How to GAN away Detector Effects

Marco Bellagente[1], Anja Butter[1], Gregor Kasieczka[2], Tilman Plehn[1], and Ramon Winterhalder[1]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Institut für Experimentalphysik, Universität Hamburg, Germany
bellagente@thphys.uni-heidelberg.de

March 18, 2020

Journal Club 31.03.2020
Janik Andrejkovic

Outline:

- What is a **GAN**?
- Why do we need to correct for **detector effects**?

- Set up the scene: What is studied?

- Naive GAN setup
- Fully conditional GAN (FCGAN) setup

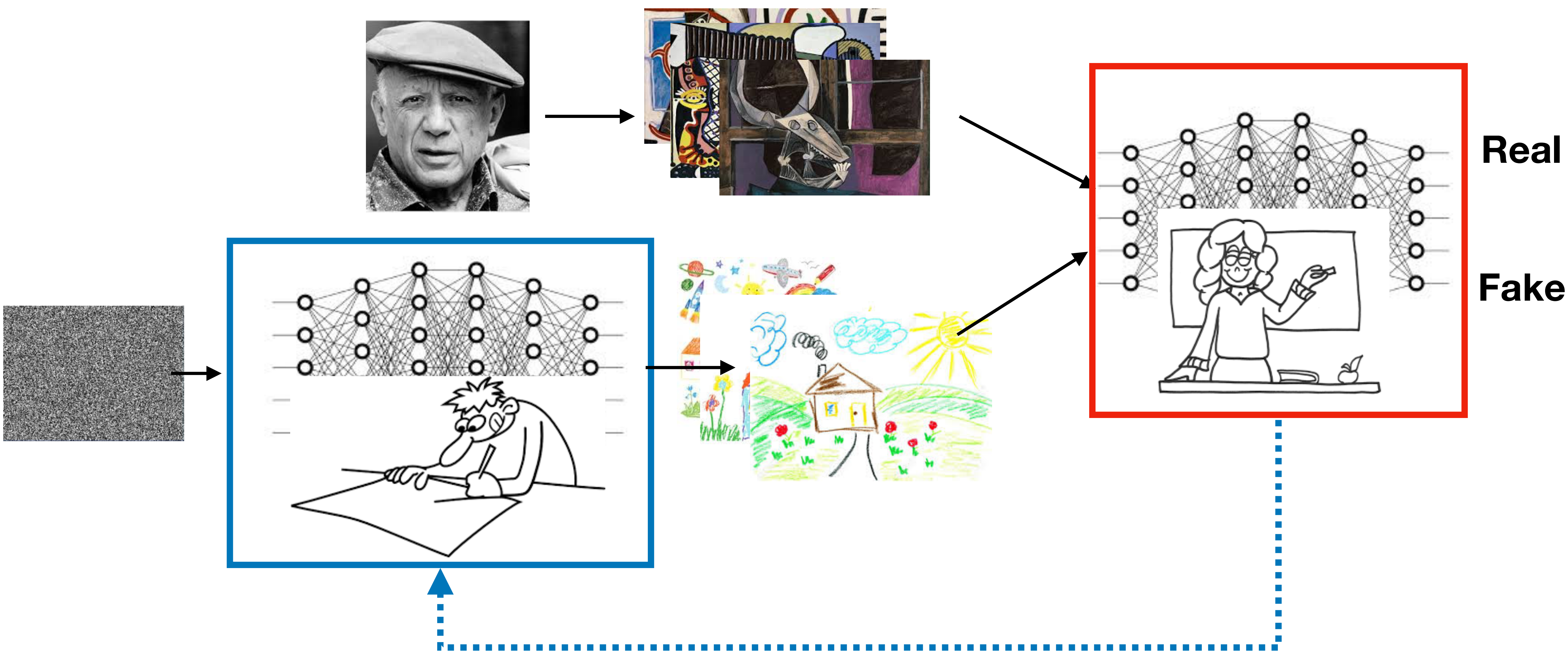
HEPHY
INSTITUT FÜR HOCHENERGIEPHYSIK

What is a GAN?

GAN : **G**enerative **A**dversarial **N**eural network

2 competing neural networks

- a **generator G**
- a **discriminator D**

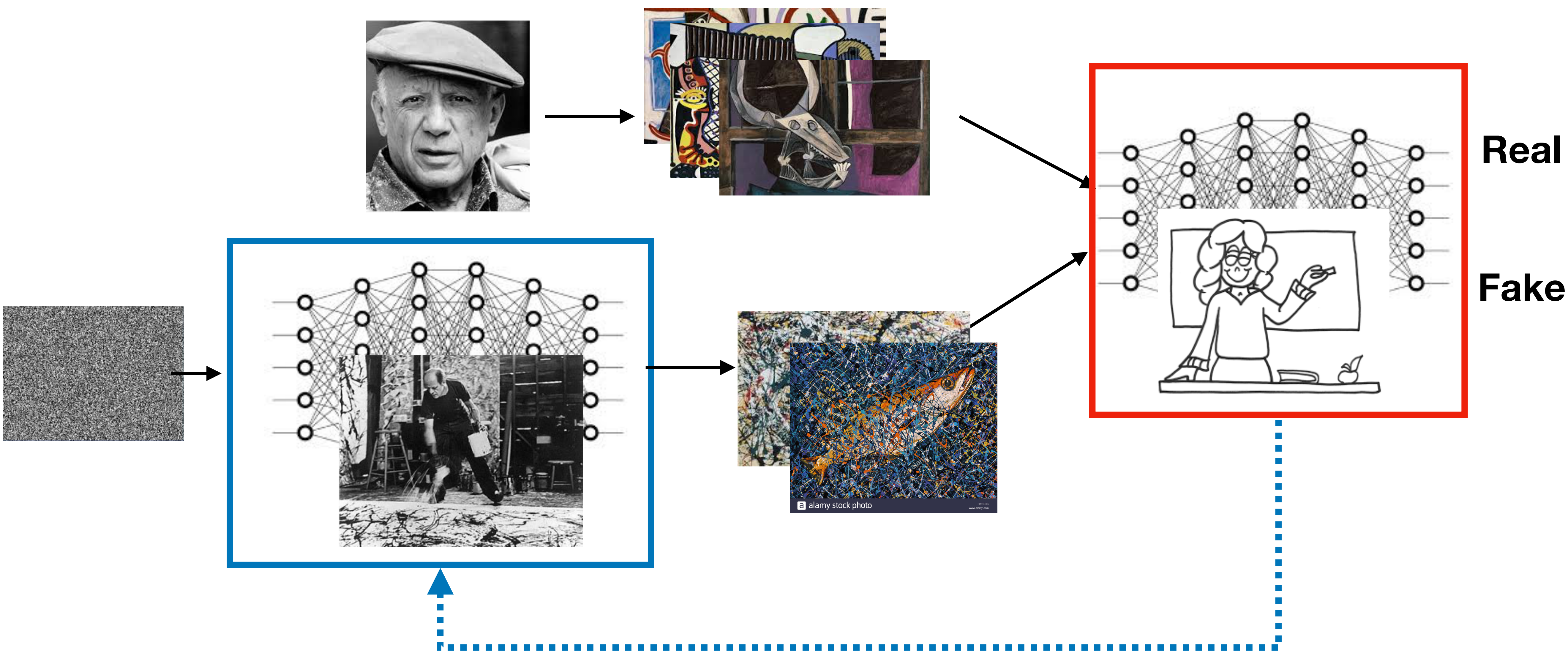


What is a GAN?

GAN : **G**enerative **A**dversarial **N**eural network

2 competing neural networks

- a **generator G**
- a **discriminator D**

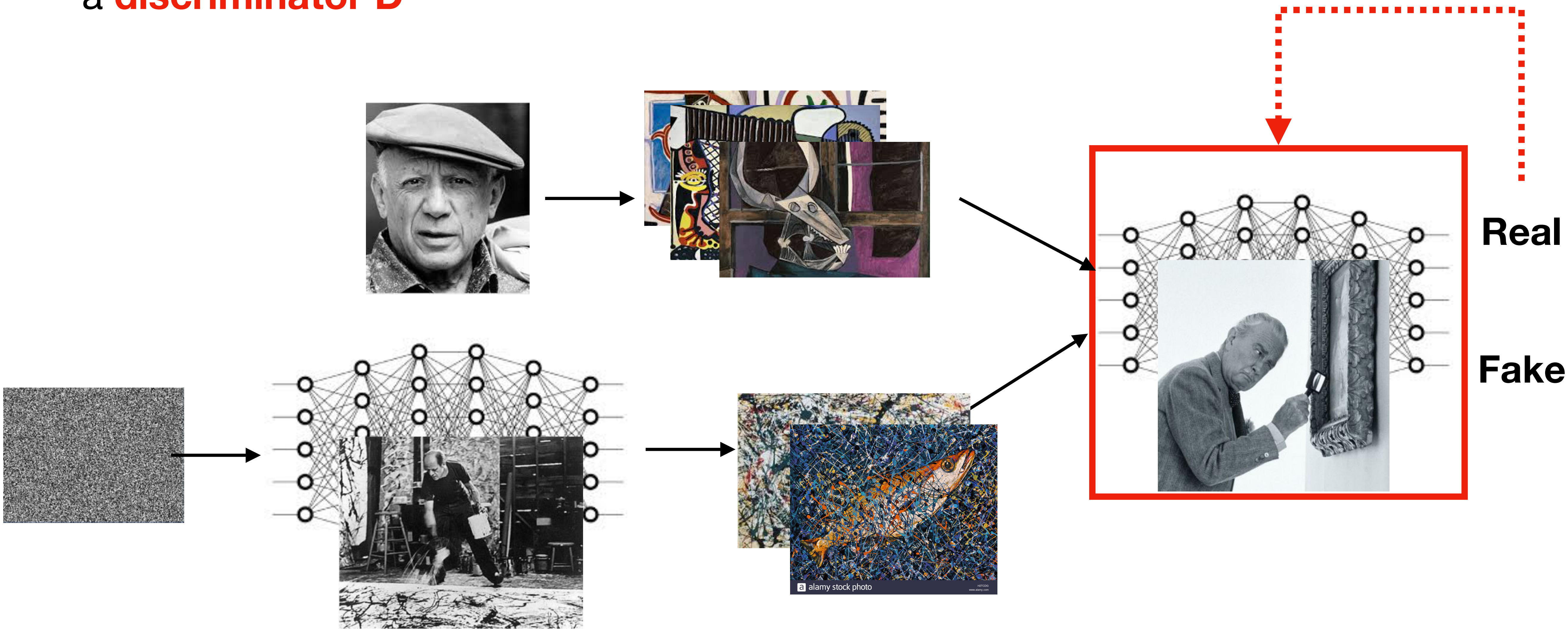


What is a GAN?

