Mind the Gap: Safely Navigating Inference through Transport Maps

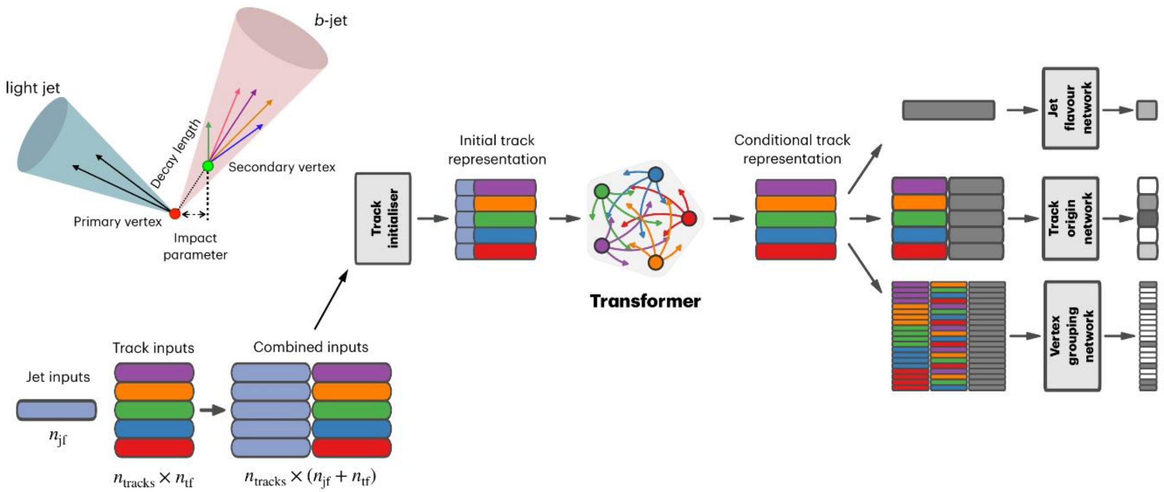
Malte Algren, Francesco Armando Di Bello,
Tobias Golling and Chris Pollard

Supervised learning in collider experiments

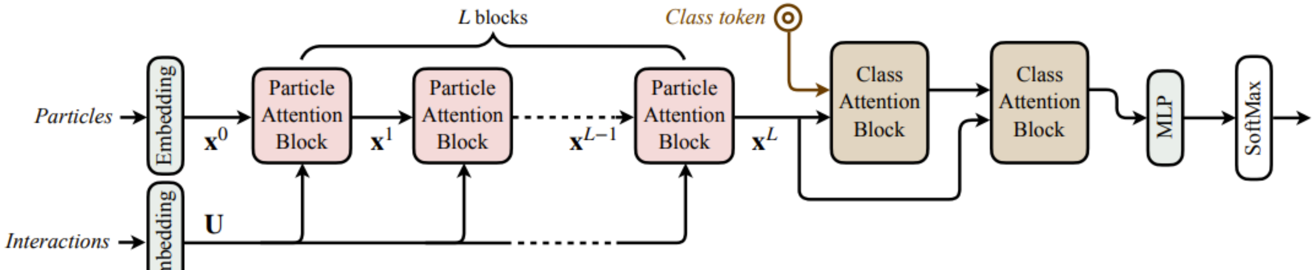
Machine learning methods have shown massive performance gain in many tasks

- Transformer, generative model etc.
- Extending the input space > point cloud
- Learning often simulation based

ATLAS flavor tagging (GN2)



CMS flavor tagging (ParT)



(a) Particle Transformer

Supervised learning in collider experiments

Machine learning methods have shown massive performance gain in many tasks

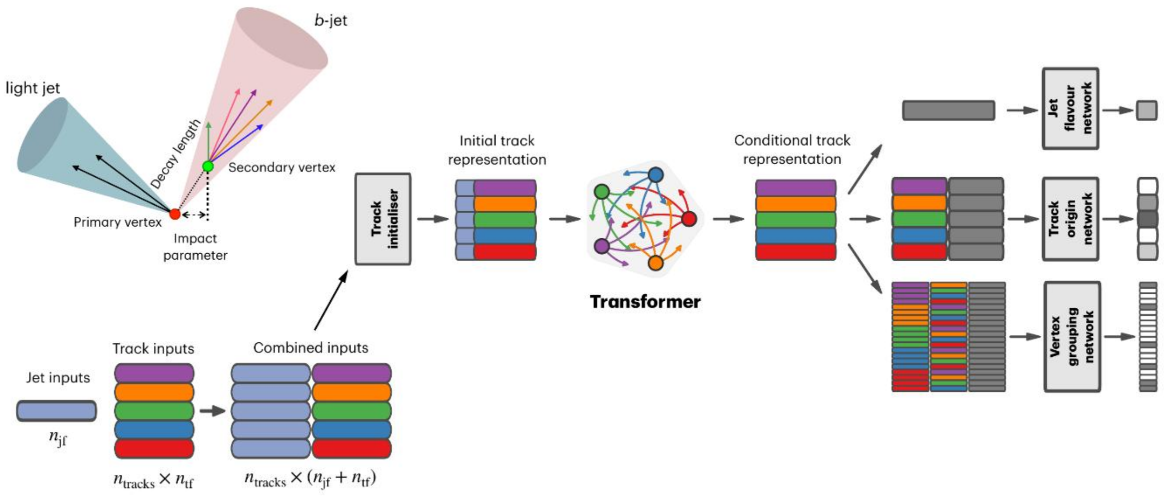
- Transformer, generative model etc.
- Extending the input space > point cloud
- Learning often simulation based

Calibration (domain shift) is often overlooked

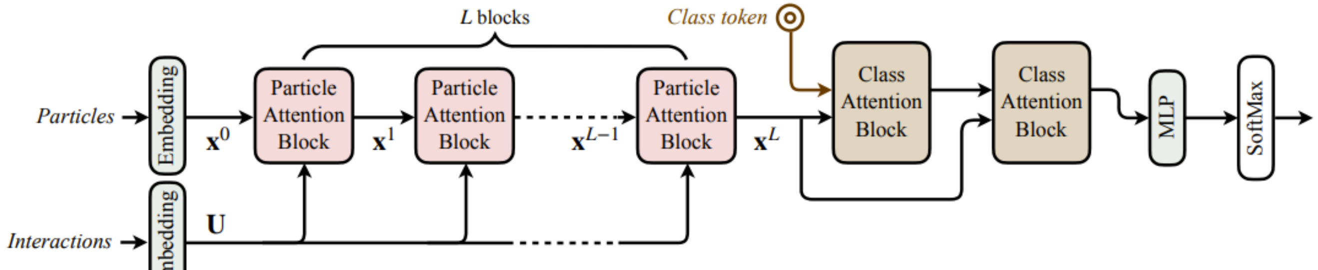
Source: MNIST **(MC)**

Target: SVHN **(Data)**

ATLAS flavor tagging (GN2)



CMS flavor tagging (ParT)



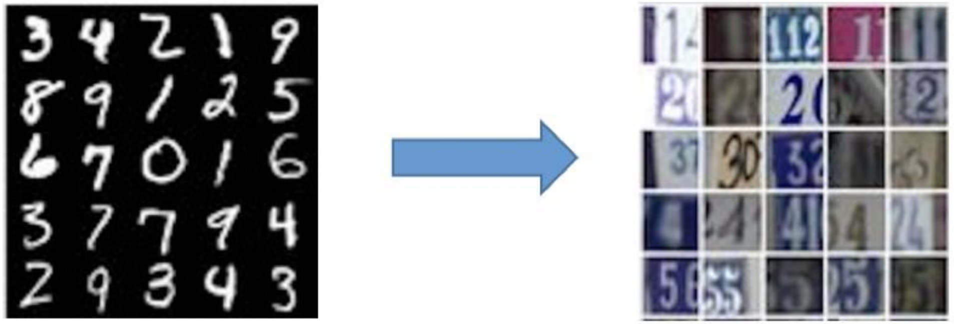
(a) Particle Transformer

Supervised learning in collider experiments

Machine learning methods have shown massive performance gain in many tasks

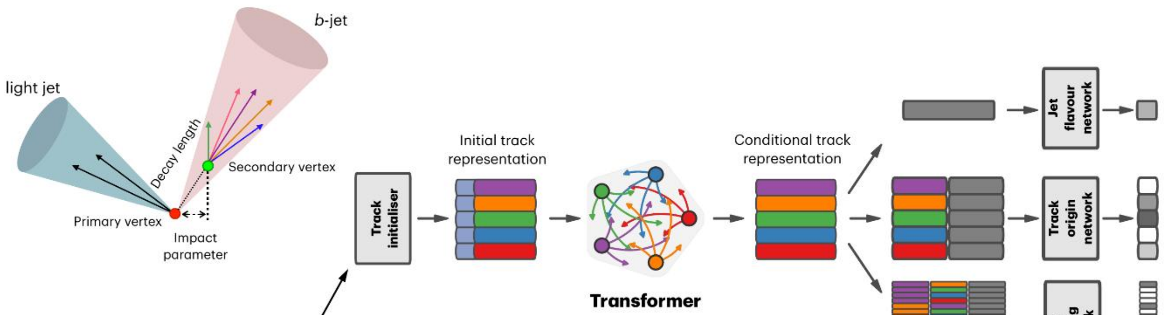
- Transformer, generative model etc.
- Extending the input space > point cloud
- Learning often simulation based

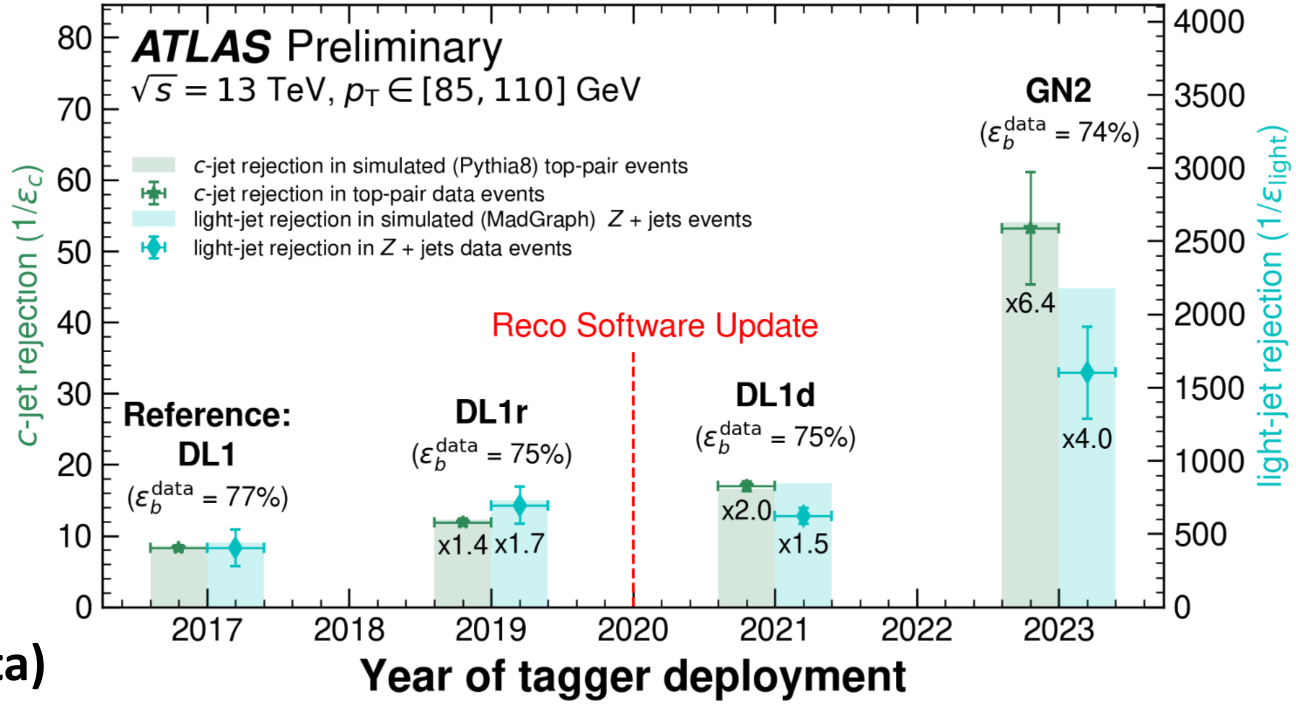
Calibration (domain shift) is often overlooked



Source: MNIST **(MC)** Target: SVHN **(Data)**

ATLAS flavor tagging (GN2)



