Learning to generate high-dimensional distributions with low-dimensional quantum Boltzmann machines

Cenk Tüysüz, Maria Demidik, Luuk Coopmans, Enrico Rinaldi
Vincent Croft, Yacine Haddad, Matthias Rosenkranz and Karl Jansen

arXiv: 2410.16363

@cenk_tuysuz

ctuysuz@mail.desy.de

DESY.

Quantum generative modelling

Challenge:

Almost all quantum machine learning (QML) algorithms suffer from trainability issues!

- Q. circuit Born machines (QCBM)
- Q. generative adversarial networks (QGAN)
- Quantum Boltzmann machines (QBM)
- …

- Recently, *fully-visible* QBMs (fv-QBM) have been shown to be sample-efficiently trainable.

L. Coopmans, et al., *Commun Phys* 7, 274 (2024)

- However, fv-QBMs are not as expressive as generic QBMs or possibly RBMs.

➔ Can fv-QBMs solve practically relevant tasks?
➔ Can we make them more expressive, while still being trainable?

Fully visible (quantum) Boltzmann machines

Gibbs state: $$\rho = \frac{e^{-\beta H}}{\mathrm{Tr}(e^{-\beta H})}$$

Classical embedding → $$\eta = \mathrm{diag}(p(s))$$

Quantum embedding → $$\eta = |\psi\rangle\langle\psi|,\ |\psi\rangle = \sqrt{p(s)}e^{i\alpha(s)}|s\rangle$$

$$H = \sum_{k\in\mathcal{P}_1}\sum_{i\in\mathcal{V}}\theta_i^k\sigma_i^k + \sum_{(k,l)\in\mathcal{P}_2}\sum_{(i,j)\in\mathcal{E}}\theta_{i,j}^{k,l}\sigma_i^k\sigma_j^l$$

Gradients are computed over the expectation values of the Hamiltonian terms → $\partial_{\theta_i} S(\eta\,||\,\rho_\theta) = \mathrm{Tr}(\eta\, H_i) - \mathrm{Tr}(\rho_\theta\, H_i)$ → Convex loss landscape

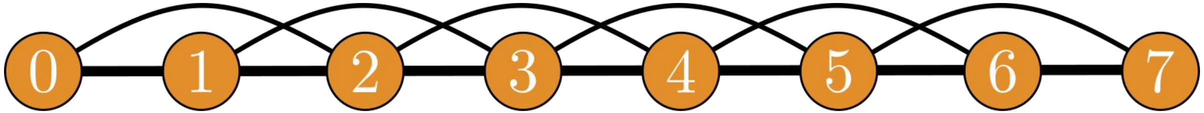
Target distribution

Learning Boltzmann distributions-1

Let us define a Boltzmann distribution on 8 sites as

$$E(s) = \sum_{i=1}^{n} s_i + \sum_{i=1}^{n-1} s_i s_{i+1} + 0.5 \times \sum_{i=1}^{n-2} s_i s_{i+2}$$

On the next nearest-neighbor connected lattice:

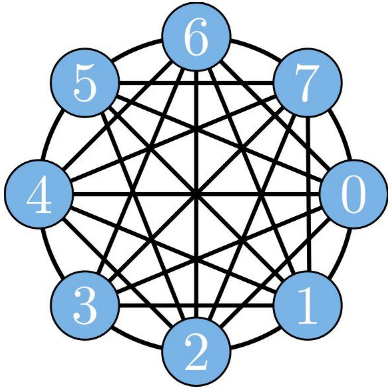


Model

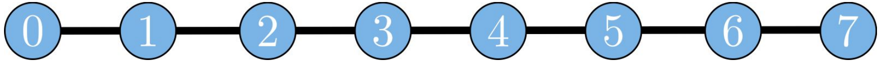
Learning Boltzmann distributions-2

Now let use define two Boltzmann machines:

All-to-all connected:



Nearest-neighbor (NN) connected:



Results

Learning Boltzmann distributions-3

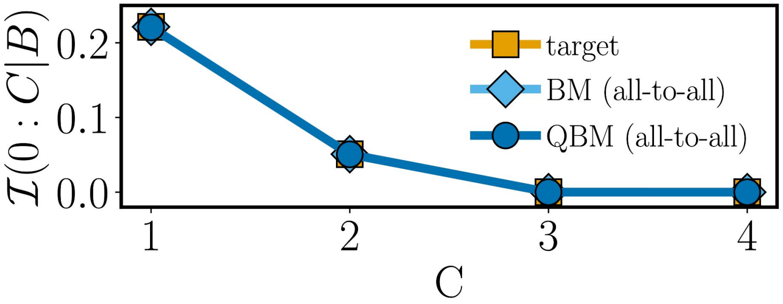
→ Both all-to-all connected models can learn the target distribution.