GAN : **G**enerative **A**dversarial **N**eural network

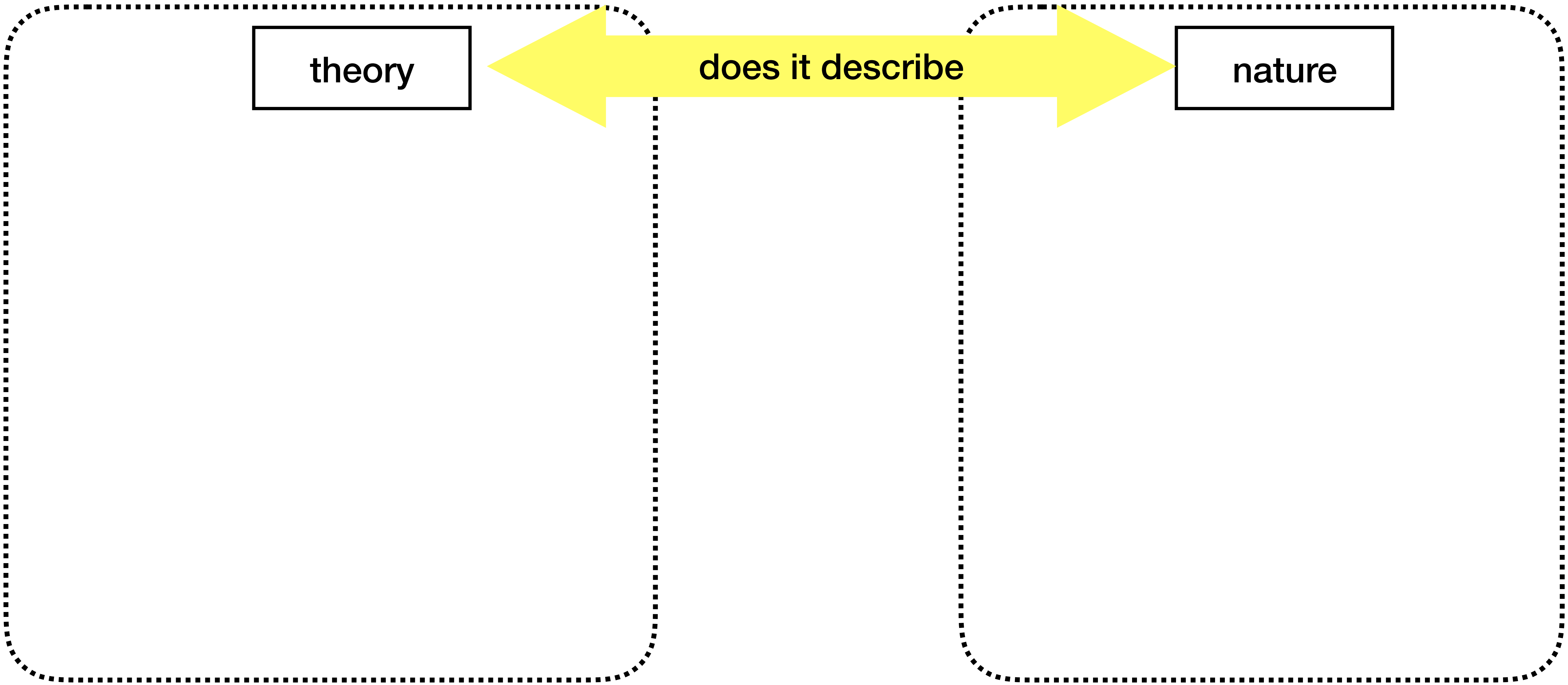
2 competing neural networks

- a **generator G**
- a **discriminator D**

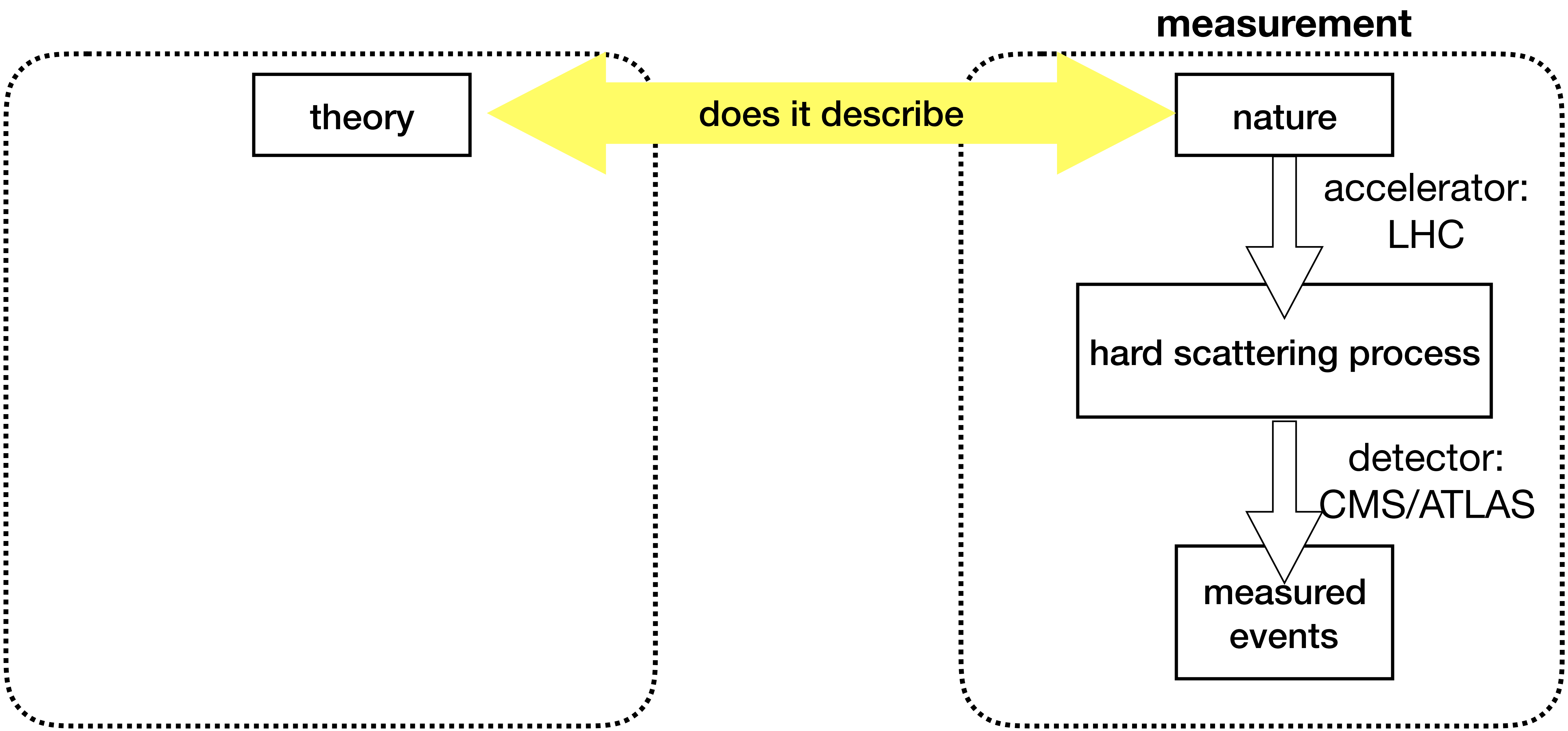


Visit https://www.thispersondoesnotexist.com
GAN creating fake images of people

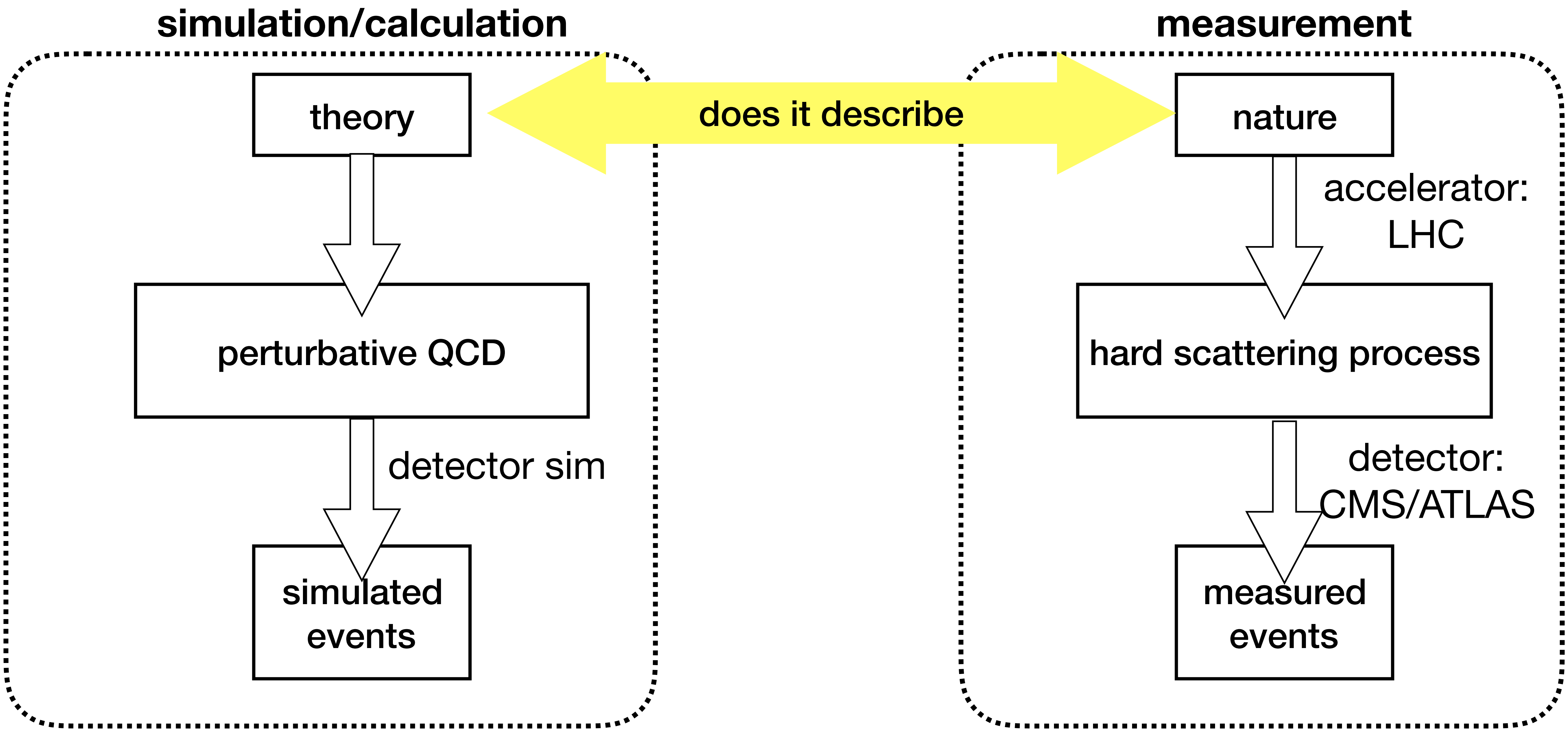
Correcting detector effects



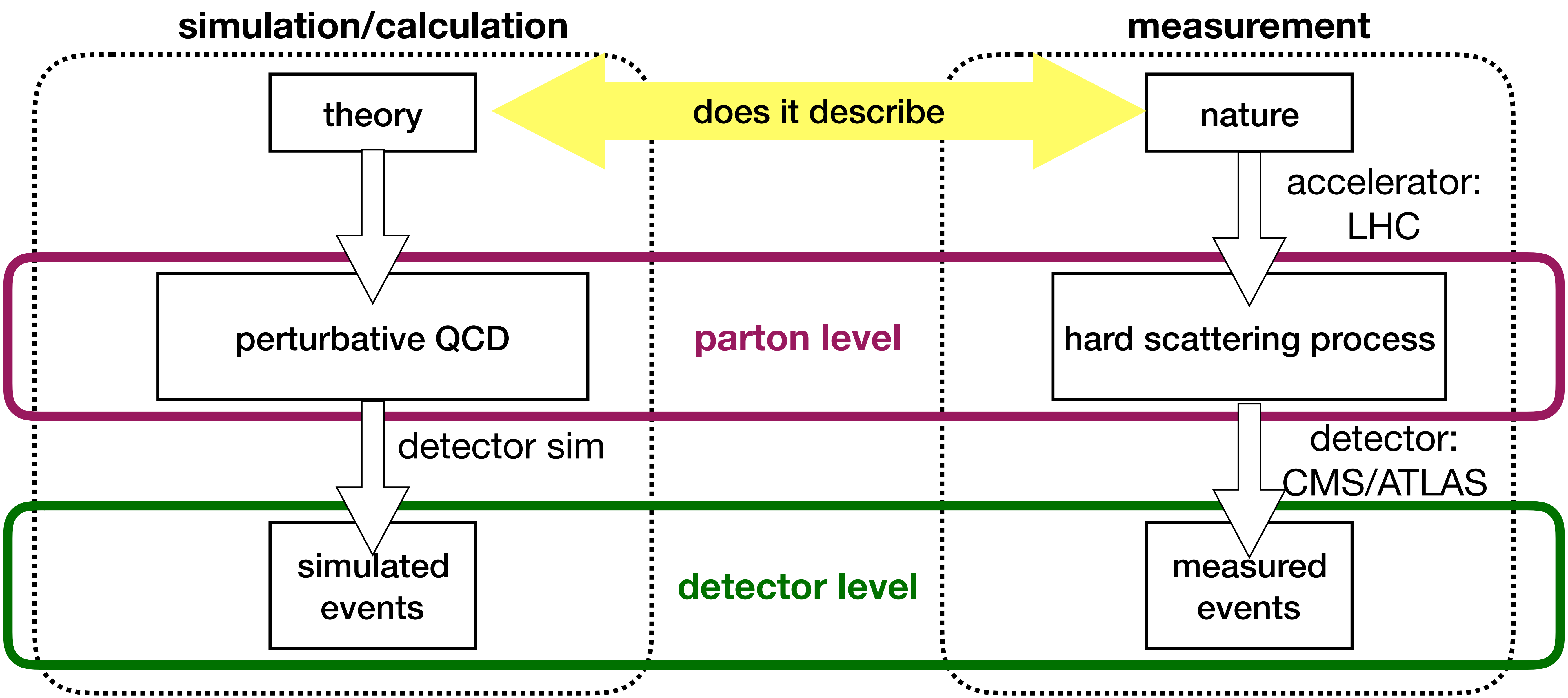
Correcting detector effects



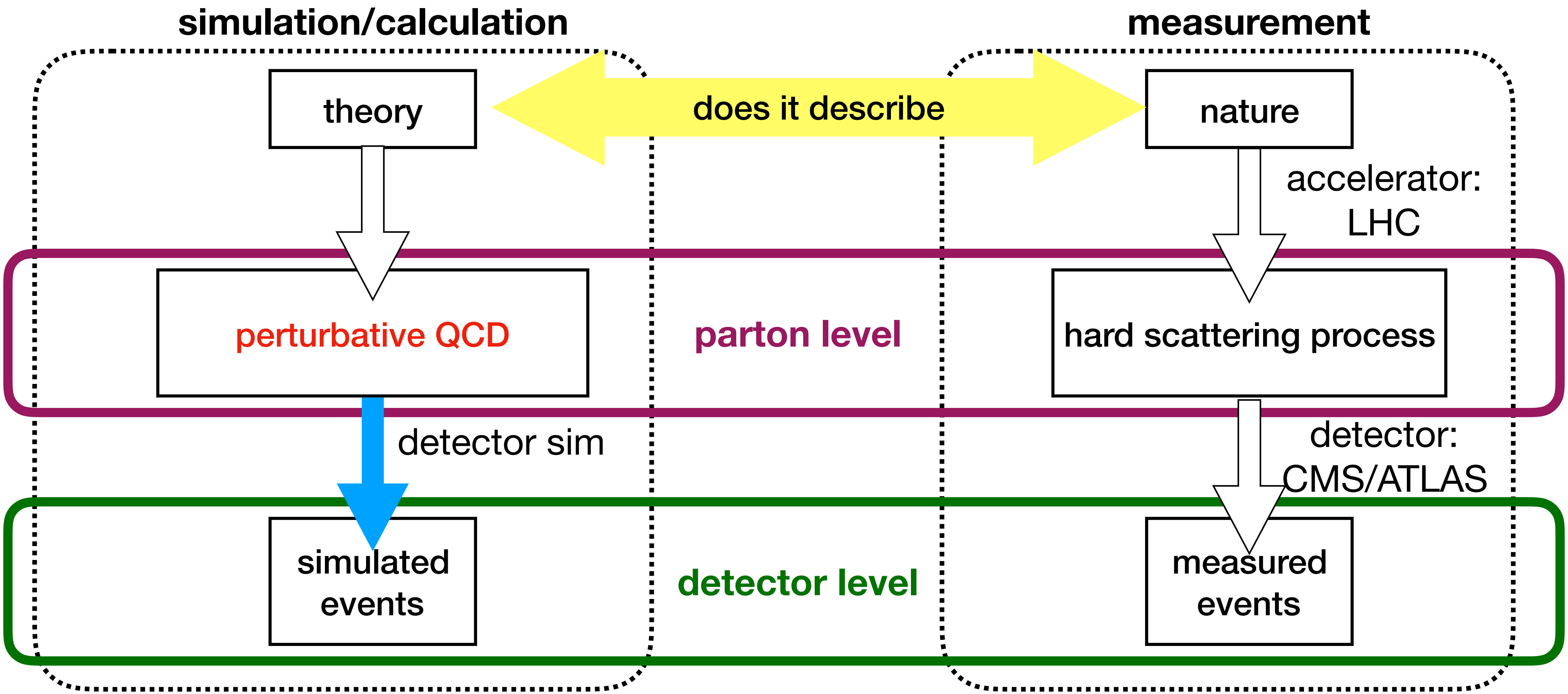
Correcting detector effects



Correcting detector effects



Correcting detector effects

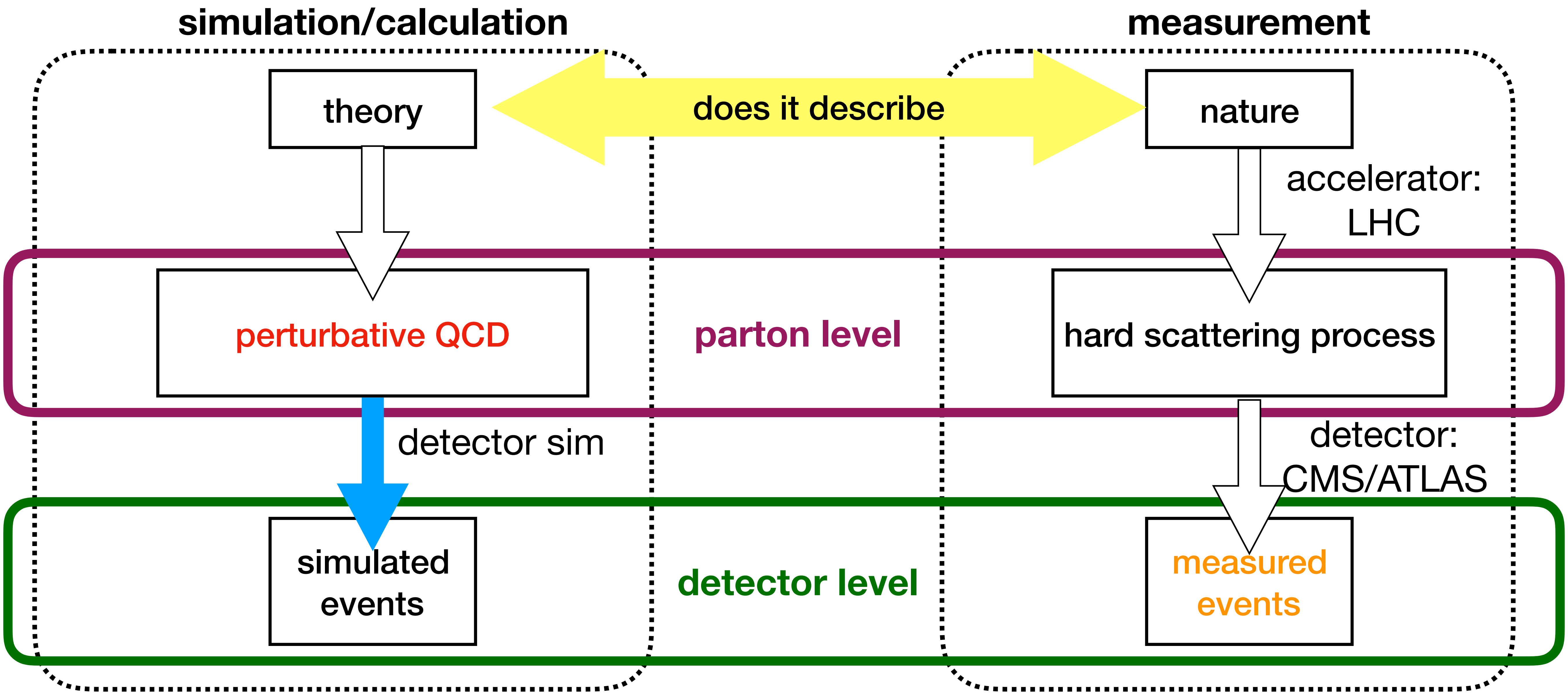


Problem(s):

First-principle predictions enter event simulation as a black-box

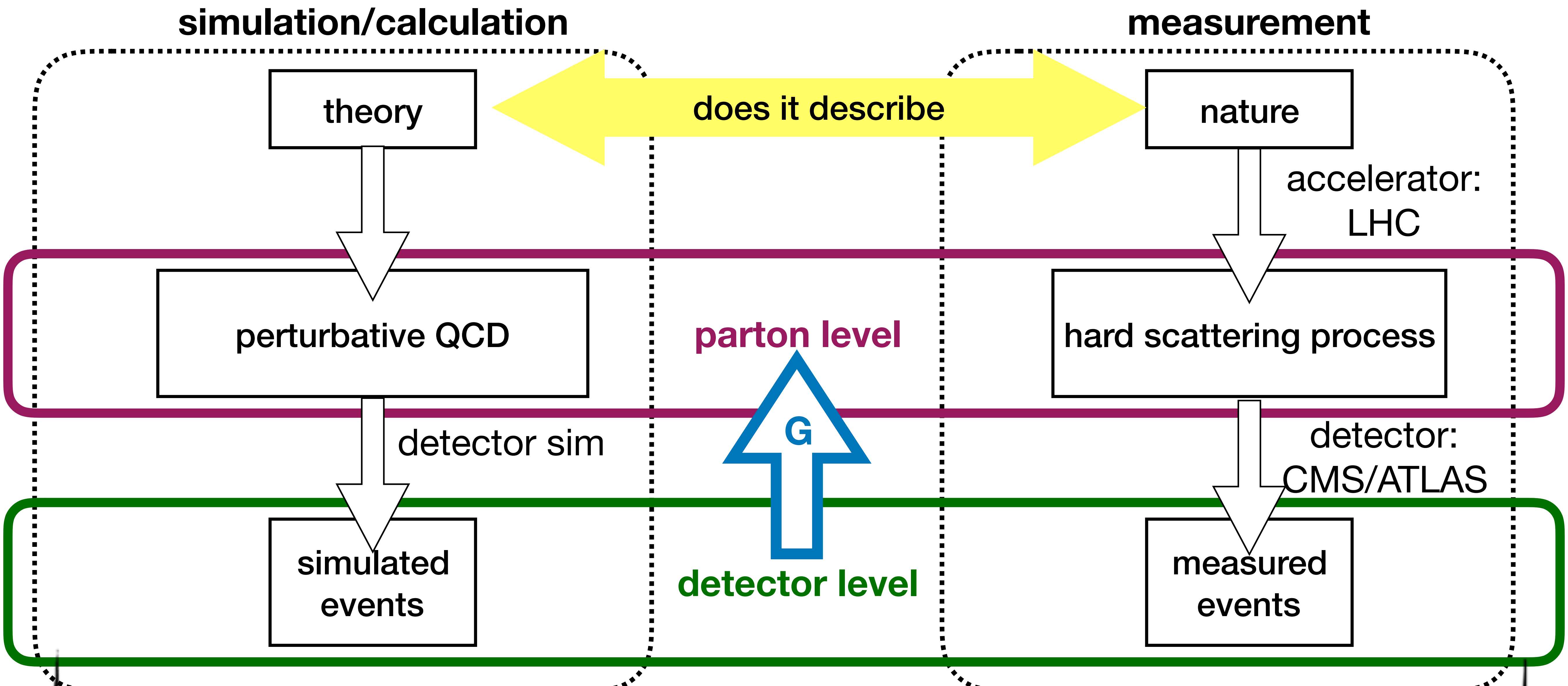
