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

生成网络和高能物理

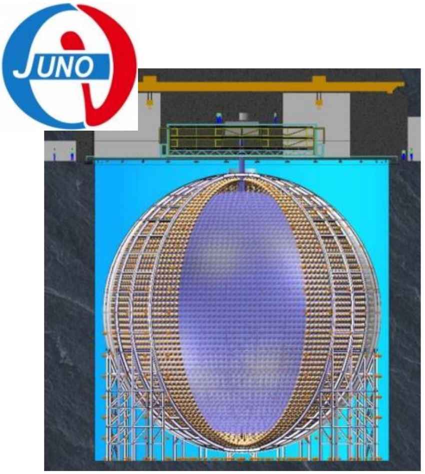
方文兴，中科院高能物理研究所

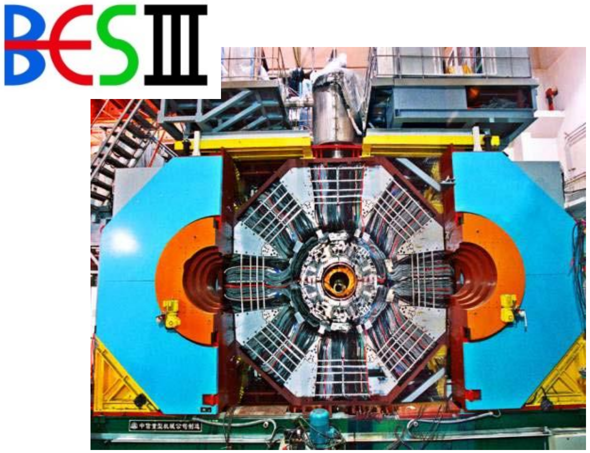
第九届华大 QCD 讲习班

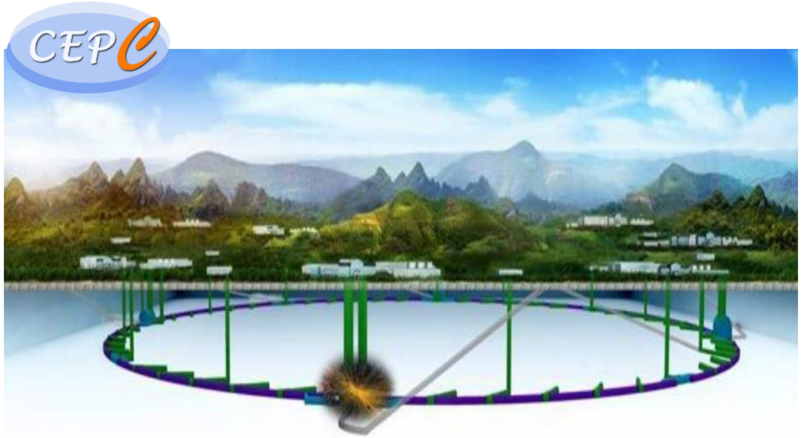
2021年10月15日

提纲

- 生成网络介绍
- JUNO实验的缪子快速模拟
- BESIII实验的dE/dx模拟和量能器模拟
- CEPC实验的量能器模拟





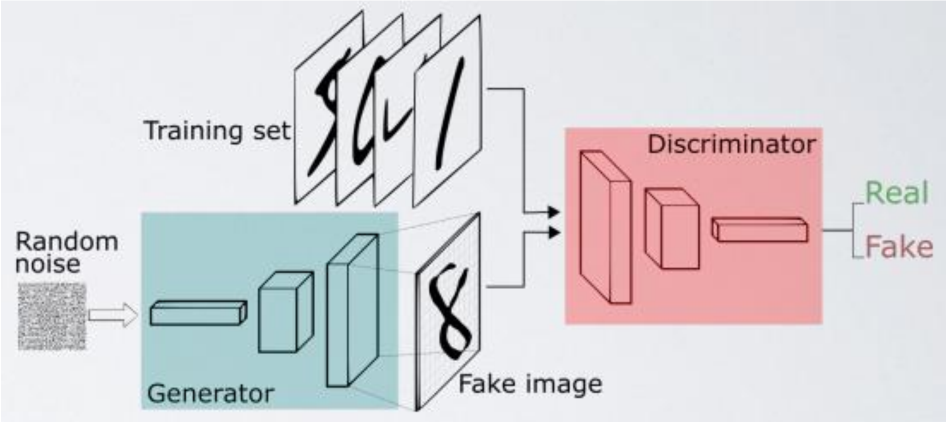


生成对抗网络（GAN）

- 近几年非常火热，主要用于图像的生成

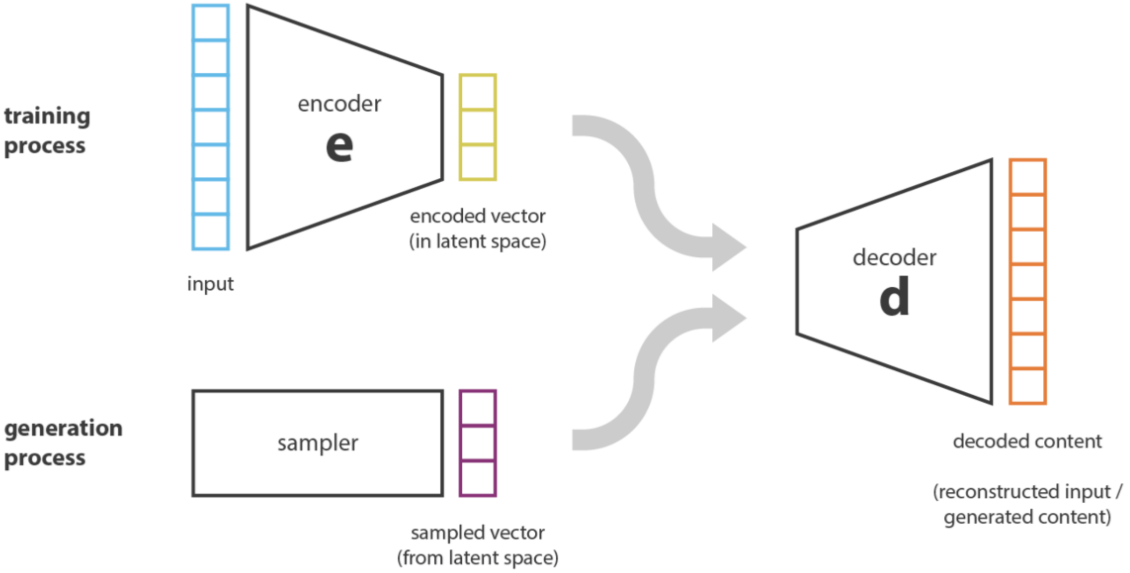
- 该网络由负责产生假数据的生成网络（G）和负责鉴别真和假数据的鉴别网络（D）构成。

- 通过两者相互对抗的训练。即G网络尽可能的产生逼真的数据，D网络尽可能的区分真和假数据。最终实现G网络能生成逼真的数据。



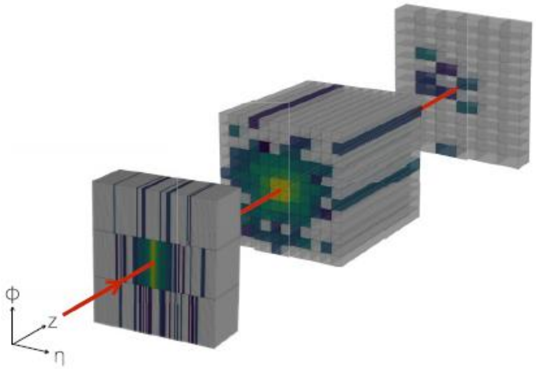
变自编码器（VAE）

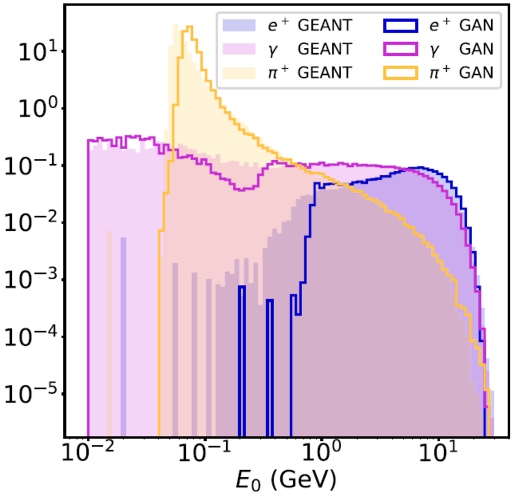
- 与GAN差不多热门
- 由编码网络（encoder）和解码网络（decoder）组成
- 训练时，将原始数据通过 encoder 压缩成低维度数据（ latent space ，同时要求 latent space 服从高斯分布）
- 生成数据时，通过在 latent space 中抽样，作为 decoder 的输入，最终生成数据

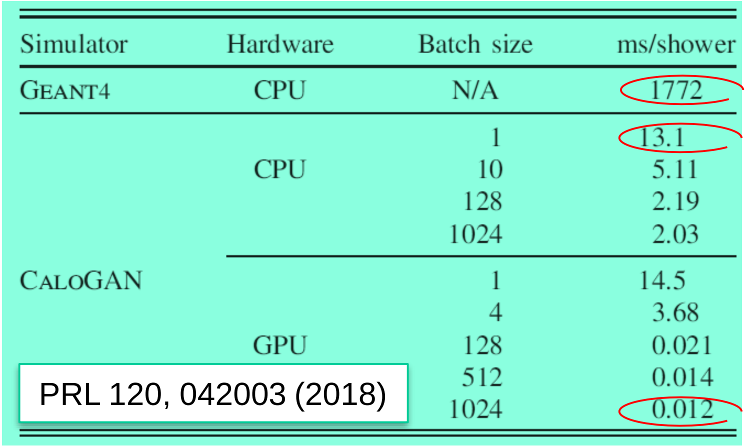


生成网络在高能物理中的例子

- 量能器的快速模拟。通过将量能器的能量沉积信息类比成图像，利用神经网络生成该图像，可以实现量能器的快速模拟（2个量级以上的加速）。



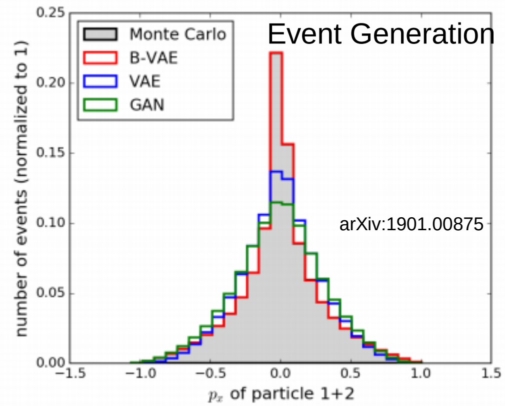




Simulator	Hardware	Batch size	ms/shower
GEANT4	CPU	N/A	1772
CALOGAN	CPU	1	13.1
		10	5.11
		128	2.19
		1024	2.03
	GPU	1	14.5
		4	3.68
		128	0.021
		512	0.014
		1024	0.012

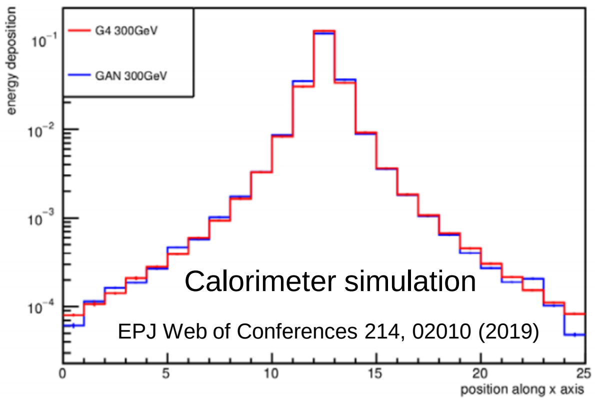
PRL 120, 042003 (2018)

- 产生子的模拟



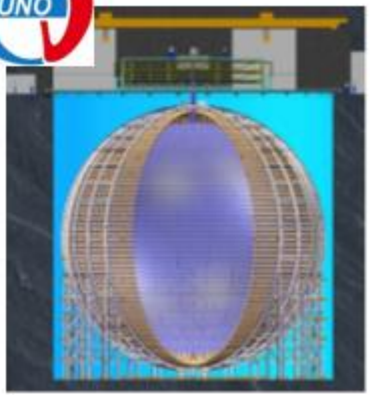
Event Generation

arXiv:1901.00875



Calorimeter simulation

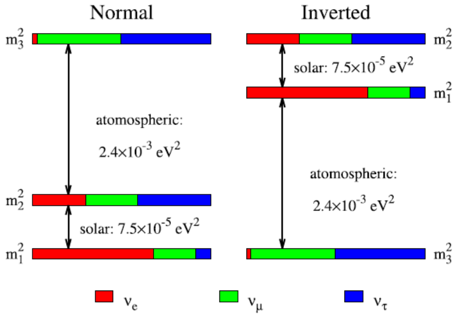
EPJ Web of Conferences 214, 02010 (2019)



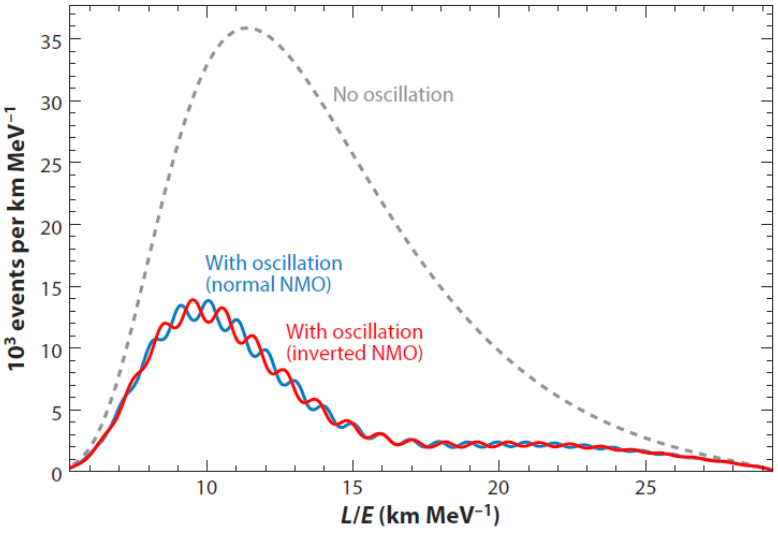
JUNO 实验的缪子快速模拟

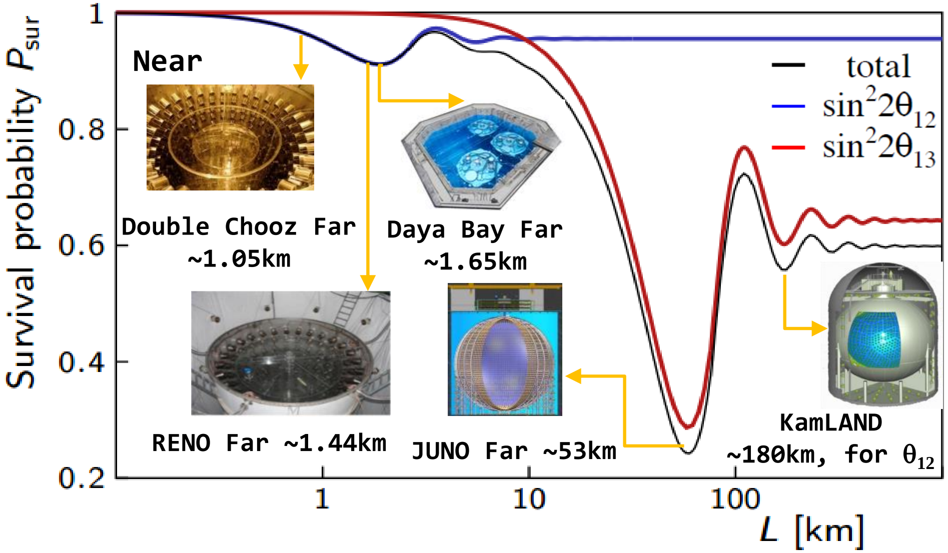
JUNO 实验物理目标

- ❑ 核心目标：测量中微子质量顺序。为最终测量中微子CP相角，破解“反物质消失之谜”奠定坚实基础。



- ❖ 精确测量电子反中微子的能谱，得出Δm^2_{31}（或Δm^2_{32}）的正负号。为了最大程度区分NH和IH，基线长度选择~53kM。

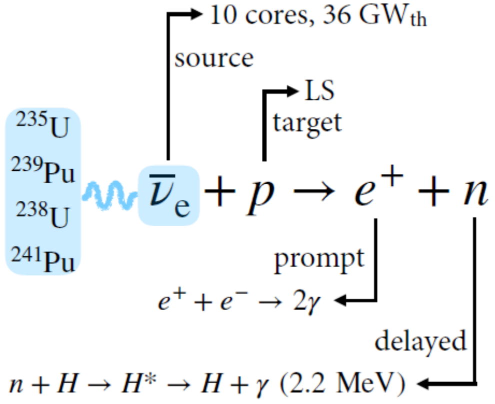




- ➢ 预计运行取数6年后，给出3至4倍σ置信水平的结果。

JUNO 实验原理

- 在JUNO探测器中，反应堆中微子（$\overline{\nu}_e$）与液体闪烁体中的氢核“发生反β衰变反应（IBD）”，产生可探测的信号。
- 因为IBD信号事例率低，所以需要尽可能地压低本底事例。利用IBD事例中的快慢信号的符合测量可以极大的消除本底。

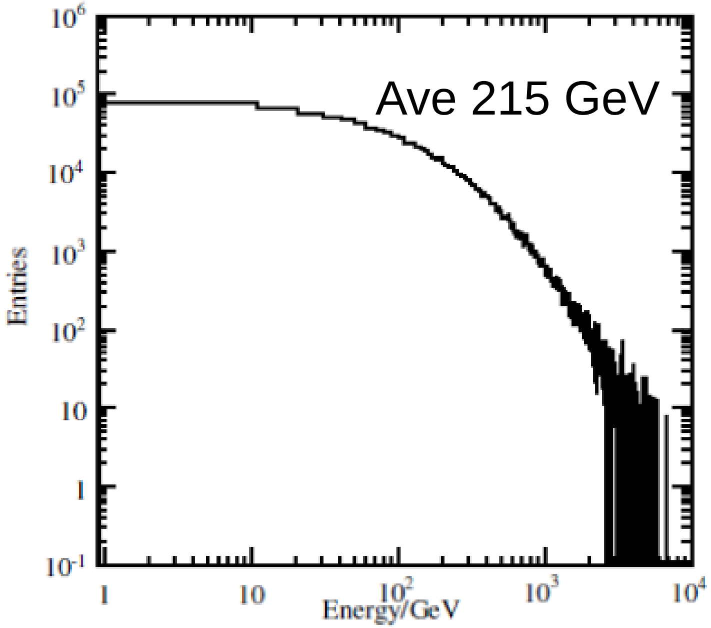


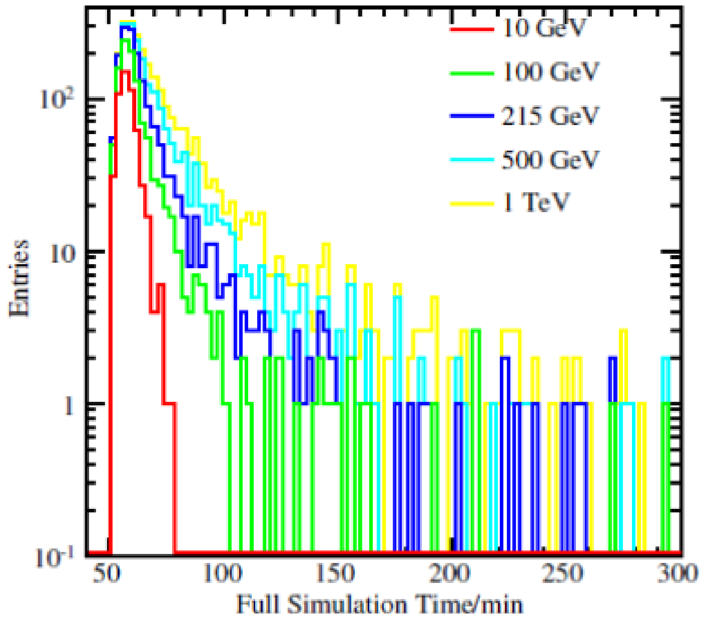
❖ 为了尽可能地降低本底信号比例，JUNO采用了一系列优化后的去本底选择条件，最终使B/S ~ 6%。剩余本底中最大贡献来自宇宙线缪子和液闪反应生成的 ^{9}Li/ ^{8}He所发生的类IBD事例的β –n关联衰变事例。

Selection	IBD efficiency	IBD	Geo-νs	Accidental	^{9}Li/^{8}He	Fast n	(α, n)
—	—	83	1.5	~5.7 × 10^4	84	—	—
Fiducial volume	91.8%	76	1.4		77	0.1	0.05
Energy cut	97.8%			410			
Time cut	99.1%	73	1.3		71		
Vertex cut	98.7%			1.1			
Muon veto	83%	60	1.1	0.9	1.6		
Combined	73%	60	3.8				

JUNO 实验缪子模拟

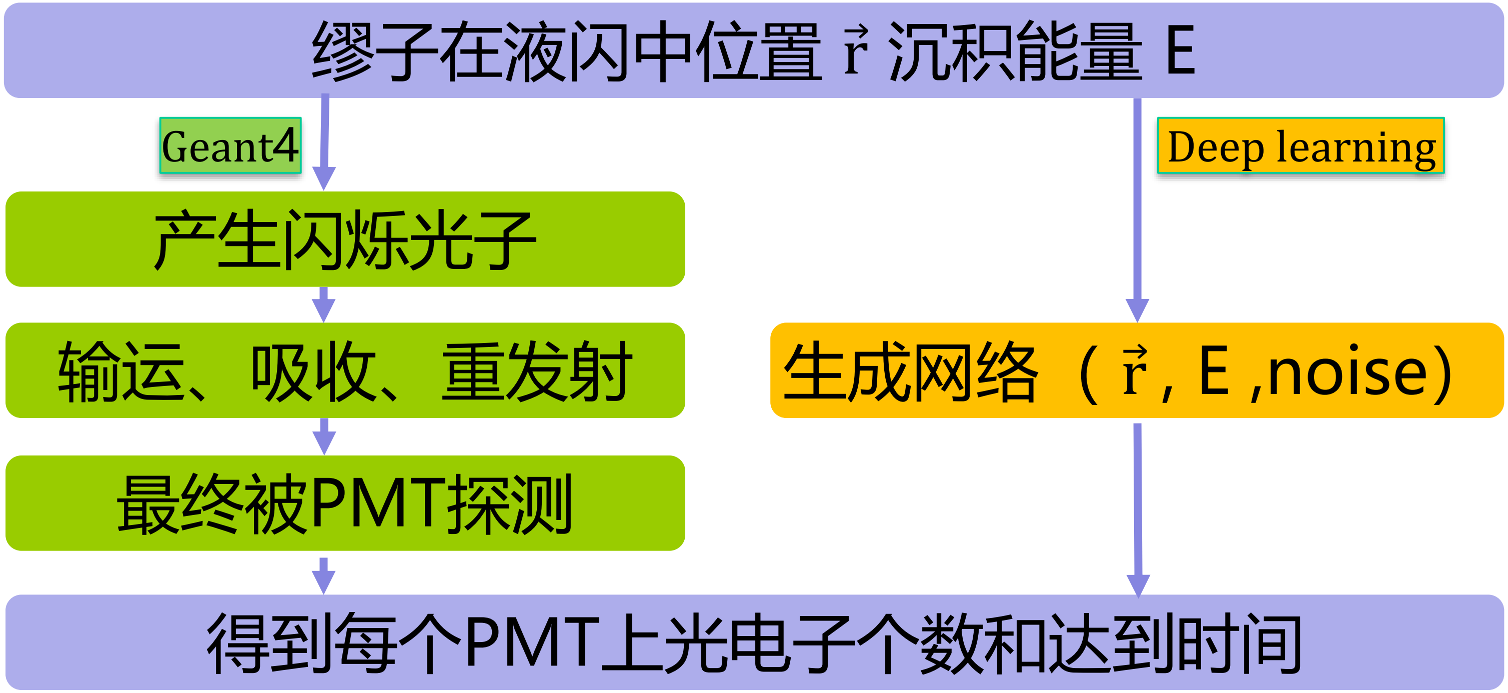
- 为了仔细研究由宇宙线缪子带来的本底（ ^{9}Li/ ^{8}He， 快中子）和利用缪子的重建信息来去除其产生的本底，需要产生大量的缪子模拟事例。
- 在Geant4模拟中，平均能量为215GeV的宇宙线缪子会在液闪中产生~ 10^7 闪烁光子。要模拟这些海量的光子在液闪中的吸收、重发射、输运等物理过程需要消耗大量计算资源。因此，发展和采用新的技术来加速缪子的模拟意义重大。





JUNO 实验缪子快速模拟

- 该研究利用生成网络代替缪子模拟中最消耗计算资源的模拟过程，即海量闪烁光子在液闪中传播及和液闪发生相互作用的过程。



缪子在液闪中位置 $\vec{r}$ 沉积能量 E

Geant4

产生闪烁光子

输运、吸收、重发射

最终被PMT探测

Deep learning

生成网络（$\vec{r}$, E ,noise)

得到每个PMT上光电子个数和达到时间

JUNO 实验缪子快速模拟

- 分别训练光电子个数模拟网络（NN_npe）和到达时间模拟网络（NN_time）。
- 假设JUNO探测器是球对称的，根据缪子能量沉积位置R以及能量沉积位置和LPMT的夹角θ，给出这个PMT上的光电子个数和到达时间。

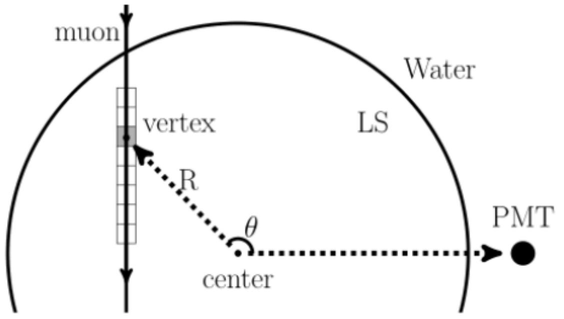
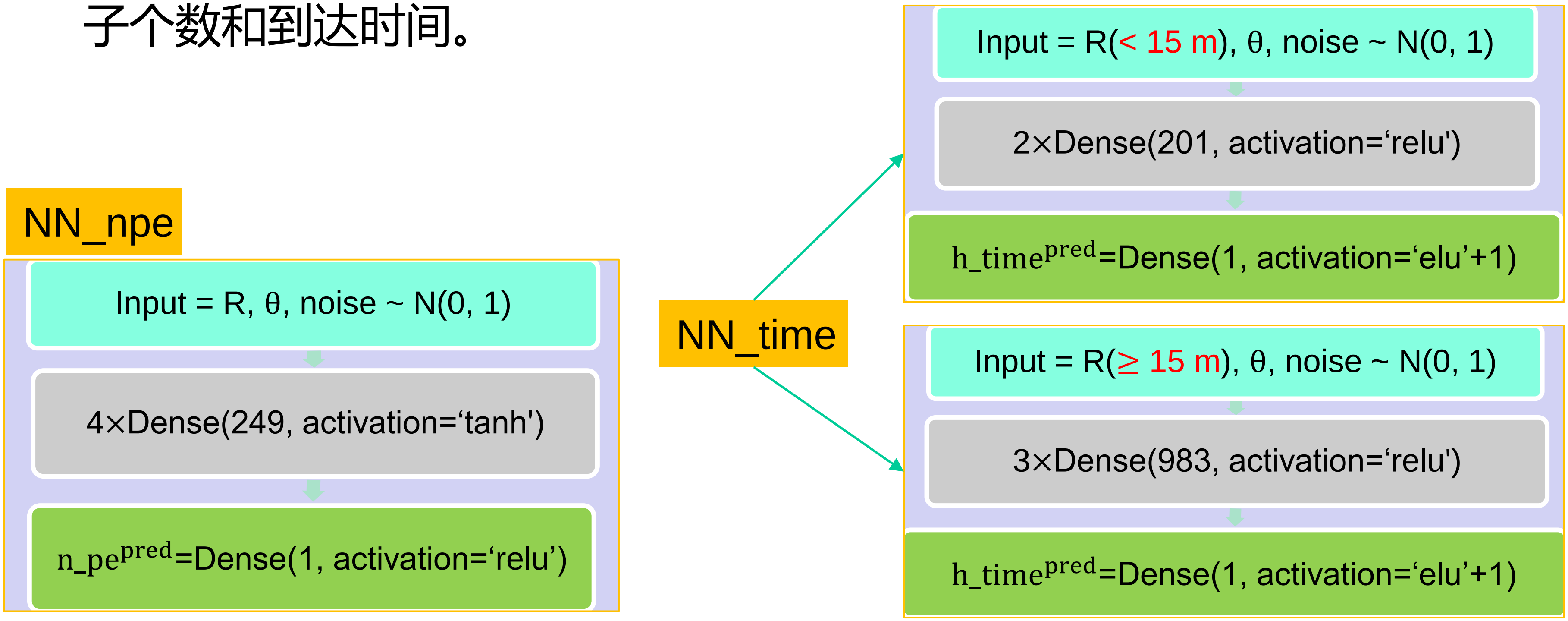


图 4.8: voxel method示意图



NN_npe

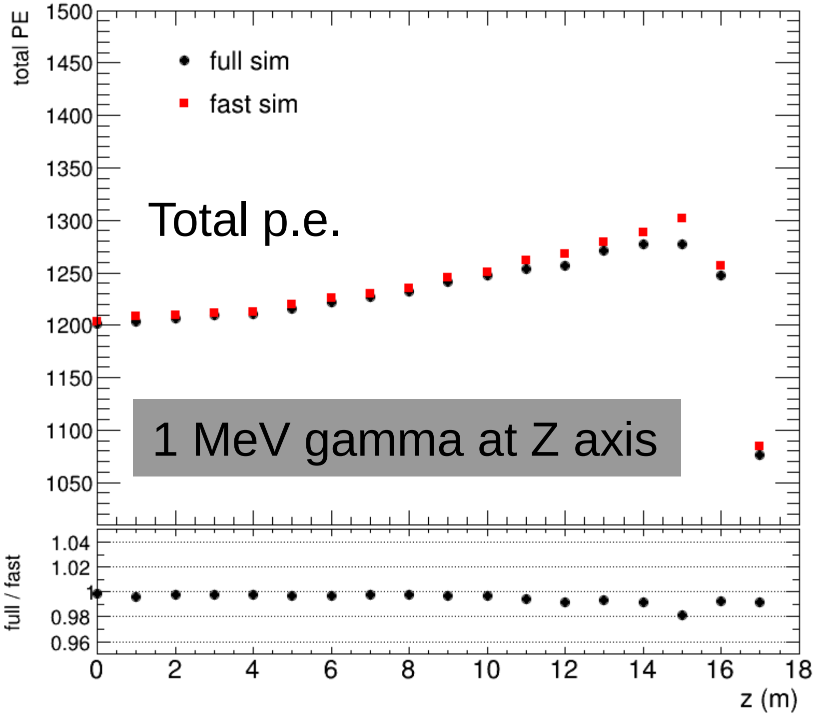
- Input = R, θ, noise ~ N(0, 1)
- 4×Dense(249, activation=‘tanh')
- n_pe^{pred}=Dense(1, activation=‘relu’)

