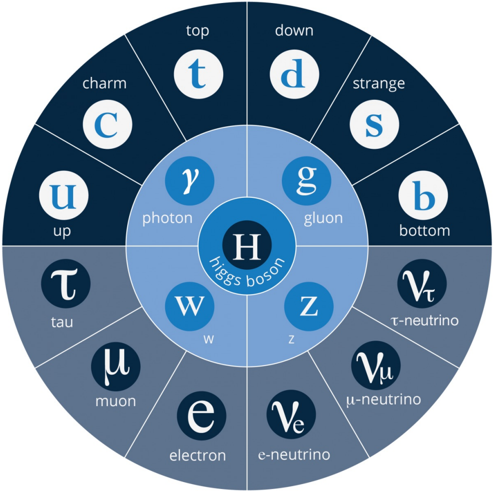
Simulating LHC events with generative networks

DESY-HU Theorie-Seminar

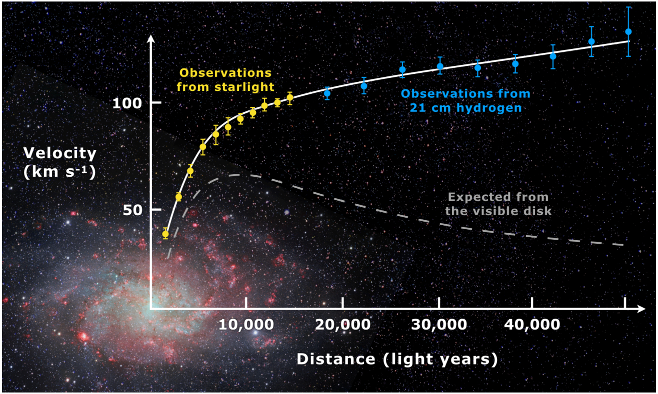
Anjа Butter

ITP, Universität Heidelberg

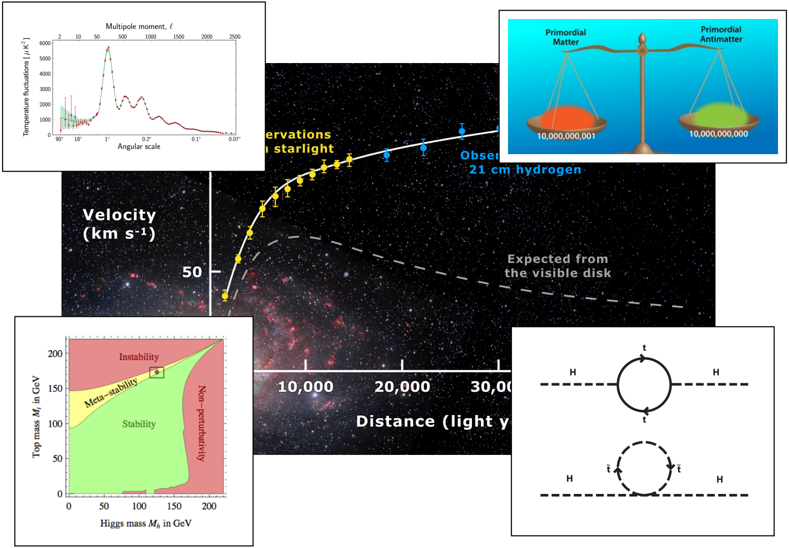
A structurally complete theory



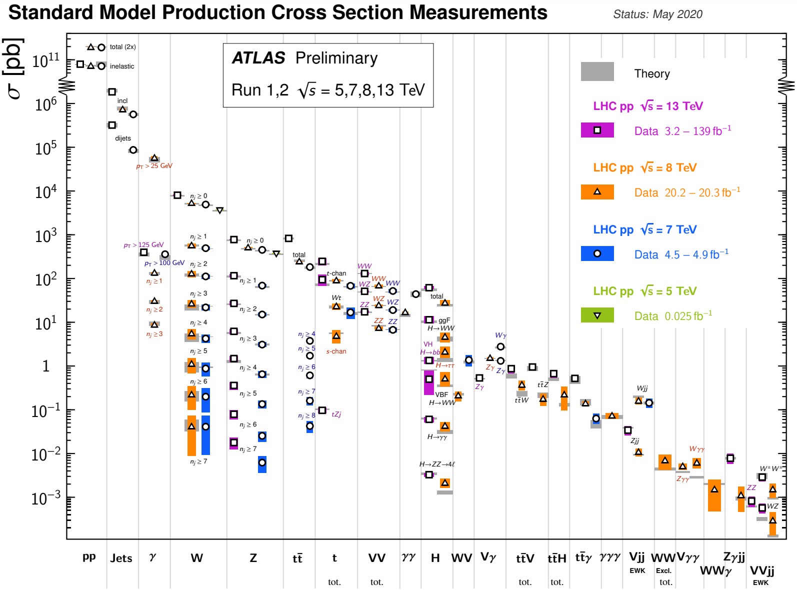
The need for new physics



The need for new physics



Era of data


Standard Model Production Cross Section Measurements
Status: May 2020
ATLAS Preliminary
Run 1,2 $\sqrt{s}$ = 5,7,8,13 TeV
Theory
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}
σ [pb]
pp
Jets
γ
W
Z
$t\bar{t}$
t
VV
