Learning probability mass functions of particle jet events with quantum Boltzmann machines

Cenk Tüysüz[1,2], Maria Demidik[1,3], Luuk Coopmans[4], Enrico Rinaldi[5,6], Vincent Croft[7], Yacine Haddad[8], Matthias Rosenkranz[4], and Karl Jansen[1,3]

[1] Deutsches Elektronen-Synchrotron DESY, Zeuthen, Germany
[2] Institut für Physik, Humboldt-Universität zu Berlin, Berlin, Germany
[3] Computation-Based Science and Technology Research Center, The Cyprus Institute, Nicosia, Cyprus
[4] Quantinuum, Partnership House, Carlisle Place, London SW1P 1BX, United Kingdom
[5] Quantinuum K.K., Otemachi Financial City Grand Cube 3F, 1-9-2 Otemachi, Chiyoda-ku, Tokyo, Japan
[6] Interdisciplinary Theoretical and Mathematical Sciences (iTHEMS) Program, RIKEN, Wako, Saitama 351-0198, Japan
[7] Leiden Institute of Advanced Computer Science LIACS, Leiden University, Netherlands
[8] Northeastern University, Boston, Massachusetts, USA

Generative learning has attracted significant interest across various fields in recent years [1]. Among their applications, generative models play a crucial role in replacing Monte Carlo-based simulation strategies in scientific computing. One notable example is simulating jet events observed in particle collider experiments [2]. The different flavors of jets induce distinct high-order correlations among their constituent particles, making them challenging to model even with advanced generative techniques [2]. In this work, we propose addressing this problem using quantum Boltzmann Machines (BMs) [3]. Although quantum BMs have been employed in generative learning tasks previously, interest from the quantum machine learning community has declined in recent years due to the need for fault-tolerant quantum hardware. However, recent developments have shown that alternative models [4, 5] to quantum BMs suffer from issues such as barren plateaus [6] and local minima [7], raising concerns about their scalability. In contrast, it has been shown that fully-visible quantum BMs can be constructed in such a way that their loss landscape becomes convex and have trainability guarantees [8]. In this work, we extend the existing literature on quantum generative learning and highlight the advantages of quantum BMs with the following contributions:

- ★ We extend the definition of Ref. [9] for mutual information and show that it can be used to predict the success of *learning* with BMs and quantum BMs.
- ★ We present numerical evidence of quantum BMs successfully learning certain probability distributions that BMs fail to learn.
- ★ Using the proposed pairwise mutual information, we numerically show that probability mass functions obtained from a real dataset from particle physics (JetNet [10]) fell into the category of *difficult* to learn distributions.
- ★ We show that quantum BM can outperform BM for the JetNet dataset [10].

Contributions that are not covered explicitly in this extended abstract:

- ■ We show the importance of the Hamiltonian choice with numerical experiments.
- ■ We extend our results to larger system sizes using the thermal pure quantum (TPQ) states method [11].
- ■ We show how a fully-visible quantum BM can learn the parity distribution with nonzero-phase, while classical or quantum BMs with zero-phase fail at this task.

A Boltzmann machine can be constructed as a generative model whose output distribution corresponds to the Gibbs state of a classical or quantum system described by a Hamiltonian. The Gibbs state ρ of a Hamiltonian H at a finite inverse temperature $\beta = 1/kT$ is given as $\rho = \exp(-\beta H)/\mathrm{Tr}(\exp(-\beta H))$. If a classical Hamiltonian is chosen, the density matrix ρ is a diagonal matrix. We denote such a model as "BM". Otherwise, if a quantum Hamiltonian is chosen, namely H, which contains non-commuting terms, then non-zero off-diagonal entries are present in the density matrix. We denote such a model as "QBM". This allows the QBM to learn a richer class of target states, including non-diagonal Gibbs states. Consider the target probability distribution $p(s)$ such that $\sum_s p(s) = 1$, where s are all possible configurations. In this work, we set s to be all possible bitstrings of length n, where n is the system size (*e.g.*, number of bits). A BM, or a QBM, is said to be fully-visible if the system size is equivalent to the output size. In this work, we only consider the fully-visible setting. Then, the fully-visible BM encodes the target distribution into a density matrix such that $\eta = \mathrm{diag}(p(s))$.

