

Run: 280464

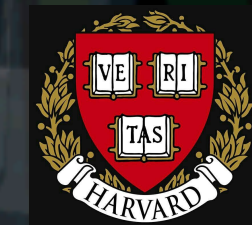
Event: 478442529

2015-09-27 22:09:07 CEST

ML for reconstruction

Our journey of deep learning impacting LHC physics

Nicole Hartman (nicomh@mit.edu)
ML4Jets, Vienna 14/9/2026



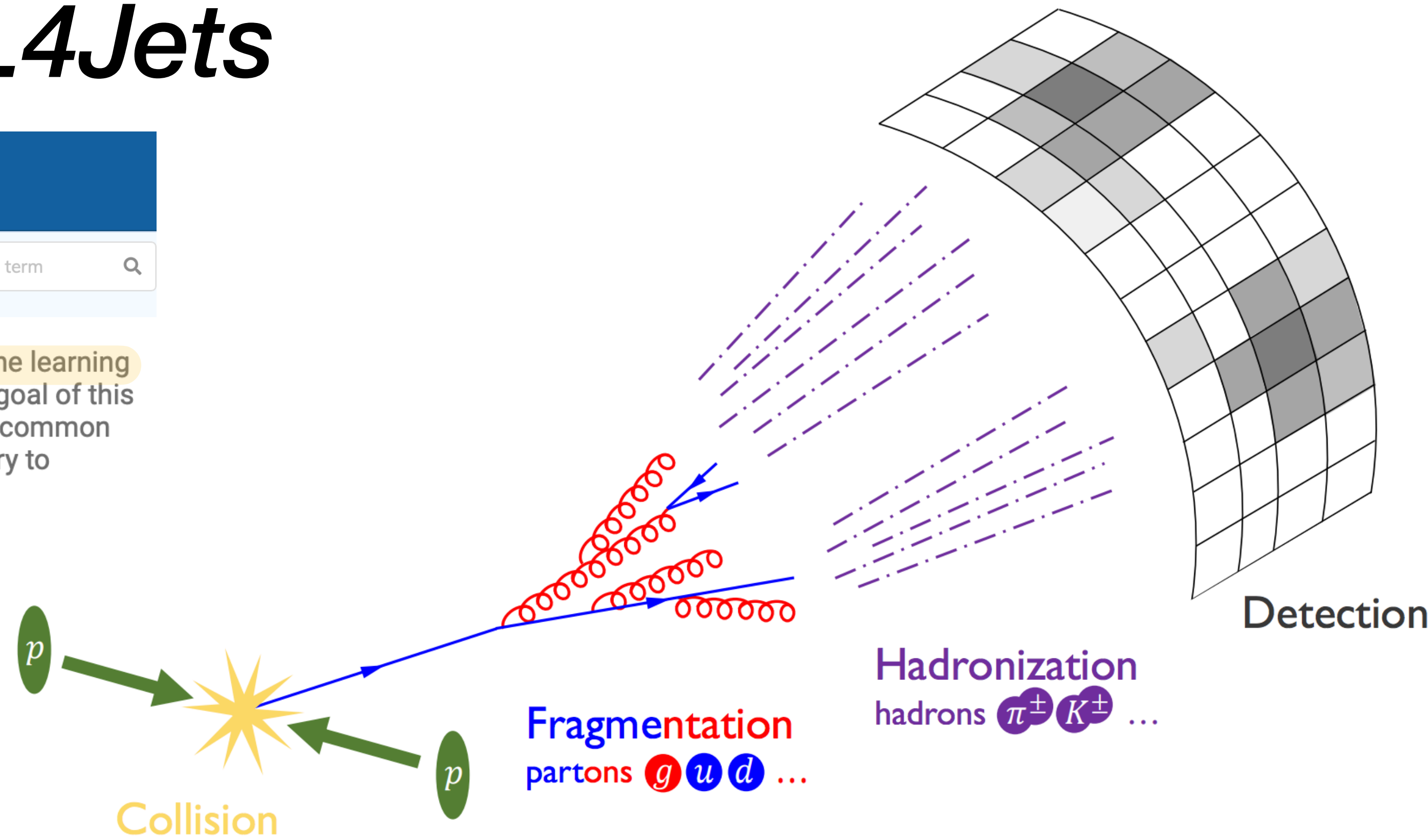
Almost a decade of ML4Jets

Machine Learning for Jet Physics

Dec 11 – 13, 2017
Lawrence Berkeley National Laboratory
US/Pacific timezone

Enter your search term

There has been a recent surge of interest in developing and applying advanced machine learning techniques in HEP, and jet physics is a domain at the forefront of the excitement. The goal of this workshop is to gather experts and new-comers to discuss progress, new ideas, and common challenges. The workshop is open to the community; we invite contributions and will try to accommodate everyone within reason.



2017

2026

Berkeley

FNAL

NYU

Heidelberg

Rutgers

DESY /
Hamburg

Paris

CalTech

Vienna

Almost a decade of ML4Jets

Machine Learning for Jet Physics

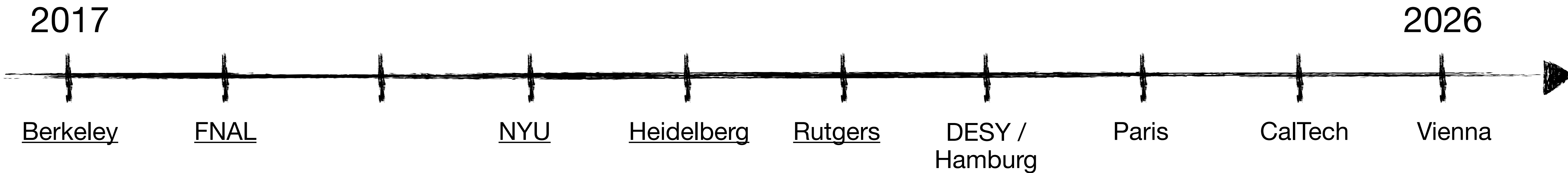
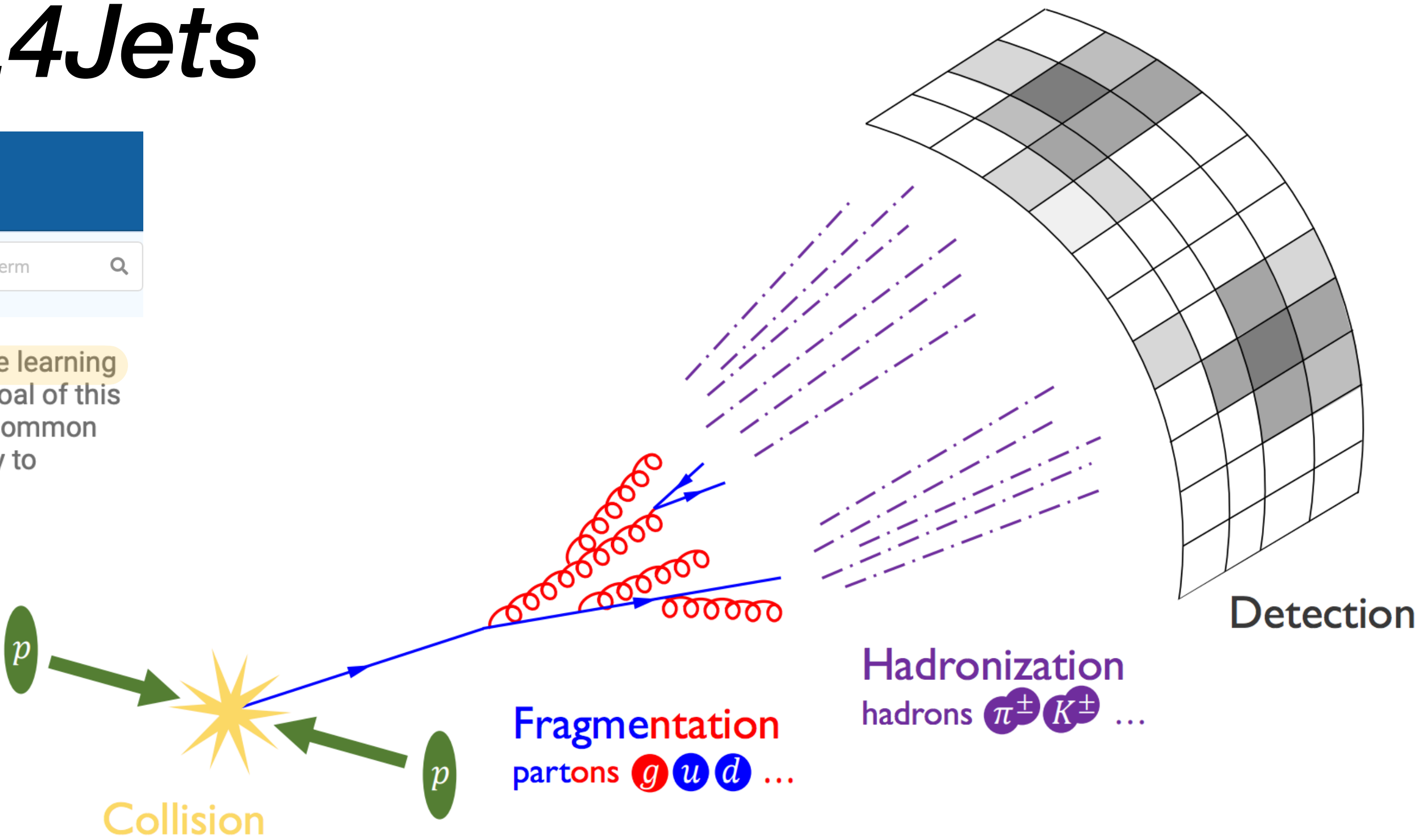
Dec 11 – 13, 2017
Lawrence Berkeley National Laboratory
US/Pacific timezone

Enter your search term

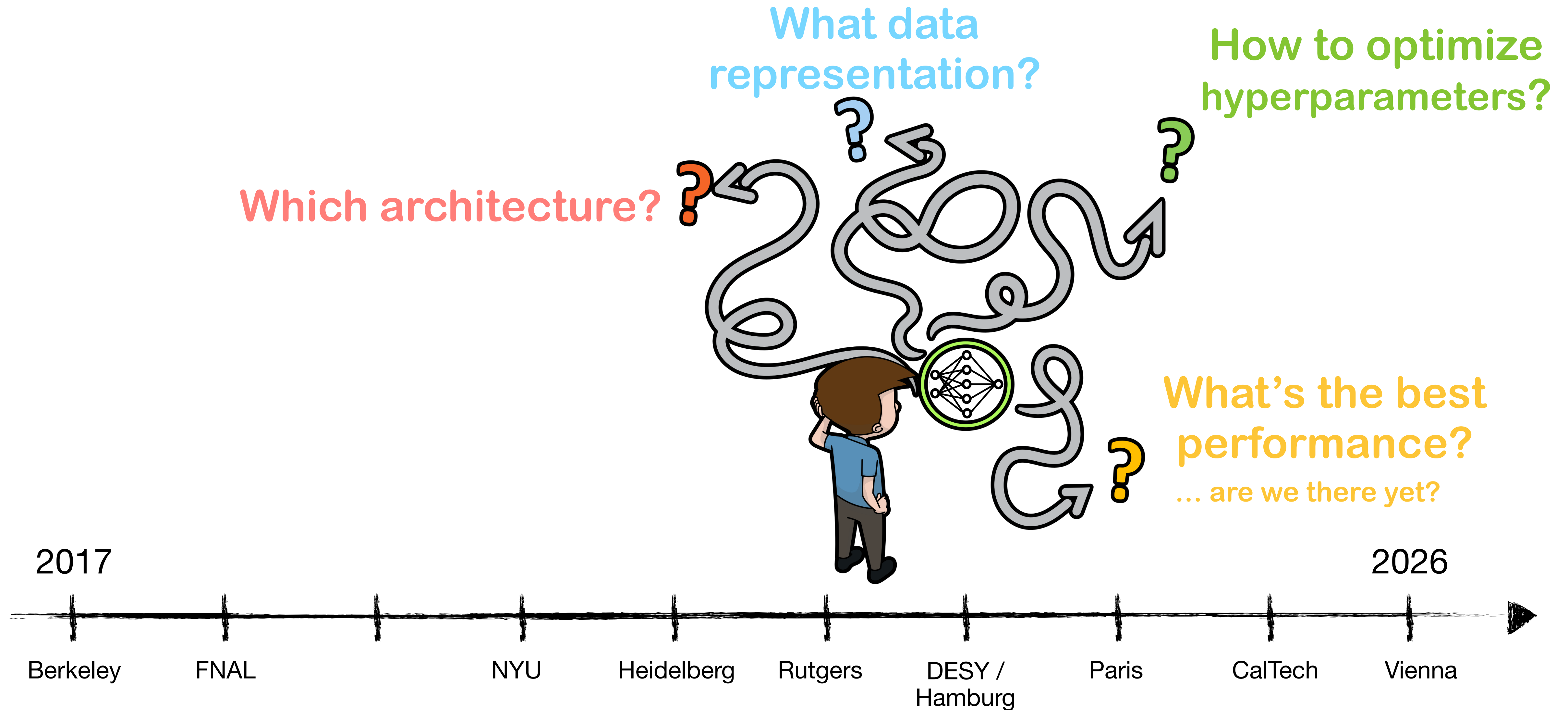
There has been a recent surge of interest in developing and applying advanced machine learning techniques in HEP, and jet physics is a domain at the forefront of the excitement. The goal of this workshop is to gather experts and new-comers to discuss progress, new ideas, and common challenges. The workshop is open to the community; we invite contributions and will try to accommodate everyone within reason.

Jets: common language between theorists and experimentalists

2506.13449

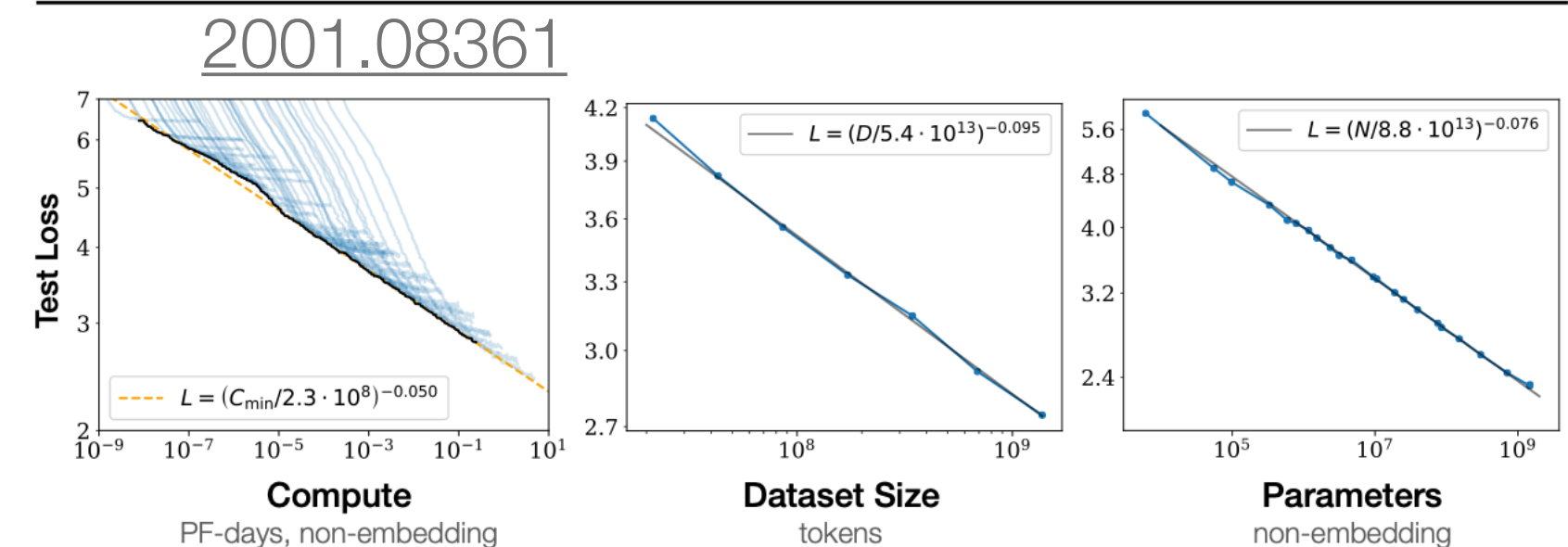


Where have we come from?

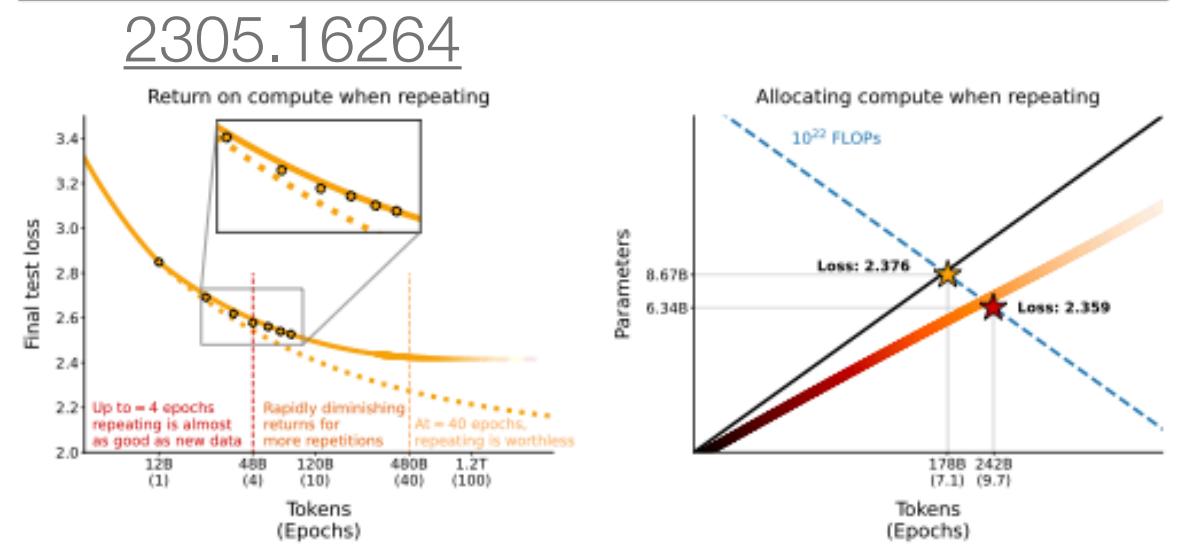


A decade of ML defying expectations

Scaling Laws for Neural Language Models



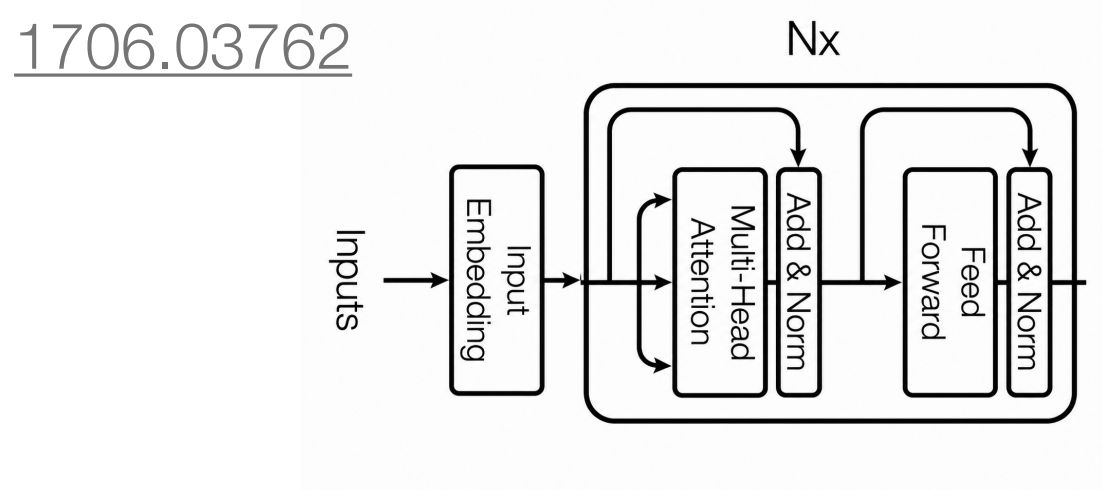
Scaling Data-Constrained Language Models



GPT3

2005.14165

Attention Is All You Need



Training Compute-Optimal Large Language Models

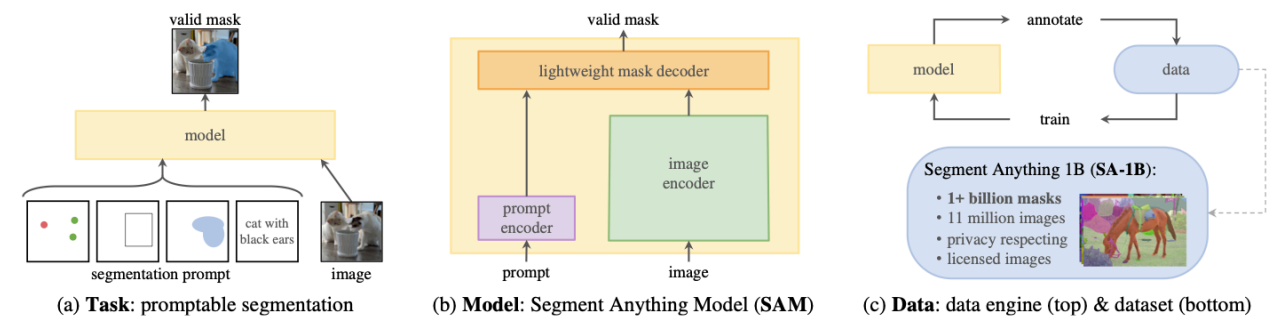
Jordan Hoffmann*, Sebastian Borgeaud*, Arthur Mensch*, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W. Rae, Oriol Vinyals and Laurent Sifre*

*Equal contributions

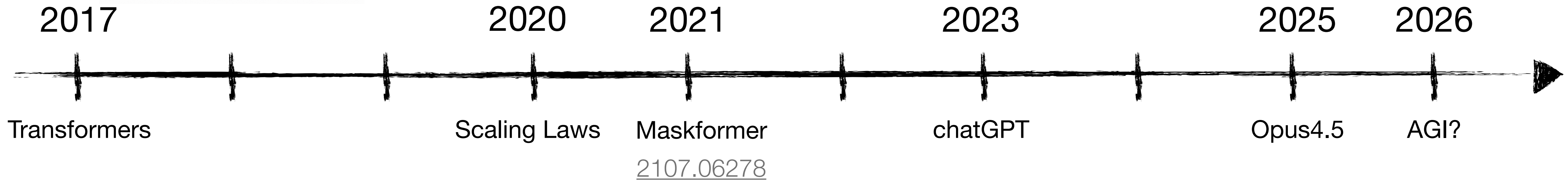
2203.15556

Segment Anything

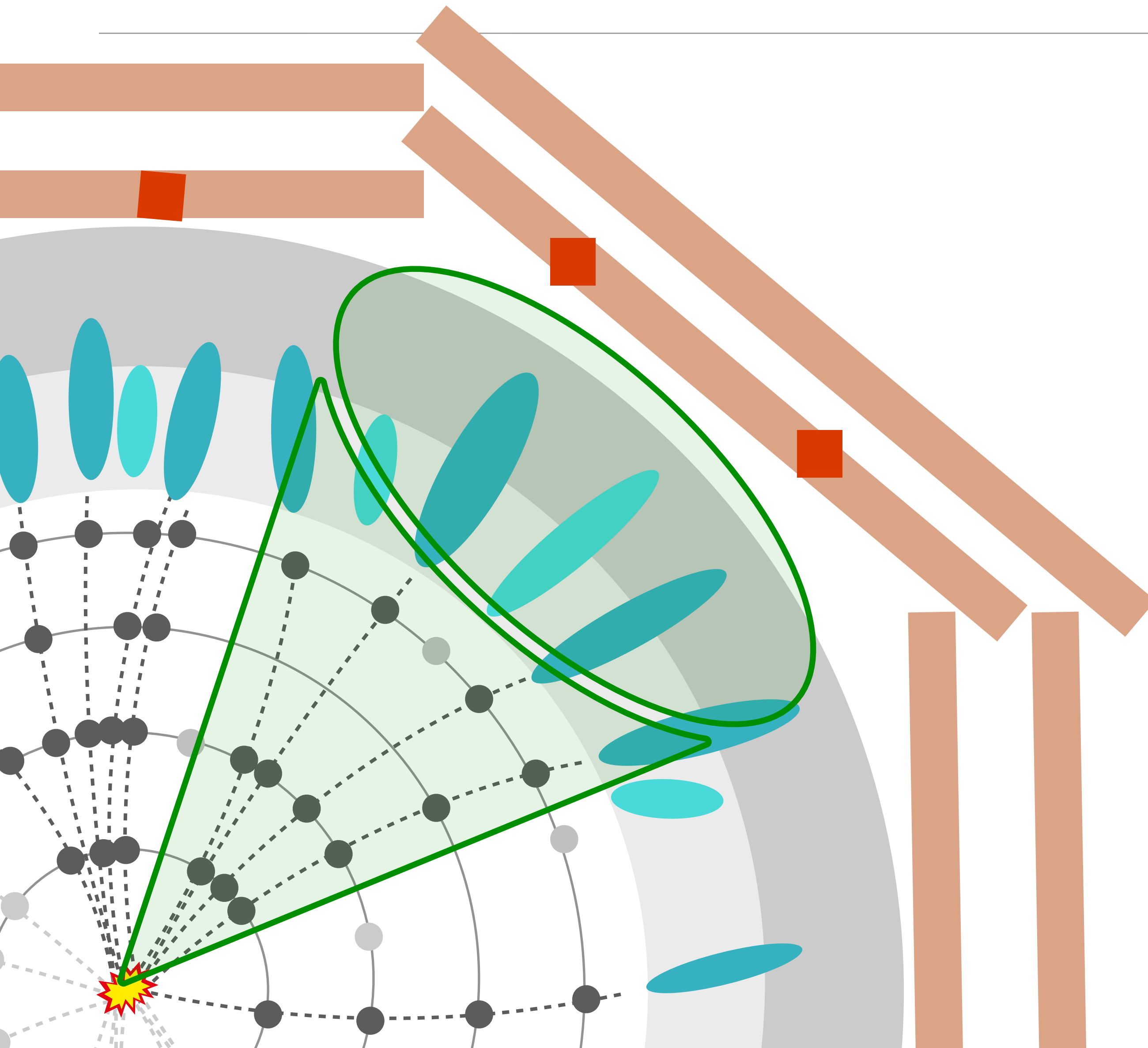
Alexander Kirillov^{1,2,4} Eric Mintun² Nikhila Ravi^{1,2} Hanzi Mao² Chloe Rolland³ Laura Gustafson³
Tete Xiao³ Spencer Whitehead¹ Alexander C. Berg² Wan-Yen Lo² Piotr Dollár⁴ Ross Girshick⁴
¹project lead ²joint first author ³equal contribution ⁴directional lead
Meta AI Research, FAIR



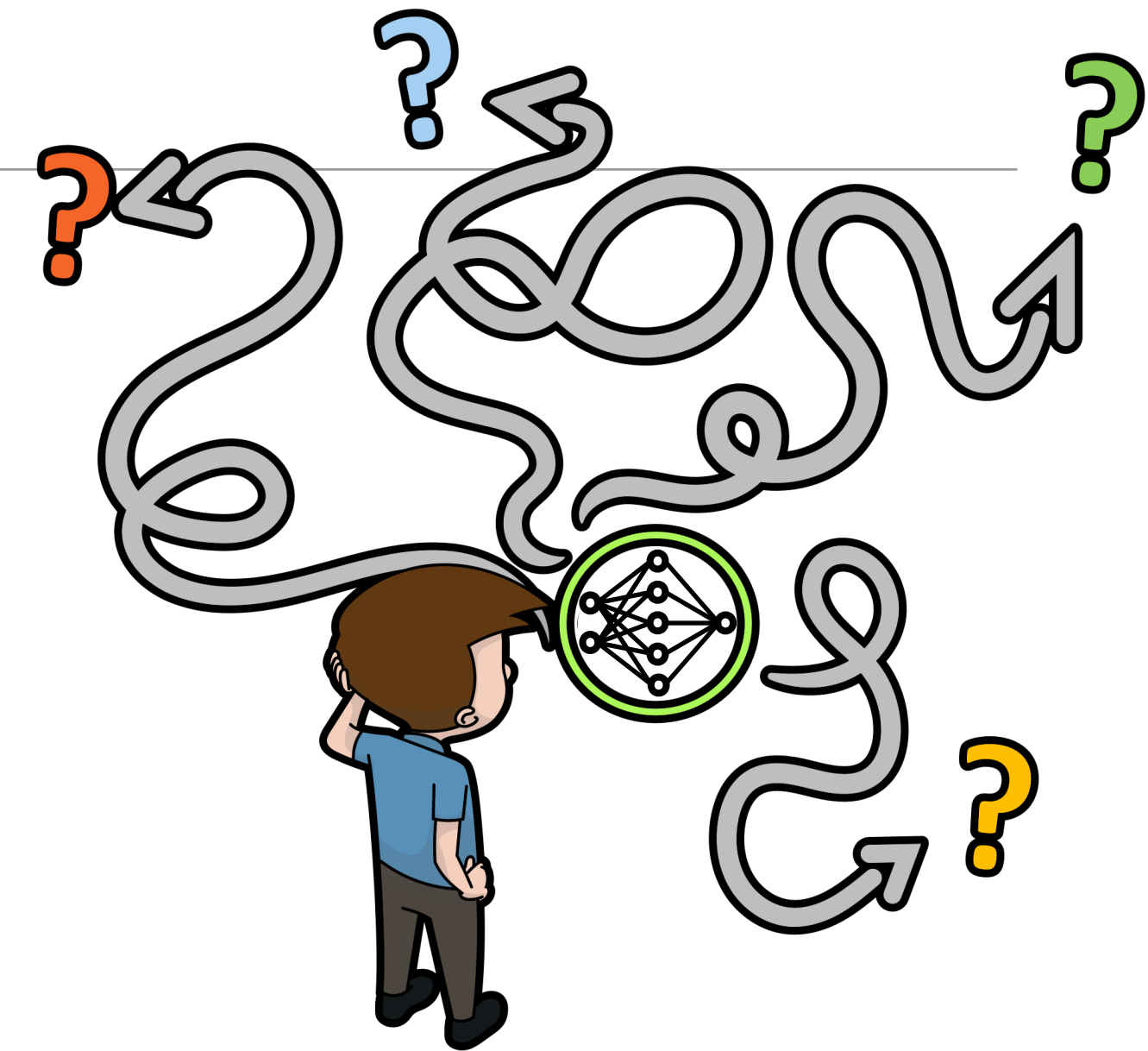
2304.02643



Data representation



What data representation?



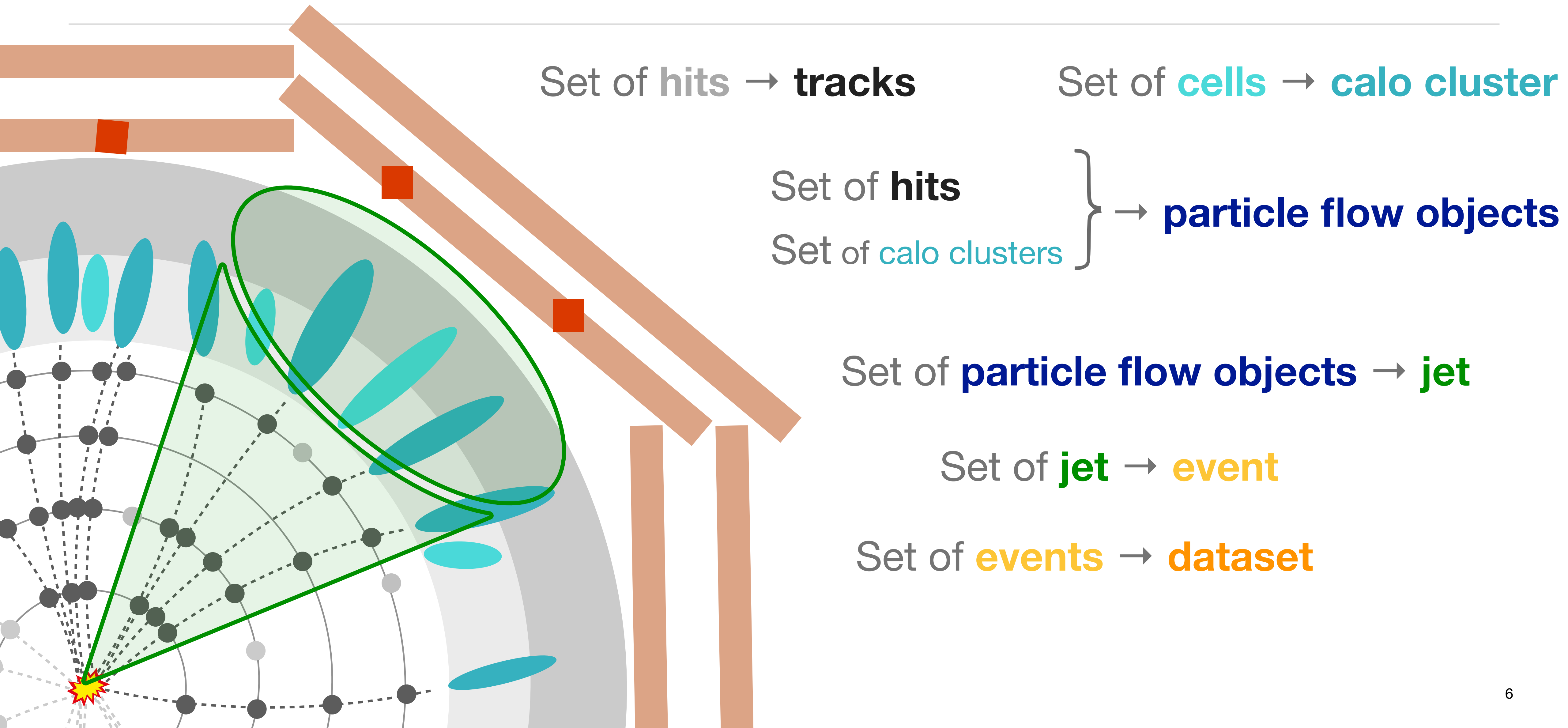
Set of **particle flow objects** → **jet**

$$X_{\text{jet}} = \{X_i\}_{i=1}^n, X_i \in \mathbb{R}^m$$

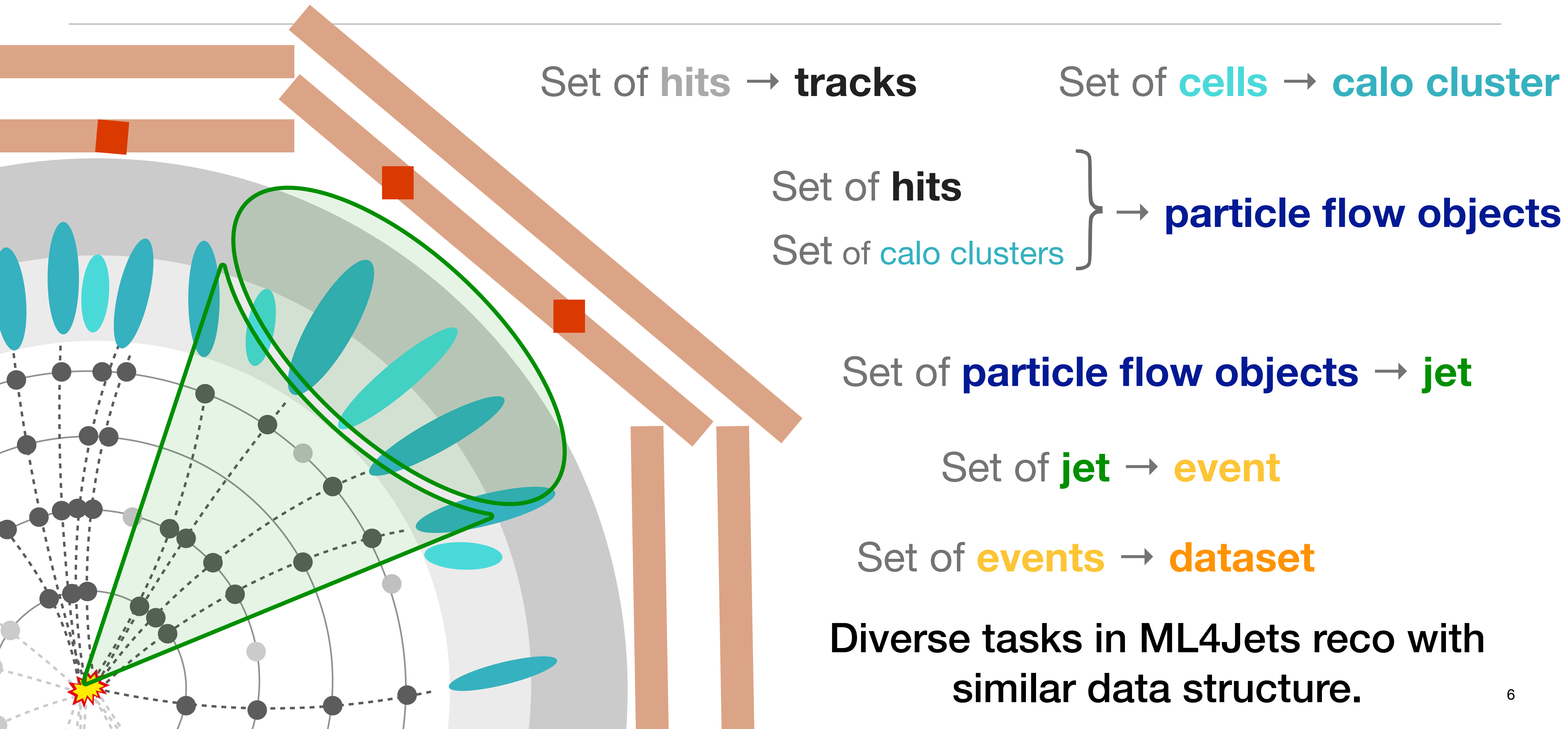
Y : labels or p_T regression

Learn: $p(Y \mid X_1, \dots, X_n)$

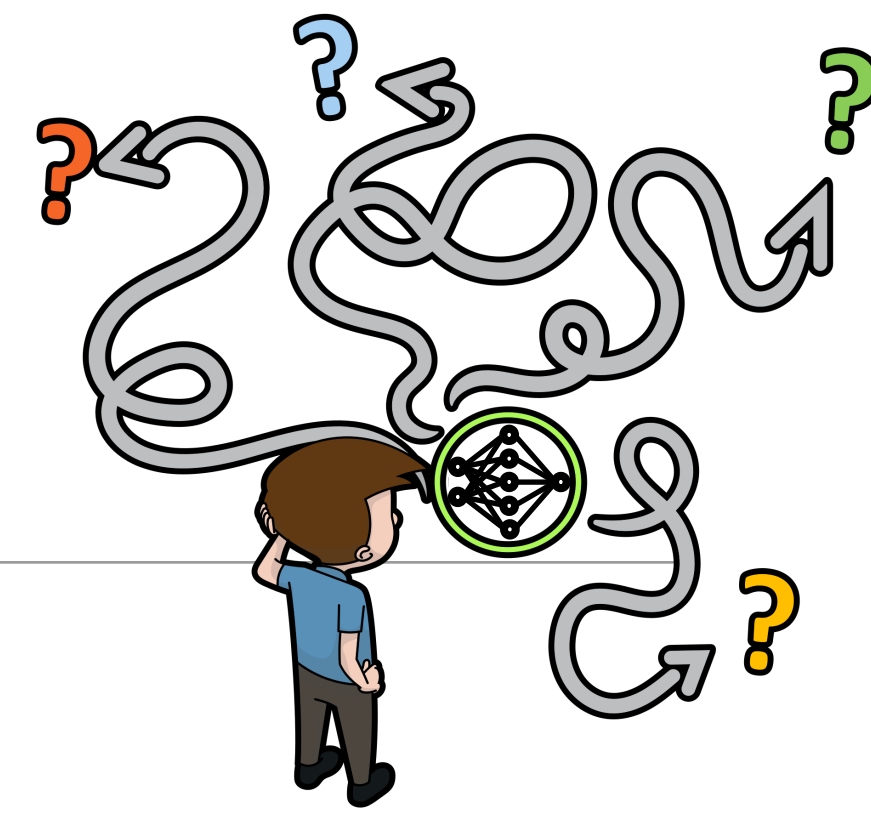
Data representation



Data representation

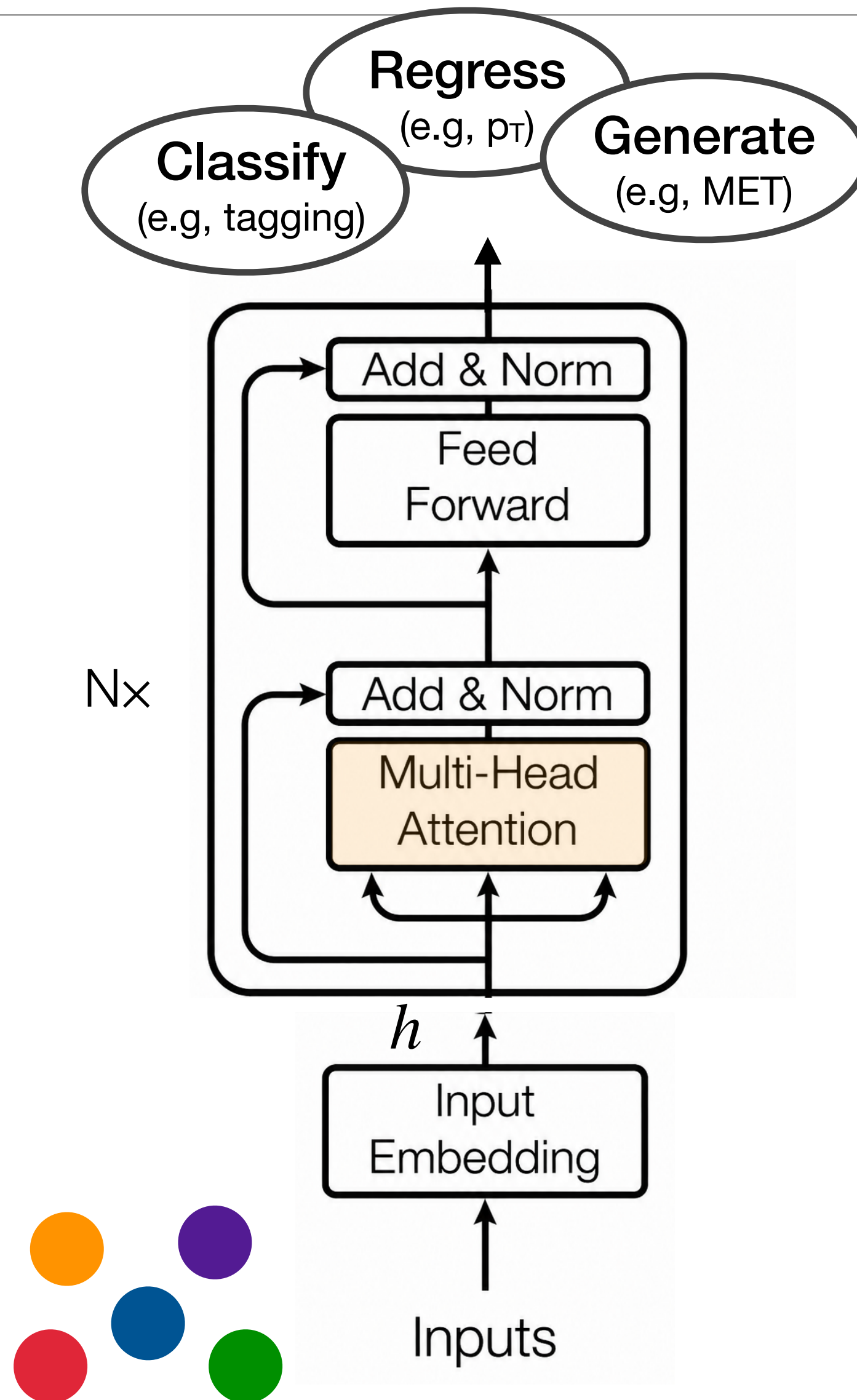
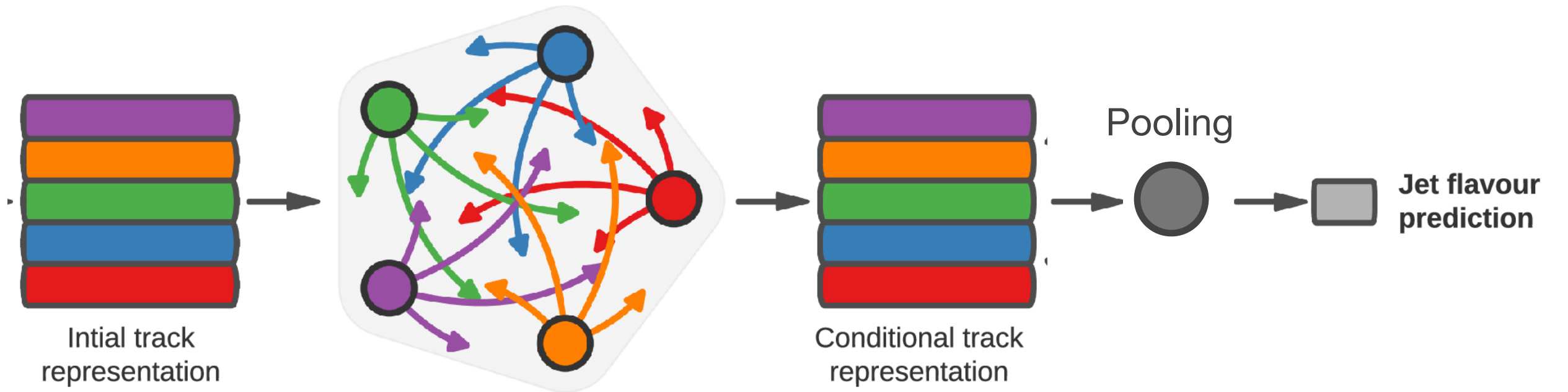


Transformer: our architecture of choice

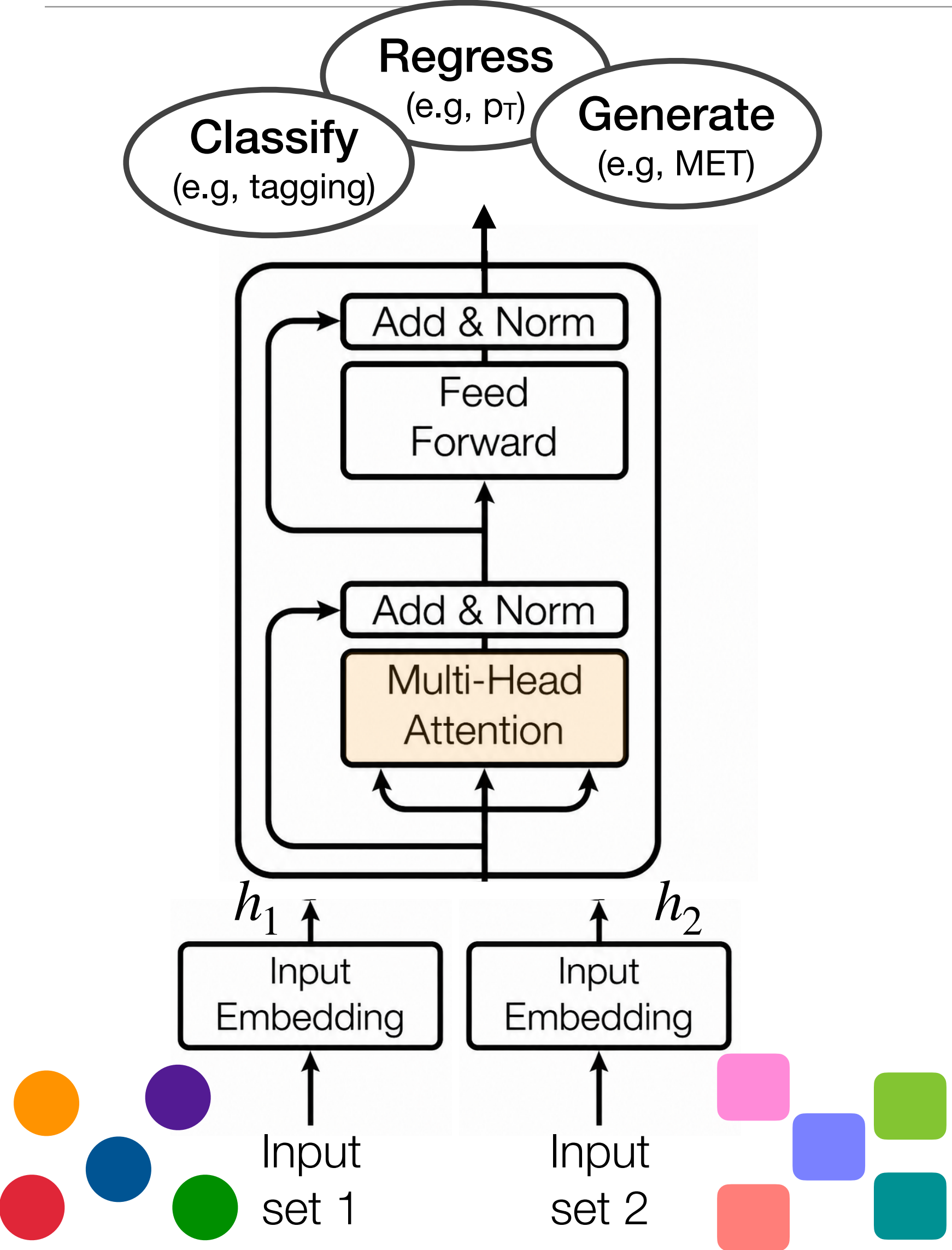
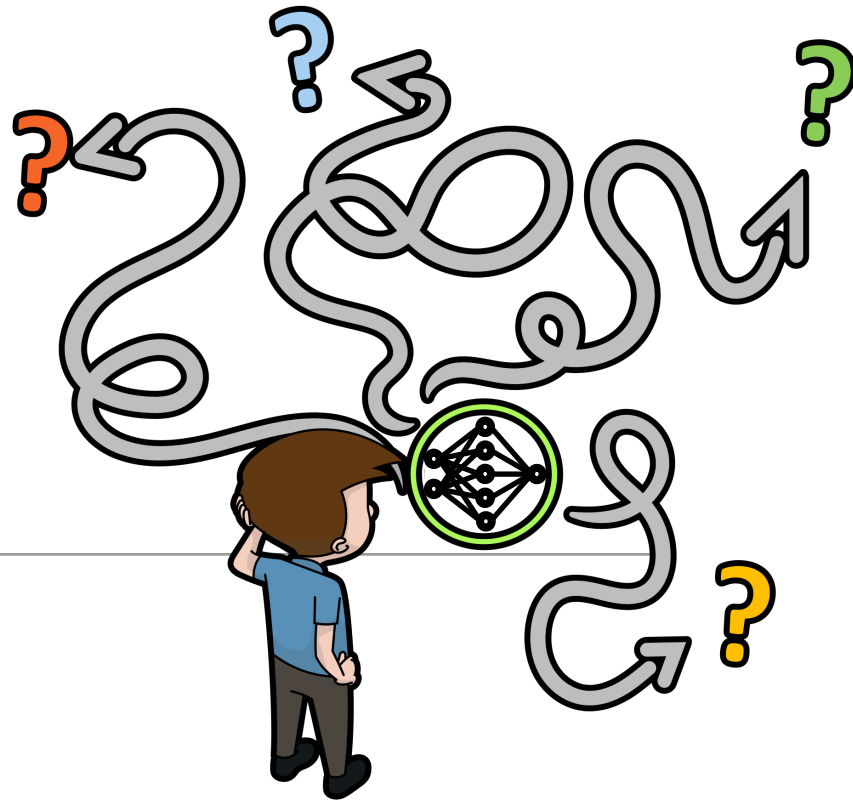


a) Self attention

All inputs condition all other input representations

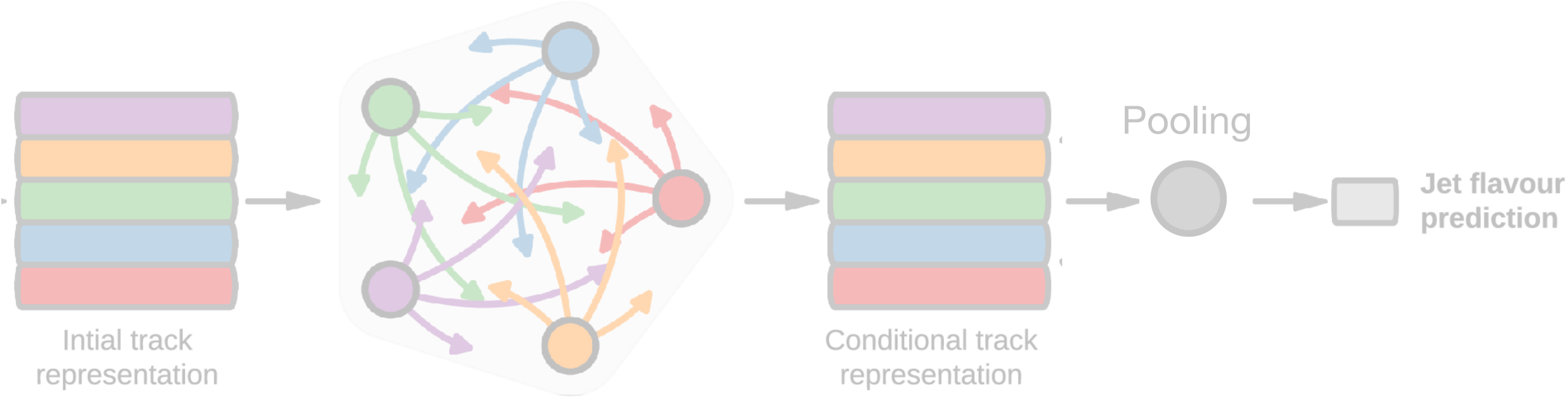


Transformer: our architecture of choice



a) Self attention

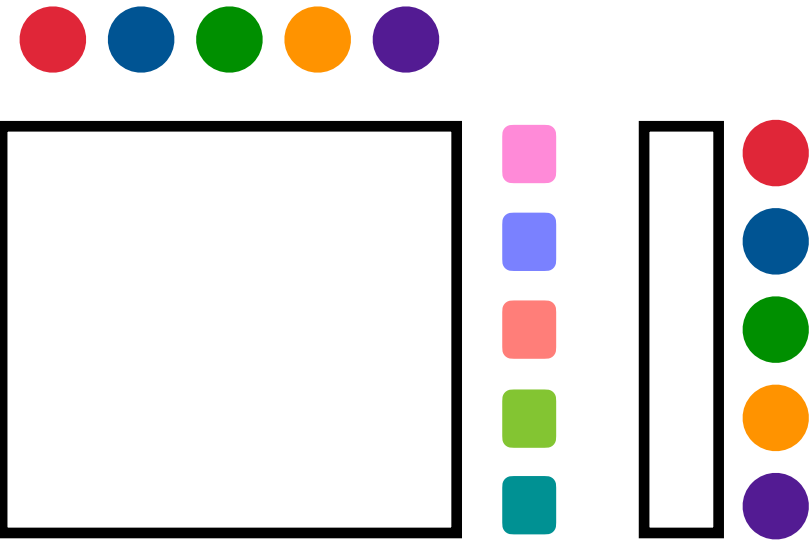
All inputs condition all other input representations



b) Cross attention

$$h_1 \leftarrow \text{Softmax} \left(\frac{q_1 k_2^T}{\sqrt{d_k}} \right) v_1$$

One input set conditions the other (constrain dimensionality of the attn matrix).



Outline



1/ Building our jets (inputs)



2/ Scaling up jet models



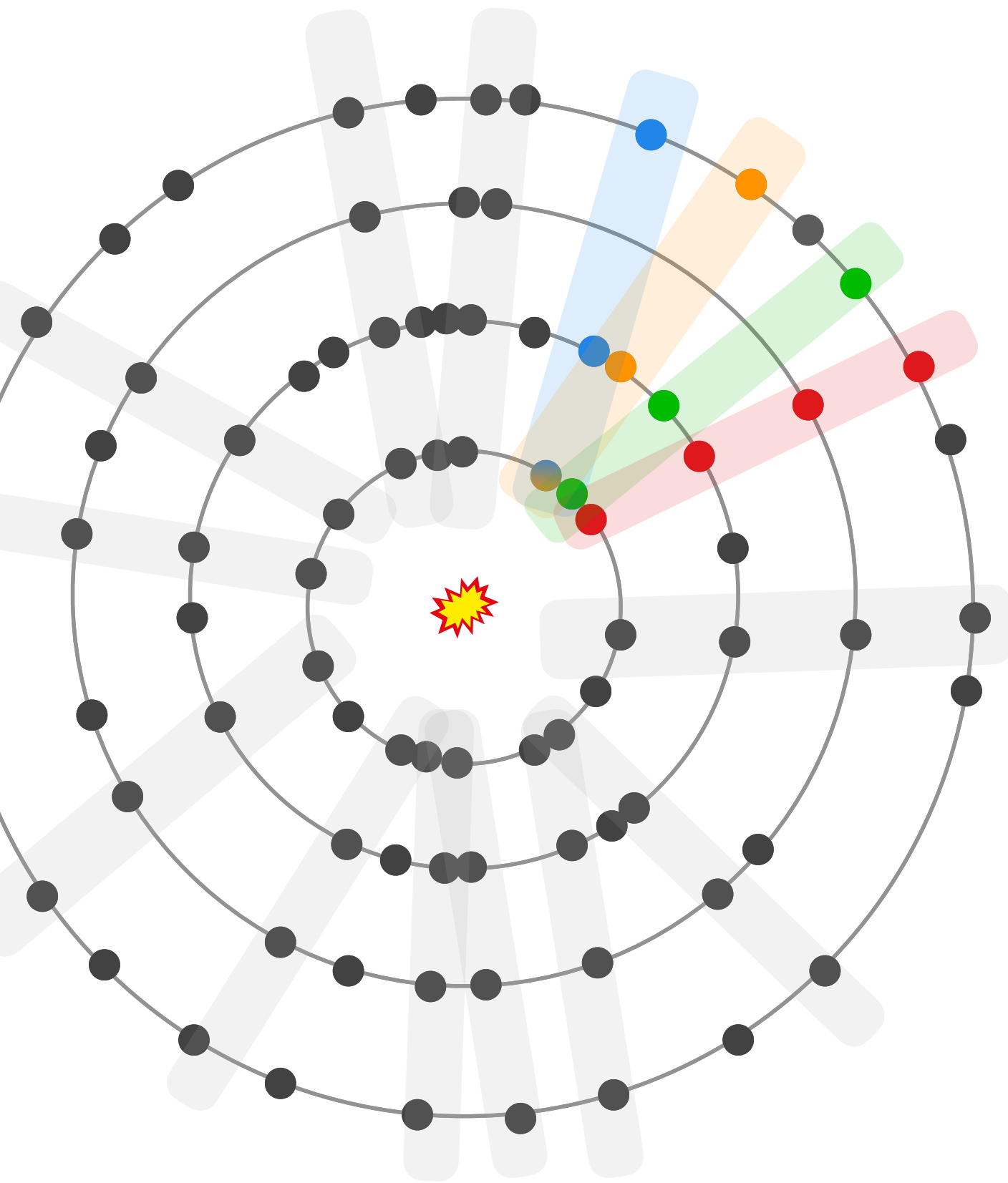
3/ Understanding jets (calibration)



