communications physics

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42005-025-02353-1

Expressive equivalence of classical and quantum restricted Boltzmann machines

Maria Demidik[1,2] ✉, Cenk Tüysüz[1,3,4], Nico Piatkowski[5], Michele Grossi[4] & Karl Jansen[1,2]

The development of generative models for quantum machine learning has faced challenges such as trainability and scalability. A notable example is the quantum restricted Boltzmann machine (QRBM), where non-commuting Hamiltonians make gradient evaluation computationally demanding, even on fault-tolerant devices. In this work, we propose a semi-quantum restricted Boltzmann machine (sqRBM), a model designed to overcome difficulties associated with QRBMs. The sqRBM Hamiltonian commutes in the visible subspace while remaining non-commuting in the hidden subspace, enabling us to derive closed-form expressions for output probabilities and gradients. Our analysis shows that, for learning a given distribution, a classical model requires three times more hidden units than an sqRBM. Numerical simulations with up to 100 units validate this prediction. With reduced resource demands, sqRBMs provide a feasible framework for early quantum generative models.

Boltzmann machines (BM) are a prominent example of energy-based machine learning models inspired by statistical physics[1,2]. They have universal approximation properties[3], and are widely applied in areas such as collaborative filtering, dimensionality reduction, pattern recognition and generative modeling[4–7]. BMs consist of binary visible and hidden units, with an energy function defined by pairwise interactions and individual biases. This energy function takes the form of an Ising Hamiltonian, a model from statistical physics that assigns lower energy to more probable configurations. Each training step requires preparing a Gibbs state that describes the joint probability distribution of visible and hidden units. This results in BM training to be computationally demanding[8].

Contrastive divergence (CD) is a widely used approximation method to accelerate training of BMs[9]. However, CD provides only a rough estimate of the true gradients, often leading to unstable convergence and limiting the practical applicability of BMs[10]. While several improvements have been proposed[11,12], efficient training of BMs remains an open challenge in machine learning research.

Quantum computing offers opportunities to facilitate the training of BMs. Researchers have proposed multiple polynomial scaling quantum algorithms for Gibbs state preparation[13]. Utilizing quantum hardware for Gibbs state preparation could not only improve the training process but also offer a sampling advantage for BMs. Consequently, quantum computing could significantly increase the practical relevance of BMs. Moreover, the ability to prepare a Gibbs state on quantum computers enables generalizing the Hamiltonian of BMs with non-commuting terms, potentially enhancing the model's representational power. A model defined by a non-commuting Hamiltonian, referred to as a quantum Boltzmann machine (QBM)[14], is part of a broader class of algorithms within the field of quantum machine learning.

Quantum machine learning aims to enhance the capabilities of machine learning models with access to quantum computers. In the domain of generative modeling, the majority of proposals in the field have been based on parametrized quantum circuits such as quantum generative adversarial networks[15] or quantum circuit Born machines[16]. Recent studies have revealed that these models encounter trainability issues, such as barren plateaus[17,18], rendering them impractical. Although certain QBM constructions are similarly affected by these limitations[19], evidence suggests that alternative QBM formulations can successfully circumvent such issues[20].

Apart from challenges related to trainability, a major challenge for QBMs is the gradient estimation. The gradients of a QBM, which is defined by a generic non-commuting Hamiltonian, are known to be computationally intractable[14]. To overcome this challenge, various frameworks impose constraints on the Hamiltonian[14,21,22]. These approaches commonly avoid training non-commuting terms, treating them instead as hyperparameters. Although this constraint makes training cheaper, it inherently restricts the model's representational power.

From a different perspective, ref. 23 has introduced a training algorithm based on the variational quantum imaginary time evolution. This approach enables training of generic QBMs; however, it encounters scalability issues of variational algorithms[17].

Last but not least, recent results have shown that fully-visible QBMs (models with no hidden units) can be trained sample-efficiently[20,24]. Many studies have followed this result to show the capabilities of

[1]Deutsches Elektronen-Synchrotron DESY, Zeuthen, Germany. [2]Computation-Based Science and Technology Research Center, The Cyprus Institute, Nicosia, Cyprus. [3]Institut für Physik, Humboldt-Universität zu Berlin, Berlin, Germany. [4]European Organisation for Nuclear Research (CERN), Geneva, Switzerland. [5]Fraunhofer IAIS, Sankt Augustin, Germany. ✉e-mail: maria.demidik@desy.de

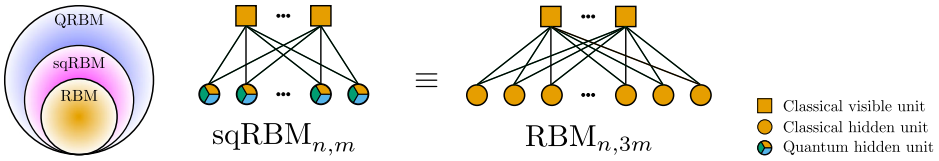

Fig. 1 | Summary of main results. This work introduces semi-quantum restricted Boltzmann machines (sqRBM) as an intermediate model, satisfying the relation QRBM ⊇ sqRBM ⊇ RBM. sqRBMs generalize RBMs by rendering the hidden units *quantum* through the use of non-commuting Hamiltonians. In Theorem 2, we show that sqRBM$_{n,m}$ ≡ RBM$_{n,3m}$, where n and m denote the number of visible and hidden units, respectively, with both models having the same number of parameters. In pedestrian terms, RBMs require three times as many hidden units as sqRBMs to learn the same target distribution.

fully-visible QBMs on learning from classical and quantum data[25–28]. While fully-visible QBMs are more expressive than their classical counterparts, their lack of hidden units considerably limits their applicability to many practical tasks[29].

Consequently, the computational cost of QBM training is fundamentally tied to the choice of Hamiltonian, highlighting a trade-off between tractability and expressiveness. In this work, we introduce semi-quantum restricted Boltzmann machines (sqRBM) designed for efficient gradient computation while enabling the training of non-commuting terms. This is achieved by defining a Hamiltonian that is diagonal in the subspace of visible units, while containing non-commuting terms in the subspace of hidden units. In this way, it serves as an intermediate model between RBMs and quantum restricted Boltzmann machines (QRBM) such that QRBM ⊇ sqRBM ⊇ RBM as illustrated in Fig. 1. This inclusion refers to model structure, not expressive power per hidden unit or per parameter. Being diagonal in the subspace of visible units allows sqRBM to provide a framework to explore the impact of non-commuting terms on learning from classical data.

In order to investigate the practical importance of sqRBMs, we establish a direct correspondence between RBMs and sqRBMs. In particular, Theorem 2 shows that an sqRBM$_{n,m}$ is expressively equivalent to an RBM$_{n,3m}$, where n and m denote the number of visible and hidden units, respectively. This result implies that an sqRBM requires only one-third of the hidden units to match the representational power of an RBM.

While this may appear to be a linear improvement, its practical impact is amplified in quantum settings. The cost of quantum Gibbs state preparation scales polynomially, often with a high polynomial degree in the number of units[13]. Therefore, reducing the number of hidden units from $3m$ to m can lead to a substantial reduction in quantum resource requirements, including circuit depth and qubit count. This makes sqRBMs a promising alternative for fault-tolerant quantum devices.

We also present numerical results across multiple datasets and models with up to 100 units, supporting the theoretical claims and illustrating the practical viability of sqRBMs. A summary of the main results is provided in Fig. 1.