$\gamma\gamma$
H
WV
Vγ
$t\bar{t}$V
$t\bar{t}$H
$t\bar{t}\gamma$
$\gamma\gamma\gamma$
Vjj
WW
V$\gamma\gamma$
WWγ
Zγjj
VVjj

First principle based event generation

$\mathcal{L}$

↓

Matrix element

↓

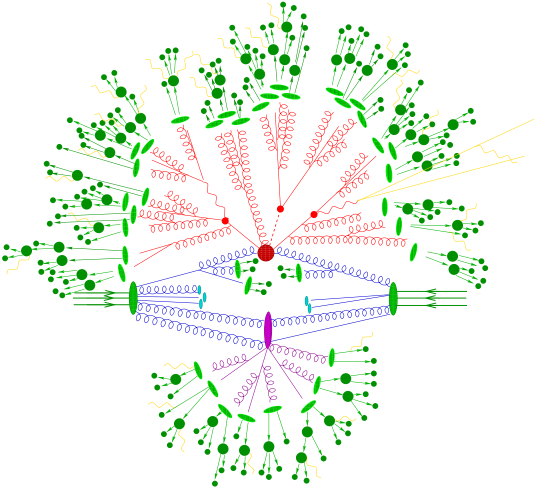
Parton shower

↓

Hadronization

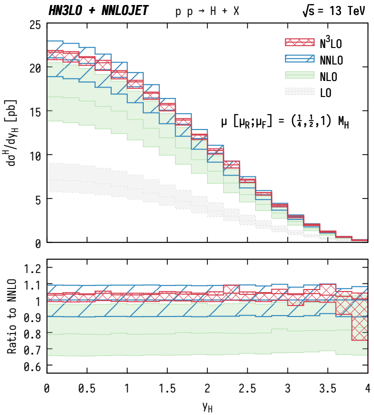
↓

Detector simulation



a sherpa artist

Precision simulations



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

New physics is hidden

Big data
HL-LHC $= 25\times$Run2

First-principle
precision simulation

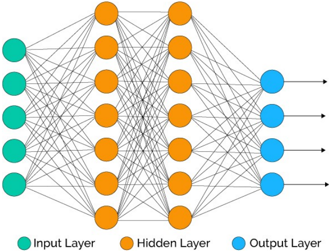
$+$

Make best use of both!

