QTML 2021

Quantum Techniques in Machine Learning

Booklet of Extended Abstracts

Thursday and Friday, November 11 and 12 2021

Barren plateaus preclude learning scramblers

Zoë Holmes, Andrew Arrasmith, Bin Yan, Patrick J. Coles, Andreas Albrecht and Andrew T. Sornborger

Scrambling [1–3], the process by which quantum information is rapidly spread through many-body quantum systems, has proven central not only to understanding quantum chaos but also to the study of the dynamics of quantum information [4–6], thermalization phenomena [7, 8], the black hole information paradox [1, 9–11], holography [12, 13], random circuits [14–16], fluctuation relations [17, 18] and entropic uncertainty relations [19]. However, the complexity of strongly-interacting many-body quantum systems makes scrambling rather challenging to study analytically. Recently, quantum machine learning (QML) has emerged as an exciting new paradigm for the study of complex physical processes [20–24]. It is therefore natural to ask whether QML could be used to study scrambling.

However, despite the high expectations placed on QML, there remain fundamental questions concerning its scalability and breadth of applicability. Of particular concern is the growing body of literature on the existence of *barren plateaus*, i.e., regions in parameter space where cost gradients vanish exponentially as the size of the system studied increases. This phenomenon, which severely limits the trainability of large scale quantum neural networks, has been demonstrated in a number of proposed architectures and classes of cost function [27–31].

In this talk we will present a no-go theorem for the use of QML to study quantum scrambling. Namely, we show that any QML approach used to learn the unitary dynamics implemented by a typical scrambler will exhibit a barren plateau and thus be untrainable in the absence of further prior knowledge. Crucially, in contrast to previously established barren plateau phenomena, which are a consequence of the ansatz structure and parameter initialization strategy, our barren plateaus holds for any choice of ansatz and any initialization of parameters. Thus, previously proposed strategies for avoiding barren plateaus do not work here.

More generally, given the close connection between scrambling and randomness, our no-go theorem also applies to learning random and pseudo-random unitaries. Consequently, our result implies that to efficiently learn an unknown unitary process using QML, prior information about that process is required. Thus, our result provides a fundamental limit on the domain of applicability of QML. Detailed analysis can be found in Ref. [32].

The machine learning task we consider can be illustrated by the famous Hayden-Preskill thought experiment [1], in which Alice attempts to destroy a secret, encoded in a quantum state, by throwing it into Nature's fastest scrambler, a black hole. The question then is: how safe is Alice's secret? Hayden and Preskill argued that if Bob knows the unitary dynamics, U, implemented by the black hole, and shares a maximally entangled state with the black hole, it is possible to decode Alice's secret by collecting a few additional photons emitted from the black hole. However, this prompts a second question, how might Bob learn the scrambling unitary in the first place? Here we investigate whether QML can be used to learn the scrambling unitary, U.

To address this, we first motivate our notion of a scrambler via the out-of-time-ordered correlator (OTOC) [3, 13],

$$f_{\text{OTOC}} \equiv \langle \tilde{X} Y \tilde{X}^\dagger Y^\dagger \rangle \,. \tag{1}$$

Here X and Y are local operators on different subsystems, $\tilde{X} = U^\dagger X(0) U$ is the Heisenberg evolved initial operator $X(0)$ and the average is taken over an infinite temperature state $\rho \propto \mathbb{1}$. After a time scale called the scrambling time, the OTOC of a chaotic system tends to a minimal value that is equivalent to taking its average over a random distribution of unitaries [14]. Since f_{OTOC} in (1) only involves the second moment of the unitary, its asymptotic smallness can be attributed to the fact that the scrambling unitary appears to be a typical element of a 2-design [14]. Hence, we can model a scrambler as a unitary that is drawn from a distribution that forms at least a unitary 2-design [1, 2].

Suppose one wants to use QML to learn an unknown target unitary V where all that is known is that it is drawn from an ensemble of scramblers $\mathbb{V}$ (i.e. a 2-design). The aim of QML is to minimize a problem-specific cost function that is evaluated on a quantum computer. In the context of learning unitaries, one considers an *ansatz* (i.e., parameterized quantum circuit) $U(\boldsymbol{\theta})$ and a target unitary V. To quantify the quality of the training, one can employ a generic cost function of the form

$$C(\boldsymbol{\theta}, V) = \langle \psi | W^2(\boldsymbol{\theta})^\dagger H W(\boldsymbol{\theta}) | \psi \rangle \,, \tag{2}$$

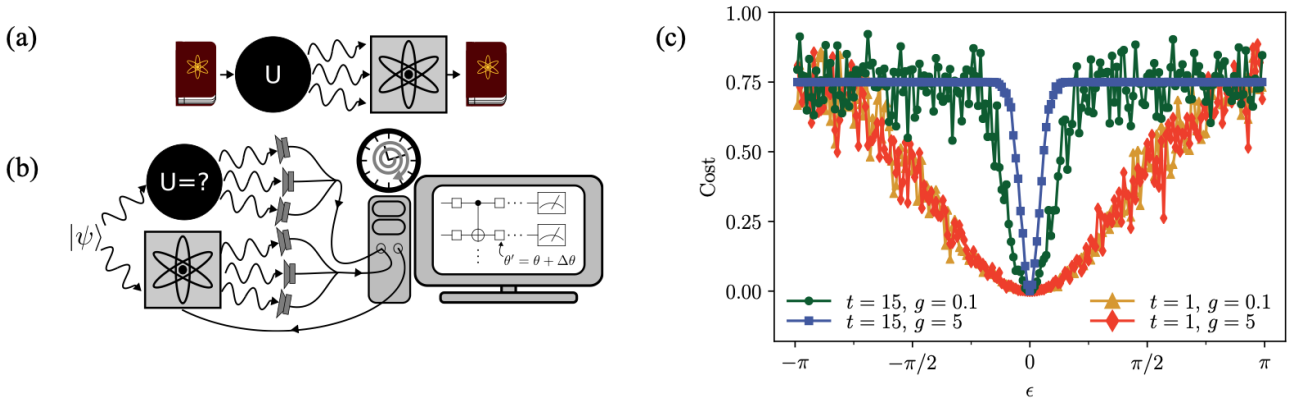


Figure 1. Panel (a) shows the setup of the classic Hayden-Preskill thought experiment. Panel (b) shows the process of attempting to learn U. Panel (c) shows numerical results for approximate scramblers. A random cut of the landscape of the LHST cost function [25] $C_{\text{LHST}}(U, V)$ where $V(g, t)$ is a randomly generated 9 qubit scrambler ($n = 9$) modelled using the minimal model introduced in [26] and $U(g, t)$ is an ansatz of the same form. Here ϵ is a noise parameter that determines the deviation of the ansatz parameters, $\boldsymbol{\theta}$, from the target scrambler's parameters, $\boldsymbol{\theta}^{\text{target}}$. The landscape for a weak scrambler with $t = 1$ and $g = 0.1$ ($g = 5$) is plotted in yellow (red). Stronger scramblers with $t = 15$ and $g = 0.1$ ($g = 5$) are plotted in green (blue).

where $|\psi\rangle$ is some state, H is some Hermitian operator and $W(\boldsymbol{\theta}) = V^\dagger U(\boldsymbol{\theta})$. We consider learning V by variationally minimizing $C(\boldsymbol{\theta}, V)$.

If the cost function gradient is vanishingly small for all parameters, then the cost landscape forms a barren plateau. It follows from Chebyshev's inequality that if the average gradient of the cost $\langle \partial_{\theta_k} C(\boldsymbol{\theta}, V)\rangle_{\mathbb{V}}$ vanishes and the variance in the gradient $\text{Var}_{\mathbb{V}}[\partial_{\theta_k} C]$ is vanishingly small for all θ_k, then the probability that the cost partial derivative is non-zero is vanishingly small for all parameters. Indeed this behavior is precisely what we find here. Specifically, our main results consists of the following proposition, theorem and corollary.

Proposition 1. *The average partial derivative of $C(\boldsymbol{\theta}, V)$, with respect to any parameter θ_k, for an ensemble of target unitaries $\mathbb{V}$ that form a 2-design, is given by*

$$\langle \partial_{\theta_k} C(\boldsymbol{\theta}, V)\rangle_{\mathbb{V}} = 0\,. \tag{3}$$

Theorem 2. *The variance of the partial derivative of $C(\boldsymbol{\theta}, V)$, with respect to any parameter θ_k, for an ensemble of target unitaries $\mathbb{V}$ that form a 2-design, is given by*

$$\text{Var}_{\mathbb{V}}[\partial_{\theta_k} C] = \left[\frac{2\,\text{Tr}[H^2]}{2^{2n}-1} - \frac{2(\text{Tr}[H])^2}{2^n(2^{2n}-1)}\right] \text{Var}_{\chi}[-iU\partial_{\theta_k} U^\dagger], \tag{4}$$

where $\text{Var}_{\mathbb{V}}$ denotes the variance over the ensemble $\mathbb{V}$, and Var_{χ} denotes the quantum-mechanical variance with respect to the ansatz-evolved state $|\chi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})\,|\psi\rangle$.

Corollary 3. *Without loss of generality, the ansatz $U(\boldsymbol{\theta})$ to learn an n-qubit target unitary V can be written in the form $U(\boldsymbol{\theta}) = \prod_{i=1}^{N} U_i(\boldsymbol{\theta_i}) W_i$ where $\{W_i\}$ is a chosen set of fixed unitaries and $U_i(\theta_i) = \exp(-i\theta_i G_i)$ with G_i an Hermitian operator. If $\text{Tr}[H^2] \in \mathcal{O}(2^n)$ and $||G_k^2||_\infty \in \mathcal{O}(1)$, then*

$$\text{Var}_{\mathbb{V}}[\partial_{\theta_k} C] \in \mathcal{O}(2^{-n})\,. \tag{5}$$

Thus these results establish that for standard costs, and any ansatze and initialization strategy, the gradient of the cost is unbiased and its variance vanishes exponentially with the size of the scrambler. Hence for large systems the gradient of the cost is vanishingly small and the landscape forms a barren plateau as claimed.

Fig. 1c) provides a nice visual representation of how the barrenness of the cost landscape depends on the degree to which the target unitary is scrambling. In the case of a highly scrambling unitary (blue) the majority of the landscape forms a *barren plateau* with only a *narrow gorge* where the cost dips down to its minimum. In contrast for weaker scramblers (yellow) the valley around the minimum is wider and the plateau more featured.

[1] Patrick Hayden and John Preskill. Black holes as mirrors: quantum information in random subsystems. *Journal of High Energy Physics*, 2007(09):120–120, sep 2007.

[2] Yasuhiro Sekino and L Susskind. Fast scramblers. *Journal of High Energy Physics*, 2008(10):065–065, oct 2008.

[3] Juan Maldacena, Stephen H. Shenker, and Douglas Stanford. A bound on chaos. *Journal of High Energy Physics*, 2016(8):106, 2016.

[4] Pavan Hosur, Xiao-Liang Qi, Daniel A. Roberts, and Beni Yoshida. Chaos in quantum channels. *Journal of High Energy Physics*, 2016(2):4, 2016.

[5] Adam Nahum, Jonathan Ruhman, Sagar Vijay, and Jeongwan Haah. Quantum entanglement growth under random unitary dynamics. *Phys. Rev. X*, 7:031016, Jul 2017.

[6] Winton Brown and Omar Fawzi. Decoupling with random quantum circuits. *Communications in Mathematical Physics*, 340(3):867–900, 2015.

[7] R. J. Lewis-Swan, A. Safavi-Naini, J. J. Bollinger, and A. M. Rey. Unifying scrambling, thermalization and entanglement through measurement of fidelity out-of-time-order correlators in the dicke model. *Nature Communications*, 10(1):1581, 2019.

[8] Markus J. Klug, Mathias S. Scheurer, and Jörg Schmalian. Hierarchy of information scrambling, thermalization, and hydrodynamic flow in graphene. *Phys. Rev. B*, 98:045102, Jul 2018.

[9] Beni Yoshida and Alexei Kitaev. Efficient decoding for the hayden-preskill protocol. *arXiv preprint arXiv:1710.03363*, 2017.

[10] Beni Yoshida and Norman Y. Yao. Disentangling scrambling and decoherence via quantum teleportation. *Phys. Rev. X*, 9:011006, Jan 2019.

[11] K. A. Landsman, C. Figgatt, T. Schuster, N. M. Linke, B. Yoshida, N. Y. Yao, and C. Monroe. Verified quantum information scrambling. *Nature*, 567(7746):61–65, 2019.

[12] Stephen H. Shenker and Douglas Stanford. Black holes and the butterfly effect. *Journal of High Energy Physics*, 2014(3):67, 2014.

[13] A. Kitaev. A simple model of quantum holography. Proceedings of the KITP Program: Entanglement in Strongly-Correlated Quantum Matter (Kavli Institute for Theoretical Physics, Santa Barbara), 2015.

[14] Daniel A. Roberts and Beni Yoshida. Chaos and complexity by design. *Journal of High Energy Physics*, 2017(4):121, 2017.

[15] Jordan Cotler, Nicholas Hunter-Jones, Junyu Liu, and Beni Yoshida. Chaos, complexity, and random matrices. *Journal of High Energy Physics*, 2017(11):48, 2017.

[16] Bruno Bertini and Lorenzo Piroli. Scrambling in random unitary circuits: Exact results. *Phys. Rev. B*, 102:064305, Aug 2020.

[17] A. Chenu, I. L. Egusquiza, J. Molina-Vilaplana, and A. del Campo. Quantum work statistics, loschmidt echo and information scrambling. *Scientific Reports*, 8(1):12634, 2018.

[18] Nicole Yunger Halpern. Jarzynski-like equality for the out-of-time-ordered correlator. *Phys. Rev. A*, 95:012120, Jan 2017.

[19] Nicole Yunger Halpern, Anthony Bartolotta, and Jason Pollack. Entropic uncertainty relations for quantum information scrambling. *Communications Physics*, 2(1):92, 2019.

[20] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018.

[21] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

[22] Kerstin Beer, Dmytro Bondarenko, Terry Farrelly, Tobias J. Osborne, Robert Salzmann, Daniel Scheiermann, and Ramona Wolf. Training deep quantum neural networks. *Nature Communications*, 11(1):808, 2020.

[23] Kyle Poland, Kerstin Beer, and Tobias J Osborne. No free lunch for quantum machine learning. *arXiv preprint arXiv:2003.14103*, 2020.

[24] Kunal Sharma, M Cerezo, Zoë Holmes, Lukasz Cincio, Andrew Sornborger, and Patrick J Coles. Reformulation of the no-free-lunch theorem for entangled data sets. *arXiv preprint arXiv:2007.04900*, 2020.

[25] Sumeet Khatri, Ryan LaRose, Alexander Poremba, Lukasz Cincio, Andrew T Sornborger, and Patrick J Coles. Quantum-assisted quantum compiling. *Quantum*, 3:140, 2019.

[26] Ron Belyansky, Przemyslaw Bienias, Yaroslav A. Kharkov, Alexey V. Gorshkov, and Brian Swingle. Minimal model for fast scrambling. *Phys. Rev. Lett.*, 125:130601, Sep 2020.

[27] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):4812, 2018.

[28] Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost-function-dependent barren plateaus in shallow quantum neural networks. *arXiv preprint arXiv:2001.00550*, 2020.

[29] Kunal Sharma, Marco Cerezo, Lukasz Cincio, and Patrick J Coles. Trainability of dissipative perceptron-based quantum neural networks. *arXiv preprint arXiv:2005.12458*, 2020.

[30] Samson Wang, Enrico Fontana, M Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J Coles. Noise-induced barren plateaus in variational quantum algorithms. *arXiv preprint arXiv:2007.14384*, 2020.

[31] M Cerezo and Patrick J Coles. Impact of barren plateaus on the hessian and higher order derivatives. *arXiv preprint arXiv:2008.07454*, 2020.

[32] Zoë Holmes, Andrew Arrasmith, Bin Yan, Patrick J. Coles, Andreas Albrecht, and Andrew T. Sornborger. Barren plateaus preclude learning scramblers. *Phys. Rev. Lett.*, 126:190501, May 2021.

Long-time simulations with high fidelity on quantum hardware

Joe Gibbs, Kaitlin Gili, Zoë Holmes, Benjamin Commeau, Andrew Arrasmith, Lukasz Cincio, Patrick J. Coles, and Andrew Sornborger

We present a new algorithm, called fixed state Variational Fast Forwarding, allowing long-time high fidelity simulations of quantum systems on NISQ devices with short coherence times. We execute it successfully on quantum hardware and in large numerical implementations, and analyse the noise resilience and scaling of simulation errors.

Motivation Relative to classical computers, quantum computers have the potential to provide exponential speed ups for simulating quantum systems. Such quantum simulations could lead to advances across a wide range of industries including pharmaceutical development and materials design. However, while quantum computers are rapidly reaching the stage where they can deliver an advantage over classical devices, we remain in the noisy intermediate-scale quantum (NISQ) era in which the length of time that may be simulated is limited by hardware noise. Here we demonstrate that, despite these limitations, it is possible to implement long-time, high fidelity simulations on current hardware.

Background Our simulations are performed using a new algorithm that we call the fixed state Variational Fast Forwarding (fsVFF) algorithm [1]. We present a significant reduction in resource requirements over the recently proposed Variational Fast Forwarding (VFF) algorithm [2]. VFF allows long time simulations to be performed using a fixed depth circuit, thus enabling a quantum simulation to be 'fast forwarded' beyond the coherence time of noisy hardware. The VFF algorithm requires finding a full diagonalization of the short time evolution operator U of the system of interest. Once found, the diagonalization enables any initial state of that system to be fast forwarded. However, for practical purposes, one is often interested in studying the evolution of a particular fixed initial state of interest. In that case a full diagonalization of U is overkill, and it suffices to find a diagonalization on the subspace spanned by the initial state, rather than on the total Hilbert space. This approach is tailored to making dynamical simulation more suitable for NISQ hardware in two key ways. First, the cost function requires half as many qubits as VFF. This not only allows larger scale simulations to be performed on current resource-limited hardware, but also has the potential to enable higher fidelity simulations since larger devices tend to be noisier. Second, fsVFF can utilize simpler ansätze than VFF both in terms of the depth of the ansatz and the number of parameters that need to be learnt. Thus, fsVFF can reduce the width, depth, and total number of circuits required to fast forward quantum simulations, hence increasing the viability of performing simulations on near-term hardware.

