GAN for CEPC study

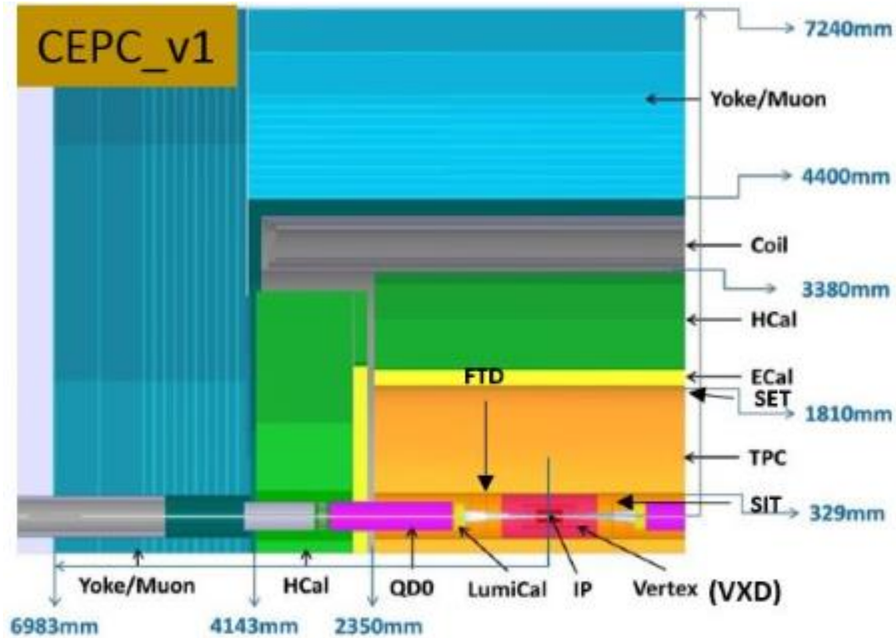
中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

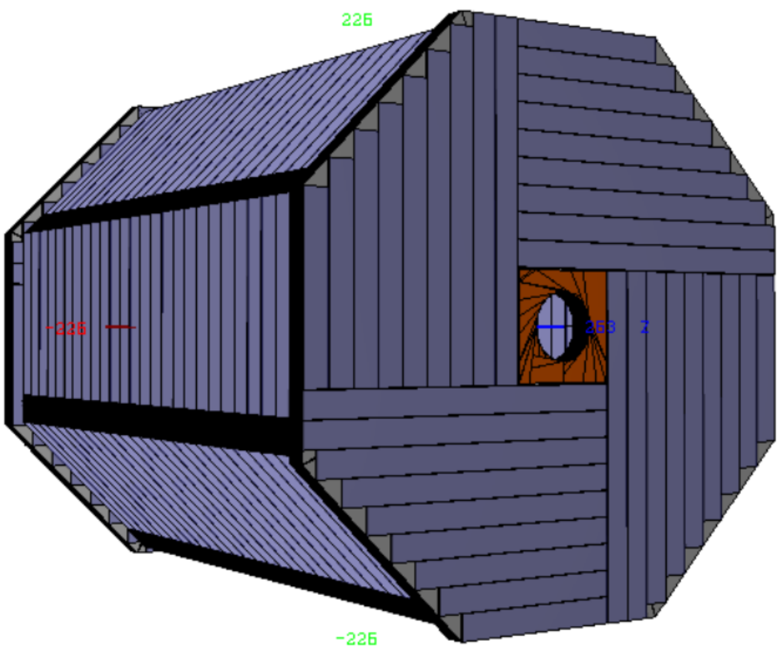
Wenxing Fang(IHEP)

Co-supervisor: Prof. Weidong Li

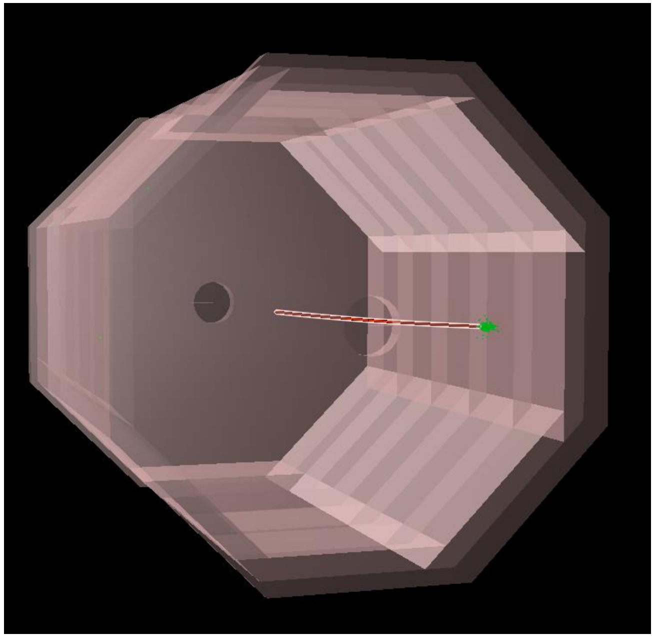
Group meeting (14th October 2019)

- The Circular Electron Positron Collider (CEPC) is a large international scientific facility proposed by the Chinese particle physics community.
- The CEPC will be hosted in China in a circular underground tunnel of approximately 100 km in circumference.
- It is designed to operate at around 91.2 GeV as a Z factory, at around 160 GeV of the WW production threshold, and at 240 GeV as a Higgs factory.

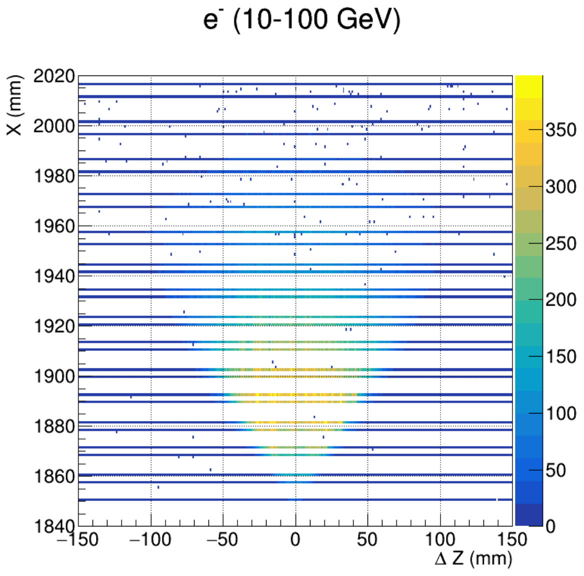


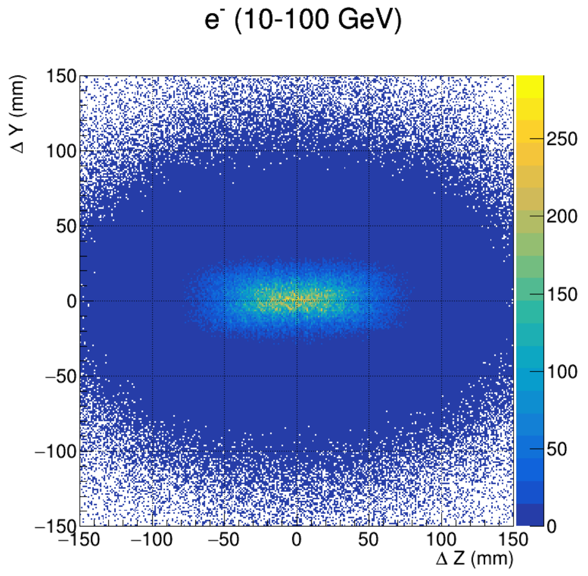


Using particle gun to hit ECAL barrel (CEPC_v4)



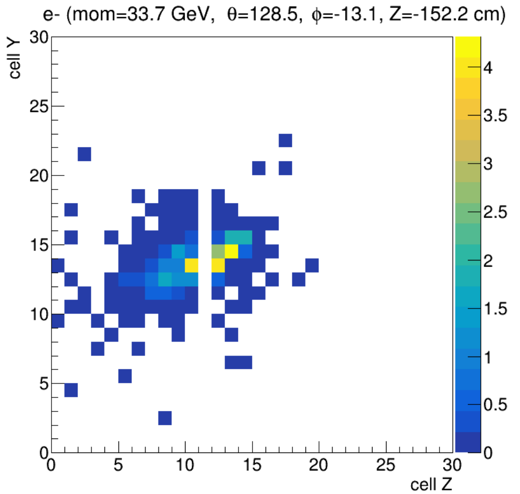
```
/generator/generator particleGun
/gun/position 0 0 0 mm
/gun/direction 1.0 0.0 0.0
/gun/momentum 55 GeV
/gun/momentumSmearing 45 GeV
/gun/phiSmearing 20 deg
/gun/thetaSmearing 50 deg
/gun/directionSmearingMode uniform
/gun/momentumSmearingMode uniform
/gun/particle e-
/run/beamOn 100000
```