Big data techniques

How can ML help to find new physics

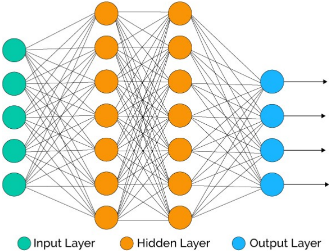
- 1.0 Classification/Regression
  $\rightarrow$ Label data



minimize $L = (y_{true} - y_{output})^2$

How can ML help to find new physics

- 1.0 Classification/Regression
  $\rightarrow$ Label data



minimize $L = (y_{true} - y_{output})^2$

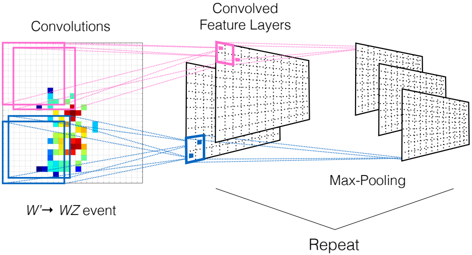
$+$ low level observables
$+$ efficient training

Why now? $\rightarrow$ GPUs
$\rightarrow$ new algorithms [convolutional networks]

First application - jet tagging

Convolutional network on W/QCD jet images

+ Physics: theoretical and experimental control
+ Straight forward from ML developments



[1511.05190] L. Oliveira, M. Kagan, L. Mackey, B. Nachman, A. Schwartzman

Top tagging with physics networks

- pixel $\rightarrow$ Lorentz vectors
- Lo(rentz) La(yer):
  Lorentz vectors $\rightarrow$ physics motivated objects

[1707.08966] AB, G. Kasieczka, T. Plehn, M. Russell

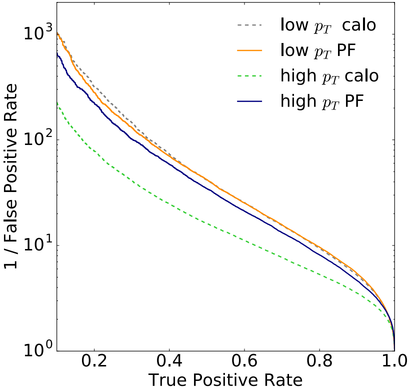
$$\tilde{k}_j \xrightarrow{\text{LoLa}} \hat{k}_j = \begin{pmatrix} m^2(\tilde{k}_j) = \tilde{k}_{j,\mu}\ \eta^{\mu\nu}\ \tilde{k}_{j,\nu} \\ p_T(\tilde{k}_j) \\ w_{jm}^{(E)}\ E(\tilde{k}_m) \\ w_{jm}^{(d)}\ d_{jm}^2 \end{pmatrix}$$

with trainable diagonal metric $\quad d_{jm}^2 = (\tilde{k}_j - \tilde{k}_m)_\mu\ \eta^{\mu\nu}\ (\tilde{k}_j - \tilde{k}_m)_\nu$

Training yields:

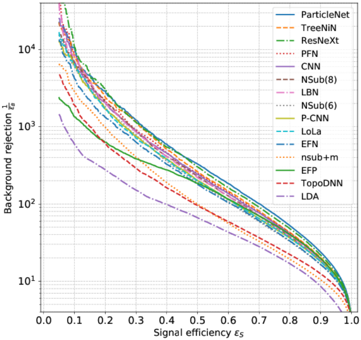
$$\boxed{\eta = \text{diag}(\quad 0.99 \pm 0.02, -1.01 \pm 0.01, -1.01 \pm 0.02, -0.99 \pm 0.02)}$$ 🙂

LoLa vs Image



- Combine tracking & calorimeter information
- Improved performance for boosted jets
- Less trainable weights

Comparative top tagging study



[1902.09914] G. Kasieczka, et al.

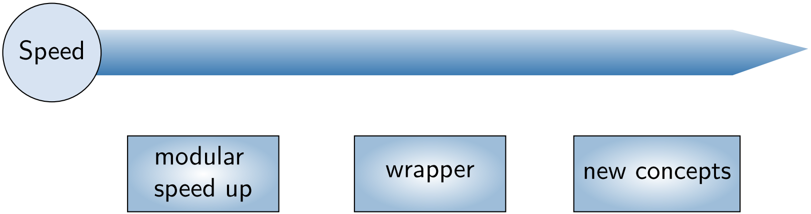
→ Other applications: jet calibration, particle identification, ...

