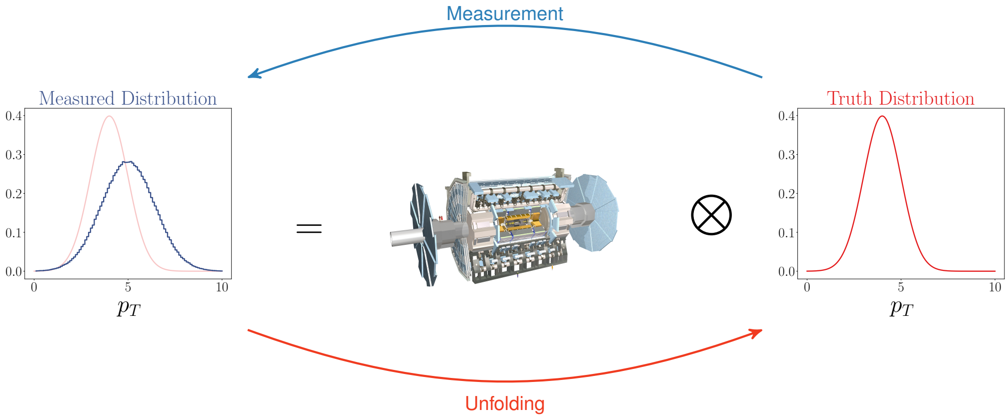
ML Unfolding based on conditional Invertible Neural Networks using Iterative Training

Mathias Backes (KIP, ITP)

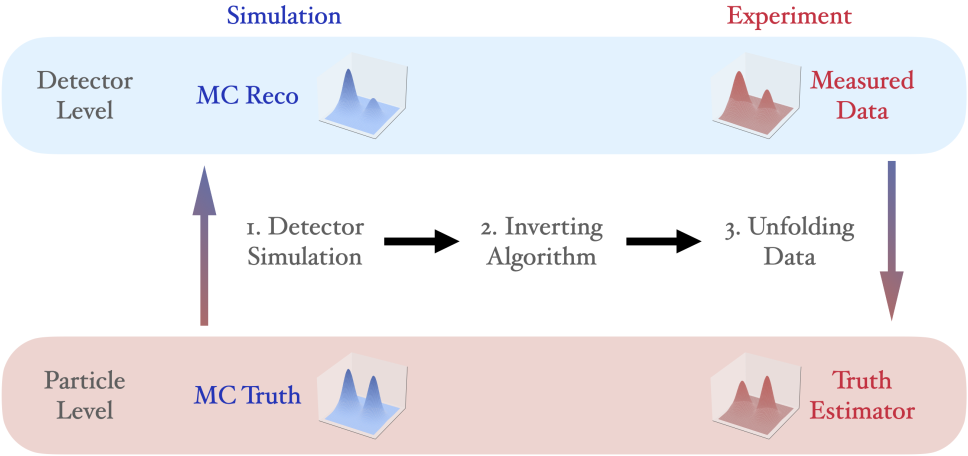
with Anja Butter (LPNHE, ITP), Monica Dunford (KIP) and Bogdan Malaescu (LPNHE).

Why Unfold?



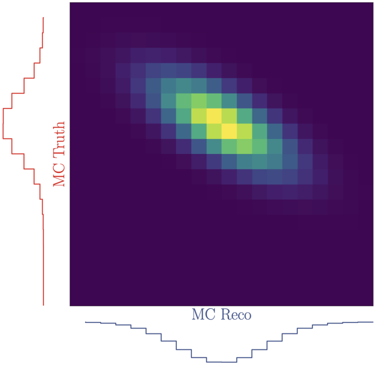
- Parameter tuning of theories
- Use of the data in the future
- Comparison to other experiments

Basic Concept



"Classical" Unfolding Algorithms

Migration matrix:

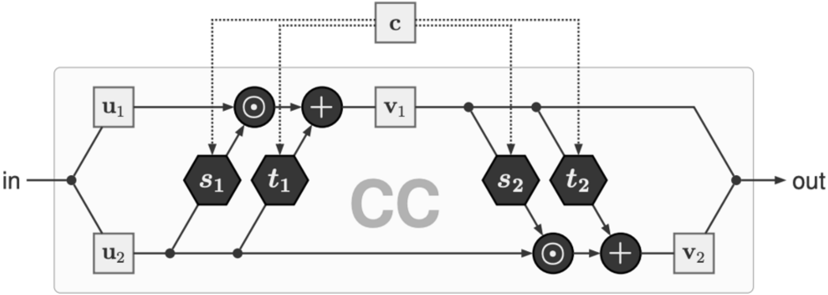


Three main problems with matrix-based unfolding algorithms:

- Binning choice involves an information loss
- No high-dimensional unfolding (only up to three dimensions)
- Sensitivity to "hidden" observables"

$\Rightarrow$ Use full phase space information with ML approaches.

