Predicting excited states from ground state wavefunction by supervised quantum machine learning

MACHINE LEARNING
Science and Technology

Predicting excited states from ground state wavefunction by supervised quantum machine learning

Hiroki Kawai[1,2] and Yuya O. Nakagawa[3]

[1] Department of Physics, Boston University, 590 Commonwealth Avenue Boston, MA, United States of America, 02215
[2] Electrical and Computer Engineering Department, Boston University, 8 Saint Mary's Street, Boston, MA, United States of America, 02215
[3] QunaSys Inc. Aqua Hakusan Building 9F, 1-13-7 Hakusan, Bunkyo, Tokyo 113-0001, Japan

E-mail: hirokik@bu.edu and nakagawa@qunasys.com

Abstract

Excited states of molecules lie in the heart of photochemistry and chemical reactions. The recent development in quantum computational chemistry leads to inventions of a variety of algorithms that calculate the excited states of molecules on near-term quantum computers, but they require more computational burdens than the algorithms for calculating the ground states. In this study, we propose a scheme of supervised quantum machine learning which predicts the excited-state properties of molecules only from their ground state wavefunction resulting in reducing the computational cost for calculating the excited states. Our model is comprised of a quantum reservoir and a classical machine learning unit which processes the measurement results of single-qubit Pauli operators with the output state from the reservoir. The quantum reservoir effectively transforms the single-qubit operators into complicated multi-qubit ones which contain essential information of the system, so that the classical machine learning unit may decode them appropriately. The number of runs for quantum computers is saved by training only the classical machine learning unit, and the whole model requires modest resources of quantum hardware that may be implemented in current experiments. We illustrate the predictive ability of our model by numerical simulations for small molecules with and without noise inevitable in near-term quantum computers. The results show that our scheme reproduces well the first and second excitation energies as well as the transition dipole moment between the ground states and excited states only from the ground states as inputs. We expect our contribution will enhance the applications of quantum computers in the study of quantum chemistry and quantum materials.

1. Introduction

The rapid growth of the machine learning technology in the last decade has revealed its potential to be utilized in various engineering fields such as image recognition, natural language processing, and outlier detection [1, 2]. Its applications to scientific fields have also attracted numerous attentions recently as well as those to engineering. One of the most active research areas is physical science [3], especially studies of quantum many-body systems including condensed matter physics and quantum chemistry. For example, one can classify a phase of matter from its wavefunction [4, 5] or predict the atomization energy of molecules [6–8] from their molecular structures with sophisticated machine learning techniques.

Most of those researches employ *classical* machine learning, with which classical data are processed by classical algorithms and computers. On the other hand, machine learning algorithms on a quantum processor have been developed since the invention of the Harrow-Hassidim-Lloyd (HHL) algorithm, and they are dubbed as 'quantum machine learning' [9–16]. In the last few years, there has been surging interest in quantum machine learning leveraging the variational method [17–22], in which a shallow quantum circuit parameterized with classical parameters such as the angles of the rotational gates is optimized with a classical optimization algorithm to find optimal parameters for performing the given objective. This is

because a primitive type of quantum computers is about to be realized in the near future, and such machines may have the potentials to outperform classical computers [23, 24]. Those near-term quantum computers are called noisy intermediate-scale quantum (NISQ) devices [25] and consist of hundreds to thousands of physical, non-fault-tolerant qubits.

So far, quantum machine learning has been mostly applied to classical computing tasks with classical data such as pattern recognition of images [18, 19, 22]. In those studies, the classical data must be encoded in quantum states to be processed by quantum computers, but the encoding is generally inefficient; it requires the exponentially large number of gates to encode classical data into a quantum state unless the data have a structure of the tensor product which is compatible with that of qubits [26–28].

Therefore, it is natural to think of performing tasks with *quantum nature*. In this study, we consider the following task: predicting excited-state properties of a given molecular system from its ground state wavefunction. Specifically, we are interested in the Hamiltonian for the electronic states of molecules. The question we raise and want to solve leveraging quantum machine learning is whether it is possible to predict properties of the excited states from the ground state wavefunction $|\psi_0\rangle$. According to the celebrated Hohenberg-Kohn theorem [29], one can determine an external potential for electrons and thereby the whole original electron Hamiltonian from its ground state electron density $\rho_0(r) = \langle\psi_0|\hat{r}|\psi_0\rangle$ up to constant, where $\hat{r}$ is the position operator. Hence, it should be also possible to predict the excited states from the ground state in principle.

The task we propose here has various practical and conceptual attractions from the viewpoint of quantum machine learning and the studies of quantum many-body systems. First, practically, computing excited states of a given Hamiltonian needs significantly larger computational cost and is more difficult than computing the ground state [30, 31]. Since the excited-state properties are essential for thermodynamics of the system and non-equilibrium dynamics such as chemical reactions, a large benefit to the studies of quantum chemistry and quantum materials is expected if one may predict the excited states only from the ground state. We note that applying classical machine learning to predict excited states of molecules from classical data (molecular structure, coulomb matrix, etc) has been widely explored in the literature [32–36]. Second, the problem of encoding data to quantum computers mentioned above can be circumvented in this setup; as we will see later, it is possible to input wavefunctions into quantum registers directly from outputs of another quantum algorithm which yields a ground state wavefunction, such as the variational quantum eigensolver (VQE), which is one of the most promising applications of the NISQ devices [37]. Third, from a conceptual point of view, the original 'data' of quantum systems are wavefunctions, which are quantum in nature, so quantum machine learning dealing with quantum data *as they are* will take advantage of the whole information contained in the wavefunctions and potentially has stronger predictability than the classical counterparts which process only classical features of quantum data in a pure classical way [38, 39].
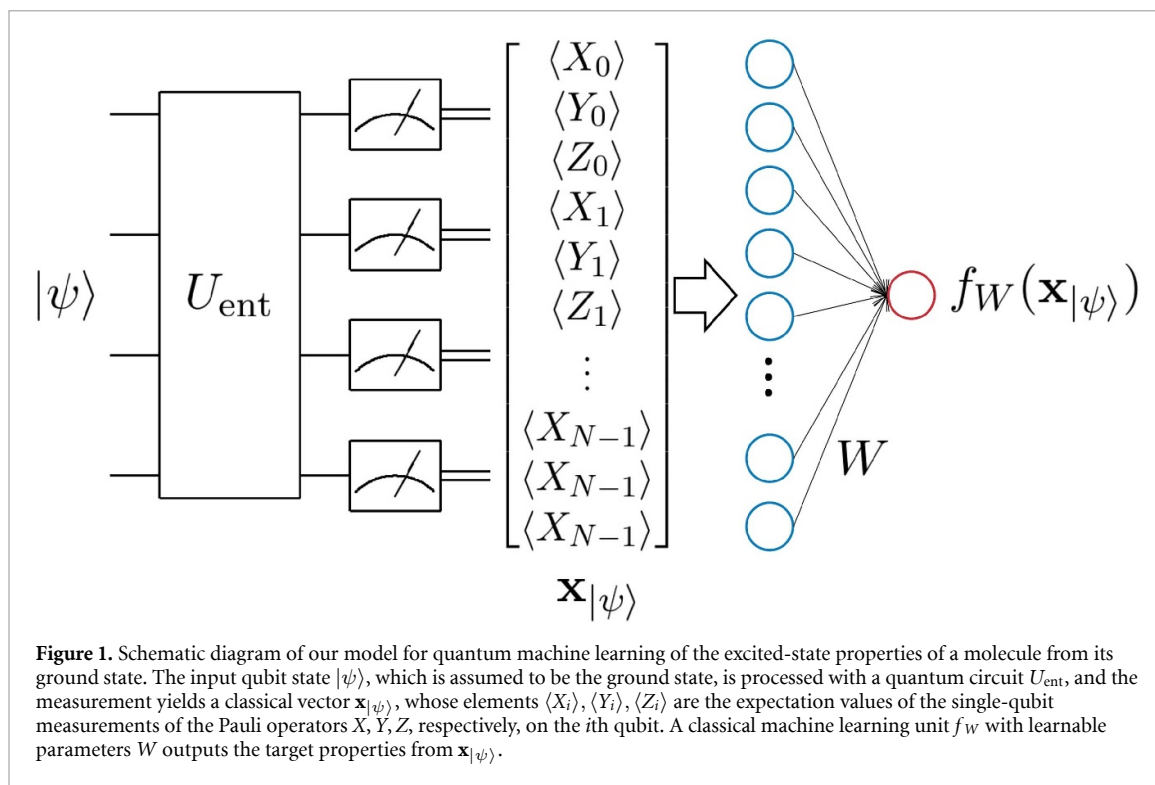
In this study, we propose a simple quantum machine learning scheme to predict the excited-state properties of the Hamiltonian of a given molecule from its ground state wavefunction. Our simulations suggest the potential that one can implement our model on the real NISQ devices being robust to the inevitable noise of outputs on such devices. In particular, we employ and generalize the quantum reservoir computing [40] and quantum reservoir processing [41] techniques.

Both techniques feed the initial quantum information to a random quantum system called a 'quantum reservoir' which evolves the initial state to another state, and the measurement results of the output state are learned by linear regression to predict some properties associated with the initial information.

Similarly, we first process an input wavefunction that is the ground state of the target molecular Hamiltonian with a random quantum circuit or the time evolution under another certain Hamiltonian and then measure the expectation values of one-qubit operators afterwards.

The measurement results are post-processed by a classical machine learning unit, and we train only the classical unit to predict the target properties of the system by supervised learning so that the overall number of runs of quantum computers is small. In the Heisenberg picture, the quantum reservoir effectively transforms the one-qubit operators into complicated multi-qubit ones which contain essential information of the system, and the classical machine learning will decode them appropriately. We numerically demonstrate the predictive power of our scheme by taking three small molecules as examples. Our model can predict the excitation energies and the transition dipole moment between the ground state and the excited state properly only from the ground state wavefunction.

The rest of the paper is organized as follows. In section 2, we explain our setup in detail and propose a model for quantum machine learning of excited states, besides presenting the way to train the model. In section 3, we show the result of numerical simulations of our scheme predicting the excited-state properties of small molecules as examples. Section 4 is dedicated for the discussion of our result. We conclude the study in section 5. Appendix A is a review of the VQE and its extension to find the excited states. Appendix B introduces the Jordan-Wigner transformation, which is used to map fermionic molecular Hamiltonians into

Figure 1. Schematic diagram of our model for quantum machine learning of the excited-state properties of a molecule from its ground state. The input qubit state $|\psi\rangle$, which is assumed to be the ground state, is processed with a quantum circuit U_{ent}, and the measurement yields a classical vector $\mathbf{x}_{|\psi\rangle}$, whose elements $\langle X_i \rangle, \langle Y_i \rangle, \langle Z_i \rangle$ are the expectation values of the single-qubit measurements of the Pauli operators X, Y, Z, respectively, on the ith qubit. A classical machine learning unit f_W with learnable parameters W outputs the target properties from $\mathbf{x}_{|\psi\rangle}$.