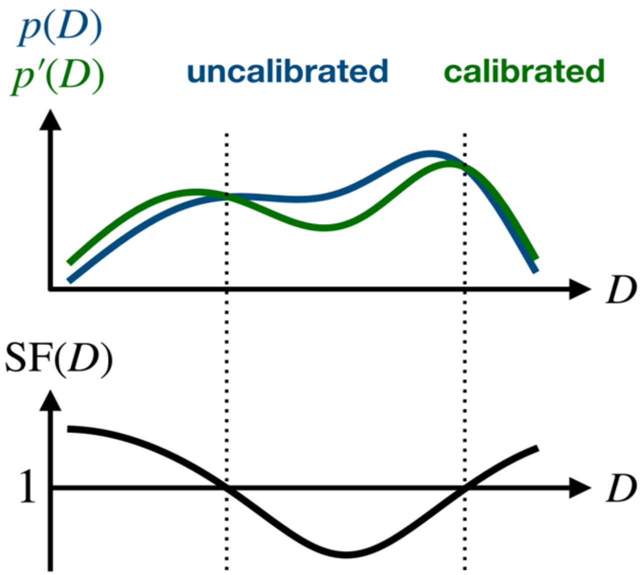


How do we account for mismodelling now

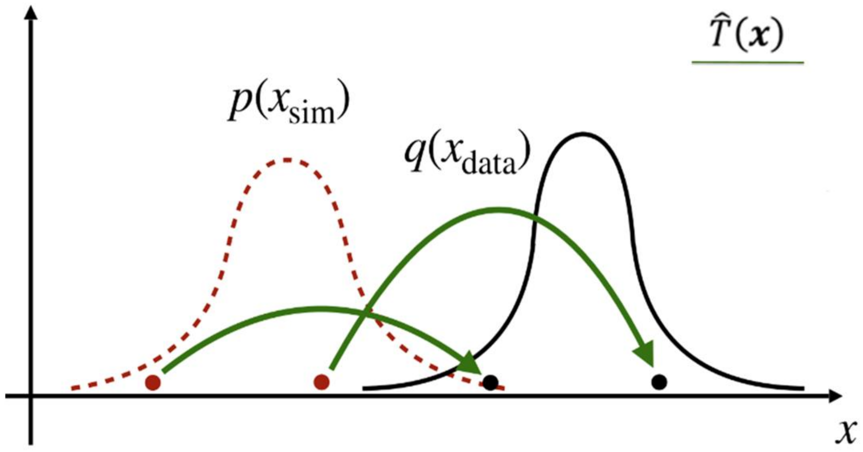
Inference with supervised trained machine learning model vulnerable to domain shifts:

- Trained on simulation and evaluated on data
- Leading to incorrect likelihoods and predictions from the models
- Traditionally, scale factors (SF) have been used to correct for this domain shift on the output
  - **Paradigm shift**: Transport your density instead of reweighting them

Scale factor (density ratio)



Optimal transport



Current implementation of continuous calibration

Implemented in ATLAS:

- Method paper is out on calibrating the output space of a tagger (4d calibration)

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

ATLAS EXPERIMENT

CERN

Submitted to: EPJC

CERN-EP-2025-026
20th May 2025

A continuous calibration of the ATLAS flavour-tagging classifiers via optimal transportation maps

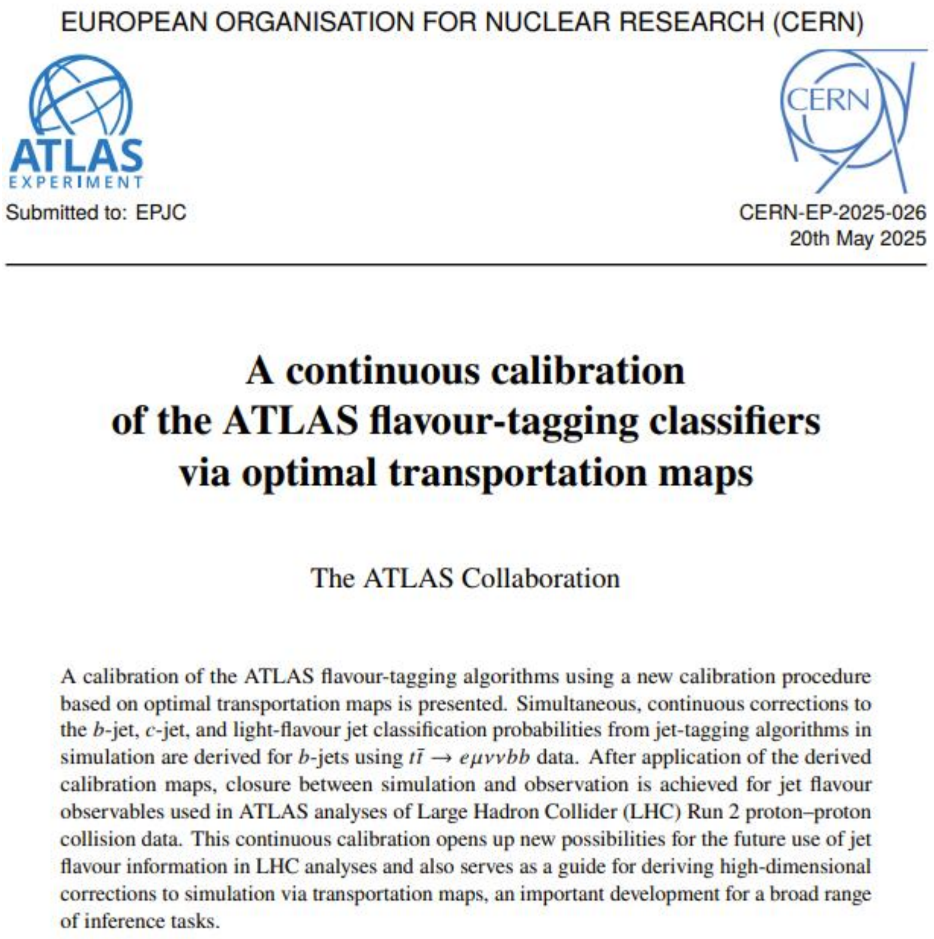
The ATLAS Collaboration

A calibration of the ATLAS flavour-tagging algorithms using a new calibration procedure based on optimal transportation maps is presented. Simultaneous, continuous corrections to the b-jet, c-jet, and light-flavour jet classification probabilities from jet-tagging algorithms in simulation are derived for b-jets using $t\bar{t} \rightarrow e\mu\nu\nu bb$ data. After application of the derived calibration maps, closure between simulation and observation is achieved for jet flavour observables used in ATLAS analyses of Large Hadron Collider (LHC) Run 2 proton–proton collision data. This continuous calibration opens up new possibilities for the future use of jet flavour information in LHC analyses and also serves as a guide for deriving high-dimensional corrections to simulation via transportation maps, an important development for a broad range of inference tasks.

Current implementation of continuous calibration

Implemented in ATLAS:

- Method paper is out on calibrating the output space of a tagger (4d calibration)


EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Submitted to: EPJC

CERN-EP-2025-026
20th May 2025

A continuous calibration of the ATLAS flavour-tagging classifiers via optimal transportation maps

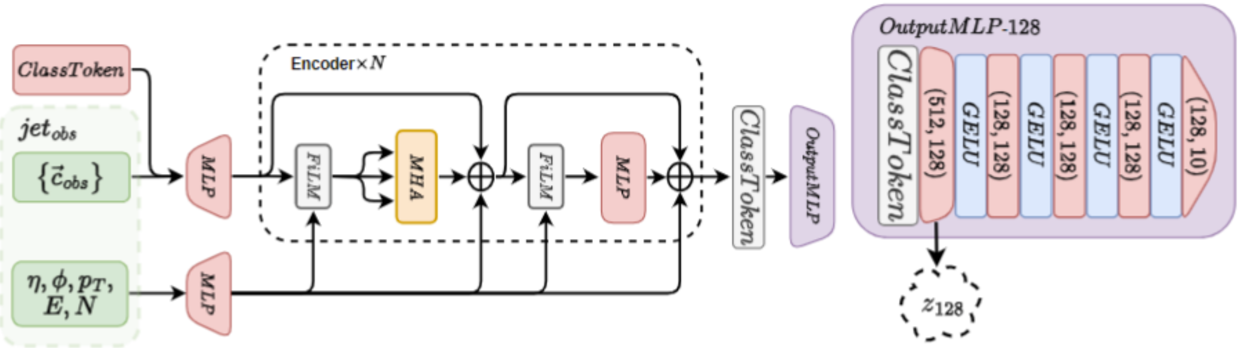
The ATLAS Collaboration

A calibration of the ATLAS flavour-tagging algorithms using a new calibration procedure based on optimal transportation maps is presented. Simultaneous, continuous corrections to the b-jet, c-jet, and light-flavour jet classification probabilities from jet-tagging algorithms in simulation are derived for b-jets using $t\bar{t} \rightarrow e\mu\nu\nu bb$ data. After application of the derived calibration maps, closure between simulation and observation is achieved for jet flavour observables used in ATLAS analyses of Large Hadron Collider (LHC) Run 2 proton–proton collision data. This continuous calibration opens up new possibilities for the future use of jet flavour information in LHC analyses and also serves as a guide for deriving high-dimensional corrections to simulation via transportation maps, an important development for a broad range of inference tasks.

Calibrate the latent representation

- Will contain more information about the jet

ATLAS & CMS inspired tagger



(a) (b)

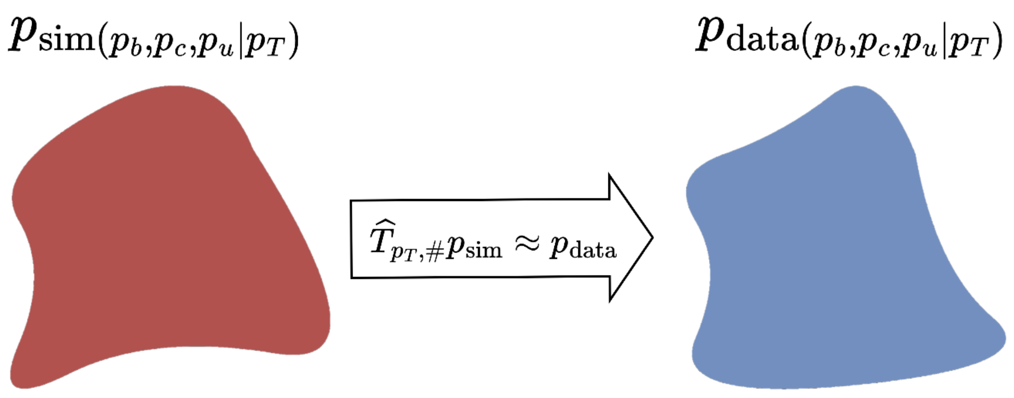
Move events instead of weighting them

- Learn a fully continuous transport map $\hat{T}_{\#}$
- That can morph $\hat{T}_{\#} p_{sim} \approx p_{data}$
- Restricted to the map that minimize:

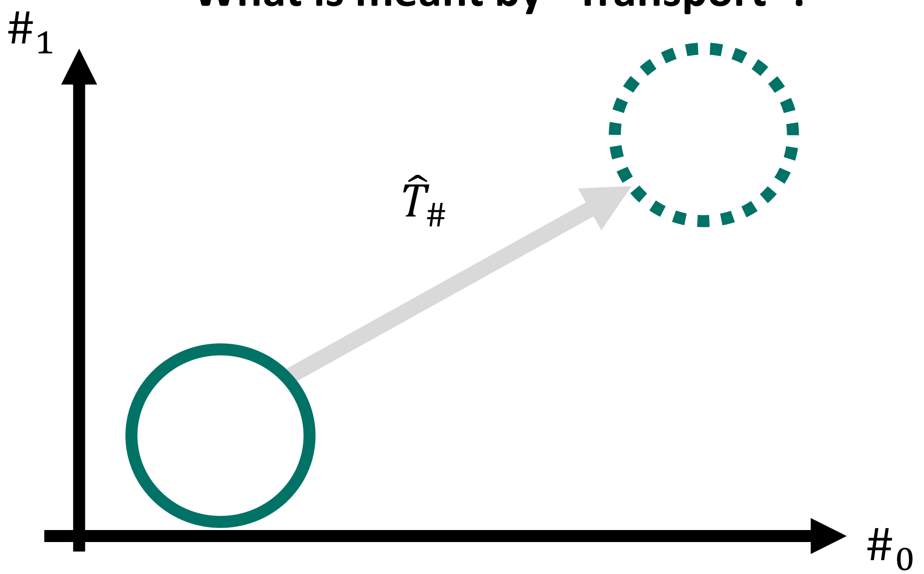
$$\vec{c}\,(\vec{x}, \vec{y}) = \left(\vec{x} - \hat{T}(\vec{x})\right)^2$$

- Advantages:
  - Continuous calibration without histograms
  - Scales to higher dimensions
  - Minimum change to existing simulation
  - Unitary (Do not alter number of jets)

OT calibrations within ATLAS



What is meant by "Transport"?