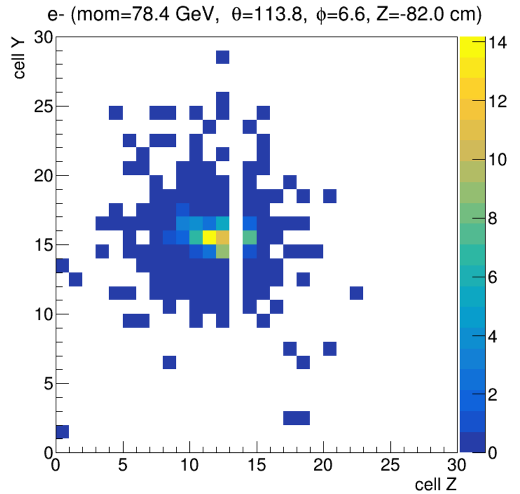


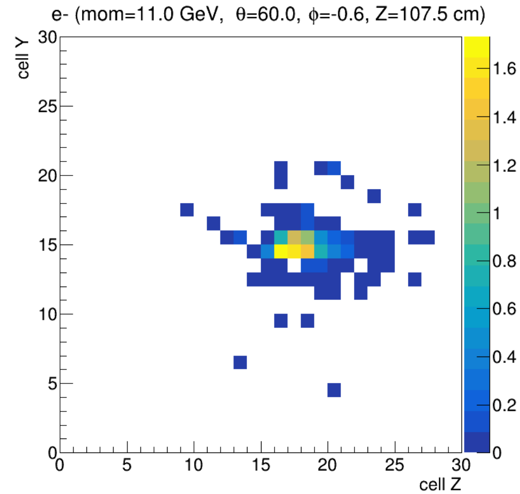


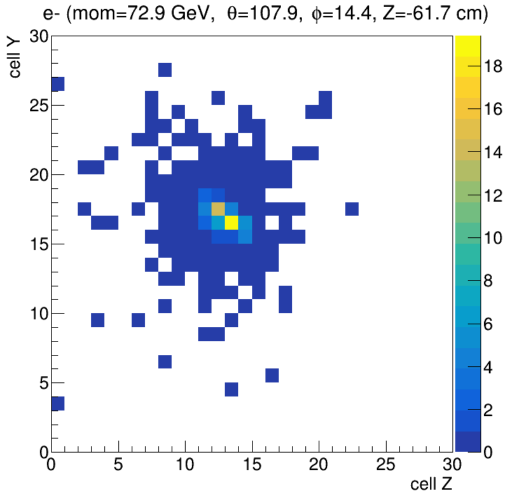
- Using ECAL only.
- Use magnetic field.
- The digitalization is applied.
- The hit point of incoming particle at first layer (x=1.85m) is chose as the center of Z-Y plane. Besides, |hit_point_y|<0.5 m and |hit_point_z|<2m is required.
- Only consider the hits within radius of 150 mm.

