Optimal and Robust In-situ Quantum Hamiltonian Learning through Parallelization

Suying Liu,[1,2,*] Xiaodi Wu,[1,2] and Murphy Yuezhen Niu[3,4,†]

[1] *Joint Center for Quantum Information and Computer Science, University of Maryland, College Park, Maryland 20742, USA*
[2] *Department of Computer Science, University of Maryland, College Park, USA*
[3] *Department of Computer Science, University of California, Santa Barbara, Santa Barbara, CA 93106, USA*
[4] *Google Quantum AI, Venice, California, CA 90291, USA*

(Dated: October 10, 2025)

Hamiltonian learning is a cornerstone for advancing accurate many-body simulations, improving quantum device performance, and enabling quantum-enhanced sensing. Existing readily deployable quantum metrology techniques primarily focus on achieving Heisenberg-limited precision in one- or two-qubit systems. In contrast, general Hamiltonian learning theories address broader classes of unknown Hamiltonian models but are highly inefficient due to the absence of prior knowledge about the Hamiltonian. There remains a lack of efficient and practically realizable Hamiltonian learning algorithms that directly exploit the known structure and prior information of the Hamiltonian, which are typically available for a given quantum computing platform. In this work, we present the first Hamiltonian learning algorithm that achieves both Cramér–Rao lower bound saturated optimal precision and robustness to realistic noise, while exploiting device structure for quadratic reduction in experimental cost for fully connected Hamiltonians. Moreover, this approach enables simultaneous in-situ estimation of all Hamiltonian parameters without requiring the decoupling of non-learnable interactions during the same experiment, thereby allowing comprehensive characterization of the system's intrinsic contextual errors. Notably, our algorithm does not require deep circuits and remains robust against both depolarizing noise and time-dependent coherent errors. We demonstrate its effectiveness with a detailed experimental proposal along with supporting numerical simulations on Rydberg atom quantum simulators, showcasing its potential for high-precision Hamiltonian learning in the NISQ era.

I. INTRODUCTION

Learning the underlying Hamiltonian of a quantum system is a foundational problem with broad implications across quantum simulation, quantum control, and fault-tolerant quantum computing. Accurate Hamiltonian characterization enables high-fidelity simulation of quantum many-body dynamics [1–6], facilitates the calibration and benchmarking of quantum hardware [7–10], and underpins the demonstration of quantum advantage in scalable fault-tolerant quantum computers [11–14]. In this context, parallel in-situ Hamiltonian learning has emerged as a critical component as the size and model complexity of current quantum devices increase dramatically for achieving fault-tolerant computing and completing practical applications. Parallelization means we can learn multiple Hamiltonian parameters with the same experiment, which is essential for efficiently processing the exponentially large Hilbert spaces encountered in large-scale quantum systems [15]. In-situ learning enables characterization of specific Hamiltonian parameters without requiring the deactivation of other couplings. It is needed for characterizing contextual and correlated errors, such as quantum crosstalk, that cannot be captured by ex-situ methods. These capabilities make parallel in-situ learning indispensable for reliable and scalable quantum computation.

Parallel and in-situ Hamiltonian learning are crucial for resolving key challenges for quantum simulators. For example, in Rydberg atom devices, atom position errors, caused by turning off optical traps during evolution, significantly impact simulation accuracy as the interaction strength scales with distance as $c_{ij} = C_6/R_{ij}^6$. This has led to notable discrepancies between theoretical predictions and experimental results in many-body dynamics [1]. A first-principles understanding of these mismatches via in-situ Hamiltonian learning is required to bridge the gap and validate error models for noisy quantum simulators, paving the way for the study of complex quantum many-body systems in the NISQ era.

Previous works on Hamiltonian learning [18–22] have yet to demonstrate parallel in-situ learning of many-body Hamiltonians and have not provided readily deployable protocols that maintain robustness under realistic noise conditions. Instead, these studies have primarily focused on establishing fundamental bounds for general learning tasks, without offering protocols that are directly applicable to current quantum devices. Existing quantum hardware suffers from diverse noise sources—including decoherence, systematic coherent errors, and time-dependent drift [23], which severely degrade the performance of learning proposals that are only robust to state preparation and measurement (SPAM) errors. Further, this line of approaches depends on the Hamiltonian Reshaping [18]

* Electronic address: syliu@umd.edu
† Electronic address: murphyniu@ucsb.edu

	Heisenberg limit	Robust	In-situ	# experiment rounds	Applicability
[16]	✓	✗	-	-	digital
[17]	✓	✓	-	-	digital
[18]	✓	✗	✗	$O(n^2)$	hybrid
This work [Algorithm 2]	✓	✓	✓	$O(n)$	digital and hybrid
This work [Algorithm 3]	✓	✓	✓	$O(n^2)$	analog

TABLE I: Comparison of this work and previous methods: (i) ability to reach the Heisenberg limit and robustness against SPAM, decoherence, and coherent errors; (ii) performance in learning fully connected Hamiltonians, including in-situ learning ability, required experiment rounds, and quantum device requirement for implementing the algorithm. '-' represents the method is not applicable at this scenario.

technique, which decomposes the many-body Hamiltonian into non-interacting clusters. As a result, these methods do not support in-situ Hamiltonian learning in general scenarios. In practice, several works in quantum metrology and quantum calibration have addressed the practical task of learning Hamiltonian coefficients as a means to certify quantum simulators [1, 3, 16, 17, 24]. However, these protocols are limited to learning two-qubit systems. For instance, the calibration method in [3] estimates coupling strengths between two superconducting qubits using pairwise measurements. Similarly, in [24], interaction terms are inferred through a four-photon process involving only two neighboring atoms. As a consequence, these methods are inherently limited to learning a single interaction term per qubit at a time. Therefore, without parallelization, $O(k)$ experiments are required to characterize couplings associated with a k-connected single qubit. Furthermore, because these approaches decompose the original Hamiltonian into pairwise interactions, they preclude in-situ learning and thus fail to capture contextual errors, such as unwanted correlation errors that emerge only when multiple qubits interact simultaneously. Despite these rapid advances, the realization of parallel and in-situ quantum computation remains an open challenge, both in theory and in practice.

In this work, we introduce the first in-situ Hamiltonian learning algorithm that incorporates both the structural prior of realistic systems and noise resilience. Specifically, we target a class of parallel-learnable Hamiltonians that cover a wide range of physical devices (Rydberg atom, trapped ion, and superconducting qubits), where subsets of parameters can be estimated simultaneously in-situ. The parallelization feature of our algorithm achieves a quadratic reduction in experimental cost with respect to the number of qubits, compared to previous approaches for estimating fully connected Hamiltonians. More specifically, our algorithm leverages parallelization to solve any set of k pairwise coupling terms that share the same qubit using only $O(1)$ experiments. Besides, our algorithm leverages Quantum Signal-Processing (QSP) to decouple the parameters, enabling robustness against SPAM, decoherence, and time-dependent coherent errors, and achieving optimal precision. The algorithm's optimality and robustness make it well-suited for learning device Hamiltonians on current quantum hardware. Moreover, the in situ feature of the learning algorithm allows us to fill in the blank in capturing the Rydberg atom distance error in many-body systems, serving as the first principles approach to validate the error model in theory and practice. To demonstrate its practicality, we present a detailed experimental proposal along with supporting numerical simulations for high-precision atom-distance estimation on Rydberg atom quantum simulators. A comprehensive comparison between our learning algorithms and previous works on learning Hamiltonian parameters is shown in Table I.

The rest of the paper is structured as follows. In Sec. II A, we define the parallel-learnable Hamiltonian and introduce specific Problem II.2 as the main focus for presenting our techniques. The main results of the work are summarized in Sec. II B. Sec. II C presents our learning algorithm with the introduction of the QSPE-based algorithm for two-atom learning and followed by the general parallel learning algorithm for n atom case. Sec. II D discusses the optimality of our algorithm with matching Cramér–Rao bound and computed variance. In Sec. III 1, we show the robustness of the algorithm against decoherence, coherent state preparation error, readout error, and time-dependent coherent error. In Sec. III 2, we present a detailed experimental proposal with corresponding parameter settings for the application of the algorithm on current Rydberg atom quantum simulators. We conclude in Sec. IV with open questions for further exploration.

II. RESULTS

A. Problem definition

Given that current quantum devices operate under well-defined Hamiltonian structures, accurately characterizing the Hamiltonian coefficients is essential for fully utilizing these systems. In this work, we address the problem of estimating the parameters of fixed-term Hamiltonians, where the structure of the terms is known, but the corresponding coefficients are unknown. Our learning technique applies to any general parallel-learnable Hamiltonian, defined as follows.

Definition II.1. *The Hamiltonian is parallel-learnable if it can be written as a direct sum of 2^{n-1} non-diagonal Hermitian matrices with dimension 2×2 up to a realizable unitary transformation. That is, for any Hamiltonian H which is parallel-learnable, there exists a realizable unitary transformation U, such that*

$$U^\dagger H U = \bigoplus_{i=1}^{2^{n-1}} A_i, \tag{1}$$

where A_i's are 2×2 non-diagonal Hermitian matrices.

From the definition, we note that the parallel-learnable structure requires the Hamiltonian H to have 2^{n-1} invariant subspaces under certain bases. The unitary transformation U is used to change it from the computational basis to which the block-diagonal structure is presented. In principle, any Hamiltonian can be transformed into a direct sum of non-diagonal 2×2 blocks through a specially designed unitary transformation U. A specific example of such a transformation is $U := RV$, where V is the eigenbasis matrix that diagonalizes the Hamiltonian, and R is a block-wise rotation that transforms each 2×2 diagonal block into a non-diagonal form. (See Supplementary Material Sec. I for more details on this construction for arbitrary Hamiltonians.) However, the unitary transformation utilized here is constructed from the eigenbasis matrix V of the Hamiltonian H, which is not always feasible to implement on current quantum hardware. Therefore, in this work, we focus on Hamiltonians that preserve the parallel-learnable structure, allowing for physically implementable unitary transformations. We present several illustrative examples of parallel-learnable Hamiltonians and their corresponding unitary transformations in the Supplementary Material Sec. I. We notice from those examples that parallel-learnable Hamiltonians frequently appear in many-body systems, particularly in the design of foundational quantum platforms. Examples include the transverse field Ising model, which captures the dynamics of Rydberg interactions [25, 26], and the XY model with local Z fields, commonly used to approximate superconducting qubit couplings and interactions in trapped ion gates [27–29]. Here, we mainly focus on the following Hamiltonian, targeting learning all the Hamiltonain coefficients in-situ and in parallel.

Definition II.2 (Problem 1)**.** *Consider the Hamiltonian of n-atom Rydberg arrays,*

$$H = \sum_{\substack{i,j \in \{1,\cdots,n\} \\ i \neq j}} c_{ij} Z_i Z_j, \tag{2}$$

which models pure atom-atom interaction between any two Rydberg atoms. There are in total $\frac{n(n-1)}{2}$ terms in the Hamiltonian, whose coefficients are characterized by the parameter vector $\mathbf{c} = [c_{11}, c_{12}, \cdots, c_{n(n-1)}]^T$. We aim to construct estimators $\hat{\mathbf{c}} = [\hat{c}_{11}, \hat{c}_{12}, \cdots, \hat{c}_{n(n-1)}]^T$ for parameter vector $\mathbf{c}$.

This problem is motivated by a fundamental challenge from the Rydberg-array-based quantum computers. The Rydberg atom quantum device is built on neutral atoms, which are trapped with optical tweezers [30]. But those traps are turned off during the evolution phase in quantum computing [25, 26], resulting in unavoidable atom position movements. Given that the interaction strength depends inversely on the atom distance, i.e.,$c_{ij} = C_6/R_{ij}^6$, small deviations in inter-atomic distances can lead to sixth-order amplification of errors in the Hamiltonian parameters. This amplification can cause significant discrepancies between theory and experiment [1]. So it is critical to measure the atom-atom interaction inside the system Hamiltonian to ensure the accuracy of the quantum simulation. However, the pure interaction Hamiltonian in Eq. (2) is already diagonal, thus does not directly satisfy the definition of parallel-learnable Hamiltonian discussed in Definition II.1. Therefore, instead of tackling the pure interaction Hamiltonian, we consider the full Hamiltonian, which is constructed by adding a single X_i term to the target Hamiltonian H.

Definition II.3. *The full Hamiltonian of the pure interaction Hamiltonian given in Eq. (2) is*

$$H = a_i X_i + \sum_{\substack{i,j \in \{1,\cdots,n\} \\ i \neq j}} c_{ij} Z_i Z_j, \tag{3}$$

where X_i denotes the application of X to the ith atom and a_i is the parameter of X_i term.

We note here that there are n different full Hamiltonians as there are n atoms inside the system. We will utilize $(n-1)$ of them for learning the full parameter vectors later. For each full Hamiltonian corresponding to a different X_i term, it has a clear invariant subspace structure. As a result, each of these Hamiltonians is parallel-learnable and can be transformed into a direct sum of 2×2 non-diagonal matrices through simple permutations. Moreover, for this particular class of Hamiltonians, the invariant subspace structure naturally emerges in the computational basis. Consequently, within the learning algorithm, we can implement a virtual permutation by preparing the appropriate encoded initial states corresponding to the defined invariant subspaces, which is effectively equivalent to permuting the Hamiltonian itself.

Lastly, the invariant subspace structure of a parallel-learnable Hamiltonian is preserved under time evolution. This leads to the following lemma concerning the form of the propagator associated with such a Hamiltonian.

Lemma II.4. *The time-evolution operator (or propagator) of the parallel-learnable Hamiltonian H is also block-diagonal with 2×2 blocks.*

Therefore, the propagator of the parallel-learnable Hamiltonian also has 2^{n-1} different invariant subspaces.

This property enables the simultaneous estimation of multiple parameters by leveraging time-evolution dynamics within different invariant subspaces in parallel. We present specific designs utilizing the propagator for parallel learning the estimators $\hat{\mathbf{c}}$ in Problem II.2 in Sec. II C. But the techniques developed in this work can also be readily applied to other parallel-learnable Hamiltonians.

B. Main Results

In this section, we present our main results, starting with the discussion on the quadratic speedup achieved by our learning algorithm for parallel-learnable Hamiltonians.

Theorem II.5. *Considering the parallel-learnable full Hamiltonian* $H_i = a_i X_i + \sum_{\substack{i,j\in\{1,\cdots,n\}\\ i\neq j}} c_{ij} Z_i Z_j$, *there exists an algorithm which involves* $O(n)$ *number of independent experiments to learn all* $O(n^2)$ *interaction parameters in Eq.*(2).

The key behind this quadratic speedup of this learning algorithm lies in the ability to exploit multiple parallel invariant subspaces of the full Hamiltonian simultaneously. In our algorithm, each independent experiment evolves a specific full Hamiltonian $H_i, \forall i \in \{1, \cdots, n-1\}$. Each single experiment round allows us to extract information about $O(n)$ parameters, thus to learn $O(n^2)$ parameters, only $O(n)$ independent experiments are required. However, parallelization involves more invariant subspaces, adding sampling overheads for each experimental round. Nevertheless, our algorithm remains favorable, as the total evolution time, accounting for both the number of rounds and the sampling per round, still achieves a $1/2$ reduction in the exponent of n compared to [18]. Further details on the total evolution time are provided in Supplementary Material Sec. VII.

For each independent experiment with Hamiltonian H_i, we utilize the parallelization of the QSPE method [17] within each invariant subspace to learn all parameters simultaneously. The QSPE method [31] was originally developed for two-qubit gate calibration, leveraging quantum signal transformation to separate parameters in the Fourier domain. It achieves the Heisenberg limit and the variance even scales as $1/d^4$ in the pre-asymptotic regime, while also demonstrating robustness to realistic experimental errors. Consequently, the estimators used in our learning algorithm retain the advantageous characteristics of the QSPE estimator.

Theorem II.6 (Optimal precision performance)**.** *The estimator* $\hat{c}_{ij}, \forall i \neq j$ *has the variance* $Var(\hat{c}_{ij}) = O(\frac{1}{Nd^4})$, *where* N *is the number of shots and* d *is the number of repetition cycles. Further, the computed variances saturate the Cramér–Rao lower bound.*

Beyond this, the estimators are inherently robust to realistic errors. The estimators decouple the parameters in all invariant subspaces for all independent experiments, resulting in the time-dependent coherent errors on one of them will not affect the estimation accuracy of the other. Moreover, the estimators are designed from Fourier coefficients, enabling simple rescaling procedures in the algorithm to effectively mitigate the depolarization and coherent state preparation errors. As a result, our learning algorithm exhibits full robustness against SPAM errors, decoherence, and time-dependent coherent noise.

The resource efficiency and robustness of our learning algorithm make it well-suited for practical deployment on current quantum hardware. In particular, our method can be applied to infer interatomic distance errors in Rydberg atom arrays, which is an important source of systematic deviation between theoretical predictions and experimental results, as highlighted in prior studies [1, 32–34]. While these deviations have been effectively modeled using distance error hypotheses, a first-principles analysis has remained elusive. This is largely due to the fact that such distance errors are difficult to directly capture in experiments, constrained by limited coherence times and the presence of realistic noise. Our algorithm addresses this challenge by requiring only short-time evolutions to amplify the parameters of interest, enabled by estimators that achieve optimal precision scaling and saturate the Cramér–Rao bound. This feature makes it compatible with the coherence times achievable on existing NISQ-era Rydberg platforms. Furthermore, the algorithm's robustness ensures reliable performance in the presence of realistic system noise. Consequently, our method enables efficient and robust estimation of multiple interatomic distances, which are essential for constructing accurate atom-distance error models and explaining the discrepancies observed in previous Rydberg-based experiments [1].

C. Methods

In this section, we provide a detailed discussion of the learning algorithm for solving Problem II.2. The techniques here are also directly applicable to any other parallel-learnable Hamiltonian. We focus on the full parallel-learnable Hamiltonian Eq. (3) with both interaction terms and the local fields to construct our learning algorithm.

1. *Two-atom Case*

First, we consider the simplest scenario for the Hamiltonian given in Eq. (3) with two atoms. The corresponding Hamiltonian is

$$H(a, c_{12}) = aX_1 + c_{12} Z_1 Z_2. \tag{4}$$

We aim to estimate $\hat{c}_{12}$ for the interaction term, while $\hat{a}$ can also be inferred via our learning algorithm as a byproduct. This Hamiltonian is native to the Rydberg atom system, where the local field is applied via Rabi frequency Ω and c_{12} is the atom-atom interaction. Therefore, the learning algorithm can be directly applied to learn the atom-atom interaction and atom distance on the Rydberg device. Details on this implementation will be discussed in Sec. III 2.

