中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

ML for fast calorimeter simulation

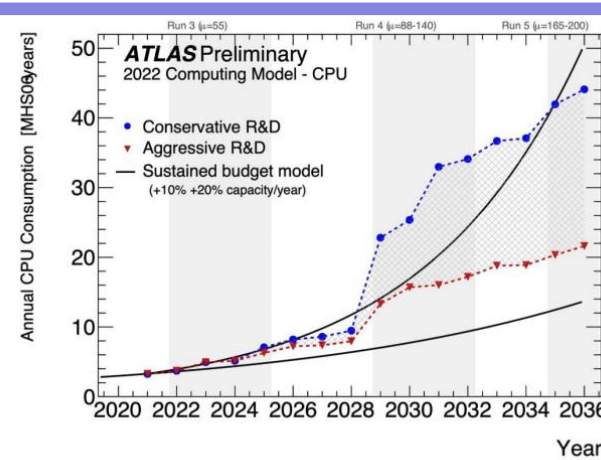
Wenxing Fang (IHEP)

IHEP ML workshop and ML group kick-off

2024.10.17

Background

- HL-LHC data challenge:
  - Event rate increased: luminosity from $2 \times 10^{34}\mathrm{s}^{-1}\mathrm{cm}^{-2}$ to $7.5 \times 10^{34}\mathrm{s}^{-1}\mathrm{cm}^{-2}$
  - Event size increased: finer-grained detector readout
  - Aggressive R&D is needed, otherwise the resources will be a problem
- The MC simulation takes most CPU resources, dominated by the detector simulation of calorimeter
  - Traditional fast calorimeter simulation methods: shower parameterization (PCA), frozen shower, …
  - In recent years, ML based methods show significant promise as a replacement



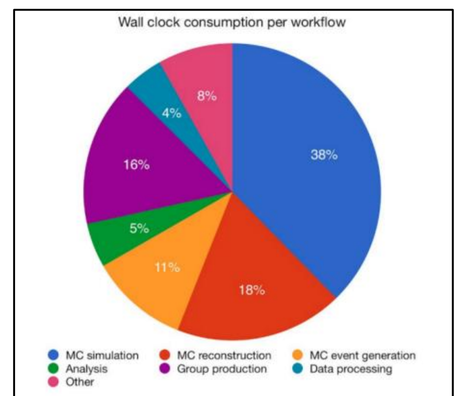
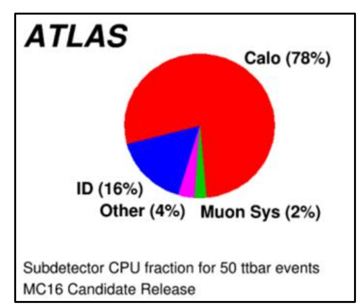
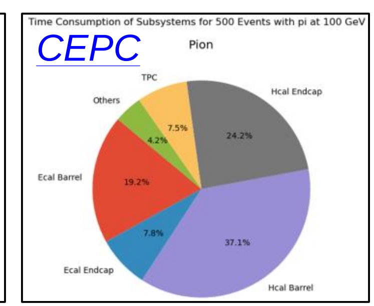


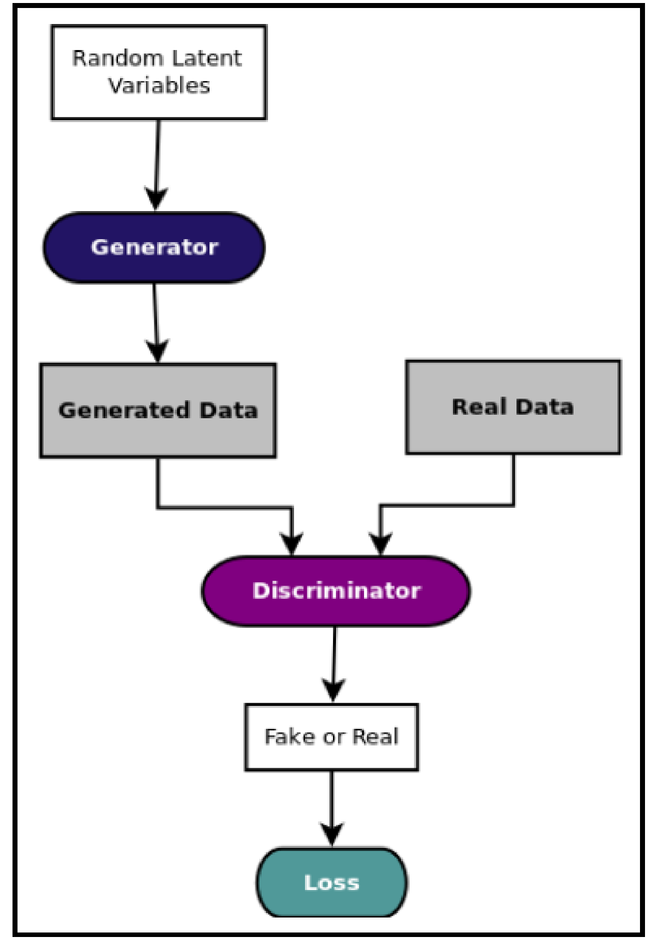
Figure 1: ATLAS CPU hours used by various activities in 2018





Generative Adversarial Networks ([GAN]())

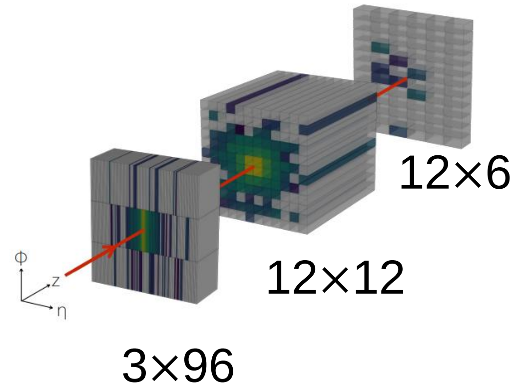
- Discriminator tries to discriminate the real data and generated data
- The generator tries to produce generated data which can confuse the discriminator
- At the end of training, the discriminator can not discriminate the real or generated data. The generator learns the true underlying data distribution

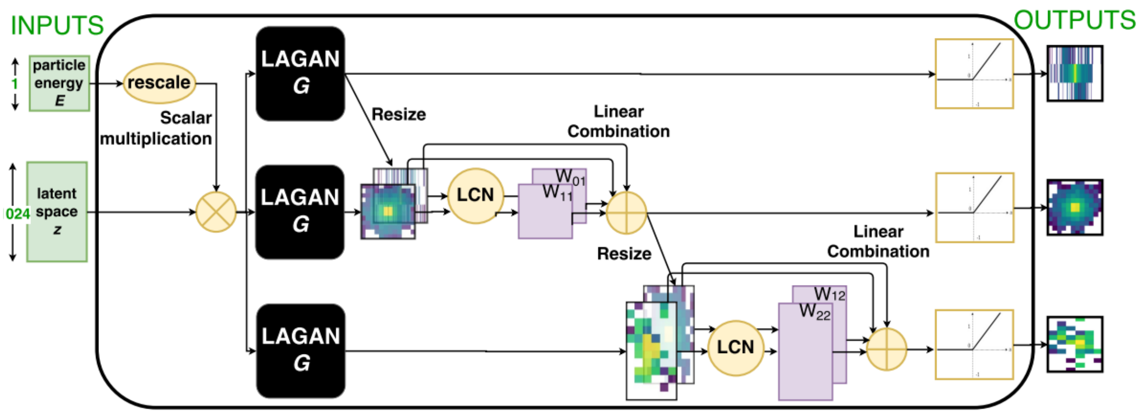


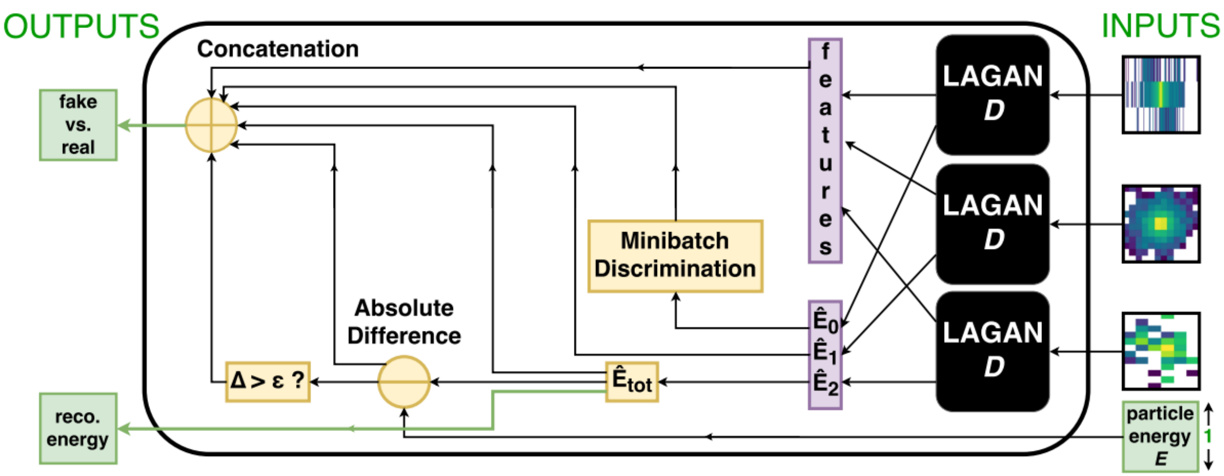
$$\min_{G} \max_{D} V(D,G) = E_{x\sim p_{data}(x)}[logD(x)] + E_{z\sim p_z(z)}[log(1-D(G(z)))]$$

CaloGAN

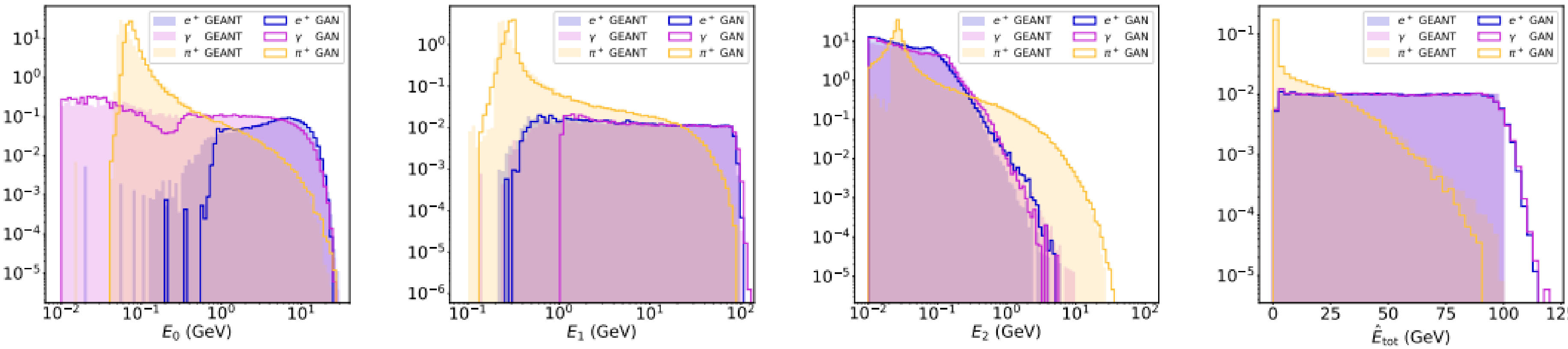
- The CaloGAN (2017) achieved a fast calorimeter simulation based on GAN







