Dijet GAN

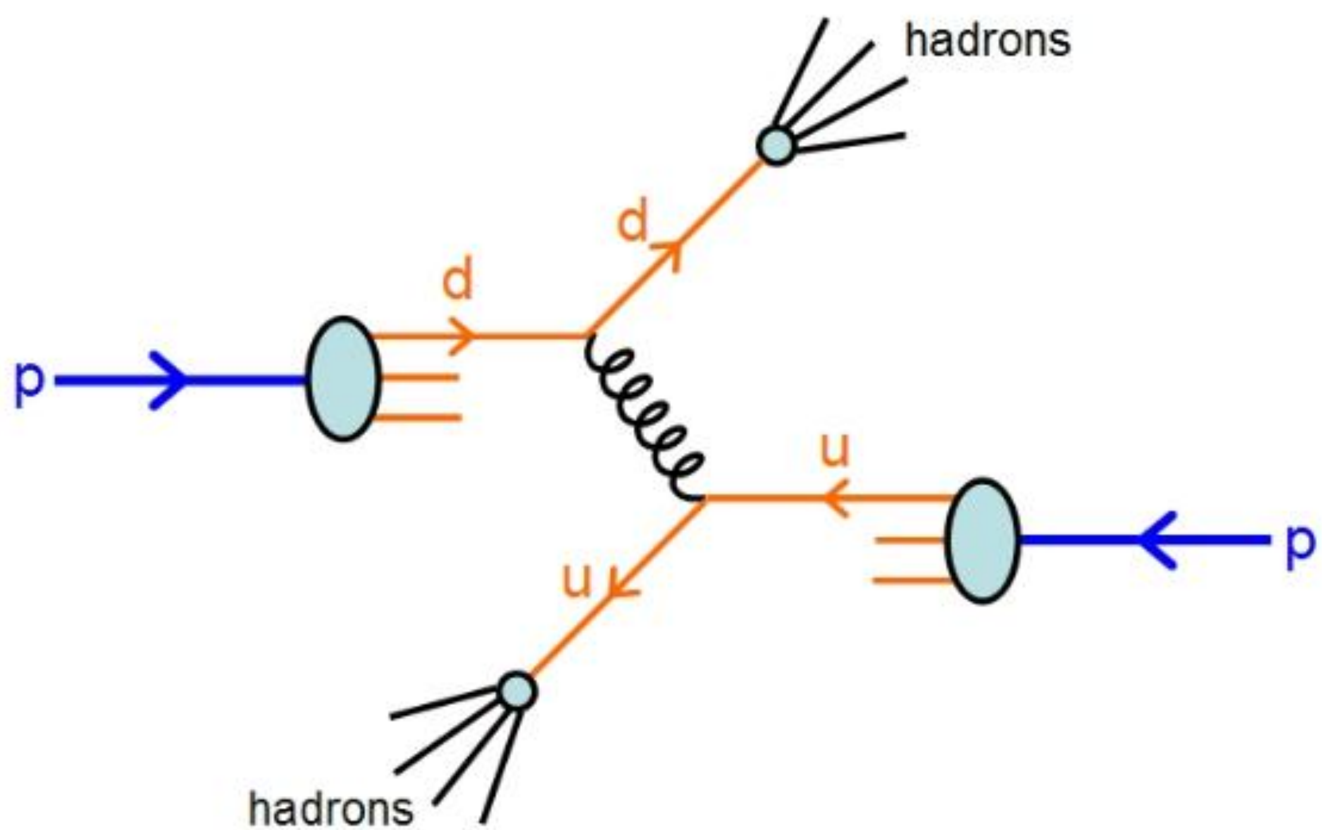
A Generative-Adversarial Network for the Simulation of QCD Dijet Events at the LHC

R. Di Sipio, M. Faucci Giannelli, S. Ketabchi, S. Palazzo

arXiv:1903.02433

LeptonPhoton 2019, 5-10 Aug, Toronto

Physics of QCD Events

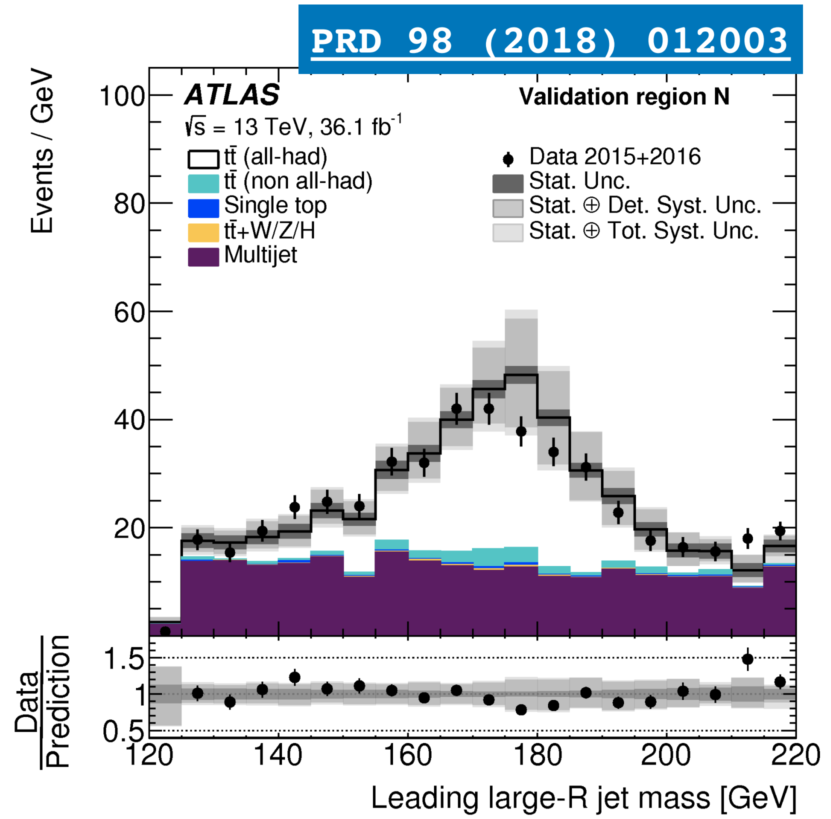


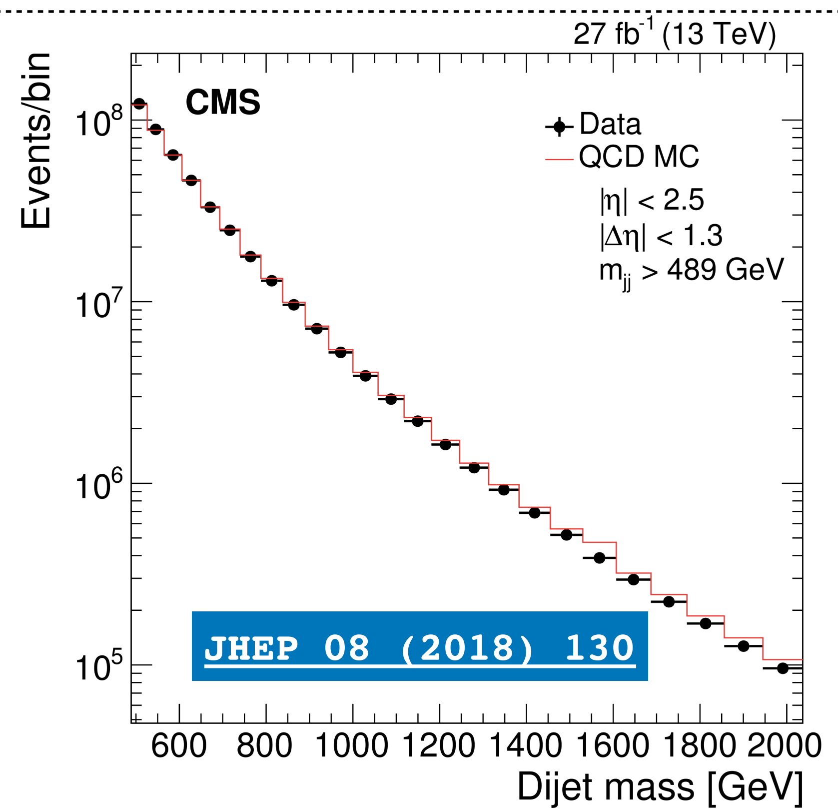
At the LHC, *pp* collisions result in the interaction of quarks and gluons (partons). 2→2 scattering processes with a pair of jets in the final state are called **dijet events**.

Jets emerge from pp collisions with **high transverse momentum** (p_T) and **large angle** (η) with respect to the incoming partons. The **jet mass** ($m = \sqrt{E^2 - P^2}$) comes mostly from the complex dynamics of **strong interactions**.