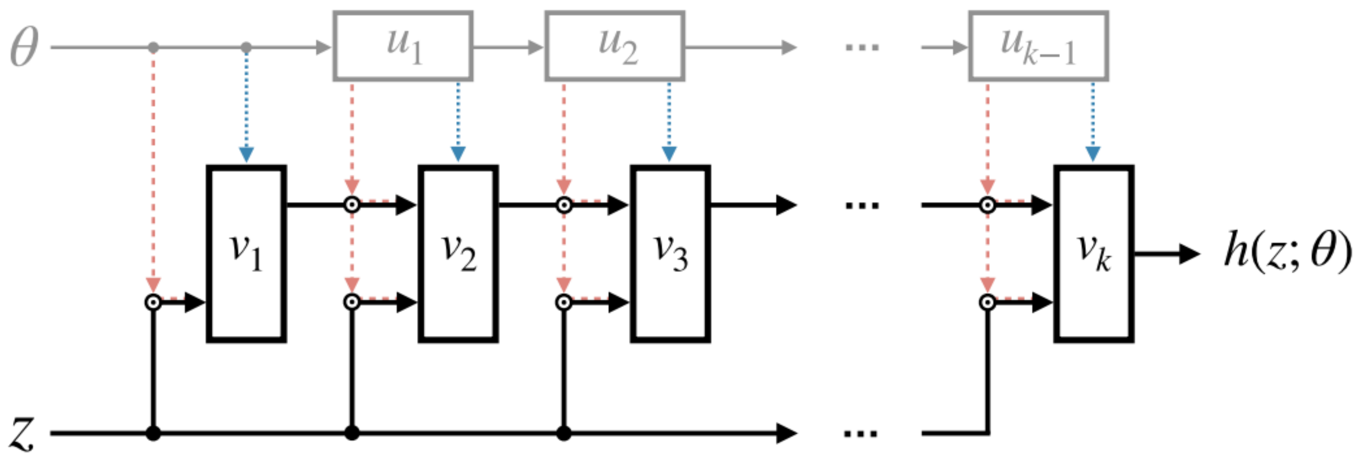
Finding the optimal transport

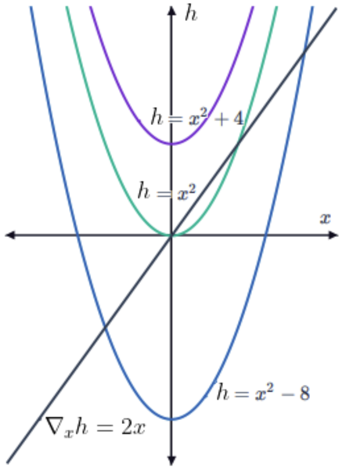
How to find the optimal transport between p_{sim} & p_{data} :

- Kantorovich duality – find the optimal transport using two **convex** function f & g

$$\mathbb{L}(f,g) = \min_f \max_g \sum f(y;\theta) + x \cdot \nabla_x g\,(x;\theta') - f(\nabla_x g\,(x;\theta'),\theta')$$

- Modern ML have solved this optimization problem by parameterizing f & g with *(P)ICNN*
  - GAN similar setup: $\nabla_x g\,(x;\theta')$ is used as a generator for the calibration

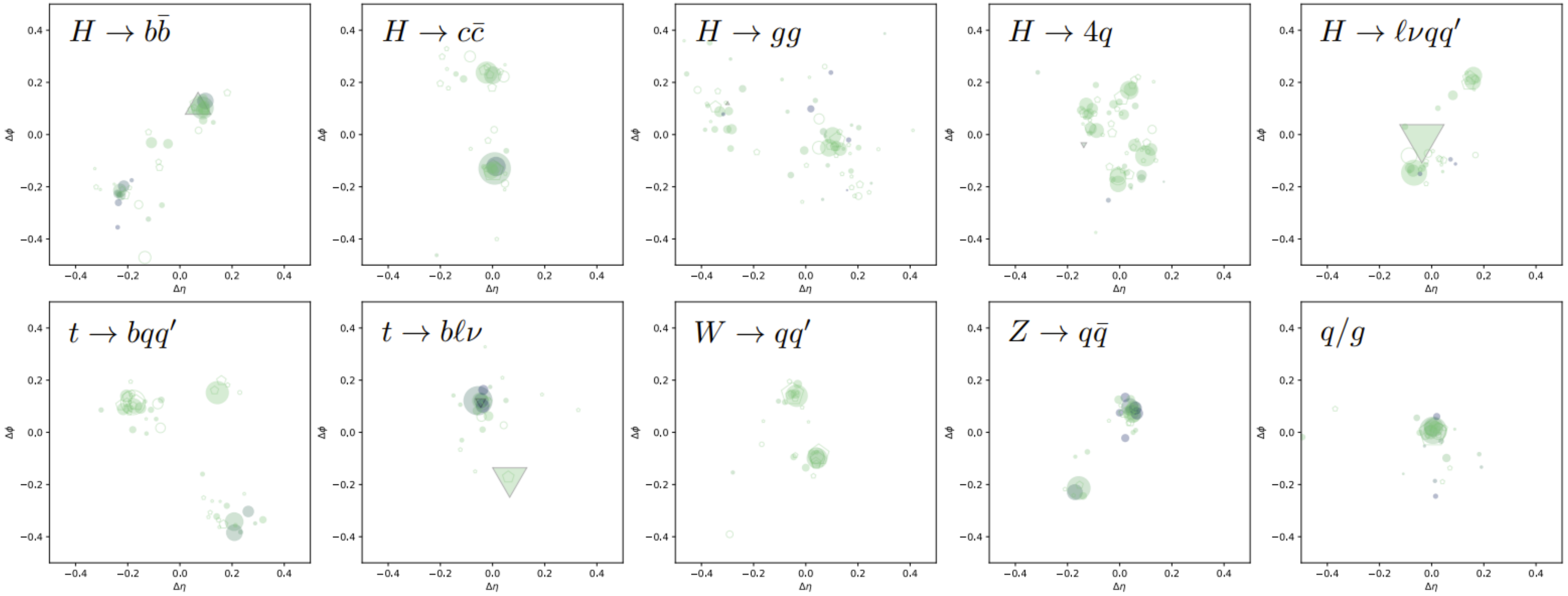




Testing group: JetClass

Open fast sim. dataset: JetClass

- Delphes-based with large R jets (0.8)
- A set of particle decays
- 10 labels

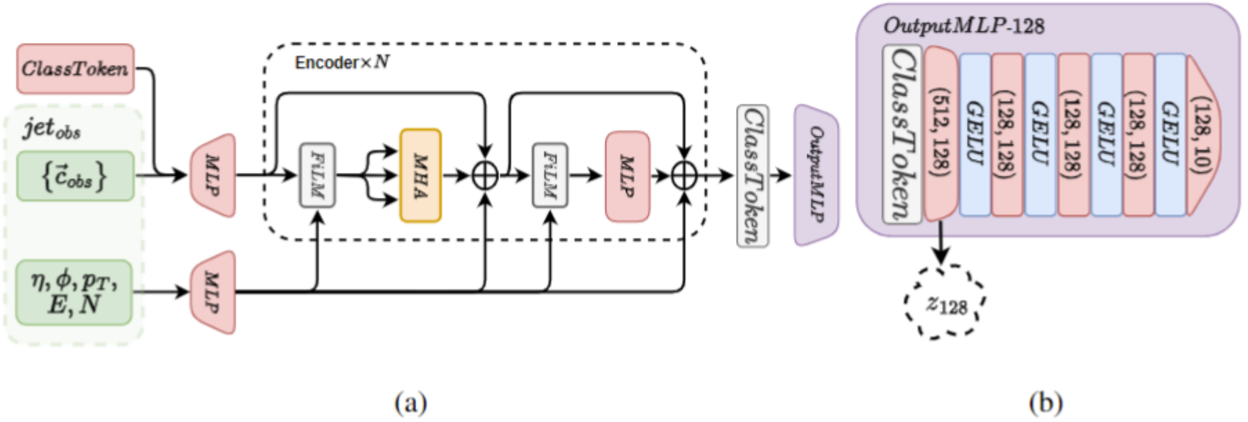


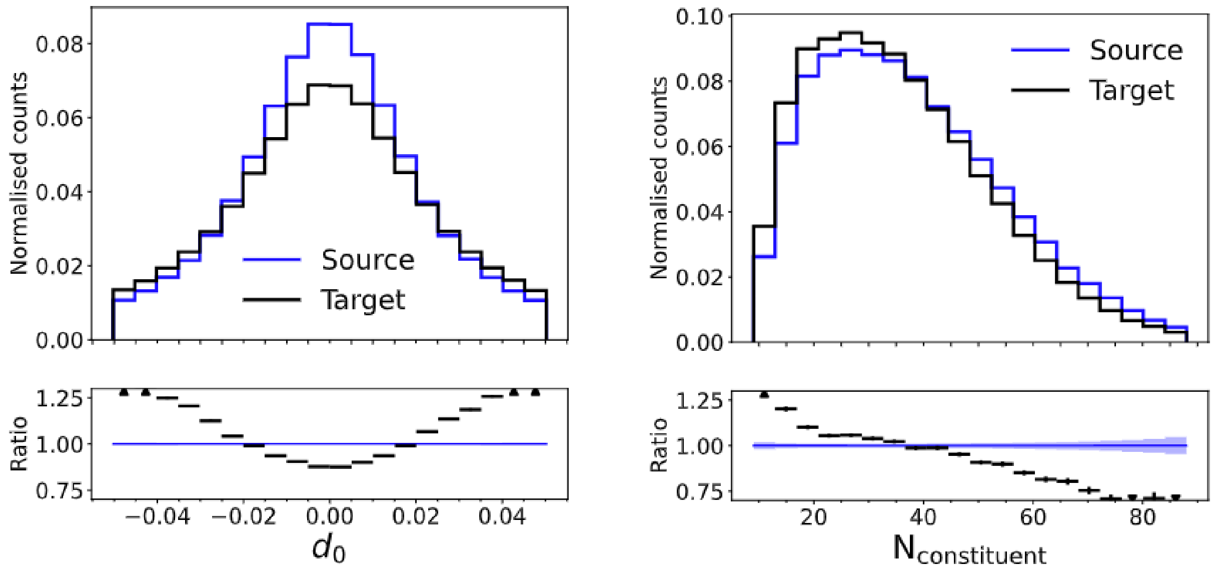
Continuous features x^{c}	
transverse momentum	p_{T}
pseudorapidity to jet axis	$\Delta\eta$
azimuthal angle to jet axis	$\Delta\phi$
transverse impact parameter	d_0
longitudinal impact parameter	z_0
uncertainty on d_0	$\sigma(d_0)$
uncertainty on z_0	$\sigma(Z_0)$
Particle type x^{id}	
photon	0
negative hadron	1
neutral hadron	2
positive hadron	3
electron	4
positron	5
muon	6
antimuon	7

Latent calibration on JetClass

Setup for Latent calibration:

1. Train tagger on original JetClass
2. Simulate two variations of JetClass
    1. Source ("MC")
    2. Target ("Data")
    - Change:
        - Experimental effects (IP smearing)
        - Theory parameters
3. Derive the calibration (only q/g)
    1. Source ==> Target



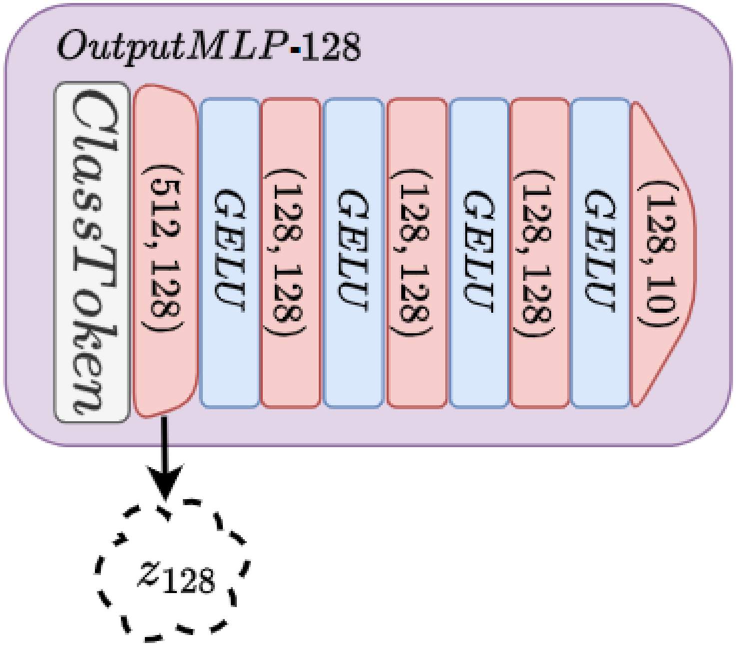


Latent calibration on JetClass

How to evaluate latent calibration:

1. 2d correlation of PCA(128>10)
2. Marginals in 10d output space
3. Physics motivated discriminators of output space
4. NN discriminators:
   - z_{128} where the calibration is derived
   - Penult. z_{128} last layer before L(128,10)
   - z_{10} output space

Frozen!

