

# A Combinatorial Search of Parameterized Quantum Circuit Learning for Chemical Applications

Grier M. Jones,<sup>†,‡</sup> Nick Taylor,<sup>†</sup> Viki Kumar Prasad,<sup>†,‡</sup> Ulrich Fekl,<sup>\*,‡</sup> and  
Hans-Arno Jacobsen<sup>\*,†</sup>

<sup>†</sup> *The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 Kings College Road, Toronto, Ontario, Canada M5S 3G4*

<sup>‡</sup> *Department of Chemical and Physical Sciences, University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, Ontario, Canada L5L 1C6*

E-mail: ulrich.fekl@utoronto.ca; jacobsen@eecg.toronto.edu

## Abstract

Within the quantum machine learning (QML) field, parameterized quantum circuits (PQCs), built using fixed and parameterized gates, offer a hybrid approach for complex machine learning tasks. While many potential use cases have been proposed, the exploration of relevant datasets for chemists is lacking. Our study seeks to understand the possible advantages and disadvantages of PQCs for two chemically relevant datasets: one based on the bond separation energies of 49 different classes of bonds, called the BSE49 dataset, and another consisting of water confirmations, where coupled-cluster singles and doubles (CCSD) wave functions are predicted using electronic structure theory data from lower-level methods using the data-driven coupled-cluster (DDCC) method. In our study, we examine a combinatorial space of 14 data encoding layers and 12 variational (ansatz) layers,

for a combined total of 168 PQCs. To calibrate our PQCs, we utilize a dataset of noisy linear, quadratic, and sine functions to explore the effects of the circuit width and depth, the effects of the feature set size, and various error mitigation techniques. Following this step, we similarly examine our chemically relevant datasets. Our work highlights the difficulties in encoding classical molecular representations in a PQC for predicting bond separation energies and the aptitude for PQCs for predicting molecular wave functions.

- Abstract
- Introduction
- Methods
- Datasets
- Results and Discussion
  - Function fitting
    - \* All (5 + 16 qubit)
    - \* RUD + AL (5 + 16 qubit)
    - \* Error mitigation on FakeQuebec
  - BSE
    - \* All 5 qubit
    - \* Truncated 16 qubit set (cost analysis 5 qubit)
    - \* RUD + AL (5 + 16 qubit)
    - \* Real device (without error mitigation and with whatever the best is from function fitting)
  - DDCC (5 qubit only because classically it can be done with 5 features!)
    - \* Truncated 5 qubit set (based on BSE 5 qubit cost analysis)
    - \* RUD + AL (5 qubit)
    - \* Real device (without error mitigation and with whatever the best is from function fitting)
- Conclusion

# Introduction

[quantum machine learning]

[parameterized circuits for chemical applications]

[What we introduce]

Our work includes the single and double encoding layers found in the paper of Suzuki and Katouda,<sup>1</sup> including the encoding of Mitarai *et al.*,<sup>2</sup> along with using instantaneous quantum polynomial (IQP) as an encoding layer.<sup>3</sup> All of our variational (ansatz) layers can be found in the study of Sim *et al.*,<sup>4</sup> where the expressibility and entanglement capability of these circuits is analyzed.

5

QML:<sup>6</sup>

- “Quantum machine learning software makes use of quantum algorithms as part of a larger implementation. By analysing the steps that quantum algorithms prescribe, it becomes clear that they have the potential to outperform classical algorithms for specific problems (that is, reduce the number of steps required). This potential is known as quantum speedup.”
- “The notion of a quantum speedup depends on whether one takes a formal computer science perspective—which demands mathematical proofs—or a perspective based on what can be done with realistic, finitesize devices—which requires solid statistical evidence of a scaling advantage over some finite range of problem sizes. For the case of quantum machine learning, the best possible performance of classical algorithms is not always known.”
- DO WE SEE THIS: “Determination of a scaling advantage contrasting quantum and classical machine learning would rely on the existence of a quantum computer and is called a ‘benchmarking’ problem. Such advantages could include improved classification accuracy and sampling of classically inaccessible systems. Accordingly, quantum speedups in machine learning are currently characterized using idealized measures from complexity theory: query complexity and gate complexity (see Box 1 and Box 1 Table). Query complexity measures the number of queries to the information source for the classical or quantum algorithm.

A quantum speedup results if the number of queries needed to solve a problem is lower for the quantum algorithm than for the classical algorithm. To determine the gate complexity, the number of elementary quantum operations (or gates) required to obtain the desired result are counted.”