→ Open questions: precision, uncertainties, visualization

Precision in forward simulations

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

Speed → more events = higher order → Precision



Boosting standard event generation...

1. Generate phase space points

2. Calculate event weight

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

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

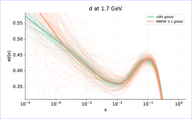
Boosting standard event generation...

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

Boosting standard event generation...

Matrix element

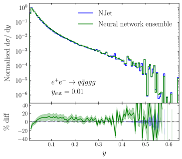
$$w_{event} = f(x_1, Q^2) f(x_2, Q^2) \times \mathcal{M}(x_1, x_2, p_1, \dots p_n) \times J(p_i(r))^{-1}$$



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

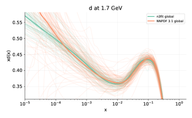
Phase space mapping

Boosting standard event generation...



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

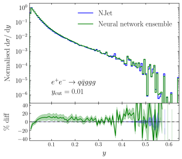
$$w_{event} = f(x_1, Q^2) f(x_2, Q^2) \times \mathcal{M}(x_1, x_2, p_1, \ldots p_n) \times J(p_i(r))^{-1}$$



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

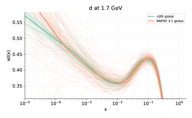
Phase space mapping

Boosting standard event generation...

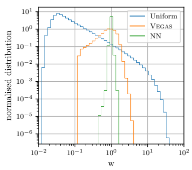


- Amplitude estimation
- S. Badger, J. Bullock [2002.07516]
- J. Bendavid [1707.00028]

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



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



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

... 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**

Detector simulation

- Jet images
- Fast shower simulation in calorimeters

[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**

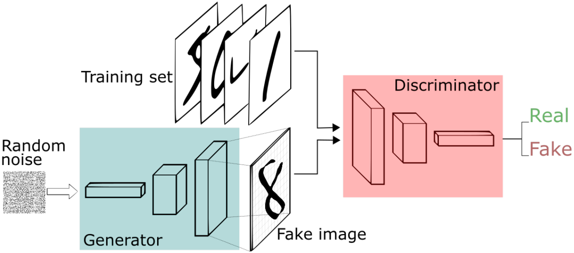
[ATL-SOFT-PUB-2018-001, ATLAS-SIM-2019-004, ATL-SOFT-PROC-2019-007] ATLAS **VAE** & **GAN**

[1909.01359] Carazza and Dreyer **GAN**

[2005.05334] Buhmann et al. **VAE**

NO claim to completeness!

Generative Adversarial Networks



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

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

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

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

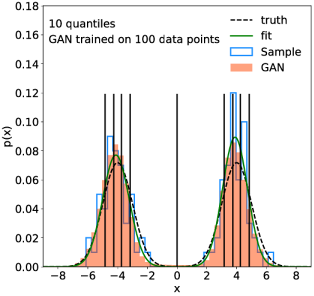
$\Rightarrow$ **New statistically independent samples**

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

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

Evaluation on quantiles:

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

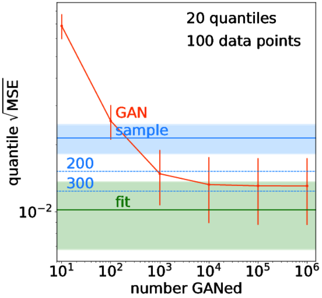


What is the statistical value of GANned events?[2008.06545]

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

Evaluation on quantiles:

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

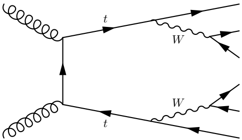


$\rightarrow$ Amplification factor 2.5

Sparser data $\rightarrow$ bigger amplification

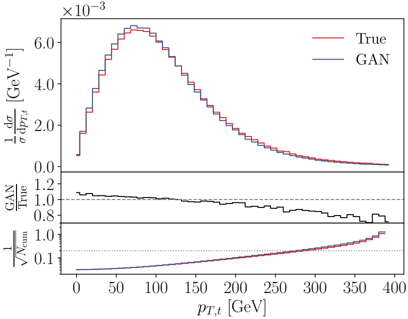
How to GAN LHC events [1907.03764]

