Optimal Ansatz-free Hamiltonian Learning In Situ

Taiqi Zhou
Department of Information Engineering
The Chinese University of Hong Kong
1155242363@link.cuhk.edu.hk

Weiyuan Gong*
John A. Paulson School of Engineering
and Applied Sciences, Harvard University
wgong@g.harvard.edu

Abstract

Characterizing the features of a Hamiltonian that governs a quantum system serves as a fundamental subroutine of quantum device calibration, signal sensing, and error correction. Recently proposed protocols have achieved the optimal Heisenberg-limited scaling for learning ansatz-free Hamiltonians from their real-time evolutions without fully specifying interaction structures. However, these protocols rely on both deep circuits with interleaving probes and control, and extremely short time resolution, making them difficult to implement on near- and intermediate-term in situ quantum experiments. In this work, we propose a computationally efficient, *control-free*, and *ancilla-free* algorithm that uses only *Pauli product state preparation and measurement*, and learns an ansatz-free Hamiltonian H with $\|H\| \leq \Lambda$ in total evolution time of $\Theta\left(\frac{\Lambda}{\varepsilon^2}\log\left(\frac{\Lambda}{\varepsilon}\right)\right)$. The evolution time cost of our algorithm is *optimal* in its worst-case (Λ, ε)-dependence for any control-free protocols as we further prove a lower bound of $\Omega\left(\frac{\Lambda}{\varepsilon^2}\log\left(\frac{\Lambda}{\varepsilon}\right)\right)$. Technically, our method introduces a randomized-sampling framework that combines band-limited kernel-based time sampling with a displacement sieve for Hamiltonian structure learning. The characteristic probe time resolution depends only on Λ, which makes our protocol especially appealing in the high-precision regime for sensing and calibration applications.

1 Introduction

Understanding the interactions governing a quantum system is a central task in quantum information science. As programmable analog simulators, NISQ devices, and early fault-tolerant quantum processors continue to mature, the need for rigorous and scalable characterization protocols has become increasingly urgent for validating engineered quantum dynamics and for practical tasks such as calibration, sensing, verification, error mitigation, and error correction [1, 2]. In closed systems, interactions of quantum systems are encoded by the Hamiltonians. A standard key subroutine for characterization is Hamiltonian learning, which infers the underlying H from accessible data. At a high level, one may access H either through equilibrium information [3, 4, 5, 6, 7, 8, 9, 10, 11], or through real-time evolutions e^{-iHt}. The latter viewpoint, which is also the access model we focus on in this work, is especially appealing for in situ characterization. Hamiltonian learning from time-evolution is deeply connected to quantum metrology and sensing [12, 13, 14], which aims to detect tiny local signal perturbations to the Hamiltonian. The early works then extend the sensing problem to a global parameter estimation task, focusing on highly structured models with heuristic guarantees, sub-optimal costs, and experimentally motivated protocols [15, 16, 17, 18, 19, 20, 21, 22, 23].

For Hamiltonians with *known* and *structured* interaction families, Huang et al. [24] proposed a protocol with total evolution time achieving the optimal Heisenberg-limited scaling. Several similar Heisenberg-limited scaling algorithms are also obtained for local Hamiltonians with unknown

*Corresponding author. A part of this work was conducted while visiting California Institute of Technology.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

supports [25, 26], interacting bosons and fermionic systems [27, 28, 29], other structured many-body systems [30, 31, 32], and testing structured Hamiltonians [33].

However, in the *ansatz-free* setting, the unknown Hamiltonian may be an ansatz-free linear combination of nonlocal Pauli strings. In this setting, Zhao [34] obtained nearly Heisenberg-limited scaling assuming queries to reversal time evolutions, ancillary qubits, and deep circuits with interleaving queries to time evolutions and unitary control gates. Later, Hu et al [35] proposed a protocol achieving Heisenberg-limited scaling total evolution time without ancillary qubits and reversal time evolutions. More recently, for ansatz-free Hamiltonian supported by M potentially nonlocal Pauli strings, Sinha and Tong [36] obtained the state-of-the-art total evolution time of $\widetilde{O}(M/\varepsilon)$. However, reliance on ingredients such as ancillary entanglements, interleaved quantum controls during time evolutions, engineered inverse dynamics, and extremely short resolved evolution times makes these protocols difficult to realize in realistic *in situ* settings. On analog hardwares, one may not be able to pause the native dynamics and insert coherent control. Ancillary entanglement can also introduce substantial experimental and calibration overhead. In sensing and calibration with high precision requirement, the short resolved evolution times proportional to the accuracy demand are also experimentally infeasible. Heisenberg-limited methods offer better ε-scaling when these additional experimental resources are available. This raises the question of whether one can retain strong evolution time (sample) guarantees in a *control-free and ancilla-free* setting. From the lower bound perspective, quantum controls can fundamentally change the complexity of Hamiltonian learning: with control one can reach Heisenberg-limited scaling, whereas without control broad classes of many-body Hamiltonians are only learnable at the standard quantum limit (SQL) of $\Omega(1/\varepsilon^2)$ [37, 38]. Yet before this work, the best known ansatz-free Hamiltonian learning protocols in the control-free setting require an $O(1/\varepsilon^4)$ evolution time scaling [34, 39, 40], leaving a substantial gap between lower and upper bounds. This motivates the main question of the present work:

Can we achieve SQL in learning ansatz-free Hamiltonian without control?

In this work, we answer this question essentially affirmatively. We show that arbitrary Hamiltonians H with a supplied bound $\|H\| \le \Lambda(n)$ can be learned using *control-free and ancilla-free* protocols achieving an SQL scaling total evolution time cost.

Theorem 1.1 (Informal, see Theorem 4.1)**.** *Let $H = \sum_u \lambda_u P_u$ be an unknown Hamiltonian on n qubits with $\|H\| \le \Lambda$. There is a computationally efficient algorithm that estimates every non-identity coefficient $\{\lambda_u\}$ up to ε with high probability using total evolution time*

$$O\left(\frac{\Lambda}{\varepsilon^2}\log\frac{\Lambda}{\varepsilon}\right). \tag{1}$$

A key feature of our result is that the protocol achieving this bound is control-free and ancilla-free and uses only *Pauli product state preparation measurement*. Moreover, the average probe time remains $O(1/\Lambda)$ and is independent of ε. This feature is particularly appealing for sensing and calibration applications, where extremely short-time resolution is often a major practical bottleneck. As a result, our protocol is substantially closer to the capabilities of near- and intermediate-term in situ experiments. We compare our algorithm with the previous results as in the following Table 1.

We also prove that this performance is optimal for the control-free setting.

Theorem 1.2 (Informal, see Theorems 5.2)**.** *Given accuracy demand ε and norm bound Λ satisfying $0 < \varepsilon \le \Lambda/16$ and $\Lambda^2/\varepsilon^2 \le O(n)$, there exists a family of Hamiltonians of bounded norm Λ for which any control-free learning protocol requires a total evolution time of at least*

$$\Omega\left(\frac{\Lambda}{\varepsilon^2}\log\frac{\Lambda}{\varepsilon}\right). \tag{2}$$

Our result clarifies the relation between Hamiltonian learning and quantum metrology. In quantum metrology, one typically detect perturbations on parameters of a Hamiltonian with a known structure. In our setting, by contrast, the support of the Hamiltonian is unknown and may lie among exponentially many Pauli strings. Our results show that this seemingly harder *structure-learning* task can be carried out at essentially the same total-time scale with only a logarithmic overhead that our lower bound shows is unavoidable.

Table 1: Comparison with representative prior works on learning n-qubit Hamiltonians from time evolutions. For ansatz-free Hamiltonians, there are two settings which use M to denote the sparsity in Pauli basis and Λ to denote the operator norm. Here $\widetilde{\Theta}$ hides $\log(\Lambda/\varepsilon)$.

Reference	Hamiltonian structure	Control	Ancilla	Total evolution time
Huang et al. [24]	Local & Known structure	Discrete	No	$O(1/\varepsilon)$
Bakshi et al. [26]	Local & Bounded-interaction	Discrete	No	$O(\log n/\varepsilon)$
Zhao [34]	Ansatz-free	Continuous / Discrete	Yes	$\widetilde{O}(M/\varepsilon)$ / $\widetilde{O}(\Lambda^3/\varepsilon^4)$
Hu et al. [35]	Ansatz-free	Discrete	Yes / No	$\widetilde{O}(M^2/\varepsilon)$ / $\widetilde{O}(M^3/\varepsilon)$
Sinha and Tong [36]	Ansatz-free	Discrete	Yes	$\widetilde{O}(M/\varepsilon)$
This work	Ansatz-free	No	No	$\widetilde{\Theta}(\Lambda/\varepsilon^2)$

2 Preliminaries

In this section, we provide the background concepts and results required throughout this paper. We use $\|A\|$ and $\|A\|_{\mathrm{HS}}$ to represent the operator norm and Hilbert-Schmidt (Frobenius) norm of matrix A. For a function $f(x)$ defined on $x \in \mathbb{R}$, we use $\|f(x)\|_1 = \int_{\mathbb{R}} |f| dx$ to represent the L^1 norm of the function. We use $\widetilde{O}$ and $\widetilde{\Theta}$ to hide polylogarithmic factors.

2.1 Pauli observables

We define $\mathcal{P}_n = \{I, X, Y, Z\}^{\otimes n}$ to be the set of Pauli observables (strings), where

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3}$$

are single-qubit Pauli operators which represent transformations on a single qubit. For any two Pauli strings $P, Q \in \mathcal{P}_n$, we have either P and Q commute, i.e. $[P, Q] = PQ - QP = 0$, or P and Q anti-commute, i.e. $\{P, Q\} = PQ + QP = 0$

Due to the relationship $Y = iXZ$, each Pauli operator on a single qubit can be also represented by a two-dimensional vector. This vector contains the parts representing the action of the X and the part representing the action of the Z. Generally, for an n-qubit Pauli string, we have a binary vector in $V = \mathbb{F}_2^{2n} = \mathbb{F}_2^n \times \mathbb{F}_2^n$ as $u = (x \mid z)$ for $x, z \in \mathbb{F}_2^n$. We call V the symplectic space and u the symplectic vector, and it can be used as the label of any Pauli string. To each label $u = (x \mid z) \in V$, we associate the canonical Hermitian Pauli string P_u with $P_u^\dagger = P_u$ and $P_u^2 = I$. The interactions between Pauli operators P_u and P_v with $u = (x \mid z)$ and $v = (x' \mid z')$ can also be represented in symplectic form. We define $\omega(u, v) \coloneqq x \cdot z' + z \cdot x' \in \mathbb{F}_2$ and the Walsh character $\chi_u(v) \coloneqq (-1)^{\omega(u,v)}$, and one can verify that

$$P_u P_v = \chi_u(v) P_v P_u, \qquad 2^{-2n} \sum_{v \in V} \chi_u(v) \chi_{u'}(v) = \mathbb{1}[u = u']. \tag{4}$$

2.2 Hamiltonians of interacting systems

We now introduce the concept of Hamiltonian, which encodes the interaction forces between quantum particles in a physical system. A Hamiltonian is a Hermitian operator $H \in \mathcal{H}^d$ with $H^\dagger = H$, with its time evolution described as e^{-iHt}. Given a Hamiltonian, we assume it has eigenvalues $\{E_j\}_j$ and corresponding eigenstates $\{|e_j\rangle\}_j$ with $H|e_j\rangle = E_j|e_j\rangle$. When the Hamiltonian H has bounded operator norm $\|H\| \leq \Lambda$, we also have $|E_j| \leq \Lambda$. Besides the above spectral decomposition of Hamiltonians according to their eigenstates and eigenvalues, Hamiltonians can also be decomposed in the operator space. A common set of complete orthogonal bases is the Pauli operator $\mathcal{P}^{\otimes n}$, and

any Hermitian operator (thus Hamiltonian) can be decomposed into the Pauli basis. A Hamiltonian is said to be M-sparse if the Hamiltonian can be written as $H = \sum_{u\in S} \lambda_u P_u$ with $|S| \leq M$. In the extreme case when $M = 4^n$, the set of 4^n-sparse Hamiltonian covers all Hamiltonians. Based on the terminology above, we can clearly define our learning objectives.

Problem 2.1 (Hamiltonian Learning)**.** *Let H be a M-sparse Hamiltonian with $\|H\| \leq \Lambda$ such that $H = \sum_{u\in S} \lambda_u P_u$, where $\lambda_u \in \mathbb{R}$ denotes the interaction strength for corresponding Pauli string P_u, and $S \subset V$ denotes the unknown support with $|S| \leq M$. Suppose we are given oracle access to the time- evolution unitary $U(t) = e^{-iHt}$, our goal is to output an estimate $\widehat{H} = \sum_{u\in\widehat{S}} \widehat{\lambda}_u P_u$ such that $|\widehat{\lambda}_u - \lambda_u| \leq \varepsilon$ for any $u \in V \setminus \{0\}$.*

Given an accuracy demand $\varepsilon > 0$, to solve Problem 2.1, we only need to recover the coefficient in ε-support $S_\varepsilon \coloneqq \{u \in V \setminus \{0\} : |\lambda_u| \geq \varepsilon\}$ up to ε, and output $\widehat{\lambda}_u = 0$ for all other $u \in V \setminus S_\varepsilon$.

2.3 The access models

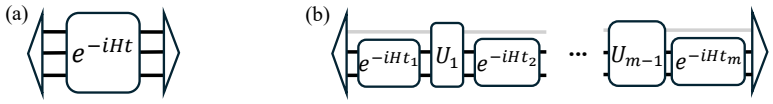

Figure 1: (a) Control-free ancilla-free Hamiltonian learning protocols considered in this work. (b) The general Hamiltonian learning settings with ancillary qubits (grey qubits) and quantum control.

In this section, we consider how we get access to the Hamiltonian and how we perform experiments to learn the Hamiltonian. The most general Hamiltonian learning protocols can be described by Figure 1 (b). In each experiment, the learner can prepare an input state on all working qubits (black qubits) and ancillary qubits (grey qubits), queries the Hamiltonian m times with evolution time $t_1, ..., t_m$ on the n working qubits interleaved with $m-1$ unitary control $U_1, ..., U_{m-1}$ between each two evolution time queries on all qubits, and make the final measurement on all qubits. However, while ancillary qubits and quantum control can improve learning efficiency, their implementation in real-world hardware is relatively costly. In this paper, we therefore consider protocols without ancillary qubits and quantum control (as shown in Figure 1 (a)). Moreover, we assume that the input state and measurement are restricted to a product state and a product measurement. Quantitatively, a single experiment can be described as follows.

Definition 2.2 (Control-free, ancilla-free, and product IO access model)**.** *We consider learning an unknown Hamiltonian H with a series of experiments described as follows: In each experiment (shot), the learner can only*

1. *prepare an arbitrary product state,*
2. *perform an evolution $U(t) = e^{-iHt}$ for some chosen duration $t \geq 0$,*
3. *and perform a single-qubit measurement on each qubit, recording all local outcomes.*

The learner can then collect all the measurement outcomes and reconstruct the coefficient of the Hamiltonian via classical postprocessing of complexity polynomial in n and Λ/ε.

We remark that our protocol only require Pauli product input states and Pauli product measurement, which is an even more restricted setting than Definition 2.2. The cost of a protocol in this model is measured by the total evolution time $T_{\text{tot}} \coloneqq \sum_{\text{all shots}} t_{\text{shot}}$ over all shots (experiments).

2.4 Same-Pauli and Cross-Pauli measurements

Suppose P and Q are two Pauli strings. We use the Heisenberg notation to describe the time evolution of a Pauli string P as $P(t) \coloneqq e^{iHt} P e^{-iHt}$. Given different choices of Pauli bases for the initial state and the final measurements, we classify the access model in Definition 2.2 into two types. The first type is referred to as the same-Pauli measurement, where the system is prepared under a Pauli string P and read out under the same Pauli basis. The corresponding statistical signal can be represented by

the auto-correlation function:

$$C_P(t) \coloneqq 2^{-n} \operatorname{Tr}(P(t)P) = 2^{-n} \operatorname{Tr}\big(e^{iHt} P e^{-iHt} P\big) \tag{5}$$

The first-order derivative of $C_P(t)$ at $t = 0$ is proportional to $\operatorname{Tr}([H, P]P)$, which is always zero for any Pauli string P. Its second-order derivative is proportional to $\operatorname{Tr}([H, [H, P]]P)$, indicating that the initial evolution of the system is entirely driven by terms that anti-commute with P. Since the first order derivative is zero, we can only read the coefficients λ_u^2 corresponding to the quadratic terms t^2, resulting in sign-free signals. Therefore, we use same-Pauli measurements only to recover $\widehat{S}_\varepsilon$.

Another type is called the cross-Pauli measurement, where the system is prepared under one Pauli string Q but read out under a different Pauli basis P. The corresponding signals are characterized by cross-correlation functions

$$F_{P,Q}(t) \coloneqq 2^{-n} \operatorname{Tr}(P(t)Q) = 2^{-n} \operatorname{Tr}\big(e^{iHt} P e^{-iHt} Q\big). \tag{6}$$

Here, the first-order derivative at $t = 0$ is proportional to $\operatorname{Tr}([H, P]Q)$, which can be non-zero. This indicates that the signal changes linearly with time t at $t = 0$, thus preserving the sign of λ_u. Therefore, cross-measurements can be used to estimate precise individual parameters after we have characterized $\widehat{S}_\varepsilon$.

2.5 The trace identity

We will need the following result to connect product-state experiments with the trace functions.

Proposition 2.3 (Trace rule)**.** *Given Pauli strings P, Q, sample uniformly from a product eigenbasis $\{|\psi\rangle\}$ of Q, let $q \in \{\pm 1\}$ be the eigenvalue of the prepared state, evolve by e^{-iHt}, and measure P with outcome $m \in \{\pm 1\}$. Then, we have*

$$\mathbb{E}[qm] = 2^{-n} \operatorname{Tr}(P(t)Q). \tag{7}$$

3 Sampling subroutines

This section provides several sampling subroutines as essential building blocks of our algorithm. To begin with, we decompose the Hamiltonian learning task into several subproblems. The first subproblem is how to design experiments with measurement results corresponding to the short-time derivatives of correlation functions. As the Hamiltonian has a bounded operator norm Λ, these can be addressed by constructing a kernel (see Section 3.1 for details). The second subproblem is how to choose the input states and measurement bases to utilize the information revealed by same-Pauli measurements and cross-Pauli measurements. To this end, we first use same-Pauli measurements to determine the support in the structure learning stage (see Section 3.2), and then use cross-Pauli measurements to accurately estimate the coefficients in the coefficient learning stage (see Section 3.3 for details). Concretely, when we represent Pauli strings in symplectic forms, same-Pauli measurements recover a candidate set of size independent of n that contains the ε-support of the Hamiltonian. Cross-Pauli measurements with randomly selected local Pauli bases recover all coefficients in the ε-support set by designing the parity variable and applying weight corrections.

3.1 Random time sampling

We now show how to obtain an unbiased estimation of the derivatives of the correlation function with a finite number of samples by constructing kernels. Consider same-Pauli measurements first, and the cross-Pauli measurement case can be addressed by a similar method. First, we diagonalize the Hamiltonian as $H|e_j\rangle = E_j|e_j\rangle$ with $|E_j| \le \Lambda$. We have $\langle e_j|P(t)|e_k\rangle = e^{i(E_j - E_k)t}\langle e_j|P|e_k\rangle$, and thus $C_P(t) = 2^{-n}\sum_{j,k} |\langle e_j|P|e_k\rangle|^2 e^{i(E_j - E_k)t}$. This indicates that $C_P(t)$ has a Fourier support of $[-2\Lambda, 2\Lambda]$, which is band-limited. One can construct a kernel to recover derivatives of $C_P(t)$ at zero as a weighted average of values at finite times. In particular, as we can rewrite $C_P(t) = \int_{-2\Lambda}^{2\Lambda} \cos(\omega t) d\mu(\omega)$ as an integral within the band limitation for some probability measure μ. Differentiating under the integral, we derive that

$$C_P''(0) = -\int_{-2\Lambda}^{2\Lambda} \omega^2 d\mu(\omega). \tag{8}$$

To obtain the sampling time, we need to rewrite the above equation to return to the time domain. We fix a real, even, smooth cutoff $\varphi \in C_c^\infty(\mathbb{R})$ satisfying $\varphi(\xi) = 1$ ($|\xi| \le 1$) and $\varphi(\xi) = 0$ ($|\xi| \ge 2$). We define $\Phi_\Lambda(\omega) \coloneqq \omega^2 \varphi(\omega/(2\Lambda))$. The cutoff is equal to one on the true frequency window $|\omega| \le 2\Lambda$, and it smoothly goes to zero before $|\omega| = 4\Lambda$. This smooth cutoff is important because it makes the kernel integrable. We define the kernel $K_\Lambda(t) \coloneqq -\frac{1}{2\pi}\int_{\mathbb{R}} \Phi_\Lambda(\omega) e^{i\omega t} d\omega$. We can thus obtain the Fourier inversion:

$$\int_{\mathbb{R}} K_\Lambda(t) C_P(t) dt = 2\int_0^\infty K_\Lambda(t) C_P(t) dt = -\int_{-2\Lambda}^{2\Lambda} \omega^2 d\mu(\omega) = C_P''(0). \tag{9}$$

We also need to regularize the kernel by defining

$$q_\Lambda(t) \coloneqq \frac{2|K_\Lambda(t)|}{\|K_\Lambda\|_1}, \qquad t \ge 0. \tag{10}$$

is the probability density corresponding to the sampling time. We are now ready to provide the random time sampling routine with its performance guarantee.

Proposition 3.1 (Unbiasedness for estimating $C_P''(0)$)**.** *Sample $\tau \sim q_\Lambda$, and then perform a single same-Pauli measurement with respect to P_v with evolution time τ. Let $q \in \{\pm 1\}$ be the eigenvalue of the prepared state of P_v, evolve by $e^{-iH\tau}$, and measure P_v with outcome $m \in \{\pm 1\}$. Define $X_{v,\tau} \coloneqq qm$ and the weighted sample $D_v \coloneqq \|K_\Lambda\|_1 \operatorname{sign}(K_\Lambda(\tau)) X_{v,\tau}$. Then the estimator D_v is an unbiased estimator of $C_{P_v}''(0)$, i.e., $\mathbb{E}[D_v] = C_{P_v}''(0)$.*

Similarly, for cross-Pauli measurements, we can also construct a kernel for estimating the first-order derivative. We first define $L_\Lambda(t) \coloneqq \frac{1}{2\pi}\int_{\mathbb{R}} \Psi_\Lambda(\omega) e^{-i\omega t} d\omega$, where $\Psi_\Lambda(\omega) \coloneqq i\omega\varphi(\omega/(2\Lambda))$, and then regularize it as

$$p_\Lambda(t) \coloneqq \frac{2|L_\Lambda(t)|}{\|L_\Lambda\|_1}, \qquad t \ge 0. \tag{11}$$

Due to the different structure of the first-order derivative, we need additional parity designs, which will be covered in Section 3.3.

3.2 Displacement sampling

We now discuss how to locate the support set using same-Pauli measurements. Note that $C_P''(0) = -4\sum_{u:\{P_u,P\}=0} \lambda_u^2$, indicating that the same-Pauli measurement signal of a Pauli string is the sum of squares of the coefficients of all terms that anti-commute with it. For simplicity, we denote $P_{(x|z)} = i^{x\cdot z}\prod_{i=1}^n X_i^{x_i} Z_i^{z_i}$ in the symplectic form. Let $|a\rangle = |a_1, \ldots, a_n\rangle$ be a computational basis state of Z such that $Z^z|a\rangle = (-1)^{z\cdot a}|a\rangle$ and $X^x|a\rangle = |a+x\rangle$. Therefore, we have $P_{(x|z)}|a\rangle = i^{x\cdot z}(-1)^{z\cdot a}|a+x\rangle$. Similarly, we define $|b\rangle_X = |b_1, \ldots, b_n\rangle_X$ as a X-basis eigenvector. Since $Z^z|b\rangle_X = |b+z\rangle_X$ and $X^x|b+z\rangle_X = (-1)^{x\cdot(b+z)}|b+z\rangle_X$, we have $P_{(x|z)}|b\rangle_X = i^{x\cdot z}(-1)^{x\cdot(b+z)}|b+z\rangle_X = (-i)^{x\cdot z}(-1)^{x\cdot b}|b+z\rangle_X$. The above results indicate that X operators result in bit flips for Z-basis same-Pauli measurements, and Z operators result in bit flips for X-basis same-Pauli measurements. We denote displacements d and e as the bit flip patterns of Z- and X-basis same-Pauli measurements, and define

$$W_x(d) \coloneqq \sum_{z\in\mathbb{F}_2^n} \lambda_{(d|z)}^2, \qquad W_z(e) \coloneqq \sum_{x\in\mathbb{F}_2^n} \lambda_{(x|e)}^2. \tag{12}$$

as sums of λ_u^2 of all terms that have corresponding displacements of the Z- and X-basis. We can show the following proposition.

