Snigdho Chakraborty

Karolos Potamianos

University of Warwick, UK

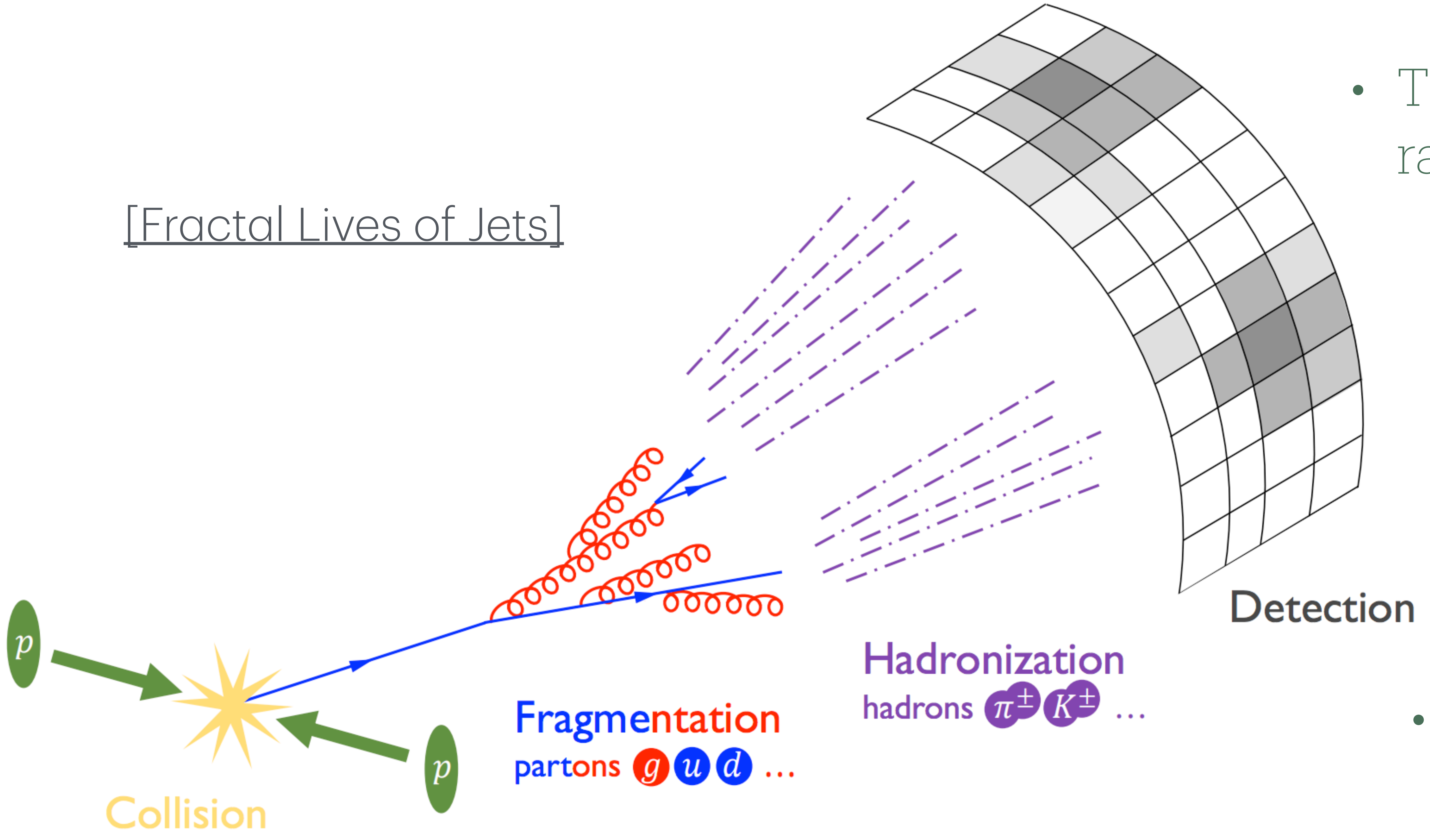
ML based Jet Calibration in HL-LHC

IOP - 2026, Edinburgh

April 8th 2026

What are jets?

[Fractal Lives of Jets]



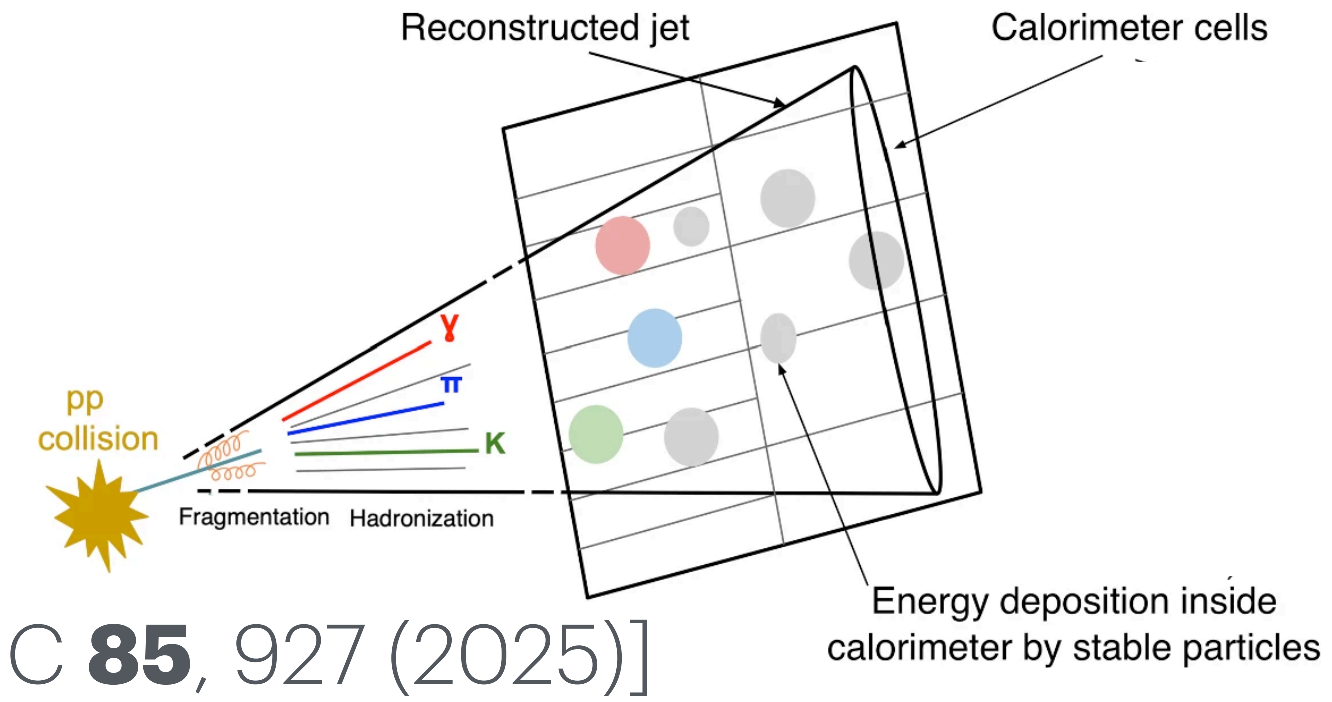
- The quarks and gluons produced in p-p collision radiate, or fragment based on laws of perturbative QCD

- When their energy reach below Λ_{QCD}, non-perturbative QCD kicks in and due to confinement, these partons hadronize to form hadrons

- We cluster these hadrons detected in a narrow section of the detector. These are what we call jets.

- But our detectors are not perfect!

Hence these jets that are clustered need to be calibrated!

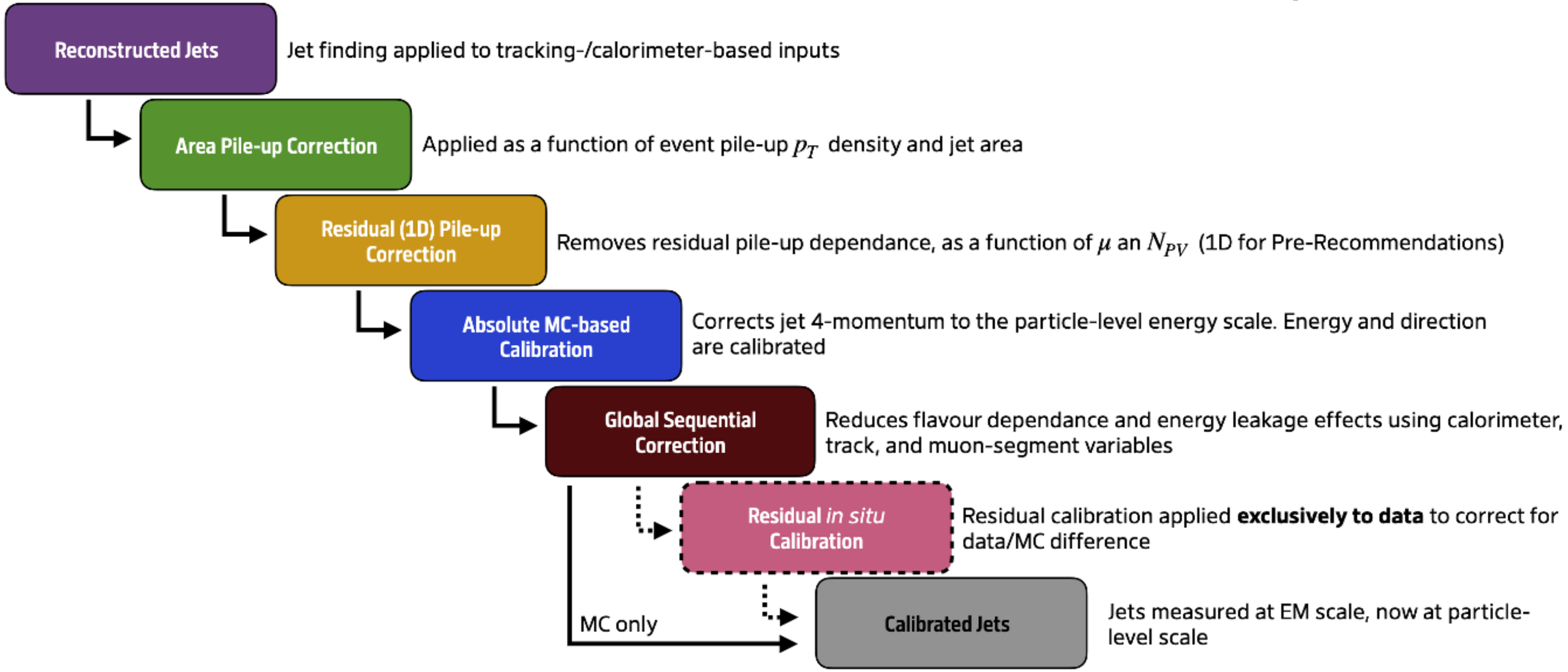


[Eur. Phys. J. C **85**, 927 (2025)]

