

# Uncertainty Estimation for LHC Event Generation

---

Sophia Vent

IRN Terascale 2021

Institute for Theoretical Physics  
Heidelberg University

# Event Generation for Precision Enthusiasts

SciPost Physics

Submission

## Generative Networks for Precision Enthusiasts

Anja Butter<sup>1</sup>, Theo Heimel<sup>1</sup>, Sander Hummerich<sup>1</sup>, Tobias Krebs<sup>1</sup>,  
Tilman Plehn<sup>1</sup>, Armand Rousselot<sup>2</sup>, and Sophia Vent<sup>1</sup>

<sup>1</sup> Institut für Theoretische Physik, Universität Heidelberg, Germany

<sup>2</sup> Heidelberg Collaboratory for Image Processing, Universität Heidelberg, Germany

November 20, 2021

## Abstract

Generative networks are opening new avenues in fast event generation for the LHC. We show how generative flow networks can reach percent-level precision for kinematic distributions, how they can be trained jointly with a discriminator, and how this discriminator improves the generation. Our joint training relies on a novel coupling of the two networks which does not require a Nash equilibrium. We then estimate the generation uncertainties through a Bayesian network setup and through conditional data augmentation, while the discriminator ensures that there are no systematic inconsistencies compared to the training data.

---

### Content

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Precision generator</b>	<b>2</b>
2.1	Data set	3
2.2	INN generator	4
<b>3</b>	<b>DiscFlow generator</b>	<b>9</b>
3.1	Discriminator reweighting	9
3.2	Joint training	11
<b>4</b>	<b>Uncertainties and control</b>	<b>15</b>
4.1	Bayesian network	15
4.2	Conditional augmentations	19
4.3	Discriminator for consistency	20
<b>5</b>	<b>Outlook</b>	<b>20</b>
<b>References</b>		<b>22</b>

---

arXiv:2110.13632

A.Butter, T.Heimel, S.Hummerich, T.Krebs,  
T.Plehn, A.Rousselot, S.V

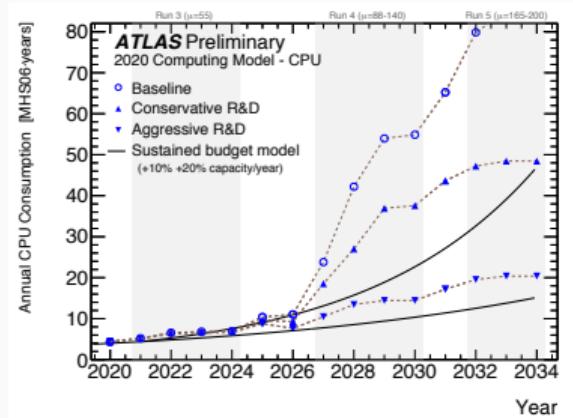
# Motivation

---

# Motivation

- Current and future LHC runs rely on first principal based simulations
- Efficiency of Monte Carlo becomes a limiting factor

→ Use neural networks (GANs, INNs) as generative models



# Motivation

- Current and future LHC runs rely on first principal based simulations
- Efficiency of Monte Carlo becomes a limiting factor

→ Use neural networks (GANs, INNs) as generative models

,

## Two main goals

1. High precision
2. Uncertainty estimation

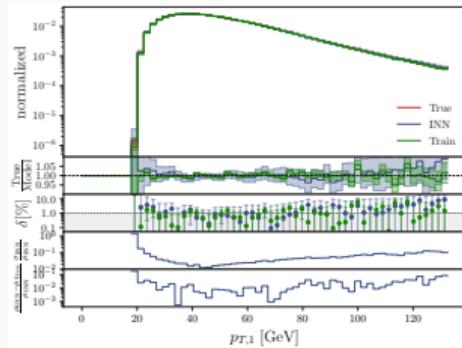
# Invertible Neural Networks

---

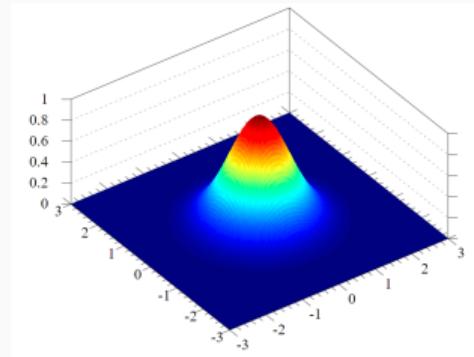
# Invertible Neural Networks

Target: learn phase space density  $p_X(x)$

Invertible mapping between  $x \in X$  and  $z \in Z$  in the latent space through a map  $f$



