QTML 2021

Quantum Techniques in Machine Learning

Booklet of Extended Abstracts

Tuesday November 9 2021

FLIP: A flexible initializer for arbitrarily-sized parametrized quantum circuits

Frederic Sauvage,[1,2] Sukin Sim,[3,4] Alexander A. Kunitsa,[3] William A. Simon,[3] Marta Mauri,[1] and Alejandro Perdomo-Ortiz[1,*]

[1]*Zapata Computing Canada Inc., 325 Front St W, Toronto, ON, M5V 2Y1*
[2]*Imperial College London, Prince Consort Road, SW7 2BW, United Kingdom*
[3]*Zapata Computing, Inc., 100 Federal Street, Boston, MA 02110, USA*
[4]*Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA*

Developing efficient methods for the training of quantum circuits is critical to the success of any variational quantum algorithms. We address this task from an initialization perspective and propose a novel meta-learning scheme which not only surpasses state-of-the-art strategies but also allows to initialize larger circuits than used during training. **Pre-print:** arXiv:2103.08572

Developing more efficient methods for the training of parametrized quantum circuits (PQCs) is urgently needed to explore further the potential offered by variational quantum algorithms (VQAs) [1, 2]. Drawing and extending ideas from the field of meta-learning [3, 4], we propose to accelerate the training of PQCs by careful initialization of their circuit parameters, and develop a FLexible Initializer for Parametrized quantum circuits (FLIP). Rather than relying on a predefined or a random set of initial parameters, FLIP is trained to discover the structure of successful initial parameters from a small number of related PQCs problems (see Figure 1); after training, it can be employed to initialize the circuit parameters of similar but new problems. Of particular appeal, FLIP is designed to accommodate, and thus to learn from, circuits of varied sizes (both in depth and width): a critical feature lacking in other meta-learning parameter initializing strategies [5, 6] and state-of-the-art techniques proposed to date.

We demonstrate the performance of FLIP in three distinct scenarios: a family of problems with proven barren plateaus, the training of quantum approximate optimization algorithm (QAOA) circuits [7], and the training of hardware-efficient quantum circuits [8] for finding the ground state energies of 1D Fermi-Hubbard models. A summary of the results is reported in Figure 2, showing that circuits initialized with FLIP are significantly easier to train – both in terms of the quality of the solutions found and the number of iterations required – compared to random initialization and even more involved initialization strategies. Additionally, in each of these examples we verify that FLIP can successfully initialize larger circuits than used during its training.

Overall, we find that informed initialization of quantum circuit parameters can substantially improve the training of PQCs and thus reduce the number of circuits to be executed. As such, we expect these techniques to be key to the current effort of extending the application of VQAs to larger problem sizes. As gate-based quantum computing technologies mature, initialization techniques which embrace this unique flexibility will be essential to mitigate the challenges in trainability posed for PQC-based models and eventually scale to their application in real-world applications settings.

* alejandro@zapatacomputing.com

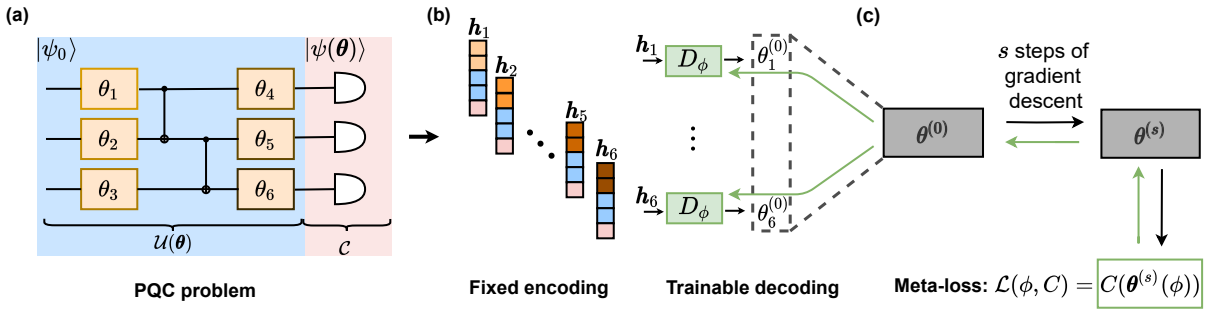


FIG. 1. **Overview of FLIP.** FLIP aims at learning the structure of good initial parameters over families of PQCs problems. **(a)** A generic PQC problem is composed of a parametrized circuit $\mathcal{U}(\boldsymbol{\theta})$ and an objective $\mathcal{C}$ which can be estimated through repeated measurements of the output state $|\psi(\boldsymbol{\theta})\rangle = \mathcal{U}(\boldsymbol{\theta})|\psi_0\rangle$. Solving a PQC problem corresponds to the minimization of the cost function $C(\boldsymbol{\theta}) = \mathcal{C}(|\psi(\boldsymbol{\theta})\rangle)$. Example for a system size of $n = 3$ qubits, and a quantum circuit with $K = 6$ parameters. **(b)** At the core of FLIP lies an encoding–decoding scheme which maps a PQC problem to a set of initial parameters $\boldsymbol{\theta}^{(0)}$ with appropriate dimension. (Encoding) Each of the K parameters of the circuit $\mathcal{U}$ is represented as an encoding vector $\boldsymbol{h}_k$ containing information regarding the parameter itself (orange squares), the overall circuit (blue) and optionally the objective (red). Importantly, these vectors are of fixed size (here $S = 5$) and uniquely represents each parameter. (Decoding) These K encodings are then decoded by a neural network, D_ϕ with weights ϕ, outputting a single value $\boldsymbol{\theta}_k^{(0)}$ per encoding $\boldsymbol{h}_k$. Taken together, the encoding and decoding produce a vector of initial parameters $\boldsymbol{\theta}^{(0)}$ which is function of the details of the circuits and objectives, and with dimension always matching the number of circuit parameters. **(c)** Training FLIP consists in adapting the weights ϕ of the decoder to produce *good initial parameters*, that is, parameters which can be quickly refined by taking a small number s of steps of gradient descent [3]. While illustrated here for a single PQC problem, in practice FLIP is trained over several PQCs sampled from a distribution of problems. After training it is used to initialize the circuits corresponding to new problems drawn from the same or a similar distribution with, for instance, problems involving larger system sizes and deeper circuits.

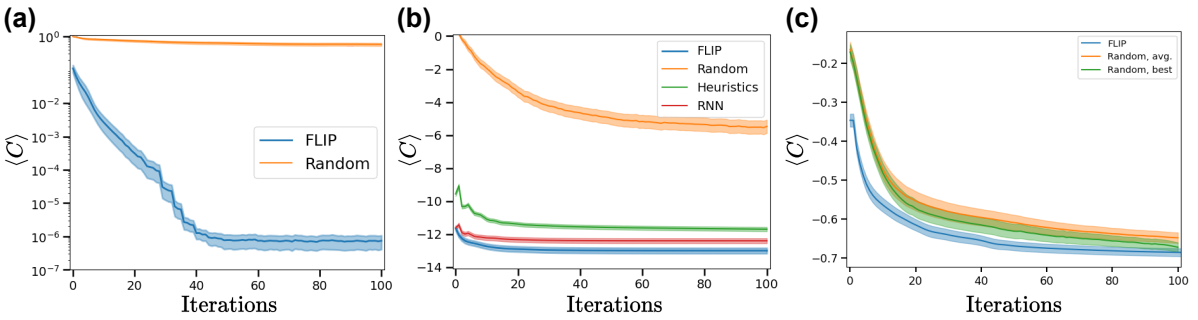


FIG. 2. **Overview of the results.** The training performances of circuits initialized with FLIP (blue) are assessed based on three distinct families of PQCs problems. In each case, converge of the corresponding costs $\langle C \rangle$ (averaged over many problems) is reported as a function of the number of optimization iterations that are needed after initialization of the circuits. **(a)** Synthetic problems of quantum state preparations which exhibit Barren plateaus. Initialization with FLIP is compared to random initialization of the circuits (orange). **(b)** Training of QAOA circuits for solving max-cut graph problems. FLIP is compared to random initialization (orange), initialization by means of heuristics inspired from [9] (green) and also more involved initialization strategies [5] (red). **(c)** Training of hardware-efficient circuits for the ground state preparation of one–dimensional Fermi-Hubbard models. In any of the problems investigated, the circuits initialized with FLIP are found to converge to better solutions and in fewer iterations than any other initialization strategy investigated.

[1] 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).

[2] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S Kottmann, Tim Menke, *et al.*, "Noisy intermediate-scale quantum (nisq) algorithms," arXiv preprint (2021).

[3] Chelsea Finn, Pieter Abbeel, and Sergey Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International Conference on Machine Learning* (PMLR, 2017) pp. 1126–1135.

[4] Alex Nichol, Joshua Achiam, and John Schulman, "On first-order meta-learning algorithms," arXiv preprint (2018).

[5] Guillaume Verdon, Michael Broughton, Jarrod R McClean, Kevin J Sung, Ryan Babbush, Zhang Jiang, Hartmut Neven, and Masoud Mohseni, "Learning to learn with quantum neural networks via classical neural networks," arXiv preprint (2019).

[6] Alba Cervera-Lierta, Jakob S. Kottmann, and Alán Aspuru-Guzik, "Meta-variational quantum eigensolver: Learning energy profiles of parameterized hamiltonians for quantum simulation," PRX Quantum **2**, 020329 (2021).

[7] Sam Gutmann Edward Farhi, Jeffrey Goldstone, "A quantum approximate optimization algorithm," arXiv preprint (2014).

[8] Pierre-Luc Dallaire-Demers, Jonathan Romero, Libor Veis, Sukin Sim, and Alán Aspuru-Guzik, "Low-depth circuit ansatz for preparing correlated fermionic states on a quantum computer," Quantum Science and Technology **4**, 045005 (2019).

[9] Fernando GSL Brandao, Michael Broughton, Edward Farhi, Sam Gutmann, and Hartmut Neven, "For fixed control parameters the quantum approximate optimization algorithm's objective function value concentrates for typical instances," arXiv preprint (2018).

Impact of Noise and Error Mitigation on Trainability of Variational Quantum Algorithms

(arXiv:2007.14384 [1] and arXiv:2109.01051 [2])

Samson Wang, Piotr Czarnik, Andrew Arrasmith, Marco Cerezo, Enrico Fontana, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J. Coles

Variational Quantum Algorithms (VQAs) are a leading algorithmic approach in the Noisy Intermediate-Scale (NISQ) era because they adapt to the constraints of NISQ devices. Specifically, VQAs prepare a cost which takes the form of an expectation value of some measurement operator over a parameterized quantum circuit, and minimize this cost by training the parameters in the circuit via a classical-quantum feedback loop [3, 4]. The idea is to partition quantum and classical resources effectively by efficiently computing the cost on a quantum computer whilst carrying out the parameter optimization classically. Different implementations of this versatile framework have been proposed for a broad spectrum of problems ranging from dynamical quantum simulation [5–15] to quantum machine learning [16–22] and beyond.

A. Noise-Induced Barren Plateaus

New challenges have recently been discovered for VQAs due to the effects of hardware noise [1, 23]. In Ref. [1] we investigate the impact of noise on the trainability of VQAs. We note that VQAs whose expectation values are strongly affected by noise but which are still trainable (i.e. one can find the optimal parameter set) can still be useful. This is due the fact that the goal of some VQAs is to extract the optimal parameter set, and for algorithms where the goal is to obtain the optimal cost value one may use the optimal parameters and approximate the optimal (noise-free) cost value by using error mitigation.

