TRIUMF | THE UNIVERSITY OF BRITISH COLUMBIA | PI | DTRC-NRC | UNIVERSITY of VIRGINIA

Calo4pQVAE: Progress and updates

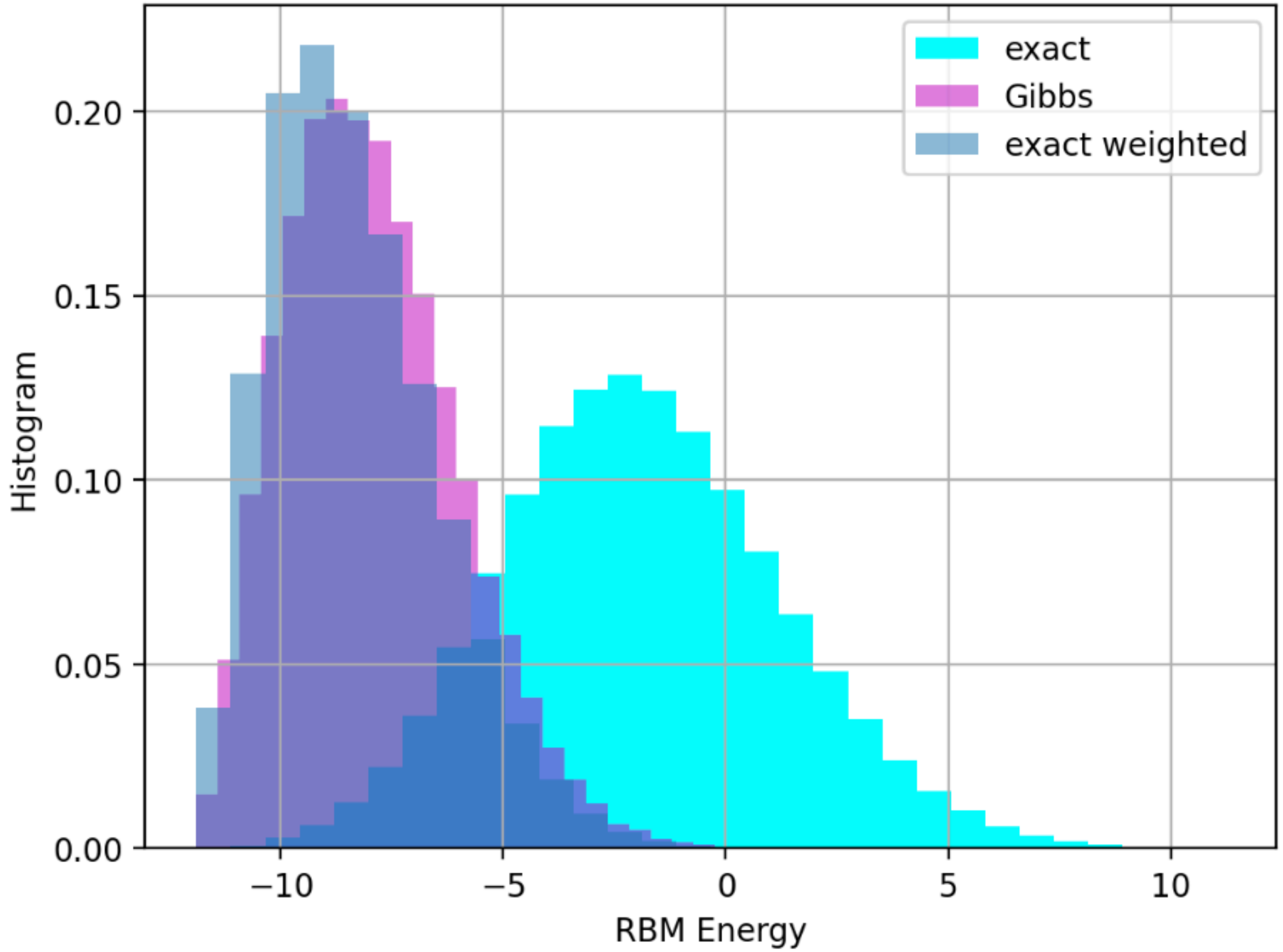
Jan 31 2025

Small RBM

Small RBM (Zephyr top)

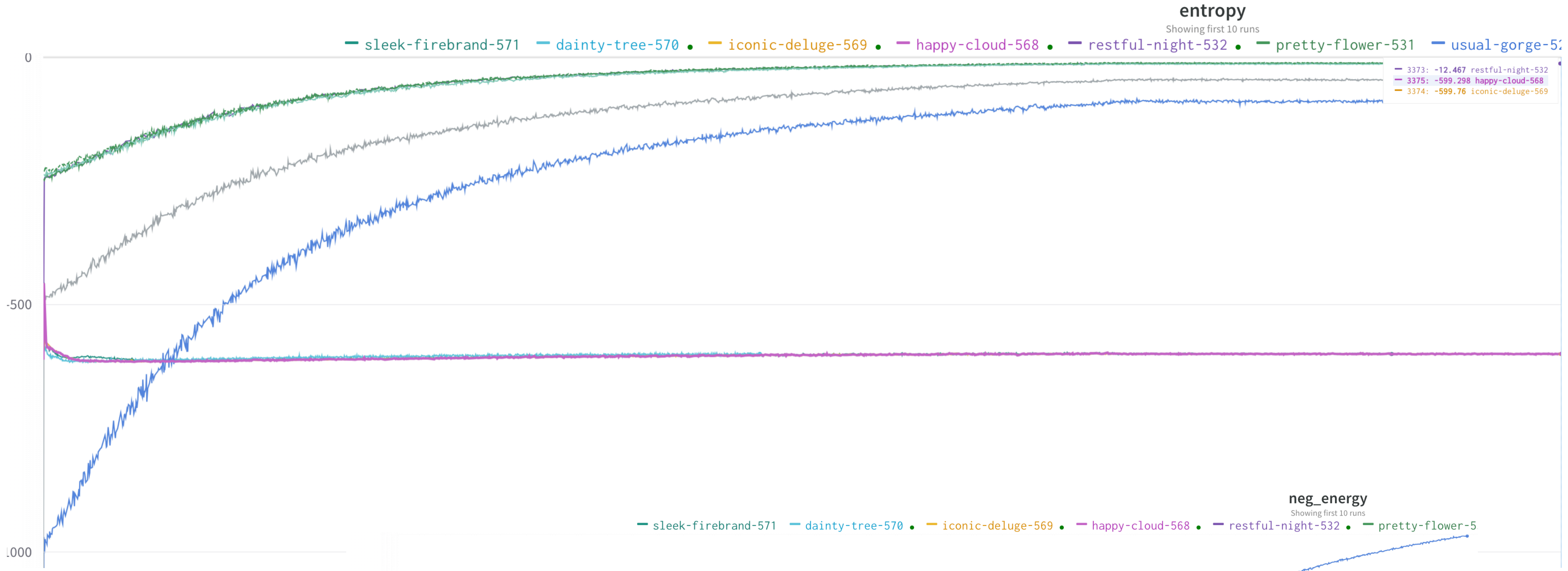
7x7x7x7. Weights and biases samples from N(0,1)

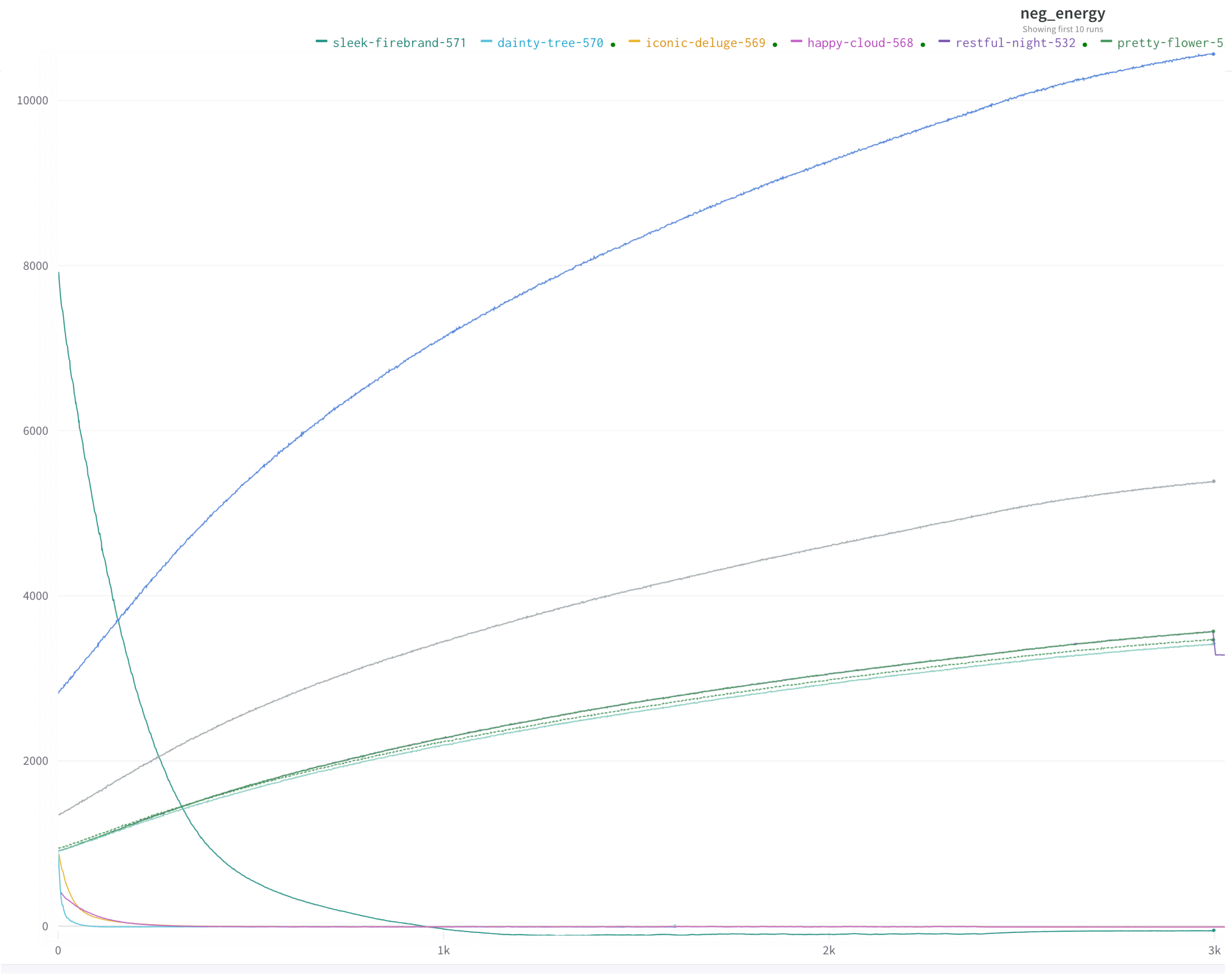
- Weighing the samples fixed the discrepancy.
- BGS and QA can accurately estimate Statistical averages.
- We can train models *reliably* with BGS and QA.