Proposition 3.2 (Distributions of Z- and X-basis displacements)**.** *Let $D \in \mathbb{F}_2^n$ be the displacement for Z-basis same-Pauli measurement with a uniformly random eigenstate of Z^n as input and $\tau \sim q_\Lambda$, and let $E \in \mathbb{F}_2^n$ be the displacement for X-basis same-Pauli measurements. Then for all nonzero $d, e \in \mathbb{F}_2^n$, we have*

$$\Pr[D = d] \ge \frac{2W_x(d)}{\|K_\Lambda\|_1}, \qquad \Pr[E = e] \ge \frac{2W_z(e)}{\|K_\Lambda\|_1}. \tag{13}$$

Given the accuracy demand ε, we define

$$\Pi_x^{(\varepsilon)} \coloneqq \{d \in \mathbb{F}_2^n : W_x(d) \geq \varepsilon^2\}, \qquad \Pi_z^{(\varepsilon)} \coloneqq \{e \in \mathbb{F}_2^n : W_z(e) \geq \varepsilon^2\}. \tag{14}$$

Note that $\sum_d W_x(d) = \sum_e W_z(e) = \sum_u \lambda_u^2 = 2^{-n} \operatorname{Tr}\big(H^2\big) \leq \Lambda^2$, each of these sets contains at most Λ^2/ε^2 nonzero elements. We then design a displacement sieve algorithm to reconstruct two sets that contain $\Pi_x^{(\varepsilon)}$ and $\Pi_z^{(\varepsilon)}$.

Lemma 3.3 (Displacement sieve algorithm)**.** *For any $0 < \varepsilon \leq \Lambda$ and failure probability $\eta \in (0,1)$, there is a protocol using same-Pauli measurements that outputs sets $\widehat{\Pi}_x, \widehat{\Pi}_z \subseteq \mathbb{F}_2^n$ with size $O(\frac{\Lambda^2}{\varepsilon^2} \log \frac{4}{\eta})$ such that, $\Pi_x^{(\varepsilon)} \setminus \{0\} \subseteq \widehat{\Pi}_x$ and $\Pi_z^{(\varepsilon)} \setminus \{0\} \subseteq \widehat{\Pi}_z$, with probability at least $1-\eta$ in deterministic total evolution time at most $O(\frac{\Lambda}{\varepsilon^2} \log \frac{4\Lambda}{\varepsilon} \log \frac{4}{\eta})$.*

After performing the displacement sieve algorithm, we formed a candidate set by combining these components using the Cartesian product.

$$\widehat{\mathcal{U}}_{\mathrm{proj}} \coloneqq (\widehat{\Pi}_x \cup \{0\}) \times (\widehat{\Pi}_z \cup \{0\}) \setminus \{0\}. \tag{15}$$

Every coefficient with $|\lambda_{(x|z)}| \geq \varepsilon$ belongs to this candidate set. Therefore, the final support is searched inside $\widehat{\mathcal{U}}_{\mathrm{proj}}$. If the displacement sieve is run with failure probability η, then the deterministic size bound is $|\widehat{\mathcal{U}}_{\mathrm{proj}}| \leq O\left(\frac{\Lambda^4}{\varepsilon^4} \log^2 \frac{4}{\eta}\right)$.

3.3 Parity sampling

To estimate the values of the coefficients given $\widehat{\Pi}_x, \widehat{\Pi}_z$, we use cross-Pauli measurements. Fix a Pauli string $c = (c_1, \ldots, c_n) \in \{X,Y,Z\}^n$. At each qubit i, we choose the two remaining Pauli bases a_i, b_i in cyclic order $(c_i, a_i, b_i) \in \{(X,Y,Z),(Y,Z,X),(Z,X,Y)\}$. The letter c_i is the basis whose coefficient we want to estimate through a cross-Pauli measurement experiment with input basis a_i and output basis b_i. Given $A, B \subseteq [n]$, we define $Q_A \coloneqq \bigotimes_{i=1}^n Q_i^{(A)}$ with $Q_i^{(A)} = a_i$ if $i \in A$ and $Q_i^{(A)} = I$ if $i \notin A$. Similarly, we define $P_B \coloneqq \bigotimes_{i=1}^n P_i^{(B)}$ with $P_i^{(B)} = b_i$ if $i \in B$ and $P_i^{(B)} = I$ if $i \notin B$. For each (A,B), we define an associated Pauli label $u_c(A,B)$ with $(P_{u_c(A,B)})_i$ being I, a_i, b_i, and c_i when $i \notin A \cup B$, $i \in A \setminus B$, $i \in B \setminus A$, and $i \in A \cap B$.

We consider the first-order derivative of the corresponding cross-correlation $F_{A,B}^{(c)}(t) \coloneqq 2^{-n} \operatorname{Tr}(P_B(t) Q_A)$. A naive observation is that $\frac{1}{2t}\Big(F_{A,B}^{(c)}(t) - F_{A,B}^{(c)}(-t)\Big)$ gives an accurate estimation of the derivative as $t \to 0$. However, our access model in Definition 2.2 does not allow inverse evolutions. To address this issue, we notice that $P_{u_c(A,B)} P_B = (-1)^{|A \cap B|} P_B P_{u_c(A,B)}$, which indicates that $(F_{A,B}^{(c)})'(0)$ is nonzero only when $|A \cap B|$ is odd. Based on this fact, we design the measurement subroutine to measure the corresponding first-order derivative when $|A \cap B|$ is odd. First, we sample τ from the cross-Pauli measurements kernel density p_Λ as in Eq. (11) and sample $S \in \{\pm 1\}$ uniformly. If $S = +1$, we prepare a random local a_i eigenstate on each qubit, evolve under the Hamiltonian for time τ, and measure along the basis b_i on every qubit, obtaining outcomes $m_i \in \{\pm 1\}$. If $S = -1$, we reverse b_i and a_i as in the case of $S = +1$, obtaining outcomes $n_i \in \{\pm 1\}$. We define the parity variable $Z_{A,B}^{(c)} = \prod_{i \in A} s_i \prod_{i \in B} m_i$ if $S = +1$ and $Z_{A,B}^{(c)} = -\prod_{i \in B} r_i \prod_{i \in A} n_i$ if $S = -1$, where s_i, r_i are the signs of the eigenstates of the input quantum states for a_i when $S = +1$ or b_i when $S = -1$. We then define the weighted parity sample

$$G_{A,B}^{(c)} \coloneqq \|L_\Lambda\|_1 \operatorname{sign}(L_\Lambda(\tau)) Z_{A,B}^{(c)}, \tag{16}$$

and show that it is an unbiased estimator of $(F_{A,B}^{(c)})'(0)$ as shown in the following proposition.

Proposition 3.4 (Unbiasedness for estimating $(F_{A,B}^{(c)})'(0)$)**.** *Let $k \coloneqq |A \cap B|$. Then $G_{A,B}^{(c)}$ is an unbiased estimator for $(F_{A,B}^{(c)})'(0)$, i.e., $\mathbb{E}[G_{A,B}^{(c)}] = (F_{A,B}^{(c)})'(0)$. We also have $(F_{A,B}^{(c)})'(0) = 2\sigma_c(A,B)\lambda_{u_c(A,B)}$ for odd k and 0 for even k, where $\sigma_c(A,B) \coloneqq i(-i)^k \in \{\pm 1\}$ when k is odd. In addition, we have $|G_{A,B}^{(c)}| \leq 2\ell_0 \Lambda$ for a constant $\ell_0 > 0$.*

Algorithm 1: Sampling subroutines

Procedure *DisplacementSieve($\varepsilon, \eta, \Lambda$) (Lemma 3.3)*:
  Set $p = \varepsilon^2/(4\kappa_0\Lambda^2)$, $r = \lceil \Lambda^2/\varepsilon^2 \rceil$, $N = \lceil C_0 p^{-1} \log(4r) \rceil$, $L = \lceil C_1 \log(4/\eta) \rceil$; $\widehat{\Pi}_x = \widehat{\Pi}_z = \varnothing$
  foreach $B \in \{Z, X\}$ *and* $\ell = 1, \ldots, L$ **do**
    Draw independent $t_1, \ldots, t_N \sim q_\Lambda$
    if $\sum_j t_j \le 4\kappa_1 N/\Lambda$ **then**
      Run the N independent B-basis same-Pauli shots at t_j, recording $h_j = a_j + b_j$
      Add nonzero labels seen $\ge Np$ times to $\widehat{\Pi}_x$ ($B = Z$) or $\widehat{\Pi}_z$ ($B = X$)
  return $\widehat{\Pi}_x, \widehat{\Pi}_z$
Procedure *ParityCoefficients($\varnothing \neq \mathcal{U} \subset V \setminus \{0\}, \alpha, \eta, \Lambda$) (Lemma 3.5)*:
  Set $N = \lceil C_0 \Lambda^2 \alpha^{-2} \log(16|\mathcal{U}|) \rceil$, $J = \lceil 16 \log(2/\eta) \rceil$; initialize $\widetilde{\lambda}_j(u) = 0$ for all j, u
  for $j = 1, \ldots, J$ **do**
    Draw independent $T_1, \ldots, T_N \sim p_\Lambda$
    if $\sum_s T_s \le 4\ell_1 N/\Lambda$ **then**
      Use Section 3.3's random frames/orientations at T_s to obtain $\widetilde{Y}_{u,s}$, zero for invisible u
      Set $\widetilde{\lambda}_j(u) = N^{-1} \sum_s \widetilde{Y}_{u,s}$ for each $u \in \mathcal{U}$
  return $\widehat{\lambda}_u = \mathrm{median}_{1 \le j \le J} \widetilde{\lambda}_j(u)$

Given Proposition 3.4, from one experiment, we can get a random sample of $Z^{(c)}_{A,B}$ for every pair (A, B) with odd $|A \cap B|$, and each $Z^{(c)}_{A,B}$ corresponds to a label $u = u_c(A, B)$. The resulting random variable provides an unbiased sample of λ_u. However, not every label appears in one experiment and different coefficients are sampled with different frequencies. A label $u \neq 0$ contributes a sample only when it matches the basis c on an odd number of indices $\mathcal{V}(c) \coloneqq \{u \in V \setminus \{0\} : |\{i : u_i = c_i\}| \text{ is odd}\}$. Let $(A_c(u), B_c(u))$ be a pair with $u = u_c(A, B)$, and define

$$Y^{(c)}_u \coloneqq \tfrac{1}{2}\sigma_c(A_c(u), B_c(u)) G^{(c)}_{A_c(u), B_c(u)}. \tag{17}$$

Then $\mathbb{E}[Y^{(c)}_u] = \lambda_u$ and $|Y^{(c)}_u| \le \ell_0\Lambda$, so each visible label contributes one unbiased sample when u is visible. To achieve global unbiasedness, we need to apply the following correction. We choose c uniformly randomly from $\{X, Y, Z\}^n$, and define $q(u) \coloneqq \Pr[u \in \mathcal{V}(c)] = \frac{1-(1/3)^{w(u)}}{2} \ge \frac{1}{3}$, where $w(u) = |\{i : u_i \neq I\}|$. We then construct the estimator

$$\widetilde{Y}_u \coloneqq q(u)^{-1} \mathbb{1}[u \in \mathcal{V}(C)] Y^{(C)}_u. \tag{18}$$

Then $\mathbb{E}[\widetilde{Y}_u] = \lambda_u$ and $|\widetilde{Y}_u| \le 3\ell_0\Lambda$, so that the estimator becomes unconditionally unbiased for every label u. Quantitatively, we have the following lemma.

Lemma 3.5 (Coefficient estimation)**.** *For any $0 < \varepsilon \le \Lambda$, given a nonempty candidate class $\widehat{\mathcal{U}}_{\mathrm{proj}} \subseteq V \setminus \{0\}$ and failure probability $\eta \in (0, 1)$, there is a protocol using cross-Pauli measurements that outputs $\{\widehat{\lambda}_u\}_{u \in \widehat{\mathcal{U}}_{\mathrm{proj}}}$ satisfying $\max_{u \in \widehat{\mathcal{U}}_{\mathrm{proj}}} |\widehat{\lambda}_u - \lambda_u| \le \varepsilon$ with probability at least $1 - \eta$ in deterministic total evolution time at most $O(\frac{\Lambda}{\varepsilon^2} \log\Big(4|\widehat{\mathcal{U}}_{\mathrm{proj}}|\Big) \log \frac{2}{\eta})$.*

4 Upper Bound

4.1 Main algorithm

We will now combine the same-Pauli measurements stage in Lemma 3.3 and the cross-Pauli measurements stage in Lemma 3.5 into a complete protocol to solve Problem 2.1 as in Algorithm 2.

The performance guarantee of Algorithm 2 is given by the following theorem.

Algorithm 2: Control-free reconstruction at detection scale ε

Input: target detection scale $\varepsilon > 0$, failure probability $\delta \in (0,1)$, certified bound $\Lambda \geq \|H\|$

Output: estimate $\widehat{S}_\varepsilon$ of the ε-detectable support and estimator $\widehat{H}_\varepsilon = \sum_{u \in \widehat{S}_\varepsilon} \widehat{\lambda}_u P_u$

Set internal scales $\varepsilon_{\mathrm{lin}} \leftarrow \varepsilon/6$ and threshold $\theta \leftarrow \varepsilon/2$; Run the displacement sieve (Lemma 3.3) with scale ε and failure $\delta/2$ to obtain $\widehat{\Pi}_x$, $\widehat{\Pi}_z$; Form the Cartesian candidate set $\widehat{\mathcal{U}}_{\mathrm{proj}} \leftarrow (\widehat{\Pi}_x \cup \{0\}) \times (\widehat{\Pi}_z \cup \{0\}) \setminus \{0\}$ **if** $\widehat{\mathcal{U}}_{\mathrm{proj}} = \varnothing$ **then**

  $\widehat{S}_\varepsilon \leftarrow \varnothing$; $\widehat{H}_\varepsilon \leftarrow 0$

else

  Run Lemma 3.5 on $\widehat{\mathcal{U}}_{\mathrm{proj}}$ at accuracy $\varepsilon_{\mathrm{lin}}$ and failure $\delta/2$ to obtain $\{\widehat{\lambda}_u\}_{u \in \widehat{\mathcal{U}}_{\mathrm{proj}}}$

  $\widehat{S}_\varepsilon \leftarrow \{u \in \widehat{\mathcal{U}}_{\mathrm{proj}} : |\widehat{\lambda}_u| > \theta\}$; $\widehat{H}_\varepsilon \leftarrow \sum_{u \in \widehat{S}_\varepsilon} \widehat{\lambda}_u P_u$

end

Theorem 4.1 (Performance guarantee of Algorithm 2)**.** *Let $H = \sum_{u \in S} \lambda_u P_u$ with $\|H\| \leq \Lambda$, and fix $0 < \varepsilon \leq \Lambda$ and $\delta \in (0,1)$. With probability at least $1 - \delta$, Algorithm 2 outputs $(\widehat{S}_\varepsilon, \widehat{H}_\varepsilon)$ satisfying*

$$S_\varepsilon \subseteq \widehat{S}_\varepsilon \subseteq \{u \in V \setminus \{0\} : |\lambda_u| > \varepsilon/3\}, \qquad \max_{u \in S_\varepsilon} |\widehat{\lambda}_u - \lambda_u| \leq \varepsilon/6, \tag{19}$$

where $S_\varepsilon \coloneqq \{u \in V \setminus \{0\} : |\lambda_u| \geq \varepsilon\}$. The total evolution time of Algorithm 2 satisfies

$$T_{\mathrm{tot}}^{\mathrm{det}} \leq C \frac{\Lambda}{\varepsilon^2} \left(1 + \log \frac{\Lambda}{\varepsilon}\right) \log \frac{4}{\delta} \log\log \frac{8}{\delta}. \tag{20}$$

We remark that any input value $\Lambda \geq \|H\|$ preserves the correctness of our algorithm. It only affects runtime via the kernel interval length.

4.2 Implementation under the discrete time assumption

Our displacement sieve sampling subroutine in Lemma 3.3 implicitly assumes we can randomly sample an arbitrary time τ and evolve under τ, which requires the ability to perform continuous time evolution of a Hamiltonian. However, real-world physical devices typically have a minimum resolvable time t_0. We will now adapt our algorithm to an even more restrictive discretization assumption where we only allow resolution time chosen from a discrete-time grid $\mathcal{T}_{t_0} \coloneqq \{0, t_0, 2t_0, \ldots\}$.

First, for any Hamiltonian learning of H when $\|H\| \leq \Lambda$, there is a fundamental barrier for uniform learning to accuracy $\varepsilon < \Lambda/4$ that requires a resolution time of at most $t_0 < \pi/\Lambda$. We provide a complete proof in the appendix. Second, we can discretize the sampling rules with only a constant overhead on the total evolution time. The key observation is that the correlation function only contains frequencies in $[-2\Lambda, 2\Lambda]$, while the kernels K_Λ and L_Λ are constructed using a smooth cutoff and therefore have frequency support $[-4\Lambda, 4\Lambda]$. As a result, the product $K_\Lambda(t)f(t)$ has a bandwidth of at most $[-6\Lambda, 6\Lambda]$. Therefore, as long as the time grid spacing satisfies $t_0 < \frac{\pi}{6\Lambda}$, we have

$$\int_{\mathbb{R}} K_\Lambda(t) f(t) dt = t_0 \sum_{m \in \mathbb{Z}} K_\Lambda(m t_0) f(m t_0). \tag{21}$$

In this scenario, we sample an integer $m \geq 0$ using the probabilities and zero-time null outcome in Appendix F, and run the same experiment at time $\tau = m t_0$. If one wants to limit the maximum evolution time per experiment, we can set a truncation parameter $R > 0$ and truncate the sum at $m \leq R/(2\Lambda t_0)$. The resulting error decays faster than any polynomial in R, while ensuring a maximum evolution time of $R/(2\Lambda)$ per experiment. Theorem F.7 gives a polylogarithmic R preserving Theorem 4.1. We further show that a resolution time of $O(1/\Lambda)$ is necessary in the grid setting.

5 Lower Bound

Here, we show that our protocol is optimal in its worst-case (Λ, ε)-dependence over all possible control-free protocols. We will illustrate this by proving two lower bounds.

Learning an unknown Hamiltonian without control or structured prior knowledge is subject to a worst-case SQL lower bound [37]. Consider the Hamiltonian family $H_\theta = \Lambda(\cos\theta Z + \sin\theta X)$ parametrized by θ. Near $\theta = 0$, the coefficient of X is $\Lambda\sin\theta \approx \Lambda\theta$. To distinguish two coefficients by ε, one must distinguish two parameters separated by ε/Λ. In a control-free experiment with probe length t, the Fisher information obtained scales at most linearly in Λt. This means that if we want to estimate the coefficient of X, it takes at least $\Omega(\Lambda/\varepsilon^2)$ time.

Lemma 5.1 (SQL lower bound)**.** *For every $0 < \varepsilon \leq \Lambda/4$, any control-free protocol with deterministic total evolution time $T_{\rm tot}$ that estimates the X-coefficient of $H_\theta = \Lambda(\cos\theta Z + \sin\theta X)$, at $|\theta| \leq \pi/6$, to accuracy ε with a success probability $\geq 2/3$ must satisfy $T_{\rm tot} = \Omega(\Lambda/\varepsilon^2)$.*

Proving a tighter lower bound is nontrivial because proofs using information theory tools usually only control how fast one can estimate a single coefficient, and can hardly constrain the strategy that learns many support locations in parallel. To improve the learning difficulty, we designed a hard family of M-sparse Hamiltonians with M pair-wise anticommuting Pauli observables. According to Proposition 9 of Ref. [41], $M \leq 2n + 1$. We index each Pauli string in the Hamiltonian by Q_{a,x_a} randomly and secretly chosen with $a = 1, 2, ..., M$ and $x_a \in \{1, 2\}$. The Hamiltonian is written as

$$H_{x,\sigma} = \frac{\Lambda}{\sqrt{M}} \sum_{a=1}^{M} (-1)^{\sigma_a} Q_{a,x_a}, \tag{22}$$

where $\sigma_a \in \{0, 1\}$ is a random sign, fixed throughout the protocol. Hence, we have $H_{x,\sigma}^2 = \Lambda^2 I$ and $\|H_{x,\sigma}\| = \Lambda$. Let $M = \Theta(\Lambda^2/\varepsilon^2)$ so that each non-zero coefficient has magnitude $\Lambda/\sqrt{M}$, which we choose to be at least 4ε. Therefore, every non-zero term is detectable for a Hamiltonian learning algorithm. Thresholding an ε-accurate estimate at 2ε recovers the support exactly. For adaptive protocols, Appendix G conditions on the complete previous transcript and simulates each shot using a coherent comparison state. A total evolution time $T_{\rm tot}$ requires comparison size $O(\Lambda T_{\rm tot})$, whereas recovery of the M hidden binary labels requires size $\Omega(M \log M)$ even under a joint measurement. After simple deductions, we can get a lower bound on total evolution time of $T_{\rm tot} = \Omega\left(\frac{\Lambda}{\varepsilon^2}\log\frac{\Lambda}{\varepsilon}\right)$.

Theorem 5.2 (Unified lower bound)**.** *Given accuracy demand ε and norm bound Λ with $0 < \varepsilon \leq \Lambda/16$ and $\Lambda^2/\varepsilon^2 \leq O(n)$, and sparsity bound $M \geq \lfloor(\Lambda/(4\varepsilon))^2\rfloor$, any control-free protocol that solves Problem 2.1 with probability $\geq 2/3$ must have deterministic total evolution time*

$$T_{\rm tot} \ \geq \Omega\left(\frac{\Lambda}{\varepsilon^2}\log\frac{\Lambda}{\varepsilon}\right). \tag{23}$$

Simulations of the complete protocol on sparse and k-local Hamiltonians are reported in Appendix H and Figure 2.

6 Discussion

In this work, we show that ansatz-free Hamiltonian learning can be performed in an in situ regime: our protocol is control-free, ancilla-free, and uses only Pauli product state preparation and measurement, while achieving the optimal total evolution time $\widetilde{\Theta}(\Lambda/\varepsilon^2)$ in the regime of Theorem 5.2, which is essentially the same total evolution time scale as SQL metrology, despite the additional burden of support recovery. However, our protocol still requires probe times with characteristic scale $O(1/\Lambda)$ even in the discretized setting. It is natural to ask whether, beyond the fixed-grid setting, one can further relax this requirement to a constant time resolution, or conversely prove that some form of short-time resolution is fundamentally necessary for control-free Hamiltonian learning, analogous to the resolution barrier recently observed in ansatz-free Lindbladian learning [39]. In addition, our lower bound is stated in terms of Λ and ε. It would be very interesting to determine whether there is a lower bound that depends explicitly on the sparsity M under any settings to decide the optimality of the previous algorithms along this line [35, 36].

Acknowledgments and Disclosure of Funding

We thank Zhiding Liang, Nikita Romanov, and Sisi Zhou for valuable feedback on the manuscript. We thank Richard Allen, Sitan Chen, Senrui Chen, Hong-Ye Hu, Hsin-Yuan Huang, and Muzhou Ma for helpful discussions during this work.

Funding. The authors received no external funding or third-party support for this work.

Competing Interests. The authors declare no competing financial interests related to this work.

References

[1] Jens Eisert, Dominik Hangleiter, Nathan Walk, Ingo Roth, Damian Markham, Rhea Parekh, Ulysse Chabaud, and Elham Kashefi. Quantum certification and benchmarking. *Nature Reviews Physics*, 2(7):382–390, 2020.

[2] Jose Carrasco, Andreas Elben, Christian Kokail, Barbara Kraus, and Peter Zoller. Theoretical and experimental perspectives of quantum verification. *PRX Quantum*, 2(1):010102, 2021.

[3] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. Sample-efficient learning of interacting quantum systems. *Nature Physics*, 17(8):931–935, 2021.