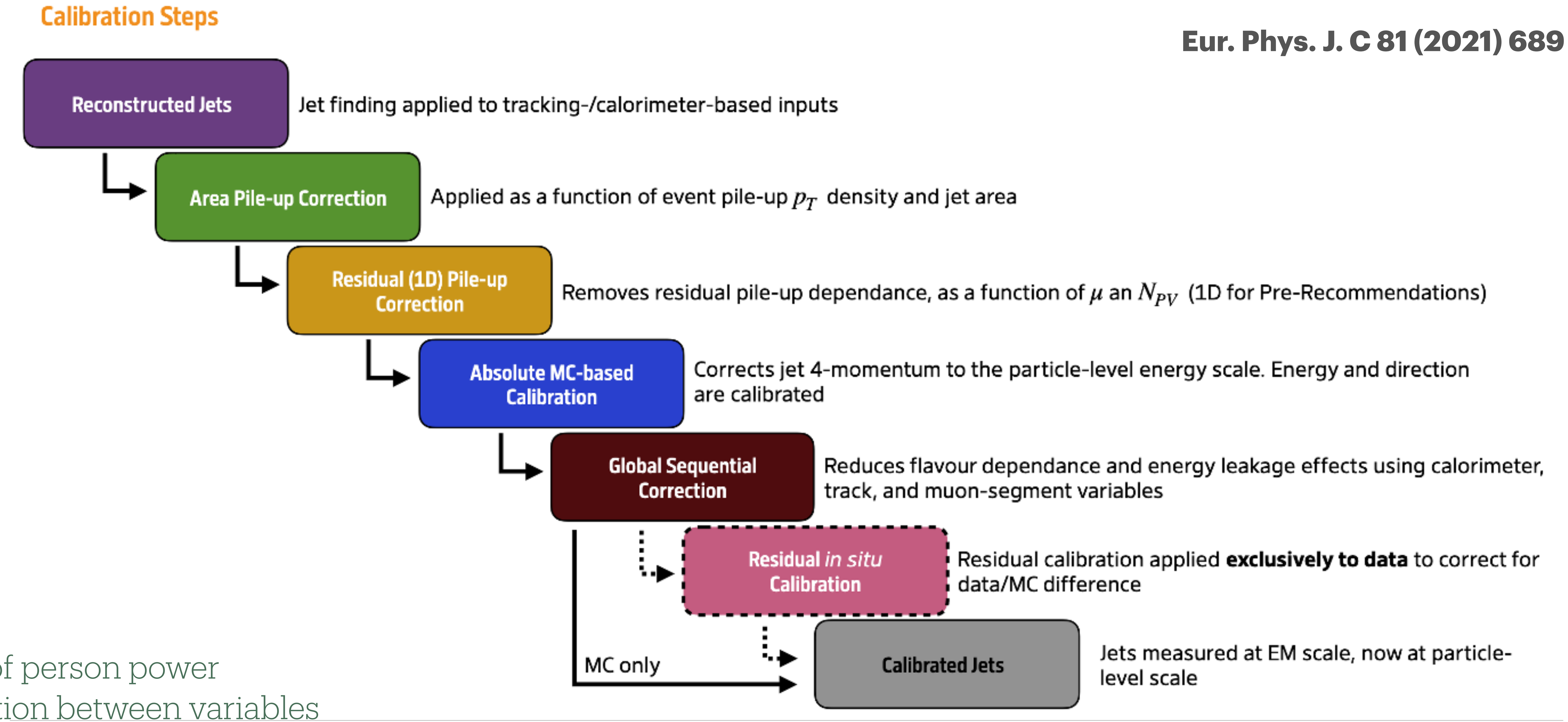
Current Jet Calibration Scheme

Calibration Steps

Eur. Phys. J. C 81 (2021) 689



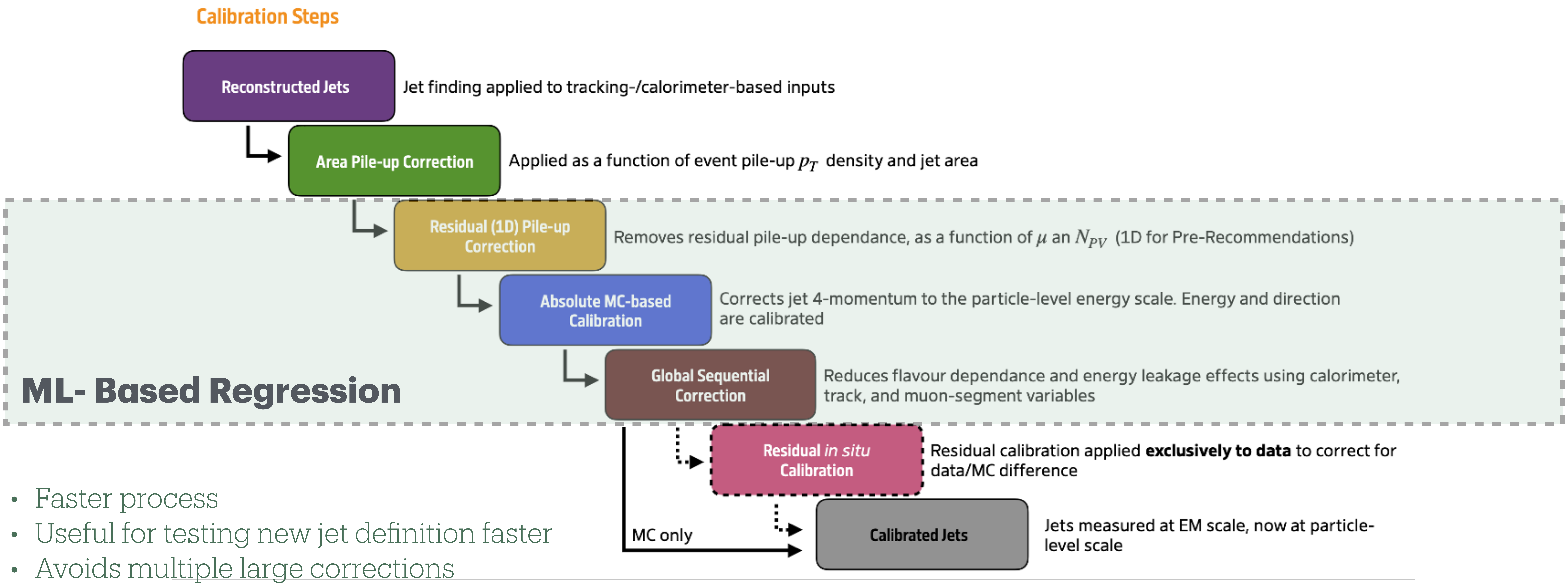
Current Jet Calibration Scheme



Eur. Phys. J. C 81 (2021) 689

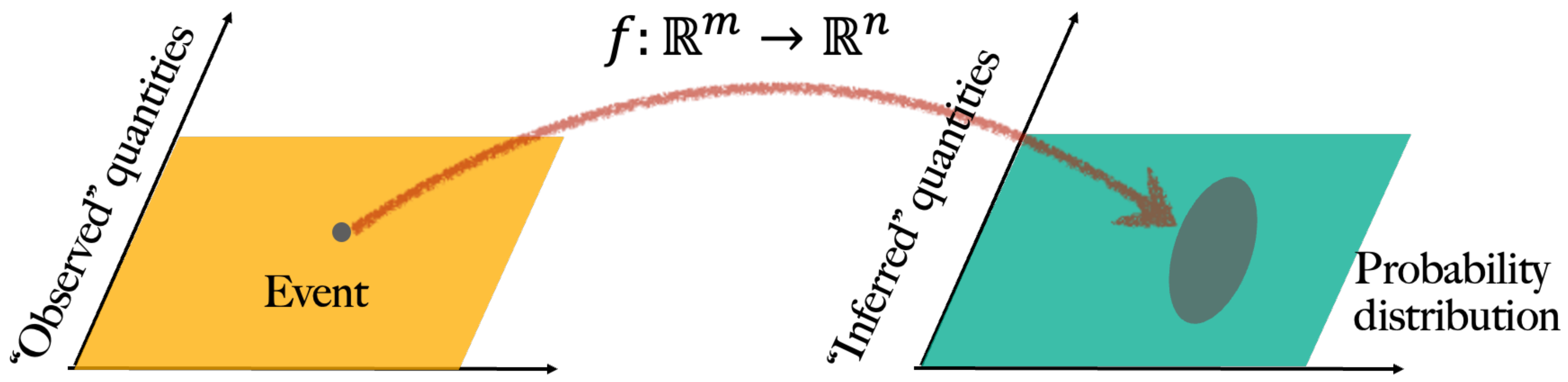
- Time taking
- Requires a lot of person power
- Loss of correlation between variables
- Introduction of artificial constraints.

The New Scheme



- Faster process
- Useful for testing new jet definition faster
- Avoids multiple large corrections

Main Idea



- From Bayes' Theorem, we have:

$$p(\vec{z}_{true} \mid \vec{x}_{measured}) \propto p(\vec{x}_{measured} \mid \vec{z}_{true}) \cdot p(\vec{z}_{true})$$

Posterior **Likelihood** **Prior**

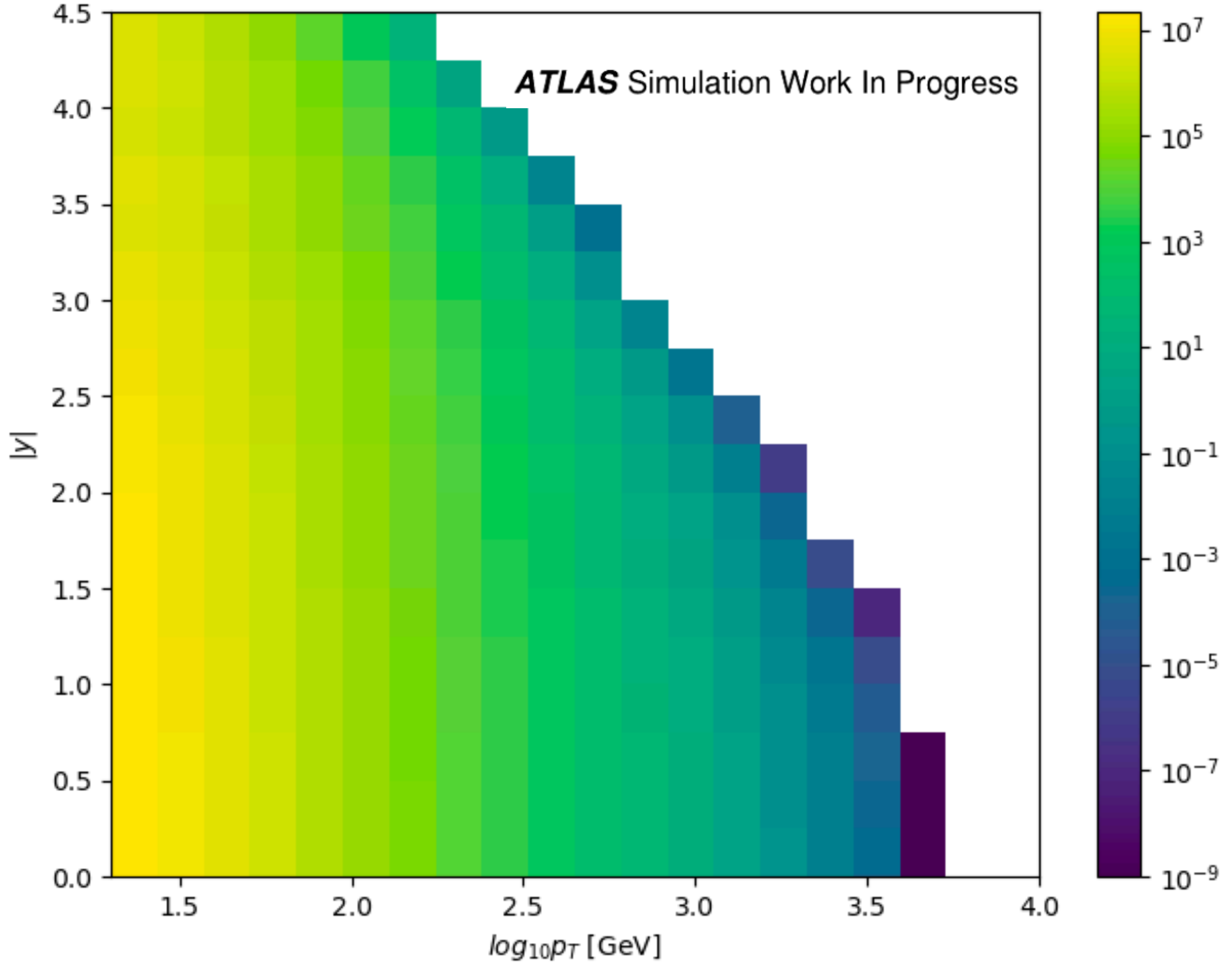
- We try to estimate the posterior density from the detector response

