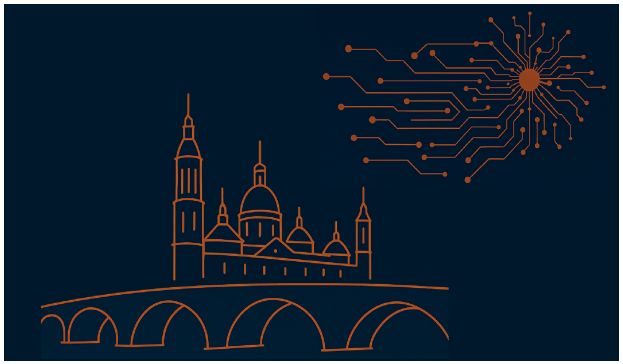


ML for HEP
Generative Models for detector simulation

David Rousseau

IJCLab-Orsay

david.rousseau@in2p3.fr @dhpmrou

COMCHA, Zaragoza, April 2026

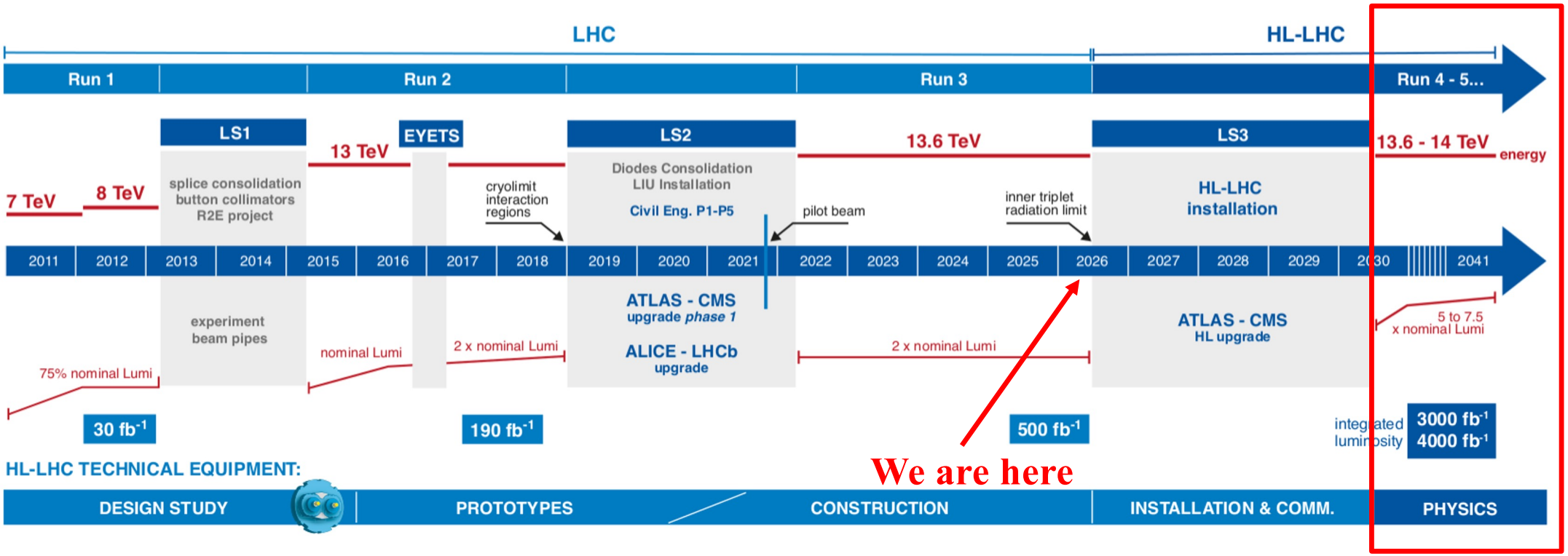
cnrs NUCLÉAIRE & PARTICULES

université PARIS-SACLAY

INSTITUT DATAIA Science des données, Intelligence & Société

IJCLab Irène Joliot-Curie Laboratoire de Physique des 2 Infinis

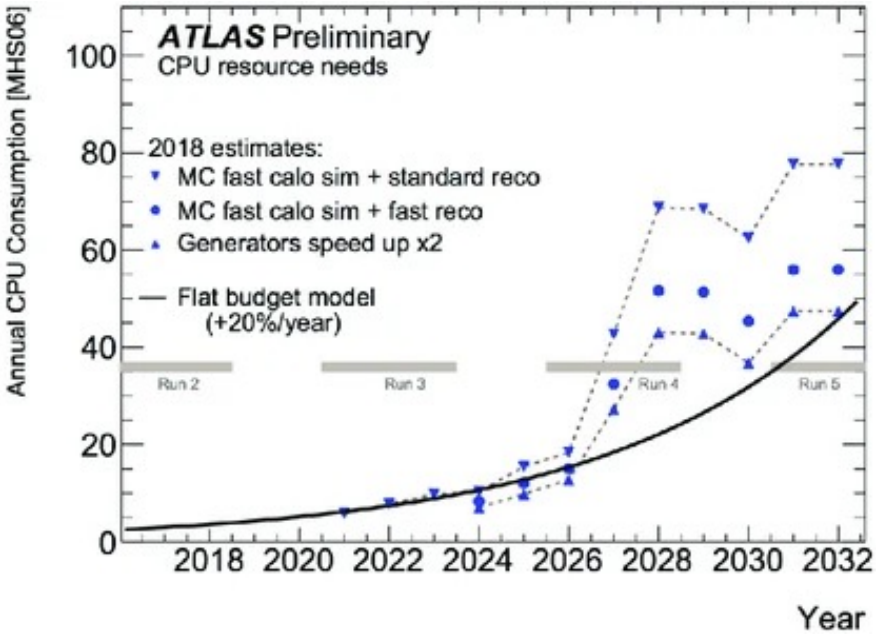
LHC / HL-LHC Plan



HL-LHC

Simulation needs

- Ever growing need of accurate simulation
- Current simulators are sophisticated but slow
- More data ⇔ more simulation



- ➔generative models to the rescue…

Conditioned Generators

Text to image GAN Reed et al. arXiv:1605.05396

2016

this small bird has a pink breast and crown, and black primaries and secondaries.

this magnificent fellow is almost all black with a red crest, and white cheek patch.

the flower has petals that are bright pinkish purple with white stigma

this white and yellow flower have thin white petals and a round yellow stamen

Nano Banana diffusion model

2026

Jan 2024

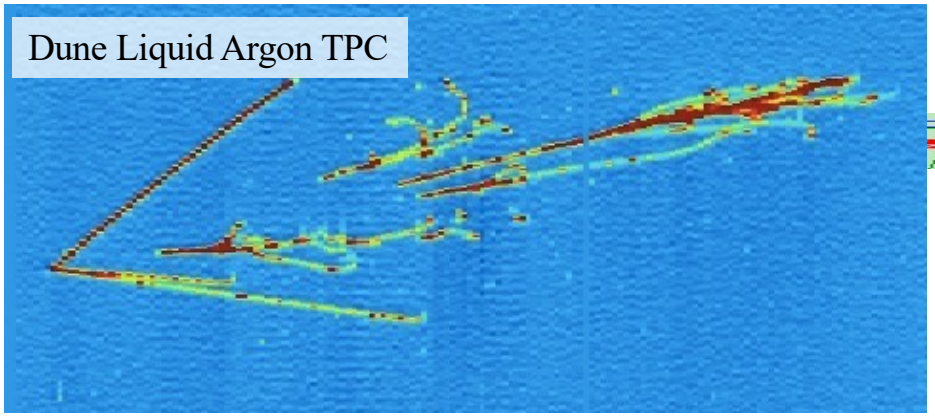
https://thispersondoesnotexist.com
https://www.whichfaceisreal.com/

