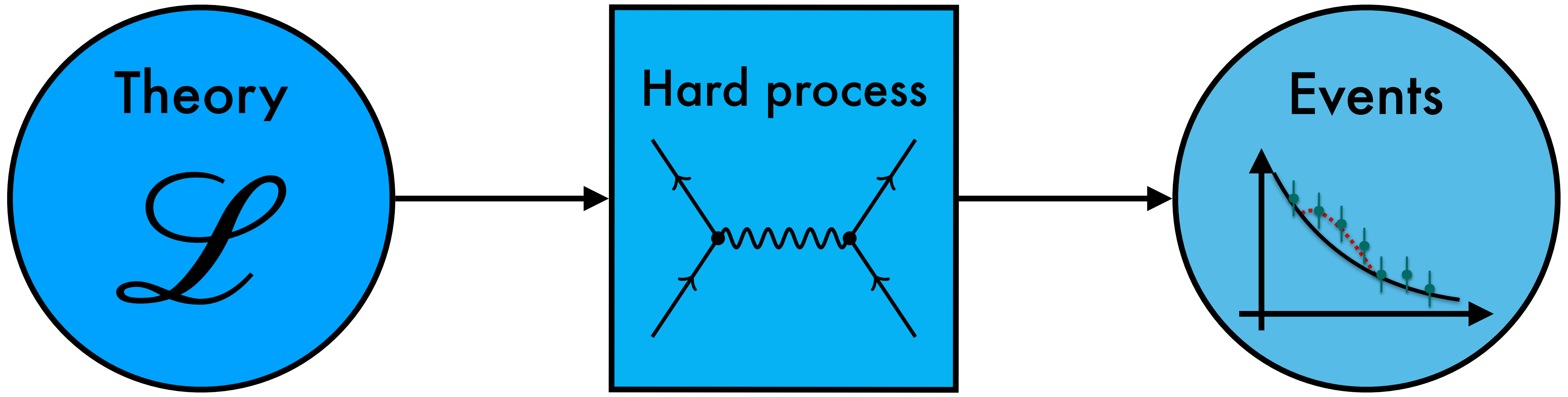
Precision Maching Learning for the Matrix Element Method

Theo Heimel
October 2023

Institut für theoretische Physik
Universität Heidelberg

[2310.07752] TH, Huetsch, Winterhalder, Plehn, Butter

Introduction



Theory parameter
α

known from theory

Hard-scattering momenta
x_{hard}

Likelihood from differential cross section

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

Classical analysis

- hand-crafted observables
- binned data

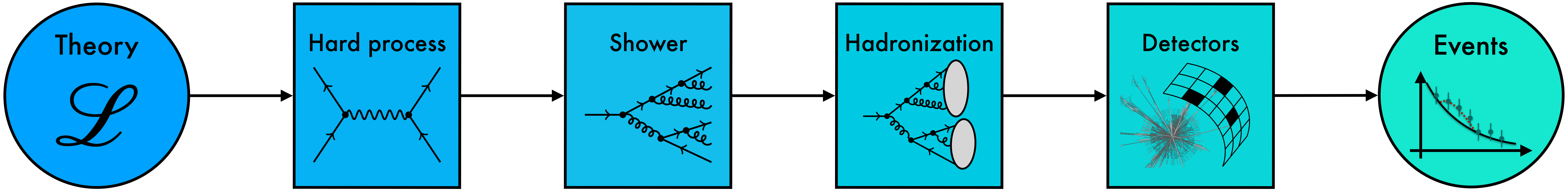
→ not all available information is used

Matrix Element Method (MEM)

- based on first principles
- estimates uncertainties reliably
- optimal use of information

→ perfect for processes with few events

Introduction



Theory parameter α

known from theory

likelihood intractable
→ learn with neural network

Reconstructed momenta x_{reco}

$$p(x_{\text{reco}} \mid \alpha) = \int \mathrm{d}x_{\text{hard}}\; p(x_{\text{hard}} \mid \alpha)\; p(x_{\text{reco}} \mid x_{\text{hard}})\; \epsilon(x_{\text{hard}})$$

