Hamiltonian learning from time dynamics using variational algorithms

Rishabh Gupta,[†] Raja Selvarajan,[‡] Manas Sajjan,[†] Raphael D. Levine,[¶,§] and Sabre Kais[*,‖]

[†]Department of Chemistry, Purdue University, West Lafayette, IN, USA

[‡]Department of Physics and Astronomy, Purdue University, West Lafayette, IN, USA

[¶]The Fritz Haber Center for Molecular Dynamics and Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

[§]Department of Chemistry and Biochemistry and Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

[‖]Department of Chemistry, Department of Physics and Astronomy, and Purdue Quantum Science and Engineering Institute, Purdue University, West Lafayette, IN, USA

E-mail: kais@purdue.edu

Abstract

The Hamiltonian of a quantum system governs the dynamics of the system via the Schrodinger equation. In this paper, the Hamiltonian is reconstructed in the Pauli basis using measurables on random states forming a time series dataset. The time propagation is implemented through Trotterization and optimized variationally with gradients computed on the quantum circuit. We validate our output by reproducing the dynamics of unseen observables on a randomly chosen state not used for the optimization. Unlike existing techniques that try and exploit the structure/properties

1

of the Hamiltonian, our scheme is general and provides freedom with regard to what observables or initial states can be used while still remaining efficient with regard to implementation. We extend our protocol to doing quantum state learning where we solve the reverse problem of doing state learning given time series data of observables generated against several Hamiltonian dynamics. We show results on Hamiltonians involving XX, ZZ couplings along with transverse field Ising Hamiltonians and propose an analytical method for the learning of Hamiltonians consisting of generators of the $SU(3)$ group. This paper is likely to pave the way toward using Hamiltonian learning for time series prediction within the context of quantum machine learning algorithms.

1 Introduction

Quantum tomography[1–6] includes the study of quantum systems (state tomography) and the dynamics (process tomography) that govern these systems. While in process tomography, we are interested in reconstructing the Hamiltonian that governs the dynamics of the system,[7–12] in quantum state tomography we characterize the quantum mechanical state by the measurement of expectation values of an informationally complete (IC) set of Hermitian operators given multiple copies of a system. Within process tomography, we have direct methods that include standard quantum process tomography,[13] ancilla assisted process[14–16] and indirect methods that make use of system dynamics, quantum entanglement, and average fidelity information.[17–19] Some of the recent work in state tomography includes using permutation invariance,[20] entropy maximization,[21–23] shadow tomography[24] , incoherent measurements[25] and ensemble averaging.[26]

In principle, the Hamiltonian learning problem requires estimating a number of parameters that increases exponentially with the size of the system. But in reality, most physical Hamiltonians can be described by only a few-body relevant interaction terms that scale polynomially with the size of the quantum systems. While the conventional Hamiltonian learning

techniques require the preparation of ground or thermal states which is still a challenge,[27,28] our Hamiltonian learning algorithm entirely circumvents this problem by providing independence on the choice of the quantum states as well as on the observables whose time dynamics can be easily recorded. For this work, we focus on n-qubit Hamiltonians involving at most two-body interactions and can be defined as $H = \sum_{i,j>i}^{n} c_{ij}\sigma_i\sigma_j$ where $\sigma's$ are the Pauli string operators (σ_x, σ_y, σ_z, σ_i). To reconstruct the Hamiltonian of such a system, it is sufficient to estimate the parameters $\{c_{ij}\}$ and one of the ways of doing so is through optimization using variational algorithms. Here, we do a study of Hamiltonian learning using the time dynamics of observables measured on different quantum states using variational algorithms run on parameterized quantum circuits (PQC's). The coefficients of the Hamiltonian are reconstructed in the Pauli basis through an optimization scheme solved on PQC's. Variational quantum algorithms (VQA)[29] are seen as one of the promising techniques that can likely explore the power of computation on quantum circuits.[30–35] Some of the applications include electronic structure calculations,[36,37] explore low energy symmetry states,[38] and supervised machine learning.[39] The interested reader may refer to[40] for a thorough study of variational algorithms and their use in quantum machine learning for chemistry and physics.

We show how the learning of the Hamiltonian is affected by increasing the number of observables, quantum states, and the sampling frequency that we choose over the time interval. We exploit the fact that given sufficient observables to characterize the dynamics of a Hamiltonian, we can uniquely converge on reconstructing the exact Hamiltonian. Further, we show examples that exploit the knowledge of the Hamiltonian locality and try reconstructing the Hamiltonian of the system with very high fidelity. In practice, one does not have access to the actual Hamiltonian. We thus create validation schemes that can be used to inform us about the statistical knowledge required for reconstructing the Hamiltonian.

Following this, we extend the scope of the method to state learning where the tools developed straightforwardly extend with a hardware efficient ansatz used to model the state. Unlike

conventional quantum state learning of providing a state description using tomography we here would like to characterize the state as parameters of an ansatz that maps to it. The methods used here are found to be robust to random parameter Hamiltonians and initial starting states used.

2 Methodology

We describe the problem in the context of spin systems as they directly map to the space of qubits over which the computations shall be performed. Let an unknown Hamiltonian act on a set of n qubits whose initial state we get to choose/prepare. The experimenter measures the time dynamics of random observables chosen on this system. We would like to be able to predict the Hamiltonian that closely approximates the dynamics of the observables. Throughout our discussion, we shall restrict ourselves to second-order Hamiltonian couplings. Generalization of this can be extended to higher-order couplings. In the absence of any knowledge of the system's Hamiltonian, we consider a generic Hamiltonian with all to all coupling. Thus,

$$H = \sum_{i,j,\beta,\gamma} J_{ij\beta\gamma}\sigma_i^\beta \sigma_j^\gamma \tag{1}$$

where β, γ indexes the Pauli matrices $\{X, Y, Z\}$. Let $\{|\psi_i\rangle\}$ be the set of states over which the Hamiltonian is allowed to act and let $\{O_\alpha\}$ be the observables whose dynamics have been recorded/sampled at discrete intervals of time. We work in the framework where $\{|\psi_i\rangle\}$ can be efficiently prepared and $\{O_\alpha\}$ can be efficiently measured. Let $O_\alpha(k\Delta t)$ refer to the measurement of observable O_α made at the timestep k, where Δt refers to the time interval between the measurements.

To start with we randomly initialize the parameters of the J_{ij} of the Hamiltonian. We then construct the unitary operator $U(t) = exp(-iHt)$ that generates the time dynamics of H

4

using Trotterization.[41] The constructed unitary U is allowed to act on each of the starting states $|\psi_i\rangle$ and observables $|O_\alpha\rangle$ is measured at intervals of Δt for a total of N timesteps, thus running for the evolution for the duration of $N\Delta t$. Having a Trotter implementation of U for time Δt allows us to trivially extend the dynamics for time $k\Delta t$ by repeating k layers of our base implementation. Just like any other machine learning algorithm, we have a scheme that implements the dynamics of a generic Hamiltonian H and data points $O_\alpha(k\Delta t)$ measured. We are now left with minimizing a cost function of our choice. We choose the 2-norm function to minimize between $O_\alpha(k\Delta t)$ and $O_{\alpha,i}^{obs}(k\Delta t)$ over the states $|\psi_i\rangle$. Thus our cost function is,

$$\text{Cost} = \sum_{\alpha,i,k}(\langle\psi_i|\,U^\dagger(k\Delta t)\,O_\alpha\,U(k\Delta t)\,|\psi_i\rangle - O_{\alpha,i,k}^{obs})^2 \tag{2}$$

where the summation α indexes the summation over observables, i indexes the summation over various starting states over which the dynamics is carried, and k indicates the time steps over the evolution. Here $O_{\alpha,i,k}^{obs}$ refers to the observable O^{obs} measured on $|\psi_i\rangle$ at the timestep k. To minimize the cost function we use standard gradient descent (SGD) to compute the gradients $\frac{d\text{Cost}}{dJ_{ij}}$. Within Trotterization one can trivially map the couplings J_{ij} directly to the circuit parameters of the constructed $U(\Delta t)$. This allows us to trivially compute the gradient to circuit parameters using parameter shift rule.[42] Alternatively, the analytical expression for gradients can be directly computed on a quantum circuit for these specific Trotter circuits as shown in Appendix 5.2.[43]

As a measure of validation, the experimenter then validates against a new observable not yet measured on the system. If the Hamiltonian fails to reproduce an approximate comparison, the experimenter chooses to either increase the number of starting states or the number of observables over which the dynamics are observed. We show a shallow-depth implementation of Trotterization for an XYZ Hamiltonian in Appendix 5.1.

Extension to State learning

We showed how using multiple starting states one can learn the Hamiltonian given the dynamics of the observables. We now solve the inverse problem. We use multiple Hamiltonians to learn/prepare a state of interest. Thus formally, let $\{H_i\}$ be the set of Hamiltonians over which the state is allowed to evolve and let $\{O_\alpha\}$ be the observables whose dynamics have been recorded/sampled at discrete intervals of time. We work in the framework where $\{U_i = exp(-iH_it)\}$ can be efficiently synthesized and $\{O_\alpha\}$ can be efficiently measured. Let $O_\alpha(k\Delta t)$ refer to the measurement of observable O_α made at timestep k, where Δt refers to the time interval of each timestep.
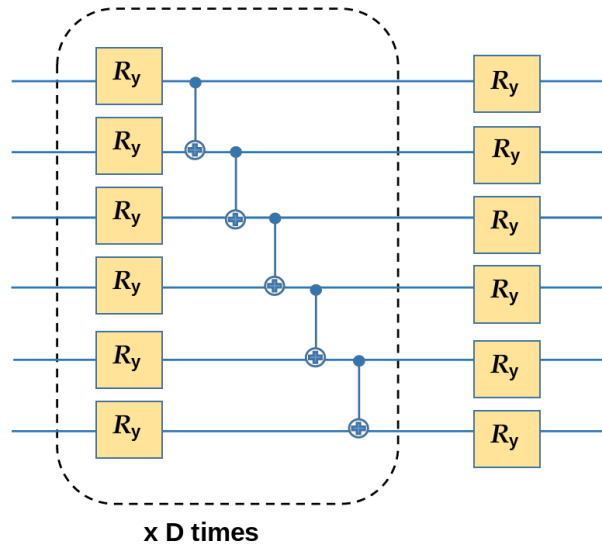

Figure 1: Ansatz $V(\vec{\beta})$ used to generate the state $\left|\psi(\vec{\beta})\right\rangle$ with real coefficients. The ansatz is made of R_y Pauli rotation gates to restrict to real coefficients and a ladder of CNOT gates that can be used to create the entanglement.

