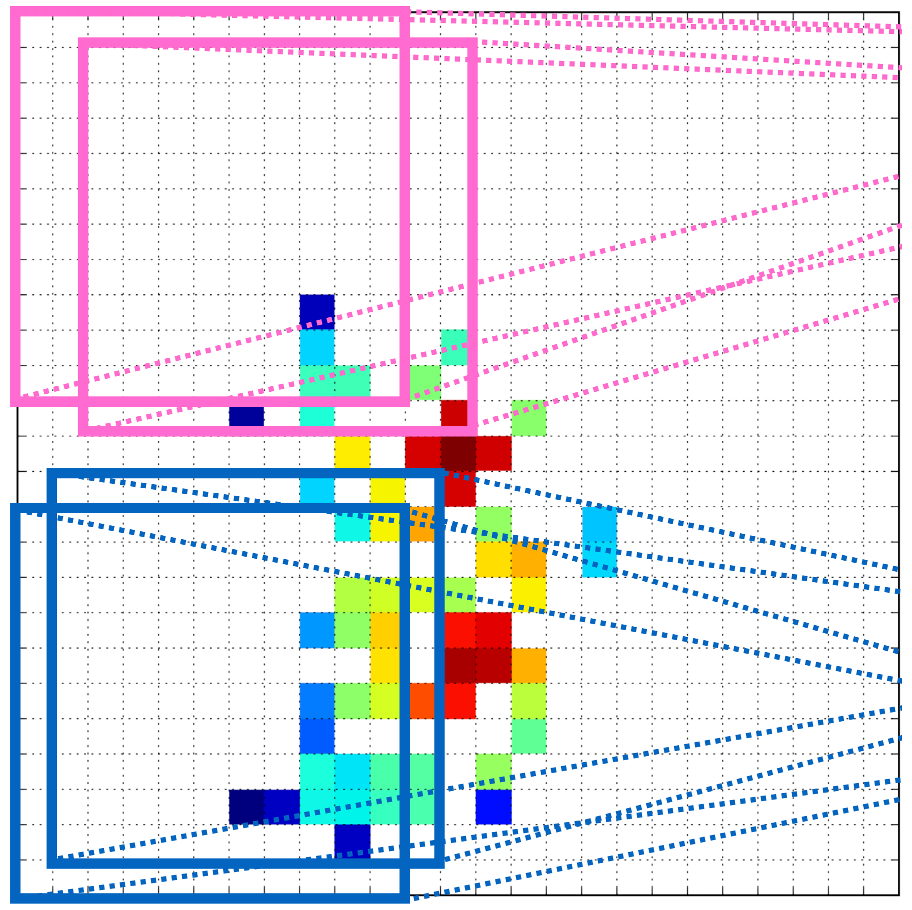


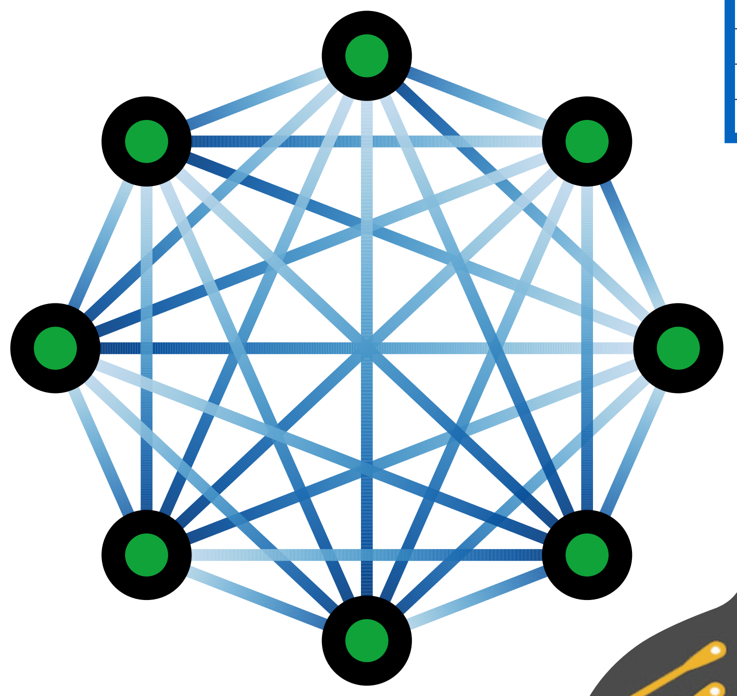
IMPROVED TAGGERS
FASTER SIMULATION
UNFOLDING
ANOMALY DETECTION
SUMMARY AND OUTLOOK