CaloGAN performance



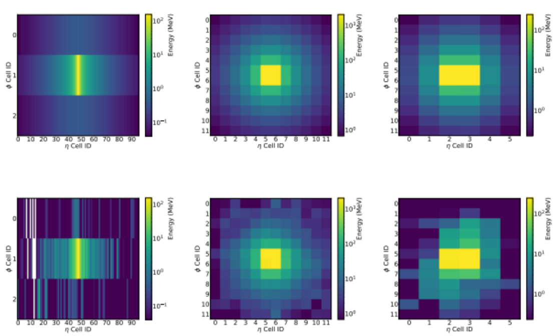


FIG. 8: Average π^+ GEANT4 shower (top), and average π^+ CALOGAN shower (bottom), with progressive calorimeter depth (left to right).

Simulator	**Hardware**	**Batch Size**	**ms/shower**
GEANT4	CPU	N/A	1772
CALOGAN	CPU	1	13.1
		10	5.11
		128	2.19
		1024	2.03
	GPU	1	14.5
		4	3.68
		128	0.021
		512	0.014
		1024	0.012

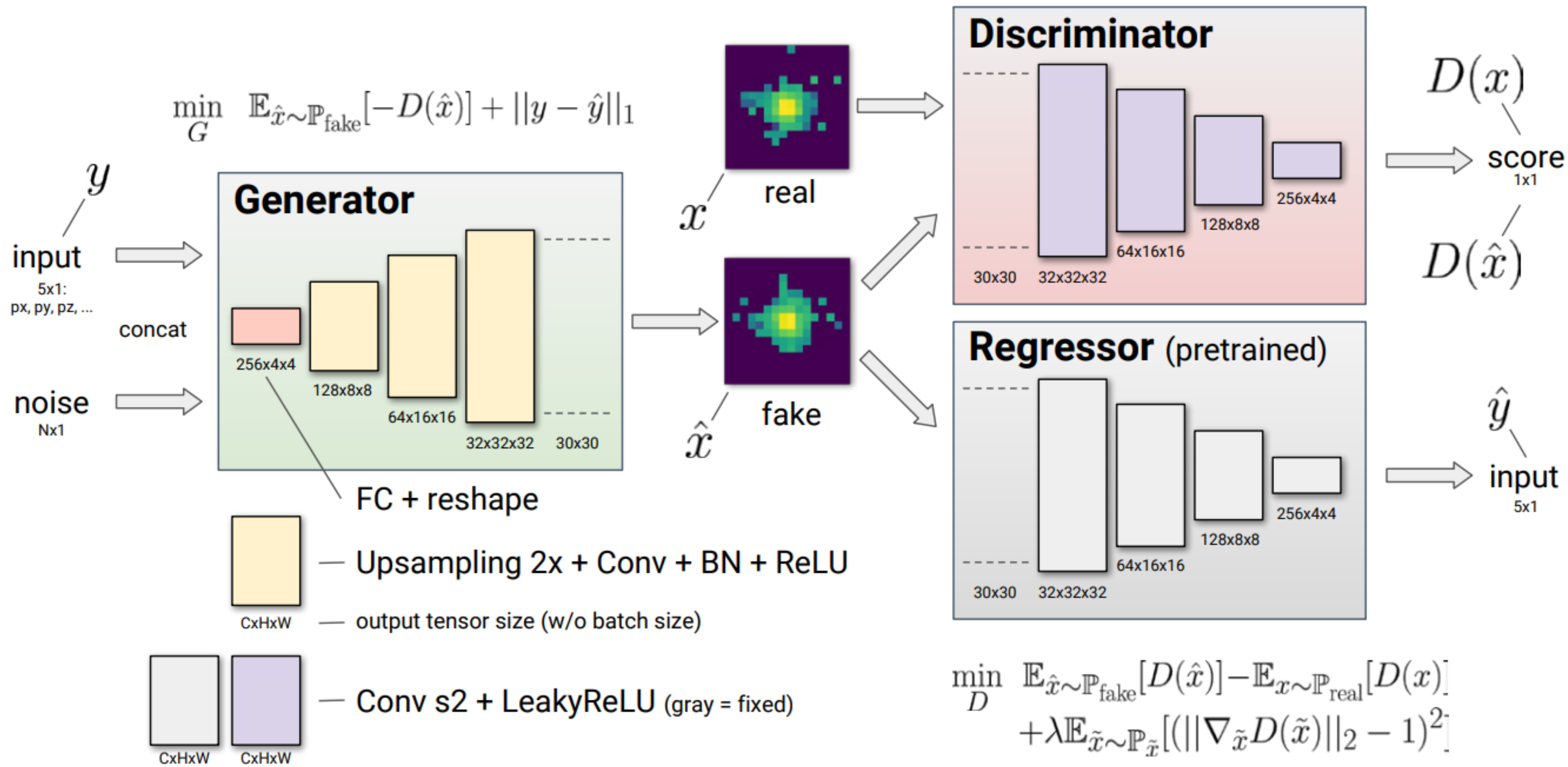
e^+ vs. π^+

		Test on	
		GEANT4	CALOGAN
Train on	GEANT4	99.6% ± 0.1%	96.5% ± 1.1%
	CALOGAN	98.2% ± 0.9%	99.9% ± 0.2%

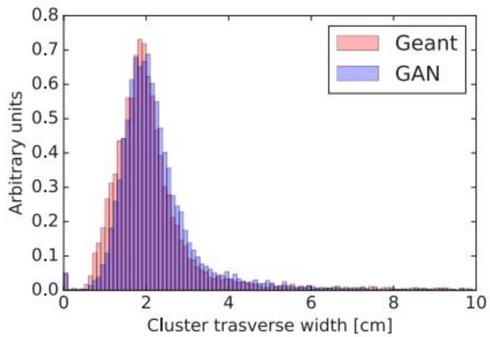
e^+ vs. γ

		Test on	
		GEANT4	CALOGAN
Train on	GEANT4	66.1% ± 1.2%	70.6% ± 2.6%
	CALOGAN	54.3% ± 0.8%	100.0% ± 0.0%

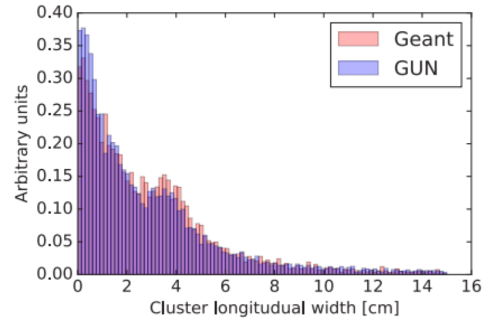
The LHCb case



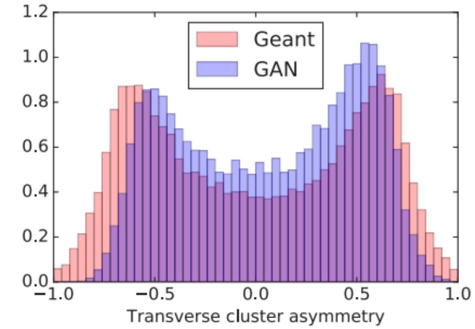
The LHCb case (performance)



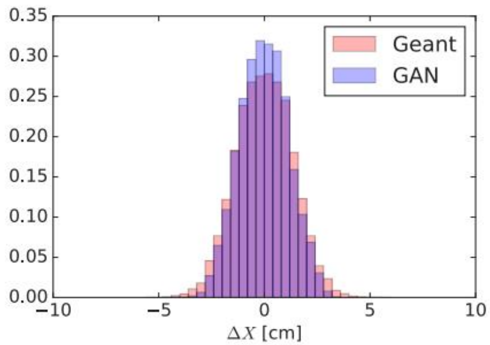
(a) The transverse width of real and generated clusters



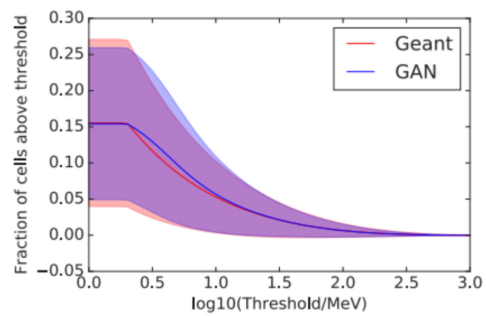
(b) The longitudinal width of real and generated clusters



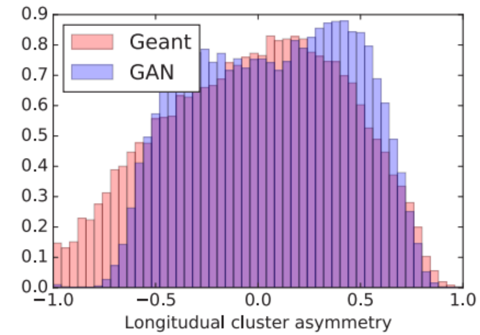
(e) The transverse asymmetry of real and generated clusters



(c) ΔX between cluster center of mass and the true particle coordinate



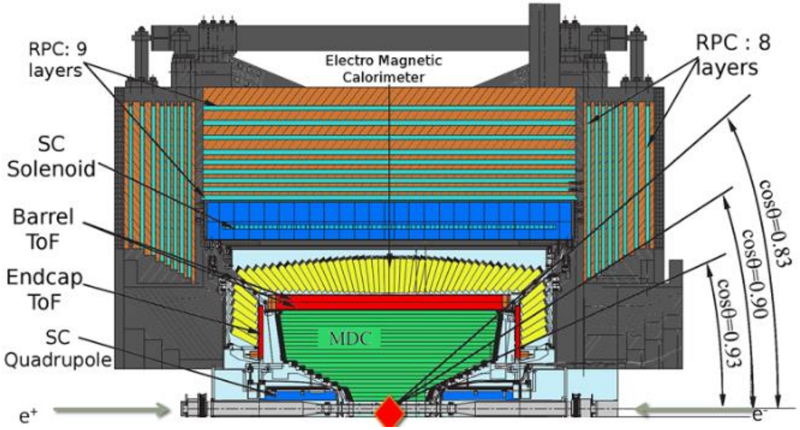
(d) The sparsity of real and generated clusters