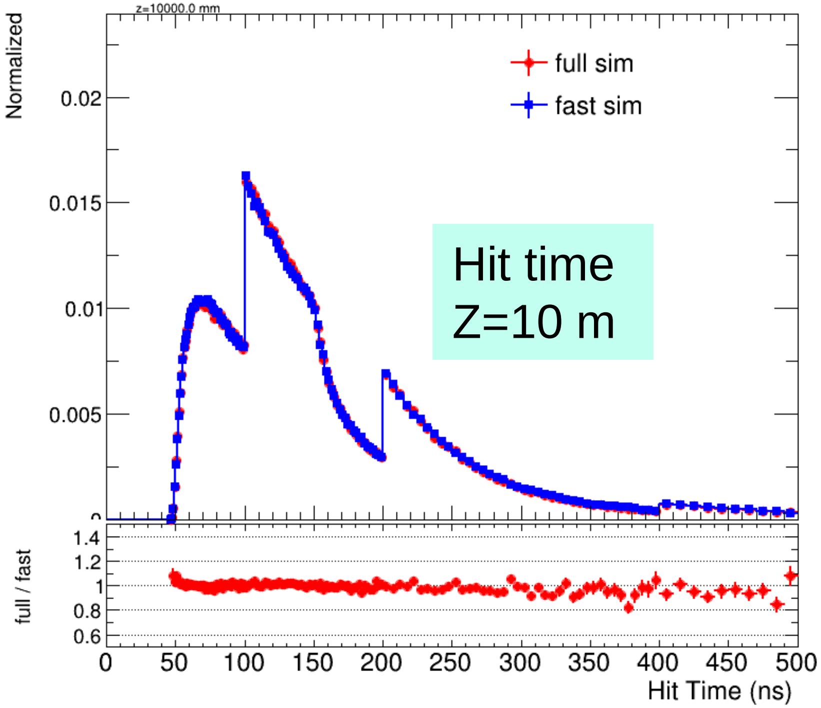
NN_time

- Input = R(< 15 m), θ, noise ~ N(0, 1)
- 2×Dense(201, activation=‘relu')
- h_time^{pred}=Dense(1, activation=‘elu’+1)

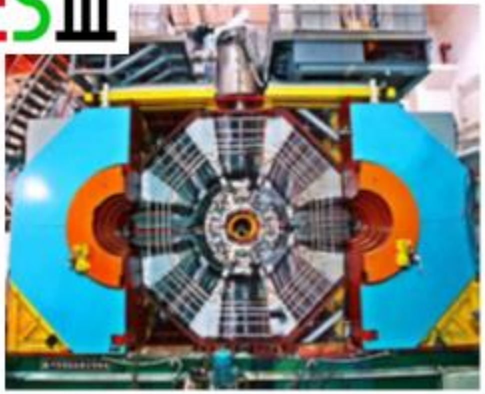
- Input = R(≥ 15 m), θ, noise ~ N(0, 1)
- 3×Dense(983, activation=‘relu')
- h_time^{pred}=Dense(1, activation=‘elu’+1)

JUNO 实验缪子快速模拟





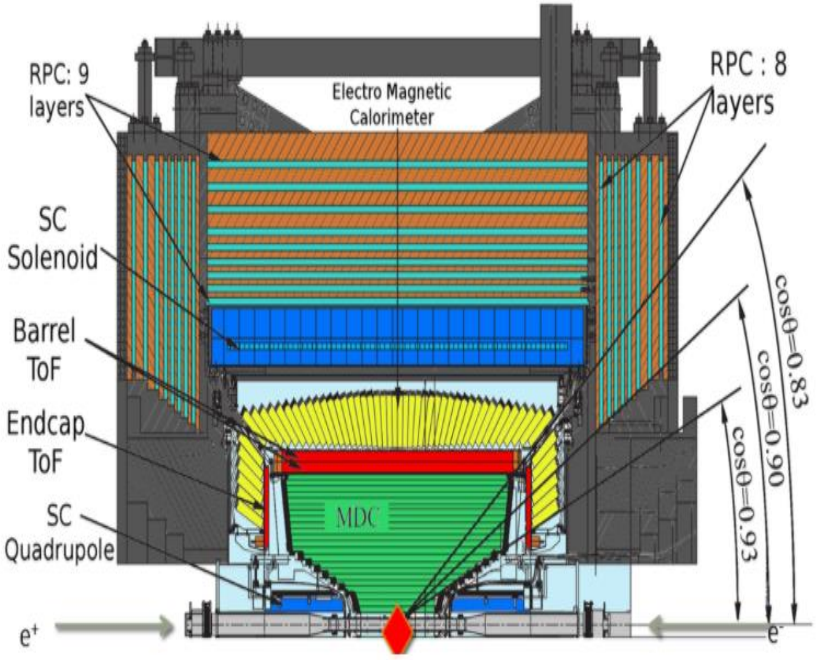
- ➢ 快速模拟和全模拟在大体上符合，有些区间存在优化的空间。
- ➢ 实现近30倍的缪子模拟的加速



BESIII 实验的 dE/dx 模拟和量能器模拟

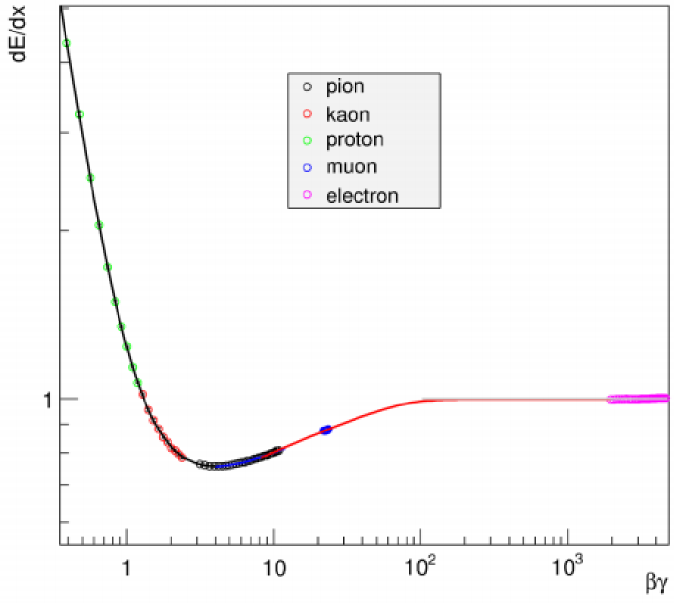
BESIII 实验物理现状

- 北京谱仪（BES）III实验通过正负电子在τ-charm能区对撞，研究陶物理、粲偶素、含粲强子等方面的物理。具有世界领先地位。
  - 面临激烈的国际竞争：BelleII，LHCb等实验
- BESIII目前积累了世界上最多的$e^+e^- \to J/\Psi$过程的数据（10^{10}个事例）
  - 统计误差大大减小
  - 系统误差成为限制物理结果精度的瓶颈，例如粒子鉴别效率。



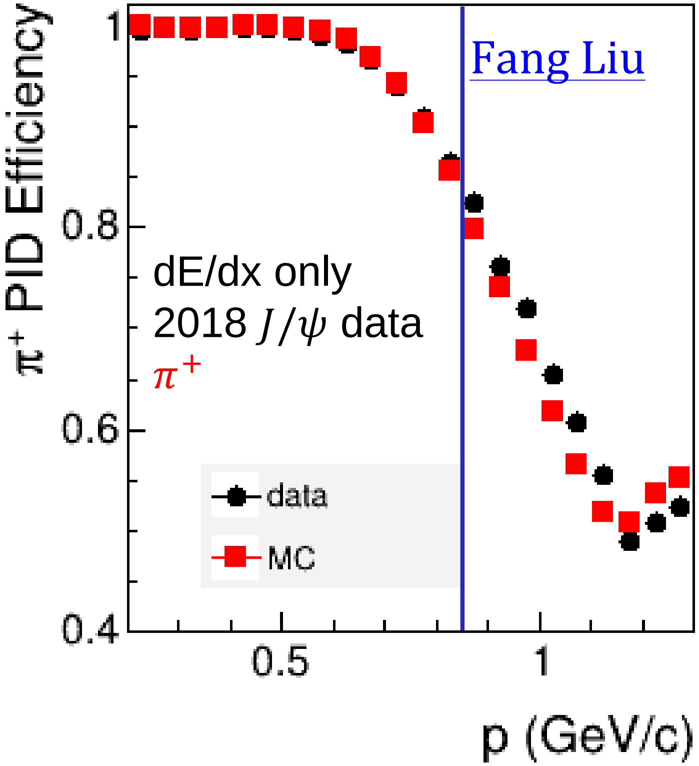
dE/dx

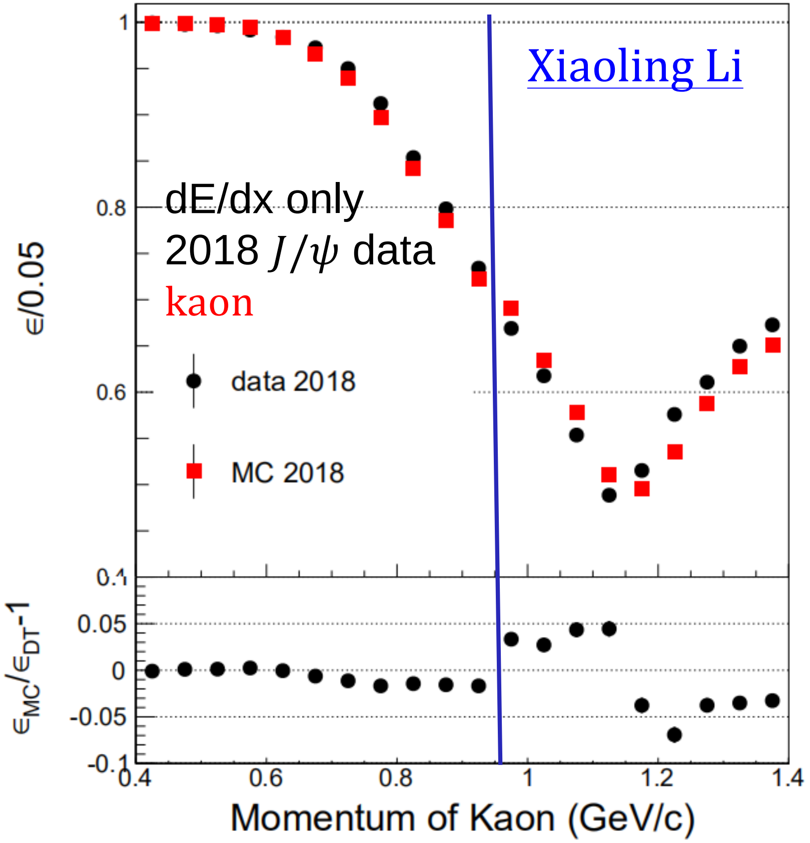
- 单位长度能量沉积，dE/dx，是用于粒子鉴别的重要测量量。$\mathrm{m} = \dfrac{p\ （\text{来自}MDC）}{v\ （\text{来自}\mathrm{d}E/dx）}$
- 影响 dE/dx 测量值的因素：
  - 粒子类型，电荷，MDC cell, 气体压力（温度）等等



- 目前通过一套复杂的 dE/dx 刻度流程，可以得到无偏的 dE/dx 测量值
- 由于 Geant4 不能很好地模拟带电粒子在稀薄气体中的能量沉积，因此对 dE/dx 的模拟，目前 BESIII 采用从真实数据中分 bin 抽样的方法

BESIII dE/dx PID 效率





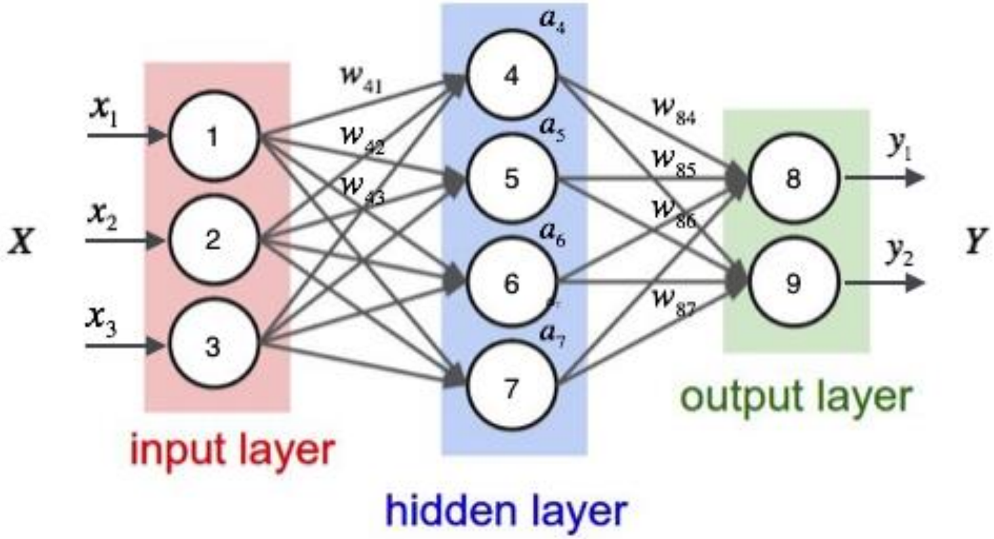
- 在高动量区间，数据和模拟的符合不是很好，有较大的改善空间
- 传统方法进展缓慢，需要用新的方法

dE/dx 的机器学习模拟方法

- 采用机器学习的方法，去模拟 dE/dx
- 步骤：
- 1, 从真实数据中学习刻度后的 dE/dx 的分布
  - dE/dx distribution = $f(\text{type}, \text{P}, \theta, \ \ldots)$, 考虑的因素越多，模拟的越细致
  - 由神经网络完成
- 2, 将训练好的神经网络在 BOSS 软件中运行，用于模拟 dE/dx
  - 检查数据和模拟 dE/dx 分布的一致性
- 3, 得到数据和模拟的 dE/dx PID 效率，检查一致性
  - 预期能改善数据和模拟的一致性，从而实现减小系统误差的目标

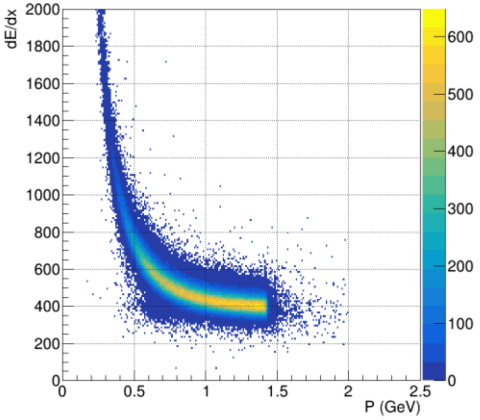
dE/dx 模拟网络

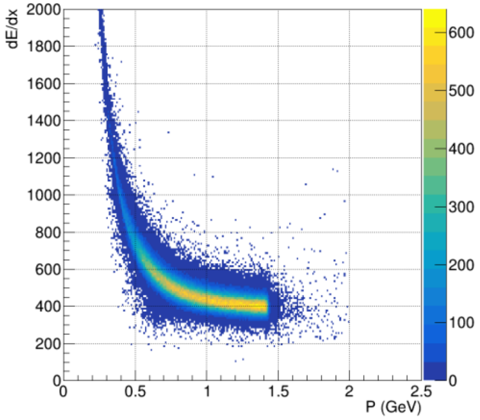
- 从真实数据中学习每条带电径迹的 dE/dx
- 模型: 全连接神经网络
  - 由输入层，隐藏层，输出层组成
  - 通过贝叶斯方法优化隐藏层的个数和每层神经元的个数
- 输入数据:
  - 粒子类别
  - 径迹的动量
  - 径迹的极角 θ
  - 径迹的好击中个数
  - 高斯分布的随机数N(0,1)
- 输出: 径迹的 dE/dx 值
- Loss: 两样本统计检验 (SmoothKNN)



dE/dx distribution (K^-)

- The distribution of dE/dx vs P (or θ) is compared between NN and data



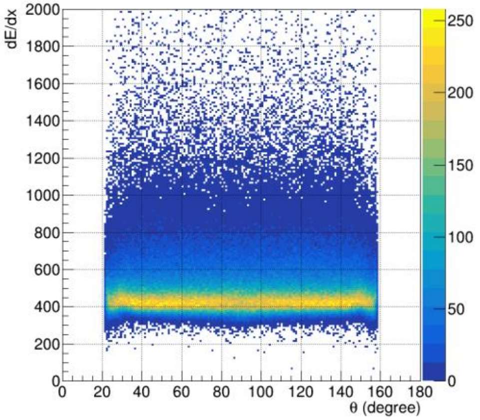


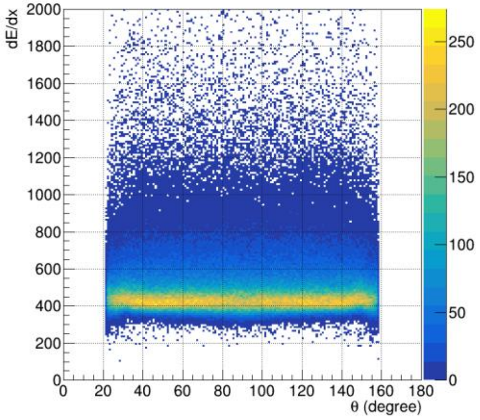
dE/dx vs P

Data

NN

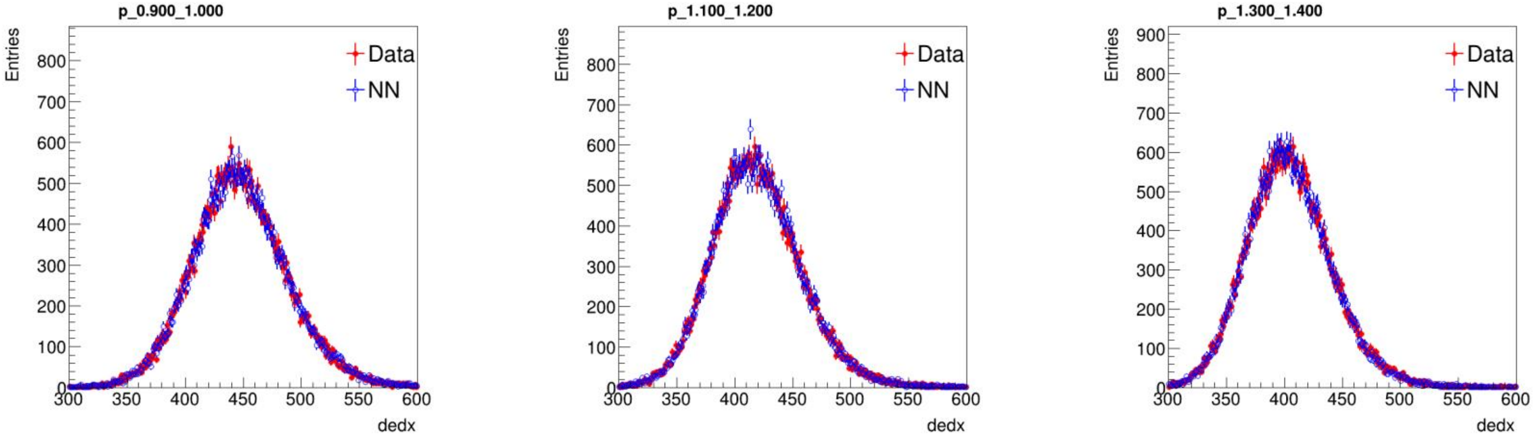
- Similar for K^+





dE/dx vs θ

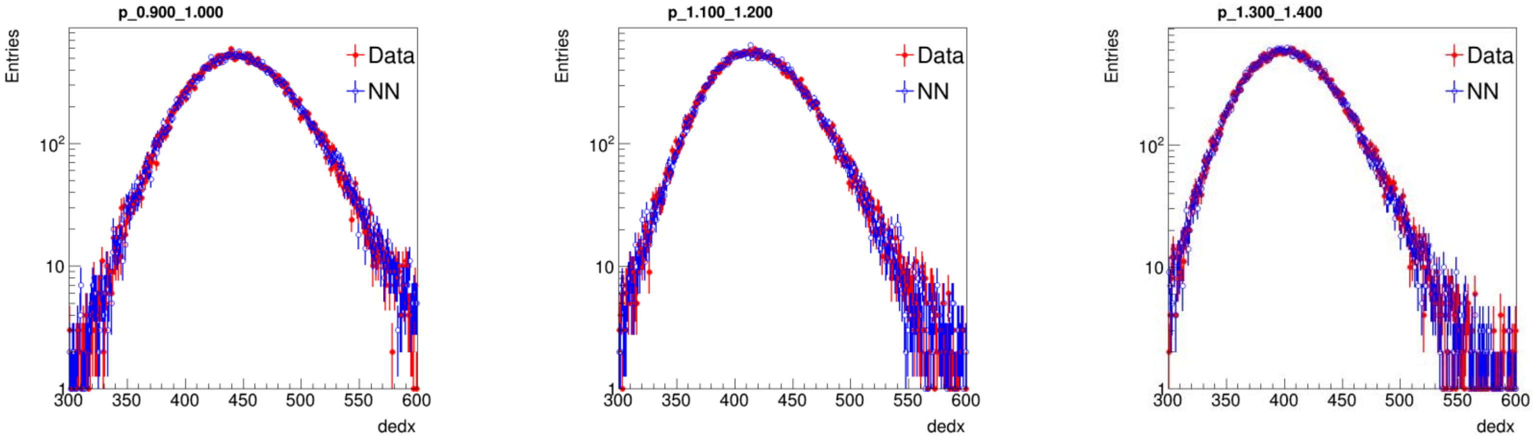
dE/dx distribution (K^-) vs P



P: 0.9-1.0 GeV

P: 1.1-1.2 GeV

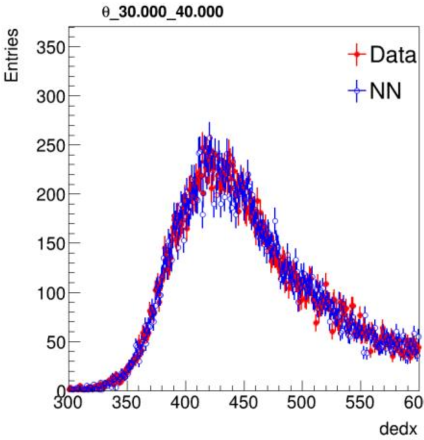
P: 1.3-1.4 GeV

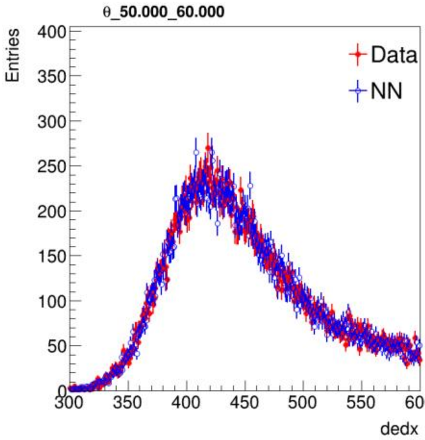


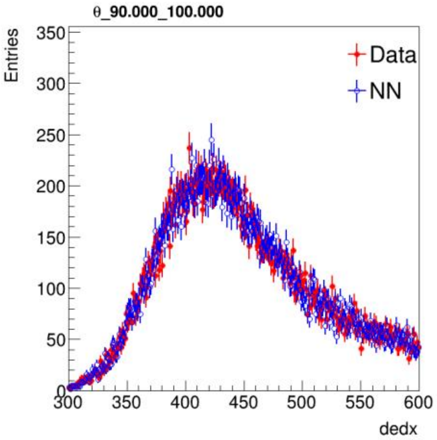
Log scale

- The nice agreement between data and NN

dE/dx distribution (K^-) vs θ



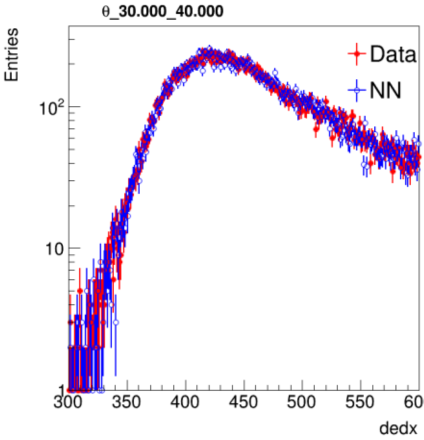


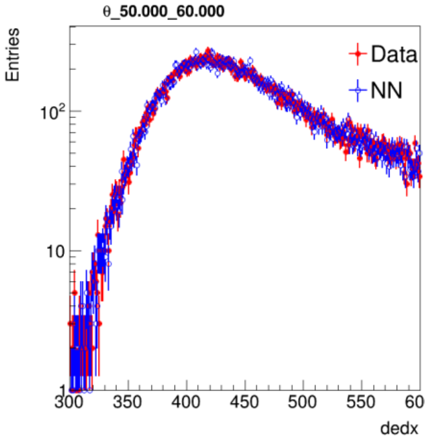


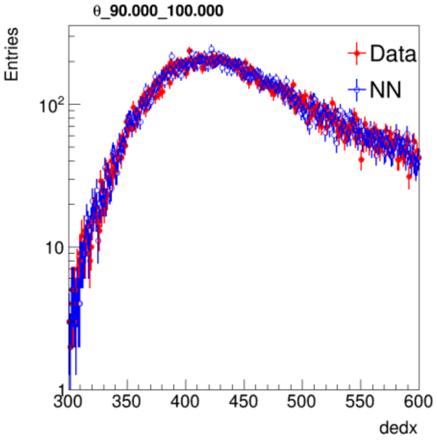
$\theta: 30^\circ - 40^\circ$

$\theta: 50^\circ - 60^\circ$

$\theta: 90^\circ - 100^\circ$



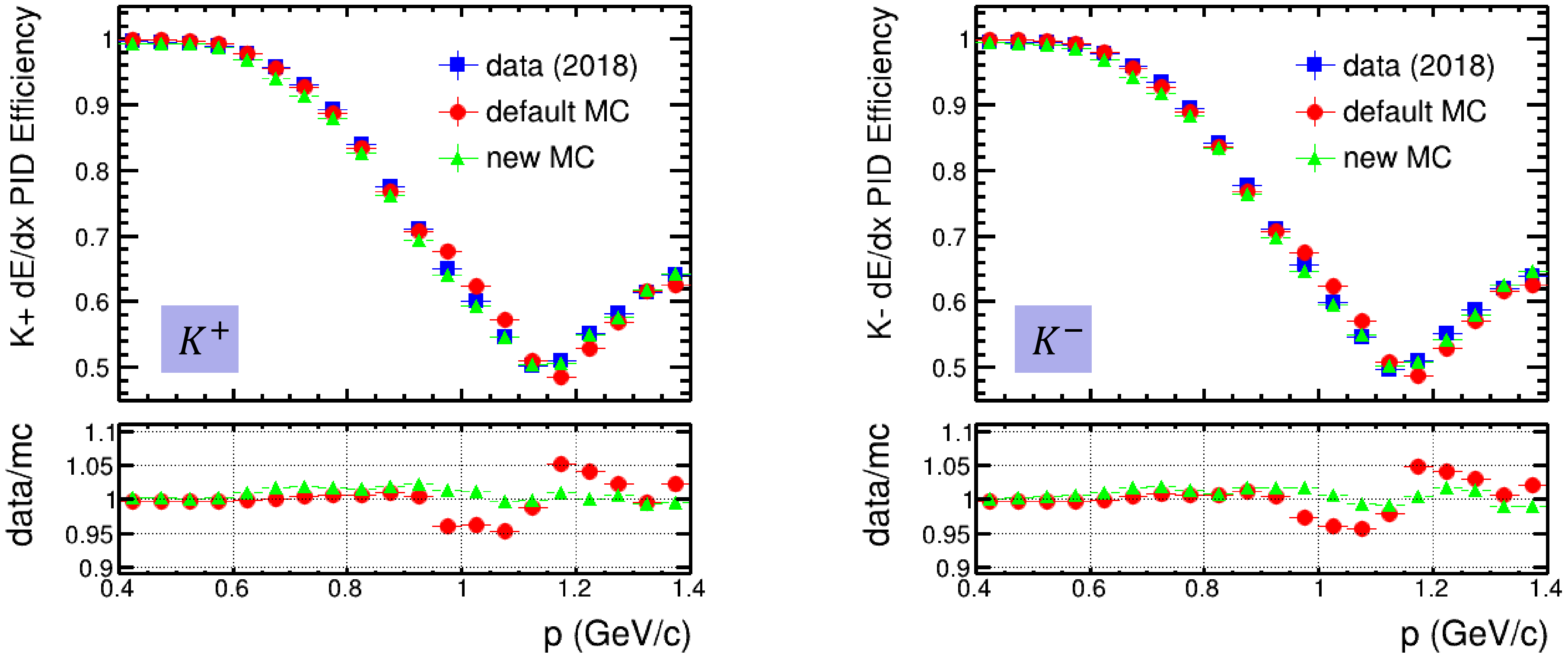




Log scale

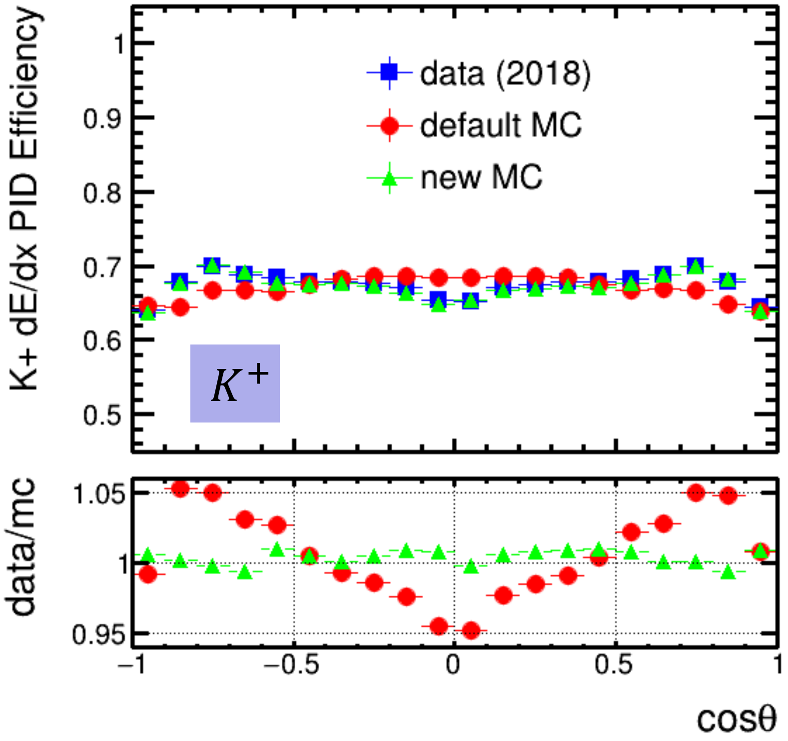
- The nice agreement between data and NN