Results

Theoretical intuition

Challenges in training generic QBMs stem from non-trivial estimation of model gradients[14]. As outlined in Methods, consider a QBM defined by a parametrized non-commuting Hamiltonian of the form $H = \sum_i \theta_i H_i$, where θ_i are real-valued trainable parameters and H_i are low-weight Pauli-strings. In this setting, one needs to compute terms of the form $\partial_{\theta_i} e^{-H}$ and $\mathrm{Tr}\left[\Lambda_\nu \partial_{\theta_i} e^{-H}\right]$, where Λ_ν is a diagonal projection operator. The structure of such gradient expressions is formalized in Proposition 4. Observe that the term $\partial_{\theta_i} e^{-H}$ admits a series expansion as follows:

$$\partial_{\theta_i} e^{-H} = e^{-H}\left(-H_i - \tfrac{1}{2}\left[H, H_i\right] - \tfrac{1}{6}\left[H, \left[H, H_i\right]\right] + \cdots\right). \tag{1}$$

We provide the details of the derivation in the Supplementary Note 3. Then, we rewrite $\mathrm{Tr}\left[\Lambda_\nu \partial_{\theta_i} e^{-H}\right]$ by substituting the derivative of the matrix exponential term. Exploiting the linearity of the trace, we obtain the following expression

$$\begin{aligned}\mathrm{Tr}\left[\Lambda_\nu \partial_{\theta_i} e^{-H}\right] = &-\mathrm{Tr}\left[\Lambda_\nu e^{-H} H_i\right] \\ &-\frac{1}{2}\mathrm{Tr}\left[\Lambda_\nu e^{-H}\left[H, H_i\right]\right] \\ &-\frac{1}{6}\mathrm{Tr}\left[\Lambda_\nu e^{-H}\left[H, \left[H, H_i\right]\right]\right] \\ &+\cdots\end{aligned} \tag{2}$$

Next, we observe that if $[\Lambda_\nu, H] = 0$ (while noting that $[H_i, H] \neq 0$ still holds), the second term in Eq. (2) simplifies to:

$$\begin{aligned}\mathrm{Tr}\left[\Lambda_\nu e^{-H}\left[H, H_i\right]\right] &= \mathrm{Tr}\left[\Lambda_\nu e^{-H} H H_i\right] - \mathrm{Tr}\left[\Lambda_\nu e^{-H} H_i H\right] \\ &= \mathrm{Tr}\left[H \Lambda_\nu e^{-H} H_i\right] - \mathrm{Tr}\left[\Lambda_\nu e^{-H} H_i H\right] \\ &= \mathrm{Tr}\left[\Lambda_\nu e^{-H} H_i H\right] - \mathrm{Tr}\left[\Lambda_\nu e^{-H} H_i H\right] \\ &= 0,\end{aligned} \tag{3}$$

where we leverage the commutation rules and the cyclic property of the trace. This can be repeated for the rest of the terms in Eq. (2) that contain commutators to observe that only $\mathrm{Tr}[\Lambda_\nu e^{-H} H_i]$ results in a non-zero value. Similarly, notice that $\mathrm{Tr}\left[\partial_{\theta_i} e^{-H}\right] = -\mathrm{Tr}\left[e^{-H} H_i\right]$ for all models, independent of the Hamiltonian definition. Then, the computation of QBM gradients can be simplified, although the Hamiltonian still contains non-commuting terms.

Recall that we assumed $[\Lambda_\nu, H] = 0$ to obtain simplified gradients for non-commuting Hamiltonians. Let us define Λ_ν for a QBM with n visible and m hidden units in the Pauli basis as

$$\Lambda_\nu = \left(\bigotimes_{i=1}^{n} \frac{1}{2}\left(I + (-1)^{\nu_i} Z\right)\right) \otimes I^{\otimes m}. \tag{4}$$

Then, we observe that the costly gradient computation of QBMs can be avoided by choosing a non-commuting Hamiltonian that commutes with Pauli-Z only over the visible units, while the remaining terms of the Hamiltonian can be defined with an arbitrary Pauli operator set. We define such a model as a semi-quantum restricted Boltzmann machine (sqRBM).

Semi-quantum Boltzmann machines

Definition 1. (Semi-quantum RBM). A semi-quantum restricted Boltzmann machine with n visible and m hidden units, denoted sqRBM$_{n,m}$, is described by a parameterized Hamiltonian of the form $H = H_\mathrm{v} + H_\mathrm{h} + H_\mathrm{int}$. The three terms of the Hamiltonian are defined as follows:

$$\begin{aligned}H_\mathrm{v} = \sum_{i=1}^{n} a_i^Z \sigma_i^Z, \quad H_\mathrm{h} = \sum_{P \in \mathcal{W}_\mathrm{h}}^{|\mathcal{W}_\mathrm{h}|} \sum_{j=1}^{m} b_j^P \sigma_{n+j}^P, \\ H_\mathrm{int} = \sum_{P \in \mathcal{W}_\mathrm{h}}^{|\mathcal{W}_\mathrm{h}|} \sum_{i=1}^{n} \sum_{j=1}^{m} w_{i,j}^{Z,P} \sigma_i^Z \sigma_{n+j}^P,\end{aligned} \tag{5}$$

Table 1 | Classification of various model

Model	$\mathcal{W}_h$	Parameters
$RBM_{n,m}$	$\{Z\}$	$n+m\,(n+1)$
$sqRBM_{n,m}$	$\{X, Z\}$	$n+2m\,(n+1)$
$sqRBM_{n,m}$	$\{Y, Z\}$	$n+2m\,(n+1)$
$sqRBM_{n,m}$	$\{X, Y, Z\}$	$n+3m\,(n+1)$

where $\boldsymbol{a} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^{|\mathcal{W}_h|\cdot m}$ and $\boldsymbol{w} \in \mathbb{R}^{|\mathcal{W}_h|\cdot n\cdot m}$ are the parameter vectors of the model. $\mathcal{W}_h$ is a non-commuting set of one-qubit Pauli operators that non-trivially act only on the hidden units.

Notice that if one chooses $\mathcal{W}_h = \{X\}$, $\mathcal{W}_h = \{Y\}$ or $\mathcal{W}_h = \{Z\}$, H is a commuting Hamiltonian, and these models are equivalent to an RBM. There are three possible non-commuting choices for an sqRBM ($\mathcal{W}_h = \{X, Z\}, \mathcal{W}_h = \{Y, Z\}$ or $\mathcal{W}_h = \{X, Y, Z\}$). Pauli sets $\mathcal{W}_h = \{X, Z\}$ and $\mathcal{W}_h = \{Y, Z\}$ result in two equivalent models that differ by a change of basis. Please refer to the Supplementary Note 2 for definitions of Pauli operators. We provide various models with their number of parameters in Table 1.

One can leverage the similarity of sqRBMs to RBMs in order to obtain a closed-form expression for the output probabilities as well as their gradients. For an sqRBM, contributions of X, Y, and Z terms are equivalent. Let us denote the state (not to be confused with a quantum state) of the hidden units for a given visible unit configuration v with $\phi_j^P(v)$ such that

$$\phi_j^P(v) = b_j^P + \sum_{i=1}^{n} (-1)^{v_i} w_{i,j}^{Z,P}, \tag{6}$$

where $P \in \{X, Y, Z\}$. Then, we define the following vector that combines the states of all possible three Pauli operators:

$$\Phi_j(v) = \begin{bmatrix} \phi_j^X(v) & \phi_j^Y(v) & \phi_j^Z(v) \end{bmatrix}. \tag{7}$$

This leads to Proposition 1, which describes the output probabilities of sqRBMs.