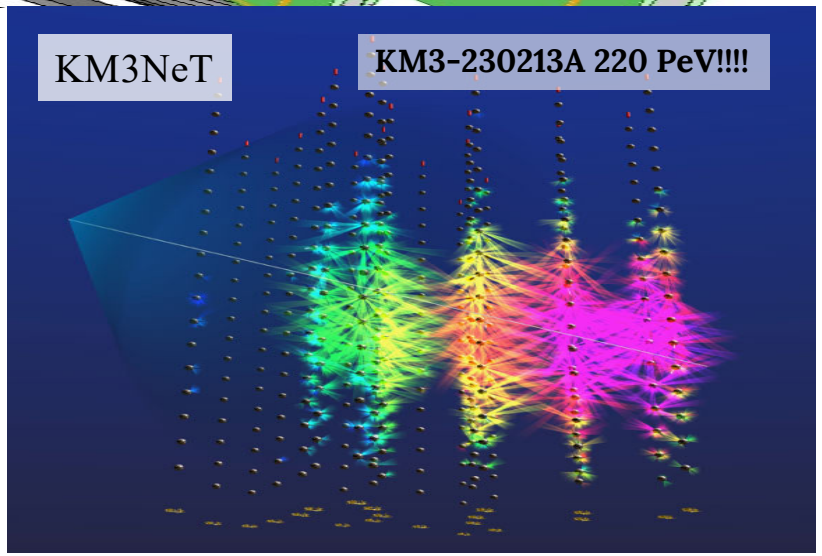
PLAY ABOUT METHODS LEARN PRESS CONTACT BOOK CALLING BS

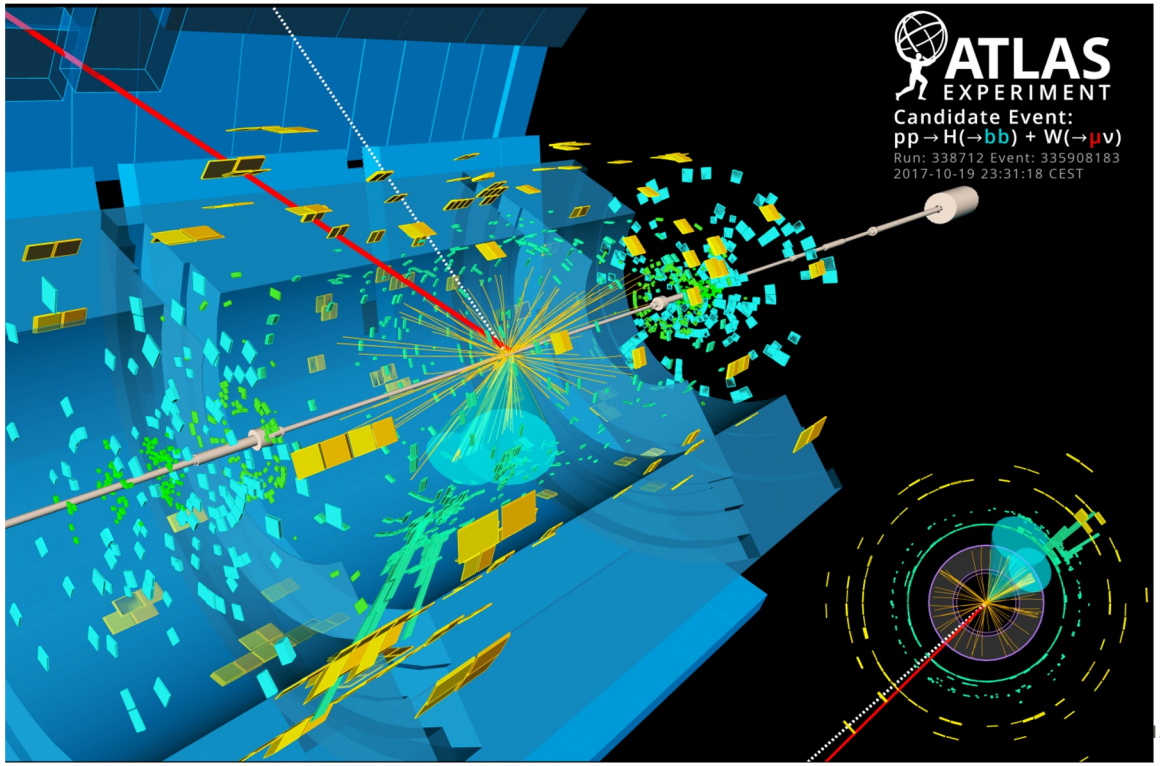
Click on the person who is real.