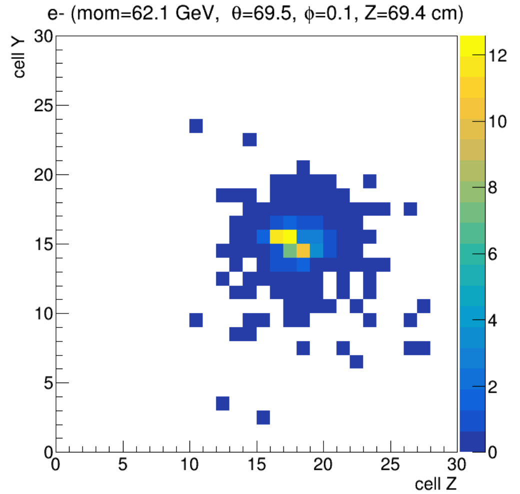
Energy deposition in Z-Y plane

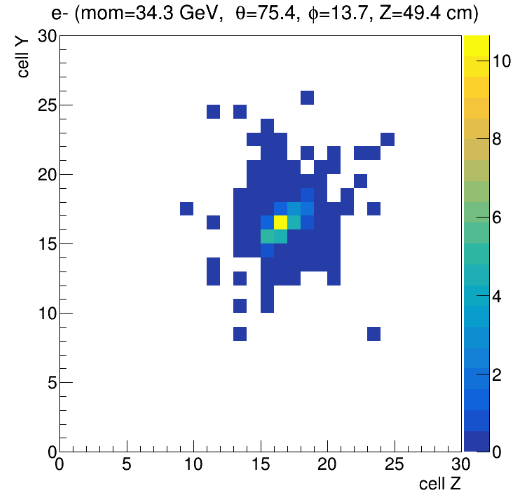




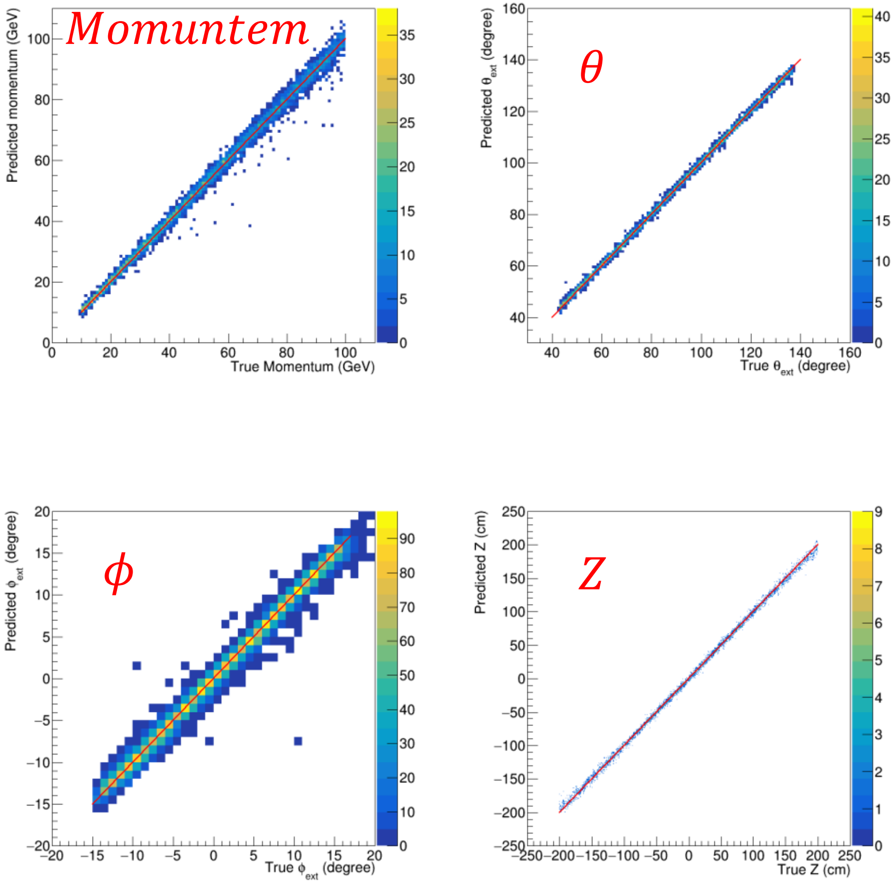


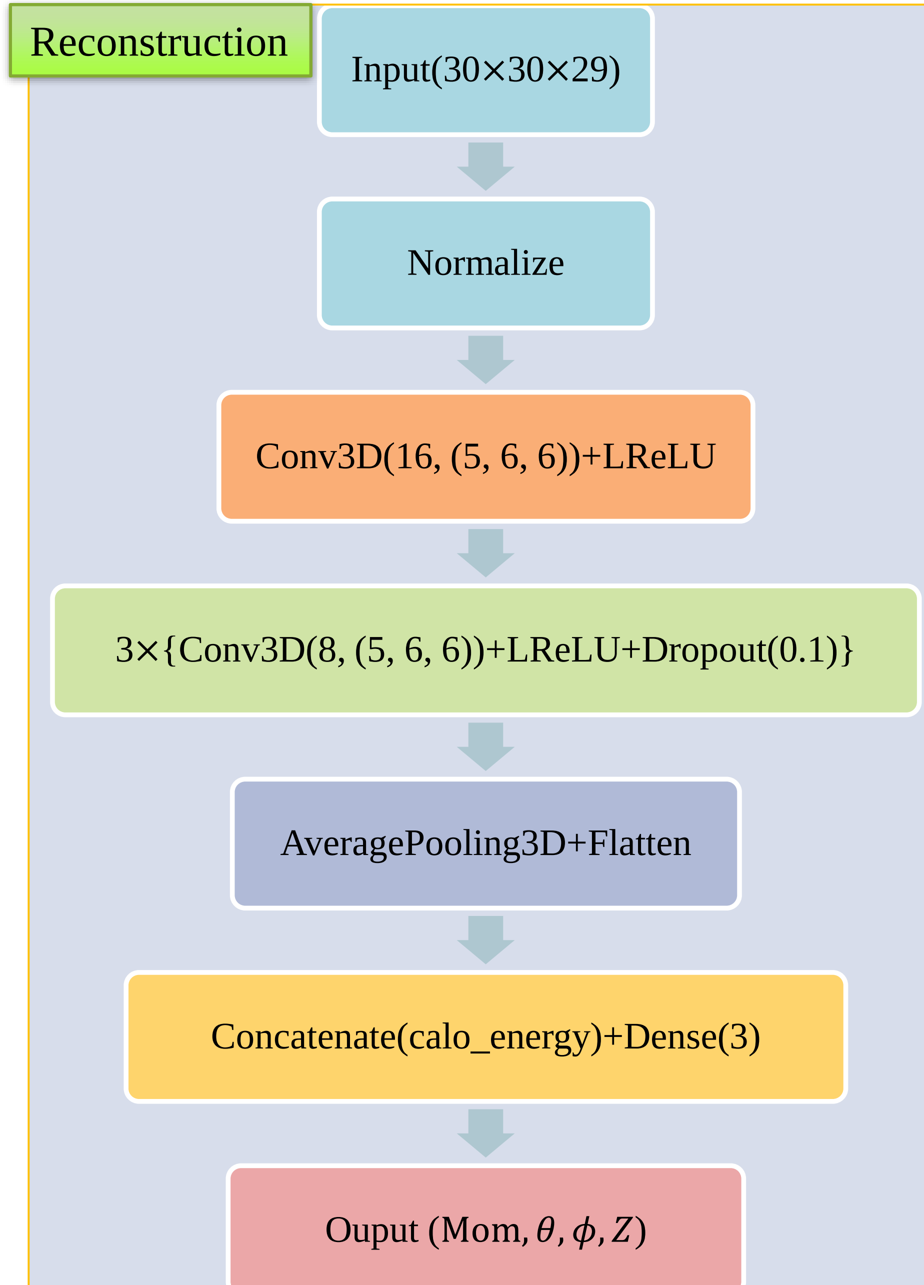




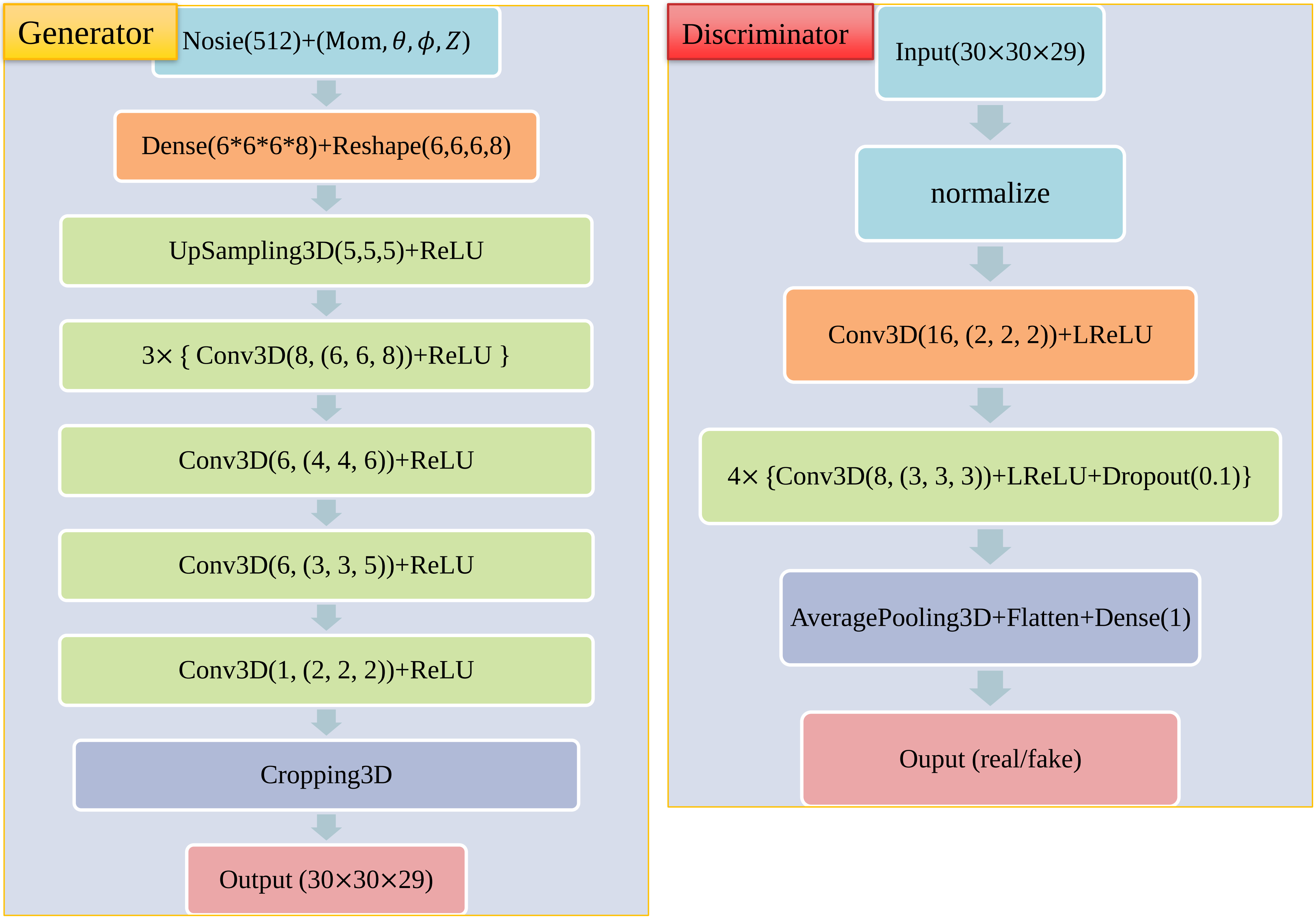


Reconstruction architecture and performance

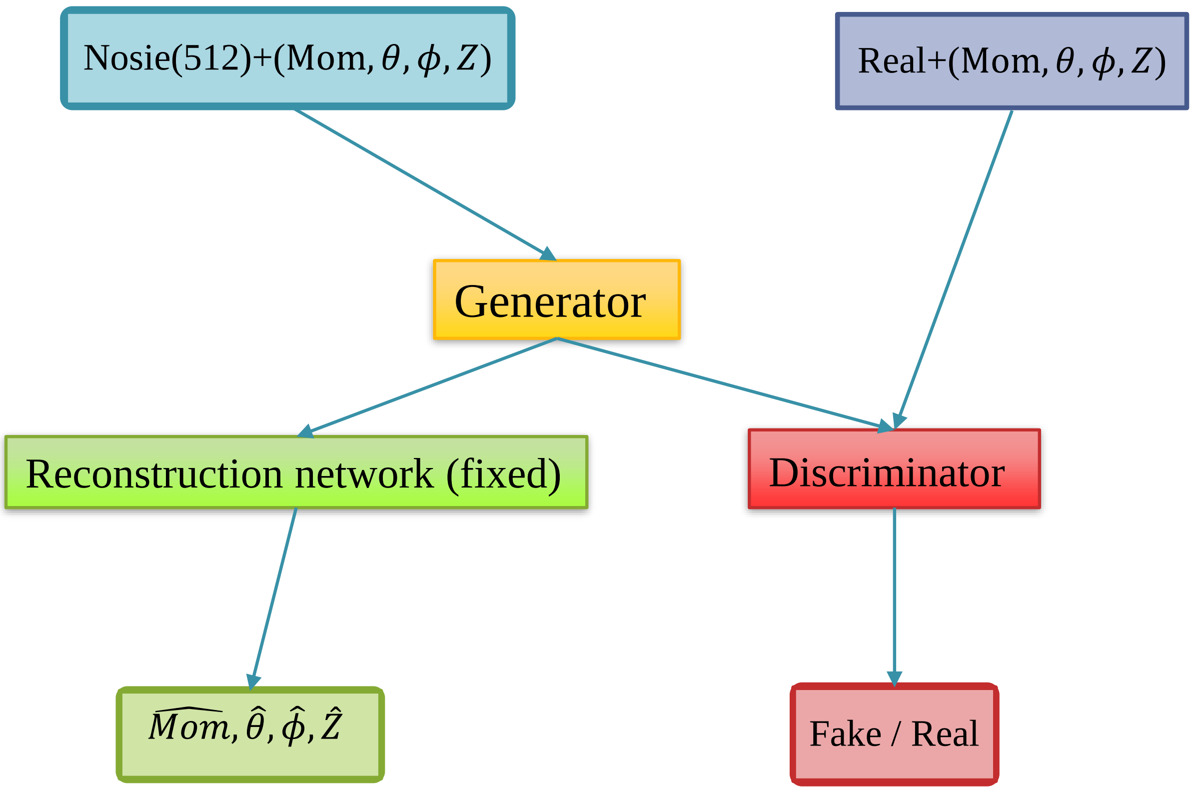




Generator and discriminator architecture



Generative Adversarial Networks (GAN)

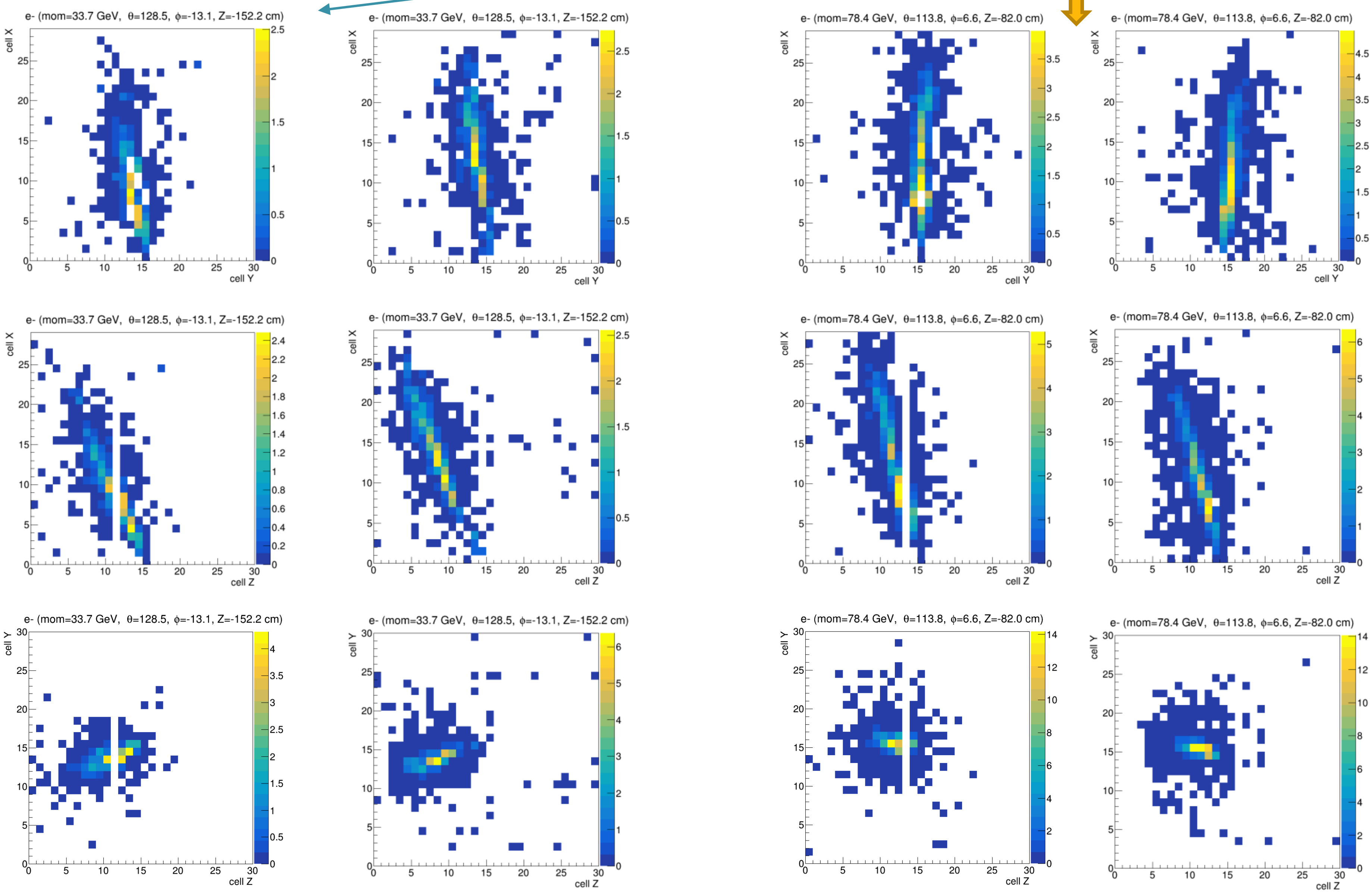


- The discriminator is designed to distinguish the image from real or generated.
- The generator is designed to fool the discriminator as well as obtains closer reconstructed parameter compared with the given one.