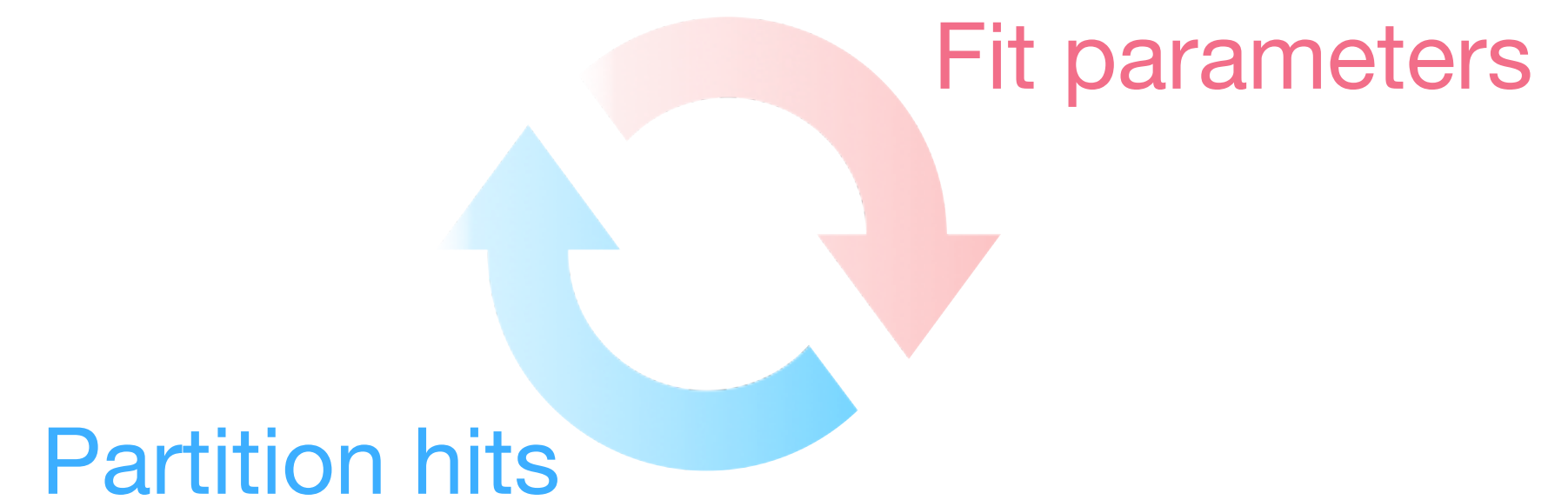
4/ Physics with our jets

Tracking

Task: Group hits into tracks

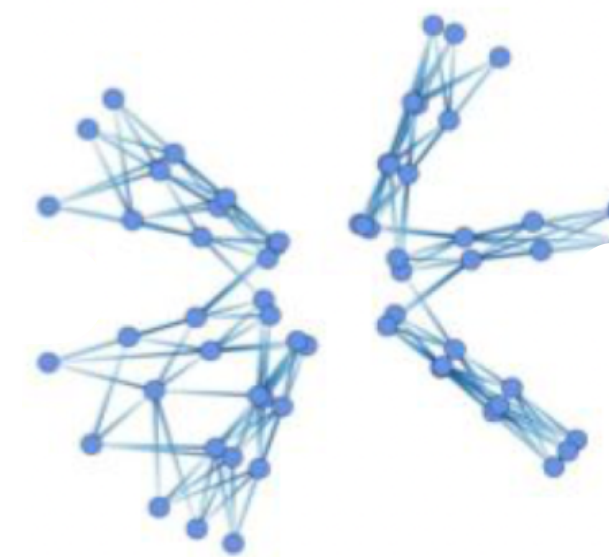


Traditionally: iterative fit



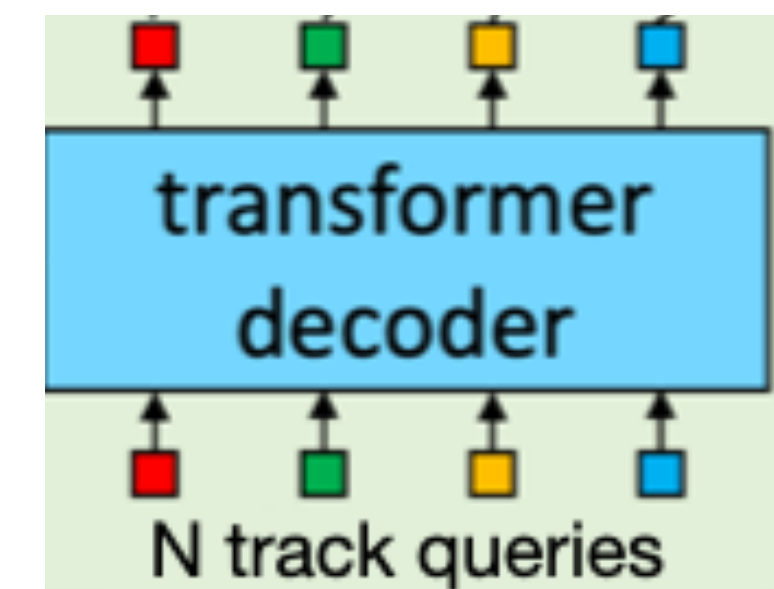
ML solutions

Graph NN



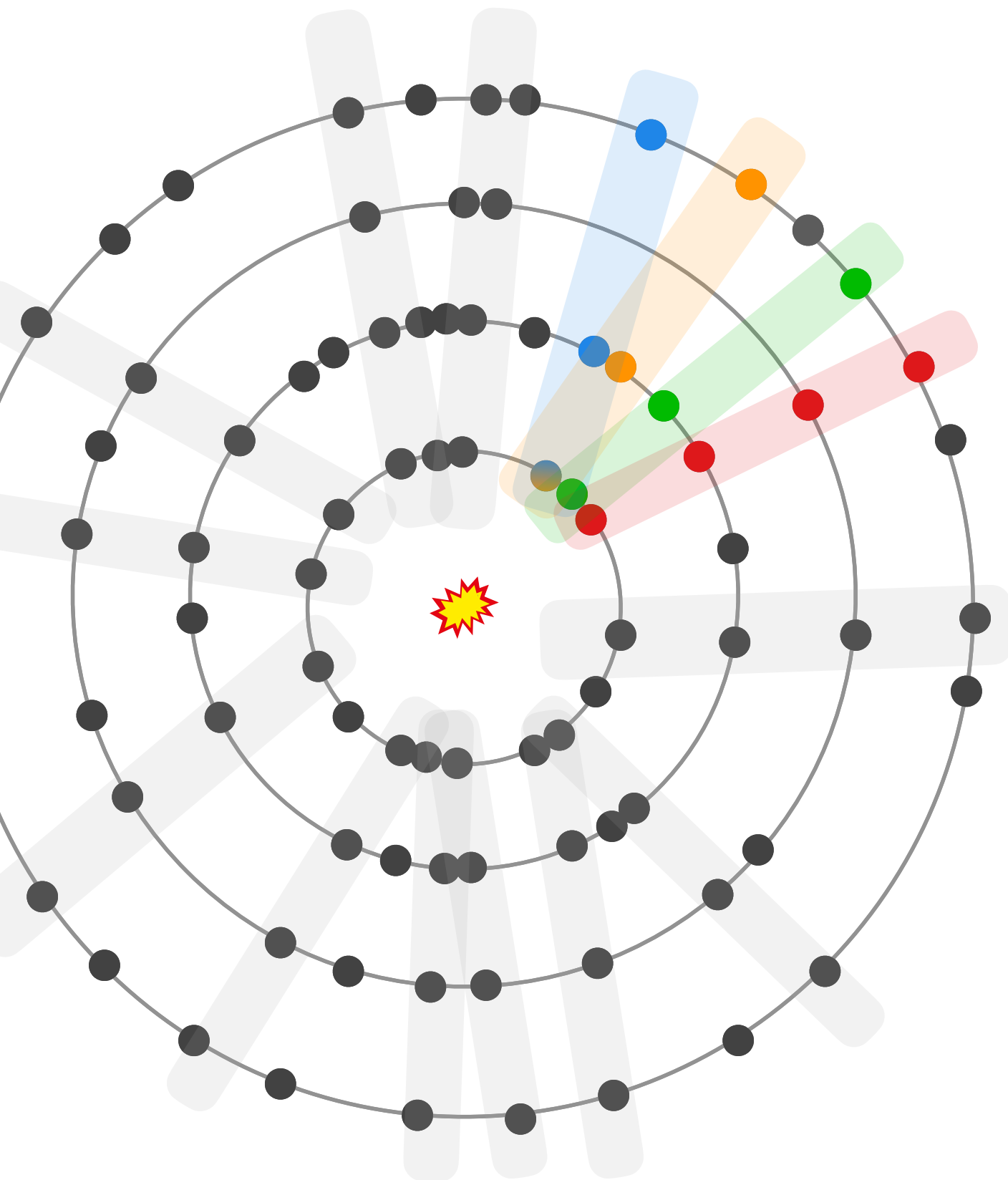
Sayed Heider Ali's [talk](#) tmr

Transformer

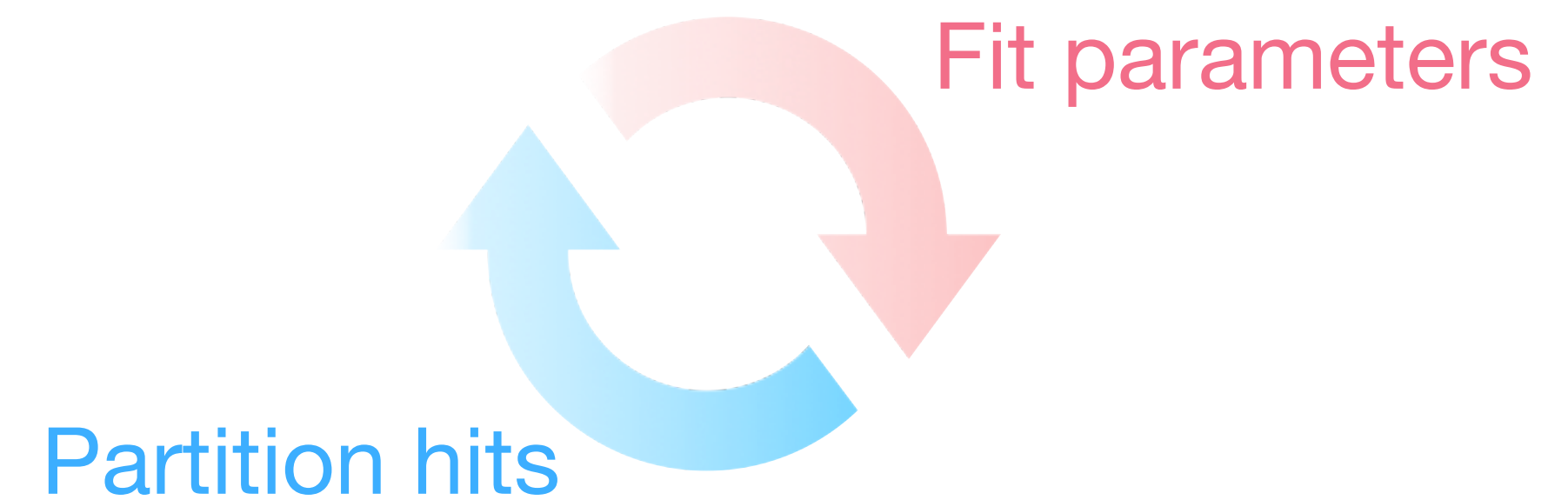


Tracking

Task: Group hits into tracks

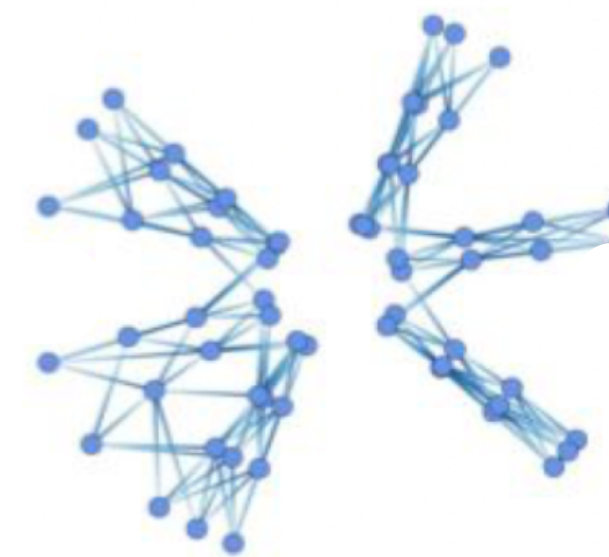


Traditionally: iterative fit



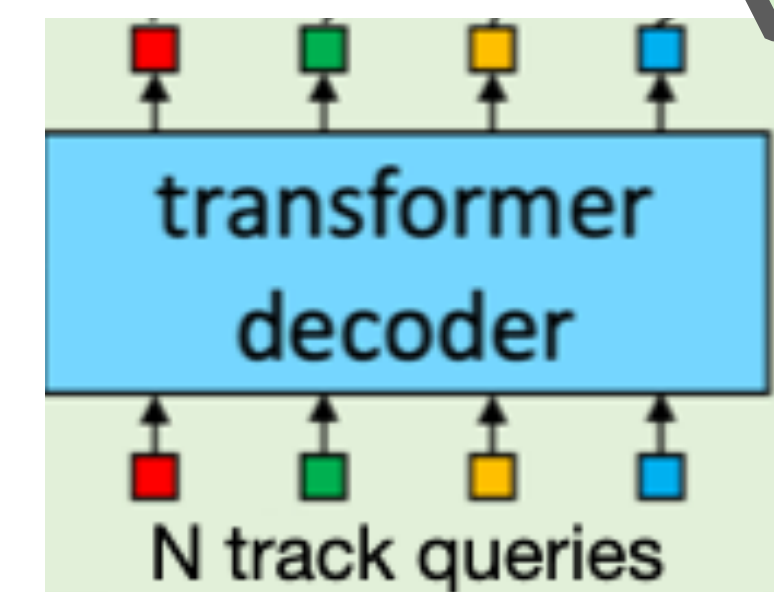
ML solutions

Graph NN

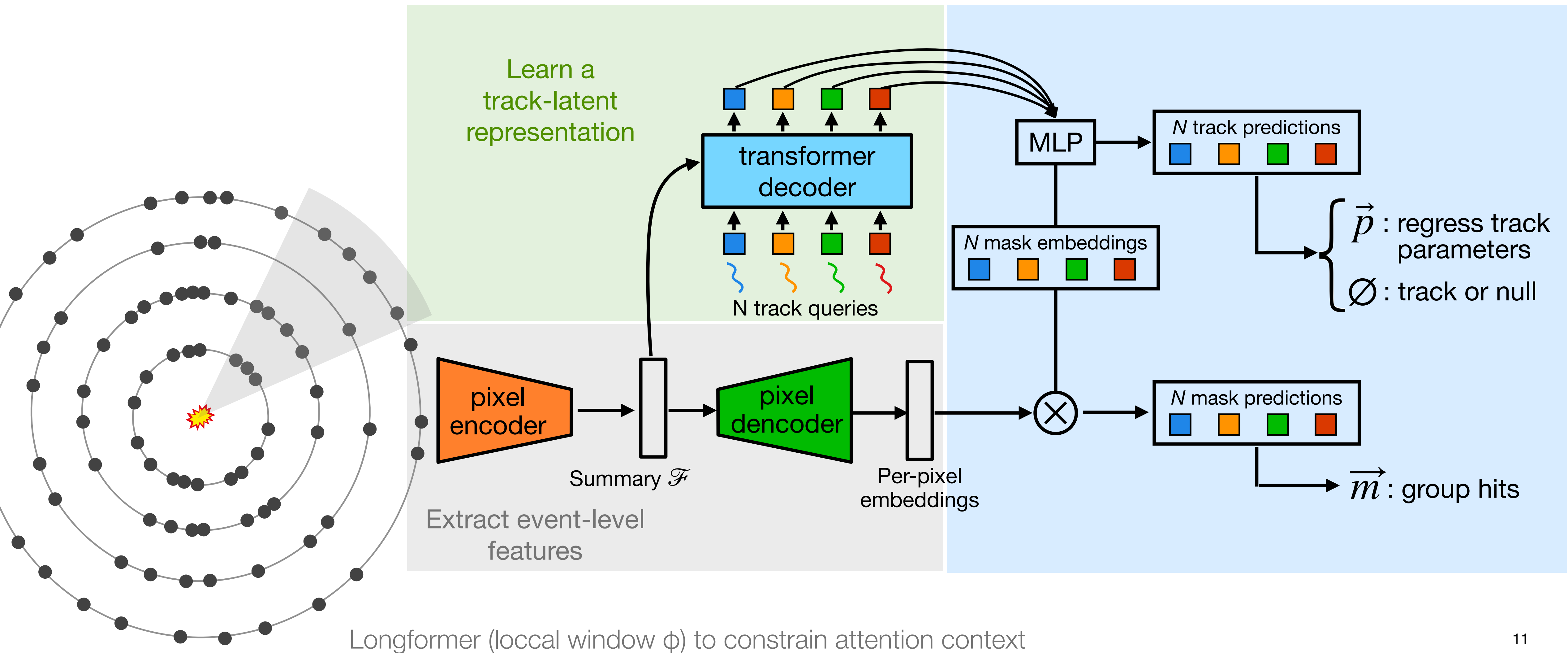


Sayed Heider Ali's [talk](#) tmr

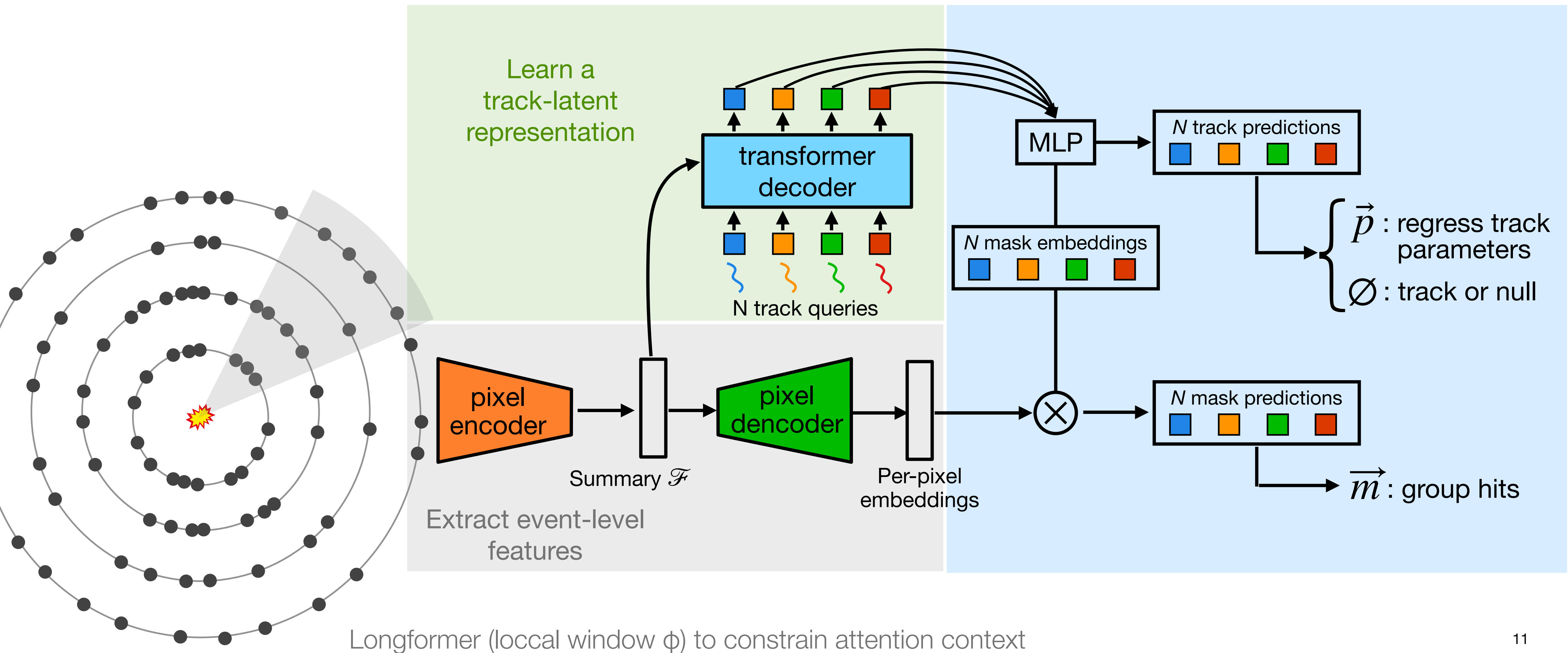
Transformer



Tracking: Maskformers

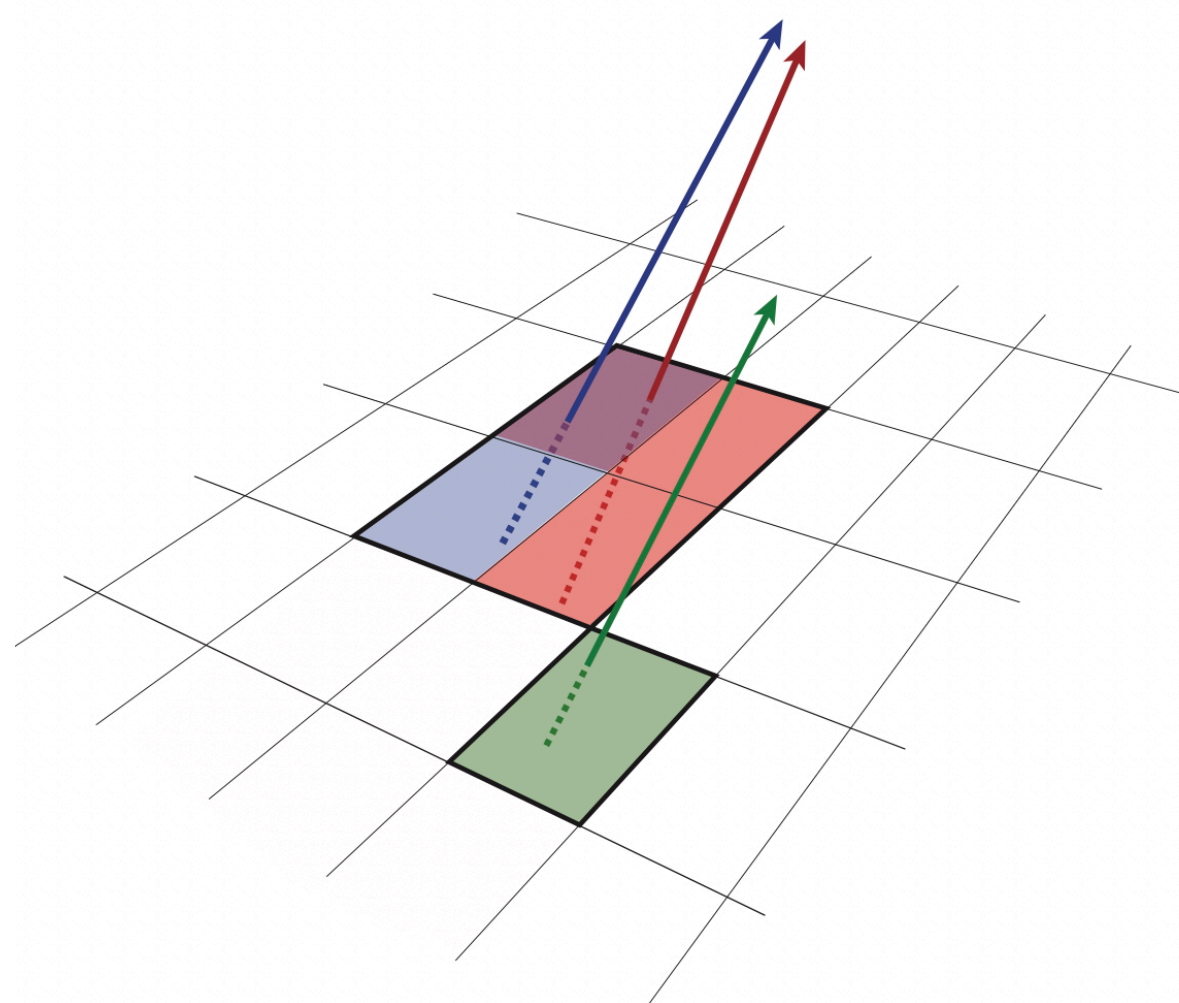


Tracking: Maskformers

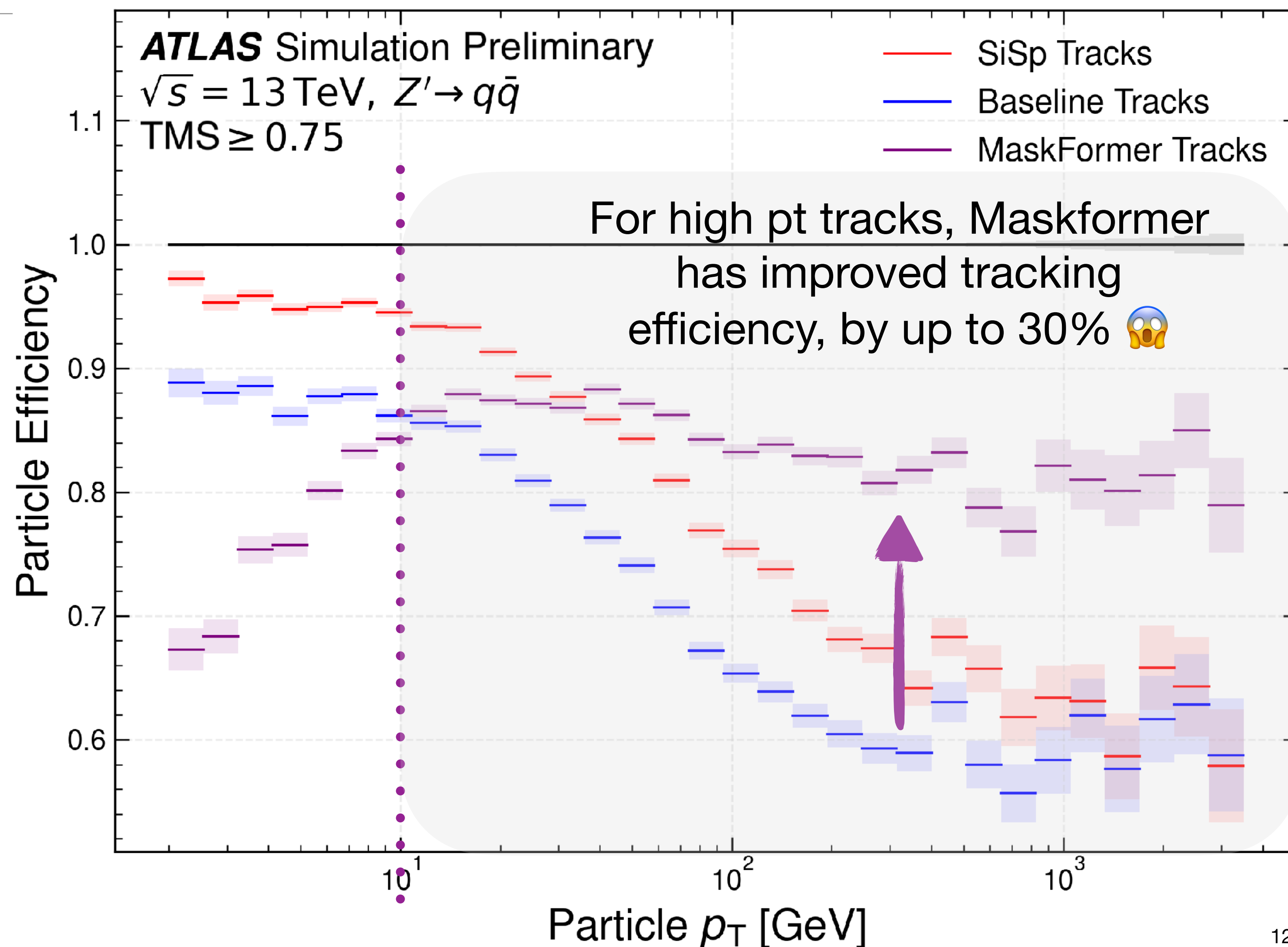


Tracking: Maskformers

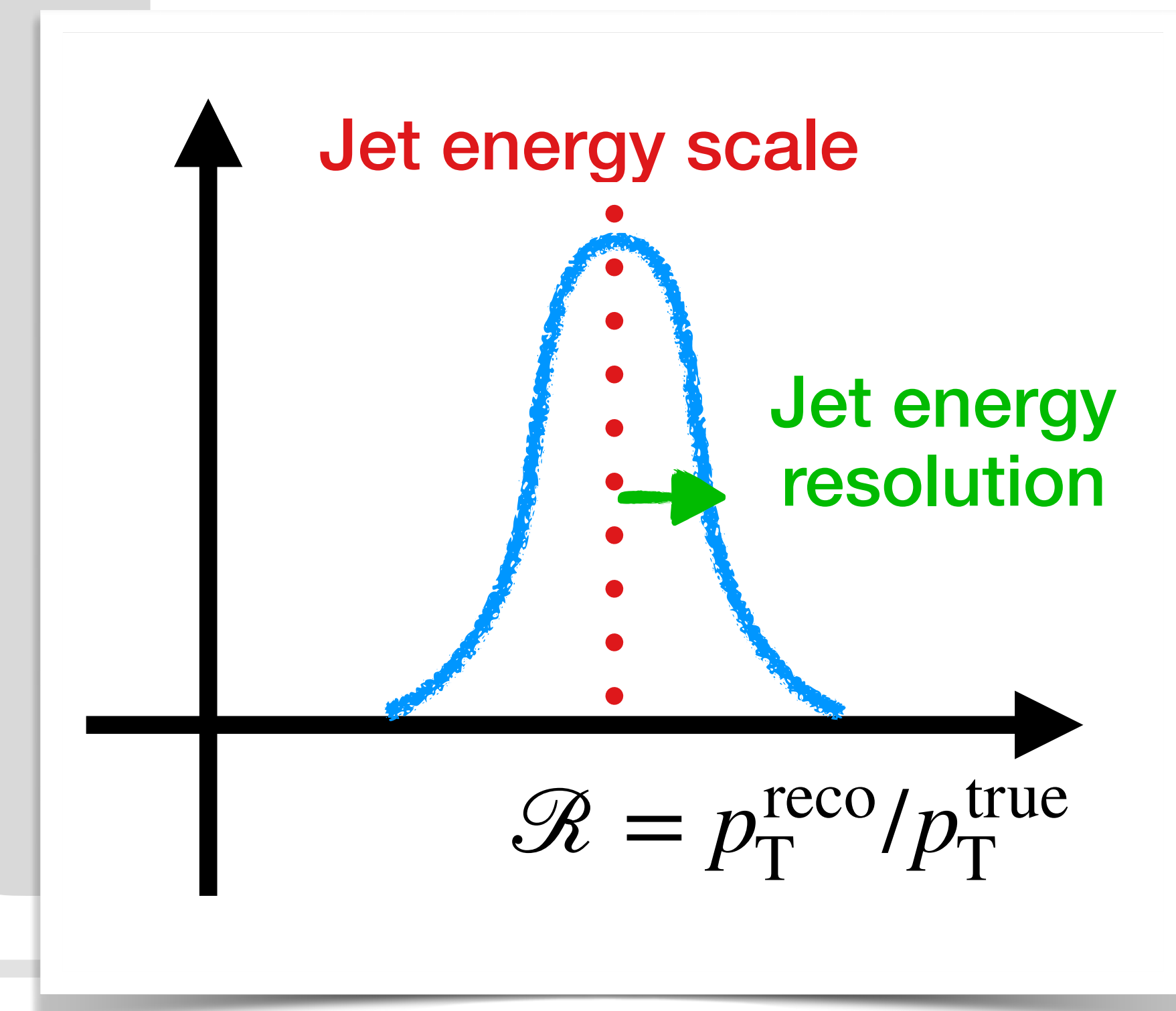
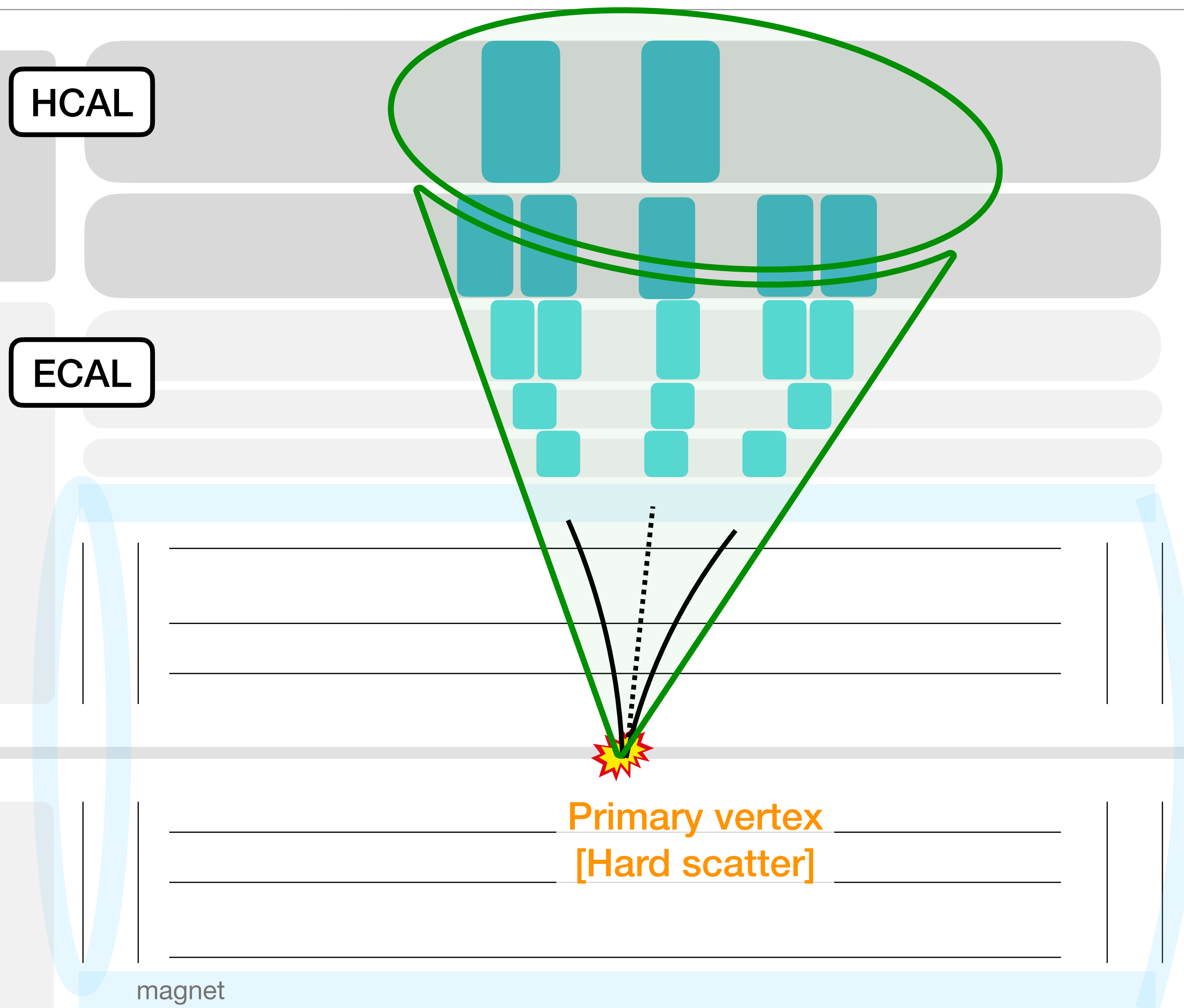
Application: Tracking in dense environments



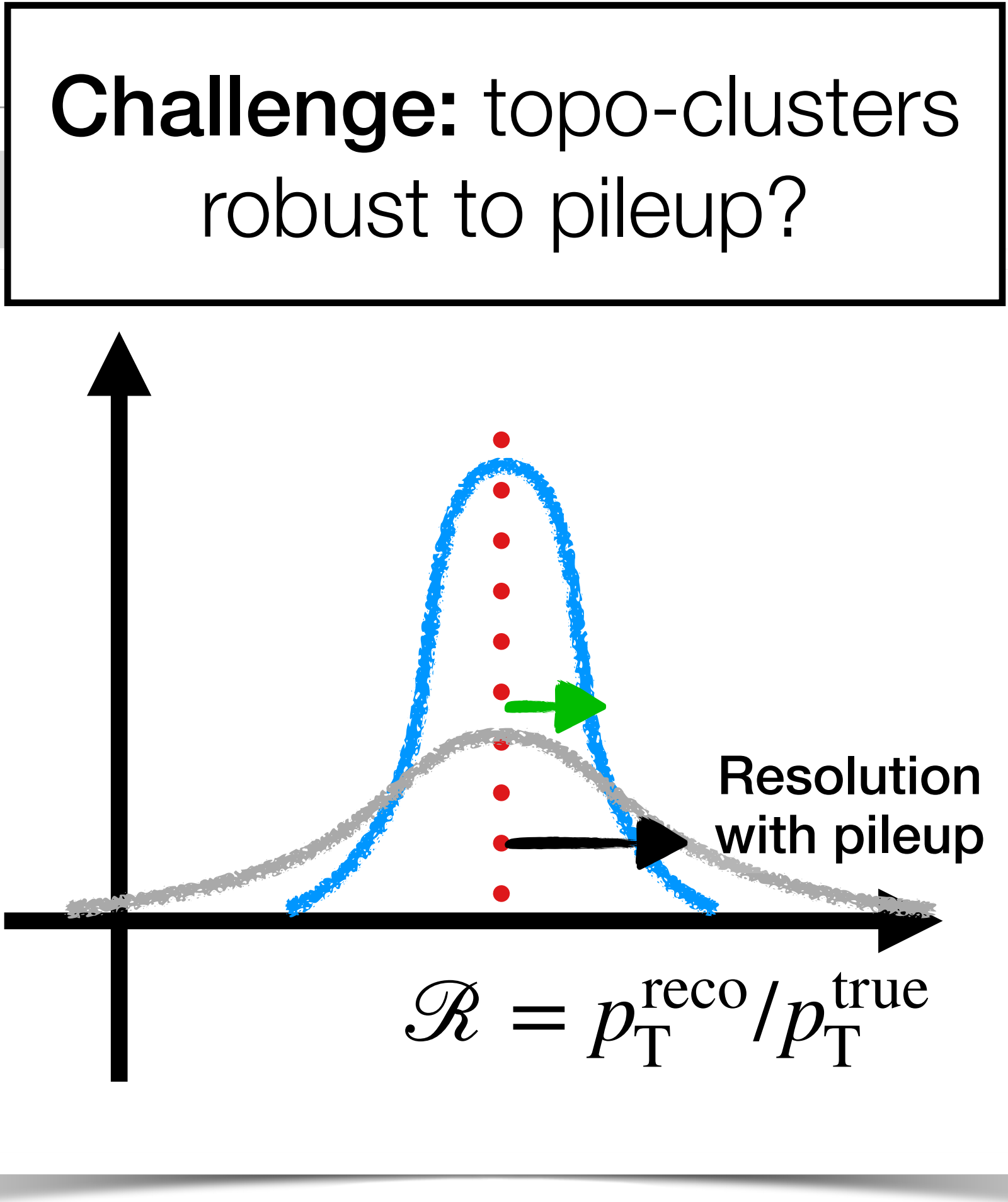
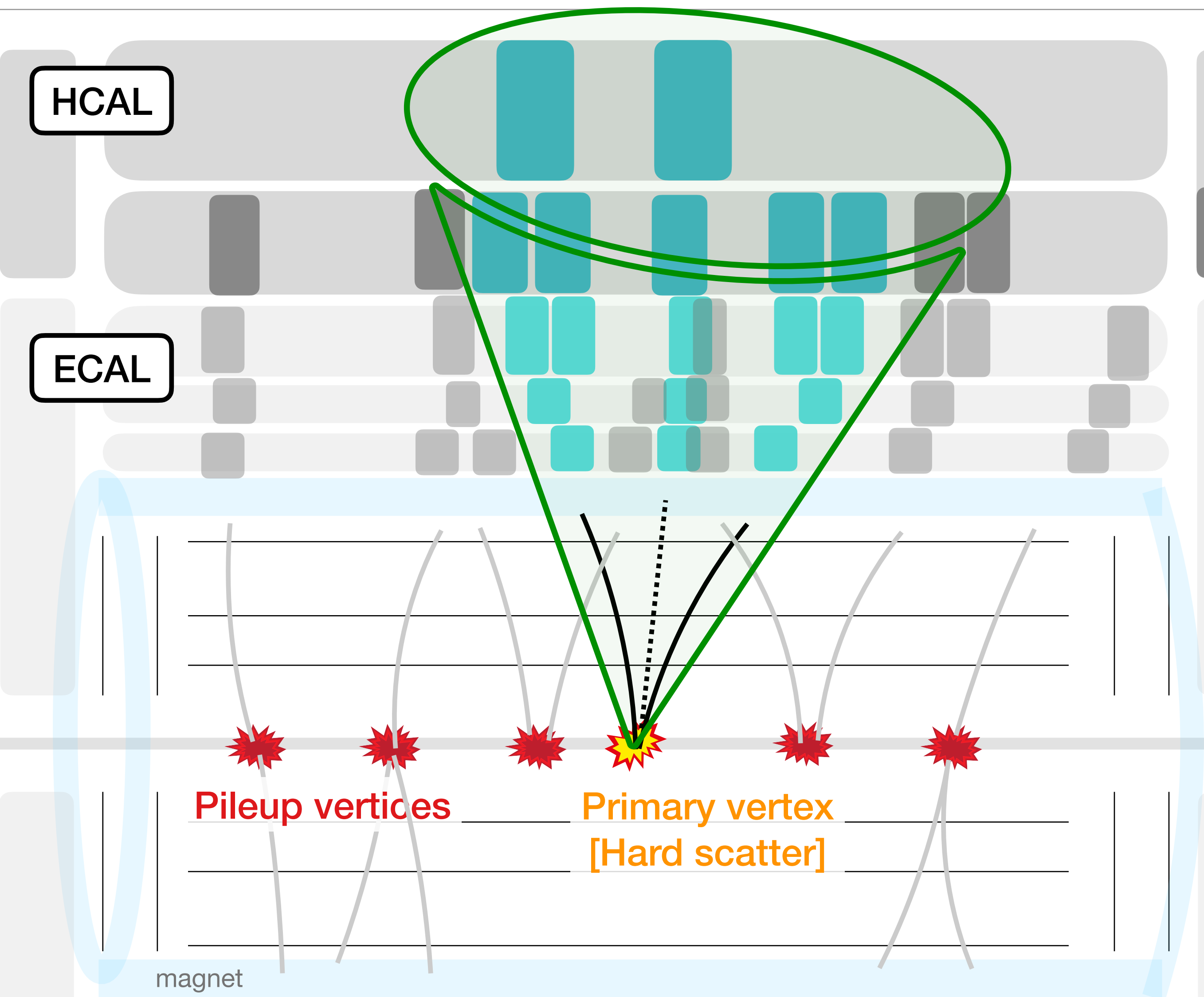
At the dense core of high p_T jets, tracks are more likely to *merge*



Challenge: Pileup robustness



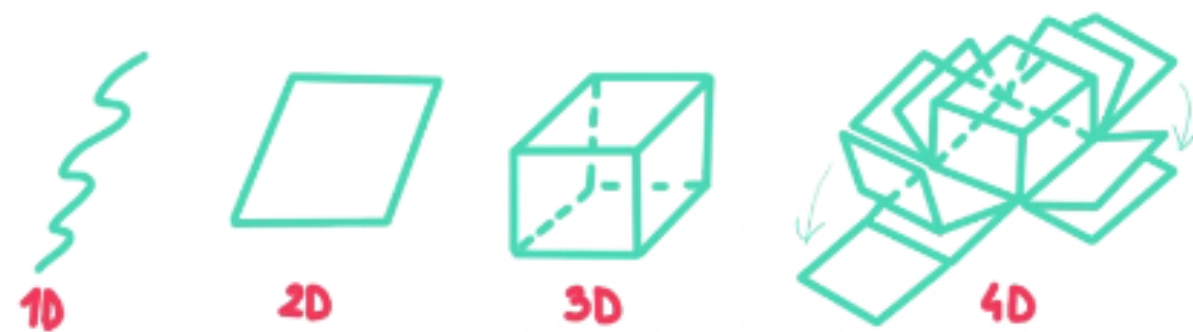
Challenge: Pileup robustness



More on PU mitigation in Nathan Suri's [talk](#)

Clustering (topocluster formation)

- 1 Cut on the cell time
EPJC 84 (2024) 455
→ Deployed in Run 3 reco



Img: V. Cairo

Challenge: topo-clusters
robust to pileup?

Clustering (topocluster formation)

1 Cut on the cell time

[EPJC 84 \(2024\) 455](#)

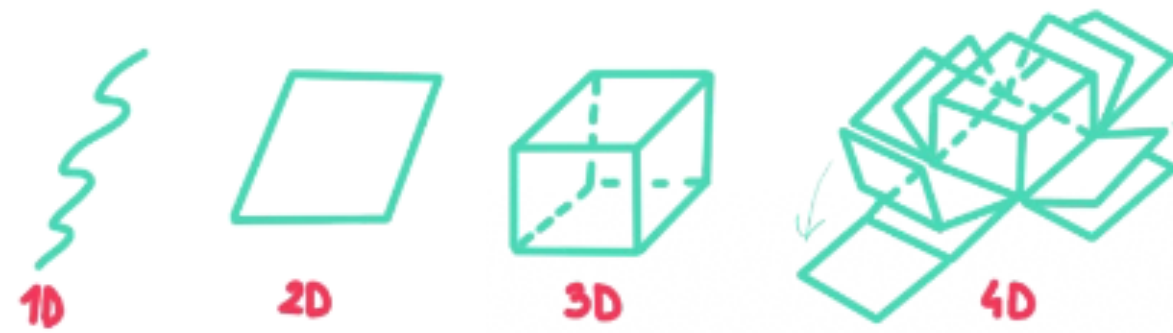
→ Deployed in Run 3 reco

2 Cut on the cluster error

[SciPost Phys 19 \(2025\) 155](#) + 

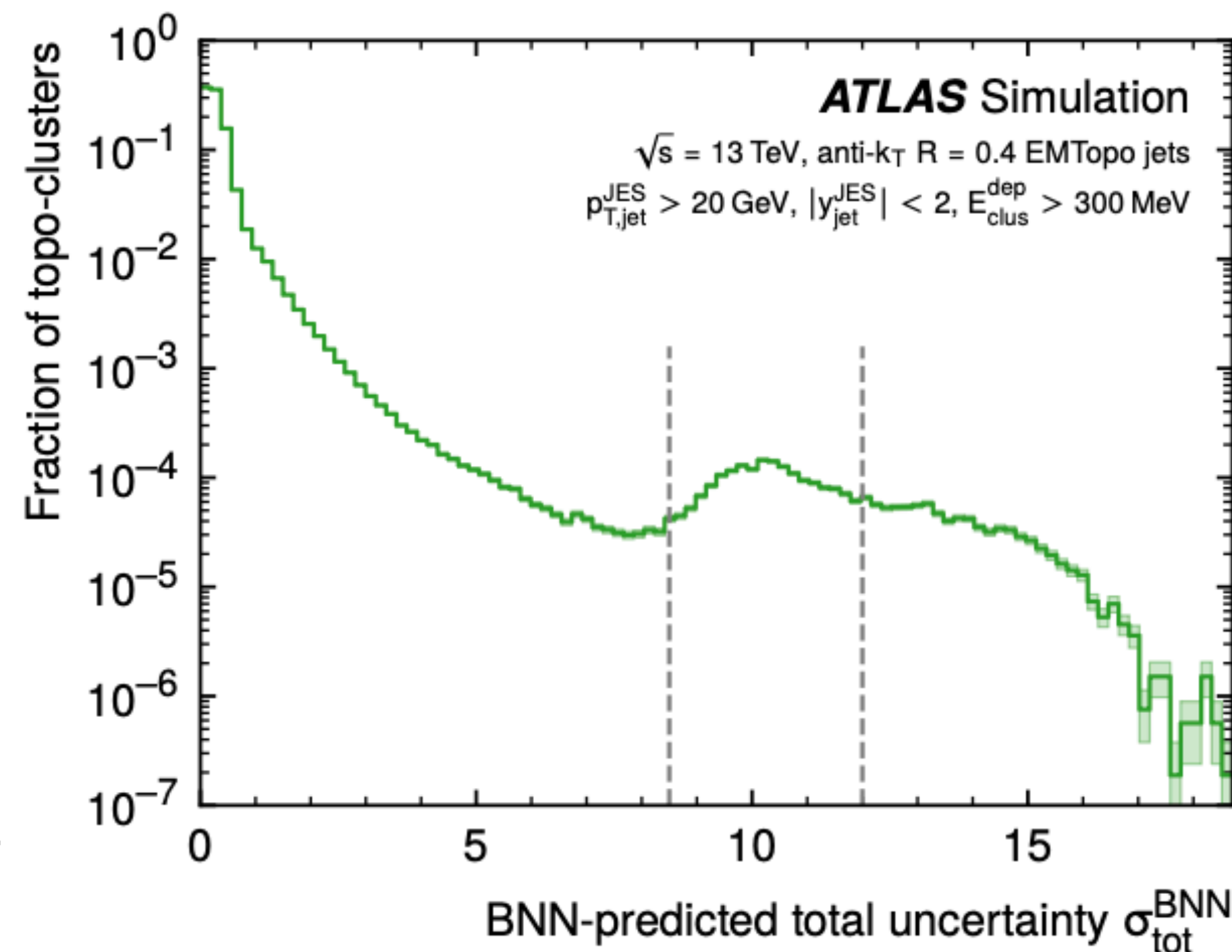
Bayesian Neural Networks /
Mixture Density Networks
define an error on the
prediction

More on calib NN unc in Lorenz's [talk](#) this morning!



Img: V. Cairo

Challenge: topo-clusters
robust to pileup?



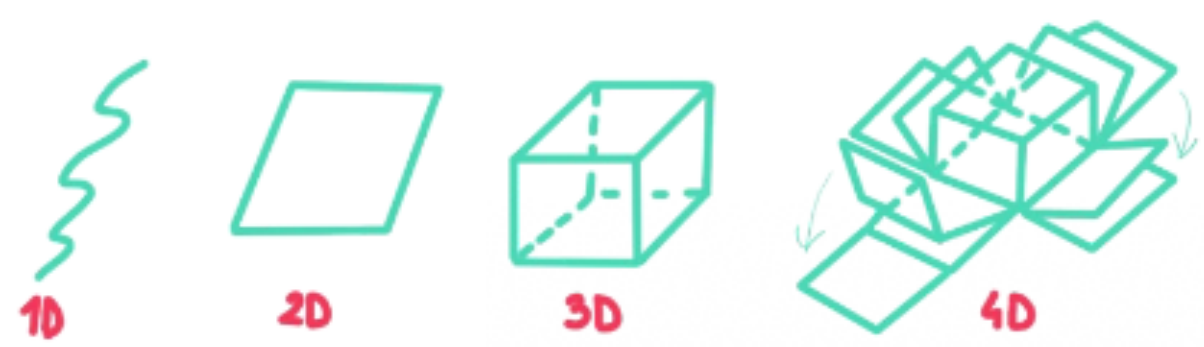
Clustering (topocluster formation)

Challenge: topo-clusters robust to pileup?

1 Cut on the cell time

EPJC 84 (2024) 455

→ Deployed in Run 3 reco



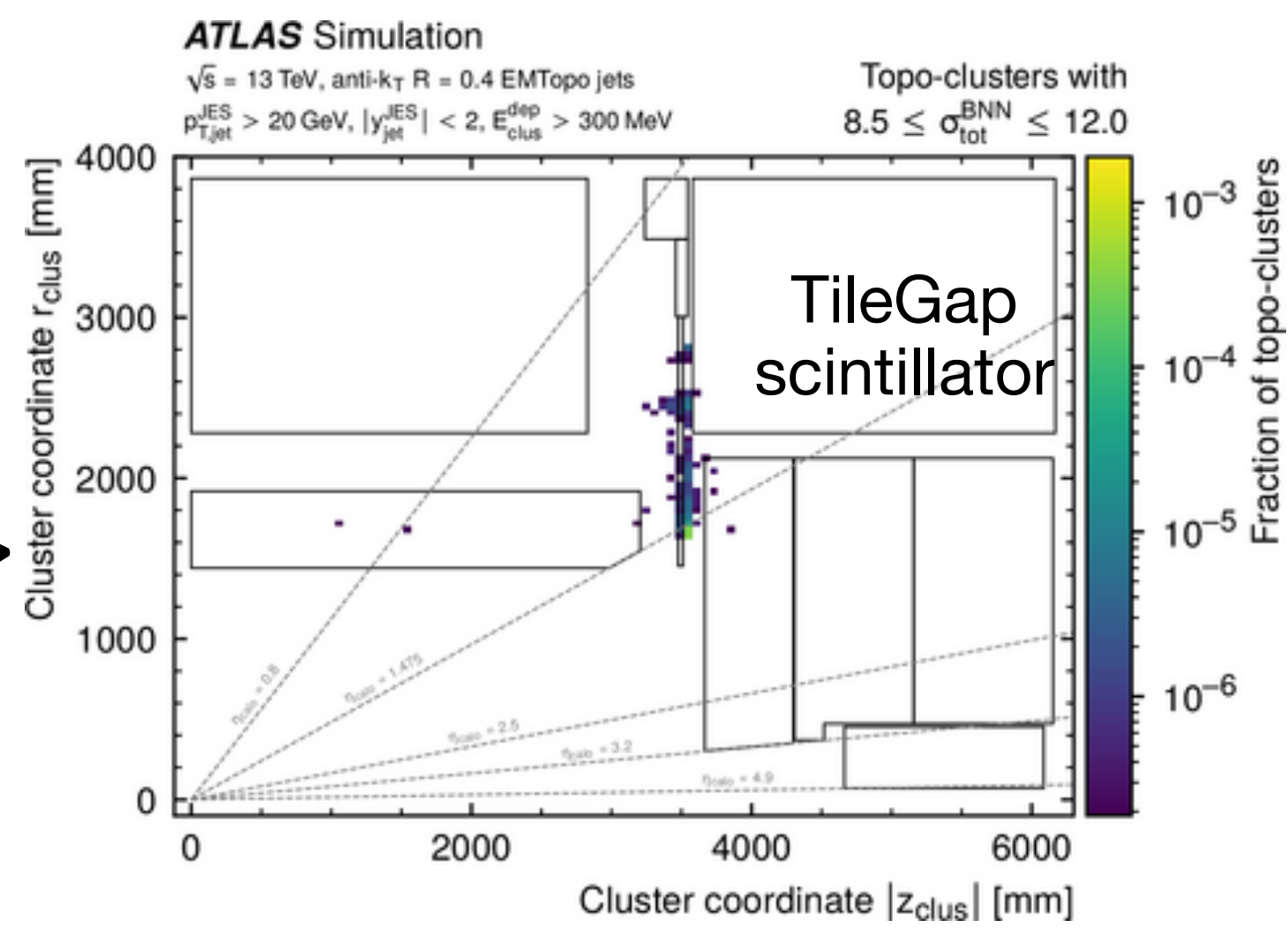
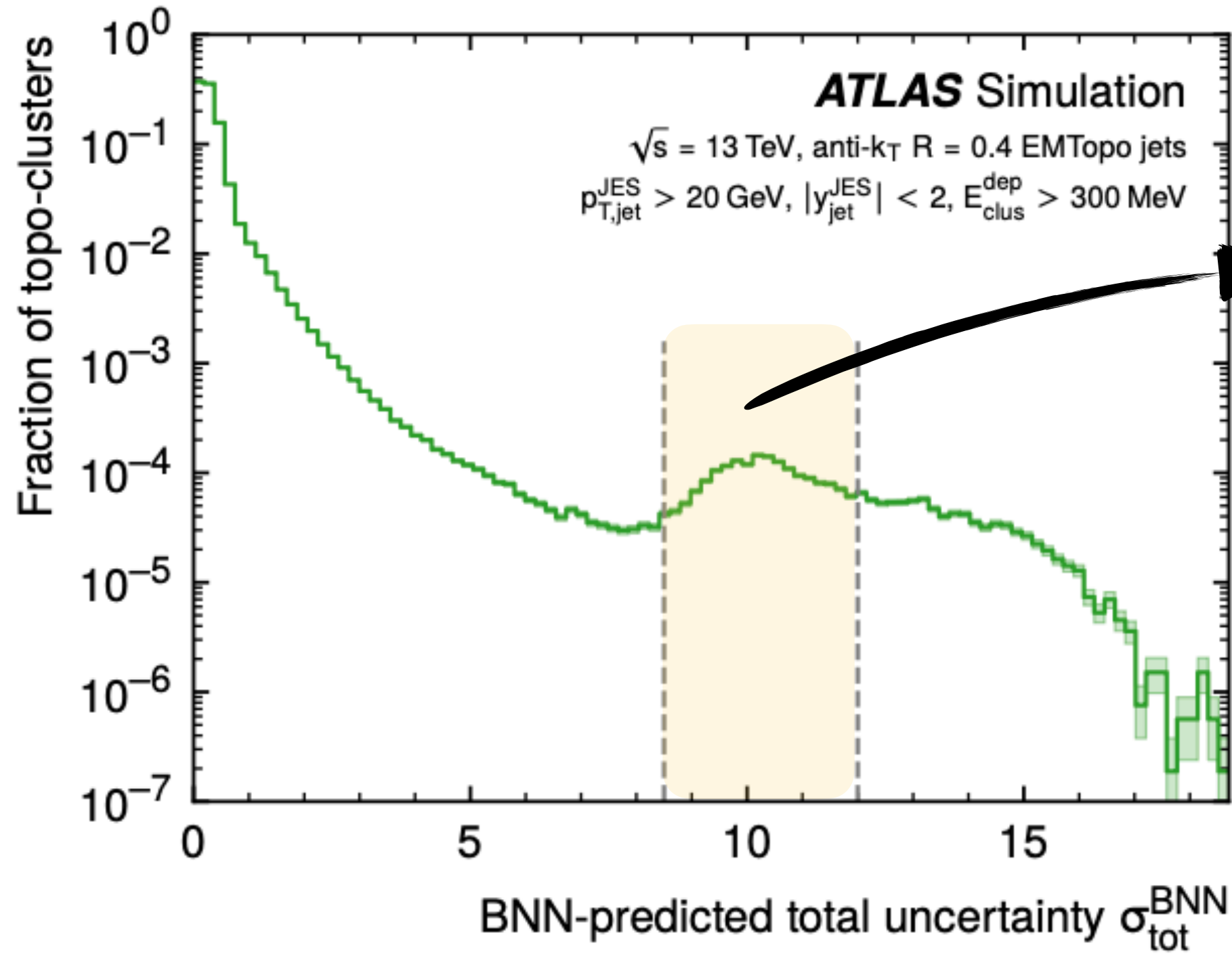
Img: V. Cairo

2 Cut on the cluster error

SciPost Phys 19 (2025) 155 + 

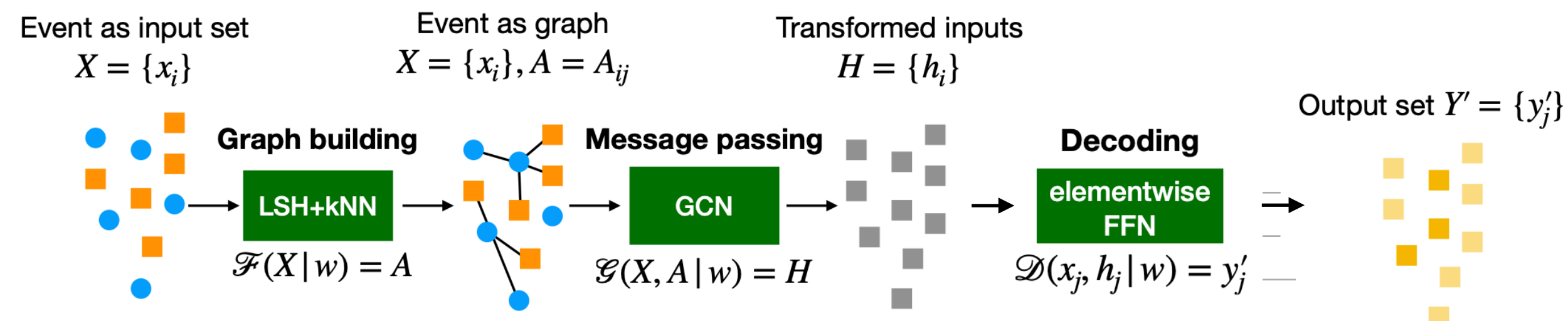
Bayesian Neural Networks /
Mixture Density Networks
define an error on the
prediction

More on calib NN unc in Lorenz's [talk](#) this morning!



Revealed missing
dead material!

Particle Flow

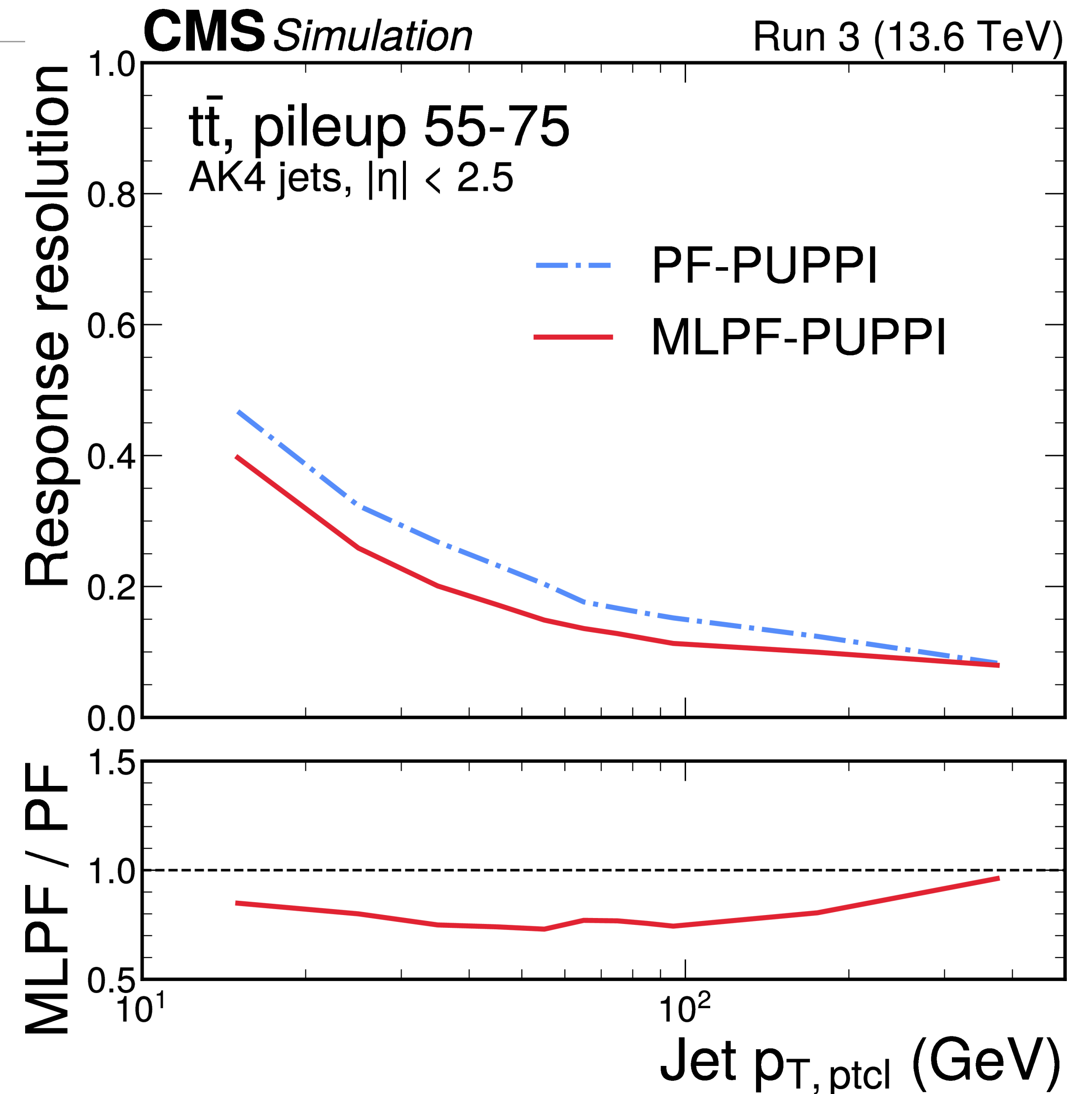


Graphic from J. Pata, et. al., EJP (2021)

Evolution: now using transformer (with global self-attention) instead of a GNN.

Impact: Improved JER

Opportunity: 5x faster than traditional approach.

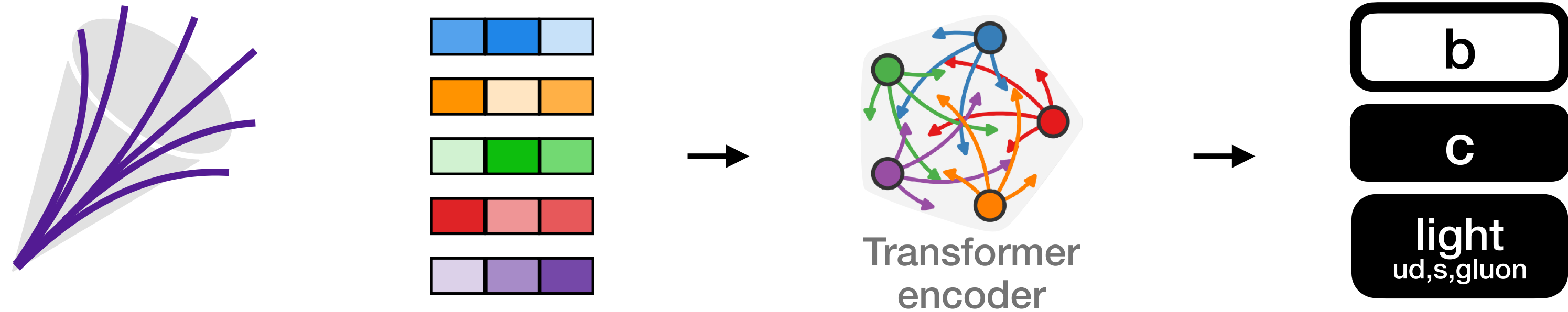


Tagging: inputs

Before (2025): Bespoke models, targetted inputs for each task

Task 1: b-tagging
Inputs: Tracker

[Eur. Phys. J. C 83 \(2023\) 681](#)
[Nature Commun. 17 \(2026\) 541](#)

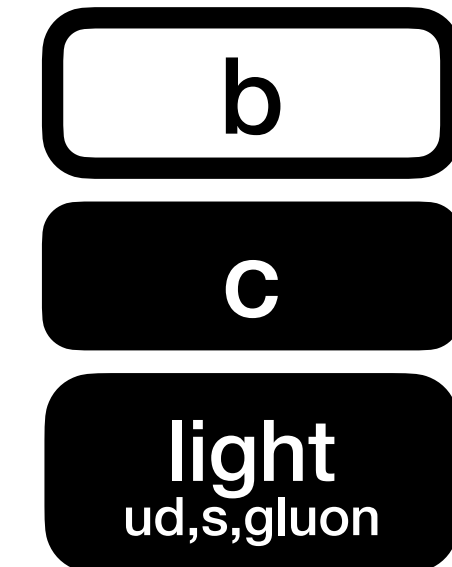
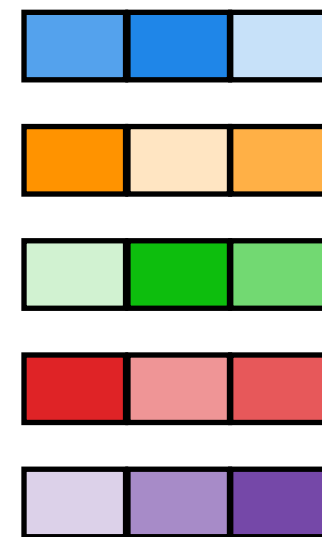


Tagging: inputs

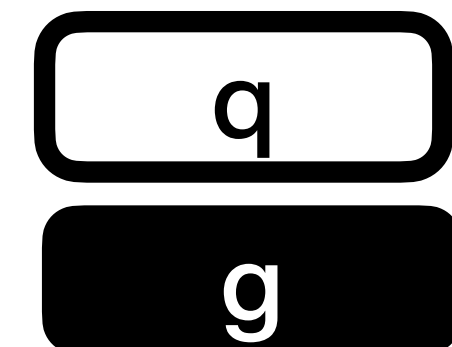
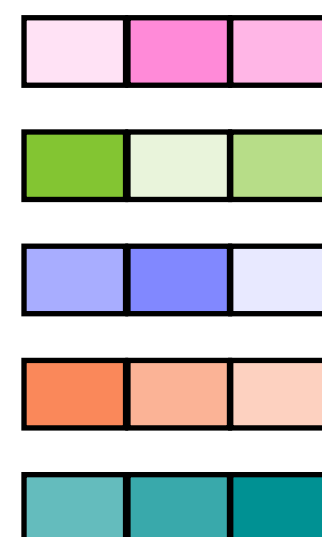
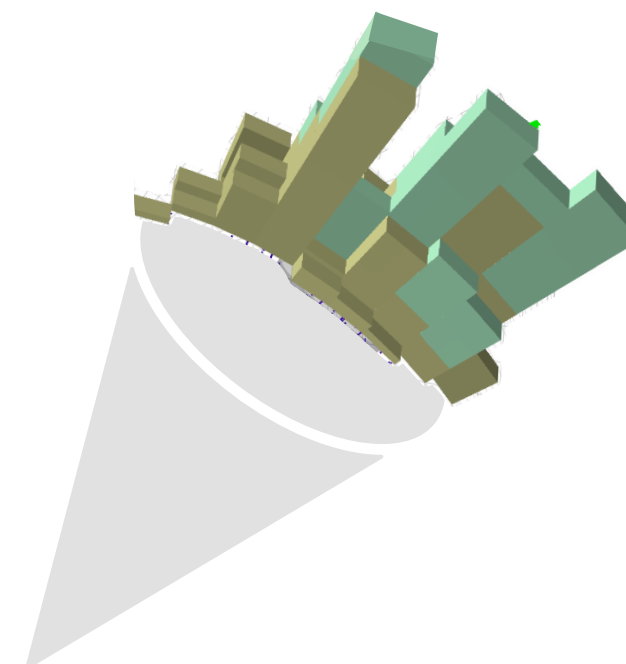
Before (2025): Bespoke models, targetted inputs for each task

Task 1: b-tagging
Inputs: Tracker

[Eur. Phys. J. C 83 \(2023\) 681](#)
[Nature Commun. 17 \(2026\) 541](#)

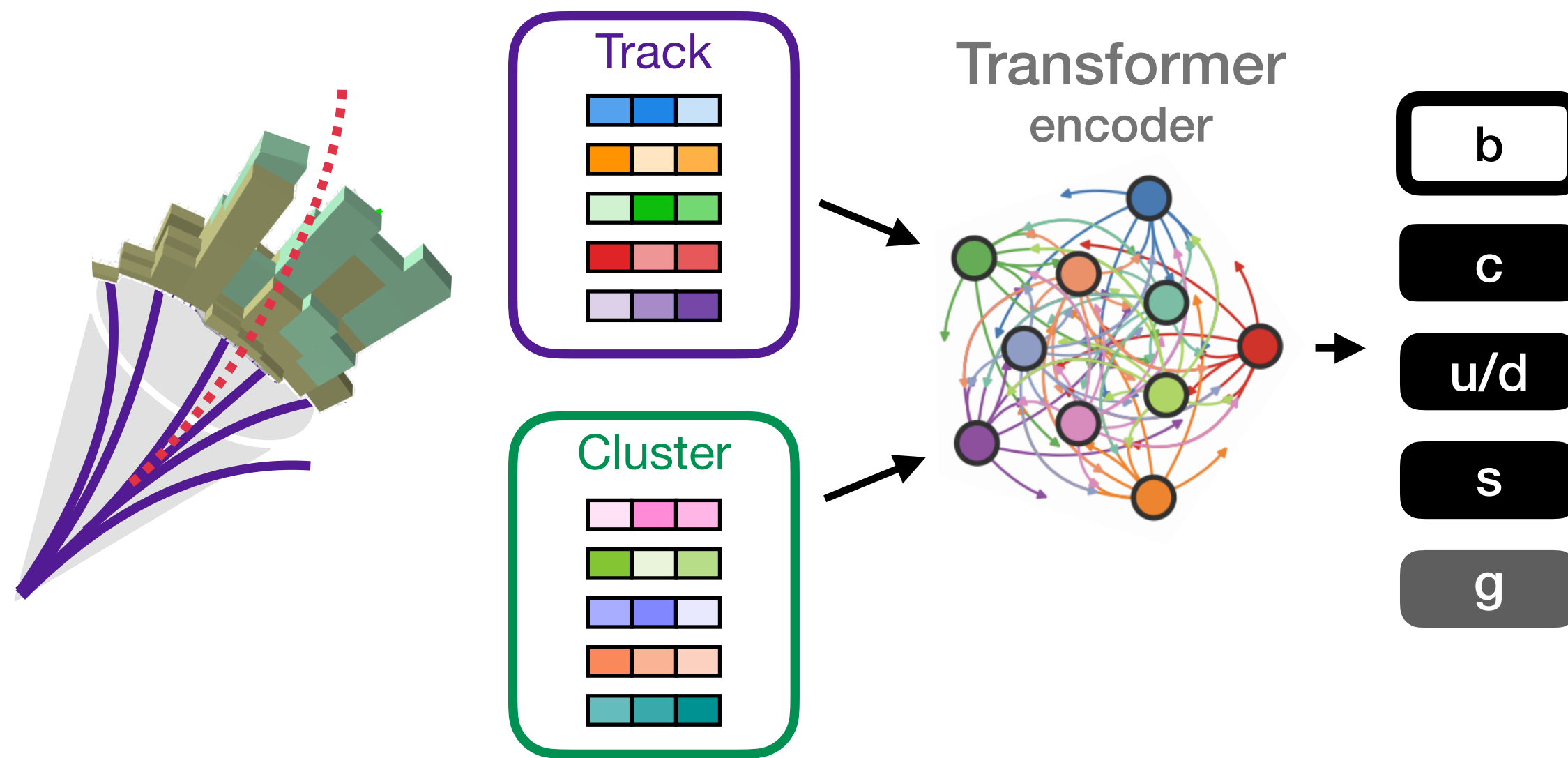


Task 2: q/g-tagging
Inputs: Calorimeter

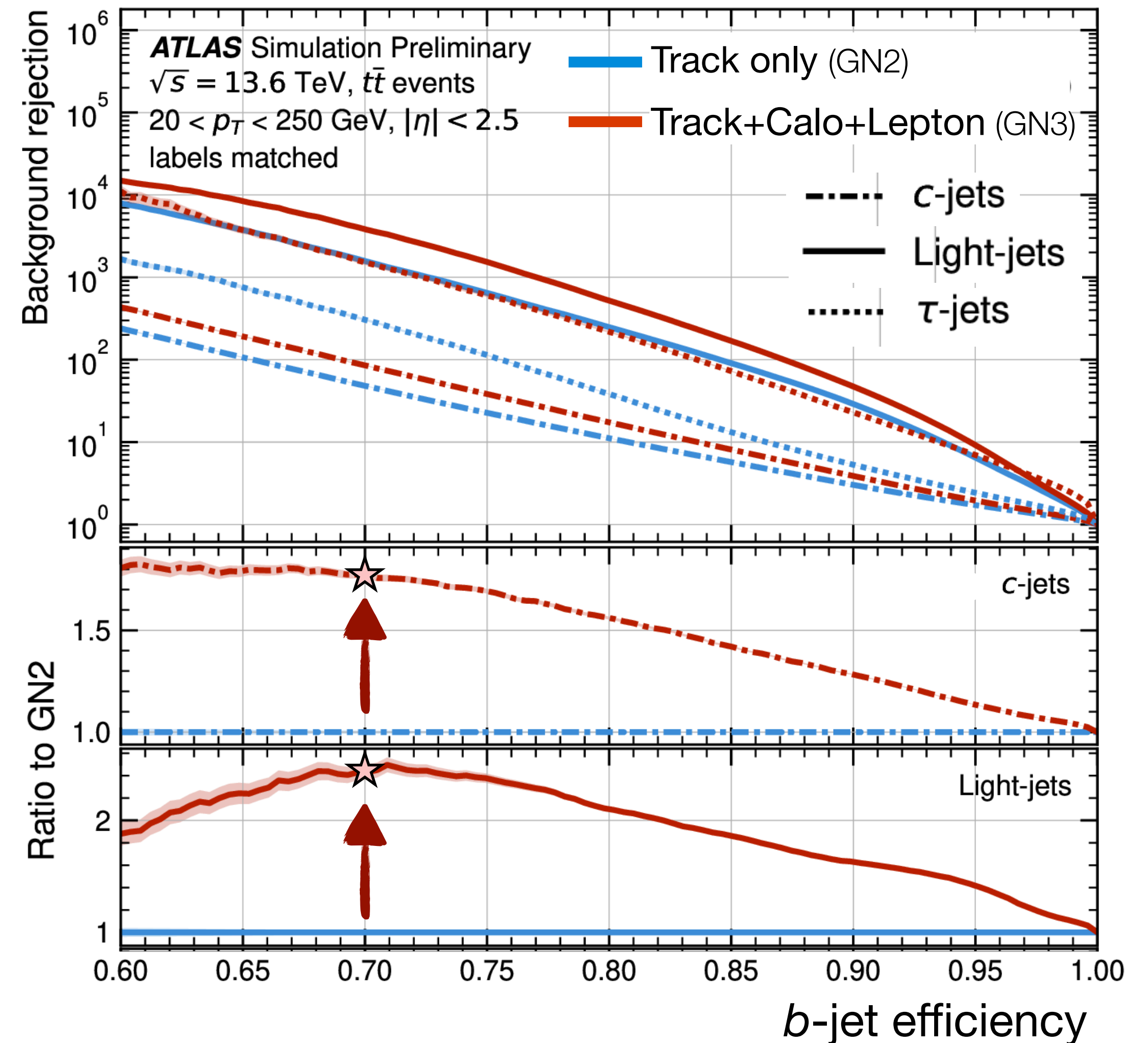


Tagging: multimodal inputs

Now (2026): Unified inputs = general purpose models



2x improvement in background rejection.



More in Alexander's [talk](#) this afternoon!

[ATL-PHYS-2026-001](#) ; also for H(bb) [ATL-PHYS-2026-021](#)

Circa 2020: Natural Language Processing

Tell information

Generative
Pre
Training



Ask questions

Comparing w/ size of SOTA
jet tagging models

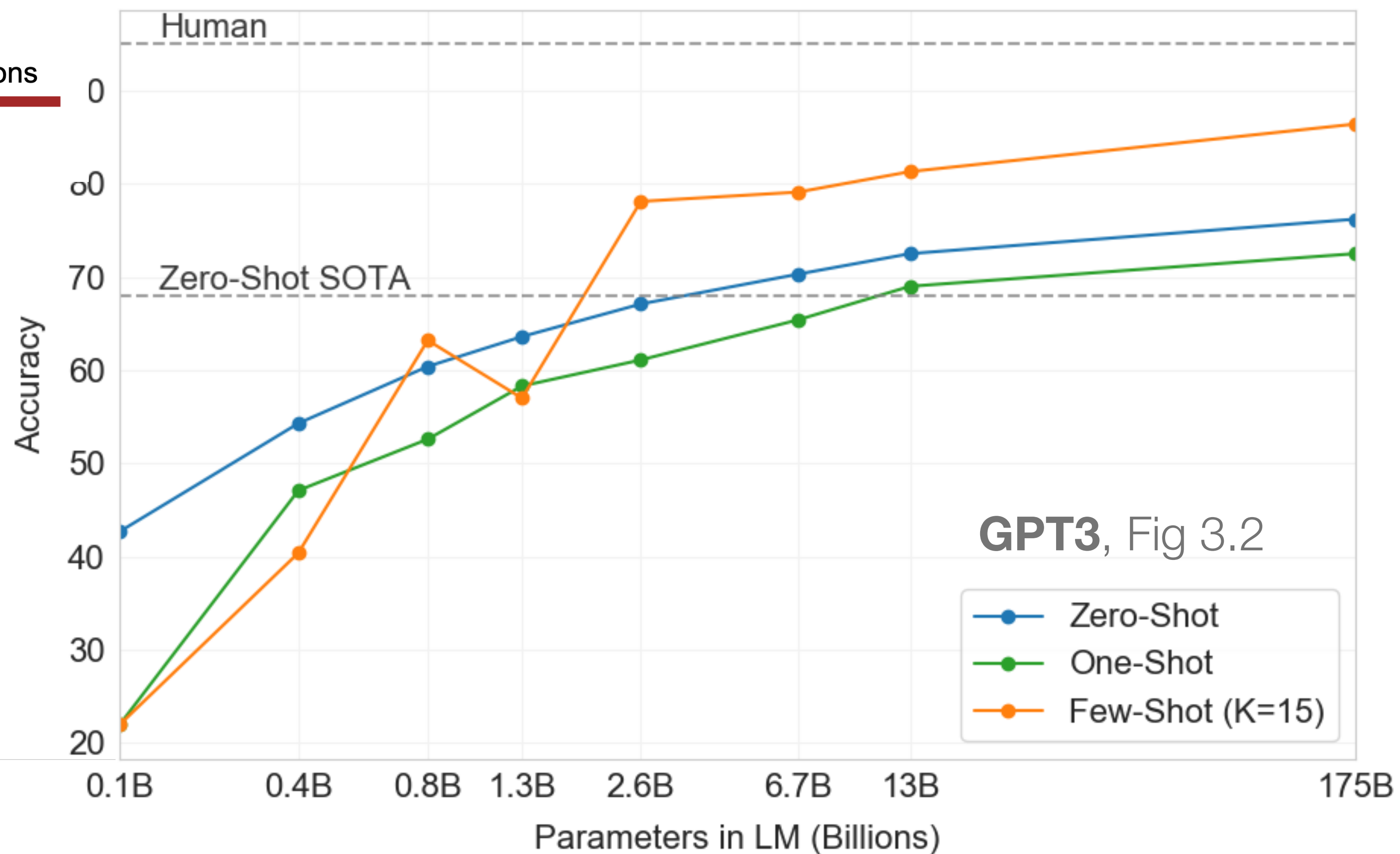
GN2
(2m)

GN3
(13m)

UParTv1
(2m)

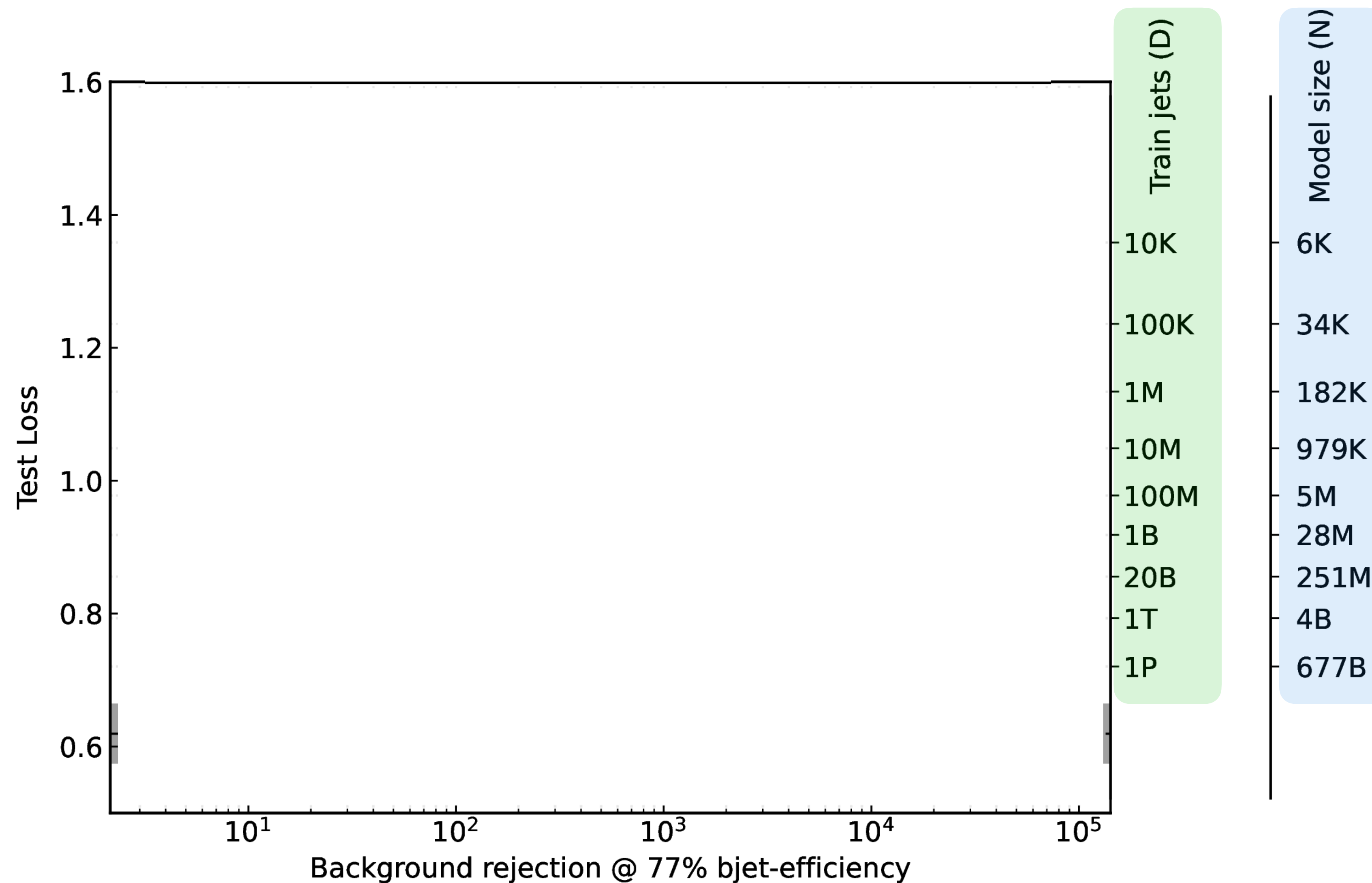
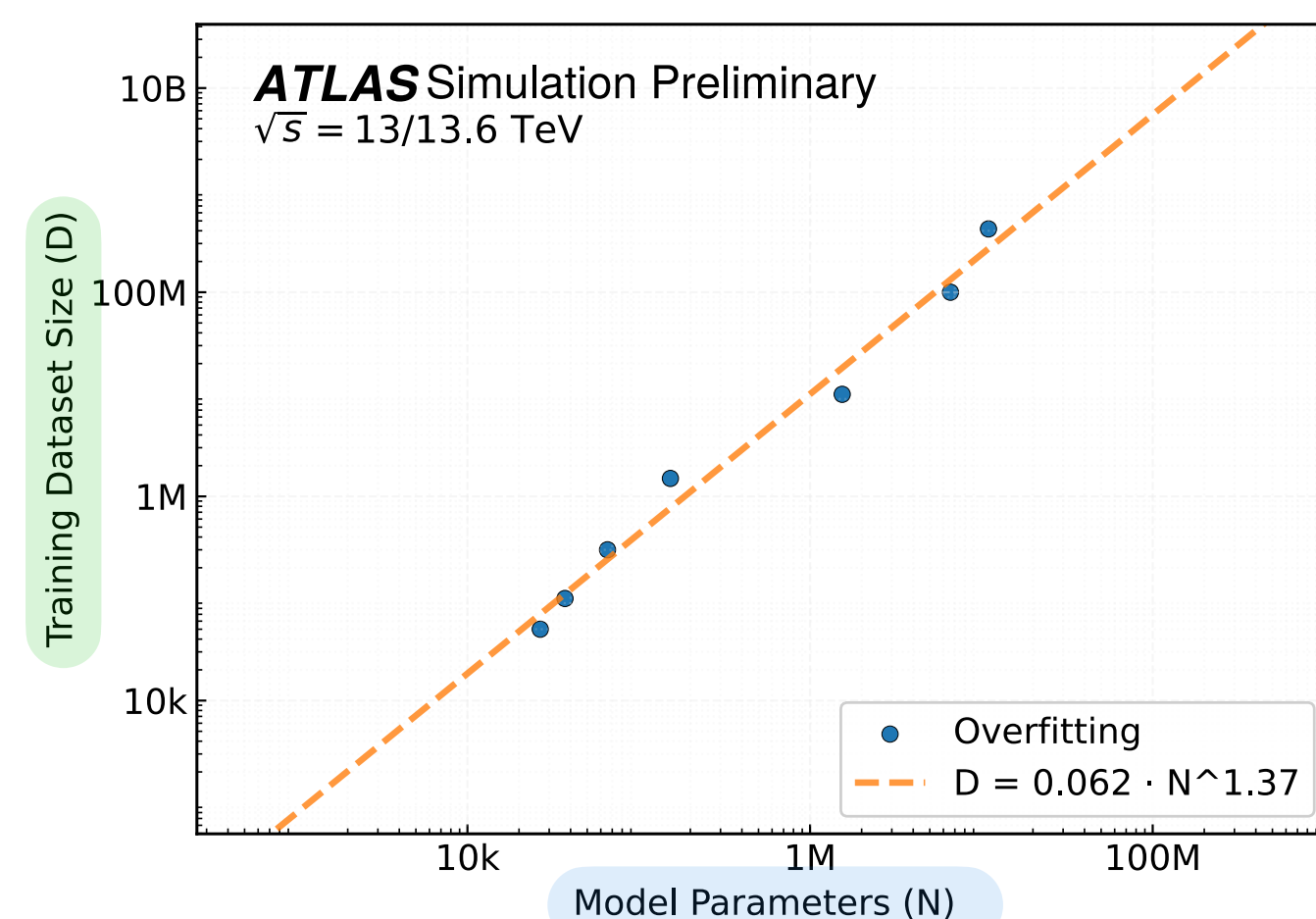
UParTv2
(6m)

GloParT
(16m)



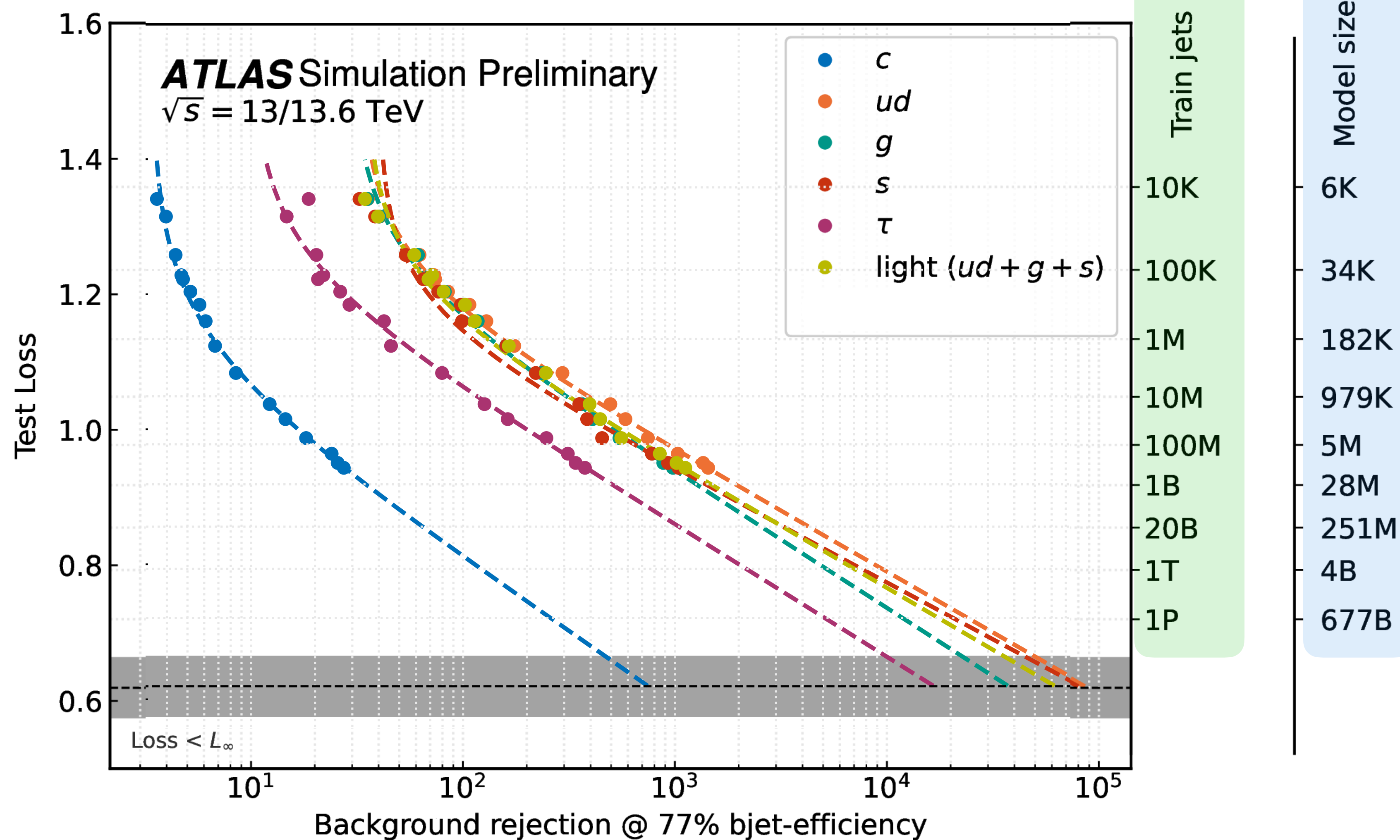
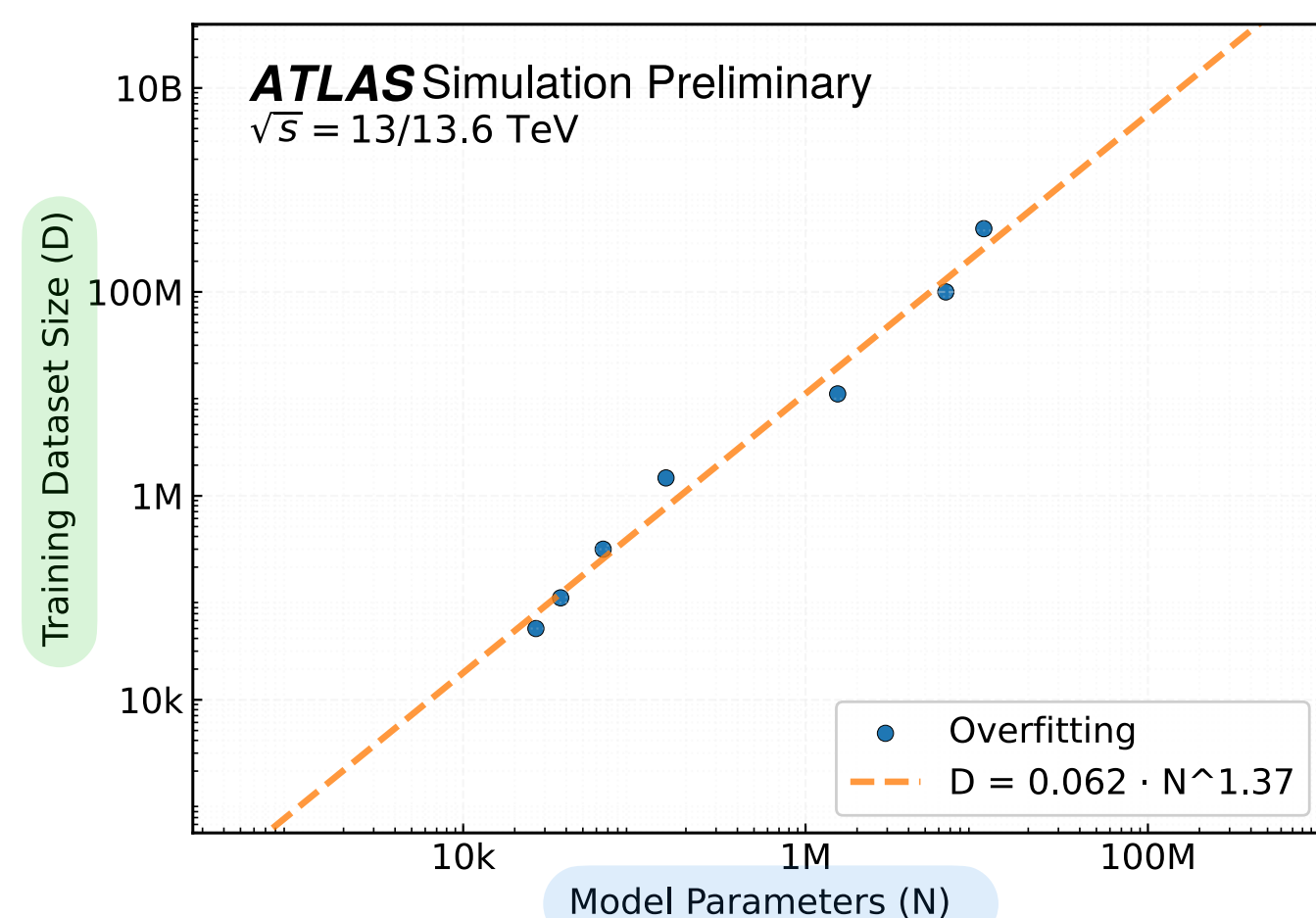
Scale translates to physics

