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

Calo4pQVAE: Progress and updates

Oct 10 2024

TRIUMF Canada's particle accelerator centre
Centre canadien d'accélération des particules

Summary

✓ Weight decay removal?

✓ Papers status

✓ Using the QPU w/ Pegasus and in Zephyr

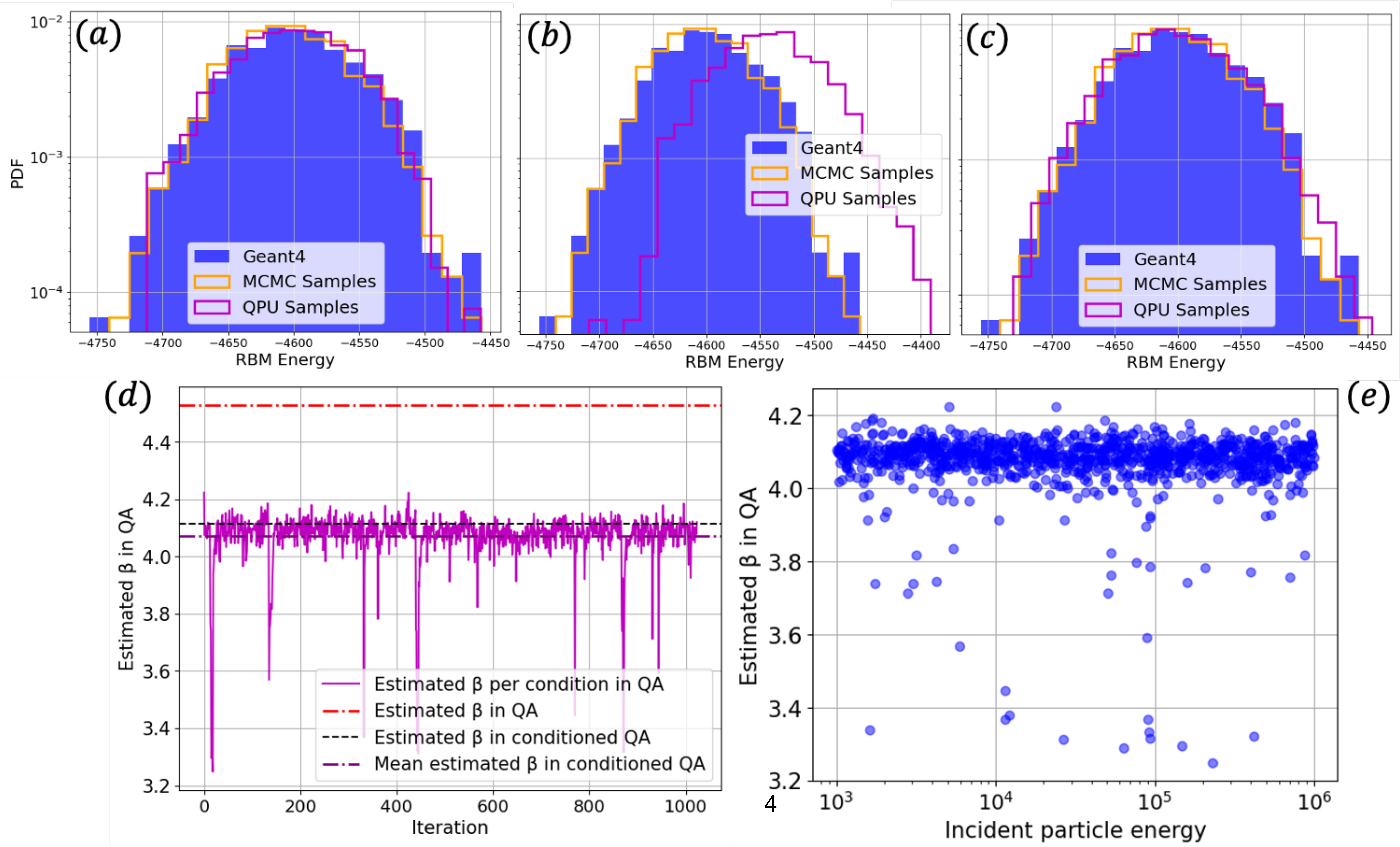
✓ High temperature gradient approximation for trained RBMs

- RBM and Diffusion model equivalence.
- Relaxation time in RBMs

Papers status

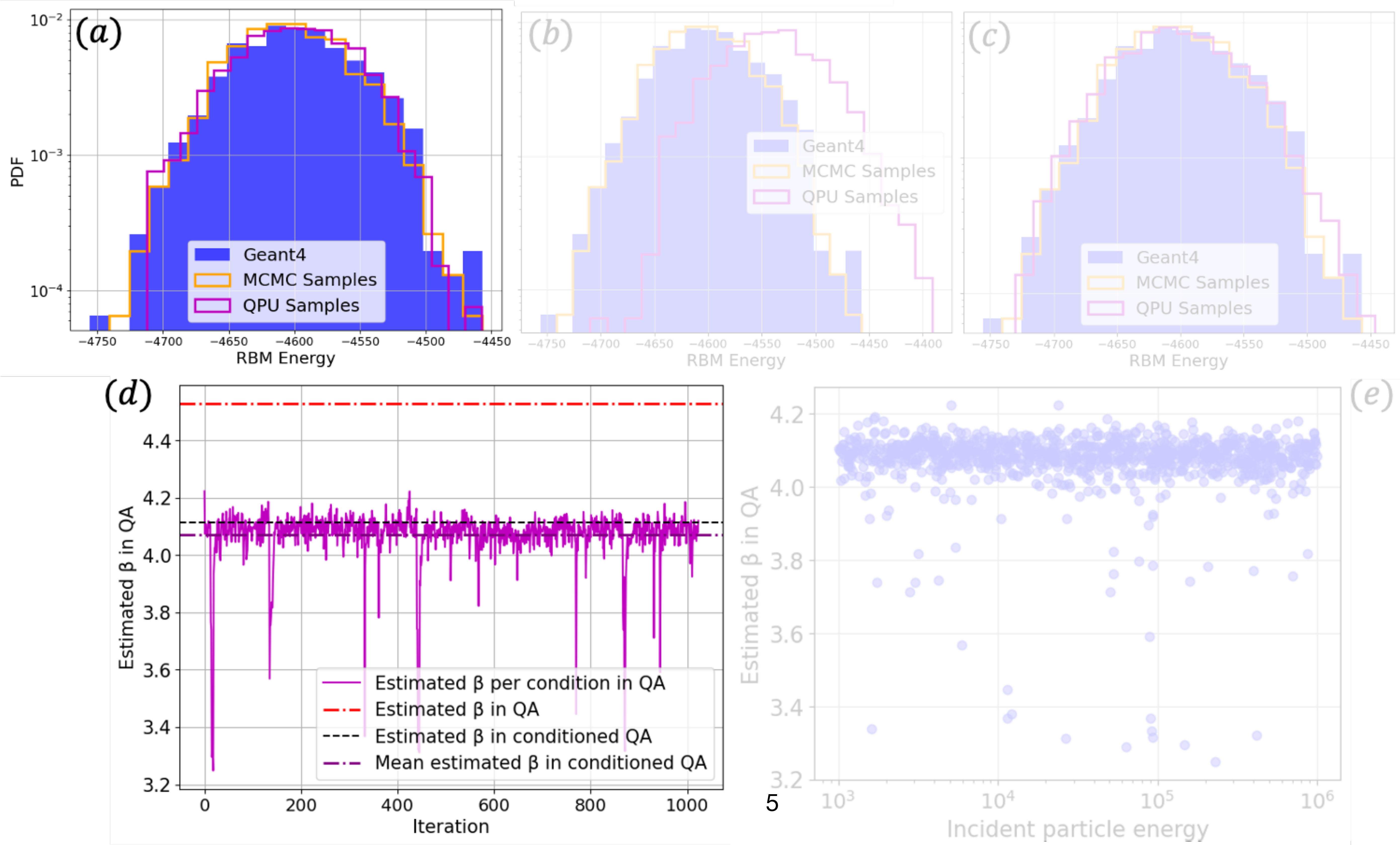
- EPJC -> Rebuttal submitted?
- iEEE QCE Conf -> haven’t seen the proceedings online yet
- Neurips ML4Phys -> Got accepted!
- PRX draft -> on countdown for submission

Using QPU w/ Pegasus



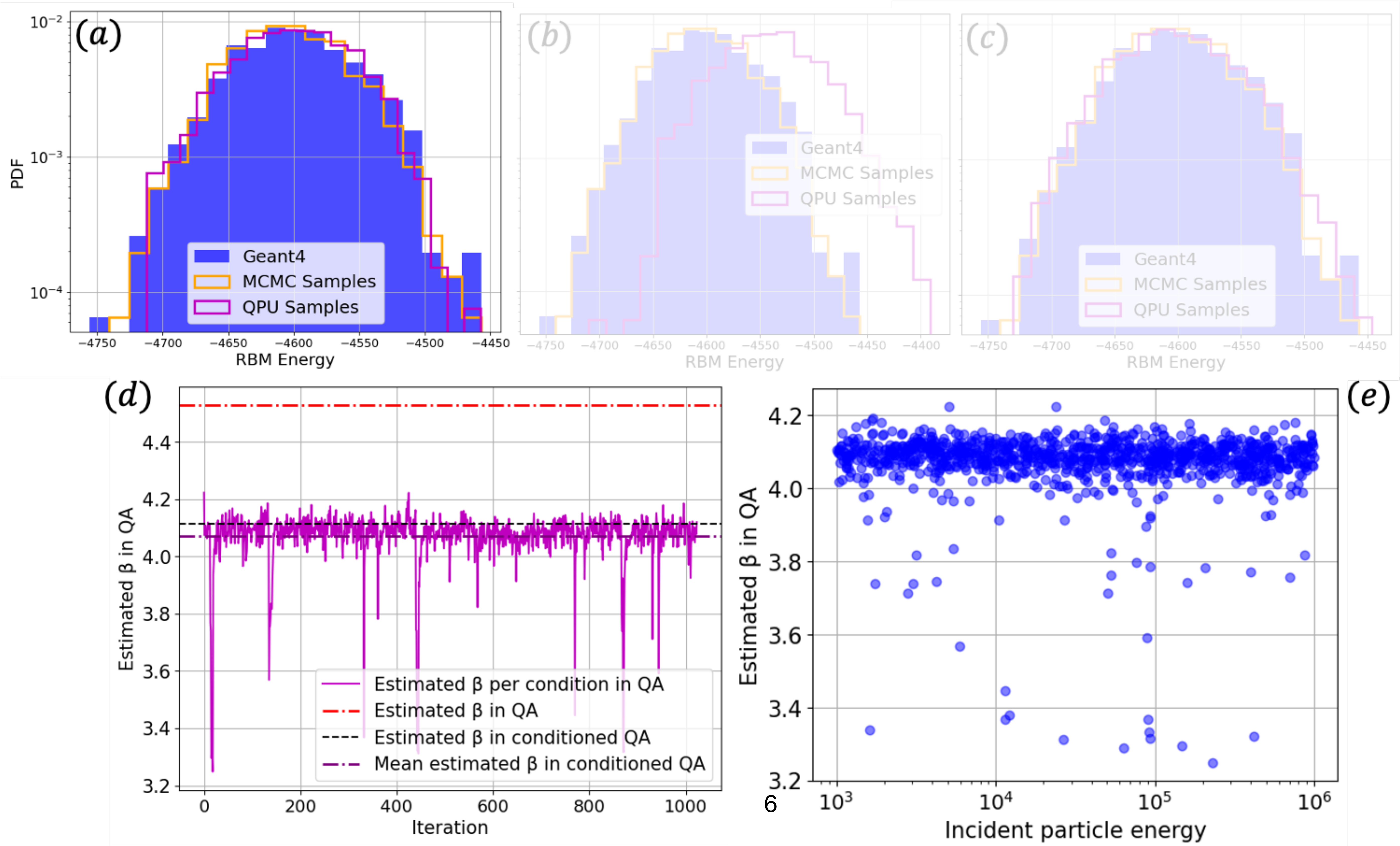
Using QPU w/ Pegasus

We estimate the QA inverse Temperature before generating each sample.



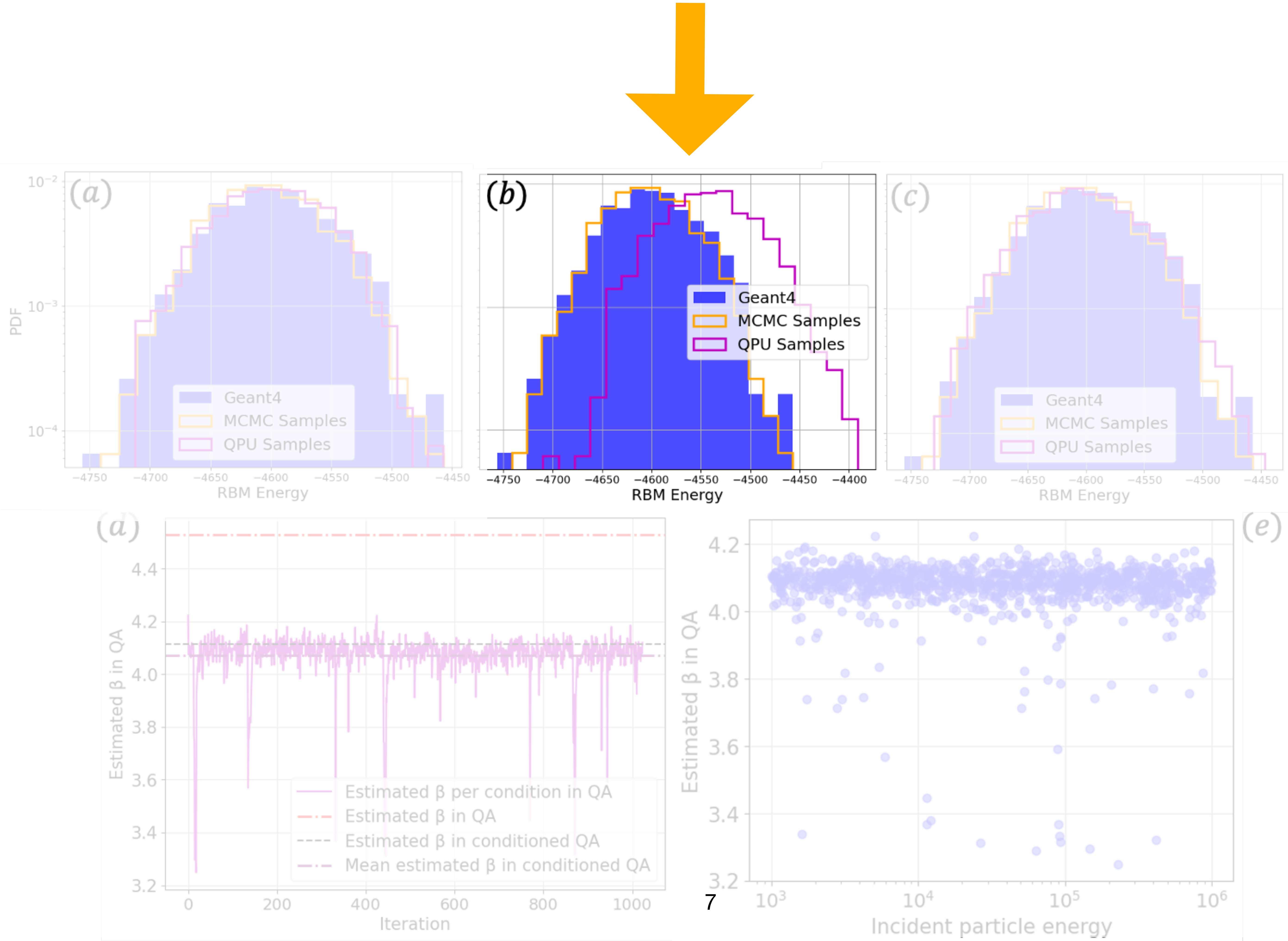
Using QPU w/ Pegasus

We estimate the QA inverse Temperature before generating each sample.



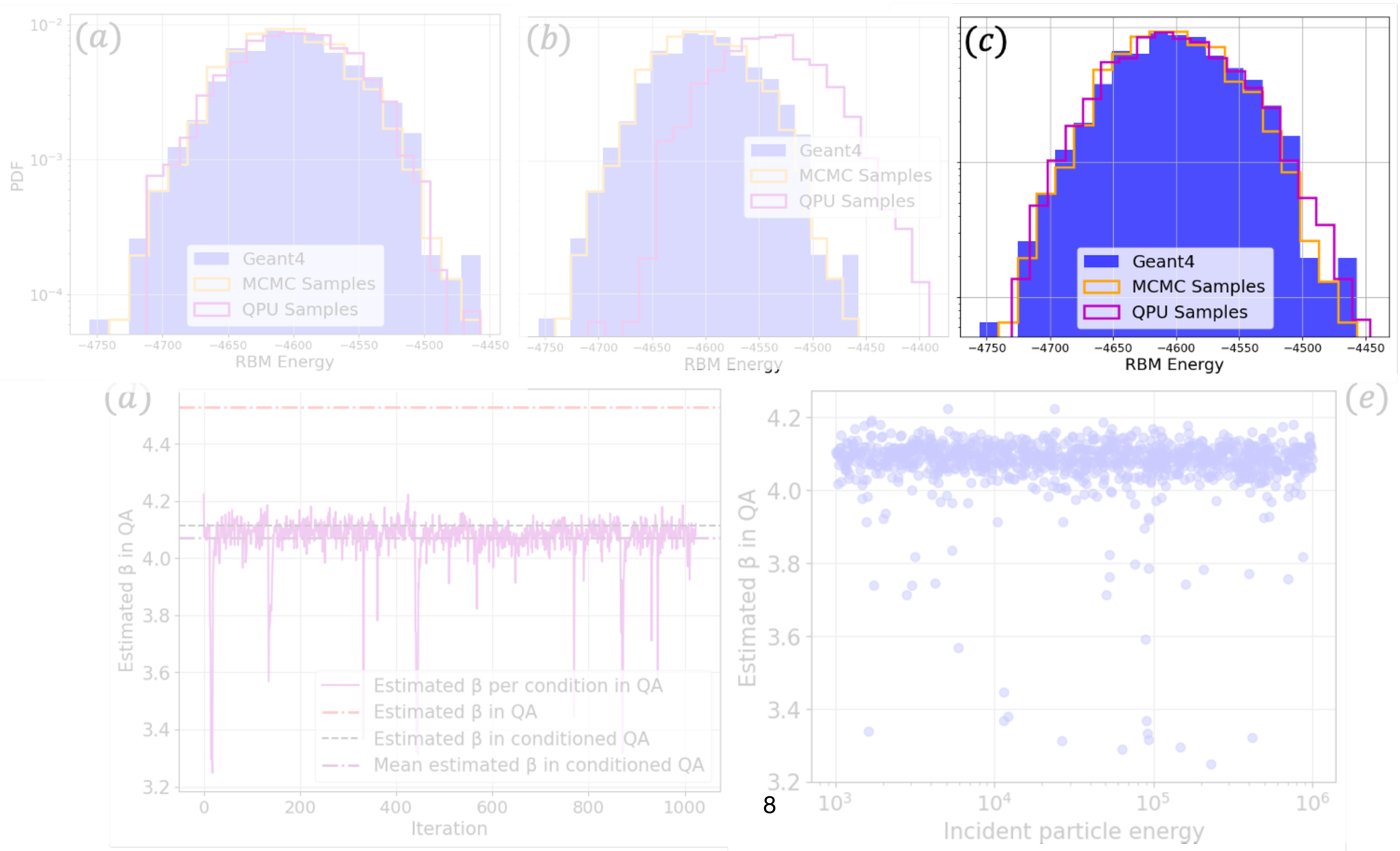
Using QPU w/ Pegasus

We estimate the QA inverse Temperature 1 time. Then generate all samples.



Using QPU w/ Pegasus

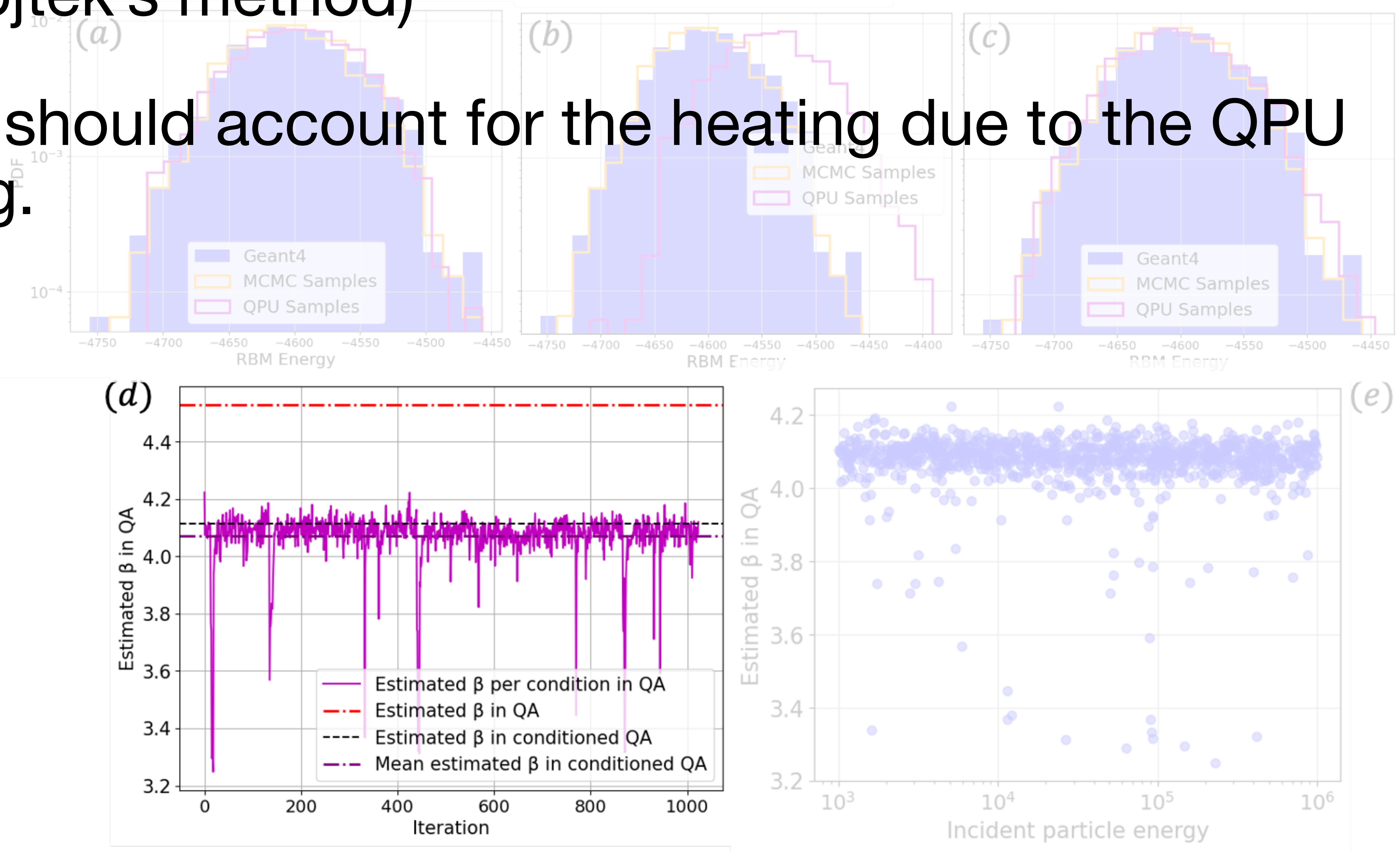
We estimate the QA inverse Temperature 1 time. Then generate all samples. Wait 2.5 seconds between samples



Using QPU w/ Pegasus

What if we estimate the QA inverse by generating 1 sample per API call? (Wojtek's method)

This way we should account for the heating due to the QPU programming.