Factorize problem → integrate out intermediate momenta

ML+MEM integral

$$p(x_{\text{reco}} \mid \alpha) = \int \mathrm{d}x_{\text{hard}} \;\; p(x_{\text{hard}} \mid \alpha) \;\; p(x_{\text{reco}} \mid x_{\text{hard}}) \;\; \epsilon(x_{\text{hard}})$$

ML+MEM integral

$$p(x_{\text{reco}} \mid \alpha) = \int \mathrm{d}x_{\text{hard}} \;\; p(x_{\text{hard}} \mid \alpha) \;\; p(x_{\text{reco}} \mid x_{\text{hard}}) \;\; \epsilon(x_{\text{hard}})$$

Efficient MC integration

importance sampling
with Normalizing Flow

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

ML+MEM integral

$$p(x_{\mathrm{reco}} \mid \alpha) = \int \mathrm{d}x_{\mathrm{hard}} \; p(x_{\mathrm{hard}} \mid \alpha) \; p(x_{\mathrm{reco}} \mid x_{\mathrm{hard}}) \;\; \epsilon(x_{\mathrm{hard}})$$

Efficient MC integration

importance sampling
with Normalizing Flow

$$x_{\mathrm{hard}} \sim p(x_{\mathrm{hard}} \mid x_{\mathrm{reco}}, \alpha)$$

Theory knowledge

differential
cross-section

$$\frac{1}{\sigma(\alpha)} \frac{\mathrm{d}\sigma(\alpha)}{\mathrm{d}x_{\mathrm{hard}}}$$

ML+MEM integral

$$p(x_{\text{reco}} \mid \alpha) = \int \mathrm{d}x_{\text{hard}} \; p(x_{\text{hard}} \mid \alpha) \; p(x_{\text{reco}} \mid x_{\text{hard}}) \; \epsilon(x_{\text{hard}})$$

Efficient MC integration

importance sampling with Normalizing Flow

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

Theory knowledge

differential cross-section

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

Transfer function

Density estimation: Normalizing Flow

Solve combinatorics: Transformer

ML+MEM integral

$$p(x_{\text{reco}} \mid \alpha) = \int \mathrm{d}x_{\text{hard}} \; p(x_{\text{hard}} \mid \alpha) \; p(x_{\text{reco}} \mid x_{\text{hard}}) \; \epsilon(x_{\text{hard}})$$