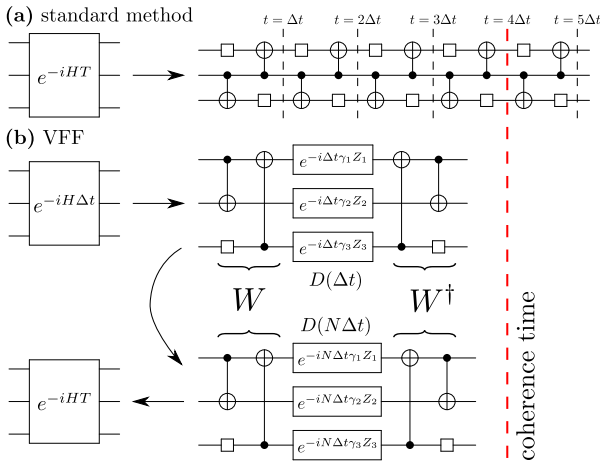


FIG. 1: **Schematic** (a) Trotterization-based quantum simulation running past the coherence time of the quantum hardware. (b) A variationally fast forwarded quantum simulation. The approximate diagonalization is used to simulate Hamiltonian evolution with a fixed depth circuit.

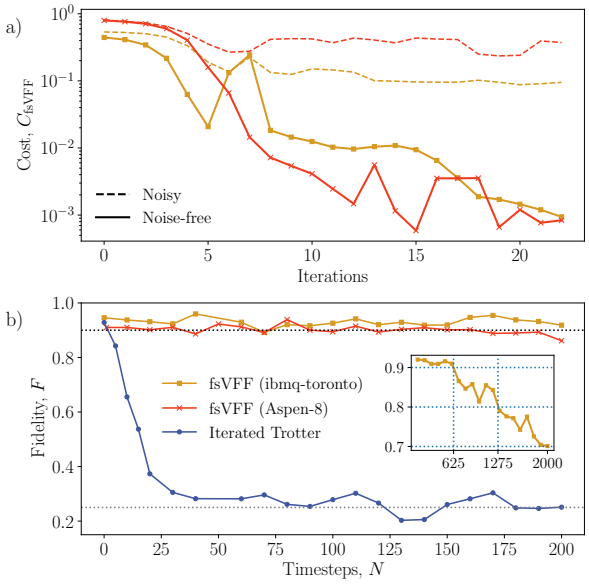


FIG. 2: **fsVFF Hardware Results** a) Training curve of the diagonalization of the 2-qubit XY Hamiltonian in the subspace spanned by $|10\rangle$. b) The initial state is evolved forwards in time using the fsVFF ansatz, plotting the fidelity $F = \langle\psi|\rho|\psi\rangle$ between the evolved state and exact evolution.

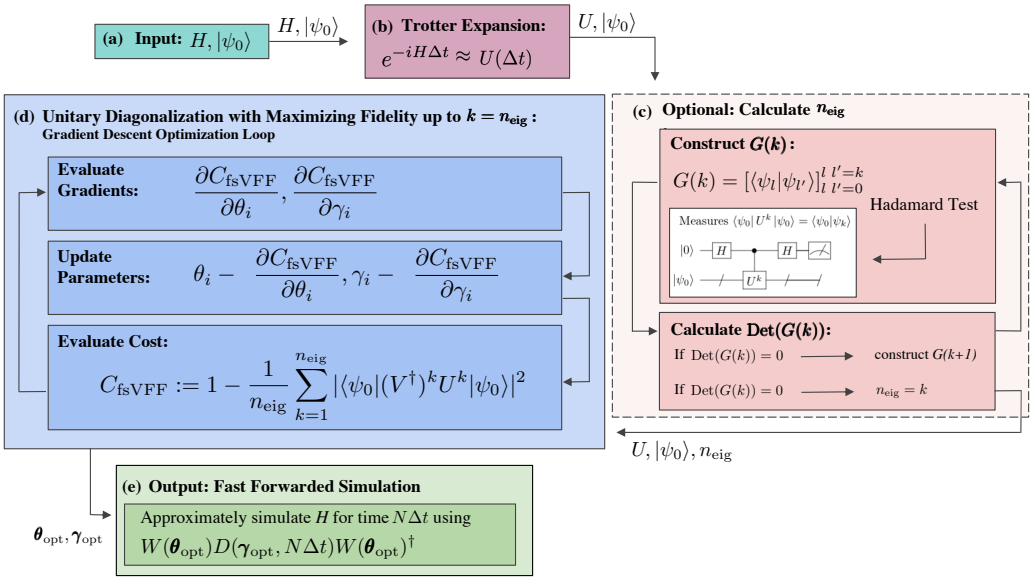


FIG. 3: **The fsVFF Algorithm**

Results We demonstrate these advantages by implementing long-time high fidelity quantum simulations of the 2-qubit XY spin chain on Rigetti's and IBM's quantum computers. Specifically, while the iterated Trotter approach has a fidelity of less than 0.9 after 4 time steps and has completely thermalized by 25 time steps, with fsVFF we achieve a simulation fidelity greater than 0.9 for over 600 time steps, as shown in Fig. 2. We further support the effectiveness of this approach for NISQ simulations, utilizing adaptive ansätze whose discrete structure evolves during training to perform 4 qubit noisy and 8 qubit noiseless numerical simulations of the XY model and Fermi-Hubbard model respectively. A method inspired by stochastic gradient descent of classical neural networks is developed to remove the dependence on the dimension of the subspace spanned from the number of circuits required per cost function evaluation.

The fsVFF Algorithm The fixed state Variational Fast Forwarding algorithm (fsVFF) is summarized in Fig. 3. We start with an initial state $|\psi_0\rangle$ that we wish to evolve under the Hamiltonian H.

1. Approximate the short time evolution using a single step Trotter approximation U.

2. Variationally search for a diagonalization $V = W(\theta)D(\gamma, \Delta t)W(\theta)^\dagger$, D is a diagonal matrix, of U over the energy subspace spanned by $|\psi_0\rangle$, using the cost function,

$$C_{\text{fsVFF}} := 1 - \frac{1}{n_{\text{eig}}}\sum_{k=1}^{n_{\text{eig}}} |\langle\psi_0|(V^\dagger)^k U^k|\psi_0\rangle|^2 . \tag{1}$$

   Here n_{eig} is the number of energy eigenstates of the Hamiltonian H spanned by the initial state, which can be calculated using an algorithm we provide. At each iteration step the gradient of the cost function is measured on the quantum computer, which is used to update the parameters using a classical optimizer. The output of the optimization loop is the set of parameters $\{\boldsymbol{\theta}_{\text{opt}}, \boldsymbol{\gamma}_{\text{opt}}\}$ that minimize C_{fsVFF}.

3. Finally, the state $|\psi_0\rangle$ can be simulated for time $T = N\Delta t$ using the circuit

$$W(\boldsymbol{\theta}_{\text{opt}})D(\boldsymbol{\gamma}_{\text{opt}}, N\Delta t)W(\boldsymbol{\theta}_{\text{opt}})^\dagger . \tag{2}$$

   That is, by simply multiplying the parameters $\boldsymbol{\gamma}_{\text{opt}}$ in the diagonalized unitary by the number of iterations N.

Noise Resilience and Energy Estimation In our analytical results, we prove the faithfulness of the fsVFF cost function by utilizing the newly developed No-Free-Lunch theorems for quantum machine learning [3, 4]. We also provide a proof of the noise resilience of the fsVFF cost function, specifically the optimal parameter resilience [5]. Finally, we perform an analysis of simulation errors under fast-forwarding. The diagonalizations obtained using fsVFF may further be useful for determining the eigenstates and eigenvalues of the Hamiltonian on the subspace spanned by the initial state. This can be done using a time series analysis, by using fsVFF to reduce the depth of the quantum phase estimation (QPE) algorithm, or using a simple sampling method. We demonstrate on IBM's quantum computer that, while standard QPE fails on real hardware, fsVFF can be used to obtain accurate estimates of the spectrum.

[1] Joe Gibbs, Kaitlin Gili, Zoë Holmes, Benjamin Commeau, Andrew Arrasmith, Lukasz Cincio, Patrick J Coles, and Andrew Sornborger, "Long-time simulations with high fidelity on quantum hardware," arXiv preprint arXiv:2102.04313 (2021).
[2] Cristina Cirstoiu, Zoe Holmes, Joseph Iosue, Lukasz Cincio, Patrick J Coles, and Andrew Sornborger, "Variational fast forwarding for quantum simulation beyond the coherence time," npj Quantum Information **6**, 1–10 (2020).
[3] Kyle Poland, Kerstin Beer, and Tobias J Osborne, "No free lunch for quantum machine learning," arXiv preprint arXiv:2003.14103 (2020).
[4] Kunal Sharma, M. Cerezo, Zoë Holmes, Lukasz Cincio, Andrew Sornborger, and Patrick J Coles, "Reformulation of the no-free-lunch theorem for entangled data sets," arXiv preprint arXiv:2007.04900 (2020).
[5] Kunal Sharma, Sumeet Khatri, M. Cerezo, and Patrick J Coles, "Noise resilience of variational quantum compiling," New Journal of Physics **22**, 043006 (2020).

Eigenstate extraction with neural-network tomography

Abhijeet Melkani[a,b,c], Clemens Gneiting[a,d], Franco Nori[a,d,e]

[a] *RIKEN Cluster for Pioneering Research, Wako, Saitama 351-0198, Japan*
[b] *Department of Physics, IIT Bombay, Mumbai, Marahashtra 400076, India*
[c] *present address: Department of Physics, University of Oregon, Eugene, OR-97405, USA*
[d] *RIKEN Center for Quantum Computing, Wako, Saitama 351-0198, Japan*
[e] *Department of Physics, University of Michigan, Ann Arbor, Michigan 48109-1040, USA*

Abstract

We implement quantum state tomography through the iterative reconstruction of the eigenstates of the experimentally produced mixed states. Our method leverages machine learning through neural-network representations of pure quantum states, which efficiently scale to large quantum systems. We verify our method with experimental data from trapped ion experiments.

Keywords: quantum state reconstruction, neural-network quantum state representation, restricted Boltzmann machines

In times where quantum experiments and quantum devices have reached unprecedented size and complexity, their verification has become increasingly hard and yet indispensable. Noise and imperfections cause deviations of the produced states from the target states, which may, in many cases, put their intended purpose in jeopardy. Quantum state tomography is the process of reconstructing the mixed states produced in quantum experiments or devices from their measurement data. Based on statistical analysis of a near-complete set of this measurement data, the realized states can, in principle, be fully reconstructed with high accuracy [1, 2, 3, 4].

However, full, unconditional quantum state tomography becomes prohibitively expensive with increasing Hilbert space dimension, both from an experimental perspective (the required number of measurements scales exponentially with the system size) and from the perspective of data post-processing. While machine learning has the potential to significantly mitigate these issues, full fledged mixed-state tomography using machine learning has remained challenging, due to the intrinsic constraints to be met by physical quantum states.

In this contribution, we propose an algorithm for the stepwise reconstruc-

tion of mixed quantum states in terms of their eigenvalues and eigenstates. Our scheme thus exploits that highly efficient, machine learning-based methods for pure-state tomography can also be used to robustly recover the eigenstates of mixed states. Tailored iteration then allows one to recover the eigenvalues and eigenstates of mixed states up to a desired rank. Such reconstruction of mixed states delivers valuable structural information about the state produced, and comes with substantially reduced costs.

To demonstrate our reconstruction scheme, we specifically use and adjust a recently developed method for pure-state tomography based on neural-network representations of quantum states [5]. Indeed, these *Neural Quantum States*, which have restricted Boltzmann machines at their core (see Fig. 1), have been shown [5] to be viable for tomography of complex, high-dimensional pure states, leveraging both the efficient and scalable representation of neural networks and their great expressional power. We can show that [6], using our iterative reconstruction scheme, these benefits carry over to the eigenstate retrieval of mixed states, and hence to full-fledged mixed-state tomography.

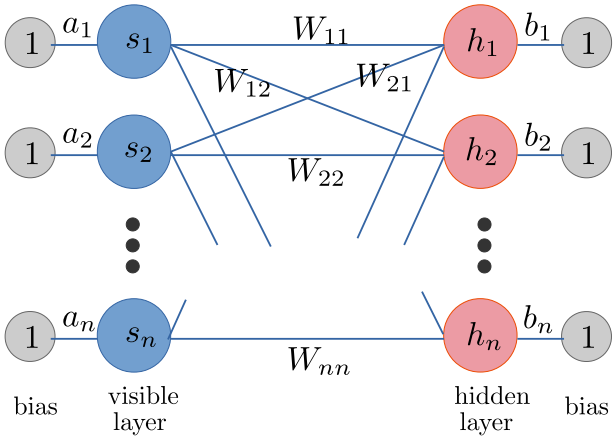


Figure 1: Schematic depiction of a restricted Boltzmann machine. Every node i in the visible layer (blue circles) is connected via a weight W_{ij} with every node j in the hidden layer (red circles). There are no intra-layer connections. In addition, all nodes are connected to bias nodes (not depicted). In our application, restricted Boltzmann machines are used to represent the amplitudes and phases of pure quantum states. Gibbs sampling can then be used to efficiently evaluate the quantum states.

References

[1] M. Paris, J. Rehacek, Quantum State Estimation, Lecture Notes in Physics, Vol. 649, Springer, Berlin, 2004.

[2] M. Hayashi, Asymptotic Theory of Quantum Statistical Inference, World Scientific, Singapore, 2005.

[3] A. I. Lvovsky, M. G. Raymer, Continuous-variable optical quantum-state tomography, Rev. Mod. Phys. 81 (2009) 299–332.

[4] Z. Hou, H.-S. Zhong, Y. Tian, D. Dong, B. Qi, L. Li, Y. Wang, F. Nori, G.-Y. Xiang, C.-F. Li, G.-C. Guo, Full reconstruction of a 14-qubit state within four hours, New J. Phys. 18 (2016) 083036.

[5] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, G. Carleo, Neural-network quantum state tomography, Nat. Phys. 14 (2018) 447.

[6] A. Melkani, C. Gneiting, F. Nori, Eigenstate extraction with neural-network tomography, Phys. Rev. A 102 (2020) 022412. URL: https://link.aps.org/doi/10.1103/PhysRevA.102.022412. doi:10.1103/PhysRevA.102.022412.

Classical variational simulation of the Quantum Approximate Optimization Algorithm

Matija Medvidović[1,2] and Giuseppe Carleo[3]

[1] *Center for Computational Quantum Physics, Flatiron Institute, New York 10010, USA*

[2] *Department of Physics, Columbia University, New York 10027, USA*

[3] *Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*

We introduce a method to simulate parametrized quantum circuits, an architecture behind many practical algorithms on near-term hardware, focusing on the Quantum Approximate Optimization Algorithm (QAOA). The algorithm can reach unexplored parameter values without requiring large-scale computational resources. Our approach can be used benchmark the next generation of quantum experiments.[1]

I. Introduction

The past decade has seen a fast development of quantum technologies and the achievement of an unprecedented level of control in quantum hardware [2], clearing the way for demonstrations of quantum computing applications for practical uses. However, near-term applications face hardware limitations (dubbed Noisy Intermediate-Scale Quantum-NISQ computers [3]) where qubit count and lack of quantum error correction constrain potential applications. Despite these limitations, hybrid classical-quantum algorithms [4–7] have been identified as candidates for practical quantum advantage [8–11]. The Quantum Approximate Optimization Algorithm (QAOA) [6] is a notable example of variational quantum algorithm with prospects of quantum speedup on NISQ devices. Built to exploit quantum effects to solve combinatorial optimization problems, it has been extensively theoretically [12–17] and experimentally [18] studied on state-of-the-art hardware and proposed as a hardware benchmark [19–22]. In this work [1], we use a (neural network) variational parametrization [23] of the many-qubit state and extend the method of Ref. [24] to simulate QAOA. This approach trades the need for expensive exact classical simulation with an approximate and accurate description. We successfully simulate the Max-Cut QAOA circuit on a 3-regular 54-qubit graph [6, 12, 18] at depth $p = 4$.

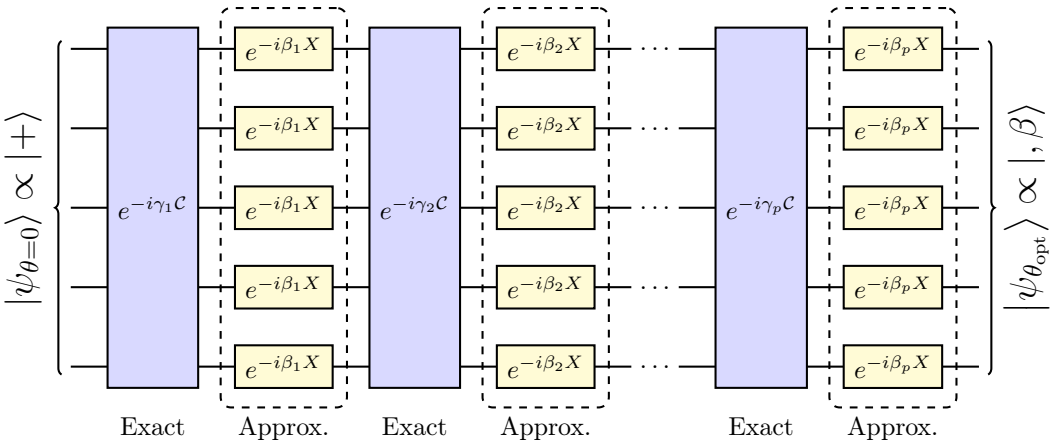


FIG. 1. **The QAOA quantum circuit.** A schematic representation of the QAOA circuit and our simulation. The input state is trivially initialized to $|+\rangle = \sum_{\mathcal{B}} |\mathcal{B}\rangle$. Next, at each p, the exchange of exactly ($U_C(\gamma) = e^{-i\gamma\mathcal{C}}$) and approximately ($RX(\beta) = e^{-i\beta X}$) applicable gates is labeled. (Here, $\mathcal{C}$ is the QAOA cost operator associated with the underlying graph.)