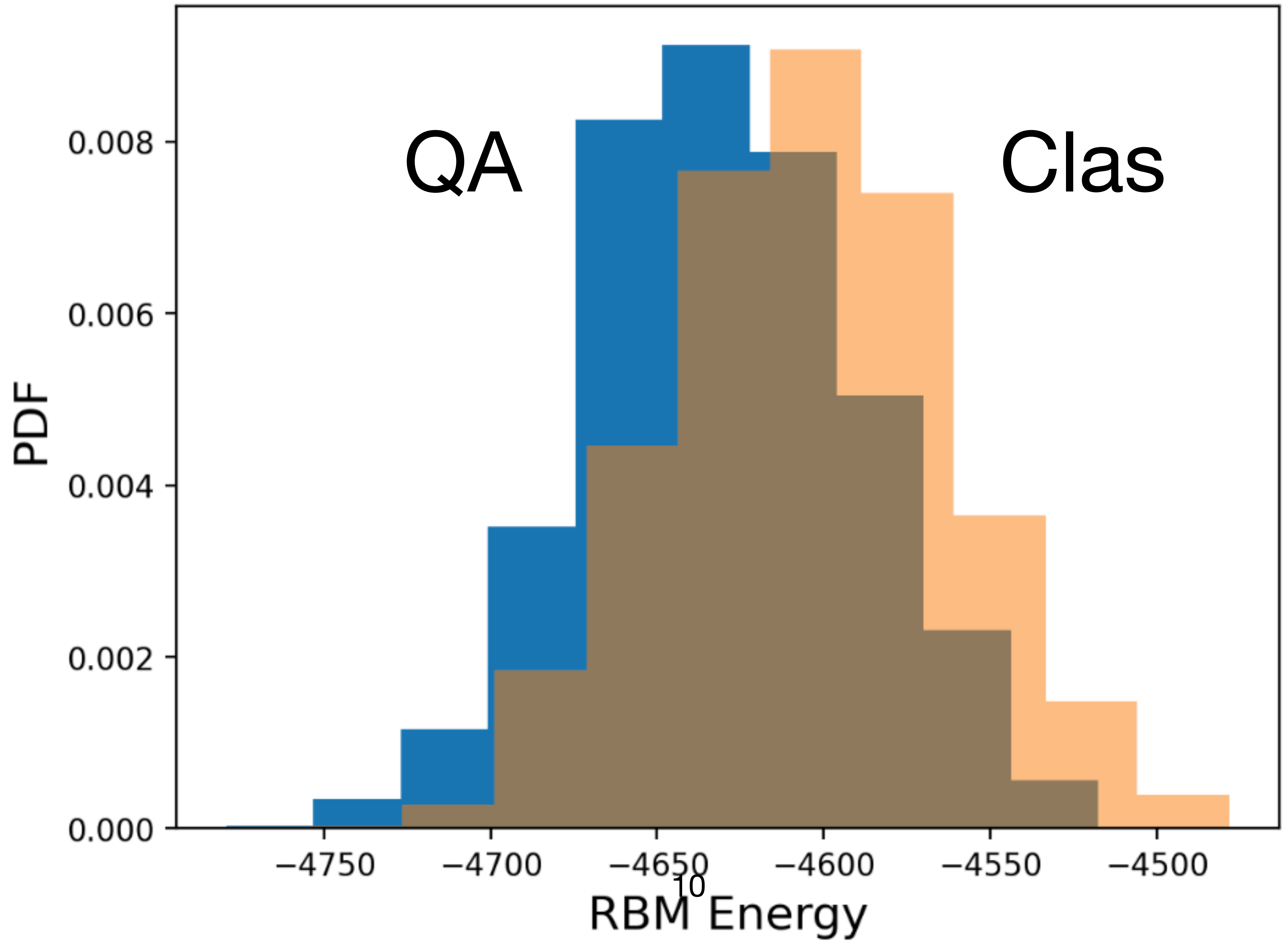


Using QPU w/ Pegasus

What if we estimate the QA inverse by generating 1 sample per API call?

This way we should account for the heating due to the QPU programming.

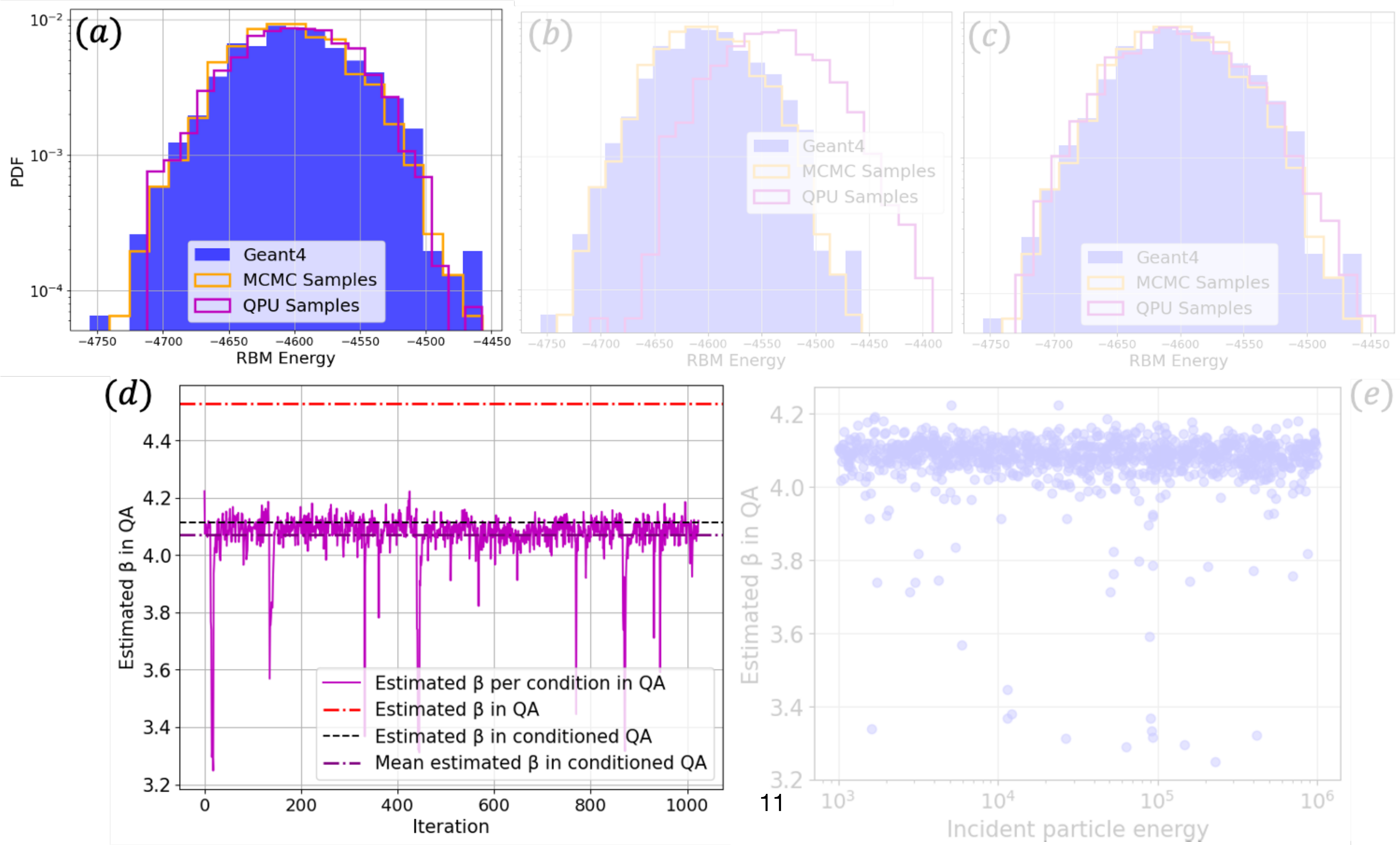
$$\beta_{QA} \approx 2.6$$



Using QPU w/ Pegasus

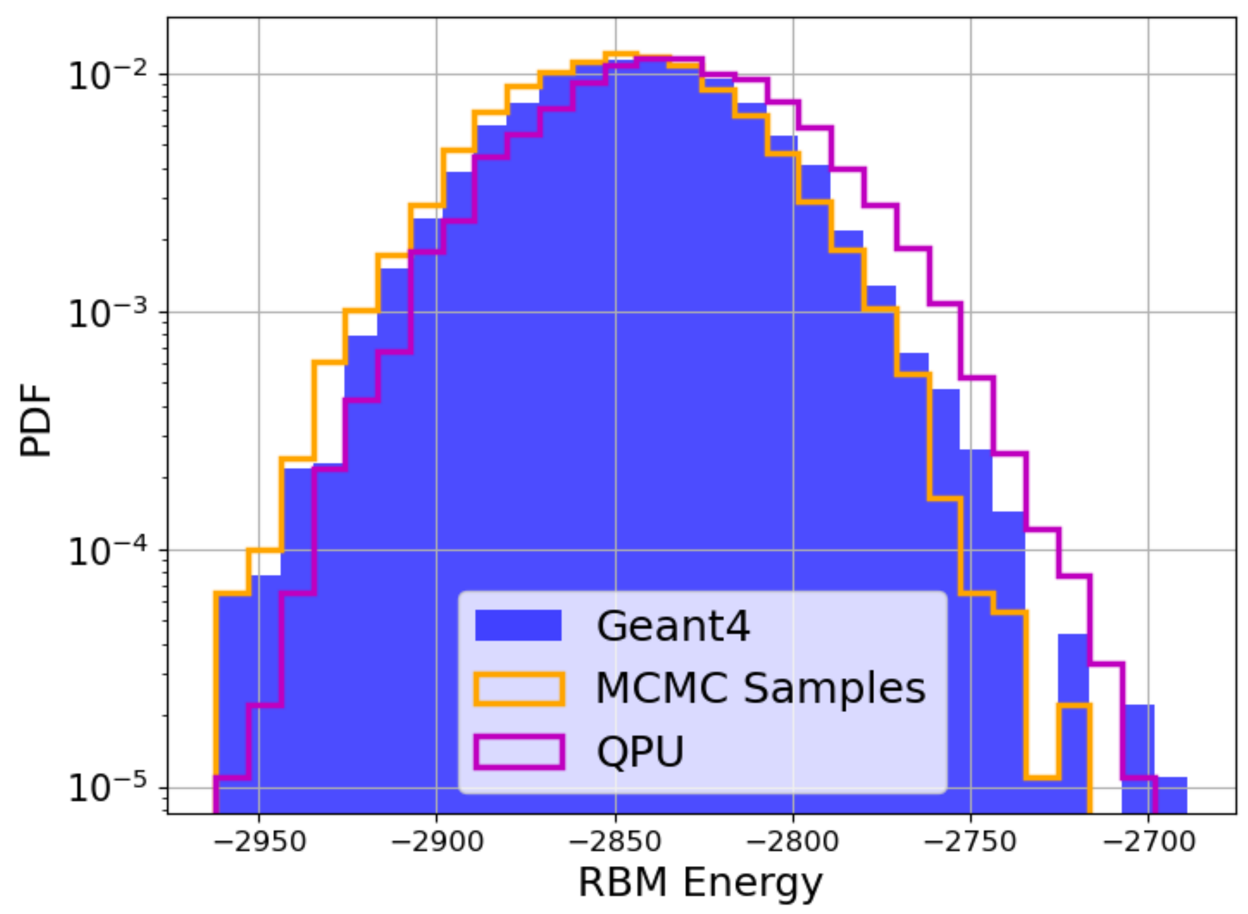
We estimate the QA inverse Temperature before generating each sample.

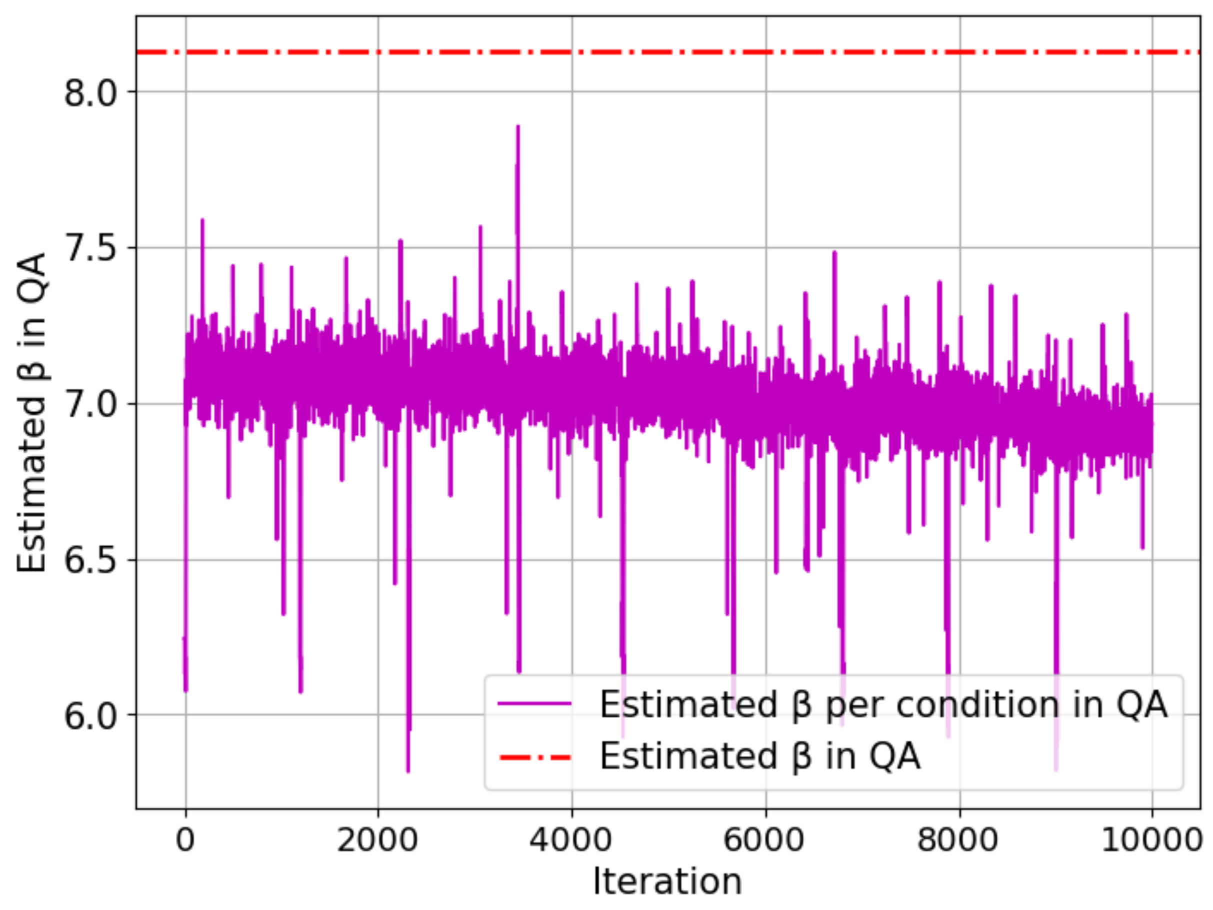
Winning method!

