How to GAN LHC Events

Ramon Winterhalder

ITP Heidelberg

Max-Planck-Institut für Kernphysik
Universität Heidelberg Zukunft seit 1386
NYU
International Max Planck Research School for Precision Tests of Fundamental Symmetries

based on arXiv:1907.03764
with Anja Butter and Tilman Plehn

Monte Carlo Simulations

Monte Carlo simulations crucial for any LHC analysis

Monte Carlo Simulations

Monte Carlo simulations crucial for any LHC analysis

Problem: high-dimensionality and sharp phase-space structures

$\rightarrow$ computationally time consuming

Monte Carlo Simulations

Monte Carlo simulations crucial for any LHC analysis

Problem: high-dimensionality and sharp phase-space structures

$\rightarrow$ computationally time consuming

How to generate events more efficiently? $\rightarrow$ GAN!

Monte Carlo Simulations

Monte Carlo simulations crucial for any LHC analysis

Problem: high-dimensionality and sharp phase-space structures

$\rightarrow$ computationally time consuming

How to generate events more efficiently? $\rightarrow$ GAN!

GANs already used in many physics applications.

Monte Carlo Simulations

Monte Carlo simulations crucial for any LHC analysis

Problem: high-dimensionality and sharp phase-space structures

$\rightarrow$ computationally time consuming

How to generate events more efficiently? $\rightarrow$ GAN!

GANs already used in many physics applications.

- Jet Images - de Oliveira et al. [1701.05927], Carazza et al. [1909.01359],
- Calorimeters - Paganini et al. [CaloGAN, 1705.02355, 1712.10321], Musella et al. [arXiv:1805.00850], Erdmann et al. [arXiv:1807.01954], ATLAS [ATL-SOFT-PUB-2018-001, ATL-SOFT-PROC-2019-007]
- Event generation - Otten et al. [1901.00875], Hashemi et al. [1901.05282], Di Sipio et al. [1903.02433], Martinez et al. [1912.02748]
- Unfolding - Datta et al. [1806.00433], Bellagente et al. [1912.0047]
- EFT models - Erbin et al. [1809.02612]
- Mass templates - Lin et al. [1903.02556]
- Event subtraction - Butter et al. [1912.08824]

GAN Approach

Task: generate events with a neural network (generator)

Require: direct comparison to data $\rightarrow$ unweighted events

Problem: in standard MC: unweighting algorithm needed $\rightarrow$ inefficient

GAN Approach

Task: generate events with a neural network (generator)

Require: direct comparison to data → unweighted events

Problem: in standard MC: unweighting algorithm needed → inefficient

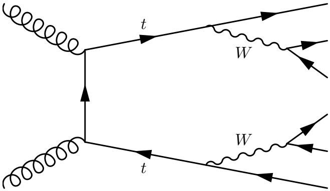
Solution: GAN

- Input: random numbers
- Output: unweighted events
- Training data:
  - unweighted MC events or real data
  - can include parton showers, hadronization and detector effects

Top-Pair Production

GAN events for the $2 \to 6$ particle production process

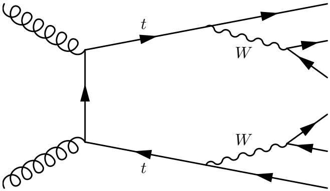
$$pp \to t\bar{t} \to (bW^-)\ (\bar{b}W^+) \to (bq_1\bar{q}_1')\ (\bar{b}q_2\bar{q}_2')\ .$$



Top-Pair Production

GAN events for the $2 \to 6$ particle production process

$$pp \to t\bar{t} \to (bW^-)\ (\bar{b}W^+) \to (bq_1\bar{q}_1')\ (\bar{b}q_2\bar{q}_2') \ .$$

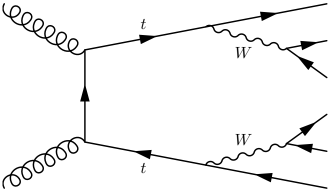


Challenges: 16-dimensional phase-space, 4 resonances, phase-space boundaries, tails