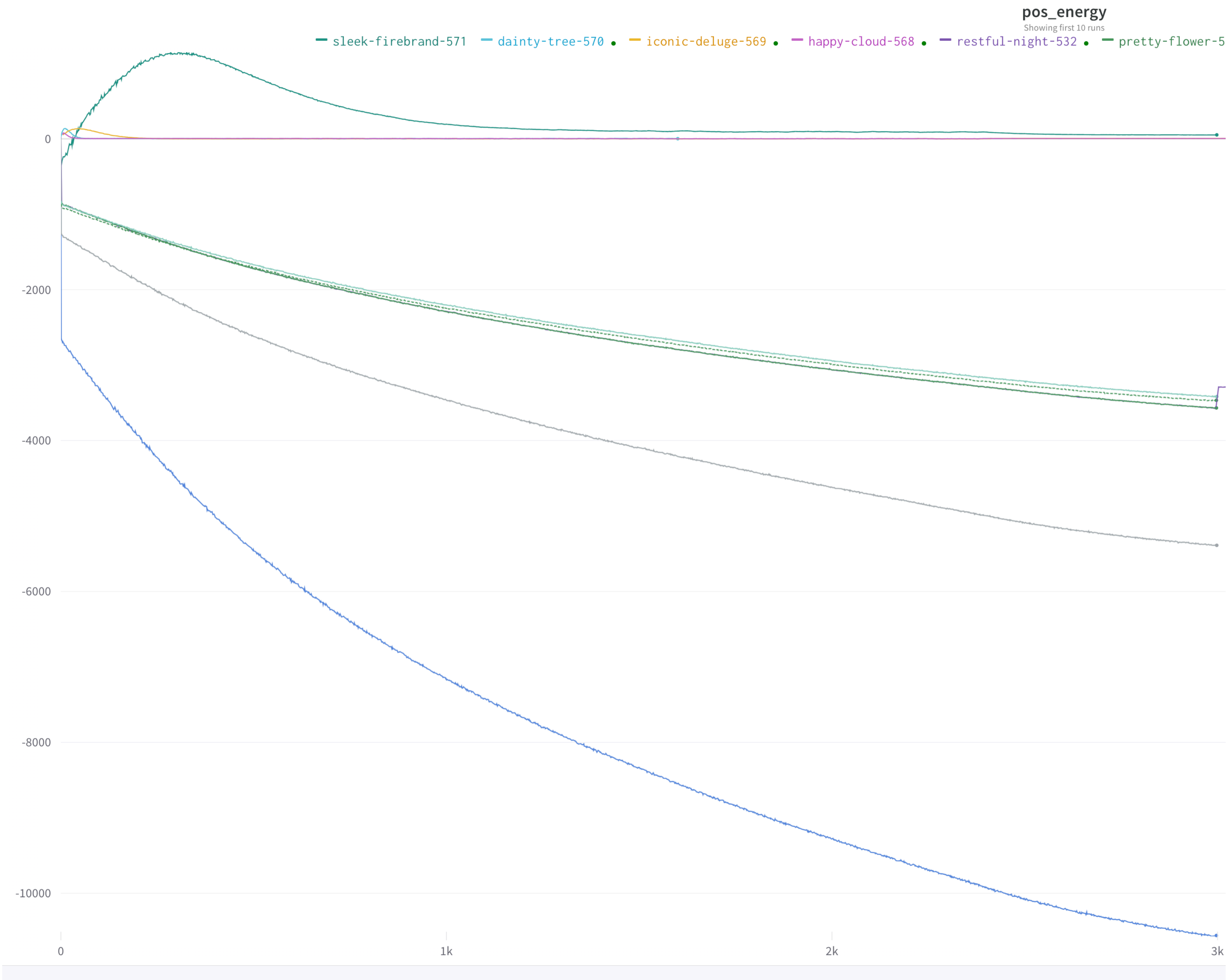


New models

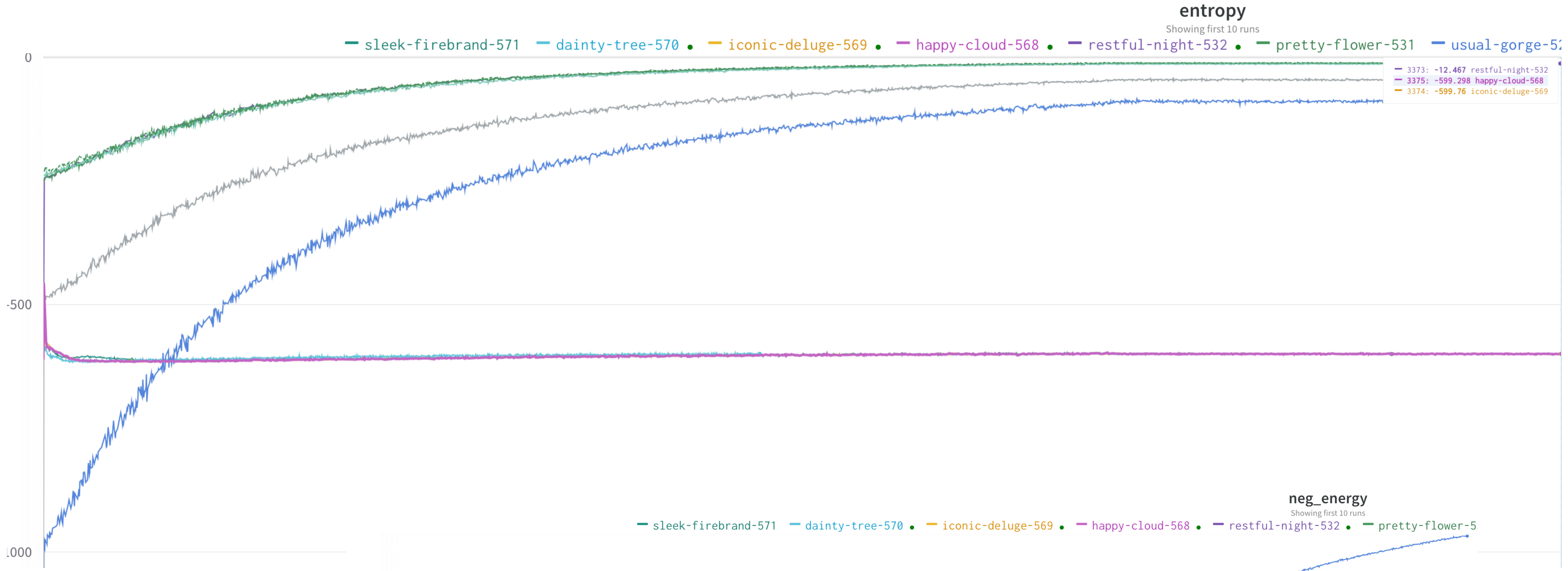
Encoder Entropy & RBM log-likelihood

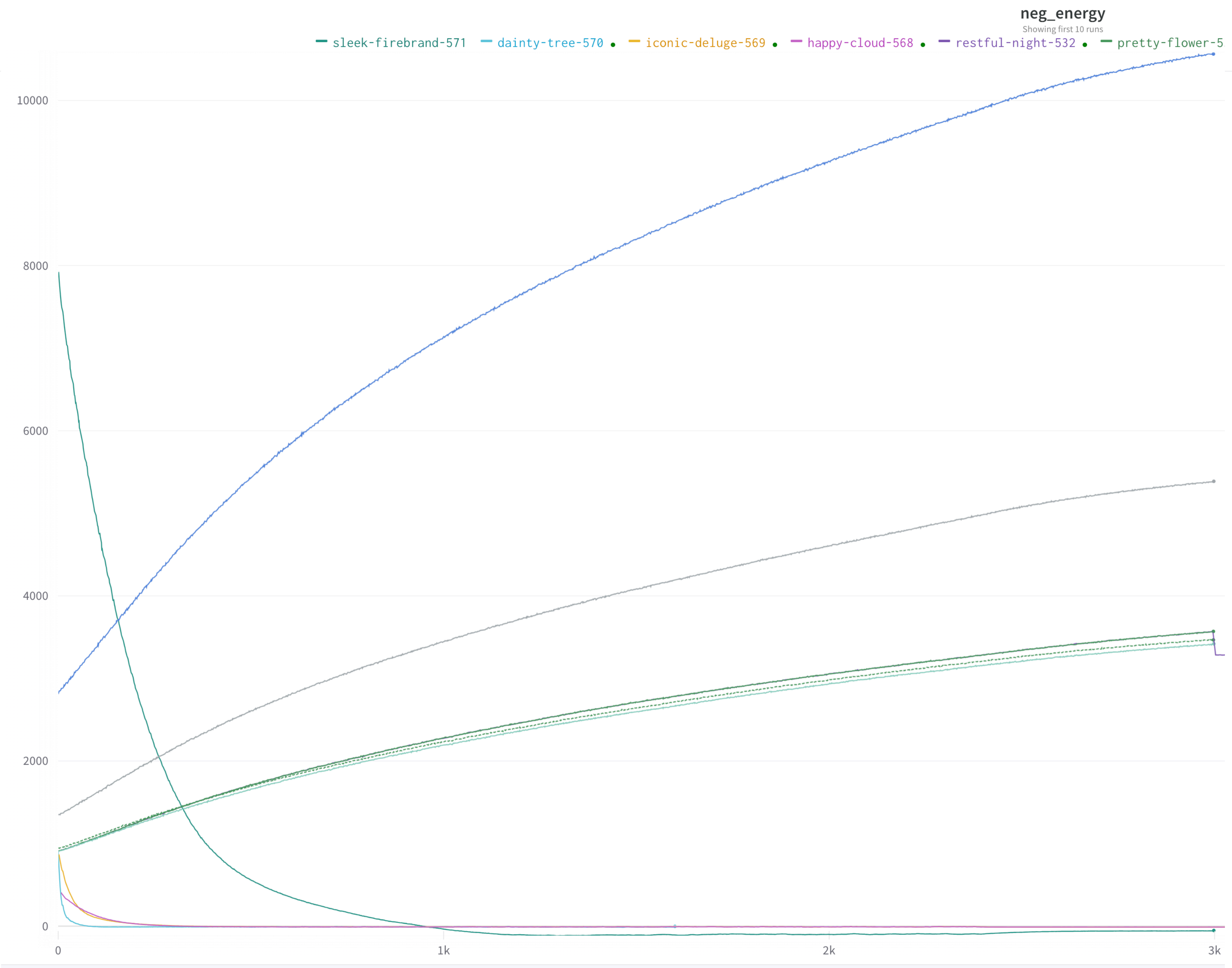


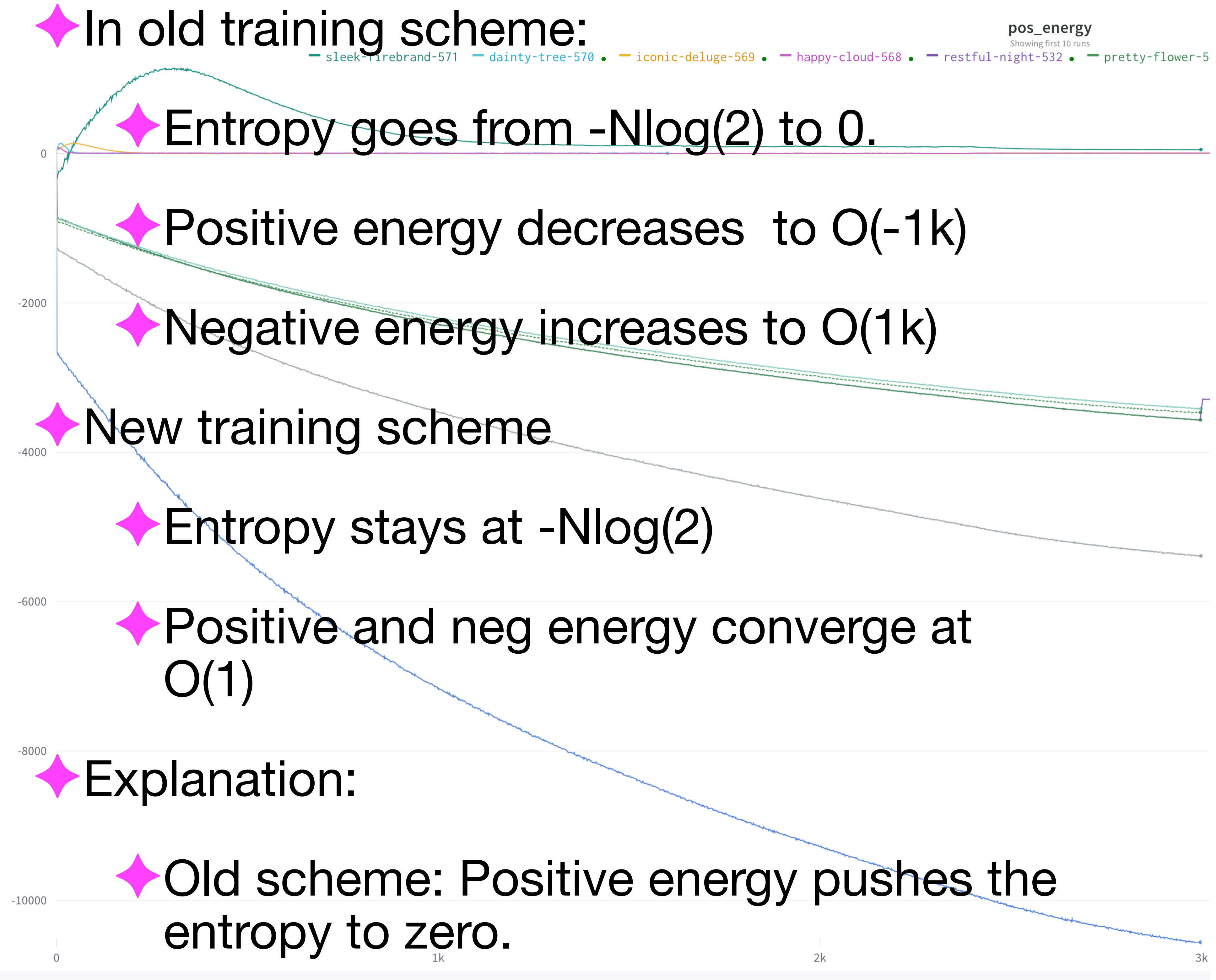




Encoder Entropy & RBM log-likelihood







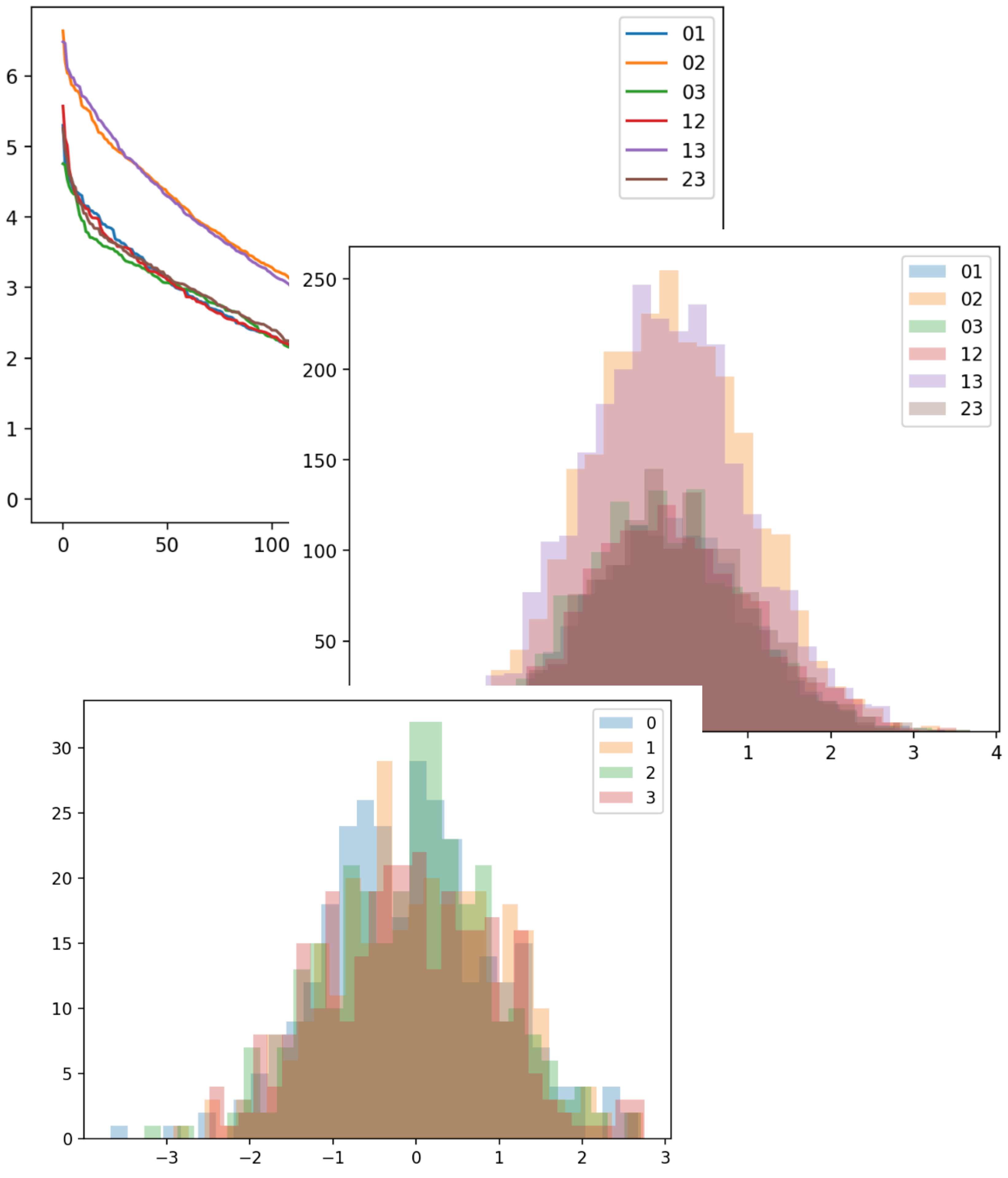
- In old training scheme:
  - Entropy goes from -Nlog(2) to 0.
  - Positive energy decreases to O(-1k)
  - Negative energy increases to O(1k)
- New training scheme
  - Entropy stays at -Nlog(2)
  - Positive and neg energy converge at O(1)
- Explanation:
  - Old scheme: Positive energy pushes the entropy to zero.
  - New scheme: Entropy stays flat

Encoder Entropy & RBM log-likelihood

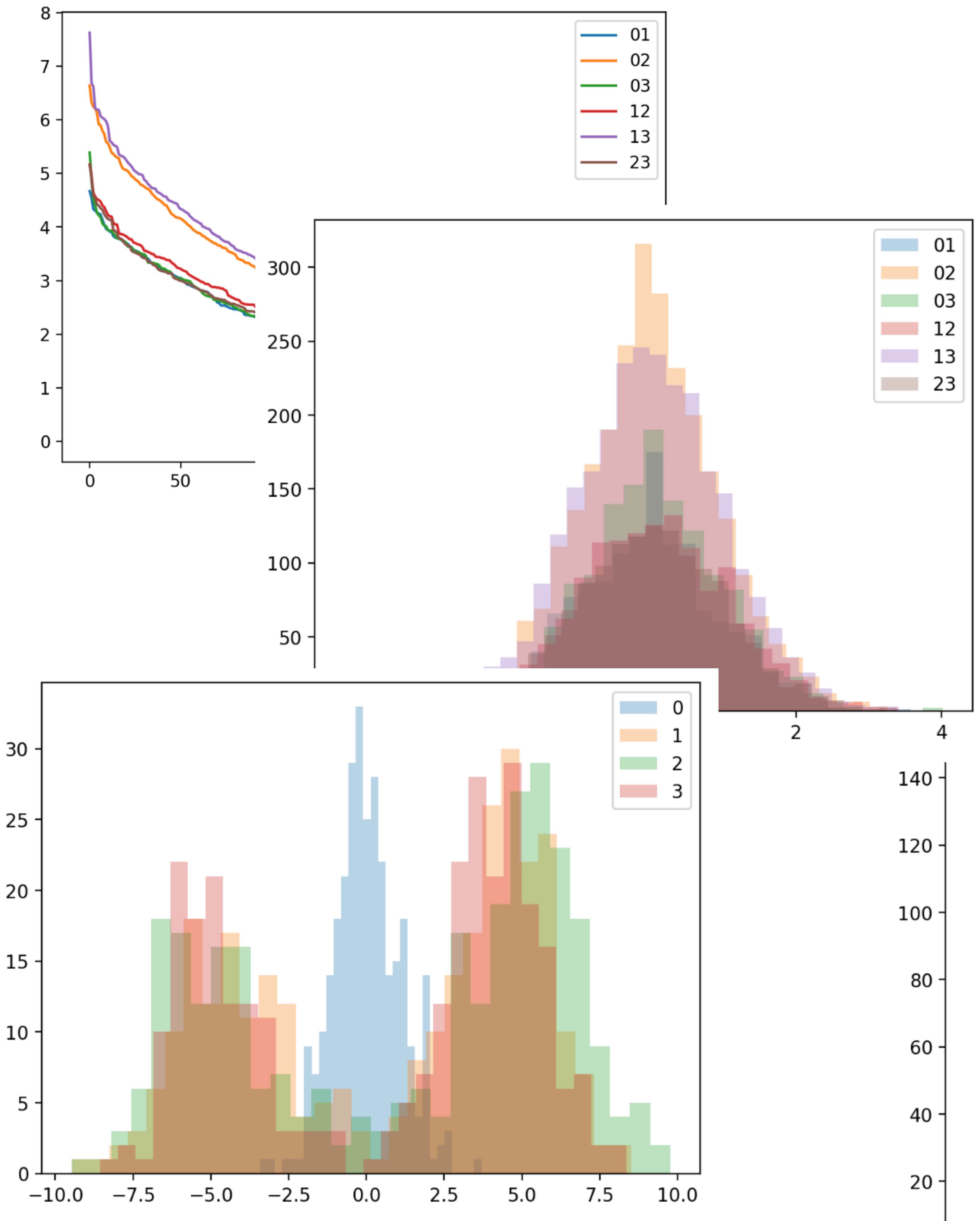
- In latent diffusion models, the encoder-decoder are first trained and the diffusion model is trained afterwards.
- In CaloQVAE, we can train the usual way (as we did in the PRX draft). Afterwards, trying the RBM using the centred gradient approach.

SVD

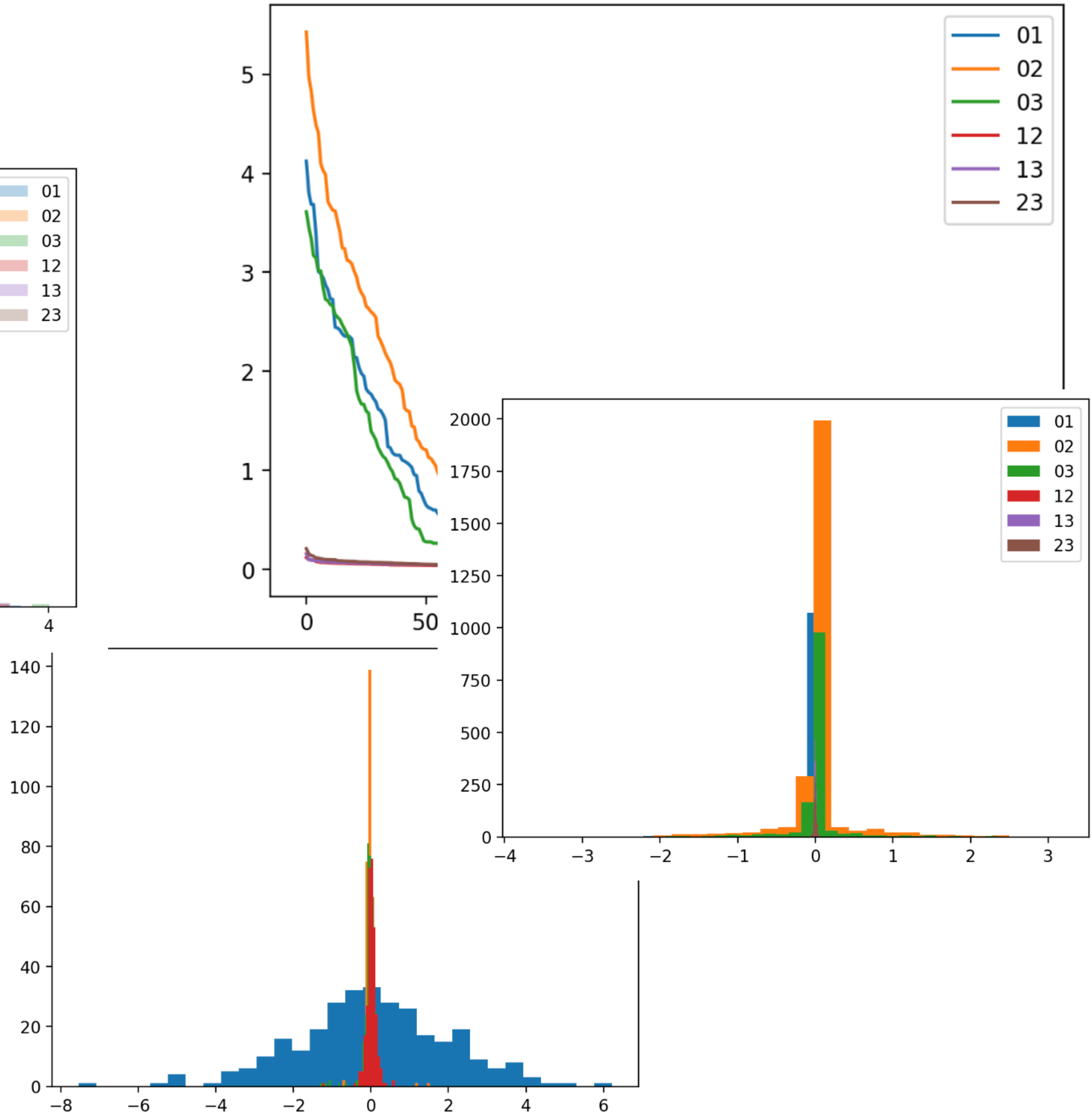
Initialized Rdm model



Trained model (old scheme)

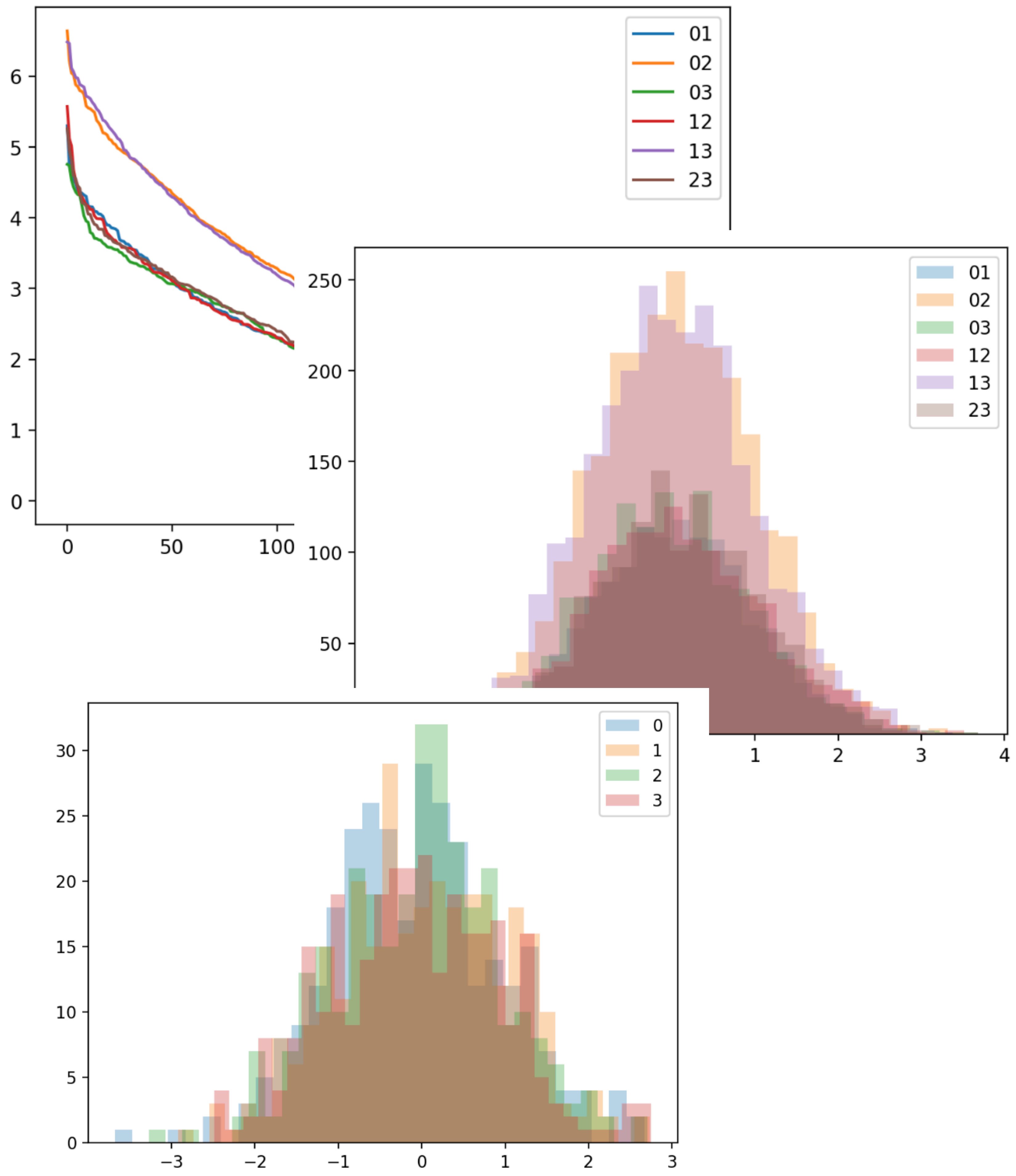


Trained model

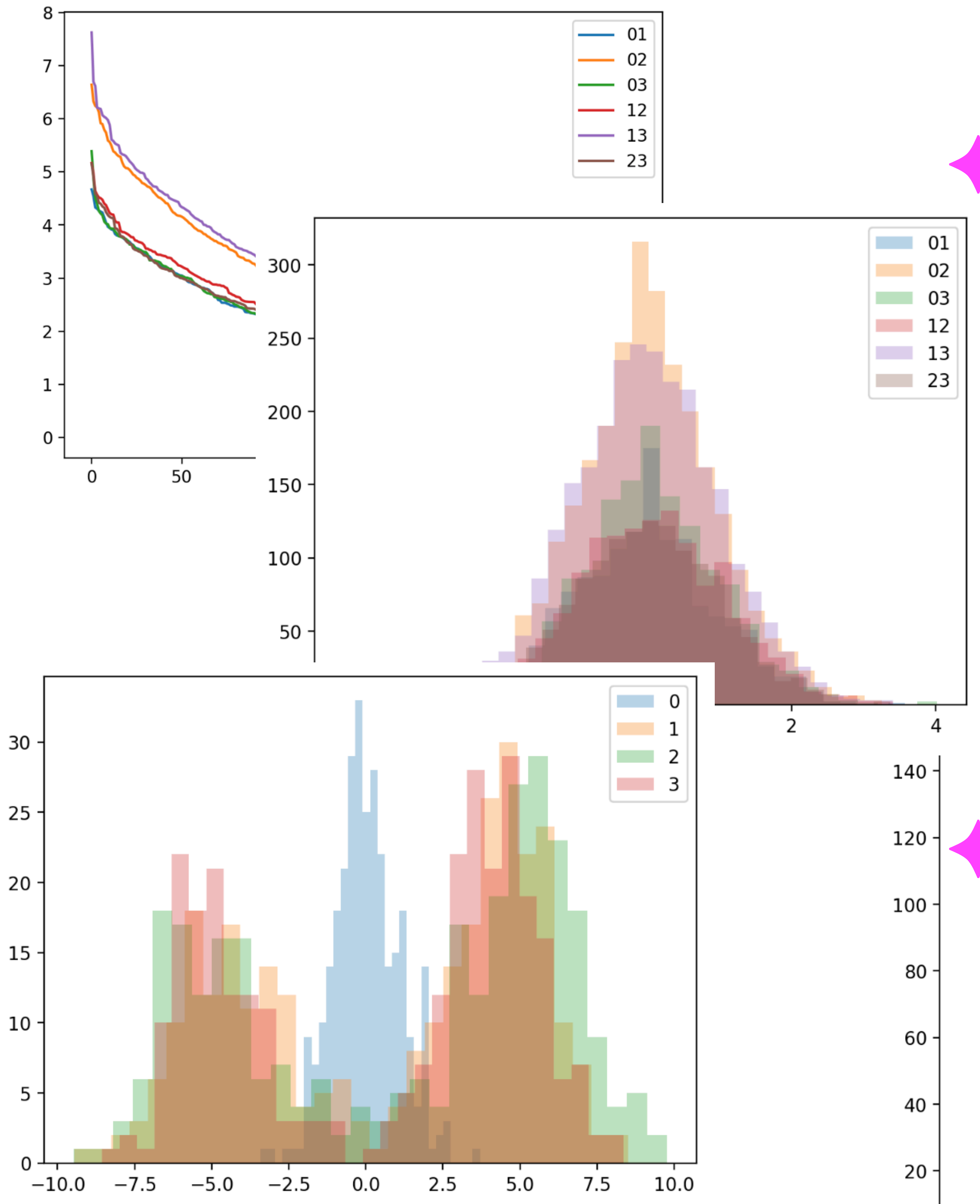


SVD

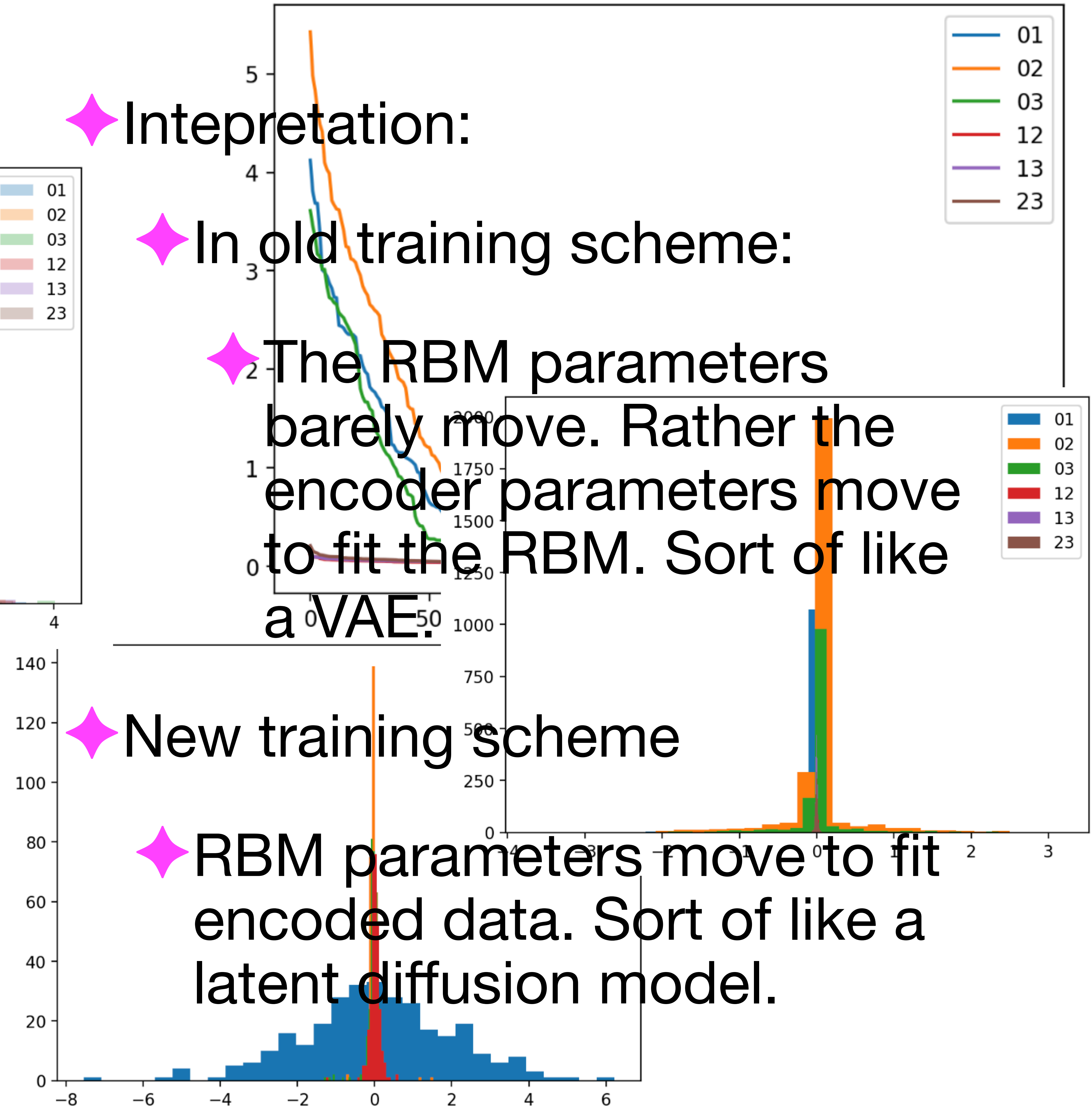
Initialized Rdm model



Trained model (old scheme)