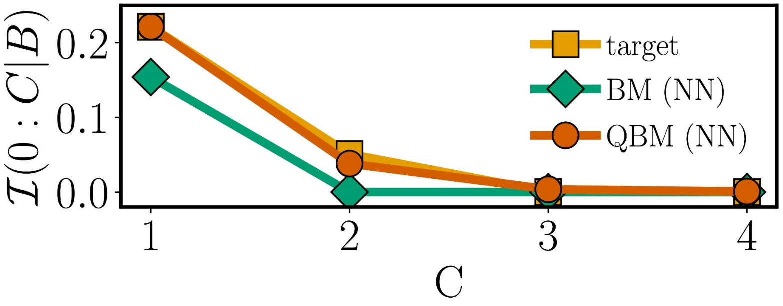
→ NN connected BM cannot learn the target distribution.

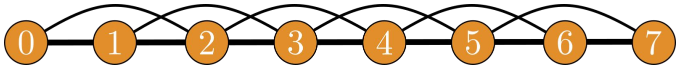
→ NN connected QBM can approximate the target distribution well.

D_{KL}	All-to-all	NN
BM	0.0	0.4
QBM	0.0	0.15

Conditional mutual information:







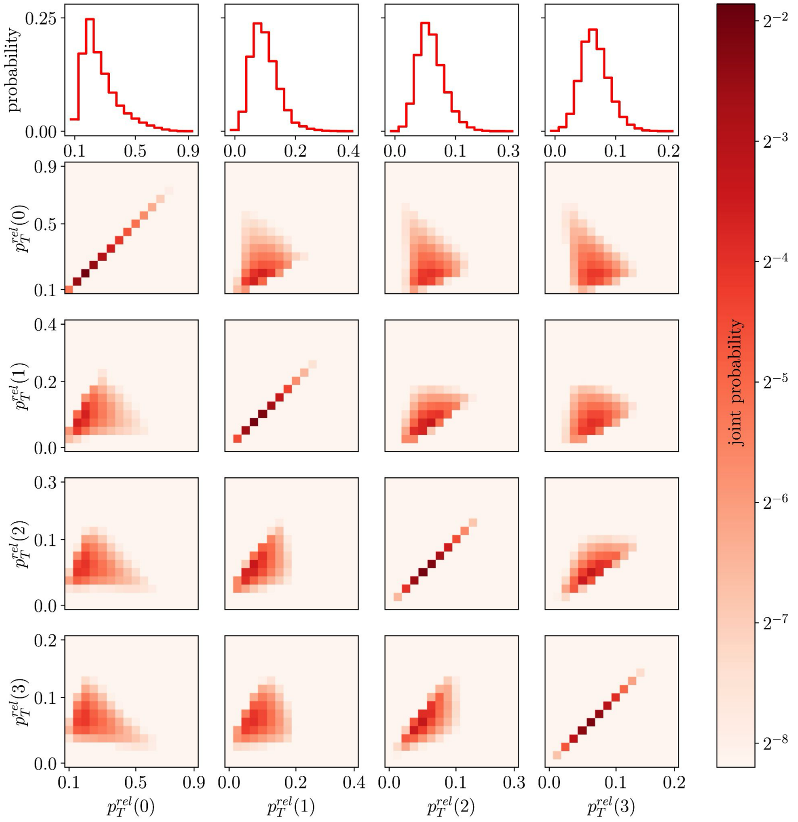
Learning particle jet events

Dataset description

To obtain the probability distribution, we create a histogram of particles' $|p_T|$ values that belongs to W bosons using the publicly available JetNet dataset.

We choose n_{particle} many leading particles to truncate the jet event.