Generated events

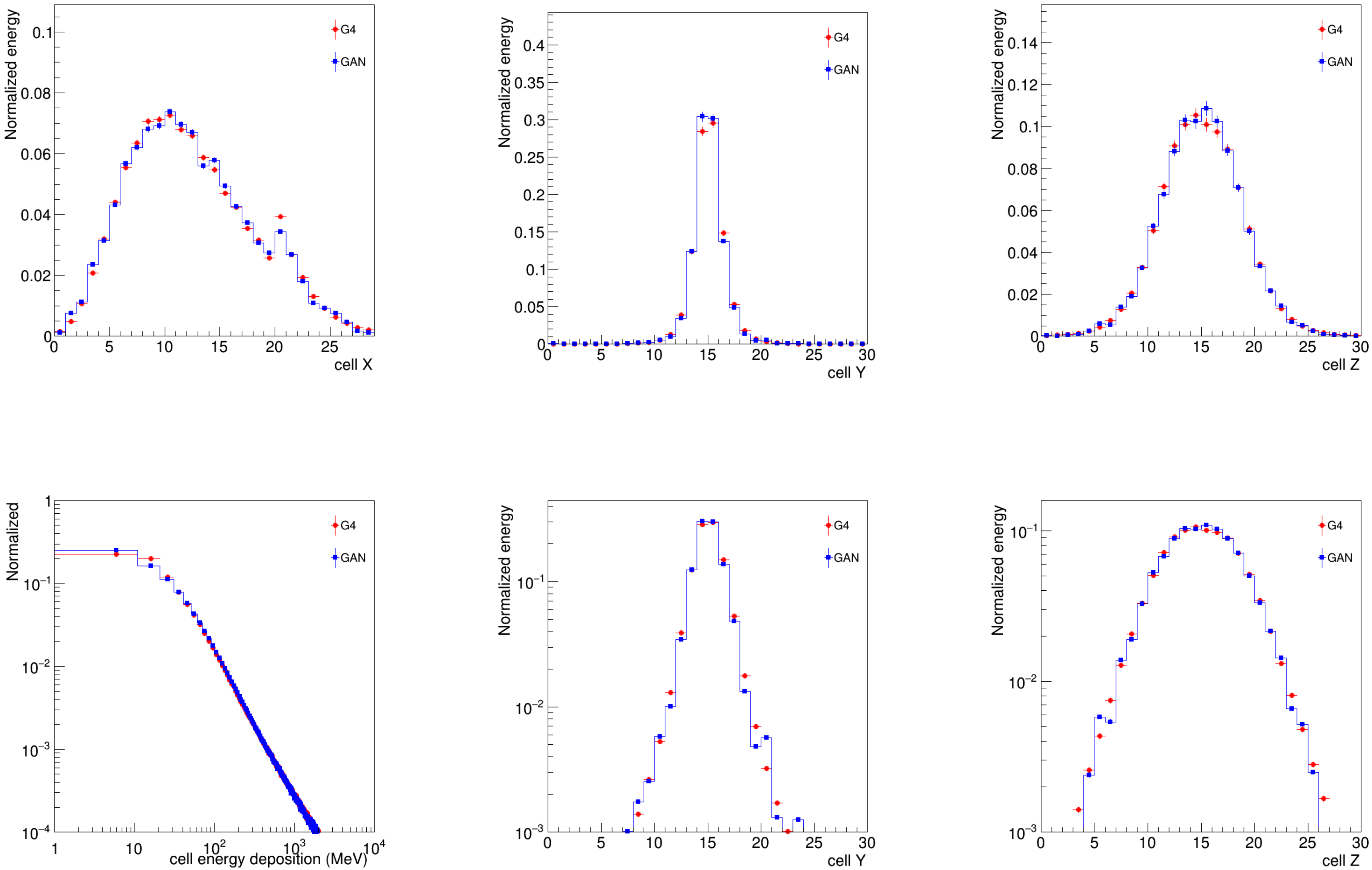
Mom=33.7 GeV, θ=128.5°, ϕ=−13.1°, Z=-152.2 cm

Mom=78.4 GeV, θ=113.8°, ϕ=6.6°, Z=-82 cm



Geant 4 | GAN | Geant 4 | GAN

Showershape variables

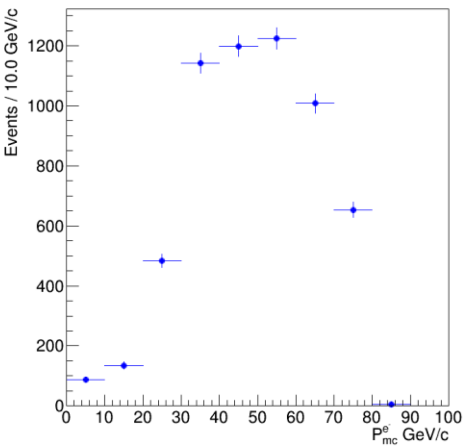


using mc events

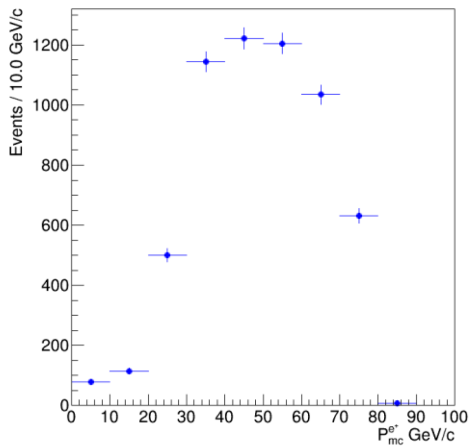
$$e^+e^- \rightarrow Z(e^+e^-)H(\mu^+\mu^-)$$

- Dataset: /cefs/data/FullSim/CEPC240/CEPC_v4/higgs/E240.Pe1e1h_e2e2.e0.p0.whizard195/e1e1h_e2e2.e0.p0.000*_sim.slcio

Momentum e^-

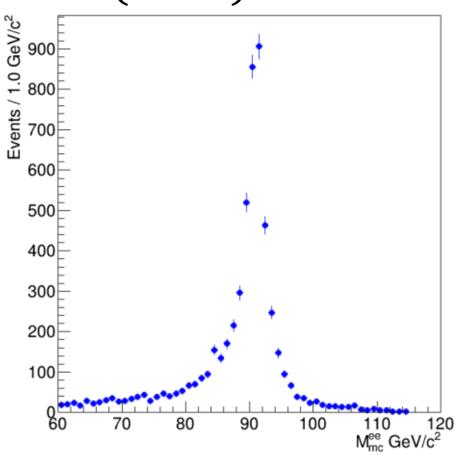


Momentum e^+

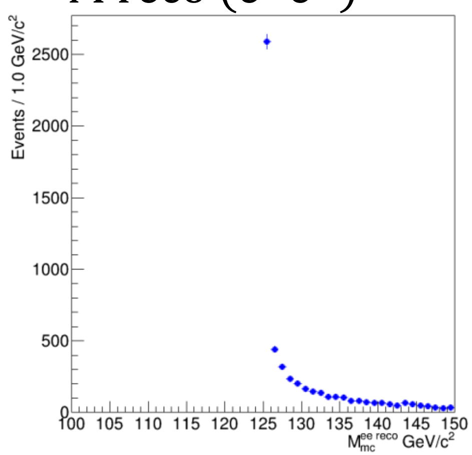


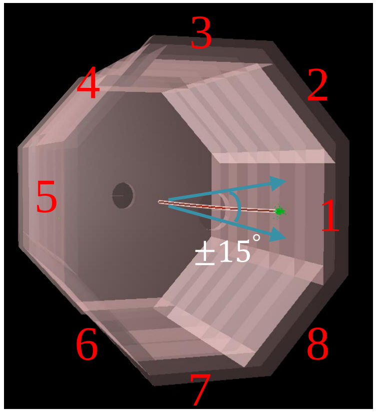
Remove the event which contains $e^{\pm}$ with energy > 10 GeV and its cluster |z| > 2 m or close to connection region of barrel Ecals.