We choose an ansatz $V(\vec{\beta})$ to prepare the state of interest $\left|\psi(\vec{\beta})\right\rangle$. Fig 1 shows a schematic representation of the ansatz used to create a state with real coefficients. Thus we have, $V(\vec{\beta})\left|0\right\rangle^{\otimes n} = \left|\psi(\vec{\beta})\right\rangle$, where n is the number of qubits used to represent the state. We then impose the constraints of the dynamics recorded by O_α. Thus we end up with an optimization problem with the following cost function to be minimized over the variables $\vec{\beta}$,

$$\text{Cost} = \sum_{\alpha,i,k} (\left\langle \psi(\vec{\beta}) \middle| U_i^{\dagger}(k\Delta t) \, O_\alpha \, U_i(k\Delta t) \middle| \psi(\vec{\beta}) \right\rangle - O_{\alpha,i,k}^{obs})^2 \qquad (3)$$

where the summation α indexes the summation over observables, i indexes the summation over various unitaries $U_i = exp(-iH_i t)$ that implements the dynamics, and k indicates the time steps over the evolution. Here $O_{\alpha,i,k}^{obs}$ refers to the observable O^{obs} measured on $|\psi_i\rangle$ at the timestep k. To minimize the cost function we use SGD to compute the gradients $\frac{d\text{Cost}}{d\beta_s}$, where s labels the indices of $\vec{\beta}$. These gradients can be computed again trivially using the parameter shift rule.

3   Results and discussion

In this current work, we propose a variational approach for Hamiltonian and quantum state learning based on the time dynamics of various observables. We considered various quantum systems and performed numerical simulations in IBM's Qiskit[44] to test and validate the proposed formalism.

3.1   2-local Hamiltonian with all $\sigma_z\sigma_z$ and $\sigma_x\sigma_x$ coupling

To test the validation of our proposed approach for Hamiltonian tomography, we first considered a general 2-local Hamiltonian of the form: $H = \sum_{i,j>i} c_{ij}\sigma_z^i\sigma_z^j + \sum_{i,j>i} d_{ij}\sigma_x^i\sigma_x^j$ where (c_{ij}, d_{ij}) are randomly chosen coupling parameters between qubit-i and qubit-j. This Hamiltonian H is the true Hamiltonian that we want to learn and thereby predict the time dynamics of certain unknown observables. The cost function of the optimization process is defined as the mean square error between the true and reconstructed time series expectation values of the different observables considered computed using random starting quantum states. To verify that the starting quantum states are chosen at random we plot in Figure 2 the state overlap and the trace distance, $\text{T}(\rho, \sigma) = \frac{1}{2}||\rho - \sigma||_1$, between the considered quantum states.

The observables correspond to the 2-point correlation functions: $\sigma_z^i \sigma_z^j$ and $\sigma_x^i \sigma_x^j$ whose time dynamics can be obtained from experiments. We, hereby, present an extensive list of results and inferences from the learning of the above Hamiltonian.
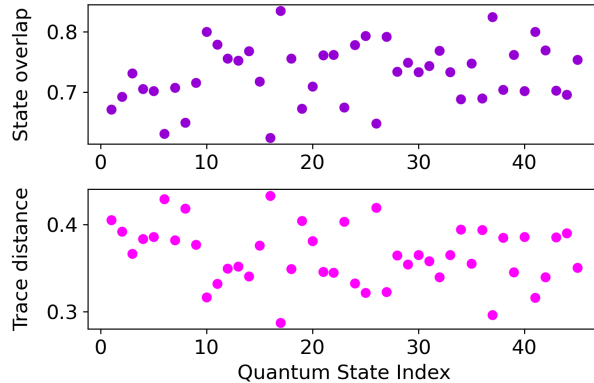

Figure 2: State overlap and trace distance between the considered quantum states to depict that the proposed protocol is independent of the choice of the starting quantum states.

One of the measures that we considered to test the performance of our approach is the trace distance between the true and the reconstructed Hamiltonian. The trace distance in the case of two Hamiltonians is defined as: if $U_H = \exp\{-iHt\}$ and $U_K = \exp\{-iKt\}$ are the evolution operators for the true (H) and the learning (K) Hamiltonians then the trace distance is $T(H, K) = ||U_H(t)^\dagger U_K(t) - I||_2$. As described in Section 2, the different variables in our approach are the number of time steps (N_T) up to which the expectation values of the different observables need to be recorded, the number of starting quantum states (N_S), and the number of observables (N_O) required for accurate learning of the Hamiltonian. Since in an experiment, not all observables can be measured with high fidelity so we try to use a limited number of observables for which the time dynamics are recorded for various starting states. We vary each of these variables and calculate the trace distance in each case to find an optimized set of (N_T,N_S,N_O) as per the requirement of the experiment.

Figure 3 shows the variation of the trace distance with the number of epochs during the optimization process. Each subplot in Figure 3 depicts the trace distance variation for a fixed number of starting quantum states but for a different number of time steps for which
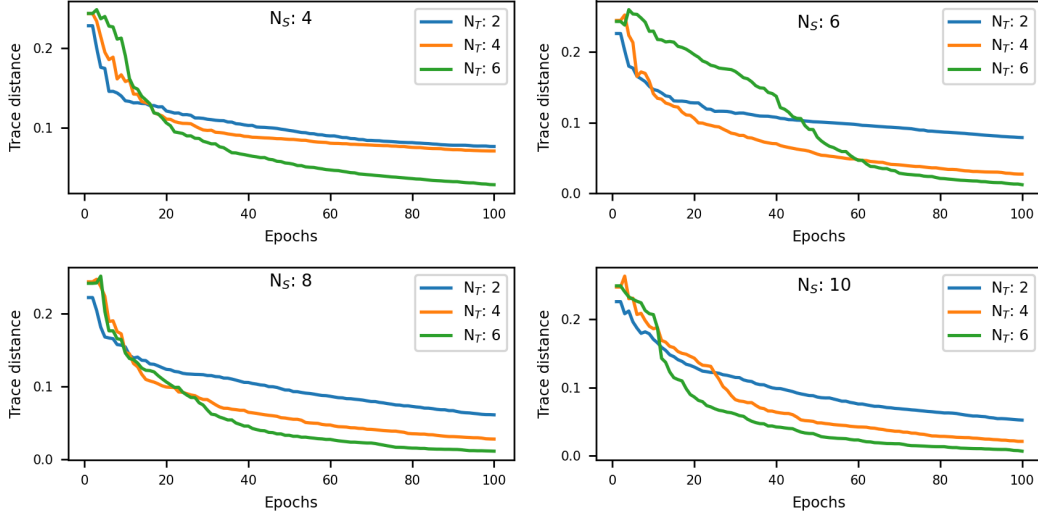
8

Figure 3: Variation of the trace distance between the true and the reconstructed Hamiltonian as a function of the number of epochs for fixed $N_O = 3$. It is evident that the Hamiltonian can be learned more efficiently if the expectation values of observables at higher time steps are also known for any number of starting quantum states.
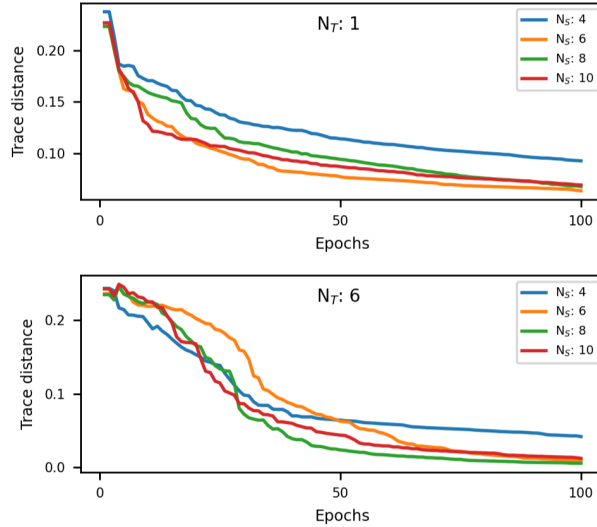

Figure 4: Convergence of the trace distance between the true and the reconstructed Hamiltonian is plotted to highlight the effect of increasing the number of starting quantum states when N_O and N_T are fixed.

the observable's expectation values are known. For this plot we fixed the number of different observables of $\sigma_k^i \sigma_k^j$ to 3 i.e. we randomly selected expectation values of 3 $\sigma_z^i \sigma_z^j$ and 3 $\sigma_x^i \sigma_x^j$ correlation functions. As can be seen in the Figure, the convergence of the trace distance improves when observables corresponding to higher time steps are included in the optimiza-

tion in all cases of N_S. To analyze the effect of the variation of N_S on the accuracy of the reconstruction of the Hamiltonian, we show in Figure 4 the trace distance convergence for fixed $N_O = 3$, and N_S being varied in each subplot for one particular N_T. We see that for $N_S \geq 6$ the converged trace distance is very close to 0 and therefore, although increasing the number of random quantum states does improve the performance up to a certain value of N_S for all N_T but beyond that, the accuracy is independent of increasing the quantum states.

