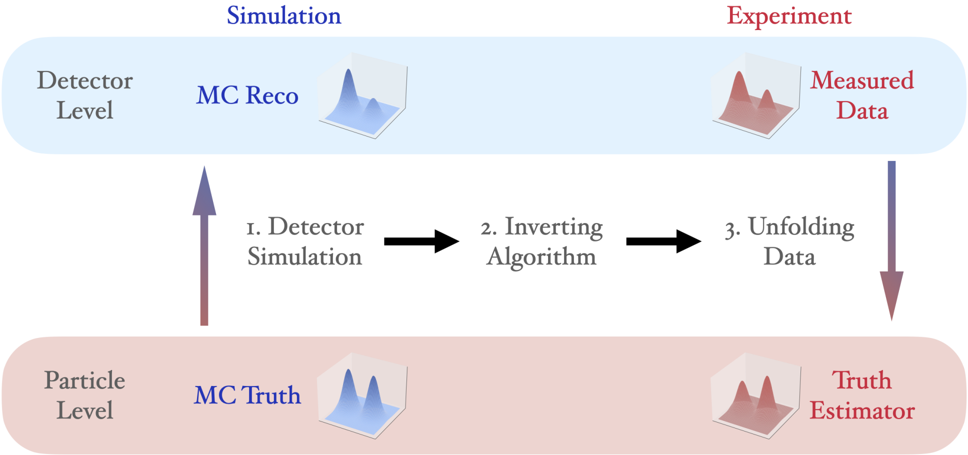
ML Unfolding based on conditional Invertible Neural Networks using Iterative Training

Mathias Backes (KIP Heidelberg)

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

Basic Concept

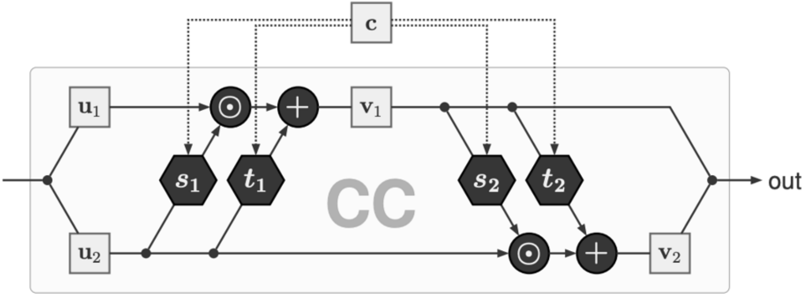


Part 1:

cINN Unfolding

Conditional Invertible Neural Networks (cINN)

$\vec{u}$



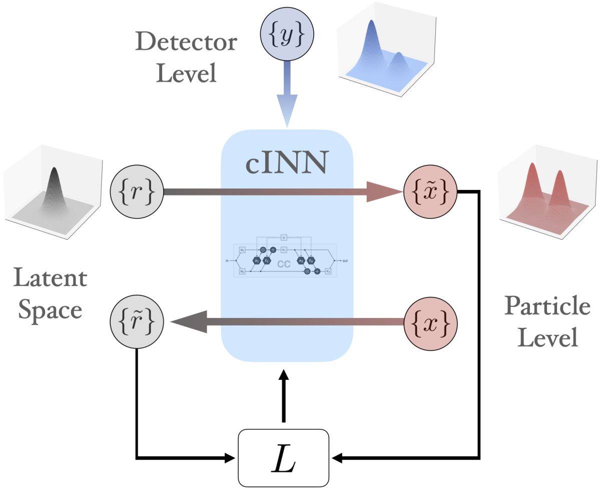
$\vec{v}$

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

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

Source: arXiv [1907.02392]

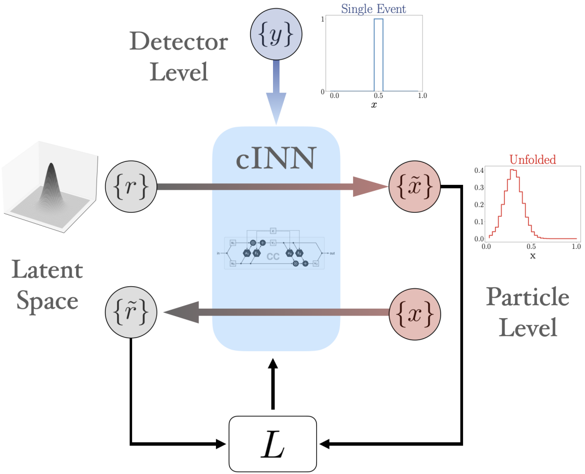
cINN Unfolding - Training



- Train on Monte Carlo simulation
- Propagate (truth, reco) event pairs through the network
- Loss forces latent space to be gaussian
- Result: conditional bijective mapping between gaussian latent space and truth-level information

Source: arXiv [2006.06685]

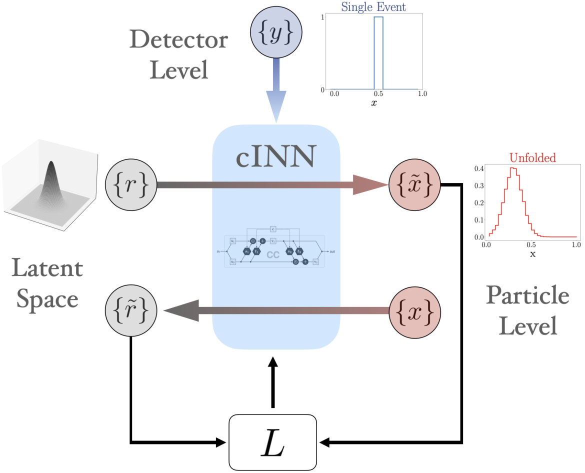
cINN Unfolding - Evaluation

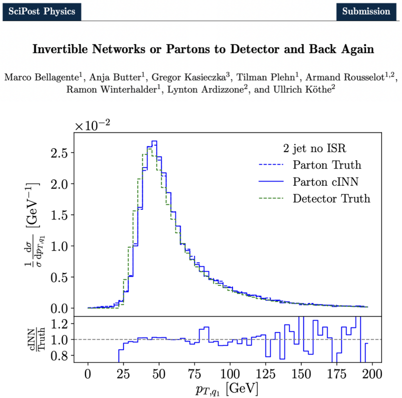


- Unfold measured data on an event-by-event basis
- Sample in gaussian latent space
- Probabilistic single-event unfolding