Proposition 1. (Output probabilities of sqRBM). The unnormalized output probabilities of an $sqRBM_{n,m}$ are as follows:

$$\tilde{p}_v = \left(\prod_{i=1}^{n} e^{-(-1)^{v_i} a_i^Z}\right)\left(\prod_{j=1}^{m} \cosh(||\Phi_j(v)||_2)\right), \tag{8}$$

where $v \in \{0, 1\}^n$ and $\Phi_j(v)$ is the vector of hidden states as described in Eq. (7). Then, the normalized output probabilities are given as

$$p_v = \tilde{p}_v / \sum_v \tilde{p}_v. \tag{9}$$

The proof is provided in the Supplementary Note 4. Next, we provide the gradients of sqRBMs following the result of Proposition 1.

Proposition 2. (Gradients of sqRBM). An $sqRBM_{n,m}$ with the output probability distribution p can be trained to minimize the negative log-likelihood with respect to the target probability distribution q using the following gradient rule:

Gradients of the parameters for the field terms acting on the visible units are given as

$$\partial_{a_i} \mathcal{L} = \sum_v q_v \left((-1)^{v_i} - \sum_v (-1)^{v_i} p_v\right). \tag{10}$$

Gradients of the parameters for the field terms acting on the hidden units are given as

$$\begin{aligned} \partial_{b_j^P} \mathcal{L} = &- \sum_v q_v \frac{\phi_j^P(v)}{||\Phi_j(v)||_2} \tanh\left(||\Phi_j(v)||_2\right) \\ &+ \sum_v q_v \sum_v \frac{\phi_j^P(v)}{||\Phi_j(v)||_2} \tanh\left(||\Phi_j(v)||_2\right) p_v. \end{aligned} \tag{11}$$

Gradients of the parameters for the interaction terms acting on both visible and hidden units are given as

$$\begin{aligned} \partial_{w_{i,j}^{Z,P}} \mathcal{L} = &- \sum_v q_v (-1)^{v_i} \frac{\phi_j^P(v)}{||\Phi_j(v)||_2} \tanh\left(||\Phi_j(v)||_2\right) \\ &+ \sum_v q_v \sum_v (-1)^{v_i} \frac{\phi_j^P(v)}{||\Phi_j(v)||_2} \tanh\left(||\Phi_j(v)||_2\right) p_v. \end{aligned} \tag{12}$$

The proof is provided in the Supplementary Note 5. Notice that both the gradients and output probabilities for RBM and sqRBM contain summations over all visible configurations, which grow exponentially in input size. This causes both of these models to be intractable for large input sizes on classical computers.

So far, we have focused on BM variants with restricted connectivity. Incorporating lateral connections within visible and hidden layers increases the computational cost of the training procedure[12]. Although additional connectivity enhances the expressive power of BMs, computational complexity limits their practical applicability. However, training fully-connected QBMs without a significant computational overhead may be discovered within the quantum computing framework. Motivated by this opportunity, we introduce a fully-connected sqBM.

Definition 2. (Semi-quantum BM). A semi-quantum Boltzmann machine (sqBM) with n visible and m hidden units is denoted $sqBM_{n,m}$. An sqBM is a generalization of an sqRBM with lateral connections in the visible and hidden layers, which is described by a parameterized Hamiltonian of the form $H = H_v + H_h + H_{int}$. The three terms of the Hamiltonian are defined as follows:

$$\begin{aligned} H_v &= \sum_{i=1}^{n} \theta_i^Z \sigma_i^Z + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \theta_{i,j}^{Z,Z} \sigma_i^Z \sigma_j^Z, \\ H_h &= \sum_{k\in\mathcal{W}_h}^{|\mathcal{W}_h|} \left(\sum_{i=n+1}^{n+m} \theta_i^k \sigma_i^k + \sum_{l\in\mathcal{W}_h}^{|\mathcal{W}_h|} \sum_{i=n+1}^{n+m-1} \sum_{j=i+1}^{n+m} \theta_{i,j}^{k,l} \sigma_i^k \sigma_j^l \right), \\ H_{int} &= \sum_{k\in\mathcal{W}_h}^{|\mathcal{W}_h|} \sum_{i=1}^{n} \sum_{j=n+1}^{n+m} \theta_{i,j}^{Z,k} \sigma_i^Z \sigma_j^k, \end{aligned} \tag{13}$$

where $\mathcal{W}_h$ is a non-commuting set of one-qubit Pauli operators that non-trivially act only on the hidden units.

Ultimately, the results from Eqs. (2) and (3) lead to a simplified expression for the gradients of a generic sqBM. Importantly, satisfying the condition $[\Lambda_v, H] = 0$ does not impose restrictions on the connectivity between units. Then, one obtains the following closed-form expression for the gradients that is significantly cheaper to compute than generic QBM gradients.

Proposition 3. (Gradients of sqBM). An $sqBM_{n,m}$ can be trained to minimize the negative log-likelihood with respect to the target probability

distribution q using the following gradient rule:

$$\partial_{\theta_i}\mathcal{L} = -\sum_{\nu} q_\nu \left(\underbrace{-\frac{\mathrm{Tr}[\Lambda_\nu e^{-H} H_i]}{\mathrm{Tr}[\Lambda_\nu e^{-H}]}}_{\text{positive phase}} + \underbrace{\frac{\mathrm{Tr}[e^{-H} H_i]}{\mathrm{Tr}[e^{-H}]}}_{\text{negative phase}} \right), \tag{14}$$

where θ_i is any real-valued parameter of the model, when the Hamiltonian terms are grouped such that $H = \sum_i \theta_i H_i$.

The proof is provided in the Supplementary Note 6. Notice that Proposition 3 provides the recipe to compute the gradients of either an sqRBM or the more general sqBM on a quantum computer. In this work, we study the relationship between RBMs and sqRBMs; therefore, in the remainder of the work, we will not consider sqBMs and leave their study as future work.

Expressive equivalence

The output probabilities derived in Proposition 1 exhibit a structural resemblance between classical RBMs and their semi-quantum counterparts. This naturally raises the question: can we establish a formal correspondence between these models? Specifically, do their functional forms, and consequently their expressivity, coincide under certain conditions?

To explore this, let us consider an $\mathrm{sqRBM}_{n,1}$ with a single hidden qubit. For simplicity, we set all field terms acting on the visible units to zero. Then, using Eq. (8), the unnormalized output probability is given by

$$\tilde{p}_\nu = \cosh\left(\sqrt{\phi_1^X(\nu)^2 + \phi_1^Y(\nu)^2 + \phi_1^Z(\nu)^2}\right), \tag{15}$$

whose Taylor expansion generates a polynomial with monomials in the span

$$\mathrm{span}\left\{\phi_1^{\mathrm{X}}(\mathrm{v})^{2\mathrm{i}}\phi_1^{\mathrm{Y}}(\mathrm{v})^{2\mathrm{j}}\phi_1^{\mathrm{Z}}(\mathrm{v})^{2\mathrm{k}} \mid \mathrm{i},\mathrm{j},\mathrm{k} \in \mathbb{N}_0\right\}, \tag{16}$$

as shown in the Supplementary Note 2.

