A Combinatorial Search of Parameterized Quantum Circuit Learning for Chemical Applications

Grier M. Jones,[†,‡] Nick Taylor,[†] Viki Kumar Prasad,[†,‡] Ulrich Fekl,[*,‡] and Hans-Arno Jacobsen[*,†]

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

‡*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

1 Dataset

The function fitting dataset consists of a noisy linear (Fig. 1a), quadratic (Fig. 1b), and sine (1c) function. For each function, 90 points are generated, where 80 are used for training set and 10 for the test set. This dataset is used for calibration of the 168 PQCs, along with providing additional insights into the effects of circuit depth, training set size, and error mitigation on model performance.
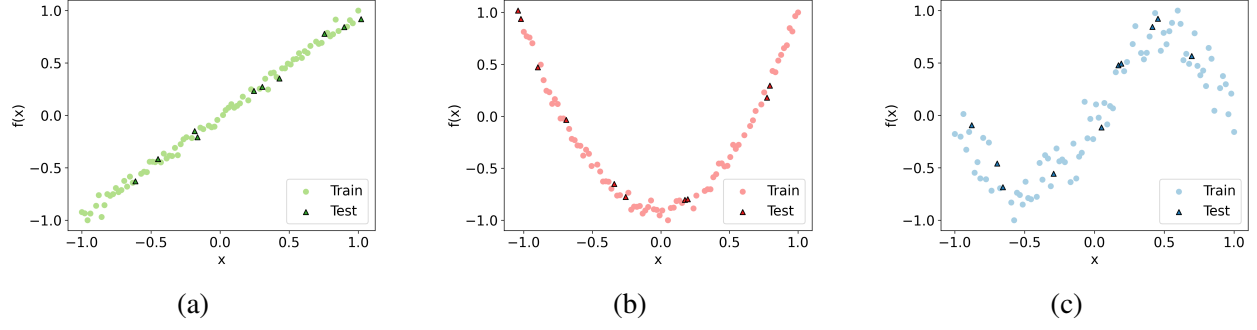
Figure 1: The function fitting dataset which consists of 90 total points, 80 for training and 10 used for testing, for the (a) linear, (b) quadratic, and (c) sine functions.

2 Function Fitting

To calibrate the PQCs, we perform an initial study of the 168 encoder-ansatz pairs using both 5 and 16 qubits. Individual heat maps showing the performance of the encoder-ansatz pairs and box plots showing the overall statistics for each encoder and ansatz are shown for the 5 and 16 qubit models in SI Section S1.

Starting with the five qubit function fitting, we found that out of the 168 encoder-ansatz pairs, for the linear function (SI Fig. S1a) is is A1_ESU2 with a train and test R^2 of 0.9886 and 0.9893, respectively. Box plots for the encoders (left) and variational layers (right) are shown in SI Fig. S1b, where, the best encoder is IQP, with a mean R^2 of 0.8434, and best ansatz layer Full-CRX, with a mean R^2 of 0.7562. For the quadratic function, the best encoder-ansatz pair is A1-A1-CNOT_Full-CRX with a train and test R^2 of 0.8809 and 0.7309, respectively, as shown in SI Fig S1c. On average the best encoder is A1-A1-CNOT with an R^2 of 0.5541 and best variational layer is Full-CRX with an R^2 of 0.5958 (SI Fig. S1d). For the sine function, SI Fig. S1e, M-M-CZ_Full-CRX is the best encoder-ansatz pair with a train and test R^2 of 0.8887 and 0.9081, respectively. As shown in SI Fig. S1f, the best encoder is M with an average R^2 of 0.6456 and best variational layer is Full-CRX with a mean R^2 of 0.7551. Across all three five qubit function fitting datasets, on average, the best encoders are M-A2-CZ with an average R^2 of 0.4776 (Fig. 2a left) and ansatz is Full-CRX with an average R^2 of 0.7023 (Fig. 2a right).

