Supplementary Material for "Machine Learning Assisted Construction of a Shallow Depth Dynamic Ansatz for Noisy Quantum Hardware"

Sonaldeep Halder, Anish Dey, Chinmay Shrikhande, and Rahul Maitra

S1. Details of Restricted Boltzmann Machines

A. Training and generating procedure

1. The configurations and their corresponding probabilities from the computational state basis are extracted after converging the energy variationally using single and double excitations pruned at the MP2 level (dUCCSD ansatz as described in section **II a** of the main article). The configurations with probabilities greater than 10^{-5} are considered for our dataset. While using statevector, exact probabilities are taken, whereas when using *shot-based* simulators (where expectation values are calculated through repeated measurements), the probabilities are obtained through measurements.
2. These configurations are used to generate a training set of 500000 points based on their corresponding probabilities. Sklearn's BernoulliRBM is used for training on the dataset. The hyperparameters of our model are given below:
   Number of hidden nodes = 30
   Learning rate = 0.01
   Number of iterations = 30
   Batch Size = 10
3. The trained parameters of the model are obtained after training. The activation probabilities of the hidden layer are used further to get the activation probabilities of the visible layer by utilizing the model's parameters. Parameters are estimated using Persistent Contrastive Divergence.
4. The "Tower Sampling Algorithm" is deployed to ensure the total spin and number of particles are conserved.
5. The tower sampled data is Gibbs Sampled for two consecutive steps to get a transformed data set. Decreasing the number of Gibbs sampling introduces more noise into the generated data, whereas increasing the number of Gibbs sampling steps makes the generated data more like the training data.
6. This study focuses on triples therefore, only the triples are extracted from the generated data. This method can be extended to other higher-excited configurations as well.
7. MP2 values are used as a measure to take into account the most dominant triples, as described in section **II C** of our main article.
8. The ansatz is updated accordingly, and the newly generated dominant triples are added to the existing training dataset. The steps from 2-9 are repeated until our model ceases to generate useful triples for 6 iterations consecutively.

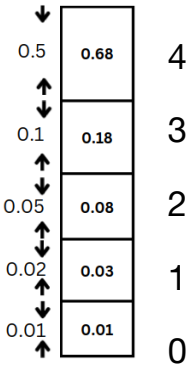
B. A Brief Description of Tower Sampling

Suppose we have the activation probabilities of the visible layer (qubit size/dimension of visible layer = 10, number of electrons = 4)

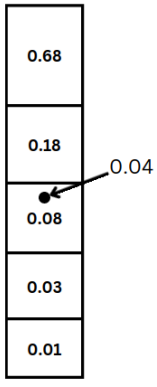
0.01	0.02	0.05	0.1	0.5	0.37	0.12	0.45	0.78	0.04
0	1	2	3	4	5	6	7	8	9

The first 5 units of the visible layer correspond to α set of spin orbitals and the later corresponds to β set of spin orbitals. To ensure that only 2 electrons (number conservation) are present in each of these halves (spin conservation), this sampling method is used. We use the algorithm on each of these halves separately.

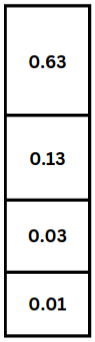
1. A tower of height $\sum_{i=0}^{n} P_i$ is constructed with the height of each block to be the activation probability of each unit of the visible layer (here n=4) -



2. A random number between 0 and $\sum_{i=0}^{4} P_i$ (height of the tower) is generated. Suppose the random number comes out to be 0.04. This lies in the 3rd block of the tower. The electron is filled in the 3rd block (label of index 2) -



3. A new tower is constructed with the probabilities of the remaining 4 units. Steps 1 to 3 are repeated until we have exhausted all the electrons (here, in this case, the number =2).



The same is done for the set of β orbitals.