We consider ansatzes that consist of layers of general unitaries $\{U_l(\boldsymbol{\theta}_l)\}_{l=1}^{L}$ where each $\boldsymbol{\theta}_l = (\theta_{lm}\}_m$ is a set of continuous parameters that are together optimized to minimize a noisy cost function $\widetilde{C}$. We suppose each unitary $U_l(\boldsymbol{\theta}_l)$ is parameterized as $U_l(\boldsymbol{\theta}_l) = \prod_m e^{-i\theta_{lm}H_{lm}}W_{lm}$, where H_{lm} are Hermitian operators and W_{lm} denote unparametrized gates. We consider a noise model where local Pauli noise channels $\mathcal{N}_j$ act on each qubit j before and after each unitary $U_l(\boldsymbol{\theta}_l)$. Finally, in this section we consider a class of local noise models where the action of each $\mathcal{N}_j$ on a local Pauli operator $\sigma \in \{X, Y, Z\}$ can be expressed as

$$\mathcal{N}_j(\sigma) = q_\sigma \sigma \,, \tag{1}$$

where $-1 < q_X, q_Y, q_Z < 1$. We characterize the noise with a single parameter $q = \sqrt{\max\{|q_X|, |q_Y|, |q_Z|\}} < 1$. Further, we denote the noisy cost function as $\widetilde{C} = \mathrm{Tr}\left[O\widetilde{\rho}\right]$ where $\widetilde{\rho}$ is the noisy state obtained after application of L layers parameterized unitaries interleaved with local noise (1), and O is some measurement operator.

Result 1: We find upper bounds that show the concurrent exponential concentration of the cost function around the corresponding value for the maximally mixed state, and exponential suppression of partial derivatives:

$$\left|\widetilde{C} - \mathrm{Tr}[O]/2^n\right| \leqslant D(q,n)\,, \tag{2}$$

$$|\partial_{lm}\widetilde{C}| \leqslant G(q,n) \;\; \forall\, l,m\,, \tag{3}$$

where if $L \in \Omega(n)$ then $D(q,n), G(q,n) \in \mathcal{O}(q^{\alpha n})$ for some positive constant α. These results hold under the assumptions that that O can be decomposed into a polynomial number of Pauli strings, and each H_{lm} that generates the unitaries in the ansatz has bounded eigenvalues, which we remark are satisfied in most studied settings. We refer to the effects of (2) and (3) as exponential cost concentration and NIBPs respectively.

NIBPs and cost concentration imply that noise severely impedes the training process of VQAs with linear or superlinear depth circuits, as in such a setting one requires an exponential number of shots per optimization step to resolve the cost landscape against finite sampling noise. Thus, we emphasize that at its core, these effects present an exponential *resolvability* issue for the VQA cost landscape. In numerical implementations we find that the training process is impeded with growing system size [1], and we corroborate our findings with an implementation of the Hamiltonian Variational Ansatz on superconducting hardware. As with other barren plateau effects [24, 25], this exponential scaling does not only arise for gradient-based optimizers but also impacts gradient-free methods [26] and optimizers that use higher-order derivatives [27]. In addition, NIBPs and cost concentration cannot be addressed by layer-wise training, correlating parameters and other strategies [28–33], all of which can help avoid noise-free barren plateaus. Thus, these effects represent a serious issue for VQA scalability, and could ultimately be a roadblock for near-term quantum advantage. It is therefore crucial to investigate potential methods to mitigate them.

B. Unification of Error Mitigation Techniques and Impact on Trainability

Given the great success of error mitigation (EM) methods in suppressing the effects of hardware noise on observable expectation values, it is natural to ask whether EM methods could address NIBPs. More generally, one could simply ask: does it help to use error mitigation during the training process for VQAs? We investigate this question the second part of our work, of which a detailed version can be found in Ref. [2].

We make two main contributions in this second part of our work. First, we consider a broad class of EM strategies that includes as special cases Zero-Noise Extrapolation [12, 34–36], Virtual Distillation [37, 38], Probabilistic Error Cancellation [34, 35] and methods that implement a linear ansatz such as Clifford Data Regression [39]. We unify and generalize such strategies by viewing them as the preparation of error-mitigated cost values of the form

$$C_m(\boldsymbol{\theta}) = \sum_{(\sigma(\boldsymbol{\theta}),X,M,k)\in T} a_{X,M,k}\, \mathrm{Tr}\left[X\left(\sigma(\boldsymbol{\theta})^{\otimes M}\otimes|0\rangle\langle 0|^{\otimes k}\right)\right]. \tag{4}$$

That is, we consider linear combinations of expectation values of the form $\mathrm{Tr}[X(\sigma^{\otimes M}\otimes|0\rangle\langle 0|^{\otimes k})]$ where σ is a n-qubit quantum state that in general can be prepared by a different circuit to that of the state of interest, M is the number of copies of that state, $|0\rangle\langle 0|^{\otimes k}$ are k clean ancillary qubits and X is a measurement operator with bounded eigenvalues acting on the full Hilbert space. By considering linear combinations of such expectation values, we account for the ability to post-process measurement results classically with a linear map, for instance, as is the case with Probabilistic Error Cancellation.

Result 2: We find that under the noise model (1), there exists $\beta > 0$ and F such that

$$\left|C_m(\boldsymbol{\theta}) - F\right| \in \mathcal{O}\left(2^{-\beta n}\right), \tag{5}$$

if the circuit depths that prepare $\sigma(\boldsymbol{\theta})$ satisfy $L_{\sigma(\boldsymbol{\theta})} \in \Omega(n)$ for all $\sigma(\boldsymbol{\theta})$ in the construction (4), and all $a_{X,M,k}, M \in \mathcal{O}(\mathrm{poly}(n))$. That is, $C_m(\boldsymbol{\theta})$ exponentially concentrates on a parameter-independent fixed point F if circuit depths are linear or superlinear, and at most a polynomial number of state copies and shots are consumed. By the triangle rule, (5) shows that any pair of cost values exponentially concentrates together. Thus, this result implies that even after applying EM, at minimum an exponential number of shots (or exponential number of state copies) are required to find a cost-minimizing optimization direction with both gradient-free (such as simplex-based methods) and gradient-based optimization methods (such as using the parameter shift rule).

Second, we consider the above discussed EM protocols in a non-asymptotic setting, and ask whether these protocols can remedy some of the effects of cost concentration, even if exponential scaling is unavoidable. To this end we define a class of quantities that we call relative resolvablilities which measure how much error mitigation improves the resolvability of the landscape. We define the relative resolvability for two cost function points corresponding to parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ as

$$\chi(\boldsymbol{\theta}_{1,2}) = \frac{N_{\mathrm{noisy}}(\boldsymbol{\theta}_{1,2})}{N_{\mathrm{EM}}(\boldsymbol{\theta}_{1,2})} = \frac{1}{\gamma}\left(\frac{\Delta C_m(\boldsymbol{\theta}_{1,2})}{\Delta\widetilde{C}(\boldsymbol{\theta}_{1,2})}\right)^2, \tag{6}$$

where $N_{\mathrm{noisy}}(\boldsymbol{\theta}_{1,2})$ and $N_{\mathrm{EM}}(\boldsymbol{\theta}_{1,2})$ are the number of shots to resolve the noisy and error mitigated cost differences $|\widetilde{C}(\boldsymbol{\theta}_1) - \widetilde{C}(\boldsymbol{\theta}_2)|$ and $|C_m(\boldsymbol{\theta}_1) - C_m(\boldsymbol{\theta}_2)|$ respectively up to some fixed precision, and γ is the error mitigation cost [40]. In order to obtain state-independent results we also consider averaged versions of (6), where the average is taken over all points in the cost landscape, or taken over a set of noisy states.

Result 3: We find that in many settings for Zero Noise Extrapolation and Virtual Distillation, χ and its averaged forms take value less than or equal to 1, under mild assumptions about the nature of the cost function under some arbitrary noise model. Moreover, in certain settings it can be a very small number. This implies that, surprisingly, in such settings the cost landscape is even harder to resolve after applying EM, and trainability is further impaired. This is corroborated by our numerics for Virtual Distillation applied to a Max-Cut problem. On the other hand, we find analytically in certain settings that Clifford Data Regression (CDR) has a neutral effect on resolvability, which opens up the possibility that CDR can overall improve trainability by remedying other corruptions to the cost landscape due to noise. Indeed, we find numerical evidence of this, which points to the possibility of engineering novel EM methods to improve trainability.

[1] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, "Noise-induced barren plateaus in variational quantum algorithms," arXiv preprint arXiv:2007.14384 (2020).

[2] S. Wang, P. Czarnik, A. Arrasmith, M. Cerezo, L. Cincio, and P. J. Coles, "Can error mitigation improve trainability of noisy variational quantum algorithms?" arXiv preprint arXiv:2109.01051 (2021).

[3] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms," New Journal of Physics **18**, 023023 (2016).

[4] 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," Nature Reviews Physics **1**, 19–40 (2021).

[5] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, "Variational ansatz-based quantum simulation of imaginary time evolution," npj Quantum Information **5**, 1–6 (2019).

[6] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, "An adaptive variational algorithm for exact molecular simulations on a quantum computer," Nature communications **10**, 1–9 (2019).

[7] C. Cirstoiu, Z. Holmes, J. Iosue, L. Cincio, P. J. Coles, and A. Sornborger, "Variational fast forwarding for quantum simulation beyond the coherence time," npj Quantum Information **6**, 1–10 (2020).

[8] B. Commeau, M. Cerezo, Z. Holmes, L. Cincio, P. J. Coles, and A. Sornborger, "Variational hamiltonian diagonalization for dynamical quantum simulation," arXiv preprint arXiv:2009.02559 (2020).

[9] J. Gibbs, K. Gili, Z. Holmes, B. Commeau, A. Arrasmith, L. Cincio, P. J. Coles, and A. Sornborger, "Long-time simulations with high fidelity on quantum hardware," arXiv preprint arXiv:2102.04313 (2021).

[10] Y.-X. Yao, N. Gomes, F. Zhang, T. Iadecola, C.-Z. Wang, K.-M. Ho, and P. P. Orth, "Adaptive variational quantum dynamics simulations," arXiv preprint arXiv:2011.00622 (2020).

[11] S. Endo, J. Sun, Y. Li, S. C. Benjamin, and X. Yuan, "Variational quantum simulation of general processes," Physical Review Letters **125**, 010501 (2020).

[12] Y. Li and S. C. Benjamin, "Efficient variational quantum simulator incorporating active error minimization," Phys. Rev. X **7**, 021050 (2017).

[13] J. W. Z. Lau, K. Bharti, T. Haug, and L. C. Kwek, "Quantum assisted simulation of time dependent hamiltonians," arXiv preprint arXiv:2101.07677 (2021).

[14] K. Heya, K. M. Nakanishi, K. Mitarai, and K. Fujii, "Subspace variational quantum simulator," arXiv preprint arXiv:1904.08566 (2019).

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

[16] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, "Circuit-centric quantum classifiers," Physical Review A **101**, 032308 (2020).

[17] G. Verdon, M. Broughton, and J. Biamonte, "A quantum algorithm to train neural networks using low-depth circuits," arXiv preprint arXiv:1712.05304 (2017).

[18] J. Romero and A. Aspuru-Guzik, "Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions," Advanced Quantum Technologies **4**, 2000003 (2021).

[19] E. Farhi and H. Neven, "Classification with quantum neural networks on near term processors," arXiv preprint arXiv:1802.06002 (2018).

[20] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, "Training deep quantum neural networks," Nature Communications **11**, 808 (2020).

[21] I. Cong, S. Choi, and M. D. Lukin, "Quantum convolutional neural networks," Nature Physics **15**, 1273–1278 (2019).

[22] E. Grant, M. Benedetti, S. Cao, A. Hallam, J. Lockhart, V. Stojevic, A. G. Green, and S. Severini, "Hierarchical quantum classifiers," npj Quantum Information **4**, 1–8 (2018).

[23] D. S. Franca and R. Garcia-Patron, "Limitations of optimization algorithms on noisy quantum devices," arXiv preprint arXiv:2009.05532 (2020).

[24] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," Nature communications **9**, 1–6 (2018).

[25] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, "Cost function dependent barren plateaus in shallow parametrized quantum circuits," Nature communications **12**, 1–12 (2021).