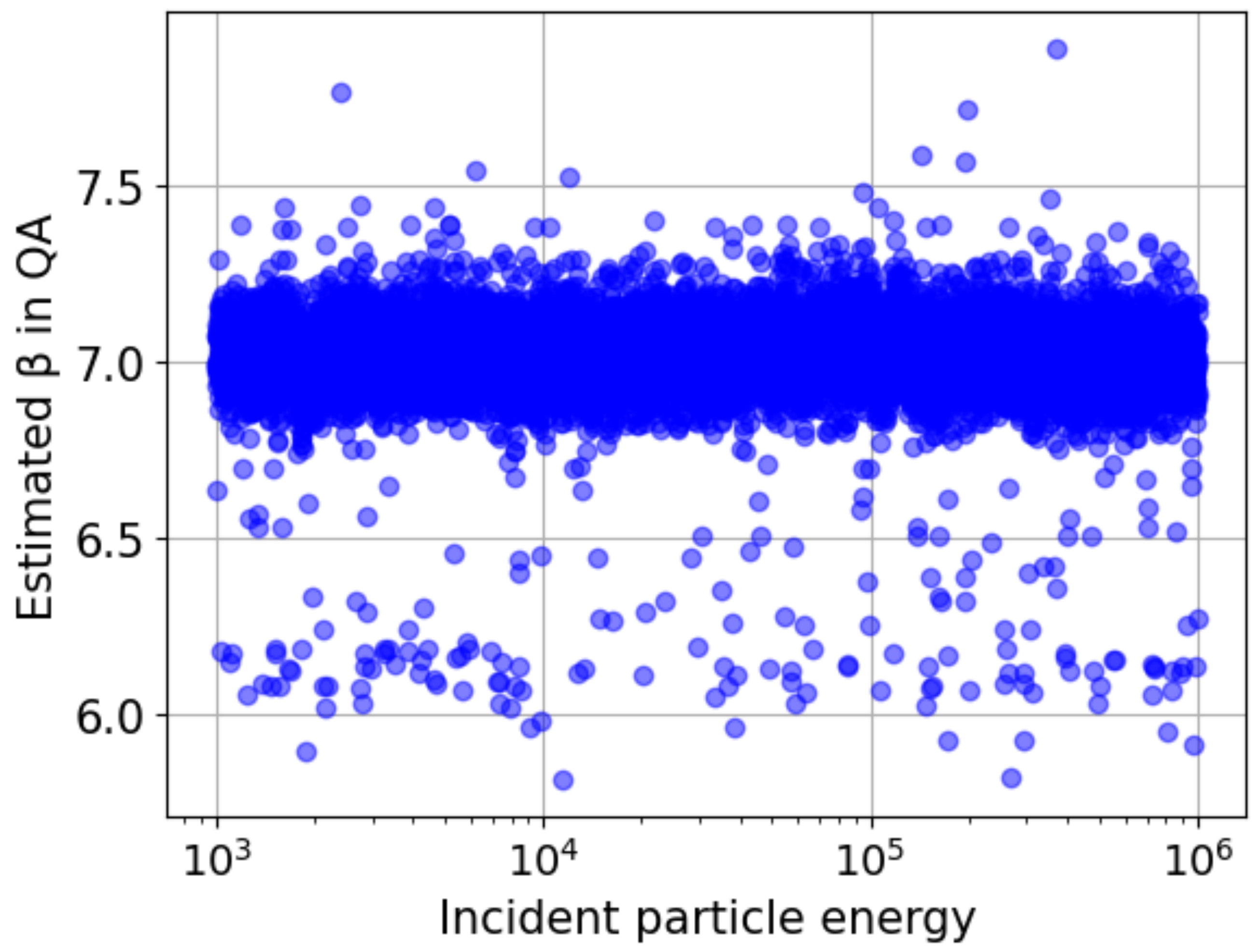


Using QPU w/ Zephyr :: Winning method

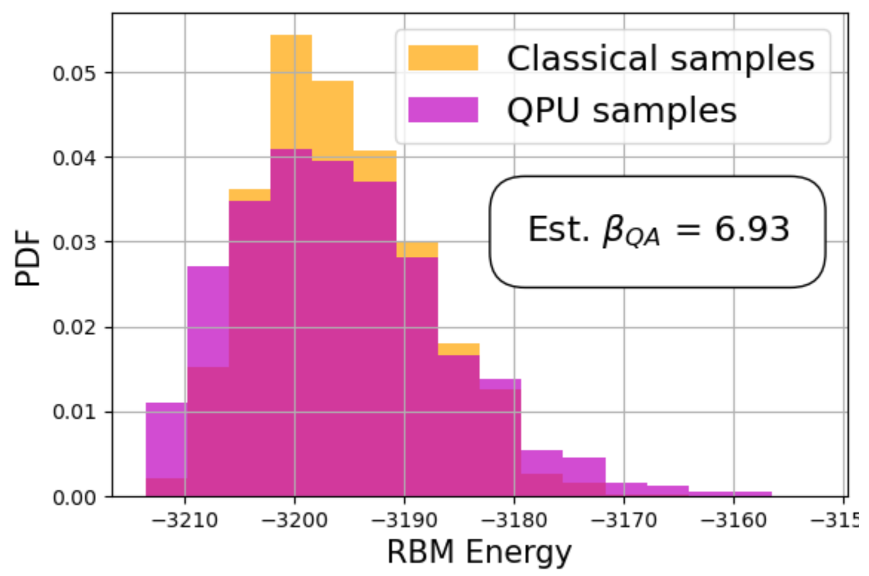
Model A

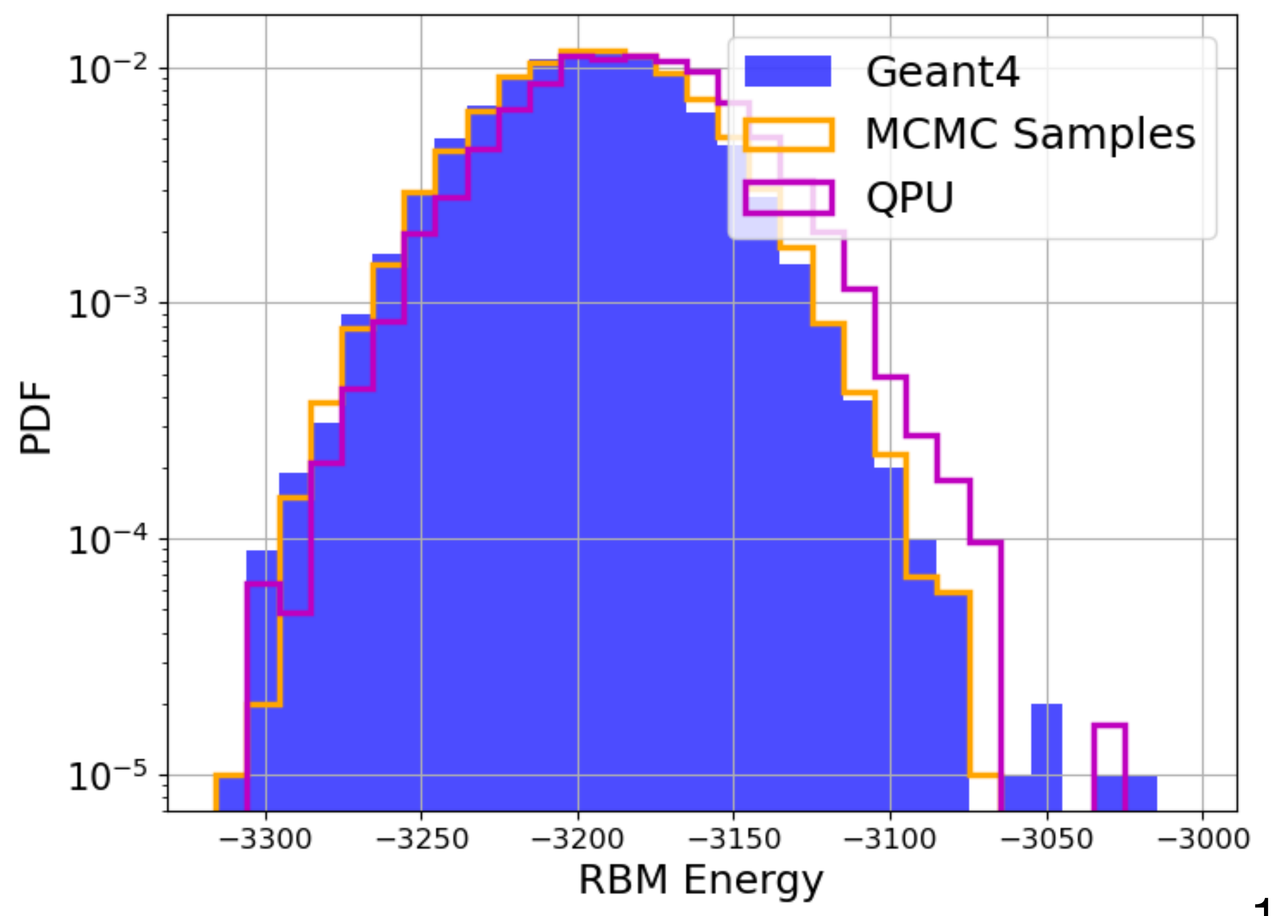


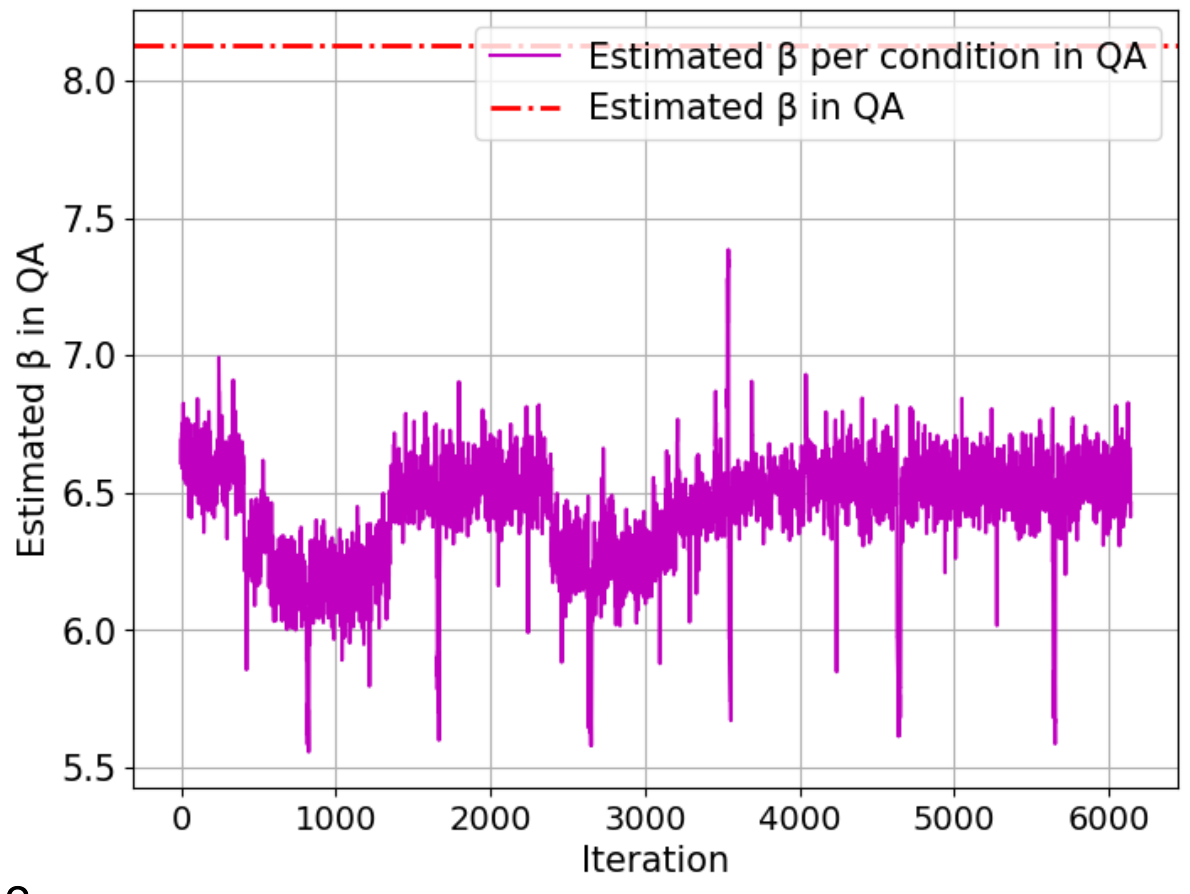


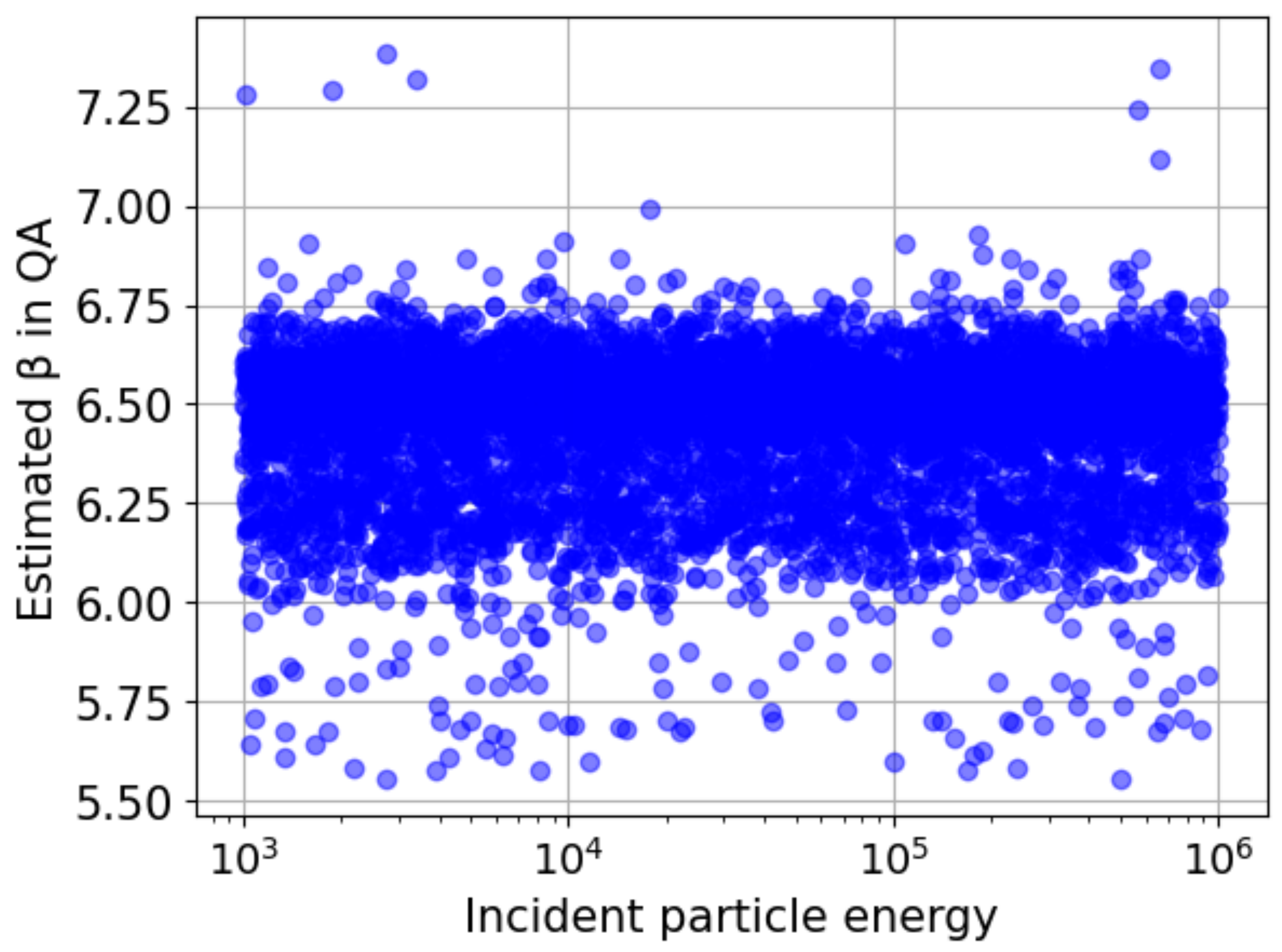


Model B



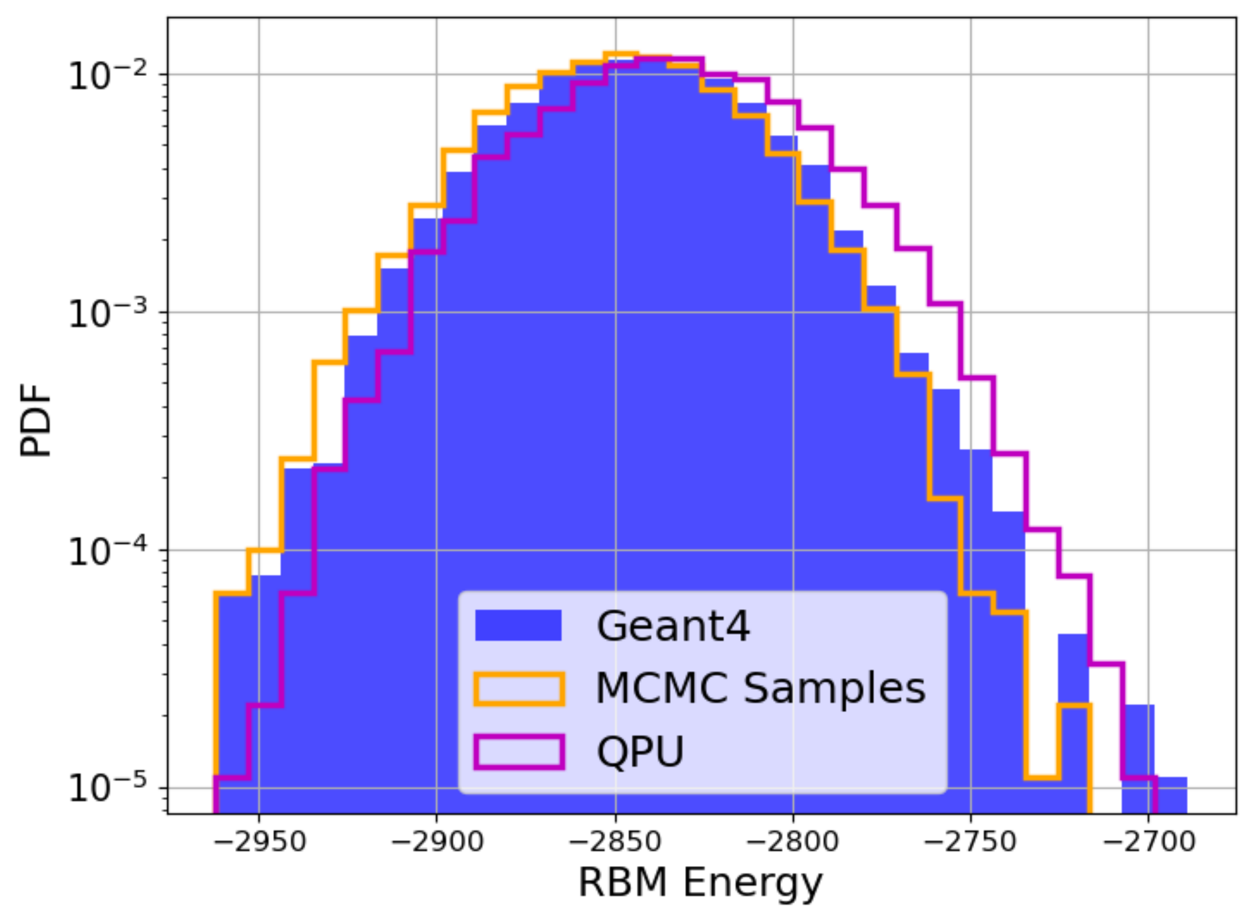


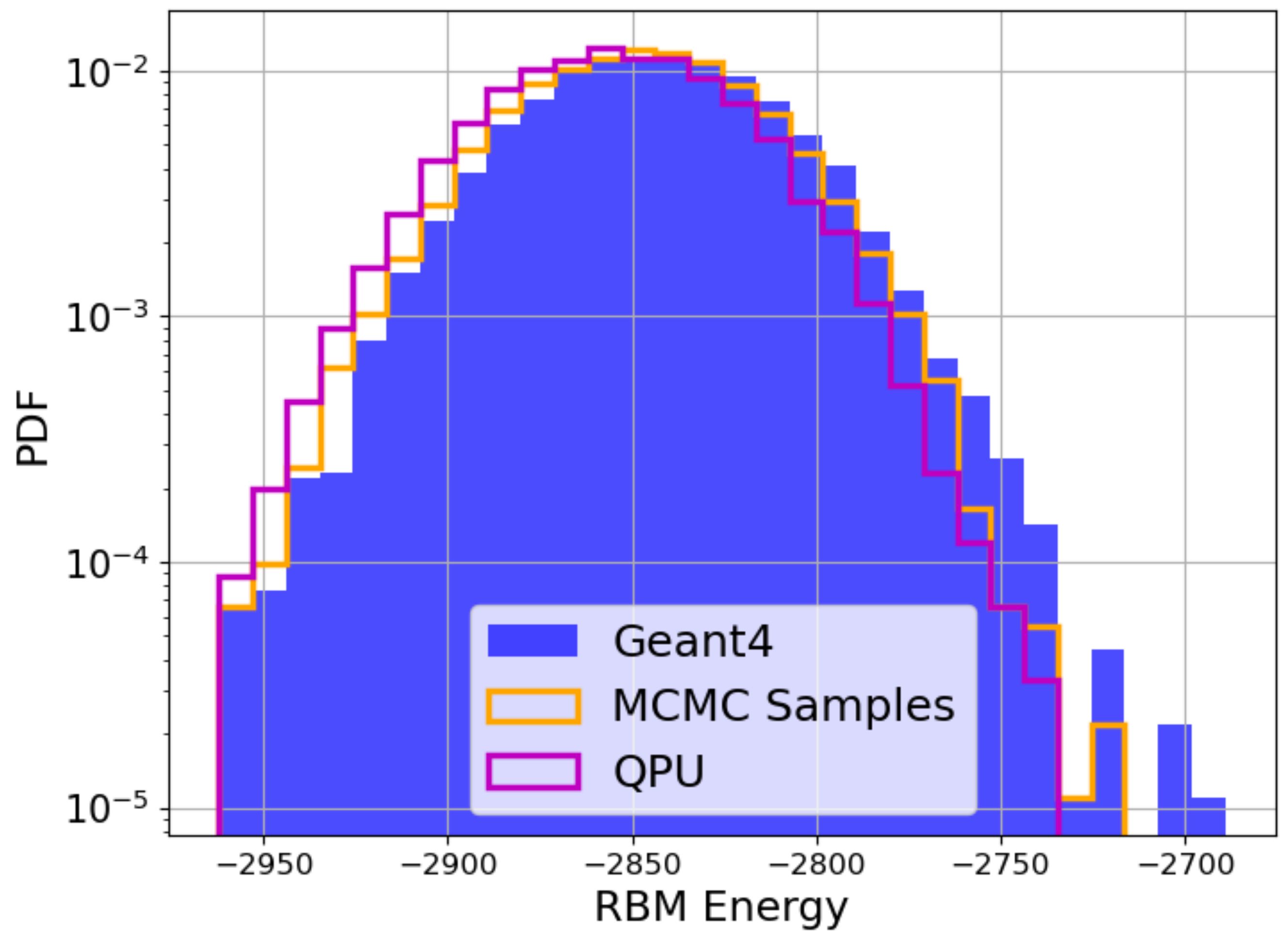




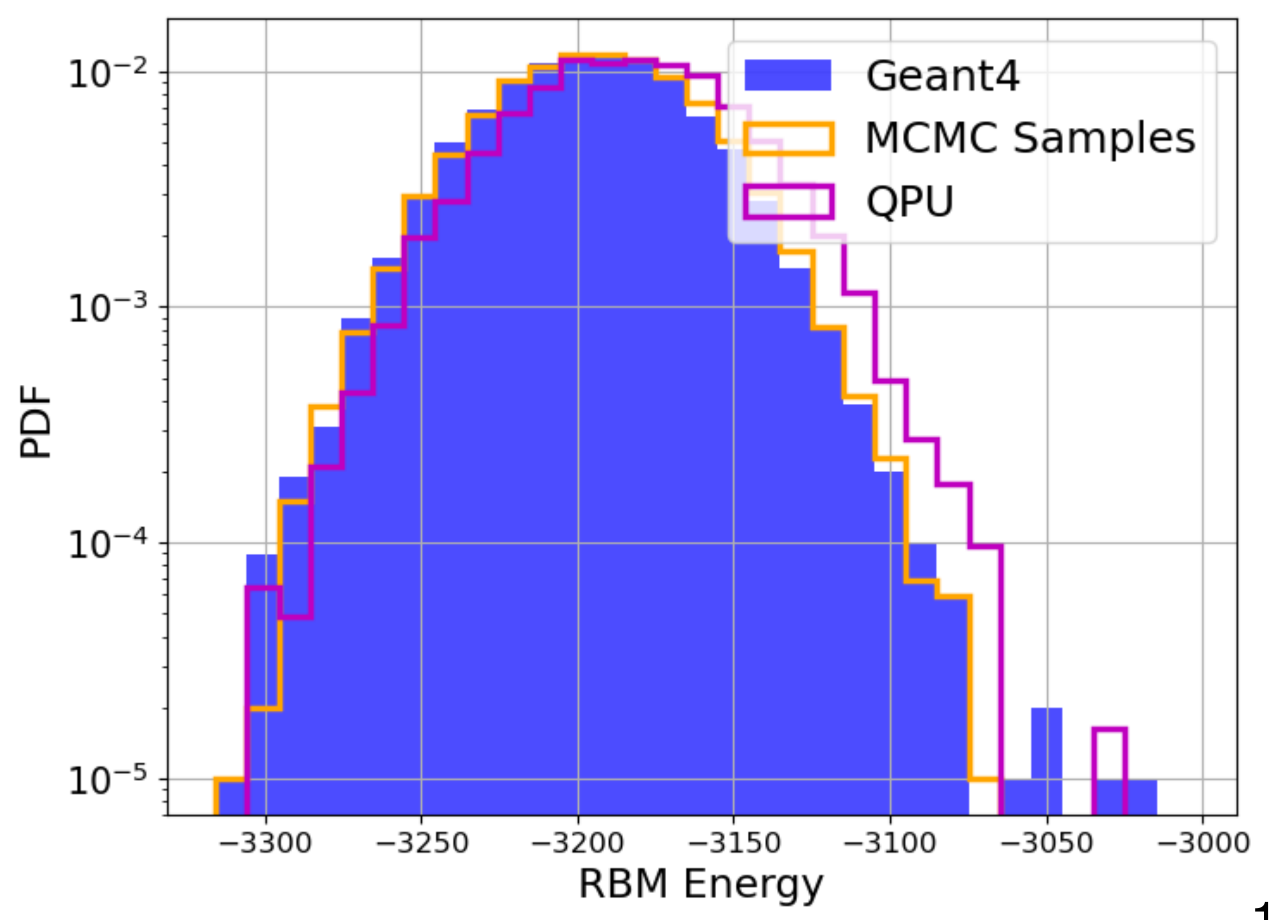
Using QPU w/ Zephyr :: vs Wojtek's method