Efficient MC integration

importance sampling
with Normalizing Flow

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

Theory knowledge

differential
cross-section

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

Transfer function

Density estimation:
Normalizing Flow

Solve combinatorics:
Transformer

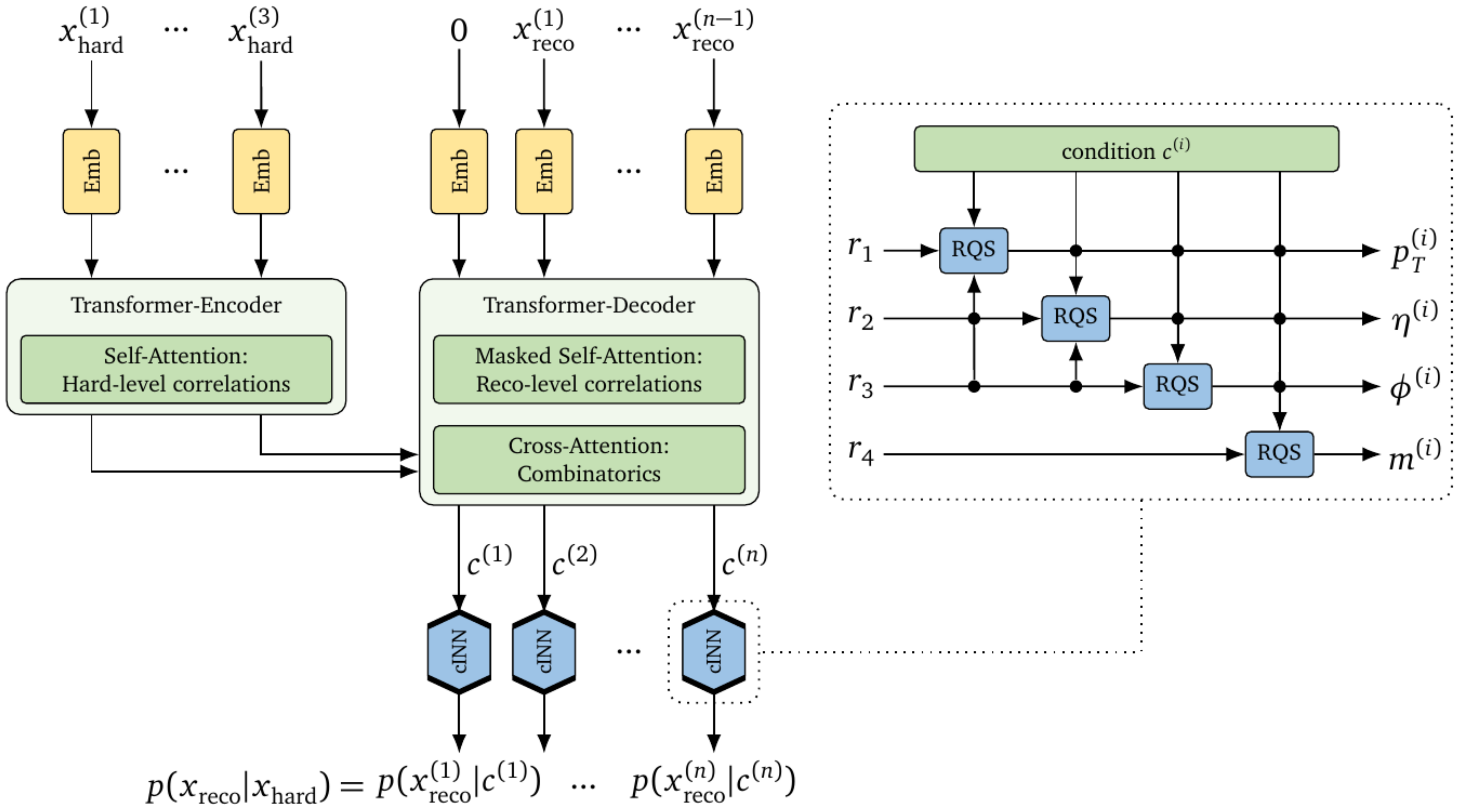
Acceptance function

learn with simple
classifier network

Learning the transfer function

Transformer
correlations between particles & combinatorics

Normalizing Flow
likelihood for individual particles



[Butter et al, 2305.10475]
[Finke et al, 2303.07364]