1) Find smallest model that overfits the data



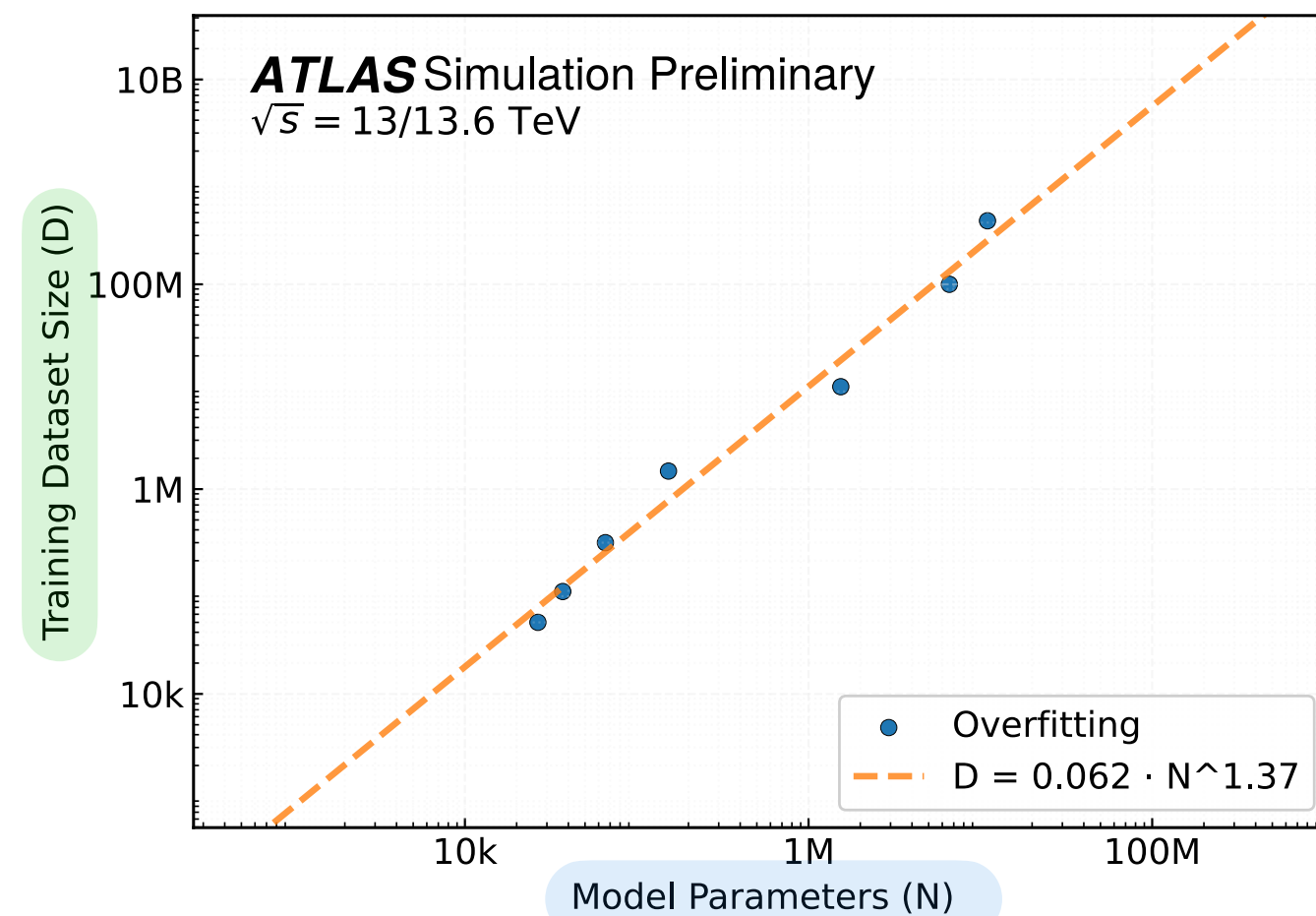
Scale translates to physics

1) Find smallest model that overfits the data



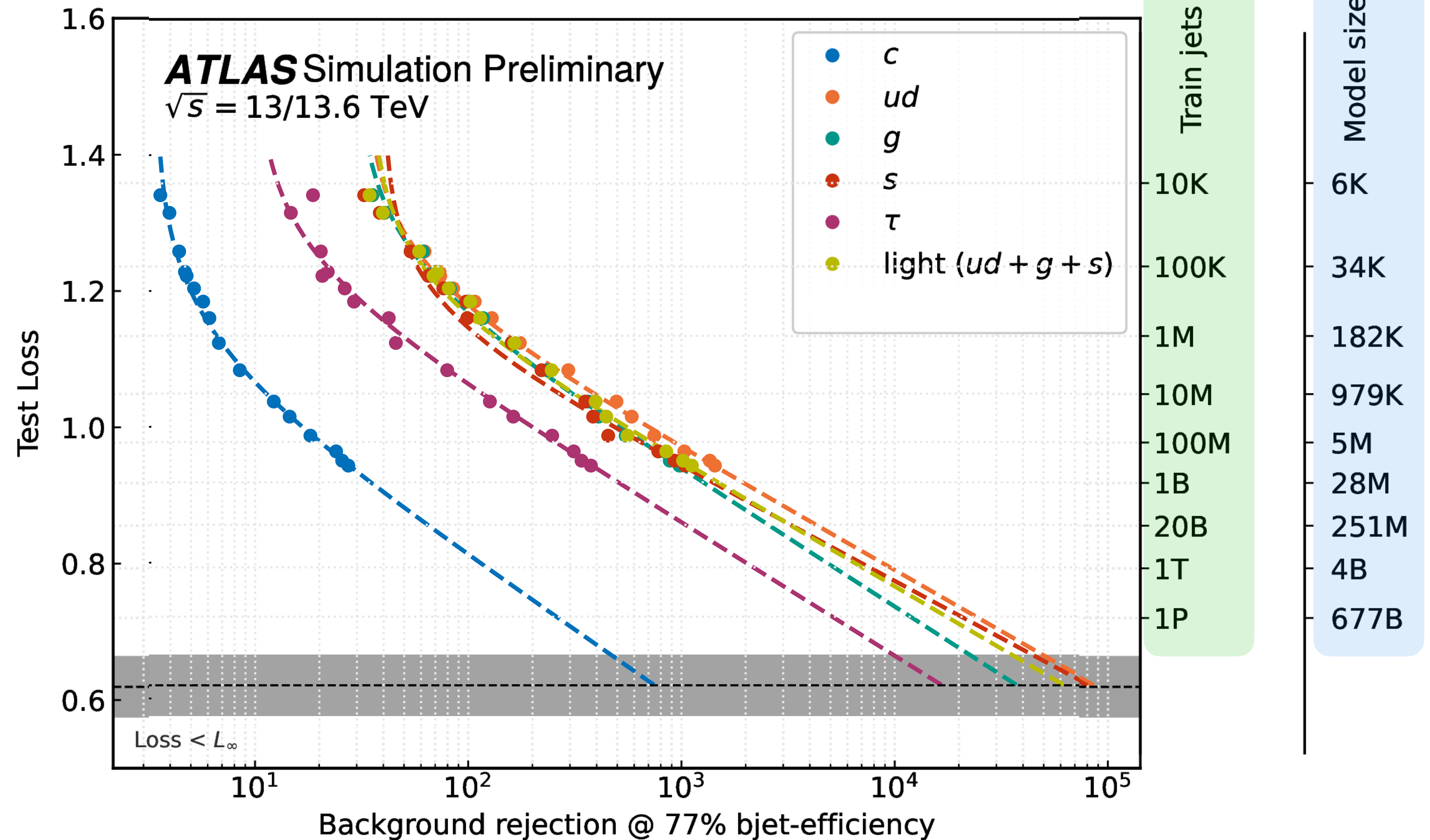
Scale translates to physics

1) Find smallest model that overfits the data



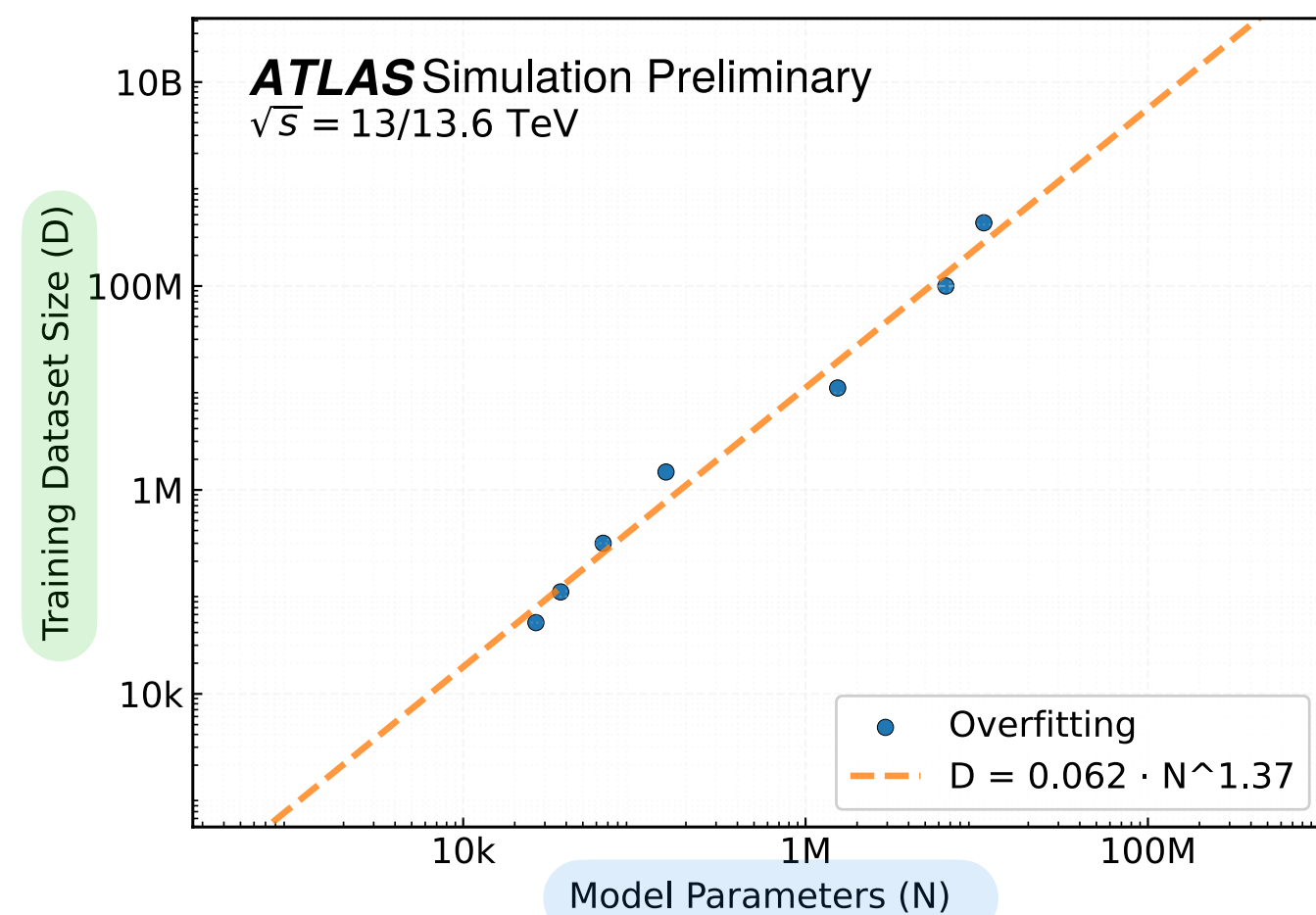
2) Dashed lines fit to:

$$R(L_i) = Ae^{BL_i} + C$$



Scale translates to physics

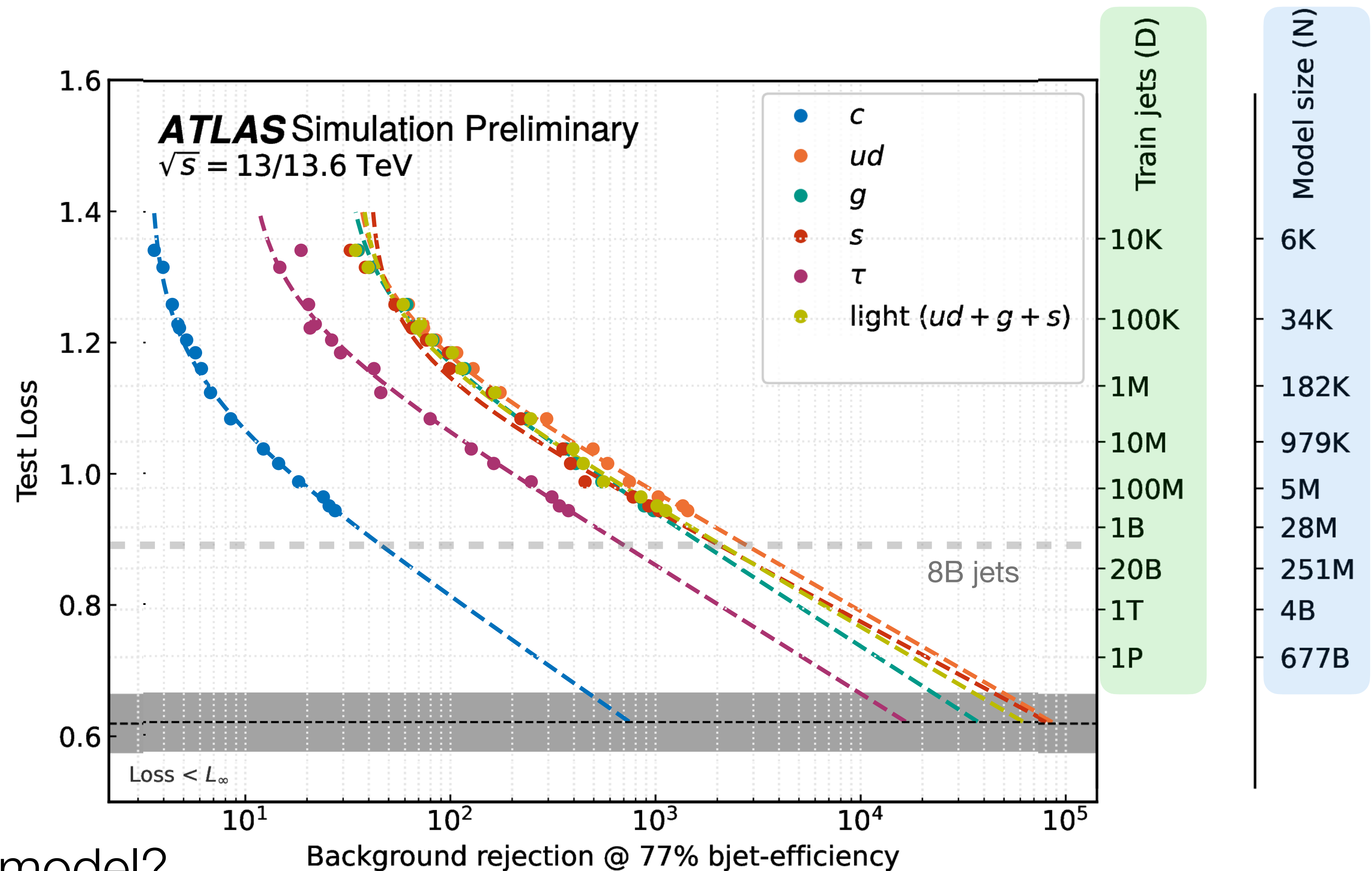
1) Find smallest model that overfits the data



2) Dashed lines fit to:

$$R(L_i) = Ae^{BL_i} + C$$

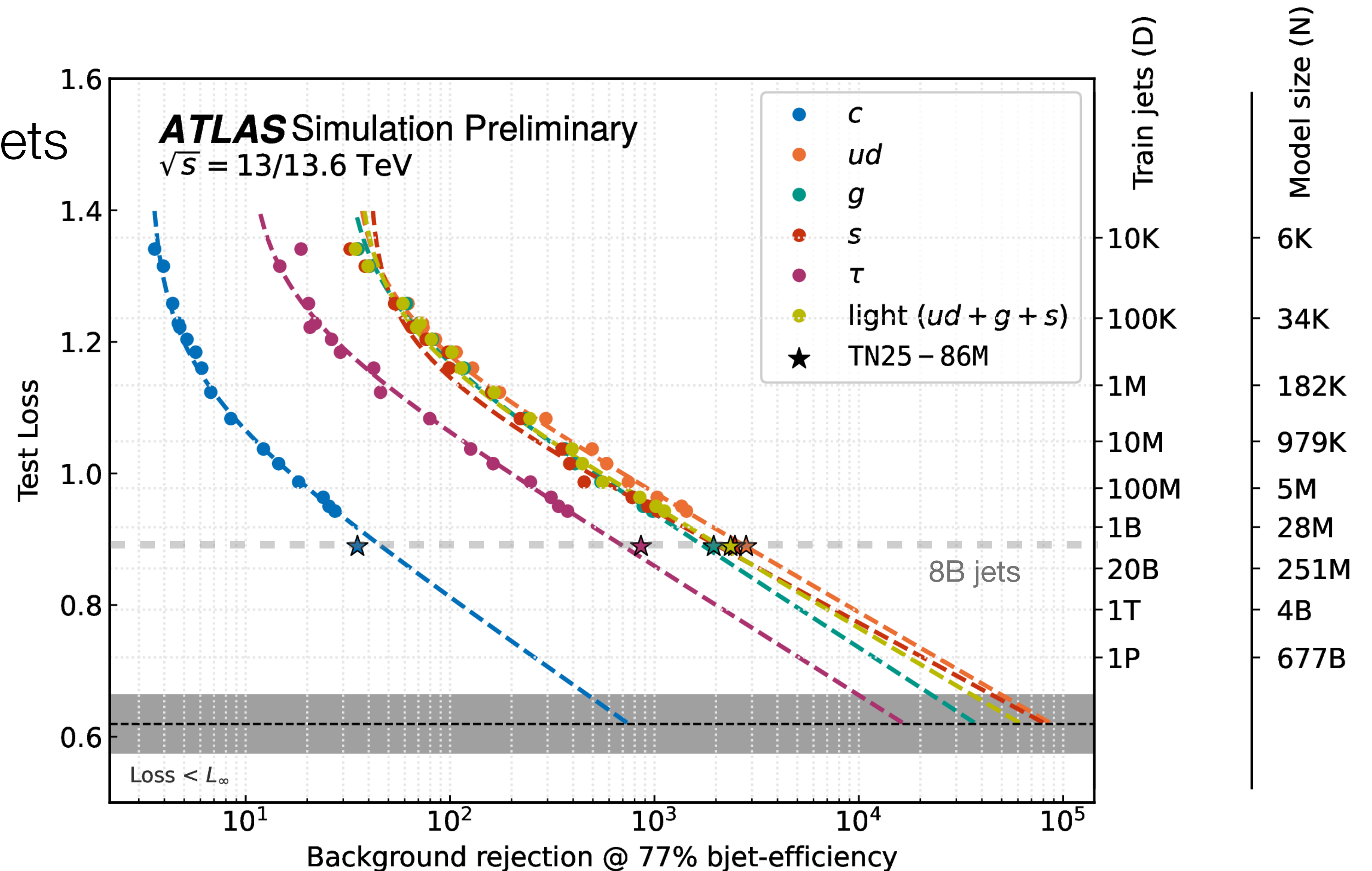
Q: Can we predict the performance of bigger model?



Scale translates to physics

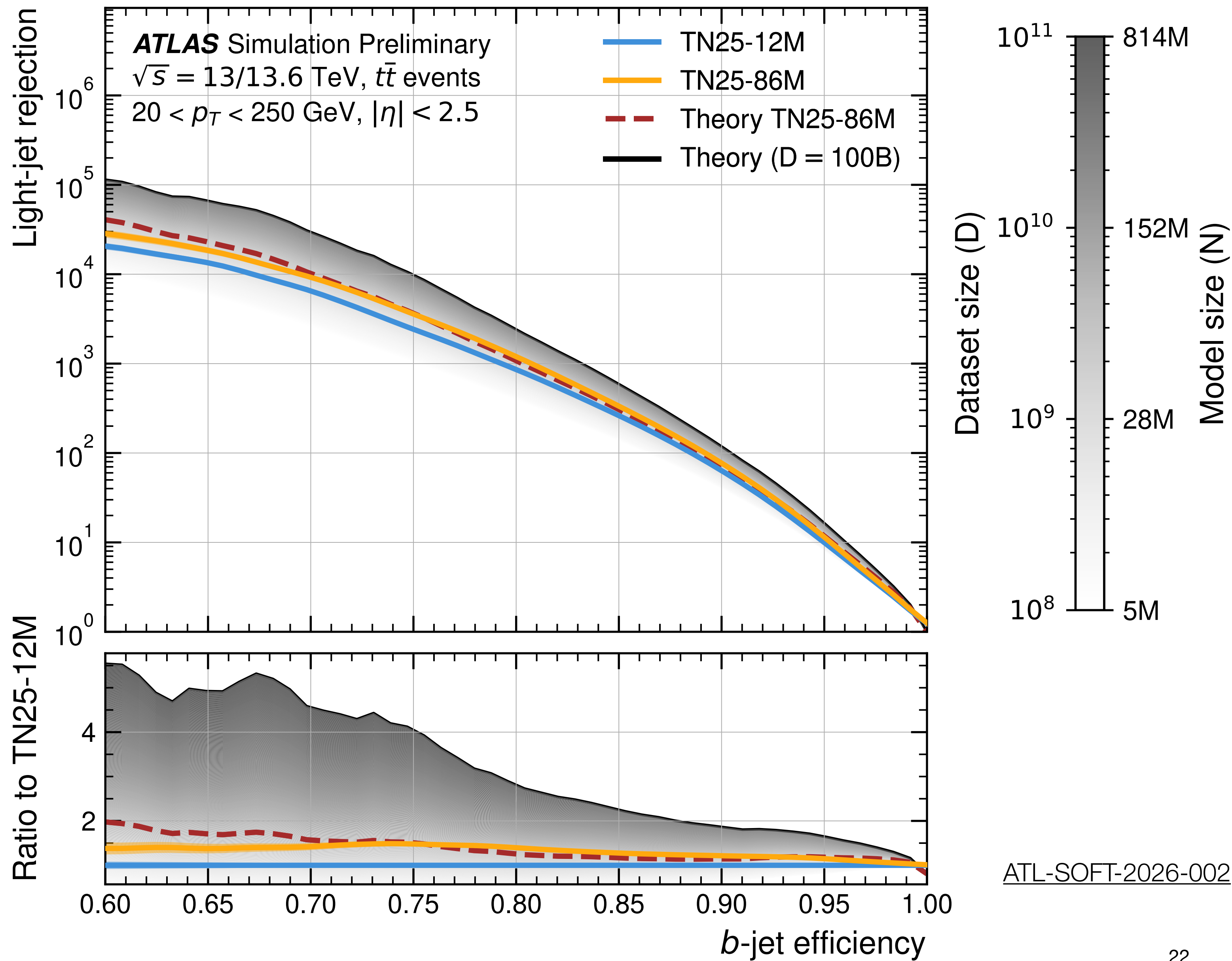
★ TN25-86M:

- New training on 8B jets
20x ↑ than GN3* dataset
- 86M param model
7x ↑ than GN3*



What could we do?

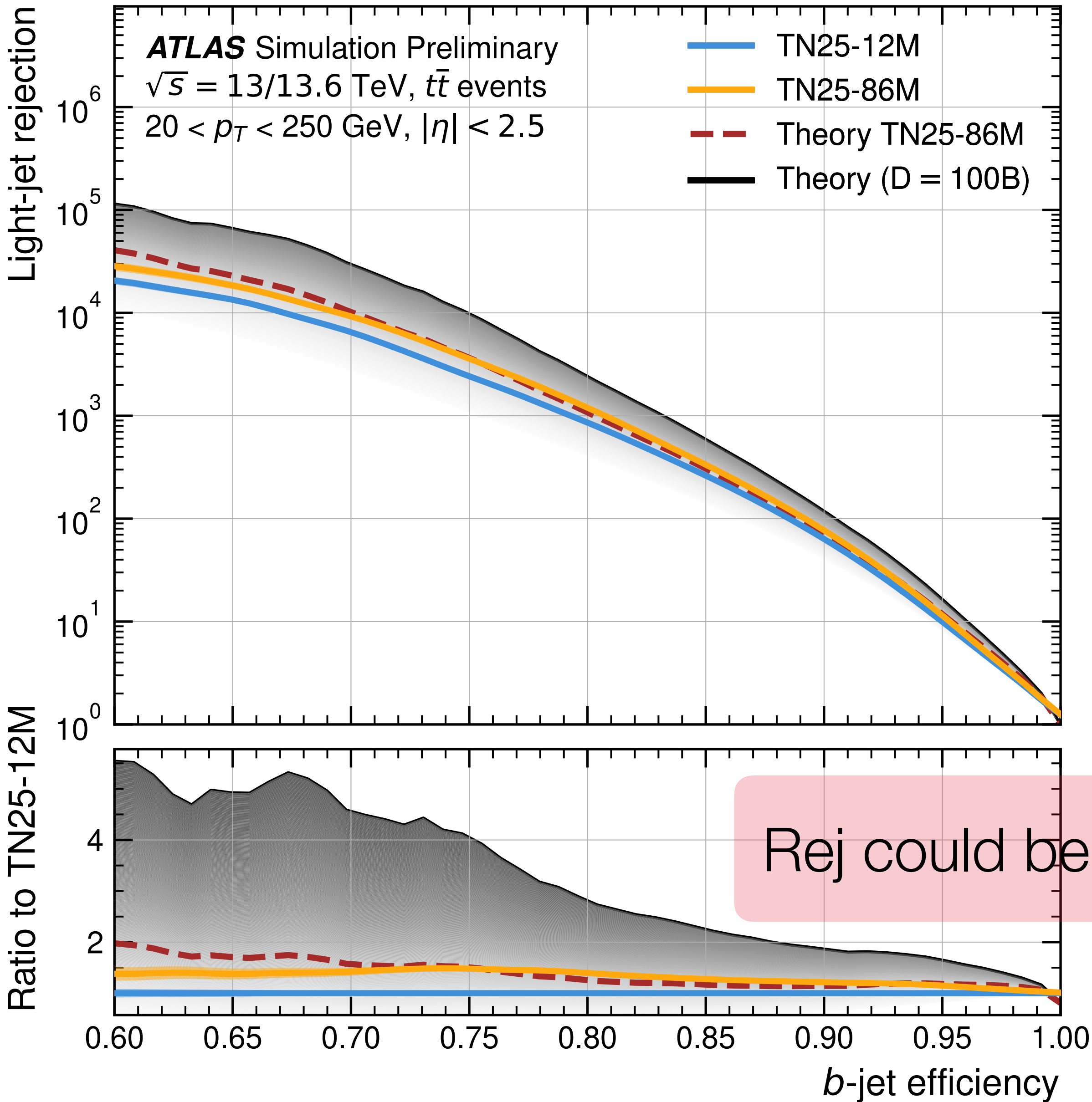
More data helps



What could we do?

More data helps

Prediction for
100 billion training jets

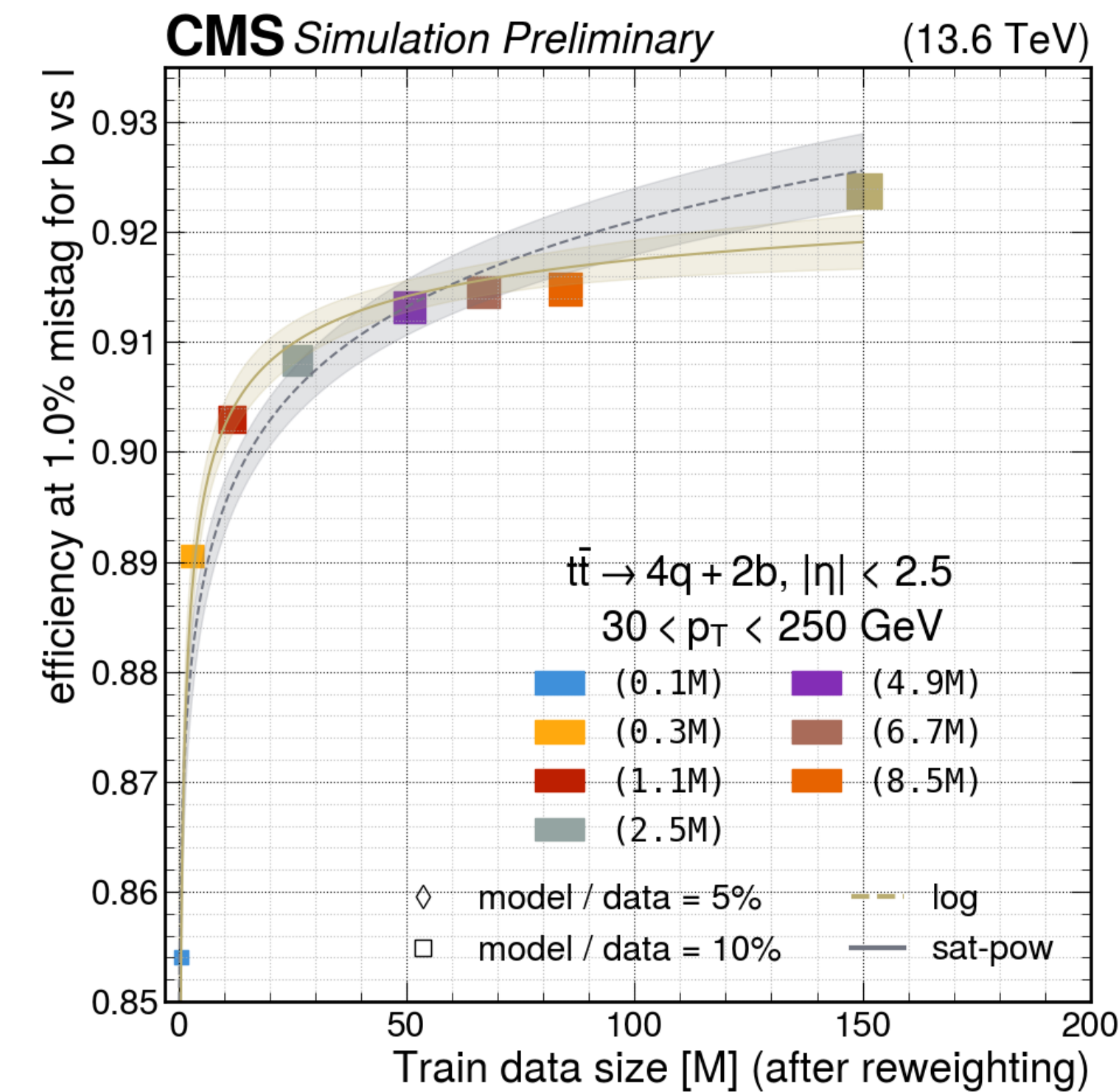


ATL-SOFT-2026-002

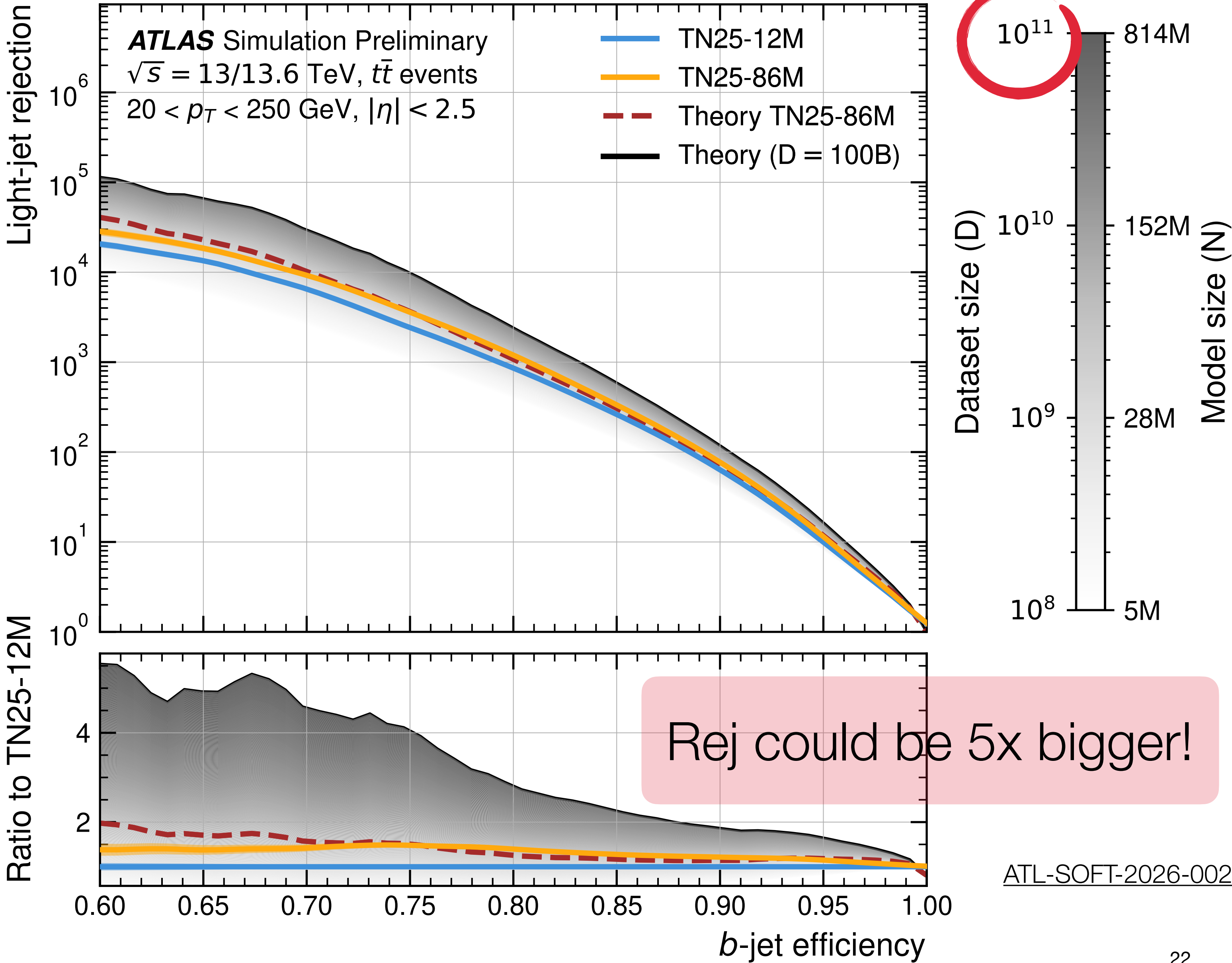
What could we do?

More data helps : CMS agrees!

CMS-DP-2025-081



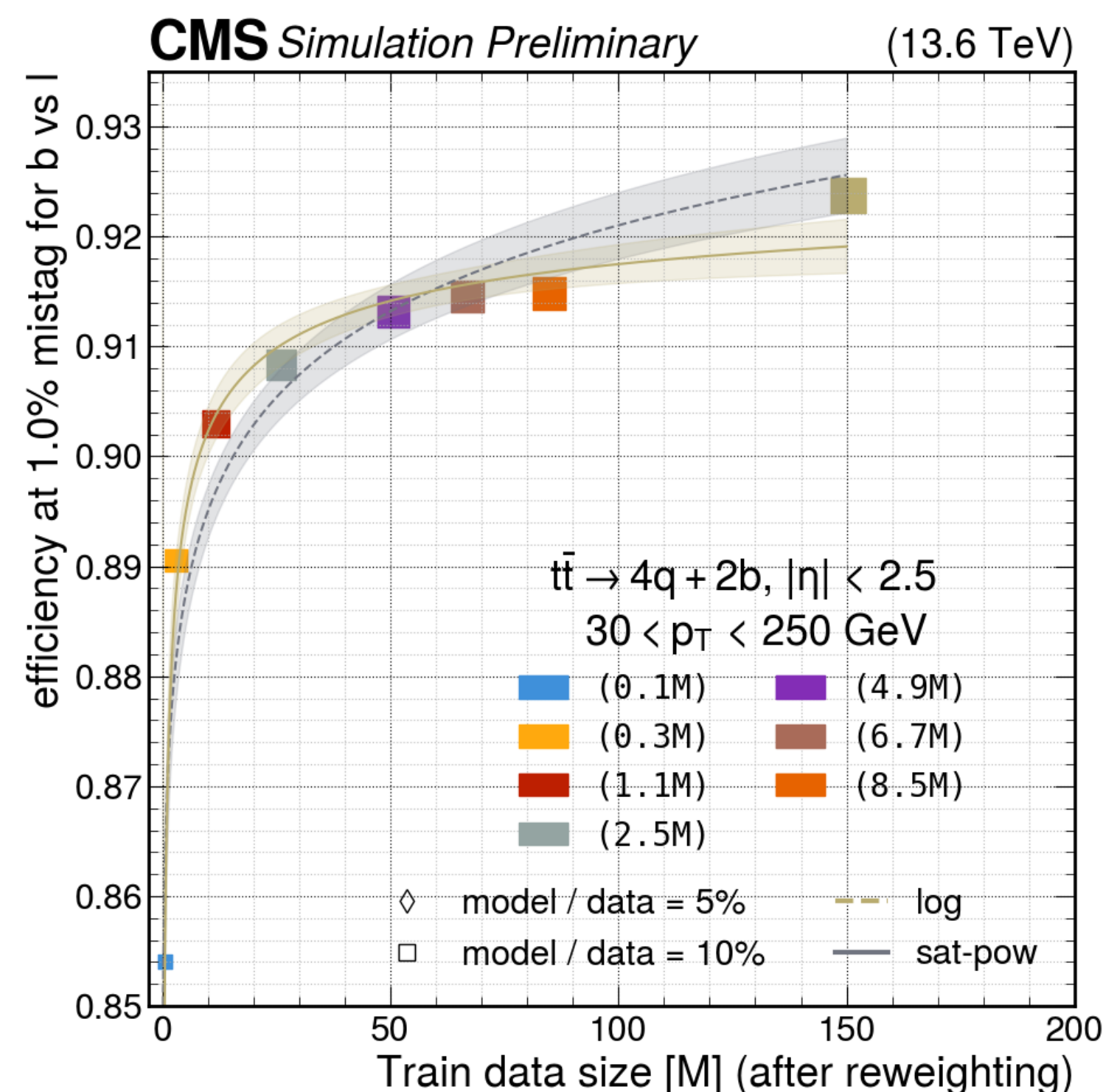
Prediction for
100 billion training jets



What could we do?

More data helps : CMS agrees!

CMS-DP-2025-081



Not only for jet tagging!

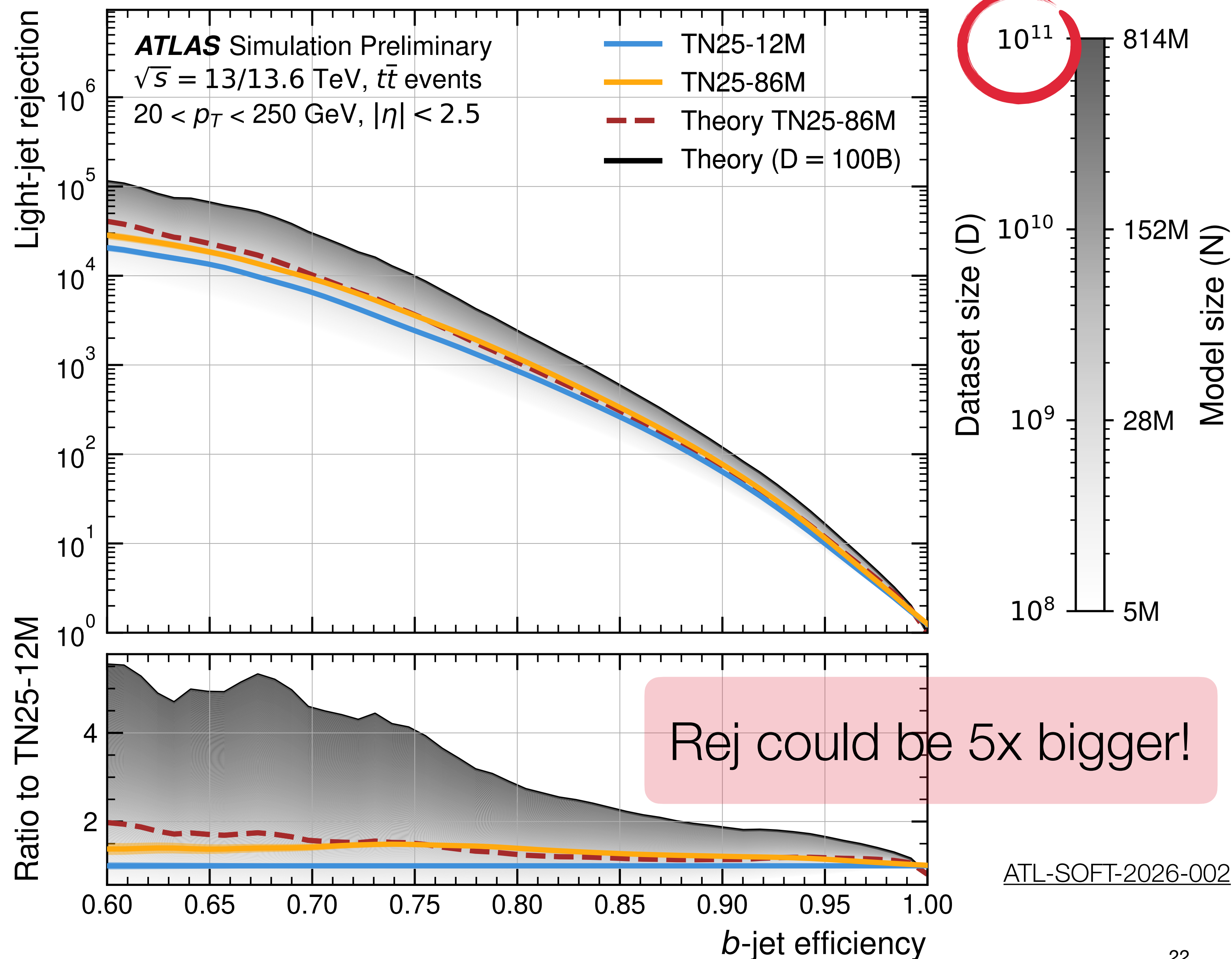
Regressing Amplitudes, [2601.13308](#)

Generative models, [2605.28940](#) + Anna's [talk](#) on Wed

Symmetric networks, [2608.02735](#) + Jonas's [talk](#) on Fri

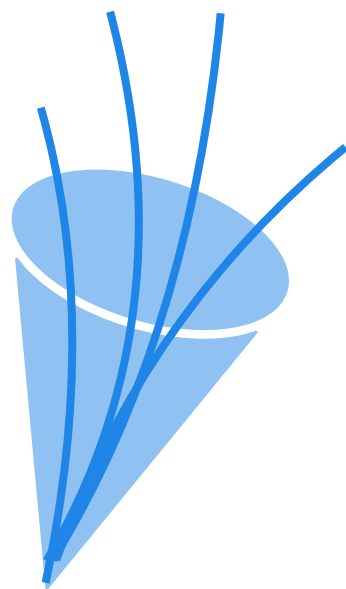
⋮

Prediction for
100 billion training jets

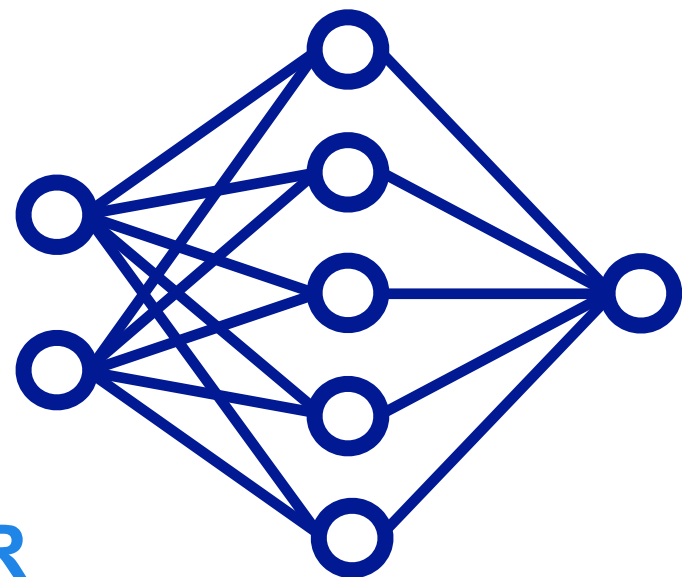




How far can we transfer?



8 billion small R
jets pretraining



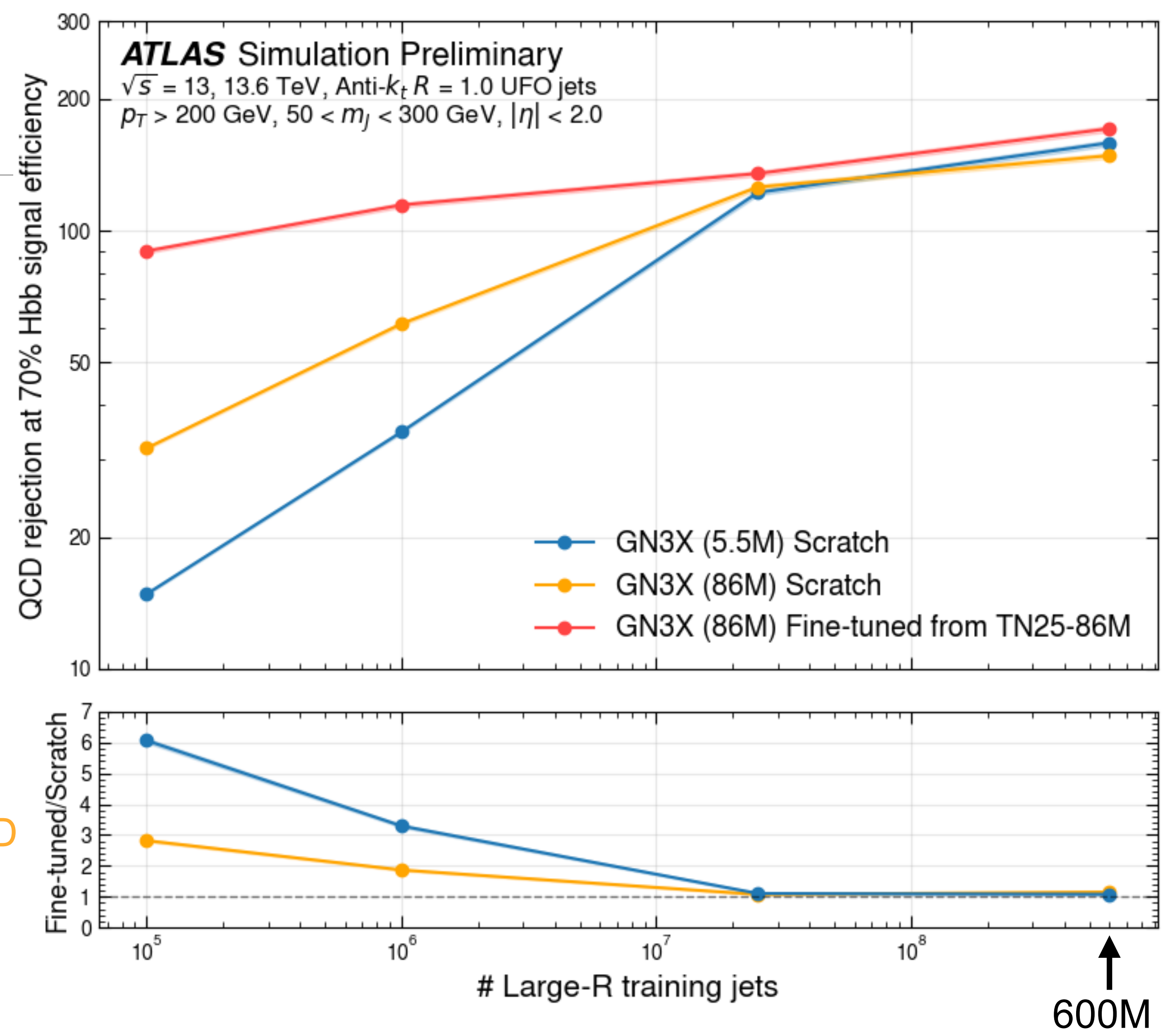
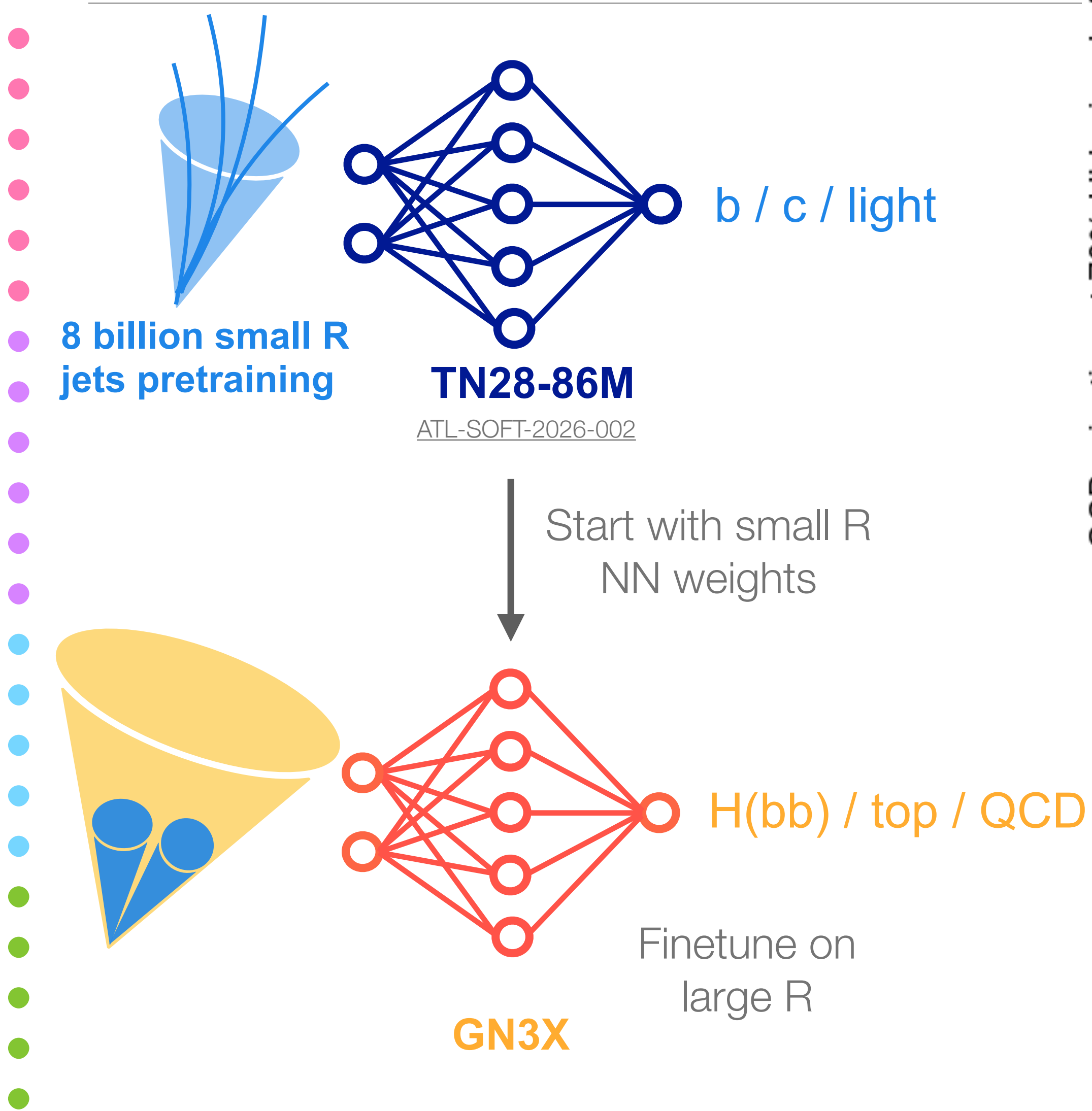
TN28-86M

ATL-SOFT-2026-002

b / c / light

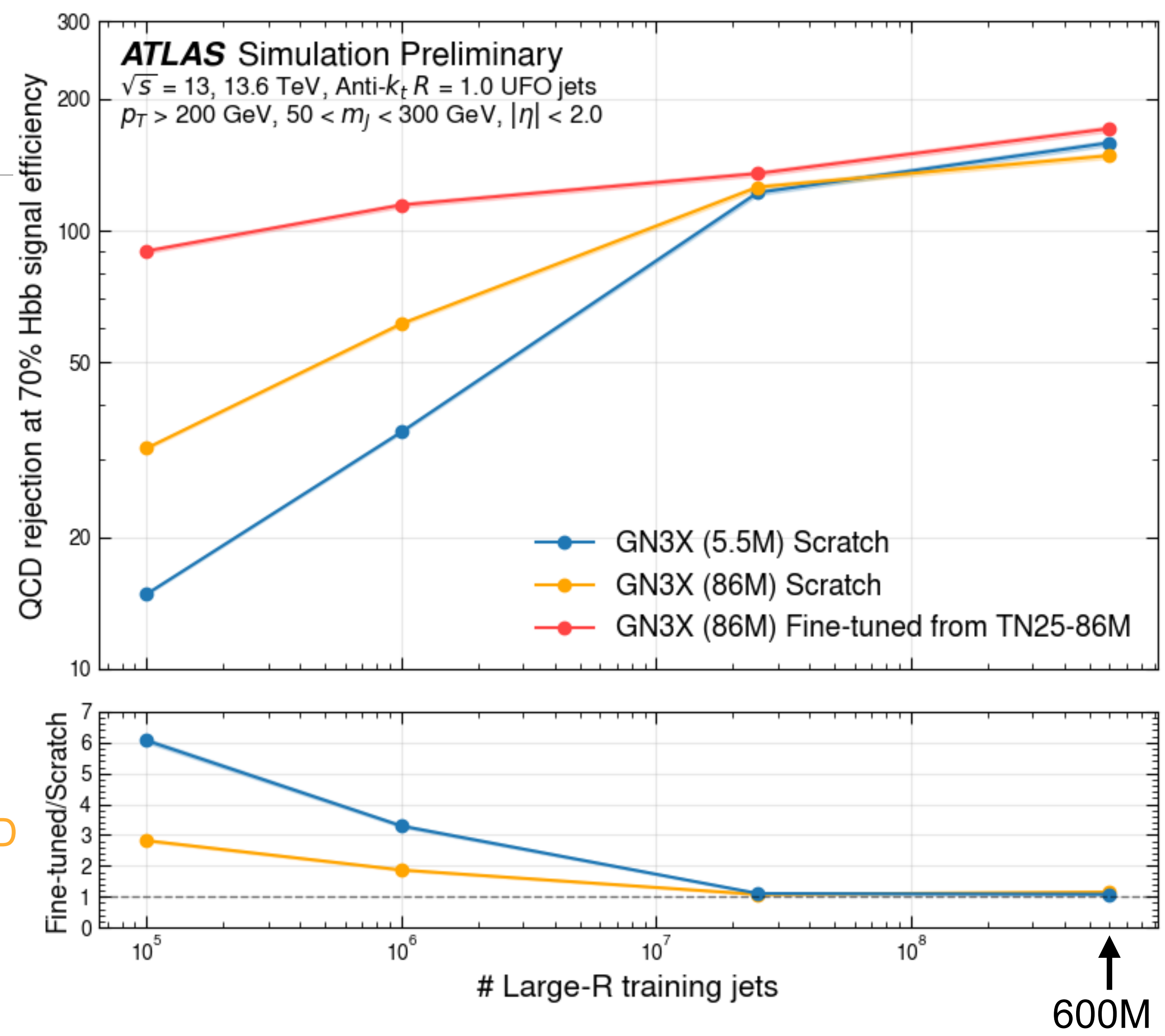
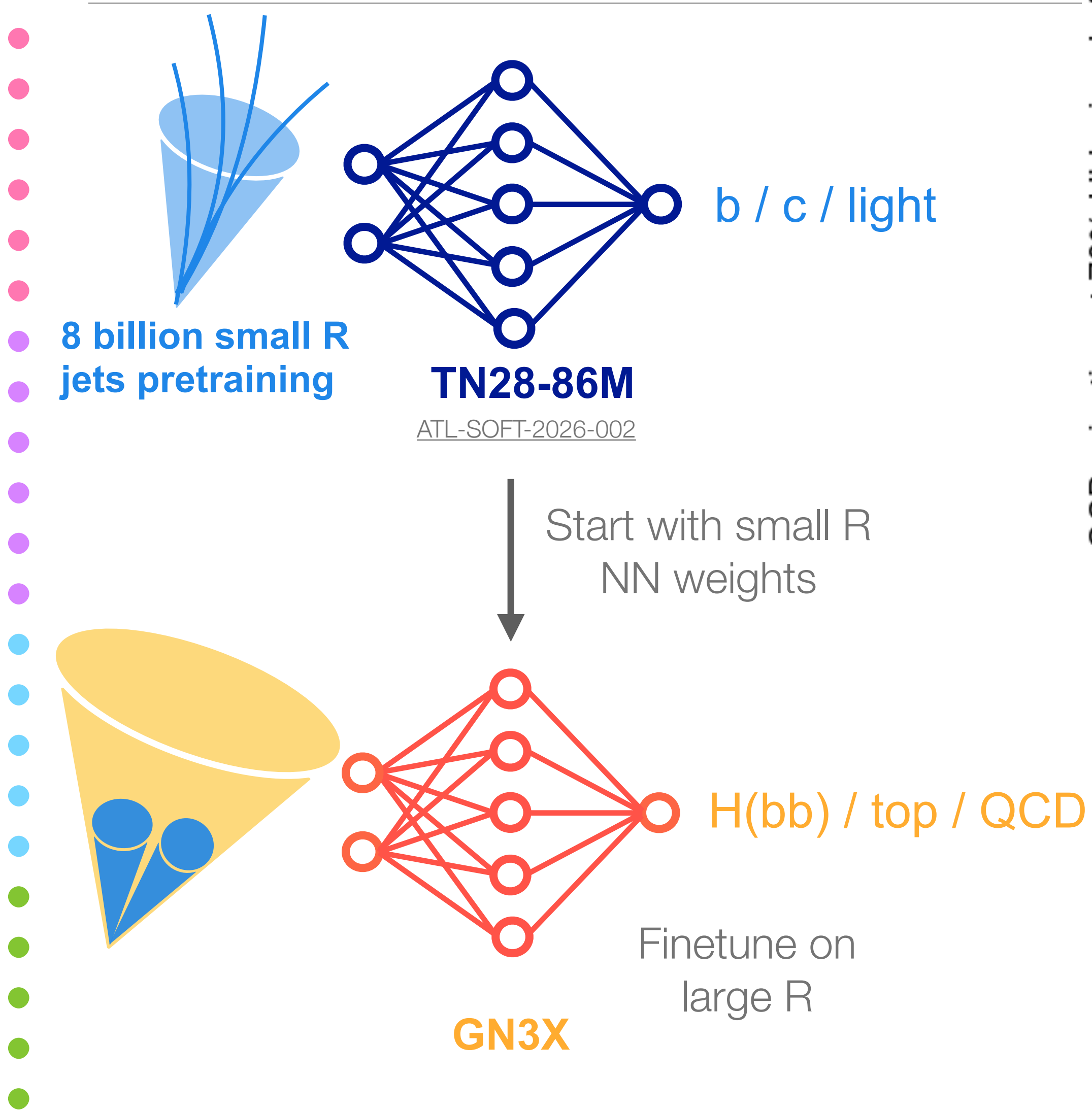
NEW!

How far can we transfer?



NEW!

How far can we transfer?



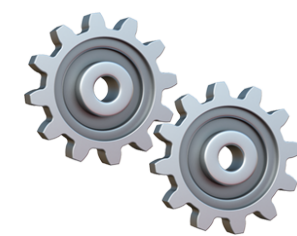
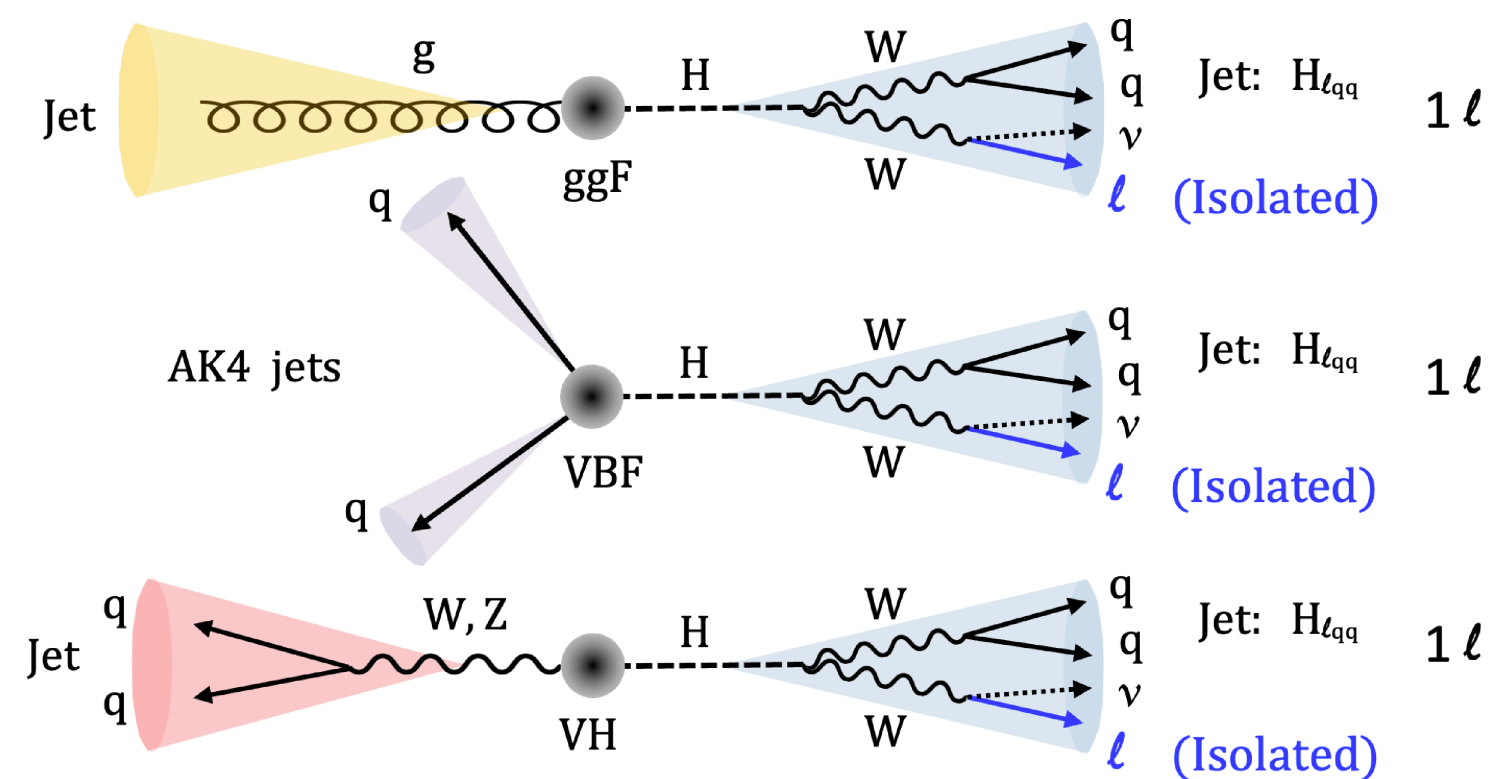
How you finetune matters, ref

Boosted H(WW) tagging

Latest GloParT model has 374 classes (general for pretraining)

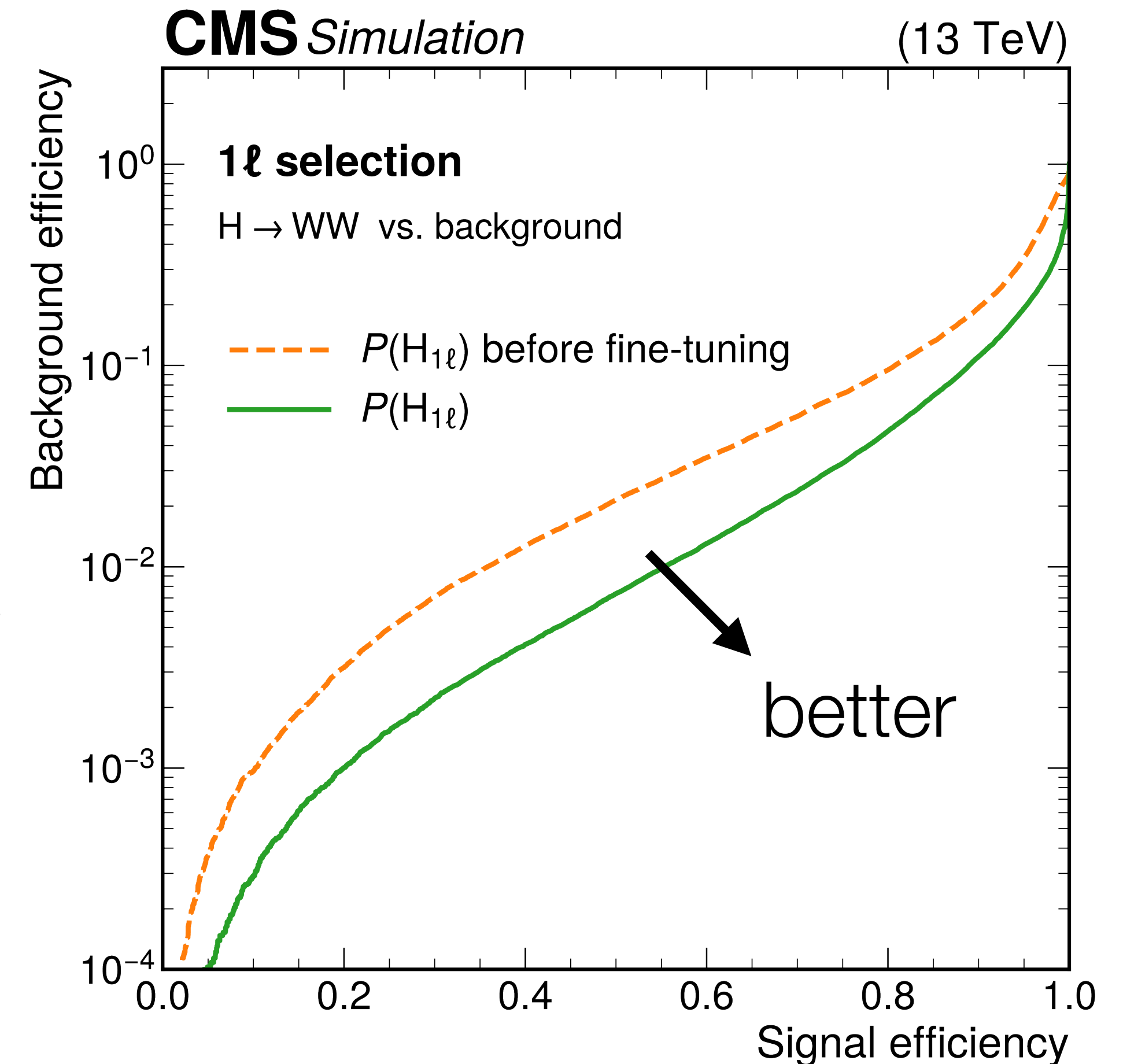
Wonky “isolated” lepton in a large R jet class not included

1 lepton

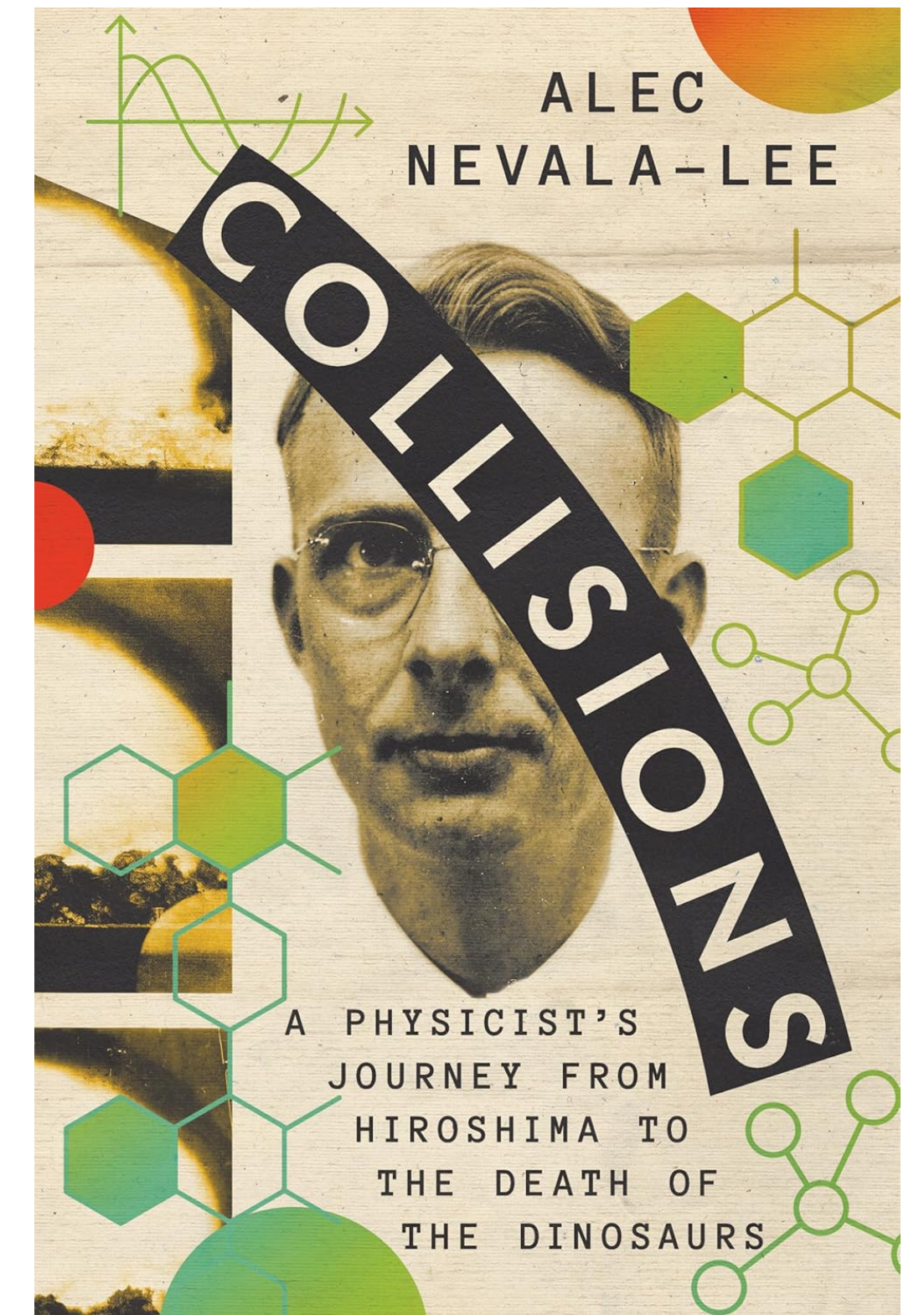
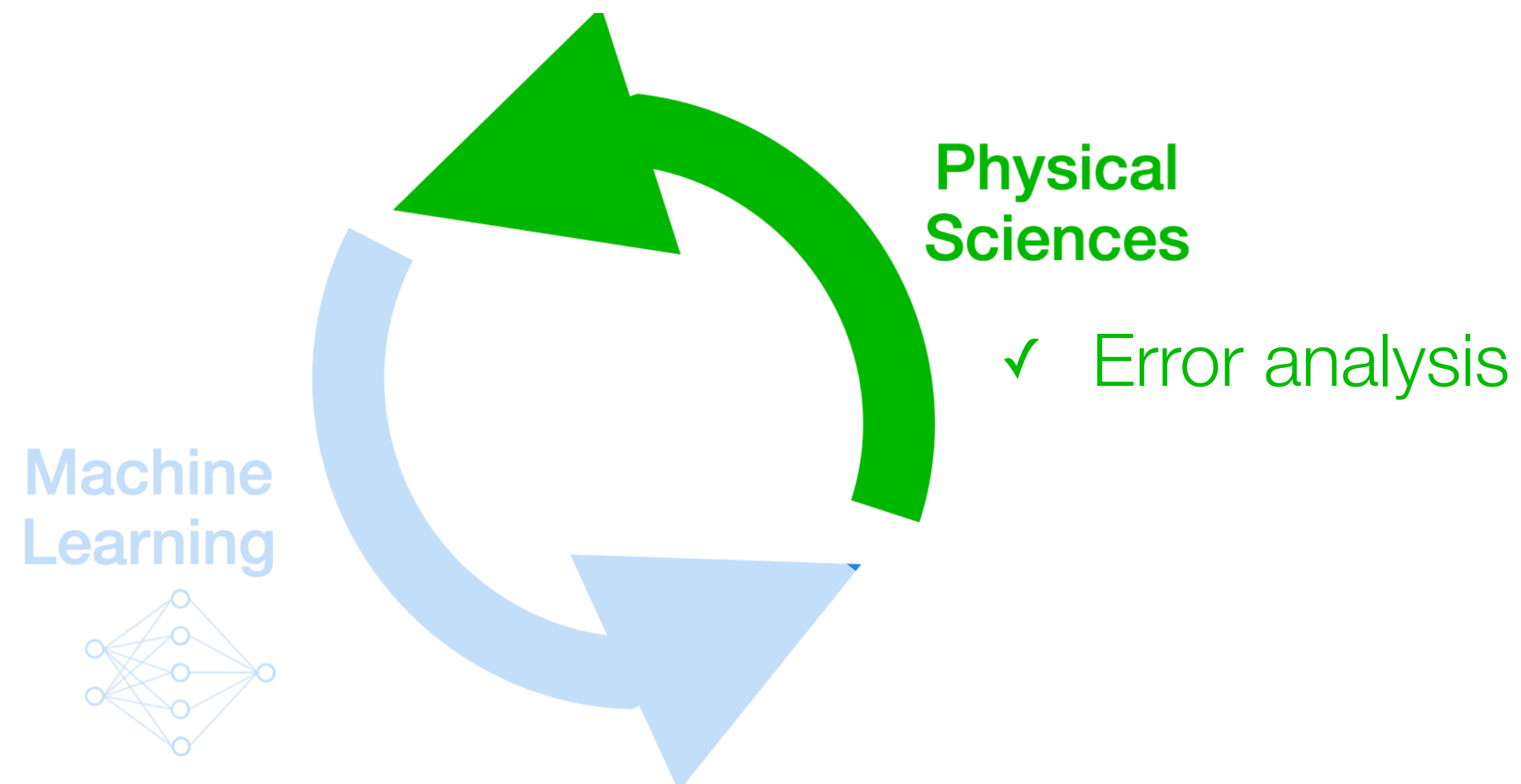


fine-tune

Impact (1ℓ): Obs: 1.36σ
Exp: 0.36σ



The real scientific method isn't hypothesis testing, but a determination to catch potential errors.



