Two Invertible Networks for the Matrix Element Method

Theo Heimel
October 2022

Institut für theoretische Physik
Universität Heidelberg

arXiv:2210.00019
Butter, Heimel, Martini, Peitzsch, Plehn

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Introduction

How can we find new physics at the LHC?
Maybe it is hidden in rare processes

$\Downarrow$

Need better analysis techniques!

Traditional analysis

- Hand-crafted observables
- Binned data

$\Downarrow$

Only fraction of information used

Matrix element method

- Based on first principles
- Estimates uncertainties reliably
- Optimal use of information

$\Downarrow$

Perfect for processes with few events

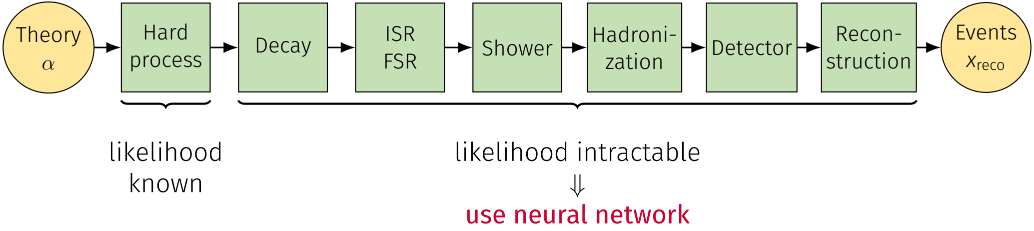
Two Invertible Networks for the Matrix Element Method

Matrix Element Method

- Process with theory parameter α, hard-scattering momenta x_{hard}
- Likelihood at hard-scattering level given by differential cross section

$$p(x_{\text{hard}}|\alpha) = \frac{1}{\sigma(\alpha)} \frac{\mathrm{d}\sigma(\alpha)}{\mathrm{d}x_{\text{hard}}}$$

- Neyman-Pearson lemma $\implies$ optimal use of information
- Differential cross section only known analytically at hard-scattering level



MEM at reconstruction level

- Integrate out hard-scattering phase space

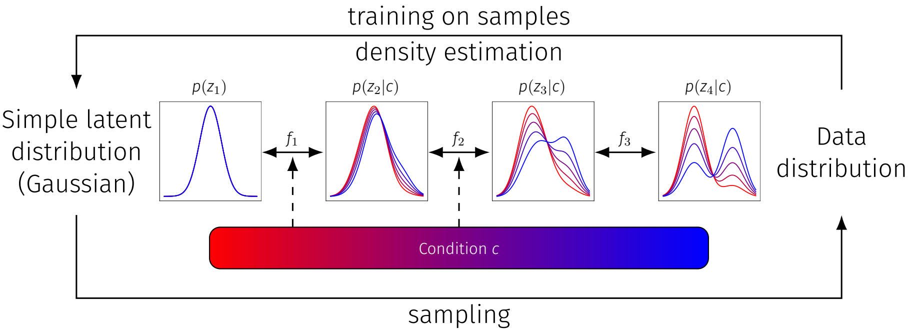
$$p(x_{\text{reco}}|\alpha) = \int \mathrm{d}x_{\text{hard}} \underbrace{p(x_{\text{hard}}|\alpha)}_{\text{diff. CS}} \underbrace{p(x_{\text{reco}}|x_{\text{hard}}, \alpha)}_{\text{estimate with network}}$$

- Need to learn probability distribution $p(x_{\text{reco}}|x_{\text{hard}}, \alpha)$
  In practice: ignore α-dependence and learn $p(x_{\text{reco}}|x_{\text{hard}})$
- Not known analytically $\rightarrow$ learn from data

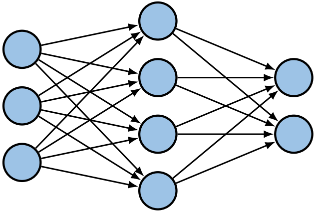
Solution:
normalizing flow $\rightarrow$ Transfer-cINN

