Normalising Flows for Particle Cloud Generation

Artwork by DALL – E · 2

IML Machine Learning Working Group, 11.10.2022

Benno Käch, Dirk Krücker, Isabell Melzer-Pellmann, Moritz Scham, Simon Schnake

benno.kaech@desy.de

HELMHOLTZAI

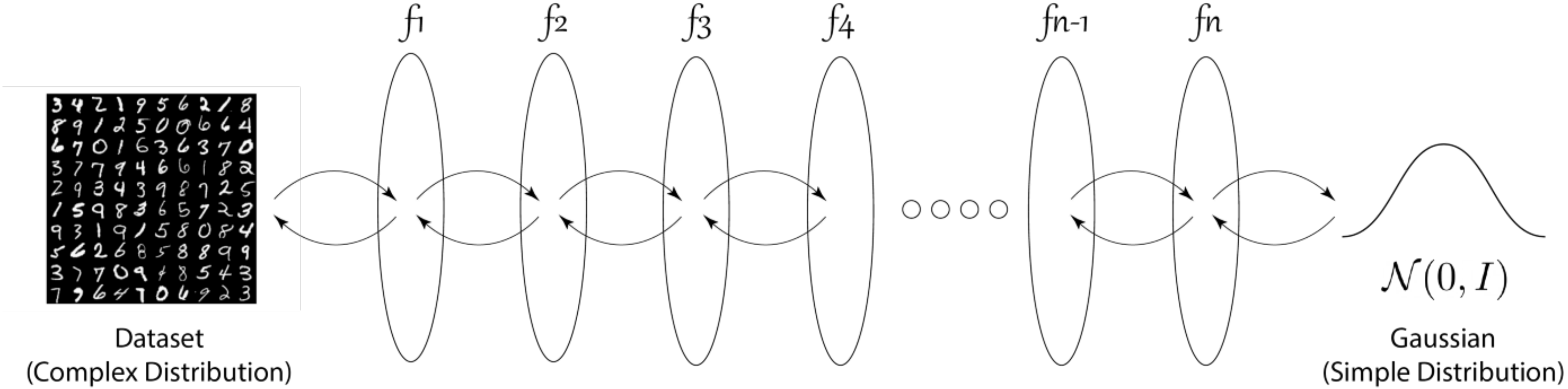
UH

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

DESY.

Normalising Flows

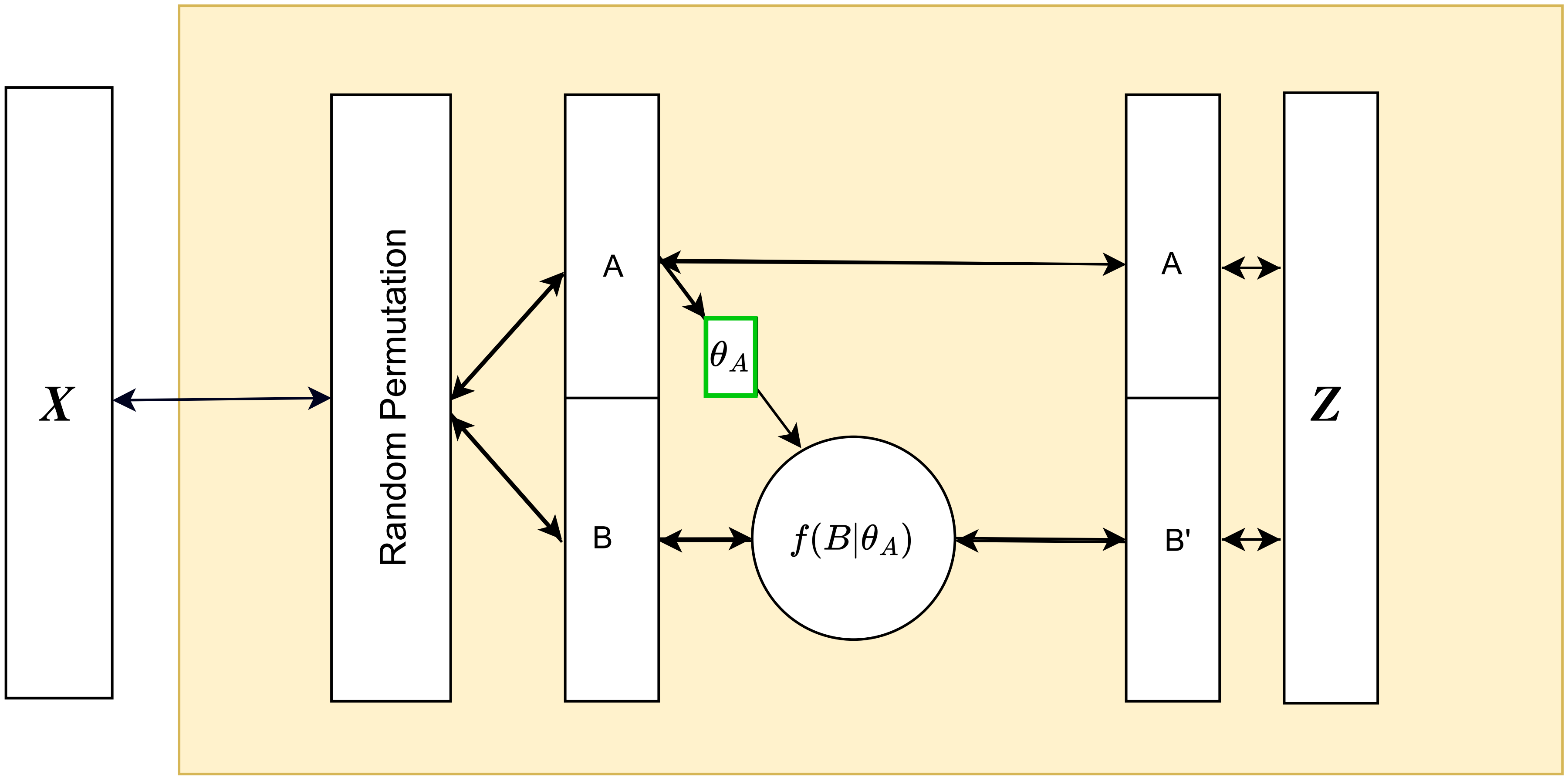
- *Find invertible functions to transform data distribution to Normal distribution*
- Invertible functions due to smart construction: **Coupling Layers**
- Stack multiple Coupling Layers for expressivity
- Contrast to GAN → Stable Maximum-Likelihood training



Coupling Layers

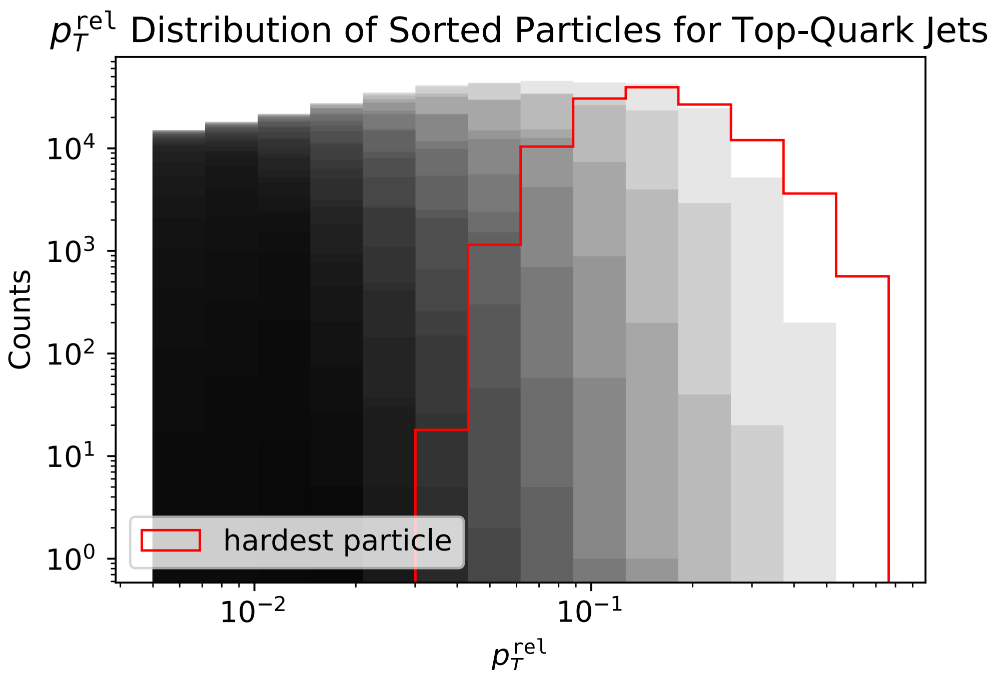
Invertibility by Construction

- Neural Networks for parameters θ of transformation f



Application: Particle Cloud Generation

JetNet [1] Datasets



- Gluon, light and top-quark Pythia jets, clustered by anti-k_T
- Jets of about $p_T^{\text{jet}} \sim 1\ TeV$
- Particles: tuples of $(\eta^{\text{rel}}, \phi^{\text{rel}}, p_T^{\text{rel}})$ relative to jet axis, massless
- p_T^{rel} normalised to total p_T of jet
- Constrained to max 30 particles/jet → 90 dimensions
- Invariant jet mass $m^2 = \sum_{i=0}^{30} \vec{p}_i^{\,2}$
- Size $\sim 150'000$ Samples
- $(70/30)$ Train/Test split

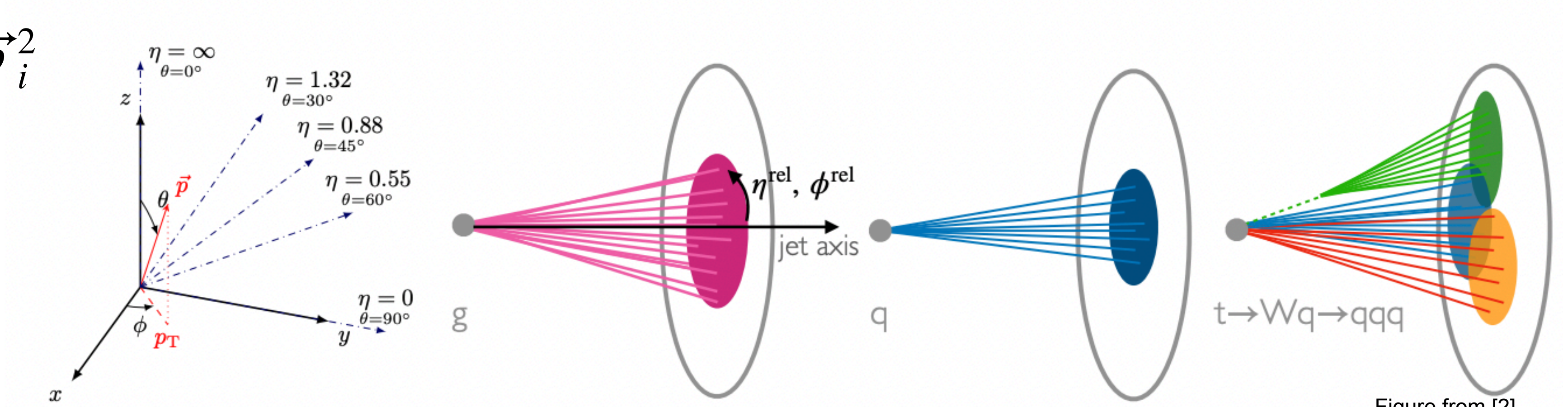
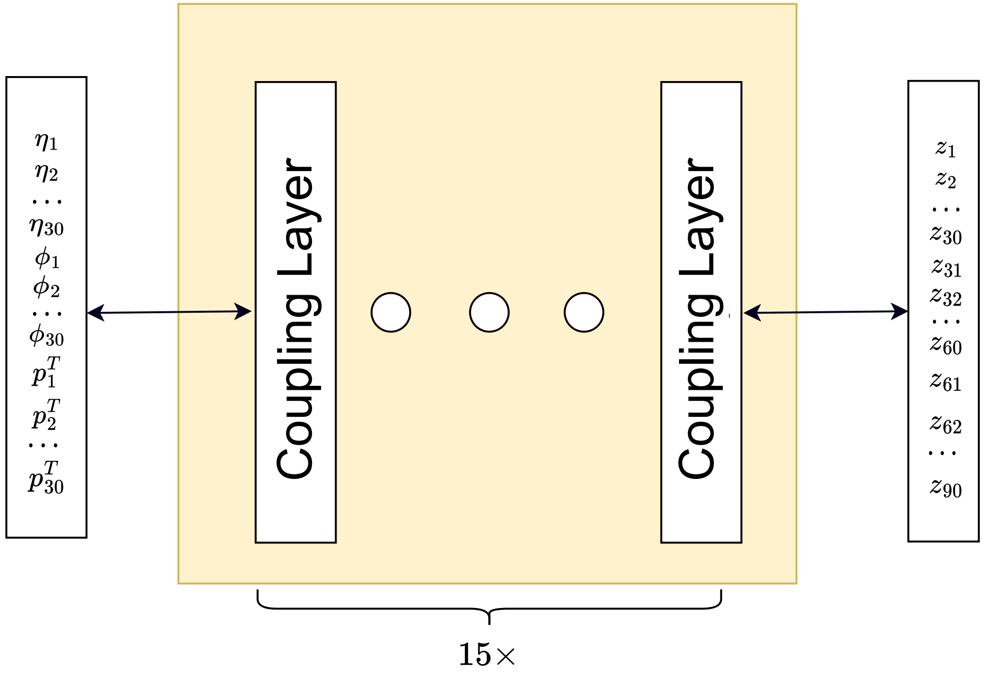


Figure from [2]

[1] Kansal et al, JetNet, PyPi
[2] Kansal et al., Particle Cloud Generation with Message Passing Generative Adversarial Networks, arxiv.org/abs/2106.11535

Normalising Flow Architecture & Training

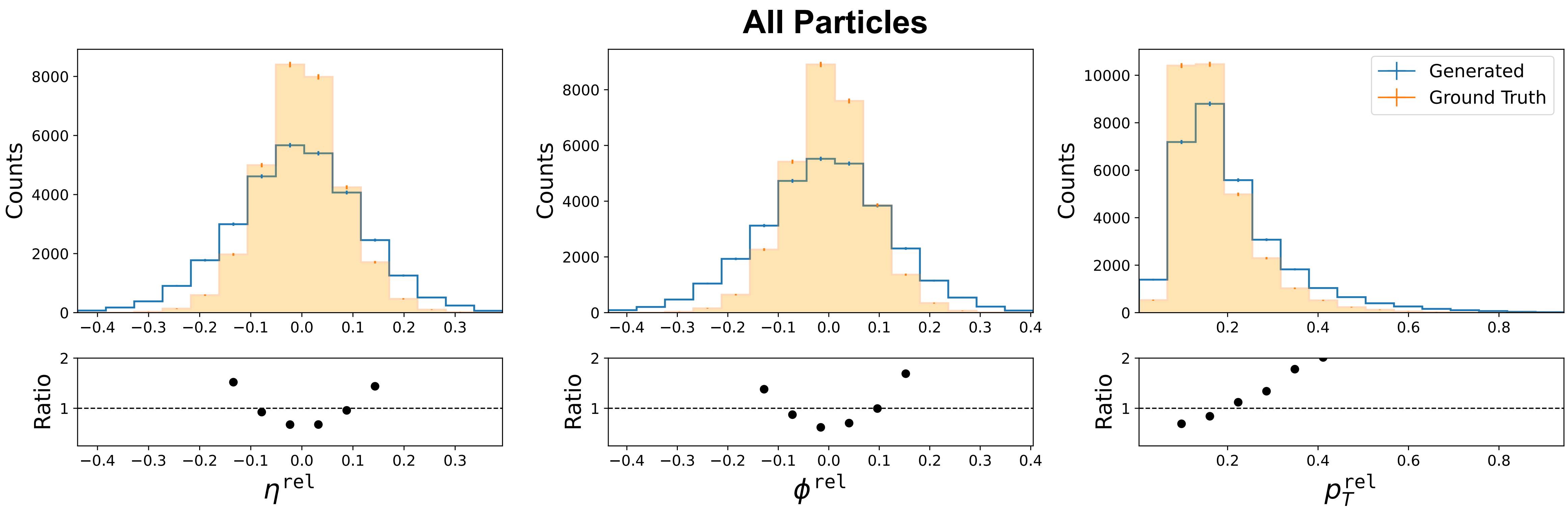
- Top-Quark → Complex jet substructure
- Flat Dense Normalising Flow: 90 dimensional latent space
- No permutation invariant encoding, particles ordered by p_T^{rel}
- Jets with less < 30 particles zero-padded & noise added $O(10^{-7})$
- **No inductive bias → contrast to other generative models**
- Adam Optimiser
- Batch Normalisation
- Dropout during training/evaluation



First Results

Implementation by `nflows` [2]

- Affine coupling layers
- 4 layer ResNets for coupling layer parameters
- Significant disagreement in inclusive marginals

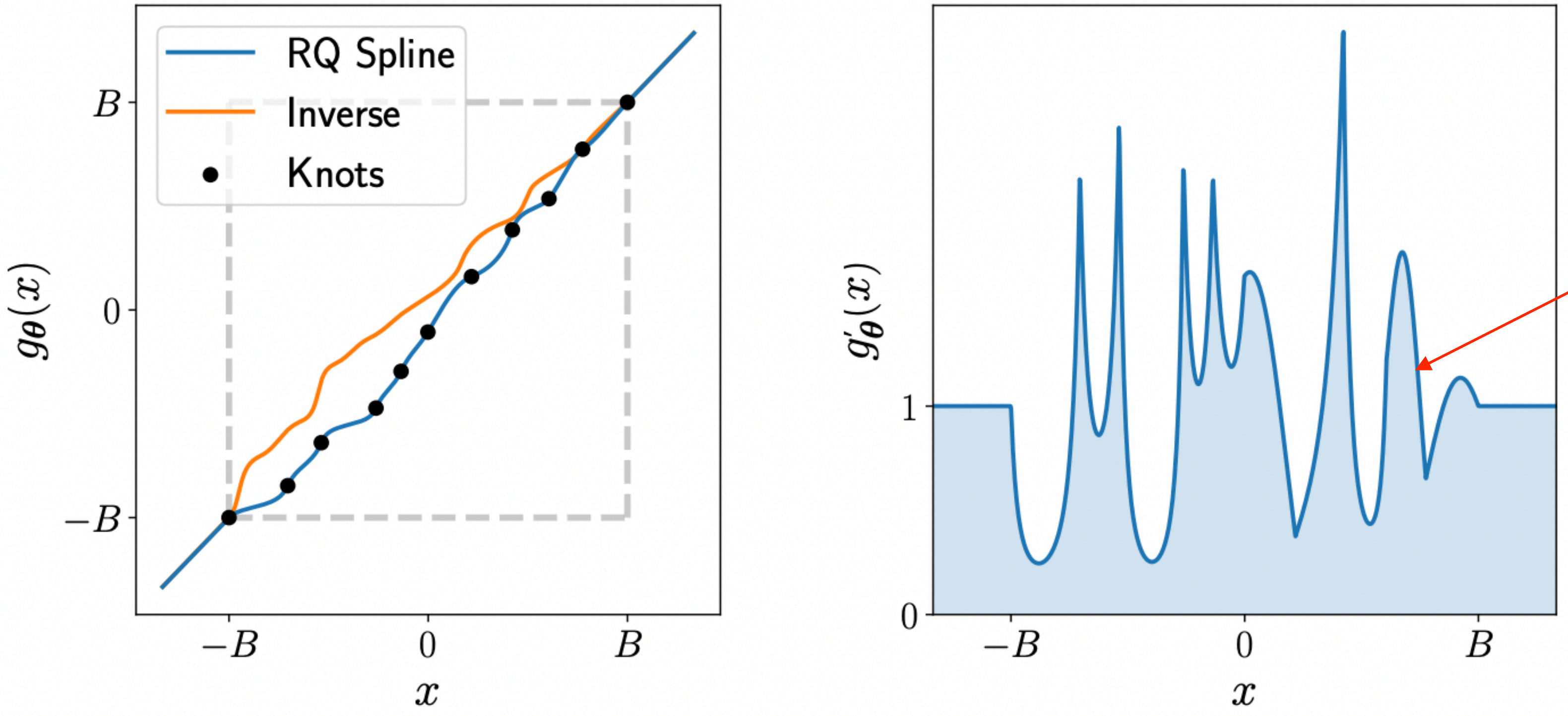


[2] Durkan and Bekasov et al., https://github.com/bayesiains/nflows

Rational Quadratic Spline Coupling

Proposed by Durkan and Bekasov et al. [3], also in `nflows` [2]

