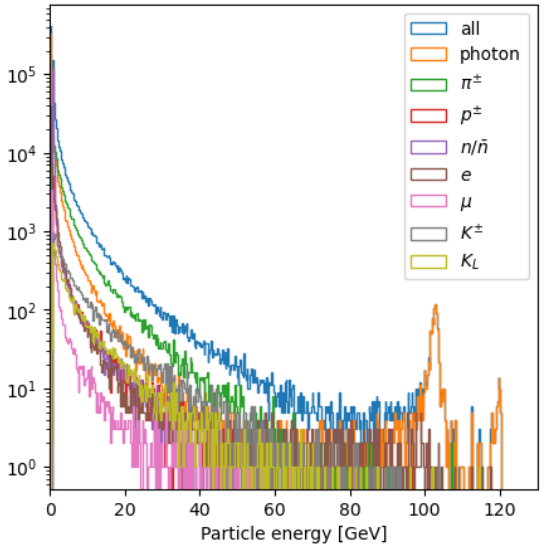
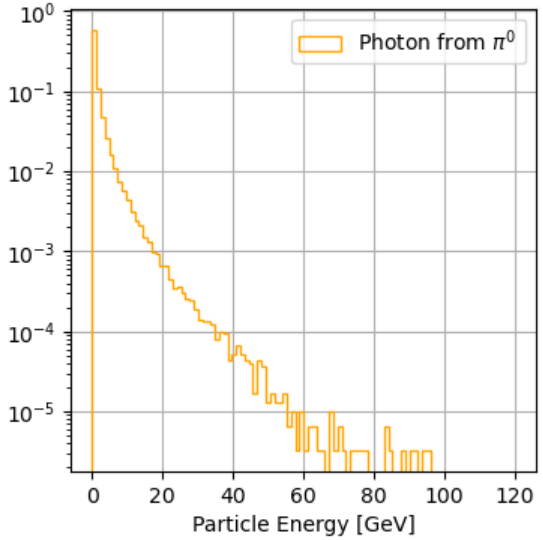
Re-optimization of SiW ECAL

Yukun Shi

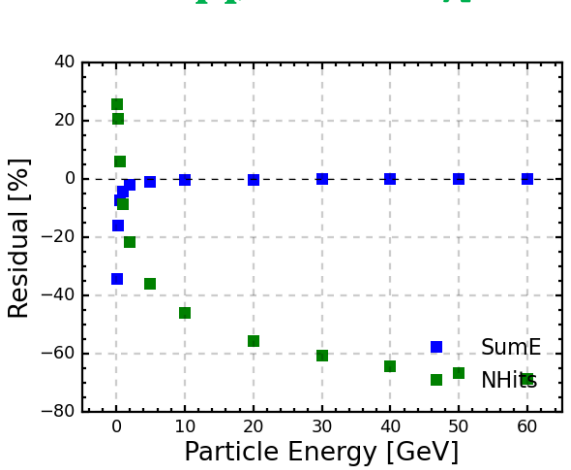
Motivation

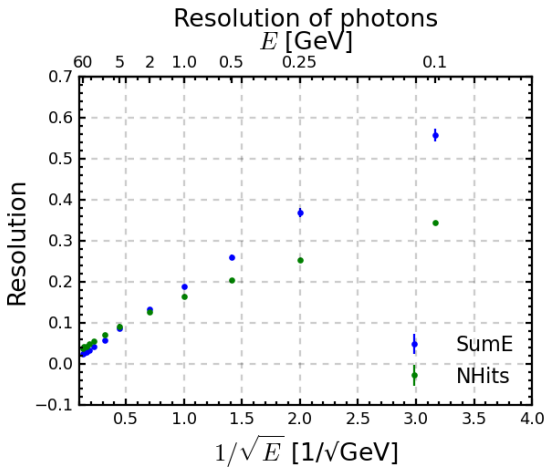
- From ILC to future circular collider
  - Continuous readout and ultra-high data flux
  - Lower CMS energy wrt ILC(500 GeV)
- Energy
  - Low energy photons in jets
  - The information provided by high granularity
- Time
  - improvement on PID and energy resolution
  - 30 ps achievable for hardware?





Particles reach Calo and photons from $\pi^0(e^+e^- \rightarrow q\bar{q}, 240\ GeV)$[Results from Hao L.]



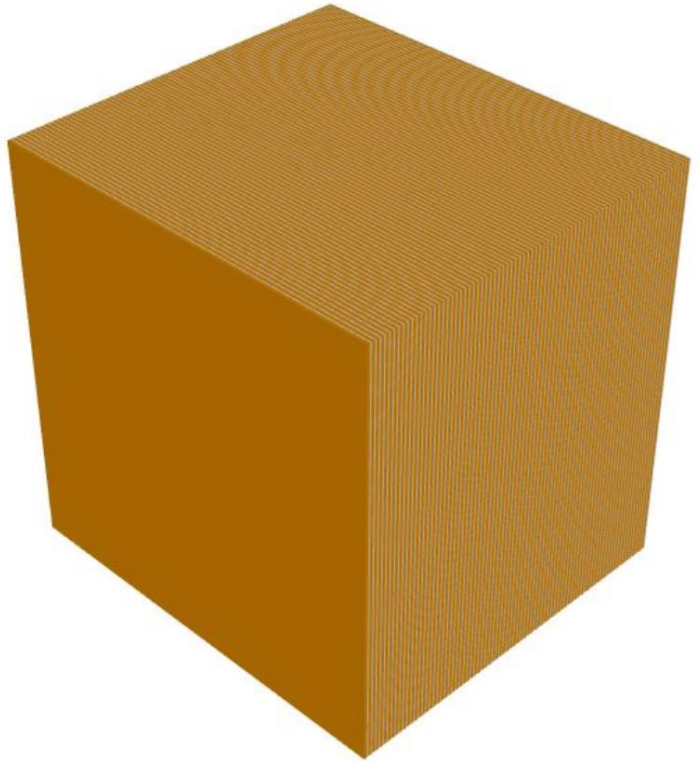


Residual and resolution for photons

Simulation setup

Geometry

- Sampling Layers: 80
- Absorber: tungsten
  - 1.5 mm per layer
  - ~34 X_0 for 80 layers
- Sensitive material: Silicon
  - 0.75 mm per layer
  - Less than 2% of W
  - Segmentation: $5 \times 5 \times 0.15\ mm^3$



ECAL prototype geometry

- Particles
  - photons with energy from 0 - 70 GeV
  - Position: (2.5 2.5 -200)(mm)
  - Direction: (0 0 1)

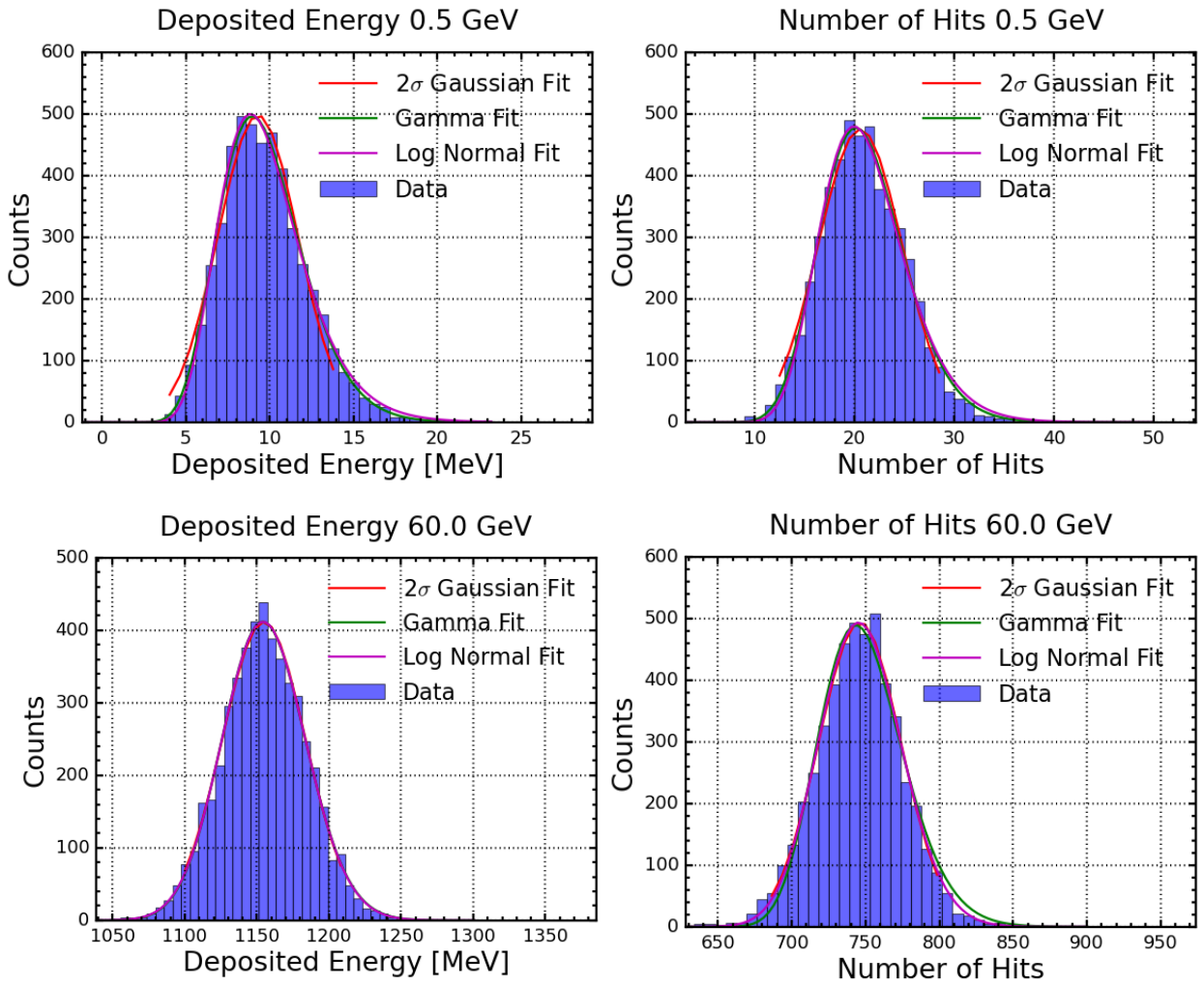
This geometry could be easily transformed into different configurations by selecting the readout cells

Energy deposition and Number of hits

Gamma function

- $f(x) = (x-\mu)^{k-1} \cdot e^{-\frac{x-\mu}{\theta}} / \theta^{k} \cdot \Gamma(k)$
- Mean :$k\theta + \mu$
- Peak: $(k-1)\theta + \mu$
- Variance: $k\theta^{2}$
- Resolution: $\sqrt{Varianve}\,/mean$

The gamma function will be used as fitting function in this study in order to describe the non- gaussian distribution in low energy range



The Fitting for Sum Energy(left) and Number of hits(right), the energy of gamma is 0.5 GeV(left) and 60 GeV(right)

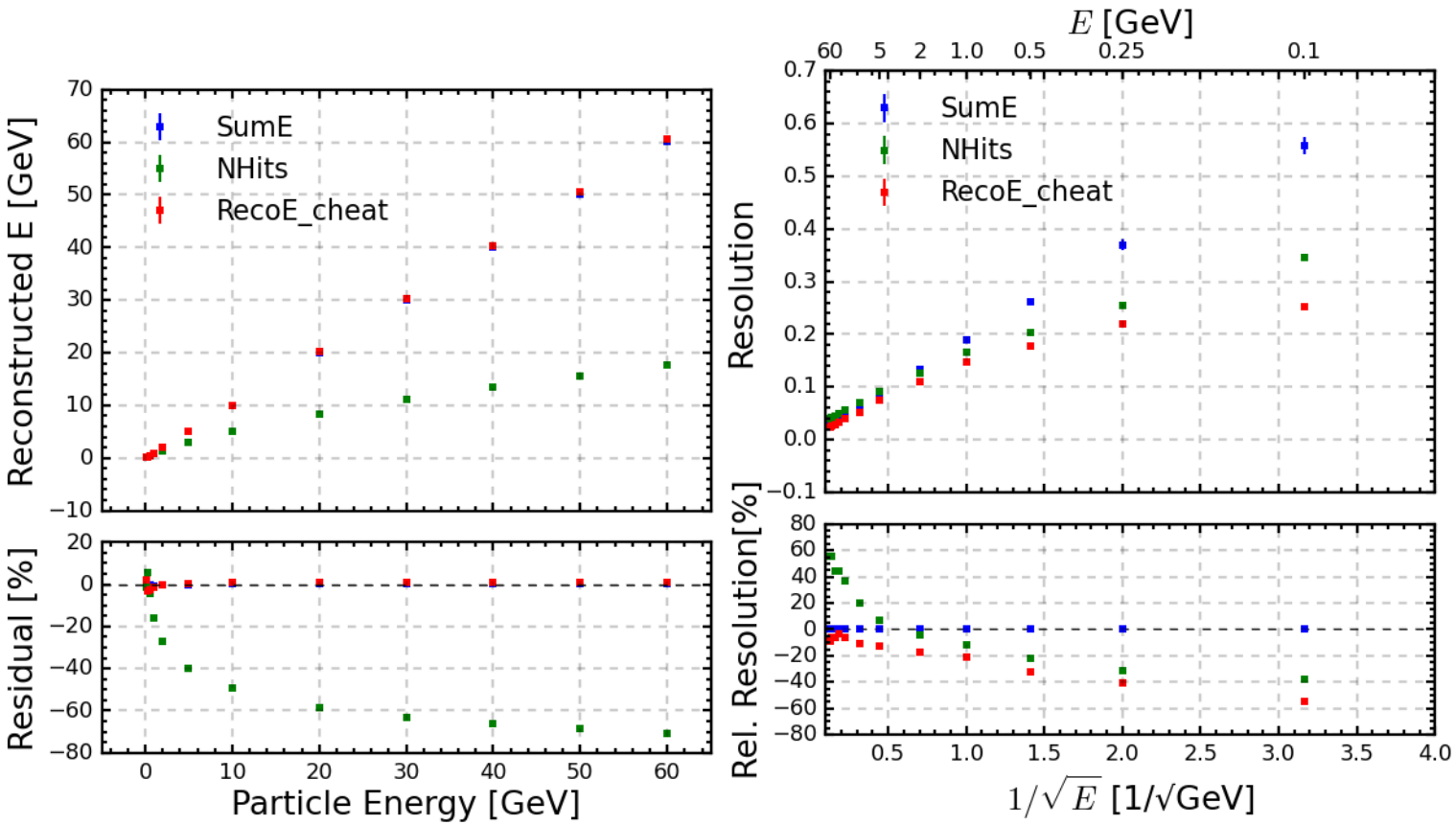
Energy reconstruction

Layer weight Method

- Reconstruction utilizing the layers
- $E_{reco} = \sum_{\mathrm{i}}^{\mathrm{Layer}} (a_i \cdot E_i + b_i \cdot N_i)$
- Parameters *a* and *b* are determined using the least chi-square method

Cheat to learn

- For each energy point, a set of parameters a and b are calculated



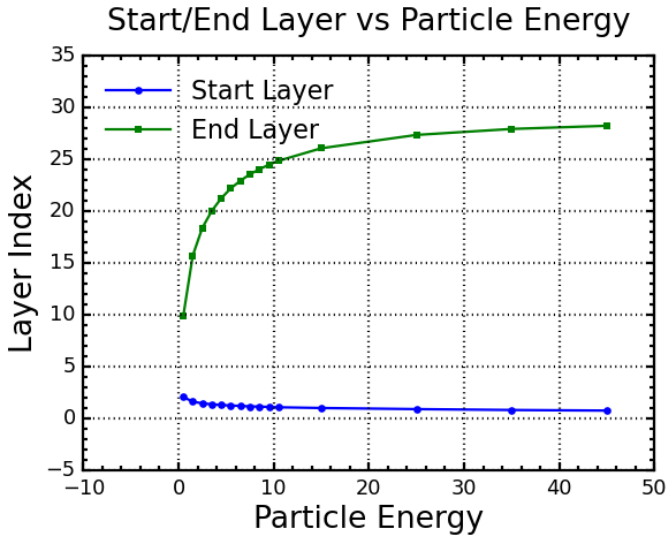
Linearity and resolution for gamma

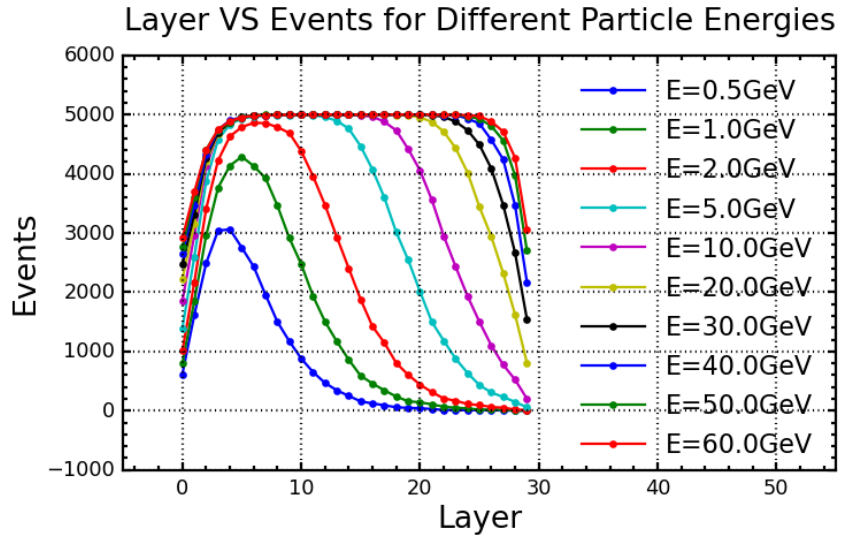
The cheat strategy shows the potential of ECAL, and the dependence of parameters a and b on the true particle energy by this method

Layer Weight Method

Shower development

- The layer 0 is set as the first hit layer, the shower start layer is the first layer with more than 2 hits
- The shower development is related to particle energy

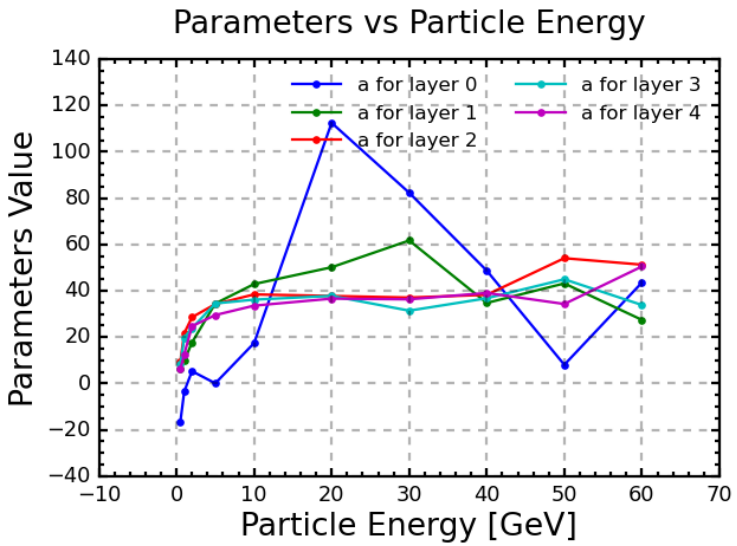


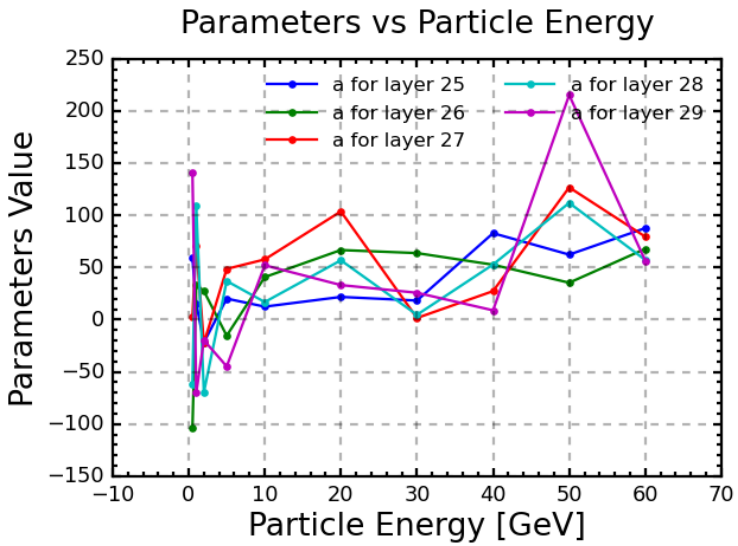


Shower patterns

Patterns of parameters

- Some parameters exhibit an exponential-like dependence on particle energy.
- Some parameters behave irregularly due to the lack of hits





Parameters patterns

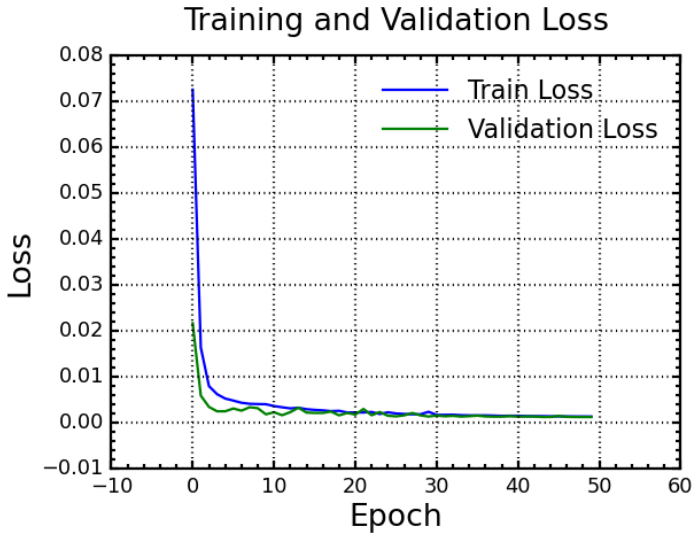
MultiLayer Perceptron

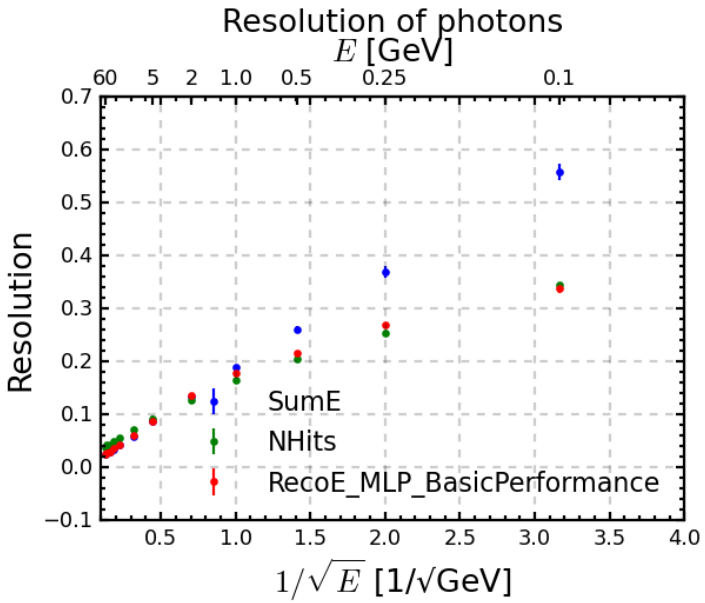
Samples

- Train & Validate
  - uniformed energy ranging 0-70 GeV
  - 1M for training, 0.2M for validate
- Test: energy points ranging 0.1,0.25,…..60 GeV
- Input Features: 152
  - $\mathrm{E_i}$ and $\mathrm{N_i}$ from each layer
  - $\mathrm{E_{sum}}$ and $\mathrm{N_{sum}}$
  - $\mathrm{E_i}/E_{\mathrm{sum}}$ and $\mathrm{N_i}/N_{sum}$
  - $\mathrm{E_i}/\mathrm{N_i}$

Model Setup

- Activation function: PReLU(Parametric Recitified Linear Unit)
- Loss function: Huber
- Layers: [64,32,16]
- Learning rate: warm up + cosine
- Batch size:4096
- Epochs: 50+ early stop(patience=10)





Training and performance

Loss function

MSE(Mean Square Error)

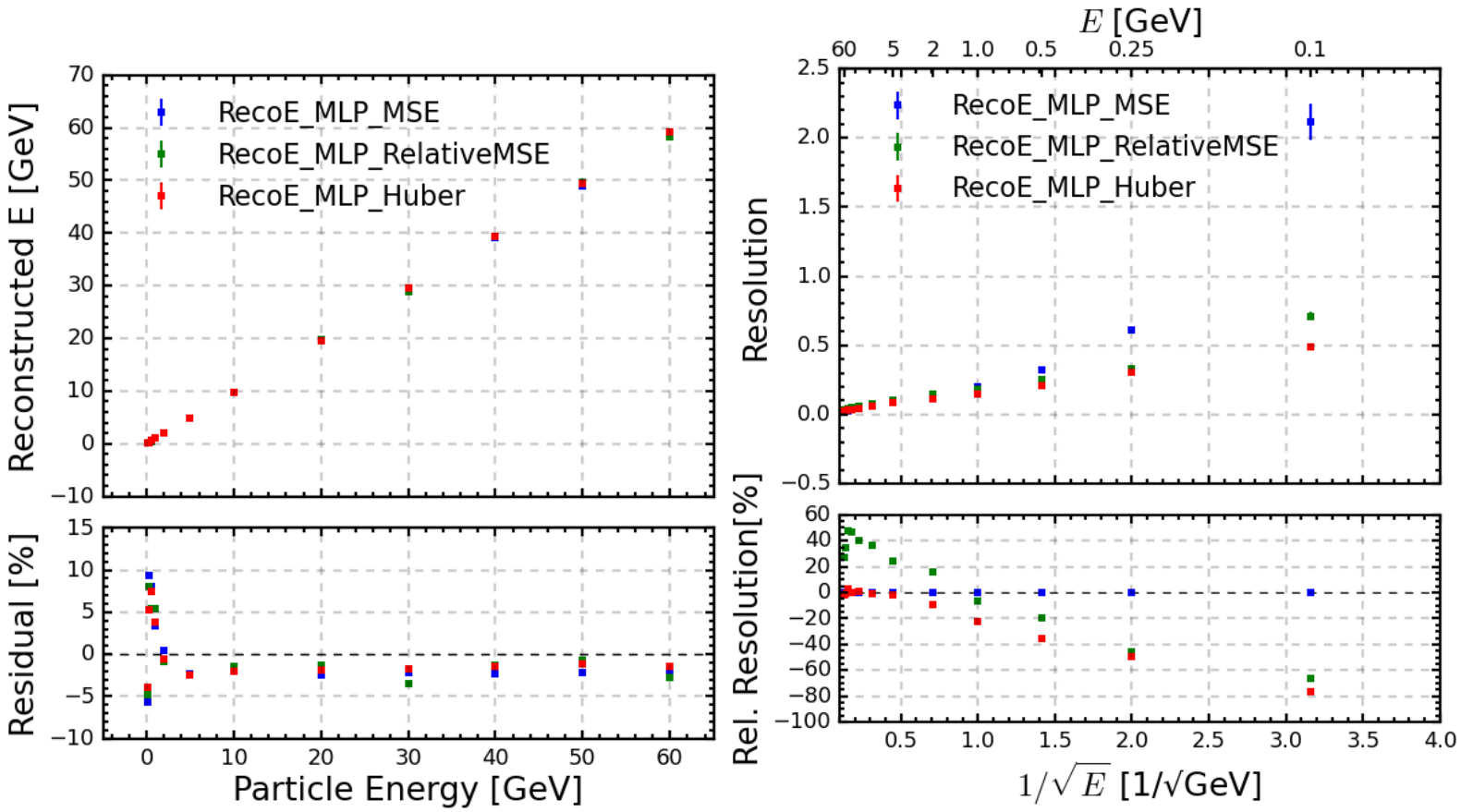
- $L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$

