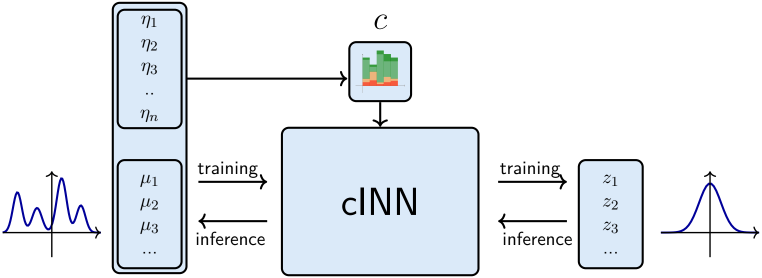


A Method for Inferring Signal Strength Modifiers by Conditional Invertible Neural Networks

CHEP 2023, Norfolk, Virginia

Máté Zoltán Farkas, Svenja Diekmann, Niclas Eich, Martin Erdmann

for the CMS Collaboration

9th May, 2023

SPONSORED BY THE
Federal Ministry
of Education
and Research

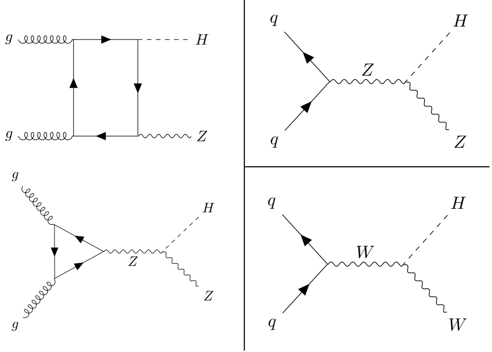
RWTH AACHEN UNIVERSITY

Table of Contents

Mate Zoltan Farkas

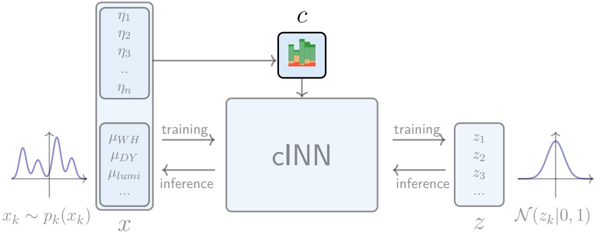
- Why conditional Invertible Neural Networks (cINNs)[1]?
  - Many parameter fits: time-consuming
    → Posterior inference with cINNs is time-efficient
  - cINN model preserve gradients
    → Applicable in differentiable workflows
- VH-Analysis at CMS:
  - 3 Signal processes:
    - $gg \to ZH$
    - $qq \to ZH$
    - $qq \to WH$
  - 13 Background processes:
    - Drell-Yan Process
    - $t\bar{t}$ production
    - Vector boson fusion Higgs production...
- Goal: infer the signal strength modifier parameters $\mu = \sigma/\sigma_{SM}$

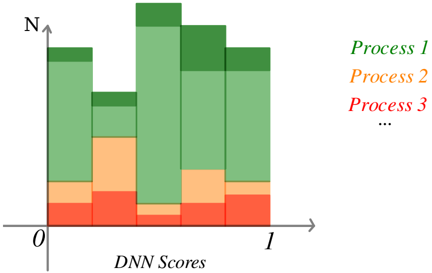
Gluon Fusion	Quark Initiated



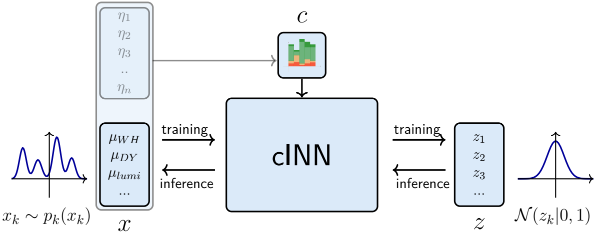
[1] 2003.06281

- Analysis performed at CMS
- Based on simulated MC samples
- Analysis workflow:
  1. Simulated events
  2. Final state object selection
  3. DNN process categorization
  4. Fit
- Maximum likelihood fit:
  - Histogrammed DNN scores
    $\rightarrow$ Used as condition c for the cINN