The MC simulation chain can at best be inverted approximatively

Correcting detector effects



Problem(s):
First-principle predictions enter event simulation as a black-box
The MC simulation chain can at best be inverted approximatively
Goal:
Wish to achieve a direct comparison of first-principles QCD predictions with modern LHC measurements at **parton level**

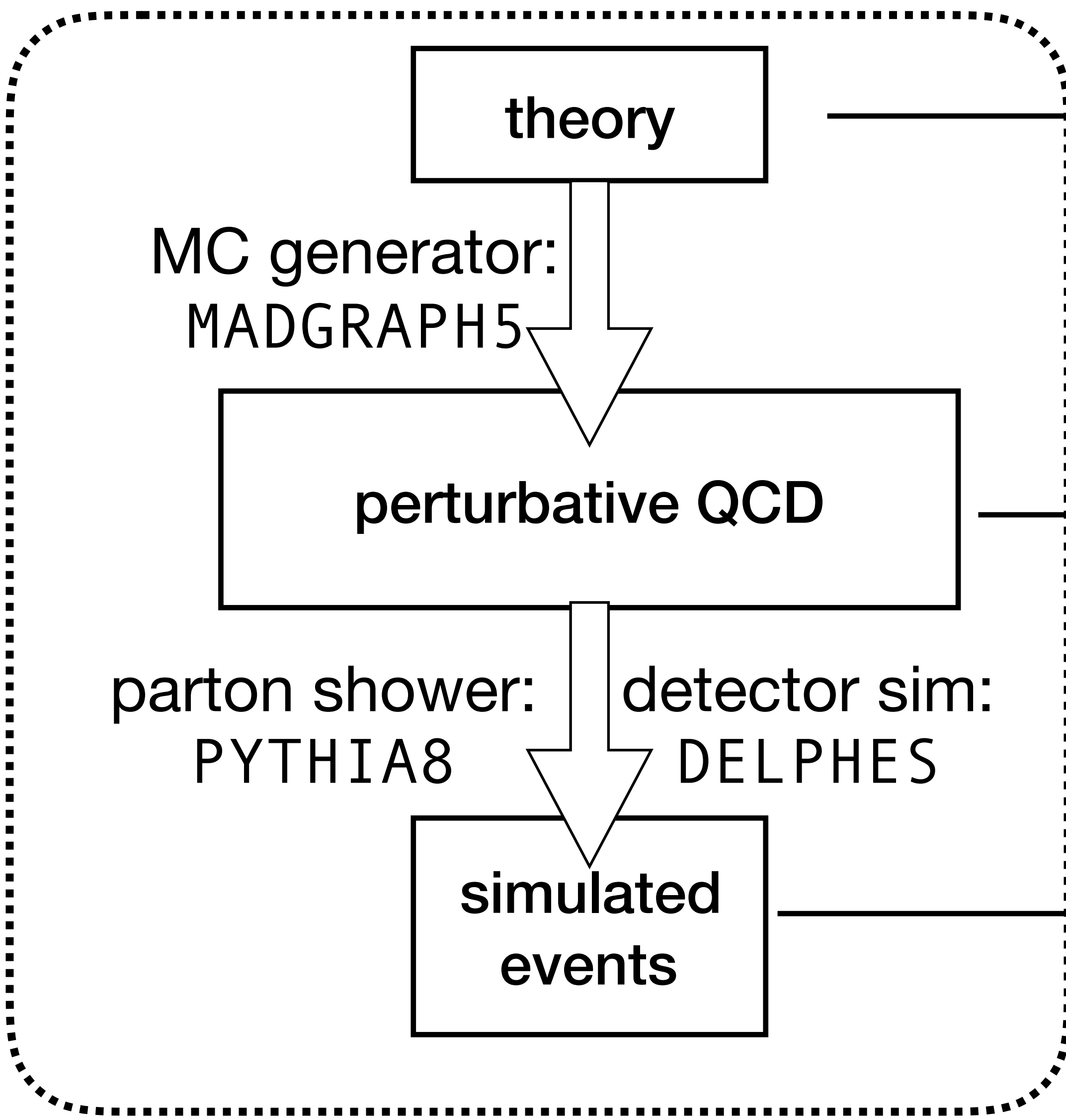
Correcting detector effects

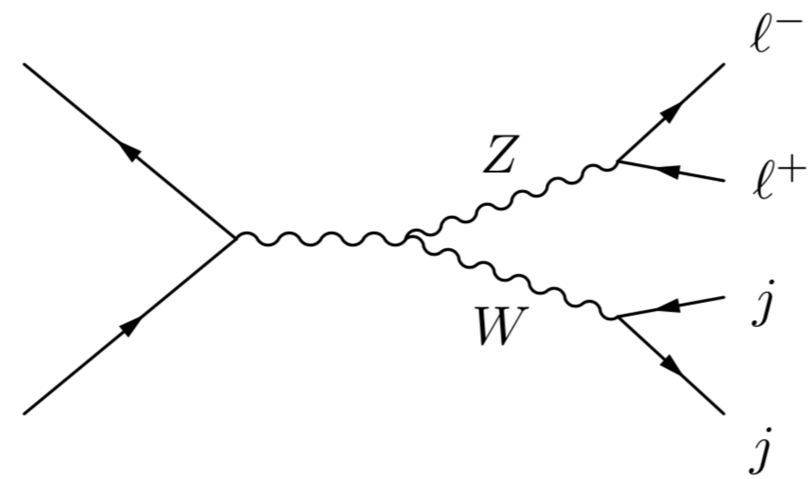


Approach:
Use a GAN, where the **generator G** takes **detector level** inputs and produces fake **parton level** distributions.
The **discriminator D** judges whether a **parton level** event is real or fake.
After training, use the **generator G** for detector unfolding