- Conditional Invertible Neural Networks (cINN): [Ardizzone et al., 1907.02392]
  chain of learnable, invertible transformations with tractable Jacobian
- Distributions linked through change of variable formula

$$p(z_n) = p(z_1) \det \frac{\partial z_1(z_n; c)}{\partial z_n}$$

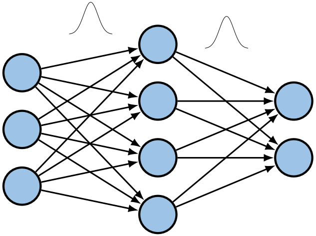


Flows with uncertainties



deterministic weights w_i

$\downarrow$



weights $w_i \sim \mathcal{N}(\mu_i, \sigma_i)$

- Quantify training uncertainty with **Bayesian Invertible Neural Networks** (BINN)

  [MacCay, 1995] [Neal, 2012] [Bellagente et al., 2104.04543]
- Simple modification of deterministic network:
  $\rightarrow$ Replace deterministic weights with distribution
  $\rightarrow$ Additional term in loss function
- Extracting uncertainties:
  sample from weight distribution
- Use as generator $\rightarrow$ Histograms with error bars
- Use as density estimator $\rightarrow$ Error on density

How to compute the integral?

- $|\mathcal{M}|^2$ spans several orders of magnitude
- Narrow distribution from Transfer-cINN

} **Integration challenging**

- Importance sampling with proposal distribution $q(x_\text{hard})$

$$p(x_\text{reco}|\alpha) = \left\langle \frac{1}{q(x_\text{hard})} \, p(x_\text{hard}|\alpha) \, p(x_\text{reco}|x_\text{hard}, \alpha) \right\rangle_{x_\text{hard} \sim q(x_\text{hard})}$$

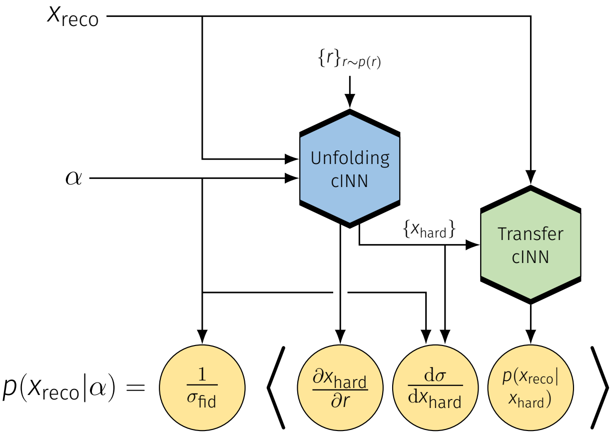
- Bayes' theorem: Integration becomes trivial if

$$x_\text{hard} \sim q(x_\text{hard}) = p(x_\text{hard}|x_\text{reco}, \alpha)$$

Solution:
normalizing flow $\rightarrow$ **Unfolding-cINN**

[Bellagente et al., 2006.06685]

Putting it all together



- Training data

$$(\alpha, x_{\text{hard}}, x_{\text{reco}})$$

- Transfer-cINN learns

$$p(x_{\text{reco}}|x_{\text{hard}})$$

$\rightarrow$ transfer function
$\rightarrow$ fast forward simulation

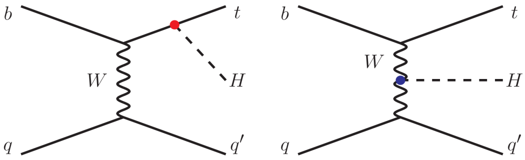
- Unfolding-cINN learns

$$p(x_{\text{hard}}|x_{\text{reco}}, \alpha)$$

$\rightarrow$ phase space sampling

Two Invertible Networks for the Matrix Element Method

LHC process



- Single Higgs production with anomalous non-CP-conserving Higgs coupling

$$\mathcal{L}_{t\bar{t}H} = -\frac{y_t}{\sqrt{2}} \left[\cos\alpha\, \bar{t}t + \frac{2}{3} \mathrm{i} \sin\alpha\, \bar{t}\gamma_5 t \right] H \qquad \text{with CP-angle } \alpha$$

[Artoisenet et al, 1306.6464] [de Aquino, Mawatari, 1307.5607] [Demartin et al, 1504.00611]

