New generative models for LHC event generation

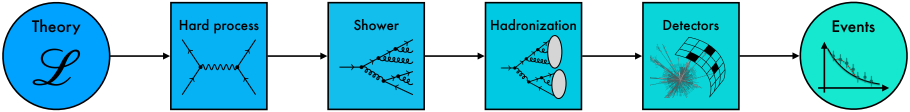
Nathan Huetsch

Institute for Theoretical Physics
Heidelberg University

A. Butter, NH, S. Palacios Schweitzer, T. Plehn, P. Sorrenson, J. Spinner
arXiv:2305.10475

IML Seminar July 2023

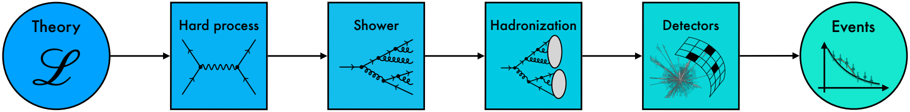
From theory to experiment in LHC physics



Fast and precise simulations are crucial for LHC physics

Figure from [A. Butter et al. 2203.07460], R. Winterhalder

Enhancing the simulation chain with ML



A non-exhaustive list:

- Loop integrals: [R. Winterhalder et al. 2112.09145]
- Importance Sampling: [T. Heimel et al. 2212.06172]
- Event unweighting: [M. Backes et al. 2012.07873]
- Hadronization: [J. Chan et al. 2305.17169]
- Detector simulation: [C. Krause et al. 2110.11377]
- **End-to-end event generation: [A. Butter et al. 2305.10475]**

Figure from [A. Butter et al. 2203.07460], R. Winterhalder

Generative Machine Learning

- Learn the **underlying density** $p_{\text{data}}(x)$ from a set of samples X_{train}

- Most generative models learn a **transformation**

$$x \sim p_{\text{model}}(x|\theta) \approx p_{\text{data}}(x) \quad \longleftarrow \quad \epsilon \sim p_{\text{latent}}(\epsilon) = \mathcal{N}(0,1)$$

- LHC event generation: Low-dimensional data, high precision required
  $\Rightarrow$ **Normalizing Flows** very successful [A. Butter et al. 2110.13632]

New generative models

Diffusion

State-of-the-art image generation
Midjourney, StableDiffusion, ...

Example: Detector Simulation
V. Mikuni et al. 2206.11898

Transformer

State-of-the-art language generation
ChatGPT, Bard, ...

Example: QCD jet generation
T. Finke et al. 2303.07364

New generative models

Diffusion

State-of-the-art image generation
Midjourney, StableDiffusion, ...

Example: Detector Simulation
V. Mikuni et al. 2206.11898

Transformer

State-of-the-art language generation
ChatGPT, Bard, ...

Example: QCD jet generation
T. Finke et al. 2303.07364

Can they challenge Normalizing Flows in LHC event generation?

Diffusion models
Model transformation as time-dependent diffusion process

$$x_0 \sim p_{\text{model}}(x_0|\theta) \quad \overset{t}{\longleftrightarrow} \quad x_T = \epsilon \sim \mathcal{N}(0,1)$$

$\Rightarrow$: Gradually **add noise to data samples** to transform them to gaussians
$\Leftarrow$: Gradually **remove noise from gaussians** to obtain data samples

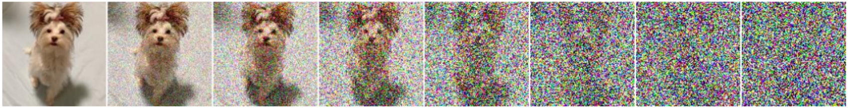


Figure from [Y. Song et al. 2011.13456]

Denoising Diffusion Probabilistic Model [J. Ho et al. 2006.11239]

- Forward: **discrete Markov process**

$$p(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\kappa_t}x_{t-1}, \kappa_t)$$
$$\Rightarrow p(x_t|x_0) = \mathcal{N}(x_t; \alpha_t x_0, \beta_t)$$
$$\text{with} \quad \alpha_t = \prod_{i=1}^{t} \sqrt{1-\kappa_t} \quad \text{and} \quad \beta_t = 1 - \prod_{i=1}^{t}(1-\kappa_t)$$

- Reverse: **approximated as same form**

$$q_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_\theta(x_t, t))$$
$$q_\theta(x_{t-1}|x_t) \approx p(x_{t-1}|x_t, x_0)$$

- NN: **predict the noise** ϵ_θ

$$\mathcal{L}_{\mathrm{DDPM}} = \left\langle C_t \left| \epsilon_\theta(x,t) - \epsilon(x,t|x_0) \right|^2 \right\rangle_{x_0 \sim p_{\mathrm{data}}, \epsilon \sim \mathcal{N}(0,1), t \sim \mathcal{U}(0,T)}$$

Conditional Flow Matching [Y. Lipman et al. 2210.02747]

- Individual samples: **continuous ODE**

$$\frac{dx(t)}{dt} = v(x,t)$$

- Density: **continuity equation**

$$\frac{\partial p(x,t)}{\partial t} + \nabla_x \left[p(x,t) v(x,t) \right] = 0$$

$$\text{s.t.} \quad p(x,t) \quad \rightarrow \quad \begin{cases} p_{\text{data}}(x) & t \to 0 \\ \mathcal{N}(x;0,1) & t \to 1 \end{cases}$$

- NN: **predict the velocity field $\mathbf{v}_\theta$**

$$\mathcal{L}_{\text{CFM}} = \left\langle \left| v_\theta(x,t) - v(x,t|x_0) \right|^2 \right\rangle_{x_0 \sim p_{\text{data}}, \epsilon \sim \mathcal{N}(0,1), t \sim \mathcal{U}([0,1])}$$