Jet topics: q/g calibration

Disentangling Jet Categories at Colliders

Machine Learning for Jet Physics Workshop

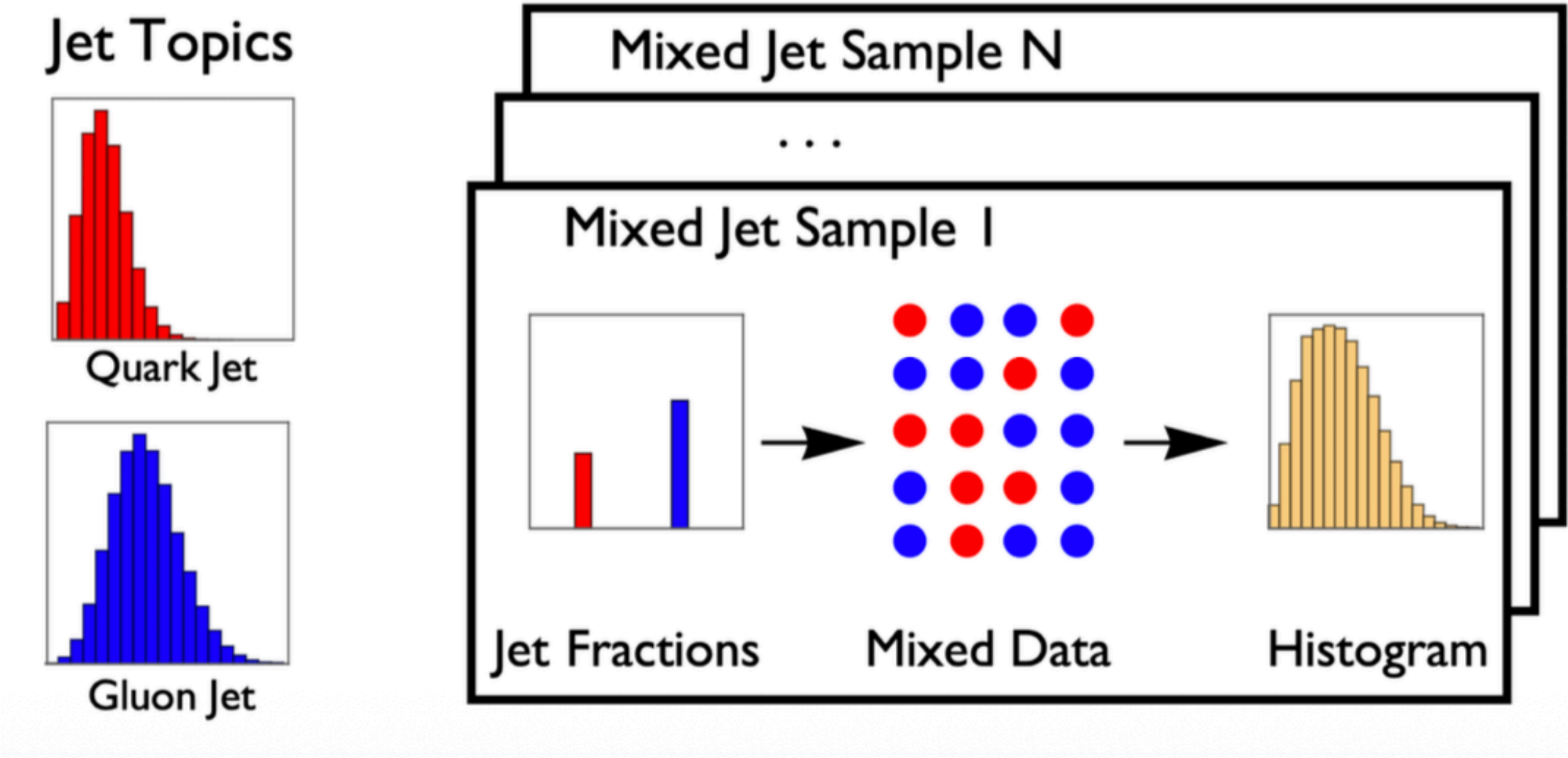
Eric M. Metodiev
Center for Theoretical Physics
Massachusetts Institute of Technology

Joint work with Patrick Komiske and Jesse Thaler

November 16, 2018

[talk](#)

Metodiev & Thaler, PRL 120 (2018) 241602
Komiske, Kryhinand, Thaler, PRD **106** (2022) 094021
LeBlanc, Nachman, Sauer, JHEP **02** (2023) 150



Key assumption: η independence

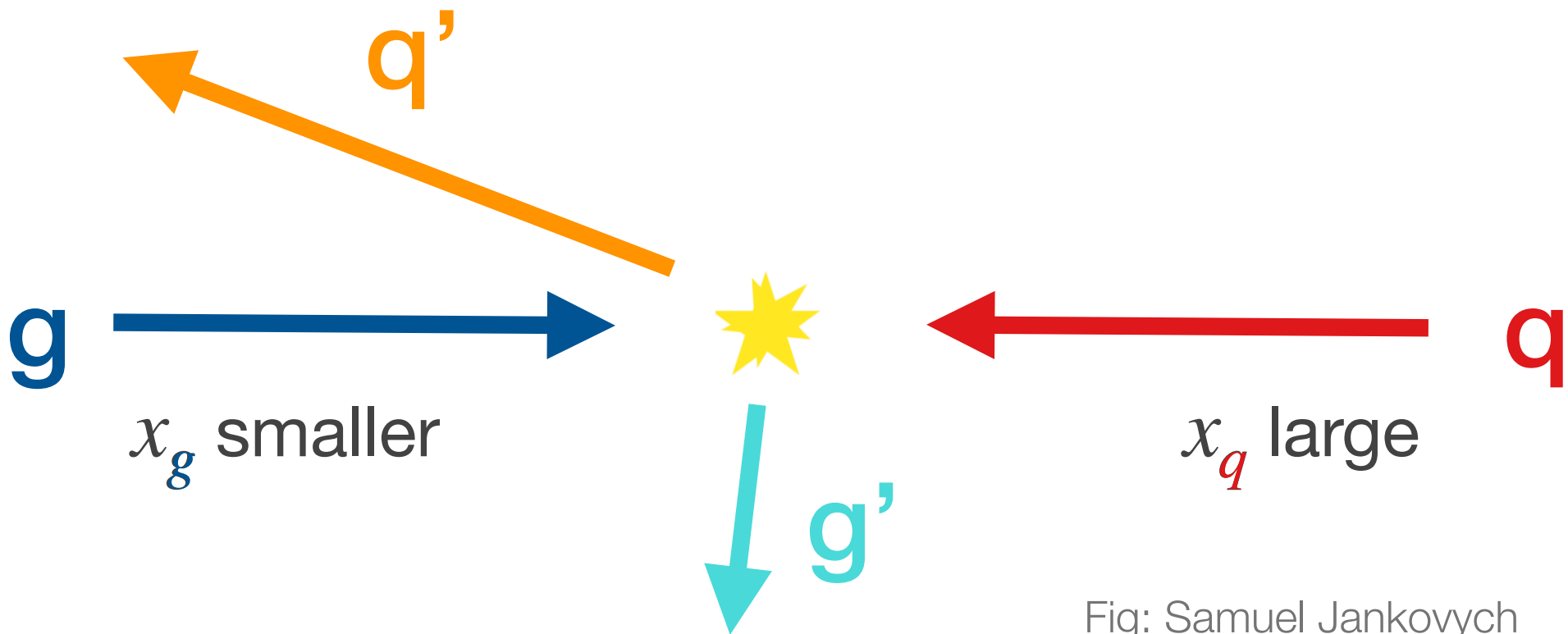


Fig: Samuel Jankovych

Jet topic: q/g calibration

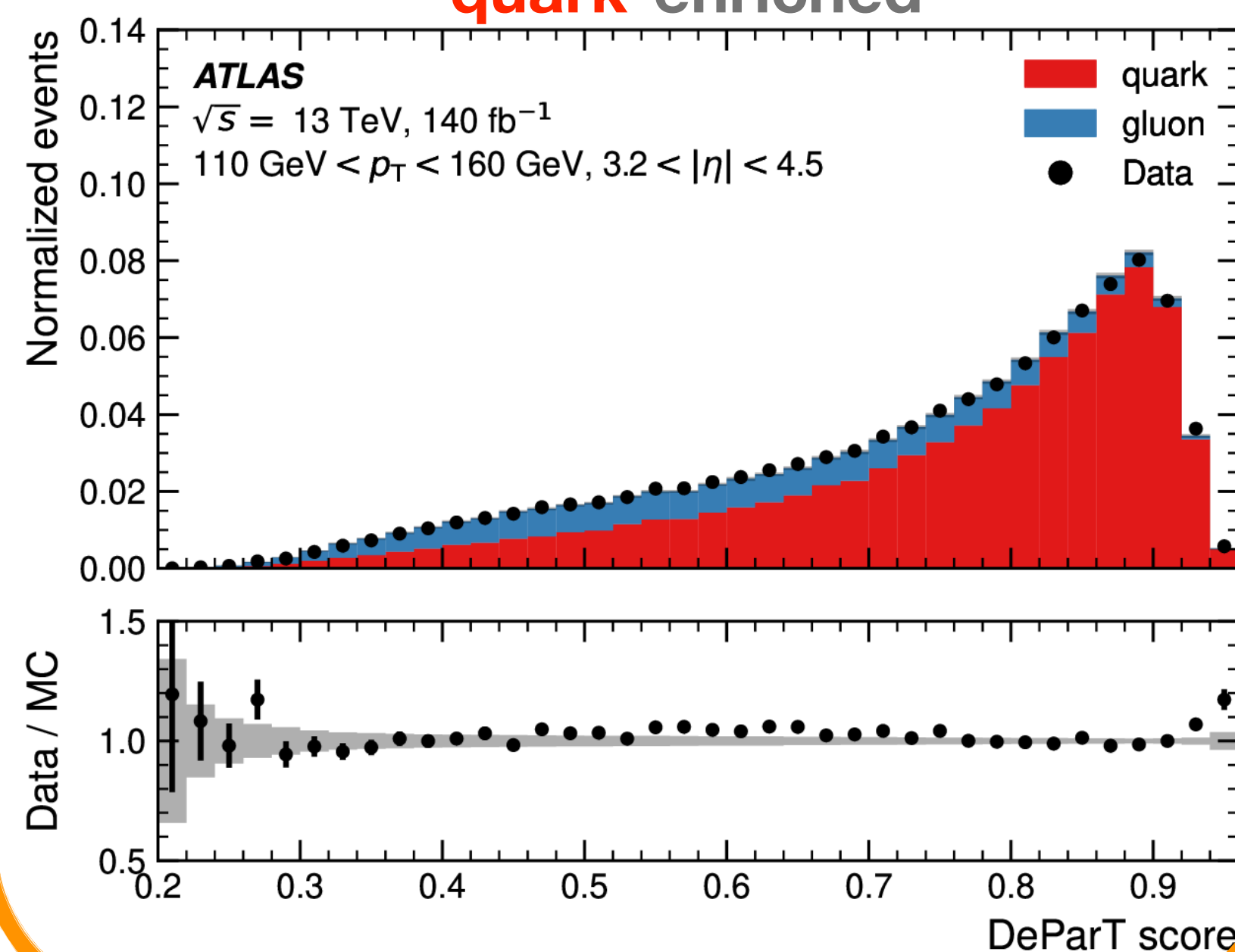
$$p_{\text{quark}}(x) \equiv \frac{p_F(x) - \kappa_{FC} p_C(x)}{1 - \kappa_{FC}}$$

$$p_{\text{gluon}}(x) \equiv \frac{p_C(x) - \kappa_{CF} p_F(x)}{1 - \kappa_{CF}}$$

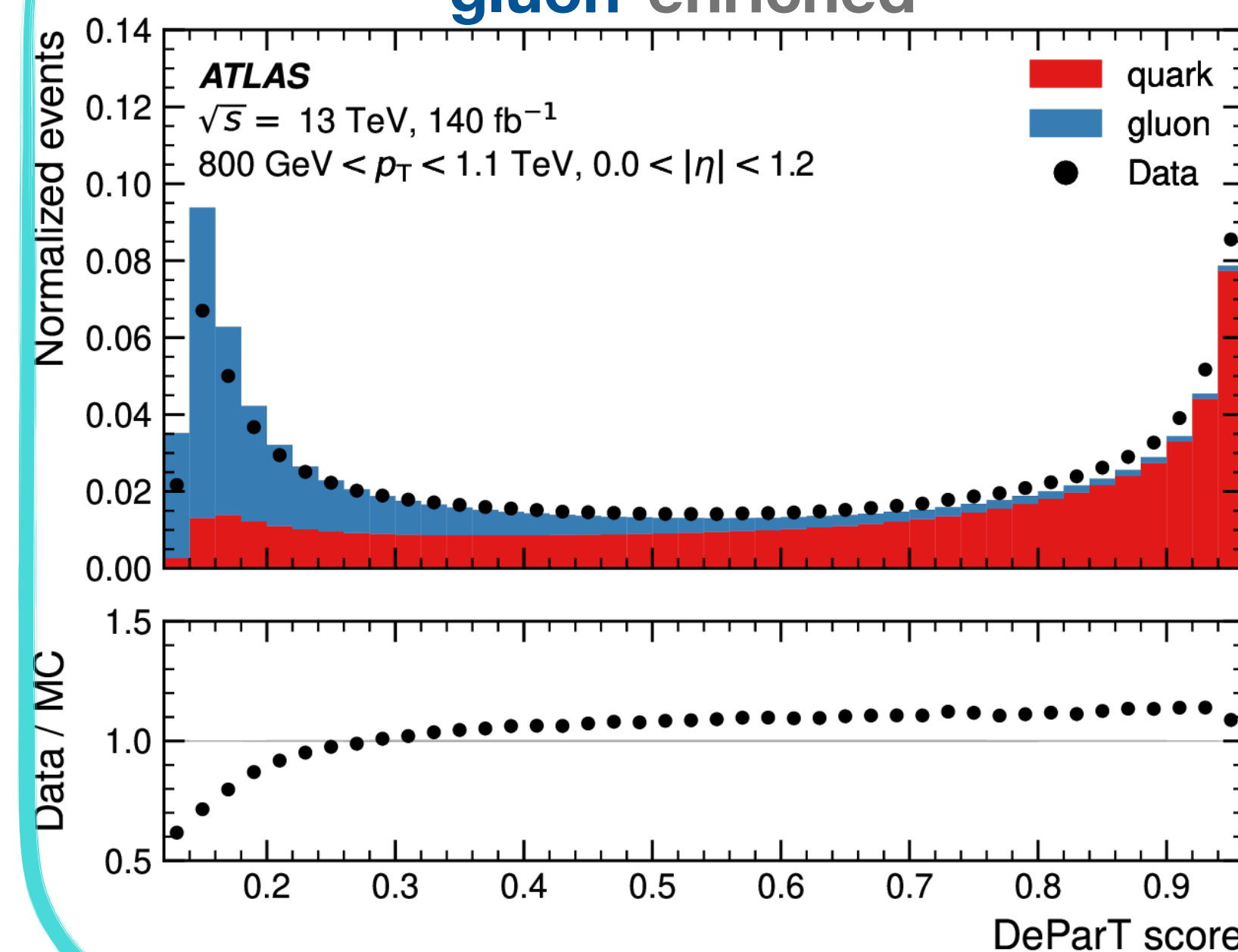
Experimental details

- η invariance not respected by detector
 - Reweight central $\rightarrow$ forward
- Need to subtract off backgrounds
 - Apply this formula twice and solve two systems of equations

Forward jets quark-enriched



Central jets gluon-enriched



Jet topic: q/g calibration

$$p_{\text{quark}}(x) \equiv \frac{p_F(x) - \kappa_{FC} p_C(x)}{1 - \kappa_{FC}}$$

$$\kappa_{FC} = \min_x \frac{p_F(x)}{p_C(x)}, \text{ and similarly for } \kappa_{CF}$$

$$p_{\text{gluon}}(x) \equiv \frac{p_C(x) - \kappa_{CF} p_F(x)}{1 - \kappa_{CF}}$$

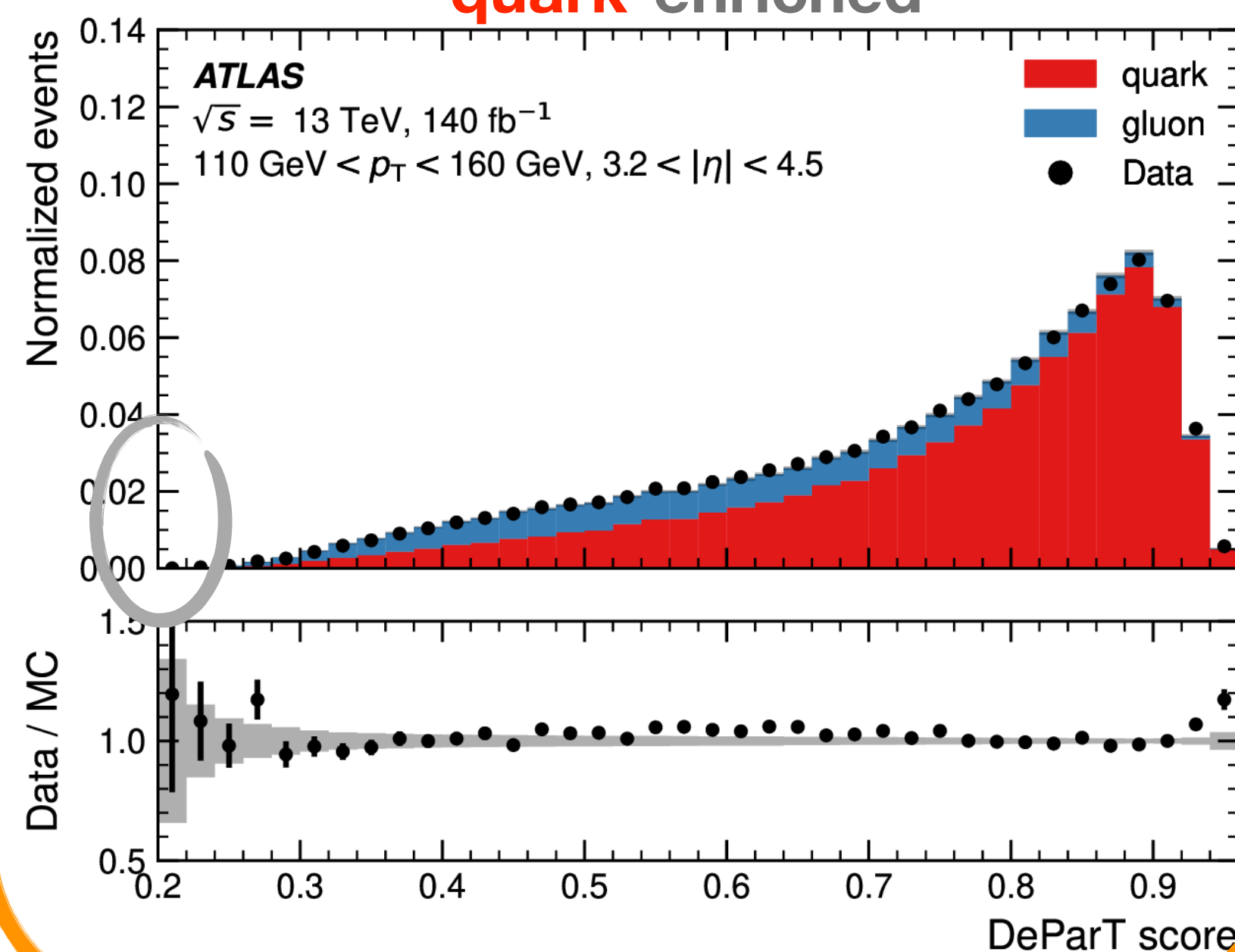


Anchor bins: extracted from (ideally) pure bin sin x

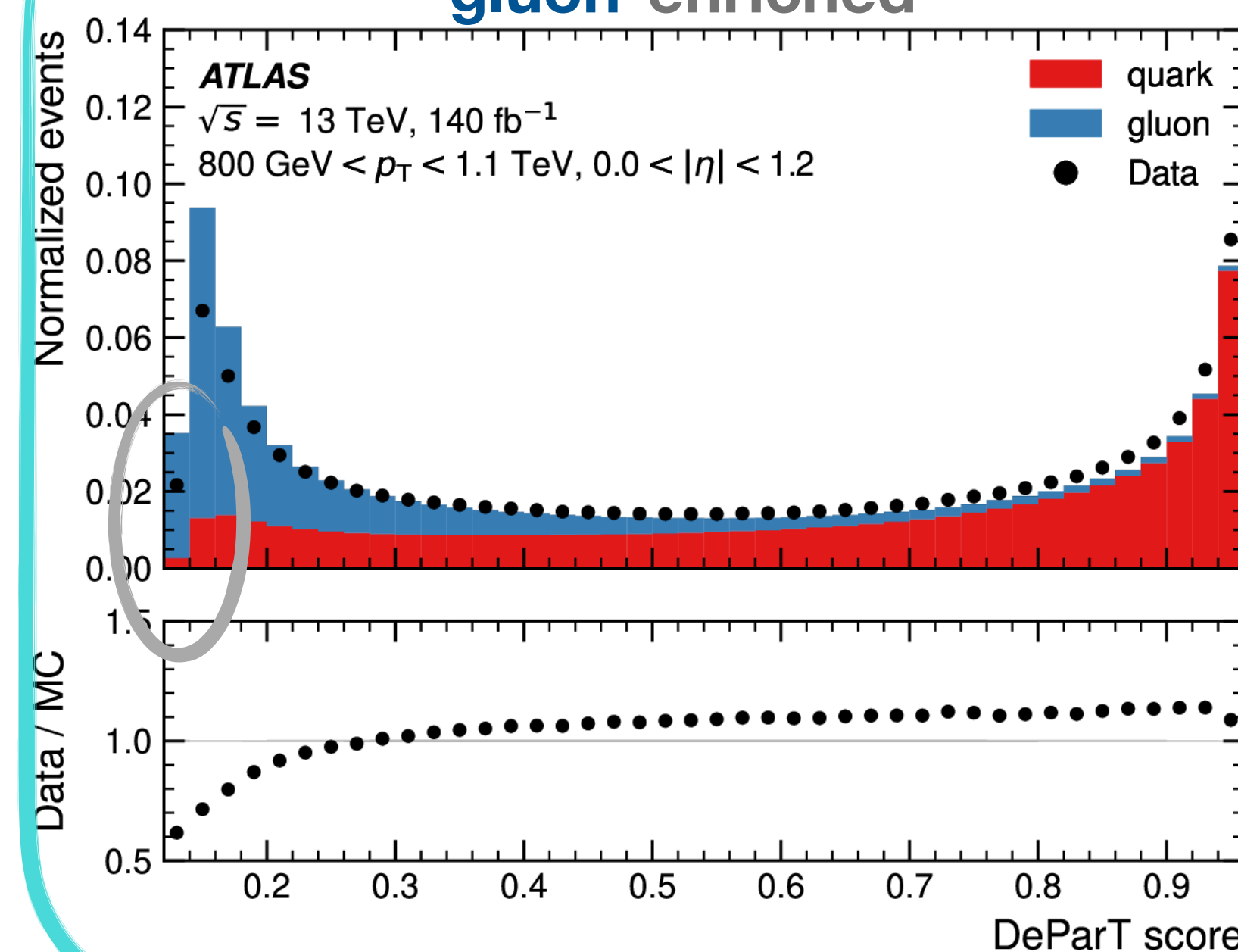
Experimental details

- η invariance not respected by detector
 - Reweight central $\rightarrow$ forward
- Need to subtract off backgrounds
 - Apply this formula twice and solve two systems of equations

Forward jets quark-enriched



Central jets gluon-enriched

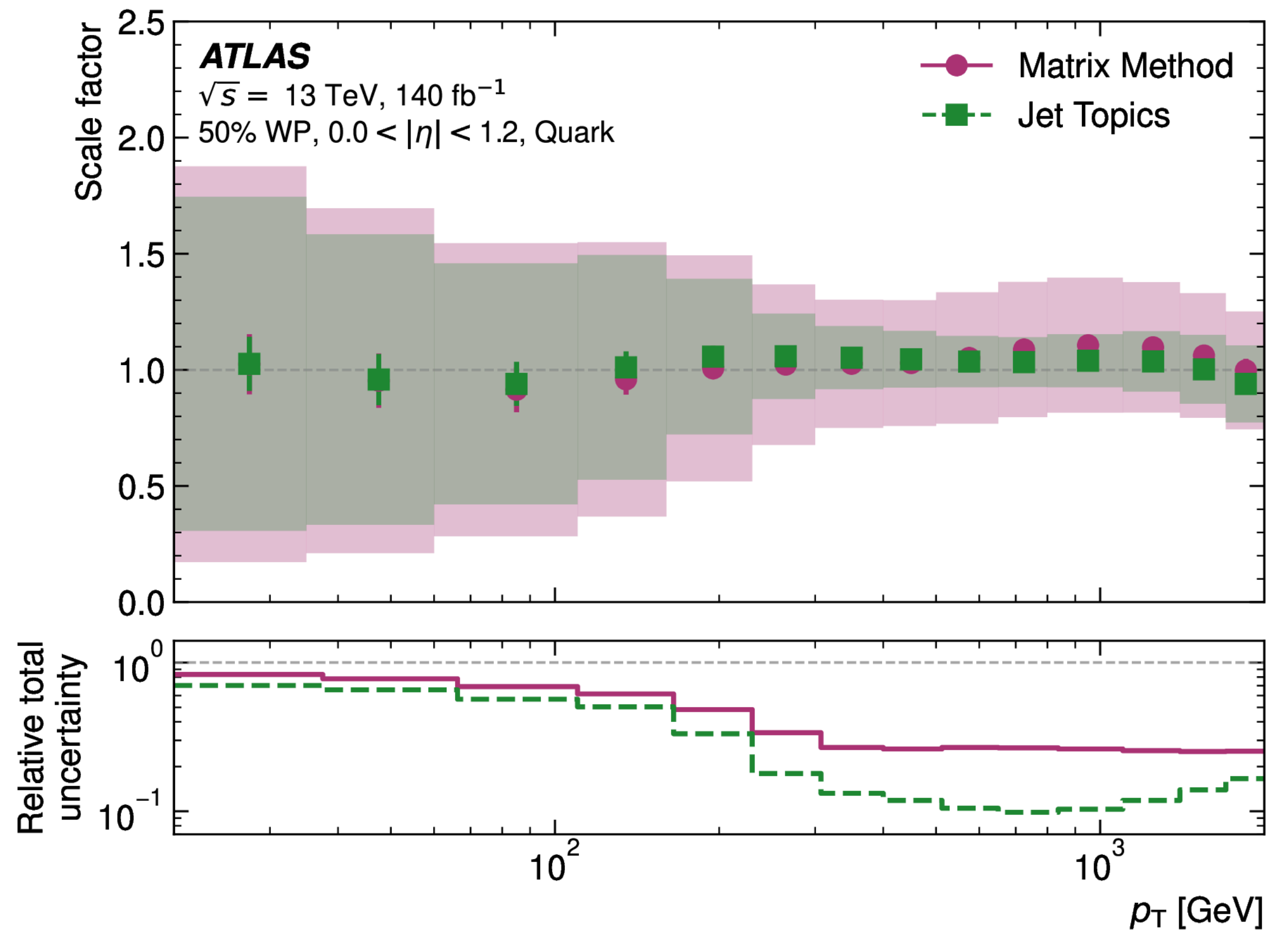


Jet topic: q/g calibration

$$p_{\text{quark}}(x) \equiv \frac{p_F(x) - \kappa_{FC} p_C(x)}{1 - \kappa_{FC}}$$

$$p_{\text{gluon}}(x) \equiv \frac{p_C(x) - \kappa_{CF} p_F(x)}{1 - \kappa_{CF}}$$

Impact: 20% decrease in experimental uncertainty

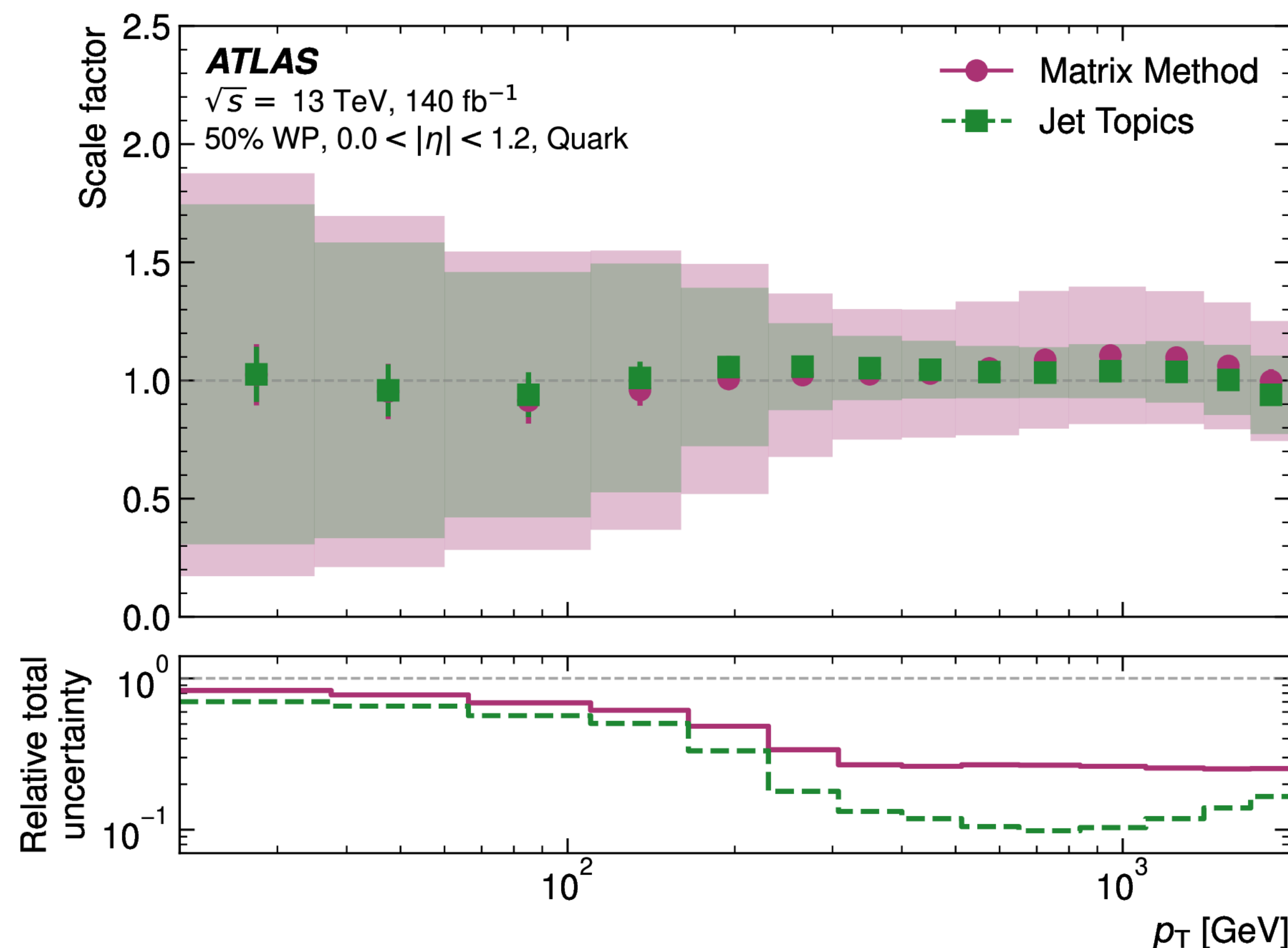
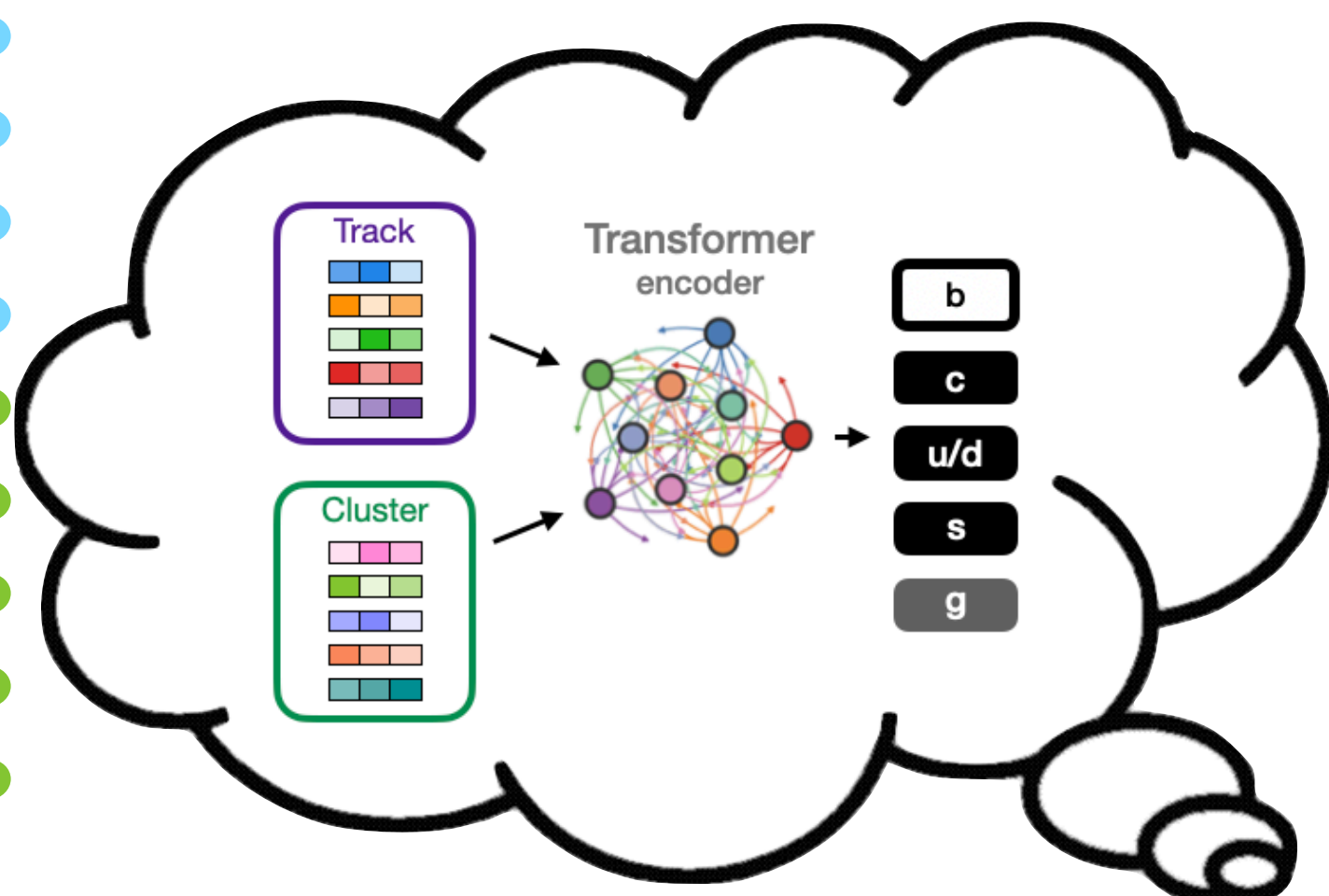


Jet topic: q/g calibration

$$p_{\text{quark}}(x) \equiv \frac{p_F(x) - \kappa_{FC} p_C(x)}{1 - \kappa_{FC}}$$

$$p_{\text{gluon}}(x) \equiv \frac{p_C(x) - \kappa_{CF} p_F(x)}{1 - \kappa_{CF}}$$

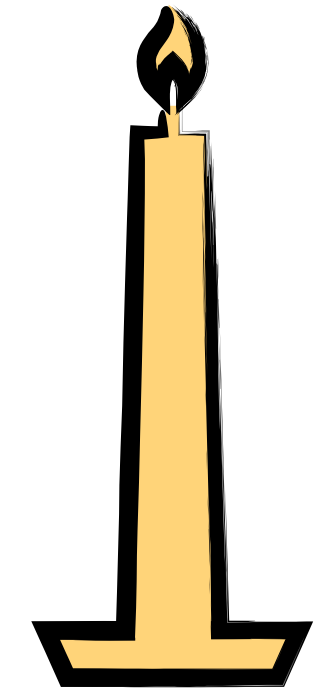
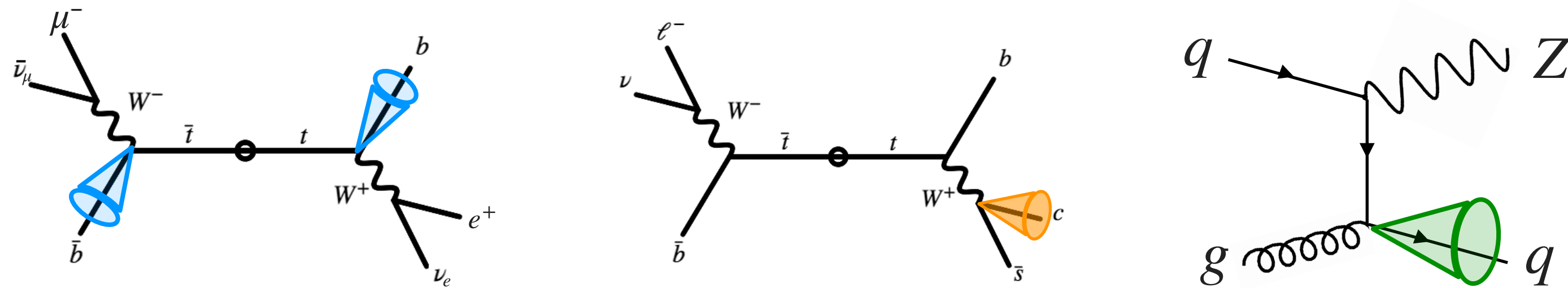
Impact: 20% decrease in experimental uncertainty



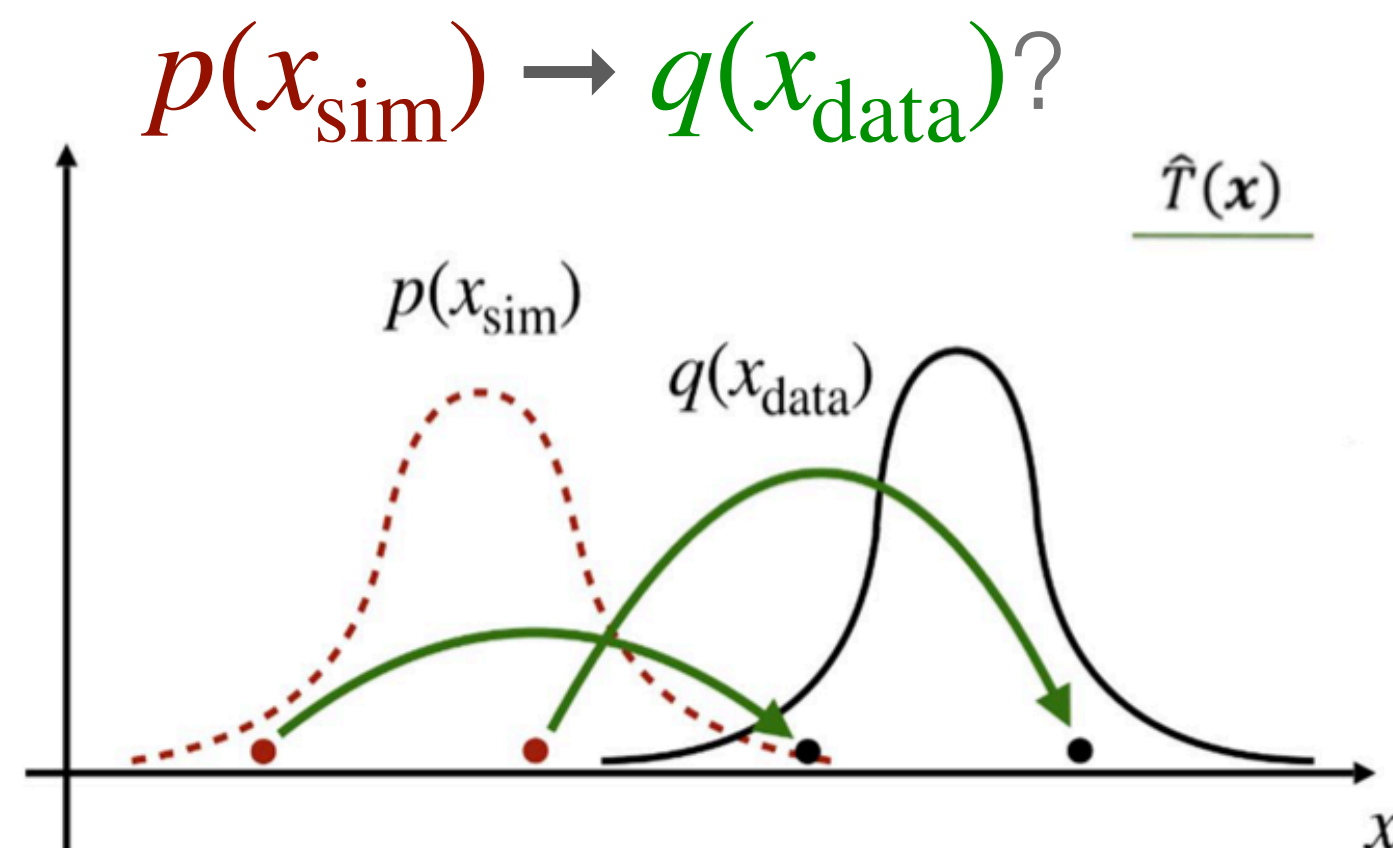
Multi-class extensions in Gregorio + Johann's talks this afternoon!

Extending to more dimensions

1) “Standard candles” to derive (and validate) uncertainties.



2) A probability dist in high dimenstions = generative models



ATLAS, [2505.13063](#)

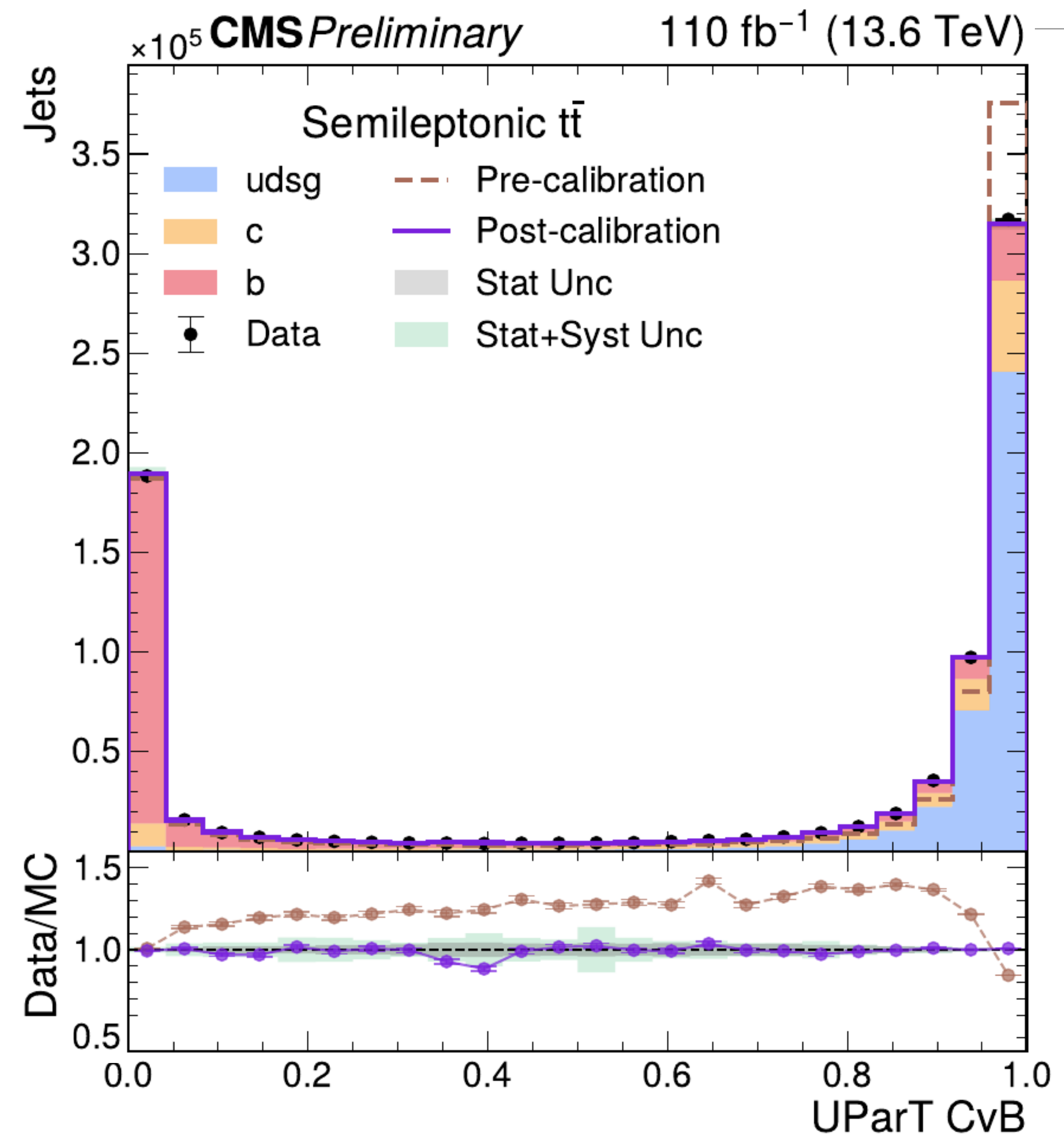
CMS, [DP-2026-106](#)



Learning Label Proportions + OT calibration

1) Correct the flavor fractions

- Train in data to learn the flavor fractions
Dery, Nachman, Rubbo, Schwartzman, [1702.00414](#)
- “Soft label” for data





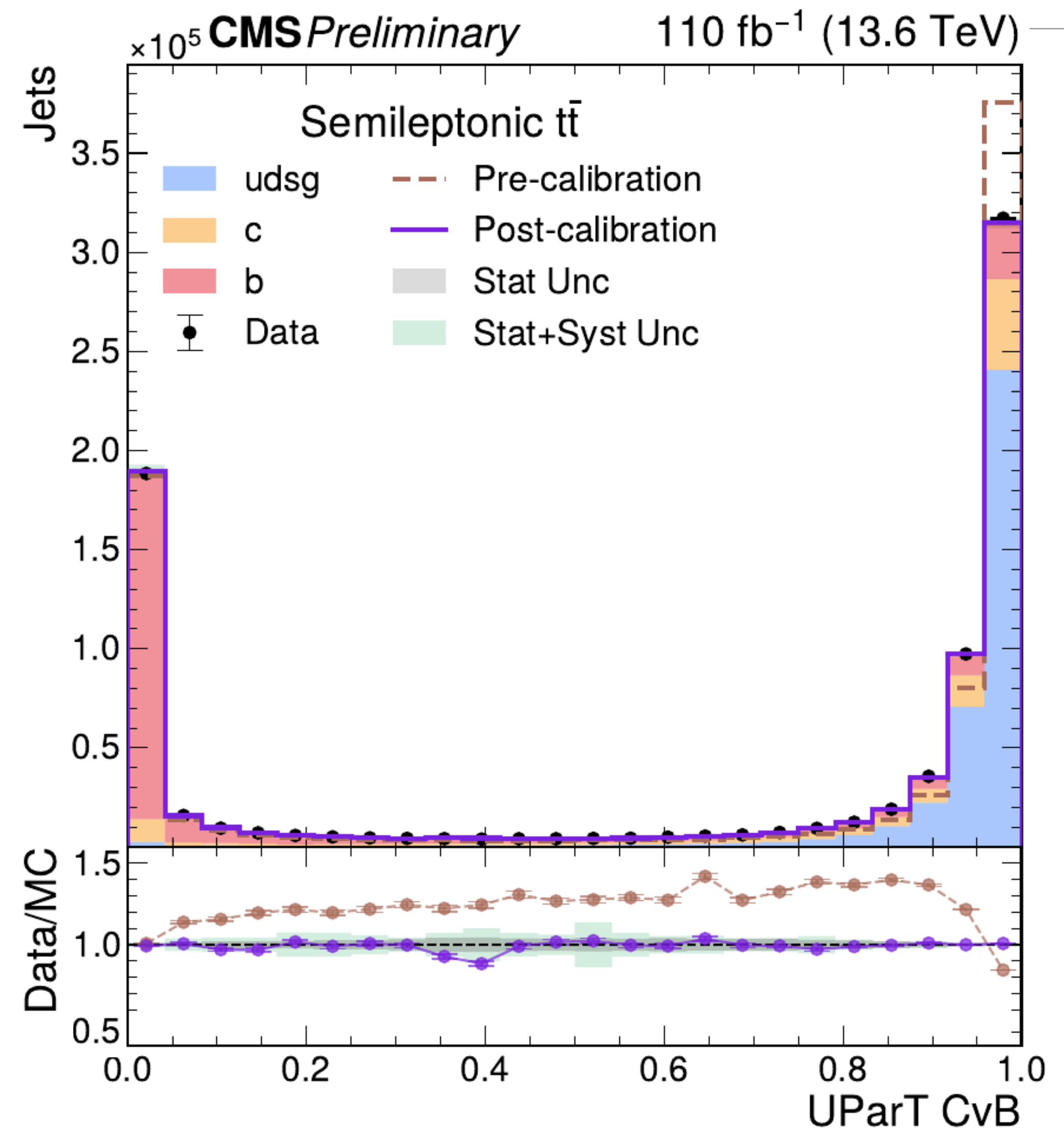
Learning Label Proportions + OT calibration

1) Correct the flavor fractions

- Train in data to learn the flavor fractions
Dery, Nachman, Rubbo, Schwartzman, [1702.00414](#)
- “Soft label” for data

2) Map from MC to data with conditional flow matching

- Learn the velocity field from MC $\rightarrow$ data
- Solve the OT per mini-batch
 - Also condition maps on pT and jet flavor





Learning Label Proportions + OT calibration

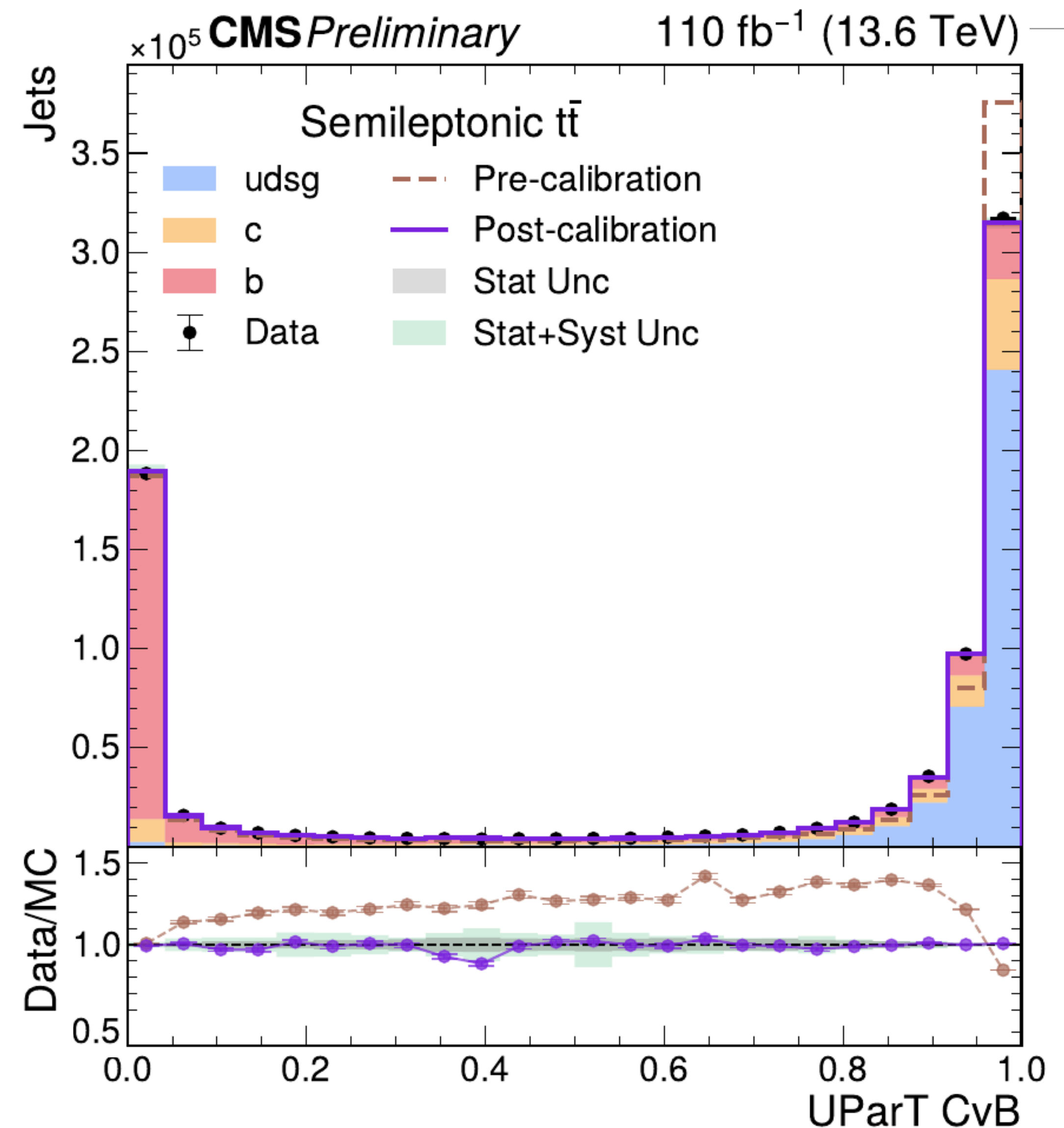
1) Correct the flavor fractions

- Train in data to learn the flavor fractions
Dery, Nachman, Rubbo, Schwartzman, [1702.00414](#)
- “Soft label” for data

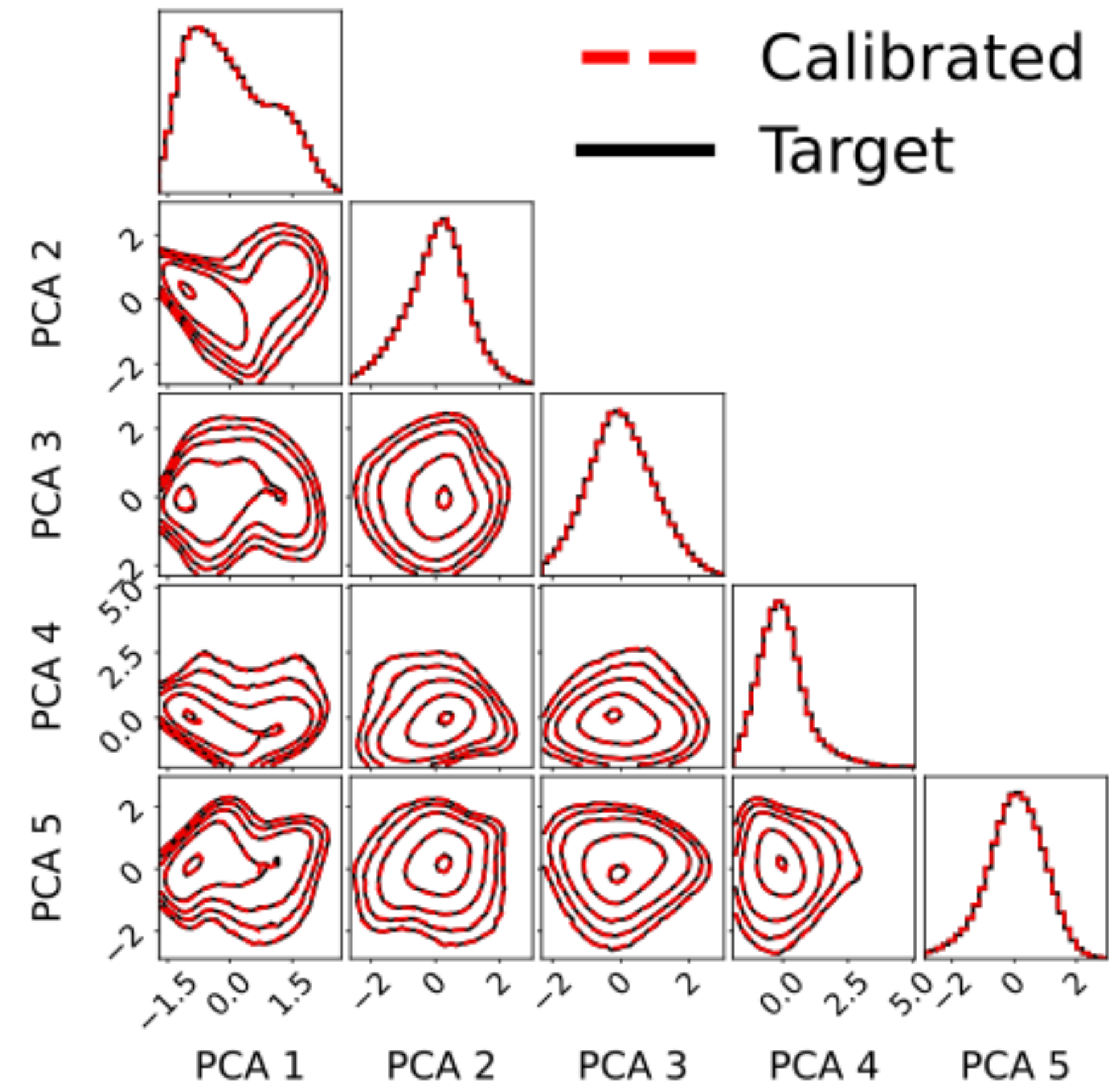
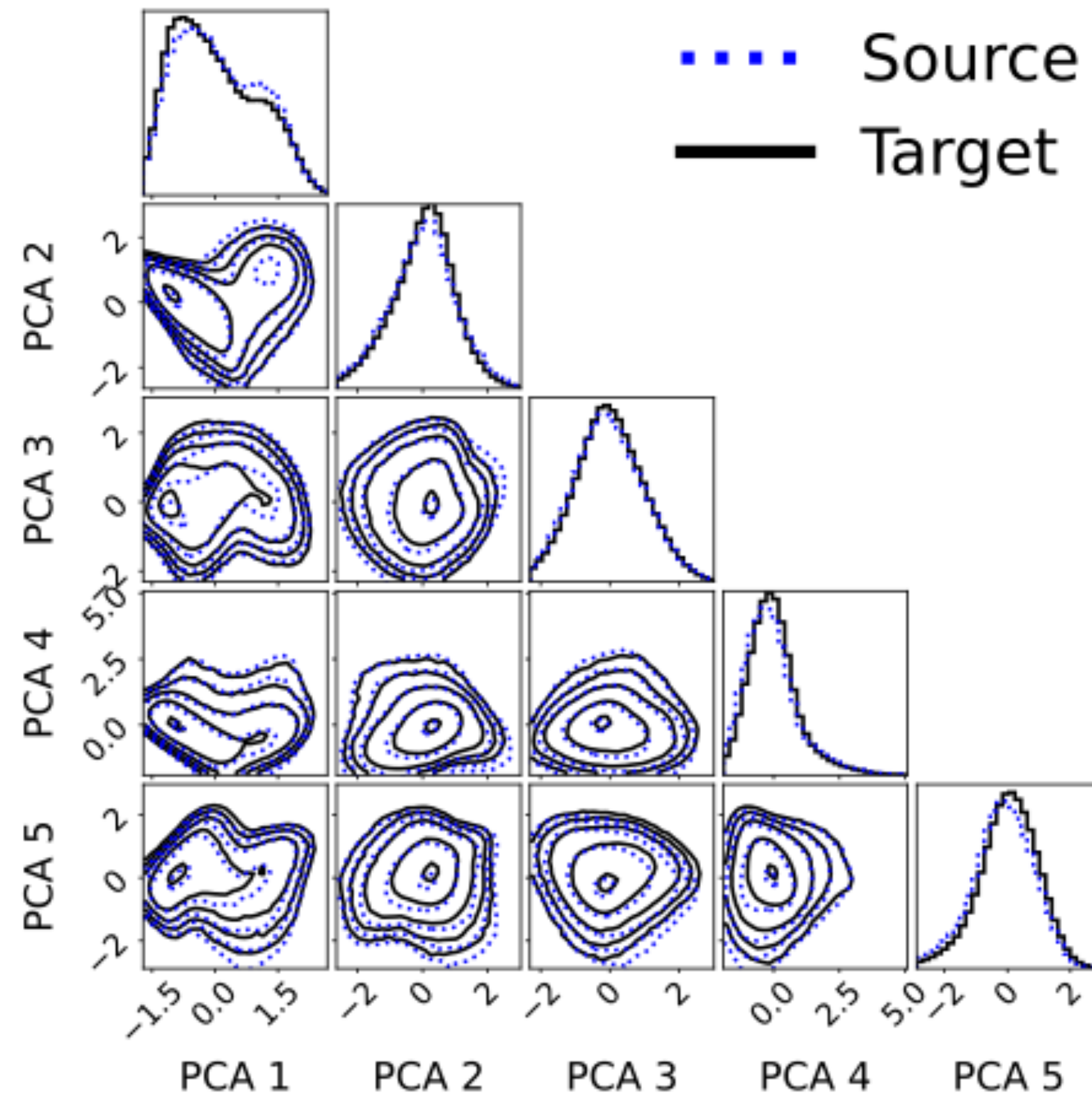
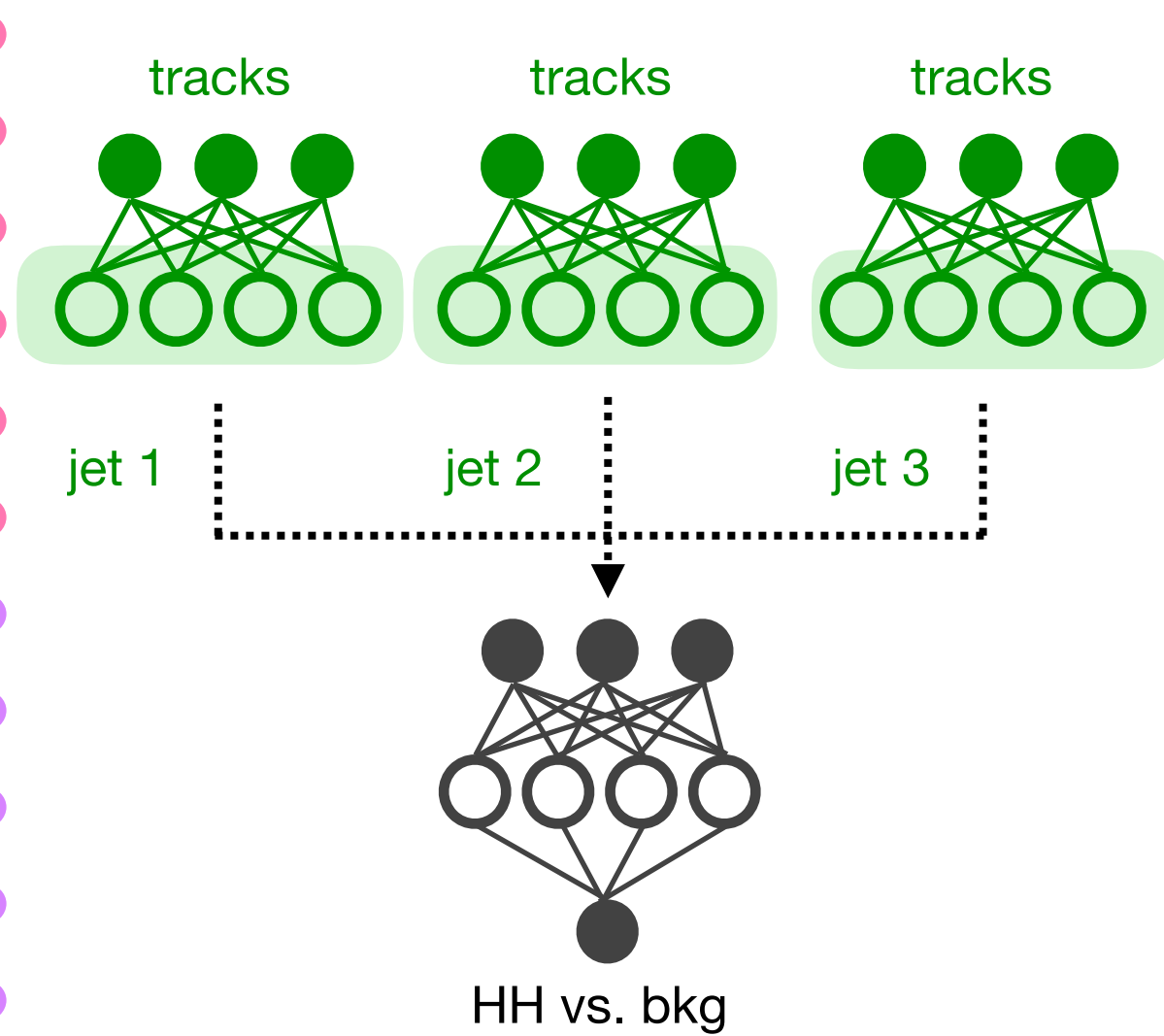
2) Map from MC to data with conditional flow matching

- Learn the velocity field from MC $\rightarrow$ data
- Solve the OT per mini-batch
 - Also condition maps on pT and jet flavor

Calibrations for *all three flavors*: b-jets ($t\bar{t}$), c-jets (W+c), light-jets (Drell-Yann)



Can we calibrate this latent space?



YES! OT (normalizing flows) can calibrate high dim latent space!

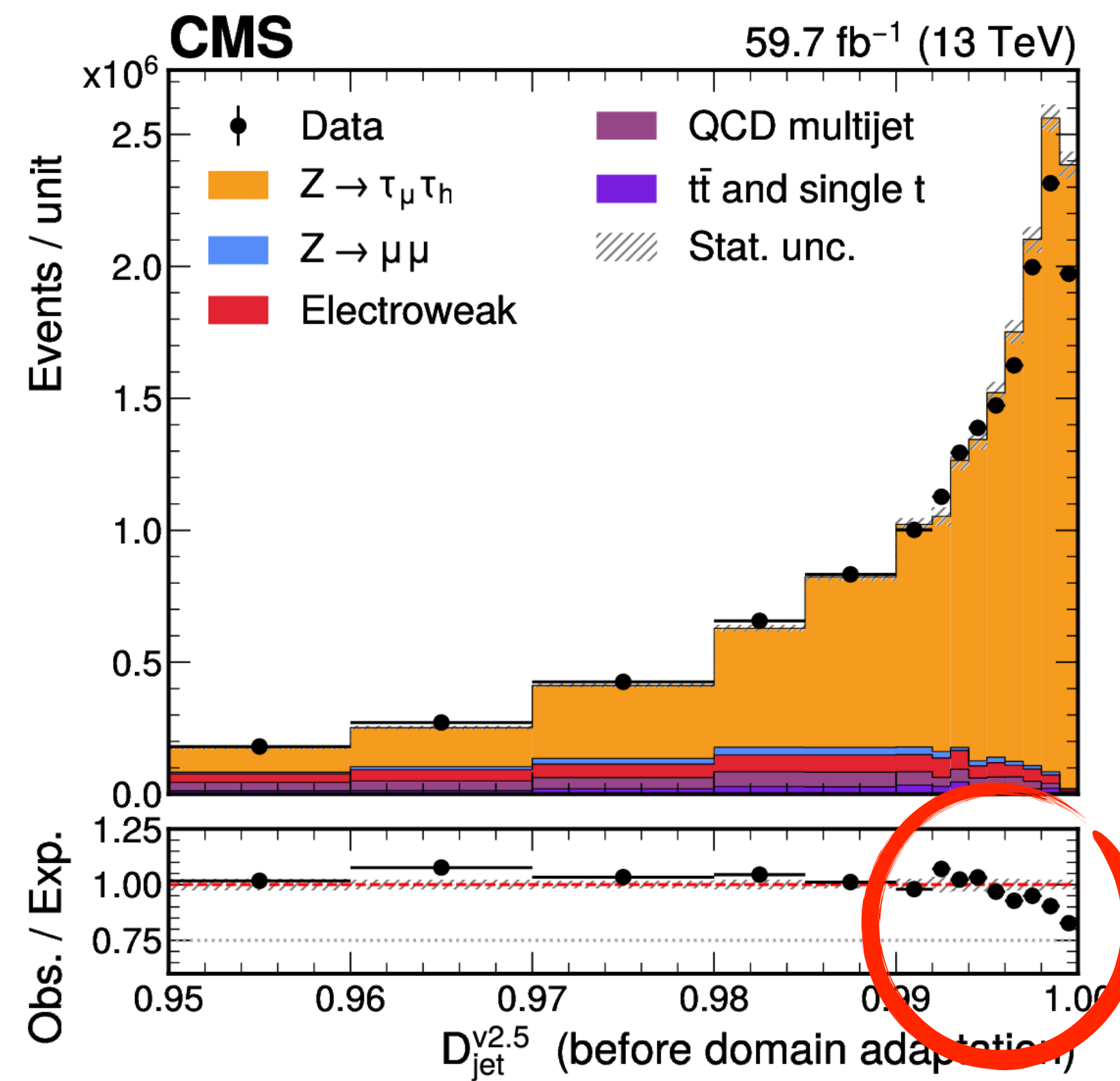
Domain adaptation

You want the data MC to agree? Put it in the loss function!

$$\mathcal{L} = \mathcal{L}_{\text{class}}$$

Standard supervised
learning loss

Ex: hadronic τ identification



Domain adaptation

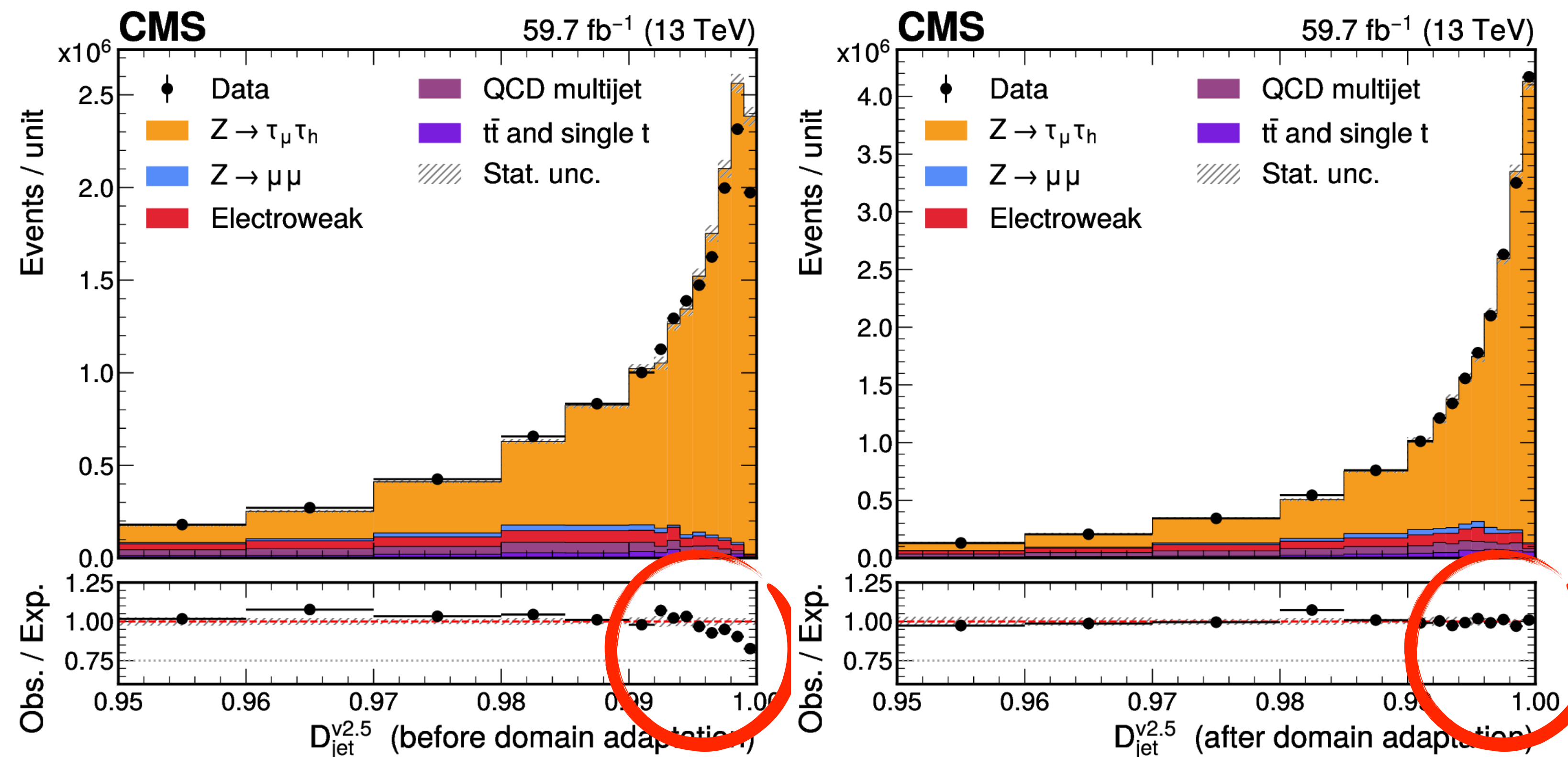
You want the data MC to agree? Put it in the loss function!

$$\mathcal{L} = \mathcal{L}_{\text{class}} - \kappa \cdot \mathcal{L}_{\text{adv}}$$

Standard supervised learning loss Adversary: fail to discriminate data / mc

Ex: hadronic τ identification

Impact: 20% non-closure improved to 1%!

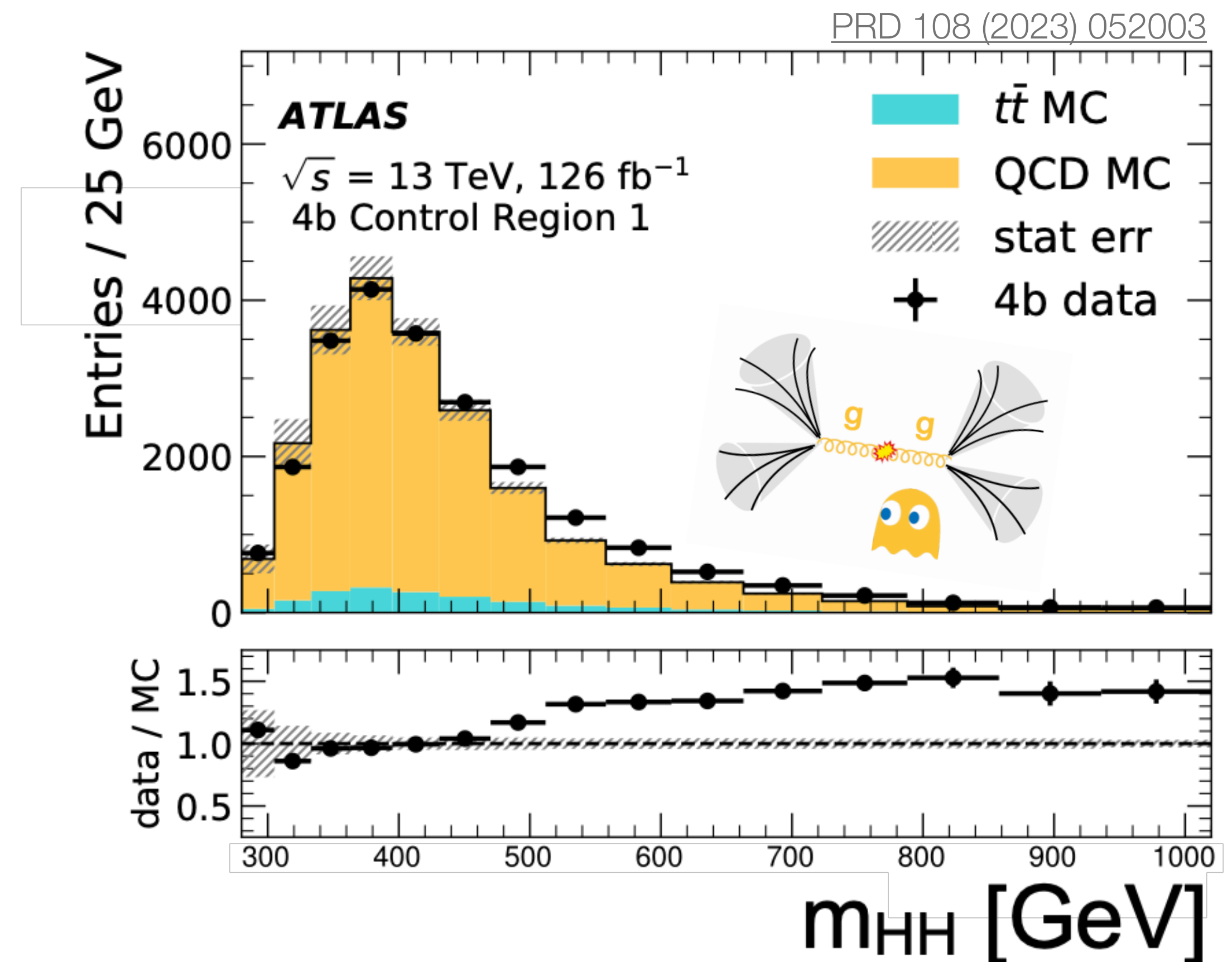


Analysis: Background estimation

Fact: No hadronic analysis relies on MC for background estimation*

Challenges: (1) Accuracy
(2) Speed

More in [Theo](#) and [Michal's](#) talks tmr!



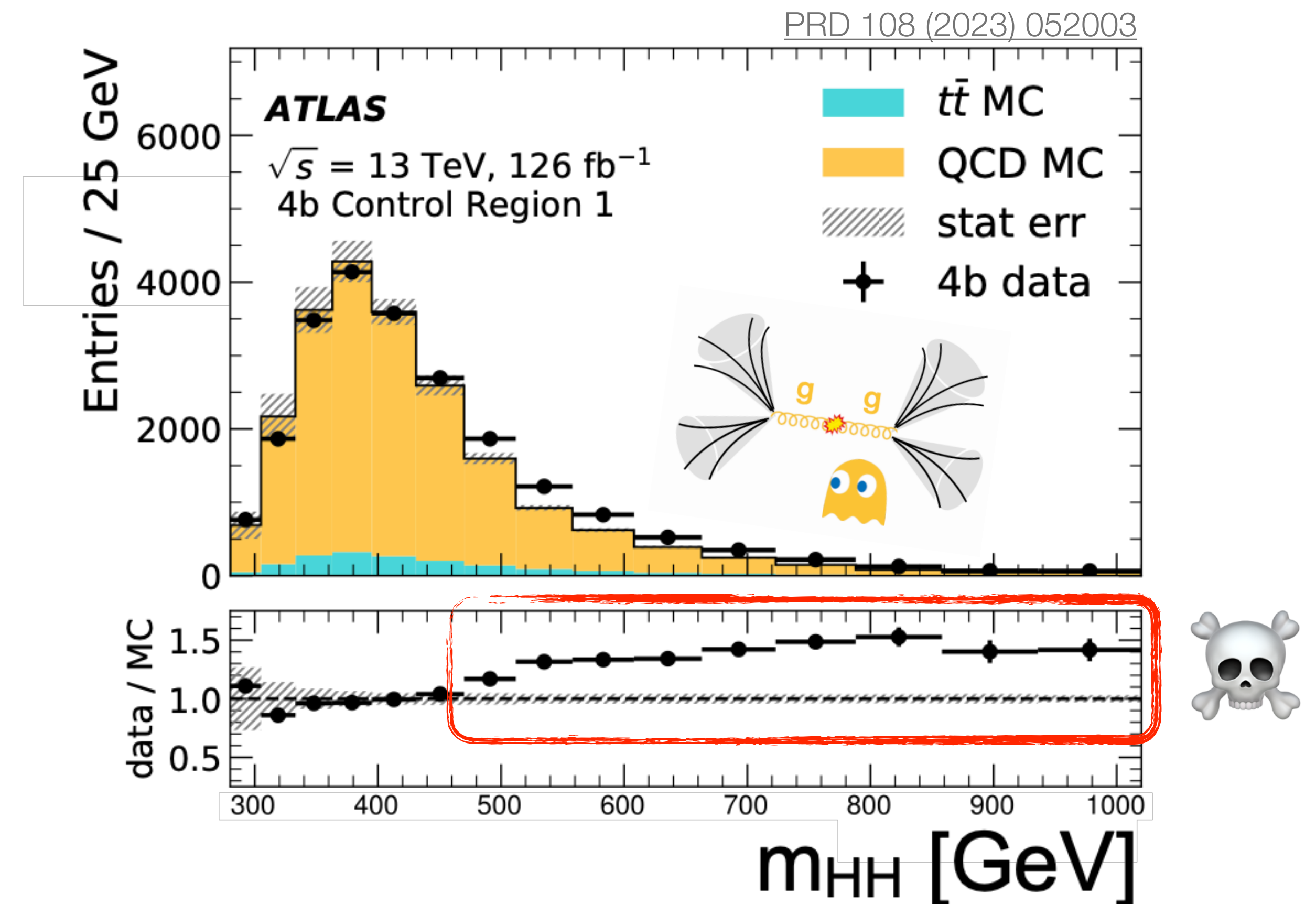
*If you have an example, please enlighten me!

Analysis: Background estimation

Fact: No hadronic analysis relies on MC for background estimation*

Challenges: (1) Accuracy
(2) Speed

More in [Theo](#) and [Michal's](#) talks tmr!



*If you have an example, please enlighten me!

Analysis: Background estimation

Fact: No hadronic analysis relies on MC for background estimation*

**Challenges: (1) Accuracy
(2) Speed**

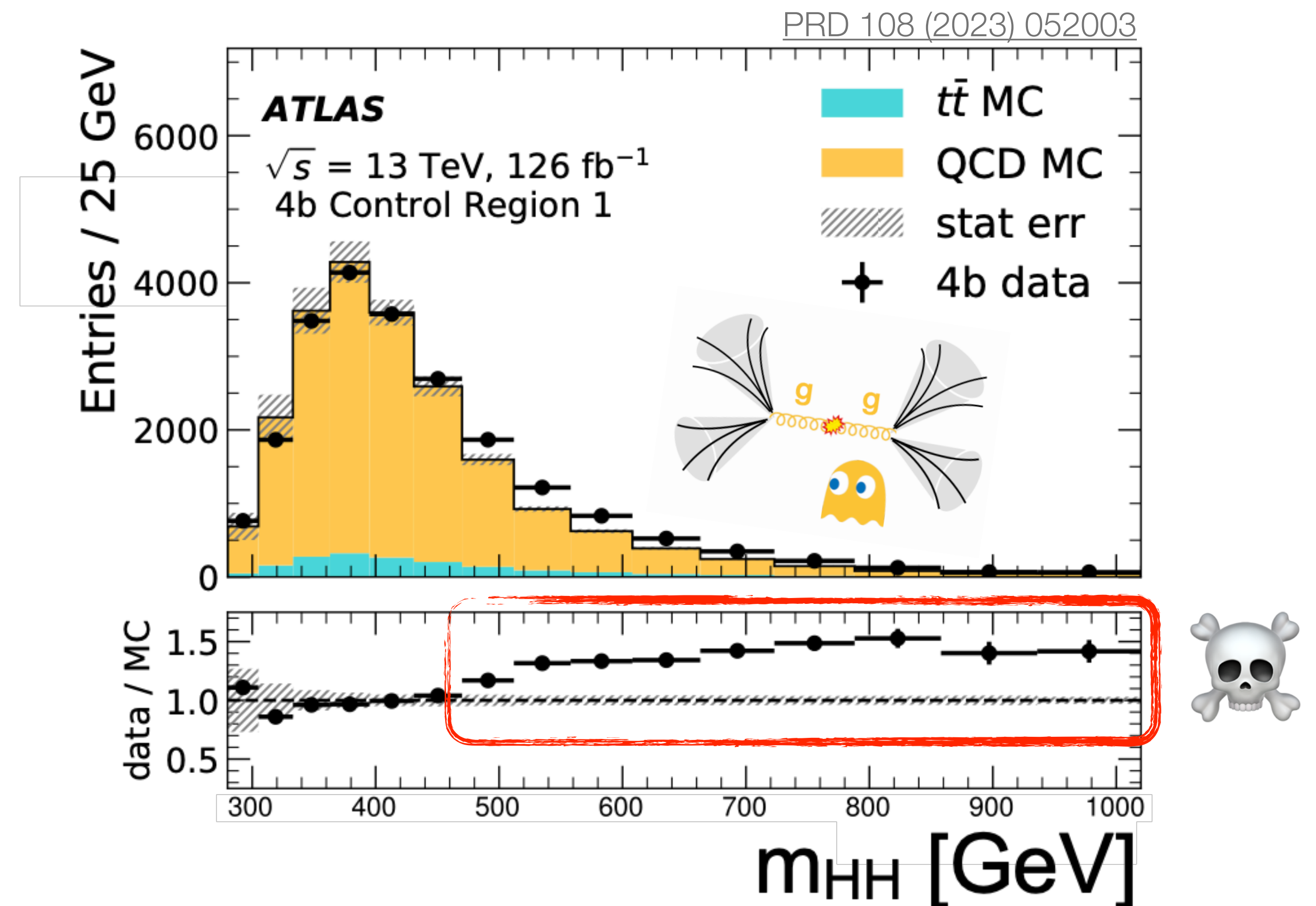
More in [Theo](#) and [Michal's](#) talks tmr!

Solution: Data driven bkg est

- Likelihood ratio estimation

ATLAS: 4b+MET [PRD 109 \(2024\) 112011](#); VH(qqbb) [PRL 132 \(2024\) 131802](#);
HH4b res [PRD 105 \(2022\) 092002](#); nonres [PRD 108 \(2023\) 052003](#)
CMS: HH4b, [PRL 129 \(2022\) 081802](#), [2604.27044](#)
+ many others

- Density estimation



*If you have an example, please enlighten me!

Analysis: Background estimation

Fact: No hadronic analysis relies on MC for background estimation*

**Challenges: (1) Accuracy
(2) Speed**

More in [Theo](#) and [Michal](#)'s talks tmr!

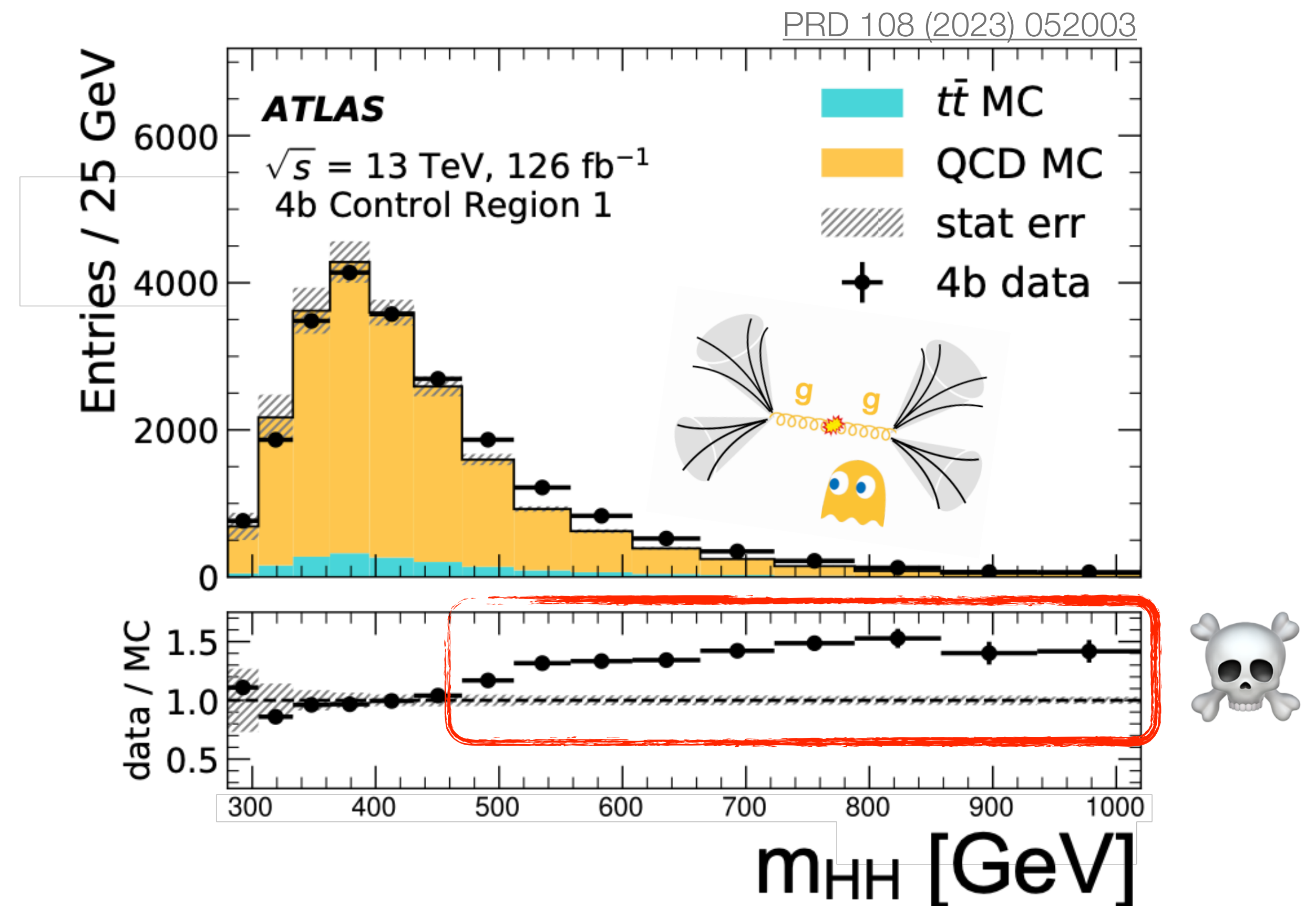
Solution: Data driven bkg est

- Likelihood ratio estimation

ATLAS: 4b+MET [PRD 109 \(2024\) 112011](#); VH(qqbb) [PRL 132 \(2024\) 131802](#);
HH4b res [PRD 105 \(2022\) 092002](#); nonres [PRD 108 \(2023\) 052003](#)
CMS: HH4b, [PRL 129 \(2022\) 081802](#), [2604.27044](#)
+ many others

- Density estimation

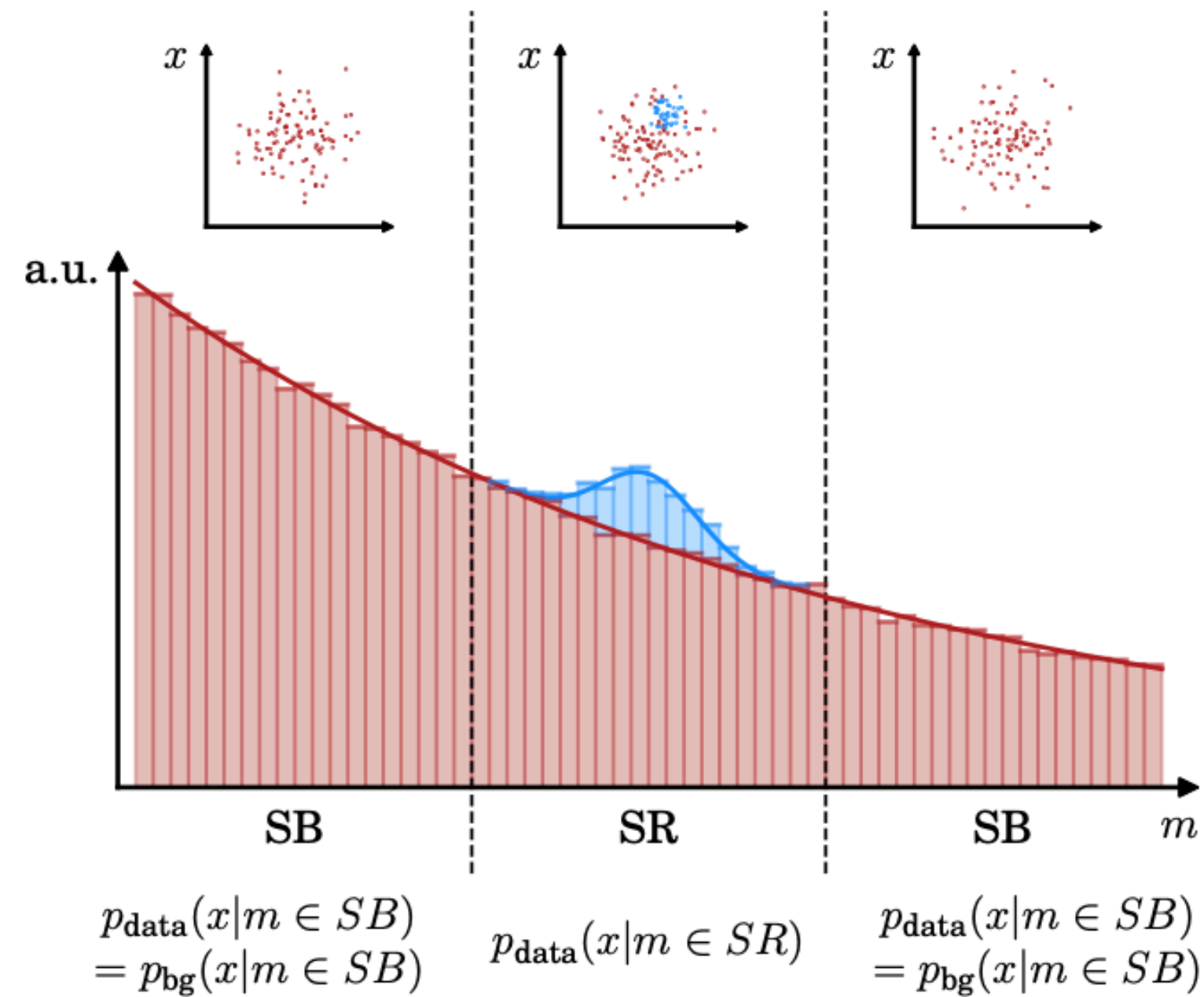
Generative models: $p_{\phi}(x)$



*If you have an example, please enlighten me!