Relative MSE

- $L_{Relative\ MSE} = \frac{1}{N}\sum_{i=1}^{N}\left((\hat{y}_i - y_i)/y_i\right)^2$

Huber

- $r_i = \left|\frac{y^i_{pred} - y^i_{true}}{y^i_{true} + \varepsilon}\right|$
- $L = \frac{1}{N}\sum_{i=1}^{N}\begin{cases} 1/2\ r_i^2, r_i < 0.05 \\ 0.05 * (r_i - 0.5 * 0.05), r_i \geq 0.05 \end{cases}$



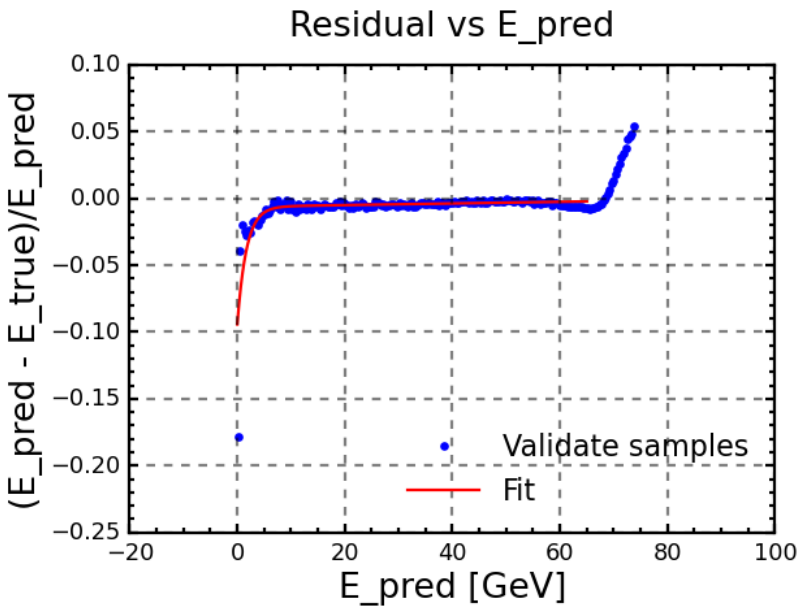
Performance of different loss functions

The resolution and linearity at low energy range should be stressed

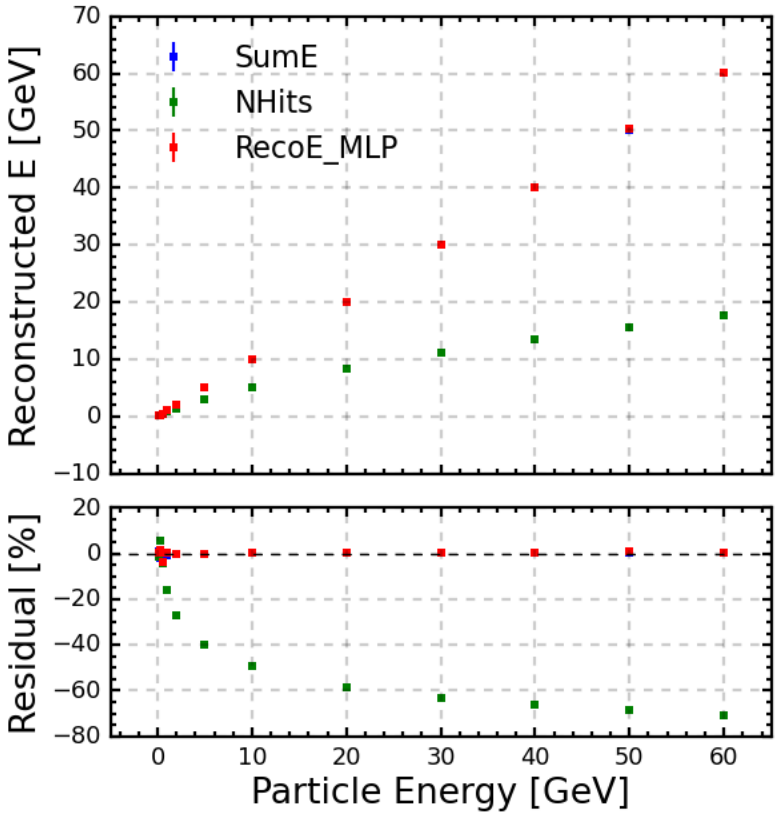
Calibration after MLP

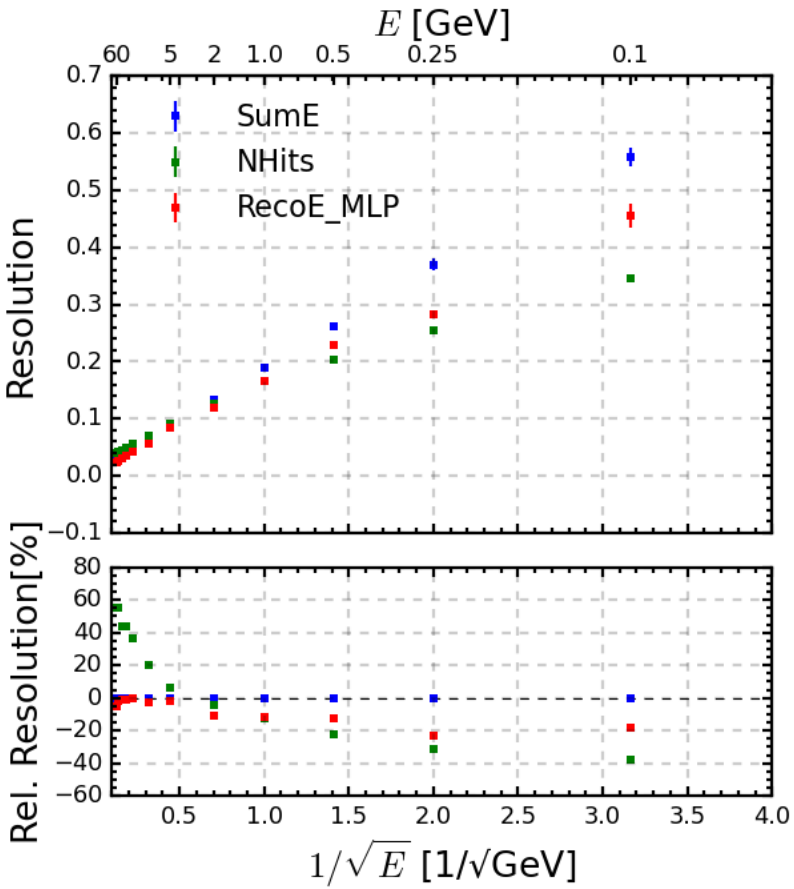
Calibration method

- $F_{fit} = a * e^{b*x+c} + d * x + e$
- Validate samples are used to fit this calibration function



- The MLP + Calibration result in good linearity and obvious improvements on resolution





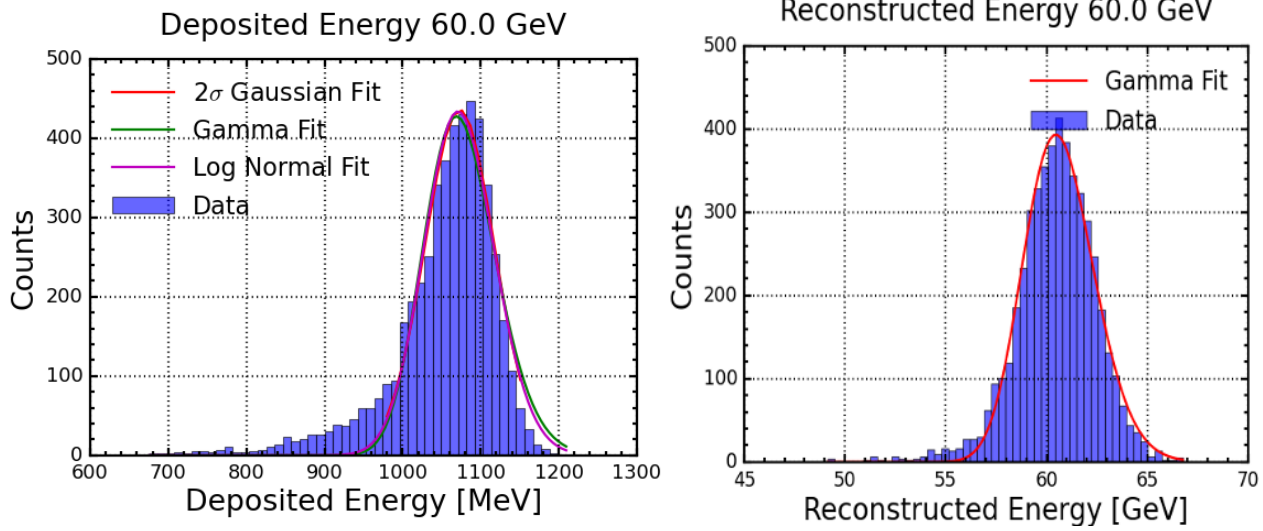
Energy leakage

40 ECAL layer geometry

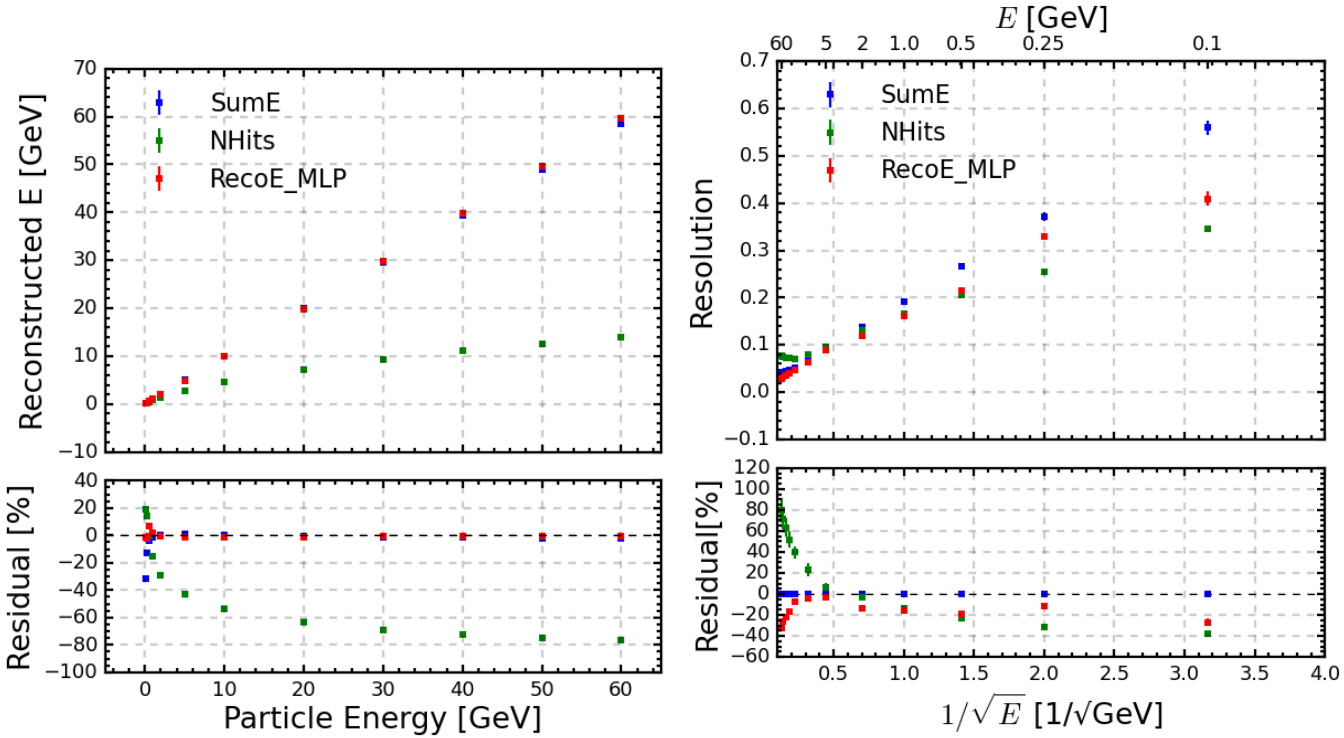
- First 40 in 80 layers were used
- Radiation length:~17 X_0
- obvious leakage could be noticed on energy distribution, linearity and resolution

MLP performance

- MLP could correct the enegy leakage to some extent, resulting in improved linearity and resolution



Deposited E and MLP reconstructed E

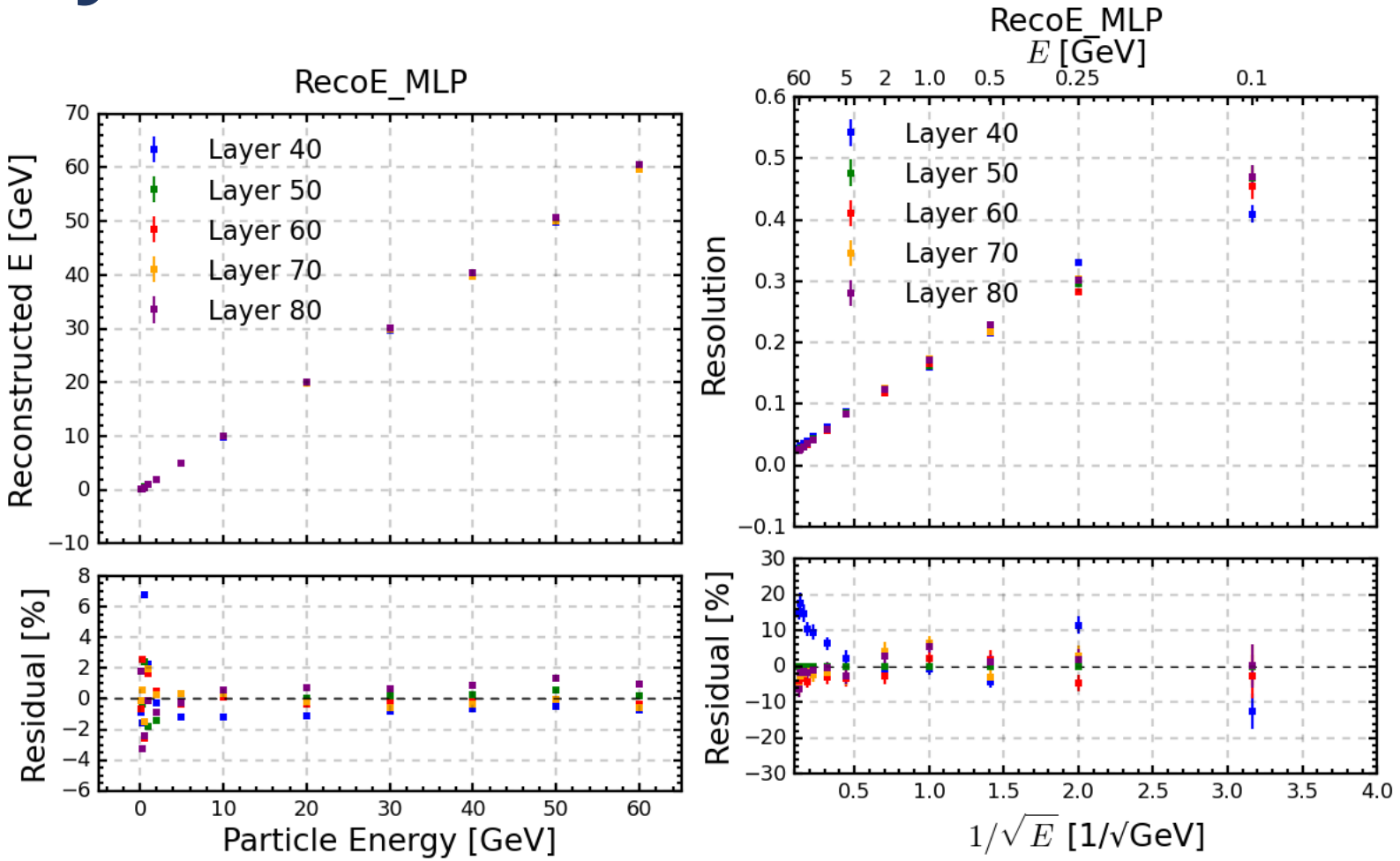


Linearity and Resolution

Number of Total Layers

MLP bias and improvement

- MLP models for different ECAL geometries exhibit varying performance at **low energies**, although in principle the information from these geometries should be equivalent.
- This bias may arise from the choice of **neuron numbers** and the effect of energy leakage.
- For all ECAL geometries with different layer numbers, the MLP improves their performance.

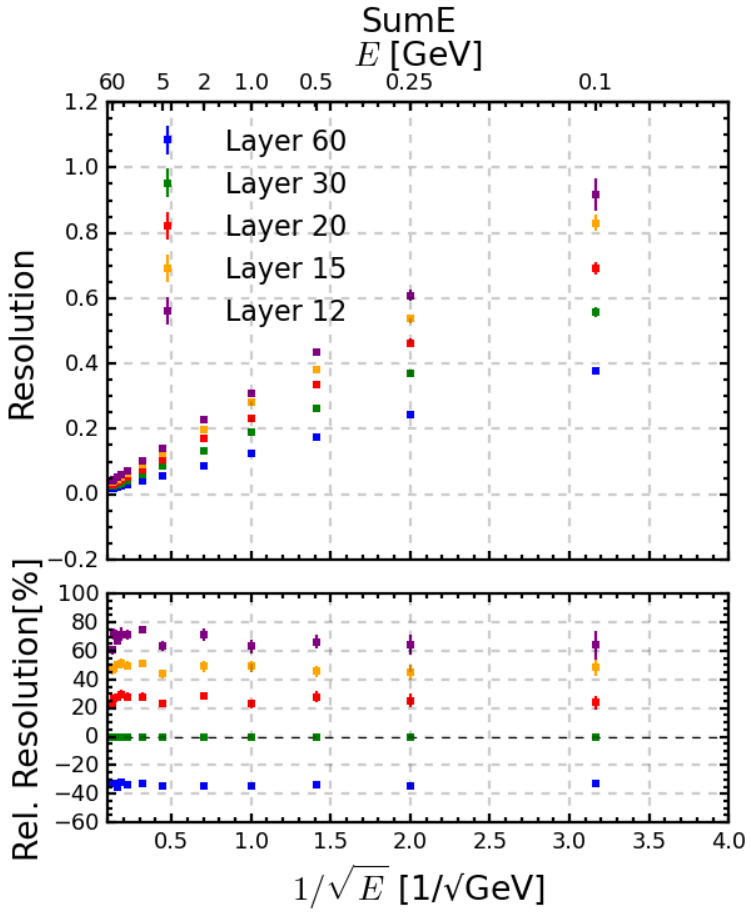


Linearity and Resolution

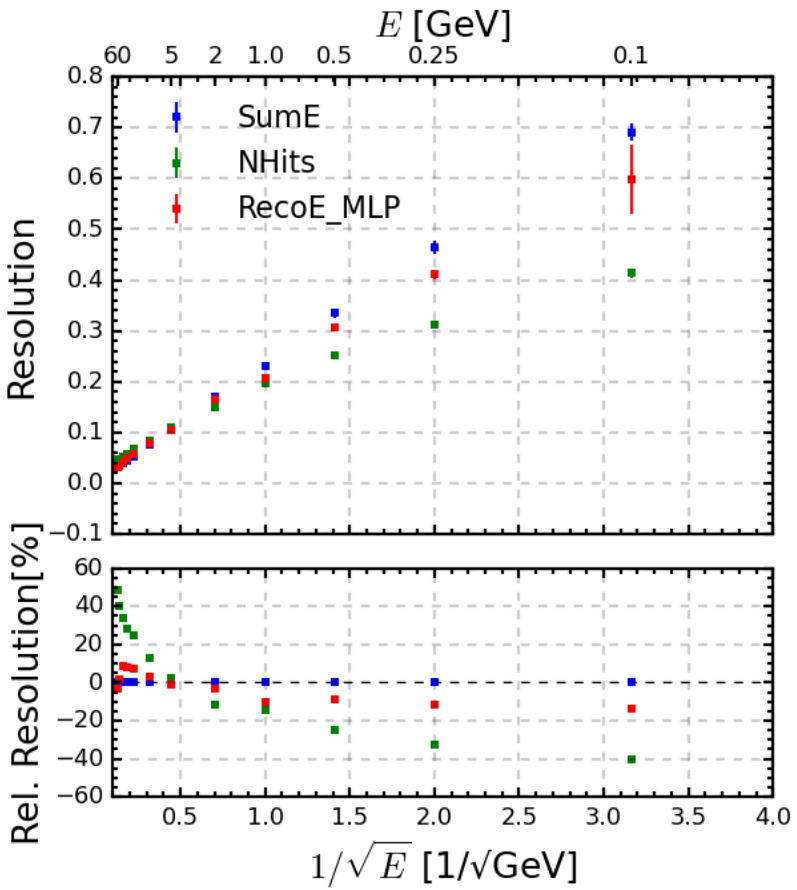
More than 40 ECAL layers were needed to suppress energy leakage

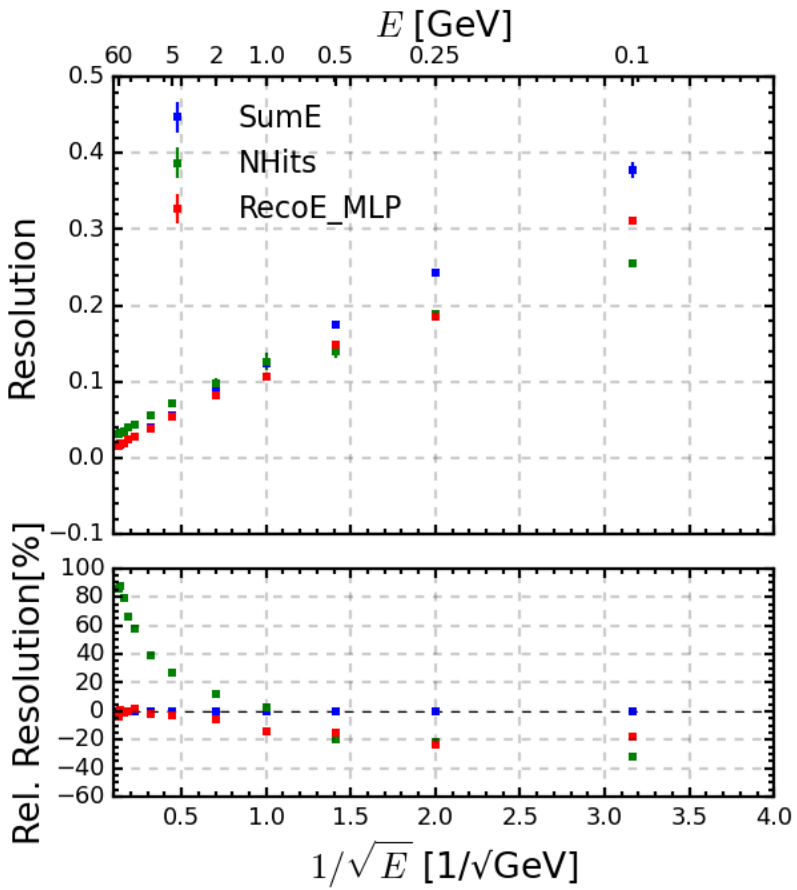
Number of Sampling Layers

- layer is read out in every 1, 2, …, 5 layers, corresponding to sampling layer numbers of 60, 30, 20, …, 12.
- The improvements achieved by the MLP are related to the number of sampling layers. MLP model is more powerful with more information provided



Resolution for SumE

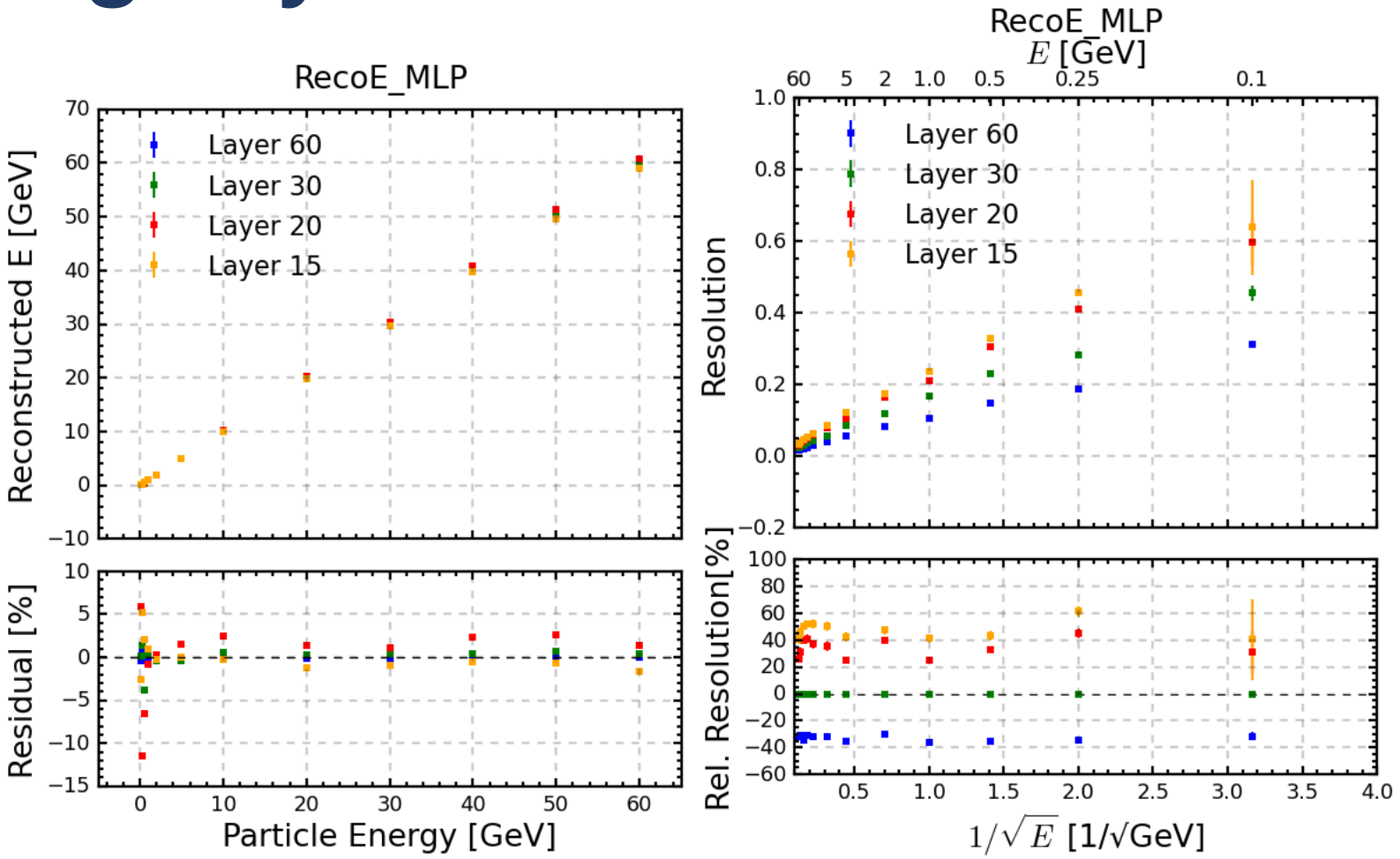




Resolution for sampling layers 20 and 60

Number of Sampling Layers

- The sampling frequency has a significant influence on ECAL performances
- This effect becomes more pronounced with the implementation of the MLP
- The difference should not be overestimated, considering that the configurations with 15, 20, 30, and 60 layers are not proportional.