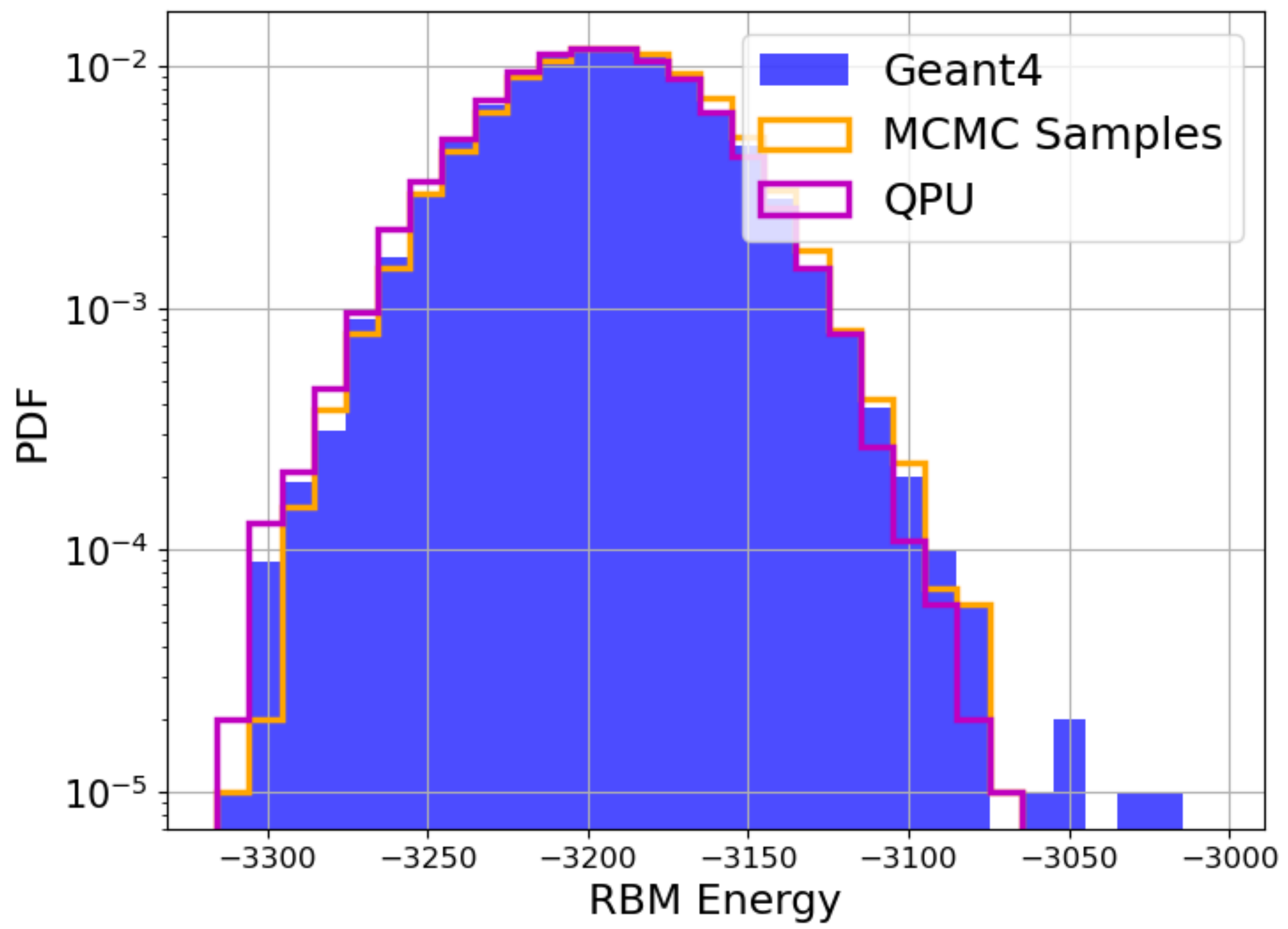
Model A





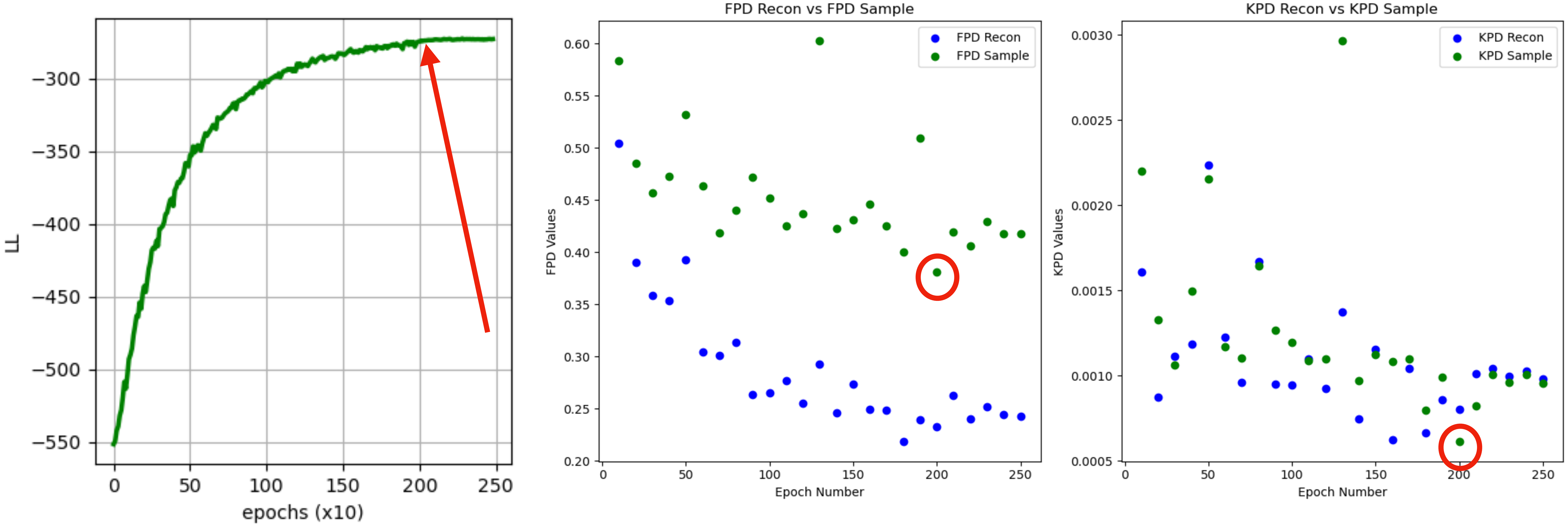
Model B





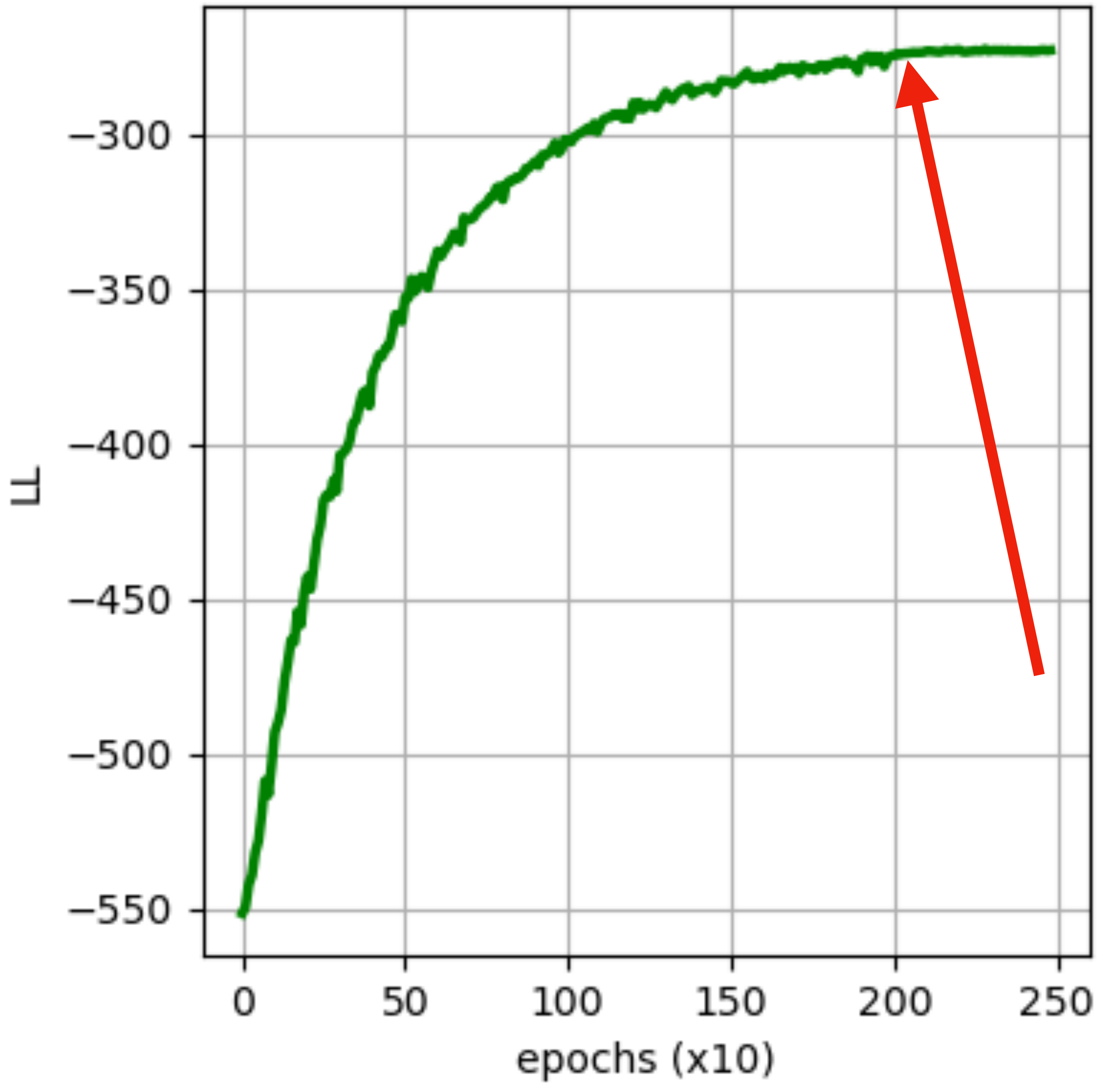
Using QPU w/ Zephyr :: vs Wojtek's method

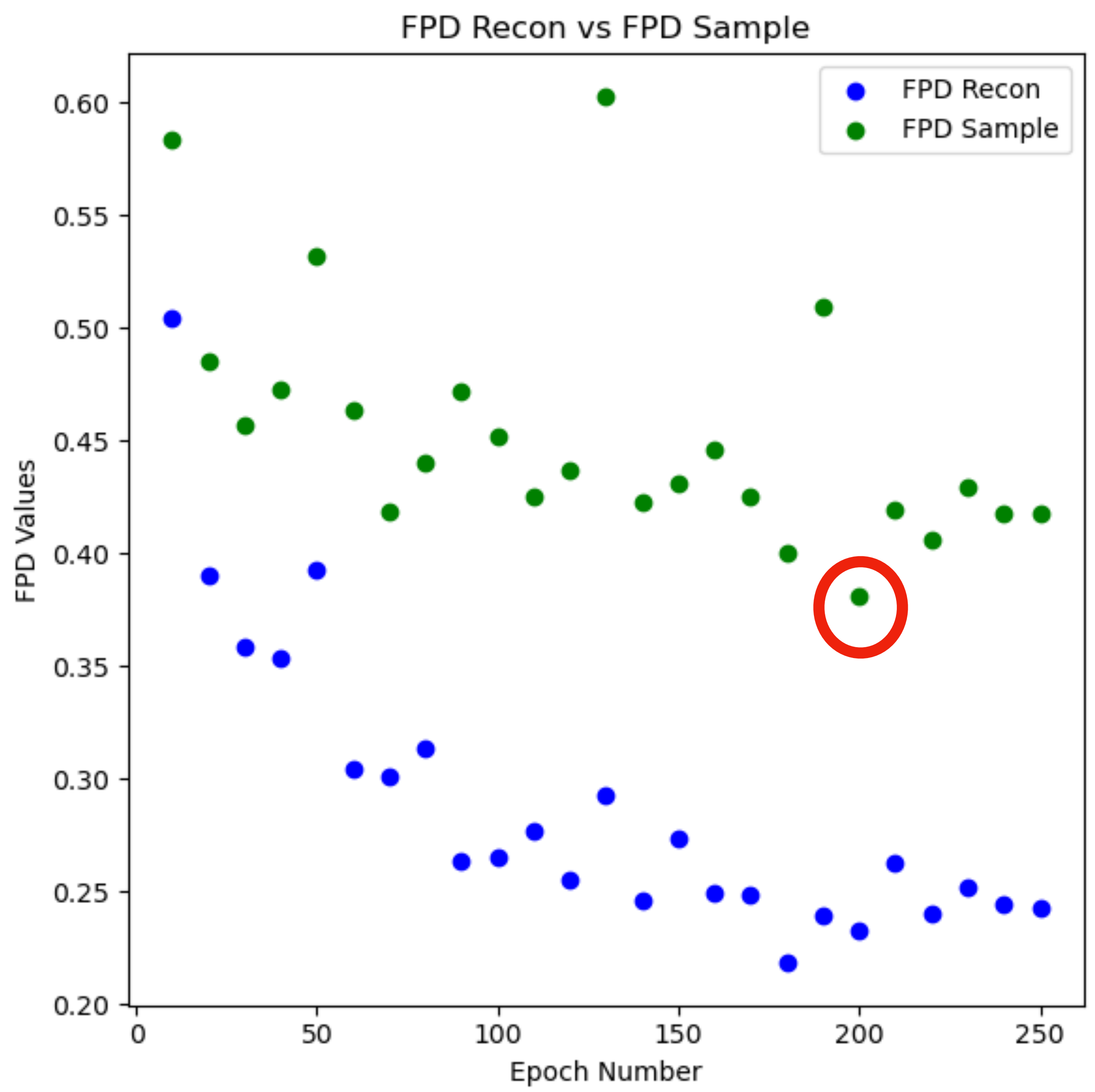
Model B

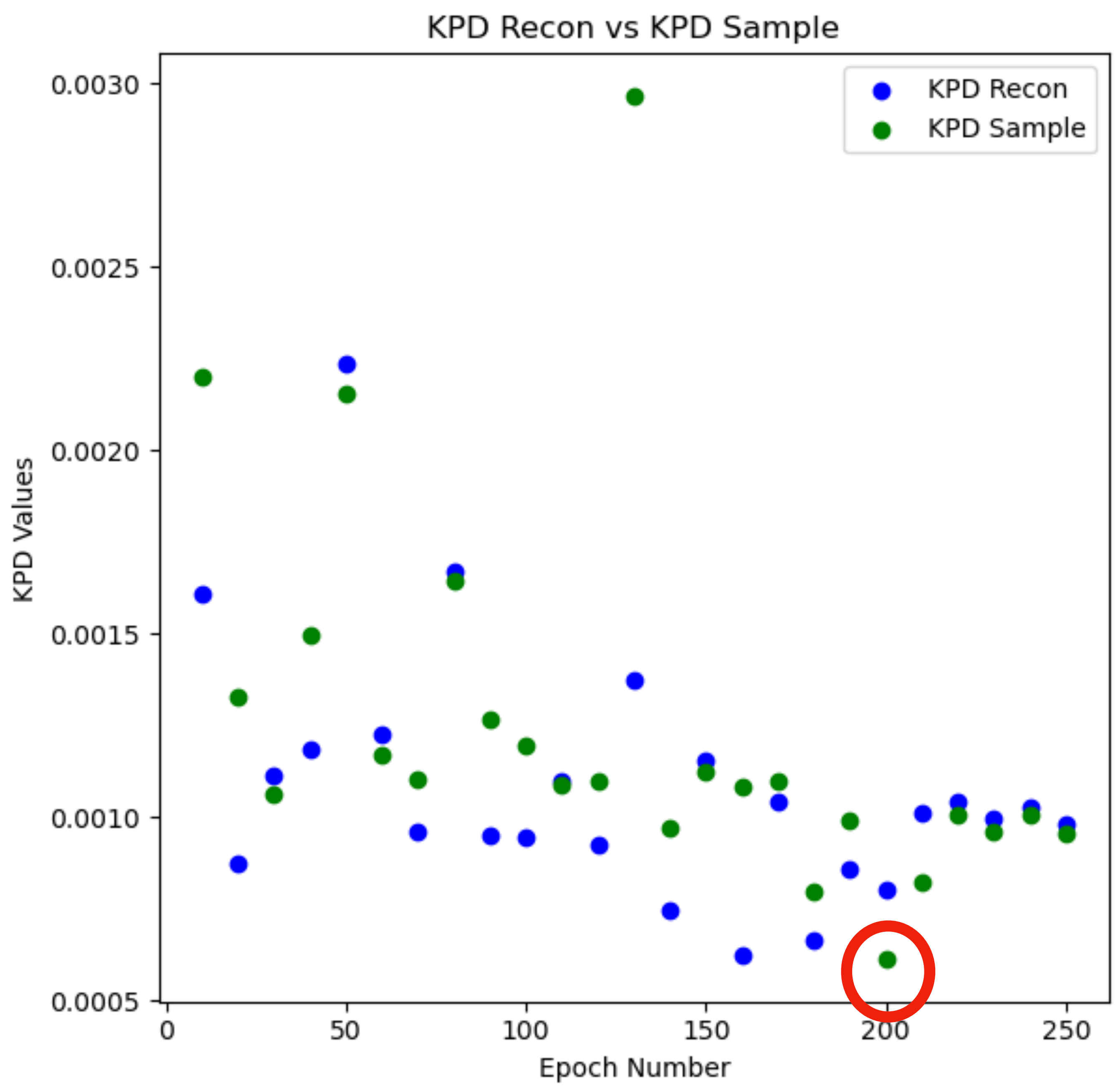


Using QPU w/ Zephyr :: vs Wojtek's method

Model B







Time required to train using QPU:

(1 sample generation time)X(# of samples)X(epochs)

(20ms)X(100k)X(200) = 111.1 hrs

Using QPU w/ Zephyr :: vs Wojtek’s method

- Train Enc and Decoder and train QPU afterwards with a smaller sample.
- Discuss with dwave options and roadmaps

High Temperature gradient approximation

$$\langle E \frac{\partial E}{\partial \Theta} \rangle - \langle E \rangle \langle \frac{\partial E}{\partial \Theta} \rangle = 0$$

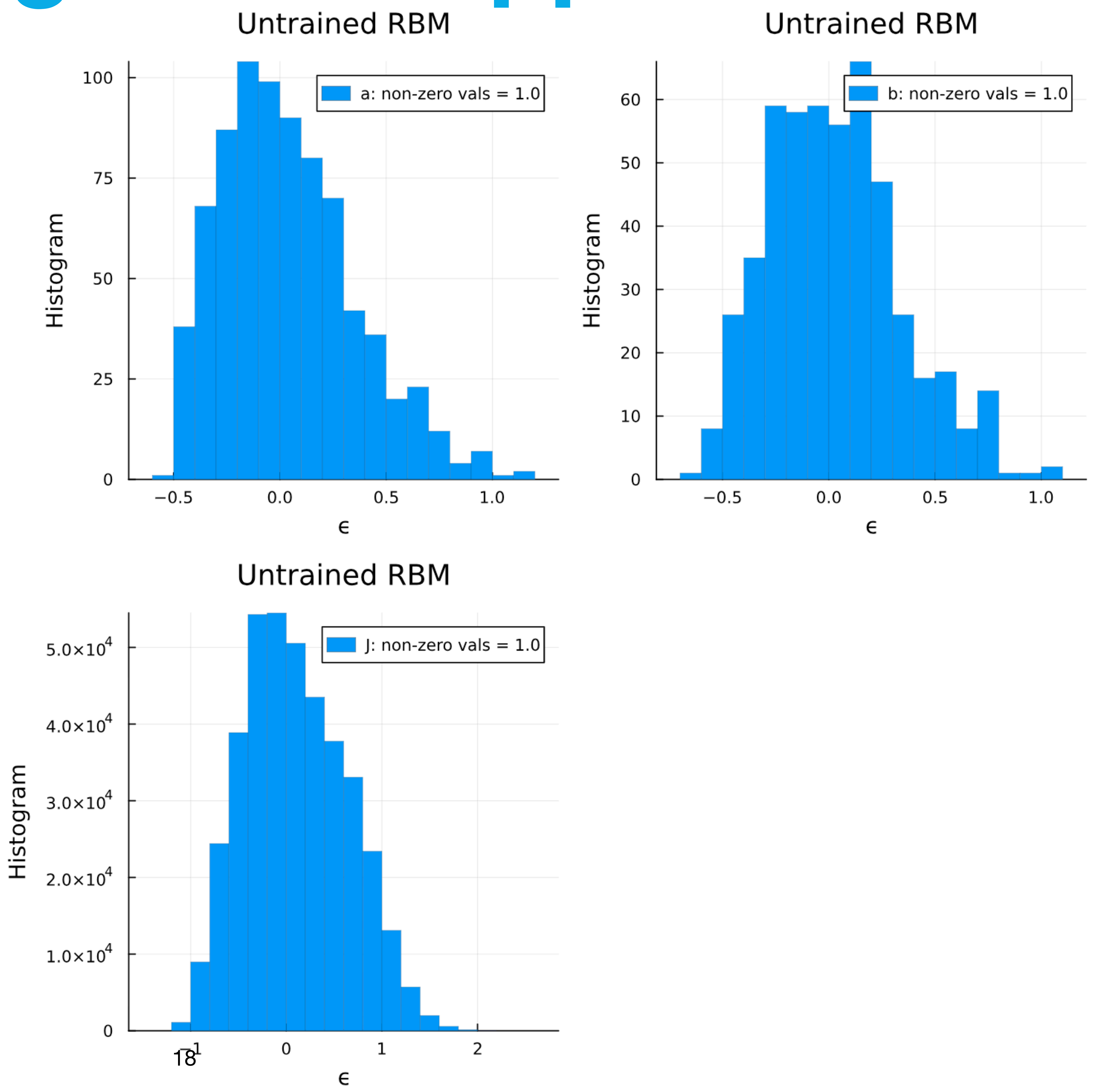
Condition

$$\epsilon = \frac{\langle E \frac{\partial E}{\partial \Theta} \rangle - \langle E \rangle \langle \frac{\partial E}{\partial \Theta} \rangle}{\langle \frac{\partial E}{\partial \Theta} \rangle}$$

1000 Gibbs sampling steps

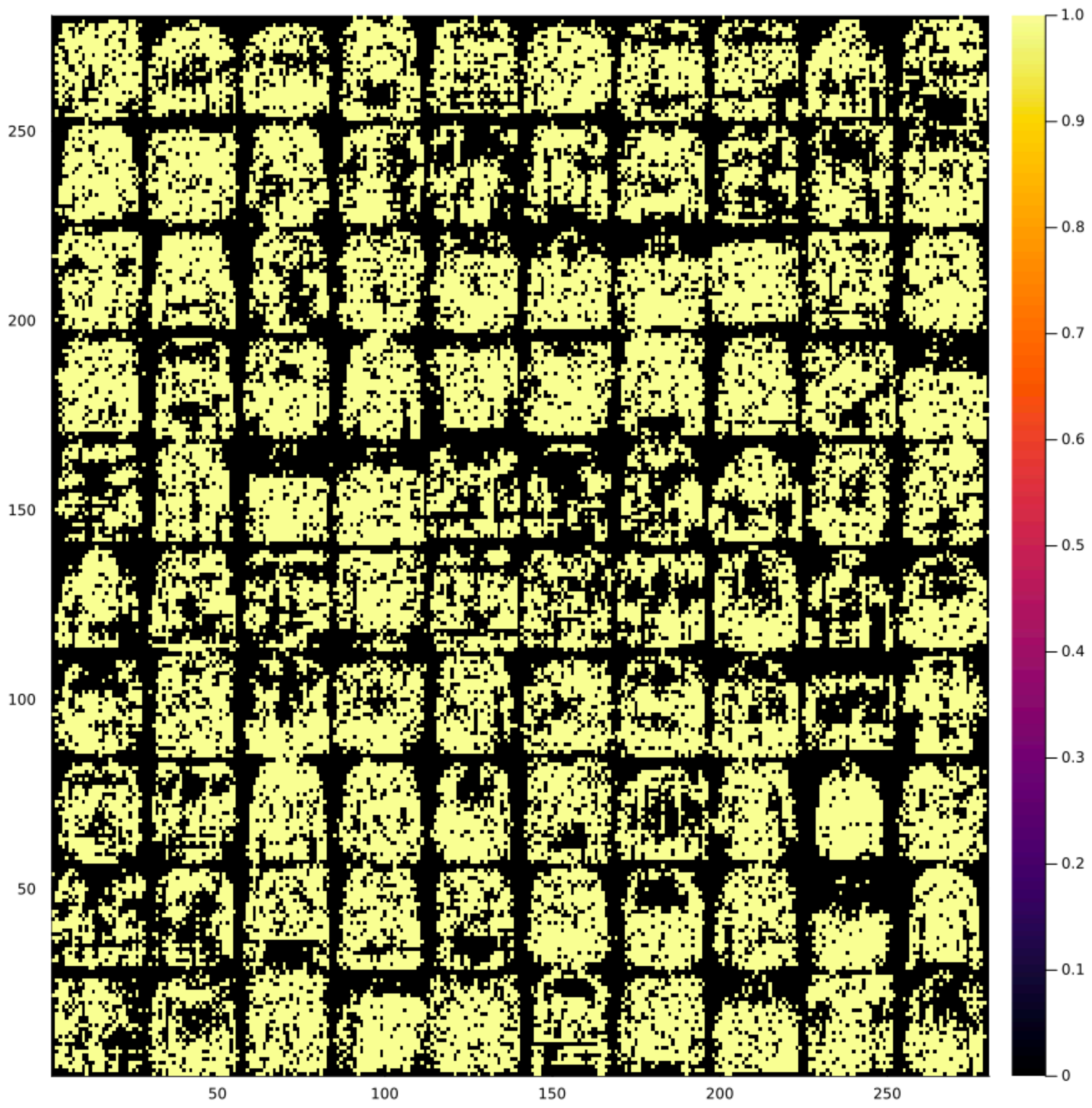
High Temperature gradient approximation

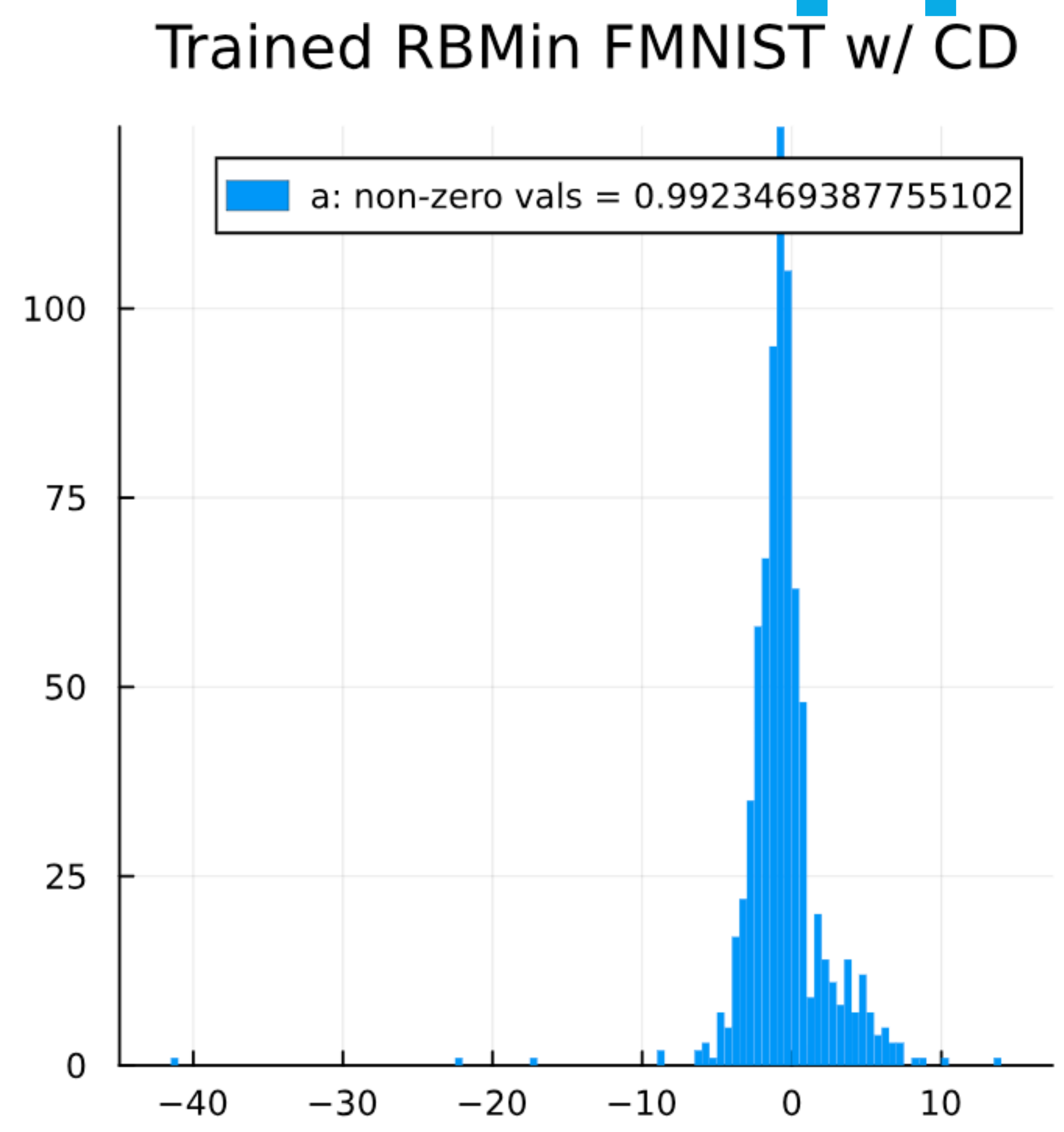
$$\epsilon = \frac{\langle E \frac{\partial E}{\partial \Theta} \rangle - \langle E \rangle \langle \frac{\partial E}{\partial \Theta} \rangle}{\langle \frac{\partial E}{\partial \Theta} \rangle}$$