S2. A brief discussion on *scatterer* operators

A more formal description regarding the *scatterers* can be obtained from the work of Halder et al. [1], where they highlight the generation of higher-order excitations as an outcome of operator non-commutativity and disentangled (factorized) structure of the unitary through a set of exponentiated nested commutator terms.

$$e^{\hat{\sigma}}e^{\hat{\kappa}} = e^{(\hat{\sigma}+\hat{\kappa})+[\hat{\sigma},\hat{\kappa}]+...} \tag{1}$$

If $\hat{\kappa}^D$ represents a rank two excitation operator, the commutator in Eq. (1) results in the generation of a rank three excitation operator ($\hat{\kappa}^T$):

$$[\hat{\sigma}, \hat{\kappa}^D] \longrightarrow \hat{\kappa}^T \tag{2}$$

Multiple exponentiated *scatterers* can act consecutively over the same low-rank excitation operator to generate excited configurations that may differ by more than rank one. This occurs due to the nested commutator that arises in such cases.

$$\begin{aligned} e^{\hat{\sigma}_2}e^{\hat{\sigma}_1}e^{\hat{\kappa}} &= e^{\hat{\sigma}_2}e^{(\hat{\sigma}_1+\hat{\kappa})+[\hat{\sigma}_1,\hat{\kappa}]+...} \\ &= e^{\hat{\sigma}_2+\hat{\sigma}_1+\hat{\kappa}+[\hat{\sigma}_2,\hat{\kappa}]+[\hat{\sigma}_1,\hat{\kappa}]+[\hat{\sigma}_2,[\hat{\sigma}_1,\hat{\kappa}]]+...} \end{aligned} \tag{3}$$

If $\hat{\kappa}^D$ represents a rank two excitation operator, the commutators in Eq. (3) result in different rank three excitation operators and even other higher rank excitation operators such as quadruples ($\hat{\kappa}^Q$) depending on the excitation structure of the two *scatterers* (σ_1 and σ_2).

$$\begin{aligned} [\hat{\sigma}_1, \hat{\kappa}^D] &\longrightarrow \hat{\kappa}_1^T \\ [\hat{\sigma}_2, \hat{\kappa}^D] &\longrightarrow \hat{\kappa}_2^T \\ [\hat{\sigma}_2, [\hat{\sigma}_1, \hat{\kappa}^D]] &\longrightarrow \hat{\kappa}^Q \end{aligned} \tag{4}$$

Eq. (4) is also provided in the main article.

S3. A comparison of ADAPT-VQE and RBM-dUCC

ADAPT-VQE[2] is among the most popular dynamic ansatz constructing protocols, which produces an extremely shallow-depth ansatz, all the while capturing excellent correlation energy during its implementation on molecular systems. However, it relies on measurements to identify dominant contributors from a pool while tailoring the ansatz. This produces a large measurement cost overhead. Apart from the cost, the construction is affected as each state preparation and subsequent measurements are associated with errors due to noise in the current quantum hardware. First, we compare the noiseless performance of ADAPT-VQE with that of RBM-dUCCSDT_S. The results are provided in Fig. S1 for BH (STO-3G). ADAPT-VQE has been implemented using the respective module from Qiskit [3]. All other parameters are kept the same as that while producing Fig. 3 (main article). The pool for ADAPT was set to be generalized Singles and Doubles excitations as provided by Qiskit[3].

From Fig. S1, it can be inferred that ADAPT produces an ansatz that is shallower than RBM-dUCCSDT_S by $\approx$ 500 CNOT gates. However, it is essential to note that this efficiency gain comes with a trade-off – ADAPT ansatz sacrifices a portion of its accuracy when contrasted with the exceptional accuracy offered by RBM-dUCCSDT_S in capturing correlation.