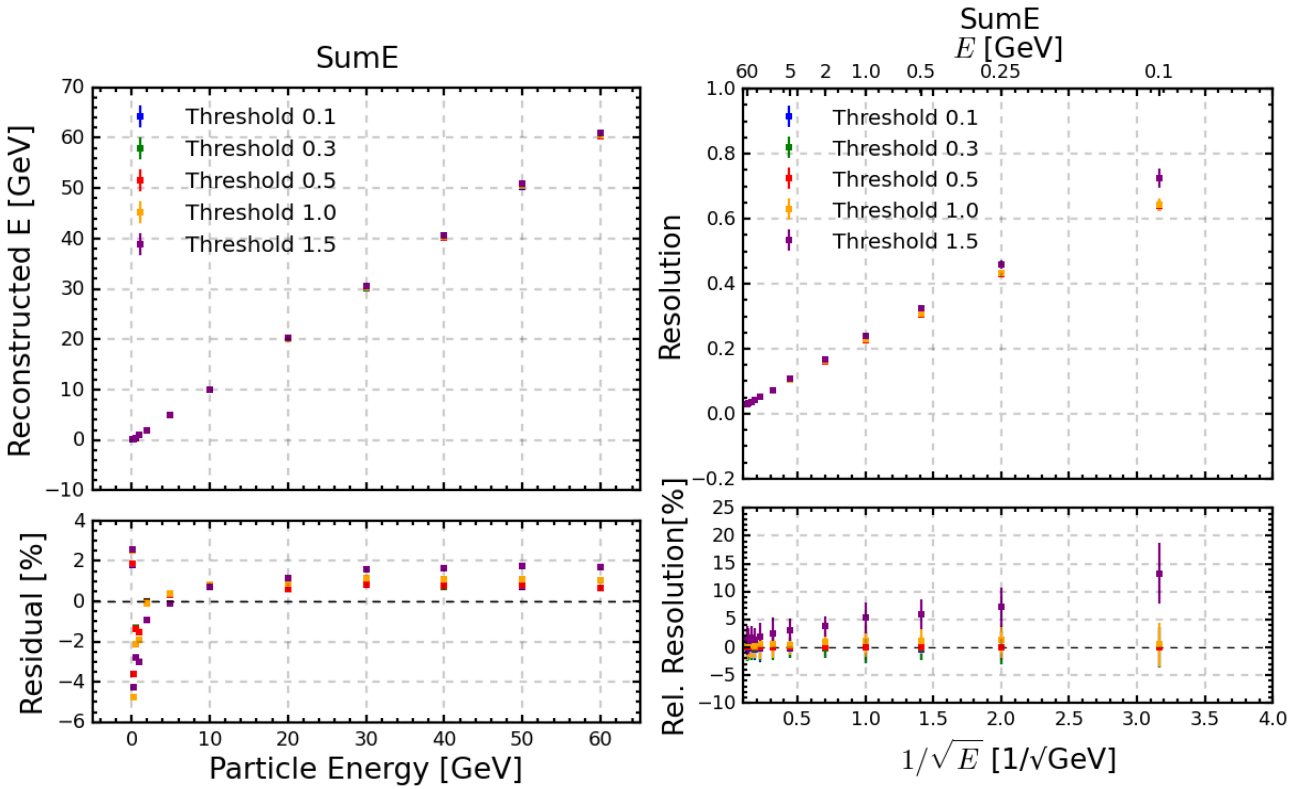


Linearity and Resolution

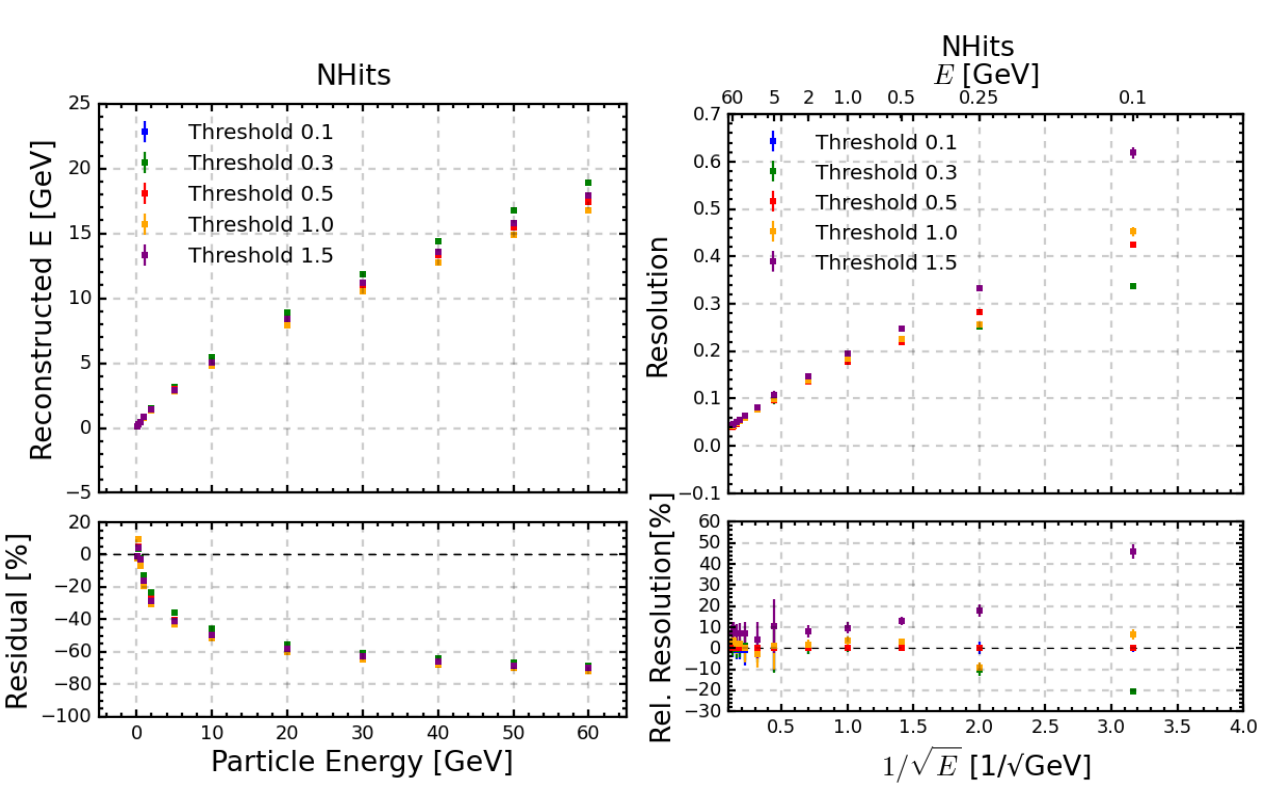
More than 20 layers are required to obtain stable benefits from the MLP, with 30 layers appearing to be a balanced choice — or 25?

Thresholds

- ECAL geometry: **0.15mm Si** with 30 sampling layers
- Sum of E: Lower thresholds result in better performance, but not significant
- Number of hits: high thresholds lead to degraded performance, with the impact being more significant than in the Sum of E method.



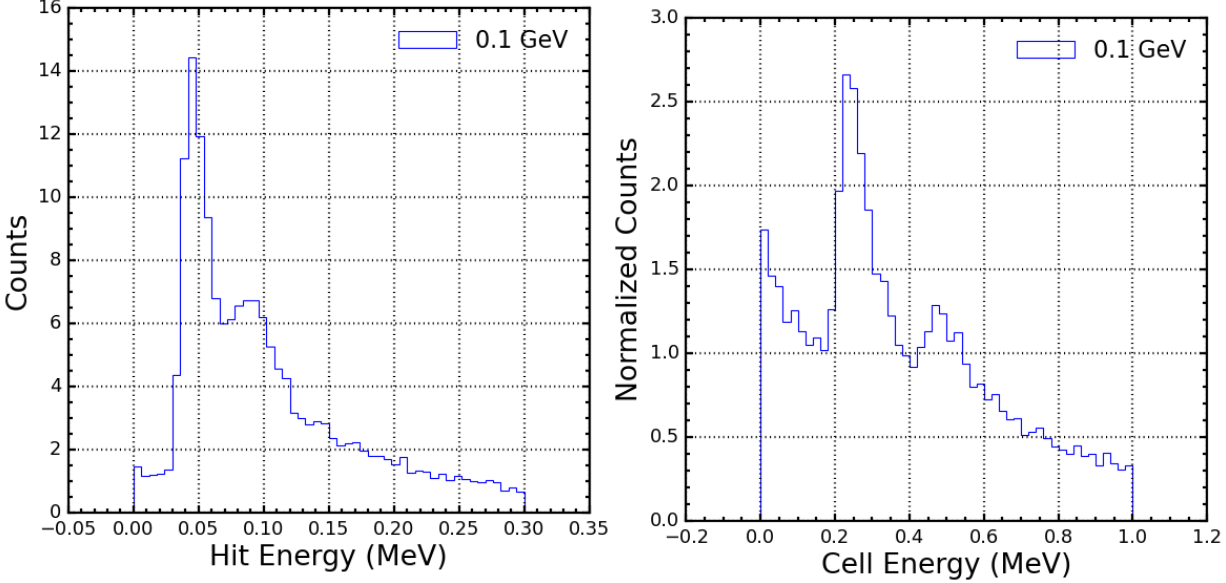
Performance for SumE



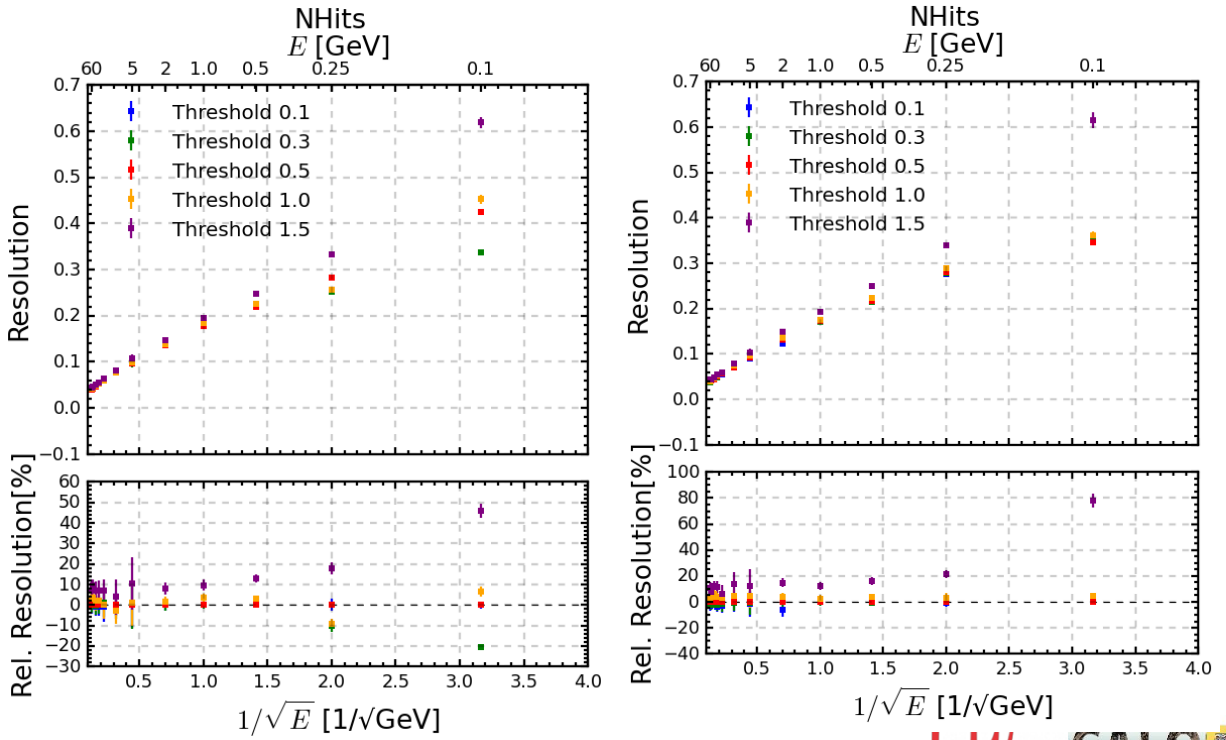
Performance for Nhits

Si thickness

- The hit energy distributions are related to the volume of the Si cells, thicker cells tend to have more hits with energy < 1MIP
- This difference in the hit energy distribution will cause the threshold to have different effects on Si layers of different thicknesses.
- As a result, a thicker Si ECAL is more sensitive to the threshold; that is, the resolution difference between high and low thresholds becomes larger.



Hit energy for 0.15mm Si and 0.75mm Si

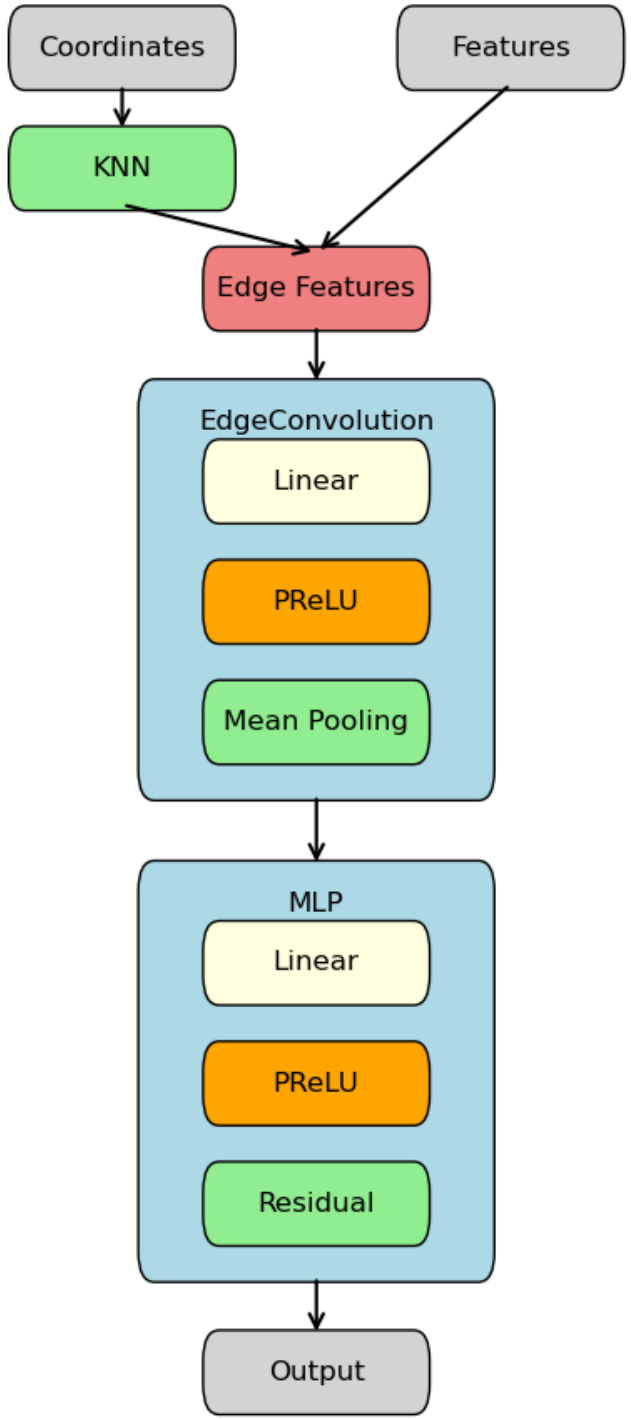


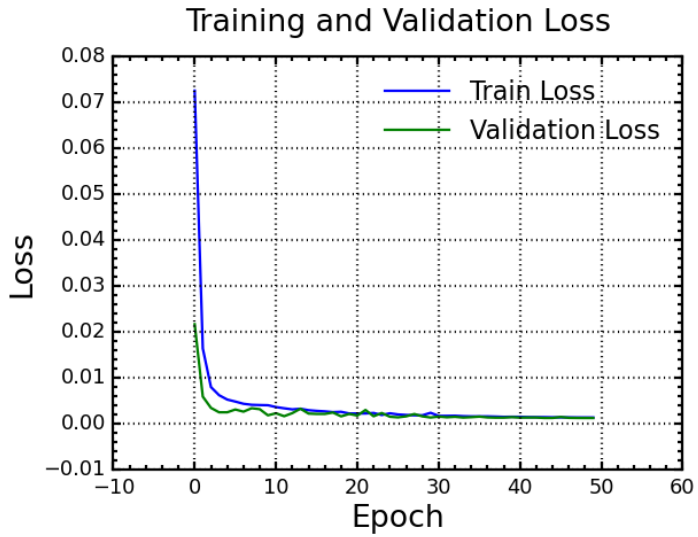
Resolution for Si 0.15mm and 0.75 mm

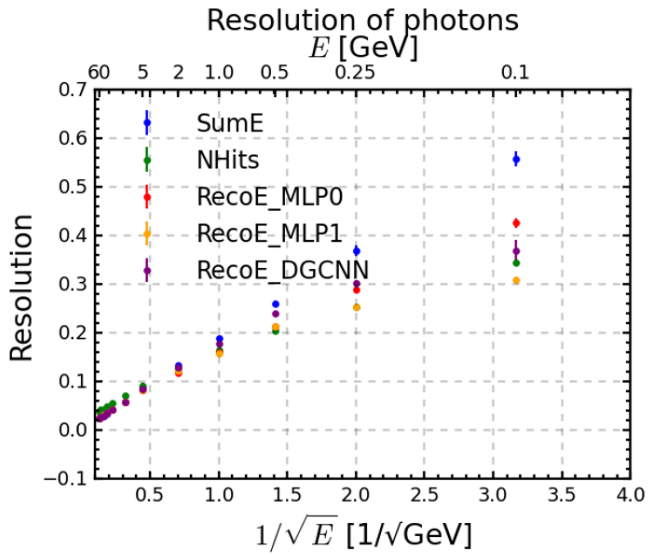
GNN based model

KNN+EdgeConvolution+MLP

- Construct edges $[x_i, x_j]$by connecting x_i with its k nearest neighbors, here i,j represents the layer idex
- For edge $[x_i, x_j]$, its features are $[E_i, N_i, Pos_i, E_j, N_j, Pos_j - Pos_i]$
- These edge features will pass through a neural network [32, 64] and then be aggregated node-wise, followed by mean pooling
- An MLP is used in the end







Training and performance

Summary and plan

Summary

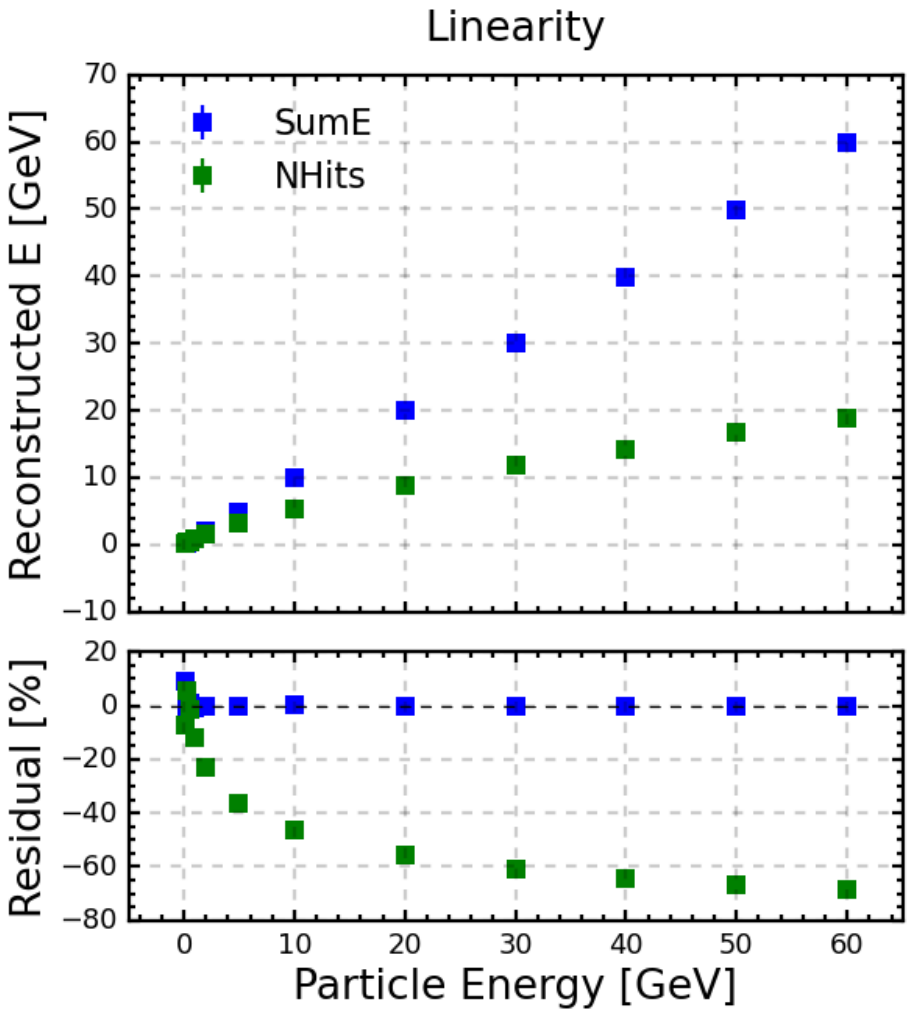
- A MLP Model was fine tuned to optimize the ECAL geometry for future circular collider
- The Total material, number of sampling layers, thresholds and thickness have been scanned

Plan

- The Si cell size would be scanned and an optimized ECAL geometry would be proposed
- Advanced model like DGCNN would be tried to evaluate the ultimate potential of this new re-optimized ECAL
- Time digitization would be implemented and time resolution would be studied in terms of different configurations
- Implement energy regression method and time digitization in Full reconstruction

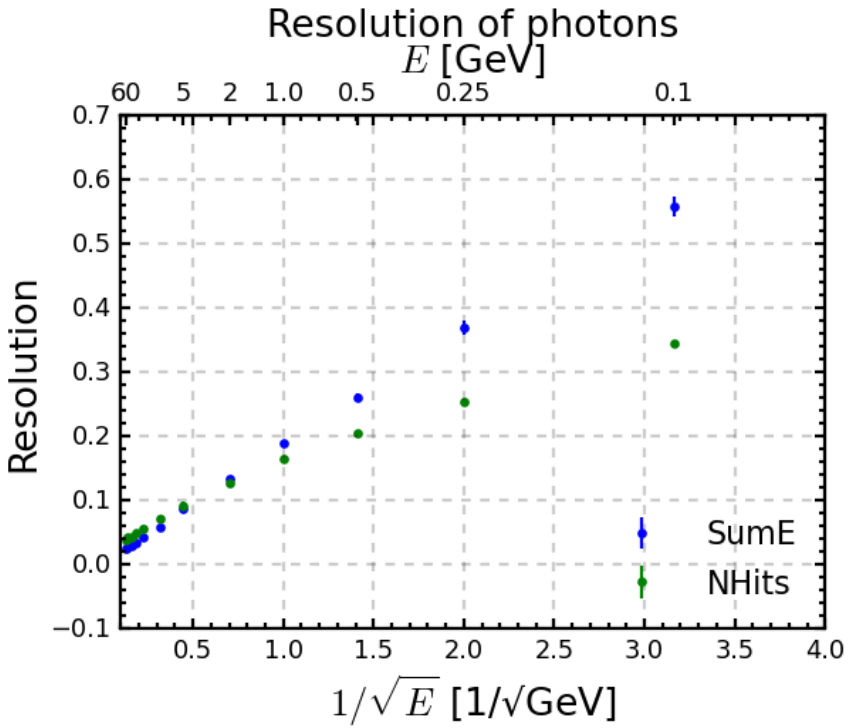
Backup

Motivation



Linearity for gamma

- The number of hits exhibit a much **better** resolution at energy range(< 5 GeV), but it will saturate soon at high energy range



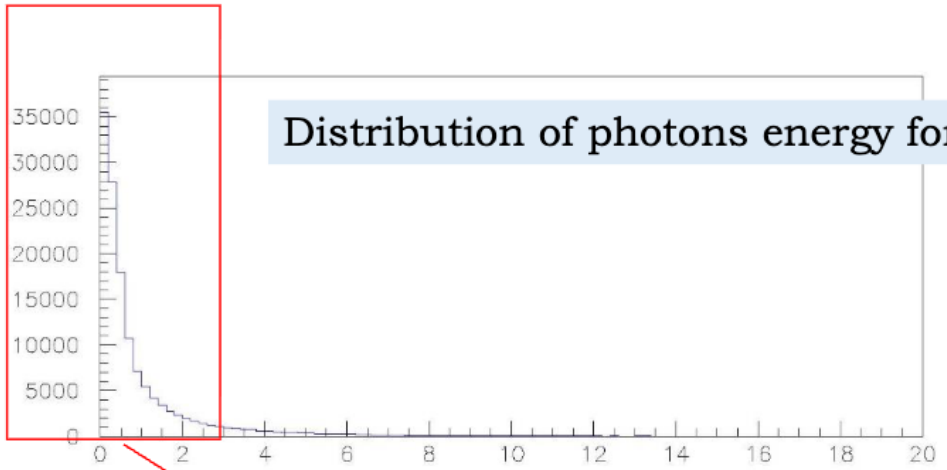
Resolution for gamma

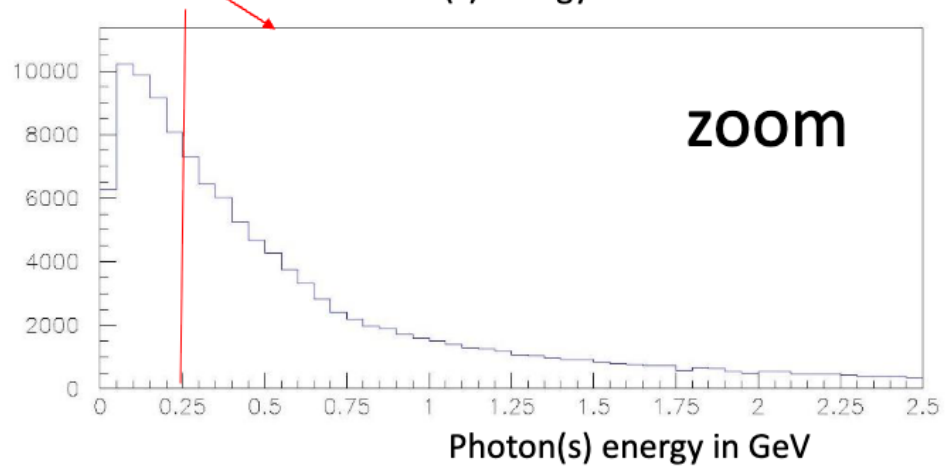
Based on the fact that the number of hits is determined by the transverse dimension, the current phenomenon should be interpreted that the potential of a high-granularity calorimeter has not yet been fully explored.