The simulation of these events is performed by **Monte Carlo generators**. The complete simulation of a single event takes **several minutes**, while **millions of events** have to be generated!

Physics of QCD Events





Dijet / multi-jet events are the **largest background** to precision measurements and searches for physics beyond the Standard Model in channels with no leptons

Monte Carlo simulations of these processes are **inaccurate or impractical**. Typical solutions involve:

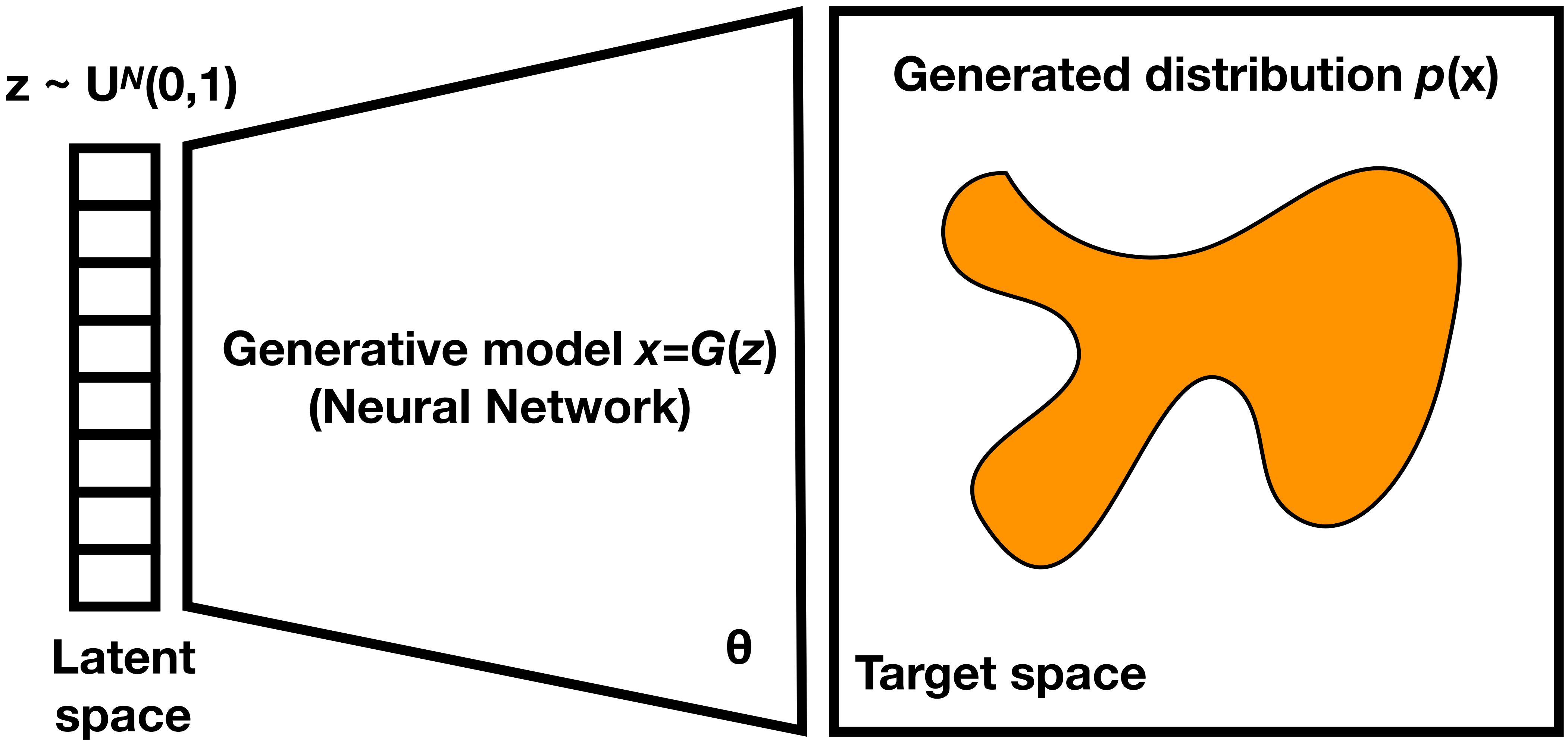
- Data driven estimations
- Fitting with parametric function

What we want eventually is a method that is:

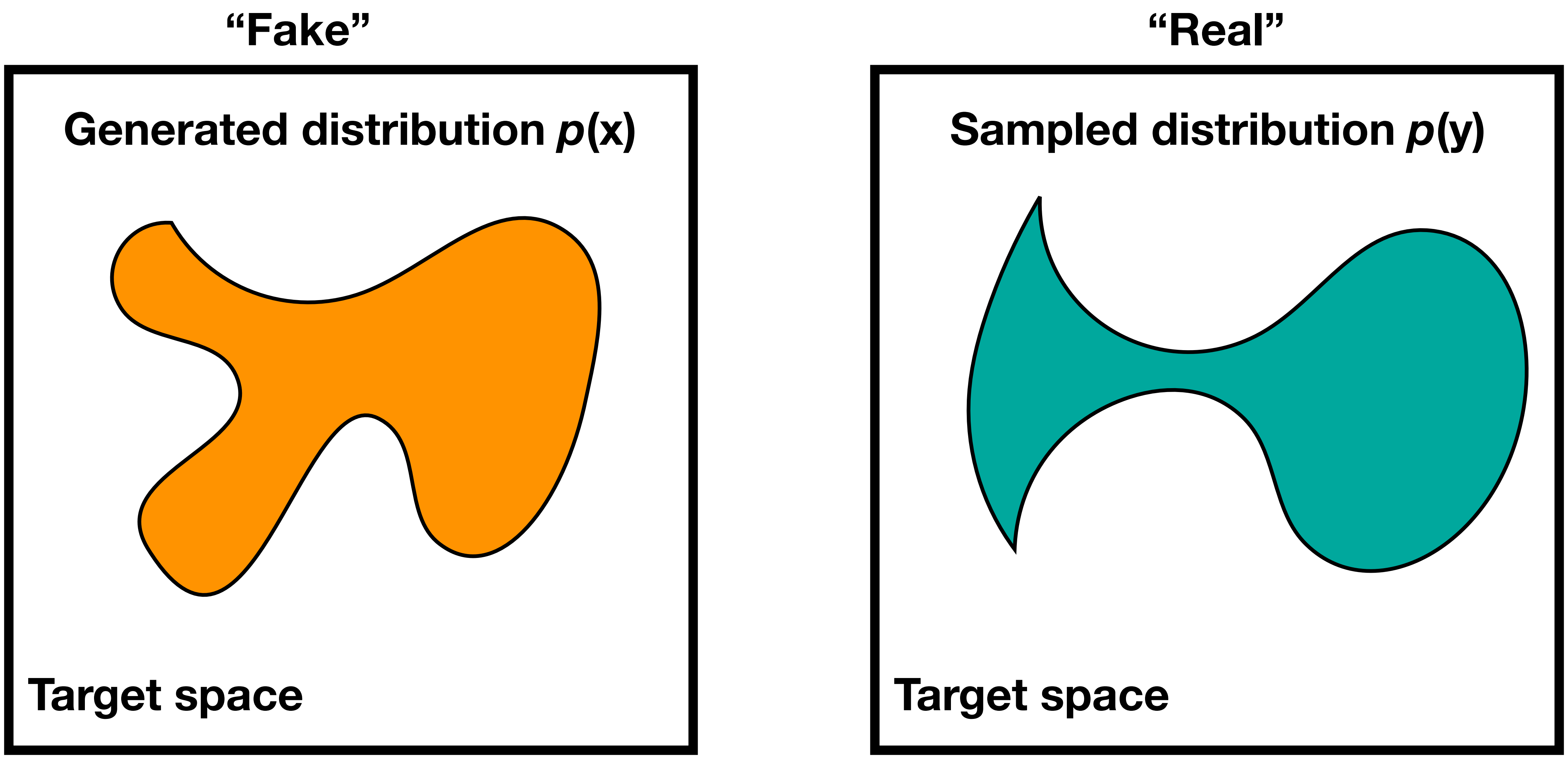
- **Generic** - dijet, $t\bar{t}$, W/Z+jets…
- **Accurate** - *e.g.* includes higher-order terms
- **Fast** - generate O(10^6) events in seconds
- **Robust** - extrapolation to tails (high p_T, m_{jj})

This talk: proof of concept with simple physics case (dijet production)