- Affine coupling lacks flexibility
- Element-wise monotonic ratio of quadratic splines
- Monotonic → analytically invertible
- K bins → $(3K-1)$ NN outputs per dimension



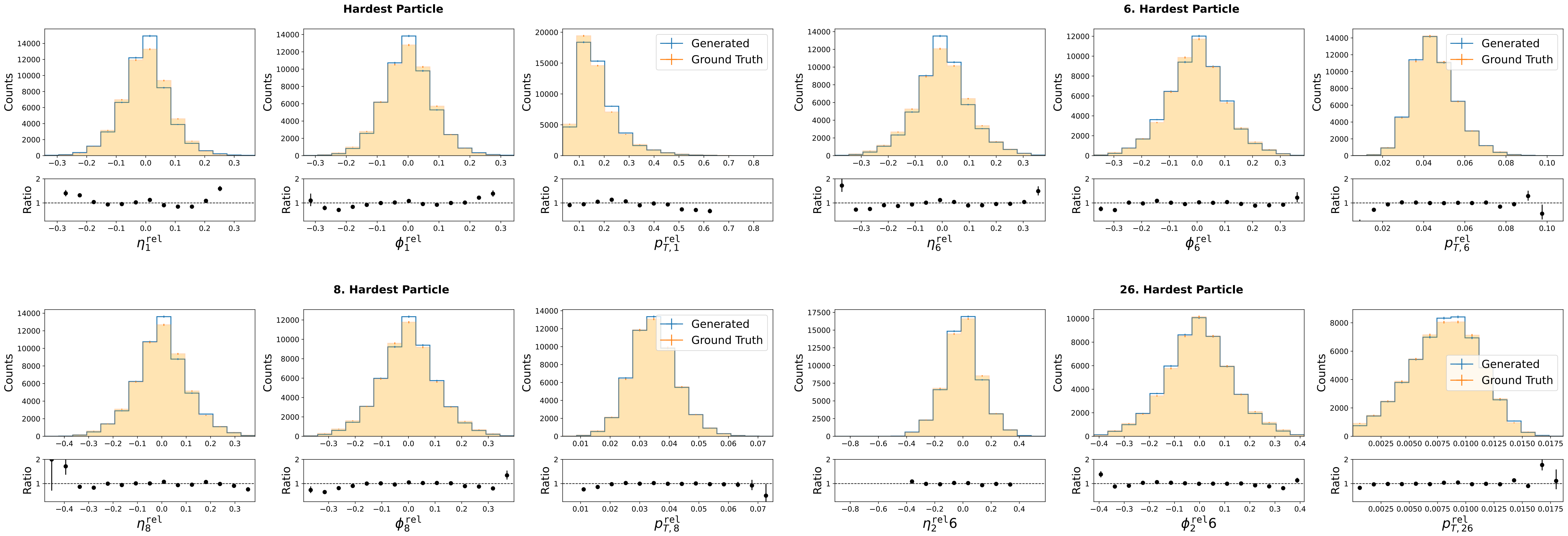
$$p_X(\boldsymbol{x}) = p_Z(\boldsymbol{f_\theta}(\boldsymbol{x}))\left|\det \frac{d\boldsymbol{f_\theta}}{d\boldsymbol{x}}\right|$$

[2] Durkan and Bekasov et al., https://github.com/bayesiains/nflows
[3] Durkan and Bekasov et al., Neural Spline Flows, arxiv.org/abs/1906.04032.pdf

Rational Quadratic Splines - Results

Distributions for Individual Particles

- Same architecture as before, RQS instead of Affine Coupling



<u>Generated individual particle distributions compatible with ground truth</u>

Assessing Performance

Same Metrics as in [2]

- Logprob well-motivated from theory → But devil is in the details
- Track multiple metrics for performance:
  - Wasserstein-1 distance W_1 on different distributions (see below)
  - Fréchet ParticleNet Distance (FPND) [2]
  - Coverage (COV)
  - Minimum Matching Distance (MMD)

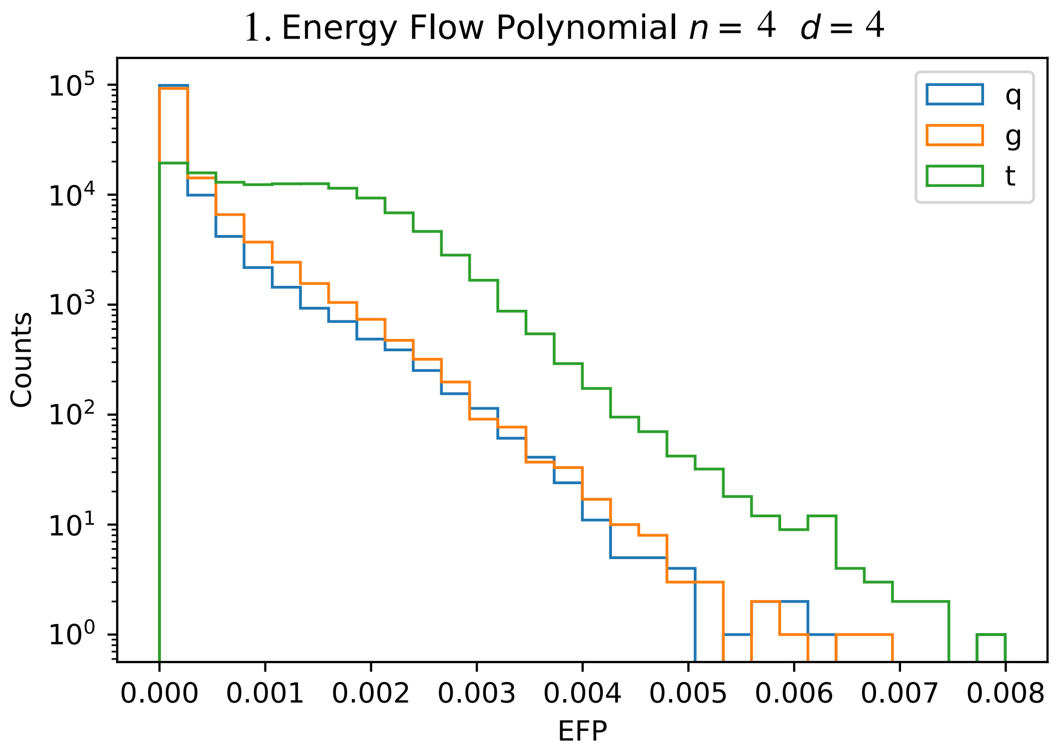
<u>In-sample distances</u>

Parton	$W_1^M (\times 10^{-3})$	$W_1^P (\times 10^{-3})$	$W_1^{EFP} (\times 10^{-5})$	FPND	COV $\uparrow$	MMD
Gluon	0.5 ± 0.1	0.4 ± 0.2	0.4 ± 0.4	0.01	0.56	0.036
Light Quark	0.42 ± 0.09	0.6 ± 0.4	0.5 ± 0.5	0.01	0.55	0.024
Top Quark	0.5 ± 0.1	0.6 ± 0.4	1.1 ± 0.4	0.03	0.56	0.072

[2] Kansal et al., Particle Cloud Generation with Message Passing Generative Adversarial Networks, arxiv.org/abs/2106.11535.pdf

Wasserstein Distance

- Metric on probability distributions
- Formally: $W_1(\mathbb{P}_r, \mathbb{P}_g) := \inf_{\gamma \in \Gamma(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(x,y)\sim\gamma}[\,|\boldsymbol{x} - \boldsymbol{y}|\,]$
- Not tractable for $\dim(\boldsymbol{X} \sim \mathbb{P}_g) > 1$
  - W_1^P: average of W_1 over (η, ϕ, p_T)
  - W_1^M: invariant jet mass
  - W_1^{EFP}: 5 Energy Flow Polynomials [4] (n=4,d=4)



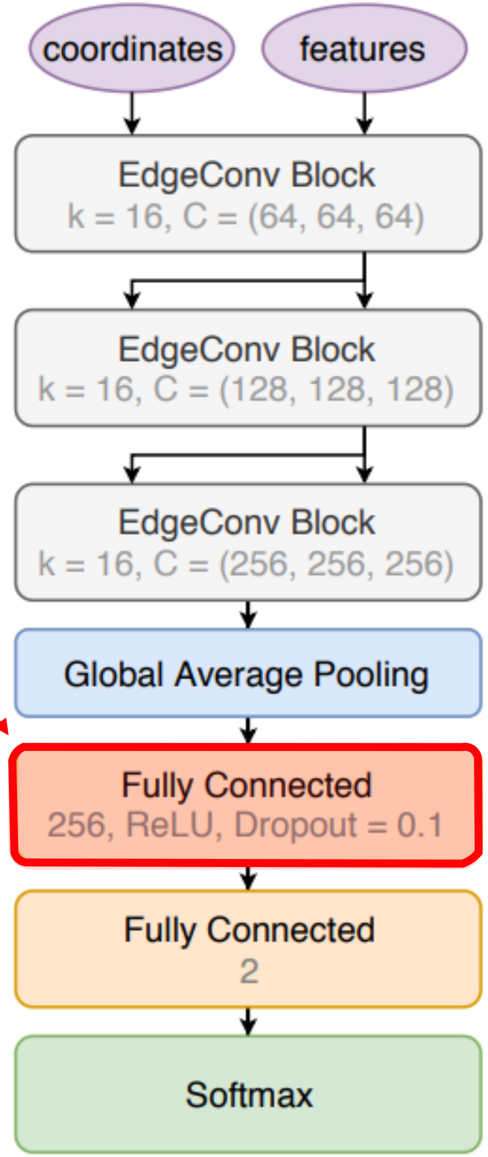
In-sample distances

Parton	$W_1^M (\times 10^{-3})$	$W_1^P (\times 10^{-3})$	$W_1^{EFP} (\times 10^{-5})$	FPND	COV $\uparrow$	MMD
Gluon	0.5 ± 0.1	0.4 ± 0.2	0.4 ± 0.4	0.01	0.56	0.036
Light Quark	0.42 ± 0.09	0.6 ± 0.4	0.5 ± 0.5	0.01	0.55	0.024
Top Quark	0.5 ± 0.1	0.6 ± 0.4	1.1 ± 0.4	0.03	0.56	0.072

[4] Komiske et al., Energy flow polynomials: A complete linear basis for jet sbstructure, arxiv.org/abs/1712.07124

Fréchet ParticleNet Distance (FPND) [2]

- Inspired from Fréchet Inception Distance (FID) for image generation [5]
- *Wasserstein-2 distance between Gaussians fitted to activations in first FC layer of ParticleNet* [6] *of MC & ML generated jets*
- Sensitive to output quality & mode collapse



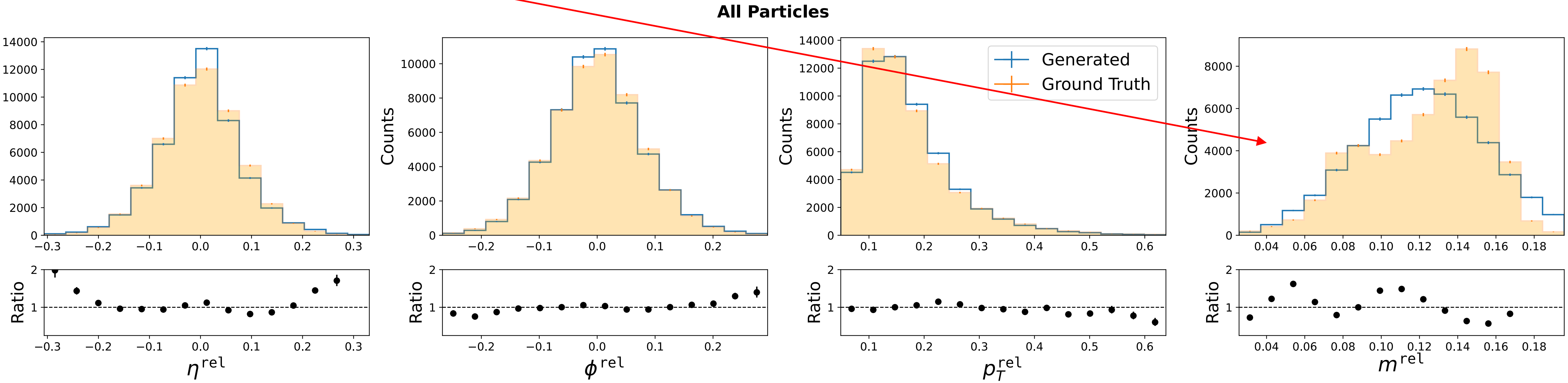
In-sample distances

Parton	$W_1^M (\times 10^{-3})$	$W_1^P (\times 10^{-3})$	$W_1^{EFP} (\times 10^{-5})$	FPND	COV $\uparrow$	MMD
Gluon	0.5 ± 0.1	0.4 ± 0.2	0.4 ± 0.4	0.01	0.56	0.036
Light Quark	0.42 ± 0.09	0.6 ± 0.4	0.5 ± 0.5	0.01	0.55	0.024
Top Quark	0.5 ± 0.1	0.6 ± 0.4	1.1 ± 0.4	0.03	0.56	0.072

[2] Kansal et al., Particle Cloud Generation with Message Passing Generative Adversarial Networks, arxiv.org/abs/2106.11535
[5] Heusel et al., GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium, arxiv.org/abs/1706.08500
[6] Qu et al., ParticleNet: Jet Tagging via Particle Clouds, arxiv.org/abs/1902.08570

Pitfall of Vanilla Normalising Flows

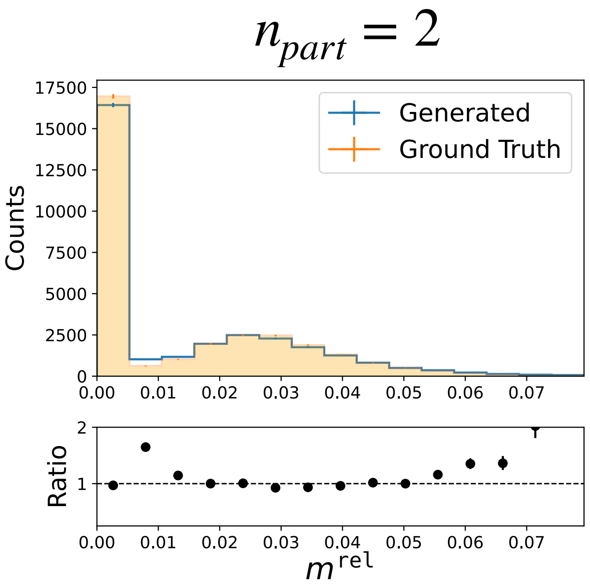
Model	$W_1^M (\times 10^{-3})$	$W_1^P (\times 10^{-3})$	$W_1^{EFP} (\times 10^{-5})$	FPND	COV ↑	MMD
VNF	6.4 ± 0.2	2.2 ± 0.2	14 ± 1	7.91	0.56	**0.071**

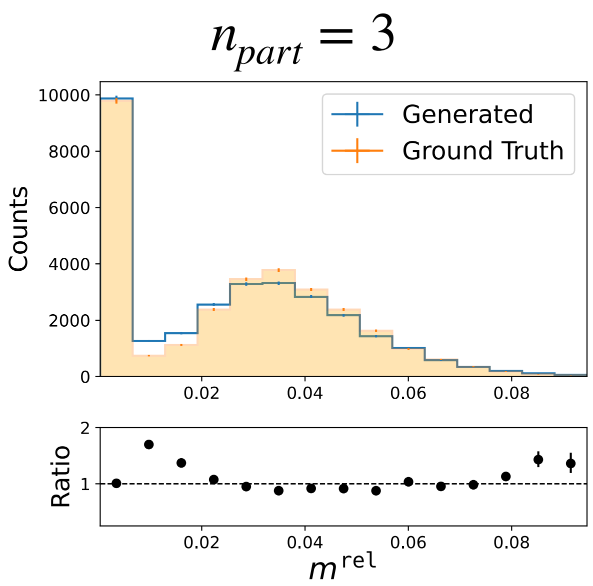


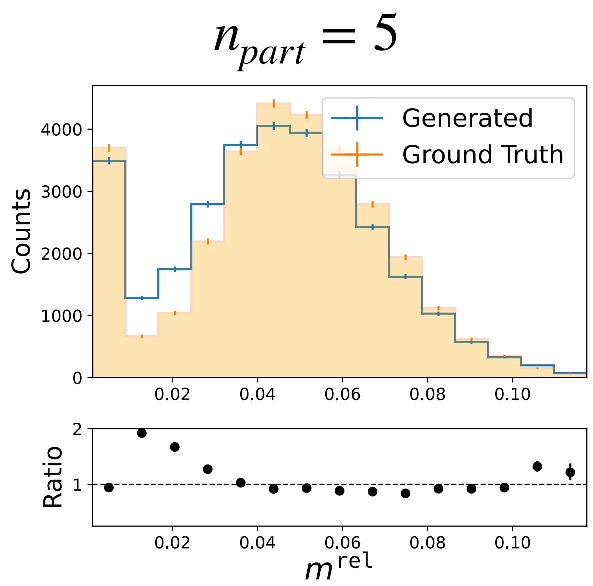
Mass Modelled Incorrectly

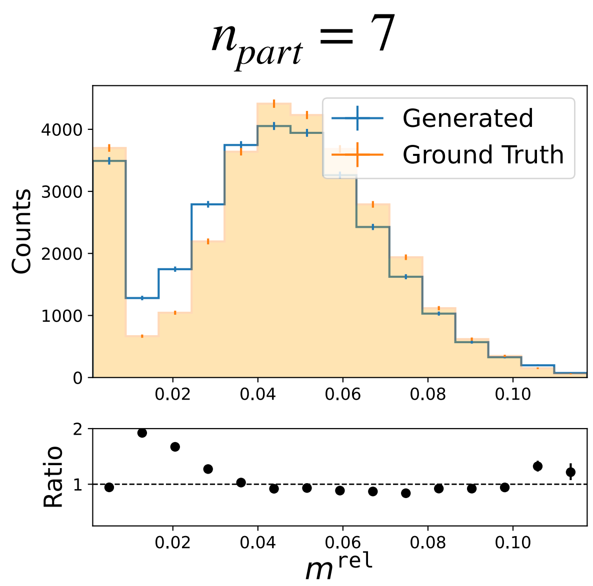
- Due to Coupling Layer construction?

