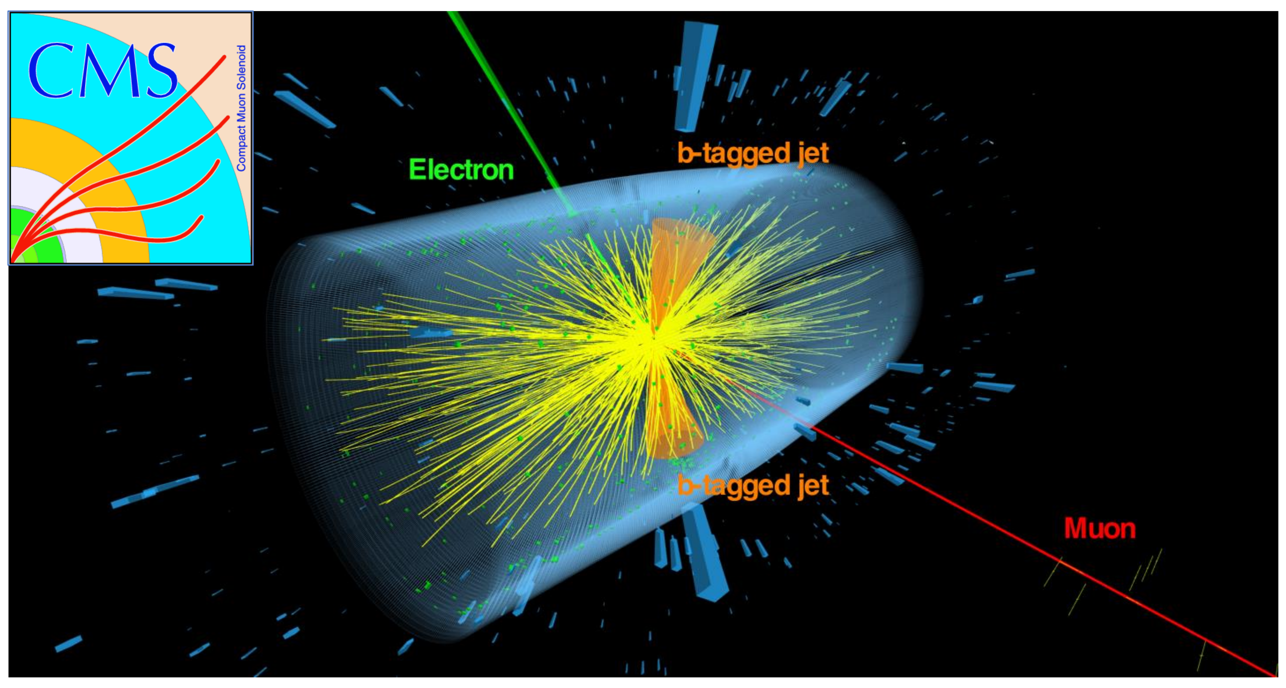
Mass Conditioned & Constrained Normalizing Flows for Particle Cloud Generation.

Benno Käch[1], Dirk Krücker[1], Isabell Melzer-Pellmann[1], Moritz Scham[1], Simon Schnake[1]

LHC & MC Simulations

- The Large Hadron collider (LHC) at CERN/Geneva probes fundamental particle physics
- Monte Carlo Simulations needed to test theories
- Uncertainty due to finite MC statistics always relevant for physics analyses
- MC slow $\sim 100\ s$ per simulated event → previous run ~ 1 billion events were simulated
- Using already $\sim 50\ \%$ of computing resources
- Future Runs higher Luminosity → Even more simulations
- Other ways for simulations needed



Jets

- Jet = narrow cone of particles
- Jets abundant in hadron colliders → simulation crucial
- Different mother particles → different kinematical properties of constituents
- In this study: gluon, top- and light quark initiated jets
- Jets described by reconstructed momenta of constituents (p_x, p_y, p_z)
- Alternatively more collider friendly variables (η, ϕ, p_T):

$$p_x = p_T \cos(\phi), \quad p_y = p_T \sin(\phi), \quad p_z = p_T \sinh(\eta), \quad \eta = -\log\frac{\theta}{2}$$