[26] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, "Effect of barren plateaus on gradient-free optimization," arXiv preprint arXiv:2011.12245 (2020).

[27] M. Cerezo and P. J. Coles, "Higher order derivatives of quantum neural networks with barren plateaus," Quantum Science and Technology **6**, 035006 (2021).

[28] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, "Cost function dependent barren plateaus in shallow parametrized quantum circuits," Nature Communications **12**, 1791 (2021).

[29] A. Uvarov and J. Biamonte, "On barren plateaus and cost function locality in variational quantum algorithms," arXiv preprint arXiv:2011.10530 (2020).

[30] T. Volkoff and P. J. Coles, "Large gradients via correlation in random parameterized quantum circuits," Quantum Science and Technology **6**, 025008 (2021).

[31] G. Verdon, M. Broughton, J. R. McClean, K. J. Sung, R. Babbush, Z. Jiang, H. Neven, and M. Mohseni, "Learning to learn with quantum neural networks via classical neural networks," arXiv preprint arXiv:1907.05415 (2019).

[32] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, "An initialization strategy for addressing barren plateaus in parametrized quantum circuits," Quantum **3**, 214 (2019).

[33] A. Skolik, J. R. McClean, M. Mohseni, P. van der Smagt, and M. Leib, "Layerwise learning for quantum neural networks," arXiv preprint arXiv:2006.14904 (2020).

[34] K. Temme, S. Bravyi, and J. M. Gambetta, "Error mitigation for short-depth quantum circuits," Phys. Rev. Lett. **119**, 180509 (2017).

[35] S. Endo, S. C. Benjamin, and Y. Li, "Practical quantum error mitigation for near-future applications," Physical Review X **8**, 031027 (2018).

[36] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor," Nature **567**, 491–495 (2019).

[37] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, "Virtual distillation for quantum error mitigation," arXiv preprint arXiv:2011.07064 (2020).

[38] B. Koczor, "Exponential error suppression for near-term quantum devices," arXiv preprint arXiv:2011.05942 (2020).

[39] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, "Error mitigation with clifford quantum-circuit data," arXiv preprint arXiv:2005.10189 (2020).

[40] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, "Hybrid quantum-classical algorithms and quantum error mitigation," Journal of the Physical Society of Japan **90**, 032001 (2021).

LA-UR-21-30610

Scalable Bosonic Random Walk Networks for Graph Learning

Shiv Shankar, Don Towsley *University of Massachusetts*

1. Summary

The success of deep learning has led to rapid progress in its usage in a wide range of applications. Similarly, in the last few years, quantum computation has experienced exponential advancement (Steinbrecher et al., 2019). Especially so-called noisy intermediate-scale quantum (NISQ) processors have witnessed major improvements in hardware. Photonic circuits are a prime candidate for both the near-term NISQ and future quantum devices (Steinbrecher et al., 2019). In this work, we explore applications of multi-particle bosonic walks for information diffusion across graphs(Dernbach et al., 2019). Our model is based on learning the operators that govern the dynamics of bosonic quantum random walkers on graphs. We demonstrate the effectiveness of our method on a temperature prediction based regression task.

2. Quantum walks on graphs

Quantum random walks (Aharonov et al., 2001) are the quantum parallel to their classical counterparts, while a classical walker is modeled by a probability distribution over positions in a graph, a quantum walker is described by a superposition over position states.

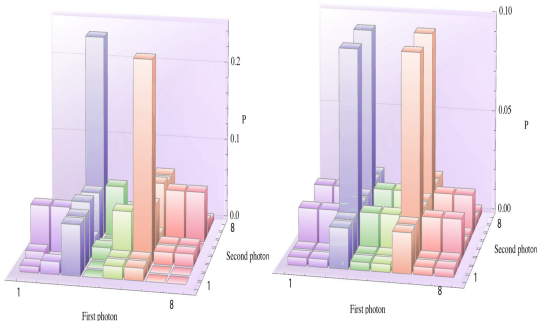


Figure 1: Two particle quantum walk with 4 steps. (left) Joint probability distribution with two indistinguishable bosons. (right) Joint probability distribution with two distinguishable particles

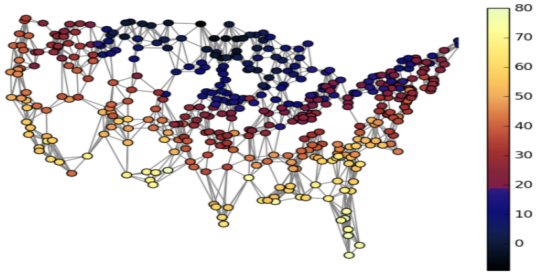


Figure 2: Example temperature predictions based on a graph constructed from the 8 nearest nodes

As a demonstration of how photonic or bosonic walks on graphs differs from the standard quantum walk; we conducted a simulation of the walks on a one-dimensional lattice induced Grover coin on (a) bosonic walkers (left) and b) standard quantum walkers(right). In Figure 1 depicts histogram plot of coincidence probabilities in the various output modes.

From the plot, it is clear that the photons show a non-independent behavior. The two photonic walkers are found clustered in nearby parts of the graph and never far from each other. At a high level, the walkers act as if they are entangled without actual entanglement. It is a discrete walk-based analog of the celebrated Hong-Ou-Mandel effect (Hong et al., 1987); and raises the possibility of performing a quantum search in a correlated manner.

3. Bosonic Walk Neural Network

A Bosonic Quantum Walker Network (QWB) incorporates bosonic quantum walk induced kernels (Bai et al., 2017) into a graph neural network. Specifically a QWB uses feature dependent coins to simulate bosonic walkers on a graph (Shankar and Towsley, 2020) and then uses the induced statistics to form a graph diffusion operator(Atwood and Towsley, 2016). This operator is used to then aggregate information from across node level representations. All of these operations are differentiable, and hence one can use backpropagation to compute the gradient of the loss with respect to all the model parameters (especially parameters of the coin matrix) .

Scaling to Large Graphs
Exact simulation of multiparticle quantum walks scales exponentially in number of walkers. As such the exact methods are difficult to use on graphs of size greater than 100. For this purpose we use an approach based motifs (Peng et al., 2019). Motifs are higher-order structures which can be thought of as capturing coarser properties of the graph. We decompose large graph into smaller subunits based on motif set and run our model on the subgraphs. Then we use motif-based attention to aggregate information across these subgraphs. This allows our model to scale on graphs which would be otherwise infeasible.

4. Experiments and Conclusion

Model	MAE
Baseline	5.929 ± 0.068
Spectral	5.148 ± 0.126
DCNN(1)	5.503 ± 0.139
DCNN(2)	5.502 ± 0.138
DCNN(3)	5.504 ± 0.139
QWNN(2)	4.383 ± 0.084
QWB(2)	$\mathbf{4.159 \pm 0.085}$

Table 1: The number in parantheses is used to denote DCNN/QWNN/QWB layer with k hops/steps/steps respectively

We ran experiments on the USTemp dataset (Williams et al., 2006) which records daily high temperatures from 409 locations across the United Sates in 2009. We form a connected graph from the stations' coordinates using 8 neighbors (Figure 2). The temperature from each station on a given day is used to predict the following day's temperatures. In our experiments, we build a networks with single quantum walk layer and vary the walk length. In Table **??** we compare our method against a diffusion convolution neural network (DCNN) , spectral graph network (Spectral) (Bruna et al., 2014), and vanilla QWNN (Dernbach et al., 2019). It is clear that the our bosonic walker QWB model is the best performing model.

The QWB model demonstrates the power of bosonic quantum techniques for deep learning. Unlike simple QWNN, this approach allows for learning significantly more powerful and complex graph diffusions.

References

Dorit Aharonov, Andris Ambainis, Julia Kempe, and Umesh Vazirani. Quantum walks on graphs. In *33rd ACM Symposium on Theory of computing*, pages 50–59, 2001.

James Atwood and Donald F. Towsley. Diffusion-convolutional neural networks. In *Proceedings of NeurIPS 2016*, 2016.

Lu Bai, Luca Rossi, Lixin Cui, Zhihong Zhang, Peng Ren, Xiao Bai, and Edwin Hancock. Quantum kernels for unattributed graphs using discrete-time quantum walks. *Pattern Recognition Letters*, 87, 02 2017.

Joan Bruna, Wojciech Zaremba, and Arthur Szlam. Spectral network learning for prediction on graphs. *ICLR 2014*, 2014.

Stefan Dernbach, Arman Mohseni-Kabir, Siddharth Pal, Miles Gepner, and Don Towsley. Quantum walk neural networks with feature dependent coins. *Applied Network Science*, 4:76, 09 2019. doi: 10.1007/s41109-019-0188-2.

Chong-Ki Hong, Zhe-Yu Ou, and Leonard Mandel. Measurement of subpicosecond time intervals between two photons by interference. *Physical review letters*, 59(18):2044, 1987.

Hao Peng, Jianxin Li, Qiran Gong, Senzhang Wang, Yuanxing Ning, and Philip S. Yu. Graph convolutional neural networks via motif-based attention, 2019.

Shiv Shankar and Don Towsley. Bosonic random walk networks for graph learning. *arXiv preprint arXiv:2101.00082*, 2020.

Gregory R Steinbrecher, Jonathan P Olson, Dirk Englund, and Jacques Carolan. Quantum optical neural networks. *npj Quantum Information*, 5(1):1–9, 2019.

CN Williams, RS Vose, DR Easterling, and MJ Menne. United states historical climatology network daily temperature, precipitation, and snow data. *Available on-line [http://cdiac.ornl.gov/epubs/ndp/ushcn/usa. html]*, 2006.

Generative Quantum Learning of Joint Probability Distribution Functions

Elton Yechao Zhu[a], Sonika Johri[b], Dave Bacon[b], Mert Esencan[a], Jungsang Kim[b], Mark Muir[a],

Nikhil Murgai[a], Jason Nguyen[b], Neal Pisenti[b], Adam Schouela[a], Ksenia Sosnova[b], Ken Wright[b]

[a]Fidelity Center for Applied Technology, FMR LLC, Boston, MA 02210, USA

[b]IonQ Inc, 4505 Campus Dr, College Park, MD 20740, USA

Link to paper: https://arxiv.org/abs/2109.06315

Abstract: We design generative learning algorithms for multivariate distributions based on a variational ansatz that we christen as a 'qopula'. We present theoretical arguments for exponential advantage in our model's expressivity over classical models. The algorithms outperform the equivalent classical generative learning algorithms when trained on IonQ quantum computers.

Modeling joint probability distributions is an important task in a wide variety of fields. A few examples of its diverse applications include risk management [1], portfolio optimization [2], reliability analysis [3], recommender systems [4], climate research [5], and medical imaging [6]. Traditionally, single-parameter quantities such as the Pearson correlation or Spearman's correlation have been used to model dependence between variables. However, such measures are good for monotonic dependence, which is frequently too simplistic for real data. Data, such as that from the financial markets, engineering reliability studies, earth/atmospheric sciences tend to exhibit tail dependence, which means they do not appear to have much correlation but exhibit dependence in extreme deviations, as in the case of a black swan event [7].

Due to the reasons above, the relationship between random variables is now commonly modeled using a dependence function between uniformly distributed variables, called a 'copula'. Sklar's theorem [8] states that any multivariate joint distribution can be written in terms of univariate marginal distributions and a copula that describes the dependence structure between the variables. Empirical copulas (copulas from real data) tend to be a mixture of copulas and are commonly modeled using parametric methods like maximum likelihood estimation [12]. As a rule, the more complex the copula and the more completely it describes the data, the more computationally challenging it becomes to extend it to higher dimensional data. More recently, generative models have been proposed for statistical modeling that learn to generate data with the same statistics as a given training dataset, effectively learning its distribution. The model can be used to output new samples that could plausibly have belonged to the original dataset [13, 14].