The next step is to analyze the performance of our approach upon varying the number of
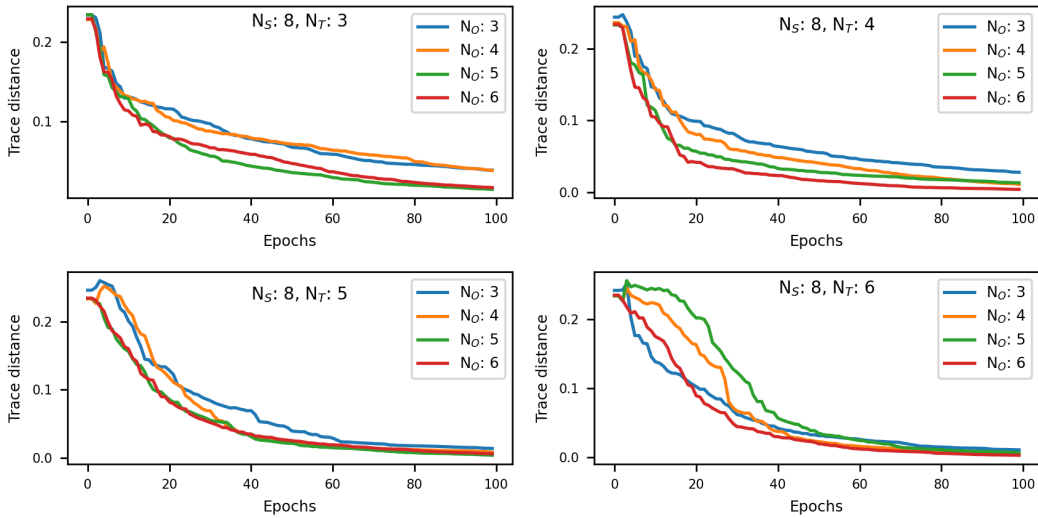

Figure 5: Plots of the trace distance upon varying the number of observables (N_O) considered for Hamiltonian learning. The convergence of the learned Hamiltonian to the true Hamiltonian improves upon increasing N_O in all cases. It is also interesting to note that we can obtain a highly accurate Hamiltonian by using fewer observables for higher N_T as is evident from the bottom plots.

observables (correlation functions) whose expectation values are considered in the Hamiltonian learning protocol. In Figure 5, we plot the trajectories of trace distance for the various number of observables (N_O) considered for a fixed N_T and N_S in each subplot. We see that the reconstructed Hamiltonian's trace distance with the true Hamiltonian converges to 0 faster upon increasing N_O in the protocol. We also change N_T in the different subplots to emphasize the advantage of using the time dynamics of observables for optimization. As can be seen in Figure 5 that $T(H, K)$ approaches 0 upon optimization even for fewer observables

upon including their expectation values for higher time steps.

Once the optimized set of parameters is obtained upon convergence, one of the ways to
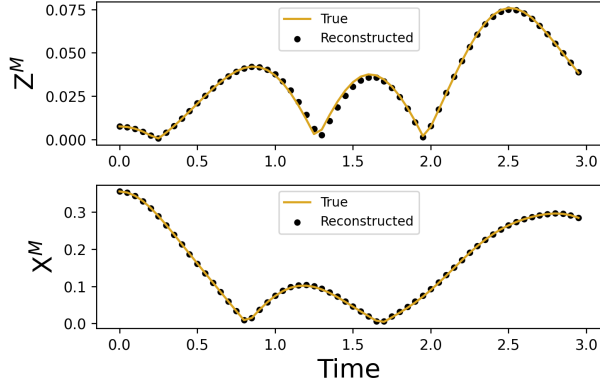

Figure 6: Time dynamics for X^M and Z^M-magnetic field observables using the true and the reconstructed Hamiltonian for a general 2-local Hamiltonian with random coupling parameters.

test the accuracy of learning is to evolve a quantum state under the true and the learned Hamiltonian separately and compare the dynamics of an observable under such an evolution. For our case, we considered the X and Z-magnetic field operators, denoted by X^M and Z^M respectively, and defined by $\hat{K} = \sum_p \sigma_k^p \sigma_i \sigma_i \sigma_i \sigma_i$ ($k = x, z$). We computed the time dynamics of these operators using the evolution of a random quantum state under the true and the learned Hamiltonian. High overlap in the trajectories in Figure 6 validates that the proposed protocol is successful in the efficient reconstruction of an unknown Hamiltonian.

3.2   Reconstructing Transverse Field Ising Model Hamiltonian

The aforementioned results correspond to the learning of a general 2-local Hamiltonian with random coupling parameters. Since every quantum system has its own Hamiltonian so we need an empirical recipe for determining the optimized set of N_O, N_S, N_T. We show the reconstruction of the Transverse Field Ising Model (TFIM) Hamiltonian for the following two systems:

11

3.2.1 Inhomogeneous 5-qubit Transverse Field Ising Model Hamiltonian

Here we present a practical approach for Hamiltonian learning for the specific case of an inhomogeneous 5-qubit TFIM Hamiltonian, defined by $H = \sum_i c_i \sigma_x^i + \sum_{i,j=i+1} d_{ij} \sigma_z^i \sigma_z^j$, but is generalizable to any Hamiltonian. We inferred from the previous section that the time dynamics of various observables is an essential element for assisting the convergence of the trace distance. Therefore, from the perspective of experiments when the true Hamiltonian is unknown, one can start with obtaining mean values of a few observables using a few random starting quantum states but for higher time steps. Then our proposed optimization method can be applied to obtain the coupling parameters in the Hamiltonian. Since for practical implementation, we won't have access to the true Hamiltonian and so, one cannot calculate the trace distance to validate that the reconstructed Hamiltonian is close to the true Hamiltonian. However, what we can have access to is the time dynamics of another observable on the real system that is not used during the optimization process. We, therefore, define the concept of validation error as the mean square error between the time series expectation values of the observable calculated using a random quantum state evolved under the true Hamiltonian and the reconstructed Hamiltonian. Figure 7 shows the validation error corresponding to the Z^M-magnetic field operator for various values of N_O, N_S, N_T for the considered TFIM Hamiltonian. Thus, the optimized set for N_O, N_S, N_T can be obtained when the validation error is below a certain threshold as per the requirement of the experiment.

For the inhomogeneous 5-qubit TFIM we obtain the converged coupling parameters corresponding to N_O, N_S, N_T for which the validation error is below 1e-6. The true and the learned Hamiltonian is then used to evolve a random quantum state to obtain the time series expectation values of the observables X' and Z' which are the 3-point correlation functions $\sigma_x^i \sigma_x^j \sigma_x^k$ and $\sigma_z^i \sigma_z^j \sigma_z^k$. The performance efficacy of our method can be validated from Figure 8 which shows the overlap at each time step in the dynamics of these observables using the true and the reconstructed Hamiltonian.
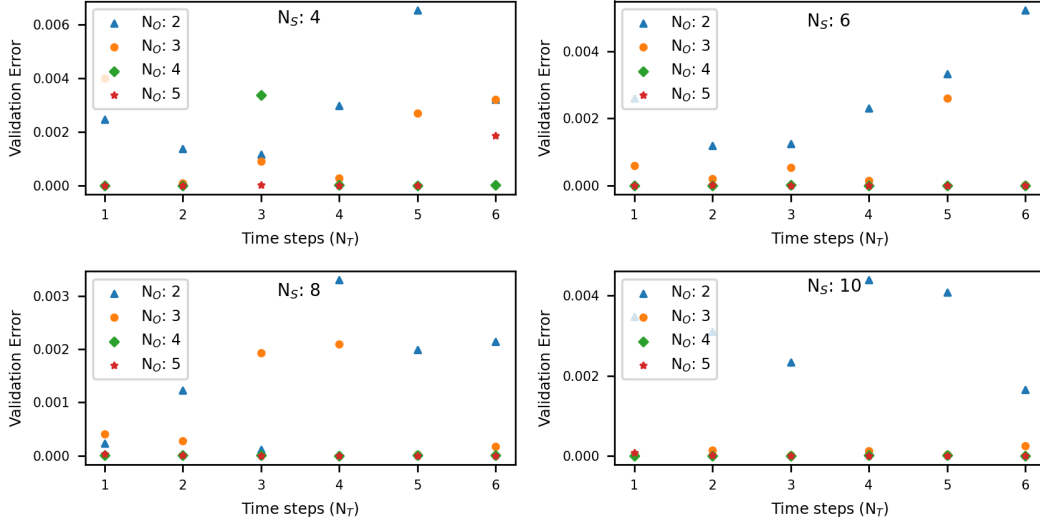
Figure 7: Validation error corresponding to the time dynamics of Z^M-magnetic field operator following the reconstruction of a 5-qubit TFIM Hamiltonian using the various values for N_O, N_S, N_T.
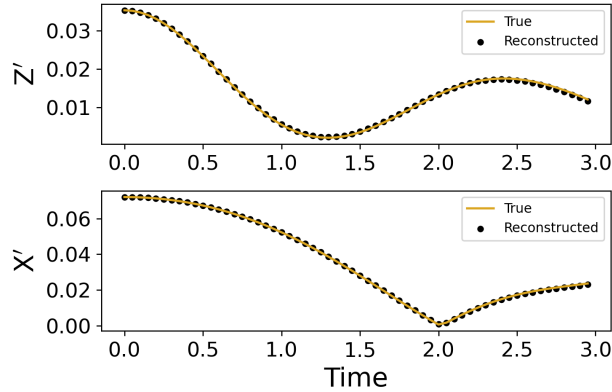

Figure 8: Time dynamics of 3-point correlation functions X' and Z' for 5-qubit TFIM Hamiltonian. The learned Hamiltonian is obtained using $N_O = 3, N_S = 8, N_T = 5$ for which the validation error is below 1e-6.