$p(\vec{x}_{measured} \mid \vec{z}_{true})$ ~ Universal across training and testing samples

BUT

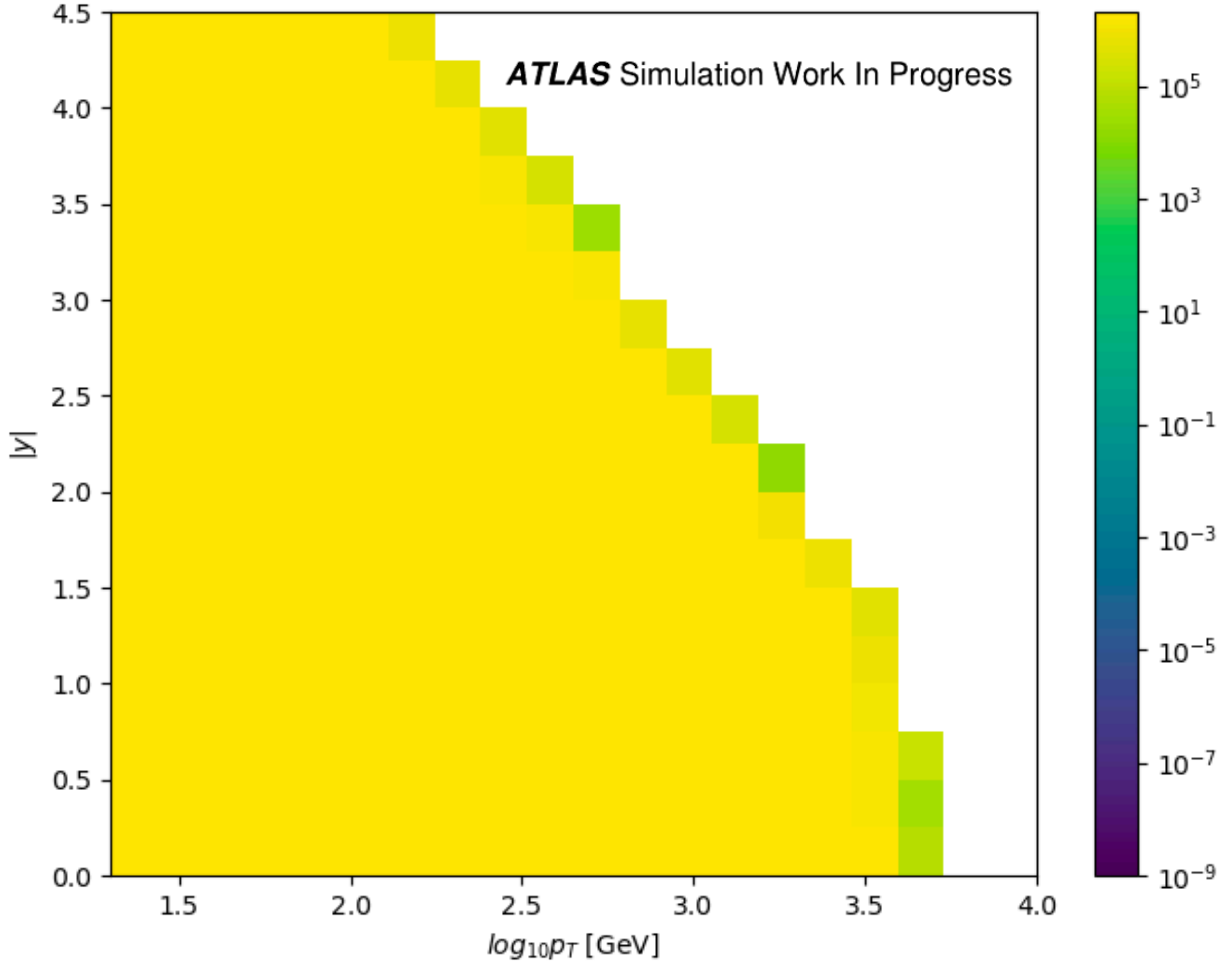
$p(\vec{z}_{true})$ **Prior might not be**

Prior Flattening

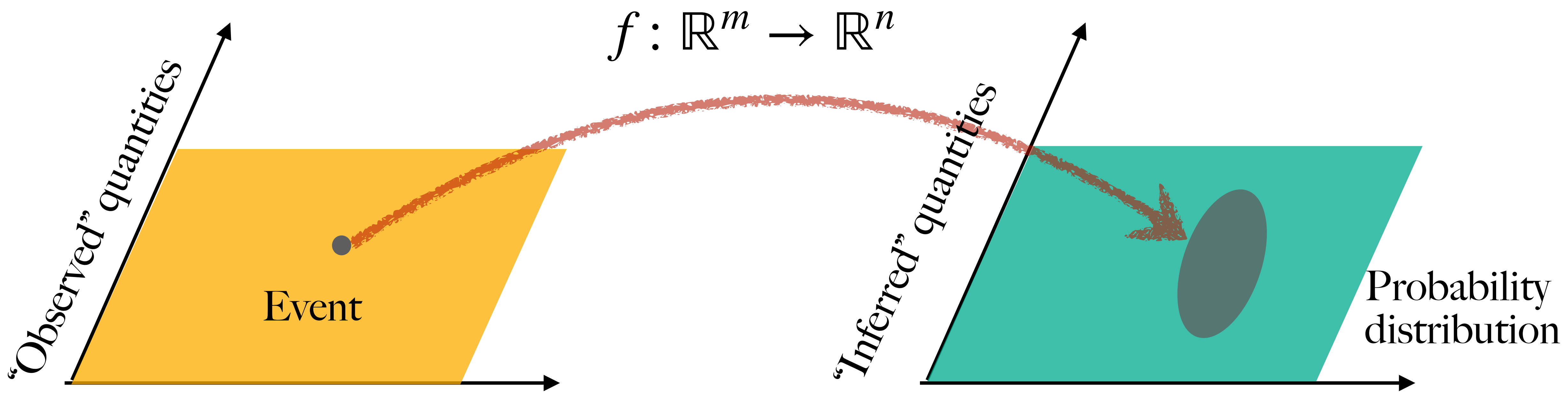


Boundary defined by yield $> 10^{-12}$

$\log_{10} p_T \oplus |y|$ **flattening**



The Network



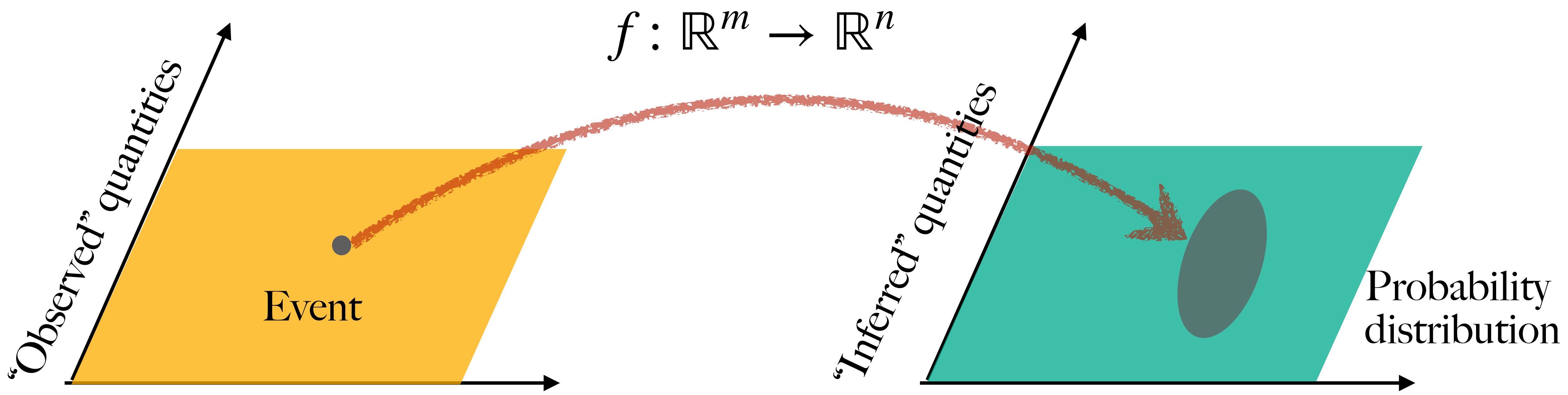
- From Bayes' Theorem, we have:

$$p(\vec{z}_t | \vec{x}_r) = \frac{p(\vec{x}_r | \vec{z}_t) \cdot p(\vec{z}_t)}{\int p(\vec{x}_r | \vec{z'}_t) \cdot p(\vec{z'}_t) \cdot d\vec{z'}_t}$$

Posterior

Intractable

The Network



- From Bayes' Theorem, we have:

$$\underbrace{p(\vec{z}_t \mid \vec{x}_r)}_{\text{Posterior}} = \frac{p(\vec{x}_r \mid \vec{z}_t) \cdot p(\vec{z}_t)}{\int p(\vec{x}_r \mid \vec{z'}_t) \cdot p(\vec{z'}_t) \cdot d\vec{z'}_t}$$