For the sixteen qubit function fitting evaluation of the 168 encoder-ansatz pairs, we found that, similarly to the five qubit data, for the linear function the best pair is A1_ESU2 with train and test R^2s of 0.9886 and 0.9893, respectively, as highlighted in SI Fig. S2a. Like the five qubit data, the best encoder is also IQP with an average R^2 of 0.7800 (SI Fig. S2b left), while the best ansatz varies from the five qubit data, Full-Pauli-CRZ with a mean R^2 of 0.3947 (SI Fig. S2b right). The A1-A1-CZ_Modified-Pauli-CRX encoder-ansatz pair has the best overall model performance, highlighted in SI Fig. S2c, where the training set has an R^2 of 0.8415 and the test set has an R^2 of 0.7722. As demonstrated in SI Fig. S2d, the best encoders and ansätze are A1-A1-CNOT and Full-Pauli-CRX, with mean R^2s of 0.5759 and 0.4474, respectively. For the last sixteen qubit dataset, the sine function data (SI Fig. S2e), the best encoder-ansatz pair is A1-A1-CZ_Full-Pauli-CRZ with an R^2 of 0.8147 for the training set and an R^2 of 0.8626 for the test set. SI Fig. S2f highlights that the best encoder is M with a mean R^2 of 0.5824 and best ansatz, on average, is Full-Pauli-CRZ with an R^2 of 0.4881. Across all three function fitting datasets, the sixteen qubit data varies where the best encoders are IQP with an average R^2 of 0.4506 and ansätze is Full-Pauli-CRZ with an average R^2 of 0.4393.

(a)   (b)

Figure 2: The combined performance of all three function fitting datasets for both the (a) five and (b) sixteen qubit models is evaluated using the R^2 (y-axis) versus the encoders (left) and ansätze (right).

Following the initially analysis of the performance of the 168 PQCs using both five and sixteen qubits, we experimented with the effect of the re-upload depth (RUD) and number of ansatz layers (AL) using the sixteen qubit circuits. We chose to use the sixteen qubit circuits over the five qubit

circuits since these circuits have complexity more similar to when the circuits are used for the chemical dataset in terms of the number of parameters and optimization times. From the initial 168 PQCs, for each function fitting dataset, we choose the five best (top row), five median (middle row), and five worst (bottom row) circuits to see the effects of increasing the RUD and AL on model performance as shown in Figs. 3a, 3b, and 3c. For the linear (Fig. 3a), quadratic (3b), and sine (Fig. 3c) data, some generalities arise for the five best circuits. We see that for these circuits, increasing the number of AL from 1 to 3 or 5 improves model performance more than RUD. For the median performers, the RUD improves model performance more than the ansatz layers. And lastly, for the worse performers a general trend does not exist and increasing the RUD and AL depends on the dataset and circuit architecture.