Consider the Hamiltonian evolution for time T, then the corresponding time evolution operator can be written as

$$\mathcal{U} = \exp(-i\int_0^T Hdt) = \exp[-i(a^T X_1 + c_{12}^T Z_1 Z_2)], \tag{5}$$

where $a^T := \int_0^T a dt = aT$ and $c_{12}^T := \int_0^T c_{12} dt = c_{12}T$ [35]. By simply dividing out the evolution time T, we can easily recover the original parameters $(\hat{a}, \hat{c}_{12})$. Taylor expanding the above Eq. (5), the corresponding unitary is $\mathcal{U} = \cos\omega I - i\sin\omega(\frac{a^T}{\omega}X_1 + \frac{c_{12}^T}{\omega}Z_1Z_2)$ where $\omega = \sqrt{(a^T)^2 + (c_{12}^T)^2}$. Defining the logical subspace as $|0\rangle_l = |00\rangle$ and $|1\rangle_l = |10\rangle$, which spans the invariant subspace $\mathcal{B}$. The unitary restricted to the subspace $\mathcal{B}$ is

$$\mathcal{U}_{\mathcal{B}} = \begin{bmatrix} \cos\omega - i\frac{c_{12}^T}{\omega}\sin\omega & -i\frac{a^T}{\omega}\sin\omega \\ -i\frac{a^T}{\omega}\sin\omega & \cos\omega + i\frac{c_{12}^T}{\omega}\sin\omega \end{bmatrix}. \tag{6}$$

The standard unitary calibrated using QSPE [17] is in the form of

$$\mathcal{U} = \begin{bmatrix} \cos(\theta)e^{-i\zeta} & -i\sin(\theta)e^{i\chi} \\ -i\sin(\theta)e^{-i\chi} & \cos(\theta)e^{i\zeta} \end{bmatrix}. \tag{7}$$

The following mapping can convert one to the other,

$$\begin{cases} \tan(\zeta) = \frac{c_{12}^T}{\sqrt{(a^T)^2+(c_{12}^T)^2}}\tan(\sqrt{(a^T)^2+(c_{12}^T)^2}) \\ \sin^2(\theta) = (\frac{a^T}{\sqrt{(a^T)^2+(c_{12}^T)^2}})^2\sin^2(\sqrt{(a^T)^2+(c_{12}^T)^2}) \end{cases}, \tag{8}$$

and $(a, c_{12}) = (a^T/T, c_{12}^T/T)$. Therefore, we can utilize the QSPE method for the 2×2 invariant subspace defined by $(|00\rangle, |10\rangle)$ for learning both parameters. The details regarding implementation are discussed in the Supplementary Material Sec. III. The step-by-step algorithm is shown as follows, see Algorithm 1.

Algorithm 1 Two-atom QSPE-Learning algorithm

Require: Choose the hyper-parameters for the number of cycles, and sample size (d, N).

1: Compute the control parameter $\omega_j = \frac{j}{2d-1}\pi$ (for $j = 0, 1, \cdots, 2d-2$).

2: Initiate a complex-valued data vector $\vec{h} \in \mathbb{C}^{2d-1}$.

Main steps:

For $j \in \{0, 1, \cdots, 2d-2\}$,

► Quantum Part:

- Prepare the logical initial state $|+_l\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |10\rangle)$ and $|i_l\rangle = \frac{1}{\sqrt{2}}(|00\rangle + i|10\rangle)$.
- Define one cycle of evolution: apply the full Hamiltonian evolution $\mathcal{U}$ and subsequently apply the logical Z rotation $Z_l(\omega_i)$ for the specific angle ω_i.
- Apply d cycles of evolution to both initial states.
- Measure in computational basis and get $x_j^{(+)}$ and $x_j^{(i)}$ 00's out of N samples for both initial states.

► Classical Part:

- Estimate the transition probability: $p_X = \frac{x_j^{(+)}}{N}$ and $p_Y = \frac{x_j^{(i)}}{N}$.
- Update $\vec{h}_j \leftarrow p_X(\omega_j) - \frac{1}{2} + i(p_Y(\omega_j) - \frac{1}{2})$.

end for

Post-processing:

3: Compute the Fourier coefficients $\vec{c} = FFT(\vec{h})$.

4: Compute the phase difference vector $\vec{\Delta}$, where $\vec{\Delta} = (\Delta_0, \cdots, \Delta_{d-2})^T$ and $\Delta_k = \text{phase}(c_k\bar{c}_{k+1})$.

5: Compute estimates $(\hat{\theta}, \hat{\zeta})$ with

$$\hat{\theta} = \frac{1}{d}\sum_{k=0}^{d-1}|c_k| \quad , \hat{\zeta} = \frac{1}{2}\frac{\vec{\mathbb{I}}^T\mathcal{D}^{-1}\vec{\Delta}}{\vec{\mathbb{I}}^T\mathcal{D}^{-1}\vec{\mathbb{I}}}, \tag{9}$$

where $\mathbb{I}$ is an all-one vector and $\mathcal{D}$ is the $(d-1)\times(d-1)$ discrete Laplacian matrix.

6: Covert the estimated parameters $(\hat{\theta}, \hat{\zeta})$ to $(\hat{a}, \hat{c}_{12})$ by the mapping given in Eq. (8).

2. *Multi-atom Case*

In this section, we extend the two-atom learning algorithm introduced earlier to construct a generic learning algorithm for the n-qubit Hamiltonian, as defined in Eq. (3). The idea is to encode the full Hamiltonian into multiple 2×2 subspaces, each solvable with the QSPE as suggested in the previous section. The learning process can be parallelized, which allows us to estimate all coefficients induced by the same qubit in parallel within a single independent experiment round. To be more specific, consider the full Hamiltonian H_i, we can define its 2×2 invariant subspaces as follows. Denote the computational bases of n-qubit Hamiltonian as $\{|v_1\rangle, |v_2\rangle, \cdots, |v_{2^n}\rangle\}$, where $|v_1\rangle = |00\cdots0\rangle$ with $|v_1| = 0$ and $|v_{2^n}\rangle = |11\cdots1\rangle$ with $|v_{2^n}| = n$. For each H_i, we define the encoding subspaces.

Definition II.7. *For an n-qubit system with the full Hamiltonian including the X_i term as given in Eq. (3), the Hilbert space decomposes into 2^{n-1} invariant subspaces. We denote these subspaces by $\{\mathcal{B}_1, \cdots, \mathcal{B}_{2^{n-1}}\}_i$. Each invariant subspace $\mathcal{B}_k$ (for $k \in \{1, \cdots, 2^{n-1}\}$) is two-dimensional and corresponds to a logical subspace. The encoding for each $\mathcal{B}_k$ is defined as:*

$$|v_m\rangle := |\bar{0}\rangle\,, \quad |v_n\rangle := |\bar{1}\rangle\,, \text{ with } v_m \oplus v_n = e_i, \quad (10)$$

where $m, n \in \{1, \cdots, 2^n\}$ with $m \neq n$, and e_i is the bitstring with 1 at the i-th position and 0 elsewhere. Each pair $(|v_m\rangle_k, |v_n\rangle_k)$ defines one of the 2^{n-1} invariant subspaces $\mathcal{B}_k$.

Under the above encoding, the full Hamiltonian in Eq.(3) can be represented as a block diagonal Hermitian matrix with block size 2. Moreover, the entries of each 2×2 matrix H_i^k restricted to the kth invariant subspace can be computed by considering the projection of the H_i onto each $\mathcal{B}_k$. The off-diagonal entries are introduced by the local bit-flip term X_i, so they are identically equal to a_i for all invariant subspaces. The diagonal entries are linear combinations of all $\frac{n(n-1)}{2}$ interaction term parameters. Specifically, we can write down the analytical form of the H_i^k as follows.

Lemma II.8. *Consider H_i^k, which is the projection of H_i onto invariant subspace $\mathcal{B}_k$ spanned by $\{(|v_m\rangle_k, |v_n\rangle_k)\}$. Let E_i denote the set of edges associated with vertex i. Specifically, the edge set is $E_i = \{(i,1),(i,2),\cdots,(i,i-1),(i,i+1),\cdots,(i,n)\}$ in the all-to-all connected system. We classify the interaction terms into two classes:*
1. The different index set, denoted by E_D, includes all operators that act on edges connected to vertex i, which result in differing signs in the diagonal elements of H_i^k. It is defined as

$$E_D = \{(p,q) \mid (p,q) \in E_i\} = \{(p,q) \mid p = i\}. \quad (11)$$

2. The same index set, denoted by E_S, includes all remaining operators that do not act on the edges connected to vertex i, and thus yield the same sign in the diagonal elements. It is defined as

$$E_S = \{(p,q) \mid (p,q) \notin E_i\} = \{(p,q) \mid p \neq i\}. \quad (12)$$

Note that the second expression in both definitions of E_D and E_S holds specifically for the case of an all-to-all connected system.

Let λ_{pq}^m denote the eigenvalue of the operator Z_pZ_q with respect to the logical zero state $|v_m\rangle$ and set it equal to λ_{pq}, then the corresponding eigenvalue of the same operator to the logical one state $|v_n\rangle$ is given by

$$\lambda_{pq}^n = \begin{cases} \lambda_{pq} & when \quad (p,q) \in E_S; \\ -\lambda_{pq} & when \quad (p,q) \in E_D. \end{cases} \quad (13)$$

The projected Hamiltonian H_i^k has the following analytical matrix form:

$$H_i^k = \begin{bmatrix} \Lambda_+ & a_i \\ a_i & \Lambda_- \end{bmatrix}. \quad (14)$$

where $\Lambda_+ = \sum_{(p,q)\in E_S} \lambda_{pq}c_{pq} + \sum_{(p,q)\in E_D} \lambda_{pq}c_{pq}$ and $\Lambda_- = \sum_{(p,q)\in E_S} \lambda_{pq}c_{pq} - \sum_{(p,q)\in E_D} \lambda_{pq}c_{pq}$.

The information of an analytical linear combination of our desired estimators is encoded at each invariant subspace of the Hamiltonian, and then we proceed to use the QSPE method to infer them with respect to each block. More specifically, let us denote the time-integrals of distinct entries in H_i^k as follows,

$$(A_i, B_i, C_i) := \int_0^T (a_i, \sum_{(p,q)\in E_D} \lambda_{pq}c_{pq}, \sum_{(p,q)\in E_S} \lambda_{pq}c_{pq})dt; \quad (15)$$

$$= (a_iT, \sum_{(p,q)\in E_D} \lambda_{pq}c_{pq}T, \sum_{(p,q)\in E_S} \lambda_{pq}c_{pq}T), \quad (16)$$

where T is the evolution time [35]. The time-evolution operator of the kth block is

$$\mathcal{U}^k = \begin{bmatrix} \cos\omega - i\sin\omega\frac{B_i}{\omega} & -i\sin\omega\frac{A_i}{\omega} \\ -i\sin\omega\frac{A_i}{\omega} & \cos\omega + i\sin\omega\frac{B_i}{\omega} \end{bmatrix}, \quad (17)$$

where $\omega = \sqrt{A_i^2 + B_i^2}$. Correspondingly, the following mapping can lead to the conversion from the unitary to the one used in the QSPE method,

$$\begin{cases} \tan(\zeta) = \frac{B_i}{\sqrt{A_i^2+B_i^2}} \tan(\sqrt{A_i^2 + B_i^2}) \\ \sin^2(\theta) = (\frac{A_i}{\sqrt{A_i^2+B_i^2}})^2 \sin^2(\sqrt{A_i^2 + B_i^2}) \end{cases}. \quad (18)$$

For each specific block, both A_i and B_i can be learned. Consequently, the linear combination of these parameters is effectively learned through Eq. (15). Since there are in total $|E_D|$ parameters connecting to the ith qubit, we require $|E_D|$ independent linear combinations of the parameters c_{pq} for $(p,q) \in E_D$, obtained from $|E_D|$ distinct invariant subspaces. Specifically, we need to identify $(n-1)$ invariant subspaces whose diagonal entries differ in a linearly independent manner for the all-to-all connected Hamiltonian. We justify the existence of such subspaces and provide a method for selecting these $(n-1)$ linearly independent subspaces in the following theorem.

Theorem II.9. *For the n-qubit full Hamiltonian H_i defined in Eq. (3), there exist 2^{n-1} distinct nonzero invariant subspaces, each associated with an encoding basis of the form $(|v_m\rangle_s, |v_n\rangle_s)$, where $s = 1, \ldots, 2^{n-1}$ and $v_m \oplus v_n = e_i$. We use one of the logical zero state $|v_m\rangle$ or logical one state $|v_n\rangle$ from each pair to label the corresponding invariant subspace. Define the set $\mathcal{S}_0 = \{|v_m\rangle_1, \ldots, |v_m\rangle_{2^{n-2}} \mid b_ib_{i+1} = 00\}$, which contains those basis states with 0 in both the i-th and $(i+1)$-th positions. The complementary set, denoted by $\mathcal{S}_1 = \{|v_n\rangle_{2^{n-2}+1}, \ldots, |v_n\rangle_{2^{n-1}} \mid b_ib_{i+1} = 11\}$, includes the remaining basis states where both the i-th and $(i+1)$-th bits are 1. Then we get the following properties:*

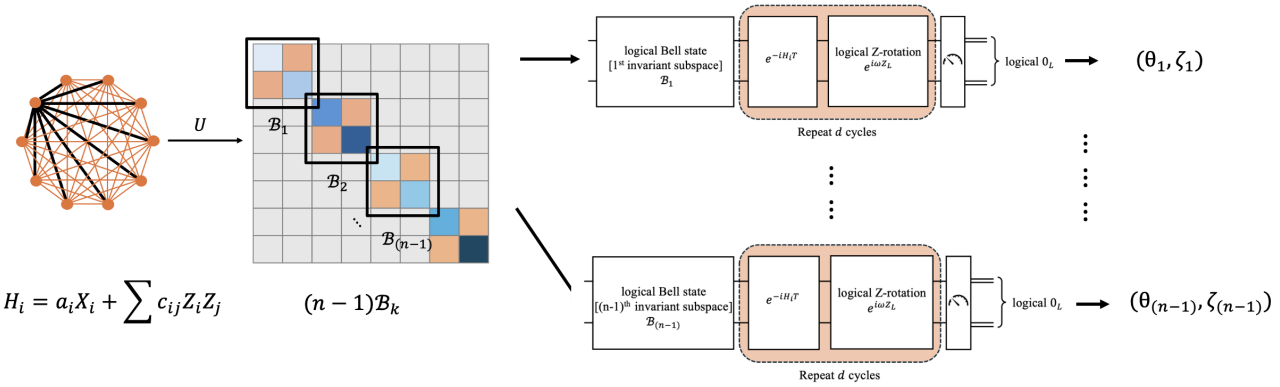


FIG. 1: General framework for multi-qubit parallel learning. The ith full Hamiltonian H_i is used to learn the $(n-1)$ parameters associated with the ith qubit. First, H_i is mapped to a block-diagonal form with block size 2, of which $(n-1)$ blocks are employed for estimation. Each block corresponds to an invariant subspace $\mathcal{B}_k$ with Definition II.7. On each block for invariant subspace $\mathcal{B}_k$, QSPE is applied to estimate a parameter pair (θ_k, ζ_k) encoding the Hamiltonian information via Eq. (18). The QSPEs across all invariant subspaces can be executed in parallel by evolving under the multi-qubit Hamiltonian, followed by the logical $Z_L(\omega)$ for phase accumulation. When projected onto each invariant subspace, the dynamics reduce to QSPE on a 2×2 matrix.

1. $|\mathcal{S}_0| + |\mathcal{S}_1| = 2^{n-1}$ *representing all non-zero invariant subspaces.*

2. *Define the coefficient vector* $w_k := [\lambda_{i1}, \cdots, \lambda_{in}]^T$ *representing the coefficient of the linear combination for the different sign term* $\sum_{(p,q)\in E_D} \lambda_{pq} c_{pq}$ *for basis state* $|v_k\rangle$. *The space spanned by the coefficient vectors associated with* $\mathcal{S}_0$ *is the same as* $\mathcal{S}_1$.

3. *Define the coefficient matrix* $W := [w_1, \cdots, w_{2^{n-2}}]$ *by putting all coefficient vector for all basis in* $\mathcal{S}_0$. *It is a full rank matrix with size* $(n-1) \times 2^{n-2}$.

4. *By selecting* $(n-1)$ *linearly independent columns indexed by a subset* $\mathcal{C} \subset \{1, \ldots, 2^{n-2}\}$, *from the coefficient matrix, one can solve for the* $(n-1)$ *parameters. The corresponding invariant subspaces chosen are those spanned by the* $(n-1)$ *distinct basis vectors* $|v_m\rangle_s$ *with* $s \in \mathcal{C}$.

Proof. The first property is straightforward. Since there are only four possible combinations of the bit pair $b_i b_{i+1}$, namely $\{00, 01, 10, 11\}$, and each encoding pair for the invariant subspace satisfies $v_m \oplus v_n = e_i$, the pairing structure effectively partitions the full basis of the Hilbert space into four equal subsets. As a result, we have $|\mathcal{S}_0| = |\mathcal{S}_1| = 2^n/4 = 2^{n-2}$.

The second property follows from the observation that

$$Z_i Z_{i+1} |s\rangle = Z_i Z_{i+1} |t\rangle ,$$

for all $|s\rangle \in \mathcal{S}_0$ and $|t\rangle \in \mathcal{S}_1$. Moreover, since the bit values at all other positions are identical between elements of $\mathcal{S}_0$ and $\mathcal{S}_1$, applying $Z_i Z_j$ with $j \in \{1, \ldots, n\}$, $j \neq i, i+1$ yields the same set of eigenvalues λ_{ij} across both sets. Therefore, the spaces spanned by coefficient vectors from $\mathcal{S}_0$ and $\mathcal{S}_1$ are the same. Therefore, it is sufficient to focus on one of the basis sets to extract all the necessary information to solve for the parameters. In the discussion of the third and fourth properties, we restrict our attention to $\mathcal{S}_0$.

Consider the $Z_i Z_{i+1}$ operator, for any basis in $\mathcal{S}_0$,

$$Z_i Z_{i+1} |b_1 \cdots 00 \cdots b_n\rangle = |b_1 \cdots 00 \cdots b_n\rangle . \tag{19}$$

So $\lambda_{i,i+1} = 1$ for all basis in $\mathcal{S}_0$, i.e., the ith term is 1 for all w_k. Each w_k contains the coefficients of the different sign terms, and thus has dimension $(n-1)$ with ith entry equal to 1. Besides, there are 2^{n-2} elements in $\mathcal{S}_0$, resulting the coefficient matrix W has the size $(n-1) \times 2^{n-2}$. To prove the matrix W is of full rank, we show there exists a submatrix $W_s \in \mathbb{R}^{(n-1)\times(n-1)}$ that is of full rank. We choose the submatrix W_s as follows.

Definition II.10 (Choice of invariant subspace and its coefficient matrix)**.** *The logical zero basis states for* $(n-1)$ *invariant subspaces are*

$$\begin{cases} |v_m\rangle_i = |00 \cdots 0 \cdots 00\rangle , & \textit{all-zero state} \\ |v_m\rangle_k = |01 \cdots 0 \cdots 00\rangle , & \textit{all entries are } 0 \textit{ except the } k\textit{th} \end{cases} \tag{20}$$

for $k \in \{1, \cdots, n\}$, $k \neq i, i+1$.

Applying $Z_i Z_j$ where $j \in \{1, \cdots, n\}$ and $j \neq i$ to the bases $\{|v_m\rangle_1, \cdots, |v_m\rangle_{n-1}\}$, the corresponding co-

efficient matrix is

$$W_s = \begin{bmatrix} -1 & 1 & 1 & \cdots & 1 \\ 1 & -1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & -1 \end{bmatrix}, \tag{21}$$

where for the kth column ($k = 1, \cdots, n$ and $k \neq i$, i+1), the vector w_k has an entry of -1 at the kth row and 1 in all other entries, while the ith column consists entirely of 1's. To prove this submatrix W_s is of full rank, we compute this determinant. Subtracting column i from all other columns, the determinant of this transformed matrix is $1 \times (-2)^{n-2} \neq 0$, thus the W_s matrix is full-rank.

