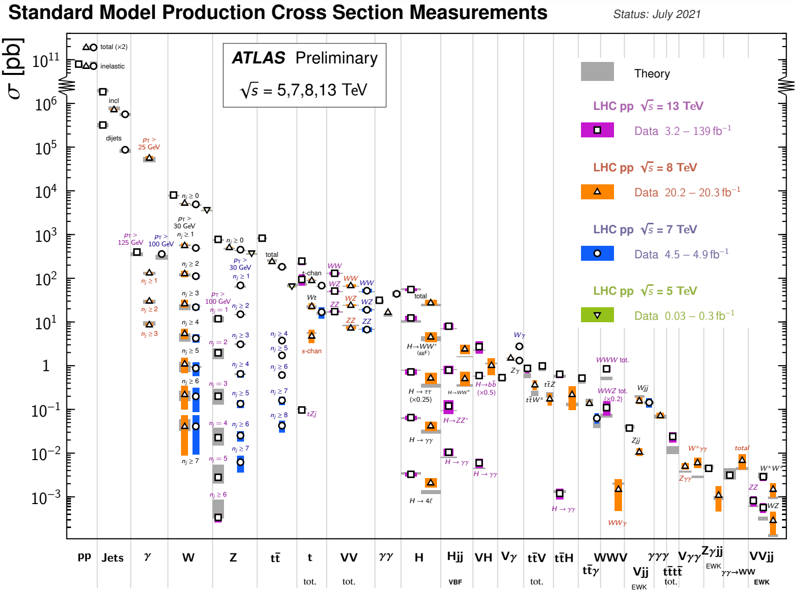
Simulating and unfolding LHC events with generative networks

DESY Theory Workshop
-
Bright ideas for a dark universe

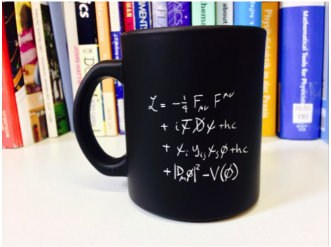
Anja Butter

ITP, Universität Heidelberg

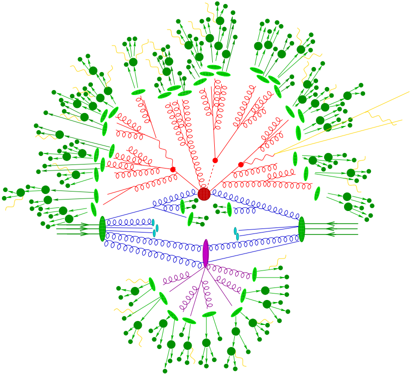
Theory in an era of data



First principle based simulations

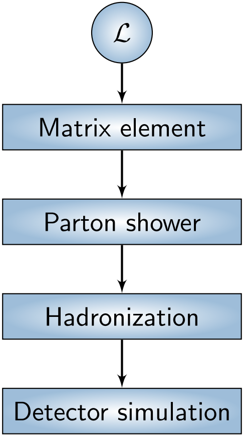


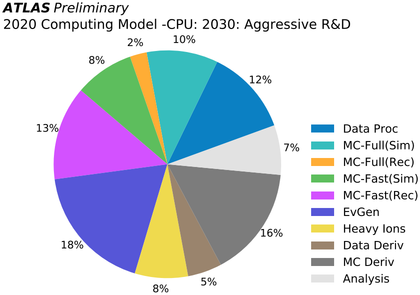
$\longrightarrow$



Need precise simulations tools

Precision simulations with limited resources



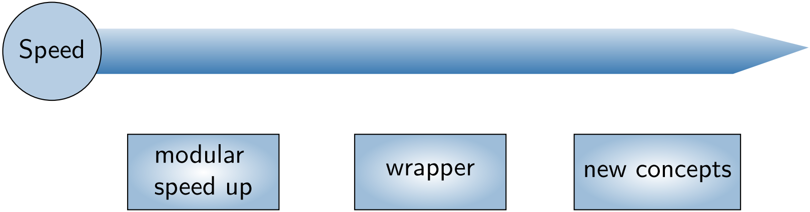


Speed = Precision

How can ML help increasing precision

- ML 2.0 Generative models
  $\rightarrow$ Can we simulate new data?