$M(e^+e^-)$



M reco (e^+e^-)

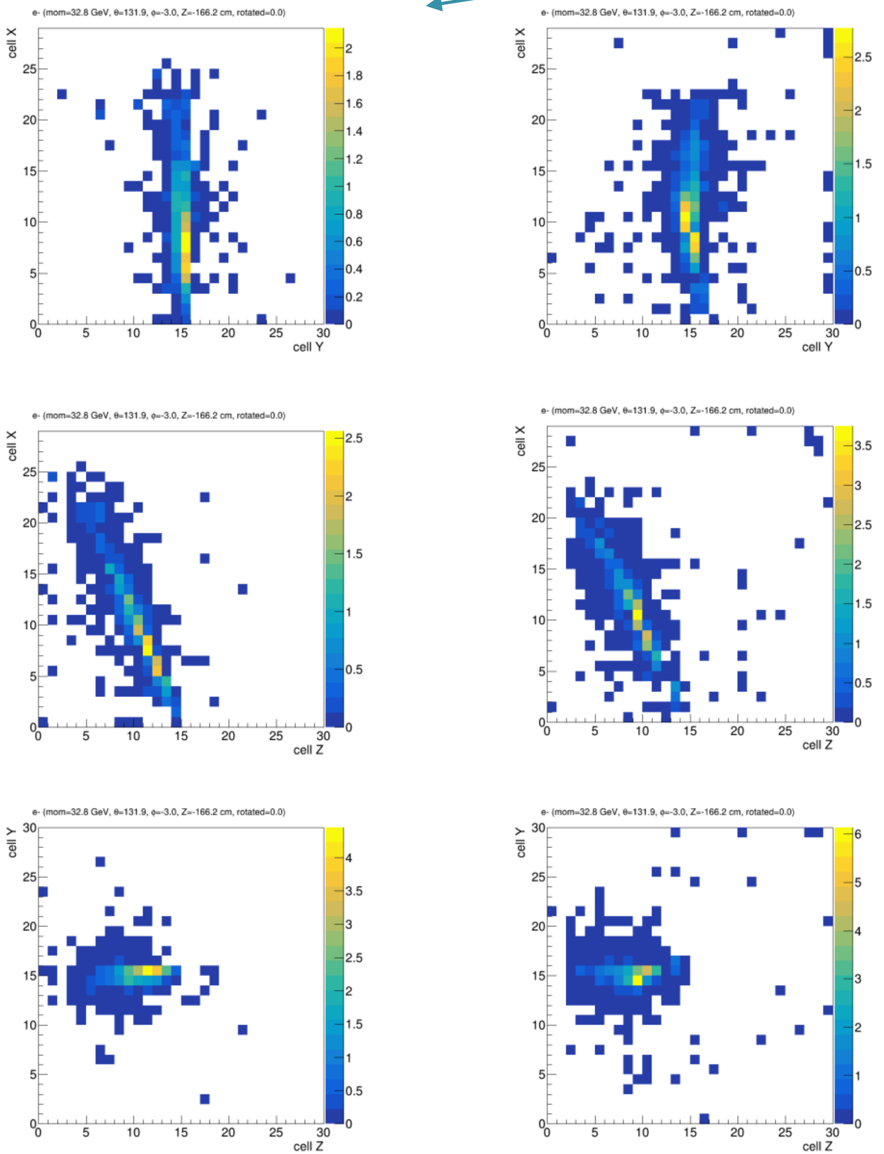


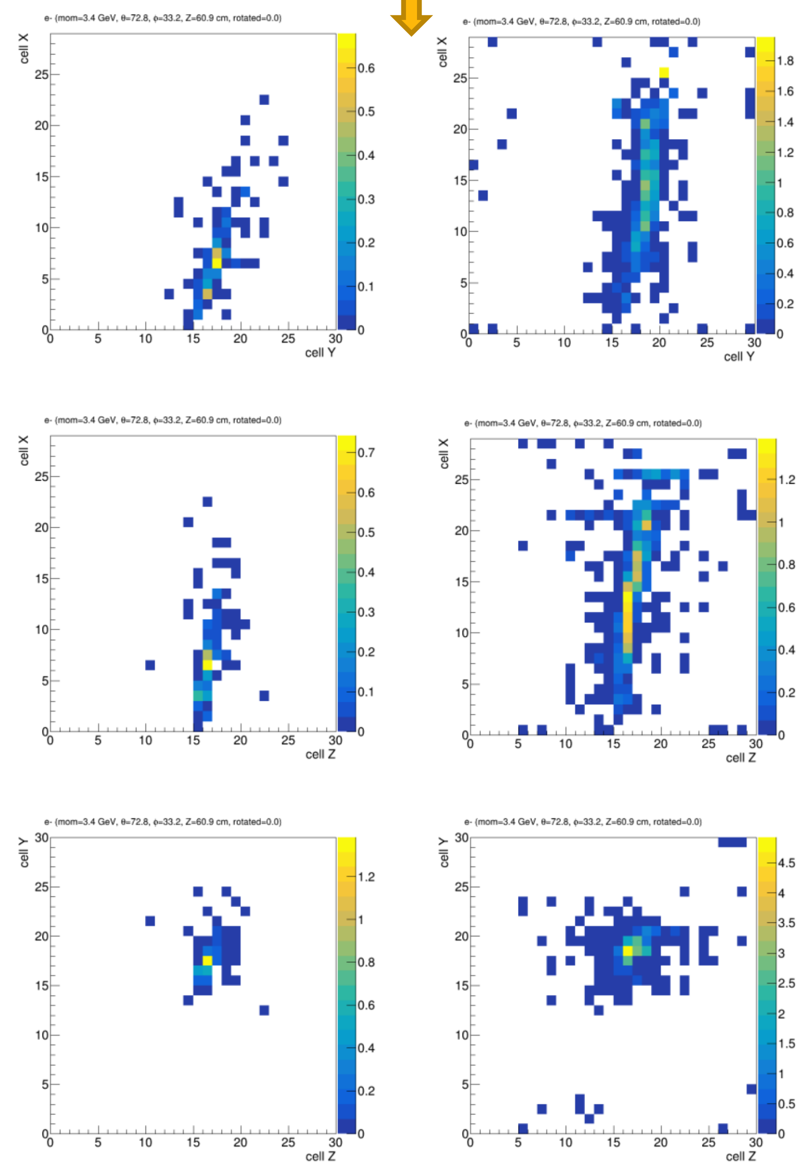


Generated events

Mom=32.8 GeV, θ=131.9°, ϕ=−3.0°, Z=-166.2 cm

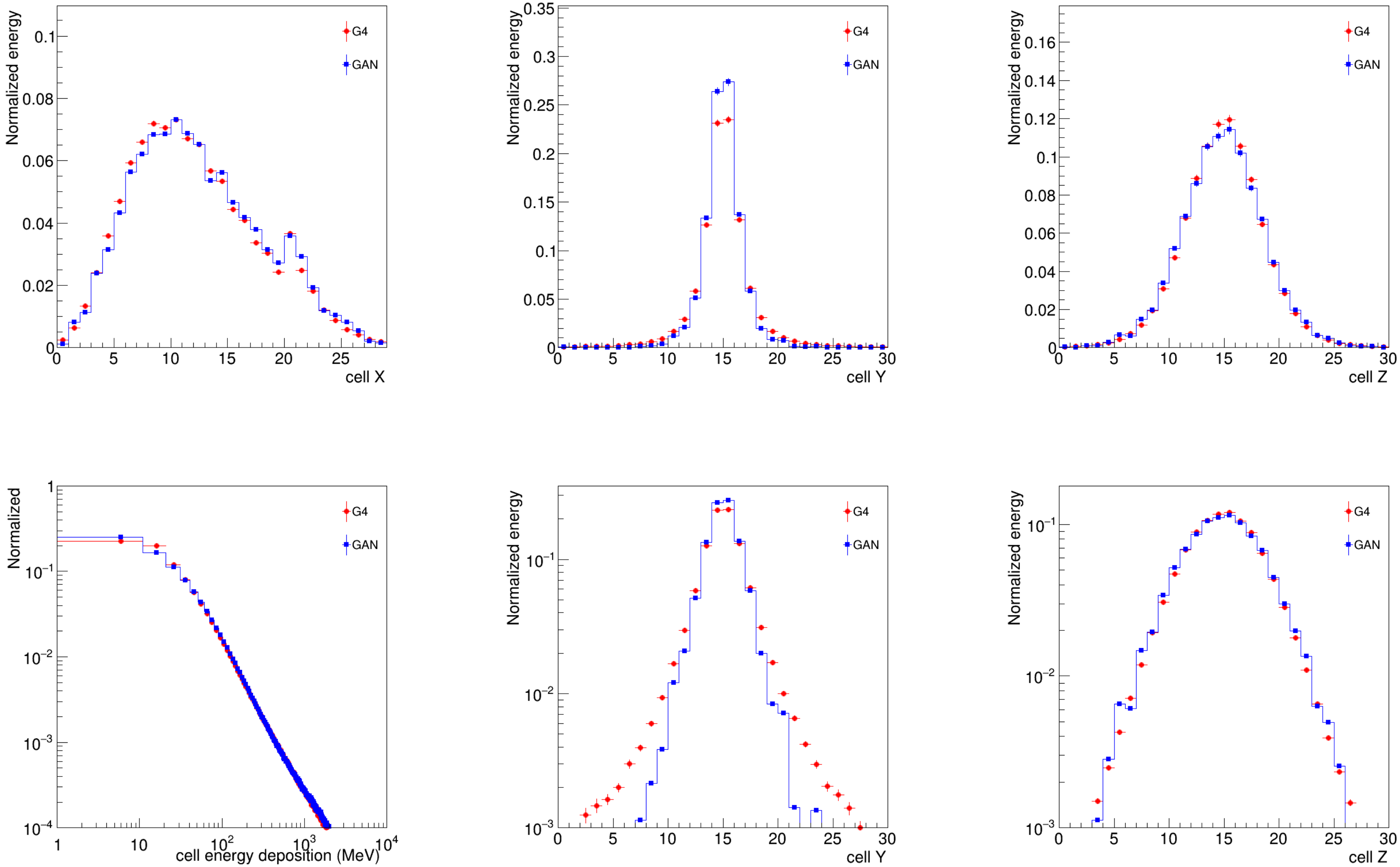
Mom=3.4 GeV, θ=72.8°, ϕ=33.2°, Z=60.9 cm