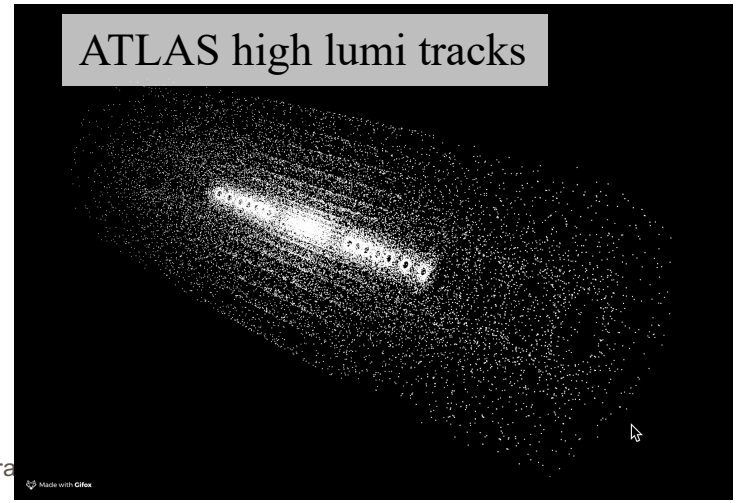
HEP ML g

Physics : an image is not the data

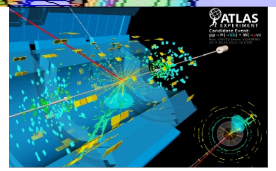


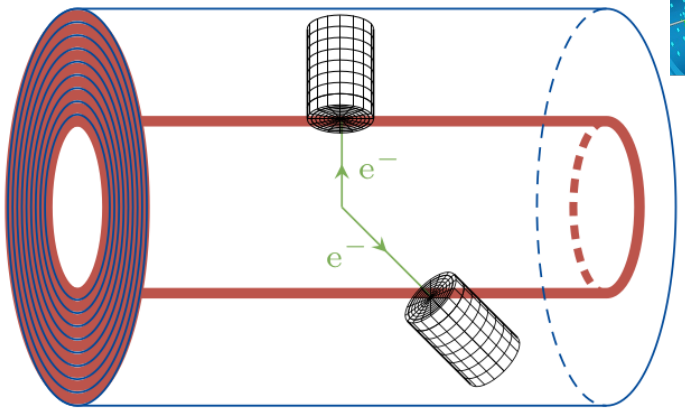


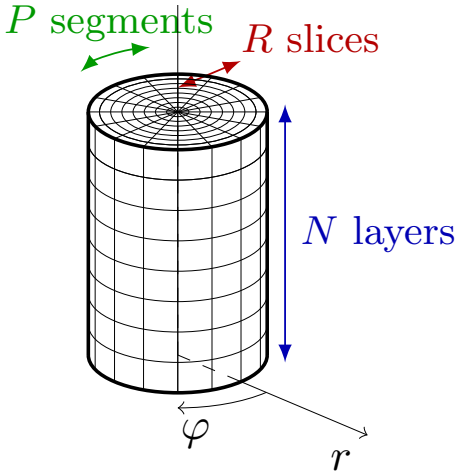


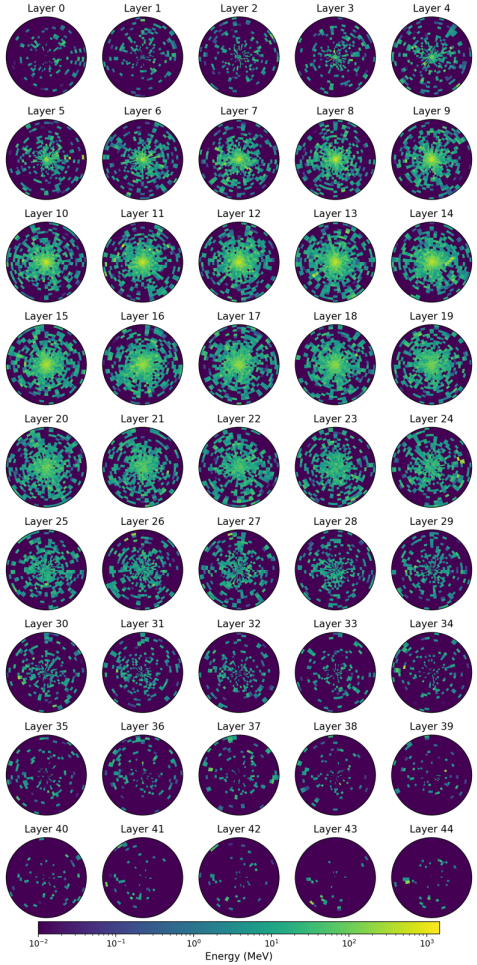


Calorimeter Simulation (CaloChallenge)





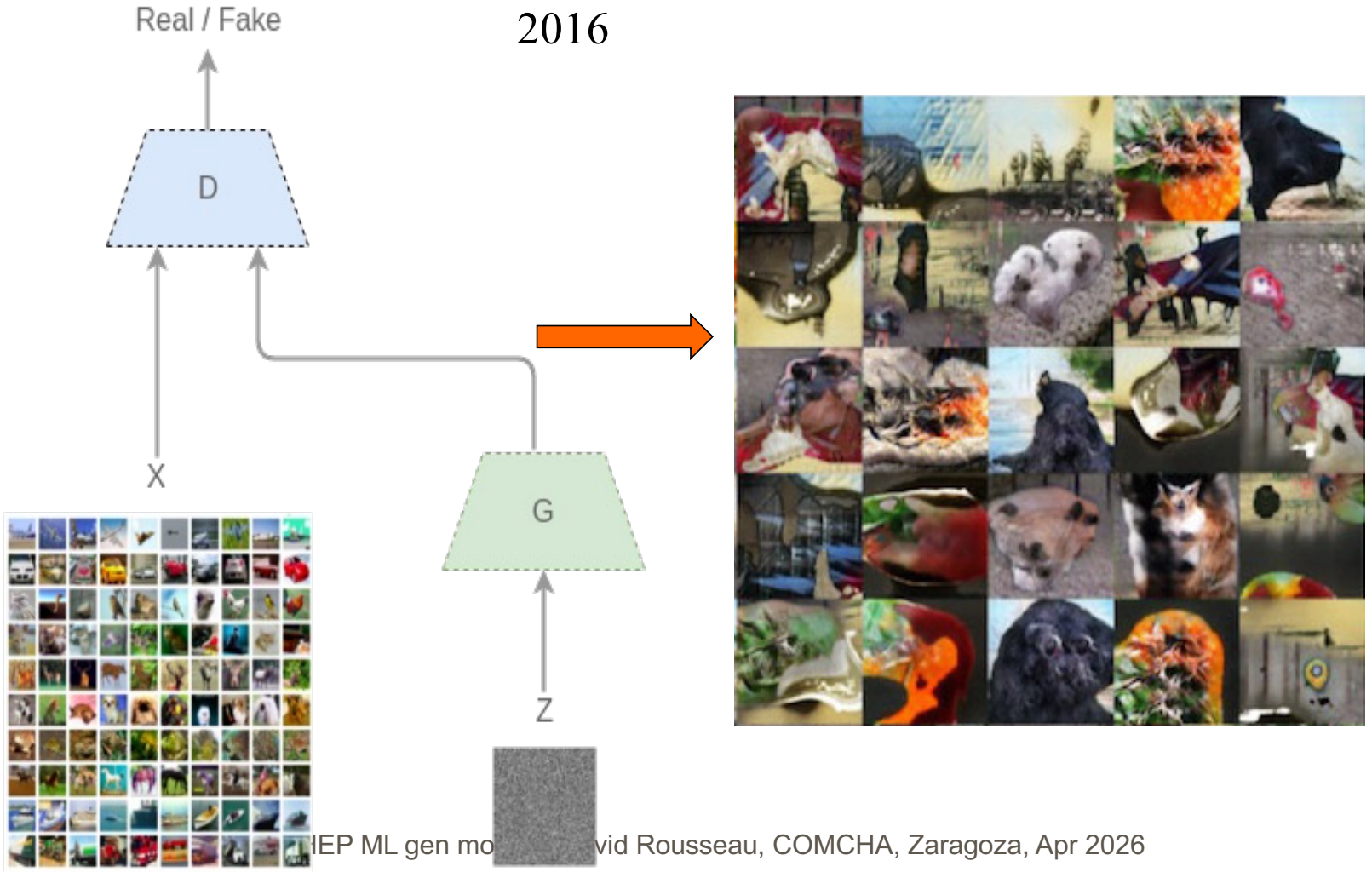




Two examples of generative models one old, one new

Generative Adversarial Network

2016

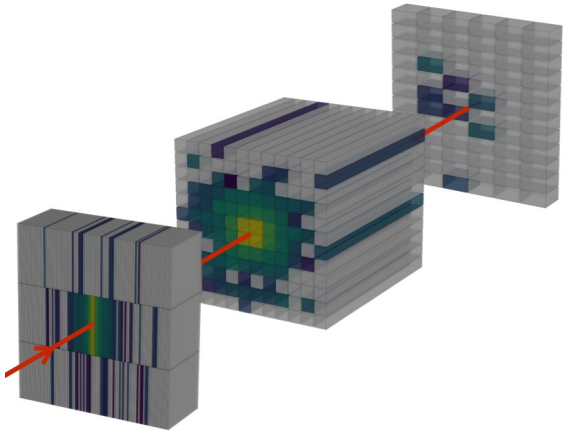


GAN for simulation

Real / Fake

D

GAN showers
(just cell energies)

