

(Chinese abstract here.)

Abstract

As the rapid evolution of quantum hardware industry, various algorithms in quantum computational chemistry are put forward for rigorous implementation. However, due to lack of quantum error correction, a promising hybrid quantum-classical algorithm to find the lowest eigenvalue of a given Hamiltonian for near-term quantum hardware is the variational quantum eigensolver (VQE). To get the best performance out of limited computational resources, using efficient basis sets to represent molecular Hamiltonian in electronic structure problems has become a crucial step. Some particular VQE calculations using Daubechies wavelet basis sets and efficient encoding schemes implemented via hardware-efficient parameterized quantum circuits (PQCs) on real quantum machines or via chemical-inspired PQCs on simulators to date are only accomplishing chemical accuracy on relatively small and simple structure molecules.

In this paper we present a hybrid quantum-classical machine learning computational framework that can be trained unsupervisedly to reliably predict molecular vibrational frequencies. A novel scheme for encoding the interatomic distances of the molecules into the quantum circuits of the quantum-classical framework is also proposed and can allow for a fewer-shot prediction of the molecular ground state potential energy curve. We demonstrate with H_2 , LiH and PN molecules that our framework exhibits high energy accuracy with the calculated results of vibrational frequencies in excellent agreement with experimental data while remains short in quantum circuit depth as the number of qubits scales up. Our approach provides a powerful hybrid quantum-classical computational architecture for the calculation of molecular properties on near-term quantum computer processors. The preliminary results have been accepted as poster by the premier conference in Quantum Information Science, 26th Quantum Information Processing.

Contents

I. Introduction	4
II. General framework	5
A. Problem formulation	5
B. Workflows	7
C. Encoding	9
D. Hamiltonian Preparation and Methods	10
III. Results	11
A. Restoring ground energy curve	11
B. Predicting Frequency	12
C. Comparison of Encoding	14
IV. Conclusion and Outlook	16
References	18
V. Appendix	20
A. Hardware specifications	20
B. Error mitigation	21
1. Twirled Readout Error eXtinction (TREX)	21
2. Zero Noise Extrapolation (ZNE)	22
3. Probabilistic Error Cancellation (PEC)	22
C. Unrolled circuit diagram	23
VI. Publication List	25

I. Introduction

Simulating many-body physics is one of the promising applications on near-term quantum computers, among which is solving the electronic structure problems efficiently, as compared to classical computers. For example, the computational complexity of exact quantum chemical computations scales exponentially with system size on classical computers, while it is in a polynomial manner for variational hybrid quantum-classical algorithms [1–4]. Most recent studies on solving electronic problem with quantum computers focus primarily on approximating the ground state energy curve of given molecules. Here we present a hybrid quantum-classical machine learning computational framework that can be trained unsupervisedly to reliably predict molecular vibrational frequencies to compare with experimental data.

In the noisy intermediate-scale quantum (NISQ) era [5], the available quantum devices to date are relatively small and prone to errors. Without the presence of fault-tolerant quantum computing, we need algorithms that require less resource, while not sacrificing the accuracy when solving electronic structure problems. Common hardware-efficient algorithms generally embed variational quantum circuits (VQC) as part of their workflows, and often feature a two-local [6] design in the VQC ansatz in order to approximate the eigenvalue of a given qubit Hamiltonian. The overall circuit with purely unitary operations, under which quantum states evolve, are necessarily linear. We are, however, engaging in a quantum machine learning (QML) task, which not only ground state energies of the given Hamiltonian are obtained, but also the energies of more unseen Hamiltonians, needed to be predicted. Therefore, non-linearity must be introduced into our quantum circuit to increase its expressibility.

Various ways exist for realizing non-linear operations in quantum circuit. For example, in the quantum convolutional neural network (QCNN) [7], they utilize partial measurements to inject non-linearity into the circuit. Circuit of this kind, extending toward the last measurement, has a much longer depth compared to that of ours with global measurement, and hence in our work the circuit is more likely to run on a NISQ device. Besides utilizing the global measurement as means for achieving nonlinear operations as stated in [8], several other techniques have been introduced in our work to increase the trainability. For example,

we propose a new encoding scheme and post an additional constraint on our circuit parameters during the encoding stage as a kind of a regularizer, which eventually leads to a better convergence, and will be addressed later in the paper.

Generally speaking, our workflow can unsupervisedly train a hybrid variational quantum circuit to calculate the ground state energy curve of a molecule, under a given set of bond lengths. After training, the circuit with fixed parameters can predict the ground state energies at unseen bond lengths, within a promising accuracy. Combined with our regularizer during the encoding stage, accurate molecular vibrational frequency can be obtained, while only few quantum resource are used. We will demonstrate our frequency calculation subroutine with H_2 , LiH and PN molecules and discuss their performance.

The outline of the paper is as follows. In Sec. II A, we formally define the electronic structure problem and the cost function associated with the optimization process. In Sec. II B, we give an illustration of our general framework regarding an advanced problem-solving procedure. In Sec. II C, we introduce the notion of adaptive encoding, and in Sec. II D we mention the better quality of the Hamiltonian used in our work as compared to the Gaussian minimal basis sets. In Sec. III A, we show the capability of our hybrid network to approximate the molecular dissociation curve within chemical accuracy, with relatively few quantum resource required. In Sec. III B and III C, our main results show great agreement with experimental data, and different encoding schemes are compared to demonstrate the advantage of our proposed method. We conclude and aim on several future directions regarding our work in Sec. IV.

II. General framework

A. Problem formulation

The framework we proposed in this paper mainly resolves a predefined electronic structure problem. The problem is formulated as follows: given a Hamiltonian of a many-body quantum system, the goal is to approximate its associated minimum eigenvalue, which is essentially the ground state energy of the system. Whenever there is a change in configu-

ration of the quantum system, a new Hamiltonian must be obtained, hence a new ground state energy. In our problem, only the distance between the constituent atoms is varied. Therefore the calculation of ground state energies at all possible bond lengths is equivalent to recovering the entire ground state potential energy curve, or the dissociation curve of that specific molecule. Once we have calculated adequate number of points near the global minimum, we can obtain the molecular vibrational frequency by taking the second order derivative of the ground state energy curve with respect to bond lengths.

In previous works [9], one would normally obtain the ground state energy of the concerned molecule at several bond lengths individually, and then derive the molecular vibrational frequency from five to seven consecutive energy points near the global minimum. Considering only the hardware-efficient methods other than the Unitary Coupled Cluster (UCC) method [10], normal workflow requires the invoking of new parametrized quantum circuit (PQC) for each individual Hamiltonian, that is, the determination of a unitary $U(\boldsymbol{\theta})$ which represents the overall quantum operation that our circuit has done on the input state, with $\boldsymbol{\theta}$ being a vector of parameters that completely specifies the action of every parametrized gate in our PQC.