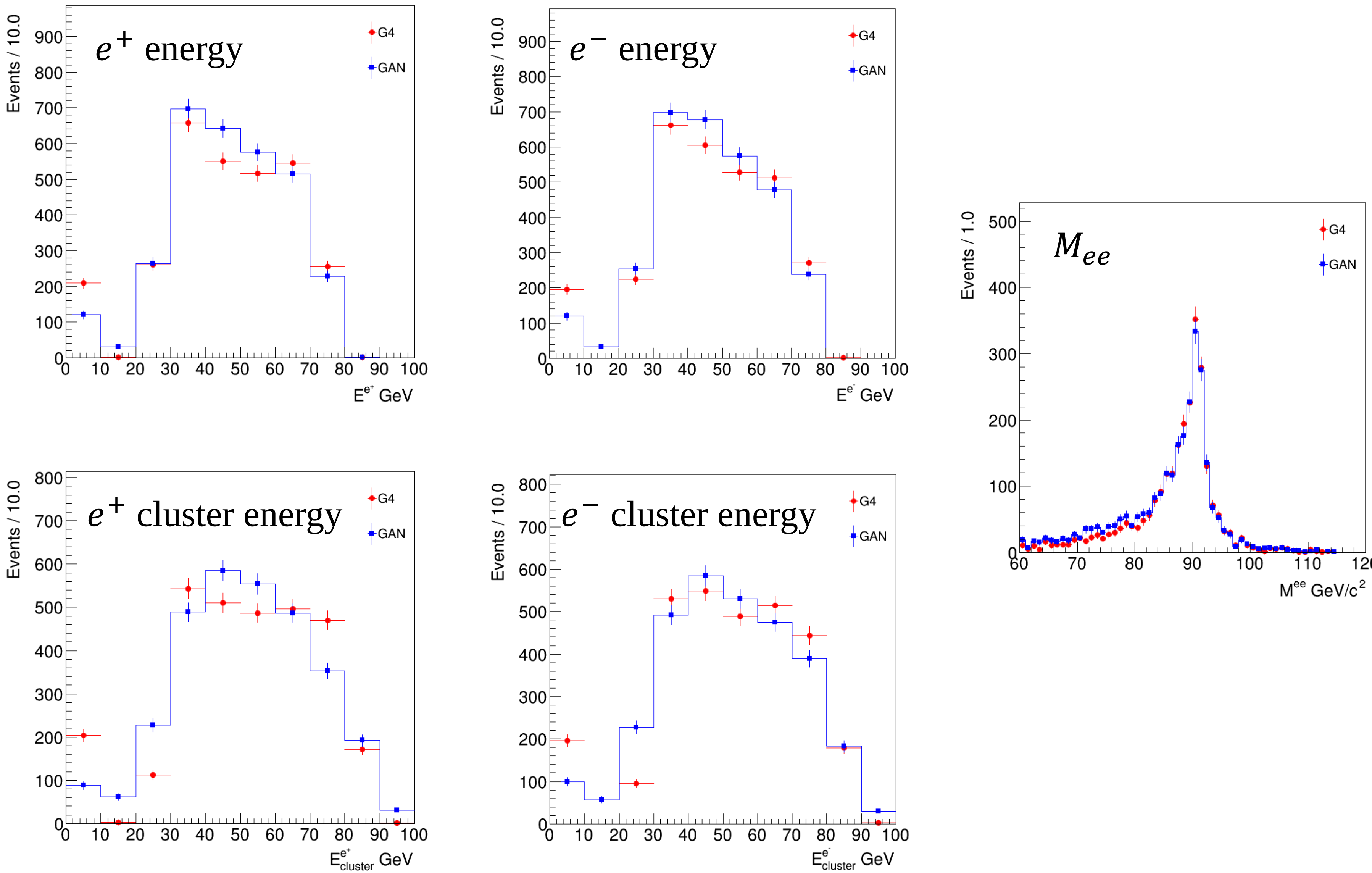
Geant 4 | GAN | Geant 4 | GAN

Showershape variables (using mc sample)



Use GAN for reconstruction

Replace the Hit in “ECALBarrel” collection. Then do reconstruction.

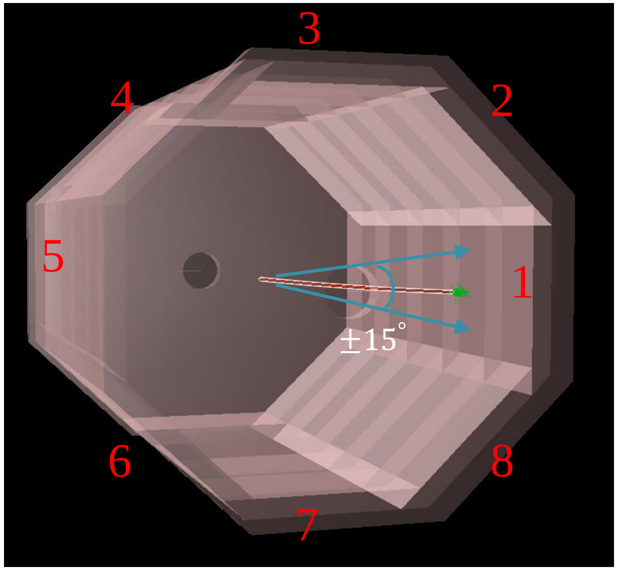


Next to do

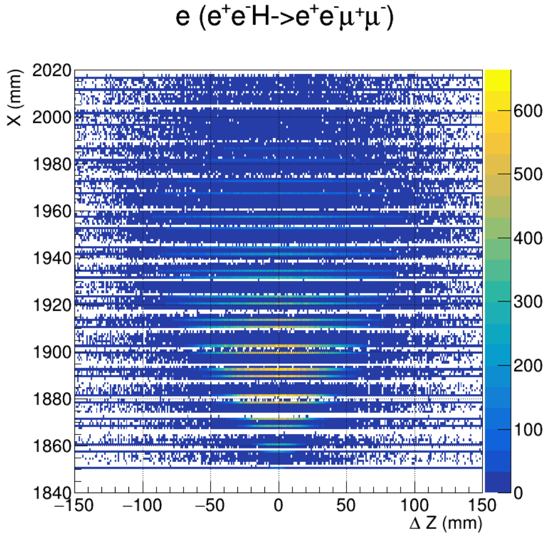
- Checking more variables like the variables used for particle identification.
- Using cell ID for selecting neighbor hits, instead of using binning.
- Extend the training energy down to 1 GeV.
- Training for gamma which is reconstructed only using ECAL information and check its behavior in $Z(\nu\nu)H(\gamma\gamma)$ event.

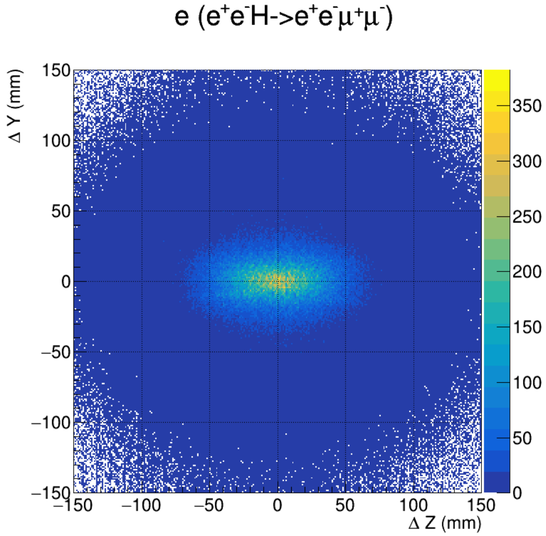
Back up

Using MC sample (CEPC_v4)



- Using full simulated mc samples.
- The digitalization is applied.
- When the particle doesn't hit region 1, then rotating it to region 1.
- The hit point of incoming particle at first layer (x=1.85m) is chose as the center of Z-Y plane. Besides, |hit_point_z|<2.3 m is required.
- Only consider the hits within radius of 150 mm.





Use Gan for reconstruction

/cvmfs/cepc.ihep.ac.cn/software/cepcsoft/x86_64-sl6-gcc49/cepcsoft/0.1.0/Framework/LCIO/02-04-03/src/cpp/src/EXAMPLE/simjob.cc

First hit x,y,z → First hit cell ID → neighbor cells ID → save neighbors cell energy

ID==LayerNum, StaveNum, SubDId …

```
string initString = "M:3,S-1:3,I:9,J:9,K-1:6"
```

/cvmfs/cepc.ihep.ac.cn/software/cepcsoft/x86_64-sl6-gcc49/cepcsoft/0.1.0/Reconstruction/PFA/Arbor/3.4.2/src/PluginMatch/MarlinArbor.cc

```
LayerNum = idDecoder(a_hit)["K-1"] + 1 + KShift;
 StaveNum = idDecoder(a_hit)["S-1"] + 1;
ArborHit a_abhit(currHitPos, LayerNum, 0, Depth, StaveNum, SubDId);
```