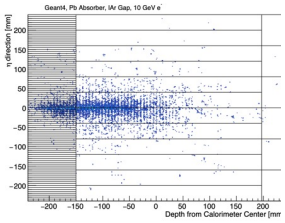


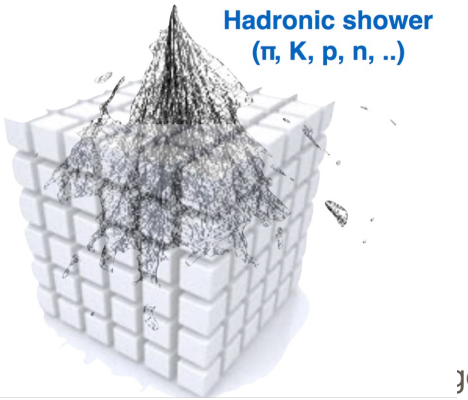
Geant4

X

G

Cells energies

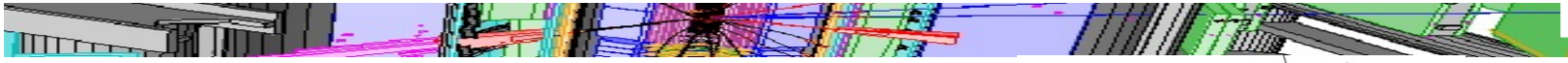




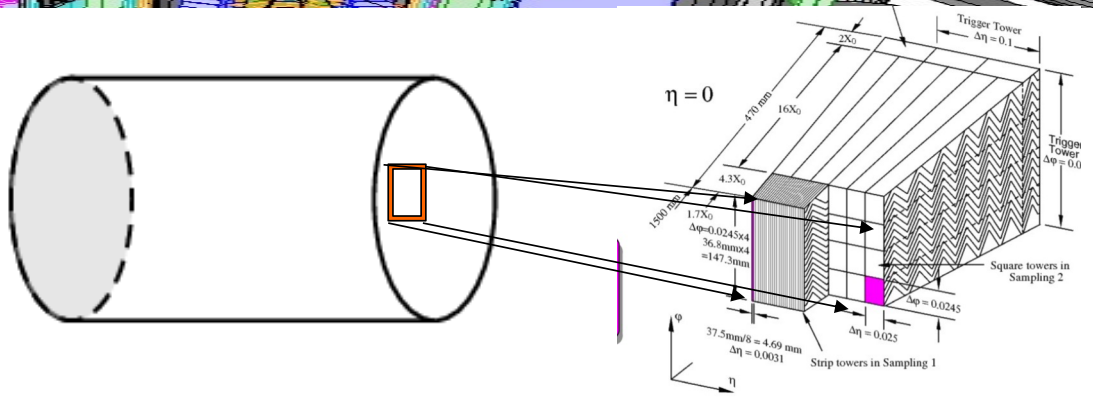
Hadronic shower
(π, K, p, n, ..)

Z

ATLAS calo simulation

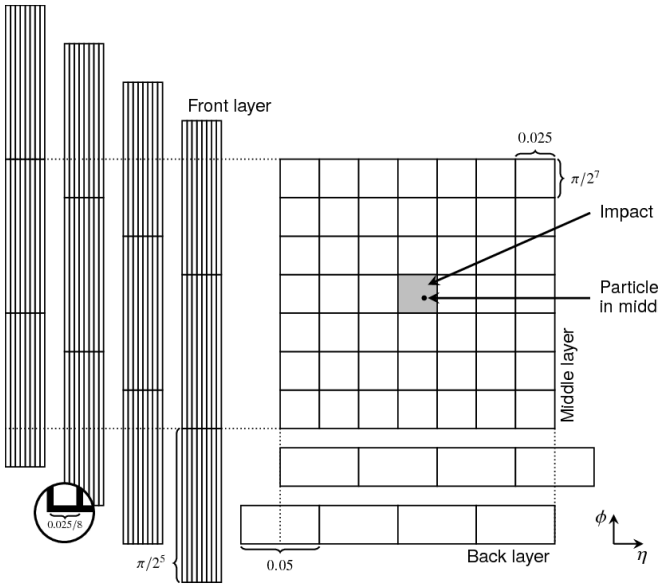


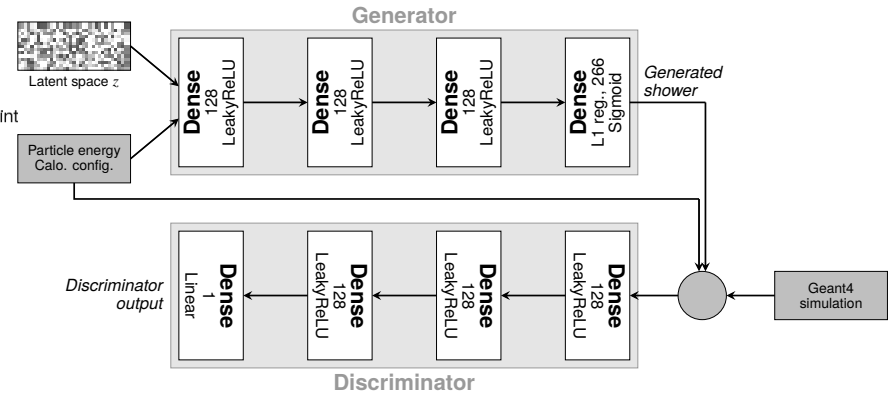
ATLAS Collaboration, arXiv:2210.06204



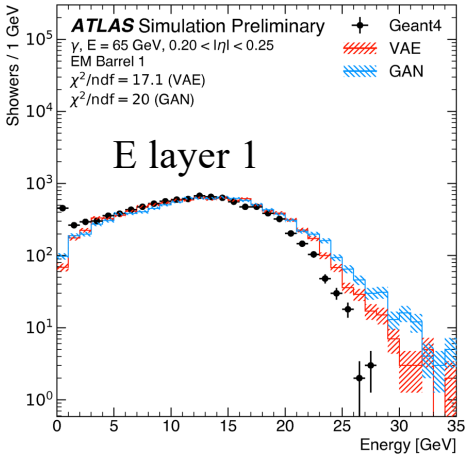
With major contribution
from Gilles Louppe U Liège

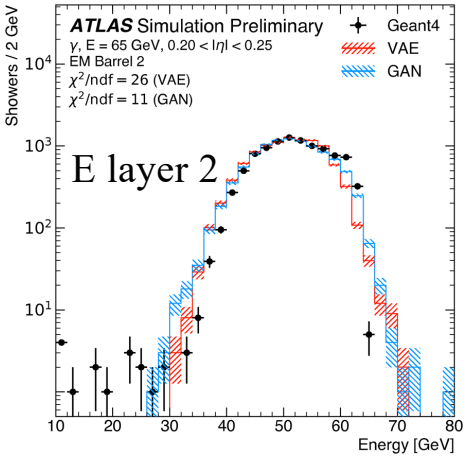
+ η, ϕ translation
177000 cells ➔266 cells