Hamiltonians written in qubit operators. Appendix C is the extension of the discussion in section 4 to demonstrate the non-linearity between the excited-state properties of the Hamiltonian and the information one may obtain from the ground state. Appendix D provides a further analysis of the effect of the entangler. The dependence of our scheme on the performance of the VQE is analyzed in Appendix E.

2. Method

In this section, we propose a model for quantum machine learning and explain its training process. The schematic diagram of our model is described in figure 1.

2.1. Model description

Let us consider an N-qubit system and a wavefunction $|\psi\rangle \in \mathbb{C}^{2^N}$ on it. Our learning model proceeds as follows. First, an input N-qubit state $|\psi\rangle$, which is assumed to be the ground state of a given Hamiltonian here, is prepared on a quantum computer and fed into a quantum circuit which is denoted as U_{ent} in figure 1. We call this circuit a quantum entangler or a quantum reservoir for its role of mixing local quantum information of the input state $|\psi\rangle$ and encoding it to the output state $U_{\text{ent}}|\psi\rangle$. U_{ent} is chosen to create enough entanglement in the wavefunction and fixed for each learning task (or an experiment). The details of U_{ent} are not so important for the quality of learning as illustrated by an exactly-solvable model in section 4, so one can use a quantum circuit easy to be realized on real quantum devices. After applying U_{ent}, we measure the expectation values of local Pauli operators $\{X_0, Y_0, Z_0, \cdots, X_{N-1}, Y_{N-1}, Z_{N-1}\}$, where X_i, Y_i, and Z_i represent a Pauli X, Y, Z operator acting on the site i, respectively. Although the total number of operators is 3N, we can measure the operators $X_0, ..., X_{N-1}$ simultaneously since they commute with each other, so can we for the cases of $Y_0, ..., Y_{N-1}$ and $Z_0, ..., Z_{N-1}$.

Hence, one can measure all operators with only three different circuits, i.e. the number of experiments to obtain the measurement data does not scale with the number of qubit N, but only with the desired precision ε as $\mathcal{O}(1/\epsilon^2)$ due to the statistical uncertainty.

After the measurements, we obtain a 3N-dimensional real-valued classical vector:

$$\mathbf{x}_{|\psi\rangle} = (\langle X_0 \rangle, \cdots, \langle Z_{N-1} \rangle)^T = \left(\langle \psi | U_{\text{ent}}^\dagger X_0 U_{\text{ent}} | \psi \rangle, \cdots, \langle \psi | U_{\text{ent}}^\dagger Z_{N-1} U_{\text{ent}} | \psi \rangle \right)^T. \qquad (1)$$

Finally, the classical data $\mathbf{x}_{|\psi\rangle}$ is fed into a classical machine learning unit with learnable parameters W, such as a linear regression model or a neural network, and the prediction $f_W(\mathbf{x}_{|\psi\rangle})$ is obtained.

We have several comments in order. First, the process to obtain $\mathbf{x}_{|\psi\rangle}$ from $|\psi\rangle$ can be viewed as compressing the data of 2^N-dimensional complex-valued vector $|\psi\rangle$ into 3N-dimensional real-valued data.

Although the way of compression is quite complicated due to the entangler U_{ent}, the classical machine learning unit can decode the information in $\mathbf{x}_{|\psi\rangle}$ and use it to predict the properties of the excited states of the Hamiltonian. More concretely, the effect of the entangler is to make the classical vector $\mathbf{x}_{|\psi\rangle}$ to contain the expectation values of complicated (generally long-ranged, many-body) observables for the original ground state $|\psi\rangle$; that is, $\mathbf{x}_{|\psi\rangle}$ can be viewed as the expectation values of the complicated operators $\{U_{\text{ent}}^{\dagger}X_i U_{\text{ent}}, U_{\text{ent}}^{\dagger}Y_i U_{\text{ent}}, U_{\text{ent}}^{\dagger}Z_i U_{\text{ent}}\}_{i=0}^{N-1}$ for the ground state $|\psi\rangle$. When we expand $U_{\text{ent}}^{\dagger}X_i U_{\text{ent}}$ as $U_{\text{ent}}^{\dagger}X_i U_{\text{ent}} = \sum_j \lambda_j^{(X_i)} P_j^{(X_i)}$, where $\lambda_j^{(X_i)}$ is some coefficient and $P_j^{(X_i)}$ is an N-qubit Pauli operator, some of $\{P_j^{(X_i)}\}_j$ are the long-ranged and many-body ones if U_{ent} creates entanglement over the whole system. This means that the classical vector $\mathbf{x}_{|\psi\rangle}$ contains a lot of detailed information of $|\psi\rangle$ as multi-point, long-ranged correlation functions even though we measure only the single-qubit Pauli operators $\bigcup_{i=0}^{N-1}\{X_i, Y_i, Z_i\}$ in reality. Although how such information is implemented in $\mathbf{x}_{|\psi\rangle}$ is not explicitly known since we do not know the actual values of coefficients $\lambda_j^{(X_i)}$, the classical machine learning unit can be trained to utilize the information to predict the excited states. An explicit example of this point is described in section 4 and Appendix C. Moreover, any U_{ent} can be written in the form of a time-evolution operator as e^{-iHT} under a certain Hamiltonian operator H. In this formulation, one can naturally interpret the entangler as an operator evolving the single-qubit Pauli operators $\bigcup_{i=0}^{N-1}\{X_i, Y_i, Z_i\}$ into a linear combination of the multi-qubit ones in the Heisenberg picture, and the linear combination consists of more variety of the multi-qubit Paulis as the evolution takes a longer time (see the details in Appendix D). Second, the model is identical to quantum reservoir computing proposed in reference [40] and quantum reservoir processing proposed in reference [41] if we choose the linear regression as the classical machine learning unit in the model. One may also consider using general classical models such as the neural network, the Gaussian process regression, etc. Even though the numerical simulations we carried out in this study leverage only a linear model, which actually gives sufficiently accurate predictions at least for the molecules we consider here, non-linearity in the classical machine learning unit may be necessary to predict the excited states for certain tasks as discussed in section 4 using an exactly-solvable toy model for the hydrogen molecule. Third, as mentioned in the previous section, we stress that this scheme is very suitable to be combined with the VQE. The VQE finds a quantum circuit that produces an approximate ground state of a given Hamiltonian by using the variational principle and has been extended to obtain the excited states recently [42–49]. Since it can handle Hamiltonians of large systems that are intractable by classical computers, the VQE is considered as one of the best approaches to utilize the NISQ devices for real-world problems. In our quantum machine learning model, one can use the quantum circuit obtained by the VQE to make an input state (approximate ground state wavefunction) for the training and the prediction of our model. There is no overhead cost at all to feed target data to the learning model in this case (see also reference [50]). We review the VQE and one of its extensions to compute the excited states in Appendix A.

2.2. Supervised learning of the model

Next, we explain the procedure for supervised learning of our model. First, we define the training set $\mathcal{R}$ whose elements $r \in \mathcal{R}$ are a set of characteristics of a molecule (e.g. name of a molecule and its atomic configuration), and we prepare the data $\{|\psi_0(r)\rangle, \mathbf{y}(r)\}_{r \in \mathcal{R}}$ for training. In the case of predicting excited states of a given Hamiltonian from its ground state, $|\psi_0(r)\rangle$ is the ground state of the molecular Hamiltonian $H(r)$ and $\mathbf{y}(r)$ contains the target properties of the excited states of $H(r)$, such as excitation energies. Next, by using the training set, the classical machine learning unit f_W is trained to predict $\{\mathbf{y}(r)\}_{r \in \mathcal{R}}$ from the classical vectors $\{\mathbf{x}_{|\psi_0(r)\rangle}\}_{r \in \mathcal{R}}$ which are calculated in the way described in the previous subsection. A typical training algorithm for the supervised learning is to minimize a cost function defined to measure the deviations of the prediction $\{f_W(\mathbf{x}_{|\psi_0(r)\rangle})\}_{r \in \mathcal{R}}$ from the training data $\{\mathbf{y}(r)\}_{r \in \mathcal{R}}$ by tuning W.

We note that our model is easier to be trained and less costly in terms of the number of runs of quantum computers compared with the so-called 'quantum circuit learning' where parameters of the quantum circuit are optimized [18–21] since once the classical representation of the quantum state $\{\mathbf{x}_{|\psi_0(r)\rangle}\}_{r \in \mathcal{R}}$ is obtained, there is no need to run the quantum device afterwards for training the model.

3. Numerical demonstration for small molecules

In this section, we numerically demonstrate the ability of our model to reproduce excited-state properties from the ground state wavefunctions by taking small molecules as examples. We consider three types of molecules: LiH molecule, H_4 molecule whose hydrogen atoms are aligned linearly with equal spacing, and H_4 molecules whose hydrogen atoms are placed in a rectangle shape. We call them as LiH, H_4 (line), H_4 (rectangle), respectively. We evaluate our model in two situations, in one of which ideal outputs of the quantum circuits are available (*noiseless*), and inevitable noise in the real NISQ devices is considered in the

other (*noisy*). The electronic ground states of those molecules with various atomic geometries are prepared by diagonalizing the Hamiltonian for the noiseless simulation and by numerically simulating the VQE for the noisy simulation. Then, we train our model with the linear regression as its classical machine learning unit to predict the first and second excitation energies and the transition dipole moment among them whose values are obtained by exactly solving the Hamiltonian. Numerical results show that our model can properly reproduce the excited states and illustrate the predictive power of our model.

3.1. Dataset

To prepare a dataset for the simulations, we consider the electronic Hamiltonians of the following configurations. For LiH molecule and H_4 (line), the atomic distances are in the range of $[0.5\text{ Å}, 3.3\text{ Å}]$. For H_4 (rectangle), we choose the two spacing of atoms (lengths of two edges) in $[0.5\text{ Å}, 2.0\text{ Å}] \times [0.5\text{ Å}, 2.0\text{ Å}]$. We perform the standard Hartree–Fock calculation by employing the STO-3G minimal basis and construct the fermionic second-quantized Hamiltonian for all of the molecules and configurations [51, 52] with open-source libraries PySCF [53] and OpenFermion [54]. Two Hartree–Fock orbitals with the highest and the second-highest energies among six orbitals of LiH molecule are removed by assuming they are vacant because they are composed almost completely from $2p_x$ and $2p_y$ atomic orbitals of LiH and do not significantly contribute to the binding energy of LiH. Then the Hamiltonian is mapped to the sum of the Pauli operators by the Jordan-Wigner transformation [55] which we denote $H(r)$ (a review of the Jordan-Wigner transformation is given in Appendix B). Then, the electric Hamiltonians for all of the molecules turn into 8-qubit Hamiltonians.