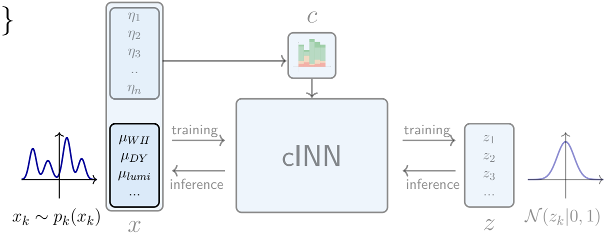
- cINN architecture:
  - Based on normalizing flows
    $\rightarrow$ invertibility and differentiability
  - Alternating network blocks:
    - Affine coupling blocks
    - Permutation blocks



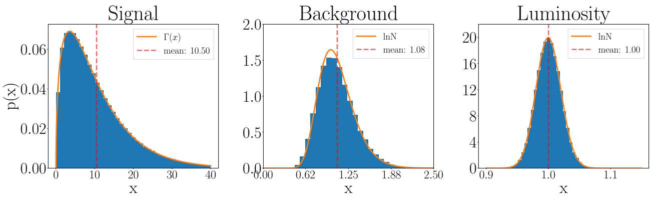
- Forward direction: training
  - Approximate the unknown true posteriors $p(x|c)$
  - Map inputs to a $\mathcal{N}(z|0,1)$

- Backward direction: inference
  - Sampling from $\mathcal{N}(z|0,1)$ and inversion
    $\rightarrow$ Posterior samples
    $\rightarrow$ Predictions

- Goal: infer signal modifier parameters $\{\mu_i\}$
  - Dataset contains expected $\{\mu_i\}$ and nuisance parameter effects
  - Scale processes with their μ_i
  - Model uncertainties

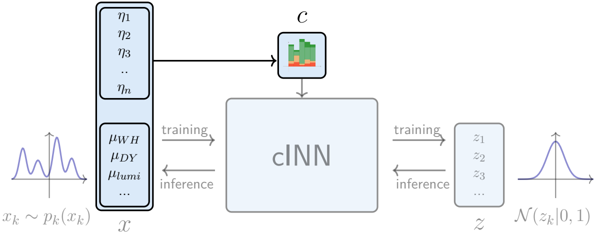


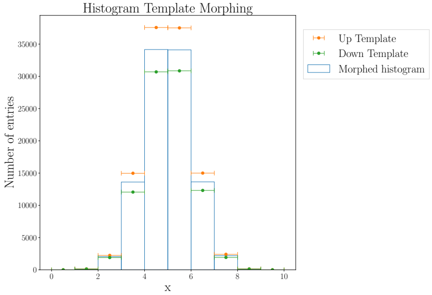
- μ Priors:
  - Signal: Γ-distribution
    - Finer sampling around expected μ
    - Sensitivity for $\mu \gtrsim 10$
  - Background:
    - Lognormal $\langle x \rangle = 1 \pm 27\%$
  - Luminosity:
    - Lognormal $\langle x \rangle = 1 \pm 2\%$
    - Affects all processes equally



Mate Zoltan Farkas

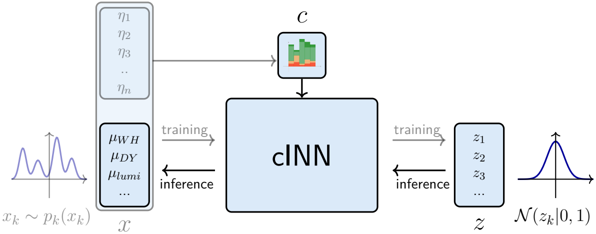
- Statistical uncertainties
  - Expected measurement uncertainty
  - MC sample size
  - Both follow a Poisson distribution
- Systematic effects:
  - Normalizing uncertainties
    $\rightarrow$ Process scaling
  - Shape-changing uncertainties
    $\rightarrow$ Histogram template morphing

