Machine Learning for LHC Theory

ACAT 2021

AI Decoded: Towards Sustainable, Diverse, Performant and Effective Scientific Computing

Anja Butter

ITP, Universität Heidelberg
First principle based precision simulations

Unique advantage of our field!
Precision simulations with limited resources

\[ \mathcal{L} \]

Matrix element → Parton shower → Hadronization → Detector simulation

Fast evaluation → higher statistics = Precision → higher order

\[ \text{Annual CPU Consumption} \ [\text{MHS06/years}] \]

ATLAS Preliminary 2020 Computing Model - CPU

- Baseline
- Conservative R&D
- Aggressive R&D
- Sustained budget model (+10% +20% capacity/year)

Run 3 (\(\mu=55\))
Run 4 (\(\mu=88-140\))
Run 5 (\(\mu=165-200\))

Year

2020 2022 2024 2026 2028 2030 2032 2034
Boosting standard event generation...

1. Generate phase space points

2. Calculate event weight

\[ w_{\text{event}} = f(x_1, Q^2)f(x_2, Q^2) \times \mathcal{M}(x_1, x_2, p_1, \ldots p_n) \times J(p_i(r))^{-1} \]

3. Unweighting via \( r \geq w/w_{\text{max}} \)
   \[ \rightarrow \text{optimal for } w \approx 1 \]
Boosting standard event generation...

\[ w_{\text{event}} = f(x_1, Q^2)f(x_2, Q^2) \times \mathcal{M}(x_1, x_2, p_1, \ldots p_n) \times J(p_i(r))^{-1} \]

Matrix element

PDF

Phase space mapping
Boosting standard event generation...

\[ w_{\text{event}} = f(x_1, Q^2)f(x_2, Q^2) \times \mathcal{M}(x_1, x_2, p_1, \ldots p_n) \times J(p_i(r))^{-1} \]

- NNPDF since 2002(!)
- S. Carrazza, J. Cruz-Martinez [1907.05075]
Boosting standard event generation...

\[ w_{\text{event}} = f(x_1, Q^2)f(x_2, Q^2) \times \mathcal{M}(x_1, x_2, p_1, \ldots p_n) \times J(p_i(r))^{-1} \]

- Amplitude estimation
- S. Badger, J. Bullock [2002.07516]
- J. Bendavid [1707.00028]

- NNPDF since 2002(!)
- S. Carrazza, J. Cruz-Martinez [1907.05075]

Phase space mapping
Boosting standard event generation...

\[ w_{\text{event}} = f(x_1, Q^2) f(x_2, Q^2) \times M(x_1, x_2, p_1, \ldots p_n) \times J(p_i(r))^{-1} \]

- Amplitude estimation
- S. Badger, J. Bullock [2002.07516]
- J. Bendavid [1707.00028]
- NNPDF since 2002(!)
- S. Carraza, J. Cruz-Martinez [1907.05075]

- Learn phase space distribution
Event simulation with generative models

1. Generative models for phase space sampling
   - Control over phase space density

2. Generative models for event generation
   - Amplification beyond training data?
   - Achieve high precision
   - Estimate uncertainties
Generative Adversarial Networks

**Discriminator** \([D(x_T) \rightarrow 1, D(x_G) \rightarrow 0]\)

\[
L_D = \langle -\log D(x) \rangle_{x \sim P_{Truth}} + \langle -\log(1 - D(x)) \rangle_{x \sim P_{Gen}} \rightarrow -2 \log 0.5
\]

**Generator** \([D(x_G) \rightarrow 1]\)

\[
L_G = \langle -\log D(x) \rangle_{x \sim P_{Gen}}
\]

⇒ New statistically independent samples
What is the statistical value of GANned events?
A.B., S. Diefenbacher, G. Kasieczka, B. Nachmann, T. Plehn, R. Winterhalder [2008.06545]

- Camel function
- Sample vs. GAN vs. 5 param.-fit

Evaluation on quantiles:

\[ \text{MSE}^* = \sum_{j=1}^{N_{\text{quant}}} \left( p_j - \frac{1}{N_{\text{quant}}} \right)^2 \]
What is the statistical value of GANned events?
A.B., S. Diefenbacher, G. Kasieczka, B. Nachmann, T. Plehn, R. Winterhalder [2008.06545]

- Camel function
- Sample vs. GAN vs. 5 param.-fit

Evaluation on quantiles:

\[
\text{MSE}^* = \sum_{j=1}^{N_{\text{quant}}} \left( p_j - \frac{1}{N_{\text{quant}}} \right)^2
\]