A common way of constructing $U(\boldsymbol{\theta})$ is to utilize the set of parameterized single-qubit rotational Pauli gates $\{R_x(\theta), R_y(\theta), R_z(\theta)\}$, combined with some multi-qubit entangling gates (such as CNOT) as the building blocks for our circuit. The circuit can then be trained on its parameters according to some optimization subroutines. Notice that the trained vector of rotation angles $\boldsymbol{\theta} = \{\theta_j\}$ has to be discarded every time when one is calculating the ground state energy $C_{\rho,H}(\boldsymbol{\theta})$ of a new Hamiltonian H ,

$$C_{\rho,H}(\boldsymbol{\theta}) = \text{tr}[HU(\boldsymbol{\theta})\rho U(\boldsymbol{\theta})^\dagger], \quad (1)$$

where $\rho = (|0\rangle\langle 0|)^{\otimes n}$ is an n -qubit input state. In order to obtain enough points for vibrational frequency calculation, the minimization of cost function $C_{\rho,H}(\boldsymbol{\theta})$ must be repeated several times, and the same number of parameter vectors $\boldsymbol{\theta}$ all serve no purpose after a series of point-wise energy calculations.

B. Workflows

In our work, we decide to take the physical value of the trained parameter vector $\boldsymbol{\theta}$ into account. We realize it by considering the loss function comprised of multiple Hamiltonian at once,

$$C_{\rho, \{H_i\}}(\boldsymbol{\theta}) = \sum_i \text{tr}[H_i \tilde{U}(\boldsymbol{\theta}) \rho \tilde{U}(\boldsymbol{\theta})^\dagger], \quad (2)$$

which is evaluated on a single quantum circuit. Notice that we have also introduced non-linear operations to the circuit for more expressibility, and hence the operator $\tilde{U}(\boldsymbol{\theta})$ need not be unitary. In addition, we anticipate our hybrid quantum-classical circuit capable of predicting ground state energies of unseen Hamiltonian $H(l)$ at different bond lengths after training, therefore the information of bond length l must somehow be encoded into the circuit. We denote the bond length encoding scheme by

$$\rho = \rho(l) = \tilde{G}(l)(|0\rangle\langle 0|)^{\otimes n} \tilde{G}(l)^\dagger, \quad (3)$$

where the encoding circuit $\tilde{G}(l)$, which depends on specific bond lengths, again need not be unitary. Now we are ready to train our hybrid network based on the above discussed new cost function formalism.

Fig. 1 shows the unrolled general framework of our computational model. The model is a multi-layer structure, each layer is essentially a parametrized quantum circuit, followed by a global Pauli-Z measurement at the end. The only information transmitting between any two adjacent layers are the set of Pauli-Z expectation values $\{\langle \sigma_z \rangle_p\}, \forall 1 \leq p \leq n$, for a n -qubit quantum circuit. In Fig. 1(a) we illustrate a K -layer structure, where all parameters $\{\boldsymbol{\theta}\}$ in each layer can be updated independently and simultaneously, according to the multidimensional gradient $\frac{\partial L}{\partial \boldsymbol{\theta}}$ of the loss function

$$L_{\{l_i\}, \{H_i\}}(\boldsymbol{\theta}) = \sum_i \text{tr}[H_i \tilde{U}(\boldsymbol{\theta}) \tilde{G}(l_i) \rho \tilde{G}(l_i)^\dagger \tilde{U}(\boldsymbol{\theta})^\dagger], \quad (4)$$

which is calculated by the classical optimizer. During each iteration, a new set of parameters $\{\boldsymbol{\theta}\}$ are obtained from some pre-defined optimization algorithm, and are updated to the quantum circuits for the evaluation of loss function during next iteration.

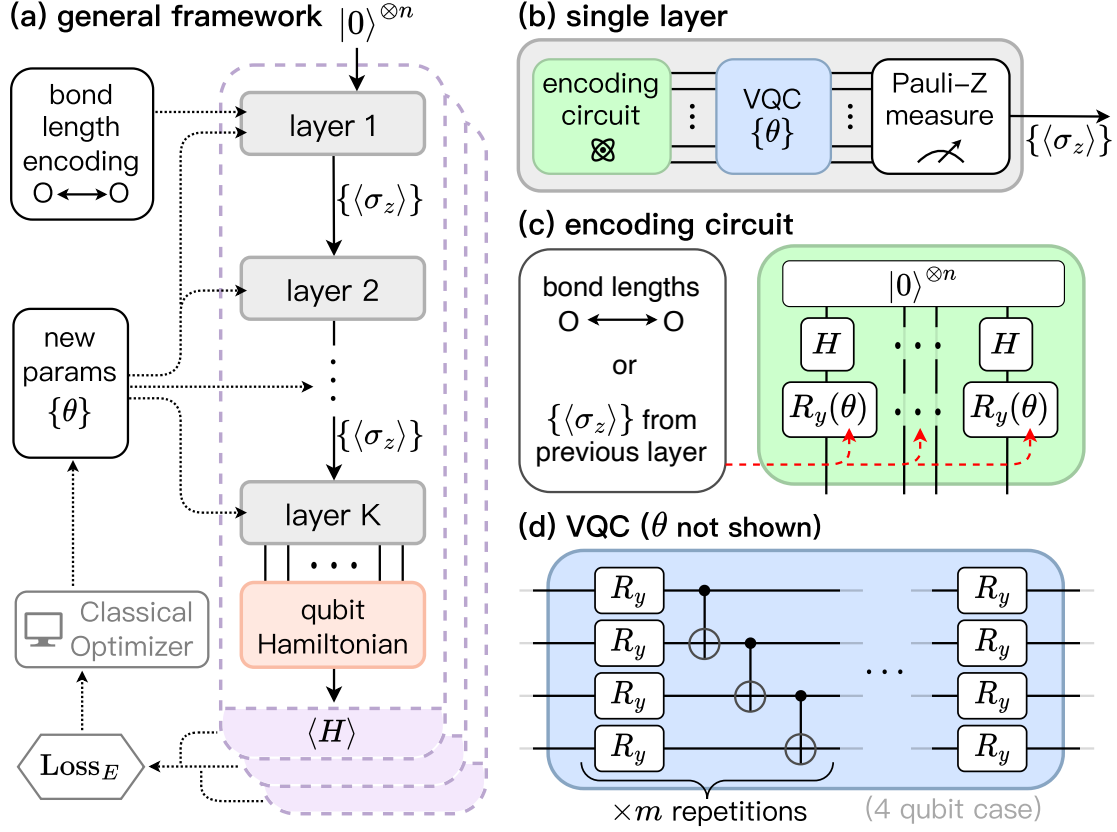


Figure 1. Schematic illustration of the hybrid quantum circuit framework with n qubits, solid lines indicate quantum wires, dotted lines indicate classical interactions with the quantum circuit. **(a)** The high-level general framework with loss-evaluation and parameter-updating routine. Note that the three stacks of $\langle H \rangle$ calculation share the same set of parameters. **(b)** Each single layer is comprised of an encoding part (quantum circuit), a variational quantum circuit, and a Pauli-Z measurement for each qubit. The measurement outcome is then passed to the next layer. **(c)** The general circuit for encoding the input scalars from previous layer (bond lengths or $\langle \sigma_z \rangle$). The red arrows represent functions that maps from input to rotation angles. **(d)** The variational quantum circuit with hardware-efficient ansatz. Linear entanglement pattern is adopted, and its repetition can be adjusted according to the desired expressibility.

