



Batched quantum state exponentiation and quantum Hebbian learning

Thomas R. Bromley¹ · Patrick Rebentrost¹

Received: 28 November 2018 / Accepted: 28 February 2019 / Published online: 15 May 2019
© Springer Nature Switzerland AG 2019

Abstract

Machine learning is a crucial aspect of artificial intelligence. This paper details an approach for quantum Hebbian learning through a batched version of quantum state exponentiation. Here, batches of quantum data are interacted with learning and processing quantum bits (qubits) by a series of elementary controlled partial swap operations, resulting in a Hamiltonian simulation of the statistical ensemble of the data. We decompose this elementary operation into one and two qubit quantum gates from the Clifford+ T set and use the decomposition to perform an efficiency analysis. Our construction of quantum Hebbian learning is motivated by extension from the established classical approach, and it can be used to find details about the data such as eigenvalues through phase estimation. This work contributes to the near-term development and implementation of quantum machine learning techniques.

Keywords Quantum learning · Quantum state exponentiation

1 Introduction

Machine learning encompasses a series of techniques that allow computers to solve problems without explicitly telling them how to do so (MacKay 2003). In supervised learning, the machine is first taught to solve the problem on a series of training data. This learning stage is a crucial element in determining the performance of a machine learning algorithm. One particularly fruitful machine learning technique is to construct an artificial neural network, represented by an interacting collection of binary valued neurons. The applications of machine learning are numerous and include, for example, finance, biotechnology, e-commerce, chemistry, insurance, and security. In particular, neural networks have been successfully used in finance for portfolio analysis (Barr and Mani 1998) and credit approval (Norris 1999), as well as in e-commerce for user ratings of online stores (Lasa and Berndt 2007).

The Hebbian approach is the most natural learning method for neural networks that are fully visible and with undirected connections (Hebb 1949). Hebbian learning specifies the connection strength between neurons according to the number of times that they fire together within training data. The output of Hebbian learning is a real symmetric weighting matrix with zero diagonal, whose elements correspond to the connection weights between neuron pairs. This matrix is then used as a component within machine learning algorithms. Undirected and fully visible neural networks, such as the canonical Hopfield network (Hopfield 1982), are often utilized as an associative memory for pattern recognition, as well as for optimization problems such as the traveling salesman problem (MacKay 2003).

Quantum machine learning (QML) combines parts of quantum mechanics, such as quantum algorithms, with machine learning (Biamonte et al. 2017). It can be split into two broad categories based upon whether the input data is of a classical nature or quantum. In classical-input QML, the data is initially classical and must be converted into quantum states for processing by a quantum device. This approach promises potentially exponential increases in data handling capacity and processing speed, as well as feasible circuit size scaling. However, such advantages hinge on the fundamental question of whether classical data can be converted efficiently into quantum data (Aaronson 2015; Giovannetti et al. 2008; Soklakov and

✉ Thomas R. Bromley
tom@xanadu.ai

Patrick Rebentrost
pr@patrickre.com

¹ Xanadu, 777 Bay Street, Toronto, Ontario, M5G 2C8, Canada

Schack 2006). Alternatively, quantum-input QML assumes that the relevant data is already of a quantum nature, taking the perspective that the QML algorithm is an element in the toolchain of a composite quantum device. Although perhaps less studied at present, quantum-input QML holds great potential given the predicted development of quantum devices over the coming years.

Nevertheless, QML algorithms can typically function in both the classical and quantum input regimes. A variety of new approaches has already been discovered, including for anomaly detection (Liu and Rebentrost 2018), data fitting (Wiebe et al. 2012), and support vector machines (Rebentrost et al. 2014). Recent works on quantum neural networks (Schuld et al. 2014) include the use of amplitude amplification (Wiebe et al. 2014), quantum annealing (Amin et al. 2018), Helmholtz machines (Benedetti et al. 2018), the alternating operator ansatz (Verdon et al. 2017), and parameterized unitaries (Farhi and Neven 2018).

Learning makes up an important stage of any QML algorithm. In this work, we detail a new controlled method of quantum state exponentiation (QSE) using batches of quantum data and show how this method can form the quantum analogue of Hebbian learning. Here, a quantum Hebbian learning (QHL) device is composed of a series of processing qubits and an ancilla learning qubit, with training data inputted from a register of data qubits. By repetitively performing partial swaps between the processing and data qubits, with control on the learning qubit, a QSE of a mixed state representing the Hebbian learning matrix can be enacted on the processing qubits. We begin by outlining our approach, called batched controlled QSE (bcQSE), in Section 2. We next breakdown the controlled partial swap operation into standard one and two qubit gates from the Clifford+ T set in Section 3, allowing us to perform a gate count as well as an analysis of errors and efficiency of bcQSE. QHL and its realization through bcQSE is formalized in Section 4. Our findings represent important precursor steps in the development of a concrete QHL device. We then conclude in Section 5 with a discussion.

2 Quantum state exponentiation

2.1 Established methods

We first outline the established results on QSE. Given any (possibly unknown) quantum state ρ , one can transform another system according to the unitary $e^{-it\rho}$ for some time t using the protocol outlined in the following theorem.

Theorem 1 (QSE) *Let ρ be a density matrix of N qubits. By partial swapping an N qubit quantum system with a*

single copy of ρ , one can enact $e^{-i\Delta t\rho}$ on the system to error $\mathcal{O}(\Delta t^2)$ in diamond norm. By repetitively partial swapping an N qubit quantum system with a number of copies $n \geq 1$ of ρ that scales as $n = \mathcal{O}(t^2/\epsilon)$, one can enact $e^{-it\rho}$ on the system to error $\epsilon > 0$ in diamond norm. This protocol is optimal in n in terms of the t and ϵ .

This theorem is proved in Refs. Lloyd et al. (2014), Kimmel et al. (2017), and Marvian and Lloyd (2016). Here, the partial swap operation is $e^{-i\Delta t\mathcal{S}}$ for some time $\Delta t = t/n$, with \mathcal{S} the swap operation. The diamond norm $\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond := \max_\rho \|[I \otimes (\mathcal{E}_1 - \mathcal{E}_2)]\rho\|_1$ quantifies the one-shot distinguishability between two channels \mathcal{E}_1 and \mathcal{E}_2 when applied to input states ρ of arbitrary dimension, where $\|\cdot\|_1$ is the trace norm (Benenti and Strini 2010). A pure-state version of QSE was highlighted in the context of emulating unitaries in Ref. Marvian and Lloyd (2016).

Importantly, QSE can be extended to a controlled version dependent upon an ancillary control qubit (Lloyd et al. 2014; Kimmel et al. 2017; Marvian and Lloyd 2016).