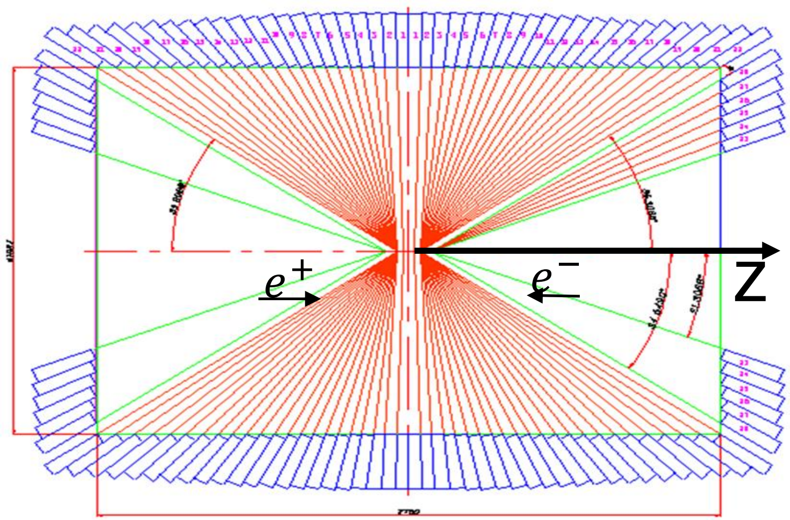
(f) The longitudinal asymmetry of real and generated clusters

BESIII experiment

- The BESIII experiment focuses on physics at the tau-charm region, such as non-perturbative QCD, exotic hadrons, BSM, …



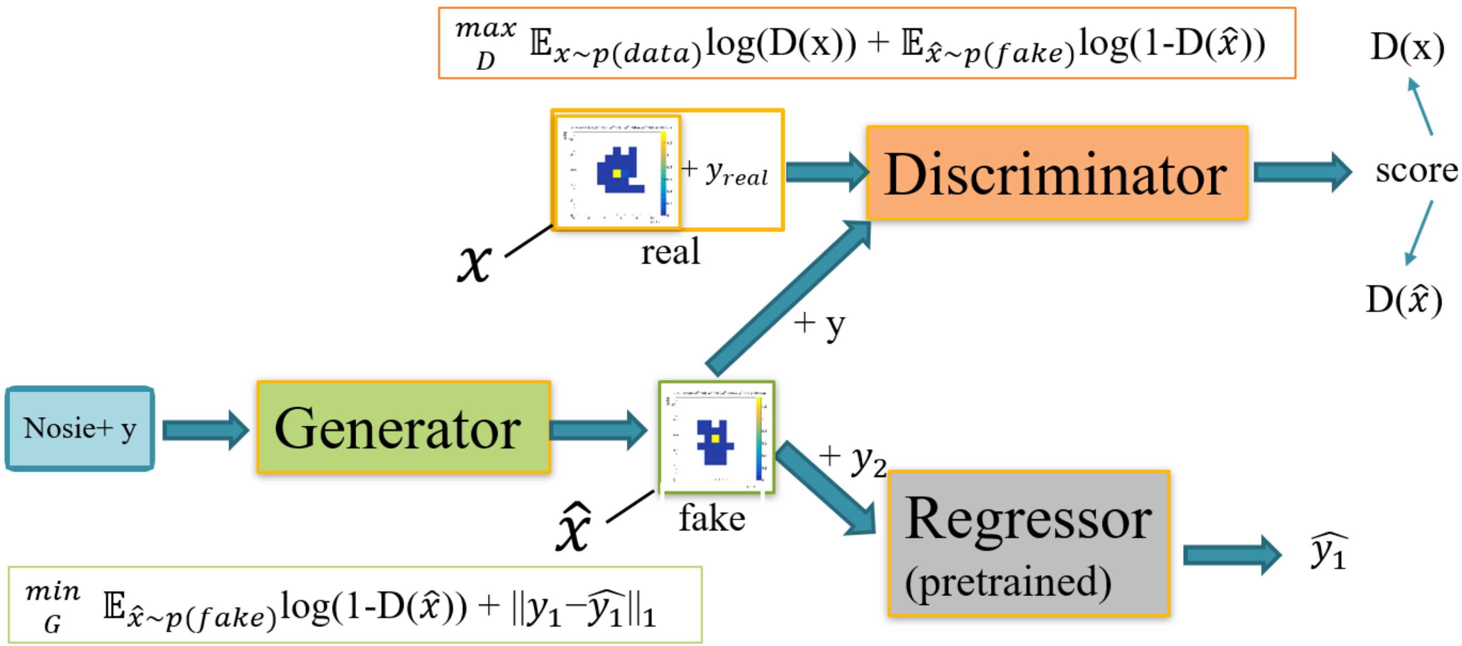
- BESIII EMC:
  - 44 rings of crystal in barrel and 120 crystals in each ring. The front size of each crystals is 5×5 cm^2
  - 6 rings of crystal in each endcap



- Apply GAN for EMC barrel fast simulation for $\mathrm{e}^{\pm}$

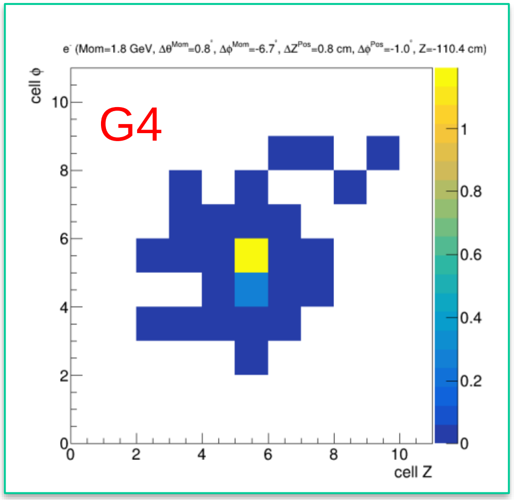
GAN model for the BESIII

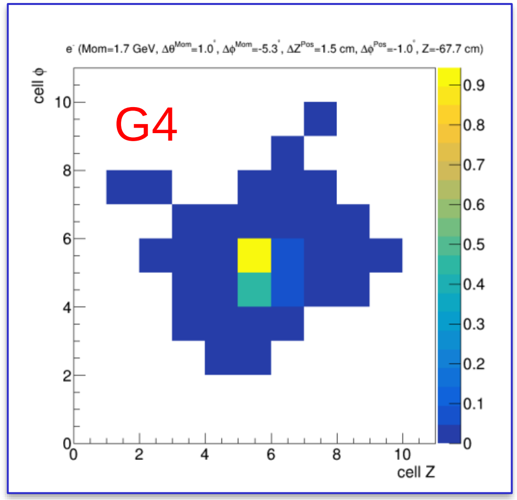
- The structure is similar to the LHCb one

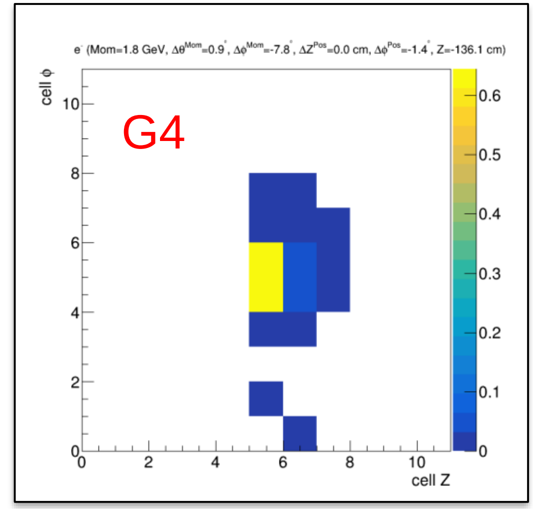


- The y ($y_1 + y_2$) contains the momentum of particle and the relative position and angular between the particle and the calorimeter.
  - y_1
    - Momentum: the momentum of the particle.
    - $\Delta\phi^{\mathrm{Mom}}$: the ϕ difference between the momentum of incoming particle and the direction of the crystal.
    - $\Delta\theta^{\mathrm{Mom}}$: the θ difference between the momentum of incoming particle and the direction of the crystal.
  - y_2
    - ΔZ^{Pos}: the Z difference between the hit point of incoming particle and the z of front center of the crystal.
    - $\Delta\phi^{\mathrm{Pos}}$: the ϕ difference between the hit point of incoming particle and the ϕ of front center of the crystal.
    - Z: the Z value of hit point

Event display (e^-)



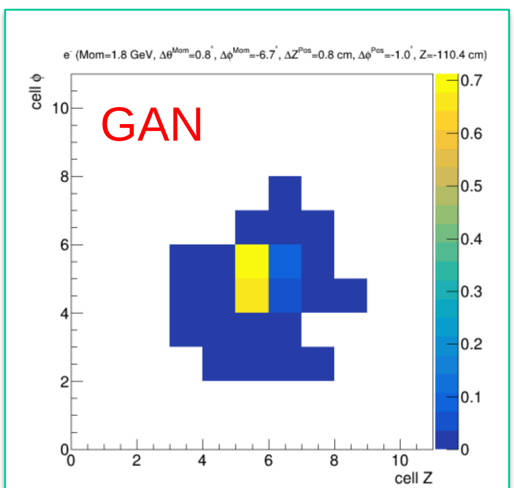


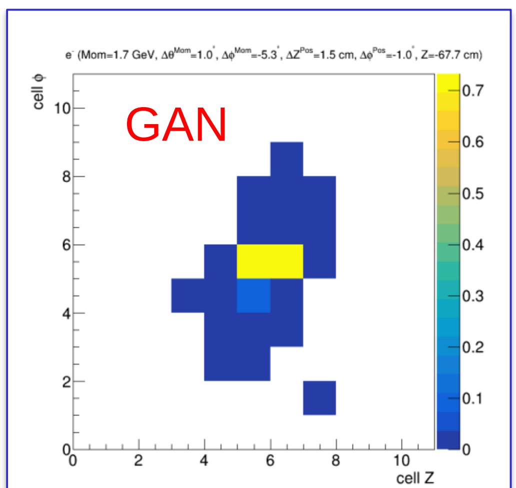


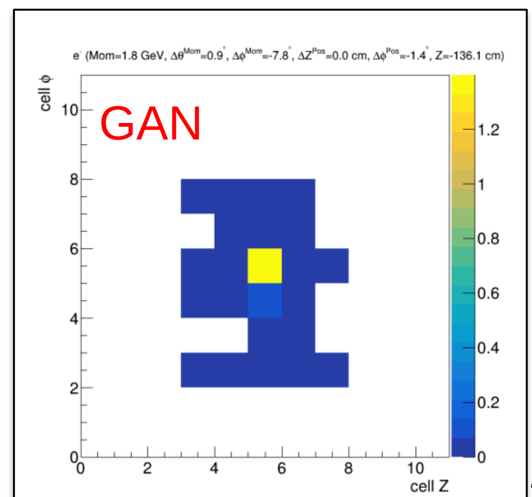
e^-(Mom = 1.8 GeV, $\Delta\theta^{\text{Mom}}$= 0.8°, $\Delta\phi^{\text{Mom}}$= −6.7°, ΔZ^{Pos} = 0.8 cm, $\Delta\phi^{\text{Pos}}$ = −1.0°, Z = −110.4 cm)