Note that we have encoded the information carried by bond lengths into the circuit at the beginning of the framework, where technical details will be discussed in Sec. II C. At the

end of the K^{th} layer, a series of Pauli string representing the qubit Hamiltonian is attached to the circuit such that we can measure the expectation value of $\{H_i\}$, thus able to calculate the loss function defined in Eq. 4. Notice that all trainable parameters are embedded in VQC part, serving as the angle for R_y gates. The hardware-efficient ansatz illustrated in Fig. 1(d) can be fine-tuned so that the repetition number m of the 2-local circuit gives our network sufficient expressibility, while maintaining its trainability.

C. Encoding

There are various ways for encoding classical bond lengths to a quantum state. In our work, the encoding part is universal in all layers, that is, it handles not only molecular bond length, but also the vector of Pauli-Z measurement outcomes from the previous layer. Therefore our encoding scheme is equivalent to the task of finding a suitable set of functions $\{f_i\}$, such that the encoded n -qubit state

$$\begin{aligned} |\psi\rangle_{\text{encode}} &= \Pi_{i=0}^{n-1} \left(\otimes R_y(\theta_i) H \right) |\psi_0\rangle \\ &= \Pi_{i=0}^{n-1} \left(\otimes R_y(f_i(\vec{a})) H \right) |0\rangle^{\otimes n}, \end{aligned} \quad (5)$$

has peak performance in vibrational frequency prediction, given a vector of scalar \vec{a} . The vector can be comprised of either specific bond length $\vec{a} = (l, \dots, l)$, or measurement outcomes from the previous layer $\vec{a} = (\langle\sigma_z\rangle_0, \dots, \langle\sigma_z\rangle_{n-1})$.

Consider the encoding formalism described in Eq. 3, one would naturally come up with a function that represents the direct encoding scheme [11]

$$\theta_i = f_i(\vec{a}) = a_i, \quad (6)$$

which is relatively simple, yet effective. Nevertheless, as we observe the continuity of the exact molecular ground energy curve, a concern has risen. Generally we need to ensure the smoothness of the ground energy curve, in order to have a reliable frequency calculation. This implies that the energy difference between adjacent bond lengths must be suppressed under a threshold, and this tiny energy gap can be accounted to the degree of overlapping of $|\psi\rangle_{\text{encode}}$ from different bond length, l_i and l_{i+1} . Hence our goal is to find a new set of

function that takes

$$\arg \max_{\{f_i\}} \sum_j \left| \langle \psi_j | \psi_{j+1} \rangle_{\text{encode}} \right|^2, \quad (7)$$

where j sums over all distinct bond lengths. Hence, we propose the *adaptive encoding* scheme.

In the adaptive encoding, one makes use of the bond length resolution Δl before training. Suppose we have picked five bond lengths for frequency prediction, $\vec{l} = \{0.70, 0.72, 0.74, 0.76, 0.78\}$ Å. The resolution in this case would be $\Delta l = |l_0 - l_1| = 0.02$ Å, and the set of encoding functions are

$$\theta_i = f_i(\vec{a}) = a_i + \left(i - \frac{n-1}{2}\right) \left(\frac{\Delta l}{2}\right) \quad (8)$$

for a n -qubit circuit. For instance, the parameter vector $\vec{\theta}$ which encode $\vec{a} = (0.74, \dots, 0.74)$ in a direct scheme will remain $(0.74, 0.74, 0.74, 0.74)$ for a 4-qubit circuit, but in the adaptive encoding scheme, the two bond lengths $l_2 = 0.74$ and $l_3 = 0.76$ will be encoded as $\vec{\theta}(l_2) = (0.71, 0.73, 0.75, 0.77)$ and $\vec{\theta}(l_3) = (0.73, 0.75, 0.77, 0.79)$, respectively. We will see in Sec. III C that the latter encoding scheme has significant improvement on the stability of vibrational frequency prediction.

D. Hamiltonian Preparation and Methods

In this work we adopted the Daubechies wavelet basis function [12] for the construction of molecular Hamiltonian, due to its capability of yielding accurate results with excellent computational efficiency, as compared with contracted Gaussian basis sets on large molecules. [13, 14] We first calculate the matrix elements, which are the orbital integrals in second quantized Hamiltonians, by using the BigDFT [15–17] package on a classical computer. Next we transform them into qubit Hamiltonians, hence compatible with quantum processors, during which we also adopt the two-qubit reduction [18]. Therefore, a n qubit Hamiltonian is transformed into a $n-2$ qubit Hamiltonian after encoding. Overall, we obtain three sets of 2, 4, and 6 qubit Hamiltonians for H_2 , LiH , and PN molecules, respectively.

The optimization is performed by the Broyden–Fletcher–Goldfarb–Shanno algorithm [19] with maximum 500 iterations and gradient norm tolerance to stop as 10^{-5} . We run our

hybrid networks on IBM’s 27-qubit quantum machine, `ibmq-montreal`, for all 3 types of molecules (H_2 , LiH , and PN).

We calculate the vibrational frequency based on the harmonic oscillator model of diatomic molecule, which requires the coefficient from the second-order fitting of the energy curve. To obtain a reliable coefficient, we need at least 5 consecutive points of energy on the energy curve. We also initialize all the variational parameters with a Gaussian distribution of mean 0 and standard deviation $\pi/2$ before training. All other molecular configuration datas are from the [Computational Chemistry Comparison and Benchmark DataBase](#).

III. Results

A. Restoring ground energy curve

Table I. Framework configurations for the potential energy curve calculation of H_2 and LiH . For the VQC part of the circuit we adopted linear entanglement throughout the entire experiment. We trained both molecules with 10 energy points. (K denotes the number of layers used in our framework, and m denotes the repetition of ansatz in the VQC part of each layer, as illustrated in Fig. 1.)

