Generative Models in Event Simulation

Anja Butter

ITP, Universität Heidelberg

arXiv:1907.03764, 1912.08824, and 1912.00477

with Marco Bellagente, Gregor Kasieczka, Tilman Plehn, and Ramon Winterhalder

The HEP trinity

Theory

Fundamental Lagrangian

- Perturbative QFT

Standard Model vs. new physics

- Matrix elements, loop integrals

Experiment

Complex detector

- ATLAS, CMS, LHCb, ALICE, ...

Reconstruction of individual events

- Big data: jet images, tracks, ...

Precision simulations

First-principle Monte Carlo generators

- Simulation of parton/particle-level events
- Herwig, Pythia, Sherpa, Madgraph, ...

Detector simulation

- Geant4, PGS, Delphes, ...

$\Rightarrow$ **Unweighted event samples**

Neural networks for precision simulations

Problems in MC simulations

- High-dimensional phase space
- Low unweighting efficiency
- Slow detector simulations [$\rightarrow$ talk by Aishik Ghosh]

Solution with neural networks

- Flexible parametrisation
- Interpolation properties
- Fast evaluation
- Multiple generative models: GAN, VAE, normalizing flow

Possibilities for ML in event generation

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**

Monte Carlo integration

- Estimating matrix element
- Neural importance sampling

[1707.00028] Bendavid, **Regression** & **GAN**

[1810.11509] Klimek and Perelstein **NF**

[1912.11055] Bishara and Montull **Regression**

[2001.05478] Bothmann et al. **NF**

[2001.05486, 2001.10028] Gao et al. **NF**

[2002.07516] Badger and Bullock **Regression**

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, ATL-SOFT-PROC-2019-007] ATLAS **VAE** & **GAN**

[1909.01359] Carazza and Dreyer **GAN**

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

Unfolding

- Detector to parton/particle level distributions

[1806.00433] Datta et al. **GAN**

[1911.09107] Andreassen et al.

[1912.0047] Bellagente et al. **GAN**

NO claim to completeness!

Generative Adversarial Networks

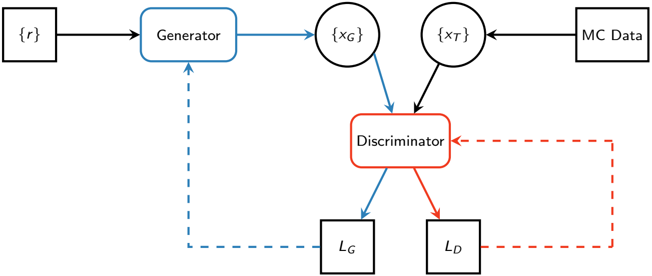
- Training data: true events $\{x_T\}$
  Output data: generated events $\{x_G\}$
- **Discriminator** distinguishes $\{x_T\}, \{x_G\}$ $[D(x_T) \to 1,\ D(x_G) \to 0]$

$$L_D = \Big\langle -\log D(x) \Big\rangle_{x\sim P_T} + \Big\langle -\log(1 - D(x)) \Big\rangle_{x\sim P_G} \xrightarrow{D(x)\to 0.5} -2\log 0.5$$

- **Generator** fools discriminator $[D(x_G) \to 1]$

$$L_G = \Big\langle -\log D(x) \Big\rangle_{x\sim P_G}$$

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



Why GANs? Features, problems and solutions

+ Generate better samples than VAE
+ Large community working on GANs

- Unstable training

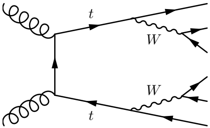
Solutions

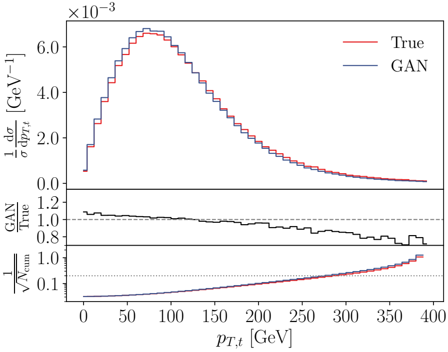
- Regularization of the discriminator, eg. gradient penalty [Ghosh, Butter et al., ...]
- Modified training objective:
  - Wasserstein GAN (incl. gradient penalty) [Lin et al., Erdmann et al., ...]
  - Least square GAN (LSGAN) [Martinez et al., ...]
  - MMD-GAN [Otten et al., ...]
  - MSGAN [Datta et al., ...]
  - Cycle GAN [Carazza et al., ...]
- Use of symmetries [Hashemi et al., ...]
- Whitening of data [Di Sipio et al., ...]
- Feature augmentation [Alanazi et al., ...]