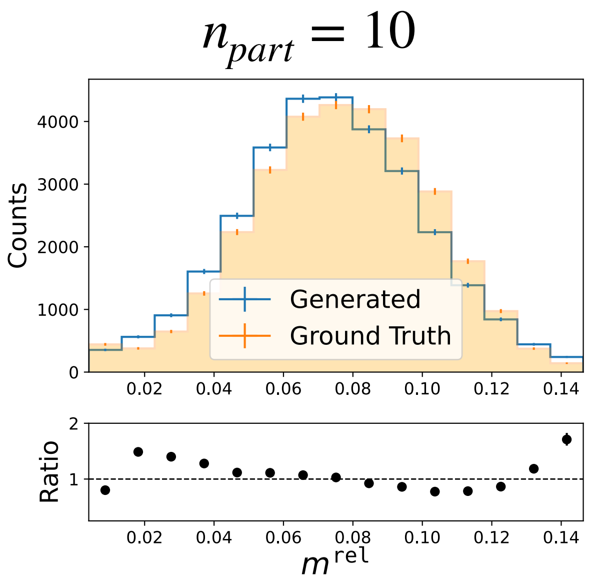
Dimensionality Scaling of Normalising Flows

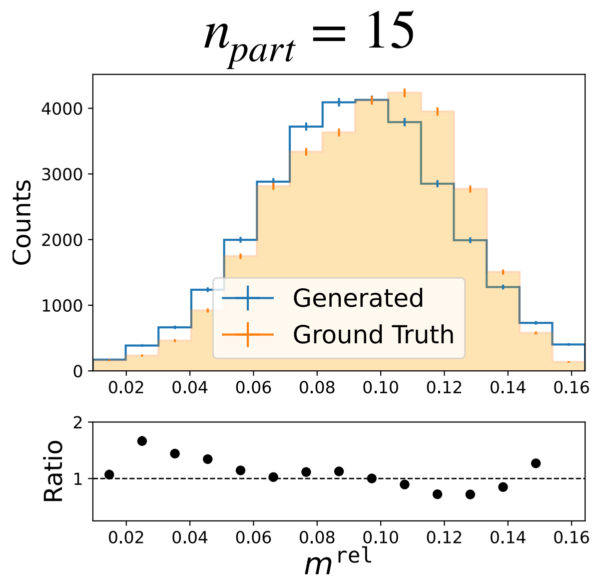


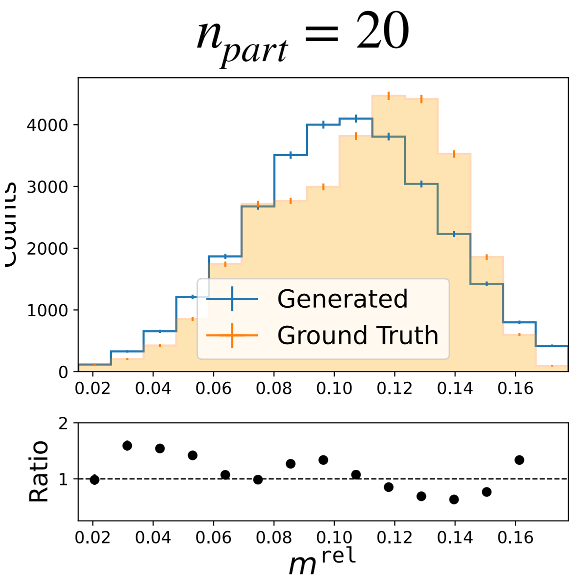








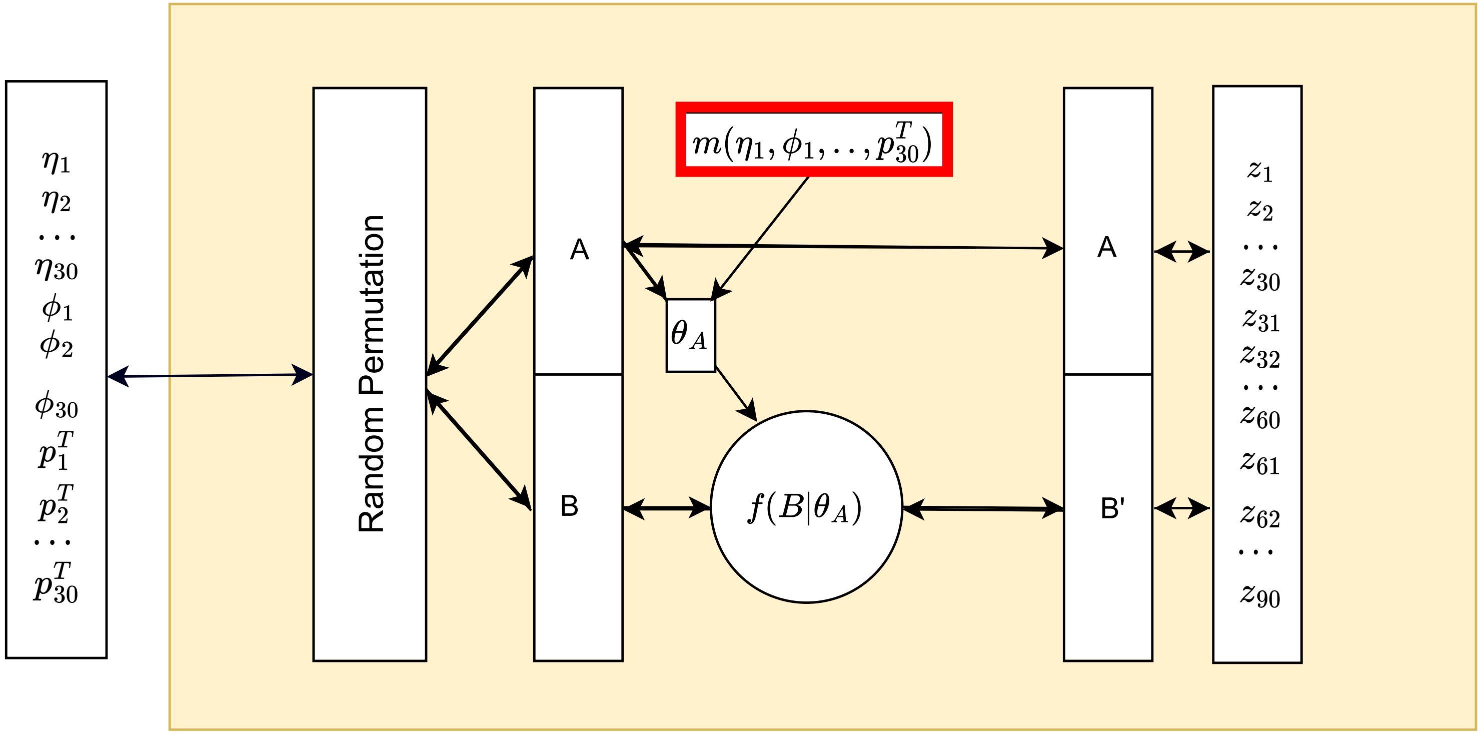




In low dimensions Normalising Flow captures correlations correctly

Conditioned Normalising Flows

- Add additional variables to parameter NN
- Increase expressivity of parameter NN
- **Needed during sampling**

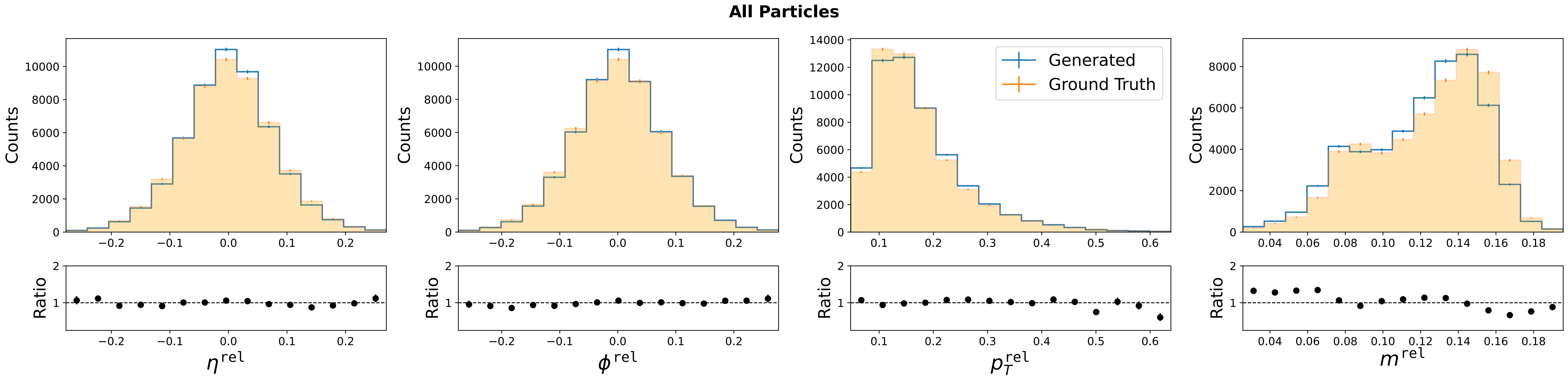


Sampling Conditioned Normalising Flows

- Mass is 1D variable → Universality of the uniform:
  one-dimensional *random variable transformed with its cumulative density function is uniformly distributed*
- Interpolate CDF with Monotone Piecewise Cubic Polynomials [7] → invertible
- Jet Generation:
  1. Sample 1D uniform
  2. Apply inverted monotonic cubic interpolation → $m_{\texttt{cond}}$
  3. Sample $D = 90$ Normal → $\boldsymbol{z}$
  4. Generate jets $\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{z} \,|\, m_{\texttt{cond}})$

Conditioned Normalising Flows

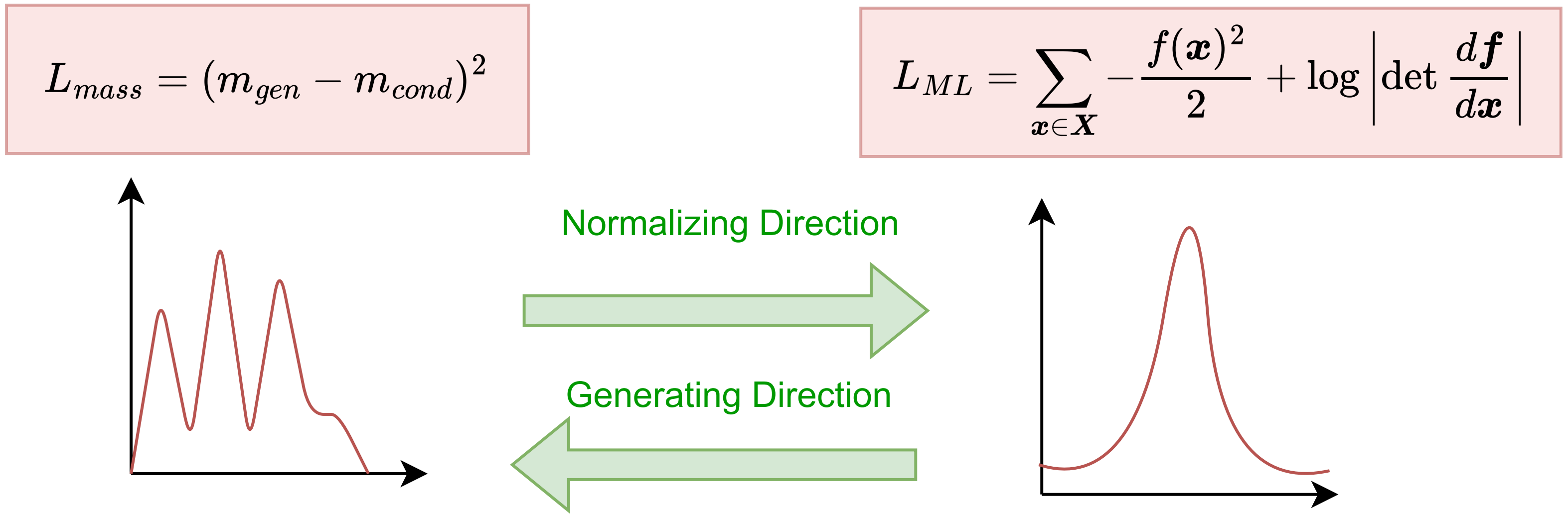
- Fixes mass modelling
- Improves nearly every other metric



Model	$W_1^M(\times 10^{-3})$	$W_1^P(\times 10^{-3})$	$W_1^{EFP}(\times 10^{-5})$	FPND	COV ↑	MMD
VNF	6.4 ± 0.2	2.2 ± 0.2	14 ± 1	7.91	0.56	**0.071**
CNF (m)	1.7 ± 0.3	$\mathbf{1.1 \pm 0.3}$	5 ± 1	4.53	0.55	0.073

Mass Constraint

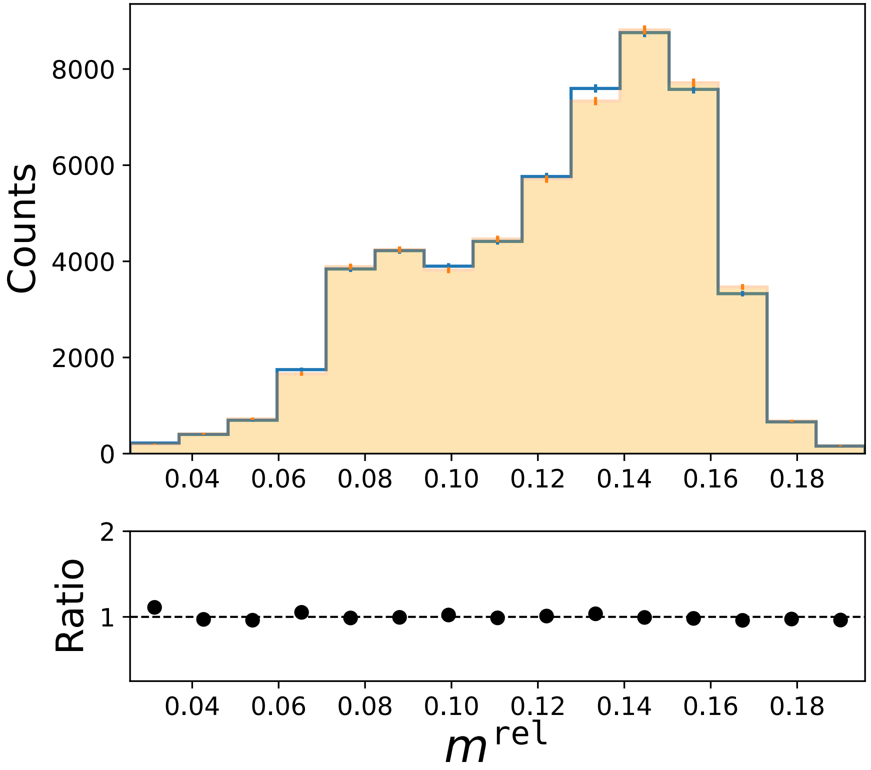
- *Constrain Normalising Flow to produce jet with mass same as condition*
  - Generate jet $\boldsymbol{x}_{\text{gen}} = \boldsymbol{f}(z \,|\, m_{\text{cond}})$
  - Calculate jet mass $m_{\text{gen}} = m(\boldsymbol{x}_{\text{gen}})$
  - Add loss $L = ||m_{\text{gen}} - m_{\text{cond}}||^2$ to nLL loss
- **During training Normalising Flow trained using both directions**

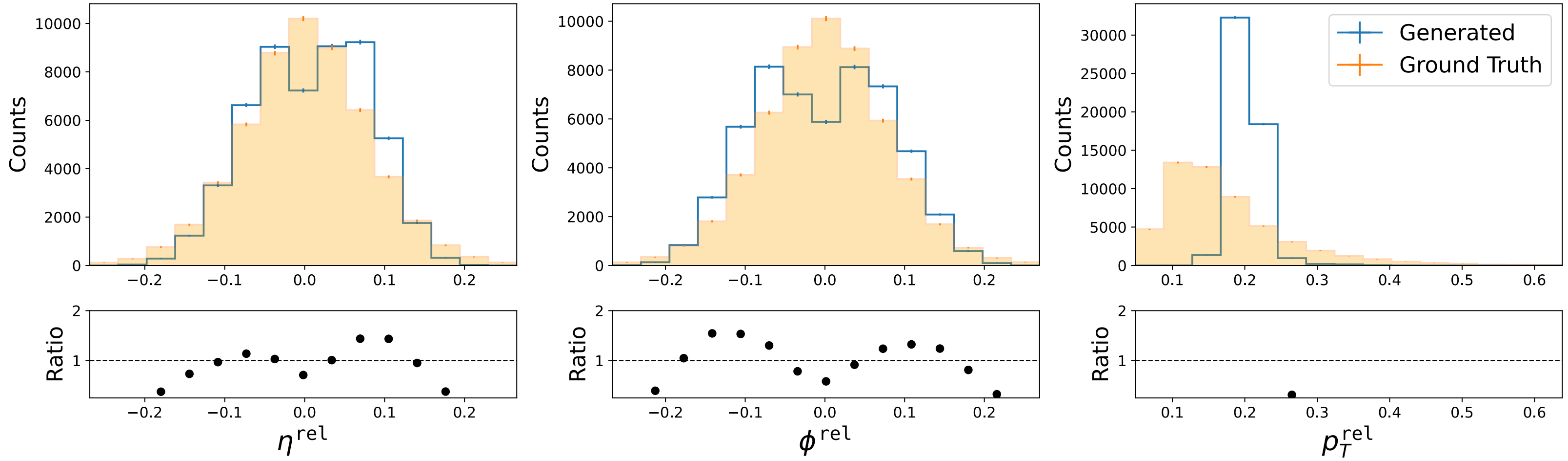


Constrained & Conditioned Normalising Flows

- Perfect mass modelling
- Tradeoff with other metrics → Hyperparameter?

Mass Distribution

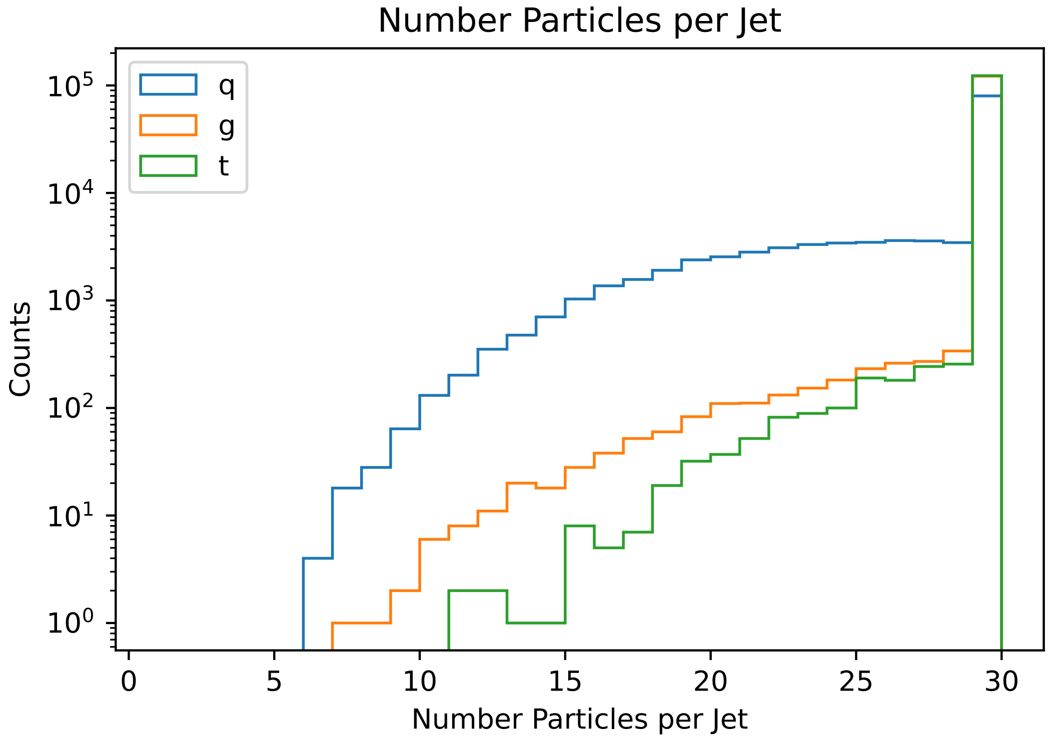




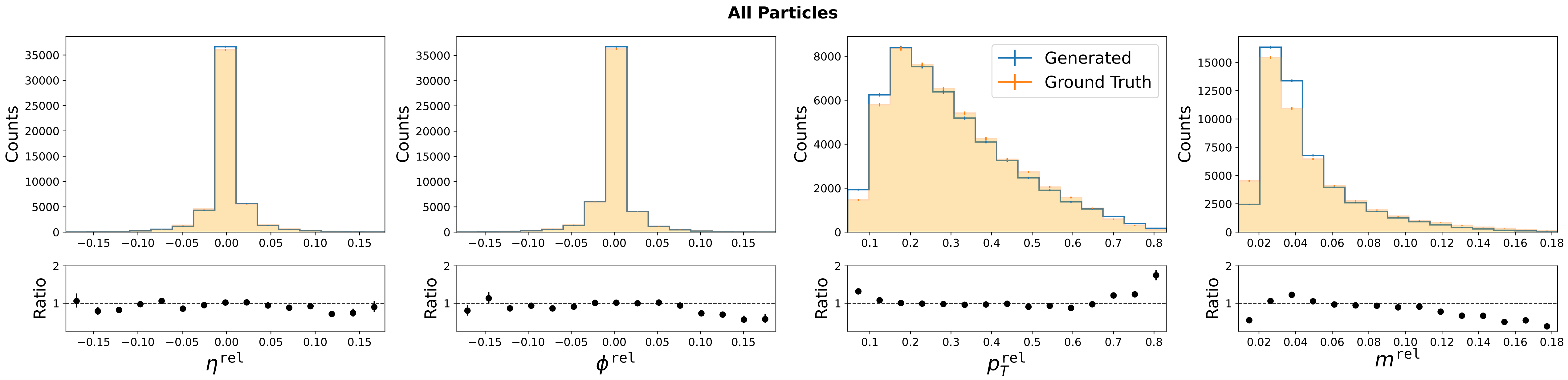
Model	$W_1^M (\times 10^{-3})$	$W_1^P (\times 10^{-3})$	$W_1^{EFP} (\times 10^{-5})$	FPND	COV ↑	MMD
VNF	6.4 ± 0.2	2.2 ± 0.2	14 ± 1	7.91	0.56	**0.071**
CNF (m)	1.7 ± 0.3	$\mathbf{1.1 \pm 0.3}$	5 ± 1	4.53	0.55	0.073
CCNF (m)	0.7 ± 0.3	12.2 ± 0.1	9 ± 1	9.16	0.38	0.083

But Point Clouds Have Variable Sizes?