In general, BMs are constructed up-to two body interactions, such that the resulting Hamiltonian is a two-local classical Hamiltonian. This formulation has been widely adopted in the classical machine learning community [12]. Although BMs are classical models, the exact computation of the underlying probability distribution becomes intractable with increasing system size. Therefore, there have been extensive efforts to use approximate methods to implement such models [13]. On the other hand, quantum computers can be used to prepare the Gibbs state of BMs [11]. Moreover, generalizing the system Hamiltonian to a quantum Hamiltonian may result in negligible overhead [8], which makes exploring QBMs intriguing.

The Gibbs state of a quantum Hamiltonian has non-zero off-diagonal entries. This results in a mismatch between the system ρ and the target η if diagonal encoding is used. To overcome this, Kappen [9] proposed the following

encoding in the computational basis, $\eta = |\psi\rangle\langle\psi|$, $|\psi\rangle = \sqrt{p(s)}e^{i\alpha(s)}|s\rangle$, where $\alpha(s)$ is an arbitrary phase that can be chosen by the user. We will choose $\forall s, \alpha(s) = 0$ here for simplicity, while this is discussed in the full manuscript version in more detail. This choice is denoted as the zero-phase embedding. Notice that this embedding results in a pure target state, in contrast to the mixed target state encoding of the diagonal BM model.

Now, let us define the Hamiltonians that describe BMs and QBMs. Consider a set of qubits with connectivity defined by a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents the qubits and $\mathcal{E}$ denotes their connectivity. Then, any two-local Hamiltonian can be expressed in the Pauli basis as follows,

$$H = \sum_{k\in\mathcal{P}_1}\sum_{i\in\mathcal{V}} \theta_i^k \sigma_i^k + \sum_{(k,l)\in\mathcal{P}_2}\sum_{(i,j)\in\mathcal{E}} \theta_{i,j}^{k,l}\sigma_i^k\sigma_j^l, \tag{1}$$

where σ_i^k denotes the Pauli matrix on the $i-$th qubit with k determining its type (X, Y, Z). The θ are real and trainable parameters of the model. Then, BM can be described with $\mathcal{P}_1 = \{Z\}, \mathcal{P}_2 = \{ZZ\}$. Common choices for QBM Hamiltonians contain physics inspired sets of operators, *e.g.*, the transversal field Ising model, which has $\mathcal{P}_1 = \{X, Z\}$ and $\mathcal{P}_2 = \{ZZ\}$. The authors of Ref. [9] propose using $\mathcal{P}_1 = \{X, Y, Z\}, \mathcal{P}_2 = \{XX, YY, ZZ\}$ and report results, outperforming classical BMs. Upon closer inspection, we have observed that the operator choice can be further improved.

Let us introduce two sets of operators with the corresponding Hamiltonians that we refer to as H_1 and H_2. H_1 contains the set $\mathcal{P}_1 = \{X, Z\}, \mathcal{P}_2 = \{XX, YY, ZZ\}$, and H_2 uses the set $\mathcal{P}_1 = \{X, Z\}, \mathcal{P}_2 = \{XX, YY, ZZ, XZ, ZX\}$. Notice that both H_1 and H_2 always have an *even* number of Y operators. Any operator with an *odd* number of Y operators will have zero expectation values when a zero-phase embedding is utilized. Hence, the operator choice of Ref. [9] is equivalent to H_1 if no phase is applied in the encoding. Under the same condition, H_2 is equivalent to the most general two-local Hamiltonian.