Generative Network

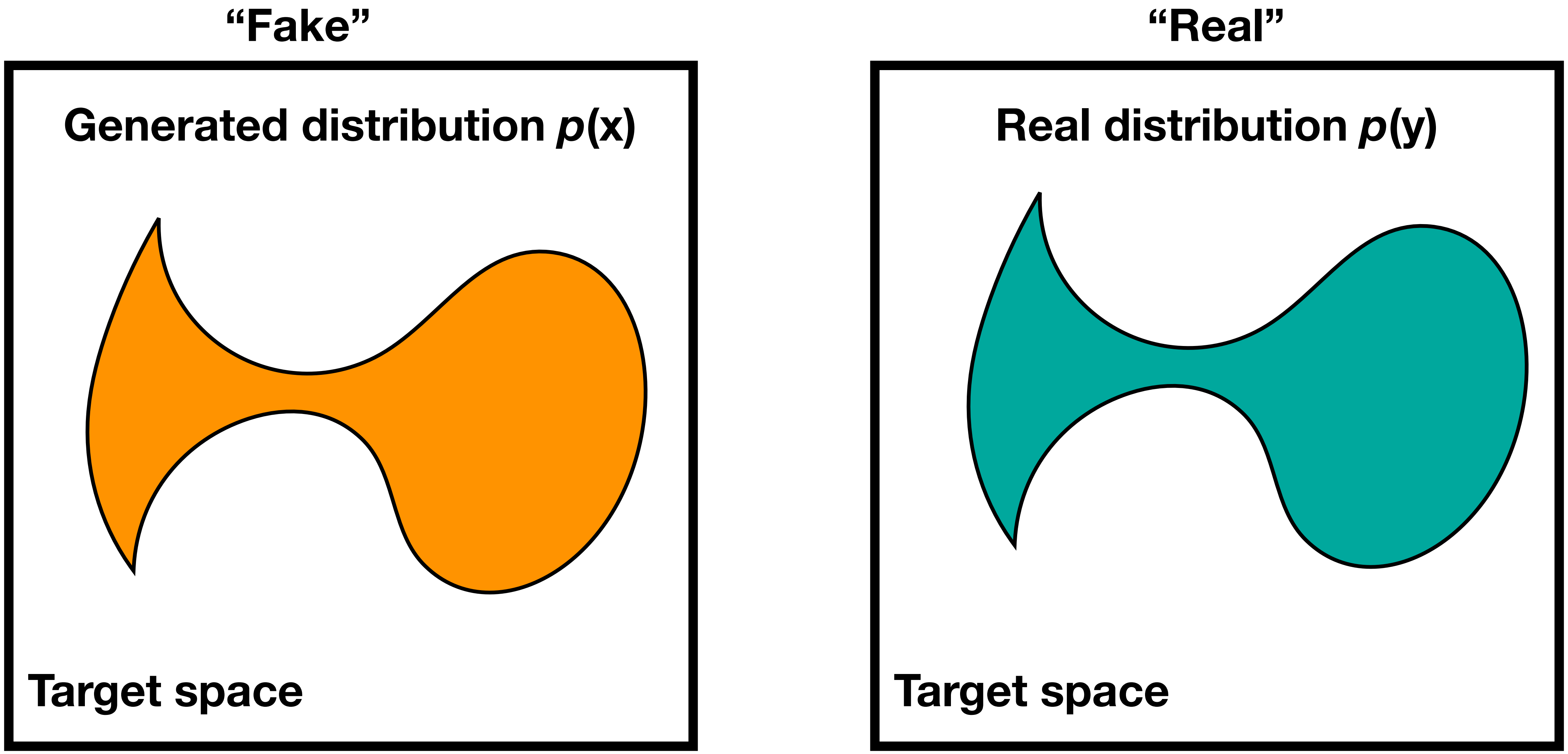


Generative Network



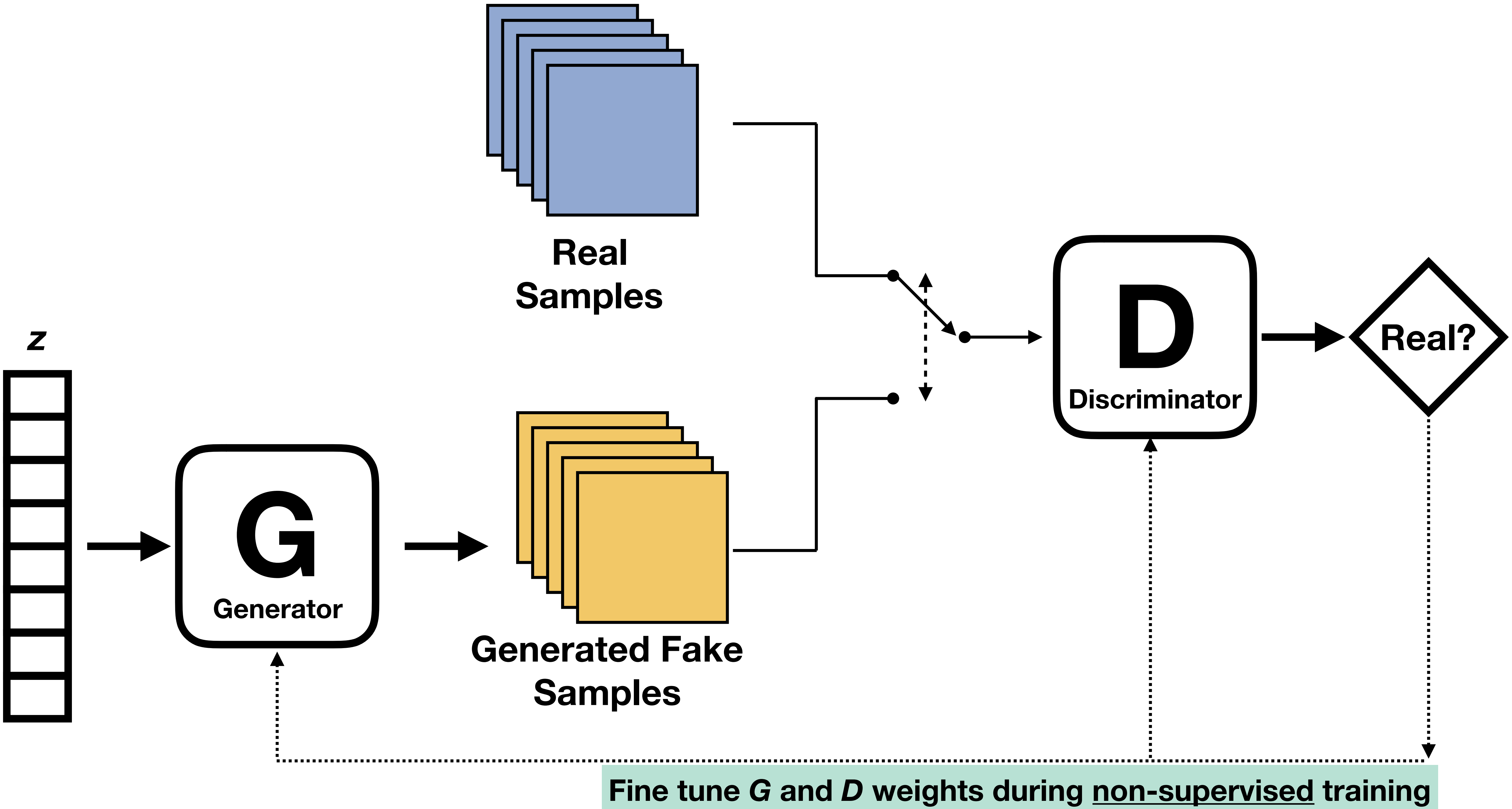
Adjust weights in $G(z)$ so that $p(x)=p(G(z)) \sim p(y)$