Currently, available Noisy Intermediate-Scale Quantum devices suffer from various sources of noise. As a result, expectation values (evaluated using state preparation followed by measurement with respect to the desired operator) deviate from the ideal one. Ansatz constructing protocols that heavily rely on such expectation values is highly sensitive to such deviations. In Table S1, we present the accuracy of the ansatz prepared using ADAPT and RBM-dUCCSDT_S under noise. It must be noted that once the ansatz is prepared (under noise), we run noiseless VQE to assess how much correlation these "distorted" ansatzes can capture. The preparation of noisy simulators can be found in Section 3.2 of the main article. We have not used any additional error mitigation techniques for the procedures mentioned in this work.

Table S1 shows that the ansatz generated by the ADAPT procedure deviates from the optimum one in terms of capturing proper correlation. This can be attributed to the selection procedure, which employs

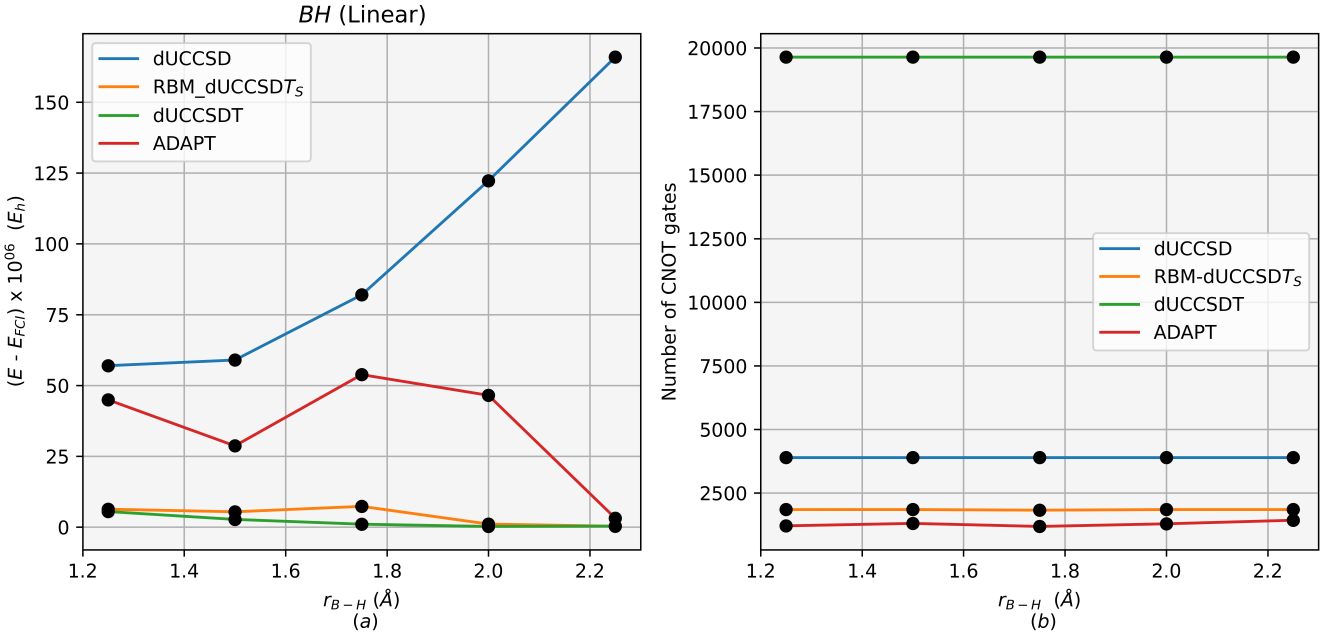


Figure S1: A comparison of accuracy(a) and gate counts(b) of ansatz produced by ADAPT-VQE with that of RBM-dUCCSDT_S under noiseless case. This comparison also includes dUCCSD and dUCCSDT ansatzes.