It has been previously shown that one fundamental difference between a BM and a QBM can be observed through the mutual information of the encoded target state [9]. The bipartite mutual information of a quantum system is given as,

$$\mathcal{I}^q(\rho, A:B) = S(\rho_A) + S(\rho_B) - S(\rho), \tag{2}$$

where $\rho_A = \mathrm{Tr}_B(\rho)$, $\rho_B = \mathrm{Tr}_A(\rho)$, $A:B$ denotes the bipartition of the subsystems and S is the von Neumann entropy. However, bipartite classical or quantum mutual information are not good metrics to evaluate capabilities of Boltzmann machines. This limitation arises because these models are in general constructed with up-to two body interactions, and the classical model is not expected to learn more than second-order statistics [9]. Therefore, we need to extend this definition. For this purpose, we introduce pairwise mutual information, which we denote as $\mathcal{I}_2$. The main idea behind this metric is to consider all single bit pairs and weight the mutual information by their mutual information with respect to the remaining $n-2$ bits. Let us consider single qubit indices k and l, we denote the remaining $n-2$ qubits as $\overline{kl}$. Then, $\mathcal{I}_2$ can be written as,

$$\mathcal{I}_2^q = \sum_{k\neq l}^{n} \mathcal{I}^q(\rho_{kl}, k:l)\mathcal{I}^q(\rho, kl:\overline{kl}), \tag{3}$$

where $\rho_{kl} = \mathrm{Tr}_{\overline{kl}}(\rho)$. Notice that the upper index in both cases denotes whether the target state is diagonal (c) or non-diagonal (q). The diagonal encoding is used to train BMs, while the non-diagonal encoding is used to train QBMs. We would like to note here that all of these mutual information metrics are in fact computationally expensive, but they provide important guidance on the capabilities of models at a small scale and whether this can be generalized and scaled up.

Next, we generate target probability distributions from random Boltzmann distributions up to third-order interactions, which can be defined as follows,

$$p(s) = \frac{e^{-\beta E(s)}}{\mathrm{Tr}(e^{-\beta E(s)})},\ E(s) = \sum_{i=1}^{n} w_i^{(1)} s_i + \sum_{i\neq j}^{n} w_{i,j}^{(2)} s_i s_j + \sum_{i\neq j\neq k}^{n} w_{i,j,k}^{(3)} s_i s_j s_k + \cdots, \tag{4}$$

where $s = s_1 s_2 \cdots s_n$ and $\forall s_i \in \{-1, +1\}$. The probability distribution p can be obtained by normalizing the given quantity, which is the partition function of the k-body interacting Ising model. Each of these terms is parametrized with $w^{(k)}$, which has $\binom{n}{k}$ unique entries. We normalize the parameters such that the contribution of each k-body interaction can be controlled. For example, we choose $\| w^{(1)} \|_1 = \| w^{(2)} \|_1 = \| w^{(3)} \|_1 = 1$ to give equal contributions to all. Alternatively, we can set the 1-norm of each term independently while sampling the unique value of each parameter from any distribution. We decide to sample the parameters from the uniform distribution between $[-1, 1]$, not to favor any configuration.

The pairwise mutual information $(\mathcal{I}_2)$ measured on different configurations of the Boltzmann distribution is reported in Fig. 1a. The cases with equal contributions of second and third-order interactions ($\| w^{(2)} \|_1 = \| w^{(3)} \|_1$) have I_2^c much closer to zero compared to the case in which the second-order dominates ($\| w^{(2)} \|_1 > \| w^{(3)} \|_1$). Moreover, in

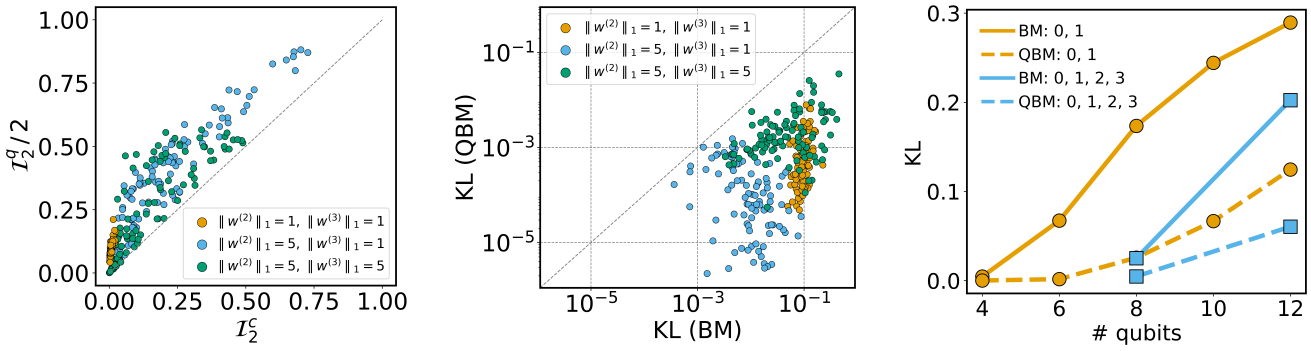
both $\| w^{(2)} \|_1 = 5$ cases I_2^c spans larger values compared to the $\| w^{(2)} \|_1 = 1$ case. Thus, we can conclude that I_2^c informs us about the second-order statistics successfully.