Set up the scene

simulation/calculation

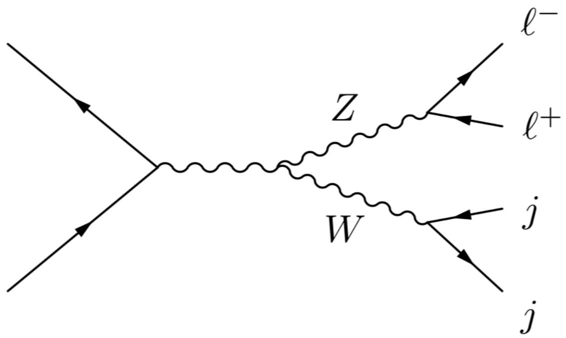




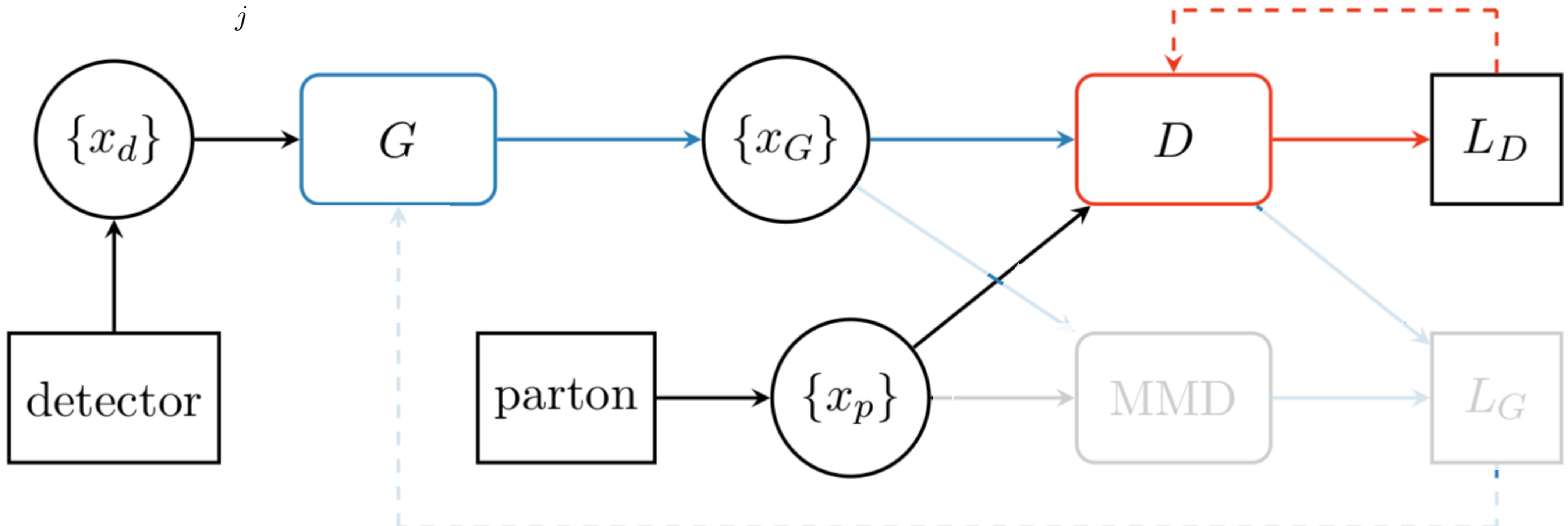
parton level (truth) distribution P_p

detector level (DELPHES) distribution P_d

selection:
2 jets
2 opposite sign electrons
$p_{T,j} > 25\,\mathrm{GeV}$
$|\eta_j| < 2.5$



Naive GAN setup

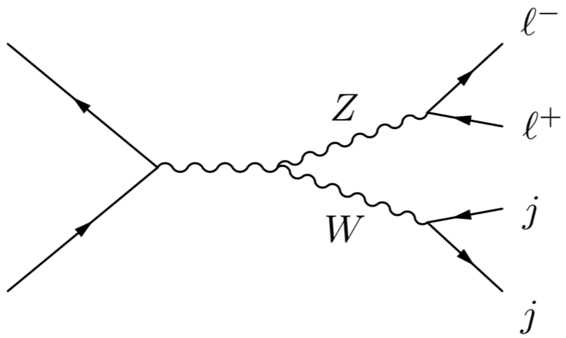


Input x are 4-momenta of final state particles.

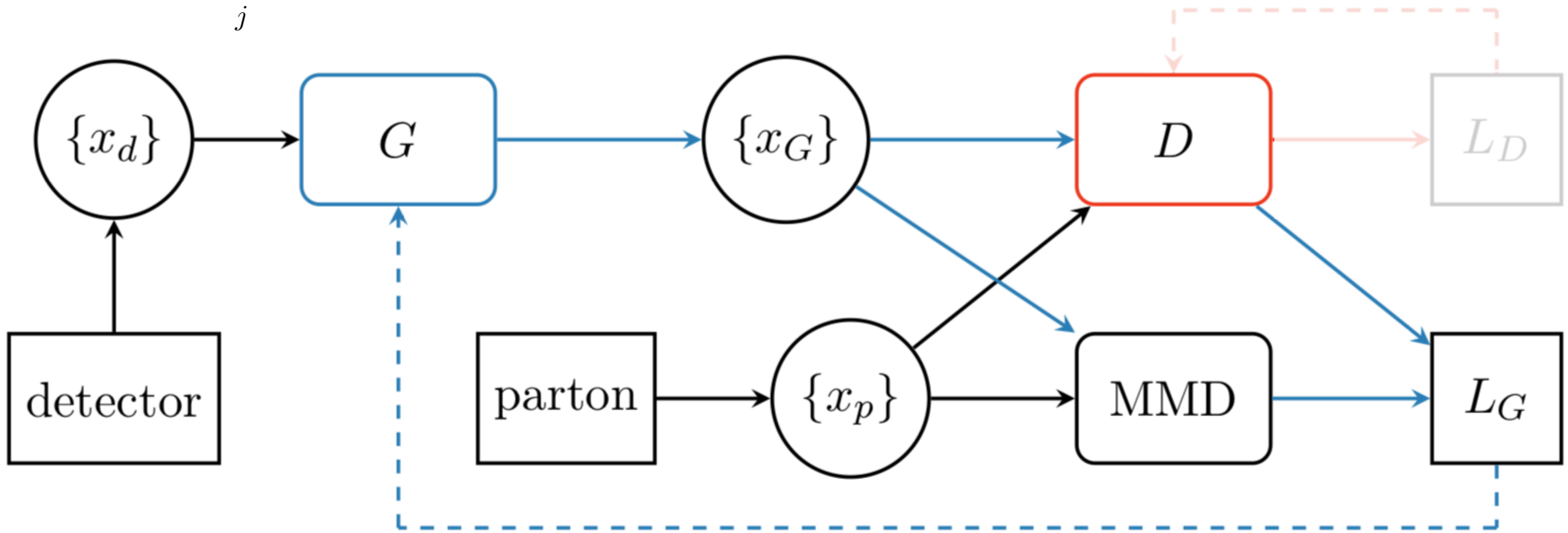
Training of the **discriminator D** follows black and red lines. **D** tries to minimise the loss function L_D

Training of the **generator G** follows black and blue lines. **G** tries to minimise the loss function L_G

Technical detail on MMD = Maximum Mean Discrepancy
MMD is a kernel-based method to compare two samples drawn from different distributions. It is used to help the GAN reproduce the invariant mass distribution of intermediate on-shell particles: arXiv:1907.03764 [hep-ph]



Naive GAN setup

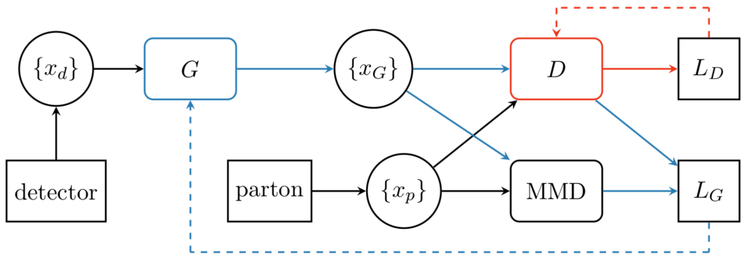


Input x are 4-momenta of final state particles.

Training of the **discriminator D** follows black and red lines. **D** tries to minimise the loss function L_D

Training of the **generator G** follows black and blue lines. **G** tries to minimise the loss function L_G

Technical detail on MMD = Maximum Mean Discrepancy
MMD is a kernel-based method to compare two samples drawn from different distributions. It is used to help the GAN reproduce the invariant mass distribution of intermediate on-shell particles: arXiv: 1907.03764 [hep-ph]

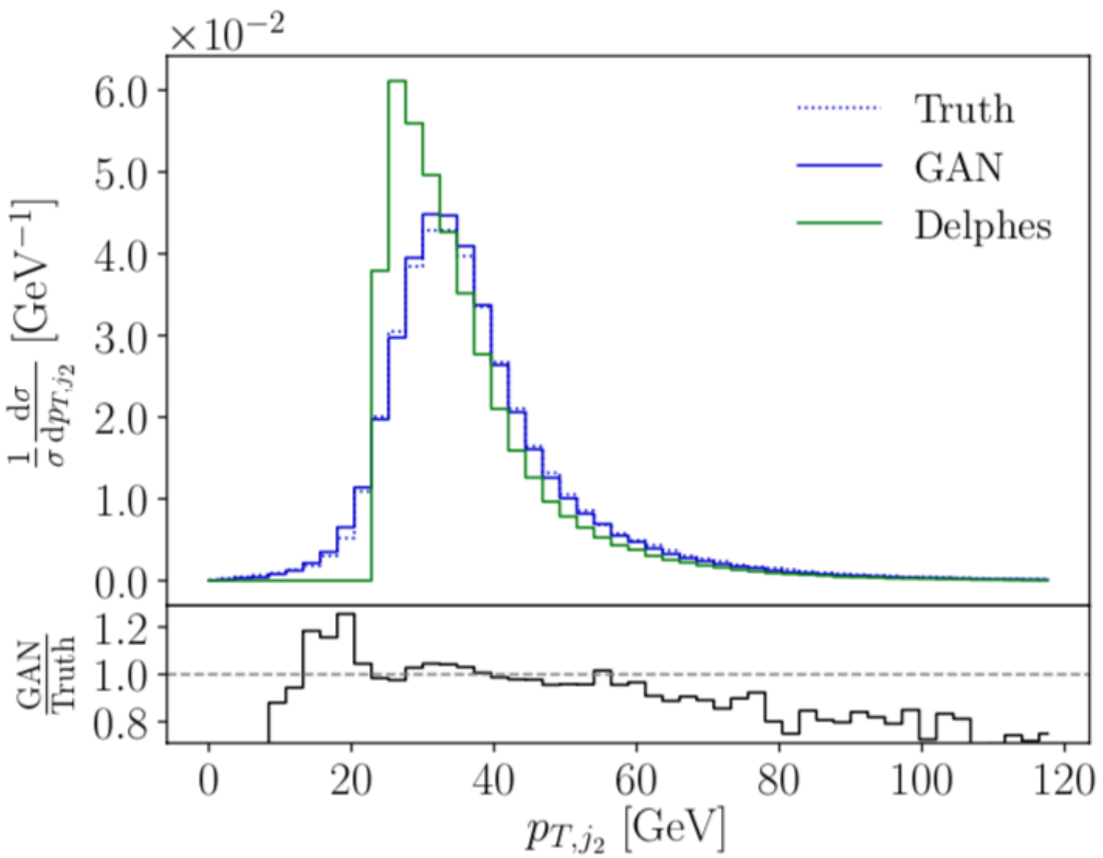


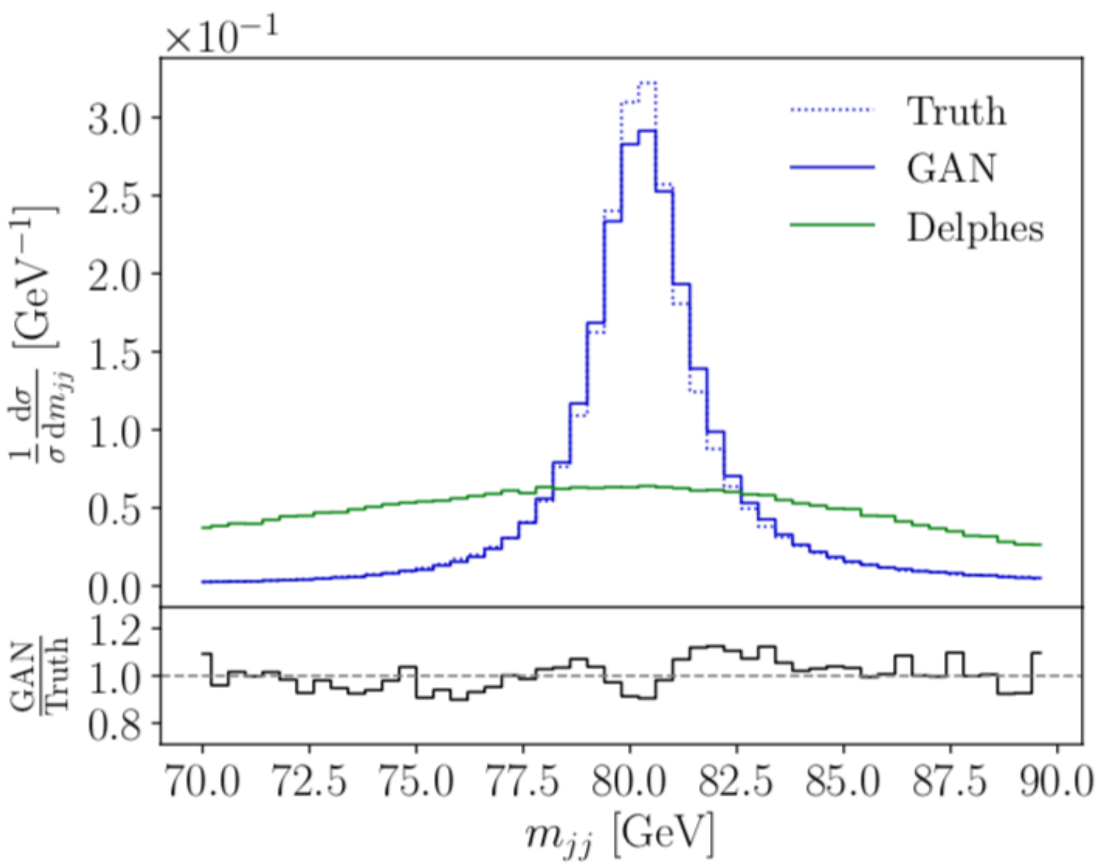
Naive GAN result

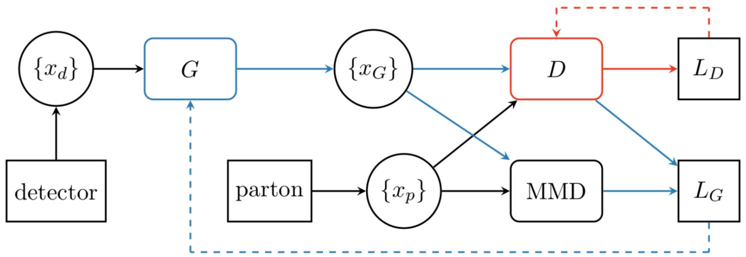
- GAN unfolding runs on statistical independent events but simulation-wise identical
- Statistical inversion of detector effect works well

Advantages
- GAN training does not require a per event parton-detector level matching