Theorem 2 (Controlled QSE) *Let ρ be a density matrix over N qubits. By partial swapping an N qubit quantum system with a single copy of ρ , one can enact $e^{-it|1\rangle\langle 1|\otimes\rho}$ on the system to error $\mathcal{O}(\Delta t^2)$ in diamond norm. By repetitively partial controlled swapping an N qubit quantum system with a number of copies $n \geq 1$ of ρ that scales as $n = \mathcal{O}(t^2/\epsilon)$, one can enact $e^{-it|1\rangle\langle 1|\otimes\rho}$ on the system to error ϵ in diamond norm. This protocol is optimal in n in terms of the t and ϵ .*

See Refs. Lloyd et al. (2014), Kimmel et al. (2017), and Marvian and Lloyd (2016) for the proof. Here, $e^{-i\Delta t|1\rangle\langle 1|\otimes\mathcal{S}}$ is the partial controlled swap operation.

2.2 Batching

We now specify our batched approach to QSE within the context of quantum data. Consider a batch of M pieces of N qubit quantum data, each represented by a $d = 2^N$ -dimensional pure quantum state $|x^{(m)}\rangle$. A statistical ensemble of this quantum data is given by the N qubit mixed state

$$\rho = \frac{1}{M} \sum_{m=1}^M |x^{(m)}\rangle\langle x^{(m)}|. \quad (1)$$

We shall see in the following that ρ is an important object, containing all the relevant information required to capture the Hebbian learning matrix. The following theorem shows that $e^{-it\rho}$ can be enacted on a series of N processing qubits by repetitively partial swapping multiple batches of quantum data. A batch partial swap constitutes M partial

swaps between the processing qubits and collections of data supplying qubits, each in the ordered quantum states $\{|x^{(m)}\rangle\}_{m=1}^M$. We focus on the situation where there is additional control on an ancilla qubit, which we refer to as the learning qubit due the link with QHL discussed in Section 4, and hence call the process batched controlled QSE (bcQSE).

Theorem 3 (bcQSE) *Let $\{|x^{(m)}\rangle\}_{m=1}^M$ be a batch of M quantum states, each quantum state being of N qubits. By repetitively batch partial swapping an N qubit quantum system with a number of batches $n \geq 1$ of this quantum data that scales as $n = \mathcal{O}(t^2/\epsilon)$, one can enact $e^{-it|1\rangle\langle 1|\otimes\rho}$ on the system to error ϵ in diamond norm.*

Proof We discuss the single step before obtaining multiple steps via repeating the single step for a single batch of M states then subsequently for n batches. For a short time Δt , which later will be set to $\Delta t = t/n$, define the M unitaries

$$\mathcal{U}_m := |0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes e^{-i\frac{\Delta t}{M}|x^{(m)}\rangle\langle x^{(m)}|}. \quad (2)$$

From Theorem 2, a transformation \mathcal{W}_m on the processing qubits can be simulated that approximates \mathcal{U}_m by partial swapping one copy of $|x^{(m)}\rangle$ from data supplying qubits. The error is $\mathcal{O}(\Delta t^2/M^2)$ in diamond norm.

For a single batch, we perform M partial swaps between processing and data supplying qubits, with the data supplying qubits sequentially prepared in the states $|x^{(1)}\rangle, |x^{(2)}\rangle, \dots, |x^{(M)}\rangle$. This simulates the sequence of unitaries $\mathcal{U}_M \dots \mathcal{U}_1$ approximately via the sequence $\mathcal{W}_M \dots \mathcal{W}_1$. The error for simulating $\mathcal{U}_M \dots \mathcal{U}_1$ compounds M times the single step error and hence is $\mathcal{O}(\Delta t^2/M)$. A single batch takes M controlled partial swaps (for a shorter time $\Delta t/M$) and M qubit registers to encode all the pure training states.

Repeating the process for $n \geq 1$ batches simulates the sequence of unitaries $(\mathcal{U}_M \dots \mathcal{U}_1)^n$. The error compounds n times to be $\mathcal{O}(n\Delta t^2/M)$. Replacing $\Delta t = t/n$ leads to an error of $\mathcal{O}(t^2/nM)$. Our target is for the partial swap steps $(\mathcal{W}_M \dots \mathcal{W}_1)^n$ to approximate $e^{-it|1\rangle\langle 1|\otimes\rho}$. The error is

$$\begin{aligned} \epsilon &:= \left\| (\mathcal{W}_M \dots \mathcal{W}_1)^n - e^{-it|1\rangle\langle 1|\otimes\rho} \right\|_{\diamond} \\ &\leq \left\| (\mathcal{W}_M \dots \mathcal{W}_1)^n - (\mathcal{U}_M \dots \mathcal{U}_1)^n \right\|_{\diamond} \\ &\quad + \left\| (\mathcal{U}_M \dots \mathcal{U}_1)^n - e^{-it|1\rangle\langle 1|\otimes\rho} \right\|_{\diamond} \\ &= \mathcal{O}\left(\frac{t^2}{nM}\right) + \mathcal{O}\left(\frac{t^2}{n}\right) = \mathcal{O}\left(\frac{t^2}{n}\right). \end{aligned} \quad (3)$$

Here, we have used the triangle inequality and, in the last step, the Lie product formula (Childs et al. 2018) to approximate $e^{-it|1\rangle\langle 1|\otimes\rho}$ with $(\mathcal{U}_M \dots \mathcal{U}_1)^n$ with the error $\mathcal{O}(\frac{t^2}{n})$. The overall error compounds the two sources and is hence $\epsilon = \mathcal{O}(t^2/n)$. \square

Figure 1 illustrates the bcQSE protocol. It requires nM collections of N data supplying qubits that are supplied in n batches of the M sequential data states. This approach is preferable when the user has access to multiple copies of the quantum data in pure state form, rather than the statistical ensemble. In the next section, we break down the controlled partial swap into standard one and two qubit gates from the Clifford+ T set (Nielsen and Chuang 2002). Being the elementary transformation of both batched and non-batched QSE, this decomposition makes more concrete the implementation of QSE. Focusing on bcQSE, we then perform an analysis of errors and efficiency.

3 Gate decomposition

The Clifford set of gates consists of the Hadamard gate H , the phase gate S , the global phase gate W , and the CNOT gate. Any quantum circuit consisting of only these gates can be simulated efficiently on a classical computer (Nielsen and Chuang 2002). The conventional approach to extending beyond classical simulability is to add the $\pi/8$ gate T . The matrix representations of these gates with respect to the computational basis are:

$$\begin{aligned} H &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \\ T &= \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{pmatrix}, \quad W = \begin{pmatrix} e^{i\frac{\pi}{4}} & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{pmatrix}, \\ \text{CNOT} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \end{aligned}$$

The ability to perform Clifford+ T gates is established for a variety of physical implementations of quantum circuits, such as those using superconducting qubits, trapped ions or nuclear magnetic resonance.