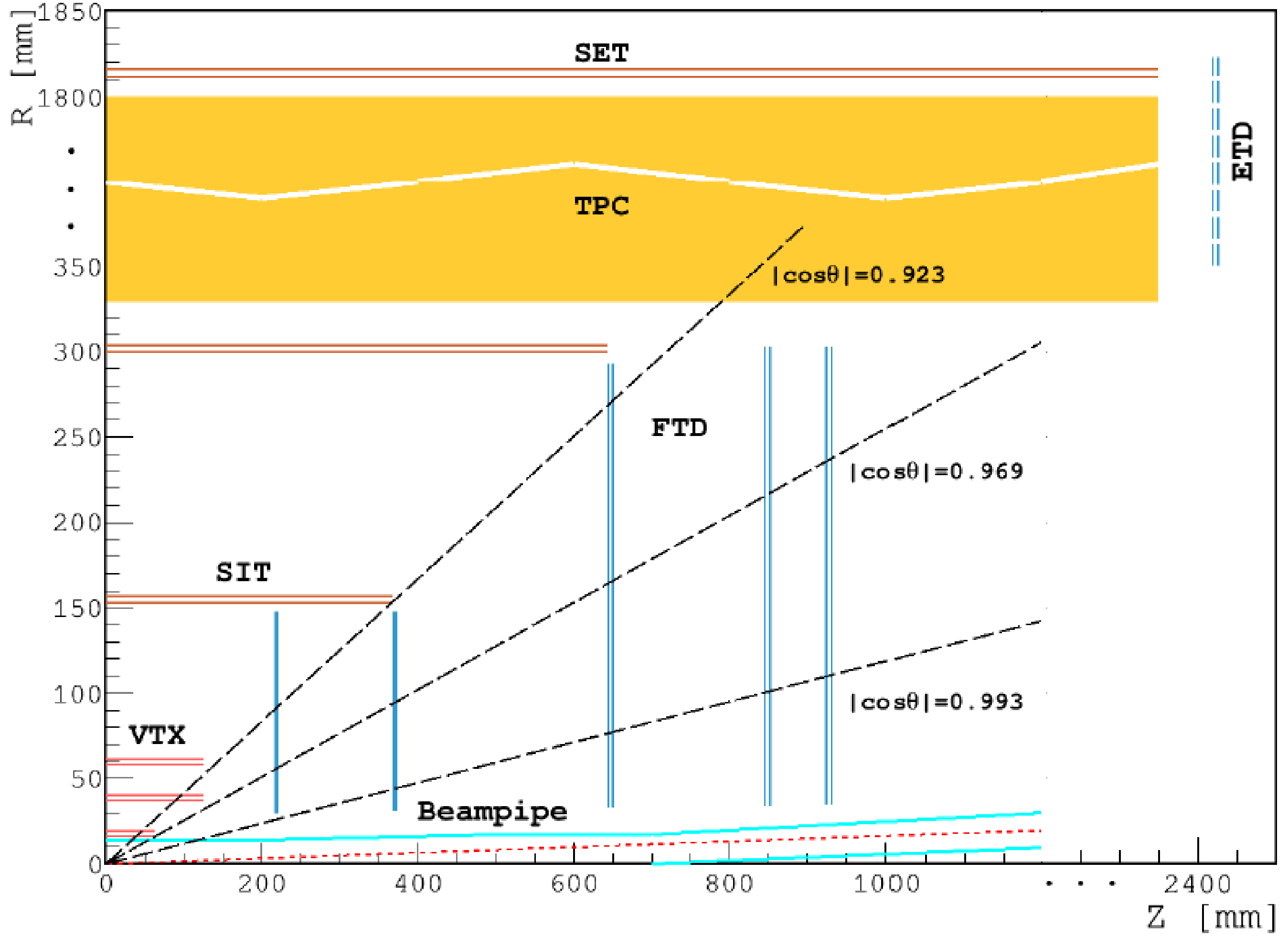


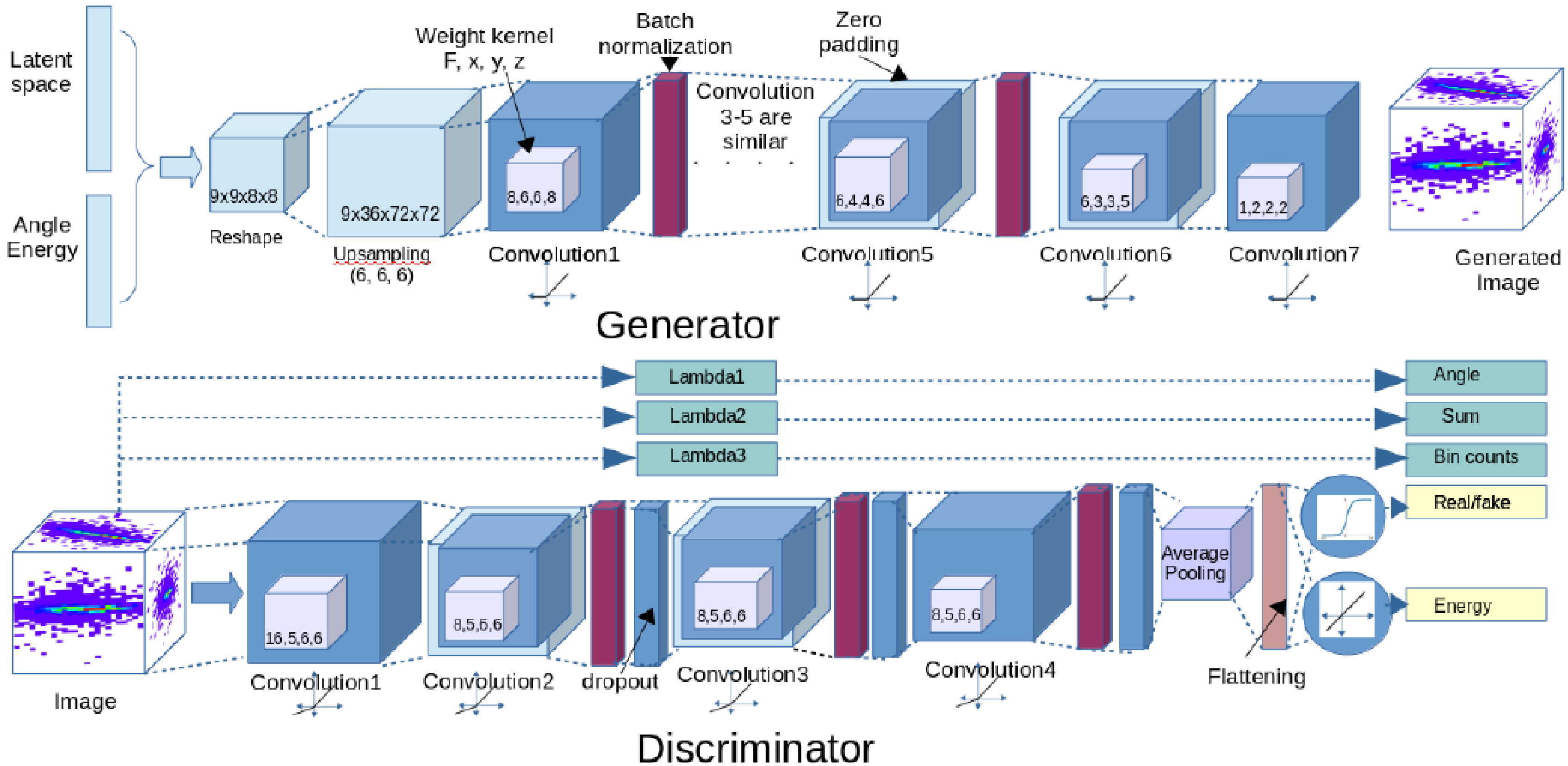
Figure 4.1: Preliminary layout of the tracking system of the CEPC baseline detector concept. The Time Projection Chamber (TPC) is embedded in a Silicon Tracker. Colored lines represent the positions of the silicon detector layers: red lines for the Vertex Detector (VTX) layers; orange lines for the Silicon Inner Tracker (SIT) and Silicon External Tracker (SET) components of the silicon tracker; gray-blue lines for the Forward Tracking Detector (FTD) and Endcap Tracking Detector (ETD) components of the silicon tracker. The cyan lines represent the beam pipe, and the dashed red line shows the beam line position with the beam crossing angle of 16.5 mrad. The ETD line is a dashed line because it is not currently in the full simulation. The radial dimension scale is broken above 350 mm for display convenience.

The model

3D convolutional Generative Adversarial Networks

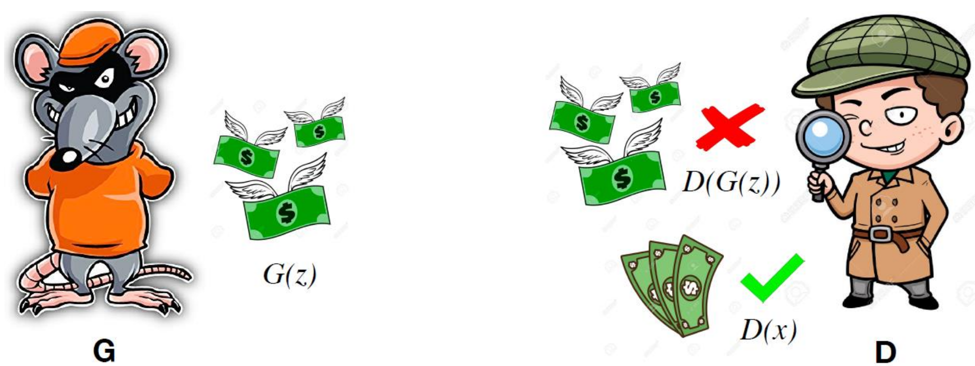
Condition training on input variables, **Custom losses**