→ Amplification factor 2.5

Sparser data → bigger amplification
Training on weighted events
M. Backes, AB, T. Plehn, R. Winterhalder [2012.07873]

Low unweighting efficiencies $\rightarrow$ bottleneck before training

$\rightarrow$ Train on weighted events

$\rightarrow L_D = \left\langle -w \log D(x) \right\rangle_{x \sim P_{Truth}} + \left\langle -\log(1 - D(x)) \right\rangle_{x \sim P_{Gen}}$

Populates high energy tails

Large amplification wrt. unweighted data!
Better control with invertible networks

\[ r \sim \mathcal{N} \]

\[ x \sim \mathcal{P}_{\text{part}} \]

+ Tractable Jacobian
+ Enable correction for perfect precision
+ Fast evaluation in both directions

\[
\begin{pmatrix}
  v_1 \\
  v_2
\end{pmatrix} =
\begin{pmatrix}
  u_1 \cdot s_2(u_2) + t_2(u_2) \\
  u_2
\end{pmatrix}
\]
Training on density

\[ z \sim \mathcal{N} \]

\[ x \sim \mathcal{P}_{\text{part}} \]

- \( z \sim \mathcal{N} \rightarrow \text{NN} \rightarrow x \sim p_x \)
- \( p_x(x) = p_z(z) \cdot J_{\text{NN}} \)
- Given target density \( t(x) \)
  - Train NN to minimize \( \log(p_z(z) \cdot J_{\text{NN}} / t(x)) \)

- Problem: Calculate \( f(x) \) each time
Training on samples


\[ x \sim \mathcal{P}_{\text{samp}} \]

\[ z \sim \mathcal{N} \]

- \( x \sim p_{\text{samples}} \rightarrow \text{NN} \rightarrow z \)
- Train NN to ensure \( z \sim \mathcal{N} \)
- Loss: Maximize posterior over network weights:

\[
- \log(p(\theta|x)) = - \log(p(x|\theta)) - \log(p(\theta)) + \text{const.} \\
= - \log(p(z|\theta)) - \log(J) - \log(p(\theta)) + \text{const.}
\]
Naive INN results

Inclusive $Z$+jets production

- INN easy to train
- Powerful baseline

![Graph showing normalized $Z + 2$ jet exclusive reweighted INN train]

How to deal with deviations?
Naive INN results

Inclusive $Z+j$-jets production

- INN easy to train
- Powerful baseline

- Challenges:
  - Topological holes
  - Sharp phase space features

How to deal with deviations?
I. Corrections through reweighting

Discriminator

\[ \mathcal{L} = - \sum_{x \sim p_{data}} \log(D(x)) - \sum_{x \sim p_{INN}} \log(1 - D(x)) \]

\[ = - \int \! dx \; p_{data}(x) \log(D(x)) + p_{inn}(x) \log(1 - D(x)) \]

From variation we obtain

\[ 0 = \frac{p_{data}(x)}{D(x)} - \frac{p_{inn}(x)}{1 - D(x)} \]

\[ \Rightarrow \frac{p_{data}(x)}{p_{inn}(x)} = \frac{D(x)}{1 - D(x)} \]
Reweighting the generated distributions

+ Close to perfect distribution after reweighting
  - Yields weighted events
II. Discriminator improved training

- Include discriminator information to improve training
- Discflow

\[
\mathcal{L}_\text{DiscFlow} = \sum_{i=1}^{B} w_{D}(x_i) \alpha \left( \frac{\psi(x_i; c_i)^2}{2} - \log J(x_i) \right) \\
\approx \int dx w_{D}(x) \alpha P(x) \left( \frac{\psi(x; c)^2}{2} - \log J(x) \right)
\]
II. Discriminator improved training

- Include discriminator information to improve training
- Discflow

\[ \mathcal{L}_{\text{DiscFlow}} = \sum_{i=1}^{B} w_{D}(x_i)^{\alpha} \left( \frac{\psi(x_i; c_i)^2}{2} - \log J(x_i) \right) \]

\[ \approx \int dx \underbrace{w_{D}(x)^{\alpha} P(x)}_{\text{rewighted truth}} \left( \frac{\psi(x; c)^2}{2} - \log J(x) \right) \]
II. Discriminator improved training

- Include discriminator information to improve training
- Discflow + Reweighting

\[
\mathcal{L}_{\text{DiscFlow}} = \sum_{i=1}^{B} w_D(x_i)^\alpha \left( \frac{\psi(x_i; c_i)^2}{2} - \log J(x_i) \right)
\]

\[
\approx \int dx \left[ w_D(x)^\alpha P(x) \left( \frac{\psi(x; c)^2}{2} - \log J(x) \right) \right] \text{ reweighted truth}
\]
Weight distribution after DiscFlow+Reweighting

![Graph showing weight distribution for different jet configurations.]