Disadvantage
- GAN unfolding is deterministic







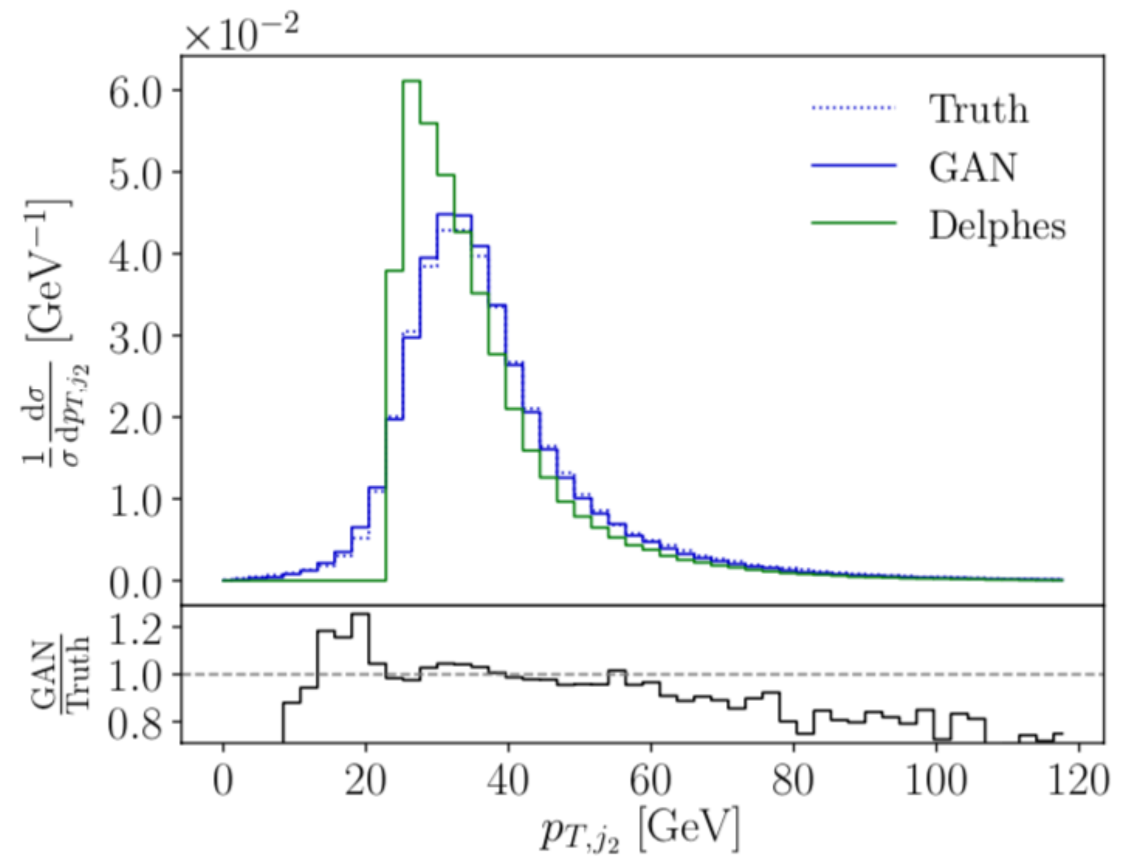
Naive GAN result

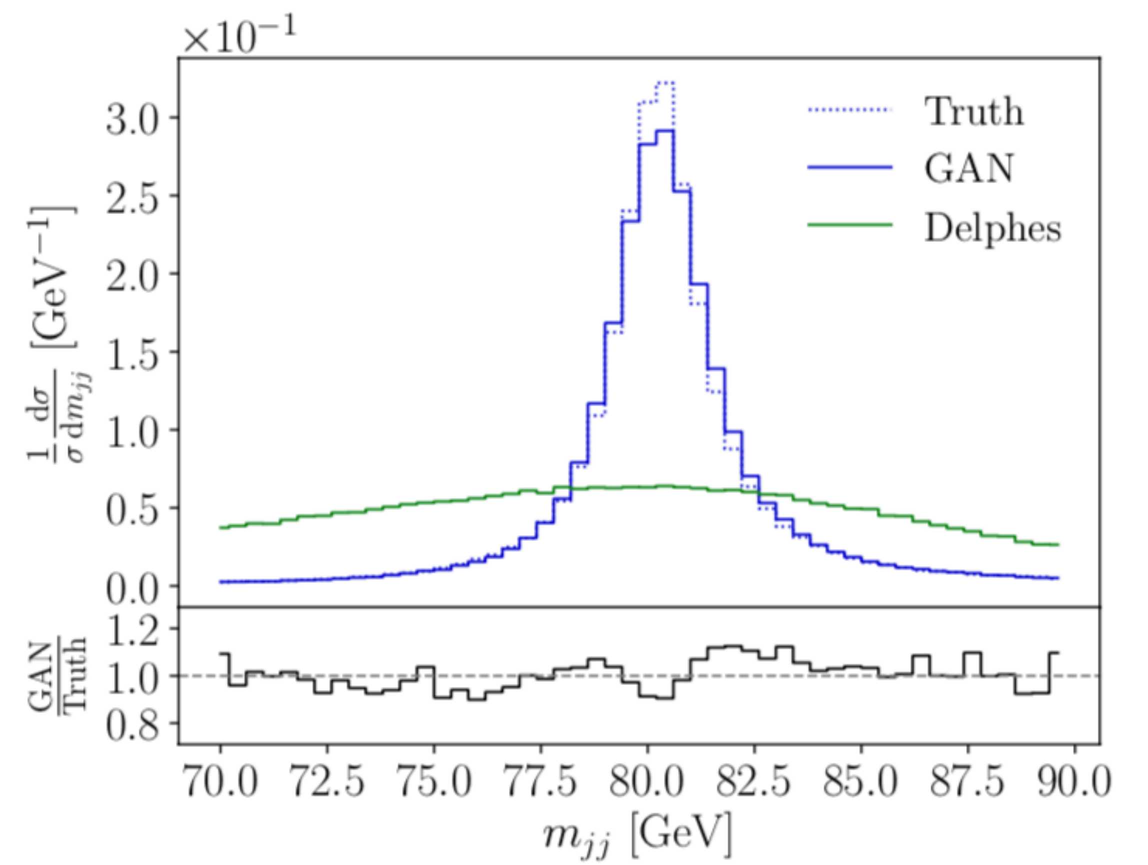
- GAN unfolding runs on statistical independent events but simulation-wise identical
- Statistical inversion of detector effect works well

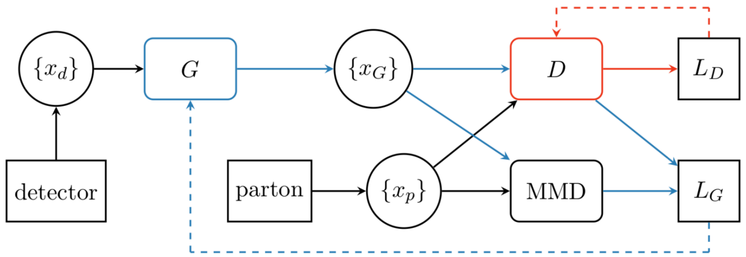
Advantages
- GAN training does not require a per event parton-detector level matching

Disadvantage
- GAN unfolding is deterministic

Question:
- What happens when GAN is used to unfold sample that only covers part of the detector-level phase space used for training
- Will the unfolding work?

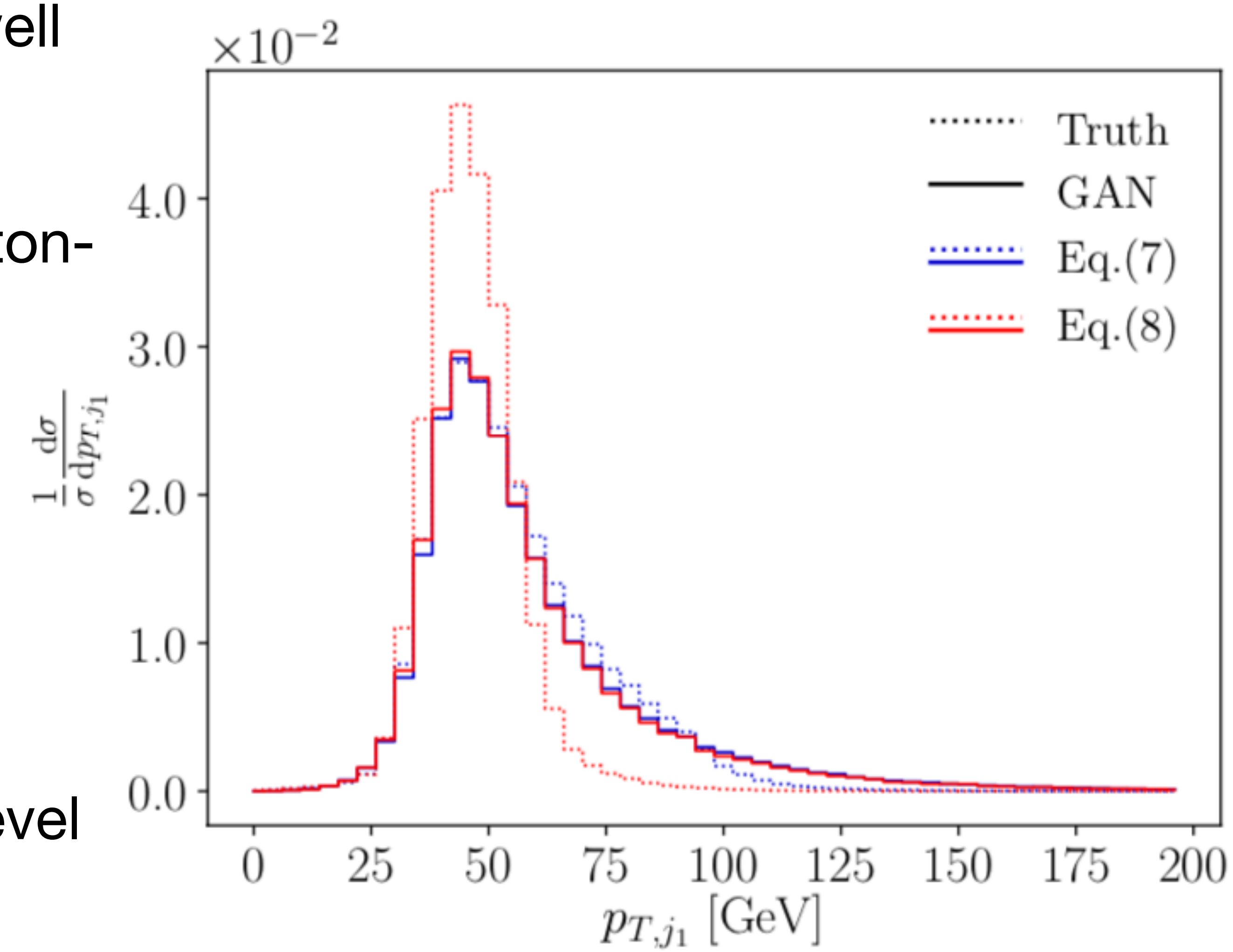






Naive GAN result

- GAN unfolding runs on statistical independent events but simulation-wise identical
- Statistical inversion of detector effect works well



Advantages
- GAN training does not require a per event parton-detector level matching

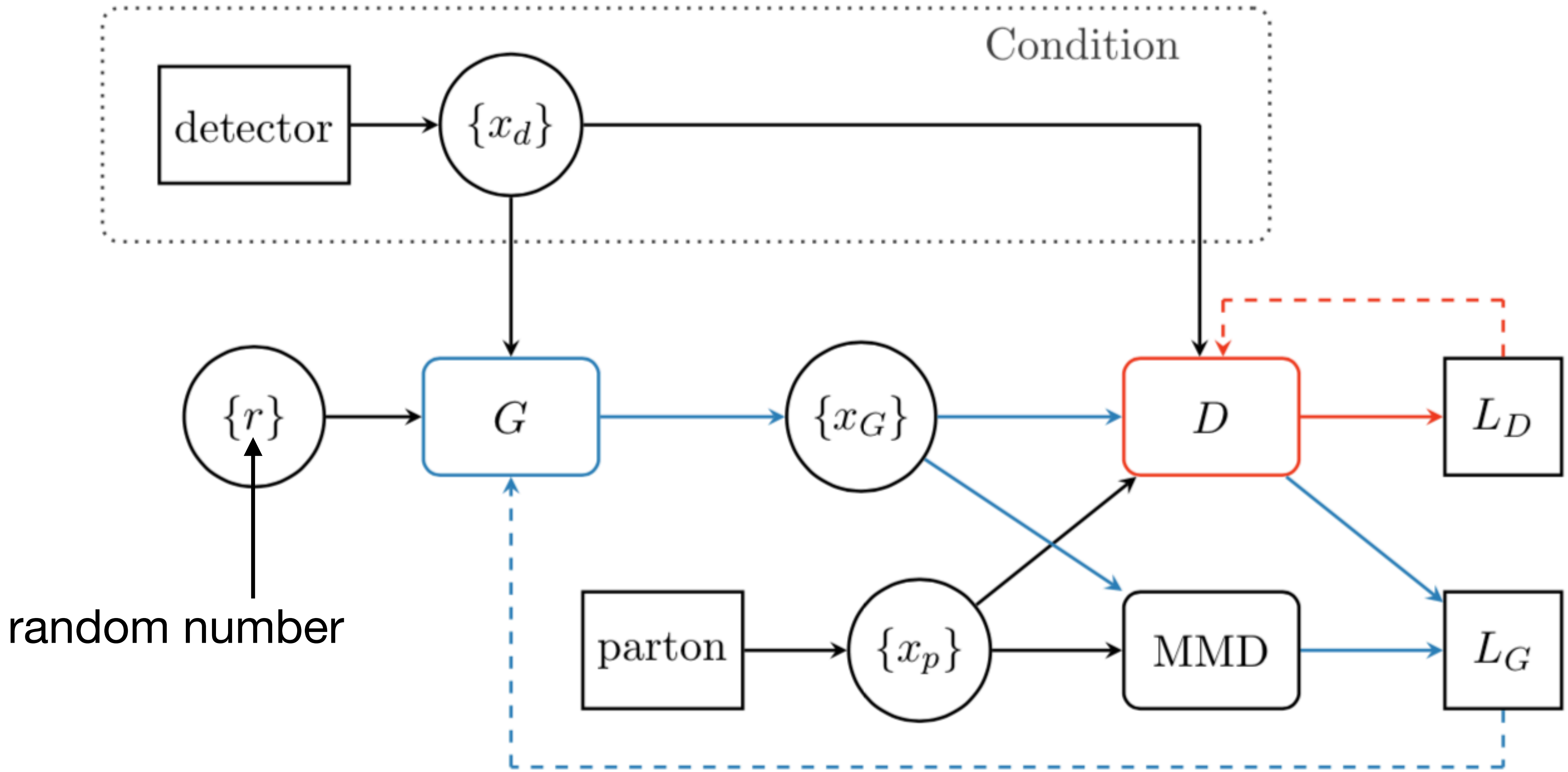
Disadvantage
- GAN unfolding is deterministic

