Simulating and unfolding LHC events with generative networks

OSU - HEP seminar

Anja Butter

ITP, Universität Heidelberg
The need for new physics
The need for new physics
Era of data

Standard Model Production Cross Section Measurements

ATLAS Preliminary
Run 1,2 $\sqrt{s} = 5,7,8,13$ TeV

Statistical details:
LHC pp $\sqrt{s} = 13$ TeV
- Data: 3.2 – 139 fb$^{-1}$
LHC pp $\sqrt{s} = 8$ TeV
- Data: 20.2 – 20.3 fb$^{-1}$
LHC pp $\sqrt{s} = 7$ TeV
- Data: 4.5 – 4.9 fb$^{-1}$
LHC pp $\sqrt{s} = 5$ TeV
- Data: 0.025 fb$^{-1}$

Status: May 2020

Anja Butter Simulating and unfolding LHC events with generative networks
Precision simulations with limited resources

\[ \mathcal{L} \]

- Matrix element
- Parton shower
- Hadronization
- Detector simulation

\[ \text{HN3LO} + \text{NNLOJET} \] \hspace{1cm} \text{p} \text{ p} \rightarrow \text{H} + X \] \hspace{1cm} \sqrt{s} = 13 \text{ TeV}

\[ \mu \left[ \mu_H, \mu_F \right] = (\frac{1}{2}, \frac{1}{2}, 1) \mu_H \]

\[ \frac{d\sigma}{dy_H} \left[ \text{pb} \right] \]

\[ \frac{\text{Ratio to NNLO}}{\text{NNLO}} \]

Table 1: Status of the theory uncertainties on Higgs boson production in gluon fusion at \( p_s = 13 \text{ TeV} \). The table is taken from Ref. [83] and the LHC Higgs WG1 TWiki, with (trunc) removed after the work of Ref. [18]. The value for (EW) was a rough estimate when Ref. [83] was published. Meanwhile the order of magnitude has been confirmed by the calculations of Refs. [84–88].

Two-loop electroweak corrections to Higgs production in gluon fusion were calculated in Refs. [89, 90, 78]. The mixed QCD-EW corrections which appear at two loops for the first time were calculated directly in Ref. [91], where however the unphysical limit \( m_Z, m_W \) \rightarrow m_H was employed. In Refs. [84–86], this restriction was lifted and the mixed QCD-EW corrections at order \( \frac{1}{m_t} \) were calculated, where the real radiation contributions were included in the soft gluon approximation. It was found that the increase in the total cross section between pure NLO QCD and NLO QCD+EW is about 5.3%. The calculation of Ref. [86] has been confirmed by Ref. [87], where also the hard real radiation was calculated, in the limit of small vector boson masses, corroborating the value of Ref. [86].

[1807.11501] Cieri, Chen, Gehrmann, Glover, Huss

\[ 10 \]
Precision simulations with limited resources

L

Matrix element

Parton shower

Hadronization

Detector simulation

ATLAS Preliminary
2020 Computing Model -CPU: 2030: Aggressive R&D

Data Proc
MC-Full(Sim)
MC-Full(Rec)
MC-Fast(Sim)
MC-Fast(Rec)
EvGen
Heavy Ions
Data Deriv
MC Deriv
Analysis

Speed = Precision
How can ML help analyzing data

• 1.0 Classification/Regression
  → Label data, eg. Signal vs Background

\[
\text{minimize } L = (y_{true} - y_{output})^2
\]
How can ML help analyzing data

- 1.0 Classification/Regression
  - Label data, eg. Signal vs Background

\[
\text{minimize } L = (y_{true} - y_{output})^2
\]
How can ML help analyzing data

• 1.0 Classification/Regression
  → Label data, eg. Signal vs Background

\[
\text{minimize } L = (y_{\text{true}} - y_{\text{output}})^2
\]

+ low level observables
+ efficient training

Why now? → GPUs
  → new algorithms [convolutional networks]
Comparative top tagging study

→ Other applications: jet calibration, particle identification, ...
→ Open questions: precision, uncertainties, visualization
How can ML help increasing precision

- ML 2.0 Generative models
  - Can we simulate new data?