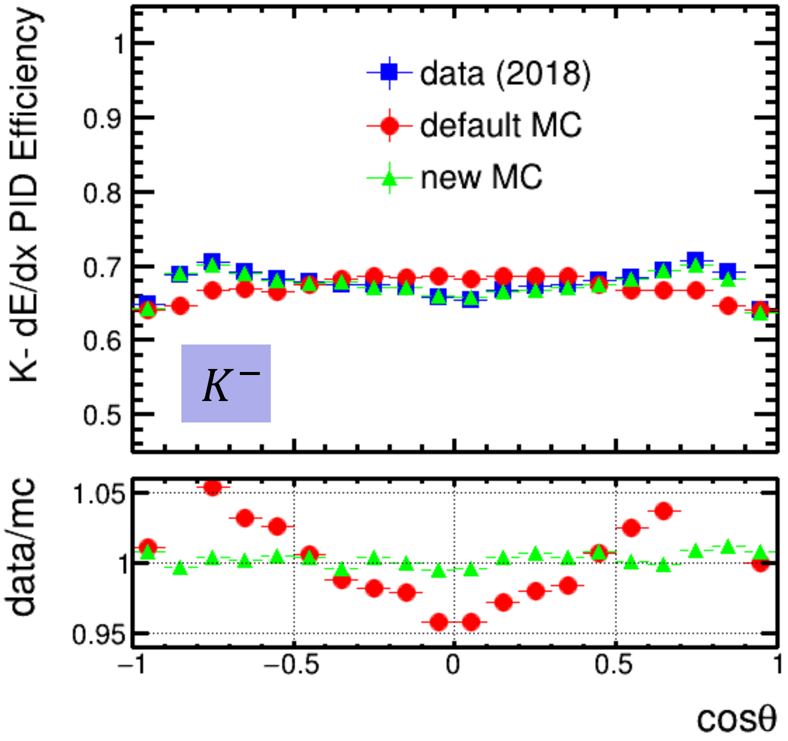
Kaon dE/dx PID efficiency vs P



- 神经网络模拟的结果与真实数据符合的更好

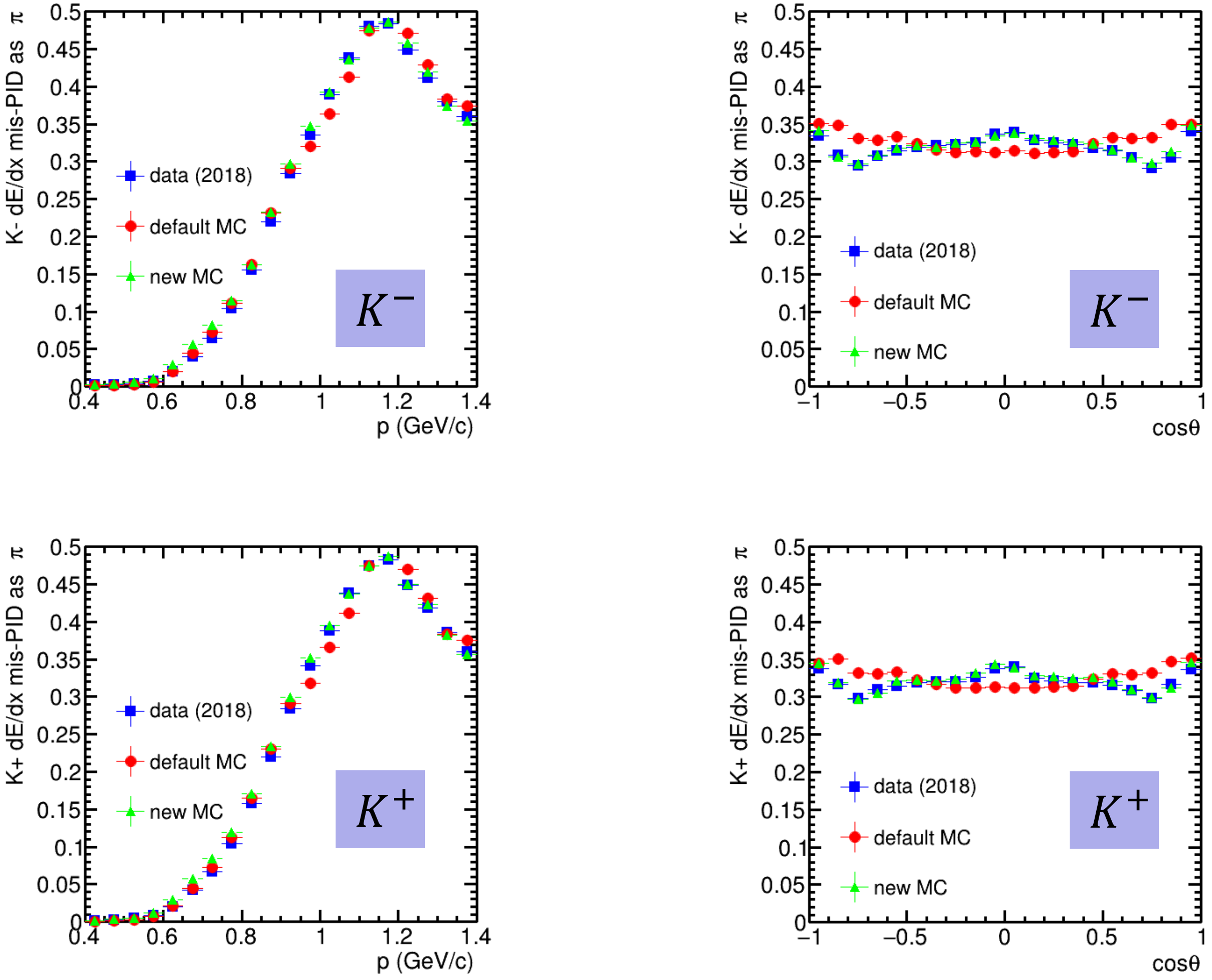
Kaon dE/dx PID efficiency vs cosθ





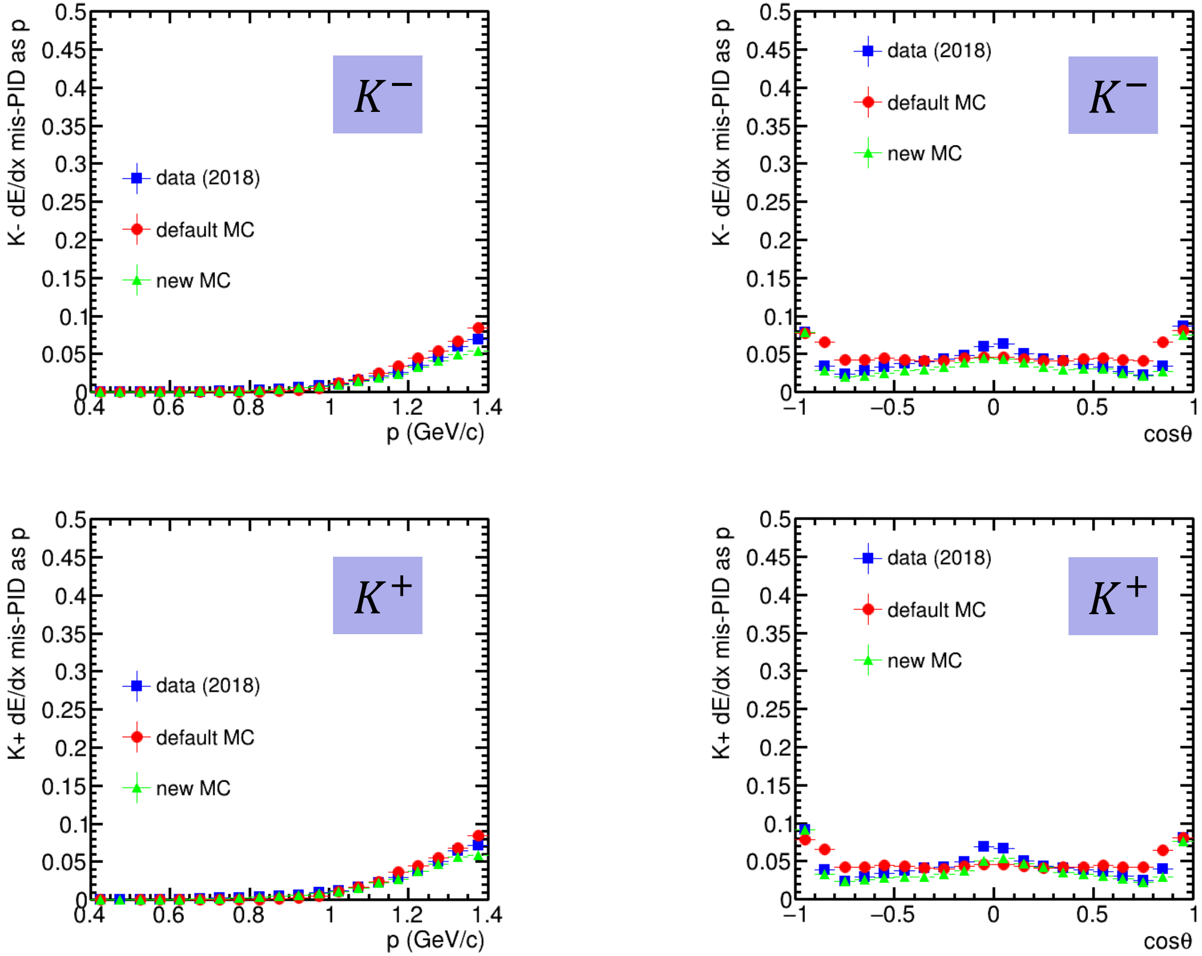
- 神经网络模拟的结果与真实数据符合的更好

Kaon dE/dx mis-PID as pion



- 对于误鉴别率，神经网络模拟结果也与真实数据符合的更好

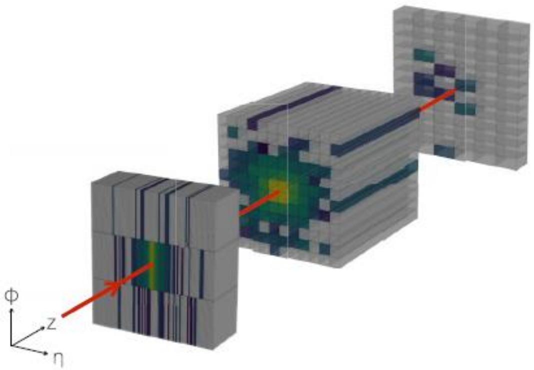
Kaon dE/dx mis-PID as (anti-)proton

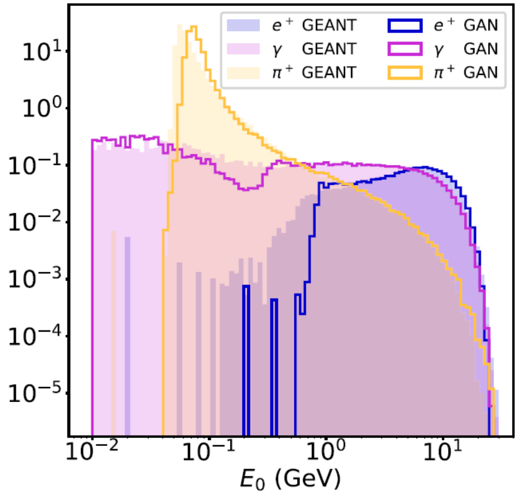


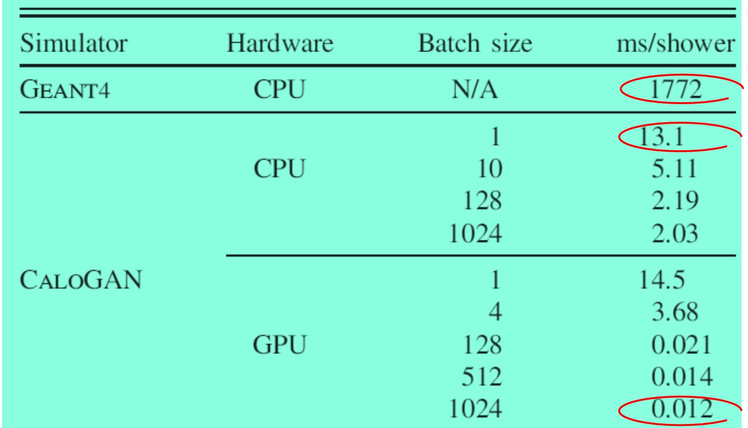
- 对于误鉴别率，神经网络模拟结果也与真实数据符合的更好
- 其它粒子的结果见 backup

量能器模拟

- 从已有的研究（CaloGAN）可知，利用神经网络去模拟粒子在量能器上的簇射能实现2个数量级以上的加速。虽然物理性能上与Geant4模拟还存在一定的差别，但是是非常值得研究的工作。
- 量能器的快速模拟对 CEPC 实验来说有重要的意义（后面会提到）
- 对于 BESIII 实验来说，可以通过学习不同粒子在量能器上的簇射来改善模拟与数据的一致性





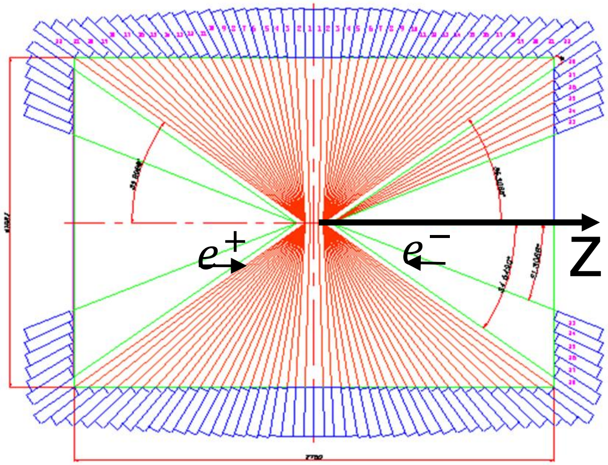


Simulator	Hardware	Batch size	ms/shower
GEANT4	CPU	N/A	1772
CALOGAN	CPU	1	13.1
		10	5.11
		128	2.19
		1024	2.03
	GPU	1	14.5
		4	3.68
		128	0.021
		512	0.014
		1024	0.012

PRL 120, 042003 (2018)

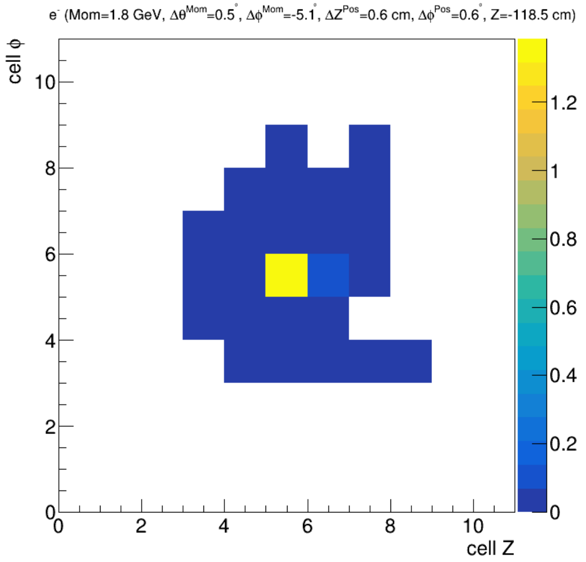
训练数据

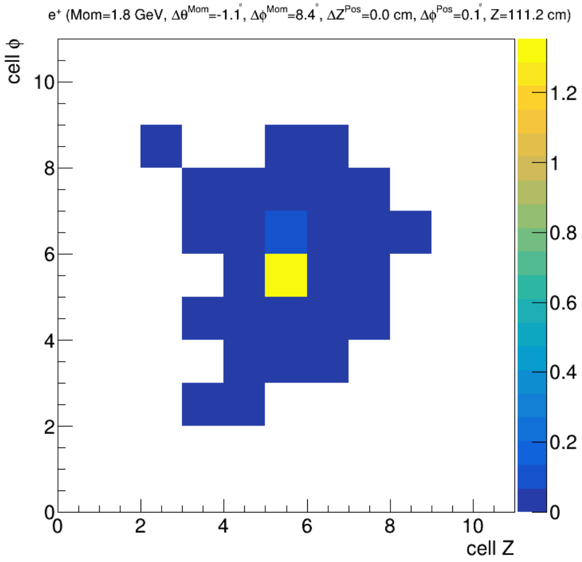
- 利用 Bhabha 模拟事例作为训练数据。模拟桶部量能器。
- 通过对$e^{\pm}$径迹的外推，得到其在量能器上的击中位置。以该位置作为中心，模拟11×11的晶体组。



$e^-(\mathrm{Mom} = 1.8\ \mathrm{GeV}, \Delta\theta^{\mathrm{Mom}} = 0.5^{\circ}, \Delta\phi^{\mathrm{Mom}} = -5.1^{\circ}, \Delta Z^{\mathrm{Pos}} = 0.6\ \mathrm{cm}, \Delta\phi^{\mathrm{Pos}} = 0.6^{\circ}, Z = -118.5\ \mathrm{cm})$

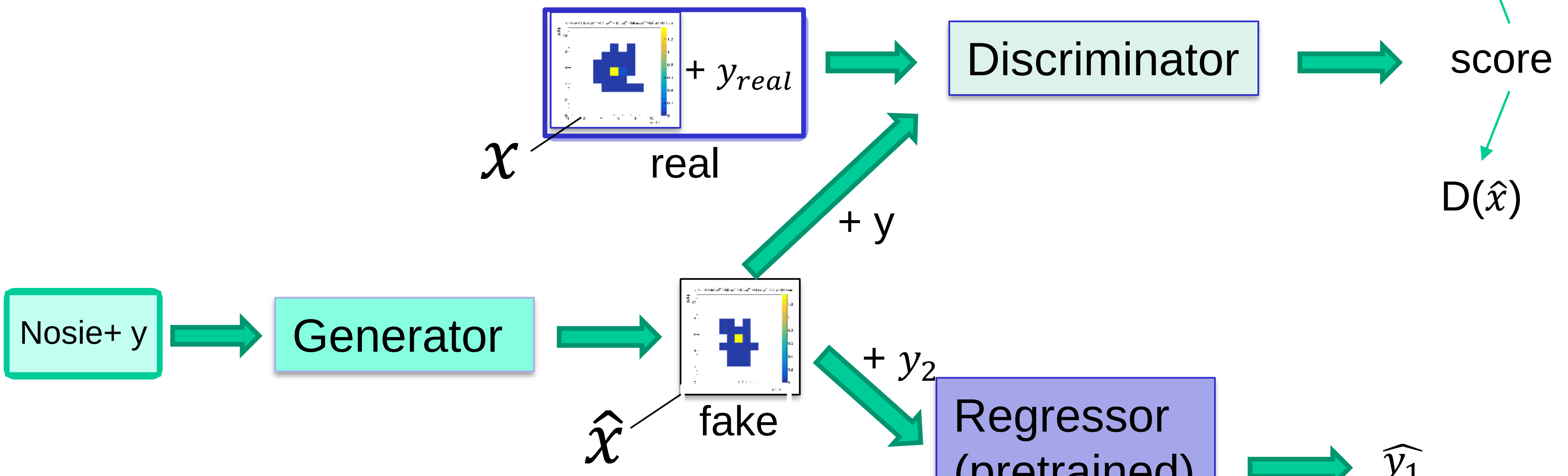
$e^+(\mathrm{Mom} = 1.8\ \mathrm{GeV}, \Delta\theta^{\mathrm{Mom}} = -1.1^{\circ}, \Delta\phi^{\mathrm{Mom}} = 8.4^{\circ}, \Delta Z^{\mathrm{Pos}} = 0.0\ \mathrm{cm}, \Delta\phi^{\mathrm{Pos}} = 0.1^{\circ}, Z = 111.2\ \mathrm{cm})$





BESIII 生成对抗网络

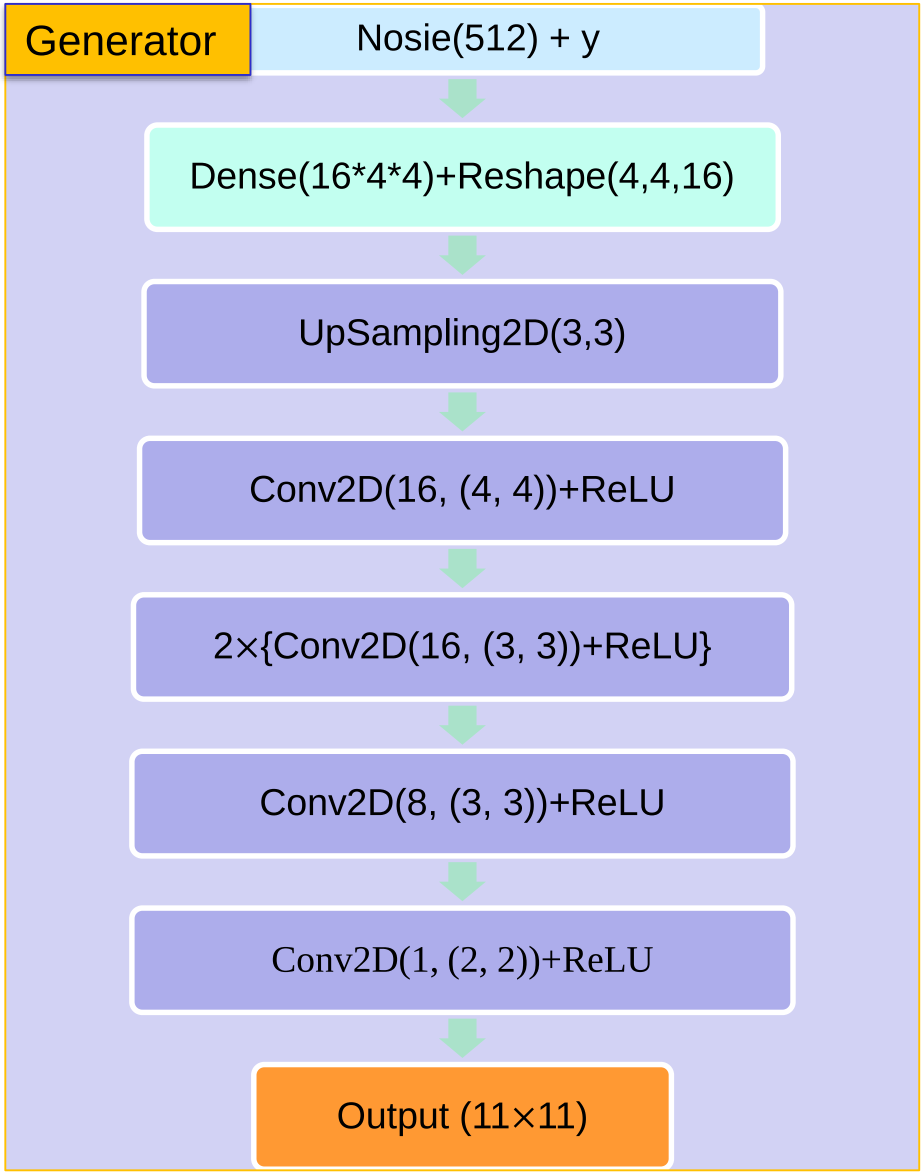
$$\max_{D} \mathbb{E}_{x\sim p(data)}\log(D(x)) + \mathbb{E}_{\hat{x}\sim p(fake)}\log(1-D(\hat{x}))$$



$$\min_{G} \mathbb{E}_{\hat{x}\sim p(fake)}\log(1-D(\hat{x})) + ||y_1-\widehat{y_1}||_1$$

- 其中 y (y_1+ y_2)包含了入射粒子动量和相对晶体位置的信息：
  - y_1：入射粒子动量大小，入射粒子方向和晶体朝向角度差别（$\Delta\phi^{\mathrm{Mom}}$， $\Delta\theta^{\mathrm{Mom}}$）
  - y_2: 入射粒子击中晶体位置和晶体前端面中心的偏差（ΔZ^{Pos}， $\Delta\phi^{\mathrm{Pos}}$）以及击中位置Z。

BESIII 生成对抗网络



Generator

Nosie(512) + y

Dense(16*4*4)+Reshape(4,4,16)

UpSampling2D(3,3)

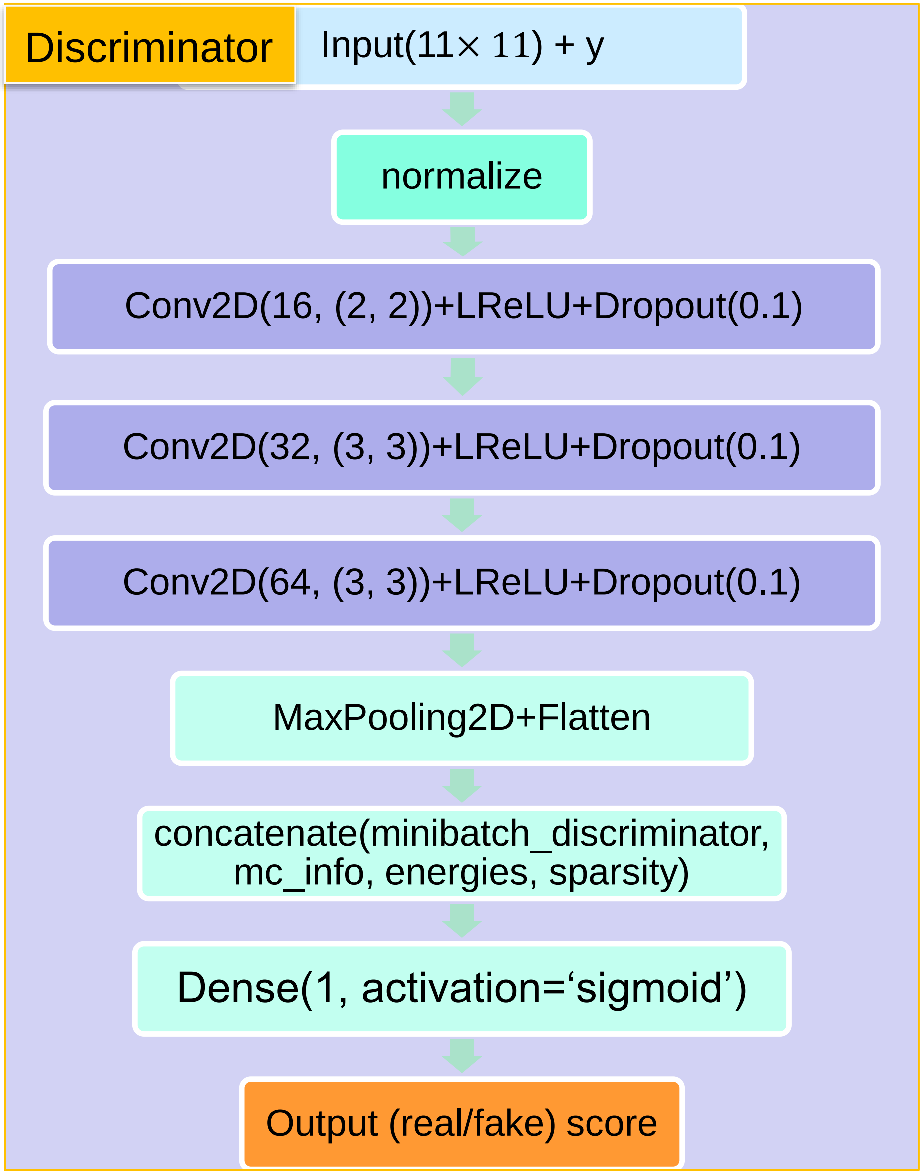
Conv2D(16, (4, 4))+ReLU

2×{Conv2D(16, (3, 3))+ReLU}

Conv2D(8, (3, 3))+ReLU

Conv2D(1, (2, 2))+ReLU

Output (11×11)



Discriminator

Input(11× 11) + y

normalize

Conv2D(16, (2, 2))+LReLU+Dropout(0.1)

Conv2D(32, (3, 3))+LReLU+Dropout(0.1)

Conv2D(64, (3, 3))+LReLU+Dropout(0.1)

MaxPooling2D+Flatten

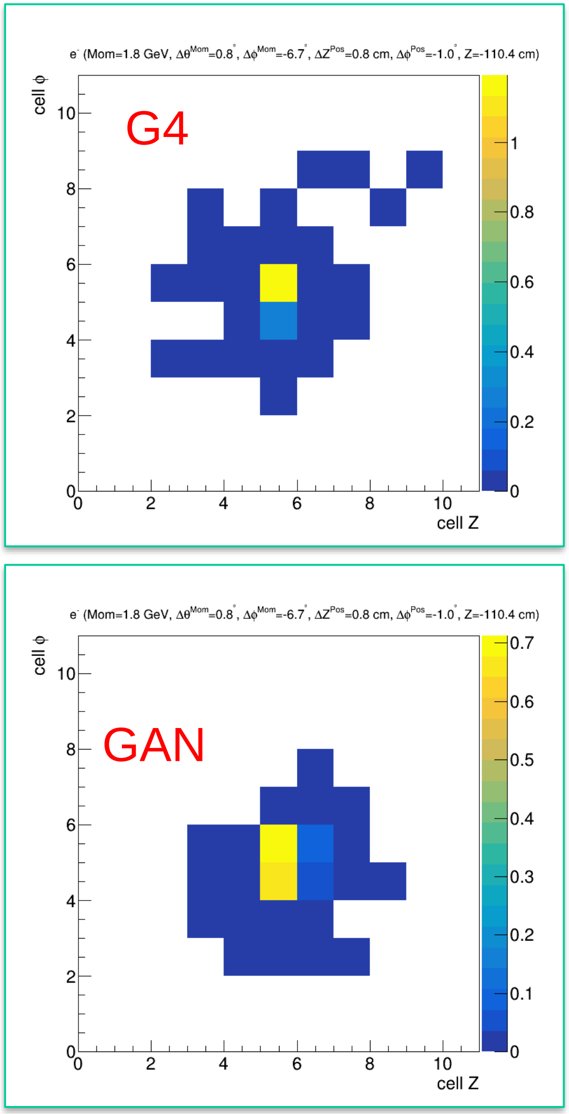
concatenate(minibatch_discriminator,
mc_info, energies, sparsity)

Dense(1, activation=‘sigmoid’)

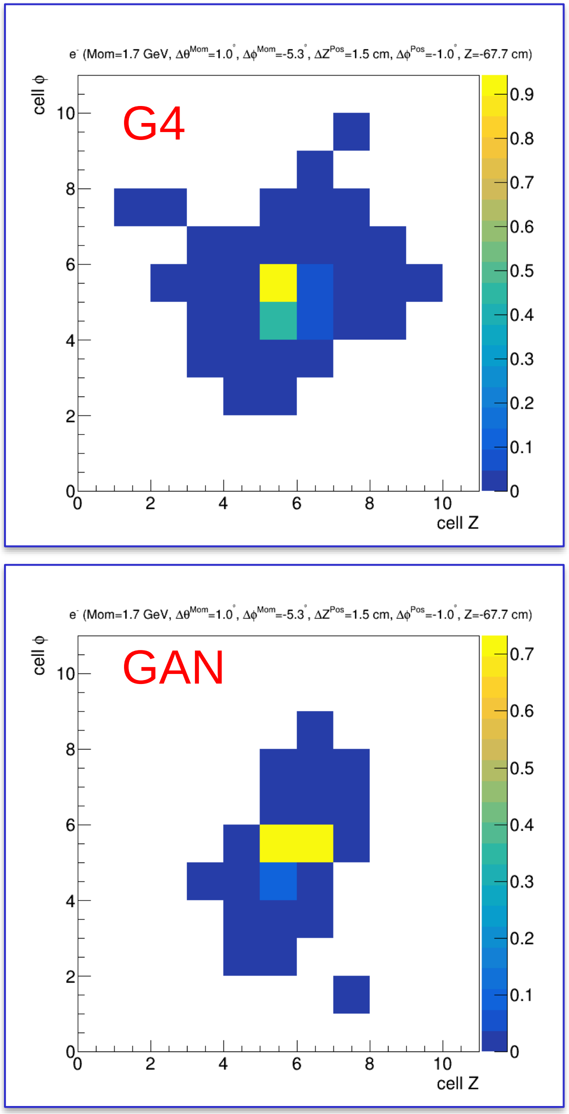
Output (real/fake) score

BESIII shower display（ e^- ）

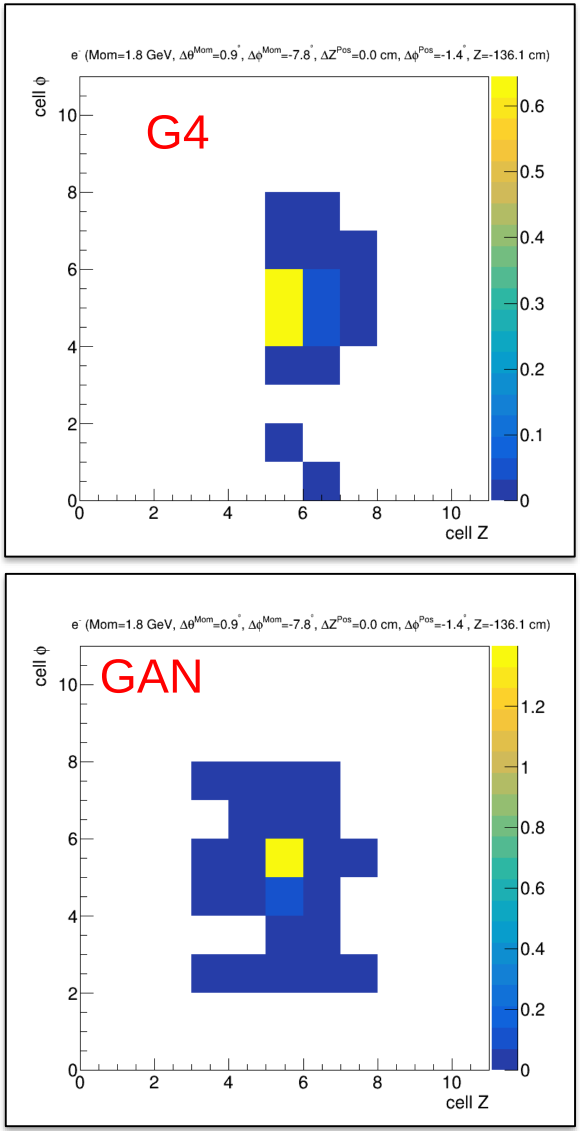
$e^-(\mathrm{Mom} = 1.8\ \mathrm{GeV}, \Delta\theta^{\mathrm{Mom}} = 0.8^\circ, \Delta\phi^{\mathrm{Mom}} = -6.7^\circ,$
$\Delta Z^{\mathrm{Pos}} = 0.8\ \mathrm{cm}, \Delta\phi^{\mathrm{Pos}} = -1.0^\circ, Z = -110.4\ \mathrm{cm})$