Molecule	Hamiltonian	Range (\AA)	K, m
H_2	2 qubit	$0.4 \rightarrow 2.48$	2, 1
LiH	4 qubit	$1.4 \rightarrow 1.8$	2, 1

Fig. 2 shows the capability of our circuit to recover the entire ground state energy curve with only few training points. Table I displays the hyperparameters we used for the calculation. In both cases (H_2 and LiH), we only utilize 2 layers and one repetition of hardware-efficient ansatz in our circuit setting. We can see from the plots that all of the energy points, whether corresponds to the bond lengths for training or testing, converge within the chemical accuracy (0.0016 Ha). As for the purpose of predicting vibrational frequency, we require even fewer training points, since the relevant prediction range shrinks toward the proximity

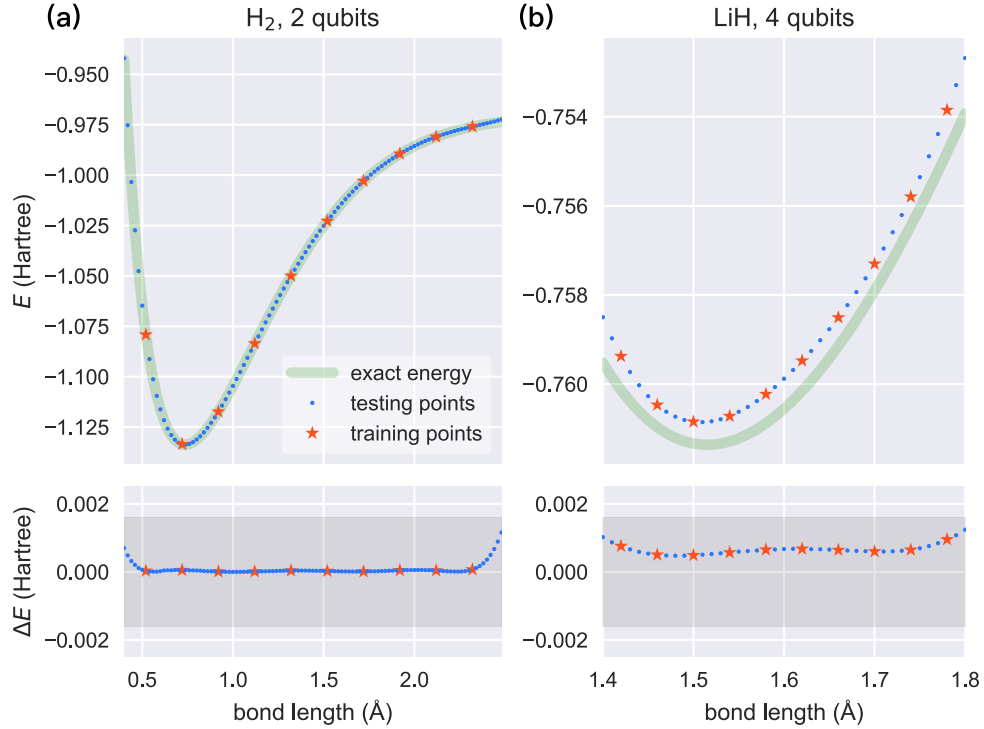


Figure 2. The capability of our proposed framework to recover the entire ground state energy curve, showcased for H_2 (2-qubit Hamiltonian) and LiH (4-qubit Hamiltonian). **(a)** We used $m = 1$ and $K = 2$ architecture for the H_2 molecule, and **(b)** We used the same $m = 1$ and $K = 2$ architecture for the LiH molecule. The trained energies, as well as the predicted ones, fitted well with the exact diagonalization of Hamiltonian. All predicted energies (testing points) converge within the chemical accuracy (indicated as shaded regions in the two bottom plots).

of global minimum. That is, a qualified calculation of frequency need not the information of the entire curve.

B. Predicting Frequency

We present our results on vibrational frequency prediction in Fig. 3. The framework settings we used during the work are listed in Table II.

For H_2 and LiH molecule, we repeated the predicting procedure for 100 times on quantum machine. For PN we repeated for 50 times due to the long queueing time. Table II shows that

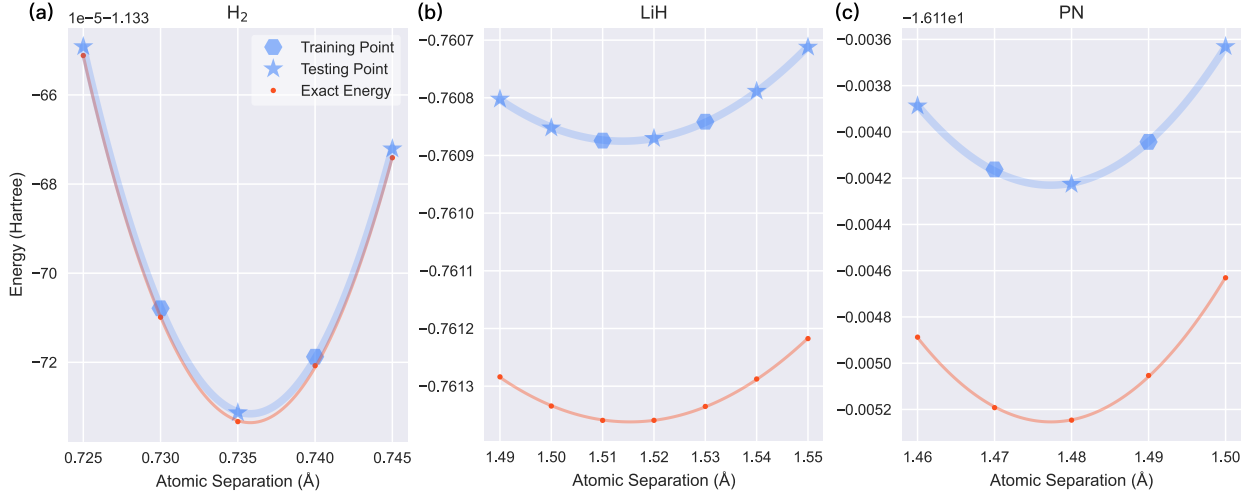


Figure 3. Ground state energy prediction near global minimum for frequency calculations. We train our hybrid network on only two distinct bond lengths for each molecule (indicated as hexagonal markers), and let it predict the energy of neighbor bond lengths (testing points, indicated as star markers) after training. We fit a set of consecutive energy points to second-order polynomial functions (shown as wide solid lines), then extract the coefficient and derive the harmonic vibrational frequency for each molecule. The results display great match with frequencies that are derived from the exact diagonalization of Hamiltonians. **(a)** For H_2 , we computed energy for 5 points, after fitting we obtain $f_{\text{H}_2} = 4460.78 \pm 2.16 \text{cm}^{-1}$. **(b)** For LiH , we computed energy for 7 points, after fitting we obtain $f_{\text{LiH}} = 1424.37 \pm 4.09 \text{cm}^{-1}$. **(c)** For PN , we computed energy for 5 points, after fitting we obtain $f_{\text{PN}} = 1368.55 \pm 9.70 \text{cm}^{-1}$. We conducted the experiment 100 times for H_2 and LiH molecules to mitigate the error from randomness (hence 100 slightly distinct frequencies generated for both molecules), and 50 times for PN molecule, due to the growing queueing time as the qubit number increases. Next, we calculate the average vibrational frequency of each molecule, and serve them as our final predictions. Overall, the above statistics show good align with our anticipated network performance. (Note that we have rescaled the y -axis of H_2 and PN for better visualization, where the scaling factor for recovering the true energy (Y) from the printed energy (y) is anchored at the top-left corner of each plot. The relations are $Y = 10^{-5} \cdot y - 1.133(\text{Hartree})$ for H_2 , and $Y = y - 16.11(\text{Hartree})$ for PN .)