3.2.2 Homogeneous 10-qubit Transverse Field Ising Model Hamiltonian

The quantum Ising chain is a classic domain for testing various ideas and methodologies from statistical mechanics and consequently, forms a classic candidate to experimentally test the efficiency of any hamiltonian learning formalism. For the learning of a homogeneous TFIM hamiltonian, defined by $H = h \sum_i \sigma_x^i + J \sum_{i,j=i+1} \sigma_z^i \sigma_z^j$, we considered a quantum system comprising of 10-qubits. Since there are only 2 parameters (h, J) that need to be learned,

13

the parameterized hamiltonian from the proposed protocol converges quickly to the true Hamiltonian. We again chronologically vary the set of (N_T, N_S, N_O) during optimization to procure that set for which the validation error is below 1e-6 following which the training parameters are obtained and used for predicting the dynamics of certain unknown observables. Figure 9 shows the dynamics of X and Z-magnetic fields' observables under the true and the learned Hamiltonian for various time steps which again validates the learning approach.
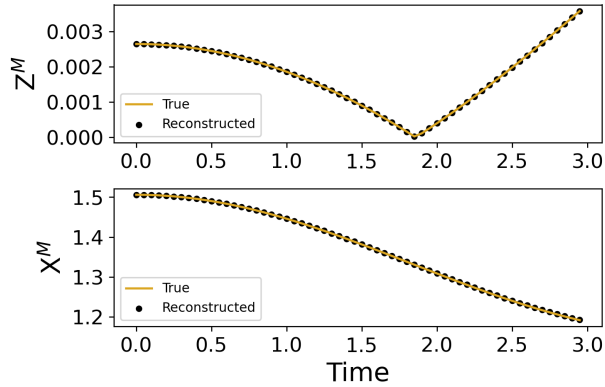

Figure 9: Time dynamics of X^M and Z^M-magnetic field observables for 10-qubit homogeneous TFIM Hamiltonian. The learned Hamiltonian is obtained using $N_O = 2, N_S = 2, N_T = 1$ for which the validation error is below 1e-6.

3.3 Quantum State Learning

Efficient characterization of a pure quantum state by learning the parameters of a parameterized quantum circuit has recently been gaining much popularity within the domain of quantum state learning owing to their use of shallow quantum circuits with low circuit depth. Here, we extend our idea of Hamiltonian learning and propose to learn the quantum state of a system using PQC based on the knowledge of the time dynamics of various observables obtained from the evolution of the state under different random Hamiltonians. We start with a hardware-efficient ansatz whose parameters are variationally optimized to satisfy the constraints of the problem. We show the learning of random n-qubit quantum states for up to 6 qubits.

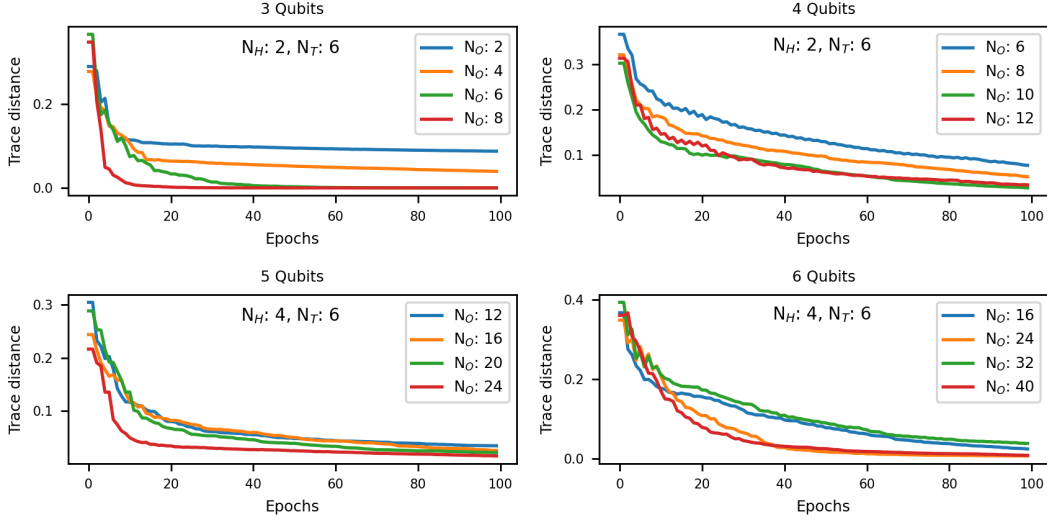As opposed to varying the number of random initial quantum states, for state learning,

Figure 10: Trace distance between the true and the learned quantum state for various n-qubit quantum states using the time dynamics approach for fixed N_H, N_T, and varying N_O. Even though an informationally incomplete set of observables are used but still the time dynamics approach is able to reconstruct the quantum state close to the true state, thereby validating the proposed formalism.

we start with a parameterized quantum state and evolve it under different Hamiltonians chosen at random for various time steps and compute the expectation values of different observables. The evolution operator for the Hamiltonian can be implemented on a quantum circuit using Trotterization as discussed in Appendix 5.1. However, for the proof of concept, we numerically evaluate the evolution operator since we are dealing with only a few qubits. Here, the choice of observables whose time dynamics we consider in the optimization protocol consists of the operators corresponding to the probabilities and coherences of the density matrix. Ideally, the number of observables required for an accurate and unique representation of a quantum state grows exponentially with the system size. Also, not all observables can be obtained with high fidelity in an experiment. This is why our method can prove to be advantageous as if we can obtain the time dynamics of fewer easily measurable observables then our protocol can learn the quantum state of a system with high fidelity. The cost function, in this case, is similar to that in Hamiltonian learning and is discussed in Section 2. We minimize the mean square error between the time dynamics of the observ-

ables obtained from the true state and the reconstructed state. The comprehensive results corresponding to the variation of the trace distance $(\mathrm{T}(\rho, \sigma) = \frac{1}{2}||\rho - \sigma||_1)$ between the true and the learned state during optimization is shown in Appendix 5.4. We vary the number of random Hamiltonians (N_H) under which the quantum state is evolved, the number of time steps (N_T) for evolution, and the number of observables (N_O) to obtain the converge set of parameters for which the trace distance is within an error threshold. However as before, in an experiment, we won't have access to the trace distance as the true state is unknown so we can again go back to the concept of a validation error. As we learn the quantum state upon optimization for one set of (N_T, N_H, N_O) we then evolve it under a known Hamiltonian and obtain the time dynamics of a new observable not used during learning and compute the validation error. Depending on the bound of the error threshold, one can choose the parameters for (N_T, N_H, N_O). In Figure 10 we see the variation of the trace distance with the number of epochs for quantum states prepared on various n-qubit quantum systems. To show the accuracy of reconstructing the quantum state, we compute the evolution of the reconstructed state and the true state under a random Hamiltonian and plot the time dynamics of X and Z-magnetic fields' observables in Figure 11.

3.4 Generalized Hamiltonian learning for SU(3) group

Quantum information processing of qutrit systems, consisting of three-dimensional quantum states, is of special significance and has drawn considerable attention both theoretically and experimentally.[45] For example, the nitrogen-vacancy center in diamond,[46] three ground states of trapped ^{171}Yb$^+$ ion, etc.[47] can serve as a qutrit system on which various quantum operations can be performed. The geometry of the special unitary SU(3^n) group can be used to define a quantum gate on n-qutrit quantum state. The use of Lie algebraic operations provides a technical advantage as certain classes of molecules can be described by common Hamiltonian, differing only by linear parameters.[48] In Appendix 5.5 we derive the mathematical expression for the calculation of analytical gradients for SU(3) group Hamiltonians
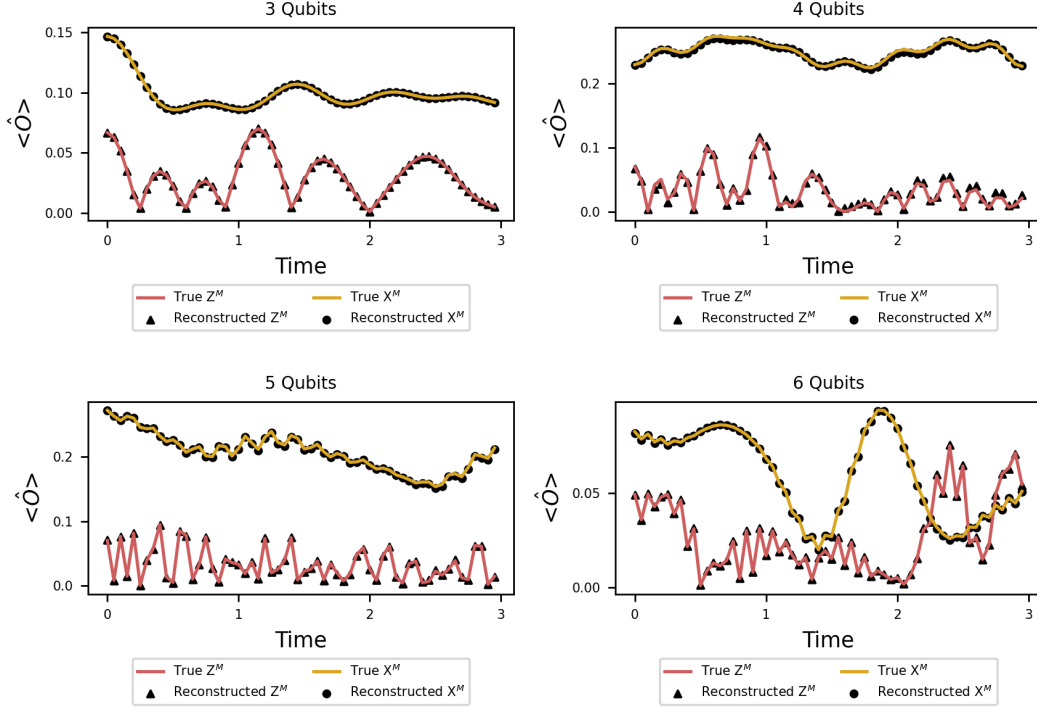
Figure 11: Time dynamics of X^M and Z^M-magnetic field observables under the evolution of a random Hamiltonian using the true and the learned quantum states for various n-qubit quantum states.