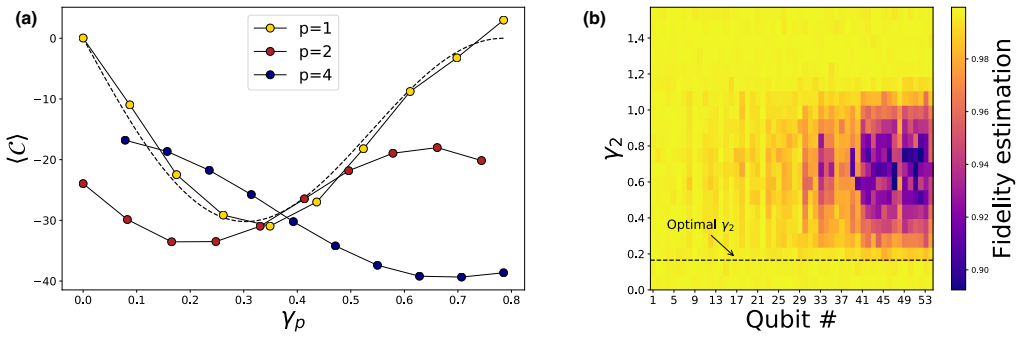


FIG. 2. **Simulating 54 qubits. a**: Randomly generated 3-regular graphs with 54 nodes are considered at $p = 1, 2, 4$. At each p, all angles were fixed except for the final γ_p. Cost dependence along a slice of the variational landscape is investigated. The dashed line represents the exact cost at $p = 1$ (see Refs. [1, 12]). **b**: Stochastic estimations of single-qubit fidelities (see Ref. [1]) in the course of optimizer progress, within a 54-qubit system at depth $p = 2$, are shown. All QAOA parameters except γ_2 are kept fixed.

II. Results

We use a neural-network representation of the many-body wavefunction $\psi_\theta(\mathcal{B})$ associated with the computational basis of classical bit strings $|\mathcal{B}\rangle$. A shallow network of the Restricted Boltzmann Machine (RBM) type is used [23, 25–27]. The advantage of using the RBM ansatz is that some one- and two-qubit gates can be applied exactly by simple analytical parameter replacements. For all other gates, we use stochastic maximization of the quantum fidelity between $\psi_\theta(\mathcal{B})$ and different target states to approximate all other gates [24, 28–30].

We can simulate QAOA by combining these two strategies (see Fig. 1). We focus on Max-Cut QAOA for 3-regular graphs [6, 13]. Details about the ansatz, optimization scheme and analytical gate implementations can be found in the full paper in Ref. [1]. Our approach can approximately simulate system sizes that are not easily reached by exact classical simulation. In Fig. 2 we show the case of $N = 54$ qubits. We closely reproduce the exact error curve using the variational optimization method. We also perform simulations at $p = 2$ and $p = 4$ and obtain corresponding approximate QAOA cost function values.

At $p = 4$, we exactly implement 324 RZZ gates and approximately implement 216 RX gates. The accuracy of our approach can be quantified using intermediate variational fidelity estimates despite the lack of exact results for systems of this size. In Fig. 2 (panel b) we show the optimal variational fidelities found when approximating the action of $RX(\beta) = e^{-i\beta X}$ gates (where X is the Pauli X gate) on the RBM wavefunction. At optimal QAOA angles, the lowest variational fidelity reached was above 98%, for a typical random graph instance shown at Fig. 2. However, for QAOA angles away from the variational optimum, we find lower fidelity estimates of the RBM variational wavefunction.

III. Conclusion

Using a novel approximate variational method, we successfully explore previously unreachable regions in the QAOA parameter space. The method is introduced as complementary to established numerical methods of classical simulation of quantum circuits because it is constrained by choices QAOA angles more than qubit counts or circuit depths. Classical variational simulations of quantum algorithms provide a natural way to both benchmark and understand the limitations of near-future quantum hardware. On the algorithmic side, our approach can help answer an open question in the field - whether QAOA can outperform classical optimization algorithms [31–33].

[1] Medvidović, M. & Carleo, G. Classical variational simulation of the Quantum Approximate Optimization Algorithm. *npj Quantum Inf.* **7**, 1–7 (2021). URL https://www.nature.com/articles/s41534-021-00440-z. 2009.01760.

[2] Arute, F. *et al.* Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).

[3] Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018). URL https://quantum-journal.org/papers/q-2018-08-06-79/. 1801.00862.

[4] Peruzzo, A. *et al.* A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**, 1–7 (2014).

[5] Farhi, E. & Neven, H. Classification with quantum neural networks on near term processors (2018). URL http://arxiv.org/abs/1802.06002. 1802.06002.

[6] Farhi, E., Goldstone, J. & Gutmann, S. A Quantum Approximate Optimization Algorithm (2014). URL http://arxiv.org/abs/1411.4028. 1411.4028.

[7] Grant, E. *et al.* Hierarchical quantum classifiers. *npj Quantum Inf.* **4**, 1–8 (2018). 1804.03680.

[8] Aspuru-Guzik, A., Dutoi, A. D., Love, P. J. & Head-Gordon, M. Chemistry: Simulated quantum computation of molecular energies. *Science* **309**, 1704–1707 (2005). URL https://science.sciencemag.org/content/309/5741/1704.

[9] O'Malley, P. J. *et al.* Scalable quantum simulation of molecular energies. *Phys. Rev. X* **6**, 031007 (2016). URL https://journals.aps.org/prx/abstract/10.1103/PhysRevX.6.031007. 1512.06860.

[10] Biamonte, J. *et al.* Quantum machine learning. *Nature* **549**, 195–202 (2017). URL https://www.nature.com/articles/nature23474. 1611.09347.

[11] Lloyd, S. Universal quantum simulators. *Science* **273**, 1073–1078 (1996).

[12] Wang, Z., Hadfield, S., Jiang, Z. & Rieffel, E. G. Quantum approximate optimization algorithm for MaxCut: A fermionic view. *Phys. Rev. A* **97**, 022304 (2018).

[13] Farhi, E., Goldstone, J. & Gutmann, S. A Quantum Approximate Optimization Algorithm Applied to a Bounded Occurrence Constraint Problem (2014). URL http://arxiv.org/abs/1412.6062. 1412.6062.

[14] Lloyd, S. Quantum approximate optimization is computationally universal (2018). URL http://arxiv.org/abs/1812.11075. 1812.11075.

[15] Jiang, Z., Rieffel, E. G. & Wang, Z. Near-optimal quantum circuit for Grover's unstructured search using a transverse field. *Phys. Rev. A* **95**, 062317 (2017). 1702.02577.

[16] Hadfield, S. *et al.* From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. *Algorithms* **12**, 34 (2019). URL http://www.mdpi.com/1999-4893/12/2/34. 1709.03489.

[17] Zhou, L., Wang, S. T., Choi, S., Pichler, H. & Lukin, M. D. Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices. *Phys. Rev. X* **10**, 21067 (2020). URL https://link.aps.org/doi/10.1103/PhysRevX.10.021067. 1812.01041.

[18] Harrigan, M. P. *et al.* Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. *Nat. Phys.* **17**, 332–336 (2021). URL http://arxiv.org/abs/2004.04197. 2004.04197.

[19] Pagano, G. *et al.* Quantum approximate optimization of the long-range Ising model with a trapped-ion quantum simulator. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 25396–25401 (2020). URL http://arxiv.org/abs/1906.02700. 1906.02700.

[20] Bengtsson, A. *et al.* Improved success probability with greater circuit depth for the quantum approximate optimization algorithm. *Phys. Rev. Appl.* **14** (2020). URL http://arxiv.org/abs/1912.10495. 1912.10495.

[21] Willsch, M., Willsch, D., Jin, F., De Raedt, H. & Michielsen, K. Benchmarking the quantum approximate optimization algorithm. *Quantum Inf. Process.* **19**, 1–24 (2020). 1907.02359.

[22] Otterbach, J. S. *et al.* Unsupervised machine learning on a hybrid quantum computer (2017). URL http://arxiv.org/abs/1712.05771. 1712.05771.

[23] Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017).

[24] Jónsson, B., Bauer, B. & Carleo, G. Neural-network states for the classical simulation of quantum computing. *arXiv* (2018). URL http://arxiv.org/abs/1808.05232. 1808.05232.

[25] Hinton, G. E. Training products of experts by minimizing contrastive divergence. *Neural Comput.* **14**, 1771–1800 (2002).
[26] Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006).
[27] Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
[28] Sorella, S. Green function monte carlo with stochastic reconfiguration. *Phys. Rev. Lett.* **80**, 4558–4561 (1998). 9803107.
[29] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).
[30] Hastings, W. K. Monte carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109 (1970).
[31] Gomes, J., Eastman, P., McKiernan, K. A. & Pande, V. S. Classical quantum optimization with neural network quantum states (2019). URL `http://arxiv.org/abs/1910.10675`. 1910.10675.
[32] Zhao, T., Carleo, G., Stokes, J. & Veerapaneni, S. Natural evolution strategies and variational Monte Carlo. *Mach. Learn. Sci. Technol.* **2**, 2–3 (2020). URL `https://doi.org/10.1088/2632-2153/abcb50`.
[33] Hibat-Allah, M., Inack, E. M., Wiersema, R., Melko, R. G. & Carrasquilla, J. Variational Neural Annealing (2021). URL `http://arxiv.org/abs/2101.10154`. 2101.10154.

Encoding-dependent generalization bounds for parametrized quantum circuits

Matthias C. Caro,[1,2] Elies Gil-Fuster,[3] Johannes Jakob Meyer,[3,4] Jens Eisert,[3,5] and Ryan Sweke[3]

[1]*Department of Mathematics, Technical University of Munich, 85748 Garching, Germany*
[2]*Munich Center for Quantum Science and Technology (MCQST), 80799 Munich, Germany*
[3]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*
[4]*QMATH, Department of Mathematical Sciences, University of Copenhagen, 2100 Copenhagen, Denmark*
[5]*Helmholtz-Zentrum Berlin für Materialien und Energie, 14109 Berlin, Germany*

(Dated: October 19, 2021)

We complement recent developments of theoretical performance guarantees for learning models based on parametrized quantum circuits (PQCs). Namely, we prove the first generalization bounds that explicitly take the encoding of the classical input into account. These bounds allow the selection of suitable PQC-based learning models through structural risk minimization.

1. INTRODUCTION AND MOTIVATION

Recent years have witnessed a surge of interest in using quantum computers for machine learning [5, 13]. Given the limitations of noisy intermediate-scale quantum (NISQ) devices, most approaches to near-term quantum-enhanced machine learning employ hybrid quantum-classical algorithms [4]. Of particular prominence are variational quantum algorithms in which a *parametrized quantum circuit* (PQC), schematically depicted in Fig. 1, defines a machine learning model which is then updated via a classical optimizer [3, 11, 18]. The flexibility in design choices for PQCs is often only perceived in terms of the structure and layout of the trainable gates [17, 24]. In comparison, the data-encoding strategy, crucial when using a PQC for machine learning with *classical data*, has received little attention. Despite this, it has recently been shown that the data-encoding strategy is directly related to the expressive power of PQC-based models [14, 20, 22, 23]. In this work, we further the study of data-encoding strategies for PQC-based supervised learning models by investigating the effect of data-encoding strategies on *generalization* performance.

More specifically, we consider the following fundamental question: Given a PQC-based model, trained on a specific data set, can we place any guarantees on its expected performance on unseen data, the *out-of-sample* performance? This question is motivated by the key insight that one should *not* choose the model or architecture which performs best on the available training data, but rather the model for which one expects the best out-of-sample performance. Typically, one refers to the difference between the accuracy of a model on a given training set and its expected out-of-sample accuracy as the *generalization gap*. We call a (probabilistic) upper bound on this generalization gap a *generalization bound*. These bounds are a central object of study in *statistical learning theory* [6, 19, 21].

2. RESULTS AND METHODS

The two main results of this work are as follows: First, we provide an explicitly encoding-dependent generalization bound for any PQC-based machine learning model. This implies rigorous out-of-sample performance guarantees, given the performance on the training data. Second, we showcase the generalization guarantee implied by our bound for different commonly used data-encoding strategies. This demonstrates how our bounds can be used to guide the choice of data-encoding.

Our general proof strategy is as follows: We bound complexity measures of the class of functions that a PQC-based machine learning model can implement. These complexity bounds then imply generalization bounds. As a first step in our proofs, we represent functions implemented by PQC-based models in terms of *generalized trigonometric polynomials* (GTPs) [23]. The

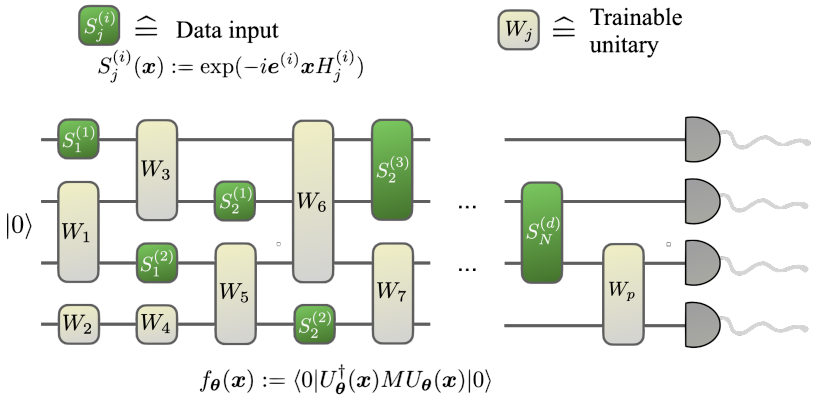


Figure 1. PQC with gates parametrized either by the data $\boldsymbol{x}$ (data-encoding gates), or the trainable parameters $\boldsymbol{\theta}$ (trainable gates).

latter have the same general form as standard trigonometric polynomials, but we allow for real-valued frequencies. Crucially, the data-encoding and the trainable part of the PQC dissociate in that representation: The admissible frequencies in the GTP depend only on the data-encoding, the coefficients in the GTP depend only on the trainable gates and the measurement in the PQC. Therefore, we can study the effect of the data-encoding on the complexity of PQC-based machine learning models via complexity bounds for GTPs in terms of the admissible frequencies. We give two proof strategies for achieving the latter, one based on known empirical Rademacher complexity bounds for classical neural networks, the other based on covering numbers.

Combining the connection between PQCs and GTPs with our generalization bounds for GTPs leads to our first main result, explicitly encoding-dependent generalization bounds for PQC-based machine learning. Thereby, we have reduced the problem of obtaining generalization guarantees for PQCs to that of bounding the number of admissible frequencies in GTP corresponding to the PQC. In our paper, we explicitly describe how to derive these frequency sets from the spectra of the data-encoding Hamiltonians, which is a purely combinatorial problem. We analyze different families of data-encodings that are commonly used in quantum machine learning. For several such strategies, we show a favourable (polynomial) worst-case scaling of the generalization error in terms of the overall number of data-encoding gates used in the PQC. However, we also demonstrate that other strategies lead to an unfavourable (exponential) scaling. This underlines the importance of the choice of encoding.

3. IMPACT

There has recently been a stream of works deriving generalization bounds for PQC-based models [1, 2, 7–10, 12, 15, 16]. However, our guarantees are the first for general PQC-based models incorporating *data re-uploading* [20], in which trainable circuit blocks are interleaved with data-encoding circuit blocks. Moreover, many prior works derived generalization bounds which depend only on properties of the trainable part of the PQC. They are thus unable to offer insight into the effect of the data-encoding on generalization. Some prior works do indeed derive generalization bounds with an encoding-dependence. However, in all the works so far, this dependence has been *implicit*, the dependence on the data-encoding strategy was not a priori clear. In this case, one cannot straightforwardly use such bounds to determine what effect a change in the data-encoding strategy will have on the generalization performance of the associated PQC-based model. Our generalization bounds for PQC-based models are *explicitly* encoding-dependent: They depend explicitly on natural hyper-parameters of the data-encoding strategy. As such, we provide a clear understanding of the effect that changes to the data-encoding strategy will have on generalization performance.

Our generalization bounds can be used complementarily to prior work. We illustrate this complementarity by suggesting "multi-dimensional structural risk minimization." Basic structural risk minimization (SRM) assumes that the model class has a tunable complexity parameter and a corresponding generalization bound. Increasing this complexity will typically allow for a smaller training error, but at the cost of larger generalization error. Therefore, as illustrated in Fig. 2, SRM balances these two phenomena to identify an optimal complexity parameter. Our suggestion of multi-dimensional SRM for quantum machine learning is the following: We can naturally consider two sources of complexity for PQC-based models, the trainable part and the data-encoding. Prior work has made progress on understanding the effect of the former on generalization, our work provides new insight into the latter. Thus, combining our encoding-dependent results with prior work leads to a rigorous foundation for architectural design choices for PQCs in machine learning that takes both of the natural sources of complexity into account.