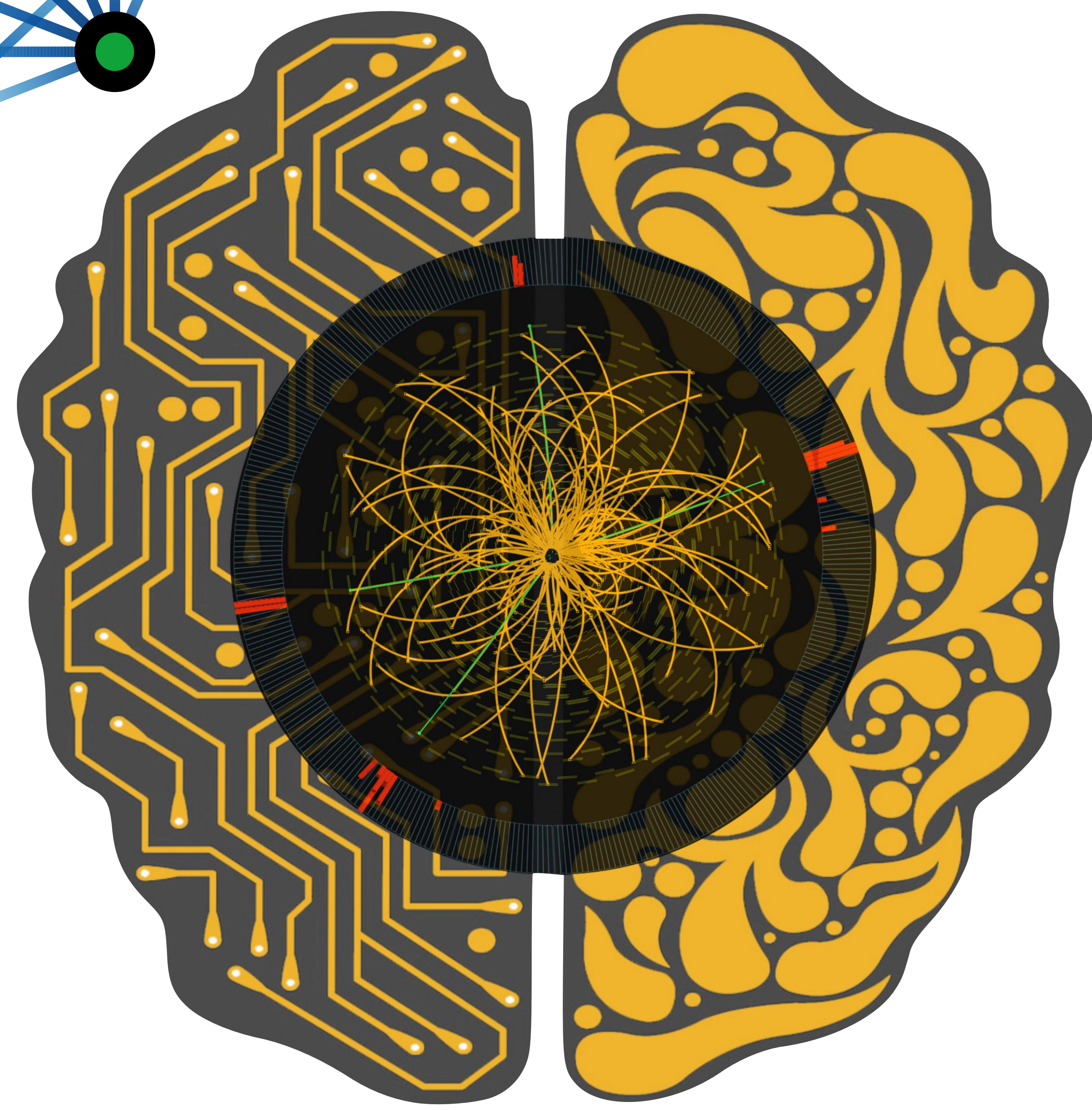


SUMMARY AND OUTLOOK

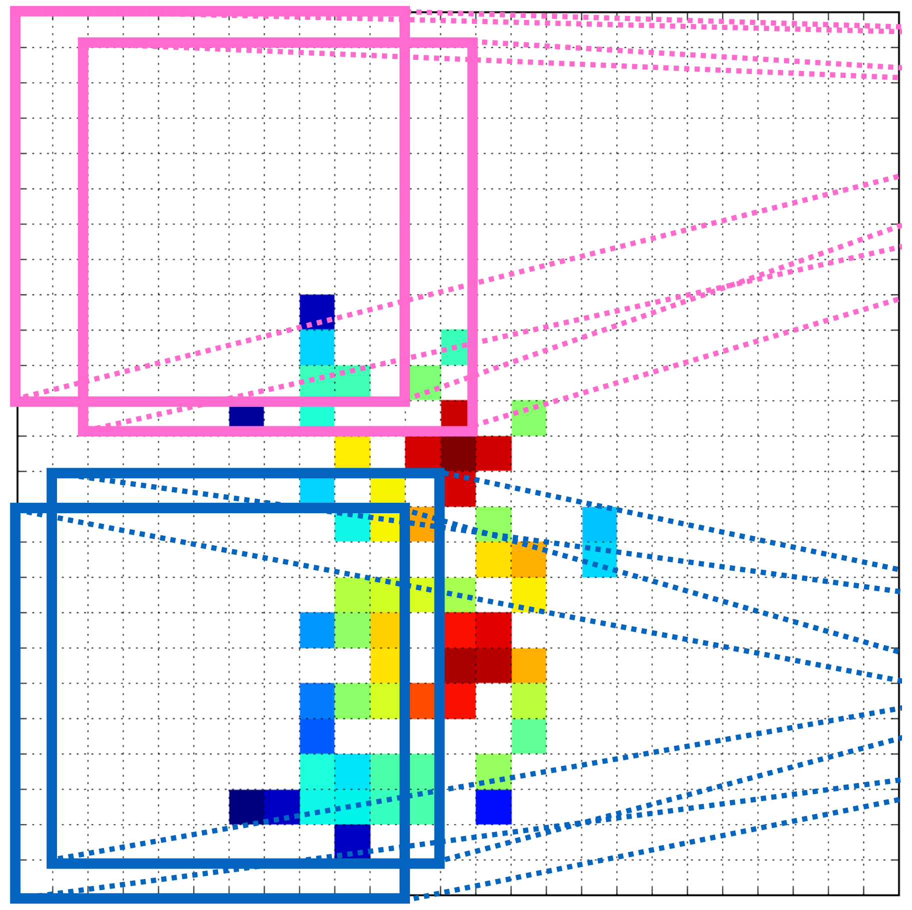
- Dizzying array of ***ML opportunities, innovations, and applications*** in ***LHC experiments***, which directly ***impact physics results***