Ansatz: $q_\phi(\vec{z}_t \mid \vec{x}_r) \longrightarrow q_\phi(\vec{z}_t, \vec{x}_r) = \int q_\phi(\vec{z}_t \mid \vec{x}_r) \cdot p(\vec{x}_r \mid \vec{x}_t) \cdot p(\vec{x}_t) \cdot d\vec{x}_t$

- We minimise the 'distance' between $q_\phi(\vec{z}_t, \vec{x}_r)$ and $p(\vec{z}_t, \vec{x}_r)$:

$$d_{KL}[p(\vec{z}_t, \vec{x}_r) \,||\, q_\phi(\vec{z}_t, \vec{x}_r)] = \int_{y \sim p(\vec{z}_t, \vec{x}_r)} p(\vec{z}_t, \vec{x}_r) \cdot \ln \left[\frac{p(\vec{z}_t, \vec{x}_r)}{q_\phi(\vec{z}_t, \vec{x}_r)} \right]$$

$$= \mathbb{E}_{p(\vec{z}_t, \vec{x}_r)} \ln \left[\frac{p(\vec{z}_t, \vec{x}_r)}{q_\phi(\vec{z}_t, \vec{x}_r)} \right]$$

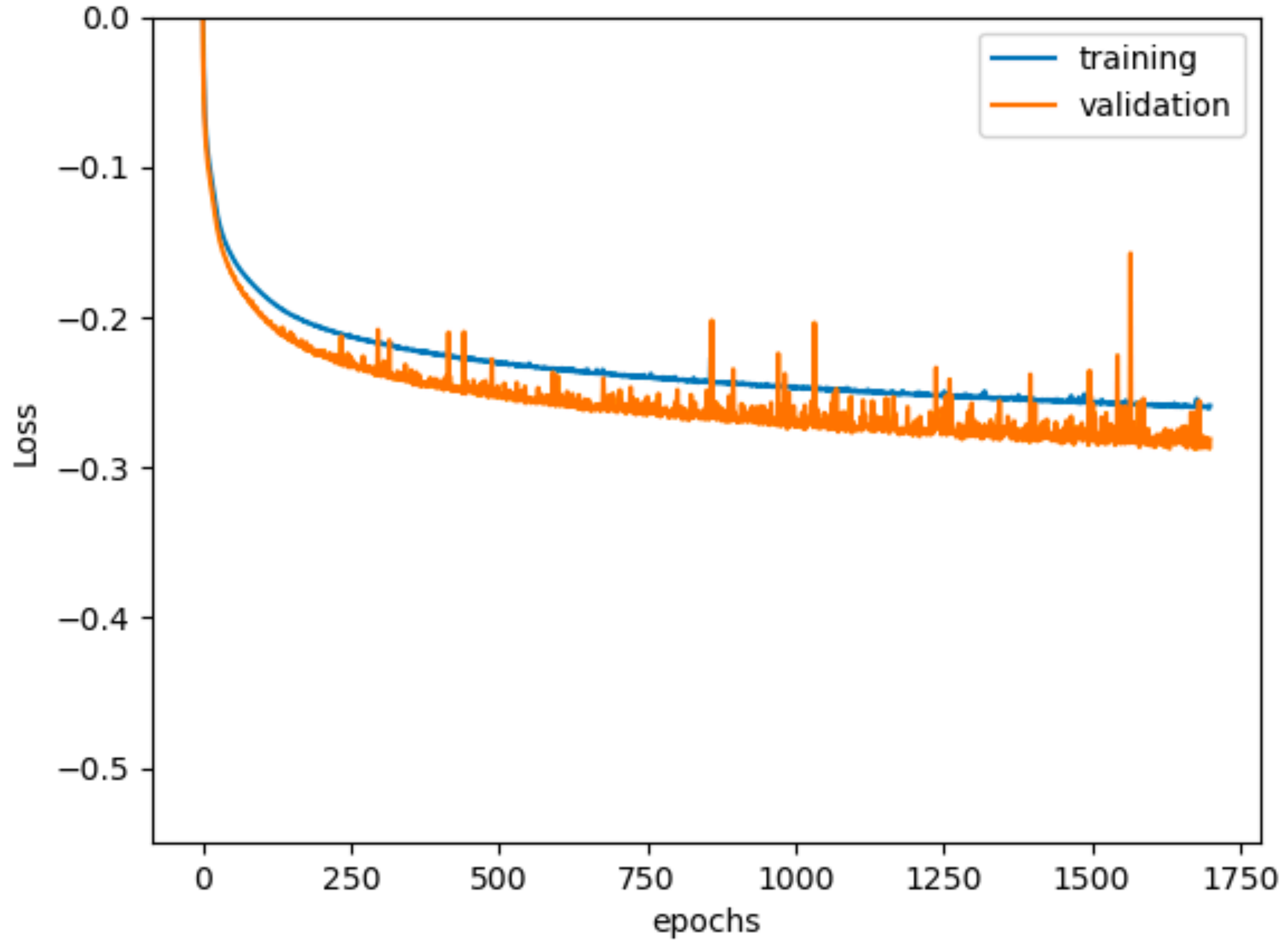
.....

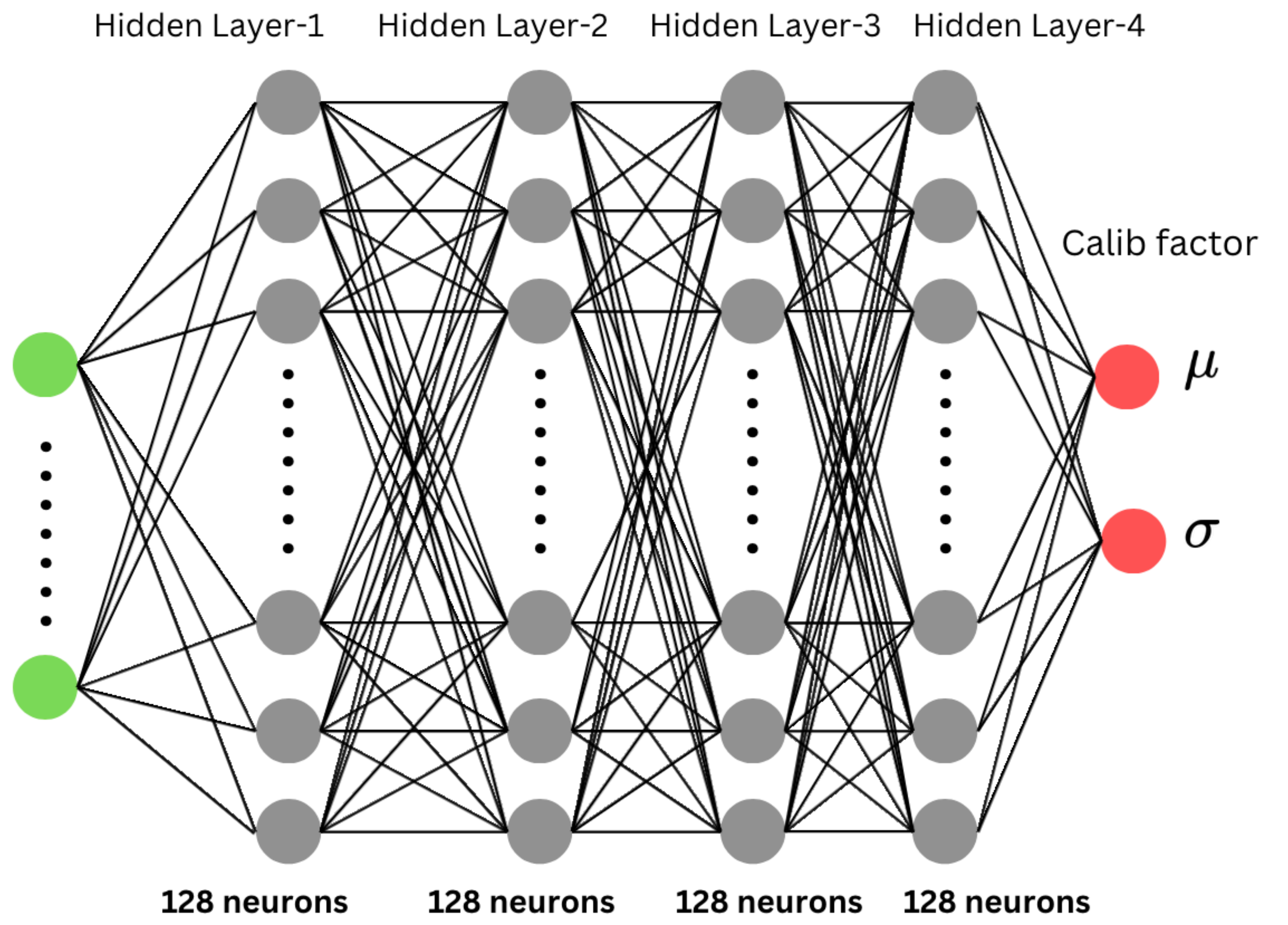
$$\underset{\phi}{argmin}\{d_{KL}[p(\vec{z}_t, \vec{x}_r) \,||\, q_\phi(\vec{z}_t, \vec{x}_r)]\} = \underset{\phi}{argmin}\{\mathbb{E}_{p(\vec{z}_t, \vec{x}_r)}[-\ln(q_\phi(\vec{z}_t \mid \vec{x}_r))]\}$$

- So we can use $-\ln(q_\phi(\vec{z}_t \mid \vec{x}_r))$ as our loss function to train a NN and learn the parameters of our ansatz posterior density.