Now consider a classical $\mathrm{RBM}_{n,3}$ with three hidden units. Its unnormalized output probability takes the form

$$\tilde{p}_\nu = \cosh\left(\phi_1^Z(\nu)\right)\cosh\left(\phi_2^Z(\nu)\right)\cosh\left(\phi_3^Z(\nu)\right), \tag{17}$$

whose Taylor expansion yields monomials of the form

$$\mathrm{span}\left\{\phi_1^{\mathrm{Z}}(\mathrm{v})^{2\mathrm{i}}\phi_2^{\mathrm{Z}}(\mathrm{v})^{2\mathrm{j}}\phi_3^{\mathrm{Z}}(\mathrm{v})^{2\mathrm{k}} \mid \mathrm{i},\mathrm{j},\mathrm{k} \in \mathbb{N}_0\right\}. \tag{18}$$

As shown in Supplementary Note 2, these functional forms coincide when ϕ^X, ϕ^Y, ϕ^Z in the sqRBM are treated analogously to independent linear maps in the RBM. This implies that the two models can generate the same class of functions over visible configurations. Moreover, both models use the same number of parameters. This correspondence is formalized below in Theorem 1 and 2.

Theorem 1. (Expressive equivalence of $\mathrm{sqRBM}_{n,1}$ and $\mathrm{RBM}_{n,|\mathcal{W}_\mathrm{h}|}$). Let $\mathcal{W}_\mathrm{h} \subseteq \{X, Y, Z\}$ be a set of Pauli operators. Then, an $\mathrm{sqRBM}_{n,1}$ with a single hidden qubit and operator set $\mathcal{W}_\mathrm{h}$ is expressively equivalent to an $\mathrm{RBM}_{n,|\mathcal{W}_\mathrm{h}|}$ with $|\mathcal{W}_\mathrm{h}|$ classical hidden units. That is, both models can represent the same class of functions over visible configurations $\nu \in \{0,1\}^n$, up to a reparameterization of the corresponding linear maps.

The proof follows Eq. (16) and (18), and it is provided in the Supplementary Note 8. Following Theorem 1, one can generalize the statement to multiple hidden units for sqRBMs and obtain Theorem 2.

Theorem 2. (Expressive equivalence of $\mathrm{sqRBM}_{n,m}$ and $\mathrm{RBM}_{n,|\mathcal{W}_\mathrm{h}|\cdot m}$). Let $\mathcal{W}_\mathrm{h} \subseteq \{X, Y, Z\}$ be a set of Pauli operators. Then, an $\mathrm{sqRBM}_{n,m}$ with m hidden qubits and operator set $\mathcal{W}_\mathrm{h}$ is expressively equivalent to an $\mathrm{RBM}_{n,|\mathcal{W}_\mathrm{h}|\cdot m}$ with $|\mathcal{W}_\mathrm{h}| \cdot m$ classical hidden units. That is, both models can represent the same class of functions over visible configurations $\nu \in \{0,1\}^n$, up to a reparameterization of the corresponding linear maps.

Proof.. The result follows directly from Theorem 1. Each hidden qubit in the sqRBM contributes a factor of the form $\cosh(\|\Phi(\nu)\|_2)$, where $\Phi(\nu) \in \mathbb{R}^{|\mathcal{W}_\mathrm{h}|}$ is a vector of linear functions corresponding to the Pauli operators in $\mathcal{W}_\mathrm{h}$. As shown in Theorem 1, each such term is expressively equivalent to a product of $|\mathcal{W}_\mathrm{h}|$ classical cosh activations.

Since the hidden units in both models contribute multiplicatively and independently (cf. Proposition 1), the expressive equivalence carries over additively across hidden units. Therefore, an $\mathrm{sqRBM}_{n,m}$ with m hidden qubits is expressively equivalent to an $\mathrm{RBM}_{n,|\mathcal{W}_\mathrm{h}|\cdot m}$, up to a reparameterization of the linear maps that define the activations.

The consequence of Theorem 2 is that an $\mathrm{sqRBM}_{n,m}$ with the operator pool $\mathcal{W}_\mathrm{h}$ can solve the tasks that an $\mathrm{RBM}_{n,|\mathcal{W}_\mathrm{h}|\cdot m}$ can solve equally well with the same number of parameters. In other words, these models have the same expressivity. In the next section, we support Theorem 2 with numerical simulations.

Numerical results

In this section, we provide numerical results to support our theoretical findings. We perform exact numerical simulations to train various RBM and sqRBM models. We use four distinct datasets, which are defined as follows:

simplified-BAS dataset: n-bit uniform probability distribution over the bitstrings that correspond to vertical or horizontal lines on a $2 \times n/2$ grid.
$\mathcal{O}(\boldsymbol{n}^2)$ dataset: n-bit uniform probability distribution over randomly chosen n^2 bitstrings.
Cardinality dataset: n-bit uniform probability distribution over the bitstrings that have $n/2$ cardinality.
Parity dataset: n-bit uniform probability distribution over the bitstrings that have even parity.

We train all models 100 times with parameters randomly initialized from a uniform distribution between $[-1, 1]$ using the AMSGRAD optimizer[30] with the hyperparameters $\{\mathrm{lr} = 0.1, \beta_1 = 0.9, \beta_2 = 0.999\}$ and D_{KL} as the loss function. We have not performed hyperparameter optimization as all models converge within a reasonable number of iterations. We emphasize that all results can be improved with dedicated hyperparameter optimization. Our goal in this work is not to find the best performing model but to treat all models the same way.

We employ an RBM and two quantum models: sqRBM$\{X, Z\}$ and sqRBM$\{X, Y, Z\}$. These quantum models are defined by two sets of operators for the hidden units, $\mathcal{W}_\mathrm{h} = \{X, Z\}$ and $\mathcal{W}_\mathrm{h} = \{X, Y, Z\}$, respectively. Since we minimize the Kullback-Leibler divergence D_{KL}, we report values of another metric, namely total variation distance ($\mathrm{TVD}(q||p) = \frac{1}{2}||p - q||_1$) for both RBM and sqRBM models across the described datasets in Fig. 2. As a reference point for practical purposes, we report the TVD = 0.2 line. It can be seen that all models can go below the TVD = 0.2 line given that they have sufficient number of hidden units.

In general, the number of hidden units required to achieve a good approximation depends on the dataset. More specifically, the support of a probability distribution is a good measure of difficulty. In Fig. 2, the datasets are ordered in increasing difficulty from left to right. There, we observe that one needs a larger value of m for all models as the support of the dataset increases. In practice, it is considered that a dataset has $\mathrm{supp}(p) \in \mathcal{O}(\mathrm{poly}(n))$, therefore, it is expected that $m \in \mathcal{O}(\mathrm{poly}(n))$ suffices to learn a distribution with good precision.

Our findings across the considered datasets indicate that sqRBM$\{X, Y, Z\}$ requires fewer hidden units to reach the reference point compared to sqRBM$\{X, Z\}$. Similarly, sqRBM$\{X, Z\}$ requires fewer hidden units to reach the reference point compared to RBM. However, all models are able to reach the same low values of TVD, provided they have a sufficient number of hidden units. Theorem 2 predicts $\mathrm{sqRBM}_{n,m}\{X, Z\}$ to perform equally as well as $\mathrm{RBM}_{n,2m}$ (with $2m$ hidden units in the RBM) and similarly