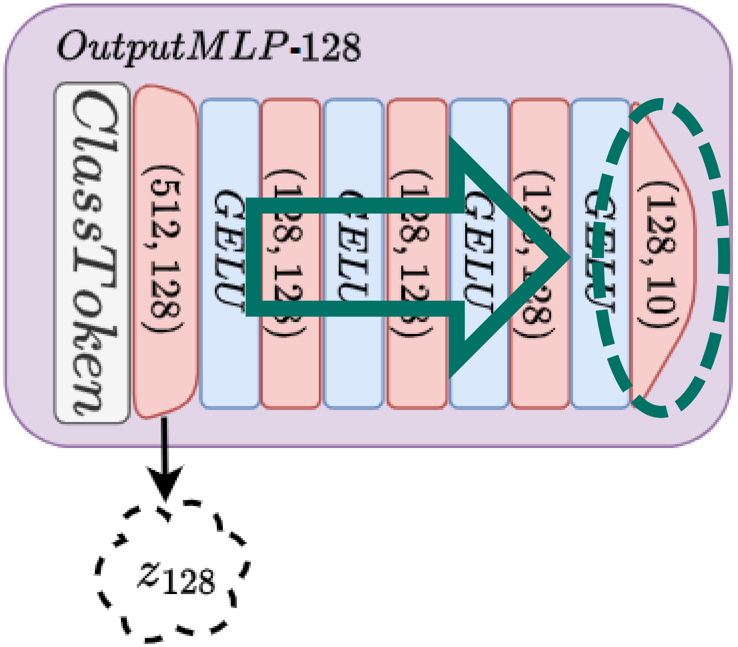


Latent calibration on JetClass

How to evaluate latent calibration:

1. 2d correlation of PCA(128>10)
2. Marginals in 10d output space
3. Physics motivated discriminators of output space
4. NN discriminators:
   - z_{128} where the calibration is derived
   - Penult. z_{128} last layer before L(128,10)
   - z_{10} output space

Frozen!

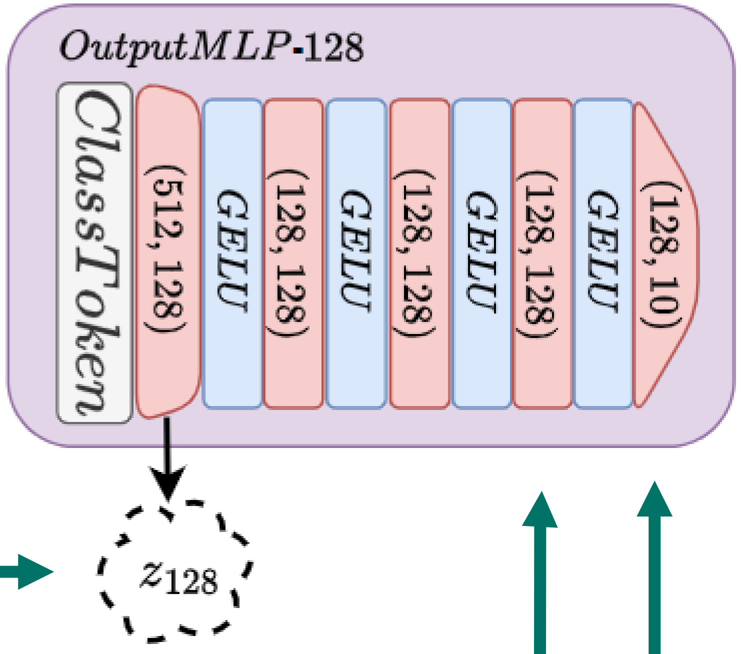


Latent calibration on JetClass

How to evaluate latent calibration:

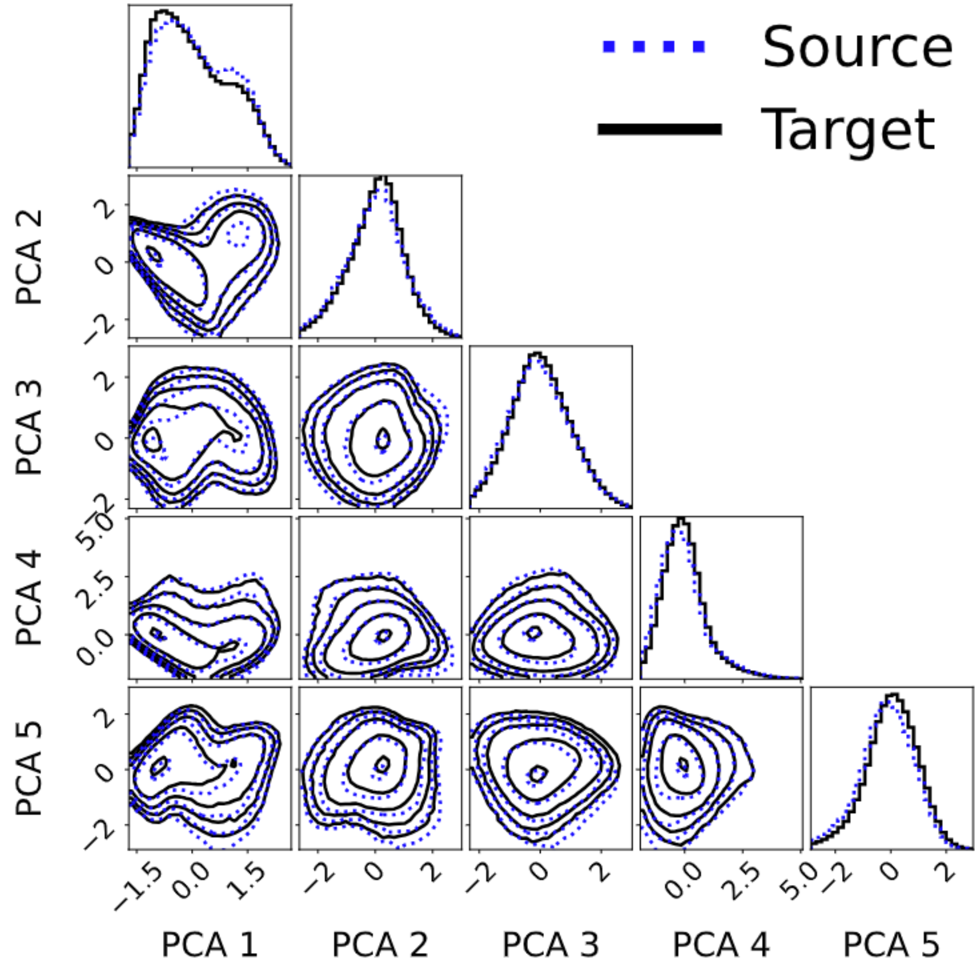
1. 2d correlation of PCA(128>10)
2. Marginals in 10d output space
3. Physics motivated discriminators of output space
4. NN discriminators:
   - z_{128} where the calibration is derived
   - Penult. z_{128} last layer before L(128,10)
   - z_{10} output space

Frozen!



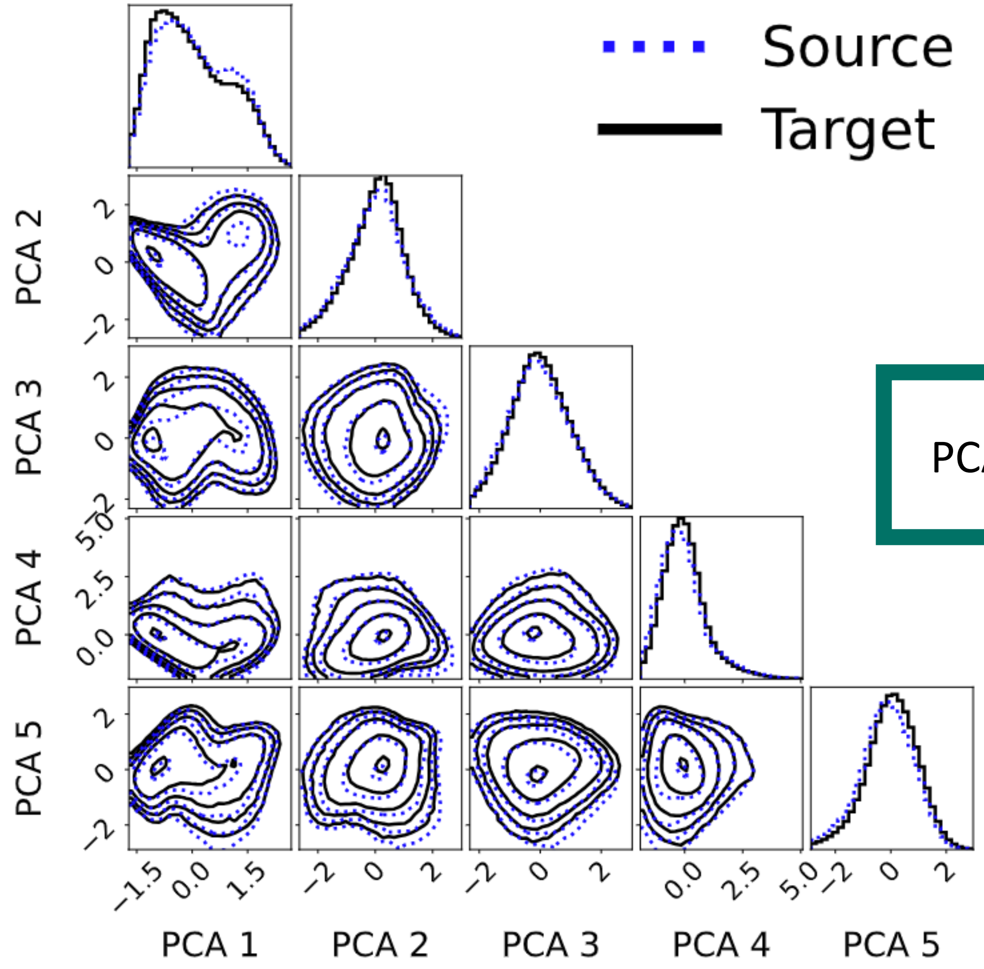
Result: PCA in 128d

2d correlation of PCA(128>10) in z_{128}

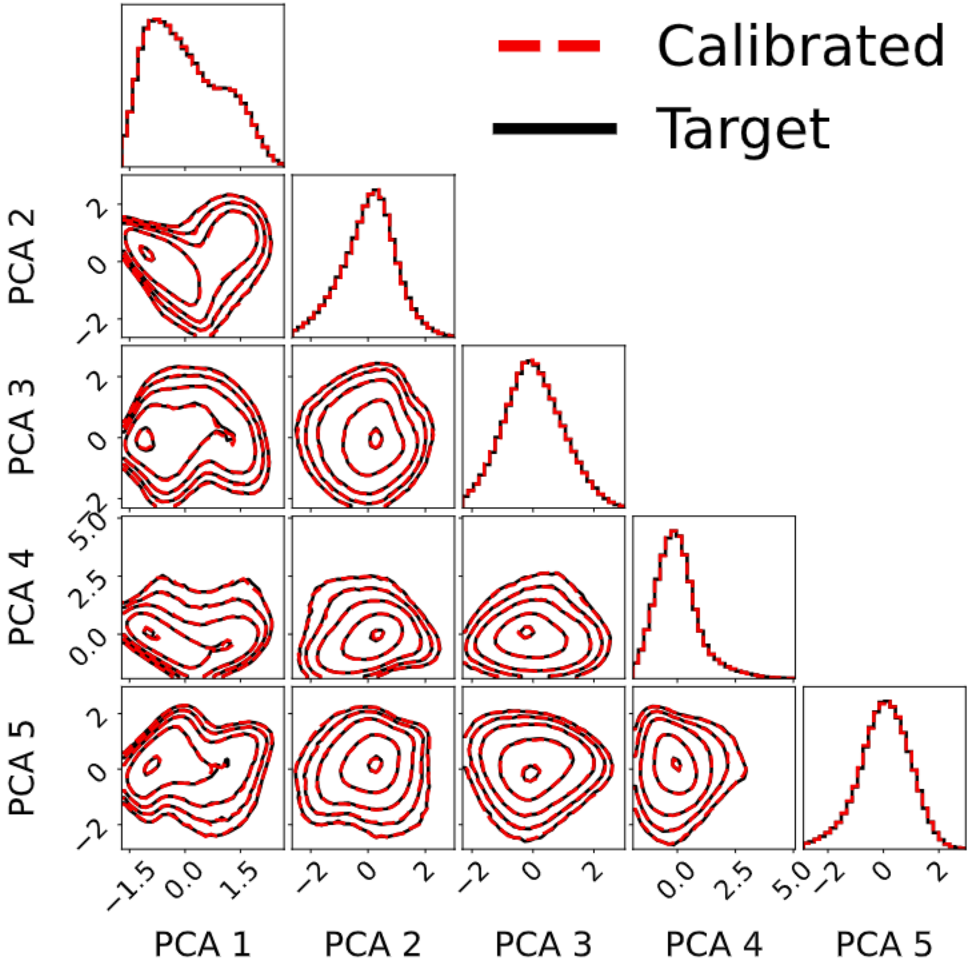


Result: PCA in 128d

2d correlation of PCA(128>10) in z_{128}



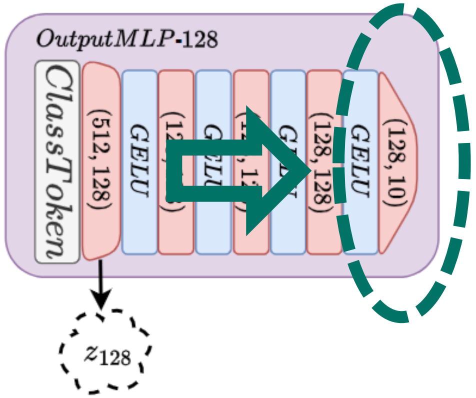
PCA($\hat{T}_{\#}p_{sim}(z_{128})$)