Finally, solving for $(n-1)$ different coefficients requires forming $(n-1)$ linearly independent equations. Since there are in total $(n-1)$ linearly independent columns from a total 2^{n-1} column matrix given property three, we can pick up any set of linearly independent ones, for example, as defined in Definition II.10. The chosen invariant subspaces $\{\mathcal{B}_1, \cdots, \mathcal{B}_{(n-1)}\}$ are then spanned by encoding pairs $\{(|v_m\rangle_1, |v_n\rangle_1), \cdots (|v_m\rangle_{n-1}, |v_n\rangle_{n-1})\}$. □

The above discussion confirms that all parameters associated with the same qubit (WOLG, saying atom i) can be recovered in one experiment round with the full Hamiltonian H_i. Therefore, the interaction parameters of all-to-all connection Hamiltonian given in Eq. (2) can be recovered with $O(n)$ independent rounds of experiments. As a result, our learning algorithm can achieve a quadratic speedup in terms of the number of experiment rounds in the whole process, as stated in the following Theorem II.5.

For each Hamiltonian term H_i, one can exploit $(n-1)$ invariant subspaces to extract the corresponding $(n-1)$ parameters associated with the ith qubit. However, when designing the learning algorithm, we can adopt an iterative approach to maximize the information gained in each round. As the iteration over i progresses, the number of unknown parameters decreases to $(n-i)$ for each H_i. Consequently, it suffices to employ only $(n-i)$ invariant subspaces at each step by preparing initial states restricted to these subspaces and evolving the system dynamics accordingly. In this case, the coefficient matrix is updated to W_s^i, an $(n-i) \times (n-1)$ matrix consisting of $(n-i)$ linearly independent rows of W_s, corresponding precisely to the selected $(n-i)$ invariant subspaces.

The learning algorithm within each invariant subspace follows the same procedure as described in the two-atom case in Algorithm 1, while extending to multiple parallel invariant subspaces. We introduce modifications to both the state preparation procedure and the implementation of the logical operation $Z_L(\omega)$, as detailed below. Furthermore, in accordance with the capabilities of the quantum hardware, we develop two distinct algorithmic variants: one designed for analog–digital hybrid architectures and another for fully analog systems. In addition, the algorithm can also be executed on a fully digital quantum device by replacing the Hamiltonian evolution step in the hybrid mode with quantum gates for learning the parameters of the quantum gates.

a. Analog-digital-hybrid learning algorithm In the analog–digital hybrid mode, the quantum dynamics is assumed to consist not solely of continuous-time Hamiltonian evolution but rather a combination of both continuous evolution and digitized gate operations. This flexibility enables the simultaneous implementation of the logical operator $Z_L(\omega)$ across $(n-i)$ invariant subspaces.

Definition II.11 (State preparation in hybrid setting)**.** *Consider the invariant subspaces spanned by*

$$\left\{(|v_m\rangle_1, |v_n\rangle_1), \cdots, (|v_m\rangle_{n-i}, |v_n\rangle_{n-i})\right\},$$

then the logical states $|+_L\rangle$ and $|i_L\rangle$ are defined as

$$\begin{cases} |+_L\rangle = \dfrac{1}{\sqrt{2(n-i)}} \displaystyle\sum_{j=1}^{n-i} (|v_m\rangle_j + |v_n\rangle_j) \\ |i_L\rangle = \dfrac{1}{\sqrt{2(n-i)}} \displaystyle\sum_{j=1}^{n-i} (|v_m\rangle_j + i\,|v_n\rangle_j) \end{cases} \tag{22}$$

The initial state corresponds to the superposition of the Bell states on $(n-i)$ invariant subspaces, where $i = 1, \cdots (n-1)$.

Definition II.12 (Logical Z_L in hybrid setting)**.** *The logical $Z_L(\omega)$ is implemented by*

$$Z_L(\omega) = \exp(-i\omega Z_i), \tag{23}$$

where $Z_i = I \otimes Z \otimes \cdots \otimes I$ acting Z on ith qubit, which performed as $Z(\omega)$ at each invariant subspaces of the full Hamiltonian.

In this setting, the logical operator $Z_L(\omega)$ is realized as a digital gate, during which the analog Hamiltonian evolution is suspended. Based on these modifications, we present the Analog-digital-hybrid parallel learning algorithm 2. for Problem II.2.

Algorithm 2 Analog-digital-hybrid parallel learning algorithm

for $i \leftarrow 1$ **to** $n-1$ **do**

1: Prepare $H_i = a_i X_i + \sum_{\substack{i,j\in\{1,\cdots,n\}\\ i\neq j}} c_{ij} Z_i Z_j$, the ith full Hamiltonian which is parallel-learnable.

2: Prepare $UH_iU^\dagger$. ▷ turning H into direct sum form by applying realizable transformation U, $U = I$ in Problem II.2

3: Choose $(n-i)$ invariant subspace of H_i. ▷ Infer $(n-i)$ parameters, $\vec{c}_i = \{c_{i(i+1)}, \cdots, c_{in}\}$.

for $j \leftarrow 0$ **to** $2d-2$ **do**

► Quantum Part:

- Prepare the initial state $|+_L\rangle$ and $|i_L\rangle$ according to Definition II.11.
- Define one cycle of evolution: apply the full Hamiltonian evolution for time T and subsequently apply the logical Z rotation gate $Z_L(\omega_j)$ according to Definition II.12.
- Apply d cycles of evolution on both initial states.
- Measure in computational basis and collect the outcome bitstrings. Count the bitstring number corresponds to the logical zero state $\{|v_m\rangle_k\}_{k=1,\cdots,n-i}$ for all k invariant subspaces and denote the number as $x^{(+)}_{\omega_j,k}$ and $x^{(i)}_{\omega_j,k}$ for both initial states.

► Classical Part:

for $k \leftarrow 1$ **to** $n-i$ **do**

- Estimate the transition probability: $\hat{p}_{X,k} = \frac{x^{(+)}_{\omega_j,k}}{N}$ and $\hat{p}_{Y,k} = \frac{x^{(i)}_{\omega_j,k}}{N}$.
- Update the reconstruction function with rescaled probability $\vec{h}^k_j \leftarrow [(n-i)\hat{p}_{X,k} - \frac{1}{2}] + i[(n-i)\hat{p}_{Y,k} - \frac{1}{2}]$.
- Use the same post-processing method in Algorithm 1 to get $\{\theta_k, \zeta_k\}$ for each subspace.
- Solve for B_k, the time integral of the linear combination of the elements in $\vec{c}_i$ via $\{\theta_k, \zeta_k\}$, according to Eq. (18).

end for

4: Solve for $(n-i)$ unknown parameters $\vec{c}_i$ with the linear equation

$$[B_1, \cdots, B_{n-i}]^T = W_s^i [\hat{c}_{i1}T, \cdots, \hat{c}_{in}T]^T, \tag{24}$$

where $[\hat{c}_{i1}, \cdots, \hat{c}_{i(i-1)}]$ is estimated in previous iterations.

end for

From the above parallel-learning algorithm, we notice that the total number of parameters can be learned through $(n-1)$ independent experiments is $(n-1) + (n-2) + \cdots + 1 = n(n-1)/2$. Thus, the algorithm learns the $O(n^2)$ coefficients in $O(n)$ rounds, leading to a quadratic speedup compared to [18]. We also emphasize that the $(n-1)$ full Hamiltonians used in the algorithm include all interactions, enabling in-situ learning of all interaction parameters simultaneously.

b. Fully analog learning algorithm The learning algorithm can also be implemented in a fully analog setting, that is, the quantum dynamics are generated by continuous-time Hamiltonian evolution without any discrete quantum gates. This setting is particularly relevant to the capabilities of Rydberg-atom quantum simulators, where the inter-atomic interactions are determined by the spatial separation between atoms and remain always-on unless the atoms are physically rearranged during the evolution. In this case, the challenge is to implement the logical $Z_L(\omega)$ rotation in the presence of the always-on ZZ interactions. Assuming limited controllability of the device, with only tunable single-qubit Z terms available, it is not possible to realize the logical operation Z_L simultaneously across all $(n-1)$ invariant subspaces. Instead, we implement the logical Z^k_L rotation targeting one specific invariant subspace $\mathcal{B}_k$ at a time.

Definition II.13 (Logical Z^k_L in analog setting)**.** *For the kth invariant subspace, the logical $Z_L(\omega)$ can be implemented with the Hamiltonian*

$$H_{Z^k_L} = b_k Z_k + \sum_{\substack{i,j\in\{1,\cdots,n\}\\ i\neq j}} c_{ij} Z_i Z_j,$$

where $b_k = \omega - \Lambda_+$. The logical operation is $Z^k_L(\omega) = \exp(-iH_{Z^k_L}t)$. In comparison with the full Hamiltonian given in Eq. (3), this Hamiltonian is realized by suspending the single-qubit X terms and activating the single-qubit Z terms, while retaining all two-local ZZ interactions.

The implementation of the logical operator Z^k_L is fully analog, thereby enabling the learning algorithms to be executed with minimal requirements on the quantum device. Moreover, since all Hamiltonian interactions remain always-on, this setting naturally supports in-situ learning of all Hamiltonian terms. However, this trade-off prevents the realization of the logical gate Z_L across all invariant subspaces simultaneously, resulting in an $O(n^2)$ overhead in the number of experiments. Fortunately, in this mode, the initial states of the algorithm can be prepared specifically for the kth invariant subspace, rather than employing a superposition of $O(n)$ Bell pairs, thereby reducing the sampling overhead. To be more concrete, the initial state for the kth invariant subspace is defined as

Definition II.14 (State preparation in analog setting)**.** *For the kth invariant subspace, which is spanned by the basis $\{|v_m\rangle_k, |v_n\rangle_k\}$, then the logical states $|+^k_L\rangle$ and $|i^k_L\rangle$ are defined as*

$$\begin{cases} |+^k_L\rangle = \dfrac{1}{\sqrt{2}}(|v_m\rangle_k + |v_n\rangle_k) \\ |i^k_L\rangle = \dfrac{1}{\sqrt{2}}(|v_m\rangle_k + i\,|v_n\rangle_k) \end{cases} \tag{25}$$

Compared to Eq.(22), the normalization factor is $(n-i)$ times larger, which will reduce the estimation

variance by a factor of n. As a result, the total evolution time required by the fully analog learning algorithm is not expected to be significantly longer than that of the analog–digital hybrid mode. More details on the total evolution time scaling can be found in Supplementary Material Sec. VII. To conclude, we present the fully analog in-situ learning algorithm as follows in Algorithm 3.

Algorithm 3 Fully analog in-situ learning algorithm

for $i \leftarrow 1$ **to** $n-1$ **do**

1: Prepare $H_i = a_i X_i + \sum_{i,j\in\{1,\cdots,n\}, i\neq j} c_{ij} Z_i Z_j$, the ith full Hamiltonian which is parallel-learnable.

2: Prepare $UH_iU^\dagger$. ▷ turning H into direct sum form by applying realizable transformation U, $U = I$ in Problem II.2

3: Choose $(n-i)$ invariant subspace of H_i. ▷ Infer $(n-i)$ parameters, $\vec{c}_i = \{c_{i(i+1)}, \cdots, c_{in}\}$

for $k \leftarrow 1$ **to** $n-i$ **do**

for $j \leftarrow 0$ **to** $2d-2$ **do**

► Quantum Part:

– Prepare the initial state $|+_L\rangle$ and $|i_L\rangle$ according to Definition II.14.

– Define one cycle of evolution: apply the full Hamiltonian evolution for time T and subsequently apply the logical Z rotation gate $Z_L(\omega_j)$ according to Definition II.13.

– Apply d cycles of evolution on both initial states.

– Measure in computational basis and collect the outcome bitstrings. Count the bitstring number corresponds to the logical zero state $\{|v_m\rangle_k\}$ and denote the number as $x^{(+)}_{\omega_j}$ and $x^{(i)}_{\omega_j}$ for both initial states.

► Classical Part:

– Estimate the transition probability: $\hat{p}_{X,k} = \frac{x^{(+)}_{\omega_j,k}}{N}$ and $\hat{p}_{Y,k} = \frac{x^{(i)}_{\omega_j,k}}{N}$.

– Update the reconstruction function with rescaled probability $\vec{h}^k_j \leftarrow [\hat{p}_{X,k} - \frac{1}{2}] + i[\hat{p}_{Y,k} - \frac{1}{2}]$.

– Use the same post-processing method in Algorithm 1 to get $\{\theta_k, \zeta_k\}$.

– Solve for B_k, the time-integral of the linear combination of the elements in $\vec{c}_i$ via $\{\theta_k, \zeta_k\}$, according to Eq. (18).

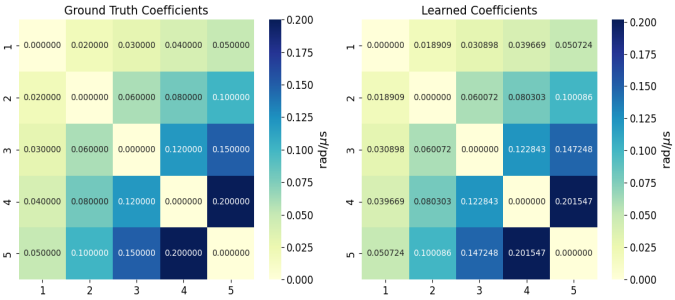
end for

end for

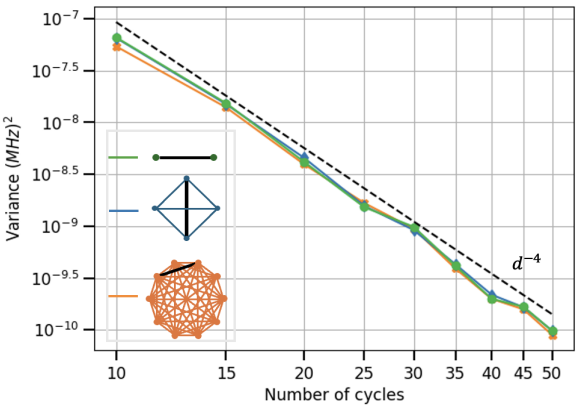
4: Solve for $(n-i)$ unknown parameters $\vec{c}_i$ with the linear equation (24), where $[\hat{c}_{i1}, \cdots, \hat{c}_{i(i-1)}]$ is estimated in previous iterations.

end for

To validate the correctness of our learning algorithm in the multi-atom setting, we simulate its performance on all-to-all connected Hamiltonians of varying sizes. In Figure. 2a, we learn the all-to-all connected Hamiltonian for 5 qubit with Algorithm 2. With $d = 10$ cycle and $N_{\text{shot}} = 10^4$ shots, the learning algorithm successfully infers all the Hamiltonian parameters. While in Figure. 2b, we further demonstrate the algorithm's stable performance across different system sizes. The black bold link in Figure. 2b indicates which interaction parameter is presented, but the performance for all of the parameters is identical. Since only one of the interactions is of interest here, we utilize Algorithm 3 while choosing two specific invariant subspaces to infer that particular parameter. With $N_{\text{shot}} = 10^5$ shots and $N_{\text{bootstrap}} = 10^3$ bootstraps, which is used to estimate the variance of the estimator, we notice the variance of the estimators scales as $1/d^4$ as suggested in Theorem II.6.



(a) Application of the learning algorithm for all-to-all connected 5 qubit system. The (i,j)th entry in the figure is the coefficient c_{ij} for the term Z_iZ_j. The simulation utilizes $d = 10$ cycles and the number of samples $N_{shot} = 10^4$.



(b) Scalability test on the learning algorithm. The simulation is done for 2-qubit (green), 4-qubit(blue) and the 10-qubit system (orange) with the number of samples $N_{\text{shot}} = 10^5$ and the number of bootstraps $N_{\text{bootstrap}} = 10^3$.

FIG. 2: Performance of the algorithm for multi-atom case.

D. Optimality discussion

In this section, we establish that the estimators in our learning algorithm achieve optimal precision under clas-

sical post-processing, with the resulting variance saturating the Cramér–Rao bound.

For both learning algorithms, the Hamiltonian parameters $\mathbf{c}$ are learned via estimators $\hat{\mathbf{c}}$ through $(n-1)$ different full Hamiltonians, each specifying by the single X_i term. Both algorithms share the same inference procedure for learning $\{\theta_k, \zeta_k\}$ on each kth invariant subspace and recovering the Hamiltonian parameters through an identical set of linear equations. The key distinction is that, in the analog–digital hybrid mode, all invariant subspaces evolve simultaneously using a superposition of the corresponding logical Bell pairs, whereas in the fully analog mode, only a single invariant subspace evolves at a time. Consequently, the sole difference in the inference analysis is that the transition probability for Algorithm 2 on each invariant subspace is reduced by a factor of $O(1/n)$ relative to Algorithm 3. Under the same number of measurement shots, this implies that the variances of the estimates $(\hat{\theta}_k, \hat{\zeta}_k)$ obtained from Algorithm 2 are larger by a factor of $O(n)$ compared to those from Algorithm 3. Since for each invariant subspace, we utilize the QSPE method [17] for learning $(\hat{\theta}_k, \hat{\zeta}_k)$, the same variance computation applies to Algorithm 3 for each 2×2 invariant subspace. Consider the accumulated angle θ_k for each subspace, which satisfies $d\theta_k \ll 1$. Here, d denotes the number of repetition cycles, and θ_k is the angle resolved by QSPE in the kth invariant subspace, computable via Eq. (15) and Eq. (18). Since θ_k depends on the time integral of the Hamiltonian coefficients, we can always select an appropriate evolution time T to ensure this condition is met. With sample size N, the variances computed for the estimated pair $(\hat{\theta}_k, \hat{\zeta}_k)$ in the kth invariant subspace are

$$\mathrm{Var}(\hat{\theta}_k) \approx \frac{1}{8Nd^2}\ , \mathrm{Var}(\hat{\zeta}_k) \approx \frac{3}{8Nd^4(\theta_k)^2}. \tag{26}$$

For Algorithm 2, we recompute the variance of $(\tilde{\theta}_k, \tilde{\zeta}_k)$ in the kth invariant subspace in Supplementary Material Sec.V, which are

$$\mathrm{Var}(\tilde{\theta}_k) \approx \frac{n}{4Nd^2}\ , \mathrm{Var}(\tilde{\zeta}_k) \approx \frac{3n}{4Nd^4(\theta_k)^2}. \tag{27}$$