Generative Network



Adjust weights in $G(z)$ so that $p(x)=p(G(z)) \sim p(y)$

Generative-Adversarial Network

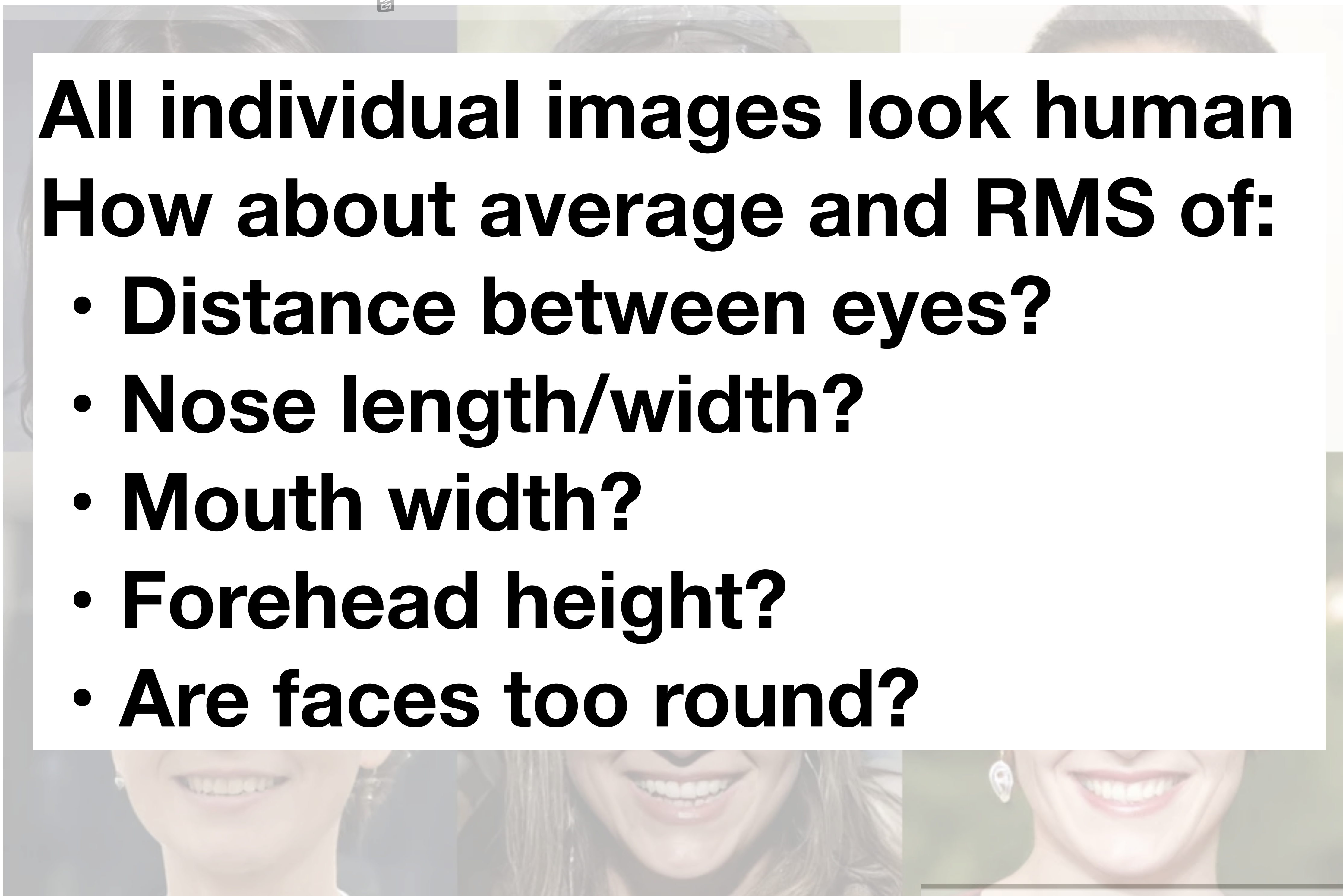


These persons do not exist

https://www.thispersondoesnotexist.com/

Imagined by a GAN (generative adversarial network)
StyleGAN (Dec 2018) - Karras et al. and Nvidia
Original GAN (2014) - Goodfellow et al.
Don't panic. Learn about how it works.
Help me figure out what was learned here.
Help this AI continue to dream
Another | Save • Cats | Articles | TV Friends - Office | x

These persons do not exist



All individual images look human

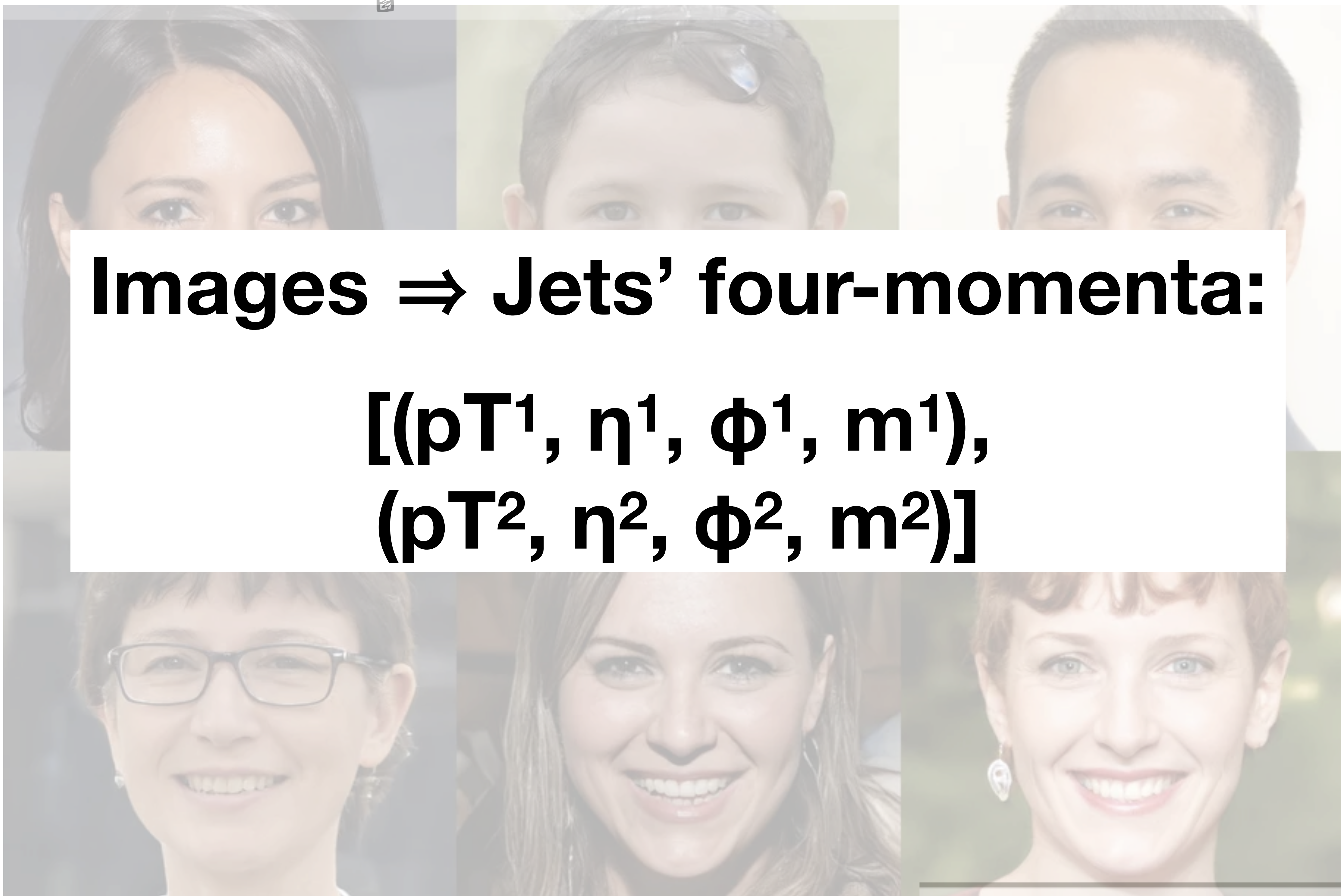
How about average and RMS of:

- **Distance between eyes?**
- **Nose length/width?**
- **Mouth width?**
- **Forehead height?**
- **Are faces too round?**

https://www.thispersondoesnotexist.com/

Imagined by a GAN (generative adversarial network)
StyleGAN (Dec 2018) - Karras et al. and Nvidia
Original GAN (2014) - Goodfellow et al.
Don't panic. Learn about how it works.
Help me figure out what was learned here.
Help this AI continue to dream
Another | Save • Cats | Articles | TV Friends - Office | x