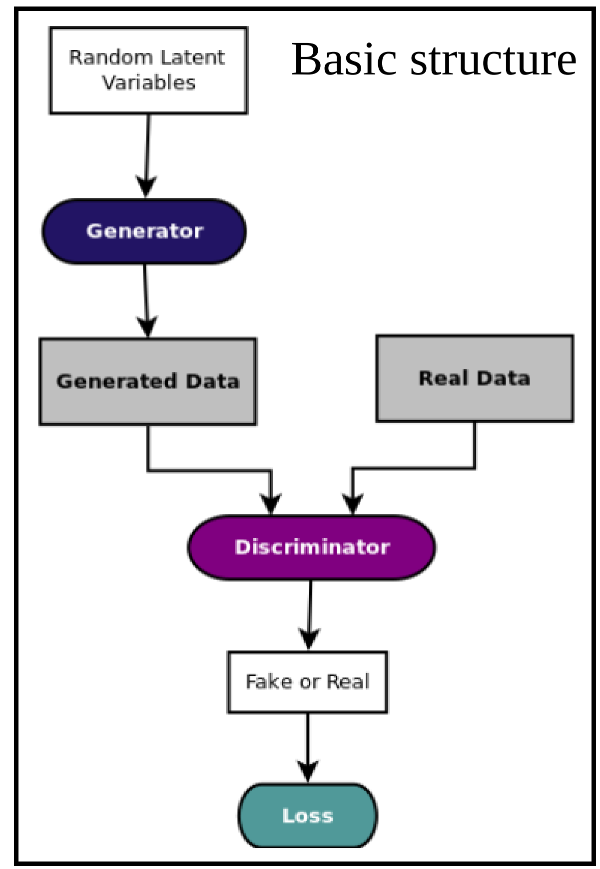
Auxiliary regression tasks assigned to the discriminator



Generative Adversarial Networks (GAN)

- Introduction:
  - As we known, using the MC method to do the full simulation, based on Geant4, requires a tremendous amount of computation resources (e.g. JUNO experiment). It will be difficult to meet a demand resulting from large quantities of data. Therefore, it is needed to develop a faster and efficient algorithm to do the particle detector simulation. Recently, some studies proved the Generative Adversarial Networks (GAN) could be used for particle detector fast simulation.





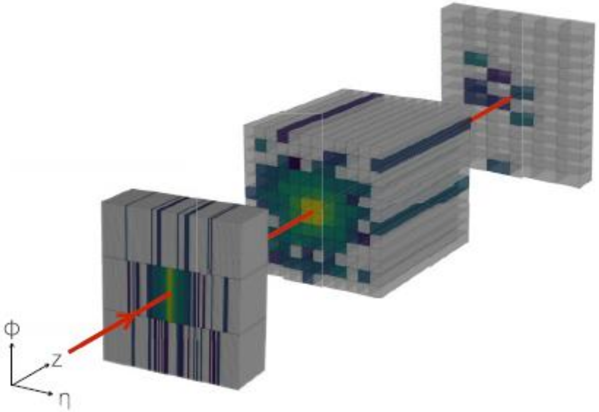
vanilla loss formulation

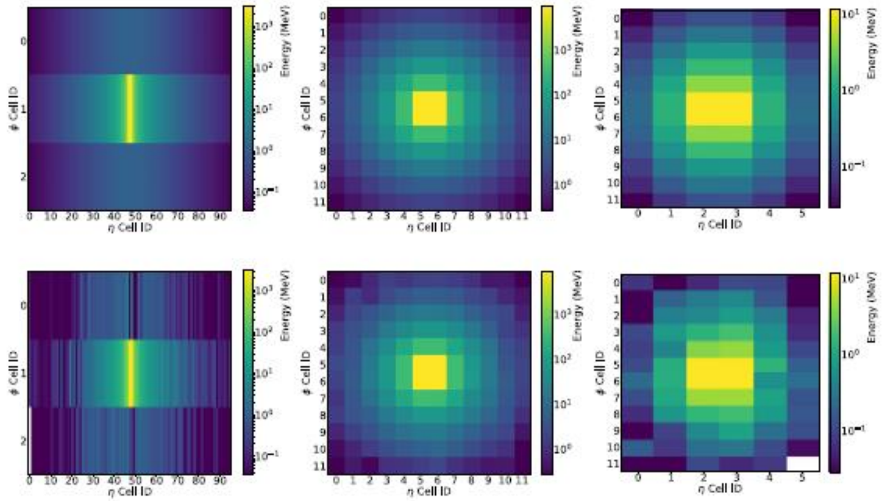
$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[logD(x)] + E_{z\sim p_z(z)}[log(1-D(G(z)))]$$

Here, x is real data, G(z) is fake data

GAN study from others

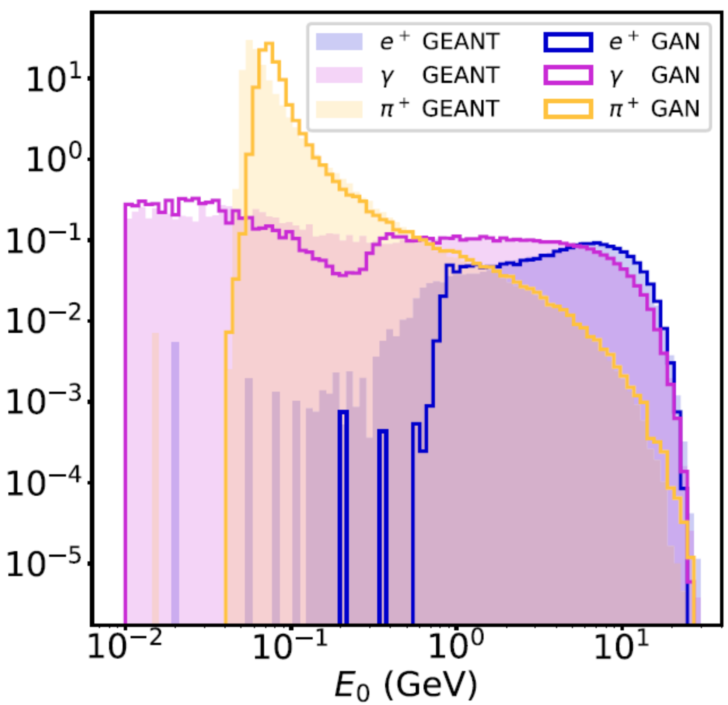
PHYSICAL REVIEW LETTERS 120, 042003 (2018)





e^+ shower from Geant4

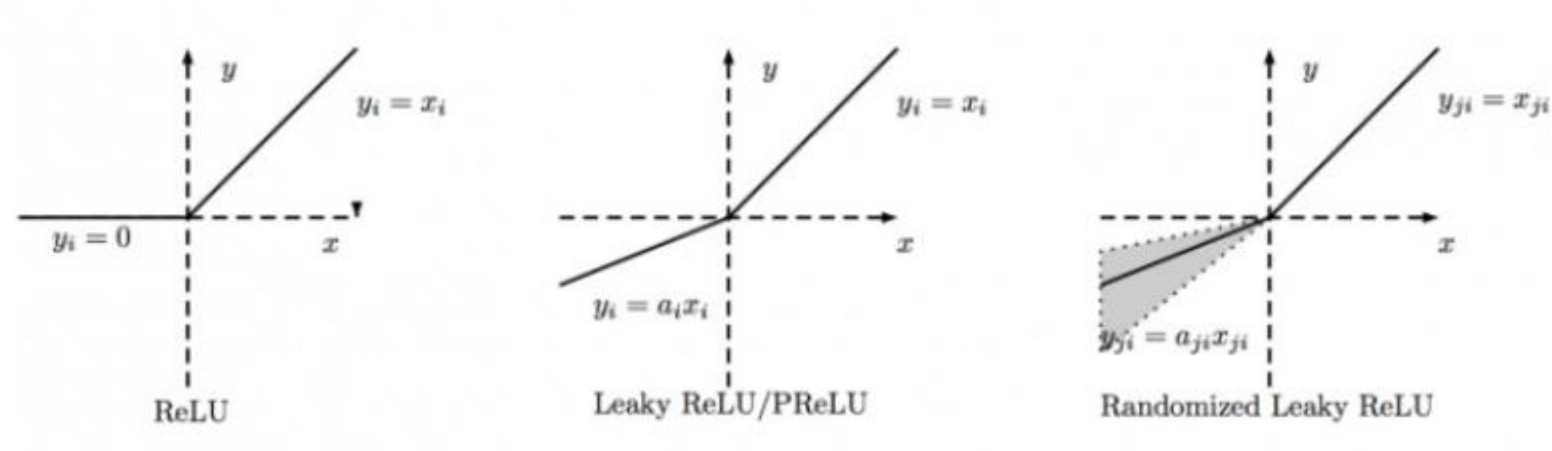
e^+ shower from GAN

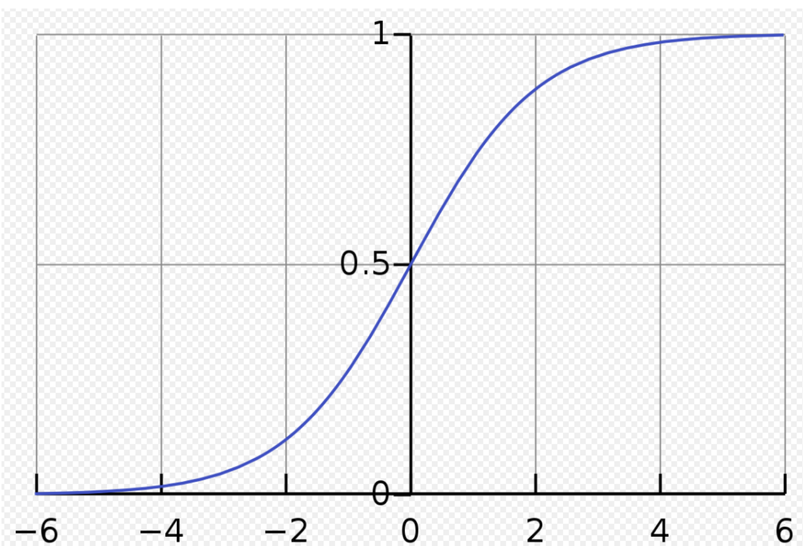


Total expected time (in milliseconds) required to generate a single shower under various algorithm-hardware combinations.

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

Activations





$$\text{sigmod} = \frac{1}{1 + e^{-x}}$$