- Invariant jet mass, assuming mass of constituents is zero:

$$m^2 = \left(\sum_{part} \sqrt{p_x^2 + p_y^2 + p_z^2}\right)^2 - \left(\sum_{part} p_x\right)^2 - \left(\sum_{part} p_y\right)^2 - \left(\sum_{part} p_z\right)^2$$

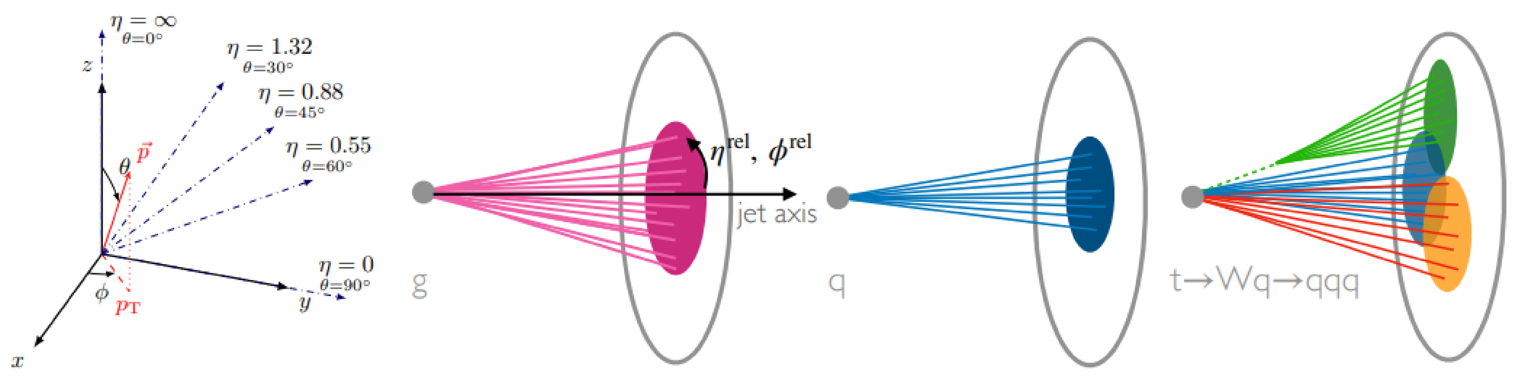


Image from arXiv:2106.11535

Goal

Learn $p_X(x)$ to generate synthetic data

Input/Output Data

- Jets of up to 30 particles (Kansal et al., arXiv/2106.11535)
- Variables relative to jet axis → $(\vec{\eta}^{rel}, \vec{\phi}^{rel}, \vec{p}_T^{rel}) = x \in X$
- Each particle tuple of $(\eta^{rel}, \phi^{rel}, p_T^{rel})$ → up to 90 features
- Events with fewer particles zero-padded
- Dataset from zenodo.org/record/6302454

ML Challenges

- Particle clouds of variable sizes
- Global features (e.g. jet mass, total transverse momentum,..) crucial for analyses
- No straight-forward performance measure for generated data

Normalizing Flows

- Change of variables for probability distributions:

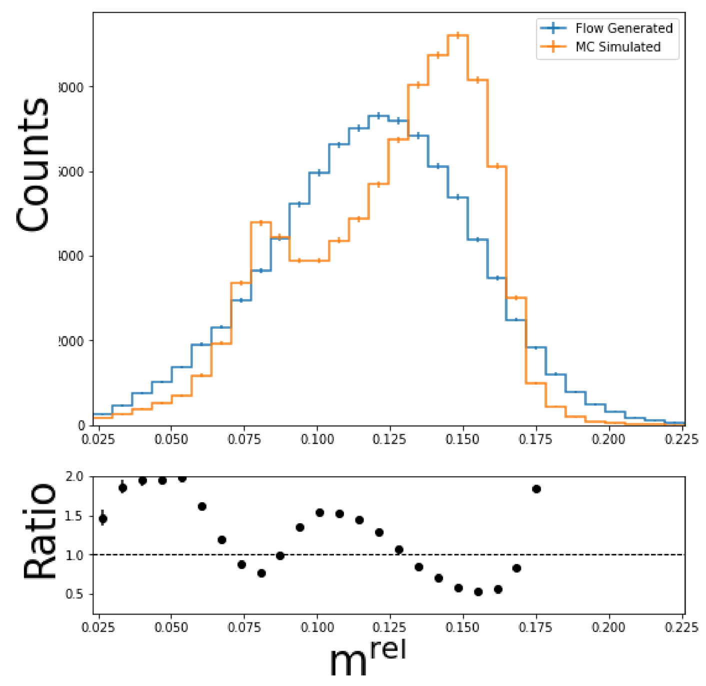
$$p_X(\vec{x}) = p_{Z=f(X)}(\vec{f}_\theta(\vec{x})) \left|\det \frac{\partial \vec{f}_\theta}{\partial \vec{x}}\right|$$