Top-Pair Production

GAN events for the $2 \to 6$ particle production process

$$pp \to t\bar{t} \to (bW^-)\ (\bar{b}W^+) \to (bq_1\bar{q}_1')\ (\bar{b}q_2\bar{q}_2')\ .$$

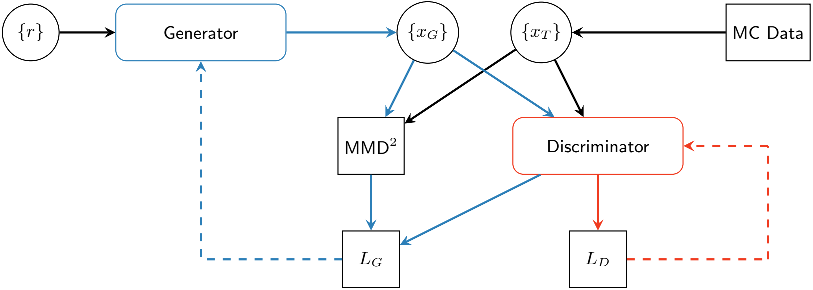


Challenges: 16-dimensional phase-space, 4 resonances, phase-space boundaries, tails

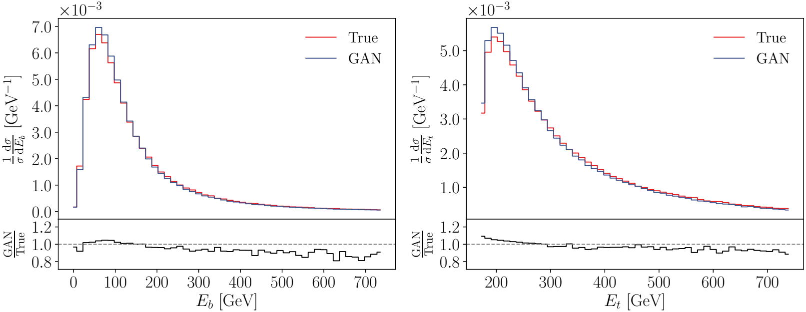
Comments: fix masses of final state particles
$\rightarrow$ generate 18 dim output