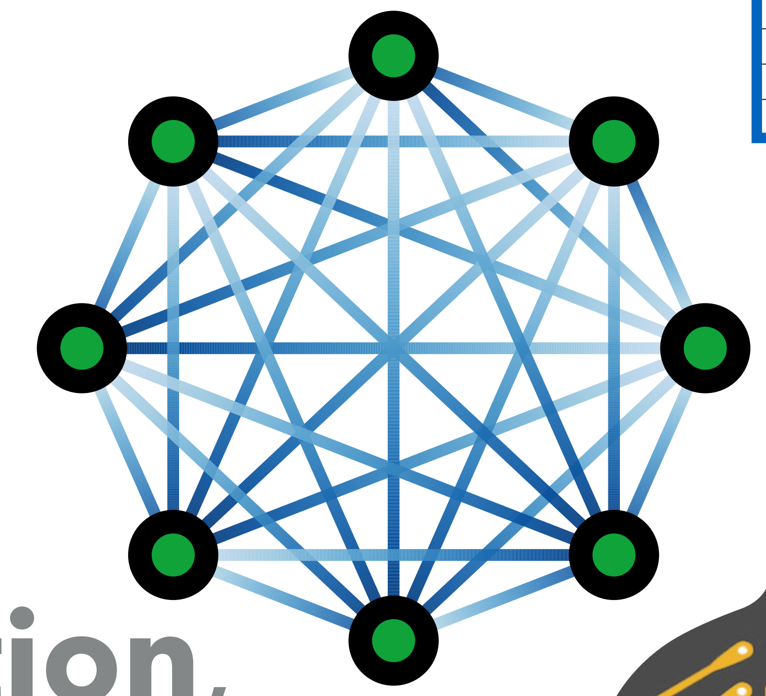


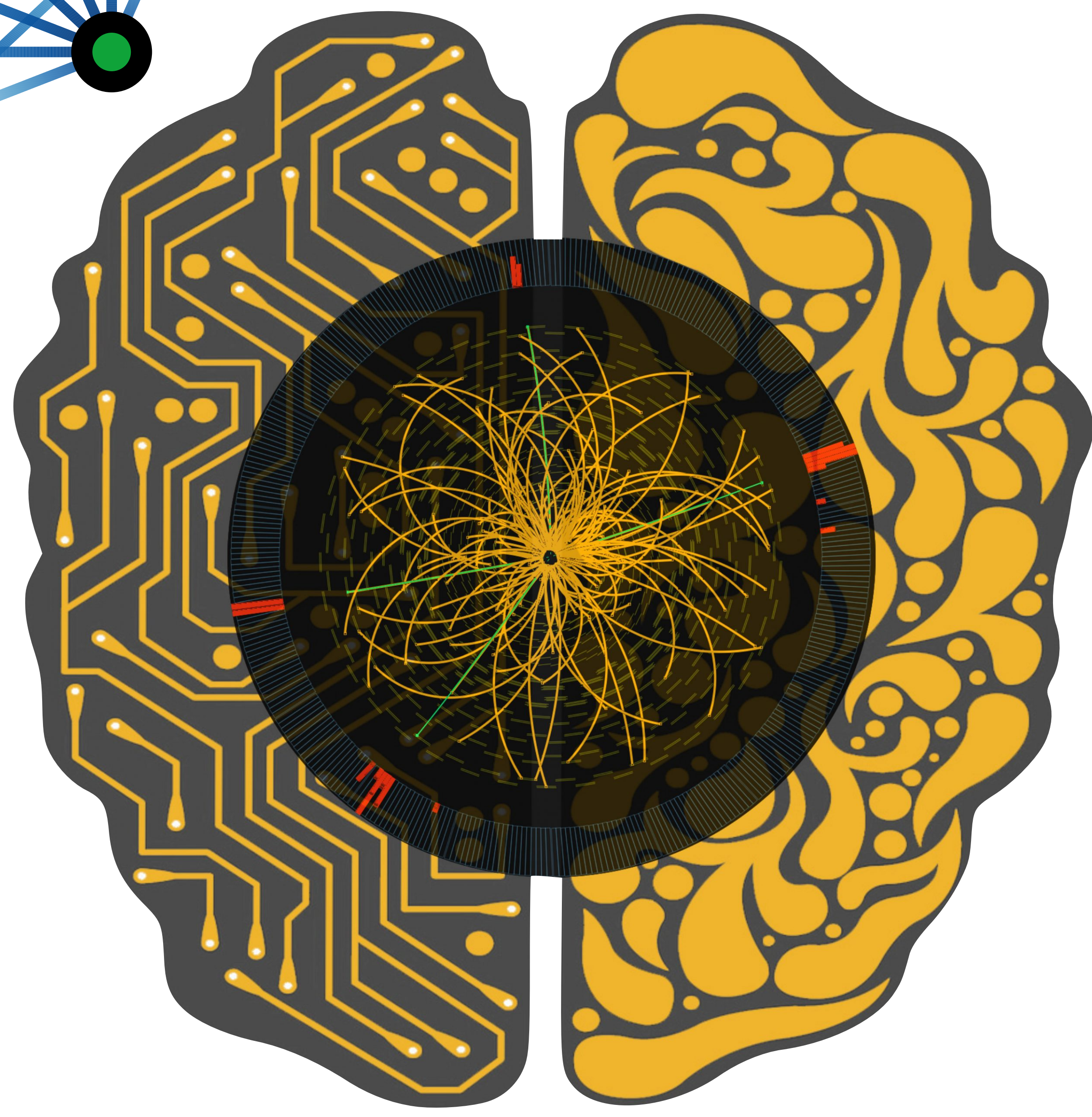


SUMMARY AND OUTLOOK

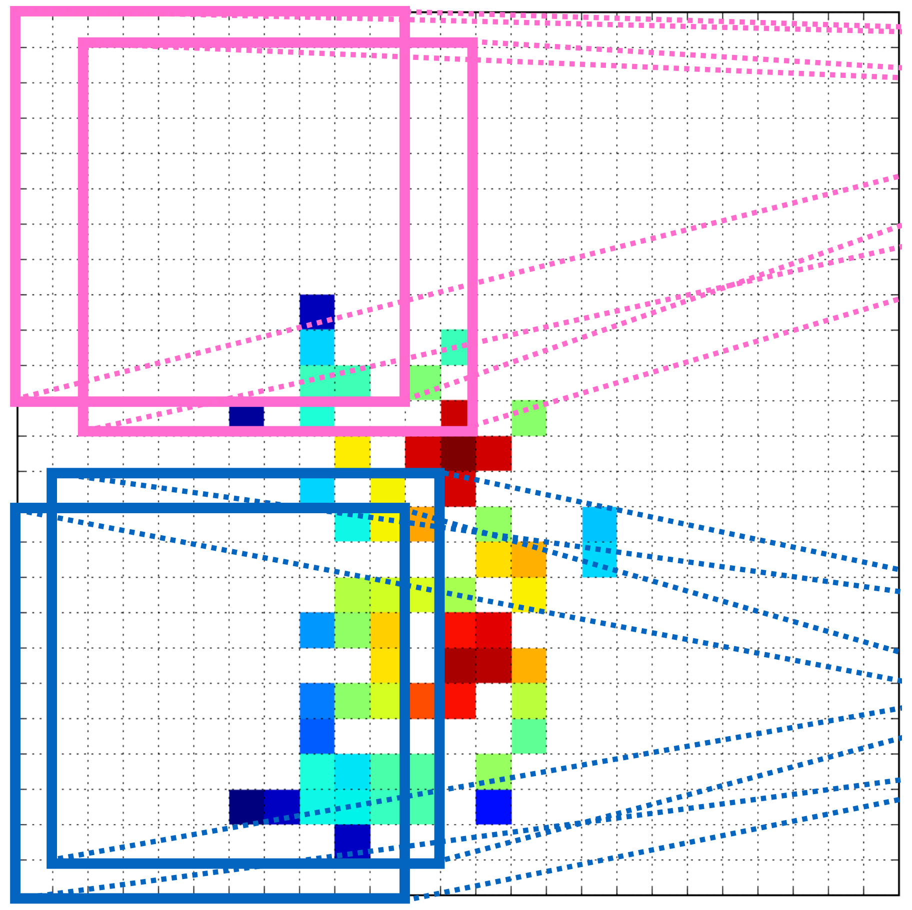


- Dizzying array of ***ML opportunities, innovations, and applications*** in ***LHC experiments***, which directly ***impact physics results***
- New ideas re-thinking what is the best way to apply ML
  - Not only concerned with performance but also **robustness**, **interpretability**, and insensitivity to modeling uncertainties…
  - Use cases **beyond classification** including **simulation**, **unfolding**, **anomaly detection**, and more…