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

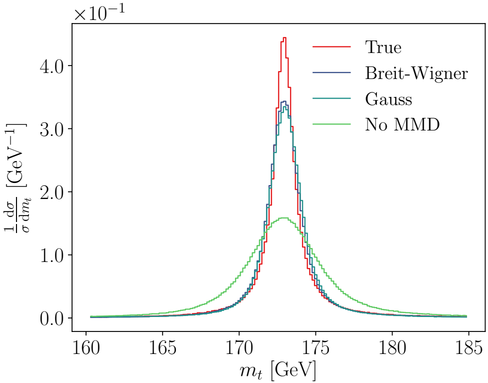


\+ Flat observables ✓

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



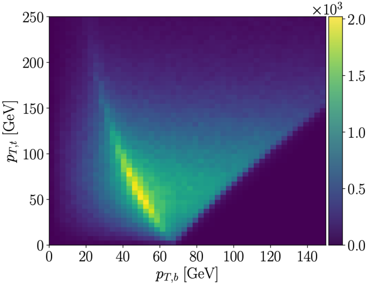
Special features

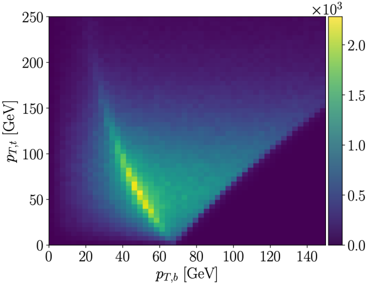


Solution: MMD kernel

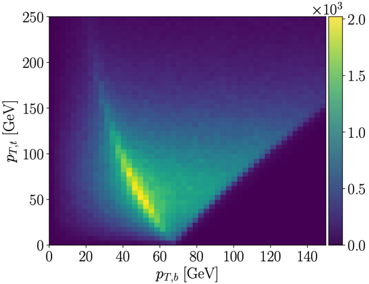
$$\text{MMD}^2(P_T, P_G) = \left\langle k(x,x')\right\rangle_{x,x'\sim P_T} + \left\langle k(y,y')\right\rangle_{y,y'\sim P_G} - 2\left\langle k(x,y)\right\rangle_{x\sim P_T, y\sim P_G}$$

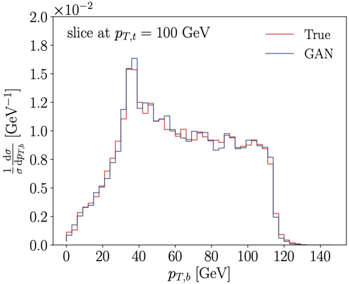
Correlations





Correlations





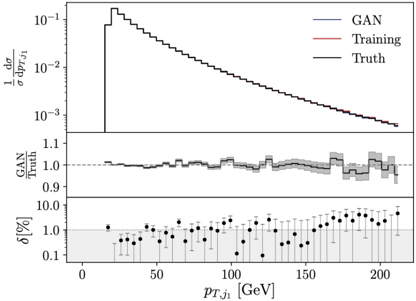
Reaching precision (preliminary)

1. Representation p_T, η, ϕ
2. Momentum conservation
3. Resolve log p_T
4. Regularization: spectral norm
5. Batch information

$\rightarrow$ 1% precision ✓

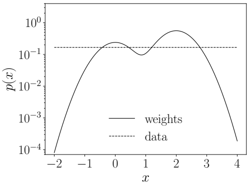
Automization?