Here we use IonQ quantum computers to train generative models called the Quantum Circuit Born Machine (QCBM) [15, 16] and Quantum Generative Adversarial Network (QGAN) [17, 18] to learn the joint distributions of two variables for historical data from the stock market: the daily return of AAPL and MSFT between 2010-2018. Table I shows how the results of quantum training compare to classical for the 2-dimensional 2-sample Kolmogorov-Smirnov (KS) test. As one can see, a classical parametric model does a good job in modeling the copula, but generative models are able to do better. The result from QGAN is consistent and outperforms classical GAN with similar number of parameters. The QCBM performs slightly worse than the QGAN but comparable to the classical GAN.

We also note that for QGAN, our model converges regardless of initialization. For classical GAN, only about 40% of our model instances are accepted by the 2D KS test with threshold 0.05. The rest failed to learn and were rejected by the test. Another advantage is that we are able to train QGAN/QCBM at a much faster learning rate and therefore conclude the training with much fewer iterations than classical GAN. In classical GAN, the learning rate used is 0.0001 and model training

concludes after 20000 iterations. Attempts to increase the learning rate failed due to non-convergence in model training. In QGAN, model training concludes after 1000 iterations. In QCBM for 6 qubits, the training converges to a good value for as little as 20 iterations.

Model	D_{KS} (the smaller the better)	p-value (threshold 0.05)
Parametric model	0.0449	0.117
Classical GAN	0.0363 - 0.0508	0.0530 - 0.309
QGAN simulation	0.0320 - 0.0396	0.226 - 0.473
QGAN experiment, QPU cloud	0.0352	0.3570
QCBM simulation	0.0425 - 0.0520	0.0511 - 0.1717
QCBM experiment, QPU cloud	0.0373 - 0.0515	0.0548 - 0.3030
QCBM experiment, QPU Next Gen	0.0330 - 0.0510	0.0578 - 0.4465

Table 1: KS statistics and p-value of KS test across multiple models. The quantum models use $N_q = 6$ qubits.

Our quantum generative algorithms are based on a variational ansatz which we christen as a 'qopula'. We show that every copula with density can be represented by a maximally entangled state. Here we construct a quantum circuit that can prepare such maximally entangled states up to relative phases, which we christen as a 'qopula' circuit. The quantum circuit for two random variables is shown in Fig. 1.

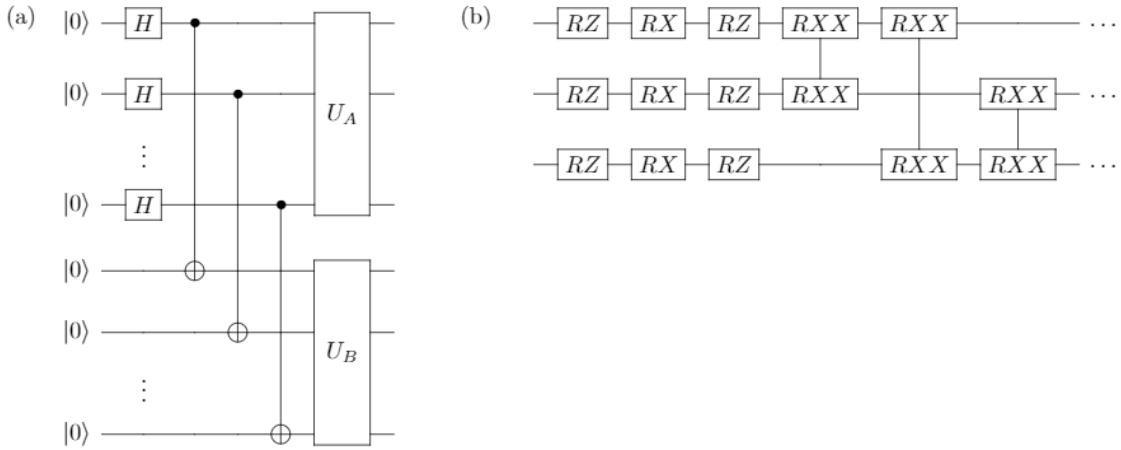


Fig. 1: (a) The 'qopula' circuit for 2 random variables. The top (bottom) half of the circuit consists of qubits that belong to the register which provides samples of the first (second) random variable. (b) The ansatz for U corresponding to 3 qubits. All gates are parametrized by angles which are optimized during the learning process. The structure can be repeated for multiple layers each with different parameters. Here the 2-qubit RXX gates represent $\exp(i\theta X_i X_j)$ acting on qubits i and j, which can be executed as $\mathrm{CNOT}(i,j)\exp(i\theta X_i)\mathrm{CNOT}(i,j)$ using standard gates available on gate-model quantum computers. The RZ and RX gates represent rotations around the Z and X axes.

Finally, we present theoretical arguments for exponential advantage in our model's expressivity over classical models based on communication and computational complexity arguments. We also project our algorithms to start having commercial value when running on quantum hardware with 50 qubits.

References:

[1] G. James, D. Witten, T. Hastie & R. Tibshirani, “An Introduction to Statistical Learning: with Applications in R”, Springer (2013).
[2] H. Markowitz, “Portfolio Selection”, The Journal of Finance 7, 77 (1952).
[3] S. Zacks, “Introduction to Reliability Analysis: Probability Models and Statistical Methods”, SpringerVerlag New York (1992).
[4] C. C. Aggarwal, “Recommender Systems: The Textbook”, 1st edition, Springer Publishing Company, Incorporated (2016).
[5] H. v. Storch & F. W. Zwiers, “Statistical Analysis in Climate Research”, Cambridge University Press (1999).
[6] C. Palmer, “Statistics of Medical Imaging”, Health Physics 104, (2013).
[7] R. Aloui, M. S. B. Aissa & D. K. Nguyen, “Global Financial Crisis, Extreme Interdependences, and Contagion Effects: The Role of Economic Structure?”, Journal of Banking and Finance 35, 130 (2011); D. Kenourgios, A. Christopoulos & D. I. Dimitriou, “Asset Markets Contagion During the Global Financial Crisis”, Mutual Funds, (2013); F. Serinaldi, A. B´ardossy & C. G. Kilsby, “Upper tail dependence in rainfall extremes: would we know it if we saw it?”, Stochastic Environmental Research and Risk Assessment 29, 1211 (2015).
[8] M. Sklar, “Fonctions de repartition a n dimensions et leurs marges”, Publications de l’Institut Statistique
de l’Universit´e de Paris (1959).
[9] R. W. van den Goorbergh, C. Genest & B. J. Werker, “Bivariate option pricing using dynamic copula models”, Insurance: Mathematics and Economics 37, 101 (2005), Papers presented at the DeMoSTAFI Conference, Qu´ebec, 20-22 May 2004.
[10] R. T. Kilgore & D. B. Thompson, “Estimating Joint Flow Probabilities at Stream Confluences by Using Copulas”, Transportation Research Record 2262, 200 (2011).
[11] J. Lapuyade-Lahorgue, J.-H. Xue & S. Ruan, “Segmenting Multi-Source Images Using Hidden Markov Fields With Copula-Based Multivariate Statistical Distributions”, IEEE Transactions on Image Processing 26, 3187 (2017).
[12] H. Joe, “Multivariate Models and Multivariate Dependence Concepts”, Chapman and Hall/CRC (1997), New York.
[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville & Y. Bengio, “Generative Adversarial Nets”, in “Advances in Neural Information Processing Systems”.
[14] D. P. Kingma & M. Welling, “Auto-Encoding Variational Bayes”, in “2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings”, ed: Y. Bengio & Y. LeCun.
[15] S. Cheng, J. Chen & L. Wang, “Information Perspective to Probabilistic Modeling: Boltzmann Machines versus Born Machines”, Entropy 20, (2018).
[16] M. Benedetti, D. Garcia-Pintos, O. Perdomo, V. Leyton-Ortega, Y. Nam & A. Perdomo-Ortiz, “A generative modeling approach for benchmarking and training shallow quantum circuits”, npj Quantum Information 5, 45 (2019).
[17] P.-L. Dallaire-Demers & N. Killoran, “Quantum generative adversarial networks”, Phys. Rev. A 98, 012324 (2018).
[18] S. Lloyd & C. Weedbrook, “Quantum Generative Adversarial Learning”, Phys. Rev. Lett. 121, 040502 (2018).

Estimating distinguishability measures on quantum computers

Rochisha Agarwal,[1] Soorya Rethinasamy,[2] Kunal Sharma,[3] and Mark M. Wilde[2]

[1]*Department of Physics, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India*
[2]*Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, and Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA*
[3]*Joint Center for Quantum Information and Computer Science, University of Maryland, College Park, Maryland 20742, USA*

(Dated: October 23, 2021)

Full version available at https://arxiv.org/abs/2108.08406

In quantum information processing, it is essential to quantify the performance of protocols by using distinguishability measures. It is typically the case that there is an ideal state to prepare or an ideal channel to simulate, but in practice, we can only realize approximations, due to experimental error. Two commonly employed distinguishability measures for states are the trace distance [1, 2] and the fidelity [3]. These distinguishability measures have generalizations to quantum channels, in the form of the diamond distance [4] and the fidelity of channels [5], as well as to strategies (sequences of channels), in the form of the strategy distance [6–8] and the fidelity of strategies [9].

Both the trace distance and the fidelity can be computed by means of semi-definite programming [10], so that they can be estimated accurately with a run-time that is polynomial in the dimension of the states. The same is true for the diamond distance [11], fidelity of channels [12, 13], the strategy distance [6–8], and the fidelity of strategies [9]. While this method of estimating these quantities is reasonable for states, channels, and strategies of small dimension, its computational complexity actually increases exponentially with the number of qubits involved, due to the well known fact that Hilbert-space dimension grows exponentially with the number of qubits.

In our paper, we provide several quantum algorithms for estimating these distinguishability measures. Some of the algorithms rely on interaction with a quantum prover, in which case they are not necessarily efficiently computable even on a quantum computer. In fact, the computational hardness results of [14–16] lend credence to the belief that estimating these quantities reliably is not generally possible in polynomial time on a quantum computer. However, as we show in our paper, by replacing the quantum prover with a parameterized circuit (see [17, 18] for reviews of variational algorithms), it is possible in some cases to estimate these quantities reliably. Already in [19], it was shown that estimating the fidelity of two quantum states is possible in quantum polynomial time when one of the states is low rank. See also [20–22] for variational algorithms that estimate fidelity of states and [21, 23] for variational algorithms to estimate trace distance. It is open to determine conditions under which precise estimation is possible for channel and strategy distinguishability measures.

One of the other results of our paper is that the problem of estimating the fidelity between a mixed state and a pure state is a BQP-complete promise problem (see [24, 25] for reviews of quantum computational complexity theory). The aforementioned result follows by demonstrating that there is an efficient quantum algorithm for this task and by showing a reduction from an arbitrary BQP algorithm to one for this task. Thus, if we believe that there is a separation between the computational power of classical and quantum computers, then this fidelity estimation problem is one for which a quantum computer has an advantage.

In our paper, we provide details of the algorithms and results mentioned above. In particular, our paper proceeds as follows:

1. We begin by establishing two quantum algorithms for estimating the fidelity of pure states,

one of which is based on a state overlap test and another that employs Bell state preparation and measurement along with a controlled unitary.

2. Next we generalize the first algorithm to estimate the fidelity of a pure state and a mixed state, and we prove that the promise version of this problem is BQP-complete.

3. We also establish several quantum algorithms for estimating the fidelity of two arbitrary states. Algorithm 4 generalizes the second algorithm. Algorithm 5 generalizes the well known swap test to the case of arbitrary states. Algorithm 6 is a variational algorithm that employs Bell measurements, as a generalization of the approach in [26, 27] for pure states. Algorithm 7 is another variational algorithm that attempts to simulate a fidelity-achieving measurement, such as the Fuchs–Caves measurement [28], in order to estimate the fidelity.