Fast evaluation $\Rightarrow$ more events $\Rightarrow$ Precision

$\Rightarrow$ higher order

- Speed
  - modular speed up
  - wrapper
  - new concepts
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 importance sampling
   \[ \rightarrow \text{optimal for } w \approx 1 \]
Boosting standard event generation...

$$w_{\text{event}} = f(x_1, Q^2)f(x_2, Q^2) \times \mathcal{M}(x_1, x_2, p_1, \ldots p_n) \times J(p_i(r))^{-1}$$
Boosting standard event generation...

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

- 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...

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

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

- Learn phase space mapping (\( \rightarrow w \approx 1 \))
- Gao et al. [2001.10028]
- Bothmann et al. [2001.05478]

- NNPDF since 2002(!)
- S. Carrazza, J. Cruz-Martinez [1907.05075]
... or training directly on event samples

**Event generation**

- Generating 4-momenta
- $Z > ll$, $pp > jj$, $pp > t\bar{t}+$decay

  [1901.00875] Otten et al. VAE & GAN  
  [1901.05282] Hashemi et al. GAN  
  [1903.02433] Di Sipio et al. GAN  
  [1903.02556] Lin et al. GAN  
  [1907.03764, 1912.08824] Butter et al. GAN  
  [1912.02748] Martinez et al. GAN  
  [2001.11103] Alanazi et al. GAN  
  [2011.13445] Stienen et al. NF  
  [2012.07873] Backes et al. GAN  
  [2101.08944] Howard et al. VAE

**Detector simulation**

- Jet images
- Fast calorimeter simulation

  [1701.05927] de Oliveira et al. GAN  
  [1705.02355, 1712.10321] Paganini et al. GAN  
  [1802.03325, 1807.01954] Erdmann et al. GAN  
  [1805.00850] Musella et al. GAN  
  [1909.01359] Carazza and Dreyer GAN  
  [1912.06794] Belayneh et al. GAN  
  [2005.05334, 2102.12491] Buhmann et al. VAE  
  [2009.03796] Diefenbacher et al. GAN  
  [2009.14017] Lu et al.

NO claim to completeness!
Generative Adversarial Networks

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

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

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

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

⇒ Nash Equilibrium

⇒ New statistically independent samples
What is the statistical value of GANned events?

- 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? [2008.06545]

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

Evaluation on quantiles:

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

→ Amplification factor 2.5

Sparser data → bigger amplification
How to GAN LHC events

- $t\bar{t} \rightarrow 6$ quarks
- 18 dim output
  - external masses fixed
  - no momentum conservation
- Flat observables ✓
  - Systematic undershoot in tails [10-20% deviation]
Correlations

![Correlation plots](image_url)
Correlations

Slice at $p_{T,t} = 100$ GeV
Reaching precision (preliminary)

1. Representation $p_T, \eta, \phi$
2. Momentum conservation
3. Resolve log $p_T$
4. Regularization: spectral norm
5. Batch information

→ 1% precision ✓

Next step automization

$W + 2$ jets
Information in distributions

Information in space distribution
(what we want)

Information in weight
(what we have)
The unweighting bottleneck

- High-multiplicity / higher-order $\rightarrow$ unweighting efficiencies $< 1\%$

$\rightarrow$ Simulate conditions with naive Monte Carlo generator
ME by Sherpa, parton densities from LHAPDF, Rambo-on-diet

$$ pp \rightarrow \mu^+ \mu^- \text{ with } m_{\mu\mu} > 50 \text{ GeV} $$

$\rightarrow$ unweighting efficiency 0.2\%
Training on weighted events

Information contained in distribution or event weights

Train on weighted events
Training on weighted events

Information contained in distribution or event weights

\[ L_D = \left\langle -w \log D(x) \right\rangle_{x \sim P_{\text{Truth}}} + \left\langle -\log(1 - D(x)) \right\rangle_{x \sim P_{\text{Gen}}} \]
Training on weighted events

Information contained in distribution or event weights

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

normalizing flow: B. Stienen, R. Verheyen [2011.13445]
uwGAN results

Populates high energy tails

Large amplification wrt. unweighted data!
Short summary

We can ..