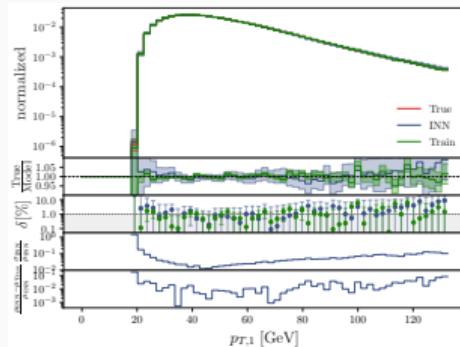
$$p_X(x) \xrightleftharpoons[\text{generating}]{\text{training}} p_Z(z)$$



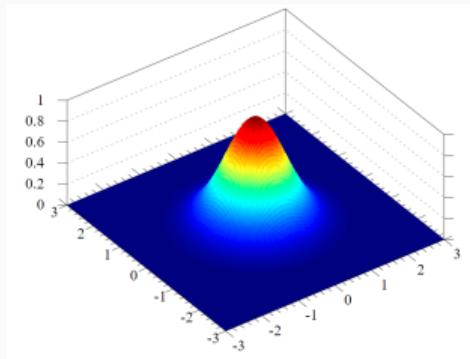
# Invertible Neural Networks

Target: learn phase space density  $p_X(x)$

Invertible mapping between  $x \in X$  and  $z \in Z$  in the latent space through a map  $f$



$$p_X(x) \xrightleftharpoons[f]{f^{-1}} p_Z(z)$$



$$p_X(x) = p_Z(z) \det \frac{\partial f(z)}{\partial z}^{-1} = p_Z(f^{-1}(x)) \det \frac{\partial f^{-1}(x)}{\partial x}$$

## Data set

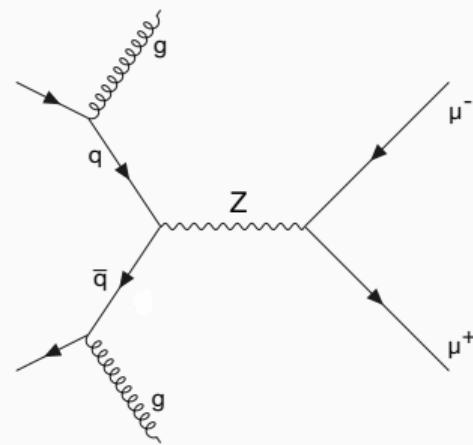
---

## Data Set

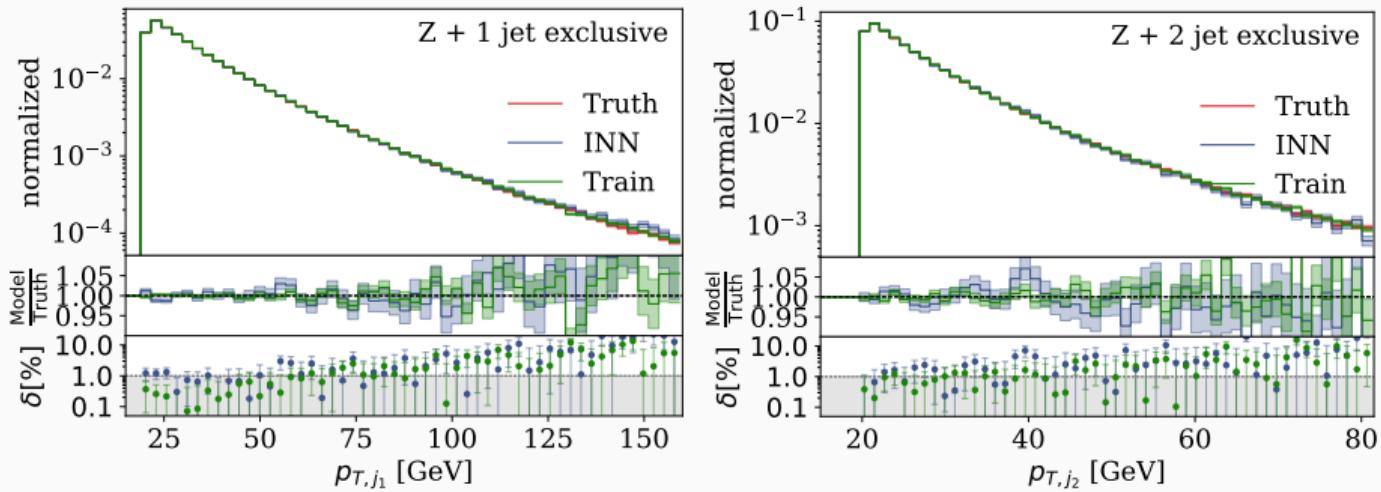
$$pp \rightarrow \mu^+ \mu^- + n \text{ jets}$$

$$n \in \{1, 2, 3\}$$

- Number of 1 jet-events 4 million
- Number of 2 jet-events 1.1 million
- Number of 3 jet-events 0.3 million