4. We generalize Algorithm 4 to a quantum algorithm for estimating the fidelity of quantum channels. This algorithm involves interaction with competing quantum provers, and interestingly, its acceptance probability is directly related to the fidelity of channels, thus giving the latter an operational meaning. Later, we replace the provers with parameterized circuits and arrive at a method for estimating fidelity of channels.

5. We also generalize the aforementioned approach in order to estimate the fidelity of strategies (a strategy is a sequence of quantum channels, which generalizes the notion of a channel).

6. Next we briefly discuss alternative methods for estimating the fidelity of channels and strategies, based on the approaches for estimating the fidelity of states.

7. We introduce a method for estimating the maximum output fidelity of two quantum channels, which has an application to generating a fixed point of a quantum channel.

8. We also generalize the whole development above to the case of testing similarity of arbitrary ensembles of states, channels, or strategies. We find that the acceptance probability of the corresponding algorithms is related to the secrecy measure from [29], which can be understood as a measure of similarity of the states in an ensemble. We then establish generalizations of this measure for an ensemble of channels and an ensemble of strategies and remark how this has applications in private quantum reading [30, 31].

9. We then move on to estimating trace-distance-based measures, for states, channels, and strategies. We stress that these various algorithms were already known, and our goal here is to investigate their performance using a variational approach.

10. Next we provide two different but related algorithms for estimating the minimum trace distance between two quantum channels. The related approaches employ competing provers.

11. We also generalize the whole development for trace-distance based algorithms to the case of multiple states, channels, and strategies.

12. Finally, we discuss the results of numerical simulations of algorithms for estimating the fidelity and trace distance of states, when using a noiseless simulator of a quantum computer along with a variational approach with parameterized circuits. We find that, among all algorithms for estimating the fidelity, Algorithm 6 that uses Bell measurement approach converges the fastest with the smallest error.

[1] Carl W. Helstrom. Detection theory and quantum mechanics. *Information and Control*, 10(3):254–291, 1967.

[2] Carl W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1:231–252, 1969.

[3] Armin Uhlmann. The "transition probability" in the state space of a *-algebra. *Reports on Mathematical Physics*, 9(2):273–279, April 1976.

[4] Alexei Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[5] Alexei Gilchrist, Nathan K. Langford, and Michael A. Nielsen. Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71(6):062310, June 2005. arXiv:quant-ph/0408063.

[6] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. Memory effects in quantum channel discrimination. *Physical Review Letters*, 101(18):180501, October 2008. arXiv:0803.3237.

[7] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. Theoretical framework for quantum networks. *Physical Review A*, 80(2):022339, August 2009. arXiv:0904.4483.

[8] Gus Gutoski. On a measure of distance for quantum strategies. *Journal of Mathematical Physics*, 53(3):032202, March 2012. arXiv:1008.4636.

[9] Gus Gutoski, Ansis Rosmanis, and Jamie Sikora. Fidelity of quantum strategies with applications to cryptography. *Quantum*, 2:89, September 2018. arXiv:1704.04033.

[10] John Watrous. Simpler semidefinite programs for completely bounded norms. *Chicago Journal of Theoretical Computer Science*, 2013(8):1–19, July 2013. arXiv:1207.5726.

[11] John Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5(11):217–238, November 2009. arXiv:0901.4709.

[12] Haidong Yuan and Chi-Hang Fred Fung. Fidelity and Fisher information on quantum channels. *New Journal of Physics*, 19(11):113039, November 2017. arXiv:1506.00819.

[13] Vishal Katariya and Mark M. Wilde. Geometric distinguishability measures limit quantum channel estimation and discrimination. *Quantum Information Processing*, 20:78, April 2021. arXiv:2004.10708.

[14] John Watrous. Limits on the power of quantum statistical zero-knowledge. *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 459–468, November 2002. arXiv:quant-ph/0202111.

[15] Bill Rosgen and John Watrous. On the hardness of distinguishing mixed-state quantum computations. *Proceedings of the 20th IEEE Conference on Computational Complexity*, pages 344–354, June 2005. arXiv:cs/0407056.

[16] John Watrous. Zero-knowledge against quantum attacks. *SIAM Journal on Computing*, 39(1):25–58, 2009. arXiv:quant-ph/0511020.

[17] 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. December 2020. arXiv:2012.09265.

[18] 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. January 2021. arXiv:2101.08448.

[19] Qisheng Wang, Zhicheng Zhang, Kean Chen, Ji Guan, Wang Fang, and Mingsheng Ying. Quantum algorithm for fidelity estimation. March 2021. arXiv:2103.09076.

[20] M. Cerezo, Alexander Poremba, Lukasz Cincio, and Patrick J. Coles. Variational quantum fidelity estimation. *Quantum*, 4:248, March 2020. arXiv:1906.09253.

[21] Ranyiliu Chen, Zhixin Song, Xuanqiang Zhao, and Xin Wang. Variational quantum algorithms for trace distance and fidelity estimation. December 2020. arXiv:2012.05768.

[22] Kok Chuan Tan and Tyler Volkoff. Variational quantum algorithms to estimate rank, quantum entropies, fidelity and Fisher information via purity minimization. March 2021. arXiv:2103.15956.

[23] Sheng-Jie Li, Jin-Min Liang, Shu-Qian Shen, and Ming Li. Variational quantum algorithms for trace norms and their applications. *Communications in Theoretical Physics*, 2021. Accepted manuscript.

[24] John Watrous. Quantum computational complexity. *Encyclopedia of Complexity and System Science*, 2009. arXiv:0804.3401.

[25] Thomas Vidick and John Watrous. Quantum proofs. *Foundations and Trends in Theoretical Computer Science*, 11(1–2):1–215, March 2016. arXiv:1610.01664.

[26] Juan Carlos Garcia-Escartin and Pedro Chamorro-Posada. SWAP test and Hong-Ou-Mandel effect are equivalent. *Physical Review A*, 87(5):052330, May 2013. arXiv:1303.6814.

[27] Yiğit Subaşı, Lukasz Cincio, and Patrick J Coles. Entanglement spectroscopy with a depth-two quantum circuit. *Journal of Physics A: Mathematical and Theoretical*, 52(4):044001, January 2019. arXiv:1806.08863.

[28] Christopher A. Fuchs and Carlton M. Caves. Mathematical techniques for quantum communication theory. *Open Systems & Information Dynamics*, 3(3):345–356, 1995. arXiv:quant-ph/9604001.

[29] Robert Koenig, Renato Renner, and Christian Schaffner. The operational meaning of min- and max-entropy. *IEEE Transactions on Information Theory*, 55(9):4337–4347, September 2009. arXiv:0807.1338.

[30] Stefan Bäuml, Siddhartha Das, and Mark M. Wilde. Fundamental limits on the capacities of bipartite quantum interactions. *Physical Review Letters*, 121(25):250504, December 2018. arXiv:1812.08223.

[31] Siddhartha Das, Stefan Bäuml, and Mark M. Wilde. Entanglement and secret-key-agreement capacities of bipartite quantum interactions and read-only memory devices. *Physical Review A*, 101(1):012344, January 2020. arXiv:1712.00827.

Storage capacity and learning capability of quantum neural networks

Maciej Lewenstein,[1,2,*] Aikaterini Gratsea,[1] Andreu Riera-Campeny,[3] Albert Aloy,[1] Valentin Kasper,[1] and Anna Sanpera[3,2]

[1]*ICFO - Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*
[2]*ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain*
[3]*Física Teòrica: Informació i Fenòmens Quàntics. Departament de Física, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain*

(Dated: October 23, 2021)

We study the storage capacity of quantum neural networks (QNNs), described by completely positive trace preserving (CPTP) maps acting on a N-dimensional Hilbert space with n qubits. We explore the number of stationary states that QNNs can store and the relative volume of CPTP by applying the Gardner program.

The full article can be found here: https://iopscience.iop.org/article/10.1088/2058-9565/ac070f .

I. INTRODUCTION

One crucial feature of NNs is their storage capacity for associative memory, that is, the number of *patterns* (stored memories/attractors) the network has for a given number of neurons n. For attractor NNs (aNNs) of the Hopfield-type [1], the relevant question is to determine how many stationary states, serving as stored memories, the network may have. In standard Hopfield models, where neurons are Ising spins and attractors correspond to metastable states resulting from two-body spin-spin interactions, the storage capacity scales $\sim O(n)$. Recently, it has been shown that if the Hopfield model is extended to p-body interactions, in the so-called Dense Associative Network model [2], the capacity of storage can be highly increased, surpassing the linear $O(n)$ behaviour and reaching $O(n^{p-1}/\ln n)$, or even beyond [3]. For feed-forward NNs, with the paradigmatic example of the perceptron [4], the corresponding question is how many attractor input-output relations can be stored. The problem of the storage capacity and learning ability of NNs was reformulated by the seminal contributions of Gardner [5, 6]. She provided the relative volume of NNs with a desired set of patterns in the full space of NNs or, equivalently, the relative volume of feed-forward NNs with desired input-output relations. Sharp shrinking of the relative volume to zero, heralds the phase transition corresponding to an overloaded NN memory.

In the recent decades, quantum information science has demonstrated that information processing can be significantly improved by exploiting quantum mechanics. Not surprisingly, both areas, ML and quantum information have merged together in the so-called quantum machine learning (QML) [7, 8]. Preliminary attempts to analyze the storage capacity of QNNs were pursued in [9]. In a different approach, an exponential increase of the storage capacity for a specific quantum search algorithm was demonstrated in [10]. More recently [11, 12], an increased storage capacity was obtained by using a feed-forward interpretation of quantum Hopfield NNs. Similarly, qudits have been studied in the context of quantum machine learning [13, 14]. Despite this progress, the storage capacity of generic QNNs remains an open problem. In this work we address and solve this question by associating QNNs to CPTP maps. We also analyze the learning capability of QNN's by applying Gardner's program to the quantum case and estimate the relative volume of QNNs realizing the desired attractor input-output relations.

II. METHODS AND RESULTS

We associate QNNs with CPTP maps transforming initial states into final states in a finite (or infinite) time. Attractors (stored memory/patterns) correspond to the stationary states of the map, i.e., $\Lambda(\rho) = \rho$. We identify the storage capacity of QNNs (number of stored memories) with the maximal number of stationary points of CPTP maps acting on density matrices in N-dimensional Hilbert spaces. We demonstrate that there exist a family of (non-trivial) CPTP maps that have $M = N$ linearly independent stationary pure states, and provide the generic expression of such

*maciej.lewenstein@icfo.eu

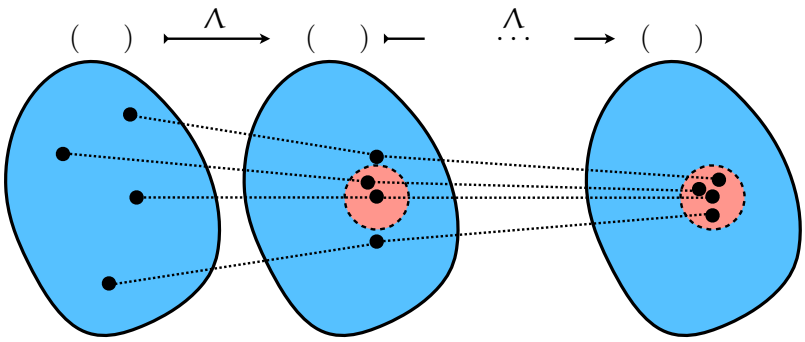


FIG. 1: Color Online. Schematic representation of the action of CPTP maps Λ with N fixed states. Successive applications of $\Lambda : \mathcal{B}(\mathcal{H}_A) \mapsto \mathcal{B}(\mathcal{H}_A)$, brings arbitrary states $\rho \in \mathcal{B}(\mathcal{H}_A)$ to the set (depicted by red area) of stationary states of the map.