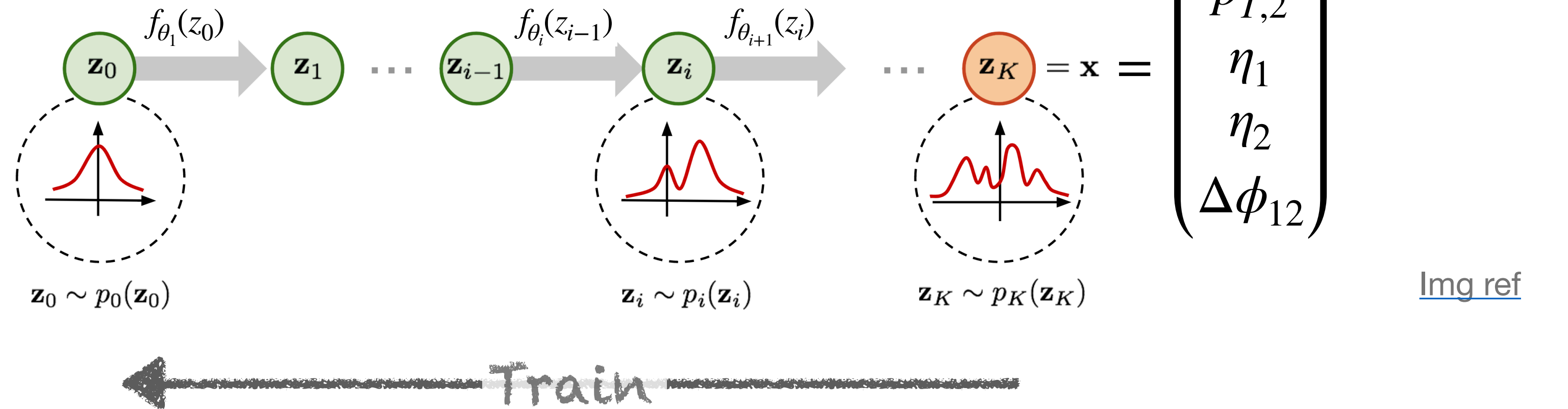
Learn $p_\phi(x | m)$: Normalising flows for background estimation

Mass: control variable to define signal-depleted regions

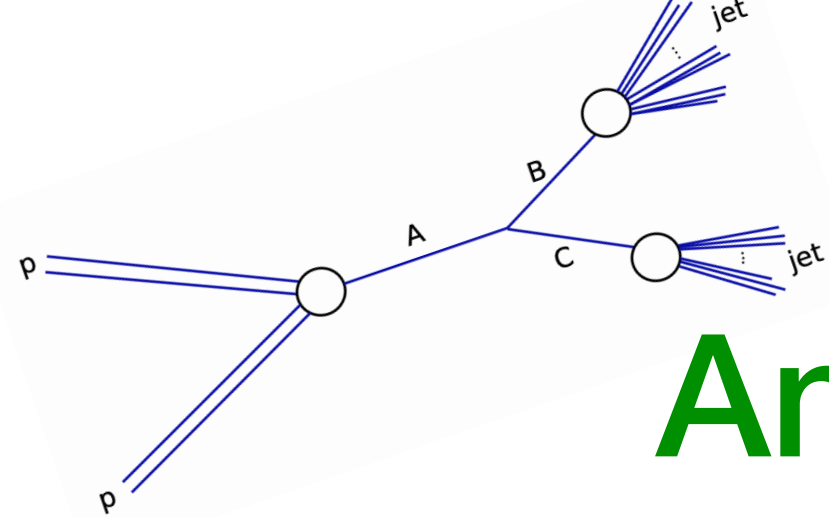


CATHODE

Hallin, et. al, PRD 106 (2022) 055006



$$\mathcal{L}oss = -\log p_\phi(x | m) = -\log p_z(f_\phi^{-1}(x | m)) - \sum_{i=1}^K \left| \frac{\partial f_{\phi_i}^{-1}}{\partial z_i^T} \right|$$

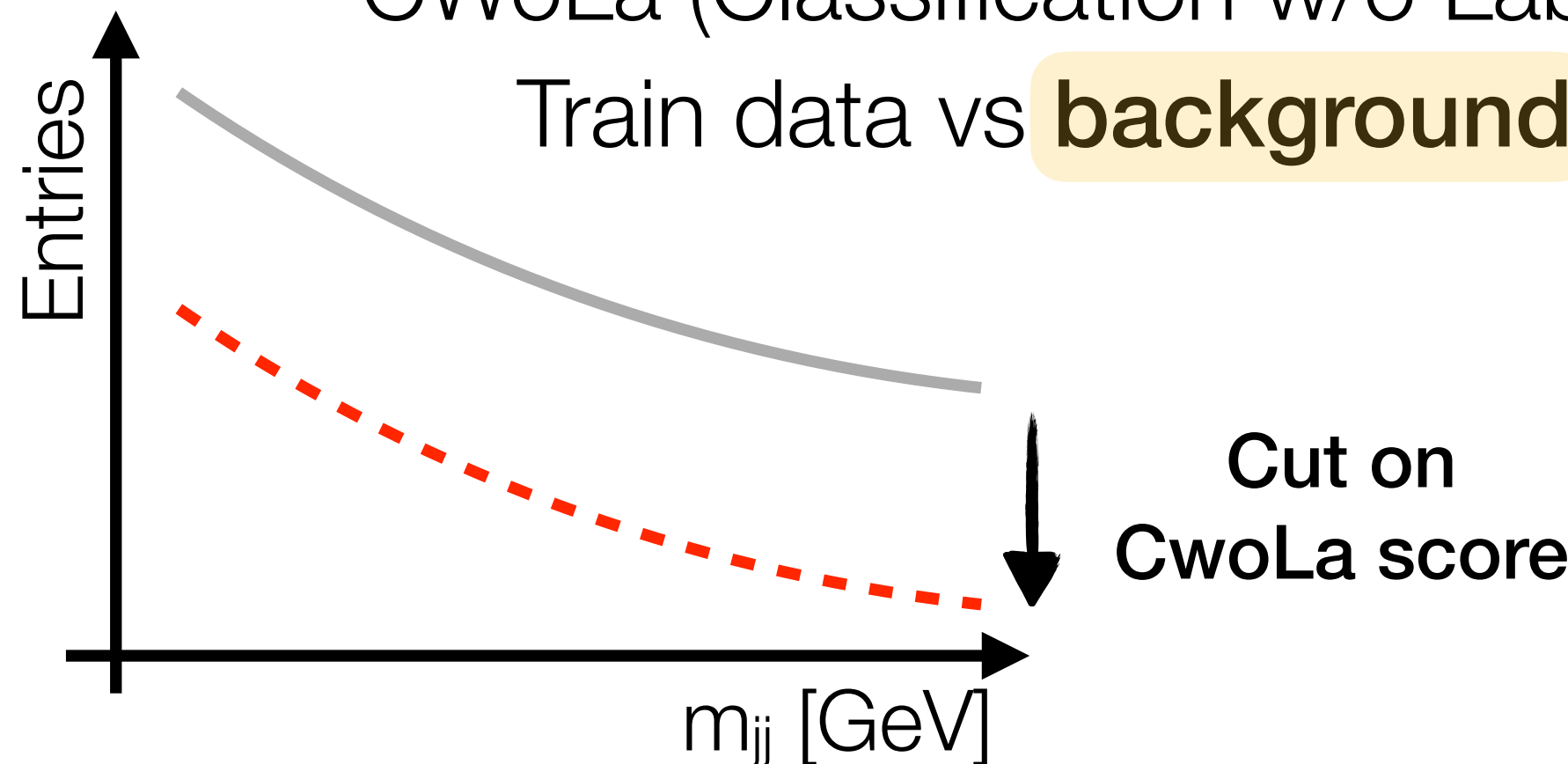


Anomaly Detection

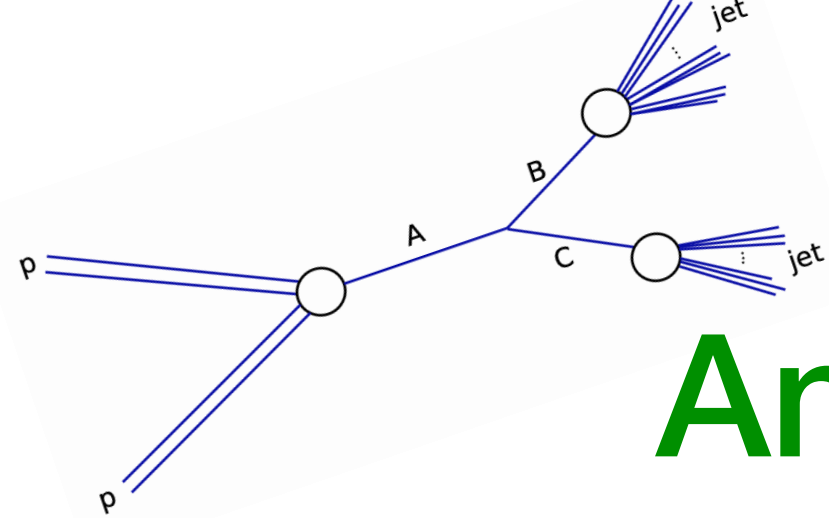
BSM Higgs searches

CWoLa (Classification w/o Labels)

Train data vs **background**

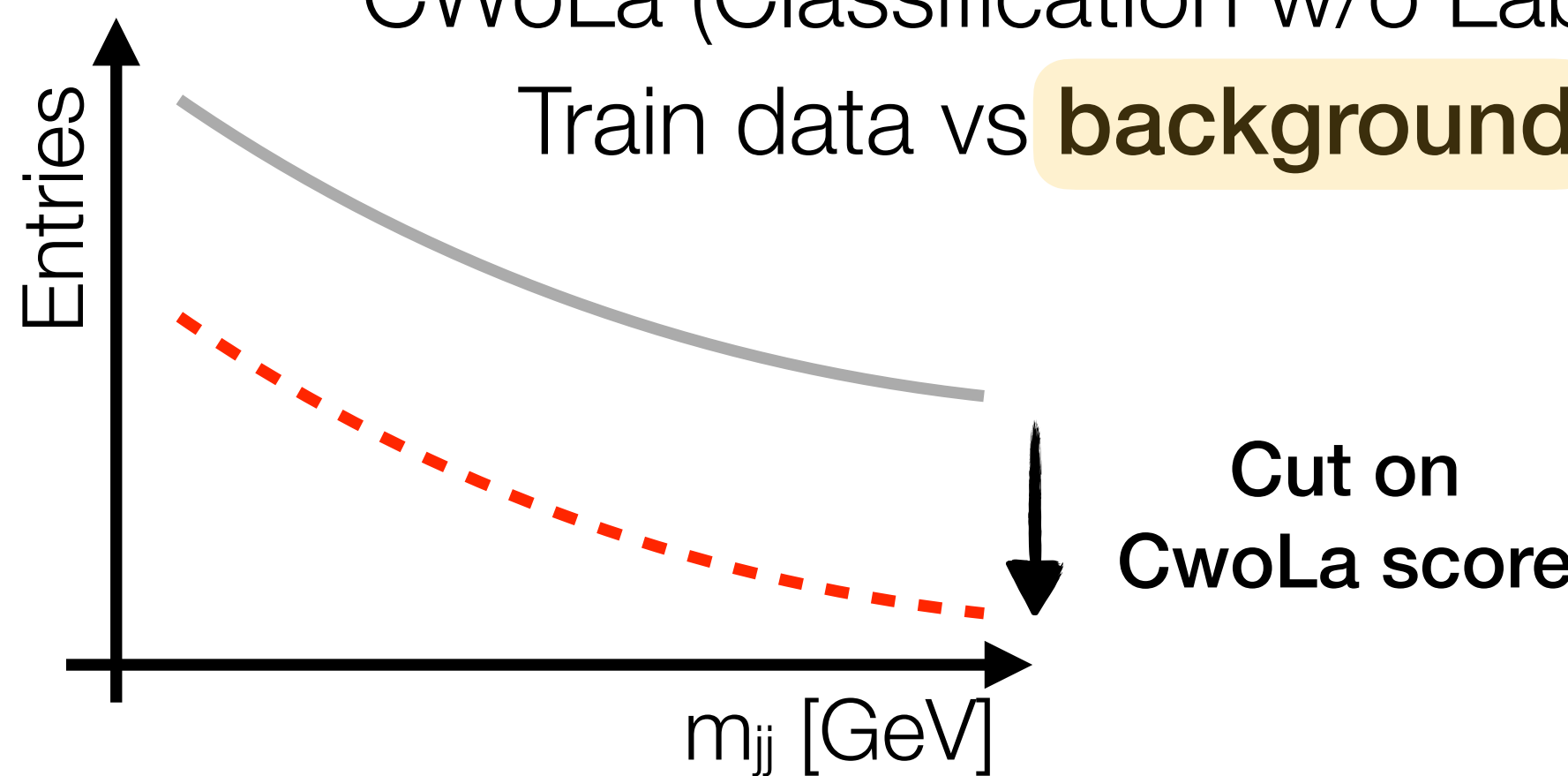


	Method	Bkg estimate
CMS Rep. Prog. Phys. 88 (2025) 067802	CWoLA	Use m_{jj} sidebands (SB)
	(TNT)	(+ “normal” jets (jet-AE))
	CATHODE	Norm flow to interpolate from m_{jj} SB
		QUAK: Norm flows estimate the density of S and B VAE-QR (1d cnn) over particle flow object $\vec{p}$
ATLAS PRD 112 (2025) 072009	CURTAINS	Norm flow & OT to map b/w m_{jj} SBs
	SALAD	NN reweight tsim to data in SBs



Anomaly Detection

CWoLa (Classification w/o Labels)
Train data vs **background**



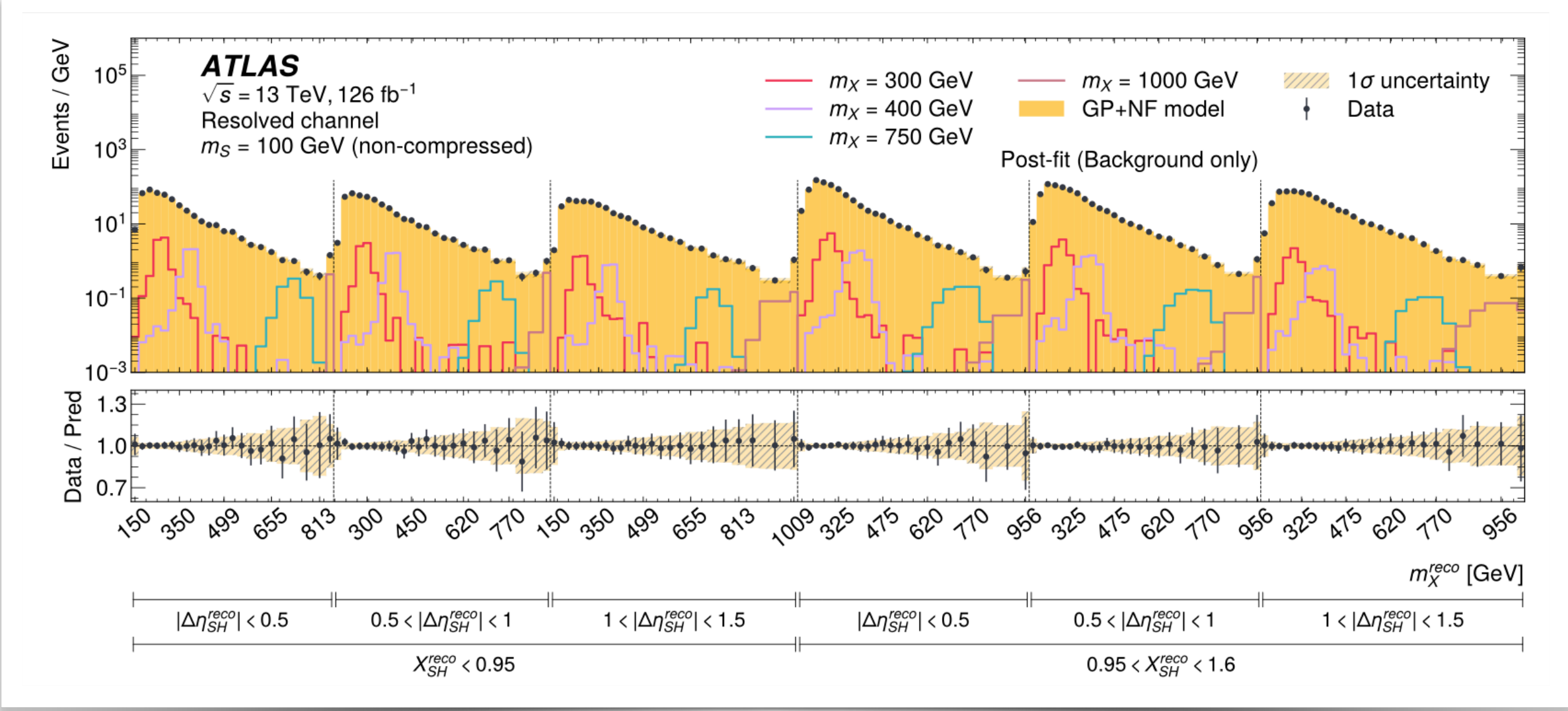
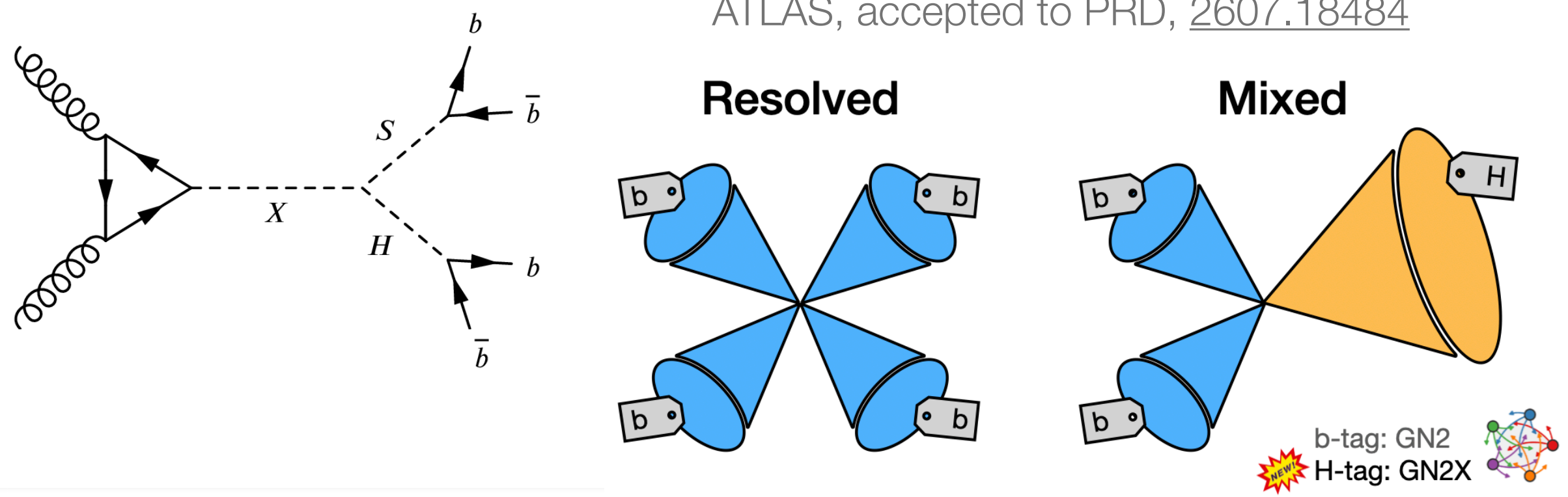
	Method	Bkg estimate
CMS Rep. Prog. Phys. 88 (2025) 067802	CWoLa (TNT)	Use m_{ij} sidebands (SB) (+ “normal” jets (jet-AE))
	CATHODE	Norm flow to interpolate from m_{ij} SB
	QUAK: Norm flows estimate the density of S and B VAE-QR (1d cnn) over particle flow object $\vec{p}$	
ATLAS PRD 112 (2025) 072009	CURTAINS	Norm flow & OT to map b/w m_{ij} SBs
	SALAD	NN reweight tsim to data in SBs

Review: Golling et. al, EPJC 84 (2024) 241; also Mark Fanselow’s talk (Thurs)!

BSM Higgs searches

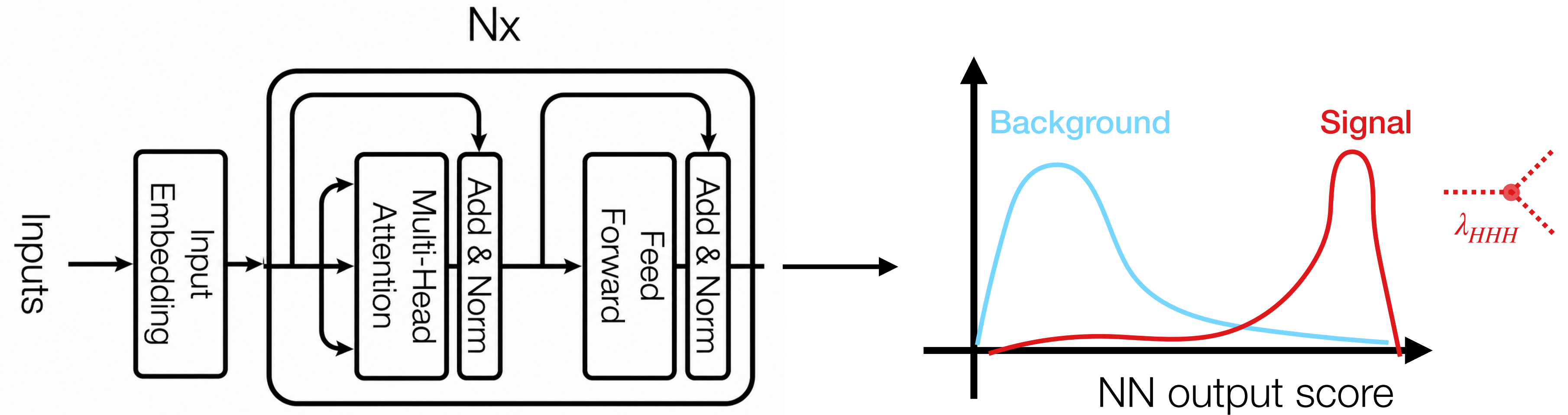
NEW!

ATLAS, accepted to PRD, 2607.18484



Excellent post-fit agreement!
Improved nonclosure c.f. NN reweighting baseline

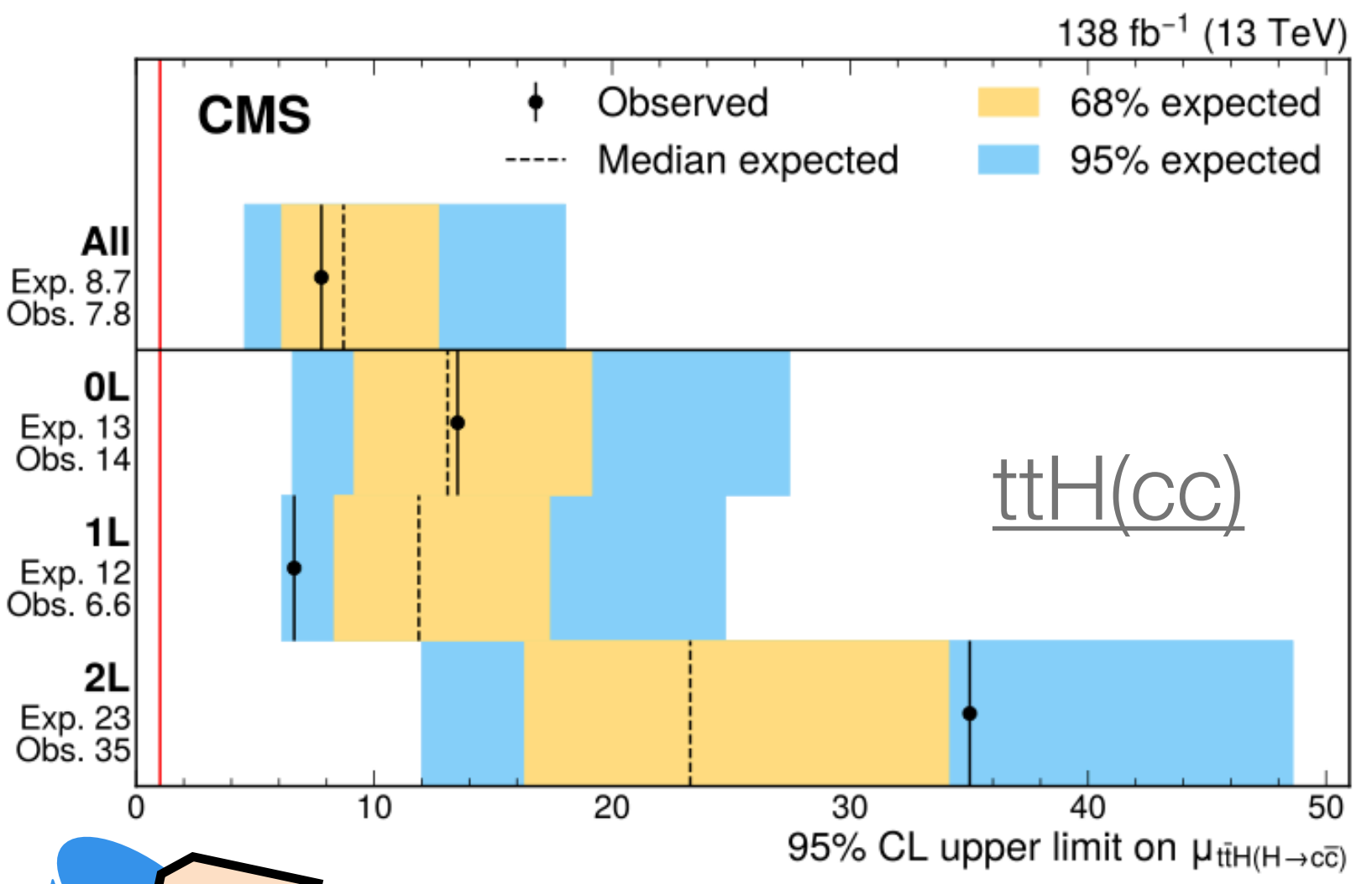
Analysis: optimal discriminants



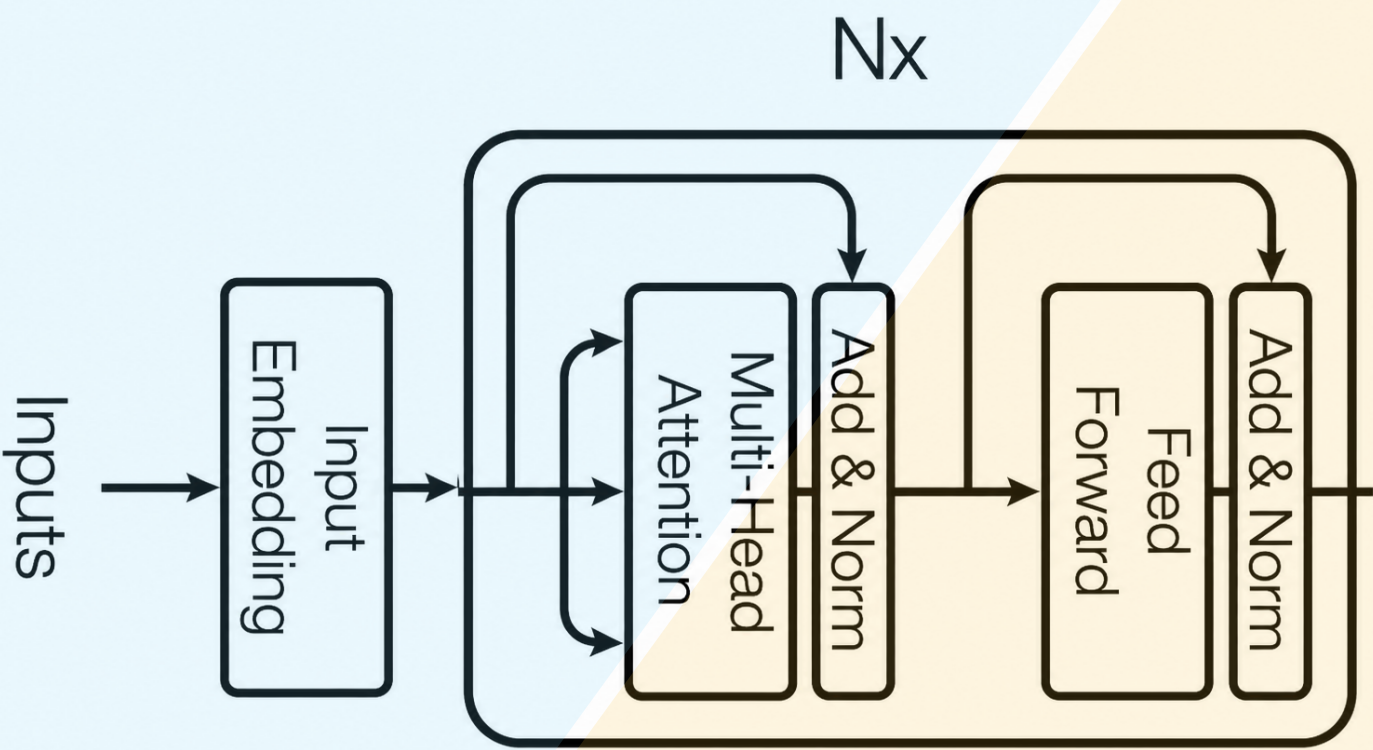
Transformer = general-purpose backbone for encoding correlations across objects in an event.

Physics impact: transformers for analysis

Charm yukawa coupling



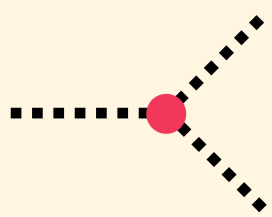
World leading limits



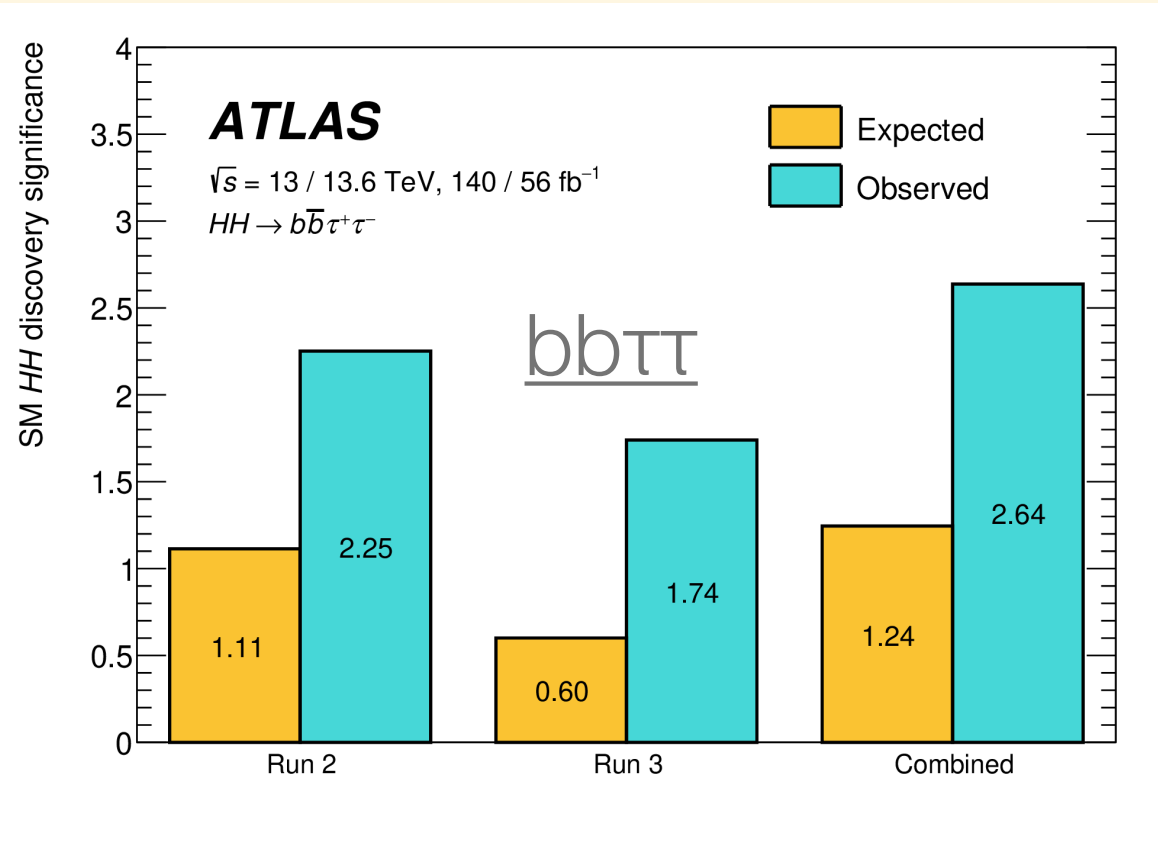
2604.27044



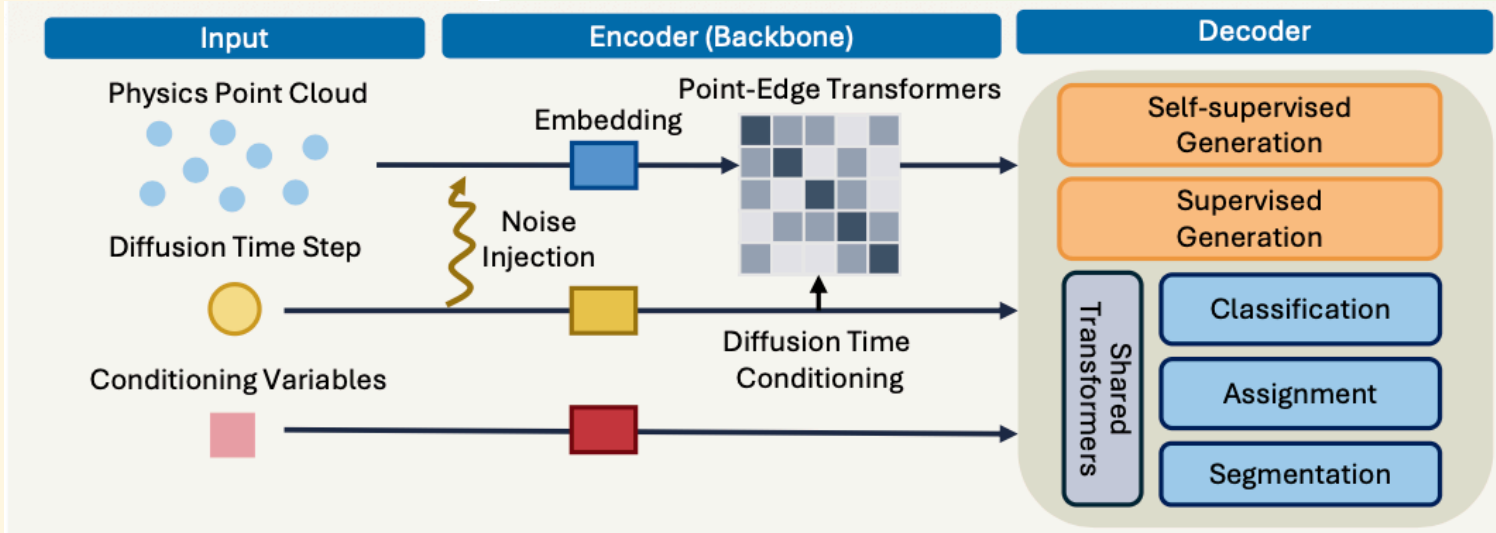
Also $HH \rightarrow$, $b\bar{b}yy$, $b\bar{b}WW$



HH sensitivity

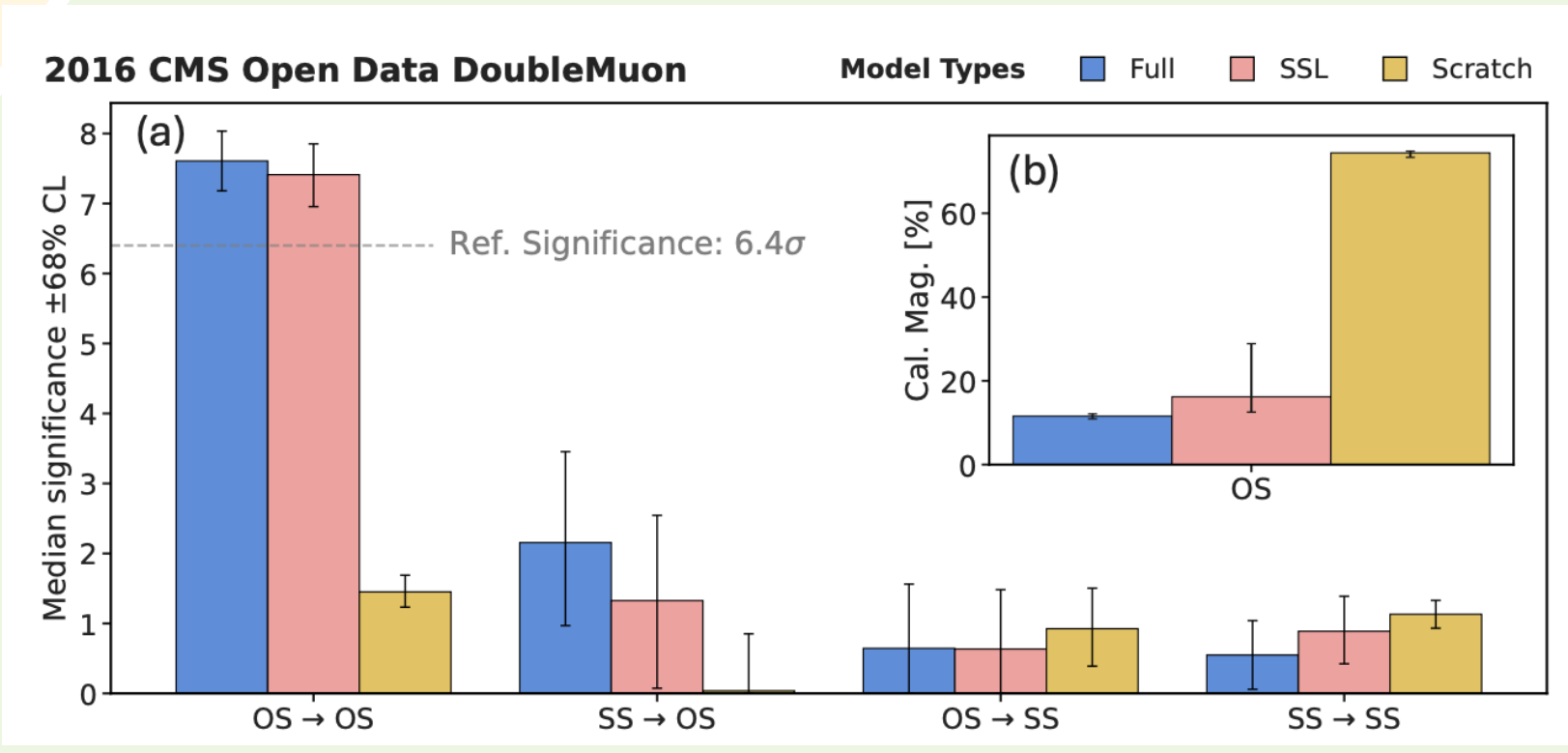
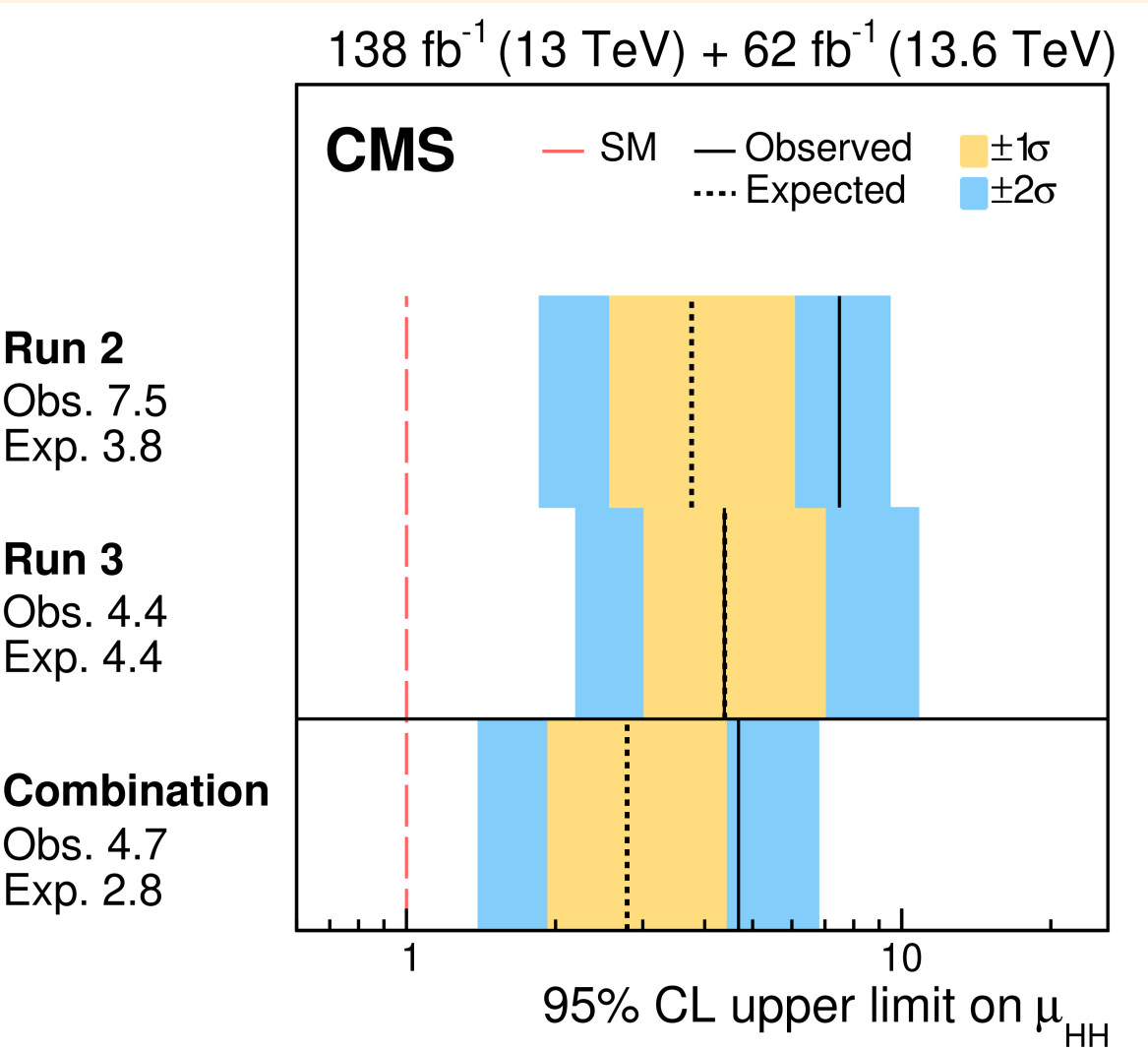


Obs (exp)
2.6σ (1.2σ)



Do *all* the tasks at once

→ Rediscover $\mathcal{I}(\mu\mu)$



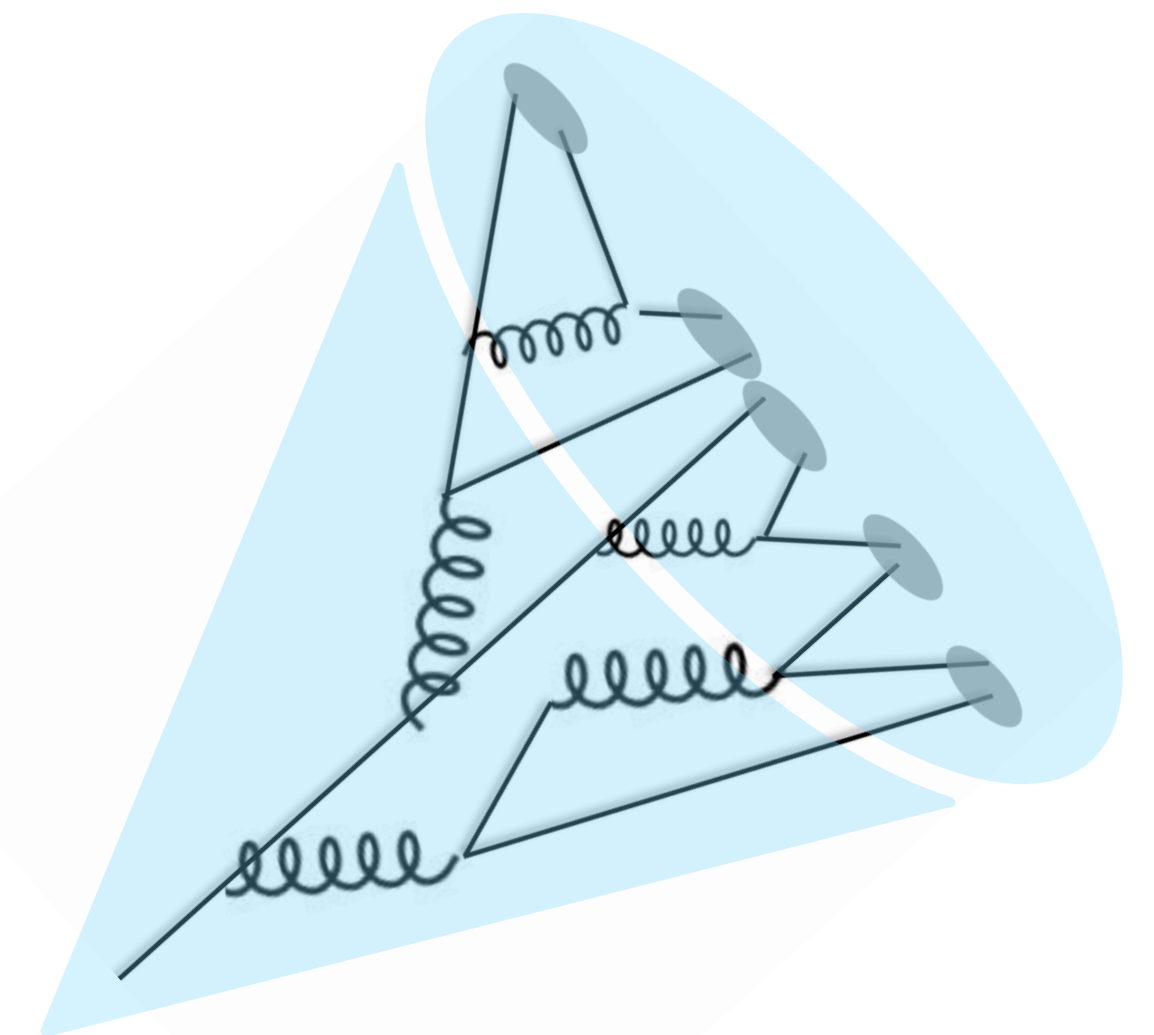
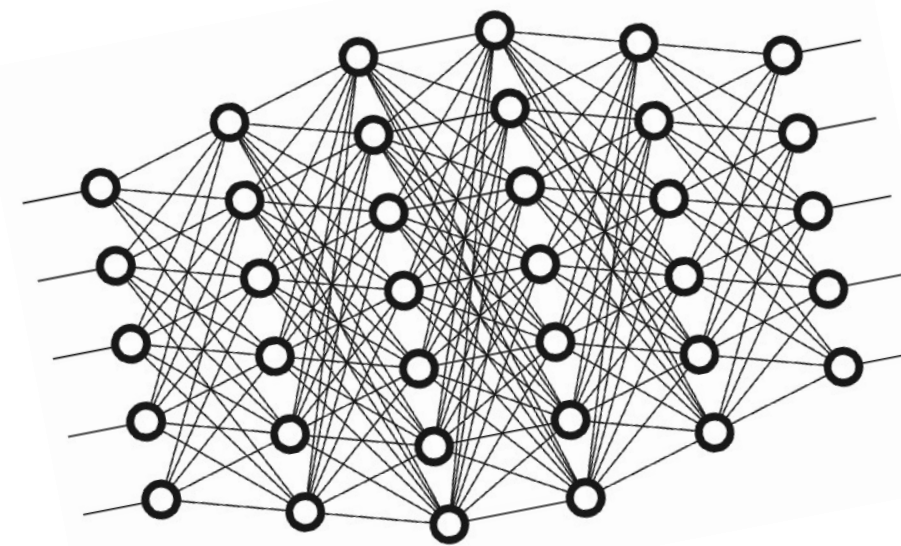
2601.17126

See also Vinny's [talk](#) & Anna's [overview](#) on FM



“There are a lot of deep mysteries about how the universe works... Most of the things we care about we don’t understand...

The nature of reality — that’s ultimately what I’m after. And I want to use AI as a tool to help us understand the nature of reality around it.”

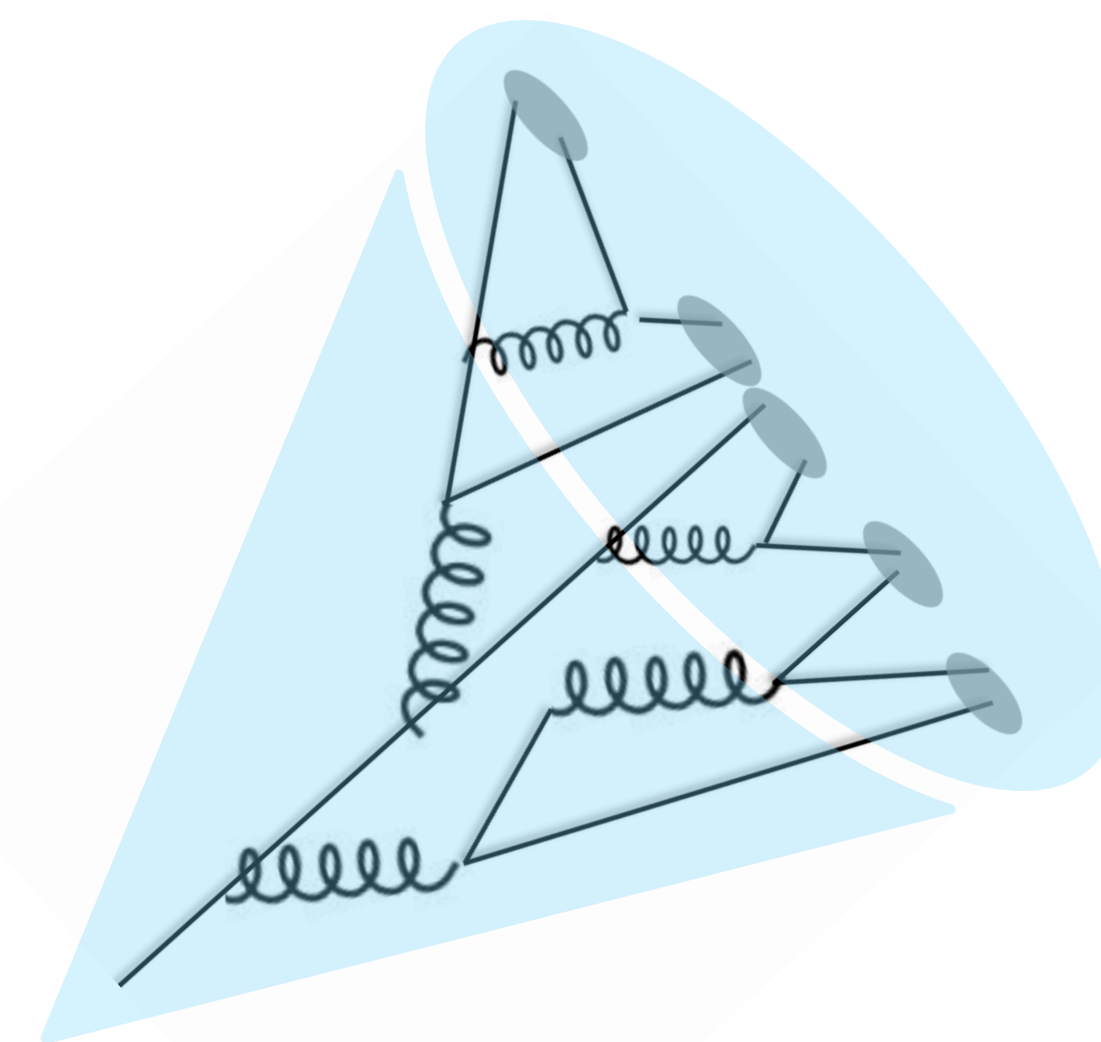
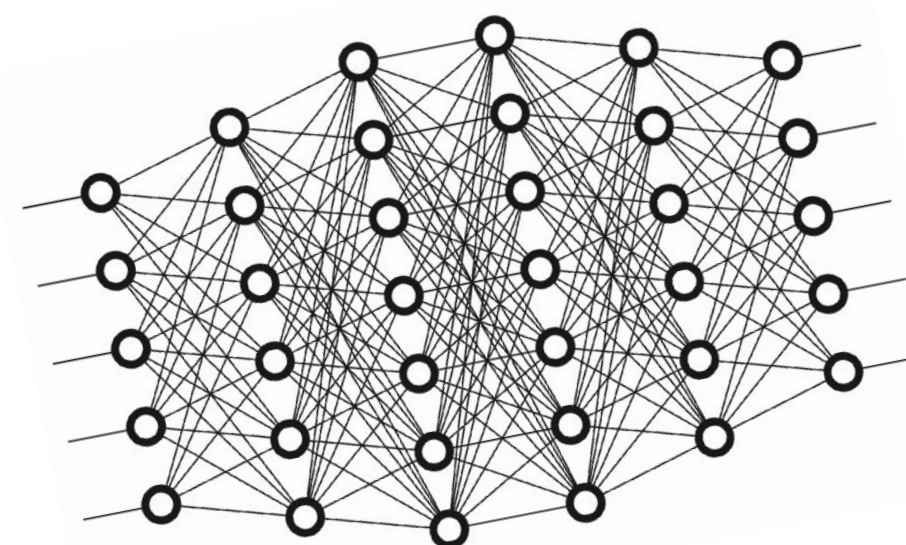




Demis Hassabis
Former CEO google DeepMind
Nobel Laureate
Interview with Cleo Abram, [video](#)

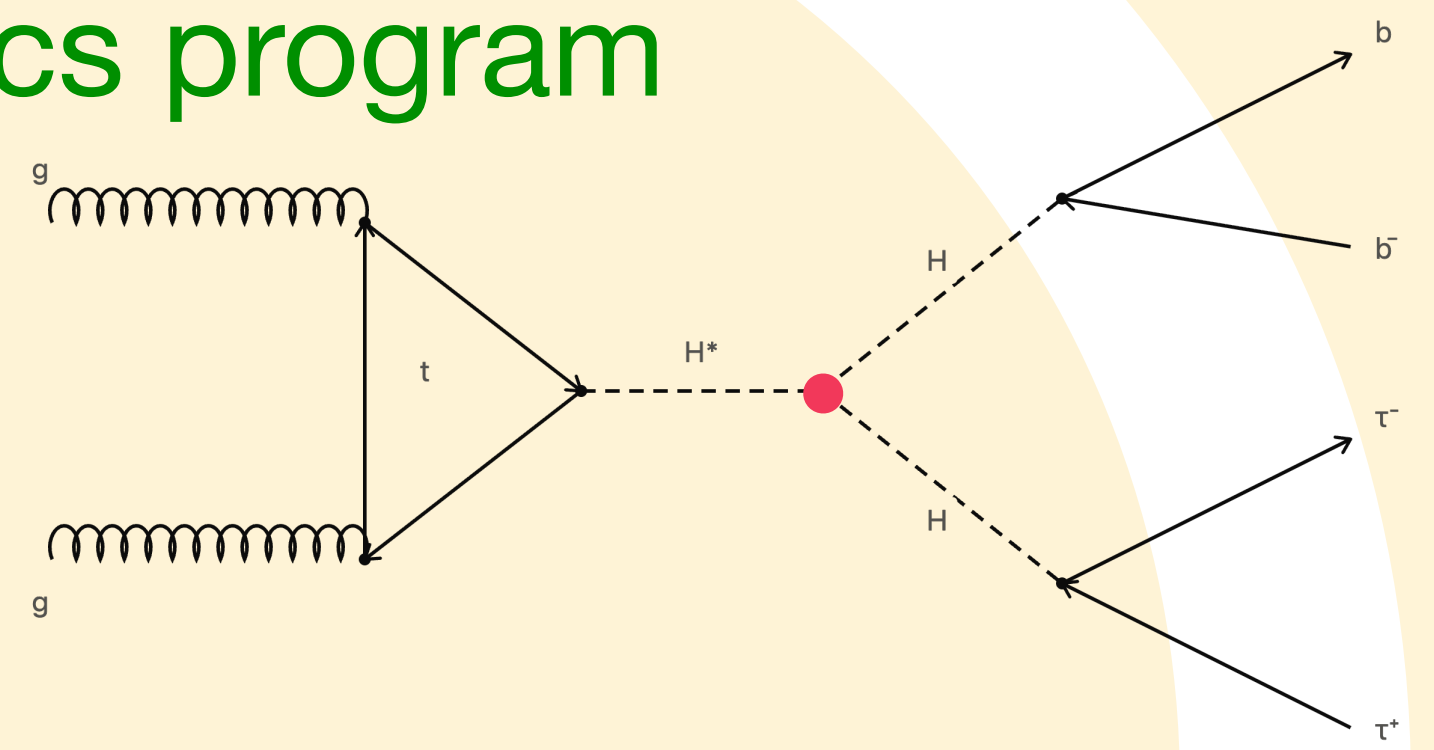
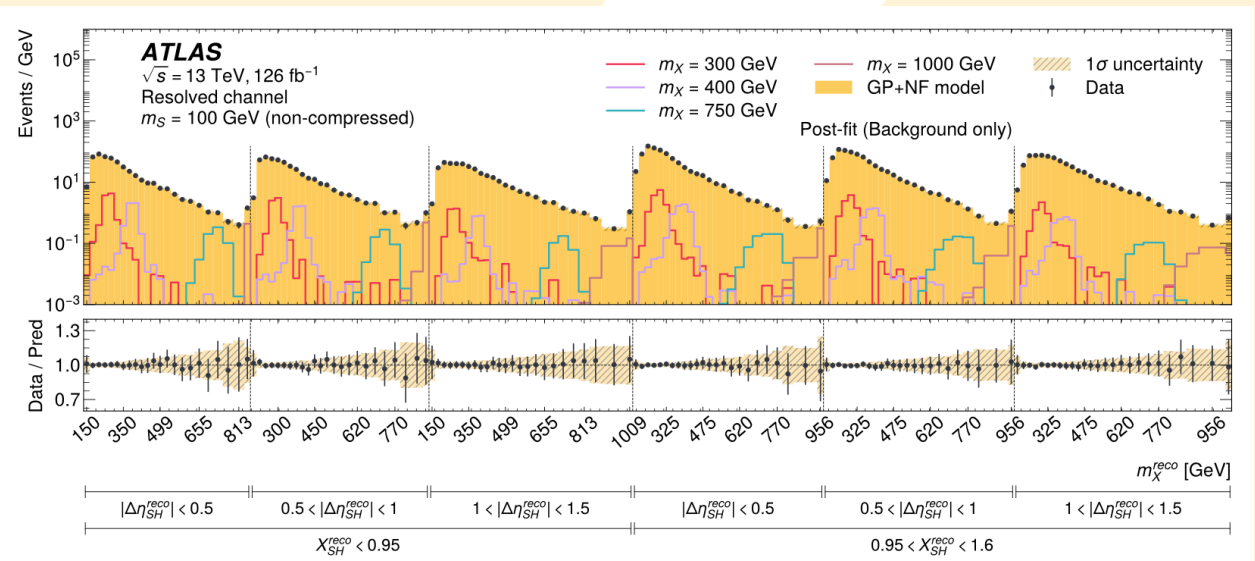
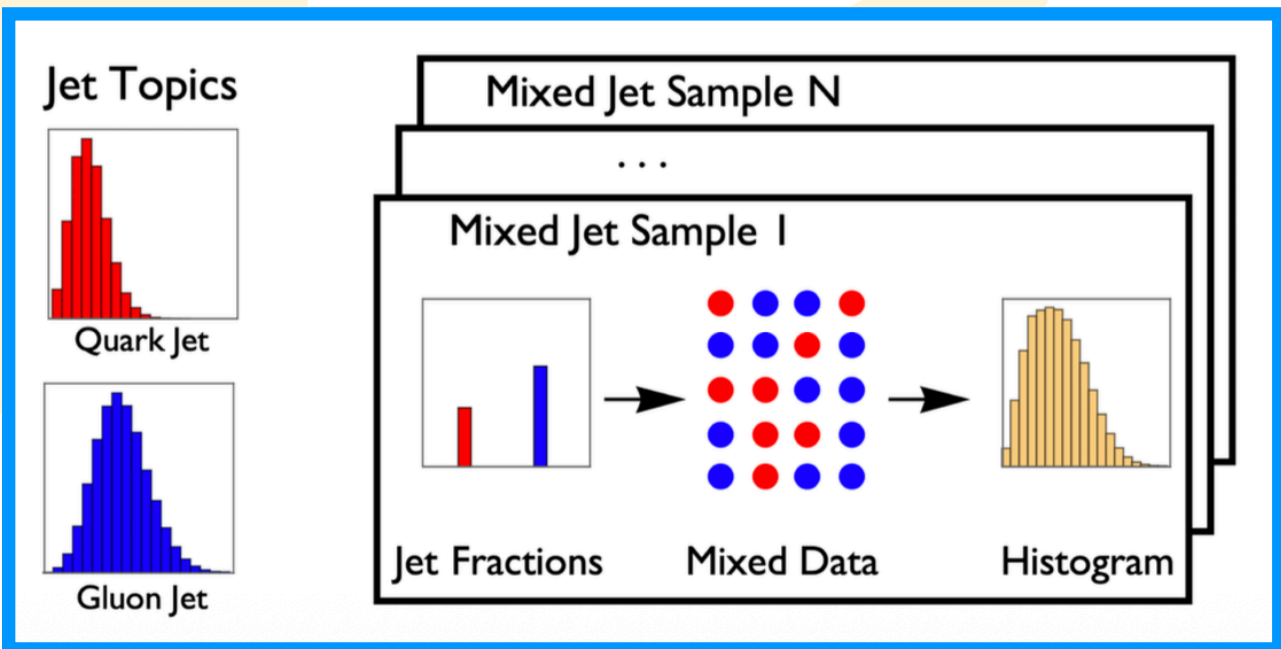
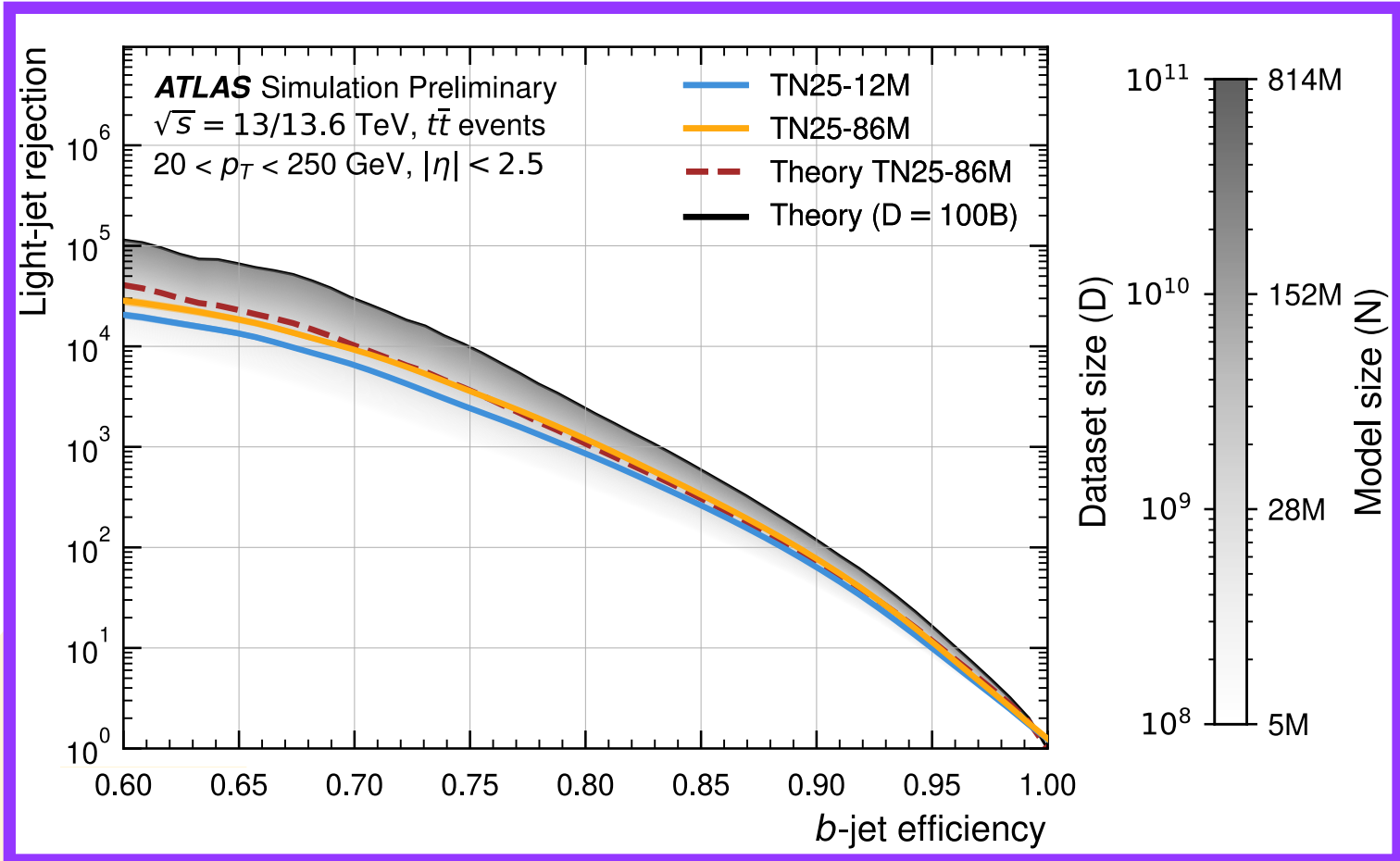
“There are a lot of deep mysteries about how the universe works... Most of the things we care about we don’t understand...”

The nature of reality — that’s ultimately what I’m after. And I want to use AI as a tool to help us understand the nature of reality around it.”

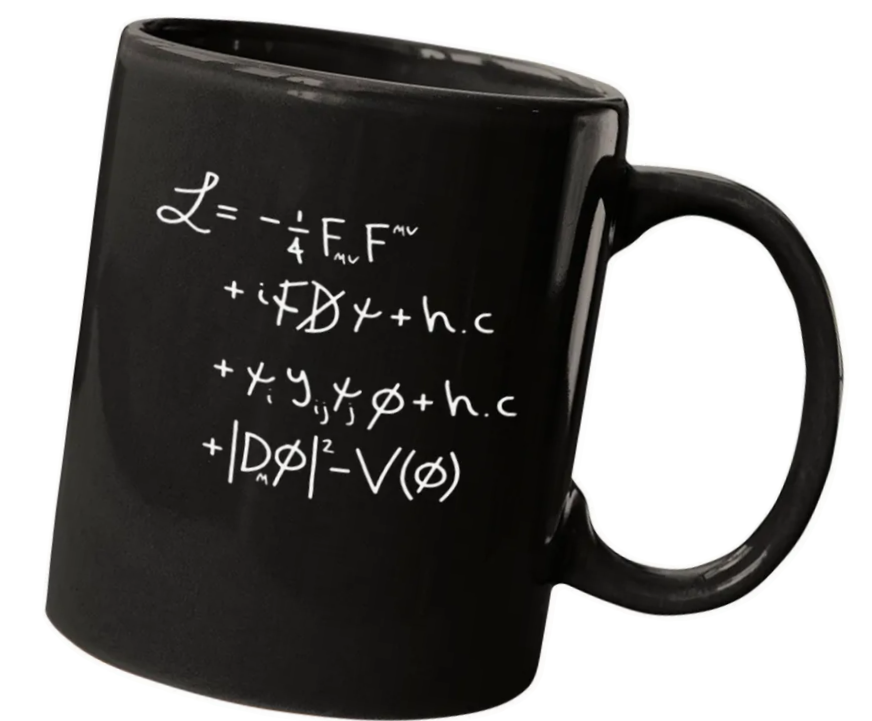


ML4Jets: a new era for LHC reconstruction+analysis

- 1/ Building jets: inputs
- 2/ Jet physics models benefit from scale
- 3/ ML in calibrations critical for performance
- 4/ ML now a vital tool for our physics program

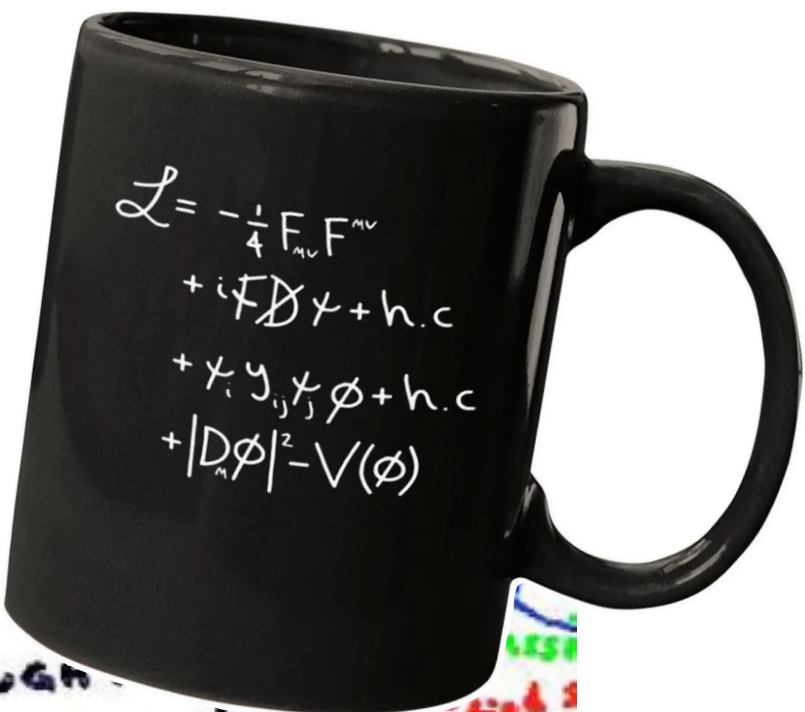


Backup

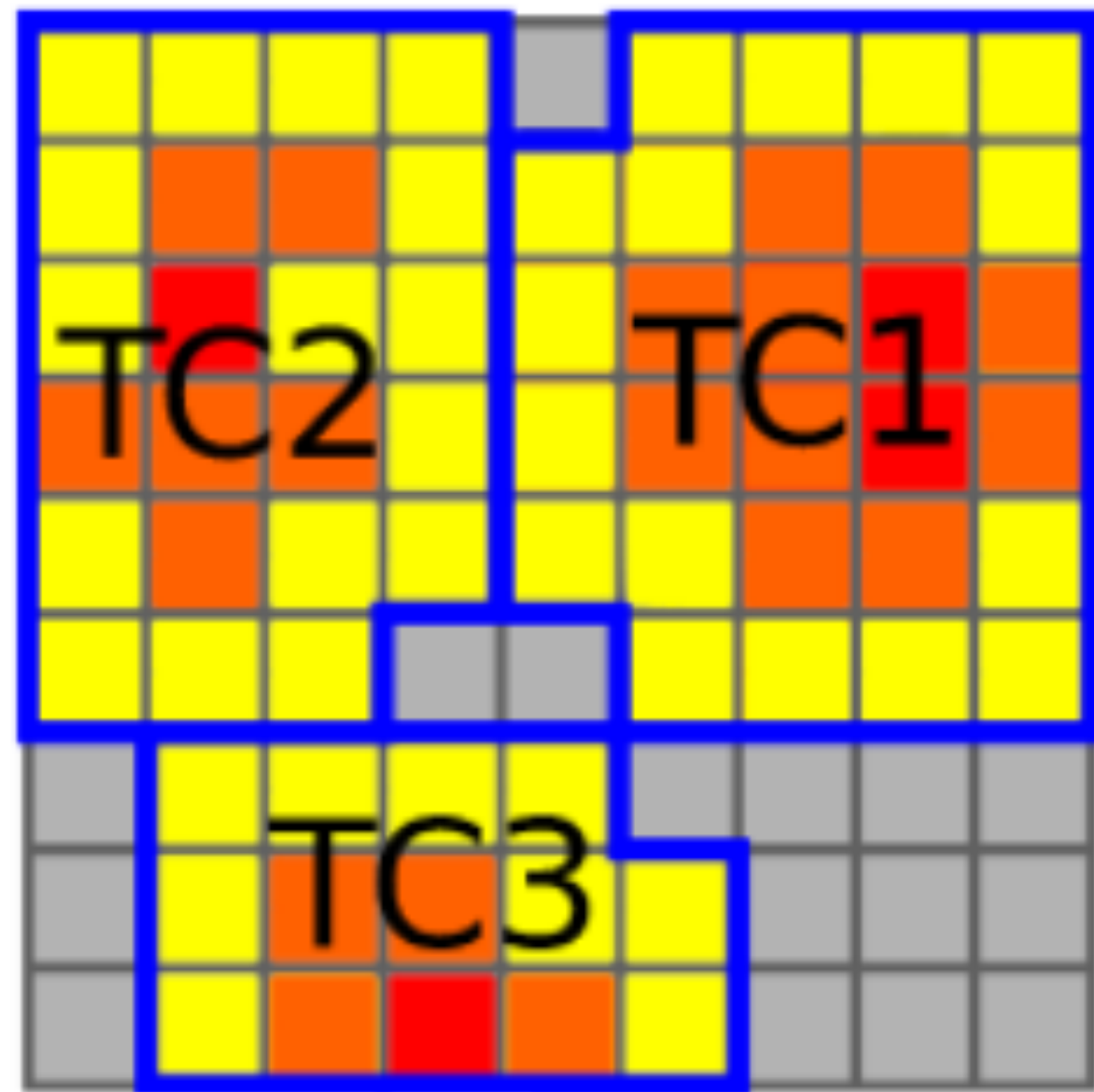


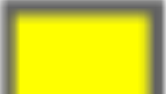
Backup

What might lie beyond?



Topoclustering



-  seed cells $|E| > 4\sigma$
-  growth cells $|E| > 2\sigma$
-  boundary cells
-  final topoclusters
 $\eta \times \phi = \text{dynamic}$

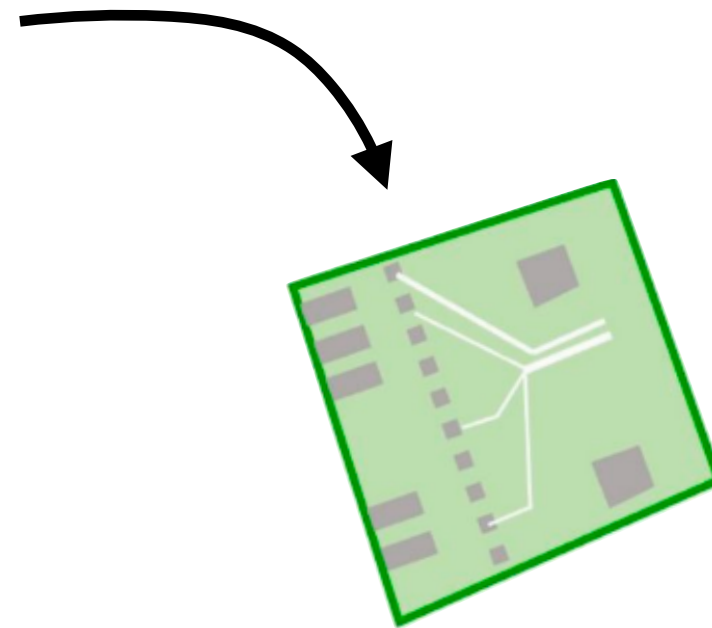
Data scouting... at L1

Demonstrated in Run 3: AXOL1TL / CICADA / GELATO

At HL-LHC CMS will have tracking at L1

↪ Jet tagging 🎉

SOTA transformer too
large to fit on an FPGA....