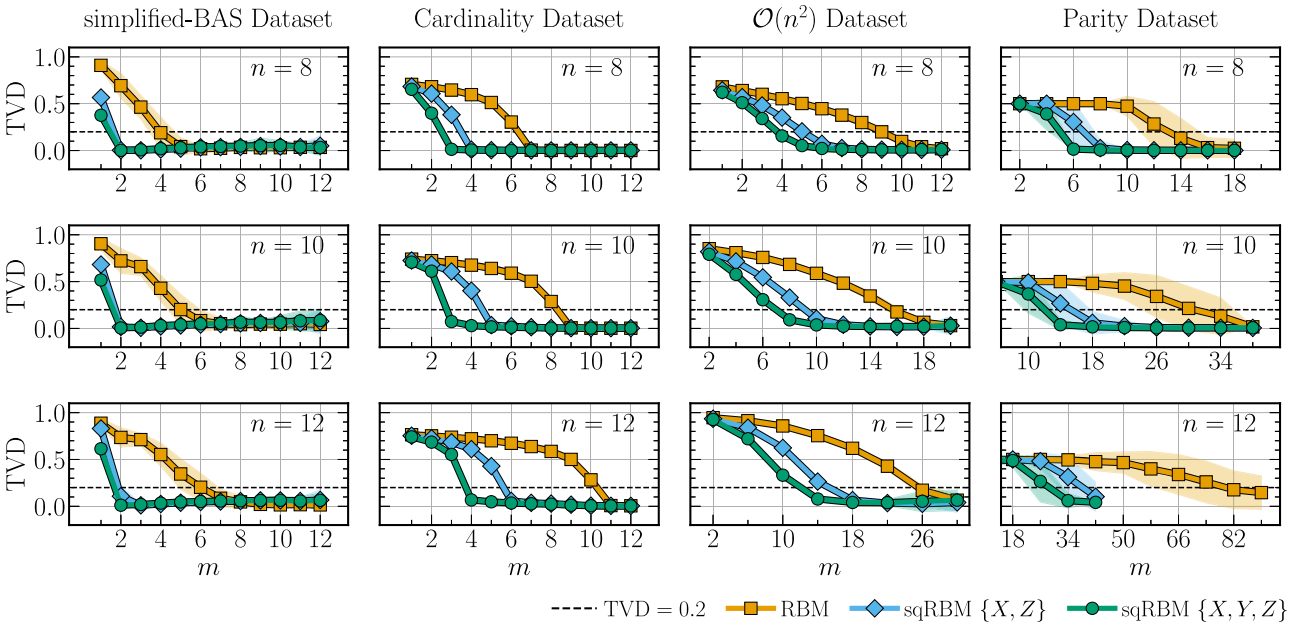

Fig. 2 | Training results. We train three models (RBM (orange square markers), sqRBM{*X*, *Z*} (blue rotated square markers) and sqRBM{*X*, *Y*, *Z*} (green circle markers)) over four datasets with three different input sizes ($n \in \{8, 10, 12\}$) and various number of hidden units in the range $m \in [1, 90]$. We report the total variation distance (TVD) measured after training all models 100 times with different initial parameters. The solid lines report the average, while the shades indicate the standard deviation. Each column reports results for a different dataset, ordered in increasing difficulty from left to right. The target probability distribution for $\mathcal{O}(n^2)$ dataset is varied for each run, using the same 100 seed for all models. The same target probability distribution is used for the other datasets in all runs. The TVD = 0.2 (black dashed line) is plotted as a reference point.

for sqRBM$_{n,m}${*X*, *Y*, *Z*} to perform equally as well as RBM$_{n,3m}$. We emphasize that the models have the same number of parameters for these settings. We numerically verify Theorem 2 by computing the ratios of the number of hidden units that are sufficient for each model to achieve TVD < 0.2 in Fig. 3. As expected, the minimum number of hidden units m, that is sufficient to go below the threshold varies with dataset and input size. Overall, results align with the prediction of Theorem 2 across four datasets and varying input sizes.

Discussion

In this work, we introduce semi-quantum restricted Boltzmann machines (sqRBM), a subclass of quantum Boltzmann machines (QBM). The sqRBM Hamiltonian acts with arbitrary non-commuting Pauli operators on hidden units, whereas operators for visible units are restricted to a commuting set. The structure provided by the Hamiltonian allows us to circumvent the expensive computation of the positive phase in Proposition 4 and enables us to derive the analytical output probabilities and gradients in Proposition 1 and Proposition 2.

In the field of quantum machine learning, efficient evaluation of gradients remains a major challenge. Expressive models based on parameterized quantum circuits exhibit trainability issues, often linked to barren plateaus of the loss landscape[17]. Barren plateaus lead to an exponential increase in the number of samples required to evaluate gradients as the system size grows. This significantly limits the scalability of circuit-based quantum machine learning models, making it difficult to demonstrate their practical viability.

QBMs offer a promising alternative to circuit-based approaches. However, when the entanglement entropy between visible and hidden units obeys the volume law, generic QRBM models are susceptible to barren plateaus[19]. Importantly, sqRBMs do not have entanglement between visible and hidden units; therefore, this problem is naturally mitigated, and the gradients can be estimated with polynomially many samples from the Gibbs state. Moreover, the similarity between RBM and sqRBM gradients also suggests the absence of vanishing gradients. For completeness, we provide numerical results that confirm this conjecture in the Supplementary Note 9.

Expressivity of QBMs is directly related to their Hamiltonian. A popular choice in the literature is to employ the transverse field Ising model (TFIM)[14] due to its similarity to the classical Ising model. However, QBMs based on TFIM have been reported to exhibit only marginal improvement in learning capacity compared to BMs, while generic Hamiltonians demonstrate significant improvement[26]. Our results help explain the poor performance of such models and provide a strategy to build more expressive ones.

Let us consider a QRBM based on the TFIM Hamiltonian. Such a model is defined by adding transverse X fields to visible and hidden units of the RBM Hamiltonian. In our notation, this can be expressed as $\mathcal{W}_\mathrm{v} = \mathcal{W}_\mathrm{h} = \{X, Z\}$ and $\mathcal{W}_\mathrm{int} = \{ZZ\}$. Recall that the state of the j-th hidden unit is described with $||\Phi_j(v)||_2$ (see Eq. (8)). Then, for a TFIM Hamiltonian, one obtains $\sqrt{\phi_j^Z(v)^2 + (b_j^X)^2}$, where b_j^X is the parameter of the transverse X field of the TFIM Hamiltonian on the j-th hidden unit. Then, it follows that the contribution of the transverse field is independent of the visible unit configuration v and can not significantly change the model's expressivity. Hence, the output probability distributions of such QBMs closely resemble those of classical models. One could improve the expressivity of such QBM models by including ZX terms in the Hamiltonian to have $\mathcal{W}_\mathrm{v} = \mathcal{W}_\mathrm{h} = \{X, Z\}$ and $\mathcal{W}_\mathrm{int} = \{ZZ, ZX\}$. Note that such a model is a QRBM and not an sqRBM. Therefore, its gradients are not efficiently computable.

The practical feasibility of machine learning models depends both on their learning capacity and the computational resources required for training and sampling. The computational cost of training an RBM$_{n,m}$ using exact analytical gradients, as well as sampling from a trained model, scales as $\mathcal{O}(\exp(n + m))$ on a classical computer[31]. However, training costs can be significantly reduced by leveraging classical techniques such as contrastive divergence (CD)[9]. Analogously, given the similarity in the output probability expressions of sqRBM and RBM, a CD-based training algorithm for sqRBM may be developed as well. However, since sqRBM is defined through a non-commuting Hamiltonian, an efficient CD algorithm would still require access to a quantum computer. This could be achieved by quantum algorithms that can condition the visible or hidden subspaces of a Gibbs state on the quantum computer, similar to how the CD algorithm works in