defined as $H = \sum_{i=1}^{8} c_i \lambda_i$, where λ_i correspond to Gell-Mann matrices that span the Lie algebra of $SU(3)$ group. Based on the cost function defined in Eq. (3), the gradients with respect to the coefficients $\{c_p\}$ are:

$$\frac{\partial Cost}{\partial c_p} = 2 \sum_{\alpha,i,k}^{N_O,N_S,N_T} \langle \psi_i | \frac{\partial U^\dagger(k\Delta t) O_\alpha U(k\Delta t)}{\partial c_p} | \psi_i \rangle \qquad (4)$$

Using the analytical gradients as shown in Appendix 5.5, it must be emphasized that the protocol so engendered is cost-effective as for a given choice of (O_α, ψ_i) (corresponding to each term within Eq. (4)) the set of quantum measurements necessary to construct Eq. (4) $\forall k$ time steps can be expressed in terms of 9 measurements for $SU(3)$ (for a general $SU(N)$ this number will be N^2) evaluated from the quantum circuit using the initial state $|\psi_i\rangle$ alone provided $O_\alpha \in \{\lambda_i\}_{i=1}^9$. For each of these 9 measurements, the primitive observables that need to be measured are the generators $\{\lambda_i\}_{i=1}^9$ defined before. One must note that

17

this obviates the need to perform repetitive measurements using the time-evolved version of $U(k\Delta t)$ for all of the successive N_T time steps as is usually necessitated for systems wherein such a Lie-algebraic structure is not present. In other words, the total number of measurements to construct Eq. (14) in its entirety will then simply be $9N_O N_S$ which is independent of N_T. Thus, using the structural properties of the SU(N) group, we can further extend this approach efficiently for Hamiltonian learning of multi-dimensional qudit quantum systems as well.

4 Concluding remarks

Measuring the time dependence of the mean values of observables enables us to solve two complementary problems. First and foremost we seek to reconstruct the Hamiltonian that generated the dynamics of various quantum systems. On the flip side we characterize the state of the evolving system. We provide numerical results showing how the method works. We used Trotterization to implement the evolution operator onto the quantum circuit and the entire workhorse of our proposed formalism is centered around variational optimization using shallow-depth quantum circuits and thus, is easily implementable on a near-term quantum device. We presented a comprehensive list of results corresponding to the Hamiltonian learning of various 2-local Hamiltonians as well as state learning of quantum systems comprising up to 6 qubits where we illustrate the significance of using time dynamics within the optimization process. The simple nature of our proposed approach can find its application in a variety of different areas where the dynamics of various observables need to be studied. As a subsidiary validation we examine a special (but general) case where the problem can be solved analytically. In Appendix 5.5, we propose to learn the Hamiltonian of multi-dimensional qudit systems by exploiting the structural properties of SU(N) groups and specifically show the analytical gradient calculation procedure for qutrit systems defined by SU(3) geometry. We also intend to further investigate the utility of the methods

developed here for studying time series datasets in typical machine learning algorithms that are commonly tackled using long short-term memory models (LSTM)[49] and recurrent neural network models (RNN).[50]

5   Appendix

5.1   Trotterized time evolution of a Hamiltonian

We consider the Heisenberg model on a 1D lattice with nearest-neighbor interactions whose Hamiltonian can be defined as $H = H_{xx} + H_{yy} + H_{zz}$ where

$$H_{xx} = \sum_i J\sigma_x^i \sigma_x^{i+1} \quad H_{yy} = \sum_j J\sigma_y^j \sigma_y^{j+1} \quad H_{zz} = \sum_k J\sigma_z^k \sigma_z^{k+1}$$

We are interested in evaluating the evolution operator $U = exp(-iHt)$ but the non-commutation of the Pauli operators doesn't allow the operator U to be written as the product of simpler exponentials as $U \neq exp(-iH_{xx}t)exp(-iH_{yy}t)exp(-iH_{zz}t)$. This is where Trotterization can be used to approximate U by splitting it into the product of simpler exponentials using N trotter steps giving an approximation up to second order in t/N as $U \approx \left[exp(-iH_{xx}t/N)exp(-iH_{yy}t/N)exp(-iH_{zz}t/N)\right]^N$. Here, we show Trotterization specifically for the nearest neighbor XYZ Hamiltonian but it can be generalized for others.

 Let us consider a Hamiltonian as defined above for an 8-qubit quantum system. For nearest neighbor coupling each qubit couples with a maximum of 2 qubits. We can thus split the execution into 2 layers such that qubits in the first layer couple with exactly one other qubit. This ensures that the overall depth of the circuit is constant and independent of the number of qubits. It scales only with the degree of the vertex which is 2 (constant) here. Figure 12 shows the quantum circuit implementation of one-step Trotterization, using only 1 and 2-qubit quantum gates, that can be repeated N times for full-time evolution. The number of 2 qubit gates scales as $4n$-4 for the construction as shown in Figure 13. The advantage
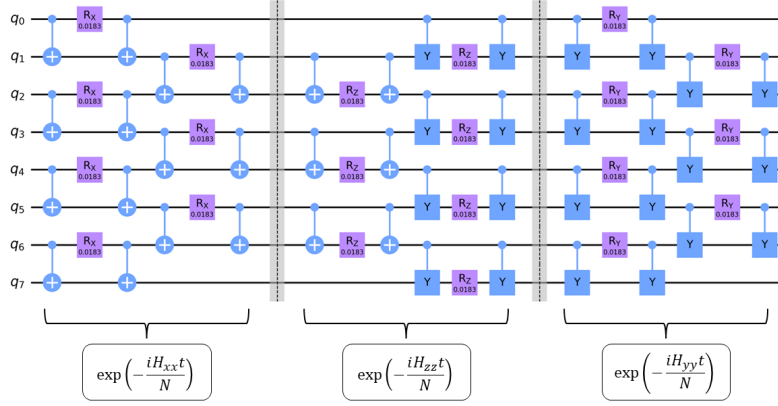
Figure 12: Quantum circuit corresponding to one step of Trotterization for the time evolution of a Hamiltonian comprising of only nearest-neighbors interactions.
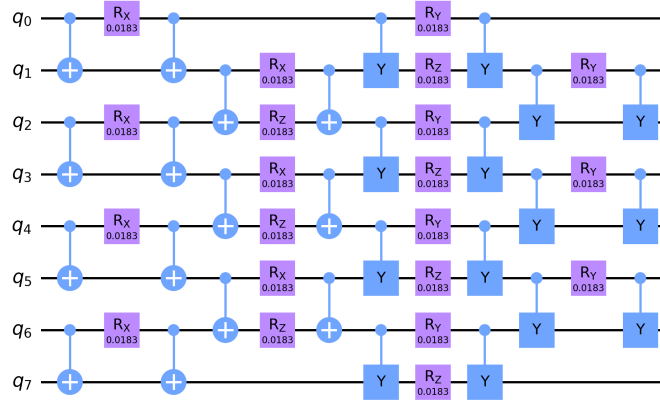

Figure 13: Quantum circuit for one step Trotterization obtained by eliminating the unnecessary CNOT and CY gates from Figure 12 to reduce the circuit depth from 18 to 12.

of using this scheme of Trotterization is evident from the circuit diagram in Figure 12 as essentially the CNOT and CY gates around the shown barrier are not required and so the circuit depth can be reduced as can be seen in Figure 13.

5.2 Gradients calculation on quantum circuit

Since our ansatz makes use of Pauli rotation gates we instead can compute them analytically. The most general Pauli rotation operator is given by $exp(-i\vec{n}.\vec{\sigma}\theta)$, where $\vec{\sigma} = (I, \sigma_x, \sigma_y, \sigma_z)$,$\vec{n} = (n_0, n_1, n_2, n_3)$ and θ is the angle of rotation. Using the identity that $exp^{-i\vec{n}.\vec{\sigma}\theta} = cos(\theta)n_0I + isin(\theta)(n_1\sigma_x + n_1\sigma_y + n_1\sigma_z)$ one can straightforwardly derive an expression for

the gradients,

$$
\begin{aligned}
\frac{d \langle \psi | H | \psi \rangle}{d\theta_i} &= \frac{d \langle \psi |}{d\theta_i} H | \psi \rangle + \langle \psi | H \frac{d | \psi \rangle}{d\theta_i} \\
&= \mathrm{Re}(\frac{d \langle \psi |}{d\theta_i} H | \psi \rangle) \quad \text{(Since H is hermitian)}
\end{aligned}
\tag{5}
$$

Now since our unitaries are a function of $R_y(\theta)$ we have,

$$
\begin{aligned}
\frac{dR_y}{d\theta} &= \frac{de^{i\sigma_y \theta/2}}{d\theta} = \frac{1}{2} i\sigma_y e^{i\sigma_y \theta/2} \\
&= \frac{1}{2} e^{i\sigma_y \pi/2} e^{i\sigma_y \theta/2} = \frac{1}{2} R_y(\pi) R_y(\theta) \\
&= \frac{1}{2} R_y(\theta + \pi)
\end{aligned}
\tag{6}
$$