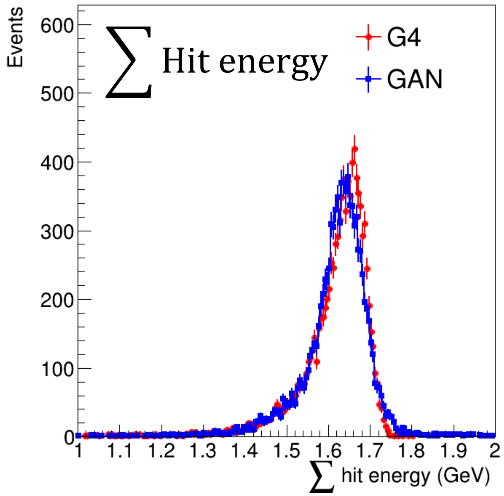
$e^-(\mathrm{Mom} = 1.7\ \mathrm{GeV}, \Delta\theta^{\mathrm{Mom}} = 1.0^\circ, \Delta\phi^{\mathrm{Mom}} = -5.3^\circ,$
$\Delta Z^{\mathrm{Pos}} = 1.5\ \mathrm{cm}, \Delta\phi^{\mathrm{Pos}} = -1.0^\circ, Z = -67.7\ \mathrm{cm})$

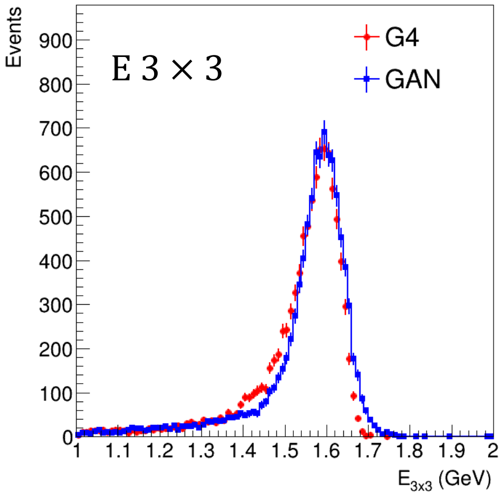


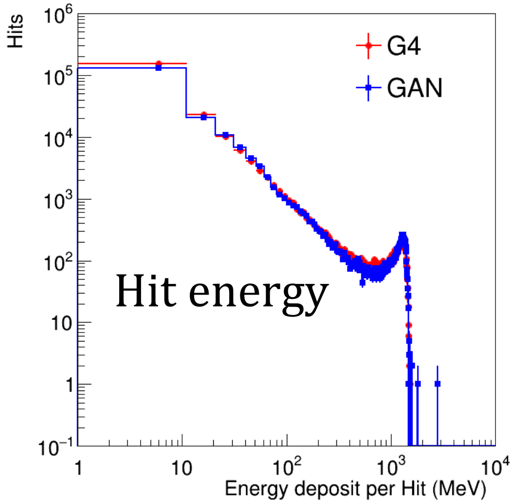
$e^-(\mathrm{Mom} = 1.8\ \mathrm{GeV}, \Delta\theta^{\mathrm{Mom}} = 0.9^\circ, \Delta\phi^{\mathrm{Mom}} = -7.8^\circ,$
$\Delta Z^{\mathrm{Pos}} = 0.\ \mathrm{cm}, \Delta\phi^{\mathrm{Pos}} = -1.4^\circ, Z = -136.1\ \mathrm{cm})$

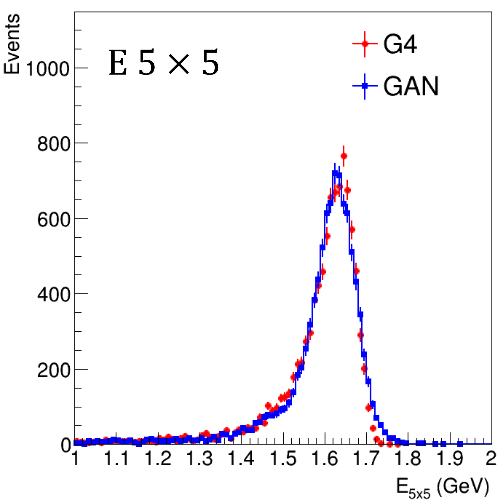


BESIII 模拟性能（ e^- ）

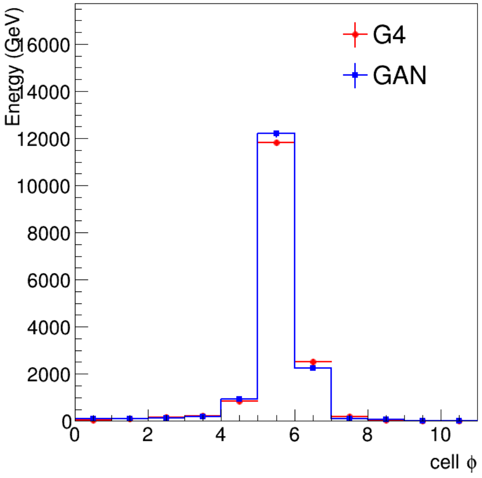




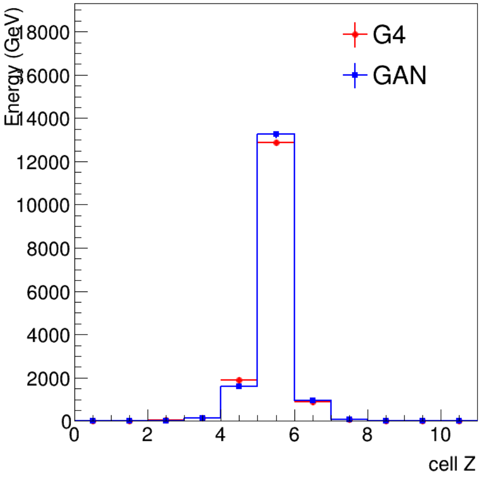




BESIII 模拟性能（ e^- ）

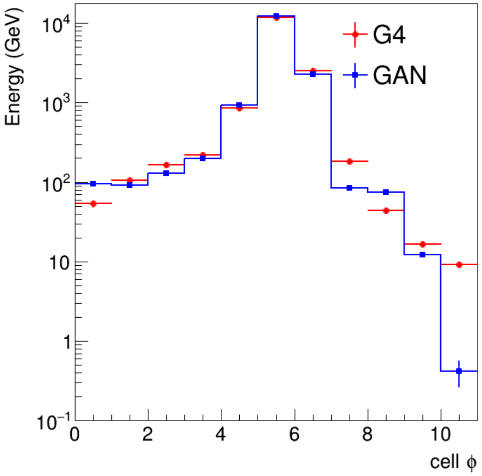


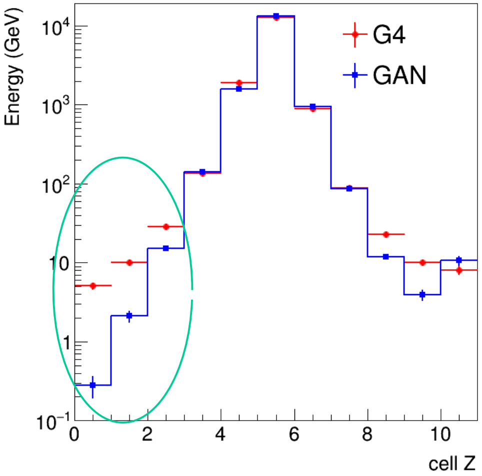
能量沉积 vs ϕ



能量沉积 vs Z

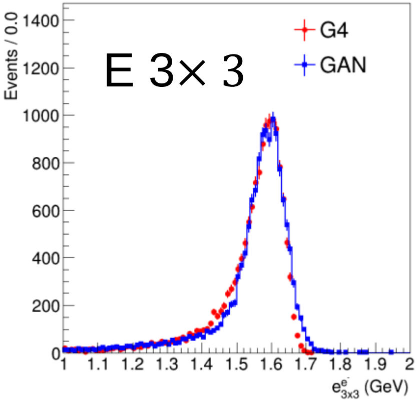
Log scale

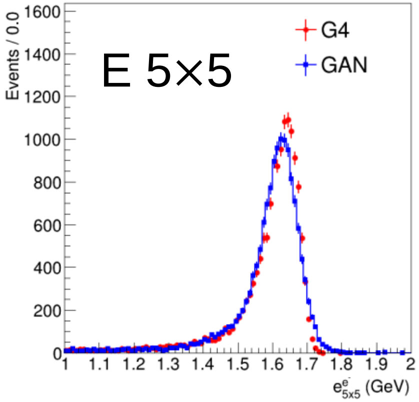


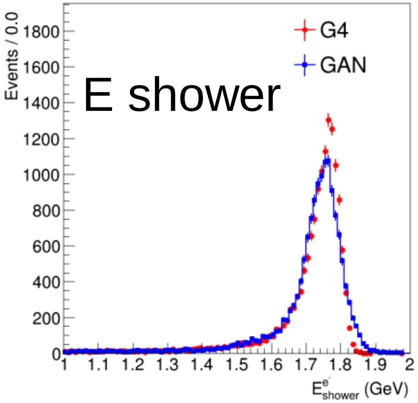


BESIII 模拟性能（ e^- ）

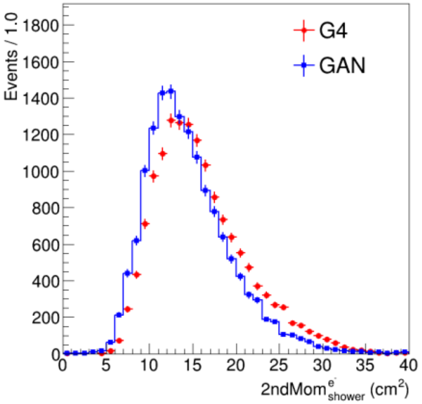
- 将生成网络加到BOSS软件中，用于模拟，然后检查粒子的重建信息。

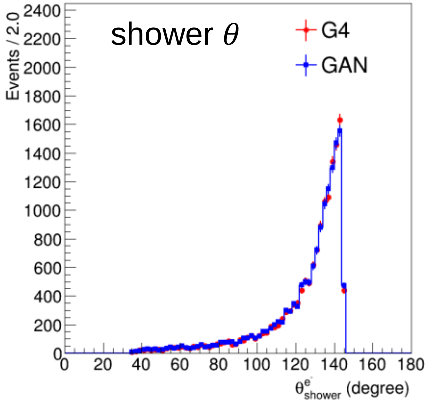


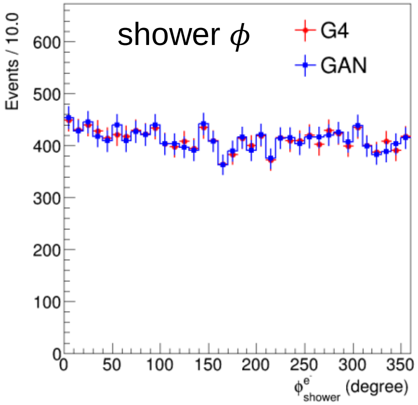




Shower 2nd mom

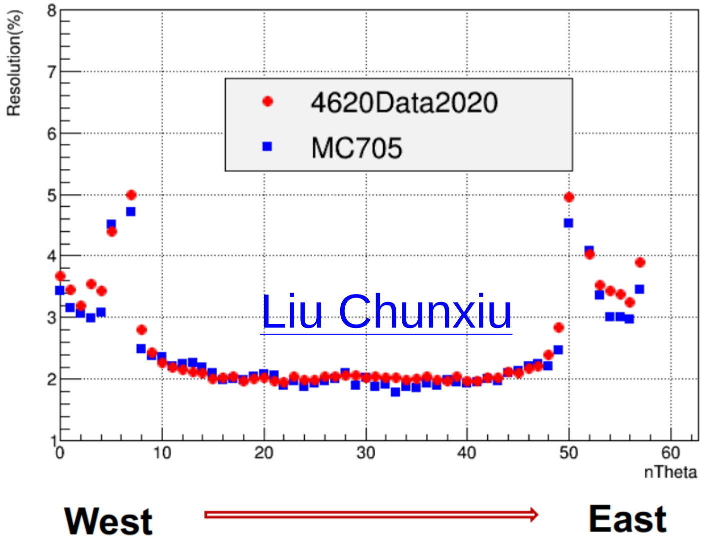




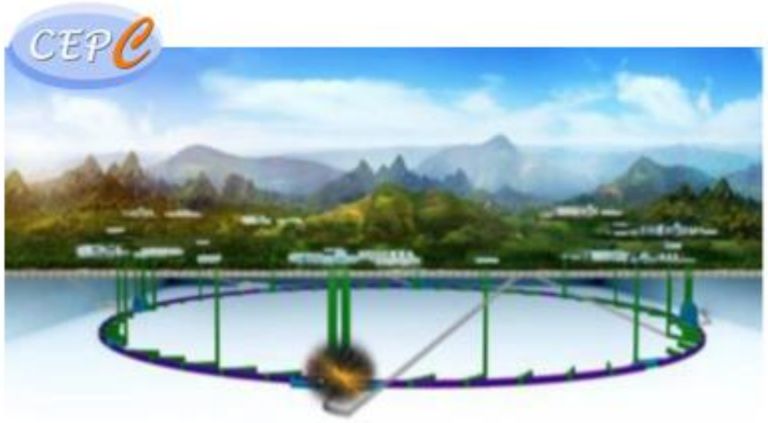


BESIII 生成对抗网络小结

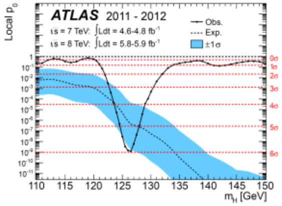
- GAN模拟和Geant4模拟在大体上能符合。
- 还需要继续改善GAN的性能，使得其最终能用于模拟。
- 对于BESIII实验，由于其桶部和端盖之间复杂的结构，使得在靠近此区间的数据和模拟符合不够好（如能量分辨率）。因此可以考虑用GAN学习数据在这个区间的能量分布，从而改善数据和MC的符合情况。



Energy resolution vs. ThetaID



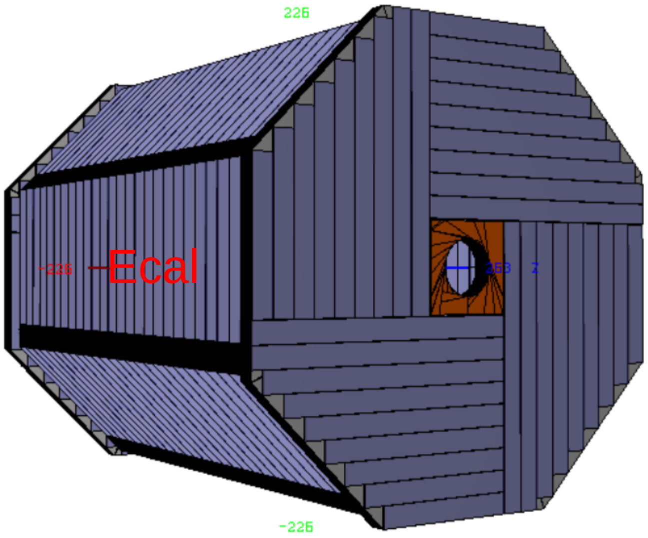
CEPC 实验的量能器模拟

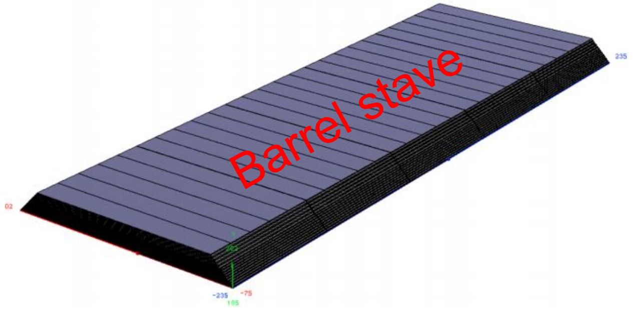


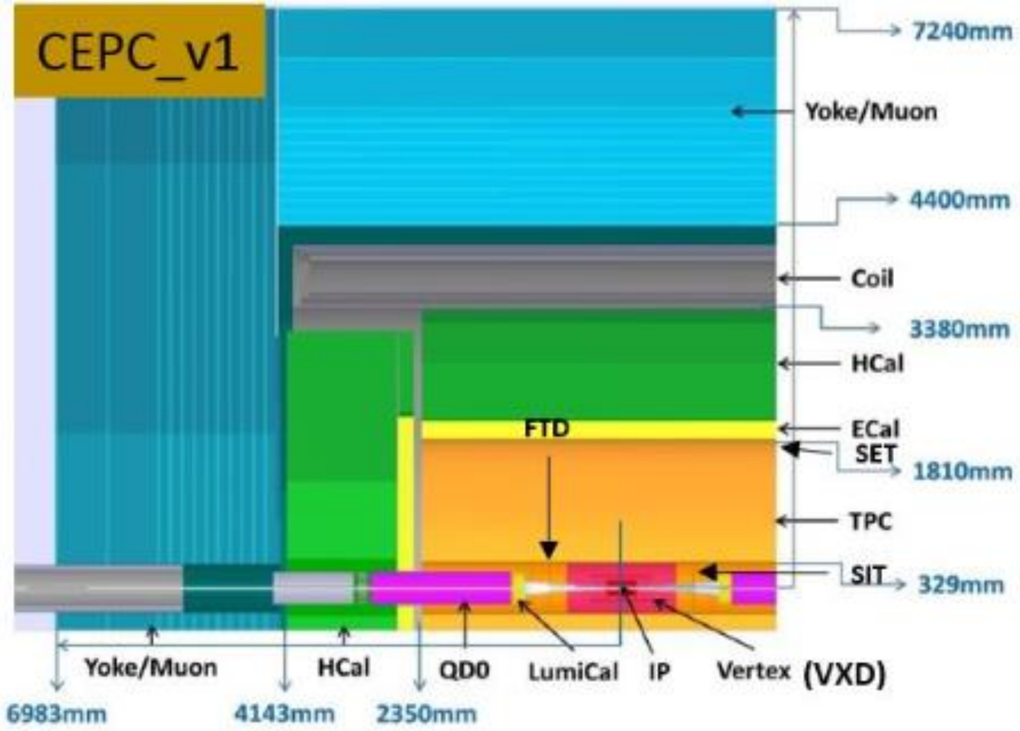
CEPC 实验

- 2012年ATLAS和CMS发现了标准模型预言的最后一个粒子，即解释质量起源的Higgs粒子。
- 通过对Higgs粒子的研究我们能深入理解标准模型，同时为寻找超标准模型的新物理提供重要线索。
- 环形正负电子对撞机（CEPC）：
  - e^+e^-质心对撞能量 250 GeV（ZH），91.2 GeV(Z)
  - 物理目标：精确测量Higgs，Z 玻色子的属性，研究味物理，寻找新物理等。
- CEPC实验将产生海量的实验数据（如~10^{12} Z 玻色子），为了降低统计误差，需要产生与之相匹配（或更多）的海量模拟数据。
- 产生如此大量的模拟数据将消耗巨大计算资源。
- 为了节约有限的计算资源，开发快速模拟方法是非常重要的工作。
- 在模拟过程中，由于大部分的计算资源用于粒子在量能器上簇射的模拟，因此加快量能器的模拟是解决该问题的关键。

CEPC 探测器



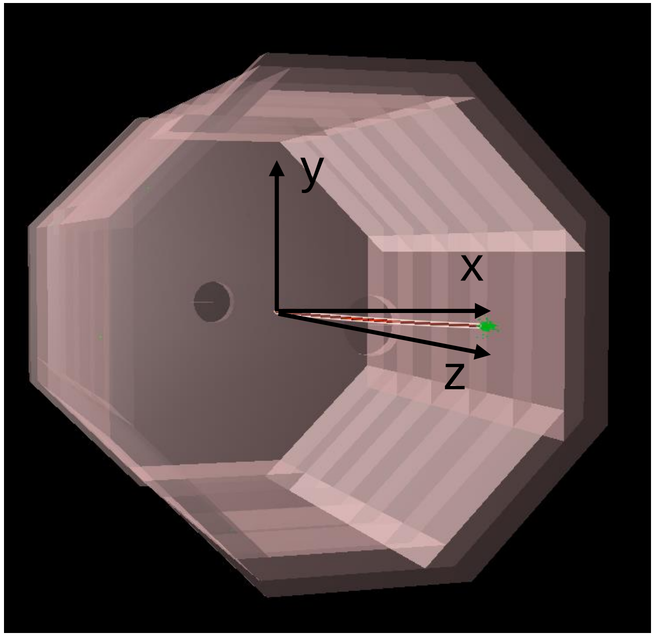


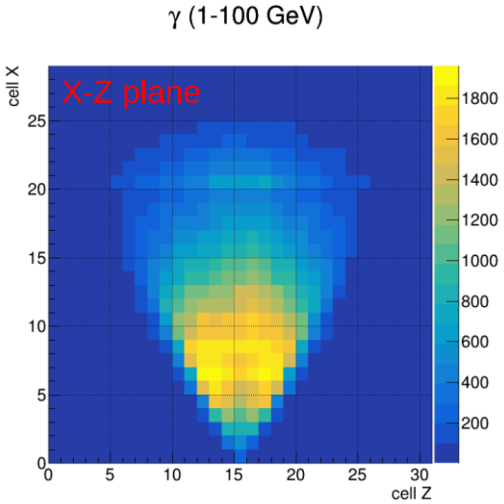


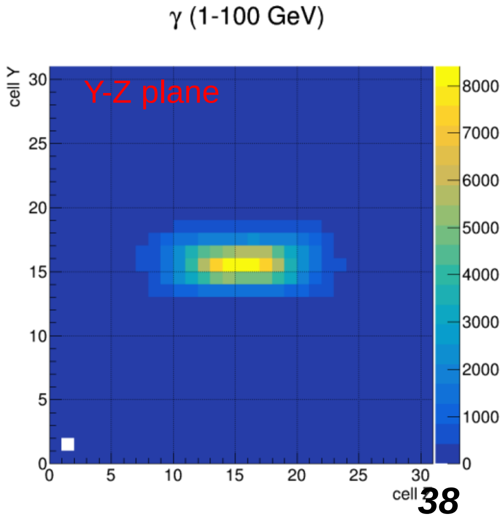
Ecal in Barrel	Cell size	Layer	Particle energy
BESIII	5 cm × 5 cm	1	$\mathcal{O}(1)$ GeV
CEPC	1 cm × 1 cm	29	$\mathcal{O}(10)$ GeV

CEPC 训练数据

- 使用单光子作为训练数据，模拟其在电磁量能器上的簇射
- 能量：1-100GeV均匀分布
- 角度： θ 在50-140度均匀分布（桶部）， ϕ 在±15度范围均匀分布。排除不同模块交界区域。



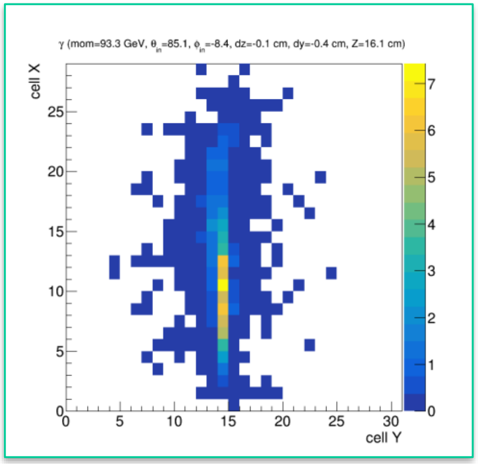


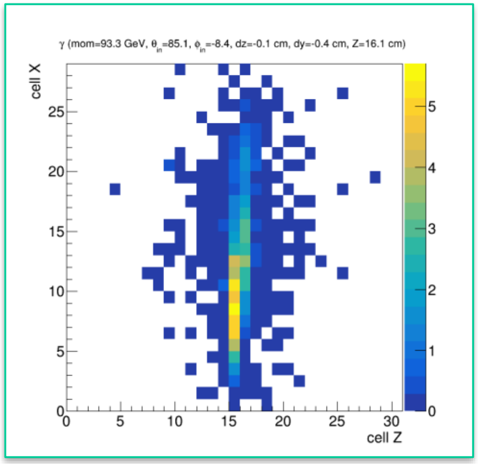


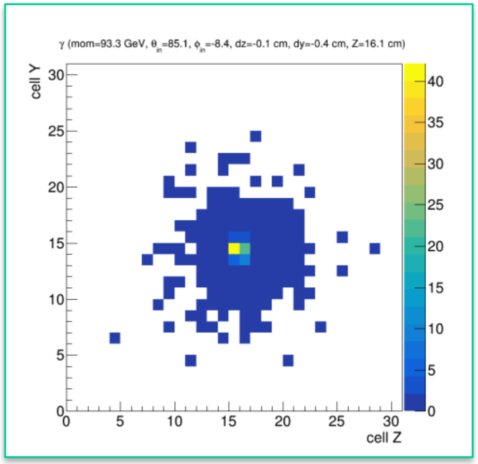
CEPC shower display

采用和BESIII实验类似的GAN网络

$\gamma\ (\mathrm{Mom} = 93.3\ \mathrm{GeV}, \theta_{in} = 85.1^{\circ}, \phi_{in} = -8.4^{\circ}, \quad \Delta Z^{Pos} = -0.1\ \mathrm{cm}, \Delta Y^{Pos} = -0.4\ \mathrm{cm}, Z = 16.1\ \mathrm{cm})$





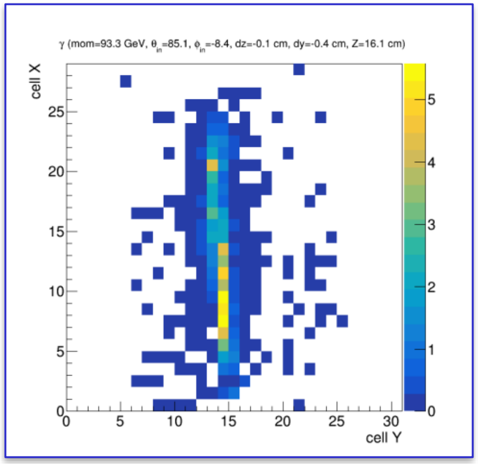


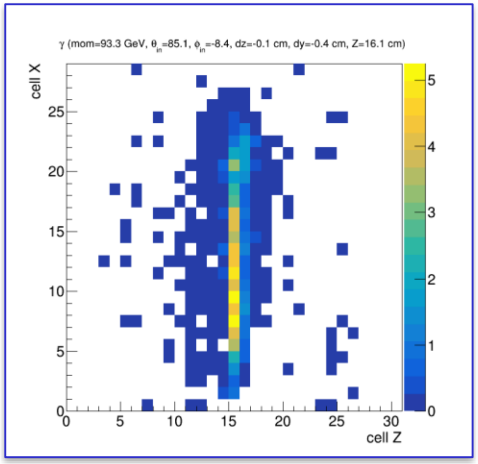
X-Y plane

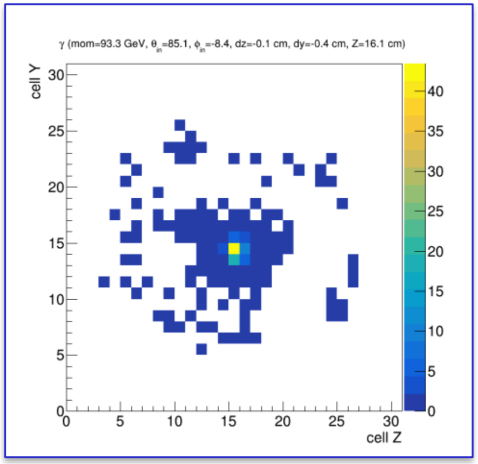
X-Z plane

Y-Z plane

Geant4

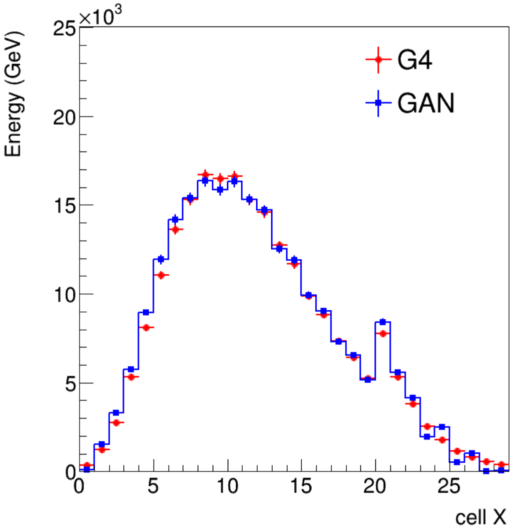




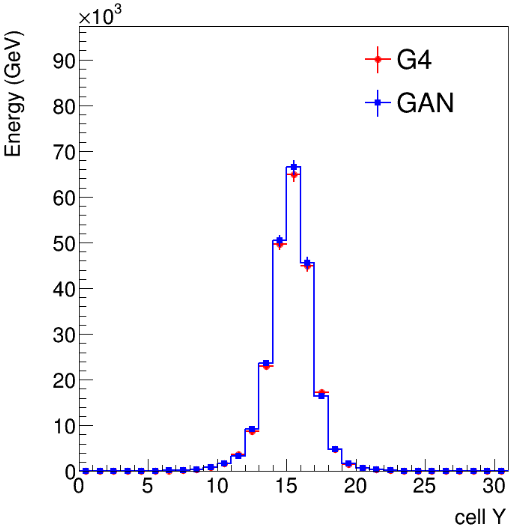


GAN

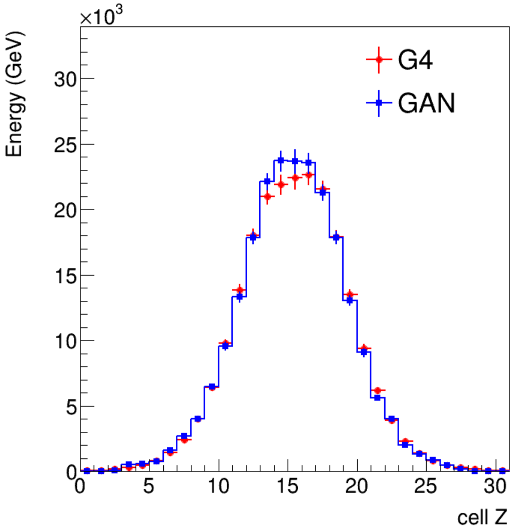
模拟性能



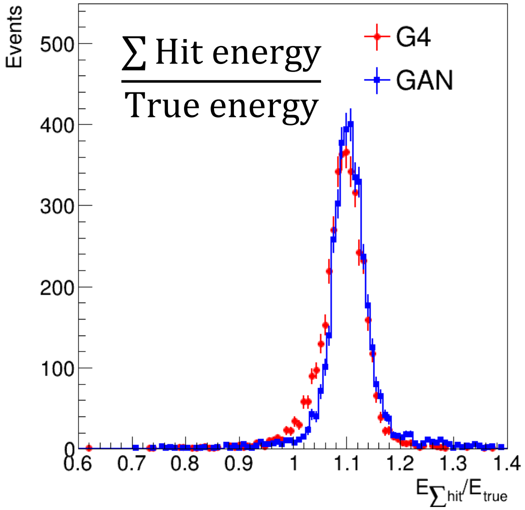
能量沉积 vs X (层数)