- Input problem: “Classical data must be input before being processed on a quantum computer. This ‘input problem’ often has little overhead but can present a serious bottleneck for certain algorithms. Likewise, the ‘output problem’ is faced when reading out data after being processed on a quantum device. Like the input problem, the output problem often causes a noticeable operational slowdown.”
- QML challenges: “These hardware challenges are technical in nature, and clear paths exist towards overcoming them. They must be overcome, however, if quantum machine learning is to become a ‘killer app’ for quantum computers. As noted previously, most of the quantum algorithms that have been identified face a number of caveats that limits their applicability. We can distill the caveats mentioned above into four fundamental problems. (1) The input problem. Although quantum algorithms can provide dramatic speedups for processing data, they seldom provide advantages in reading data. This means that the cost of reading in the input can in some cases dominate the cost of quantum algorithms. Understanding this factor is an ongoing challenge. (2) The output problem. Obtaining the full solution from some quantum algorithms as a string of bits requires learning an exponential number of bits. This makes some applications of quantum machine learning algorithms infeasible. This problem can potentially be sidestepped by learning only summary statistics for the solution state. (3) The costing problem. Closely related to the input/output problems, at present very little is known about the true number of gates required by quantum machine learning algorithms. Bounds on the complexity suggest that for sufficiently large problems they will offer huge advantages, but it is still unclear when that crossover point occurs. (4) The benchmarking problem. It is often difficult to assert that a quantum algorithm is ever better than all known

classical machine algorithms in practice because this would require extensive benchmarking against modern heuristic methods. Establishing lower bounds for quantum machine learning would partially address this issue.”

PQCs:<sup>7</sup>

- “Parameterized quantum circuits (PQCs) offer a concrete way to implement algorithms and demonstrate quantum supremacy in the NISQ era. PQCs are typically composed of fixed gates, e.g. controlled NOTs, and adjustable gates, e.g. qubit rotations. Even at low circuit depth, some classes of PQCs are capable of generating highly non-trivial outputs. For example, under well-believed complexity-theoretic assumptions, the class of PQCs called instantaneous quantum polynomial-time cannot be efficiently simulated by classical resources (see Lund et al [3] and Harrow and Montanaro [4] for accessible Reviews of quantum supremacy proposals). The demonstration of quantum supremacy is an important milestone in the development of quantum computers. In practice, however, it is highly desirable to demonstrate a quantum advantage on applications. The main approach taken by the community consists in formalizing problems of interest as variational optimization problems and use hybrid systems of quantum and classical hardware to find approximate solutions. The intuition is that by implementing some subroutines on classical hardware, the requirement of quantum resources is significantly reduced, particularly the number of qubits, circuit depth, and coherence time. Therefore, in the hybrid algorithmic approach NISQ hardware focuses entirely on the classically intractable part of the problem.”

PQCs:<sup>1</sup>

- “More recently, quantum machine learning (QML) [26–30] is a rapidly growing research field that combines near-term quantum algorithms and machine learning techniques. In particular, parameterized quantum circuits (PQCs) have been considered as machine learning models with high expressive power within the hybrid quantum–classical framework [31, 32]. PQCs are typically composed of fixed quantum gates (e.g., qubit rotations and entangling

gates) in a shallow circuit layout, with variable parameters optimized in a classical feedback loop.”

- “To our knowledge, however, the application of QML to regression tasks has not been fully investigated in the literature. It remains unclear what kinds of quantum states should be used in order to generate the feature map with high expressibility that is suited for real-world data sets.”
- “quantitative structure–toxicity relationship (QSTR) models for predicting the toxicity of 221 phenols. While there are a variety of QSAR/QSTR models (e.g., 3D-QSAR [1, 4]), as a first step we employ QSAR/QSTR models including molecular descriptors such as hydrophobicity, acidity constant, and frontier orbital energies. There have been quantum computations in biochemical and pharmaceutical areas, such as protein folding [43–45], molecular similarity [46], and biological data [47]; yet, there has been no study on quantum application to QSAR modeling, albeit an important part of ligand-based computer-aided drug design.”
- “PQCs have been regarded as machine learning models with high expressive power within the framework of the hybrid quantum–classical approach. PQCs are usually composed of one-qubit rotations and two-qubit entangling operations in a shallow circuit layout, with parameters optimized in a feedback loop.”
- “Combining near-term quantum algorithms and machine learning, QML using the framework of PQCs is sometimes referred to as quantum circuit learning (QCL)”
- “PQCs consist of three components: the encoder circuit, the variational circuit, and the measurement for the estimation of the loss function”