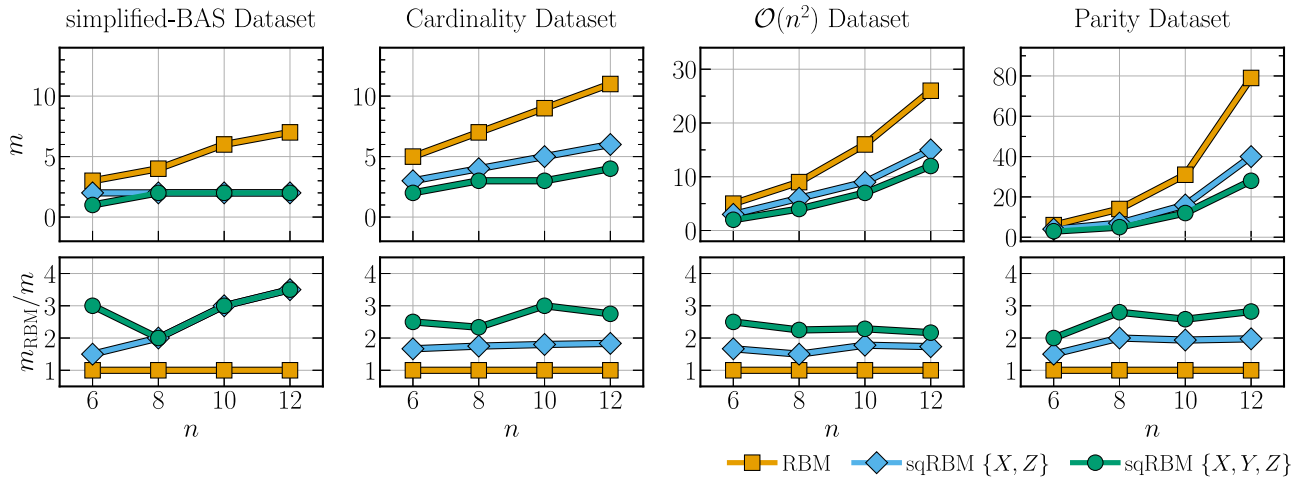

Fig. 3 | Minimum number of hidden units required to learn target probability distributions on average. We report the minimum number of hidden units m required to achieve total variation distance, TVD < 0.2 on average, over four datasets for various input sizes ($n \in \{6, 8, 10, 12\}$) using three models (RBM (orange square markers), sqRBM $\{X, Z\}$ (blue rotated square markers) and sqRBM$\{X, Y, Z\}$ (green circle markers)) as in Fig. 2. In the bottom panel we provide the ratio of m_{RBM} to m of the other models. The difficulty of each dataset results in a different scaling behavior. Recall that RBM has the same number of parameters and expressivity according to Theorem 2 as sqRBM $\{X, Z\}$ for the ratio $m_{\mathrm{RBM}} = 2m$ and similarly sqRBM$\{X, Y, Z\}$ for $m_{\mathrm{RBM}} = 3m$.

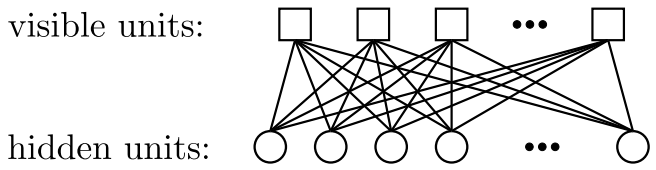

Fig. 4 | Connectivity graph of restricted Boltzmann machines (RBM). An RBM model has connections only between visible and hidden units. Lateral connections (e.g., visible to visible) are not permitted.

the classical setting. We leave the study of CD algorithms to train sqRBMs for future work.

While the training cost of both RBMs and sqRBMs can be reduced using approximate methods, sampling remains a major bottleneck. Several quantum algorithms have been proposed to prepare Gibbs states with polynomial complexity in the number of units[12,13,32], but these costs often involve high-degree scaling. This makes the number of hidden units a critical resource constraint in quantum implementations.

In this context, the reduced hidden-layer size of sqRBMs, enabled by their enhanced expressivity through non-commuting hidden interactions, translates into meaningful efficiency gains. As shown in our equivalence result, sqRBMs can achieve the same representational power as RBMs using only one-third as many hidden units. This leads to a nontrivial reduction in quantum resource requirements, including qubit count and circuit depth, which are key limiting factors in fault-tolerant settings.

More broadly, sqRBMs represent an intermediate regime between classical RBMs and fully quantum BMs. By maintaining a classical interface to the visible units while exploiting quantum structure in the hidden layer, sqRBMs may offer a more feasible route for early quantum generative modeling. These models could serve as effective testbeds for evaluating the capabilities of quantum Gibbs samplers, or as components in hybrid classical-quantum architectures for feature learning and structured data modeling.

One of the goals of quantum machine learning is to process both classical and *quantum* data efficiently. Although sqRBMs are quantum models, they only support classical data as input because their Hamiltonian is commuting within the subspace of visible units. Additionally, there are proposals in the literature for QBMs that are suitable for quantum data[33].

In future work, adding lateral connections between hidden units could be utilized to further enhance the expressivity of sqRBMs. To this end, we introduce the Definition 2 of sqBM that has additional connectivity within visible and hidden units. However, increased connectivity within hidden units may result in entanglement-induced barren plateaus[19]. Another promising extension to improve the expressivity of sqRBMs is to incorporate additional hidden layers, similar to deep Boltzmann machines[6]. More broadly, given that RBMs belong to the class of undirected graphical models, it is natural to conjecture that similar expressive equivalence results hold for other graphical models, such as quantum and classical Markov random fields (MRF)[34]. In particular, a partially observed MRF parameterized by Pauli-Z and Pauli-X terms may be equivalent to a classical MRF with more hidden variables, suggesting a pathway for future research on structured quantum models.

Methods

Model definitions

A Boltzmann machine (BM) can be described by a Hamiltonian H and its corresponding Gibbs state ρ. BMs consist of two types of units: visible and hidden. Visible units are the ones that are observed and used for input/output purposes, while hidden units form the latent dimension, giving the model its representation power. We denote the number of visible units with n and the number of hidden units with m.

The Hamiltonian that describes BMs can be written as a sum of three terms as follows:

$$H = H_{\mathrm{v}} + H_{\mathrm{h}} + H_{\mathrm{int}}, \tag{19}$$

where H_{v} and H_{h} act on visible and hidden units, respectively, while H_{int} represents the interaction between visible and hidden units. Consequently, the corresponding Gibbs state ρ of H is given as

$$\rho = e^{-\beta H}/\mathcal{Z}, \quad \mathcal{Z} = \mathrm{Tr}\,[e^{-\beta H}], \tag{20}$$

where $\mathcal{Z}$ is the partition function that ensures normalization ($\mathrm{Tr}[\rho] = 1$) and β is the inverse temperature, which we set as $\beta = 1$ to simplify subsequent equations. In this work, we focus on restricted BM configurations, allowing interactions only between visible and hidden units. We denote the number of visible units with n and the number of hidden units with m. A visualization of the restricted BM (RBM) configuration can be seen in Fig. 4.