	ADAPT Ansatz	RBM-dUCCSDT_S Ansatz
Energy (in E_h; Ansatz generated under no noise)	-24.69159675	-24.69159959
Energy (in E_h; Ansatz generated under noise)	-24.67943577	-24.69159972

Table S1: A comparison of the accuracy of ADAPT and RBM-dUCCSDT_S ansatz for BH (2.25Å, STO-3G, core frozen). The energy is obtained after noiseless VQE optimization of the generated ansatz. FCI energy for this system is -24.69159992 (E_h).

measurements (expectation value evaluations) for the calculation of energy gradients[2]. Under noise, such evaluations would produce erroneous values impacting the ansatz construction. As a result, the ansatz is likely to be less expressible[4]. To elucidate operator selection discrepancies (without any additional error mitigation techniques), we compare each operator's gradient in the selection pool at an intermediate step of the ADAPT protocol under a noiseless and noisy scenario.

From Fig. S2, it can be discerned that the maxima or minima of the gradient values occur at different points of the operator pool for noise and noiseless cases. As the ansatz keeps growing, the optimum choice of operators from the pool will become more and more difficult. This will result in a deviation of the final ansatz from the optimum one, possibly by a large margin (especially as the size of the system becomes larger).

References

[1]D. Halder, V. Prasannaa, and R. Maitra, "Dual exponential coupled cluster theory: unitary adaptation, implementation in the variational quantum eigensolver framework and pilot applications", The Journal of Chemical Physics **157** (2022).

[2]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**, 3007 (2019).

[3]H. Abraham *et. al*, *Qiskit: an open-source framework for quantum computing*, 2021.

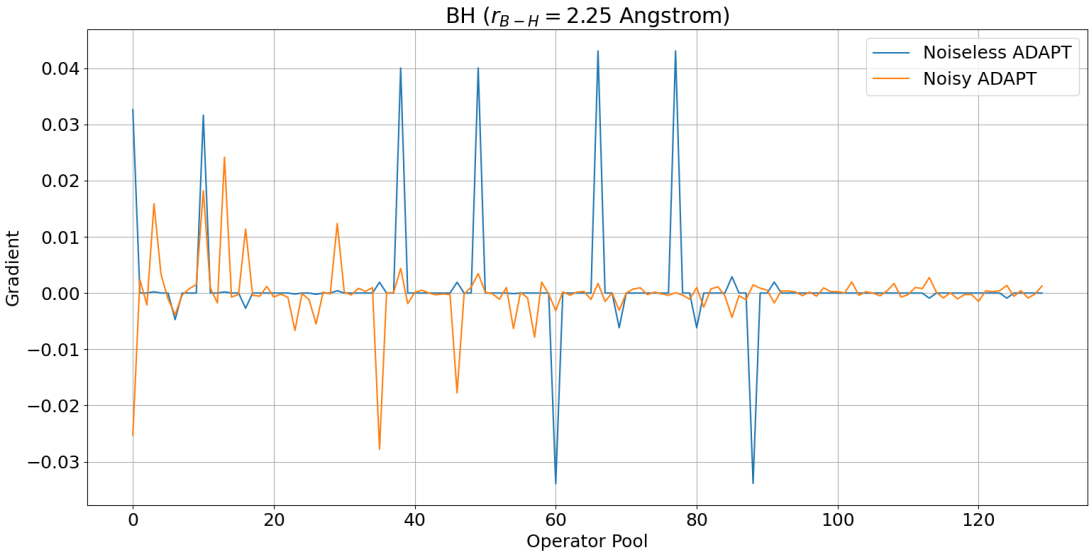


Figure S2: A comparison of gradient values during an intermediate growth step of ADAPT. Each gradient value is obtained for the noisy case by averaging 20 runs. The gradient has been calculated after the ADAPT is arbitrarily allowed to grow the ansatz (in a noiseless simulation) till it has 7 parameters. In the noiseless scenario, if ADAPT is allowed to run to completion, the ansatz contains 19 parameters. Under noise, ADAPT protocol terminates after the ansatz in 11 parameters long.

[4]S. Sim, P. D. Johnson, and A. Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms", Advanced Quantum Technologies **2**, 1900070 (2019).