Source: arXiv [2006.06685]

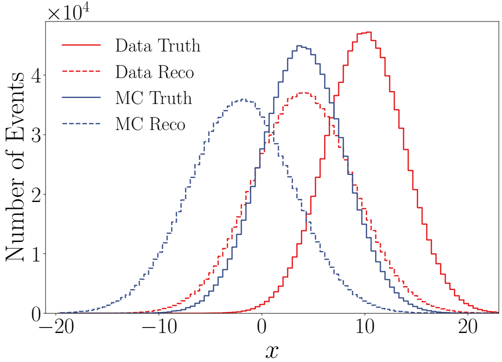
cINN Unfolding - Evaluation

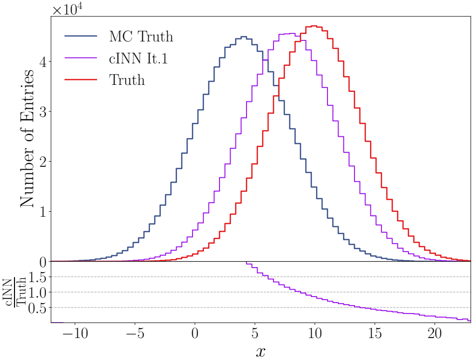




Source: arXiv [2006.06685]

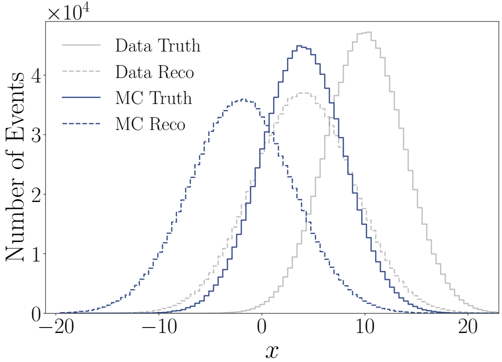
cINN Unfolding

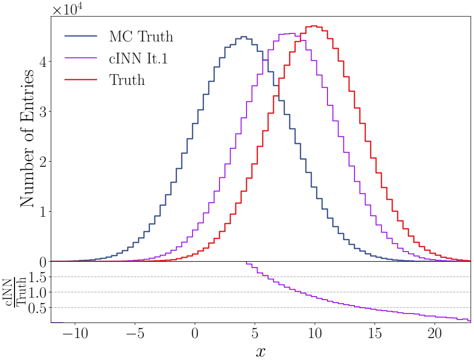




Differences between Data and MC induce biases in the unfolding result
$\Rightarrow$ Iterative approach needed

cINN Unfolding

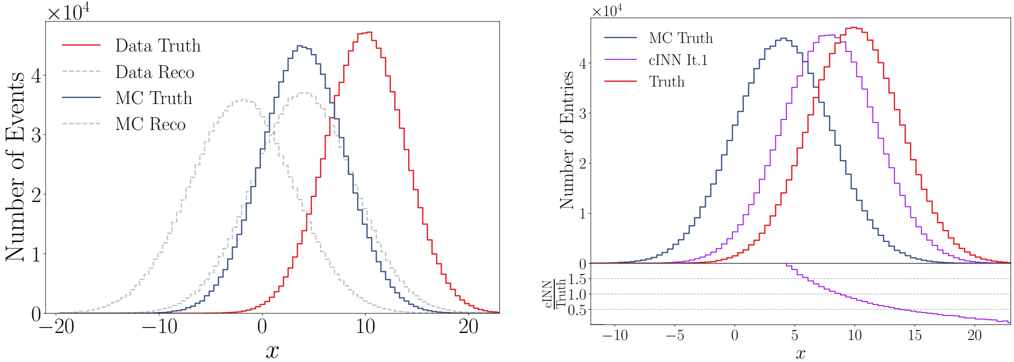




Differences between Data and MC induce biases in the unfolding result

$\Rightarrow$ Iterative approach needed

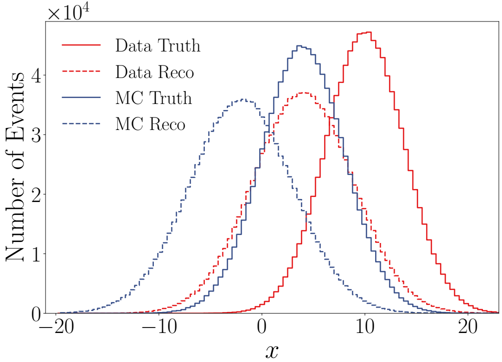
cINN Unfolding

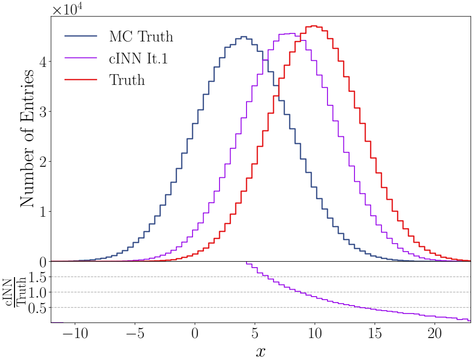


Differences between Data and MC induce biases in the unfolding result

⇒ Iterative approach needed

cINN Unfolding





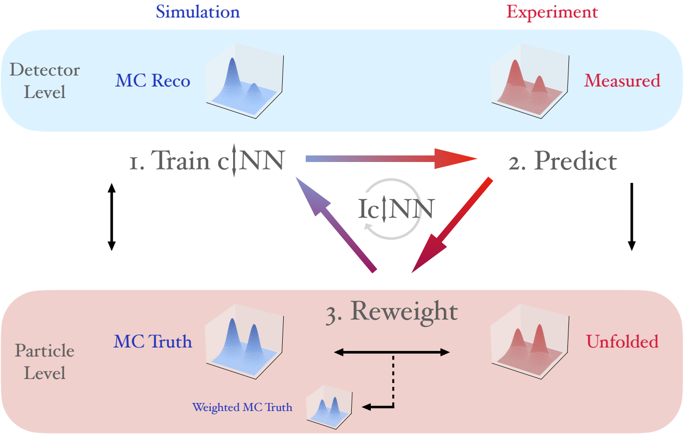
Differences between Data and MC induce biases in the unfolding result