e^-(Mom = 1.7 GeV, $\Delta\theta^{\text{Mom}}$= 1.0°, $\Delta\phi^{\text{Mom}}$= −5.3°, ΔZ^{Pos} = 1.5 cm, $\Delta\phi^{\text{Pos}}$ = −1.0°, Z = −67.7 cm)

e^-(Mom = 1.8 GeV, $\Delta\theta^{\text{Mom}}$= 0.9°, $\Delta\phi^{\text{Mom}}$= −7.8°, ΔZ^{Pos} = 0. cm, $\Delta\phi^{\text{Pos}}$ = −1.4°, Z = −136.1 cm)



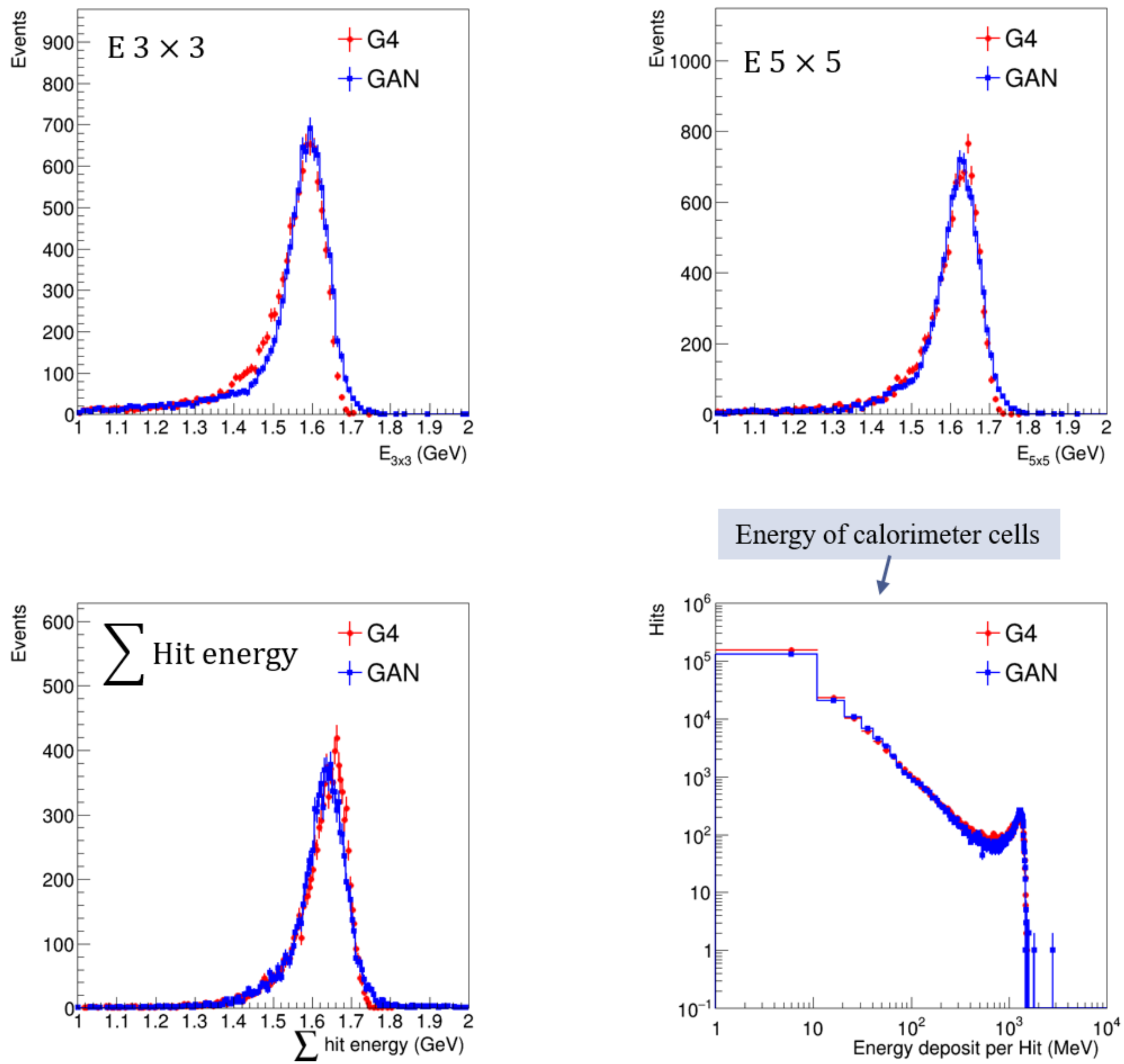




The BESIII case (performance)

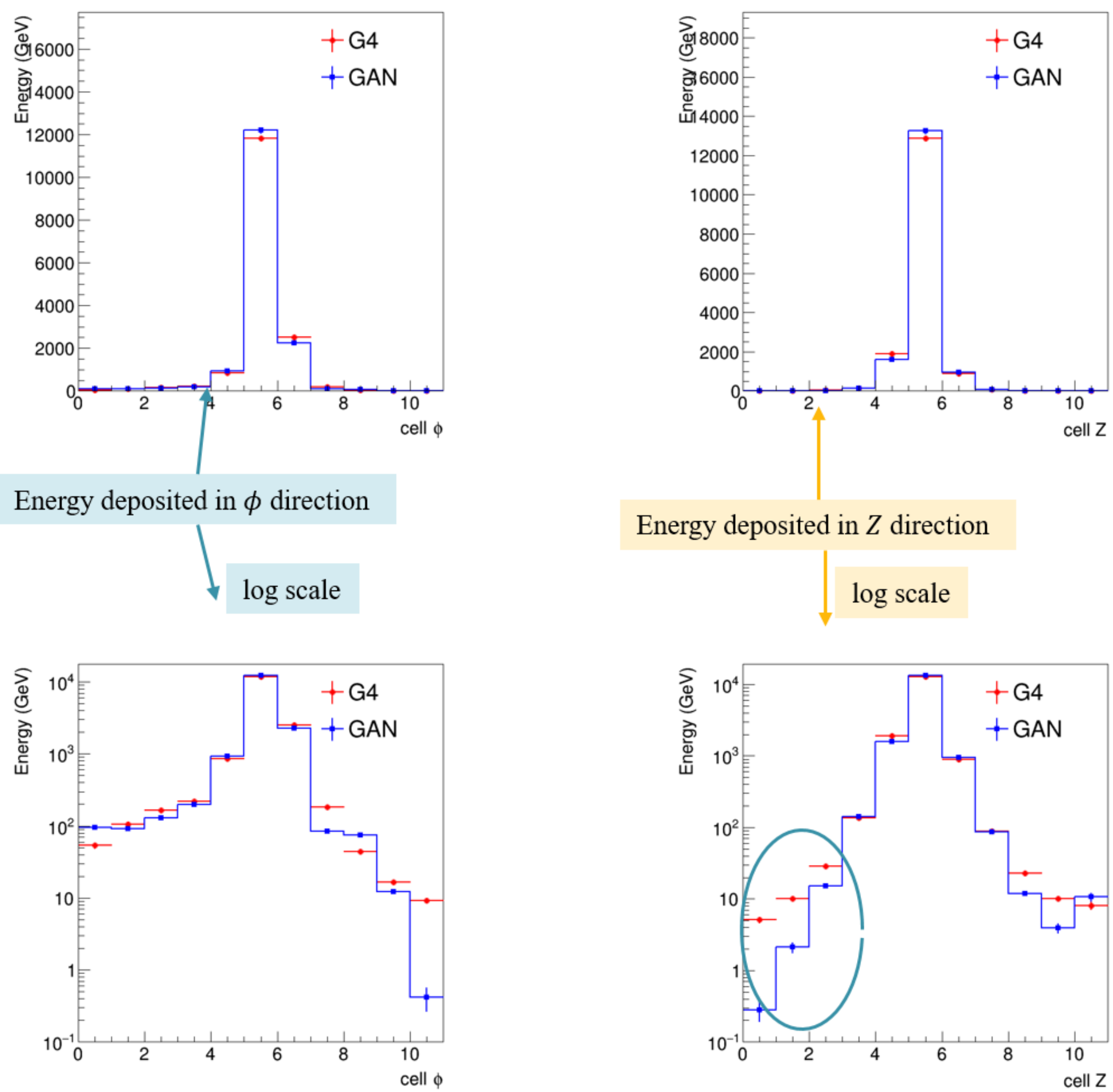
Dataset:
- MC Bhabha events

e^-



The BESIII case (performance)