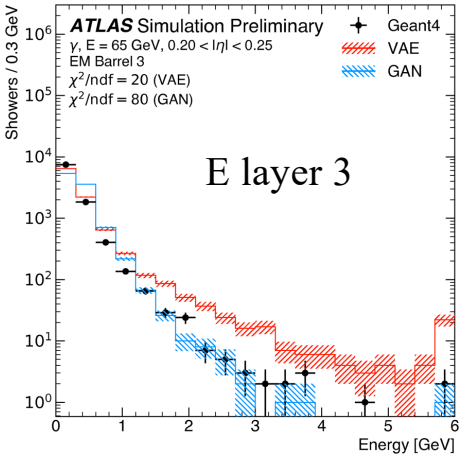


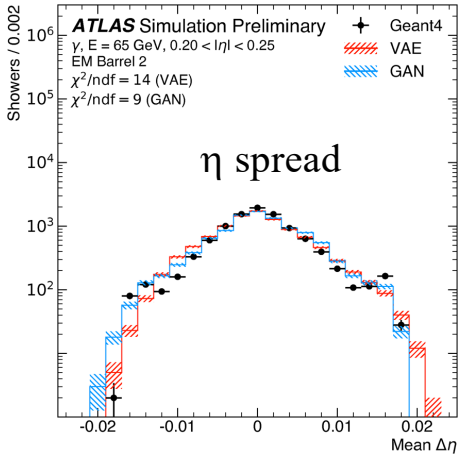


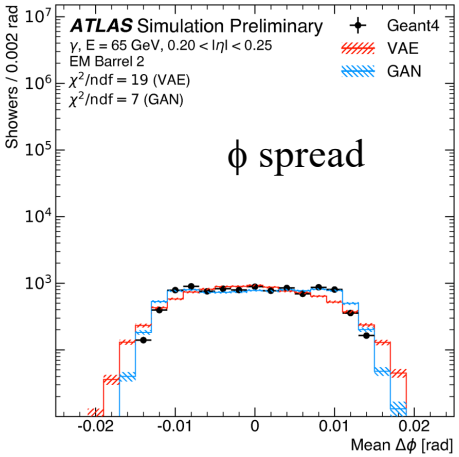
Results







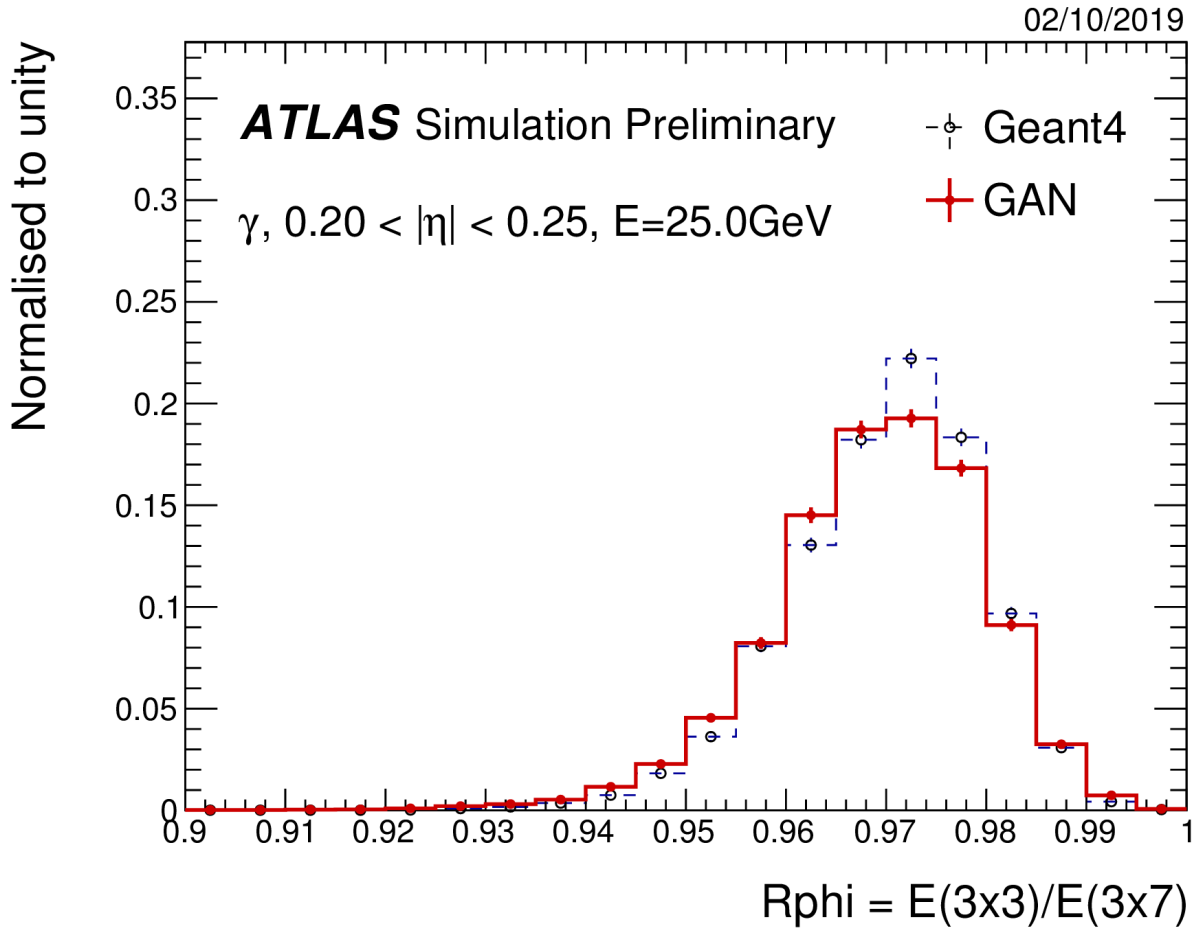




Reasonably accurate but not sufficient
Training tricky
Speed: <1ms compared to 10s

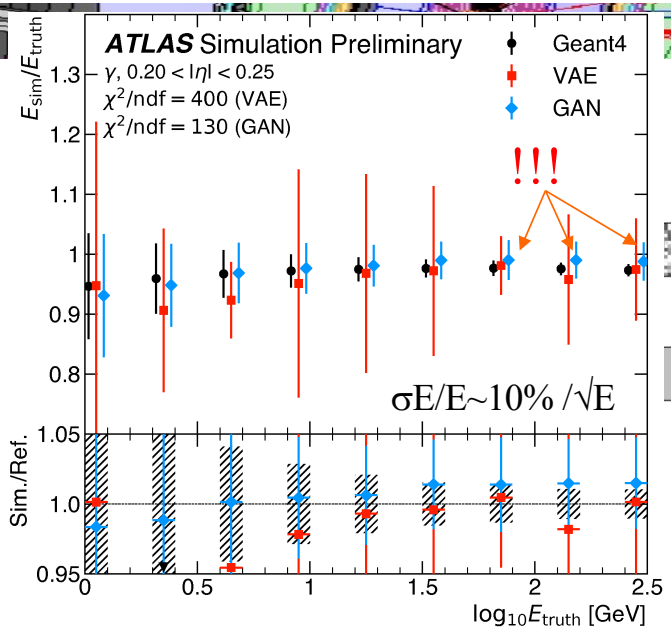
Interpolation

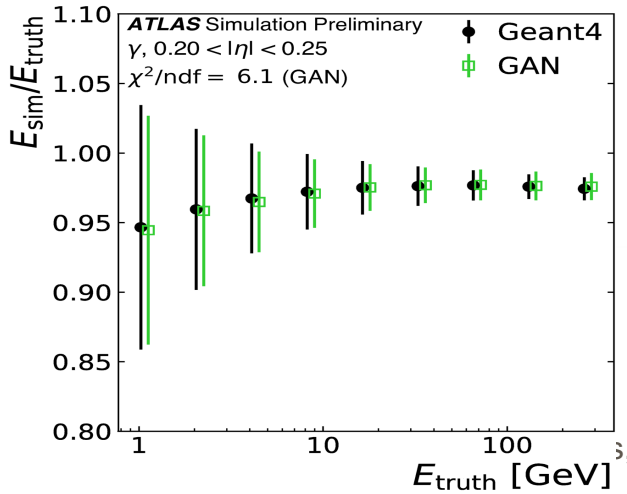
GAN trained at 1 2 4 8 16 32 65 131 262 GeV, tested at 25 GeV