- Introduce number particles in jet as 2^{nd} condition
- Switch datasets: $99\,\%$ of Top-quark jets have 30 particles → Light Quarks
- Particles $[k+1,...,n]$ set to zero for jet with $n_{cond} = k$
- Sampling Condition: Number particles discrete → Decompose $p(m,n) = p(n)p(m\,|\,n)$

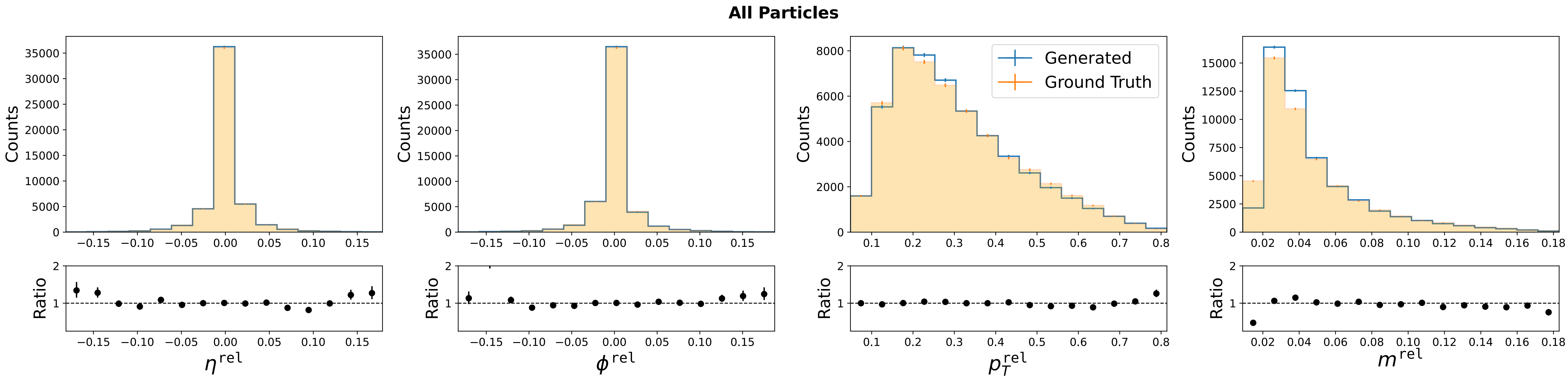


Results on Light Quark Initiated Jets - No Condition



Model	$W_1^M (\times 10^{-3})$	$W_1^P (\times 10^{-3})$	$W_1^{EFP} (\times 10^{-5})$	FPND	COV ↑	MMD
VNF	3.5 ± 0.3	2.3 ± 0.4	3 ± 1	2.10	0.54	0.025

Results on Light Quark Initiated Jets - Two Conditions



Model	$W_1^M(\times 10^{-3})$	$W_1^P(\times 10^{-3})$	$W_1^{EFP}(\times 10^{-5})$	FPND	COV $\uparrow$	MMD
VNF	3.5 ± 0.3	2.3 ± 0.4	3 ± 1	2.10	0.54	0.025
CNF (m)	0.9 ± 0.1	0.6 ± 0.3	0.9 ± 0.5	0.68	0.54	0.036
CNF (m,n)	1.8 ± 0.2	0.9 ± 0.5	0.9 ± 0.4	0.64	0.54	0.025

→ Conditioning with number particles → No significant improvement

Results

Jet Class	Model	$W_1^M (\times 10^{-3})$	$W_1^P (\times 10^{-3})$	$W_1^{EFP} (\times 10^{-5})$	FPND	COV ↑	MMD
Gluon	MP-MP	0.7 ± 0.2	0.9 ± 0.3	$\mathbf{0.7 \pm 0.7}$	**0.12**	**0.56**	0.037
	MP_LFC-MP	0.69 ± 0.07	1.8 ± 0.3	0.9 ± 0.2	0.20	0.54	0.037
	VNF	5.5 ± 0.7	2.9 ± 0.7	4.4 ± 0.8	2.32	0.54	0.035
	CNF (m)	0.9 ± 0.1	0.6 ± 0.3	0.9 ± 0.5	0.68	0.54	0.036
	CNF (m,n)	0.9 ± 0.3	0.7 ± 0.2	0.9 ± 0.5	0.70	0.55	**0.034**
	CCNF (m)	0.9 ± 0.2	$\mathbf{0.5 \pm 0.2}$	1.1 ± 0.5	1.10	0.55	0.036
	CCNF (m,n)	$\mathbf{0.6 \pm 0.5}$	0.8 ± 0.1	0.9 ± 0.8	0.72	**0.56**	0.035
Light Quark	MP-MP	$\mathbf{0.6 \pm 0.2}$	4.9 ± 0.5	$\mathbf{0.7 \pm 0.4}$	0.35	0.50	0.026
	MP_LFC-MP	0.7 ± 0.2	2.6 ± 0.4	0.9 ± 0.9	**0.08**	0.52	0.037
	VNF	3.5 ± 0.3	2.3 ± 0.4	3 ± 1	2.10	0.54	0.025
	CNF (m)	1.1 ± 0.2	1.3 ± 0.3	$\mathbf{0.7 \pm 0.3}$	0.78	**0.55**	**0.024**
	CNF (m,n)	1.8 ± 0.2	0.9 ± 0.5	0.9 ± 0.4	0.64	0.54	0.025
	CCNF (m)	0.9 ± 0.3	$\mathbf{0.6 \pm 0.2}$	$\mathbf{0.7 \pm 0.3}$	0.71	0.54	**0.024**
	CCNF (m,n)	1.0 ± 0.3	1.2 ± 0.4	0.8 ± 0.5	1.99	0.51	0.026
Top Quark	MP-MP	$\mathbf{0.6 \pm 0.2}$	2.3 ± 0.3	$\mathbf{2 \pm 1}$	**0.37**	**0.57**	**0.071**
	MP_LFC-MP	0.9 ± 0.3	2.2 ± 0.7	$\mathbf{2 \pm 1}$	0.93	0.56	0.073
	VNF	6.4 ± 0.2	2.2 ± 0.2	14 ± 1	7.91	0.56	**0.071**
	CNF (m)	1.7 ± 0.3	$\mathbf{1.1 \pm 0.3}$	5 ± 1	4.53	0.55	0.073
	CNF (m,n)	0.9 ± 0.5	1.5 ± 0.1	7 ± 2	3.46	0.56	**0.071**
	CCNF (m)	0.7 ± 0.3	12.2 ± 0.1	9 ± 1	9.16	0.38	0.083
	CCNF (m,n)	1.1 ± 0.5	3.7 ± 0.3	13 ± 3	6.34	0.55	0.073

Results - Strength

Jet Class	Model	$W_1^M (\times 10^{-3})$	$W_1^P (\times 10^{-3})$	$W_1^{EFP} (\times 10^{-5})$	FPND	COV ↑	MMD
Gluon	MP-MP	0.7 ± 0.2	0.9 ± 0.3	$\mathbf{0.7 \pm 0.7}$	**0.12**	**0.56**	0.037
	MP_LFC-MP	0.69 ± 0.07	1.8 ± 0.3	0.9 ± 0.2	0.20	0.54	0.037
	VNF	5.5 ± 0.7	2.9 ± 0.7	4.4 ± 0.8	2.32	0.54	0.035
	CNF (m)	0.9 ± 0.1	0.6 ± 0.3	0.9 ± 0.5	0.68	0.54	0.036
	CNF (m,n)	0.9 ± 0.3	0.7 ± 0.2	0.9 ± 0.5	0.70	0.55	**0.034**
	CCNF (m)	0.9 ± 0.2	$\mathbf{0.5 \pm 0.2}$	1.1 ± 0.5	1.10	0.55	0.036
	CCNF (m,n)	$\mathbf{0.6 \pm 0.5}$	0.8 ± 0.1	0.9 ± 0.8	0.72	**0.56**	0.035
Light Quark	MP-MP	$\mathbf{0.6 \pm 0.2}$	4.9 ± 0.5	$\mathbf{0.7 \pm 0.4}$	0.35	0.50	0.026
	MP_LFC-MP	0.7 ± 0.2	2.6 ± 0.4	0.9 ± 0.9	**0.08**	0.52	0.037
	VNF	3.5 ± 0.3	2.3 ± 0.4	3 ± 1	2.10	0.54	0.025
	CNF (m)	1.1 ± 0.2	1.3 ± 0.3	$\mathbf{0.7 \pm 0.3}$	0.78	**0.55**	**0.024**
	CNF (m,n)	1.8 ± 0.2	0.9 ± 0.5	0.9 ± 0.4	0.64	0.54	0.025
	CCNF (m)	0.9 ± 0.3	$\mathbf{0.6 \pm 0.2}$	$\mathbf{0.7 \pm 0.3}$	0.71	0.54	**0.024**
	CCNF (m,n)	1.0 ± 0.3	1.2 ± 0.4	0.8 ± 0.5	1.99	0.51	0.026
Top Quark	MP-MP	$\mathbf{0.6 \pm 0.2}$	2.3 ± 0.3	$\mathbf{2 \pm 1}$	**0.37**	**0.57**	**0.071**
	MP_LFC-MP	0.9 ± 0.3	2.2 ± 0.7	$\mathbf{2 \pm 1}$	0.93	0.56	0.073
	VNF	6.4 ± 0.2	2.2 ± 0.2	14 ± 1	7.91	0.56	**0.071**
	CNF (m)	1.7 ± 0.3	$\mathbf{1.1 \pm 0.3}$	5 ± 1	4.53	0.55	0.073
	CNF (m,n)	0.9 ± 0.5	1.5 ± 0.1	7 ± 2	3.46	0.56	**0.071**
	CCNF (m)	0.7 ± 0.3	12.2 ± 0.1	9 ± 1	9.16	0.38	0.083
	CCNF (m,n)	1.1 ± 0.5	3.7 ± 0.3	13 ± 3	6.34	0.55	0.073

Normalising Flows handle marginal distributions well

Results - Weaknesses

Jet Class	Model	$W_1^M (\times 10^{-3})$	$W_1^P (\times 10^{-3})$	$W_1^{EFP} (\times 10^{-5})$	FPND	COV ↑	MMD
Gluon	MP-MP	0.7 ± 0.2	0.9 ± 0.3	$\mathbf{0.7 \pm 0.7}$	**0.12**	**0.56**	0.037
	MP_LFC-MP	0.69 ± 0.07	1.8 ± 0.3	0.9 ± 0.2	0.20	0.54	0.037
	VNF	5.5 ± 0.7	2.9 ± 0.7	4.4 ± 0.8	2.32	0.54	0.035
	CNF (m)	0.9 ± 0.1	0.6 ± 0.3	0.9 ± 0.5	0.68	0.54	0.036
	CNF (m,n)	0.9 ± 0.3	0.7 ± 0.2	0.9 ± 0.5	0.70	0.55	**0.034**
	CCNF (m)	0.9 ± 0.2	$\mathbf{0.5 \pm 0.2}$	1.1 ± 0.5	1.10	0.55	0.036
	CCNF (m,n)	$\mathbf{0.6 \pm 0.5}$	0.8 ± 0.1	0.9 ± 0.8	0.72	**0.56**	0.035
Light Quark	MP-MP	$\mathbf{0.6 \pm 0.2}$	4.9 ± 0.5	$\mathbf{0.7 \pm 0.4}$	0.35	0.50	0.026
	MP_LFC-MP	0.7 ± 0.2	2.6 ± 0.4	0.9 ± 0.9	**0.08**	0.52	0.037
	VNF	3.5 ± 0.3	2.3 ± 0.4	3 ± 1	2.10	0.54	0.025
	CNF (m)	1.1 ± 0.2	1.3 ± 0.3	$\mathbf{0.7 \pm 0.3}$	0.78	**0.55**	**0.024**
	CNF (m,n)	1.8 ± 0.2	0.9 ± 0.5	0.9 ± 0.4	0.64	0.54	0.025
	CCNF (m)	0.9 ± 0.3	$\mathbf{0.6 \pm 0.2}$	$\mathbf{0.7 \pm 0.3}$	0.71	0.54	**0.024**
	CCNF (m,n)	1.0 ± 0.3	1.2 ± 0.4	0.8 ± 0.5	1.99	0.51	0.026
Top Quark	MP-MP	$\mathbf{0.6 \pm 0.2}$	2.3 ± 0.3	$\mathbf{2 \pm 1}$	**0.37**	**0.57**	**0.071**
	MP_LFC-MP	0.9 ± 0.3	2.2 ± 0.7	$\mathbf{2 \pm 1}$	0.93	0.56	0.073
	VNF	6.4 ± 0.2	2.2 ± 0.2	14 ± 1	7.91	0.56	**0.071**
	CNF (m)	1.7 ± 0.3	$\mathbf{1.1 \pm 0.3}$	5 ± 1	4.53	0.55	0.073
	CNF (m,n)	0.9 ± 0.5	1.5 ± 0.1	7 ± 2	3.46	0.56	**0.071**
	CCNF (m)	0.7 ± 0.3	12.2 ± 0.1	9 ± 1	9.16	0.38	0.083
	CCNF (m,n)	1.1 ± 0.5	3.7 ± 0.3	13 ± 3	6.34	0.55	0.073

But still significantly worse on some metrics

Summary

- Normalising Flows light, quick & stable alternative to GANs
- Training duration $\sim 1-2\ h$ on $\texttt{NVIDIA P100}$
- Significant differences on FPND
- Difficulties in modelling global features
- Conditioning enhances expressivity
- Normalising Flows handle variable number particles well
- Mass Constraint → mass modelling ✅ overall performance 📉

Any Questions?

Backup

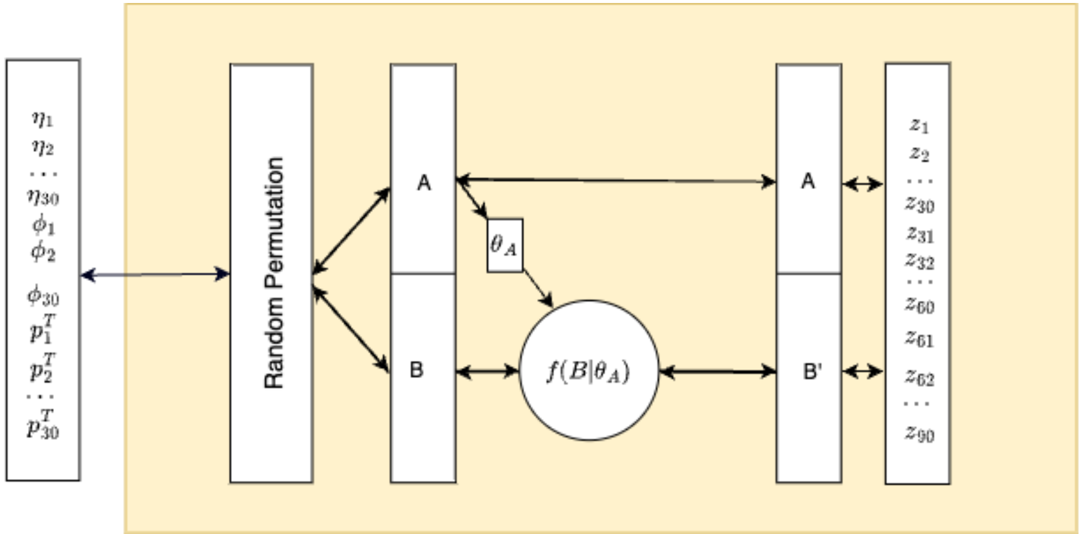
Coupling Layers

How to Construct Invertible Functions with a Tractable Determinant

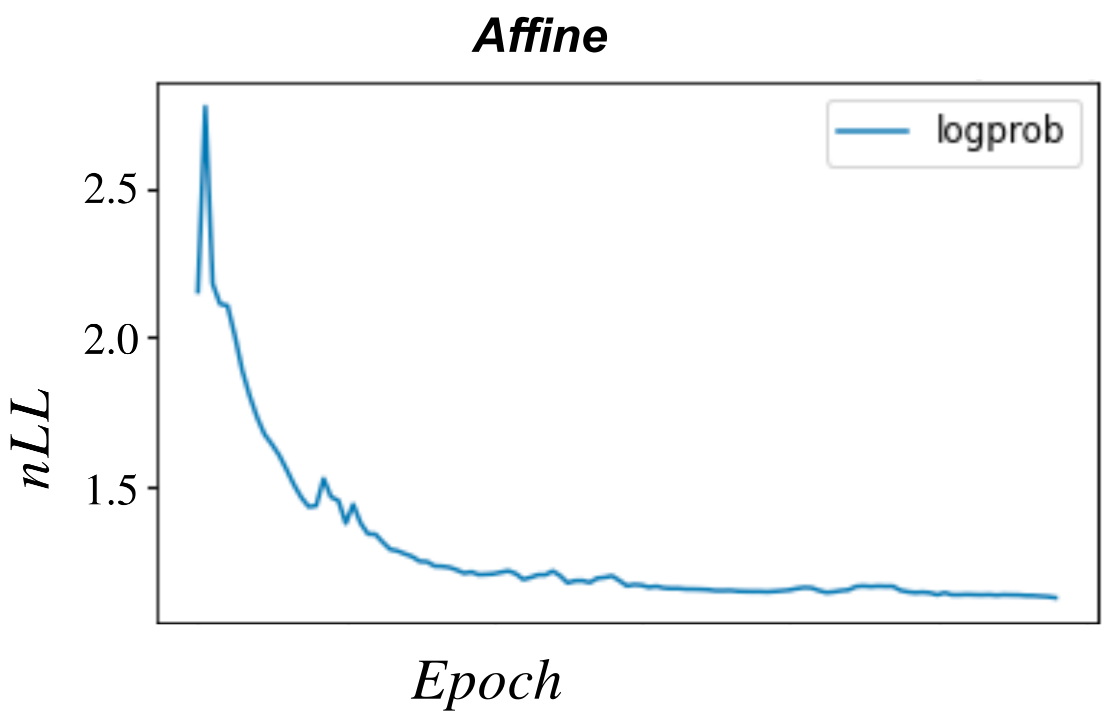
- Partition input into $(\mathbf{x}^A, \mathbf{x}^B) \in \mathbb{R}^d \times \mathbb{R}^{D-d}, D = 90,\ d = 45$
- Construct map element-wise: $f_\theta(x) = \begin{cases} z_i^A = x_i^A \\ z_i^B = s_{\theta(x^A)}(x_i^B) \end{cases} \Leftrightarrow f_\theta^{-1}(z) = \begin{cases} x_i^A = z_i^A \\ x_i^B = s^{-1}_{\theta(z^A)}(z_i^B) \end{cases}$