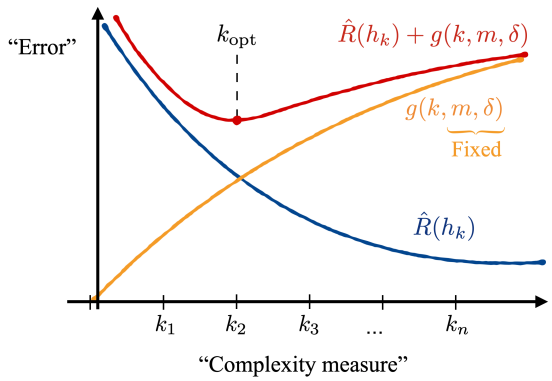


Figure 2. Illustration of SRM [19].

Variational QML based on PQCs is a promising area of application for NISQ devices. However, a theoretical understanding of the prospects of variational QML is only just developing. Our work is the first to explicitly show that the choice of classical-to-quantum data-encoding for PQCs is crucial to the generalization capabilities of the corresponding QML model. Thereby, we enrich theoretical research in QML by a novel perspective on generalization. As we demonstrate with our proposal of multi-dimensional SRM, this paves the way for more advanced studies of generalization in QML, where the complementary aspects of trainable part and data-encoding are combined. Finally, based on our theory, we also propose numerical experiments, which we envision to further illuminate the expressivity of PQCs, in particular compared to classical machine learning models.

[1] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner. The power of quantum neural networks. *Nature Computational Science*, 1(6):403–409, 2021.
[2] L. Banchi, J. Pereira, and S. Pirandola. Generalization in quantum machine learning: A quantum information perspective. *arXiv:2102.08991*, 2021.
[3] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini. Parameterized quantum circuits as machine learning models. *Quant. Sc. Tech.*, 4(4):043001, Nov 2019.
[4] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik. Noisy intermediate-scale quantum (NISQ) algorithms. *arXiv:2101.08448*, 2021.
[5] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd. Quantum machine learning. *Nature*, 549:195–202, 2017.
[6] C. Bishop. *Pattern recognition and machine learning*. Springer, Berlin, 2006.
[7] K. Bu, D. E. Koh, L. L., Q. Luo, and Y. Zhang. Rademacher complexity of noisy quantum circuits. *arXiv:2103.03139*, 2021.
[8] K. Bu, D. E. Koh, L. Li, Q. Luo, and Y. Zhang. Effects of quantum resources on the statistical complexity of quantum circuits. *arXiv:2102.03282*, 2021.
[9] K. Bu, D. E. Koh, L. Li, Q. Luo, and Y. Zhang. On the statistical complexity of quantum circuits. *arXiv:2101.06154*, 2021.
[10] M. C. Caro and I. Datta. Pseudo-dimension of quantum circuits. *Quant. Mach. Int.*, 2:172, 2020.
[11] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles. Variational quantum algorithms. *arXiv:2012.09265*, 2020.
[12] Y. Du, Z. Tu, X. Yuan, and D. Tao. An efficient measure for the expressivity of variational quantum algorithms. *arXiv:2104.09961*, 2021.
[13] V. Dunjko and H. J. Briegel. Machine learning & artificial intelligence in the quantum domain: A review of recent progress. *Rep. Prog. Phys.*, 81(7):074001, 2018.
[14] F. J. Gil Vidal and D. O. Theis. Input redundancy for parameterized quantum circuits. *Front. Phys.*, 8, 2020. Publisher: Frontiers.
[15] C. Gyurik, D. van Vreumingen, and V. Dunjko. Structural risk minimization for quantum linear classifiers. *arXiv:2105.05566*, 2021.
[16] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean. Power of data in quantum machine learning. *Nature Comm.*, 12:1–9, 2021.
[17] T. Hubregtsen, J. Pichlmeier, P. Stecher, and K. Bertels. Evaluation of parameterized quantum circuits: on the relation between classification accuracy, expressibility, and entangling capability. *Quant. Mach. Int.*, 3(1):1–19, 2021.
[18] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.*, 18:023023, 2016.
[19] M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of machine learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2 edition, 2018.
[20] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre. Data re-uploading for a universal quantum classifier. *Quantum*, 4:226, 2020.
[21] B. Schölkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Adaptive computation and machine learning. MIT Press, Cambridge, MA, 2002.
[22] M. Schuld. Quantum machine learning models are kernel methods. *arXiv:2101.11020*, 2021.
[23] M. Schuld, R. Sweke, and J. J. Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Phys. Rev. A*, 103:032430, 2021.
[24] S. Sim, P. D. Johnson, and A. Aspuru-Guzik. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Adv. Quant. Tech.*, 2:1900070, 2019.

Note: A technical version of this work is available as arXiv:2106.03880.

Generalization in quantum machine learning from few training data

Matthias C. Caro,[1,2] Hsin-Yuan Huang,[3,4] Marco Cerezo,[5,6] Kunal Sharma,[7] Andrew Sornborger,[8,9] Lukasz Cincio,[5] and Patrick J. Coles[5]

[1] *Department of Mathematics, Technical University of Munich, Garching, Germany*
[2] *Munich Center for Quantum Science and Technology (MCQST), Munich, Germany*
[3] *Institute for Quantum Information and Matter, Caltech, Pasadena, CA, USA*
[4] *Department of Computing and Mathematical Sciences, Caltech, Pasadena, CA, USA*
[5] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[6] *Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, USA*
[7] *Joint Center for Quantum Information and Computer Science, University of Maryland, College Park, MD 20742, USA*
[8] *Information Sciences, Los Alamos National Laboratory, Los Alamos, NM, USA*
[9] *Quantum Science Center, Oak Ridge, TN, 37931, USA*

Modern quantum machine learning (QML) involves optimizing a parameterized quantum circuit on training data, and subsequently makes predictions on testing data. We comprehensively study the generalization performance in QML after training on few data. We support our theory by numerical experiments for quantum phase recognition and unitary compiling.

I. Background and Motivation

The ultimate goal of machine learning (ML) is to make accurate predictions on unseen data. This is known as generalization, and significant effort has been expended to understand the generalization capabilities of classical ML models. For example, theoretical results have been formulated as upper bounds on the generalization error as a function of the training data size and the model complexity. Such bounds provide guidance as to how much training data is required and/or sufficient to achieve accurate generalization.

Quantum machine learning (QML) is an emerging field that has generated great excitement [1–4]. Modern QML involves training a parameterized quantum circuit in order to analyze either classical or quantum data sets [5–9]. Early results indicate that QML models (QMLMs) may offer some advantage over classical models under certain circumstances for classical data analysis [10, 11]. It has also been proven that QMLMs provide an exponential advantage in analyzing quantum data [12, 13]. However, little is known about the conditions needed for QMLMs to accurately generalize from training data to previously unseen data. Significant progress has been made in understanding the trainability of QMLMs [11, 14–26], but trainability is a separate question from generalization [11, 27, 28]. Overfitting of training data could be an issue for QML, just as it is for classical machine learning. We provide a comprehensive study for the generalization performance in QML, and thereby show how QML can avoid overfitting.

II. Results and Methods

We prove bounds on the generalization error in variational QML: The difference between the true performance and the training performance is approximately upper bounded by $\sqrt{T/N}$, with T the number of trainable gates and N the training data size. Importantly, this implies that an efficiently implementable QMLM, with $T \in \mathcal{O}(\text{poly}\, n)$, only requires an efficient amount of training data, $N \in \mathcal{O}(\text{poly}\, n)$, to obtain good generalization. Here, n is the number of qubits. This implication, by itself, will improve the efficiency guarantees of variational quantum algorithms [5, 29, 30] that employ training data, such as quantum autoencoders [8], quantum generative adversarial networks [31], variational quantum error correction [32, 33], variational quantum compiling [34, 35], and variational dynamical simulation [36–39]. It also yields improved guarantees for classical algorithms that simulate QMLMs. We furthermore refine our bounds to account for gate-sharing, variable circuit architecture, and the optimization process, see Fig. 1(a)-(e).

We showcase two applications of our theory. First we use quantum convolutional neural networks (QCNNs) [33] for quantum phase recognition (QPR). QCNNs have only $T = \mathcal{O}(\log n)$ parameters, thus our results give a strong generalization guarantee. In support of this, we demonstrate numerically that QCNNs have good generalization error for QPR with only logarithmic training resources, $N \in \mathcal{O}(\log^2 n)$. Namely, we use QCNN architectures for QPR on ground states of the generalized cluster Hamiltonian of up to 64 qubits. The ground state phase diagram of this Hamiltonian has four phases: trivial, ferromagnetic phase, anti-ferromagnetic, and symmetry-protected topological [40]. Fig. 1(f) shows the performance of the QCNN in classifying ground states into these four phases. Second, we highlight the task of quantum compiling, a crucial application for the quantum computing industry. State-of-the-art classical methods for approximate optimal compiling of unitaries often employ exponentially large training data sets [41–43]. However,

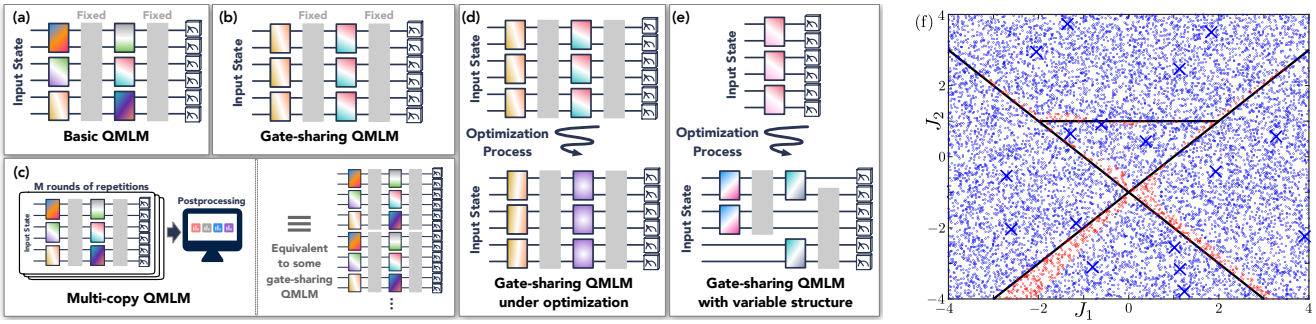


FIG. 1: (a)-(e) Illustration for the various types of QMLM considered in this work. (f) Performance of the QCNN for quantum phase recognition on ground states of the generalized cluster Hamiltonian. Blue (red) indicates correct (incorrect) classification. Crosses denote the training data points. Thick black lines delimit the five regions of the underlying phase diagram, with four different phases.

our work indicates that only polynomial-sized data sets are needed, suggesting that state-of-the-art compilers could be further improved. Indeed, we numerically demonstrate that compiling the quantum Fourier transform requires only linearly many training data points. With this significantly reduced cost, we achieve successful compilation for up to 10 qubits. Assuming favourably initialized training, we compile the QFT on up to 40 qubits.

III. Impact and Outlook

Our work injects new hope into the field of QML, as good generalization is guaranteed from few training data for any efficiently implementable QMLM. We have supported our general theoretical results by promising numerical demonstrations for two specific applications of interest. Already these two applications carry a lot of potential.

Quantum phase classification is an exciting application of QML. While Ref. [33] has already successfully applied QCNNs to this problem, our work is the first to rigorously prove their good performance, as a special case of our general theory. Moreover, our analysis allows us to go beyond QCNNs and extract general principles for how to ensure good generalization. As generating training data for this problem asks an experimenter to prepare a variety of states from different phases of matter, which will require careful tuning of different parameters in the underlying Hamiltonian, good generalization guarantees for small training data sizes are crucial to allow for the implementation of phase classification through QML in actual physical experiments.

Several successful protocols for unitary compiling make use of training data [41–43]. However, prior work has relied on exponentially large training data sets. Such large data sets are problematic: They lead to similarly high computational complexity and are expensive to obtain in physical experiments. Our results provide guarantees on the performance of unitary compiling with only polynomial-size training data, for the relevant case of efficiently implementable unitaries. As we have numerically demonstrated for the Quantum Fourier Transform, this reduction in training data size makes unitary compiling scalable. Moreover, our results provide new insight into why the VAns algorithm [44] is successful for unitary compiling. We believe that the QML perspective on unitary compiling advocated for in this work will lead to new and improved ansätze, which could scale to even larger systems.

We envision a variety of further applications for our theory First, recent methods for variational dynamical simulation rely on quantum compiling to compile a Trotterized unitary into a structured ansatz with the form of a diagonalization [36, 37, 45, 46]. This allows for quantum simulations of times longer than an iterated Trotterization. We expect our quantum compiling results to carry over to this application and thus allow these variational quantum simulation methods to use fewer training resources. Second, discovering quantum error correcting codes can be viewed as a ML problem [32, 33, 47–51]. Both classical [47–51] and near-term quantum approaches [32, 33] to this problem can benefit from our generalization bounds and enjoy reduced training data requirements. Finally, autoencoders and generative adversarial networks (GANs) have recently been generalized to the quantum setting [8, 31, 52, 53]. Both employ training data, and hence our generalization bounds provide quantitative guidance for how much training data to employ in these applications. Moreover, our results can provide guidance for ansatz design in these settings.

Our results do not prove a quantum advantage of quantum over classical machine learning. However, generalization bounds for QMLMs are necessary to understand their potential for quantum advantage: QMLMs can outperform classical methods, assuming both achieve small training error, only when QMLMs generalize well, but classical ML methods do not. We therefore consider our results a guide in the search for quantum advantage of QML.

[1] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd, "Quantum machine learning," Nature **549**, 195–202 (2017).
[2] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione, "An introduction to quantum machine learning," Contemporary Physics **56**, 172–185 (2015).
[3] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione, "The quest for a quantum neural network," Quantum Information Processing **13**, 2567–2586 (2014).
[4] Vedran Dunjko and Hans J Briegel, "Machine learning & artificial intelligence in the quantum domain: a review of recent progress," Reports on Progress in Physics **81**, 074001 (2018).
[5] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles, "Variational quantum algorithms," Nature Reviews Physics **3**, 625–644 (2021).
[6] Vojtěch Havlíček, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta, "Supervised learning with quantum-enhanced feature spaces," Nature **567**, 209–212 (2019).
[7] Edward Farhi and Hartmut Neven, "Classification with quantum neural networks on near term processors," arXiv preprint arXiv:1802.06002 (2018).
[8] Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik, "Quantum autoencoders for efficient compression of quantum data," Quantum Science and Technology **2**, 045001 (2017).
[9] Kwok Ho Wan, Oscar Dahlsten, Hlér Kristjánsson, Robert Gardner, and MS Kim, "Quantum generalisation of feedforward neural networks," npj Quantum information **3**, 1–8 (2017).
[10] Hsin-Yuan Huang, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, and Jarrod R McClean, "Power of data in quantum machine learning," Nature Communications **12**, 1–9 (2021).
[11] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner, "The power of quantum neural networks," Nature Computational Science **1**, 403–409 (2021).
[12] Hsin-Yuan Huang, Richard Kueng, and John Preskill, "Information-theoretic bounds on quantum advantage in machine learning," Phys. Rev. Lett. **126**, 190505 (2021).
[13] Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi, "Quantum algorithmic measurement," arXiv preprint arXiv:2101.04634 (2021).
[14] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven, "Barren plateaus in quantum neural network training landscapes," Nature communications **9**, 1–6 (2018).
[15] M Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles, "Cost function dependent barren plateaus in shallow parametrized quantum circuits," Nature communications **12**, 1–12 (2021).
[16] M. Cerezo and Patrick J Coles, "Higher order derivatives of quantum neural networks with barren plateaus," Quantum Science and Technology **6**, 035006 (2021).
[17] Andrew Arrasmith, M. Cerezo, Piotr Czarnik, Lukasz Cincio, and Patrick J Coles, "Effect of barren plateaus on gradient-free optimization," Quantum **5**, 558 (2021).
[18] Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J Coles, "Connecting ansatz expressibility to gradient magnitudes and barren plateaus," arXiv preprint arXiv:2101.02138 (2021).
[19] Arthur Pesah, M. Cerezo, Samson Wang, Tyler Volkoff, Andrew T Sornborger, and Patrick J Coles, "Absence of barren plateaus in quantum convolutional neural networks," arXiv preprint arXiv:2011.02966 (2020).
[20] Tyler Volkoff and Patrick J Coles, "Large gradients via correlation in random parameterized quantum circuits," Quantum Science and Technology **6**, 025008 (2021).
[21] Kunal Sharma, M. Cerezo, Lukasz Cincio, and Patrick J Coles, "Trainability of dissipative perceptron-based quantum neural networks," arXiv preprint arXiv:2005.12458 (2020).
[22] Zoë Holmes, Andrew Arrasmith, Bin Yan, Patrick J Coles, Andreas Albrecht, and Andrew T Sornborger, "Barren plateaus preclude learning scramblers," Physical Review Letters **126**, 190501 (2021).
[23] Carlos Ortiz Marrero, Mária Kieferová, and Nathan Wiebe, "Entanglement induced barren plateaus," arXiv preprint arXiv:2010.15968 (2020).
[24] AV Uvarov and Jacob D Biamonte, "On barren plateaus and cost function locality in variational quantum algorithms," Journal of Physics A: Mathematical and Theoretical **54**, 245301 (2021).
[25] Taylor L Patti, Khadijeh Najafi, Xun Gao, and Susanne F Yelin, "Entanglement devised barren plateau mitigation," Physical Review Research **3**, 033090 (2021).
[26] Samson Wang, Enrico Fontana, M. Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J Coles, "Noise-induced barren plateaus in variational quantum algorithms," arXiv preprint arXiv:2007.14384 (2020).
[27] Leonardo Banchi, Jason Pereira, and Stefano Pirandola, "Generalization in quantum machine learning: a quantum information perspective," arXiv preprint arXiv:2102.08991 (2021).
[28] Yuxuan Du, Zhuozhuo Tu, Xiao Yuan, and Dacheng Tao, "An efficient measure for the expressivity of variational quantum algorithms," arXiv preprint arXiv:2104.09961 (2021).
[29] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote,