Fast evaluation → more events = higher order → Precision



Boosting standard event generation...

1. Generate phase space points

2. Calculate event weight

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

3. Unweighting via importance sampling
$\rightarrow$ optimal for $w \approx 1$

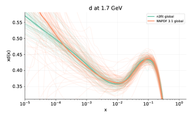
Boosting standard event generation...

$$w_{event} = \underbrace{f(x_1, Q^2) f(x_2, Q^2)}_{\text{PDF}} \times \overbrace{\mathcal{M}(x_1, x_2, p_1, \dots p_n)}^{\text{Matrix element}} \times \underbrace{J(p_i(r))^{-1}}_{\text{Phase space mapping}}$$

Boosting standard event generation...

Matrix element

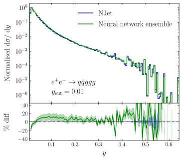
$$w_{event} = f(x_1, Q^2) f(x_2, Q^2) \times \mathcal{M}(x_1, x_2, p_1, \dots p_n) \times J(p_i(r))^{-1}$$



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

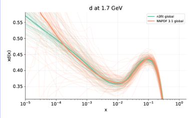
Phase space mapping

Boosting standard event generation...



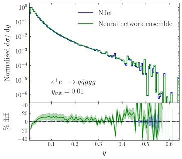
- Amplitude estimation
- S. Badger, J. Bullock [2002.07516]
- J. Bendavid [1707.00028]

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



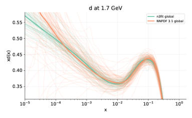
- NNPDF since 2002(!)
- S. Carrazza, J. Cruz-Martinez [1907.05075]

Phase space mapping

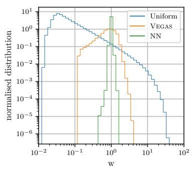


- Amplitude estimation
- S. Badger, J. Bullock [2002.07516]
- J. Bendavid [1707.00028]

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

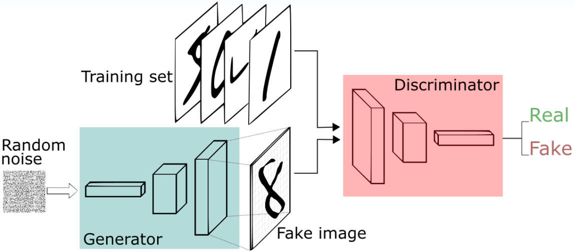


- NNPDF since 2002(!)
- S. Carrazza, J. Cruz-Martinez [1907.05075]



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

Generative Adversarial Networks



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

$$L_D = \left\langle -\log D(x) \right\rangle_{x \sim P_{Truth}} + \left\langle -\log(1 - D(x)) \right\rangle_{x \sim P_{Gen}} \to -2\log 0.5$$

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

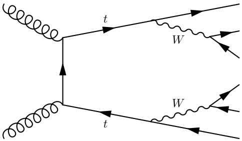
$$L_G = \left\langle -\log D(x) \right\rangle_{x \sim P_{Gen}}$$

$\Rightarrow$ **Equilibrium**

$\Rightarrow$ **New statistically independent samples**

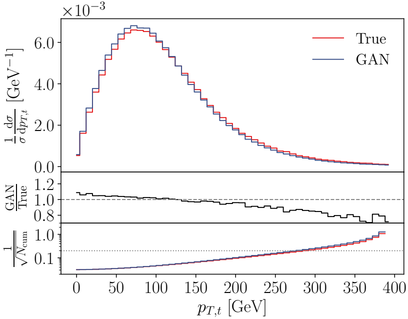
How to GAN LHC events [1907.03764]

- $t\bar{t} \rightarrow 6$ quarks
- 18 dim output
  - external masses fixed
  - no momentum conservation

