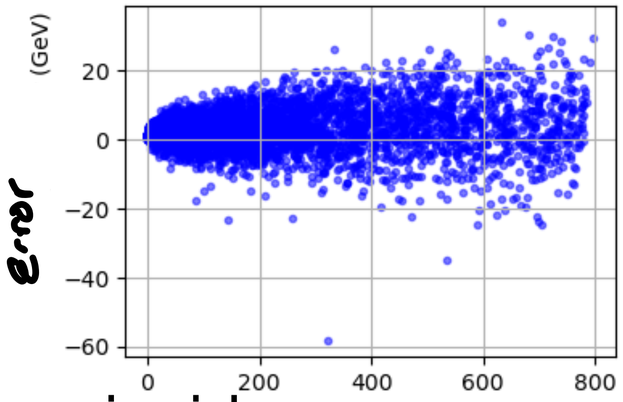
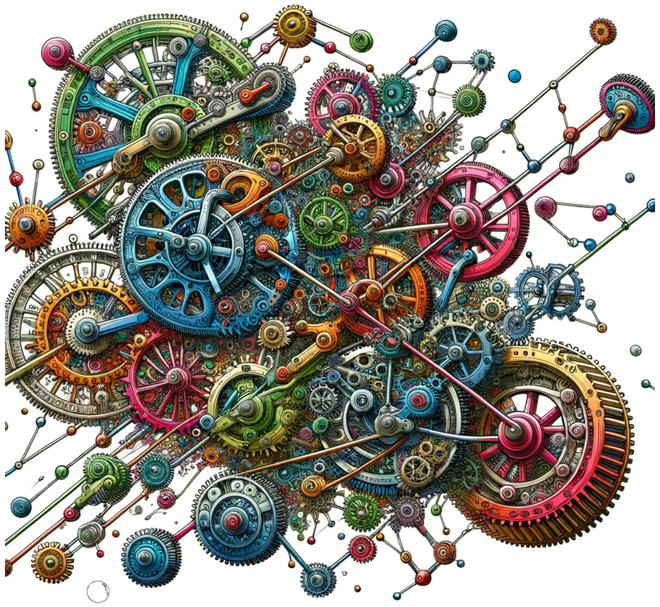
QVAE w/ Pegasus