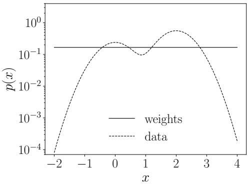
W + 2 jets



Training on weighted events

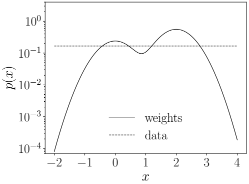
Information contained in distribution or event weights

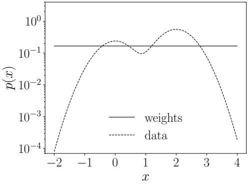




Training on weighted events

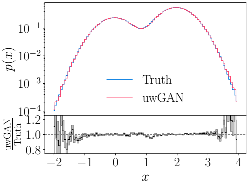
Information contained in distribution or event weights

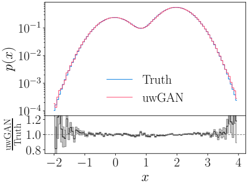




Train on weighted $\rightarrow$ generate unweighted events

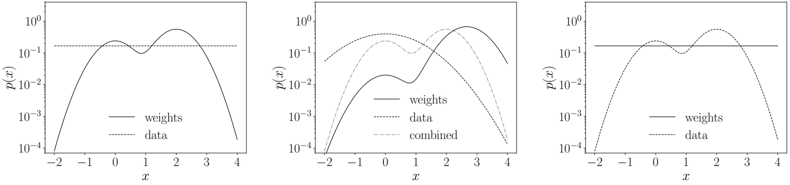
$$L_D = \Big\langle -\, w \log D(x) \Big\rangle_{x\sim P_{Truth}} + \Big\langle -\log(1 - D(x)) \Big\rangle_{x\sim P_{Gen}}$$





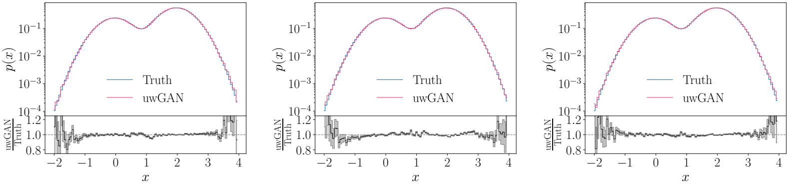
Training on weighted events

Information contained in distribution or event weights



Train on weighted $\rightarrow$ generate unweighted events

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

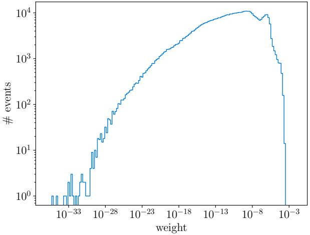


The unweighting bottleneck

- High-multiplicity processes & higher-order calculations
  $\rightarrow$ unweighting efficiency below 1%

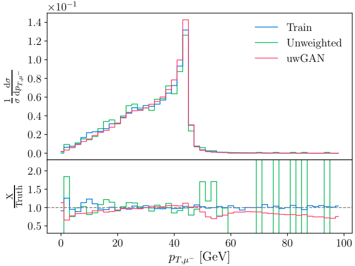
$\rightarrow$ Simulate conditions with naive Monte Carlo generator
[ME by `Sherpa`, parton densities from `LHAPDF`, Rambo-on-diet]

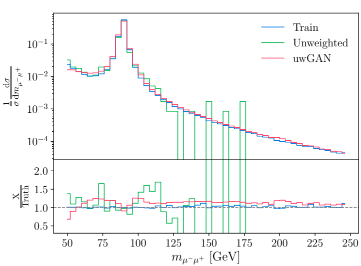
$pp \rightarrow \mu^+\mu^-$ with $m_{\mu\mu} > 50$ GeV

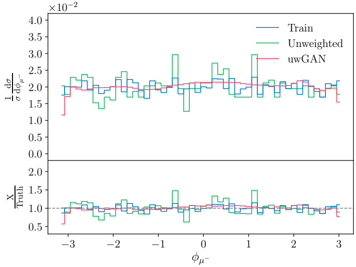


unweighting efficieny $4 \cdot 10^{-3}$

Results





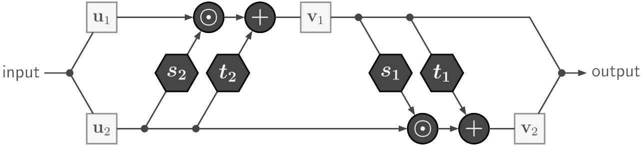


Large amplification factor

Can we invert a Markov process?