The probability distribution of the model, denoted with p, can be obtained by marginalizing over hidden units, such that

$$p_v = \mathrm{Tr}\left[\Lambda_v \rho\right], \tag{21}$$

where $v \in \{0,1\}^n$ is a length n bitstring and Λ_v is a projective measurement with respect to the computational basis of the visible units given as

$$\Lambda_v = |v\rangle\langle v| \otimes \mathbb{1}_{2^m,2^m}. \tag{22}$$

Of particular interest, let us define a Pauli string of length $n+m$ as the tensor product of operators from the set of Pauli matrices, including the identity $\{I, X, Y, Z\}$ (see the Supplementary Note 2 for definitions). A k-body operator acts on k many qubits non-trivially, meaning it has k many non-identity operators in the Pauli string representation. We write the $n+m$-qubit Pauli string of the Pauli-Z operator acting on the i-th qubit as an example

$$\sigma_i^Z = \underbrace{I \otimes \cdots \otimes I}_{i-1} \otimes Z \otimes \underbrace{I \otimes \cdots \otimes I}_{n+m-i}. \tag{23}$$

We provide a formal definition of a classical restricted Boltzmann machine (RBM) as follows:

Definition 3. (Restricted Boltzmann machine (RBM)). A restricted Boltzmann machine with n visible and m hidden units, denoted as $\mathrm{RBM}_{n,m}$, is described by a parameterized Hamiltonian according to Eq. (19). The three terms of the Hamiltonian are defined as follows:

$$\begin{aligned} H_{\mathrm{v}} &= \sum_{i=1}^{n} a_i^Z \sigma_i^Z, \quad H_{\mathrm{h}} = \sum_{j=1}^{m} b_j^Z \sigma_{n+j}^Z, \\ H_{\mathrm{int}} &= \sum_{i=1}^{n}\sum_{j=1}^{m} w_{i,j}^{Z,Z} \sigma_i^Z \sigma_{n+j}^Z, \end{aligned} \tag{24}$$

where $\boldsymbol{a} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$ and $\boldsymbol{w} \in \mathbb{R}^{nm}$ are the parameter vectors of the model.

The Hamiltonian of an $\mathrm{RBM}_{n,m}$ corresponds to a classical Ising model and contains only commuting operators ($\forall(i,j)$, $\left[\sigma_i^Z, \sigma_i^Z \sigma_{n+j}^Z\right] = 0$). An $\mathrm{RBM}_{n,m}$ can be extended to a quantum model by incorporating non-commuting operators in the Hamiltonian. We refer to a generic model with a non-commuting Hamiltonian as a quantum restricted Boltzmann machine (QRBM).

Definition 4. (Quantum restricted Boltzmann machine (QRBM)). A quantum restricted Boltzmann machine with n visible and m hidden units, denoted as $\mathrm{QRBM}_{n,m}$, is described by a parameterized Hamiltonian according to Eq. (19). The three terms of the Hamiltonian are defined as follows:

$$\begin{aligned} H_{\mathrm{v}} &= \sum_{P \in \mathcal{W}_{\mathrm{v}}}^{|\mathcal{W}_{\mathrm{v}}|} \sum_{i=1}^{n} a_i^P \sigma_i^P, \quad H_{\mathrm{h}} = \sum_{P \in \mathcal{W}_{\mathrm{h}}}^{|\mathcal{W}_{\mathrm{h}}|} \sum_{j=1}^{m} b_j^P \sigma_{n+j}^P, \\ H_{\mathrm{int}} &= \sum_{(P,Q) \in \mathcal{W}_{\mathrm{int}}}^{|\mathcal{W}_{\mathrm{int}}|} \sum_{i=1}^{n}\sum_{j=1}^{m} w_{i,j}^{P,Q} \sigma_i^P \sigma_{n+j}^Q, \end{aligned} \tag{25}$$

where $\boldsymbol{a} \in \mathbb{R}^{|\mathcal{W}_{\mathrm{v}}|\cdot n}$, $\boldsymbol{b} \in \mathbb{R}^{|\mathcal{W}_{\mathrm{h}}|\cdot m}$ and $\boldsymbol{w} \in \mathbb{R}^{|\mathcal{W}_{\mathrm{int}}|\cdot n\cdot m}$ are the parameter vectors of the model. $\mathcal{W}_{\mathrm{v}}$, $\mathcal{W}_{\mathrm{h}}$ and $\mathcal{W}_{\mathrm{int}}$ are the sets of Pauli operators that describe the model and $|\mathcal{W}|$ denotes the cardinality of set $\mathcal{W}$.

A generic QRBM contains all possible one- and two-body Pauli operators when $\mathcal{W}_{\mathrm{v}} = \mathcal{W}_{\mathrm{h}} = \{X, Y, Z\}$ and $\mathcal{W}_{\mathrm{int}} = \mathcal{W}_{\mathrm{v}} \otimes \mathcal{W}_{\mathrm{h}}$. A common choice in the literature is the set that corresponds to the transverse field Ising model, such that $\mathcal{W}_{\mathrm{v}} = \mathcal{W}_{\mathrm{h}} = \{X, Z\}$ and $\mathcal{W}_{\mathrm{int}} = \{ZZ\}$[14]. Notice that the definition of QRBM also includes the RBM when $\mathcal{W}_{\mathrm{v}} = \mathcal{W}_{\mathrm{h}} = \{Z\}$ and $\mathcal{W}_{\mathrm{int}} = \{ZZ\}$.

Training Boltzmann machines

Boltzmann machines can be trained by minimizing the negative log-likelihood, which is defined as

$$\mathcal{L} = -\sum_v q_v \log p_v, \tag{26}$$

where p and q are the probability distributions of the model and the target, respectively ($\sum_v q_v = \sum_v p_v = 1$). Minimizing the negative log-likelihood is equivalent to minimizing D_{KL}, which is given as

$$\begin{aligned} D_{\mathrm{KL}}(q||p) &= \sum_v q_v \log\left(\frac{q_v}{p_v}\right) \\ &= -\sum_v q_v \log p_v + \sum_v q_v \log q_v. \end{aligned} \tag{27}$$

Then, gradients of both RBMs and QRBMs with respect to the negative log-likelihood can be obtained using a single formula that has two parts: positive phase and negative phase. The negative phase is obtained by measuring expectation values over the Gibbs state of the model, while the positive phase requires projective measurements on the visible unit subspace. Although these terms may look different in the classical machine learning literature, we keep the naming convention. The following proposition is a restatement of a result from Ref. 14.

Proposition 4. (Gradients of QRBM). A $\mathrm{QRBM}_{n,m}$ can be trained to minimize the negative log-likelihood with respect to the target probability distribution q using the following gradient rule:

$$\partial_{\theta_i}\mathcal{L} = -\sum_v q_v \left(\underbrace{\frac{\mathrm{Tr}[\Lambda_v \partial_{\theta_i} e^{-H}]}{\mathrm{Tr}[\Lambda_v e^{-H}]}}_{\text{positive phase}} - \underbrace{\frac{\mathrm{Tr}[\partial_{\theta_i} e^{-H}]}{\mathrm{Tr}[e^{-H}]}}_{\text{negative phase}} \right), \tag{28}$$

where $\theta_i \in \boldsymbol{\theta}$ is any real-valued parameter of the model, when the Hamiltonian terms are grouped such that $H = \sum_i \theta_i H_i$ and $\boldsymbol{\theta} \in \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{w}\}$.

The proof is provided in the Supplementary Note 7. In the case of a commuting Hamiltonian (e.g., RBM), the gradients of the negative log-likelihood with respect to the parameters take a fairly simple form. This follows the fact that if $[\partial_{\theta_i} H, H] = 0$, then $\partial_{\theta_i} e^{-H} = e^{-H}(-H_i)$, where $H_i = \partial_{\theta_i} H$ and $H = \sum_i \theta_i H_i$. Then, the gradients can be computed by measuring the expectation values of the Hamiltonian terms on the Gibbs state of the model. Although this appears relatively straightforward, preparing the Gibbs state is still exponentially expensive with respect to the system size $n+m$ on a classical computer. For this reason, in practice, alternative methods are often employed to avoid this step. One of the most popular approaches is called contrastive divergence (CD)[9].

For a generic non-commuting Hamiltonian, computing gradients becomes expensive, primarily because the derivative of the Hamiltonian does not commute with itself ($[\partial_{\theta_i} H, H] \neq 0$), requiring explicit computation of $\partial_{\theta_i} e^{-H}$. As a result, evaluating the positive phase is costly, even when the Gibbs state can be prepared efficiently. Consequently, a generic QRBM cannot be trained efficiently using standard gradient descent. In contrast, sqRBM overcomes this limitation, enabling a more tractable training process.