PQCs:<sup>8</sup>

- What they do: “Introduce quantum circuit learning (QCL) as an emerging regression algorithm for chemo- and materials-informatics. The supervised model, functioning on the rule of quantum mechanics, can process linear and smooth non-linear functions from small

datasets (<100 records). Compared with conventional algorithms, such as random forest, support vector machine, and linear regressions, the QCL can offer better predictions with some one-dimensional functions and experimental chemical databases”

- “Here, we introduce quantum circuit learning (QCL), an emerging algorithm for supervised learning.<sup>15–18</sup> QCL works on the rule of quantum mechanics. It can predict various parameters likewise to classical models. The current quantum systems (noisy intermediate-scale quantum computers: NISQ)<sup>19,20</sup> face the problems of calculation noise and the limited processable number of information units (qubits).”
- “Since QCL is a frontier for machine learning, few reports have been published on authentic regression tasks. The success of prediction with a toxicity dataset of organic molecules was reported,<sup>17</sup> whereas the study was still conceptual. The prediction processes and advantages have been unclear, especially from chemical and material viewpoints. Here, we conducted a more comprehensive study on QCL with standard datasets of one-dimensional functions and chemical properties. Both simulated and actual quantum computing was undertaken to clarify the challenges of QCL. Various hyperparameters were optimized for higher accuracy. The comparison with conventional models contrasted the benefits of the new approach: capable of learning both linear and non-linear functions even from small datasets. The property was also favorable for predicting the so-called extrapolating data region, which is essential for chemical and material research.”
- They cite<sup>1,2,9</sup> as examples of PQCs/quantum circuit learning for regression task
- Justification for starting with 1D regression: “Preliminary optimization of quantum circuits<sup>15,17</sup> and our optimization revealed that the following configuration was concise and practical for regression tasks (for details, see results and explanations in Fig. S1”



# Methods

Use a mix of<sup>8</sup> and<sup>1</sup> for the methods.

Format from:<sup>1</sup>

- Feature set  $\mathbf{x} = (x_1, x_2, \dots, x_d)^T \in \mathbb{R}^d$
- Feature map (encoding)  $U_{\Phi(\mathbf{x})}$  produces the quantum state  $U_{\Phi(\mathbf{x})} |0\rangle^{\otimes n}$
- Variational circuit (ansatz)  $U(\theta)$ , leads to the parameterized quantum circuit  $U(\theta)U_{\Phi(\mathbf{x})} |0\rangle^{\otimes n}$

Quantum machine learning is a hybrid classical-quantum approach utilizing parameterized quantum circuits. Typical hybrid algorithms consist of shallow circuit depth of one qubit rotations and two qubit entangling gates. In the context of machine learning, we can subdivide our circuit into three components: the encoder, the variational circuit, and the act of measuring to estimate predictions. The encoder takes our  $n$ -dimensional data  $\mathbf{x} \in \mathbb{R}^n$  and encodes it into a unitary feature map  $U(\mathbf{x})$  in our working Hilbert Space. This unitary acts on our initialized state  $|0\rangle^{\otimes n'}$  to produce the state  $U(\mathbf{x})|0\rangle^{\otimes n'}$ . It is usual to take  $n' = n$  and for each feature  $x_i$  to correspond to a qubit. With  $n < n'$ , features can be repeated across the qubits present in the system. Further elaboration will be discussed with specific encoding circuits. After the encoder produces the state  $U(\mathbf{x})|0\rangle^{\otimes n}$ , we then tune the state with a parameterized unitary  $U(\theta)$  to transform the state into a meaningful prediction. Finally, some configuration of qubits are measured to produce expectation values  $\hat{y}_k$  for our data labels  $y_i$  which we then use to minimize a loss function. An obstacle in creating quantum models is there is no way to determine the best encoding or variational circuit for the problem. As such, this paper explores many different encoding and variational circuits.

## Encoder Circuit

Representation of the data in any machine learning context is a crucial factor in its success. In a classical sense, this can be thought of as generating features and selecting the appropriate features to input into your model. In the quantum context, however, there is the additional question of how

to translate the data into the Hilbert space. Different proposed methods include angle encoding, and amplitude encoding. Following the work done by Suzuki and Katouda, we explore three base encoders composed together with entangling operations. Additionally, we explore the IQP encoding as given in PennyLane.