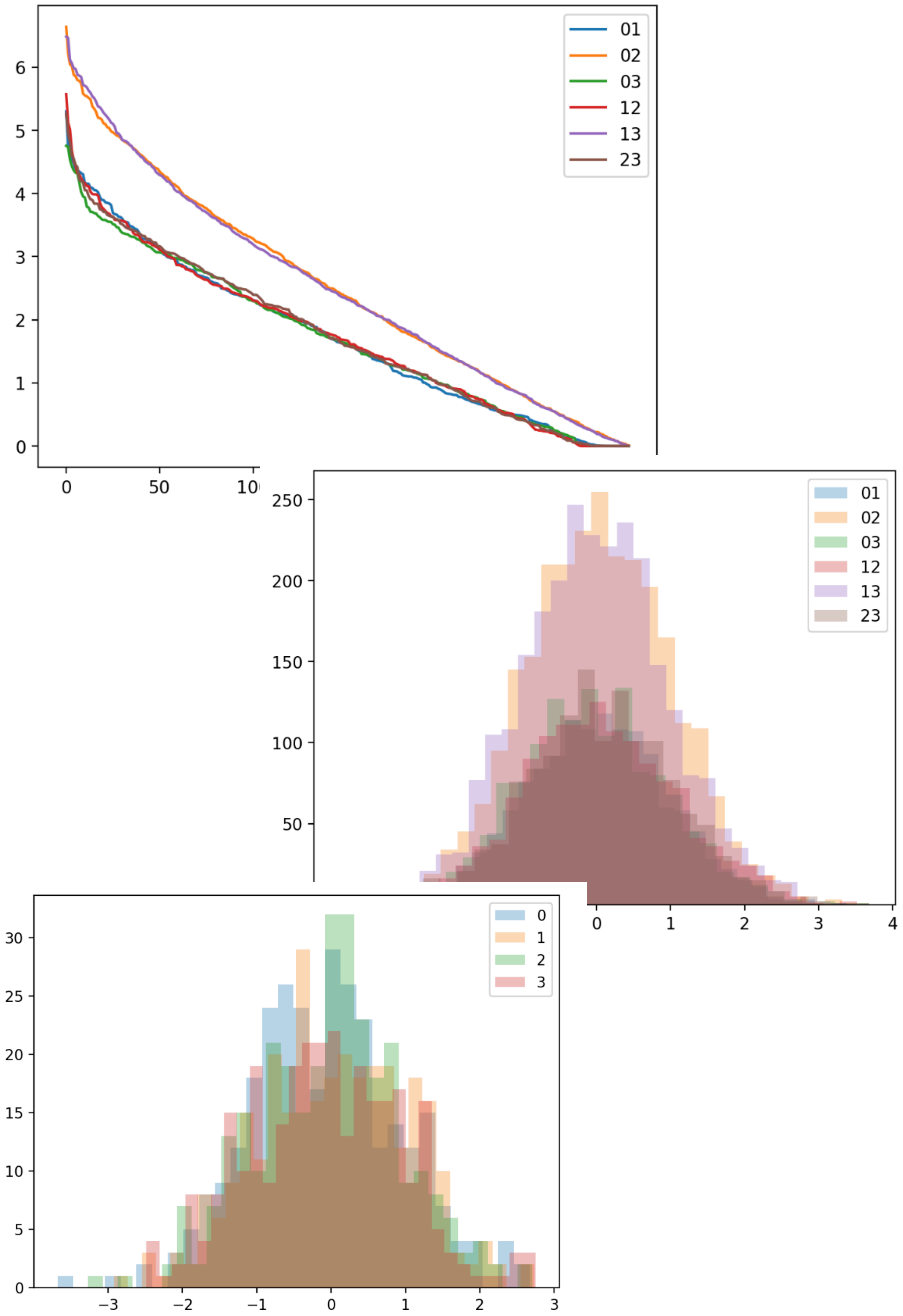
Trained model



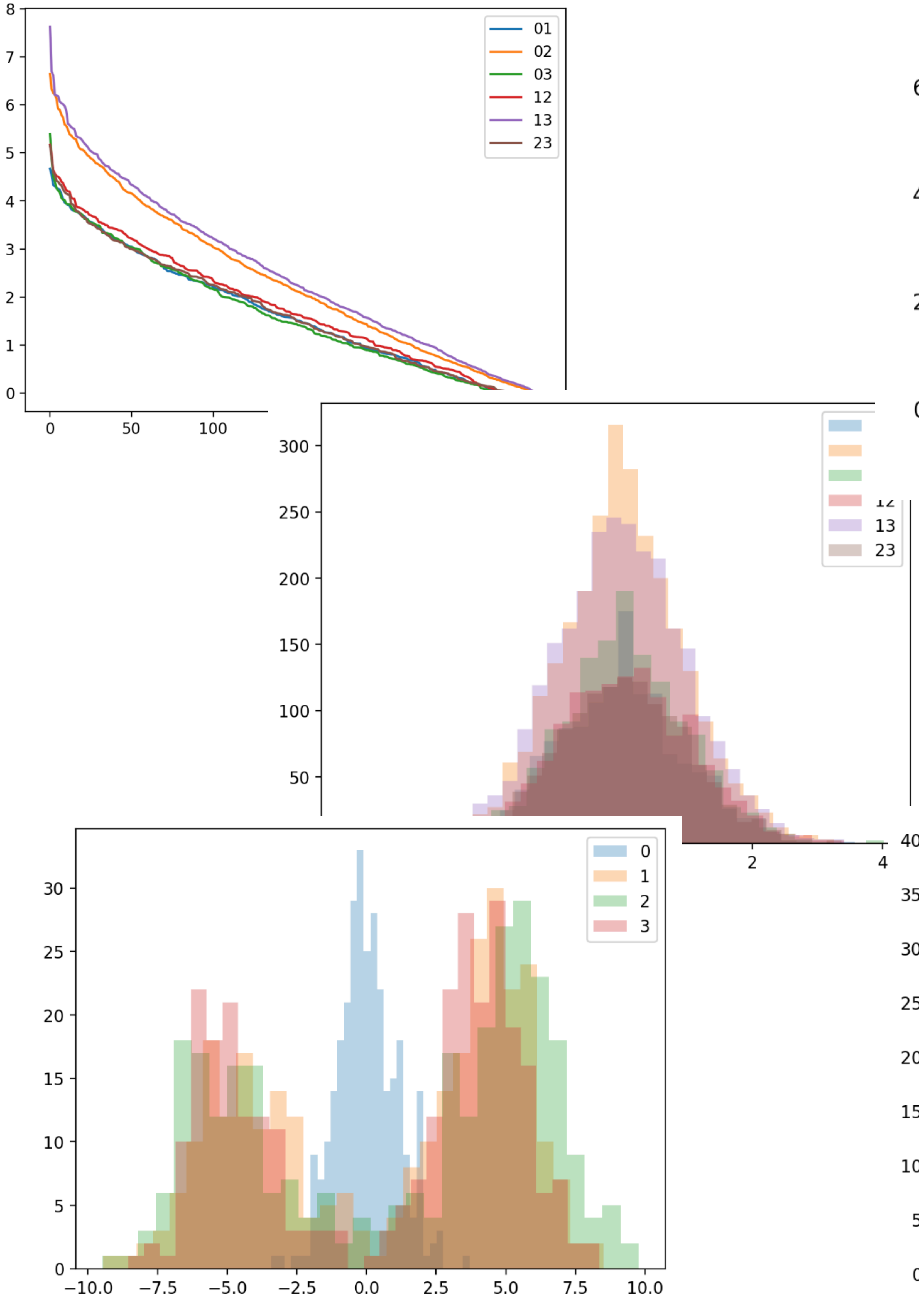
- Intepretation:
  - In old training scheme:
    - The RBM parameters barely move. Rather the encoder parameters move to fit the RBM. Sort of like a VAE.
- New training scheme
  - RBM parameters move to fit encoded data. Sort of like a latent diffusion model.

SVD

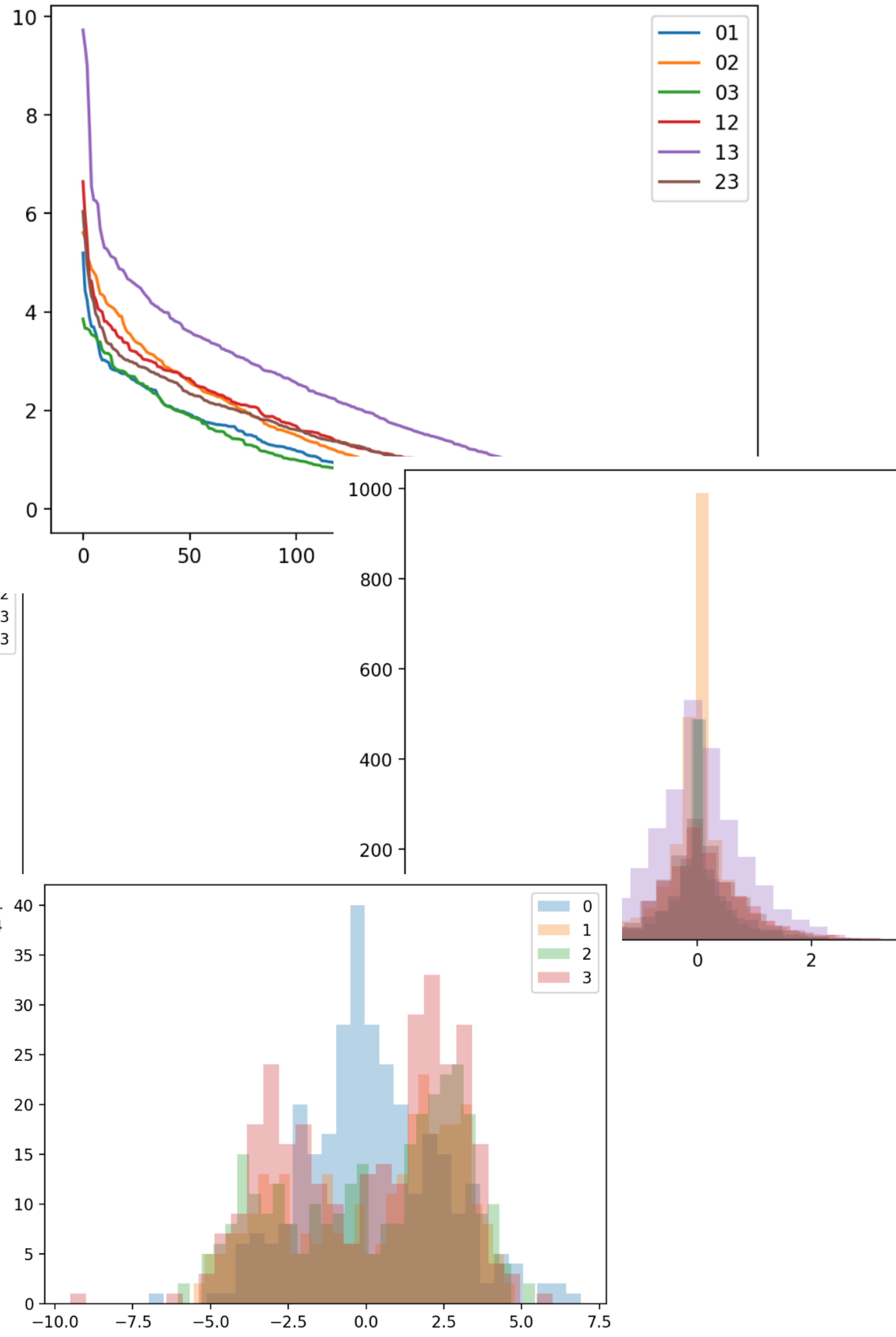
Initialized Rdm model



Trained model (old scheme)



Trained model



Loss functions

$$\langle \sum_{i=1}^{n} (\chi_i(\theta) \cdot \Theta(x_i) - x_i)^2 \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} - \langle \Theta(x_i) \ln(p_{\xi_i}(\theta)) + (1 - \Theta(x_i)) \ln(1 - p_{\xi_i}(\theta)) \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} + \langle \ln q_\phi(\boldsymbol{z}|\boldsymbol{x}) \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} + \langle E(\boldsymbol{z}) \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} + \ln Z$$

MSE | Hit loss | Entropy | Positive Energy | Negative Energy

Loss functions

$$\langle \sum_{i=1}^{n} (\chi_i(\theta) \cdot \Theta(x_i) - x_i)^2 \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} - \langle \Theta(x_i) \ln(p_{\xi_i}(\theta)) + (1-\Theta(x_i)) \ln(1 - p_{\xi_i}(\theta)) \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} + \langle \ln q_\phi(\boldsymbol{z}|\boldsymbol{x}) \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} + \langle E(\boldsymbol{z}) \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} + \ln Z$$

Loss functions

Old training scheme

$$\langle \sum_{i=1}^{n} (\chi_i(\theta) \cdot \Theta(x_i) - x_i)^2 \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} - \langle \Theta(x_i) \ln(p_{\xi_i}(\theta)) + (1 - \Theta(x_i)) \ln(1 - p_{\xi_i}(\theta)) \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} + \langle \ln q_\phi(\boldsymbol{z}|\boldsymbol{x}) \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} + \langle E(\boldsymbol{z}) \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} + \ln Z$$