How to GAN LHC events

Idea: generate hard process

- Realistic LHC final state $t\bar{t} \to 6$ jets [1907.03764]
- 18 dim output
  - external masses fixed
  - no momentum conservation
- Flat observables precise
- Systematic undershoot in tails [10-20% deviation]



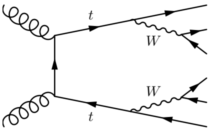


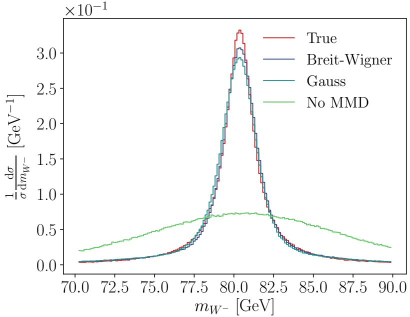
How to GAN LHC events

Idea: generate hard process

- Realistic LHC final state $t\bar{t} \to 6$ jets [1907.03764]
- 18 dim output
- Flat observables precise
- Systematic undershoot in tails [10-20% deviation]
- Sharp phase-space structures, not using Γ_W

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

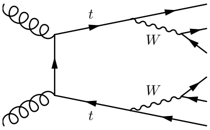


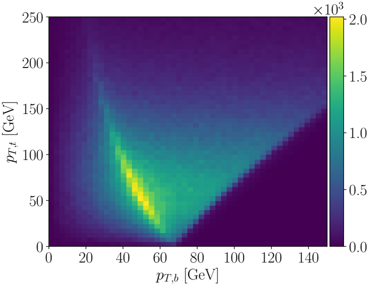


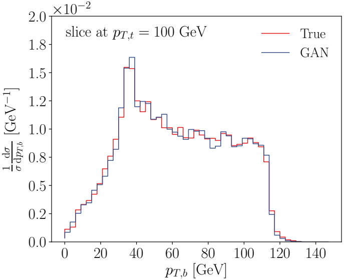
How to GAN LHC events

Idea: generate hard process

- Realistic LHC final state $t\bar{t} \to 6$ jets [1907.03764]
- 18 dim output
- Flat observables precise
- Systematic undershoot in tails [10-20% deviation]
- Sharp phase-space structures, not using Γ_W [MMD-loss]
- 2D correlations







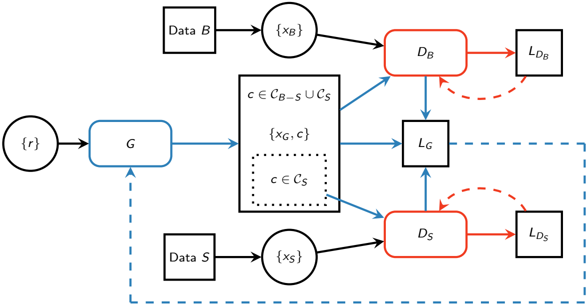
How to GAN event subtraction

Idea: sample based subtraction of distributions [1912.08824]

1. Consistent multidimensional difference between two distributions
2. Beat bin-induced statistical uncertainty [interpolation of distributions]

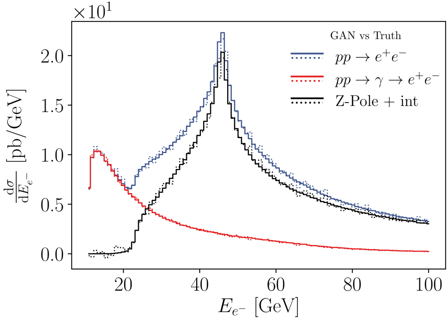
$$\Delta_{B-S} = \sqrt{n_B^2 N_B + n_S^2 N_S} > \max(\Delta_B, \Delta_S)$$

- Many applications:
  - Soft-collinear subtraction, multi-jet merging, on-shell subtraction
  - Background subtraction [4-body decays $\longrightarrow$ preserves correlations]