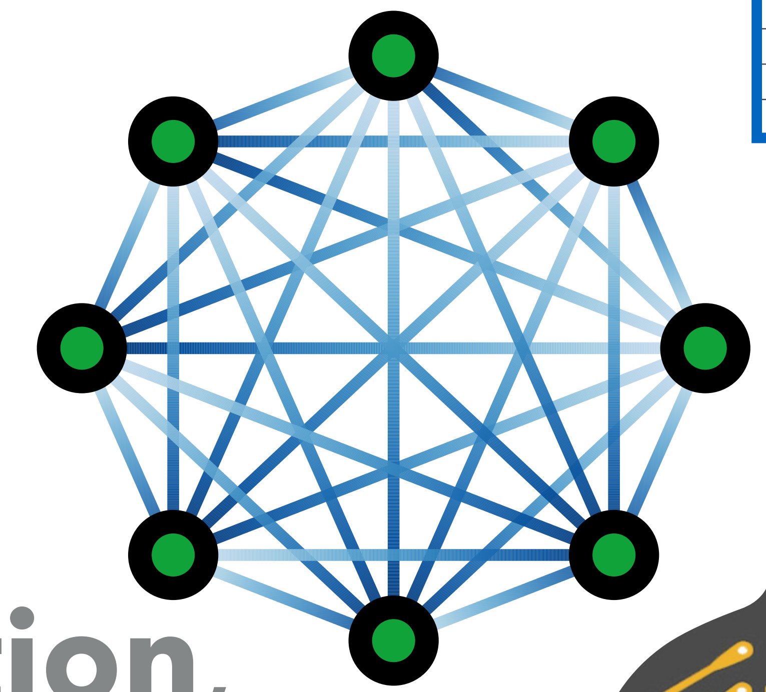


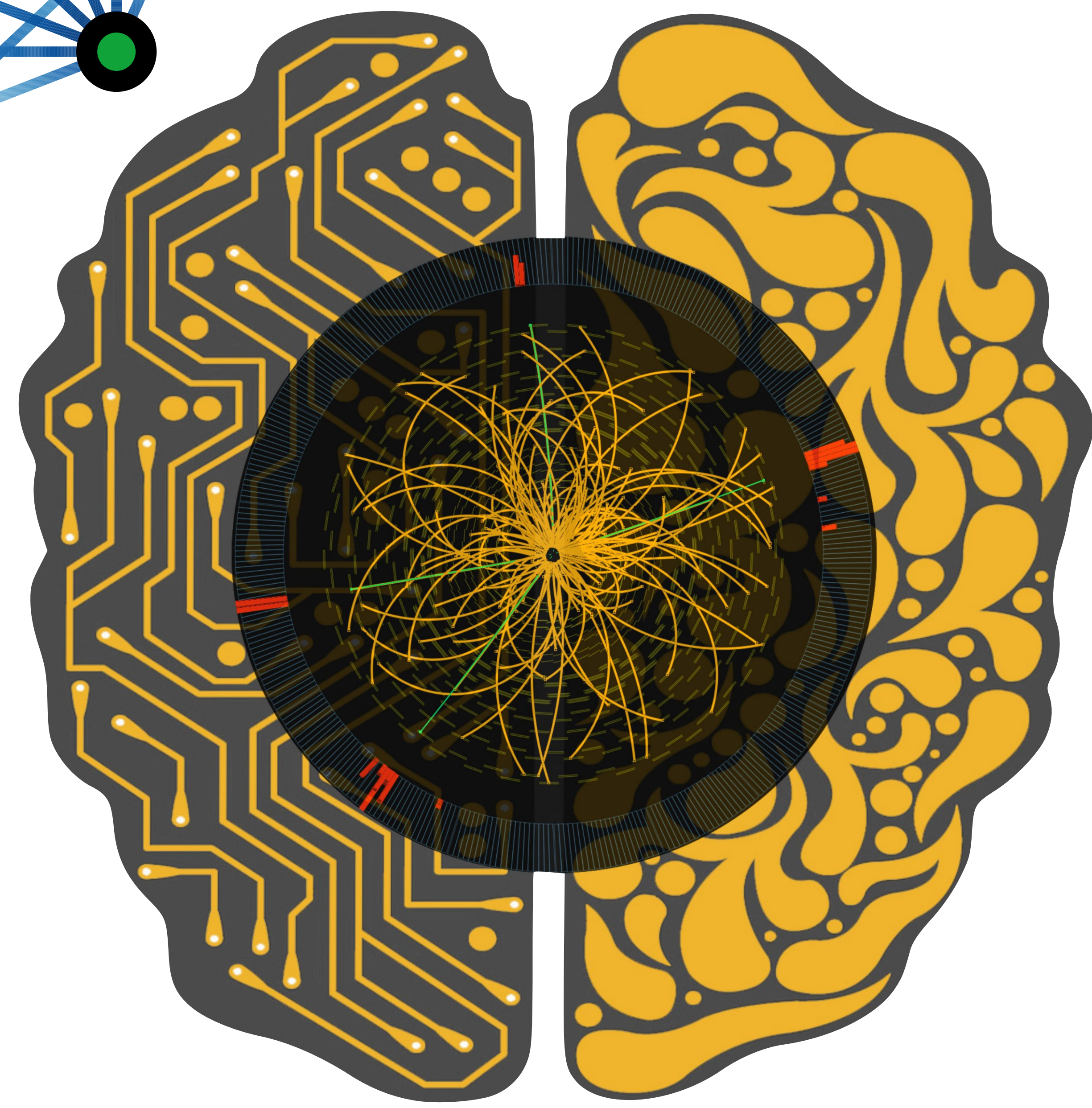


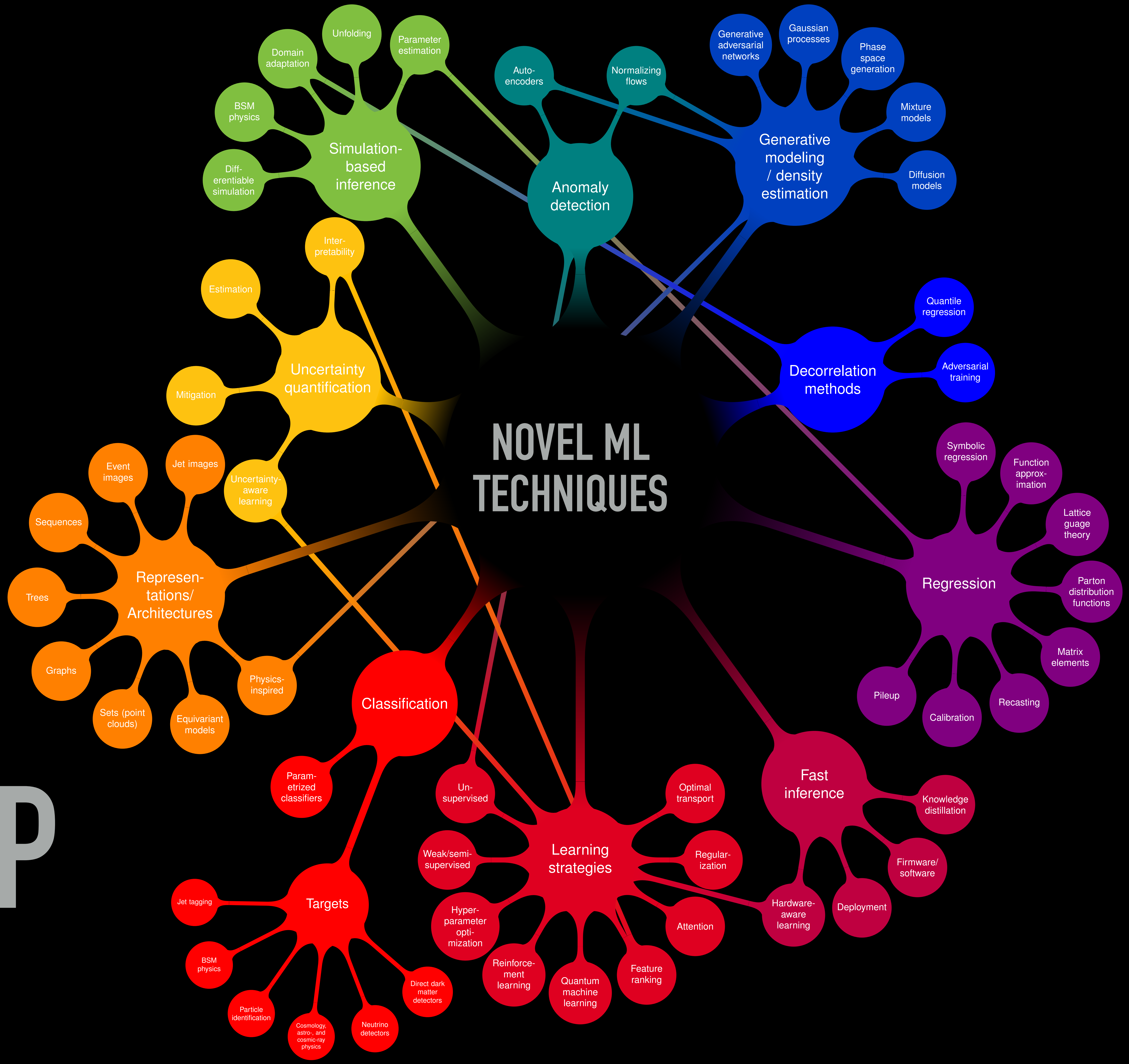
SUMMARY AND OUTLOOK



- Dizzying array of ***ML opportunities, innovations, and applications*** in ***LHC experiments***, which directly ***impact physics results***
- New ideas re-thinking what is the best way to apply ML
  - Not only concerned with performance but also **robustness**, **interpretability**, and insensitivity to modeling uncertainties…
  - Use cases **beyond classification** including **simulation**, **unfolding**, **anomaly detection**, and more…
- ML enables **new searches and measurements** that were impossible before