Despite these trends, we found that for the linear function, the best performing model is IQP_Full-Pauli-CRZ with a RUD and AL of 1. This circuit also has a training and test R^2 of 0.9883 and 0.9891, respectively, and a depth of 95. Unlike the best linear circuit, the best circuits for both the quadratic and sine data, IQP_Full-Pauli-CRX and A2_HWE-CZ, respectively, benefit from an increased RUD, with the best models having a RUD of 5 and AL of 1. These two circuits differ in depth and performance, with the IQP_Full-Pauli-CRX for the quadratic data having a circuit depth of 520, training R^2 of 0.9817, and test R^2 of 0.9609 and the A2_HWE-CZ circuit for the sine data having a depth of 105 , training R^2 of 0.9156, and test R^2 of 0.9576. Following these insights, we applied these circuits for analyzing the effects of the training set size on model performance using learning curves and the effects of error mitigation using noisy simulation.
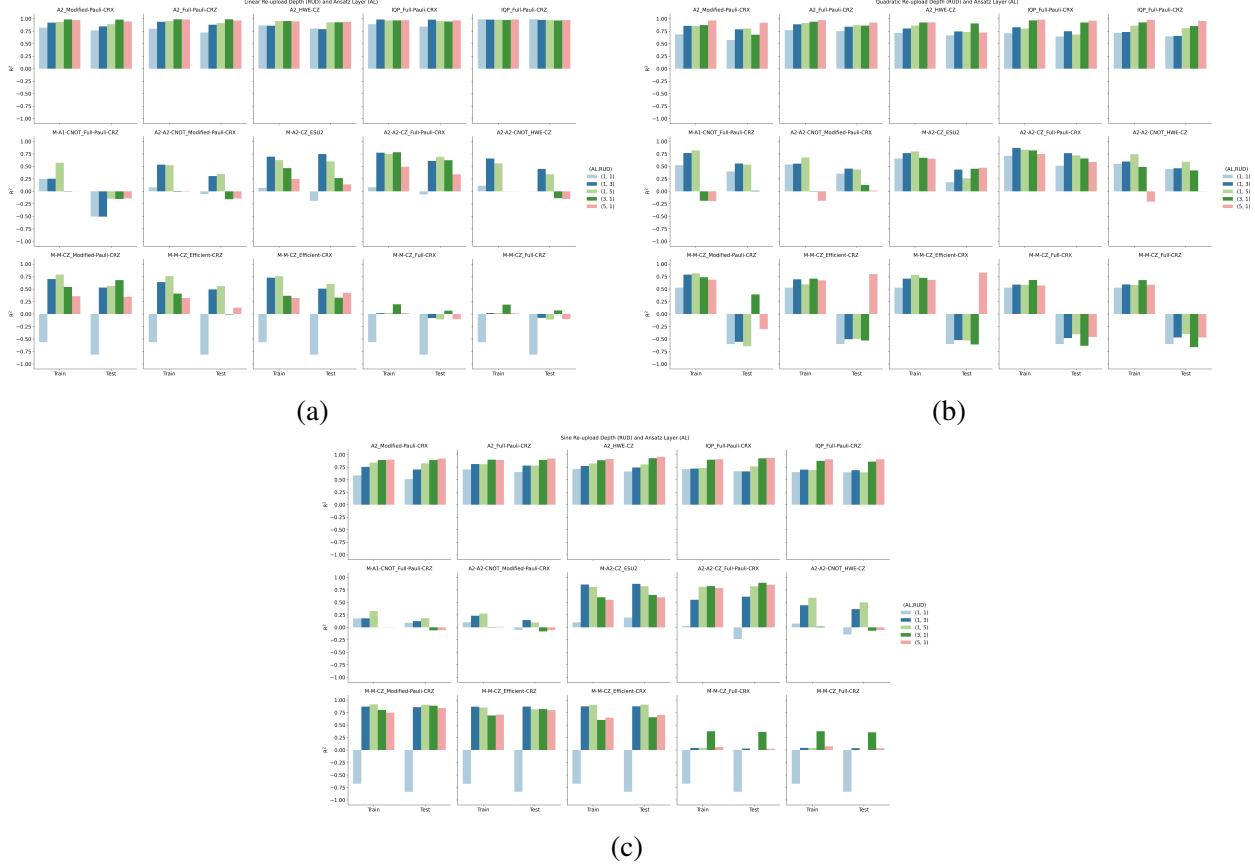
Figure 3: The best five (top row), median five (middle row), and worst five (bottom row) circuits for the 16 qubit (a) linear, (b) quadratic, and (c) sine data show that, in general, an increase in the re-upload depth (RUD) or ansatz layers (AL) increases model performance (R^2 shown on the y-axis).

One proposed advantage of using PQCs for ML tasks is that, in some cases, they can outperform classical ML models using less training data.[1] We examine this phenomena with learning curves using the best performing circuits previously described. The learning curves are generated by varying the percentage of training data from 10-80% of the available data, while holding the number of test points to 10% of the total data. The three PQCs for each data set are compared to the classical models mentioned in subsection **??** in a statistical manner (left sub-figures in Figs. 4a, 4b, and 4c). Across the full learning curve, IQP_Full-Pauli-CRZ (linear function fitting) outperforms seven out of nine classical models for the training set and all of the classical models on the test set. On the quadratic dataset, using IQP_Full-Pauli-CRX, offers similar performance to the classical models, despite being outperformed on the training set across the learning curve. Similar to the

IQP_Full-Pauli-CRZ using the linear data, IQP_Full-Pauli-CRX outperforms seven out of the nine classical models on the test set across the full learning curve. On the sine data, A2_HWE-CZ is consistent across the learning curve, outperforming six of the nine classical models on the training set and all of the models on the test set. For all three datasets, this implies that the PQCs offer similar performance to classical models that do well on these datasets but provide better results on the test set.