Table II. Framework configurations as well as results for the vibrational frequency calculation of H_2 , LiH and PN . We adopted linear entanglement for the VQC ansatz throughout the entire simulation process to reduce the circuit depth. f_{predict} , f_{diag} , and f_{exp} represent the vibrational frequencies from our network prediction, the exact diagonalization of Hamiltonian, and the experimental data, respectively, and $\text{error}_{\text{diag}}$ and $\text{error}_{\text{exp}}$ denote the deviations of our predicted frequency from the values of the exact diagonalization and experimental data, respectively.

Molecule	K, m	l_{training} (\AA)	l_{testing} (\AA)		
H_2 , 2 qubit	2, 1	0.73, 0.74	0.725, 0.735, 0.745		
LiH , 4 qubit	2, 1	1.51, 1.53	1.49, 1.50, 1.52, 1.54, 1.55		
PN , 6 qubit	3, 2	1.47, 1.49	1.46, 1.48, 1.50		

Molecule	f_{predict} (cm^{-1})	f_{diag} (cm^{-1})	f_{exp} (cm^{-1})	$\text{error}_{\text{diag}}$	$\text{error}_{\text{exp}}$
H_2 , 2 qubit	4460.78 ± 2.16	4462.40	4401.21	-0.03%	1.35%
LiH , 4 qubit	1424.37 ± 4.09	1426.23	1405.50	-0.13%	1.34%
PN , 6 qubit	1368.55 ± 9.70	1365.63	1336.95	0.21%	2.36%

the results are consistent with the exact values obtained from Hamiltonian diagonalization. We train our hybrid network on only 2 bond lengths for all three molecules (indicated as blue hexagonal markers in Fig. 3), and left the other for prediction. Overall, we generated 5, 7, and 5 energy points for H_2 , LiH and PN respectively, which are essential for second-order function fitting, and hence the calculation of frequency. We can see from Table II that the predicted frequencies are very accurate. Note that even as we increased the ansatz repetition and the layer number for 6-qubit PN to increase the network’s expressibility, the overall circuit depth still remains shallow.

C. Comparison of Encoding

We are curious about the degree of performance improvement that the adaptive encoding scheme brings about eventually. In Fig. 4 we show the difference between our proposed

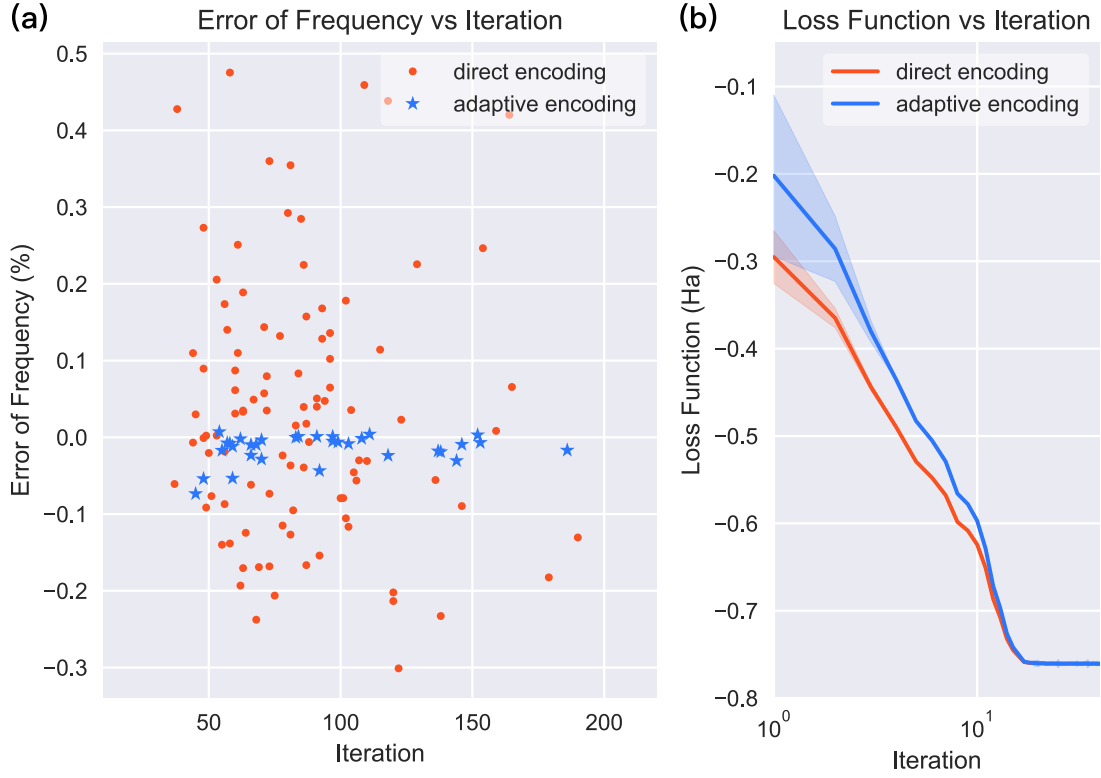


Figure 4. **(a)** Scatter plot for the vibrational frequency calculation of LiH molecule on a 4-qubit hybrid network. We conduct the experiment for 100 times on both direct encoding and adaptive encoding scheme, the advantage of the latter over the former has been statistically visualized on the plot. In terms of frequency error rate, adaptive encoding has a value of -0.014 ± 0.018 , which is significantly more stable and more accurate than that of the direct encoding (-0.034 ± 0.169). **(b)** The loss function trajectory over iterations for both encoding schemes. We can see that the two methods take essentially the same order of magnitude of steps to converge, making the adoption of adaptive encoding from direct encoding with nearly no cost.

encoding method and the direct encoding scheme. From the scatter plot we notice that the adaptive encoding scheme guarantees a more stable prediction of vibrational frequency than the original direct encoding scheme. Such stableness is visualized from a much higher concentration of blue star markers near the zero-error line, as compared to the sparse distribution of orange dots in Fig. 4 (a). Whereas the iteration counts remain same for both

schemes, making the adoption of adaptive encoding from direct encoding with nearly no cost, as illustrated in Fig. 4 (b).

IV. Conclusion and Outlook

In this work, we proposed a hybrid framework that shows great potential in frequency prediction for various molecules. We have shown that our hybrid quantum-classical neural network, combined with adaptive encoding scheme can be trained to obtain the electronic energies at certain bond lengths and then generate the whole potential energy curve. The results of H_2 , LiH and PN are very accurate and the corresponding errors are over an order of magnitude smaller compared to previous work [11]. The circuit’s width and depth are also shortened due to the advantage of Daubechies wavelet basis set over the Gaussian based basis set, and the adoption of global measurement over partial measurement as non-linear operations. The fact that our network succeeds at predicting ground state energy on previously unseen bond lengths may implies that it has been trained to memorize the relation between individual bond length and their corresponding ground state wave function (or at least the state with energy approximating the true ground state energy).