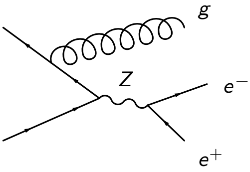
Example I: Z pole

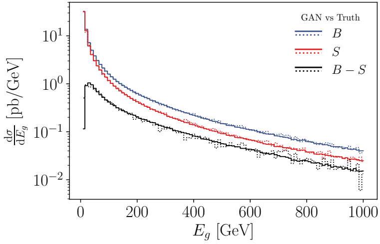
- Training data:
  - $pp \to e^+e^-$
  - $pp \to \gamma \to e^+e^-$
  - 1 M events per dataset, MadGraph5
- Generated events: Z-Pole + interference

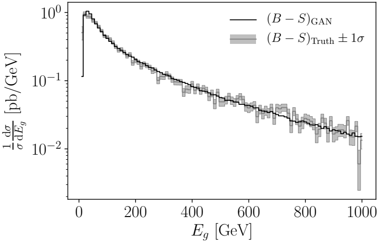


Example II: Dipole subtraction

- Theory uncertainties → limiting factor for HL-LHC
- Higher order: Subtract diverging Catany Seymour Dipole from real emission term
- 1 M events per dataset, SHERPA







How to GAN away detector effects

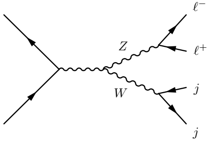
Idea: invert Markov process [1912.00477]

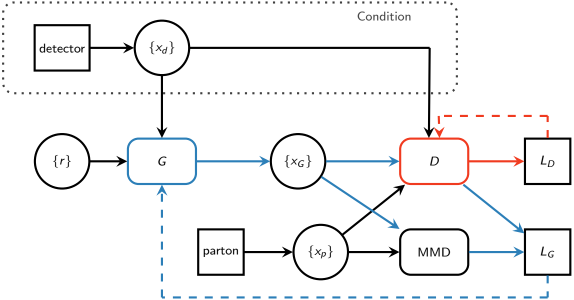
Detector simulation

- Typical Markov process
- Prior dependent inversion possible [Datta et al.]
- Aim: unfolding multidimensional phase space

Reconstruct parton level $pp \to ZW \to (ll)(jj)$

- GAN: no connection between input and discr.
  $\to$ use **fully conditional** GAN (FCGAN)



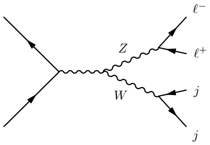


How to GAN away detector effects

Idea: invert Markov process [1912.00477]

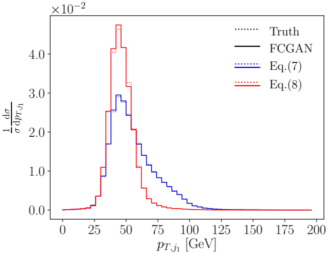
Reconstruct parton level $pp \to ZW \to (ll)(jj)$

- Use **fully conditional** GAN (FCGAN)
- Inversion works ✓



$$\text{Eq.}(7): \quad p_{T,j_1} = 30 \ldots 100 \text{ GeV} \quad (\sim 88\%)$$

$$\text{Eq.}(8): \quad p_{T,j_1} = 30 \ldots 60 \text{ GeV} \quad \text{and} \quad p_{T,j_2} = 30 \ldots 50 \text{ GeV} \quad (\sim 38\%)$$

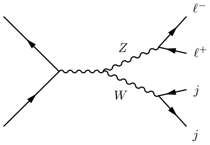


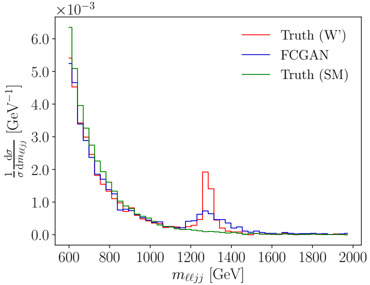
How to GAN away detector effects

Idea: invert Markov process [1912.00477]

Reconstruct parton level $pp \to ZW \to (ll)(jj)$

- Use **fully conditional** GAN (FCGAN)
- Inversion works ✓
- BSM injection ✓
  - train: SM events
  - test: 10% events with W' in s-channel





Summary

- GANs can learn underlying distributions from event samples
- MMD improves performance for special features
- Possibilities to stabilize GAN training: gradient penalty, WGAN-GP, LSGAN,...

- Successful sample based subtraction implemented
- Applications: background subtraction, soft-collinear subtraction, . . .

- Unfold high-dimensional detector level distributions with GAN
- Employ FCGAN for notion of locality to enable meaningful slicing