Then, we learn these distributions using the BM and the QBM equipped with H_2, and report the final KL divergences in Fig. 1b. KL divergence is defined as $\sum_s p(s) \log(p(s)/q(s))$, where $p(s)$ is the target distribution and $q(s)$ is the output distribution of the model. It is evident that QBM outperforms BM in all cases. This is attributed to the fact that all the distributions we consider here have non-zero third-order correlations. BM suffers significantly when the second-order statistics are limited ($\| w^{(2)} \|_1 = 1$), while QBM can reach much lower KL values. BM can show better performance on the distributions with $\| w^{(2)} \|_1 = 5$, while QBM still achieves lower KL. We note that the amount of second-order statistics also impacts QBM performance, although much milder compared to BMs. These insights provide a way to determine whether a distribution is suitable for learning with BMs.

Next, we target datasets that originate from particle physics to diversify our results. Jets are groups of particles that are observed at particle collider experiments such as the Large Hadron Collider [1]. There are many flavors of jets, and the study of each of them provides a fundamental understanding of the Standard Model and beyond [14]. Simulating jet events is essential to testing existing and new theories, but the computational cost of this task is a limiting factor [15]. To overcome this, researchers have employed classical generative deep learning techniques such as graph neural networks in recent years [1]. Despite their early success, these methods are not able to learn the correlations between the particles accurately [2]. In recent years, quantum machine learning researchers have begun to show interest in this problem [16–18]. However, most prior work either used methods such as feature reduction techniques that do not capture the high-order correlations or variational methods that were shown to suffer from issues such as barren plateaus [6,19,20]. In this work, we apply QBMs to the particle jet event generation problem for the first time.

We use the JetNet dataset [10] to produce probability distributions for investigating the performance of BMs and QBMs. For this purpose, for each jet we consider the p_T^{rel} (relative transversal momentum) values of the n particles with the largest p_T^{rel}. Then, we discretize the continuous values such that they are split into the k equal bins bounded by the minimum and maximum values observed in the dataset. This way, we represent one feature of a jet that consists of n particles with $n \times \log_2(k)$ bits and $\mathcal{O}(1/k)$ precision. This allows us to use a small system to represent the same problem, but with low precision. Using $\sim$ 100k samples, we first obtain the joint distribution of particles, which we then reshape and convert to a 1D probability distribution.

For the demonstration in this extended abstract, we chose the first two and four particles with the largest p_T^{rel} and obtain results with various numbers of bins. We have observed that $\mathcal{I}_2$ values measured on the data always remains small but non-zero. This agrees with prior work [2], which has emphasized the existence of high-order correlations in the data. Then, we train the BM and QBM on the target distributions and report the KL divergences in Fig. 1c. In all cases, we observe that QBM outperforms BM, resulting in much lower KL values. It is important to note here that the increase in KL values is not due to declining learning performance but rather the result of an increase in the size of the target distributions. Although not explicitly shown here, both models fit the continuous distribution better as the number of bins (or qubits) grows. These results are obtained using exact methods, hence, we don't present results for more than 12 qubits. In the full version of the manuscript, we scale our results using the recently introduced TPQ states method to train BMs and QBMs [11].



(a) (b) (c)

Fig. 1: (a) Pairwise mutual information ($\mathcal{I}_2$) of 100 four-bit Boltzmann distributions with different interaction strengths. (b) KL divergences measured after training on the same four-qubit Boltzmann distributions. (c) KL divergences measured after training on datasets obtained for two particles (0, 1) and four particles (0,1,2,3) with different number of bins using BM and QBM.