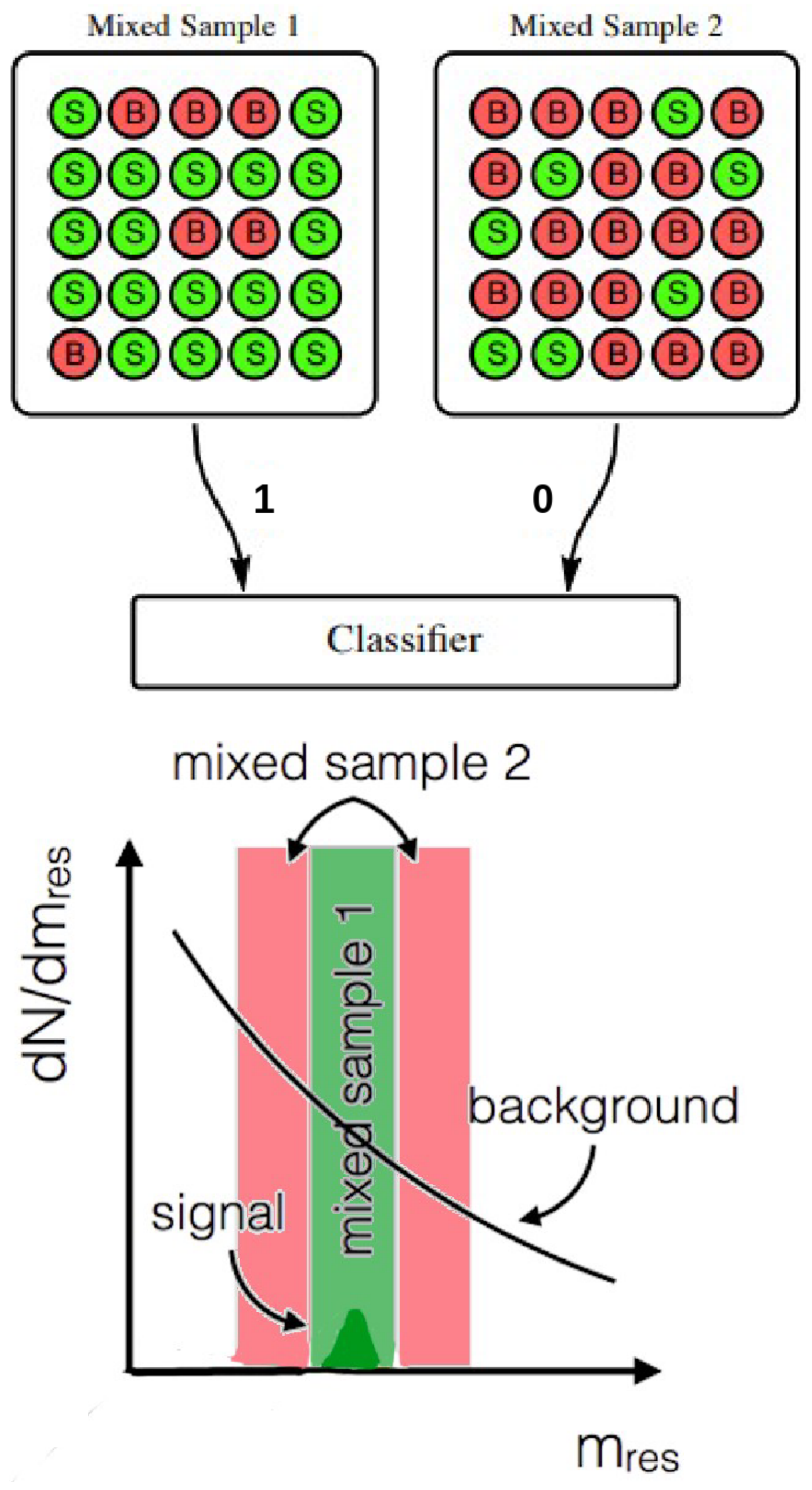



NOVEL ML TECHNIQUES
Simulation-based inference
Unfolding
Domain adaptation
Parameter estimation
BSM physics
Diff-erentiable simulation
Anomaly detection
Auto-encoders
Normalizing flows
Generative modeling / density estimation
Generative adversarial networks
Gaussian processes
Phase space generation
Mixture models
Diffusion models
Decorrelation methods
Quantile regression
Adversarial training
Regression
Symbolic regression
Function approx-imation
Lattice guage theory
Parton distribution functions
Matrix elements
Recasting
Calibration
Pileup
Fast inference
Knowledge distillation
Firmware/ software
Deployment
Hardware-aware learning
Learning strategies
Optimal transport
Regular-ization
Attention
Feature ranking
Quantum machine learning
Reinforce-ment learning
Hyper-parameter opti-mization
Weak/semi-supervised
Un-supervised
Classification
Param-etrized classifiers
Targets
Jet tagging
BSM physics
Particle identification
Cosmology, astro-, and cosmic-ray physics
Neutrino detectors
Direct dark matter detectors
Represen-tations/ Architectures
Event images
Jet images
Sequences
Trees
Graphs
Sets (point clouds)
Equivariant models
Physics-inspired
Uncertainty quantification
Inter-pretability
Estimation
Mitigation
Uncertainty-aware learning

BACKUP

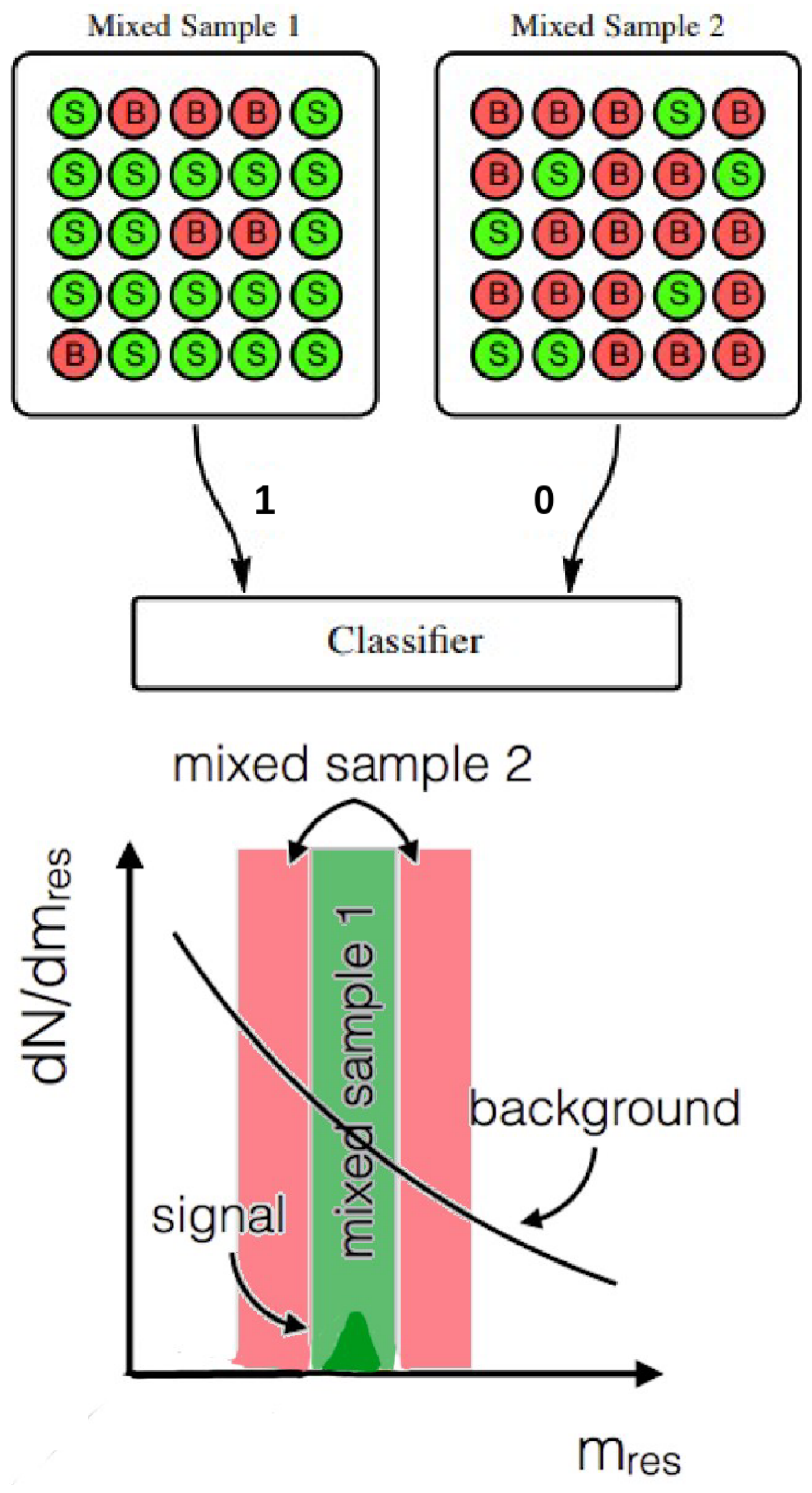
ANOMALY DETECTION: 5 COMPLEMENTARY METHODS