Mar 4th

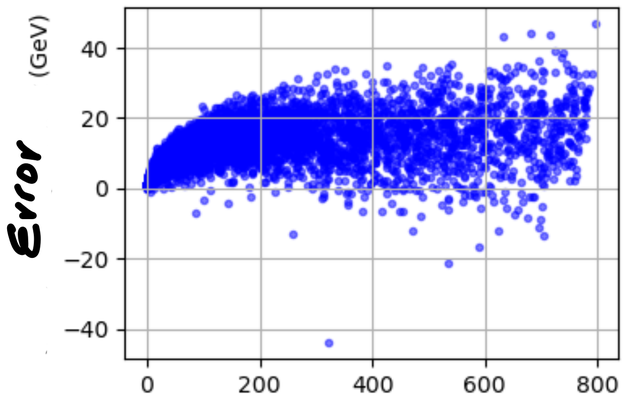
No hits

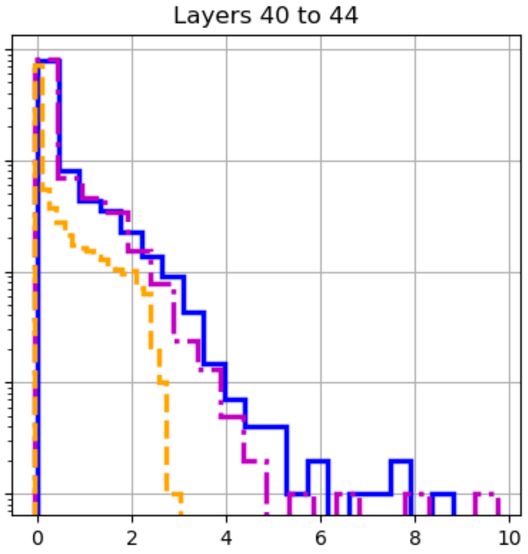


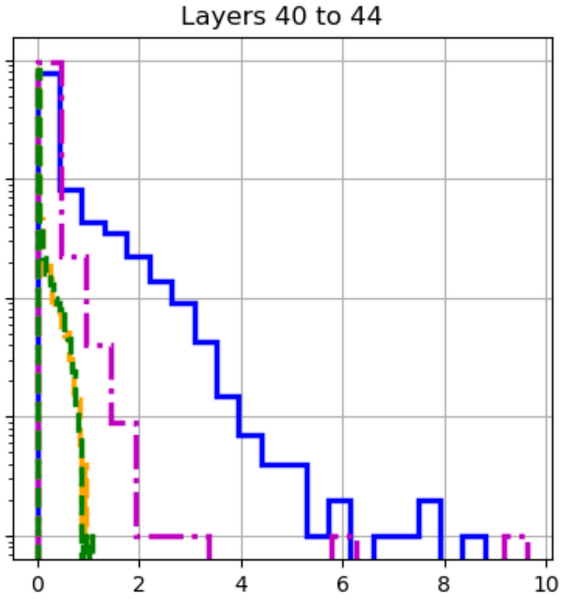
incidence energy



w/ hits

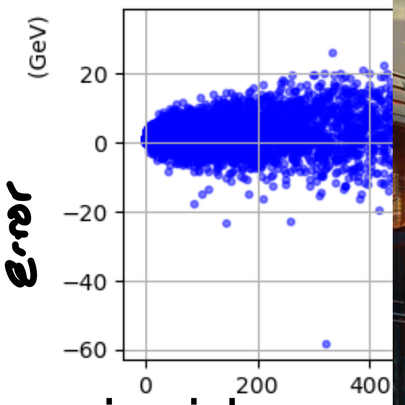






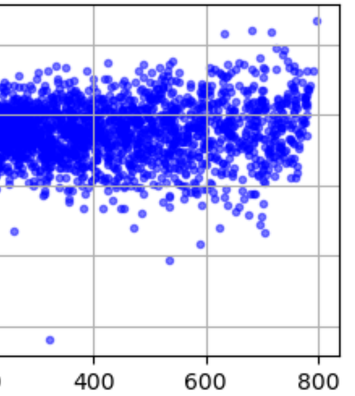
No hits

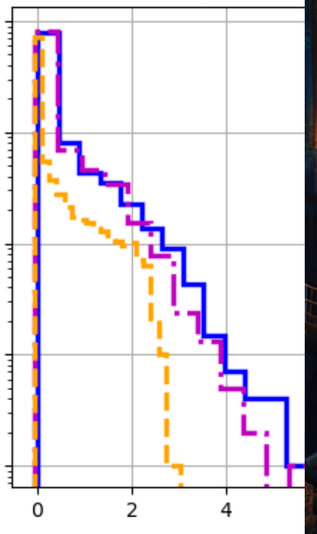
its

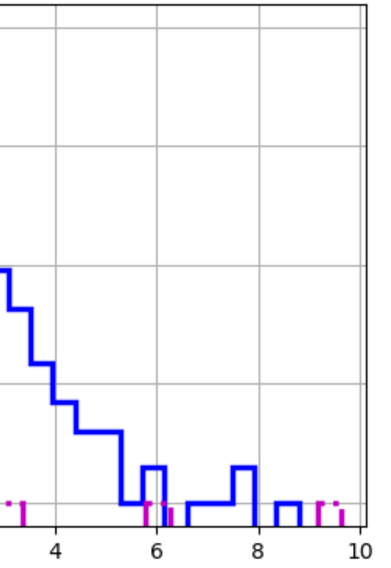


Error

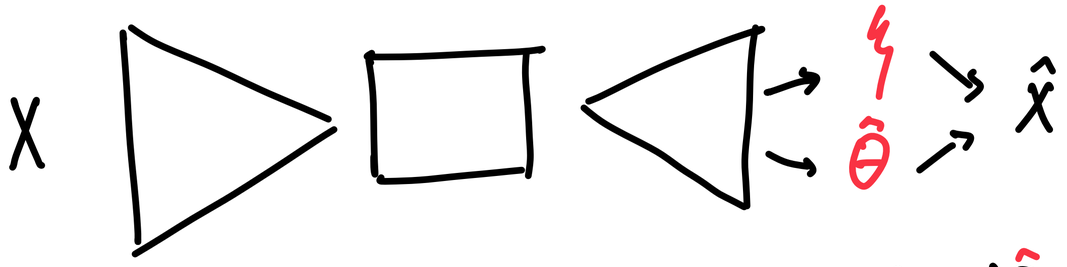
incidence







Model bias



Hits: $P(\theta) = \prod_{i=1}^{N} p_i^{\theta_i} (1-p_i)^{1-\theta_i}$

BCE: $\theta_i \ln p_i + (1-\theta_i) \ln(1-p_i)$

$p_i = \sigma(\hat{\theta}_i)$

$\mathcal{L} = MSE(x, \hat{x}) + BCE(\hat{\theta}, \theta(x))$

$+ KL_Div$

ID	x	$\mathcal{L}$	$\hat{\theta}$
1	1	1	0
2	100	100	0
3	0	0	0

Error from ID 2 is 100 times larger than error from ID 1

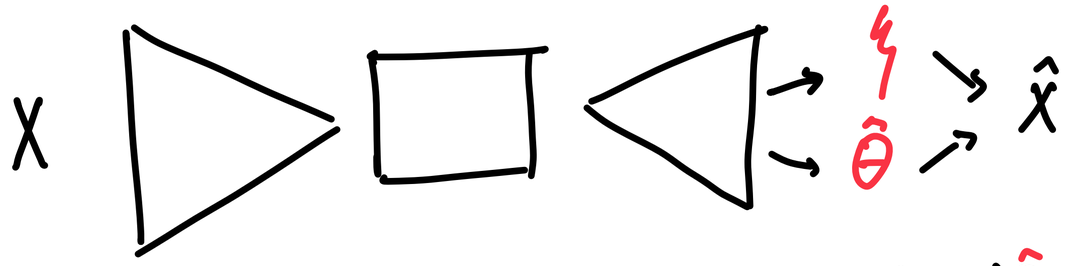
But BCE in its current form doesn't take that into consideration