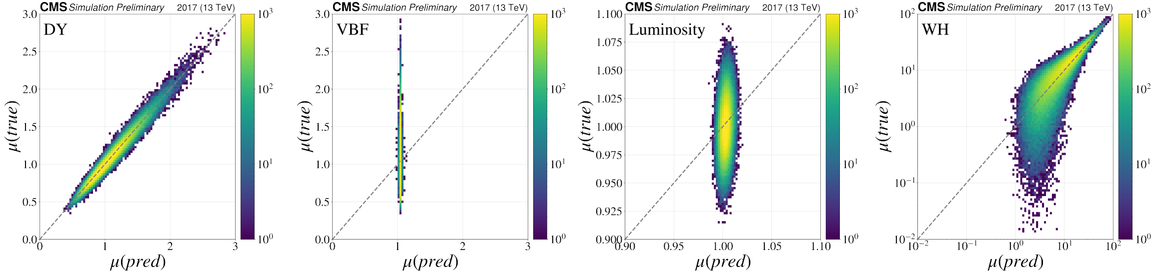




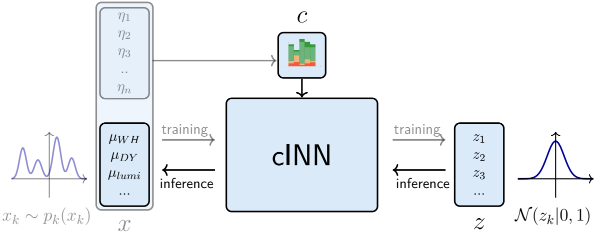
Signal Modifier Parameter Inference – Predictions

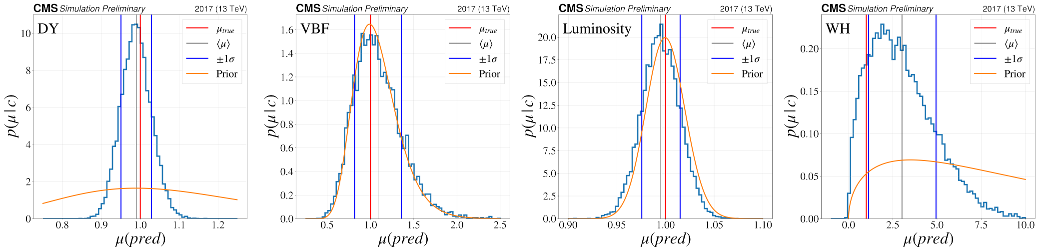
- 3 parameters groups per sensitivity:
  - well-reconstructed parameters (μ_{DY} ...)
  - unrecognized parameters (μ_{VBF} ...)
  - weakly-recognized parameters (μ_{lumi}...)
- Signal: sensitivity threshold for small μ



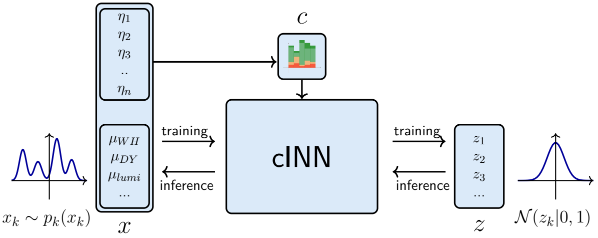


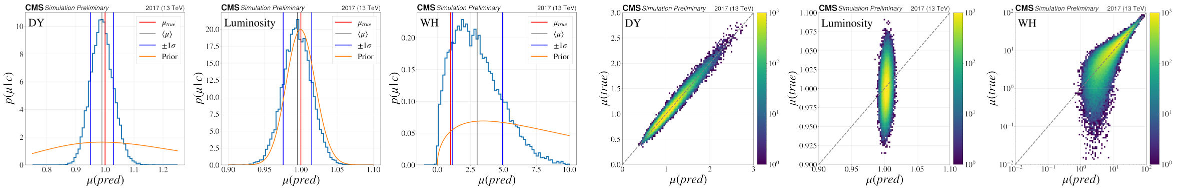
- Background (`DY`, ...):
  - narrow posteriors = high sensitivity
- Unrecognized (`VBF`, ...):
  - posteriors = priors
- Luminosity nuisance: weakly recognised
- Signal: broad posteriors → high uncertainty
- Comparable results to likelihood fit





- cINN is able to infer the signal strength modifiers
  $\rightarrow$ applicable in HEP
- Good prediction performance
  - Posterior width reflects network sensitivity
  - Signal: sensitivity drop for signal for small μ
  - Comparable results to likelihood fit