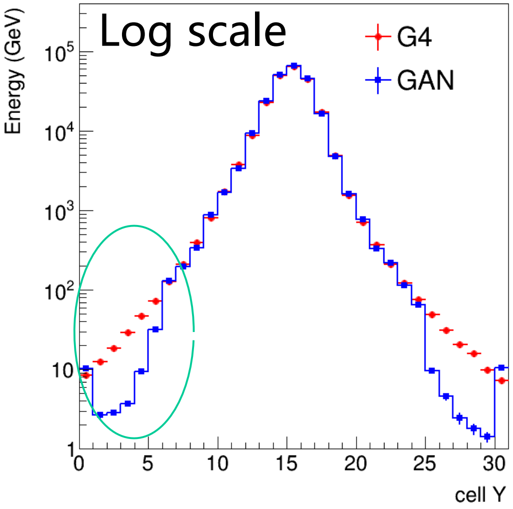


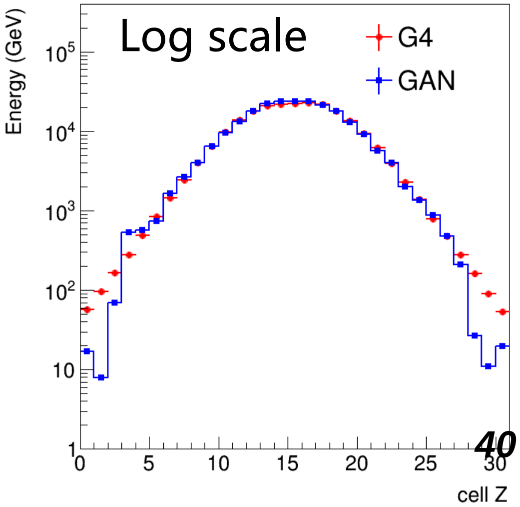
能量沉积 vs Y 方向



能量沉积 vs Z 方向

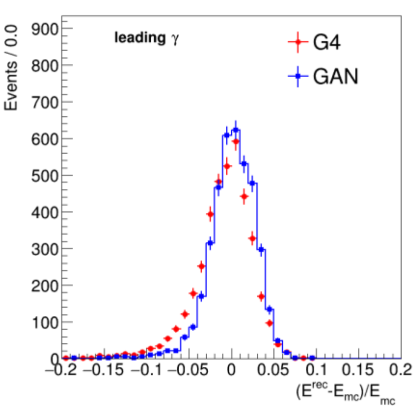




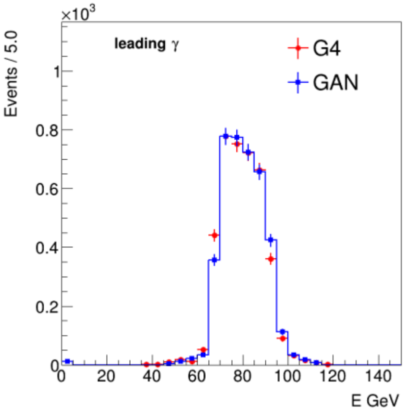


模拟性能

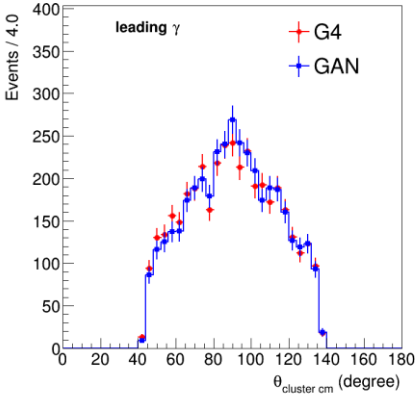
将生成网络用于$e^+e^- \rightarrow Z(\nu\nu)H(\gamma\gamma)$过程的模拟，检查光子后光子性能。



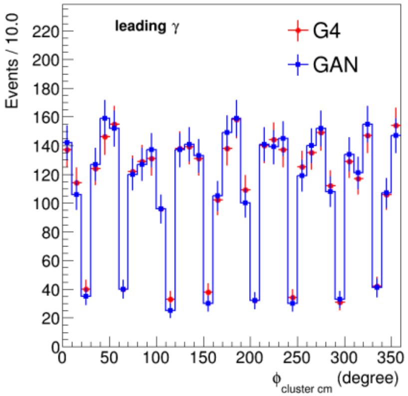
$$\frac{E_{rec} - E_{mc}}{E_{mc}}$$



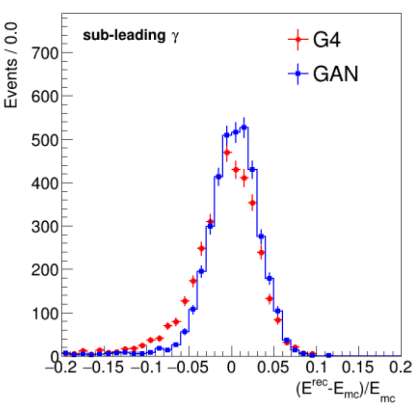
E_{rec}

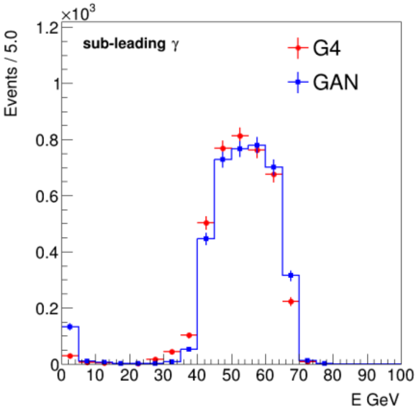


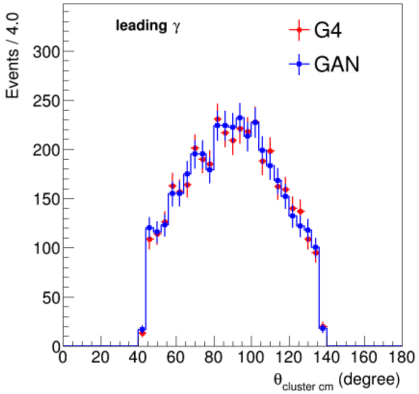
θ

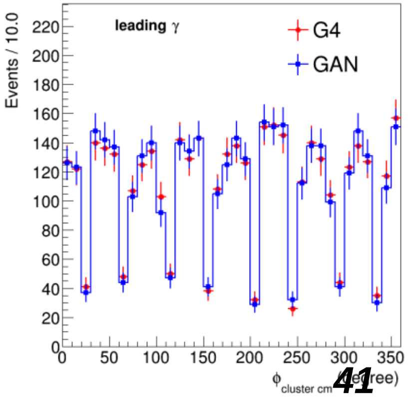


ϕ









总结

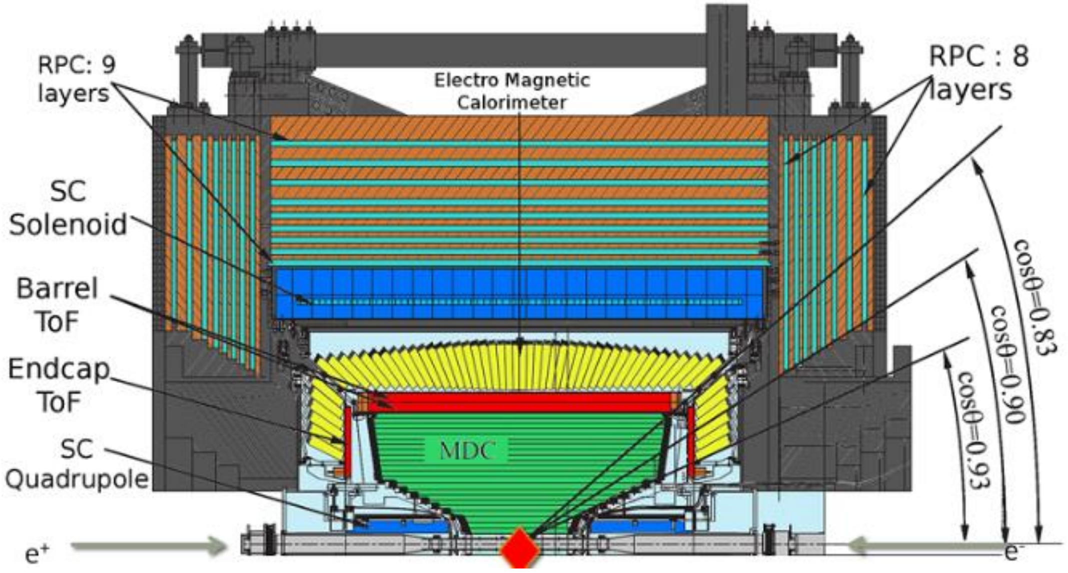
- 介绍了近些年比较热门的生成网络 GAN 和 VAE，其中主要介绍了 GAN 在 BESIII 和 CEPC 实验量能器模拟中的应用。目前 GAN 的结果能与 Geant4 模拟结果大致相符，细节部分存在改善的空间。
- 此外，介绍了全链接生成网络在 JUNO 实验的缪子快速模拟以及 BESIII 实验 dE/dx 模拟中的应用。通过采用从数据中学习 dE/dx 的分布用于模拟，可以改善数据和模拟的一致性，从而减小系统误差。
- 在未来，生成网络在高能物理的模拟中发挥越来越重要的作用!

谢谢！

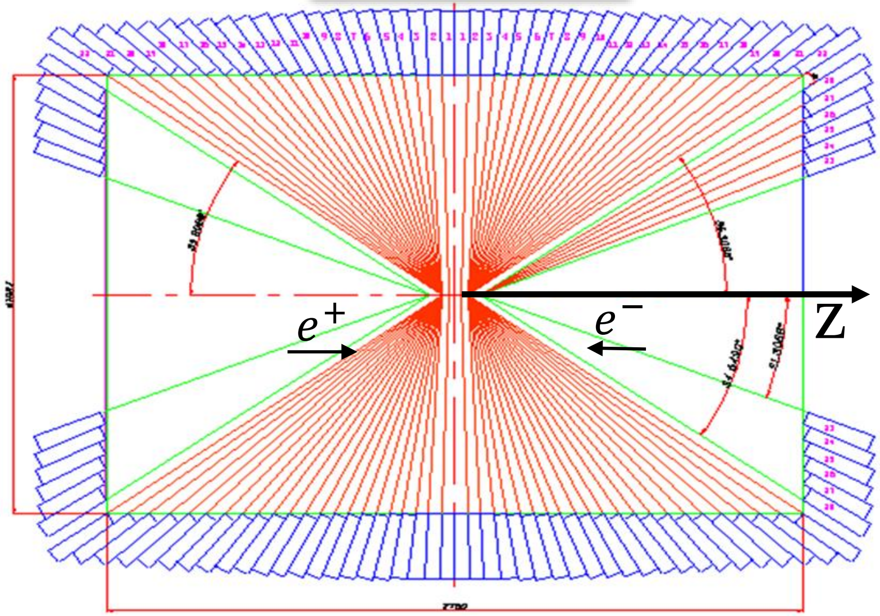
Back up

BESIII experiment

- ❑ The BESIII detector is designed to study physics in the τ-charm energy region utilizing the high luminosity BEPCII double ring $e^{+}e^{-}$collider which has peak luminosity $10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ at center-of-mass energy 3.78 GeV.



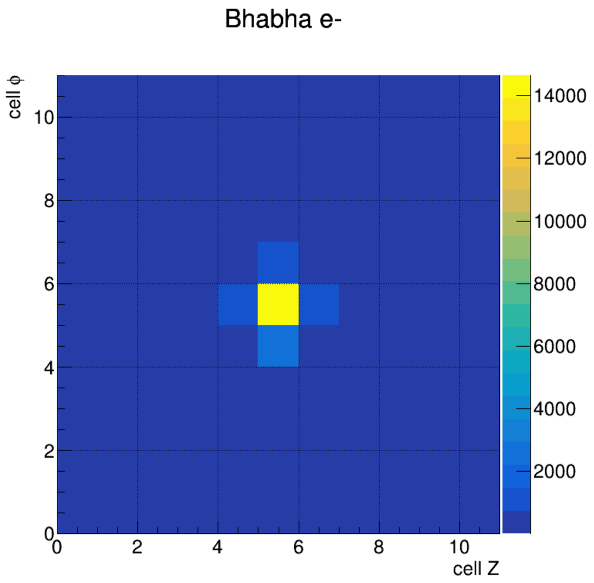
Zoomed EMC

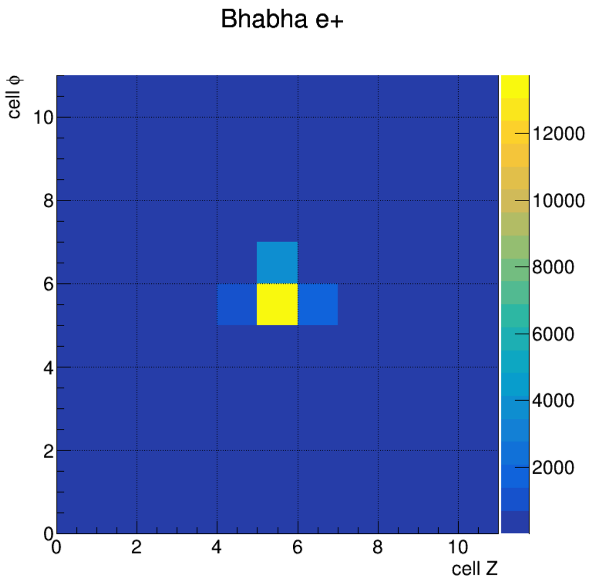


- ➢ 44 rings of crystal in barrel and 120 crystals in each ring. The front size of each crystals is 5×5 cm^2, the crystal length is 28 cm .
- ➢ 6 rings of crystal in each endcap.

Dataset

- /besfs/groups/cal/emc/liucx/BhabhaCalib/mcdata/bb703/bb1776_703_2017*
- Select $e^{\pm}$:
  - EvtRecTrack. isMdcTrackValid && EvtRecTrack. isExtTrackValid .
  - EvtRecTrack. isEmcShowerValid && RecEmcShower .energy > 40 MeV .
  - RecEmcShower. getCluster != 0 .
  - RecMdcTrack. Charge > 0 for e^{+} and < 0 for e^{-} .
  - Select one e^{+} and one e^{-} with highest momentum according to RecMdcTrack. P .
  - Finally the $|\cos\theta| < 0.83$ is asked for selected $e^{\pm}$.
- ~ 450000 training events.
- The position of MDC track extends to EMC is chose as the center. Hit energy in 11×11 calorimeter cells are considered.



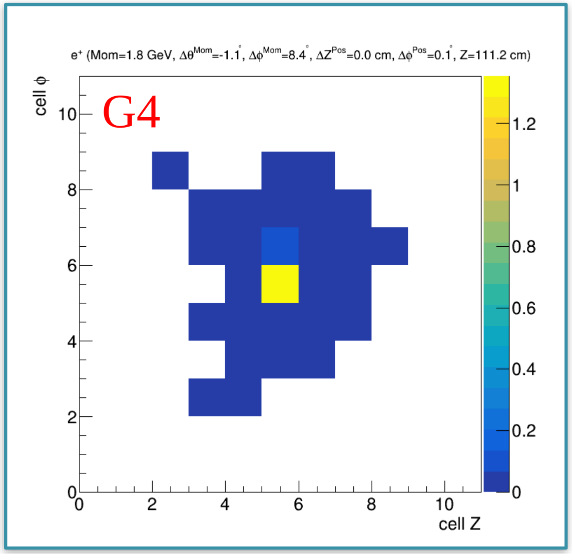


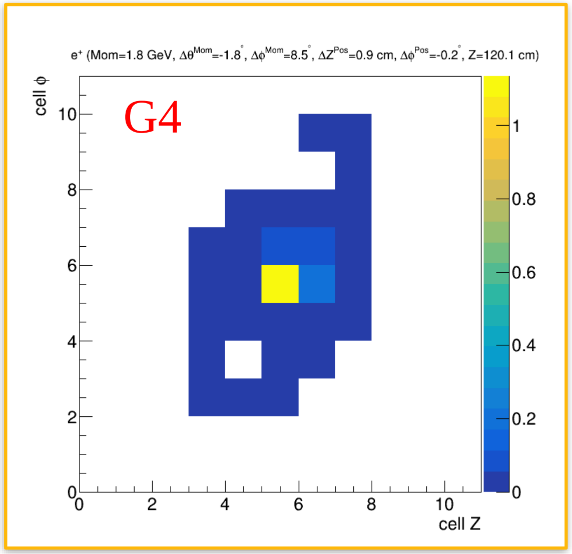
Event display (e^+)

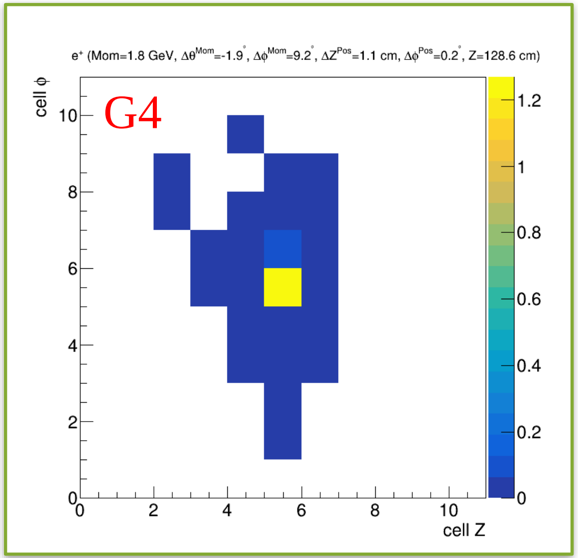
e^+(Mom = 1.8 GeV, $\Delta\theta^{Mom} = -1.1^\circ$, $\Delta\phi^{Mom} = 8.4^\circ$, $\Delta Z^{Pos} = 0.0$ cm, $\Delta\phi^{Pos} = 0.1^\circ$, Z = 111.2 cm)

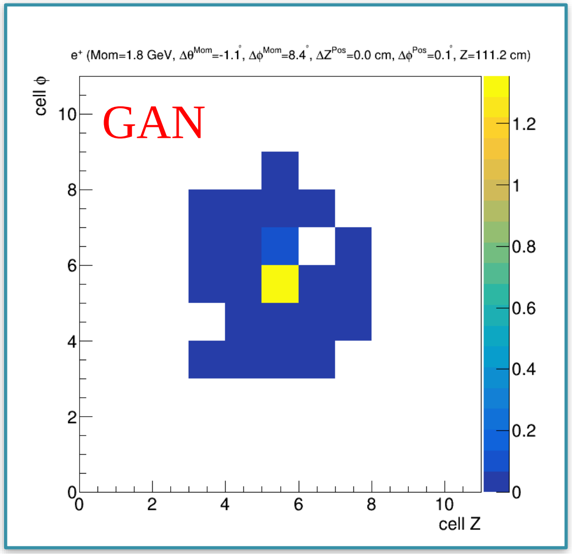
e^+(Mom = 1.8 GeV, $\Delta\theta^{Mom} = -1.8^\circ$, $\Delta\phi^{Mom} = 8.5^\circ$, $\Delta Z^{Pos} = 0.9$ cm, $\Delta\phi^{Pos} = -0.2^\circ$, Z = 120.1 cm)

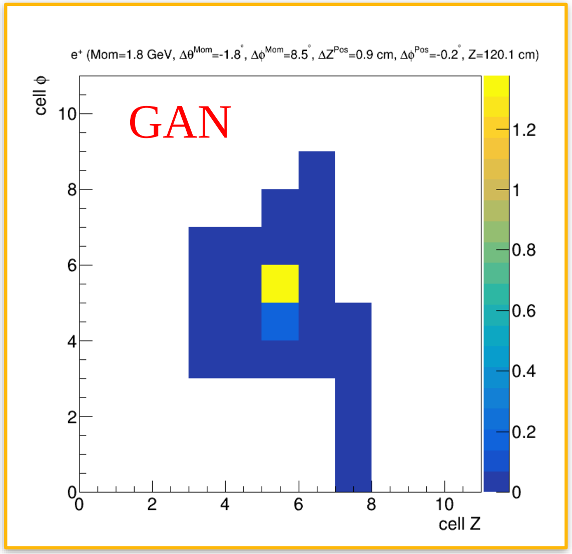
e^+(Mom = 1.8 GeV, $\Delta\theta^{Mom} = -1.9^\circ$, $\Delta\phi^{Mom} = 9.2^\circ$, $\Delta Z^{Pos} = 1.1$ cm, $\Delta\phi^{Pos} = 0.2^\circ$, Z = 128.6 cm)

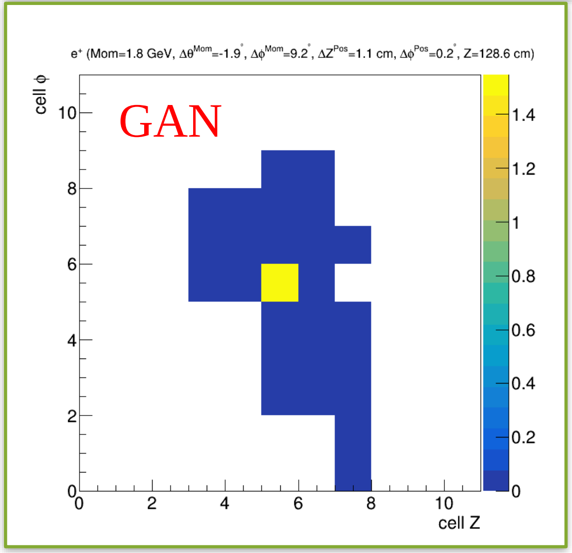






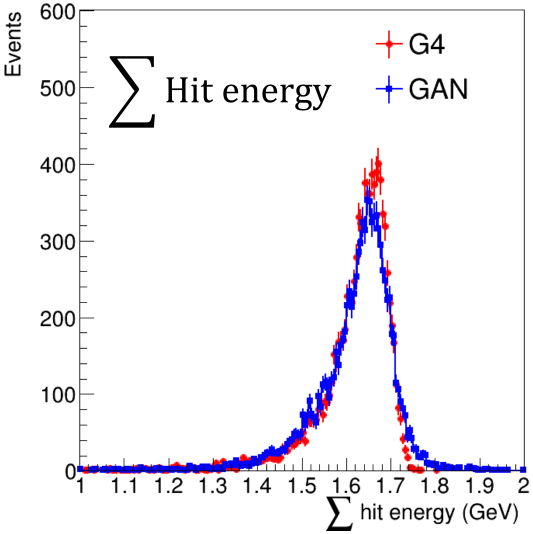


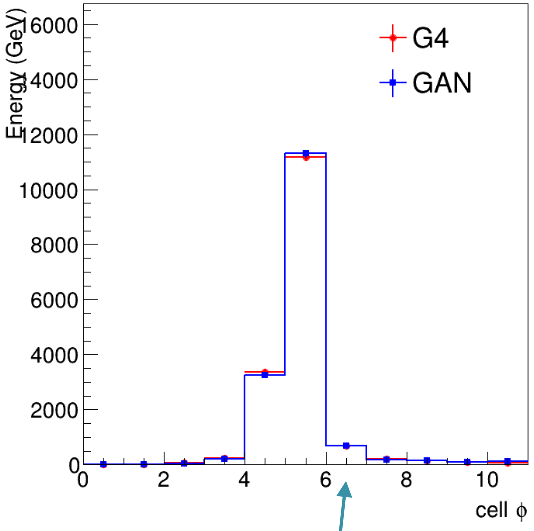


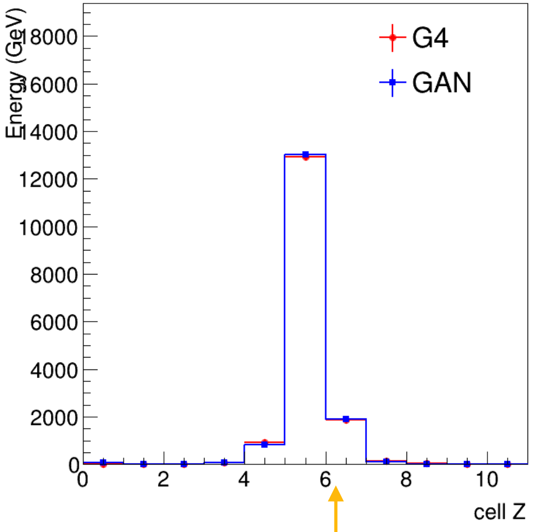


See distributions in following.

Some distributions (e^+)



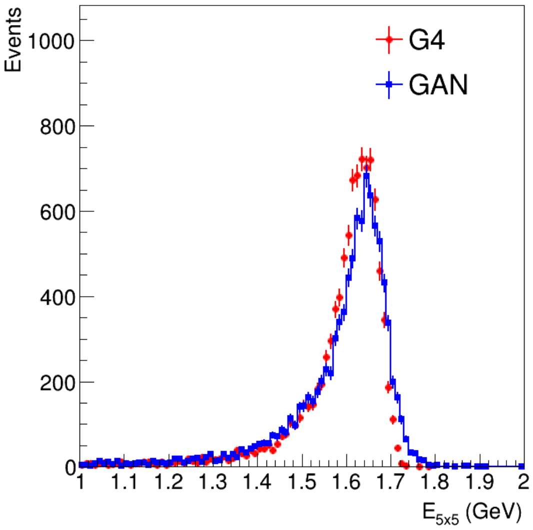


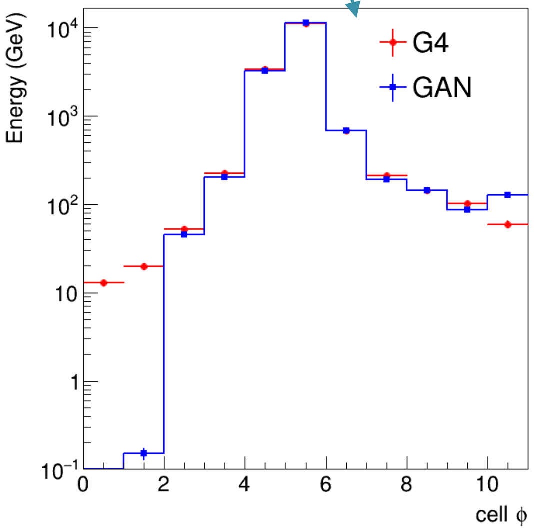


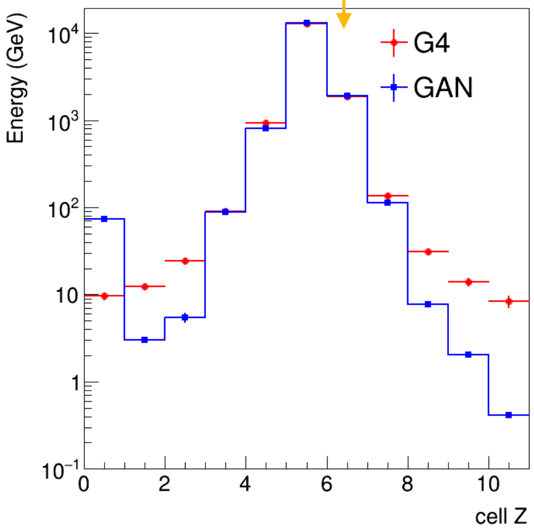
Energy of calorimeter cells

Energy deposited in ϕ direction

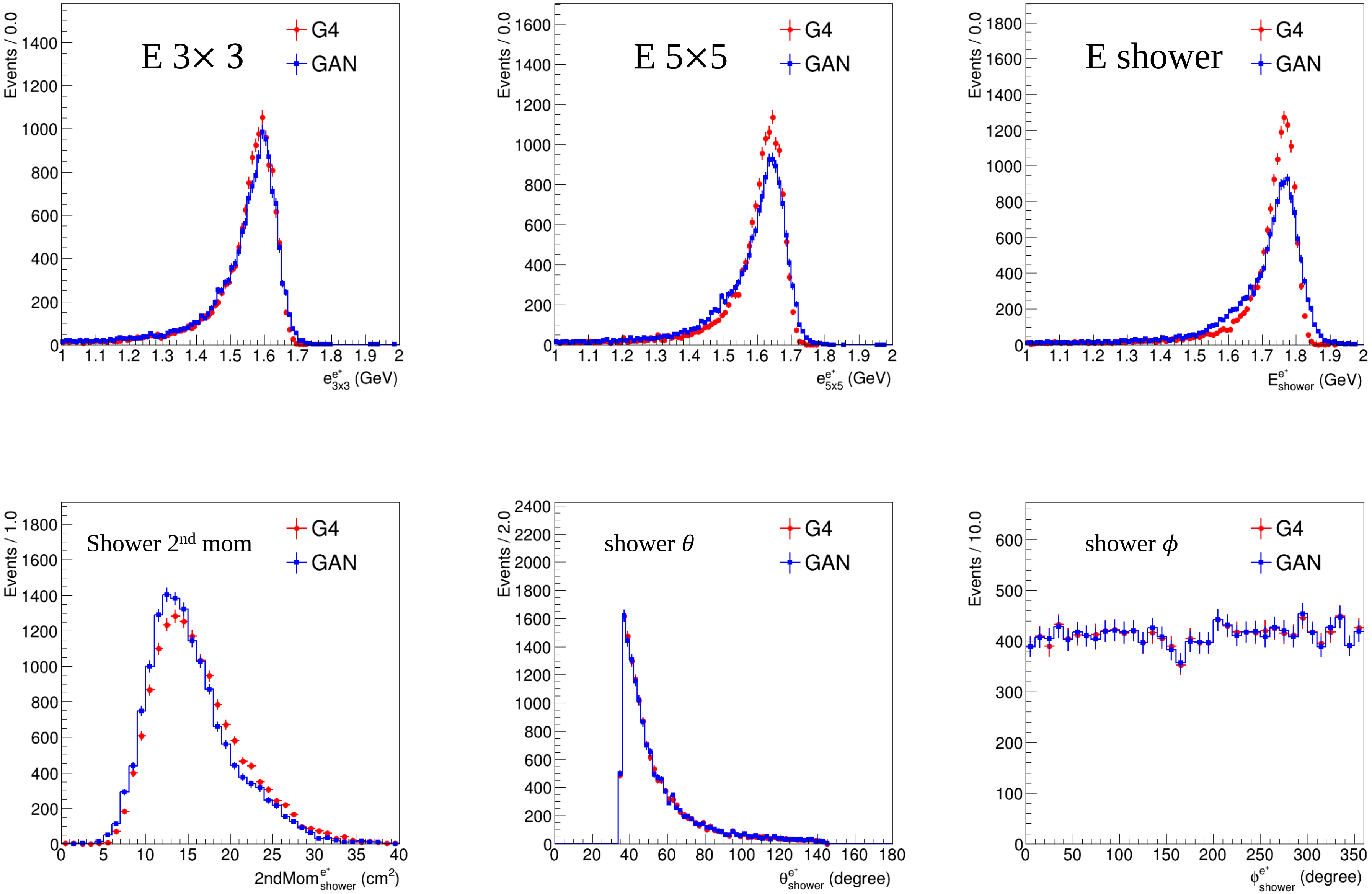
Energy deposited in Z direction







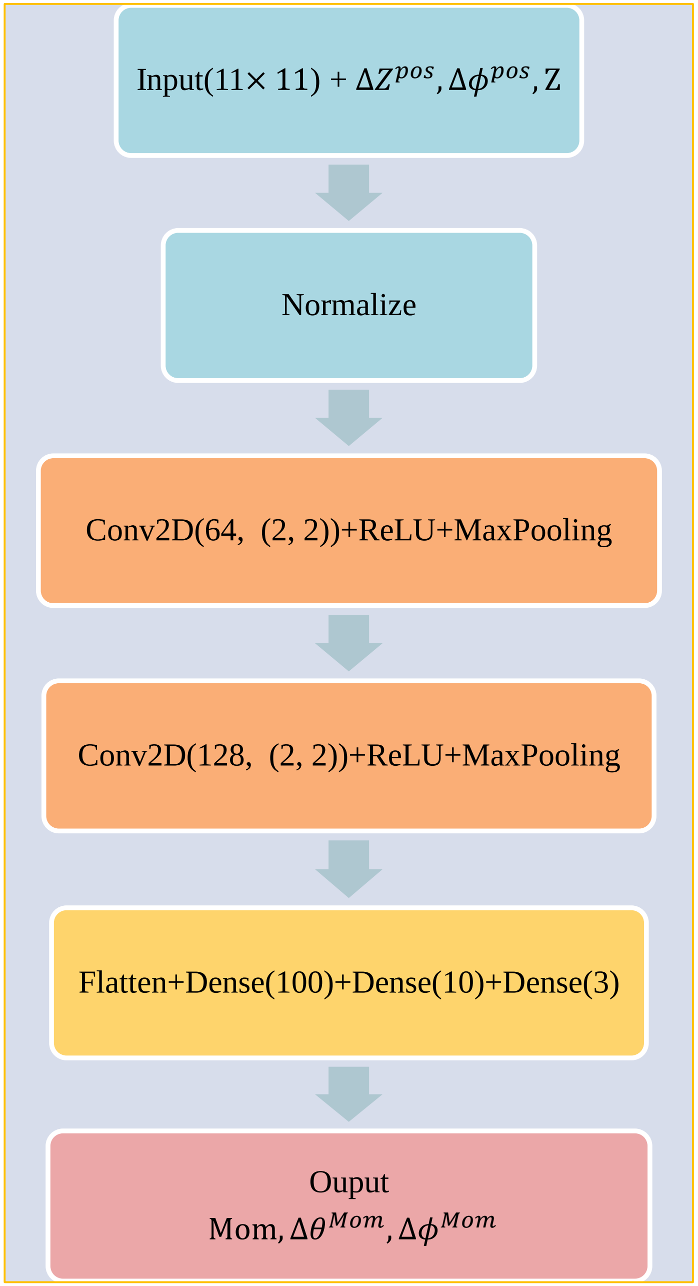
Some variables for reconstruced particle (e^+)