$\Longrightarrow$ ADD Weights $\propto f(x)$

Model bias

$$MSE(x, \hat{x}) = (x - \hat{x})^2$$

$$\nabla_{\Omega} MSE(x, \hat{x}) = 2(x - \hat{x}) \cdot \hat{\theta} \frac{\partial \psi}{\partial \Omega}$$



$$\mathcal{L} = MSE(x, \hat{x}) + BCE(\hat{\theta}, \Theta(x)) + KL_Div$$

ID	x	ψ	θ̂
1	1	2	0
2	100	1	0
3	0	0	0
4	0	100	1

$\nabla_{\Omega} MSE$

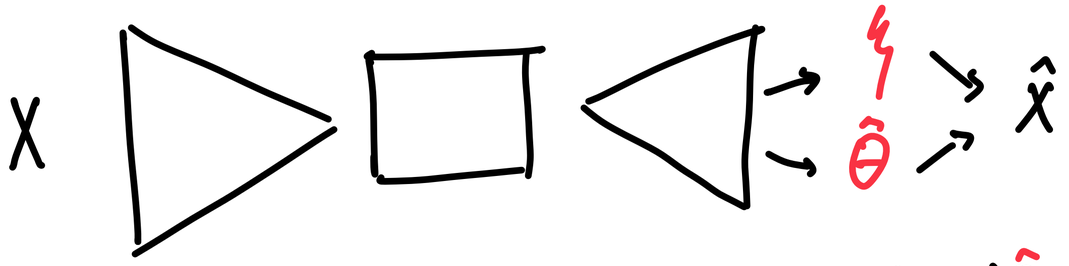
$2(1-2) \cdot 0 \cdot \frac{\partial \psi}{\partial \Omega} = 0$

$2(100-1) \cdot 0 \cdot \frac{\partial \psi}{\partial \Omega} = 0$

$2(0-100) \cdot 1 \cdot \frac{\partial \psi}{\partial \Omega} \neq 0$

Θ(x)	θ̂	Effect on MSE
1	0	Null
0	1	force activation to zero

Model bias



$$\mathcal{L} = MSE(x, \hat{x}) + BCE(\hat{\theta}, \Theta(x)) + KL_Div$$

$$MSE(x, \hat{x}) = (x - \hat{x})^2$$

$$\nabla_{\phi} MSE(x, \hat{x}) = 2(x - \hat{x}) \cdot \hat{\theta} \frac{\partial \phi}{\partial z}$$

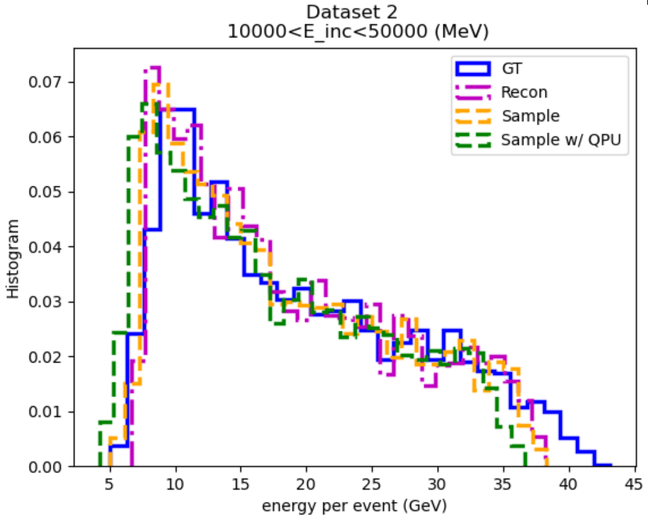
Solution: During training: $\hat{x} = \phi \otimes \Theta(x)$

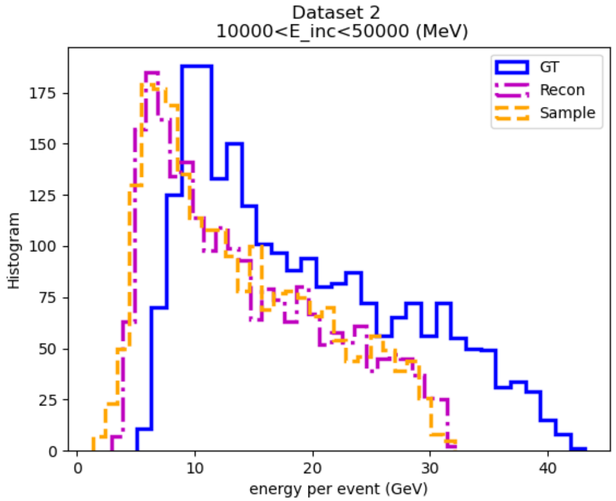
During testing: $\hat{x} = \phi \otimes \hat{\theta}$

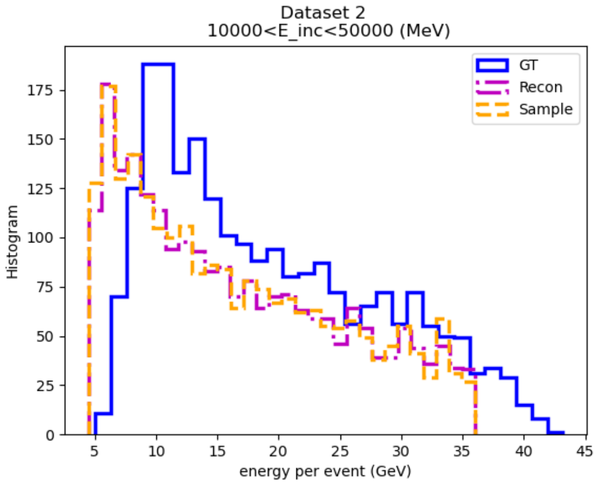
$\Theta(x)$	$\hat{\theta}$	Effect on MSE
1	0	Null
0	1	force activation to zero