These persons do not exist



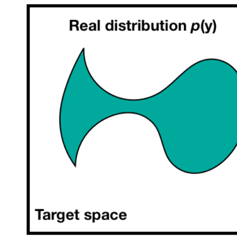
Images ⇒ Jets' four-momenta:

$$[(pT^1, \eta^1, \phi^1, m^1), (pT^2, \eta^2, \phi^2, m^2)]$$

https://www.thispersondoesnotexist.com/

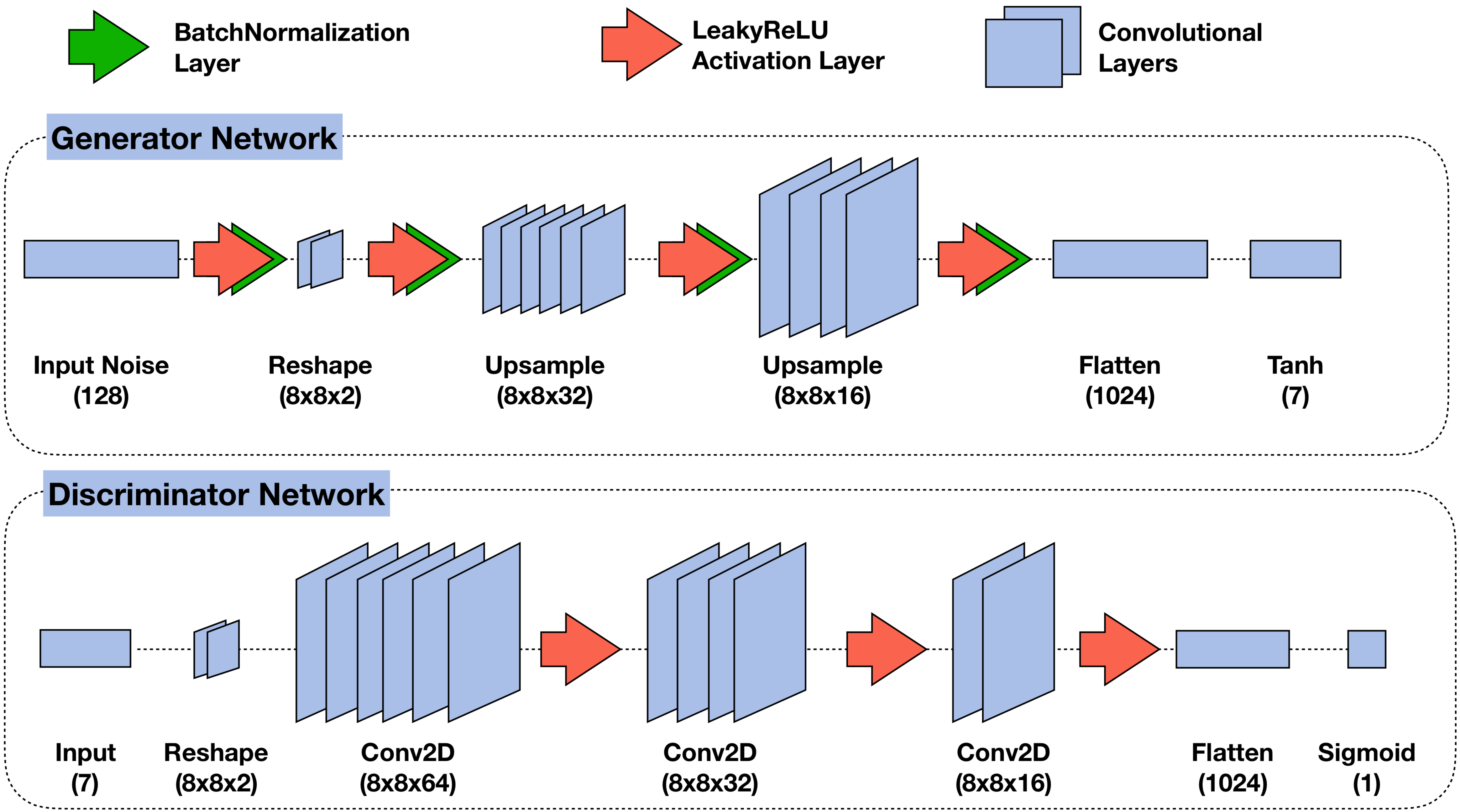
Imagined by a GAN (generative adversarial network)
StyleGAN (Dec 2018) - Karras et al. and Nvidia
Original GAN (2014) - Goodfellow et al.
Don't panic. Learn about how it works.
Help me figure out what was learned here.
Help this AI continue to dream
Another | Save • Cats | Articles | TV Friends - Office | x

Real Monte Carlo Sample



- MadGraph5 + Pythia8, 10 million events ~ 0.5 fb-1
- Parton-level filter: H_T > 500 GeV
- Fast detector simulation (Delphes3) with pileup $<\mu>=25$
- Anti-k_T R=1.0 jets, p_T > 250 GeV
- Approx 7.5M events generated events (particle level), 4.5M after detector simulation (reco level)
    - 1M events in ~80 sec

Network Architecture

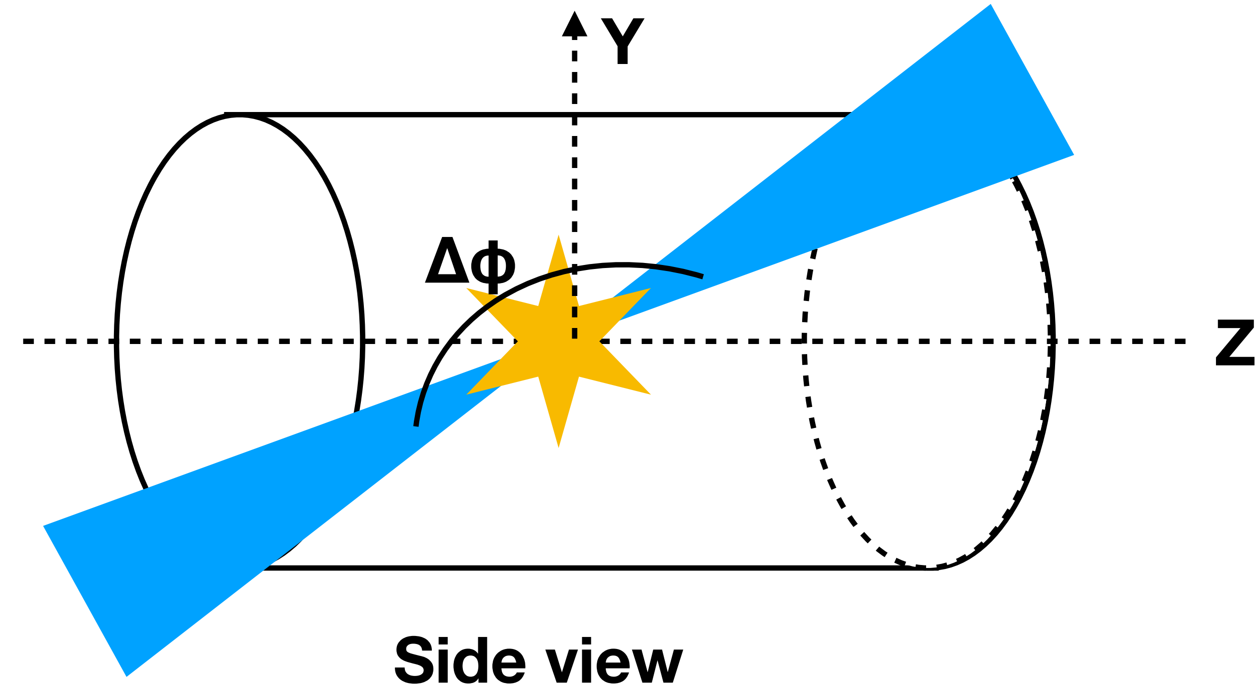


Other architectures tried (fully connected, RNNs) but CNNs yielded best results thanks to their superior ability to “learn" complex patterns

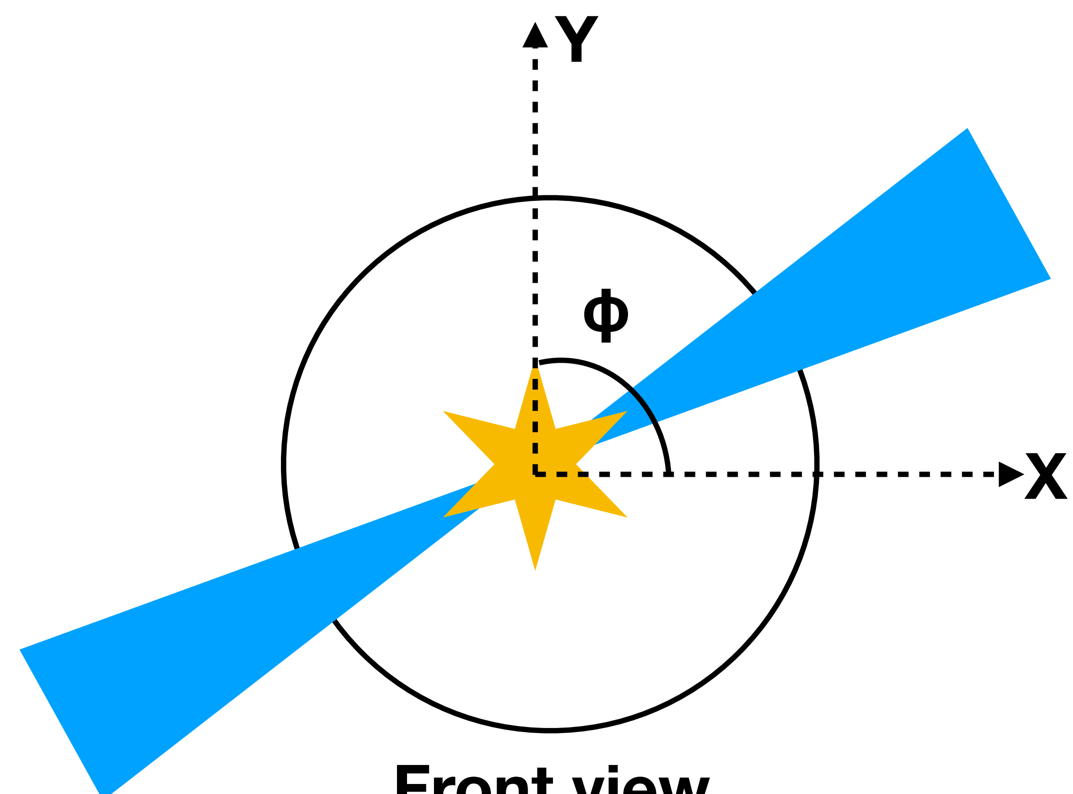
Pre-processing

- Dijet events have a number of **intrinsic symmetries**
- Symmetries are **hard to learn** just by throwing events at the network
- Preprocessing:
    - Rotation so that $\phi_1 = 0$ (azimuthal symmetry, also removes one degree of freedom)
    - Mirror so that $\Delta\phi > 0$
    - Flip pseudo-rapidity η (left-right symmetry)