Regressor

- $\Delta\phi^{Mom}$: the ϕ difference between the momentum of incoming particle and the direction of the crystal.
- $\Delta\theta^{Mom}$: the θ difference between the momentum of incoming particle and the direction of the crystal.
- ΔZ^{Pos}: the Z difference between the hit point of incoming particle and the z of front center of the crystal.
- $\Delta\phi^{Pos}$: the ϕ difference between the hit point of incoming particle and the ϕ of front center of the crystal.
- Momentum: the momentum of the particle ($P_{MDC} - E_{TOF}$).
- Z

❖ Due to the e^- (e^+) is mostly at negative (positive) Z region, the e^- (e^+) at positive (negative) is not used.



Input(11× 11) + $\Delta Z^{pos}, \Delta\phi^{pos}$, Z

Normalize

Conv2D(64, (2, 2))+ReLU+MaxPooling

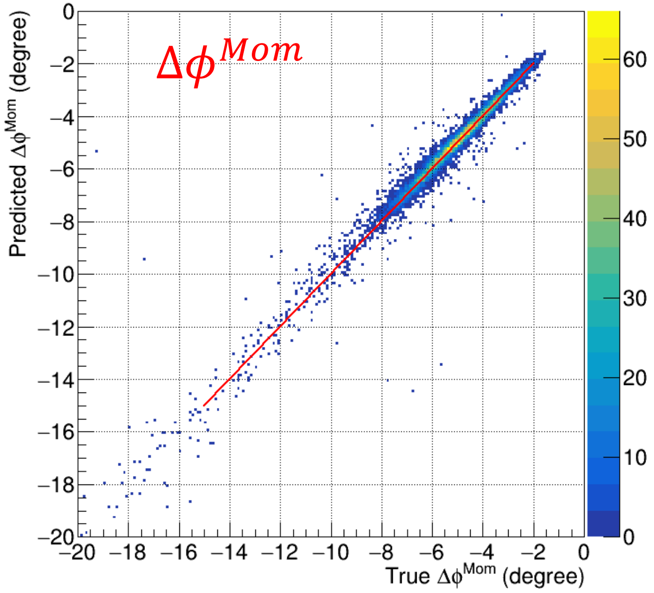
Conv2D(128, (2, 2))+ReLU+MaxPooling

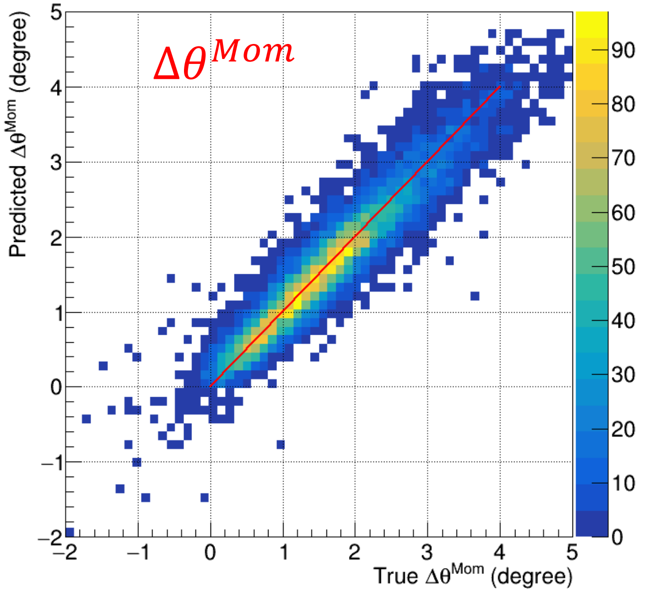
Flatten+Dense(100)+Dense(10)+Dense(3)

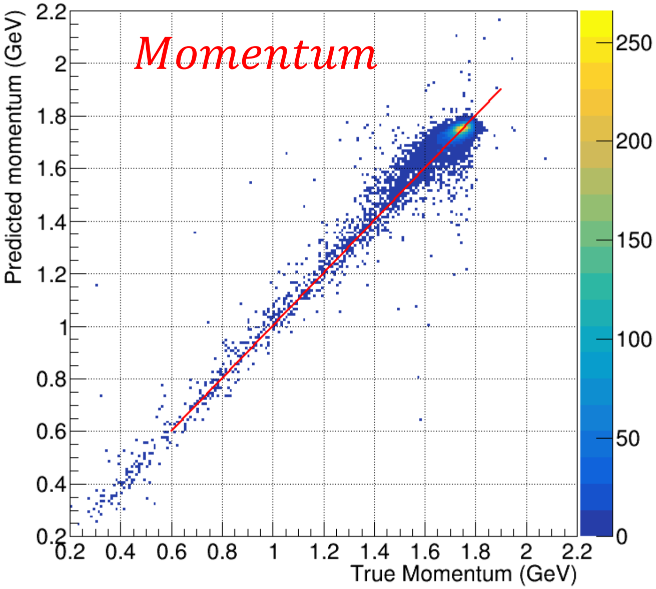
Ouput

Mom, $\Delta\theta^{Mom}, \Delta\phi^{Mom}$

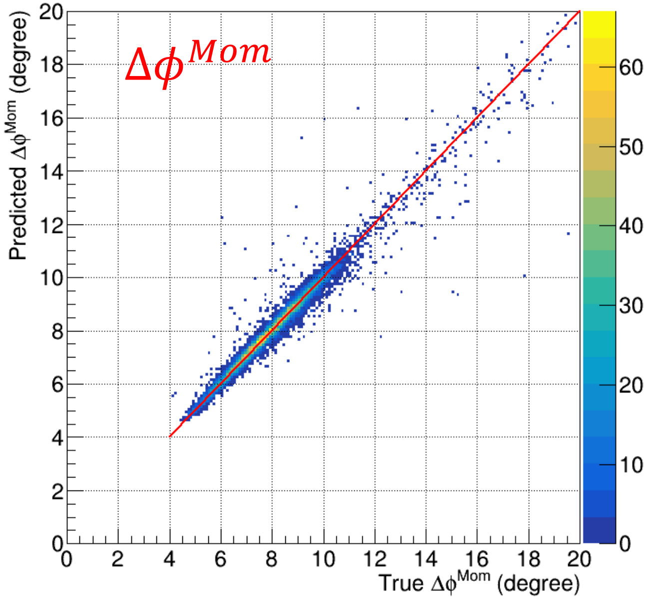
Regressor performance (e^-)

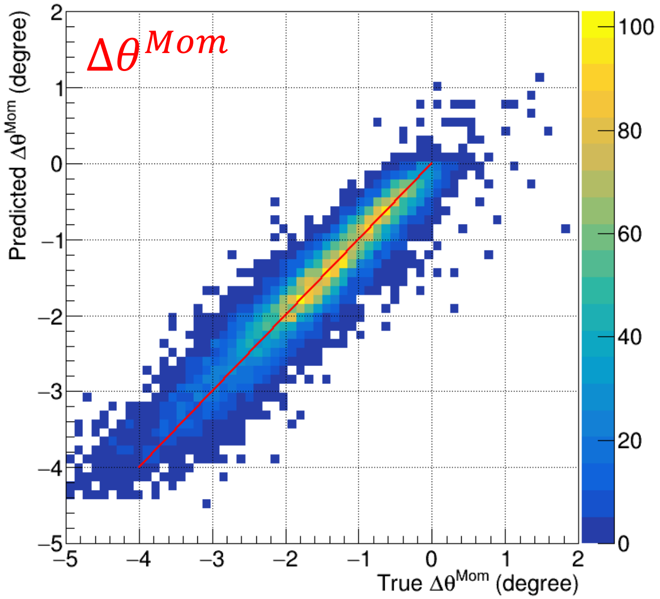


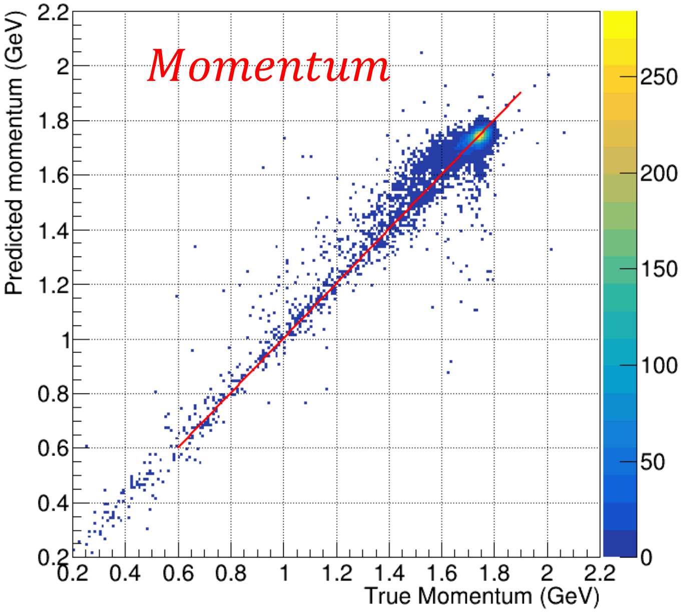




Regressor performance (e^+)







Generator and discriminator architecture (CEPC)

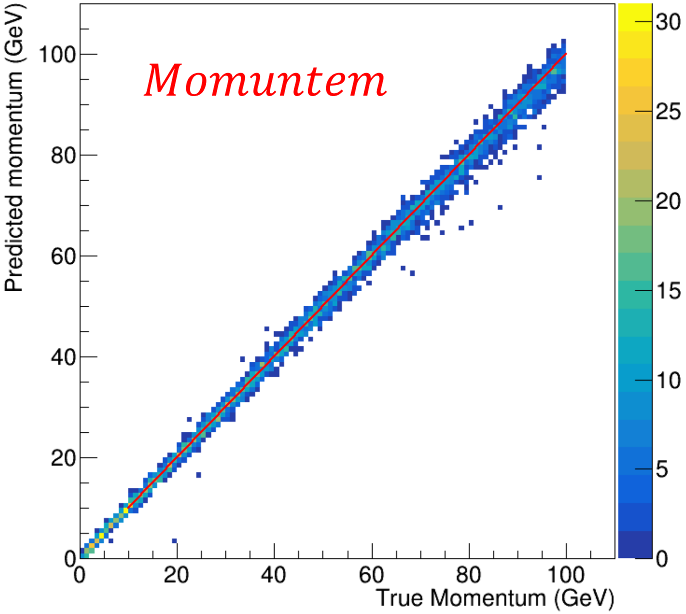
Generator

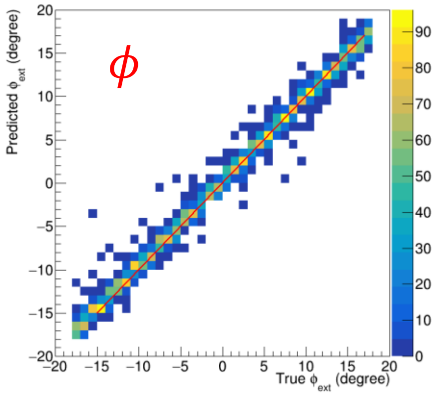
- Nosie(512)+(Mom, θ, ϕ, Z)
- Dense(6*6*6*8)+Reshape(6,6,6,8)
- UpSampling3D(5,5,5)+ReLU
- 3× { Conv3D(8, (6, 6, 8))+ReLU }
- Conv3D(6, (4, 4, 6))+ReLU
- Conv3D(6, (3, 3, 5))+ReLU
- Conv3D(1, (2, 2, 2))+ReLU
- Cropping3D
- Output (30×30×29)

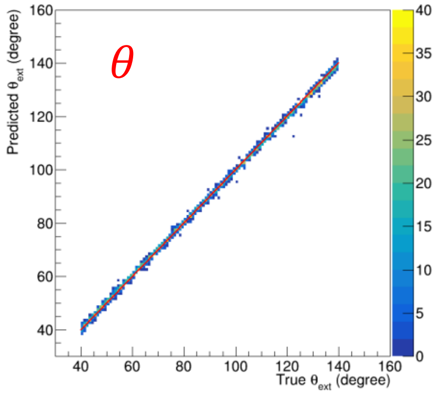
Discriminator

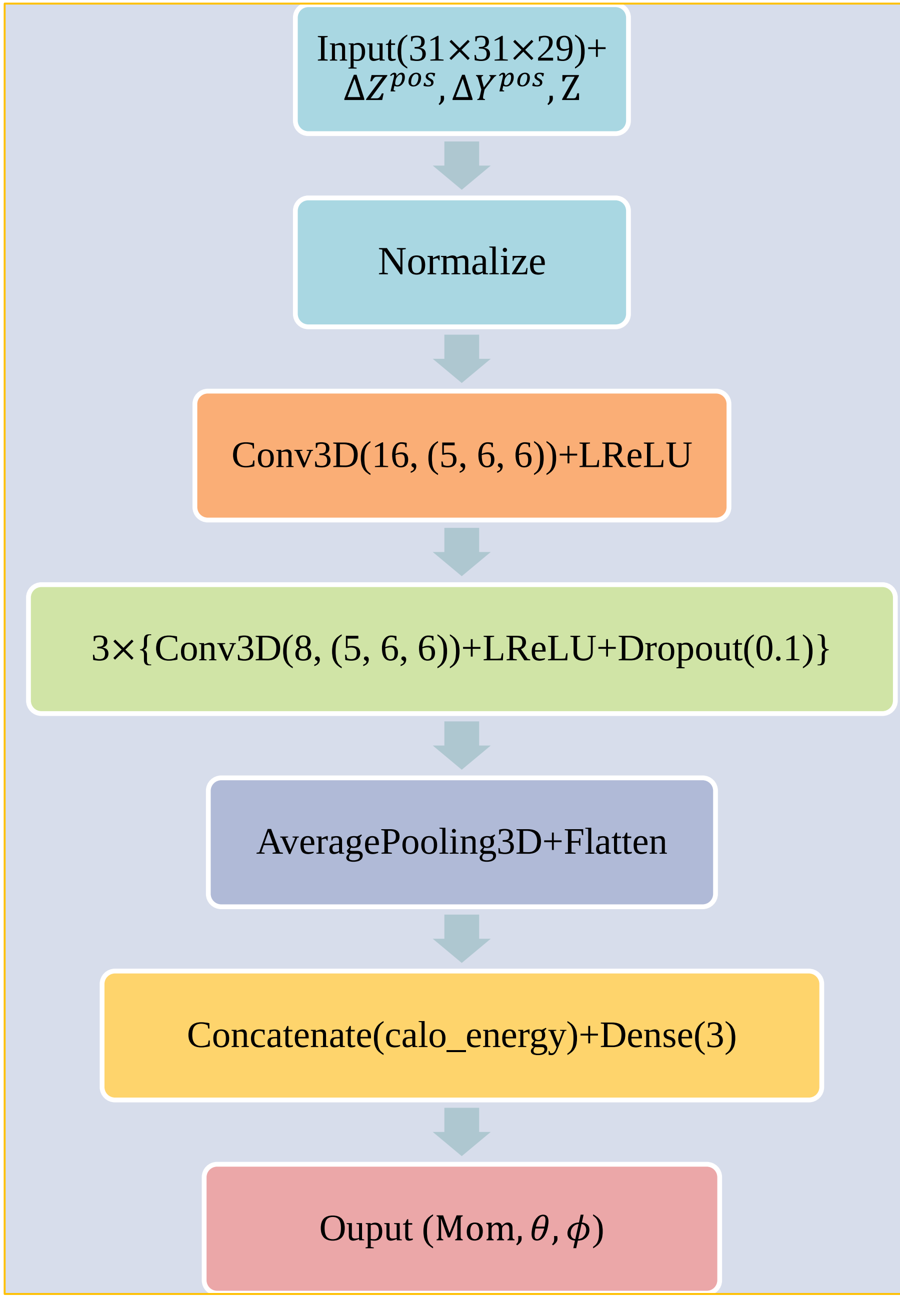
- Input(30×30×29)
- normalize
- Conv3D(16, (2, 2, 2))+LReLU
- 4× {Conv3D(8, (3, 3, 3))+LReLU+Dropout(0.1)}
- AveragePooling3D+Flatten+Dense(1)
- Ouput (real/fake)

Regressor architecture and performance (γ)









Input(31×31×29)+ $\Delta Z^{pos}, \Delta Y^{pos}$, Z

Normalize

Conv3D(16, (5, 6, 6))+LReLU

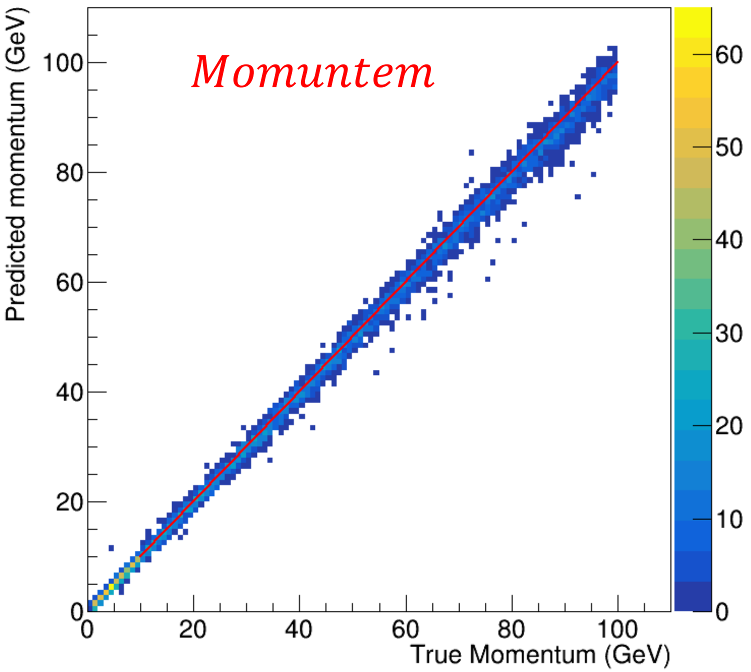
3×{Conv3D(8, (5, 6, 6))+LReLU+Dropout(0.1)}

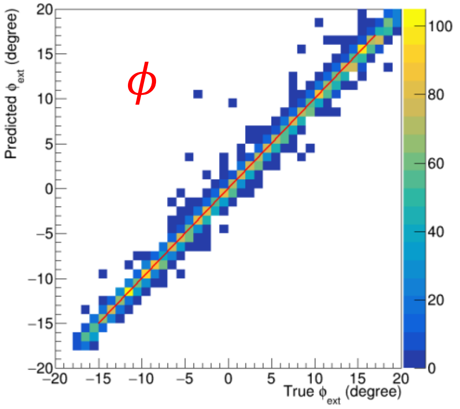
AveragePooling3D+Flatten

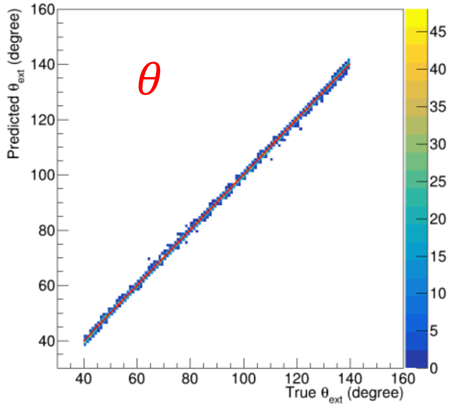
Concatenate(calo_energy)+Dense(3)

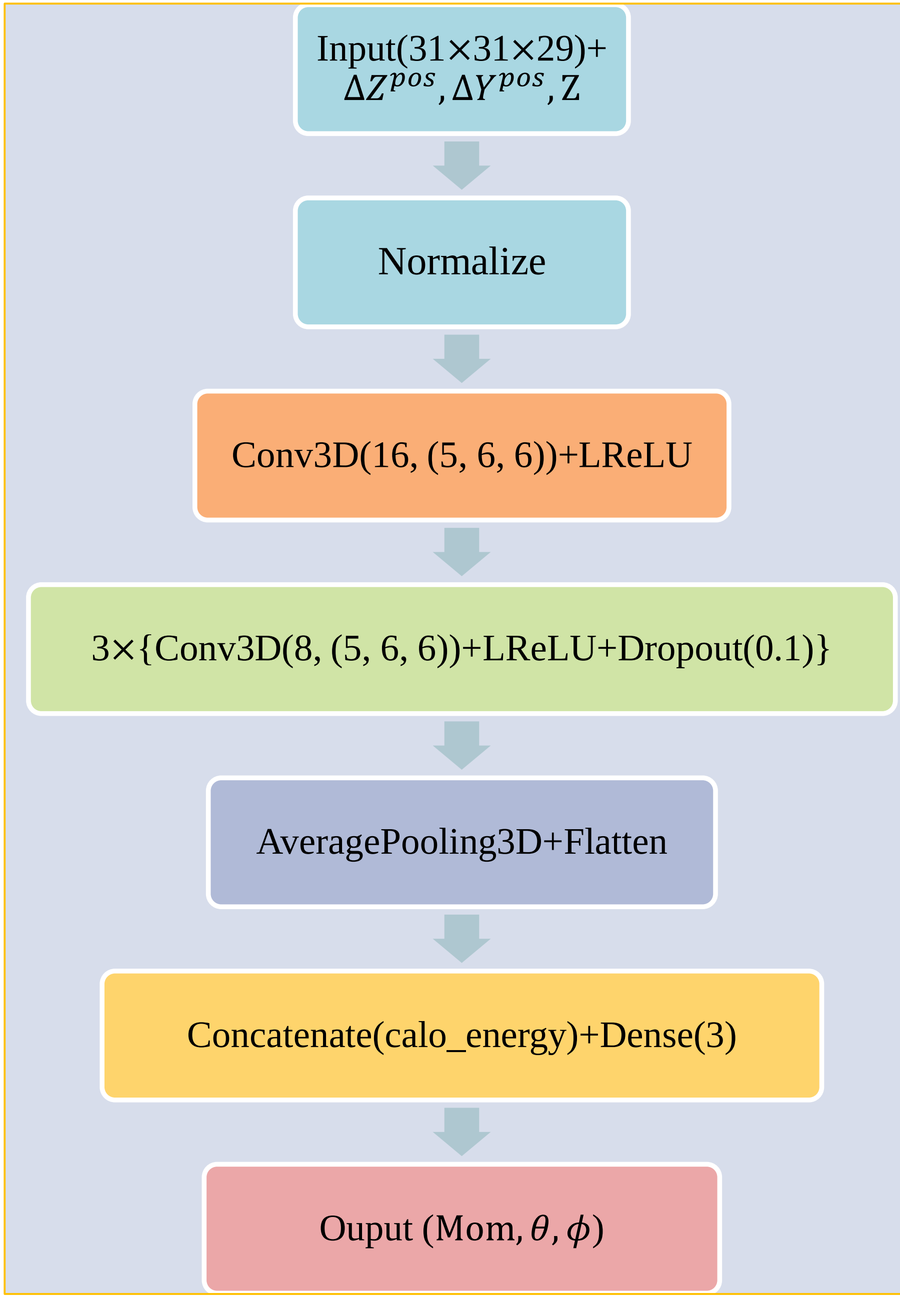
Ouput (Mom, θ, ϕ)

Regressor architecture and performance (e^-)









Input(31×31×29)+ $\Delta Z^{pos}, \Delta Y^{pos}$, Z

Normalize

Conv3D(16, (5, 6, 6))+LReLU

3×{Conv3D(8, (5, 6, 6))+LReLU+Dropout(0.1)}

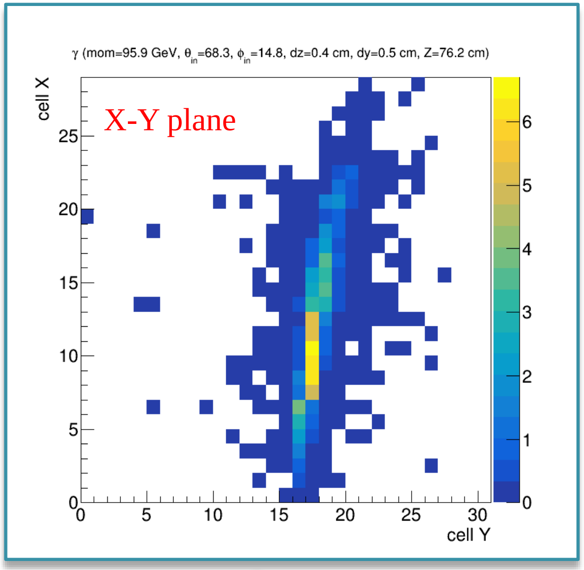
AveragePooling3D+Flatten

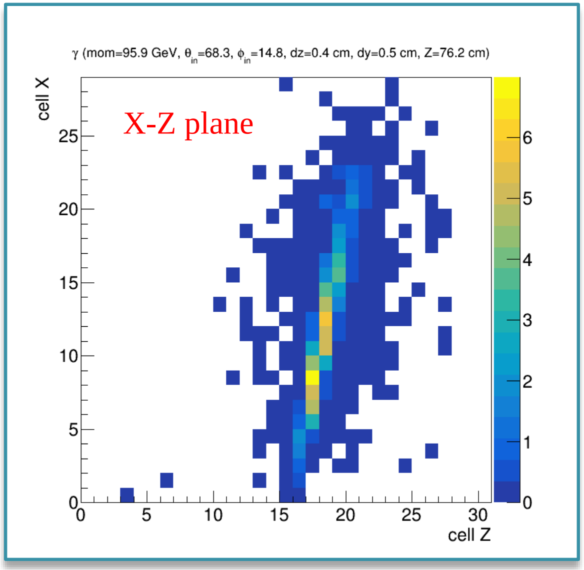
Concatenate(calo_energy)+Dense(3)

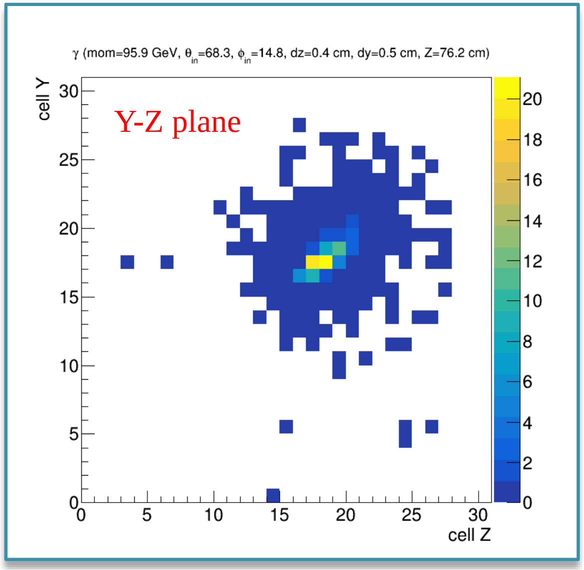
Ouput (Mom, θ, ϕ)

Event display

γ (Mom $= 95.9$ GeV, $\theta_{in} = 63.8^{\circ}$, $\phi_{in} = 14.8^{\circ}$, $\Delta Z^{Pos} = 0.4$ cm, $\Delta Y^{Pos} = 0.5$ cm, $Z = 76.2$ cm)