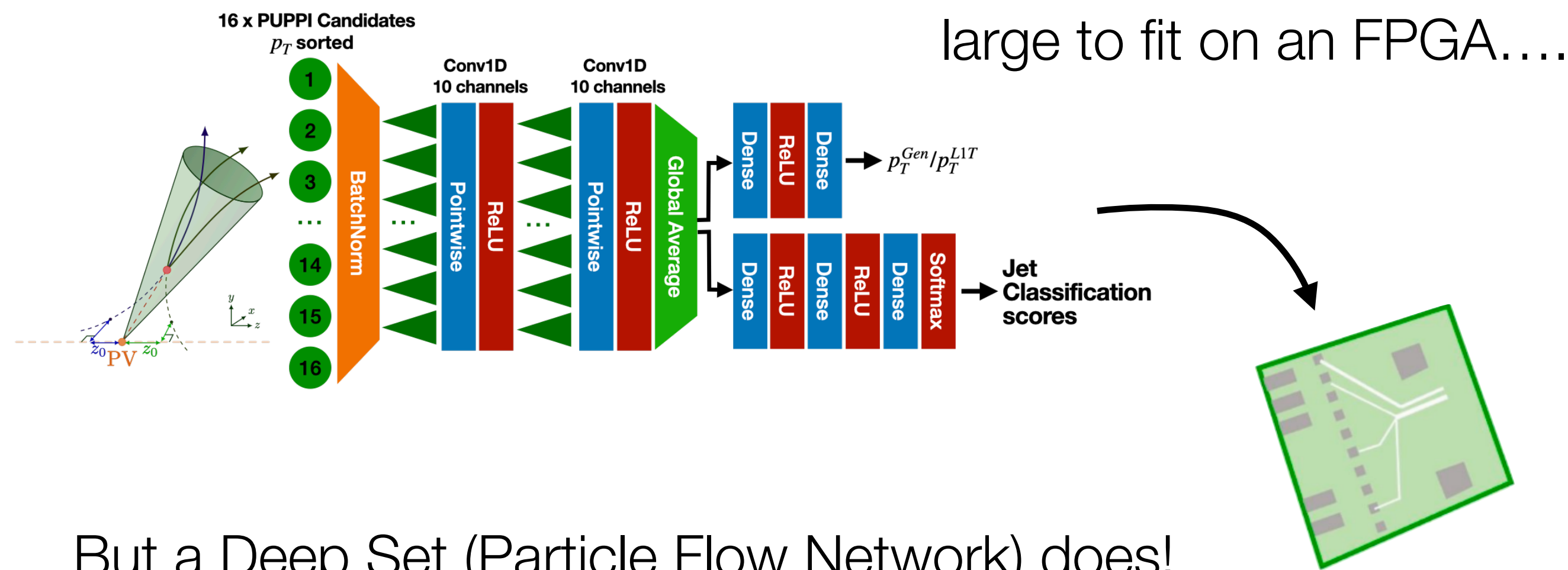


Data scouting... at L1

Demonstrated in Run 3: [AXOL1TL](#) / [CICADA](#) / [GELATO](#)

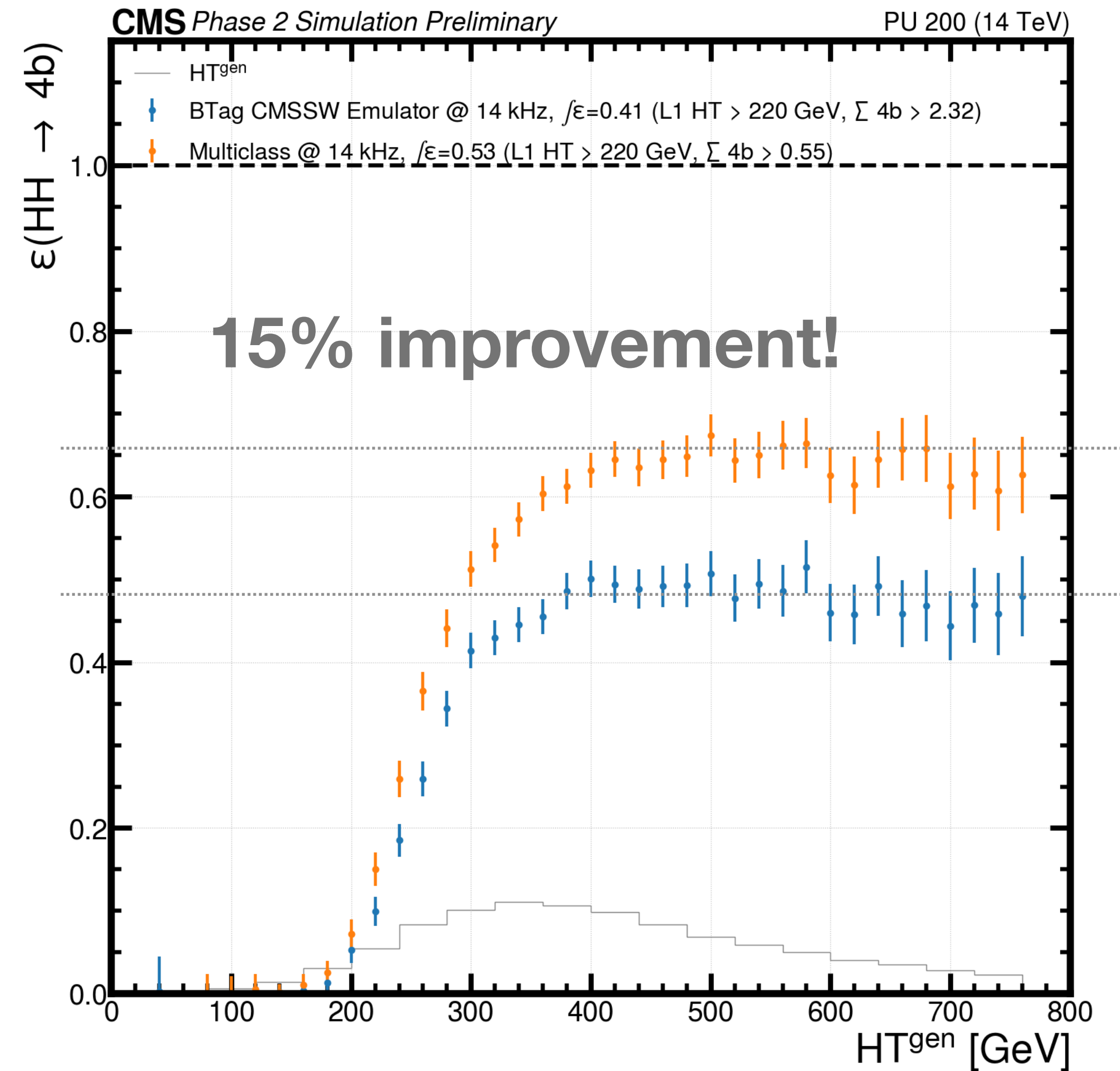
At HL-LHC CMS will have tracking at L1

↪ Jet tagging 🎉



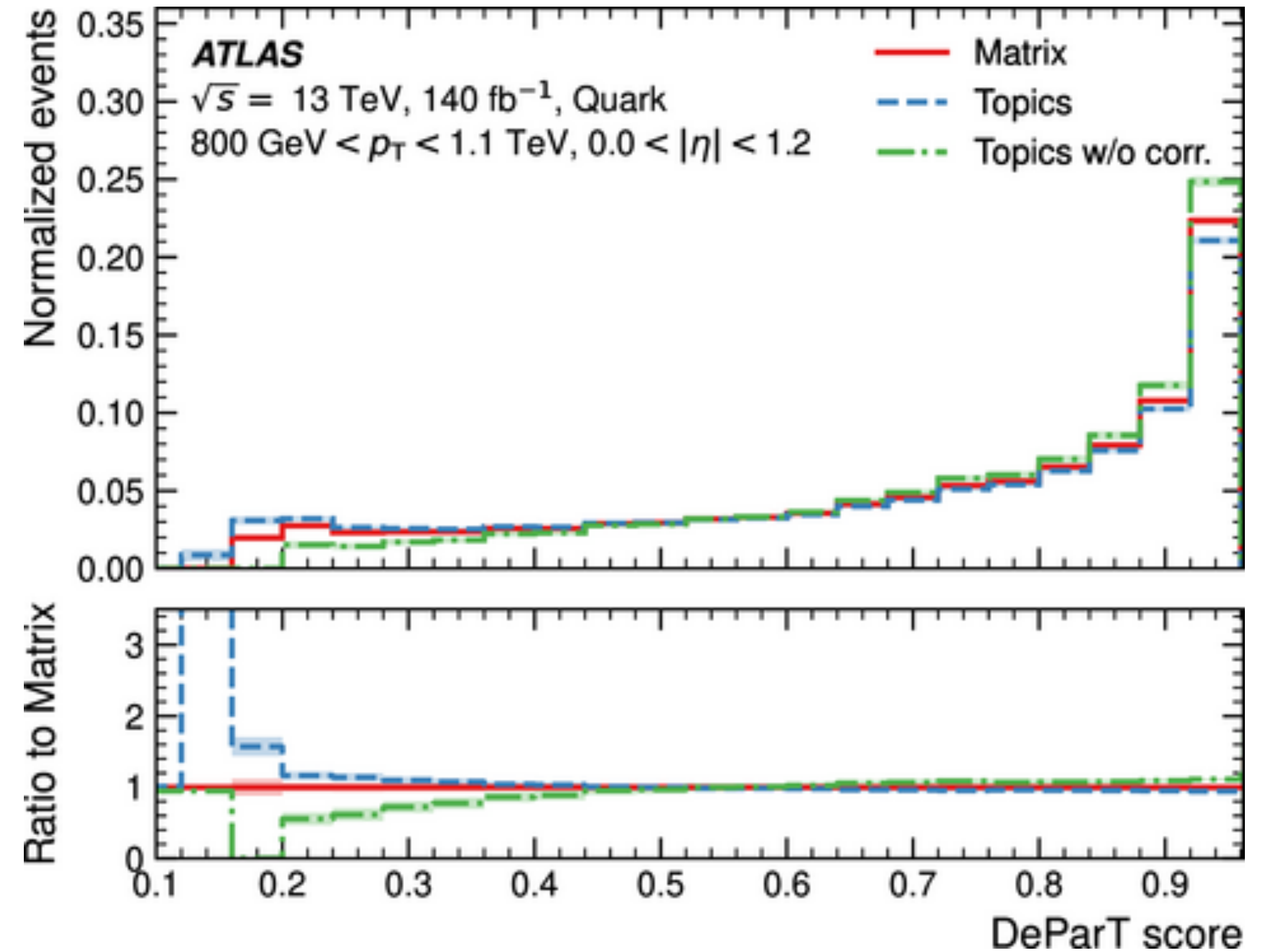
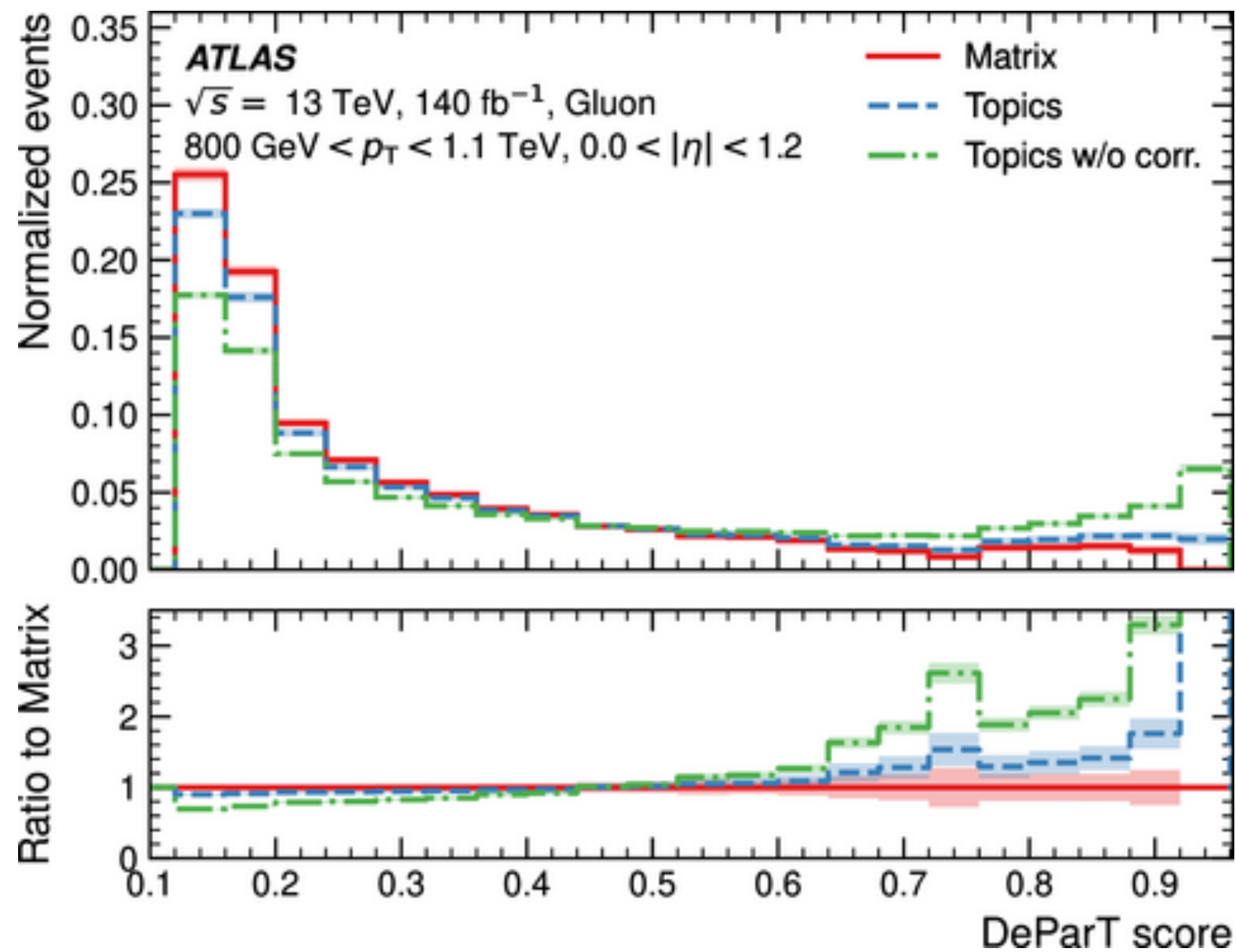
But a Deep Set (Particle Flow Network) does!

- Introduced Patrick Komiske, [BOOST 2018](#)

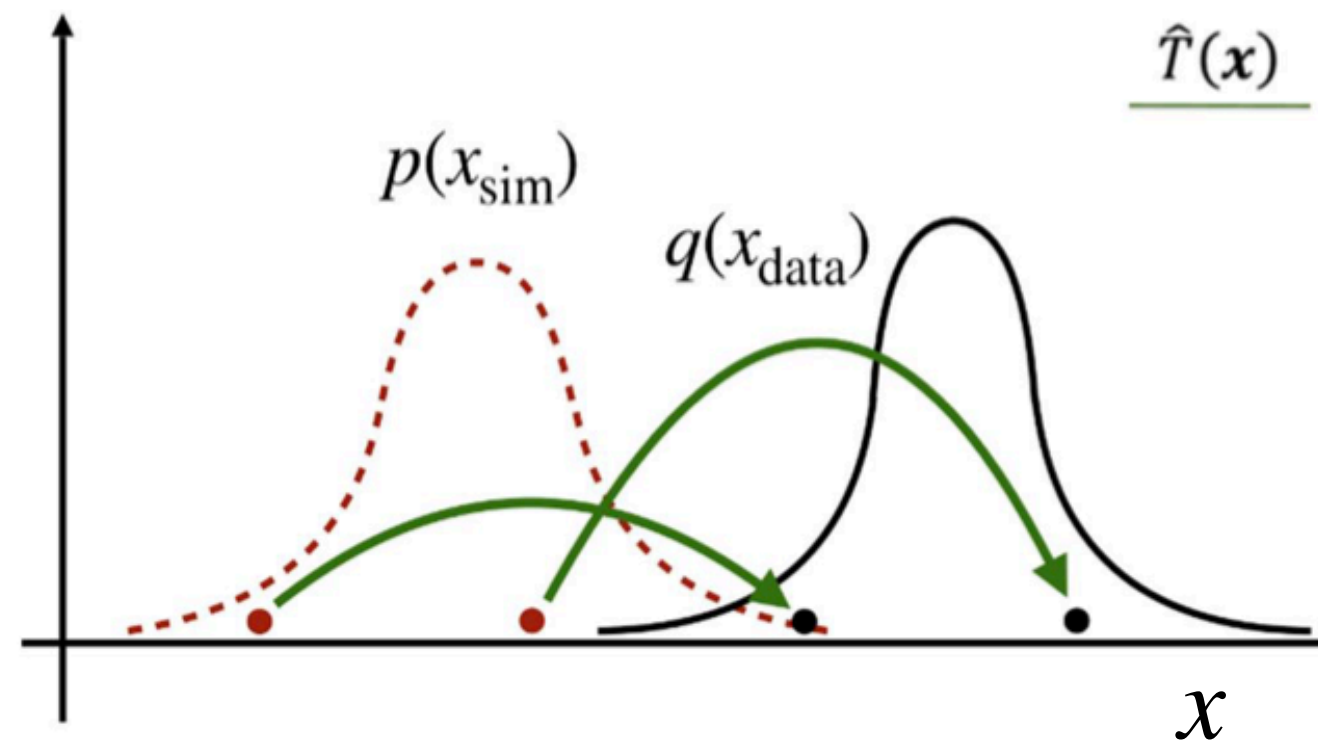


Jet topics: q/g calibration

Extracted probability distributions



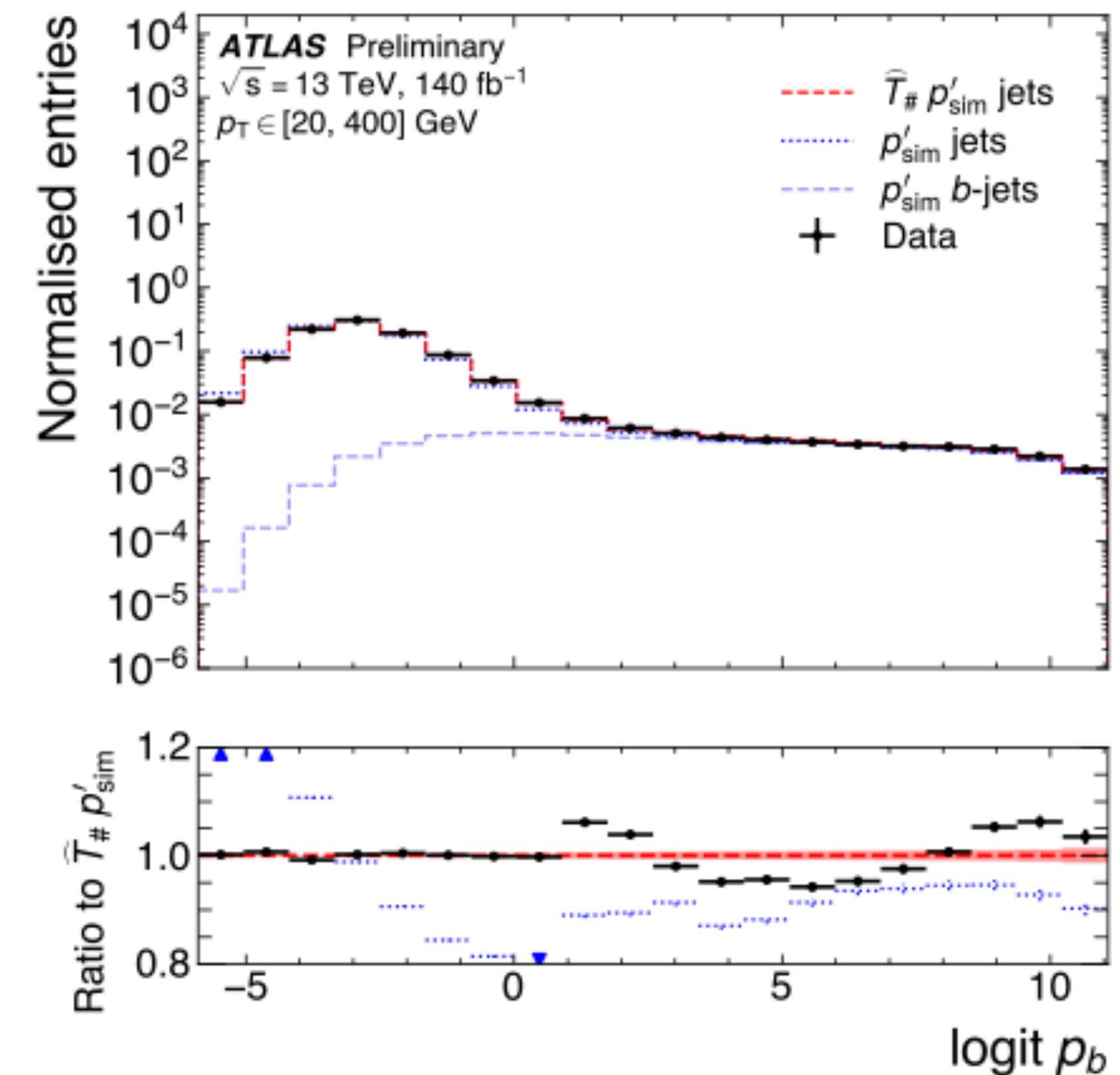
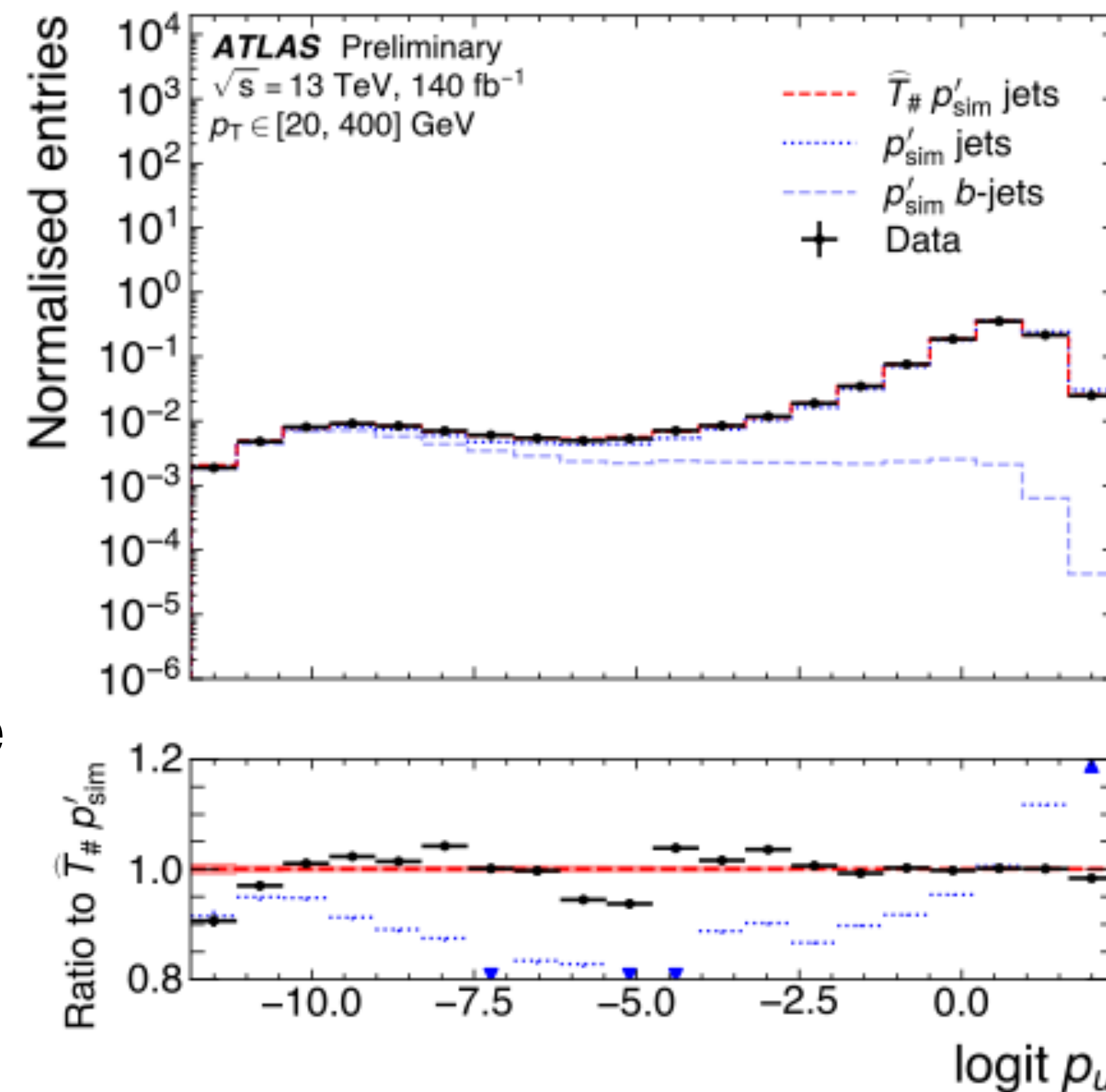
Optimal transport for dat $\rightarrow$ MC corrections



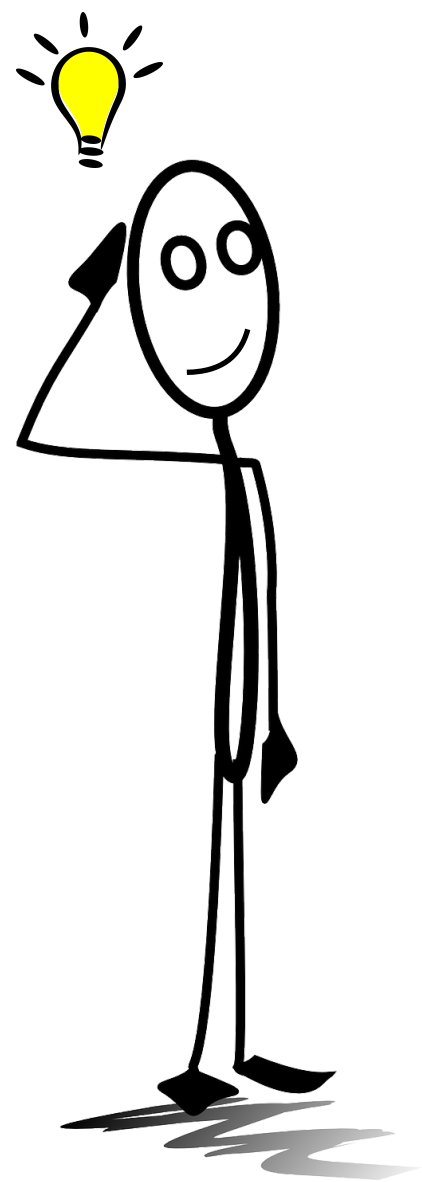
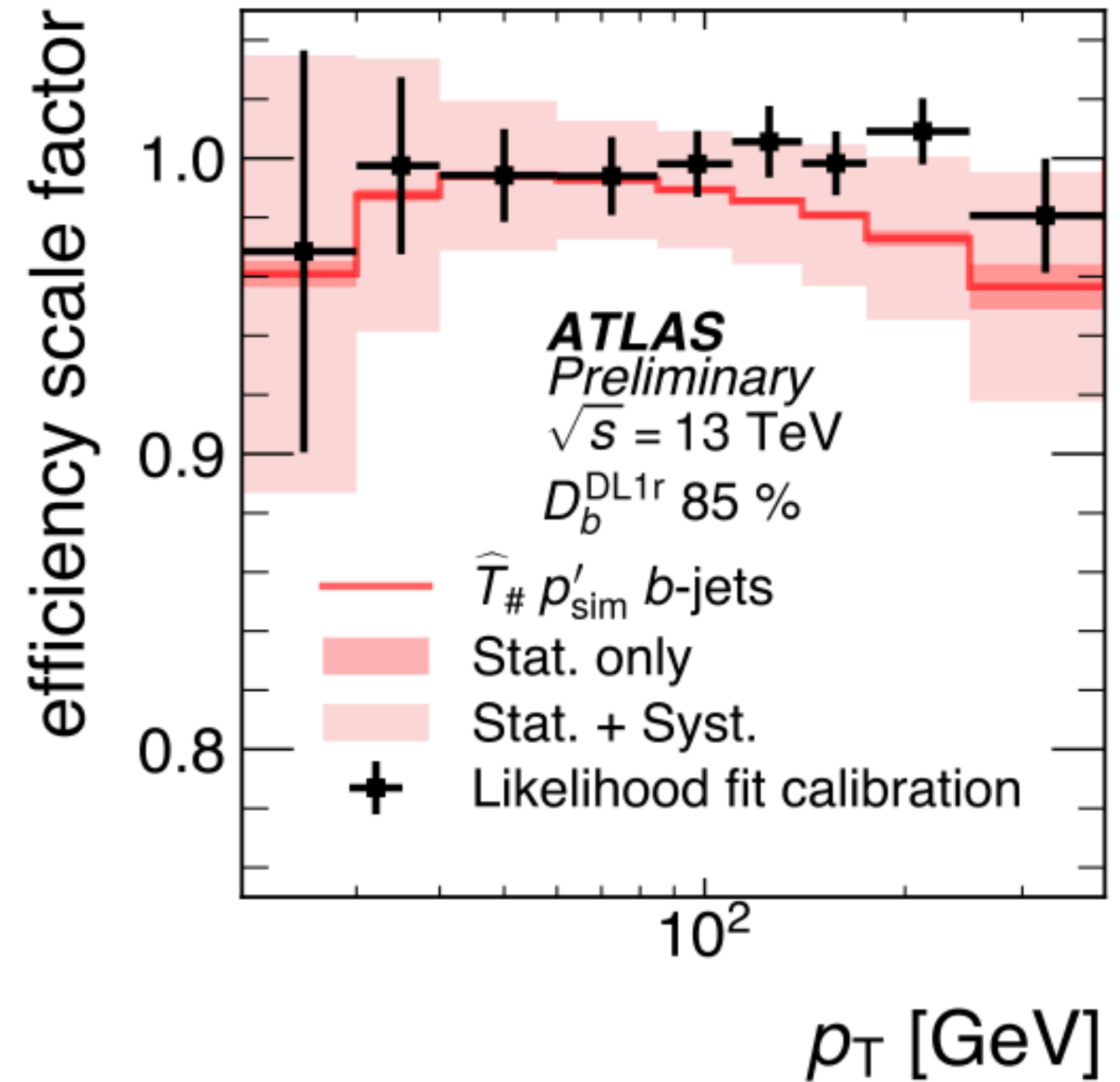
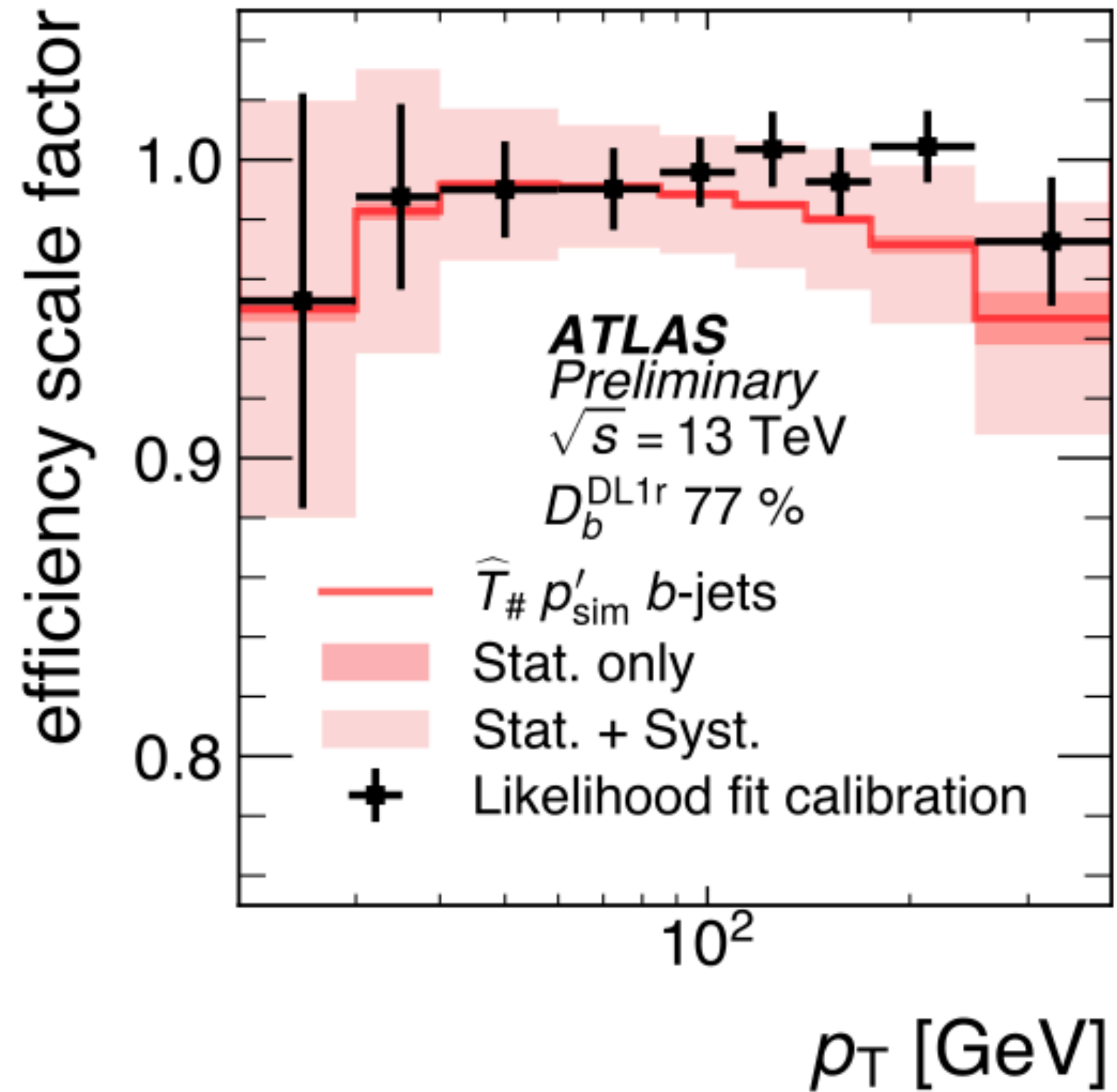
Can we use ML to map from $p(x_{\text{sim}}) \rightarrow q(x_{\text{data}})$?

Architecture: Normalizing flow

Constraint to enforce this is the optimal transport mapping.

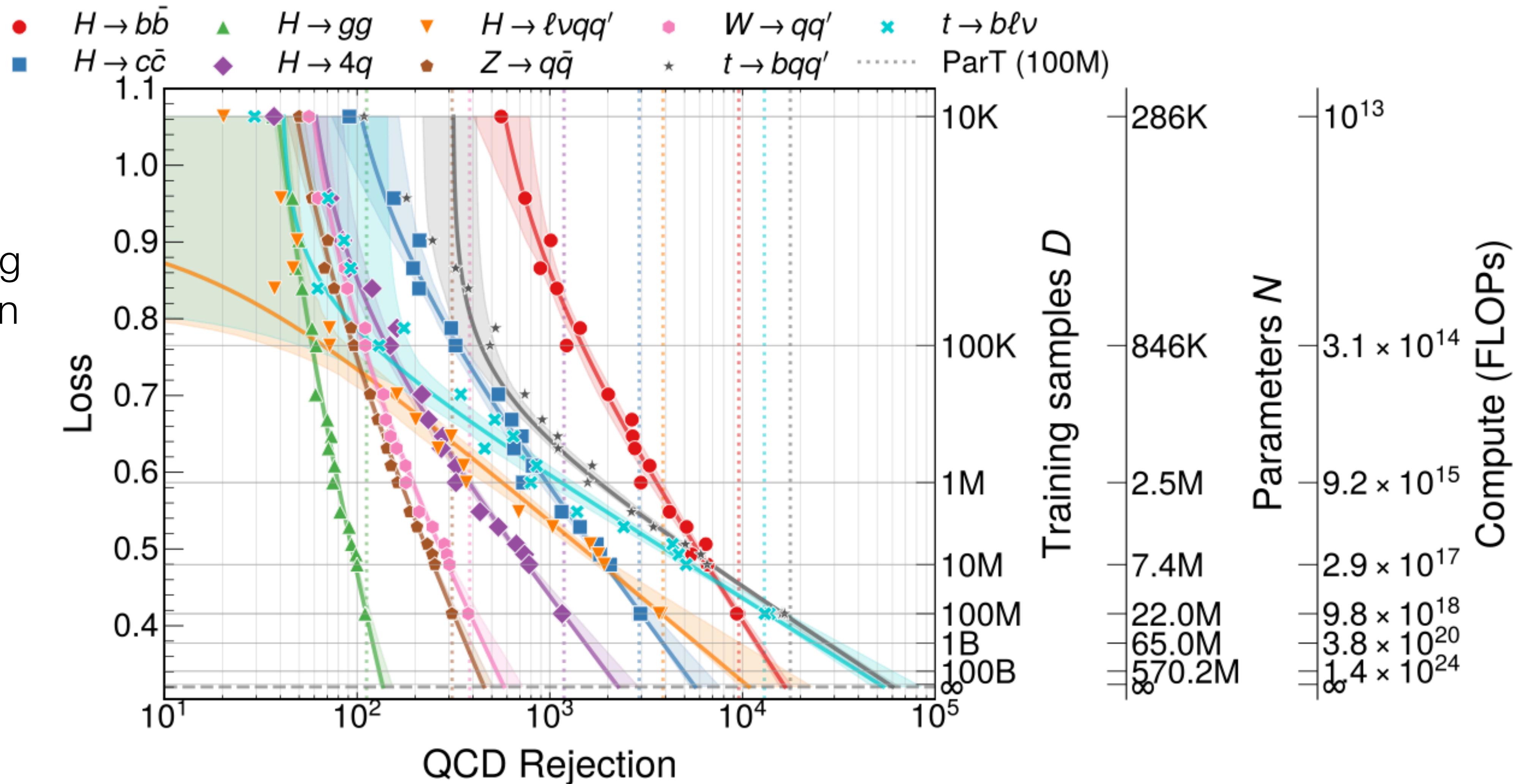


It works!! (Reproduces standard calibration)



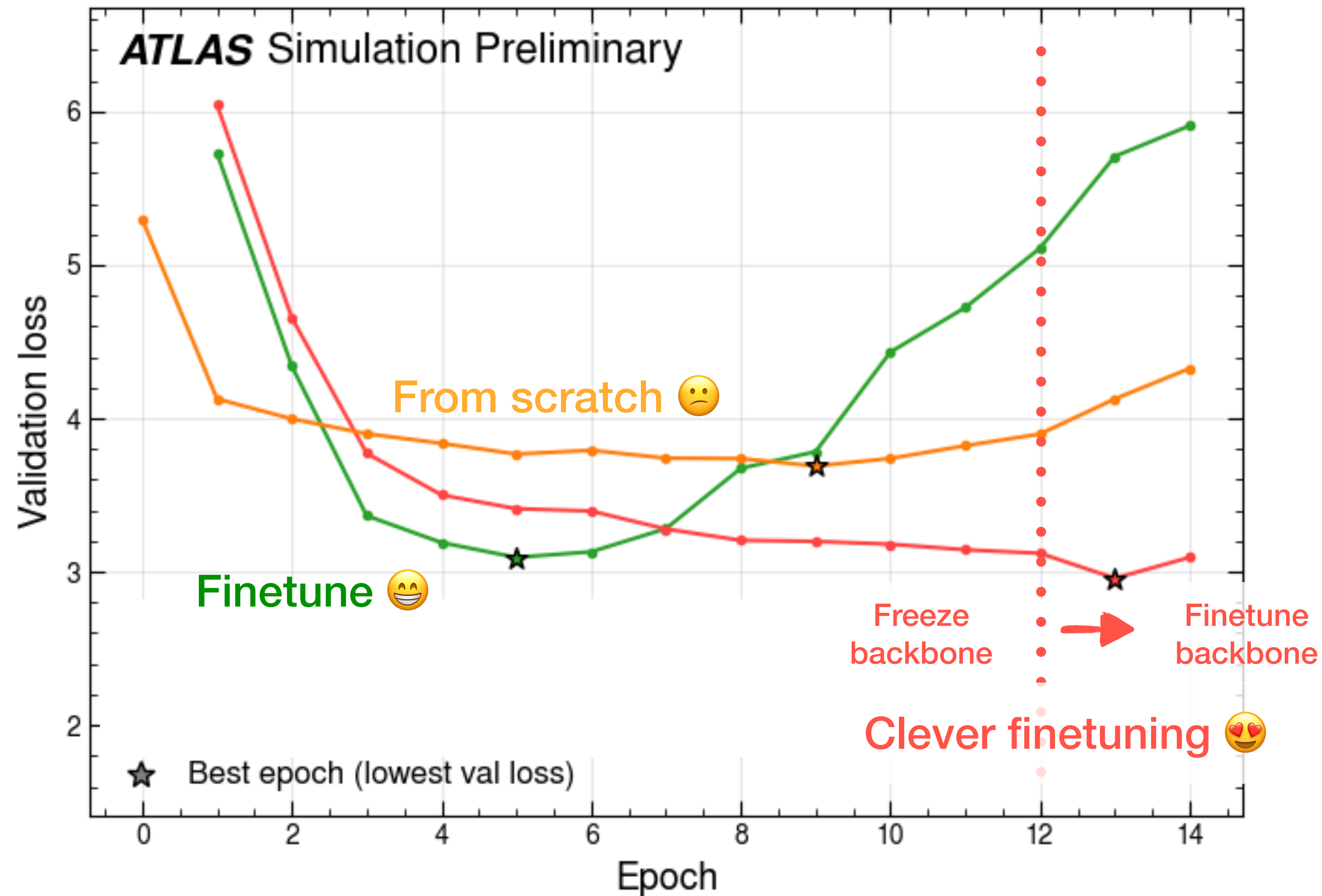
Performance of ParT on JetClass

One JetClass
(Delphes) we're
close to saturating
what we can learn
(w/ classification
metric)





When to finetune

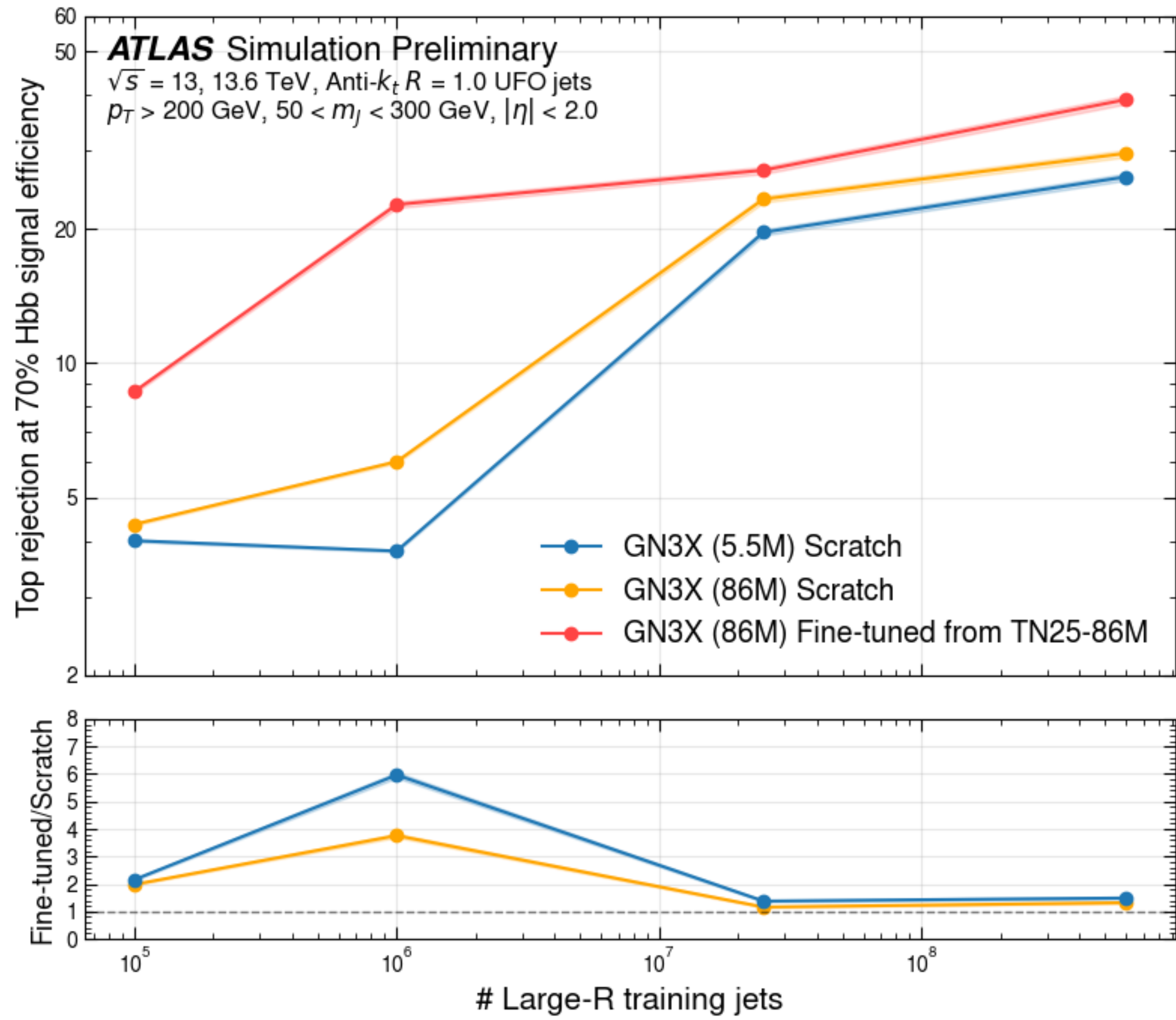
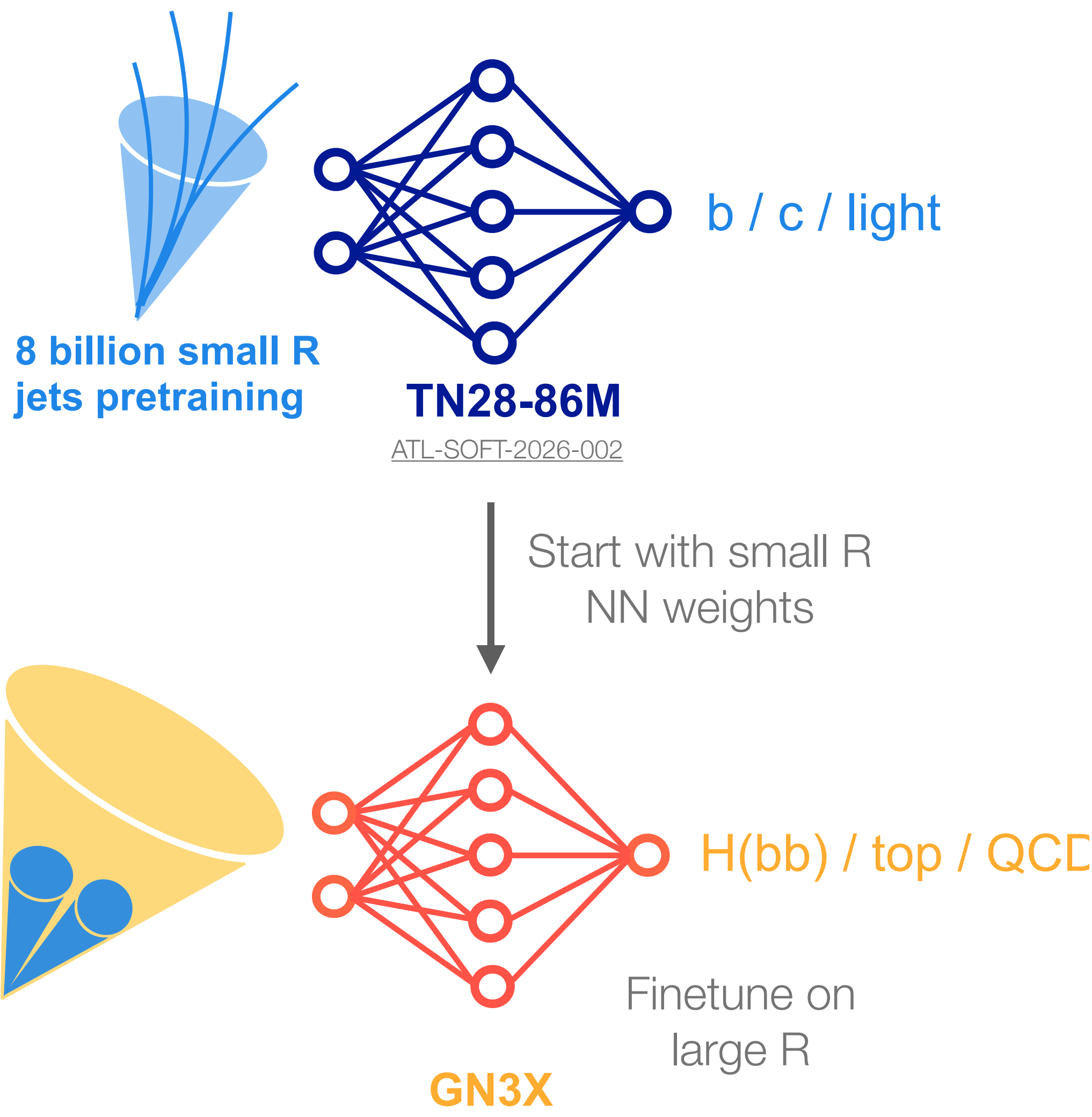


Ex: Train / finetune on 1M large R jets

Impact: improved rejection

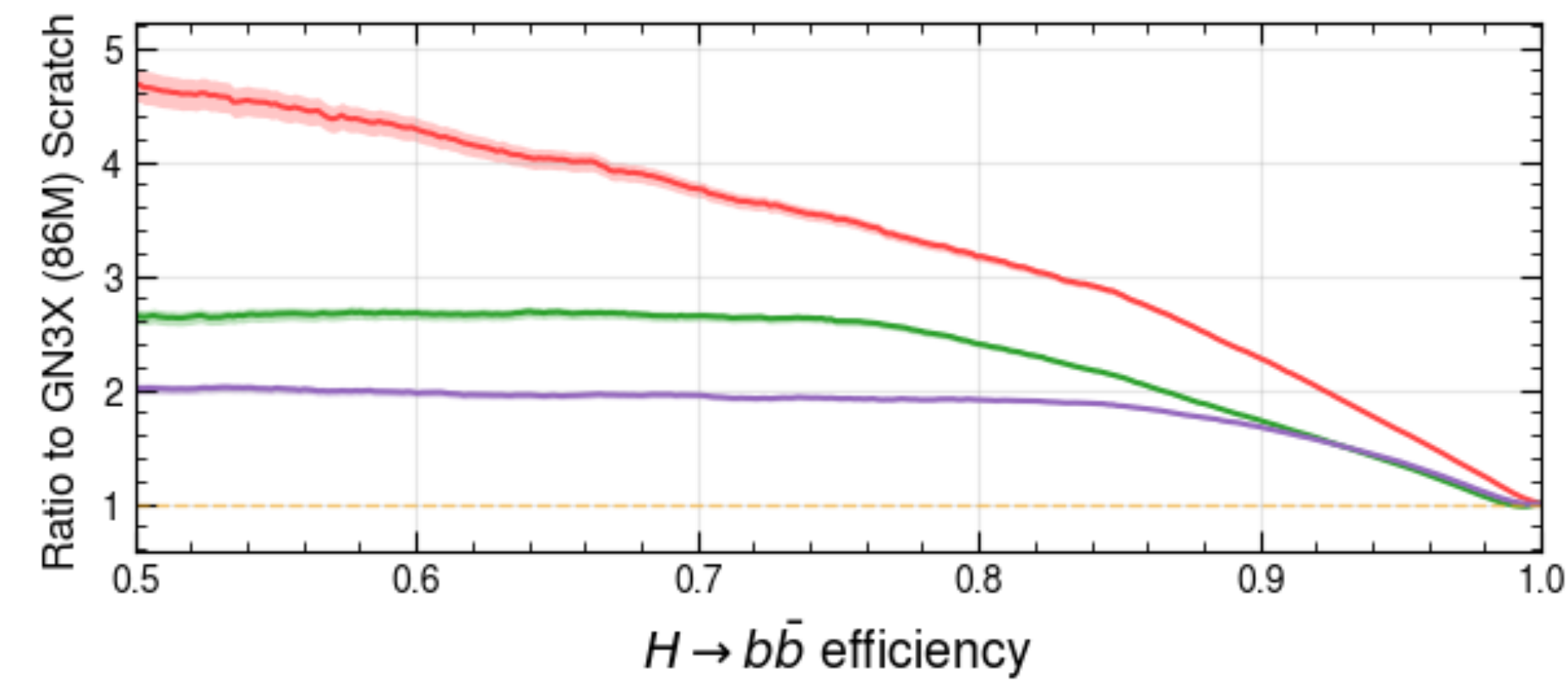
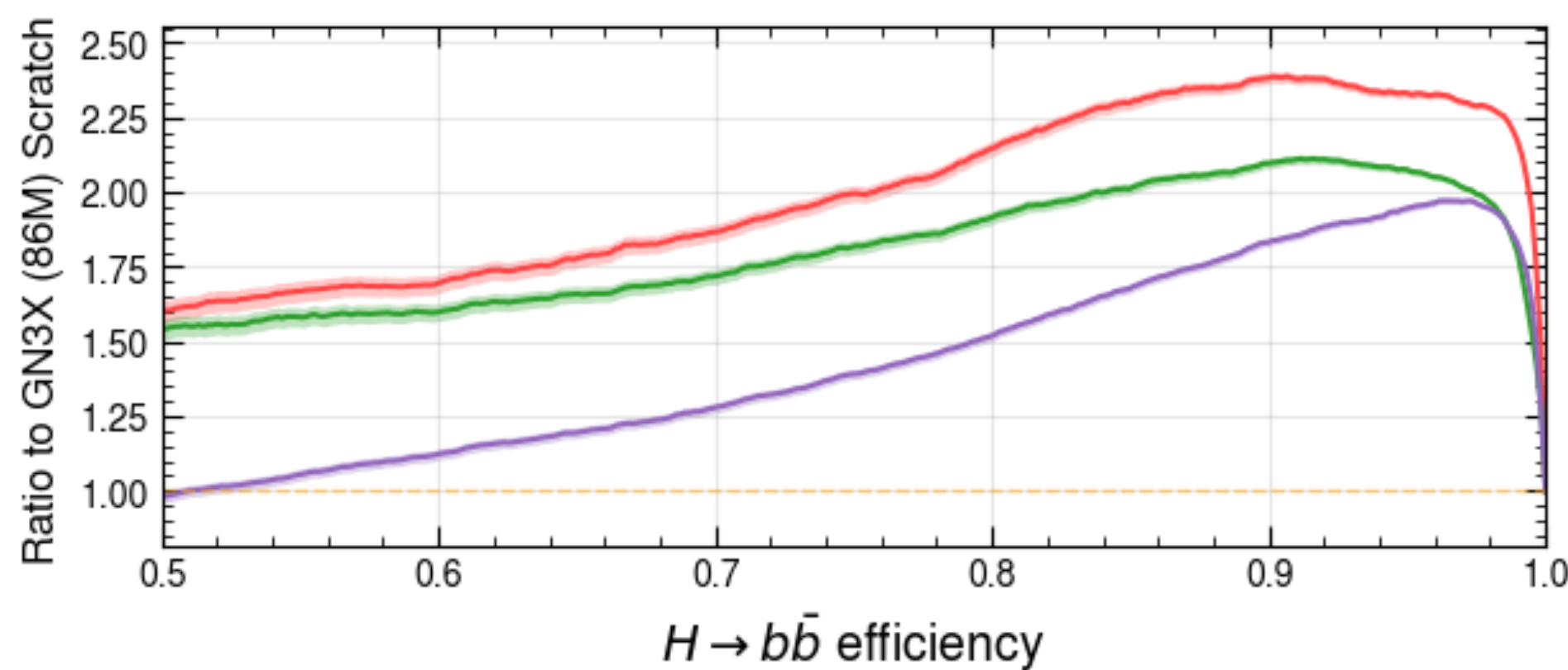
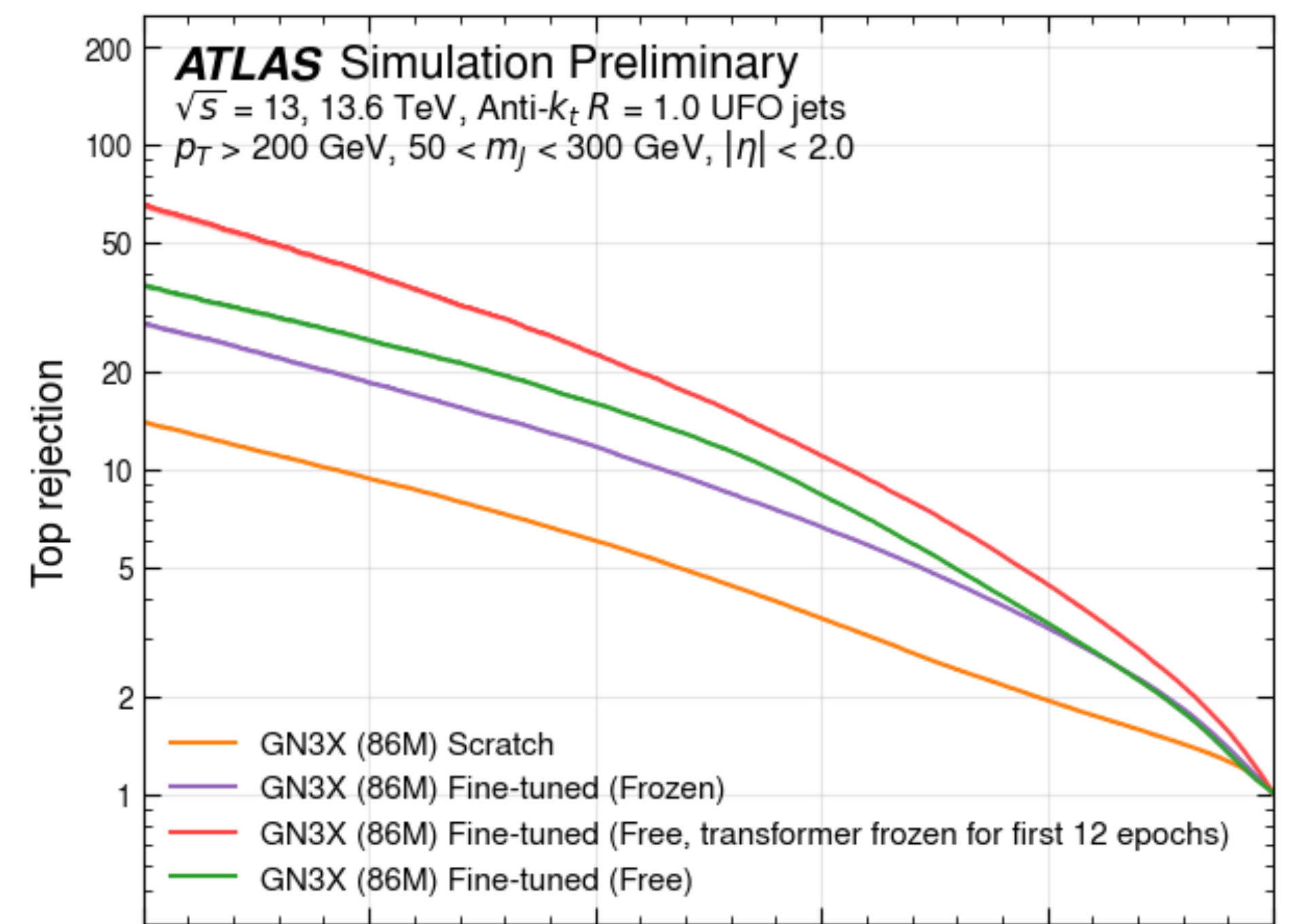
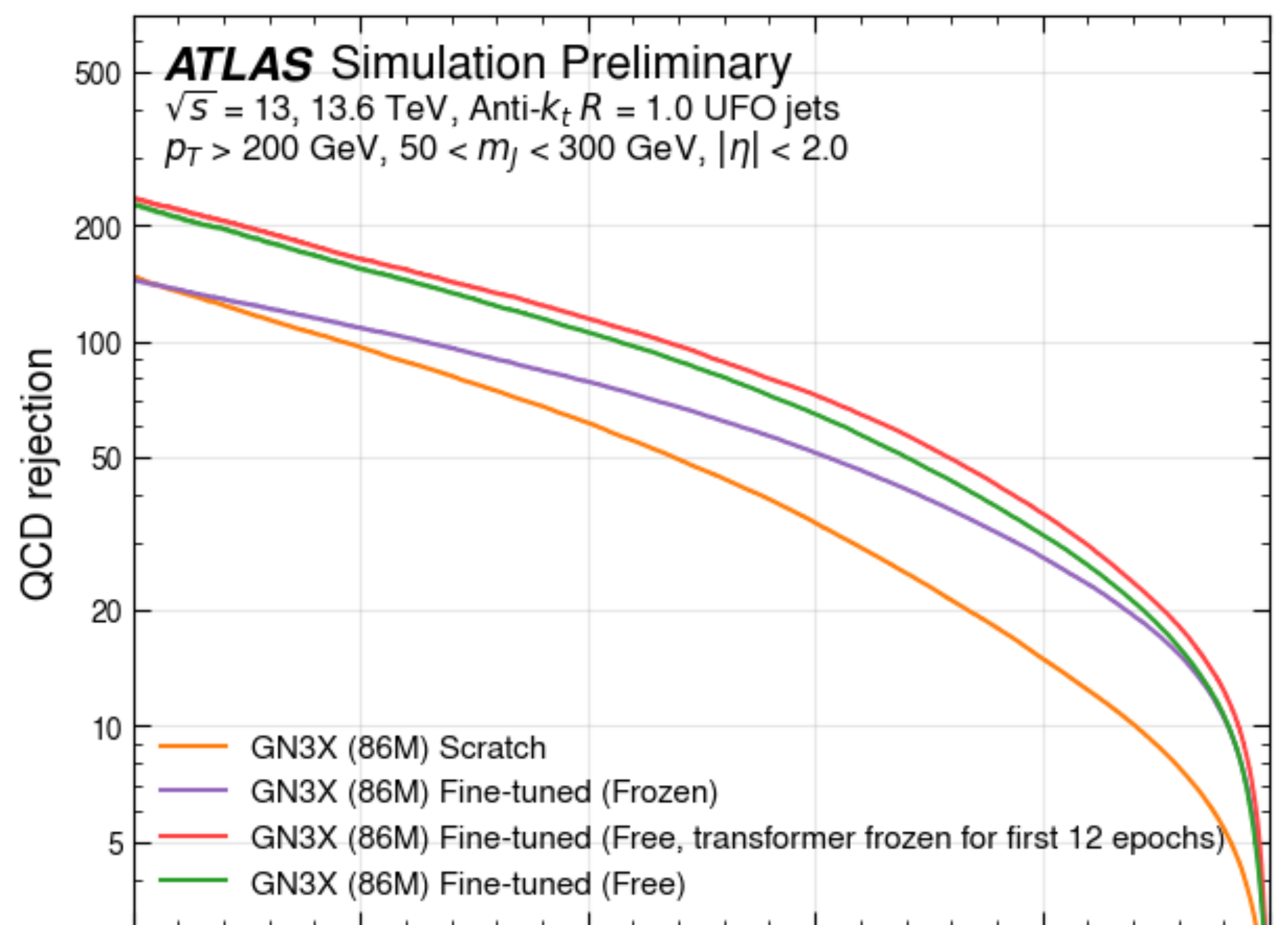


How far can we transfer?



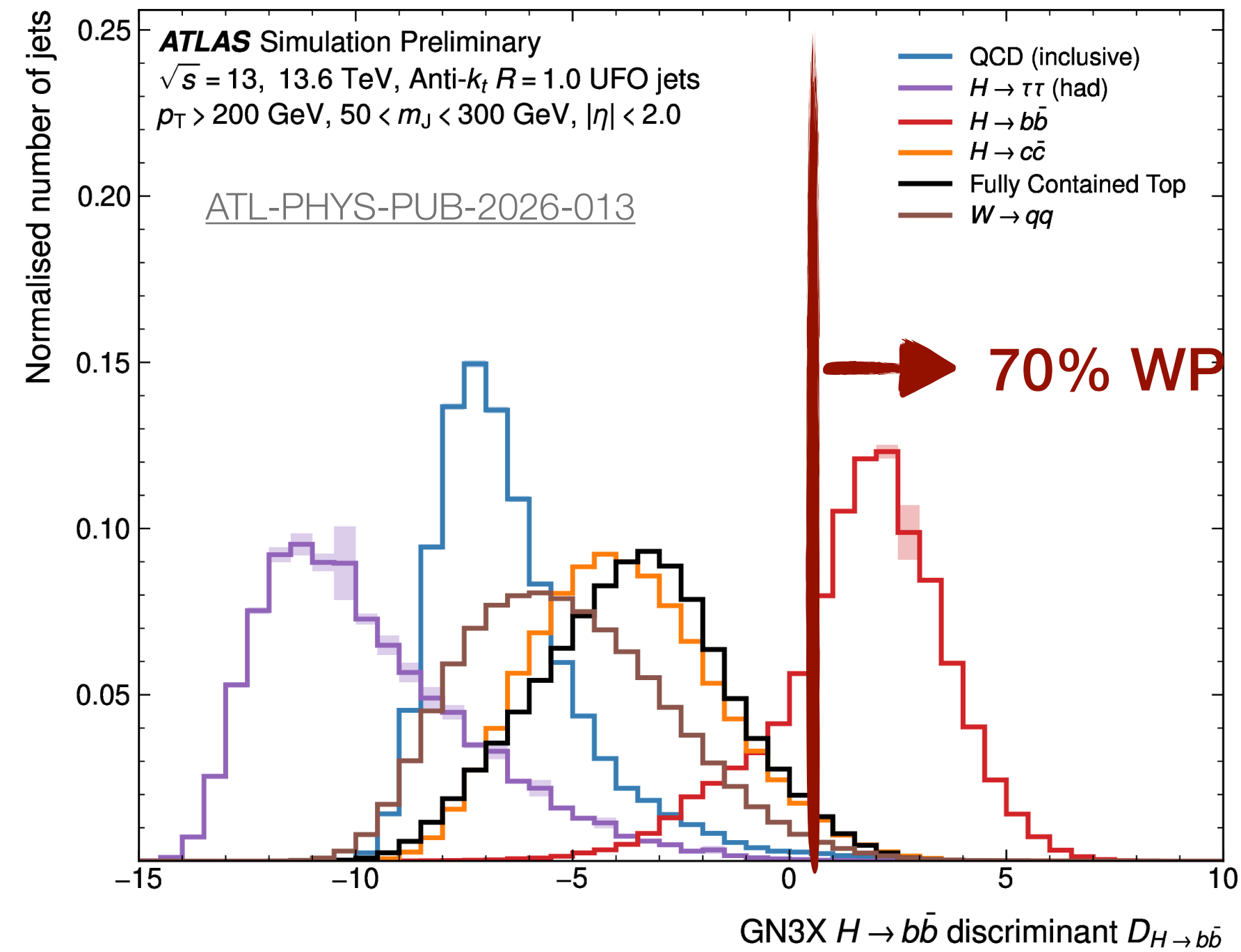


How to finetune

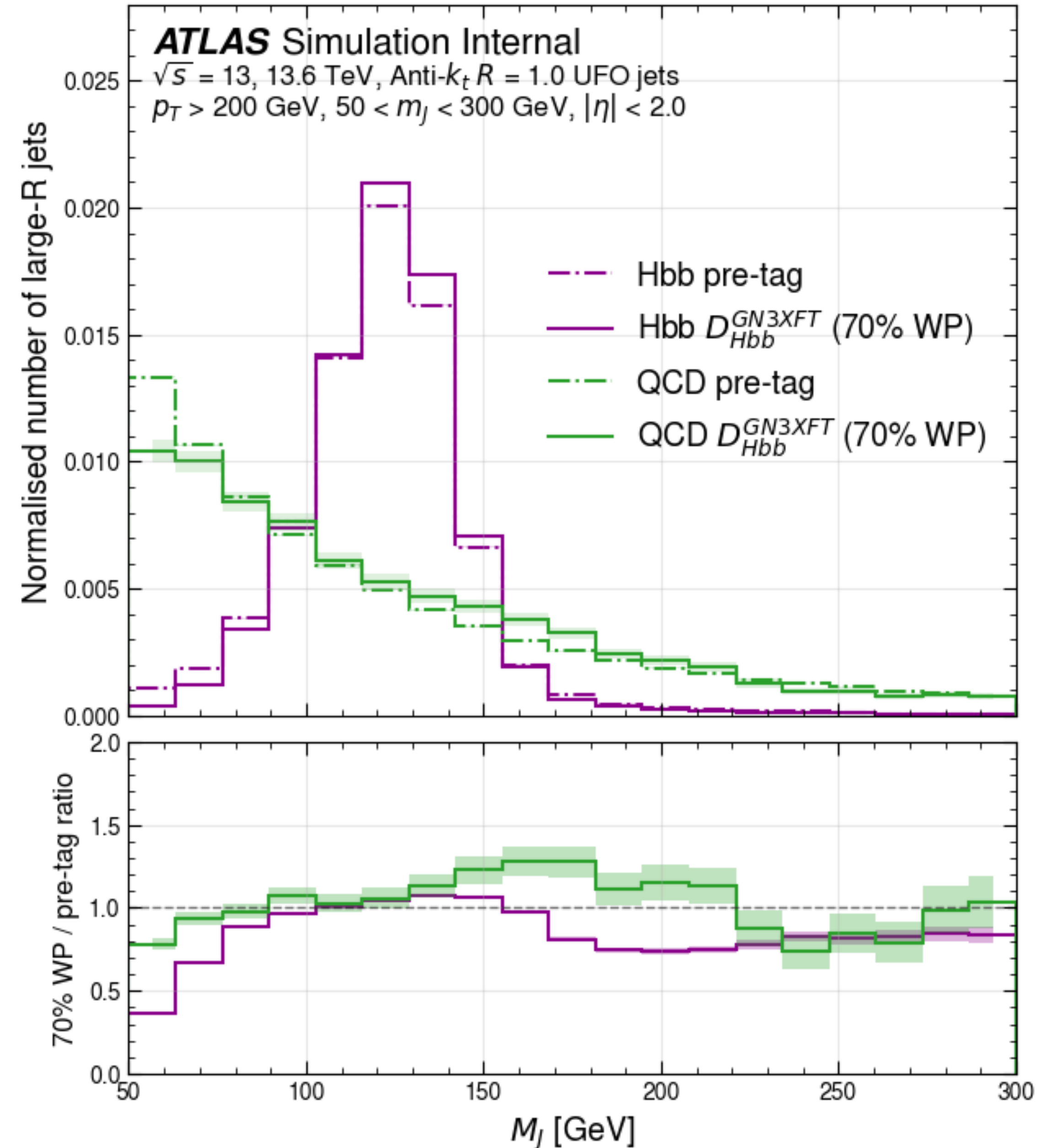


Train/finetune
1M large R jets

Mass sculpting (or lack thereof)



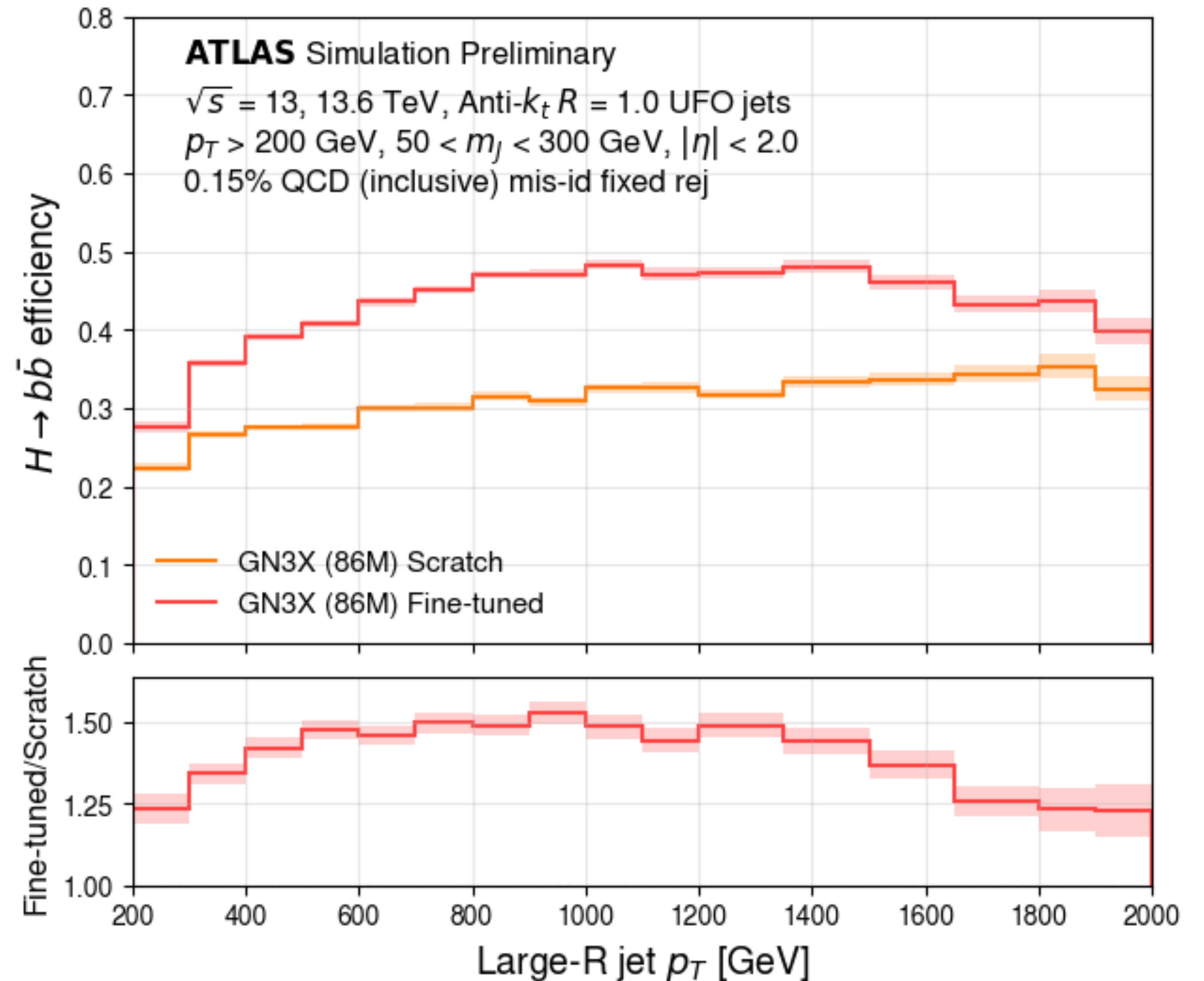
Sanity check: no significant mass sculpting after finetuning.



Performance vs pT

Train/finetune
1M large R jets

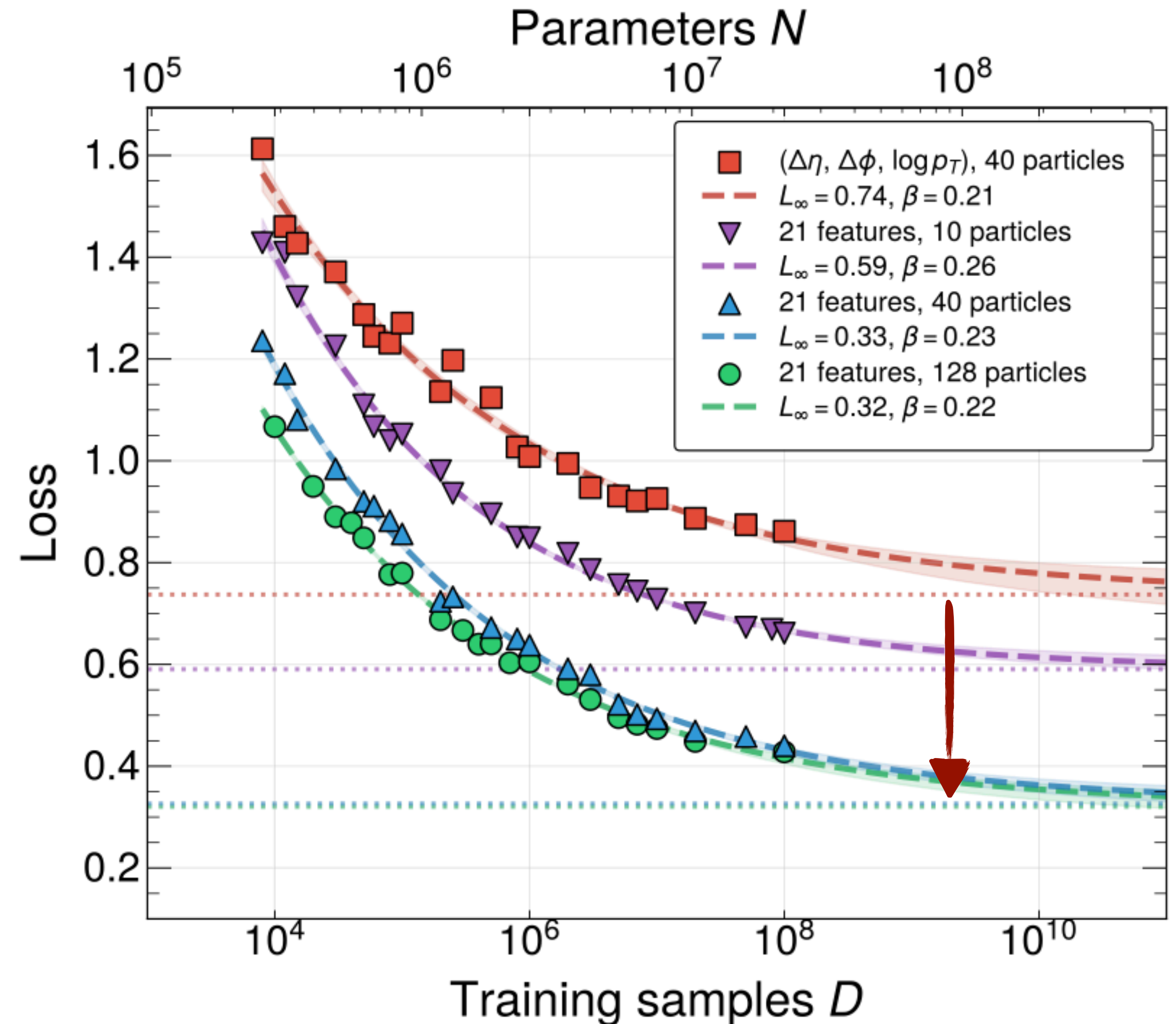
Performance gains persists
across large R jet pT spectrum



How else to improve? MORE INPUTS

Increasing the number of input features LOWERS the $\mathcal{L}_\infty$ floor

Dictates a research program we should follow for the next few years
 → Use the highest-dim inputs!