Hermanni Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik, "Noisy intermediate-scale quantum (nisq) algorithms," arXiv preprint arXiv:2101.08448 (2021).
[30] Suguru Endo, Zhenyu Cai, Simon C Benjamin, and Xiao Yuan, "Hybrid quantum-classical algorithms and quantum error mitigation," Journal of the Physical Society of Japan **90**, 032001 (2021).
[31] Jonathan Romero and Alán Aspuru-Guzik, "Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions," Advanced Quantum Technologies **4**, 2000003 (2021).
[32] Peter D Johnson, Jonathan Romero, Jonathan Olson, Yudong Cao, and Alán Aspuru-Guzik, "Qvector: an algorithm for device-tailored quantum error correction," arXiv preprint arXiv:1711.02249 (2017).
[33] Iris Cong, Soonwon Choi, and Mikhail D Lukin, "Quantum convolutional neural networks," Nature Physics **15**, 1273–1278 (2019).
[34] Sumeet Khatri, Ryan LaRose, Alexander Poremba, Lukasz Cincio, Andrew T Sornborger, and Patrick J Coles, "Quantum-assisted quantum compiling," Quantum **3**, 140 (2019).
[35] Kunal Sharma, Sumeet Khatri, M. Cerezo, and Patrick J Coles, "Noise resilience of variational quantum compiling," New Journal of Physics **22**, 043006 (2020).
[36] Cristina Cirstoiu, Zoe Holmes, Joseph Iosue, Lukasz Cincio, Patrick J Coles, and Andrew Sornborger, "Variational fast forwarding for quantum simulation beyond the coherence time," npj Quantum Information **6**, 1–10 (2020).
[37] Benjamin Commeau, M. Cerezo, Zoë Holmes, Lukasz Cincio, Patrick J Coles, and Andrew Sornborger, "Variational hamiltonian diagonalization for dynamical quantum simulation," arXiv preprint arXiv:2009.02559 (2020).
[38] Suguru Endo, Jinzhao Sun, Ying Li, Simon C Benjamin, and Xiao Yuan, "Variational quantum simulation of general processes," Physical Review Letters **125**, 010501 (2020).
[39] Y. Li and S. C. Benjamin, "Efficient variational quantum simulator incorporating active error minimization," Phys. Rev. X **7**, 021050 (2017).
[40] Ruben Verresen, Roderich Moessner, and Frank Pollmann, "One-dimensional symmetry protected topological phases and their transitions," Physical Review B **96**, 165124 (2017).
[41] Lukasz Cincio, Yiğit Subaşı, Andrew T Sornborger, and Patrick J Coles, "Learning the quantum algorithm for state overlap," New Journal of Physics **20**, 113022 (2018).
[42] Lukasz Cincio, Kenneth Rudinger, Mohan Sarovar, and Patrick J. Coles, "Machine learning of noise-resilient quantum circuits," PRX Quantum **2**, 010324 (2021).
[43] E. Younis and L. Cincio, "Quantum Fast Circuit Optimizer (qFactor)," .
[44] M Bilkis, M Cerezo, Guillaume Verdon, Patrick J Coles, and Lukasz Cincio, "A semi-agnostic ansatz with variable structure for quantum machine learning," arXiv preprint arXiv:2103.06712 (2021).
[45] Joe Gibbs, Kaitlin Gili, Zoë Holmes, Benjamin Commeau, Andrew Arrasmith, Lukasz Cincio, Patrick J Coles, and Andrew Sornborger, "Long-time simulations with high fidelity on quantum hardware," arXiv preprint arXiv:2102.04313 (2021).
[46] Michael R Geller, Zoë Holmes, Patrick J Coles, and Andrew Sornborger, "Experimental quantum learning of a spectral decomposition," arXiv preprint arXiv:2104.03295 (2021).
[47] Andrew S Fletcher, Peter W Shor, and Moe Z Win, "Channel-adapted quantum error correction for the amplitude damping channel," IEEE Transactions on Information Theory **54**, 5705–5718 (2008).
[48] Andrew S Fletcher, Peter W Shor, and Moe Z Win, "Structured near-optimal channel-adapted quantum error correction," Physical Review A **77**, 012320 (2008).
[49] Robert L Kosut, Alireza Shabani, and Daniel A Lidar, "Robust quantum error correction via convex optimization," Physical review letters **100**, 020502 (2008).
[50] Robert L Kosut and Daniel A Lidar, "Quantum error correction via convex optimization," Quantum Information Processing **8**, 443–459 (2009).
[51] Soraya Taghavi, Robert L Kosut, and Daniel A Lidar, "Channel-optimized quantum error correction," IEEE transactions on information theory **56**, 1461–1473 (2010).
[52] Seth Lloyd and Christian Weedbrook, "Quantum generative adversarial learning," Physical review letters **121**, 040502 (2018).
[53] Pierre-Luc Dallaire-Demers and Nathan Killoran, "Quantum generative adversarial networks," Physical Review A **98**, 012324 (2018).

Simultaneous Perturbation Stochastic Approximation of the Quantum Fisher Information

Julien Gacon,[1,2,*] Christa Zoufal,[1,3] Giuseppe Carleo,[2] and Stefan Woerner[1]

[1] *IBM Quantum, IBM Research – Zurich*
[2] *Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL)*
[3] *Institute for Theoretical Physics, ETH Zurich*

(Dated: August 9, 2021)

The Quantum Fisher Information matrix (QFIM) is a central metric in promising near-term algorithms, such as Variational Quantum Imaginary Time Evolution. Calculating the QFIM, however, is computationally expensive and scales quadratically with the number of model parameters. Here we propose sampling techniques to approximate the QFIM at a constant cost.

While fault-tolerant quantum computers are not yet available, a computational paradigm particularly suitable for near-term, noisy quantum devices is that of variational quantum algorithms. In this context, Variational Imaginary Time Evolution (VarQITE) and, closely related, variational ground state search using Quantum Natural Gradient Descent (QNG) are particularly promising algorithms that have recently received a lot of interest [1–5]. These iterative algorithms employ a parameterized quantum circuit as a trainable model where the number of tunable parameters d grows with the complexity of the molecule or the size of the dataset. The computational cost of each iteration step is quickly limited by the evaluation of the Quantum Fisher Information matrix (QFIM), which, for a model with d parameters, generally requires $\mathcal{O}(d^2)$ function evaluations. While still manageable for small problems we are approaching hardware dimensions that allow us to simulate increasingly larger systems where this quadratic scaling becomes the bottleneck. Diagonal and block-diagonal approximations of the QFIM have been proposed [3], which reduce the cost from quadratic to linear in d. However, these methods do not allow access to the full QFIM but remain restricted to the (block-)diagonals and thereby do not properly capture parameter correlations.

To remedy the increasing costs in high-dimensional parameter spaces, we propose using simultaneous perturbation stochastic approximation (SPSA) [6, 7] techniques to approximate the QFIM at a constant cost. A similar idea has previously been explored for the classical Fisher Information in context of the Expectation-Maximization algorithm [8]. By tuning the number of stochastic samples, our approach allows a flexible trade-off between asymptotically exact evaluation of the QFIM, which is important for time evolution algorithms like VarQITE, and reduced computational cost, which can be advantageous for ground-state searches.

Our approach is particularly efficient if no exact state evolution is required, as is e.g. the case for ground-state search. This allows us to iteratively construct a QFIM estimate with only a small number of stochastic samples in each iteration. Further it's important to note that the estimate of the QFIM is independent of the Hamiltonian we evolve. If the system's Hamiltonian consists of many non-commuting Pauli groups, and therefore requires sampling a large number of circuits, the overhead to estimate the QFIM is negligible.

We test our algorithm first in simulations and then on real hardware. In the simulations, we find the solution of a classical MAXCUT problem and learn a probability distribution using generative learning and Variational Quantum Boltzmann machines [4]. On real hardware, we prepare the ground state of the LiH molecule for a given bond distance. The experiments show that our method is able to approximate VarQITE and QNG reliably at a fraction of the computational

* jul@zurich.ibm.com

cost. As a consequence, we show that our proposed technique outperforms other standard gradient-based algorithms for variational ground state search. The advantage is especially prominent if the expectation value calculation of the system energy requires a large number of circuits, as is frequently the case for molecular Hamiltonians.

The preprint of our paper is available at [9].

---

[1] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Theory of variational quantum simulation, Quantum **3**, 191 (2019).

[2] S. McArdle *et al.*, Variational ansatz-based quantum simulation of imaginary time evolution, npj Quantum Information **5**, 10.1038/s41534-019-0187-2 (2019).

[3] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum natural gradient, Quantum **4**, 269 (2020).

[4] C. Zoufal, A. Lucchi, and S. Woerner, Variational quantum boltzmann machines, Quantum Machine Intelligence **3**, 7 (2020).

[5] D. Wierichs, C. Gogolin, and M. Kastoryano, Avoiding local minima in variational quantum eigensolvers with the natural gradient optimizer, Physical Review Research **2**, 043246 (2020).

[6] J. C. Spall, An overview of the simultaneous perturbation method for efficient optimization, Johns Hopkins APL Technical Digest **19**, 11 (1998).

[7] J. C. Spall, Accelerated second-order stochastic optimization using only function measurements, in *Proceedings of the 36th IEEE Conference on Decision and Control*, Vol. 2 (1997) pp. 1417–1424 vol.2, iSSN: 0191-2216.

[8] L. Meng and J. C. Spall, Efficient computation of the fisher information matrix in the em algorithm, in *2017 51st Annual Conference on Information Sciences and Systems (CISS)* (2017) pp. 1–6.

[9] J. Gacon, C. Zoufal, G. Carleo, and S. Woerner, Simultaneous Perturbation Stochastic Approximation of the Quantum Fisher Information, Preprint at arXiv:2103.09232 (2021).

Configurable sublinear circuits for quantum state preparation

Israel F. Araujo[1], Daniel K. Park[2], Teresa B. Ludermir[1], Wilson R. Oliveira[3], Francesco Petruccione[4], and Adenilton J. da Silva[1]

[1]Centro de Informática, Universidade Federal de Pernambuco, Recife, PE, Brazil
[2]Sungkyunkwan University Advanced Institute of Nanotechnology, Suwon, South Korea
[3]Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Recife, PE, Brazil
[4]Quantum Research Group, School of Chemistry and Physics, University of KwaZulu-Natal, Durban, KwaZulu-Natal, South Africa

Full paper: https://arxiv.org/abs/2108.10182.

The theory of quantum algorithms promises unprecedented benefits of harnessing the laws of quantum mechanics for solving certain computational problems. A persistent obstacle to using such algorithms for solving a wide range of real-world problems is the cost of loading classical data to a quantum state. The worst case complexity of preparing an arbitrary quantum state is exponential with the number of qubits [1]. For this reason, the most significant quantum speed-ups occur when the quantum algorithm [2–7] operates on an input state that is easy to prepare, such as the uniform superposition of all computational basis states. For algorithms that rely on loading data into an arbitrary quantum superposition state, an efficient means to prepare input states is a prerequisite to quantum speed-ups [8–11].

Several solutions to the problem of quantum state preparation have been proposed [1, 12–17], but all produce circuits with width or depth growing at least linearly with the size of the input vector [1]. For example, the top-down method proposed in Ref. [12] achieves the exponential compression of the quantum circuit width while requiring $O(N)$ quantum circuit depth for N-dimensional data. On the other extreme end, the bottom-up method [17] achieves the exponential compression of the quantum circuit depth while requiring $O(N)$ quantum circuit width and entangled information in ancillary qubits. Since there is an extra resource overhead in many quantum algorithms due to the quantum measurement postulate [18, 19], such linear cost can impose restrictions on possible speed-ups, dominating the computational cost of the intended quantum application.

In this work, we present a quantum state preparation method that achieves sublinear scaling on both quantum circuit resources. More specifically, we develop a bidirectional strategy that effectively combines the aforementioned approaches in a way that the trade-off between computational time and space can be configured. Both temporal and spatial complexities depend on the parameter $s \in [1, n]$, which adjusts the trade-off between computational time and space. Given an N-dimensional input vector, the total time complexity of the bidirectional algorithm is $O_c(N)+O_d(2^s+\log_2^2(N)-s^2)$, where $O_c(N)$ is the time of the classical preprocessing to create the quantum circuit and $O_d(2^s+\log_2^2(N)-s^2)$ is the quantum circuit depth. Typically the same input vector is loaded $l \gg N$ times, and hence the amortized computational time is $O_d(2^s + \log_2^2(N) - s^2)$. Note that classical preprocessing is also common in classical computing and is necessary in other quantum state preparation methods as well. The spatial complexity (i.e. the width) of the circuit is $O_w((s+1)^{N/2^s})$.

Quantum state preparation algorithms aim to create a state $\sum_p |x_p|e^{i\omega_p}|p\rangle$ that encodes a normalized vector $\mathbf{x} = (|x_0|e^{i\omega_0}, \ldots, |x_{N-1}|e^{i\omega_{N-1}})$ as the probability amplitudes. Several of the existing methods can be understood as a walk on a binary tree [1, 15, 17, 20]. Each tree node corresponds to a controlled gate operation and the height increases with the number of qubits. Two edges stemming from each node indicate that each controlled gate operation splits the Hilbert space into two

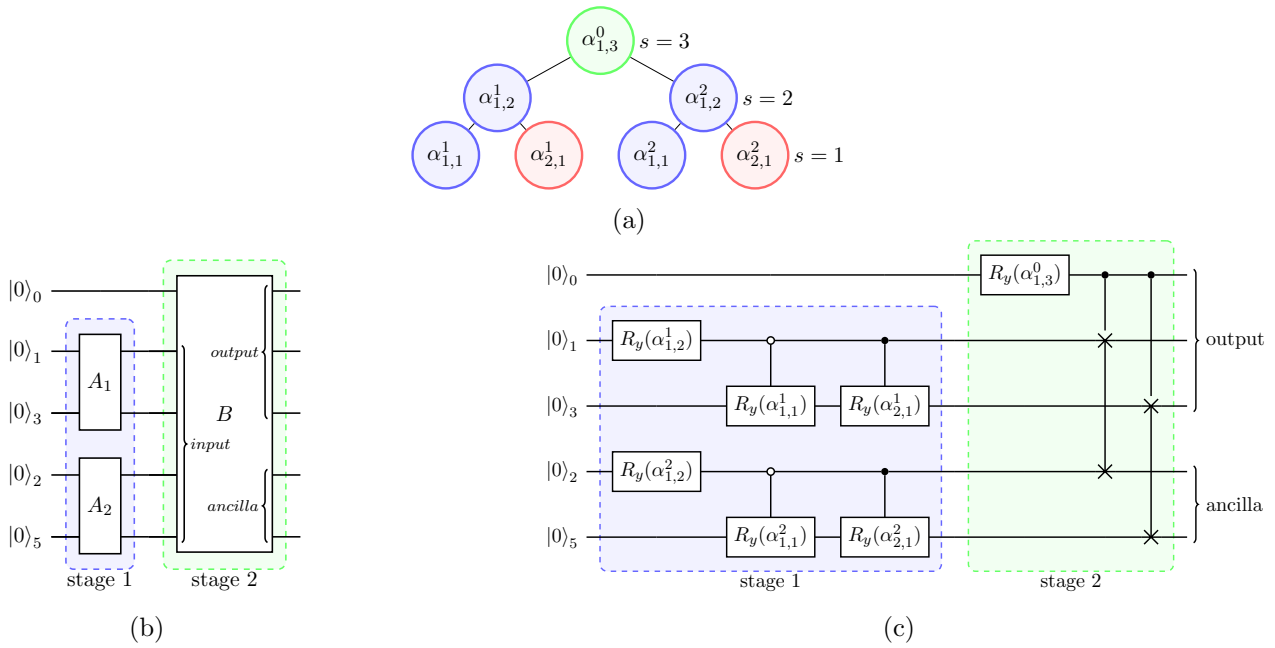


Figure 1: Schematics of the bidirectional algorithm. (a) Angle tree example with a split at level $s = 2$. The blue and red nodes (α^1 and α^2) correspond to the bidirectional procedure first stage. In each of the two sub-trees of the first stage, 4 of the 8 amplitudes expected as input by stage 2 are encoded using a top-down method. The green node (α^0) above the tree split correspond to the second stage single sub-tree, subject to a partial DCSP bottom-up procedure. The first stage red nodes ($j > 1$) are no longer associated with an ancilla since they are now encoded through a top-down approach. (b) Block diagram circuit, corresponding to the tree in (a). In stage 1, the A_k operators (the index k is related to angle vectors α^k upper index) are responsible for encoding the amplitudes that will be used as input by stage 2. In this example, each A_k operator encodes 4 amplitudes from a total of 8. The B operator is the partial DCSP for 8 amplitudes, which is initialized with the expected state for the split level 2 and continues with the traditional algorithm. (c) Detailed view of (b), generated by the bidirectional algorithm for a real and positive 8-dimensional input vector.

subspaces. Therefore, after n layers, there can be 2^n subspaces with distinct probability amplitudes. Depending on the choice of the walk direction, different state preparation strategies, such as top-down and bottom-up approaches, can be constructed. The bidirectional state preparation (BDSP) method combines both bottom-up and top-down strategies as walking on the tree in both directions.