$\Rightarrow$ Iterative approach needed

Part 2:

Iterative cINN Unfolding

Iterative Approach

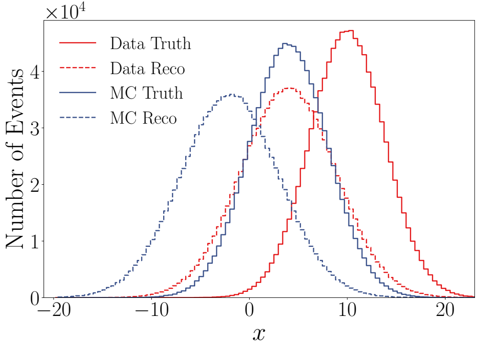


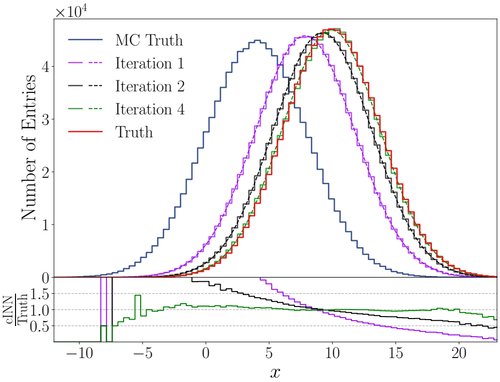
Features:

- 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

Publication: [2212.08674]

Results for the Iterative Approach





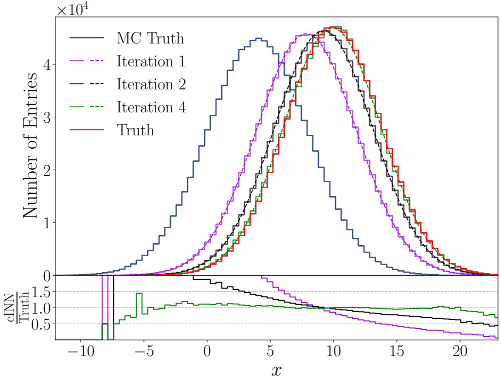
Results for the Iterative Approach

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

$$p(t|r) = \frac{p(r|t) \cdot p(t)}{\int p(r|t) \cdot p(t)\,\mathrm{d}t},$$

- Apply pseudo-inverse to measured distribution:

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



Statistical Uncertainties and Correlations

Sources of statistical uncertainties:

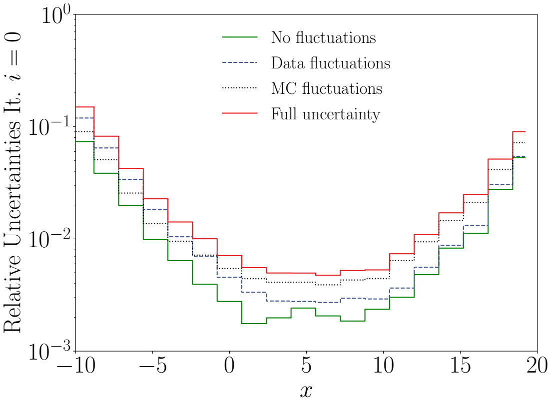
- Training of the network initialized randomly
- Fluctuations in the data
- Fluctuations in the Monte Carlo simulation

Calculation of covariance matrices with fluctuated pseudo-experiments (bootstrap method):

$$\text{cov}_{ij} = \frac{1}{N_f}\sum_{1}^{N_f}\left(t_i^{\text{Unf}} - \overline{t_i^{\text{Unf}}}\right)\left(t_j^{\text{Unf}} - \overline{t_j^{\text{Unf}}}\right)$$

$$\sigma_i = \sqrt{\text{cov}_{ii}}$$

Relative uncertainties without reweighting



Statistical Uncertainties and Correlations

Sources of statistical uncertainties:

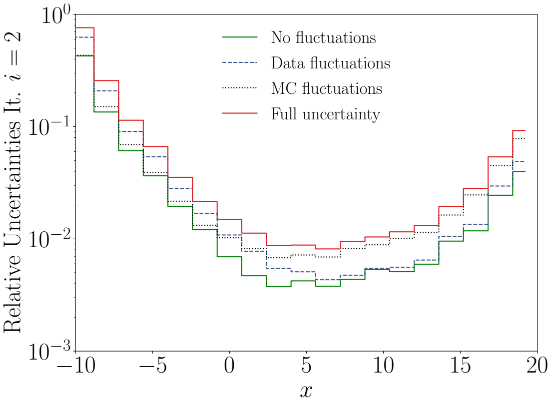
- Training of the network initialized randomly
- Fluctuations in the data
- Fluctuations in the Monte Carlo simulation

Calculation of covariance matrices with fluctuated pseudo-experiments (bootstrap method):

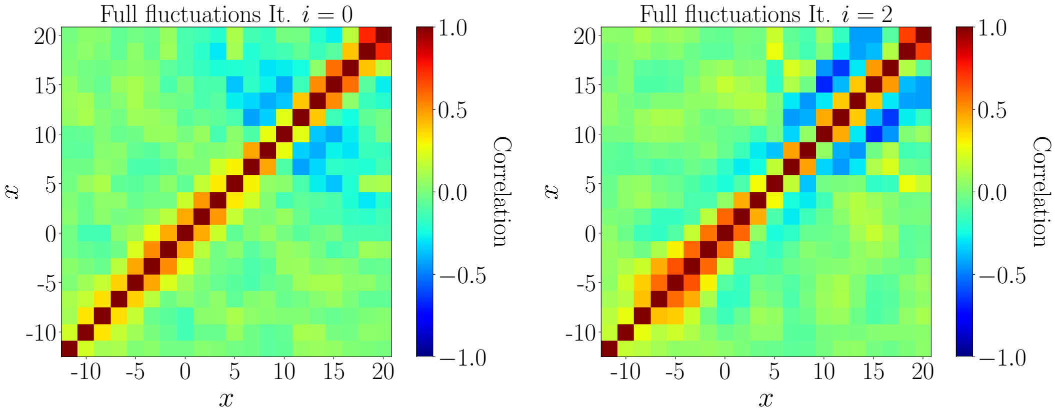
$$\mathrm{cov}_{ij} = \frac{1}{N_f} \sum_{1}^{N_f} \left(t_i^{\mathrm{Unf}} - \overline{t_i^{\mathrm{Unf}}} \right) \left(t_j^{\mathrm{Unf}} - \overline{t_j^{\mathrm{Unf}}} \right)$$