[4] Jeongwan Haah, Robin Kothari, and Ewin Tang. Optimal learning of quantum Hamiltonians from high-temperature Gibbs states. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 135–146. IEEE, 2022.

[5] Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. Learning quantum Hamiltonians at any temperature in polynomial time. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1470–1477, 2024.

[6] Andi Gu, Lukasz Cincio, and Patrick J Coles. Practical Hamiltonian learning with unitary dynamics and Gibbs states. *Nature Communications*, 15(1):312, 2024.

[7] Xiao-Liang Qi and Daniel Ranard. Determining a local Hamiltonian from a single eigenstate. *Quantum*, 3:159, 2019.

[8] Zhi Li, Liujun Zou, and Timothy H Hsieh. Hamiltonian tomography via quantum quench. *Physical Review Letters*, 124(16):160502, 2020.

[9] Tim J Evans, Robin Harper, and Steven T Flammia. Scalable bayesian Hamiltonian learning. *arXiv:1912.07636*, 2019.

[10] Sitan Chen, Jordan Cotler, and Hsin-Yuan Huang. Quantum probe tomography. *arXiv:2510.08499*, 2025.

[11] Chi-Fang Chen, Anurag Anshu, and Quynh T. Nguyen. Learning quantum Gibbs states locally and efficiently. In *2025 IEEE 66th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1458–1464, 2025.

[12] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum metrology. *Physical Review Letters*, 96(1):010401, 2006.

[13] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Advances in quantum metrology. *Nature Photonics*, 5(4):222–229, 2011.

[14] Christian L Degen, Friedemann Reinhard, and Paola Cappellaro. Quantum sensing. *Reviews of Modern Physics*, 89(3):035002, 2017.

[15] Christopher E Granade, Christopher Ferrie, Nathan Wiebe, and David G Cory. Robust online Hamiltonian learning. *New Journal of Physics*, 14(10):103013, 2012.

[16] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and David G Cory. Hamiltonian learning and certification using quantum resources. *Physical Review Letters*, 112(19):190501, 2014.

[17] Sheng-Tao Wang, Dong-Ling Deng, and Lu-Ming Duan. Hamiltonian tomography for quantum many-body systems with arbitrary couplings. *New Journal of Physics*, 17(9):093017, 2015.

[18] Milan Holzäpfel, Tillmann Baumgratz, Marcus Cramer, and Martin B Plenio. Scalable reconstruction of unitary processes and Hamiltonians. *Physical Review A*, 91(4):042129, 2015.

[19] Liangyu Che, Chao Wei, Yulei Huang, Dafa Zhao, Shunzhong Xue, Xinfang Nie, Jun Li, Dawei Lu, and Tao Xin. Learning quantum Hamiltonians from single-qubit measurements. *Physical Review Research*, 3(2):023246, 2021.

[20] Christian Kokail, Rick van Bijnen, Andreas Elben, Benoît Vermersch, and Peter Zoller. Entanglement Hamiltonian tomography in quantum simulation. *Nature Physics*, 17(8):936–942, 2021.

[21] Frederik Wilde, Augustine Kshetrimayum, Ingo Roth, Dominik Hangleiter, Ryan Sweke, and Jens Eisert. Scalably learning quantum many-body Hamiltonians from dynamical data. *arXiv:2209.14328*, 2022.

[22] Dominik Hangleiter, Ingo Roth, Jonáš Fuksa, Jens Eisert, and Pedram Roushan. Robustly learning the Hamiltonian dynamics of a superconducting quantum processor. *Nature Communications*, 15(1):9595, 2024.

[23] Shi-An Guo, Yu-Kai Wu, Jing Ye, Lin Zhang, Ye Wang, Wen-Qian Lian, Rui Yao, Yu-Lin Xu, Chi Zhang, Yu-Zi Xu, et al. Hamiltonian learning for 300 trapped ion qubits with long-range couplings. *Science Advances*, 11(5):eadt4713, 2025.

[24] Hsin-Yuan Huang, Yu Tong, Di Fang, and Yuan Su. Learning many-body Hamiltonians with Heisenberg-limited scaling. *Physical Review Letters*, 130(20):200403, 2023.

[25] Muzhou Ma, Steven T Flammia, John Preskill, and Yu Tong. Learning k-body Hamiltonians via compressed sensing. *arXiv:2410.18928*, 2024.

[26] Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. Structure learning of Hamiltonians from real-time evolution. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1037–1050. IEEE, 2024.

[27] Haoya Li, Yu Tong, Tuvia Gefen, Hongkang Ni, and Lexing Ying. Heisenberg-limited Hamiltonian learning for interacting bosons. *npj Quantum Information*, 10(1):83, 2024.

[28] Hongkang Ni, Haoya Li, and Lexing Ying. Quantum Hamiltonian learning for the fermi-hubbard model. *Acta Applicandae Mathematicae*, 191(1):2, 2024.

[29] Arjun Mirani and Patrick Hayden. Learning interacting fermionic Hamiltonians at the Heisenberg limit. *Physical Review A*, 110(6):062421, 2024.

[30] Wenjun Yu, Jinzhao Sun, Zeyao Han, and Xiao Yuan. Robust and efficient Hamiltonian learning. *Quantum*, 7:1045, 2023.

[31] Daniel Stilck França, Liubov A. Markovich, V. V. Dobrovitski, Albert H. Werner, and Johannes Borregaard. Efficient and robust estimation of many-qubit hamiltonians. *Nature Communications*, 15(1):311, 2024.

[32] Daniel Stilck França, Tim Möbus, Cambyse Rouzé, and Albert H Werner. Learning and certification of local time-dependent quantum dynamics and noise. *arXiv:2510.08500*, 2025.

[33] Srinivasan Arunachalam, Arkopal Dutt, and Francisco Escudero Gutiérrez. Testing and learning structured quantum Hamiltonians. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 1263–1270, 2025.

[34] Andrew Zhao. Learning the structure of any Hamiltonian from minimal assumptions. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 1201–1211, 2025.

[35] Hong-Ye Hu, Muzhou Ma, Weiyuan Gong, Qi Ye, Yu Tong, Steven T. Flammia, and Susanne F. Yelin. Ansatz-free Hamiltonian learning with Heisenberg-limited scaling. *PRX Quantum*, 6:040315, Oct 2025.

[36] Savar D Sinha and Yu Tong. Improved Hamiltonian learning and sparsity testing through bell sampling. *arXiv:2509.07937*, 2025.

[37] Alicja Dutkiewicz, Thomas E O'Brien, and Thomas Schuster. The advantage of quantum control in many-body Hamiltonian learning. *Quantum*, 8:1537, 2024.

[38] Ziyun Chen and Jerry Li. Lower bounds for Hamiltonian parameter learning from time evolution. *arXiv:2509.20665*, 2025.

[39] Petr Ivashkov, Nikita Romanov, Weiyuan Gong, Andi Gu, Hong-Ye Hu, and Susanne F Yelin. Ansatz-free learning of Lindbladian dynamics in situ. *arXiv:2603.05492*, 2026.

[40] Matthias C Caro. Learning quantum processes and Hamiltonians via the pauli transfer matrix. *ACM Transactions on Quantum Computing*, 5(2):1–53, 2024.

[41] Pavel Hrubeš. On families of anticommuting matrices. *Linear Algebra and its Applications*, 493:494–507, 2016.

[42] William K Wootters. Statistical distance and hilbert space. *Physical Review D*, 23(2):357, 1981.

[43] Samuel L Braunstein and Carlton M Caves. Statistical distance and the geometry of quantum states. *Physical Review Letters*, 72(22):3439, 1994.

[44] Samuel L Braunstein, Carlton M Caves, and Gerard J Milburn. Generalized uncertainty relations: theory, examples, and lorentz invariance. *Annals of Physics*, 247(1):135–173, 1996.

A Extended Preliminaries

A.1 Hamiltonian and time evolution

Here, we provide more primary results on Hamiltonian and time evolution. In physics, the Hamiltonian describes the dynamics of a system. In quantum mechanics, it is a Hermitian operator that determines how a state evolves over time. The evolution is governed by the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle. \tag{24}$$

When H is time-independent and we set $\hbar = 1$, the solution is

$$|\Psi(t)\rangle = e^{-iHt}|\Psi(0)\rangle. \tag{25}$$

For an n-qubit system, H is a $2^n \times 2^n$ Hermitian matrix. We measure the size of the Hamiltonian by its operator norm

$$\|H\| \coloneqq \sup_{\|\psi\|_2=1} \|H|\psi\rangle\|. \tag{26}$$

Since H is Hermitian, $\|H\|$ is simply the absolute value of the largest eigenvalue of H as $\|H\| = \max_j |E_j|$, where $\{E_j\}$ are the eigenvalues of H.

We also use the Hilbert-Schmidt (Frobenius) norm for operators, which is defined as

$$\|A\|_{\mathrm{HS}}^2 \coloneqq \mathrm{Tr}\big(A^\dagger A\big). \tag{27}$$

Furthermore, for any Hamiltonian on an n-qubit system, the operator norm and the Hilbert-Schmidt norms satisfy $\|H\|^2 \le \|H\|_{HS}^2 \le 2^n\|H\|^2$.

The above description on the evolution of the state $|\Psi(t)\rangle$ over time is known as the Schrödinger picture. An alternative approach to describe a Hamiltonian dynamics is to keep the state constant and transfer the time dependency to the observable being measured, which is known as the Heisenberg picture.

Definition A.1 (Heisenberg picture)**.** *Let $H = H^\dagger$ be a time-independent Hamiltonian and let $U(t) \coloneqq e^{-iHt}$ be the corresponding unitary time-evolution operator. For an observable $A = A^\dagger$, its Heisenberg evolution is defined by*

$$A_H(t) \coloneqq U(t)^\dagger A U(t) = e^{iHt} A e^{-iHt}. \tag{28}$$

Let $|\Psi(t)\rangle = U(t)|\Psi(0)\rangle$. Then

$$\langle\Psi(t)|A|\Psi(t)\rangle = \langle\Psi(0)|U(t)^\dagger A U(t)|\Psi(0)\rangle = \langle\Psi(0)|A_H(t)|\Psi(0)\rangle. \tag{29}$$

Therefore, measuring A at time t in the Schrödinger picture is equivalent to measuring the $A_H(t)$ on the initial state. For example, in our learning protocol, our direct learning object is the Pauli operator. We thus write the observed time-dependent signal as $P_H(t) = e^{iHt}Pe^{-iHt}$, and we omit H and write it as $P(t)$ when H is clear from the context.

A certified bound on $\min_c \|H - cI\| = (E_{\max} - E_{\min})/2$ can be used for Λ, since the scalar shift contributes only an unobservable global phase. Controlled-$U(t)$ access could reveal this shift.

The total evolution-time bound depends on n and M only through the supplied norm bound Λ. For $\Lambda = \Theta(n)$, the constant-confidence time is $O(n\varepsilon^{-2}\log(n/\varepsilon))$ in the high-precision regime and the mean probe time is $O(1/n)$. Classical processing depends on n and the candidate-set size.

A.2 Symplectic algebra

Here, we provide an extended introduction to Pauli observables and symplectic algebra. Note that the four single-qubit Pauli operators $\{I, X, Y, Z\}$, along with their phases, form a group:

$$\mathcal{P}_1 = \{\pm I, \pm iI, \pm X, \pm iX, \pm Y, \pm iY, \pm Z, \pm iZ\} \tag{30}$$

where multiplication in this group is matrix multiplication. We call $\mathcal{P}_1$ the single-qubit Pauli group. Accordingly, by replacing the group elements with tensor products of single-qubit Pauli operators, we obtain the multi-qubit Pauli group $\mathcal{P}_n$. Due to the presence of phase, the Pauli group $\mathcal{P}_n$ is not an abelian group. To simplify the operation, we can divide it by the center of the Pauli group $\{+I, -I, +iI, -iI\}$, thus obtaining the quotient group

$$V \coloneqq \mathcal{P}_n/\{\pm I, \pm iI\} \cong (\mathbb{F}_2^{2n}, \oplus). \tag{31}$$

V is a set of $2n$-bit binary vectors, and the corresponding group operation is the bitwise XOR addition. We call a vector u in the quotient group a label of the corresponding P_u in the original Pauli group. Concretely, we represent each vector as

$$u = (x, z) \in \mathbb{F}_2^n \times \mathbb{F}_2^n, \tag{32}$$

where the two binary strings x and z record the X- and Z-components of the Pauli operator. Because $Y = iXZ$, the effect of Y can be decomposed into the sequential effects of X and Z. On qubit j, the pair (x_j, z_j) specifies

$$(0,0) \leftrightarrow I, \qquad (1,0) \leftrightarrow X, \qquad (0,1) \leftrightarrow Z, \qquad (1,1) \leftrightarrow Y, \tag{33}$$

Therefore, this label fully preserves the tensor product pattern of the Pauli operator while discarding its global phase.

Symplectic labels allow us to represent commutation and anti-commutation without multiplying matrices directly. We define $\omega(u, v) \coloneqq x \cdot z' + z \cdot x' \pmod 2$, where the dot products are over $\mathbb{F}_2$. This number records the parity of the local anti-commutations between the two Pauli strings: if $\omega(u, v) = 0$, then P_u and P_v commute; if $\omega(u, v) = 1$, then they anticommute. For each $u \in V$, we can also define the Walsh character

$$\chi_u(v) \coloneqq (-1)^{\omega(u,v)} \in \{\pm 1\}, \tag{34}$$

and thus

$$\chi_u(v) = +1 \iff [P_u, P_v] = 0, \qquad \chi_u(v) = -1 \iff \{P_u, P_v\} = 0. \tag{35}$$

Hence, we have

$$\mathbb{1}[\{P_u, P_v\} = 0] = \frac{1 - \chi_u(v)}{2}. \tag{36}$$

Formally, under the symplectic notation, we have the following fact.

Fact A.2 (Pauli product rule)**.** *For every $u, v \in V$, there exists a phase $\zeta(u, v) \in \{\pm 1, \pm i\}$ such that $P_u P_v = \zeta(u, v) P_{u+v}$. If $\omega(u, v) = 0$, then $\zeta(u, v) \in \{\pm 1\}$. If $\omega(u, v) = 1$, then $\zeta(u, v) = is(u, v)$ for some sign $s(u, v) \in \{\pm 1\}$, and $[P_u, P_v] = 2is(u, v)P_{u+v}$.*

Since P_u can be a Pauli string formed by tensor products, we can also expand it into a qubit-by-qubit form $P_u = \bigotimes_{i=1}^n P_{u_i}$, for $P_{u_i} \in \{I, X, Y, Z\}$. The addition $u + v$ on it is still the bitwise XOR addition over $\mathbb{F}_2$, and $(u + v)_i$ denotes the ith local Pauli letter. For each fixed label u, one can assign a $\{\pm 1\}$ value to every $v \in V$ depending on their commutation relation. When these values are summed over a subspace, the result is either a sum of all 1 or cancellation out completely to 0. Quantitatively, we have the following lemma.

Lemma A.3 (Orthogonality of symplectic characters)**.** *If $K \leq V$ has dimension r, then for every $u \in V$,*

$$\sum_{v \in K} \chi_u(v) = \begin{cases} 2^r, & u \in K^\perp, \\ 0, & u \notin K^\perp. \end{cases} \tag{37}$$

Proof. If $u \in K^\perp$, then $\chi_u(v) = 1$ for all $v \in K$, so the sum is $|K| = 2^r$. If $u \notin K^\perp$, choose $v_0 \in K$ such that $\omega(u, v_0) = 1$. Then the map $v \mapsto v + v_0$ is a bijection of K, and

$$\chi_u(v + v_0) = \chi_u(v)\chi_u(v_0) = -\chi_u(v), \tag{38}$$

resulting in the sum canceling out in pairs. □

A.3 Proof of Proposition 2.3

Let $\rho_\psi \coloneqq |\psi\rangle\langle\psi|$. Consider the conditional probability of m given ψ

$$\mathbb{E}[m \mid \psi] = \operatorname{Tr}(\rho_\psi P(t)). \tag{39}$$

Therefore, we have

$$\mathbb{E}[qm] = 2^{-n} \sum_\psi q(\psi) \operatorname{Tr}(\rho_\psi P(t)) = 2^{-n} \operatorname{Tr}\left(P(t) \sum_\psi q(\psi) \rho_\psi \right). \tag{40}$$

Since $\{|\psi\rangle\}$ is an eigenbasis of Q and $Q|\psi\rangle = q(\psi)|\psi\rangle$, we have $Q = \sum_\psi q(\psi)\rho_\psi$. Namely

$$\mathbb{E}[qm] = 2^{-n} \operatorname{Tr}(P(t)Q). \tag{41}$$

B Same-Pauli Measurement

Same-Pauli measurements require obtaining the short-time second-order derivative of the auto-correlation function in Eq. (5) at the initial time. We will now show how to obtain the support of the Hamiltonian using these measurements.

B.1 Complementarity of anti-commuting probes

In this subsection, we will prove the following three points:

1. The short-time second-order derivative of the auto-correlation function is proportional to the sum of the squared coefficients of all anti-commutative terms with the probe operator in the Hamiltonian.
2. The raw signal obtained from the same-Pauli measurement can be written as a linear sum of the squares of the coefficients in the Hamiltonian.
3. Any single probe suffers from invisibility of some Hamiltonian terms. As a result, we need complementary bases to extract all Pauli terms in the Hamiltonian.

Proposition B.1 (Same-Pauli signal)**.** *For every Hermitian Pauli string P and Pauli Hamiltonian $H = \sum_{u\in V} \lambda_u P_u$, we have*

$$C_P''(0) = -4 \sum_{u:\{P_u,P\}=0} \lambda_u^2. \tag{42}$$

Proof. Note that $\frac{d}{dt}P(t) = i[H, P(t)]$ and $P''(0) = -[H,[H,P]]$, we have

$$C_P''(0) = 2^{-n} \operatorname{Tr}(P''(0)P) = -2^{-n} \operatorname{Tr}([H,[H,P]]P). \tag{43}$$

Write $X \coloneqq [H,P]$, we have

$$\begin{aligned} \operatorname{Tr}([H,[H,P]]P) &= \operatorname{Tr}((HX - XH)P) = \operatorname{Tr}(HXP) - \operatorname{Tr}(XHP) = \operatorname{Tr}(XPH) - \operatorname{Tr}(XHP) \\ &= \operatorname{Tr}(X(PH - HP)) = \operatorname{Tr}(X[P,H]) = -\operatorname{Tr}(X[H,P]) \\ &= -\operatorname{Tr}\big(X^2\big). \end{aligned} \tag{44}$$

Because H and P are Hermitian, $X = [H,P]$ is anti-Hermitian, so $X^\dagger = -X$ and therefore

$$-\operatorname{Tr}\big(X^2\big) = \operatorname{Tr}\big(X^\dagger X\big) = \|X\|_{HS}^2. \tag{45}$$

Thus

$$2^{-n} \operatorname{Tr}([H,[H,P]]P) = 2^{-n} \|[H,P]\|_{HS}^2. \tag{46}$$

As we can write $[H,P]$ as

$$[H,P] = \sum_{u\in S} \lambda_u [P_u, P] = \sum_{u:\{P_u,P\}=0} 2\lambda_u P_u P. \tag{47}$$

For $u \neq u'$, the Hilbert-Schmidt inner product of the corresponding terms is

$$\mathrm{Tr}\big((P_u P)^\dagger (P_{u'} P)\big) = \mathrm{Tr}(P P_u P_{u'} P) = \mathrm{Tr}(P_u P_{u'}), \tag{48}$$

where we used fact that $P = P^\dagger = P^{-1}$ and cyclicity of trace. Since $P_u P_{u'}$ is a phase times a non-identity Pauli string when $u \neq u'$, it is traceless. For $u = u'$, the trace equals $\mathrm{Tr}(I) = 2^n$. Hence the family $\{P_u P : \{P_u, P\} = 0\}$ is orthogonal in Hilbert-Schmidt inner product. Therefore

$$\|[H, P]\|_{HS}^2 = 4 \sum_{u:\{P_u,P\}=0} \lambda_u^2 \|P_u P\|_{HS}^2 = 4 \cdot 2^n \sum_{u:\{P_u,P\}=0} \lambda_u^2, \tag{49}$$

After substituting the above into Eq. (46), we have

$$C_P''(0) = -4 \sum_{u:\{P_u,P\}=0} \lambda_u^2. \tag{50}$$

□

Based on the above principle, we can extend the representation of the same-Pauli measurement signal to the entire V and give the form of a linear sum. We denote

$$x_u \coloneqq \lambda_u^2 \geq 0, \qquad \text{and} \qquad W_{\mathrm{tot}} \coloneqq \sum_{u \in V} x_u = 2^{-n} \mathrm{Tr}\big(H^2\big). \tag{51}$$

For each $v \in V$, set P_v and define

$$y(v) \coloneqq \sum_{u \in V} \mathbb{1}[\{P_u, P_v\} = 0] x_u = -\frac{1}{4} C_{P_v}''(0). \tag{52}$$

according to Proposition B.1. We also introduce the notation

$$g(v) \coloneqq W_{\mathrm{tot}} - 2y(v) = \sum_{u \in V} x_u \chi_u(v). \tag{53}$$

In this way, each probe direction is assigned a number, and the Walsh characters form the coordinate system that converts those numbers back into squared Pauli coefficients.

Proposition B.2 (Walsh inversion for results from auto-correlation function values)**.** *For every nonzero $u \in V$,*

$$x_u = 2^{-2n} \sum_{v \in V} g(v) \chi_u(v) = \frac{1}{2^{2n+1}} \sum_{v \in V} C_{P_v}''(0) \chi_u(v). \tag{54}$$

Proof. As we have shown in the main text that

$$2^{-2n} \sum_{v \in V} \chi_u(v) \chi_{u'}(v) = \mathbb{1}\{u = u'\}, \tag{55}$$

and we recall that

$$g(v) = \sum_{u' \in V} x_{u'} \chi_{u'}(v), \tag{56}$$

we obtain

$$\begin{aligned} 2^{-2n} \sum_{v \in V} g(v) \chi_u(v) &= 2^{-2n} \sum_{v \in V} \sum_{u' \in V} x_{u'} \chi_{u'}(v) \chi_u(v) \\ &= \sum_{u' \in V} x_{u'} \left(2^{-2n} \sum_{v \in V} \chi_{u'}(v) \chi_u(v) \right) \\ &= x_u. \end{aligned} \tag{57}$$

This proves the first step. For the second step, we use

$$y(v) = -\frac{1}{4} C_{P_v}''(0), \qquad g(v) = W_{\mathrm{tot}} - 2y(v) = W_{\mathrm{tot}} + \frac{1}{2} C_{P_v}''(0). \tag{58}$$

Therefore, we have

$$2^{-2n} \sum_{v \in V} g(v) \chi_u(v) = 2^{-2n} W_{\mathrm{tot}} \sum_{v \in V} \chi_u(v) + \frac{1}{2^{2n+1}} \sum_{v \in V} C_{P_v}''(0) \chi_u(v). \tag{59}$$

When $u \neq 0$, the character χ_u is nontrivial, so $\sum_{v \in V} \chi_u(v) = 0$. This gives the claimed result. □

Proposition B.2 indicates there is no information loss in the full field. However, one can never query all 4^n Pauli directions at a same time. The following theorem explains what a whole subspace of probes sees and what it misses.

Lemma B.3. *Let $K \leq V$ have dimension r, and define $\mathcal{K}(K) \coloneqq \frac{1}{2^{r+1}} \sum_{v \in K} |C''_{P_v}(0)|$ and $\mathcal{S}(K) \coloneqq \sum_{u \in K^\perp} x_u$. Then*

$$\mathcal{K}(K) = \sum_{u \notin K^\perp} x_u, \qquad \mathcal{K}(K) + \mathcal{S}(K) = W_{\mathrm{tot}}. \tag{60}$$

Proof. By Proposition B.1, each $C''_{P_v}(0)$ is non-positive, and hence $|C''_{P_v}(0)| = 4y(v)$. Using $y(v) = \sum_{u \in V} \frac{1-\chi_u(v)}{2} x_u$, we get

$$\sum_{v \in K} |C''_{P_v}(0)| = 4 \sum_{v \in K} y(v) = 2 \sum_{u \in V} x_u \left(|K| - \sum_{v \in K} \chi_u(v) \right). \tag{61}$$

By Lemma A.3, the character sum equals 2^r if $u \in K^\perp$ and equals 0 otherwise. Thus all $u \in K^\perp$ contribute zero, while each $u \notin K^\perp$ contributes $2^{r+1} x_u$. Dividing by 2^{r+1} gives

$$\mathcal{K}(K) = \sum_{u \notin K^\perp} x_u. \tag{62}$$

We then obtain the second step in the claimed result by the fact that

$$W_{\mathrm{tot}} = \sum_{u \notin K^\perp} x_u + \sum_{u \in K^\perp} x_u. \tag{63}$$

□

In short, a family of same-Pauli probes cannot see the coefficients in $K^\perp$, namely the terms commuting with all probes in the family. This is why the protocol uses complementary product bases: the Z basis reveals the X-projection of a Pauli label, while the X basis reveals the Z-projection.