Question:
- What happens when GAN is used to unfold sample that only covers part of the detector-level phase space used for training
- Will the unfolding work?
  - **NO!**

Cut I : $p_{T,j_1} = 30 \dots 100$ GeV (Eq 7 - 88%)

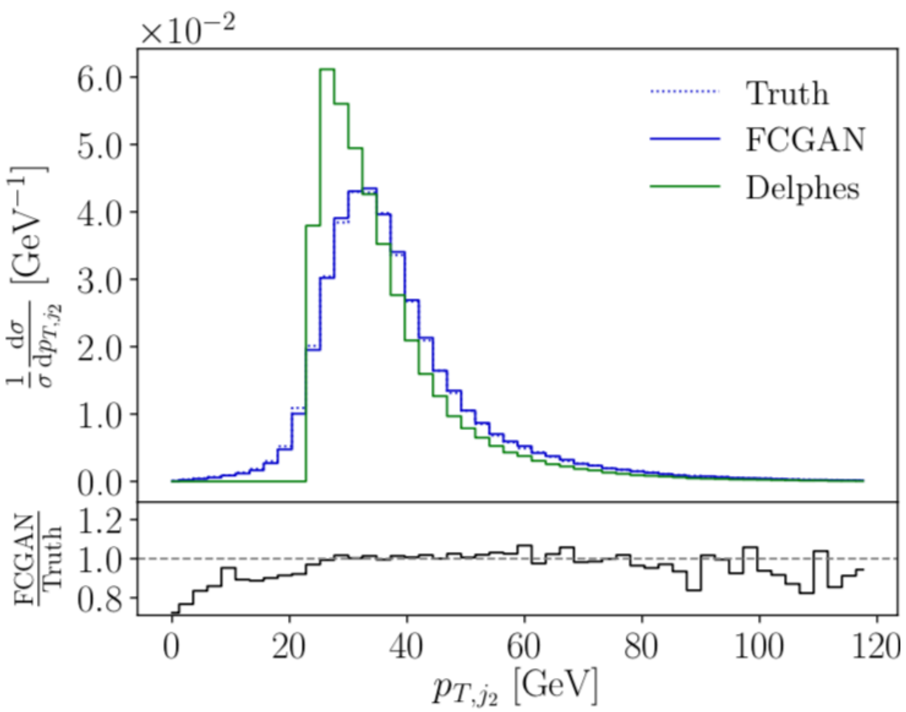
Cut II : $p_{T,j_1} = 30 \dots 60$ GeV and $p_{T,j_2} = 30 \dots 50$ GeV (Eq 8 - 38%)

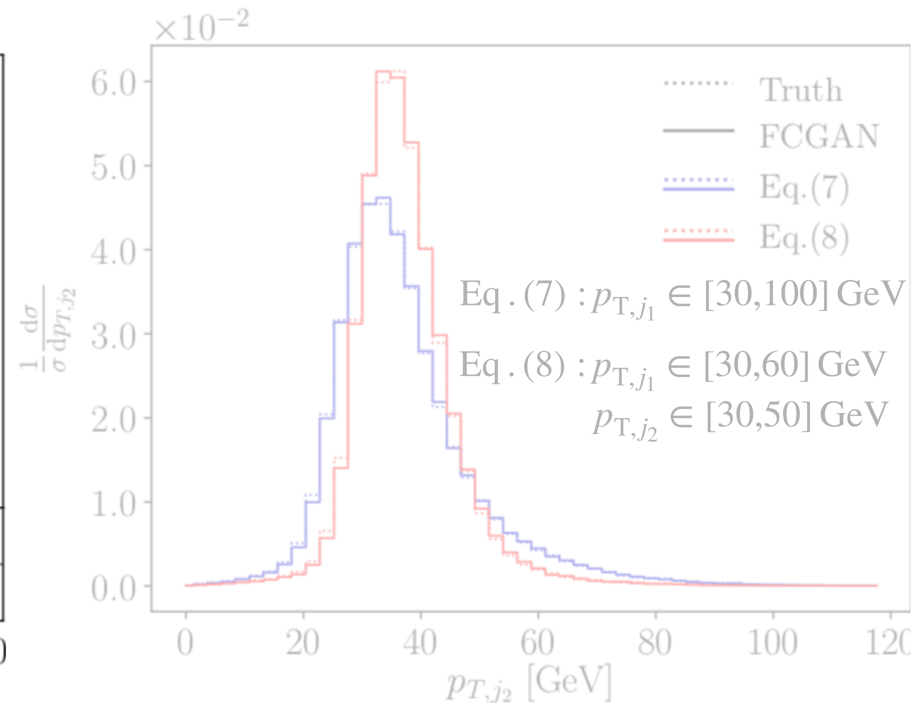
Fully conditional GAN (FCGAN) setup

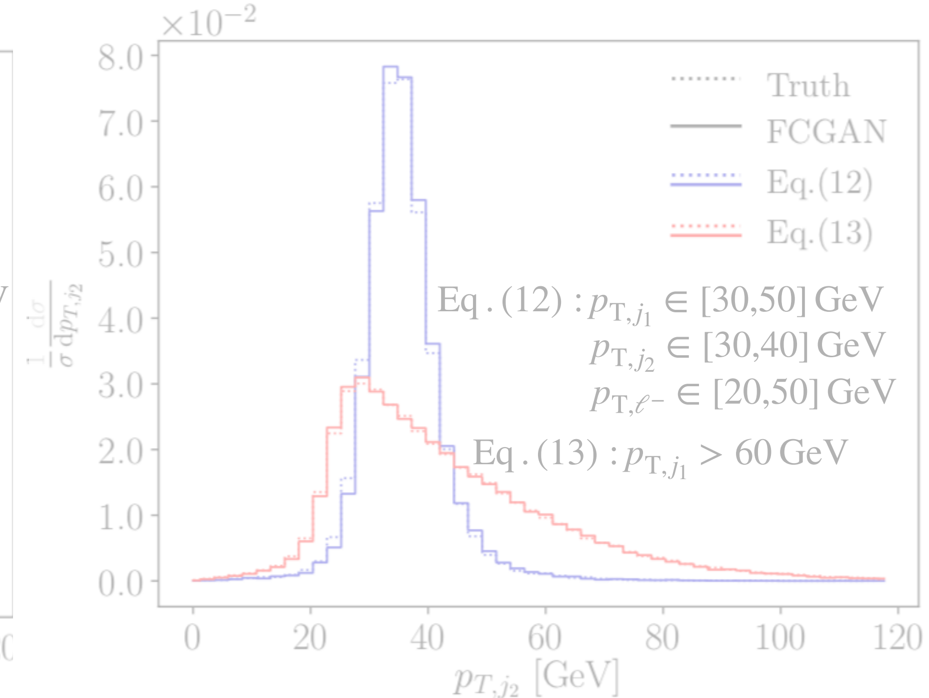


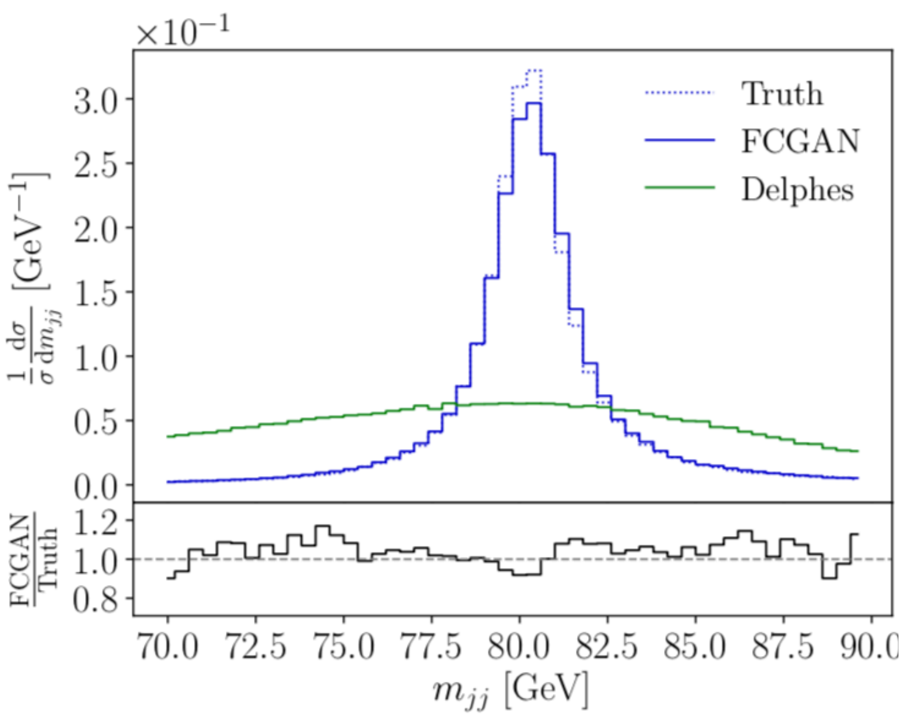
While the naive GAN only required event batches to be matched between parton level and detector level, the training of the FCGAN actually requires event-by-event matching.