→ Makes Maximum Likelihood training (nLL) viable:

$$L_{nll}(\theta) = -\sum_X \log(\underbrace{p_X(\vec{x})}_{unknown}) \underbrace{=}_{\vec{z}=\vec{f}_\theta(\vec{x})} -\sum_X \log(p_Z(\vec{z})) + \log\left|\det \frac{\partial \vec{f}_\theta}{\partial \vec{x}}\right|$$

- p_Z closed form (e.g. Gaussian), f invertible
- Coupling layers candidates for invertible transformations
- Chaining together multiple coupling layers → Enhanced expressivity
- Generating: Sample known p_Z and apply inverse transformations
- Add noise $\sim \mathcal{O}(10^{-7})$ to zero-padded particles
- nflows implementation used - github.com/bayesiains/nflows

Vanilla Normalizing Flows



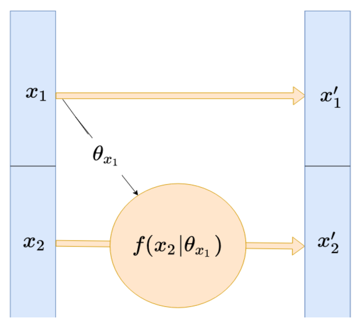
Global feature not modelled correctly

Coupling Layers

- Coupling layer split features in two sets → Parameters for transformation of one set from other
- Simple affine coupling layer:

$$f(\vec{x}) = \begin{cases} \vec{x}_1' = \vec{x}_1 \\ \vec{x}_2' = \vec{a}(\vec{x}_1) \odot \vec{x}_2 + \vec{b}(\vec{x}_1) \end{cases}$$

$$\Rightarrow f^{-1}(\vec{x}') = \begin{cases} \vec{x}_1 = \vec{x}'_1 \\ \vec{x}_2 = \dfrac{\vec{x}_2' - \vec{b}(\vec{x}'_1)}{\vec{a}(\vec{x}'_1)} \end{cases}$$



- Parameters a, b output of deep residual network
- Network parameters optimized by gradient descent on nLL
- Features permutated after each coupling layer
- In this study more complicated coupling layers → Rational Quadratic Splines

Conditioning

- Additional inputs to neural networks in coupling layers → Needed during sampling
- Here: number of particles n, invariant mass m
- Conditioning enhances expressivity of flows and modelling of global features