The training and test datasets for the simulations are prepared for each Hamiltonian $H(r)$ in the following way. First, in the case of the noiseless simulation, the ground state of $H(r)$ is prepared by the exact diagonalization. In the case of the noisy simulation, the VQE algorithm is applied to $H(r)$, and the approximate ground state is obtained as $|\tilde{\psi}_0(r)\rangle = U(\vec{\theta})|0\rangle$. Here $U(\vec{\theta})$ is a variational quantum circuit (ansatz) with classical parameters $\vec{\theta}$ and $|0\rangle$ is a reference Hartree-Fock state. We adapt the unitary coupled-cluster singles and doubles ansatz [37, 56] as $U(\vec{\theta})$. Next, we compute the quantities of the excited-state properties to be predicted,

$$\mathbf{y}(r) = (\Delta E_1(r), \Delta E_2(r), \|\boldsymbol{\mu}_{\text{eg}}(r)\|)^T, \tag{2}$$

where $\Delta E_{1(2)}(r) = E_{1(2)}(r) - E_0(r)$ is the first (second) excitation energy of $H(r)$ in the sector of neutral charge, where $E_{0,1,2}(r)$ are three lowest eigenenergies of $H(r)$ in the same sector ignoring degeneracy. For our choice of the molecules and configurations, $E_0(r)$ is the energy of the spin-singlet ground state S_0, and $E_1(r)$ is the energy of the spin-triplet excited state T_1. $E_2(r)$ is the energy of the spin-singlet excited state S_1 or the spin-triplet excited state T_2 depending on the configurations of the molecule. The transition dipole moment between the ground state and the excited state $\boldsymbol{\mu}_{\text{eg}}(r)$ is defined as

$$\boldsymbol{\mu}_{\text{eg}}(r) = \langle\psi_0(r)|\boldsymbol{\mu}|\psi_{\text{ex}}(r)\rangle, \tag{3}$$

where $|\psi_0(r)\rangle$ is the exact ground state of $H(r)$ (the singlet state S_0), $|\psi_{\text{ex}}(r)\rangle$ is the exact excited state of $H(r)$ which has the lowest energy among those having a non-zero transition dipole moment from the ground state (typically S_1 state), and $\boldsymbol{\mu} = -e(\hat{x}, \hat{y}, \hat{z})^T$ is the dipole moment operator with electronic charge e. In this study, we use its L2-norm $\|\boldsymbol{\mu}_{\text{eg}}(r)\|$ for the learning tasks. The calculation of each value of $\mathbf{y}(r)$ is performed by the exact diagonalization of $H(r)$ for both of the noiseless and noisy simulations. To stabilize the learning process, we scale those calculated values to fit them into the $[-1, 1]$ range, so that the maximum value $y_{\text{max}}^{(k)}$ and the minimum value $y_{\text{min}}^{(k)}$ in the training dataset are scaled as $y_{\text{max}}^{(k)} = \max_{r \in \mathcal{R}} y^{(k)}(r) \to 1$ and $y_{\text{min}}^{(k)} = \min_{r \in \mathcal{R}} y^{(k)}(r) \to -1$ where $y(r)^{(k)}$ denotes the k-th element of $\mathbf{y}(r)$ for each $k = 1, 2, 3$, and other values, including those in the test dataset, are mapped as $y(r)^{(k)} \to 2\frac{y(r)^{(k)} - y_{\text{max}}^{(k)}}{y(r)_{\text{max}}^{(k)} - y(r)_{\text{min}}^{(k)}} - 1$.

For the numerical experiments, we randomly split those obtained data $\{|\tilde{\psi}_0(r)\rangle, \mathbf{y}(r)\}_r$ into the training set and the test set for the evaluation of the model. We used 30 training data points and 50 test points, respectively, for the tasks of the LiH and H_4 (linear) molecules, and 250 training data points and 1250 test data points for the H_4 (rectangle) molecules.

3.2. Model for the simulations

The entangler U_{ent} in the model is chosen to be the time-evolution operator $e^{-iH_{\text{TFIM}}T}$ under the random transverse-field Ising model (TFIM),

$$H_{\text{TFIM}} = \sum_{i,j=0}^{N-1} J_{ij} Z_i Z_j + \sum_{i=0}^{N-1} h_i X_i, \tag{4}$$

where X_i and Z_j are Pauli operators acting on the site i, j-th qubit, coefficients h_i and J_{ij} are sampled from the Gaussian distributions $N(1, 0.1)$ and $N(0.75, 0.1)$, respectively, and we set $T = 10$. These coefficients are fixed during each of the numerical simulations. This type of the entangler can be implemented on various types of the NISQ devices; for example, in the case of superconducting qubits, it can be realized by a sequence of the cross resonance gates [57, 58] or simply tuning the resonance frequency of the qubits [18]. We note that a similar kind of the quantum reservoir has recently been implemented on a real NISQ device [59].

In our numerical simulations, this time evolution is exactly simulated as the unitary operation $e^{-iH_{\text{TFIM}}T}$ acting on the input state.

For the classical machine learning unit for the numerical demonstration, we employ the linear regression (LR) [60]. Although the LR does not have non-linearity which is in principle necessary to compute the excited-state properties (see section 4), it performs well enough for the molecular Hamiltonians we consider for the simulations as shown in section 3.4, so it serves as a nice demonstrative model to evaluate the concept of our model.

The output function of the LR is

$$f^{(k)}\left(\mathbf{x}_{|\psi\rangle}\right) = \mathbf{w}_{\text{out}}^{(k)} \cdot \mathbf{x}_{|\psi\rangle}, \tag{5}$$

where $\mathbf{w}_{\text{out}}^{(k)}$ is a 3N-dimensional vector, or parameters of the model, to be optimized, and $k = 0, 1, 2$ corresponds to the component of the prediction for $\mathbf{y} = (y^{(0)}, y^{(1)}, y^{(2)})^T$. The model is trained to minimize the mean squared error (MSE) cost function

$$L_{\text{LR}}(\{\mathbf{w}_{\text{out}}^{(k)}\}) = \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} \left| \mathbf{w}_{\text{out}}^{(k)} \cdot \mathbf{x}_{|\psi_0(r)\rangle} - y(r)^{(k)} \right|^2, \tag{6}$$

where $\mathcal{R}$ represents the training dataset, respectively, for each target property. The exact optimum of the cost function can be obtained as

$$\mathbf{w}_{\text{out}}^{(k)*} = \left(V^T V\right)^{-1} V^T \mathbf{Y}^{(k)}, \tag{7}$$

where V is a $|\mathcal{R}| \times 3N$ dimensional matrix whose ith row is $\mathbf{x}_{|\psi_0(r_i)\rangle}^T$, and $\mathbf{Y}^{(k)}$ is a $|\mathcal{R}|$ dimensional column vector whose ith component is $y(r_i)^{(k)}$, where r_i is the ith element of $\mathcal{R}$. The whole classical process requires the computational complexity of $\mathcal{O}(N^3 + N^2 |\mathcal{R}|)$.

3.3. Simulation of quantum circuits

To check the practical advantage of our model with the NISQ devices, we numerically simulate quantum circuits of the model considering the noiseless and noisy situations (including preparation of the ground state wavefunctions by the VQE for the noisy case). The latter reflects a more realistic situation of experiments on a real NISQ device, but we stress that the former still serves as a reference point to judge whether the model has the capability of performing the learning task or not.

In the noiseless simulation, the expectation value of the Pauli operator $\langle \psi | P_i | \psi \rangle$, where the $|\psi\rangle$ is a quantum state and P_i is the Pauli operator acting on ith qubit, is estimated exactly by calculating the inner product. In the noisy simulation, we consider two error sources which make estimations of those expectation values deviate from the exact ones. One of them is a sequence of the depolarizing noise channels [61] that transform the quantum state $\rho = U_{\text{ent}} |\psi\rangle \langle \psi | U_{\text{ent}}^\dagger$ from the reservoir into $\rho' = \mathcal{E}_{N-1} \circ \ldots \circ \mathcal{E}_0(\rho)$, where $\mathcal{E}_i$ is the depolarizing channel that acts on ith qubit as $\mathcal{E}_i(\sigma) = (1 - p)\sigma + \frac{p}{3}(X_i \sigma X_i + Y_i \sigma Y_i + Z_i \sigma Z_i)$. We take $p = 0.01$ in the simulations. The other source is the so-called shot noise that stems from the finite number of shots for the projective measurements of the Pauli operator P_i. Each measurement returns ± 1 according to the probability distribution determined by the exact values of $\text{Tr}(\rho' P_i)$. We sample 10^4 shots of measurements to compute the ground state with the VQE, and 10^6 shots for each Pauli operator to construct the vector in equation (1). These are feasible numbers in experiments [57].

3.4. Results

The model described in the previous subsections is trained by the training dataset and evaluated by the test set. The evaluation is performed based on the mean absolute error (MAE) for the test set $\mathcal{T}$,

$$C_{\mathcal{T}} = \frac{1}{3|\mathcal{T}|} \sum_{r \in \mathcal{T}} \sum_{k=0}^{2} \left| \mathbf{f}_W(\mathbf{x}_{|\psi_0(r)\rangle}) - \mathbf{y}(r)^{(k)} \right|, \tag{8}$$

where $\mathbf{f}_W(\mathbf{x})$ is the output of the model considered as a vector. We train and evaluate the model for each molecule separately.

Table 1. The individual MAEs estimating the properties ΔE_1 and ΔE_2 in the units of Hartree, compared with the chemical accuracy 1.6mHa for both of the noisy and the noiseless cases.

		LiH	H4 (line)	H4 (rectangle)	Chemical Accuracy
MAE (noiseless) (mHa)	ΔE_1	0.1	1.4	21.3	1.6
	ΔE_2	0.1	3.2	28.0	
MAE (noisy) (mHa)	ΔE_1	15.8	39.2	109.1	
	ΔE_2	15.4	50.3	103.8	

Table 2. The MAEs evaluated with the test set for the trained models in the noiseless situation with and without the entangler. The mean values are taken over all the three properties to be estimated. The output values and the excited-state properties are in the same standardized scale as explained in section 3.1. The MAEs for the random guess are also presented as a reference.

	LiH	H4 (line)	H4 (rectangle)
Test MAE with entangler U_{ent}	0.018 1	0.020 3	0.083 6
Test MAE without entangler U_{ent}	0.172	0.324	0.300
Random Guess	0.673	0.544	0.444

3.4.1. Noiseless simulation

The prediction results by the trained model in the noiseless numerical simulation are shown in figures 2 and 3. In these figures, the excited-state properties $\mathbf{y}(r)$ are scaled back to the original scale. Our model obviously reproduces the exact values of the excited-state properties $\mathbf{y}(r)$ for all of the three molecule types.

To quantify it, in the upper rows of table 1, we summarize the MAEs of ΔE_1 and ΔE_2 for the test data, equation (8), between the predictions and the exact values scaled back to the original scale. For the LiH and H_4 (linear) molecules, the MAEs are below or in the comparable scale to the chemical accuracy 1.6×10^{-3}Ha. The error is larger than the chemical accuracy for H_4 (rectangle), and it is probably because the degrees of freedom of the molecular structure of H_4 (rectangle) are larger, and the LR model may not have a sufficient expressive space. Utilizing other machine learning methods (e.g. neural networks) is one possible way to achieve more accurate results.

Also, we investigate the necessity of the entangler U_{ent} by comparing the values of the MAE (equation (8)) for the test set after the training. The evaluations of the learners with and without the entangler are summarized in table 2, indicating that the entangler significantly enhances the predictive power of the model. To treat all the excited-state properties on an equal footing, here we use the scaled values of $\mathbf{y}(r)$. In section 4 and Appendix C, another supporting result for the necessity of the entangler is presented by using an exactly solvable model for the hydrogen molecule.