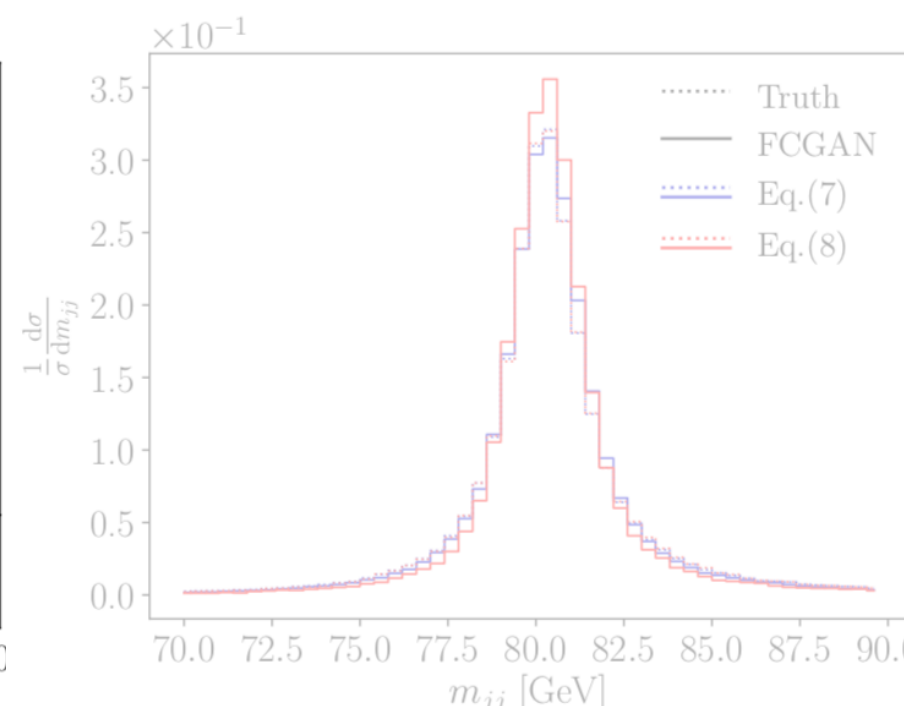
Fully conditional GAN (FCGAN) results

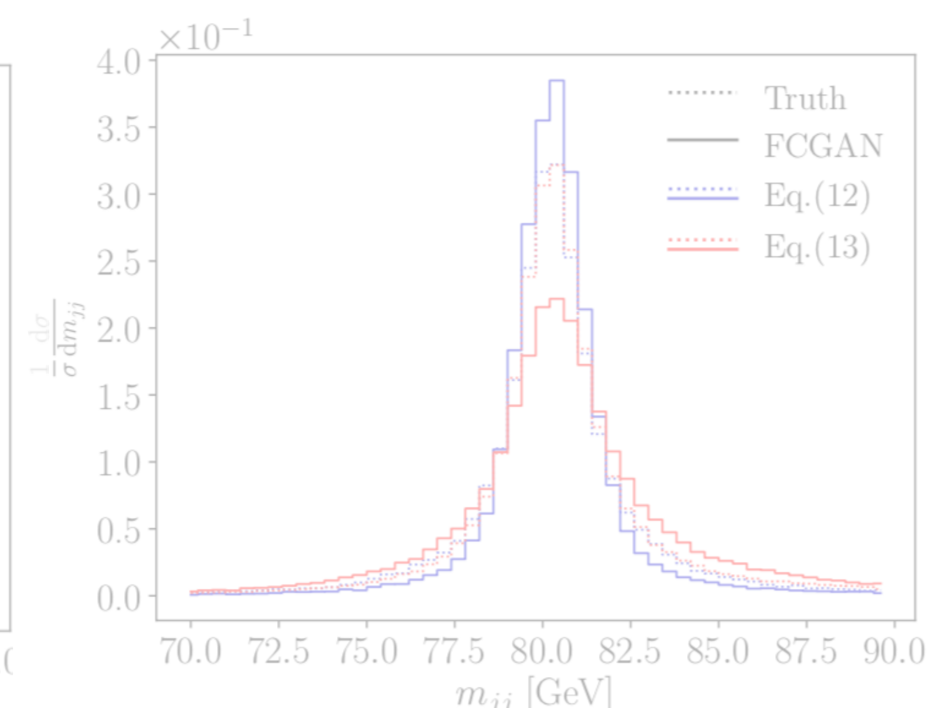










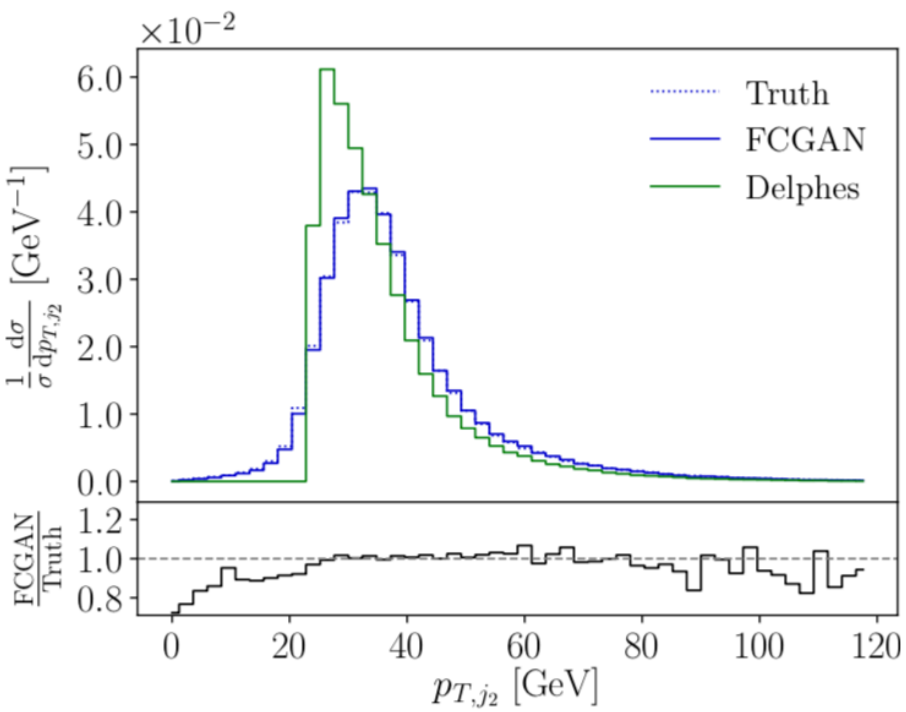


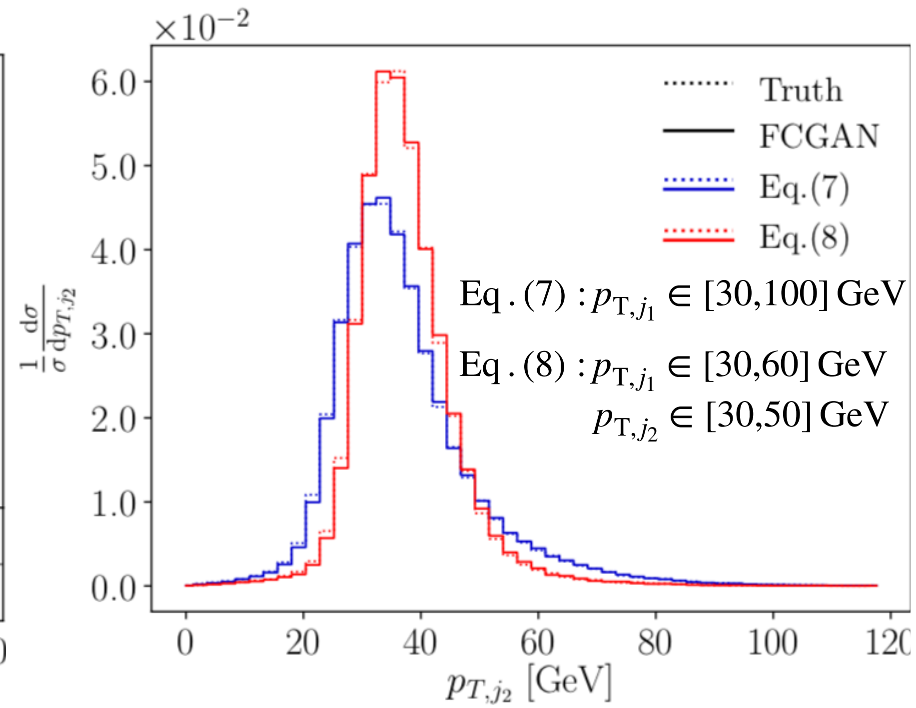
Unfolding full phase space works even better than naive GAN

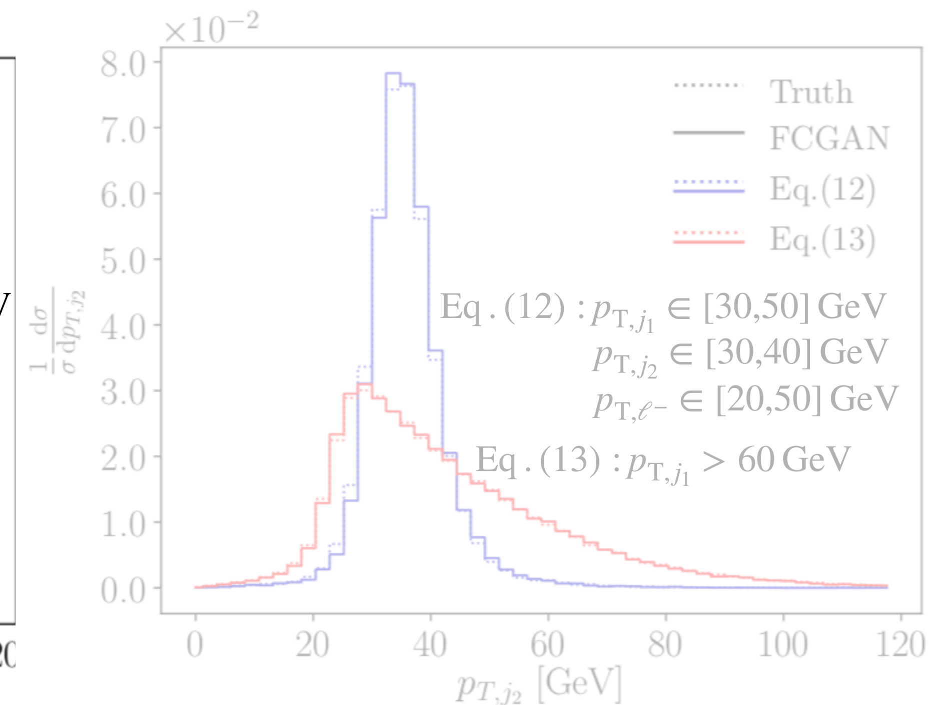
Unfolding a part of phase space works now

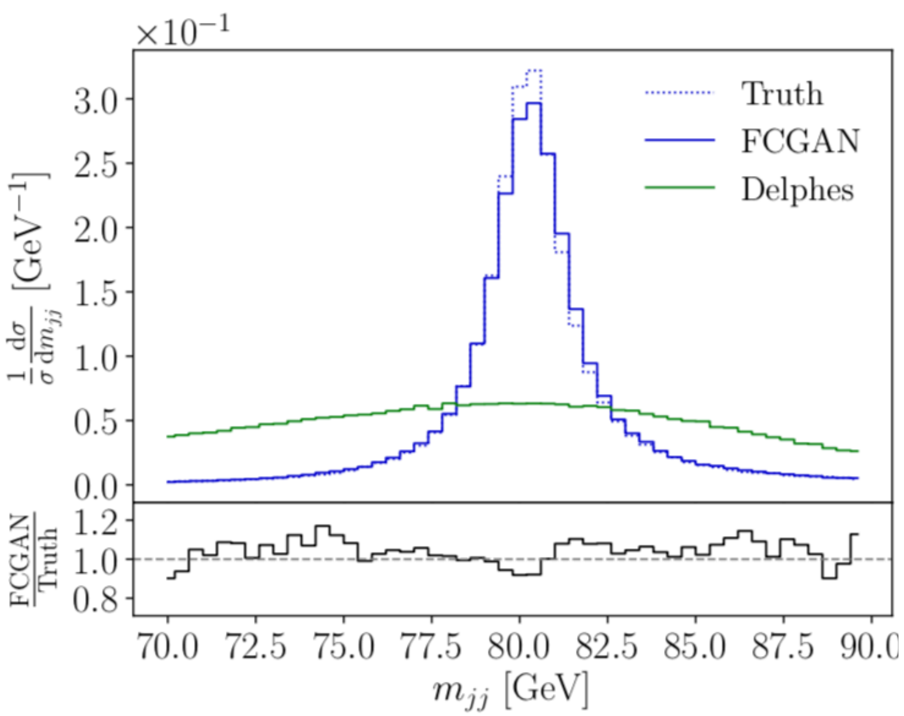
Eq. (12) 14% of phase space still works. Cutting off one side with eq (13) only spoils m_{jj} could maybe fixed with conditional MMD