Furthermore, we show the importance of problem-specific encoding technique in comparison with direct encoding scheme. Often we are not really concerned about the absolute value of the ground state energy, but are rather more interested in the energy difference, when the interatomic distance is varied, for example, the molecular vibrational frequency near equilibrium bond length. Through the design of new encoding methods, we aim to maintain the correlation between energies under various bond lengths, which is illustrated in Fig. 5. The idea is similar to the regularization technique in the context of classical machine learning. This approach may also be generalized to accomplish the encoding of two-dimensional molecular configuration data in future work. On the other hand, the intermediate nonlinear measurements can reduce the circuit depth, which are more suitable for NISQ devices, and the layer-structure proposed in this work may serve as a fundamental building block for hybrid VQCs in the near-term future.

During our work, Barren plateau has emerged as the qubit number scales up. From

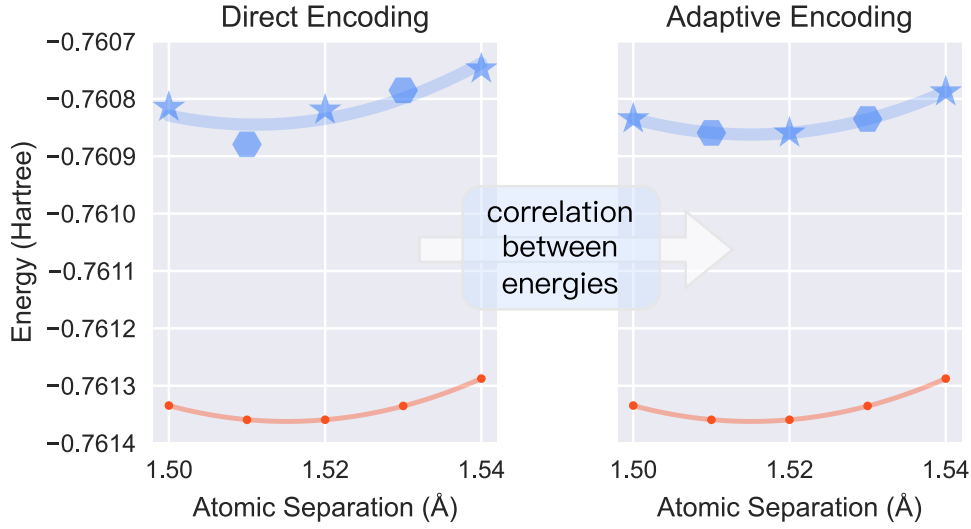


Figure 5. The adaptive encoding method proposed in our work serves as one of the means to establish correlation between energies of neighboring bond lengths. As one can see, high correlation makes energies (blue marks) distribute less randomly along the fitted curve (blue curve). Thus, the figure shows the advantage of adaptive encoding over direct encoding, under the same environmental settings (LiH, 4-qubit Hamiltonian).

Ref. [8] we know that *strongly correlating parameters* and/or initializing close to the solution (and then restricting the ansatz to explore the region close to the initialization) are the most effective approaches to avoid exponentially vanishing cost gradients. Thus, in the future work, we can post a stronger constraint on the circuit parameters to maintain their correlation, to avoid low trainability. We will also keep on to investigate the performance of our proposed framework on quantum hardware with the presence of noise in the future.

Acknowledgments:

The author thanks Prof. Hsi-Sheng Goan for invaluable guidance and support throughout this project. The author also thanks Jyh-Pin Chou, Yuan-Chung Cheng for great suggestions and discussions. The author also thanks IBM Quantum Hub at NTU for providing computational resources and accesses for conducting the experiments on real quantum device. Chia-Tung Chu is supported by the National Science and Technology Council under Grants

- [1] I. M. Georgescu, S. Ashhab, and F. Nori, Quantum simulation, *Reviews of Modern Physics* **86**, 153 (2014).
- [2] 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).
- [3] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, *Reviews of Modern Physics* **92**, 015003 (2020).
- [4] D. S. Abrams and S. Lloyd, Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors, *Physical Review Letters* **83**, 5162 (1999).
- [5] J. Preskill, Quantum computing in the nisq era and beyond, *Quantum* **2**, 79 (2018).
- [6] F. G. Brandao, A. W. Harrow, and M. Horodecki, Local random quantum circuits are approximate polynomial-designs, *Communications in Mathematical Physics* **346**, 397 (2016).
- [7] I. Cong, S. Choi, and M. D. Lukin, Quantum convolutional neural networks, *Nature Physics* **15**, 1273 (2019).
- [8] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, *PRX Quantum* **3**, 010313 (2022).
- [9] S. McArdle, A. Mayorov, X. Shan, S. Benjamin, and X. Yuan, Digital quantum simulation of molecular vibrations, *Chemical science* **10**, 5725 (2019).
- [10] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, Generalized unitary coupled cluster wave functions for quantum computation, *Journal of chemical theory and computation* **15**, 311 (2018).
- [11] R. Xia and S. Kais, Hybrid quantum-classical neural network for calculating ground state energies of molecules, *Entropy* **22**, 828 (2020).
- [12] C.-L. Hong, T. Tsai, J.-P. Chou, P.-J. Chen, P.-K. Tsai, Y.-C. Chen, E.-J. Kuo, D. Srolovitz, A. Hu, Y.-C. Cheng, *et al.*, Accurate and efficient quantum computations of molecular properties using daubechies wavelet molecular orbitals: A benchmark study against experimental data, *PRX Quantum* (2022).

- [13] S. Mohr, L. E. Ratcliff, L. Genovese, D. Caliste, P. Boulanger, S. Goedecker, and T. Deutsch, Accurate and efficient linear scaling dft calculations with universal applicability, *Physical Chemistry Chemical Physics* **17**, 31360 (2015).
- [14] L. E. Ratcliff, W. Dawson, G. Fisicaro, D. Caliste, S. Mohr, A. Degomme, B. Videau, V. Cristiglio, M. Stella, M. D'Alessandro, *et al.*, Flexibilities of wavelets as a computational basis set for large-scale electronic structure calculations, *The Journal of chemical physics* **152**, 194110 (2020).
- [15] L. Genovese, A. Neelov, S. Goedecker, T. Deutsch, S. A. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson, A. Bergman, *et al.*, Daubechies wavelets as a basis set for density functional pseudopotential calculations, *The Journal of chemical physics* **129**, 014109 (2008).
- [16] S. Mohr, L. E. Ratcliff, P. Boulanger, L. Genovese, D. Caliste, T. Deutsch, and S. Goedecker, Daubechies wavelets for linear scaling density functional theory, *The Journal of chemical physics* **140**, 204110 (2014).
- [17] L. Genovese, B. Videau, M. Ospici, T. Deutsch, S. Goedecker, and J.-F. Méhaut, Daubechies wavelets for high performance electronic structure calculations: The bigdft project, *Comptes Rendus Mécanique* **339**, 149 (2011).
- [18] M. Steudtner and S. Wehner, Fermion-to-qubit mappings with varying resource requirements for quantum simulation, *New Journal of Physics* **20**, 063010 (2018).
- [19] C. G. Broyden, The convergence of a class of double-rank minimization algorithms 1. general considerations, *IMA Journal of Applied Mathematics* **6**, 76 (1970).