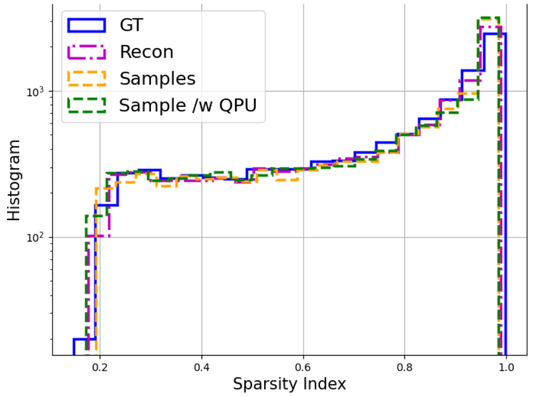
This effectively biases the model towards low energies! (which would explain the undershooting)

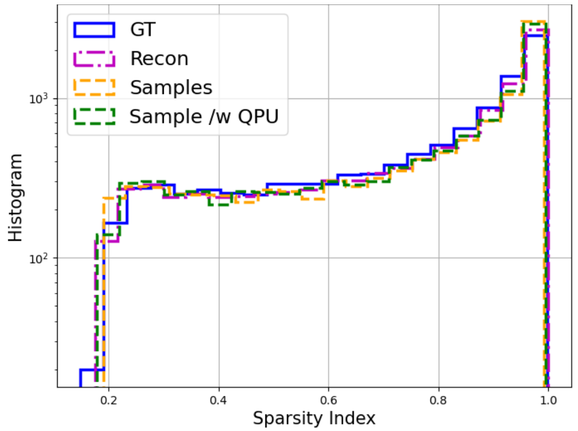
Model comparison

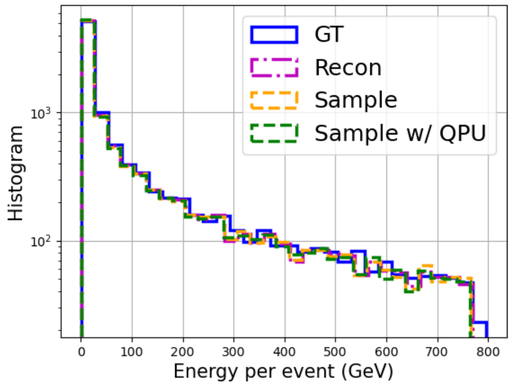


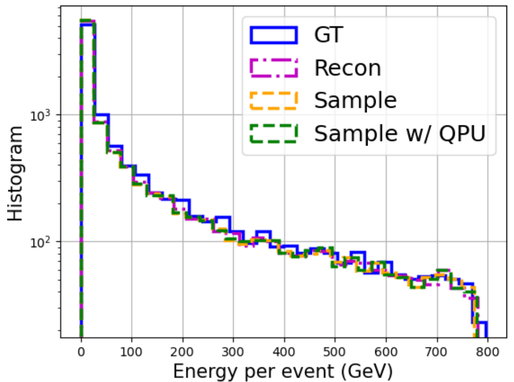