References

1. A. Butter, T. Plehn, S. Schumann, S. Badger, S. Caron, K. Cranmer, F. A. Di Bello, E. Dreyer, S. Forte, S. Ganguly, D. Gonçalves, E. Gross, T. Heimel, G. Heinrich, L. Heinrich, A. Held, S. Höche, J. N. Howard, P. Ilten, J. Isaacson, T. Janßen, S. Jones, M. Kado, M. Kagan, G. Kasieczka, F. Kling, S. Kraml, C. Krause, F. Krauss, K. Kröninger, R. K. Barman, M. Luchmann, V. Magerya, D. Maitre, B. Malaescu, F. Maltoni, T. Martini, O. Mattelaer, B. Nachman, S. Pitz, J. Rojo, M. Schwartz, D. Shih, F. Siegert, R. Stegeman, B. Stienen, J. Thaler, R. Verheyen, D. Whiteson, R. Winterhalder, and J. Zupan, "Machine Learning and LHC Event Generation," *SciPost Physics*, vol. 14, no. 4, p. 079, Apr. 2023, arXiv:2203.07460 [hep-ex, physics:hep-ph]. [Online]. Available: http://arxiv.org/abs/2203.07460
2. R. Kansal, J. Duarte, H. Su, B. Orzari, T. Tomei, M. Pierini, M. Touranakou, J.-R. Vlimant, and D. Gunopulos, "Particle Cloud Generation with Message Passing Generative Adversarial Networks," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. W. Vaughan, Eds., vol. 34, 2021, pp. 23 858–23 871. [Online]. Available: https://proceedings.neurips.cc/paper/2021/file/c8512d142a2d849725f31a9a7a361ab9-Paper.pdf
3. M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytskyy, and R. Melko, "Quantum Boltzmann Machine," *Physical Review X*, vol. 8, no. 2, p. 021050, May 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.8.021050
4. J.-G. Liu and L. Wang, "Differentiable learning of quantum circuit born machines," *Phys. Rev. A*, vol. 98, p. 062324, Dec 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.98.062324
5. B. Coyle, D. Mills, V. Danos, and E. Kashefi, "The Born supremacy: quantum advantage and training of an Ising Born machine," *npj Quantum Information*, vol. 6, no. 1, pp. 1–11, Jul. 2020, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41534-020-00288-9
6. M. S. Rudolph, S. Lerch, S. Thanasilp, O. Kiss, S. Vallecorsa, M. Grossi, and Z. Holmes, "Trainability barriers and opportunities in quantum generative modeling," May 2023, arXiv:2305.02881 [hep-ex, physics:quant-ph, stat]. [Online]. Available: http://arxiv.org/abs/2305.02881
7. X. You and X. Wu, "Exponentially Many Local Minima in Quantum Neural Networks," *arXiv:2110.02479 [quant-ph]*, Oct. 2021, arXiv: 2110.02479. [Online]. Available: http://arxiv.org/abs/2110.02479
8. L. Coopmans and M. Benedetti, "On the Sample Complexity of Quantum Boltzmann Machine Learning," 2023.
9. H. J. Kappen, "Learning quantum models from quantum or classical data," *Journal of Physics A: Mathematical and Theoretical*, vol. 53, no. 21, p. 214001, May 2020. [Online]. Available: https://dx.doi.org/10.1088/1751-8121/ab7df6
10. R. Kansal, J. Duarte, H. Su, B. Orzari, T. Tomei, M. Pierini, M. Touranakou, J.-R. Vlimant, and D. Gunopulos, "JetNet," Aug. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6975118
11. L. Coopmans, Y. Kikuchi, and M. Benedetti, "Predicting Gibbs-State Expectation Values with Pure Thermal Shadows," *PRX Quantum*, vol. 4, no. 1, p. 010305, Jan. 2023, publisher: American Physical Society. [Online]. Available: https://link.aps.org/doi/10.1103/PRXQuantum.4.010305