These results from the noiseless simulations illustrate the predictive power of our model for the difficult task to predict the excited-state properties only from the ground state.

3.4.2. Noisy simulation

In order to evaluate our scheme in a realistic situation with a quantum device, we add two noise sources to the simulation as described in section 3.3. In this case, to enhance the noise robustness, we make two modifications to the noiseless case as follows. First, after obtaining the classical vectors $\{\mathbf{x}_{|\psi_0(r)\rangle}\}_{r \in \mathcal{R}}$ by processing the training dataset with the noisy quantum circuit, we make 100 copies of every vector $\mathbf{x}_{|\psi_0(r)\rangle}$ on a classical computer, and add a gaussian noise $N(0, 2 \times 10^{-3})$ to each component of it. We stack these vectors, and now we have a new $100|\mathcal{R}| \times 3N$ dimensional matrix V'. The vector $\mathbf{Y}^{(k)}$ is also duplicated 100 times to match up with V' (let us call this new vector $\mathbf{Y}'^{(k)}$ for later use). Notice that this modification does not affect the required number of measurements of the quantum circuit. Second, we add the L2 regularization term into the cost function of the LR, particularly saying the cost function becomes
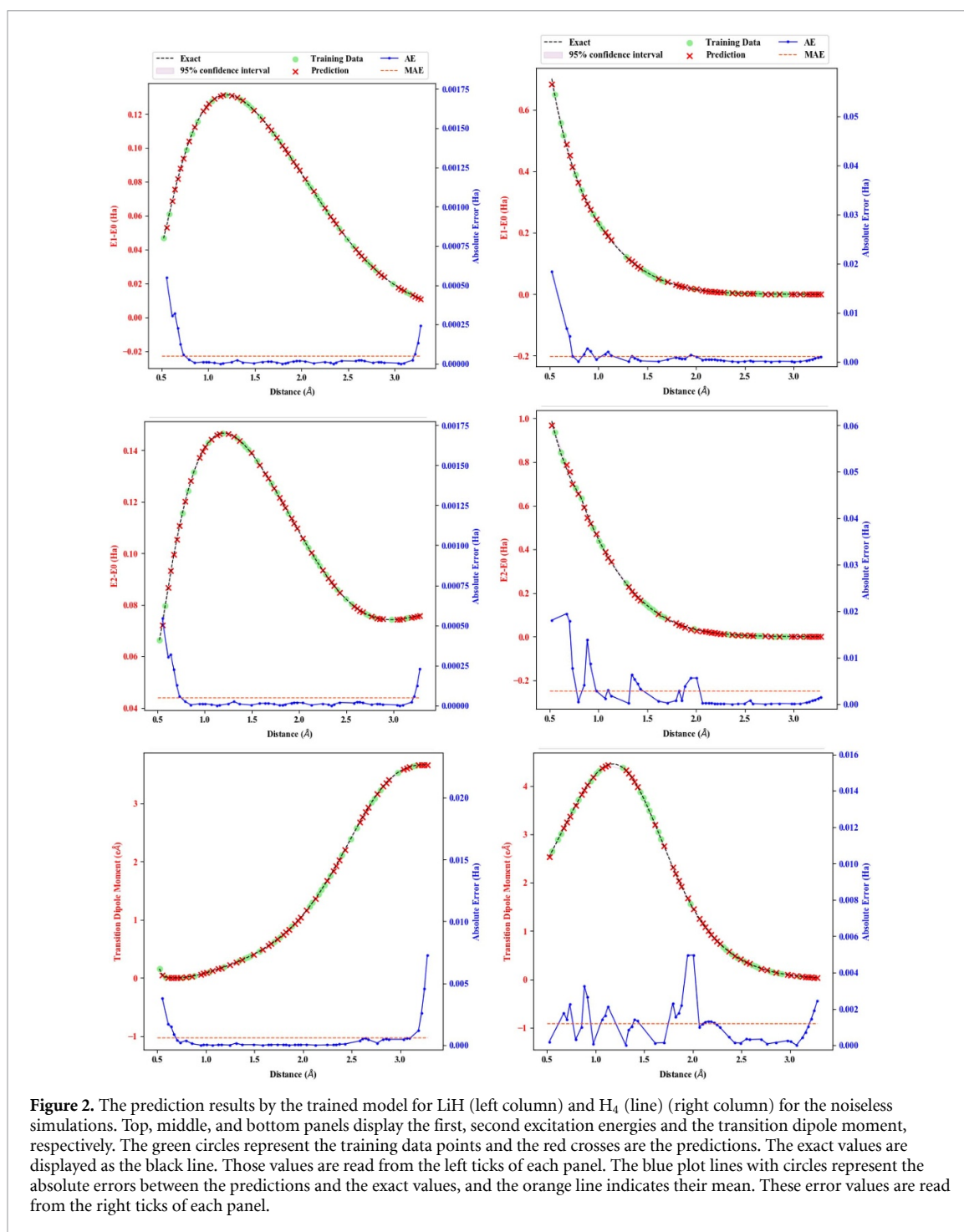
$$L'_{\mathrm{LR}}(\{\mathbf{w}_{\mathrm{out}}^{(k)}\}) = L_{\mathrm{LR}}(\{\mathbf{w}_{\mathrm{out}}^{(k)}\}) + \alpha \|\mathbf{w}_{\mathrm{out}}^{(k)}\|^2, \qquad (9)$$

where we used $\alpha = 10^{-3}$ in the simulations. We may obtain the exact optimum by computing

$$\mathbf{w}_{\mathrm{out}}^{(k)*} = (V'^T V' + \alpha I)^{-1} V'^T \mathbf{Y}'^{(k)}, \qquad (10)$$

where I is an identity matrix of $3N \times 3N$ dimensions. Both of these two modifications work as regularizations preventing the model from overfitting due to the outliers with large noises.

The prediction results are presented in figures 4 and 5. We see that the model still predicts $\mathbf{y}(r)$ well even in this noisy case.

Figure 2. The prediction results by the trained model for LiH (left column) and H_4 (line) (right column) for the noiseless simulations. Top, middle, and bottom panels display the first, second excitation energies and the transition dipole moment, respectively. The green circles represent the training data points and the red crosses are the predictions. The exact values are displayed as the black line. Those values are read from the left ticks of each panel. The blue plot lines with circles represent the absolute errors between the predictions and the exact values, and the orange line indicates their mean. These error values are read from the right ticks of each panel.

We attribute this noise-robustness to the regularization technique of the LR in equation (9). The MAEs for the predictions of ΔE_1 and ΔE_2 are summarized in the lower rows of table 1 in the same way as the noiseless cases. The noise makes the accuracy of the predictions worse than those of the noiseless cases, and all of the errors become larger than the chemical accuracy. A part of the reason for this is because the noise hinders obtaining sufficiently precise ground states of the molecular Hamiltonians. Indeed, for example in the case of LiH, we find that the ground-state energies computed by the VQE in the noisy situation already have a larger error (0.007 4 Ha) than the chemical accuracy, and the MAEs of the predictions (~ 0.015 Ha) are in the similar order. Adapting the error mitigation techniques [62, 63] to the VQE can remove the effect of the noise and will yield more accurate results even in the noisy situation. In Appendix E, we present how the accuracy of the predictions for the excited-state properties varies as the function of the number of the shots used to perform the VQE.
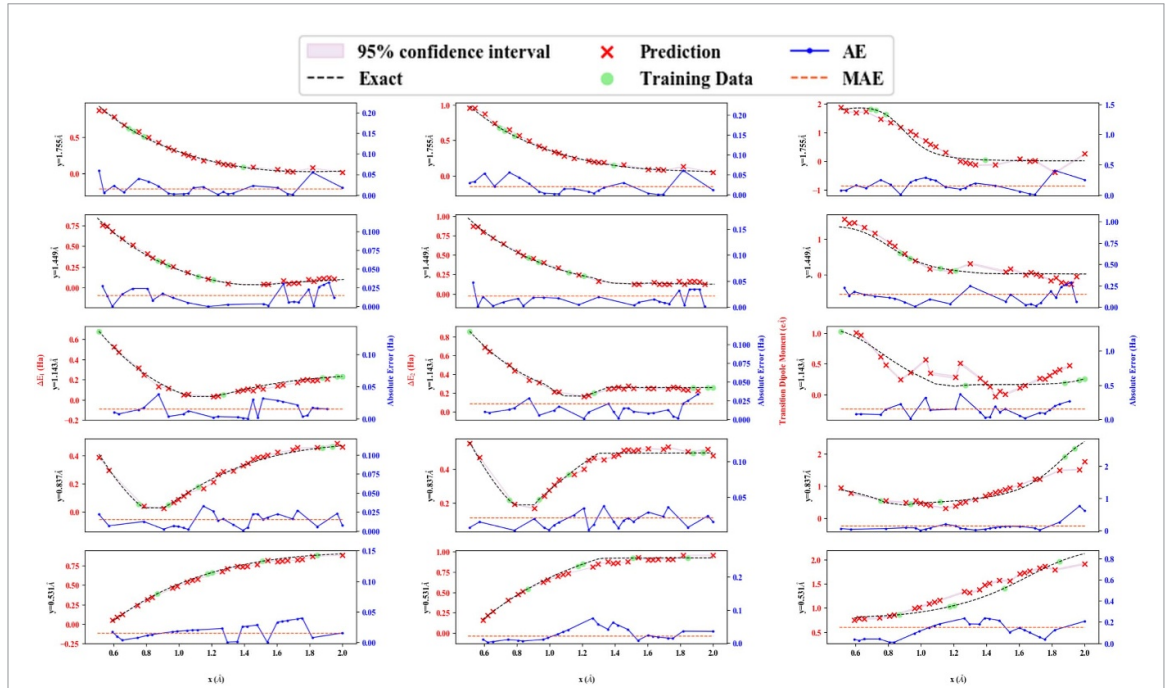
Figure 3. The same figures as figure 2 for H_4 (rectangle). x and y denote the two atomic spacings of the rectangular geometry.

4. Discussion

4.1. Necessity of the entangler U_{ent} and the non-linearity of f_W

Here we discuss the necessity of the entangler U_{ent} and the non-linearity in the classical machine learning unit f_W by considering an exactly solvable model of fermions, namely, the 2-site fermion Hubbard model at half-filling [64].