- Decays $tHj \to (bW)\ (\gamma\gamma)\ j$. Test on different datasets

Leptonic decay
$(b\mu^+\nu_\mu)\ (\gamma\gamma)\ j$
Challenge:
reconstruct ν

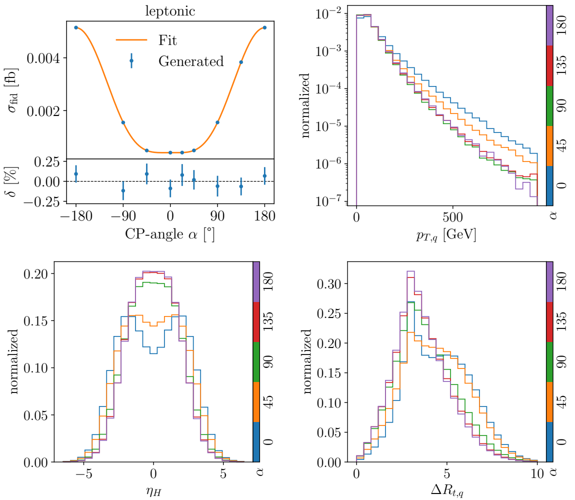
→

Hadronic decay
$(bjj)\ (\gamma\gamma)\ j$
Challenge:
jet combinatorics

→

Hadronic decay + ISR
$(bjj)\ (\gamma\gamma)\ j +$ QCD jets
Challenge:
no clear combinatorics

Why we need the MEM



Around the SM, $\alpha = 0°$:

low total cross section (few events)

$+$

low variation of rate

$+$

kinematic observables still sensitive

$\Downarrow$

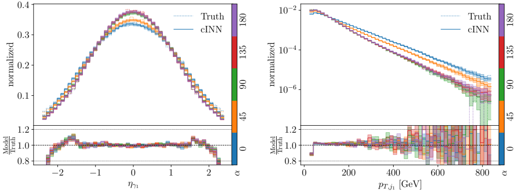
need kinematic observables
to use all available information

$\Downarrow$

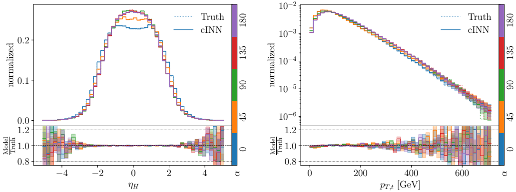
ideal use case for MEM

Two Invertible Networks for the Matrix Element Method

Testing the cINNs



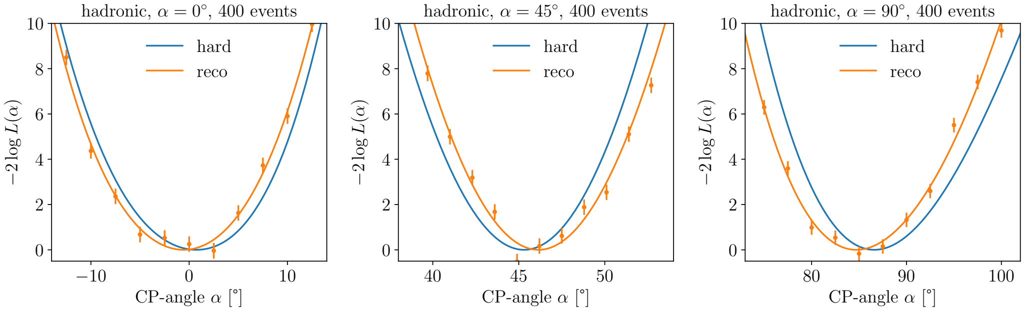
Transfer-cINN (reco level plots)



Unfolding-cINN (hard-scattering level plots)

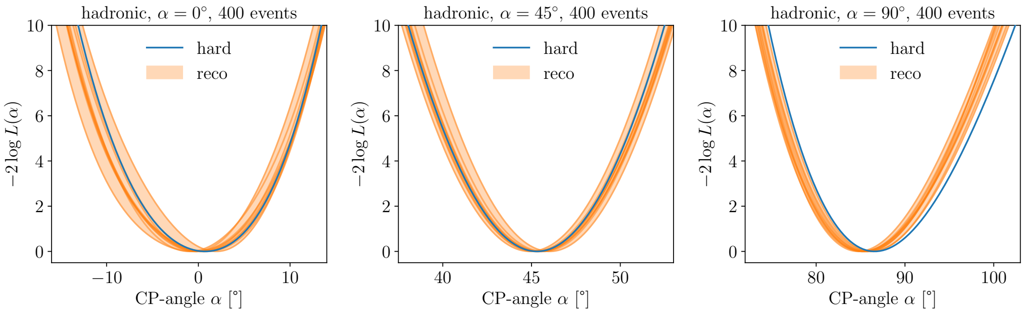
- Check performance on test dataset
  $\rightarrow$ Transfer-cINN as forward simulator
  $\rightarrow$ Unfolding-cINN: once for each event
- Good agreement with Truth
- Error bars from Bayesian network
  $\rightarrow$ Within BINN errors in bulk
- deterministic Unfolding-cINN used for integration