- **Z + 1 jet**
- **Z + 2 jet**
- **Z + 3 jet**

The graph displays the normalized distribution of event weights after DiscFlow+Reweighting for different jet configurations, with the x-axis representing the event weight $w_D$ and the y-axis showing the normalized distribution.
III. Addressing uncertainties

\[ \mathcal{L} = \mathcal{L}_{\text{INN}} + KL_{\text{prior}} \]
BINN results

⇒ BINN uncertainty captures convergence of the network ✓
⇒ BINN uncertainty does NOT capture where network fails
IV. Including external uncertainties through conditioning

\[ w = 1 + a \left( \frac{pT,j_1 - 15 \text{ GeV}}{100 \text{ GeV}} \right)^2 \]

\[ \sigma_{\text{BINN}} \]

\[ \mu_{\text{BINN}} \]

\[ \text{without conditioning} \]

→ Include prior over \( \alpha \) in BINN sampling
Overview on uncertainties

$Z + 1$ jet exclusive

Reweighted
Train

$\delta$ [%]

$w_D$

BINN
Truth

Conditioned
Truth

$a \in [0, 6, 12]$

$\frac{p_T,\mu}{[GeV]}$
Can we invert the simulation chain?

What we want to know

What we measure or simulate

wish list:  □ multi-dimensional
□ bin independent
□ statistically well defined
Inverting detector effects

\[
\begin{pmatrix}
X_{\text{part}} \\
r_{\text{part}}
\end{pmatrix}
\xleftarrow{\text{unfolding: } \tilde{g}}
\begin{pmatrix}
X_{\text{det}} \\
r_{\text{det}}
\end{pmatrix}
\]

\[
\text{Pythia, Delphes: } g \rightarrow
\]

multi-dimensional ✓ bin independent ✓ statistically well defined ?
Asking the right question

Given an event $x_d$, what is the probability distribution at parton level? → event generation conditioned on $x_d$

$$x_p \leftarrow g(x_p, f(x_d)) \rightarrow r$$

← unfolding: $\bar{g}(r, f(x_d))$

Minimizing the posterior

$$L = \langle 0.5 \| \bar{g}(x_p, f(x_d)) \|_2^2 - \log |J| \rangle_{x_p \sim P_p, x_d \sim P_d} - \log p(\theta)$$

$r \sim \mathcal{N}$

$x \sim \mathcal{P}_{part}$
Condition INN on detector data

\[ g(x_p, f(x_d)) \rightarrow X_p \leftarrow \text{unfolding: } \bar{g}(r, f(x_d)) \rightarrow r \]

events normalized
cINN eINN
FCGAN
single detector event
3200 unfoldings

fraction of events
cINN eINN FCGAN
0.0 0.2 0.4 0.6 0.8 1.0
quantile \( p_{T,q1} \)
0.0 0.2 0.4 0.6 0.8 1.0
cINN eINN eINN FCGAN

0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
fraction of events
cINN eINN FCGAN
0.0 1.0
Inverting the full event

\[ pp > WZ > q\bar{q}l^+l^- + ISR \rightarrow 2/3/4 \text{ jet events} \]

Train on inclusive dataset

Evaluate exclusive 2/3/4 jet channels

multi-dimensional ✓  bin independent ✓  statistically well defined ✓
Application to MEM

current work in progress with T. Martini, T. Heimel, S. Peitzsch, T. Plehn

- Single top production in association with Higgs
- Measure CP-phase in the top Yukawa coupling

\[
\mathcal{L}(\alpha) = \prod_{i=1}^{N} \frac{1}{\sigma_{\text{fid}}(\alpha)} \int d^m z \frac{d^m \sigma(\alpha)}{dz_1 \ldots dz_m} T(\vec{y}^{(i)}, \vec{z}).
\]
We can use ML ... 

... to improve precision simulations in forward direction

... to **amplify** underlying statistics

... to achieve **precision** with discriminators

... to estimate the corresponding **uncertainties**

... to **invert** the simulation chain statistically