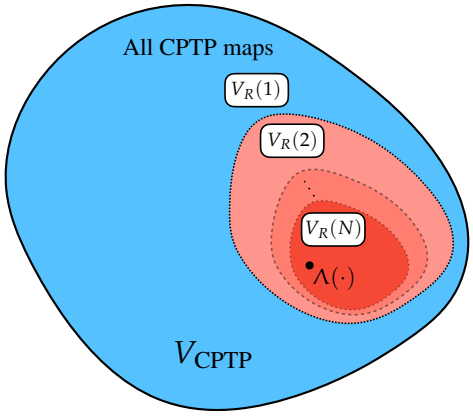


FIG. 2: Color online. Representation of the relative volume $V_R(M)$ of CPTP maps acting as aQNN and storing M stationary pure states. The volume shrinks as we increase the number of stationary states from $V_{\text{CPTP}} = V_R(1)$ for $M = 1$, to $V_R(N)$ for $M = N$.

maps. These maps act as attractors in the space of states, i.e., the successive application of the map brings an arbitrary state to the set of its fixed points, see Fig. 1. We interpret this class of maps as attractors QNNs (aQNNs). Further, we estimate the relative volume of CPTP maps that have exactly $1 < M \leq N$ pure stationary states. This calculation corresponds to a quantum version of the Gardner program. We show that, in the limit of large N, the relative volume in the space of CPTP maps capable to store M patterns decreases very slowly with M as $\exp(-M^2/(N^4 - N^2))$. Our results signal quantum advantage meaning that CPTP maps acting on n-qubit states may reach a storage capacity of $\mathcal{O}(2^n)$, surpassing the storage capacity of classical neural networks, including Dense Associative Networks. We derive analogous results for bilayer QNNs and their respective attractor input-output relations. Here, we do not restrict the CPTP maps by any notion of k-locality, which may reduce the number of stored states. Finally, we discuss the extension of our results to generic feed-forward NN and provide the details in the Appendix.

III. CONCLUSIONS AND OUTLOOK

We have demonstrated, using CPTP maps acting on a Hilbert space of dimension N, that aQNN's can store up to N linear independent pure states. For n qubits, quantum channels reach thus the capacity 2^n, clearly outperforming the storage capacity of standard classical neural networks $\sim O(n)$, where n is the number of binary neurons, or the best Dense Associative Networks whose storage capacity $\sim O(2^{(n/2)})$. Applying Gardner's program to the quantum case, we have related the learning capability of aQNN's to the relative volume $V_R(M)$ of CPTP maps with M stationary pure states, and show that this volume decreases very slowly with the number of stored patterns M. Finally, we have applied our procedure also to feed-forward QNN with different input and output spaces. Our results are simple and mathematically rigorous. Furthermore, they open the path to study the relation between the storage capacity of QNNs and the quantum features, such as coherence and entanglement, of the desired attractor input-output relations.

[1] J. J. Hopfield, Proc. Natl. Acad. Sci. **79**, 2554 (1982).
[2] D. Krotov and J. J. Hopfield, (2016), arXiv:1606.01164 .
[3] M. Demircigil, J. Heusel, M. Löwe, S. Upgang, and F. Vermet, Journal of Statistical Physics **168**, 288 (2017).
[4] M. Minsky, S. A. Papert, and L. Bottou, *Perceptrons: An Introduction to Computational Geometry* (MIT Press, Cambridge, 2017).
[5] E. Gardner, J. Phys. A. Math. Gen. **21**, 257 (1988).
[6] E. Gardner and B. Derrida, J. Phys. A Math. Gen. **21**, 271 (1988).
[7] P. Wittek, *Quantum Machine Learning: What Quantum Computing Means to Data Mining* (Elsevier Science & Technology, San Diego, 2014).
[8] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Nature **549**, 195 (2017).
[9] M. Lewenstein, J. Mod. Opt. **41**, 2491 (1994).
[10] D. Ventura and T. Martinez, *1998 IEEE Int. Jt. Conf. Neural Networks Proceedings*, Vol. 1 (1998) pp. 509–513.
[11] P. Rebentrost, T. R. Bromley, C. Weedbrook, and S. Lloyd, Phys. Rev. A **98**, 42308 (2018).
[12] N. Meinhardt, N. M. P. Neumann, and F. Phillipson, *Quantum Hopfield Neural Networks*, edited by V. V. Krzhizhanovskaya, G. Závodszky, M. H. Lees, J. J. Dongarra, P. M. A. Sloot, S. Brissos, and J. Teixeira (Springer International Publishing, 2020) pp. 576–590.
[13] G. Verdon, T. McCourt, E. Luzhnica, V. Singh, S. Leichenauer, and J. Hidary, "Quantum graph neural networks," (2019), arXiv:1909.12264 [quant-ph] .
[14] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Nature Communications **11** (2020), 10.1038/s41467-020-14454-2.

Quantum Circuits for Graph Representation Learning*

Peter Mernyei[1], Konstantinos Meichanetzidis[1,2], İsmail İlkan Ceylan[1]

[1]University of Oxford, United Kingdom
[2]Cambridge Quantum Computing, United Kingdom

Abstract

We investigate quantum circuits for learning functions over graphs, proposing a unifying framework of *Equivariant Quantum Graph Circuits* subsuming earlier work as special cases and discuss other possible subclasses. We prove that these circuits are universal approximators for functions over the graph domain and provide experimental evidence.

Background

Graph neural networks (GNNs) are deep neural networks that are used to predict properties of nodes or entire graphs in a way that respects their structural invariances: the ordering of nodes and edges in the representation should not matter [5, 3]. GNNs achieved impressive results in graph representation learning on a wide range of benchmarks, and there has also been significant work on the theoretical capabilities of such models. This includes results showing that the expressive power of a broad class of popular GNN models is bounded by the 1-dimensional Weisfeiler Leman heuristic (for graph isomorphism testing), and, as a result, some pairs of non-isomorphic graphs always result in the same predictions with these methods (see, e.g., [2, 11, 6]). Recent work has focused on alleviating this limitation at the cost of more computational power [8, 6] or by introducing randomisation [1, 9].

Meanwhile, there have been a number of proposals for quantum analogues of GNNs, typically using sets of qubits to represent information about each node, and performing multi-node entangling operations according to the structure of the graph in question [10, 12]. This lets us make use of the exponentially large Hilbert space of the joint system to model complex interactions via entanglement between the node states. Such methods have been supported by empirical data from small-scale experiments, particularly when applied to modelling quantum systems, but to date there has been no clarity on how their expressivity compares to classical approaches.

Equivariant Quantum Graph Circuits

We propose a unifying framework of *Equivariant Quantum Graph Circuits* (EQGCs) to formalise the notion of quantum circuits that respect the invariances of the graph domain, show that it subsumes existing methods as special subclasses, and prove powerful results about the expressive power of such models, applying to several of these classes.

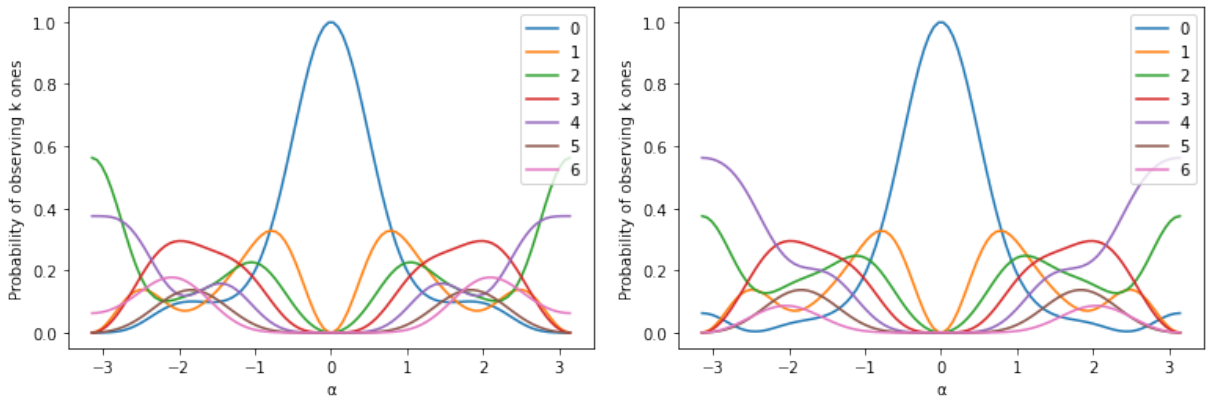
Similarly to Verdon et al. [10] and Zheng et al. [12], we represent nodes in a quantum circuit as a joint system given by the tensor product of subsystems for each node. We set the initial state to be a product of node states encoding each node's features: $|\psi\rangle = \bigotimes_{i=1}^{n} |v_i\rangle \in \mathbb{C}^{s^n}$. We only make measurements at the end of the process, so we wish to learn unitary transformations mapping this input to a useful output state where we can measure each node. This unitary may depend on the adjacency matrix, but the ordering of nodes and edges should not matter. This is formalised in the following definition:

Definition 1. An EQGC is an arbitrary function $\boldsymbol{C}(\cdot)$ mapping an adjacency matrix $\boldsymbol{A} \in \mathbb{B}^{n \times n}$ to a unitary $\boldsymbol{C}(\boldsymbol{A}) \in \mathbb{C}^{s^n \times s^n}$ for any n that behaves equivariantly in the following sense:

For any permutation p over n elements, consider its usual representation by $\boldsymbol{P} \in \mathbb{B}^{n \times n}$ as well as a larger matrix $\tilde{\boldsymbol{P}} \in \mathbb{B}^{s^n \times s^n}$ that reorders the tensor product, mapping any $|v_1\rangle |v_2\rangle \dots |v_n\rangle$ with $|v_i\rangle \in \mathbb{C}^s$ to $|v_{p(1)}\rangle |v_{p(2)}\rangle \dots |v_{p(n)}\rangle$. Then $\boldsymbol{C}$ must satisfy:

$$\boldsymbol{C}(\boldsymbol{A}) = \tilde{\boldsymbol{P}}^T \boldsymbol{C}(\boldsymbol{P}^T \boldsymbol{A} \boldsymbol{P}) \tilde{\boldsymbol{P}} \tag{1}$$

*This is an extended abstract of the results first presented as part of the thesis of the first author [7], available here.



(a) Graph G_1: two triangles

(b) Graph G_2: one 6-cycle

Figure 1: Distribution of the number of measured $|1\rangle$s out of six qubits when applying a restricted EDU-QGC over two 1-WL indistinguishable graphs as a function of the model parameter α.

We consider models for graph classification and regression by encoding node features as a product state, applying an EQGC, measuring all nodes and classically aggregating the results to provide predictions.

This definition is very general due to $\boldsymbol{C}$ being an arbitrary function of $\boldsymbol{A}$. In principle, this allows the unitaries given for non-isomorphic graphs to be chosen completely independently. While this gives us a flexible framework to discuss possible approaches, we clearly need to choose some further restricted subclass if we hope to inductively generalise to unseen graphs, and to represent models with a finite number of parameters. We therefore propose two subclasses: *Equiariant Hamiltonian Quantum Graph Circuits* (EH-QGC) closely related to the Quantum Graph Convolutional Neural Networks proposed by Verdon et al. and *Equivariantly Diagonalisable Unitary Quantum Graph Networks* (EDU-QGC) that can be seen as a restriction of the work of Zheng et al. that ensures Equation 1 is respected [12, 10]. We prove that EDU-QGCs are contained in EH-QGCs.

Overview of the Results

We prove powerful results about the expressivity of our quantum models similar to recent work on randomisation in classical GNNs [1, 9]. Note that unlike in the classical case, where this randomisation had to be explicitly added to extend model capacity, we can do this *without modifying our model definition* in the quantum case – our results apply to EDU-QGCs and their superclasses.