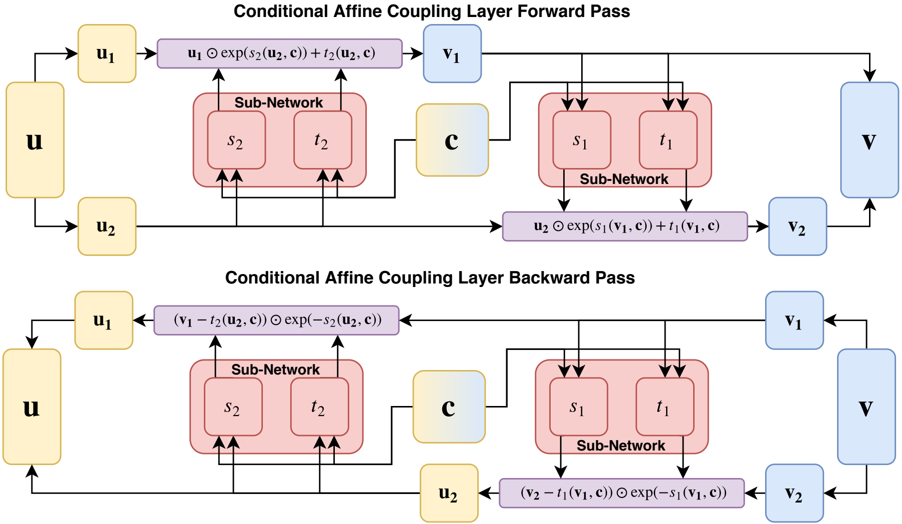




Backup

Network Setup – cINN Architecture

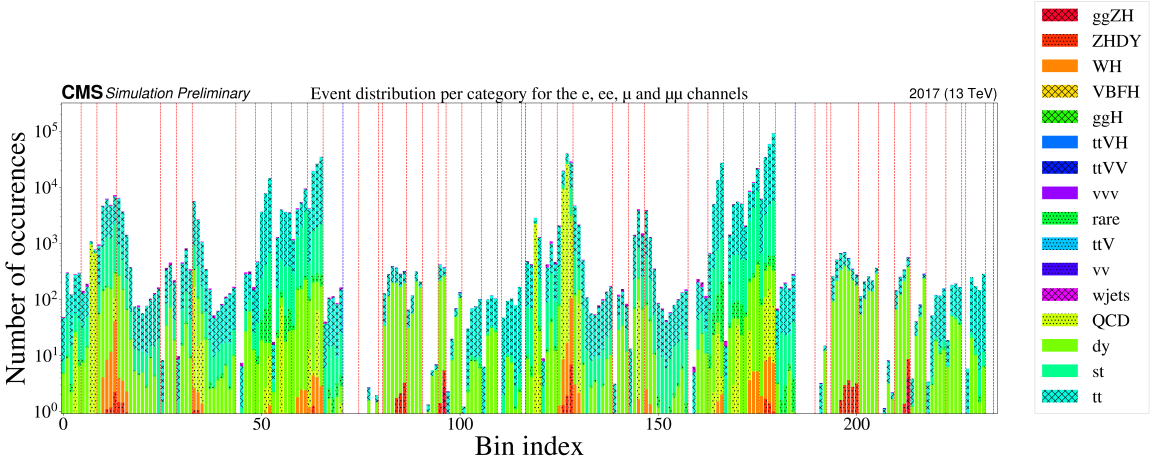
- $\dim c = 235$
- $\dim x = 17$
  - 3 signal modifier parameters
  - 13 background modifier parameters
  - 1 nuisance parameter (luminosity)
- 12 GLOW Blocks with permutation layers
- Subnetworks with 3 layers à 128 nodes with ReLU



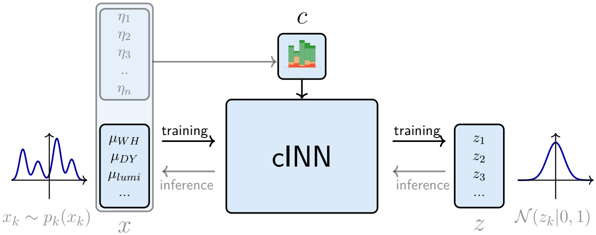
from [2007.08391]

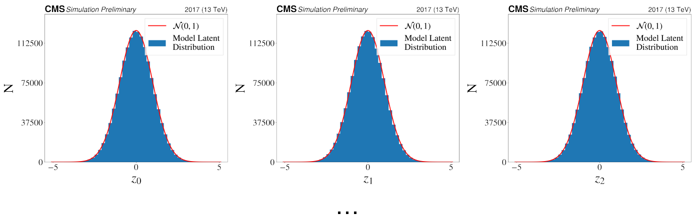
The Conditions – Nominal with $\mu = 1$

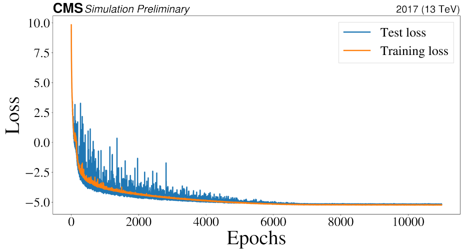
- 4 channels: 2 SL + 2 DL
- 13 subcategories