LHC use case: Z+jets

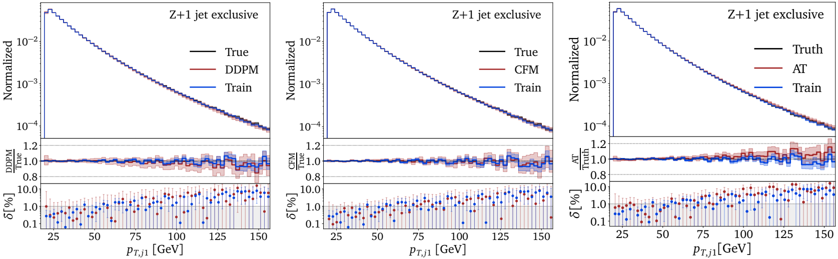
- Follow example process from [A. Butter et al. 2110.13632]:
  Leptonically decaying **Z boson with variable number of QCD jets**

$$pp \to Z_{\mu\mu} + \{1, 2, 3\} \text{ jets} .$$

- Events generated with **Sherpa at 13 TeV**.
  Jets are defined with anti-k_T algorithm and applying basic cuts

$$p_{T,j} > 20 \text{ GeV} \qquad \text{and} \qquad \Delta R_{jj} > 0.4$$

Z+jets: Transverse momenta



→ Sub-1-percent precision in the bulk of the distribution

→ Traindata precision over the complete distribution

What about uncertainties?

Learned phase space density comes with **uncertainty**

→ Lack of training data

→ Insufficient model flexibility

→ Stochastic optimization of model parameters

Bayesian Neural Networks

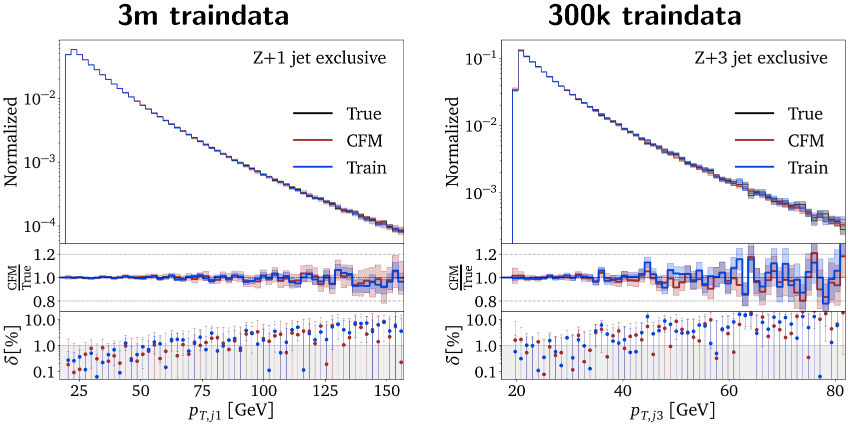
1. Replace deterministic weights θ with **distributions over weights**
2. Place a **(meaningless) prior** $p(\theta)$ over the weights
3. Train network via **variational approximation** of the posterior

$$q(\theta) \approx p(\theta|X_{\text{train}}) = \frac{p(X_{\text{train}}|\theta)p(\theta)}{p(X_{\text{train}})}$$

4. Evaluate the network by calculating the **posterior expectation**

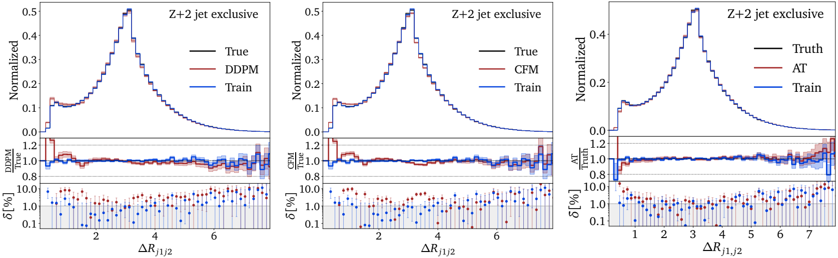
$$\langle p \rangle(x) = \int d\theta \; p(x|\theta)p(\theta|X_{\text{train}})$$

Z+jets: Bayesian Network Uncertainty



→ Bayesian NN correctly captures uncertainty from lack of training data

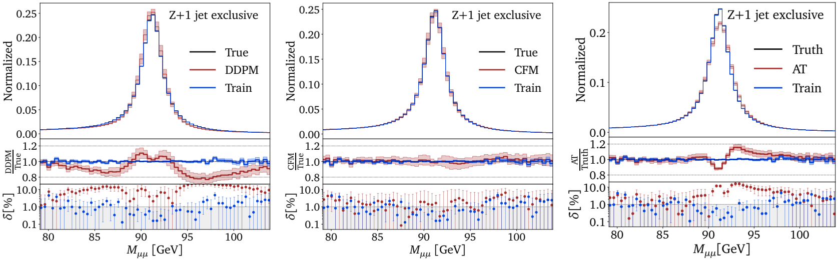
Z+jets: Jet seperation



→ Percent-level precision in the bulk of the distribution

→ Hard cut at $\Delta R = 0.4$ is learned well but not perfect

Z+jets: Lepton mass peak



→ Percent-level precision in the $\mu\mu$ mass peak!

Summary

- We adapted two diffusion models and an autoregressive transformer model to LHC event generation
- Bayesian versions allows us to quantify the uncertainties
- These new models match or even surpass the percent-level precision of Normalizing Flows in end-to-end LHC event generation

Outlook

- The next step is to incorporate these models into different parts of the LHC simulation and analysis chain
- This includes, but is not limited to Importance Sampling, Matrix Element Methods, Unfolding, ...
- We expect that LHC physics will benefit from different model classes