Conditional Invertible Neural Networks (cINN)

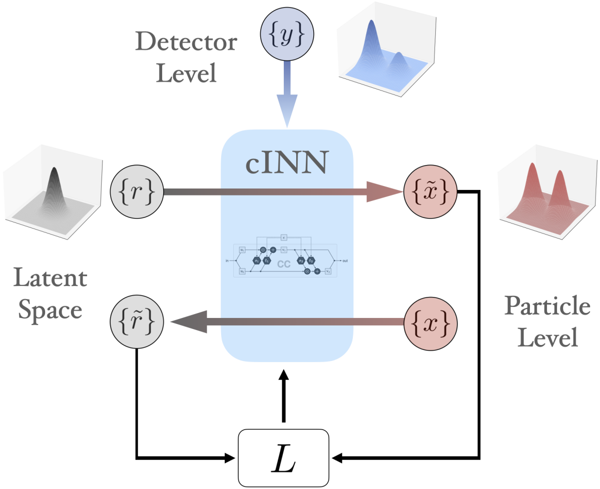
$\vec{u}$  $\vec{v}$

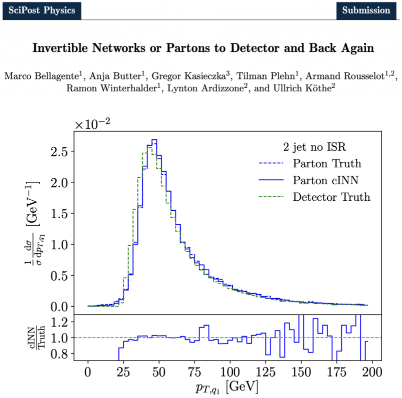
$$u_1 = (v_1 - t_1(u_2, c)) \oslash \exp(s_1(u_2, c)) \qquad v_1 = u_1 \odot \exp(s_1(u_2, c)) + t_1(u_2, c)$$

$$u_2 = (v_2 - t_2(v_1, c)) \oslash \exp(s_2(v_1, c)) \qquad v_2 = u_2 \odot \exp(s_2(v_2, c)) + t_2(v_1, c)$$

Source: arXiv [1907.02392]

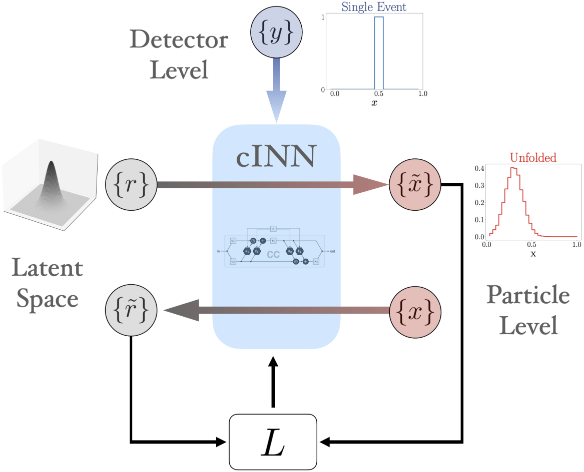
cINN Unfolding

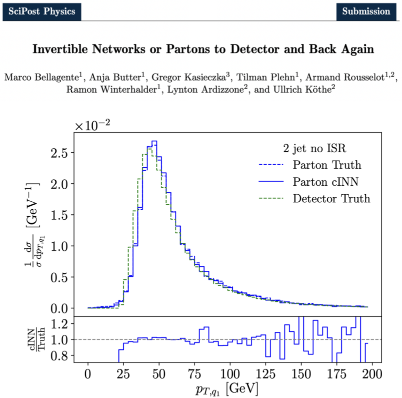




Source: arXiv [2006.06685]