- Training: loss converges
- Latent space distribution follows $\mathcal{N}(0,1)$
- Sampling from $\mathcal{N}(0,1)$ yields well-approximated posteriors







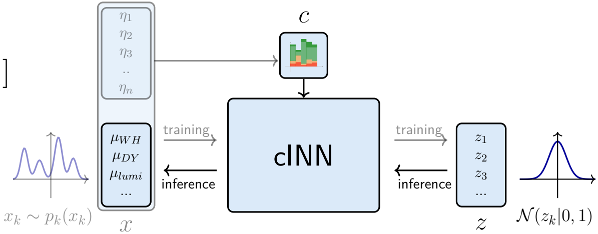
Signal Modifier Parameter Inference – Calibration Curves

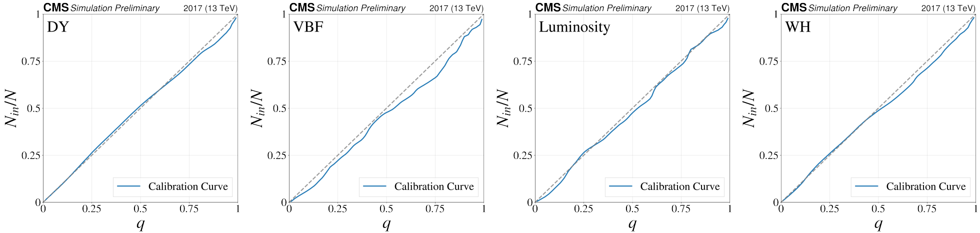
RWTH AACHEN UNIVERSITY
Mate Zoltan Farkas

- Calibration error: measure of model bias

$$e_{cal}(q) = \frac{N^{in}}{N} - q, \qquad q \text{ quantile}; q \in [0,1]$$

- N^{in}: number of posteriors containing the true MC value in their q quantile
- Ideal calibration: $e_{cal}(q) = 0$



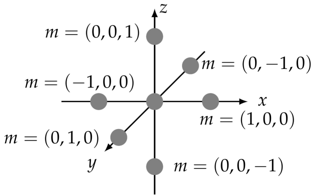


...

- Max median absolute calibration error $\lesssim \mathcal{O}(0.04)$
  $\Rightarrow$ No strong biases in the network model

$$f(x|\mathbf{m}_i) = \underbrace{f(x|0)}_{f_0} + \sum_{j=1}^{T} \underbrace{\left.\frac{\partial f(x|\mathbf{m})}{\partial m_j}\right|_{\mathbf{m}=0}}_{f'_j} (m_i)_j + \sum_{j=1}^{T} \underbrace{\frac{1}{2!} \left.\frac{\partial^2 f(x|\mathbf{m})}{\partial m_j^2}\right|_{\mathbf{m}=0}}_{f'_{jj}} (m_i)_j^2 + \mathcal{O}\left((m_i)_j^3\right)$$

- Express the unknown derivatives f'_j, f'_{jj}
- 24 templates: 24 shape changing uncertainties



$$\begin{pmatrix} f\Big(x|(0, \ \ 0, 0, ..., \ \ 0)\Big) \\ f\Big(x|(0, \ \ 1, 0, ..., \ \ 0)\Big) \\ f\Big(x|(0, -1, 0, ..., \ \ 0)\Big) \\ \vdots \\ f\Big(x|(0, \ \ 0, 0, ..., \ \ 1\Big) \\ f\Big(x|(0, \ \ 0, 0, ..., -1\Big) \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & & 0 & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & -1 & 1 & 0 & & 0 & 0 \\ & & \vdots & & \ddots & \vdots & \\ 1 & 0 & 0 & 0 & & 1 & 1 \\ 1 & 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}}_{M} \begin{pmatrix} f_0 \\ f'_1 \\ f'_{11} \\ \vdots \\ f'_T \\ f'_{TT} \end{pmatrix}$$

$$f(x|\mathbf{m}) = (1, m_1, m_1^2, ..., m_T, m_T^2)M^{-1}\mathbf{f}$$