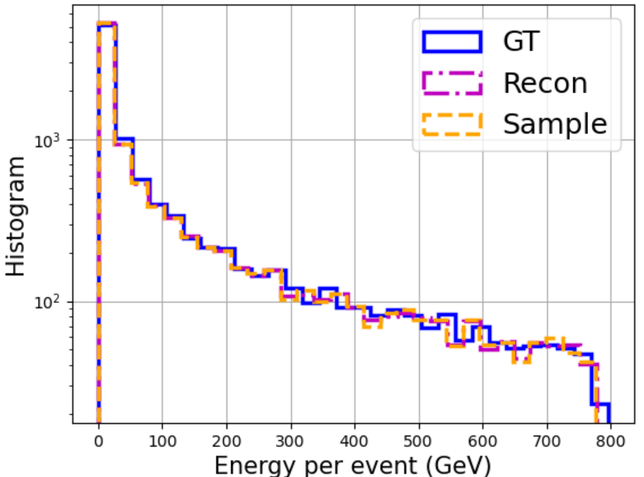


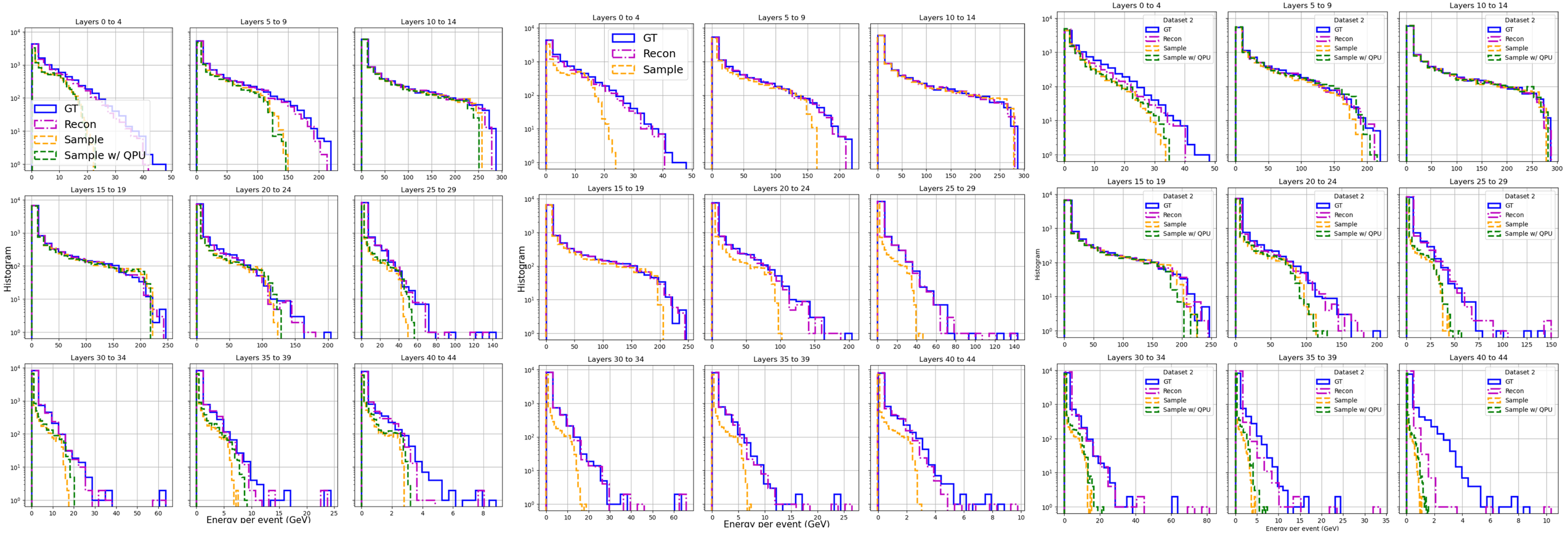


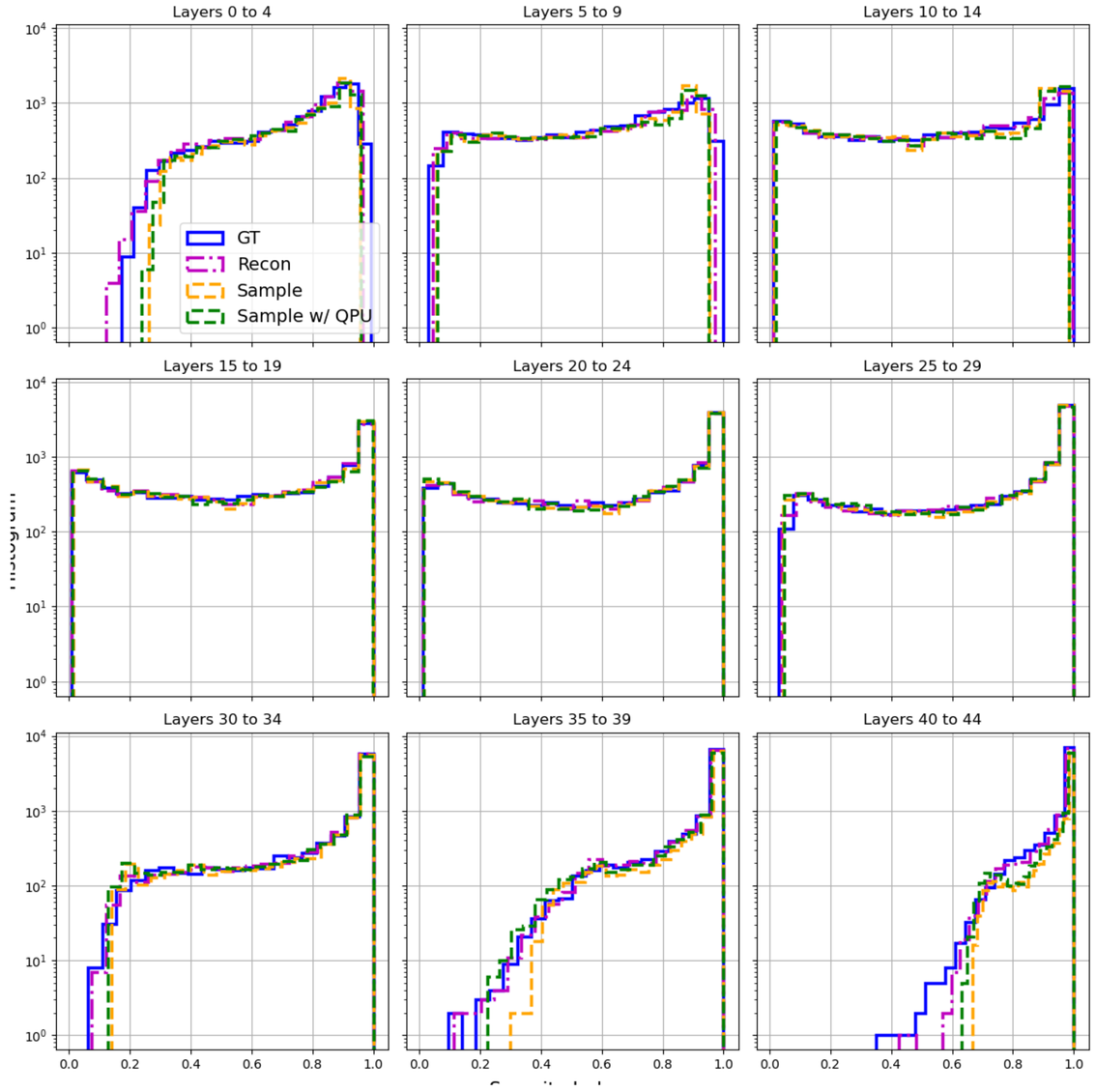


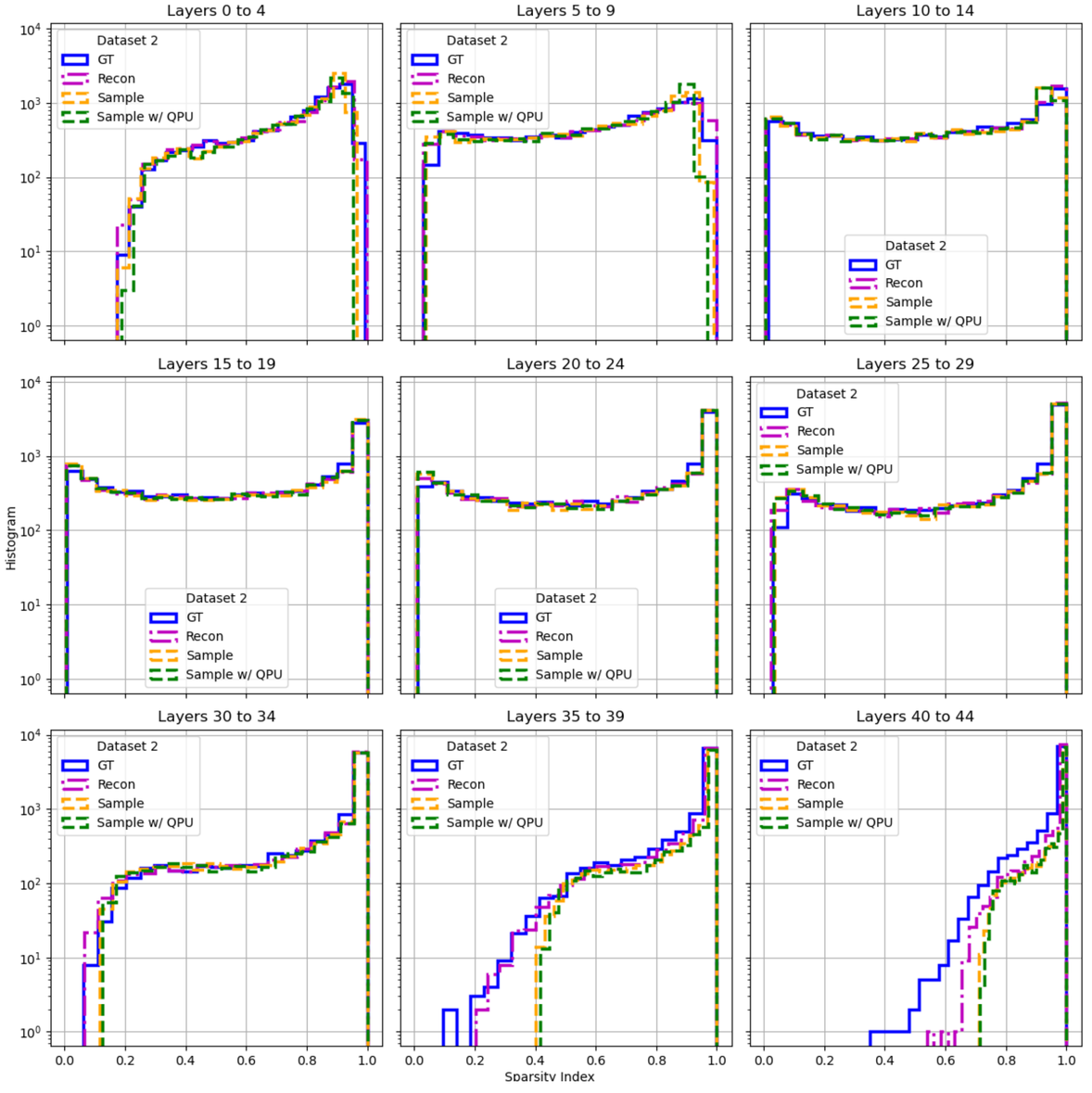


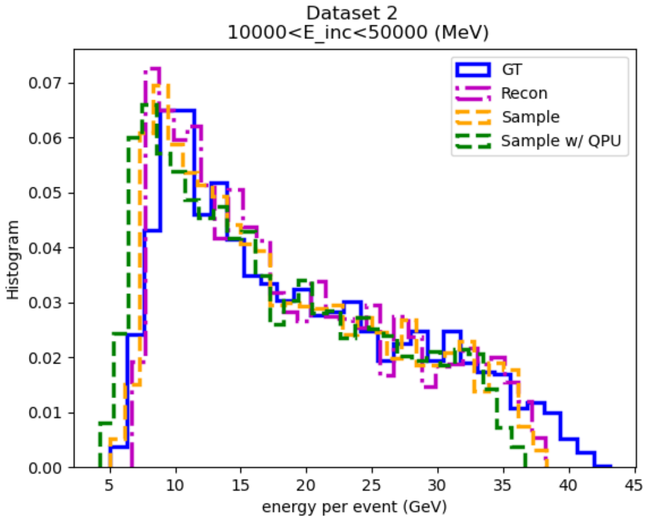




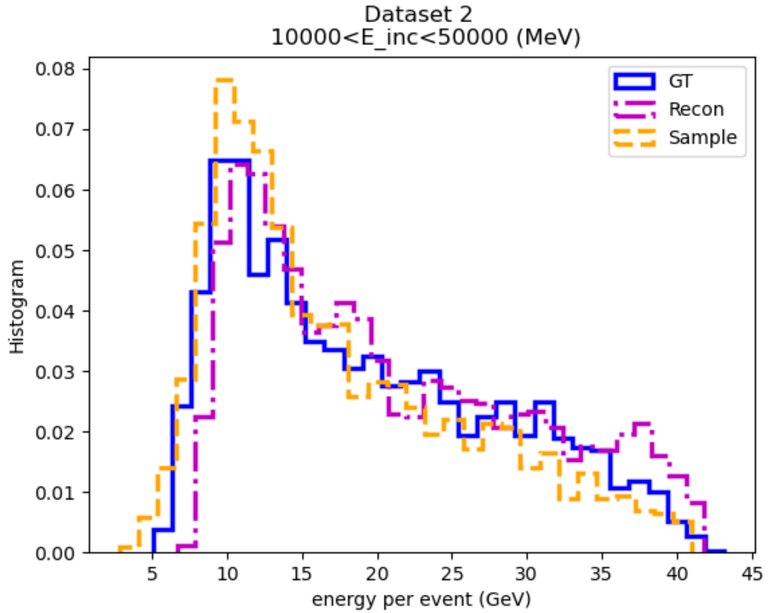

Layers 0 to 4
Layers 5 to 9
Layers 10 to 14