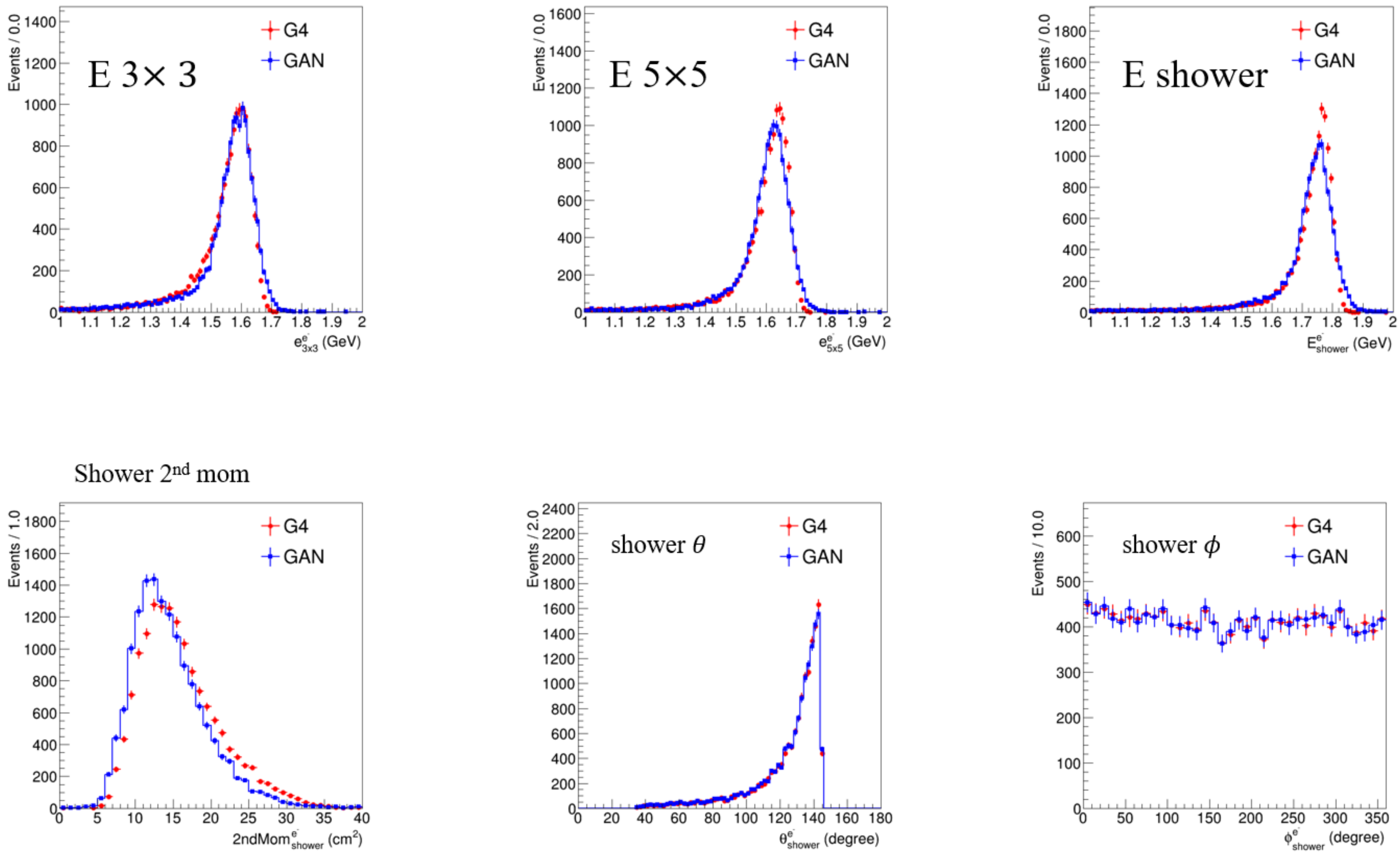
e^-



The BESIII case (performance)

e^{-}

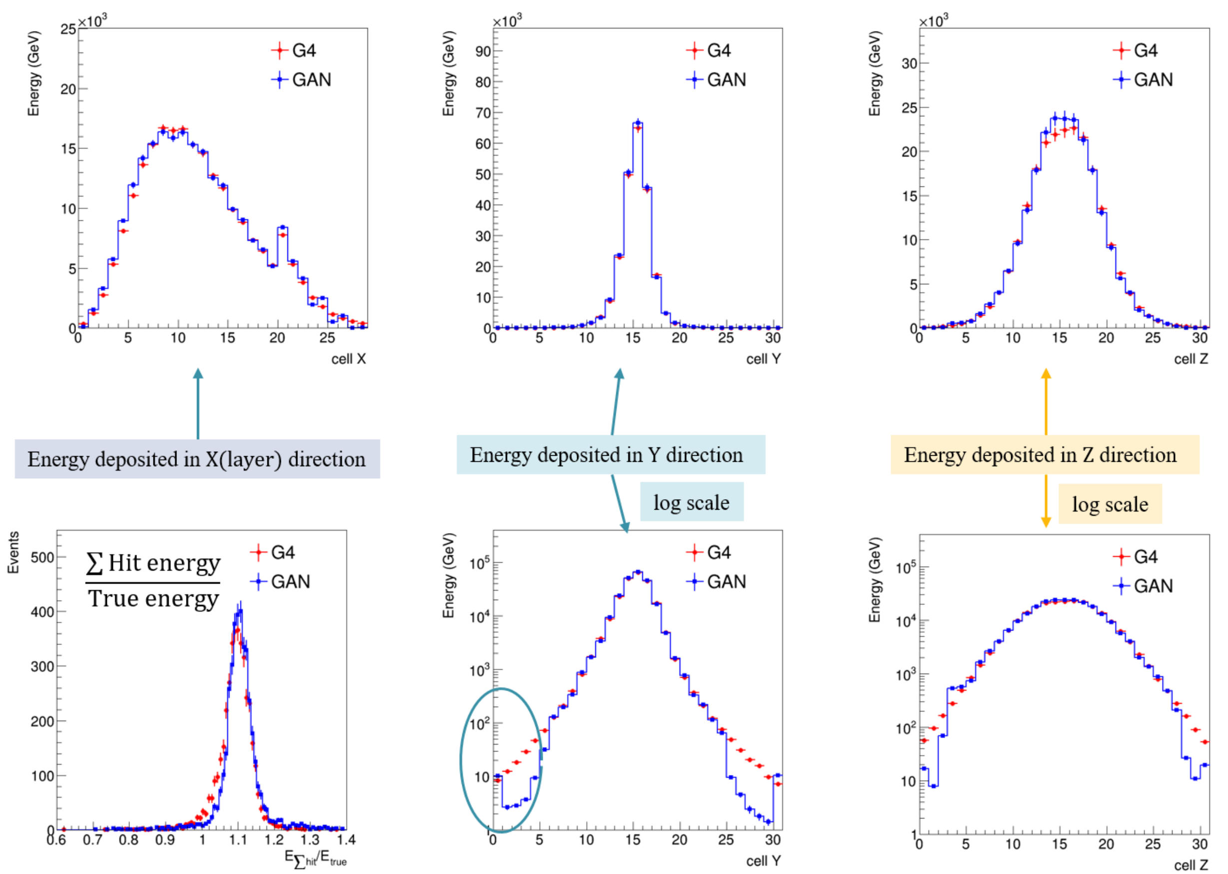
- Apply the GAN simulation in BESIII offline software



For CEPC experiment

[CEPC]

❖ Apply the model for silicon-tungsten ECAL (29 layers). The model is extended from 2D to 3D (mainly replace 2D convolution operation by 3D convolution)



Dataset:

- photon showers in ECAL Barrel
- 31x31x29 voxels

Summary

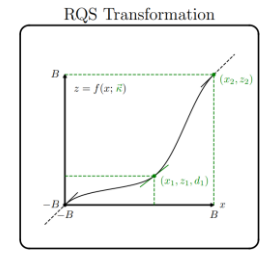
- ML based fast simulation calorimeter is widely studied in different experiments
- Many promising results and also challenges
- The field is in a rapid development stage, e.g. using the latest diffusion model (CaloDiffusion)
- Please stay tuned!

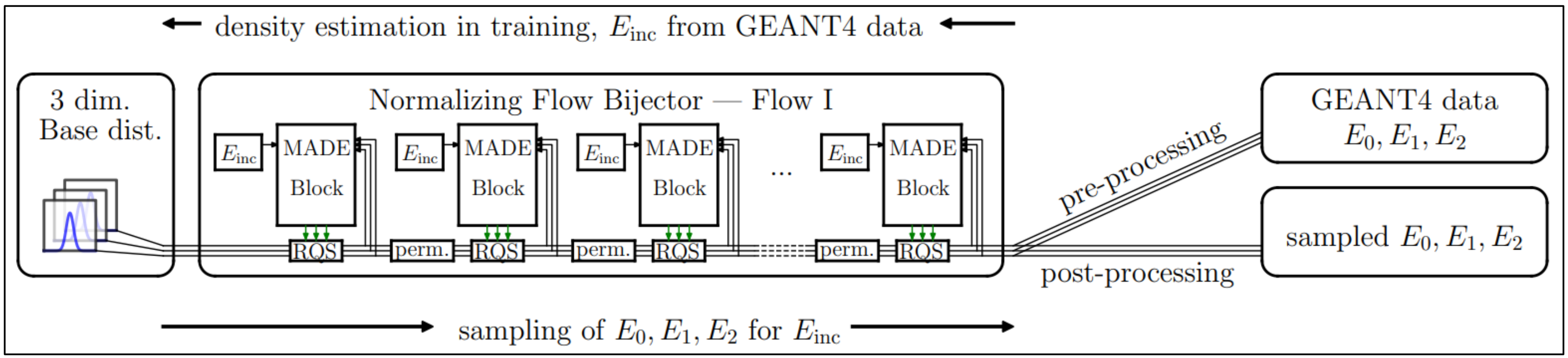
Thanks for your attention!

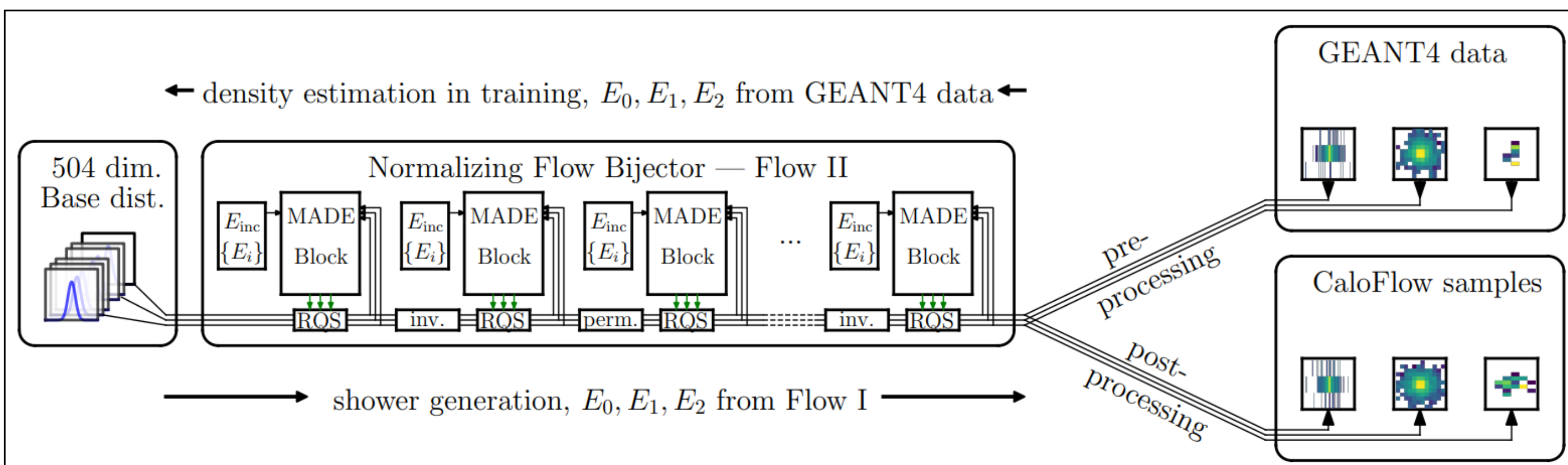
Backup

CaloFlow

- The CaloFlow (2021) uses the same dataset as CaloGAN and shows much better physics performance