Given the above calculation, we proceed to compute the variance for our desired estimators $\hat{\mathbf{c}}$ for both algorithms. Without loss of generality, we first focus on the full Hamiltonian $H_1 = a_1 X_1 + \sum c_{ij} Z_i Z_j$, the first learned Hamiltonian in the whole learning procedure. Let us denote the $(n-1)$ parameters associated with the first atom as $\mathbf{c_1} = (c_{12}, c_{13}, \cdots, c_{1n})$. We utilize $(n-1)$ invariant subspaces to solve for the estimator $\hat{\mathbf{c}}_\mathbf{1} = (\hat{c}_{12}, \hat{c}_{13}, \cdots, \hat{c}_{1n})$. While for each kth subspace, we utilize the QSPE method to solve for (θ_1^k, ζ_1^k). Here we use the superscript k to represent the kth invariant subspace and the subscript 1 to represent the H_1. After which, we follow the mapping

$$\begin{cases} \tan(\zeta_1^k) = \frac{m_1^k}{\sqrt{(a_1^T)^2+m_1^{k\,2}}} \tan(\sqrt{(a_1^T)^2 + m_1^{k\,2}}); \\ \sin^2(\theta_1^k) = (\frac{a_1^T}{\sqrt{(a_1^T)^2+m_1^{k\,2}}})^2 \sin^2(\sqrt{(a_1^T)^2 + m_1^{k\,2}}), \end{cases} \tag{28}$$

to estimate the parameter set $(\hat{a}_1^T, \hat{m}_1^k)$, where $\hat{m}_1^k = \hat{B}_i^k := \sum_{1,j} \lambda_{1j}^k \hat{c}_{1j} T$ and λ_{1j}^k is the coefficient for applying $Z_1 Z_j$ towards the basis $|v_m\rangle_k$. Given the variance for both algorithms in Eq. (26) and (27), the performance of the estimator pair $(\hat{a}_1^T, \hat{m}_1^k)$ for each invariant subspace can be derived with the relation given in Eq. (28). Since the estimator pair $(\hat{a}_1^T, \hat{m}_1^k)$ accounts for the time-integral of the Hamiltonian parameters, we can utilize a proper evolution time T for the Hamiltonian, forcing $p = \sqrt{(a_1^T)^2 + m_1^{k\,2}}$ to be small. Under the small-angle approximation ($\sin p \approx p$ and $\tan p \approx p$), the variances for the pair $(\hat{a}_1^T, \hat{m}_1^k)$ are approximately

$$\mathrm{Var}(\hat{a}_1^T) \approx \mathrm{Var}(\hat{\theta}_1^k) \approx \frac{1}{8Nd^2}, \tag{29}$$

$$\mathrm{Var}(\hat{m}_1^k) \approx \mathrm{Var}(\hat{\zeta}_1^k) \approx \frac{3}{8Nd^4 a_1^2}, \tag{30}$$

for each of the $(n-1)$ invariant subspaces, assuming Algorithm 3. The same analysis also applies to Algorithm 2, that is,

$$\mathrm{Var}(\tilde{a}_1^T) \approx \mathrm{Var}(\tilde{\theta}_1^k) \approx \frac{n}{4Nd^2}; \tag{31}$$

$$\mathrm{Var}(\tilde{m}_1^k) \approx \mathrm{Var}(\tilde{\zeta}_1^k) \approx \frac{3n}{4Nd^4(a_1)^2}, \tag{32}$$

for all $(n-1)$ invariant subspaces. Lastly, we estimate our desired quantities $\hat{c}_{1i}$'s by solving the linear system comprising $\hat{m}_1^k$'s. Therefore, to further provide the variance for the Hamiltonian parameters, we should consider the linear relation between c_{1i}'s and m_1^k's, i.e.

$$\Lambda \hat{\mathbf{c}}_1 = \hat{\mathbf{m}}_1, \tag{33}$$

where $\hat{\mathbf{m}}_1 := (\hat{m}_1^1, \hat{m}_1^2, \cdots, \hat{m}_1^{n-1})$. The coefficient matrix Λ, with all entries equal to $\pm T$, encodes the linear combination relationships. To ensure that it provides both necessary and sufficient conditions for solving all $\hat{\mathbf{m}}_1$ terms, Λ must be of full rank. The choice of Λ is determined by the selection of $(n-1)$ invariant subspaces, one example of which is discussed in Definition II.10. Based on the linear relation, the covariance matrix of $\hat{\mathbf{c}}_1$ is

$$\Sigma_{\hat{\mathbf{c}}_1} = \Lambda^{-1} \Sigma_{\hat{\mathbf{m}}_1} \Lambda^{-T}, \tag{34}$$

where $\Sigma_{\hat{\mathbf{m}}_1} = \mathrm{diag}(\mathrm{Var}(\hat{m}_1^1), \mathrm{Var}(\hat{m}_1^2), \cdots, \mathrm{Var}(\hat{m}_1^{n-1}))$. Therefore, the variance of $\hat{c}_{1i}$ is the (i,i)th entry of the covariance matrix $\Sigma_{\hat{\mathbf{c}}_1}$, i.e.

$$\mathrm{Var}(\hat{c}_{1i}) = \sum_{j=1}^{n-1} (\Lambda^{-1}[i,j])^2 \mathrm{Var}(\hat{m}_1^j). \tag{35}$$

For Algorithm 3, we plug in Eq. (30). The variance of $\hat{c}_{1i}$ terms is in the form of

$$\mathrm{Var}(\hat{c}_{1i}) \approx \frac{3}{8Nd^4a_1^2}\sum_{j=1}^{n-1}(\Lambda^{-1}[i,j])^2. \quad (36)$$

This leads to a general bound on the variance of $\hat{\mathbf{c}}_\mathbf{1}$ as the average entry of Λ^{-1} generally scales as $O(1)$. Therefore,

$$\mathrm{Var}(\hat{c}_{1i}) \approx O(\frac{3}{8Nd^4a_1^2}), \quad (37)$$

where N is the number of shots, d is the number of repetition cycles. Similarly, for Algorithm 2, Eq. (32) is utilized and leads to

$$\mathrm{Var}(\tilde{c}_{1i}) \approx O(\frac{3n}{4Nd^4a_1^2}). \quad (38)$$

Further, the whole learning algorithm involves $(n-1)$ different full Hamiltonian H_i for which the above analysis can be directly applied to the estimator $\hat{\mathbf{c}}_\mathbf{i} = (\hat{c}_{i(i+1)}, \cdots, \hat{c}_{in})$ corresponds to each H_i. In general, the variance of the estimator $\hat{\mathbf{c}}$ for both algorithms can be summarized as

$$\mathrm{Var}(\hat{c}_{ij}) \approx O\left(\frac{3}{8Nd^4a_i^2}\right), \quad \text{for Algorithm 3} \quad (39)$$

$$\mathrm{Var}(\hat{c}_{ij}) \approx O\left(\frac{3n}{4Nd^4a_i^2}\right), \quad \text{for Algorithm 2} \quad (40)$$

where a_i is the coefficient of the single X_i term for H_i and $i \in \{1, \cdots, (n-1)\}, j \in \{(i+1), \cdots, n\}$.

Based on the computed variance, we show that the Heisenberg limit performance of the estimators is attained in our learning algorithm in both scenarios. The classical resource used in the learning algorithm is $N \times 2(2d-1)$, and the quantum resource used is d for each parameter. The Heisenberg limit suggests the variance scales as

$$\Omega\left(\frac{1}{(\text{Classical resource} \times \text{Quantum resource}^2)}\right) \quad (41)$$

$$= \frac{1}{(N \times 2(2d-1)) \times (d)^2} = \Omega(\frac{1}{Nd^3}). \quad (42)$$

Our estimators, with variances given in Eq. (39) and Eq. (40), achieve the Heisenberg limit, exhibiting a faster variance reduction that scales as $O(1/d^4)$, similar to the performance achieved by the QSPE method [17].

Moreover, we further demonstrate the optimality of the estimators with respect to classical post-processing by showing that the calculated variance attains the Cramér–Rao lower bound to complete the proof of the Theorem II.6. With the same setting in the discussion of the variance computation, we assume the full Hamiltonian H_i for estimating $(n-i)$ Hamiltonian parameters via $(n-i)$ invariant subspaces. Each kth invariant subspace solves a pair (θ_i^k, ζ_i^k) which can be used to infer (a_i, m_i^k) based on the mapping given in Eq. (28). Similarly, for Algorithm 3, the optimal variances of the estimator pair $(\hat{\theta}_i^k, \hat{\zeta}_i^k)$ are computed from the Cramér–Rao lower bound discussed in [17], that is,

$$\mathrm{Var}(\hat{\theta}_i^k)_{opt} \approx \frac{1}{8Nd^2}, \mathrm{Var}(\hat{\zeta}_i^k)_{opt} \approx \frac{3}{8Nd^4(\theta_i^k)^2}. \quad (43)$$

For Algorithm 2, we recompute the optimal variance of $(\tilde{\theta}_i^k, \tilde{\zeta}_i^k)$ in the kth invariant subspace in Supplementary Material Sec.VI, which are

$$\mathrm{Var}(\tilde{\theta}_i^k)_{opt} \approx \frac{n}{4Nd^2}, \mathrm{Var}(\tilde{\zeta}_i^k)_{opt} \approx \frac{3n}{4Nd^4(\theta_1^k)^2}. \quad (44)$$

Based on the same analysis that we developed for variance computation for estimators, the optimal variance for the Hamiltonian parameter estimators can be derived based on Eq. (43) and Eq. (44). To conclude, the optimal variance of the estimator $\hat{\mathbf{c}}$ for both algorithms can be summarized as

$$\mathrm{Var}(\hat{c}_{ij})_{opt} \approx O\left(\frac{3}{8Nd^4a_i^2}\right), \quad \text{for Algorithm 3} \quad (45)$$

$$\mathrm{Var}(\hat{c}_{ij})_{opt} \approx O\left(\frac{3n}{4Nd^4a_i^2}\right), \quad \text{for Algorithm 2} \quad (46)$$

where a_i is the coefficient of the single X_i term for H_i and $i \in \{1, \cdots, (n-1)\}, j \in \{(i+1), \cdots, n\}$. It is straightforward to verify that the optimal variance given by the Cramér-Rao lower bound agrees with the computed variance, indicating that the classical post-processing component of our learning algorithms is optimal.

III. EXPERIMENT DEPLOYMENT

In this section, we consider the realistic scenarios for deploying the learning algorithms on NISQ devices, taking into account various noise sources commonly present in current hardware and providing a detailed experimental proposal for NISQ demonstration of our algorithm.

1. *Robustness against realistic error*

a. Decoherence The unavoidable disturbance from the environment influences the quantum dynamics, thus might lead to the wrong estimation using our algorithms. Here we model the decoherence effect by the depolarizing channel resulting in the mismatch of the probability from the quantum part of Algorithm 1, i.e.

$$p'_{X(Y)} = \alpha p_{X(Y)} + \frac{1-\alpha}{4}, \quad (47)$$

where $\alpha \in [0,1]$ quantifies the fidelity of the quantum evolution. Here, we analyze the performance of one specific invariant subspace, but due to the parallelization structure, the analysis can be generalized to the whole process. The probabilities coming from the noisy dynamics suffer a rescaled factor and a constant shift, thus the corresponding reconstructed function h used in Algorithm 1 also changes accordingly, i.e. $h' = \alpha h - \frac{1-\alpha}{4}(1+i)$. The Fourier coefficients of h are expected to have an additional rescaling factor α, and the constant shift only contributes to the change of zeroth Fourier coefficients, namely,

$$|c_0'| = |\alpha c_0 - \frac{1-\alpha}{4}(1+i)| \approx \alpha\theta + \frac{1-\alpha}{2\sqrt{2}}, \tag{48}$$

$$|c_k| \approx \alpha\theta, \ \forall k = 1, \cdots, d-1. \tag{49}$$

Therefore, utilizing the relation between the zeroth order and the remaining Fourier coefficients [17], the estimator in Algorithm 1 can be redefined as follows:

$$\hat{\alpha} = 1 - 2\sqrt{2}(|c_0| - \frac{1}{d-1}\sum_{k=1}^{d-1}|c_k|), \quad \hat{\theta} = \frac{1}{\hat{\alpha}} \times \frac{1}{d-1}\sum_{k=1}^{d-1}|c_k|. \tag{50}$$

The rescaling estimator mitigates decoherence errors in the algorithm by leveraging the fact that the Fourier coefficients encode information about the estimated fidelity.

b. Coherent state preparation error The state preparation error is also dominant in quantum systems. Especially, in our state preparation step of Algorithm 1, we need to accurately applying $-X(\frac{\pi}{4})$ and $Y(\frac{\pi}{4})$ pulse towards the first atom (See Supplementary Material Sec. III for details), where miscalibration of the pulses leads to imperfect initial states and thus affects the estimation process. Here, we test the robustness of the algorithm against state preparation error in the following setting. Instead of the perfect pulse in Eq. (S15) & (S14), we assume a coherent error happening in the pulse, which leads to over-rotation, i.e.

$$\begin{aligned} e^{-i(\theta+\alpha)X_1\otimes I}|00\rangle &= \cos(\theta+\alpha)|00\rangle - i\sin(\theta+\alpha)|10\rangle \\ e^{-i(\theta+\alpha)Y_1\otimes I}|00\rangle &= \cos(\theta+\alpha)|00\rangle + \sin(\theta+\alpha)|10\rangle; \end{aligned} \tag{51}$$

where α denotes the error rate. Here we provide a bound on the difference of the estimation with and without coherent state preparation error using the original estimator and the scaled estimator.

Theorem III.1. *Consider the coherent error Eq. (51) with error rate α in the state preparation step of Algorithm 1, the difference between the noisy estimation $\hat{\theta}' = \frac{1}{d}\sum_{k=0}^{d-1}|c_k'|$ and noiseless estimation $\hat{\theta} = \frac{1}{d}\sum_{k=0}^{d-1}|c_k|$ can be bounded by*

$$D = |\hat{\theta}' - \hat{\theta}| \leq \sqrt{2}(d+1)^2 \sin 2\alpha \sin^2\theta + (\cos 2\alpha - 1)\theta, \tag{52}$$

which scales with $O(\theta)$. The difference can be improved by introducing a scaled estimator $\hat{\theta}_s' = \frac{1}{\cos 2\alpha} \times \frac{1}{d}\sum_{k=0}^{d-1}|c_k'|$. The corresponding bound then reduces to

$$D_s = |\hat{\theta}_s' - \hat{\theta}| \leq \sqrt{2}(d+1)^2 \tan 2\alpha \sin^2\theta, \tag{53}$$

where scales only with $O(\theta^2)$.

Unlike the simple scaling and constant shift introduced by the depolarizing channel to the function h, coherent state preparation errors contribute both a rescaling factor and a shift that depend on the true parameter value θ. However, by applying the same rescaling technique used to address decoherence, we reduce the error from first order to second order. As a result, when θ is sufficiently small, the induced error becomes negligible. The detail is discussion in Supplementary Material Sec. IV.

c. Readout error Another predominant error in practice comes from state measurement. For example, atom loss and imperfect re-trapping can introduce measurement errors in Rydberg atom quantum simulators. More specifically, a ground state ($|0\rangle$) can be misidentified as a Rydberg state ($|1\rangle$) due to atom loss, with probability $P_{\text{loss}} = \mathbb{P}(1|0) = 0.01$; and a Rydberg state ($|1\rangle$) can be misread as a ground state ($|0\rangle$) due to imperfect anti-trapping, with probability $P_{\text{anti-trapping}} = \mathbb{P}(0|1) = 0.08$ [32].

We mitigate the readout error via the confusion matrix, which stores the conditional probability of measurement outcome being the bit-string b_i given the quantum state is $|b_j\rangle$, i.e.

$$R := [\mathbb{P}(b_i|b_j)_{b_i,b_j\in\{0,1\}^n}], \tag{54}$$

for n qubit system. Take the two-atom system as an example, the confusion matrix is

$$R = \begin{bmatrix} \mathbb{P}(00|00) & \mathbb{P}(01|00) & \mathbb{P}(10|00) & \mathbb{P}(11|00) \\ \mathbb{P}(00|01) & \mathbb{P}(01|01) & \mathbb{P}(10|01) & \mathbb{P}(11|01) \\ \mathbb{P}(00|10) & \mathbb{P}(01|10) & \mathbb{P}(10|10) & \mathbb{P}(11|10) \\ \mathbb{P}(00|11) & \mathbb{P}(01|11) & \mathbb{P}(10|11) & \mathbb{P}(11|11) \end{bmatrix}.$$

The readout error in the measured outcome vector $\mathbf{q} = (q_{00}, q_{01}, q_{10}, q_{11})^{\mathrm{T}}$ can be mitigated by applying a confusion matrix correction, given by

$$\mathbf{p} = (R^T)^{-1}\mathbf{q}, \tag{55}$$

where $\mathbf{p}$ denotes the corrected probability vector.

d. Time-dependent coherent error Time-dependent coherent error is the most challenging adversarial in the learning task. In previous methods that operate in the real space, such as robust phase estimation [16] and periodic calibration [2], the relationship between the unknown parameters becomes entangled through complex trigonometric and inverse trigonometric functions. During inference, this leads to a highly non-convex optimization problem involving both parameters. Consequently, an error in estimating

one parameter can severely affect the accuracy of the other. However, our learning algorithm does the inference on the Fourier space, decoupling the multiple parameters into largely orthogonal components. To be more specific, the estimators designed in our learning algorithm are based on the relation between the Fourier coefficients and the parameters in each invariant subspace as follows,

$$c_k \approx i\theta e^{-i(2k+1)\zeta}, \forall k = 0, \cdots, d-1. \tag{56}$$

Here we completely separate two parameters, where θ depends on the amplitude of the Fourier coefficients and the phase of the coefficients implies ζ. Further, the desired estimators of each invariant subspace can be inferred with the mapping Eq. (28), which are also decoupled under small-angle approximation [36]. As a result, this decoupling allows accurate estimation of one of them (e.g. c_{ij} term) while the other is suffering from time-dependent coherent error (e.g. the drifting error on the local field a_i). To model the coherent error, we consider that there exists a γ shift on the local fields, resulting in the coherent change in Hamiltonian to $H_i = (a_i + \gamma a_i)X_i + \sum Z_i Z_j$ in the simulation.

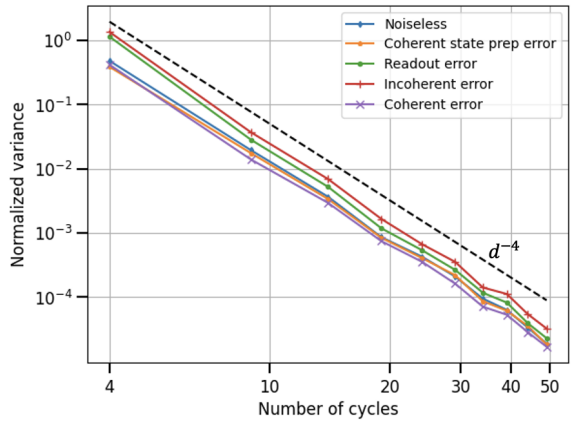


FIG. 3: Robustness test against coherent state preparation error(orange), readout error(green), depolarizing error(red) and time-dependent coherent error(purple) with the number of samples $N_{\text{shot}} = 10^5$ and the number of boostraps $N_{\text{bootstrap}} = 10^3$. The noise model is chosen based on realistic noise model for Rydberg system [32], where the readout error is $(p_{\text{loss}}, p_{\text{anti-trapping}}) = (0.01, 0.08)$, the depolarizing noise has fidelity $\alpha = 0.8$, coherent state preparation error $\alpha = 0.01$ and the coherent error on X_i is $\gamma = 10\%$. The normalized variance is defined as the relative variance $\text{Var}(\hat{c})/c^2$. The simulation results exhibit $\propto d^{-4}$ scaling in both noisy and noiseless scenarios, suggesting the robustness of the learning algorithm against experimental noise.

To conclude the discussion on the robustness of our learning algorithm, we conducted a comprehensive study on the robustness of our learning algorithm under various realistic noise sources. The simulation results are presented in Figure. 3, which showcases the algorithm's performance within an invariant subspace of a 10-qubit all-to-all connected system governed by the Hamiltonian in Eq. (3). Due to the parallelized structure of the algorithm, its performance remains consistent across systems of different sizes. All curves in the figure exhibit a $1/d^4$ scaling, consistent with Theorem II.6, indicating that the algorithm maintains its effectiveness in both noiseless and noisy environments. These results confirm the algorithm's robustness against readout errors, state preparation errors, depolarizing noise, and time-dependent coherent errors.