The Gaussian Density Network

- Used Keras with Tensorflow Backend
  - Network Learns the Calibration factor: $c = \frac{p_T^{true}}{p_T^{reco}}$
    - With a gaussian posterior : $q_\phi(\vec{z}_t | \vec{x}_r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z_t - \mu)^2}{2\sigma}\right]$
    - Learns μ, σ
  - 4 Hidden Layers, each with 128 neurons
  - Adam Optimiser with LR= 10^{-6}





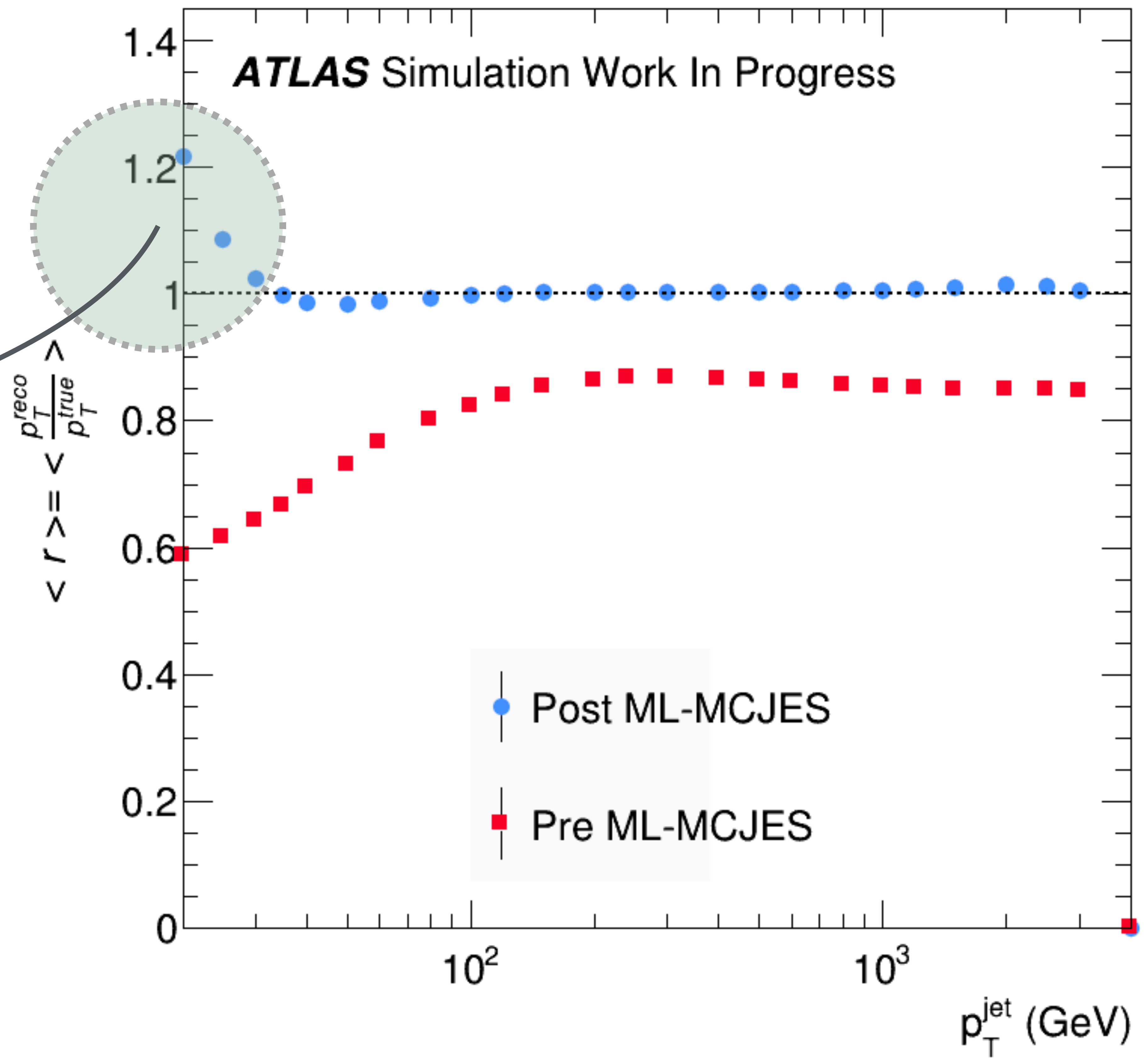
Inputs

Category	Variable	Description
Calorimeter	$N_{90\%}$	The minimum number of clusters containing 90% of the jet energy
Jet kinematics	p_T^*	The jet p_T after the MCJES calibration
	y_{det}	The detector y
Tracking	w_{track}^*	The average p_T-weighted transverse distance in the η-ϕ plane between the jet axis and all tracks of $p_T > 1$ GeV ghost-associated with the jet
	N_{track}^*	The number of tracks with $p_T > 1$ GeV ghost-associated with the jet
	$f_{charged}^*$	The fraction of the jet p_T measured from ghost-associated tracks
Muon segments	$N_{segments}^*$	The number of muon track segments ghost-associated with the jet
Pile-up	μ_{act}	The actual number of interactions per bunch crossing
	N_{PV}	The number of reconstructed primary vertices

Performance of the Network

- We look at the closure in the jet response $r = \frac{p_T^{reco}}{p_T^{truth}}$ in bins of p_T^{truth}
- In ideal case $r = 1$

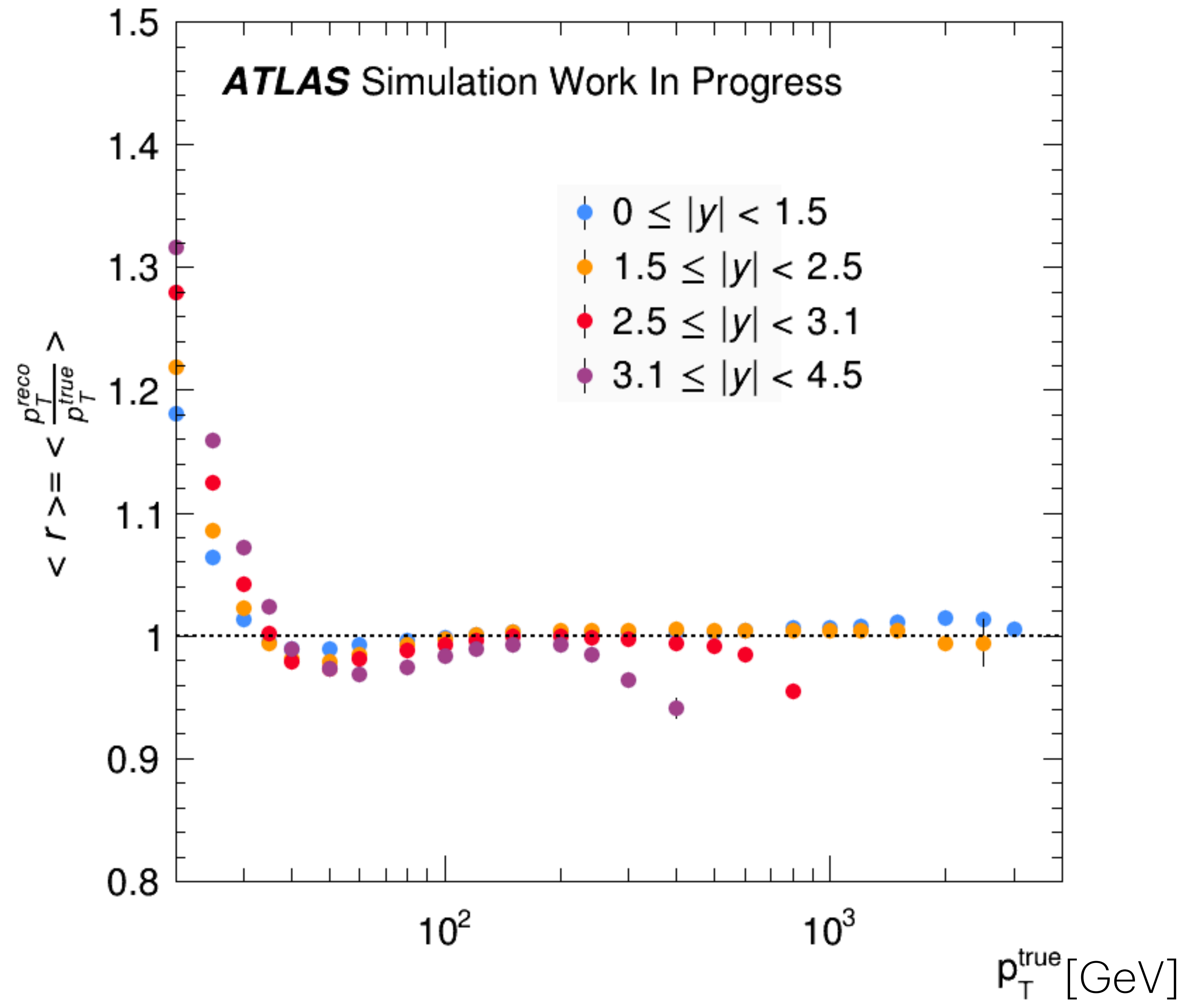
- Average response closes very well in bins of p_T^{truth}
  - Some non-closure in low p_T^{truth} bins



- Why do we have a over-correction in low p_T jets?
  - Explored in slide 18 + 19

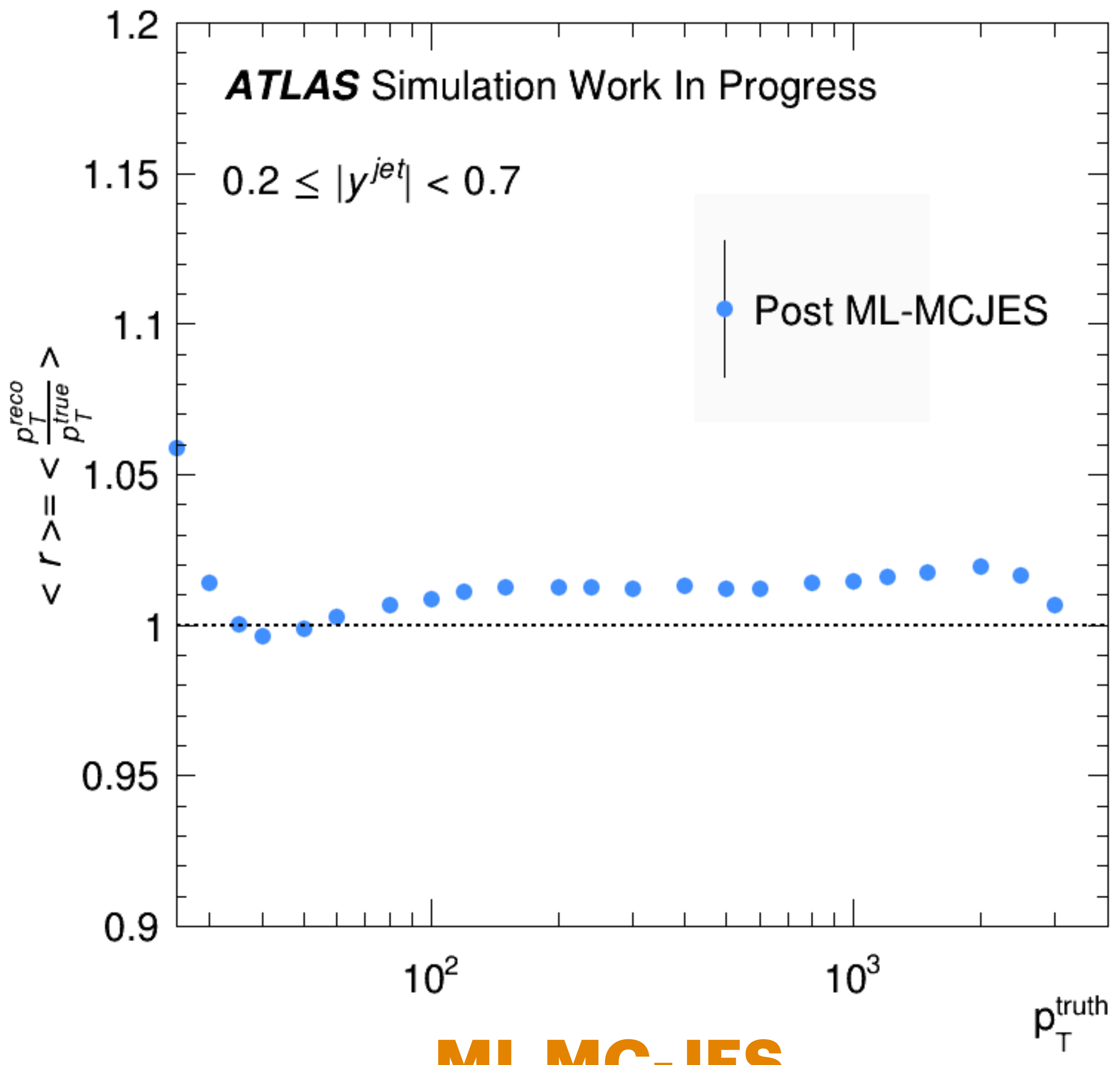
Performance of the Network

Rapidity dependence of calibration



- Slightly more non-closure in high rapidity jets across all p_T bins.