We first show that EDU-QGCs can simulate a large and popular class of classical graph neural networks called *message-passing neural networks* (MPNNs) [4]:

Theorem 1. *EDU-QGCs can represent any MPNN with sum aggregation. For an MPNN with k layers with an embedding dimensionality of w, with a fixed-point real representation of b bits per real number, this EDU-QGC needs $(2k+1)wb$ qubits per node.*

Building on this result, and establishing a correspondence between randomised MPNNs and quantum circuits, we show universality over bounded-size graphs:

Theorem 2. *For any real function f defined over graphs up to size n, and any $\epsilon > 0$, there is an EDU-QGC that calculates $f(\mathcal{G})$ with probability $(1-\epsilon)$ for any graph $\mathcal{G}$.*

To empirically verify our findings, we consider a pair of graphs, known to be indistinguishable by MPNNs, apply certain restricted EDU-QGCs with a single parameter α to these graphs, and compare the probability of possible measurements. As shown in Figure 1, the distributions of outcomes do differ, particularly around $\alpha \approx \pm\pi$.

Although our constructions do not show quantum advantage over classical models with randomisation, since the specific constructions we give could be simply executed classically, our results serve as important steps towards a better theoretical understanding of quantum methods. Additionally, through the EQGC framework, we provide a unifying perspective on possible quantum architectures that can help inform future work. If we can find a class of useful functions that are significantly more costly to learn classically than in the quantum setting, perhaps related to the behaviour of molecules or other quantum systems, that would be very relevant for applications and a good candidate for practical quantum advantage.

References

[1] Ralph Abboud et al. "The Surprising Power of Graph Neural Networks with Random Node Initialization". In: *IJCAI*. 2021.

[2] Pablo Barceló et al. "The Logical Expressiveness of Graph Neural Networks". In: *ICLR*. 2020.

[3] Peter W Battaglia et al. "Relational inductive biases, deep learning, and graph networks". In: *arXiv preprint arXiv:1806.01261* (2018).

[4] Justin Gilmer et al. "Neural Message Passing for Quantum Chemistry". In: *ICML-17*. 2017.

[5] William L. Hamilton. "Graph Representation Learning". In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* 14.3 ().

[6] Haggai Maron et al. "Provably Powerful Graph Networks". In: *NeurIPS*. 2019.

[7] Peter Mernyei. "Quantum Computing for Graph Representation Learning". Master's Thesis. University of Oxford, 2021.

[8] Christopher Morris et al. "Weisfeiler and Leman Go Neural: Higher-order Graph Neural Networks". In: *AAAI*. 2019.

[9] Ryoma Sato, Makoto Yamada, and Hisashi Kashima. "Random features strengthen graph neural networks". In: *SDM*. 2021.

[10] Guillaume Verdon et al. "Quantum graph neural networks". In: *arXiv preprint arXiv:1909.12264* (2019).

[11] Keyulu Xu et al. "How Powerful are Graph Neural Networks?" In: *ICLR*. 2019.

[12] Jin Zheng et al. "Quantum Graph Convolutional Neural Networks". In: *arXiv preprint arXiv:2107.03257* (2021).

Large-scale quantum machine learning

Tobias Haug,[1,*] Chris N. Self,[1] and M. S. Kim[1]

[1] *QOLS, Blackett Laboratory, Imperial College London SW7 2AZ, UK*

We substantially speed up quantum machine learning of large datasets via randomized measurements of quantum kernels. We identify a high-dimensional data encoding that is characterized by quantum geometric measures. Our technique successfully classifies handwritten images using IBM quantum computers with parallel processing and a complementary noise mitigation method.

Quantum computers promise to enhance machine learning for practical applications [1, 2]. Quantum machine learning for real-world data has to handle extensive amounts of high-dimensional data. However, conventional methods for measuring quantum kernels are impractical for large datasets as they scale with the square of the dataset size. Here, we measure quantum kernels using randomized measurements to gain a quadratic speedup in quantum computation time and quickly process large datasets [3]. Further, we efficiently encode high-dimensional data into quantum computers with the number of features scaling linearly with the circuit depth. The encoding is characterized by the quantum Fisher information metric [4] and is related to the radial basis function kernel [5, 6]. We demonstrate the advantages of our methods by classifying images of handwritten digits with the IBM quantum computer. To achieve further speedups we distribute the quantum computational tasks between different quantum computers. Our approach is exceptionally robust to noise via a complementary error mitigation scheme. Using currently available quantum computers [7], the MNIST database can be processed within 220 hours instead of 10 years which opens up industrial applications of quantum machine learning.

* thaug@ic.ac.uk

[1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature **549**, 195 (2017).

[2] M. Schuld and F. Petruccione, *Supervised learning with quantum computers*, Vol. 17 (Springer, 2018).

[3] T. Haug, C. N. Self, and M. S. Kim, Large-scale quantum machine learning, arXiv:2108.01039 (2021).

[4] T. Haug, K. Bharti, and M. Kim, Capacity and quantum geometry of parametrized quantum circuits, PRX Quantum **2**, 040309 (2021).

[5] T. Haug and M. S. Kim, Optimal training of variational quantum algorithms without barren plateaus,

arXiv:2104.14543 (2021).

[6] T. Haug and M. S. Kim, Natural parameterized quantum circuit, arXiv:2107.14063 (2021).

[7] 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).

The Inductive Bias of Quantum Kernels*

Jonas M. Kübler* Simon Buchholz* Bernhard Schölkopf
* equal contribution
Max Planck Institute for Intelligent Systems, Tübingen, Germany
{jmkuebler, sbuchholz, bs}@tue.mpg.de

Abstract

We analyze the spectral properties of quantum kernels and show how they determine the potential for a quantum advantage in supervised machine learning. We show a no-free lunch theorem and furthermore that kernel evaluations might block the road to quantum advantages, similarly to Barren Plateaus in QNNs.

1 Extended Abstract

In recent years, much attention has been dedicated to studies of how small and noisy quantum devices [1] could be used for near term applications to showcase the power of quantum computers. Besides fundamental demonstrations [2], potential applications that have been discussed are in quantum chemistry [3], discrete optimization [4] and machine learning (ML) [5–12].

Initiated by the seminal HHL algorithm [13], early work in quantum machine learning (QML) was focused on speeding up linear algebra subroutines, commonly used in ML, offering the perspective of a runtime logarithmic in the problem size [14–17].

Rather than speeding up linear algebra subroutines, we focus on more recent suggestions that use a quantum device to define and implement the function class and do the optimization on a classical computer. There are two ways to that: the first are so-called *Quantum Neural Networks* (QNN) or parametrized quantum circuits [5–7] which can be trained via gradient based optimization [5, 18–22]. The second approach is to use a predefined way of encoding the data in the quantum system and defining a *quantum kernel* as the inner product of two quantum states [7–11]. We here focus on quantum kernels, which allow for convex problems and thus lend themselves more readily to theoretical analysis.

Only recently first steps where taken to theoretically understand when quantum kernels are advantageous [10, 12, 23]. In this work, we relate the discussion of quantum advantages to the classical concept of *inductive bias*. The *no free lunch* theorem informally states that no learning algorithm can outperform other algorithms on all problems. This

*A preprint is available: arXiv:2106.03747 and the work is accepted to NeurIPS 2021.

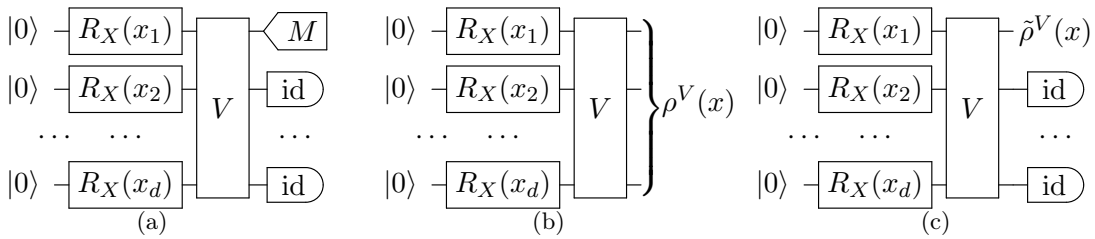


Figure 1: **Quantum advantage via inductive bias:** (a) Data generating quantum circuit $f(x) = \mathrm{Tr}\left[\rho^V(x)(M \otimes \mathrm{id})\right] = \mathrm{Tr}\left[\tilde{\rho}^V(x)M\right]$. (b) The full quantum kernel $k(x, x') = \mathrm{Tr}\left[\rho^V(x)\rho^V(x')\right]$ is too general and cannot learn f efficiently. (c) The biased quantum kernel $q(x, x') = \mathrm{Tr}\left[\tilde{\rho}^V(x)\tilde{\rho}^V(x')\right]$ meaningfully constrains the function space and allows to learn f with little data. However, estimating q to sufficient precision generally requires exponentially many measurements. This blocks the road to an easy quantum advantage.

implies that an algorithm that performs well on one type of problem necessarily performs poorly on other problems. A standard inductive bias in ML is to prefer functions that are continuous. An algorithm with that bias, however, will then struggle to learn functions that are discontinuous. For a QML model to have an edge over classical ML models, we could thus ensure that it is equipped with an inductive bias that cannot be encoded (efficiently) with a classical machine. If a given dataset fits this inductive bias, the QML model will outperform any classical algorithm. For kernel methods, the qualitative concept of inductive bias can be formalized by analyzing the spectrum of the kernel and relating it to the target function [24–29].

Our main contribution is the analysis of the inductive bias of quantum machine learning models based on the spectral properties of quantum kernels. First, we show that quantum kernel methods will fail to generalize as soon as the data embedding into the quantum Hilbert space is too expressive. Then we note that projecting the quantum kernel appropriately allows to construct inductive biases that are hard to create classically (Figure 1). However, our second Theorem also implies that estimating the biased kernel requires exponential measurements, a phenomenon reminiscent of the Barren plateaus observed in quantum neural networks. Finally we show experiments supporting our main claims.

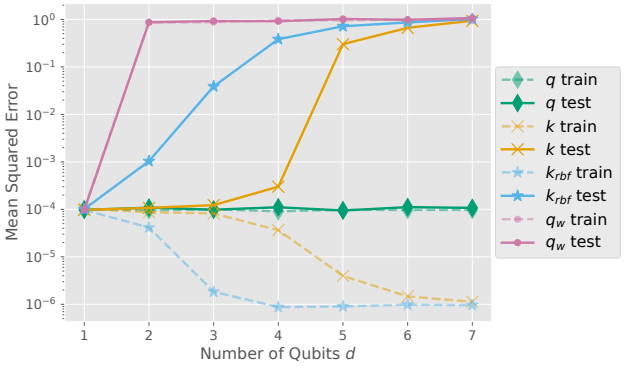


Figure 2: The biased kernel q, equipped with prior knowledge, easily learns the function for arbitrary number of qubits and achieves optimal mean squared error (MSE). Models that are ignorant to the structure of f fail to learn the function.

References

[1] John Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, 2018.

[2] Frank Arute, Kunal Arya, Ryan Babbush, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[3] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5:4213, 2014.

[4] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv:1411.4028*, 2014.

[5] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *Phys. Rev. A*, 98:032309, 2018.

[6] Edward Farhi and Hartmut Neven. Classification with quantum neural networks on near term processors. *arXiv:1802.06002*, 2018.

[7] 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(7747):209, 2019.

[8] Maria Schuld and Nathan Killoran. Quantum machine learning in feature Hilbert spaces. *Phys. Rev. Lett.*, 122:040504, 2019.

[9] Jonas M. Kübler, Krikamol Muandet, and Bernhard Schölkopf. Quantum mean embedding of probability distributions. *Phys. Rev. Research*, 1:033159, 2019.

[10] Yunchao Liu, Srinivasan Arunachalam, and Kristan Temme. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics*, 17(9):1013–1017, 2021.

[11] Maria Schuld. Quantum machine learning models are kernel methods. *arXiv:2101.11020*, 2021.

[12] 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):2631, 2021.

[13] Aram W Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.*, 103(15):150502, 2009.

[14] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. Quantum support vector machine for big data classification. *Phys. Rev. Lett.*, 113(13), 2014.

[15] Nathan Wiebe, Daniel Braun, and Seth Lloyd. Quantum algorithm for data fitting. *Phys. Rev. Lett.*, 109:050505, 2012.