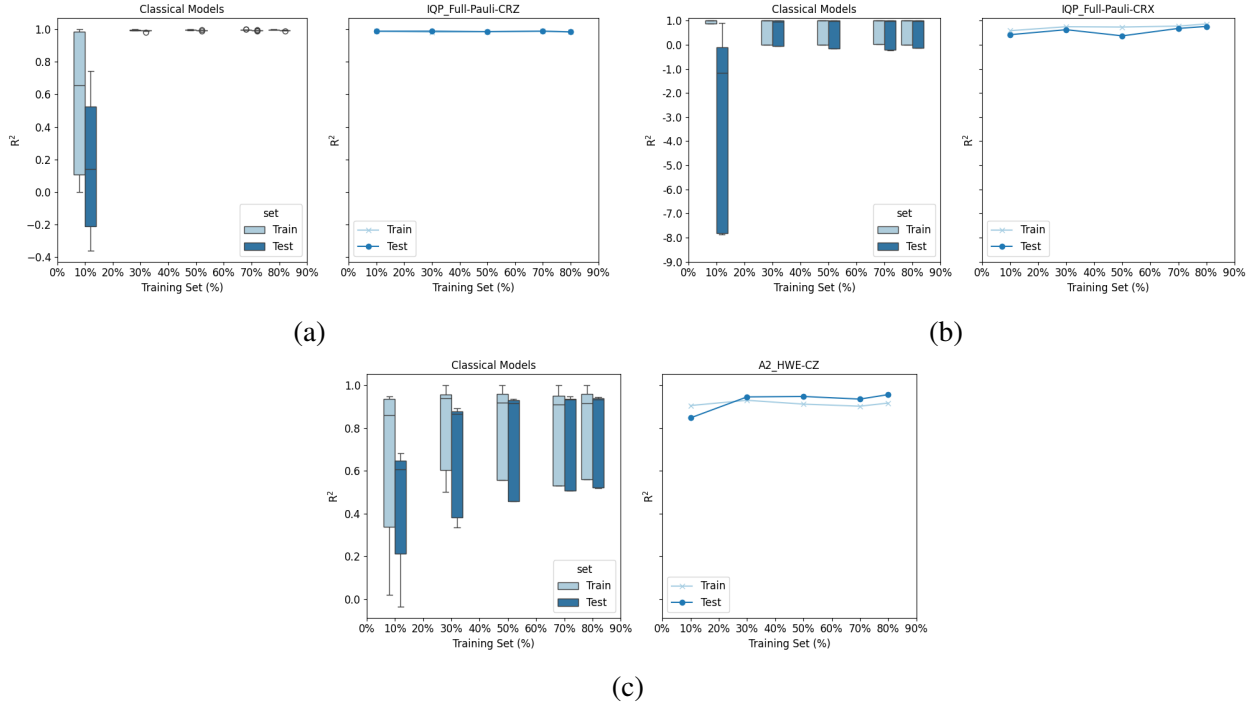(a)   (b)

(c)

Figure 4: Learning curves of the best PQCs (right side in the subfigures) versus various classical models (left side in the subfigure) for the (a) linear, (b) quadratic, and (c) sine function datasets.

optimization level needs to be set to 0!

Lastly, using the function fitting datasets we analyze the effects of error mitigation on the model accuracy by comparing state vector simulations with noisy and error mitigated simulations using the *fake_quebec* backend. Initially, we explored two different methods of error mitigation zero-noise extrapolation (ZNE) using MITIQ[2] and TREX, as implemented in Qiskit. We found that both linear and Richardson ZNE were too costly to simulate on classical computers using 16 qubits for all three function fitting datasets and chose only to use TREX for error mitigation. Our analysis