V. Appendix

A. Hardware specifications

All quantum circuits in our work are run on the IBM's 27-qubit quantum computer, `ibmq-montreal`. The calibration datas of relevant qubits are listed in Table III, while Fig. 6 shows the hardware's qubit map view. The experiments are chosen to conduct on qubits $\#\{1,2,3,4,5,8\}$, so that the actual CNOT gates number can be minimized hence mitigate the errors.

Table III. Calibration data for 27-qubit quantum machine `ibmq-montreal` with regard to all qubits used. For each qubit, six types of relevant properties are listed, including relaxation time (T_1), dephasing time (T_2), qubit operation frequency, anharmonicity, Pauli-X error and CNOT error. We run our quantum circuits on qubits $\#\{1,4\}$ for H_2 , qubits $\#\{1,2,3,4\}$ for LiH , and qubits $\#\{1,2,3,4,5,8\}$ for PN , respectively. All listed values have been averaged within the time span of our experiments, from January 14th, 2023 to February 26th, 2023. The notation for CNOT error A_B represents the error of CNOT gate connecting qubits $\#A$ and $\#B$.

Qubit #	T_1 (μs)	T_2 (μs)	Frequency (GHz)	Anharmonicity (GHz)	Pauli-X error (10^{-4})	CNOT error (10^{-3})
1	161.8354	30.7795	4.8327	-0.3247	2.1961	$1_4:9.336354071$; $1_2:8.006437467$
2	119.6453	140.7872	4.9821	-0.3386	2.3030	$2_3:7.8294181379$; $2_1:8.006437467$
3	64.8695	41.4315	5.1040	-0.3352	3.8469	$3_5:11.082183326$; $3_2:7.8294181379$
4	146.0053	131.9639	5.0024	-0.3387	2.0058	$4_1:9.336354071$
5	124.2324	105.1834	5.0339	-0.3372	2.6063	$5_3:11.082183326$; $5_8:6.1232952963$
8	92.3221	99.0577	4.9072	-0.3243	2.4756	$8_5:6.1232952963$

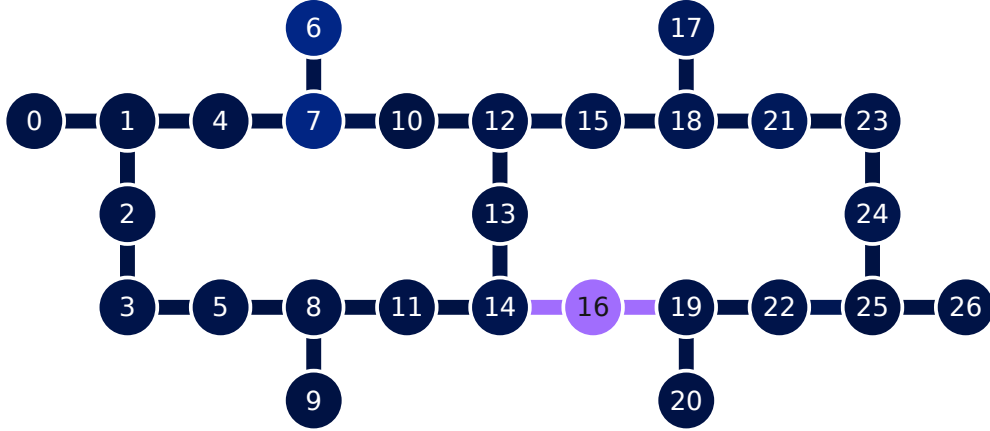


Figure 6. Map view of the 27-qubit `ibmq-montreal` quantum machine, we exploited six connected qubits ($\#\{1,2,3,4,5,8\}$) during our experiments.

B. Error mitigation

Quantum computers are subject to various types of errors that can occur due to environmental factors such as thermal noise, imperfect control pulses, and decoherence. Therefore, quantum error mitigation is a crucial part in our work, as it is essential for achieving reliable circuit operations, since untreated errors can significantly affect the accuracy of quantum computations, leading to incorrect results. To address these challenges, we exploited various quantum error mitigation techniques in `Qiskit Runtime` primitive (provided by IBM) to minimize the impact of errors on our proposed framework. In the following subsections, we briefly elaborate all three techniques that are used in our experiments.

1. *Twirled Readout Error eXtinction (TREX)*

The Twirled Readout Error eXtinction method reduces measurement error by diagonalizing the noise channel associated with measurement by randomly flipping qubits via X gates immediately prior to measurement, and flipping the corresponding measured bit if an X gate was applied. A rescaling term from the diagonal noise channel is learned by benchmarking random circuits initialized in the $|0\rangle$ state. Consequently, the bias from expectation values

which is resulted from readout noise can be removed.

2. Zero Noise Extrapolation (ZNE)

The Zero Noise Extrapolation method computes an expectation value of the observable for different noise factors (amplification stage), and then uses the measured expectation values to infer the ideal expectation value at the zero-noise limit (extrapolation stage). This procedure is visualized in Fig. 7. This approach tends to reduce errors in expectation values, but is not guaranteed to produce an unbiased result. The overhead of this method scales with the number of noise factors (which is 3 times overhead in our experiment’s setting).

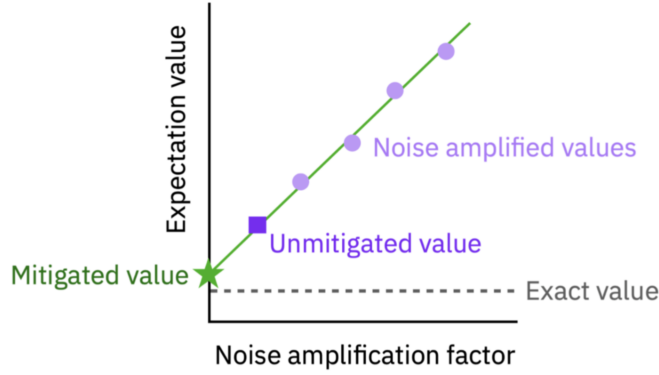


Figure 7. Illustration of the Zero Noise Extrapolation (ZNE) method.