The BDSP algorithm starts by informing a level $v = s$ (enumerated from bottom to top, where $1 \leq s \leq n$) at which the tree is split, followed by two stages. In the first stage, it segments the tree section below s into 2^{n-s} sub-trees of height s. The 2^{n-s} nodes at level s are the roots of these sub-trees. The number of sub-trees determines how many initial sub-states should be prepared in the first stage of the algorithm. The amplitude values of these sub-states $a_j = (a_{j,1}, \ldots, a_{j,2^s})$ $(1 \leq j \leq 2^{n-s})$ are loaded concurrently using a sequential algorithm [1, 15, 20] based on the TDSP method. The initial sub-states are the input of the second stage of BDSP. They reproduce the state that would be created by the bottom-up steps up to the split level s. In the second stage, the sub-states are combined to generate the complete state by the divide-and-conquer approach. The bottom-up algorithm takes the state prepared in the first stage as the input, and starts walking on the tree from the split level.

The free parameter $s \in [1, n]$ determines the balance between the top-down and the bottom-up approaches. At two extreme cases of setting $s = n$ and $s = 1$, the top-down and the bottom-up approaches are respectively recovered. At the equilibrium point $s = \lceil n/2 \rceil$, quadratic reduction in both quantum circuit depth and width can be achieved. The configuration parameter can be viewed as a hyperparameter that can tune circuit sizes and the number of CNOT gates according to the compound of application and hardware properties.

References

[1] V.V. Shende, S.S. Bullock, and I.L. Markov. Synthesis of quantum-logic circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 25(6):1000–1010, 2006.

[2] David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907): 553–558, 1992.

[3] Tad Hogg, Bernardo A. Huberman, and Colin P. Williams. Phase transitions and the search problem. *Artificial Intelligence*, 81(1):1–15, 1996. Frontiers in Problem Solving: Phase Transitions and Complexity.

[4] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of Computing*, STOC '96, pages 212–219, Philadelphia, Pennsylvania, USA, 1996. Association for Computing Machinery.

[5] Daniel R. Simon. On the Power of Quantum Computation. *SIAM Journal on Computing*, 26(5): 1474–1483, 1997.

[6] Barbara M. Terhal and John A. Smolin. Single quantum querying of a database. *Physical Review A*, 58(3):1822–1826, 1998.

[7] Peter W. Shor. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM Review*, 41(2):303–332, 1999.

[8] Scott Aaronson. Read the fine print. *Nature Physics*, 11(4):291–293, 2015.

[9] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

[10] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum Algorithm for Linear Systems of Equations. *Phys. Rev. Lett.*, 103:150502, 2009.

[11] Ewin Tang. Quantum Principal Component Analysis Only Achieves an Exponential Speedup Because of Its State Preparation Assumptions. *Phys. Rev. Lett.*, 127:060503, 2021.

[12] Dan Ventura and Tony Martinez. Initializing the Amplitude Distribution of a Quantum State. *Foundations of Physics Letters*, 12(6):547–559, 1999.

[13] Lov K. Grover. Synthesis of Quantum Superpositions by Quantum Computation. *Physical Review Letters*, 85(6):1334–1337, 2000.

[14] Gui-Lu Long and Yang Sun. Efficient scheme for initializing a quantum register with an arbitrary superposed state. *Physical Review A*, 64(1):014303, 2001.

[15] Mikko Mottonen, Juha J. Vartiainen, Ville Bergholm, and Martti M. Salomaa. Transformation of Quantum States Using Uniformly Controlled Rotations. *Quantum Info. Comput.*, 5(6):467–473, 2005.

[16] Martin Plesch and Časlav Brukner. Quantum-state preparation with universal gate decompositions. *Physical Review A*, 83(3):032302, 2011.

[17] Israel F. Araujo, Daniel K. Park, Francesco Petruccione, and Adenilton J. da Silva. A divide-and-conquer algorithm for quantum state preparation. *Scientific Reports*, 11(1):6329, 2021.

[18] Daniel K. Park, Francesco Petruccione, and June-Koo Kevin Rhee. Circuit-Based Quantum Random Access Memory for Classical Data. *Scientific Reports*, 9(1):3949, 2019.

[19] Daniel K Park, Ilya Sinayskiy, Mark Fingerhuth, Francesco Petruccione, and June-Koo Kevin Rhee. Parallel quantum trajectories via forking for sampling without redundancy. *New Journal of Physics*, 21(8):083024, 2019.

[20] Ville Bergholm, Juha J. Vartiainen, Mikko Möttönen, and Martti M. Salomaa. Quantum circuits with uniformly controlled one-qubit gates. *Physical Review A*, 71(5):052330, 2005.

Variational Quantum Reinforcement Learning via Evolutionary Optimization

Samuel Yen-Chi Chen[1], Chih-Min Huang[2], Chia-Wei Hsing[2], Hsi-Sheng Goan[2], and Ying-Jer Kao[2]

[1]**Computational Science Initiative, Brookhaven National Laboratory, Upton, NY 11973, USA**
[2]**Department of Physics, National Taiwan University, Taipei 10617, Taiwan**

ABSTRACT

We present a quantum reinforcement learning framework optimized by evolutionary algorithm. The model combines a quantum-inspired tensor network and variational quantum circuits. We demonstrate via numerical simulation that the proposed model can process input data with dimensions larger than the capacity of existing quantum devices.

Keywords: Quantum neural networks, Reinforcement learning, Evolutionary optimization

INTRODUCTION

Recent advance in classical reinforcement learning (RL) and quantum computation (QC) points to a promising direction of performing RL on a quantum computer. However, potential applications in quantum RL are limited by the number of qubits available in the existing quantum devices. For example, most of the interesting RL testing environments are with observation or state in a dimension which is much larger than the capabilities of existing quantum computers. It is extremely difficult to load such data directly into a quantum computer. Significant improvements are needed in order to construct quantum RL agent which can successfully deal with such testing environments.

In this work we present two frameworks of deep quantum RL tasks using a gradient-free evolution optimization: First, we apply the amplitude encoding scheme to the Cart-Pole problem; Second, we propose a hybrid framework where the quantum RL agents are equipped with hybrid tensor network-variational quantum circuit (TN-VQC) architecture to handle inputs with dimensions exceeding the number of qubits. This allows us to perform quantum RL on the more complicated MiniGrid environment.

In the first part of the experiment, we employ the *amplitude encoding* to encode the classical input vector into a quantum state such that an n-dimensional vector can be encoded into a $(\log_2 n)$-qubit system. We demonstrate the quantum advantage of parameter-saving with the amplitude encoding in the Cart-Pole problem. In the second part, we construct a quantum-classical hybrid TN-VQC architecture. The tensor network part, operating on a classical computer, is used to compress the incoming state vector into a smaller vector suitable for a quantum computer to process. This smaller vector is then processed by the VQC to produce the action for the next move. The whole TN-VQC model is trained in an end-to-end manner similar to the ones described in the work Chen et al. (2021b). However, different from the previous work, here we optimize the quantum RL model via gradient-free evolutionary method which has the potential to perform better in environments with sparse rewards. We demonstrate via numerical simulation that such a hybrid TN-VQC quantum RL model can successfully learn the proper policy in the MiniGrid environments with a 147-dimension state/observation vector. In summary, the hybrid TN-VQC

architecture provides a natural way to perform efficient compression or feature extraction of observation vectors, enabling further quantum RL applications on noisy intermediate-scale quantum devices through leveraging the both the classical and quantum computing paradigms.

The result described in this abstract has been posted on the arXiv Chen et al. (2021a).

ACKNOWLEDGMENTS

This work is supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics program under Award Number DE-SC-0012704 and the Brookhaven National Laboratory LDRD #20-024 (S.Y.-C.C.), the Ministry of Science and Technology (MOST) of Taiwan under grants No. 107-2112-M-002-016-MY3, 108-2112-M-002-020-MY3 (Y.-J.K.), 109-2112-M-002-023-MY3, 109-2627-M-002-003, 107-2627-E-002-001-MY3, 109-2622-8-002-003 (H.S.G.) and the U.S. Air Force Office of Scientific Research under award number FA2386-20-1-4033, and by the National Taiwan University under Grant No. NTU-CC-110L890102 (H.S.G.).

REFERENCES

Chen, S. Y.-C., Huang, C.-M., Hsing, C.-W., Goan, H.-S., and Kao, Y.-J. (2021a). Variational quantum reinforcement learning via evolutionary optimization. *arXiv preprint arXiv:2109.00540*.

Chen, S. Y.-C., Huang, C.-M., Hsing, C.-W., and Kao, Y.-J. (2021b). An end-to-end trainable hybrid classical-quantum classifier. *Machine Learning: Science and Technology*, 2:045021.

Optimal control of quantum thermal machines with differentiable programming

Ilia Khait[1], Juan Carrasquilla[2,3], Dvira Segal[1,4]

[1] Department of Physics, University of Toronto, 60 Saint George St., Toronto, Ontario, Canada

[2] Vector Institute, MaRS Centre, Toronto, Ontario, Canada

[3] Department of Physics \& Astronomy, University of Waterloo, Waterloo, ON, Canada

[4] Department of Chemistry and Centre for Quantum Information and Quantum Control,

University of Toronto, 80 Saint George St., Toronto, Ontario, Canada

Abstract:

Identifying optimal thermodynamic processes has been the essence of thermodynamics since its inception. We harness differentiable programming to optimize finite-time processes in a quantum thermal machine. Overcoming hard physical constraints, our scheme discovers profiles that are superior to previously suggested protocols and finds flaws in a previously employed thermodynamic quantity.

Extended abstract:

Since its inception, thermodynamics has been concerned with performance optimization by identifying constraints and bounds on energy conversion processes. For example, the ideal Carnot engine is designed to reach maximal efficiency, but this upper bound is theoretically obtained for arbitrarily slow, quasistatic processes. In contrast to arbitrarily slow processes, real thermal devices operate on finite-time cycles, and they are naturally described in terms of finite-time thermodynamics [1,2]. This theory is concerned with e.g., how the efficiency of thermal machines erodes when heat-to-work conversion processes take place in finite-time cycles [3,4].

Quantum thermal machines, in which e.g., quantum coherences, correlations, and quantum statistics play a decisive role, cater fundamental understanding of thermodynamics at the nano and atomistic scale [5-6]. Beyond fundamental interest, quantum thermal machines promise compact, fast, and efficient work extraction and refrigeration schemes for quantum devices. It remains, however, a challenge to harness such effects and achieve a quantum advantage in thermal machines [7-11].

Optimizing the performance of *nanoscale quantum* thermal machines is a central problem in the rapidly emerging field of quantum thermodynamics. Techniques such as shortcut-to-adiabaticity (STA) allow the

design of finite-time protocols, which reproduce the same final state of an adiabatic time evolution, yet at a price of supplemental work on the system [12-16].

In this work, we harness state-of-the-art machine learning (ML) techniques to optimize the performance of quantum thermal machines. Specifically, we adopt Differentiable Programming (DP) [17,18], to find optimal refrigeration schemes for the quantum Otto cycle under STA conditions (see Fig. 1). We consider a specific class of STA protocols - local counterdiabatic driving (LCD), which are advantageous to the realization of quantum engines since they only require the application of local time-dependent potentials. The STA conditions serve as constraints that are imposed on our optimization problem. From a reinforcement learning (RL) perspective, in this scheme an agent plays a ``game", where the time-dependent frequency $\omega(t)$ of the harmonic oscillator acting as the working medium of the refrigerator can be varied in the interval $t \in [0,\tau)$. For each attempted strategy, $\omega(t)$, the agent receives a reward designed to minimize the energetic cost of the protocol while subjected to the physical constraints imposed by the LCD condition.

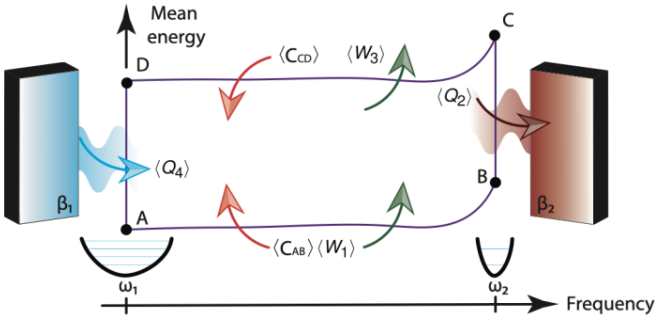


Figure 1 - Scheme of the Otto refrigerator in the energy-frequency domain. A cycle includes a compression stroke (AB) of duration τ, an instant isochoric stroke (BC) with the system coupled to a hot bath, an expansion stroke (CD) of duration τ, and another instant isochoric stroke with a cold bath (DA). Refrigeration corresponds to the withdrawal of heat $\langle Q_4 \rangle$ from the cold bath. The energetic cost of the cycle is the sum of the work contributions $\langle W_1 \rangle$ and $\langle W_3 \rangle$, along with the energetic cost of the STA driving, $\langle C_{AB} \rangle$ and $\langle C_{CD} \rangle$.

The driving profiles that are discovered by the ML scheme are exemplified in Fig. 2. Our method discovers driving protocols of the strokes of an Otto engine that perform twice as better compared to previously conceived solutions [19,20]. Our scheme finds a nontrivial family of functions where the first and second derivatives follow each other; we conclude that this is a crucial property induced by the cost function we have utilized. Furthermore, employing the DP-ML optimization scheme enabled us to uncover a flaw in a previous definition for the energetic cost of STA driving [21]. In this case, the ML optimization identified violations of thermodynamic laws such as the Carnot bound. In contrast, we use a modified definition of the energetic cost of the STA protocol based on the time-averaged Schmidt norm of driving Hamiltonian [22,23] which provided physically transparent results that obeyed the Carnot bound and reached the adiabatic limit.

The advantage of the DP-ML scheme derives from it using the exact gradients of the quantity of interest with respect to variational parameters, hence reducing the number of required iterations to reach an extremum [24,25]. This enabled an effective search in the large multidimensional variational parameter space of functions fulfilling STA conditions.

Since our optimization method is general, it can be easily turned to optimize other cost functions and figures of merits with little effort. Our framework could be directly applied in other control problems, such as entropy reduction in closed systems [26], dynamical decoherence control [27], steering chemical reactions [28], and for the design of quantum electronic and thermal machines [29]. We paved the way for solving hard-constrained problems using state-of-the-art ML tools, by orchestrating an objective as the minimum of a cost function. More generally, our study shows that ML has an advantage over standard theoretical tools in designing quantum devices, thus making them favorable for an experimental realization.

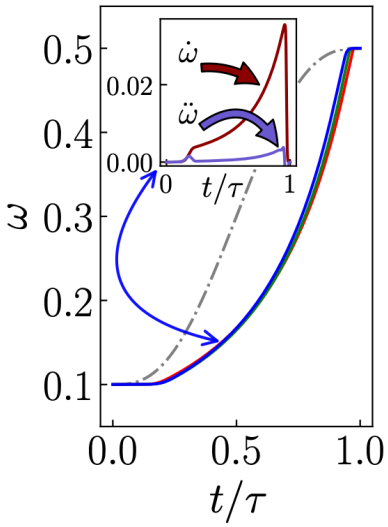


Figure 2 - Examples of frequency profiles $\omega(t)$ discovered by the DP-ML scheme normalized by the compression or expansion stroke duration τ. The gray dashed-dotted line displays the benchmark.

Link to preprint

References

[1] Gemmer J. et al, Quantum thermodynamics: Emergence of thermodynamic behavior within composite quantum systems, Springer (20090
[2] Shubhash K. et al, Finite Time Thermodynamics of Power and Refrigeration Cycles, Springer (2018)
[3] Esposito M. et al, Efficiency at Maximum Power of Low-Dissipation Carnot Engines, Phys. Rev. Lett. 105 15 150603 (2010)
[4] Esposito M. et al, Finite-time thermodynamics for a single-level quantum dot, Europhysics Letters 89 20003 (2010)
[5] Deffner S., Quantum Thermodynamics: An introduction to the thermodynamics of quantum information, Morgan & Claypool Publishers (2019)
[6] Bhattacharjee S. et al, arXiv 2008.07889 (2020)
[7] Kosloff R. et al, Quantum Heat Engines and Refrigerators: Continuous Devices, Annual Review of Physical Chemistry 65 1 365-393 (2014)
[8] Vinjanampathy S. et al, Quantum thermodynamics, Contemporary Physics, 57, 545 (2016)
[9] Das A. et al, Quantum-enhanced finite-time Otto cycle, Phys. Rev. Research 2, 033083 (2020)
[10] Klatzow J. et al, Experimental Demonstration of Quantum Effects in the Operation of Microscopic Heat Engines, Phys. Rev. Lett. 122 110601 (2019)
[11] Revathy S. et al, Universal finite-time thermodynamics of many-body quantum machines from Kibble-Zurek scaling, Phys. Rev. Research 2, 043247 (2020)
[12] Chen X. et al, Fast Optimal Frictionless Atom Cooling in Harmonic Traps: Shortcut to Adiabaticity, Phys. Rev. Lett. 104, 063002 (2010)
[13] Torrontegui E. et al, Shortcuts to Adiabaticity, Advances in Atomic, Molecular, and Optical Physics pages 117–169 Elsevier (2013)
[14] Chen X. et al, Transient energy excitation in shortcuts to adiabaticity for the time-dependent harmonic oscillator, Phys. Rev. A 82 5 053403 (2010)
[15] Muga J. et al, Transitionless quantum drivings for the harmonic oscillator, Journal of Physics B: Atomic, Molecular and Optical Physics 43 8 (2010)
[16] Cui Y. et al, Transient Particle Energies in Shortcuts to Adiabatic Expansions of Harmonic Traps, The Journal of Physical Chemistry A 120 19 (2016)
[17] Baydin A. et al, Automatic Differentiation in Machine Learning: a Survey, Journal of Machine Learning Research 18 153 (2018)
[18] Schäfer F. et al, A differentiable programming method for quantum control, Machine Learning: Science and Technology 1 035009 (2020)
[19] Del Campo A., Shortcuts to adiabaticity by counter-diabatic driving, Phys. Rev. Lett. 111, 100502 (2013)
[20] Beau M. et al, Scaling-Up Quantum Heat Engines Efficiently via Shortcuts to Adiabaticity, Entropy 18, 168 (2016)
[21] Abah O. et al, Shortcut-to-adiabaticity quantum Otto refrigerator, Phys. Rev. Research 2 023120 (2020)
[22] Zheng Y. et al, Cost of counterdiabatic driving and work output, Phys. Rev. A 94, 042132 (2016)
[23] Campbell S. et al, Trade-Off Between Speed and Cost in Shortcuts to Adiabaticity, Phys. Rev. Lett. 118 100601 (2017)
[24] Carrasquilla J., Machine learning for quantum matter, Advances in Physics: X 5 1 1797528 (2020)
[25] Coopmans L., Protocol Discovery for the Quantum Control of Majoranas by Differentiable Programming and Natural Evolution Strategies, PRX Quantum 2 020332 (2021)
[26] Sgroi P. et al, Reinforcement Learning Approach to Nonequilibrium Quantum Thermodynamics, Phys. Rev. Lett. 126 020601 (2021)
[27] Porotti R. et al, Coherent transport of quantum states by deep reinforcement learning, Communications Physics 2 1 (2019)