Mass Constraint

- Enforce correct mass distribution by adding penalty term

$$L_{mass} = (m_{gen} - m_{cond})^2$$

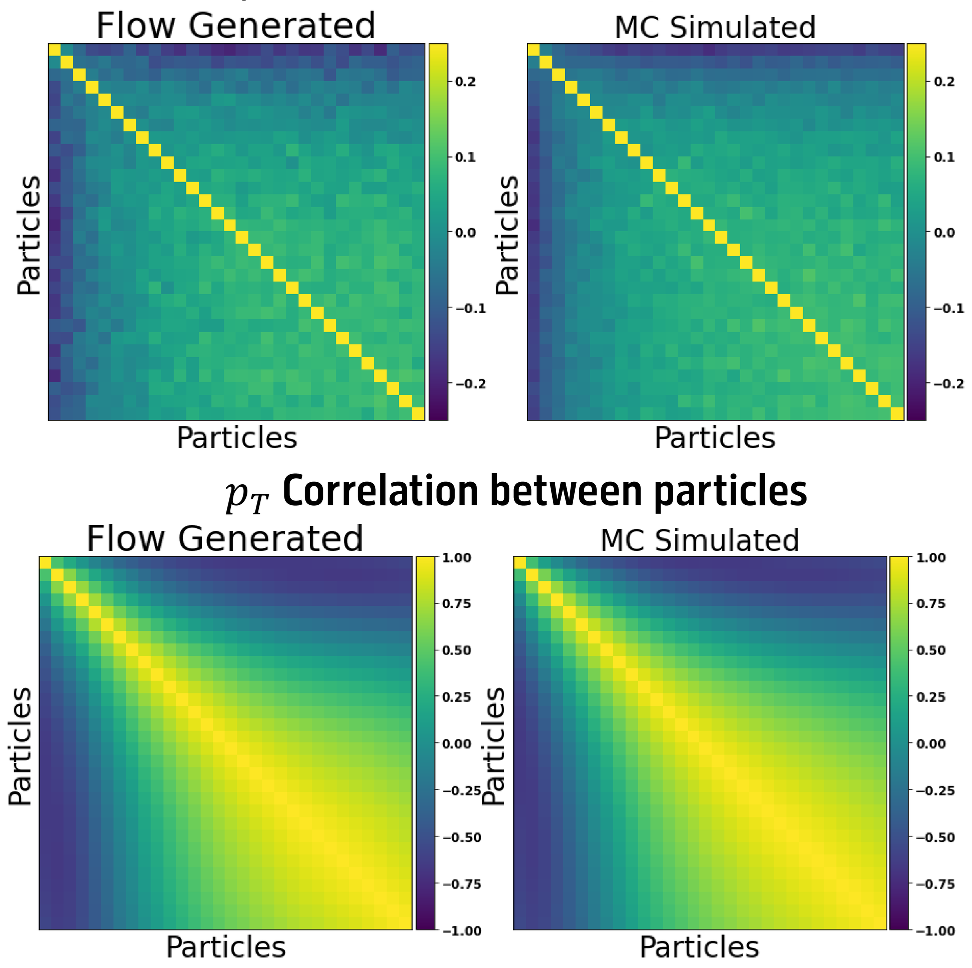
- Loss from latent space and input space:

$$L_{Tot} = \underbrace{L_{nll}}_{latent\ space} + \lambda \cdot \underbrace{L_{mass}}_{input\ space}$$