B.2 Proof of Proposition 3.1

By assuming $\|H\| \leq \Lambda$, all correlation functions become band-limited. Here, we explain why a fixed kernel can combine finite discrete values to obtain a derivative at zero. We first show that every same-Pauli auto-correlation function is a bounded-frequency signal that we can recover its second-order derivative from its finite-time values.

Proposition B.4 (A spectral representation of the auto-correlation function)**.** *Assume $\|H\| \leq \Lambda$. For each $v \in V$, there exists a symmetric probability measure μ_v supported on $[-2\Lambda, 2\Lambda]$ such that*

$$C_{P_v}(t) = \int_{-2\Lambda}^{2\Lambda} \cos(\omega t) d\mu_v(\omega). \tag{64}$$

Consequently,

$$C''_{P_v}(0) = -\int_{-2\Lambda}^{2\Lambda} \omega^2 d\mu_v(\omega). \tag{65}$$

Proof. We defined the normalized Hilbert-Schmidt inner product $\langle A, B \rangle_{\mathrm{HS}} \coloneqq 2^{-n} \operatorname{Tr}\big(A^\dagger B\big)$. The commutator map $\mathcal{A}_H \coloneqq \mathrm{ad}_H$ is self-adjoint on this Hilbert space because

$$\langle A, [H, B] \rangle_{\mathrm{HS}} = 2^{-n} \operatorname{Tr}\big((A^\dagger H - H A^\dagger) B\big) = \langle [H, A], B \rangle_{\mathrm{HS}}. \tag{66}$$

As $H|e_j\rangle = E_j |e_j\rangle$, we have

$$\mathcal{A}_H(|e_j\rangle\langle e_k|) = (E_j - E_k)|e_j\rangle\langle e_k|. \tag{67}$$

Since $|E_j| \leq \Lambda$, the spectrum of $\mathcal{A}_H$ lies in $[-2\Lambda, 2\Lambda]$.

Let $E(d\omega)$ be the spectral measure of $\mathcal{A}_H$, and define

$$\mu_v(\Omega) \coloneqq \langle P_v, E(\Omega) P_v \rangle_{\mathrm{HS}}. \tag{68}$$

This is a probability measure supported on $[-2\Lambda, 2\Lambda]$. Moreover,

$$C_{P_v}(t) = \langle P_v, e^{it\mathcal{A}_H} P_v \rangle_{\mathrm{HS}} = \int_{-2\Lambda}^{2\Lambda} e^{i\omega t} d\mu_v(\omega). \tag{69}$$

The function $C_{P_v}(t)$ is real and even. We may average the frequency measure with its reflection ω and $-\omega$ and get:

$$\frac{1}{2}\big(C_{P_v}(t) + C_{P_v}(-t)\big) = C_{P_v}(t). \tag{70}$$

After this averaging, the measure is symmetric, so the sine terms cancel and only the cosine part remains:

$$C_{P_v}(t) = \int_{-2\Lambda}^{2\Lambda} \cos(\omega t) d\mu_v(\omega). \tag{71}$$

Since the measure is supported on a bounded interval, we can differentiate inside the integral, giving

$$C''_{P_v}(0) = -\int_{-2\Lambda}^{2\Lambda} \omega^2 d\mu_v(\omega). \tag{72}$$

□

In other words, we only need to choose K_Λ such that on the entire frequency range where the signal can live, and any convolution against K_Λ is equal to taking a second-order derivative at zero. To build such a kernel with good decay, we fix a smooth cutoff $\varphi \in C_c^\infty(\mathbb{R})$ satisfying

$$\varphi(\xi) = 1 \quad (|\xi| \le 1), \qquad \varphi(\xi) = 0 \quad (|\xi| \ge 2). \tag{73}$$

We set $B_\Lambda \coloneqq 2\Lambda$ and $\Phi_\Lambda(\omega) \coloneqq \omega^2 \varphi(\omega / B_\Lambda)$, and define

$$K_\Lambda(t) \coloneqq -\frac{1}{2\pi} \int_{\mathbb{R}} \Phi_\Lambda(\omega) e^{i\omega t} d\omega. \tag{74}$$

The cutoff does not change anything on the true spectral window as $\varphi(\omega/B_\Lambda) = 1$ whenever $|\omega| \le B_\Lambda = 2\Lambda$. We thus have

$$\Phi_\Lambda(\omega) = \omega^2 \qquad (|\omega| \le 2\Lambda). \tag{75}$$

Beyond this range, the $\Phi_\Lambda(\omega)$ gently fades to zero. This is just a mathematical technique to let the integral behave well. Any construct that meets the requirements will not affect the complexity of the protocol in terms of order. As a simple example, we may define:

$$\rho(s) \coloneqq \begin{cases} e^{-1/s}, & s > 0, \\ 0, & s \le 0, \end{cases} \qquad \theta(s) \coloneqq \frac{\rho(s)}{\rho(s) + \rho(1-s)}. \tag{76}$$

Then $\theta \in C^\infty(\mathbb{R})$, $\theta(s) = 0$ for $s \le 0$, and $\theta(s) = 1$ for $s \ge 1$. We then define the even cutoff:

$$\varphi(\xi) \coloneqq \theta\left(\frac{4 - \xi^2}{3}\right). \tag{77}$$

It is easy to verify that the example kernel constructed in this way meets the above requirements. The next lemma illustrates that the effect on the bounded frequency window is equivalent to taking the second-order derivative at zero.

Lemma B.5 (Even kernel for the second-order derivative of the auto-correlation function)**.** *If $f(t) = \int_{-2\Lambda}^{2\Lambda} \cos(\omega t) d\nu(\omega)$ for a finite symmetric signed measure ν, then*

$$f''(0) = \int_{\mathbb{R}} K_\Lambda(t) f(t) dt = 2 \int_0^\infty K_\Lambda(t) f(t) dt. \tag{78}$$

Proof. Since $K_\Lambda(t)$ is integrable over the entire $\mathbb{R}$ and f is bounded, we can interchange the order of inner and outer integrals according to Fubini's theorem. Using the definition of K_Λ, we have

$$\int_{\mathbb{R}} K_\Lambda(t) f(t) dt = \int_{-2\Lambda}^{2\Lambda} \left(\int_{\mathbb{R}} K_\Lambda(t) e^{i\omega t} dt \right) d\nu(\omega). \tag{79}$$

The inner transform equals $-\Phi_\Lambda(\omega)$, which is $-\omega^2$ on $[-2\Lambda, 2\Lambda]$. Hence, we have

$$\int_{\mathbb{R}} K_\Lambda(t) f(t) dt = -\int_{-2\Lambda}^{2\Lambda} \omega^2 d\nu(\omega) = f''(0). \tag{80}$$

The second equality follows because both K_Λ and f are even. □

Now we have a method to express the derivative at zero as a weighted average of the values of $C_{P_v}(t)$ at positive time points. The remaining problem is that $K_\Lambda(t)$ is a signed function. Therefore, we need to sample time from the absolute value $|K_\Lambda(t)|$, and then put the sign of $K_\Lambda(t)$ back into the estimator. Because the cutoff φ is fixed once and for all, the kernels K_Λ's are just rescalings of one specific fixed kernel. Hence, there exist absolute constants $\kappa_0, \kappa_1 > 0$, depending only on φ, such that

$$\|K_\Lambda\|_1 = 4\kappa_0 \Lambda^2, \qquad \frac{2}{\|K_\Lambda\|_1} \int_0^\infty t |K_\Lambda(t)| dt = \frac{\kappa_1}{2\Lambda}. \tag{81}$$

where $\|K_\Lambda\|_1 \coloneqq \int_{\mathbb{R}} |K_\Lambda(t)| dt$ is the function L^1 norm. The first equality controls the size of one sample. The second equality means that the average sampled evolution time is of order $1/\Lambda$.

Since K_Λ is even, we can define a probability density on $[0, \infty)$ by

$$q_\Lambda(t) \coloneqq \frac{2|K_\Lambda(t)|}{\|K_\Lambda\|_1} = \frac{|K_\Lambda(t)|}{2\kappa_0 \Lambda^2}. \tag{82}$$

Now sample $\tau \sim q_\Lambda$. For a same-Pauli shot with probe P_v and time τ, let $X_v \in \{\pm 1\}$ be the product of the prepared eigenvalue and the final measurement outcome. By Proposition 2.3, we have

$$\mathbb{E}[X_v \mid T = t] = C_{P_v}(t). \tag{83}$$

Define

$$D_v \coloneqq \|K_\Lambda\|_1 \operatorname{sign}(K_\Lambda(T)) X_v = 4\kappa_0 \Lambda^2 \operatorname{sign}(K_\Lambda(T)) X_v, \qquad Y_v \coloneqq -\frac{1}{4} D_v. \tag{84}$$

We now convert the kernel into a randomized experiment to let one same-Pauli shot give an unbiased estimate of the squared coefficient as claimed in Proposition 3.1.

Proposition B.6 (Sample for same-Pauli measurements)**.** *For every $v \in V$, we have*

$$\mathbb{E}[D_v] = C''_{P_v}(0), \qquad \mathbb{E}[Y_v] = y(v), \qquad |Y_v| \le \kappa_0 \Lambda^2, \qquad \mathbb{E}[\tau] = \frac{\kappa_1}{2\Lambda}. \tag{85}$$

Proof. Conditioning on τ and using the density q_Λ,

$$\mathbb{E}[D_v] = \int_0^\infty \|K_\Lambda\|_1 \operatorname{sign}(K_\Lambda(t)) C_{P_v}(t) \frac{2|K_\Lambda(t)|}{\|K_\Lambda\|_1} dt = 2\int_0^\infty K_\Lambda(t) C_{P_v}(t) dt = C''_{P_v}(0), \tag{86}$$

where the last equality is from Lemma B.5. Therefore, we have

$$\mathbb{E}[Y_v] = -\frac{1}{4} C''_{P_v}(0) = y(v), \qquad |Y_v| \le \frac{1}{4} \|K_\Lambda\|_1 = \kappa_0 \Lambda^2, \tag{87}$$

and the mean-time identity is exactly the definition of κ_1. □

B.3 Proof of Proposition 3.2

As discussed in Lemma B.3, we need a pair of complementary bases to fully recover all the squared coefficient information. Since Pauli labels are described by the symplectic form, a natural selection is the Z product basis and the X product basis. However, a direct search over the full Pauli label space $V = \mathbb{F}_2^{2n}$ with $|V| = 4^n$ would be too expensive. We therefore design a displacement sieve method on the Boolean hypercube $\mathbb{F}_2^n$: the same-Pauli measurement first finds which x- and which z-coordinates appear in Hamiltonian terms.

In the Z product basis, a basis vector is labeled by $a \in \mathbb{F}_2^n$. For a Pauli label $u = (x \mid z)$, the x part records which Z-basis labels are flipped by P_u. Indeed, the i-th Z-basis label is flipped exactly when P_u anticommutes with the corresponding Z_i observable, i.e., when

$$\omega(u, (0 \mid e_i)) = x_i. \tag{88}$$

Thus

$$P_{(x|z)}|a\rangle_Z \propto |a + x\rangle_Z. \tag{89}$$

So if we prepare $|a\rangle_Z$, evolve, measure again in the Z basis, and obtain $|b\rangle_Z$, then the observed displacement $d \coloneqq a + b$ points to Pauli terms with $x = d$.

We will first prove that when the time is sampled from the same kernel density q_Λ defined above, this observed displacement probability has a lower bound.

Lemma B.7 (Displacement bound)**.** *Let $\{|a\rangle_Z\}_{a\in\mathbb{F}_2^n}$ be any orthonormal basis indexed by $\mathbb{F}_2^n$, and fix a nonzero displacement $d \in \mathbb{F}_2^n$. Define*

$$f_d(t) \coloneqq 2^{-n} \sum_{a\in\mathbb{F}_2^n} \left|\langle a + d|e^{-iHt}|a\rangle_Z\right|^2. \tag{90}$$

Then f_d is even, nonnegative, and spectrally supported in $[-2\Lambda, 2\Lambda]$. Moreover, we have

$$\mathbb{E}_{\tau\sim q_\Lambda}[f_d(\tau)] \geq \frac{1}{2\kappa_0\Lambda^2} 2^{-n} \sum_{a\in\mathbb{F}_2^n} |\langle a + d|H|a\rangle_Z|^2. \tag{91}$$

Proof. Let $H = \sum_j E_j|\phi_j\rangle\langle\phi_j|$ with $|E_j| \leq \Lambda$. For each a,

$$\langle a + d|e^{-iHt}|a\rangle = \sum_j \langle a + d|\phi_j\rangle_Z\langle\phi_j|a\rangle_Z e^{-iE_jt}. \tag{92}$$

The squared modulus has frequencies $E_j - E_k \in [-2\Lambda, 2\Lambda]$. Taking the average value of a preserves this spectral support and makes it nonnegative.

We next show that f_d is even. In fact, the transition probability from a to $a + d$ at time $-t$ equals the transition probability from $a + d$ to a at time t since $U(-t) = U(t)^\dagger$:

$$\begin{aligned} f_d(-t) &= 2^{-n} \sum_a |\langle a + d|e^{iHt}|a\rangle_Z|^2 \\ &= 2^{-n} \sum_a |\langle a|e^{-iHt}|a + d\rangle_Z|^2 = 2^{-n} \sum_b |\langle b + d|e^{-iHt}|b\rangle_Z|^2 = f_d(t). \end{aligned} \tag{93}$$

Because $d \neq 0$ and $\langle a + d|a\rangle = 0$, denote $\alpha_a(t) \coloneqq \langle a + d|e^{-iHt}|a\rangle$, then $\alpha_a(0) = 0$ and $\alpha_a'(0) = -i\langle a + d|H|a\rangle$. Therefore, we have

$$\frac{d^2}{dt^2}|\alpha_a(t)|^2\Big|_{t=0} = 2|\langle a + d|H|a\rangle_Z|^2, \qquad f_d''(0) = 2^{1-n} \sum_a |\langle a + d|H|a\rangle_Z|^2. \tag{94}$$

Applying Lemma B.5 to f_d gives

$$2\int_0^\infty K_\Lambda(t) f_d(t) dt = f_d''(0). \tag{95}$$

Since $f_d(t) \geq 0$, we have

$$\int_0^\infty |K_\Lambda(t)| f_d(t) dt \geq \left|\int_0^\infty K_\Lambda(t) f_d(t) dt\right| = \frac{1}{2} f_d''(0). \tag{96}$$

Using $q_\Lambda(t) = 2|K_\Lambda(t)|/\|K_\Lambda\|_1$ and $\|K_\Lambda\|_1 = 4\kappa_0\Lambda^2$ leads to the claim. □

For $u = (x \mid z)$, write $x(u) = x$ and $z(u) = z$. We define the projection intensities

$$W_x(d) \coloneqq \sum_{z \in \mathbb{F}_2^n} \lambda^2_{(d|z)}, \qquad W_z(e) \coloneqq \sum_{x \in \mathbb{F}_2^n} \lambda^2_{(x|e)}. \tag{97}$$

A Z-basis displacement shot samples a uniformly random computational basis state $|A\rangle_Z$, evolves for $\tau \sim q_\Lambda$, measures every qubit in the Z basis, obtains $B \in \mathbb{F}_2^n$, and records $D = A + B$. We now prove the following two propositions, which together prove Proposition 3.2.

Proposition B.8 (Z-basis displacements)**.** *For every nonzero $d \in \mathbb{F}_2^n$, we have*

$$\Pr[D = d] \geq \frac{W_x(d)}{2\kappa_0 \Lambda^2}. \tag{98}$$

Proof. For a fixed t, we have

$$\Pr[D = d \mid T = t] = 2^{-n} \sum_{a \in \mathbb{F}_2^n} |\langle a + d|e^{-iHt}|a\rangle_Z|^2. \tag{99}$$

Lemma B.7 gives

$$\Pr[D = d] \geq \frac{1}{2\kappa_0 \Lambda^2} 2^{-n} \sum_a |\langle a + d|H|a\rangle_Z|^2. \tag{100}$$

For the canonical Pauli representative, there is a unit phase $\zeta_Z(d, z)$, independent of a, such that

$$P_{(d|z)}|a\rangle_Z = \zeta_Z(d, z)(-1)^{z \cdot a}|a + d\rangle_Z. \tag{101}$$

Thus

$$\langle a + d|H|a\rangle_Z = \sum_{z \in \mathbb{F}_2^n} \lambda_{(d|z)} \zeta_Z(d, z)(-1)^{z \cdot a}. \tag{102}$$

Averaging the squared modulus over uniform a and using Walsh orthogonality on $\mathbb{F}_2^n$ gives

$$2^{-n} \sum_a |\langle a + d|H|a\rangle_Z|^2 = \sum_{z \in \mathbb{F}_2^n} \lambda^2_{(d|z)} = W_x(d). \tag{103}$$

Substitution proves the proposition. □

The experiment based on the X basis is the same as that based on the Z basis, except that the roles of x and z are reversed. The Pauli Z component flips the eigenstates of the X basis, so the observed displacement is a z projection.

Proposition B.9 (X-basis displacements)**.** *For every nonzero $e \in \mathbb{F}_2^n$, we have*

$$\Pr[E = e] \geq \frac{W_z(e)}{2\kappa_0 \Lambda^2}. \tag{104}$$

Proof. Define the X-basis displacement shot exactly as above, but prepare and measure in the X basis and record $E = A + B$. Lemma B.7 gives

$$\Pr[E = e] \geq \frac{1}{2\kappa_0 \Lambda^2} 2^{-n} \sum_a |\langle a + e|H|a\rangle_X|^2. \tag{105}$$

For each fixed $(x \mid e)$ there is a unit phase $\zeta_X(x, e)$, independent of a, such that

$$P_{(x|e)}|a\rangle_X = \zeta_X(x, e)(-1)^{x \cdot a}|a + e\rangle_X. \tag{106}$$

Therefore, we have

$$\langle a + e|H|a\rangle_X = \sum_{x \in \mathbb{F}_2^n} \lambda_{(x|e)} \zeta_X(x, e)(-1)^{x \cdot a}. \tag{107}$$

Walsh orthogonality over a gives

$$2^{-n} \sum_a |\langle a + e|H|a\rangle_X|^2 = \sum_{x \in \mathbb{F}_2^n} \lambda^2_{(x|e)} = W_z(e), \tag{108}$$

which proves the claim. □

B.4 Proof of Lemma 3.3

According to the definitions, $W_x(d) = \sum_z \lambda^2_{(d|z)}$ is the total squared coefficient of all Pauli terms whose x-part is d, and $W_z(e)$ is the total squared coefficient of all Pauli terms whose z-part is e. We now show that all these sums together are bounded by Λ^2, so there can be only a few coordinates d or e whose sum is at least ε^2.

Proposition B.10 (Projection counts)**.** *One has*

$$\sum_{d\in\mathbb{F}_2^n} W_x(d) = \sum_{e\in\mathbb{F}_2^n} W_z(e) = W_{\mathrm{tot}} \le \Lambda^2. \tag{109}$$

Consequently, for every target detection scale $\varepsilon > 0$, we have

$$|\Pi_x^{(\varepsilon)} \setminus \{0\}| \le \frac{\Lambda^2}{\varepsilon^2}, \qquad |\Pi_z^{(\varepsilon)} \setminus \{0\}| \le \frac{\Lambda^2}{\varepsilon^2}. \tag{110}$$

Proof. If we sum $W_x(d)$ over all $d \in \mathbb{F}_2^n$, every squared Pauli coefficient is counted exactly once:

$$\sum_{d\in\mathbb{F}_2^n} W_x(d) = \sum_{d\in\mathbb{F}_2^n}\sum_{z\in\mathbb{F}_2^n} \lambda^2_{(d|z)} = \sum_{u\in V} \lambda_u^2 = W_{\mathrm{tot}}. \tag{111}$$

Similarly, we have $\sum_{e\in\mathbb{F}_2^n} W_z(e) = W_{\mathrm{tot}}$. Recall that $W_{\mathrm{tot}} = 2^{-n}\operatorname{Tr}\big(H^2\big)$. We have $W_{\mathrm{tot}} = 2^{-n}\operatorname{Tr}\big(H^2\big) \le \|H\|^2 \le \Lambda^2$.

Now let $A_x \coloneqq \Pi_x^{(\varepsilon)} \setminus \{0\}$. For every $d \in A_x$, the definition of $\Pi_x^{(\varepsilon)}$ gives $W_x(d) \ge \varepsilon^2$. Hence

$$|A_x|\varepsilon^2 \le \sum_{d\in A_x} W_x(d) \le \sum_{d\in\mathbb{F}_2^n} W_x(d) = W_{\mathrm{tot}} \le \Lambda^2. \tag{112}$$

Dividing by ε^2 gives $|\Pi_x^{(\varepsilon)} \setminus \{0\}| \le \frac{\Lambda^2}{\varepsilon^2}$. Similarly, we have $|\Pi_z^{(\varepsilon)} \setminus \{0\}| \le \frac{\Lambda^2}{\varepsilon^2}$. □

Now, we are ready to prove Lemma 3.3, which we rewrite as the following for convenience.

Lemma B.11 (Displacement sieve)**.** *There is an absolute constant $C > 0$ that satisfies the following conditions. For any $0 < \varepsilon \le \Lambda$ and $\eta \in (0, 1)$, there is a protocol that outputs sets $\widehat{\Pi}_x, \widehat{\Pi}_z \subseteq \mathbb{F}_2^n$ with probability at least $1 - \eta$ and size $O(\frac{\Lambda^2}{\varepsilon^2}\log\frac{4}{\eta})$ within deterministic total evolution time at most*

$$C\frac{\Lambda}{\varepsilon^2}\log\frac{4\Lambda}{\varepsilon}\log\frac{4}{\eta}. \tag{113}$$

Proof. We will complete the proof based on repeated sampling. We first show that every coordinate d with $W_x(d) \ge \varepsilon^2$ appears often enough in the Z-basis displacement experiment. The same argument will then apply to the X basis.

By Proposition B.8, every nonzero $d \in \Pi_x^{(\varepsilon)}$ appears in one Z-basis displacement shot with probability at least $\frac{\varepsilon^2}{2\kappa_0\Lambda^2}$. Similarly, by Proposition B.9, every nonzero $e \in \Pi_z^{(\varepsilon)}$ appears in one X-basis displacement shot with the same lower bound. Set $p \coloneqq \frac{\varepsilon^2}{4\kappa_0\Lambda^2}$, then every nonzero coordinate in $\Pi_x^{(\varepsilon)}$ or $\Pi_z^{(\varepsilon)}$ has appearance probability at least $2p$ in the corresponding basis.

Next, Proposition B.10 gives

$$|\Pi_x^{(\varepsilon)} \setminus \{0\}| \le \frac{\Lambda^2}{\varepsilon^2}, \qquad |\Pi_z^{(\varepsilon)} \setminus \{0\}| \le \frac{\Lambda^2}{\varepsilon^2}. \tag{114}$$

We let

$$r \coloneqq \left\lceil \frac{\Lambda^2}{\varepsilon^2} \right\rceil, \qquad N \coloneqq \left\lceil C_0 p^{-1}\log(4r) \right\rceil, \qquad L \coloneqq \left\lceil C_1 \log\frac{4}{\eta} \right\rceil, \tag{115}$$

where $C_0, C_1 > 0$ are sufficiently large absolute constants.