Motivation

Jean Claude's report on DRD6 September 2025
https://indico.cern.ch/event/1551941/contributions/6656570/attachments/3138669/5569958

Why low energy is so important





E cms = 91.2 GeV

$Z \rightarrow b\ \bar{b} \rightarrow B_s + X$

$B_s \rightarrow D_s \pi^{\pm}$ and $D_s \rightarrow \phi\rho$

- $\phi \rightarrow K^+K^-$
- $\rho \rightarrow \pi^{\pm}\pi^{\circ}$
- $\pi^{\circ} \rightarrow \gamma\gamma$

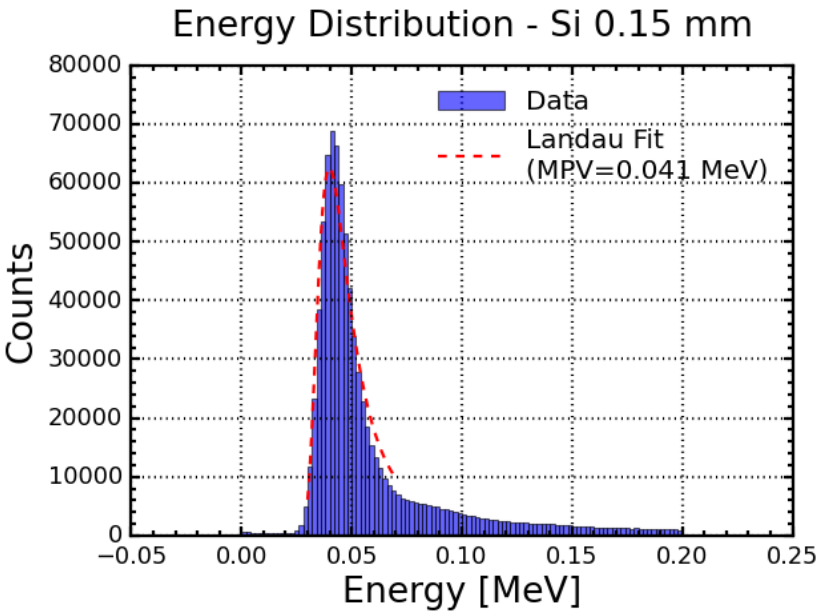
Final state $k^+k^-\pi^+\pi^-\ \gamma\gamma$

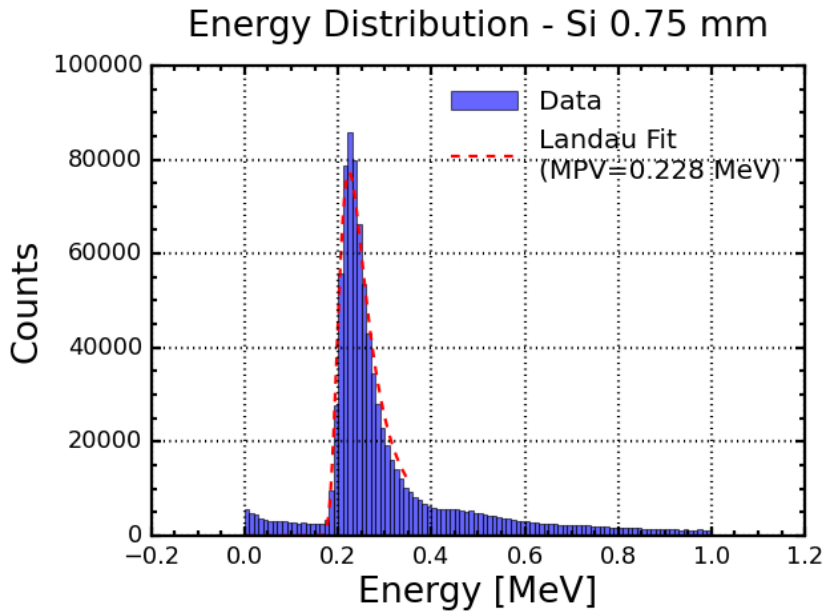
Recall :
The threshold in ALEPH was at 250 MeV

MIP Calibration

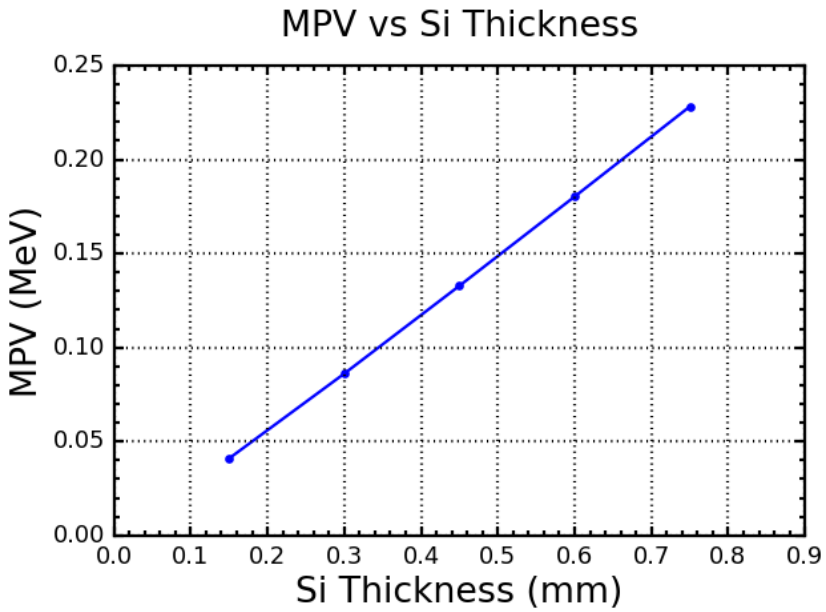
- 100 GeV muon were simulated for this calibration
- Landau fitting: the MPV value will be used as the unit of MIP

0.5 MIP threshold would be implemented in the analysis





Energy deposition in different thickness of Si, and their laudau fittings



Landau MPV vs Si thickness

Fitting Functions

Gaussian

- $f(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Mean : μ
- Peak:μ
- Variance: σ^2
- Resolution: $\sigma/\ \mu$
- 2σ gaussian is to fit gaussian in $\mu \pm 2\sigma$

Gamma

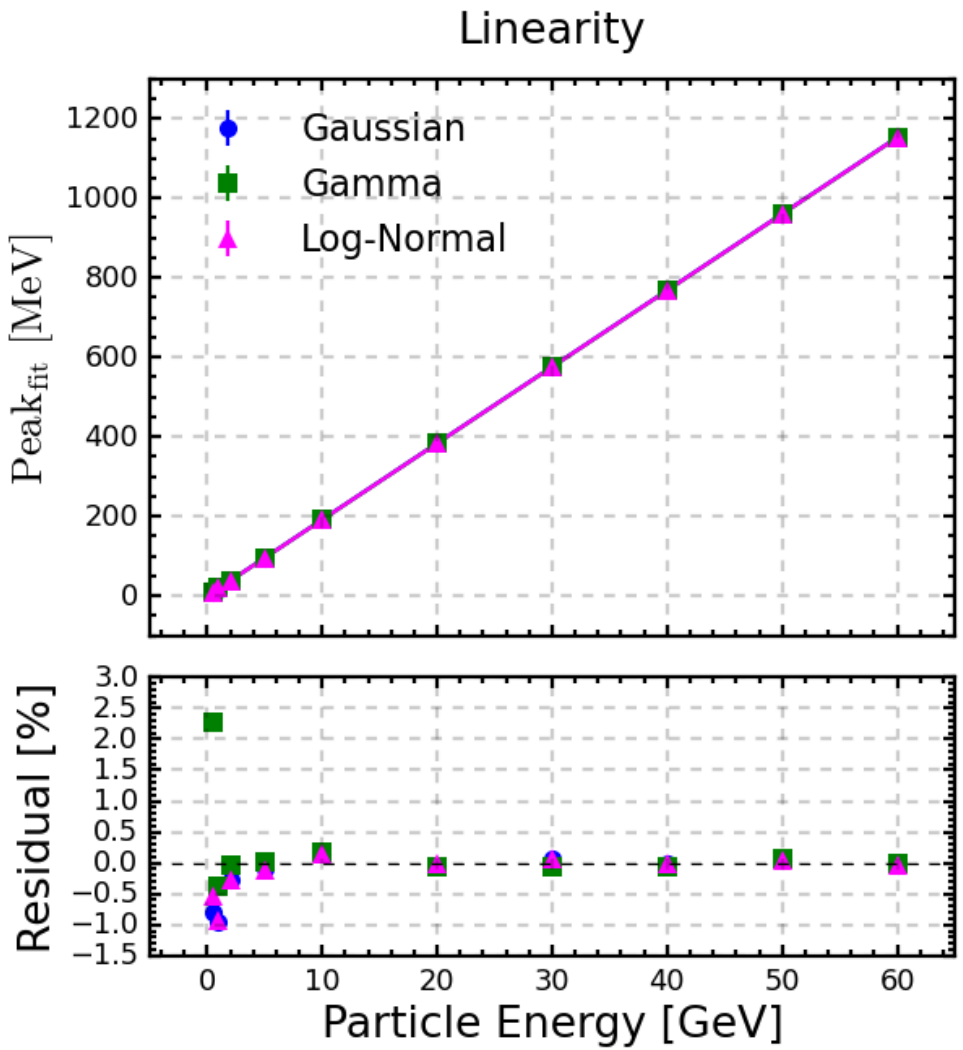
- $f(x) = (x-\mu)^{k-1} \cdot e^{-\frac{x-\mu}{\theta}} / \theta^k \cdot \Gamma(k)$
- Mean :$k\theta + \mu$
- Peak: $(k-1)\theta + \mu$
- Variance: $k\theta^{\ 2}$
- Resolution: $\sqrt{Varianve}\ /mean$

Log-Normal

- $f(x) = \frac{1}{\sqrt{2\pi}\cdot x\sigma} \cdot e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$
- Mean :$e^{\mu+\sigma^2/2}$
- Peak: $e^{\mu-\sigma^2}$
- Variance: $(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$
- Resolution: $\sqrt{Varianve}\ /mean$

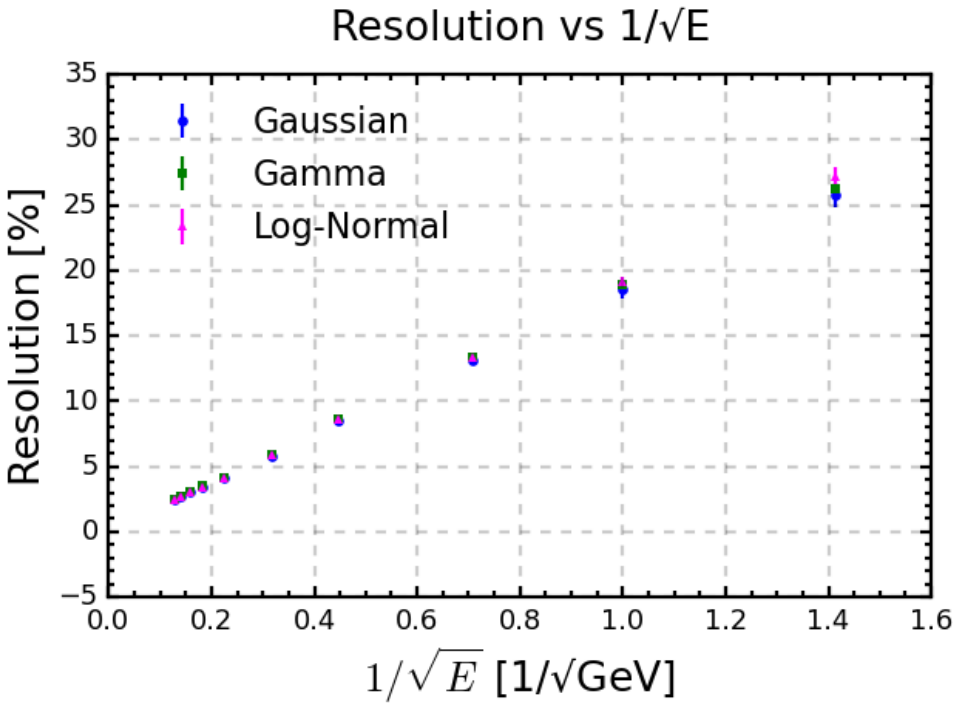
The gamma fitting has the best performance at low energy range. And at high energy range, the gamma fitting and gaussian fitting have the same result. This study will use gamma fitting for the linearity and resolution

Linearity and resolution



Linearity for gamma

- Different fitting functions derive different linearities and resolutions at low energy range, approximately below 5 GeV, but converge beyond this range



Resolution for gamma

Gamma function will be used for fitting in this study

ML methods

CNN

- Require fixed grid input and fixed spatial order
- calo hits would be highly redundant in CNN input

GNN/Point Net

- Graph input: nodes and edges
- Point cloud: a set of points
- Input data doesn’t have order

Attention mechanism

- Point transformer V3, CVPR 2024

Existing works

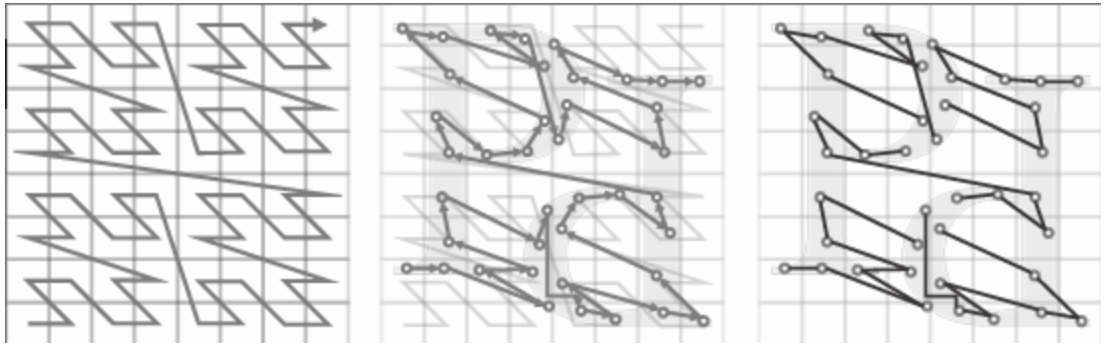
- 3d CNN, ICCV 2019 workshop
- GNN, ZDC for future EIC neutron, 2025 NIMA
- DeepSC, CMS photon, 2023 NIMA
  - GNN + attention module
- GNN & CNN, ATLAS pion, 2022 Note
- KNN, CALICE AHCAL pion, 2024 JINST

Point transformer 3

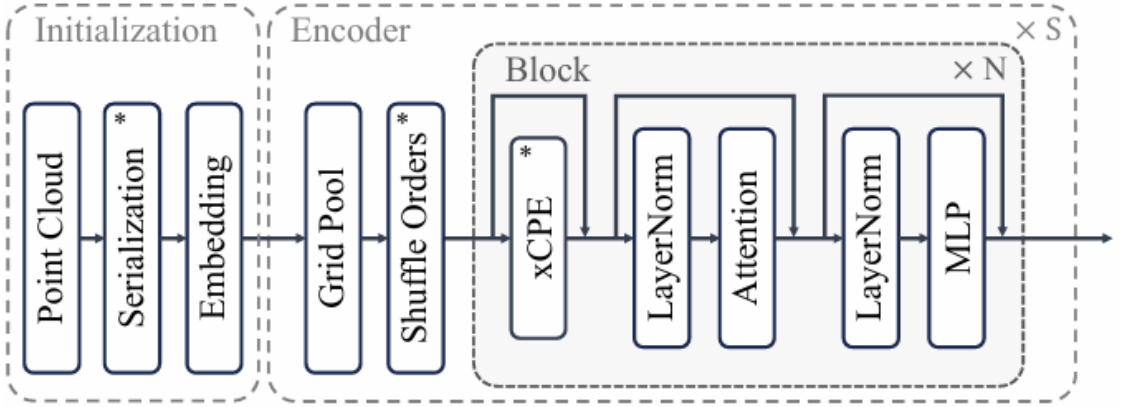
[Point Transformer V3: Simpler Faster Stronger]
This slide refer

Serialized point cloud

- Point cloud is a set of points without sequence
- Point transformer 3 use serialized point cloud, this feature could benefit 5d calo hits, especially the **time**?



Point cloud serialization

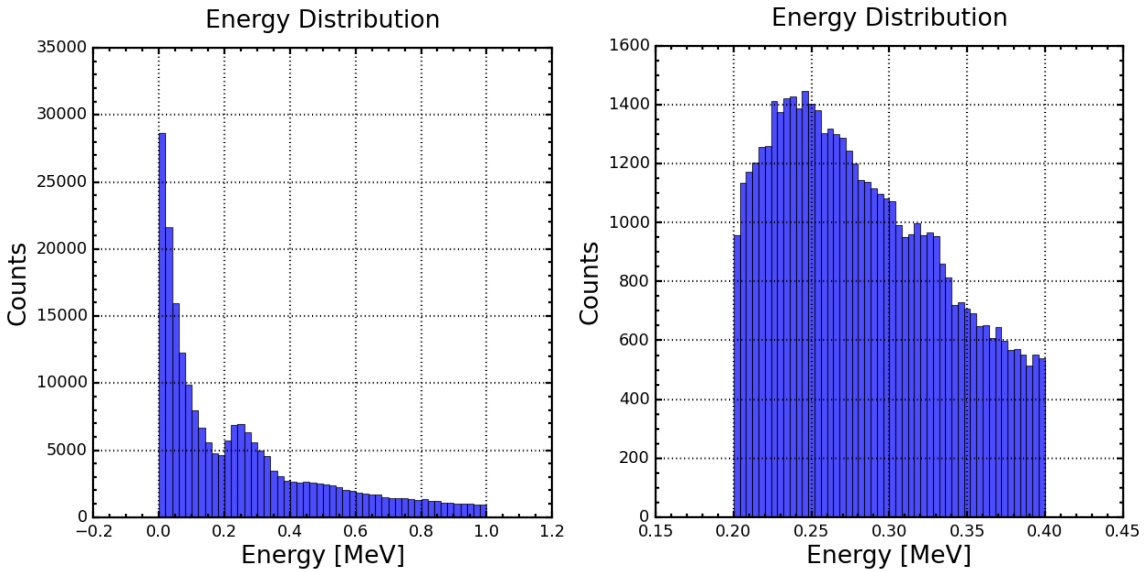


Architecture of Point transformer3

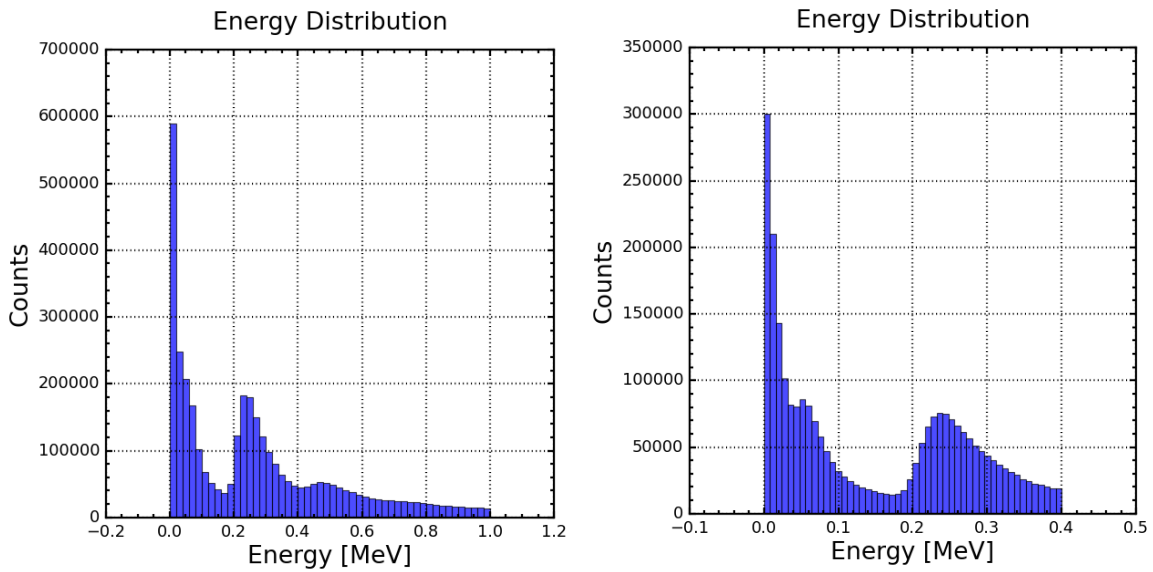
Serialized attention

In general the gamma fit has the best performance

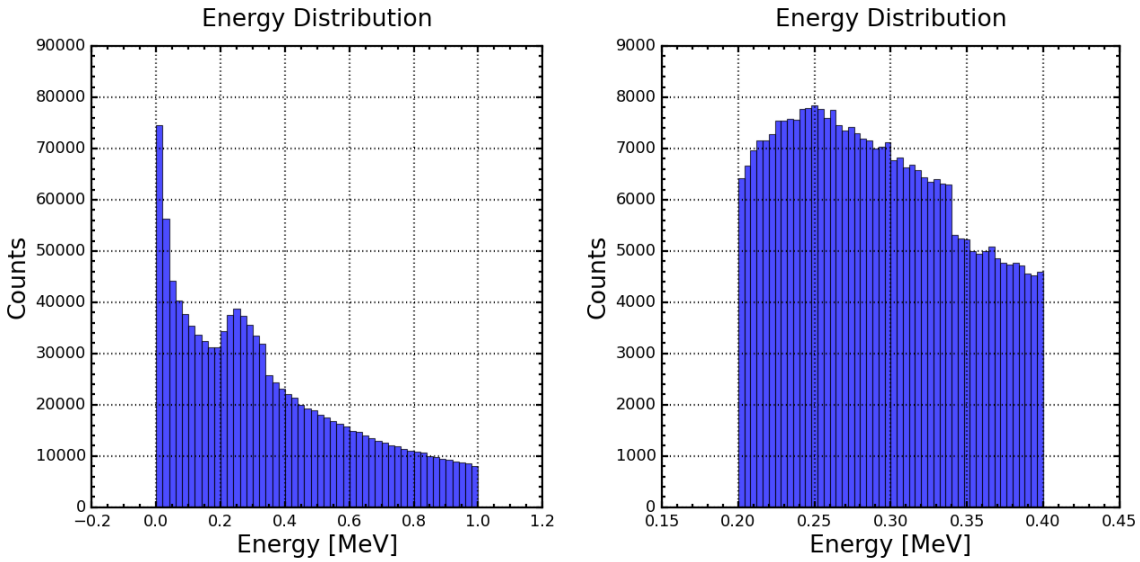
Abnormal hit E



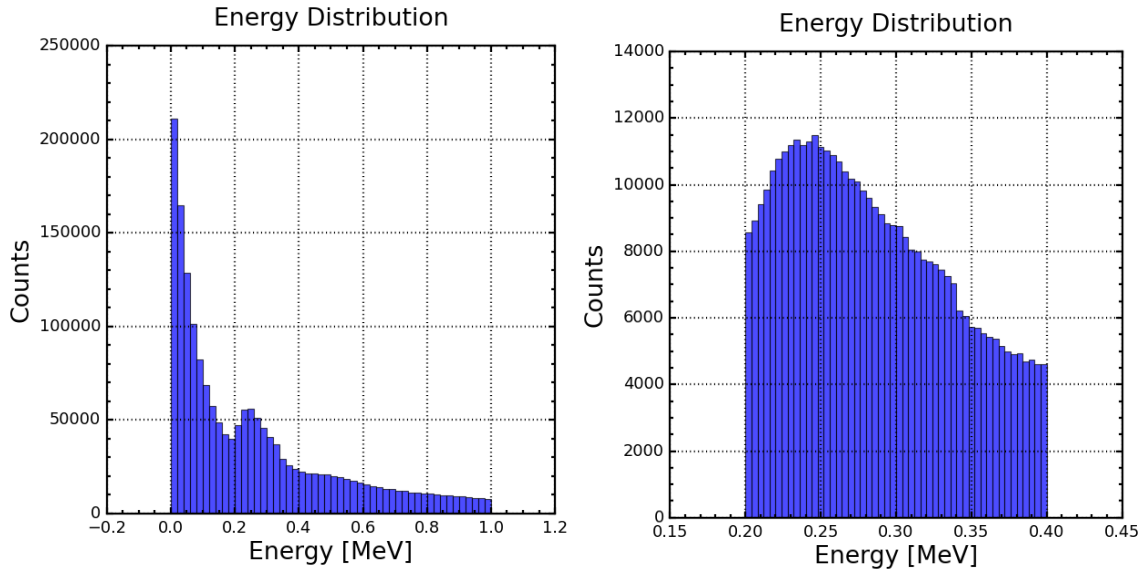
Fe:5 mm



Al: 15mm

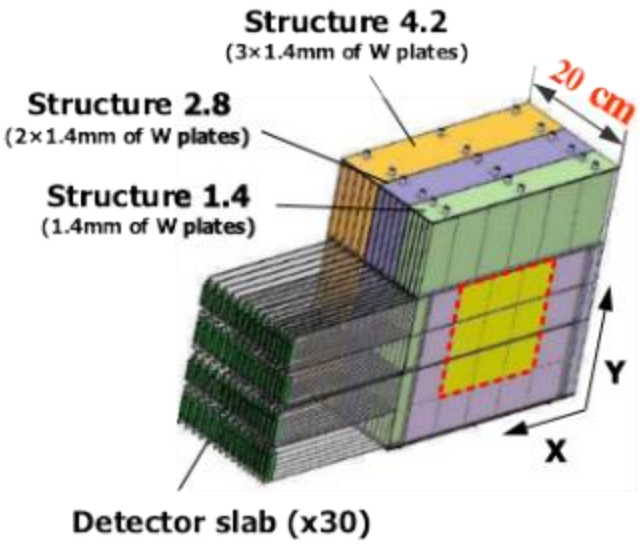


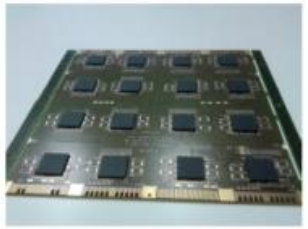
Tungsten: 1.5 mm



Cu:5 mm

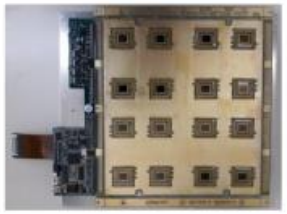
SiW ECAL





FEV10, 11, 12

- BGA packaging
- Incremental modifications
- From v10 -> v12
- Main "Working horses" since 2014