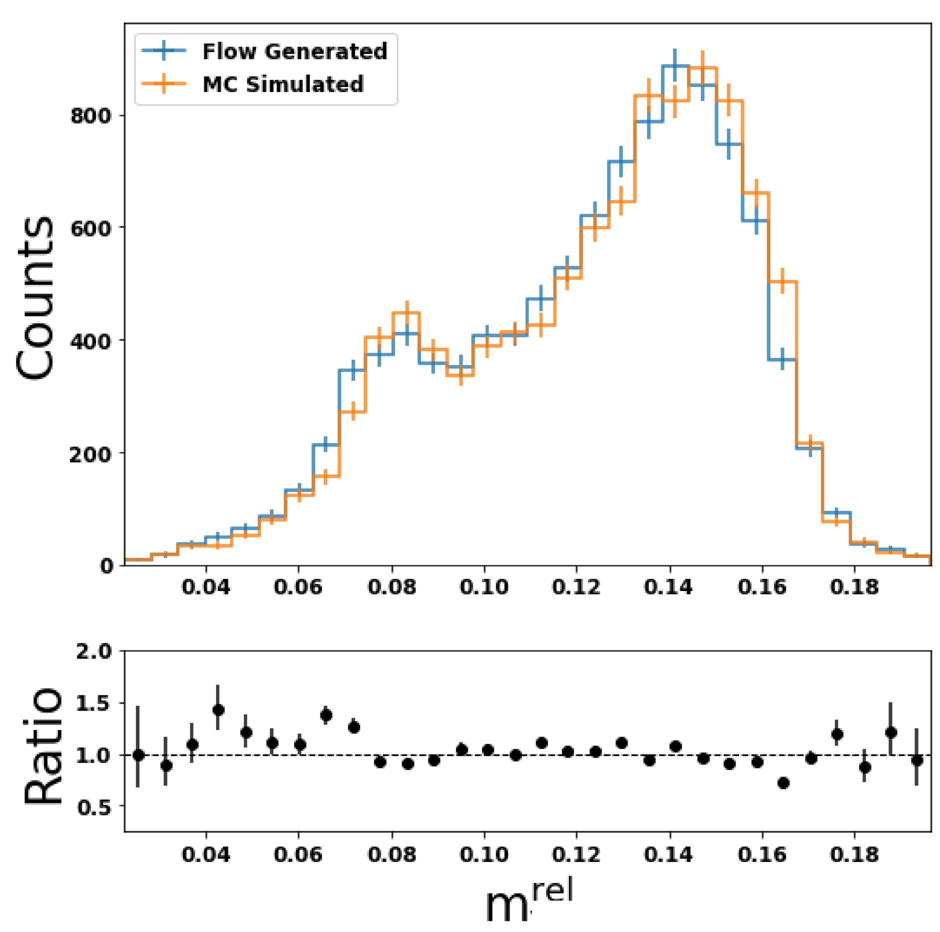
Results

ϕ Correlation between particles

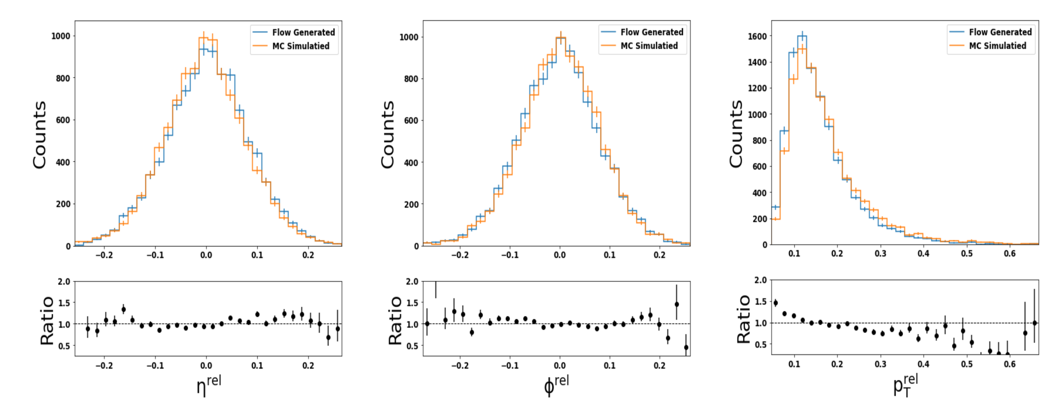
p_T Correlation between particles



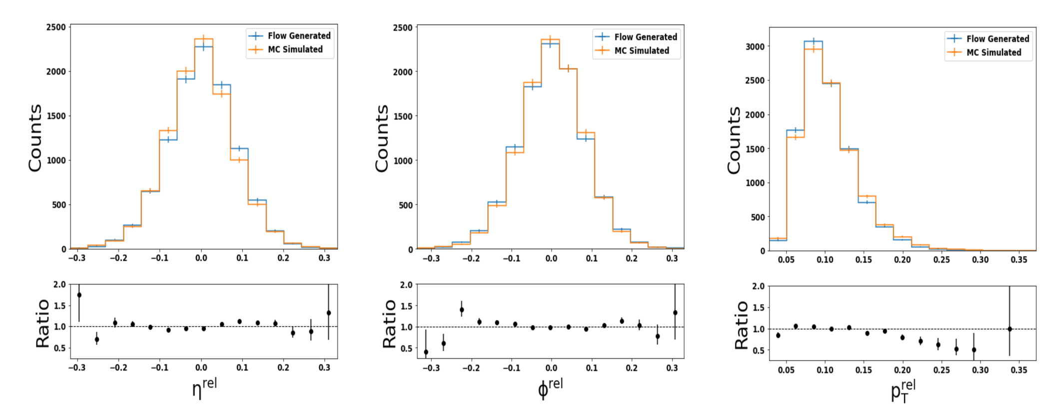
Jet Mass Distribution modelled correctly



Inclusive Distributions (η, ϕ, p_T) over all particles and all jets



(η, ϕ, p_T) Distribution of particle with highest p_T



(1) DESY

HELMHOLTZ

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DESY.