Thus computing the gradients of the cost function with respect to parameters of Pauli rotation gates can be straightforwardly evaluated on the circuit with π shift of the corresponding parameter. Alternatively for a unitary $U = exp(-i\frac{\theta}{2}P_i)$, where P_i is a pauli operator, one could use the expression from parameter shift rules[51] to compute the gradient of a cost function $C(\alpha, \beta) = \langle \psi | U^\dagger(\alpha, \beta) O U(\alpha, \beta) | \psi \rangle$ given by,

$$
\frac{dC(\alpha, \beta)}{d\alpha} = \frac{1}{2} \left[C(\alpha + \frac{\pi}{2}, \beta) - C(\alpha - \frac{\pi}{2}, \beta) \right]
\tag{7}
$$

5.3 Cost function and convergence criteria for Hamiltonian learning

The recipe for Hamiltonian learning in our case involves the experimental knowledge of the time dynamics of expectation values of correlation functions starting with multiple random quantum states. The number of correlation functions and random initial states required for Hamiltonian reconstruction depends on the error threshold ϵ, that the experimentalist is interested in, between the true and the reconstructed Hamiltonian and can be evaluated

using the validation error. As discussed in Section 2, let $O_\alpha(k\Delta t)$ refer to the measurement of observable O_α made at timestep k, where Δt refers to the time interval of each timestep. Then, the cost function is defined as the 2-norm function between $O_\alpha(k\Delta t)$ and $O_{\alpha,i}^{obs}(k\Delta t)$ over the states $|\psi_i\rangle$:

$$\text{Cost} = \sum_{\alpha,i,k} (\langle\psi_i|\, U^\dagger(k\Delta t)\, O_\alpha\, U(k\Delta t)\, |\psi_i\rangle - O_{\alpha,i,k}^{obs})^2 \tag{8}$$

where $U(k\Delta t)$ is the unitary operator that evolves the quantum state under a Hamiltonian H. As shown in Section 3, the variables in our proposed protocol that affect the accuracy of Hamiltonian learning are the number of quantum states (N_S), the number of time steps (N_T) up to which we consider the time dynamics, and the number of observables (N_O) whose time dynamics we consider in the cost function. In section 3 we have shown the results corresponding to the trace distance between the true and the learned Hamiltonian and also defined the validation error for choosing the correct set of (N_S, N_T, N_O). In Figure 14 the cost function is plotted as a function of the epochs for different sets of (N_S, N_T, N_O), and the corresponding validation error is plotted in Figure 15 for the inhomogeneous 5-qubit TFIM Hamiltonian defined in Section 3.2.1. As can be seen the mean square error loss i.e. the cost converges to 0 in all cases yet the validation error is significantly high for accurate learning in various cases. This is because even though the cost converges to 0 for fewer starting states, observables, and time steps but that information is insufficient to have accurate learning of the Hamiltonian. Thus the concept of validation error is important as based on it we can check whether or not the learned Hamiltonian, obtained from a given set of (N_S, N_T, N_O), is close to the true Hamiltonian.

5.4 Quantum state learning from time dynamics of observables

A natural extension of the idea of using time dynamics of various observables for Hamiltonian learning can be applied to solving the inverse problem of learning an unknown quantum
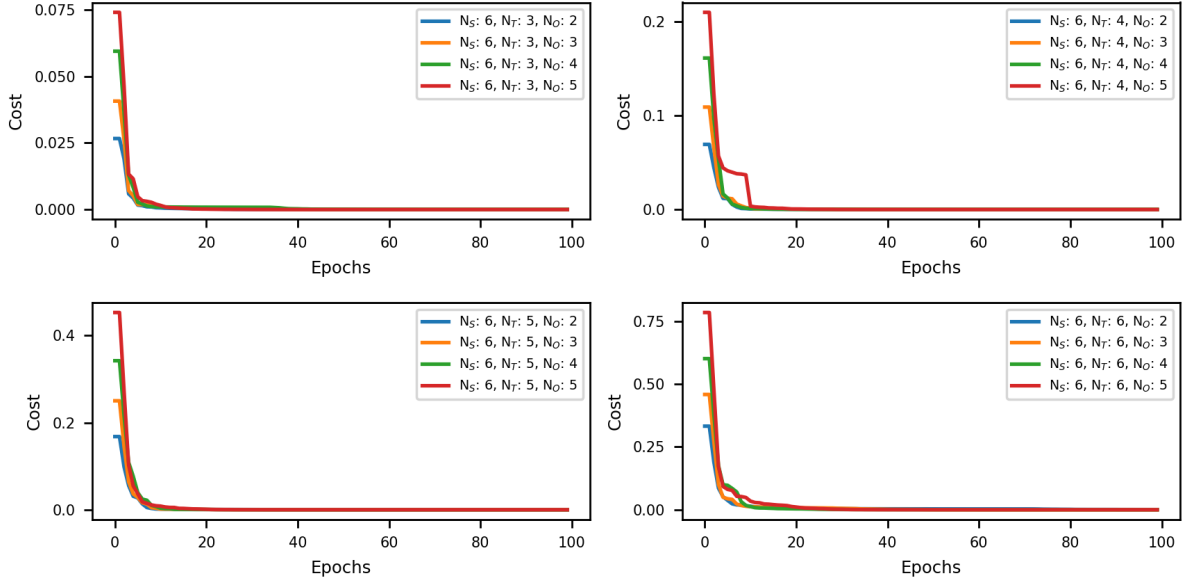
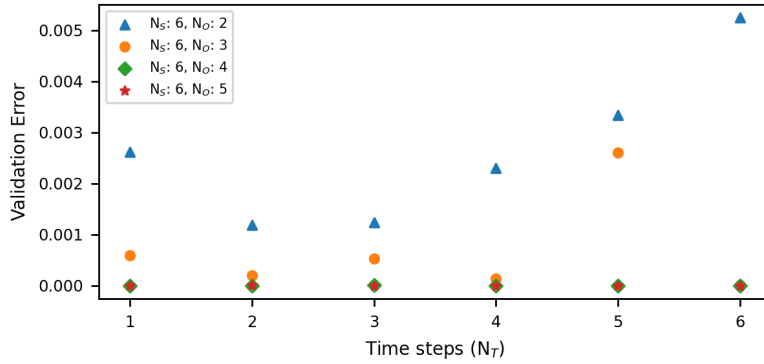Figure 14: The mean square error (cost) as a function of the epochs for various sets of (N_S, N_T, N_O).

Figure 15: The validation error plot showing that even though the cost function converges to 0 in Figure 14 for all sets of (N_S, N_T, N_O) yet the corresponding learned Hamiltonian is close to the true Hamiltonian for those sets for which the validation error converges to 0.

state of a given system. Here, instead of starting with multiple random quantum states, we have at our disposal multiple random Hamiltonians (N_H) that are used for the evolution of the quantum state and the expectation values of different observables (N_O) are recorded at various time steps (N_T). The observables here correspond to the probabilities and coherences

of the n-qubit quantum system:

$$\{|1\rangle\langle 1|, |2\rangle\langle 2|, |3\rangle\langle 3|, |4\rangle\langle 4|, (|1\rangle\langle 2| \pm |2\rangle\langle 1|), (|1\rangle\langle 3| \pm |3\rangle\langle 1|), (|1\rangle\langle 4| \pm |4\rangle\langle 1|),$$

$$(|2\rangle\langle 3| \pm |3\rangle\langle 2|), (|2\rangle\langle 4| \pm |4\rangle\langle 2|), (|3\rangle\langle 4| \pm |4\rangle\langle 3|)\} \tag{9}$$

The operators in Eq. (9) form the informationally complete (IC) Hermitian operators
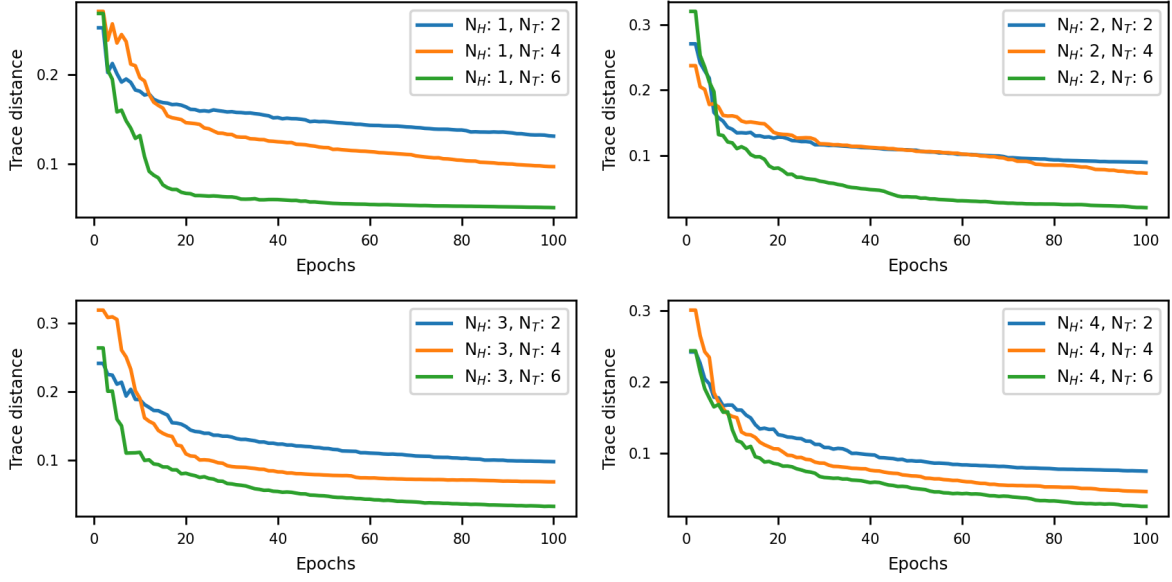

