QVAE w/ Pegasus

Dec 8th

- QVAE
    - Architectures
        - CNN
        - FCN
    - Energy incidence
        - Condition on encoder and decoder
        - Condition on encoder
        - Unconditionalized
        - Modulated energy => Can lead to learning how to modulate more features, position of voxels, angles, etc.
    - Results/metrics
        - Energy histogram
        - Sparsity histogram
        - Conditionalized energy and sparsity histogram (NOT GOOD)
- RBM
    - Topology
        - Chimera-like
        - Pegasus
    - Metrics
        - Energy distribution for encoded and RBM Gibbs samples
        - Zais and Zrais estimates for partition function => log-likelihood of model
    - Dwave
        - Sehmi’s method
        - Fast stein. Not robust but could be helpful?
        - Hao’s method
- Theory. Work in progress

Things to do

- Make code lighter. **(Progress has been made)**
- More work needed on the Positional Encoding **(Testing different approaches)**
- I suggest increasing annealing rates. (**Done**)
- Save loss function data. (**Wandb does this already but the data is weird***)
- Save backup model after each epoch. (**Pending**)
- What defines our best model during training? (**Loss AHEP = AE + Hits + Ent + Pos**)

zany-cloud-260

CNN+ cond
VAE+posEnc on
voxels+scaled data

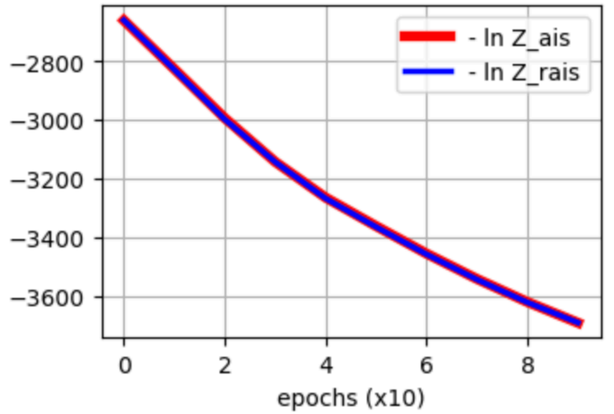
$\{z_i\}_{i=1}^N$: validation dataset

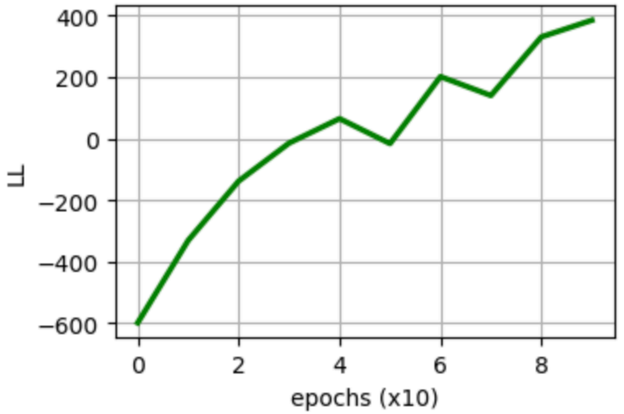
L: $p(\{z_i\}) = \prod_{i=1}^{N} p(z_i) = \frac{1}{Z^N} \prod_{i=1}^{N} e^{-E(z_i)}$