$$\sigma_i = \sqrt{\mathrm{cov}_{ii}}$$

Relative uncertainties with two reweightings



Correlations from Full Uncertainties

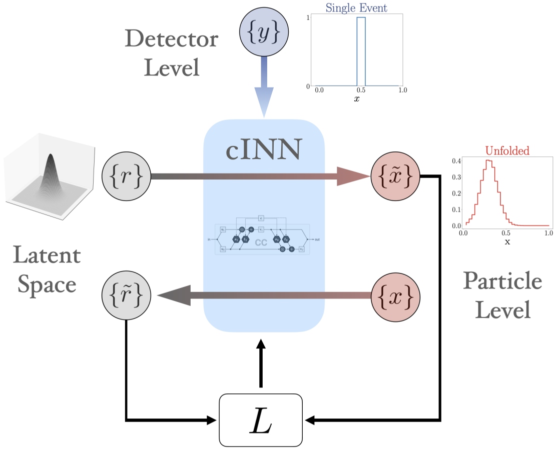


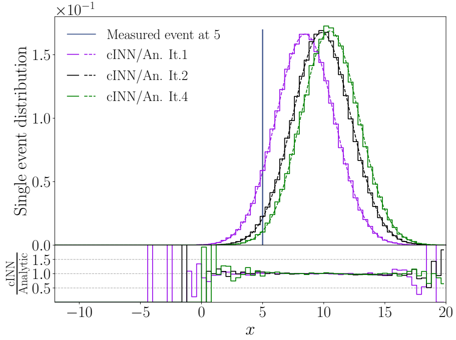
- Size of correlations increases for more iterations
- Range of significant bin-to-bin correlations is driven by the resolution

Part 3:

Probabilistic Unfolding

Unfolding a Single Event



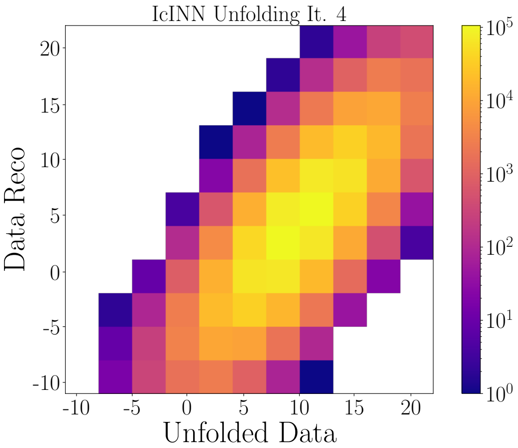


Reasons for Single Event Unfolding

Several interesting features:

- Keep track of unfolded- and reco-level quantities
- Possibility to implement reco cuts after the unfolding the event
- Simple derivation of secondary observables

$\Rightarrow$ Problem: validation needed



Reminder: Matrix-Based Unfolding Algorithms

- Probabilistic response matrix

$$R_{ij}^{(MC)} = p(r^{(MC)} \in (\text{bin})_i | t^{(MC)} \in (\text{bin})_j)$$

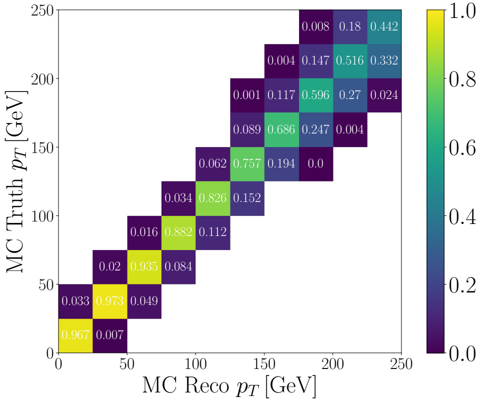
- Folding equation to connect truth-level and reco-level

$$r_i^{(MC)} = \sum_j R_{ij}^{(MC)} t_j^{(MC)}$$

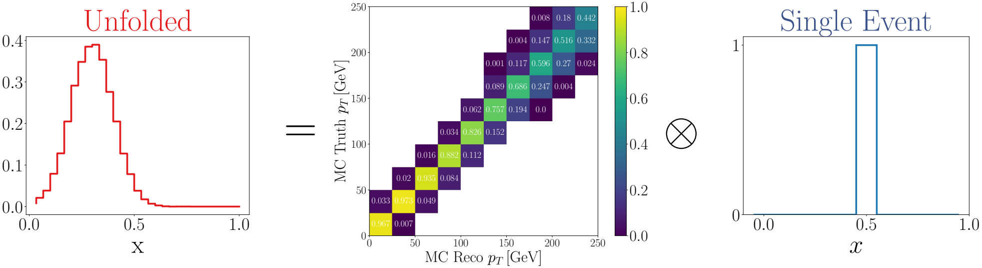
- Iterative Bayesian Unfolding calculates pseudo-inverse

$$\tilde{R}_{ji}^{(n)} = \frac{R_{ij}^{(MC)}\, t_j^{\text{Unf},(n-1)}}{\sum_k R_{ik}^{(MC)}\, t_k^{\text{Unf},(n-1)}}$$

$$t^{\text{Unf},(n)} = \tilde{R}^{(n)}\, r^{\text{Meas}}$$



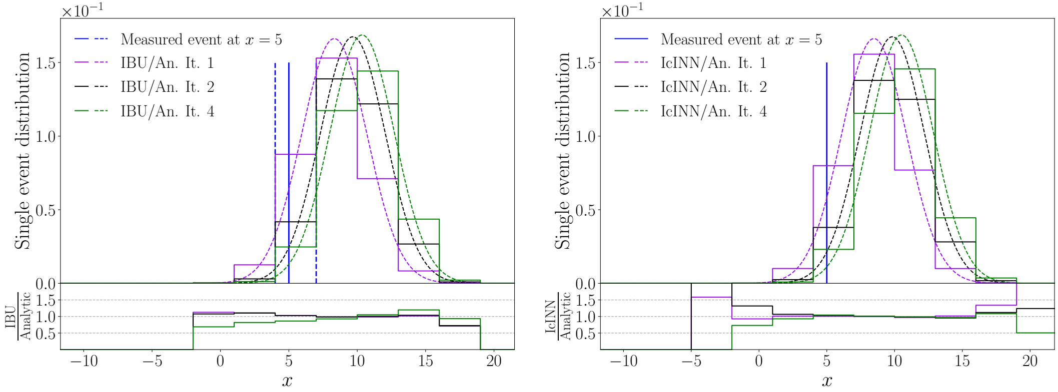
Matrix-based Single Event Unfolding



- Standard output of matrix-based unfolding is the full unfolded distribution
- New implementation allows to unfold also single events
- Also implemented for different matrix-based unfolding algorithms (Publication: [2310.17037])