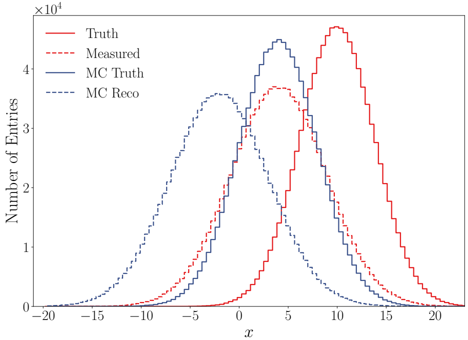
cINN Unfolding

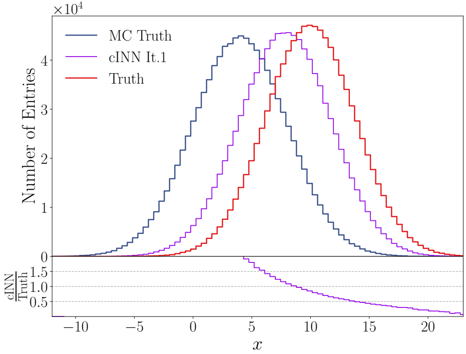




Source: arXiv [2006.06685]

cINN Unfolding



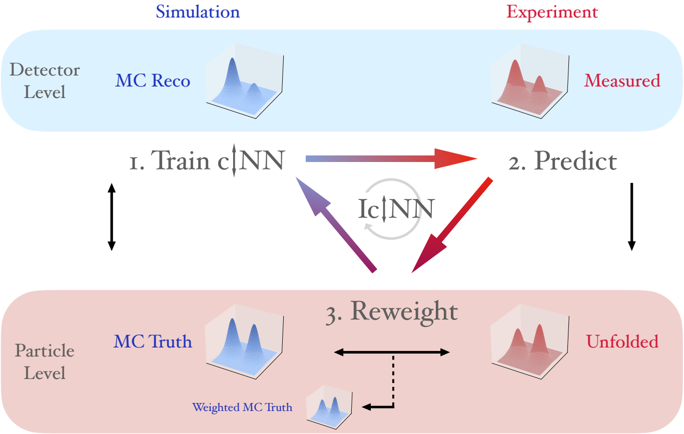


- Extreme toy example with large MC-data-differences and significant detector effects
- cINN unfolded distribution shows a strong bias towards the MC truth

$$p(t|r) = \frac{p(r|t) \cdot p(t)}{p(r)}, \quad \text{with} \quad t = \text{truth}, \quad r = \text{reco}$$

$\Rightarrow$ Iterative approach needed

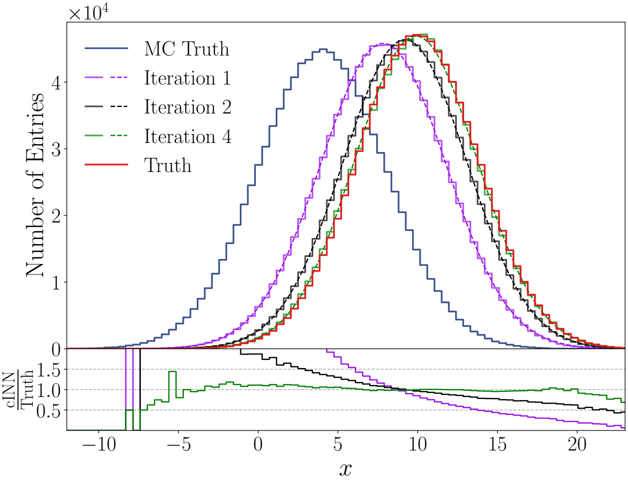
Iterative Approach



Advantages:

- Structures present in the data are encoded implicitly in the MC Truth
- General similarities to matrix based iterative bayesian-like unfolding
- Maintain event-wise probabilistic distributions

Results for the Iterative Approach



- Construct an analytically solvable toy model
- Use Bayes theorem to construct pseudo-inverse:

$$p(t|r) = \frac{p(r|t) \cdot p(t)}{p(r)}$$

- Apply pseudo-inverse to measured distribution:

$$p_u(t) = \int p(t|r) p_M(r) \mathrm{d}r$$

Unfolding an EFT Process

- Simulating the process

  $pp \to Z\gamma\gamma$ with $Z \to \mu^- \mu^+$

- MC $\to$ pure SM

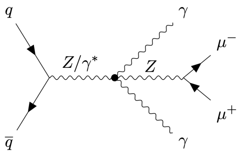
- Data $\to$ SM + EFT contribution of

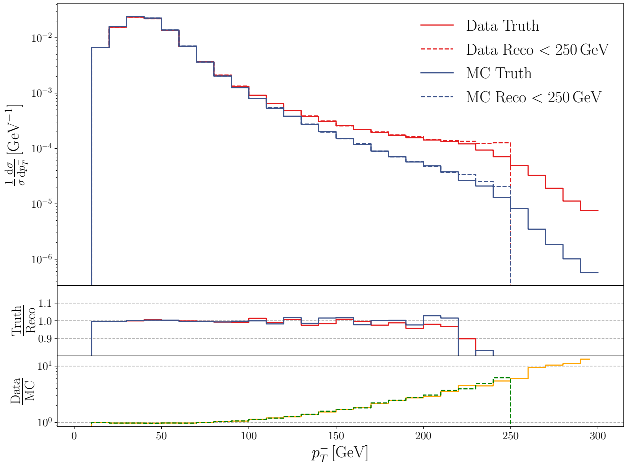
  $\mathcal{L}_{T,8} = \frac{C_{T,8}}{\Lambda^4} B_{\mu\nu} B^{\mu\nu} B_{\alpha\beta} B^{\alpha\beta}$

  with $\frac{C_{T,8}}{\Lambda^4} = \frac{2}{\mathrm{TeV}^4}$

- Applied detector smearing:

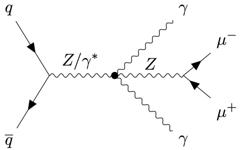
  $\Delta p_T = p_T \cdot \sqrt{0.025^2 + p_T^2 \cdot 3.5 \cdot 10^{-8}}$

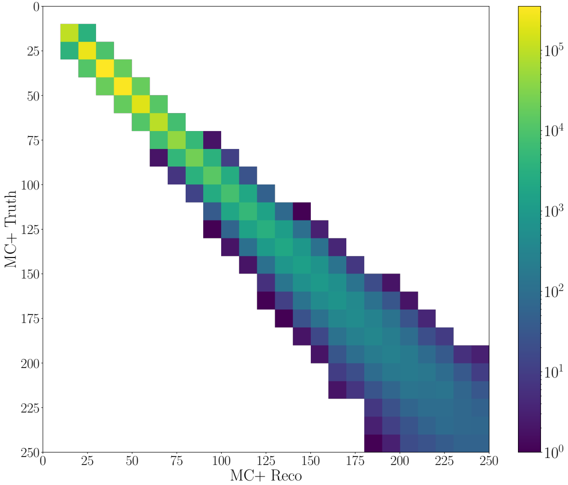




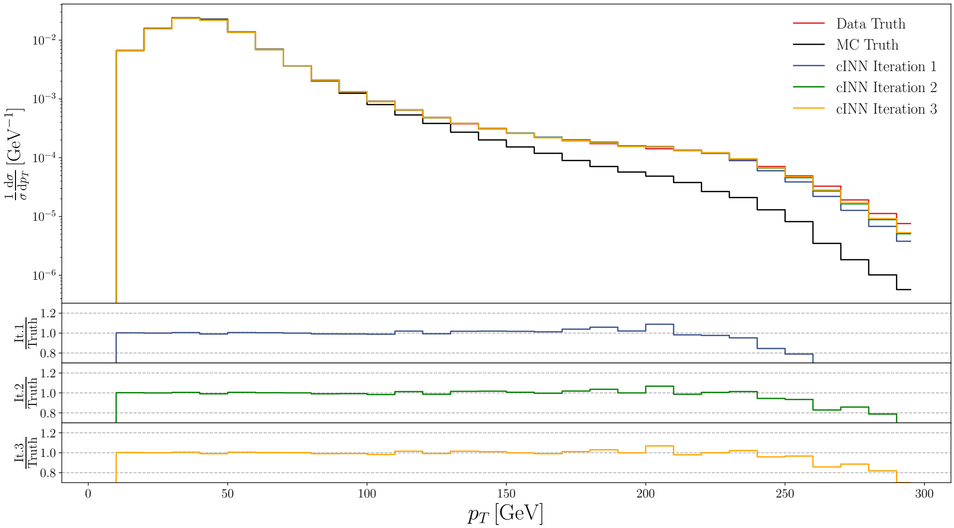
Unfolding an EFT Process

- Simulating the process
  $pp \to Z\gamma\gamma$ with $Z \to \mu^- \mu^+$
- MC $\to$ pure SM
- Data $\to$ SM + EFT contribution of
  $\mathcal{L}_{T,8} = \frac{C_{T,8}}{\Lambda^4} B_{\mu\nu} B^{\mu\nu} B_{\alpha\beta} B^{\alpha\beta}$
  with $\frac{C_{T,8}}{\Lambda^4} = \frac{2}{\text{TeV}^4}$
- Applied detector smearing:
  $\Delta p_T = p_T \cdot \sqrt{0.025^2 + p_T^2 \cdot 3.5 \cdot 10^{-8}}$