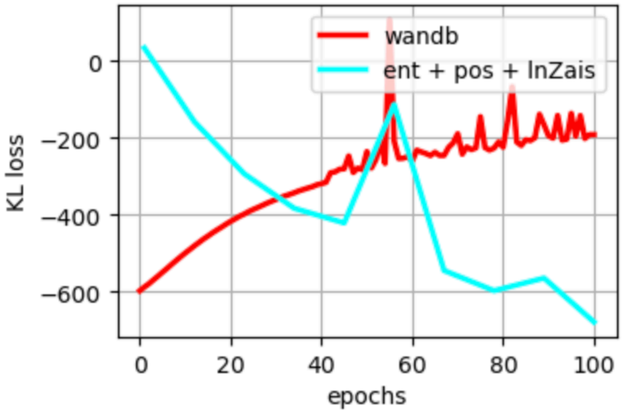
Z : partition function

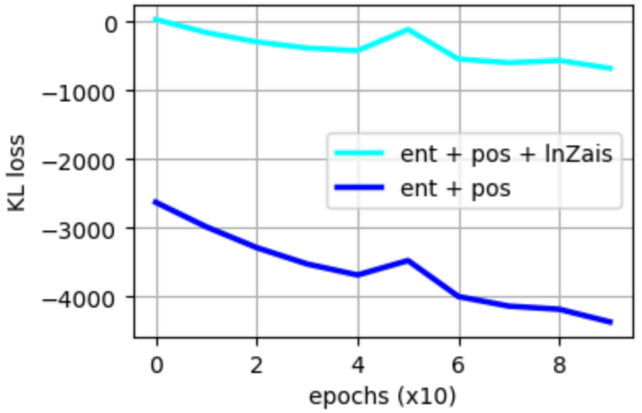
$LL = \langle \ln p(\{z_i\}) \rangle = \frac{1}{N} \left[-\sum_{i=1}^{N} E(z_i) - N \ln Z \right]$

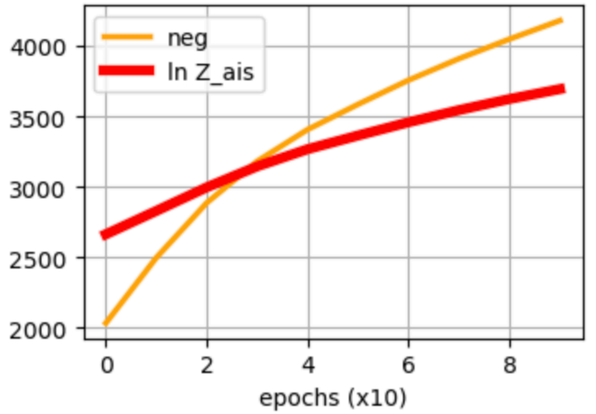
$= - \langle E(z) \rangle_{\{z_i\}} - \ln Z$