$\Rightarrow$ Possibility for cross-checks with IcINN Unfolding

Matrix-based Single Event Unfolding



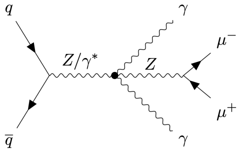
$\Rightarrow$ Per-event unfolding enables detailed comparisons: similar results for the two methods in this example

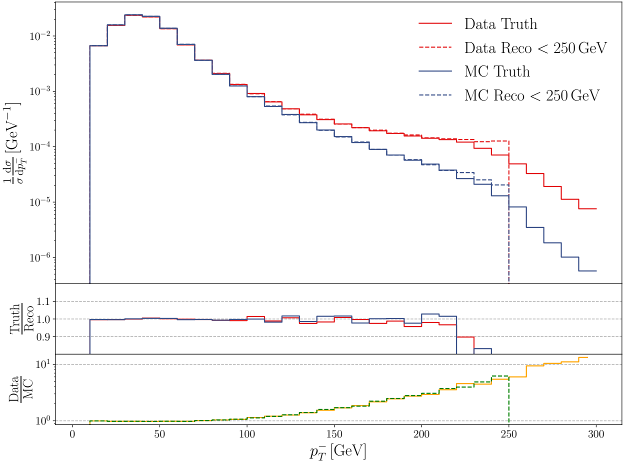
Part 4:

Unfolding an EFT Process

Unfolding an EFT Process

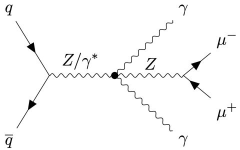
- Simulating the process
  $pp \to Z\gamma\gamma$ with $Z \to \mu^- \mu^+$
- MC $\to$ pure SM
- Data $\to$ SM + EFT contribution of
  $\mathcal{L}_{T,8} = \frac{C_{T,8}}{\Lambda^4} B_{\mu\nu} B^{\mu\nu} B_{\alpha\beta} B^{\alpha\beta}$
  with $\frac{C_{T,8}}{\Lambda^4} = \frac{2}{\mathrm{TeV}^4}$
- Applied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

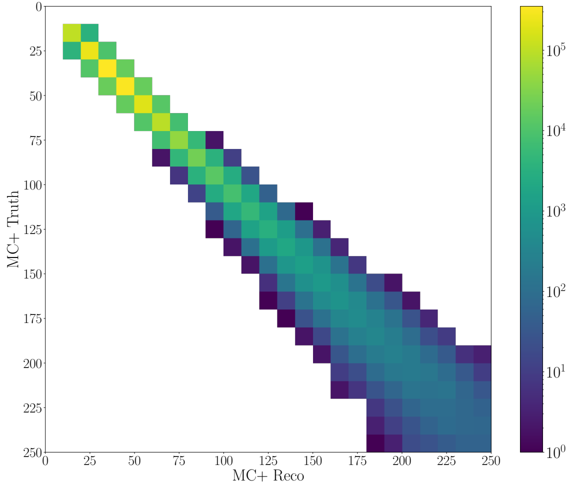




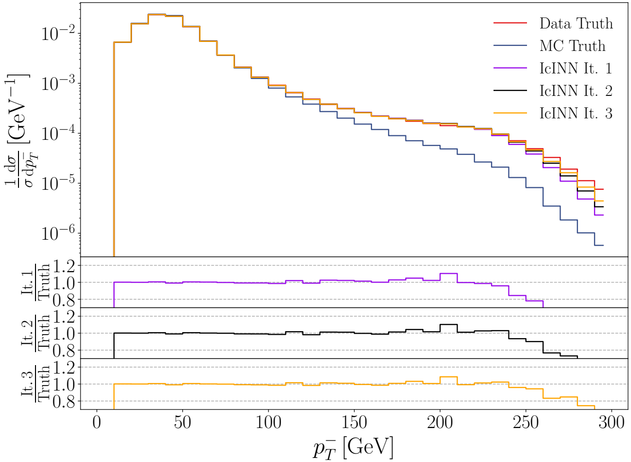
Unfolding an EFT Process

- 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 $\frac{C_{T,8}}{\Lambda^4} = \frac{2}{\mathrm{TeV}^4}$
- Applied detector smearing:
  $$\Delta p_T = p_T \cdot \sqrt{0.025^2 + p_T^2 \cdot 3.5 \cdot 10^{-8}}$$