2. *Experimental proposal*

The algorithm's robustness against all major sources of realistic error highlights its potential for practical deployment on current NISQ devices to learn device-specific parameters. In this section, we provide a detailed implementation strategy for the Rydberg atom platform, including the experimental procedure, parameter settings, and simulation results. While the proposal is specifically tailored for learning two-atom Rydberg interactions, it can be naturally extended to systems with more atoms following Algorithm 3. Additionally, the same framework can be applied to superconducting platforms for learning XY couplings with the parallel-learnable Hamiltonain discussed in Supplementary Material Sec. I.

a. Experiment procedure Consider the scenario for the Rydberg Hamiltonian involving two atoms with the corresponding Hamiltonian

$$H(\Omega, \phi, R) = \frac{\Omega}{2}(\cos\phi X_1 - \sin\phi Y_1) + c_{12} Z_1 Z_2, \tag{57}$$

where Ω is the Rabi frequency, ϕ is the phase and $c_{12} = \frac{C_6}{R^6}$ is the interaction strength. This is a general version of the full Hamiltonian given in Eq. (3) with phase ϕ between X_1 and Y_1, we can set the $\phi = 0$ to get back to Eq. (3) as the phase information is irrelevant here. For this Hamiltonian, we are interested in inferring the atom-atom distance R.

Similarly, consider the time-evolution operator for time T, where the accumulated angles are $\int_0^T \frac{\Omega}{2} dt = a$ and $\int_0^T c_{12} dt = b$. The corresponding mapping for accumulated parameters is,

$$\begin{cases} \tan(\zeta) = \frac{b}{\sqrt{a^2+b^2}} \tan(\sqrt{a^2+b^2}) \\ \tan(\chi) = \tan(\phi) \\ \sin^2(\theta) = (\frac{a}{\sqrt{a^2+b^2}})^2 \sin^2(\sqrt{a^2+b^2}) \end{cases} \tag{58}$$

Dividing evolution time T gives back to our desired terms, i.e., $(\Omega, c_{12}) = (a/T, b/T)$.

The physical implementation of the learning algorithm for the Rydberg system can be realized by considering the two main components of Algorithm 1.

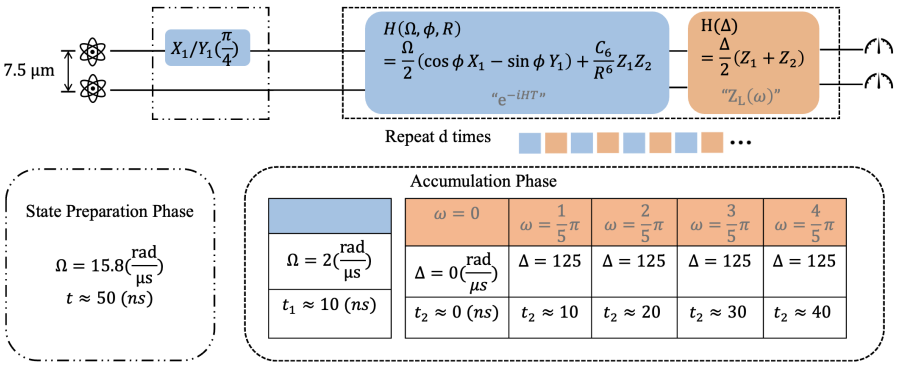


FIG. 4: Physical implementation on the Rydberg quantum simulator. The first dotted block is for state preparation, where $|+\rangle_l$ and $|i\rangle_l$ should be prepared according to Supplementary Material Sec. III. The second dotted block evolves d-repetition of target quantum dynamics and Z_l phase accumulation. The whole evolution should run for $\omega = \frac{j}{2d-1}\pi, j = 0, \cdots, 2d-2$. $d = 3$ is picked here for reference.

atom distance R	coefficient $V = \frac{C_6}{R^6}$	evolution time t	effective integral $b = \int_0^T \frac{C_6}{R^6} dt$
$7.16\mu m$	$40\ rad/\mu s$	$1e-03\mu s$	$0.04\ rad$
$7.52\mu m$	$30\ rad/\mu s$	$1e-03\mu s$	$0.03\ rad$
$8.04\mu m$	$20\ rad/\mu s$	$1e-03\mu s$	$0.02\ rad$

TABLE II: Table of benchmark cases for testing the learning algorithm on Rydberg device. The evolution time can be chosen flexibly while satisfying the hardware constraint. The parameters listed here are used in simulation for reference.

1. Quantum part: The experimental realization of the algorithm is shown in the Figure. 4. By adding up the time take for two phases in the Figure. 4, the total evolution time required on the device for one tunable ω point can be roughly analyzed as,

$$T_{\text{total}} = t_{\text{state prep}} + t_{\text{repetition}} = t + (t_1 + t_2) \times d \approx 180ns. \tag{59}$$

As a result, the total evolution time for all $(2d-1)$ sampled ω for the complete inference steps is then

$$T_{\text{round}} = 2 \times (2d-1) \times T_{\text{total}} \approx 1.8\mu s, \tag{60}$$

where 2 stands for two different initial states and $(2d-1)$ is for different tunable ω terms.

2. Classical post-processing: Following the Algorithm 1, the estimators $(\hat{\theta}, \hat{\zeta})$ converts back to the accumulated angles $(\hat{a}, \hat{b})$. The estimator of the atom distance R follows

$$\hat{R} = (\frac{C_6}{\hat{b}/T})^{1/6}, \tag{61}$$

where T is the evolution time and $C_6 = 5,420,503\ \mu m^6 rad/\mu s$ is a constant. Notably, under the proposed parameters, the total evolution time per experimental round is approximately $T_{\text{round}} \approx 1.8\mu s$, which is significantly shorter than the system's T_1 and T_2 time [32], highlighting the algorithm's feasibility for near-term implementation.

b. Parameter setting The Table II summarizes the parameters used for testing on Rydberg devices. The selected inter-atomic distances are motivated by the findings in [1], which show significant deviations between experimental performance and theoretical predictions at those distances. Set the local field Ω such that the accumulated angle $a = \int_0^T \frac{\Omega(t)}{2} dt = 0.01\ rad$, we ensure that $d\theta \ll 1$, under which the variance of $\hat{b}$ scales with the repetition depth as $\text{Var}(\hat{b}) \sim O(1/d^4)$. Further, since the atom-atom interaction strength in the Rydberg system is in the form of $b(R) = C_6/R^6$, the variance relation between them is

$$\text{Var}(b(R)) \approx b(R)' \text{Var}(R) = (\frac{6b(R)}{R})^2 \text{Var}(R). \tag{62}$$

As a result, the scaling between the variance of $\hat{R}$ and repetition depth d is also $\text{Var}(\hat{R}) \sim O(1/d^4)$.

c. Simulation results For the parameters listed in the Table. II, we simulate the performance of our learning algorithm according to the experiment procedure discussed in the previous section. The results are shown in the Figure. 5, where the atom distances are $(R_1, R_2, R_3) = (7.16\mu m, 7.52\mu m, 8.04\mu m)$. The figure shows the identical scaling with respect to $\hat{b}$ and $\hat{R}$ satisfying the relation in Eq. (62). Besides, we notice from the simulation result that with the variance scaling as $O(1/d^4)$, an accurate estimation of the atom distance only requires shallow depths of dynamics repetition. From the Figure. 5, to learn the distance with standard deviation $\delta R \approx 10^{-7}$, i.e. normalized variance

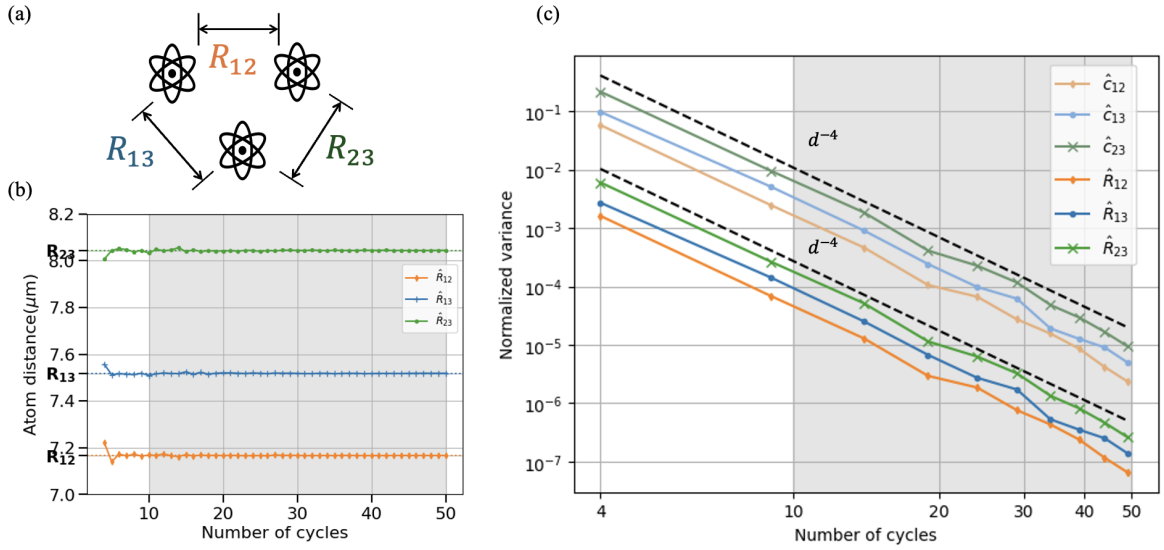


FIG. 5: Application of the learning algorithm on Rydberg quantum simulators for atom distance. (a) Atom distance learning with three different distance ($R_{12} = 7.16\mu m, R_{13} = 7.52\mu m, R_{23} = 8.04\mu m$) (b) The estimated distance of three estimators ($\hat{R}_{12}, \hat{R}_{13}, \hat{R}_{23}$). The bold x-ticks are the true value (R_{12}, R_{13}, R_{23}). After $d = 10$ cycles, the estimations of all distances are consistently correct. (c) The upper block of curves (light-green, light-blue, light-orange) represents the variance of the estimators ($\hat{c}_{12}, \hat{c}_{13}, \hat{c}_{23}$) with respect to the number of cycles d, where c_{ij} is the parameter of term Z_iZ_j in Eq. (3). While the lower block of curves(green, blue, orange) is the variance of atom distance estimators ($\hat{R}_{12}, \hat{R}_{13}, \hat{R}_{23}$) with respect to d after conversion given in Eq. (61). The simulation results exhibit $\propto d^{-4}$ scaling, in agreement with Theorem II.6 for all estimators.

$\frac{\delta R}{R} \approx 0.01$, only $d = 10$ repetition is needed.

IV. CONCLUSION

In this work, we present the first in-situ Hamiltonian learning algorithm that is robust against a broad range of systematic and environmental noise. By exploiting the system Hamiltonian's parallel structure, the algorithm achieves a quadratic speedup in sample complexity relative to the number of parameters compared with existing approaches for general Hamiltonian models [18–22]. This demonstrates the potential of leveraging known Hamiltonian structures to efficiently certify quantum devices. Another important advantage of our learning algorithm is its ability to perform in-situ learning: all couplings remain active during the process, allowing us to capture crosstalk effects that arise only in the full many-body setting. This leads to a more accurate characterization of noise in current quantum systems compared to previous pairwise calibration approaches [1, 3, 24]. Looking ahead, this capability could be extended to enable calibration protocols dealing with correlated noise, provide deeper insights into non-Markovian noise processes, and guide the design of error-mitigation strategies that consider many-body interaction effects.

Besides, the algorithm attains the optimal precision and remains resilient to SPAM, decoherence, and time-dependent coherent errors. These results highlight its potential for practical deployment on NISQ devices, enabling high-precision learning of structured Hamiltonians in realistic settings. We provide the end-to-end proposal for learning distance error for Rydberg atom arrays. Unlike the previous approaches, which are only capable to learn distance pairwisely, we introduce the first in-situ learning algorithm for Rydberg atom distance learning. This feature is particularly crucial for Rydberg systems, where couplings cannot be naturally switched off without physically moving the atoms. Moreover, since our approach does not rely on long-time evolutions and is applicable to both hybrid and fully analog devices, the proposal is directly compatible with current experimental platforms. Our method is also applicable to other platforms with a parallel-learnable Hamiltonian with no additional overhead. It will also be interesting to check whether other experiment-induced Hamiltonians can be converted to parallel-learnable via realizable unitary transformations [37]. Additionally, the broader question of how prior structural knowledge can further reduce learning costs remains an open and valuable direction for future research. A comprehensive evaluation across various physical devices, through either simulation or real quantum experiments, is also left for future investigation.

ACKNOWLEDGMENTS

We thank Grecia Castelazo and Srijan Nikhar for helpful discussions. S.L. and X.W. were supported by Air Force Office of Scientific Research under award number FA9550-21-1-0209, the U.S. National Science Foundation grant CCF-1942837 (CAREER), and a Sloan research fellowship. M. N. is supported by the U.S. National Science Foundation grant CCF-2441912(CAREER), Air Force Office of Scientific Research under award number FA9550-25-1-0146, and the U.S. Department of Energy, Office of Advanced Scientific Computing Research under Award Number DE-SC0025430.

[1] C. B. Dag, H. Ma, P. M. Eugenio, F. Fang, and S. F. Yelin, Emergent disorder and sub-ballistic dynamics in quantum simulations of the ising model using rydberg atom arrays, arXiv preprint arXiv:2411.13643 (2024).

[2] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, A. Bengtsson, S. Boixo, M. Broughton, B. B. Buckley, *et al.*, Observation of separated dynamics of charge and spin in the fermi-hubbard model, arXiv preprint arXiv:2010.07965 (2020).

[3] T. I. Andersen, N. Astrakhantsev, A. H. Karamlou, J. Berndtsson, J. Motruk, A. Szasz, J. A. Gross, A. Schuckert, T. Westerhout, Y. Zhang, *et al.*, Thermalization and criticality on an analogue–digital quantum simulator, Nature **638**, 79 (2025).

[4] A. H. Karamlou, I. T. Rosen, S. E. Muschinske, C. N. Barrett, A. Di Paolo, L. Ding, P. M. Harrington, M. Hays, R. Das, D. K. Kim, *et al.*, Probing entanglement in a 2d hard-core bose–hubbard lattice, Nature **629**, 561 (2024).

[5] T. Manovitz, S. H. Li, S. Ebadi, R. Samajdar, A. A. Geim, S. J. Evered, D. Bluvstein, H. Zhou, N. U. Koyluoglu, J. Feldmeier, *et al.*, Quantum coarsening and collective dynamics on a programmable simulator, Nature **638**, 86 (2025).

[6] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, *et al.*, Logical quantum processor based on reconfigurable atom arrays, Nature **626**, 58 (2024).

[7] M. D. Shulman, S. P. Harvey, J. M. Nichol, S. D. Bartlett, A. C. Doherty, V. Umansky, and A. Yacoby, Suppressing qubit dephasing using real-time hamiltonian estimation, Nature communications **5**, 5156 (2014).

[8] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, Universal quantum control through deep reinforcement learning, npj Quantum Information **5**, 33 (2019).

[9] D. Hangleiter, I. Roth, J. Fuksa, J. Eisert, and P. Roushan, Robustly learning the hamiltonian dynamics of a superconducting quantum processor, Nature Communications **15**, 9595 (2024).

[10] T. Olsacher, T. Kraft, C. Kokail, B. Kraus, and P. Zoller, Hamiltonian and liouvillian learning in weakly-dissipative quantum many-body systems, Quantum Science and Technology **10**, 015065 (2025).

[11] M. De Burgh and S. D. Bartlett, Quantum methods for clock synchronization: Beating the standard quantum limit without entanglement, Physical Review A—Atomic, Molecular, and Optical Physics **72**, 042301 (2005).

[12] D. Leibfried, M. D. Barrett, T. Schaetz, J. Britton, J. Chiaverini, W. M. Itano, J. D. Jost, C. Langer, and D. J. Wineland, Toward heisenberg-limited spectroscopy with multiparticle entangled states, Science **304**, 1476 (2004).

[13] K. McKenzie, D. A. Shaddock, D. E. McClelland, B. C. Buchler, and P. K. Lam, Experimental demonstration of a squeezing-enhanced power-recycled michelson interferometer for gravitational wave detection, Physical review letters **88**, 231102 (2002).

[14] R. R. Allen, F. Machado, I. L. Chuang, H.-Y. Huang, and S. Choi, Quantum computing enhanced sensing, arXiv preprint arXiv:2501.07625 (2025).

[15] J. Preskill, Fault-tolerant quantum computation, in *Introduction to quantum computation and information* (World Scientific, 1998) pp. 213–269.

[16] S. Kimmel, G. H. Low, and T. J. Yoder, Robust calibration of a universal single-qubit gate set via robust phase estimation, Physical Review A **92**, 062315 (2015).

[17] Y. Dong, J. A. Gross, and M. Y. Niu, Optimal low-depth quantum signal-processing phase estimation, Nature Communications **16**, 1504 (2025).

[18] H.-Y. Huang, Y. Tong, D. Fang, and Y. Su, Learning many-body hamiltonians with heisenberg-limited scaling, Physical Review Letters **130**, 200403 (2023).

[19] H.-Y. Hu, M. Ma, W. Gong, Q. Ye, Y. Tong, S. T. Flammia, and S. F. Yelin, Ansatz-free hamiltonian learning with heisenberg-limited scaling, arXiv preprint arXiv:2502.11900 (2025).

[20] M. Ma, S. T. Flammia, J. Preskill, and Y. Tong, Learning k-body hamiltonians via compressed sensing, arXiv preprint arXiv:2410.18928 (2024).

[21] H. Li, Y. Tong, T. Gefen, H. Ni, and L. Ying, Heisenberg-limited hamiltonian learning for interacting bosons, npj Quantum Information **10**, 83 (2024).

[22] H. Ni, H. Li, and L. Ying, Quantum hamiltonian learning for the fermi-hubbard model, Acta Applicandae Mathematicae **191**, 2 (2024).

[23] A. L. Shaw, Z. Chen, J. Choi, D. K. Mark, P. Scholl, R. Finkelstein, A. Elben, S. Choi, and M. Endres, Benchmarking highly entangled states on a 60-atom analogue quantum simulator, Nature **628**, 71 (2024).

[24] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, *et al.*, Probing many-body dynamics on a 51-atom quantum simulator, Nature **551**, 579 (2017).

[25] D. Bluvstein, A. Omran, H. Levine, A. Keesling, G. Semeghini, S. Ebadi, T. T. Wang, A. A. Michailidis, N. Maskara, W. W. Ho, *et al.*, Controlling quantum many-body dynamics in driven rydberg atom arrays, Science **371**, 1355 (2021).

[26] A. Browaeys and T. Lahaye, Many-body physics with individually controlled rydberg atoms, Nature Physics **16**, 132 (2020).

[27] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, Annual Review of Condensed Matter Physics **11**, 369 (2020).
[28] I. Siddiqi, Engineering high-coherence superconducting qubits, Nature Reviews Materials **6**, 875 (2021).
[29] A. Sørensen and K. Mølmer, Entanglement and quantum computation with ions in thermal motion, Physical Review A **62**, 022311 (2000).
[30] M. Morgado and S. Whitlock, Quantum simulation and computing with rydberg-interacting qubits, AVS Quantum Science **3** (2021).
[31] The introduction to the QSPE method can be found in the Supplementary Material Sec. II.
[32] J. Wurtz, A. Bylinskii, B. Braverman, J. Amato-Grill, S. H. Cantu, F. Huber, A. Lukin, F. Liu, P. Weinberg, J. Long, *et al.*, Aquila: Quera's 256-qubit neutral-atom quantum computer, arXiv preprint arXiv:2306.11727 (2023).
[33] H. Tamura, T. Yamakoshi, and K. Nakagawa, Analysis of coherent dynamics of a rydberg-atom quantum simulator, Physical Review A **101**, 043421 (2020).
[34] M. Marcuzzi, J. Minář, D. Barredo, S. De Léséleuc, H. Labuhn, T. Lahaye, A. Browaeys, E. Levi, and I. Lesanovsky, Facilitation dynamics and localization phenomena in rydberg lattice gases with position disorder, Physical Review Letters **118**, 063606 (2017).
[35] Here we apply the time-independent pulse to the system, the accumulation angle is just the time-integral for the square pulse.
[36] The Eq. (28) maps the estimated pair (θ, ζ) to the time-integral of our desired estimators, we can choose evolution time properly to approximately estimate a^T from θ and m^k via ζ.
[37] T. S. Cubitt, A. Montanaro, and S. Piddock, Universal quantum hamiltonians, Proceedings of the National Academy of Sciences **115**, 9497 (2018).

