# Training Quantum Boltzmann Machines to learn physical Gibbs states

Enrico Rinaldi[1], Yuta Kikuchi[1], Matthias Rosenkranz[1], Ryuji Sakata[2]
1 Quantinuum Ltd. 2 Panasonic Holdings Corporation

## 1. Abstract

Gibbs states play central roles in understanding the equilibrium properties of quantum many-body systems, and also find applications in optimization and machine learning, where they are known as quantum Boltzmann machines (QBMs). The fact that the model Hamiltonian may contain non-commuting terms can make QBM more expressive than the classical Boltzmann machine, and in stark contrast to many other quantum machine learning models, training the QBM with an appropriate choice of objective function does not suffer from the vanishing gradient problem. In this work we numerically investigate the performance of QBMs for different target Gibbs states, model Hamiltonians, training parameters, and training strategies, used as a quantum generative model. We have developed a software package that performs the QBM training with various combinations of target, model, and parameters using exact diagonalization on classical computers.

## 2. Methods[1]

Quantum Boltzmann Machine

$$\rho_\theta = \frac{e^{H_\theta}}{Z}, \qquad Z = \mathrm{Tr}[e^{H_\theta}], \qquad H_\theta = \sum_{i=1}^{m} \theta_i h_i$$

Objective: Quantum Relative Entropy

$$S(\eta \parallel \rho_\theta) = \mathrm{Tr}[\eta \log \eta] - \mathrm{Tr}[\eta \log \rho_\theta]$$

Optimize model parameters by gradient descent

$$\frac{\partial S(\eta \parallel \rho_\theta)}{\partial \theta_i} = \langle H_i \rangle_{\rho_\theta} - \langle H_i \rangle_\eta + \sigma\varepsilon, \qquad \sigma\text{: intensity of shot noise}, \ \varepsilon \sim N(0,1)$$

## 3. Model Hamiltonians

| Name | Hamiltonian | # parameters (8 qubits) |
|---|---|---|
| 1D Transverse-field Ising model (TFI) | $H = J\sum_{i=1}^{n-1} Z_i Z_{i+1} + B\sum_{i=1}^{n} X_i$ | 2 |
| 1D Hubbard model (Hubbard) | $H = -t\sum_{j=1}^{\frac{n}{2}-1}\sum_{\sigma\in\{\uparrow,\downarrow\}}(a_{j,\sigma}^\dagger a_{j+1,\sigma} + a_{j+1,\sigma}^\dagger a_{j,\sigma}) + u\sum_{j=1}^{\frac{n}{2}}\sum_{\sigma\in\{\uparrow,\downarrow\}}\left(a_{j,\uparrow}^\dagger a_{j,\uparrow} - \frac{1}{2}\right)\left(a_{j,\downarrow}^\dagger a_{j,\downarrow} - \frac{1}{2}\right)$, <br> $a_{j,\uparrow} = \frac{X_{2j-1}+iY_{2j-1}}{2} Z_{2j-2}\cdots Z_1, a_{j,\downarrow} = \frac{X_{2j}+iY_{2j}}{2} Z_{2j-1}\cdots Z_1$ (Jordan-Wigner transformation) | 2 |
| 1D Geometrically local model (GL) | $H = \sum_{i=1}^{n-1}(J_{i,i+1}^X X_i X_{i+1} + J_{i,i+1}^Y Y_i Y_{i+1} + J_{i,i+1}^Z Z_i Z_{i+1}) + \sum_{i=1}^{n}(B_i^X X_i + B_i^Y Y_i + B_i^Z Z_i)$ | 45 |
| Fully connected model (FC) | $H = \sum_{i>j}(J_{i,j}^X X_i X_j + J_{i,j}^Y Y_i Y_j + J_{i,j}^Z Z_i Z_j) + \sum_{i=1}^{n}(B_i^X X_i + B_i^Y Y_i + B_i^Z Z_i)$ | 108 |