Transfer function + Transformer = Transfermer

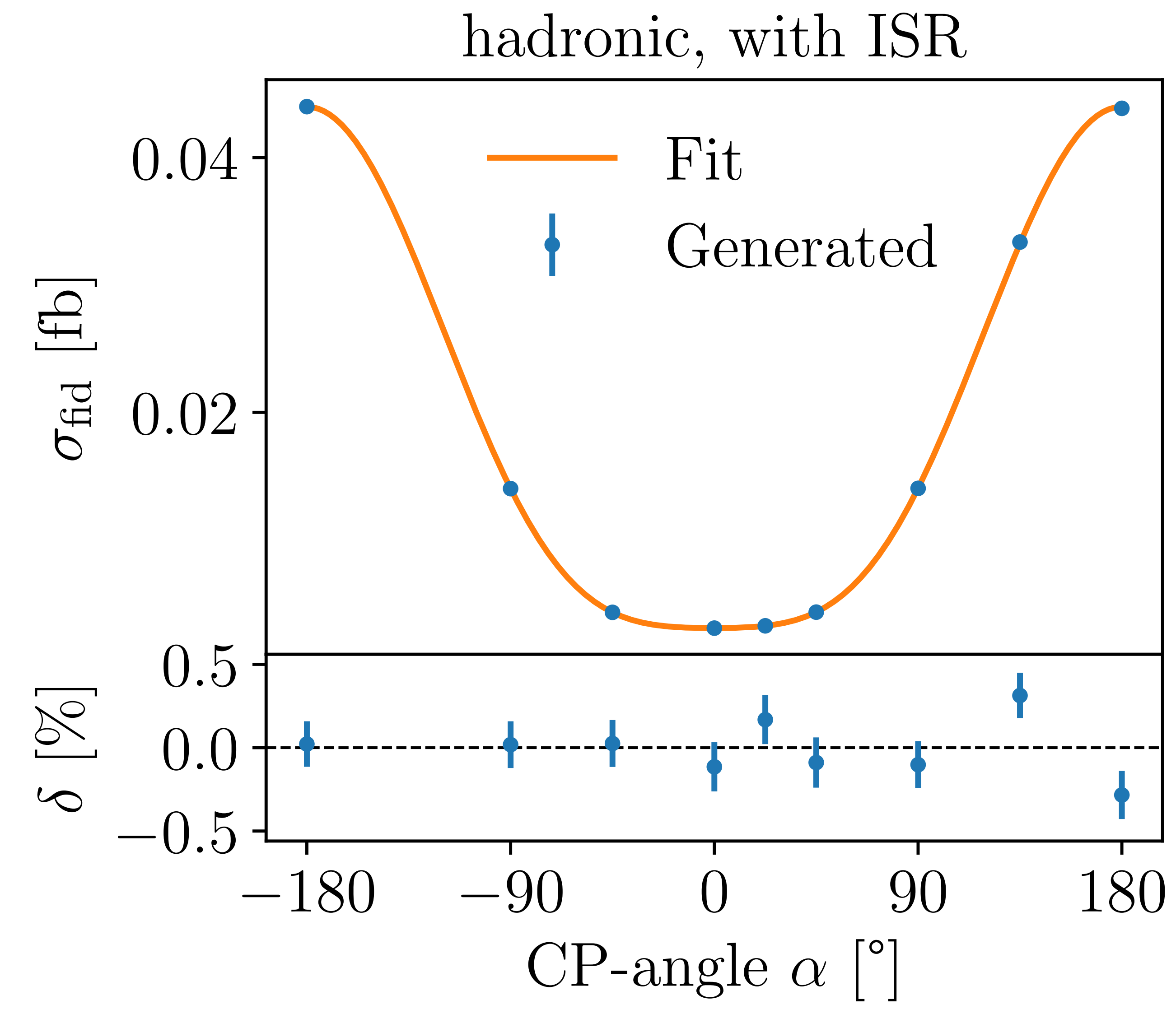
LHC example

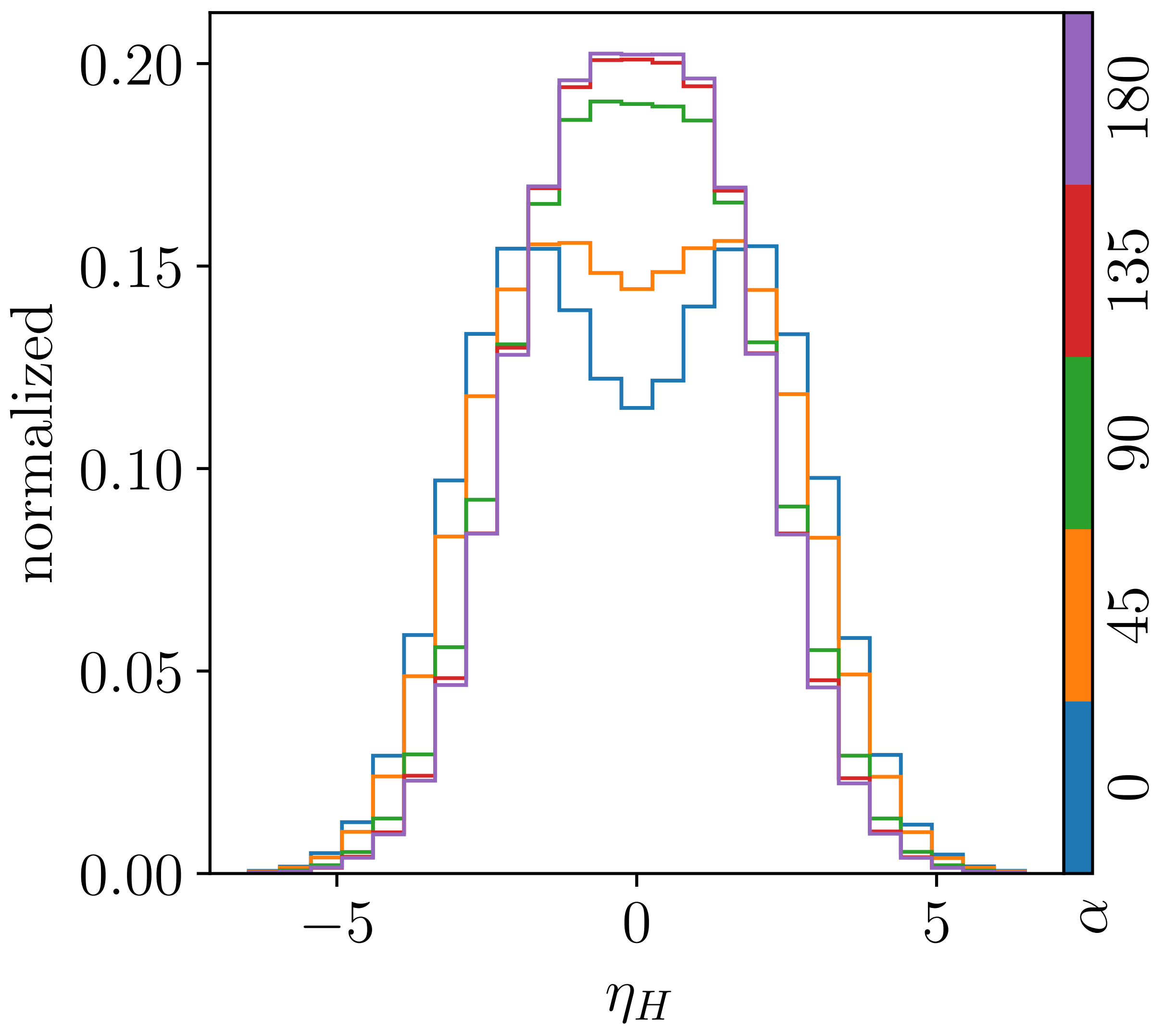
Single Higgs production with anomalous non-CP-conserving Higgs coupling

$$\mathscr{L}_{t\bar{t}H} = -\frac{y_t}{\sqrt{2}}\left[\cos\alpha\ \bar{t}t + \frac{2}{3}\mathrm{i}\sin\alpha\ \bar{t}\gamma_5 t\right]H$$

Hadronic decay of top + ISR:

$\mathrm{tHq} \to (\mathrm{bjj})\ (\gamma\gamma)\ \mathrm{j}$ + QCD jets





Around the SM, CP-phase $\alpha = 0°$:

low total cross section (few events)