$[(pT^1, \eta^1, \phi^1, m^1), (pT^2, \eta^2, \phi^2, m^2)] \rightarrow [(pT^1, \eta^1, m^1), (pT^2, \eta^2, \phi^2, m^2)]$



Side view



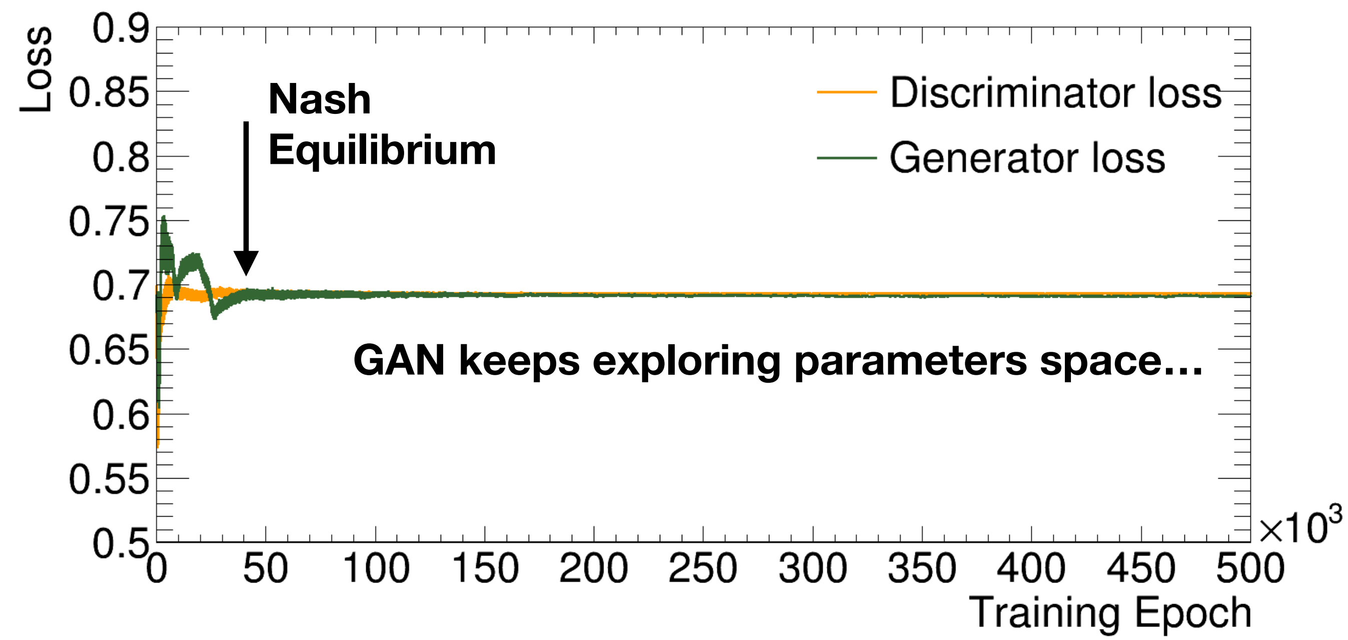
Front view (transverse plane)

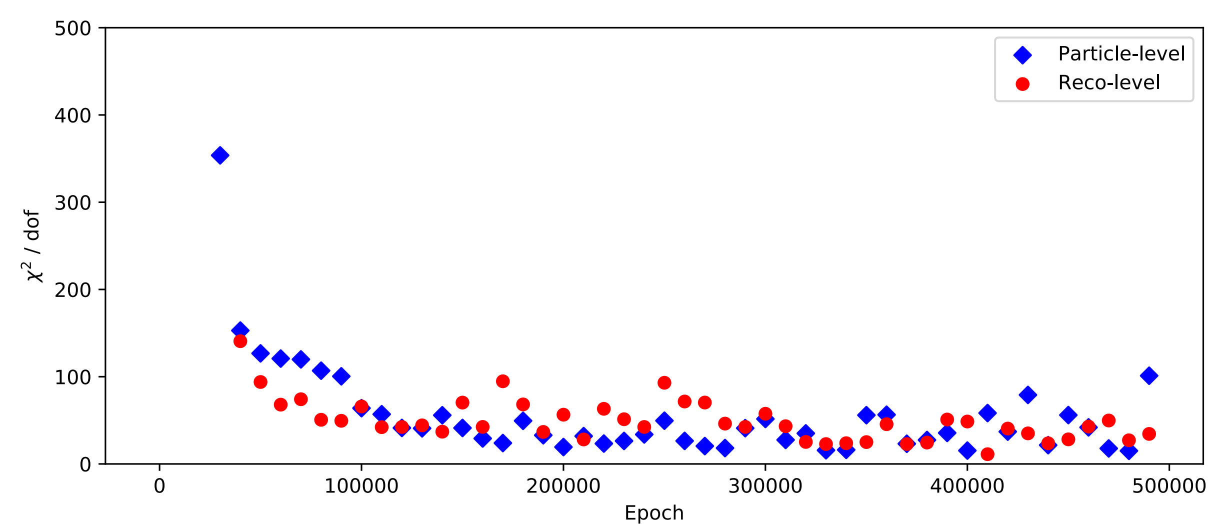
Training

- Training is a **minmax** game, not a minimization
  ⇒ **stabilization** of loss