We composed circuits out of primary rotation blocks and entangling blocks. These rotation blocks include the one proposed by Mitrai et al.,<sup>2</sup> a single angle rotation, and a double rotation encoder:

$$U_{A1} = \prod_{i=1}^n R_i^Y(x_i) \quad (1)$$

$$U_{A2} = \prod_{i=1}^n R_i^Y(x_i) R_i^Z(x_i) \quad (2)$$

$$U_M = \prod_{i=1}^n R_i^Z(\arccos x_i^2) R_i^Y(\arcsin x_i^2) \quad (3)$$

Where  $R_i^j$  is a rotation about the  $j$  axis on qubit  $i$ . It's clear to see that this method of encoding imposes a restriction on the number of features to the number of qubits. With this encoding method, it is possible to use more features than qubits and to re-encode the features along the remaining qubits. Then, to compose these encoders together we use entangling circuits that entangle adjacent qubits using either CNOT or CZ entangling operations. These encoding circuits can be seen presented in Figure ??.

## Variational Circuit

After encoding the data, we have a quantum circuit in a state  $U_{\text{enc}}(\mathbf{x})|0\rangle^{\otimes n}$ . We need to act on this state with a unitary that has parameters which can be tuned in our machine learning model. Choosing this circuit ansatz can greatly dictate the performance of a given model. Model performance can be greatly influenced by the variational circuit chosen as well as the depth of the circuit as we will later show. It is not obvious which circuit we should pick for our given problem, especially given that our algorithm is not physically motivated like other popular parameterized

quantum circuit algorithms. We have gathered a list of 12 variational circuit blocks from previous works and libraries which we will then benchmark for our particular problem. These circuits each can be decomposed into individual rotation gates and two-qubit entangling gates.

After our variational circuit acts on our state we have the state:

$$|\psi\rangle = U_{\text{var}}(\boldsymbol{\theta})U_{\text{enc}}(\mathbf{x})|0\rangle^{\otimes n} \quad (4)$$

where  $\boldsymbol{\theta}$  is our parameter vector to be tuned in learning. We can further modify this state; by repeating variational circuit blocks we create a more rich and complex model with a better ability to explore the Hilbert space. This would give us the state:

$$|\psi\rangle = U_{\text{var}}(\boldsymbol{\phi})U_{\text{var}}(\boldsymbol{\theta})U_{\text{enc}}(\mathbf{x})|0\rangle^{\otimes n} \quad (5)$$

However, as we are interested in NISQ applications, it should be noted that increasing layer depths creates decoherence in quantum devices. Additionally, we can use the idea of data re-uploading to create redundancies in the encoding of the model to explore a richer feature space.

$$|\psi\rangle = U_{\text{var}}(\boldsymbol{\phi})U_{\text{enc}}(\mathbf{x})U_{\text{var}}(\boldsymbol{\theta})U_{\text{enc}}(\mathbf{x})|0\rangle^{\otimes n} \quad (6)$$

This paper uses and compares both the data re-uploading and increased depth approaches of creating more complicated models.

## Measurement

After our system is in state  $|\psi\rangle$  as given by either (4), (5) or (6) we can measure our system to generate a predicted value  $\hat{y}_i$ . At this point in the algorithm, any loss function can be used. In our simplest system, we measured the first qubit so that,

$$\hat{y}_i = \langle Z_0 \rangle_{\psi} \quad (7)$$

for a  $\psi$  generated by the  $k$ th data point. Then, we have the loss function we are trying to minimize as,

$$\mathcal{L}(\hat{\mathbf{y}}) = \sum_i^N \frac{(y_i - \hat{y}_i)^2}{N} \quad (8)$$

## Implementation

Our work was implemented in Python using PennyLane for constructing the quantum circuits, and Qulacs as a primary backend for simulating the circuits. The optimization was implemented using the `scipy.optimize` module from the SciPy library. Classical models were implemented using `scikit-learn` while processing of the data was handled with `Pandas`, `NumPy`, and `SciPy` libraries.

Simulated using FakeQuebecbackend

## Dataset

We trained our models using three datasets: a function fitting dataset, consisting of a noisy linear, quadratic, and sine function, used for model calibration; a dataset consisting of electronic structure features to predict wavefunctions using the data-driven coupled-cluster scheme of Townsend and Vogiatzis;<sup>10</sup> and a dataset of bond separation energies (BSE) of molecules, where the feature set encodes structural information of each molecule.

For the function fitting dataset, there is only one feature per target value. For the DDCC dataset, there are initially  $X$  features, reduced down to 5 and 16 features, respectively, using SHapley Additive exPlanation (SHAP) values.<sup>11</sup>

$$t_{ij(\text{MP2})}^{ab} = \frac{\langle ij || ab \rangle}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \quad (9)$$

For each  $t_{ij(\text{CCSD})}^{ab}$ , orbital energies ( $\epsilon_i, \epsilon_j, \epsilon_a, \epsilon_b$ ), Coulomb and exchange integrals ( $J_a^i, J_b^j, K_i^a, K_j^b$ ), binary feature whether two electrons are promoted to the same virtual orbital, the initial MP2 amplitudes, along with the numerator (two-electron integrals ( $\langle ij || ab \rangle$ )) and denominator  $\epsilon_i + \epsilon_j -$

$$\varepsilon_a - \varepsilon_b.$$

For the BSE dataset, various molecular representations are explored, and we settled on using Morgan fingerprints with X parameters. For the feature reduction, we use PCA, as explained in Section ??.