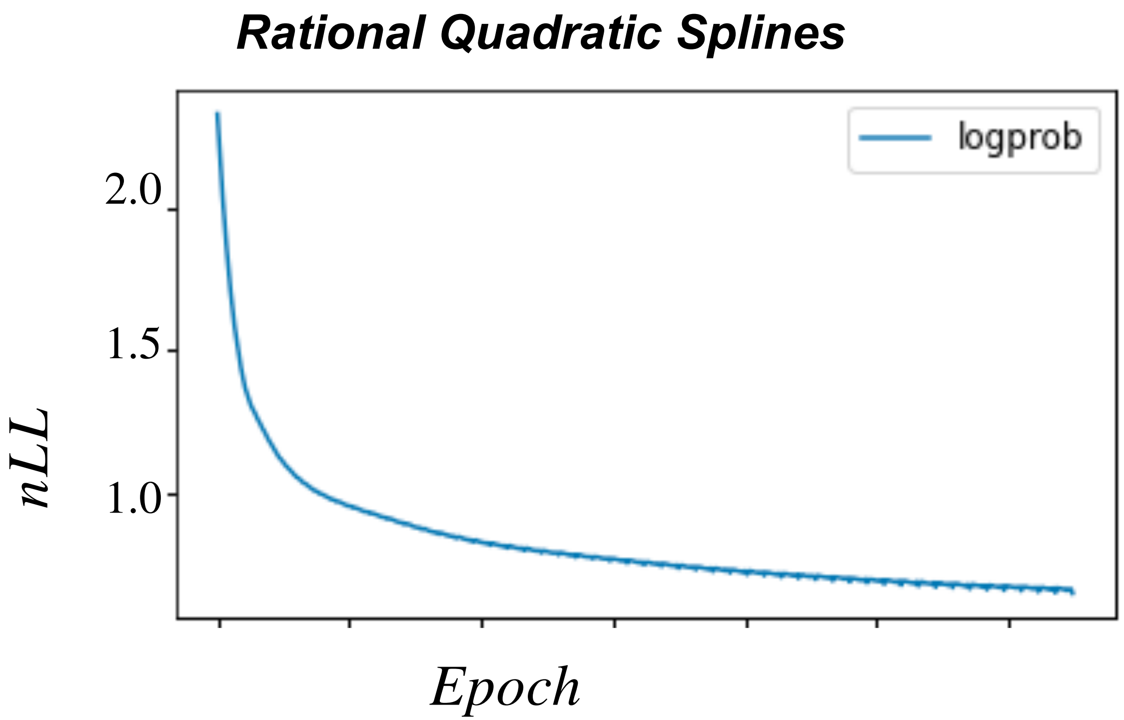
$$\Rightarrow \frac{df}{dx} = \begin{bmatrix} \mathbb{I} & 0 \\ \frac{dz_B}{dx_A} & \frac{ds_{\theta(x^A)}}{dx^B} \end{bmatrix} \Rightarrow \det \frac{df}{dx} = \prod_{i=d+1}^{D} \frac{ds_i^{\theta(x_A)}}{dx_i^B}$$

- $s_\theta(x)$ simple parametrised function but θ arbitrarily complex → **NN for parameters θ**
- Affine: $s_{\theta=(\theta_1,\theta_2)}(x) = x_B \odot \theta_1(x_A) + \theta_2(x_A)$



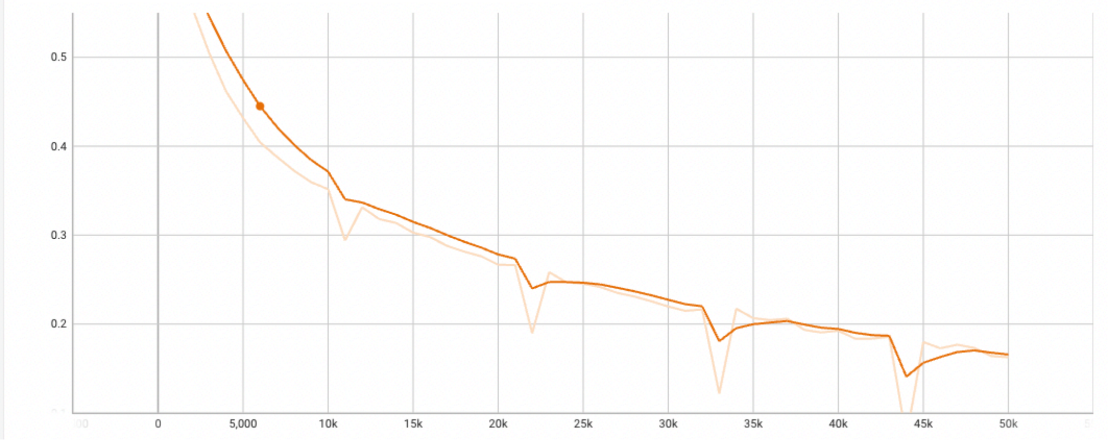
Affine vs Spline



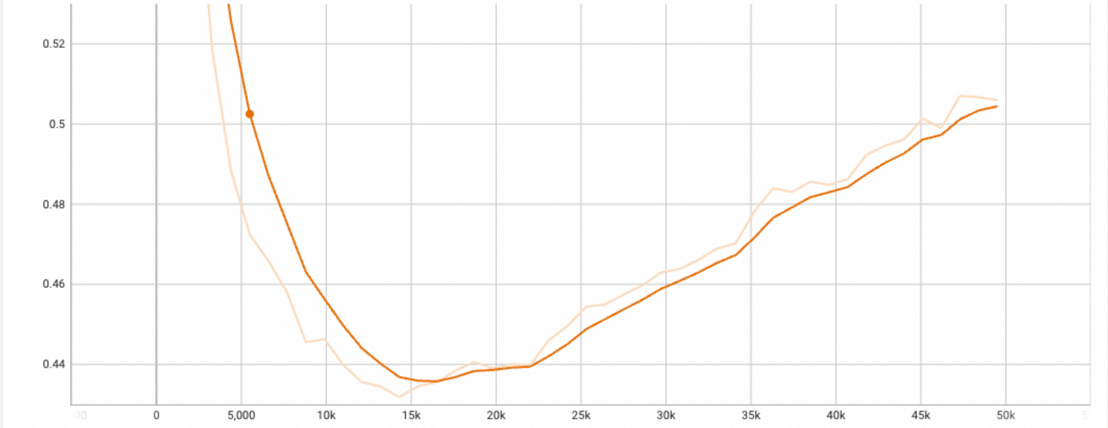


Training vs Validation Logprobs, and Metrics

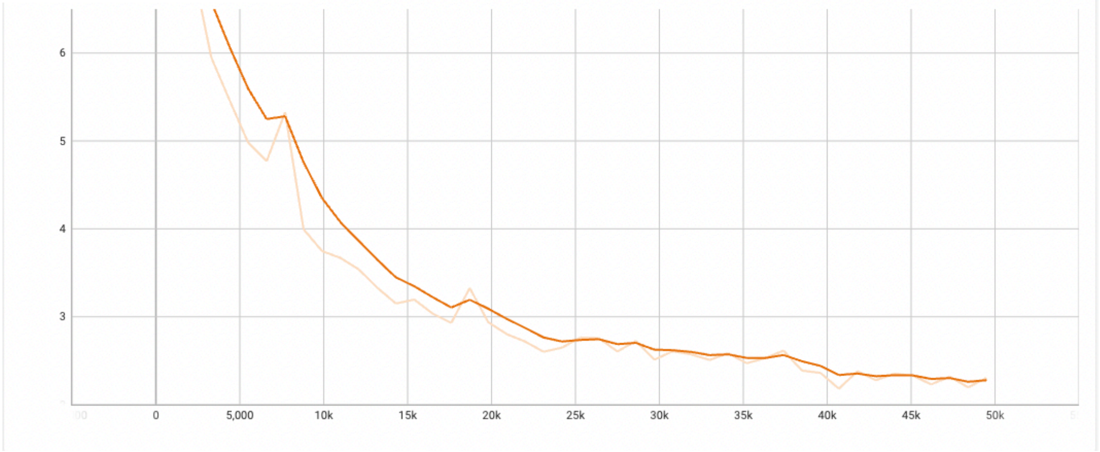
Training Logprob



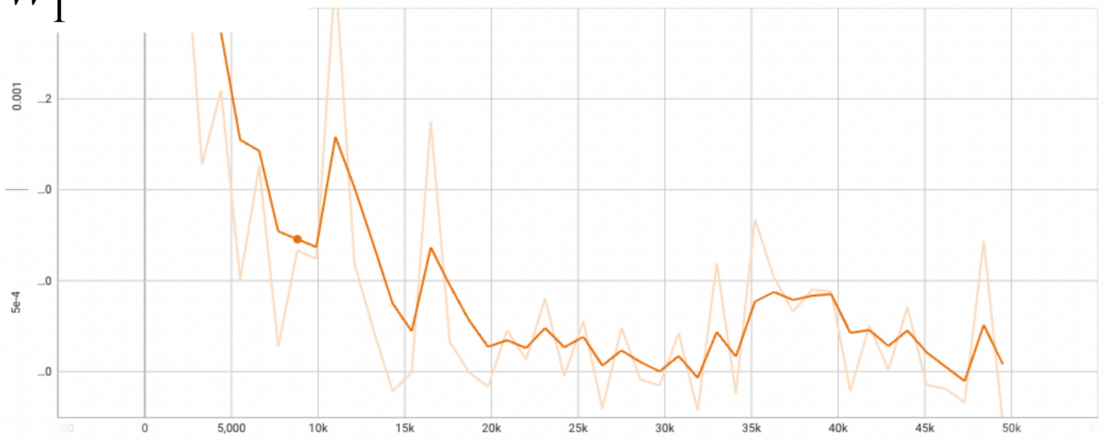
Validation Logprob



FPND



W_1^M

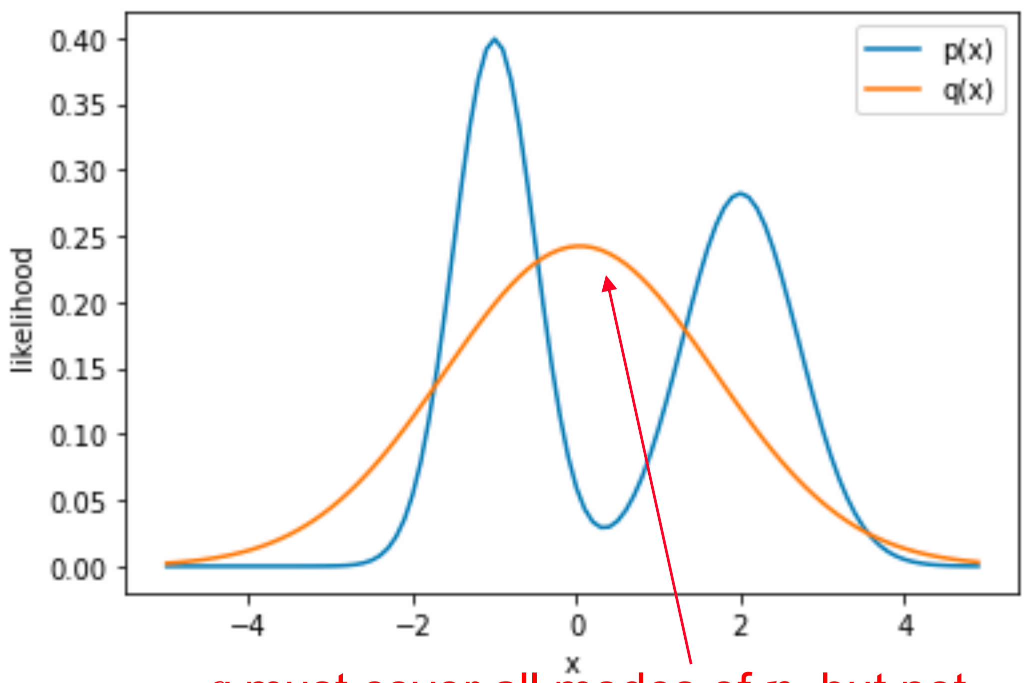


<u>Log Probability seems to not capture quality of generated data</u>

Possible Explanation why Max-Likelihood is not Enough

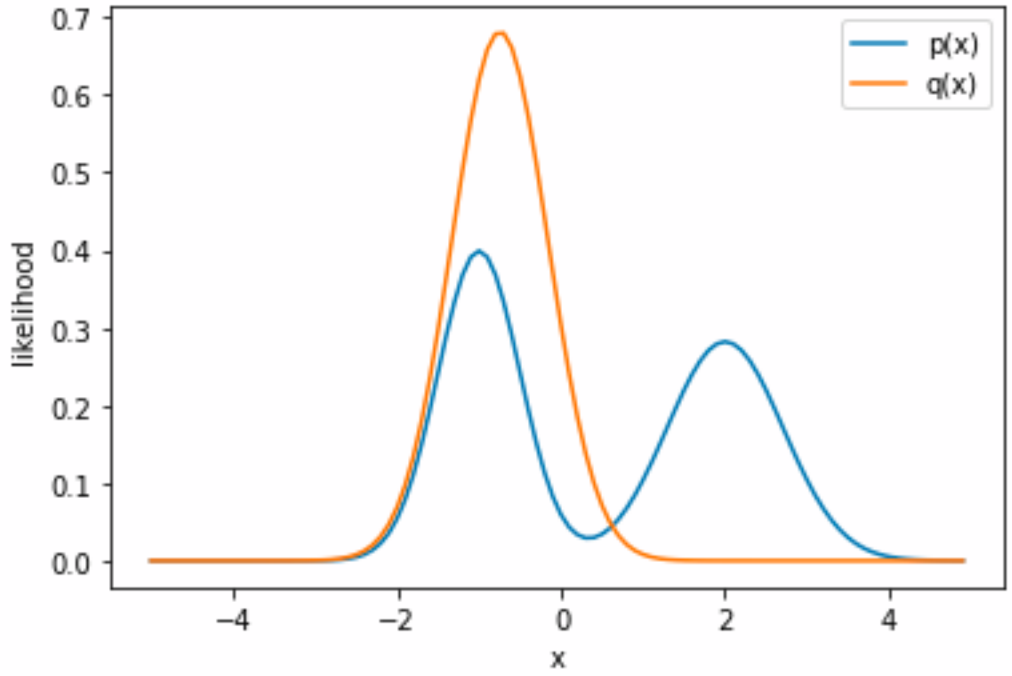
- Maximum likelihood consistent → can learn any distribution given **infinite** data & perfect model class
- Under model misspecification and finite data→ produces models that overgeneralise
- Minimising Forward KL-Divergence: equivalent to Maximum Likelihood

Forward $D_{KL}(p||q)$



q must cover all modes of p, but not penalised for having high q where p is low

Backward $D_{KL}(q||p)$



No punishment for mode collapse

Normalising Flows

In more formal language

- Main foundation: Change of Variables formula, $z = \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x})$

$$p_X(\boldsymbol{x}) = p_Z(\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}))\left|\det\frac{d\boldsymbol{f}_{\boldsymbol{\theta}}}{d\boldsymbol{x}}\right| = p_{\mathbf{Z}}(z)\left|\det\frac{d{\boldsymbol{f}_{\boldsymbol{\theta}}}^{-1}}{dz}\right|^{-1}$$

  - 2 Constraints: Invertible functions, Jacobi-Matrix tractable

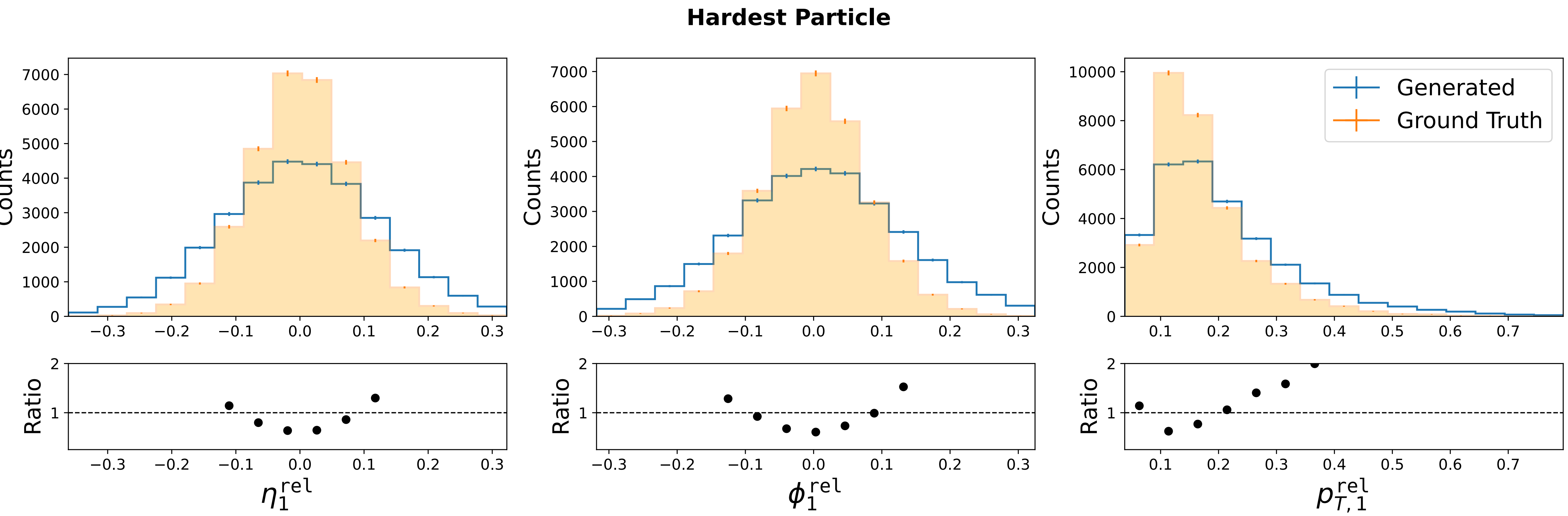
- Stack transformations: $z = z_K = \boldsymbol{f}^{(K)} \circ \cdots \circ \boldsymbol{f}^{(0)}(z_0 = \boldsymbol{x})$

$$\rightarrow \text{Invertible with determinant } \prod_{i=0}^{K} \det\left|\frac{d{\boldsymbol{f}_{\boldsymbol{\theta}}}^{(i)}}{d\mathbf{x}_i}\right|$$