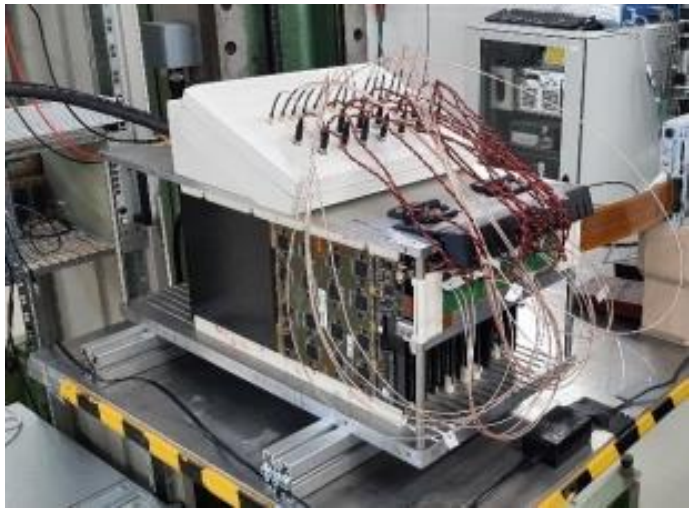


FEV-COB

- Chip-On-Board : ASICs wirebonded in cavities
  - Thinner than FEV with BGA
- Based on FEV11
  - External connectivity compatible

FEV13

- BGA packaging
  - Improved routing
  - Local power storage
  - Different external connectivity



Physical(2005-11)

- 1×1 cm² on 500µm 6×6 cm² Pad glued on PCB Floating GR
- 30 layers (10k chan).
- External readout
- Proof of principe
- SKIROC2 chips

Technological (now)

- Embedded electronics
  - Power-Pulsed, Auto-Trig, delayed RO
  - **S/N = (MPV/σNoise) ≥ ~12 (trig)**
- Compatible w/ 8+ modules-slab
- 0.5×0.5 cm² on **320–650µm** 9×9 cm²
- 26–30 layers, 8k (slab) ~ 30k (calo) channels

Full Detector(future)

- 1M ➞ 70M channels
- on **725µm** 12×12 cm² 8'' Wafers ?
- Pre-industrial building
- Full integration (⊃ cooling)
- **Final ASIC**

Digitization

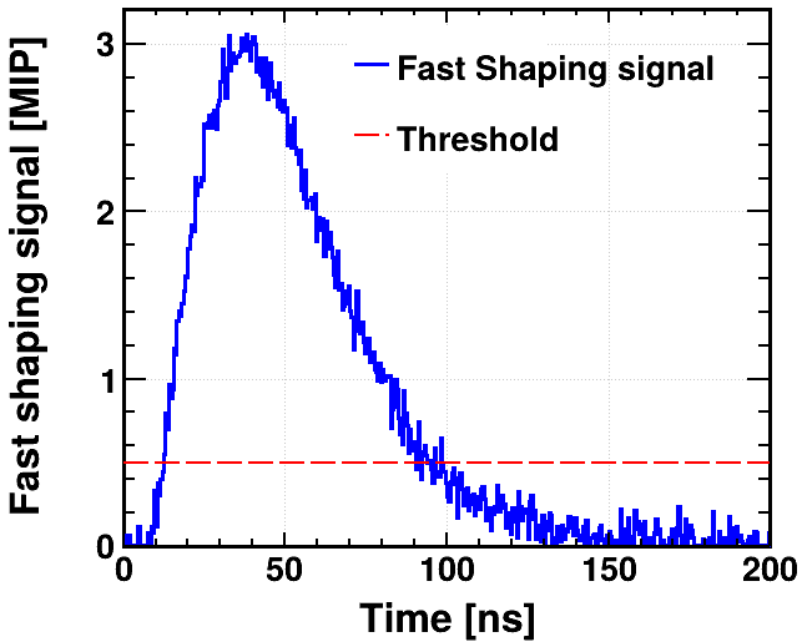
Shaping signal function

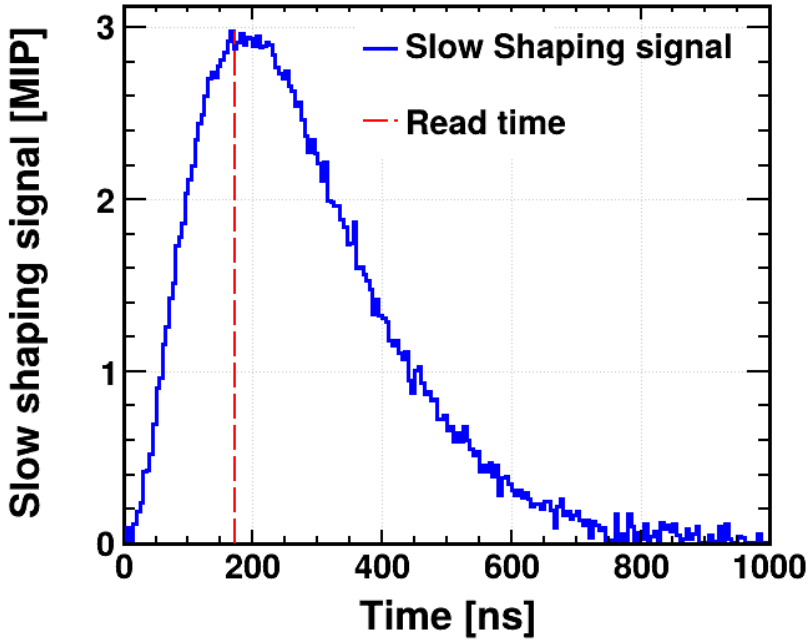
- $S(t) = \sum_i^{subhits} f_{scale} \frac{E_i \cdot T_i^n(t) \cdot e^{-T_i(t)}}{n!} \otimes gauss(0, noise)$
- f_{scale}=4
- $T_i(t) = (t - t_i^{hit})/\tau$
  - fast shaping:$\tau_{\mathrm{fast}} = 30$ ns
  - slow shaping:$\tau_{\mathrm{slow}} = 180$ ns

- Order of CR_RC filter
  - $\mathrm{n_{fast}} = 2$
  - $\mathrm{n_{slow}} = 2$
- Noise
  - Fast: 1/12 MIP
  - Slow: 0.05 MIP

Digi Hit

- t_{Digi}: the time when $S_{fast}(t_{Digi}) = 0.5\mathrm{MIP}$
- Energy: Slow shaping signal after delay $S_{slow}(t_{Digi} + t_{delay})$
- delay time: 160ns

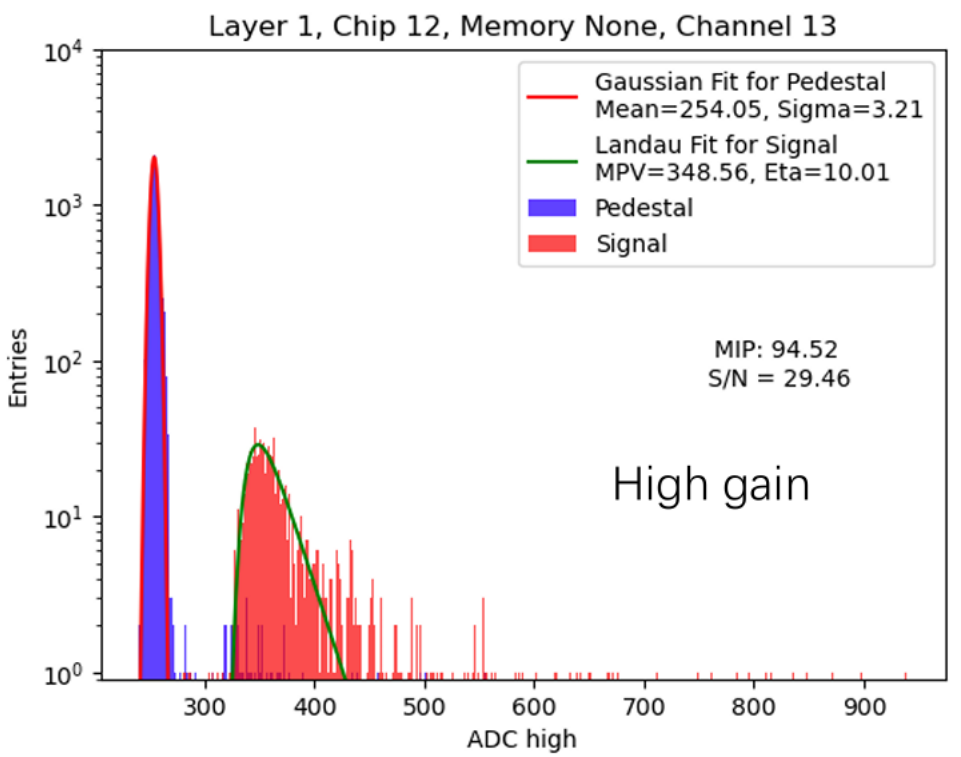


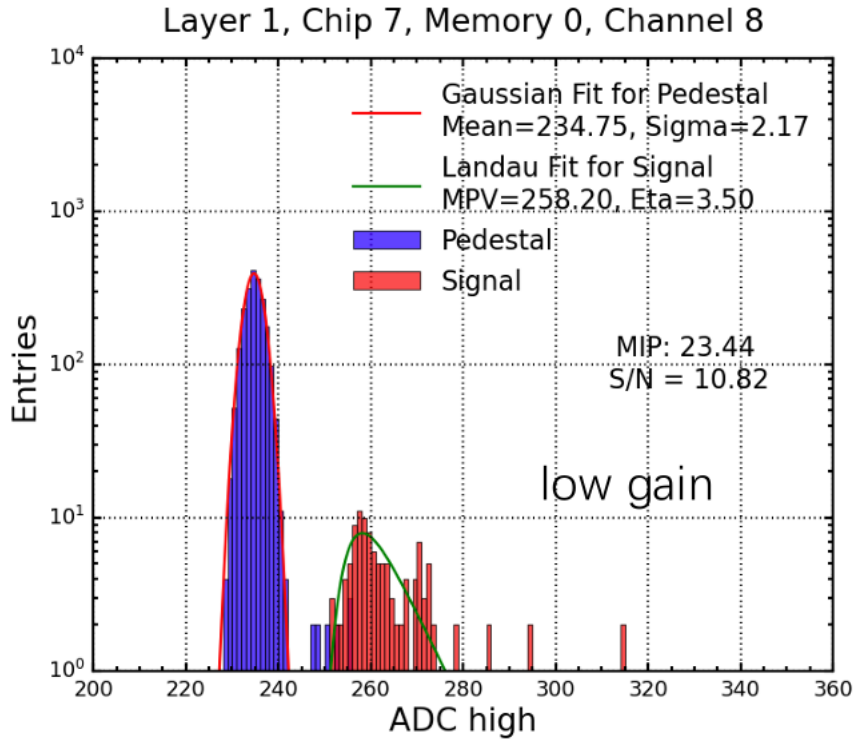


Digitization for Fast shaping(left) and slow shaping(right)

Signal Noise Ratio(S/N)

- Pedestal is obtained by the beam data with hitbit=0
- High gain: MPV ~90 ADC(after pedestal extraction), S/N ~ 30
- Low gain: MPV~ 23 ADC, S/N~ 11





Single channel's S/N for high gain and low gain

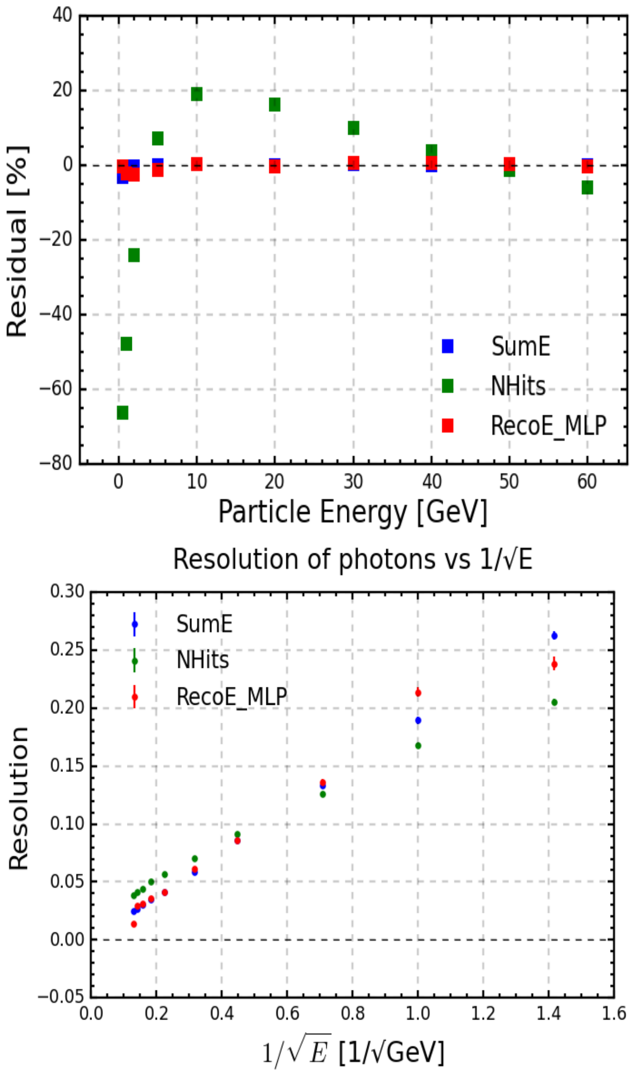
Activation function

- RELU: cant work on x<0
- Sigmoid: can solve non linearity, but not powerful enough
- Tanh: somehow break
- SELU: fit the case for energy, better add parameter b to exp(bx)
- Swish:failed

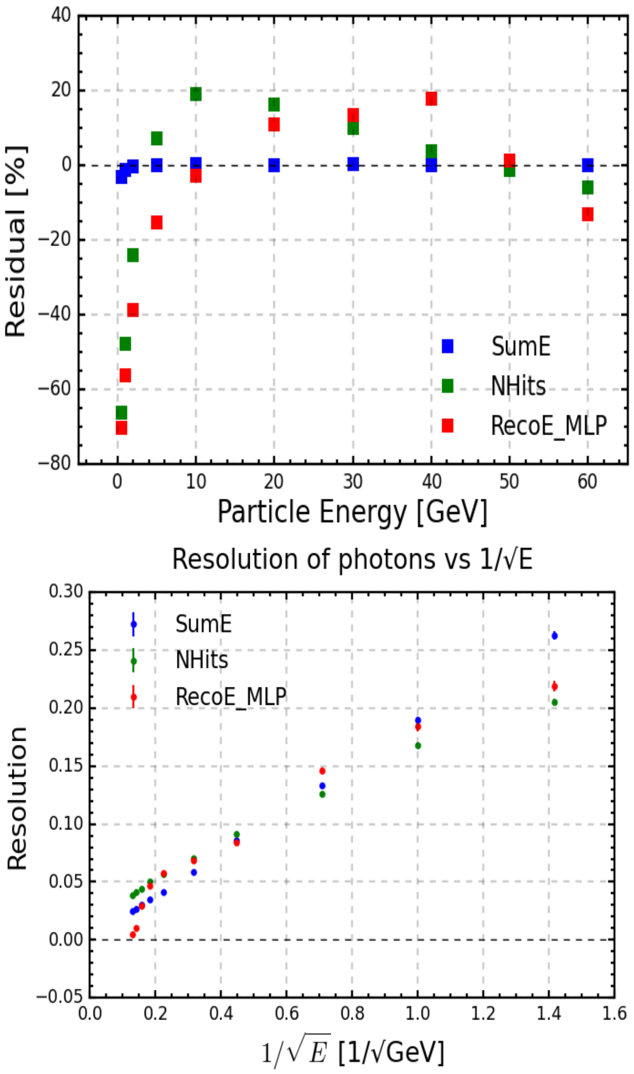
- **Sigmoid**: $f(x) = \frac{1}{1+e^{-x}}$
- **Tanh**: $f(x) = \tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$
- **SELU**: $f(x) = \lambda \begin{cases} x, & x > 0 \\ \alpha(e^{x}-1), & x \leq 0 \end{cases}$
- **Swish**: $f(x) = x \cdot \frac{1}{1+e^{-\beta x}}$

Function	Full Name	Definition
ReLU	Rectified Linear Unit	$\mathrm{ReLU}(x) = \max(0, x)$
LeakyReLU	Leaky Rectified Linear Unit	$\mathrm{LeakyReLU}(x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases}$
PReLU	Parametric Rectified Linear Unit	$\mathrm{PReLU}(x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases}$