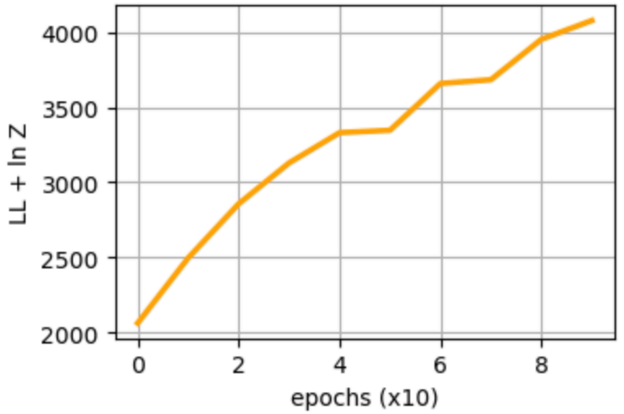


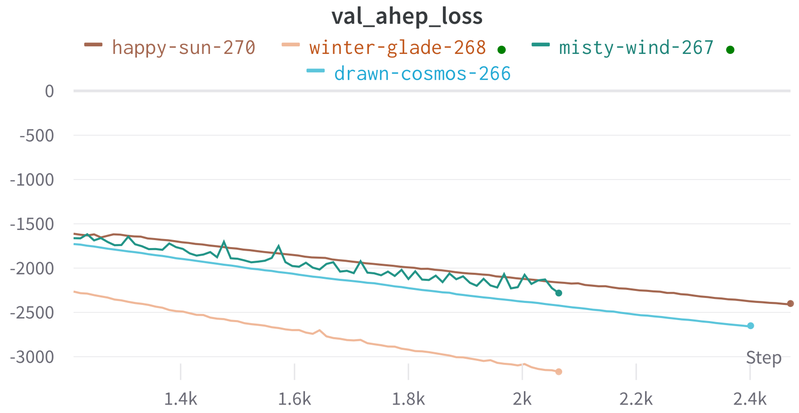


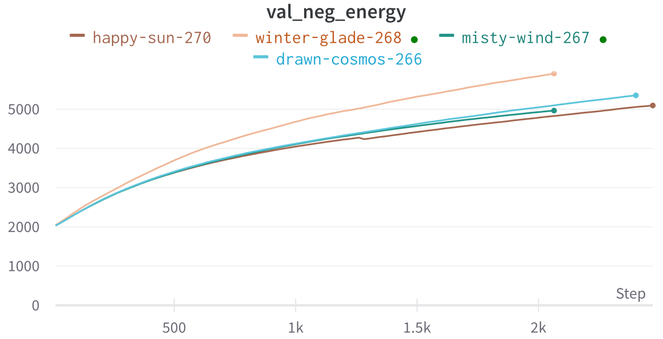


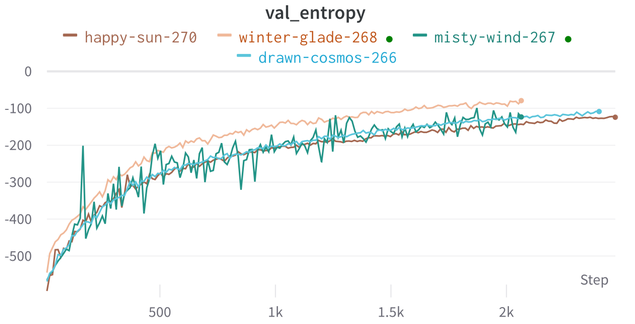


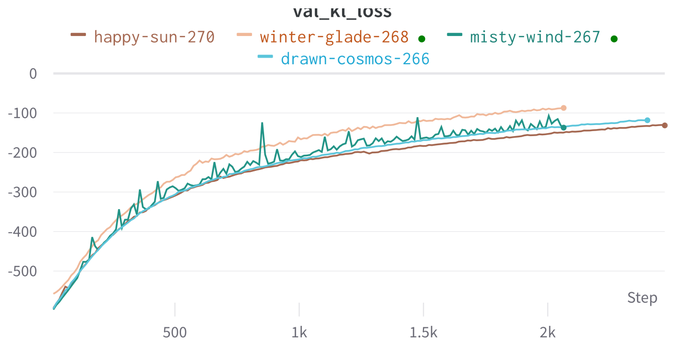


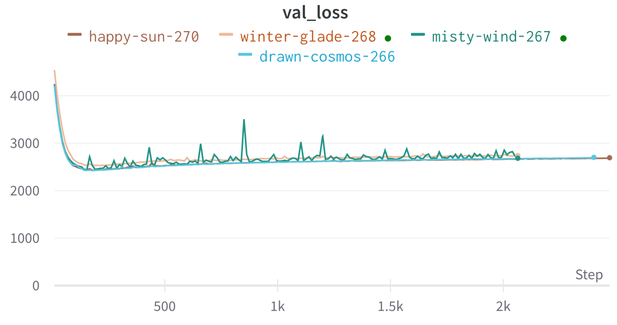


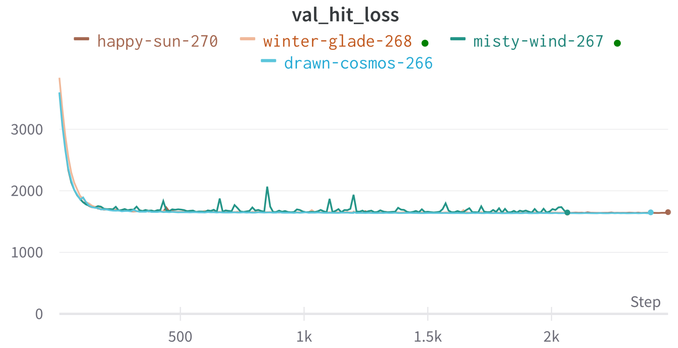

val_ahep_loss
happy-sun-270
winter-glade-268
misty-wind-267
drawn-cosmos-266
0
-500
-1000
-1500
-2000
-2500
-3000
Step
1.4k
1.6k
1.8k
2k
2.2k
2.4k