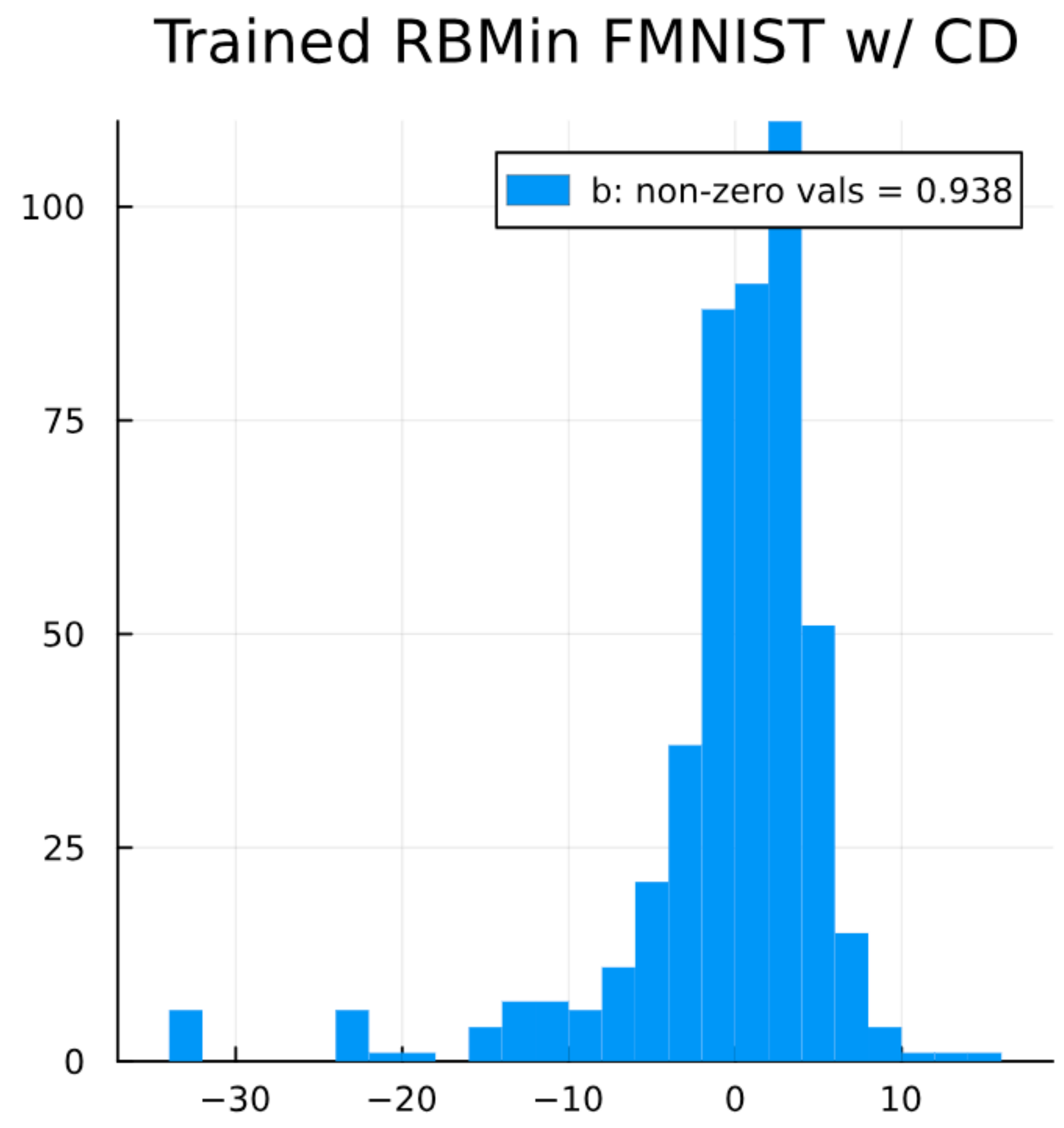


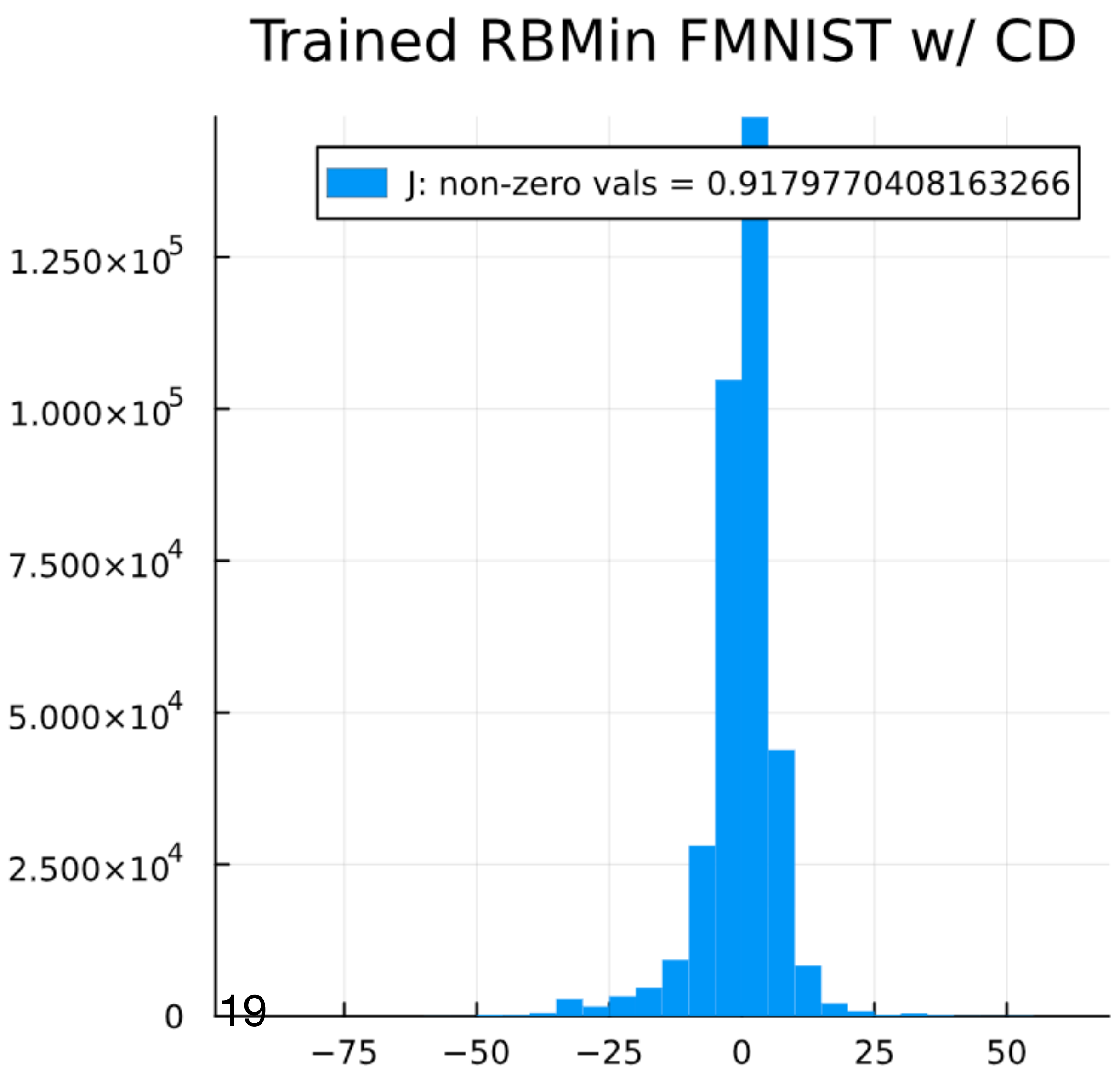
High Temperature gradient approximation

$$\epsilon = \frac{\langle E\frac{\partial E}{\partial \Theta}\rangle - \langle E\rangle\langle\frac{\partial E}{\partial \Theta}\rangle}{\langle\frac{\partial E}{\partial \Theta}\rangle}$$



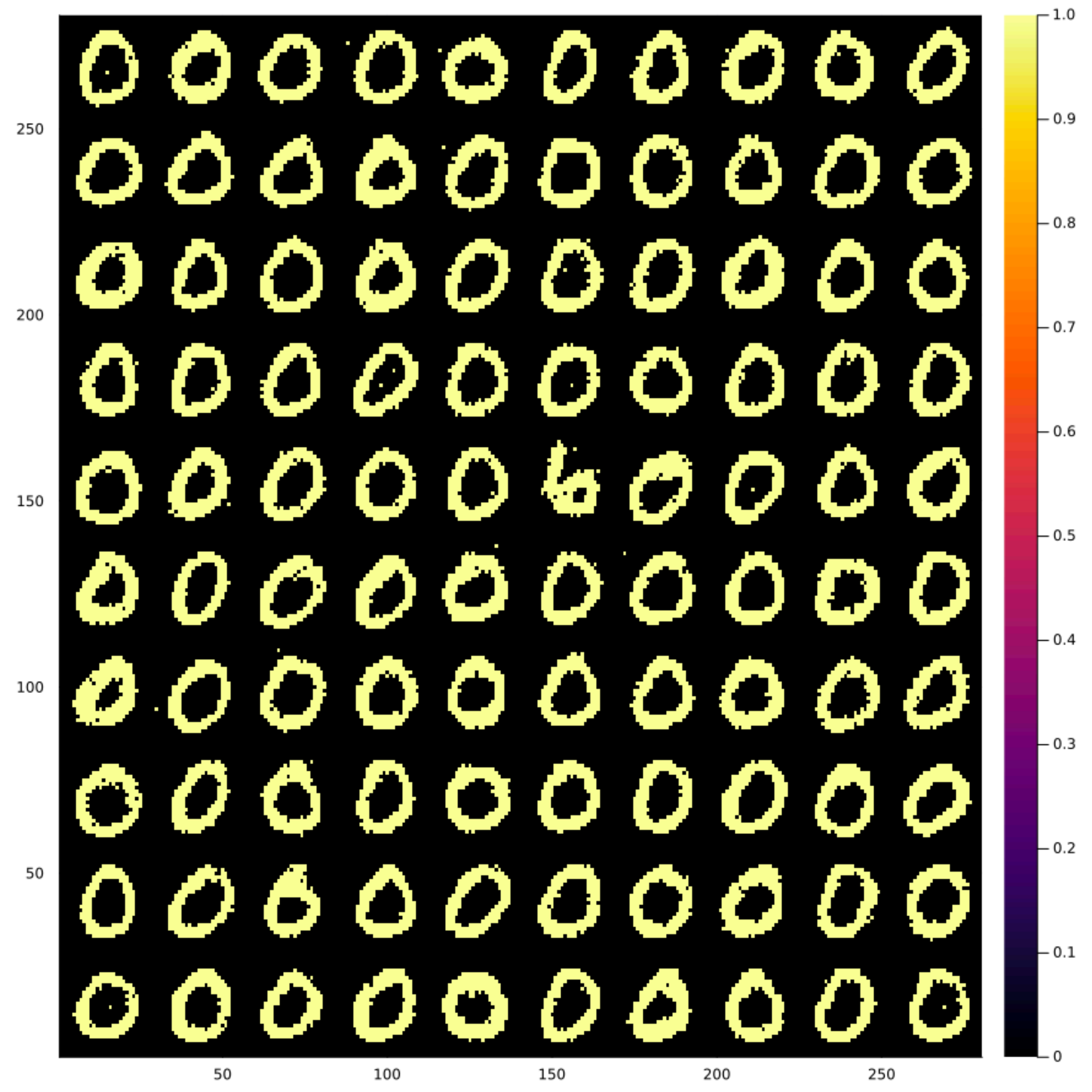


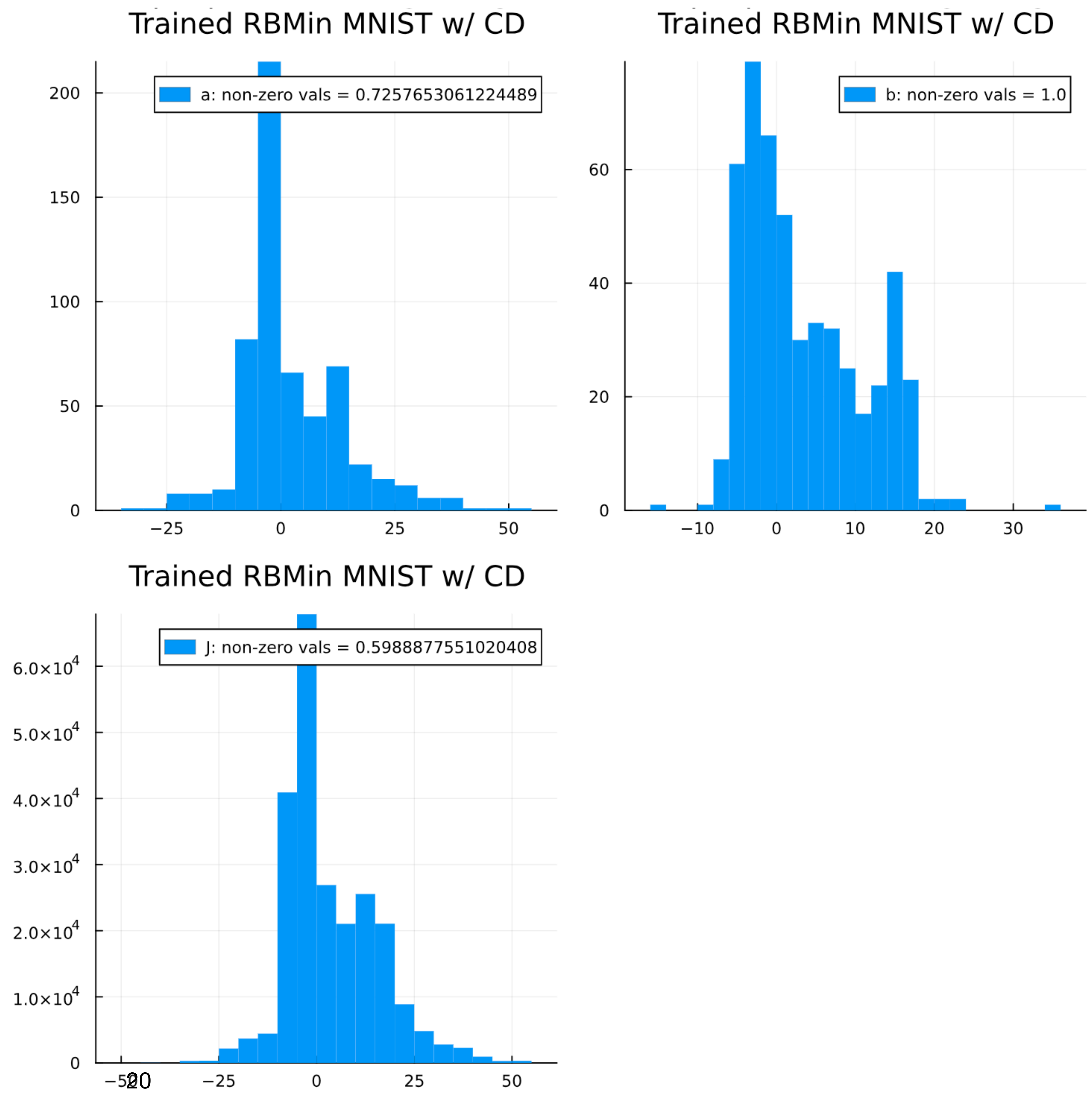




High Temperature gradient approximation

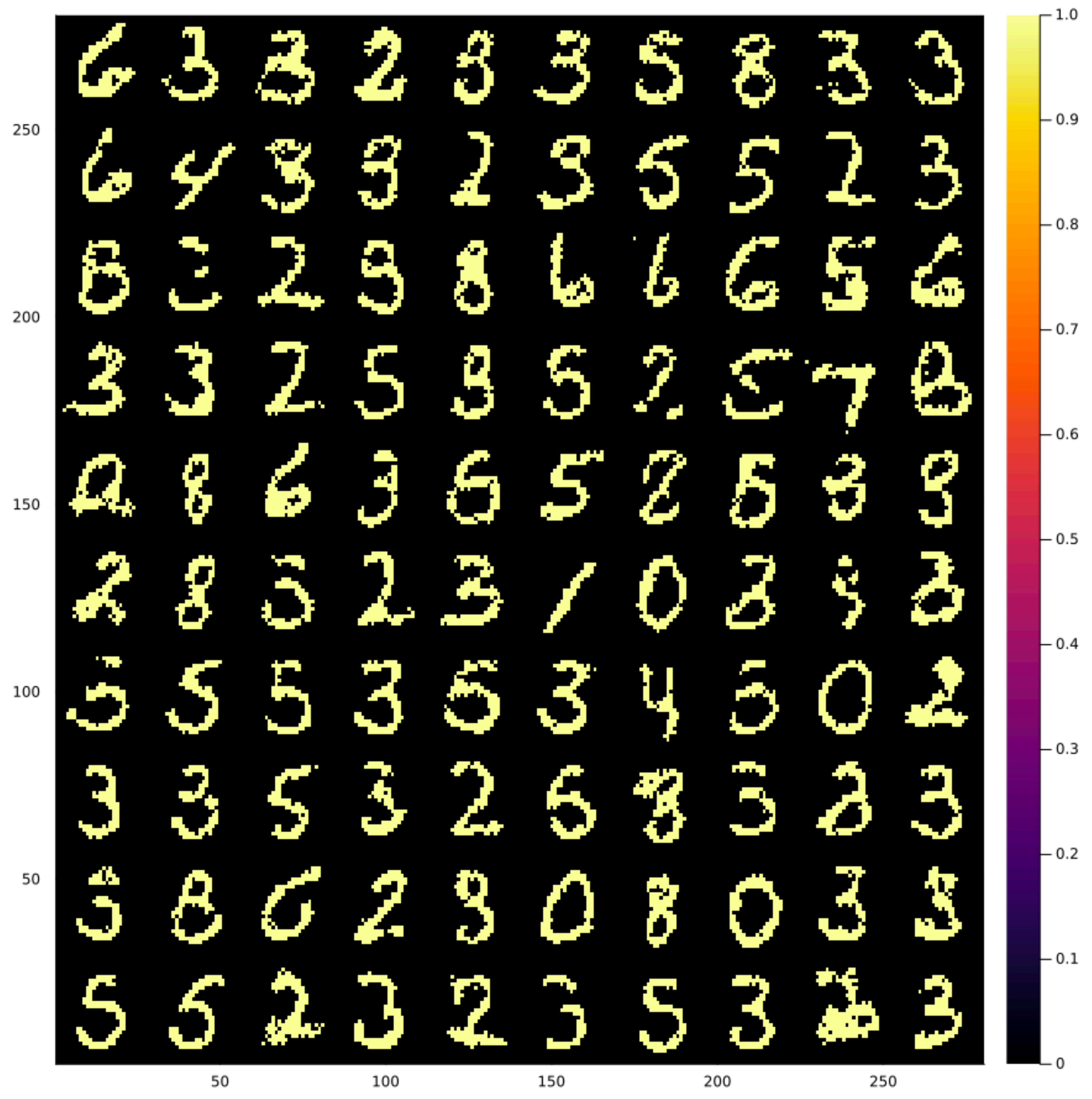
$$\epsilon = \frac{\langle E\frac{\partial E}{\partial \Theta}\rangle - \langle E\rangle\langle\frac{\partial E}{\partial \Theta}\rangle}{\langle\frac{\partial E}{\partial \Theta}\rangle}$$