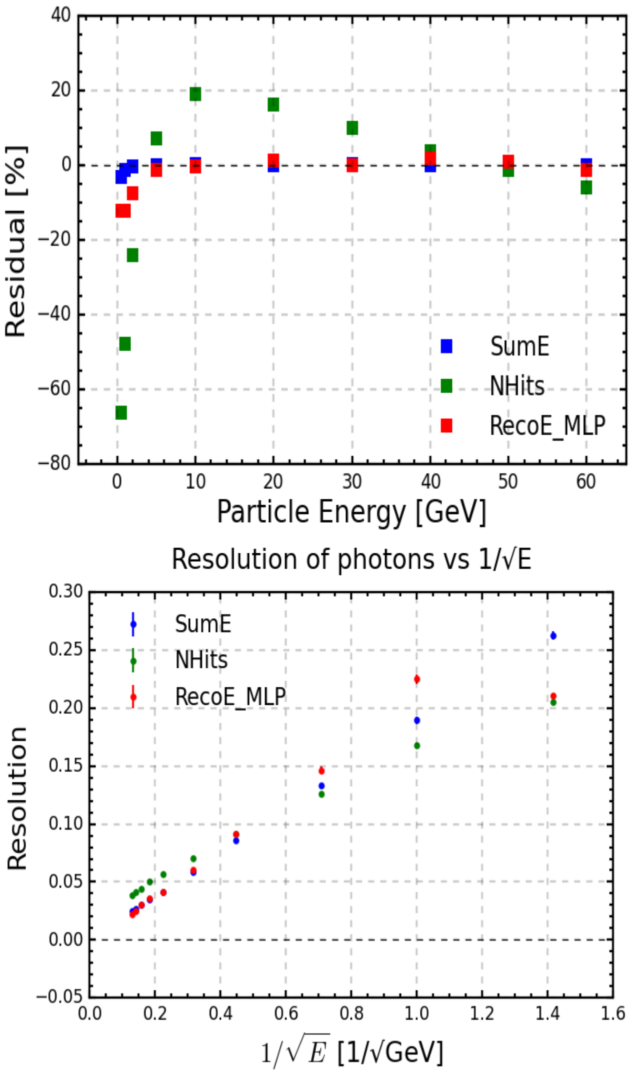
Activation function



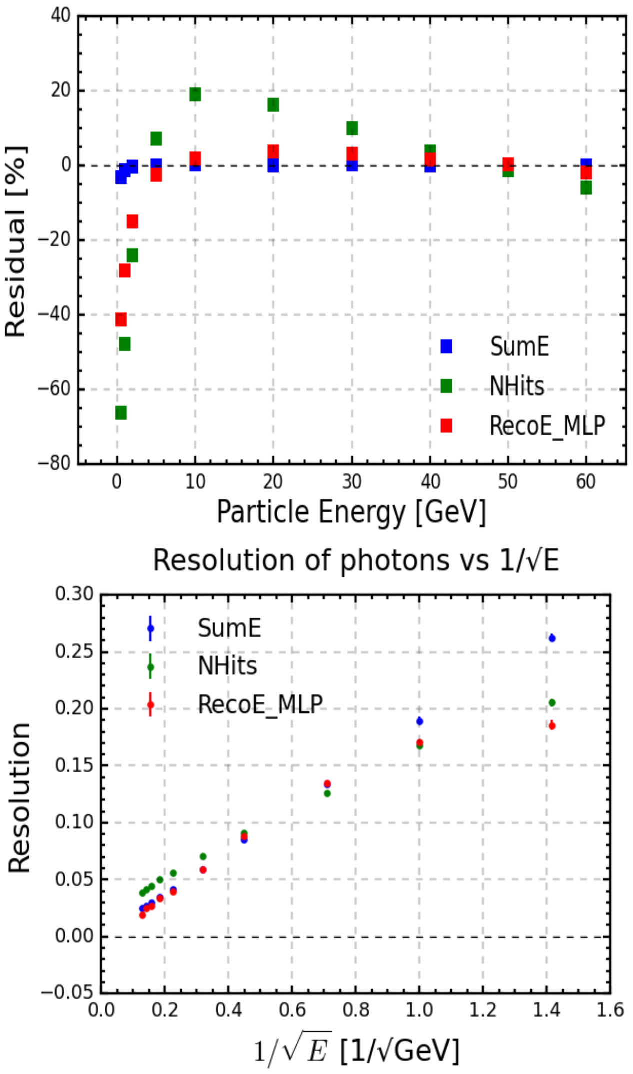
Sigmoid



tanh



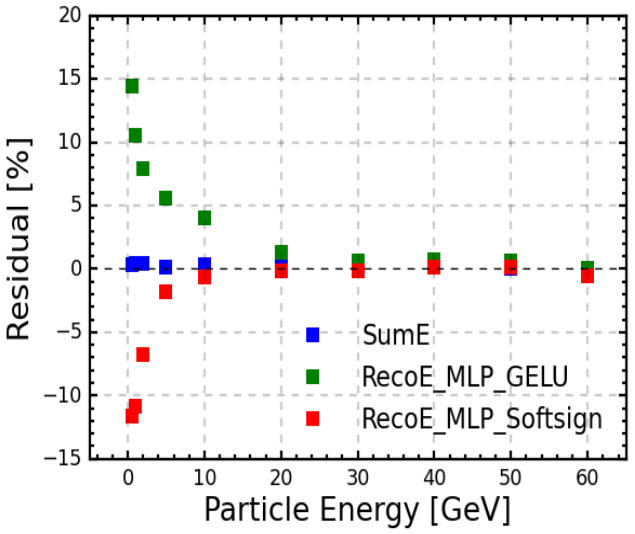
SELU

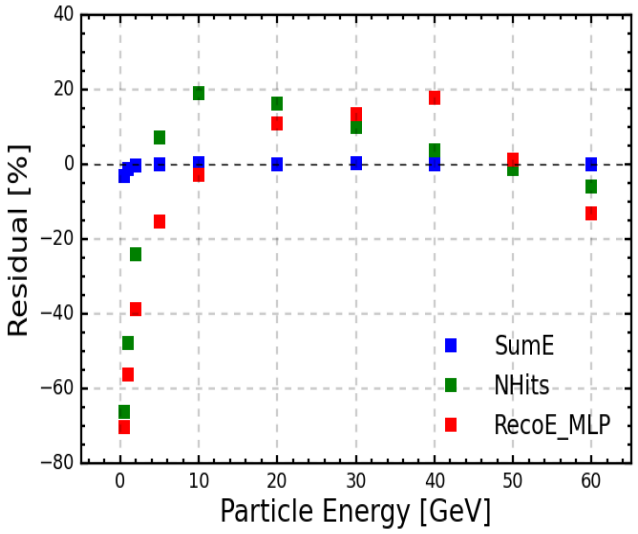


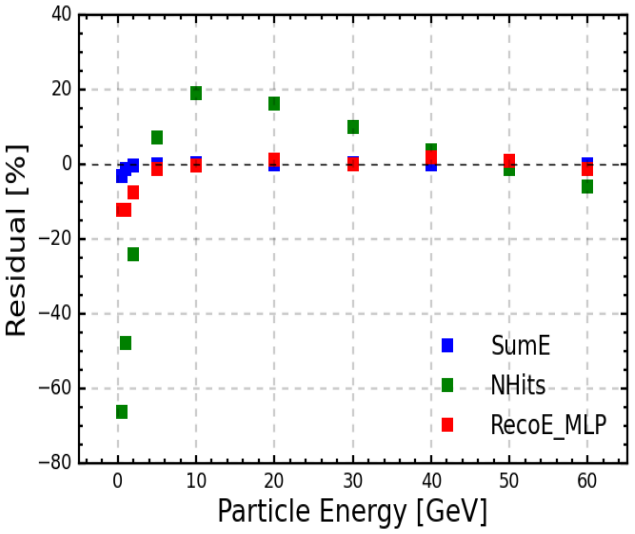
Swish

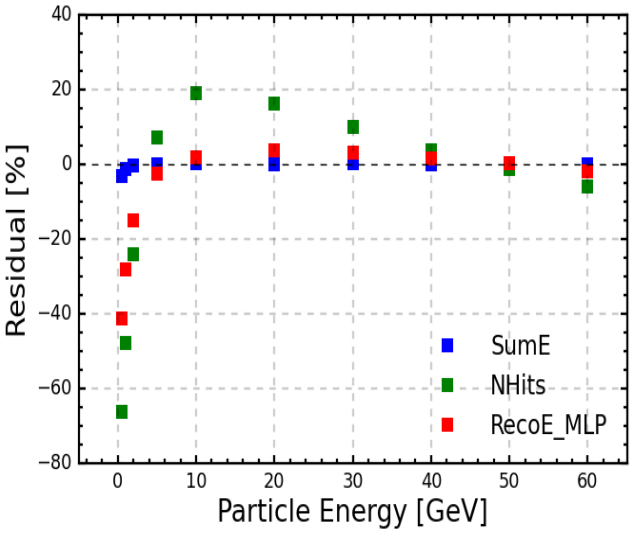
Continue with

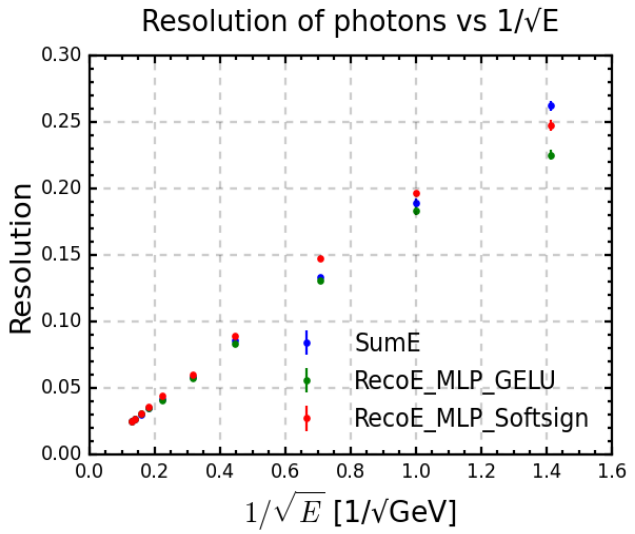
Activation function

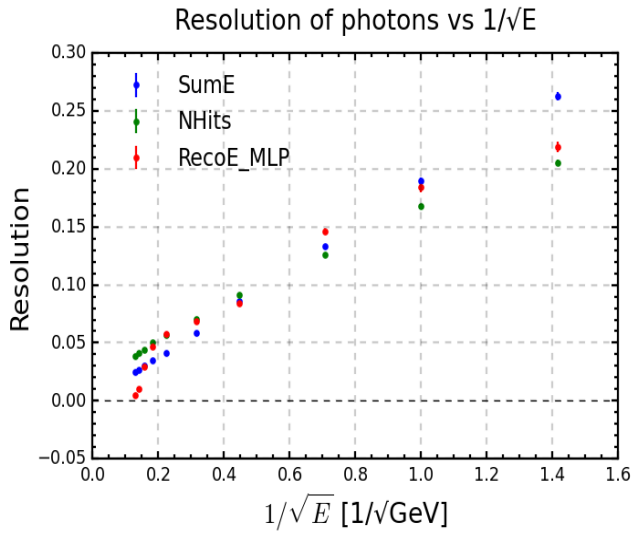


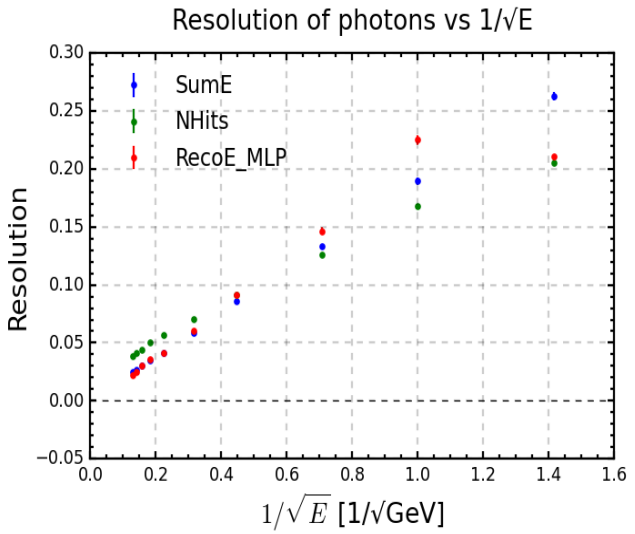


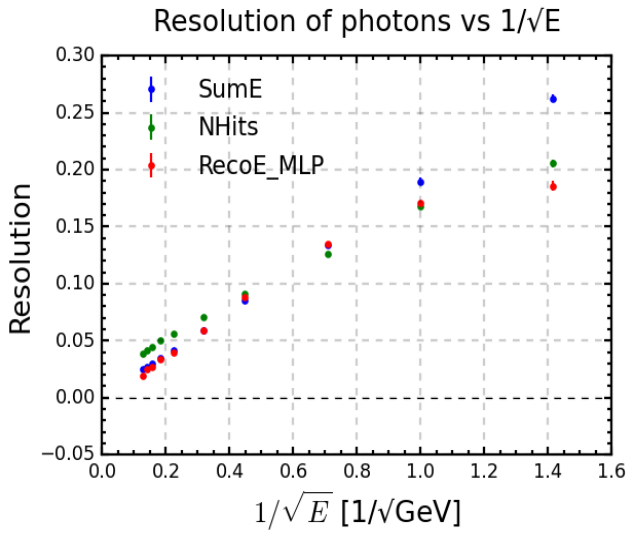








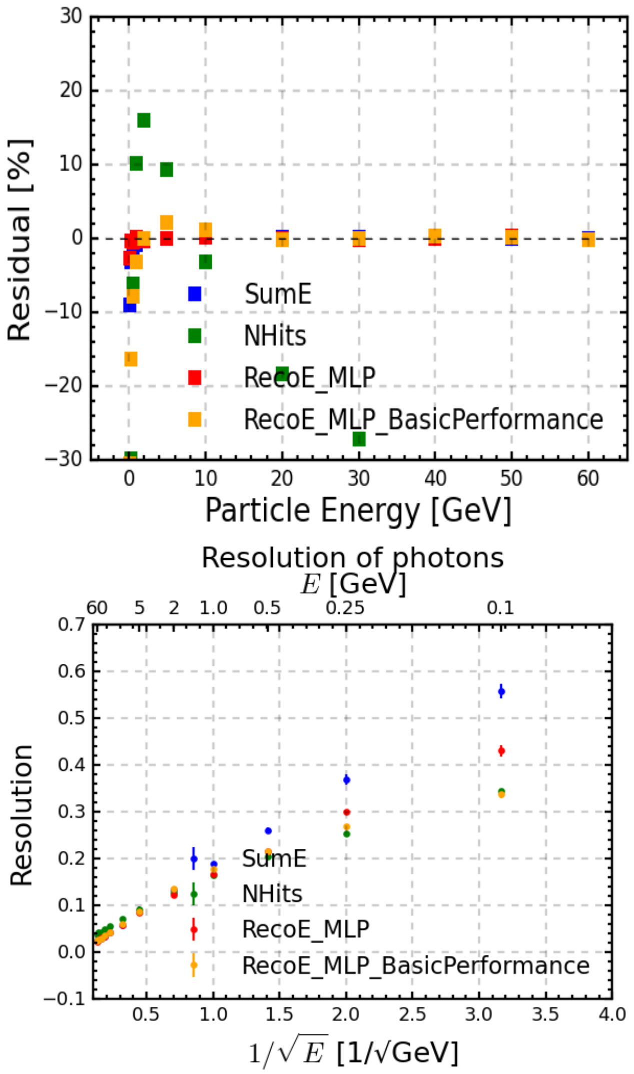




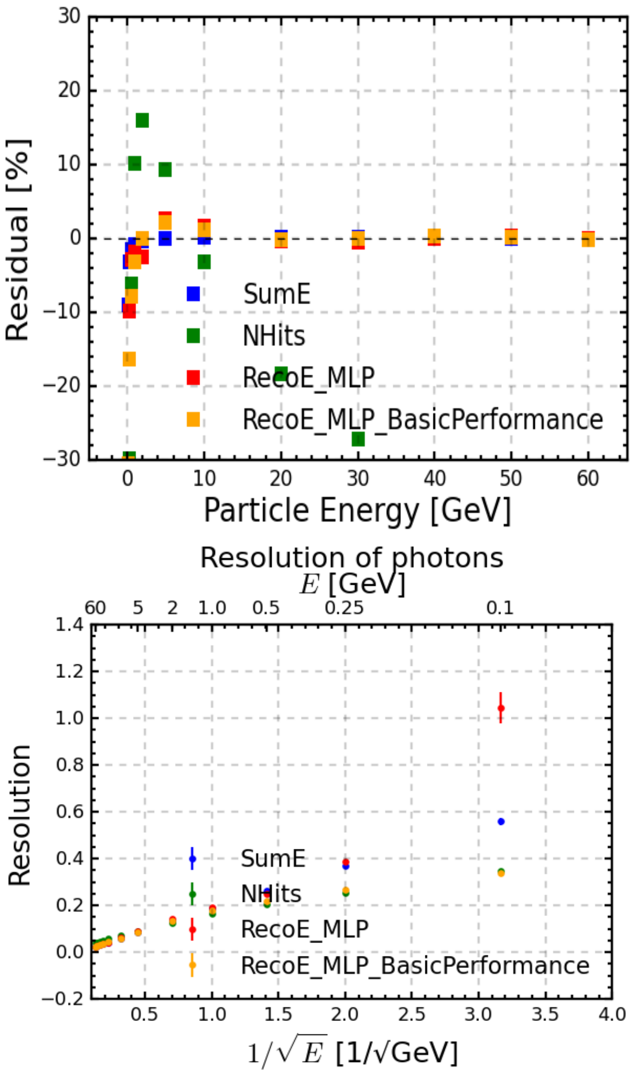
Sigmoid **tanh** **SELU** **Swish**

Continue with

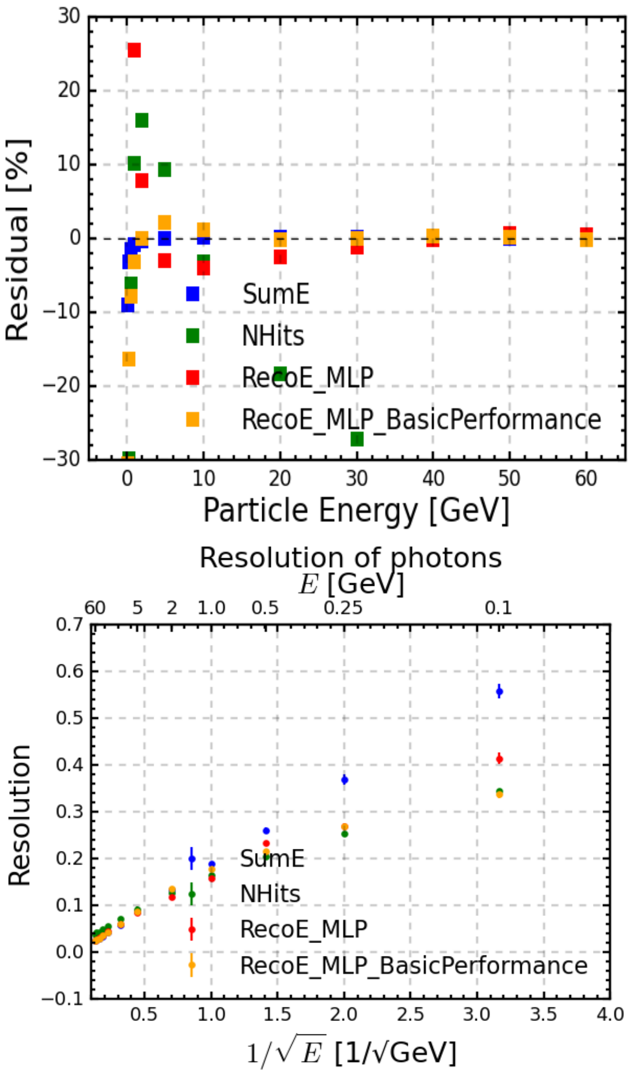
X



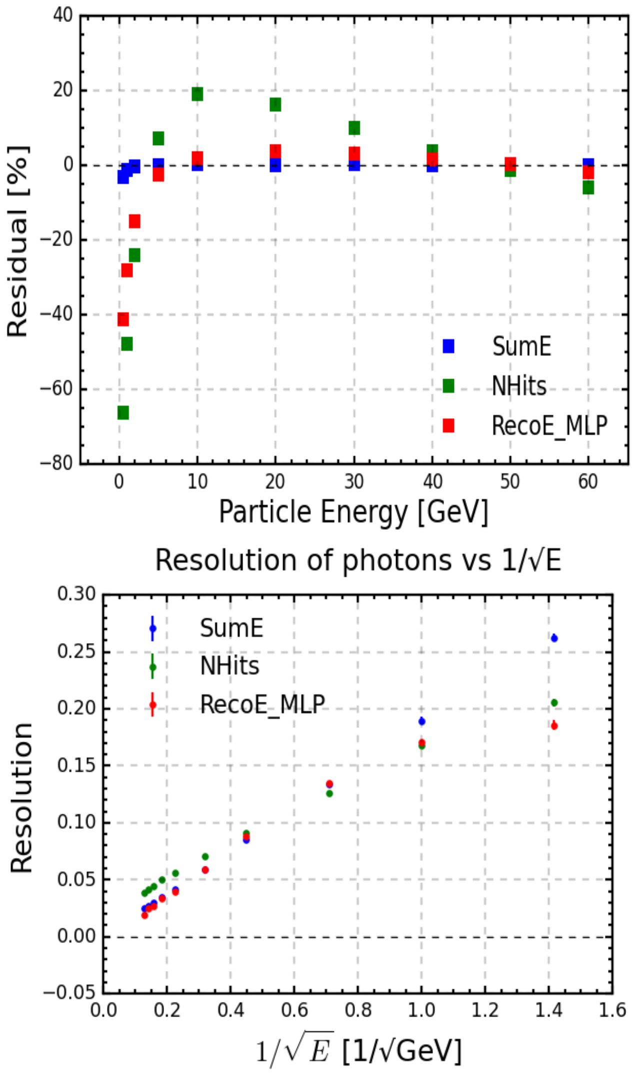
E/N



E*N



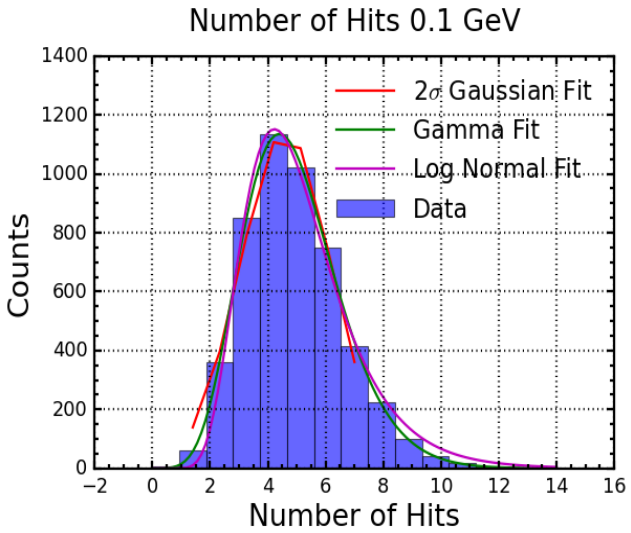
E/N+E*N

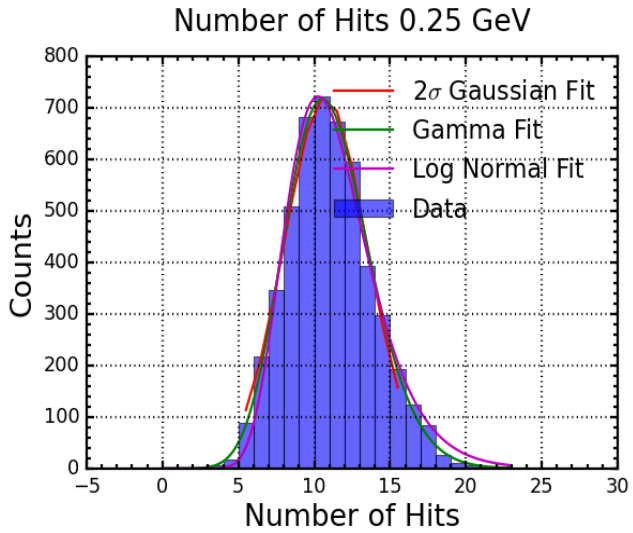


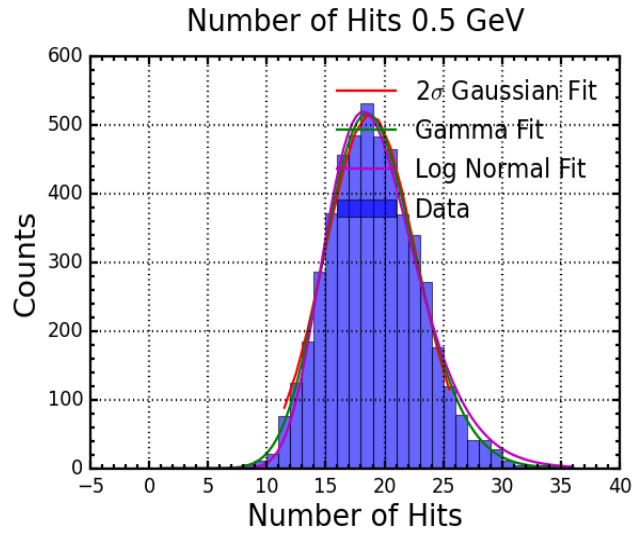
Swish

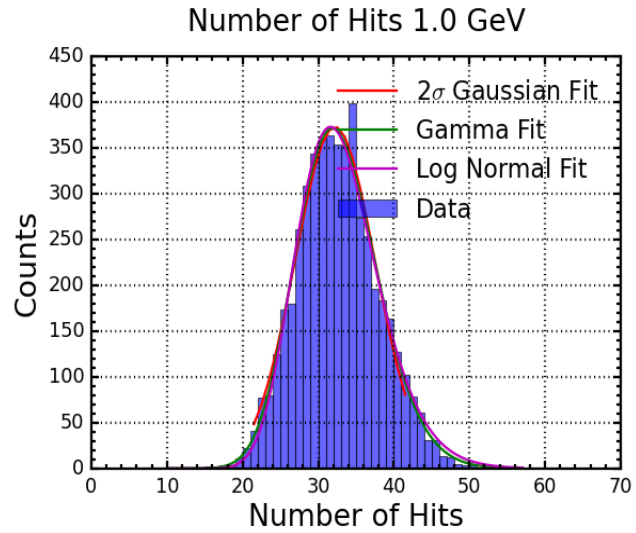
Dont use E*N

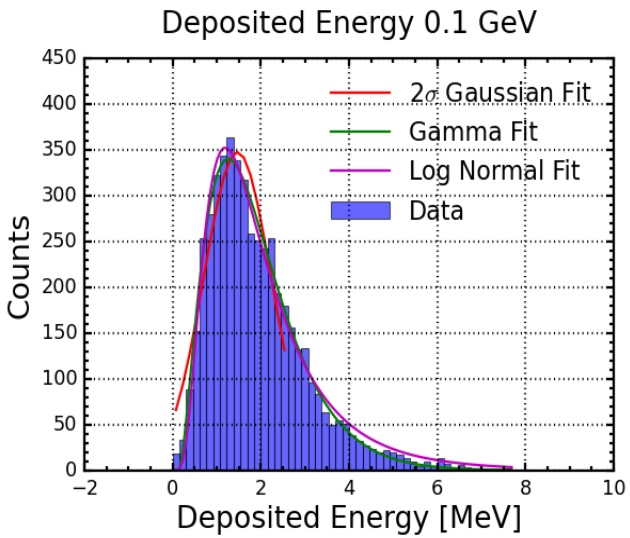
Fittings

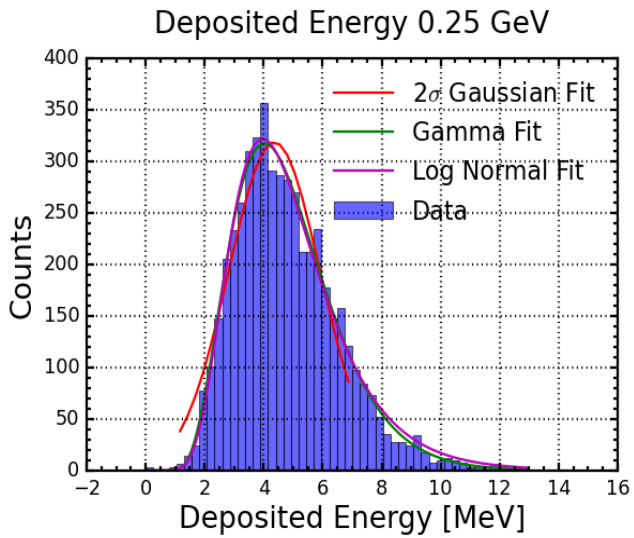


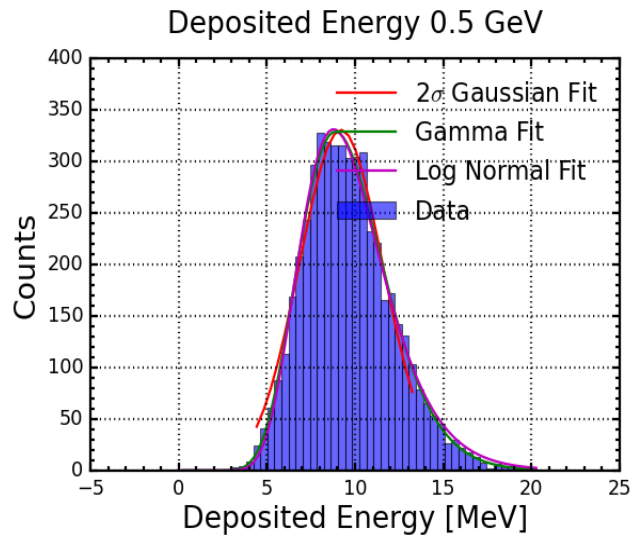