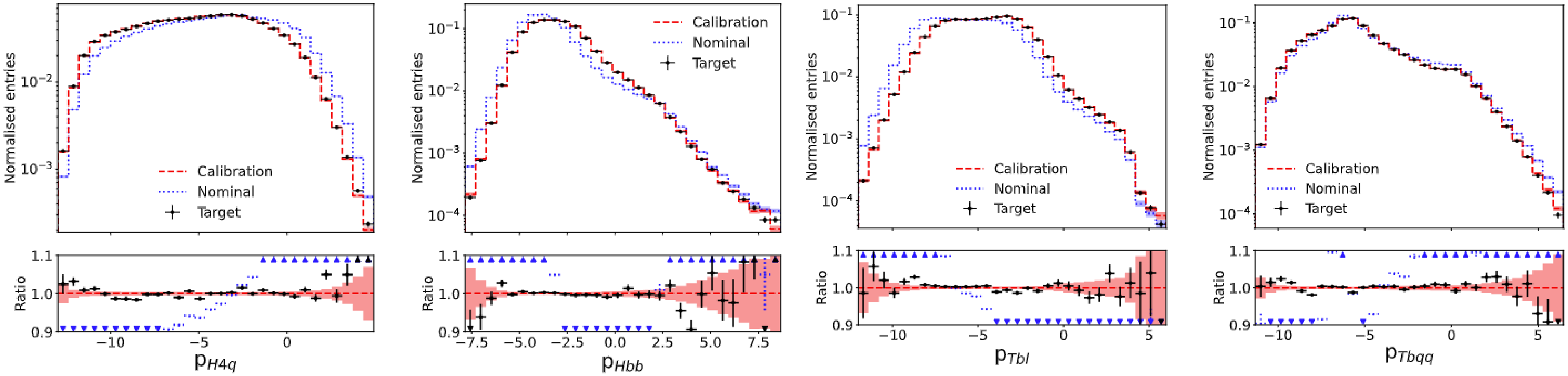


Result: Output space

Marginals in 10d output space

- Here we have a subset of the marginals
- General good closure in the marginals (red vs black)

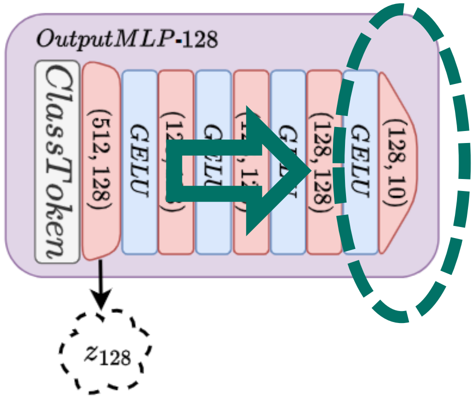


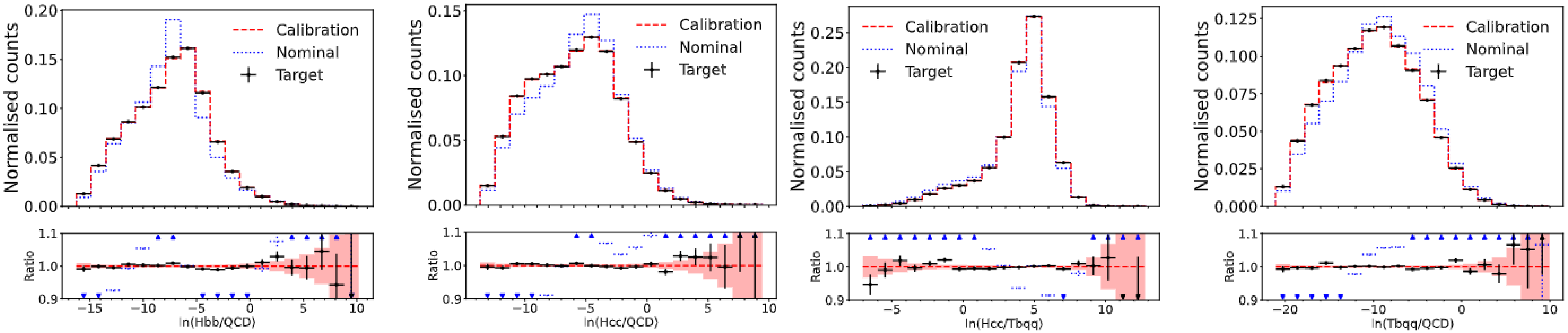


Result: Physics discriminators

Physics motivated discriminators of output space

- Testing a subset of 2d correlations
- Closure still looking good (red vs black)

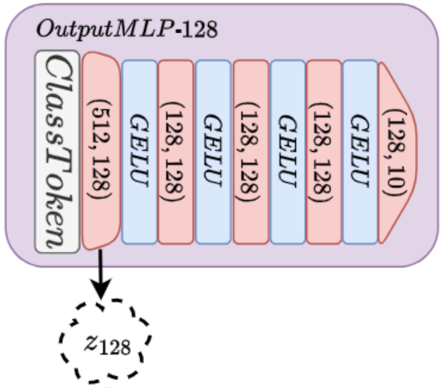


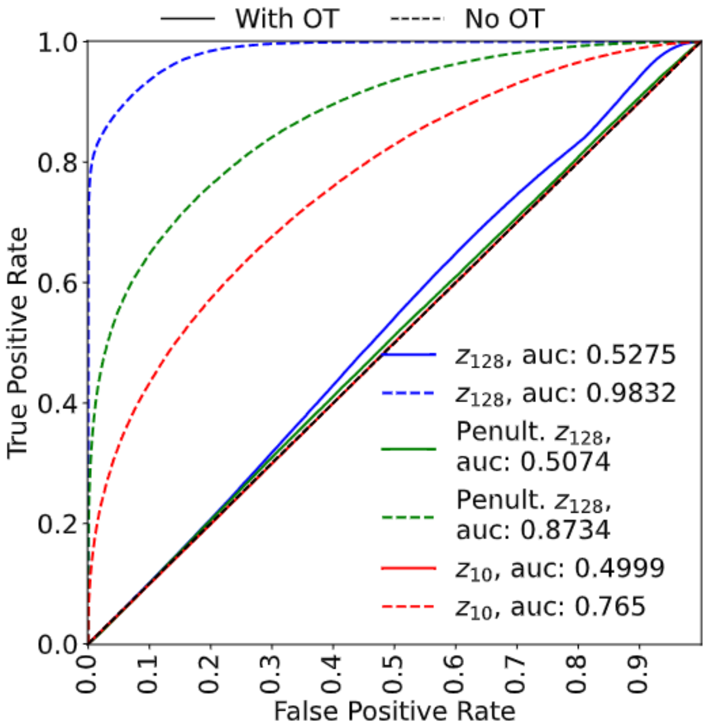


Result: Latent calibration

NN discriminators:

- **Train a discriminator between Nominal and Target**
    - Discriminator evaluate between calibration and Target
- The calibration remove discrepancies present in the nominal

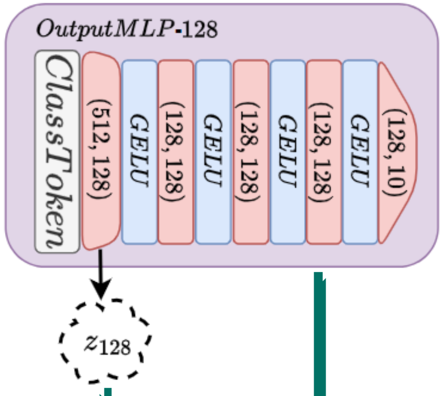


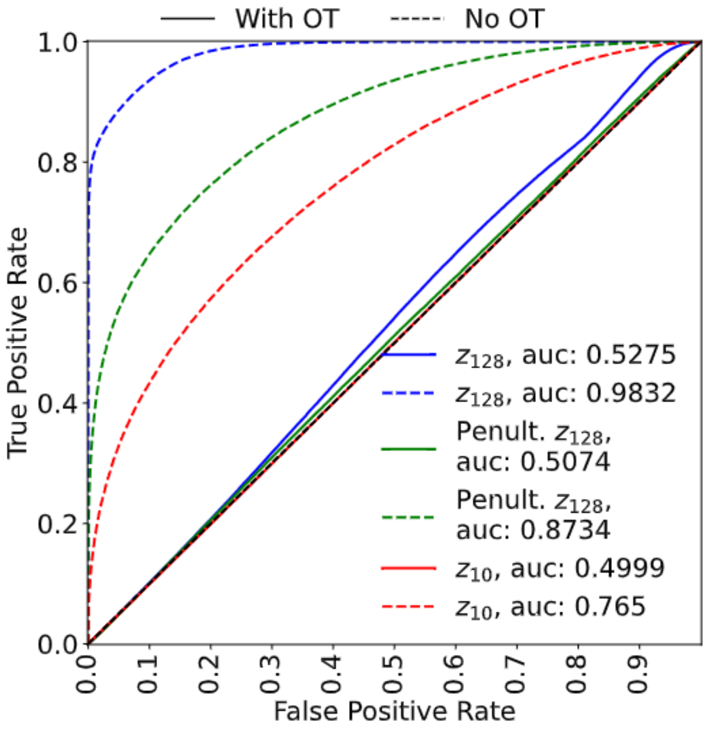


Result: Latent calibration

NN discriminators:

- **Train a discriminator between Nominal and Target**
    - Discriminator evaluate between calibration and Target
- The calibration remove discrepancies present in the nominal