→ use GANs to learn event distributions and correlations
  → amplify underlying statistics
  → achieve precision
  → train directly on weighted events

→ boost precision simulations with generative networks
Can we invert the simulation chain?

<table>
<thead>
<tr>
<th>What we want to know</th>
<th>What we measure or simulate</th>
</tr>
</thead>
<tbody>
<tr>
<td>scattering</td>
<td>detectors</td>
</tr>
<tr>
<td>QCD</td>
<td></td>
</tr>
<tr>
<td>shower</td>
<td></td>
</tr>
<tr>
<td>fragmentation</td>
<td></td>
</tr>
</tbody>
</table>

wish list: □ multi-dimensional  
□ bin independent  
□ statistically well defined
Invertible networks

\[ \text{Pythia, Delphes: } g \rightarrow \bar{g} \leftarrow \text{unfolding: } \overline{\text{det}} \]

- Bijective mapping
- Tractable Jacobian
- Fast evaluation in both directions
- Arbitrary networks \( s \) and \( t \)

Inverting detector effects

- \( pp \rightarrow ZW \rightarrow (ll)(jj) \)
- Train: parton \( \rightarrow \) detector
- Evaluate: parton \( \leftarrow \) detector

multi-dimensional ✓ bin independent ✓ statistically well defined ?
Including stochastical effects

\[
\begin{pmatrix}
    x_p \\
    r_p
\end{pmatrix}
\xrightarrow{\text{unfolding: } \bar{g}}
\begin{pmatrix}
    x_d \\
    r_d
\end{pmatrix}
\]

Sample \( r_d \) for fixed detector event

How often is Truth included in distribution quantile?

- Problem: arbitrary balance of many loss functions
Taking a different angle

Given an event $x_d$, what is the probability distribution at parton level?

→ sample over $r$, condition on $x_d$

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

$X_p \leftarrow$ unfolding: $\tilde{g}(r, f(x_d))$

$\rightarrow r$

$\mathcal{J}$acobian of bijective mapping
Taking a different angle

Given an event $x_d$, what is the probability distribution at parton level?
→ sample over $r$, condition on $x_d$

\[ g(x_p, f(x_d)) \quad \xrightarrow{\text{unfolding: } \bar{g}(r, f(x_d))} \quad r \]

→ Training: Maximize posterior over model parameters

\[ L = - \langle \log p(\theta | x_p, x_d) \rangle_{x_p \sim P_p, x_d \sim P_d} \]
\[ = - \langle \log p(x_p | \theta, x_d) \rangle_{x_p \sim P_p, x_d \sim P_d} - \log p(\theta) + \text{const.} \quad \leftarrow \text{Bayes} \]
\[ = - \left\langle \log p(\bar{g}(x_p, x_d)) + \log \left| \frac{\partial \bar{g}(x_p, x_d)}{\partial x_p} \right| \right\rangle - \log p(\theta) \quad \leftarrow \text{change of var} \]
\[ = \langle 0.5 \| \bar{g}(x_p, f(x_d)) \|^2 - \log |J| \rangle_{x_p \sim P_p, x_d \sim P_d} - \log p(\theta) \]

→ Jacobian of bijective mapping
Cross check distributions

2 jet no ISR

\[ \frac{1}{\sigma} \frac{d\sigma}{dp_{T,q}} \text{ [GeV}^{-1}] \times 10^{-2} \]

\( p_{T,q} \) [GeV]