+

low variation of rate

+

kinematic observables still sensitive

↓

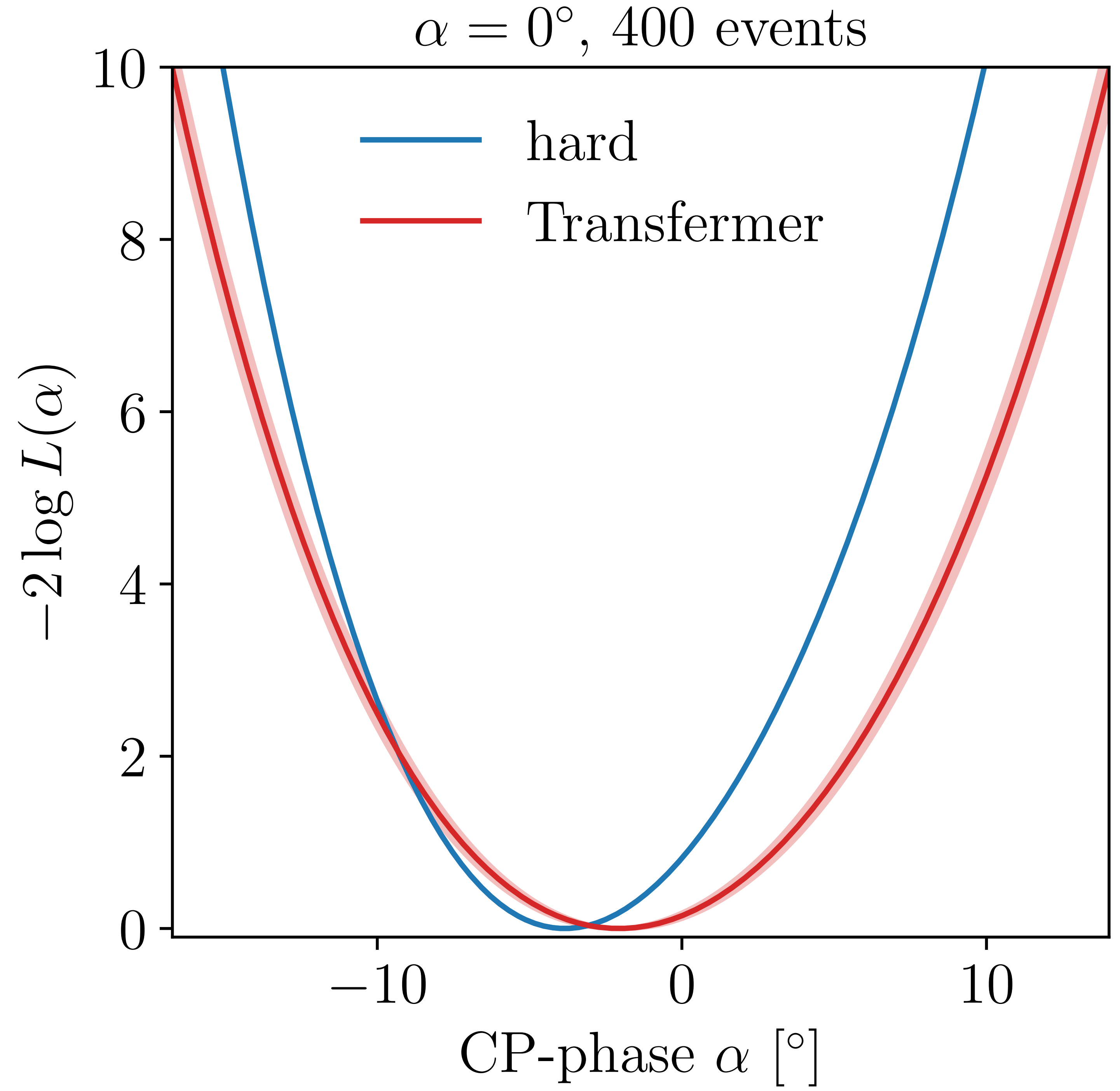
need kinematic observables to use all available information