The controlled partial swap operation used in Fig. 1 between N processing qubits and N data supplying qubits for a time $\frac{\Delta t}{M}$, with control on the learning qubit, can be first decomposed into two and three qubit gates following similar steps to Ref. Childs et al. (2003). The result is presented in Fig. 2. Here, we label the N processing qubits as a_1, a_2, \dots, a_N and the N data supplying qubits as b_1, b_2, \dots, b_N (although the choice of labeling is symmetric). The next step is to further decompose these two and three qubit gates into gates from the Clifford+ T set. In Fig. 3, we show such decompositions for (a) the controlled $e^{-i\frac{\pi}{4}Y}$ unitary, (b) the Toffoli gate, and (c) the time-dependent controlled $e^{-i\frac{\Delta t}{M}Z}$ unitary, where Y and Z

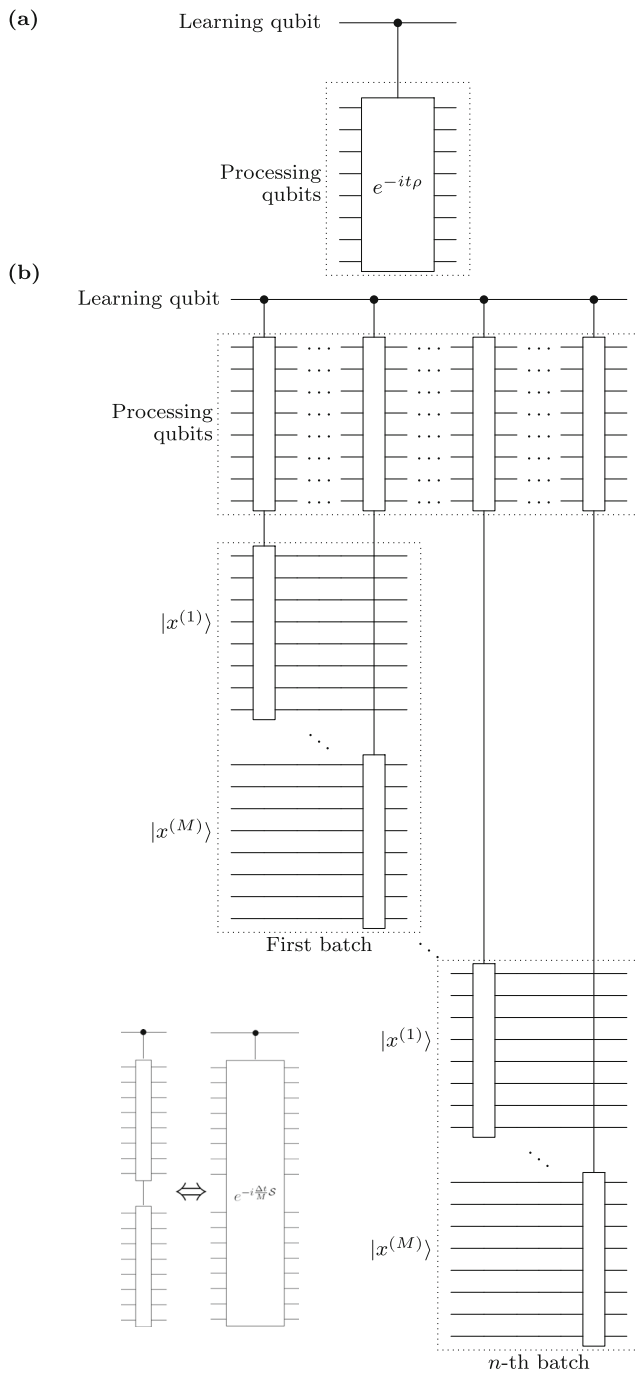


Fig. 1 **a** Quantum state exponentiation (QSE) of a state ρ onto a series of processing qubits, with control from an ancilla learning qubit. **b** Batched controlled quantum state exponentiation (bcQSE) of a batch of quantum states. Quantum data is organized into n batches of M pure states. The first collection of N qubits is prepared in $|x^{(1)}\rangle$, then the next collection in $|x^{(2)}\rangle$, and so forth until the M -th collection which is in $|x^{(M)}\rangle$. This represents the first batch, and the process is repeated n times

are qubit Pauli gates. In Fig. 3c, we show the controlled $e^{-i\frac{\Delta t}{M}Z}$ as a composition of two CNOT gates and the two single qubit unitaries $e^{\pm i\frac{\Delta t}{2M}Z}$, which can be approximately

decomposed into Clifford+ T gates using a technique given in Ref. Ross and Selinger (2015) (see the following for further details).

We now count the total number of Clifford+ T gates required to perform bcQSE. The gate count number is recorded in the vector $(H, S, W, \text{CNOT}, T)$. Herein, we assume that the inverse gates T^\dagger and S^\dagger have the same overhead as the corresponding T and S gates. Let us first consider the time-dependent qubit unitaries $e^{\pm i\frac{\Delta t}{2M}Z}$, which can be reduced using the following result (Ross and Selinger 2015).

Result 1 A unitary $e^{-i\tau Z}$ for any time τ can be approximated to error η in operator norm using a probabilistically generated sequence of H , S , W , and T gates with an approximate gate count number $(3g_\eta, 2g_\eta, g, 0, 3g_\eta)$, where $g_\eta := \tilde{O}(\log_2(1/\eta))$ and $g \approx 10$.

This result can be confirmed numerically using an algorithm developed by Ross and Selinger (2015), who also showed rigorously that the number of required T gates scales as $\tilde{O}(\log(\frac{1}{\eta}))$, with $\tilde{O}(\cdot)$ indicating the presence of more slowly growing terms in η . Note that one can typically have a lower gate count requirement by selecting the best output from multiple runs of the algorithm. Hence, the two $e^{\pm i\frac{\Delta t}{2M}Z}$ gates can be performed to a combined error 2η using approximately $(6g_\eta, 4g_\eta, 2g, 0, 6g_\eta)$ gates.