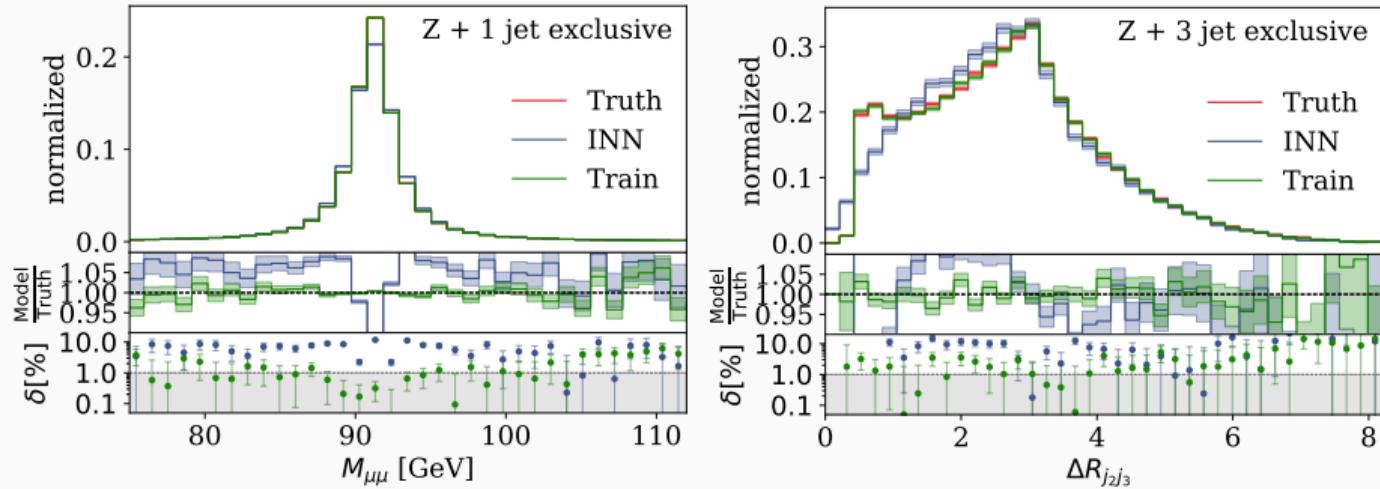


# Results



- High precision at  $\approx 1\%$

## Results INN Baseline



- Peak structure and non-trivial phase space topologies pose a challenge

## Increasing the precision

---

## Discriminator Training

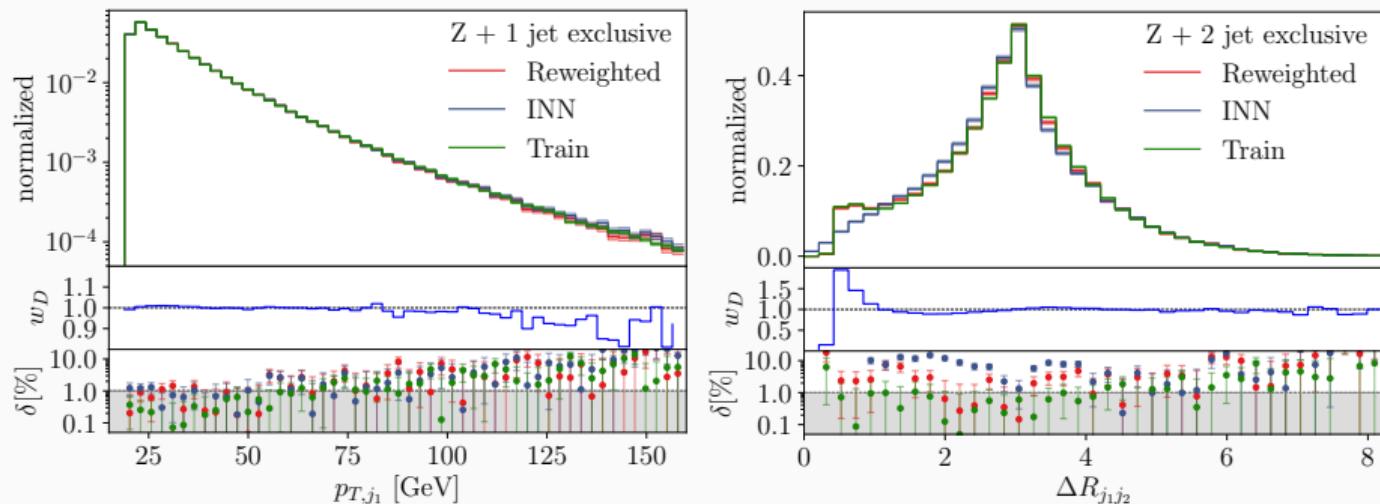
Discriminator:

- Train the Discriminator by minimizing the cross entropy
- Extract probabilities  $D(x_i)$  for an identified event ( $x_i$ )
- Reweight the event weights with the Discriminator output

$$w_D(x_i) = \frac{D(x_i)}{1 - D(x_i)}$$

# Results Discriminator Reweight

## Discriminator Training

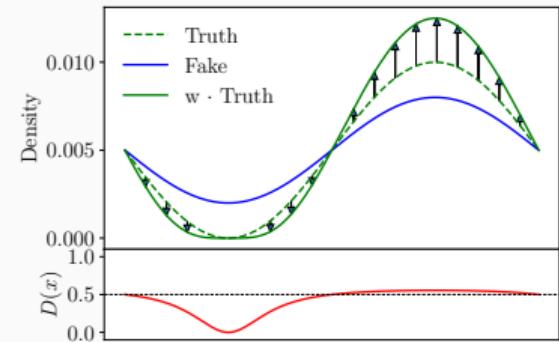


- Close to perfect reweighting but left with weighted events

# Discflow

- Train INN and Discriminator parallel and give access to each other

$$\begin{aligned}\mathcal{L}_{\text{DiscFlow}} &= \sum_{i=1}^B w_D(x_i)^\alpha \left( \frac{\psi(x_i; c_i)^2}{2} - \log J(x_i) \right) \\ &\approx \int dx \underbrace{w_D(x)^\alpha P(x)}_{\text{reweighted truth}} \left( \frac{\psi(x; c)^2}{2} - \log J(x) \right) .\end{aligned}$$

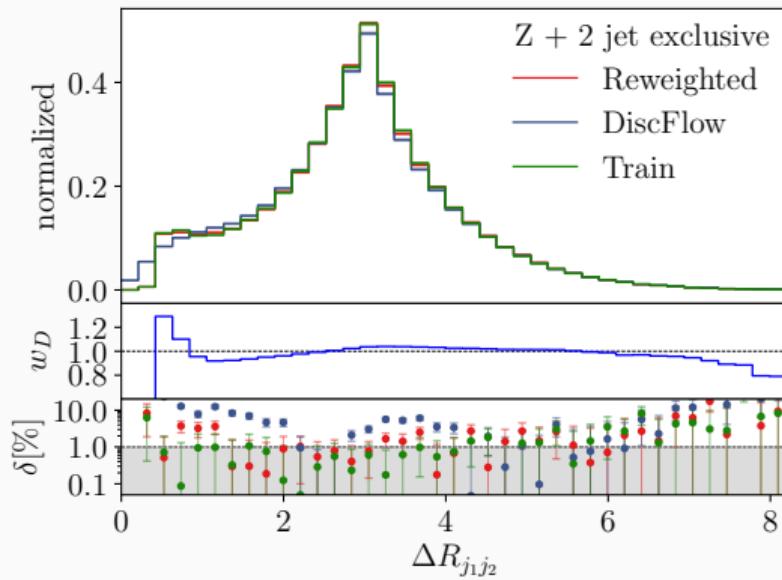
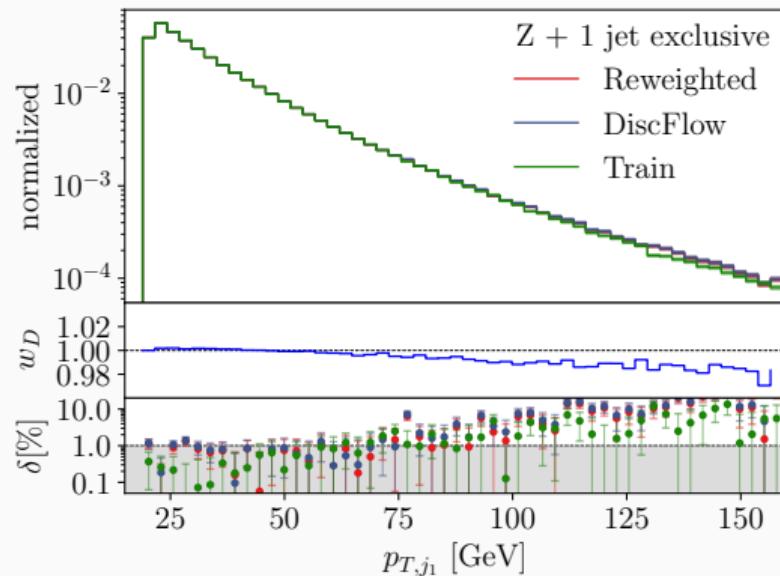