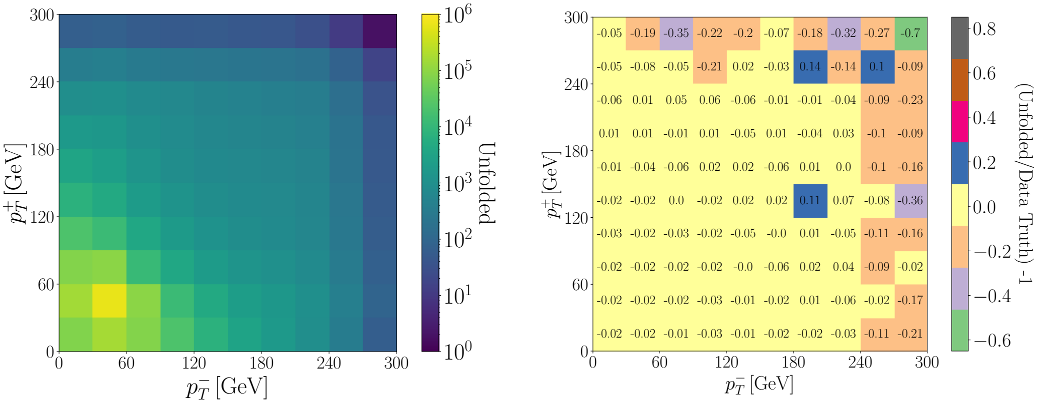


IcINN Unfolding Result



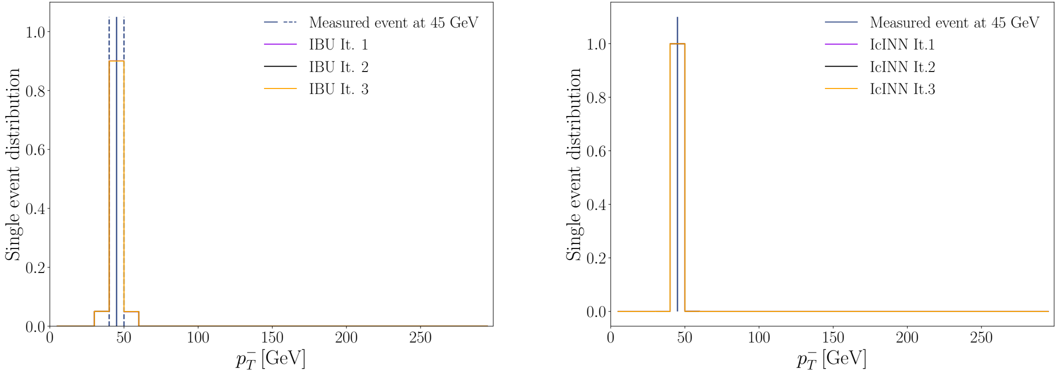
- Good convergence of the cINN unfolding already without reweighting
- Nevertheless, high-p_T region needs further iterations including reweighting
- Expected behaviour: Iterations most effective in regions with strong detector response and MC to data differences

2D IcINN Unfolding Result



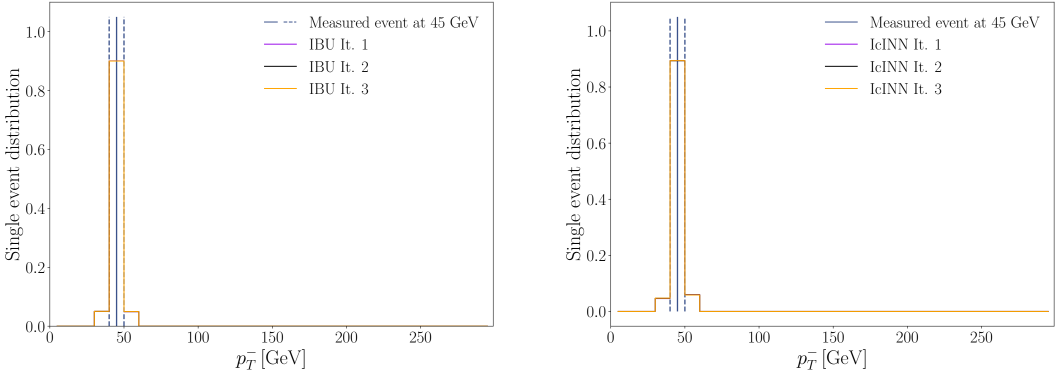
⇒ Unfolding of the p_T distributions of both muons simultaneously is possible

Single Event Unfolding



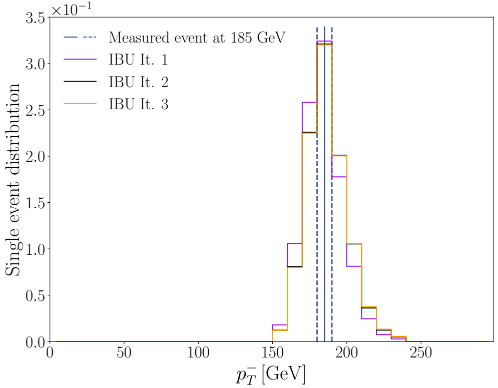
- Problem: bin size is broader than the detector resolution ⇒ Distributions not really comparable
- Solution: sample single events for IcINN uniformly over the bin width

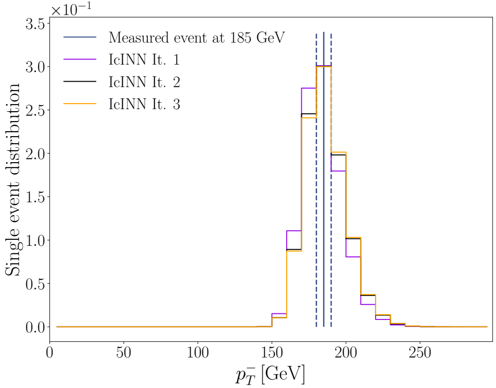
Single Event Unfolding



- Problem: bin size is broader than the detector resolution ⇒ Distributions not really comparable
- Solution: sample single events for IcINN uniformly over the bin width

Single Event Unfolding





- Similarities between single event IBU and IcINN in the overall distributions
- Same behaviour with more iterations: shift events from low-p_t to high-p_T

Conclusion / Outlook

- Implementation of an iterative cINN unfolding algorithm and application to a physical example [2212.08674]
- Introduction of a single-event matrix-based unfolding [2310.17037]
- Central Ideas:
  - Iteratively reduce bias towards Monte Carlo
  - Reweight on truth level
  - Keep probabilistic cINN unfolding
- Next steps:
  - Compare IcINN to other unfolding algorithms
  - Apply IcINN to real data

Thank you for your attention!