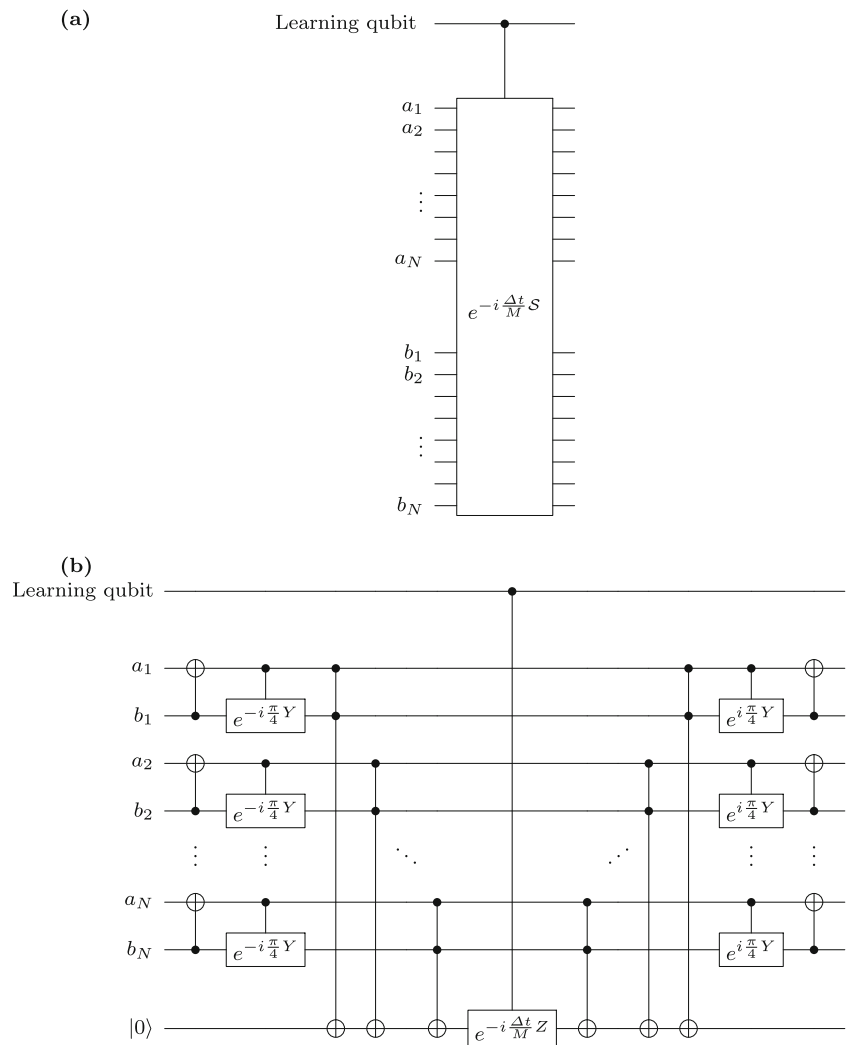
This result can be combined with the decomposition in Figs. 2 and 3 to arrive at the following result.

Theorem 4 (Controlled partial swap gate count) Consider two N qubit systems along with a single control qubit as well as an ancillary qubit prepared in $|0\rangle$. Assuming error-free Clifford+ T gates, there exists a quantum circuit performing the controlled partial swap operation $e^{-i\frac{\Delta t}{M}S}$ to error 2η in operator norm using a gate count of

$$\mathbf{g}(\eta) = (12N + 6g_\eta, 10N + 4g_\eta, 2g, 18N + 2, 18N + 6g_\eta). \quad (4)$$

Proof The controlled partial swap in Fig. 2a is broken down into Clifford+ T gates in Fig. 2b, which is adapted from Ref. Childs et al. (2003), and Fig. 3. The controlled unitary $e^{-i\frac{\pi}{4}Y}$ has a gate count of $(4, 4, 0, 2, 2)$ and is applied $2N$ times in Fig. 2b (N of those times being the conjugate gate, with the same gate count). The Toffoli gate has a gate count of $(2, 1, 0, 6, 7)$ and is also applied $2N$ times. Next, the controlled $e^{-i\frac{\Delta t}{M}Z}$ unitary has a gate count of $(6g_\eta, 4g_\eta, 2g, 2, 6g_\eta)$ using Result 1 and is applied once. Finally, there are $2N$ additional CNOT gates in Fig. 2b. Combining all of these elements results in a gate count of $\mathbf{g}(\eta) = (12N + 6g_\eta, 10N + 4g_\eta, 2g, 18N + 2, 18N + 6g_\eta)$.

Fig. 2 The controlled partial swap operation (a) between the N qubit collection a_1, a_2, \dots, a_N and the N qubit collection b_1, b_2, \dots, b_N , with learning qubit control, can be decomposed into two and three qubit quantum gates according to (b), following steps from Ref. Childs et al. (2003). Here, an additional ancilla qubit initialized in the state $|0\rangle$ is utilized to perform the transformation. The symbols used follow standard conventions (Nielsen and Chuang 2002)



As all Clifford+ T gates are assumed to be error-free, the only source of error is in approximating the controlled $e^{-i \frac{\Delta t}{M} Z}$. This results in an overall error of 2η . \square

We can now perform a gate count of bcQSE and provide a detailed account of the errors arising from both the controlled partial swap and bcQSE. Errors can arise from two sources: (a) Clifford+ T gate errors and (b) unitary simulation errors. We denote the error (in operator norm) arising from a single application of each Clifford+ T gate by the vector $\epsilon_g := (\epsilon_H, \epsilon_S, \epsilon_W, \epsilon_{\text{CNOT}}, \epsilon_T)$. Theorem 4 then says that the controlled partial swap can be performed with imperfect gates to an error $\epsilon_g \cdot \mathbf{g}(\eta) + 2\eta$, combining errors both from the elementary gates and the simulations of $e^{\pm i \frac{\Delta t}{2M} Z}$. We now consider the overall error in bcQSE.

Theorem 5 (bcQSE gate count and error) *Consider N processing qubits, a single learning qubit, nM collections of N data supplying qubits and nM ancillary qubits in state $|0\rangle$. There exists a quantum circuit of controlled*

partial swap operations which performs bcQSE to error 2η in operator norm using a gate count of $nM\mathbf{g}(\eta)$. Using imperfect Clifford+ T gates with errors $\epsilon_g = (\epsilon_H, \epsilon_S, \epsilon_W, \epsilon_{\text{CNOT}}, \epsilon_T)$, such bcQSE is performed to an error in the operator norm of

$$\epsilon = \alpha \frac{t^2}{n} + nM (\epsilon_g \cdot \mathbf{g}(\eta) + 2\eta) \quad (5)$$

for some constant $\alpha > 0$.