  ⇒ No natural way to measure the agreement for choosing the best training epoch

  - Take the one with lowest χ^2

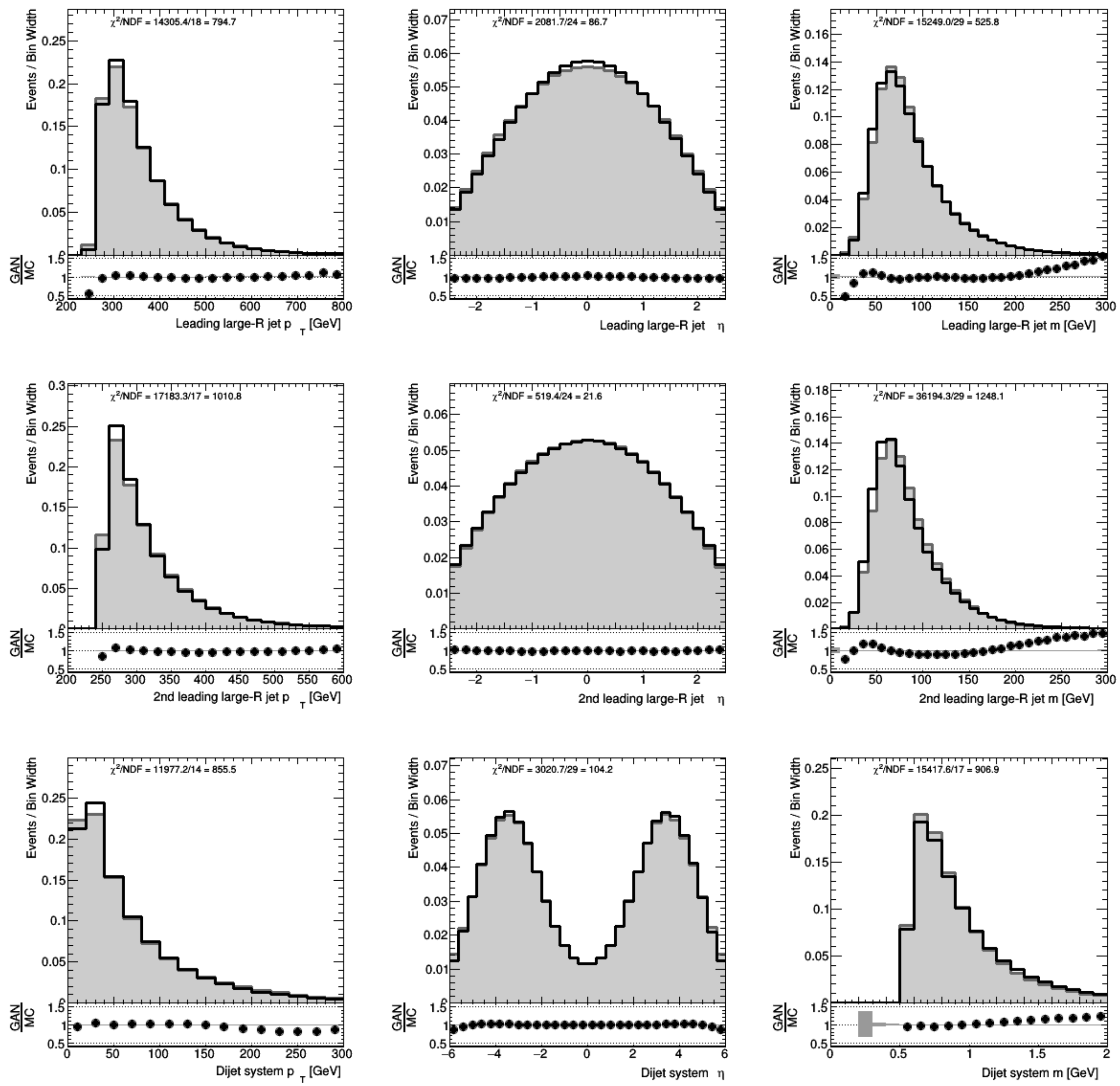




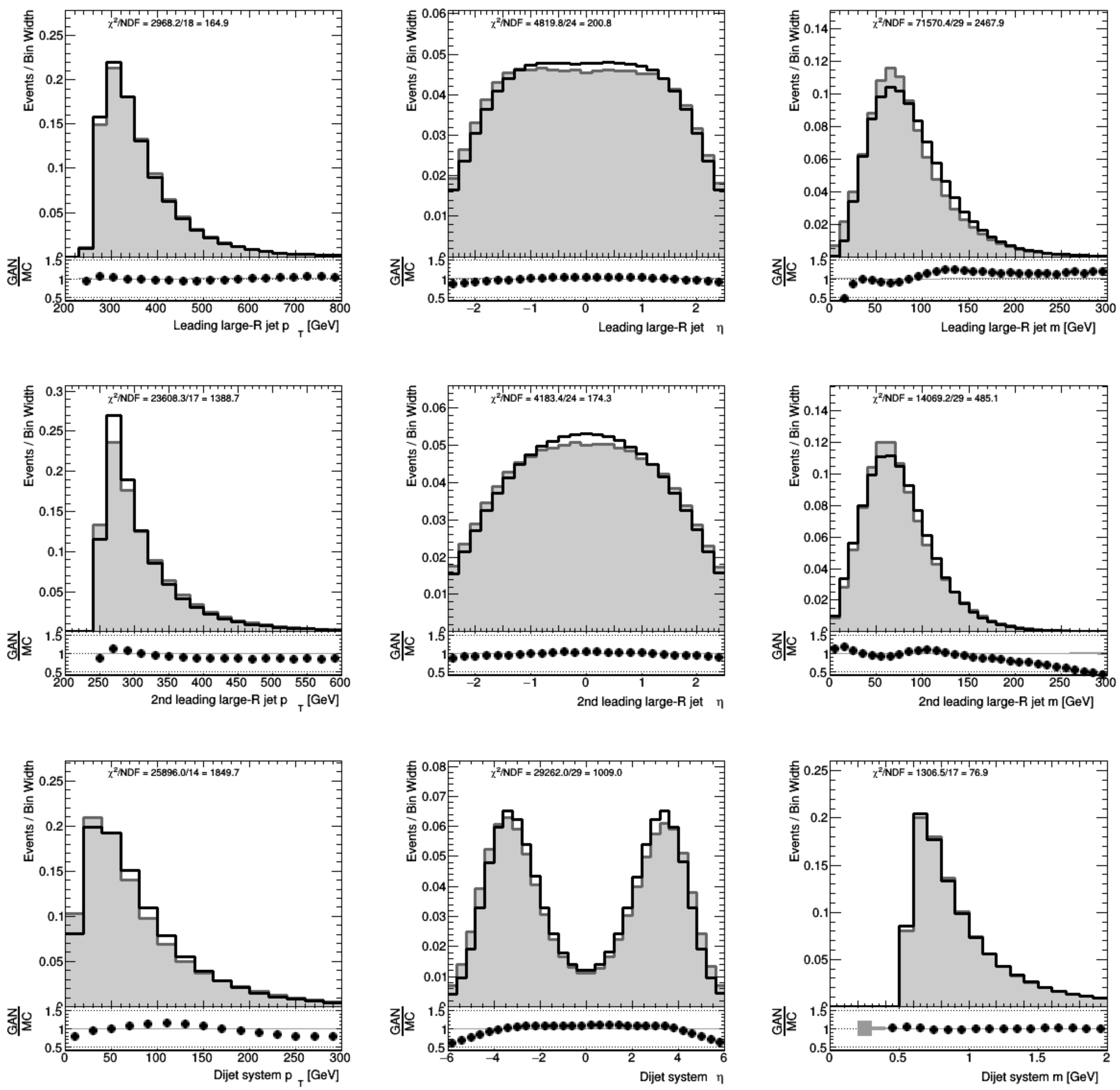
- Cross-entropy loss quickly converges to -ln(0.5)=0.693