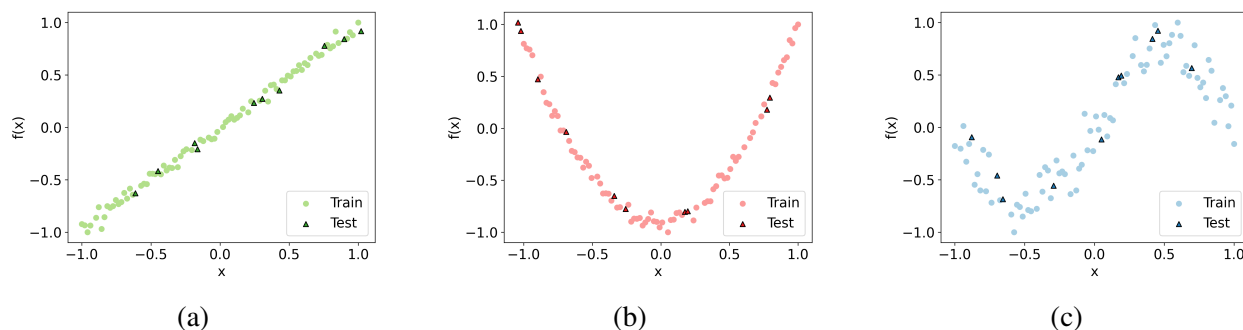


Figure 1

## Results

Ansaetze analysis<sup>4</sup> “In particular, a substantial improvement in performance of two-qubit gates in a ring or all-to-all connected arrangement, compared to that of those on a line, is observed.”

“Furthermore, improvement in both descriptors is achieved by sequences of controlled X-rotation gates compared to sequences of controlled Z-rotation gates.”

“investigated how expressibility “saturates” with increased circuit depth, finding that the rate and saturated value appear to be distinguishing features of a PQC”

## Conclusion

Depth is not always better! Molecular representations specifically for QML Distributed QC to incorporate more features

## References

- (1) Suzuki, T.; Katouda, M. Predicting toxicity by quantum machine learning. *J. Phys. Commun.* **2020**, *4*, 125012.
- (2) Mitarai, K.; Negoro, M.; Kitagawa, M.; Fujii, K. Quantum circuit learning. *Phys. Rev. A* **2018**, *98*, 032309.
- (3) Bremner, M. J.; Montanaro, A.; Shepherd, D. J. Average-case complexity versus approximate simulation of commuting quantum computations. 2015; <http://arxiv.org/abs/1504.07999>, arXiv:1504.07999.
- (4) Sim, S.; Johnson, P. D.; Aspuru-Guzik, A. Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms. *Advanced Quantum Technologies* **2019**, *2*, 1900070.
- (5) Krenn, M.; Landgraf, J.; Foesel, T.; Marquardt, F. Artificial intelligence and machine learning for quantum technologies. *Phys. Rev. A* **2023**, *107*, 010101.
- (6) Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum machine learning. *Nature* **2017**, *549*, 195–202.
- (7) Benedetti, M.; Lloyd, E.; Sack, S.; Fiorentini, M. Parameterized quantum circuits as machine learning models. *Quantum Sci. Technol.* **2019**, *4*, 043001.
- (8) Hatakeyama-Sato, K.; Igarashi, Y.; Kashikawa, T.; Kimura, K.; Oyaizu, K. Quantum circuit learning as a potential algorithm to predict experimental chemical properties. *Digital Discovery* **2023**, *2*, 165–176.
- (9) Takaki, Y.; Mitarai, K.; Negoro, M.; Fujii, K.; Kitagawa, M. Learning temporal data with a variational quantum recurrent neural network. *Phys. Rev. A* **2021**, *103*, 052414, Publisher: American Physical Society.

- (10) Townsend, J.; Vogiatzis, K. D. Data-Driven Acceleration of the Coupled-Cluster Singles and Doubles Iterative Solver. *J. Phys. Chem. Lett.* **2019**, *10*, 4129–4135, Publisher: American Chemical Society.
- (11) Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*. 2017.