- **Optimise with negative Log-Likelihood:**

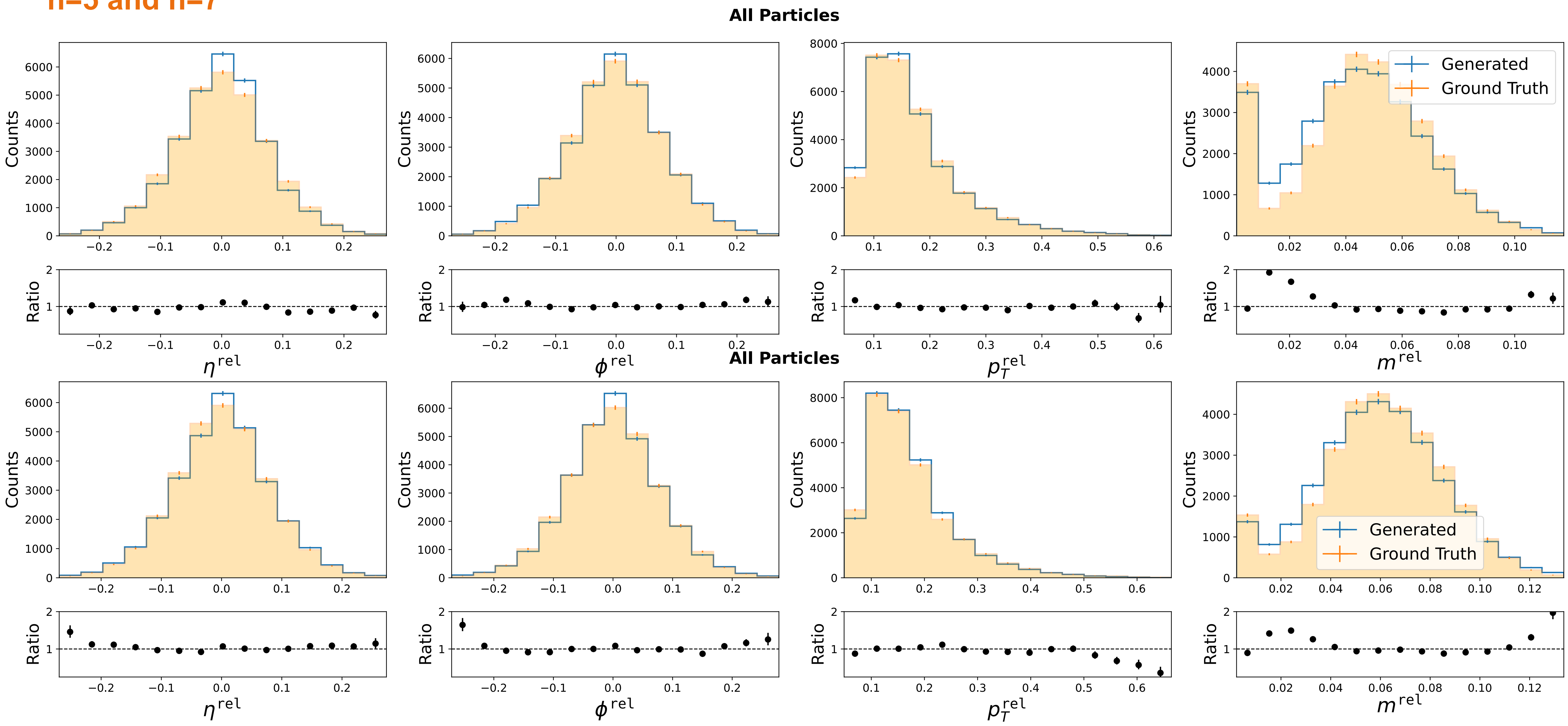
$$\boldsymbol{\theta} = -\arg\min_{\boldsymbol{\theta}} \sum_{x\in X} \log p_{\boldsymbol{x}}(\boldsymbol{x}) = \arg\min_{\boldsymbol{\theta}} \sum_{x\in X}\left(\frac{\boldsymbol{f}(\boldsymbol{x})^2}{\mathbf{2}} - \sum_{i=0}^{K}\left|\det\frac{d\boldsymbol{f}_{\boldsymbol{\theta}}^{(i)}}{d\boldsymbol{x}_i}\right|\right)$$

Affine Marginals



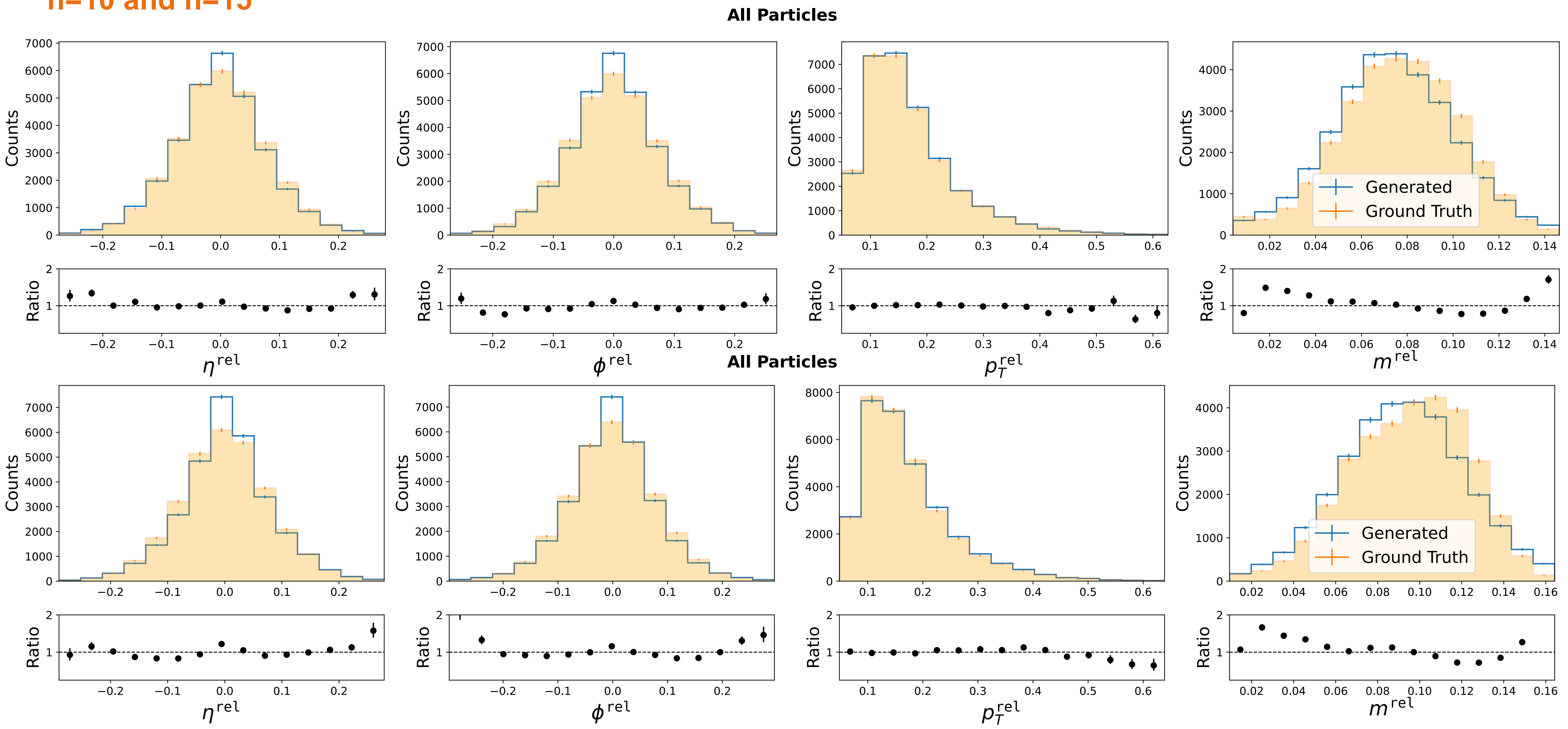
Inclusive Distributions for Fewer Particles