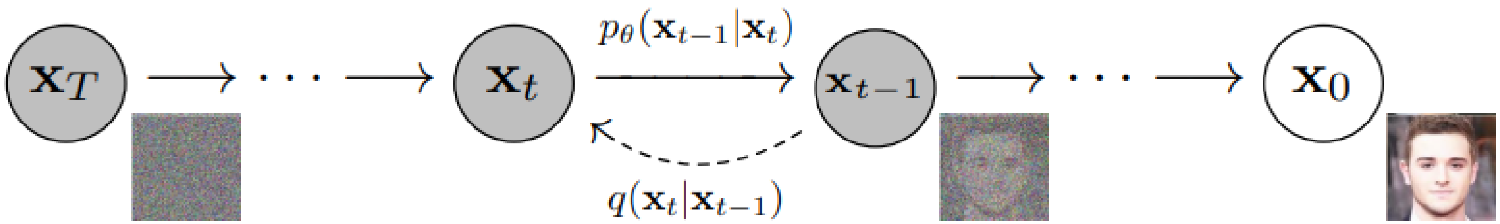






Diffusion model

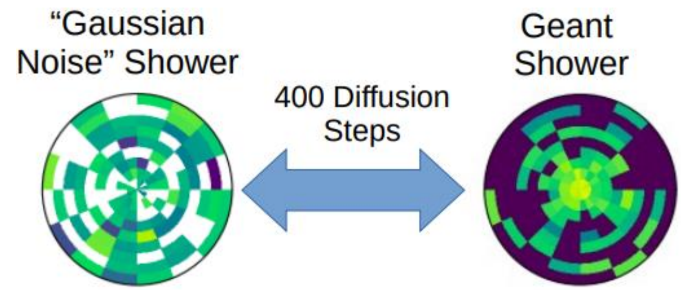
- The diffusion model is proposed in 2020

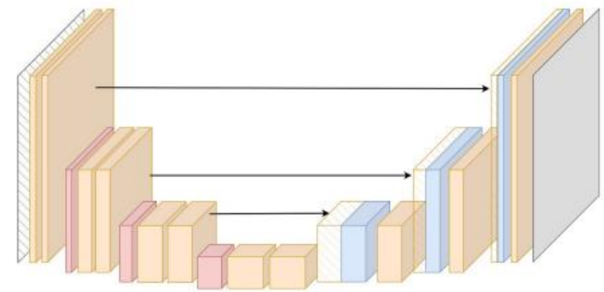


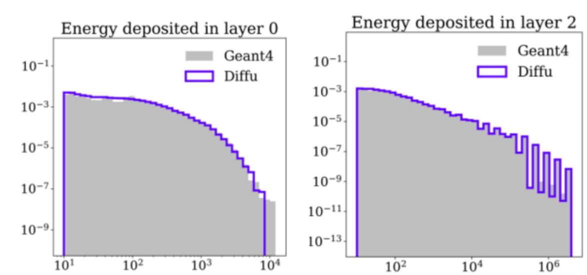
- Diffusion process: $x_0 \rightarrow x_T$
  - Adding noise step by step, making $x_T \sim \mathcal{N}(0, \mathbf{I})$
- Train a model to invert the diffusion process
- When do simulation, start from $\mathcal{N}(0, \mathbf{I})$ and denoise it step by step using the trained model

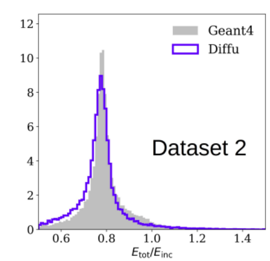
CaloDiffusion

- CaloDiffusion (a fast calorimeter simulation method based on diffusion model)
- Dataset:
  - ATLAS-like geometry, 5 layer cylinder with irregular binning, 368 voxels
- Denoise model:
  - U-net architecture with 3D convolutions
  - Input: Noisy shower
  - Condition inputs: incident particle energy, diffusion step
  - Output: noise
- Good agreement with Geant4, some properties (e.g. total shower energy), can still be improved
- Generation time is slower than other ML approaches (still faster than Geant4)



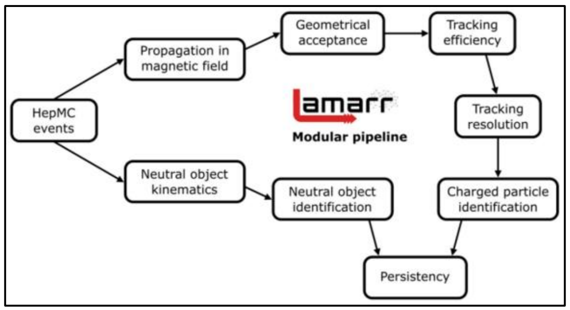


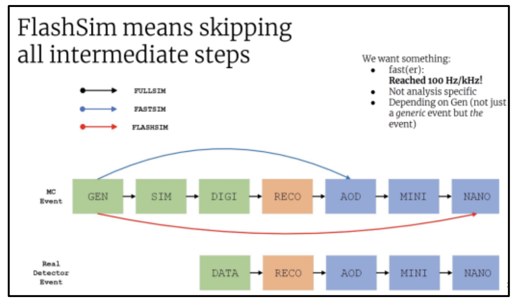


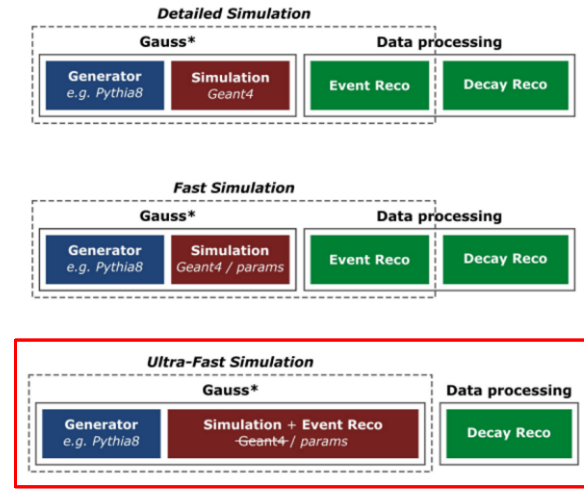


Ultra-Fast Simulation

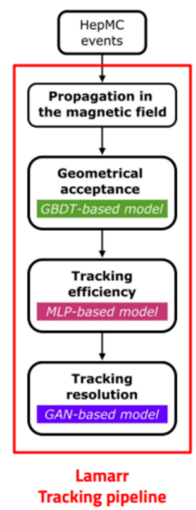
- Without Geant4 simulation, from MC particle to physics analysis object simulation
  - Such as LHCb [Lamarr](), CMS [FlashSim]()

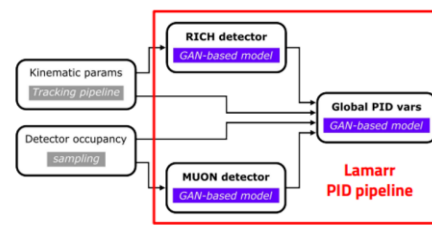


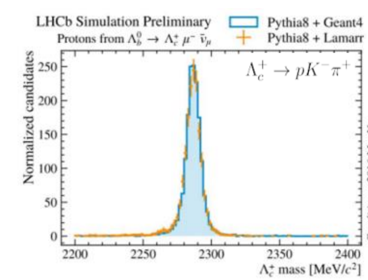




- Most parts are ML-based:
  - GBDT for acceptance
  - MLP for tracking efficiency
  - GAN for tracking resolution and PID







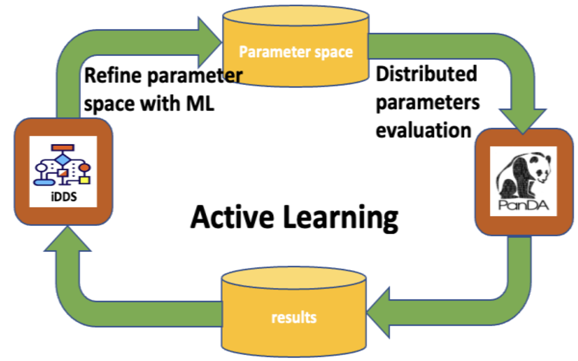
- Simulating ECAL with an ultra-fast approach requires to face the particle-to-particle correlation problem:
  - Sequence of N generated photons → sequence of N′ reconstructed clusters
  - Approached as a language translation problem

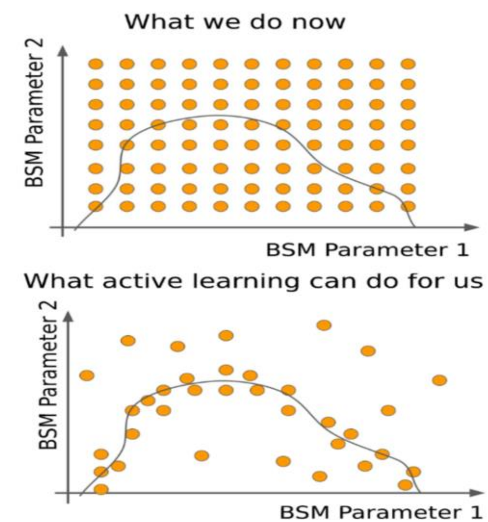
Detector simulation

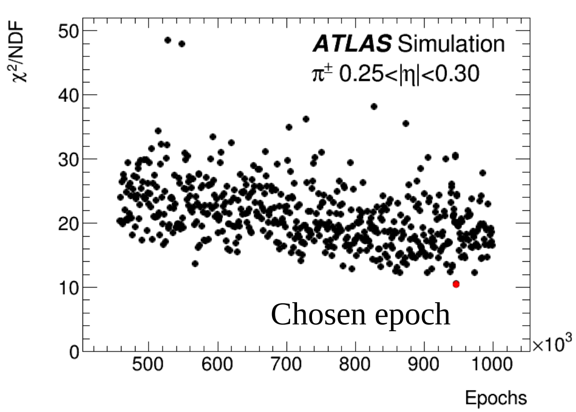
- Calorimeter Fast simulation:
  - FastCaloGAN: a fast simulation of the ATLAS Calorimeter with GANs
  - gaede_chep23_caloml_v01 (jlab.org)
  - Generating Accurate Showers in Highly Granular Calorimeters Using Normalizing Flows
  - Fast and Accurate Calorimeter Simulation with Diffusion Models
  - Transformers for Generalized Fast Shower Simulation
- Ultra-fast simulation
  - THE LHCB ULTRA-FAST SIMULATION OPTION, LAMARR
  - Flashsim: an ML simulation framework
- Refining fast simulation using machine learning
- Hadronic Simulation with conditional Masked Autoregressive Flow