$$\left(x_p\right) \xleftrightarrow[\leftarrow \text{ unfolding:}\bar{g}]{\textsc{Pythia},\textsc{Delphes}:g\rightarrow} \left(x_d\right)$$

Invertible networks



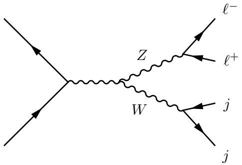
[1808.04730] L. Ardizzone, J. Kruse, S. Wirkert, D. Rahner,

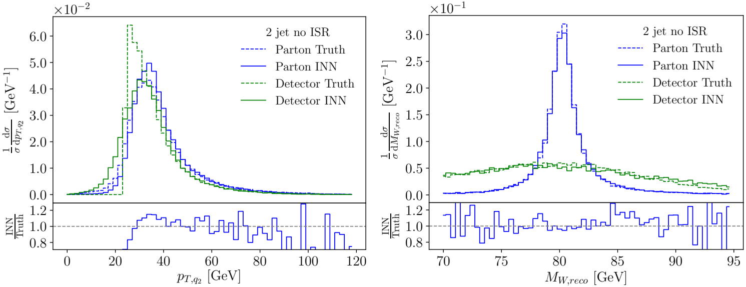
E. W. Pellegrini, R. S. Klessen, L. Maier-Hein, C. Rother, U. Köthe

- Fast evaluation in both directions
- Tractable Jacobian
- Arbitrary networks *s* and *t*

Inverting detector effects

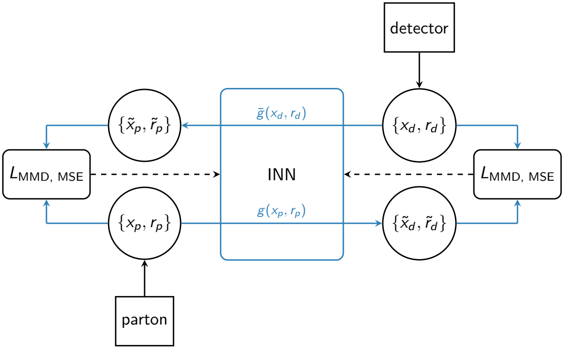
- $pp \to ZW \to (ll)(jj)$
- Train parton $\to$ detector
- Evaluate detector $\to$ parton





Including stochastical effects

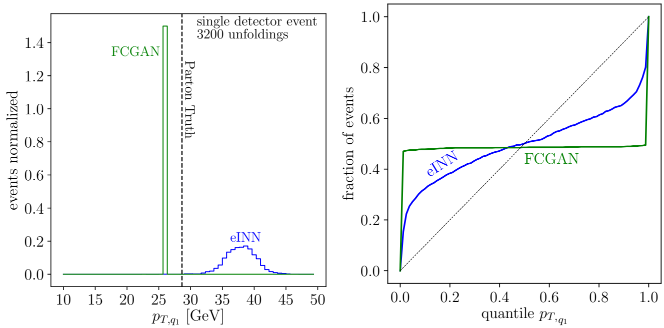
- So far only mapping of mean values
- Extend with noise to include probabilistic nature



- Improved stability via training in both directions
- MSE fixes mean values
- MMD fixes distributions

Calibration curves

$$\begin{pmatrix} x_p \\ r_p \end{pmatrix} \xleftrightarrow[\leftarrow \text{unfolding:}\bar{g}]{\textsc{Pythia,Delphes}:g\rightarrow} \begin{pmatrix} x_d \\ r_d \end{pmatrix}$$



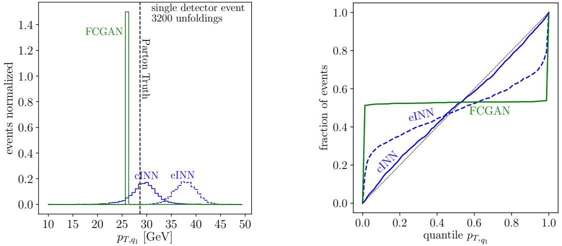
- Mean correct, distribution too narrow
- Problem: arbitrary balance of many loss functions