Loss functions

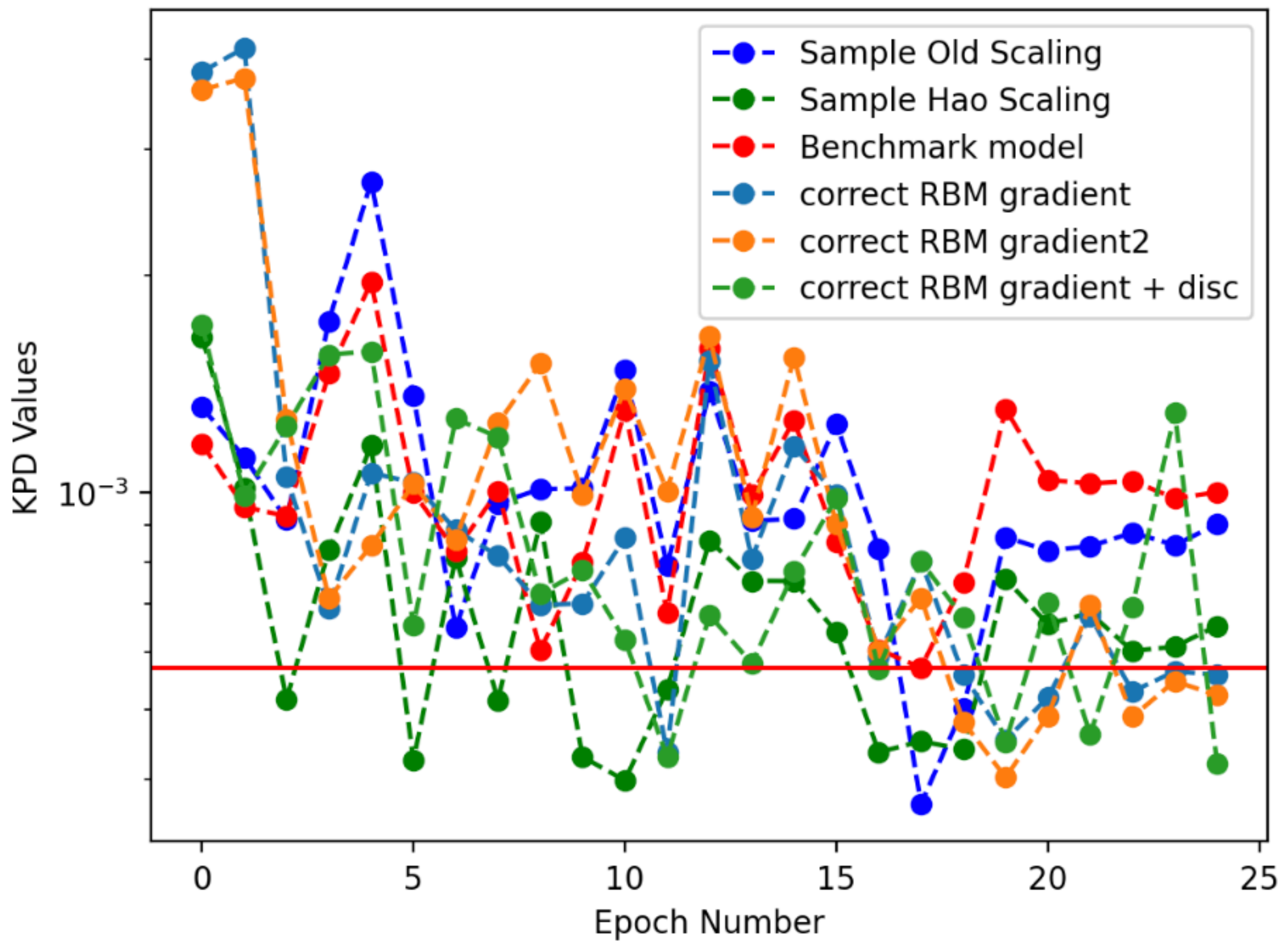
Training scheme from last week

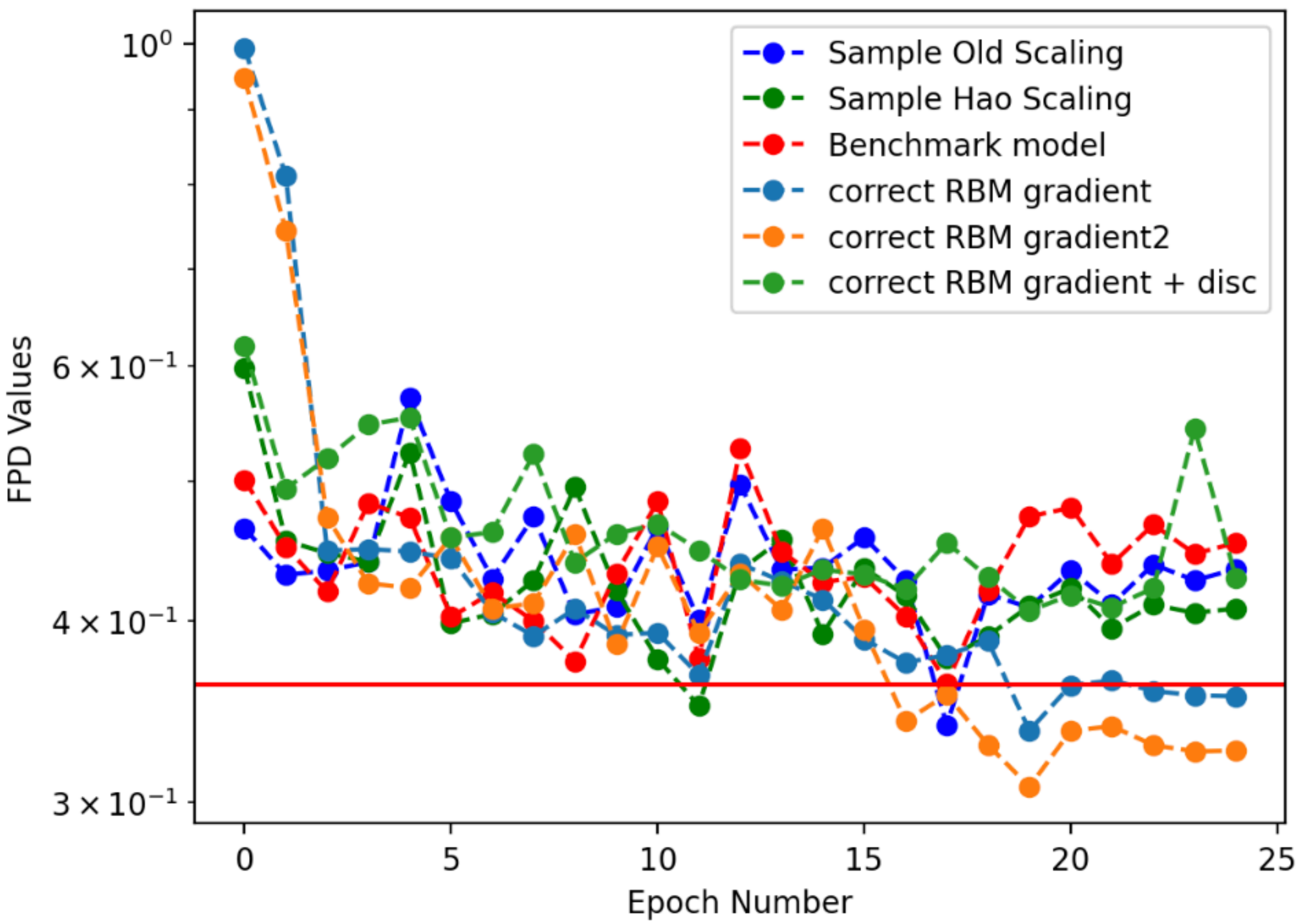
Loss functions

New Training scheme

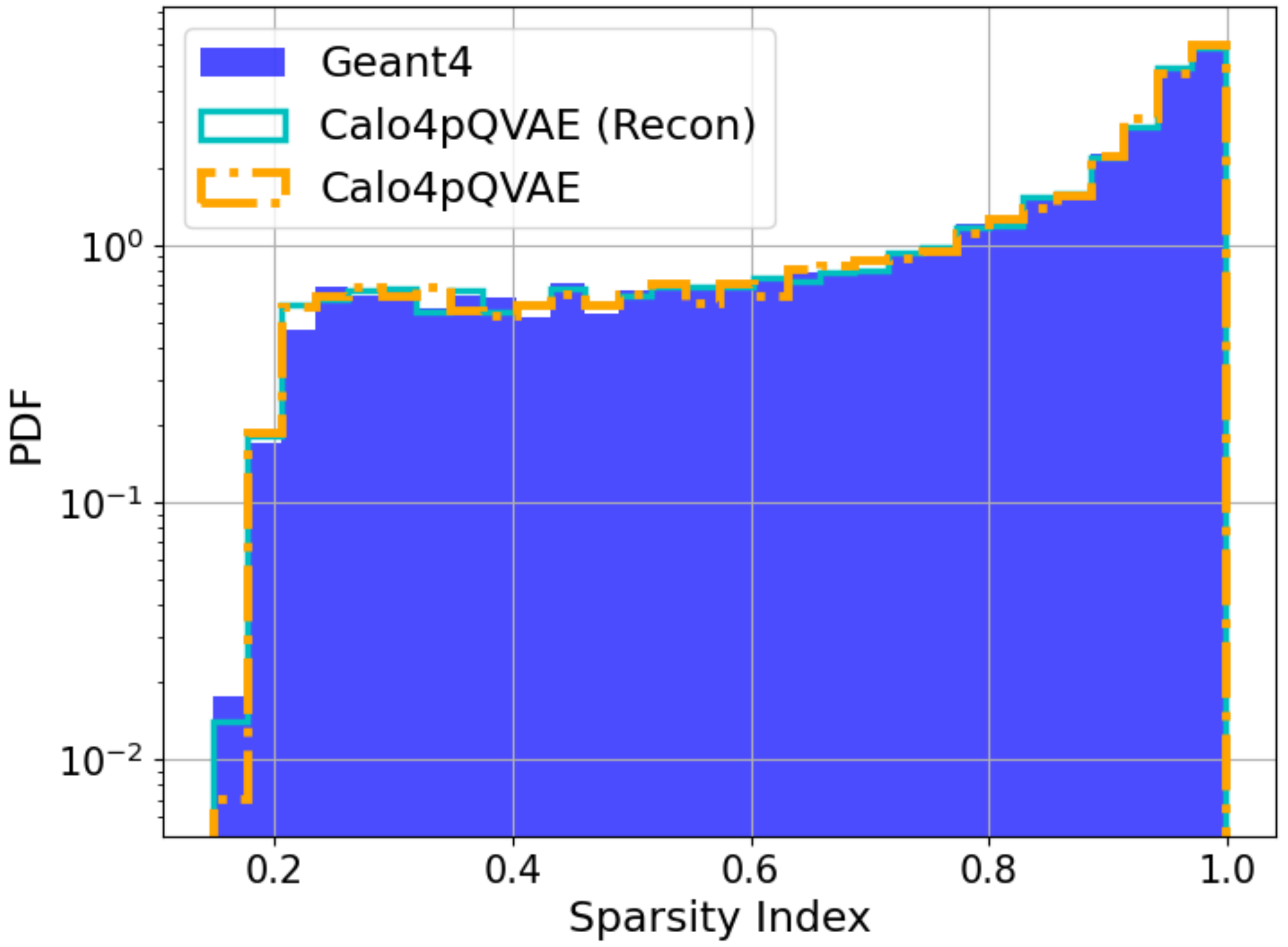
$$\langle \sum_{i=1}^{n} (\chi_i(\theta) \cdot \Theta(x_i) - x_i)^2 \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} - \langle \Theta(x_i) \ln(p_{\xi_i}(\theta)) + (1 - \Theta(x_i)) \ln(1 - p_{\xi_i}(\theta)) \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} + \langle \ln q_\phi(\boldsymbol{z}|\boldsymbol{x}) \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} + \langle E(\boldsymbol{z}) \rangle_{q_\phi(\boldsymbol{z}|\boldsymbol{x})} + \ln Z$$

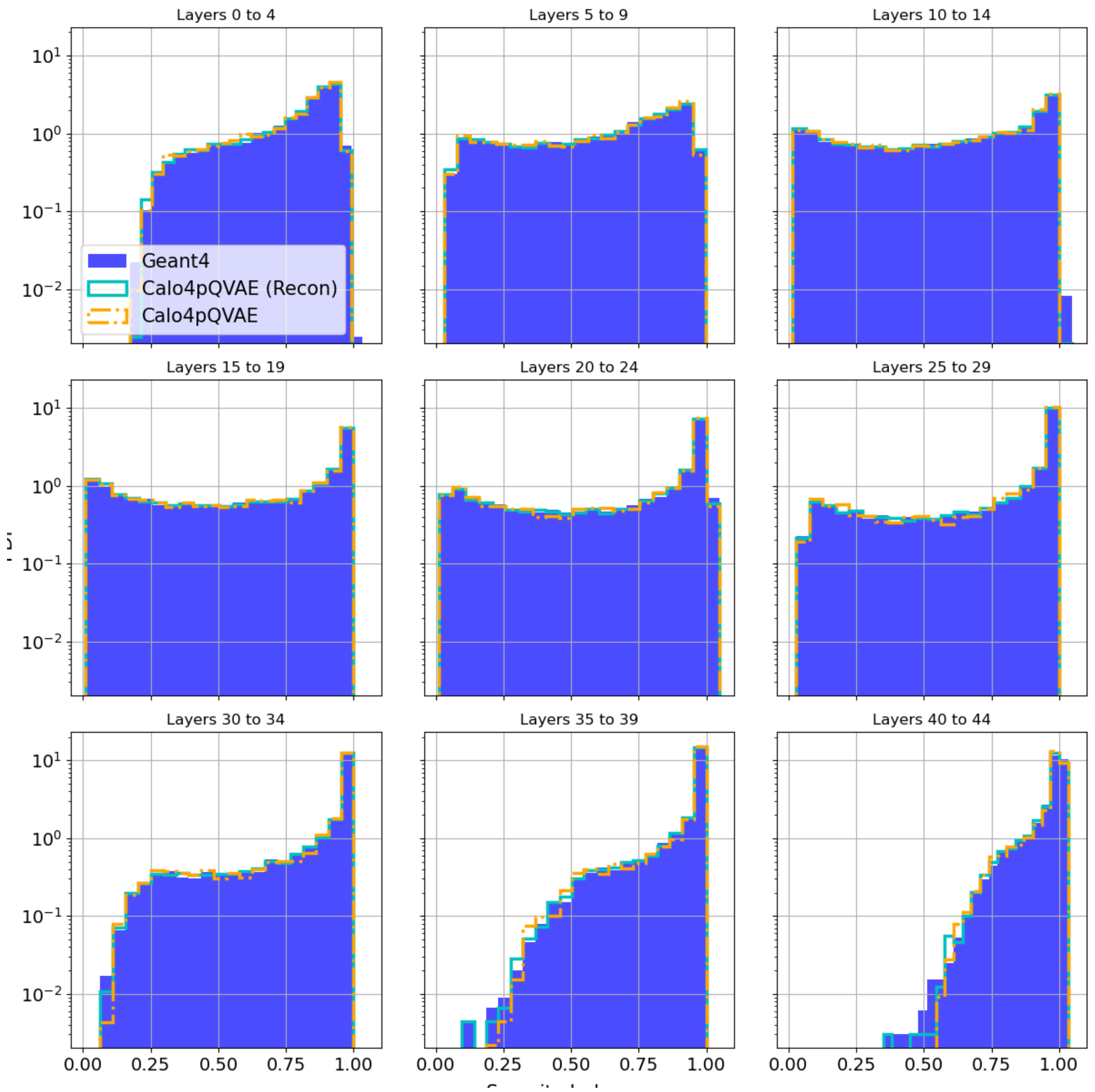
New model

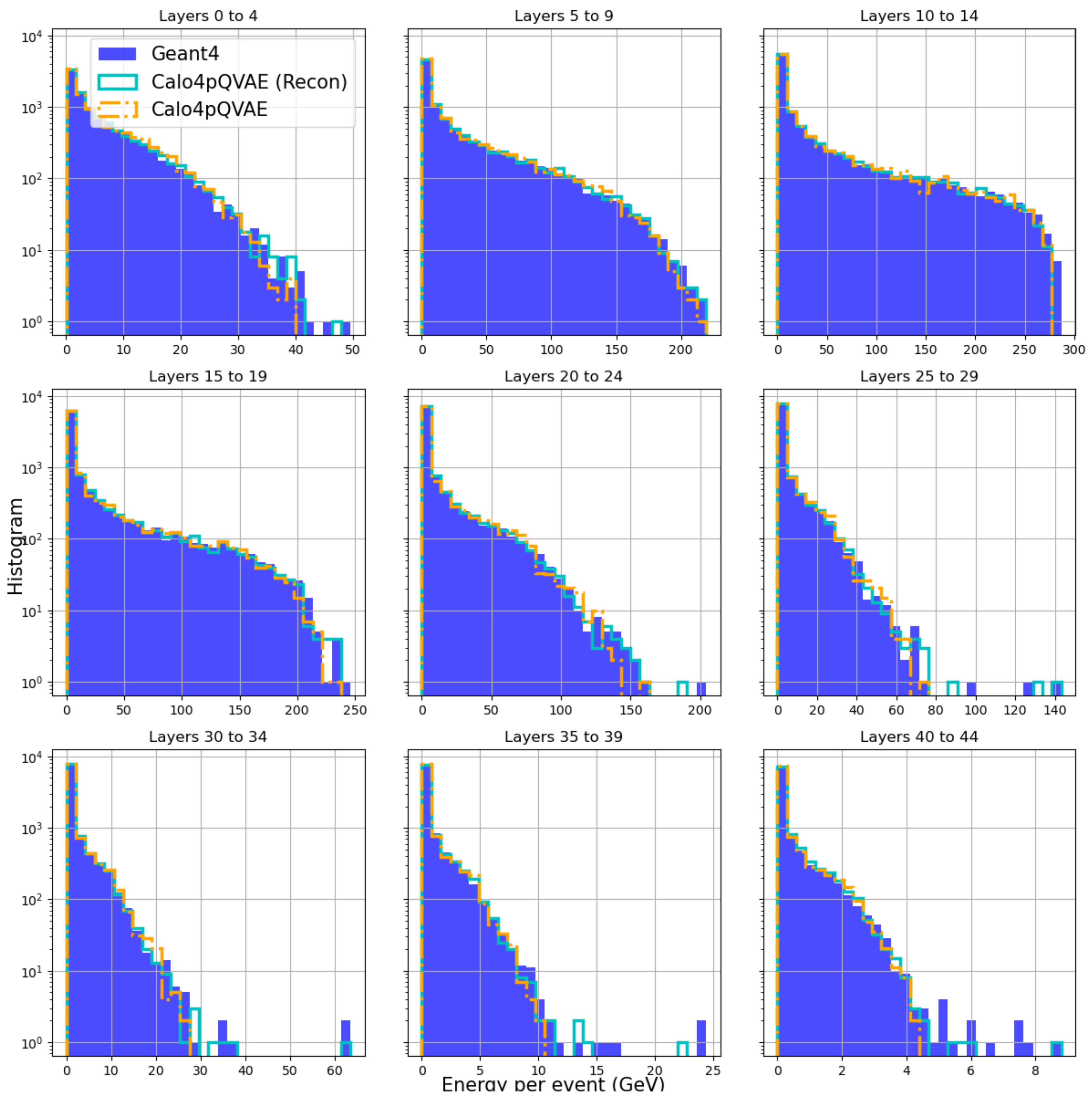

Sample Old Scaling
Sample Hao Scaling
Benchmark model
correct RBM gradient
correct RBM gradient2
correct RBM gradient + disc
KPD Values
10^{-3}
0
5
10
15
20
25
Epoch Number

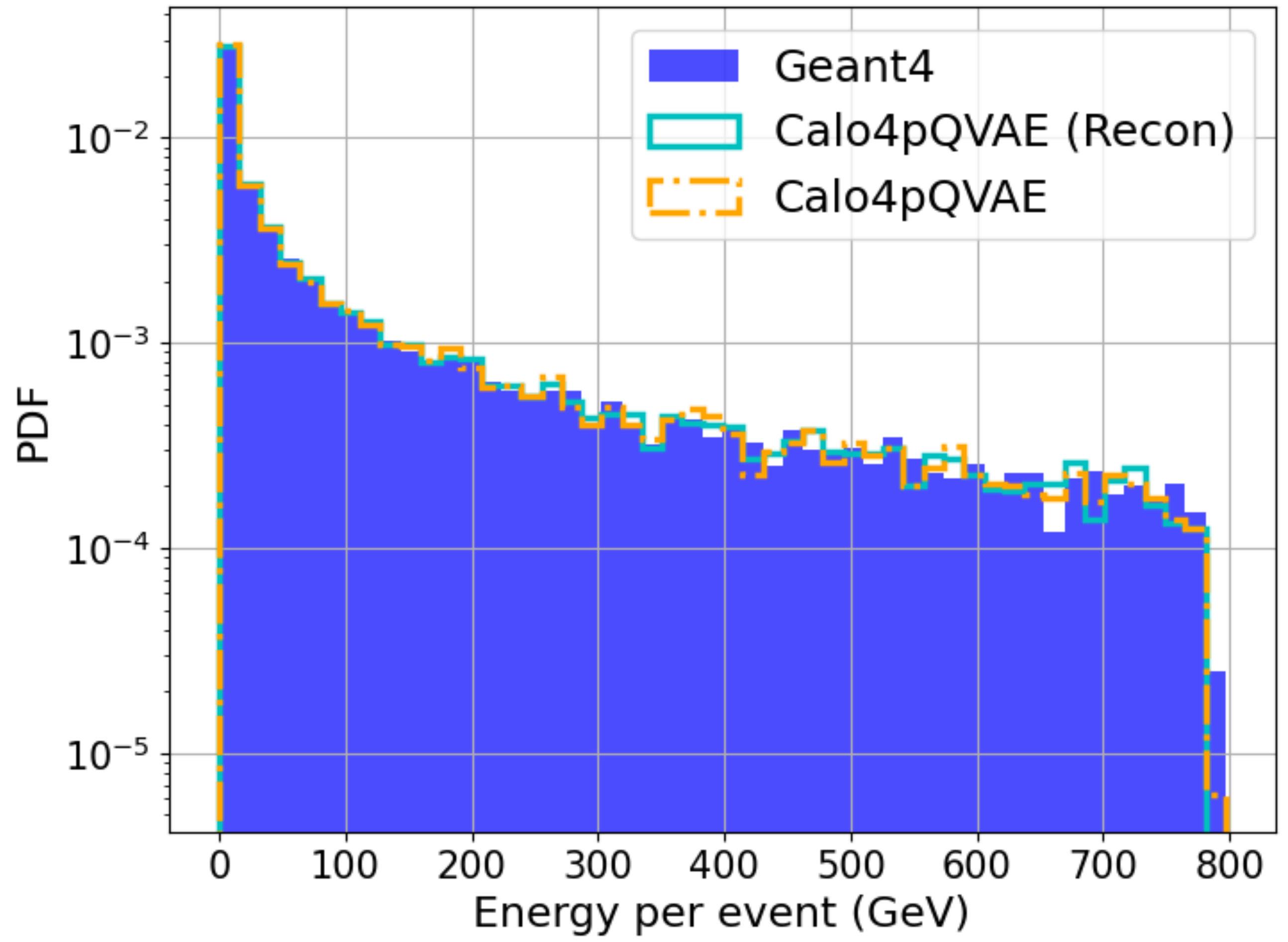

Sample Old Scaling
Sample Hao Scaling
Benchmark model
correct RBM gradient
correct RBM gradient2
correct RBM gradient + disc
FPD Values
10^{0}
6×10^{-1}
4×10^{-1}
3×10^{-1}
0
5
10
15
20
25
Epoch Number

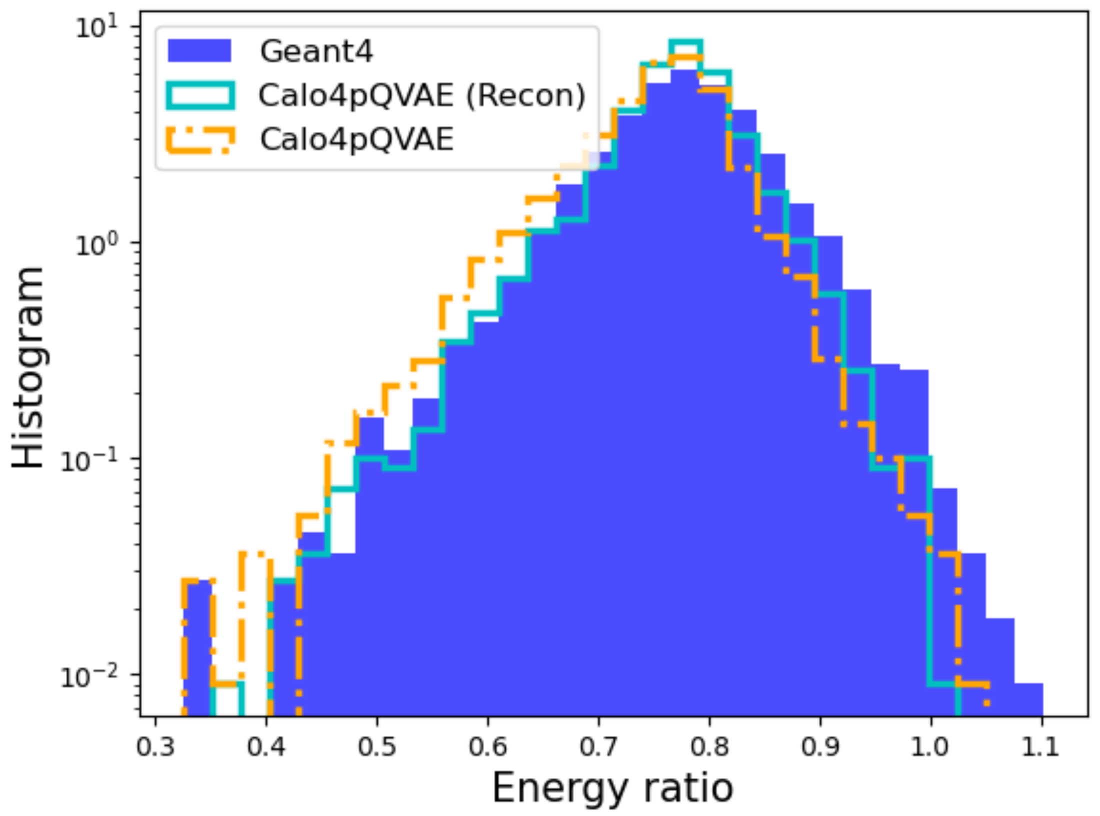
New model

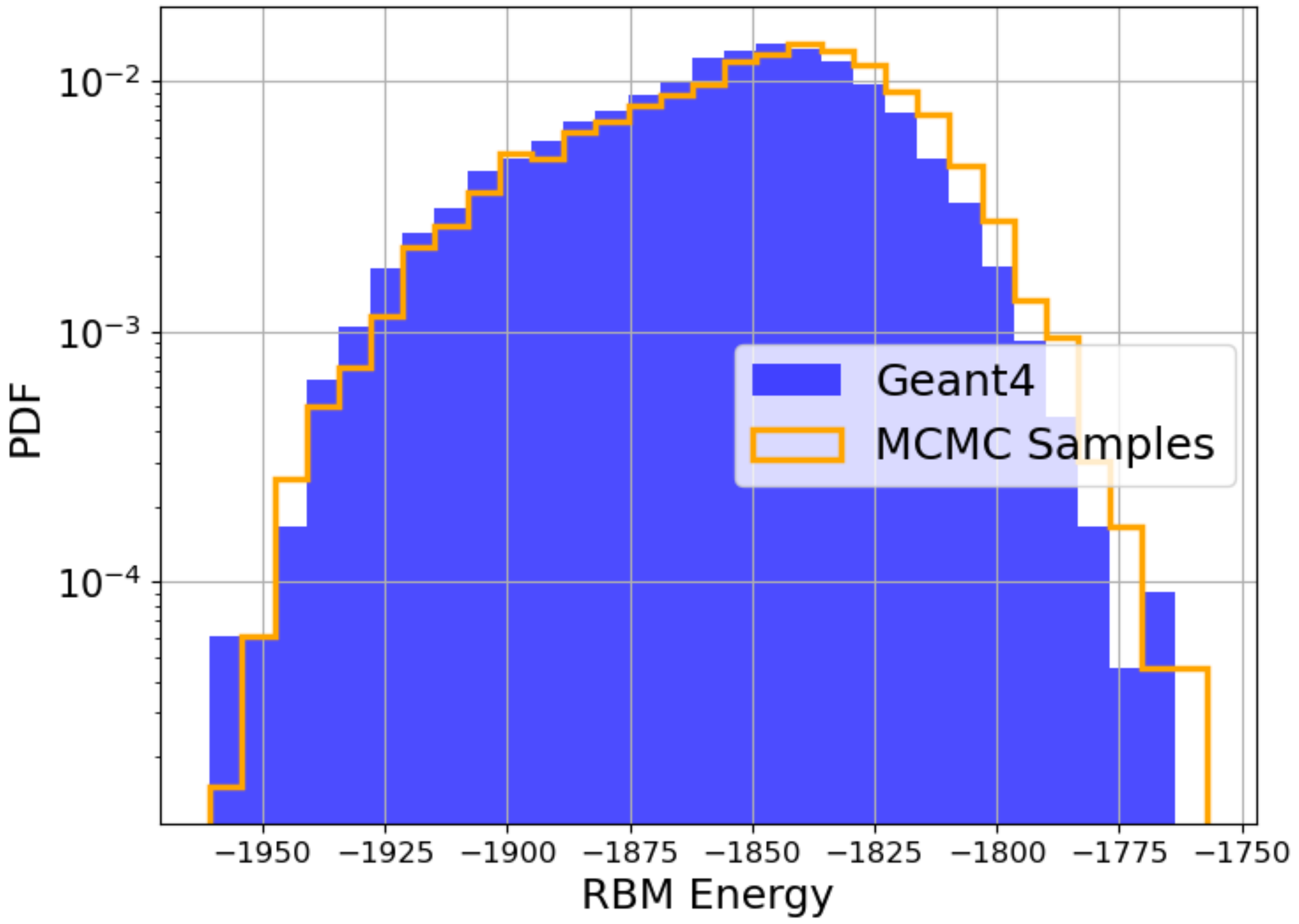

Geant4
Calo4pQVAE (Recon)
Calo4pQVAE
PDF
Sparsity Index


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
Geant4
Calo4pQVAE (Recon)
Calo4pQVAE