3. Probabilistic Error Cancellation (PEC)

The Probabilistic Error Cancellation method mitigates error by learning and inverting a sparse noise model that is able to capture correlated noise. PEC returns an unbiased estimate of an expectation value so long as learned noise model faithfully represents the actual noise model at the time of mitigation. In practice, the experimental procedure for learning the noise model has ambiguities due to certain error terms that cannot be independently distinguished. These are resolved by a symmetry assumption, which depending on the true underlying noise may lead a biased estimate of the mitigated expectation values due to using an imperfect noise model. In our experiments’ setting, this method specifically addresses

noise in self-inverse two-qubit gates, so it first stratifies each input circuit into an alternating sequence of simultaneous 1-qubit gates followed by a layer of simultaneous 2-qubit gates. Then it learns the noise model associated with each unique two-qubit gate layer, as illustrated in Fig. 7.

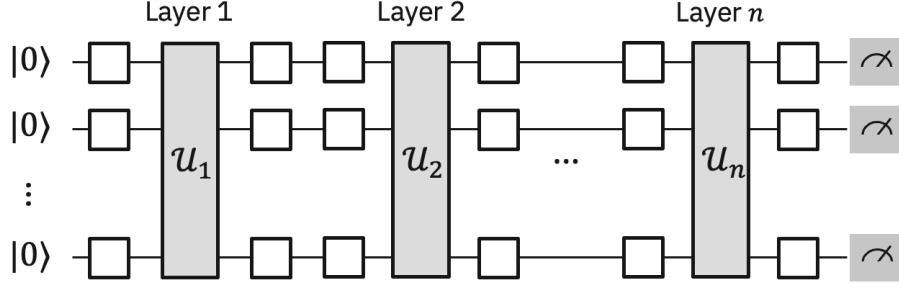


Figure 8. An example of a stratified circuit to achieve Probabilistic Error Cancellation, where the layers of two-qubit gates are labeled layer 1 through n . Note that each U_i is composed of two-qubit gates on the native connectivity graph of the quantum processor. The open boxes represent arbitrary single-qubit gates.

The overhead of this method scales with the number of noise factors. In our experiments the expectation value is sampled at three noise factors, leading to a roughly 3-times overhead.

Since PEC uses a quasi-probability method to mimic the effect of inverting the learned noise, it requires sampling from a randomized circuit family associated with the user’s original circuit. Therefore, applying PEC will increase the variability of the returned expectation value estimates unless the number of samples per circuit is also increased for both input and characterization circuits. The amount of samples required to counter this variability scales exponentially with the noise strength of the mitigated circuit, thus it is a relatively expensive error-mitigation method.

C. Unrolled circuit diagram

To give a concrete example of the circuit being run on the hardware, in Fig. 6 we explicitly print out the quantum part of our hybrid network for PN molecule with the same parameters

($K = 3$ and $m = 2$) listed in Table II.

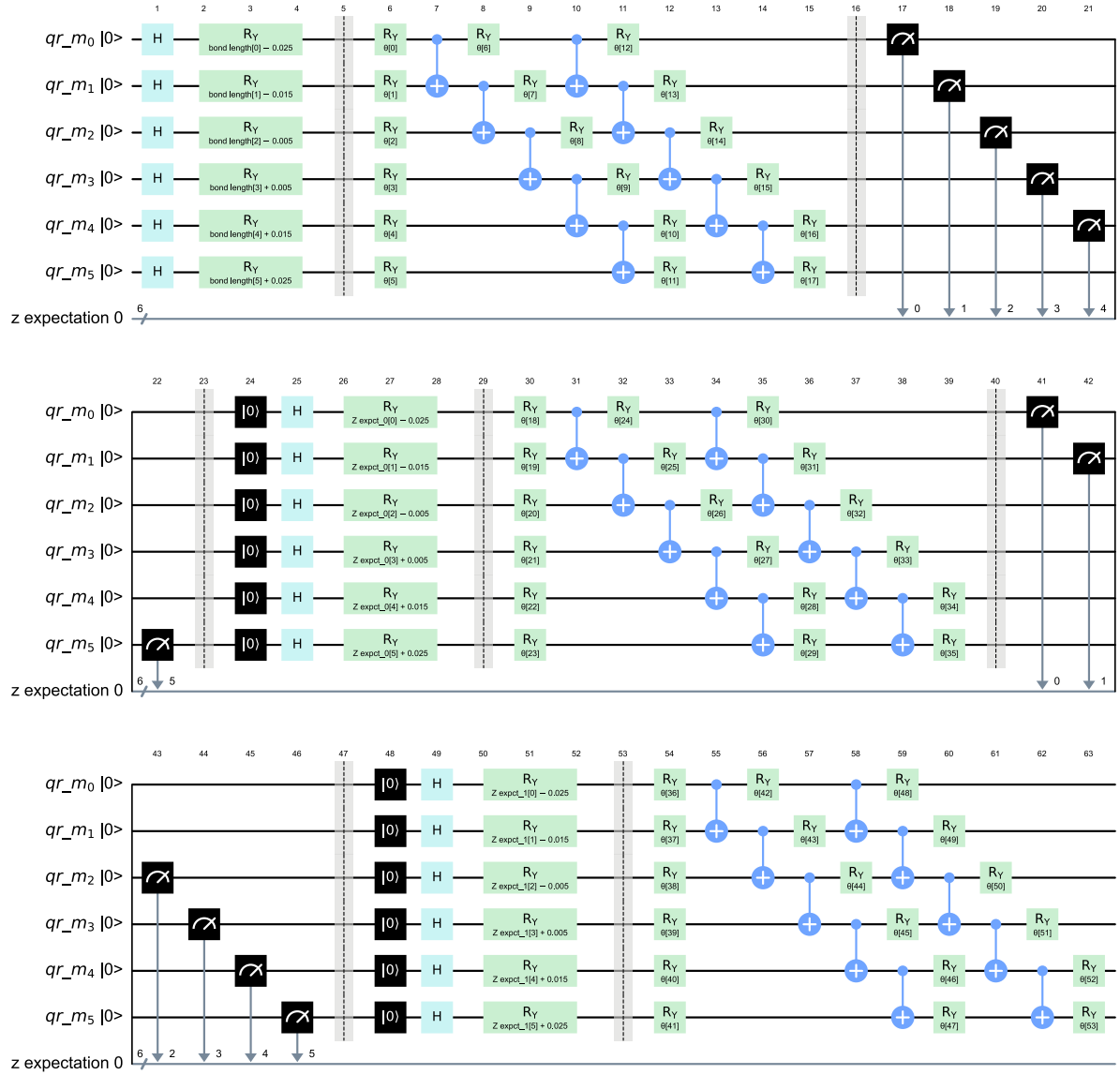


Figure 9. The quantum part (unrolled) of our hybrid network for PN molecule, which is essentially a 6-qubit circuit with framework parameters $K = 3$ and $m = 2$ (referring to Fig. 1).

VI. Publication List

[1] Chia-Tung Chu, Shi-Kai Chou, and Hsi-Sheng Goan, "Hardware-Efficient and Accurate Unsupervised Quantum Machine Learning on Electronic Structure Problems", in *proceedings of 26th Quantum Information Processing*, Ghent, Belgium, February 2023. [\[link\]](#)