[28] del Almeida A. et al, Synthetic organic chemistry driven by artificial intelligence, Nature Reviews Chemistry volume 3, pages 589–604 (2019)
[29] Santos A. et al, Superadiabatic controlled evolutions and universal quantum computation, Scientific Reports 5 15775 (2015)

QNLP: Compositional Models of Meaning on a Quantum Computer

Konstantinos Meichanetzidis[†], Robin Lorenz[†], Anna Pearson[†],
Alexis Toumi[†‡], Giovanni de Felice[†‡], Dimitri Kartsaklis[†], Bob Coecke[†]

[†] *Cambridge Quantum Computing Ltd, 17 Beaumont Street, Oxford OX1 2NA, UK*
[‡] *Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3QD, UK*

Quantum Natural Language Processing (QNLP) deals with the design and implementation of NLP models intended to be run on quantum hardware. The categorical compositional distributional model of meaning that combines vector space semantics with compositional syntax and grammar, compels a formal analogy of the tensor structure it features with the mathematical structure of quantum theory. This encourages its use for QNLP since grammatical sentences can then naturally be represented as quantum processes. Here we present experimental results for simple sentence classification tasks with small to medium scale datasets, from implementations on noisy quantum computers provided by IBMQ. (Extended Abstract based on Refs. [1] and [2])

Introduction. DisCoCat (DIStributional COmpositional CATegorical) [3] is a framework for models of natural language meaning. It comes with a rigorous treatment of the interplay between syntax and semantics and with a convenient diagrammatic representation in terms of string diagrams. The conception of this framework was the fruit of recognising the shared formal structure between *pregroup grammar* [4] and compact closed categories like that of finite-dimensional Hilbert spaces (*FHilb*). Sentences are here represented as string diagrams with an open wire carrying the sentence meaning interpreted in the chosen semantics category. The motivation for such a framework stems from the ambition to reconcile vector space semantics with formal approaches to linguistics, and to address the question of how the meaning of a sentence arises from the meanings of its words. In particular, a clear separation is made between syntax and semantics, where compositionality is made explicit in grammar. Over the past 10 years DisCoCat also attracted interest as it was demonstrated to be useful for capturing linguistic phenomena such as ambiguity and entailment [5, 6]. Furthermore, part of the motivation for DisCoCat models is the interpretability of language models, which is a quality that does not trivially characterise modern language models.

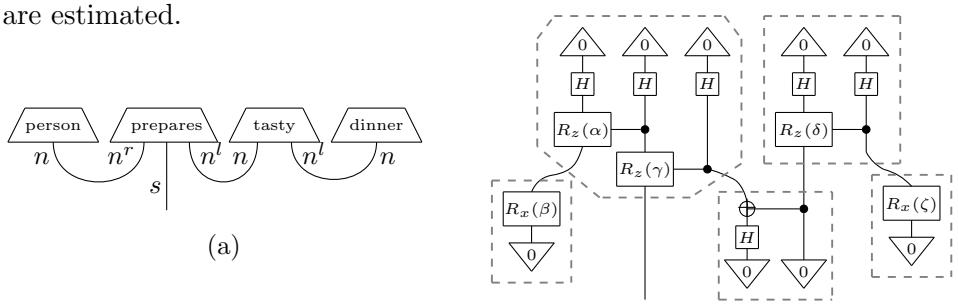
Choosing to interpret the string diagrams in *FHilb* as *quantum circuits* then the 'computation of meaning' – a tensor contraction – would naturally be estimated on a *quantum* computer. In this case there is a correspondence between words and quantum states and between grammatical structure and Bell effects. Notably, the work of Zeng and Coecke [7] built on this idea and presented a quantum algorithm for sentence similarity by reduction to the closest vector problem. Today, NISQ processors are readily available and provide an opportunity for implementing simple NLP tasks using DisCoCat.

Here, we report on a series of experiments, presented in Refs. [1] and [2], which are the first experiments that implement a DisCoCat model – in fact any NLP model – on an actual quantum machine. All three experiments successfully address simple *binary classification* tasks.

The tasks. The first task as a basic proof of concept, addressed in Ref. [1], takes the labels of the sentences in a very small dataset of 16 sentences as 'truth values'. The second task concerns a 'meaning classification' of 130 sentences ('food' vs 'IT'), and the third task concerns the syntactical role of 'relative pronouns' in noun phrases – whether the relative pronoun replaces the subject or the object of the respective relative subclause – with a dataset of 105 sentences [2].

The experiments. Our pipeline begins by parsing each sentence using a pregroup grammar built out of atomic types n for *nouns* and s for *sentence*. The name of a pregroup grammar stems from the fact that every atomic type $t \in \{n, s\}$ has a left- and a right- adjoint type (t^l and t^r), with the property that $t^l \cdot t \to 1$ and $t \cdot t^r \to 1$, where 1 is the trivial type. The existence of two different inverses is motivated by the fact that in language word-order can carry meaning. The pipeline then involves translating each type-tagged sentence into the induced DisCoCat diagram, a string diagram. This diagram represents the grammatical reduction of the sentence and can be seen as a graphical proof that the sentence is grammatical, as witnessed by the reduction of the product of types of all words

to the s-type. Finally the boxes are filled with parametrised quantum circuits, giving them *semantics*. Using a quantum compiler [8], running the circuits on a quantum computer returns outcome statistics from which the labels are estimated.

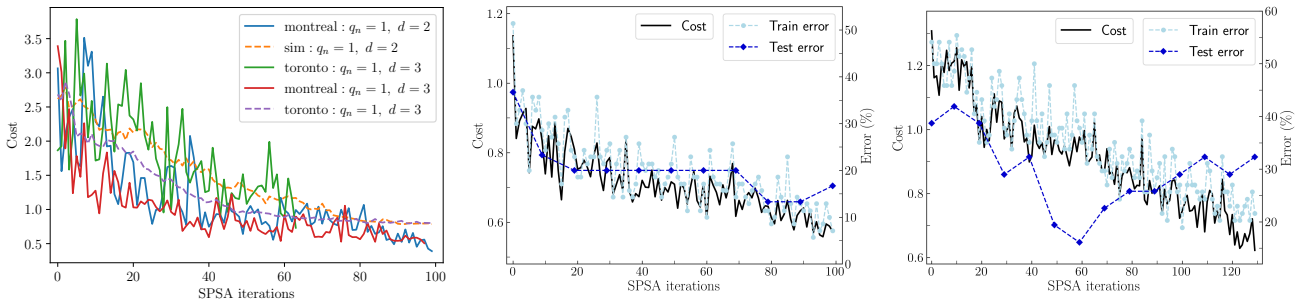


(a) (b)

Figure 1: Example sentence 'person prepares tasty dinner' with in (a) its DisCoCat diagram reflecting the pregroup parsing and in (b) a parametrised quantum circuit it can be mapped to.

Consider the sentence 'person prepares tasty dinner' (from the second task's dataset). Fig. 1a shows its DisCoCat diagram based on the pregroup parsing with the cups corresponding to the pregroup reductions $n \cdot n^r \to 1$ and $n^l \cdot n \to 1$ and the output type indeed being that of a sentence. Fig. 1b shows a corresponding quantum circuit. A hyperparameter regards assigning a number of qubits to each pregroup type. Here each type is assigned one qubit due to the devices' limitations. These qubits define Hilbert spaces in which the word meanings are represented by states prepared from a trivial reference state by parameterised quantum circuits (ansaetze). Thus, a word state is *defined by the parameters* of the quantum circuit that prepares it. Further, states corresponding to *person* and *dinner* were turned into effects, by 'bending them down', in order to reduce circuit width and the amount of required post-selection to implement non-deterministic effects.

Having split the datasets into respective train and test subsets, the model parameters are trained on the former subset via the SPSA optimiser against a Cost function measuring the discrepancy between predicted and actual labels. Despite the typical noise that comes with currently available NISQ machines, in all three experiments the model converges well, i.e. the Cost is minimised successfully [1]. In addition, classical simulations were performed to see the projected behaviour of the model in a noise-free set-up. The typical errors after 100 SPSA iterations are around 8-25% on the training data and 17-37% on the test data, depending on which of the three tasks.



(a) (b) (c)

Figure 2: Results for the first, second and third task in (a), (b) and (c), respectively (q_n and d are hyperparameters determining the ansatz). See Refs. [1],[2] for details.

Conclusions. From a quantum machine learning perspective, this is an instance of a variational quantum circuit approach, where, importantly, the structure of the circuit, that is its connectivity, is not rooted in mere heuristics, *but in fact dictated by the sentence's syntax*. From an NLP perspective, contemplating an obvious question today, namely whether one can do NLP on a quantum computer, the work serves as proof of concept and indeed paves the way to such QNLP. Future work may further scale up the NLP tasks one can consider as the available quantum machines improve, do comparative analyses with approaches that do *not* employ compositionality, and design scalable experiments towards demonstrating quantum advantage.

[1] We used IBM's machines `ibmq_montreal`, `ibmq_toronto` and `ibmq_bogota` with log_2QuantumVolume $= 5$.

References

[1] K. Meichanetzidis, A. Toumi, G. de Felice, and B. Coecke, "Grammar-Aware Question-Answering on Quantum Computers," 2020.

[2] R. Lorenz, A. Pearson, K. Meichanetzidis, D. Kartsaklis, and B. Coecke, "QNLP in Practice: Running Compositional Models of Meaning on a Quantum Computer," `arXiv:2102.12846 [cs.CL]`.

[3] B. Coecke, M. Sadrzadeh, and S. Clark, "Mathematical Foundations for a Compositional Distributional Model of Meaning," *Linguistic Analysis* **36** (2010) 345–384.

[4] J. Lambek, *From Word to Sentence*. Polimetrica, Milan, 2008.

[5] D. Bankova, B. Coecke, M. Lewis, and D. Marsden, "Graded Entailment for Compositional Distributional Semantics," *Journal of Language Modelling* **6** no. 2, (2019) 225–260.

[6] D. Kartsaklis and M. Sadrzadeh, "Prior Disambiguation of Word Tensors for Constructing Sentence Vectors," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNL)*, pp. 1590–1601. 2013.

[7] W. Zeng and B. Coecke, "Quantum Algorithms for Compositional Natural Language Processing," *Electronic Proceedings in Theoretical Computer Science* **221** (Aug, 2016) 67–75. `http://dx.doi.org/10.4204/EPTCS.221.8`.

[8] S. Sivarajah, S. Dilkes, A. Cowtan, W. Simmons, A. Edgington, and R. Duncan, "t|ket>: a Retargetable Compiler for NISQ Devices," *Quantum Science and Technology* **6** no. 1, (2020) 014003.

Regression and Classification with Optimized Random Features: Applications of Exponential Speedup by Quantum Machine Learning without Sparsity and Low-Rankness Assumptions

We develop a quantum algorithm for sampling from an optimized probability distribution of random features, in runtime $O(D)$ that is linear in dimension D of input data, so as to significantly reduce and provably minimize the required number of random features for achieving common learning tasks, regression and classification.

Background.— Kernel methods are a class of the most important methods in ML [1]. However, typical kernel methods, which compute an $N \times N$ Gram matrix, are not scalable as N gets large. To scale up kernel methods to big data, random features [2, 3] are one of the most highly appreciated technique of central use in practice. By random features, we will represent a non-linear function $f : \mathbb{R}^D \to \mathbb{R}$ to be learned as a linear combination $f(x) \approx \hat{f}(x) = \sum_{m=1}^{M} \alpha_m \varphi(v_m, x)$ of non-linear feature maps $\varphi(v, x) = \mathrm{e}^{-2\pi \mathrm{i} v \cdot x}$, that is, sin and cos. Note that D-dimensional data are considered. Learning with random features is achieved by sampling many feature maps $\varphi(v_m, \cdot)$ at random parameters $v_1, \ldots, v_M \in \mathbb{R}^D$, followed by finding appropriate coefficients $\alpha_1, \ldots, \alpha_M$ by convex optimization using the N examples. Conventionally, $v_1, \ldots, v_M$ are sampled from a *data-independent* probability distribution $d\tau(v)$ depending only on a kernel function k in kernel methods, but this may *require a large number of features* $M = \widetilde{O}(1/\epsilon^2)$ for achieving learning to accuracy ϵ [2, 3]. The requirement of large M slows down the decision of all M features, the regression over M coefficients, and the evaluation of the learned function $\hat{f}(x)$ in using it after the learning. Acceleration of kernel methods with random features is not a specific problem but of central importance for their various applications, e.g., to computer vision, natural language processing, marketing, robotics, and ML-assisted investigation of physics.

Problem and Result.— To achieve the acceleration, we will aim at minimizing M required for the learning. Recent work [4] has proposed to sample features from a *data-optimized* probability distribution, that is, a distribution that puts greater weight on important features optimized for the data. The optimized distribution is given by a weighted distribution $q_\epsilon^*(v)d\tau(v) \propto \langle \varphi(v, \cdot) | (\Sigma + \epsilon \mathbb{1})^{-1} \varphi(v, \cdot) \rangle_{L_2(d\rho)} d\tau(v)$, where Σ is a linear operator on a space $L_2(d\rho)$ of functions, depending both on the kernel k and the data distribution $d\rho(x)$ [4]. The use of $q_\epsilon^*(v)d\tau(v)$ can significantly reduce the required number M of features for the learning, which is indeed provably optimal up to a logarithmic gap [4]. We call features sampled from this data-optimized distribution *optimized random features*. However, a problematic computational bottleneck arises from each sampling step from $q_\epsilon^*(v)d\tau(v)$ due to the inversion of $\Sigma + \epsilon \mathbb{1}$, which is infinite-dimensional. A discretized approximation of Σ using bits yields an $O(\exp(D))$-dimensional operator for data dimension D [5]. As a result, sampling from $q_\epsilon^*(v)d\tau(v)$ requires an *exponential runtime* $O(\exp(D))$ for inverting this operator as long as we use the existing classical algorithms [4, 6, 7]. To achieve our aim, the problem is how to perform sampling from $q_\epsilon^*(v)d\tau(v)$ efficiently. One may wonder why Ref. [4] is appreciated despite the exponential runtime, but Ref. [4] is significant in the context of statistical learning theory. In contrast, our results indeed make this sampling from $q_\epsilon^*(v)d\tau(v)$ possible with quantum computation in as fast as linear runtime $O(D)$, and also demonstrate how to use this quantum algorithm to speed up central ML tasks, regression and classification, as summarized in the following (see also the technical version [5, 8]).

1. ([5]) We develop a quantum algorithm for sampling from $q_\epsilon^*(v)d\tau(v)$ with exponential speedup in D:

   Runtime of our quantum algorithm: $O(D)$ $\leftrightarrow$ Runtime of classical algorithms [4, 6, 7]: $O(\exp(D))$.

2. ([5]) We show that we can combine M features sampled by our quantum algorithm with optimization of coefficients by a well-established classical algorithm, i.e., stochastic gradient descent (SGD), to achieve the regression task as a whole in time $O(MD/\epsilon^2)$, without canceling out our exponential speedup in D. Importantly, M can be much smaller than conventional random features [2, 3], achieving a significant acceleration as discussed below.

3. ([8]) We also show that the combination of our quantum algorithm with the SGD can achieve a classification task (under a well-studied low-noise condition [9–11]) in time $O(MD\,\mathrm{polylog}(1/\epsilon'))$, where ϵ' is an excess classification error [8]. This M can also be much smaller than conventional random features [2, 3]. This significantly accelerates leading classification algorithms [9–11] based on kernel methods in terms of M, without ruining their exponential error-convergence speed in ϵ'.

Advantage.— Our quantum algorithm makes it possible to sample optimized random features minimizing M, by improving the bottleneck faced by classical sampling algorithms [4, 6, 7]. In the following, we clarify this advantage for the regression task [5], while Sec. 3.2 of our technical version [8] also shows a similar advantage for the classification task. The minimal M for achieving the regression task is given by the degree of freedom $d(\epsilon) = \mathrm{Tr}\,\Sigma(\Sigma+\epsilon\mathbb{1})^{-1}$, which is determined depending both on kernel k and data distribution $d\rho$ [4]. In the worst case, we may only have $M = \widetilde{O}(d(\epsilon)) = \widetilde{O}(1/\epsilon^2)$ [4], the same scaling as conventional random features [2, 3]. But for a Gaussian kernel and a sub-Gaussian data distribution, we have an exponential advantage in ϵ [4]

$$\text{Our quantum algorithm: } M = \widetilde{O}(d(\epsilon)) = O(\mathrm{polylog}(1/\epsilon)) \leftrightarrow \text{Conventional random features [2, 3]: } M = \widetilde{O}(1/\epsilon^2).$$

The advantage becomes significant especially *when we use* the function learned with optimized random features. Once the learning has finished, the learned function $\hat{f}(x) = \sum_{m=1}^{M} \alpha_m \varphi(v_m, x)$ can be evaluated within runtime $O(MD) = O\left(D\,\mathrm{polylog}\,(1/\epsilon)\right)$ in the above case. For comparison, conventional random features require $O(MD) = \widetilde{O}\left(D/\epsilon^2\right)$ [2, 3]. Our quantum algorithm makes it possible to achieve this exponential advantage in ϵ in using the function learned with optimized random features, feasibly in linear runtime $O(D)$. This advantage is crucial especially for applications of ML that requires real-time computing, e.g., an embedded system, robotics, feedback control in physical experiments, and machine-learning-based decoders for quantum error correction.