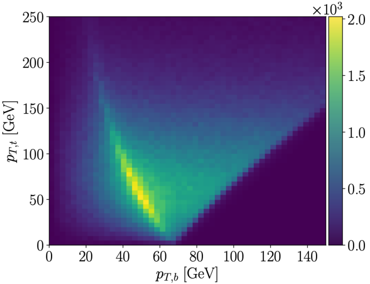


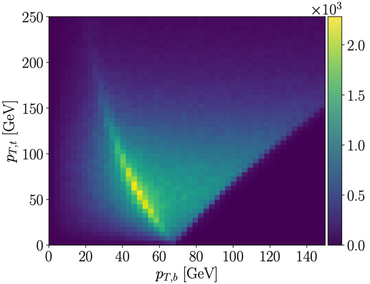
\+ Flat observables ✓

– Systematic undershoot in tails [10-20% deviation]



Correlations





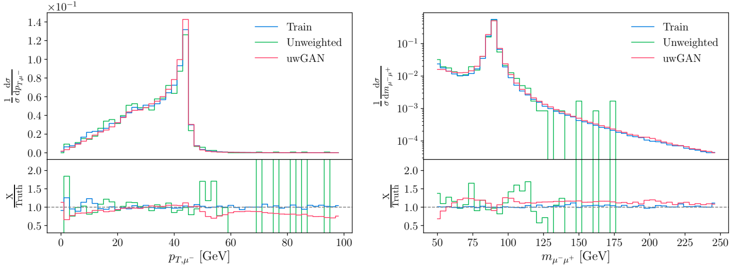
✓

Training on weighted events

Low unweighting efficiencies $\rightarrow$ bottleneck before training

$\rightarrow$ Train on weighted events

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



Populates high energy tails

Large amplification wrt. unweighted data!

Short summary

GANs can ..

$\rightarrow$ learn event distributions and correlations

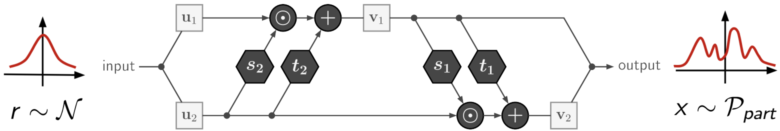
$\rightarrow$ amplify underlying statistics

$\rightarrow$ train on weighted events

$\rightarrow$ How to achieve precision?

$\rightarrow$ How to quantify uncertainties?

Invertible networks



+ Bijective mapping

+ Fast evaluation in both directions

+ Tractable Jacobian

+ Enable correction for perfect precision

+ Extendable to Bayesian invertible networks

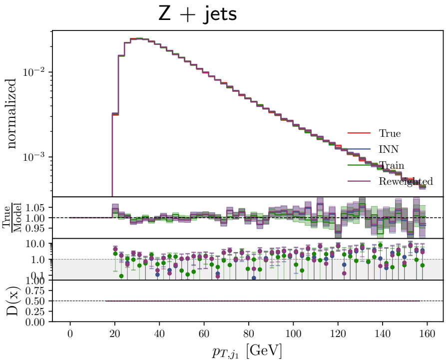
+ Trainable on either density or samples

Preliminary

Inclusive Z+jets production

- INN easy trainable, powerful baseline

- Challenges:
  - Variable number of jets
  - Topological holes
  - 1% precision in correlations
  - Associated uncertainties



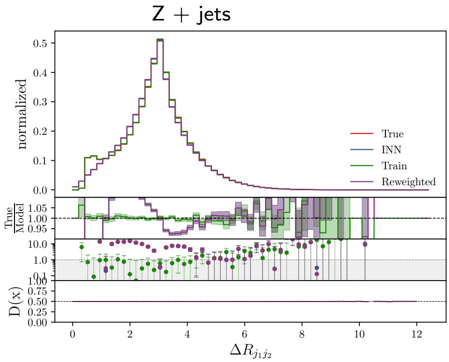
arXiv:2110.XXXXX

Preliminary

Inclusive Z+jets production

- INN easy trainable, powerful baseline

- Challenges:
  - Variable number of jets
  - Topological holes
  - 1% precision in correlations
  - Associated uncertainties