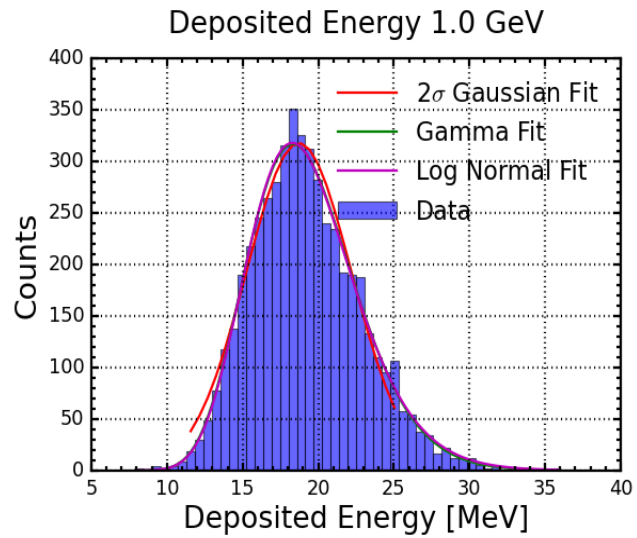






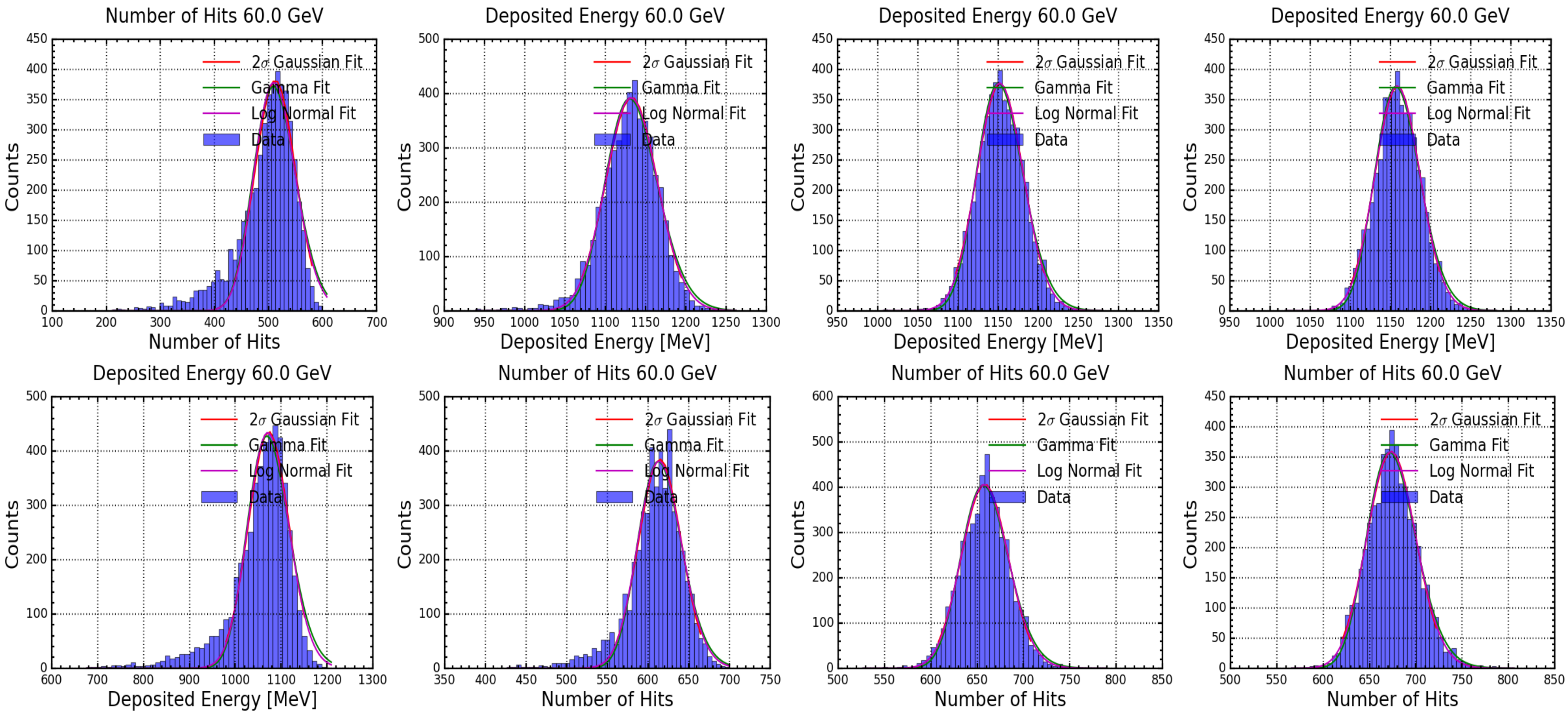






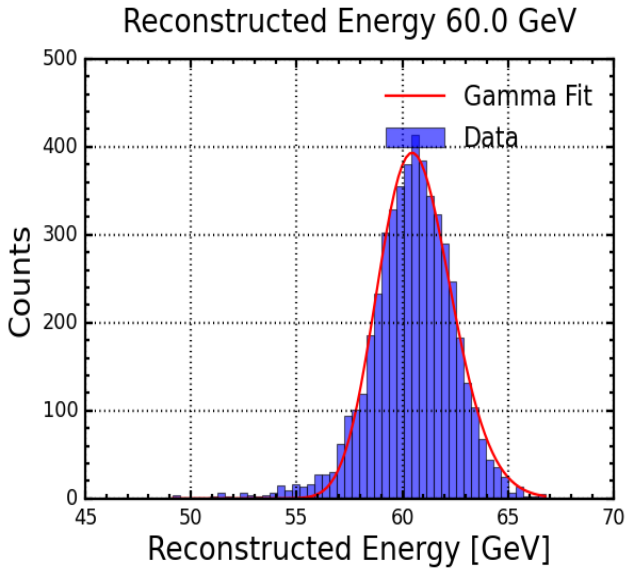
0. **0.25** **0.5** **1**

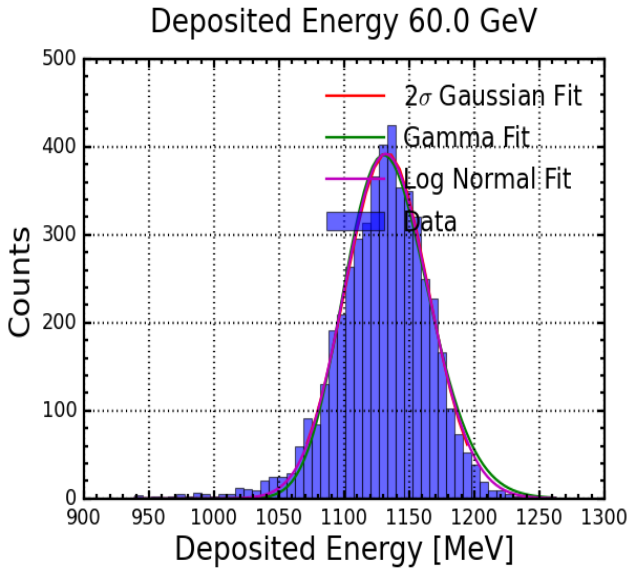
Fitting

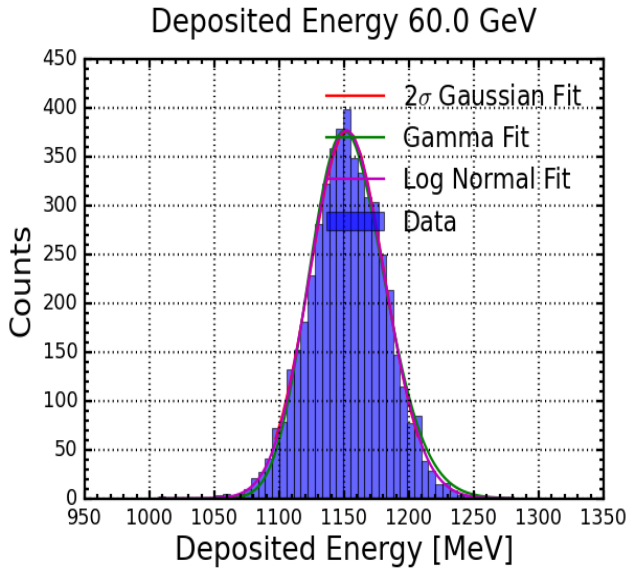


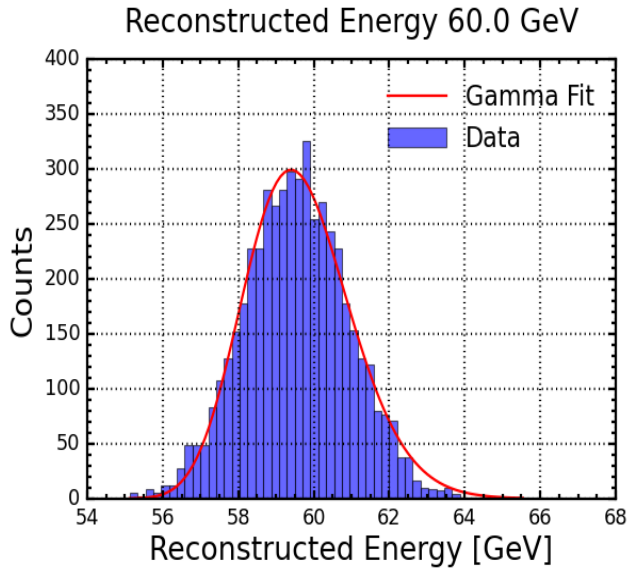
Layer20 **Layer25** **Layer30** **Layer35**

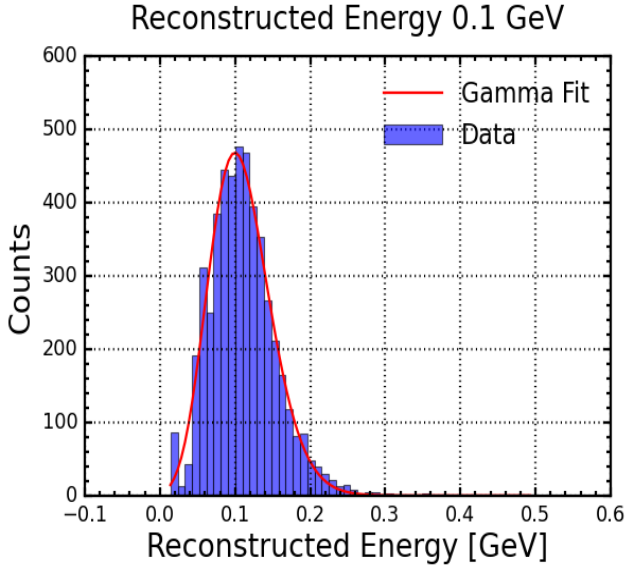
Fitting MLP

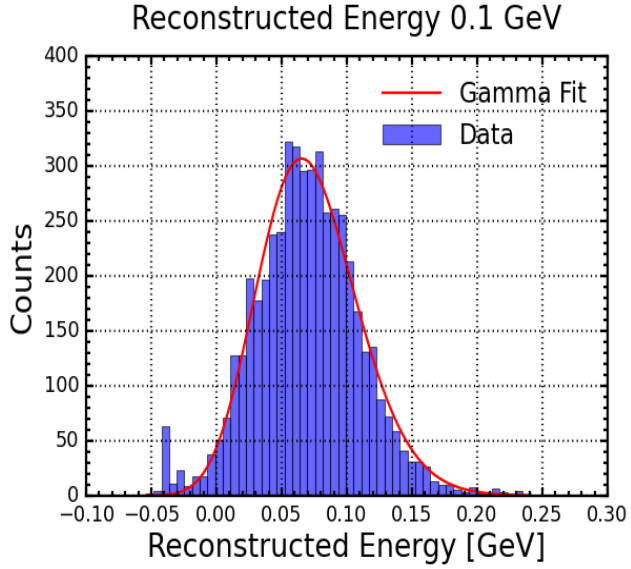






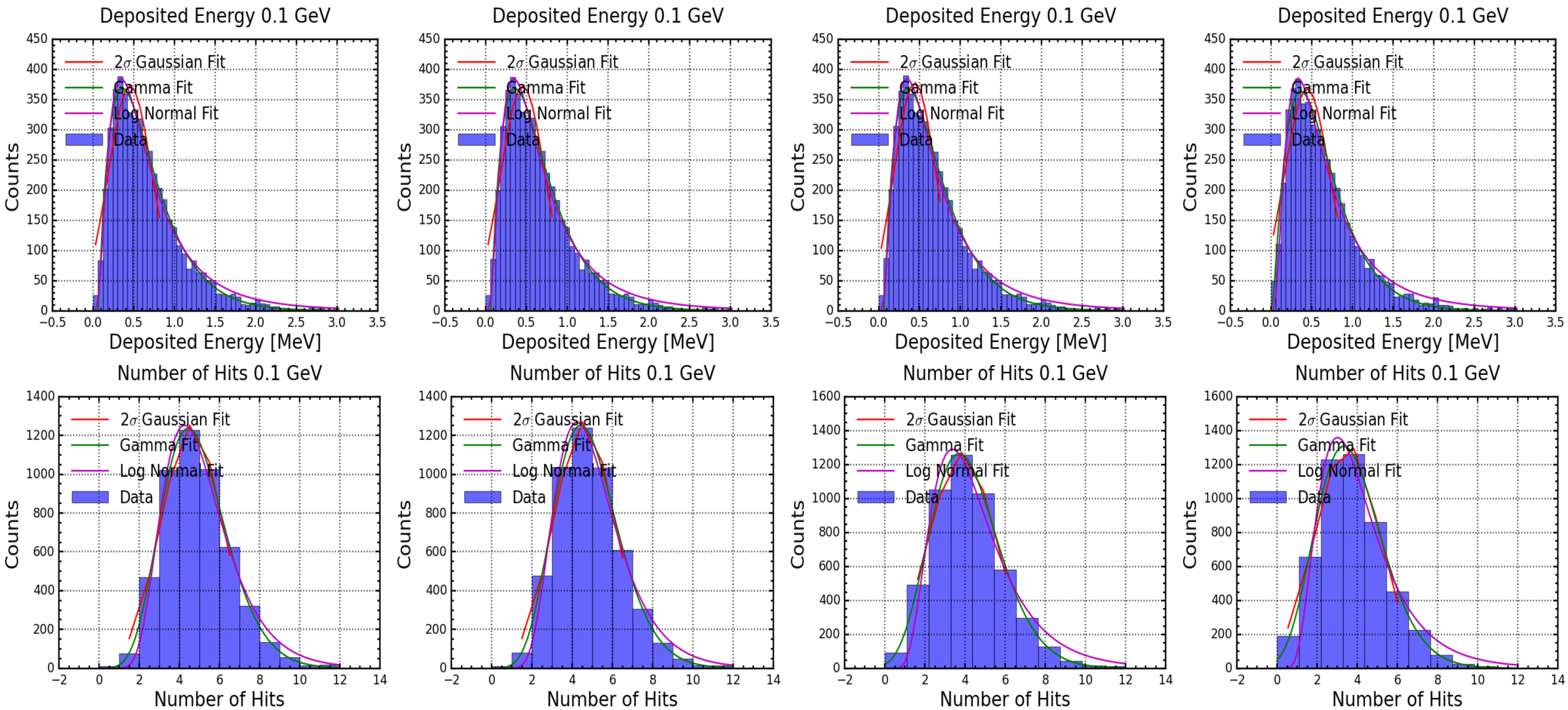






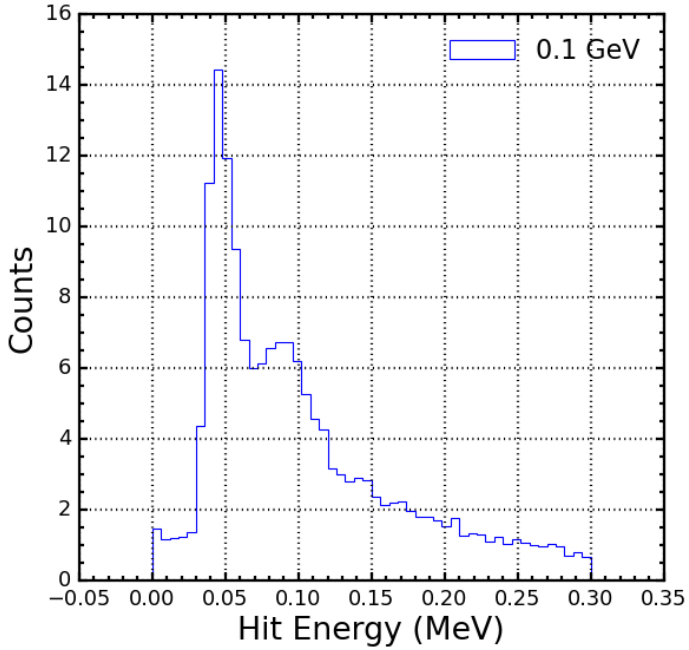
Layer20 Layer25 Layer30 Layer35

Thresholds 0.15mmSi

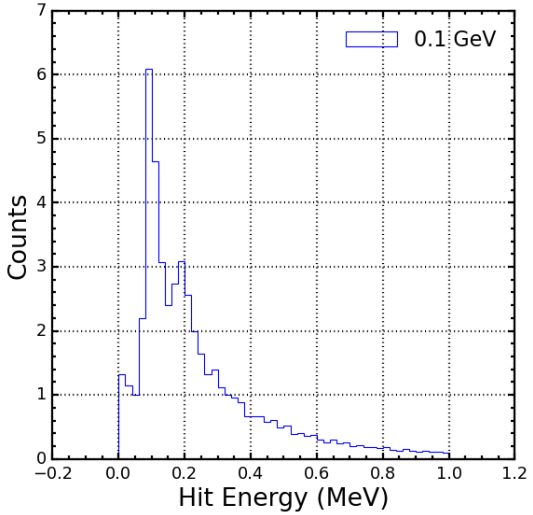


0.1 0.3 0.5 1

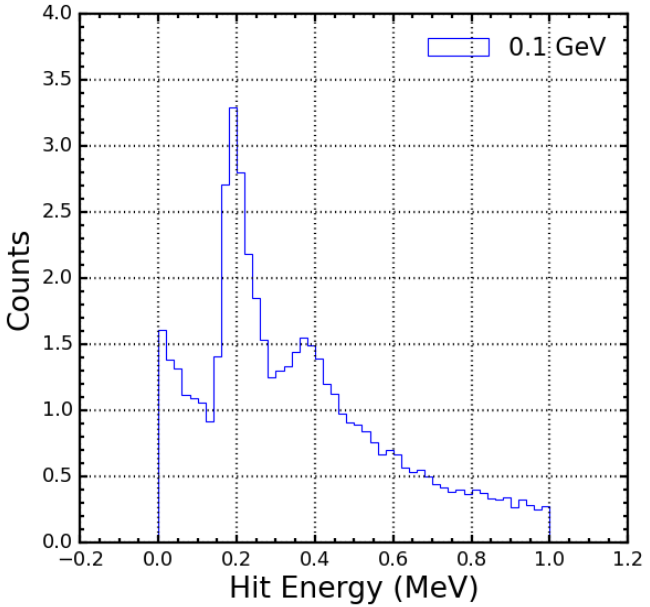
Hit E distribution



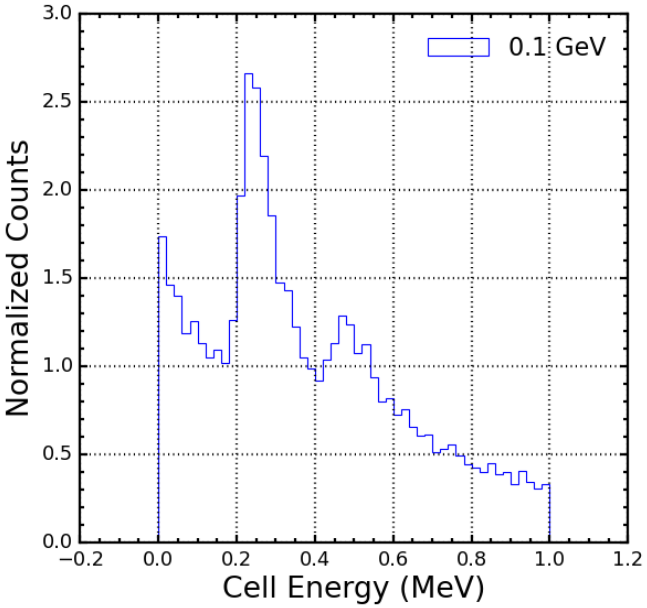
0.15mm



0.30mm

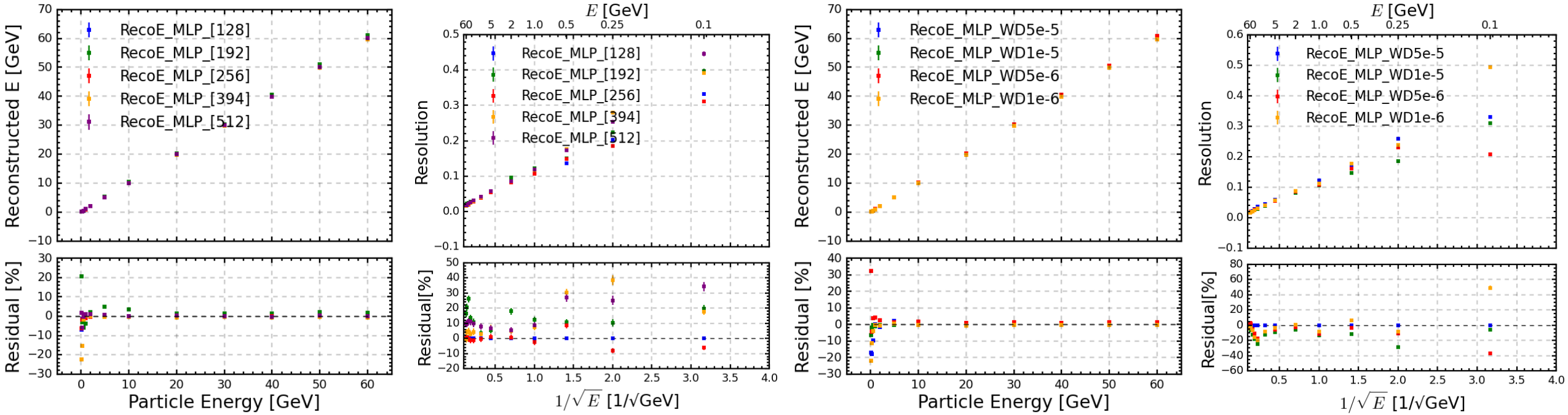


0.60mm



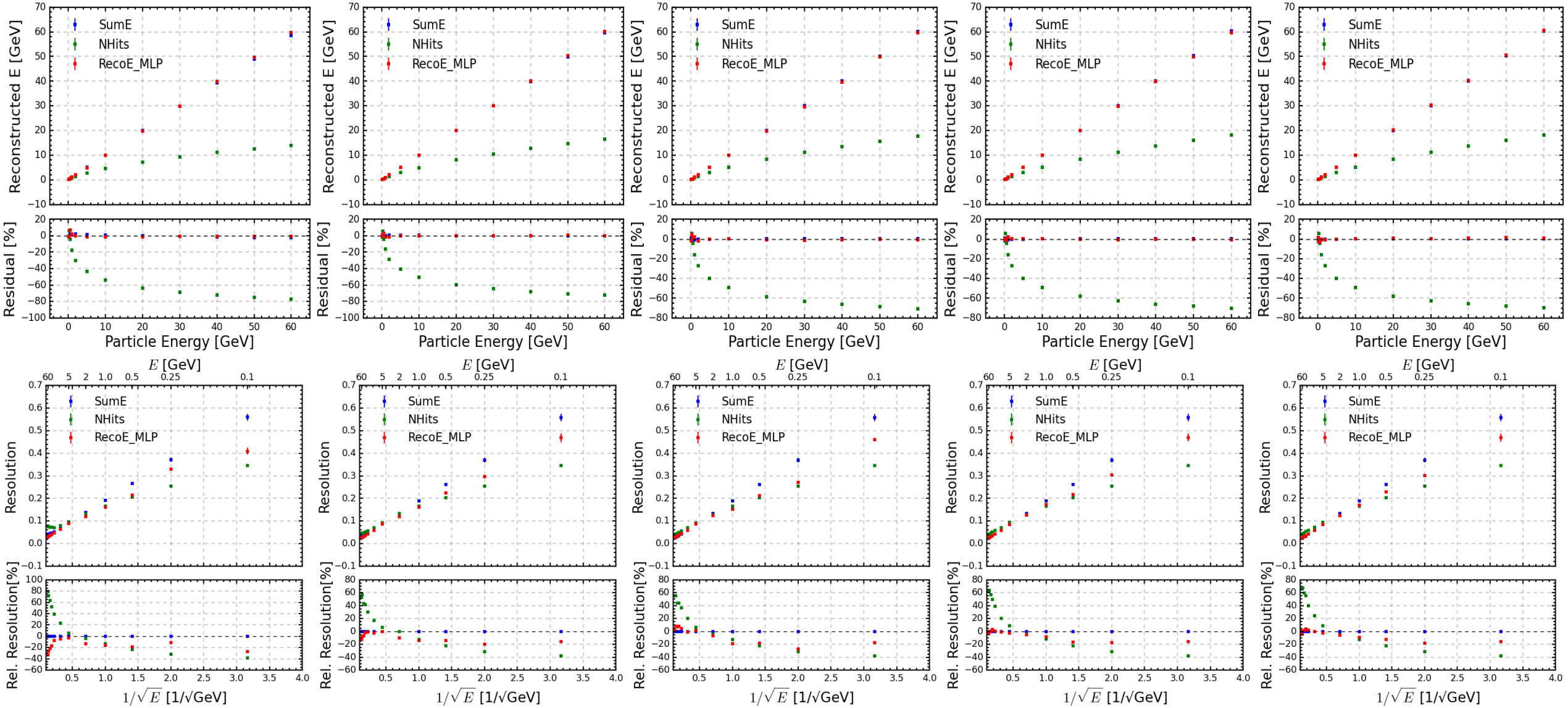
0.75mm

MLP Optimization ECAL Layer 60



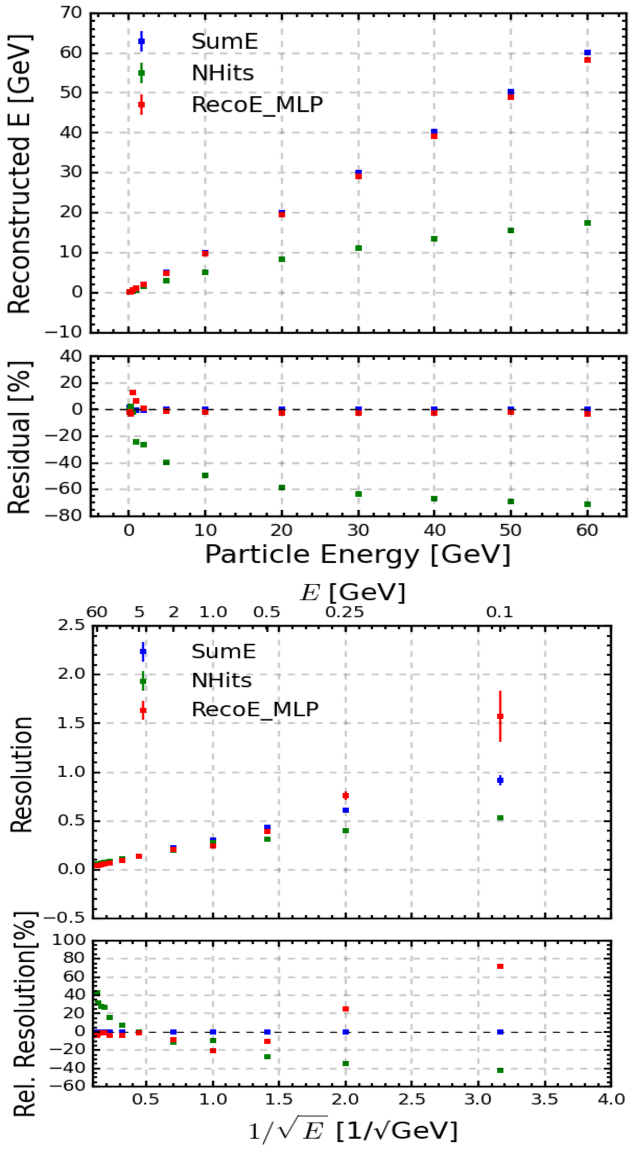
- The NN layer should be better setup as 2^N
- Weight decay and other regularization parameters should decrease while the model gets larger