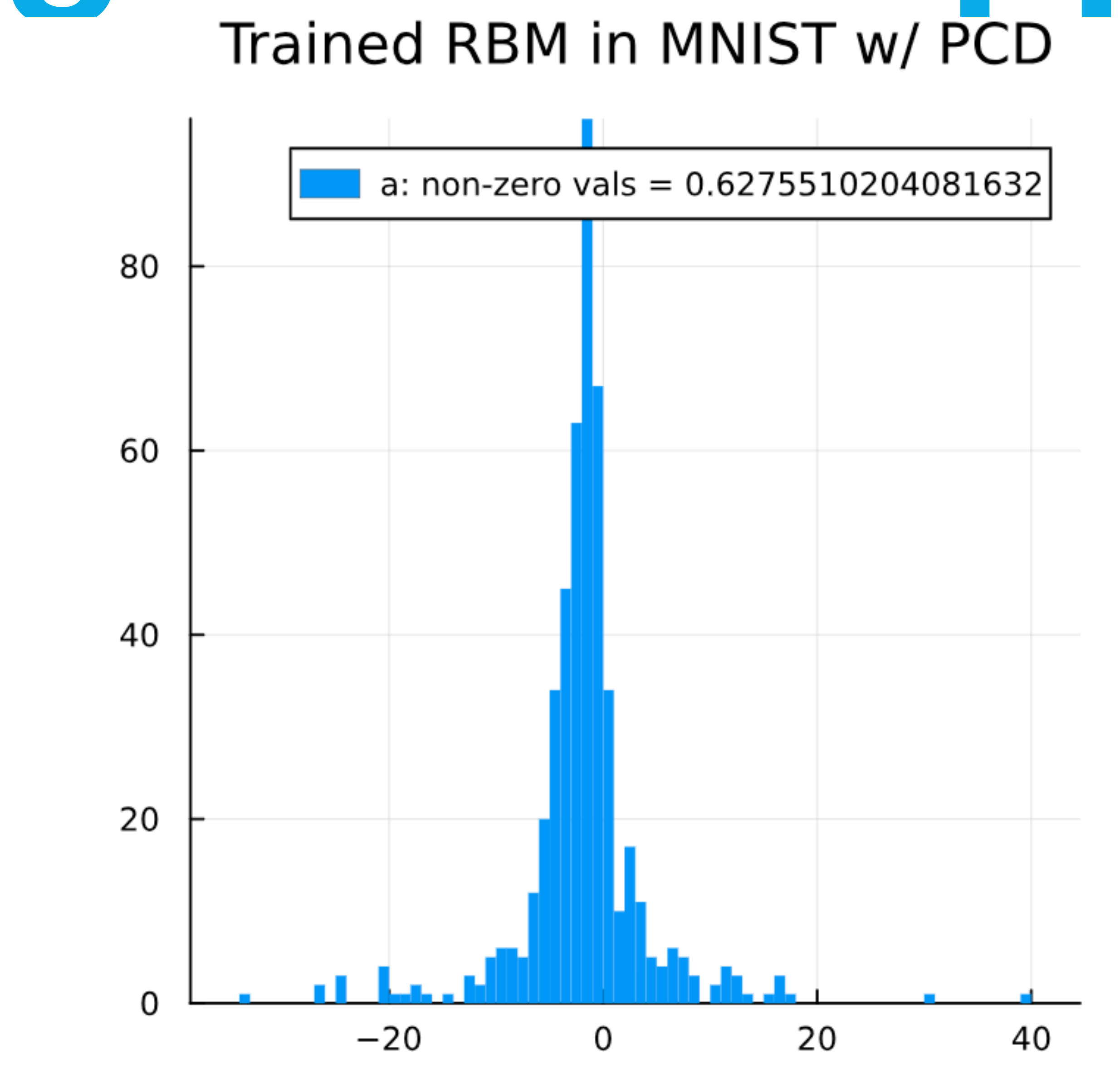


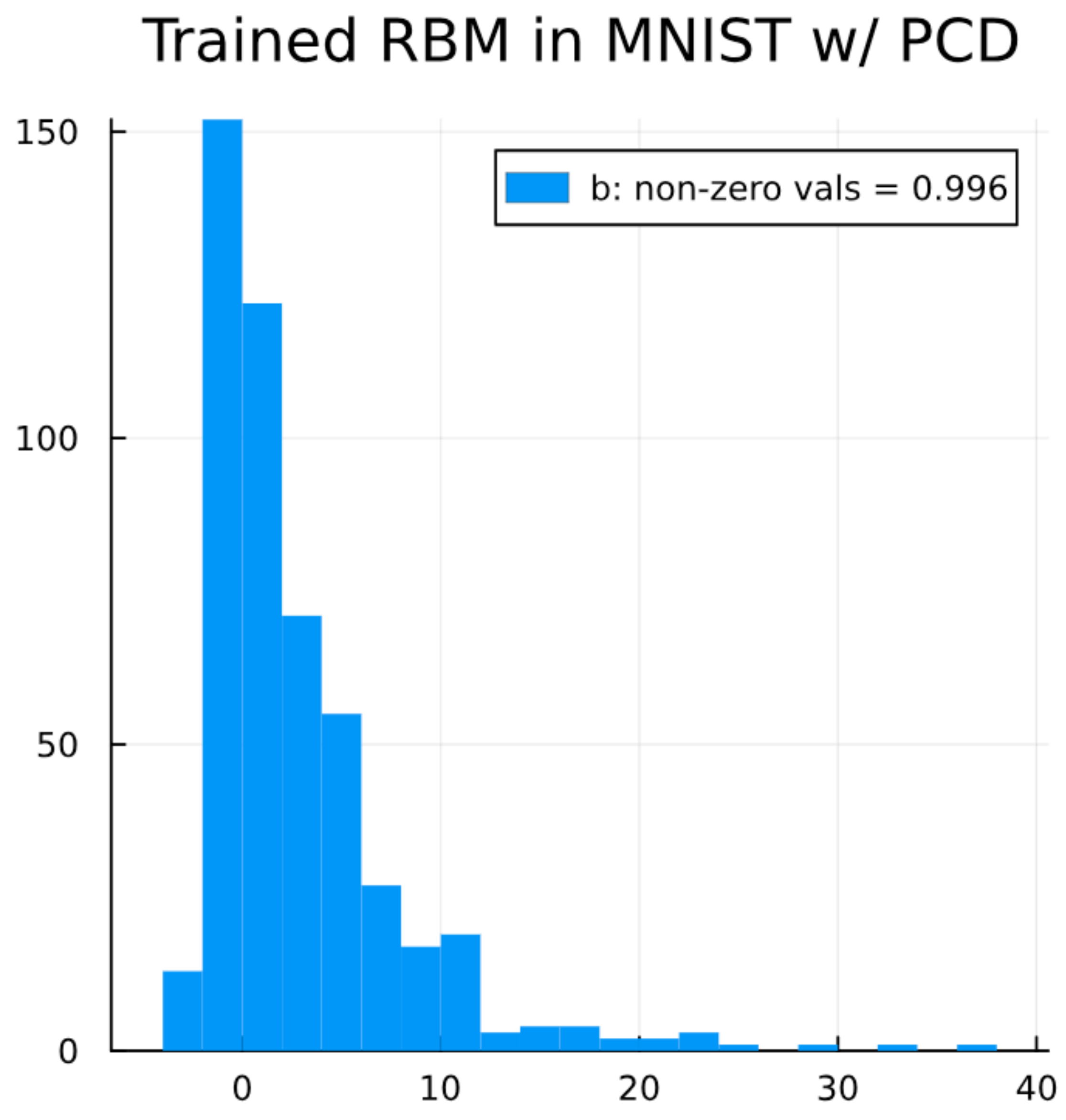


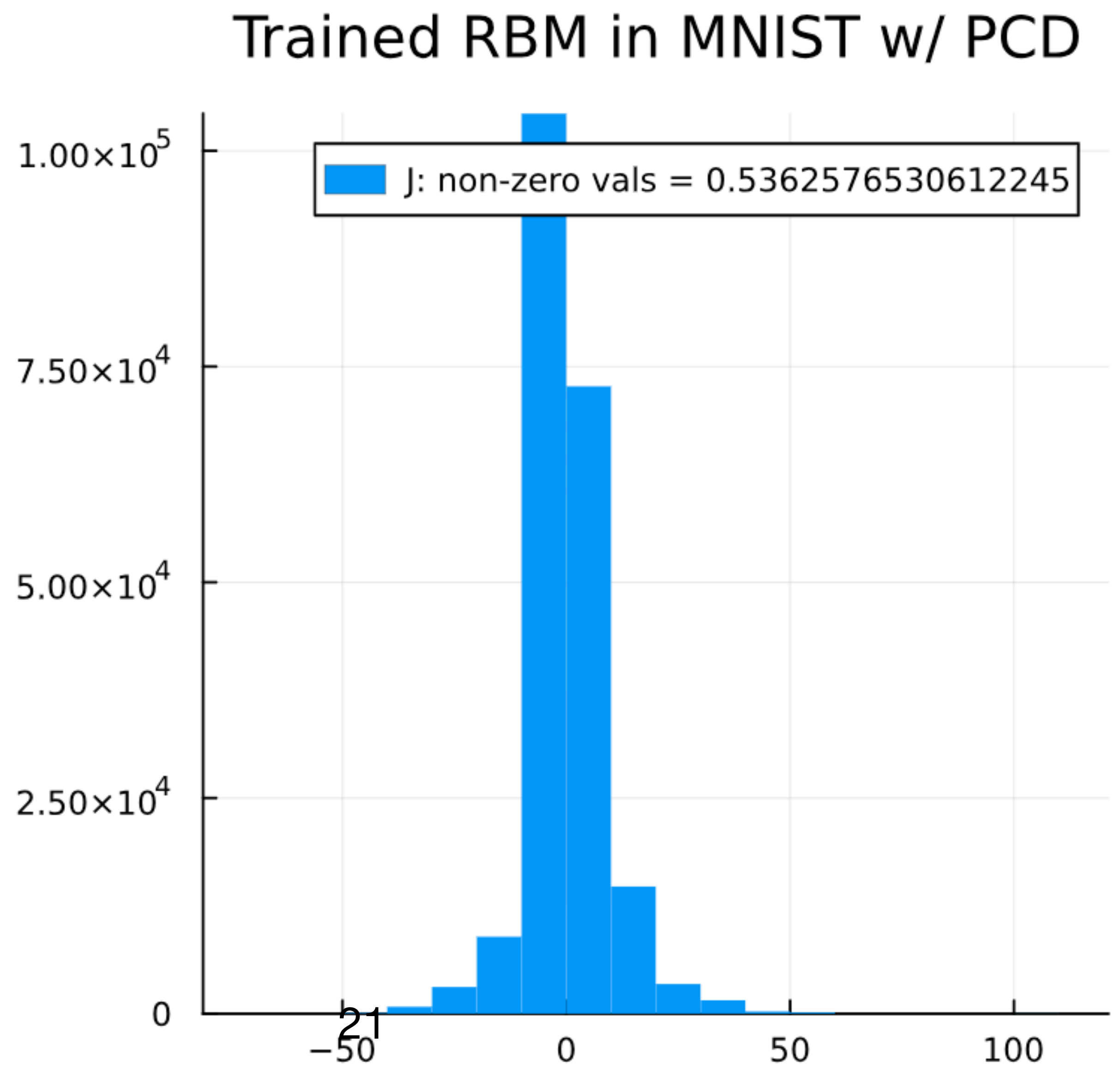
High Temperature gradient approximation

$$\epsilon = \frac{\langle E \frac{\partial E}{\partial \Theta} \rangle - \langle E \rangle \langle \frac{\partial E}{\partial \Theta} \rangle}{\langle \frac{\partial E}{\partial \Theta} \rangle}$$







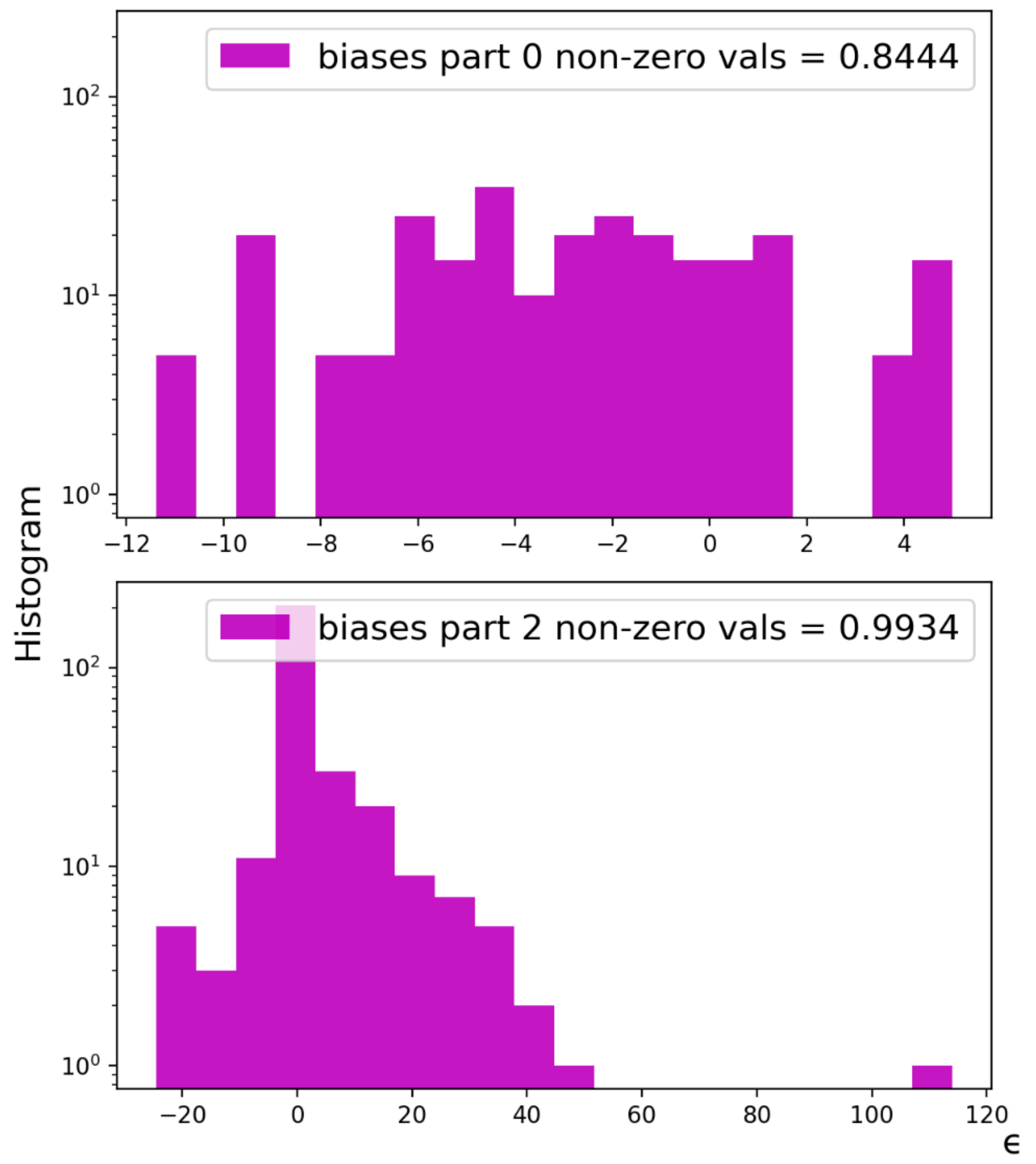


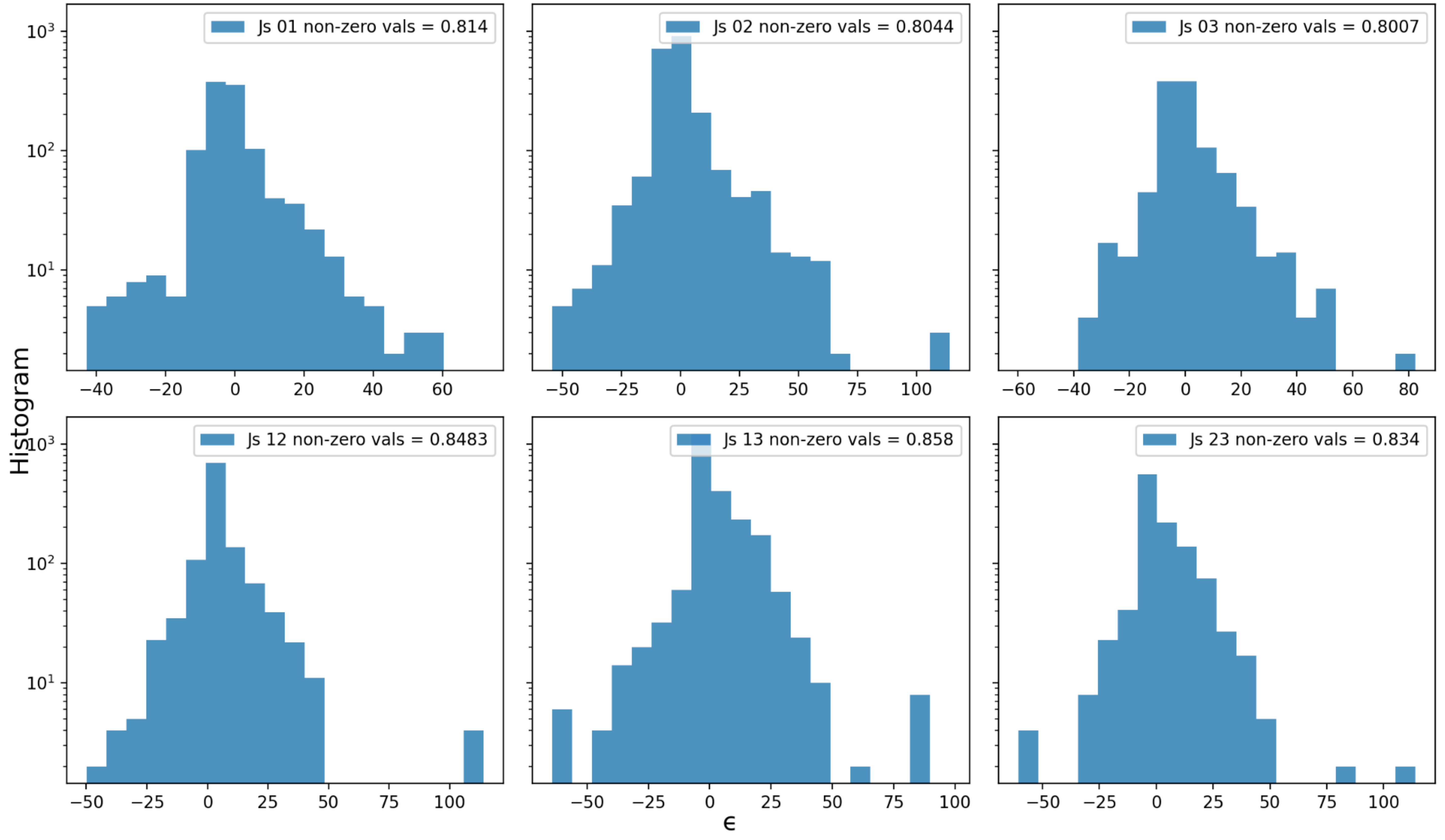
High Temperature gradient approximation

Each point in the histogram corresponds to a parameter being updated *separately* from the rest.

High Temperature gradient approximation

CaloQVAE Model B





High Temperature gradient approximation

Small RBMs

Let's assume an RBM w/ 10 visible and 10 hidden nodes.

$$\epsilon = \frac{\langle E \frac{\partial E}{\partial \Theta} \rangle - \langle E \rangle \langle \frac{\partial E}{\partial \Theta} \rangle}{\langle \frac{\partial E}{\partial \Theta} \rangle}$$

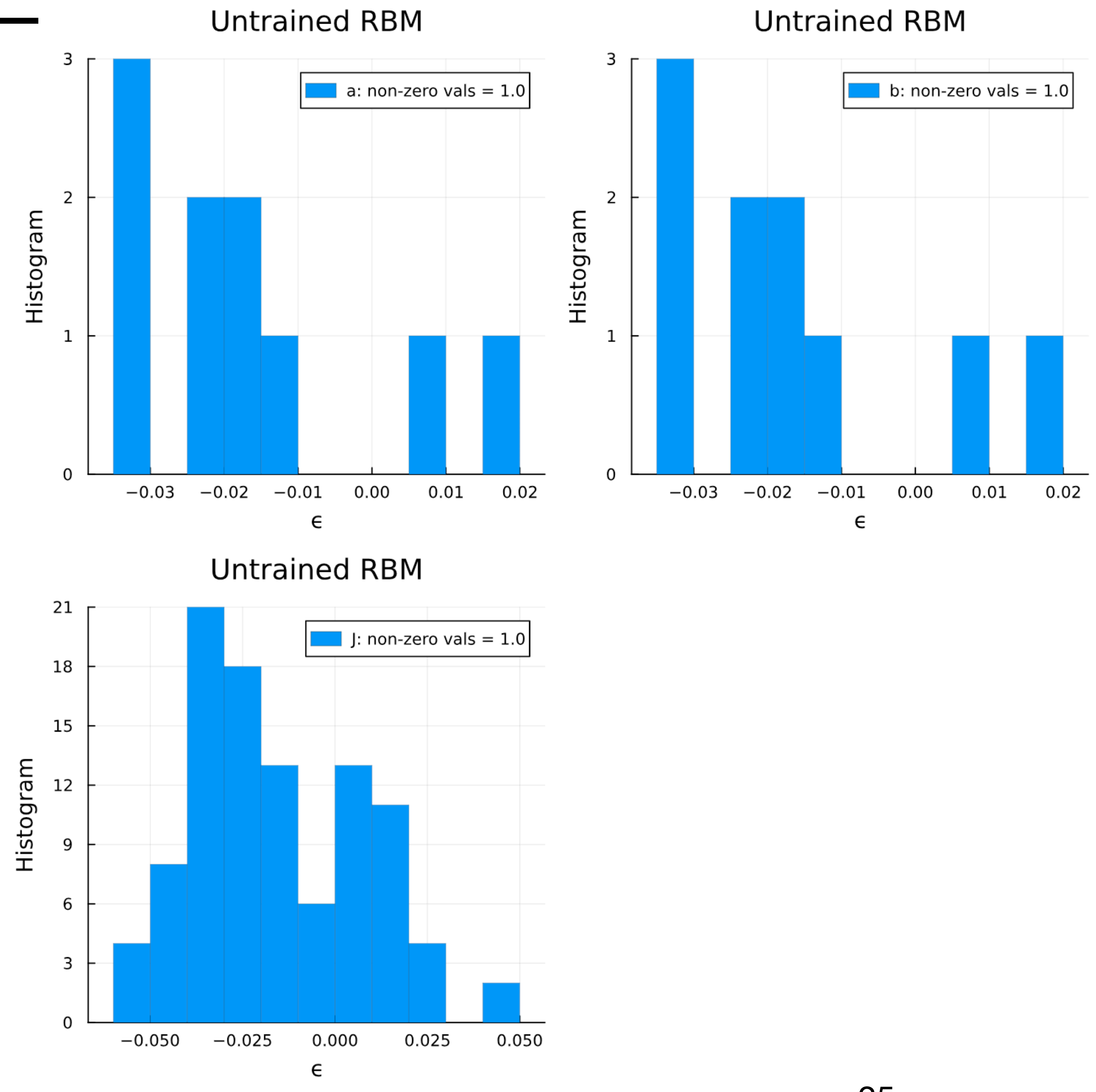
High Temperature gradient approximation

Small RBMs

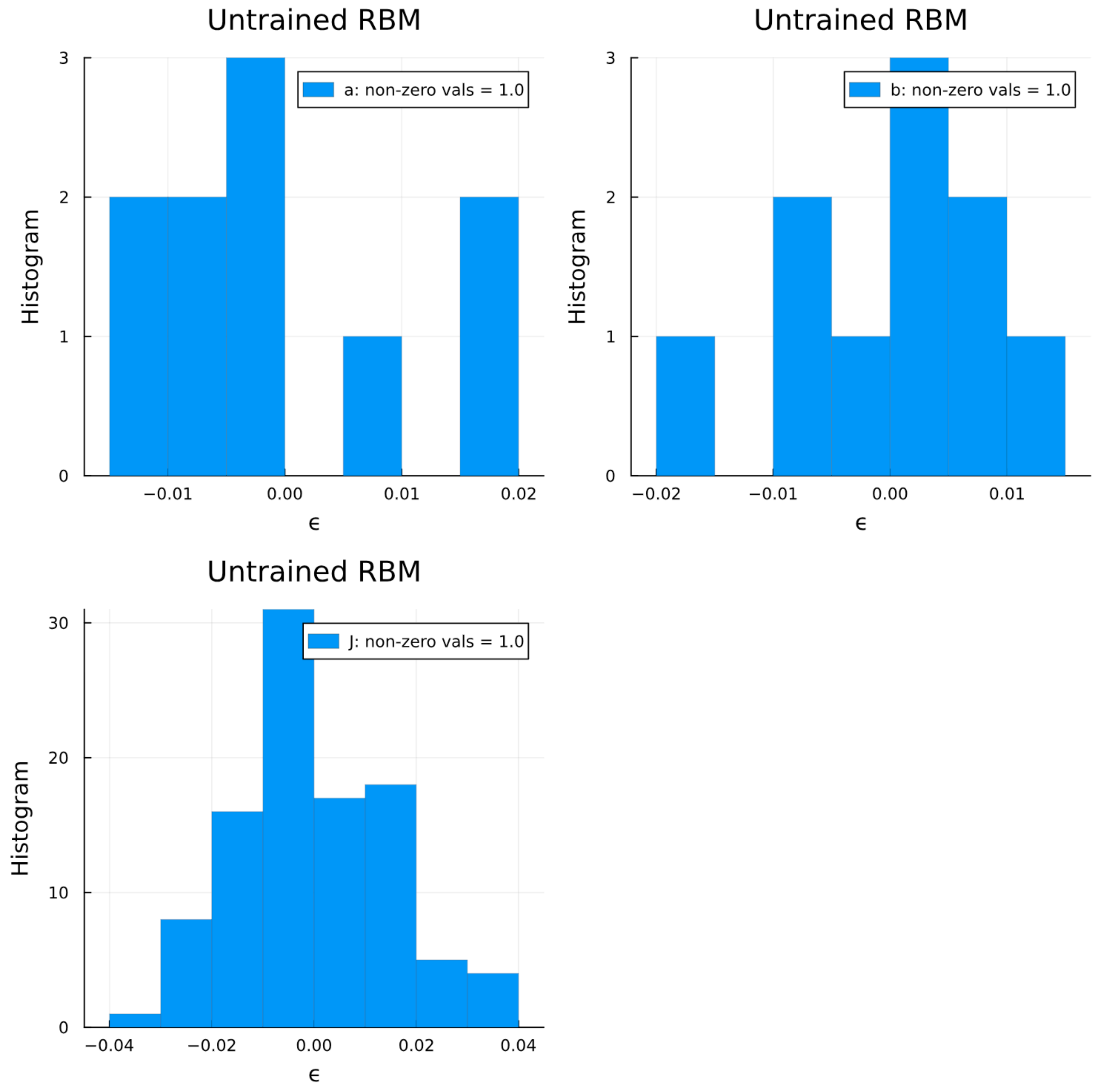
Let’s assume the weights and biases are sampled from a normal N(0,0.1)

$$\epsilon = \frac{\langle E \frac{\partial E}{\partial \Theta} \rangle - \langle E \rangle \langle \frac{\partial E}{\partial \Theta} \rangle}{\langle \frac{\partial E}{\partial \Theta} \rangle}$$