The 2-site Hubbard model is defined as

$$H_{\text{Hub}}(U) = -\sum_{\sigma=\uparrow,\downarrow} \left(c_{0,\sigma}^\dagger c_{1,\sigma} + \text{h.c.} \right) + U \sum_{i=0,1} n_{i,\uparrow} n_{i,\downarrow} \tag{11}$$

where $c_{i,\sigma}, c_{i,\sigma}^\dagger$ are fermionic annihilation and creation operators acting on an electron with spin $\sigma = \uparrow, \downarrow$ located at ith site ($i = 0, 1$), and $n_{i,\sigma} = c_{i,\sigma}^\dagger c_{i,\sigma}$ is the number operator of an electron with spin σ at ith site. The parameter $U > 0$ determines the strength of electron repulsion. This system can be considered as a simplified model of a hydrogen molecule whereas it also serves as a prototype of strongly-correlated materials. When we restrict ourselves into the sector where the number of electrons is two, i.e. the neutral hydrogen states, the first, second excitation energies and the transition dipole moment are
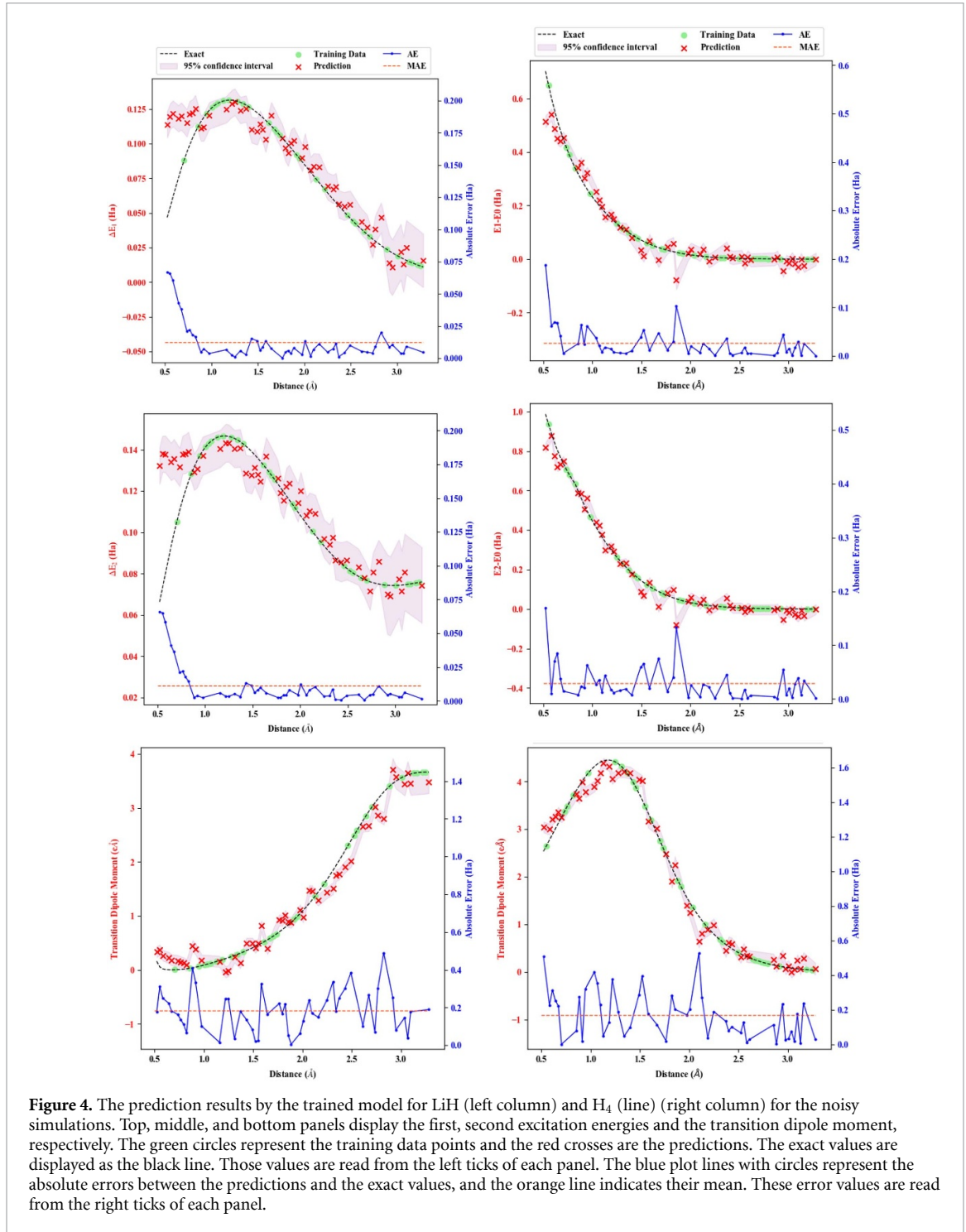
$$\Delta E_1 = \frac{U}{2} \left(-1 + \sqrt{1 + \frac{16}{U^2}} \right), \tag{12}$$

$$\Delta E_2 = \frac{U}{2} \left(1 + \sqrt{1 + \frac{16}{U^2}} \right), \tag{13}$$

$$\|\boldsymbol{\mu}_{\text{eg}}\| = \sqrt{\frac{1 - (1 + 16/U^2)^{-1/2}}{2}}, \tag{14}$$

respectively. We note that the dipole moment operator is defined as $\boldsymbol{\mu} = \frac{1}{2}(n_{1,\uparrow} + n_{1,\downarrow} - n_{0,\uparrow} - n_{0,\downarrow})$.

Applying the Jordan-Wigner transformation [55] to the system (equation 11), we obtain the 4-qubit Hamiltonian $H_{\text{qubit}}(U)$. We denote the ground state of $H_{\text{qubit}}(U)$ as $|\psi_0(U)\rangle$. When there is no entangler, the classical vector $\mathbf{x}_{|\psi_0(U)\rangle}$ is trivial because

Figure 4. The prediction results by the trained model for LiH (left column) and H$_4$ (line) (right column) for the noisy simulations. Top, middle, and bottom panels display the first, second excitation energies and the transition dipole moment, respectively. The green circles represent the training data points and the red crosses are the predictions. The exact values are displayed as the black line. Those values are read from the left ticks of each panel. The blue plot lines with circles represent the absolute errors between the predictions and the exact values, and the orange line indicates their mean. These error values are read from the right ticks of each panel.

$$\langle X_j \rangle_{\mathrm{GS}} = \langle Y_j \rangle_{\mathrm{GS}} = \langle Z_j \rangle_{\mathrm{GS}} = 0 \tag{15}$$

holds for all qubit sites $j = 0, 1, 2, 3$, where we define $\langle \ldots \rangle_{\mathrm{GS}} = \langle \psi_0(U) | \ldots | \psi_0(U) \rangle$. In contrast, when there is an entangler U_{ent} in our model, it converts the Pauli operators X_j, Y_j, Z_j into a sum of more complicated Pauli strings as $U_{\mathrm{ent}}^\dagger Z_0 U_{\mathrm{ent}} = Z_0 Z_1 + 0.2 Z_1 X_1 Z_2 Y_2 + \ldots$ in the Heisenberg picture. Then, the classical vector $\mathbf{x}_{|\psi_0(U)\rangle}$ contains contributions from the terms like $\langle Z_0 Z_1 \rangle_{\mathrm{GS}}$. It follows that

$$\langle Z_0 Z_1 \rangle_{\mathrm{GS}} = - \left(1 + \frac{16}{U^2} \right)^{-1/2}, \tag{16}$$

and

$$\Delta E_1 = \frac{2 \left(1 + \langle Z_0 Z_1 \rangle_{\mathrm{GS}} \right)}{\sqrt{1 - \langle Z_0 Z_1 \rangle_{\mathrm{GS}}^2}}, \tag{17}$$
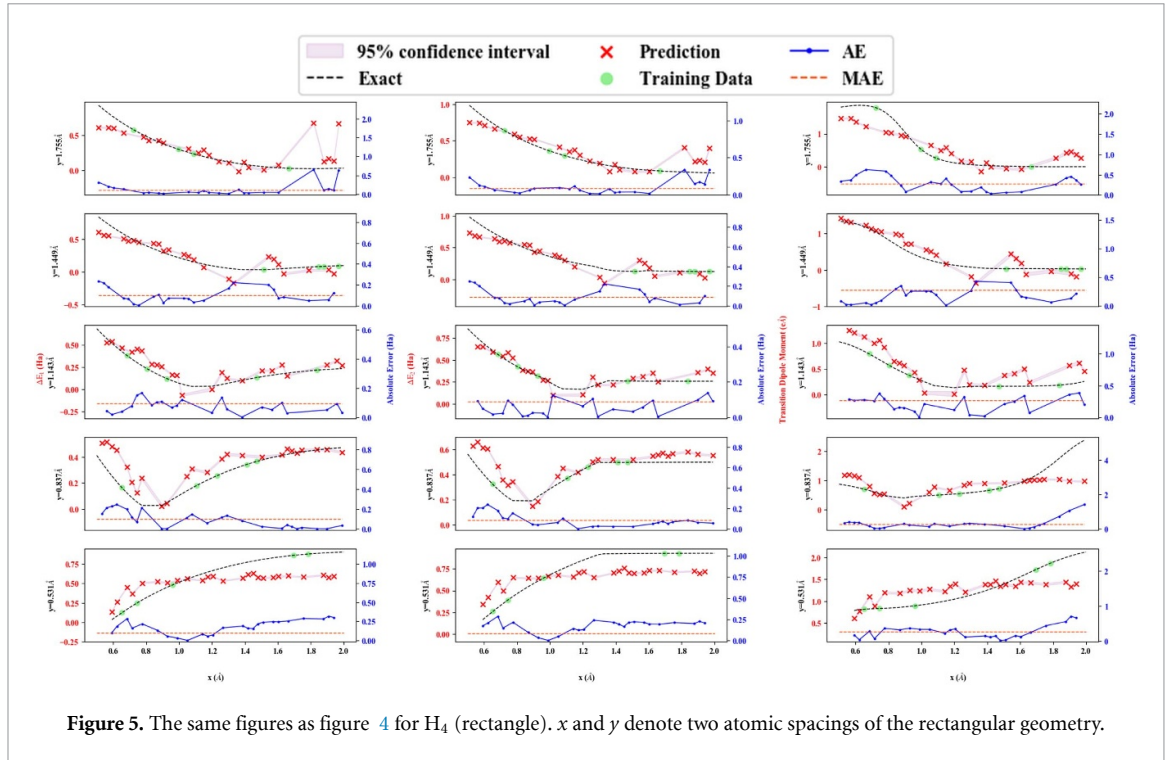
Figure 5. The same figures as figure 4 for H_4 (rectangle). x and y denote two atomic spacings of the rectangular geometry.

$$\Delta E_2 = \frac{2\left(1 - \langle Z_0 Z_1 \rangle_{\mathrm{GS}}\right)}{\sqrt{1 - \langle Z_0 Z_1 \rangle_{\mathrm{GS}}^2}}, \tag{18}$$

$$\|\boldsymbol{\mu}_{\mathrm{eg}}\| = \sqrt{\frac{1 + \langle Z_0 Z_1 \rangle_{\mathrm{GS}}}{2}}. \tag{19}$$

These equations indicate that the excitation energies can be predicted by utilizing the values of $\langle Z_0 Z_1 \rangle_{\mathrm{GS}}$ appropriately. Therefore, one can see that it is possible to predict the excitation energies from the classical vector $\mathbf{x}_{|\psi_0(U)\rangle}$ if the ground state vector is processed by an entangler, and the classical machine learning unit f_W has enough non-linearity. These equations also imply that the details of the entangler, which determine coefficients of the terms like $\langle Z_0 Z_1 \rangle_{\mathrm{GS}}$ in $\mathbf{x}_{|\psi_0(U)\rangle}$, is not so important for predictions; the classical machine learning unit can compensate the difference of such coefficients. In the appendix, we provide further analysis of the 2-site Hubbard model including the necessity of non-linearity.