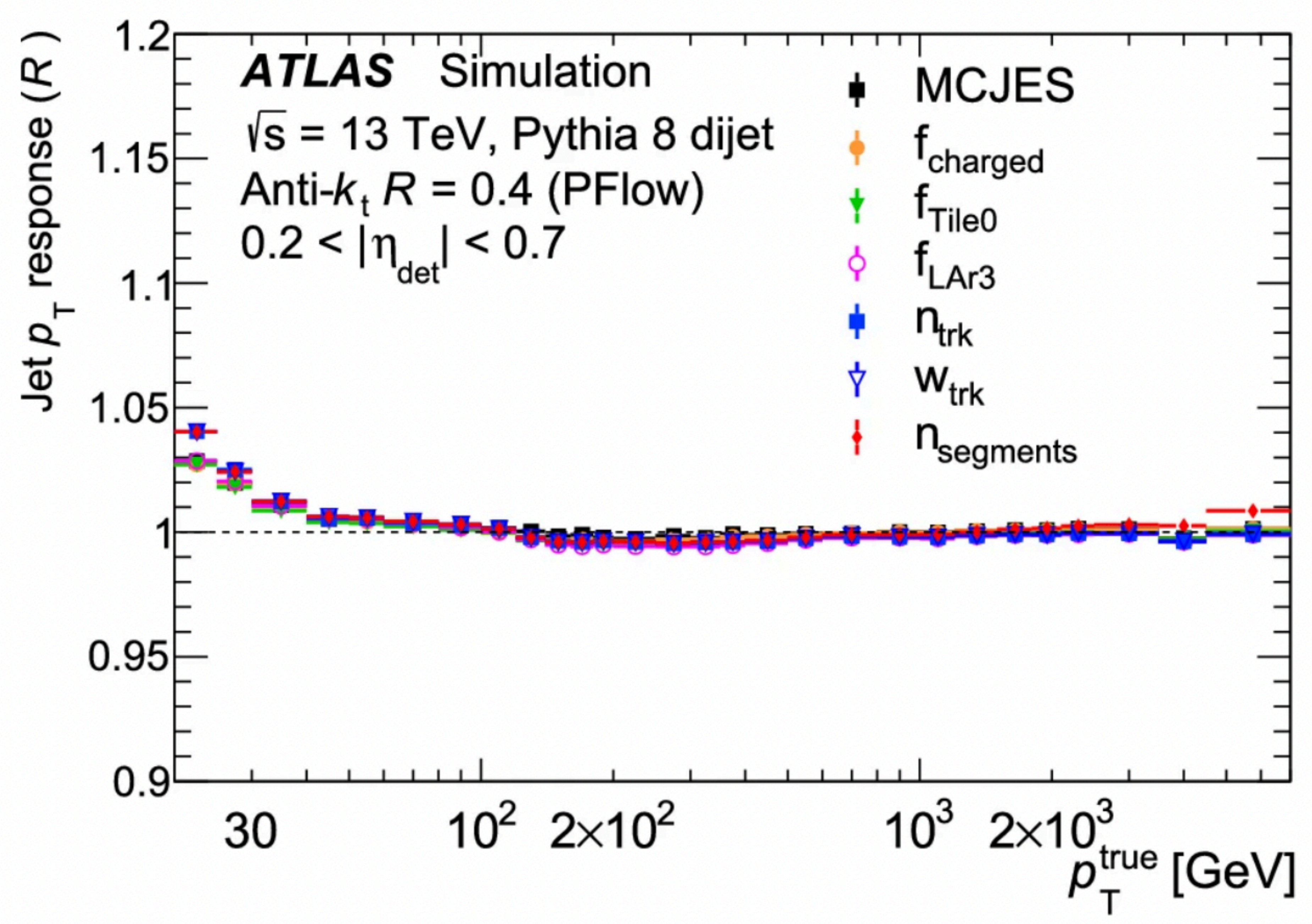
Performance of the Network



ML MC-JES

Comparison to current MC-JES

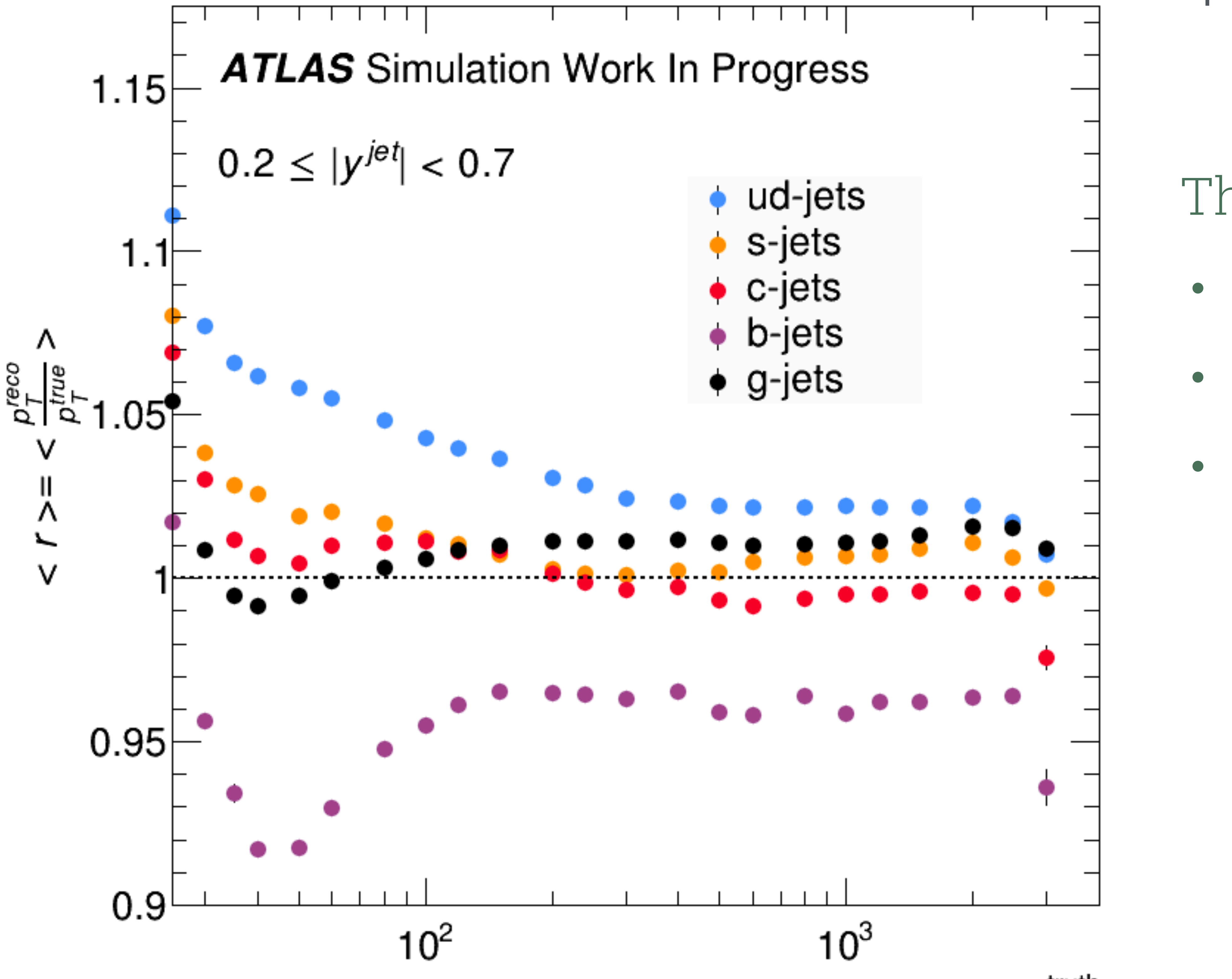
- Remarkable closure for jets with $p_T \geq 25$ Gev in ML-MCJES.
- Closure comparable to the current MC-JES with (upto 5% non-closure)



Current MC-JES: Eur. Phys. J. C (2023) 83:761

Performance of the Network

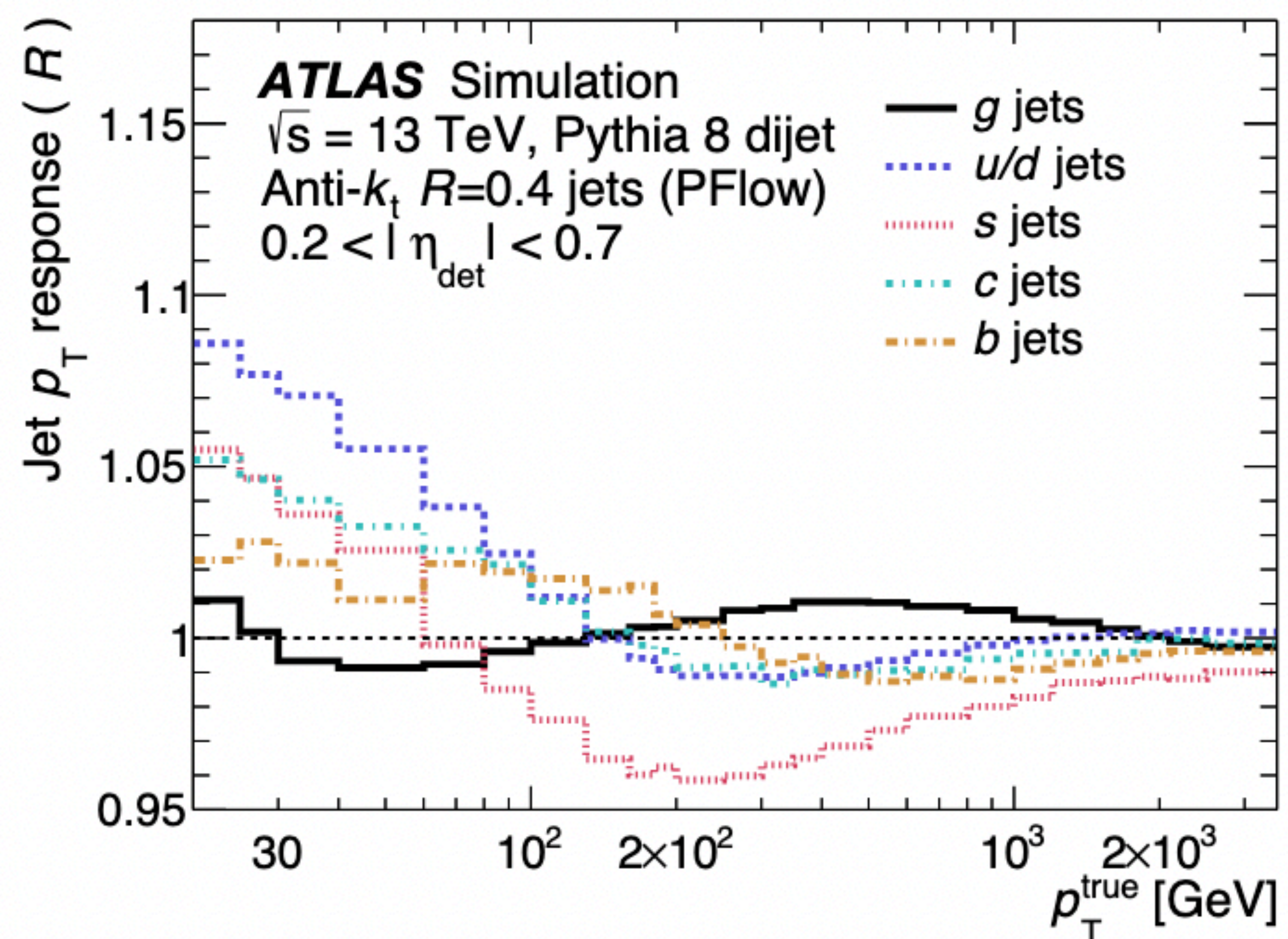
Flavor dependence of calibration



ML MC-JES

There are flavour dependencies on the Average response of the calibrated jets.