Proof This result holds since one needs to perform nM controlled partial swaps to realize the bcQSE in Theorem 3. A single application of the controlled partial swap operation can be carried out with an error in operator norm of $\epsilon_g \cdot \mathbf{g}(\eta) + 2\eta$. This operation must be repeated nM times in bcQSE. Theorem 3 tells us that $\alpha \frac{t^2}{n}$ additional error (in operator norm, which is upper bounded by the diamond norm) arises due to approximating controlled QSE with the batched process, with $\alpha > 0$ a constant factor determined by ρ . \square

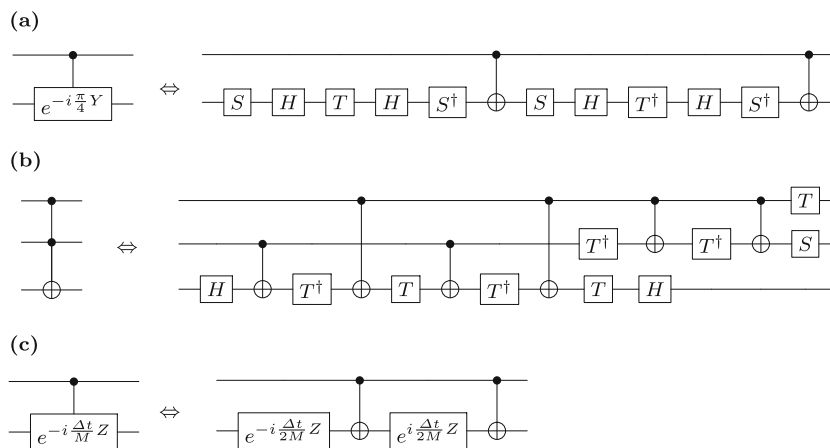


Fig. 3 A decomposition of the two and three qubit gates in Fig. 2b into standard one and two qubit gates from the Clifford+ T set for **a** the controlled unitary $e^{-i\frac{\pi}{4}Y}$ and **b** the Toffoli gate, as given in (Nielsen and Chuang 2002). The inverse gates T^\dagger and S^\dagger are assumed to be part of the Clifford+ T set and that $e^{i\frac{\pi}{4}Y}$ can be realized by simply swapping

the positions of T and T^\dagger in (a). In (c), the time-dependent controlled unitary $e^{-i\frac{\Delta t}{M}Z}$ is decomposed into two CNOT gates and two single qubit unitaries $e^{\pm i\frac{\Delta t}{2M}Z}$. We discuss in the main text how these unitaries can be approximately decomposed into Clifford+ T gates (Ross and Selinger 2015)

Performing bcQSE with imperfect Clifford+ T gates means that the error cannot be brought arbitrarily close to zero simply by increasing the number of batches n . Indeed, Eq. 5 shows that the dominant error for large n stems from the need to perform a large number of imperfect controlled partial swap operations. One can find an optimal regime by using a specific n to minimize the resultant error. For a fixed choice of η , one can find the optimal n via $\partial\epsilon/\partial n = 0$, which results in

$$n_{\text{opt}} = t \sqrt{\frac{\alpha}{M(\epsilon_g \cdot \mathbf{g}(\eta) + 2\eta)}}. \quad (6)$$

with the constraint that $n_{\text{opt}} \geq 1$. If the gate error $M(\epsilon_g \cdot \mathbf{g}(\eta) + 2\eta)$ is too large, then the optimized n_{opt} is pushed below 1 and the user must fix $n = 1$. In this scenario, batched quantum state exponentiation fails. Practically, this constraint places limitations on the batch size M and the individual gate errors for successful bcQSE.

Note that we find $n_{\text{opt}} = \mathcal{O}(t/\sqrt{MN})$ since α , ϵ_g , and η are constants and we have the lower bound $\mathbf{g}(\eta) = \Omega(N)$ from Eq. 4. As $N, M \geq 1$, we can conveniently use $n_{\text{opt}} = \mathcal{O}(t)$, keeping in mind that $n_{\text{opt}} \geq 1$. Using Eq. 6 in Eq. 5 results in the error $\epsilon = 2t\sqrt{\alpha M(\epsilon_g \cdot \mathbf{g}(\eta) + 2\eta)} = \mathcal{O}(t\sqrt{MN})$. This error is independent of the number of discretization steps taken and is the best one can do in the presence of constant gate errors. The user can also vary η to choose a satisfactory compromise between the number of batches (which influences the gate count) and the overall error.

However, using quantum error correction the user may be free to change the Clifford+ T gate errors at a cost

of increased physical resources, e.g., by combining ancilla qubits and measurements with imperfect physical gates to perform a low-error logical gate (Nielsen and Chuang 2002). We can then represent the additional gate resources required to enact the Clifford+ T gates to an error ϵ_g by the 5×5 matrix $G_{\text{ec}}(\epsilon_g)$, with the obtainable error being arbitrarily small provided that the physical gates have an error below a constant threshold. Each column of $G_{\text{ec}}(\epsilon_g)$ is set to contain the physical gates required to perform one of the elementary logical gates to the corresponding error in ϵ_g . For example, the user can decrease a logical CNOT gate error ϵ_{CNOT} using a quantum error-correcting code that requires multiple noisy CNOTs along with other physical gates. The fourth column of $G_{\text{ec}}(\epsilon_g)$ then determines the number of basic gates required for such a logical CNOT. We also denote the additional qubits required by $\mathbf{q}_{\text{ec}}(\epsilon_g) := (q_H(\epsilon_H), q_S(\epsilon_S), q_W(\epsilon_W), q_{\text{CNOT}}(\epsilon_{\text{CNOT}}), q_T(\epsilon_T))$.

Recall that the error 2η associated with approximating the two time-dependent gates $e^{\pm i\frac{\Delta t}{2M}Z}$ can be freely chosen. We know from Ross and Selinger (2015) that the cost in terms of T gates to perform a time-dependent Z unitary to error η is $\tilde{\mathcal{O}}(\log(1/\eta))$. Using quantum error correction methods, it is reasonable to assume that for the elementary gates themselves, every bit in precision also takes a linear cost to achieve, i.e., $G_{\text{ec}}(\epsilon_g) = \mathcal{O}(\log(\frac{1}{\epsilon_g}))$ and $\mathbf{q}_{\text{ec}}(\epsilon_g) = \mathcal{O}(\log(\frac{1}{\epsilon_g}))$, both element wise (Gottesman 2009). This setting allows for controlling the overall error ϵ of performing bcQSE. We now account for the qubit and gate count in this case.

Theorem 6 (Performing bcQSE to arbitrary error) *Let $\epsilon > 0$ be the desired accuracy of bcQSE for M quantum*

states per batch and N qubits encoding every vector in a batch. Assume Clifford+ T gate errors ϵ_g , which take additional resources in terms of gates $G_{ec}(\epsilon_g) = \mathcal{O}\left(\log\left(\frac{1}{\epsilon_g}\right)\right)$ and qubits $q_{ec}(\epsilon_g) = \mathcal{O}\left(\log\left(\frac{1}{\epsilon_g}\right)\right)$ to achieve via quantum error correction. The number of batches required to perform bcQSE can be set to $n = \mathcal{O}\left(\frac{t^2+1}{\epsilon}\right)$, where $n \geq 1$, requiring per single partial swap a number of physical gates and physical qubits of both $\mathcal{O}((N + \log(nM)) \log(nM(N + \log nM))) = \tilde{\mathcal{O}}(N \log(nMN))$.