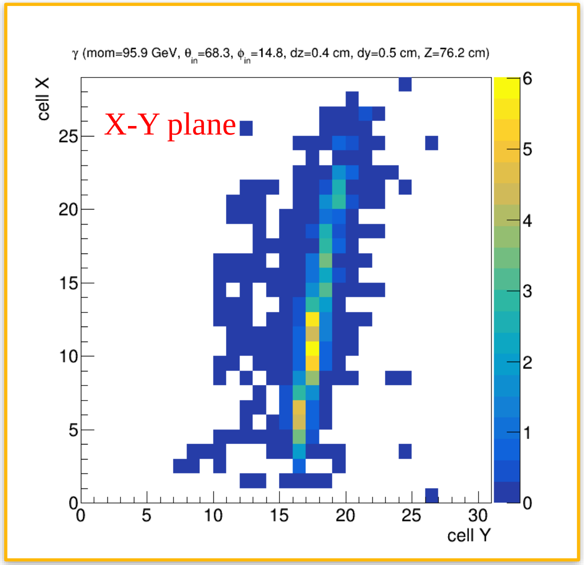


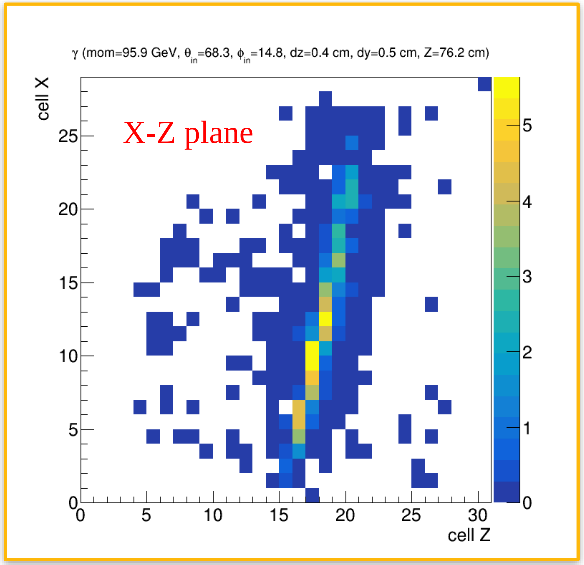


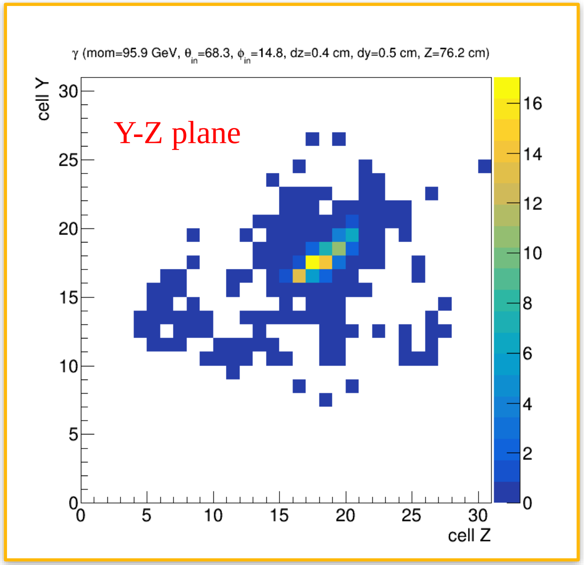


Geant4

GAN

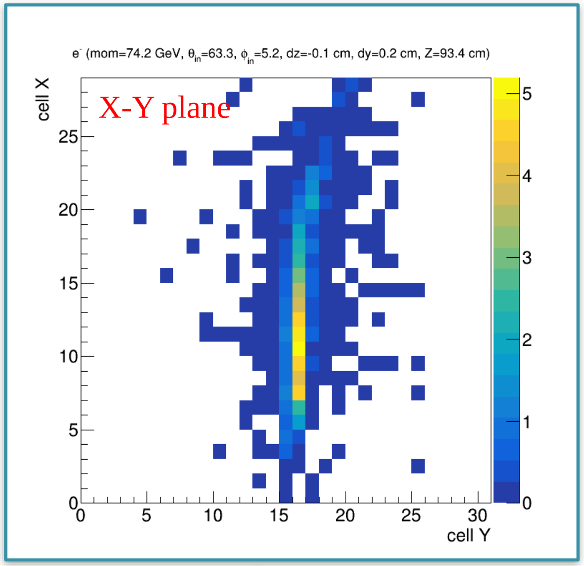


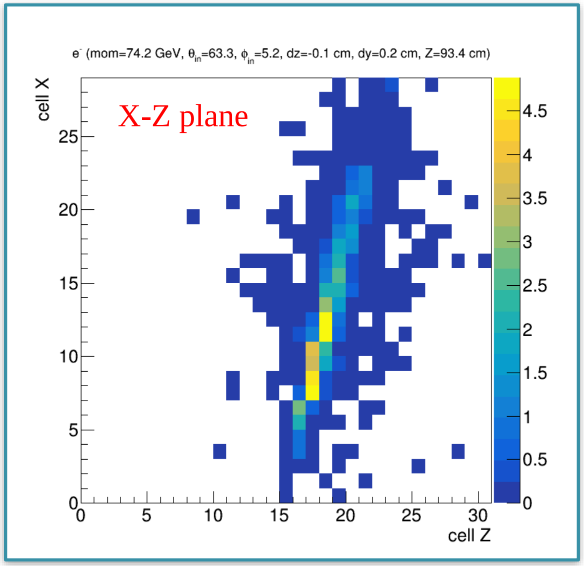


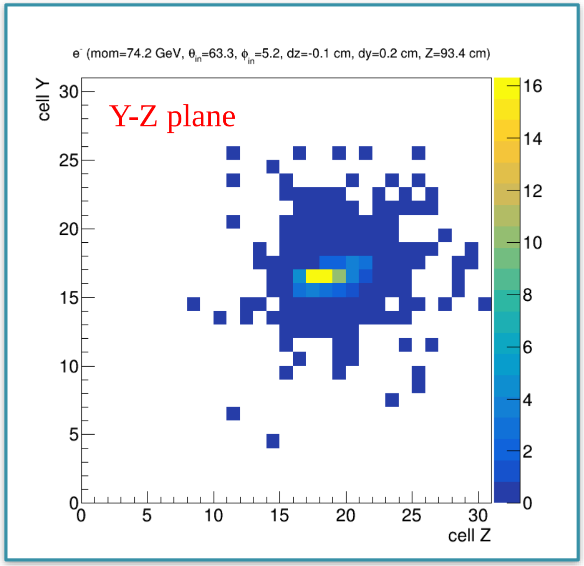


Event display of e^-

e^-(Mom = 74.2 GeV, $\theta_{\text{in}} = 63.3^\circ$, $\phi_{\text{in}} = 5.2^\circ$, $\Delta Z^{\text{Pos}} = -0.1$ cm, $\Delta Y^{\text{Pos}} = 0.2$ cm, Z = 93.4 cm)

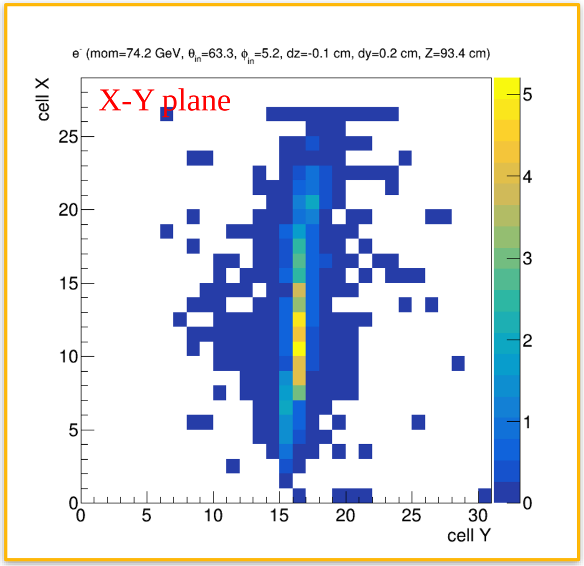


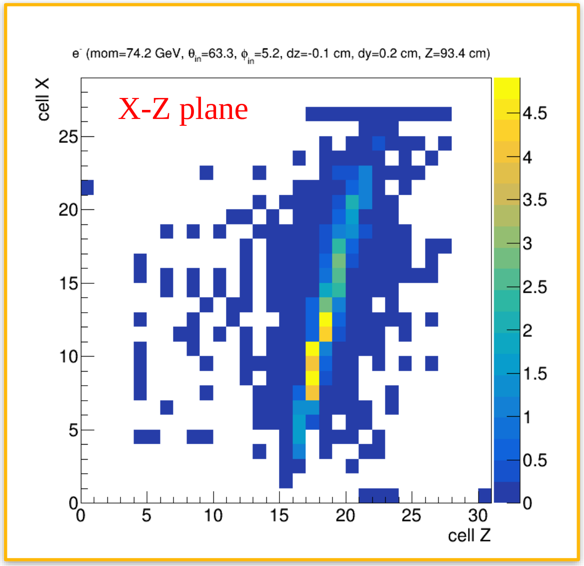


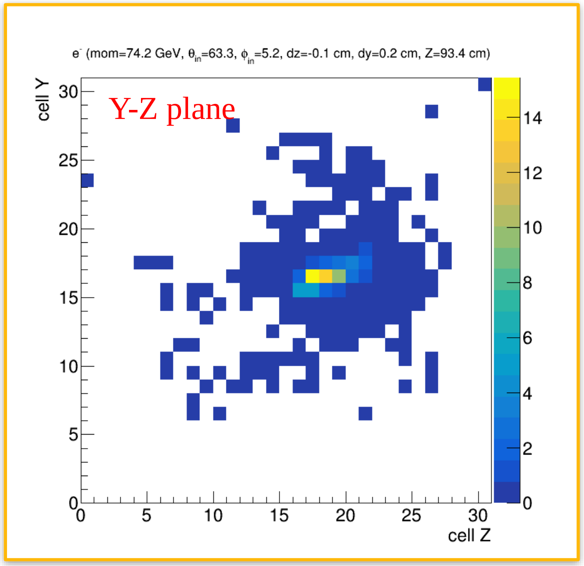


Geant4

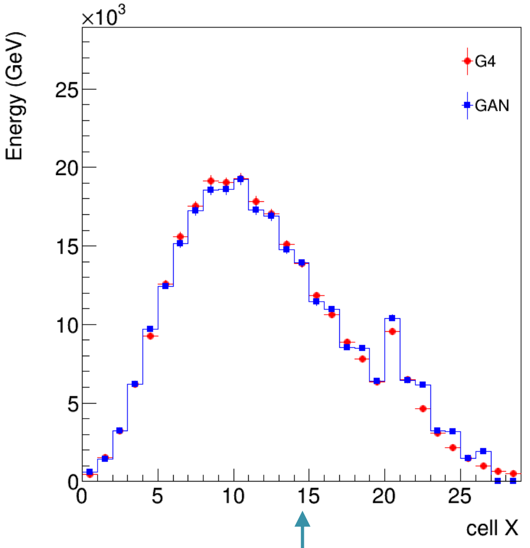
GAN

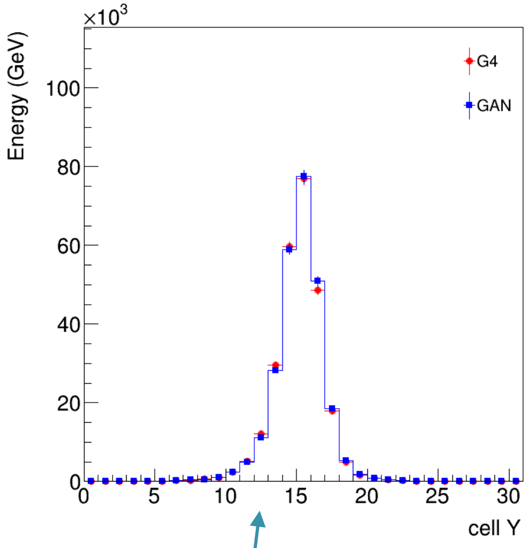


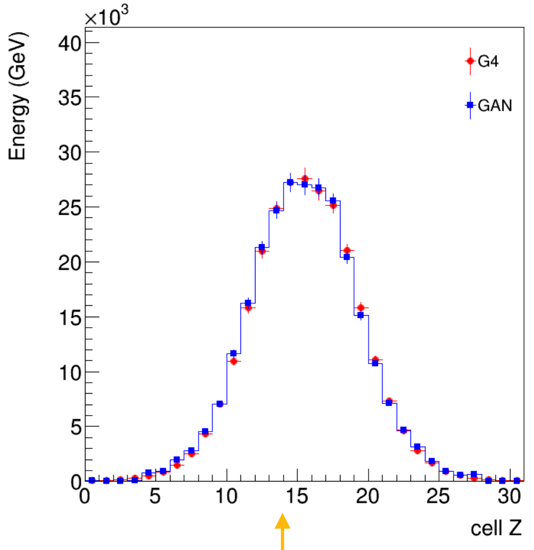




Some variables for e^- (cepc)







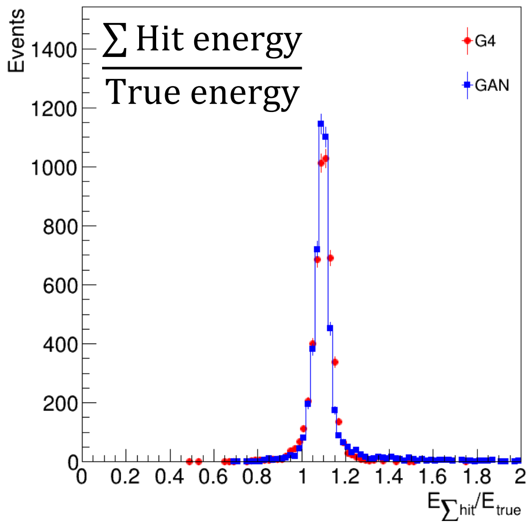
Energy deposited in X(layer) direction

Energy deposited in Y direction

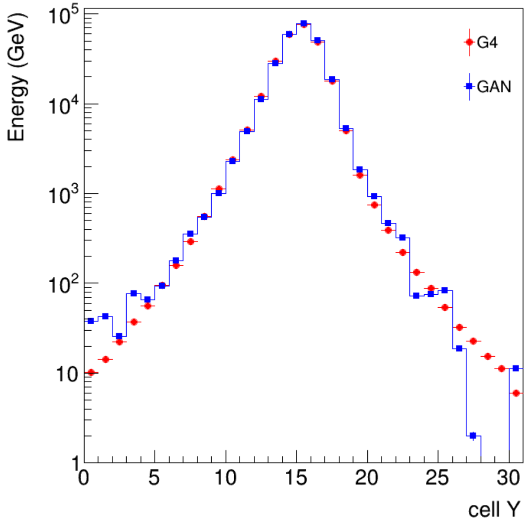
Energy deposited in Z direction

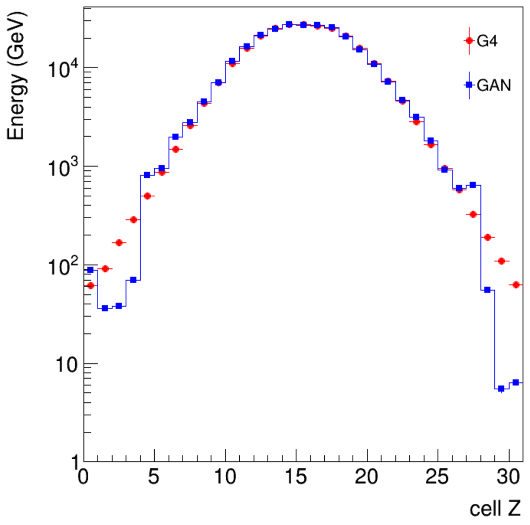
log scale

log scale



$$\frac{\sum \text{Hit energy}}{\text{True energy}}$$



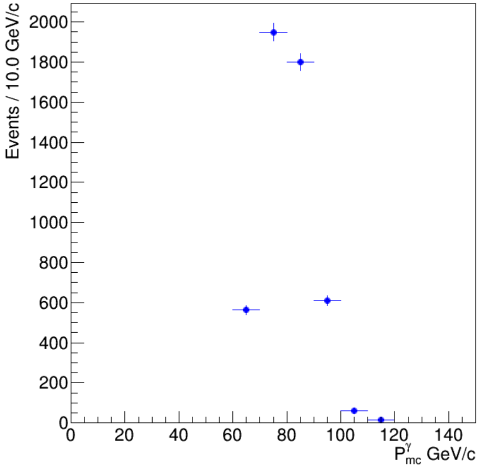


using mc samples

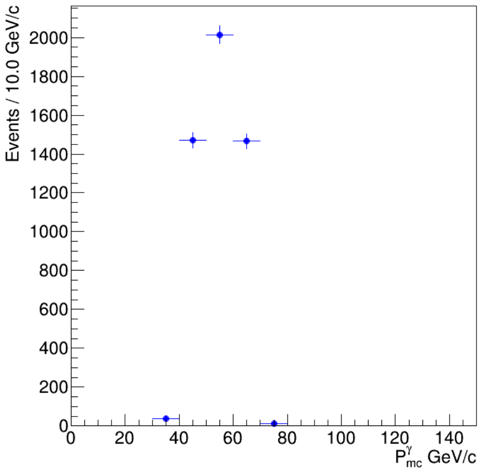
$e^+e^- \rightarrow Z(\nu\nu)H(\gamma\gamma)$

- Dataset: /cefs/data/FullSim/CEPC240/CEPC_v4/higgs/E240.Pnnh_aa.e0.p0.whizard195/nnh_aa.e0.p0.0000*_sim.slcio

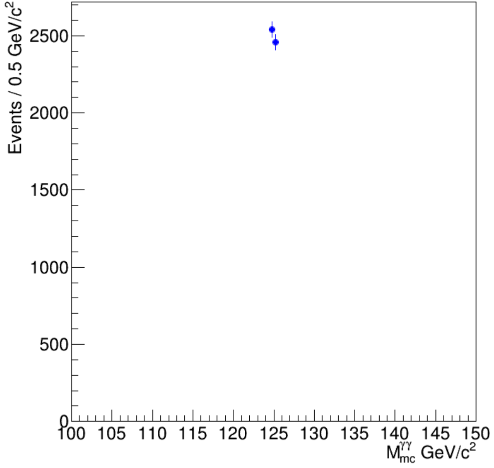
Momentum leading γ



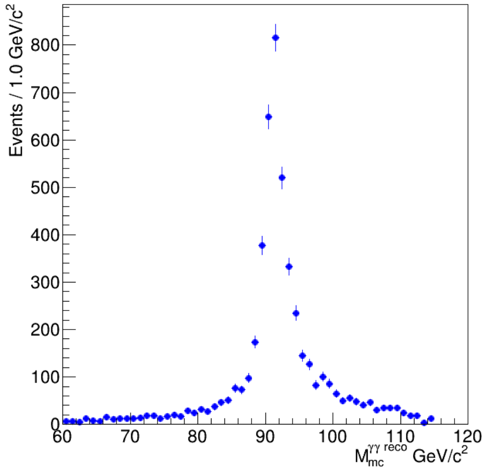
Momentum sub − leading γ



M$(\gamma\gamma)$



M reco $(\gamma\gamma)$

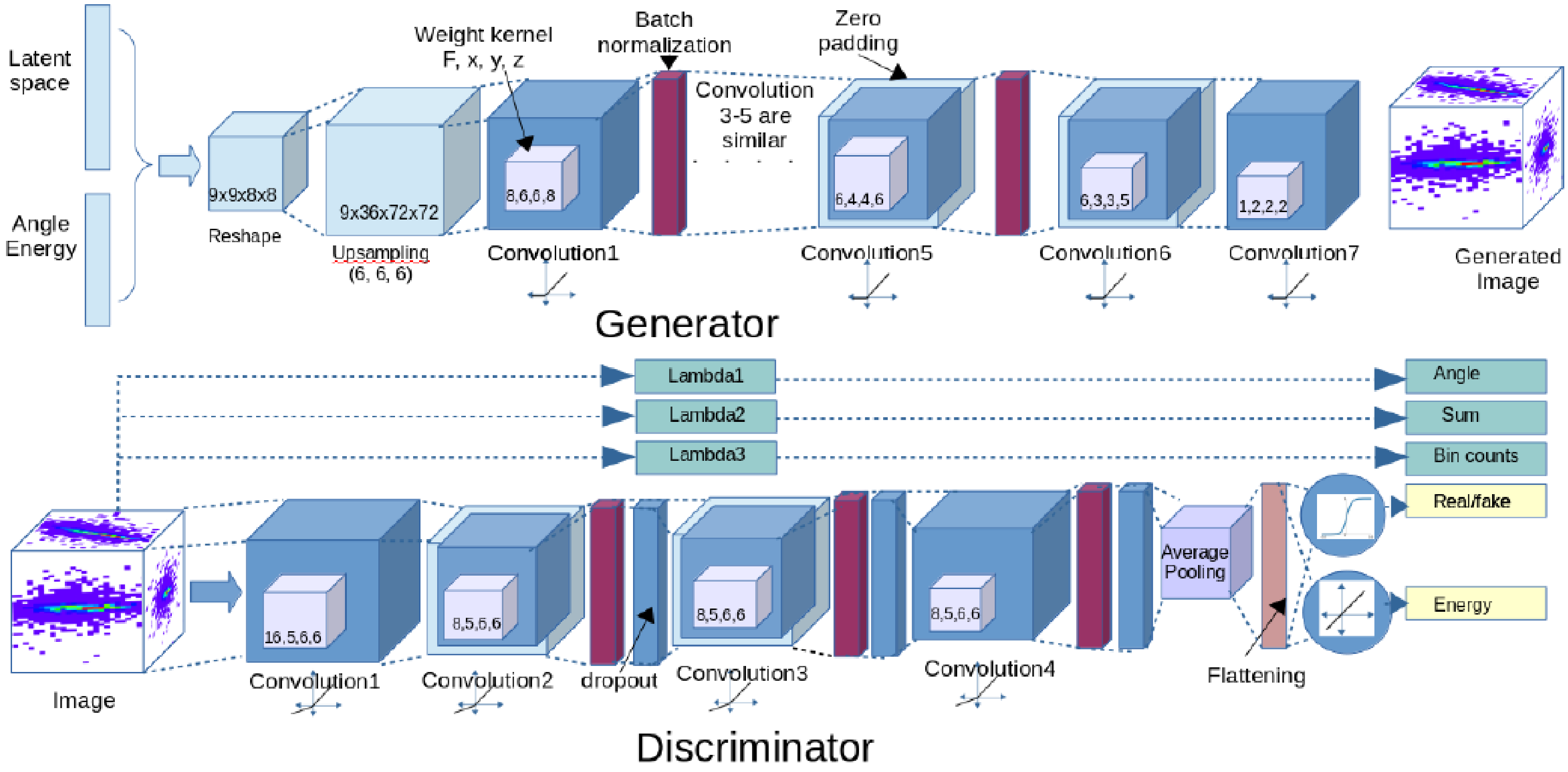


The model

3D convolutional Generative Adversarial Networks

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

Auxiliary regression tasks assigned to the discriminator



Publications

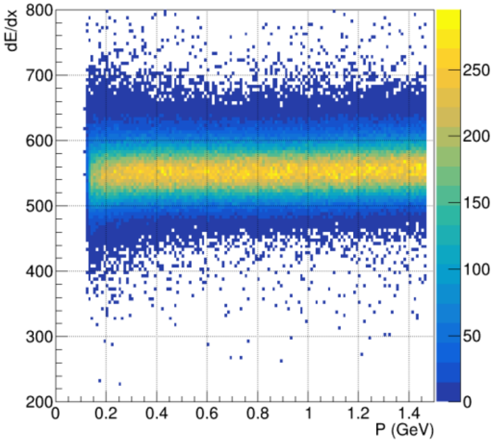
- CMS Collaboration, “Search for narrow resonances in dilepton mass spectra in proton-proton collisions at $\sqrt{S}$ = 13 TeV and combination with 8 TeV data”, PLB, DOI:10.1016/j.physletb.2017.02.010
- CMS Collaboration , “Search for high-mass resonances in dilepton final states in proton-proton collisions at $\sqrt{S}$= 13 TeV”, JHEP, DOI:10.1007/JHEP06(2018)120
- CMS Collaboration, “Search for high mass resonances in dielectron final state”, CMS-PAS-EXO-18-006
- CMS Collaboration , “Search for new physics via top quark production in dilepton final state at 13 TeV”, Eur. Phys. J. C (2019) 79:886
- Clerbaux B., Fang W., Giammanco A. et al , “Model-independent constraints on the CKM matrix elements $|V_{tb}|$, $|V_{ts}|$ and $|V_{td}|$”, JHEP, DOI:10.1007/JHEP03(2019)022

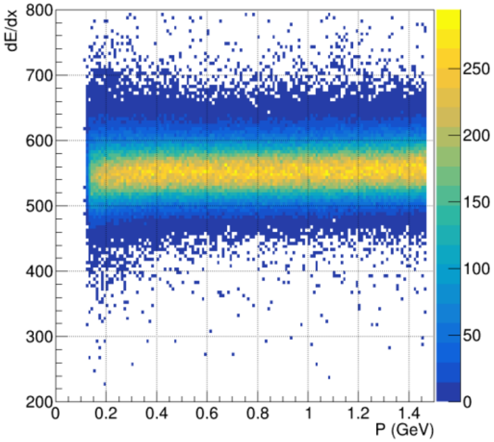
Personal contributions

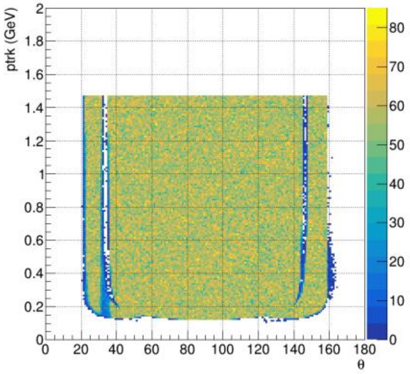
- $Z' \rightarrow ee$ study:
    - Calculating High Energy Electron Pair (HEEP) selection efficiency and the scale factor between data and MC.
    - Measuring mass scale and resolution.
    - Doing electron saturation study.
    - Doing background fit.
    - Calculating systematic uncertainty.
    - Drawing official dielectron invariant mass.
- Phenomenological work:
    - Calculating systematic uncertainty.
    - Performing a global fit of V_{tb}, V_{ts} and V_{td}.
    - Providing all plots.
- Top EFT study:
    - Drawing data and MC distributions for different variables.
    - Doing trigger scale factor study between data and MC.
    - Background estimation.
    - Calculating systematic uncertainty.
    - Draw final NN distributions.
    - Cross check of tW cross section measurement.
    - Calculating limits for EFT couplings.