4.2. Generalizablity
In the numerical simulations in section 3, the models are trained and evaluated for each molecule separately. The generalizability of our model to predict the properties of various molecules simultaneously is one possibility of our model for future extensions.

To make our model more powerful and capable of taking various molecules as inputs, several modifications can be considered. First, including the ground state energy which can also be calculated by the VQE besides the ground state in the input of the classical machine learning unit f_W will be necessary since otherwise, one may not determine the energy scale of an input molecule. Second, replacing the entangler U_{ent} with a parametrized quantum circuit $V(\theta)$ and optimizing the circuit parameters θ along with the classical machine learning unit increase the degree of freedoms of the model and may result in a better predictive power, with a possible drawback that the number of required experiments on the NISQ devices would increase in the training step. Exploring these ideas is an interesting future direction of the work.

5. Conclusion

In this study, we introduce a new quantum machine learning framework for predicting the excited-state properties of a molecule from its ground state wavefunction. By employing the quantum reservoir and choosing simple one-qubit observables for measurements accompanied by post-processing with classical

machine learning, one may process our framework easily on the NISQ devices requiring the realistic number of runs of them. The numerical simulations with and without the noise in outputs of quantum circuits demonstrate that our model accurately predicts the excited states. Although our framework is tested only with small molecules to illustrate its potential in the numerical simulations, we expect that it will benefit the calculation of excited states of larger molecules by reducing the computational cost from calculating exact solutions. Our result opens up the further possibility to utilize the NISQ devices in the study of quantum chemistry and quantum material fields.

Acknowledgment

Appendix A. Variational Quantum Eigensolver (VQE) and its extension to the excited states

In this appendix, we first review the VQE algorithm [37] which finds the ground state of a given Hamiltonian by using the near-term quantum computers. We use it to prepare the ground states of the molecular Hamiltonians considering the realistic noisy situation in section 3. Next, to give the readers an insight on how costly it is to find the excited states of a given Hamiltonian on the near-term quantum computers compared with the computations for the ground states, we review the subspace-search VQE (SSVQE) algorithm [44] as one example of such algorithms.

A.1. VQE algorithm

The VQE tries to compute the minimum eigenvalue and its corresponding eigenstate of a given observable H by minimizing the expectation value of H with the ansatz state $|\psi(\theta)\rangle = U(\vec{\theta})|0\rangle$, where $U(\vec{\theta})$ is the parameterized unitary circuit on a quantum computer with classical parameters $\vec{\theta}$ and $|0\rangle$ is some reference state. When the expectation value $E(\vec{\theta}) = \langle\psi(\vec{\theta})|H|\psi(\vec{\theta})\rangle$ reaches the minimum at $\vec{\theta}_{\mathrm{opt}}$ by optimizing the parameters $\vec{\theta}$, $E(\vec{\theta}_{\mathrm{opt}})$ and $|\psi(\vec{\theta}_{\mathrm{opt}})\rangle$ are the closest approximation of the lowest eigenvalue and its corresponding eigenstate, respectively. Evaluation of $E(\vec{\theta})$ for a given $\vec{\theta}$ is performed by the near-term quantum computers, and one uses a classical optimization algorithm to iteratively update the values of $\vec{\theta}$ to find the minimum. This classical-quantum hybrid architecture of the VQE algorithm requires less computational/experimental abilities for quantum computers than the long-term, pure-quantum algorithms such as the phase estimation, so that one may run it on the near-term quantum computers.

When applying the VQE to the molecular Hamiltonian, first we prepare the observable H as the second-quantized Hamiltonian of a given molecule by using the finite number of orbitals. Typically, the Hartree–Fock molecular orbitals are used for the second-quantization and each spin orbital corresponds one qubit [51, 52]. Since the second-quantized Hamiltonian is written in fermionic operators while quantum computers can handle with qubit operators $P \in \{I, X, Y, Z\}^{\otimes N}$ only, it is then mapped into the linear combination of qubit operators, $H \to \sum_P h_P P$ where h_P is a coefficient corresponding to the operator P. One example of such the fermion-spin mapping is the Jordan-Wigner transformation which is reviewed in B.

A.2. Subspace-search variational quantum eigensolver (SSVQE) for excited states

Here, we also review the SSVQE algorithm [44], which is one of the algorithms to find the eigenstates corresponding to the higher eigenvalues of an observable H on the near-term quantum computers. Suppose we would like to find the k lowest eigenvalues and eigenstates of H. The SSVQE employs the k reference states $\{|\phi_i\rangle\}_{i=1}^k$ which are mutually orthogonal and prepares the k ansatz states with a parameterized unitary circuit $U(\vec{\theta})$ as $\{|\psi_i(\vec{\theta})\rangle = U(\vec{\theta})|\phi_i\rangle\}_{i=1}^k$. It was shown in [44] that when the following cost function

$$C(\vec{\theta}) = \sum_{i=1}^k w_i \langle\psi_i(\vec{\theta})|H|\psi_i(\vec{\theta})\rangle \tag{A1}$$

takes the minimum at $\vec{\theta}_{\mathrm{opt}}$ for appropriate weights $\{w_i\}_{i=1}^k$ satisfying $i < j \Rightarrow w_i > w_j$, ith eigenvalue and eigenstate are approximated by $\langle\psi_i(\vec{\theta})|H|\psi_i(\vec{\theta})\rangle$ and $|\psi_i(\vec{\theta})\rangle$, respectively. Compared with the VQE for the

ground state, evaluating the cost function of the SSVQE takes more computational cost (runs of quantum circuits and measurements) by k times because one need to evaluate $\langle \psi_i(\vec{\theta}) | H | \psi_i(\vec{\theta}) \rangle$ for each $i = 1, \ldots, k$ separately and combine them. Moreover, the parameterized unitary circuit must be deeper to express the excited states because they are generally more entangled than the ground state. To implement the deeper unitary circuit, the fidelity required for the near-term quantum computers is tougher than that for the VQE, and more parameters need to be optimized so it will take longer for the cost function to converge.

Appendix B. Jordan-Wigner transformation

The Jordan-Wigner transformation [55] converts the fermionic creation and annihilation operators to the spin (qubit) operators faithfully preserving the algebra. It regards the vacuum state $|0\rangle$ as the down spin $|\downarrow\rangle$ and the occupied state $|1\rangle$ as the up spin $|\uparrow\rangle$. The algebra of the fermionic operators follows the anti-commutation relations

$$\{c_m, c_n^\dagger\} = \delta_{mn}, \qquad \{c_m, c_n\} = \{c_m^\dagger, c_n^\dagger\} = 0 \tag{B2}$$

where $c_m^\dagger$ and c_m are the fermionic creation and annihilation operators acting on the m-th lattice site, respectively, and δ_{mn} is the Kronecker delta. These relations can be represented in terms of the spin operators if one replaces the fermionic operators as

$$c_m^\dagger \rightarrow (-1)^{m-1} Z_1 Z_2 \cdots Z_{m-1} \sigma_m^+, \qquad c_m \rightarrow (-1)^{m-1} Z_1 Z_2 \cdots Z_{m-1} \sigma_m^- \tag{B3}$$

where $\sigma_m^+ = (X_m + iY_m)/2$ and $\sigma_m^- = (X_m - iY_m)/2$.

Appendix C. Nonlinearity of excited-state properies

In this appendix, we present further analysis of the 2-site Hubbard model discussed in section 4. The exact expressions of the excited-state properties of the Hubbard model in terms of the elements of the classical vector $\mathbf{x}_{|\psi_0(r)\rangle}$ present specific examples demonstrating that they may and may not be approximated with a linear model, given the classical vector.

There are $4^4 = 256$ Pauli operators (from $I_0 I_1 I_2 I_3$ to $Z_0 Z_1 Z_2 Z_3$) which may act on the Hilbert space for the 2-site Hubbard model (11). Exhaustive search for all of these Pauli operators reveals that only two functions of U appear as the ground state expectation values:

$$f_1(U) := \frac{1}{\sqrt{1 + \frac{16}{U^2}}} = -\langle Z_0 Z_1 \rangle_{\mathrm{GS}} = \langle Z_0 Z_3 \rangle_{\mathrm{GS}} = \langle Z_1 Z_2 \rangle_{\mathrm{GS}} = -\langle Z_2 Z_3 \rangle_{\mathrm{GS}} \tag{C4}$$

$$= -\langle X_0 X_1 X_2 X_3 \rangle_{\mathrm{GS}} = -\langle X_0 Y_1 X_2 Y_3 \rangle_{\mathrm{GS}} = -\langle Y_0 X_1 Y_2 X_3 \rangle_{\mathrm{GS}} = -\langle Y_0 Y_1 Y_2 Y_3 \rangle_{\mathrm{GS}}, \tag{C5}$$

$$f_2(U) := \frac{1}{\sqrt{1 + \frac{U^2}{16}}} = \langle X_0 Z_1 X_2 \rangle_{\mathrm{GS}} = \langle Y_0 Z_1 Y_2 \rangle_{\mathrm{GS}} = \langle X_1 Z_2 X_3 \rangle_{\mathrm{GS}} = \langle Y_1 Z_2 Y_3 \rangle_{\mathrm{GS}} \tag{C6}$$

$$= -\langle X_0 X_2 Z_3 \rangle_{\mathrm{GS}} = -\langle Y_0 Y_2 Z_3 \rangle_{\mathrm{GS}} = -\langle Z_0 X_1 X_3 \rangle_{\mathrm{GS}} = -\langle Z_0 Y_1 Y_3 \rangle_{\mathrm{GS}}. \tag{C7}$$

In other words, for any choice of the entangler U_{ent}, all components of the classical vector $\mathbf{x}_{|\psi_0(U)\rangle} = \left(\langle \psi_0(U) | U_{\mathrm{ent}}^\dagger X_0 U_{\mathrm{ent}} | \psi_0(U) \rangle, \cdots, \langle \psi_0(U) | U_{\mathrm{ent}}^\dagger Z_{N-1} U_{\mathrm{ent}} | \psi_0(U) \rangle \right)^T$ will be written as a linear combination of $f_1(U)$ and $f_2(U)$.

Now, we can see that the non-linearity in the classical unit is not necessary for a small value of U, but it is for a large U. For $0 < U \ll 1$, $f_1(U) \approx U/4$ and $f_2(U) \approx 1$, and the excited-state properties in terms of these functions are

$$\Delta E_1 \approx 2 - \frac{U}{2}, \quad \Delta E_2 \approx 2 + \frac{U}{2}, \quad \|\boldsymbol{\mu}_{\mathrm{eg}}\| \approx \frac{1}{\sqrt{2}} - \frac{U}{8\sqrt{2}}, \tag{C8}$$

ignoring $\mathcal{O}(U^2)$ terms. In this case, the excited-state properties can be expressed easily as the linear combination of $f_1(U)$ and $f_2(U)$. On the other hand, when U is large, i.e. when $0 < 1/U \ll 1$, it follows that $f_1(U) \approx 1$ and $f_2(U) \approx 4/U$, and the excited-state properties can be expressed as

$$\Delta E_1 \approx \frac{4}{U}, \quad \Delta E_2 \approx U, \quad \|\boldsymbol{\mu}_{\mathrm{eg}}\| \approx \frac{2}{U}, \tag{C9}$$

ignoring $O(1/U^2)$ terms. ΔE_2 may not be expressed as a linear combination of f_1 and f_2.