Flows for high dimensional interpolation

Hierarchical model: $p(x, m_{H1}, m_{H2}) =$ $p(x | m_{H1}, m_{H2})$ $p(m_{H1}, m_{H2})$

Normalizing flow

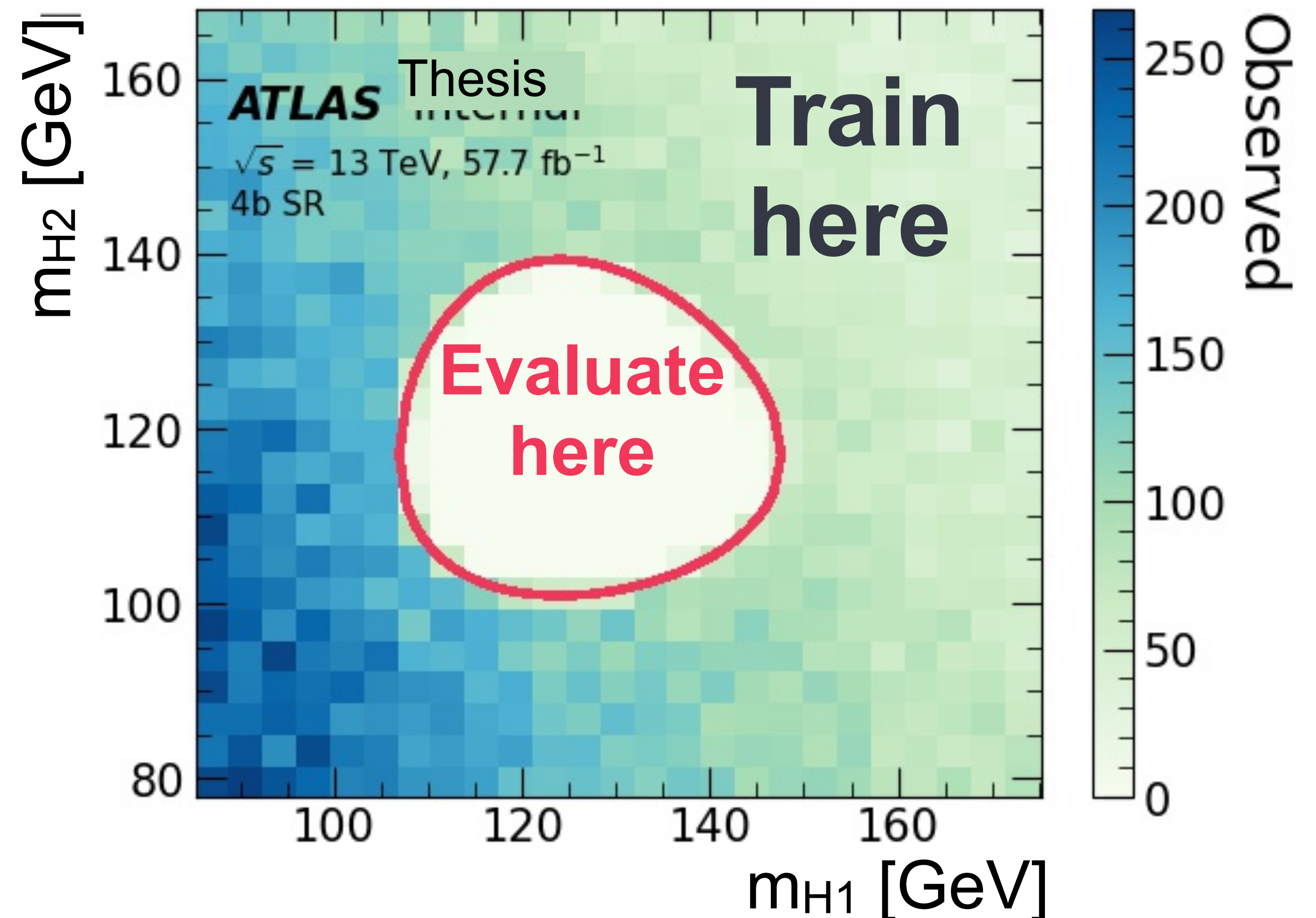
Gaussian process

x : Event kinematics

$p_{T,H1}, p_{T,H2}, \eta_{H1}, \eta_{H2}, \Delta\phi_{HH}, X_{Wt}$ [top veto]

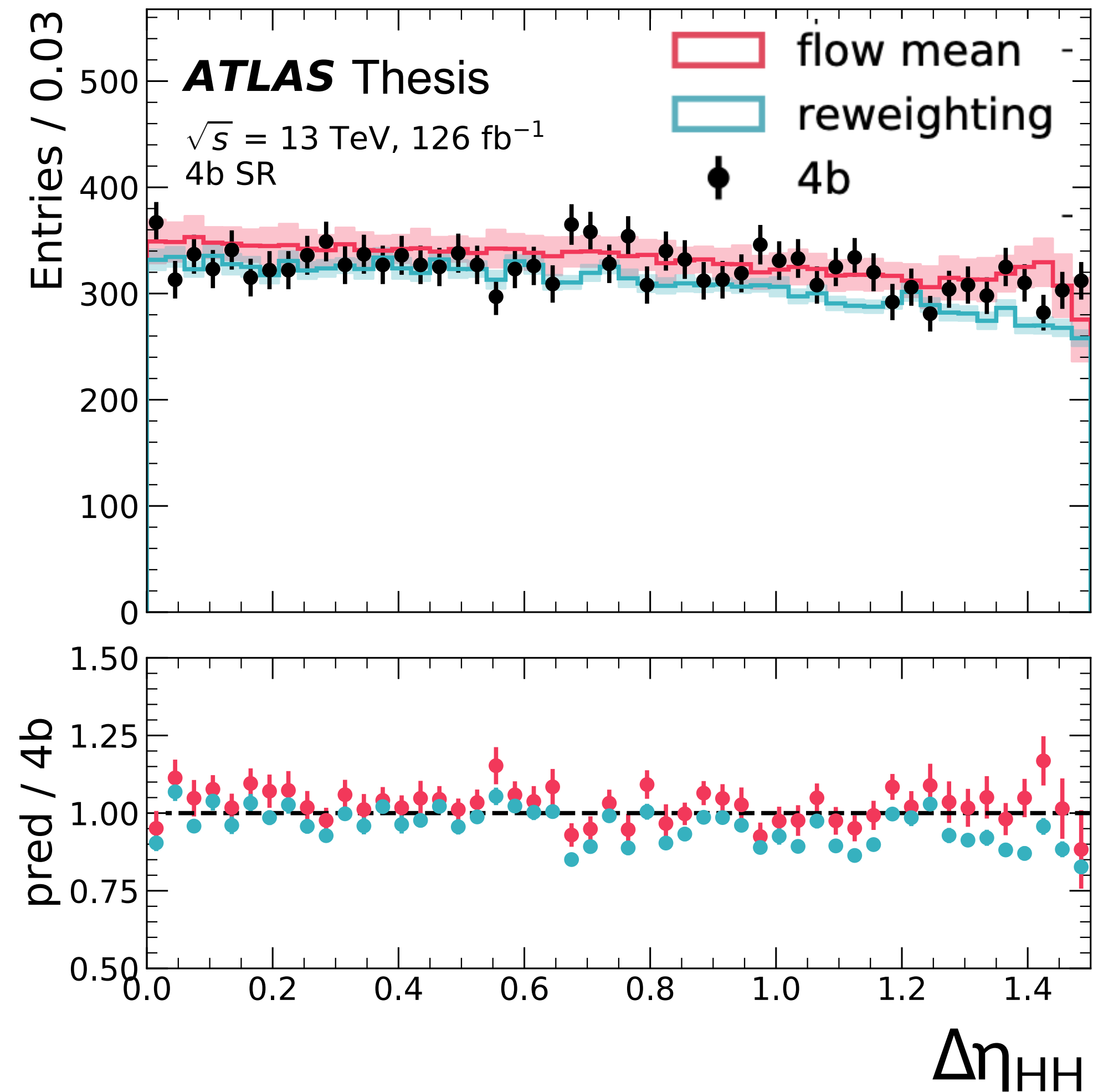
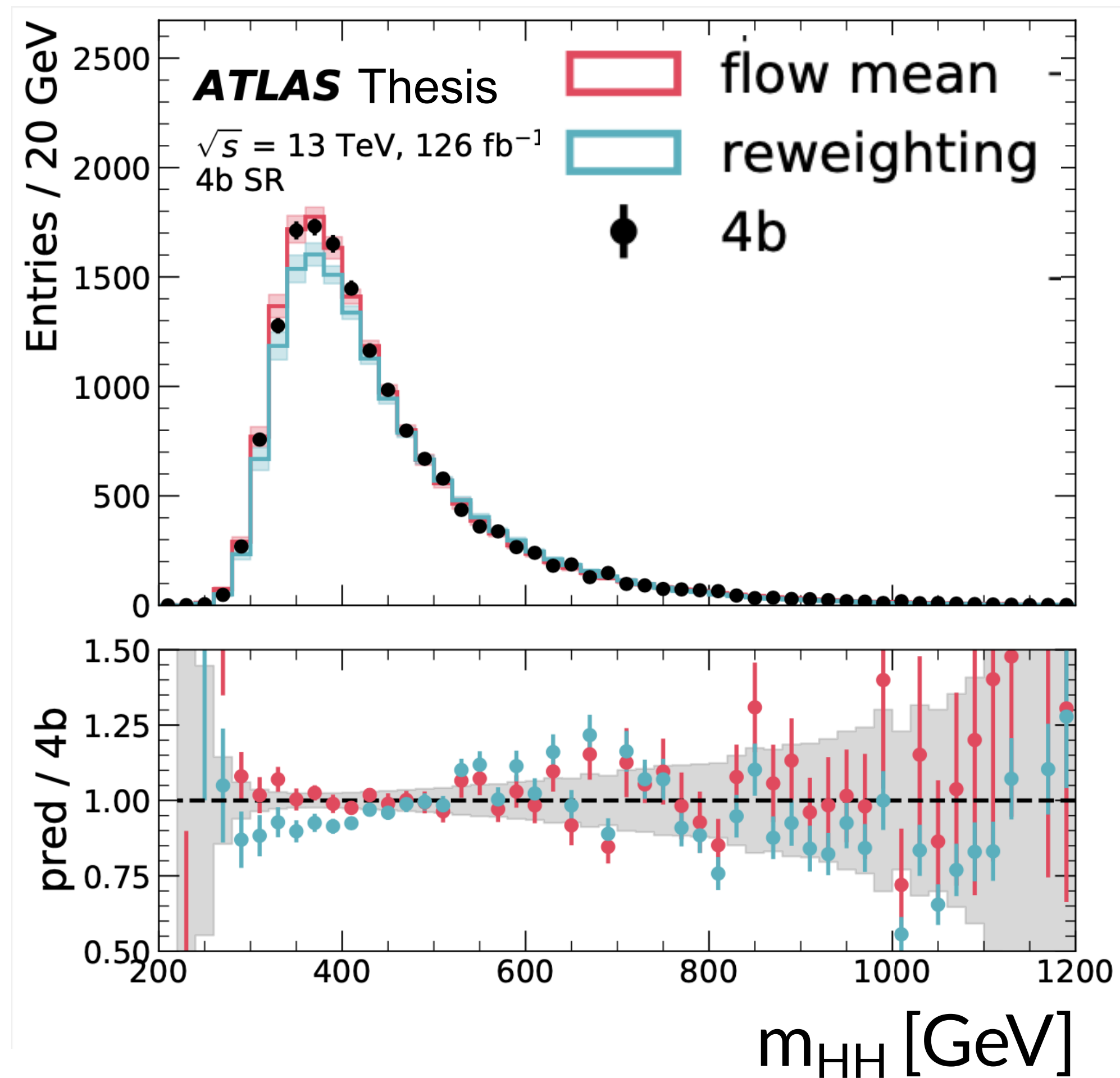


Use the smoothly varying (m_{H1}, m_{H2}) to predict **out of domain** kinematics.



Input processing : $HH \rightarrow 4b$ background modeling

Variable transformations

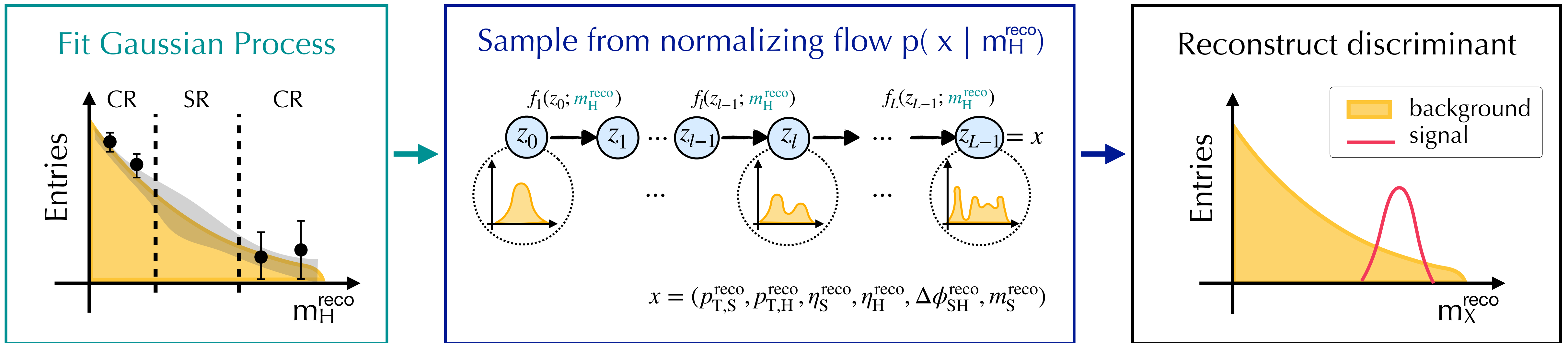


Background estimation

Instead of learning a density ratio, p_{4b}/p_{2b}
 $\hookrightarrow$ Learn the 4b density directly

$$p_{4b}(x, m_H^{\text{reco}}) = p_{4b}(m_H^{\text{reco}}) \cdot p_{4b}(x | m_H^{\text{reco}})$$

Fig 4



Same strategy for resolved and mixed!

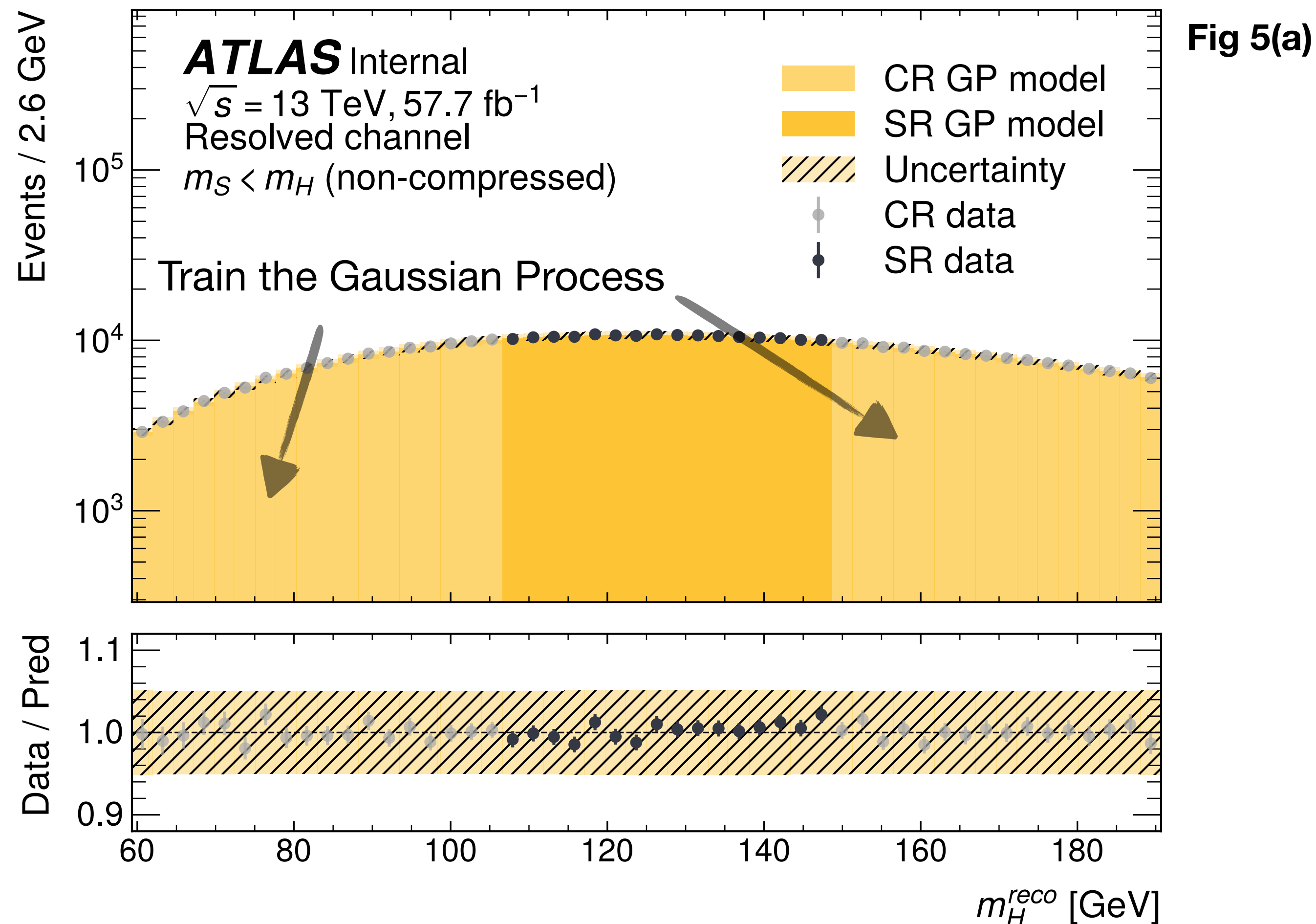
Background estimation

Instead of learning a density ratio, p_{4b}/p_{2b}

↪ Learn the 4b density directly

$$p_{4b}(x, m_H^{\text{reco}}) = p_{4b}(m_H^{\text{reco}}) \cdot p_{4b}(x | m_H^{\text{reco}})$$

Gaussian Process
Normalization + m_H shape

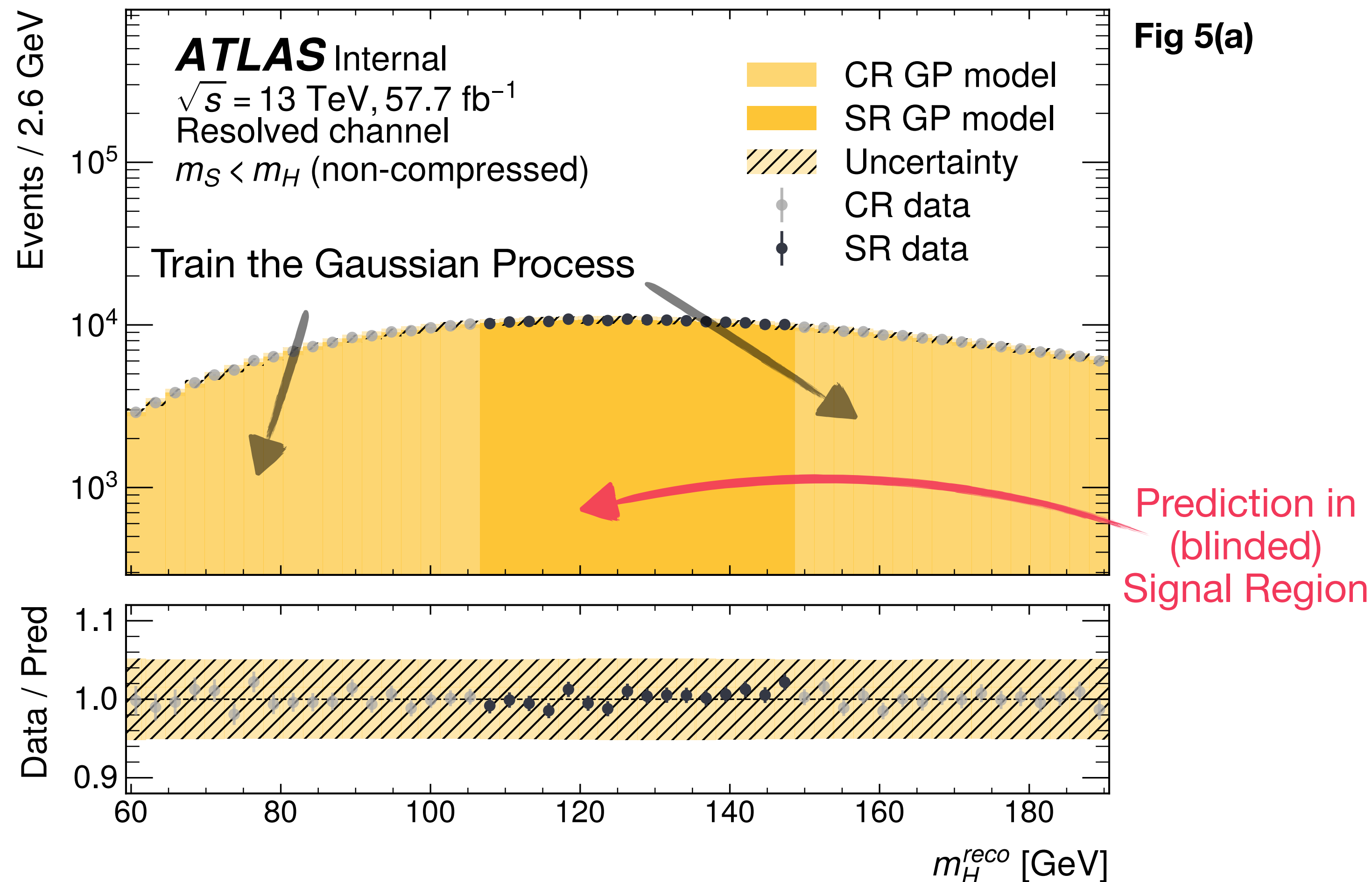


Background estimation

Instead of learning a density ratio, p_{4b}/p_{2b}
 $\hookrightarrow$ Learn the 4b density directly

$$p_{4b}(x, m_H^{\text{reco}}) = p_{4b}(m_H^{\text{reco}}) \cdot p_{4b}(x | m_H^{\text{reco}})$$

Gaussian Process
 Normalization + m_H shape



Background estimation

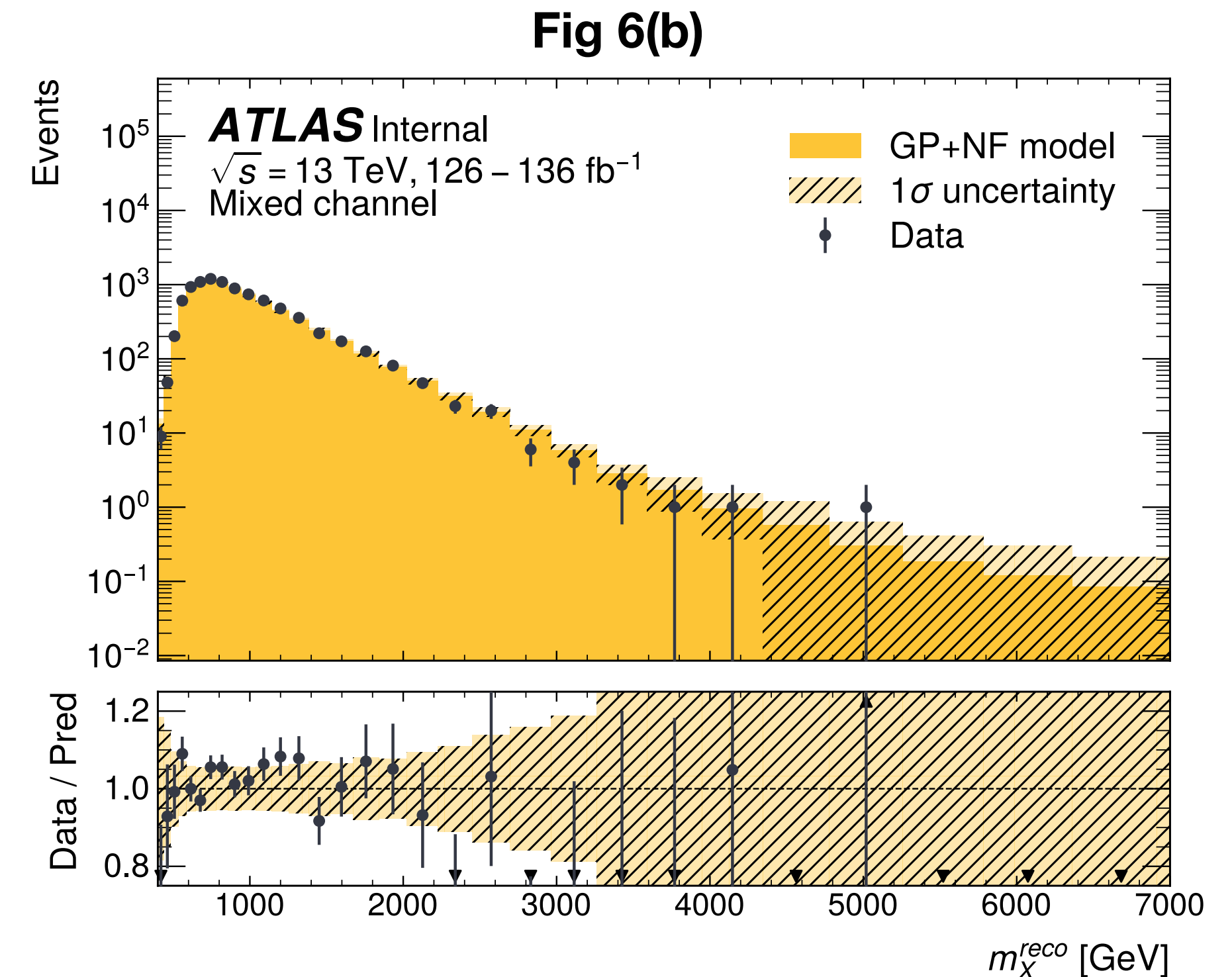
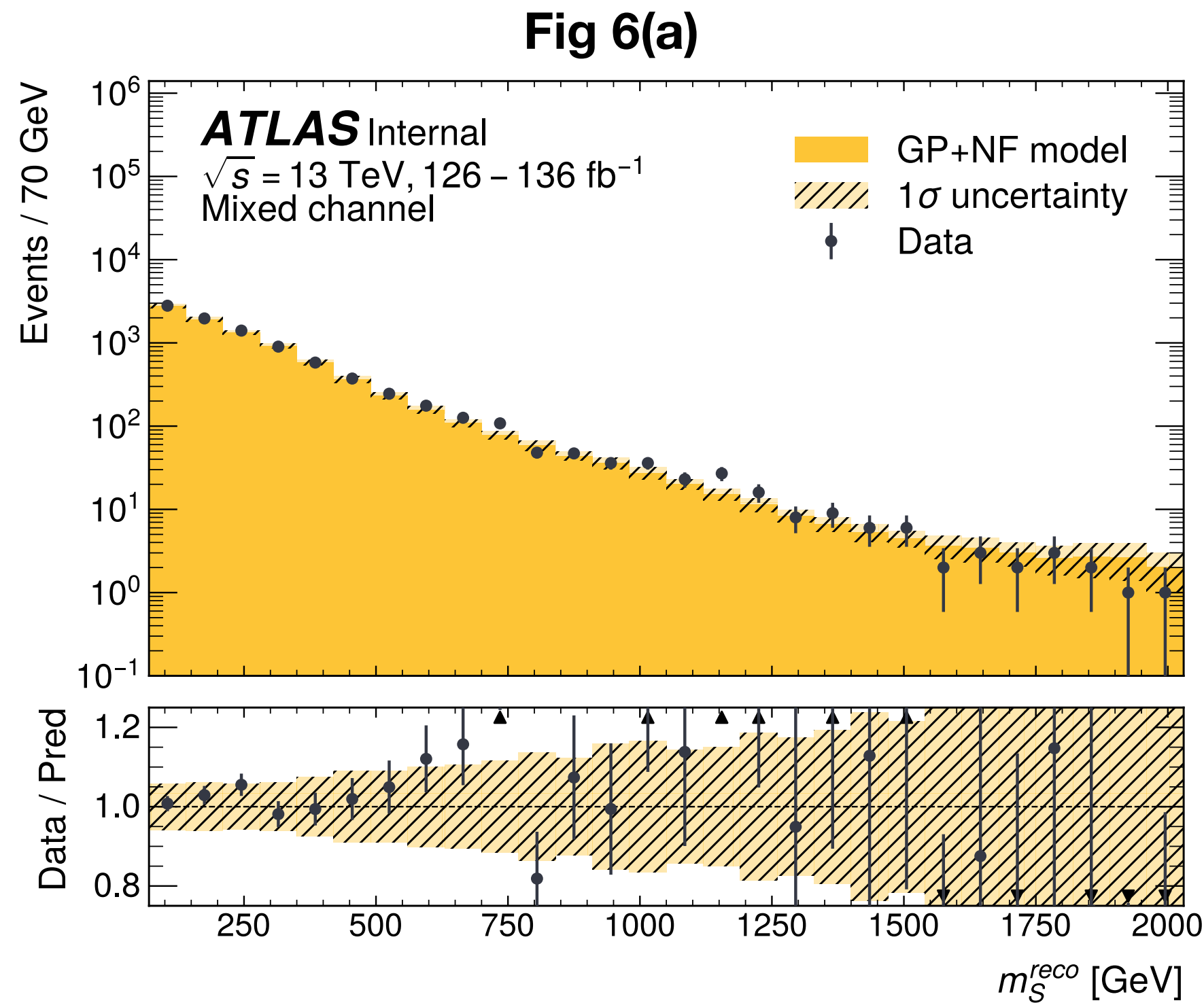
Instead of learning a density ratio, p_{4b}/p_{2b}
 $\hookrightarrow$ Learn the 4b density directly

$$p_{4b}(x, m_H^{\text{reco}}) = p_{4b}(m_H^{\text{reco}}) \cdot p_{4b}(x | m_H^{\text{reco}})$$

Conditional
Normalizing flow

Density estimation: way to generalize a 'histogram' in high dim, §2.2

$$x = \begin{pmatrix} p_{T,S}^{\text{reco}} \\ p_{T,H}^{\text{reco}} \\ \eta_S^{\text{reco}} \\ \eta_H^{\text{reco}} \\ \Delta\phi_{SH}^{\text{reco}} \\ m_S^{\text{reco}} \end{pmatrix}$$



Background estimation

Instead of learning a density ratio, p_{4b}/p_{2b}

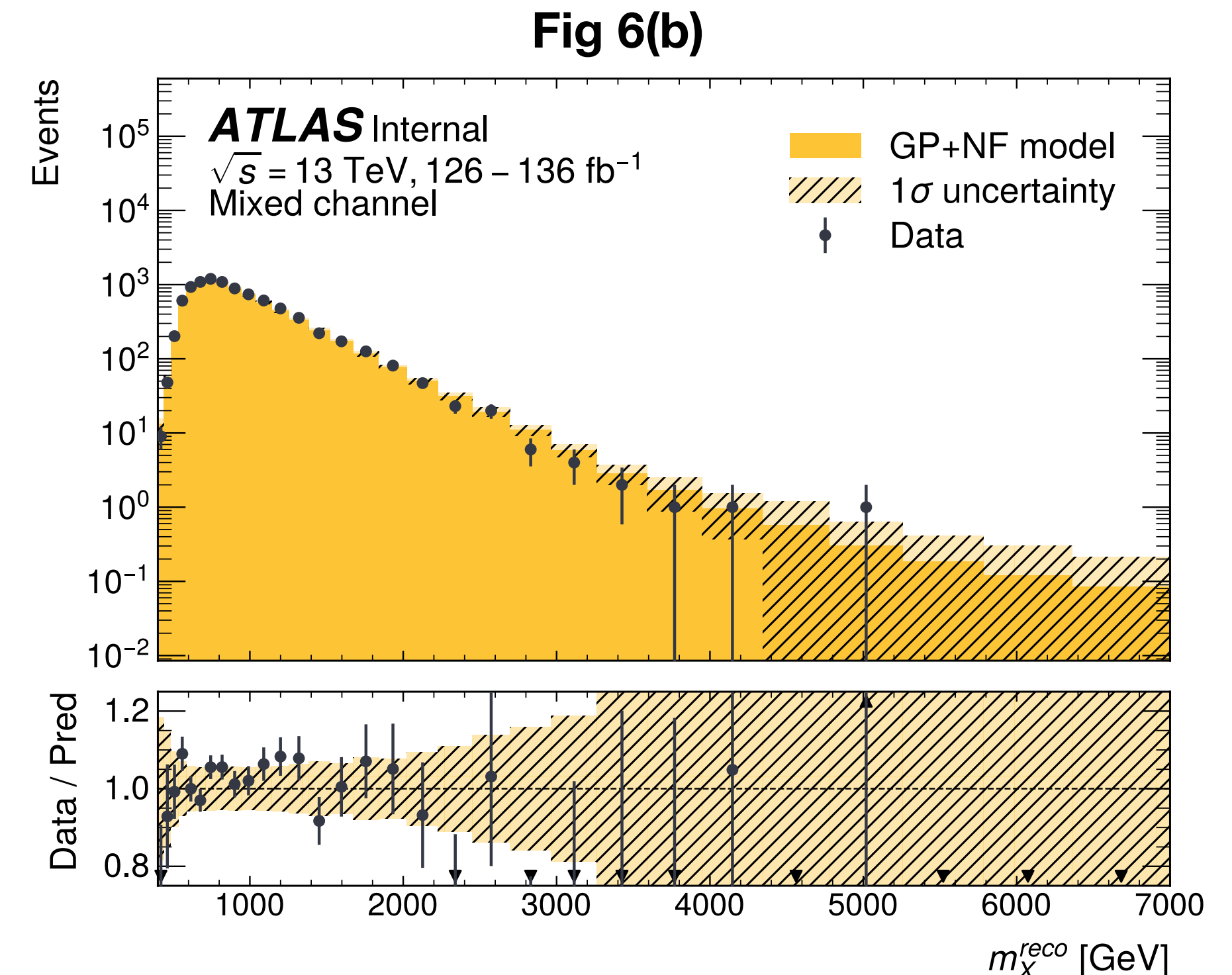
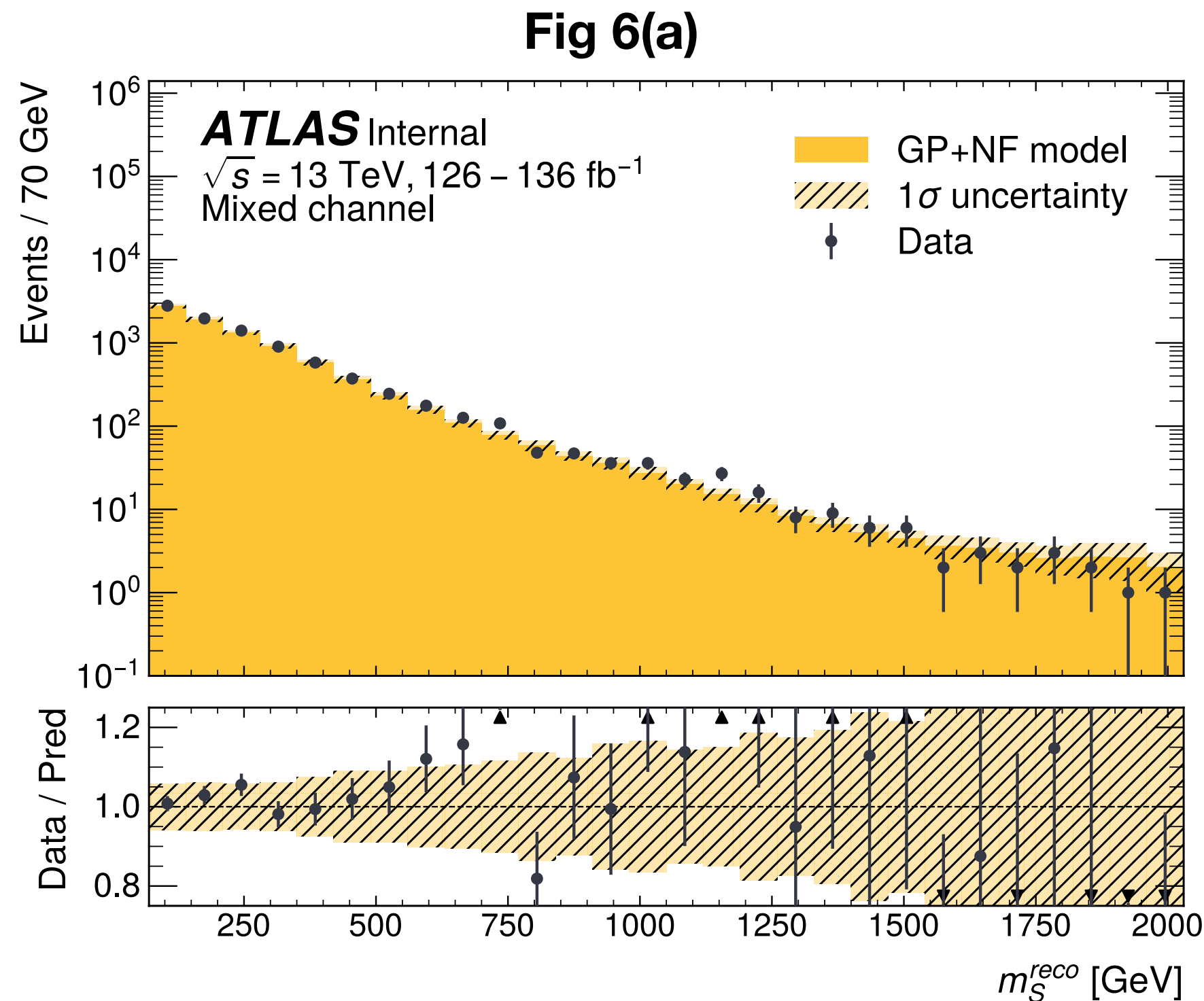
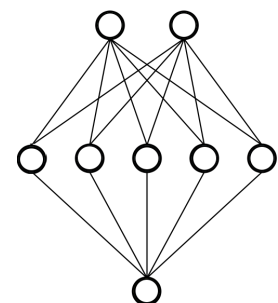
↪ Learn the 4b density directly

$$p_{4b}(x, m_H^{\text{reco}}) = p_{4b}(m_H^{\text{reco}}) \cdot p_{4b}(x | m_H^{\text{reco}})$$

Conditional
Normalizing flow

Density estimation: way to generalize a 'histogram' in high dim, §2.2

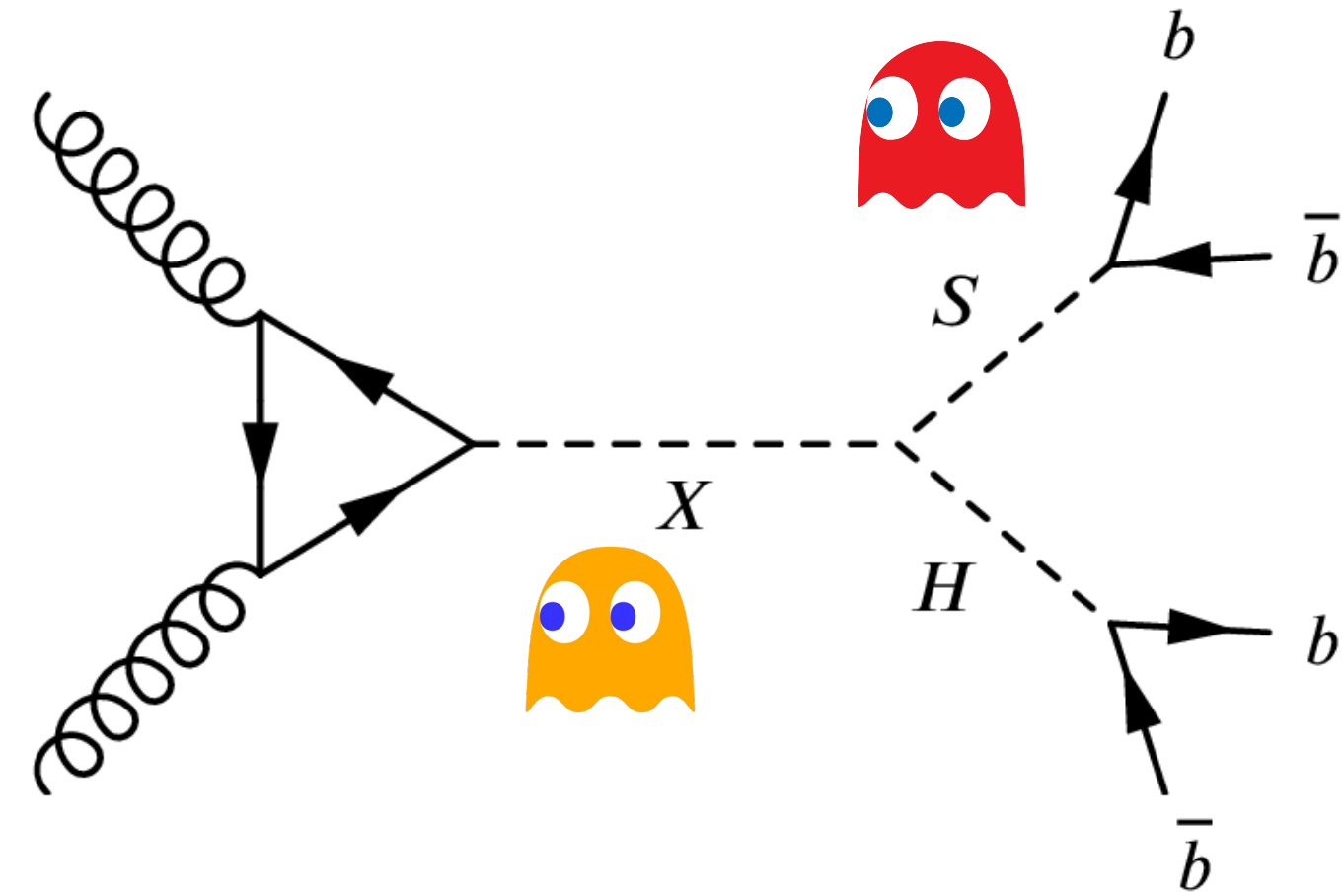
$$x = \begin{pmatrix} p_{T,S}^{\text{reco}} \\ p_{T,H}^{\text{reco}} \\ \eta_S^{\text{reco}} \\ \eta_H^{\text{reco}} \\ \Delta\phi_{SH}^{\text{reco}} \\ m_S^{\text{reco}} \end{pmatrix}$$



Retrained 25x → error bar (visualized above)

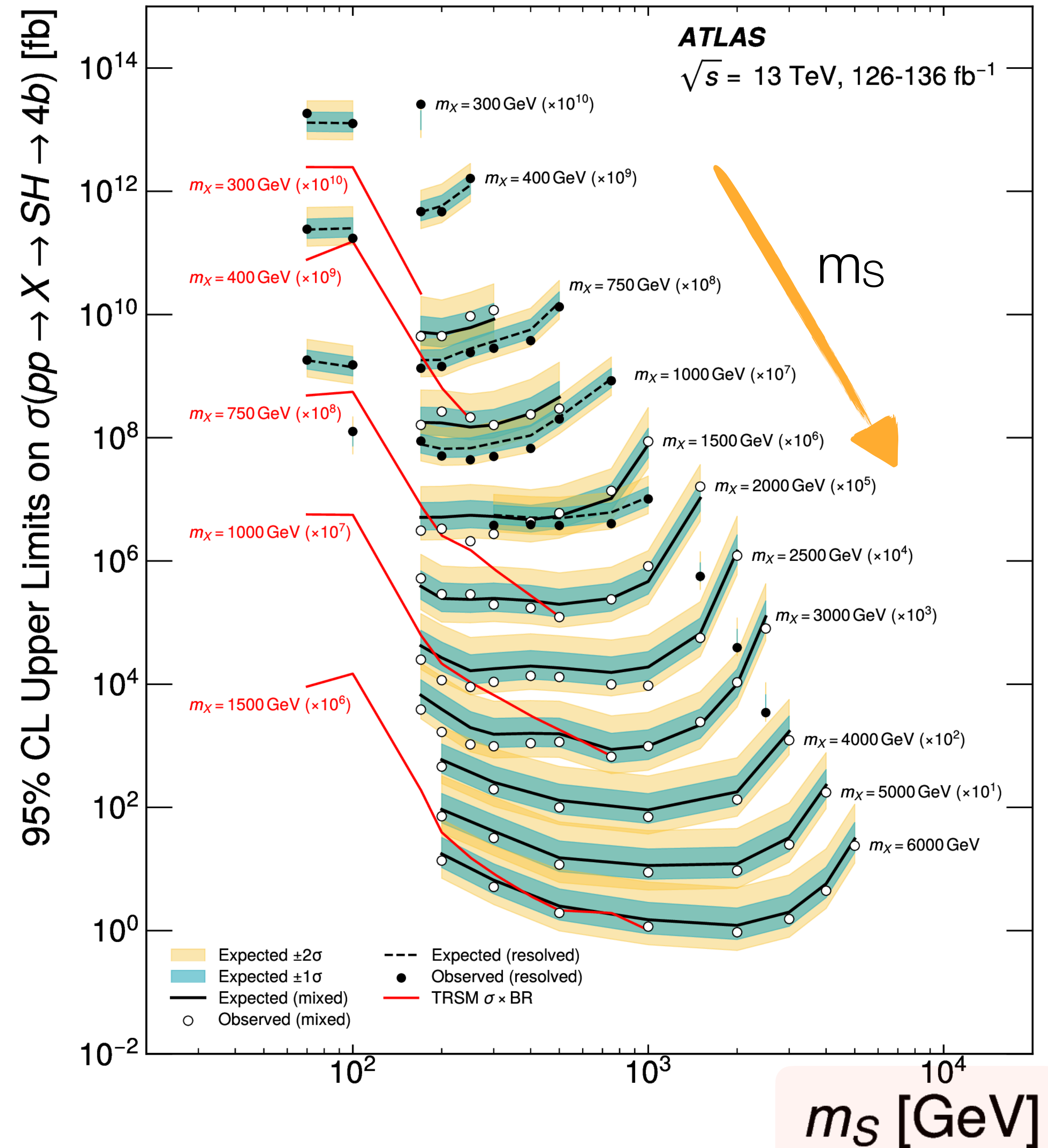
Results

$$X \rightarrow SH \rightarrow 4b$$

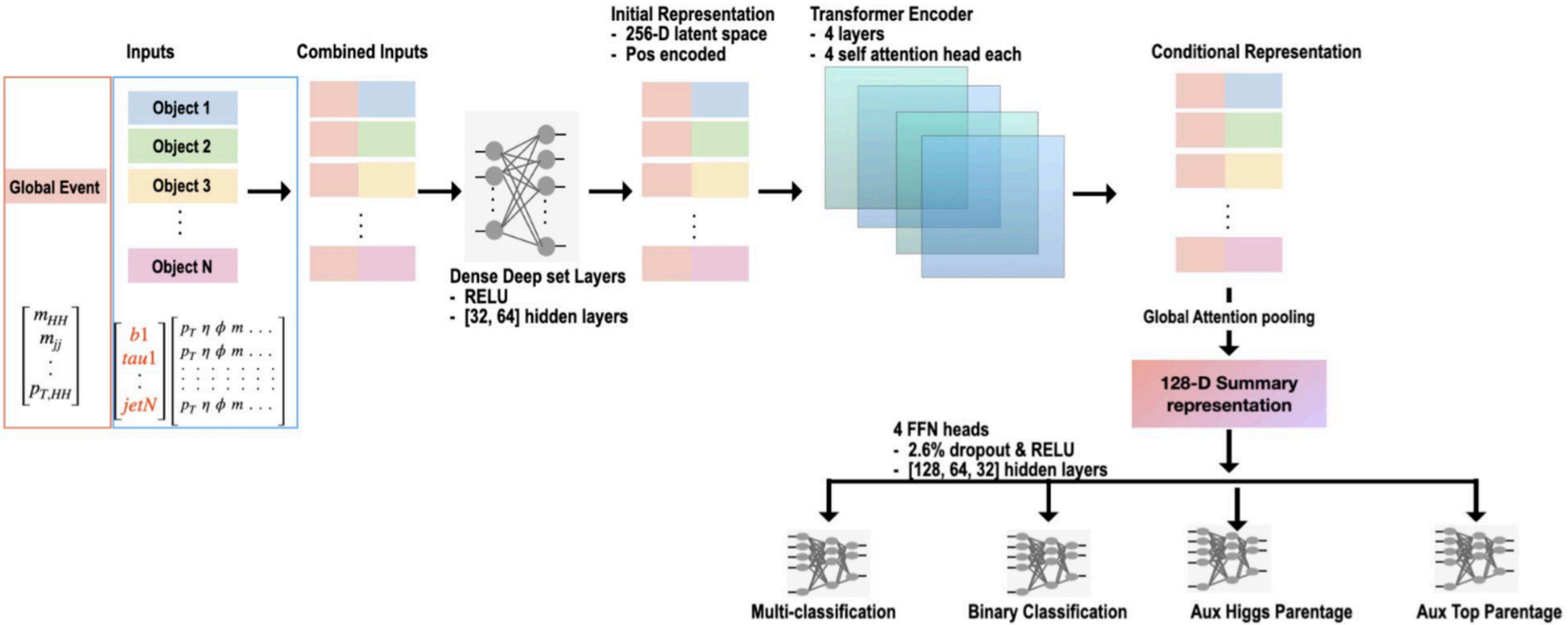


What have we learned?

- (1) We haven't found the new physics (yet)!
 - (2) Very good agreement between the expectation and observation.
- Triumph for the norm flow + UQ



ATLAS HH(bbττ)



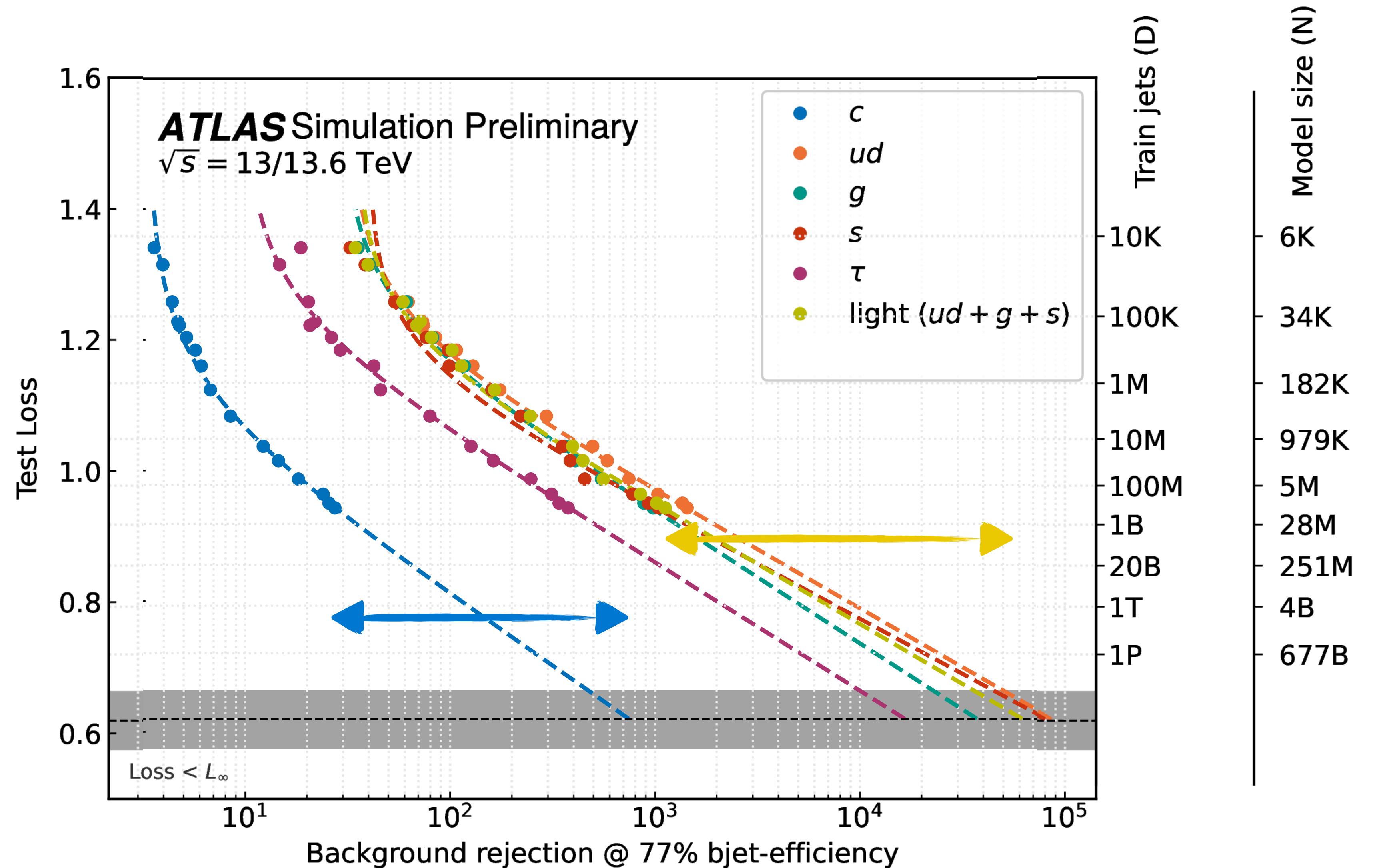
Performance translates to physics

Dashed lines fit to:

$$R(L_i) = Ae^{BL_i} + C$$

Rejection improvement could be as large as **100x!** 🤯

Only gain factor of 3 on Delphes
Vigl, **NH**, Kagan, Heinrich, [2602.15781](#)



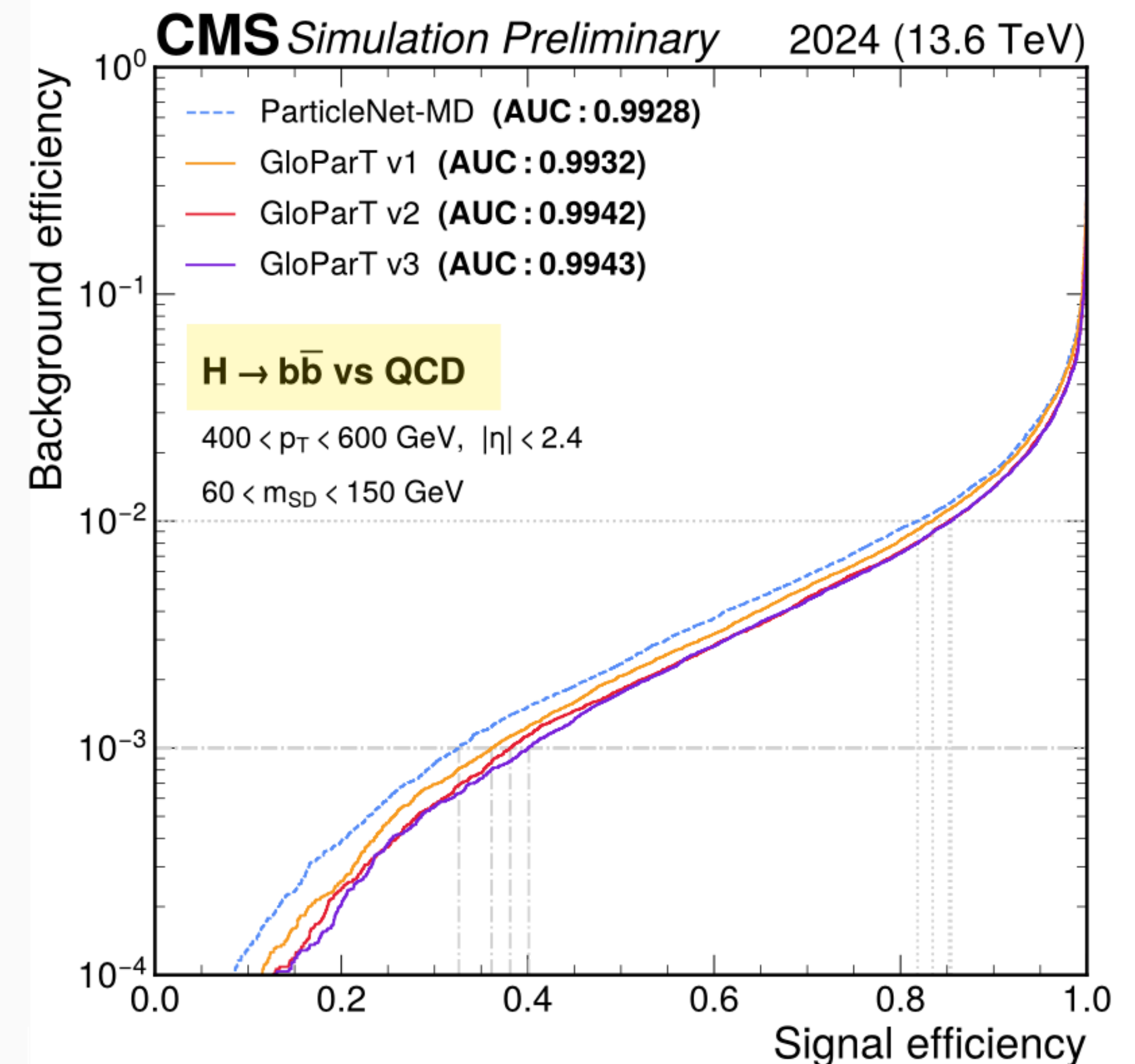


Boosted Jet tagging

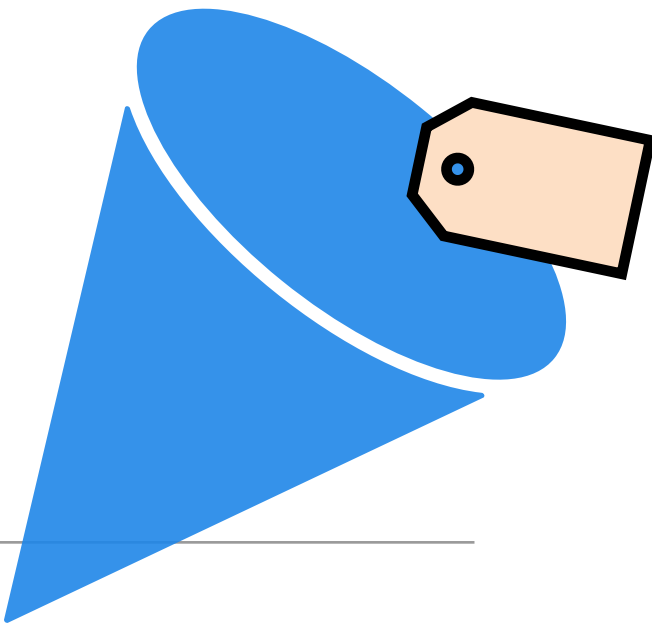
Latest GloParT model has *even more* classes: 374 classes (+ regression nodes!)

↪ General for pretraining for downstream tasks

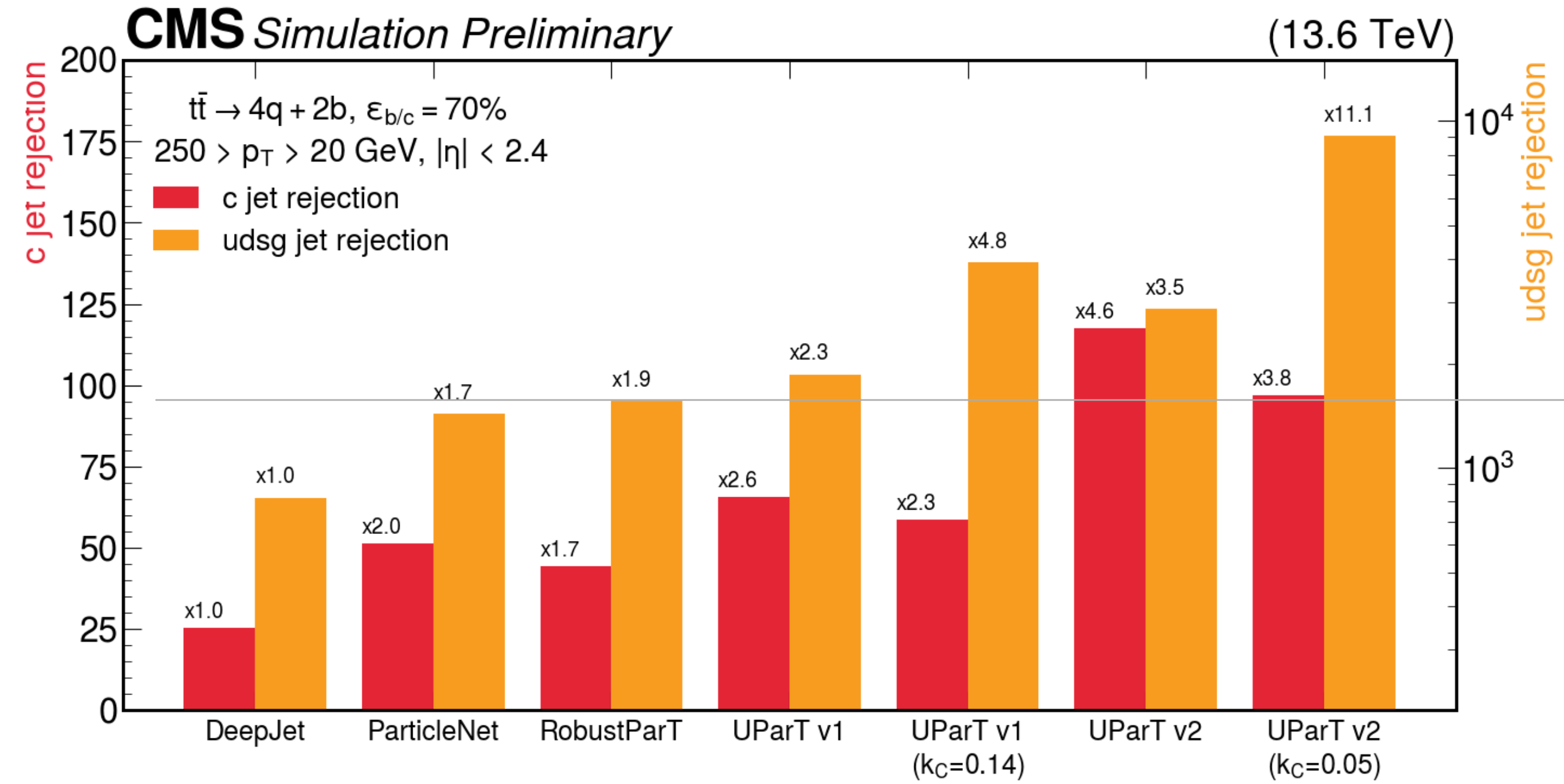
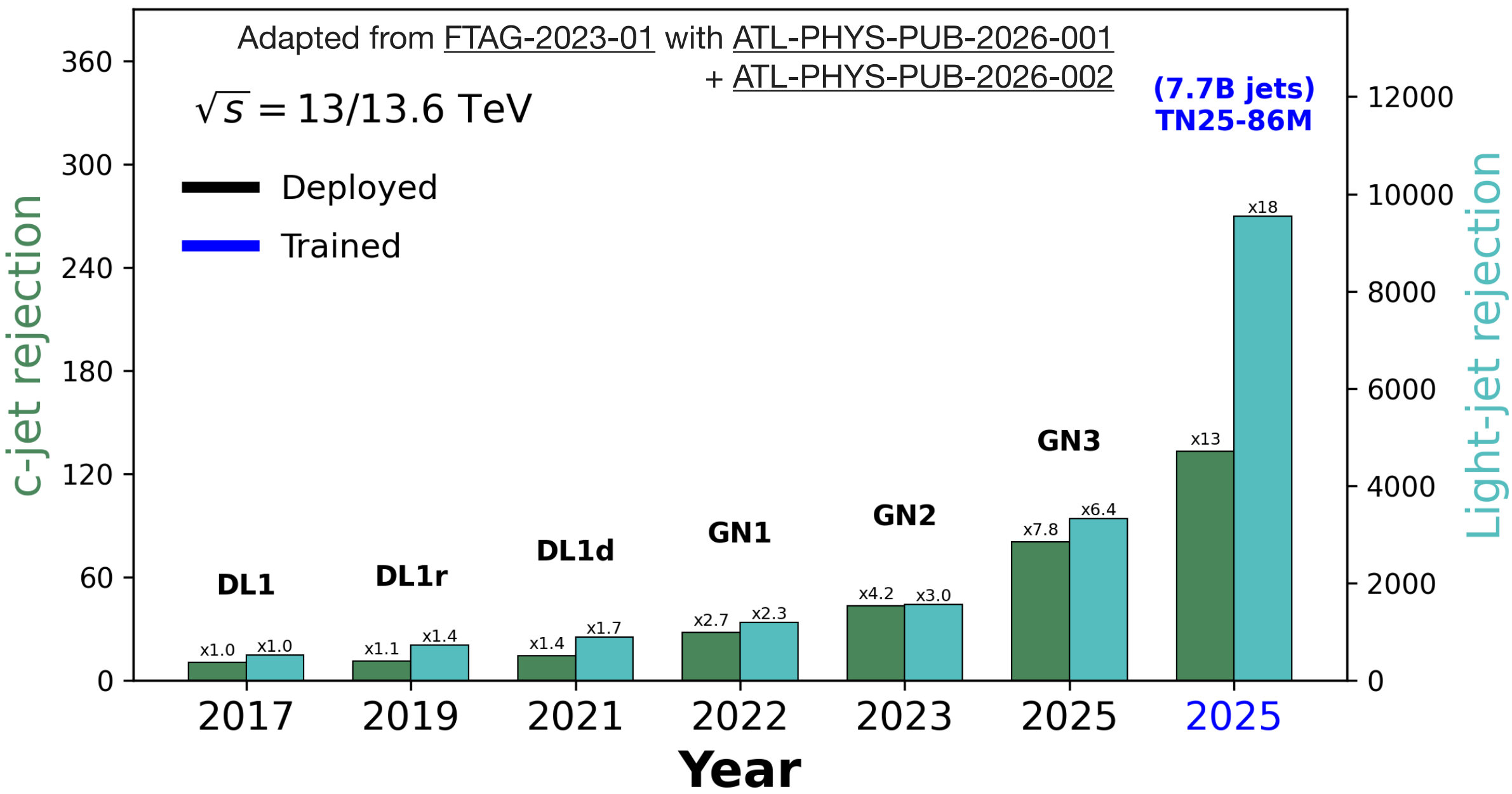
Impact: 5% improvement in signal eff
at a constant background rejection



Small-R (b-tagging)



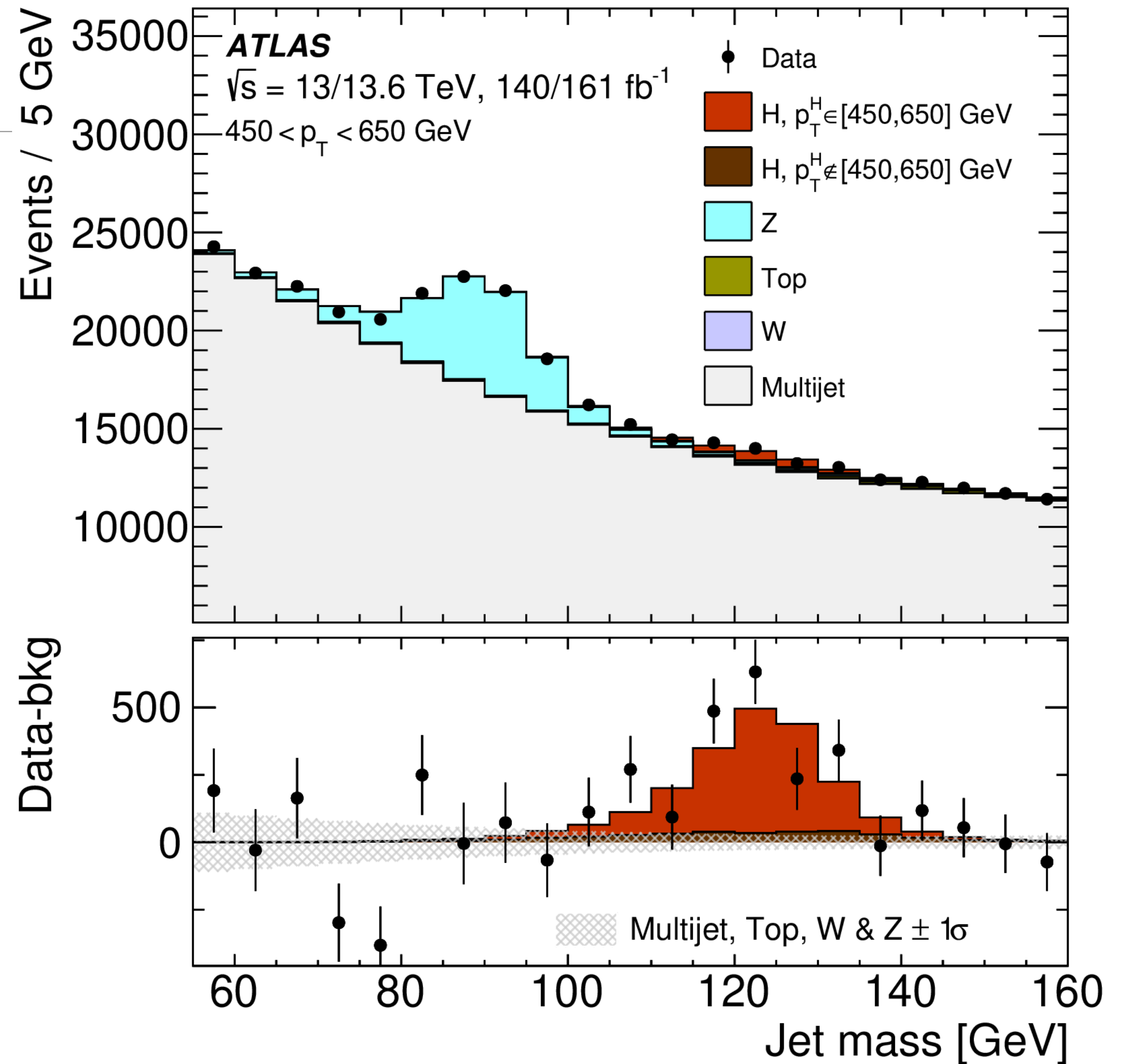
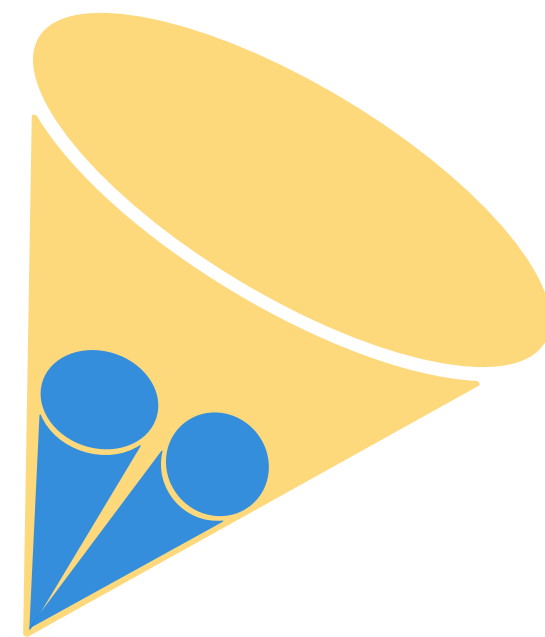
DP-2025-081



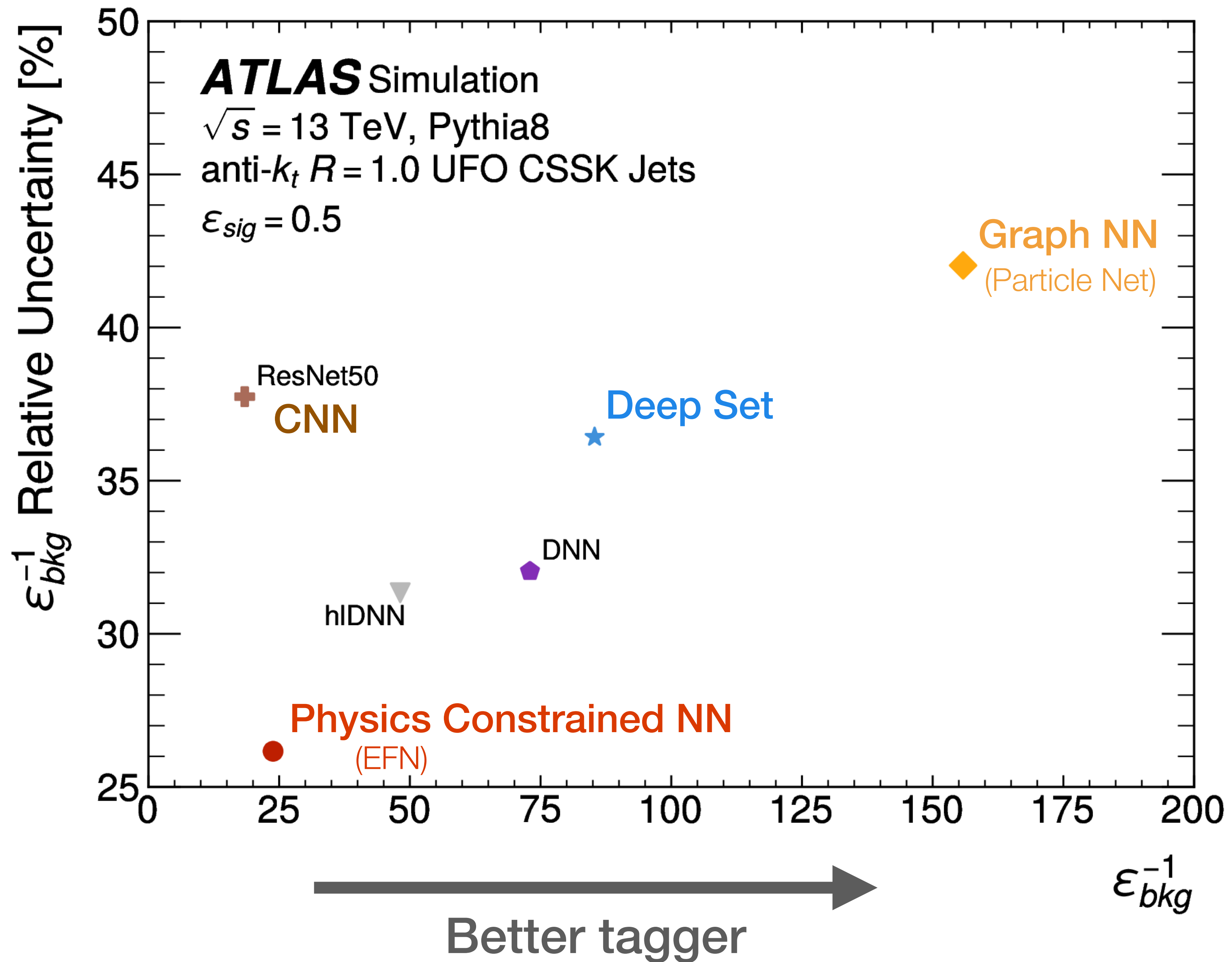
Physics impact

Boosted Higgs evidence

Obs (exp) 3.8σ (2.5σ)



Larger
uncertainties



Where is our HEP training data?



Eur. Phys. J. C 85 (2025) 234
DOI: [10.1140/epjc/s10052-024-13701-w](https://doi.org/10.1140/epjc/s10052-024-13701-w)



CERN-EP-2024-100
17th July 2025

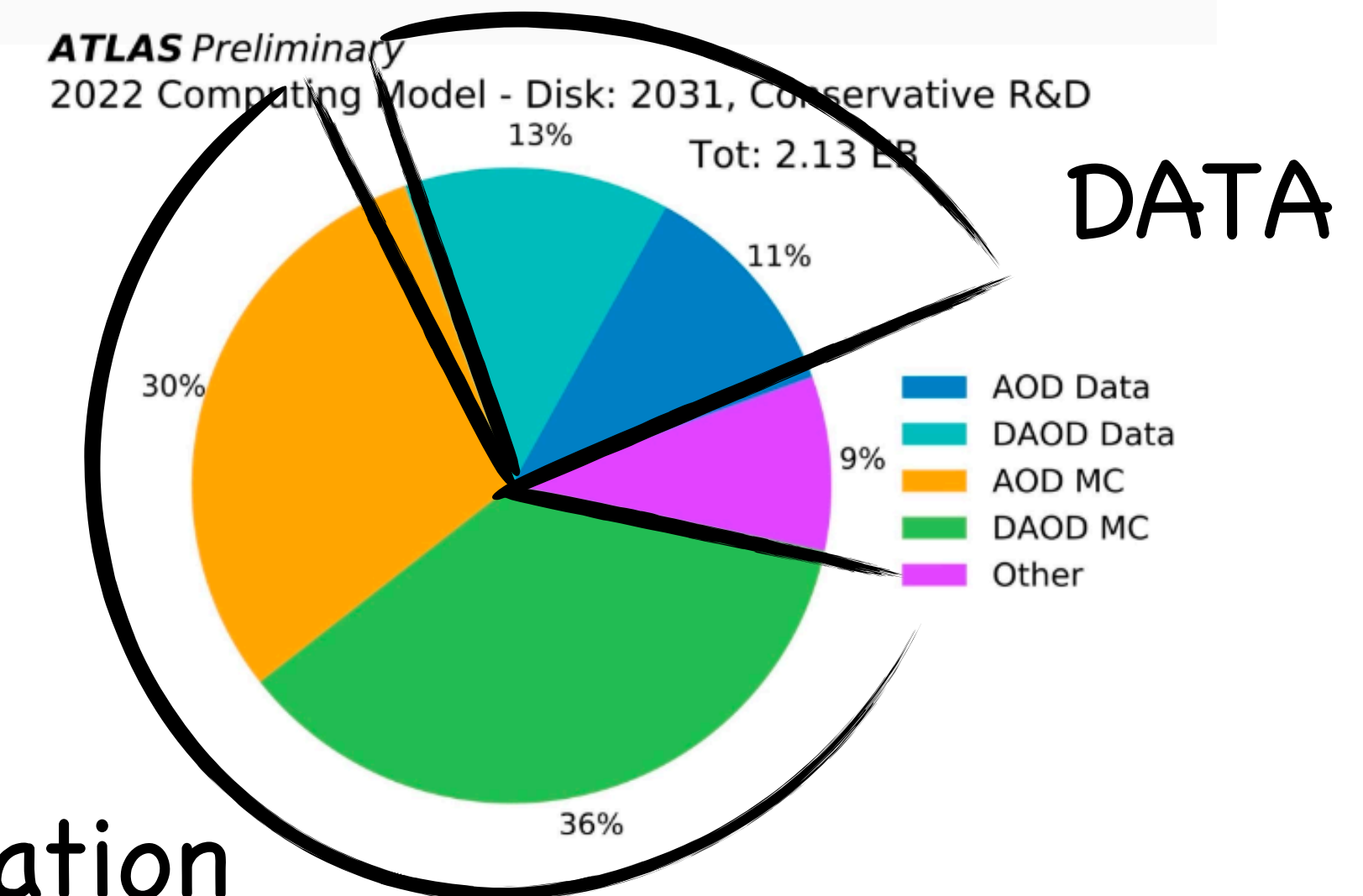
Software and computing for Run 3 of the ATLAS experiment at the LHC

During Run 2, ATLAS recorded about 18 billion complete events. Generally, 2–3 times more simulated events are produced than real detector events are recorded. This volume of MC simulation is required to keep statistical uncertainties in predictions small, and to provide capacity for simulation of new physics signals.

[Eur. Phys. J. C 85 \(2025\) 234](#)

This will change what a particle physics foundation model looks like!

Simulation



[LHCC-G-182](#)

How to get more simulation?

Key idea: just simulate little bricks

$$p(E_{\text{dep}}, \{x_{\text{sec}}\}, \{n\} \mid x_{\text{inc}}, \lambda)$$

Deposited
energy

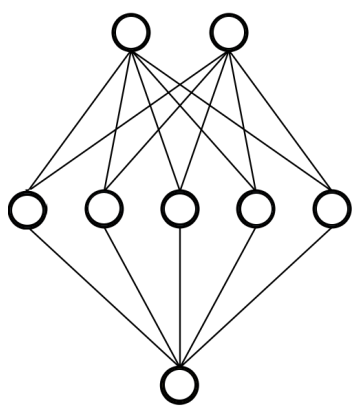
Set of
secondaries

Number of
secondaries

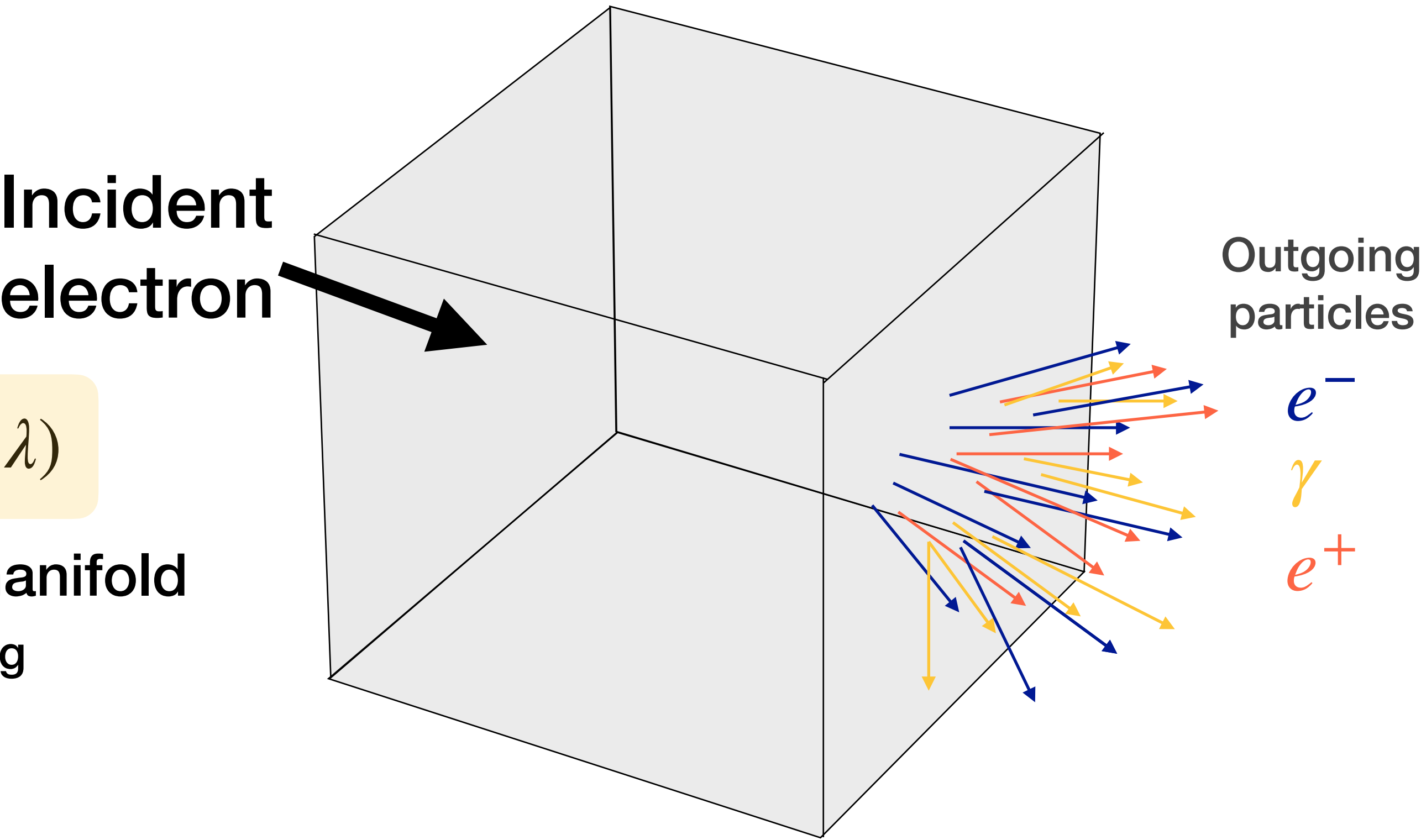
Incident
particle

Material
properties

$$= p(\{n\} \mid x_{\text{inc}}, \lambda) \, p(E_{\text{dep}}, \{x_{\text{sec}}\} \mid \{n\}, x_{\text{inc}}, \lambda)$$



Generative model on the manifold
Riemannian Flow Matching



How to get more simulation?