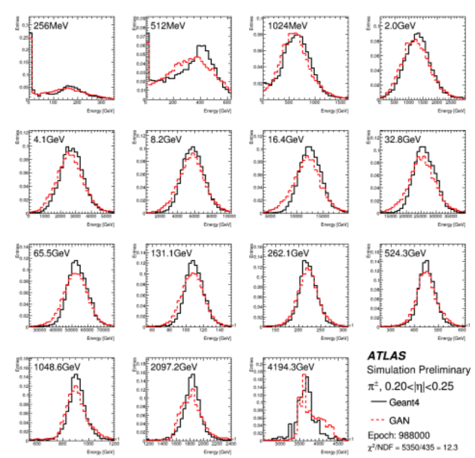
量能器快速模拟（1）

- 由于要训练的 GAN 模型个数多（500 个），且每个 GAN 模型的训练也不容易（训练过程不稳定、需要优化超参数）。因此，需要解决 GAN 模型训练的问题
- 为此 ATLAS 实验利用 Active Learning 的技术，实现网络模型的自动训练和超参数的优化：
  - iDDS（intelligent Data Delivery Service）负责根据当前模型训练的结果产生下一批模型训练作业（如利用贝叶斯算法、GP 等算法缩小超参数范围）
  - PanDA 系统将作业调度到分布式的异构计算资源上进行模型的训练，返回训练结果



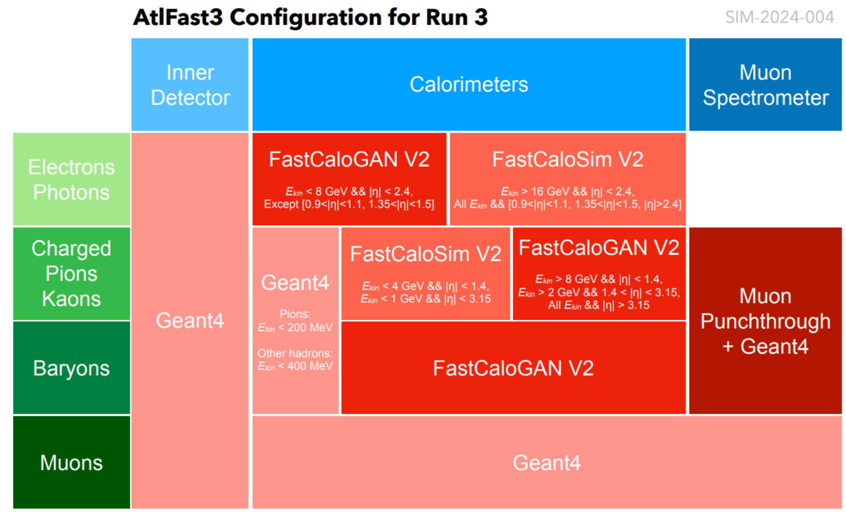


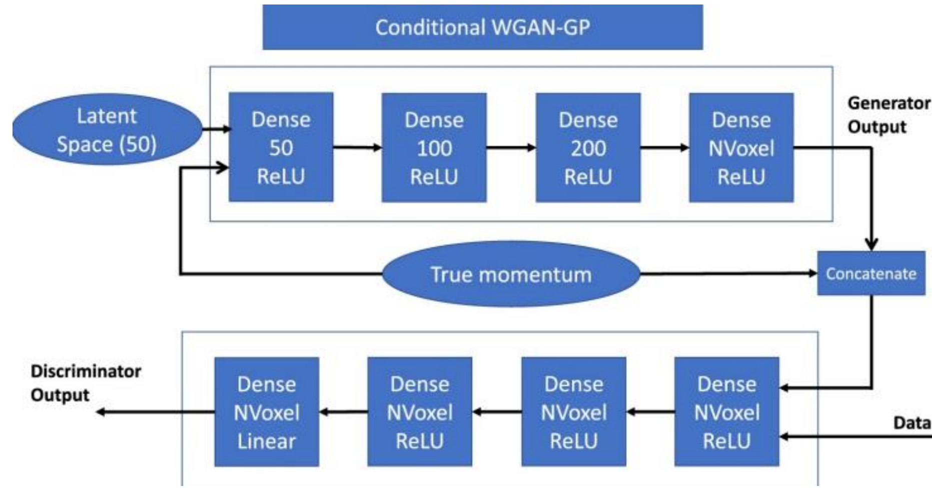




The ATLAS case

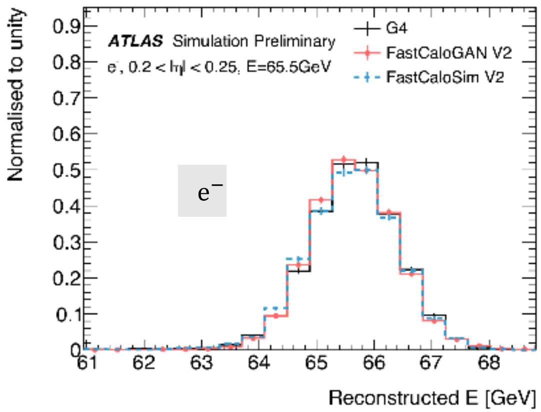
- AltFast3 (a detector response fast simulation system):
  - FastCaloGAN V2 (ML-based)
  - FastCaloSim V2 (parametrization-based)
  - Geant4 (limited to specific cases)
- FastCaloGAN:
  - WGANs trained for each particle type, for each | η | slice.
  - Conditioned on truth momentum
  - Total 300 GANs

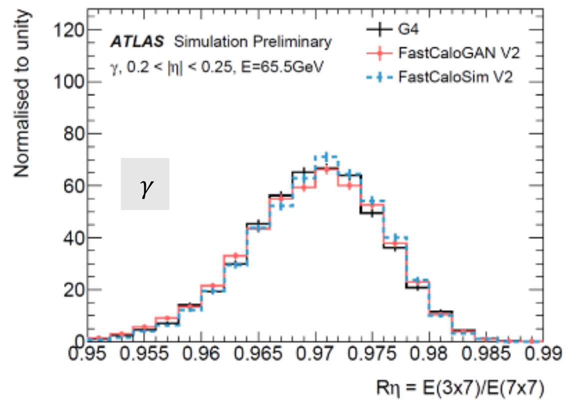


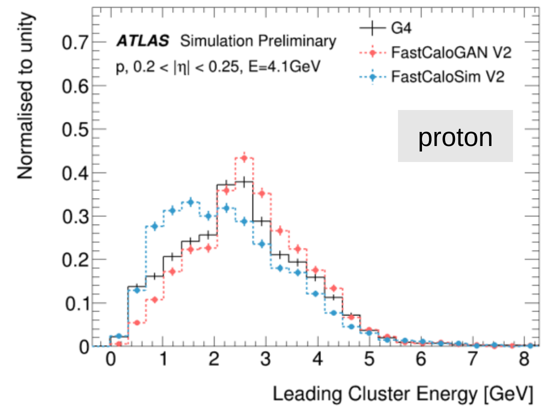


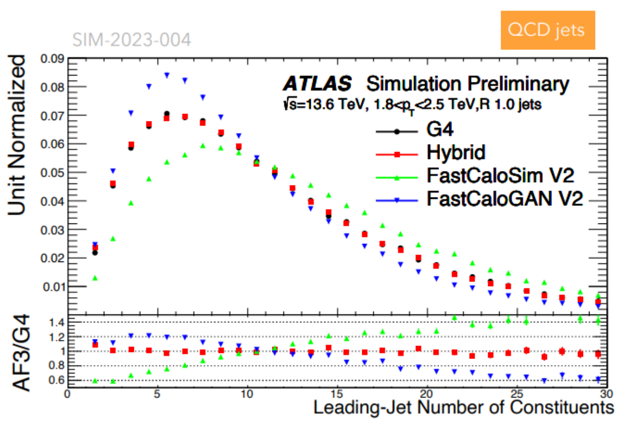
[ACAT2024]

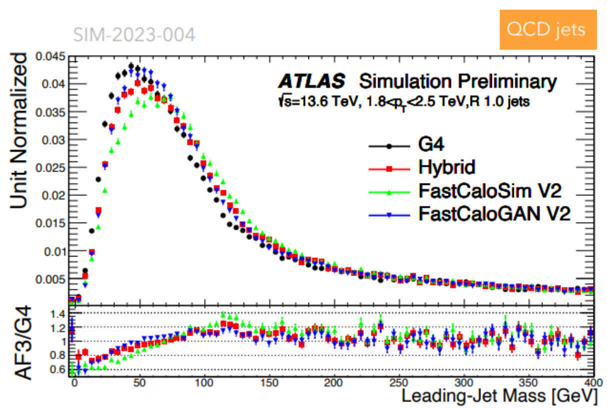
The ATLAS case (performance)

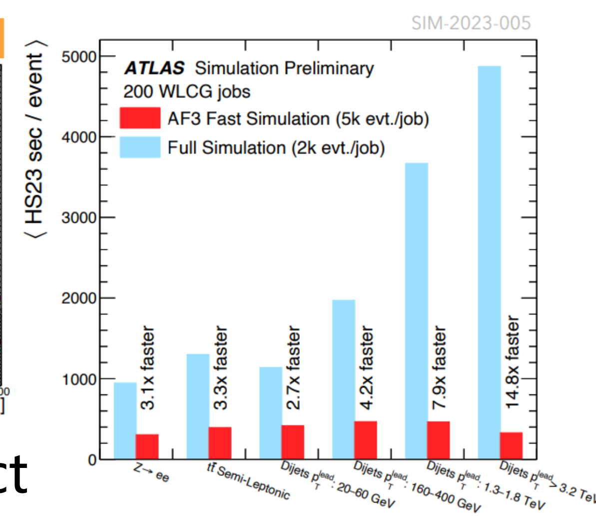








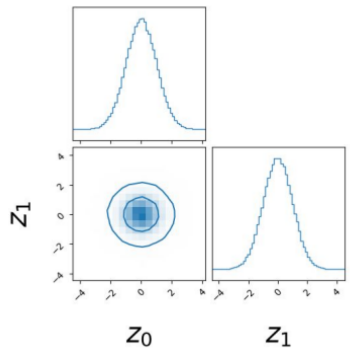




- 3 – 15 speed-up in simulation time with respect to Geant4, depending on the physics process
- Simulation time in AtlFast3 completely dominated by full simulation of the Inner Detector