Data availability

The data that is generated in the numerical experiments of this study is available in Ref. 35.

Code availability

The code to reproduce results of this study is available in ref. 36.

Received: 14 April 2025; Accepted: 6 October 2025;
Published online: 29 October 2025

References

1. Smolensky, P. Information processing in dynamical systems: foundations of harmony theory. In *Proc. Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1: Foundations (MIT Press, 1986).
2. Ackley, D. H., Hinton, G. E. & Sejnowski, T. J. A learning algorithm for Boltzmann machines. *Cogn. Sci.* **9**, 147 (1985).
3. Le Roux, N. & Bengio, Y. Representational power of restricted Boltzmann machines and deep belief networks. *Neural Comput.* **20**, 1631 (2008).
4. Salakhutdinov, R., Mnih, A. & Hinton, G. Restricted Boltzmann machines for collaborative filtering. In *Proc. 24th International Conference on Machine Learning* 791–798 (ACM, 2007).
5. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504 (2006).
6. Salakhutdinov, R. & Hinton, G. Deep Boltzmann machines. In *Proc. Twelfth International Conference on Artificial Intelligence and Statistics* 448–455 (PMLR, 2009). https://proceedings.mlr.press/v5/salakhutdinov09a.html
7. Yichuan, T., Salakhutdinov, R. & Hinton, G. Robust Boltzmann Machines for recognition and denoising. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 2264–2271 (IEEE, 2012).
8. Long, P. M. & Servedio, R. A. Restricted Boltzmann machines are hard to approximately evaluate or simulate. In *Proc. 27th International Conference on International Conference on Machine Learning* (Omnipress, 2010).
9. Hinton, G. E. Training products of experts by minimizing contrastive divergence. *Neural Comput.* **14**, 1771–1800 (2002).
10. Carreira-Perpiñán, M. Á. & Hinton, G. On contrastive divergence learning. In *Proc. International Workshop on Artificial Intelligence and Statistics* (PMLR, 2005).
11. Bengio, Y. & Delalleau, O. Justifying and generalizing contrastive divergence. *Neural Comput.* **21**, 1601 (2009).
12. Salakhutdinov, R. & Hinton, G. An efficient learning procedure for deep Boltzmann machines. *Neural Comput.* **24**, 1967 (2012).
13. Chen, C.-F., Kastoryano, M. J., Brandão, F. G. S. L. & Gilyén, A. Quantum thermal state preparation https://arxiv.org/abs/2303.18224 arXiv:2303.18224 (2023).
14. Amin, M. H., Andriyash, E., Rolfe, J., Kulchytskyy, B. & Melko, R. Quantum Boltzmann machine. *Phys. Rev. X* **8**, 021050 (2018).
15. Dallaire-Demers, P.-L. & Killoran, N. Quantum generative adversarial networks. *Phys. Rev. A* **98**, 012324 (2018).
16. Benedetti, M. et al. A generative modeling approach for benchmarking and training shallow quantum circuits. *npj Quantum Inf.* **5**, 1 (2019).
17. McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R. & Neven, H. Barren plateaus in quantum neural network training landscapes. *Nat. Commun.* **9**, 4812 (2018).
18. Rudolph, M. S. et al. Trainability barriers and opportunities in quantum generative modeling. *npj Quantum Inf.* **10**, 116 (2024).
19. Ortiz Marrero, C., Kieferová, M. & Wiebe, N. Entanglement-induced barren plateaus. *PRX Quantum* **2**, 040316 (2021).
20. Coopmans, L. & Benedetti, M. On the sample complexity of quantum Boltzmann machine learning. *Commun. Phys.* **7**, 274 (2024).
21. Anschuetz, E. R. & Cao, Y. Realizing quantum Boltzmann machines through eigenstate thermalization https://arxiv.org/abs/1903.01359 arXiv (2019).
22. Kieferová, M. & Wiebe, N. Tomography and generative training with quantum Boltzmann machines. *Phys. Rev. A* **96**, 062327 (2017).
23. Zoufal, C., Lucchi, A. & Woerner, S. Variational quantum Boltzmann machines. *Quantum Mach. Intell.* **3**, 7 (2021).
24. Kappen, H. J. Learning quantum models from quantum or classical data. *J. Phys. A Math. Theor.* **53**, 214001 (2020).
25. Patel, D., Koch, D., Patel, S. & Wilde, M. M. Quantum Boltzmann machine learning of ground-state energies https://arxiv.org/abs/2410.12935 arXiv:2410.12935 (2024).
26. Tüysüz, C. et al. Learning to generate high-dimensional distributions with low-dimensional quantum Boltzmann machines https://arxiv.org/abs/2410.16363 arXiv:2410.16363 (2024).
27. Patel, D. & Wilde, M. M. Natural gradient and parameter estimation for quantum Boltzmann machines https://arxiv.org/abs/2410.24058 arXiv:2410.24058 (2024).
28. Minervini, M., Patel, D. & Wilde, M. M. Evolved quantum Boltzmann machines https://arxiv.org/abs/2501.03367 arXiv:2501.03367 (2025).
29. Montufar, G., Rauh, J. & Ay, N. Expressive power and approximation errors of restricted Boltzmann machines (2014).
30. Tran, P. T. & Phong, L. T. On the convergence proof of amsgrad and a new version. *IEEE Access* **7**, 61706 (2019).
31. Kappen, H. J. & Rodríguez, F. B. Efficient learning in Boltzmann machines using linear response theory. *Neural Comput.* **10**, 1137 (1998).
32. Kappen, H. J. & Rodríguez, F. B. Boltzmann machine learning using mean field theory and linear response correction. In *Proc. 11th International Conference on Neural Information Processing Systems*, NIPS'97 280–286 (MIT Press, Cambridge, 1997).
33. Wiebe, N. & Wossnig, L. Generative training of quantum Boltzmann machines with hidden units, https://arxiv.org/abs/1905.09902 arXiv:1905.09902 (2019).
34. Piatkowski, N. & Zoufal, C. Quantum circuits for discrete graphical models. *Quantum Mach. Intell.* **6**, 37 (2024).
35. Data repository (2025)
36. Github repository (2025)

Acknowledgements

C.T. is supported in part by the Helmholtz Association -"Innopool Project Variational Quantum Computer Simulations (VQCS)". This work is supported with funds from the Ministry of Science, Research, and Culture of the State of Brandenburg within the Centre for Quantum Technologies and Applications (CQTA). This work is funded within the framework of QUEST by the European Union's Horizon Europe Framework Programme (HORIZON) under the ERA Chair scheme with grant agreement No. 101087126. Parts of this research have been funded by the Federal Ministry of Education and Research of Germany and the state of North Rhine-Westphalia as part of the Lamarr-Institute for Machine Learning and Artificial Intelligence. C.T. and M.G. are supported by CERN through the CERN Quantum Technology Initiative.

Author contributions

M.D. conceived the study and designed the research. M.D. and C.T. developed the theoretical framework. M.D. conducted the numerical simulations. M.D. and C.T. analyzed the data. M.D. and C.T. wrote the manuscript with input from N.P., M.G. and K.J.. M.D., C.T., N.P., M.G. and K.J. contributed to discussions, provided feedback, and reviewed the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-025-02353-1.

Correspondence and requests for materials should be addressed to Maria Demidik.

Peer review information *Communications Physics* thanks Luuk Coopmans and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. [A peer review file is available.]