- GPU NVIDIA Quadro P6000, ~100k epochs / 1hr

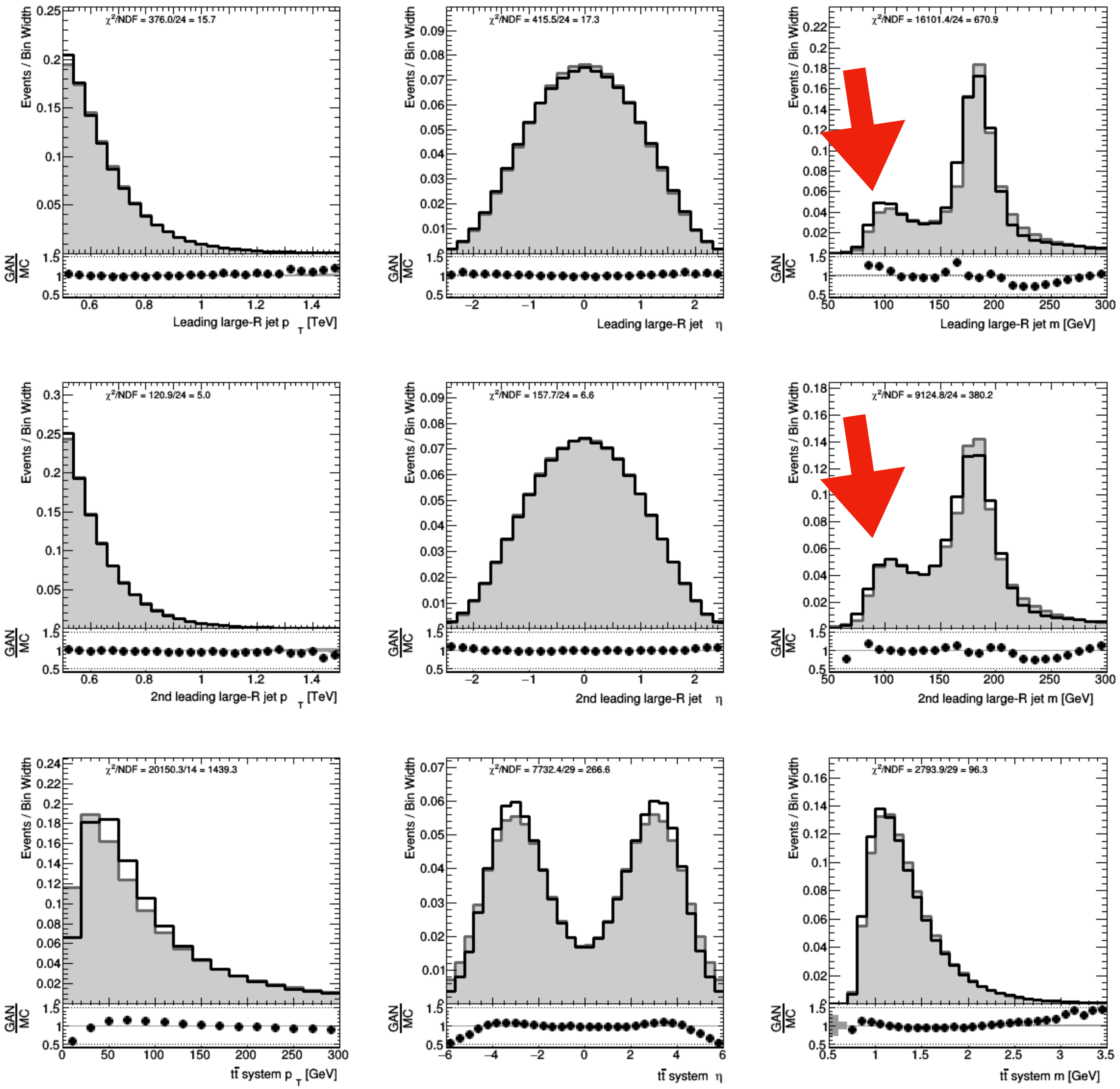
Results - Particle level

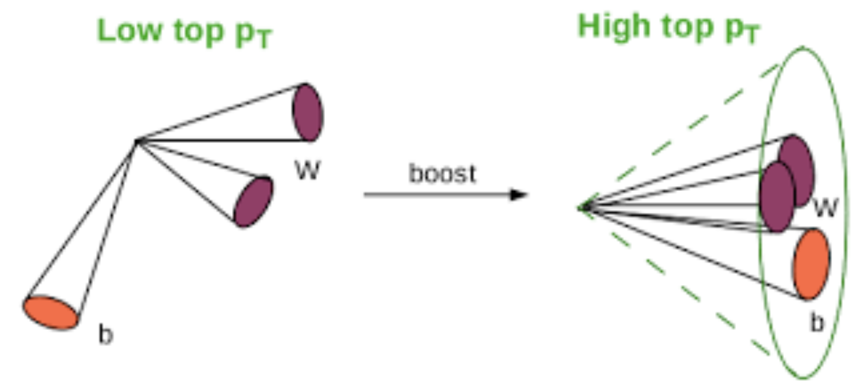


Results - Reco level



Results - boosted top quarks

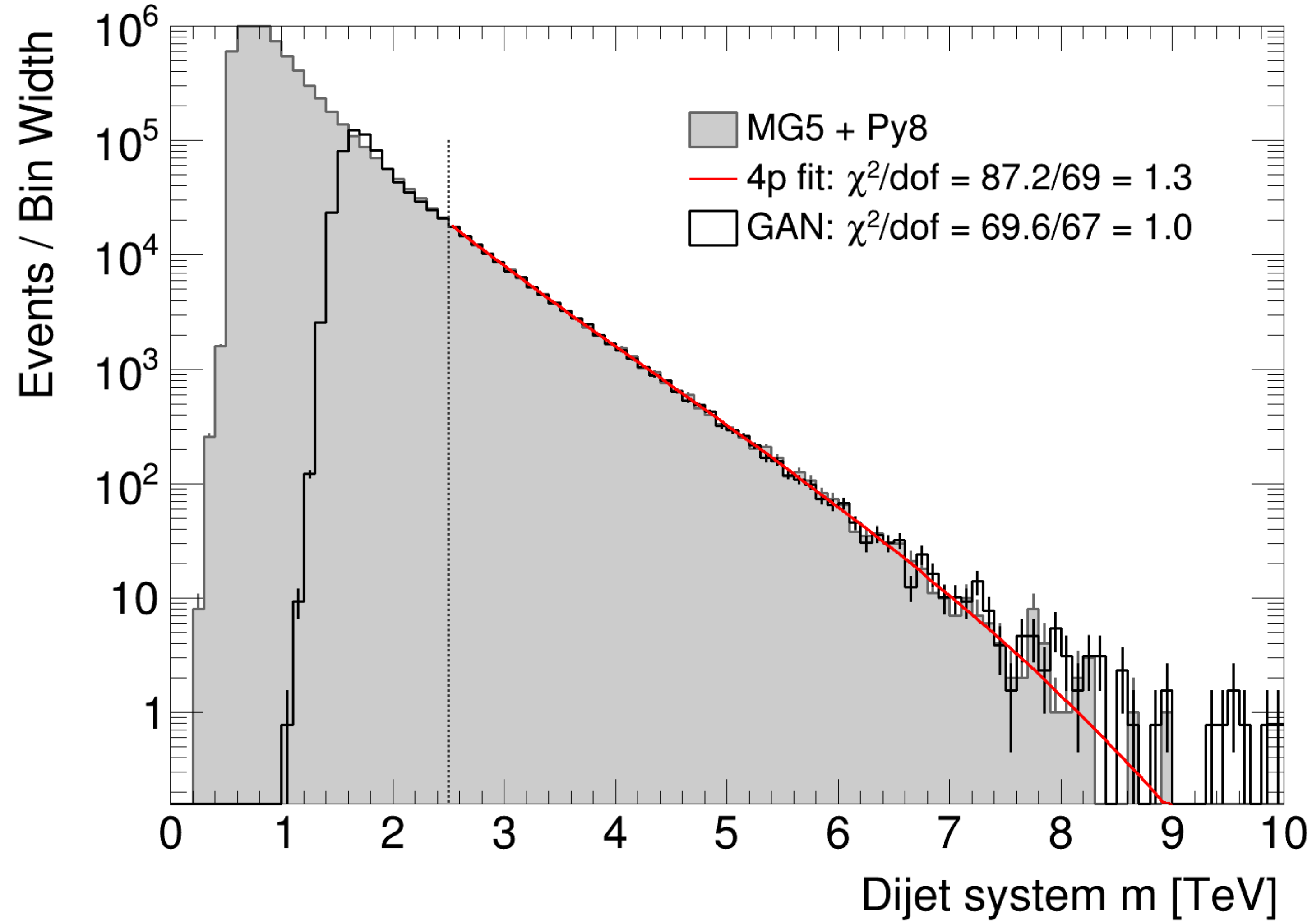




- $t\bar{t} \rightarrow WbWb \rightarrow bq\bar{q}b\bar{q}$
- Non trivial to fit mass distribution with two peaks:
  - Semi-boosted: $m_J \sim m_W \sim 80$ GeV
  - Fully-boosted: $m_J \sim m_t \sim 173$ GeV

Results - Extrapolation

- Use only 150k events with m_{jj} > 1.5 TeV to train the network
- Generate 1M events
- Compare against:
  - 4-params fit $f(x) = \frac{p_0(1-x)^{p_1}}{x^{(p_2+p_3 \log x)}}$
  - Real MC



Conclusions and Outlook

- Machine learning applications to HEP is blooming
- Generative networks to **speed up** generation of large Monte Carlo samples
    - Generate small MC with high **accuracy**, use GAN to increase **statistics**
    - Quality of the **inputs** >> the network parameters
    - **Symmetry** ⇒ exploit it!
    - GANs are not a minimization, picking up the **best** epoch is **not trivial** (e.g. lowest χ^2)
- **Generic** method, apply to other processes (top quarks, W/Z+jets, Higgs…)
- Possible extensions:
    - Variable **conditioning** to populate more some regions of the phase-space (*e.g.* p_T-slicing, high-m_{jj})
    - Use **auto-encoder** (non-linear PCA) to handle arbitrary number of input variables
- Competing methods (β-VAE) also under study by other groups

Backup

Monte Carlo Simulations

Hard-scattering process

Parton Shower / Hadronization / Decay

proton-proton Interaction

Particle level

Detector Simulation

Final State Objects reconstruction

Reco level

- Particle-level simulations: LO(fast), NLO(slower), NNLO(slow, not always matched to parton shower)
  - Ultimate accuracy (e.g. MENLOPS) usually too slow to produce O(100M) events are needed at the LHC Run3
- Detector simulation: Full/Accurate/Slow (10 mins / event), Fast/LessAccurate
  - Attempts to speed up fast simulations (e.g. calorimeter) with GANs, tracking with other ML methods
- All in one sweep?

Software packages

- Keras v2.2.4
- Tensorflow v1.12
- Scikit-learn, Pandas, other libraries
- Input scaled in the [-1,1] range

Parameters

- Generators: 128 random number ~U(0,1) → 7 physics quantities:
  - p_T, η, m of the leading jet
  - p_T, η, ϕ, m of the sub-leading jet
- Loss functions:
  - Generator: mean square error (MSE)
  - Discriminator/GAN: binary cross-entropy
- Optimizer: Adam, $lr=10^{-5}$, $\beta_1 = 0.5$, $\beta_2 = 0.9$ (slow gradient descent with momentum)