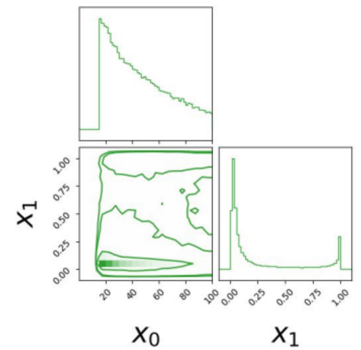
[ACAT2024]

Normalizing Flows



reversible

$\xleftarrow{\;f^{-1}(\mathbf{x})\;}$

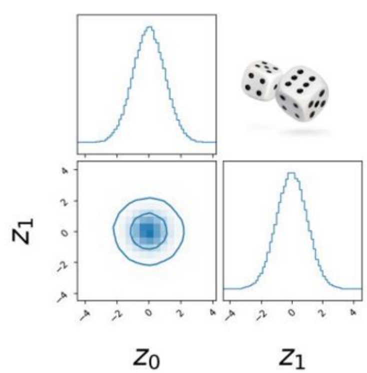


$$\mathbf{x} = f(\mathbf{z})$$

$$p_x(\mathbf{x}) = p_z(\mathbf{z}) \det \left| \frac{d\mathbf{z}}{d\mathbf{x}} \right|$$

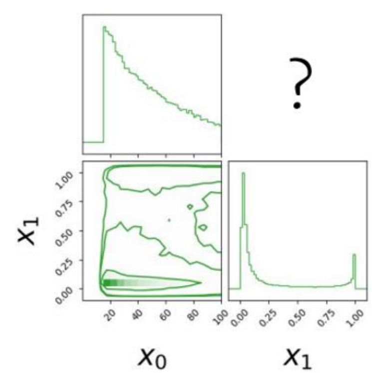
$$\log(p_x(x)) = \boxed{\log(p_z(f^{-1}(\mathbf{x})))} + \boxed{\log\left(\det \mathbb{J}_{f^{-1}}(\mathbf{x})\right)}$$

What we know



$\xrightarrow{\;f(\mathbf{z})\;}$

What we need



$\mathbf{Z}$

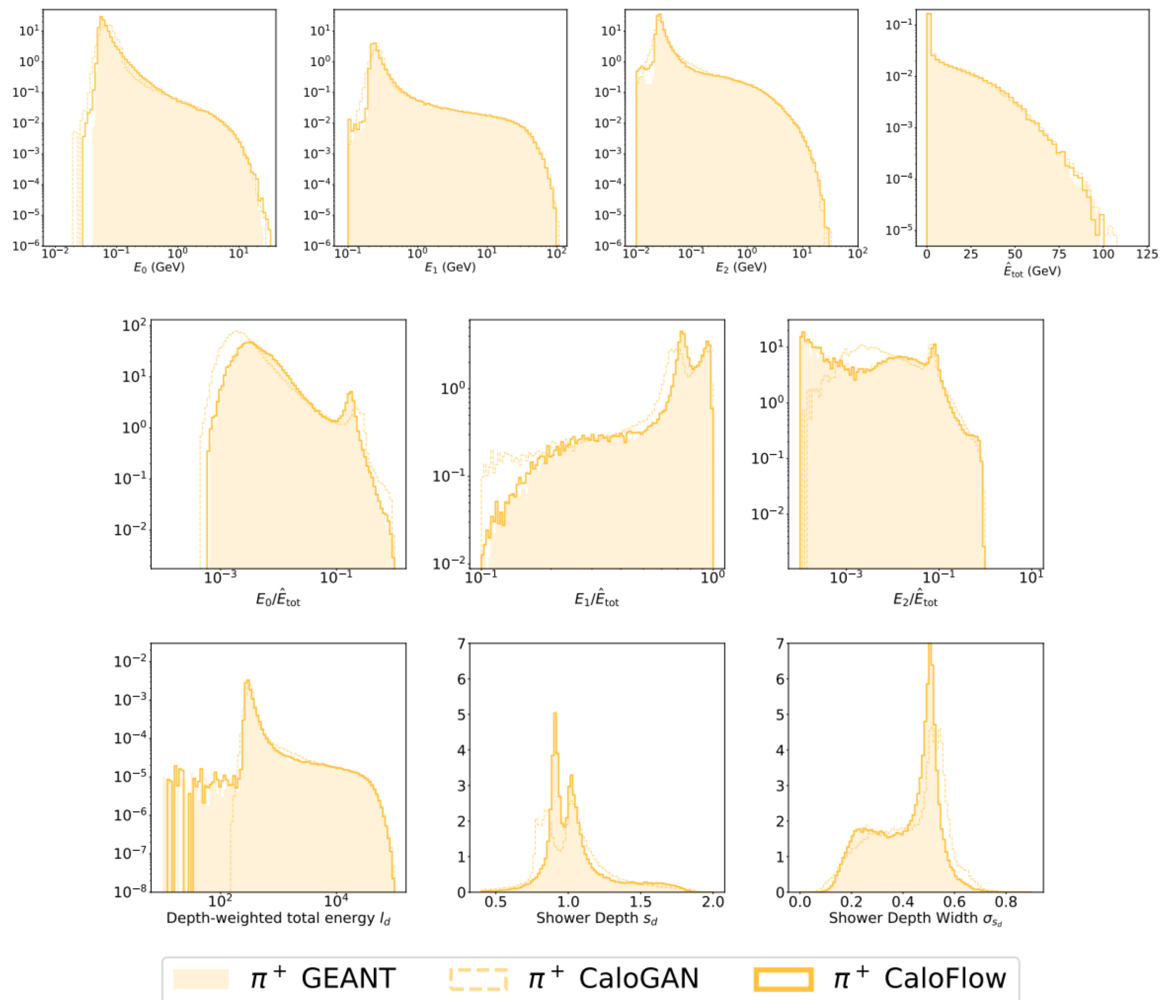
multi-dimensional gaussian

$\mathbf{X}$

FullSim data, pdf unknown!

CaloFlow

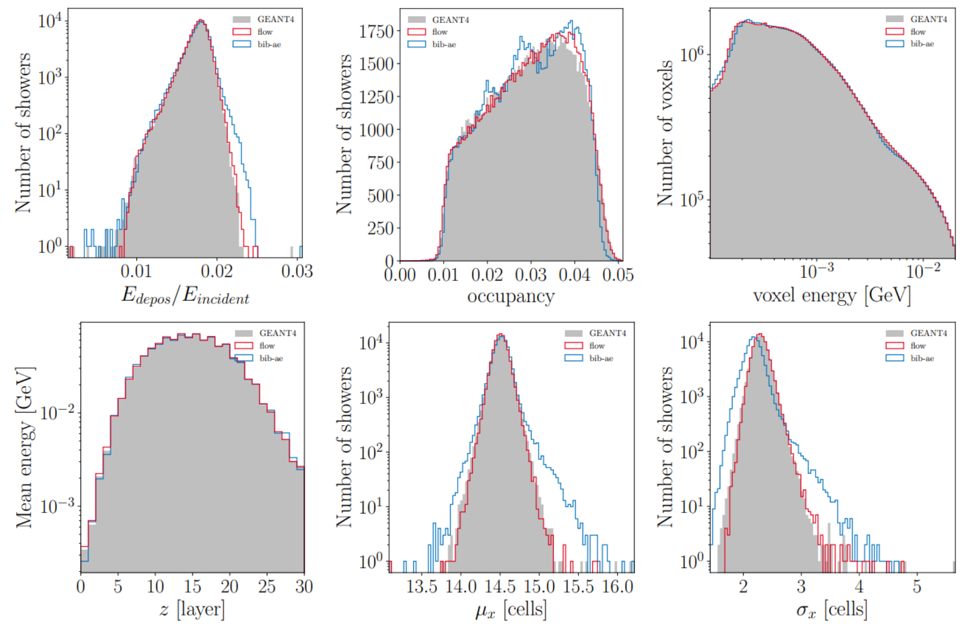
- The CaloFlow uses the same dataset as CaloGAN and shows better physics performance



The ILC case

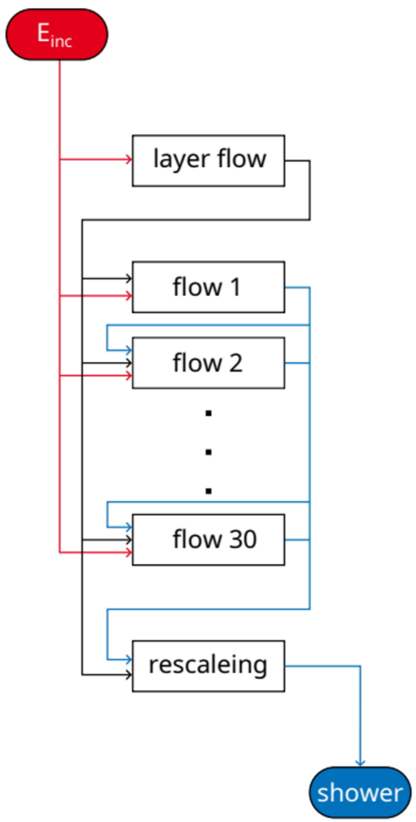
- Dataset:
  - photon showers in ECAL
  - 30x30x30 voxels

Preliminary Results



Simulator	Hardware	Batch size	time [ms]	Speedup
GEANT4	CPU	1	4081.53 ± 169.92	×1.0
BIB-AE	CPU	1	102.25 ± 0.64	×40.0
		10	37.81 ± 0.13	×108.0
		100	48.51 ± 0.01	×84.1
		1000	48.19 ± 0.01	×84.7
Flow	CPU	1	1746.61 ± 64.50	x2.3
		10	392.61 ± 0.34	x10.4
		100	228.86 ± 7.09	x17.8
		1000	275.55 ± 3.01	x14.8
BIB-AE	GPU	1	74.22 ± 3.18	×42.5
		1000	0.249 ± 0.002	×16326.1
Flow	GPU	1	2471.07 ± 70.20	x1.7
		1000	3.39 ± 0.09	×1202.3

Architecture



Link