We now describe one round of experiments in the Z basis. First sample N times $t_1, \dots, t_N \sim q_\Lambda$ independently, and let $\Theta \coloneqq \sum_{j=1}^{N} t_j$ represent the total evolution time planned for this round. To prevent the random sampling time from becoming too long, we need to set an upper limit on the total evolution time. A reasonable choice is 8 times the average total time $\tau_{\max} \coloneqq 4\kappa_1 \frac{N}{\Lambda}$. If $\Theta > \tau_{\max}$, the round is aborted before any quantum evolution and outputs the empty set. If $\Theta \le \tau_{\max}$, the round performs the N Z-basis displacement shots with time $t_1, \dots, t_N$. For every nonzero displacement d, let A_d be the number of times d appears. This round outputs all nonzero d such that $A_d \ge Np$. The X-basis round is defined in the same way. The final sets $\widehat{\Pi}_x$ and $\widehat{\Pi}_z$ are the unions of the outputs of L independent rounds in the corresponding basis.

Now fix one Z-basis round and one coordinate $d \in \Pi_x^{(\varepsilon)} \setminus \{0\}$. Since d appears in each shot with probability at least $2p$, the count A_d has a mean of at least $2Np$. The Chernoff bound gives

$$\Pr[A_d < Np] \le \exp(-Np/4). \tag{116}$$

As there are at most r nonzero coordinates d with $W_x(d) \ge \varepsilon^2$, we have

$$\Pr\left[\text{some } d \in \Pi_x^{(\varepsilon)} \setminus \{0\} \text{ has } A_d < Np\right] \le r \exp(-Np/4). \tag{117}$$

By choosing C_0 large enough, this probability is at most $1/16$. Hence, with probability at least $15/16$, the round records every element of $\Pi_x^{(\varepsilon)} \setminus \{0\}$, provided it is not aborted.

It remains to bound the probability of abortion. By Proposition B.6, a sampled time length from q_Λ has the mean value $\frac{\kappa_1}{2\Lambda}$. Hence, we have

$$\mathbb{E}[\Theta] = N \frac{\kappa_1}{2\Lambda}. \tag{118}$$

By Markov's inequality,

$$\Pr[\Theta > \tau] \le \frac{N\kappa_1/(2\Lambda)}{4\kappa_1 N/\Lambda} = \frac{1}{8}. \tag{119}$$

Then one Z-basis round succeeds, meaning that it is not aborted, and it records every element of $\Pi_x^{(\varepsilon)} \setminus \{0\}$, with probability at least $1 - \frac{1}{16} - \frac{1}{8} = \frac{13}{16}$. The same proof applies to one X-basis round and the set $\Pi_z^{(\varepsilon)} \setminus \{0\}$.

Since the rounds are independent, the probability that all L Z-basis rounds fail is at most $\left(\frac{3}{16}\right)^L$. By choosing C_1 large enough, we have $\left(\frac{3}{16}\right)^L \le \frac{\eta}{2}$. The same bound holds for the X basis. A union bound over the two bases gives

$$\Pi_x^{(\varepsilon)} \setminus \{0\} \subseteq \widehat{\Pi}_x, \qquad \Pi_z^{(\varepsilon)} \setminus \{0\} \subseteq \widehat{\Pi}_z \tag{120}$$

with probability at least $1 - \eta$.

We next bound the output size. In any non-aborted round, every recorded coordinate appears at least Np times, while the total number of observations is exactly N. Hence, one round can record at most $\frac{N}{Np} = \frac{1}{p}$ coordinates. There are L rounds in the Z basis and L rounds in the X basis, so

$$|\widehat{\Pi}_x| + |\widehat{\Pi}_z| \le \frac{2L}{p} \le C \frac{\Lambda^2}{\varepsilon^2} \log \frac{4}{\eta}. \tag{121}$$

Finally, we bound the deterministic total evolution time. Each round uses at most τ evolution time. Therefore, the total time over both bases are at most

$$2L\tau = 8\kappa_1 L \frac{N}{\Lambda}. \tag{122}$$

Using $p^{-1} = 4\kappa_0 \frac{\Lambda^2}{\varepsilon^2}$ and $\log(4r) = O(\log \frac{4\Lambda}{\varepsilon})$, we get

$$2L\tau \le C \frac{\Lambda}{\varepsilon^2} \log \frac{4\Lambda}{\varepsilon} \log \frac{4}{\eta} \tag{123}$$

for some large enough constant C as claimed. □

Assuming that we succeed in Lemma B.11, the Cartesian candidate set of the output

$$\widehat{\mathcal{U}}_{\text{proj}} \coloneqq (\widehat{\Pi}_x \cup \{0\}) \times (\widehat{\Pi}_z \cup \{0\}) \setminus \{0\} \tag{124}$$

contains every $u = (x \mid z)$ with $|\lambda_u| \ge \varepsilon$: if $x \ne 0$, then $W_x(x) \ge \lambda_u^2 \ge \varepsilon^2$, and if $x = 0$ then x belongs to $\widehat{\Pi}_x \cup \{0\}$. The same applies to z.

C Cross-Pauli Measurement

C.1 Frame and parity blocks

The same-Pauli measurement finds where the coefficients with magnitude at least ε are. In this section, we use the cross-Pauli measurement to estimate their signs and magnitudes. The idea is to choose, on every qubit, one local direction c_i whose coefficient letter we want to detect, and use the two remaining directions as prepare and measure bases. Finally, we use the input and output to construct parity variables, making them equal to the first-order derivative of the cross-correlation function in Eq. (6).

Definition C.1 (Target Pauli basis)**.** *A target Pauli basis is a Pauli string $c = (c_1, \dots, c_n) \in \{X, Y, Z\}^n$. For each qubit i, define (a_i, b_i) by the cyclic convention*

$$(c_i, a_i, b_i) \in \{(X, Y, Z), (Y, Z, X), (Z, X, Y)\}. \tag{125}$$

Thus $a_i b_i = i c_i$, $b_i c_i = i a_i$, and $c_i a_i = i b_i$. For subsets $A, B \subseteq [n]$, define

$$\begin{aligned} Q_A &\coloneqq \bigotimes_{i=1}^{n} Q_i^{(A)}, \qquad Q_i^{(A)} \coloneqq \begin{cases} a_i, & i \in A, \\ I, & i \notin A, \end{cases} \\ P_B &\coloneqq \bigotimes_{i=1}^{n} P_i^{(B)}, \qquad P_i^{(B)} \coloneqq \begin{cases} b_i, & i \in B, \\ I, & i \notin B. \end{cases} \end{aligned} \tag{126}$$

The label $u_c(A, B) \in V$ is defined sitewise by

$$(u_c(A, B))_i = \begin{cases} I, & i \notin A \cup B, \\ a_i, & i \in A \setminus B, \\ b_i, & i \in B \setminus A, \\ c_i, & i \in A \cap B. \end{cases} \tag{127}$$

The visible set of the target Pauli basis is

$$\mathcal{V}(c) \coloneqq \{u \in V \setminus \{0\} : |\{i : u_i = c_i\}| \text{ is odd}\}. \tag{128}$$

For every $u \in \mathcal{V}(c)$, there is a unique pair $(A_c(u), B_c(u))$ such that $u = u_c(A_c(u), B_c(u))$. In fact, the set A records where the input basis is used, and the set B records where the output basis is used. When A and B overlap, the local product of the input and output directions leaves exactly the missing direction c_i.

Definition C.2 (Parity Blocks)**.** *Fix a target Pauli basis c. For $A, B \subseteq [n]$, set*

$$F_{A,B}^{(c)}(t) \coloneqq 2^{-n} \operatorname{Tr}(P_B(t) Q_A). \tag{129}$$

A target Pauli basis shot is performed as follows. First sample

$$\tau \sim p_\Lambda, \qquad S \sim \mathrm{Unif}\{\pm 1\}, \tag{130}$$

independently, where p_Λ is defined in Eq. (11) *and again below in Lemma C.5. If $S = +1$, prepare independent random eigenstates of the a_i bases, with eigenvalue signs $s_i \in \{\pm 1\}$, evolve for time τ, and measure every qubit in the b_i basis, obtaining signs m_i. If $S = -1$, swap the two bases: prepare b_i eigenstates with signs r_i, evolve for time τ, and measure in the a_i basis, obtaining signs n_i.*

For $A, B \subseteq [n]$, define

$$s_A \coloneqq \prod_{i \in A} s_i, \quad m_B \coloneqq \prod_{i \in B} m_i, \quad r_B \coloneqq \prod_{i \in B} r_i, \quad n_A \coloneqq \prod_{i \in A} n_i, \tag{131}$$

and the parity variable

$$Z_{A,B}^{(c)} \coloneqq \begin{cases} s_A m_B, & S = +1, \\ -r_B n_A, & S = -1. \end{cases} \tag{132}$$

We now prove the following two propositions.

Proposition C.3 (Extracts the odd part of cross-correlation functions)**.** *For every target Pauli basis c, subsets $A, B \subseteq [n]$, and time $t \geq 0$,*

$$\mathbb{E}[Z^{(c)}_{A,B} \mid \tau = t] = \frac{1}{2}\Big(F^{(c)}_{A,B}(t) - F^{(c)}_{A,B}(-t)\Big). \tag{133}$$

Proof. If $S = +1$, the random local eigenvalue signs define a uniformly random eigenvector of Q_A with eigenvalue s_A. By Proposition 2.3,

$$\mathbb{E}[s_A m_B \mid T = t, S = +1] = 2^{-n} \operatorname{Tr}(P_B(t)Q_A) = F^{(c)}_{A,B}(t). \tag{134}$$

If $S = -1$, the same trace rule, now with preparation observable P_B and measurement observable Q_A, gives

$$\mathbb{E}[-r_B n_A \mid T = t, S = -1] = -2^{-n} \operatorname{Tr}(Q_A(t)P_B). \tag{135}$$

By the cyclicity of trace,

$$2^{-n} \operatorname{Tr}(Q_A(t)P_B) = 2^{-n} \operatorname{Tr}(P_B(-t)Q_A) = F^{(c)}_{A,B}(-t). \tag{136}$$

Averaging the two equally likely signs $S = \pm 1$ proves the claimed result. □

Proposition C.4 (A spectral representation of the cross-correlation function)**.** *Assume $\|H\| \leq \Lambda$. For every Hermitian Pauli pair (P, Q), there exists a finite complex measure $\nu_{P,Q}$ supported on $[-2\Lambda, 2\Lambda]$ such that*

$$F_{P,Q}(t) = \int_{-2\Lambda}^{2\Lambda} e^{i\omega t} d\nu_{P,Q}(\omega), \qquad |\nu_{P,Q}|([-2\Lambda, 2\Lambda]) \leq 1. \tag{137}$$

Consequently,

$$F'_{P,Q}(0) = \int_{-2\Lambda}^{2\Lambda} i\omega d\nu_{P,Q}(\omega). \tag{138}$$

Proof. As in Proposition B.4, the superoperator $\mathcal{A}_H := \operatorname{ad}_H$ is self-adjoint on the normalized Hilbert-Schmidt space, i.e., given $H|e_j\rangle = E_j|e_j\rangle$ with $|E_j| \leq \Lambda$, then

$$\mathcal{A}_H(|e_j\rangle\langle e_k|) = (E_j - E_k)|e_j\rangle\langle e_k|. \tag{139}$$

Thus the spectrum of $\mathcal{A}_H$ is contained in $[-2\Lambda, 2\Lambda]$. Let $E(d\omega)$ be its spectral measure. Define

$$\nu_{P,Q}(\Omega) := \langle Q, E(\Omega)P\rangle_{\mathrm{HS}}.$$

Then, we have

$$F_{P,Q}(t) = \langle Q, e^{it\mathcal{A}_H} P\rangle_{\mathrm{HS}} = \int_{-2\Lambda}^{2\Lambda} e^{i\omega t}\, d\nu_{P,Q}(\omega). \tag{140}$$

To bound the total variation, let $\{\Omega_j\}$ be any finite measurable partition of $[-2\Lambda, 2\Lambda]$. Since the ranges of the spectral projections $E(\Omega_j)$ are orthogonal, we have

$$\begin{aligned}
\sum_j |\nu_{P,Q}(\Omega_j)| &= \sum_j \big|\langle Q, E(\Omega_j)P\rangle_{\mathrm{HS}}\big| \\
&= \sum_j \big|\langle E(\Omega_j)Q, E(\Omega_j)P\rangle_{\mathrm{HS}}\big| \\
&\leq \left(\sum_j \|E(\Omega_j)Q\|^2_{\mathrm{HS}}\right)^{1/2} \left(\sum_j \|E(\Omega_j)P\|^2_{\mathrm{HS}}\right)^{1/2} \\
&= \|Q\|_{\mathrm{HS}}\|P\|_{\mathrm{HS}} = 1.
\end{aligned} \tag{141}$$

Taking the supremum over all finite partitions gives

$$|\nu_{P,Q}|([-2\Lambda, 2\Lambda]) \leq 1. \tag{142}$$

Differentiation under the integral is allowed because the support is bounded. After differentiation, we obtain the claimed result. □

Fix the same smooth cutoff φ as in the even-kernel construction and set

$$\Psi_\Lambda(\omega) \coloneqq i\omega\varphi(\omega/(2\Lambda)). \tag{143}$$

Now we can define the odd kernel

$$L_\Lambda(t) \coloneqq \frac{1}{2\pi}\int_{\mathbb{R}} \Psi_\Lambda(\omega)e^{-i\omega t}d\omega. \tag{144}$$

The odd kernel above satisfies the following lemma.

Lemma C.5 (Odd kernel for the first-order derivative of the cross-correlation function)**.** *If $f(t) = \int_{-2\Lambda}^{2\Lambda} e^{i\omega t}d\nu(\omega)$ for a finite complex measure ν, then*

$$f'(0) = \int_{\mathbb{R}} L_\Lambda(t)f(t)dt = \int_0^\infty L_\Lambda(t)\big(f(t) - f(-t)\big)dt. \tag{145}$$

Proof. Because $L_\Lambda \in L^1(\mathbb{R})$ and $|f(t)| \le |\nu|([-2\Lambda, 2\Lambda])$, we apply Fubini's theorem:

$$\int_{\mathbb{R}} L_\Lambda(t)f(t)dt = \int_{-2\Lambda}^{2\Lambda}\left(\int_{\mathbb{R}} L_\Lambda(t)e^{i\omega t}dt\right)d\nu(\omega). \tag{146}$$

The inner Fourier transform equals $\Psi_\Lambda(\omega)$, and on $[-2\Lambda, 2\Lambda]$ we have $\varphi(\omega/(2\Lambda)) = 1$, hence $\Psi_\Lambda(\omega) = i\omega$. Therefore, we have

$$\int_{\mathbb{R}} L_\Lambda(t)f(t)dt = \int_{-2\Lambda}^{2\Lambda} i\omega d\nu(\omega) = f'(0). \tag{147}$$

Since L_Λ is odd, we have

$$\int_{\mathbb{R}} L_\Lambda(t)f(t)dt = \int_0^\infty L_\Lambda(t)\big(f(t) - f(-t)\big)dt. \tag{148}$$

□

As explained earlier, the kernel we will use is just a rescaling of the fixed kernel. So there exist absolute constants $\ell_0, \ell_1 > 0$, depending only on the cutoff, namely

$$\|L_\Lambda\|_1 = 2\ell_0\Lambda, \qquad \frac{2}{\|L_\Lambda\|_1}\int_0^\infty t|L_\Lambda(t)|dt = \frac{\ell_1}{2\Lambda}. \tag{149}$$

Now we define the odd sampling density as

$$p_\Lambda(t) \coloneqq \frac{2|L_\Lambda(t)|}{\|L_\Lambda\|_1}, \qquad t \ge 0. \tag{150}$$

Similar to the same-Pauli measurement, we define the weighted parity sample value as

$$G^{(c)}_{A,B} \coloneqq \|L_\Lambda\|_1 \operatorname{sign}(L_\Lambda(\tau))Z^{(c)}_{A,B}. \tag{151}$$

C.2 Proof of Proposition 3.4

We first prove that the sampled parity block has an expectation equal to the first-order derivative at the initial time. By Proposition C.3 and Lemma C.5, we have

$$\begin{aligned}\mathbb{E}\big[G^{(c)}_{A,B}\big] &= \int_0^\infty \|L_\Lambda\|_1 \operatorname{sign}(L_\Lambda(t)) \cdot \frac{1}{2}\big(F^{(c)}_{A,B}(t) - F^{(c)}_{A,B}(-t)\big) \cdot \frac{2|L_\Lambda(t)|}{\|L_\Lambda\|_1}dt \\ &= \int_0^\infty L_\Lambda(t)\big(F^{(c)}_{A,B}(t) - F^{(c)}_{A,B}(-t)\big)dt \\ &= \big(F^{(c)}_{A,B}\big)'(0).\end{aligned} \tag{152}$$

It remains to compute this derivative. Since $P'_B(0) = i[H, P_B]$, we have

$$\big(F^{(c)}_{A,B}\big)'(0) = \frac{i}{2^n}\sum_{u\in V}\lambda_u \operatorname{Tr}([P_u, P_B]Q_A). \tag{153}$$

If P_u commutes with P_B, the term contributes nothing. If it anticommutes with P_B, then $[P_u, P_B] = 2P_uP_B$. Therefore, the only possible nonzero trace is controlled by $\mathrm{Tr}(P_uP_BQ_A)$. Looking qubit-by-qubit, as we have

$$(P_u)_i = \begin{cases} I, & i \notin A \cup B, \\ a_i, & i \in A \setminus B, \\ b_i, & i \in B \setminus A, \\ c_i, & i \in A \cap B. \end{cases} \tag{154}$$

Hence, at most one label can contribute, namely $u = u_c(A, B)$. For this label, P_u and P_B anticommute exactly on the index $i \in A \cap B$, for $(P_u)_i = c_i$, and $(P_B)_i = b_i$. Therefore, $[P_u, P_B] \neq 0$ means $k = |A \cap B|$ is odd. If k is even, the derivative is zero. If k is odd, then

$$\big(F^{(c)}_{A,B}\big)'(0) = \frac{2i}{2^n}\lambda_{u_c(A,B)}\,\mathrm{Tr}\big(P_{u_c(A,B)}P_BQ_A\big). \tag{155}$$

It remains only to evaluate the phase. For $u = u_c(A, B)$, we have

$$(P_u)_i(P_B)_i(Q_A)_i = \begin{cases} I, & i \notin A \cup B, \\ a_i^2 = I, & i \in A \setminus B, \\ b_i^2 = I, & i \in B \setminus A, \\ c_ib_ia_i = -iI, & i \in A \cap B. \end{cases} \tag{156}$$

This indicates that

$$P_{u_c(A,B)}P_BQ_A = (-i)^kI, \qquad \mathrm{Tr}\big(P_{u_c(A,B)}P_BQ_A\big) = 2^n(-i)^k. \tag{157}$$

We can thus compute the first-order derivative at zero time as

$$\big(F^{(c)}_{A,B}\big)'(0) = 2i(-i)^k\lambda_{u_c(A,B)} = 2\sigma_c(A, B)\lambda_{u_c(A,B)}. \tag{158}$$

Combining the definition of $G^{(c)}_{A,B}$ and the identity $\|L_\Lambda\|_1 = 2\ell_0\Lambda$, we finally get $|G^{(c)}_{A,B}| \leq 2\ell_0\Lambda$ as $|Z^{(c)}_{A,B}| \leq 1$. For a visible label u, the block with $(A, B) = (A_c(u), B_c(u))$ has expectation $\mathbb{E}\big[G^{(c)}_{A,B}\big] = 2\sigma_c(A, B)\lambda_u$. Since $\sigma_c(A, B) \in \{\pm 1\}$. Multiplying this block by $\sigma_c(A, B)/2$ removes the known sign and the factor 2, leaving an unbiased sample of λ_u. Specifically, for $u \in \mathcal{V}(c)$, we define

$$Y^{(c)}_u \coloneqq \frac{1}{2}\sigma_c(A_c(u), B_c(u))G^{(c)}_{A_c(u),B_c(u)}. \tag{159}$$

Then Proposition 3.4 gives

$$\mathbb{E}[Y^{(c)}_u] = \lambda_u, \qquad |Y^{(c)}_u| \leq \ell_0\Lambda. \tag{160}$$

Therefore, proposition 3.4 is proved.

C.3 Visibility correction

For a fixed target Pauli basis c, the block we constructed above only gives a useful sample for labels in $\mathcal{V}(c)$. Concretely, this means the label has an odd number of indices where its Pauli letter is the target Pauli basis letter c_i. In this case, the corrected block $Y^{(c)}_u$ has mean λ_u. But if the number is even, Proposition 3.4 has shown that the corresponding first-order signal is zero, so this target Pauli basis is blind to that label. To address this issue, we choose the target Pauli basis at random. We need to make sure that every nonzero label becomes visible with a known probability $q(u)$. To achieve this, we rescale the sample by $q(u)^{-1}$ on the shots where it is visible. This will compensate for the shots where the label is invisible, so the averaged sample still outputs λ_u.

Proposition C.6 (Visibility probability of a random target Pauli basis)**.** *Let $c \in \{X, Y, Z\}^n$ be a uniformly random chosen Pauli string with no identity site. For every fixed nonzero label $u \in V \setminus \{0\}$, we have*

$$\Pr[u \in \mathcal{V}(c)] = \frac{1 - (1/3)^{w(u)}}{2} \geq \frac{1}{3}. \tag{161}$$

where $w(u)$ is the Pauli weight of u, i.e. the number of qubits on which $u_i \neq I$. Then for independent random target Pauli bases $c^{(1)}, \ldots, c^{(R)}$, we have

$$\Pr\big[u \notin \mathcal{V}(c^{(r)}) \textit{ for all } r = 1, \ldots, R\big] \leq \left(\frac{2}{3}\right)^R. \tag{162}$$

Proof. On every qubit where $u_i \neq I$, the random target Pauli basis letter satisfies $c_i = u_i$ with probability $1/3$. Therefore, the number of matches is

$$X \sim \mathrm{Bin}(w(u), 1/3). \tag{163}$$

The label is visible exactly when X is odd. For a binomial random variable X with success probability p, we have

$$\Pr[X \text{ is odd}] = \frac{1 - (1-2p)^{w(u)}}{2}. \tag{164}$$

Taking $p = 1/3$ gives

$$\Pr[u \in \mathcal{V}(c)] = \frac{1 - (1/3)^{w(u)}}{2}. \tag{165}$$

Since $u \neq 0$, we have $w(u) \geq 1$, and the probability is at least $1/3$. The claimed result can be derived from the independence of every single site. □

The method for correcting for missed target Pauli bases is simple. For a variable u, each observation has a probability of $q(u)$ of seeing it. Therefore, we only look at the data that actually captures u, and then scale the calculated value up by $q(u)^{-1}$. We define this explicitly below.

Definition C.7 (Visibility correction)**.** *Fix a nonzero label $u \in V \setminus \{0\}$ and let $C \sim \mathrm{Unif}\{X, Y, Z\}^n$ be an independent random target Pauli basis. The random variable*

$$\widetilde{Y}_u \coloneqq q(u)^{-1} \mathbb{1}\{u \in \mathcal{V}(C)\} Y_u^{(C)} \tag{166}$$

is called the visibility-corrected sample for the label u.

The following proposition will demonstrate that even if we only receive a sample on the visible target Pauli bases, the average value obtained after visibility correction will equal the true coefficient.

Proposition C.8 (Visibility correction preserves the mean value of the coefficient)**.** *For every nonzero label $u \in V \setminus \{0\}$, we have $\mathbb{E}[\widetilde{Y}_u] = \lambda_u$ and $|\widetilde{Y}_u| \leq 3\ell_0\Lambda$.*

Proof. By Proposition C.6, $q(u) > 0$. Using the identity(160), we have

$$\mathbb{E}[\widetilde{Y}_u] = q(u)^{-1} \sum_{c: u \in \mathcal{V}(c)} \Pr[C = c] \mathbb{E}[Y_u^{(c)}] = q(u)^{-1} \sum_{c: u \in \mathcal{V}(c)} \Pr[C = c] \lambda_u = \lambda_u. \tag{167}$$

Moreover, as $q(u) \geq 1/3$, we have

$$|\widetilde{Y}_u| \leq q(u)^{-1} \ell_0 \Lambda \leq 3\ell_0\Lambda. \tag{168}$$

□

C.4 Proof of Lemma 3.5

Previously, we only discussed the case of one unbiased sampling. However, similar to the situation in the same-Pauli measurement, each sample may take an unreasonably long time. To prevent the program from running indefinitely, we break the computation task into several blocks and also set an abortion rule. We then take the median of the results from all blocks, which yields a higher accuracy and success rate.