Condition INN on detector data [2006.06685]

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

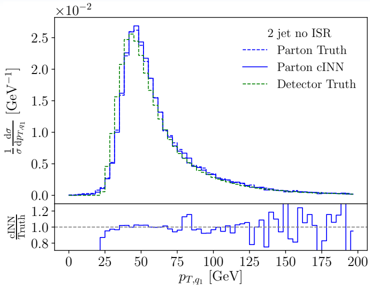
Training: Maximize posterior over model parameters

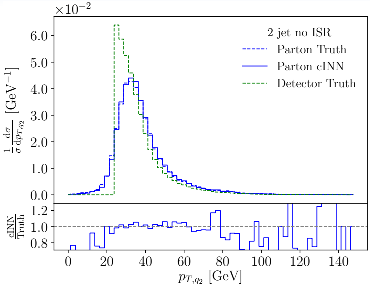
$$\text{Minimizing } L = -\left\langle \log p(\theta | x_p, x_d) \right\rangle_{x_p \sim P_p, x_d \sim P_d}$$
$$= \left\langle 0.5 ||g(x_p, f(x_d)))||_2^2 - \log |J| \right\rangle_{x_p \sim P_p, x_d \sim P_d} - \log p(\theta)$$

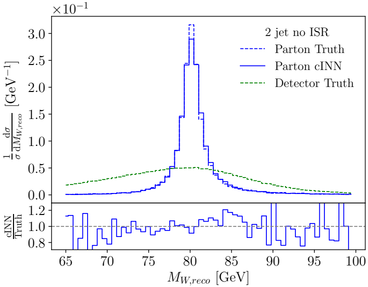


$\rightarrow$ calibrated parton level distributions

Cross check distributions



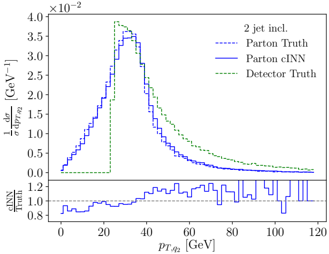




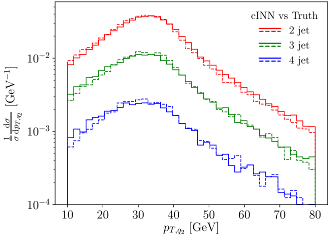
Inverting the full event

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

Train on inclusive dataset



Evaluate
exclusive 2/3/4 jet channels



We can use ML to ...

... improve analyses with optimized S vs B classification

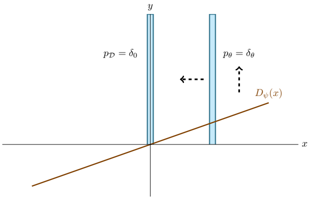
... enable precision simulations in forward direction

... unfold high dimensions

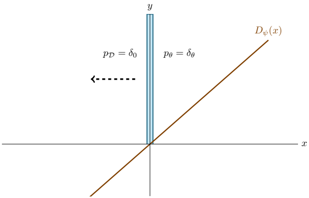
... learn more about particle physics!

BACK UP

The GAN challenge
or
Why do we need regularization?



(a) $t = t_0$



(b) $t = t_1$

Solutions:
Additional loss or restricted network parameters

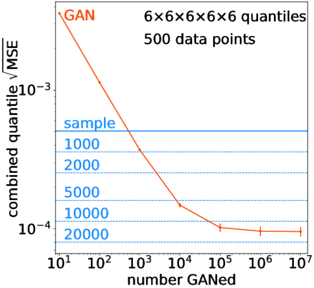
Improving GAN training

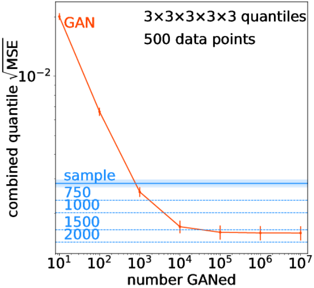
Solutions

- 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., ...]

Amplification

5-dim sphere





Model dependence