Proof Take a constant $\delta' > 0$ and set $\eta = \frac{\delta'}{n^2 M}$. From Theorem 4, this implies that $\mathbf{g}\left(\frac{\delta'}{n^2 M}\right) = \mathcal{O}(N + \log(nM))$. With another constant $\delta'' > 0$, take $\epsilon_g = \delta'' / (n^2 M \mathbf{g}\left(\frac{\delta'}{n^2 M}\right))$, where the inverse of \mathbf{g} is defined element wise. With $\mathbf{1}$ the vector of ones, this implies that the gate count of a single partial swap is given by $\mathbf{1}^T G_{ec}(\epsilon_g) \mathbf{g}\left(\frac{\delta'}{n^2 M}\right) = \mathcal{O}\left(\log\left(\frac{1}{\epsilon_g}\right) (N + \log(nM))\right) = \mathcal{O}((N + \log(nM)) \log(nM(N + \log nM)))$ and the number of error correction qubits is $q_{ec}^T(\epsilon_g) \mathbf{g}\left(\frac{\delta'}{n^2 M}\right) = \mathcal{O}((N + \log(nM)) \log(nM(N + \log nM)))$. With $\delta = 2\delta' + 5\delta''$ and $\alpha > 0$, we then have from Theorem 5 for the total error that $\epsilon = (\alpha t^2 + \delta)/n$, which implies $n = \mathcal{O}\left(\frac{t^2+1}{\epsilon}\right)$. \square

It is important to consider the number of physical qubits and the number of physical gates for any implementation of a quantum device, along with the resultant error. For bcQSE, one can consider how these quantities scale as a function of the number of qubits N in a piece of quantum data, the number of quantum data states M , and the desired simulation time t . This can be achieved by first considering the number of batches n required to perform bcQSE. We have discussed two regimes for n , for which always $n \geq 1$. They are, (i) when the overall error ϵ is partly determined by the fixed-error Clifford+ T gates, so that $n = \mathcal{O}(t)$ according to Eq. 6, and (ii) when the overall error ϵ can be determined by the user by altering the individual errors in the Clifford+ T gates via quantum error correction, so that $n = \mathcal{O}\left(\frac{t^2+1}{\epsilon}\right)$.

The total number of logical qubits required to perform bcQSE is $(nM + 1)(N + 1) = \mathcal{O}(nMN)$. Indeed, we require N processing qubits and a single learning qubit, which receive nM partial swaps requiring nMN qubits along with nM ancilla qubits in $|0\rangle$. On the other hand, we have seen in Theorem 5 that the total number of gates required for bcQSE is $nM \mathbf{g}(\eta) = \mathcal{O}(nMN)$. In the error-corrected setting, each logical qubit is replaced by multiple physical qubits and additional physical gates are required to implement the logical gates. The scaling of the relevant

quantities is summarized in Table 1. In each case, scaling with N , M , and t , along with ϵ for regime (ii), is never worse than polynomial, indicating efficiency. A physical implementation can, in principle, handle exponentially large vectors without hitting a “brick wall” of efficiency. The second regime using quantum error correction is more costly than the first regime because the user must pay a qubit and gate cost to control the error.

4 Application to quantum Hebbian learning

Controlled QSE allows a quantum state ρ to be simulated on another system for a chosen time t , i.e., for application of the unitary transformation $e^{-it|1\rangle\langle 1| \otimes \rho}$. This transformation can provide a way to use ρ operationally within other quantum algorithms. For example, one can perform matrix multiplication with ρ (Wiebe et al. 2014), or find its eigenvalues using quantum phase estimation (Harrow et al. 2009). Our protocol for bcQSE is suited to the case where there is access to the pure quantum states composing the ensemble. Indeed, we show in the following how bcQSE can be thought of as a prototype quantum version of Hebbian learning.

4.1 Classical Hebbian learning

We first introduce the conventional approach to Hebbian learning to help motivate our following definition of quantum Hebbian learning. Consider an artificial neural network consisting of d binary-valued neurons $x_i \in \{1, -1\}$ with $i \in \{1, 2, \dots, d\}$ (McCulloch and Pitts 1943) that are together described by the activation pattern vector $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$. Each neuron is visible and every pair of neurons can be connected with an undirected weight. Suppose that we are supplied with M activation patterns $\mathbf{x}^{(m)}$ that consist the training data, which may be, e.g., the pixel data of images. These activation patterns are summarized by the Hebbian weight matrix.

Table 1 Efficiency analysis of performing bcQSE using Clifford+ T gates for relevant figures of merit in terms of the number of processing qubits N , the number of data states M , the simulation time t , and the error ϵ

Quantity	Gate error regime	
	(i) Fixed	(ii) Error corrected
Error	$\mathcal{O}(t\sqrt{MN})$	ϵ (user specified)
$n \geq 1$	$\mathcal{O}(t)$	$\mathcal{O}\left(\frac{t^2+1}{\epsilon}\right)$
Qubit number	$\mathcal{O}(tMN)$	$\tilde{\mathcal{O}}\left(\frac{t^2+1}{\epsilon} MN \log\left(\frac{t^2+1}{\epsilon} MN\right)\right)$
Gate count	$\mathcal{O}(tMN)$	$\tilde{\mathcal{O}}\left(\frac{t^2+1}{\epsilon} MN \log\left(\frac{t^2+1}{\epsilon} MN\right)\right)$

Definition 1 The normalized Hebbian weight matrix W is defined as $W_{ij} = \frac{1}{Md} \sum_{m=1}^M x_i^{(m)} x_j^{(m)}$ for $i \neq j$ and $W_{ii} = 0$.

Hence, W is a square d -dimensional, real, and symmetric matrix. It encodes the simple principle from Hebb that neurons that fire together (i.e., are both +1 or both -1 at the same time) also wire together (Hebb 1949; MacKay 2003). The normalization of W is such that the operator norm satisfies $\|W\| = \mathcal{O}(1)$. More generally, we do not have to restrict the training data $\mathbf{x}^{(m)}$ to be activation patterns of neural networks, but can instead allow them to be arbitrary vectors corresponding to relevant data.

We can then define Hebbian learning.

Definition 2 Hebbian learning is the process of:

1. Constructing the Hebbian weight matrix W from the set of training data $\mathbf{x}^{(m)}$.
2. Using the Hebbian weight matrix within the operation of a machine learning algorithm.

4.2 Quantum Hebbian learning

Quantum Hebbian learning (QHL) can be thought of as the quantum analogy to Hebbian learning.

Definition 3 We define quantum Hebbian learning to include any quantum algorithm that performs the following:

1. Provides a quantum analogue representation of the Hebbian weight matrix W .
2. Allows access to this quantum representation and makes the contained data operationally available in a quantum algorithm.

In addition, we call a quantum Hebbian learning method dimension efficient if (1) and (2) require $\mathcal{O}(\text{poly}(\log(d)))$ quantum bits and operations. We call a quantum Hebbian learning method data efficient if (1) and (2) require $\mathcal{O}(\text{poly}(\log(M)))$ quantum bits and operations.