Exact



BGS



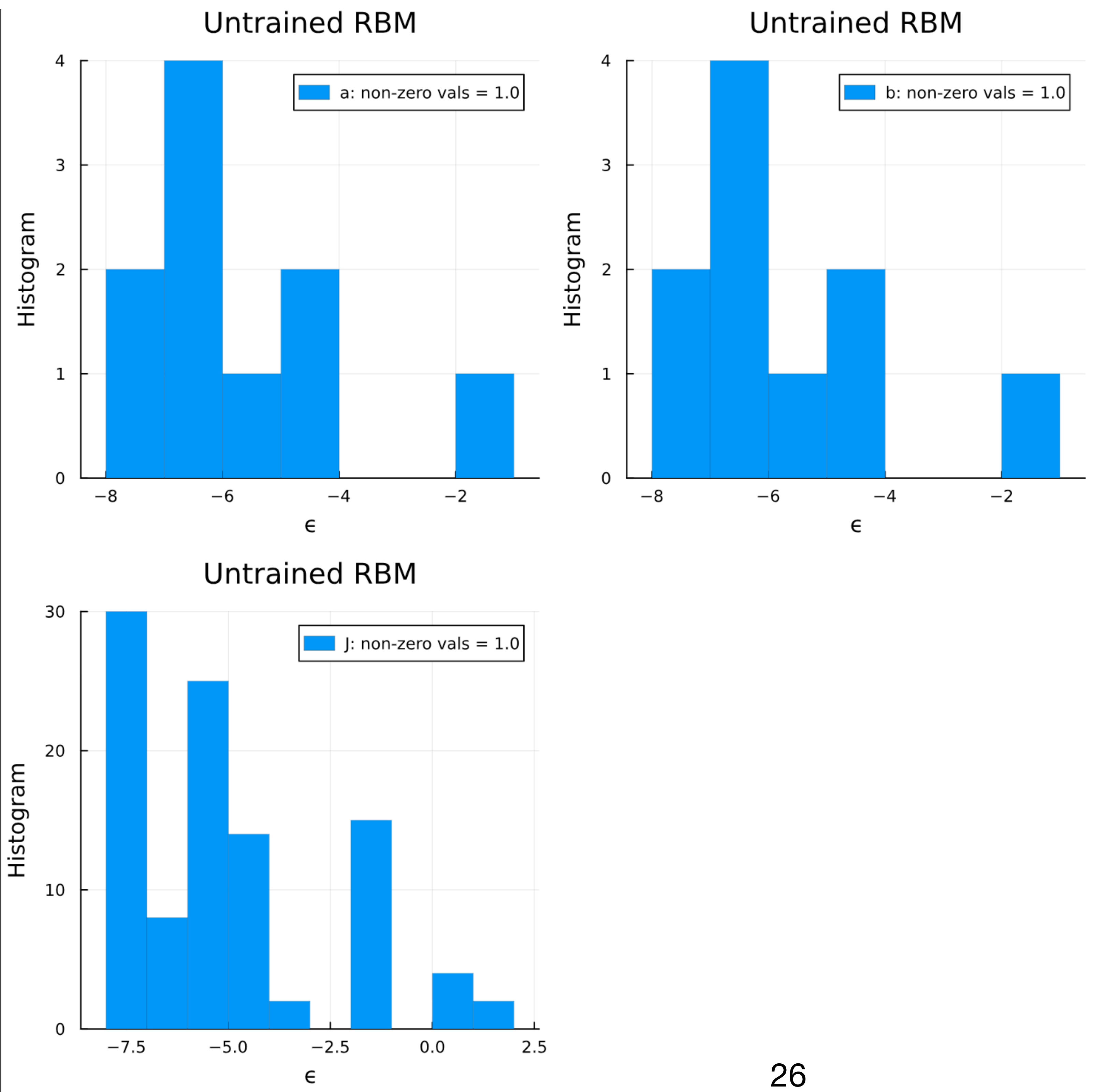
High Temperature gradient approximation

Small RBMs

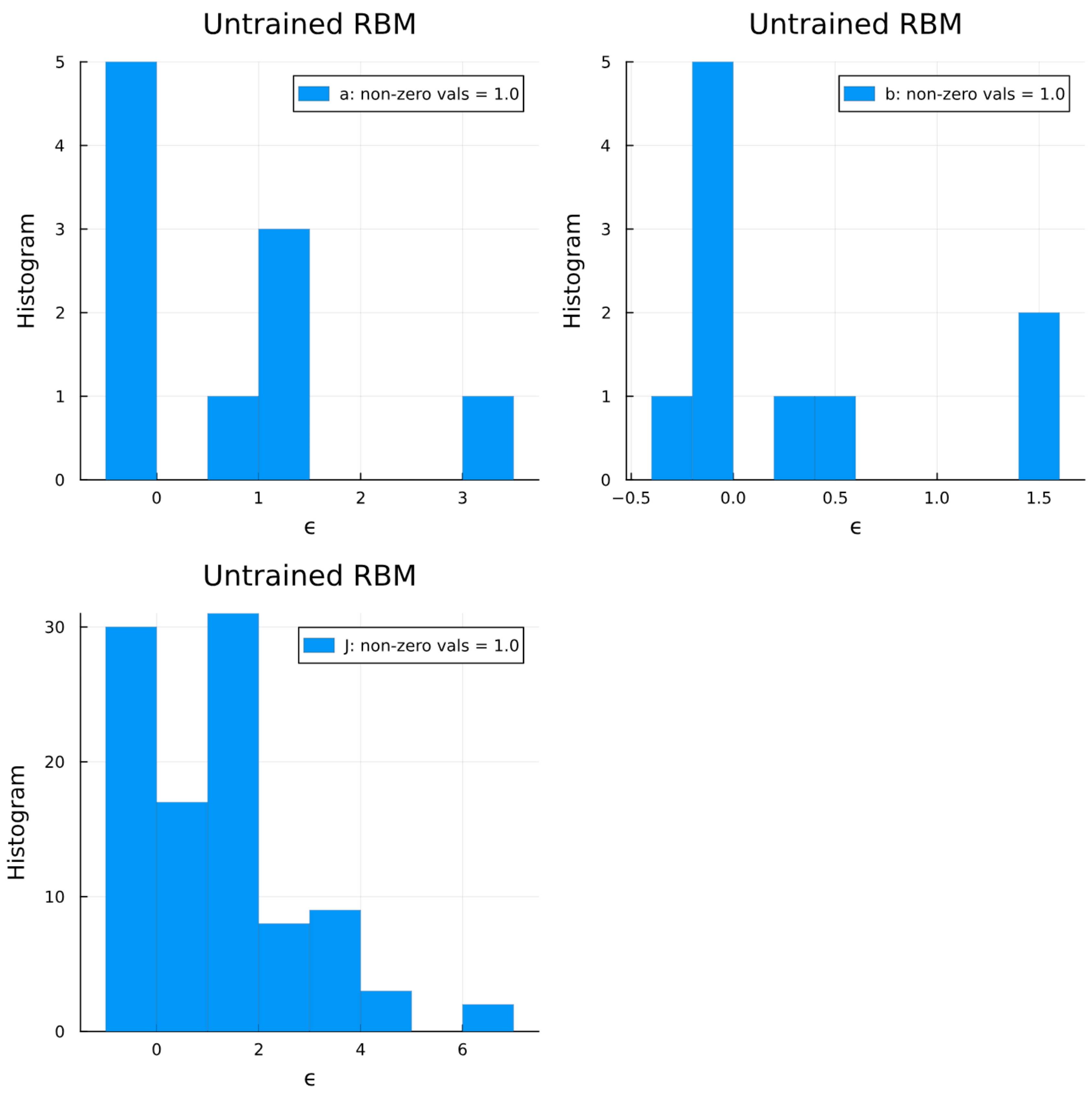
Let's assume the weights and biases are sampled from a Uni(-1,1)

$$\epsilon = \frac{\langle E \frac{\partial E}{\partial \Theta} \rangle - \langle E \rangle \langle \frac{\partial E}{\partial \Theta} \rangle}{\langle \frac{\partial E}{\partial \Theta} \rangle}$$

Exact



BGS



High Temperature gradient approximation

Small RBMs

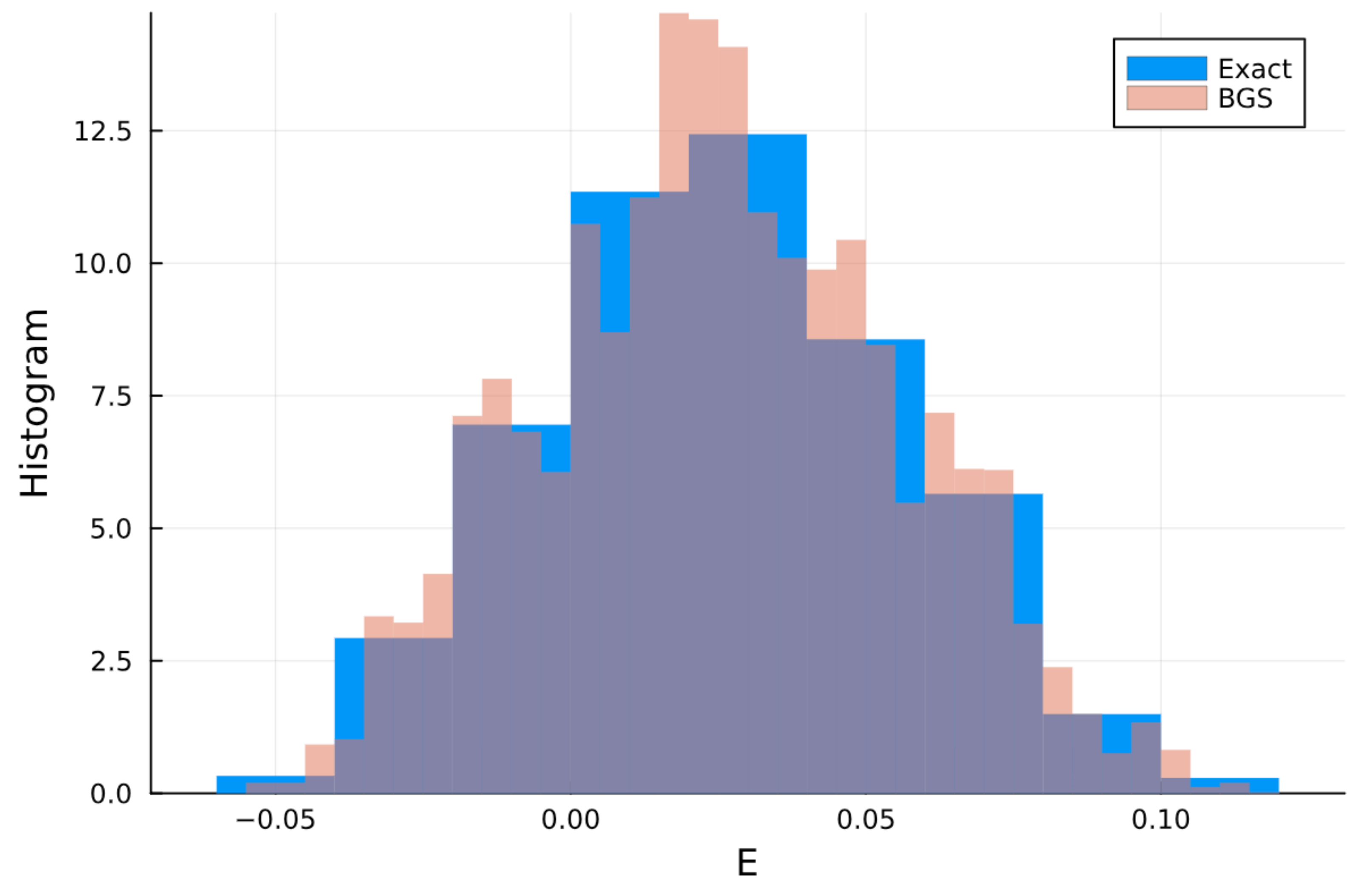
For Uni(-1,1) distributed weights and biases, the range of epsilons do not match between exact and BGS.