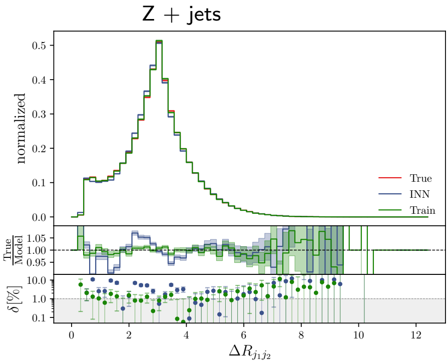
arXiv:2110.XXXXX

Preliminary

Inclusive Z+jets production

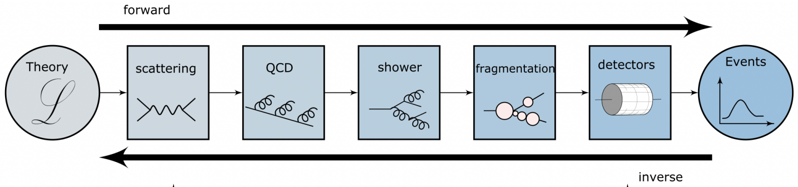
- INN easy trainable, powerful baseline

- Challenges:
  - Variable number of jets
  - Topological holes
  - 1% precision in correlations
  - Associated uncertainties



arXiv:2110.XXXXX

Can we invert the simulation chain?



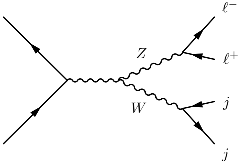
What we
want to know

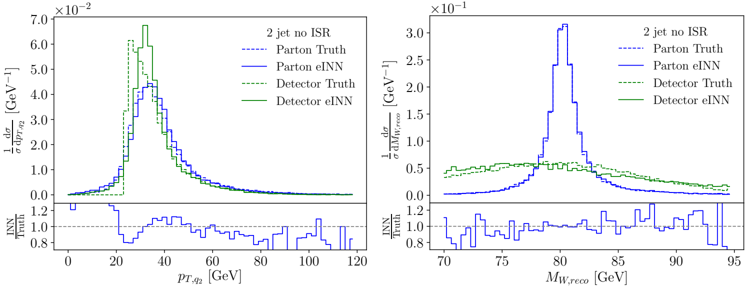
What we
measure or simulate

wish list: ☐ multi-dimensional
☐ bin independent
☐ statistically well defined

Inverting detector effects

$$\begin{pmatrix} x_{part} \\ r_{part} \end{pmatrix} \xleftrightarrow[\leftarrow \text{ unfolding:}\bar{g}]{\text{PYTHIA,DELPHES:}g\rightarrow} \begin{pmatrix} x_{det} \\ r_{det} \end{pmatrix}$$



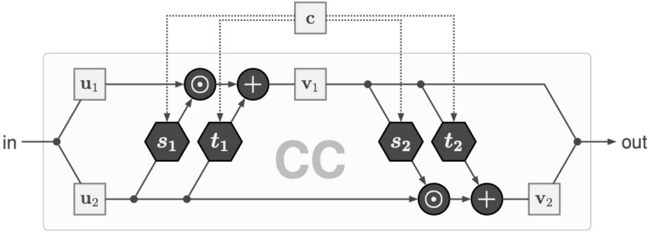


multi-dimensional ✓ bin independent ✓ statistically well defined ?

Taking a different angle

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

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

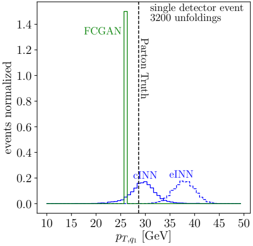


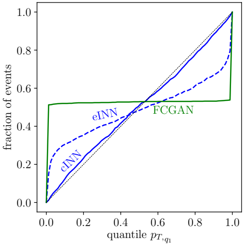
Condition INN on detector data [2006.06685]

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

Minimizing the posterior

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



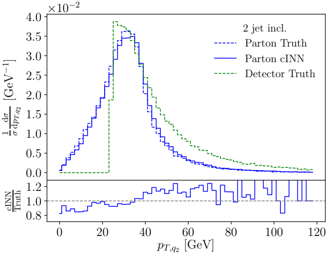


multi-dimensional ✓ bin independent ✓ statistically well defined ✓

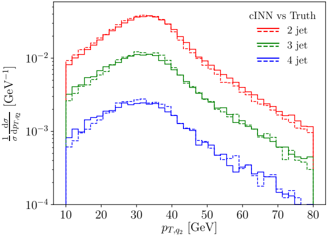
Inverting the full event

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

Train on inclusive dataset



Evaluate
exclusive 2/3/4 jet channels



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 :)