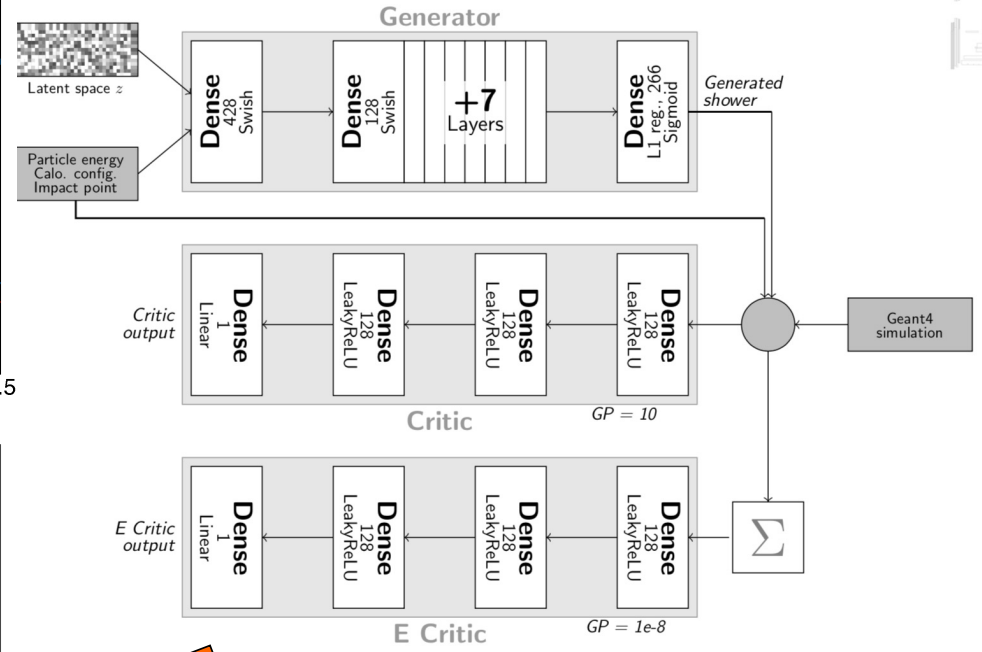
good interpolation (usually the case with Neural Networks)

Simulation of energy resolution



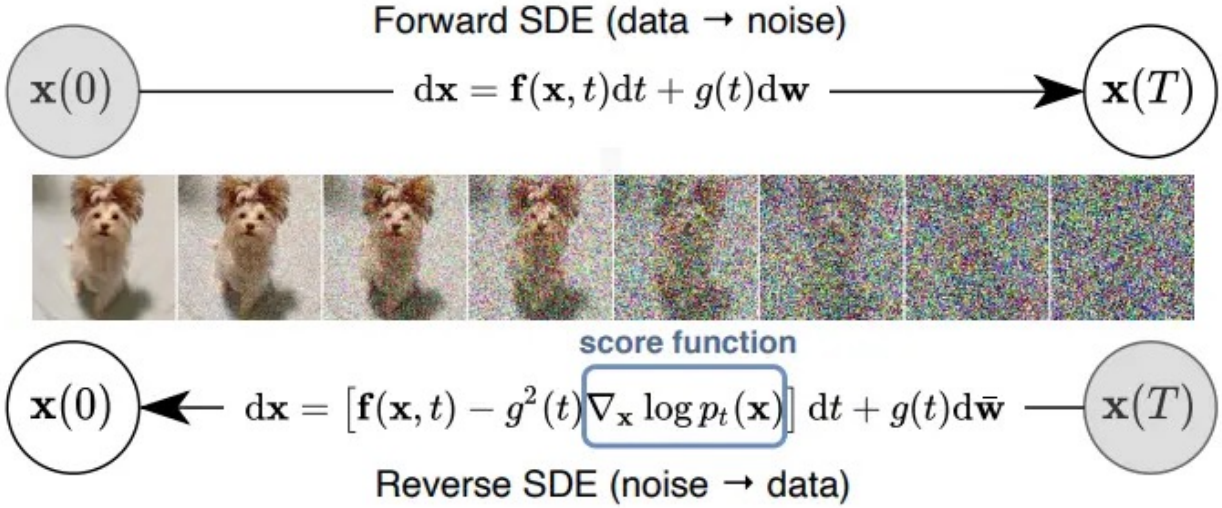


Updated architecture



So much progress since 2016

See nice introduction #Dall-e Diffusion model



an espresso machine that makes coffee from human souls, artstation

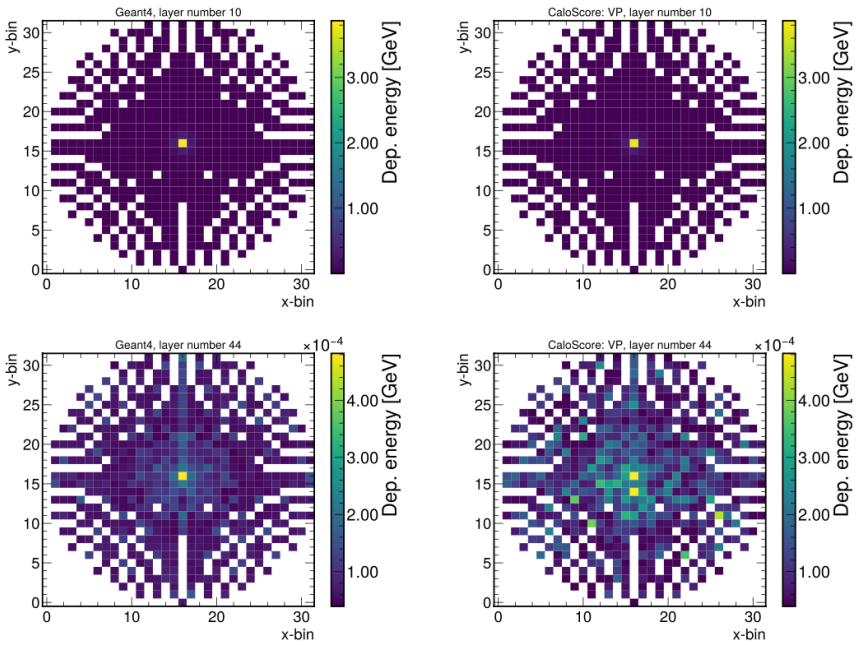
panda mad scientist mixing sparkling chemicals, artstation