revealed that the state vector, TREX, and unmitigated IQP_Full-Pauli-CRZ (linear function fitting data) models all have R^2 between 0.98 and 0.99, for both the training and test sets, as highlighted in Fig. 5a. While the performance of the linear data can be attributed to the simiplicity of the dataset, for the quadratic function fitting data using IQP_Full-Pauli-CRX (Fig. 5b), we found that the model with unmitigated error has an R^2 of 0.65 for the training set and 0.45 for the test set. This model is improved slightly by the incorporation of error mitigation using TREX, where the training set R^2 is 0.69 and test set R^2 is 0.60. Overall, the model using TREX offers results more comparable to the state vector simulations, which has a R^2s of 0.80 and 0.68 for the training and test set, respectively. On the sine function fitting data, using A2_HWE-CZ (Fig. 5c), for the state vector, TREX, and unmitigated models, the training sets have R^2s of 0.92, 0.87, and 0.87, respectively, while the test sets have R^2s of 0.96, 0.89, and 0.91, respectively. Overall, while these results are for simple function fitting data, we concluded that TREX offers a computational efficient error mitigation method, with results similar to the state vector calculations.0
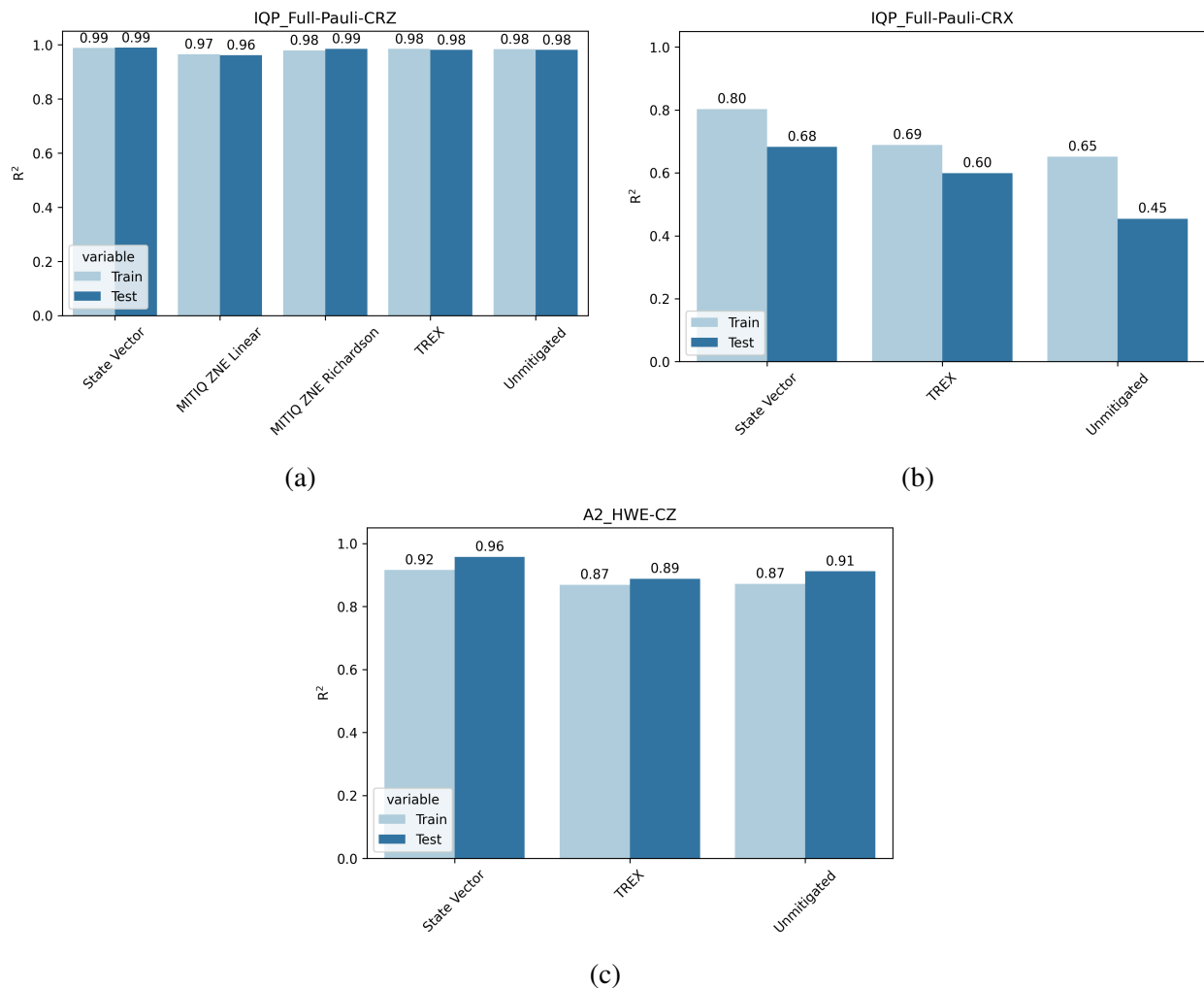
Figure 5: Error mitigation benchmarking for the (a) linear, (b) quadratic, and (c) sine function comparing unmitigated and mitigated error using TREX with state vector simulations.

References

(1) 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.

(2) LaRose, R. et al. Mitiq: A software package for error mitigation on noisy quantum computers. *Quantum* **2022**, *6*, 774, Publisher: Verein zur Forderung des Open Access Publizierens in den Quantenwissenschaften.