We note that classical Hebbian learning in its simple formulation requires $\mathcal{O}(\text{poly}(d, M))$ operations as every neuron and every training example is processed at least once. Hence, when quantum Hebbian learning is dimension efficient, it is exponentially faster than the naive/simple classical counterpart. However, we additionally note that classical sampling techniques such as those discussed in recent quantum-inspired algorithms (Tang 2018; Gilyén et al. 2017) may also lead to enhanced performance of classical Hebbian learning.

The link to bcQSE can now be elucidated by considering learning with generic quantum training data $\{|\mathbf{x}^{(m)}\rangle\}_{m=1}^M$.

Theorem 7 (QHL using bcQSE) *With access to multiple copies of quantum training data $\{|\mathbf{x}^{(m)}\rangle\}_{m=1}^M$, QHL can be achieved through bcQSE.*

Proof We first note by referring to Definition 1 that $\rho = \mathbb{I}_d/d$ forms the quantum analogue of the Hebbian weight matrix, where $\rho = (1/M) \sum_{m=1}^M |\mathbf{x}^{(m)}\rangle \langle \mathbf{x}^{(m)}|$ and with \mathbb{I}_d the identity operator. Then, using Theorem 3, we see that ρ is operationally accessible through the transformation $e^{-it|1\rangle\langle 1| \otimes \rho}$ acting on a learning qubit and N processing qubits. Furthermore, the explicit operational availability of W is guaranteed from ρ since

$$e^{-i\Delta t|1\rangle\langle 1| \otimes W} = e^{-i\Delta t|1\rangle\langle 1| \otimes \rho} e^{i\Delta t|1\rangle\langle 1| \otimes \frac{\mathbb{I}_d}{d}} + \mathcal{O}(\Delta t^2), \quad (7)$$

i.e., so in every batch of bcQSE, one can apply a conditional phase $e^{i\Delta t|1\rangle\langle 1| \otimes \frac{\mathbb{I}_d}{d}}$, which is equivalent to applying a single qubit time-dependent phase and Pauli Z gate for each batch. Note that this step can typically be skipped as the incurred phase error is small $\mathcal{O}(\frac{1}{d})$. \square

Thus, we have shown that, from our definitions, QHL and batched quantum state exponentiation are closely related. The given error analysis of bcQSE applies similarly to QHL and results in a *dimension-efficient* simulation. Data efficiency is not achieved with the methods shown in this work.

Theorem 8 *If $t = \mathcal{O}(\text{poly}(NM))$ and $1/\epsilon = \mathcal{O}(\text{poly}(NM))$, then QHL realized through bcQSE is dimension efficient.*

Proof From Table 1, using bcQSE to perform QHL with Clifford+T gates is polynomial in N for the required number of qubits and gate count. \square

Let us elaborate further on the nature of the quantum training data. On the one hand, we can assume the training data to be inherently quantum, originating from another quantum device. Alternatively, we can attempt to encode classical training data into quantum states. Suppose one is given the d -dimensional training data $\mathbf{x}^{(m)}$ where without loss of generality $d = 2^N$. The Hebbian weight matrix is written as in the quantum case

$$W = \frac{1}{Md} \left[\sum_{m=1}^M \mathbf{x}^{(m)} (\mathbf{x}^{(m)})^\top \right] - \frac{\mathbb{I}_d}{d}. \quad (8)$$

Assume the following oracle.

Data Input 1 *We are given M oracles such that the operation $|0 \dots 0\rangle \rightarrow |\mathbf{x}^{(m)}\rangle$ on $N = \lceil \log d \rceil$ qubits can be performed, with $|\mathbf{x}^{(m)}\rangle = \frac{1}{\sqrt{d}} \sum_i x_i^{(m)} |i\rangle$ and $m = 1, \dots, M$. Each oracle has a runtime of at most T_{data} .*

Since the oracle generates quantum states as required for batched quantum state exponentiation, we have the following corollary.

Corollary 1 *Given Oracle 1, classical-input QHL can be performed. If $t = \mathcal{O}(\text{poly}(NM))$, $1/\epsilon = \mathcal{O}(\text{poly}(NM))$, and also $T_{\text{data}} = \mathcal{O}(\text{poly}(NM))$ then classical-input QHL is dimension efficient.*

Efficiency of classical-input QHL requires that the qubit training states $|x^{(m)}\rangle$ can be produced efficiently. This is a topical question that requires future development, but has also not been conclusively ruled out (Aaronson 2015; Giovannetti et al. 2008; Soklakov and Schack 2006).

4.3 Phase estimation

Here, we focus on achieving QHL using bcQSE and elaborate further on the operational availability of ρ through phase estimation, allowing for calculation of eigenvalues (and corresponding eigenvectors) of ρ .

Theorem 9 (Phase estimation with bcQSE) *An eigenvalue of ρ can be estimated to error ϵ through phase estimation using $\mathcal{O}(\frac{1}{\epsilon})$ applications of $e^{-i\rho}$. In the setting of Theorem 6, realizing bcQSE using Clifford+T gates results in an overall overhead for phase estimation of:*

- $\mathcal{O}(\frac{1}{\epsilon^3})$ batches of quantum data;
- $\tilde{\mathcal{O}}(\frac{MN}{\epsilon^3} \log(\frac{MN}{\epsilon^3}))$ qubits;
- $\tilde{\mathcal{O}}(\frac{MN}{\epsilon^3} \log(\frac{MN}{\epsilon^3}))$ Clifford+T gates;

Proof Consider the spectral decomposition $\rho = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|$ and suppose N processing qubits are prepared in the eigenstate $|\lambda_i\rangle$. Using Kitaev's version Ref. Kitaev (1995), Kimmel et al. (2017), phase estimation allows λ_i to be estimated to precision ϵ using $\mathcal{O}(\frac{1}{\epsilon})$ controlled unitary applications of $e^{-i\rho}$. If the processing qubits are prepared in an arbitrary state $|\psi\rangle = \sum_i c_i |\lambda_i\rangle$, $\sum_i |c_i|^2 = 1$, then the eigenvalue λ_i will be measured with probability $|c_i|^2$ and the resultant state of the processing qubits is $|\lambda_i\rangle$ (Nielsen and Chuang 2002).

Now, if each application of bcQSE for time $t = 1$ can be achieved to error ϵ^2 , then the overall simulation error after $\mathcal{O}(\frac{1}{\epsilon})$ repetitions is $\mathcal{O}(\epsilon)$. The number of steps required in any single bcQSE interaction for phase estimation is thus $n = \mathcal{O}(\frac{1}{\epsilon^2})$. Referring to Table 1, we see that all the bcQSE operations combined require $\mathcal{O}(\frac{1}{\epsilon^3})$ batches of quantum data, $\tilde{\mathcal{O}}(\frac{MN}{\epsilon^3} \log(\frac{MN}{\epsilon^3}))$ qubits and gates.