n=5 and n=7



Inclusive Distributions for Fewer Particles

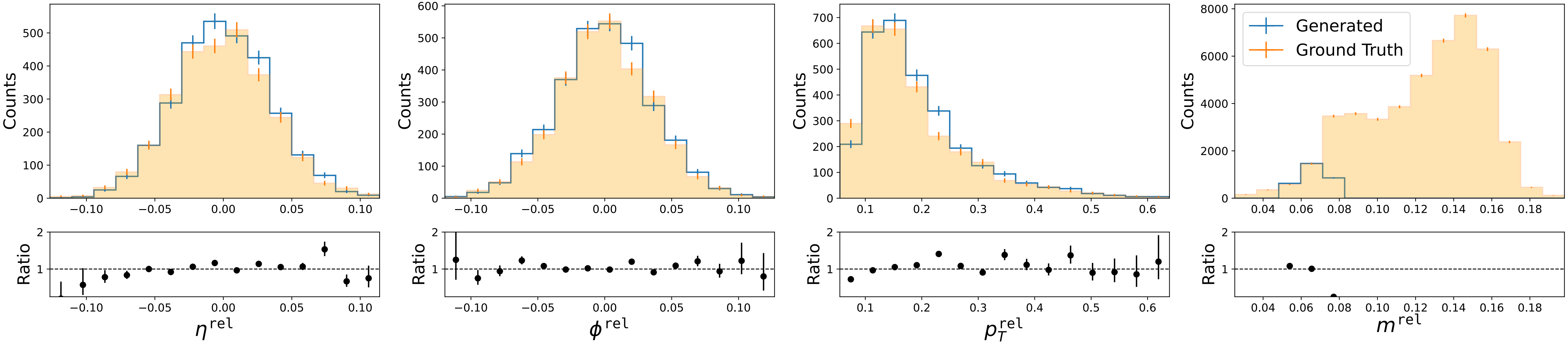
n=10 and n=15



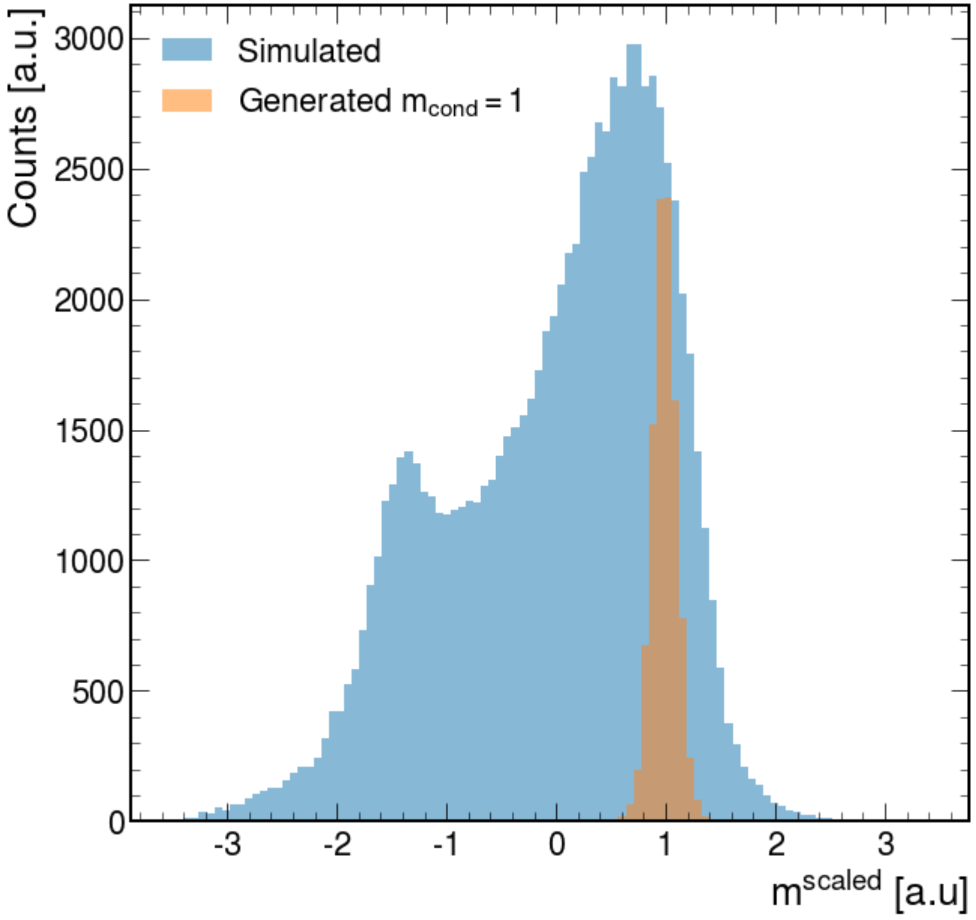
Limited Condition

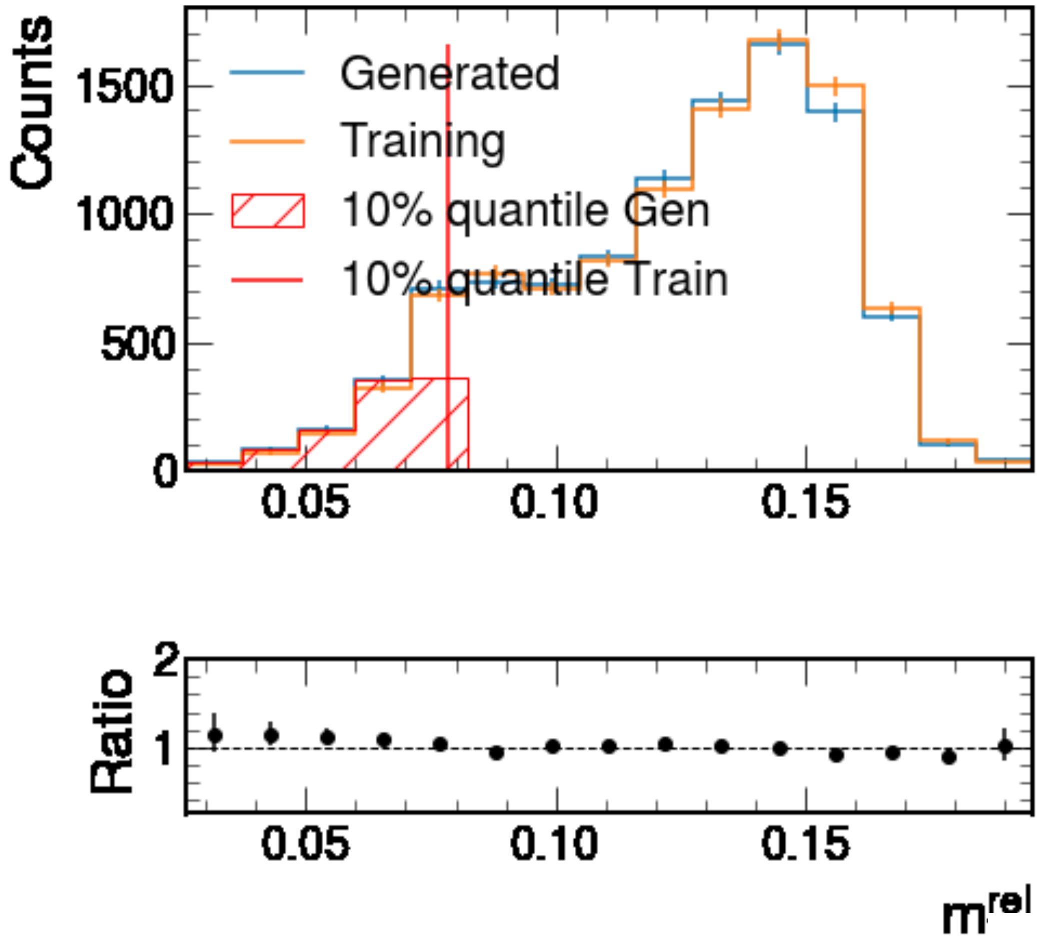
$0.05 < m_{conditon} < 0.075$

Mass Window

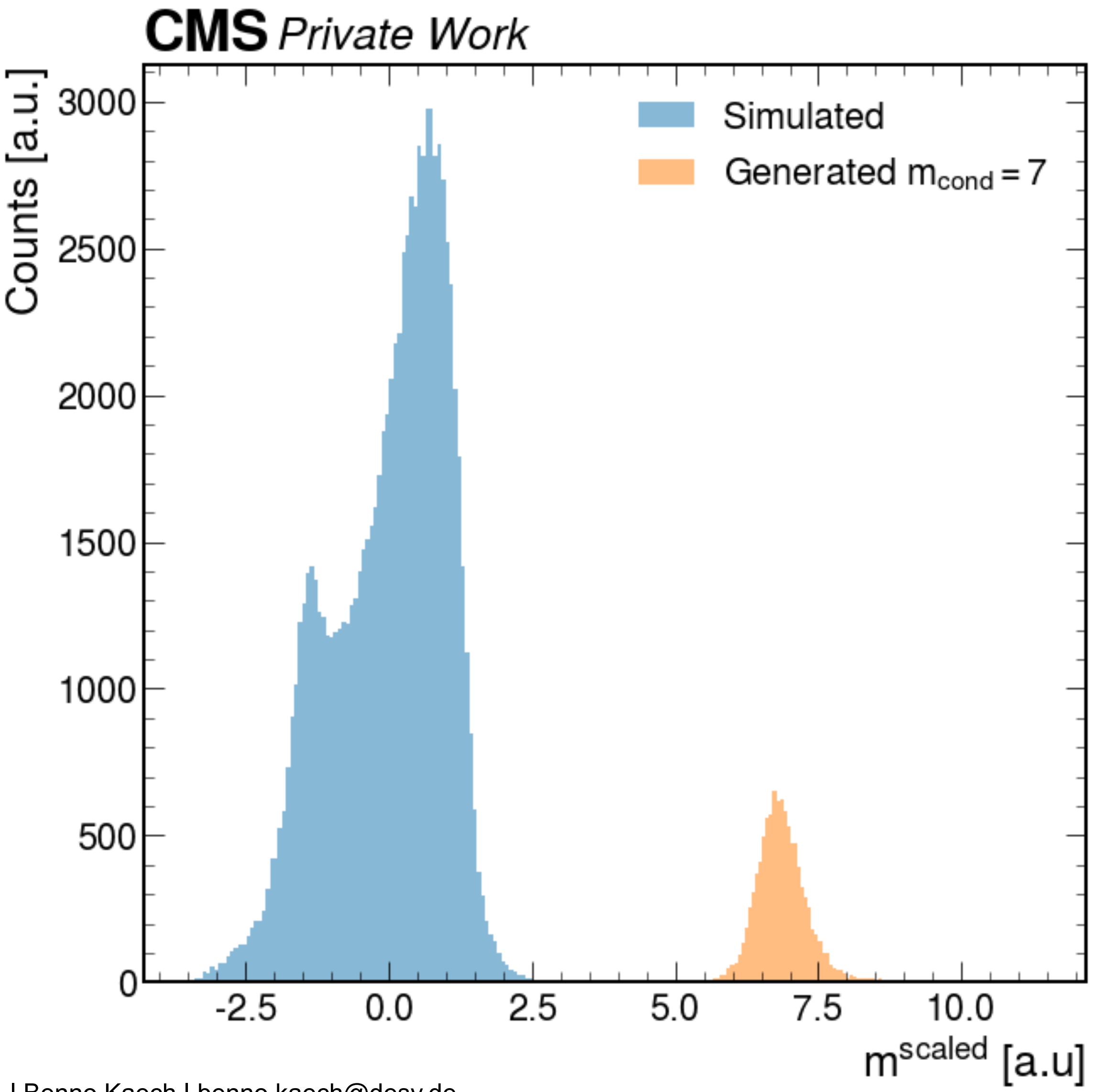


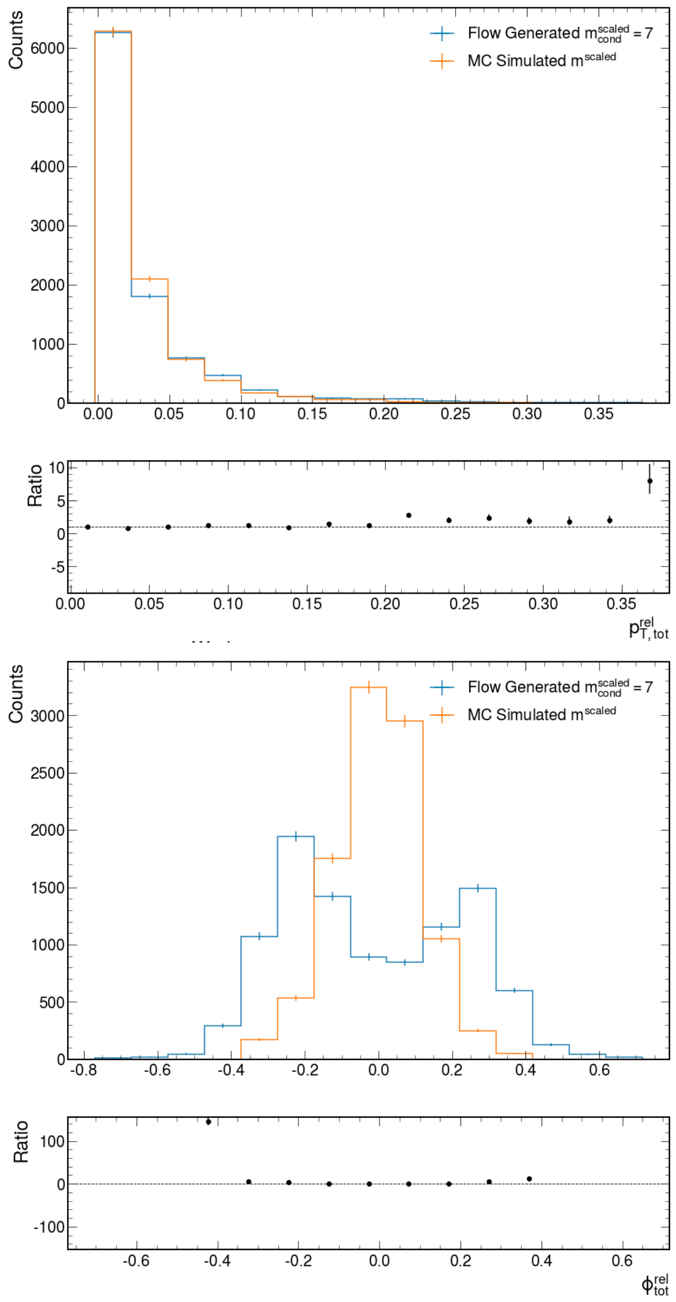
Precision of Mass Conditioning





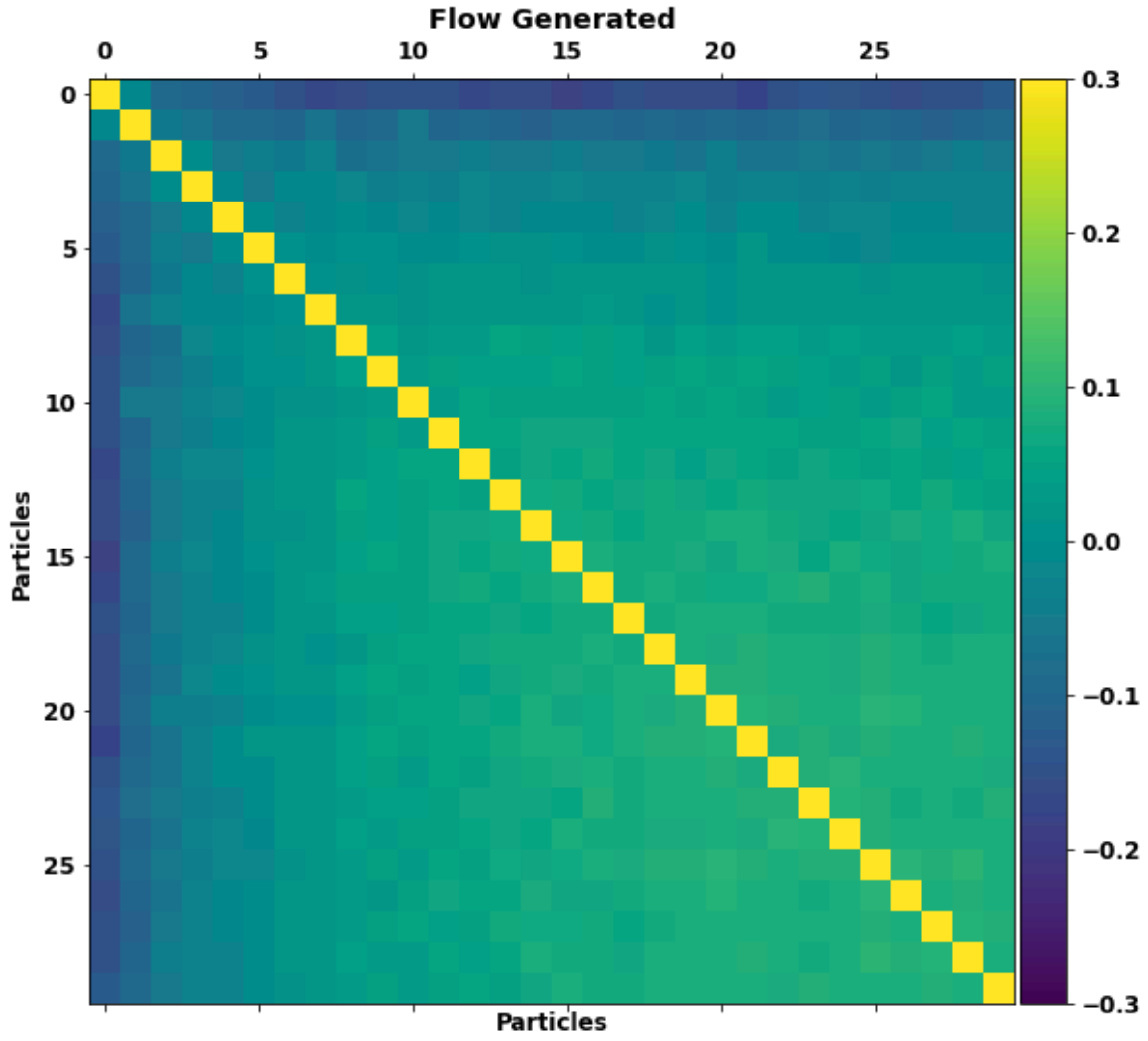
Accuracy of Mass Conditioning

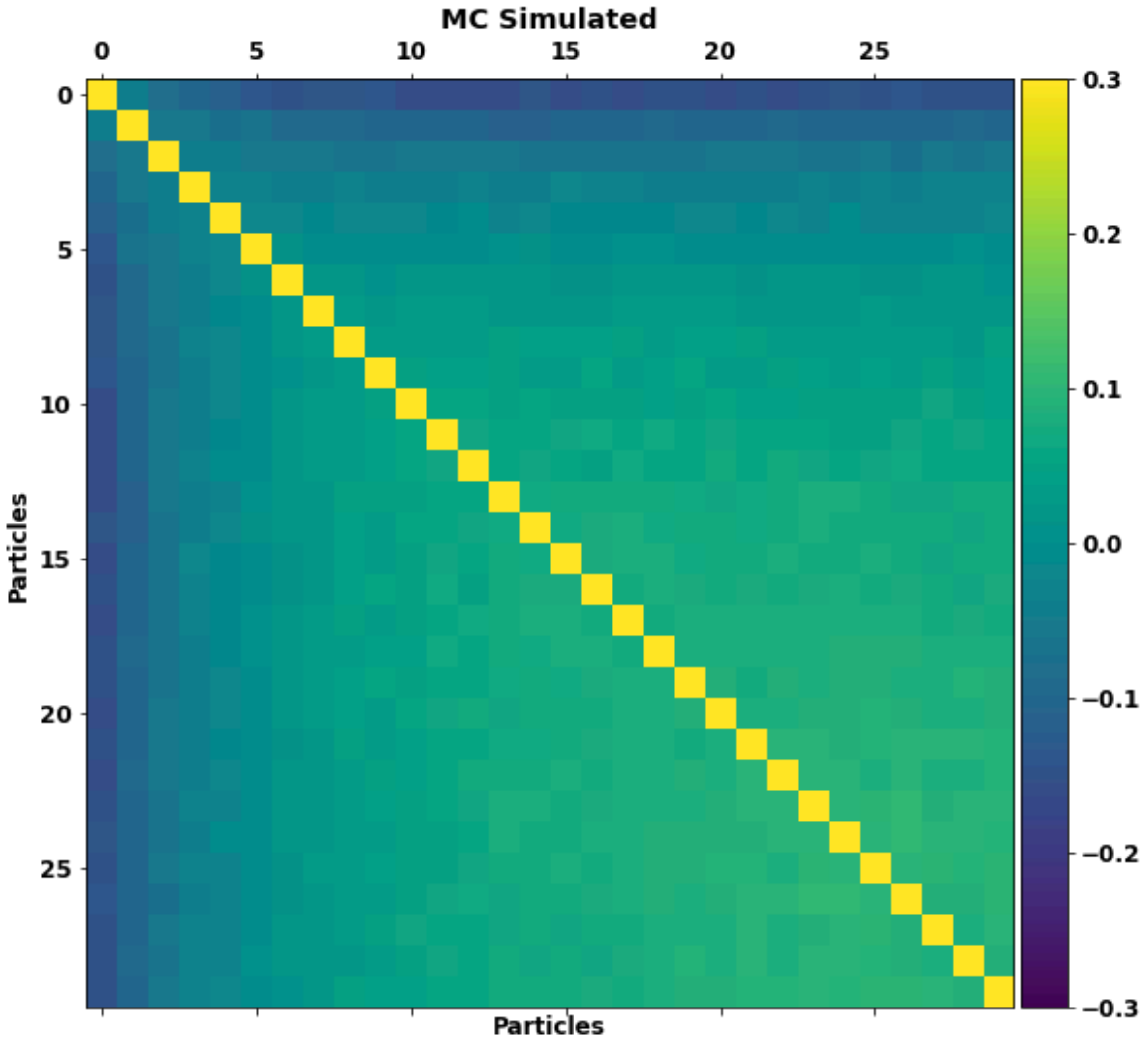




Correlation Plots

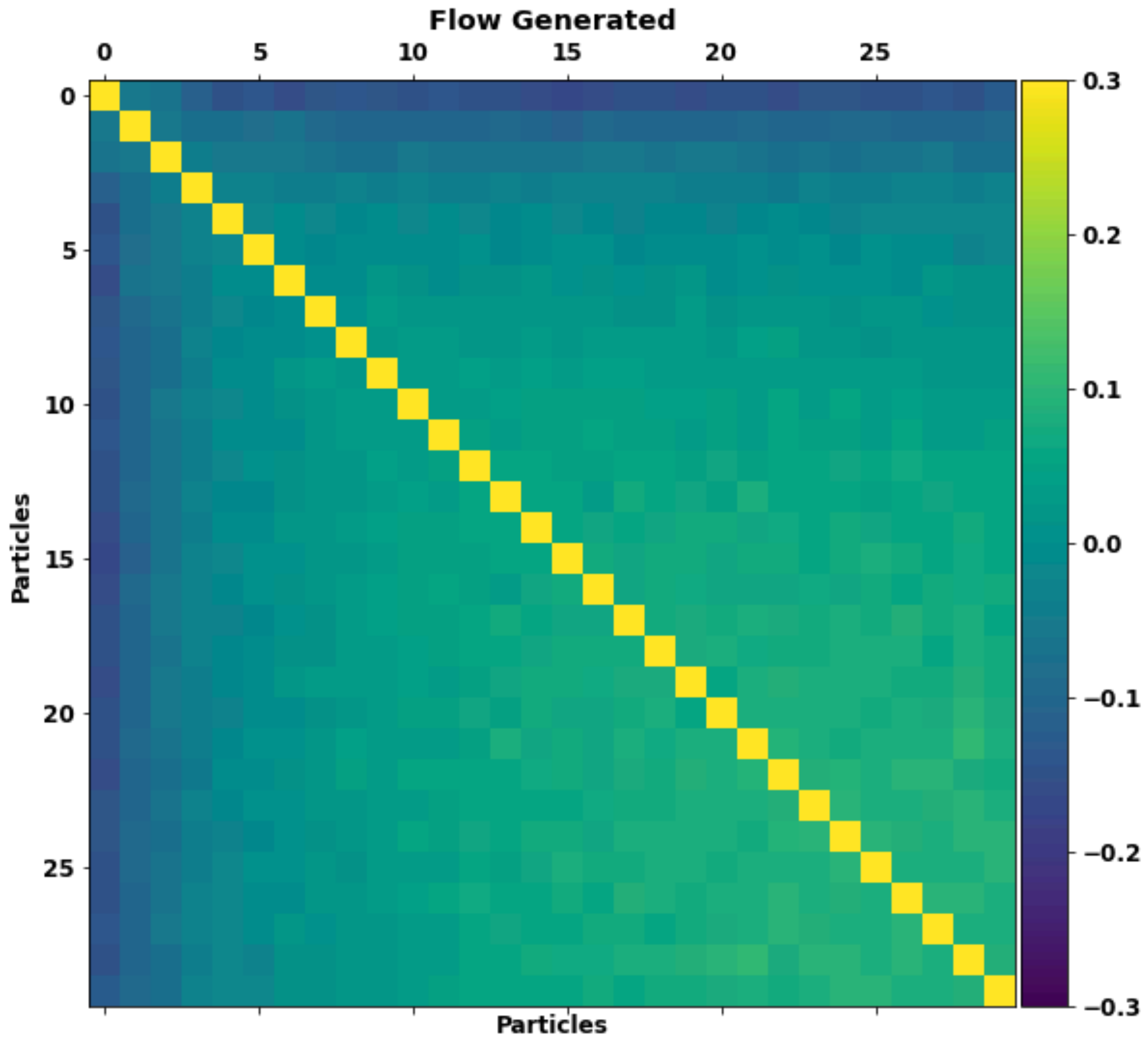
η_{rel}

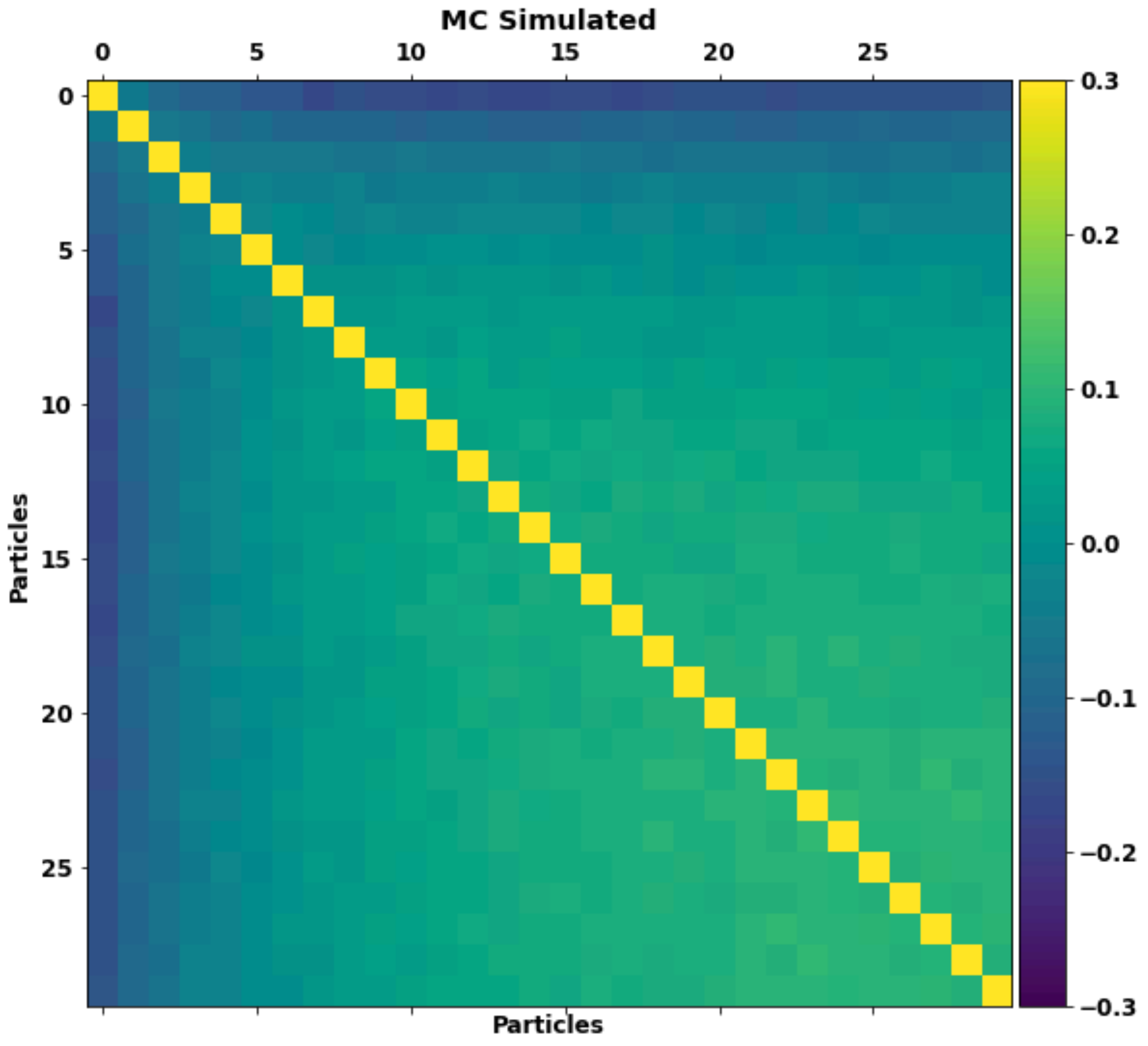




Correlation Plots

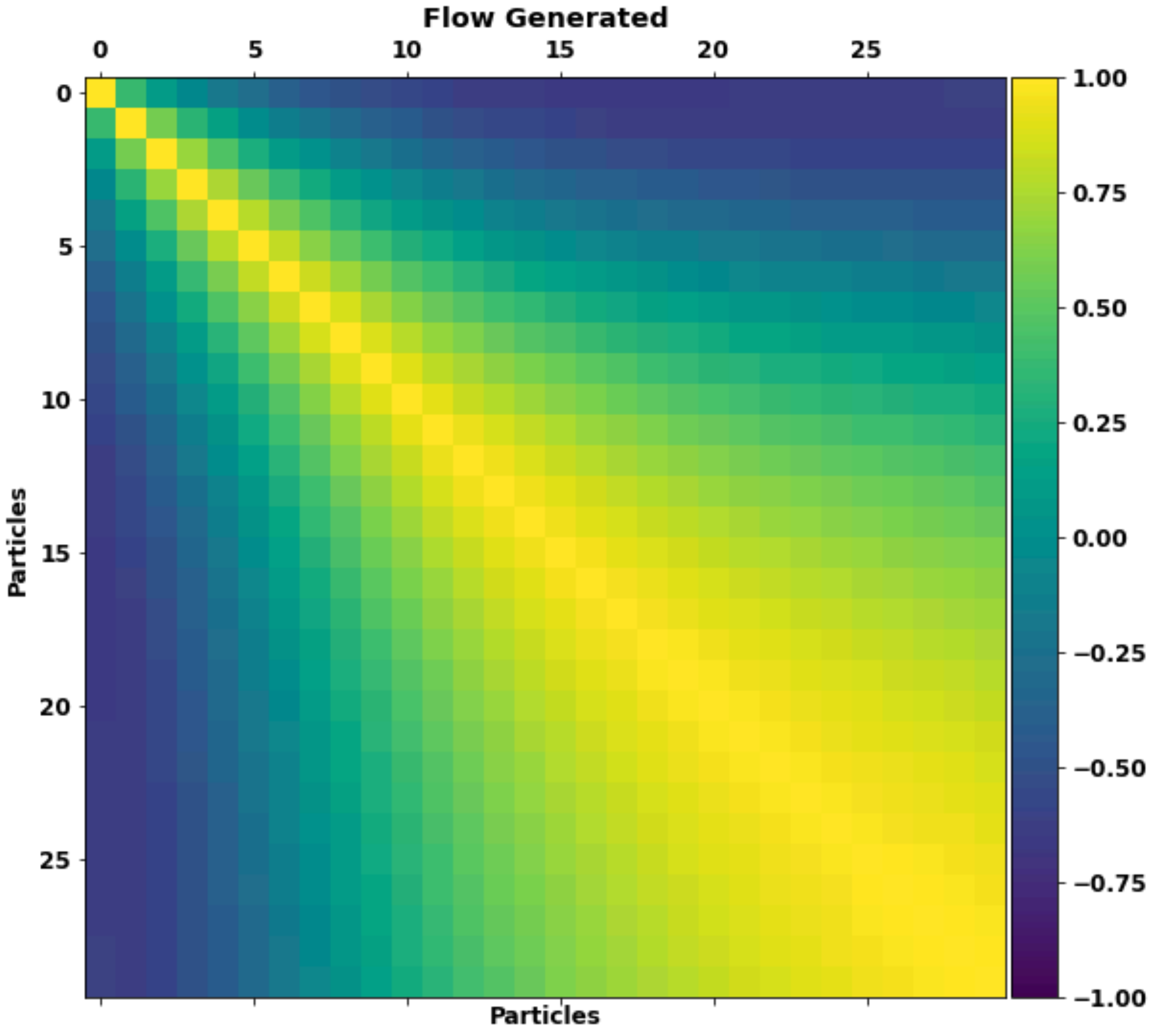
ϕ_{rel}

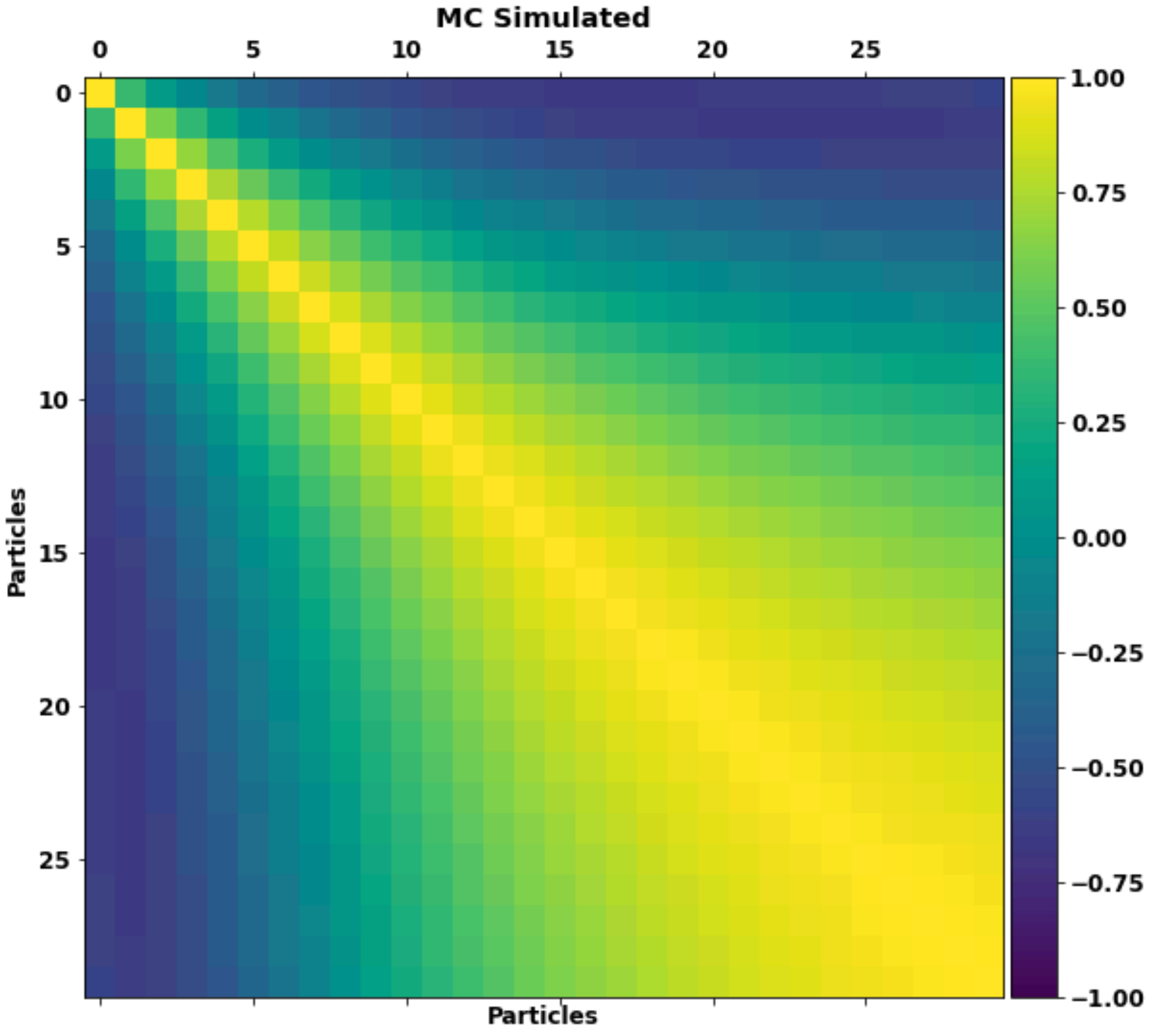




Correlation Plots

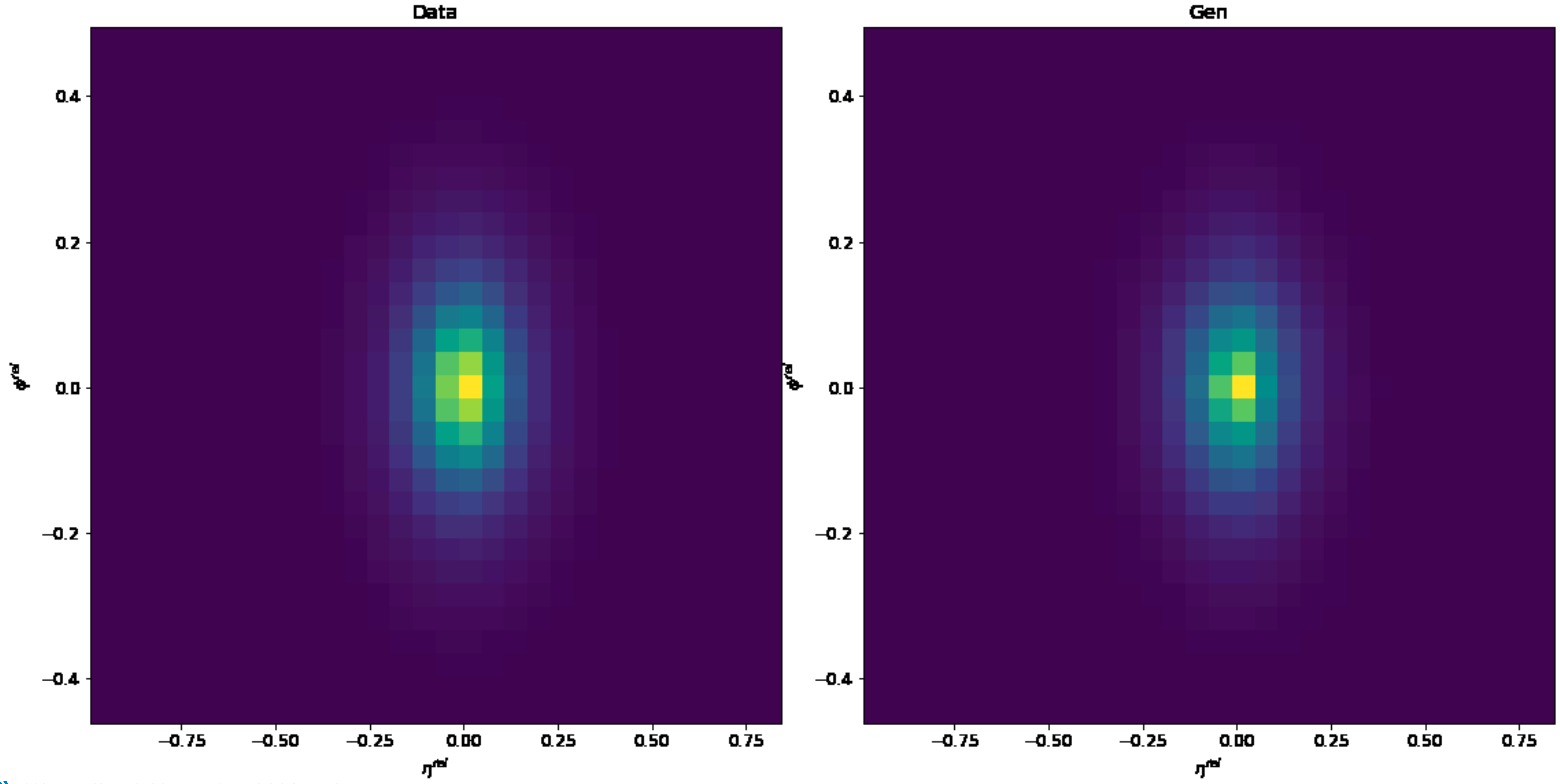
p_T





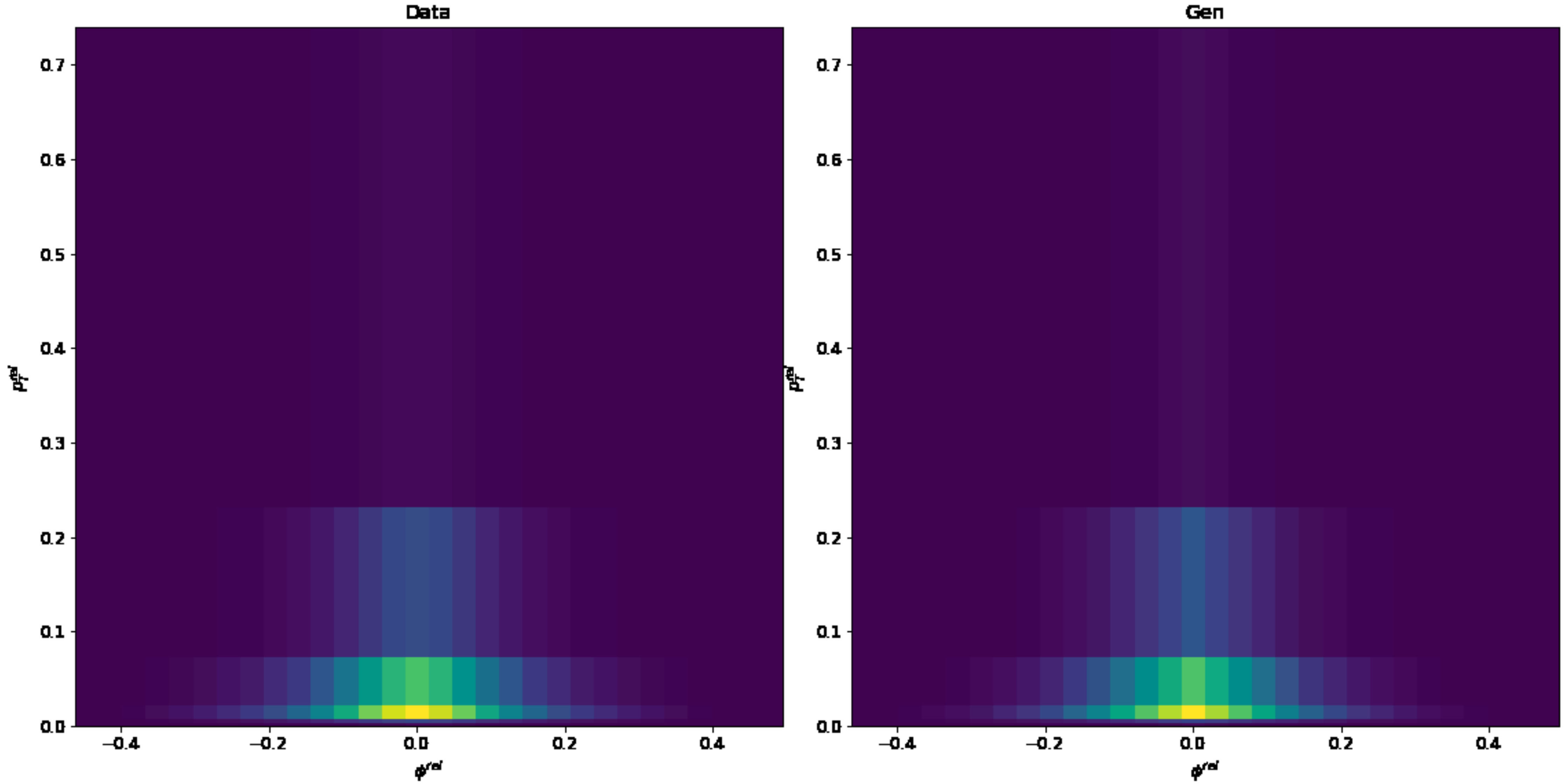
2D Histograms

$\eta^{rel}\phi^{rel}$



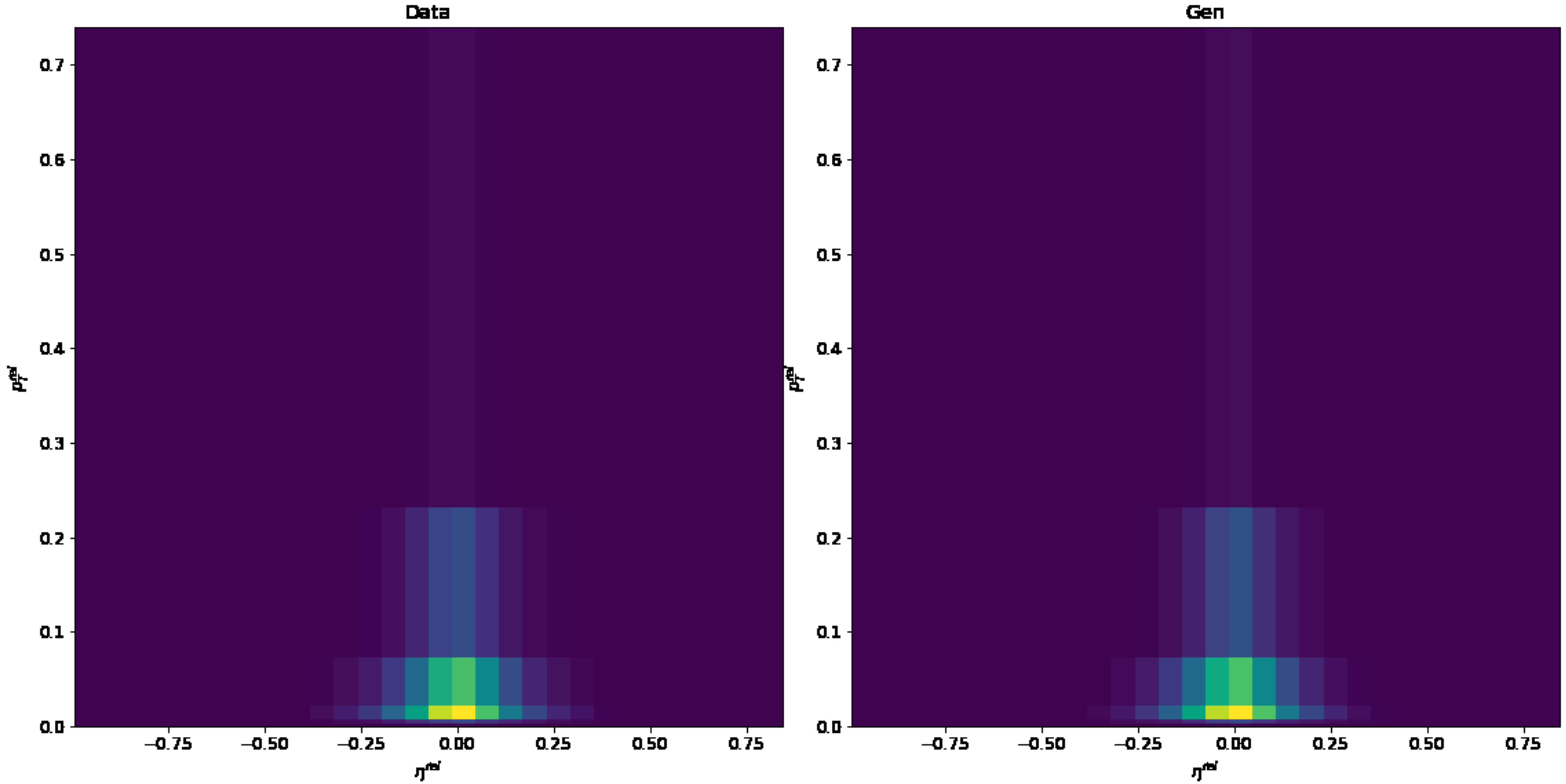
2D Histograms

$\phi^{rel}\, p_T^{rel}$



2D Histograms

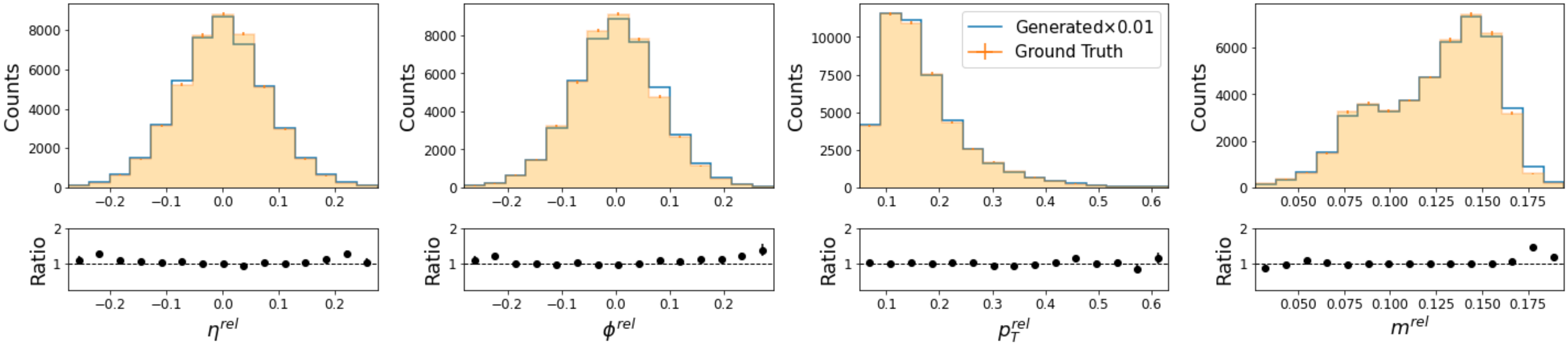
$\eta^{rel}\, p_T^{rel}$



Oversampling

Using 1 Condition

Oversampling by factor 100



Normalising Flows are BIG

```
==========================================================================================
Layer (type:depth-idx)                              Param #
==========================================================================================
TransGan                                            --
├─StandardNormal: 1-1                               --
├─CompositeTransform: 1-2                           --
│    └─ModuleList: 2-1                              --