Supplementary Material

I. PARALLEL-LEARNABLE HAMILTONIAN

In this section, we provide additional details on the structure of parallel-learnable Hamiltonians and the construction of corresponding unitary transformations. As described in the main text, an n-qubit Hamiltonian is said to be *parallel-learnable* if it admits a decomposition into 2^{n-1} invariant subspaces, under a specific basis. This structure enables the parallel estimation of multiple parameters by tackling them within independent 2×2 subspaces simultaneously. To obtain such a block-diagonal form, we apply a unitary transformation U that maps the Hamiltonian from the computational basis to a basis where the invariant subspaces are presented. In principle, any Hamiltonian H can be brought into a block-diagonal form consisting of non-diagonal 2×2 blocks through a specially designed transformation $U := RV$, where V is the eigenbasis matrix that diagonalizes H, and R is a block-wise rotation that converts the diagonal 2×2 blocks into non-diagonal ones. This transformation results in a Hamiltonian of the form described in Eq. (1) of the main text. In the following, we elaborate on this construction for general Hamiltonians.

Consider the diagnolization of the Hamiltonain H, i.e.,

$$H = V \Lambda V^\dagger,$$

where V is the eigenbasis matrix, consisting all the eigenvectors of H as the columns. Changing the basis of H from computational basis to eigenbases results in the diagonal structure,

$$\Lambda = V^\dagger H V = \oplus_{k=1}^{2^{n-1}} \Lambda_k, \quad \text{where} \quad \Lambda_k = \begin{bmatrix} \lambda_k^1 & 0 \\ 0 & \lambda_k^2 \end{bmatrix} \quad \text{and} \quad \{\lambda_k^i\} \quad \text{are eigenvalues of } H. \tag{S1}$$

The resultant matrix Λ after basis change is a direct sum of 2×2 diagonal matrices, which does not fit in the definition. So we define another unitary transformation matrix,

$$R = \oplus_{k=1}^{2^{n-1}} R_k(\theta_k), \quad \text{where} \quad R_k(\theta_k) = \begin{bmatrix} \cos\theta_k & -\sin\theta_k \\ \sin\theta_k & \cos\theta_k \end{bmatrix}. \tag{S2}$$

R is a block-diagonal rotation matrix acting on each 2×2 block. So after applying the unitary transformation R, the matrix is transformed to

$$H' = R \Lambda R^\dagger = \oplus_{k=1}^{2^{n-1}} H_k, \quad \text{where} \quad H_k = \begin{bmatrix} \lambda_k^1 \cos^2\theta_k + \lambda_k^2 \sin^2\theta_k & (\lambda_k^1 - \lambda_k^2)\cos\theta_k \sin\theta_k \\ (\lambda_k^1 - \lambda_k^2)\cos\theta_k \sin\theta_k & \lambda_k^1 \sin^2\theta_k + \lambda_k^2 \cos^2\theta_k \end{bmatrix}. \tag{S3}$$

H_k's are non-diagonal if $\lambda_k^1 \neq \lambda_k^2$ and $\theta_k \notin \{0, \pi/2\}$. As a result, the overall unitary transformation $U := RV$ can transform any Hamiltonian H into the desired block structured form. It is important to note that this transformation is not unique and there exist multiple choices of unitary transformations that achieve the same structural transformation. The particular construction presented here relies on the eigenbasis matrix V, which may be challenging to implement on current quantum hardware. To ensure the practicality of our learning algorithm, we therefore restrict the definition of parallel-learnable Hamiltonians to those admitting realizable unitary transformations.

Here we present some examples of parallel-learnable Hamiltonians with their corresponding realizable unitary transformations. These examples demonstrate how certain structured Hamiltonians can be transformed into a direct sum of 2×2 non-diagonal blocks using unitary operations that are feasible to implement on current quantum hardware. For each case, we explicitly construct the unitary transformations.

a. Hadarmad-transformed Consider the Hamiltonian in the following form

$$H = \frac{1}{4} \begin{bmatrix} 2a+2b+2c+2d & 2a-2b & 2c-2d & 0 \\ 2a-2b & 2a+2b-2c-2d & 0 & -2c+2d \\ 2c-2d & 0 & 2a+2b+2c+2d & 2a-2b \\ 0 & -2c+2d & 2a-2b & 2a+2b-2c-2d \end{bmatrix}, \tag{S4}$$

it does not have clear invariant subspaces under the computational basis. However, it is still parallel-learnable since under the transformation

$$U = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}, \tag{S5}$$

it can be written in the direct sum of 2×2 non-diagonal matrix, i.e.,

$$U^\dagger H U = \begin{bmatrix} a & c & 0 & 0 \\ c & b & 0 & 0 \\ 0 & 0 & a & d \\ 0 & 0 & d & b \end{bmatrix}. \tag{S6}$$

Also, the unitary transformation U is easy to implement using just Hadarmard gates, i.e., $U = \mathrm{H} \otimes \mathrm{H}$.

b. Permutation-transformed Consider the Hamiltonian of the general XY model with local Z term for two qubits as an example, i.e.,

$$H = g_x X_1 X_2 + g_y Y_1 Y_2 + g_1 Z_1 + g_2 Z_2 = \begin{bmatrix} g_1+g_2 & 0 & 0 & g_x+g_y \\ 0 & g_1-g_2 & g_x+g_y & 0 \\ 0 & g_1-g_2 & -g_1+g_2 & 0 \\ g_x-g_y & 0 & 0 & -g_1-g_2 \end{bmatrix}, \tag{S7}$$

it cannot be written as a direct sum in the normal computational bases. However, it has clear invariant subspaces under the computational basis. Therefore, the unitary transformations to make it into a direct sum are just permutations. Under the permutation transformation,

$$U = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \tag{S8}$$

the resulting matrix will be

$$U^\dagger H U = \begin{bmatrix} g_1+g_2 & g_x-g_y & 0 & 0 \\ g_x-g_y & -g_1-g_2 & 0 & 0 \\ 0 & 0 & g_1-g_2 & g_x+g_y \\ 0 & 0 & g_x+g_y & -g_1+g_2 \end{bmatrix}. \tag{S9}$$

This U transformation can be realized with CNOT gates, i.e., $U = \mathrm{CNOT}_{1\to 0} \cdot \mathrm{CNOT}_{0\to 1}$. This Hamiltonian is commonly used to approximate the superconducting qubit Hamiltonian and interactions in Trapped Ion gates [27–29]

II. QUANTUM SIGNAL PROCESSING ESTIMATION

Parallel structure of the time-evolution operator allows the usage of the QSPE method [17] in different 2×2 invariant subspace simultaneously. In this section, we briefly introduce the idea of the QSPE method. The QSPE method is designed for estimating the angles of the gate containing a two-level invariant subspace. The unitary gate restricted to the invariant subspace is parametrized as

$$U = \begin{bmatrix} \cos(\theta) e^{-i\zeta} & -i\sin(\theta) e^{i\chi} \\ -i\sin(\theta) e^{-i\chi} & \cos(\theta) e^{i\zeta} \end{bmatrix},$$

where the θ is the swap angle, ζ is the phase difference and χ is the phase accumulated during the swap. Denote the $\{|0_l\rangle, |1_l\rangle\}$ as the logical basis of the invariant subspace, the quantum circuits utilize in QSPE methods are constructed as follows. First, the Bell state $|+_l\rangle$ and $|i_l\rangle$ is prepared. Then d repetitions of the target gate U followed by a tunable Z rotation with respect to parameter ω are applied to the circuit. Finally, the circuit is measured in computational basis for the transition probability p_x (for initial state $|+_l\rangle$) and probability p_y (for initial state $|i_l\rangle$). These steps are used to amplify the angle of θ and ζ to reach the Heisenberg limit in the following inference step.

After the quantum stage, the QSPE method focuses on a reconstructed function $h(\omega) = p_x - \frac{1}{2} + i(p_y - \frac{1}{2})$. It admits an approximated expansion, $h(\omega) = \sum_{-d+1}^{d-1} c_k e^{2ik\omega}$ and if $d\theta << 1$, the Fourier coefficients

$$c_k \approx i\theta e^{-i\chi} e^{-i(2k+1)\zeta} \quad \text{with } k = 0, \cdots, d-1. \tag{S10}$$

To characterize the information completely, sampling the reconstructed function h on $2d-1$ distinct ω points is required from the Fourier expansion. In QSPE method, the set of ω points is chosen with equal space for the

quantum stage. Moreover, two amplified angles θ and ζ are completely decoupled in Eq. (S10). As a result, the complicated inference problem becomes two independent linear regression problems with respect to the amplitude and phase of Fourier coefficients, leading to the estimators in QSPE method as follows.

$$\hat{\theta} = \frac{1}{d}\sum_{k=0}^{d-1}|c_k| \quad , \hat{\zeta} = \frac{1}{2}\frac{\vec{\mathbb{I}}^T\mathcal{D}^{-1}\vec{\Delta}}{\vec{\mathbb{I}}^T\mathcal{D}^{-1}\vec{\mathbb{I}}}; \tag{S11}$$

where $\mathbb{I}$ is an all-one vector and $\mathcal{D}$ is the $(d-1)\times(d-1)$discrete Laplacian matrix as

$$\mathcal{D} = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{bmatrix}, \tag{S12}$$

and $\mathbf{\Delta} := (\Delta_0, \cdots, \Delta_{d-2})^T$ with the sequential phase difference is $\Delta_k := \text{phase}(c_k\bar{c}_{k+1})$.

The variances of $(\hat{\theta}, \hat{\zeta})$ in QSPE are

$$\text{Var}(\hat{\theta}) \approx \frac{1}{8Nd^2} \ , \text{Var}(\zeta) \approx \frac{3}{8Nd^4\theta^2}, \tag{S13}$$

where N is the number of shots and d is the repetition depths.

III. IMPLEMENTATION DETAILS OF ALGORITHM 1

To apply the QSPE method [17] in Algorithm 1, adaptations of the original digital-gate-based QSPE protocols are required for physical implementation on quantum simulators. While the Rydberg atom quantum simulator is used as the implementation platform in this work, the proposed methods are equally applicable to other quantum simulation platforms. Since $\mathcal{U}_\mathcal{B}$ in Eq. (6) is equivalent to the standard unitary $\mathcal{U}$ in Eq. (7) corresponding to the mapping from Eq. (8), thus adapting the QSPE to the defined logical subspace provides the characterization of unknown Hamiltonian coefficients. The adaptations contain: 1) prepare the logical $|+_l\rangle$ and $|i_l\rangle$ states; 2) applying logical control Z_l term; 3) measuring transition probability with respect to the logical $|0_l\rangle$ state.

1) Adaptation 1 To prepare logical Bell state $|+_l\rangle = \frac{1}{\sqrt{2}}(|0_l\rangle + |1_l\rangle) = \frac{1}{\sqrt{2}}(|00\rangle + |10\rangle)$ and $|i_l\rangle = \frac{1}{\sqrt{2}}(|0_l\rangle + i\,|1_l\rangle) = \frac{1}{\sqrt{2}}(|00\rangle + i\,|10\rangle)$, we make use of X and Y pulse correspondingly. First, we initialize both atoms in the ground state, i.e. $|00\rangle$, then consider X and Y rotations on the first atom, that is

$$e^{-i\theta X_1\otimes I}\,|00\rangle = \cos\theta\,|00\rangle - i\sin\theta\,|10\rangle \tag{S14}$$

$$e^{-i\theta Y_1\otimes I}\,|00\rangle = \cos\theta\,|00\rangle + \sin\theta\,|10\rangle\,; \tag{S15}$$

Thus, applying $-X(\frac{\pi}{4})$ leads to $|i_l\rangle$ and applying $Y(\frac{\pi}{4})$ leads to $|0_l\rangle$, where the negative sign is implemented by tuning phase $\phi = \pi$.

2) Adaptation 2 Considering the invariant subspace $\mathcal{B}$, we want to effectively implement Z_l acting on it. There are two options to achieve that, based on the control capability of the quantum system. In the analog-digital hybrid setting, applying global detuning to two atoms gives us the desired control unitary. Consider the global detuning $H = -\frac{\Delta(t)}{2}(Z_1 + Z_2)$, the time-evolution operator is

$$\exp(-i\int_0^T H dt) = \exp(-i\int_0^T -\frac{\Delta(t)}{2}dt(Z_1 + Z_2)) \tag{S16}$$

$$= \exp(-ic(Z_1 + Z_2)) \tag{S17}$$

$$= \exp(-icZ_1)\exp(-icZ_2) \tag{S18}$$

$$= \begin{bmatrix} e^{-i2c} & & & \\ & 1 & & \\ & & 1 & \\ & & & e^{i2c} \end{bmatrix} \tag{S19}$$

$$= \begin{bmatrix} e^{-ic} & & & \\ & e^{ic} & & \\ & & e^{ic} & \\ & & & e^{i3c} \end{bmatrix} \text{ up to a global phase} \tag{S20}$$

where $c = \int_0^T \Delta(t)/2dt$. Therefore, restricting to the invariant subspace gives us the control Z_l with control parameter c.

In the fully analog setting, where we can not turn down the ZZ interaction, then the control part can be provided by having global Z interaction and atom-atom interaction together. Since Z_1,Z_2 and Z_1Z_2 commute, then we can write down the time evolution operator as:

$$\exp(-ic(Z_1+Z_2)-ibZ_1Z_2) = \begin{bmatrix} e^{-ic} & & & \\ & e^{ic} & & \\ & & e^{ic} & \\ & & & e^{i3c} \end{bmatrix} \begin{bmatrix} e^{-ib} & & & \\ & e^{ib} & & \\ & & e^{ib} & \\ & & & e^{-ib} \end{bmatrix} \tag{S21}$$

$$= \begin{bmatrix} e^{-i(b+c)} & & & \\ & e^{i(b+c)} & & \\ & & e^{i(b+c)} & \\ & & & e^{i(3c-b)} \end{bmatrix} \tag{S22}$$

where $c = \int_0^T \Delta(t)/2dt$ and $b = \int_0^T c_{12}dt$. Therefore, restricting to the invariant subspace gives us the control Z_l with control parameter $b+c$.

3) Adaptation 3 Measurements can also be done in the computational basis. We estimate the probability of getting $|0_l\rangle = |00\rangle$ by $\Pr(00) = N_{00}/N$, where N_{00} is the number of string 00 occurs and N is the total shots.

Upon these three adaptations, the learning algorithm 1 can be implemented on the current Rydberg quantum device without any additional functionality required.

IV. ROBUSTNESS AGAINST COHERENT STATE PREPARATION ERROR

Recall that the coherent error model considered here is the miscalibration in the state preparation step (discussed in the previous section in Adaptation 1. More specifically,

$$e^{-i(\theta+\alpha)X_1\otimes I}|00\rangle = \cos(\theta+\alpha)|00\rangle - i\sin(\theta+\alpha)|10\rangle$$
$$e^{-i(\theta+\alpha)Y_1\otimes I}|00\rangle = \cos(\theta+\alpha)|00\rangle + \sin(\theta+\alpha)|10\rangle;$$

where α denotes the error rate. Inspired by the trick used for depolarizing noise, we construct a rescaled estimator reduce the error from first order to second order.

Theorem (Restatement of Theorem III.1)**.** *Consider the coherent error Eq. (51) with error rate α in the state preparation step of Algorithm 1, the difference between the noisy estimation $\hat{\theta}' = \frac{1}{d}\sum_{k=0}^{d-1}|c'_k|$ and noiseless estimation $\hat{\theta} = \frac{1}{d}\sum_{k=0}^{d-1}|c_k|$ can be bounded by*

$$D = |\hat{\theta}' - \hat{\theta}| \le \sqrt{2}(d+1)^2 \sin 2\alpha \sin^2\theta + (\cos 2\alpha - 1)\theta, \tag{S23}$$

which scales with $O(\theta)$. The difference can be improved by introducing new type of estimator,i.e. scaled estimator $\hat{\theta}'_s = \frac{1}{\cos 2\alpha} \times \frac{1}{d}\sum_{k=0}^{d-1}|c'_k|$. The corresponding bound then reduces to

$$D_s = |\hat{\theta_s}' - \hat{\theta}| \le \sqrt{2}(d+1)^2 \tan 2\alpha \sin^2\theta, \tag{S24}$$

where scales only with $O(\theta^2)$.

Proof. Consider the coherent error Eq. (51),the corresponding effective initial states $|+_l\rangle'$ and $|i_l\rangle'$ are in the forms of

$$|+_l\rangle' = \begin{bmatrix} \cos(\theta+\alpha) \\ 0 \\ \sin(\theta+\alpha) \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \cos\alpha - \sin\alpha \\ 0 \\ \cos\alpha + \sin\alpha \\ 0 \end{bmatrix}, \tag{S25}$$

$$|i_l\rangle' = \begin{bmatrix} \cos(\theta+\alpha) \\ 0 \\ -i\sin(\theta+\alpha) \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \cos\alpha + \sin\alpha \\ 0 \\ i(\cos\alpha - \sin\alpha) \\ 0 \end{bmatrix}. \tag{S26}$$

(S27)

Recall the repeated circuits used in Algorithm 1 satisfy the form of

$$U^{(d)} = \begin{bmatrix} P^{(d)}(\cos\theta) & i\sin\theta Q^{(d)}(\cos\theta) \\ i\sin(\theta)Q^{(d)}(\cos\theta) & P^{(d)*}(\cos\theta)), \end{bmatrix} \tag{S28}$$

therefore the corresponding probability p'_X and p'_Y will be

$$p'_X = |e^{i\frac{\chi+\pi+\varphi}{2}Z} \langle 0_l| U^{(d)}(\omega-\varphi,\theta)e^{-i(\omega+\frac{\chi+\pi+\varphi}{2})Z} |+_l\rangle' |^2 \tag{S29}$$

$$= |\frac{1}{\sqrt{2}}e^{i\frac{\chi+\pi+\varphi}{2}Z} \langle 0_l| U^{(d)}(\omega-\varphi,\theta)(e^{-i(\omega+\frac{\chi+\pi+\varphi}{2})}(\cos\alpha-\sin\alpha)|0_l\rangle + e^{i(\omega+\frac{\chi+\pi+\varphi}{2})}(\cos\alpha+\sin\alpha)|1_l\rangle)|^2 \tag{S30}$$

$$= \frac{1}{2}|(e^{-i(\omega+\frac{\chi+\pi+\varphi}{2})}(\cos\alpha-\sin\alpha)P^{(d)}(\cos\theta) + i\sin\theta e^{i(\omega+\frac{\chi+\pi+\varphi}{2})}(\cos\alpha+\sin\alpha)Q^{(d)}(\cos\theta))|^2 \tag{S31}$$

$$= \frac{1}{2}\left((\cos\alpha-\sin\alpha)^2PP^* + \sin\theta^2(\cos\alpha+\sin\alpha)^2Q^2\right)$$

$$- \frac{1}{2}\left(i\sin\theta e^{-i2(\omega+\frac{\chi+\pi+\varphi}{2})}(\cos\alpha^2-\sin\alpha^2)QP + i\sin\theta e^{i2(\omega+\frac{\chi+\pi+\varphi}{2})}(\cos\alpha^2-\sin\alpha^2)P^*Q\right) \tag{S32}$$

$$= \frac{1}{2}\left((\cos\alpha-\sin\alpha)^2PP^* + \sin\theta^2(\cos\alpha+\sin\alpha)^2Q^2 + 2(\cos\alpha^2-\sin\alpha^2)\mathrm{Re}(i\sin\theta e^{i(-2\omega-\chi+\varphi)}QP)\right) \tag{S33}$$

Then, we notice that

$$P^*P = \cos^2((d+1)\sigma) + \frac{\sin^2((d+1)\sigma)}{\sin^2\sigma}\sin^2\omega\cos^2\theta \tag{S34}$$

$$\sin^2\theta Q^2 = \frac{\sin^2((d+1)\sigma)}{\sin^2\sigma}\sin^2\theta \tag{S35}$$

$$P^*P + \sin^2\theta Q^2 = 1 \tag{S36}$$

$$\tag{S37}$$

where $\cos\sigma = \cos\omega\cos\theta$. Expanding the first term in the probability expression,

$$(\cos\alpha-\sin\alpha)^2PP^* + \sin\theta^2(\cos\alpha+\sin\alpha)^2Q^2 = P^*P + \sin^2\theta Q^2 - 2\cos\alpha\sin\alpha(P^*P - \sin^2\theta Q^2) \tag{S38}$$

$$= 1 - 2\cos\alpha\sin\alpha(P^*P - \sin^2\theta Q^2). \tag{S39}$$

Also, the minus term can be computed as

$$P^*P - \sin^2\theta Q^2 = \cos^2((d+1)\sigma) + \frac{\sin^2((d+1)\sigma)}{\sin^2\sigma}(\sin^2\omega\cos^2\theta - \sin^2\theta) \tag{S40}$$

$$= 1 - \sin^2((d+1)\sigma)\frac{2\sin^2\theta}{\sin^2\sigma} \tag{S41}$$

$$= 1 - C_{\omega,d,\theta} \tag{S42}$$

$$\tag{S43}$$

where we denote $C_{\omega,d,\theta} := \frac{\sin^2((d+1)\omega)}{\sin^2\omega}$.