Phase estimation also has an overhead. It requires $\mathcal{O}(\frac{1}{\epsilon})$ additional qubits (referred to here as the learning qubits) containing the eigenvalues and an additional $\mathcal{O}(\frac{1}{\epsilon^2})$ gates. Both of these numbers are dominated by their counterparts from bcQSE. \square

Phase estimation is the prototypical operational usage of ρ , leading us to the following definition.

Definition 4 Quantum Hebbian eigensystem learning is any quantum algorithm that uses QHL to derive the eigenvalues and/or eigenvectors of the corresponding Hebbian weight matrix.

We note that the above Corollary 1 implies that classical-input quantum Hebbian eigensystem learning can be dimension efficient using the encoding in Eq. 8.

5 Discussion

This work presents a method of quantum state exponentiation by using batches of pure quantum data (bcQSE). Here, the controlled partial swap between processing and data qubits acts as the elementary transformation which must be repeated multiple times for varying registers of data qubits. We have decomposed this elementary transformation into Clifford+T gates, hence allowing for the realization of bcQSE through a compilation of Clifford+T gates. An analysis of the efficiency of our decomposition was also provided. It must be emphasized that this decomposition is not necessarily optimal, and it would be of interest to compare to the results of quantum compilers—which can aim to minimize the gate cost overhead (Heyfron and Campbell 2017).

We have presented bcQSE within the context of quantum Hebbian learning. A general formalism for quantum Hebbian learning was constructed based upon ideas established in classical Hebbian learning. However, the application of bcQSE is not restricted to classical-input quantum Hebbian learning: a neuroscience analogy allows bcQSE to be thought of representing multiple levels of perception (as in the human brain). In this analogy, the data qubits represent the environment, which is continually interacting with processing qubits acting as low-level senses. The learning qubits are then the higher levels of perception, using information from the processing qubits to learn about the environment. We further extended the concept of Hebbian learning to quantum phase estimation, showing how the elementary controlled partial swap operation can be built up to achieve important quantum algorithms with applicability for quantum learning.

Quantum Hebbian learning holds relevance for the teaching of quantum and classical neural networks, such as the Hopfield network (and its quantum versions Rebentrost et al. 2018), with application in pattern recognition and optimization. Future work can consider quantum analogues of more advanced classical learning techniques such as the Storkey learning rule (Storkey and Valabregue 1999), or more general applications of the batching process in machine learning.

Acknowledgements We acknowledge Seth Lloyd, Iman Marvian, and George Siopsis for insightful discussions.

References

- Aaronson S (2015) Read the fine print. *Nat Phys* 11(4):291
- Amin MH, Andriyash E, Rolfe J, Kulchitsky B, Melko R (2018) Quantum Boltzmann Machine. *Phys Rev X* 8(2):021050
- Barr DS, Mani G (1998) Predictive neural network means and method for selecting a portfolio of securities wherein each network has been trained using data relating to a corresponding security. US Patent 5,761,442
- Benedetti M, Realpe-Gómez J, Perdomo-Ortiz A (2018) Quantum-assisted Helmholtz machines: a quantum–classical deep learning framework for industrial datasets in near-term devices. *Quantum Science and Technology* 3(3):034007
- Benenti G, Strini G (2010) Computing the distance between quantum channels: usefulness of the Fano representation. *J Phys B Atomic Mol Phys* 43:215508
- Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S (2017) Quantum machine learning. *Nature* 549:195
- Childs AM, Cleve R, Deotto E, Farhi E, Gutmann S, Spielman DA (2003) In: *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. ACM, pp 59–68
- Childs AM, Cleve R, Deotto E, Farhi E, Gutmann S, Spielman DA (2018) Toward the first quantum simulation with quantum speedup. *Proc Natl Acad Sci* 115(38):9456–9461
- Farhi E, Neven H (2018) arXiv:1802.06002
- Gilyén A, Lloyd S, Tang E (2017) arXiv:1811.04909
- Giovannetti V, Lloyd S, Maccone L (2008) Quantum random access memory. *Phys Rev Lett* 100(16):160501
- Gottesman D (2009) arXiv:0904.2557
- Harrow AW, Hassidim A, Lloyd S (2009) Quantum algorithm for linear systems of equations. *Phys Rev Lett* 103(15):150502
- Hebb DO (1949) *The organization of behavior*. Wiley, Hoboken
- Heyfron L, Campbell ET (2017) arXiv:1712.01557
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci* 79(8):2554
- Kimmel S, Lin CYY, Low GH, Ozols M, Yoder TJ (2017) Hamiltonian simulation with optimal sample complexity. *NPJ Quantum Inf* 3(1):13
- Kitaev AY (1995) arXiv:9511026
- Lasa R, Berndt D (2007) System for rating quality of online visitors. US Patent App. 11/759,889
- Liu N, Rebentrost P (2018) Quantum machine learning for quantum anomaly detection. *Phys Rev A* 97(4):042315
- Lloyd S, Mohseni M, Rebentrost P (2014) Quantum principal component analysis. *Nat Phys* 10:631
- MacKay DJ (2003) *Information theory, inference and learning algorithms*. Cambridge University Press, Cambridge
- Marvian I, Lloyd S (2016) arXiv:1606.02734
- McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5(4):115
- Nielsen MA, Chuang I (2002) *Quantum computation and quantum information*. Cambridge University Press, Cambridge
- Norris JA (1999) System and method for real time loan approval. US Patent 5,870,721
- Rebentrost P, Mohseni M, Lloyd S (2014) Quantum support vector machine for big data classification. *Phys Rev Lett* 113(13):130503
- Rebentrost P, Bromley TR, Weedbrook C, Lloyd S (2018) Quantum Hopfield neural network. *Phys Rev A* 98(4):042308
- Ross NJ, Selinger P (2015) Optimal ancilla-free Clifford+T approximation of z-rotations. *Quantum Inf Comput* 15(11–12):901
- Schuld M, Sinayskiy I, Petruccione F (2014) The quest for a quantum neural network. *Quantum Inf Process* 13(11):2567
- Soklakov AN, Schack R (2006) *Phys Rev A* 73(1):012307
- Storkey AJ, Valabregue R (1999) The basins of attraction of a new Hopfield learning rule. *Neural Netw* 12(6):869
- Tang E (2018) arXiv:1807.04271
- Verdon G, Broughton M, Biamonte J (2017) arXiv:1712.05304
- Wiebe N, Braun D, Lloyd S (2012) Quantum algorithm for data fitting. *Phys Rev Lett* 109(5):050505
- Wiebe N, Kapoor A, Svore KM (2014) arXiv:1412.3489

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.