To support the observation, we also perform a numerical simulation for the 2-site Hubbard model. We randomly sample 30 distinct values of U for the training data and 50 distinct values of it for the test data in the range of $U \in [0.1, 6]$ (Case 1) and $U \in [0.1, 20]$ (Case 2). The ground state wavefunction of $H_{\mathrm{Hub}}(U)$ is prepared by the exact diagonalization. Instead of using an entangler, here we define the classical vector $\mathbf{x}_{|\psi_0(U)\rangle}$ as $(\langle Z_0 Z_1 \rangle_{\mathrm{GS}}, \langle X_0 Z_1 X_2 \rangle_{\mathrm{GS}})^T$. The linear regression to learn the excited-state properties $\mathbf{y} = (\Delta E_1, \Delta E_2, \|\boldsymbol{\mu}_{\mathrm{eg}}\|)^T$ from $\mathbf{x}_{|\psi_0(U)\rangle}$ is performed both for Case 1 and Case 2. All values of $\mathbf{x}_{|\psi_0(U)\rangle}$ and $\mathbf{y}$ are standardized by using the mean and the standard deviation of the training dataset, respectively, during the training process of the LR. The results are shown in figure C1. The LR predicts the excites state properties almost perfectly for small values of U as one may see in the results for Case 1, whereas it fails once one tries to learn and predict from the data with large U values as shown in the results for Case 2, especially evident from the prediction of ΔE_2. Those results support our expectation that a linear classical unit can sufficiently approximate the excited-state properties in the case of small U, but it may not for a large U. We consider a similar mechanism applies to the numerical simulations of small molecules in section 3, where atomic-spacings are not very small so that the Coulomb repulsion U is not large.
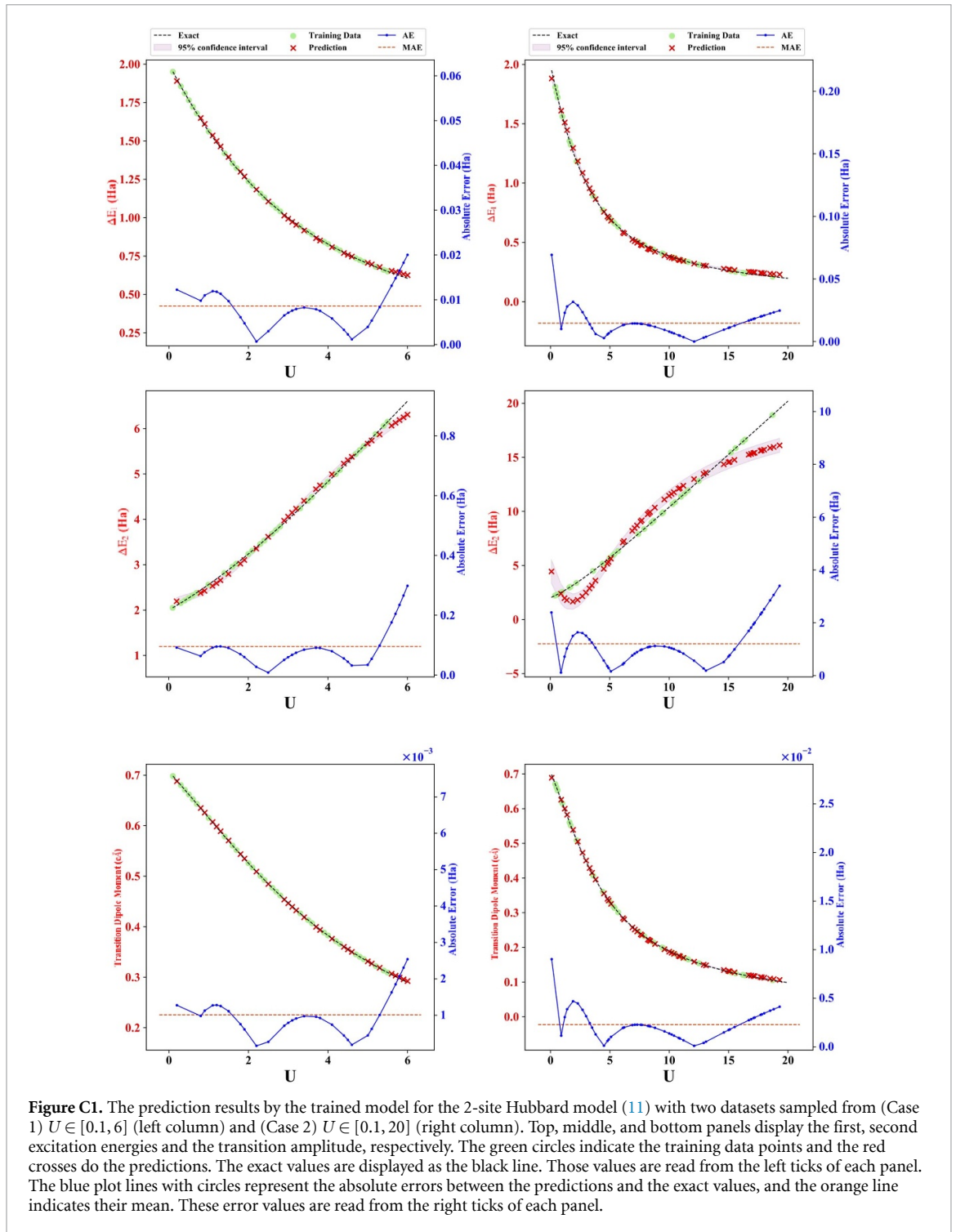
Appendix D. Time evolution operator as the entangler: its action on the single qubit operators

In the numerical simulations in section 3, we adopt the time evolution operator $e^{-iH_{\mathrm{rand}}T}$, where H_{rand} is a random Hamiltonian and T is a fixed time for the evolution, as the entangler U_{ent}. In this case, we can intuitively understand the effect of the entangler by considering the time evolution of the single-qubit Pauli operators $\bigcup_{i=0}^{N-1} \{X_i, Y_i, Z_i\}$ in the Heisenberg picture, where N is the number of qubits in the system.

As discussed in section 2.1, the information we obtain as the outputs of the quantum circuit in our model are the expectation values of the complicated operators $\bigcup_{i=0}^{N-1} \{U_{\mathrm{ent}}^\dagger X_i U_{\mathrm{ent}}, U_{\mathrm{ent}}^\dagger Y_i U_{\mathrm{ent}}, U_{\mathrm{ent}}^\dagger Z_i U_{\mathrm{ent}}\}$ for the ground state $|\psi\rangle$. If we expand the random Hamiltonian as $H_{\mathrm{rand}} = \sum_{P \in \{I,X,Y,Z\}^{\otimes N}} h_P P$ in the basis of N-qubit Pauli operators P and coefficients h_P, it follows

$$U_{\mathrm{ent}}^\dagger X_i U_{\mathrm{ent}} = e^{iH_{\mathrm{rand}}T} X_i e^{-iH_{\mathrm{rand}}T} = X_i + iT \sum_{P \in \{I,X,Y,Z\}^{\otimes N}, P \neq I} h_P [P, X_i] + O(T^2), \tag{D10}$$

where $[A, B] = AB - BA$. Same for the Y_i and Z_i operators. $i[P, X_i]$ is another N-qubit Pauli operator with larger support (i.e. the number of qubits on which $i[P, X_i]$ acts nontrivially is larger) than X_i, if P nontrivially acts on the ith qubit and one or more other qubits. As seen in section 4.1 and Appendix C, to estimate the excited-state properties, we generally need the information of the expectation values of certain Pauli operators nontrivially acting on multiple qubits. Hence, if the set of the Pauli operators $\{i[P, X_i]\}_P$ includes such required operators, the machine learning unit may automatically find them and construct the excited-state properties as a function of the expectation values. We note that $O(T^2)$ terms contain the terms like $[[P', [P, X_i]]$, $O(T^3)$ terms contain the terms like $[[P'', [P', [P, X_i]]]$, and so on, so even when $\{i[P, X_i]\}_P$ does not contain the required operators, they may be contained in these higher-order terms, and the machine learning unit may find them if T is large enough so that the higher-order terms in T contribute enough to $U_{\mathrm{ent}}^\dagger X_i U_{\mathrm{ent}}$. Hence, the operators $\{U_{\mathrm{ent}}^\dagger X_i U_{\mathrm{ent}}, U_{\mathrm{ent}}^\dagger Y_i U_{\mathrm{ent}}, U_{\mathrm{ent}}^\dagger Z_i U_{\mathrm{ent}}\}_{i=0}^{N-1}$ are constituted from more long-ranged, multi-qubit Pauli operators if (1) the random Hamiltonian H_{rand} contains stronger and longer-ranged interactions and/or (2) the time for the evolution becomes larger. This means that the expectation values of $\bigcup_{i=0}^{N-1} \{U_{\mathrm{ent}}^\dagger X_i U_{\mathrm{ent}}, U_{\mathrm{ent}}^\dagger Y_i U_{\mathrm{ent}}, U_{\mathrm{ent}}^\dagger Z_i U_{\mathrm{ent}}\}$ for the ground state $|\psi\rangle$ bring more information of $|\psi\rangle$ and the original molecular Hamiltonian than those of $\bigcup_{i=0}^{N-1} \{X_i, Y_i, Z_i\}$ for the ground state, and there is more chance for the machine learning unit to successfully predict the excited-state properties from the information. We note that the physical picture of the spreading of the single-qubit Pauli operators over the whole system under chaotic Hamiltonians was discussed in reference [65].

Figure C1. The prediction results by the trained model for the 2-site Hubbard model (11) with two datasets sampled from (Case 1) $U \in [0.1, 6]$ (left column) and (Case 2) $U \in [0.1, 20]$ (right column). Top, middle, and bottom panels display the first, second excitation energies and the transition amplitude, respectively. The green circles indicate the training data points and the red crosses do the predictions. The exact values are displayed as the black line. Those values are read from the left ticks of each panel. The blue plot lines with circles represent the absolute errors between the predictions and the exact values, and the orange line indicates their mean. These error values are read from the right ticks of each panel.

Appendix E. Dependence of the excited-state prediction on the number of shots for the VQE

In this appendix, we present how the accuracy of the predictions from our model varies with the number of shots used to perform the VQE for preparing the dataset of the ground states. We carried out the same numerical simulation for the LiH molecules in the noisy situation as described in sections 3.3 and 3.4.2, but with the various numbers of shots ranging from 100 to 10^6 for the computation of the VQE, instead of fixing it to 10 000 shots. Left panel of figure E2 displays the MAE for the predictions of ΔE_1 versus the number of shots for the VQE, showing that the MAE decreases almost monotonically with the number of shots. Right panel of figure E2 shows the MAE between the ground state energy computed by the VQE and the exact one obtained by diagonalization of the molecular Hamiltonians of LiH, as a function of the number of shots.