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
Geant4
Calo4pQVAE (Recon)
Calo4pQVAE
Histogram
Energy per event (GeV)

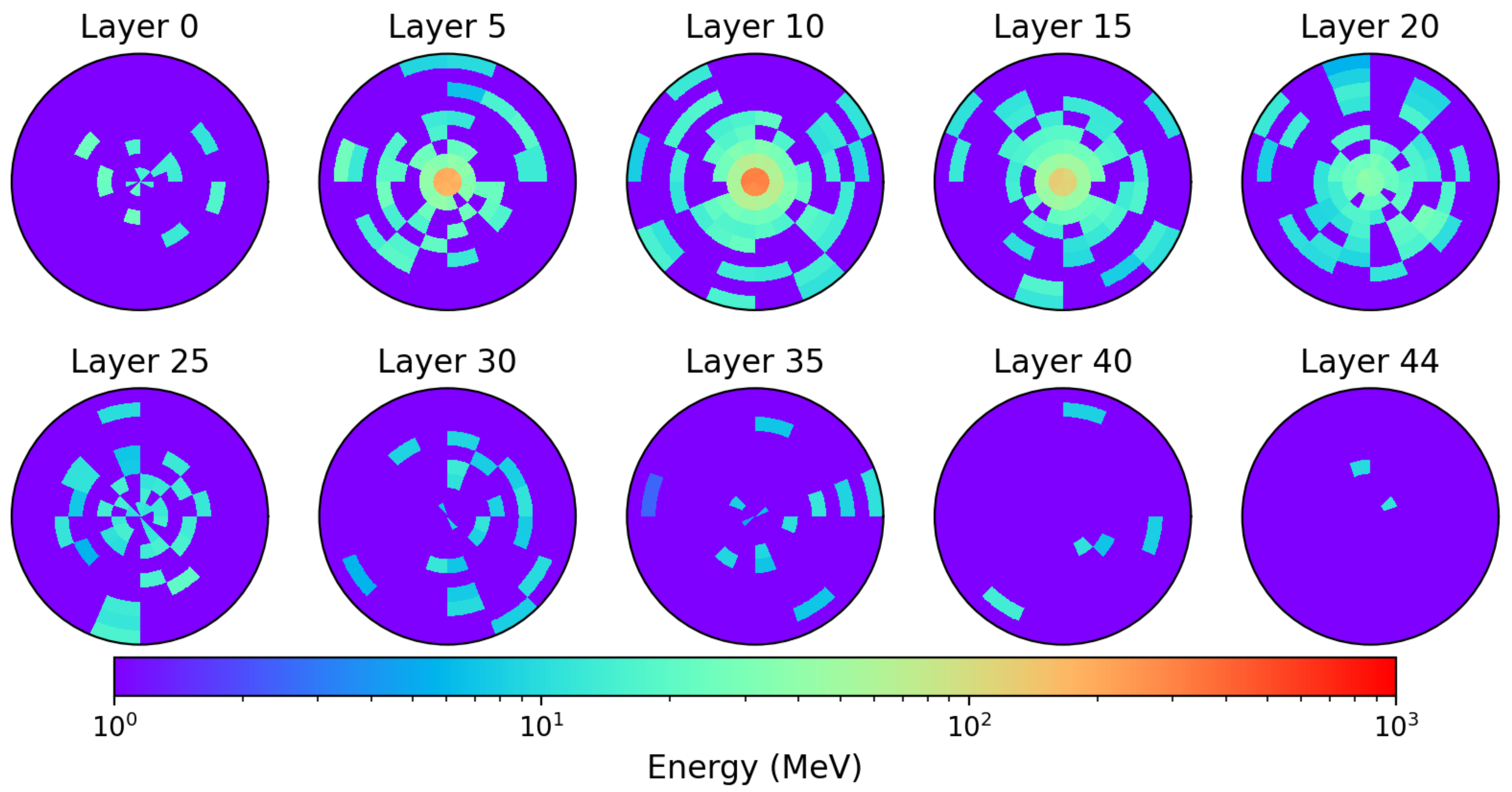

Geant4
Calo4pQVAE (Recon)
Calo4pQVAE
PDF
Energy per event (GeV)