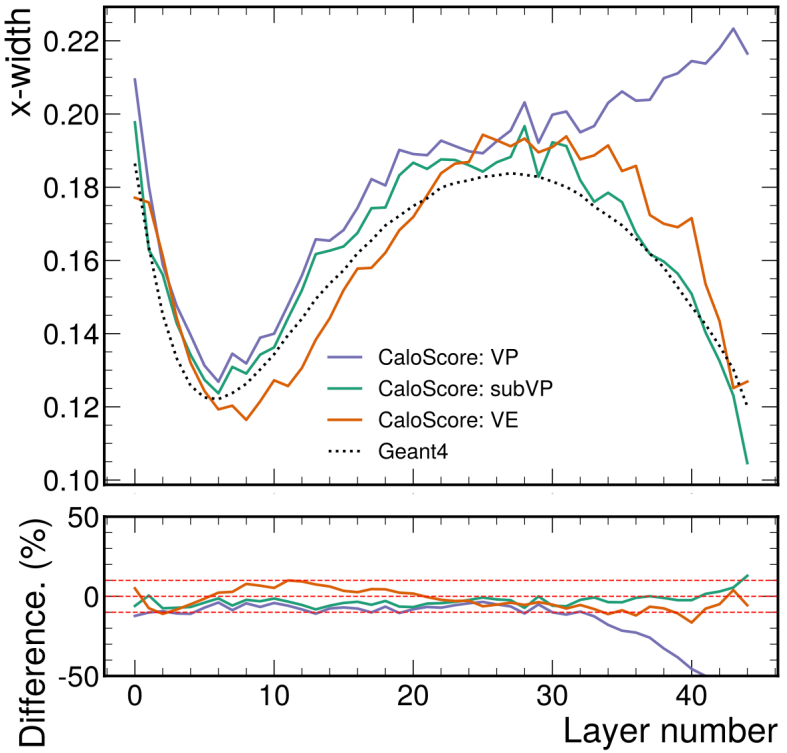
GAN becoming obsolete, can we do something (serious) with diffusion model ?

Simulation with diffusion model

Mikuli, Nachman

- On CaloChallenge dataset
- ➔seems to work, claim it is easier to train
- However inference (=generation) slower



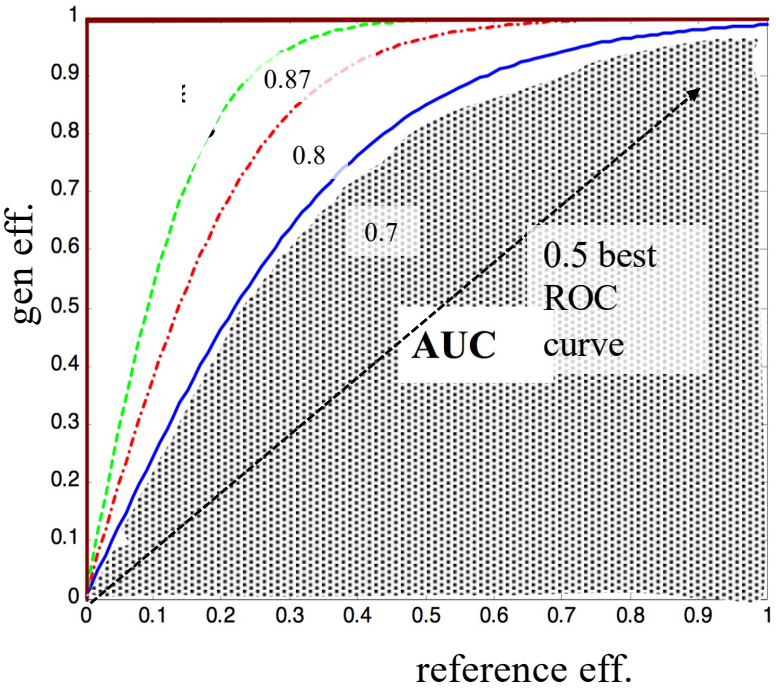


CaloChallenge

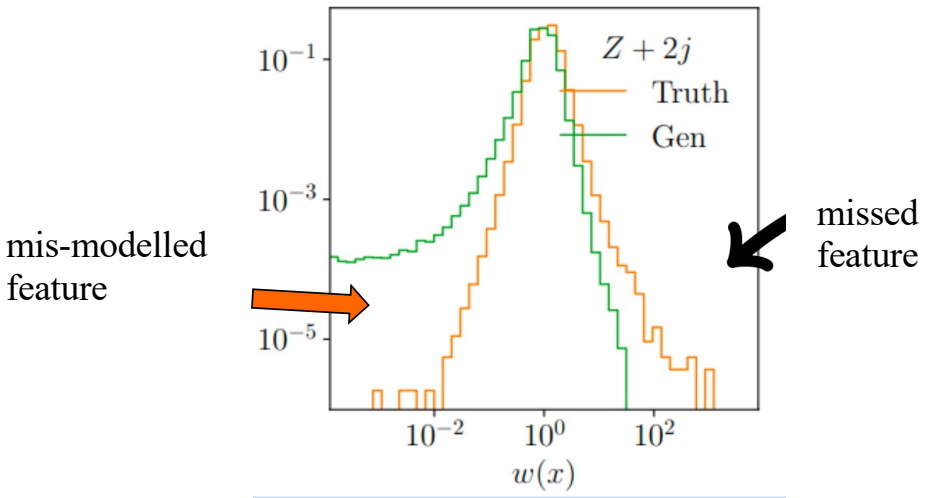
How to evaluate a generative model

- For text or image generation, subjective human evaluation « by eye »
- For science we should be more quantitative:
- train a classifier to distinguish the generative model from reference=truth :

AUC : Area Under the (ROC) Curve

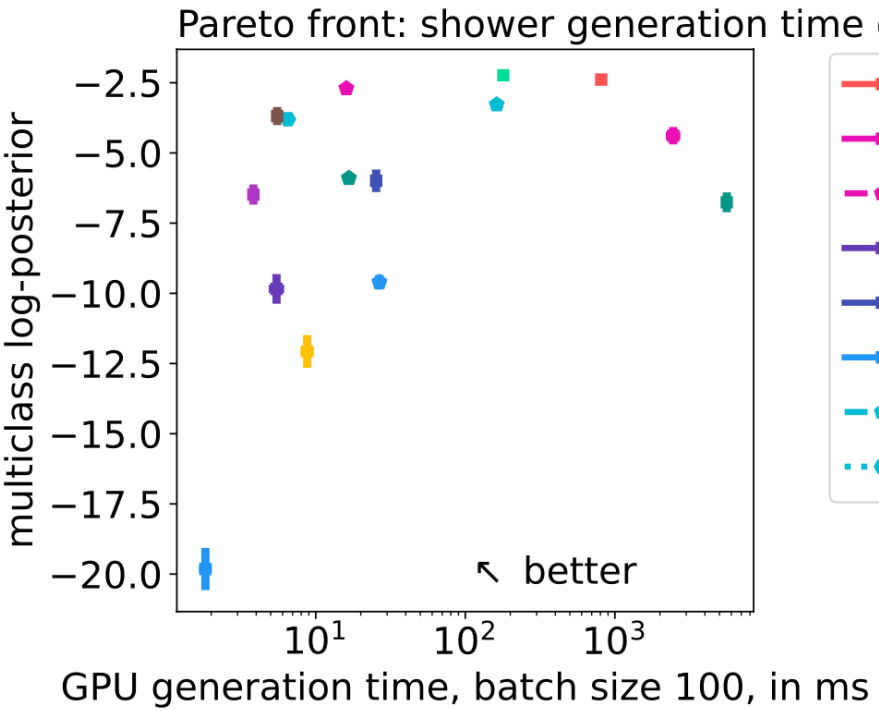


$$w = \frac{p_{gen}}{p_{truth}}$$ with calibrated classifier



CaloChallenge results

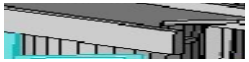
C. Krause et al, CaloChallenge 2022 final paper arXiv:2410.21611