- Hyperparameter  $a$  determines the impact of the discriminator weights

$$\alpha = \alpha_0 \left| \frac{1}{2} - D(x) \right| .$$

# Results Discflow

## Discflow



- Improved precision in the unweighted distribution

# Estimating Uncertainties with Bayesian Invertible Neural Networks

---

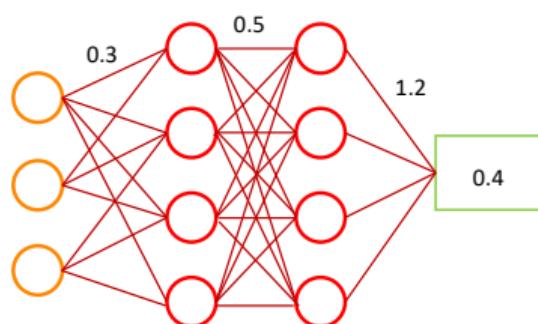
# Bayesian Neural Networks

Basic NN

Input layer

hidden layer

output layer

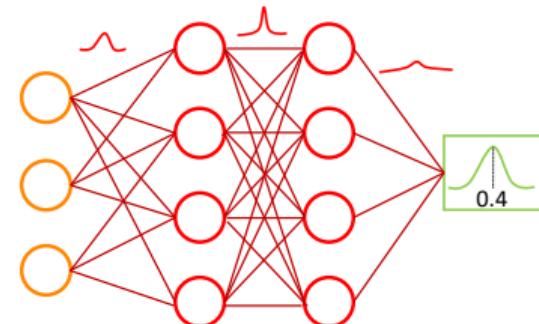


Bayesian NN

Input layer

hidden layer

output layer

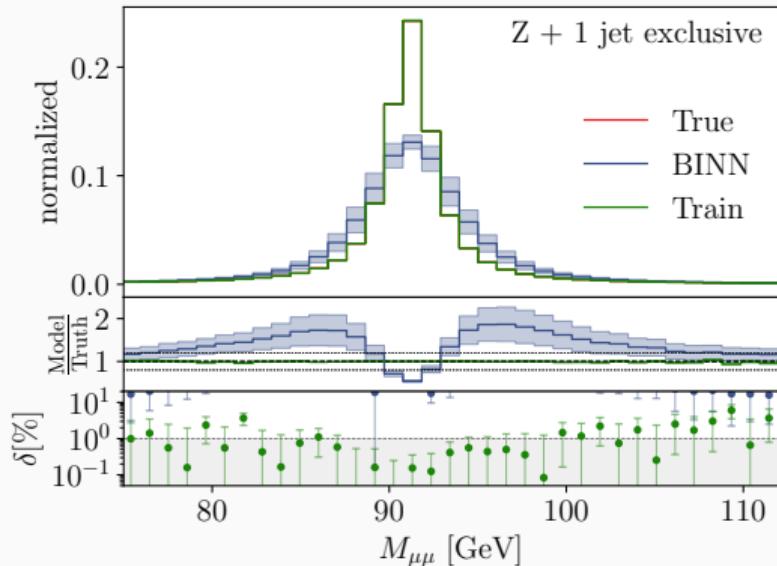
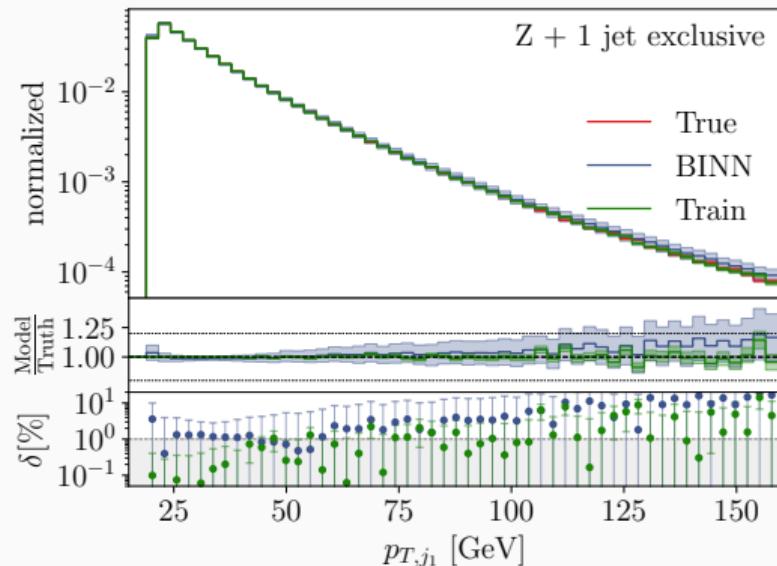


## How to estimate Uncertainties

$$\mathcal{L} = \sum_{i=1}^N \langle \log(p_Z(f^{-1}(x_i; \theta)) + \log \det \frac{\partial f^{-1}(x_i; \theta)}{\partial x_i} \rangle - \text{KL}(q_\phi(\theta), p(\theta))$$