Figure E2. (Left) Dependence of the MAEs of the predictions of ΔE_1 for LiH on the number of shots employed in performing the VQE. (Right) Dependence of the MAEs between the ground state energy obtained by the VQE and the exact one computed by diagonalization of the molecular Hamiltonian of LiH on the number of shots to perform the VQE.
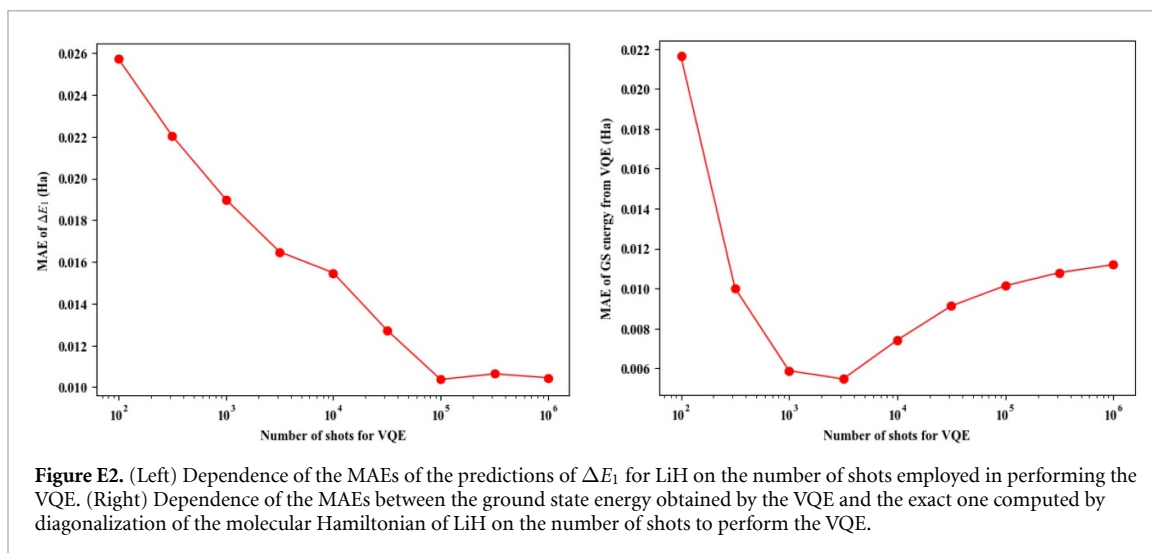
Interestingly, the accuracy of the VQE has an empirical overhead at around $10^3 - 10^4$ shots and gradually saturates the infinite-shots limit which is non-zero because of the presence of the noise. Two panels of figure E2 suggest that the accuracy of the predictions of the excited-state properties is almost independent of the precision of the computation result from the VQE. Rather, it depends on the number of shots, and one may simply increase the shots to obtain estimations with higher accuracy. We leave a deeper analysis of this curious dependence as future work.

ORCID iDs

Hiroki Kawai ⓘ https://orcid.org/0000-0002-3957-3231
Yuya O. Nakagawa ⓘ https://orcid.org/0000-0001-6191-4592

References

[1] LeCun Y, Bengio Y and Hinton G 2015 *Nature* **521** 436–44
[2] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (Cambridge, MA: MIT Press)
[3] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L and Zdeborová L 2019 *Rev. Mod. Phys.* **91** 045002
[4] Carrasquilla J and Melko R G 2017 *Nat. Phys.* **13** 431
[5] van Nieuwenburg E P L, Liu Y H and Huber S D 2017 *Nat. Phys.* **13** 435–9
[6] Rupp M, Tkatchenko A, Müller K R and von Lilienfeld O A 2012 *Phys. Rev. Lett.* **108** 058301
[7] Montavon G, Hansen K, Fazli S, Rupp M, Biegler F, Ziehe A, Tkatchenko A, Lilienfeld A V and Müller K R 2012 Learning invariant representations of molecules for atomization energy prediction *Advances in Neural Information Processing Systems 25* (Red Hook, NY: Curran Associates, Inc.) ed Pereira F, Burges C J C, Bottou L and Weinberger K Q pp 440–8
http://papers.nips.cc/paper/4830-learning-invariant-representations-of-molecules-for-atomization-energy-prediction.pdf
[8] Hansen K, Montavon G, Biegler F, Fazli S, Rupp M, Scheffler M, von Lilienfeld O A, Tkatchenko A and Müller K R 2013 *Journal of Chemical Theory and Computation* **9** 3404–19
[9] Harrow A W, Hassidim A and Lloyd S 2009 *Phys. Rev. Lett.* **103** 150502
[10] Wiebe N, Braun D and Lloyd S 2012 *Phys. Rev. Lett.* **109** 050505
[11] Lloyd S, Mohseni M and Rebentrost P 2014 *Nat. Phys.* **10** 631
[12] Rebentrost P, Mohseni M and Lloyd S 2014 *Phys. Rev. Lett.* **113** 130503
[13] Schuld M, Sinayskiy I and Petruccione F 2016 *Phys. Rev. A* **94** 022342
[14] Kerenidis I and Prakash A 2020 *Phys. Rev. A* **101** 022316
[15] Wang G 2017 *Phys. Rev. A* **96** 012335
[16] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N and Lloyd S 2017 *Nature* **549** 195–202
[17] Cong I, Choi S and Lukin M D 2019 *Nat. Phys.* **15** 1273–8
[18] Havlíček V, Córcoles A D, Temme K, Harrow A W, Kandala A, Chow J M and Gambetta J M 2019 *Nature* **567** 209–12
[19] Kusumoto T, Mitarai K, Fujii K, Kitagawa M and Negoro M 2019 *arXiv:1911.12021*
[20] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 *Phys. Rev. A* **98** 032309
[21] Farhi E and Neven H 2018 *arXiv:1802.06002*
[22] Wilson C, Otterbach J, Tezak N, Smith R, Crooks G and da Silva M 2018 *arXiv:1806.08321*
[23] Harrow A W and Montanaro A 2017 *Nature* **549** 203
[24] Arute F *et al* 2019 *Nature* **574** 505–10
[25] Preskill J 2018 *Quantum* **2** 79
[26] Giovannetti V, Lloyd S and Maccone L 2008 *Phys. Rev. Lett.* **100** 160501
[27] Giovannetti V, Lloyd S and Maccone L 2008 *Phys. Rev. A* **78** 052310
[28] Prakash A 2014 Quantum algorithms for linear algebra and machine learning *PhD Thesis* UC Berkeley
[29] Hohenberg P and Kohn W 1964 *Phys. Rev.* **136** B864–B871

[30] Roos B O 2012 *Lecture Notes in Quantum Chemistry: European Summer School in Quantum Chemistry* (MIT Press: Springer Science & Business Media) vol 58

[31] Helgaker T, Jorgensen P and Olsen J 2014 *Molecular Electronic-Structure Theory* (New York: Wiley)

[32] Goh G B, Hodas N O and Vishnu A 2017 *J. Comput. Chem.* **38** 1291–307

[33] Montavon G, Rupp M, Gobre V, Vazquez-Mayagoitia A, Hansen K, Tkatchenko A, Müller K R and von Lilienfeld O A 2013 *New J. Phys.* **15** 095003

[34] Ramakrishnan R, Hartmann M, Tapavicza E and von Lilienfeld O A 2015 *J. Chem. Phys.* **143** 084111

[35] Häse F, Valleau S, Pyzer-Knapp E and Aspuru-Guzik A 2016 *Chemical Science* **7** 5139–47

[36] Wu Z, Ramsundar B, Feinberg E, Gomes J, Geniesse C, Pappu A S, Leswing K and Pande V 2018 *Chem. Sci.* **9** 513–30

[37] Peruzzo A, McClean J, Shadbolt P, Yung M H, Zhou X Q, Love P J, Aspuru-Guzik A and O'Brien J L 2014 *Nat. Commun.* **5** 4213

[38] Sasaki M, Carlini A and Jozsa R 2001 *Phys. Rev.* A **64** 022317

[39] Sasaki M and Carlini A 2002 *Phys. Rev.* A **66** 022303

[40] Fujii K and Nakajima K 2017 *Phys. Rev.* A **8** 024030

[41] Ghosh S, Opala A, Matuszewski M, Paterek T and Liew T C H 2019 *npj Quantum Information* **5** 35

[42] McClean J R, Kimchi-Schwartz M E, Carter J and de Jong W A 2017 *Phys. Rev.* A **95** 042308

[43] Colless J I, Ramasesh V V, Dahlen D, Blok M S, Kimchi-Schwartz M E, McClean J R, Carter J, de Jong W A and Siddiqi I 2018 *Phys. Rev.* X **8** 011021

[44] Nakanishi K M, Mitarai K and Fujii K 2019 *Phys. Rev. Research* **1** 033062

[45] Parrish R M, Hohenstein E G, McMahon P L and Martínez T J 2019 *Phys. Rev. Lett.* **122** 230401

[46] Higgott O, Wang D and Brierley S 2019 *Quantum* **3** 156

[47] Jones T, Endo S, McArdle S, Yuan X and Benjamin S C 2019 *Phys. Rev.* A **99** 062304

[48] Ollitrault P J *et al* 2019 *arXiv:1910.12890*

[49] Tilly J, Jones G, Chen H, Wossnig L and Grant E 2020 *arXiv:2001.04941*

[50] Uvarov A, Kardashin A and Biamonte J 2020 *Phys. Rev.* A **102** 012415

[51] McArdle S, Endo S, Aspuru-Guzik A, Benjamin S and Yuan X 2020 *Rev. Mod. Phys.* **92** 015003

[52] Cao Y *et al* 2019 *Chem. Rev.* **119** 10856–10915

[53] Sun Q *et al* 2017 *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8** e1340

[54] McClean J R *et al* 2020 *Quantum Sci. Technol.* **5** 034014

[55] Jordan P and Wigner E 1928 *Zeitschrift für Physik* **47** 631–51

[56] Lee J, Huggins W J, Head-Gordon M and Whaley K B 2019 *J. Chem. Theory Comput.* **15** 311–24

[57] Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow J M and Gambetta J M 2017 *Nature* **549** 242–6

[58] Kandala A, Temme K, Córcoles A D, Mezzacapo A, Chow J M and Gambetta J M 2019 *Nature* **567** 491–5

[59] Chen J, Nurdin H I and Yamamoto N 2020 *Phys. Rev. Applied* **14** 024065

[60] Bishop C M 2006 *Pattern Recognition and Machine Learning* (New York, NY: Springer) https://cds.cern.ch/record/998831

[61] Nielsen M A and Chuang I L 2011 *Quantum Computation and Quantum Information: 10th edn* (New York, NY, USA: Cambridge University Press) 10th ed

[62] Temme K, Bravyi S and Gambetta J M 2017 *Phys. Rev. Lett.* **119** 180509

[63] Endo S, Benjamin S C and Li Y 2018 *Phys. Rev.* X **8** 031027

[64] Hubbard J and Flowers B H 1963 *Proc. of the Royal Society of London. Series A. Mathematical and Physical Sciences* **276** 238–57

[65] Ho W W and Abanin D A 2017 *Phys. Rev.* B **95** 094302