Figure 16: Trace distance between the true and the learned state is plotted as a function of the epochs during the optimization process for fixed $N_O = 16$ and varying N_H, N_T. As is evident, using time dynamics of observables in the optimization process improves the learning of the quantum state.
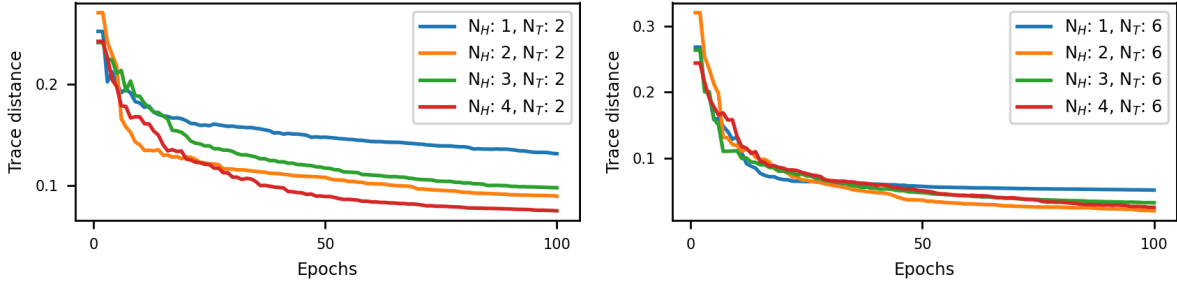

Figure 17: The effect of varying N_H for fixed N_T on state learning.

whose expectation values are the probabilities and coherences of a 2-qubit quantum system. We consider the time dynamics of a few of these observables in our optimization protocol

24

as discussed in Section 2. Here, we present the results corresponding to the learning of a random 5-qubit quantum state. As was evident in the Hamiltonian learning protocol, for state learning as well, including the expectation values of observables at higher time steps in the optimization process improves the convergence of the learned state to the true state as shown in Figure 16. Also, to analyze the effect of varying the number of random Hamiltonians on state learning we see in Figure 17 that as N_H increases the learning improves significantly when N_T is small but for higher N_T the state can be learned using fewer Hamiltonians. Lastly, in Figure 18 we vary the number of observables for various N_T and as expected, we can learn the quantum state using fewer observables than the IC set if we include their time dynamics in the optimization protocol.
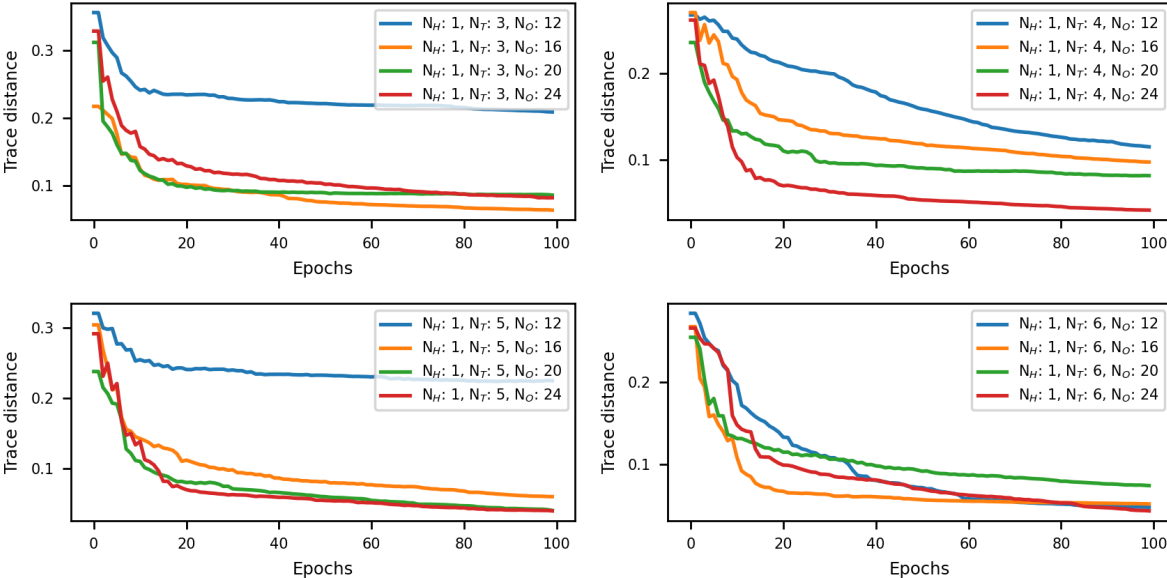

Figure 18: Varying the number of observables included during optimization shows that the accuracy of state learning improves upon increasing N_O. However, we can achieve similar accuracy using fewer observables if we have the knowledge of observables at higher time steps.

5.5 Quantum Hamiltonian learning of the dynamics generated by the SU(3) group using analytical gradients

The Hamiltonian of such a system can be expressed in terms of Gell-Mann matrices: $H = \sum_{i=1}^{8} c_i \lambda_i$, where λ_i are the generators of the SU(3) group:

$$\lambda_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \lambda_2 = \begin{bmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \lambda_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \lambda_4 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

$$\lambda_5 = \begin{bmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{bmatrix}, \lambda_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \lambda_7 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{bmatrix}, \lambda_8 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

Now, in order to learn the Hamiltonian of such a quantum system we can perform optimization to learn the coefficients $\{c_i\}$. One of the ways to do so is to use our proposed formalism of using time dynamics of various observables sampled using different quantum states and optimizing the coefficients using the numerical approach of gradient calculation as discussed in Section 5.2 based on the cost function defined in Eq. (8). However, as a generalized approach, we can also calculate the gradients analytically. There can be different approaches to doing so. We, hereby, propose an approach based on a version of the Baker–Campbell–Hausdorff (BCH) formula:

$$e^A B e^{-A} = B + [A, B] + \frac{1}{2!}[A, [A, B]] + \frac{1}{3!}[A, [A, [A, B]] + \ldots + \frac{1}{n!}[A, [A, \ldots, [A, B]], \ldots] \quad (10)$$

The unitary operator for the state evolution under the Hamiltonian H is $U = \exp\{-i \sum_{i=1}^{8} c_i \lambda_i t\}$. To compute the gradients of the cost function in Eq. (8) analytically, we use the BCH ex-

pansion of $U^\dagger O_\alpha U$ as:

$$U^\dagger O_\alpha U = O_\alpha + it \left[\sum_k c_k \lambda_k, O_\alpha \right] + \frac{(it)^2}{2!} \left[\sum_k c_k \lambda_k, \left[\sum_{k'} c_{k'} \lambda_{k'}, O_\alpha \right] \right] + \ldots$$
$$+ \frac{(it)^n}{n!} \left[\sum_k c_k \lambda_k, \left[\sum_{k'} c_{k'} \lambda_{k'}, \ldots, \left[\sum_{k^{n'}} c_{k^{n'}} \lambda_{k^{n'}}, O_\alpha \right] \right] \right], \ldots, \right] \tag{11}$$

Now, we take the derivative of $U^\dagger O_\alpha U$ with respect to the coefficients $\{c_p\}$:

$$\frac{\partial U^\dagger O_\alpha U}{\partial c_p} = it \left[\lambda_p, O_\alpha \right] + \frac{(it)^2}{2!} \left(\left[\sum_{k' \neq p} c_{k'} \lambda_p, [\lambda_{k'}, O_\alpha] \right] + 2c_p [\lambda_p, [\lambda_p, O_\alpha]] \right) + \ldots$$
$$+ \frac{(it)^n}{n!} (\left[\sum_{k' \neq p} \sum_{k'' \neq p} \cdots \sum_{k^{n'} \neq p} c_{k'} c_{k''} \ldots c_{k^{n'}} \lambda_p [\lambda_{k'}, [\lambda_{k''}, \ldots, [\lambda_{k^{n'}}, O_\alpha]] \ldots \right]$$
$$+ \left[\sum_{k'' \neq p} \sum_{k''' \neq p} \cdots \sum_{k^{n'} \neq p} 2c_p c_{k''} c_{k'''} \ldots c_{k^{n'}} \lambda_p, [\lambda_p, [\lambda_k'', \ldots, [\lambda_{k^{n'}}, O_\alpha]], \ldots \right]$$
$$+ [nc_p^{n-1} \lambda_p, [\lambda_p, \ldots [\lambda_p, O_\alpha]] \ldots]) \tag{12}$$

The Gell-Mann matrices satisfy the commutation relationship:

$$[\lambda_a, \lambda_b] = 2i \sum_c f^{abc} \lambda_c \tag{13}$$

where f^{abc} are the structure constants given by: $f^{abc} = -\frac{1}{4} i (tr(\lambda_a [\lambda_b, \lambda_c]))$. Thus, depending on the required accuracy level, one can fix the value of n and calculate $\frac{\partial U^\dagger O_\alpha U}{\partial c_p}$. Thereby, we can calculate the gradients as:

$$\frac{\partial Cost}{\partial c_p} = 2 \sum_{\alpha,i,k}^{N_O,N_S,N_T} \langle \psi_i | \frac{\partial U^\dagger(k\Delta t) O_\alpha U(k\Delta t)}{\partial c_p} | \psi_i \rangle \tag{14}$$

Thus, substituting Eq. (12) in Eq. (14) and taking advantage of the structural relationship of SU(N) systems, the gradients can be computed with measurements that are independent of the time steps of the evolution. This protocol can be extended to SU(N) systems as well

and we intend to investigate it further in future works.

Acknowledgement

References

(1) Nielsen, M. A.; Chuang, I. L. *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed.; Cambridge University Press: USA, 2011.

(2) Kais, S. *Quantum Information and Computation for Chemistry*; Wiley and Sons: Hoboken: NJ, 2014; Vol. 154.