Thus, combing all above, the effective probability p'_X is

$$p'_X = \frac{1}{2}[1 - 2\cos\alpha\sin\alpha(1-C_{\omega,d,\theta})] + (\cos\alpha^2-\sin\alpha^2)\mathrm{Re}(i\sin\theta e^{i(-2\omega-\chi+\varphi)}QP) \tag{S44}$$

$$= \cos 2\alpha p_X + [\frac{1-\cos 2\alpha - \sin 2\alpha(1-C_{\omega,d,\theta})}{2}], \tag{S45}$$

where p_X is the probability without noise.

Similarly, for the initial state $|i_l\rangle'$, we end up with

$$p'_Y = \frac{1}{2}\left((\cos\alpha+\sin\alpha)^2PP^* + \sin\theta^2(\cos\alpha-\sin\alpha)^2Q^2 + 2(\cos\alpha^2-\sin\alpha^2)\mathrm{Im}(i\sin\theta e^{i(-2\omega-\chi+\varphi)}QP)\right) \tag{S46}$$

$$= \frac{1}{2}[1 + 2\cos\alpha\sin\alpha(1-C_{\omega,d,\theta})] + (\cos\alpha^2-\sin\alpha^2)\mathrm{Im}(i\sin\theta e^{i(-2\omega-\chi+\varphi)}QP) \tag{S47}$$

$$= \cos 2\alpha p_Y + ([\frac{1-\cos 2\alpha + \sin 2\alpha(1-C_{\omega,d,\theta})}{2}]. \tag{S48}$$

Notice here that the effective probability is not simply constant shifted and scaled, so the trick for depolarizing channel does not work here.

Since $h = p_X - \frac{1}{2} + i(p_Y - \frac{1}{2})$, the effective h' is

$$h' = p'_X - \frac{1}{2} + i(p'_Y - \frac{1}{2}) \tag{S49}$$

$$= \cos 2\alpha h - \frac{\cos 2\alpha + \sin 2\alpha(1 - C_{\omega,d,\theta})}{2} - \frac{\cos 2\alpha - \sin 2\alpha(1 - C_{\omega,d,\theta})}{2} i. \tag{S50}$$

Recall that the estimator in the QSPE algorithm 1 is of the form of:

$$\hat{\theta} = \frac{1}{d} \sum_{k=0}^{d-1} |c_k|, \tag{S51}$$

where the Fourier transform gives the desired coefficient for inferring our target parameters, i.e.

$$c_k = \frac{1}{\pi} \int_0^\pi e^{-i(2k\omega)} h d\omega. \tag{S52}$$

Original estimator Coherent error in state preparation will induce errors in probability p_X and p_Y, resulting in errors in h. Thus, using the same estimator given in Eq. (S51), we can measure the performance of this estimator as follows.

First, we do the Fourier transform on h' to get coefficients c'_k for the original estimator.

$$c'_k = \frac{1}{\pi} \int_0^\pi e^{-i(2k\omega)} h' d\omega \tag{S53}$$

$$= \frac{1}{\pi} \int_0^\pi e^{-i(2k\omega)} [\cos 2\alpha h - \frac{\cos 2\alpha + \sin 2\alpha(1 - C_{\omega,d,\theta})}{2} - \frac{\cos 2\alpha - \sin 2\alpha(1 - C_{\omega,d,\theta})}{2} i] d\omega \tag{S54}$$

$$= \cos 2\alpha c_k - \frac{1}{\pi} \int_0^\pi e^{-i(2k\omega)} [\frac{\cos 2\alpha + \sin 2\alpha(1 - C_{\omega,d,\theta})}{2} - \frac{\cos 2\alpha - \sin 2\alpha(1 - C_{\omega,d,\theta})}{2} i] d\omega \tag{S55}$$

$$= \cos 2\alpha c_k - \frac{1}{\pi} \int_0^\pi e^{-i(2k\omega)} [\frac{\cos 2\alpha + \sin 2\alpha}{2} + \frac{\cos 2\alpha - \sin 2\alpha}{2} i] - e^{-i(2k\omega)} [\frac{\sin 2\alpha C_{\omega,d,\theta}}{2} + \frac{\sin 2\alpha C_{\omega,d,\theta}}{2} i] d\omega \tag{S56}$$

$$= \cos 2\alpha c_k + \frac{\sin 2\alpha}{2\pi} \int_0^\pi e^{-i(2k\omega)} (C_{\omega,d,\theta} + C_{\omega,d,\theta} i) d\omega \tag{S57}$$

where $C_{\omega,d,\theta} = 2\sin^2\theta \frac{\sin^2[(d+1)\sigma]}{\sin^2\sigma} \leq 2\sin^2\theta(d+1)^2$ and $\cos\sigma = \cos\omega\cos\theta$.

Then the amplitude of the coefficients can be bounded as follows:

$$|c'_k| \leq \cos 2\alpha |c_k| + \frac{\sin 2\alpha}{2\pi} \int_0^\pi |e^{-i(2k\omega)} (C_{\omega,d,\theta} + C_{\omega,d,\theta} i)| d\omega \tag{S58}$$

$$= \cos 2\alpha |c_k| + \frac{\sin 2\alpha}{2\pi} \int_0^\pi \sqrt{2} |(C_{\omega,d,\theta})| d\omega \tag{S59}$$

$$\leq \cos 2\alpha |c_k| + \frac{\sin 2\alpha}{\sqrt{2}} \max_{[0,\pi]} |(C_{\omega,d,\theta})| \tag{S60}$$

$$\leq \cos 2\alpha |c_k| + \sqrt{2} \sin 2\alpha \sin^2\theta (d+1)^2 \tag{S61}$$

$$\tag{S62}$$

Therefore, the difference between the original estimator under coherent state preparation error $\hat{\theta}' = \frac{1}{d} \sum_{k=0}^{d-1} |c'_k|$

and the noiseless estimator $\hat{\theta} = \frac{1}{d}\sum_{k=0}^{d-1}|c_k|$ will be:

$$|\hat{\theta}' - \hat{\theta}| = \frac{1}{d}\sum_{k=0}^{d-1}|c'_k| - |c_k| \tag{S63}$$

$$\leq \frac{1}{d}\sum_{k=0}^{d-1}|c'_k - c_k| \tag{S64}$$

$$= \frac{1}{d}\sum_{k=0}^{d-1}|c'_k - \cos 2\alpha c_k + \cos 2\alpha c_k - c_k| \tag{S65}$$

$$\leq \frac{1}{d}\sum_{k=0}^{d-1}|c'_k - \cos 2\alpha c_k| + |\cos 2\alpha c_k - c_k| \tag{S66}$$

$$\leq \frac{1}{d}\sum_{k=0}^{d-1}\sqrt{2}(d+1)^2 \sin 2\alpha \sin^2\theta + \frac{1}{d}\sum_{k=0}^{d-1}|\cos 2\alpha c_k - c_k| \tag{S67}$$

$$= \sqrt{2}(d+1)^2 \sin 2\alpha \sin^2\theta + (\cos 2\alpha - 1)\theta \tag{S68}$$

Scaled Estimator Inspired by the above analysis, we introduce a new estimator by scaling the original one with $1/\cos 2\alpha$, i.e. $\hat{\theta}' = \frac{1}{\cos 2\alpha} \times \frac{1}{d}\sum_{k=0}^{d-1}|c'_k|$.

Then the corresponding bias will be:

$$|\hat{\theta}' - \hat{\theta}| = \frac{1}{d}\sum_{k=0}^{d-1}|\frac{1}{\cos 2\alpha}c'_k| - |c_k| \tag{S69}$$

$$\leq \frac{1}{d}\sum_{k=0}^{d-1}|\frac{1}{\cos 2\alpha}c'_k - c_k| \tag{S70}$$

$$= \frac{1}{d}\sum_{k=0}^{d-1}\sqrt{2}\sin 2\alpha \sin^2\theta(d+1)^2/\cos 2\alpha \tag{S71}$$

$$= \sqrt{2}(d+1)^2 \tan 2\alpha \sin^2\theta. \tag{S72}$$

□

V. VARIANCE COMPUTATION FOR ALGORITHM 2

For each full Hamiltonian, we simultaneously work on $(n-1)$ invariant subspaces as defined in Eq. (II.10). As a result, we should prepare the initial logical Bell states with respect to each invariant subspace and superpose them. Due to the normalization factor, the final state restricted to each invariant subspace gains a $1/\sqrt{(n-1)}$ factor compared to the standard Bell pair. Without loss of generality, let us focus on the full Hamiltonian H_1 and the kth invariant subspace of the Hamiltonian. Denote the transition probability with initial plus state as $p_k^M(\omega_j;\theta,\phi,\chi)$, the relation between the transition probability with superposition of the Bell pairs and the standard Bell pair is

$$p_k^M(\omega_j;\theta,\phi,\chi) = p^B(\omega_j;\theta,\phi,\chi)/(n-1). \tag{S73}$$

Consider the sampling error in the system. Let N be the number of measurements in one experiment with the set of parameters $(\omega_j;\theta,\phi,\chi)$, the i.i.d. measurement outcomes are sampled from a Multinomial distribution such that

$$\sum_{i=1}^{(n-1)} p_i^M(b_i) + p_i^M(1-b_i) = 1, \tag{S74}$$

where b_i is the bitstring corresponding to the basis that spans the ith invariant subspace and 1 is the all-one bitstring.

Given N large enough samples, the transition probability corresponding to each invariant subspace is estimated by

$$\hat{p}_k^M(\omega_j) = \frac{\#(x = b_k)}{N}, \tag{S75}$$

where b_k is the logical zero state for the kth invariant subspace. Focusing on this specific transition probability for the kth invariant subspace, the Multinomial distribution can be treated as a Bernoulli distribution with probability p_k^M. As a result, when the sample size N is large enough, the central limit theorem leads to the normal distribution with mean $\mathbb{E}[\hat{p}_k^M(\omega_j)] = p_k^M$ and variance follows

$$\Sigma_k^2 = \frac{p_k^M(1-p_k^M)}{N} = \frac{1}{4N} - \frac{(p_k^M - \frac{1}{2})^2}{N}. \tag{S76}$$

Considering the standard Bell pair as the initial state for this specific invariant subspace, the transition probability is $p^B := (n-1)p_k^M$. Let the sample size still be N, the MLE for the transition probability will be $\hat{p}^B = \frac{\#(x=0)}{N}$ with expectation $\mathbb{E}(\hat{p}^B) = p^B$ and variance $\mathrm{Var}(\hat{p}^B) = \frac{p^B(1-p^B)}{N}$. From [17], we notice the variance is bounded by $[\frac{1}{4N} - \frac{(d\theta)^2}{M}, \frac{1}{4N}]$. For the inference process, we utilize the sample data from the Multinomial senario to estimate the starndard transition probability following Eq. (S73), the estimator is defined as

$$\hat{p}_{k,res}^M = (n-1)\hat{p}_k^M(\omega_j) = \frac{(n-1)\#(x = b_k)}{N}, \tag{S77}$$

where $\hat{p}_{k,res}^M$ denotes the rescaled probability. Then the expectation and variance value of $\hat{p}_{k,res}^M$ can be computed as follows.

$$\mathbb{E}(\hat{p}_{k,res}^M) = (n-1)\mathbb{E}[\hat{p}_k^M] = (n-1)p_k^M = p^B; \tag{S78}$$

$$\mathrm{Var}(\hat{p}_{k,res}^M) = (n-1)^2\mathrm{Var}(\hat{p}_k^M) = (n-1)^2\frac{p_k^M(1-p_k^M)}{N} = \frac{p^B(n-1-p^B)}{N}, \tag{S79}$$

where the last inequality follows Eq. (S73). Rewrite the Eq. (S79), we derive the upper and lower bounds of it as follows.

$$\begin{aligned}
\mathrm{Var}(\hat{p}_{k,res}^M) &= \frac{p^B(n-1-p^B)}{N} && \text{(S80)}\\
&= -\frac{(p^B - \frac{1}{2})^2}{N} - \frac{p^B}{N} + \frac{1}{4N} + \frac{(n-1)p^B}{N} && \text{(S81)}\\
&= -\frac{(p^B - \frac{1}{2})^2}{N} + (n-2)\frac{(p^B - \frac{1}{2})}{N} + \frac{(2n-3)}{4N}. && \text{(S82)}\\
& && \text{(S83)}
\end{aligned}$$

Given that $(p^B - \frac{1}{2})^2 \leq (d\theta)^2$, we can bound the variance as

$$\frac{(2n-3)}{4N} - [\frac{(d\theta)^2 + (n-2)(d\theta)}{N}] \leq \mathrm{Var}(\hat{p}_{k,res}^M) \leq \frac{n-2}{N}, \tag{S84}$$

where the left inequality matches when $(p^B - \frac{1}{2}) = -d\theta$ and right inequality matches when $(p^B - \frac{1}{2}) = \frac{1}{2}$.

Then, we compute the Fourier coefficients of the transition probability $\{\hat{p}_{k,res}^M(\omega_j)\}$ from the experimental data. The FFT is performed on the reconstructed function for the kth invariant subspace, i.e. $\hat{h}_j := \hat{p}_{X,res}^M + i\hat{p}_{Y,res}^M - \frac{1+i}{2}$. The expectation value of this reconstructed function follows $\mathbb{E}[\hat{h}_j] = p_X^B + ip_Y^B - \frac{1+i}{2}$ admitting the analytical form. We compute the covariance matrix of the reconstructed function as follows. Let $u_j = \hat{h}_j - \mathbb{E}[\hat{h}_j]$, it holds that

$$\begin{aligned}
&\text{for any } j\ , \mathbb{E}(u_j) = 0, \mathbb{E}(|u_j|^2) = \mathrm{Var}(\hat{p}_{X,res}^M) + \mathrm{Var}(\hat{p}_{Y,res}^M) = \frac{p_X^B(n-1-p_X^B)}{N} + \frac{p_Y^B(n-1-p_Y^B)}{N} && \text{(S85)}\\
&\text{for any } j \neq j', \mathbb{E}(u_j u_{j'}) = 0. && \text{(S86)}
\end{aligned}$$

By doing FFT, the coefficients extracted are

$$\begin{bmatrix} \hat{c}_0 \\ \vdots \\ \hat{c}_{d-1} \\ \hat{c}_{-d+1} \\ \vdots \\ \hat{c}_{-1} \end{bmatrix} = \frac{1}{2d-1}\Omega^\dagger \begin{bmatrix} \hat{h}_0 \\ \vdots \\ \hat{h}_{2d-2} \end{bmatrix}, \text{ where } \Omega_{jk} = e^{i\frac{2\pi jk}{2d-1}}. \tag{S87}$$

Assume the Fourier coefficients are approximately distributed following the complex normal distribution as

$$\hat{c}_k = c_k + \begin{cases} v_k, & k = 0, \cdots, d-1; \\ v_{2d-1+k}, & k = -d+1, \cdots, -1, \end{cases} \tag{S88}$$

where v_k's are complex normally distributed random variables. Next, we compute the mean and covariance matrix of the v_k's, which offers insight into the error introduced by finite sampling directly. Since for all rescaled transition probabilities, they are be written as

$$\hat{p}^M_{X/Y,res}(\omega_j) = p^B_{X/Y} + st(\hat{p}^M_{X/Y,res})\epsilon_i, \tag{S89}$$

where $\epsilon_i \sim N(0,1)$ and $\frac{(2n-3)}{4N} - [\frac{(d\theta)^2+(n-2)(d\theta)}{N}] \leq \mathrm{Var}(\hat{p}^M_{k,res}) \leq \frac{n-2}{N}$. Then the reconstructed function can also be interpreted as

$$\hat{h}_j := \hat{p}^M_{X,res} + i\hat{p}^M_{Y,res} - \frac{1+i}{2} = p^B_X + ip^B_Y - \frac{1+i}{2} + st(\hat{p}^M_{X/Y,res})\epsilon_i. \tag{S90}$$

By linearity,

$$v_k = \frac{1}{2d-1}(\Omega^\dagger u) = \frac{1}{2d-1}\sum_{j=0}^{2d-2} \overline{\Omega_{kj}} u_j, \tag{S91}$$

where $u = [u_0 \cdots u_{2d-2}]^T$, the vector contains all $u_j = \hat{h}_j - \mathbb{E}[\hat{h}_j]$.