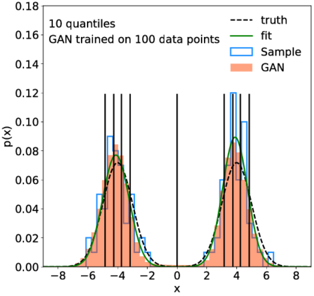
BACK UP

What is the statistical value of GANned events? [2008.06545]

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

Evaluation on quantiles:

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

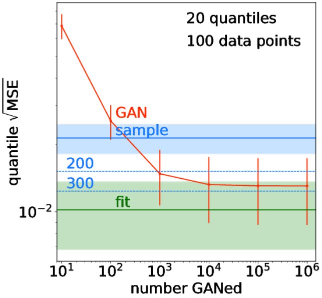


What is the statistical value of GANned events?[2008.06545]

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

Evaluation on quantiles:

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

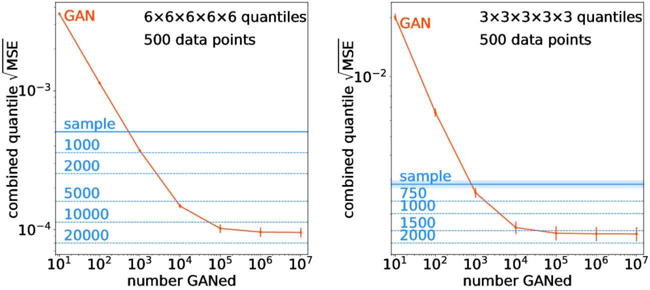


$\rightarrow$ Amplification factor 2.5

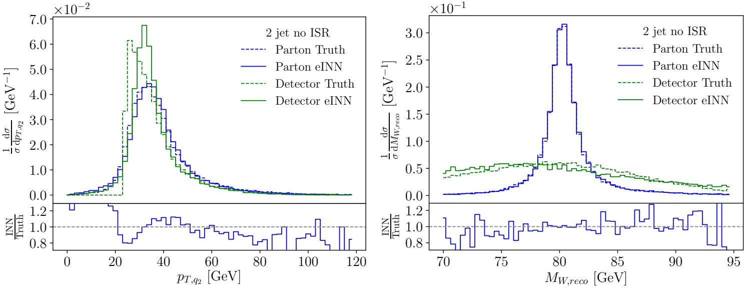
Sparser data $\rightarrow$ bigger amplification

Amplification

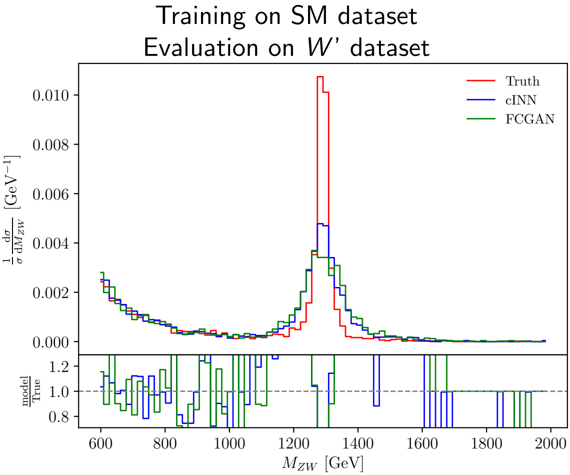
5-dim sphere



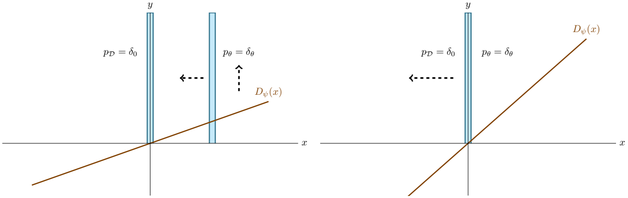
Noise extended INN



Model dependence



The GAN challenge
or
Why do we need regularization?



(a) $t = t_0$ (b) $t = t_1$

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., ...]