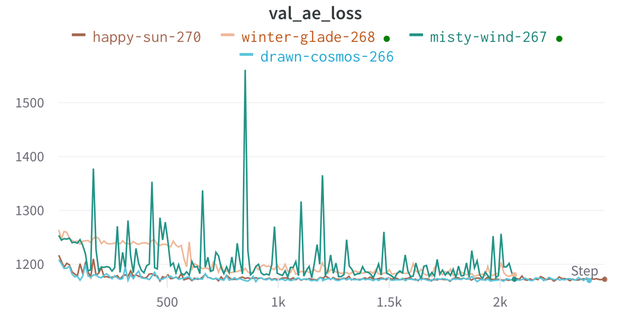

val_neg_energy
happy-sun-270
winter-glade-268
misty-wind-267
drawn-cosmos-266
5000
4000
3000
2000
1000
0
Step
500
1k
1.5k
2k


val_entropy
happy-sun-270
winter-glade-268
misty-wind-267
drawn-cosmos-266
0
-100
-200
-300
-400
-500
Step
500
1k
1.5k
2k


val_kl_loss
happy-sun-270
winter-glade-268
misty-wind-267
drawn-cosmos-266
0
-100
-200
-300
-400
-500
Step
500
1k
1.5k
2k

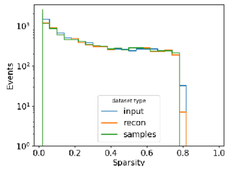

val_loss
happy-sun-270
winter-glade-268
misty-wind-267
drawn-cosmos-266
4000
3000
2000
1000
0
Step
500
1k
1.5k
2k


val_hit_loss
happy-sun-270
winter-glade-268
misty-wind-267
drawn-cosmos-266
3000
2000
1000
0
Step
500
1k
1.5k
2k

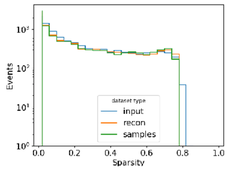

val_ae_loss
happy-sun-270
winter-glade-268
misty-wind-267
drawn-cosmos-266
1500
1400
1300
1200
Step
500
1k
1.5k
2k