1. Sample each  $\theta_i$  50 times from the learned Normal Distribution  $\theta_i \sim \mathcal{N}(\mu_i, \sigma_i)$
2. Generate histograms for each sampled  $\theta_i$
3. Calculate the mean value and standard deviation for each bin

# BINN Results



- Increasing uncertainty in areas of lower statistics
- The lack of performance is not captured through increasing uncertainties

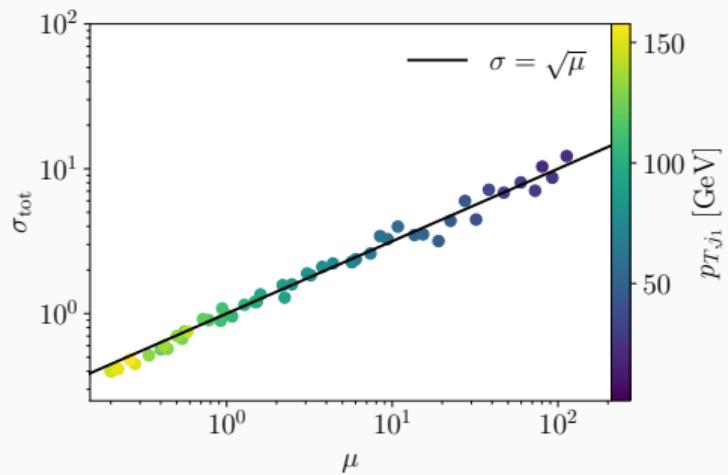
## Quantifying the Uncertainties

---

## Effect of the generated statistic

Verify effect of number of generated events

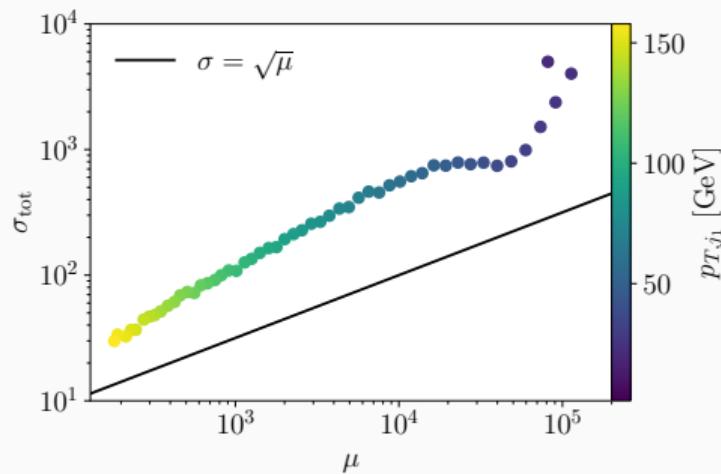
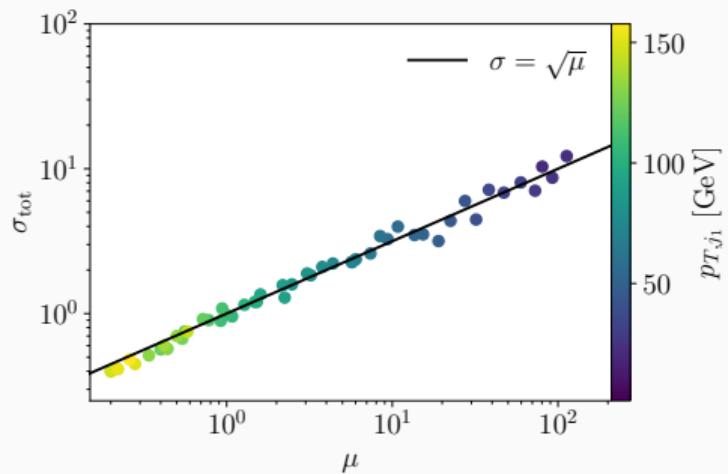
$$\sigma_{\text{tot}}^2 = \int d\theta q(\theta) [\langle n^2 \rangle_\theta - \langle n \rangle_\theta^2 + (\langle n \rangle_\theta - \langle n \rangle)^2] \equiv \sigma_{\text{stoch}}^2 + \sigma_{\text{pred}}^2.$$