Definition C.9 (One estimation block)**.** *Fix a candidate set $\mathcal{U} \subseteq V \setminus \{0\}$. If $\mathcal{U} = \varnothing$, the estimator returns no values and uses zero evolution time. Therefore, we assume that $|\mathcal{U}| \geq 1$. For target accuracy $\varepsilon > 0$, we set*

$$N \coloneqq \left\lceil C_0 \frac{\Lambda^2}{\varepsilon^2} \log\bigl(16|\mathcal{U}|\bigr) \right\rceil, \tag{169}$$

where $C_0 > 0$ is a sufficiently large absolute constant.

One estimation block consists of N independent planned target Pauli basis shots. For shot s, we sample

$$C_s \sim \mathrm{Unif}\{X, Y, Z\}^n, \qquad T_s \sim p_\Lambda, \tag{170}$$

independently, let $\Theta \coloneqq \sum_{s=1}^N T_s$ be the planned total time, and set the deterministic limit 8 times the expected value of the planned total time as $\tau \coloneqq 4\ell_1 \frac{N}{\Lambda}$. If $\Theta > \tau$, we abort the block before any quantum evolution and return nothing on $\mathcal{U}$. Otherwise, we execute the planned shots. For every $u \in \mathcal{U}$, we define

$$W_s(u) \coloneqq q(u)^{-1} \mathbb{1}\{u \in \mathcal{V}(C_s)\} Y_u^{(C_s)}, \qquad s = 1, \ldots, N. \tag{171}$$

The block then outputs

$$\widetilde{\lambda}(u) \coloneqq \frac{1}{N} \sum_{s=1}^N W_s(u), \qquad u \in \mathcal{U}. \tag{172}$$

We are now ready to prove Lemma 3.5, which is restated below for convenience.

Lemma C.10 (Coefficient estimation on a candidate set)**.** *There exists an absolute constant $C > 0$ such that the following holds. Let $\mathcal{U} \subseteq V \setminus \{0\}$ be a candidate set, let $0 < \varepsilon \leq \Lambda$, and let $\eta \in (0, 1)$. If $\mathcal{U} = \varnothing$, the protocol returns no value and uses zero time. If $|\mathcal{U}| \geq 1$, run $R \coloneqq \left\lceil 16 \log \frac{2}{\eta} \right\rceil$ independent estimation blocks. If a block is aborted, assign it the value 0 for every $u \in \mathcal{U}$. For each $u \in \mathcal{U}$, let $\widehat{\lambda}_{\mathcal{U}}(u)$ be the median of the R reported values. Then, with probability at least $1 - \eta$,*

$$\max_{u \in \mathcal{U}} |\widehat{\lambda}_{\mathcal{U}}(u) - \lambda_u| \leq \varepsilon. \tag{173}$$

Moreover, the total evolution time is at most

$$C \frac{\Lambda}{\varepsilon^2} \log\bigl(4|\mathcal{U}|\bigr) \log \frac{2}{\eta} \tag{174}$$

when $|\mathcal{U}| \geq 1$, and is zero when $\mathcal{U} = \varnothing$.

Proof. The theorem clearly holds for the case of an empty set, so we consider the case where $|\mathcal{U}| \geq 1$. Fix $u \in \mathcal{U}$, the single-shot variable $W_s(u)$ is exactly the visibility-corrected sample from Definition C.7. Proposition C.8 gives

$$\mathbb{E}[W_s(u)] = \lambda_u, \qquad |W_s(u)| \leq 3\ell_0 \Lambda. \tag{175}$$

On the planned sample space, we define the raw average $\bar{\lambda}(u) \coloneqq \frac{1}{N} \sum_{s=1}^N W_s(u)$. If the block is not aborted, this is exactly the value reported by the block. After applying Hoeffding's inequality, we get

$$\Pr\Bigl[|\bar{\lambda}(u) - \lambda_u| > \varepsilon\Bigr] \leq 2 \exp\left(-\frac{N\varepsilon^2}{18\ell_0^2 \Lambda^2}\right). \tag{176}$$

By choosing C_0 sufficiently large, the right-hand side is at most $1/(16|\mathcal{U}|)$. A union bound over all $u \in \mathcal{U}$ yields

$$\Pr\left[\max_{u \in \mathcal{U}} |\bar{\lambda}(u) - \lambda_u| > \varepsilon\right] \leq \frac{1}{16}. \tag{177}$$

The planned time of one block satisfies $\mathbb{E}[\Theta] = N \frac{\ell_1}{2\Lambda}$. Since $\tau = 4\ell_1 \frac{N}{\Lambda} = 8N \frac{\ell_1}{2\Lambda}$, Markov's inequality gives

$$\Pr[\Theta > \tau] \leq \frac{1}{8}. \tag{178}$$

Therefore, a single block succeeds simultaneously for all labels in $\mathcal{U}$ with probability at least $1 - \frac{1}{16} - \frac{1}{8} = \frac{13}{16}$.

Let I_j be the indicator that block j fails. The variables I_j are independent and satisfy $\mathbb{E}[I_j] \le \frac{3}{16}$. If fewer than half of the blocks fail, then more than half of the block outputs are ε-accurate for every $u \in \mathcal{U}$. Therefore, by taking the median for each coordinate, the result guarantees the precision of ε for $u \in U$ under all circumstances. Hence, we have

$$\Pr\left[\max_{u\in\mathcal{U}} |\widehat{\lambda}_{\mathcal{U}}(u) - \lambda_u| > \varepsilon\right] \le \Pr\left[\sum_{j=1}^{R} I_j \ge \frac{R}{2}\right]. \tag{179}$$

By the choice of R, the Chernoff bound gives

$$\Pr\left[\sum_{j=1}^{R} I_j \ge \frac{R}{2}\right] \le \exp\left(-\frac{25}{128}R\right) \le \eta. \tag{180}$$

It remains to bound the total evolution time. Each block uses at most τ, so the total time is at most $R\tau$. Using $0 < \varepsilon \le \Lambda$ and $|\mathcal{U}| \ge 1$, we have

$$N \le C' \frac{\Lambda^2}{\varepsilon^2} \log(16|\mathcal{U}|) \tag{181}$$

for an absolute constant C'. We also note that $R \le C' \log \frac{2}{\eta}$ after enlarging C'. Therefore, we have

$$R\tau = 4\ell_1 R \frac{N}{\Lambda} \le C \frac{\Lambda}{\varepsilon^2} \log\big(4|\mathcal{U}|\big) \log \frac{2}{\eta} \tag{182}$$

after enlarging the absolute constant C. This proves the theorem. □

D Proof of Theorem 4.1

We have now clearly demonstrated two different measurements and their functions. The same-Pauli displacement sieve gives a small Cartesian candidate set that contains all detectable labels. The cross-Pauli estimator then estimates all coefficients in that candidate set. Now we combine them to give the complete process of our protocol.

We note that the displacement sieve is only a projection sieve. At the population level, it identifies the large x- and z-projections

$$\Pi_x^{(\varepsilon)} \coloneqq \{d \in \mathbb{F}_2^n : W_x(d) \ge \varepsilon^2\}, \qquad \Pi_z^{(\varepsilon)} \coloneqq \{e \in \mathbb{F}_2^n : W_z(e) \ge \varepsilon^2\}. \tag{183}$$

These projections define the Cartesian candidate set

$$\mathcal{U}_{\mathrm{proj}} \coloneqq (\Pi_x^{(\varepsilon)} \cup \{0\}) \times (\Pi_z^{(\varepsilon)} \cup \{0\}) \setminus \{0\}. \tag{184}$$

The true ε-detectable support $S_\varepsilon \coloneqq \{u \in V \setminus \{0\} : |\lambda_u| \ge \varepsilon\}$ is contained in this candidate set $S_\varepsilon \subseteq \mathcal{U}_{\mathrm{proj}}$. This means $\mathcal{U}_{\mathrm{proj}}$ can be larger than S_ε. Therefore, in the second stage, we need to estimate the coefficient λ_u in the candidate set $\widehat{\mathcal{U}}_{\mathrm{proj}} \coloneqq (\widehat{\Pi}_x \cup \{0\}) \times (\widehat{\Pi}_z \cup \{0\}) \setminus \{0\}$. We define the precision of the second stage as $\varepsilon_{\mathrm{lin}}$, i.e.,

$$\max_{u\in\widehat{\mathcal{U}}_{\mathrm{proj}}} |\widehat{\lambda}_u - \lambda_u| \le \varepsilon_{\mathrm{lin}}. \tag{185}$$

Thus, we obtain an estimation of the true support S_ε as

$$\widehat{S}_\varepsilon \coloneqq \{u \in \widehat{\mathcal{U}}_{\mathrm{proj}} : |\widehat{\lambda}_u| > \theta\}. \tag{186}$$

The remaining question is what values should θ and $\varepsilon_{\mathrm{lin}}$ take to satisfy the requirement. We only need to note that we ultimately need every coefficient with $|\lambda_u| \ge \varepsilon$ to pass the final threshold. Therefore, a reasonable choice is to take

$$\varepsilon_{\mathrm{lin}} \coloneqq \frac{\varepsilon}{6}, \qquad \theta \coloneqq \frac{\varepsilon}{2}. \tag{187}$$

Now, we are ready to show Theorem 4.1, which we restate as the following for convenience.

Theorem D.1 (Upper bound restated)**.** *Let*

$$H = \sum_{u \in S} \lambda_u P_u, \qquad S \subseteq V \setminus \{0\}, \qquad \|H\| \le \Lambda, \tag{188}$$

and fix $0 < \varepsilon \le \Lambda$ *and* $\delta \in (0,1)$*. Then Algorithm 2 outputs* $(\widehat{S}_\varepsilon, \widehat{H}_\varepsilon)$ *such that, with probability at least* $1-\delta$*,*

$$S_\varepsilon \subseteq \widehat{S}_\varepsilon \subseteq \{u \in V \setminus \{0\} : |\lambda_u| > \varepsilon/3\}, \qquad \max_{u \in S_\varepsilon} |\widehat{\lambda}_u - \lambda_u| \le \varepsilon/6. \tag{189}$$

Moreover, for an absolute constant $C > 0$*,*

$$T_{\mathrm{tot}}^{\mathrm{det}} \le C \frac{\Lambda}{\varepsilon^2} \left(1 + \log \frac{\Lambda}{\varepsilon}\right) \log \frac{4}{\delta} \log\log \frac{8}{\delta}. \tag{190}$$

Proof. First, we run the displacement sieve in Lemma B.11 with scale ε and failure probability $\delta/2$. Let the outputs be $\widehat{\Pi}_x$ and $\widehat{\Pi}_z$, and define

$$\widehat{\mathcal{U}}_{\mathrm{proj}} \coloneqq (\widehat{\Pi}_x \cup \{0\}) \times (\widehat{\Pi}_z \cup \{0\}) \setminus \{0\}. \tag{191}$$

On the projection success event, which has probability at least $1 - \delta/2$, every label in S_ε is contained in $\widehat{\mathcal{U}}_{\mathrm{proj}}$. Indeed, if $u = (x \mid z)$ and $|\lambda_u| \ge \varepsilon$, then the corresponding projection intensity is at least ε^2 in every nonzero coordinate. Hence, the sieve includes the nonzero x and z coordinates, while zero coordinates are also included.

We also need the size of the candidate set. The proof of Lemma B.11 gives the deterministic bound

$$|\widehat{\Pi}_x| + |\widehat{\Pi}_z| \le C_1 \frac{\Lambda^2}{\varepsilon^2} \log \frac{8}{\delta}, \tag{192}$$

for an absolute constant C_1. Therefore, whenever $\widehat{\mathcal{U}}_{\mathrm{proj}}$ is nonempty, we have

$$|\widehat{\mathcal{U}}_{\mathrm{proj}}| \le \big(|\widehat{\Pi}_x| + 1\big)\big(|\widehat{\Pi}_z| + 1\big) - 1 \le \left(C_1 \frac{\Lambda^2}{\varepsilon^2} \log \frac{8}{\delta} + 1\right)^2. \tag{193}$$

Taking the logarithm keeps the product inside the logarithm as a sum. Hence

$$\log\Big(4|\widehat{\mathcal{U}}_{\mathrm{proj}}|\Big) \le C_2 \left(\log \frac{4\Lambda}{\varepsilon} + \log\log \frac{8}{\delta}\right). \tag{194}$$

Now consider any candidate set. If $\widehat{\mathcal{U}}_{\mathrm{proj}} = \varnothing$, then there are no Pauli labels to estimate in the second stage. The theorem is clearly true. Otherwise, apply Lemma C.10 on this fixed candidate set with accuracy $\varepsilon/6$ and failure probability $\delta/2$. Together with Eq. (194), it gives

$$\Pr\left[\max_{u \in \widehat{\mathcal{U}}_{\mathrm{proj}}} |\widehat{\lambda}_u - \lambda_u| \le \frac{\varepsilon}{6} \middle| \widehat{\mathcal{U}}_{\mathrm{proj}}\right] \ge 1 - \frac{\delta}{2}. \tag{195}$$

Let E_{proj} be the projection success event and let E_{lin} be the event that the linear estimates are $\varepsilon/6$-accurate on the candidate set. The above conditional statement holds for every candidate set, so

$$\Pr(E_{\mathrm{proj}} \cap E_{\mathrm{lin}}) \ge 1 - \delta. \tag{196}$$

It remains to check the threshold. On $E_{\mathrm{proj}} \cap E_{\mathrm{lin}}$, we take any $u \in S_\varepsilon$. Then $u \in \widehat{\mathcal{U}}_{\mathrm{proj}}$ and

$$|\widehat{\lambda}_u| \ge |\lambda_u| - \frac{\varepsilon}{6} \ge \frac{5\varepsilon}{6} > \frac{\varepsilon}{2} = \theta. \tag{197}$$

Therefore, we have $u \in \widehat{S}_\varepsilon$. This proves $S_\varepsilon \subseteq \widehat{S}_\varepsilon$. Conversely, if $u \in \widehat{S}_\varepsilon$, then $u \in \widehat{\mathcal{U}}_{\mathrm{proj}}$ and

$$|\lambda_u| \ge |\widehat{\lambda}_u| - \frac{\varepsilon}{6} > \frac{\varepsilon}{2} - \frac{\varepsilon}{6} = \frac{\varepsilon}{3}. \tag{198}$$

Therefore, we have

$$\widehat{S}_\varepsilon \subseteq \{u \in V \setminus \{0\} : |\lambda_u| > \varepsilon/3\}. \tag{199}$$

Finally, we bound the deterministic evolution time. The projection stage uses

$$T_{\mathrm{proj}}^{\mathrm{det}} \le C_3 \frac{\Lambda}{\varepsilon^2} \log \frac{4\Lambda}{\varepsilon} \log \frac{8}{\delta}. \tag{200}$$

The linear stage uses accuracy $\varepsilon/6$, so by Lemma C.10 and Eq. (194), we have

$$T_{\mathrm{lin}}^{\mathrm{det}} \le C_4 \frac{\Lambda}{\varepsilon^2} \left(1 + \log \frac{\Lambda}{\varepsilon}\right) \log \frac{4}{\delta} \log\log \frac{8}{\delta}. \tag{201}$$

Since $\log(4\Lambda/\varepsilon) \le C(1 + \log(\Lambda/\varepsilon))$, $\log\log(8/\delta) \ge \log\log 8 > 0$, and $\log(8/\delta) \le \frac{3}{2}\log(4/\delta)$, the sum of the two bounds gives

$$T_{\mathrm{tot}}^{\mathrm{det}} \le C \frac{\Lambda}{\varepsilon^2} \left(1 + \log \frac{\Lambda}{\varepsilon}\right) \log \frac{4}{\delta} \log\log \frac{8}{\delta}. \tag{202}$$

This proves the claimed result and thus Theorem 4.1. □

E Robustness to calibrated SPAM errors

In this appendix, we consider the impact of calibrated SPAM errors.

E.1 The calibrated SPAM error model

For the coefficient-estimation stage, we use the following calibrated local depolarizing model. For a single qubit, let the preparation and measurement error be local depolarization noise on each qubit as

$$\mathcal{P}(\rho) = r_{\mathrm{p}}\rho + (1 - r_{\mathrm{p}})\frac{I}{2}, \qquad \mathcal{M}(\rho) = r_{\mathrm{m}}\rho + (1 - r_{\mathrm{m}})\frac{I}{2}, \tag{203}$$

where $0 < r_{\mathrm{p}}, r_{\mathrm{m}} \le 1$ are known calibration numbers. The preparation noise $\mathcal{P}^{\otimes n}$ is applied after the ideal product-state preparation, and the measurement noise $\mathcal{M}^{\otimes n}$ is applied immediately before the ideal terminal Pauli measurement. The noise maps and all planned random settings are applied independently from shot to shot. For any Pauli string P,

$$\mathcal{P}^{\otimes n}(P) = r_{\mathrm{p}}^{w(P)} P, \qquad (\mathcal{M}^{\otimes n})^{\dagger}(P) = r_{\mathrm{m}}^{w(P)} P, \tag{204}$$

where $w(P)$ is the weight of P as the number of non-identity single-qubit Paulis in P.

The projection stage uses same-basis displacement records. We assume a calibrated bound for the full recorded displacement labels. Let D_Z, E_X denote the ideal displacement labels generated by fresh Z- and X-basis displacement shots, including the fresh random time $T \sim q_\Lambda$. Let $\widetilde{D}_{Z,j}, \widetilde{E}_{X,j}$ denote the recorded labels in the j-th shot in the presence of SPAM noise. We assume that there is a known number $\xi_1 \ge 0$ such that, for every block, every planned shot j, every history transcript on previous preparation and measurement records in the block so far (denoted as $\mathcal{F}_{j-1}$), and every nonzero displacement label,

$$\Pr\left[\widetilde{D}_{Z,j} = d \mid \mathcal{F}_{j-1}\right] \ge \Pr[D_Z = d] - \xi_1, \quad \Pr\left[\widetilde{E}_{X,j} = e \mid \mathcal{F}_{j-1}\right] \ge \Pr[E_X = e] - \xi_1. \tag{205}$$

For the displacement stage, the calibration bound is assumed shot by shot on the planned samples, before conditioning on block execution. Different blocks, including their SPAM noise, are independent. Here, ξ_1 bounds the full n-qubit displacement record. Therefore, any local calibration must first be converted to this full-record bound.

For Hermitian Pauli strings P, Q, recall that $F_{P,Q}(t) \coloneqq 2^{-n}\operatorname{Tr}(P(t)Q)$ with $P(t) = e^{iHt}Pe^{-iHt}$. We have the following SPAM-noisy version of estimating $F_{P,Q}(t)$.

Lemma E.1 (SPAM-noise-scaled estimation of $F_{P,Q}(t)$)**.** *Run the experiment with the intended preparation Pauli Q and intended measurement Pauli P. Let $X \in \{\pm 1\}$ be the usual product of the prepared eigenvalue sign and the measured eigenvalue sign. Under the calibrated depolarizing SPAM error model above, we have*

$$\mathbb{E}[X \mid t] = r_{\mathrm{p}}^{w(Q)} r_{\mathrm{m}}^{w(P)} F_{P,Q}(t). \tag{206}$$

Consequently, the rescaled variable $X^{\sharp} \coloneqq r_{\mathrm{p}}^{-w(Q)} r_{\mathrm{m}}^{-w(P)} X$ satisfies

$$\mathbb{E}[X^{\sharp} \mid t] = F_{P,Q}(t), \qquad |X^{\sharp}| \le r_{\mathrm{p}}^{-w(Q)} r_{\mathrm{m}}^{-w(P)}. \tag{207}$$

Proof. For the ideal experiment, the signed average over the prepared eigenbasis of Q gives $2^{-n}Q$. After preparation noise, this signed operator becomes $2^{-n}\mathcal{P}^{\otimes n}(Q) = 2^{-n}r_{\mathrm{p}}^{w(Q)}Q$. Measurement noise before an ideal measurement of P is equivalent, in the Heisenberg picture, to measuring $(\mathcal{M}^{\otimes n})^{\dagger}(P) = r_{\mathrm{m}}^{w(P)}P$. Therefore, the observed signed expectation equals

$$2^{-n}\operatorname{Tr}\left(r_{\mathrm{m}}^{w(P)}P(t)r_{\mathrm{p}}^{w(Q)}Q\right) = r_{\mathrm{p}}^{w(Q)}r_{\mathrm{m}}^{w(P)}F_{P,Q}(t). \tag{208}$$

Dividing by the known factor gives the last two claims. □

E.2 SPAM error in the projection stage

Let $\rho_\varepsilon \coloneqq \min\left\{1, \frac{\varepsilon^2}{2\kappa_0\Lambda^2}\right\}$. Assume $0 \le \xi_1 < \rho_\varepsilon$, and set $\rho_1 \coloneqq \rho_\varepsilon - \xi_1$. We have the following proposition.

Proposition E.2 (SPAM-robust displacement stage)**.** *Fix $0 < \varepsilon \le \Lambda$ and $\eta \in (0,1)$. We set variables*

$$N_1 \coloneqq \left\lceil C_1^{(0)}\rho_1^{-1}\log\frac{16\Lambda^2}{\varepsilon^2}\right\rceil, \quad L_1 \coloneqq \left\lceil 16\log\frac{4}{\eta}\right\rceil, \quad \tau_1 \coloneqq 8N_1\frac{\kappa_1}{2\Lambda}, \tag{209}$$

where $C_1^{(0)}$ is a sufficiently large absolute constant. Replace the block length in the ideal displacement stage by N_1. Use L_1 independent blocks for each of the two bases Z and X, and use the deterministic block cap τ_1. Let $\widehat{\Pi}_x, \widehat{\Pi}_z$ be the unions of all nonzero recorded displacement labels from the non-aborted blocks in the two bases. Then, with probability at least $1-\eta$,

$$\Pi_x^{(\varepsilon)} \setminus \{0\} \subseteq \widehat{\Pi}_x, \qquad \Pi_z^{(\varepsilon)} \setminus \{0\} \subseteq \widehat{\Pi}_z. \tag{210}$$

The deterministic projection-stage time is bounded by

$$T_1 \le C_1\frac{1}{\Lambda\rho_1}\log\frac{16\Lambda^2}{\varepsilon^2}\log\frac{4}{\eta}, \quad \Rightarrow \quad T_1 \le C_1'\frac{\Lambda}{\varepsilon^2}\log\frac{16\Lambda^2}{\varepsilon^2}\log\frac{4}{\eta} \tag{211}$$

if $\xi_1 \le \rho_\varepsilon/2$ for constants C_1, C_1'.

Proof. We prove the Z-basis statement, and the X-basis statement follows an identical argument. For every nonzero $d \in \Pi_x^{(\varepsilon)}$, Proposition B.8 gives

$$\Pr[D_Z = d] \ge \frac{W_x(d)}{2\kappa_0\Lambda^2} \ge \rho_\varepsilon. \tag{212}$$

By Eq. (205), for every planned shot j,

$$\Pr\left[\widetilde{D}_{Z,j} = d \mid \mathcal{F}_{j-1}\right] \ge \rho_\varepsilon - \xi_1 = \rho_1. \tag{213}$$

Therefore, by the chain rule, a fixed $d \in \Pi_x^{(\varepsilon)} \setminus \{0\}$ is missed by all N_1 planned shots in one block with probability at most

$$\Pr\left[\widetilde{D}_{Z,1} \ne d, \dots, \widetilde{D}_{Z,N_1} \ne d\right] \le (1-\rho_1)^{N_1} \le e^{-N_1\rho_1}. \tag{214}$$

The choice of N_1 makes this at most $\varepsilon^2/(16\Lambda^2)$. Since $|\Pi_x^{(\varepsilon)} \setminus \{0\}| \le \frac{\Lambda^2}{\varepsilon^2}$, a union bound shows that the planned displacement samples in one Z-basis block contain every nonzero element of $\Pi_x^{(\varepsilon)}$ with probability at least $15/16$.

The scheduled cumulative time in one block has the expectation $N_1\kappa_1/(2\Lambda)$. By Markov's inequality and the definition of τ_1, the block aborts with probability at most $1/8$. Hence, a single block succeeds with probability at least $1 - 1/16 - 1/8 = 13/16$. With $L_1 = \lceil 16\log(4/\eta)\rceil$ independent blocks, the probability that no Z-basis block succeeds is at most $\eta/4$. The same estimate holds in the X basis. A union bound over the two bases proves the two inclusions.