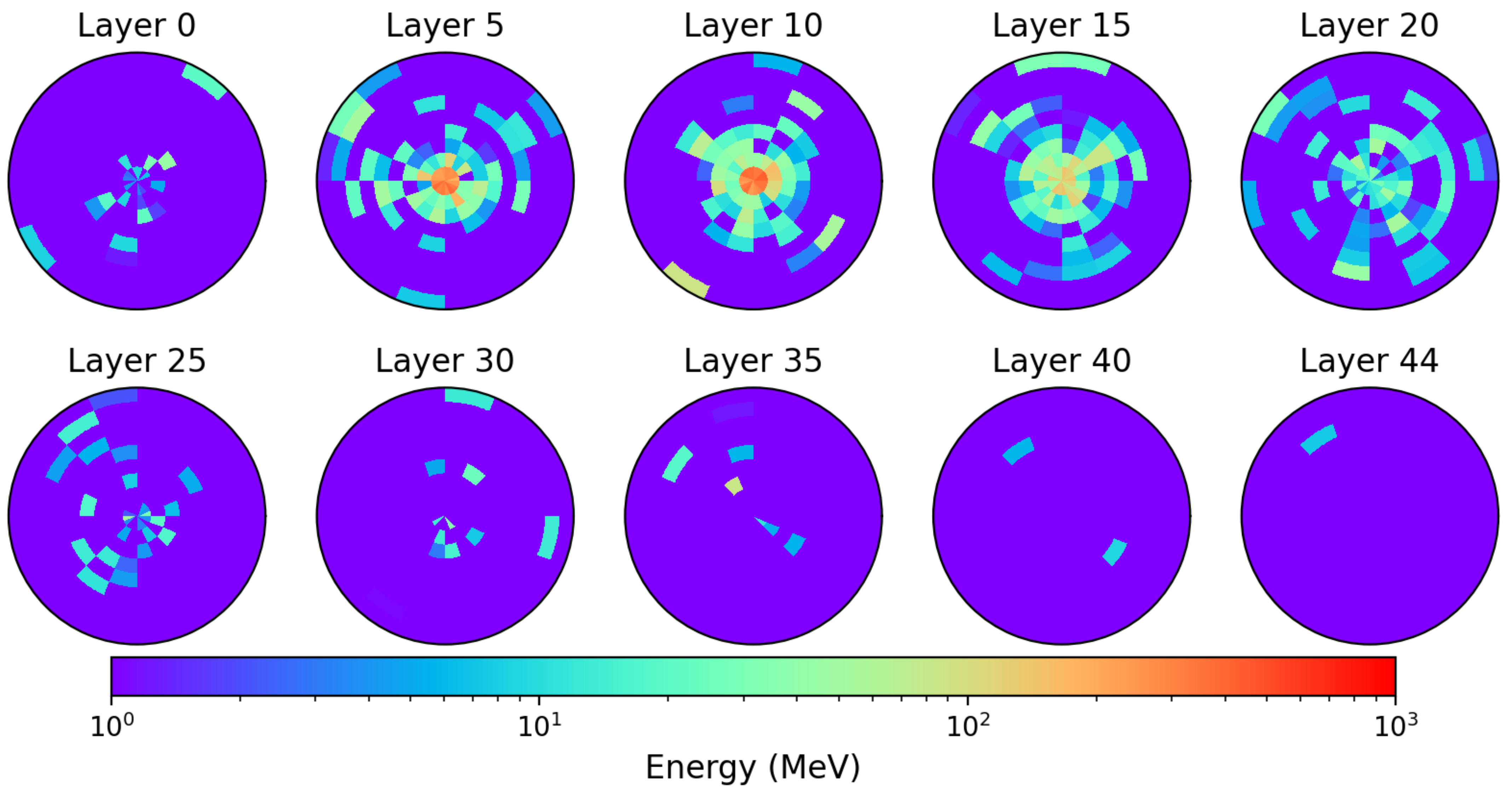

Geant4
Calo4pQVAE (Recon)
Calo4pQVAE
Histogram
Energy ratio


Geant4
MCMC Samples
PDF
RBM Energy

New model

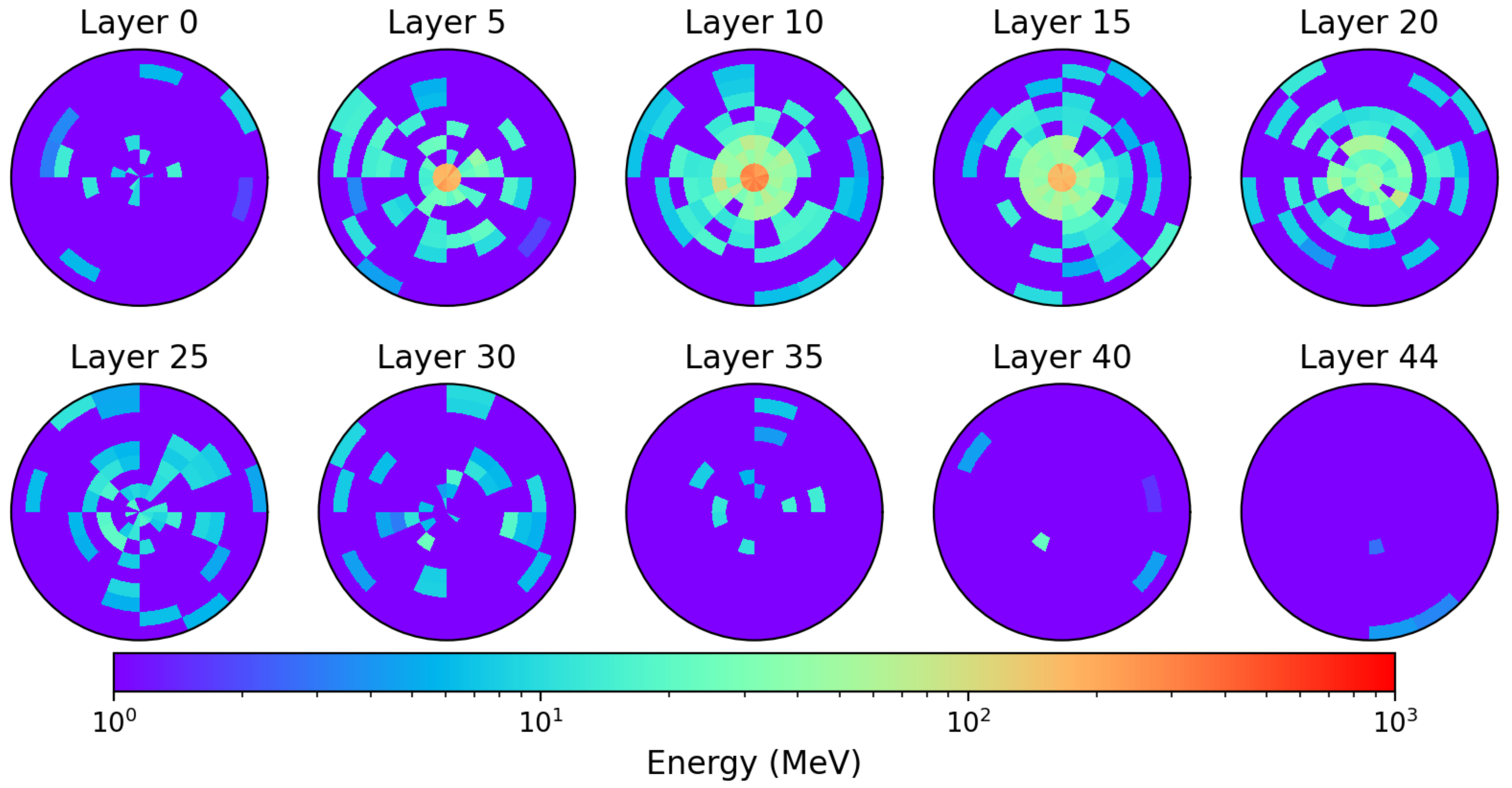
But the granularity is still poor

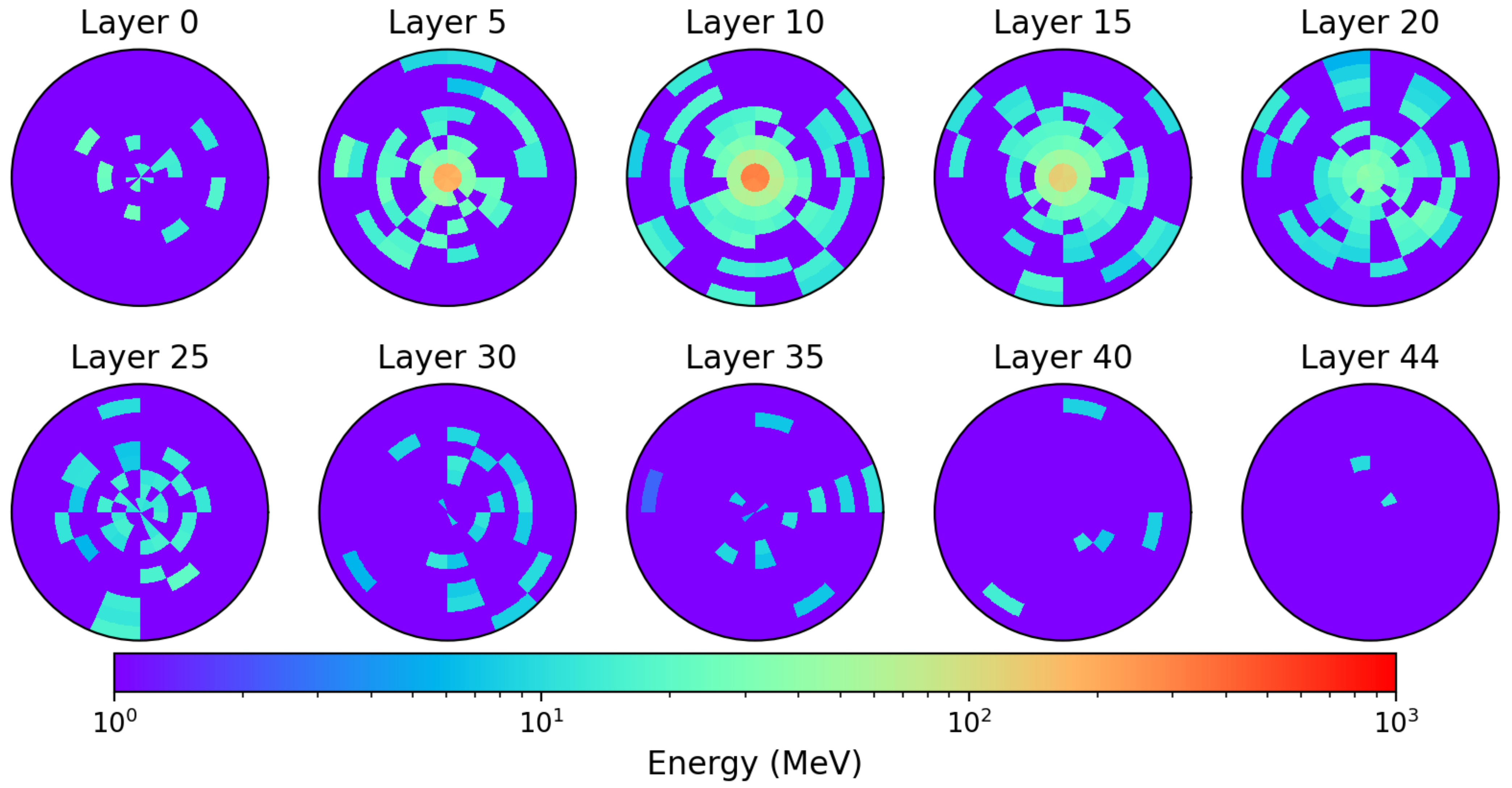


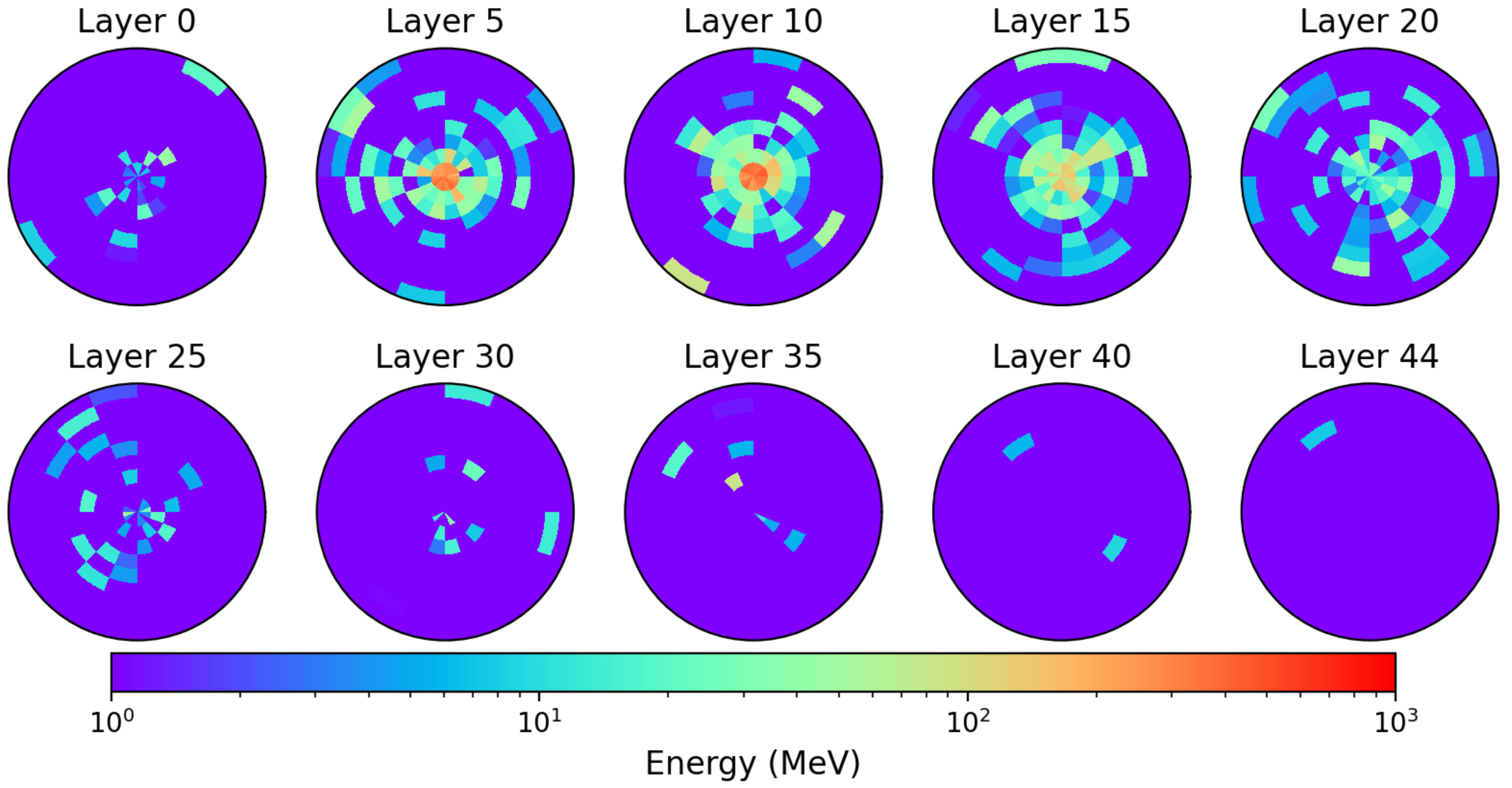


New model

But the granularity is still poor
But the discriminator might be the answer…

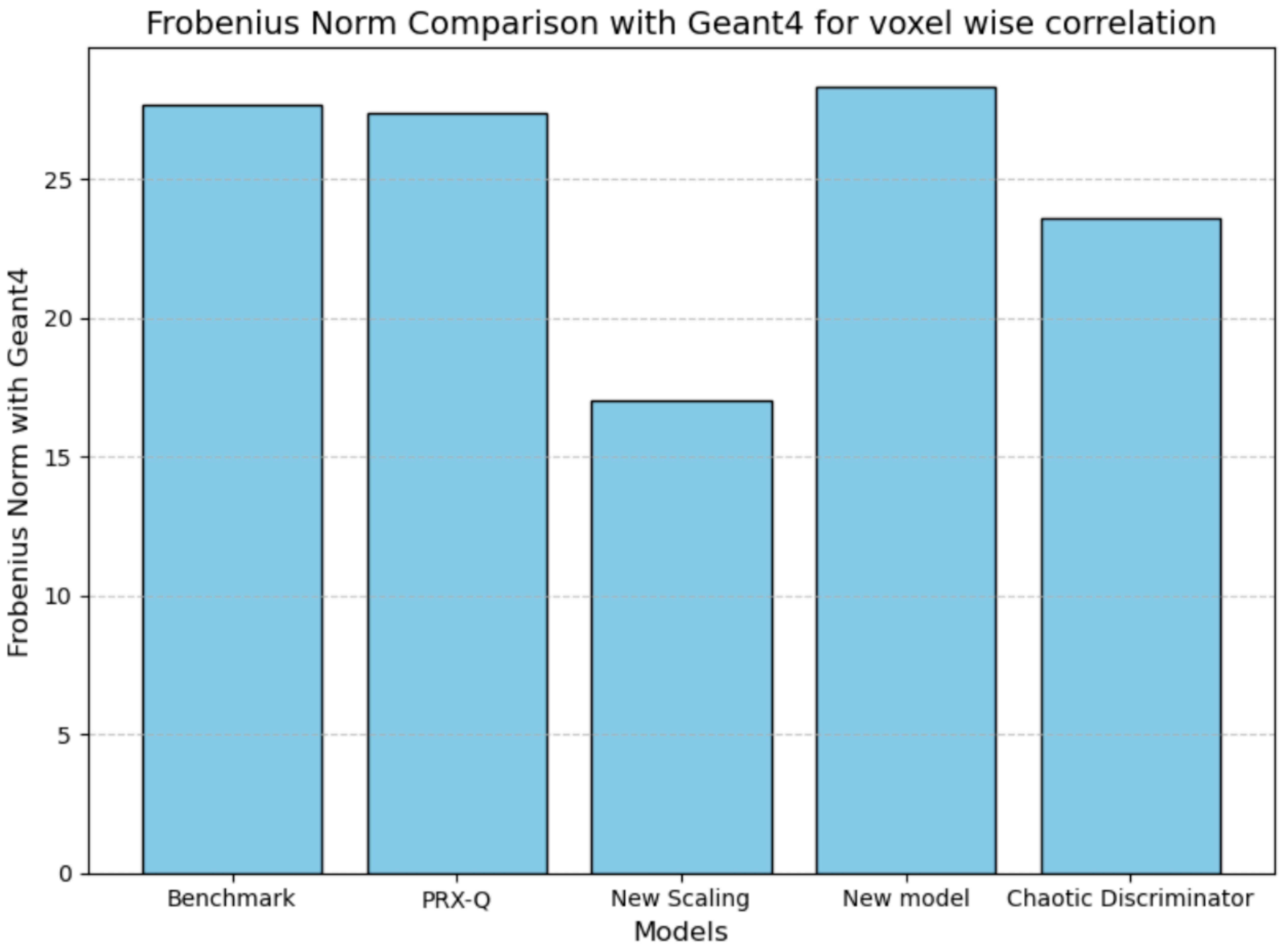


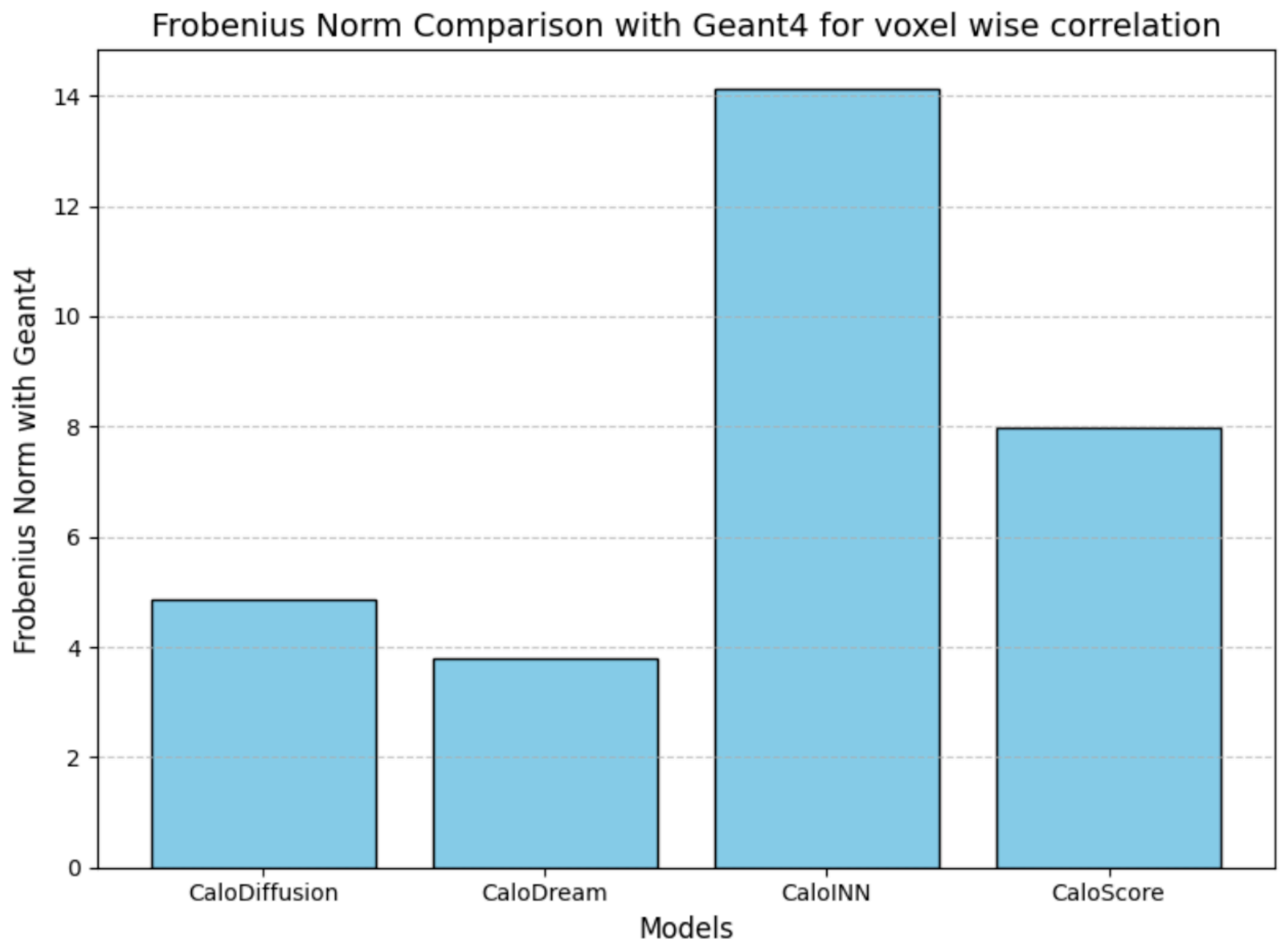




Frobenius metric

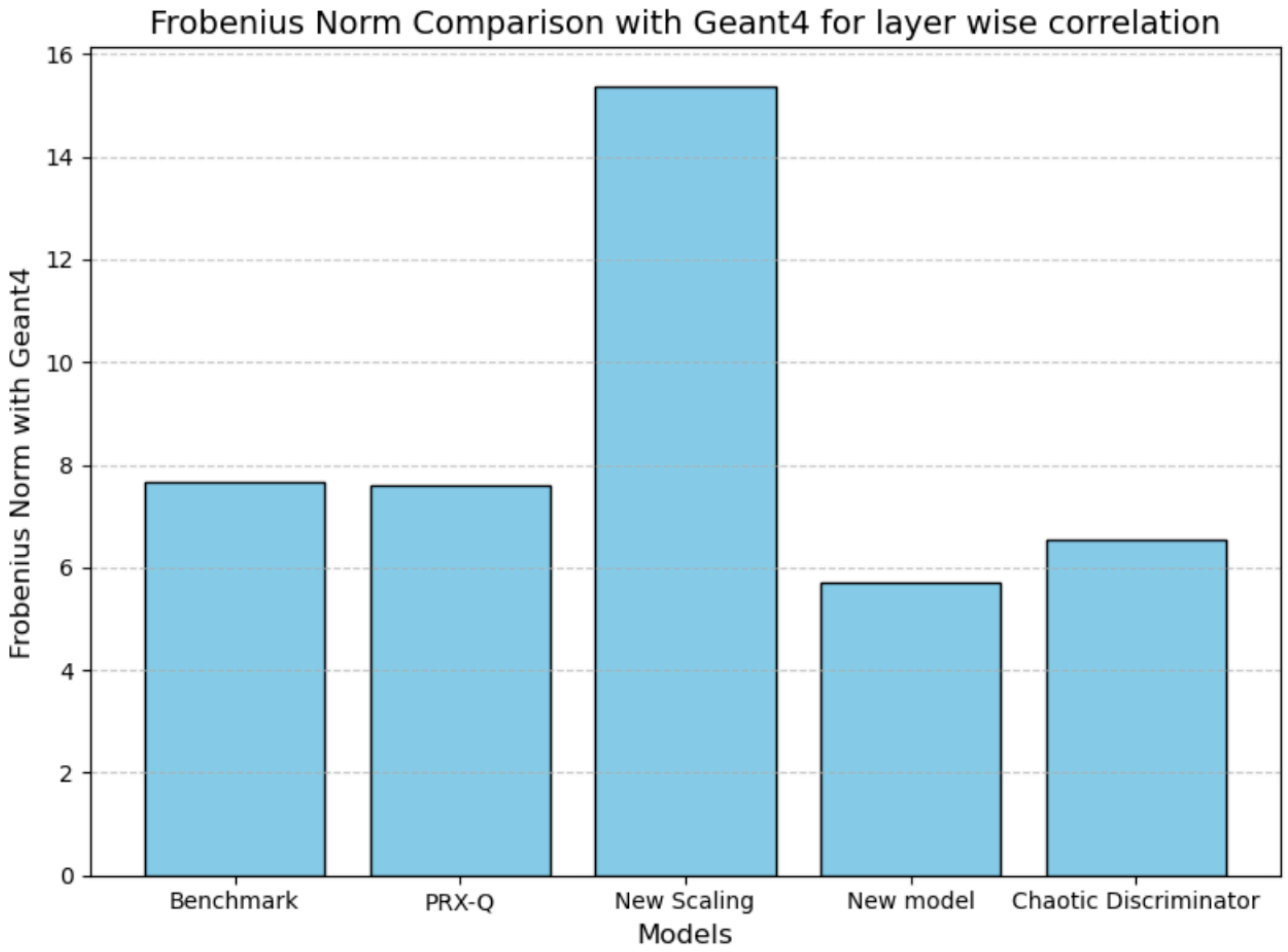
Code written and shared by Farzana

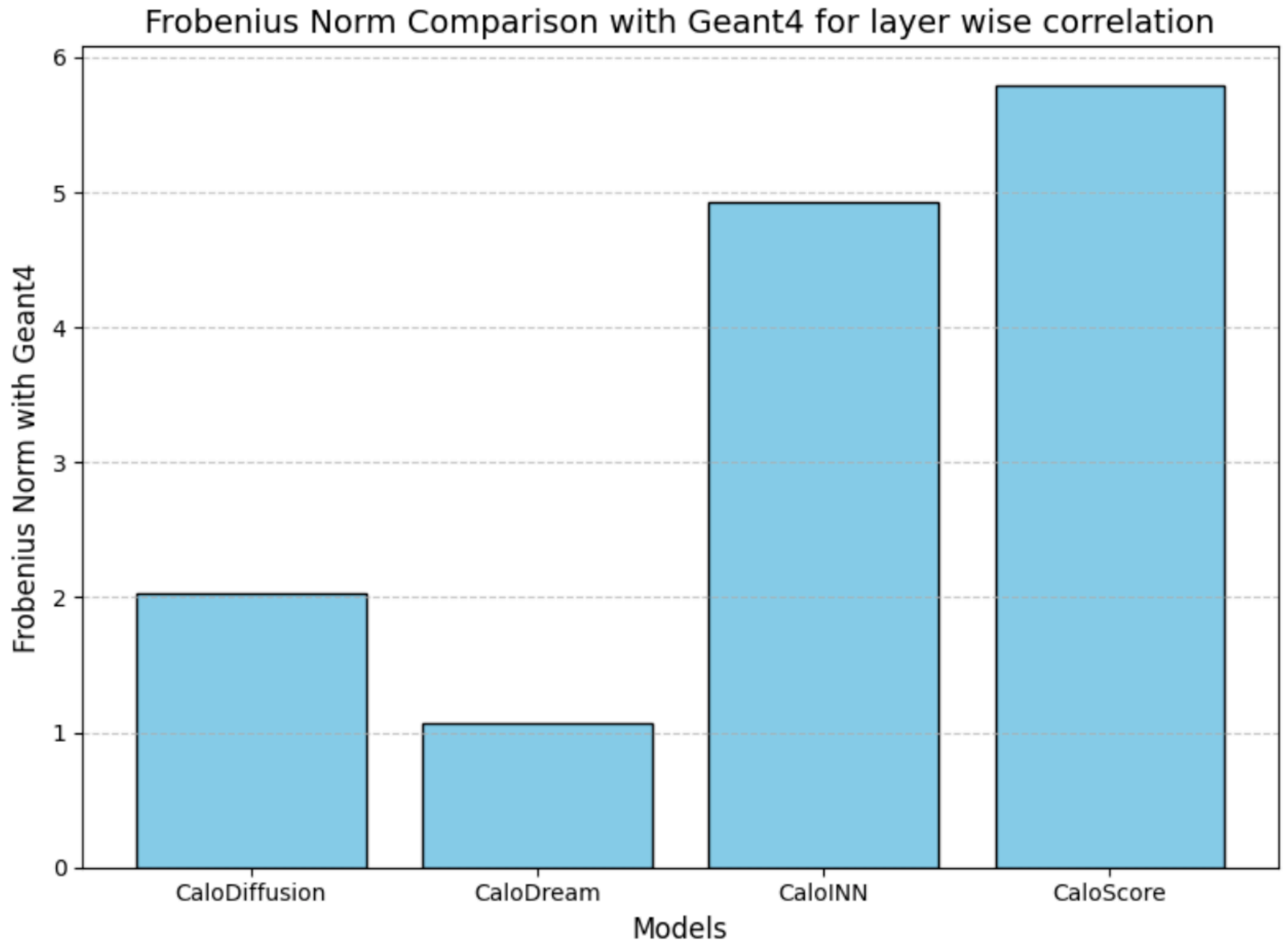




Frobenius metric

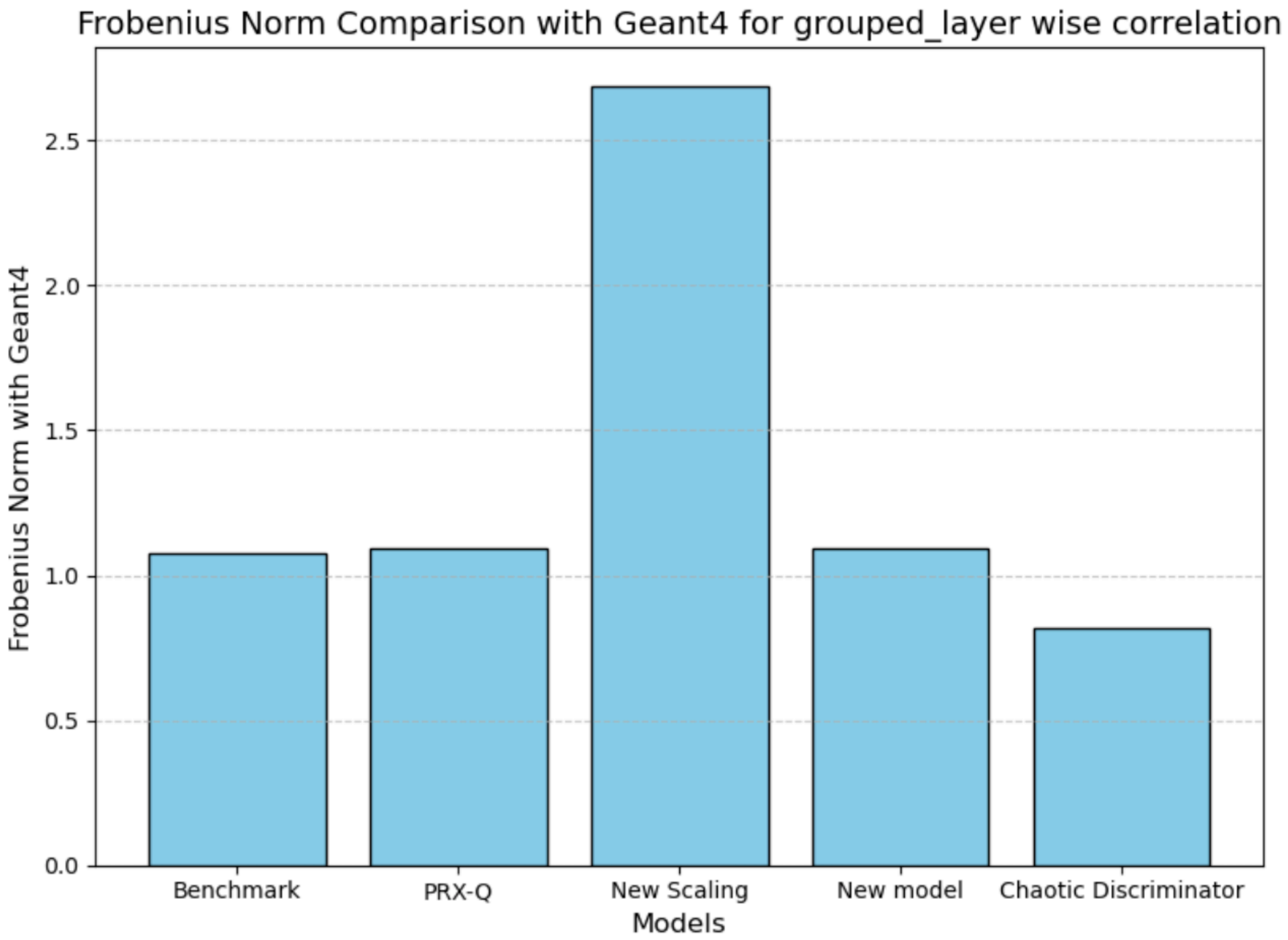
Code written and shared by Farzana

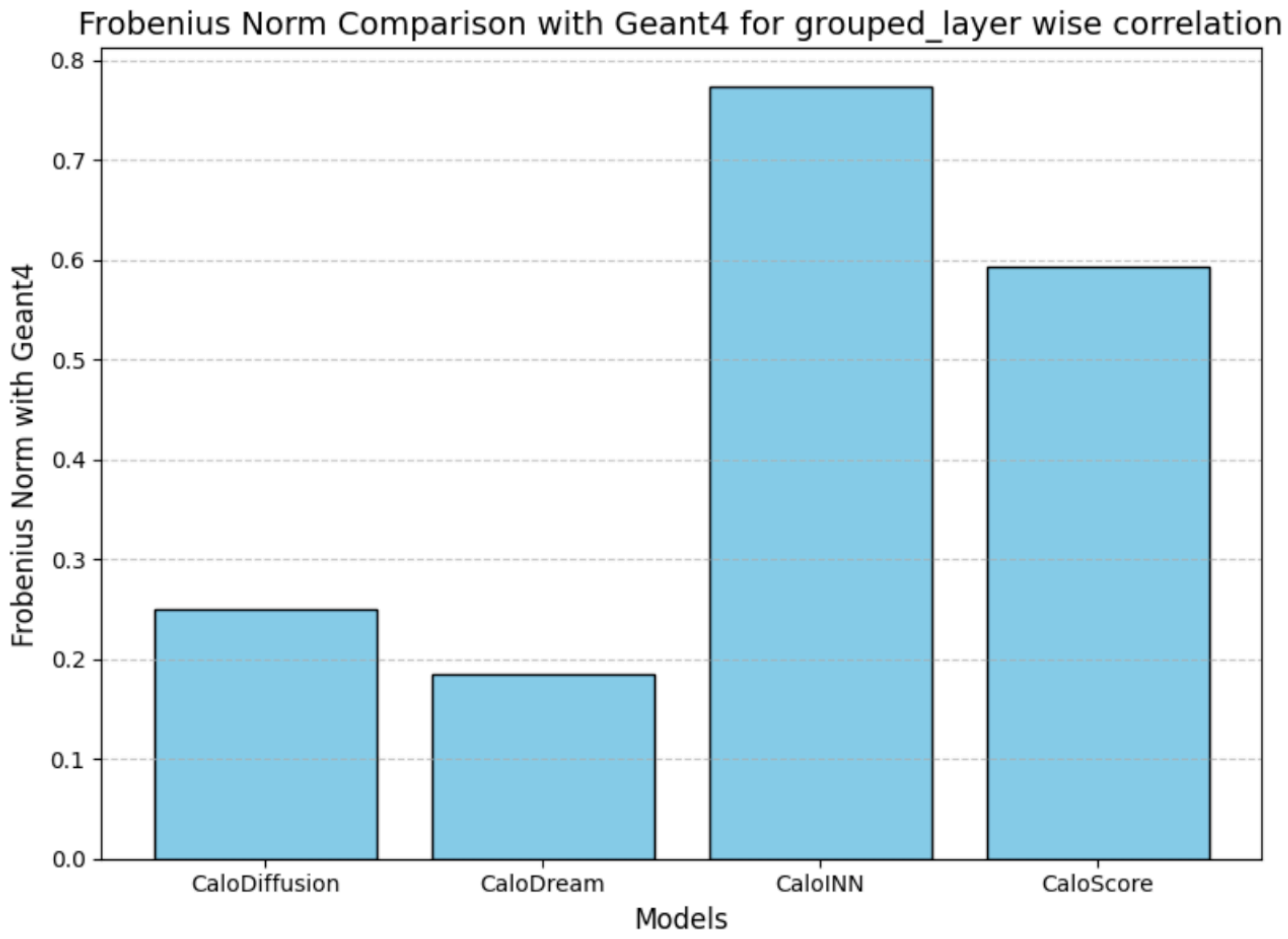




Frobenius metric

Code written and shared by Farzana





Correlation function

Correlation function

Definitions

- Auto-correlation function (ACF). <> is average over ensemble

$$ACF(t,\tau) = \langle (f(t+\tau) - \langle f(t+\tau)\rangle)(f(t) - \langle f(t)\rangle)\rangle$$

- Normalized ACF

$$ACF_N(t,\tau) = \frac{ACF(t,\tau)}{\langle f(t)^2\rangle - \langle f(t)\rangle^2}$$

- Time-average Normalized ACF

$$C(\tau) \equiv ACF_N(\tau) = \frac{1}{t_{max} - \tau}\sum_{t=0}^{t_{max}-\tau-1} ACF_N(t,\tau)$$

Correlation function

Magnetization

- We'll look at the ACF of the magnetization over time.