0 25 50 75 100 125 150 175 200

2 jet no ISR

\[ \frac{1}{\sigma} \frac{d\sigma}{dp_{T,q}} \text{ [GeV}^{-1}] \times 10^{-2} \]

\( p_{T,q} \) [GeV]

0 20 40 60 80 100 120 140

2 jet no ISR

\[ \frac{1}{\sigma} \frac{d\sigma}{dM_{W,\text{reco}}} \text{ [GeV}^{-1}] \times 10^{-1} \]

\( M_{W,\text{reco}} \) [GeV]

65 70 75 80 85 90 95 100
Condition INN on detector data

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

Minimizing \( L = \left\langle 0.5 \left| \left| \bar{g}(x_p, f(x_d)) \right| \right|^2 - \log |J| \right\rangle_{x_p \sim p_p, x_d \sim p_d} - \log p(\theta) \)

multi-dimensional ✓  bin independent ✓  statistically well defined ✓
Inverting the full event I

- \( pp > WZ > q\bar{q}l^+l^- + \text{ISR} \)
  - ISR leads to large fraction of 2/3/4 jet events
- Train and test on exclusive channels
Inverting the full event II

\[ pp > WZ > q\bar{q}l^+l^- + \text{ISR} \]

Train on inclusive dataset

Evaluate exclusive 2/3/4 jet channels
Going beyond unfolding

**Same principle for inference**

Measure parton shower parameters

**Infere splitting kernels**

\[
P_{qq}(z, y) = C_F \left[ D_{qq} \frac{2z(1-y)}{1-z(1-y)} + F_{qq}(1-z) + C_{qq}yz(1-z) \right]
\]

![Diagram of the inference setup](image)

Inference

- LHC jets
- \{x\} → Summary net
- Summary net: \( h \)
- QCD measurement
- \( m \) → cINN
- cINN: \( \tilde{g}(z; h) \)
- Gaussian sampling
- \( z \sim P(z) \)

Posterior Gaussian fit
Relative error of 2%

3

Idealized jet measurements

Before applying BayesFlow to LHC jets including hadronization and detector simulation, we define our theory assumptions and test the corresponding model on an idealized data set using a toy shower. That will give us an idea what kind of measurement we could aim for and will also allow for some simple benchmarking. We have checked that this toy shower agrees with the full Sherpa shower, except that we do not include the e\( \rightarrow \) effects from the 2-loop cusp anomalous dimension.
We can use ML ...

... to enable precision simulations in forward direction

... to turn weighted into unweighted events

... to invert the simulation chain statistically

... for fun and precision :)}
BACK UP
5-dim sphere

 Amplification

6x6x6x6x6 quantiles
500 data points

3x3x3x3x3 quantiles
500 data points
Noise extended INN

![Graphs showing distributions of INN and other variables with legends for Parton Truth, Parton eINN, Detector Truth, and Detector eINN.](image)

Anja Butter Simulating and unfolding LHC events with generative networks 30 / 30
Model dependence

Training on SM dataset
Evaluation on $W'$ dataset

Truth
cINN
FCGAN

$\frac{1}{\sigma} \frac{d\sigma}{dM_{ZW}}$ [GeV$^{-1}$]

$M_{ZW}$ [GeV]

0.8
1.0
1.2
model
True

model / truth

0.8
1.0
1.2
The GAN challenge

or

Why do we need regularization?

Solutions:
Additional loss or restricted network parameters
Improving GAN training

Solutions

• Regularization of the discriminator, eg. gradient penalty [Ghosh, Butter et al., ...]

• Modified training objective:
  • Wasserstein GAN (incl. gradient penalty) [Lin et al., Erdmann et al., ...]
  • Least square GAN (LSGAN) [Martinez et al., ...]
  • MMD-GAN [Otten et al., ...]
  • MSGAN [Datta et al., ...]
  • Cycle GAN [Carazza et al., ...]

• Use of symmetries [Hashemi et al., ...]

• Whitening of data [Di Sipio et al., ...]

• Feature augmentation [Alanazi et al., ...]