additional loss focusing on phase-space structures
$\rightarrow$ MMD Loss

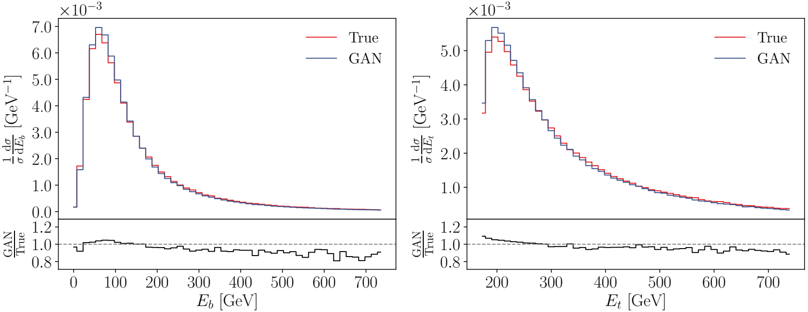
GAN Workflow



Energy Distributions

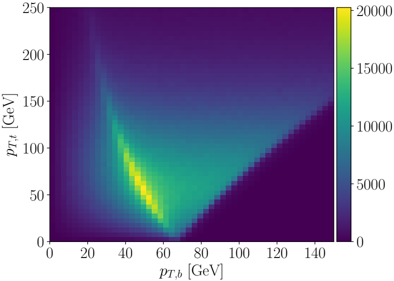


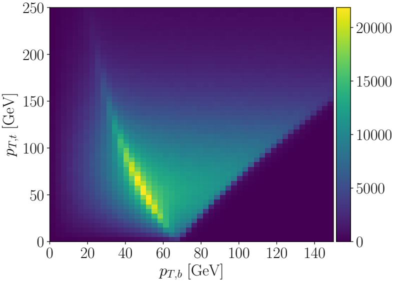
Energy Distributions



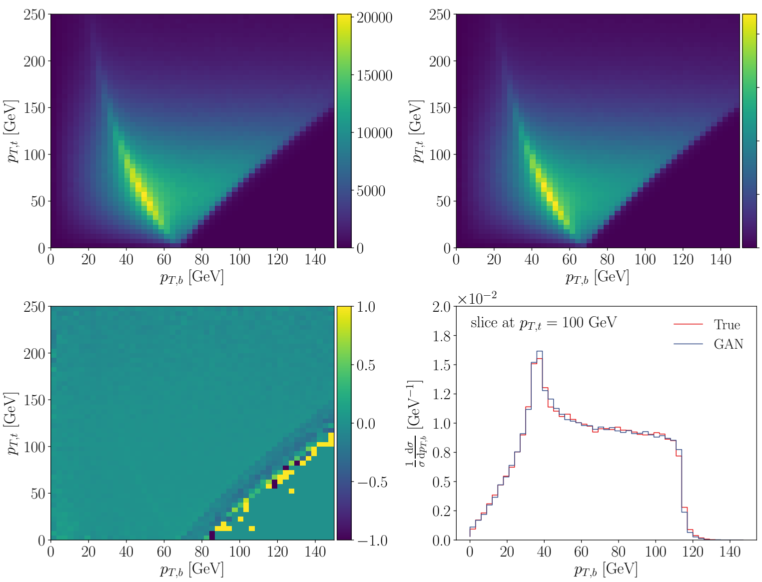
$\rightarrow$ flat distributions easy to learn!

2-dimensional Correlations





2-dimensional Correlations

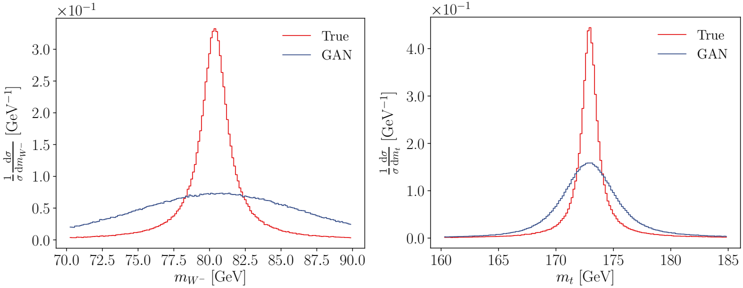


Invariant Mass Peaks

What about the resonances?

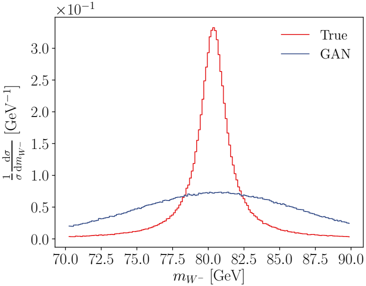
Invariant Mass Peaks

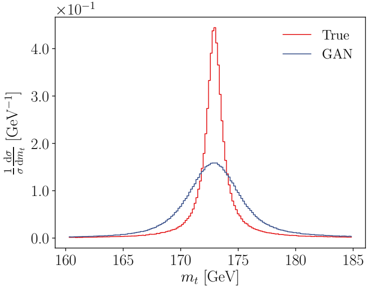
Without the additional loss:



Invariant Mass Peaks

Without the additional loss:

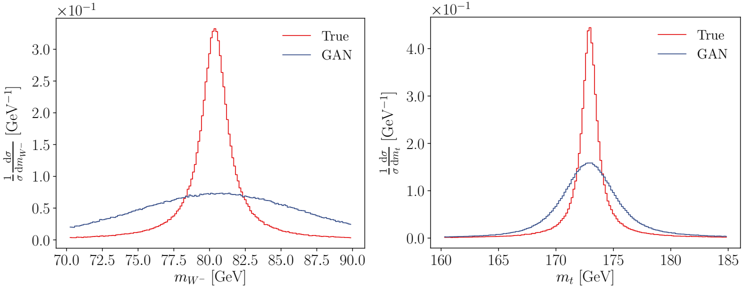




Challenge: resolve the mass peaks

Invariant Mass Peaks

Without the additional loss:



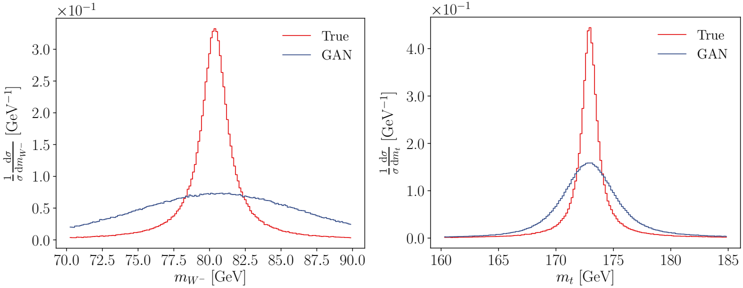
Challenge: resolve the mass peaks

Standard solution: phase-space remapping

$$\int \mathrm{d}s \frac{F(s)}{(s-m^2)^2+m^2\Gamma^2} = \frac{1}{m\Gamma}\int \mathrm{d}z\ F(s) \quad \text{with} \quad z = \arctan\frac{s-m^2}{m\Gamma}\ .$$

Invariant Mass Peaks

Without the additional loss:



Challenge: resolve the mass peaks

Standard solution: phase-space remapping

$$\int ds \frac{F(s)}{(s-m^2)^2 + m^2\Gamma^2} = \frac{1}{m\Gamma}\int dz\, F(s) \quad \text{with} \quad z = \arctan\frac{s-m^2}{m\Gamma}\,.$$

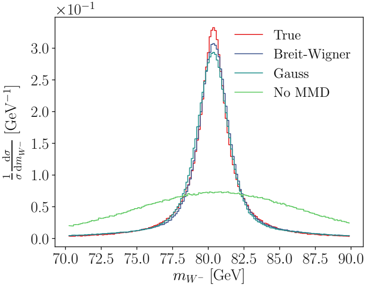
However: knowledge of m and Γ needed

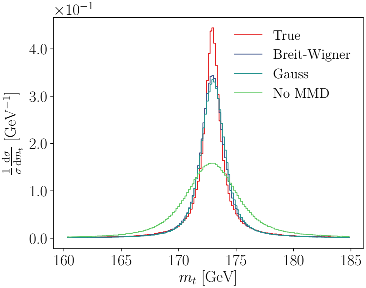
Invariant Mass Peaks

Can we do it better?

Invariant Mass Peaks

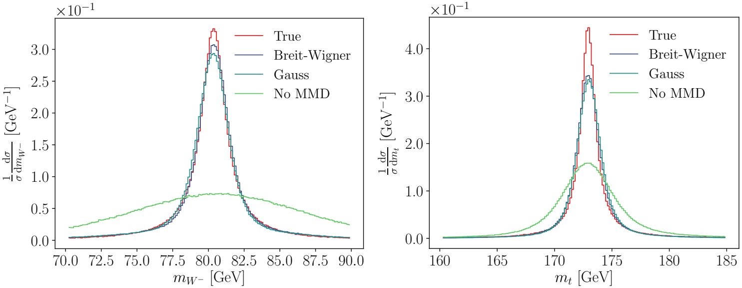
Including the MMD Loss





Invariant Mass Peaks

Including the MMD Loss



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

- free kernel choice $\rightarrow$ stable results
- **no** knowledge of m and Γ needed

Summary

- The GAN is able to reproduce the full phase space structure of a realistic LHC process
- Flat distributions can be reproduced at arbitrary precison, limited only by statistics
- Using the MMD loss, we can even describe rich peaking resonances
- The same setup will allow us to GAN events from an actual LHC event sample

Appendix

Network Parameters

Parameter	Value
Input dimension G	$18+6$
Layers	10
Units per layer	512
Trainable weights G	2382866
Trainable weights D	2377217
λ_D	10^{-3}
λ_G	1
Batch size	1024
Epochs	1000
Iterations per epoch	1000
Training time	26h
Size of trainings data	10^6

Phase-space coverage