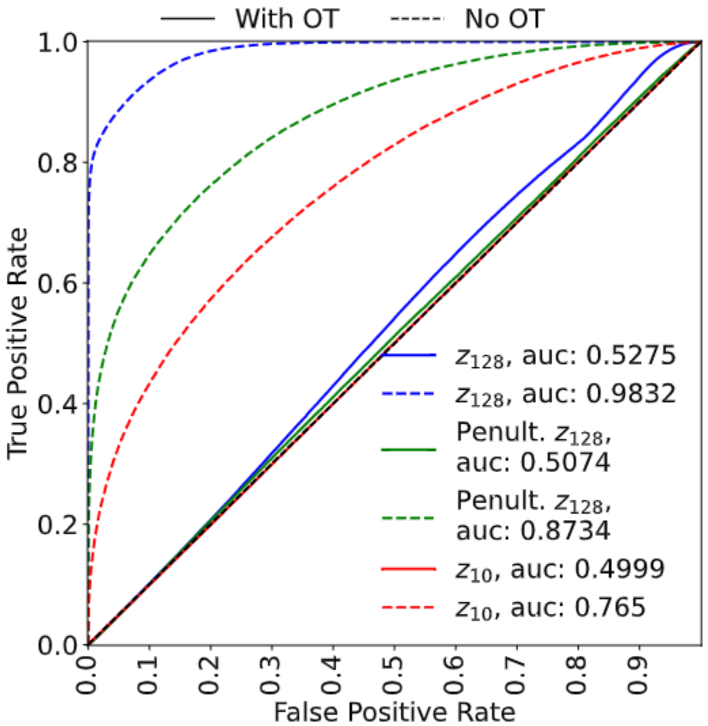


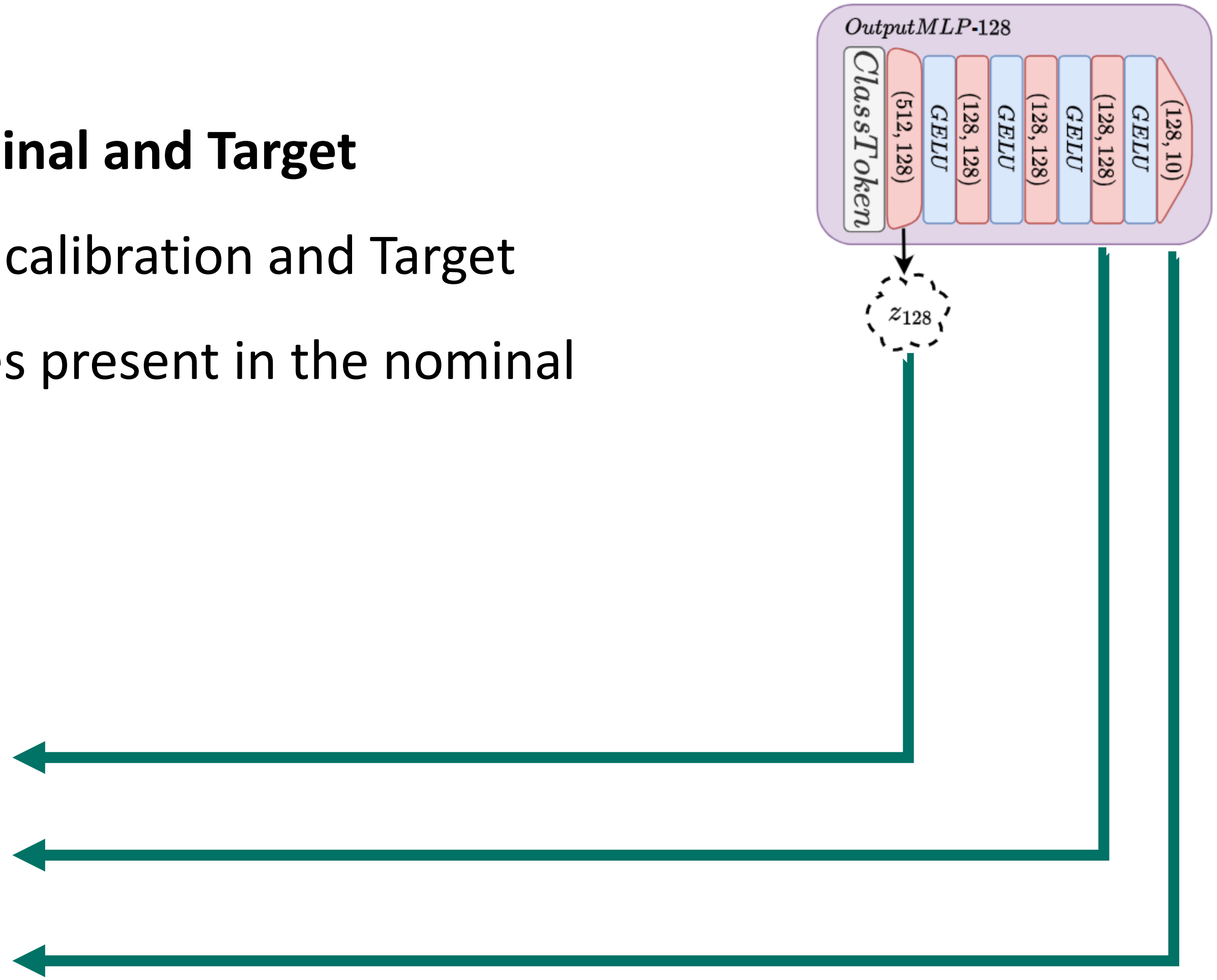


Result: Latent calibration

NN discriminators:

- **Train a discriminator between Nominal and Target**
  - Discriminator evaluate between calibration and Target
- The calibration remove discrepancies present in the nominal

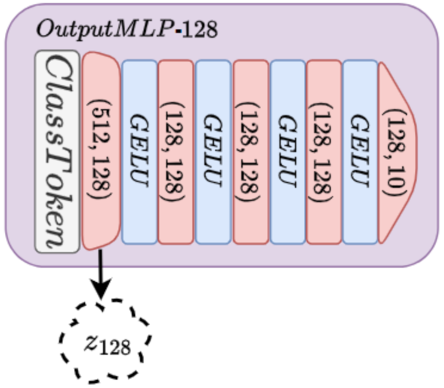


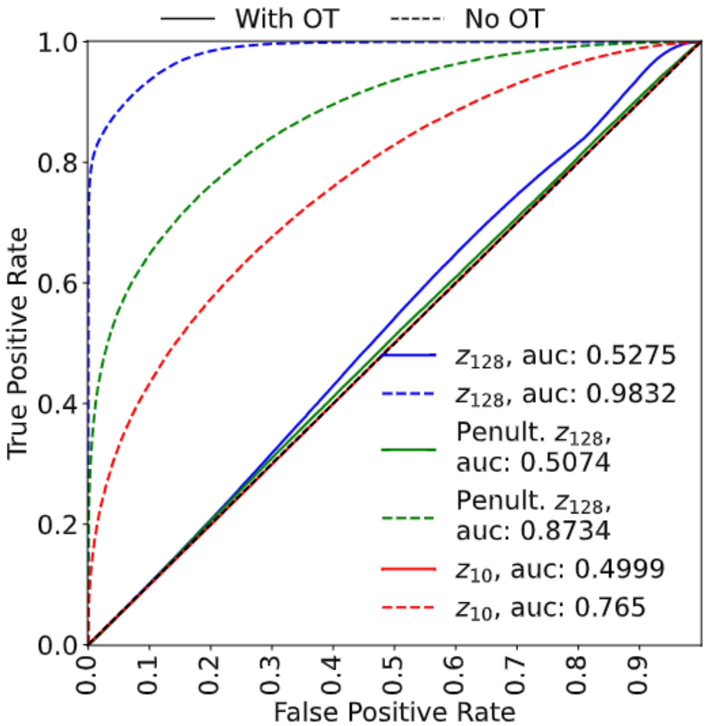


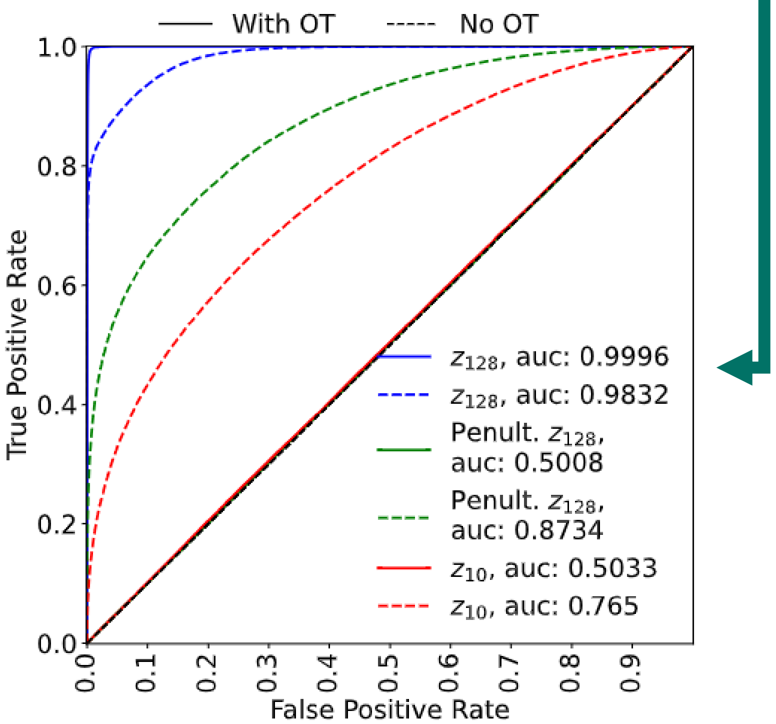
Result: Latent calibration

NN discriminators:

- Train a discriminator between Calibration and Target
- Non-closure or noise have been introduced by OT
- However, the noise is not propagated



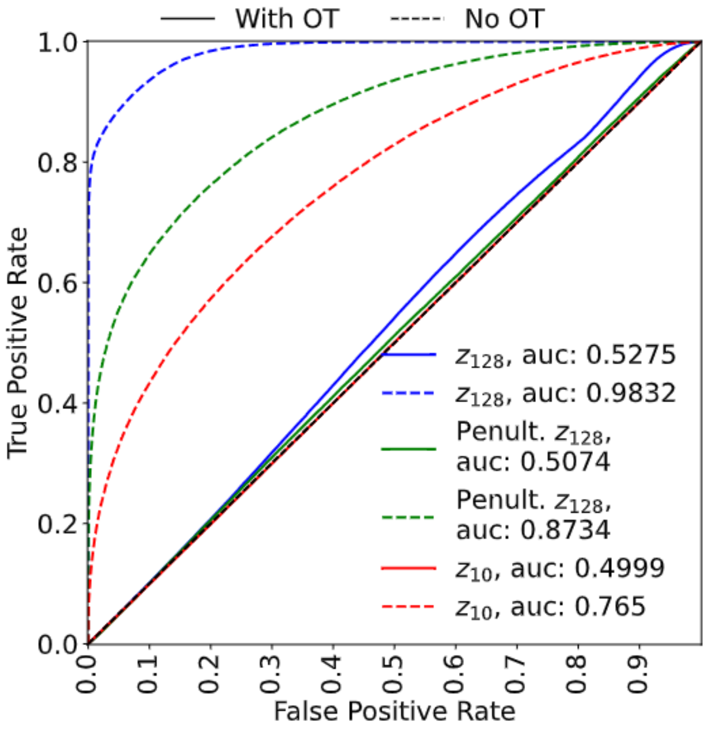


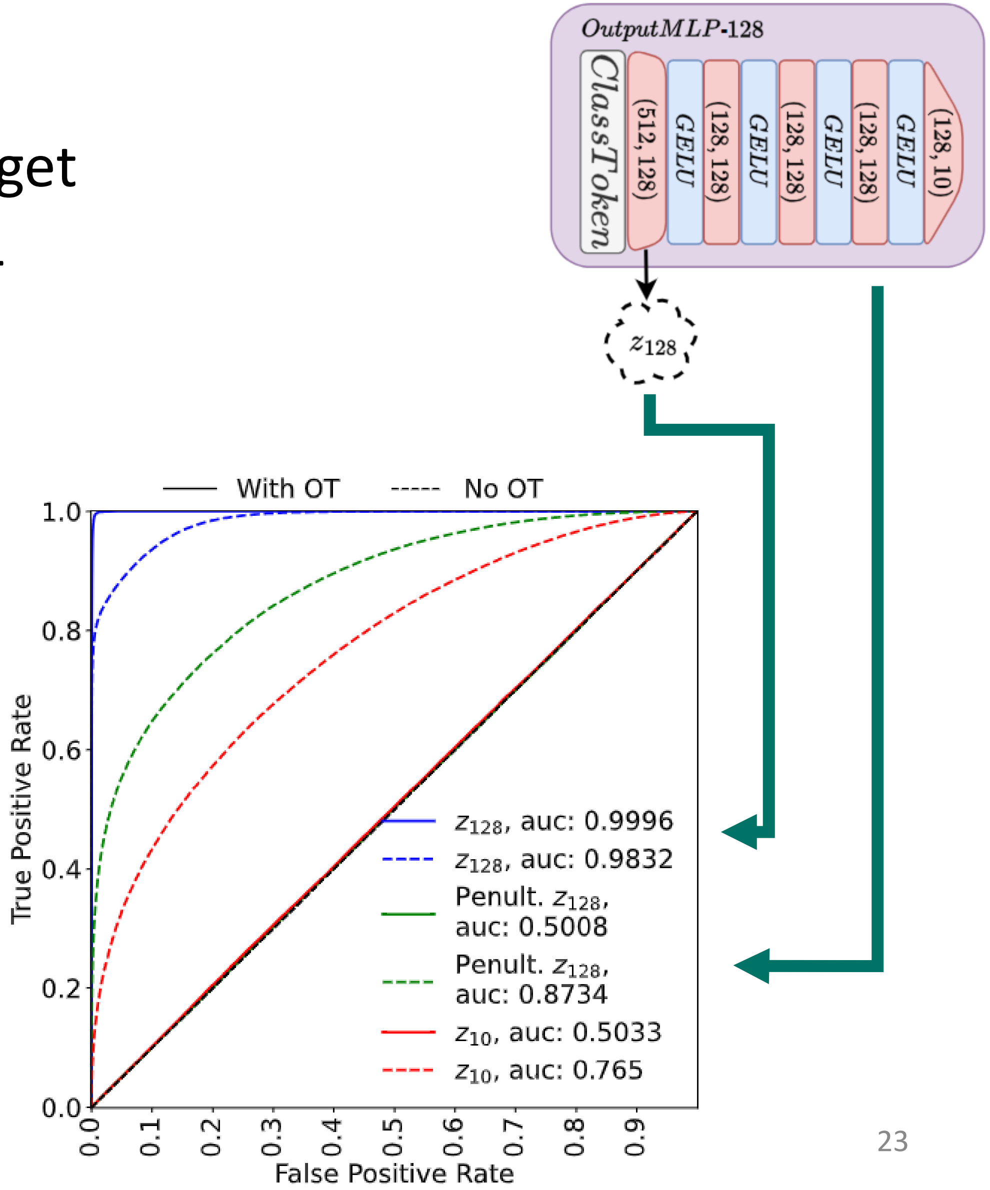


Result: Latent calibration

NN discriminators:

- Train a discriminator between Calibration and Target
- Non-closure or noise have been introduced by OT
- However, the noise is not propagated

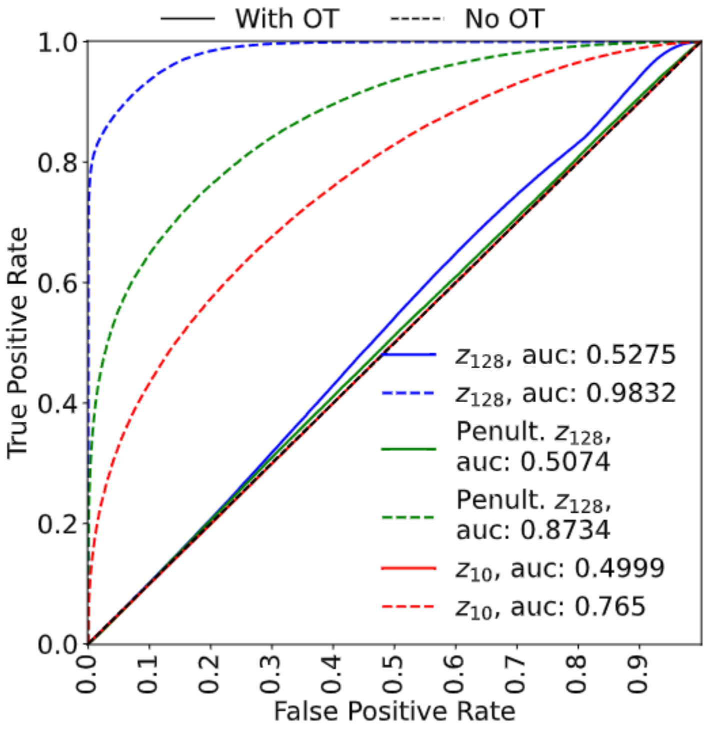


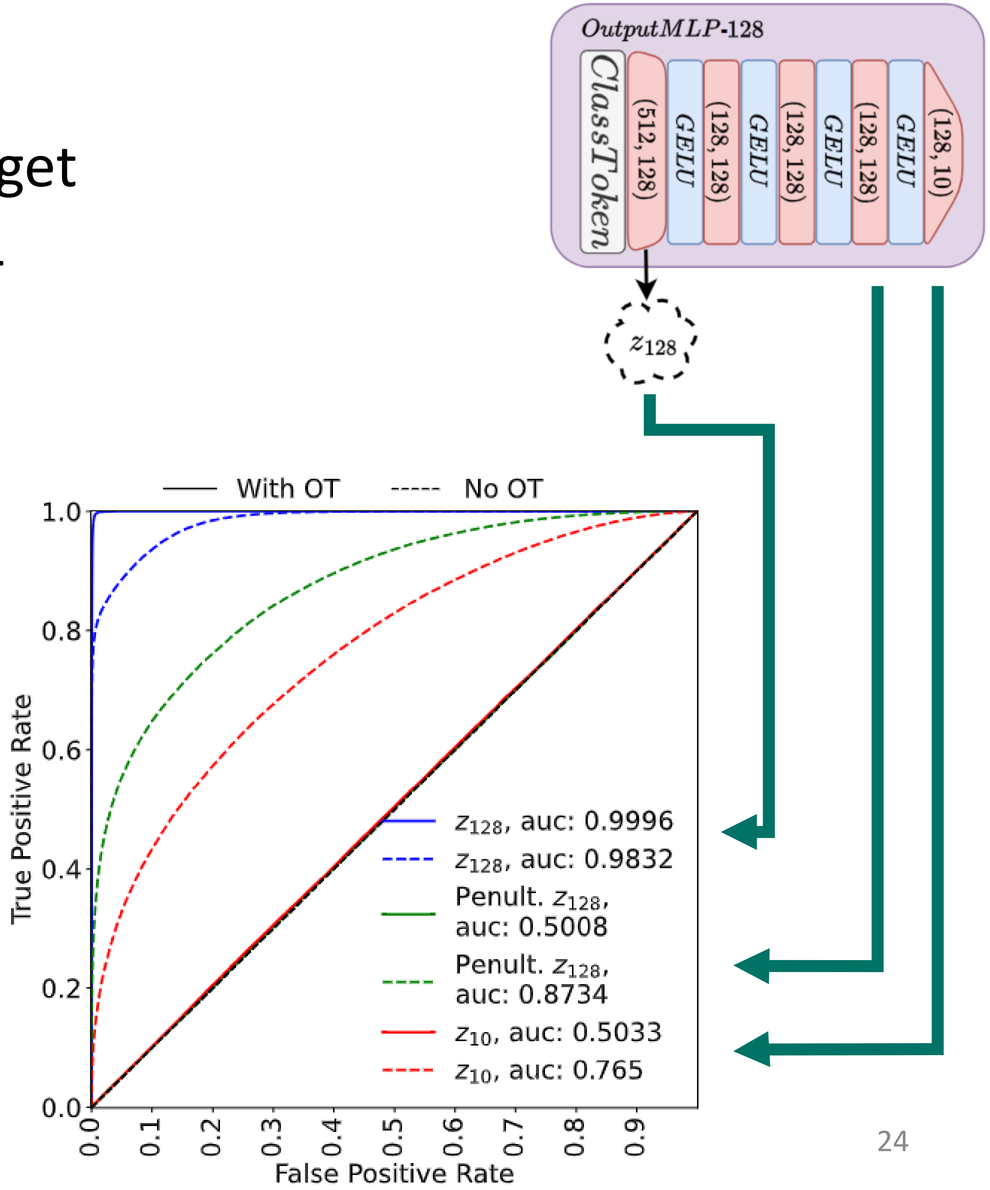


Result: Latent calibration

NN discriminators:

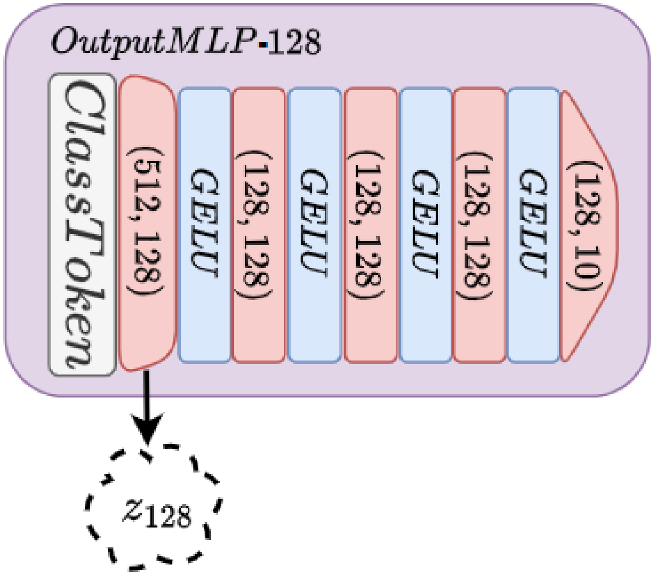
- Train a discriminator between Calibration and Target
- Non-closure or noise have been introduced by OT
- However, the noise is not propagated

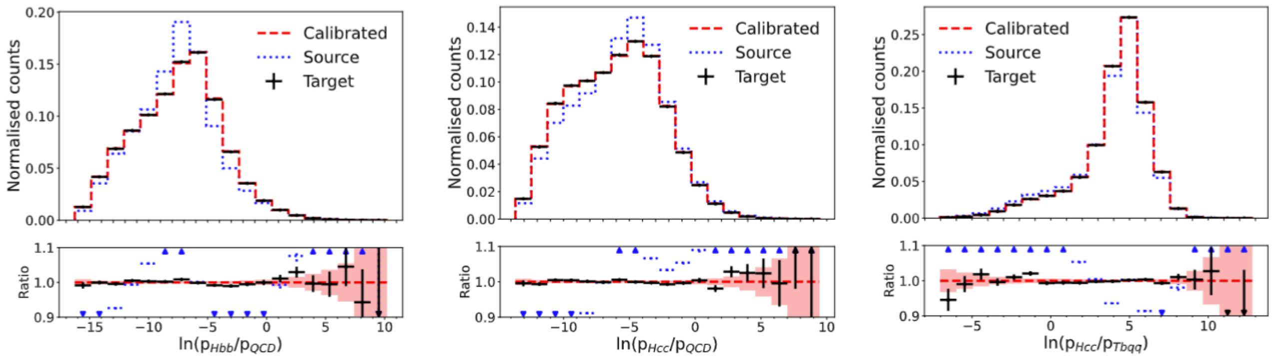




Summary

- We can derive a continuous calibration using OT
- It can calibrate the latent features of a backbone
- Despite noise/non-closure being introduced in the calibration – this information is not propagated forward in subsequent layers!
- Implementation within ATLAS (on the output)





Backup slides

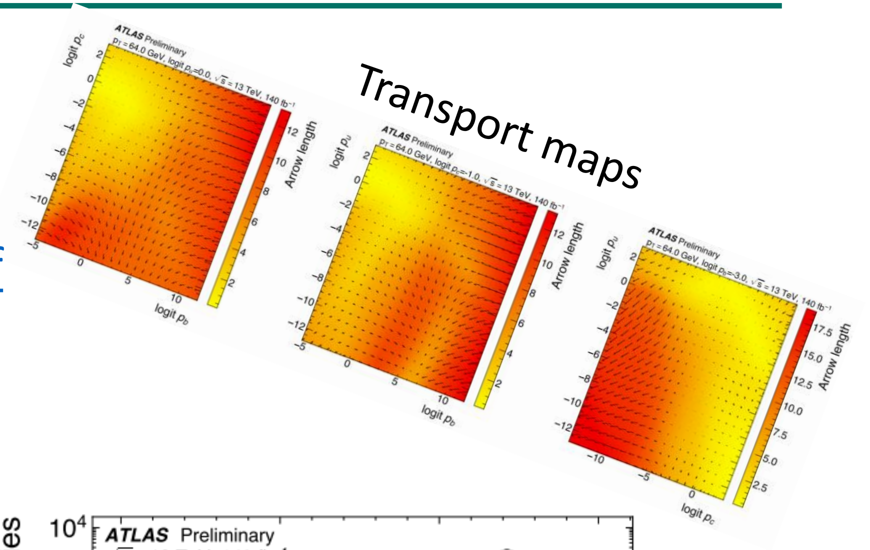
Current calibration of FTAG

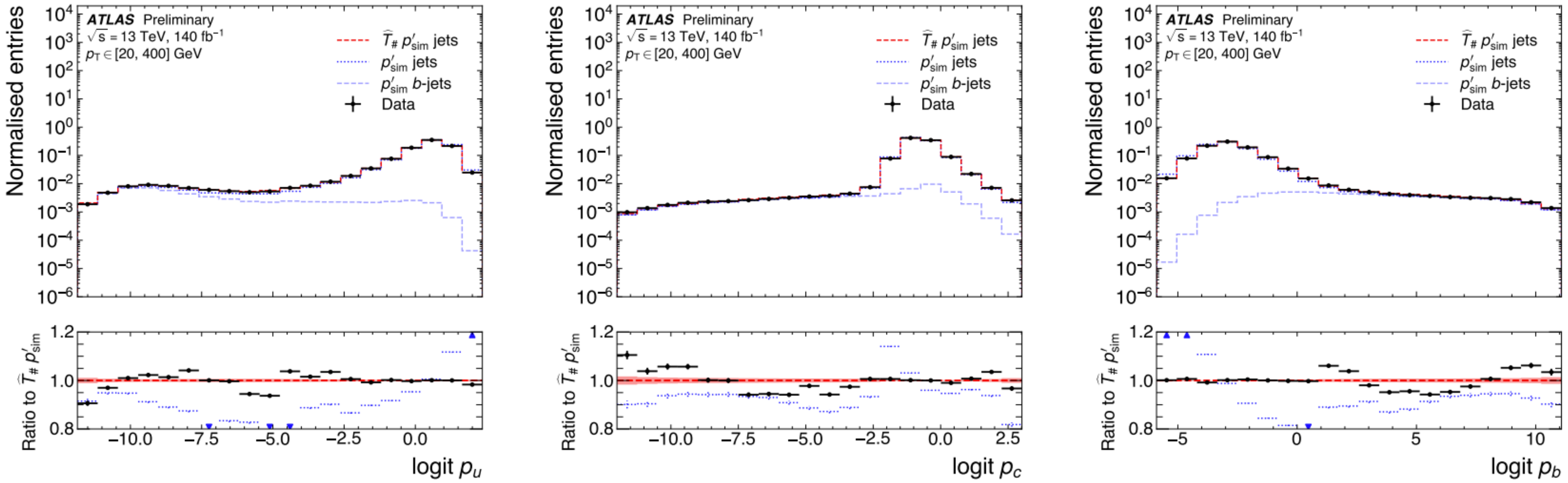
Optimal transport-based calibration

- Current OT DL1r calibration:

  https://cds.cern.ch/record/2911642/files/ATLAS-CONF-2024-014.pdf

Transport maps





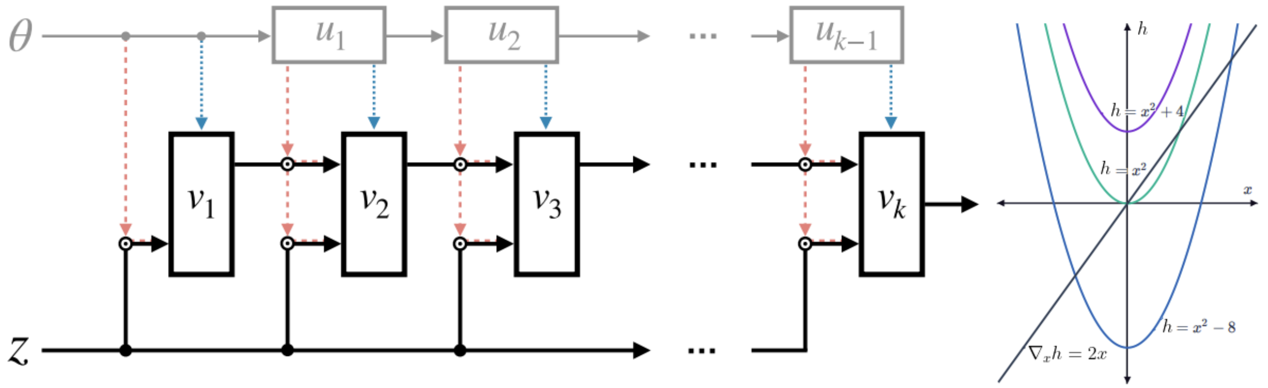
Finding the optimal transport

How to find the optimal transport between p_{sim} & p_{data} :

- Kantorovich duality – find the optimal transport using two **convex** function f & g

$$\mathbb{L}(f,g) = \min_f \max_g \sum f(y;\theta) + x \cdot \nabla_x g\,(x;\theta') - f(\nabla_x g\,(x;\theta'),\theta')$$

- Modern ML have solved this optimization problem by parameterizing f & g with *(P)ICNN*
    - GAN similar setup:
        - $\nabla_x g\,(x;\theta')$ is used as a generator for the calibration
- *Caveats for training* (details in backup)
    - $\theta' = \theta$
    - $p_{sim}(\cdot\,|\mathrm{background}) < p_{data}$



Finding the optimal transport

$$\text{logit}\, p_i = \log\left(\frac{p_i}{1 - p_i}\right)$$

How to find the optimal transport between p_{sim} & p_{data} :

- Kantorovich duality – find the optimal transport using two **convex** function f & g

$$\mathbb{L}(f, g) = \min_f \max_g \sum f(y; \theta) + x \cdot \nabla_x g\,(x; \theta') - f(\nabla_x g\,(x; \theta'), \theta')$$

- x: Probabilities of p_{sim}, y: Probabilities of p_{data} and $\theta = p_{\mathrm{T}}$
- At convergence, the optimal transport between p_{sim} & p_{data} will be
- Only *b*-jets will be calibration/moved

$$\hat{T}_{p_{\mathrm{T}}}\, \vec{q} = \begin{cases} \vec{\nabla}\, \hat{g}(\vec{q}) & \text{for } b\text{-jets} \\ \vec{q} & \text{for non-}b\text{-jets} \end{cases}$$

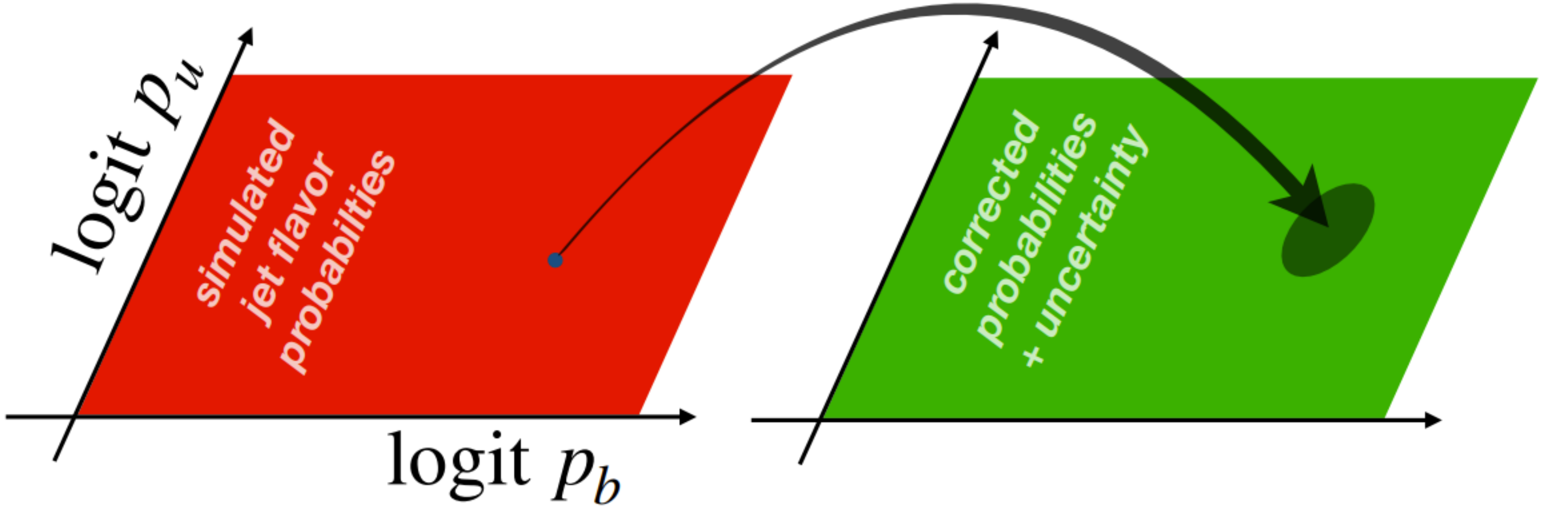
$$(\vec{\nabla}\, \hat{g})_{\#}\, p_{\mathrm{sim}}(\vec{q}|p_{\mathrm{T}}) \approx p_{\mathrm{data}}(\vec{q}|p_{\mathrm{T}})$$

$$(\vec{\nabla}\, \hat{f})_{\#}\, p_{\mathrm{data}}(\vec{q}|p_{\mathrm{T}}) \approx p_{\mathrm{sim}}(\vec{q}|p_{\mathrm{T}})$$

How to estimate uncertainties

Transport maps are deterministic (no stochasticity)

- Find a $T(\vec{\boldsymbol{q}}|p_T)$ for each variation

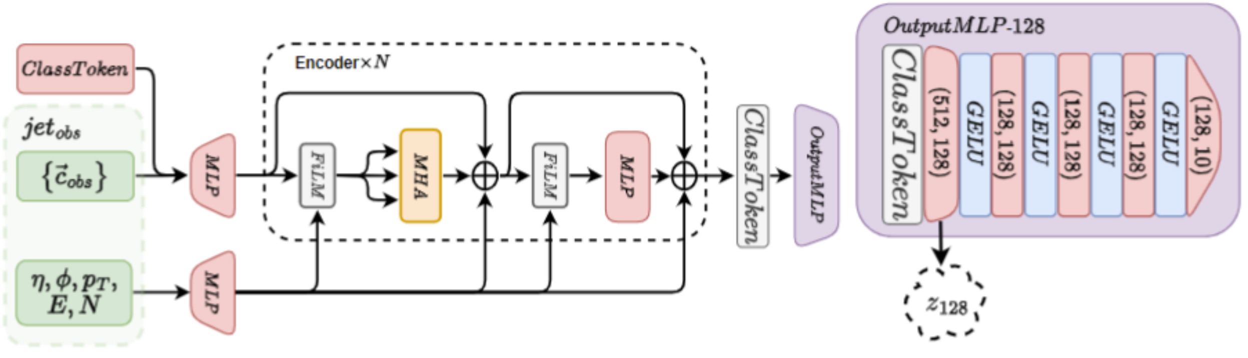


Latent calibration on JetClass

Setup for Latent calibration:

1. Train tagger on original JetClass

GN2 like architecture



(a)

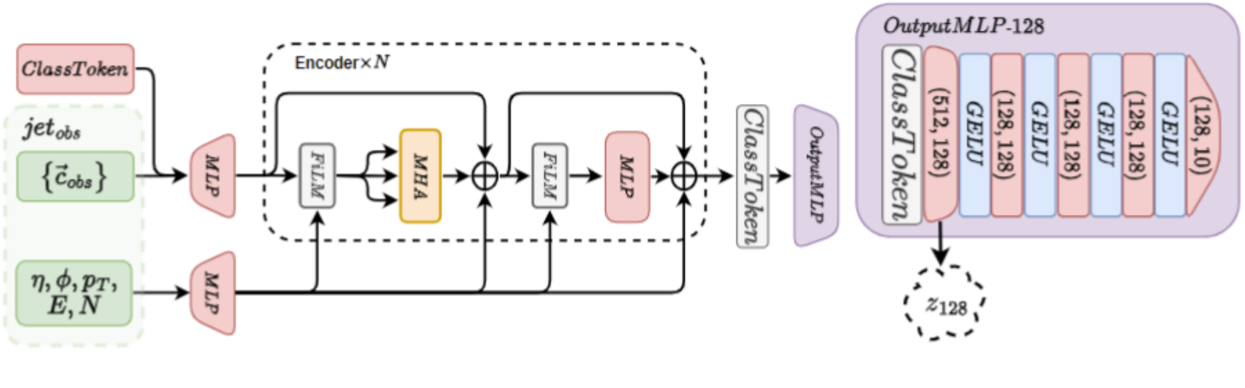
(b)

Latent calibration on JetClass

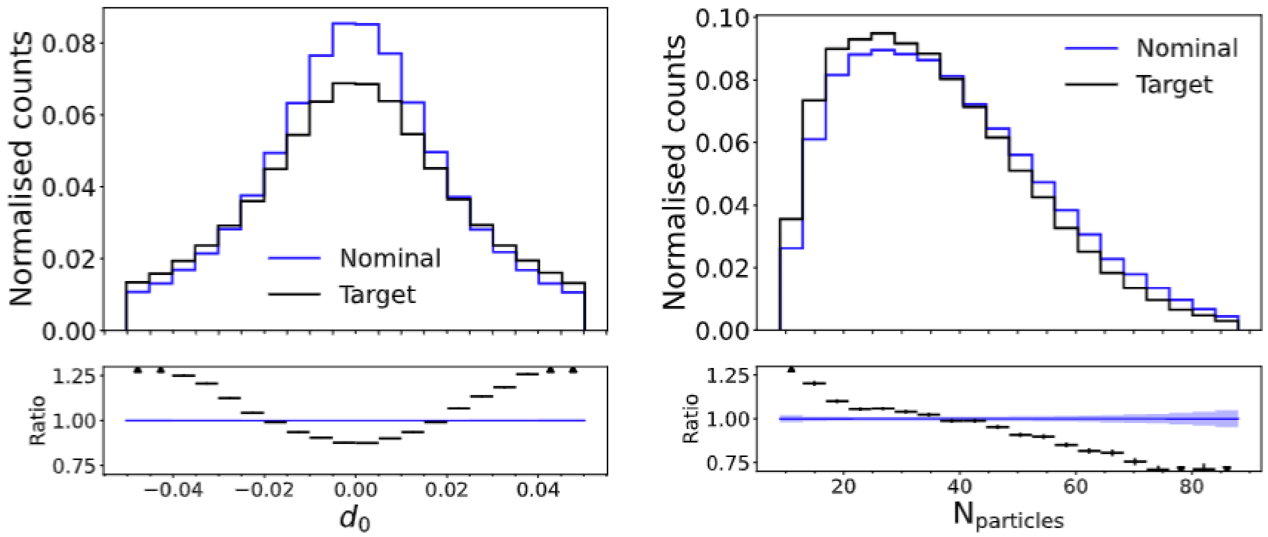
Setup for Latent calibration:

1. Train tagger on original JetClass
2. Simulate two variation of JetClass
    1. Nominal (MC)
    2. Target (Data)
    - Change:
        - Experimental effects
            - IP smearing
        - Theory parameters

GN2 like architecture



(a) (b)



(a) (b)