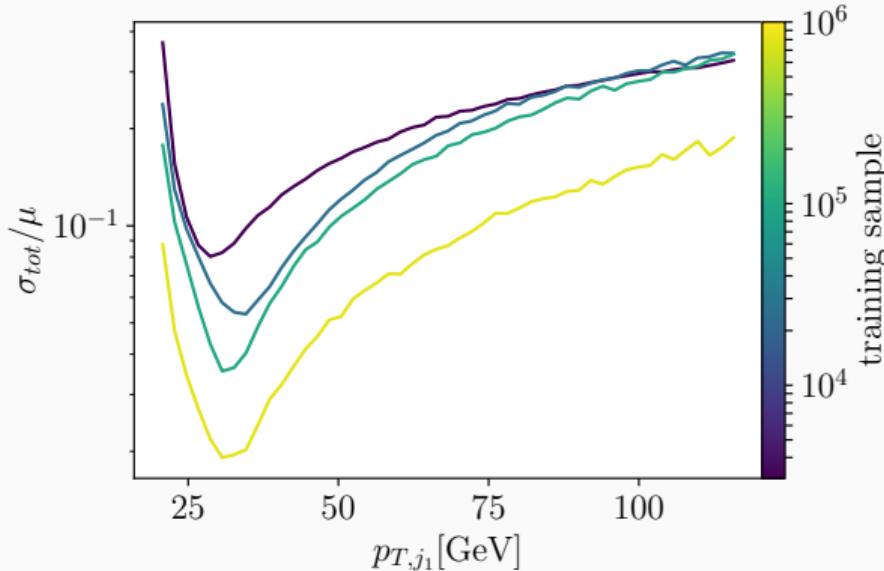
# Effect of the generated statistic

Verify effect of number of generated events

$$\sigma_{\text{tot}}^2 = \int d\theta q(\theta) [\langle n^2 \rangle_\theta - \langle n \rangle_\theta^2 + (\langle n \rangle_\theta - \langle n \rangle)^2] \equiv \sigma_{\text{stoch}}^2 + \sigma_{\text{pred}}^2.$$



## Effect of the training statistics



- The color encodes the size of the data set used for training
- The uncertainties decrease with a increasing number of training points

## Systematic Uncertainties

---

# Systematic uncertainties

## Augmentation of the transverse momentum

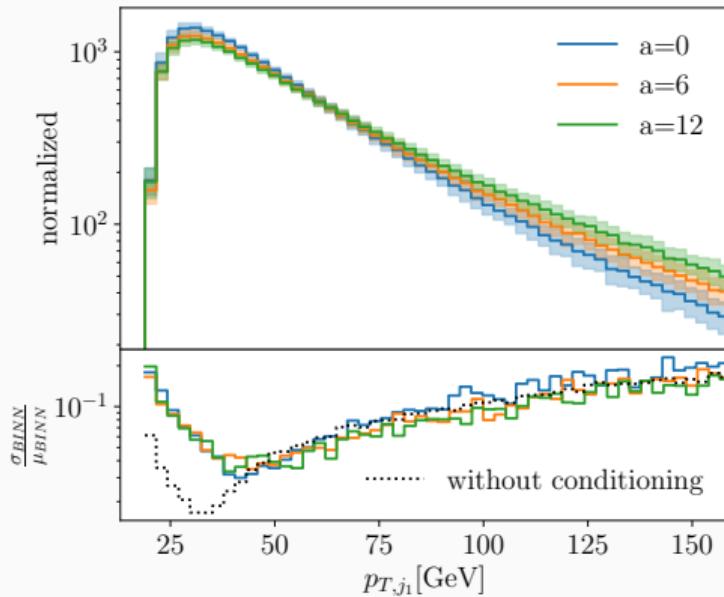
$$p'_T = p_T + f(p_T)$$

$$w = 1 + a \cdot \frac{(p_T - 15\text{GeV})^2}{100\text{GeV}^2} \quad a \in \{0, 3, 6, 9, \dots, 30\}$$