Likelihoods for hadronic decay without ISR



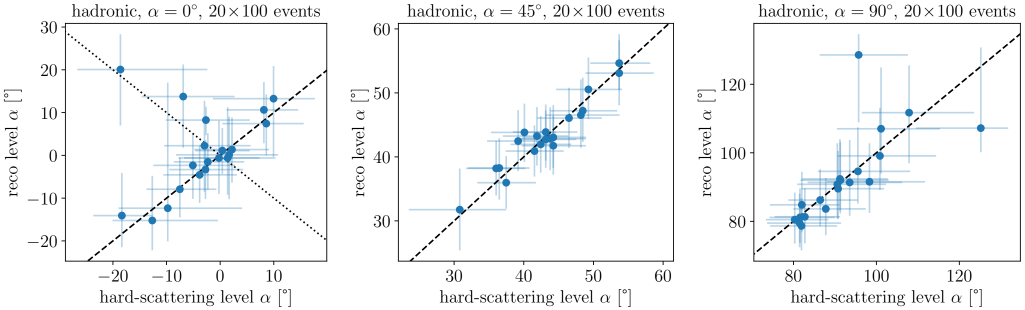
- Deterministic network, $\alpha = 0°, 45°, 90°$, 400 events each
- Extract likelihood for different α, sum events, fit polynomial (orange line)
- Compare to likelihood from hard-scattering data (blue line)
- **Good agreement between hard-scattering and reco-level**
  $\rightarrow$ But how large is the systematic uncertainty from training?

Training uncertainty from BINN



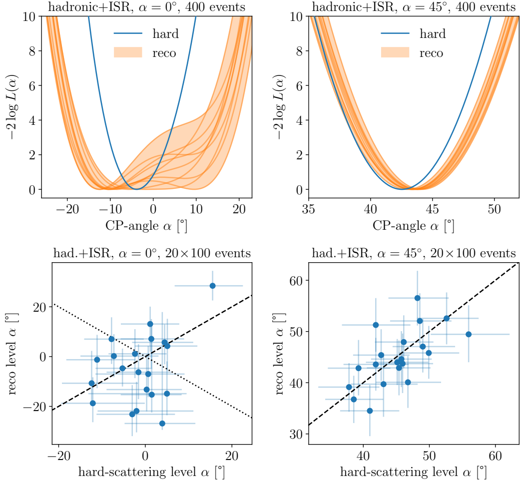
- Extract likelihood for 10 sampled networks
  $\rightarrow$ estimate of systematic error from training
- Only uncertainty from finite training data
  $\rightarrow$ lack of expressivity not captured

Calibration checks



- Minimum and 68% confidence intervals for 20×100 events
- Good correlation betwen reco- and hard-scattering level
- Slight bias can be removed by calibration
- Lagrangian almost symmetric around $\alpha = 0°$
  $\rightarrow$ sometimes wrong sign

Hadronic decay with ISR



- Final state $(bjj)\ (\gamma\gamma)\ j$
  + additional jets from ISR and FSR
- Can't resolve between relevant jets and ISR jets during reconstruction
  $\rightarrow$ combinatorics more difficult
- Loss of sensitivity around $\alpha = 0°$
- Worse calibration, more bias
- Increased systematic uncertainty captured by Bayesian network

Two Invertible Networks for the Matrix Element Method

Outlook

- Measure fundamental Lagrangian parameters from small numbers of events
- Transfer-cINN: encode QCD and detector effects
- Unfolding-cINN: efficient integration over hard-scattering phase space
- Without ISR: close to hard-scattering truth
- With ISR: worse performance from more challenging combinatorics
- Promising for extracting maximal information from small event numbers