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-1      437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-2      437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-3      437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-4      437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-5      437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-6      437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-7      437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-8      437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-9      437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-10     437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-11     437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-12     437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-13     437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-14     437,366
│    │    └─PiecewiseRationalQuadraticCouplingTransform: 3-15     437,366
├─Flow: 1-3                                         6,560,490
│    └─CompositeTransform: 2-2                      (recursive)
│    │    └─ModuleList: 3-16                        (recursive)
│    └─StandardNormal: 2-3                          —
│    └─Identity: 2-4                                --
```

```
├─Flow: 1-3                                         6,560,490
│    └─CompositeTransform: 2-2                      (recursive)
│    │    └─ModuleList: 3-16                        (recursive)
│    │    │    └─PiecewiseRationalQuadraticCouplingTransform: 4-16     (recursive)
│    │    │    │    └─ResidualNet: 5-46             (recursive)
│    │    │    │    │    └─Linear: 6-31             (recursive)
│    │    │    │    │    └─ModuleList: 6-32         (recursive)
│    │    │    │    │    │    └─ResidualBlock: 7-91      (recursive)
│    │    │    │    │    │    └─ResidualBlock: 7-92      (recursive)
│    │    │    │    │    └─Linear: 6-33             (recursive)

│    │    │    │    └─Dropout: 5-99                 --
│    │    │    │    └─Dropout: 5-100                --
```