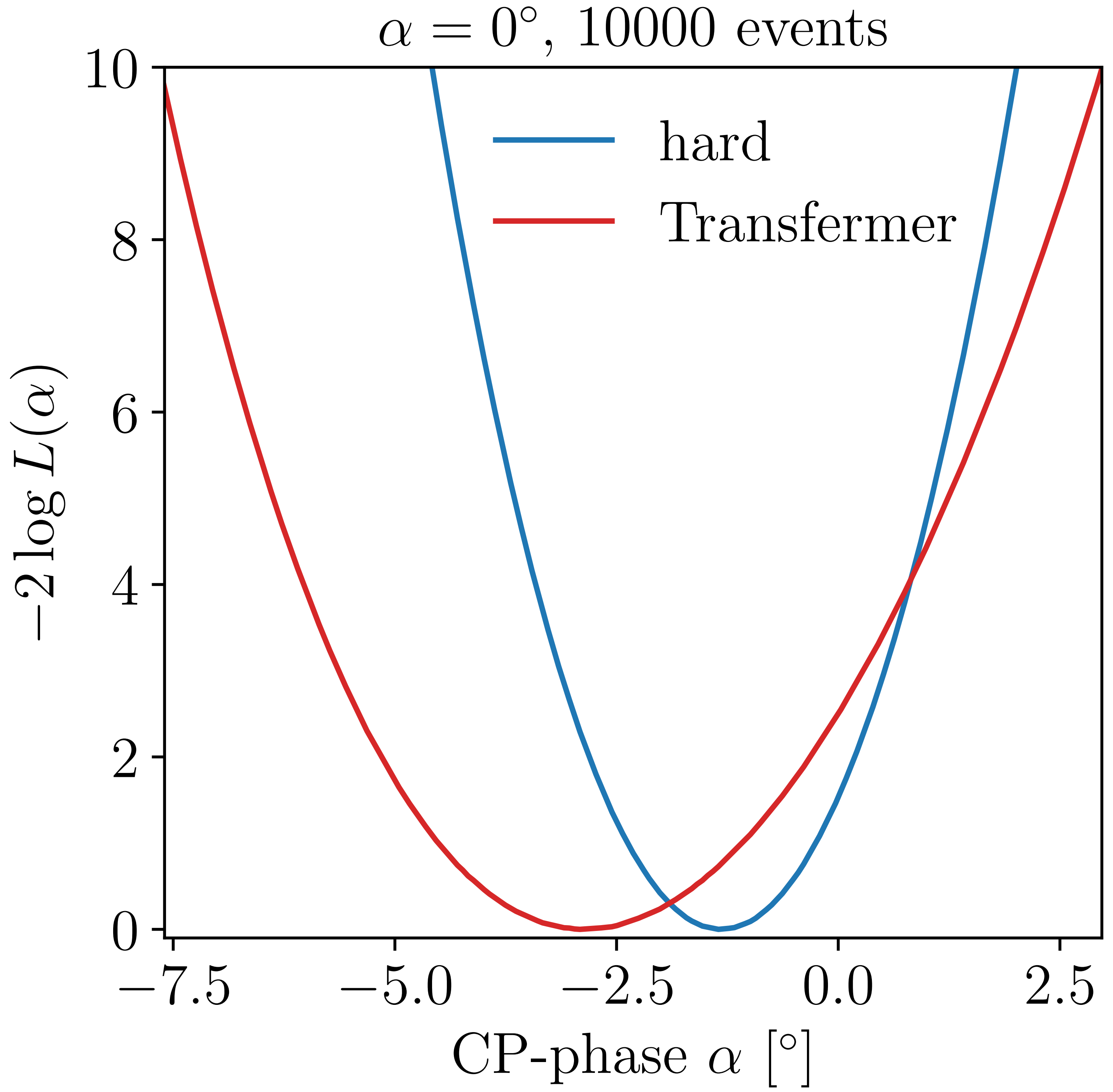
↓

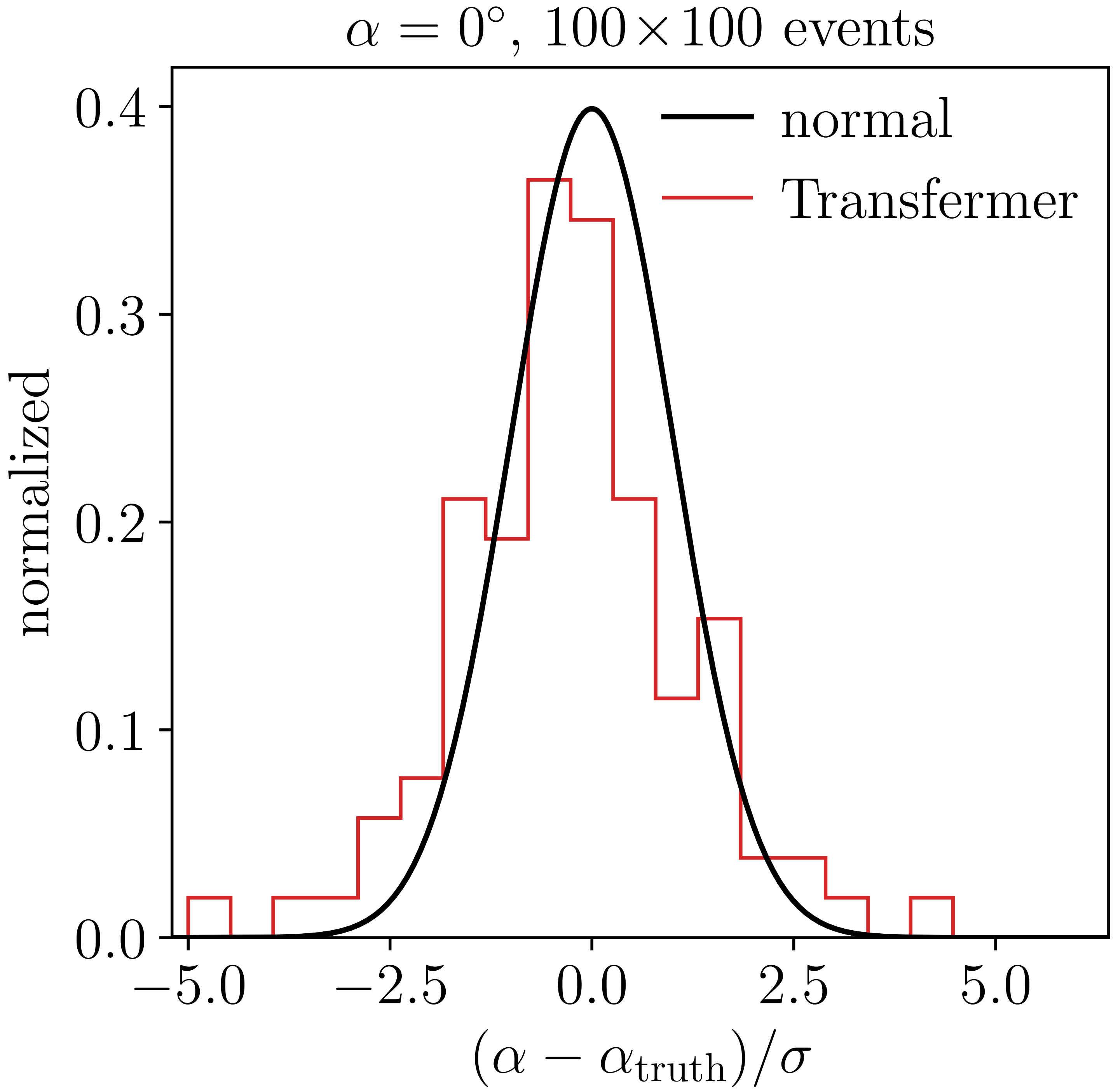
ideal use case for MEM

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

Inference results







→ smooth and well-calibrated likelihoods, both for low and high event count
→ close to optimal information
→ Uncertainty bands: MC integration error &
syst. error from limited training statistics (Bayesian NN)

Outlook

Comparison to other simulation-based inference methods

⊕ Matrix element information used during inference
→ does not need to be learned by network

⊕ Factorization of ME, transfer function, acceptance
→ less “black box”, more control of uncertainties

⊙ Computes phase space integral for every event
→ can be made fast, efficient & stable (see paper)

Outlook

→ extend to NLO QCD
→ discuss analysis applications with experimentalists