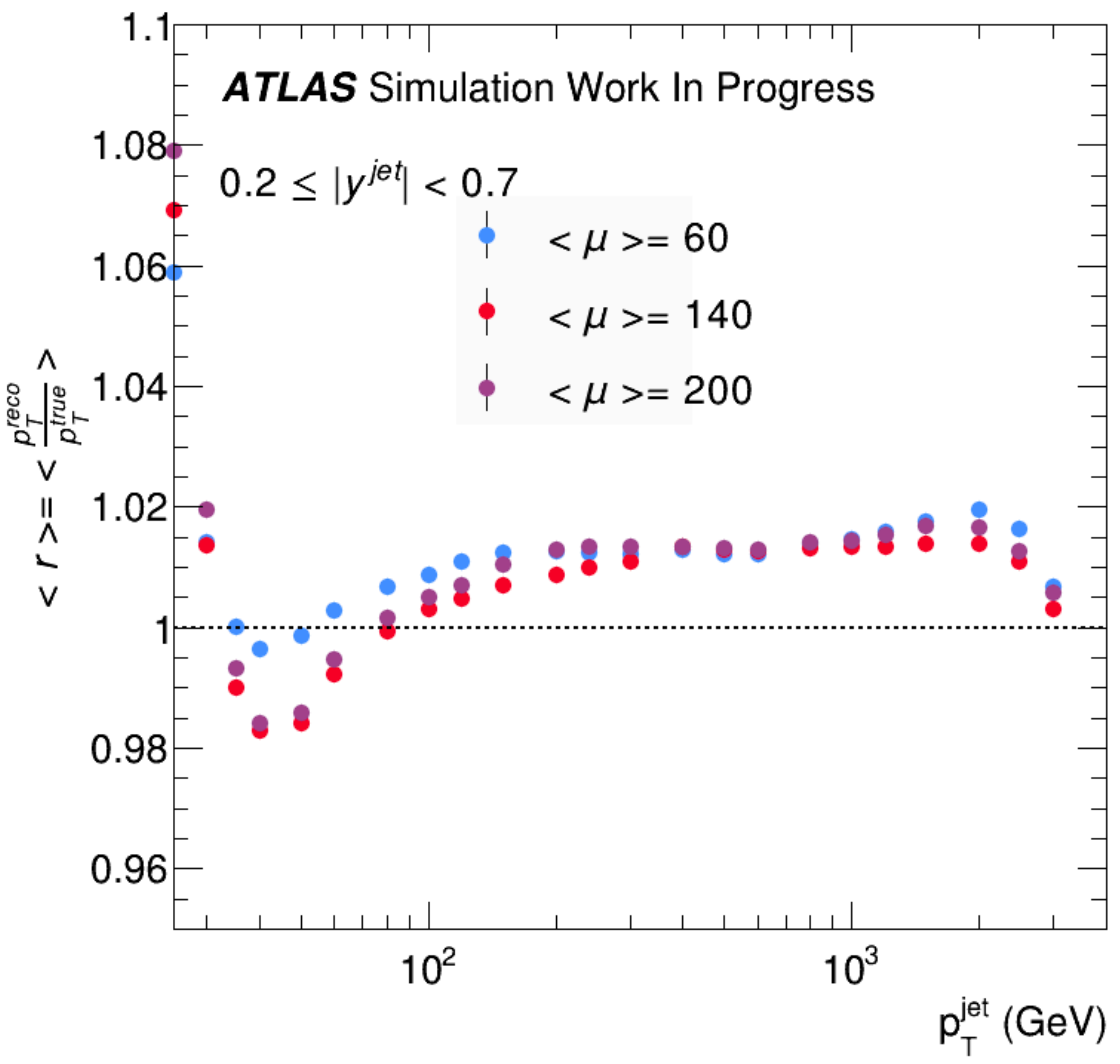
- Gluon jets have better closure that quark jets.
- b-jets have the worst closure of all jet flavours.
- Similar trends seen in current GNNC Calibrations in Run-3 jets.



Current Calibration: **Eur. Phys. J. C (2023) 83:761**

Performance of the Network

Pile-up dependence of ML-MCJES



Most important for the HI-LHC era

- A separate NN was trained for different pileup conditions
- Very good closure (Non-closure < 5%) in most p_T bins.
- Low p_T jets coming from high pileup conditions have worse closure than low pileup ones.
- Pileup dependence is very low.

Conclusions

- Closure is extremely good in jets with $p_T > 25$ GeV. But why do we have non-closures in the low p_T bins?

Conclusions

Response distributions in jet p_T bins

- Non-gaussian tails in the response distribution - shifts the mean response upwards
  - As we go to higher p_T bins, the response looks more and more gaussian - the gaussian ansatz works better!

Conclusions

- Closure is extremely good in jets with $p_T > 25$ GeV.
- We can make use of Mixture Density Networks (MDNs)
    - Modelling the non-gaussian tails of the posterior density in low p_T jets

Conclusions

- Closure is extremely good in jets with $p_T > 25$ GeV.
- We can make use of Mixture Density Networks (MDNs)
  - Modelling the non-gaussian tails of the posterior density in low p_T jets
- Use of calorimeter cluster information as inputs can improve the calibration

Calorimeter	$f_{\mathrm{LAr0-3*}}$	The E_{frac} measured in the 0th–3rd layer of the EM LAr calorimeter
	$f_{\mathrm{Tile0*-2}}$	The E_{frac} measured in the 0th–2nd layer of the hadronic tile calorimeter
	$f_{\mathrm{HEC,0-3}}$	The E_{frac} measured in the 0th–3rd layer of the hadronic end cap calorimeter
	$f_{\mathrm{FCAL,0-2}}$	The E_{frac} measured in the 0th–2nd layer of the forward calorimeter

Can we do better?

- Closure is extremely good in jets with $p_T > 25$ GeV.
- We can make use of Mixture Density Networks (MDNs)
    - Modelling the non-gaussian tails of the posterior density in low p_T jets
- Use of calorimeter cluster information as inputs can improve the calibration

Calorimeter	$f_{\mathrm{LAr0-3*}}$	The E_{frac} measured in the 0th–3rd layer of the EM LAr calorimeter
	$f_{\mathrm{Tile0*-2}}$	The E_{frac} measured in the 0th–2nd layer of the hadronic tile calorimeter
	$f_{\mathrm{HEC,0-3}}$	The E_{frac} measured in the 0th–3rd layer of the hadronic end cap calorimeter
	$f_{\mathrm{FCAL,0-2}}$	The E_{frac} measured in the 0th–2nd layer of the forward calorimeter

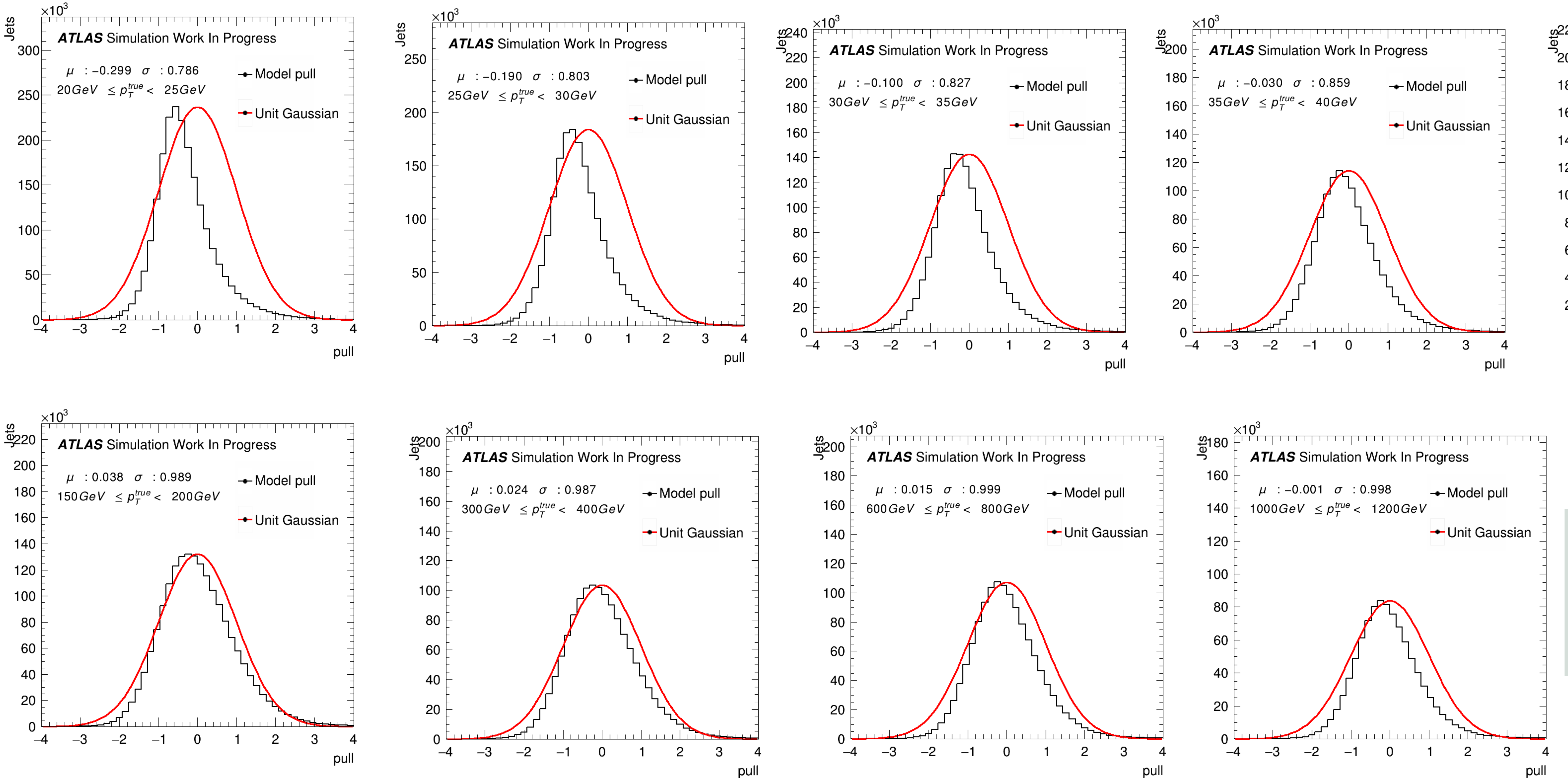
But we are already doing very well!