The time bound is the deterministic cap $2L_1\tau_1$. The final simplified bound follows from $\xi_1 \le \rho_\varepsilon/2$, which gives $\rho_1^{-1} \le 2\rho_\varepsilon^{-1}$, and from

$$\rho_\varepsilon^{-1} = \max\left\{1, \frac{2\kappa_0\Lambda^2}{\varepsilon^2}\right\} \le C_\kappa\frac{\Lambda^2}{\varepsilon^2} \tag{215}$$

for $0 < \varepsilon \le \Lambda$ and an absolute constant C_κ. □

E.3 SPAM errors in the coefficient learning stage

Fix a target Pauli basis $c \in \{X, Y, Z\}^n$. For $A, B \subseteq [n]$, keep the same Pauli strings Q_A and P_B used in the ideal stage. Define $a_{A,B} \coloneqq r_{\mathrm{p}}^{|A|} r_{\mathrm{m}}^{|B|}$ and $b_{A,B} \coloneqq r_{\mathrm{p}}^{|B|} r_{\mathrm{m}}^{|A|}$. The parity variable under SPAM noise is

$$Z_{A,B}^{(c),\sharp} \coloneqq \begin{cases} a_{A,B}^{-1} s_A m_B, & S = +1, \\ -b_{A,B}^{-1} r_B n_A, & S = -1. \end{cases} \tag{216}$$

Lemma E.3 (Identity on parity variable and weighted parity sample value under SPAM noise)**.** *For every target Pauli basis c, every $A, B \subseteq [n]$, and every $t \geq 0$,*

$$\mathbb{E}[Z_{A,B}^{(c),\sharp} \mid T = t] = \frac{1}{2}\left(F_{A,B}^{(c)}(t) - F_{A,B}^{(c)}(-t)\right). \tag{217}$$

Consequently, with $G_{A,B}^{(c),\sharp} \coloneqq \|L_\Lambda\|_1 \operatorname{sign}(L_\Lambda(T)) Z_{A,B}^{(c),\sharp}$ the corrected weighted parity sample value, one has

$$\mathbb{E}[G_{A,B}^{(c),\sharp}] = \left(F_{A,B}^{(c)}\right)'(0). \tag{218}$$

Proof. Given $S = +1$, the intended preparation Pauli is Q_A, and the intended measurement Pauli is P_B. Lemma E.1 gives

$$\mathbb{E}[s_A m_B \mid T = t, S = +1] = a_{A,B} F_{A,B}^{(c)}(t). \tag{219}$$

After division by $a_{A,B}$, the contribution is $F_{A,B}^{(c)}(t)$. Given $S = -1$, the intended preparation Pauli is P_B, and the intended measurement Pauli is Q_A. Lemma E.1 and cyclicity of trace give

$$\mathbb{E}[r_B n_A \mid T = t, S = -1] = b_{A,B} F_{A,B}^{(c)}(-t). \tag{220}$$

The definition includes a minus sign, so this branch contributes $-F_{A,B}^{(c)}(-t)$. Averaging the two equally likely branches gives the first identity. The second follows the same kernel calculation as in the ideal stage. □

Let $\mathcal{U} \subseteq V \setminus \{0\}$ be the candidate set used in the coefficient stage. Define

$$\zeta_2(\mathcal{U}) \coloneqq \inf\left\{ r_{\mathrm{p}}^{|A_c(u)|} r_{\mathrm{m}}^{|B_c(u)|}, r_{\mathrm{p}}^{|B_c(u)|} r_{\mathrm{m}}^{|A_c(u)|} : u \in \mathcal{U},\ c \in \{X, Y, Z\}^n,\ u \in \mathcal{V}(c) \right\}. \tag{221}$$

If $\mathcal{U} = \emptyset$, we set $\zeta_2(\mathcal{U}) = 1$. If every $u \in \mathcal{U}$ has Pauli weight at most k, then $\zeta_2(\mathcal{U}) \geq (r_{\mathrm{p}} r_{\mathrm{m}})^k$. If a k-local ansatz is known and desired, one may first replace $\mathcal{U}$ by $\mathcal{U}_{\leq k} \coloneqq \{u \in \mathcal{U} : w(u) \leq k\}$, and without this filtering, projection false positives can have weight larger than that of the true terms.

Proposition E.4 (SPAM-robust coefficient estimation)**.** *Let $\mathcal{U} \subseteq V \setminus \{0\}$ be finite and let $\zeta_2 = \zeta_2(\mathcal{U}) > 0$. If $\mathcal{U} = \emptyset$, the protocol returns no coefficients and uses zero time. Otherwise, run the coefficient estimation stage with variables under SPAM noise from Lemma E.3. Then, for every target accuracy $\varepsilon > 0$ and failure probability $\eta \in (0, 1)$, there is an estimator $\widehat{\lambda}_{\mathcal{U}}$ such that*

$$\Pr\left[\max_{u \in \mathcal{U}} |\widehat{\lambda}_{\mathcal{U}}(u) - \lambda_u| \leq \varepsilon\right] \geq 1 - \eta, \tag{222}$$

and the deterministic time satisfies

$$T_{\mathrm{lin}}^{\sharp}(\mathcal{U}; \varepsilon, \eta) \leq C_2 \frac{\Lambda}{\varepsilon^2 \zeta_2^2} \log(4 \max\{|\mathcal{U}|, 1\}) \log \frac{2}{\eta}. \tag{223}$$

If every $u \in \mathcal{U}$ has Pauli weight at most k, then the variance overhead is at most $(r_{\mathrm{p}} r_{\mathrm{m}})^{-2k}$.

Proof. If $\mathcal{U} = \emptyset$, the conclusion is straightforward. Assume $\mathcal{U} \neq \emptyset$. For a visible label $u \in \mathcal{V}(c)$, define the corrected target Pauli sample exactly as in the ideal proof, but with $G_{A,B}^{(c),\sharp}$ in place of $G_{A,B}^{(c)}$:

$$Y_u^{(c),\sharp} \coloneqq \frac{1}{2} \sigma_c(A_c(u), B_c(u)) G_{A_c(u), B_c(u)}^{(c),\sharp}. \tag{224}$$

If $u \notin \mathcal{V}(c)$, set $Y_u^{(c),\sharp} = 0$. Lemma E.3 and the ideal case calculation give

$$\mathbb{E}[Y_u^{(c),\sharp}] = \lambda_u \quad (u \in \mathcal{V}(c)), \qquad |Y_u^{(c),\sharp}| \le \ell_0 \Lambda \zeta_2^{-1}. \tag{225}$$

After the correction over randomly chosen c as $\widetilde{Y}_u^{\sharp} \coloneqq q(u)^{-1} \mathbb{1}\{u \in \mathcal{V}(C)\} Y_u^{(C),\sharp}$, we still have

$$\mathbb{E}[\widetilde{Y}_u^{\sharp}] = \lambda_u, \qquad |\widetilde{Y}_u^{\sharp}| \le 3\ell_0 \Lambda \zeta_2^{-1}, \tag{226}$$

because $q(u) \ge 1/3$.

The rest is the same Hoeffding and median-of-blocks argument as in the ideal stage, with the single-shot bound $3\ell_0\Lambda$ replaced by $3\ell_0\Lambda\zeta_2^{-1}$. Thus, the number of shots and the deterministic time are multiplied by ζ_2^{-2}. If all labels in $\mathcal{U}$ have weight at most k, then $|A_c(u)| \le k$ and $|B_c(u)| \le k$, so every preparation-measurement factor is at least $(r_{\mathrm{p}} r_{\mathrm{m}})^k$. Hence, we have $\zeta_2^{-2} \le (r_{\mathrm{p}} r_{\mathrm{m}})^{-2k}$. □

E.4 SPAM-robust Hamiltonian reconstruction

We now combine the arguments in the previous two stages and obtain the following theorem.

Theorem E.5 (SPAM-robust Hamiltonian reconstruction)**.** *Let $H = \sum_{u \in V \setminus \{0\}} \lambda_u P_u$ with $\|H\| \le \Lambda$, and fix $0 < \varepsilon \le \Lambda$ and $\delta \in (0,1)$. Assume the calibrated displacement condition Eq.* (205) *with $\xi_1 < \rho_\varepsilon = \min\{1, \frac{\varepsilon^2}{2\kappa_0 \Lambda^2}\}$. Run the SPAM-robust projection stage with failure probability $\delta/2$ and form $\widehat{\mathcal{U}} \coloneqq \big((\widehat{\Pi}_x \cup \{0\}) \times (\widehat{\Pi}_z \cup \{0\})\big) \setminus \{0\}$, and then run the SPAM-robust coefficient estimation stage on $\widehat{\mathcal{U}}$, using fresh noise and settings independent of the projection stage, with target accuracy $\varepsilon/6$ and failure probability $\delta/2$. Finally, we set*

$$\widehat{S}_\varepsilon \coloneqq \left\{ u \in \widehat{\mathcal{U}} : |\widehat{\lambda}_{\widehat{\mathcal{U}}}(u)| > \frac{\varepsilon}{2} \right\}, \qquad \widehat{H}_\varepsilon \coloneqq \sum_{u \in \widehat{S}_\varepsilon} \widehat{\lambda}_{\widehat{\mathcal{U}}}(u) P_u. \tag{227}$$

Then, with probability at least $1 - \delta$,

$$S_\varepsilon \subseteq \widehat{S}_\varepsilon \subseteq \{u \in V \setminus \{0\} : |\lambda_u| > \varepsilon/3\}, \qquad \max_{u \in S_\varepsilon} |\widehat{\lambda}_{\widehat{\mathcal{U}}}(u) - \lambda_u| \le \varepsilon/6. \tag{228}$$

For every realized candidate set $\widehat{\mathcal{U}}$, the following pathwise scheduled-time bound holds:

$$T_{\mathrm{tot}}^{\sharp} \le C \left[\frac{1}{\Lambda(\rho_\varepsilon - \xi_1)} \log \frac{16\Lambda^2}{\varepsilon^2} \log \frac{8}{\delta} + \frac{\Lambda}{\varepsilon^2 \zeta_2(\widehat{\mathcal{U}})^2} \log\Big(4 \max\{|\widehat{\mathcal{U}}|, 1\}\Big) \log \frac{4}{\delta} \right]. \tag{229}$$

Moreover, $|\widehat{\Pi}_x|, |\widehat{\Pi}_z| \le L_1 N_1$, so the logarithmic candidate-size factor can be made nonrandom by replacing $|\widehat{\mathcal{U}}|$ with $(L_1 N_1 + 1)^2 - 1$. In particular, if $\xi_1 \le \rho_\varepsilon/2$ and every label in the realized candidate set $\widehat{\mathcal{U}}$ has Pauli weight at most k, then

$$T_{\mathrm{tot}}^{\sharp} = \widetilde{O}\left(\frac{\Lambda}{\varepsilon^2} \left(1 + (r_{\mathrm{p}} r_{\mathrm{m}})^{-2k}\right) \right). \tag{230}$$

Proof. By Proposition E.2, with probability at least $1 - \delta/2$, we have $\Pi_x^{(\varepsilon)} \setminus \{0\} \subseteq \widehat{\Pi}_x$ and $\Pi_z^{(\varepsilon)} \setminus \{0\} \subseteq \widehat{\Pi}_z$. On this event, every $u = (x(u) \mid z(u)) \in S_\varepsilon$ belongs to $\widehat{\mathcal{U}}$. Indeed, if $x(u) \ne 0$, then $W_x(x(u)) \ge \lambda_u^2 \ge \varepsilon^2$, so $x(u) \in \Pi_x^{(\varepsilon)} \subseteq \widehat{\Pi}_x$. If $x(u) = 0$, then it is included by the definition of $\widehat{\mathcal{U}}$. The same argument applies to $z(u)$.

Conditioned on the realized set $\widehat{\mathcal{U}}$, proposition E.4, applied with target accuracy $\varepsilon/6$, gives

$$\Pr\left[\max_{u \in \widehat{\mathcal{U}}} |\widehat{\lambda}_{\widehat{\mathcal{U}}}(u) - \lambda_u| \le \varepsilon/6 \,\middle|\, \widehat{\mathcal{U}} \right] \ge 1 - \delta/2. \tag{231}$$

A union bound gives simultaneous success of the projection and coefficient stages with probability at least $1 - \delta$.

On this event, if $u \in S_\varepsilon$, then $u \in \widehat{\mathcal{U}}$ and

$$|\widehat{\lambda}_{\widehat{\mathcal{U}}}(u)| \ge |\lambda_u| - \varepsilon/6 \ge 5\varepsilon/6 > \varepsilon/2, \tag{232}$$

so $u \in \widehat{S}_\varepsilon$. Conversely, if $u \in \widehat{S}_\varepsilon$, then

$$|\lambda_u| \geq |\widehat{\lambda}_{\widehat{\mathcal{U}}}(u)| - \varepsilon/6 > \varepsilon/2 - \varepsilon/6 = \varepsilon/3. \tag{233}$$

This proves the support inclusions. The coefficient error bound on S_ε is exactly the coefficient recovery stage accuracy.

The pathwise time bound is the sum of the projection-stage bound in Proposition E.2 and the coefficient-stage bound in Proposition E.4, after conditioning on the realized candidate set. The final displayed form follows from $\xi_1 \leq \rho_\varepsilon/2$ and $\zeta_2(\widehat{\mathcal{U}}) \geq (r_{\mathrm{p}} r_{\mathrm{m}})^k$ when all labels in the realized $\widehat{\mathcal{U}}$ have weight at most k. □

F Discrete Time Assumption

F.1 Sampling rules on the time grid

The previous analysis allows each shot to use evolution times of arbitrary lengths randomly sampled from a distribution. As real-world experiments cannot perform extremely short evolutions, we need to discuss the feasibility of our protocol under the discretization assumption. To this end, we consider a more restrictive setting where we only allow resolution time chosen from a discrete-time grid $\mathcal{T}_{t_0} \coloneqq \{0, t_0, 2t_0, \ldots\}$.

For a Hamiltonian class with norm bound Λ, uniform coefficient learning to accuracy $0 < \varepsilon < \Lambda/2$ with success probability greater than $1/2$ requires $t_0 < \pi/\Lambda$ (Corollary F.10). Under the sufficient grid condition in Eq. (236), our protocol can still work with only constant factors overhead over Theorem 4.1. The intuition is that signals are still band-limited, so we can use Poisson summation to convert the continuous kernel identities into exact discrete identities. If the grid is too coarse, it is possible for two different Hamiltonians to have identical evolutions at every allowed time. In that case, support recovery is impossible. In other words, the resolution $t_0 = O(1/\Lambda)$ is unavoidable under the setting where we only allow resolution time chosen from a discrete-time grid $\mathcal{T}_{t_0} \coloneqq \{0, t_0, 2t_0, \ldots\}$.

Definition F.1 (Resolution time in the grid setting)**.** *A resolution time is a number $t_0 > 0$ such that every allowed evolution time belongs to the grid*

$$\mathcal{T}_{t_0} \coloneqq \{0, t_0, 2t_0, \ldots\}. \tag{234}$$

The discrete time control-free model is the same access model as Definition 2.2, except that every shot must use a time in $\mathcal{T}_{t_0}$.

For our discretized algorithm, we need to fix t_0 based on a frequency count. The signals produced by the Hamiltonian use only frequencies in $[-2\Lambda, 2\Lambda]$. The kernels K_Λ and L_Λ use only frequencies in $[-4\Lambda, 4\Lambda]$. Therefore, after multiplying a signal by a kernel, the product involves frequencies in $[-6\Lambda, 6\Lambda]$. Poisson summation compares a continuous integral with a grid sum by looking at the Fourier transform at the points

$$0, \quad \pm\frac{2\pi}{t_0}, \quad \pm\frac{4\pi}{t_0}, \quad \ldots. \tag{235}$$

The point 0 gives the integral. We want every other point to lie outside $[-6\Lambda, 6\Lambda]$, because the Fourier transform is zero there. It would already be enough to assume $\frac{2\pi}{t_0} > 6\Lambda$, namely $t_0 < \frac{\pi}{3\Lambda}$. In the following parts of this section, we will use the slightly stronger condition

$$0 < t_0 < \frac{\pi}{6\Lambda}. \tag{236}$$

This only changes absolute constants.

F.2 Discretized kernel

Before describing the specific discretized kernel, we further explain some principles of kernel construction. Using the even kernel as an example

$$B_\Lambda \coloneqq 2\Lambda, \qquad \Phi_\Lambda(\omega) \coloneqq \omega^2 \varphi(\omega/B_\Lambda), \qquad K_\Lambda(t) \coloneqq -\frac{1}{2\pi} \int_{\mathbb{R}} \Phi_\Lambda(\omega) e^{i\omega t} d\omega. \tag{237}$$

Define the fixed profile kernel

$$K_*(s) \coloneqq -\frac{1}{2\pi}\int_{\mathbb{R}} \xi^2 \varphi(\xi) e^{i\xi s} d\xi. \tag{238}$$

Then K_* is independent of Λ. By changing the variables into $\omega = B_\Lambda \xi$, we get

$$K_\Lambda(t) = -\frac{1}{2\pi}\int_{\mathbb{R}} (B_\Lambda \xi)^2 \varphi(\xi) e^{iB_\Lambda \xi t} B_\Lambda d\xi = B_\Lambda^3 K_*(B_\Lambda t). \tag{239}$$

Since $B_\Lambda = 2\Lambda$, this is

$$K_\Lambda(t) = (2\Lambda)^3 K_*(2\Lambda t). \tag{240}$$

The power 3 comes from two sources: the factor ω^2 contributes B_Λ^2, and the measure $d\omega$ contributes one more B_Λ. Similarly, recall from the odd-kernel construction that

$$\Psi_\Lambda(\omega) \coloneqq i\omega\varphi(\omega/B_\Lambda), \qquad L_\Lambda(t) \coloneqq \frac{1}{2\pi}\int_{\mathbb{R}} \Psi_\Lambda(\omega) e^{-i\omega t} d\omega. \tag{241}$$

Define the fixed profile kernel

$$L_*(s) \coloneqq \frac{1}{2\pi}\int_{\mathbb{R}} i\xi\varphi(\xi) e^{-i\xi s} d\xi. \tag{242}$$

Again using $\omega = B_\Lambda \xi$, we obtain

$$L_\Lambda(t) = \frac{1}{2\pi}\int_{\mathbb{R}} iB_\Lambda \xi \varphi(\xi) e^{-iB_\Lambda \xi t} B_\Lambda d\xi = B_\Lambda^2 L_*(B_\Lambda t). \tag{243}$$

Thus

$$L_\Lambda(t) = (2\Lambda)^2 L_*(2\Lambda t). \tag{244}$$

The power 2 appears because the first derivative kernel has only one factor of ω, and $d\omega$ gives the other factor.

With these detailed principles, we can give an exact summation rule that replaces continuous integrals. For clarity, we assume that a function has frequencies in $[-B, B]$, i.e., its Fourier transform is zero outside $[-B, B]$.

Proposition F.2 (Summation on the time grid)**.** *Assume Eq.* (236)*, and let h be a rapidly decaying function that uses only frequencies in $[-6\Lambda, 6\Lambda]$. Then, we have*

$$\int_{\mathbb{R}} h(t)dt = t_0 \sum_{m\in\mathbb{Z}} h(mt_0). \tag{245}$$

Proof. The Poisson summation formula gives

$$t_0 \sum_{m\in\mathbb{Z}} h(mt_0) = \sum_{\ell\in\mathbb{Z}} \widehat{h}\left(\frac{2\pi\ell}{t_0}\right). \tag{246}$$

The right-hand side checks $\widehat{h}$ at the points $0, \pm\frac{2\pi}{t_0}, \pm\frac{4\pi}{t_0}, \ldots$. The point 0 gives

$$\widehat{h}(0) = \int_{\mathbb{R}} h(t)dt. \tag{247}$$

Now take any nonzero integer ℓ. Since $t_0 < \pi/(6\Lambda)$, we have

$$\left|\frac{2\pi\ell}{t_0}\right| \geq \frac{2\pi}{t_0} > 12\Lambda. \tag{248}$$

This point is outside $[-6\Lambda, 6\Lambda]$, while h uses no frequencies outside this interval. Therefore

$$\widehat{h}\left(\frac{2\pi\ell}{t_0}\right) = 0 \qquad (\ell \neq 0). \tag{249}$$

So only the $\ell = 0$ term remains and thus

$$t_0 \sum_{m\in\mathbb{Z}} h(mt_0) = \widehat{h}(0) = \int_{\mathbb{R}} h(t)dt. \tag{250}$$

□

We now apply this rule to the signals used by the algorithm. For a nonzero $d \in \mathbb{F}_2^n$, we define

$$f_d^{(Z)}(t) \coloneqq 2^{-n} \sum_{a \in \mathbb{F}_2^n} \left| \langle a + d | e^{-iHt} | a \rangle_Z \right|^2 . \tag{251}$$

Also for a nonzero $e \in \mathbb{F}_2^n$, we define

$$f_e^{(X)}(t) \coloneqq 2^{-n} \sum_{a \in \mathbb{F}_2^n} \left| \langle a + e | e^{-iHt} | a \rangle_X \right|^2 . \tag{252}$$

Thus $f_d^{(Z)}(t)$ and $f_e^{(X)}(t)$ are the fixed-time displacement probabilities in the Z and X bases. For a target Pauli basis c and subsets $A, B \subseteq [n]$, we define

$$g_{A,B}^{(c)}(t) \coloneqq F_{A,B}^{(c)}(t) - F_{A,B}^{(c)}(-t). \tag{253}$$

which is used to extract the odd time inversion part. We can prove the following proposition.

Proposition F.3 (Kernel identities on the time grid)**.** *Assume Eq.* (236)*. Then for every $v \in V$, every nonzero $d, e \in \mathbb{F}_2^n$, every target Pauli basis c, and every $A, B \subseteq [n]$, we have*

$$C''_{P_v}(0) = t_0 K_\Lambda(0) C_{P_v}(0) + 2t_0 \sum_{m \geq 1} K_\Lambda(mt_0) C_{P_v}(mt_0), \tag{254}$$

$$2W_x(d) = 2t_0 \sum_{m \geq 1} K_\Lambda(mt_0) f_d^{(Z)}(mt_0), \tag{255}$$

$$2W_z(e) = 2t_0 \sum_{m \geq 1} K_\Lambda(mt_0) f_e^{(X)}(mt_0), \tag{256}$$

$$\left(F_{A,B}^{(c)}\right)'(0) = t_0 \sum_{m \geq 1} L_\Lambda(mt_0) g_{A,B}^{(c)}(mt_0). \tag{257}$$

Proof. We apply Proposition F.2 to the four signals in the claimed proposition. By Proposition B.4 and Lemma B.7, the functions C_{P_v}, $f_d^{(Z)}$, and $f_e^{(X)}$ use only frequencies in $[-2\Lambda, 2\Lambda]$. By Proposition C.4, the same is true for $F_{A,B}^{(c)}$. The kernels use only frequencies in $[-4\Lambda, 4\Lambda]$ by construction of the cutoff. Therefore, the four products $K_\Lambda C_{P_v}$, $K_\Lambda f_d^{(Z)}$, $K_\Lambda f_e^{(X)}$, and $L_\Lambda F_{A,B}^{(c)}$ use only frequencies in $[-6\Lambda, 6\Lambda]$. Applying the exact grid identity to these products and using Lemmas B.5 and C.5 gives the displayed formulas. The first three formulas use the evenness of K_Λ, C_{P_v}, $f_d^{(Z)}$, and $f_e^{(X)}$. For the displacement formulas, we also use $f_d^{(Z)}(0) = 0$ and $f_e^{(X)}(0) = 0$ because $d, e \neq 0$. Finally, the last formula uses that L_Λ is odd, so the positive and negative grid points combine into $F_{A,B}^{(c)}(t) - F_{A,B}^{(c)}(-t)$. □

In the continuous time setting, we proved two bounds: one for the size of a one weighted sample and the other for the average sampling time of a single shot. Now we give their versions under the discrete-time assumption.