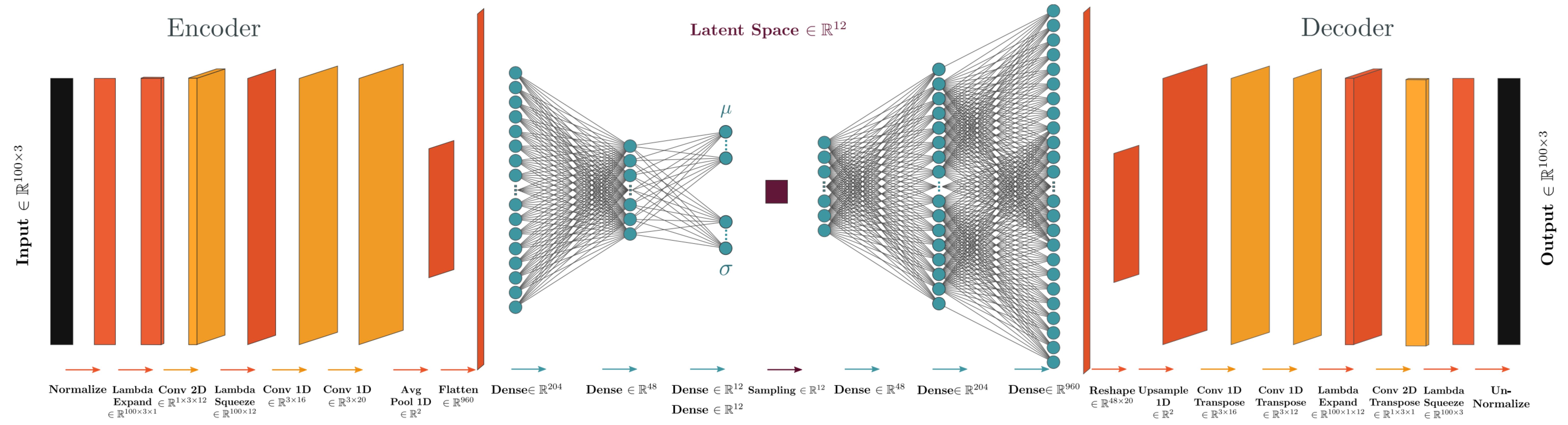
- Weakly supervised
  - **CWoLa Hunting**
  - Tag N' Train (TNT)
  - Classifying anomalies through outer density estimation (CATHODE)



ANOMALY DETECTION: 5 COMPLEMENTARY METHODS

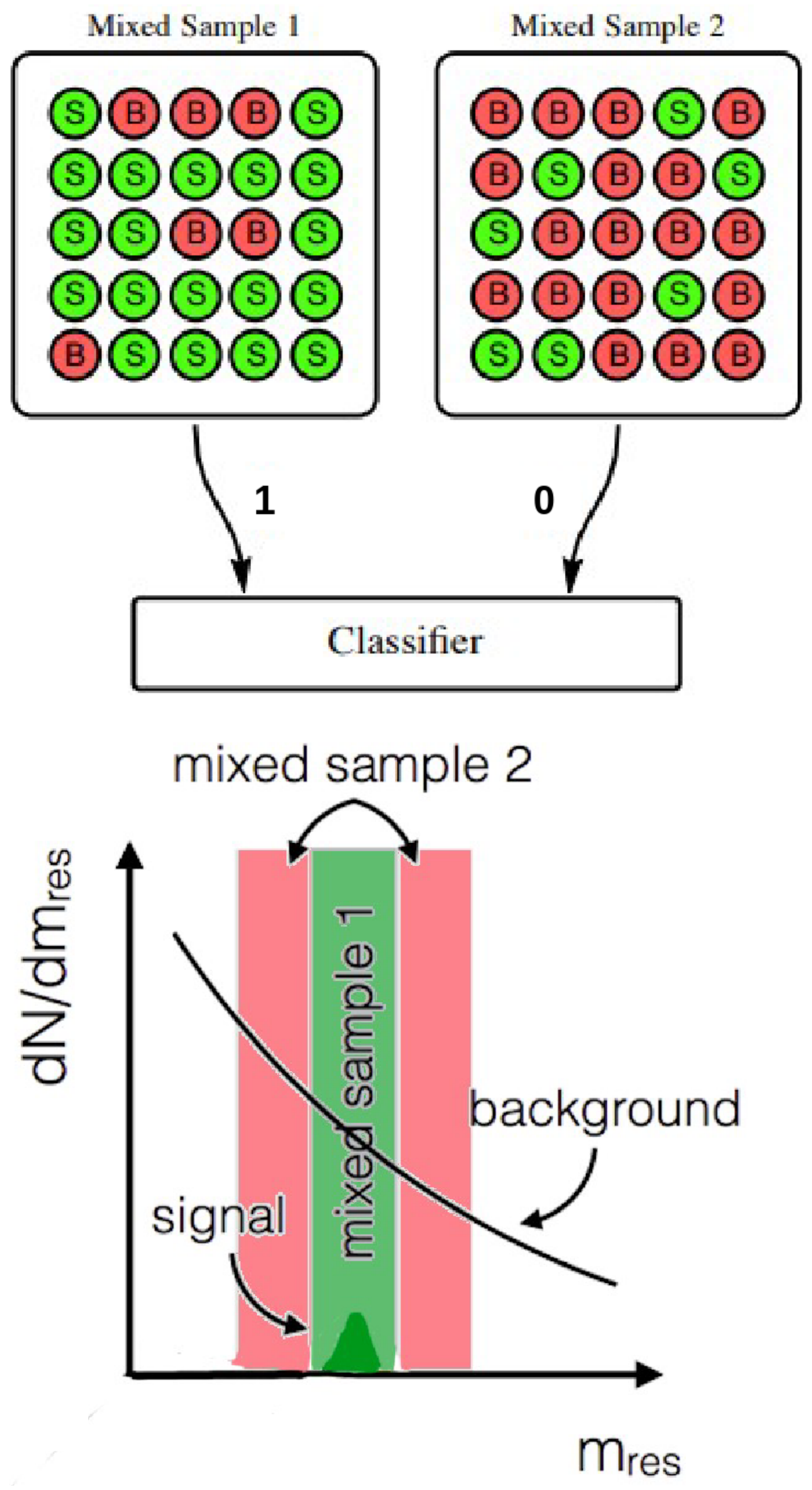
- Weakly supervised
  - **CWoLa Hunting**
  - Tag N' Train (TNT)
  - Classifying anomalies through outer density estimation (CATHODE)
- Unsupervised
  - **Variational autoencoder with quantile regression (VAE-QR)**