- Time taken for training the NN for 1700 epochs: 12hrs
- Time taken for inference: 1 min

} Much faster than current jet calibration scheme

Backups

Pull Plots



$$\text{pull} = \frac{p_T^{calib} - p_T^{truth}}{\sigma_{p_T}^{calib}}$$

Snigdho Chakraborty

What is SHAP value?

- **SHAP (SHapley Additive exPlanations)** is a method to explain individual predictions of an ML model. — Based on Llyod Shapley's concept of Shapley value in co-operative games.
- Each input feature is a "player" in our game that tries to produce the model's prediction.
- SHAP value determines how each player contributes to this final prediction.
- SHAP decomposes the prediction as:

$$f_{prediction} = f_{base} + \sum_i \phi_i$$

Average prediction of the model

Contribution of the feature i

What is SHAP value?

- **For a trained network $f(x)$, SHAP defines:**

$$\phi_i = \mathbb{E}_{S \subseteq F \setminus i}[f(x_{S \cup i}) - f(x_S)]$$

- Take a subset of features S.
- Compute model prediction without feature i : $f(x_S)$
- Add the feature i: $f(x_{S \cup i})$
- Measure the change in prediction.
- Average over all subsets.

Game Theory Version

But that would be computationally intensive for large NNs

- SHAP calculates the baseline.

$$f(\text{baseline})$$

- For a given jet total prediction shift is calculated

$$\Delta f = f(x) - f(\text{baseline})$$

- SHAP propagates contribution score through the network:
  - At each neuron:
    - $\Delta\text{output} = \sum \text{contribution from each input}$
- This is done layer by layer.
- Finally SHAP assigns ϕ_i of each feature i to Δf

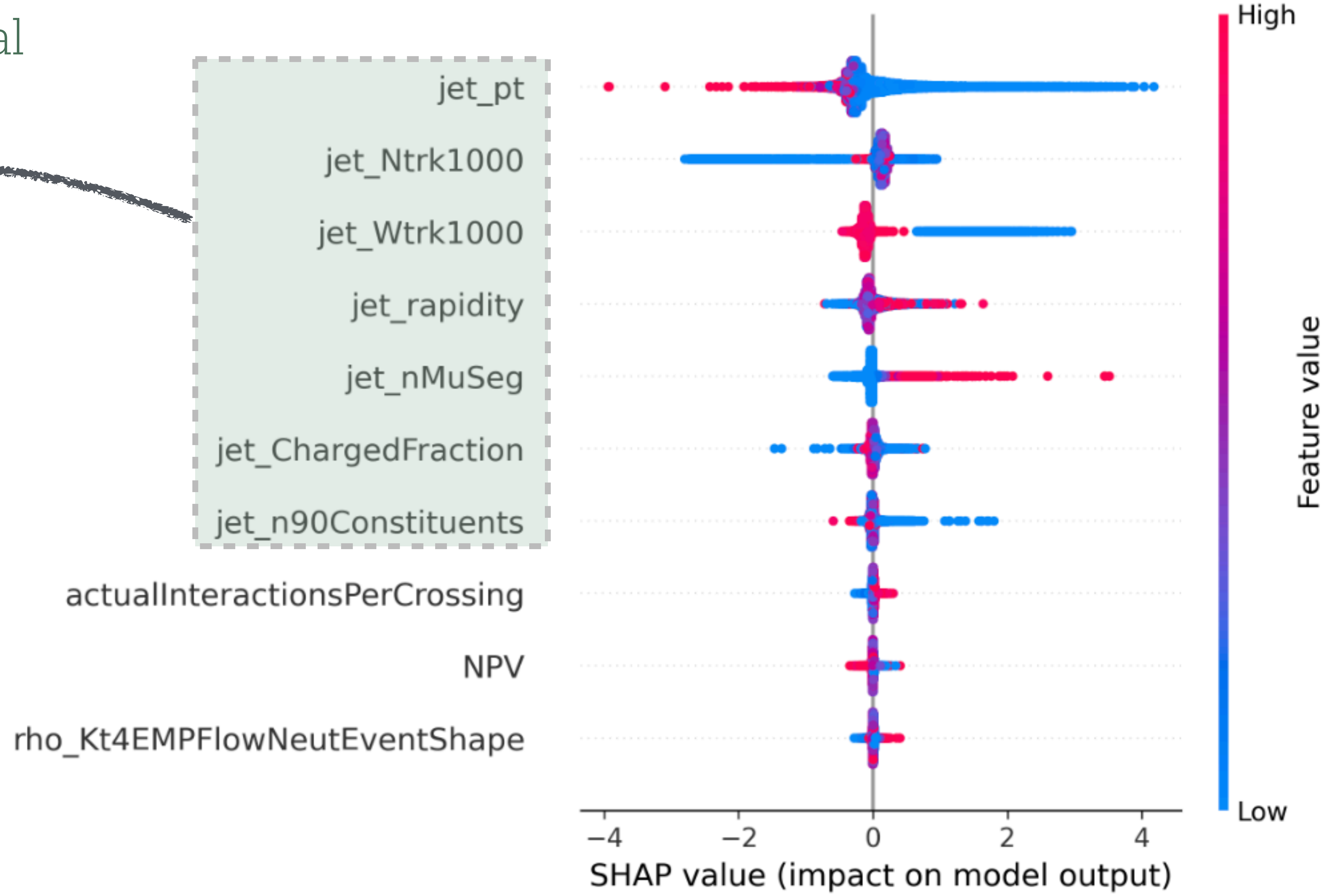
Are you what I think you are?

- We use SHAP to look at the feature importance to investigate if the network is leaning something unphysical

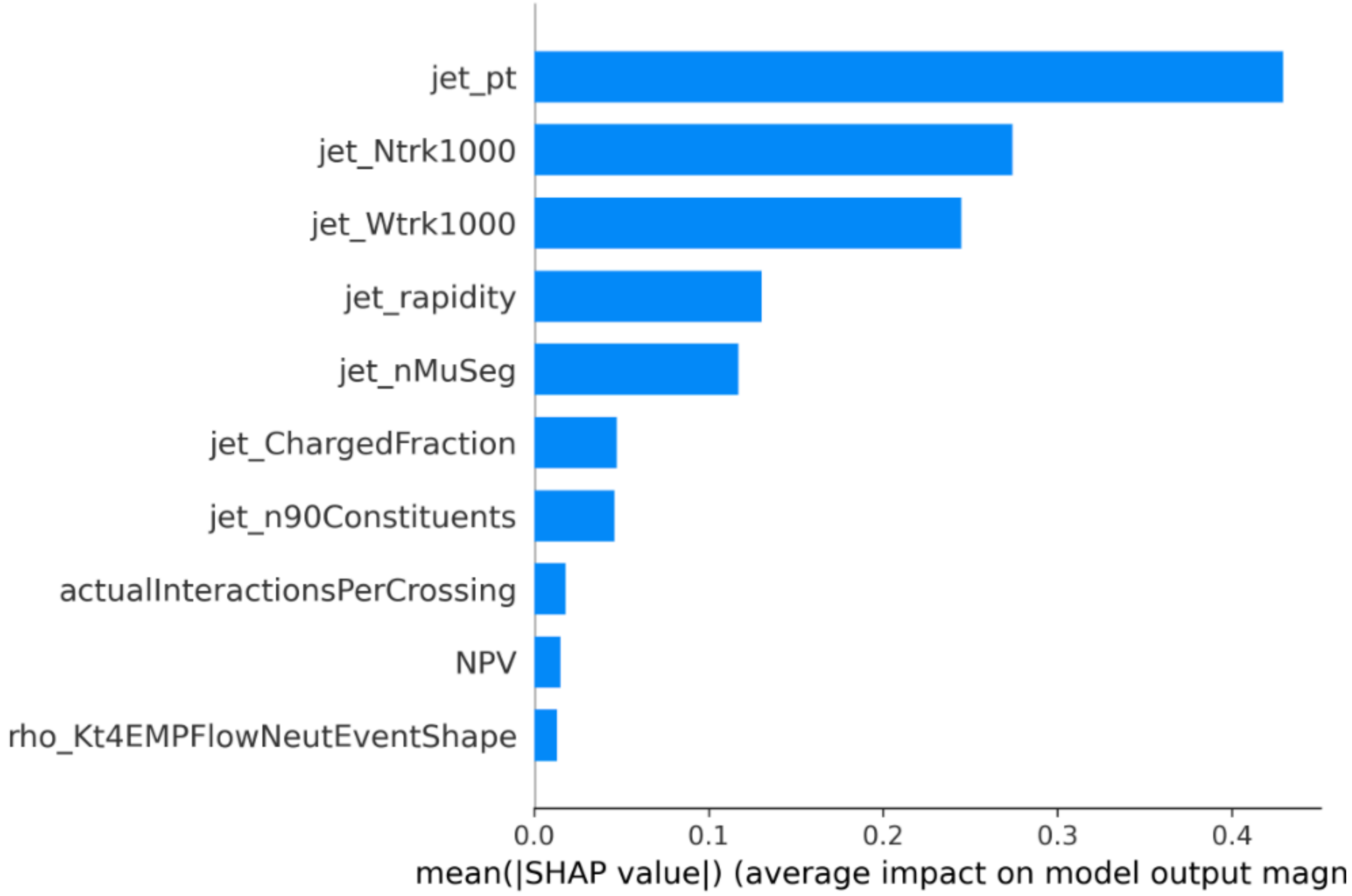
    - High importance to p_T and y: like MC-JES
    - N_{trk} , w_{trk} , $N_{\mu-seg}$, f^{Ch} , $N_{90\%}$: like GSC

Absolute MC-based Calibration — Corrects jet 4-momentum to the particle-level energy scale. Energy and direction are calibrated

↳ **Global Sequential Correction** — Reduces flavour dependance and energy leakage effects using calorimeter, track, and muon-segment variables



SHAP Feature Importance



- High importance to jet kinematic variables (p_T and y) and jet constituent variables ($N_{trk}, w_{trk}, N_{\mu-seg}, f^{Ch}, N_{90\%}$)
- Lower importance to pileup variables (μ_{act}, N_{P_V}, ρ)