Unfolding an EFT Process



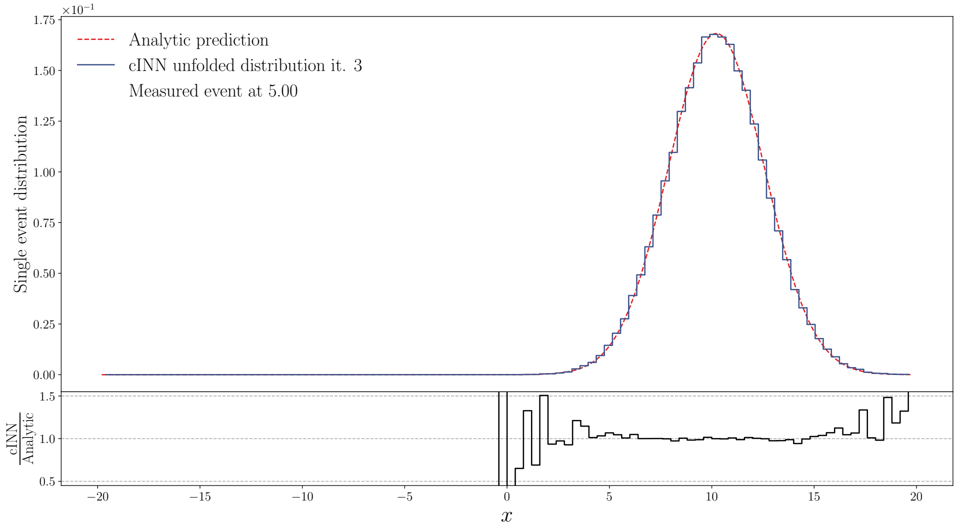
Conclusion / Outlook

- Implementation of an iterative cINN unfolding algorithm and application to a physical example
- Central Idea: still obtain the cINN result of a probabilistic unfolded **distribution** while iteratively reducing the bias towards the MC simulation
- Next step: application to real experimental data

Thank you for your attention!

Additional Material

Single Event Unfolded



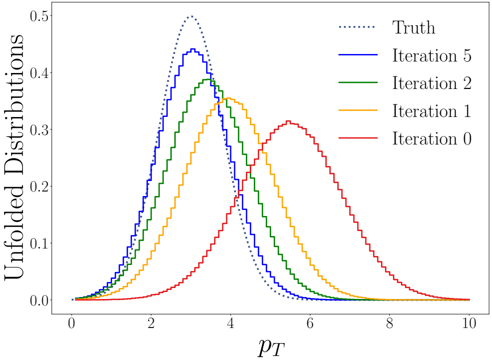
"Classical" Unfolding: Iterative Bayesian Unfolding

1. Choose an initial prior $t_i^{(0)}$
2. Calculate the unfolding function

$$R^{(n)}(t_j|r_i) = \frac{R(r_i|t_j)\, t_i^{(n-1)}}{\sum_k R(r_k|t_j)\, t_k^{(n-1)}}$$