[16] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum algorithms for supervised and unsupervised machine learning. *arXiv:1307.0411*, 2013.

[17] Iordanis Kerenidis and Anupam Prakash. Quantum gradient descent for linear systems and least squares. *Phys. Rev. A*, 101:022316, 2020.

[18] Gian Giacomo Guerreschi and Mikhail Smelyanskiy. Practical optimization for hybrid quantum-classical algorithms. *arXiv:1701.01450*, 2017.

[19] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99(3): 032331, 2019.

[20] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. Quantum natural gradient. *Quantum*, 4:269, 2020.

[21] Ryan Sweke, Frederik Wilde, Johannes Jakob Meyer, Maria Schuld, Paul K Fährmann, Barthélémy Meynard-Piganeau, and Jens Eisert. Stochastic gradient descent for hybrid quantum-classical optimization. *Quantum*, 4:314, 2020.

[22] Jonas M. Kübler, Andrew Arrasmith, Lukasz Cincio, and Patrick J. Coles. An Adaptive Optimizer for Measurement-Frugal Variational Algorithms. *Quantum*, 4: 263, 2020.

[23] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Phys. Rev. Lett.*, 126:190505, 2021.

[24] R. C. Williamson, A. J. Smola, and B. Schölkopf. Generalization performance of regularization networks and support vector machines via entropy numbers of compact operators. *IEEE Transactions on Information Theory*, 47(6):2516–2532, 2001.

[25] B. Schölkopf and A. J. Smola. *Learning with Kernels*. MIT Press, Cambridge, MA, USA, 2002.

[26] Nello Cristianini, Jaz Kandola, Andre Elisseeff, and John Shawe-Taylor. On kernel target alignment. In Dawn E. Holmes and Lakhmi C. Jain, editors, *Innovations in Machine Learning: Theory and Applications*, pages 205–256. Springer Berlin Heidelberg, 2006.

[27] Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh. Algorithms for learning kernels based on centered alignment. *Journal of Machine Learning Research*, 13(28): 795–828, 2012.

[28] Abdulkadir Canatar, Blake Bordelon, and Cengiz Pehlevan. Spectral bias and task-model alignment explain generalization in kernel regression and infinitely wide neural networks. *Nature Communications*, 12(1):2914, 2021.

[29] Arthur Jacot, Berfin Simsek, Francesco Spadaro, Clement Hongler, and Franck Gabriel. Kernel alignment risk estimator: Risk prediction from training data. In *NeurIPS*, 2020.

Universal Approximation Property via Quantum Feature Maps

Quoc Hoan Tran,[1,2,*] Takahiro Goto,[2,*] and Kohei Nakajima[1,2]

[1] *Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan*
[2] *Reservoir Computing Seminar Group, Nagase Hongo Building F8, 5-24-5, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

The quantum Hilbert space can be used as a quantum-enhanced feature space in machine learning (ML) via the quantum feature map to encode classical data into quantum states. We prove the ability to approximate any continuous function with optimal approximation rate via quantum ML models in typical quantum feature maps.

The expectation in utilizing advantages of quantum effects to surpass the classical ML techniques has led to the advent of quantum machine learning (QML) research [1]. QML is currently benefiting from the arrival of noisy intermediate-scale quantum (NISQ) devices that may include a few tens to hundreds of qubits with no error correction capability [2, 3]. One of motivations for QML is that quantum systems can explore a larger class of features than can classical systems. Hence, the quantum Hilbert space can be used as a quantum-enhanced feature space for classical data, which can be efficiently manipulated from NISQ devices. Here, the input data is encoded in a quantum state via a quantum feature map, a nonlinear feature map that maps data to the quantum Hilbert space (Fig. 1). A quantum computer can analyse the input data in this feature space, where a classifier, such as a linear support vector machine (SVM), can gain power in finding a hyperplane to separate the data. The quantum feature map is first proposed and implemented as a fixed quantum circuit, followed by a variational circuit that adapts the measurement basis with trainable parameters [4, 5]. Quantum feature maps underscore the QML advantage; there may be a provable exponential speed-up due to the classical intractability of generating correlations for a particular learning problem [6]. Still, little is known about the relation between the classical intractability of quantum feature maps and the generalization learning performance.

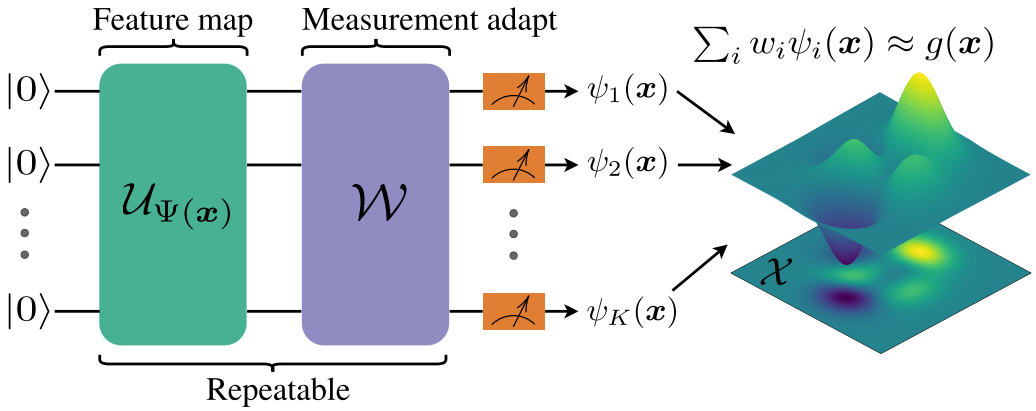


FIG. 1. A quantum feature framework consists of a feature map circuit $\mathcal{U}_{\Psi(\boldsymbol{x})}$ that realizes $\Psi(\boldsymbol{x})$ to map the classical data $\boldsymbol{x} \in \mathcal{X}$ to a quantum state in the Hilbert space and a quantum circuit $\mathcal{W}$ to adapt the measurement basis. The combination of $\mathcal{U}_{\Psi(\boldsymbol{x})}$ and $\mathcal{W}$ can be repeated as a sequence with different parameters. This framework has the universal approximation property if the linear combining of measurement results can approximate any continuous function $g : \mathcal{X} \to \mathbb{R}$.

[*] These authors contributed equally to this work.

We focus to answer whether a QML model based on a quantum feature map can obtain expressivity that is as powerful as, or is more powerful than, classical ML schemes. This can be investigated from the perspective of the universal approximation property (UAP) and the classification capability, which have been extensively explored in feedforward classical neural networks [7, 8]. Here, UAP refers to the ability to approximate any continuous real function on a dataset $\mathcal{X}$. The classification capability implies that the function constructed from quantum feature maps can form disjoint decision regions [9]. Quantum neural networks, which employ qubits as quantum perceptrons with nonlinear excitation responses [10], can be emulated on a photonic quantum computer to obtain UAP [11]. In Ref. [12], the expressivity of a quantum model with a variational circuit is characterized in terms of a partial Fourier series in the data. However, the study of UAP and classification capability of QML models with quantum feature maps still remains challenging.

In this research, we present a provable UAP and classification capability in two typical scenarios when setting the quantum feature map. Here, we combine quantum feature maps with an appropriate possible set of K observables $O_1, O_2, \ldots, O_K$, which are Hermitian operators applied to the output state $|\Psi(\boldsymbol{x})\rangle$ of the quantum circuit. If we measure O_i, we can obtain the expectation value of this observable and consider it as the basis function $\psi_i(\boldsymbol{x}) = \langle\Psi(\boldsymbol{x})|O_i|\Psi(\boldsymbol{x})\rangle = \mathrm{Tr}[O_i\,|\Psi(\boldsymbol{x})\rangle\,\langle\Psi(\boldsymbol{x})|]$. We focus on the UAP of functions $f : \mathcal{X} \to \mathbb{R}$, where each f is the linear combination of the basis functions ψ_i (Fig. 1). In the first scenario, which is defined as the *parallel scenario*, the quantum feature map is a tensor product of multiple quantum circuits; each circuit acts on a subsystem, and the number of qubits can be set freely. In the second scenario, which is defined as the *sequential scenario*, the quantum feature map is the repetition of a simple fixed quantum circuit, and the number of qubits is fixed. We obtain the UAP in the first scenario and prove the UAP for the second in single-qubit circuits of the finite input space.

We further describe relative goodness or badness in a universal approximation via the approximation rate. This rate refers to the speed at which the approximation error decreases when the parameters, such as the number of qubits N and the input dimension d, are increased. The approximation rate strongly depends on the nature of the target function $g : \mathcal{X} \to \mathbb{R}$ to be approximated and the type of the input set $\mathcal{X}$. If $\mathcal{X} = [0,1]^d$ and the target function g is Lipschitz continuous with respect to the Euclidean norm, we can construct an explicit form of the approximator to g in the parallel scenario by N qubits with the error $\varepsilon = O(d^{7/6}N^{-1/3})$. Furthermore, we can achieve an approximation error with a better approximation rate in terms of N as $\varepsilon = O(d^{3/2}N^{-1})$. It implies that $O(d^{3/2}\varepsilon^{-1})$ qubits are enough to obtain an approximation with ε-error. More detailed proofs for our results can be found in Ref. [13].

The approximation rate provides a method to compare the asymptotic universality between our quantum feature framework and the classical neural networks. The number of observables K corresponds with the number of parameters in the classical neural networks. In the parallel scenario, we can write our best approximation error as $\varepsilon = O(K^{-1/d})$ if we fix d and focus on K. Interestingly, this is also the best approximation when using a classical neural network to approximate a Lipschitz continuous function [14, 15]. This result suggests a strong guarantee that the QML models in quantum-enhanced feature spaces can exhibit at least the same expressivity as the classical ML models. Furthermore, our work enables an important theoretical analysis to ensure that ML algorithms based on quantum feature maps can handle a broad class of ML tasks.

K.N. and Q.H.T. were supported by MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) Grant Nos. JPMXS0118067394 and JPMXS0120319794.

[1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature **549**, 195 (2017).
[2] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).
[3] G. Torlai and R. G. Melko, Machine-learning quantum states in the NISQ era, Annu. Rev. Condens. Matter Phys. **11**, 325 (2020).
[4] 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).
[5] M. Schuld and N. Killoran, Quantum machine learning in feature Hilbert spaces, Phys. Rev. Lett. **122**, 040504 (2019).
[6] Y. Liu, S. Arunachalam, and K. Temme, A rigorous and robust quantum speed-up in supervised machine learning, Nat. Phys. (2021).
[7] G.-B. Huang, Y.-Q. Chen, and H. Babri, Classification ability of single hidden layer feedforward neural networks, IEEE Trans. Neural Netw. **11**, 799 (2000).
[8] G.-B. Huang, L. Chen, C. K. Siew, *et al.*, Universal approximation using incremental constructive feedforward networks with random hidden nodes, IEEE Trans. Neural Netw. **17**, 879 (2006).
[9] W. Y. Huang and R. P. Lippmann, Neural net and traditional classifiers, in *Neural Information Processing Systems*, edited by D. Z. Anderson (American Institute of Physics, New York, 1988) pp. 387–396.
[10] E. Torrontegui and J. J. García-Ripoll, Unitary quantum perceptron as efficient universal approximator, EPL (Europhysics Letters) **125**, 30004 (2019).
[11] N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd, Continuous-variable quantum neural networks, Phys. Rev. Research **1**, 033063 (2019).
[12] 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).
[13] T. Goto, Q. H. Tran, and K. Nakajima, Universal approximation property of quantum machine learning models in quantum-enhanced feature spaces, Phys. Rev. Lett. **127**, 090506 (2021).
[14] H. N. Mhaskar, Neural networks for optimal approximation of smooth and analytic functions, Neural Computation **8**, 164 (1996).
[15] D. Yarotsky, Error bounds for approximations with deep ReLU networks, Neural Networks **94**, 103 (2017).