## 4. Results

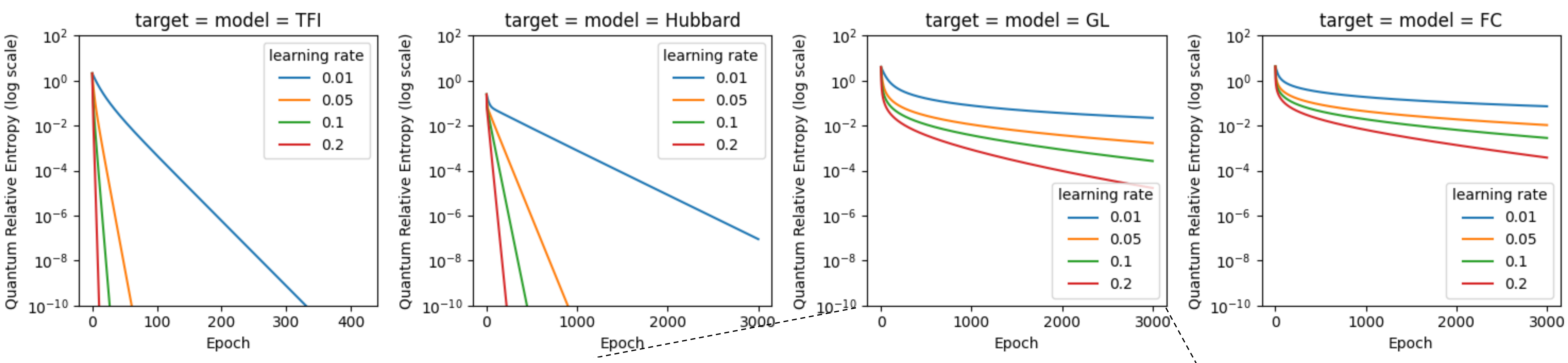


Fig. 1. Difference in learning curves with Hamiltonian in the absence of noise

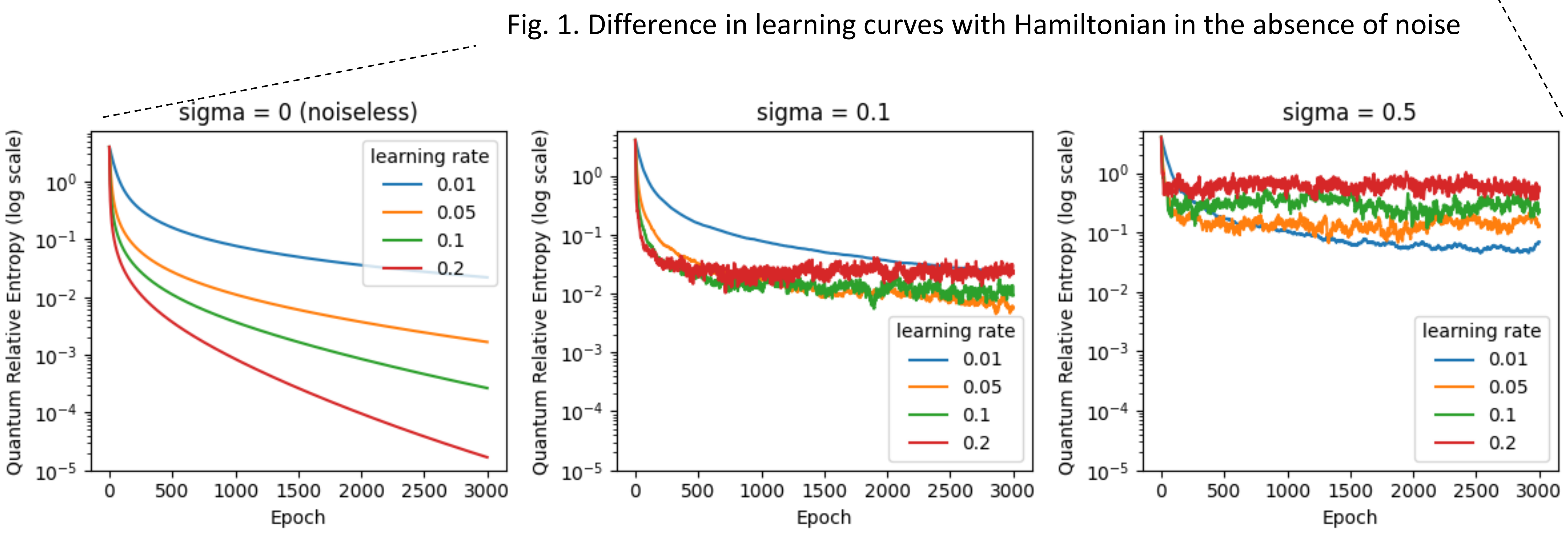


Fig. 2. Difference in learning curves due to noise when the Hamiltonian is GL

- Fig. 1 shows that as the number of parameters to be learned increases, convergence takes longer.
- Fig. 2 shows that under ideal conditions, a high learning rate can accelerate convergence, but in the presence of noise, a high learning rate can actually degrade learning performance.

**In noisy situations, the learning rate should be adjusted according to the intensity of noise. This is one evidence of the theoretical results in Ref [2].**

The above result is just one example. With our repository, you can perform various experiments and gain knowledge about QBMs!
https://github.com/CQCL/qbm_benchmark_dataset

## 5. References

[1] H. J. Kappen, Journal of Physics A: Mathematical and Theoretical 53, 214001 (2020).
[2] L. Coopmans and M. Benedetti, On the Sample Complexity of Quantum Boltzmann Machine Learning (2023), arXiv:2306.14969 [quant-ph].