12. G. E. Hinton, "A Practical Guide to Training Restricted Boltzmann Machines," in *Neural Networks: Tricks of the Trade: Second Edition*, G. Montavon, G. B. Orr, and K.-R. Müller, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 599–619. [Online]. Available: https://doi.org/10.1007/978-3-642-35289-8_32
13. M. A. Carreira-Perpiñán and G. Hinton, "On contrastive divergence learning," in *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, R. G. Cowell and Z. Ghahramani, Eds., vol. R5. PMLR, 06–08 Jan 2005, pp. 33–40, reissued by PMLR on 30 March 2021. [Online]. Available: https://proceedings.mlr.press/r5/carreira-perpinan05a.html
14. C. Collaboration, "Search for physics beyond the standard model in events with jets and two same-sign or at least three charged leptons in proton-proton collisions at $\sqrt{s} =$ 13 TeV," *The European Physical Journal C*, vol. 80, no. 8, p. 752, Aug. 2020, arXiv:2001.10086 [hep-ex]. [Online]. Available: http://arxiv.org/abs/2001.10086
15. A. Butter and T. Plehn, *Generative Networks for LHC Events*, ch. Chapter 7, pp. 191–240. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/9789811234033_0007
16. A. Delgado, K. E. Hamilton, P. Date, J.-R. Vlimant, D. Magano, Y. Omar, P. Bargassa, A. Francis, A. Gianelle, L. Sestini, D. Lucchesi, D. Zuliani, D. Nicotra, J. de Vries, D. Dibenedetto, M. L. Martinez, E. Rodrigues, C. V. Sierra, S. Vallecorsa, J. Thaler, C. Bravo-Prieto, s. Y. Chang, J. Lazar, C. A. Argüelles, and J. J. M. de Lejarza, "Quantum computing for data analysis in high energy physics," Dec. 2022, arXiv:2203.08805 [hep-ex, physics:physics, physics:quant-ph]. [Online]. Available: http://arxiv.org/abs/2203.08805
17. A. Delgado and K. E. Hamilton, "Quantum Machine Learning Applications in High-Energy Physics," in *Proceedings of the 41st IEEE/ACM International Conference on Computer-Aided Design*, ser. ICCAD '22. New York, NY, USA: Association for Computing Machinery, Dec. 2022, pp. 1–5. [Online]. Available: https://dl.acm.org/doi/10.1145/3508352.3561114
18. A. Di Meglio, K. Jansen, I. Tavernelli, C. Alexandrou, S. Arunachalam, C. W. Bauer, K. Borras, S. Carrazza, A. Crippa, V. Croft, R. de Putter, A. Delgado, V. Dunjko, D. J. Egger, E. Fernandez-Combarro, E. Fuchs, L. Funcke, D. Gonzalez-Cuadra, M. Grossi, J. C. Halimeh, Z. Holmes, S. Kuhn, D. Lacroix, R. Lewis, D. Lucchesi, M. L. Martinez, F. Meloni, A. Mezzacapo, S. Montangero, L. Nagano, V. Radescu, E. R. Ortega, A. Roggero, J. Schuhmacher, J. Seixas, P. Silvi, P. Spentzouris, F. Tacchino, K. Temme, K. Terashi, J. Tura, C. Tuysuz, S. Vallecorsa, U.-J. Wiese, S. Yoo, and J. Zhang, "Quantum Computing for High-Energy Physics: State of the Art and Challenges. Summary of the QC4HEP Working Group," Jul. 2023, arXiv:2307.03236 [hep-ex, physics:hep-lat, physics:hep-th, physics:quant-ph]. [Online]. Available: http://arxiv.org/abs/2307.03236
19. A. Delgado and K. E. Hamilton, "Unsupervised Quantum Circuit Learning in High Energy Physics," *Physical Review D*, vol. 106, no. 9, p. 096006, Nov. 2022, arXiv:2203.03578 [hep-ex, physics:quant-ph]. [Online]. Available: http://arxiv.org/abs/2203.03578
20. A. Rousselot and M. Spannowsky, "Generative Invertible Quantum Neural Networks," Mar. 2023, arXiv:2302.12906 [hep-ph, physics:quant-ph]. [Online]. Available: http://arxiv.org/abs/2302.12906