$$n_{\text{particle}} \in \{2, 3, 4\}$$

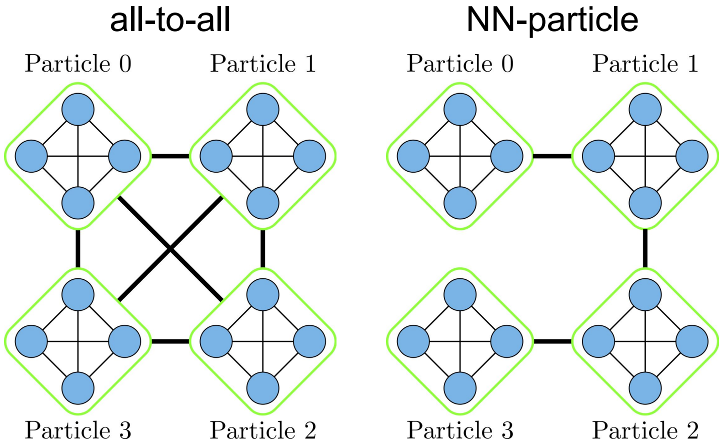


Learning particle jet events

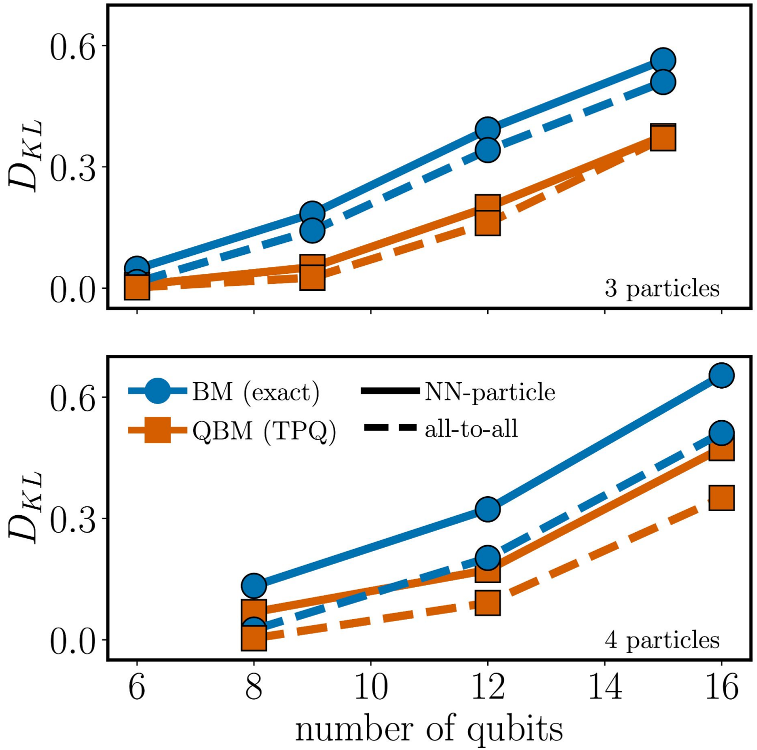
Comparing model connectivity

→ all-to-all connected QBM outperforms all other models.

→ NN-particle connected QBM either outperforms or matches the all-to-all connected BM.



Training results:



Learning particle jet events

Impact of operator pool

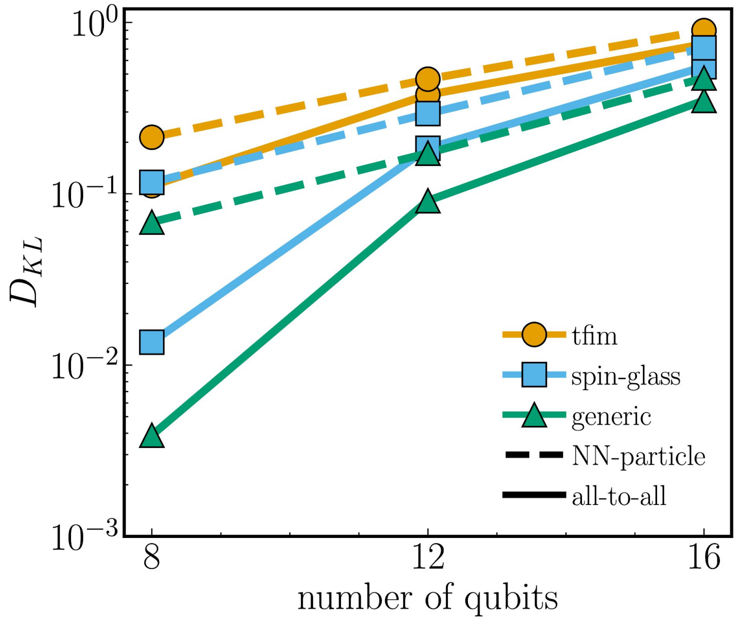
→ Performance depends on Hamiltonian choice.

→ Connectivity and Hamiltonian terms can be alternated based on the available resources.

→ Transversal field Ising model (tfim), often used in literature, is not the best choice!

$$H = \sum_{k \in \mathcal{P}_1} \sum_{i \in \mathcal{V}} \theta_i^k \sigma_i^k + \sum_{(k,l) \in \mathcal{P}_2} \sum_{(i,j) \in \mathcal{E}} \theta_{i,j}^{k,l} \sigma_i^k \sigma_j^l$$

	$\mathscr{P}_1$	$\mathscr{P}_2$
tfim	*X, Z*	*ZZ*
spin-glass	*X, Y, Z*	*XX, YY, ZZ*
generic	*X, Y, Z*	*XX, XY, XZ, YX, YY, YZ, ZX, ZY, ZZ*



Summary

More details available:

arXiv:2410.16363

- There are probability distributions that can be learned better with fv-QBMs compared to fv-BMs.
- fv-QBMs can learn higher dimensional probability distributions, while fv-BMs can only learn the distributions matching their dimension.
- fv-QBMs perform better with a larger operator pool and the benefit is interchangeable with its dimension.