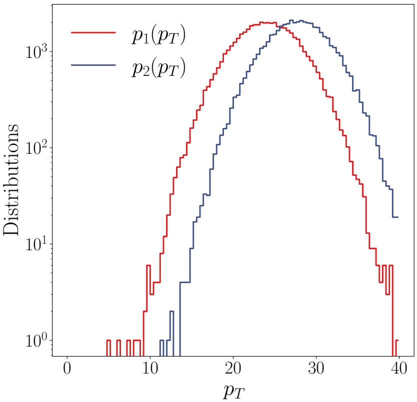
3. Recalculate the truth distribution

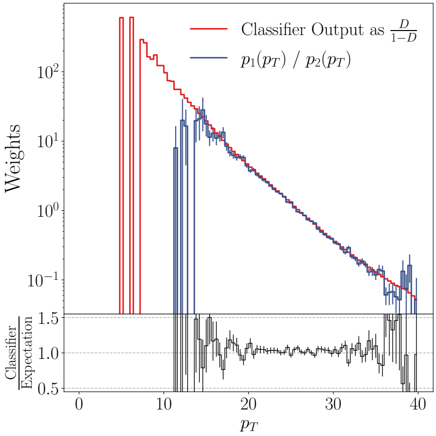
$$t_j^{(n)}(t) = \sum_k R^{(n)}(t_j|r_k)\, r_{\mathrm{Meas},k}$$



$\Rightarrow$ Balance between bias and uncertainties

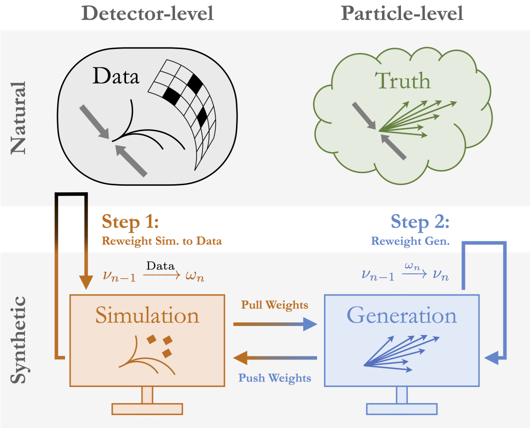
Reweighting Distributions





$$p_1(p_T) = \frac{D}{1-D}\, p_2(p_T)$$

Omnifold



Problems:

- MC and data need to cover the same phase space
- E.g. observables based on high jet multiplicities
  $\Rightarrow$ Not necessarily multi-jet-event in MC
- Range of validity?

Source: arXiv [1911.09107]

Analytic Toy Example

- Gaussian smearing:

$$p(r|t) = \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp\left(-\frac{(r-(t+\mu_s))^2}{2\sigma_s^2}\right).$$

- Bayes theorem:

$$p(t|r) = \frac{p(r|t)\cdot p(t)}{p(r)}.$$

- Unfolding a measured distribution $p_M(r)$ using Gaussian functions for $p(r)$, $p(t)$ and $p_M(r)$:

$$p_u(t) = \int p(t|r)p_M(r)\mathrm{d}r = \frac{1}{2\pi}\sqrt{\frac{\sigma_r^2}{\sigma_t^2\sigma_s^2\sigma_M^2}} \int \mathrm{d}r \exp\left(-\frac{(r-(t+\mu_s))^2}{2\sigma_s^2} - \frac{(t-\mu_t)^2}{2\sigma_t^2} + \frac{(r-\mu_r)^2}{2\sigma_r^2} - \frac{(r-\mu_M)^2}{2\sigma_M^2}\right)$$

- Evaluating leads to gaussian unfolded distribution with:

$$\mu_u = \frac{\mu_m\sigma_t^2 + \mu_t\sigma_s^2 - \mu_s\sigma_t^2}{\sigma_s^2+\sigma_t^2}, \qquad \sigma_u = \frac{\sqrt{\sigma_t^2\sigma_M^2 + \sigma_t^2\sigma_s^2 + \sigma_s^4}\sigma_t}{\sigma_s^2+\sigma_t^2}.$$

cINN Loss function

Minimize loss function:

$$\begin{aligned}
\mathcal{L} &= -\langle \log p(\theta|x,y)\rangle_{x\sim f, y\sim g} \\
&= -\langle \log p(x|\theta,y)\rangle_{x\sim f, y\sim g} - \langle \log p(\theta|y)\rangle_{y\sim g} + \langle \log p(x|y)\rangle_{x\sim f, y\sim g} \\
&= -\langle \log p(x|\theta,y)\rangle_{x\sim f, y\sim g} - \lambda\theta^2 + \text{const.} \\
&= -\langle \log p(z(x)|\theta,y)\rangle_{x\sim f, y\sim g} - \langle \log \left|\frac{\mathrm{d}z}{\mathrm{d}x}\right|\rangle_{x\sim f, y\sim g} - \lambda\,\theta^2 + \text{const.}
\end{aligned}$$

θ = cINN parameter, x = Parton Level, y = Detector level, z = Latent space variable

Source: arXiv [1907.02392]