## Conditional training

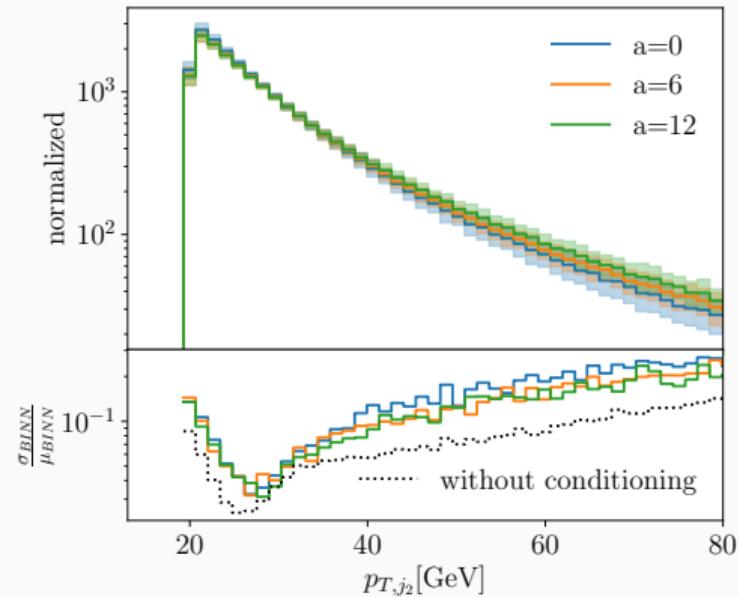
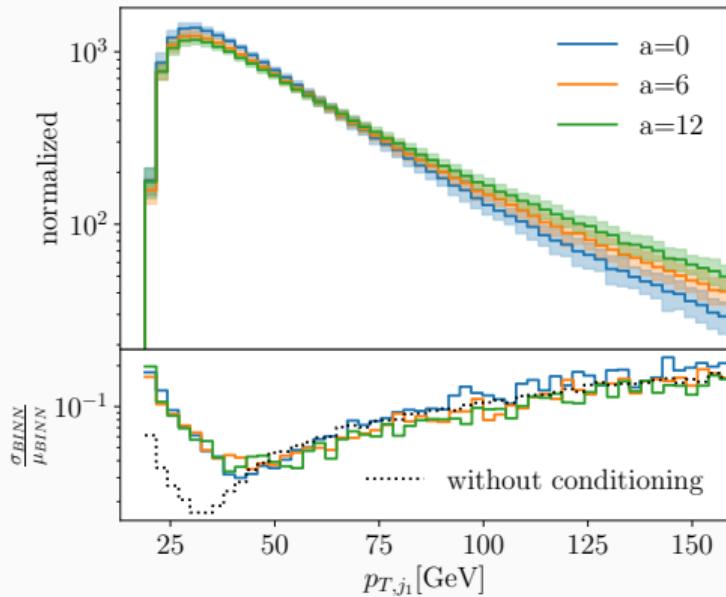
- $a$  randomly drawn for each event
- During event generation  $d'$  is fixed

## Systematic uncertainties



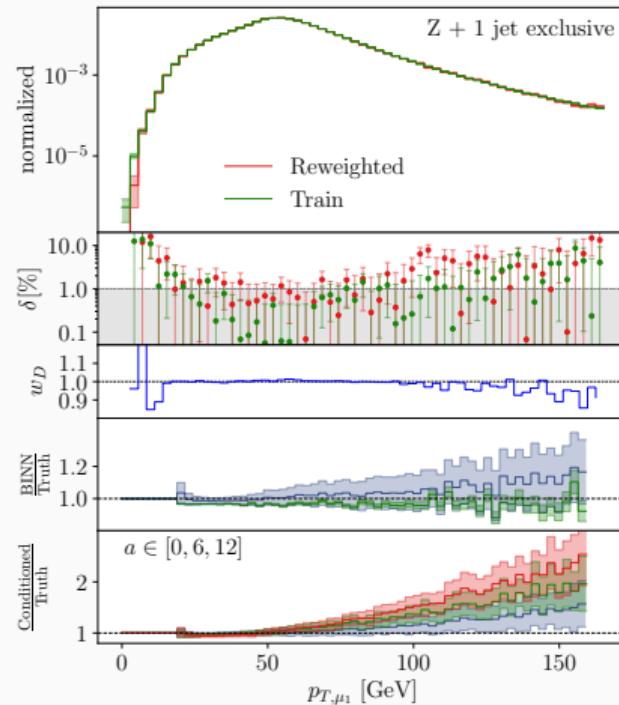
- Size of augmentation has no effect on the relative uncertainty

# Systematic uncertainties

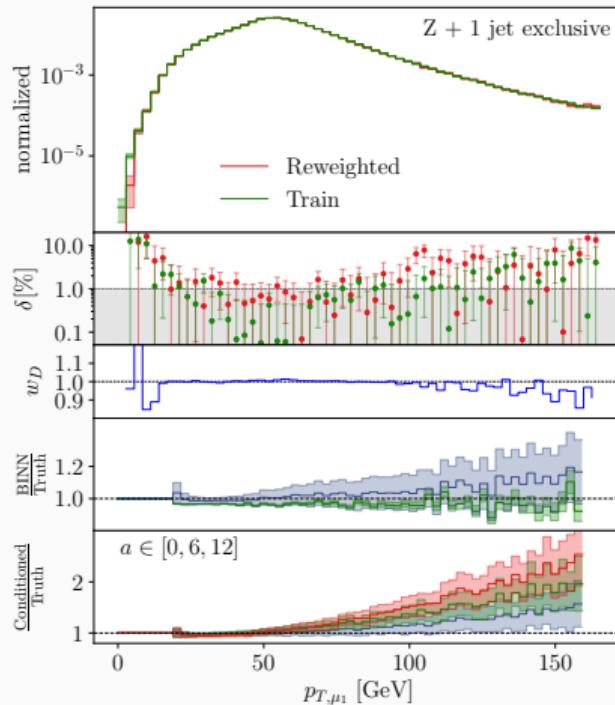


- Size of augmentation has no effect on the relative uncertainty
- Increasing uncertainties in  $p_{T,j_2}$  because of the additional source of correlations

# Final Results



# Final Results



Questions?