High Temperature gradient approximation

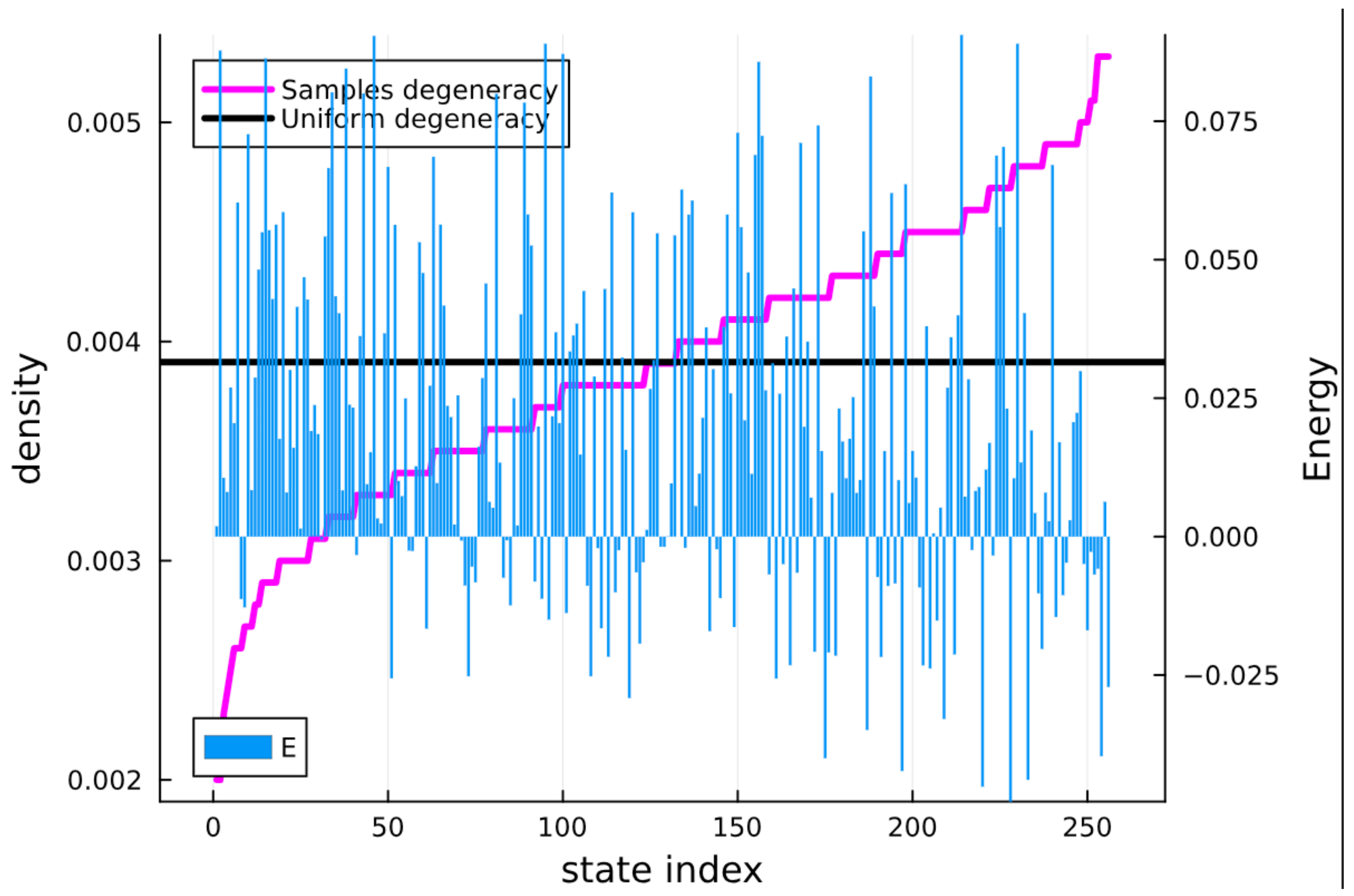
Small RBMs

Let's look at the energy histograms

N(0,1) weights and biases



10k states sampled via BGS. We measure the state degeneracy

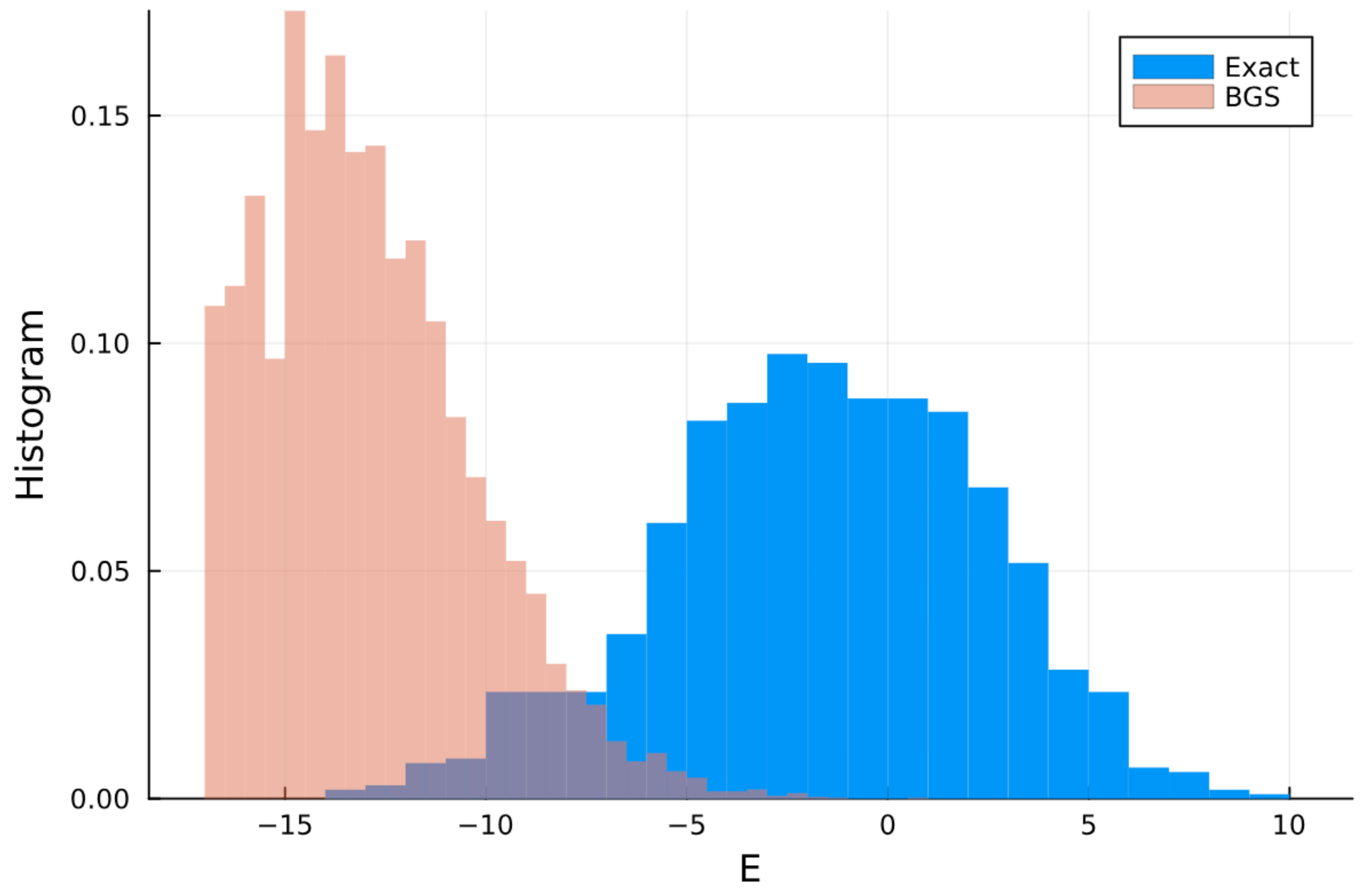


High Temperature gradient approximation

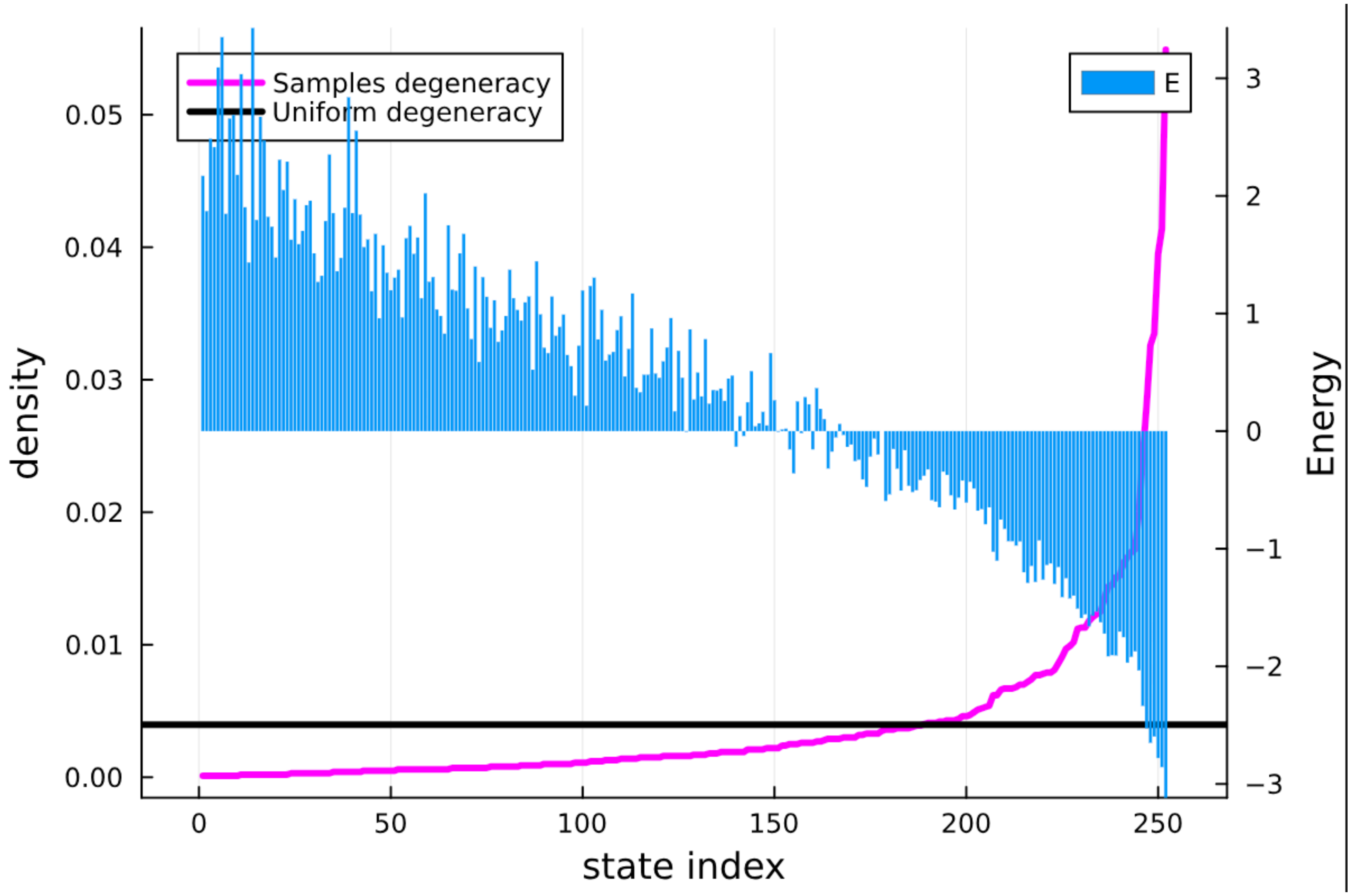
Small RBMs

Let's look at the energy histograms

U(-1,1) weights and biases



10k states sampled via BGS. We measure the state degeneracy



High Temperature gradient approximation

Small RBMs

$$\mathcal{N}(J_0, J_1)$$

$$\tilde{J}_0 = NJ_0 \rightarrow 20 \cdot \delta$$

$$\tilde{J} = N^{1/2}J \rightarrow \sqrt{20} \cdot 1$$

$$kT = 1$$

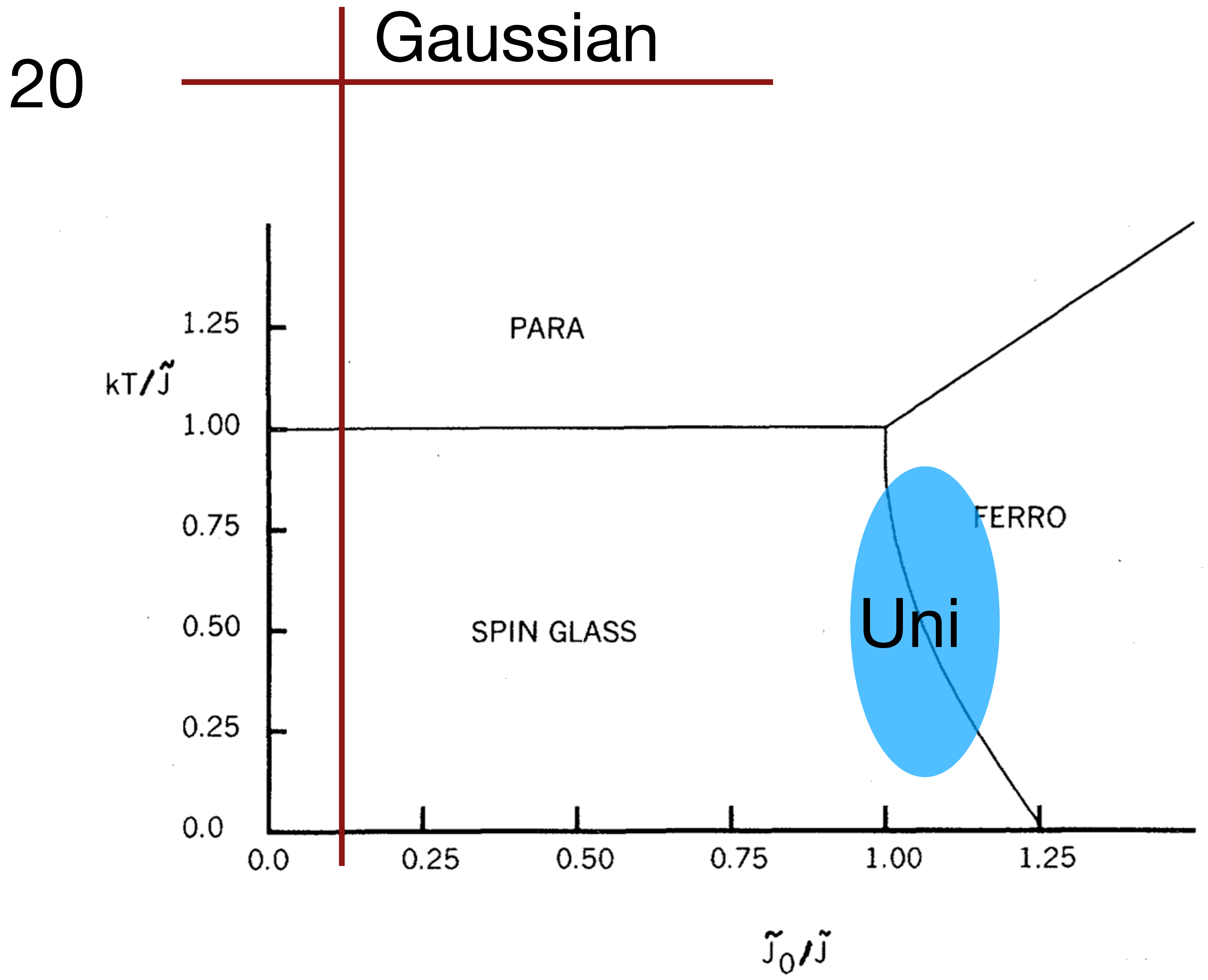


FIG. 1. Phase diagram of spin-glass ferromagnet.

Solvable Model of Spin glass, Kirkpatrick, Sherrington

High Temperature gradient approximation

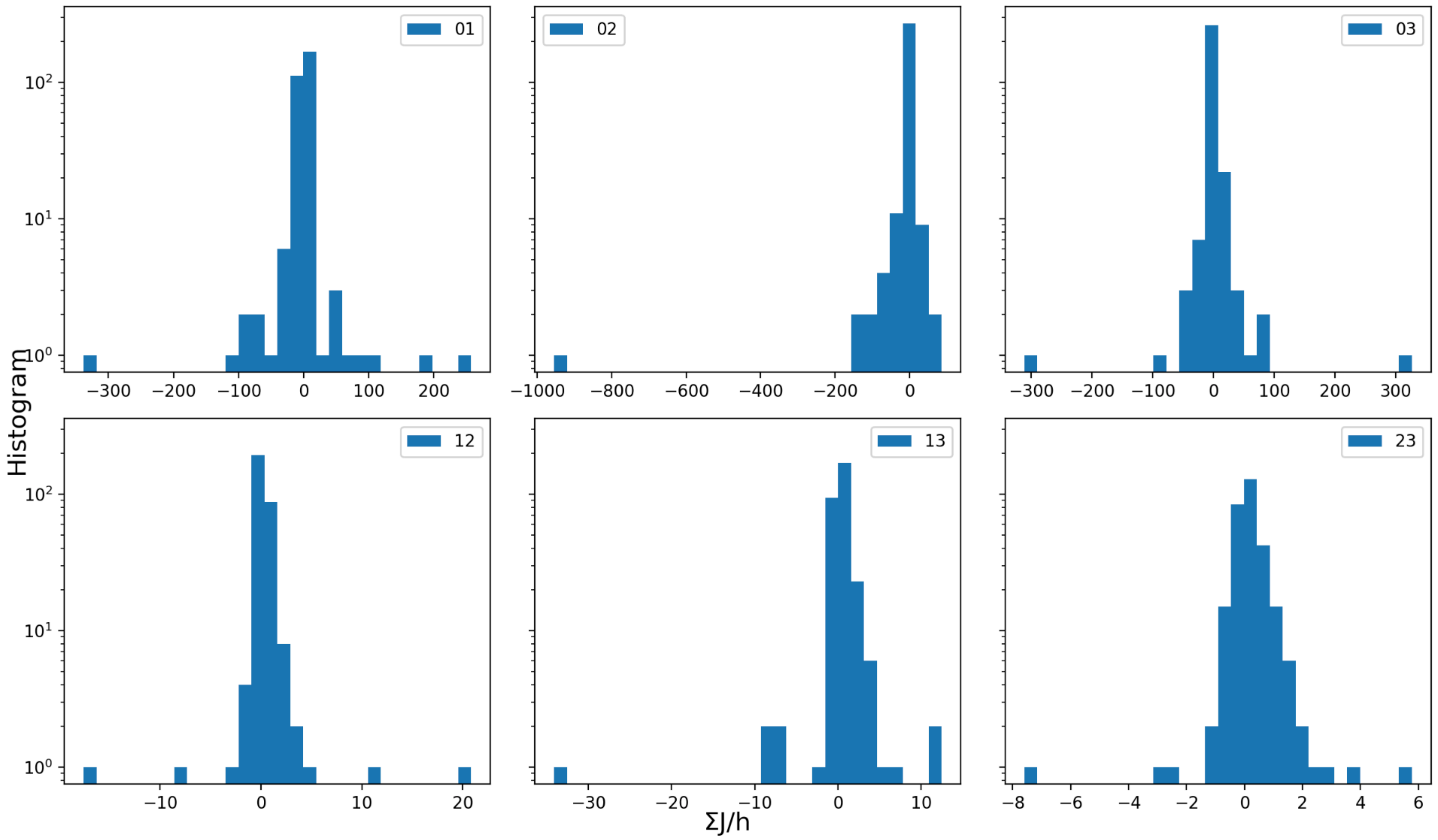
- Next step: Train model w/o this approximation.

Ratio between couplers and fields

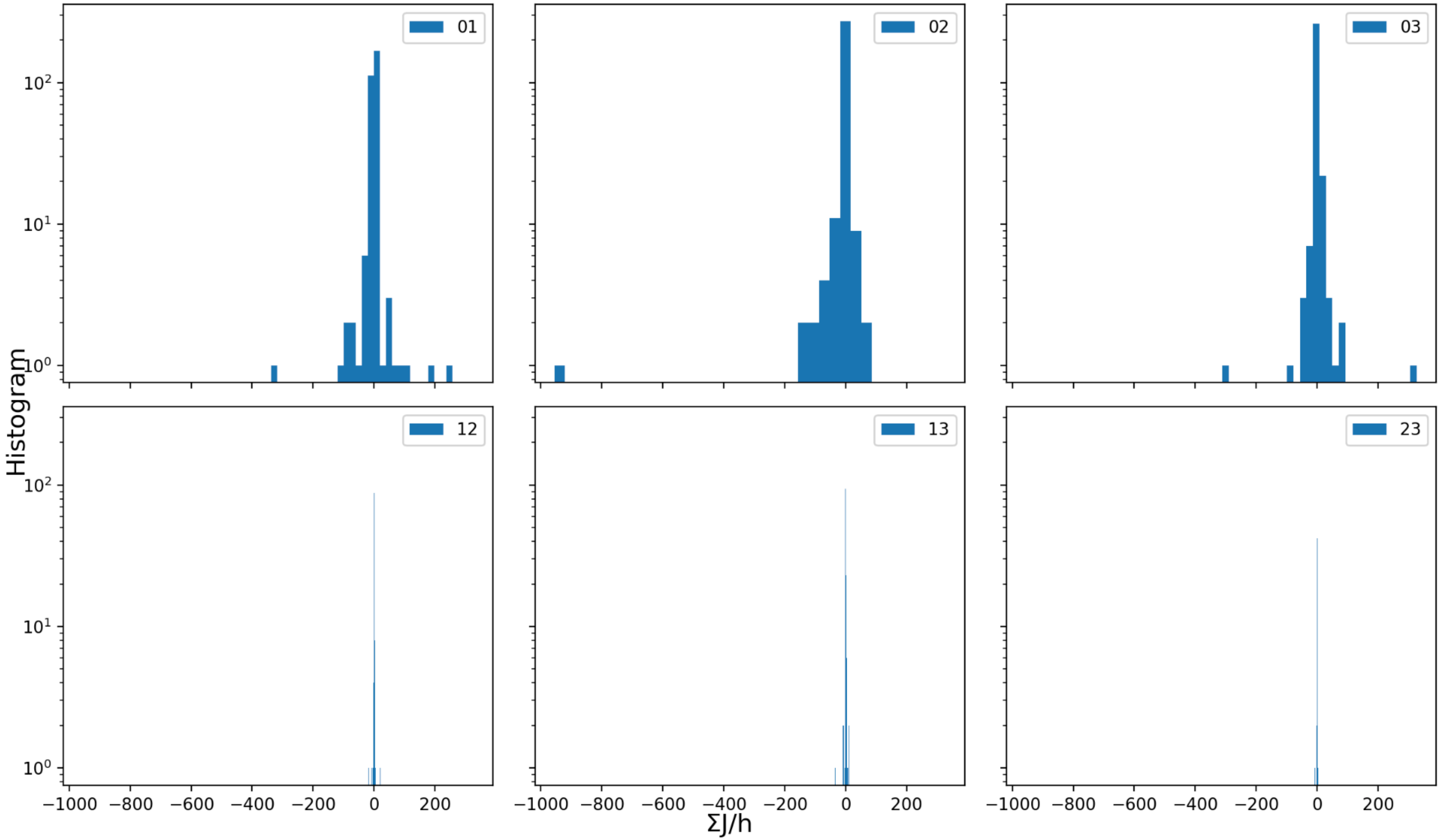
For a given spin i, what is the ratio between the some over its couplers and the self-field?

$$\frac{1}{h_i}\sum_j J_{ij}$$

Ratio between couplers and fields



Ratio between couplers and fields



RBM to Diffusion Model equivalence

$$Z = \sum_{v,h} e^{-\beta E(v,h)}$$

$$E(v,h) = -\langle v | a_0 \rangle - \langle b_0 | h \rangle - \langle v | W | h \rangle$$

$W = U \Sigma V^t$ (SVD)

$$|x\rangle = U|v\rangle, \qquad |y\rangle = V|h\rangle$$

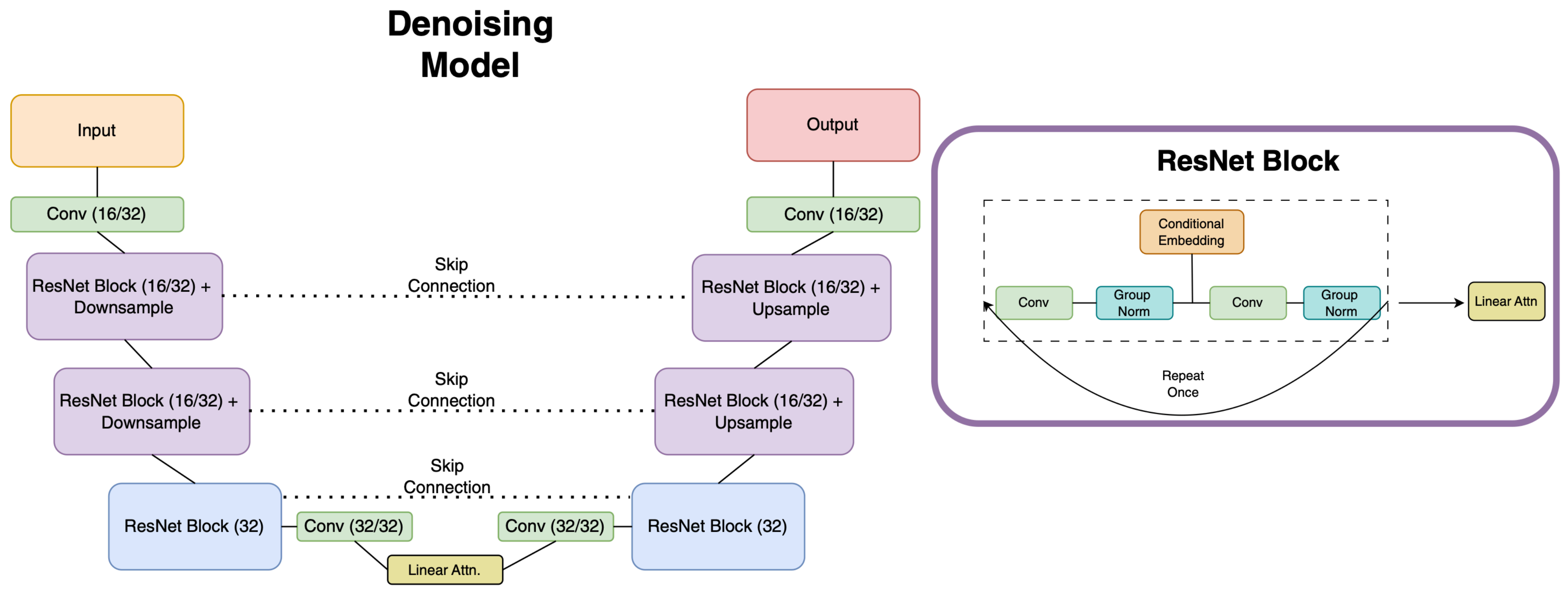
$$Z = \sum_{x,y} e^{-\beta E(x,y)}$$

Diffusion models

How good are they?

CaloDiffusion

Kevin Pedro, Oz Amram



CaloLatentDiffusion

Madula, Mikuni

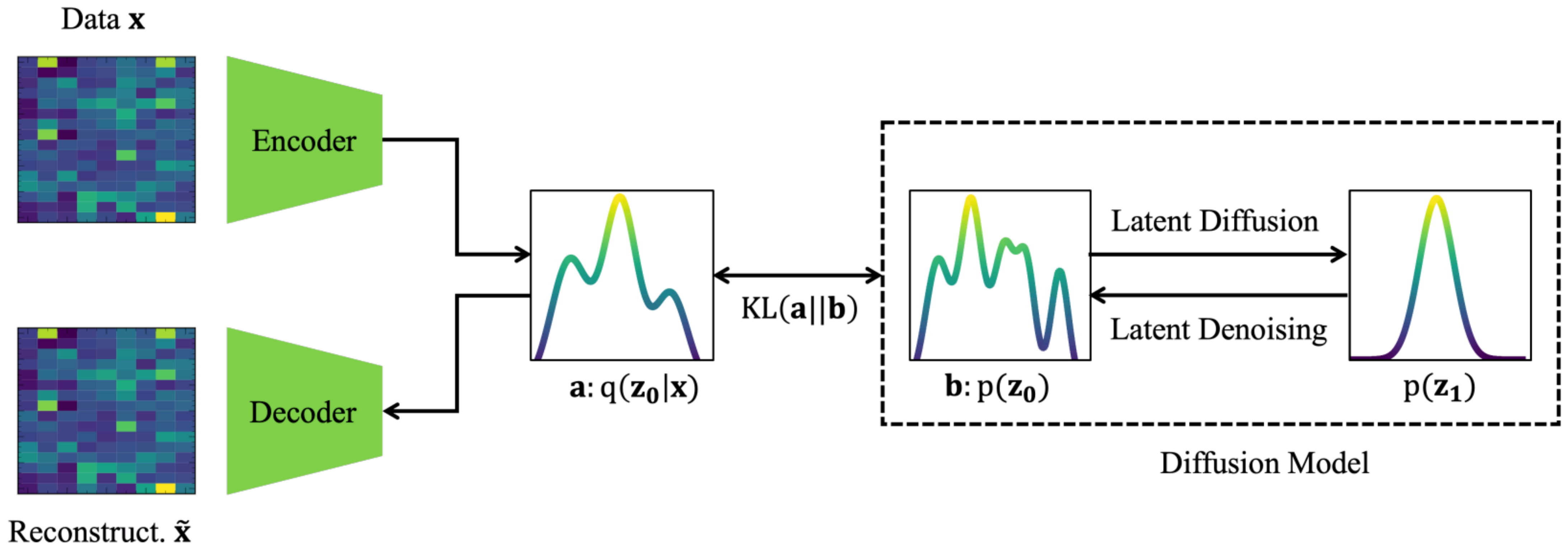
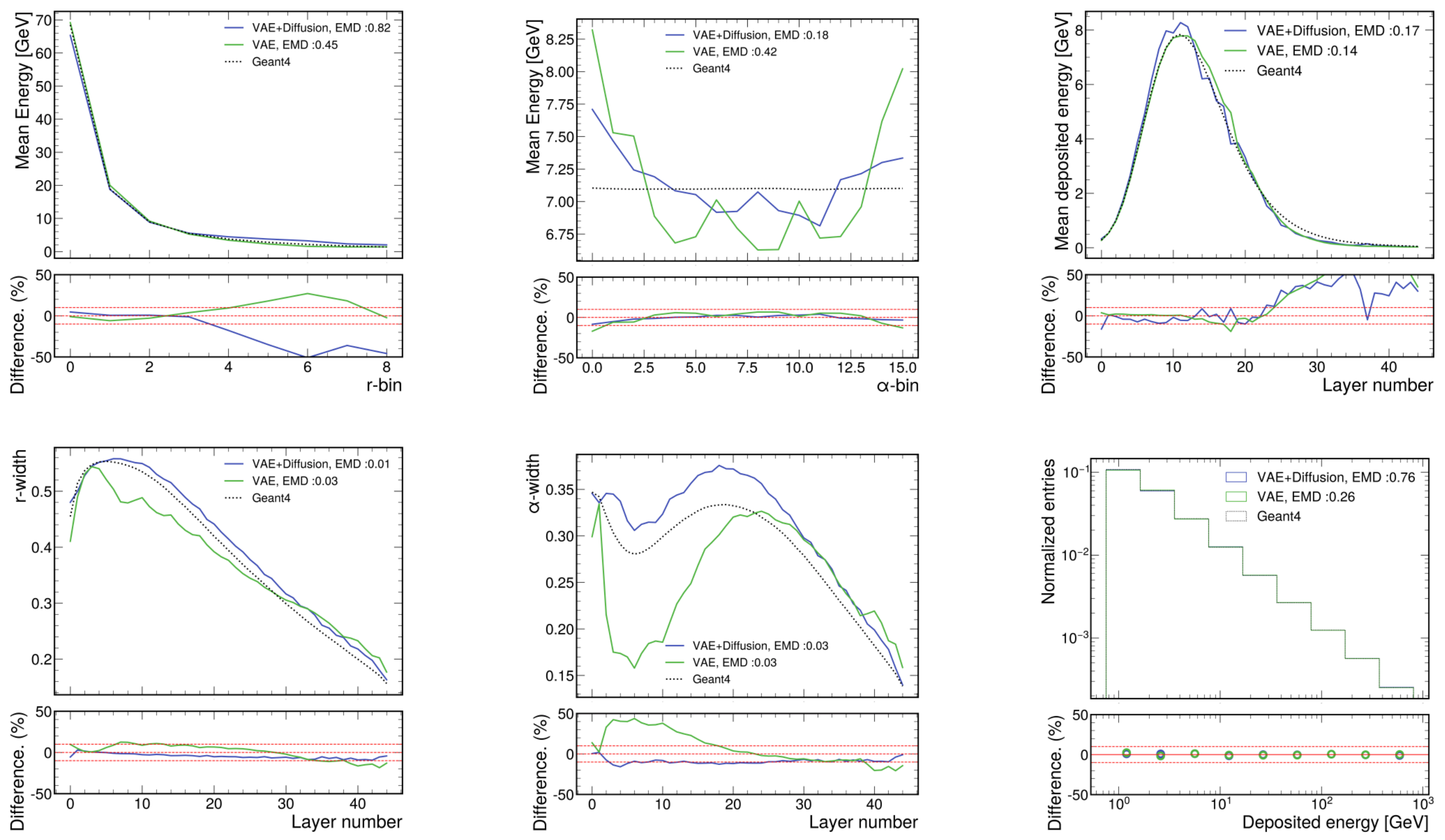


Figure 1: A schematic diagram of latent diffusion



Summary

✓ Papers status

✓ Using the QPU w/ Pegasus and in Zephyr

  ➡ Meet with dwave to discuss strategy?

✓ High temperature gradient approximation for trained RBMs

  ➡ Train model w/o this approximation

✓ Weight decay removal?

- RBM and Diffusion model equivalence.
- Relaxation time in RBMs

ToDo

✓ UNet for CaloQVAE — Ian

✓ CaloQVAE w/ linear attention layers

✓ Train current model with large RBM in Pegasus

✓ From RBM to diffusion models. Equivalence.

✓ Relaxation times in RBMs

✓ Training model with average of the gradient as opposed to gradient of the average in RBM

✓ Associative mem in GAN — Coherent samples from Zephyr