$$f(t) \to m(t) = \frac{1}{N}\sum_{i=1}^{N} \sigma_i(t)$$

- $\sigma_i(t)$ is the i-th spin value at time t

Correlation function

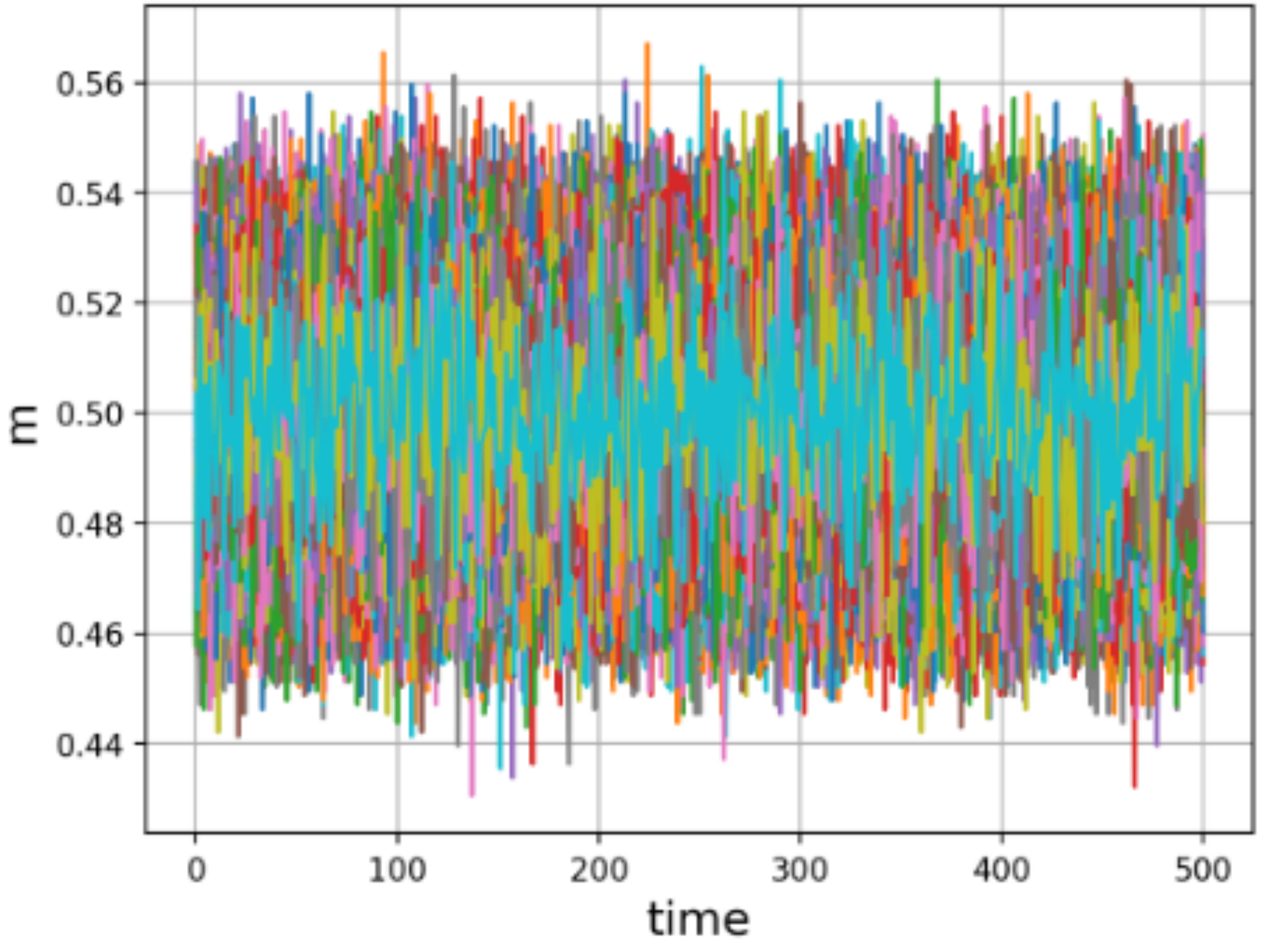
Simple example: Random Bernoulli-distributed noise

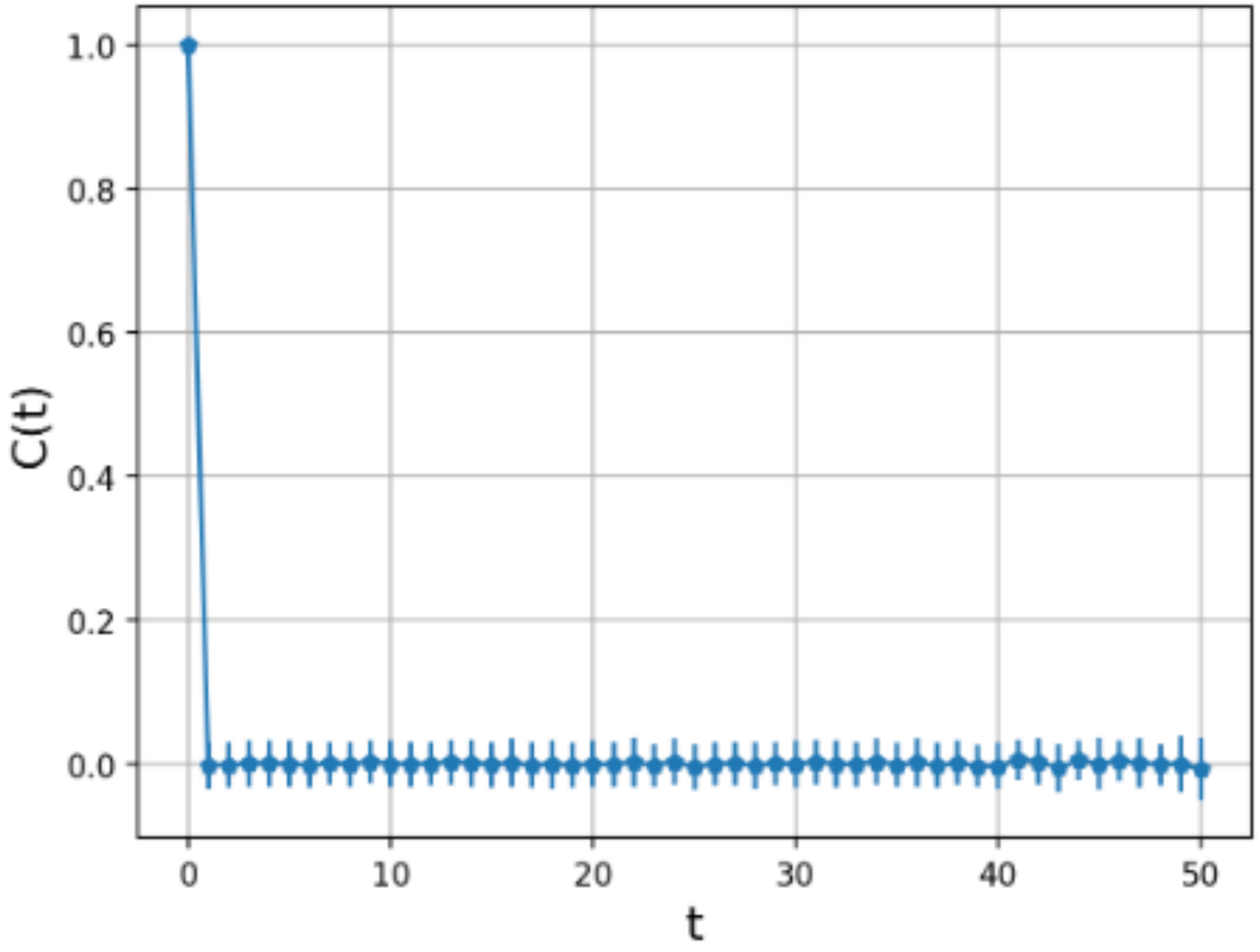
302x4 spins

Ensemble of size 1k

500 time-steps

(MC sweep = time-step)





Correlation function

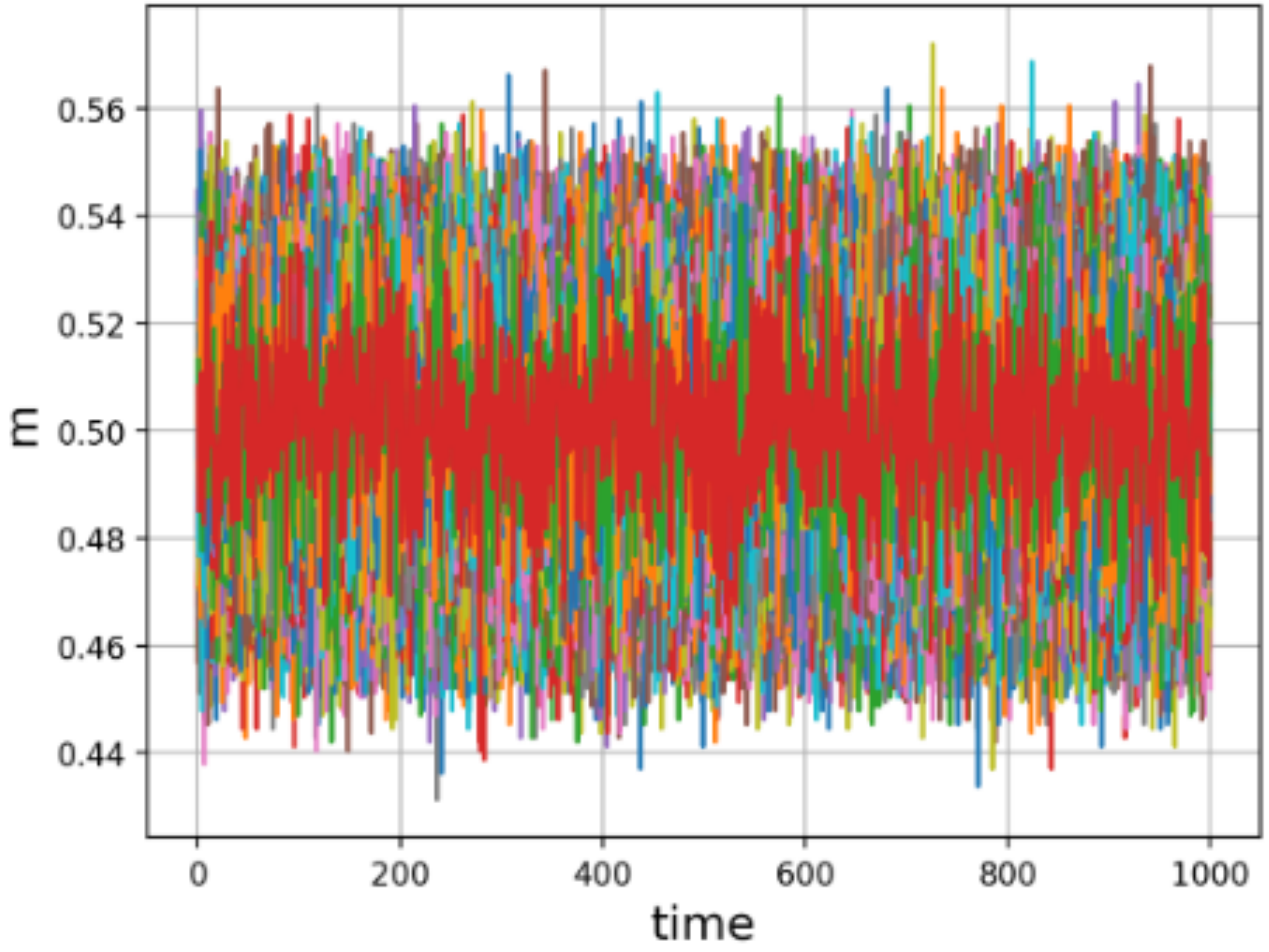
Random Initialized RBM
(fully connected)
Weight and biases sampled from N(0,0.01)

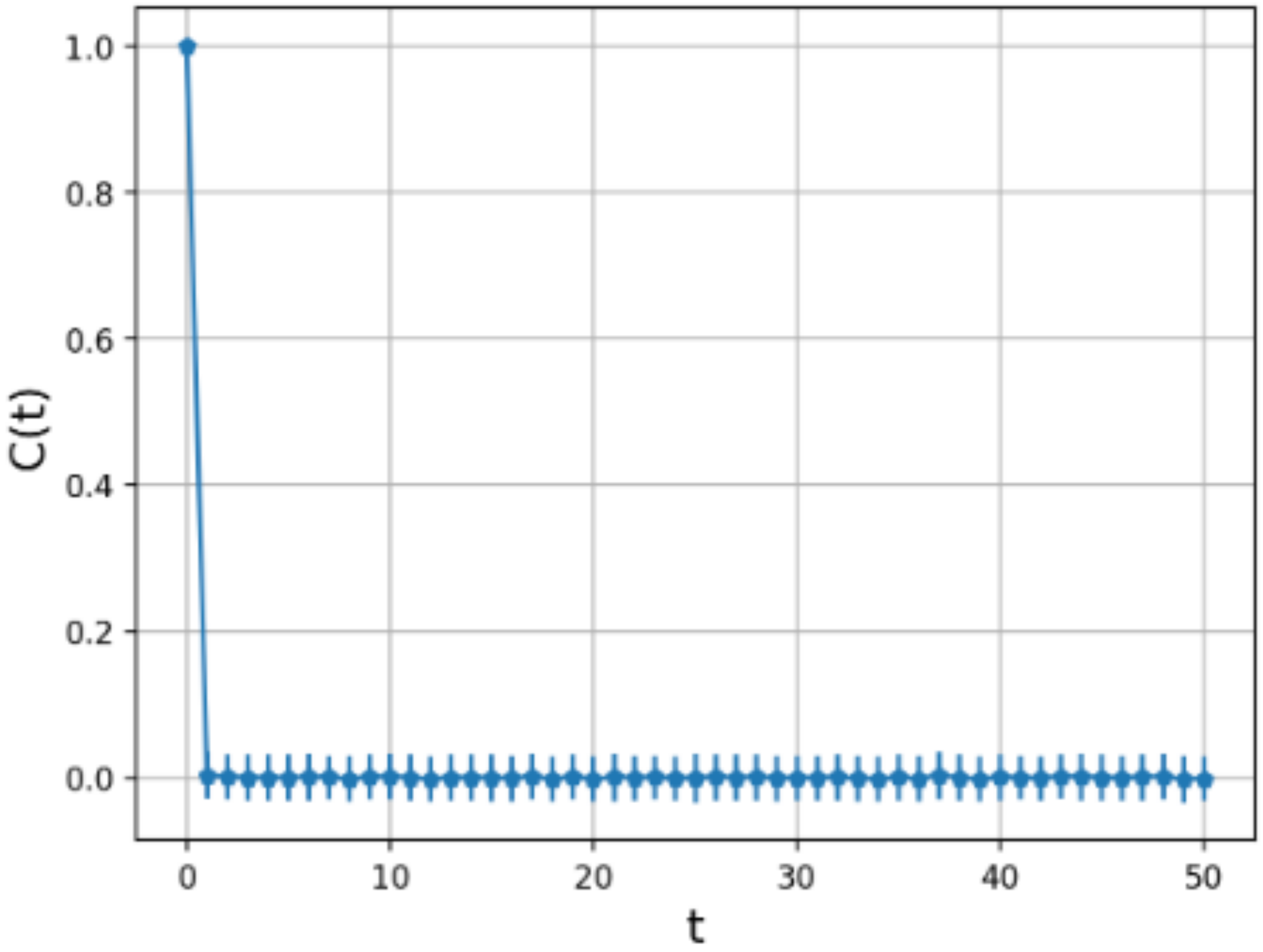
302x4 spins

Ensemble of size 100

1000 time-steps

(MC sweep = time-step)





Correlation function

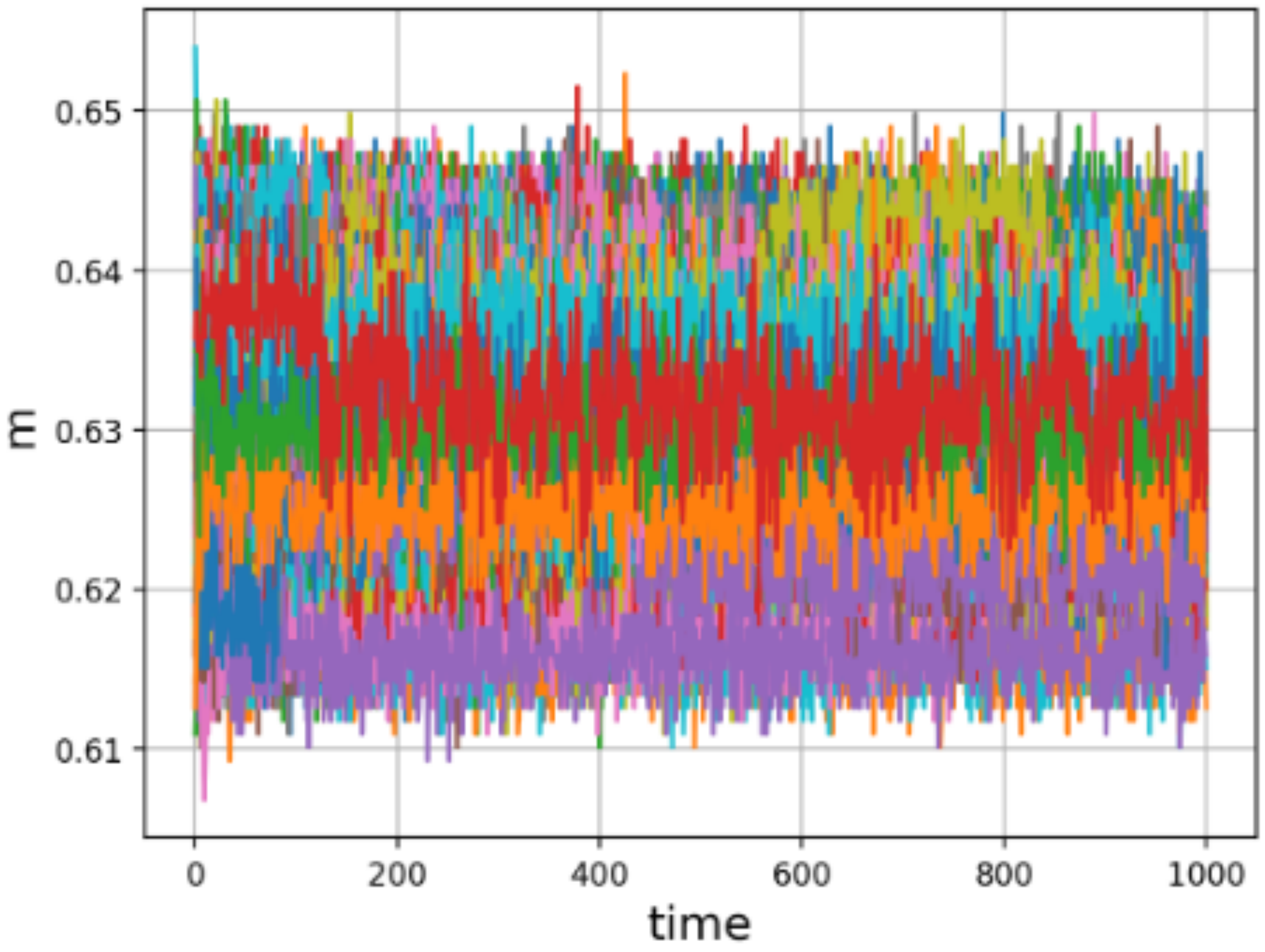
Random Initialized RBM
(fully connected)

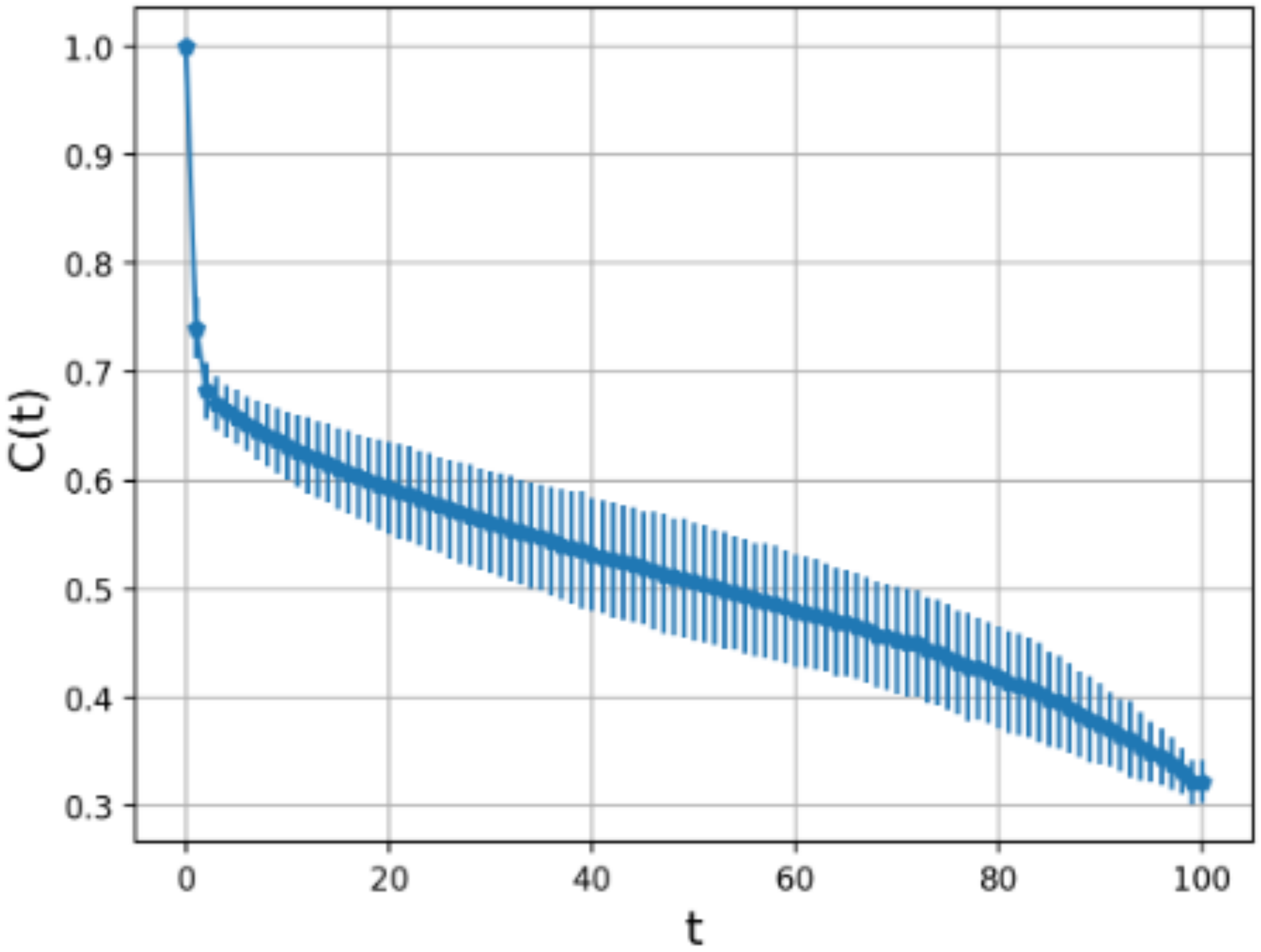
302x4 spins

Ensemble of size 100

1000 time-steps ($\Delta t = 1000$)

(MC sweep = time-step)