For electron (positron) study

- Radiative Bhabha events from 2018 data





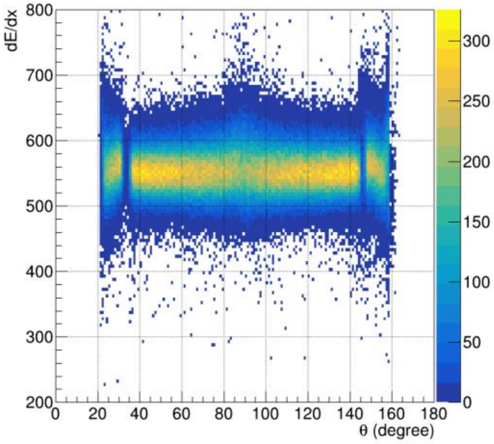


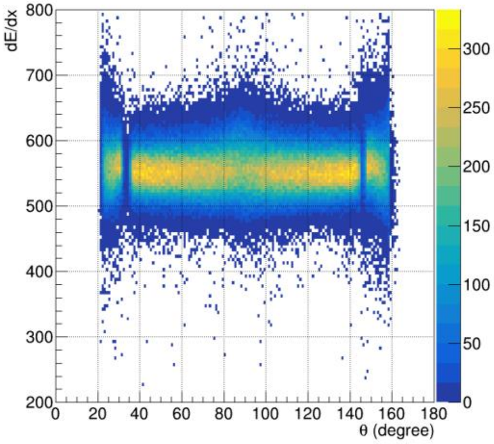
Data: e⁻

NN: e⁻

dE/dx vs P

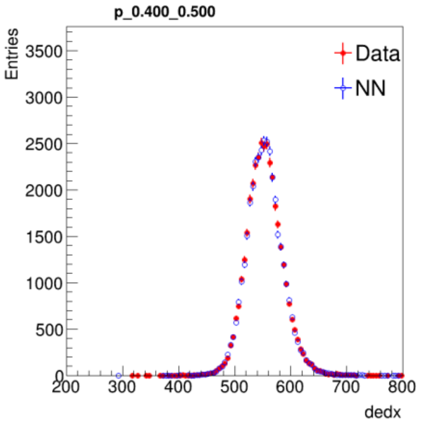
- Similar for e⁺



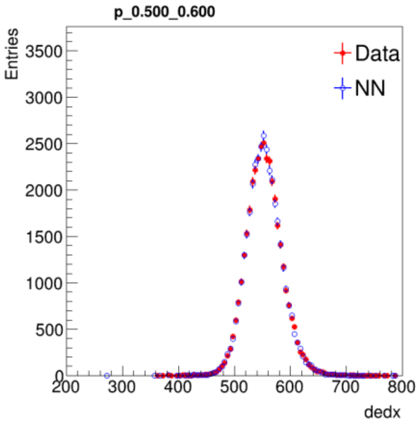


dE/dx vs θ

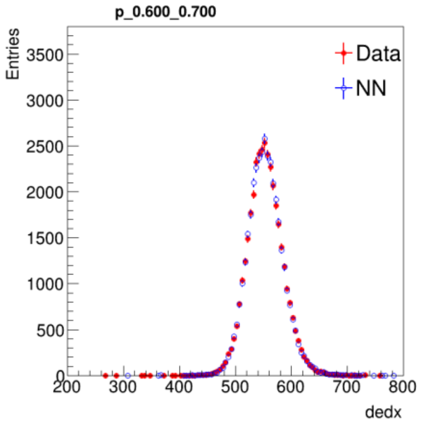
dE/dx distribution (e^-) vs P



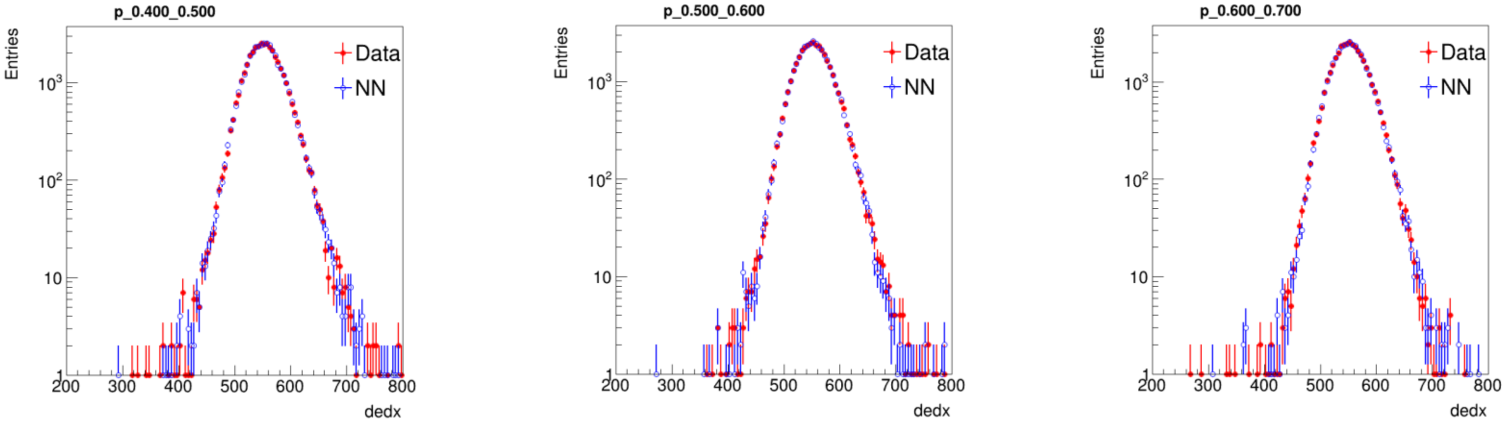
P: 0.4-0.5 GeV



P: 0.5-0.6 GeV



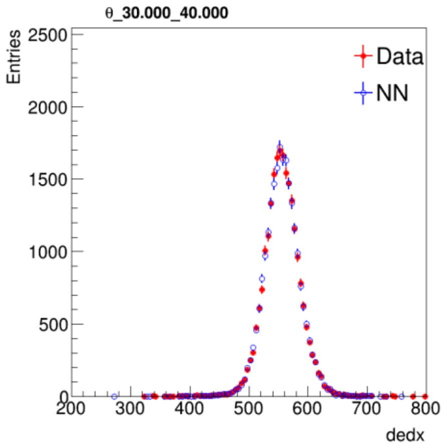
P: 0.6-0.7 GeV



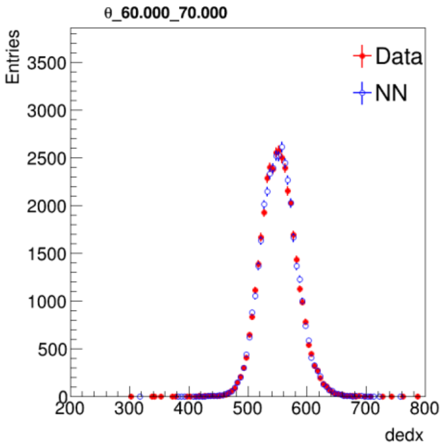
Log scale

- The nice agreement between data and NN

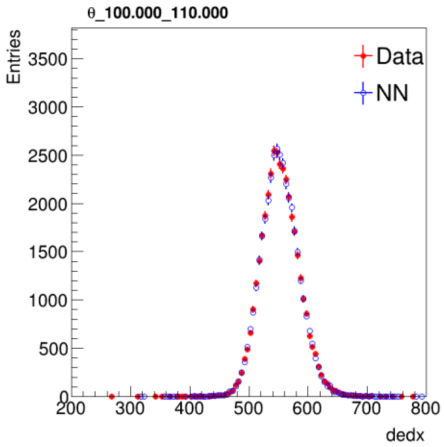
dE/dx distribution (e^-) vs θ



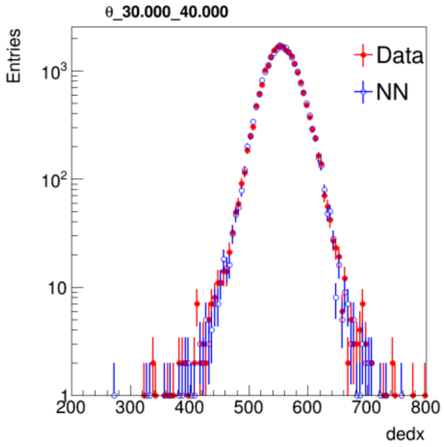
$\theta: 30^\circ - 40^\circ$

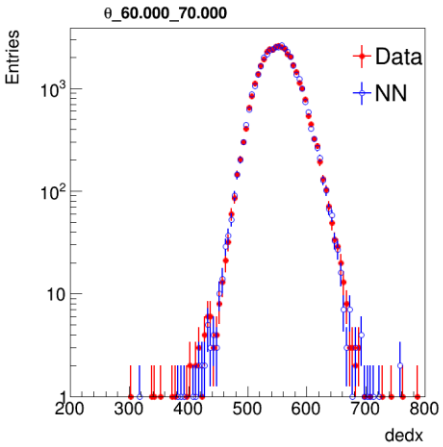


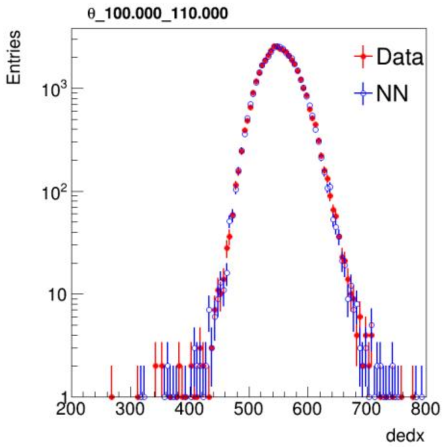
$\theta: 60^\circ - 70^\circ$



$\theta: 100^\circ - 110^\circ$





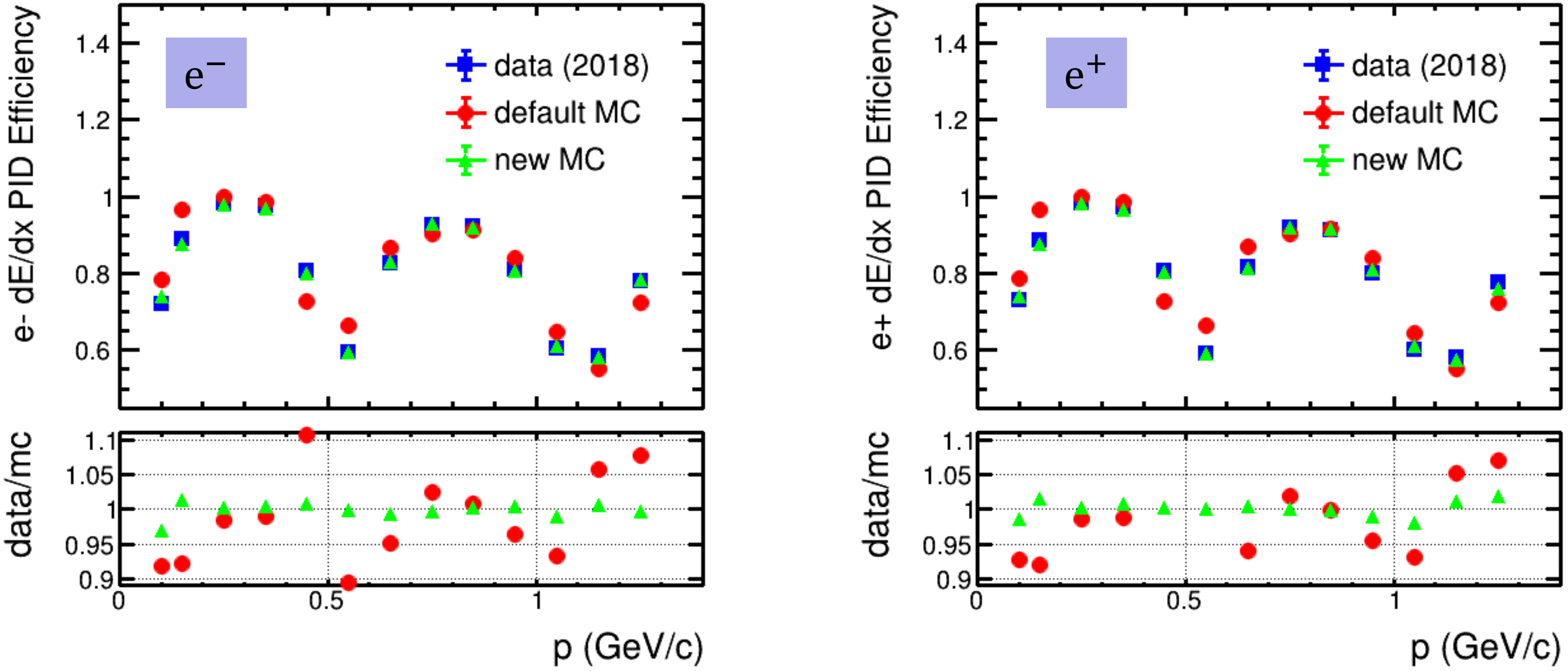


Log scale

- The nice agreement between data and NN

dE/dx PID efficiency for $e^{\pm}$

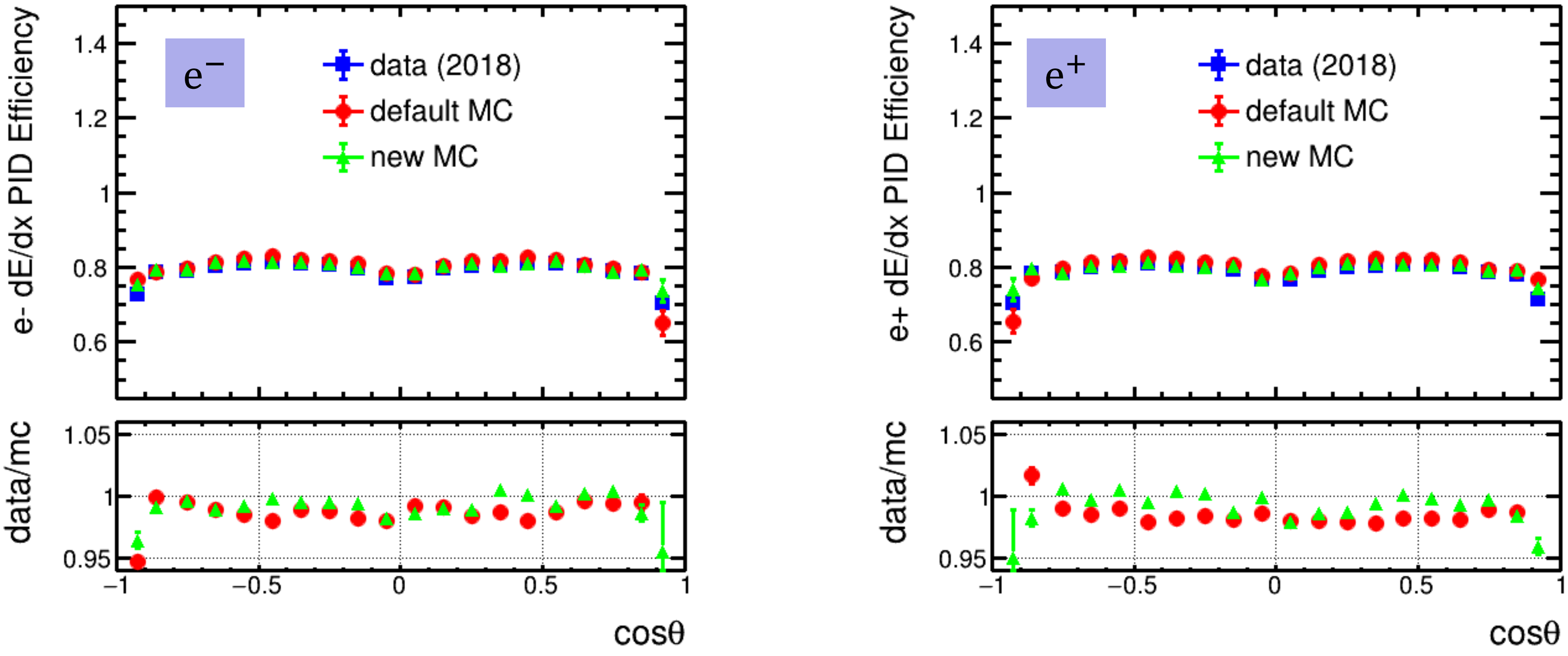
Measured by Mengzhen Wang



- Improved agreement with data when using new MC

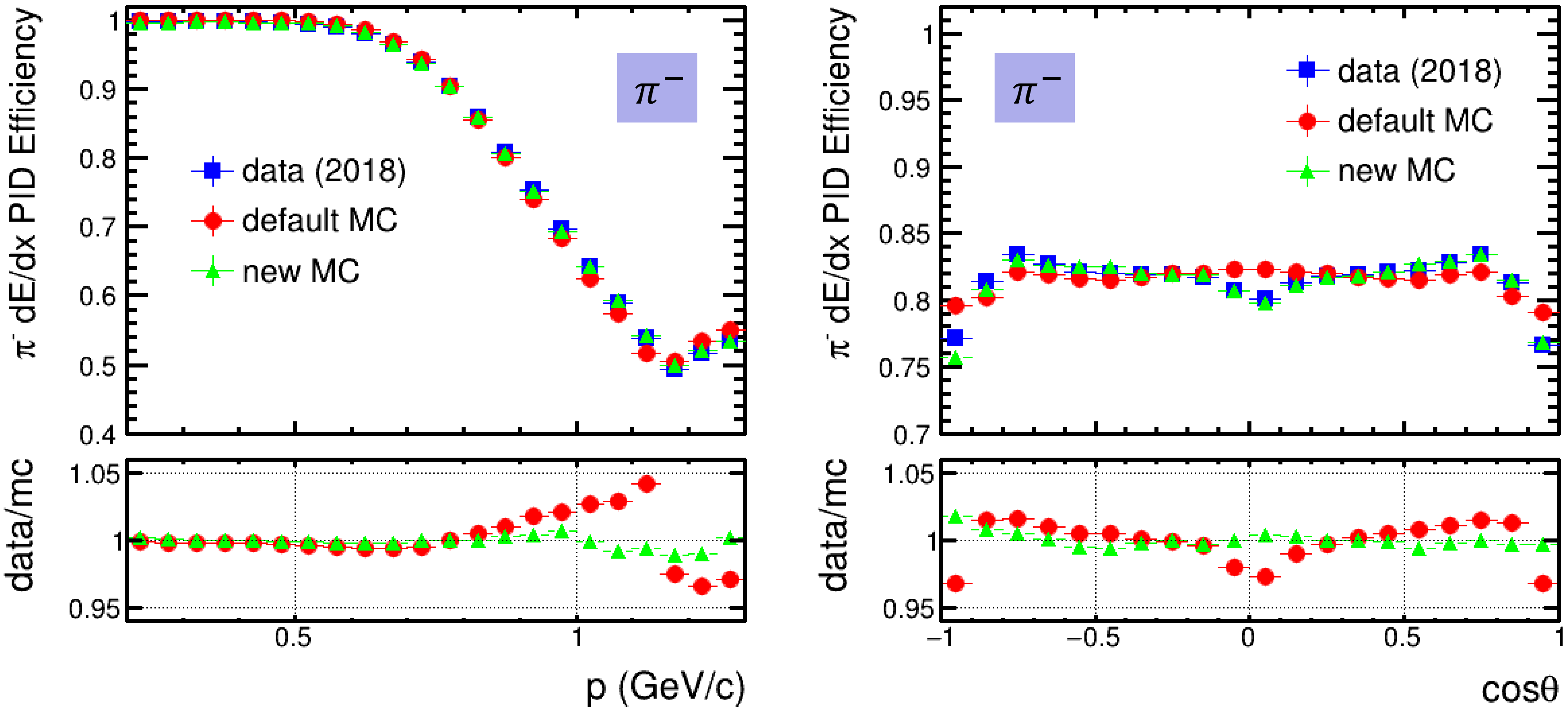
dE/dx PID efficiency for $e^{\pm}$

Measured by Mengzhen Wang



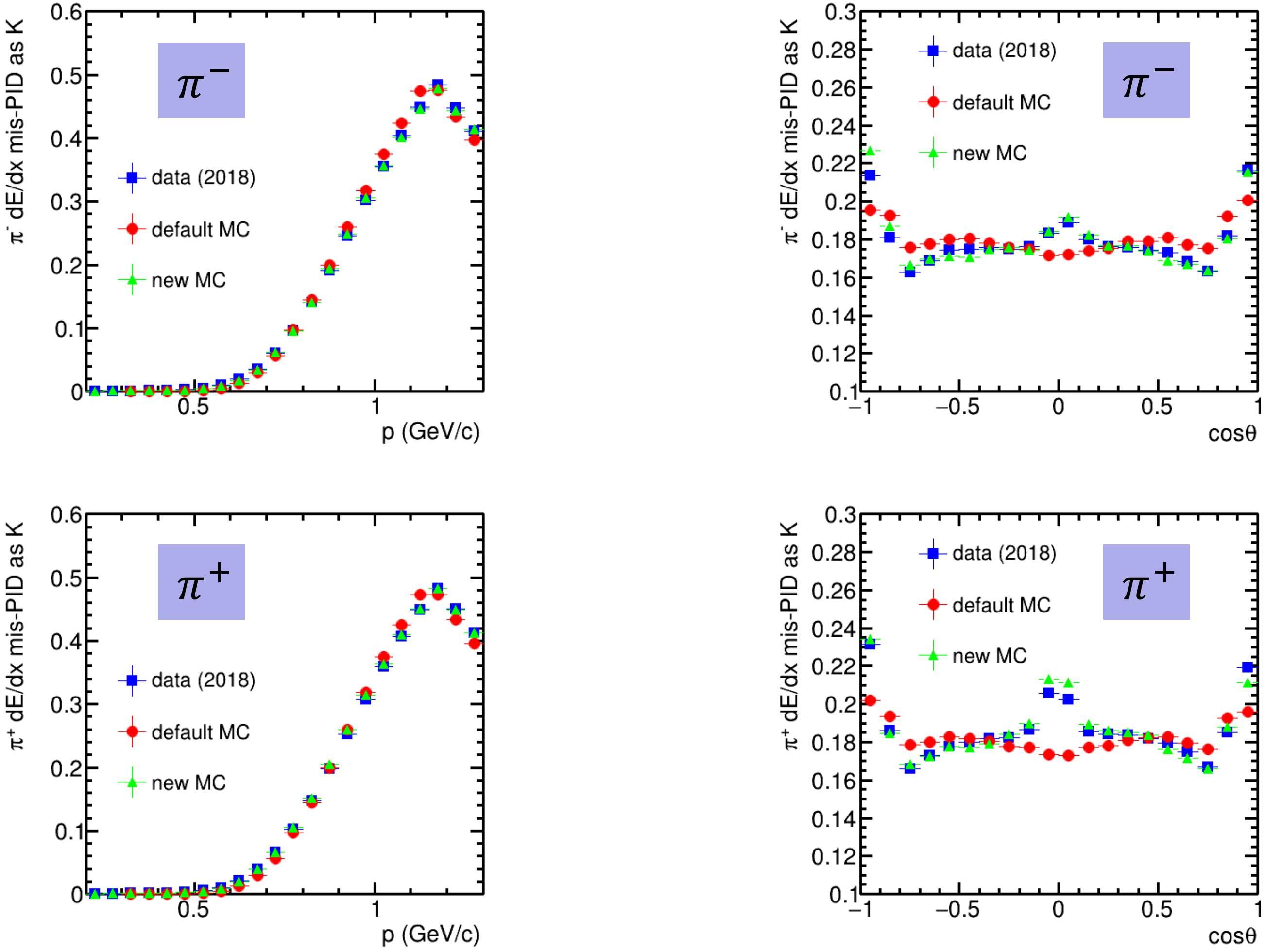
- Improved agreement with data when using new MC

Pion dE/dx PID 效率



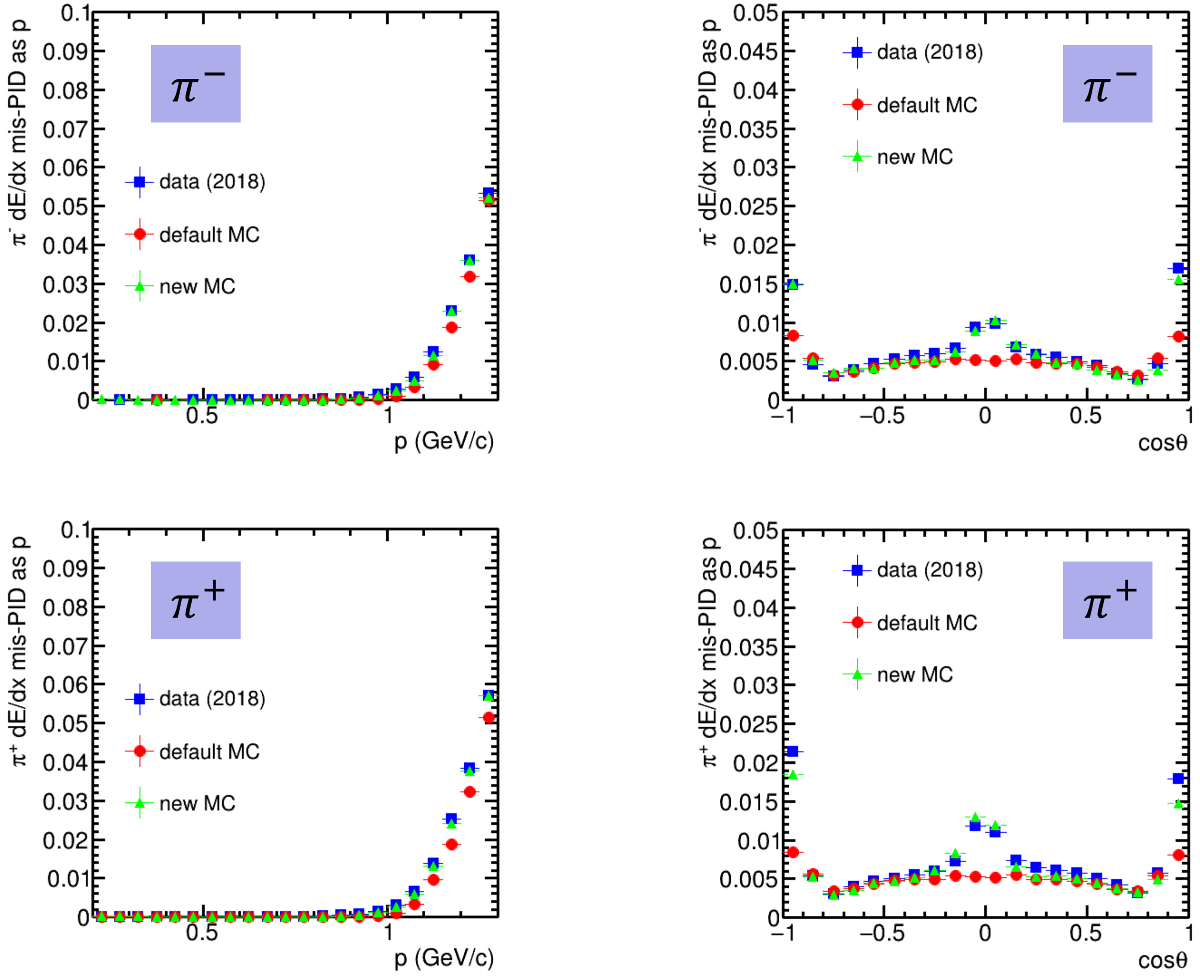
- 神经网络模拟的结果与真实数据符合的更好

Pion dE/dx mis-PID as K



- 对于误鉴别率，神经网络模拟结果也与真实数据符合的更好

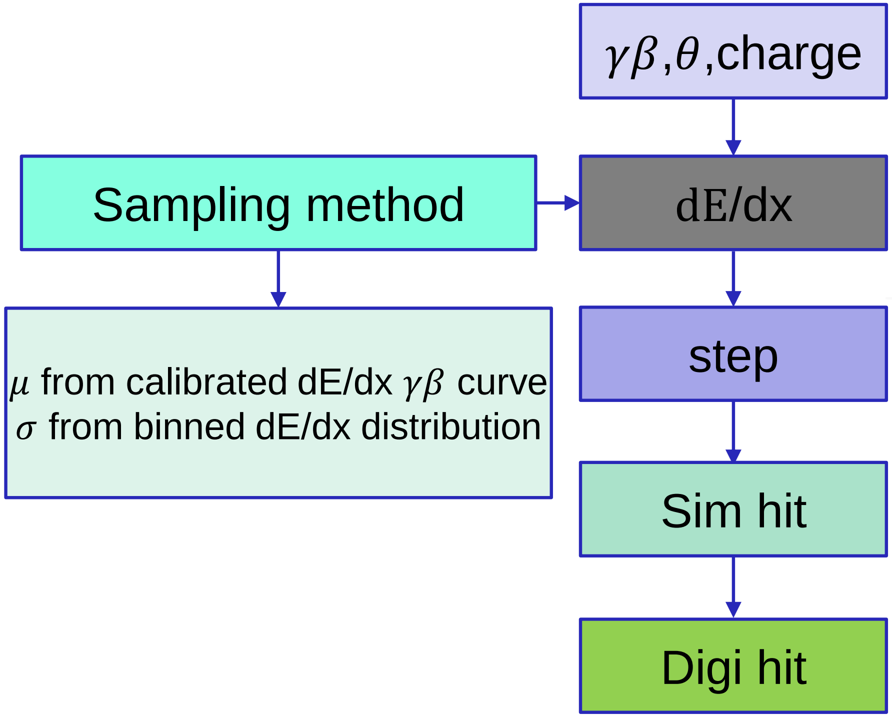
Pion dE/dx mis-PID as (anti-)proton



- 对于误鉴别率，神经网络模拟结果也与真实数据符合的更好

BESIII dE/dx 模拟和重建

Simulation



$\gamma\beta$,θ,charge

Sampling method

dE/dx

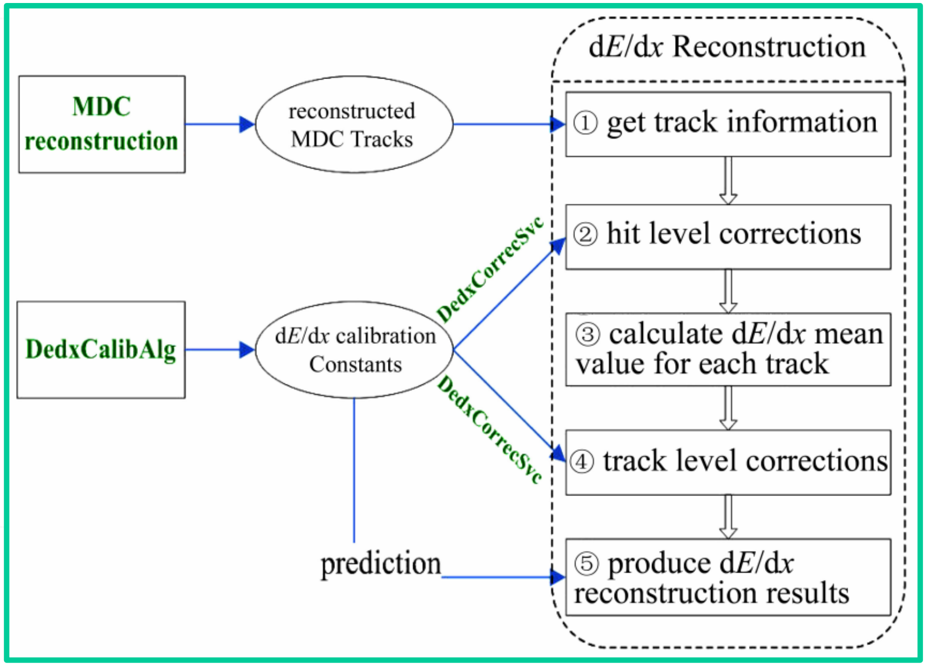
μ from calibrated dE/dx $\gamma\beta$ curve
σ from binned dE/dx distribution

step

Sim hit

Digi hit

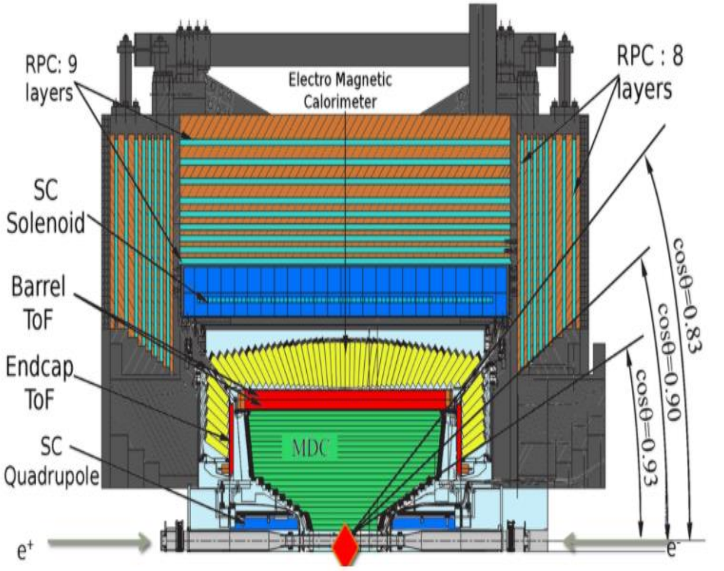
Reconstruction

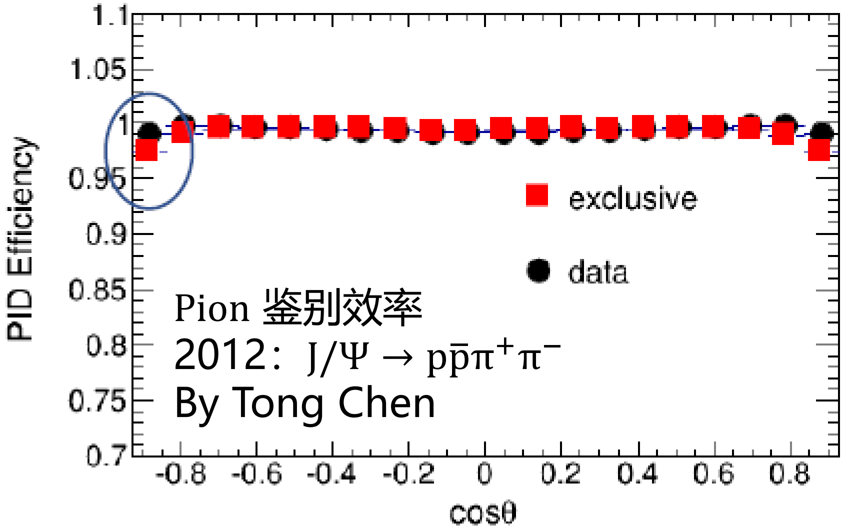


- 该方法有效的工作了很多年，依然存在改善的空间，特别是对于带电强子的模拟 (pion, kaon, proton)

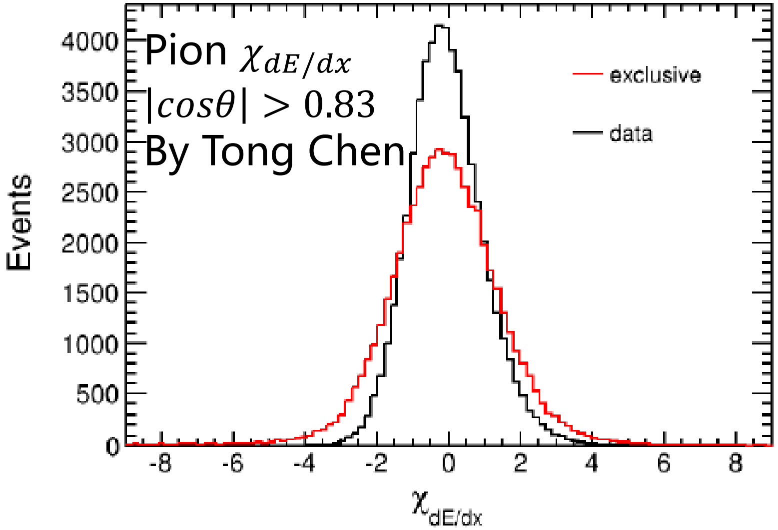
BESIII 实验物理现状

- 北京谱仪（BES）III实验通过正负电子在τ-charm能区对撞，研究陶物理、粲偶素、含粲强子等方面的物理。具有世界领先地位。
  - 面临激烈的国际竞争：BelleII，LHCb等实验
- BESIII目前积累了世界上最多的$e^+e^- \to J/\Psi$过程的数据（10^{10}个事例）
  - 统计误差大大减小
  - 系统误差成为限制物理结果精度的瓶颈，例如粒子鉴别效率。





Pion 鉴别效率
2012：$J/\Psi \to p\bar{p}\pi^+\pi^-$
By Tong Chen



Pion $\chi_{dE/dx}$
$|cos\theta| > 0.83$
By Tong Chen