val_showers_sparsityHist

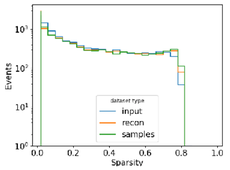
happy-sun-270



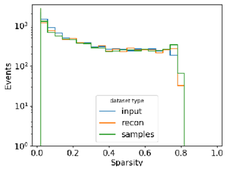
winter-glade-268



misty-wind-267



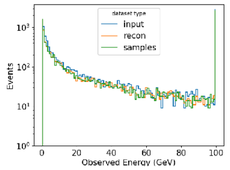
drawn-cosmos-266



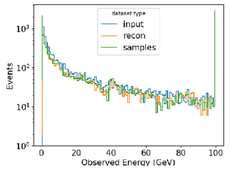
Step 2471

val_totalEnergyHist

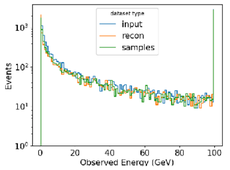
happy-sun-270



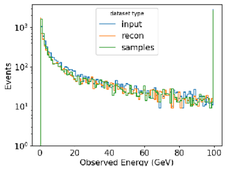
winter-glade-268



misty-wind-267



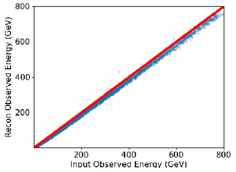
drawn-cosmos-266



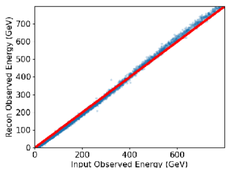
Step 2471

val_totalEnergyHistScatter