Total layers

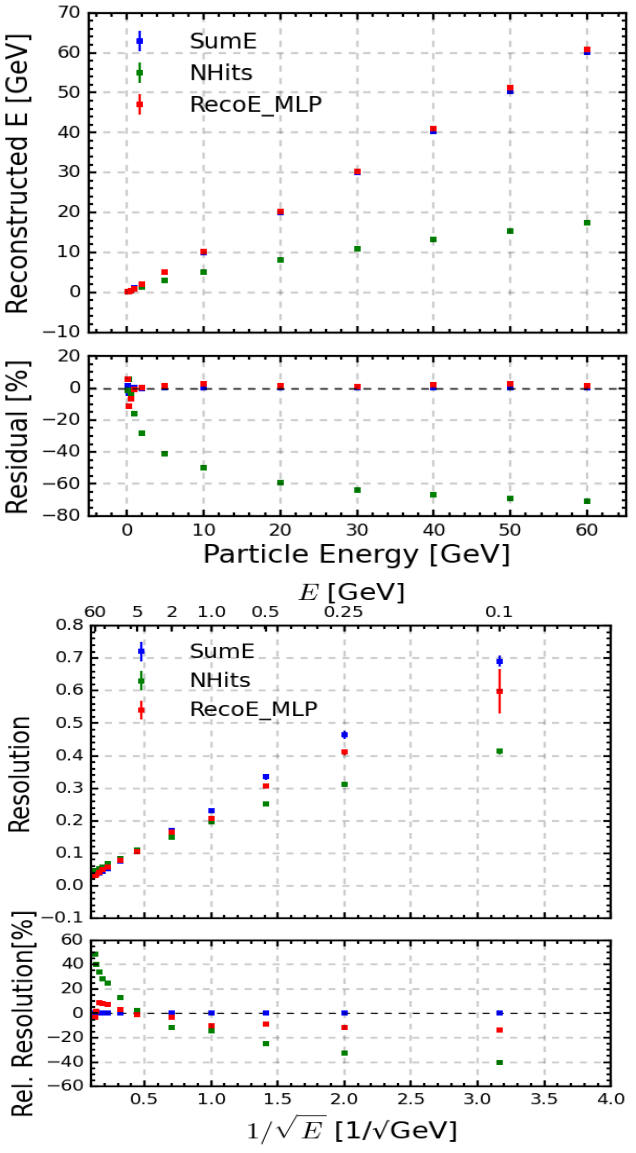


Layer40 Layer50 Layer60 Layer70

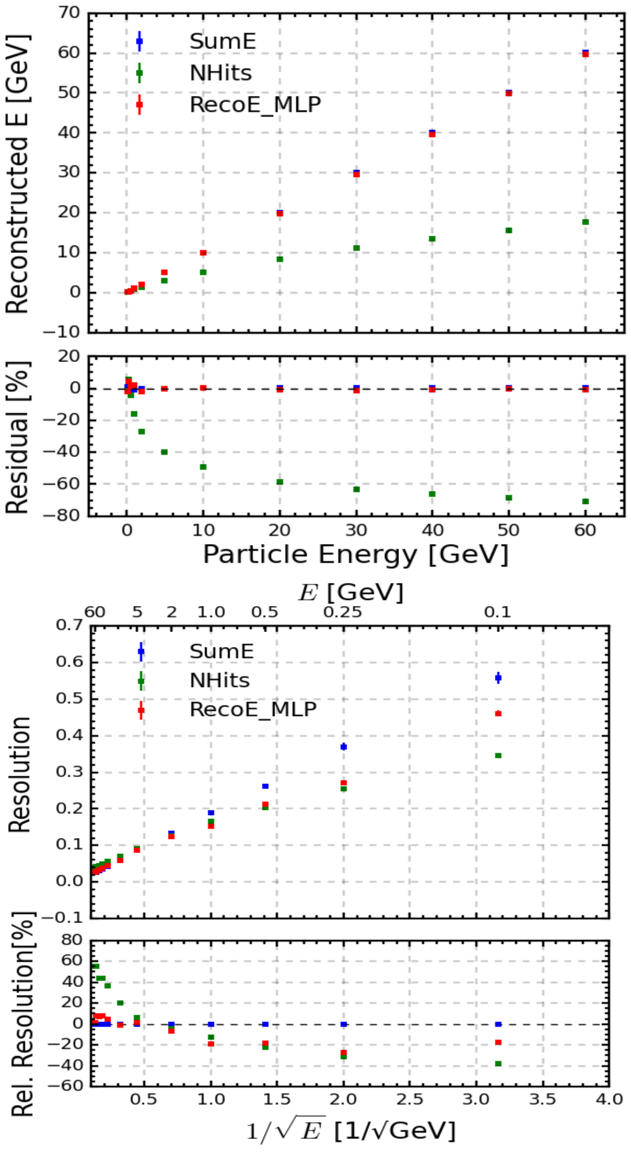
Sampling layers



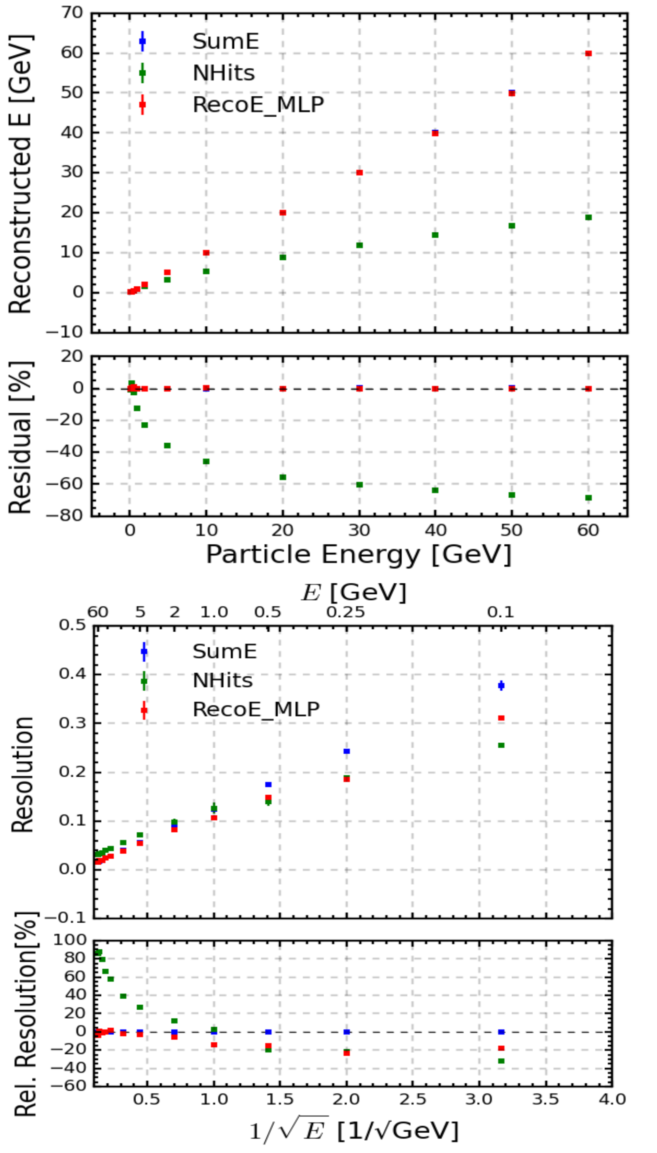
Layer12



Layer20



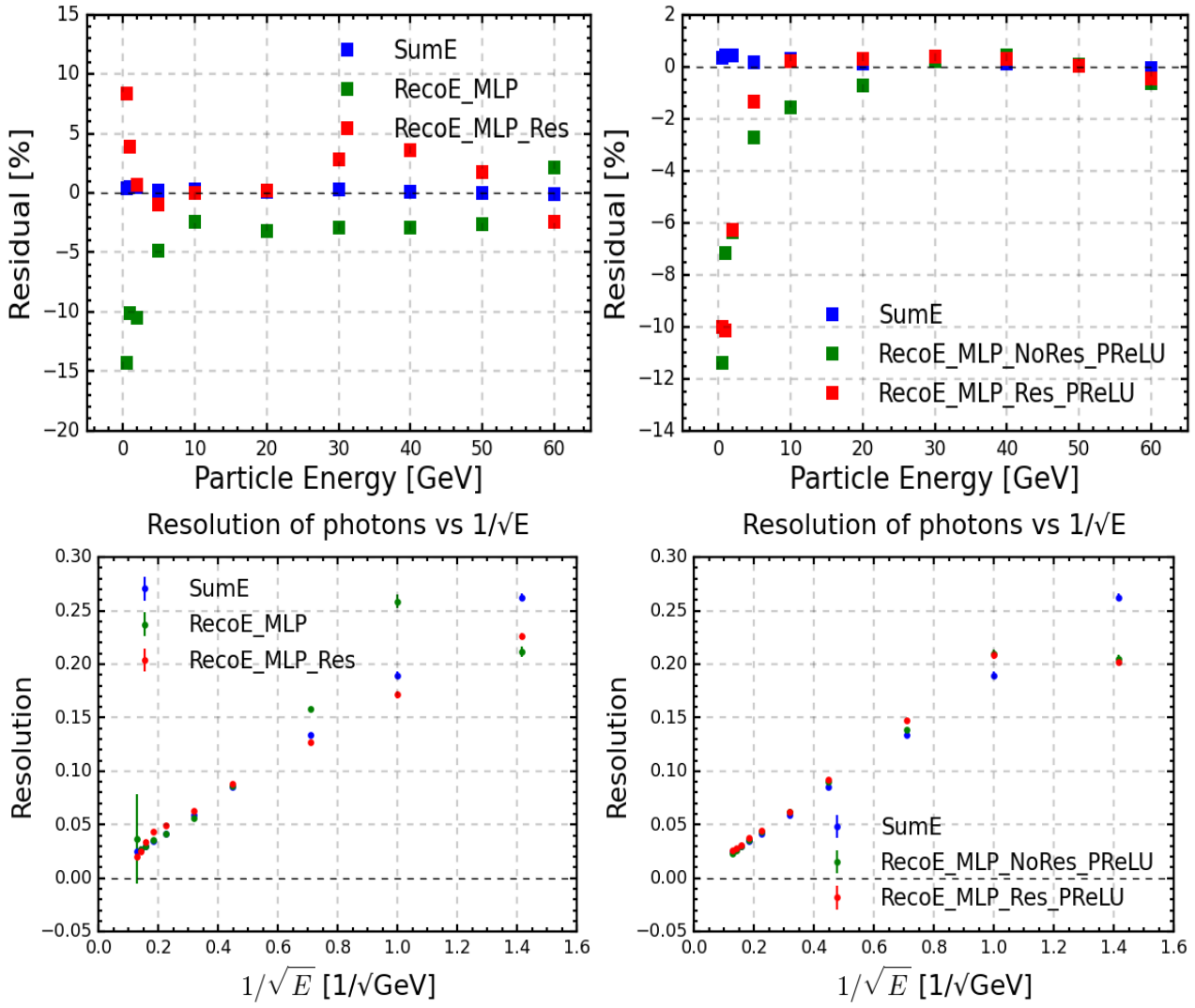
Layer30



Layer60

Add Residual connection

- y=F(x)+x, the output of a layer contains the input
- Helps the network learn **residual mappings** instead of full transformations
- Improves **training stability** and **convergence speed**
- **Try attention in the future?**



Sigmoid **PReLU**

Note for 0

LR down ✅

GNN net down ✅

Add extra seems to be worse?!

Make the extra slim first ❌

Make the model smaller ❌

Update dropout and weightdecay ❌

Slim extra to two没啥

Pool后进行activation ❌

MLP最开始进行Batch norm ✅

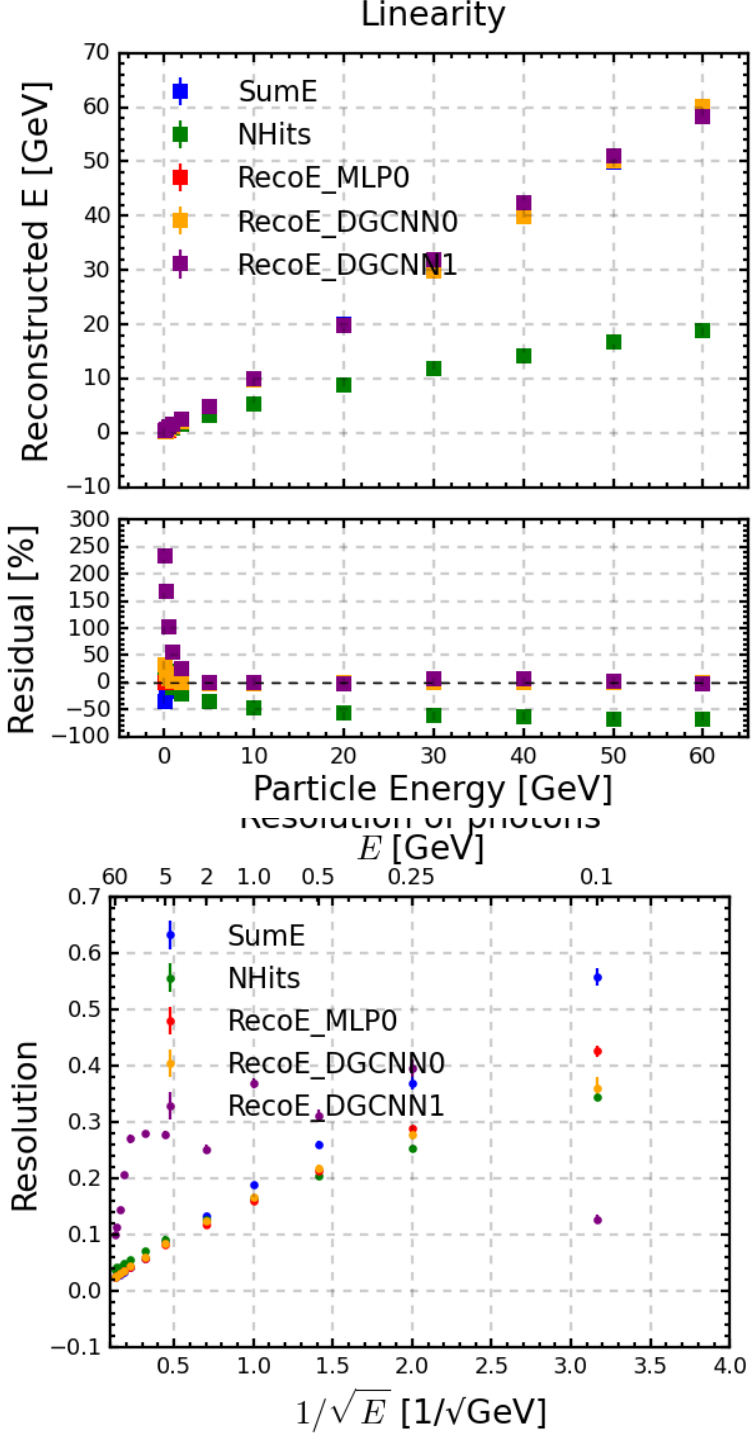
加入E_N不行？

去掉extra，这个专门不用extra

Edge加一个 ❌

Drop out->0.1不行？

先不管了，去调GATEdge



Note for 1

NN layers 128 ❌

Drop out 0.2 ✅

Weight decay 1e-4高能区不行

KNN=6，辣鸡，之后再试

Mean mean还行，其它都是傻逼，试试attention

Reduce GNN layers？之后再试试

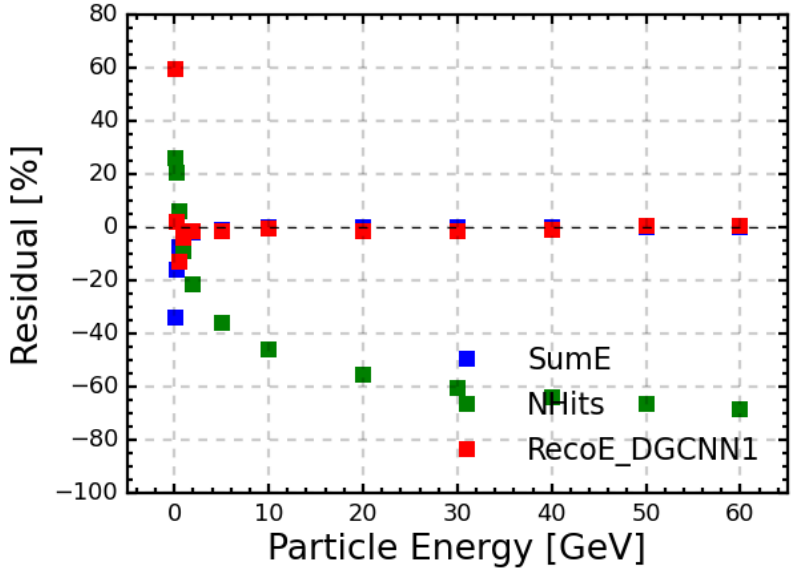
Activation after scatter ❌

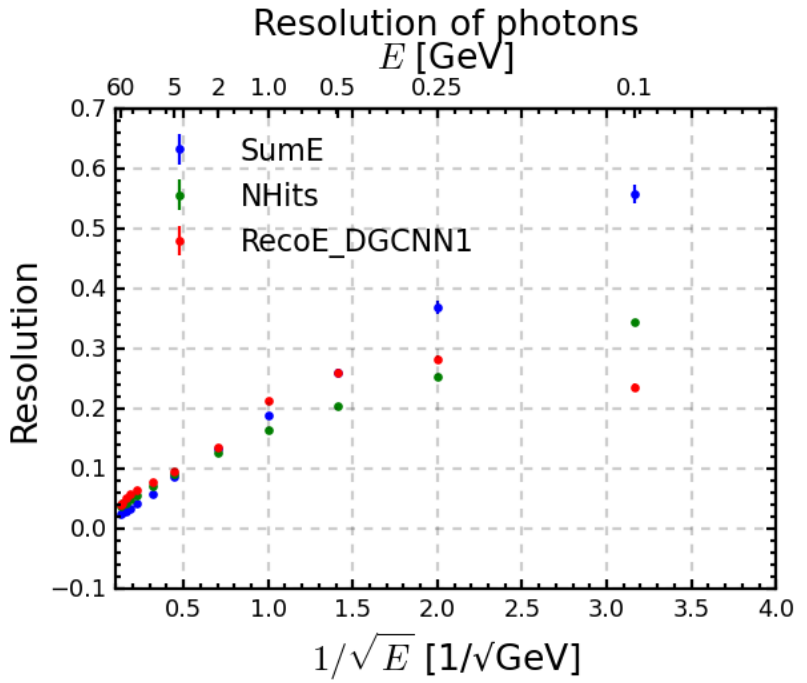
Node feature加入E_N

Edge NN改为1层效果不明显

到底是什么让线性这么差，貌似是dropout和weight decay

试试多加activation，或者换activation函数





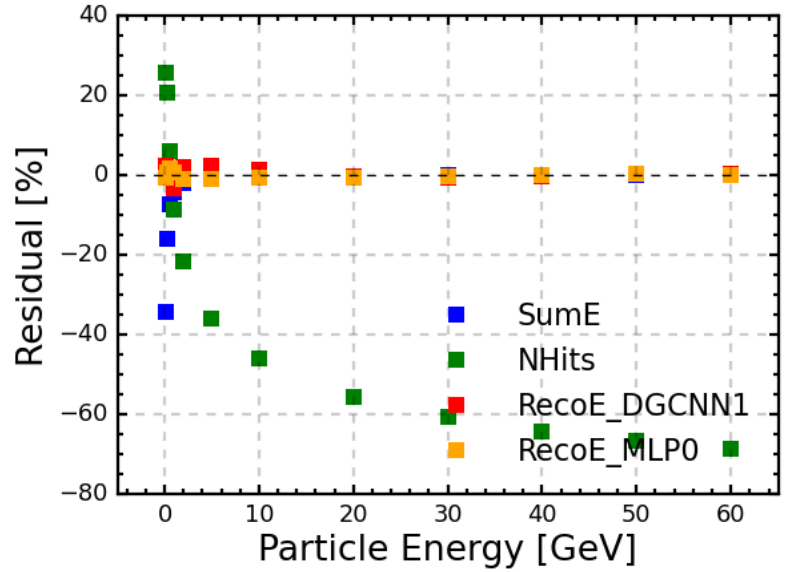
Note for 1

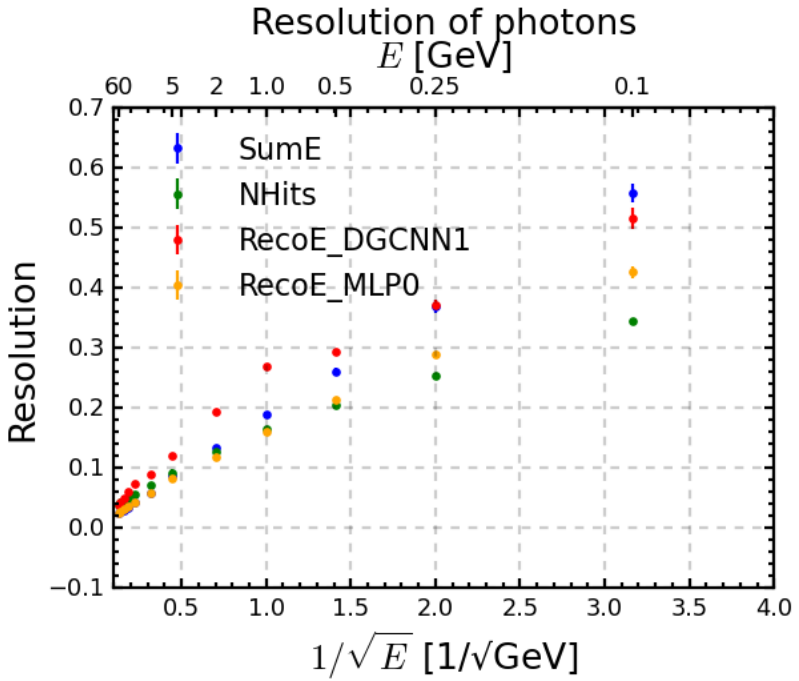
加activation，先这样吧

ReLU不行，GNN部分用ReLU,也不行

Dropout=0感觉区别不大，测试0.05,还行

New huber 0.1





Note for 2

With extra features

MLP 258->128

Extra简化到2个✅

Extra增加E_N✅

E_N加入node feature✅

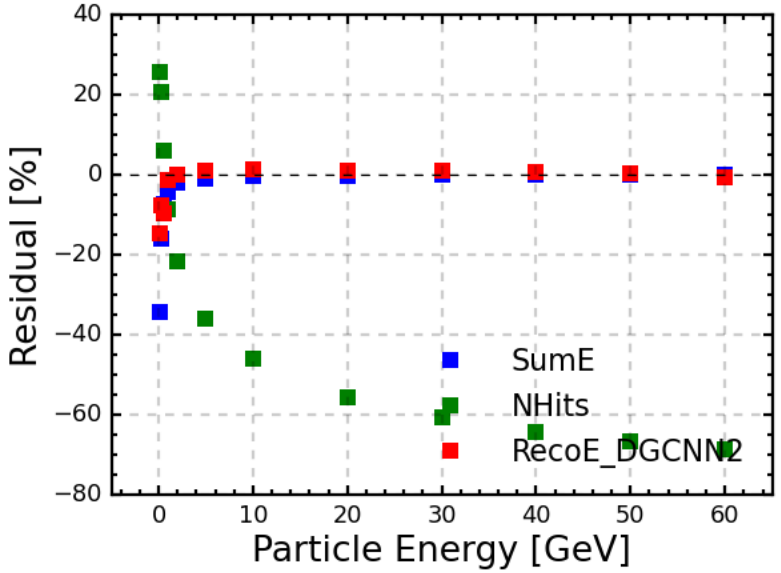
E_ratio加入node feature❌，N_ratio也不行

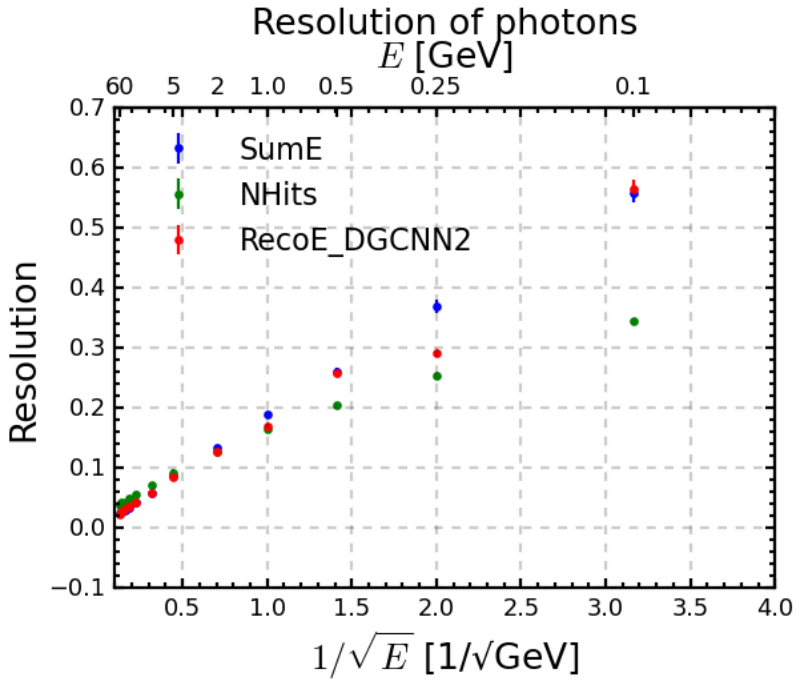
增加Edgeconv数量,貌似有用

E和N加入extra有用

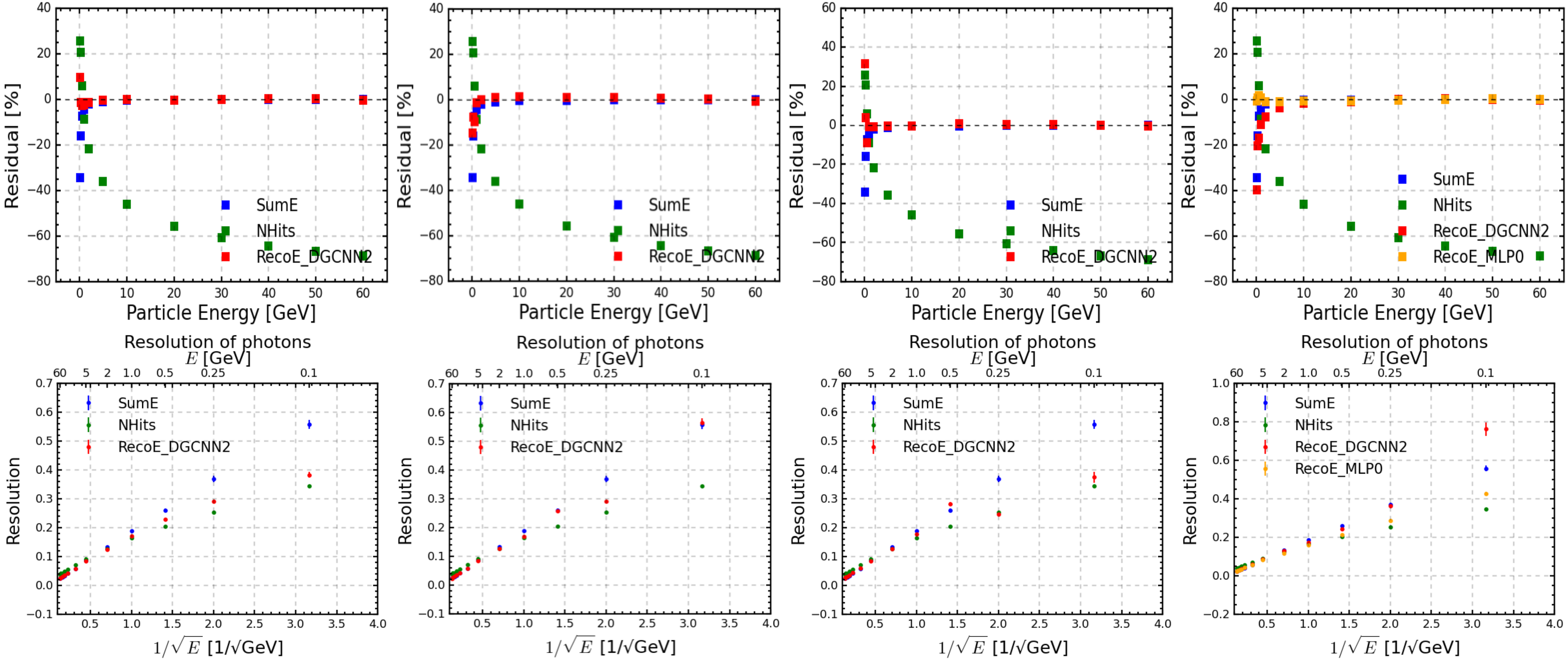
之前的MLP+少量GNN features

增加GNN?GNN得到的featur质量不够好？改Huber





Note for 2 Huber



0.05 **0.1** **0.01** **0.03**

Note for 3

Heads改为4