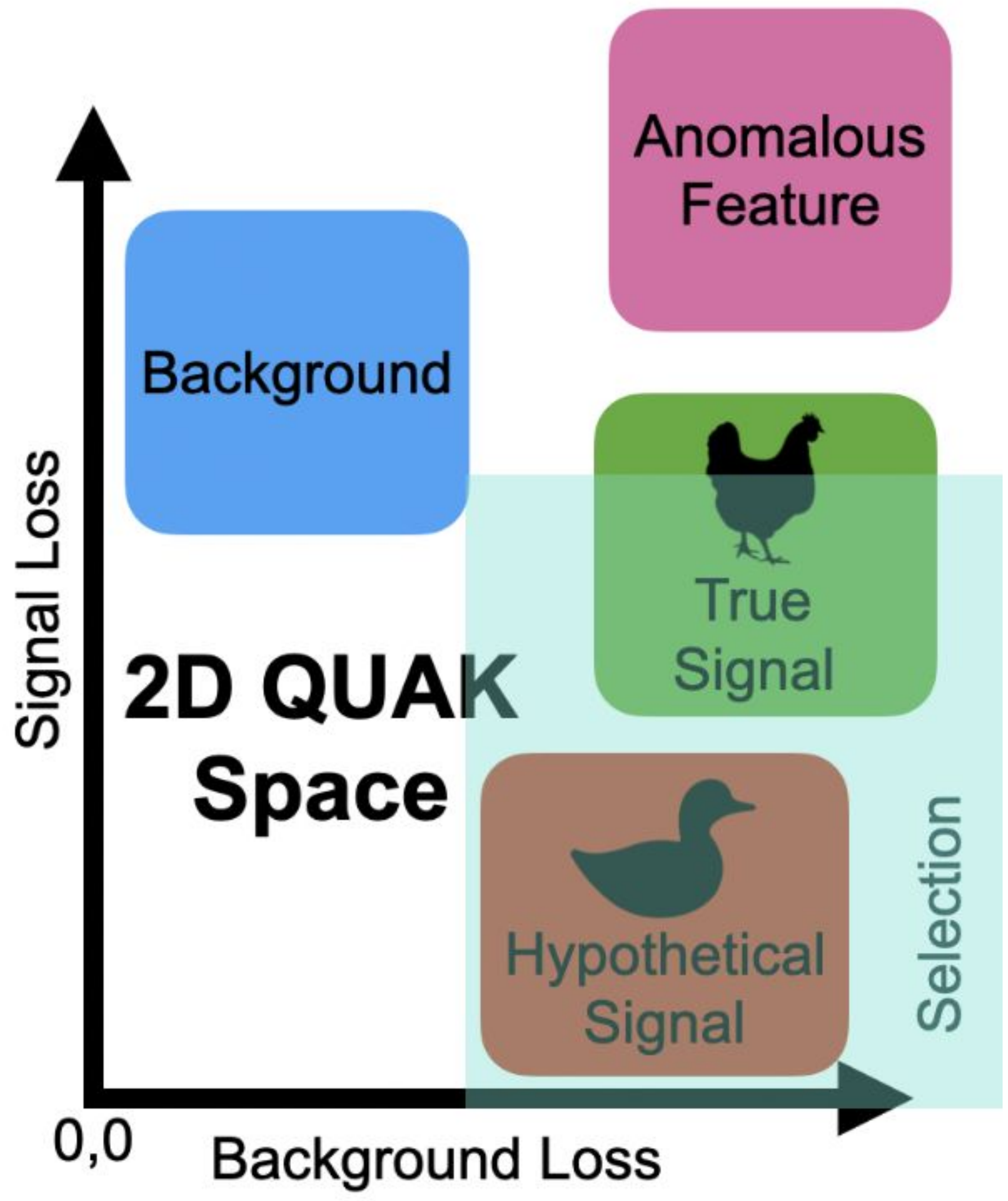


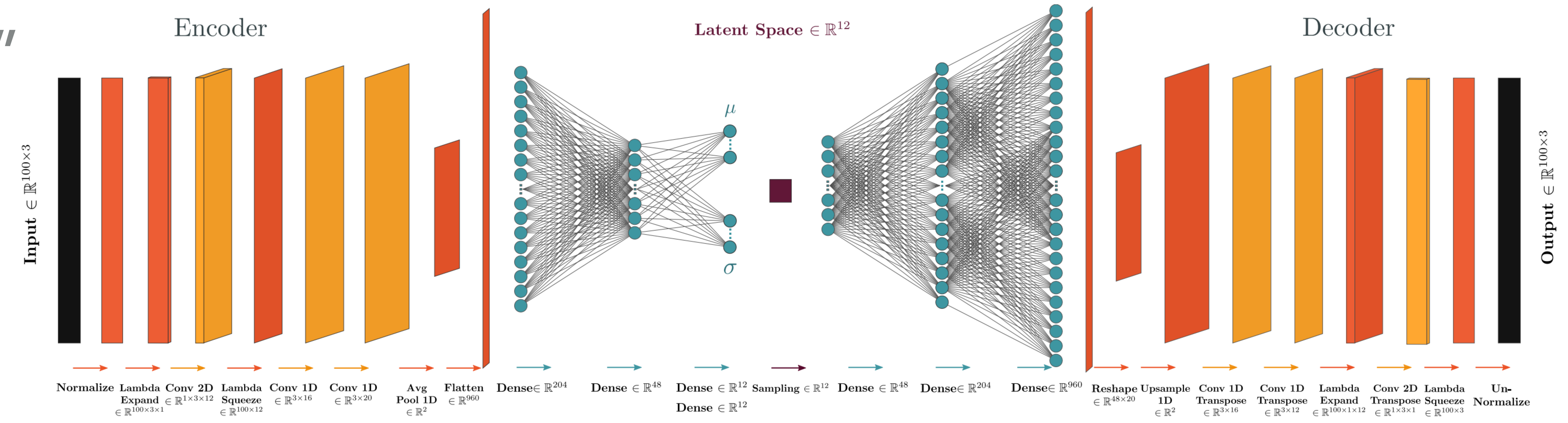


ANOMALY DETECTION: 5 COMPLEMENTARY METHODS

- Weakly supervised
  - **CWoLa Hunting**
  - Tag N' Train (TNT)
  - Classifying anomalies through outer density estimation (CATHODE)
- Unsupervised
  - **Variational autoencoder with quantile regression (VAE-QR)**
- Hybrid: encode "signal prior" based on expected signals
  - Quasianomalous knowledge (QUAK)







- GNN4ITK track reconstruction approaching standard CKF approach