The expectation value is $\mathbb{E}(v_k) = \frac{1}{2d-1}\sum_{j=0}^{2d-2} \overline{\Omega_{kj}}\mathbb{E}(u_j) = 0$, and the covariance is

$$\mathbb{E}(v_k\overline{v}_{k'}) = \frac{1}{(2d-1)^2}\sum_{j,j'=0}^{2d-2} \overline{\Omega_{kj}}\Omega_{k'j'}\mathbb{E}(u_j\overline{u_{j'}}) = \frac{1}{(2d-1)^2}\sum_{j,j'=0}^{2d-2} e^{i\frac{2\pi}{2d-1}(k'-k)j}\mathbb{E}(|u_j|^2). \tag{S92}$$

When $k = k'$, the variance terms are

$$\mathbb{E}(|v_k|^2) = \frac{1}{(2d-1)^2}\sum_{j=0}^{2d-2} \mathbb{E}(|u_j|^2) \tag{S93}$$

$$= \frac{1}{(2d-1)^2}\sum_{j=0}^{2d-2} -\frac{(p^B_X - \frac{1}{2})^2 + (p^B_Y - \frac{1}{2})^2}{N} + (n-2)\frac{(p^B_X - \frac{1}{2}) + (p^B_Y - \frac{1}{2})}{N} + \frac{(2n-3)}{2N} \tag{S94}$$

$$= \frac{(2n-3)}{2N(2d-1)} + \frac{1}{(2d-1)^2}\sum_{j=0}^{2d-2} -\frac{(p^B_X - \frac{1}{2})^2 + (p^B_Y - \frac{1}{2})^2}{N} + (n-2)\frac{(p^B_X - \frac{1}{2}) + (p^B_Y - \frac{1}{2})}{N}. \tag{S95}$$

It can be bounded as

$$\frac{(2n-3) - 4(d\theta)^2 - 4(n-2)(d\theta)}{2N(2d-1)} \leq \mathbb{E}(|v_k|^2) \leq \frac{(2n-3) - 4(d\theta)^2 + 4(n-2)(d\theta)}{2N(2d-1)}, \tag{S96}$$

where the conditions $(p^B_{X/Y} - \frac{1}{2})^2 \leq (d\theta)^2$ are used.

When $k \neq k'$, since $\sum_{j=0}^{2d-2} e^{i\frac{2\pi}{2d-1}(k'-k)j} = (2d-1)\delta_{kk'}$, the constant term $\frac{2n-3}{2N}$ vanishes. Thus,

$$|\mathbb{E}(v_k\overline{v}_{k'})| = \left|\frac{1}{(2d-1)^2}\sum_{j=0}^{2d-2} e^{i\frac{2\pi}{2d-1}(k'-k)j}\left(\frac{-[(p_X^B-\frac{1}{2})^2+(p_Y^B-\frac{1}{2})^2]}{N}+(n-2)\frac{(p_X^B-\frac{1}{2})+(p_Y^B-\frac{1}{2})}{N}\right)\right| \tag{S97}$$

$$\leq \frac{1}{N(2d-1)^2}\sum_{j=0}^{2d-2}\left|\left(\frac{-[(p_X^B-\frac{1}{2})^2+(p_Y^B-\frac{1}{2})^2]}{N}+(n-2)\frac{(p_X^B-\frac{1}{2})+(p_Y^B-\frac{1}{2})}{N}\right)\right| \tag{S98}$$

$$\leq \frac{2(d\theta)^2+2(n-2)(d\theta)}{N(2d-1)}. \tag{S99}$$

Consider the signal-to-noise ratio(SNR) for each Fourier coefficients

$$\mathrm{SNR}_k := \frac{|c_k|^2}{\mathbb{E}(|v_k|^2)} \geq N(2d-1)\sin^2\theta(1-\frac{4}{3}(d\theta)^2(1+3d^3\theta^2))/(2n-4), \tag{S100}$$

this inequality comes from Theorem 12 [17] and Eq. (S96).

When SNR is high, that is, $dM\theta^2/n >> 1$, the decoupling of the θ and ζ dependence is well captured and for all $k = 0, \cdots, d-1$, $d^5\theta^4 << 1$, the estimators are defined as

$$\text{amplitude}(\hat{c}_k) \approx \theta + v_k^{(amp)}; \tag{S101}$$

$$\text{phase}(\hat{c}_k) = \frac{\pi}{2} - (2k+1)\zeta + v_k^{(pha)} \quad (\text{up to } 2\pi - \text{periodicity}), \tag{S102}$$

where $v_k^{(amp)}$ and $v_k^{(pha)}$ are normal distributed and can be approximately computed by $v_k^{(amp)} = \mathrm{Re}(v_k)$, $v_k^{(pha)} = \mathrm{Im}(v_k)/c_k(\theta)$ [17]. Then we can write the covariance matrices as $\mathrm{Cov}_{kk'}^{(amp)} := \mathbb{E}(v_k^{(amp)}v_{k'}^{(amp)})$ and $\mathrm{Cov}_{kk'}^{(pha)} := \mathbb{E}(v_k^{(pha)}v_{k'}^{(pha)})$.

The covariance matrices can be computed from the real and imaginary components of the sampling error v_k in Fourier coefficients. For any $k \neq k'$,

$$|\mathbb{E}(\mathrm{Re}(v_k)\mathrm{Re}(v_{k'}) + \mathrm{Im}(v_k)\mathrm{Im}(v_{k'}))| = |\mathbb{E}(v_k\overline{v}_{k'})| \leq \frac{2(d\theta)^2+2(n-2)(d\theta)}{N(2d-1)}. \tag{S103}$$

For any k,

$$\frac{(2n-3)-4(d\theta)^2-4(n-2)(d\theta)}{2N(2d-1)} \leq \mathbb{E}(\mathrm{Re}^2(v_k)+\mathrm{Im}^2(v_k)) = \mathbb{E}(|v_k|^2) \tag{S104}$$

$$\leq \frac{(2n-3)-4(d\theta)^2+4(n-2)(d\theta)}{2N(2d-1)}. \tag{S105}$$

Then for any k,

$$|\mathbb{E}(\mathrm{Re}^2(v_k)-\mathrm{Im}^2(v_k))| \leq |\mathbb{E}(\mathrm{Re}^2(v_k)-\mathrm{Im}^2(v_k)+2i\mathrm{Re}(v_k)\mathrm{Im}(v_k))| \tag{S106}$$

$$= |\mathbb{E}(v_k^2)| \leq \frac{1}{(2d-1)^2}\sum_{j=0}^{2d-2}\mathbb{E}(|u_j^2|) \tag{S107}$$

$$= \frac{1}{(2d-1)^2}\sum_{j=0}^{2d-2}|\mathrm{Var}(\hat{p}_{X,res}^M) - \mathrm{Var}(\hat{p}_{Y,res}^M)| \leq \frac{2(d\theta)^2+2(n-2)(d\theta)}{N(2d-1)}. \tag{S108}$$

By the triangle inequality,

$$|\mathbb{E}(\mathrm{Re}^2(v_k)) - \frac{(2n-3)-4(d\theta)^2+4(n-2)(d\theta)}{4N(2d-1)}| \leq \frac{1}{2}|\mathbb{E}(\mathrm{Re}^2(v_k)+\mathrm{Im}^2(v_k))$$
$$-\frac{(2n-3)-4(d\theta)^2+4(n-2)(d\theta)}{4N(2d-1)}| + \frac{1}{2}|\mathbb{E}(\mathrm{Re}^2(v_k)-\mathrm{Im}^2(v_k))| \tag{S109}$$

$$\leq \frac{2(d\theta)^2+2(n-2)(d\theta)}{N(2d-1)}. \tag{S110}$$

Similarly, for the imaginary part, we also have the above bound, i.e.,

$$|\mathbb{E}(\mathrm{Im}^2(v_k)) - \frac{(2n-3)-4(d\theta)^2+4(n-2)(d\theta)}{4N(2d-1)}| \leq \frac{2(d\theta)^2+2(n-2)(d\theta)}{N(2d-1)}. \tag{S111}$$

Also, for any $k \neq k'$,

$$\mathbb{E}(\mathrm{Im}(v_k)\mathrm{Im}(v_{k'})) \leq \frac{1}{2}|\mathbb{E}(\mathrm{Re}(v_k)\mathrm{Re}(v_{k'}) + \mathrm{Im}(v_k)\mathrm{Im}(v_{k'}))| + \frac{1}{2}|\mathbb{E}(-\mathrm{Re}(v_k)\mathrm{Re}(v_{k'}) + \mathrm{Im}(v_k)\mathrm{Im}(v_{k'}))| \tag{S112}$$

$$\leq \frac{2(d\theta)^2+2(n-2)(d\theta)}{N(2d-1)}. \tag{S113}$$

From [17] Eq. (82), when $d^3\theta^2 << 1$ and $d\theta \leq \frac{1}{5}$, the following inequality holds,

$$|\frac{\sin\theta}{c_k\theta}| \leq \frac{1}{1-\frac{8}{3}(d\theta)^2} < \sqrt{\frac{4}{3}}. \tag{S114}$$

So for any $k \neq k'$,

$$|\mathbb{E}(v_k^{(pha)}v_{k'}^{(pha)})| = \frac{1}{|c_k(\theta)c_{k'}(\theta)|}|\mathbb{E}(\mathrm{Im}(v_k)\mathrm{Im}(v_{k'}))| \leq \frac{4[2(d\theta)^2+2(n-2)(d\theta)]}{3N(2d-1)(\sin\theta)^2}. \tag{S115}$$

Also,

$$|\mathbb{E}((v_k^{(pha)})^2) - \frac{(2n-3)-4(d\theta)^2+4(n-2)(d\theta)}{4N(2d-1)\sin^2\theta}| \leq \frac{1}{c_k^2(\theta)}|\mathbb{E}(\mathrm{Im}^2(v_k)) - \frac{(2n-3)-4(d\theta)^2+4(n-2)(d\theta)}{4N(2d-1)}| \tag{S116}$$

$$+ \frac{(2n-3)-4(d\theta)^2+4(n-2)(d\theta)}{4N(2d-1)\sin^2\theta}\frac{|c_k(\theta)^2-\sin^2\theta|}{c_k^2(\theta)} \tag{S117}$$

$$\leq \frac{(2n-3)}{12N(2d-1)\theta^2} + \frac{7d^2}{3(2d-1)N} + \frac{3(n-2)d}{N(2d-1)\theta}. \tag{S118}$$

Finally, we conclude the variance for the estimator

$$\hat{\theta} = \frac{1}{d}\sum_{k=0}^{d-1}|c_k| \quad , \hat{\zeta} = \frac{1}{2}\frac{\vec{\mathbb{I}}^T\mathcal{D}^{-1}\vec{\Delta}}{\vec{\mathbb{I}}^T\mathcal{D}^{-1}\vec{\mathbb{I}}}; \tag{S119}$$

where $\mathcal{D}$ is given in the Eq. (S12).

The variances are

$$\mathrm{Var}(\hat{\theta}) \approx \frac{2n-3}{4Nd(2d-1)} \approx \frac{n}{4Nd^2} \quad \text{and} \quad \mathrm{Var}(\hat{\zeta}) \approx \frac{3(2n-3)}{4Nd(2d-1)(d^2-1)\theta^2} \approx \frac{3n}{4Nd^4\theta^2}. \tag{S120}$$

VI. CRAMER-RAO BOUND FOR ALGORITHM 2

In this setting, the estimation is performed using d repetitions with N samples per circuit. Altogether, the learning procedure involves $2(2d-1)$ circuits, each with depth $\Theta(d)$. We compute the optimal variance bounded by Cramer-Rao lower bound for pre-asymptotic regime, when $d\theta << 1$, following the same analysis from [17].

For the parallel learning algorithm, we simultaneously work on $(n-1)$ invariant subspaces and infer the (θ, ζ) pair for each subspace independently. In the following discussion, we focus on the kth invariant subspace which can be parametrized with the same parameter set introduced in [17], i.e. $\Xi = (\varphi_k) = (\theta, \zeta, \chi)$ with $\chi = 0$ in our scenario. In the large sample limit, where the experimental estimated probability can be approximated by a normal

distribution with variance bounded as in Eq. (S84), then the Fisher information matrix for this specific subspace is

$$I_{kk'}(\Xi) = \sum_{j=0}^{2d-2} \mathrm{Var}^{-1}(p^M_{k,res,X}) \frac{\partial p^B_X(\omega_j,\Xi)}{\partial \varphi_k} \frac{\partial p^B_X(\omega_j,\Xi)}{\partial \varphi_{k'}} + \sum_{j=0}^{2d-2} \mathrm{Var}^{-1}(p^M_{k,res,Y}) \frac{\partial p^B_Y(\omega_j,\Xi)}{\partial \varphi_k} \frac{\partial p^B_X(\omega_j,\Xi)}{\partial \varphi_{k'}} \tag{S121}$$

$$\approx \left(\frac{4N}{(2n-3)+4(d\theta)^2+4(n-2)(d\theta))}\right) \sum_{j=0}^{2d-2} \frac{\partial p^B_X(\omega_j,\Xi)}{\partial \varphi_k} \frac{\partial p^B_X(\omega_j,\Xi)}{\partial \varphi_{k'}} + \sum_{j=0}^{2d-2} \frac{\partial p^B_Y(\omega_j,\Xi)}{\partial \varphi_k} \frac{\partial p^B_X(\omega_j,\Xi)}{\partial \varphi_{k'}} \tag{S122}$$

$$\approx \left(\frac{4N(2d-1)}{(2n-3)+4(d\theta)^2+4(n-2)(d\theta))}\right) \mathrm{Re}\left(\sum_{-d+1}^{d-1} \frac{\partial c_j(\Xi)}{\partial \varphi_k} \frac{\partial c_j(\Xi)}{\partial \varphi_{k'}} \right). \tag{S123}$$

$$\tag{S124}$$

Following the relation $c_j(\Xi) \approx ie^{-i\chi} e^{-i(2j+1)\zeta} \theta \mathbb{I}_{j>0}$ when $d\theta << 1$, we can compute the fisher information matrix as

$$I(\Xi) \approx \left(\frac{4N(2d-1)}{(2n-3)+4(d\theta)^2+4(n-2)(d\theta))}\right) \begin{bmatrix} d & 0 & 0 \\ 0 & \frac{d(4d^2-1)\theta^2}{3} & d^2\theta^2 \\ 0 & d^2\theta^2 & d\theta^2 \end{bmatrix}. \tag{S125}$$

The covariance matrix of the estimator set (Ξ) is bounded as follows according to Cramer-Rao lower bound,

$$\mathrm{Cov}(\hat{\theta},\hat{\zeta},\hat{\chi}) \geq I^{-1}(\Xi) \approx \frac{(2n-3)}{4Nd(2d-1)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{(d^2-1)\theta^2} & \frac{-3d}{(d^2-1)\theta^2} \\ 0 & \frac{-3d}{(d^2-1)\theta^2} & \frac{4d^2-1}{(d^2-1)\theta^2} \end{bmatrix}. \tag{S126}$$

As a result, the optimal variances for $(\hat{\theta},\hat{\zeta})$ are

$$\mathrm{Var}_{opt}(\hat{\theta}) \approx \frac{n}{4Nd^2}; \qquad \mathrm{Var}_{opt}(\hat{\zeta}) \approx \frac{3n}{4Nd^4\theta^2}. \tag{S127}$$

VII. TOTAL EVOLUTION TIME SCALING COMPARISON

The total evolution time of the whole learning algorithm is defined as

$$T = \sum_{i=1}^{r} (\tau_i \times N), \tag{S128}$$

where r is the total number of experiment rounds, τ_i is the running time for each experiment, and N is the number of samples taken for each experiment. For analog-digital-hybrid learning algorithm 2 in which all the invariant subspaces are learned in parallel, we need to run $r = (n-1)$ experiments with each one taking $\tau_i = 2(2d-1) \times d$ running time and N samples for each circuit. Thus, the total evolution time can be computed as follows.

Theorem (Upper bound for the total evolution time for algorithm 2)**.** *Assuming the quantum system is evolving under a 2-body Hamiltonian specifying as Hamiltonian Eq. (2), our algorithm provides estimators* $\hat{\mathbf{c}}$ *with probability at least* $1-\delta$ *satisfying* $|\hat{c}_{ij} - c_{ij}| \leq \epsilon$ *for all* $i,j \in \{1,\cdots,n\}$ *and* $i \neq j$ *with total evolution time* $T = O(\frac{n^{3/2}}{\epsilon})$.

Proof. According to Chebyshev equality, for any individual estimator $\hat{c}_{ij}$,

$$\Pr(|\hat{c}_{ij} - c_{ij}| \leq \epsilon) \geq 1 - \frac{\mathrm{Var}(\hat{c}_{ij})}{\epsilon^2}. \tag{S129}$$

We plug in the variance given in Eq. (40) and set $\frac{\mathrm{Var}(\hat{c}_{ij})}{\epsilon^2} = \delta$, we get $d^4 = \frac{3n}{4Na_i^2\epsilon^2\delta}$, then the total evolution time $T = \sum_{i=1}^{r}(\tau_i \times N) = (n-1) \times (2(2d-1) \times d) \times N \approx nd^2N$ scales as

$$T = O(\frac{n^{3/2}}{\epsilon}). \tag{S130}$$

□

We have the quadratic speedup result which indicates $O(n)$ experimental rounds are required for learning $O(n^2)$ terms as indicated in Theorem II.5, but the total evolution time scales with $O(n^{3/2})$ which is due to the sampling overhead for parallelizing $O(n)$ invariant subspaces.

We also consider here applying the approach proposed in [18] for learning all-to-all connected Hamiltonians. In this method, the many-body Hamiltonian is first decoupled into non-interacting patches by applying random Pauli operations on the qubits that connect different patches. Once decoupled, the remaining non-interacting pairs can be learned in parallel using robust phase estimation. However, in the all-to-all connected case, the Hamiltonian is no longer geometrically local. To fully disconnect the interaction graph, one must eliminate all $(n-2)$ residual couplings, leaving only a single interaction active at a time. This requires $O(n^2)$ experimental rounds, i.e., $r = O(n^2)$. Each experiment achieves Heisenberg-limited precision, with $\tau \times N \sim O(\epsilon^{-1})$, so the total evolution time scales as $O(n^2/\epsilon)$. Thus, the total evolution time of our parallel learning algorithm still scales better than the $O(n^2)$ suggests in [18] where $O(n^2)$ experimental rounds are required for learning all-to-all connected Hamiltonian for n qubits.

Besides, for the fully-analog learning algorithm 3, we increase the experimental rounds to focus on each invariant subspace one at a time for ensuring the whole algorithm is implementable on a fully analog quantum simulator. So here the number of rounds is $r = \frac{n(n-1)}{2}$ with each one taking $\tau_i = 2(2d-1) \times d$ running time and N samples for each circuit.

Theorem VII.1 (Upper bound for the total evolution time for algorithm 3)**.** *Assuming the quantum system is evolving under a 2-body Hamiltonian specifying as Hamiltonian Eq. (2), our algorithm provides estimators* $\hat{\mathbf{c}}$ *with probability at least* $1-\delta$ *satisfying* $|\hat{c}_{ij} - c_{ij}| \leq \epsilon$ *for each* $i, j \in \{1, \cdots, n\}$ *and* $i \neq j$ *with total evolution time* $T = O(\frac{n^2}{\epsilon})$.

Proof. According to Chebyshev equality, for any individual estimator $\hat{c}_{ij}$,

$$\Pr(|\hat{c}_{ij} - c_{ij}| \leq \epsilon) \geq 1 - \frac{\mathrm{Var}(\hat{c}_{ij})}{\epsilon^2}. \tag{S131}$$

We plug in the variance given in Eq. (40) and set $\frac{\mathrm{Var}(\hat{c}_{ij})}{\epsilon^2} = \delta$, we get $d^4 = \frac{3}{8Na_i^2\epsilon^2\delta}$, then the total evolution time $T = \sum_{i=1}^{r}(\tau_i \times N) = (n-1)n/2 \times (2(2d-1) \times d) \times N \approx n^2 d^2 N$ scales as

$$T = O(\frac{n^2}{\epsilon}). \tag{S132}$$

□

This scaling is consistent with the total evolution time reported in [18], as both approaches adopt a sequential strategy. The key distinction, however, is that our learning algorithm supports in-situ learning and operates in a fully analog manner, whereas the method in [18] relies on digital single-qubit gates and reshapes the Hamiltonian, thereby precluding in-situ learning.