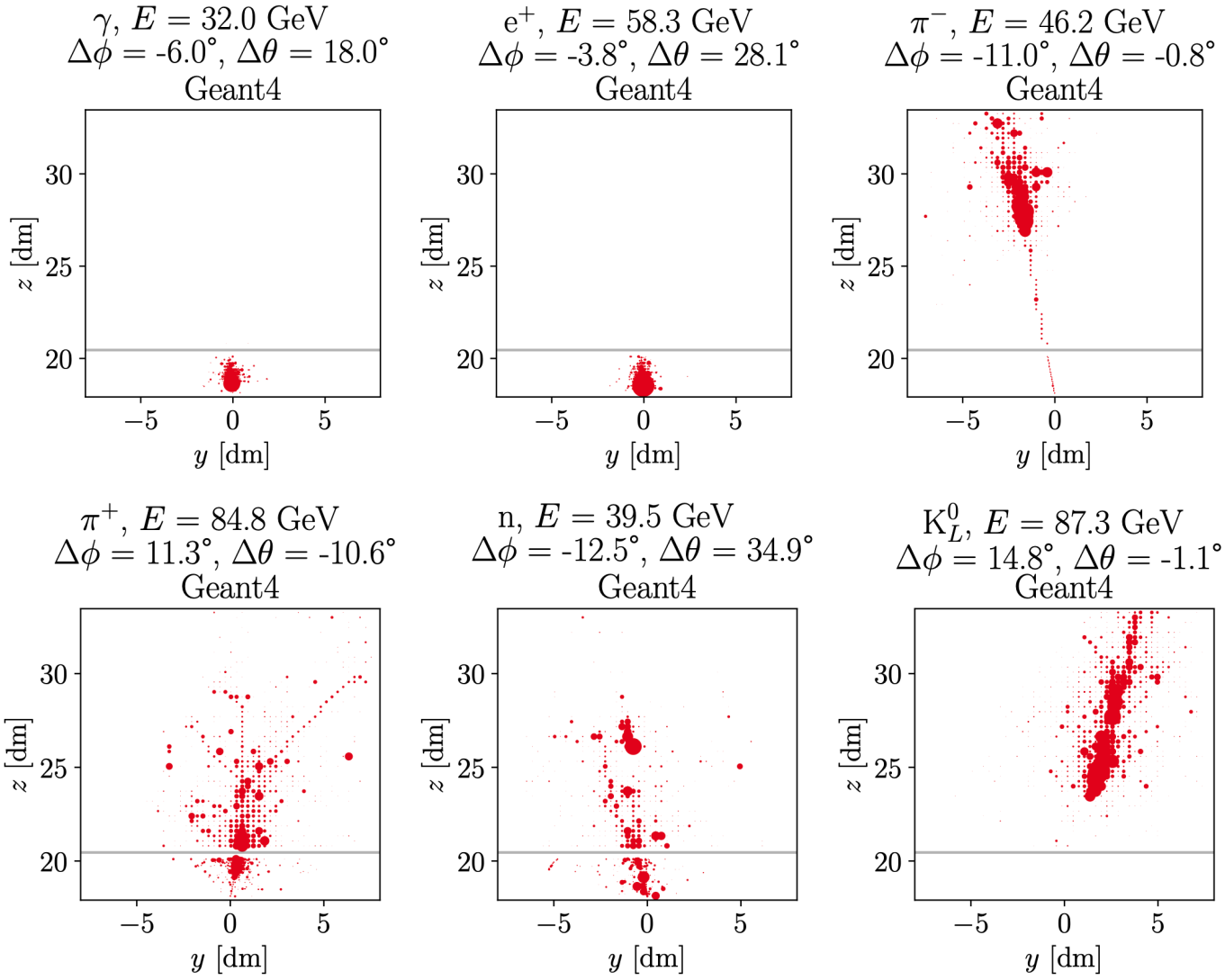
- Diffusion models more accurate but slower
- Very active sub-field
- Community building

Outlook

High granularity calorimeter (FCC)



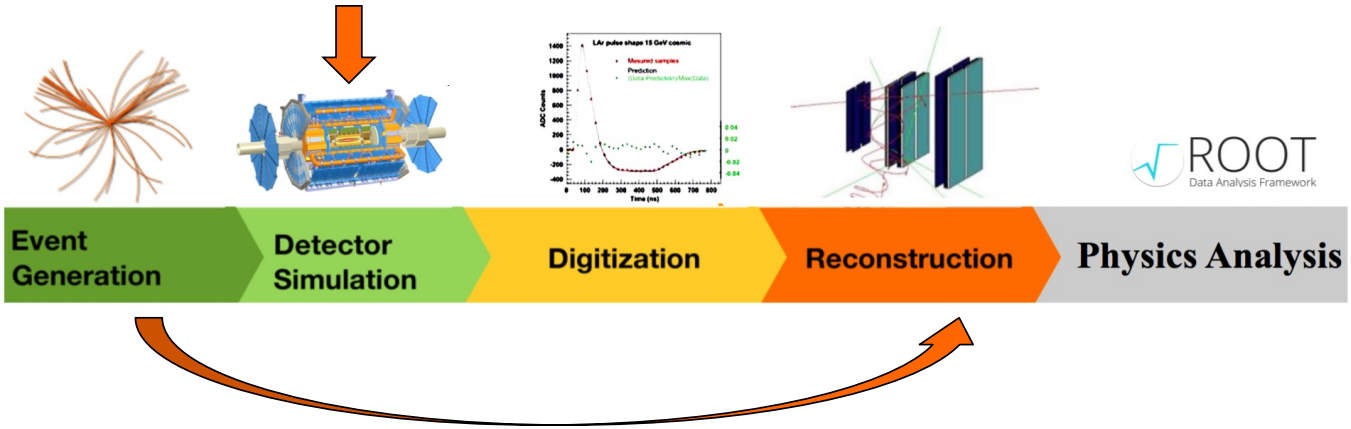
arXiv:2601.11716



Jet to reco GAN by-pass

Musella & Pandolfi arXiv 1805.00850

- Simulation strategy shown so far:
  - Replace each Geant4 showers by NN simulation
  - Still need to simulate all particles



- Why not go directly from truth to reconstruction or even Physics Analysis?

What's next

- 10 years after GAN paper, generative models are still only marginally used in production in experiments…
- Challenges:
  - What accuracy is « good enough » ?
  - Validation on high-level variables, g/e identification but also jet substructure
  - Hadron/jet are background to photon/electron➔tails of hadron/jet simulation to be accurate to 10^3 -10^4
  - Accuracy versus inference speed
  - Wide range of particle energy
  - Tricky corners of detector geometry
  - Scaffolding of simulation engine : truth to calorimeter cell, truth to analysis object