(3) James, D. F. V.; Kwiat, P. G.; Munro, W. J.; White, A. G. Measurement of qubits. *Phys. Rev. A* **2001**, *64*, 052312.

(4) Banaszek, K.; Cramer, M.; Gross, D. Focus on quantum tomography. *New J. Phys.* **2013**, *15*, 125020.

(5) Paris, M.; Rehacek, J. *Quantum state estimation*; Springer Science & Business Media, 2004; Vol. 649.

(6) Artiles, L. M.; Gill, R. D.; Guṭă, M. An invitation to quantum tomography. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2005**, *67*, 109–134.

(7) Yu, W.; Sun, J.; Han, Z.; Yuan, X. Practical and Efficient Hamiltonian Learning. *arXiv preprint arXiv:2201.00190* **2022**,

(8) Haah, J.; Kothari, R.; Tang, E. Optimal learning of quantum Hamiltonians from high-temperature Gibbs states. *arXiv preprint arXiv:2108.04842* **2021**,

(9) Krastanov, S.; Zhou, S.; Flammia, S. T.; Jiang, L. Stochastic estimation of dynamical variables. *Quantum Science and Technology* **2019**, *4*, 035003.

(10) Evans, T. J.; Harper, R.; Flammia, S. T. Scalable bayesian hamiltonian learning. *arXiv preprint arXiv:1912.07636* **2019**,

(11) Bairey, E.; Arad, I.; Lindner, N. H. Learning a local Hamiltonian from local measurements. *Physical review letters* **2019**, *122*, 020504.

(12) Qi, X.-L.; Ranard, D. Determining a local Hamiltonian from a single eigenstate. *Quantum* **2019**, *3*, 159.

(13) Poyatos, J. F.; Cirac, J. I.; Zoller, P. Complete Characterization of a Quantum Process: The Two-Bit Quantum Gate. *Phys. Rev. Lett.* **1997**, *78*, 390–393.

(14) D'Ariano, G. M.; Lo Presti, P. Quantum Tomography for Measuring Experimentally the Matrix Elements of an Arbitrary Quantum Operation. *Phys. Rev. Lett.* **2001**, *86*, 4195–4198.

(15) Altepeter, J. B.; Branning, D.; Jeffrey, E.; Wei, T. C.; Kwiat, P. G.; Thew, R. T.; O'Brien, J. L.; Nielsen, M. A.; White, A. G. Ancilla-Assisted Quantum Process Tomography. *Phys. Rev. Lett.* **2003**, *90*, 193601.

(16) D'Ariano, G. M.; Lo Presti, P. Imprinting Complete Information about a Quantum Channel on its Output State. *Phys. Rev. Lett.* **2003**, *91*, 047902.

(17) Ekert, A. K.; Alves, C. M.; Oi, D. K. L.; Horodecki, M.; Horodecki, P.; Kwek, L. C. Direct Estimations of Linear and Nonlinear Functionals of a Quantum State. *Phys. Rev. Lett.* **2002**, *88*, 217901.

(18) Horodecki, P.; Ekert, A. Method for Direct Detection of Quantum Entanglement. *Phys. Rev. Lett.* **2002**, *89*, 127902.

(19) Bovino, F. A.; Castagnoli, G.; Ekert, A.; Horodecki, P.; Alves, C. M.; Sergienko, A. V. Direct Measurement of Nonlinear Properties of Bipartite Quantum States. *Phys. Rev. Lett.* **2005**, *95*, 240407.

(20) Tóth, G.; Wieczorek, W.; Gross, D.; Krischek, R.; Schwemmer, C.; Weinfurter, H. Permutationally Invariant Quantum Tomography. *Phys. Rev. Lett.* **2010**, *105*, 250403.

(21) Gupta, R.; Levine, R. D.; Kais, S. Convergence of a Reconstructed Density Matrix to a Pure State Using the Maximal Entropy Approach. *The Journal of Physical Chemistry A* **2021**, *125*, 7588–7595.

(22) Gupta, R.; Xia, R.; Levine, R. D.; Kais, S. Maximal entropy approach for quantum state tomography. *PRX Quantum* **2021**, *2*, 010318.

(23) Gupta, R.; Sajjan, M.; Levine, R. D.; Kais, S. Variational approach to quantum state tomography based on maximal entropy formalism. *Phys. Chem. Chem. Phys.* **2022**, *24*, 28870–28877.

(24) Aaronson, S. Shadow Tomography of Quantum States. 2017; `https://arxiv.org/abs/1711.01053`.

(25) Chen, S.; Huang, B.; Li, J.; Liu, A.; Sellke, M. Tight bounds for state tomography with incoherent measurements. *arXiv preprint arXiv:2206.05265* **2022**,

(26) Huang, H.-Y.; Kueng, R.; Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nature Physics* **2020**, *16*, 1050–1057.

(27) Schuch, N.; Cirac, I.; Verstraete, F. Computational difficulty of finding matrix product ground states. *Physical review letters* **2008**, *100*, 250501.

(28) Bilgin, E.; Boixo, S. Preparing thermal states of quantum systems by dimension reduction. *Physical review letters* **2010**, *105*, 170405.

(29) McClean, J. R.; Romero, J.; Babbush, R.; Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics* **2016**, *18*, 023023.

(30) Bravo-Prieto, C.; LaRose, R.; Cerezo, M.; Subasi, Y.; Cincio, L.; Coles, P. J. Variational quantum linear solver. *arXiv preprint arXiv:1909.05820* **2019**,

(31) Huang, H.-Y.; Bharti, K.; Rebentrost, P. Near-term quantum algorithms for linear systems of equations. *arXiv preprint arXiv:1909.07344* **2019**,

(32) LaRose, R.; Tikku, A.; O'Neel-Judy, É.; Cincio, L.; Coles, P. J. Variational quantum state diagonalization. *npj Quantum Information* **2019**, *5*, 1–10.

(33) Cerezo, M.; Sharma, K.; Arrasmith, A.; Coles, P. J. Variational quantum state eigensolver. *arXiv preprint arXiv:2004.01372* **2020**,

(34) Peruzzo, A.; McClean, J.; Shadbolt, P.; Yung, M.-H.; Zhou, X.-Q.; Love, P. J.; Aspuru-Guzik, A.; O'brien, J. L. A variational eigenvalue solver on a photonic quantum processor. *Nature communications* **2014**, *5*, 1–7.

(35) Chen, R.; Song, Z.; Zhao, X.; Wang, X. Variational quantum algorithms for trace distance and fidelity estimation. *Quantum Science and Technology* **2021**, *7*, 015019.

(36) Sureshbabu, S. H.; Sajjan, M.; Oh, S.; Kais, S. Implementation of Quantum Machine Learning for Electronic Structure Calculations of Periodic Systems on Quantum Computing Devices. *Journal of Chemical Information and Modeling* **2021**,

(37) Sajjan, M.; Sureshbabu, S. H.; Kais, S. Quantum machine-learning for eigenstate filtration in two-dimensional materials. *Journal of the American Chemical Society* **2021**, *143*, 18426–18445.

(38) Selvarajan, R.; Sajjan, M.; Kais, S. Variational quantum circuits to prepare low energy symmetry states. *Symmetry* **2022**, *14*, 457.

(39) Dixit, V.; Selvarajan, R.; Aldwairi, T.; Koshka, Y.; Novotny, M. A.; Humble, T. S.; Alam, M. A.; Kais, S. Training a quantum annealing based restricted boltzmann machine on cybersecurity data. *IEEE Transactions on Emerging Topics in Computational Intelligence* **2021**, *6*, 417–428.

(40) Sajjan, M.; Li, J.; Selvarajan, R.; Sureshbabu, S. H.; Kale, S. S.; Gupta, R.; Singh, V.; Kais, S. Quantum machine learning for chemistry and physics. *Chem. Soc. Rev.* **2022**, *51*, 6475–6573.

(41) Cao, Y.; Romero, J.; Olson, J. P.; Degroote, M.; Johnson, P. D.; Kieferová, M.; Kivlichan, I. D.; Menke, T.; Peropadre, B.; Sawaya, N. P., et al. Quantum chemistry in the age of quantum computing. *Chemical reviews* **2019**, *119*, 10856–10915.

(42) Crooks, G. E. Gradients of parameterized quantum gates using the parameter-shift rule and gate decomposition. *arXiv preprint arXiv:1905.13311* **2019**,

(43) Selvarajan, R.; Dixit, V.; Cui, X.; Humble, T. S.; Kais, S. Prime factorization using quantum variational imaginary time evolution. *Scientific reports* **2021**, *11*, 1–8.

(44) Abraham, H. et al. Qiskit: An Open-source Framework for Quantum Computing. 2019.

(45) Li, B.; Yu, Z.-H.; Fei, S.-M. Geometry of quantum computation with qutrits. *Scientific reports* **2013**, *3*, 1–6.

(46) Gardill, A.; Cambria, M. C.; Kolkowitz, S. Fast relaxation on qutrit transitions of nitrogen-vacancy centers in nanodiamonds. *Physical Review Applied* **2020**, *13*, 034010.

(47) Lucarelli, D. Chow's theorem and universal holonomic quantum computation. *Journal of Physics A: Mathematical and General* **2002**, *35*, 5107.

(48) Iachello, F.; Levine, R. D. *Algebraic theory of molecules*; Oxford University Press, 1995.

(49) Staudemeyer, R. C.; Morris, E. R. Understanding LSTM – a tutorial into Long Short-Term Memory Recurrent Neural Networks. 2019; `https://arxiv.org/abs/1909.09586`.

(50) Sherstinsky, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena* **2020**, *404*, 132306.

(51) Schuld, M.; Bergholm, V.; Gogolin, C.; Izaac, J.; Killoran, N. Evaluating analytic gradients on quantum hardware. *Physical Review A* **2019**, *99*, 032331.