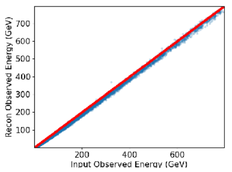
happy-sun-270



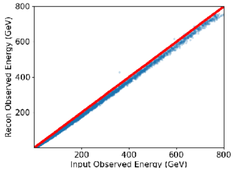
winter-glade-268



misty-wind-267



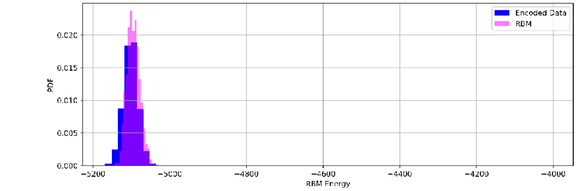
drawn-cosmos-266



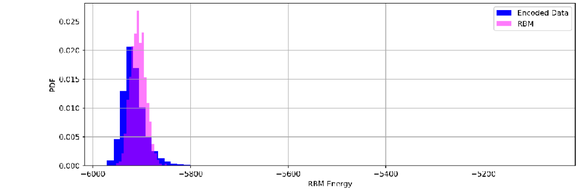
Step 2471

RBM energy

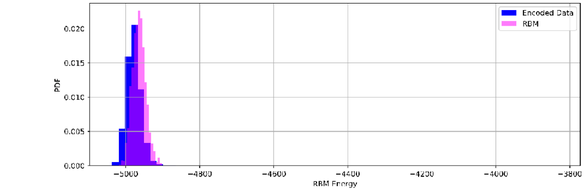
happy-sun-270



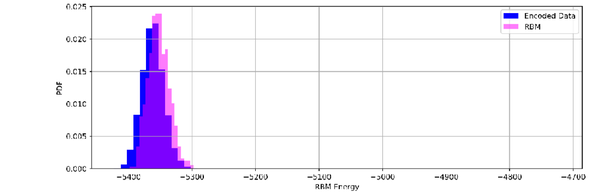
winter-glade-268



misty-wind-267

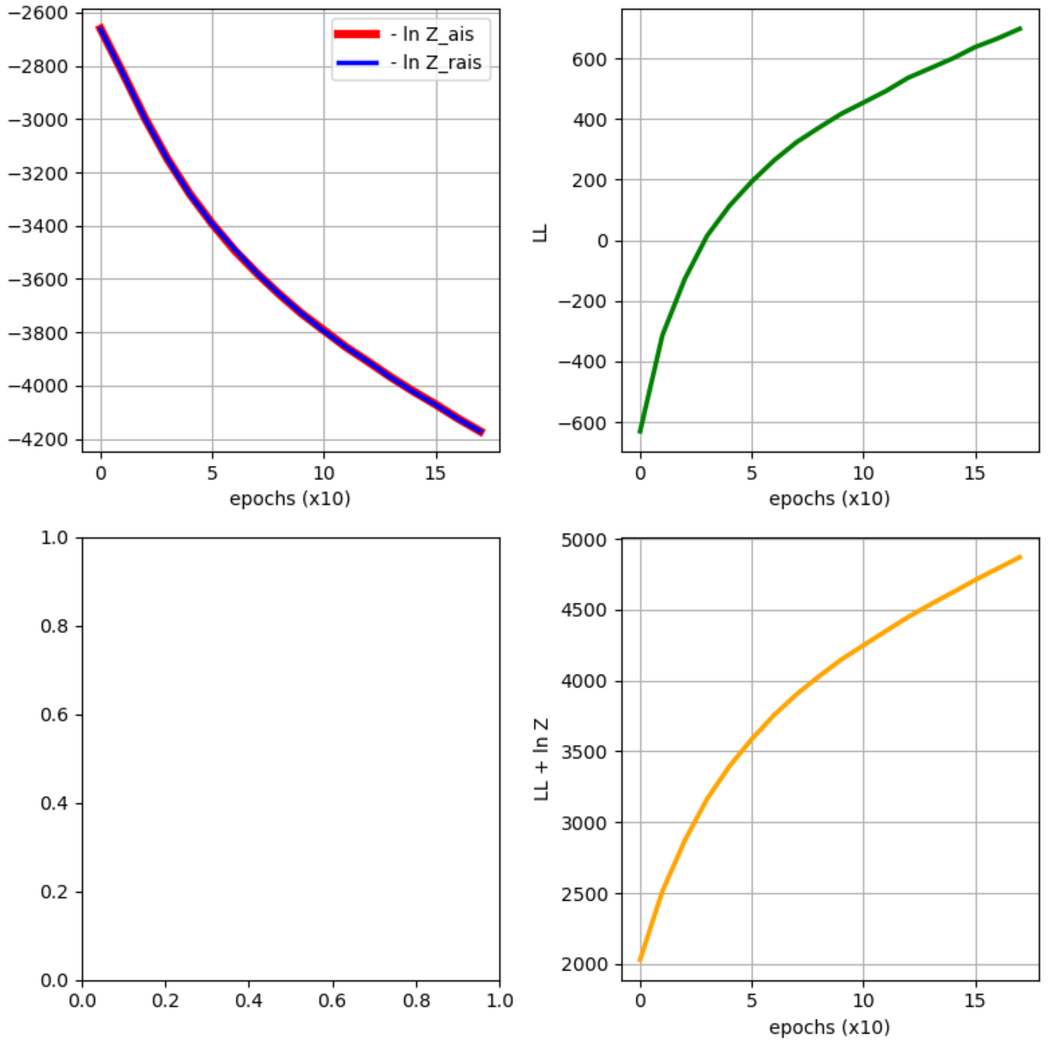


drawn-cosmos-266



Happy-sun-270

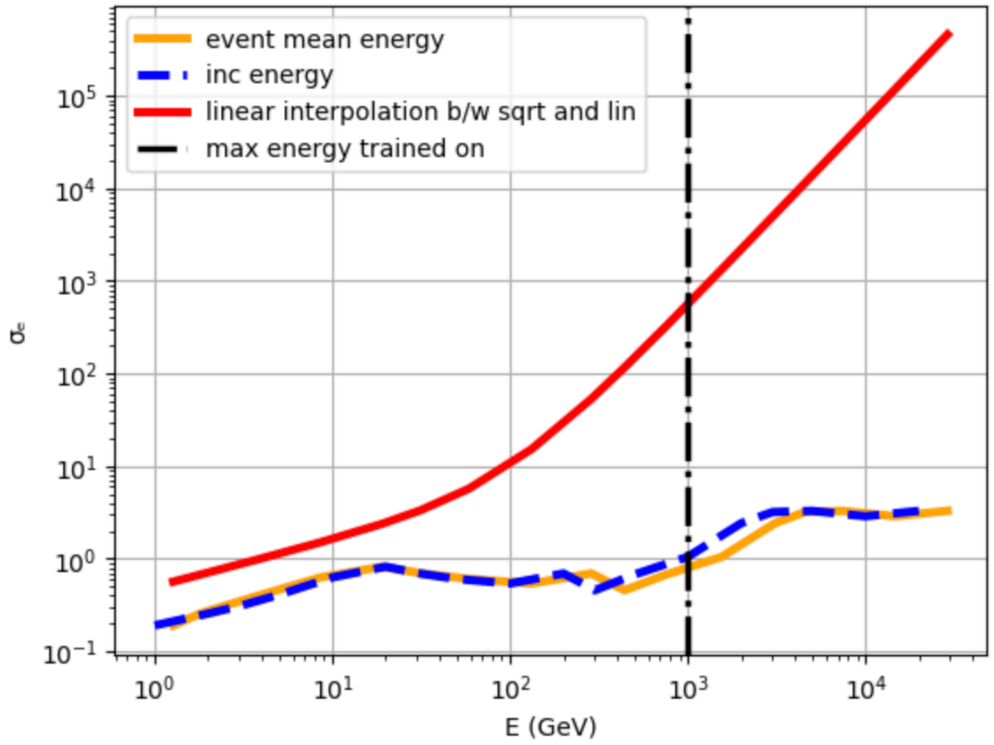
CNN+ cond VAE+posEnc on voxels+scaled data