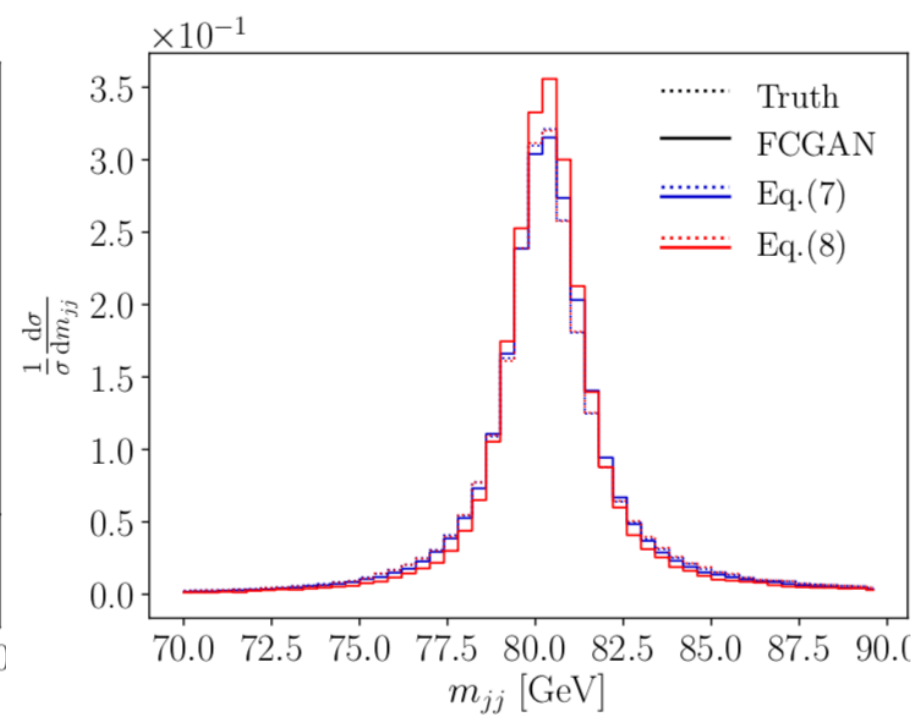
Fully conditional GAN (FCGAN) results

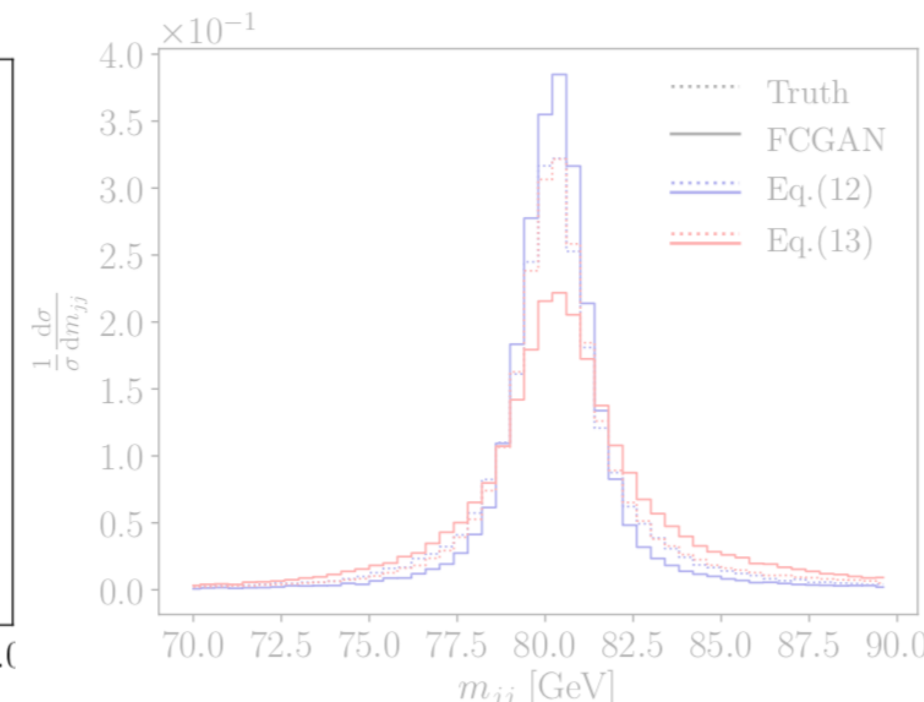










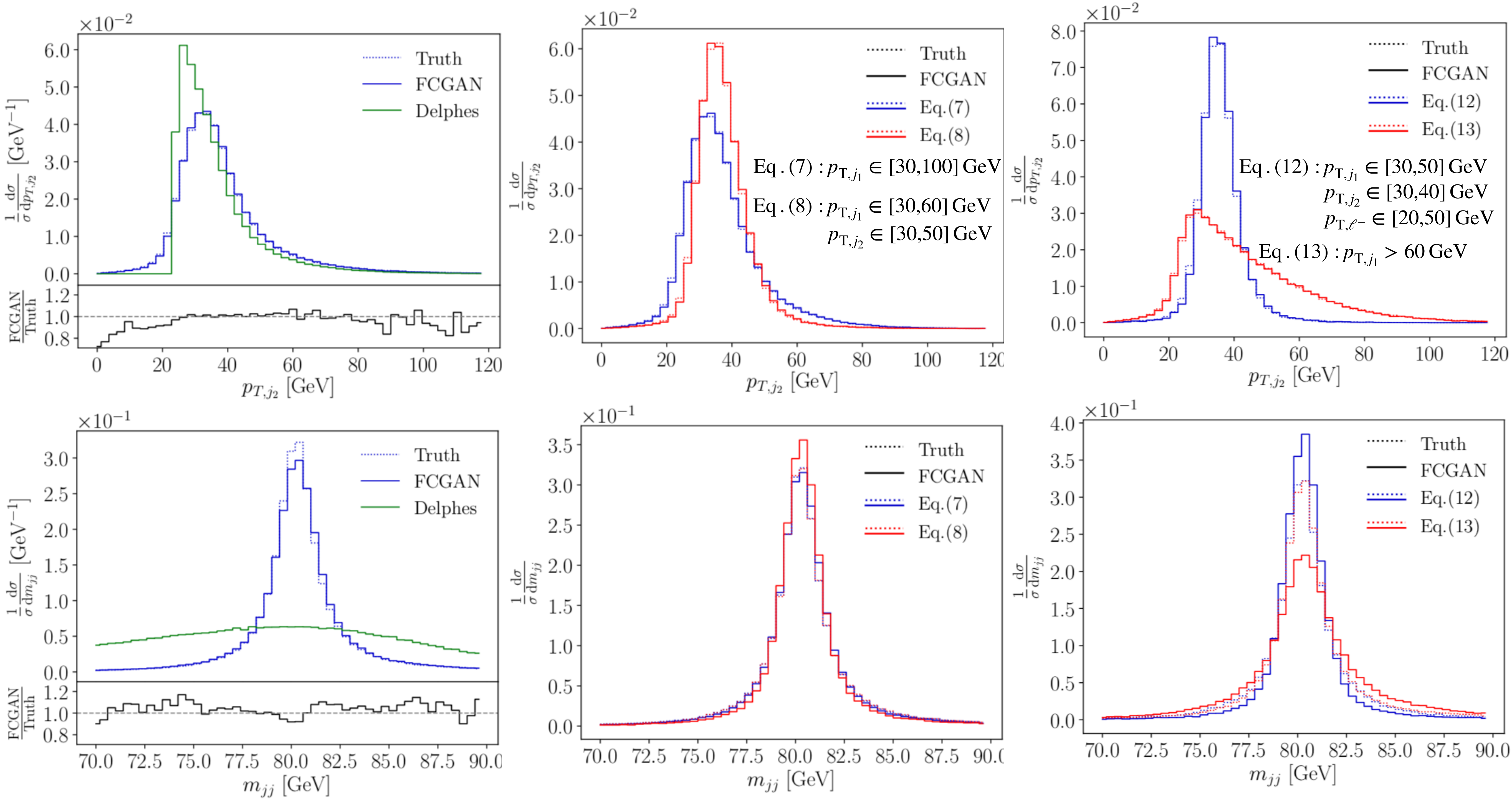


Unfolding full phase space works even better than naive GAN

Unfolding a part of phase space works now

Eq. (12) 14% of phase space still works. Cutting off one side with eq (13) only spoils m_{jj} could maybe fixed with conditional MMD

Fully conditional GAN (FCGAN) results



Unfolding full phase space works even better than naive GAN

Unfolding a part of phase space works now

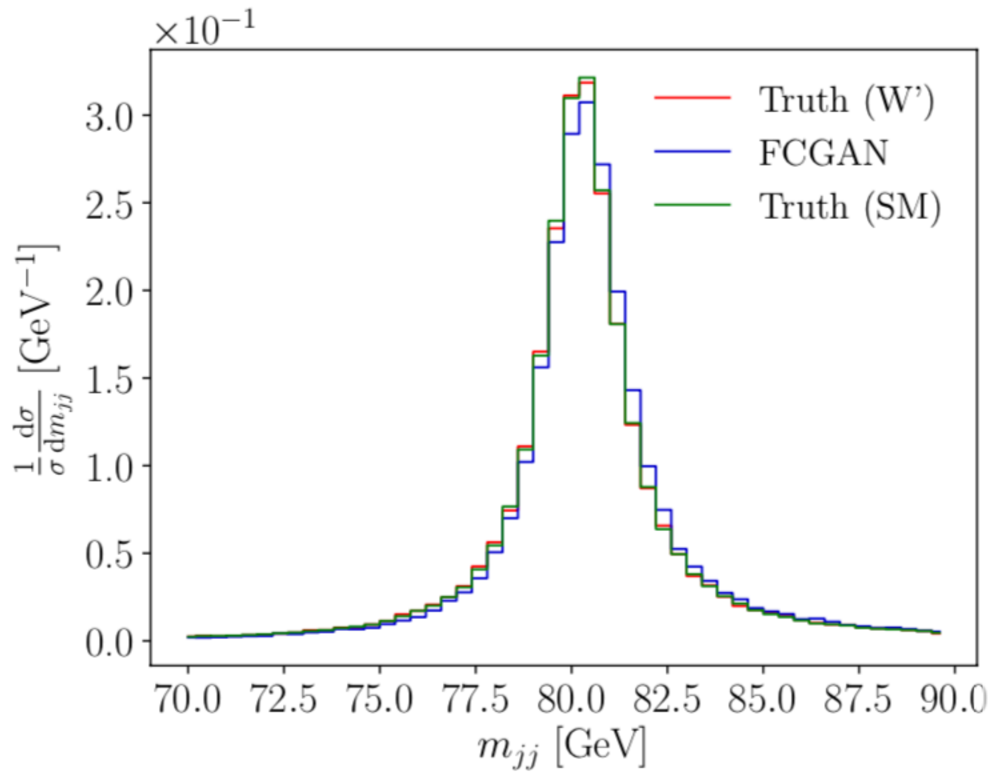
Eq. (12) 14% of phase space still works. Cutting off one side with eq (13) only spoils m_{jj} could be maybe fixed with conditional MMD

Bonus: new physics injection

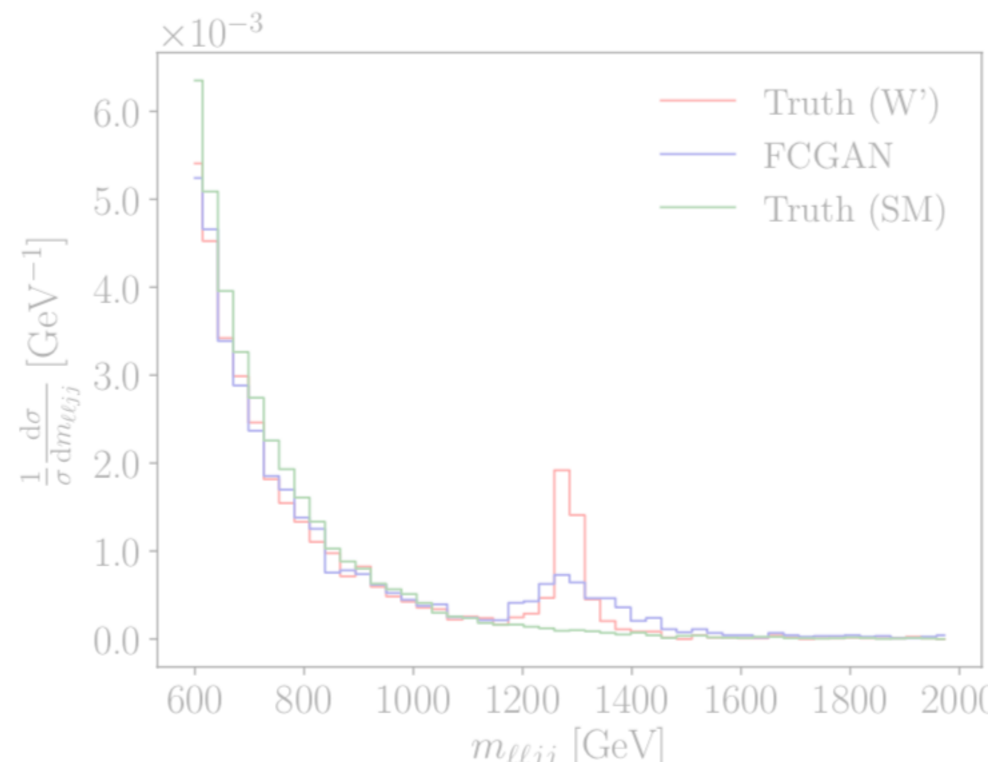
Test model dependence on FCGAN-unfolding:

What happens if we train our FCGAN on Standard Model data, but apply it to a different hypothesis?

- Use W$'$ with mass of 1.3 TeV and width of 15 GeV $pp \to W'^* \to ZW^\pm \to (\ell^-\ell^+)\,(jj)$



Invariant mass of the hadronically decaying W-boson hardly changes



Reproduces the W$'$ peak faithfully

W$'$-mass as the central peak position very well learned

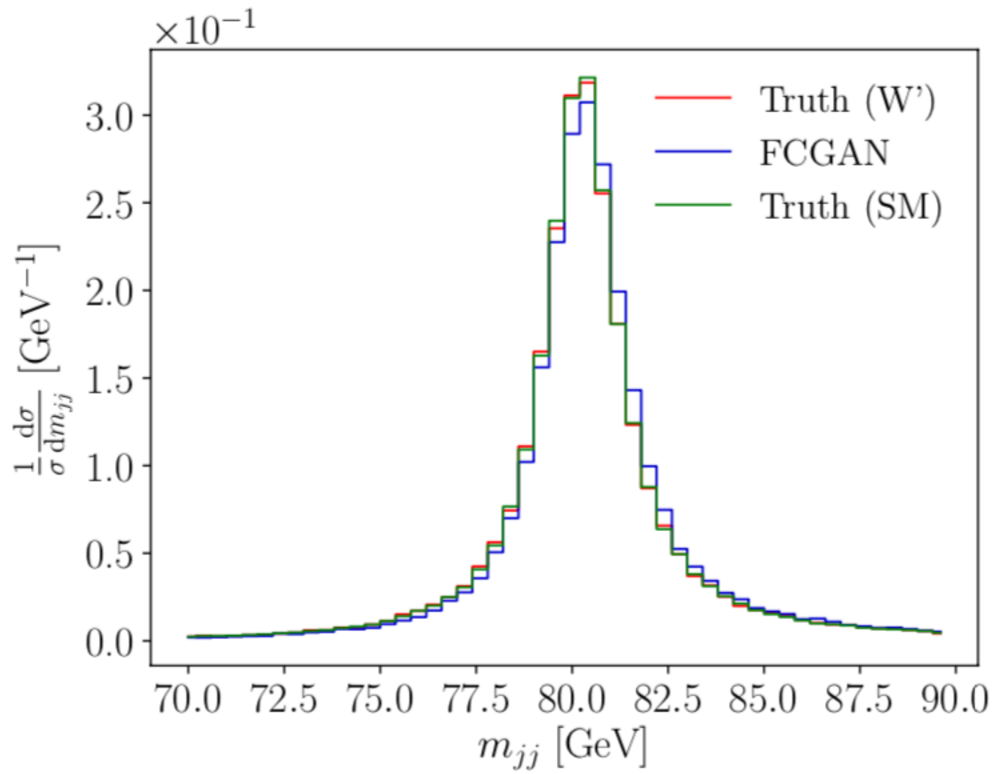
Only issue is the W$'$-width, which the network over-estimates

Bonus: new physics injection

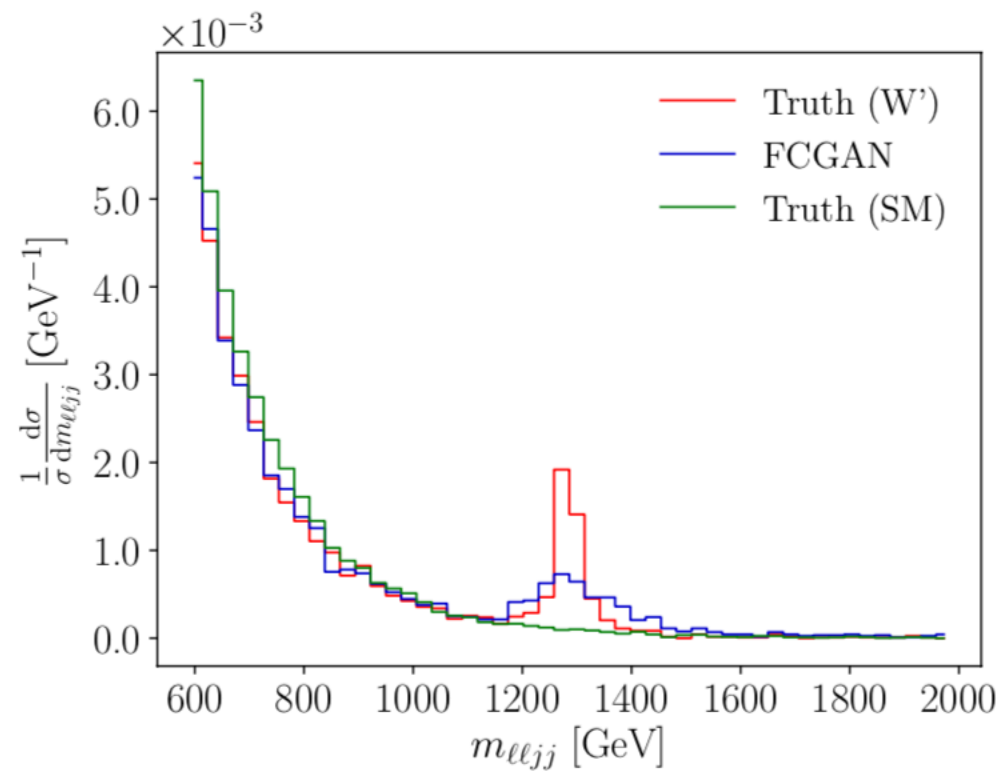
Test model dependence on FCGAN-unfolding:

What happens if we train our FCGAN on Standard Model data, but apply it to a different hypothesis?

- Use W′ with mass of 1.3 TeV and width of 15 GeV $pp \to W'^* \to ZW^\pm \to (\ell^- \ell^+)\,(jj)$



Invariant mass of the hadronically decaying W-boson hardly changes



Reproduces the W′ peak faithfully

W′-mass as the central peak position very well learned

Only issue is the W′-width, which the network over-estimates

Summary

- Short introduction to GANs
- Unfolding - GAN unfolding
- Results using a naive GAN
- Results using FCGAN
- Testing on model-independence of FCGAN

Backup

Parameter	Value	Parameter	Value
Layers	12	Batch size	512
Units per layer	512	Epochs	1200
Trainable weights G	3M	Iterations per epoch	500
Trainable weights D	3M	Number of training events	3×10^5
λ_G	1		
λ_D	10^{-3}		

Table 1: FCGAN setup.