As a result, we provide a promising framework of quantum machine learning (QML) that leverages our quantum algorithm for sampling optimized random features, to achieve the optimal M among all algorithms using random features. The optimized random features can be used for a general class of kernel-based regression tasks [5] as shown above, and also kernel-based classification tasks as we show in Ref. [8], which are of central importance in ML. The significant advantage indeed appears for the Gaussian kernel and the sub-Gaussian data distribution, in both the regression and classification tasks [5, 8]. These results show end-to-end applications of quantum computation in the central problems of ML.

Impact on QML.— The novelty of our results is that our QML algorithm is *exponentially faster* than the existing classical sampling algorithms [4, 6, 7], yet is still *free from sparsity and low-rankness assumptions.* QML algorithms such as Refs. [12–15] may achieve exponential speedups over classical algorithms only if matrices involved in the algorithms are sparse. Another class of QML algorithms such as Refs. [16–19] do not require sparsity but may attain large speedups only if the matrices have low rank. Note that recent "quantum-inspired" classical algorithms such as Refs. [20–22] also require low rank. But power and applicability of these QML algorithms are restricted by the assumptions on sparsity and low-rankness; that is, to attain large speedups, careful justifications of these assumptions have been needed [23].

Our key technical contribution is to develop an approach for *circumventing the sparsity and low-rankness assumptions in our QML algorithm, broadening applicability of QML.* A difficulty of our sampling task arises from the fact that the discretized description of $q_\epsilon^*(v)d\tau(v)$ includes an inverse of $O(\exp(D))$-dimensional linear operator $\hat{\mathbf{\Sigma}}_\epsilon$ to approximate $(\Sigma+\epsilon\mathbb{1})^{-1}$, and $\hat{\mathbf{\Sigma}}_\epsilon$ may *not be sparse or of low rank* (see Secs. 3.1 and E of Ref. [5] for precise definition). Remarkably, our technique does not directly use the conventional ways of implementing a linear operator that requires sparsity or low-rankness, yet implements $\hat{\mathbf{\Sigma}}_\epsilon$ efficiently in time $O(D)$. The sparse and low-rankness assumptions can be avoided because we prove that in our sampling task, we can explicitly decompose $O(\exp(D))$-*dimensional non-sparse full-rank* $\hat{\mathbf{\Sigma}}_\epsilon$ (to be inverted by quantum singular-value transformation (QSVT) [24]) into addition and multiplication of *efficiently implementable* building blocks, i.e., block encodings of diagonal (i.e., sparse) operators and quantum Fourier transform (QFT).

Wide applicability.— Owing to our circumventing the sparsity and low-rankness assumptions, our QML algorithm is *widely applicable to learning with representative choices of kernels*, e.g., the Gaussian kernel and the Laplacian kernel [5]. Remarkably, the advantage of our QML algorithm can be obtained in the same model of functions $f(x)$ as conventional random features [5, 8]. As a result, the advantage is expected to appear for practical classes of data sets, e.g., those given by a sub-Gaussian data distribution to be learned with a Gaussian kernel.

Consequently, our results open a route to a *widely applicable* framework of QML without sparsity and low-rankness assumptions yet taking advantage of large quantum speedups in the central ML tasks, regression and classification. QSVT and QFT may make our algorithm hard to simulate by classical computation, and hard to perform even on near-term noisy quantum devices. However, in contrast to heuristic QML algorithms for noisy quantum devices such as Ref. [25] where no proof bounds its runtime, our QML algorithm aims at applications on large scales; to achieve this aim, we analytically prove the exponential quantum speedup over the existing classical algorithms. The wide applicability of our QML algorithm makes it a promising candidate for "killer applications" of universal quantum computers in the long run; after all, large-scale ML will be eventually needed in practice. Therefore, our results establish a more solid theoretical motivation based on QML for further technological development toward realizing the quantum computer.

[1] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, Cambridge, MA, USA, 2001).

[2] A. Rahimi and B. Recht, Random features for large-scale kernel machines, in *Advances in Neural Information Processing Systems 20*, edited by J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis (Curran Associates, Inc., 2008) pp. 1177–1184.

[3] A. Rahimi and B. Recht, Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning, in *Advances in Neural Information Processing Systems 21*, edited by D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou (Curran Associates, Inc., 2009) pp. 1313–1320.

[4] F. Bach, On the equivalence between kernel quadrature rules and random feature expansions, Journal of Machine Learning Research **18**, 1 (2017).

[5] H. Yamasaki, S. Subramanian, S. Sonoda, and M. Koashi, Learning with optimized random features: Exponential speedup by quantum machine learning without sparsity and low-rank assumptions, in *Advances in Neural Information Processing Systems*, Vol. 33, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Curran Associates, Inc., 2020) pp. 13674–13687, arXiv:2004.10756.

[6] Y. Sun, A. Gilbert, and A. Tewari, But how does it work in theory? linear svm with random features, in *Advances in Neural Information Processing Systems 31*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., 2018) pp. 3379–3388.

[7] S. Shahrampour and S. Kolouri, On sampling random features from empirical leverage scores: Implementation and theoretical guarantees (2019), arXiv:1903.08329.

[8] H. Yamasaki and S. Sonoda, Exponential error convergence in data classification with optimized random features: Acceleration by quantum machine learning (2021), arXiv:2106.09028.

[9] L. Pillaud-Vivien, A. Rudi, and F. Bach, Exponential convergence of testing error for stochastic gradient methods, in *Proceedings of the 31st Conference On Learning Theory*, Proceedings of Machine Learning Research, Vol. 75, edited by S. Bubeck, V. Perchet, and P. Rigollet (PMLR, 2018) pp. 250–296.

[10] A. Nitanda and T. Suzuki, Stochastic gradient descent with exponential convergence rates of expected classification errors, in *Proceedings of Machine Learning Research*, Proceedings of Machine Learning Research, Vol. 89, edited by K. Chaudhuri and M. Sugiyama (PMLR, 2019) pp. 1417–1426.

[11] S. Yashima, A. Nitanda, and T. Suzuki, Exponential convergence rates of classification errors on learning with sgd and random features (2019), arXiv:1911.05350.

[12] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum algorithm for linear systems of equations, Phys. Rev. Lett. **103**, 150502 (2009).

[13] N. Wiebe, D. Braun, and S. Lloyd, Quantum algorithm for data fitting, Phys. Rev. Lett. **109**, 050505 (2012).

[14] S. Lloyd, S. Garnerone, and P. Zanardi, Quantum algorithms for topological and geometric analysis of data, Nature communications **7**, 10138 (2016).

[15] Z. Zhao, J. K. Fitzsimons, and J. F. Fitzsimons, Quantum-assisted gaussian process regression, Phys. Rev. A **99**, 052331 (2019).

[16] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum principal component analysis, Nature Physics **10**, 631 (2014).

[17] P. Rebentrost, M. Mohseni, and S. Lloyd, Quantum support vector machine for big data classification, Phys. Rev. Lett. **113**, 130503 (2014).

[18] I. Kerenidis and A. Prakash, Quantum Recommendation Systems, in *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 67, edited by C. H. Papadimitriou (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2017) pp. 49:1–49:21.

[19] L. Wossnig, Z. Zhao, and A. Prakash, Quantum linear system algorithm for dense matrices, Phys. Rev. Lett. **120**, 050502 (2018).

[20] E. Tang, A quantum-inspired classical algorithm for recommendation systems, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019 (Association for Computing Machinery, New York, NY, USA, 2019) p. 217–228.

[21] D. Jethwani, F. L. Gall, and S. K. Singh, Quantum-Inspired Classical Algorithms for Singular Value Transformation, in *45th International Symposium on Mathematical Foundations of Computer Science (MFCS 2020)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 170, edited by J. Esparza and D. Kráľ (Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2020) pp. 53:1–53:14.

[22] N.-H. Chia, A. Gilyén, T. Li, H.-H. Lin, E. Tang, and C. Wang, Sampling-based sublinear low-rank matrix arithmetic framework for dequantizing quantum machine learning, in *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020 (Association for Computing Machinery, New York, NY, USA, 2020) p. 387–400.

[23] S. Aaronson, Read the fine print, Nature Physics **11**, 291 (2015).

[24] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019 (Association for Computing Machinery, New York, NY, USA, 2019) p. 193–204.

[25] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, Nature **567**, 209 (2019).

Automated NV-centre calibration for quantum internet nodes

Eliska Greplova, Achmed Marif, Matteo Pompili, Ronald Hanson

Abstract: Quantum internet is one of the exciting uses of quantum technologies. NV centres have emerged as an excellent platform for distributed entanglement links key for its realisation. In this work we introduce a physics guided reinforcement learning algorithm for their automated tuning and operation completely free of human operator surveillance.

Recent explosive growth of quantum technologies promises a range of exciting technological applications, from digital quantum computing to quantum networks for quantum communication [1]. Contemporary quantum devices are reaching the boundaries of both classical simulatability and scalable classical control. As a possible resolution of these challenges, the field of artificial intelligence (AI)-driven quantum control has emerged.

Generally the interpretation of the data acquired by measuring quantum devices is a complex task that requires the analysis by a skilled human operator. This fact is also a central reason why the automation of large scale quantum experiments has been evading effective solutions via standard algorithmic techniques in favour of direct control by human experts. While such approaches function well for a small scale devices in university and industry labs, they are not scalable towards large-scale devices and industry production.

The adoption of AI and machine learning techniques across many science and technology fields fuelled the creation of a new field within the quantum devices community: AI-driven tuning of quantum devices. The goal of these efforts is to employ advanced feature recognition and decision making techniques of artificial intelligence to analyse the features in the incoming quantum devices data and realise the tuning of the chip in the real time in a completely autonomous way.

The first efforts towards useful practical utilisation of these methods have been established within the field of semiconductor quantum dots, where all stages of operator-free tuning have been experimentally established [2-6]. In addition to that, a tuning speed outperforming human operators has been achieved [7].

In the context of automated tuning, NV centres [8], a leading platform for quantum communication, remained unexplored until now because of their intricate electronic structure that is yet not fully theoretically understood. This theoretical complexity presents an obstacle to formulating good policy for training of AI algorithms.

In this work, we take an approach of radically simplifying the NV-centre model to only consider a handful of energy levels that are most relevant for distributed entanglement experiments. This approach allows us to formulate 6-level Born-Markov master equation description of the system. At the same time, this high-level theoretical formalism appears to be sufficient for the generation of experiment-like measurement data traces, which, in turn, provide us with the wealth of data for theoretical analysis as well as for training of the machine learning models.

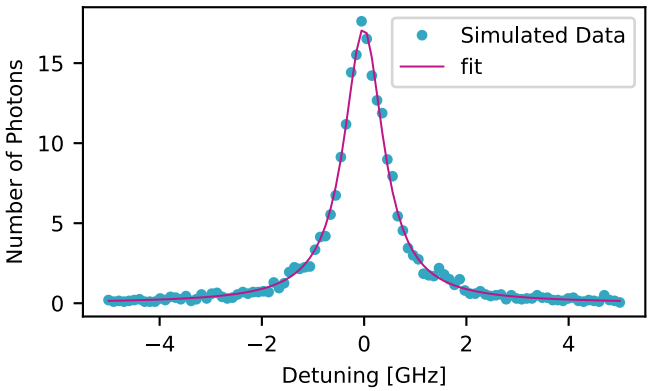


Fig. 1: Theoretical analysis of the number of measured photons as a function of the measurement laser detuning from the resonant frequency on the NV centre qubit.

Our master equation model allows us to understand the relation between the measured number of photons and detuning of the measured

frequency from the resonance (example for one of the controlling lasers in Fig. 1) and therefore formulate a rigorous bound on the distance towards resonance frequency. The fit shown in Fig. 1 alone could, of course, provide us with sufficient information to tune a single laser frequency. However, even a single NV centre is controlled by multiple mutually dependent parameters, which makes the deterministic exploration of the parameter space exceedingly costly in terms of measurement time.

In this work we approach the unfavourable scaling of NV centre frequency tuning via a combination of reinforcement learning (RL) and the physical modelling described above. Specifically, the fits like the one shown in Fig. 1 allow us to define a laser frequency search domain as well as a suitable way to discretize it. Then we employ Deep-Q Networks, a type of reinforcement learning, to adjust the control laser frequencies based on the input of the measured photon counts.

The results of our algorithm for one of the NV-centre control lasers is illustrated in Fig. 2. The upper panel shows the number of measured photon counts as a function of the measurement number. The desirable result is to keep the photon count above the threshold of 10 photons, which corresponds to the laser frequency being on resonance with the NV centre qubit. One can observe that our RL agent recovers the photon count in just a few iterations and for the most part of the measurement times retains the measured photon number consistently up. The bottom panel of the Fig. 2 shows a benchmark of our method against standard error correction technique based on the sampling from the photon count distribution by visualising detuning for both methods as function of the measurement iteration.

In conclusion, we present physics modelling guided reinforcement learning technique for automated on-device real time NV-centre tuning. Our method presents the first step towards fully autonomous operation of distributed entanglement links for quantum internet.

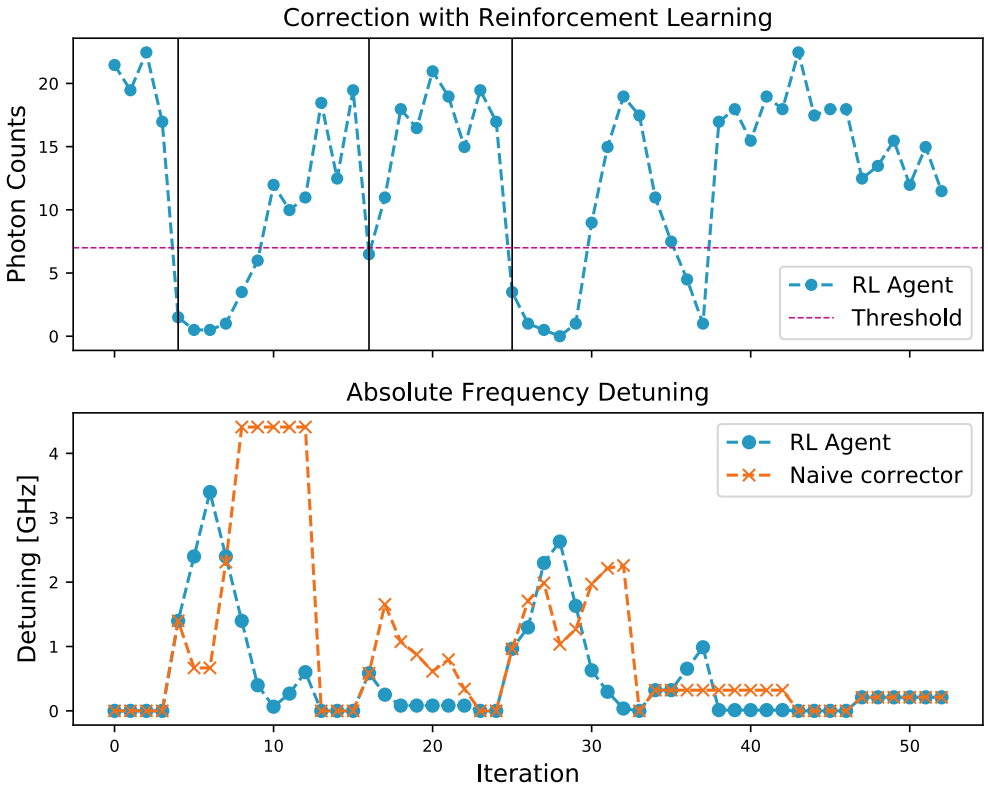


Fig. 2: The upper panel illustrates the actions of the RL agent based on the incoming photon count. The lower panel shows the comparison of the detuning from the resonance frequency of the RL agent and a benchmark naive corrector.

1. Wehner, Stephanie, David Elkouss, and Ronald Hanson. "Quantum internet: A vision for the road ahead." *Science* 362.6412 (2018).

2. Zwolak, Justyna P., et al. "Autotuning of double-dot devices in situ with machine learning." *Physical review applied* 13.3 (2020): 034075.

3. Kalantre, Sandesh S., et al. "Machine learning techniques for state recognition and auto-tuning in quantum dots." *npj Quantum Information* 5.1 (2019): 1-10.

4. Durrer, Renato, et al. "Automated tuning of double quantum dots into specific charge states using neural networks." *Physical Review Applied* 13.5 (2020): 054019.

5. Lennon, Dominic T., et al. "Efficiently measuring a quantum device using machine learning." *npj Quantum Information* 5.1 (2019): 1-8.

6. Nguyen, V., et al. "Deep reinforcement learning for efficient measurement of quantum devices." *npj Quantum Information* 7.1 (2021): 1-9.

7. Moon, Hyungil, et al. "Machine learning enables completely automatic tuning of a quantum device faster than human experts." *Nature communications* 11.1 (2020): 1-10.

8. Pompili, Matteo, et al. "Realization of a multinode quantum network of remote solid-state qubits." *Science* 372.6539 (2021): 259-264.