Correlation function

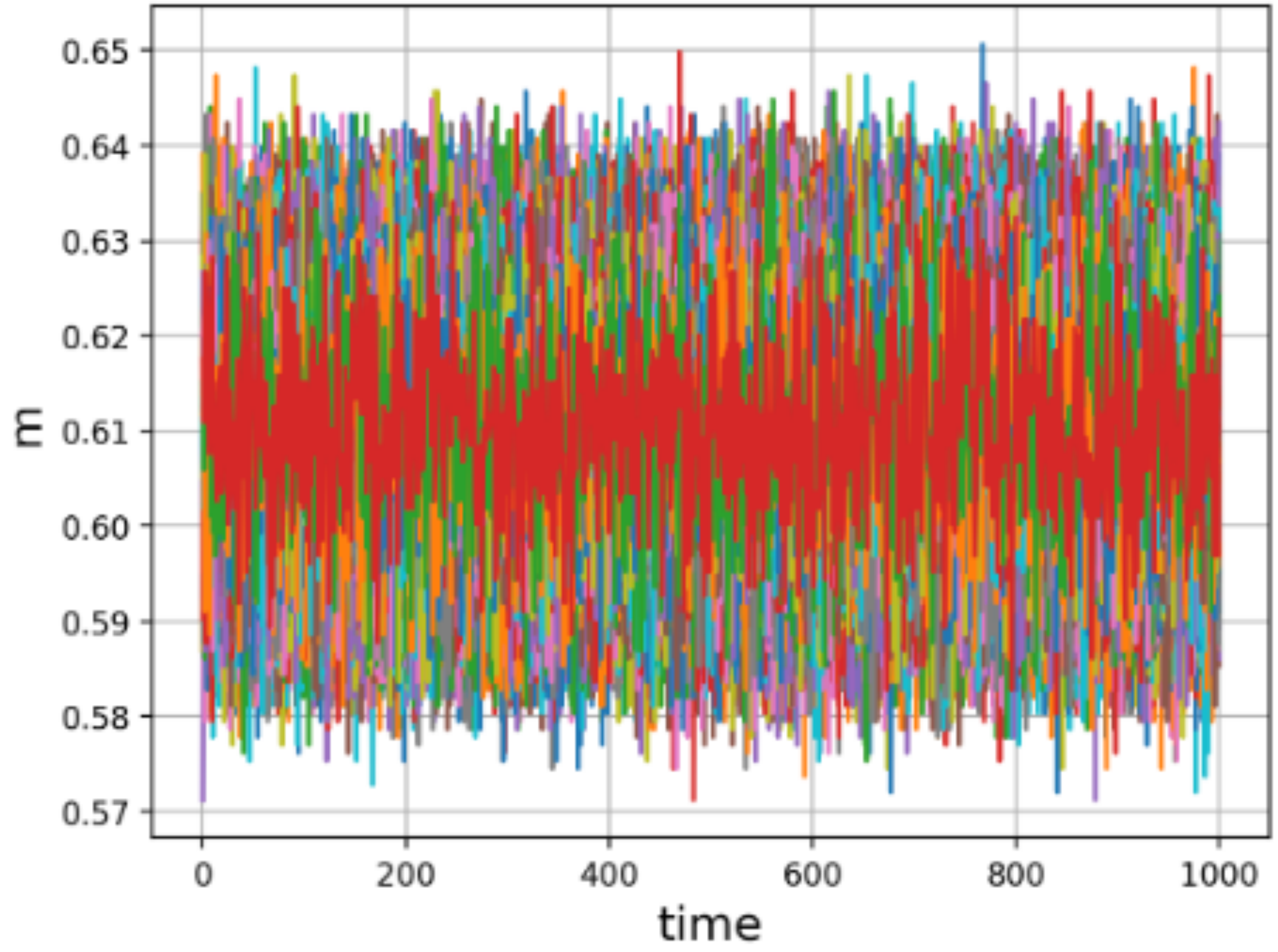
Initialized RBM (Zephyr top)

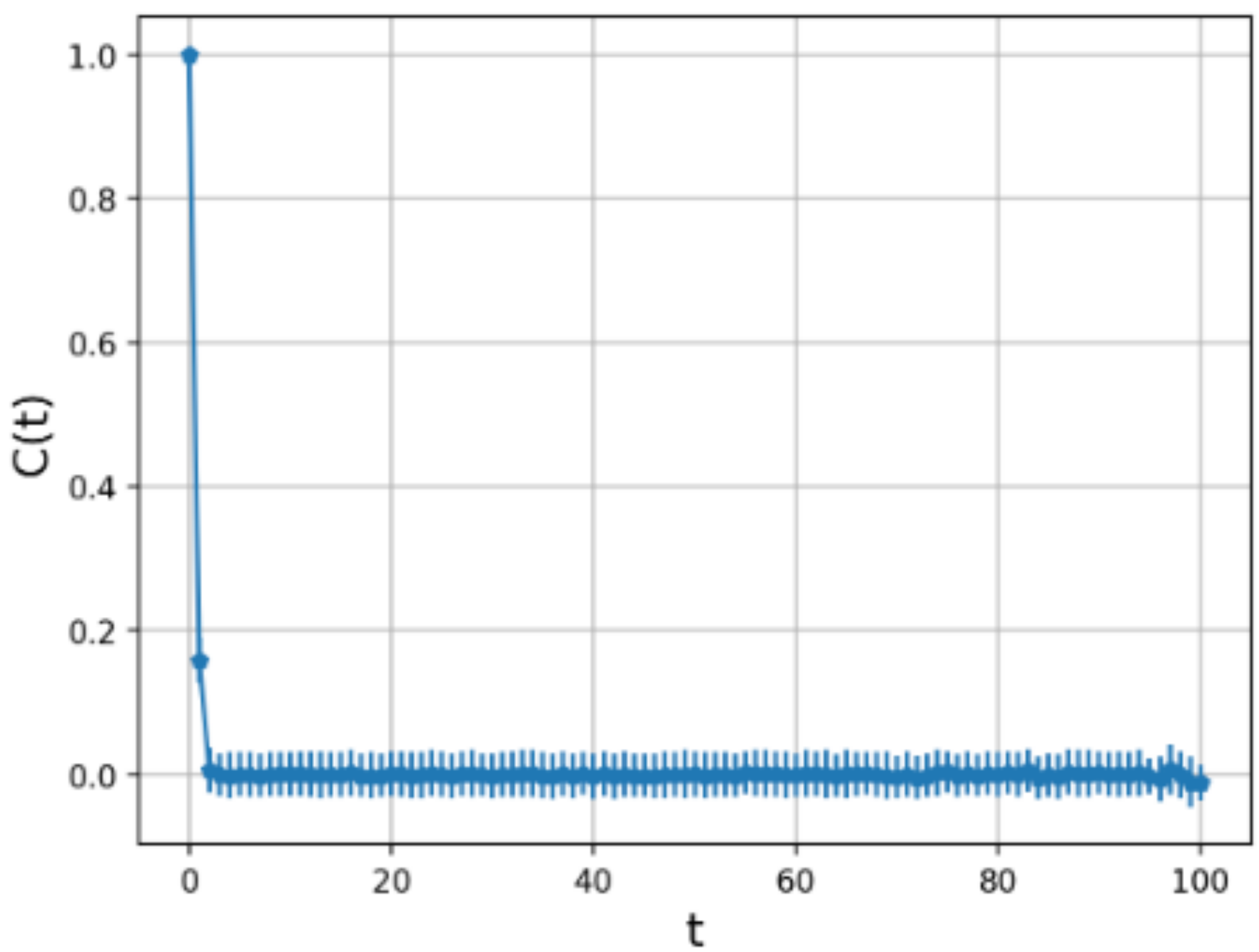
302x4 spins

Ensemble of size 100

1000 time-steps ($\Delta t = 1$)

(MC sweep = time-step)





Correlation function

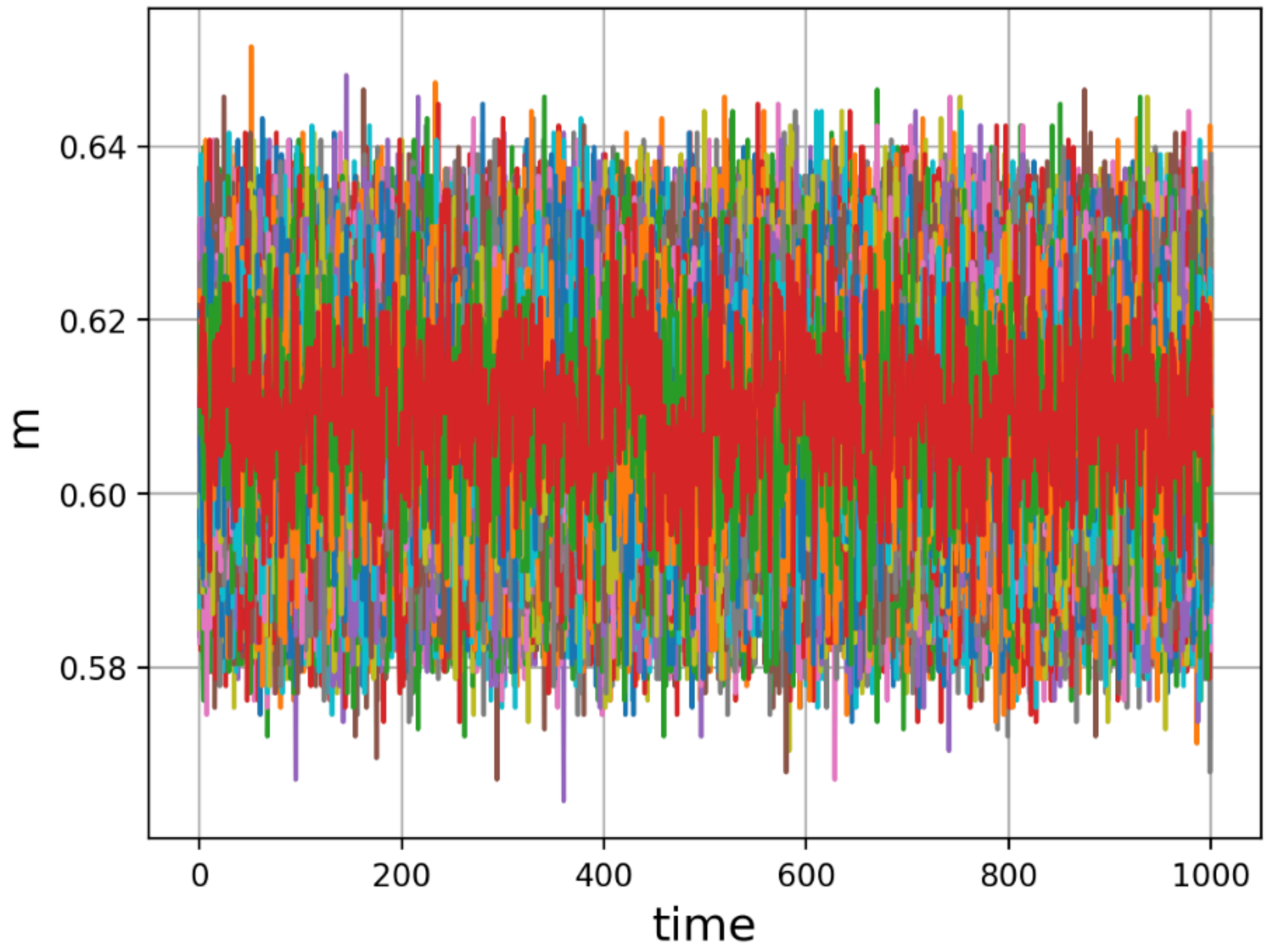
Initialized RBM (Zephyr top)
W/ biases=0

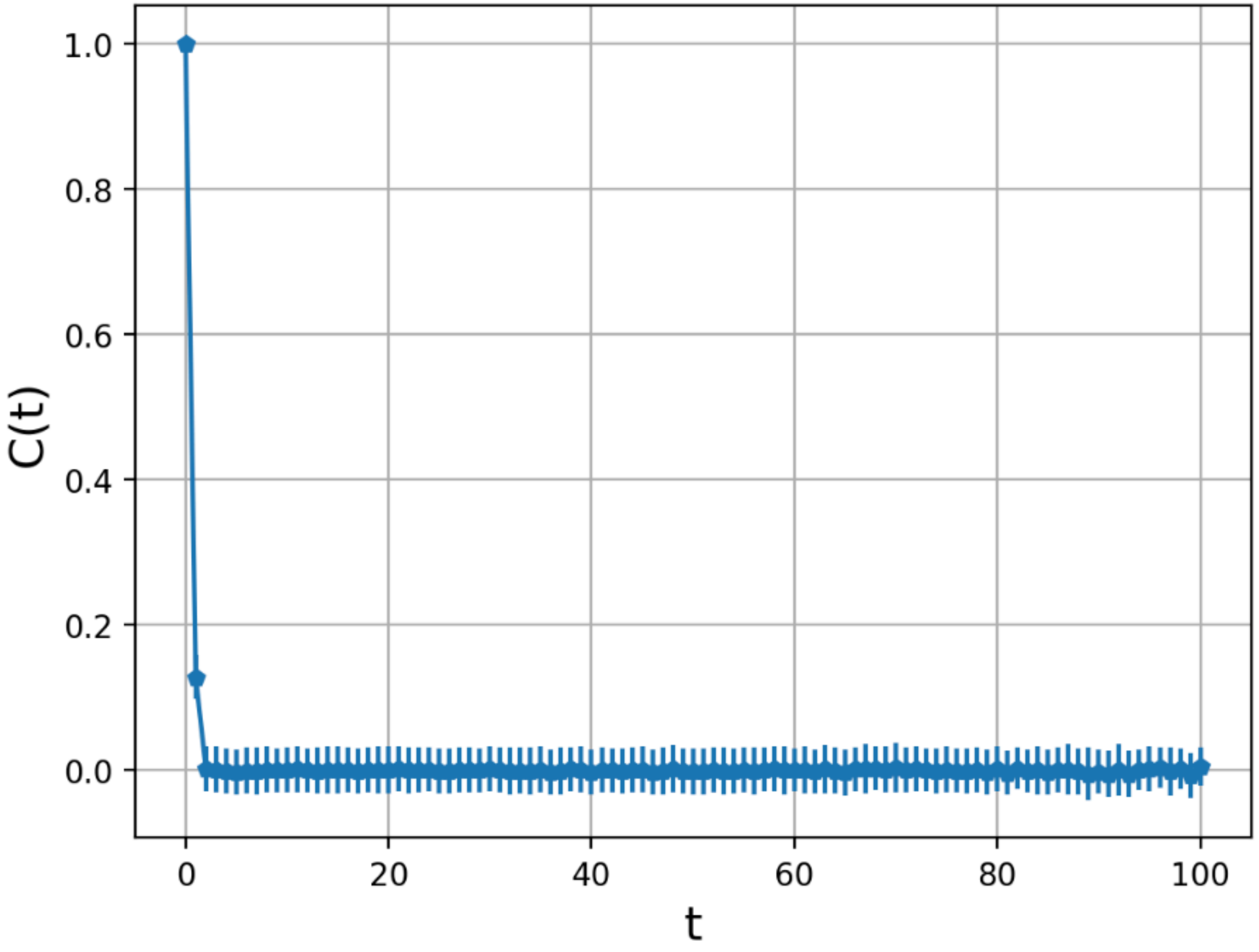
302x4 spins

Ensemble of size 100

1000 time-steps ($\Delta t = 1$)

(MC sweep = time-step)





Correlation function

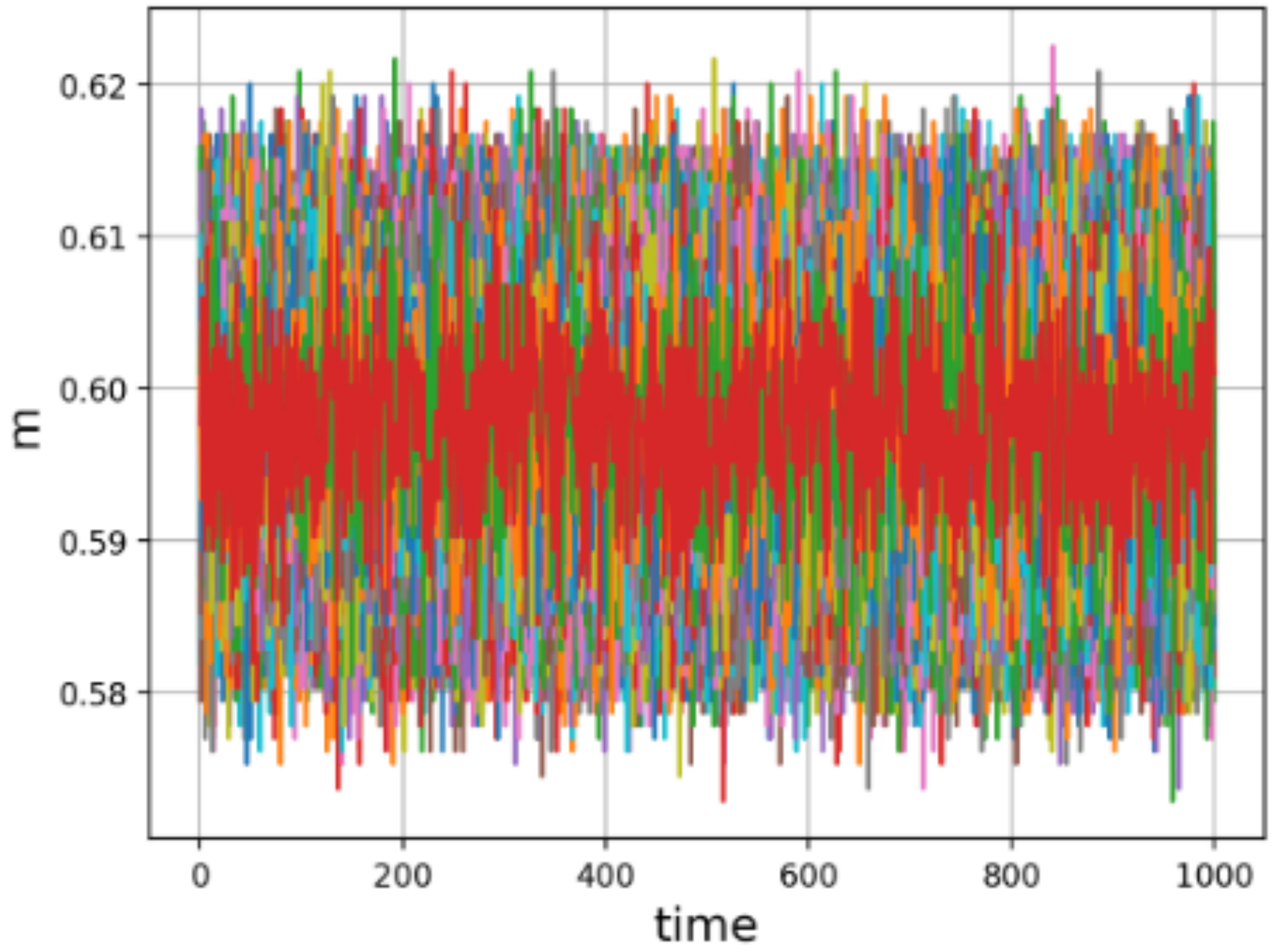
Trained RBM (Zephyr top)
(Making biases=0 yields same result)

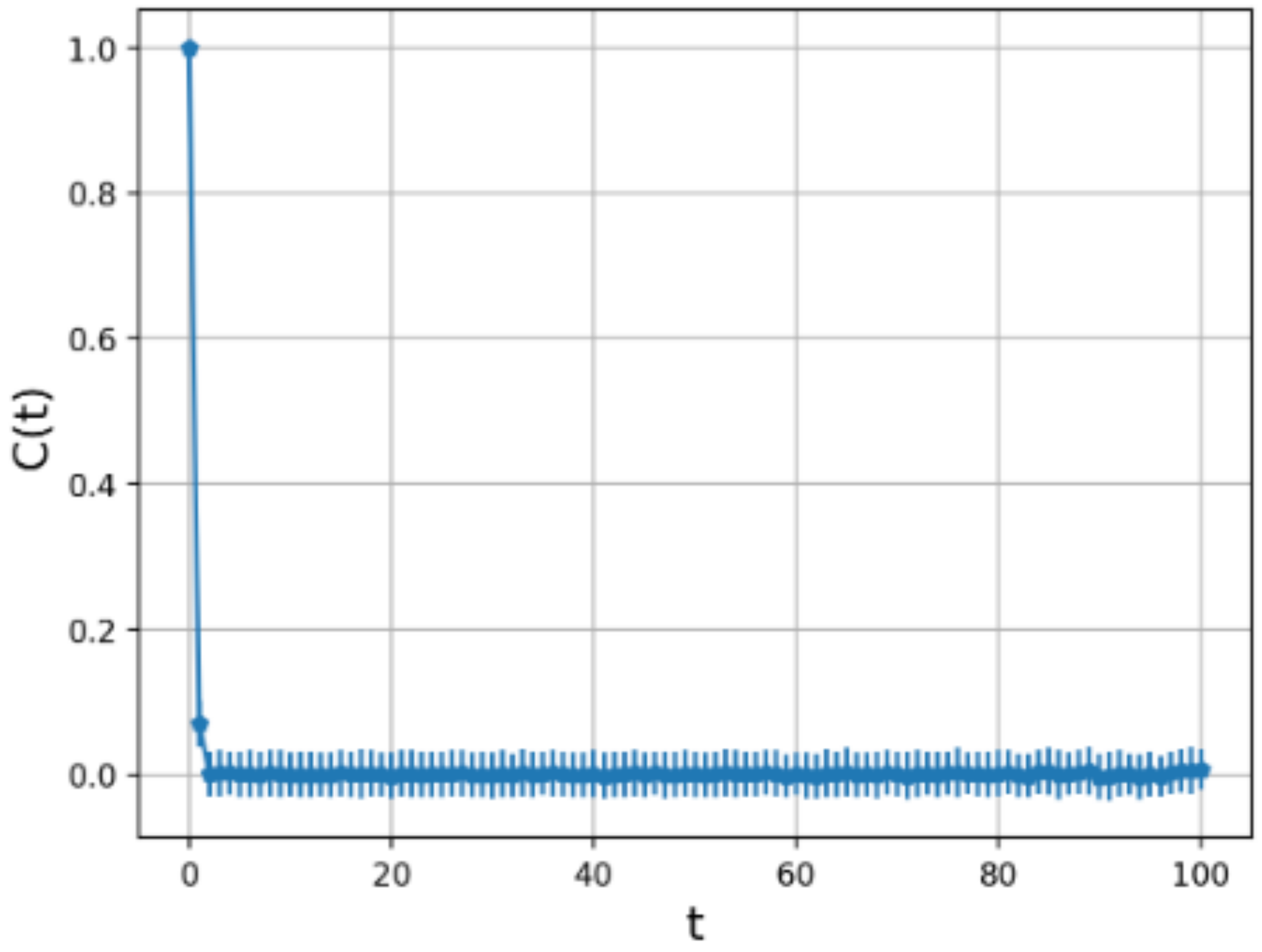
302x4 spins

Ensemble of size 100

1000 time-steps ($\Delta t = 1$)

(MC sweep = time-step)





Correlation function

Trained RBM New scheme (Zephyr top)
(Making biases=0 yields same result)

302x4 spins

Ensemble of size 100

1000 time-steps ($\Delta t = 1$)

(MC sweep = time-step)