- Ideally, if all shower particles counted:
- In practice

$$E \propto N \qquad \sigma_E \approx \sqrt{N} \approx \sqrt{E}$$

$$\sigma_E = a\sqrt{E} \oplus bE \oplus c \qquad \frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{c}{E}$$



X : Ground truth ; $\mathcal{E}$: Event energy

$\hat{X}$: Reconstruction ; $\hat{\mathcal{E}}$: Reconstructed Event energy

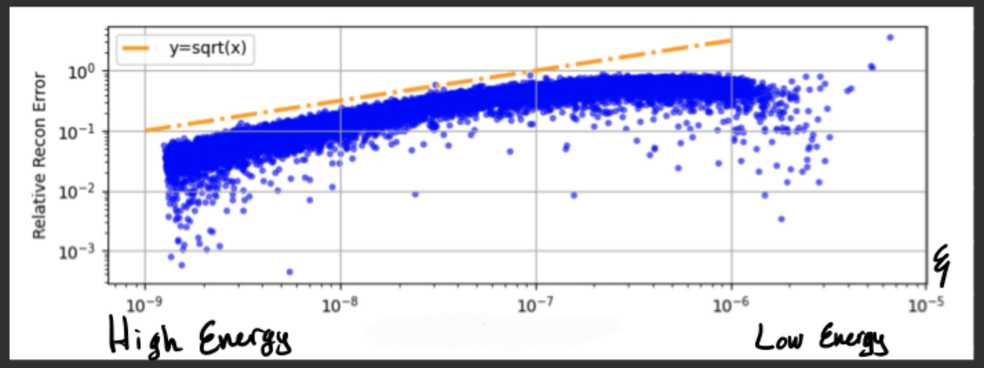
$$\hat{X} = f_{AE}(X)$$

$$\hat{\mathcal{E}} = \sum_{i \in Voxels} \hat{X}_i$$

$$\mathcal{E} = \sum_{i \in Voxels} X_i$$

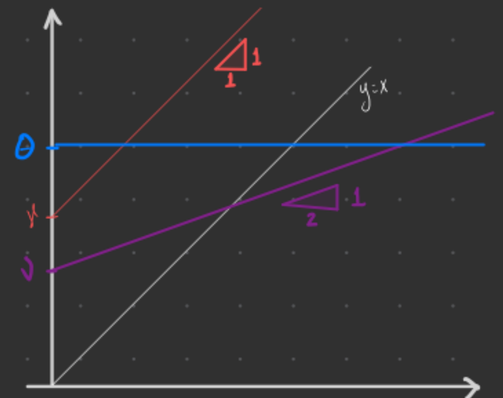
Relative Error

$$Rel(\hat{\mathcal{E}}, \mathcal{E}) = \frac{|\mathcal{E} - \hat{\mathcal{E}}|}{\mathcal{E}}$$



Suppose we plot:

$\ln Rel(\hat{\mathcal{E}}, \mathcal{E})$ vs $\ln \frac{1}{\mathcal{E}}$



Possible scenarios

$|\mathcal{E} - \hat{\mathcal{E}}| = Const = \gamma$

$|\mathcal{E} - \hat{\mathcal{E}}| = \delta\sqrt{\mathcal{E}}$

$|\mathcal{E} - \hat{\mathcal{E}}| = \theta\mathcal{E}$

$|\mathcal{E} - \hat{\mathcal{E}}| = \alpha\mathcal{E}^{\delta}$

High Energy

Low Energy

$\xi = \ln \frac{1}{\mathcal{E}}$

- For High Energy, the error goes as $\sim\sqrt{E}$
- For Low Energy, the error goes as $\sim E$

- Training QVAE on QAs **protocol**