Layers 15 to 19
Layers 20 to 24
Layers 25 to 29
Layers 30 to 34
Layers 35 to 39
Layers 40 to 44
GT
Recon
Sample
Sample w/ QPU
Dataset 2
Histogram
Energy per event (GeV)


Layers 0 to 4
Layers 5 to 9
Layers 10 to 14
Layers 15 to 19
Layers 20 to 24
Layers 25 to 29
Layers 30 to 34
Layers 35 to 39
Layers 40 to 44
GT
Recon
Sample
Sample w/ QPU

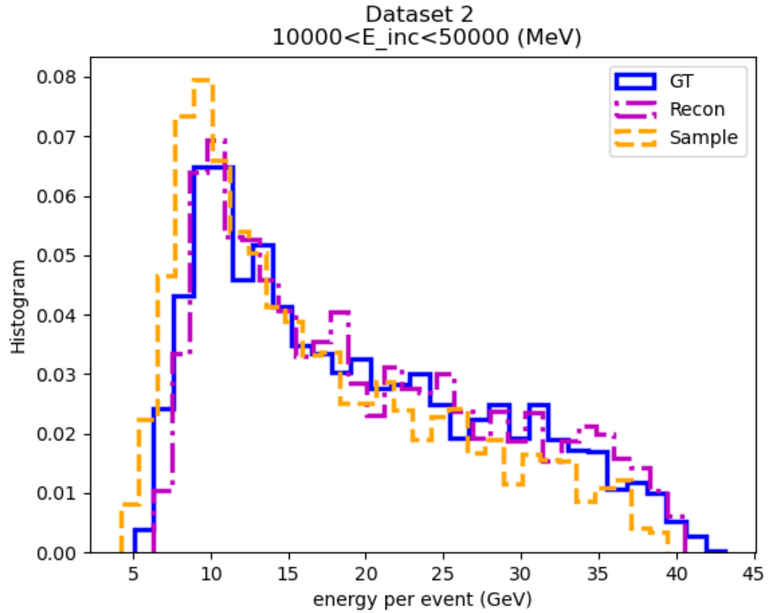

Layers 0 to 4
Layers 5 to 9
Layers 10 to 14
Layers 15 to 19
Layers 20 to 24
Layers 25 to 29
Layers 30 to 34
Layers 35 to 39
Layers 40 to 44
Dataset 2
GT
Recon
Sample
Sample w/ QPU
Histogram
Sparsity Index