Additional Material

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

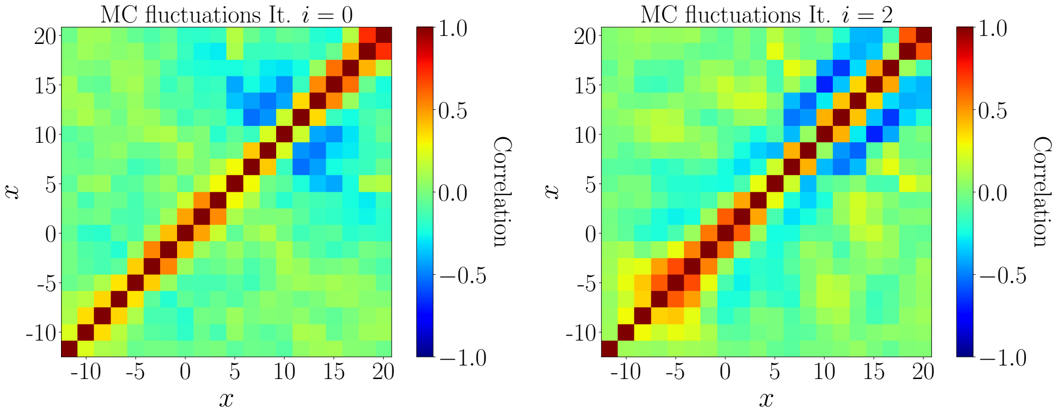
Minimize loss function:

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

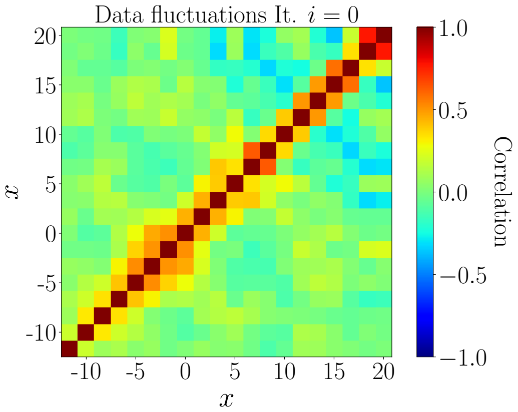
θ = cINN parameter, x = Parton Level, y = Detector level, z = Latent space variable

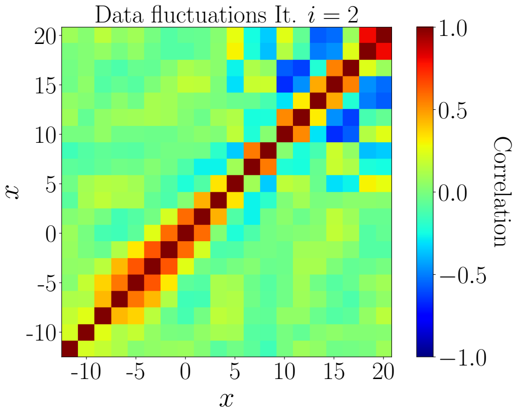
Source: arXiv [1907.02392]