Key idea: just simulate little bricks

$$p(E_{\text{dep}}, \{x_{\text{sec}}\}, \{n\} \mid x_{\text{inc}}, \lambda)$$

Deposited
energy

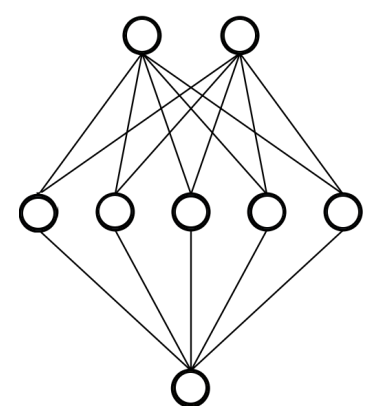
Set of
secondaries

Number of
secondaries

Incident
particle

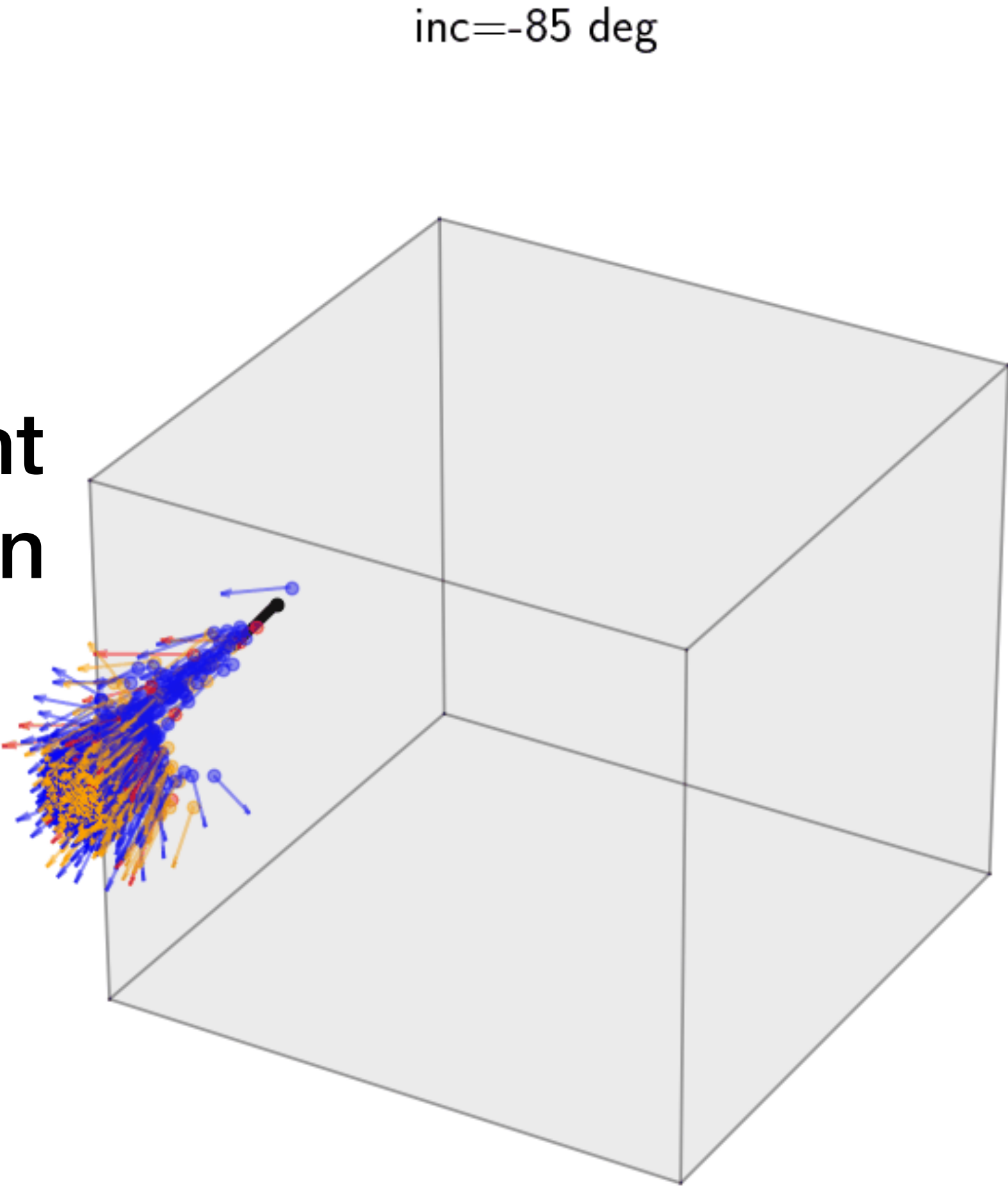
Material
properties

$$= p(\{n\} \mid x_{\text{inc}}, \lambda) \, p(E_{\text{dep}}, \{x_{\text{sec}}\} \mid \{n\}, x_{\text{inc}}, \lambda)$$



Generative model on the manifold
Riemannian Flow Matching

**Incident
electron**



**Outgoing
particles**

e^-
 γ
 e^+

How to get more simulation?

Key idea: just simulate little bricks

$$p(E_{\text{dep}}, \{x_{\text{sec}}\}, \{n\} \mid x_{\text{inc}}, \lambda)$$

Deposited
energy

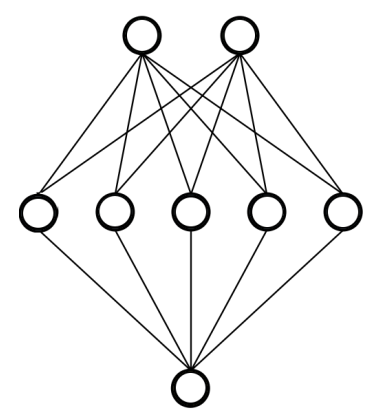
Set of
secondaries

Number of
secondaries

Incident
particle

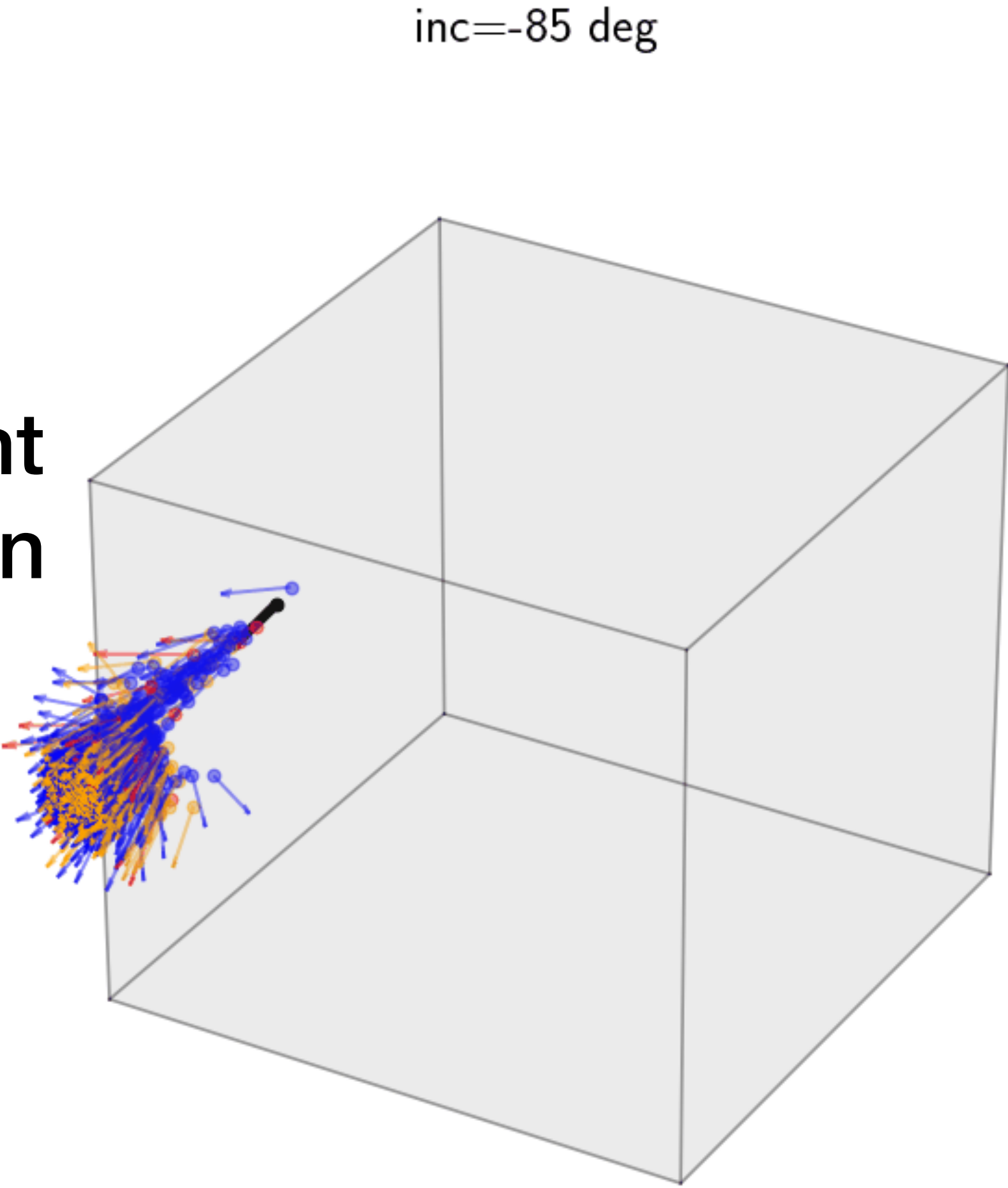
Material
properties

$$= p(\{n\} \mid x_{\text{inc}}, \lambda) \, p(E_{\text{dep}}, \{x_{\text{sec}}\} \mid \{n\}, x_{\text{inc}}, \lambda)$$



Generative model on the manifold
Riemannian Flow Matching

**Incident
electron**

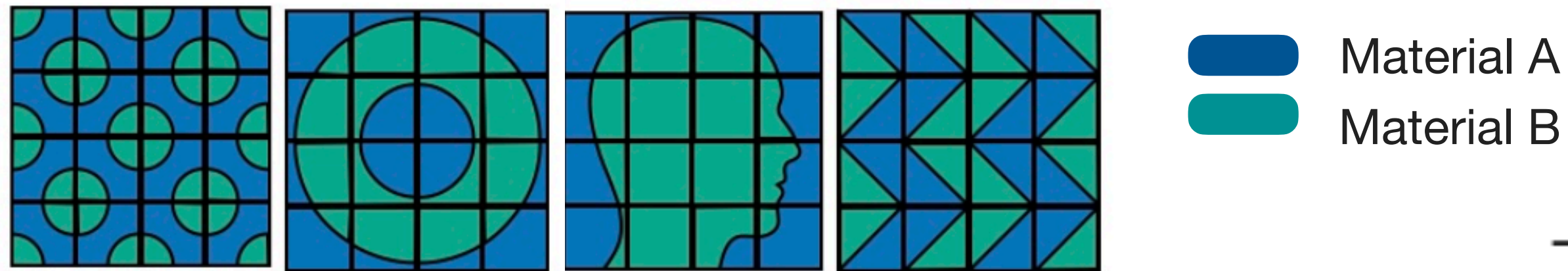


**Outgoing
particles**

e^-
 γ
 e^+

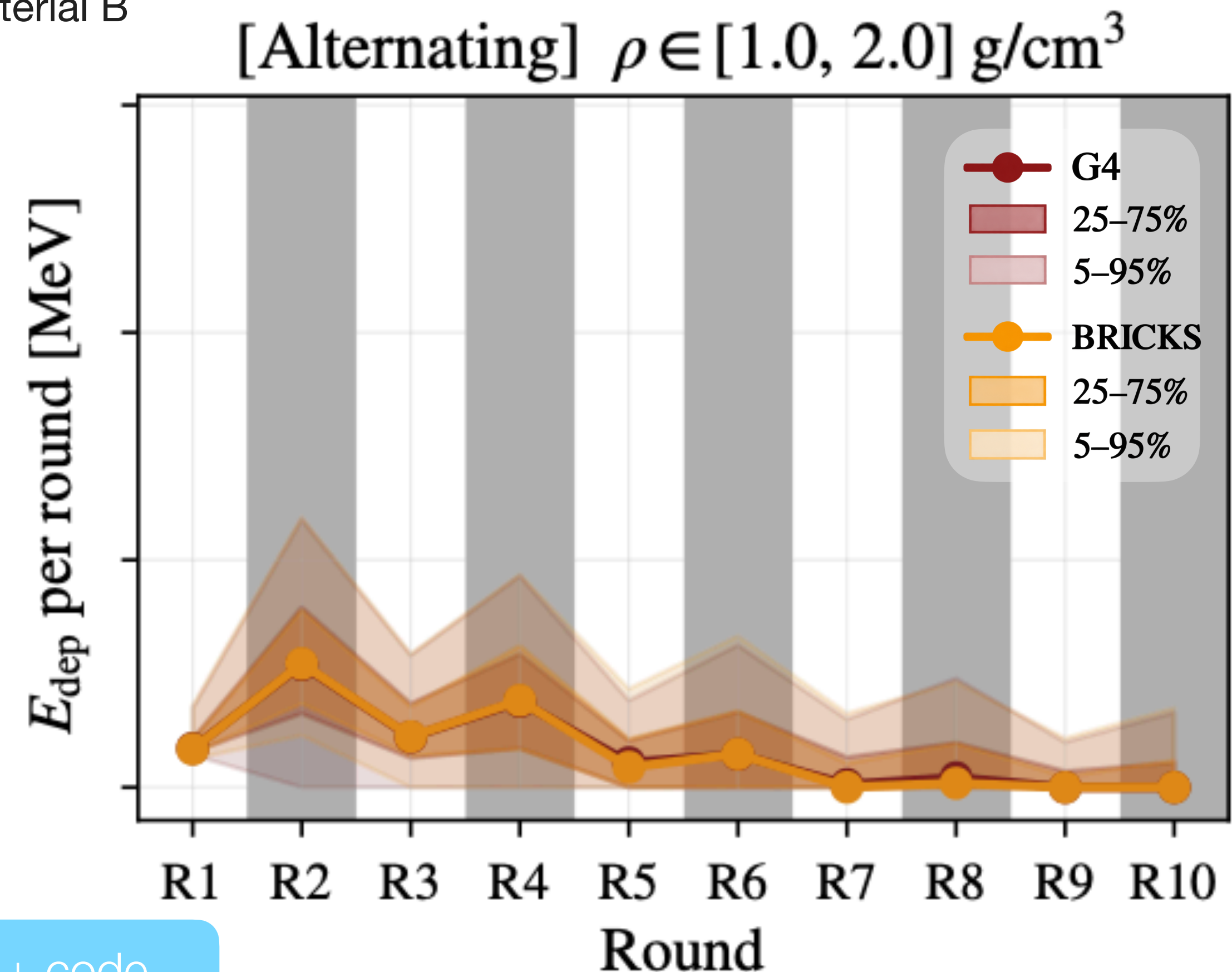
How to get more simulation?

Impact: Stitch bricks together to build complicated structures

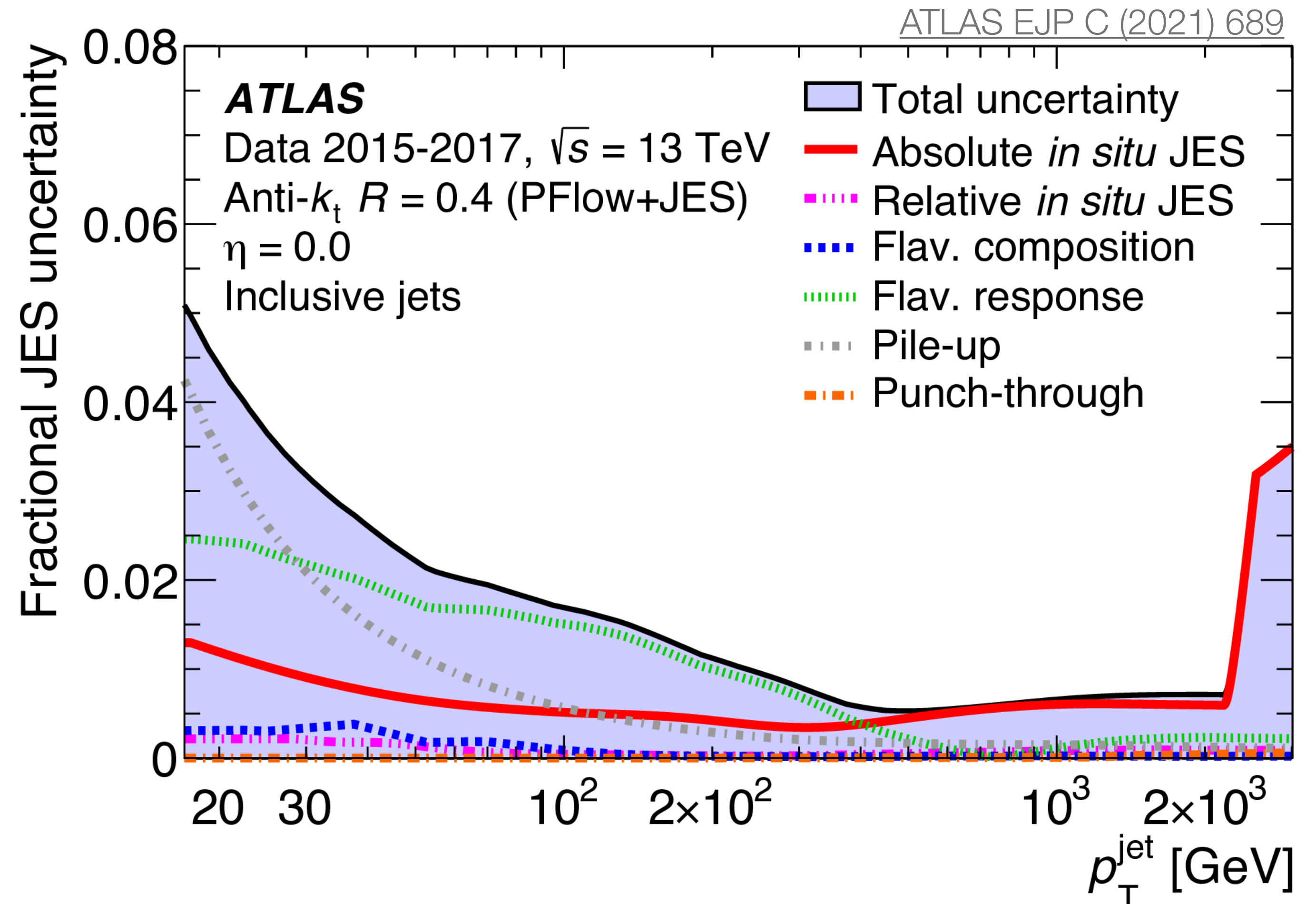
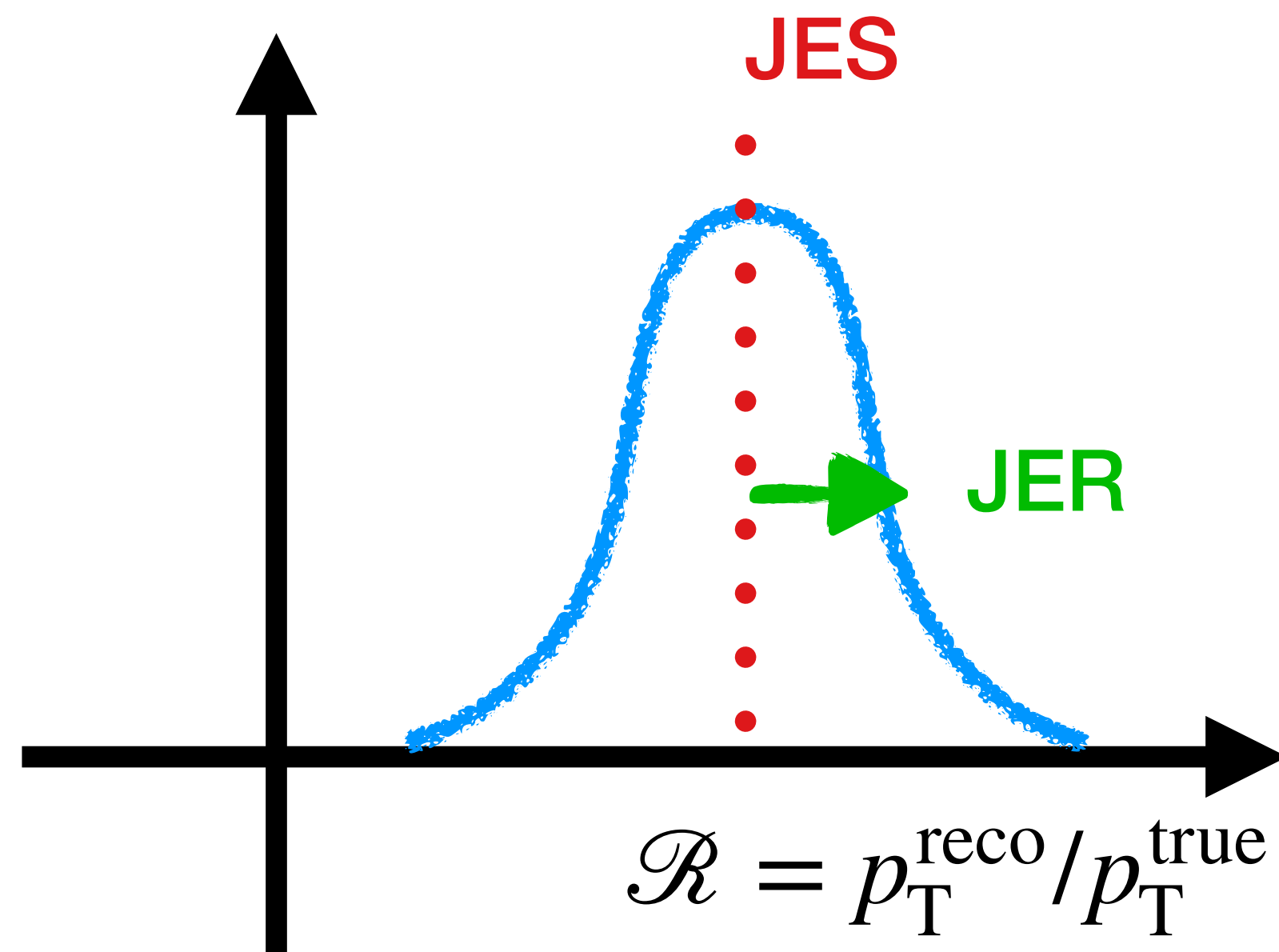


→ Faster than GEANT4

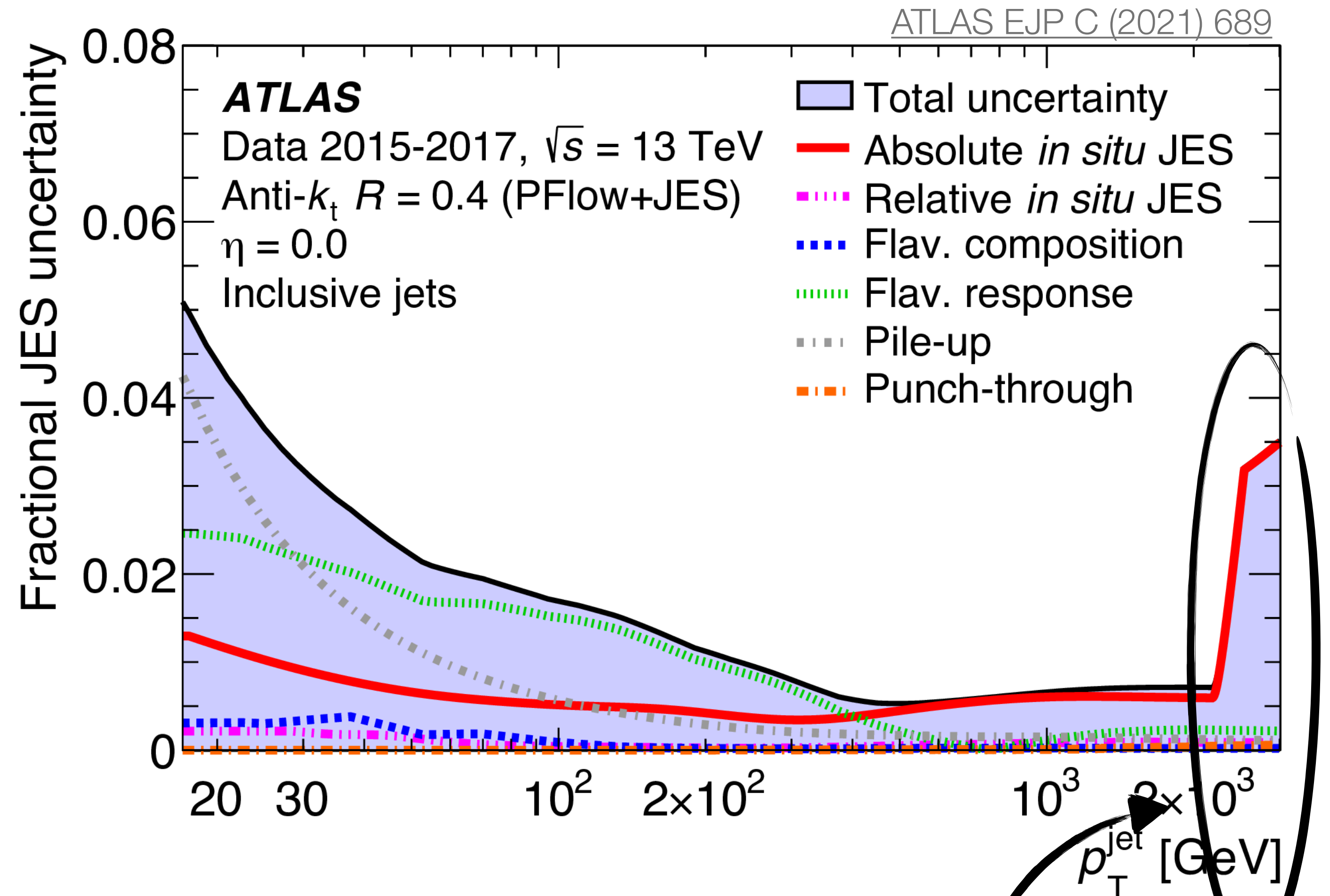
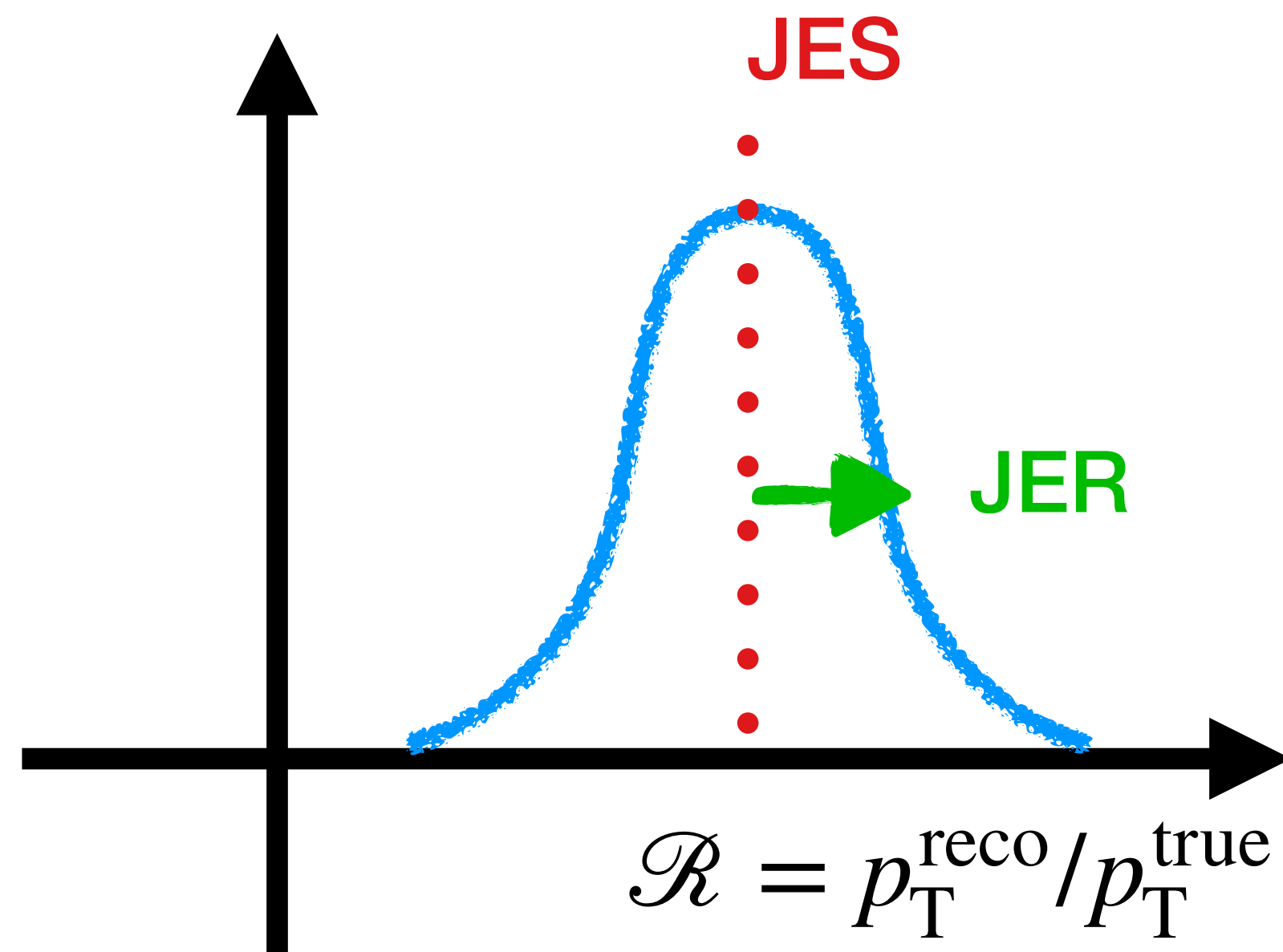
Neural surrogates with 0-shot
transfer across detector geometries!



How do we calibrate jets?



How do we calibrate jets?

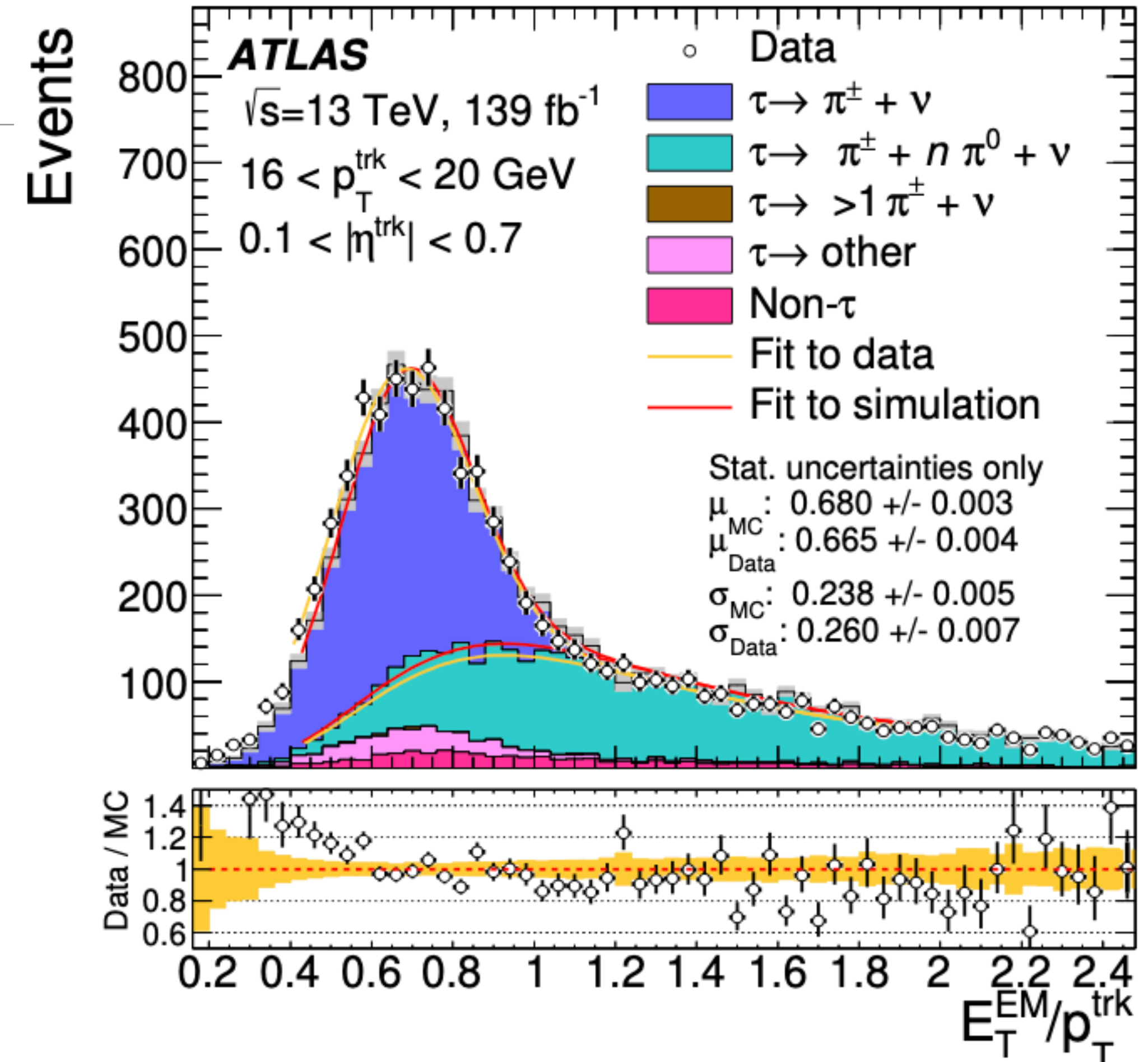
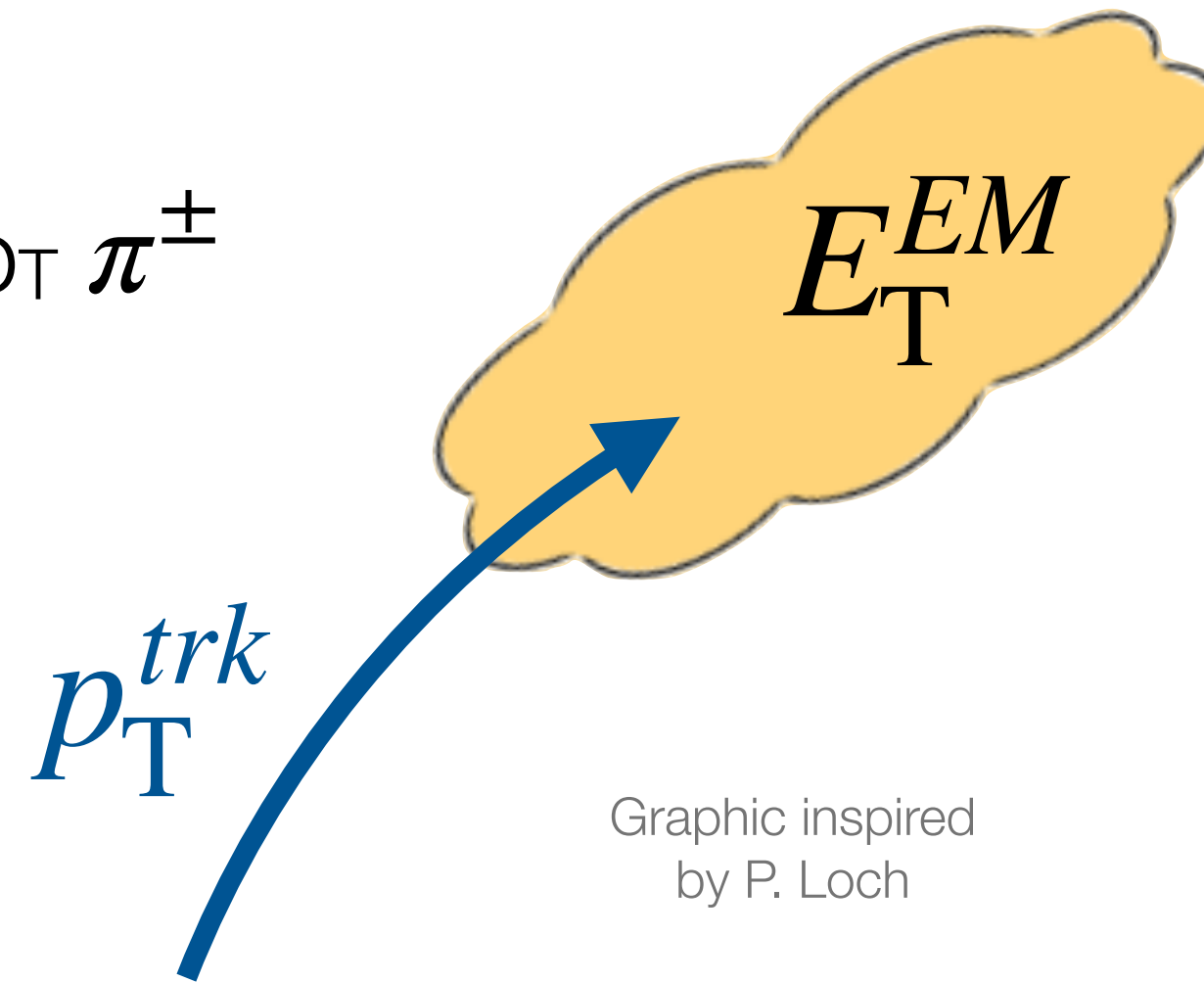


How to decrease this uncertainty??

What about calibrated inputs?

$$W^{\pm} \rightarrow \tau^{\pm} (\rightarrow \pi^{\pm} \nu_{\tau}) \nu_{\tau}$$

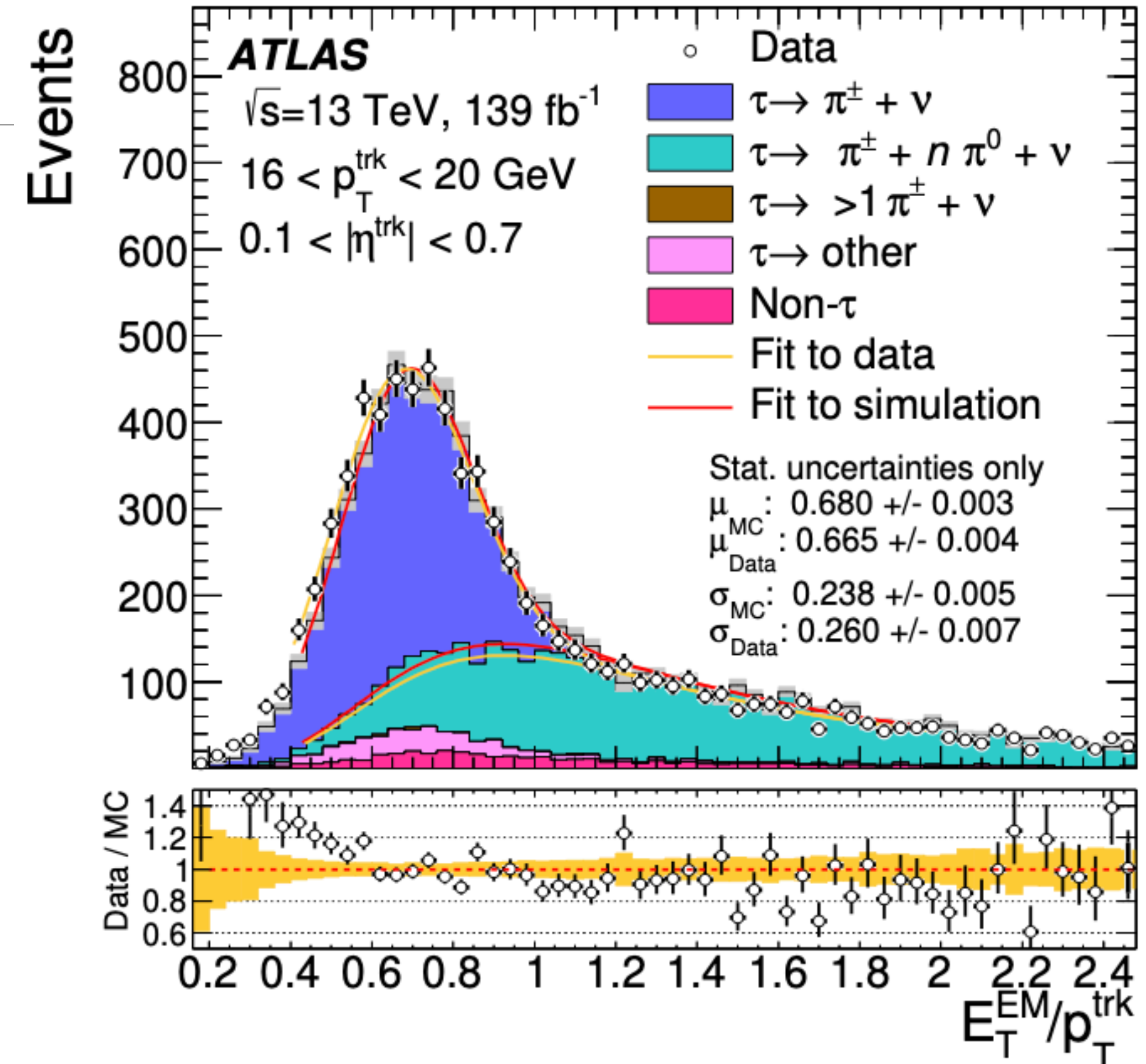
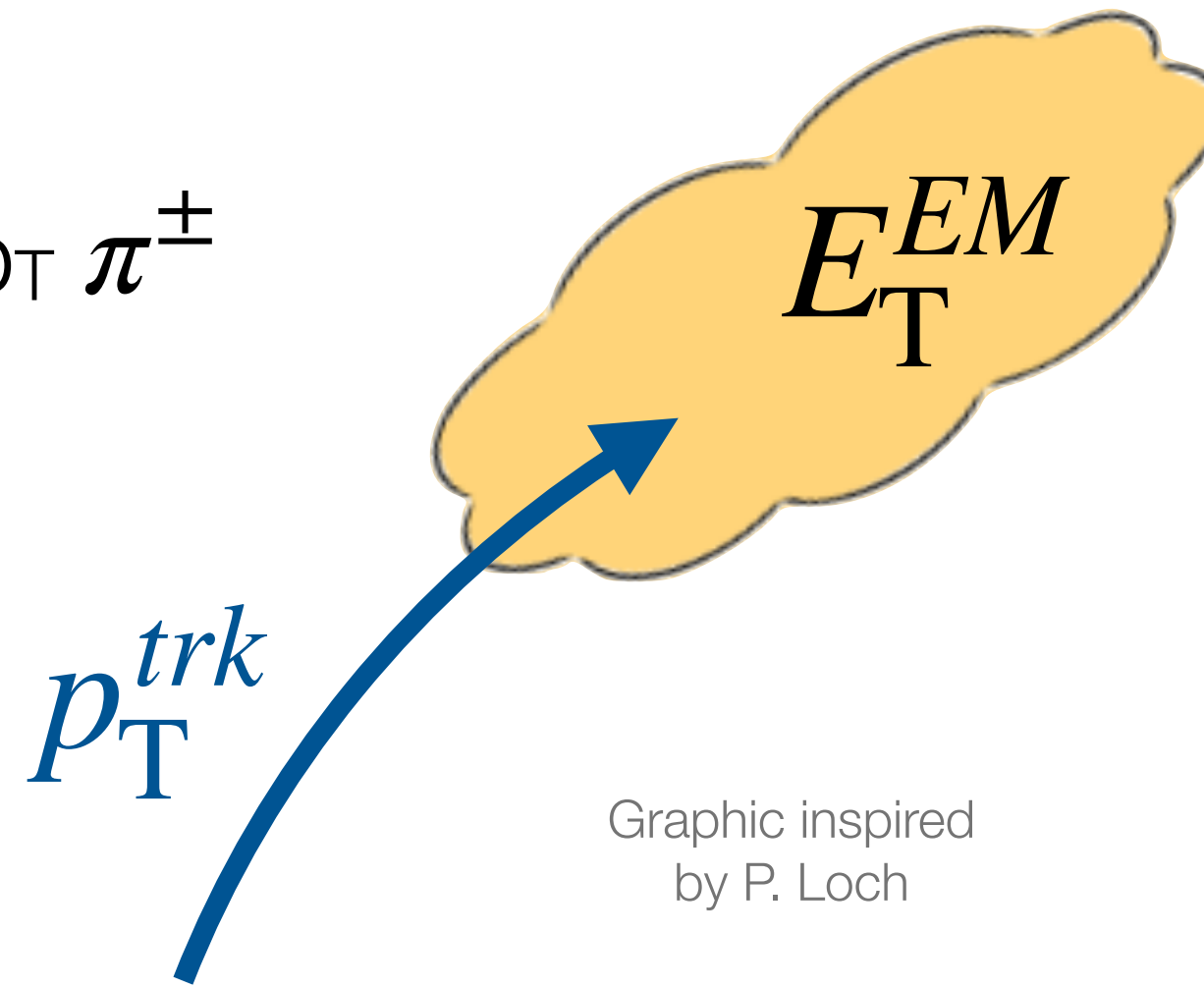
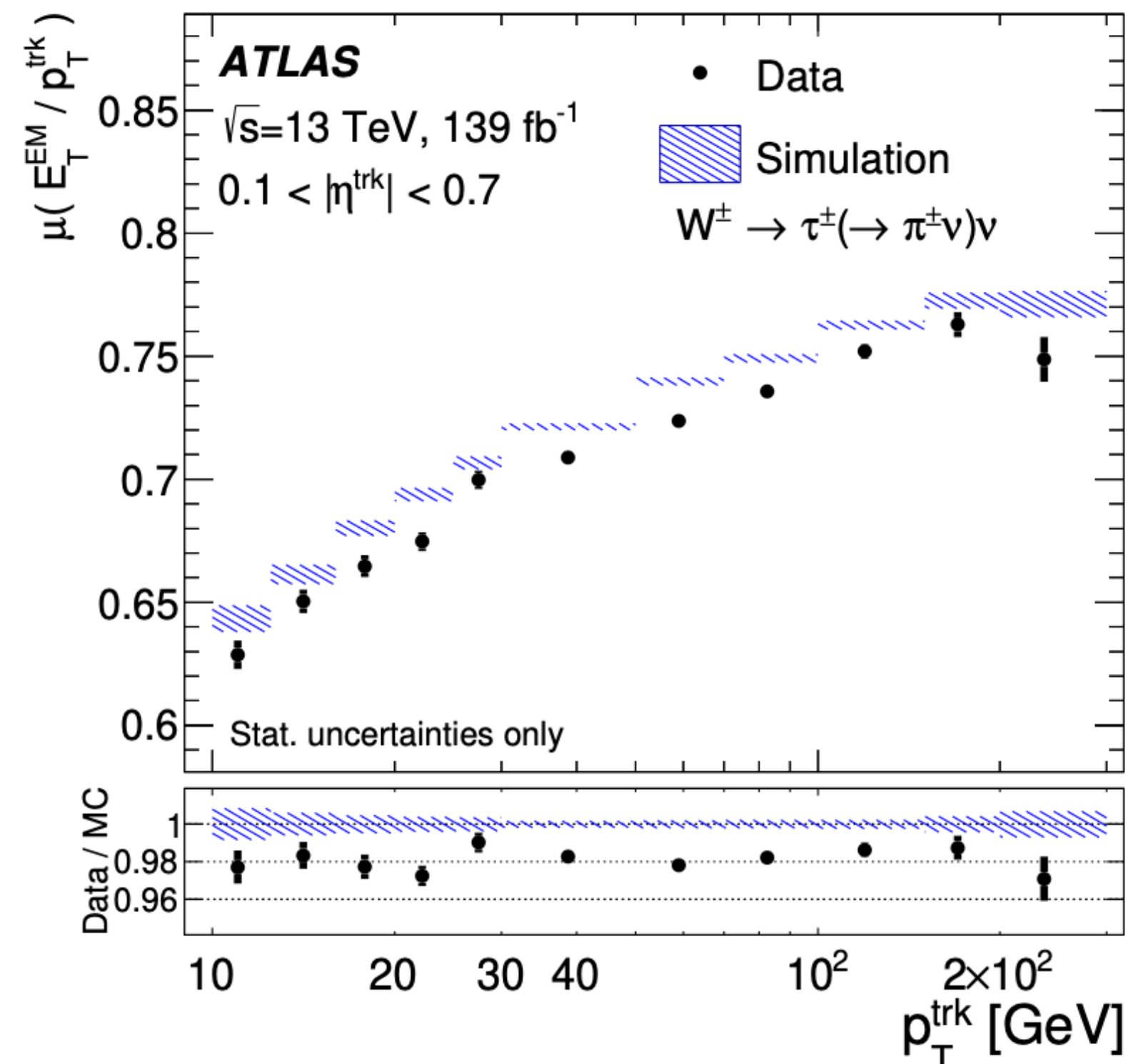
High purity selection for high p_T $\pi^{\pm}$



What about calibrated inputs?

$$W^{\pm} \rightarrow \tau^{\pm}(\rightarrow \pi^{\pm}\nu_{\tau})\nu_{\tau}$$

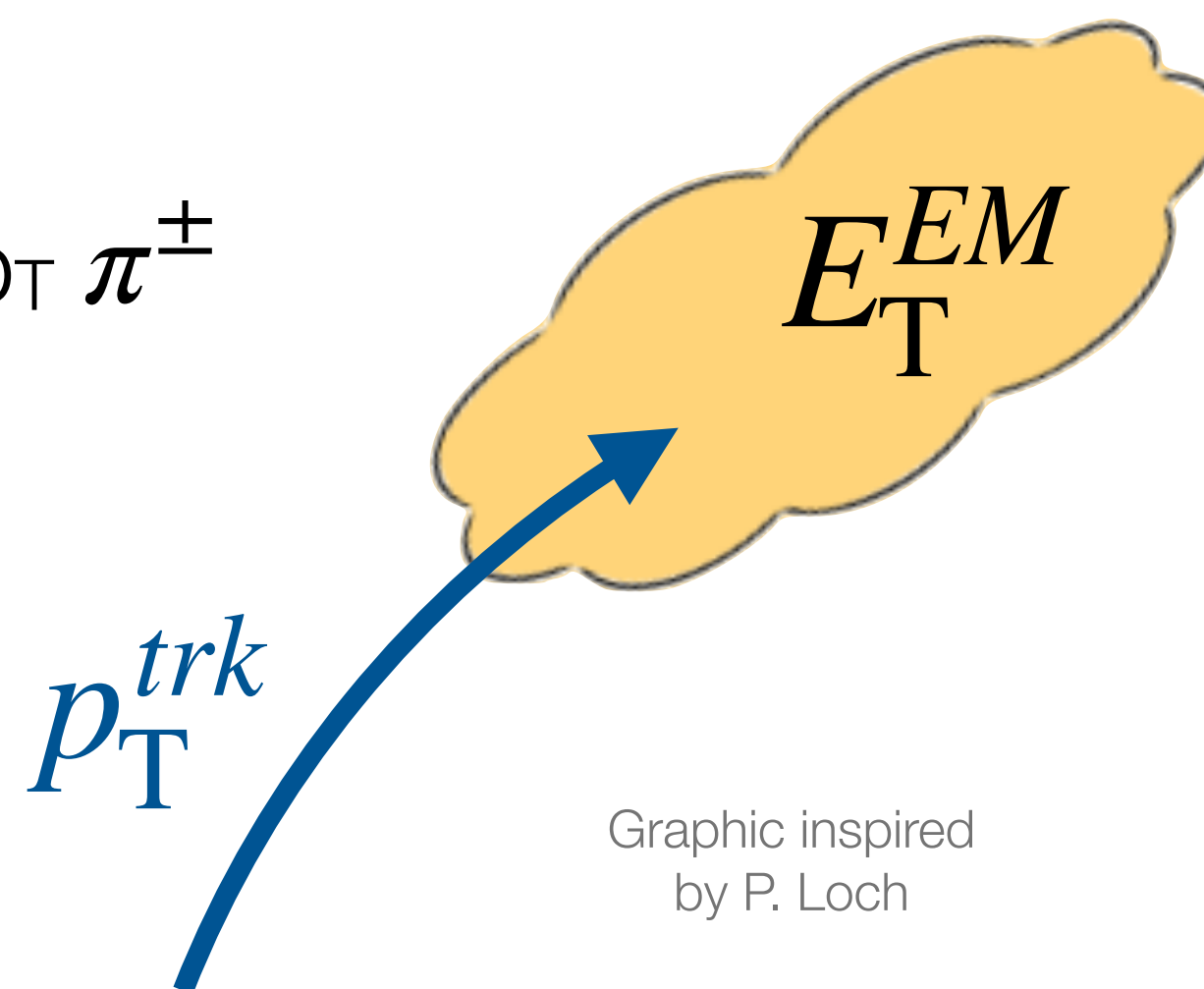
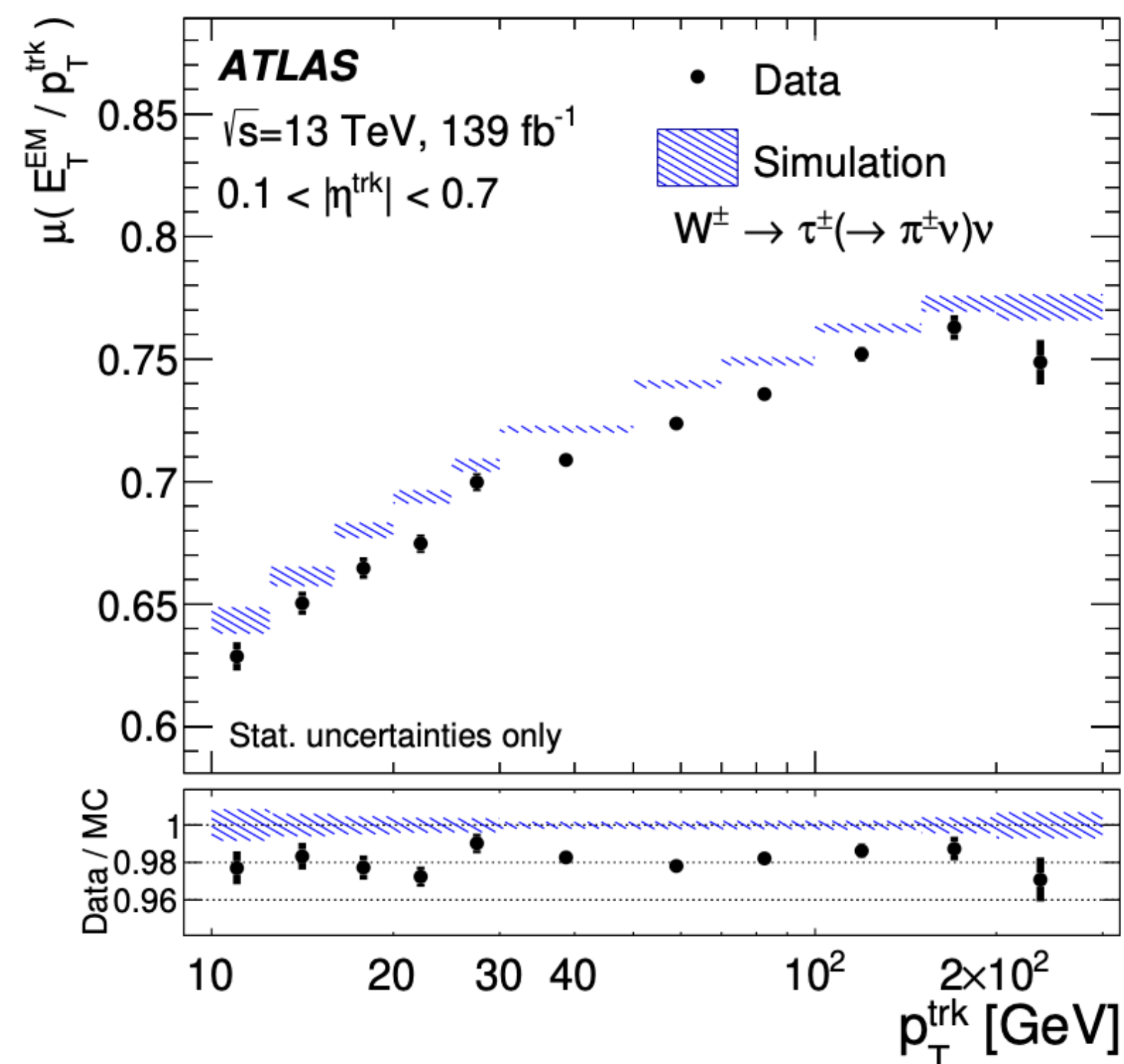
High purity selection for high p_T $\pi^{\pm}$



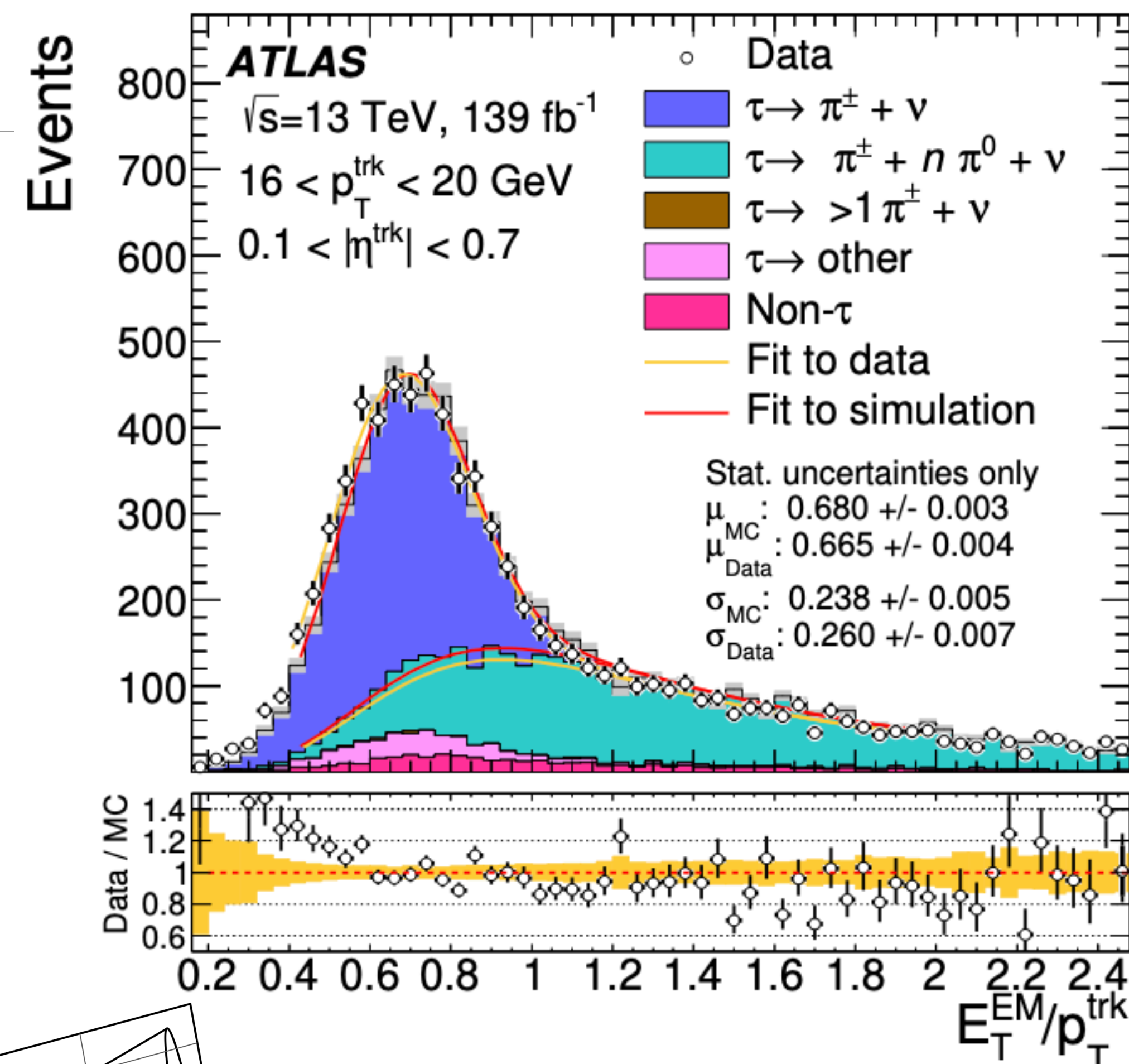
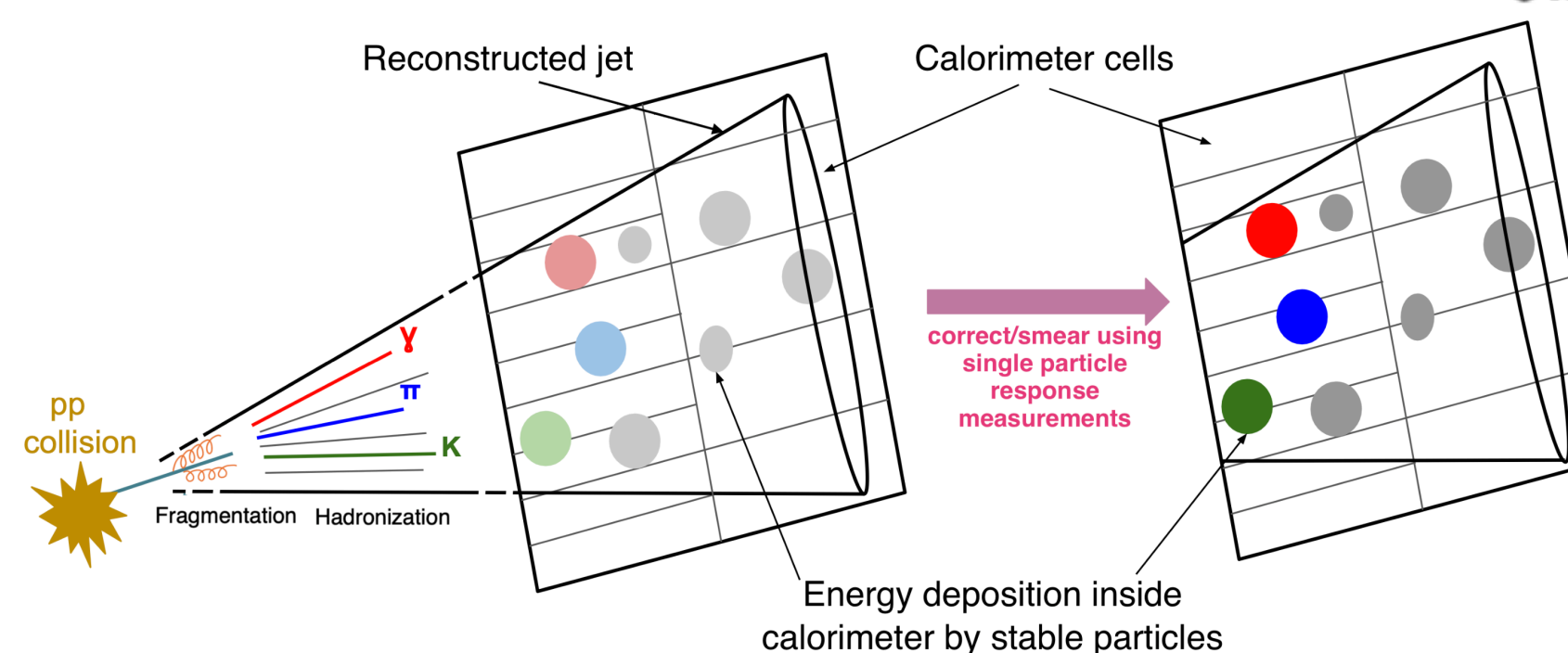
What about calibrated inputs?

$$W^\pm \rightarrow \tau^\pm (\rightarrow \pi^\pm \nu_\tau) \nu_\tau$$

High purity selection for high p_T $\pi^\pm$



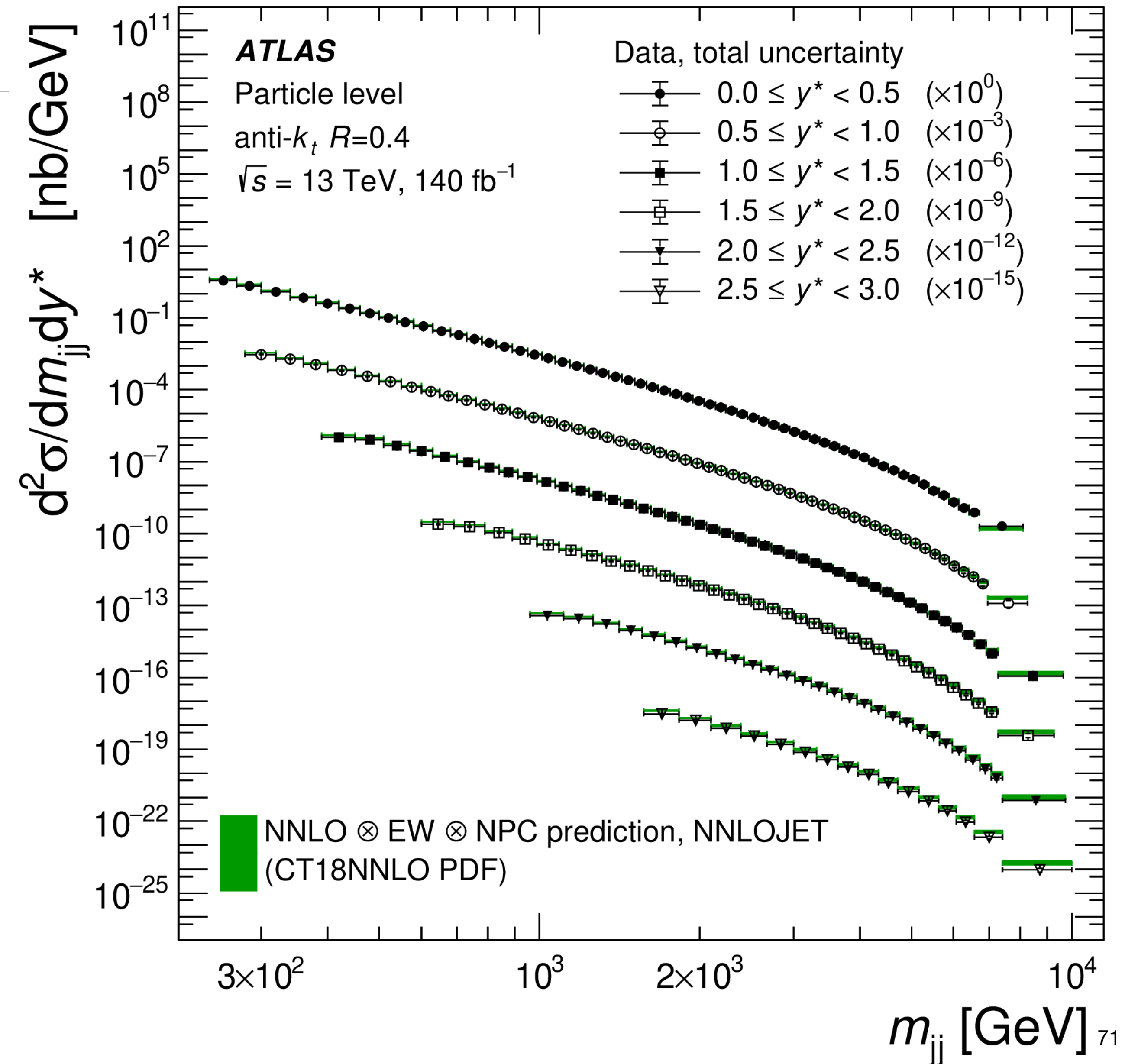
Graphic inspired by P. Loch



Then can derive a bottom-up uncertainty by correcting & smearing the inputs.

Impact: dijet analyses (!)

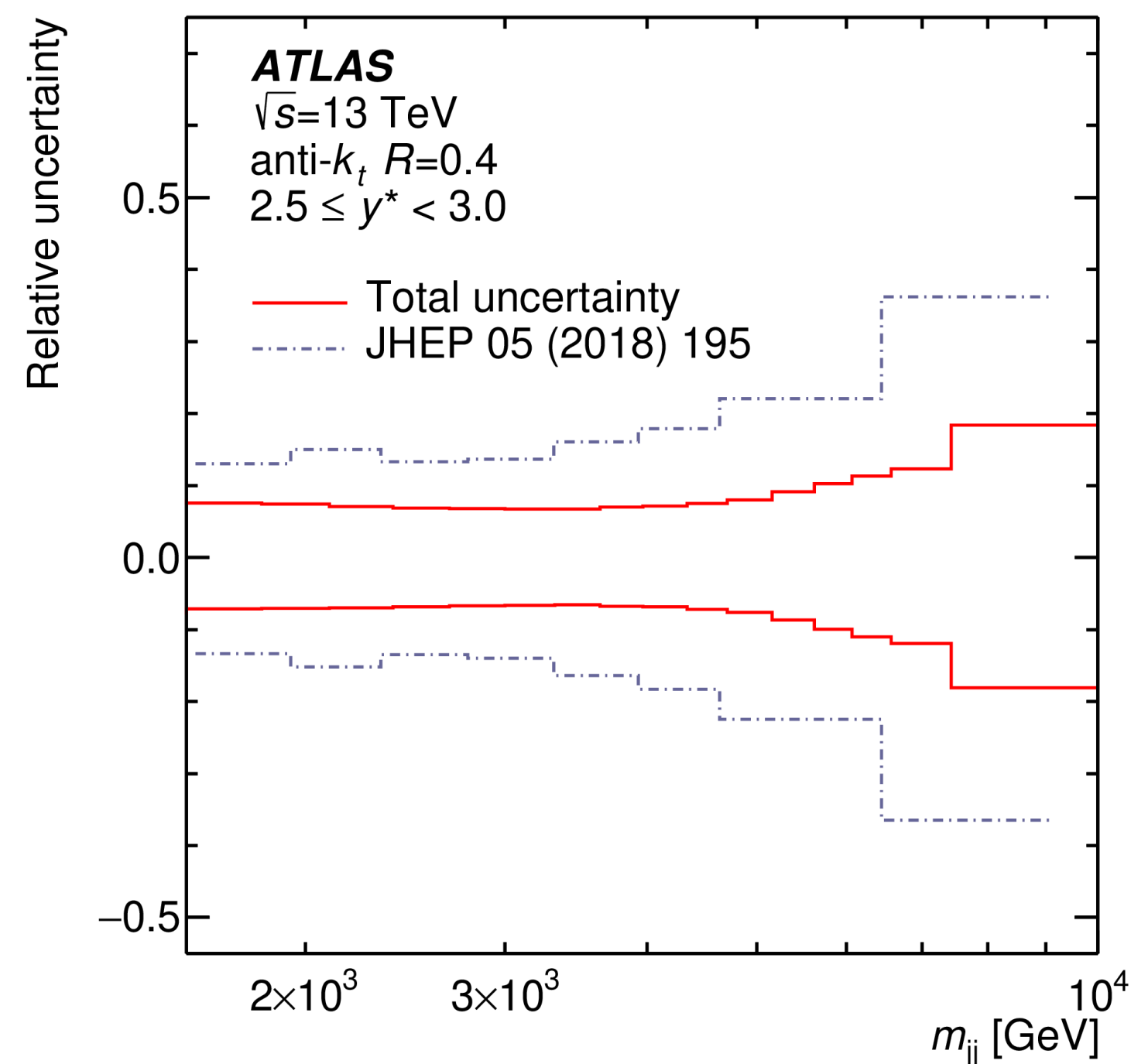
JES uncertainties = leading systematic for dijet measurements.



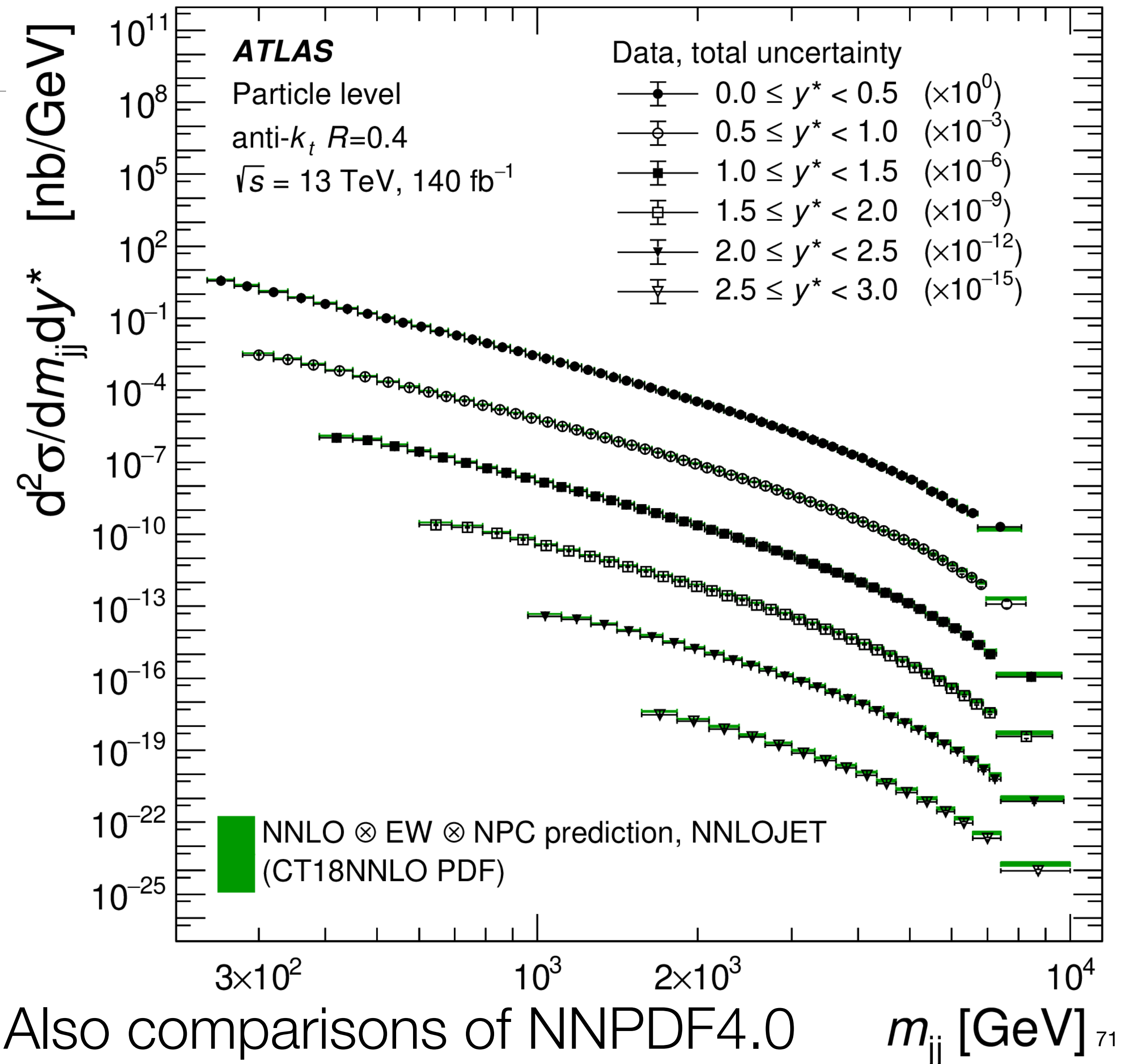
Impact: dijet analyses (!)

JES uncertainties = leading systematic for dijet measurements.

↪ latest ATLAS dijets measurement includes $\pi^\pm$ response uncertainty improvement.



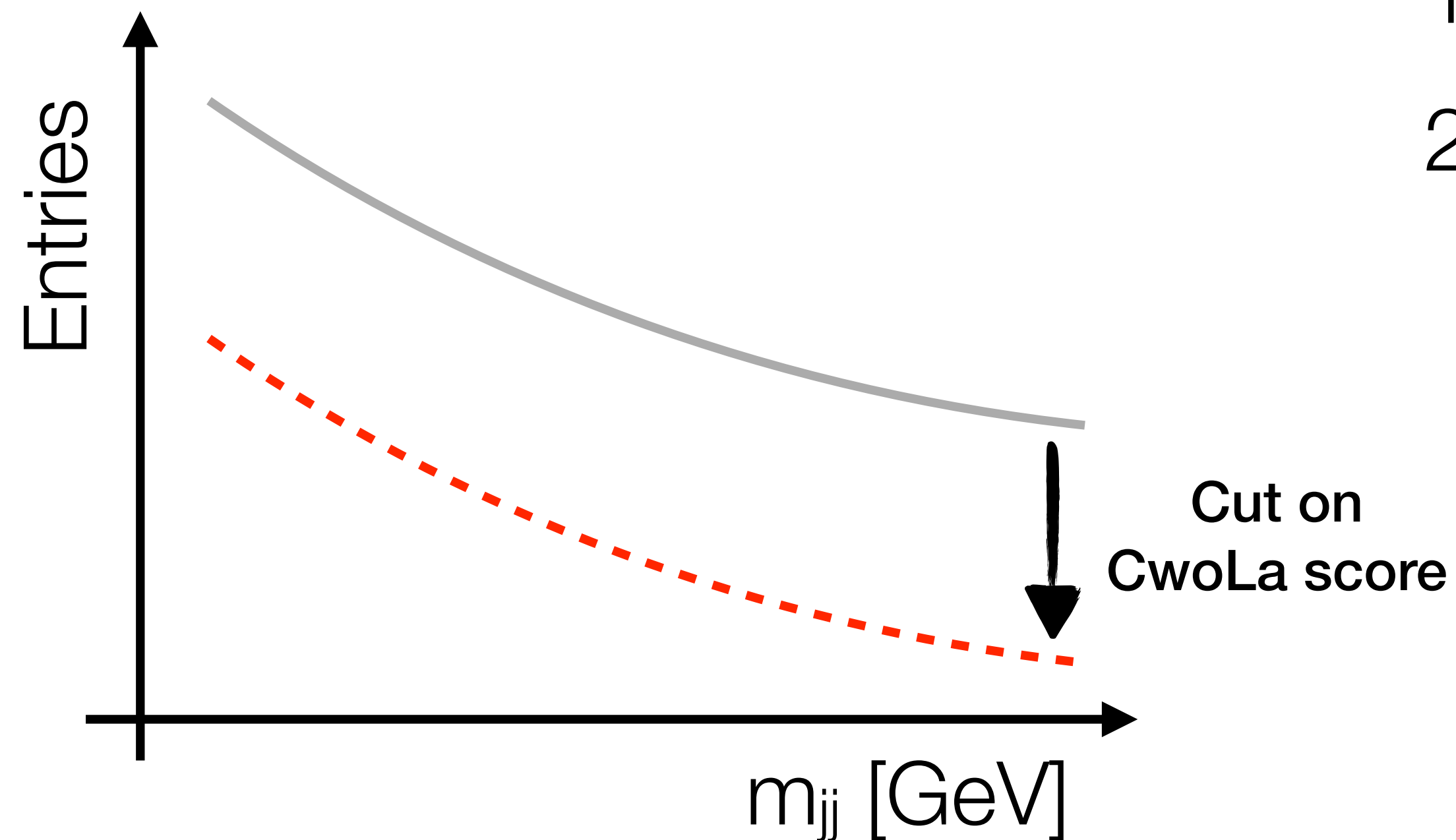
Improves the uncertainty by 2-3x! 🎉



Also comparisons of NNPDF4.0

m_{jj} [GeV] 71

Anomaly detection (further details)



- 1) Define a SR (blinding strategy)
- 2) Train a bkg est
 - (Vanilla) CWoLa: m_{jj} sidebands (SB)
 - TNT: m_{jj} sidebands + “normal” looking jets (as defined by jet level AE)
 - CATHODE: NF conditioned on m_{jj} SB
 - CURTAINS: NF trained w/ OT loss to map SB to SR
 - FETA: NF to map sim to data
 - SALAD: NN reweighting to map sim to data
- 3) Train **data** vs bkg estimate classifier (CWoLa, **C**lassification **W**/o **L**abels)
- 4) Cut on this CWoLa score ($\downarrow$ qcd bkg)
- 5) Fit the m_{jj} spectrum as usual