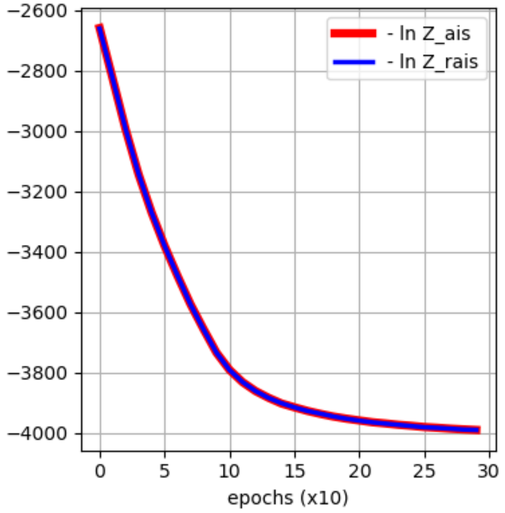
No weights on BCE

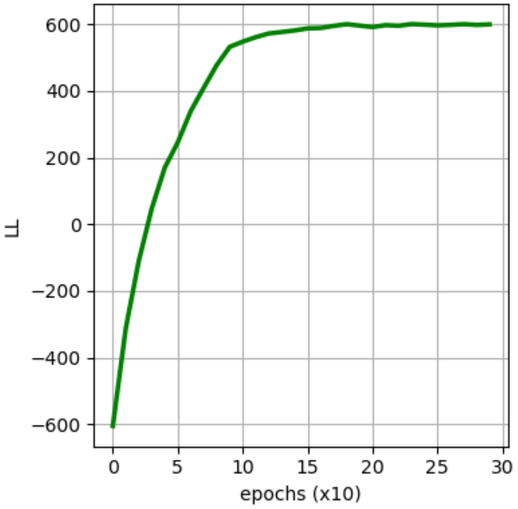


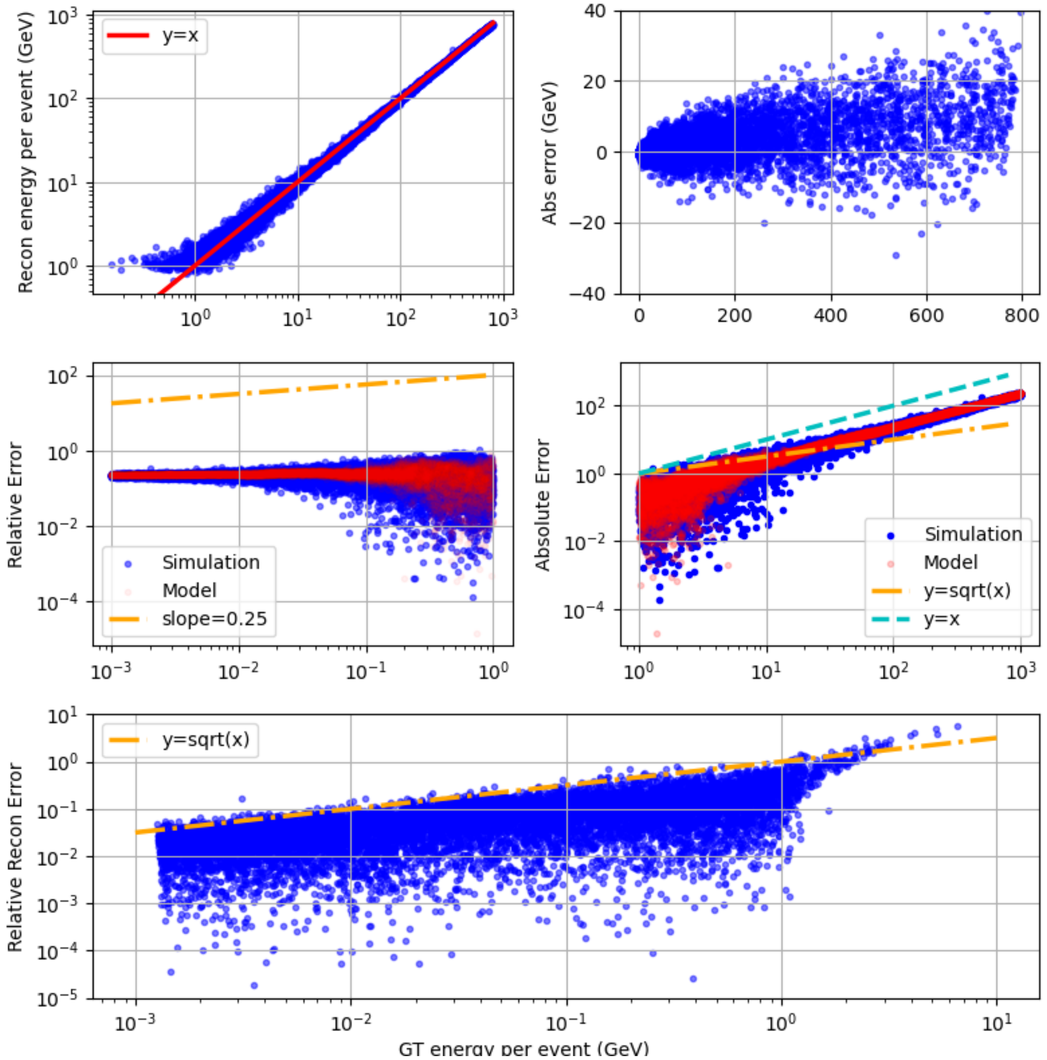
$w(x) = 1 + x$



$w(x) = \sqrt{1+x}$





$$\ln \frac{x-\hat{x}}{x} \propto \frac{1}{2} \ln \frac{1}{x}$$

$$x - \hat{x} \propto \sqrt{x}$$