Correlation Matrices

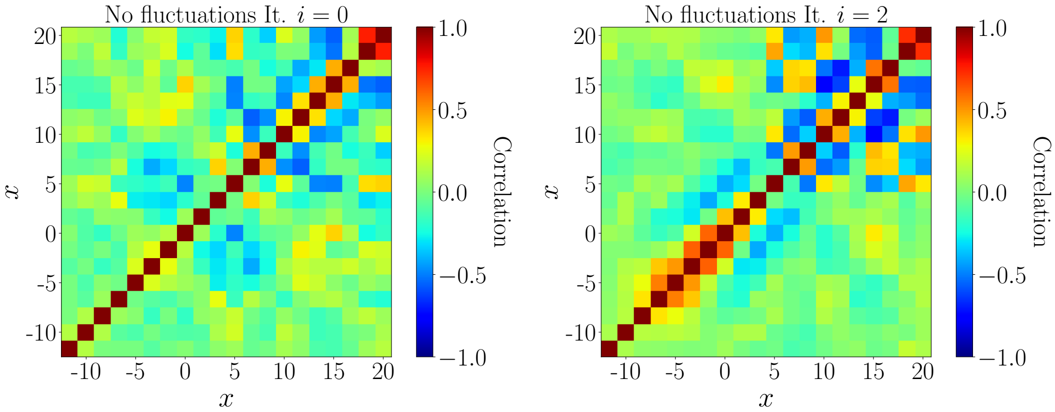


Correlation Matrices

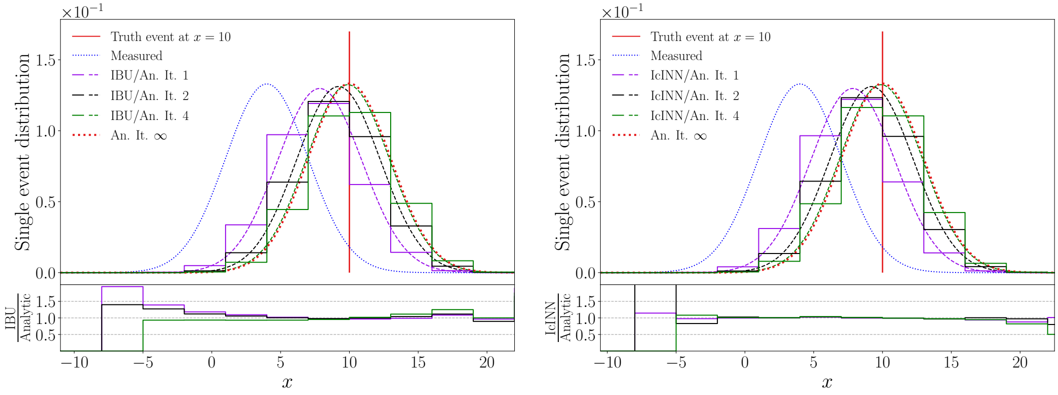




Correlation Matrices

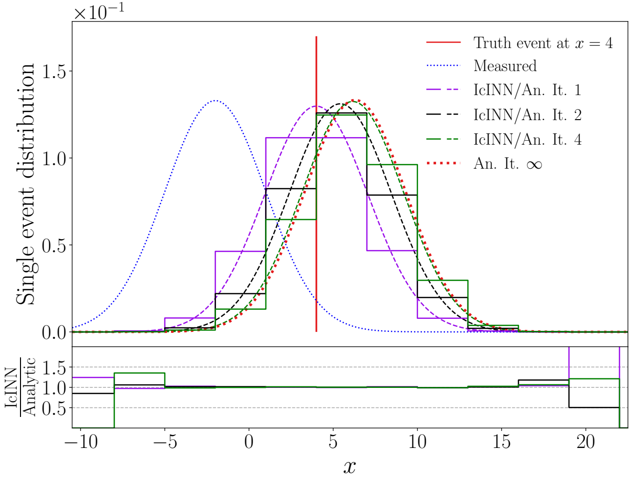


Truth to Reco and Back Again

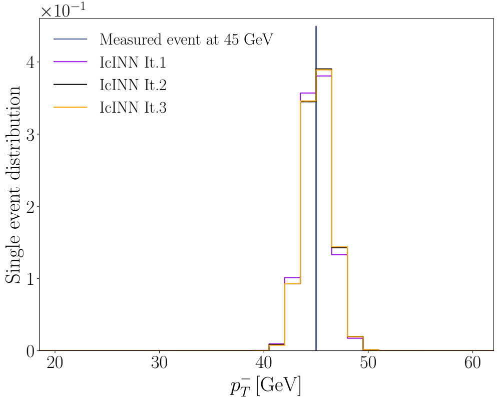


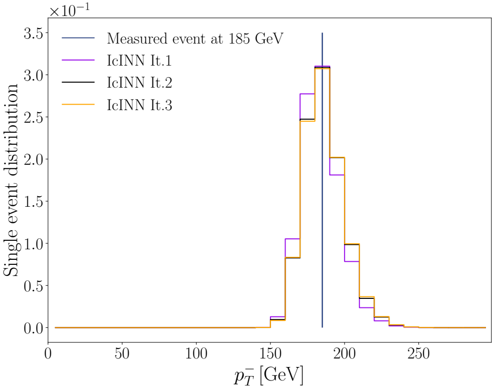
Truth to Reco and Back Again

- Analytically predictable behavior
- Unfolded distribution does not converge towards truth event
- Exact position of truth event is lost in the smearing
- Unfolded distributions are in accordance with overall unfolding result
- Careful estimation of systematic uncertainties is needed

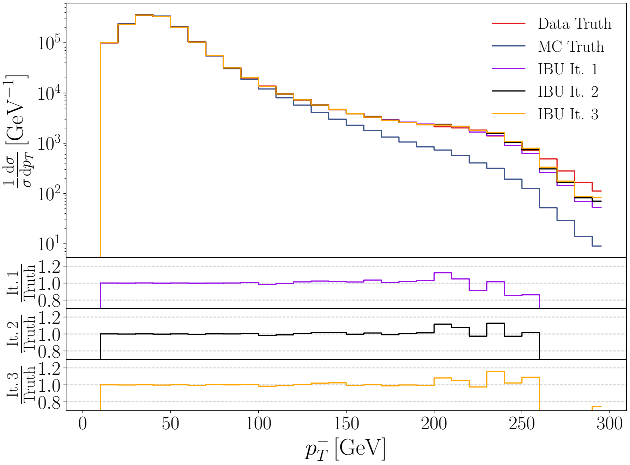


Single Event Unfolding

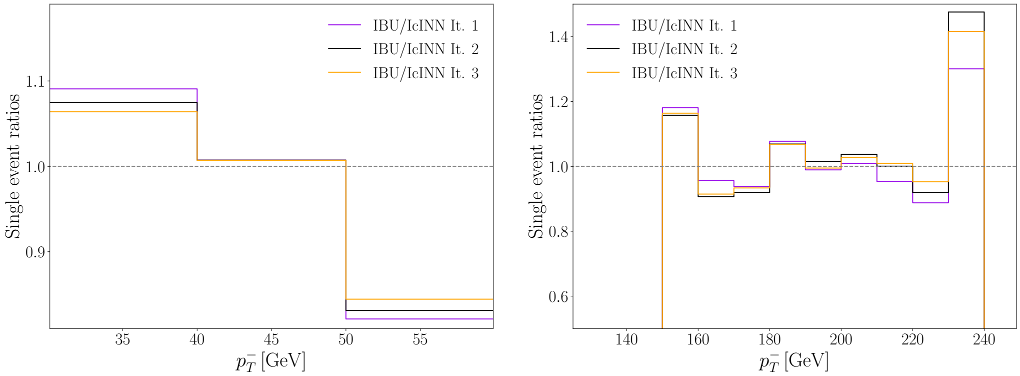




Unfolding Result IBU



Single Event Unfolding Ratios



Iterative Bayesian Unfolding (IBU)

1. Choose an initial prior $t_j^{\mathrm{Unf},(0)}$
2. Use Bayes theorem

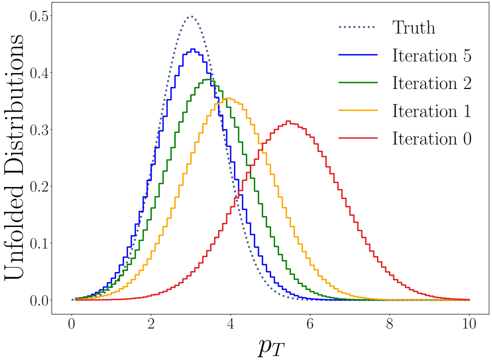
$$p(t|r) = \frac{p(r|t) \cdot p(t)}{\int p(r|t) \cdot p(t)\,\mathrm{d}t},$$

to calculate the pseudo-inverse

$$\tilde{R}_{ji}^{(n)} = \frac{R_{ij}^{(MC)}\, t_j^{\mathrm{Unf},(n-1)}}{\sum_k R_{ik}^{(MC)}\, t_k^{\mathrm{Unf},(n-1)}}$$

3. Update the prior

$$t_j^{\mathrm{Unf},(n)} = \sum_l \tilde{R}_{jl}^{(n)}\, r_l^{\mathrm{Meas}}$$



⇒ Balance between Bias and Uncertainties

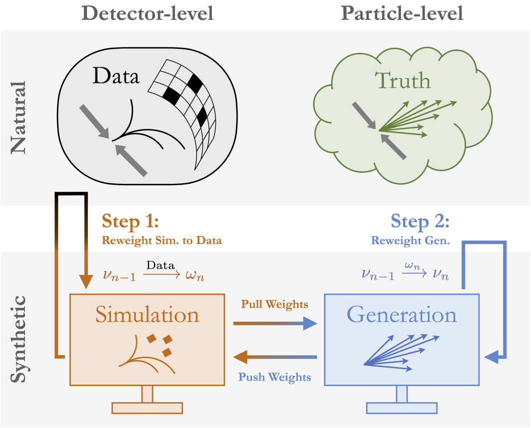
Classical Unfolding Algorithms

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.

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]