Lemma F.4 (Bounds under discrete time assumption)**.** *Assume Eq.* (236)*. There exist constants κ_0^{grid}, κ_1^{grid}, ℓ_0^{grid}, and ℓ_1^{grid}, depending only on the cutoff φ, such that*

$$t_0 |K_\Lambda(0)| + 2t_0 \sum_{m \geq 1} |K_\Lambda(mt_0)| \leq 4\kappa_0^{\mathrm{grid}} \Lambda^2, \tag{258}$$

$$2t_0 \sum_{m \geq 1} mt_0 |K_\Lambda(mt_0)| \leq 2\kappa_0^{\mathrm{grid}} \kappa_1^{\mathrm{grid}} \Lambda, \tag{259}$$

$$2t_0 \sum_{m \geq 1} |L_\Lambda(mt_0)| \leq 2\ell_0^{\mathrm{grid}} \Lambda, \tag{260}$$

$$2t_0 \sum_{m \geq 1} mt_0 |L_\Lambda(mt_0)| \leq \ell_0^{\mathrm{grid}} \ell_1^{\mathrm{grid}}. \tag{261}$$

Proof. Set $\Delta = 2\Lambda t_0$. Then $0 < \Delta < \pi/3$. Using Eq. (240) and Eq. (244), we have

$$t_0|K_\Lambda(0)| + 2t_0 \sum_{m\geq 1} |K_\Lambda(mt_0)| = 4\Lambda^2 \left(\Delta|K_*(0)| + 2\Delta \sum_{m\geq 1} |K_*(m\Delta)| \right). \tag{262}$$

The other three displayed bounds reduce in the same way to sums of K_* and L_* on a grid of mesh Δ. Since K_* and L_* decay rapidly, these rescaled sums and their first moments are bounded uniformly for $0 < \Delta < \pi/3$. Enlarging the four constants gives (258)-(261). □

Finally, we propose a discrete version of the projection sieving algorithm.

Corollary F.5 (Discrete displacement sieve)**.** *Assume Eq.* (236)*. Define one discrete displacement shot as follows. Sample $M \in \{0, 1, 2, \ldots\}$ by*

$$\Pr[M = 0] := 1 - \frac{2t_0 \sum_{m\geq 1} |K_\Lambda(mt_0)|}{4\kappa_0^{\mathrm{grid}}\Lambda^2}, \tag{263}$$

$$\Pr[M = m] := \frac{2t_0|K_\Lambda(mt_0)|}{4\kappa_0^{\mathrm{grid}}\Lambda^2}, \qquad m \geq 1. \tag{264}$$

If $M = 0$, record a null symbol and perform no evolution. If $M = m \geq 1$, run the usual Z-basis or X-basis displacement shot at time mt_0.

Let D^{grid} be the displacement returned by the Z-basis experiment, and let E^{grid} be the displacement returned by the X-basis experiment. Then for every nonzero $d, e \in \mathbb{F}_2^n$,

$$\Pr\big[D^{\mathrm{grid}} = d\big] \geq \frac{W_x(d)}{2\kappa_0^{\mathrm{grid}}\Lambda^2}, \qquad \Pr\big[E^{\mathrm{grid}} = e\big] \geq \frac{W_z(e)}{2\kappa_0^{\mathrm{grid}}\Lambda^2}. \tag{265}$$

Moreover,

$$\mathbb{E}[T_{\mathrm{proj}}^{\mathrm{grid}}] \leq \frac{\kappa_1^{\mathrm{grid}}}{2\Lambda}. \tag{266}$$

Proof. The distribution of M is valid by Eq. (258). We prove the Z-basis claim here, and the X-basis claim is identical. For a fixed $m \geq 1$, the probability of displacement d at time mt_0 is $f_d^{(Z)}(mt_0)$. Hence, we have

$$\Pr\big[D^{\mathrm{grid}} = d\big] = \frac{1}{2\kappa_0^{\mathrm{grid}}\Lambda^2} \sum_{m\geq 1} t_0|K_\Lambda(mt_0)|f_d^{(Z)}(mt_0). \tag{267}$$

Since $f_d^{(Z)}(t) \geq 0$, we have

$$\sum_{m\geq 1} t_0|K_\Lambda(mt_0)|f_d^{(Z)}(mt_0) \geq \left| \sum_{m\geq 1} t_0 K_\Lambda(mt_0) f_d^{(Z)}(mt_0) \right|. \tag{268}$$

By Eq. (255), the last signed sum is $W_x(d)$. This proves the lower bound for D^{grid}. The proof for E^{grid} uses Eq. (256) and follows a similar approach.

The mean-time bound follows directly from Eq. (259) as

$$\mathbb{E}[T_{\mathrm{proj}}^{\mathrm{grid}}] = \sum_{m\geq 1} \frac{2t_0|K_\Lambda(mt_0)|}{4\kappa_0^{\mathrm{grid}}\Lambda^2}(mt_0) \leq \frac{\kappa_1^{\mathrm{grid}}}{2\Lambda}. \tag{269}$$

□

We next replace one continuous-time coefficient estimation block with a discrete-time shot. The parity block is exactly the same as before. The only change is the choice of evolution time. Fix a target Pauli basis c and sets $A, B \subseteq [n]$. We sample an integer $M \in \{0, 1, 2, \ldots\}$ by

$$\Pr[M = 0] := 1 - \frac{2t_0 \sum_{m\geq 1} |L_\Lambda(mt_0)|}{2\ell_0^{\mathrm{grid}}\Lambda}, \tag{270}$$

$$\Pr[M = m] := \frac{2t_0|L_\Lambda(mt_0)|}{2\ell_0^{\mathrm{grid}}\Lambda}, \qquad m \geq 1. \tag{271}$$

The outcome $M = 0$ is null. In that case, we output 0 and do not evolve the system. If $M = m \geq 1$, we run the same parity block experiment as in Proposition C.3 but at time mt_0.

Let $Z^{(c),\mathrm{grid}}_{A,B,m}$ be the parity variable produced by this block. We put back the sign of $L_\Lambda(mt_0)$ and define

$$G^{(c),\mathrm{grid}}_{A,B} \coloneqq \begin{cases} 2\ell_0^{\mathrm{grid}}\Lambda \operatorname{sign}(L_\Lambda(mt_0))Z^{(c),\mathrm{grid}}_{A,B,m}, & M = m \geq 1, \\ 0, & M = 0. \end{cases} \tag{272}$$

This is the discrete version of the continuous linear sample. The factor $2\ell_0^{\mathrm{grid}}\Lambda$ only cancels the normalization in Eq. (271).

We will also use the same visibility correction as before. For a label $u \neq 0$ and a target Pauli basis c, define

$$Y^{(c),\mathrm{grid}}_u \coloneqq \begin{cases} \frac{1}{2}\sigma_c(A_c(u), B_c(u))G^{(c),\mathrm{grid}}_{A_c(u),B_c(u)}, & u \in \mathcal{V}(c), \\ 0, & u \notin \mathcal{V}(c). \end{cases} \tag{273}$$

If $C \sim \mathrm{Unif}\{X, Y, Z\}^n$ and $q(u) \coloneqq \Pr[u \in \mathcal{V}(C)]$, then the discrete sample after applying visibility correction is

$$\widetilde{Y}^{\mathrm{grid}}_u \coloneqq q(u)^{-1}Y^{(C),\mathrm{grid}}_u. \tag{274}$$

Proposition F.6 (Discrete linear sample)**.** *Assume Eq.* (236)*. For the grid shot defined above,*

$$\mathbb{E}\left[G^{(c),\mathrm{grid}}_{A,B}\right] = \left(F^{(c)}_{A,B}\right)'(0), \qquad \left|G^{(c),\mathrm{grid}}_{A,B}\right| \leq 2\ell_0^{\mathrm{grid}}\Lambda, \qquad \mathbb{E}[T^{\mathrm{grid}}_{\mathrm{lin}}] \leq \frac{\ell_1^{\mathrm{grid}}}{2\Lambda}. \tag{275}$$

Moreover, for every nonzero label u, we have

$$\mathbb{E}[\widetilde{Y}^{\mathrm{grid}}_u] = \lambda_u, \qquad |\widetilde{Y}^{\mathrm{grid}}_u| \leq 3\ell_0^{\mathrm{grid}}\Lambda. \tag{276}$$

Proof. The distribution of M is valid by Eq. (260). For $m \geq 1$, Proposition C.3 gives

$$\mathbb{E}\left[Z^{(c),\mathrm{grid}}_{A,B,m} \mid M = m\right] = \frac{1}{2}g^{(c)}_{A,B}(mt_0). \tag{277}$$

Therefore, we have

$$\begin{aligned} \mathbb{E}\left[G^{(c),\mathrm{grid}}_{A,B}\right] &= \sum_{m\geq 1} \frac{2t_0|L_\Lambda(mt_0)|}{2\ell_0^{\mathrm{grid}}\Lambda} \cdot 2\ell_0^{\mathrm{grid}}\Lambda \operatorname{sign}(L_\Lambda(mt_0)) \cdot \frac{1}{2}g^{(c)}_{A,B}(mt_0) \\ &= t_0 \sum_{m\geq 1} L_\Lambda(mt_0)g^{(c)}_{A,B}(mt_0) \\ &= \left(F^{(c)}_{A,B}\right)'(0), \end{aligned} \tag{278}$$

where the last step is Eq. (257). The bound on $G^{(c),\mathrm{grid}}_{A,B}$ follows from $|Z^{(c),\mathrm{grid}}_{A,B,m}| \leq 1$.

The mean time is also immediate from the definition of M as

$$\mathbb{E}[T^{\mathrm{grid}}_{\mathrm{lin}}] = \sum_{m\geq 1} \frac{2t_0|L_\Lambda(mt_0)|}{2\ell_0^{\mathrm{grid}}\Lambda}(mt_0) \leq \frac{\ell_1^{\mathrm{grid}}}{2\Lambda}, \tag{279}$$

where we used Eq. (261).

It remains to check the visibility correction. If $u \in \mathcal{V}(c)$, the derivative calculation in Proposition 3.4 gives

$$\left(F^{(c)}_{A_c(u),B_c(u)}\right)'(0) = 2\sigma_c(A_c(u), B_c(u))\lambda_u. \tag{280}$$

Thus,

$$\mathbb{E}\left[Y^{(c),\mathrm{grid}}_u\right] = \lambda_u, \qquad \left|Y^{(c),\mathrm{grid}}_u\right| \leq \ell_0^{\mathrm{grid}}\Lambda \tag{281}$$

whenever $u \in \mathcal{V}(c)$. If $u \notin \mathcal{V}(c)$, then $Y_u^{(c),\mathrm{grid}} = 0$ by definition.

Finally, Proposition C.6 gives $q(u) \geq 1/3$. Hence, we have

$$\begin{aligned} \mathbb{E}[\widetilde{Y}_u^{\mathrm{grid}}] &= q(u)^{-1} \Pr[u \in \mathcal{V}(C)]\lambda_u = \lambda_u, \\ |\widetilde{Y}_u^{\mathrm{grid}}| &\leq q(u)^{-1}\ell_0^{\mathrm{grid}}\Lambda \leq 3\ell_0^{\mathrm{grid}}\Lambda. \end{aligned} \tag{282}$$

□

The grid distributions above have infinite support, but their mean time is finite. If we want a deterministic limit on the duration of each shot, fix $R > 0$ and set

$$M_R \coloneqq \left\lfloor \frac{R}{2\Lambda t_0} \right\rfloor. \tag{283}$$

as the new cutoff threshold. Keep only outcomes $m \leq M_R$ and move the remaining probability to the null outcome. Then every executed shot has time at most

$$M_R t_0 \leq \frac{R}{2\Lambda}. \tag{284}$$

The only new loss is the tail:

$$2t_0 \sum_{m > M_R} |K_\Lambda(mt_0)| \quad \text{or} \quad 2t_0 \sum_{m > M_R} |L_\Lambda(mt_0)|. \tag{285}$$

Using Eq. (240), Eq. (244), and $\Delta = 2\Lambda t_0$, these tails are bounded by

$$C_N \Lambda^2 (1+R)^{-N} \quad \text{and} \quad C_N \Lambda (1+R)^{-N} \tag{286}$$

Therefore, R controls the tradeoff. A larger R allows a longer maximum shot time $R/(2\Lambda)$, but discards less probability mass. A smaller R gives a shorter maximum shot time, but discards more of the tail.

F.3 Truncated sampling

The grid identities above are exact before truncation. For finite sampling, we retain the original positive-grid probabilities and transfer the omitted probability to the null outcome. A null outcome contributes zero to the coefficient estimate and counts toward the prescribed block size. For $R \geq 4$, define

$$M_R \coloneqq \left\lfloor \frac{R}{2\Lambda t_0} \right\rfloor, \tag{287}$$

$$B_K(R) \coloneqq 2t_0 \sum_{m > M_R} |K_\Lambda(mt_0)|, \qquad B_L(R) \coloneqq 2t_0 \sum_{m > M_R} |L_\Lambda(mt_0)|. \tag{288}$$

Theorem F.7 (Finite-grid reconstruction)**.** *Let $\|H\| \leq \Lambda$, $0 < \varepsilon \leq \Lambda$, $\delta \in (0,1)$, and $0 < t_0 < \pi/(6\Lambda)$. Suppose that*

$$B_K(R) \leq \varepsilon^2/2, \qquad B_L(R) \leq \varepsilon/6. \tag{289}$$

For the specified cutoff, the explicit choice in Eq. (297) *is sufficient by Lemma F.8. Use the finite grid distributions*

$$\pi_K(m) = \frac{2t_0|K_\Lambda(mt_0)|}{4\kappa_0^{\mathrm{grid}}\Lambda^2}, \qquad \pi_L(m) = \frac{2t_0|L_\Lambda(mt_0)|}{2\ell_0^{\mathrm{grid}}\Lambda}, \qquad 1 \leq m \leq M_R, \tag{290}$$

$$\pi_K(0) = 1 - \sum_{m=1}^{M_R} \pi_K(m), \qquad \pi_L(0) = 1 - \sum_{m=1}^{M_R} \pi_L(m). \tag{291}$$

An outcome $m = 0$ is null and entails no evolution. Run the displacement sieve with the grid constants, count threshold parameter $p_R = \varepsilon^2/(8\kappa_0^{\mathrm{grid}}\Lambda^2)$, and failure budget $\delta/2$. On its candidate set, use fresh truncated parity samples, the same visibility correction and median-of-blocks procedure,

statistical radius $\varepsilon/12$, and failure budget $\delta/2$. Keep the final threshold $\varepsilon/2$ of Algorithm 2. Then, with probability at least $1-\delta$,

$$S_\varepsilon \subseteq \widehat{S}_\varepsilon \subseteq \{u \neq 0 : |\lambda_u| > \varepsilon/3\}, \qquad \max_{u\in S_\varepsilon} |\widehat{\lambda}_u - \lambda_u| \le \varepsilon/6. \tag{292}$$

Setting the omitted coefficients to zero also gives $\max_{u\neq 0} |\widehat{\lambda}_u - \lambda_u| \le \varepsilon$. Every executed evolution time is a grid multiple at most $R/(2\Lambda)$, and for a constant C depending only on the cutoff,

$$T_{\mathrm{tot}}^{\mathrm{det}} \le C\frac{\Lambda}{\varepsilon^2}\left(1+\log\frac{\Lambda}{\varepsilon}\right)\log\frac{4}{\delta}\log\log\frac{8}{\delta}. \tag{293}$$

The admissible truncation level does not depend on n or δ.

Proof. Write $\kappa = \kappa_0^{\mathrm{grid}}$. For a nonzero Z-basis displacement d, removing the tail decreases its occurrence probability by at most $B_K(R)/(4\kappa\Lambda^2)$. Consequently,

$$\Pr[D_R = d] \ge \frac{W_x(d)}{2\kappa\Lambda^2} - \frac{B_K(R)}{4\kappa\Lambda^2} = \frac{2W_x(d) - B_K(R)}{4\kappa\Lambda^2}. \tag{294}$$

If $W_x(d) \ge \varepsilon^2$, this is at least $3\varepsilon^2/(8\kappa\Lambda^2) \ge 2p_R$. The same argument applies in the X basis. Thus the proof of Lemma B.11 applies with p_R and the grid constants. Truncation can only decrease the planned mean time, so the same deterministic block caps and abortion argument remain valid. It follows that, with probability at least $1-\delta/2$, the candidate set contains S_ε and has deterministic size $O((\Lambda/\varepsilon)^4\log^2(8/\delta))$.

For a fixed frame c and visible label u, the bias in the truncated weighted derivative sample is at most $B_L(R)$, because $|F^{(c)}_{A,B}(t) - F^{(c)}_{A,B}(-t)| \le 2$. The subsequent factor $\sigma_c(A,B)/2$ makes the bias at most $B_L(R)/2$. Averaging over frames and applying the visibility correction gives

$$\left|\mathbb{E}[\widetilde{Y}_{u,R}] - \lambda_u\right| \le q(u)^{-1}\Pr[u\in\mathcal{V}(C)]\frac{B_L(R)}{2} = \frac{B_L(R)}{2} \le \frac{\varepsilon}{12}. \tag{295}$$

The sample magnitude remains at most $3\ell_0^{\mathrm{grid}}\Lambda$, and its mean evolution time is at most $\ell_1^{\mathrm{grid}}/(2\Lambda)$. Applying the Hoeffding, union-bound, abortion, and median arguments from Lemma C.10 to the truncated sample means, with statistical radius $\varepsilon/12$, bounds the total coefficient error by $\varepsilon/6$ simultaneously on the candidate set, with conditional probability at least $1-\delta/2$. The conditioning is on the first-stage candidate set; the second-stage randomness is fresh.

A union bound over the two stages and the same threshold calculation as in Theorem 4.1 prove Eq. (292). Any coefficient omitted from the output has magnitude less than ε, which proves the stated uniform coefficient guarantee. The sample sizes change only by absolute factors, and $\log\Big(4\max\{|\widehat{\mathcal{U}}_{\mathrm{proj}}|,1\}\Big)$ is bounded by a constant times $\log(4\Lambda/\varepsilon)+\log\log(8/\delta)$. As in the proof of Theorem D.1, summing the deterministic caps proves Eq. (293). Finally, $m \le M_R$ gives the claimed bound on every executed shot. □

Lemma F.8 (Kernel tail bounds)**.** *For the explicit cutoff $\varphi(\xi) = \theta((4-\xi^2)/3)$ constructed above from $\rho(s) = e^{-1/s}\mathbb{1}_{s>0}$ and $\theta(s) = \rho(s)/(\rho(s)+\rho(1-s))$, there are fixed constants $A_K, A_L, a > 0$ such that, uniformly over Eq.* (236)*,*

$$B_K(R) \le A_K\Lambda^2 e^{-a\sqrt{R}}, \qquad B_L(R) \le A_L\Lambda e^{-a\sqrt{R}}, \qquad R \ge 4. \tag{296}$$

Consequently, Eq. (289) *holds for*

$$R \ge \max\left\{4, a^{-2}\log^2\left(A\left(\frac{\Lambda}{\varepsilon}\right)^2\right)\right\}, \qquad A := \max\{e, 2A_K, 6A_L\}. \tag{297}$$

In particular, $R = O(\log^2(2\Lambda/\varepsilon))$ is sufficient.

Proof. The specified cutoff is Gevrey of order two: its derivatives satisfy $\|\varphi^{(j)}\|_\infty \le A_0 B_0^j (j!)^2$ for fixed constants A_0, B_0. To see this directly, for $s > 0$ apply Cauchy's derivative estimate to $e^{-1/z}$ on $|z-s| = s/2$; on this circle $\mathrm{Re}(1/z) \ge 2/(3s)$. This gives $|\rho^{(j)}(s)| \le j!(2/s)^j e^{-2/(3s)} \le 3^j (j!)^2$.

The derivatives extend by zero at $s = 0$. On $[0, 1]$, $\rho(s) + \rho(1 - s) \geq e^{-2}$. Leibniz's rule and the derivative recursion for its reciprocal therefore preserve bounds of the form $AB^j(j!)^2$; composition with the fixed quadratic $(4 - \xi^2)/3$ does so as well. The same bounds, with possibly different constants, apply to $\xi^2\varphi(\xi)$ and $i\xi\varphi(\xi)$ and to the L^1 norms of their derivatives.

Integrating the Fourier formulas for K_* and L_* by parts j times gives a bound $A_1 B_1^j (j!)^2 |s|^{-j}$. Taking j proportional to $\sqrt{|s|}$ yields $|K_*(s)| + |L_*(s)| \leq A_2 e^{-a_2\sqrt{|s|}}$. For $\Delta = 2\Lambda t_0 \in (0, \pi/3)$ and $R \geq 4$, comparison with the integral of this decreasing envelope gives

$$\Delta \sum_{m > \lfloor R/\Delta \rfloor} e^{-a_2\sqrt{m\Delta}} \leq \int_{R-\Delta}^{\infty} e^{-a_2\sqrt{s}}\, ds \leq A_3 e^{-a\sqrt{R}} \tag{298}$$

for fixed $A_3, a > 0$. The scaling identities Eqs. (240) and (244) yield the tail bounds. Substitution of Eq. (297), using $\Lambda/\varepsilon \geq 1$, proves the final claim. □

F.4 A minimal resolution time lower bound in the grid setting

The upper bound above assumes that the time grid spacing is smaller than a constant times $1/\Lambda$. We now show that this scale is necessary in the grid setting.

Proposition F.9 (A universal hard family)**.** *Fix $t_0 > 0$. Suppose every allowed shot time belongs to $\mathcal{T}_{t_0}$ and the admissible Hamiltonian class contains the one-sparse Hamiltonians*

$$H_u \coloneqq \frac{\pi}{t_0} P_u, \qquad u \in V \setminus \{0\}. \tag{299}$$

Then no protocol can identify the support label u with success probability better than random guessing, even with arbitrarily many shots and arbitrary adaptivity.

Proof. For every integer $m \geq 0$,

$$e^{-iH_u(mt_0)} = e^{-im\pi P_u} = \cos(m\pi) I - i \sin(m\pi) P_u = (-1)^m I. \tag{300}$$

Therefore, every allowed experiment has exactly the same outcome distribution for every u. So the support label can not be identified. □

The proposition uses Hamiltonians of norm π/t_0. Therefore, simply requiring the norm bound to be greater than π/t_0 is not enough. We will also give a necessary resolution condition depending on the norm bound.

Corollary F.10 (Necessary grid spacing for approximate learning)**.** *Fix $0 < \varepsilon < \Lambda/2$. If $t_0 \geq \pi/\Lambda$, no protocol restricted to $\mathcal{T}_{t_0}$ can estimate every non-identity coefficient of every Hamiltonian with $\|H\| \leq \Lambda$ to accuracy ε with success probability greater than $1/2$.*

Proof. Set $k = \lfloor \Lambda t_0/\pi \rfloor \geq 1$ and $a = k\pi/t_0$. Then $\Lambda/2 < a \leq \Lambda$. For any fixed non-identity Pauli string P, the two admissible Hamiltonians $H_+ = aP$ and $H_- = -aP$ have identical allowed evolutions,

$$e^{-iH_+(mt_0)} = e^{-iH_-(mt_0)} = (-1)^{km} I, \qquad m \geq 0. \tag{301}$$

Their outcome distributions are identical under arbitrary adaptive protocols, while their coefficientwise ε-balls are disjoint because $\varepsilon < a$. A common output distribution cannot put probability greater than $1/2$ in each of these disjoint success sets. For distinct labels $u \neq v$, the positive pair aP_u, aP_v likewise rules out uniform support recovery and coefficientwise learning whenever $\varepsilon < a/2$. □

G Lower Bound

G.1 Information theoretic tools

The proof of the lower bound consists of two parts. First, in a control-free protocol, even estimating a very small coefficient already reaches the standard quantum limit in time complexity. Second, learning a large number of hidden support locations leads to a coupon collection problem, thus adding a log term to the final lower bound. We begin with two fundamental information-theoretic inequalities as facts from quantum Fisher information for single-parameter estimations (see, e.g., Refs. [42, 43, 44]).

Fact G.1 (Total variation from Fisher information, see, e.g. [42, 43, 44])**.** *Let P_θ be the distribution of the full history of a protocol for a parametrization estimation problem on a single parameter θ, and let $I(\theta)$ be its classical Fisher information. Then for any $a < b$,*

$$\mathrm{TV}(P_a, P_b) \le \frac{1}{2}\int_a^b \sqrt{I(\xi)}d\xi. \tag{302}$$

Fact G.2 (One-shot Fisher information bound)**.** *Let $\rho_\theta(t) = e^{-iH_\theta t}\rho_0 e^{iH_\theta t}$ be followed by any measurement and classical postprocessing. Then the classical Fisher information is bounded by*

$$I_{\mathrm{cl}}(\theta) \le F_Q(\rho_\theta(t)), \tag{303}$$

where F_Q is the quantum Fisher information. If H_θ is differentiable and

$$\mathcal{G}_\theta(t) \coloneqq \int_0^t e^{iH_\theta s}(\partial_\theta H_\theta)e^{-iH_\theta s}ds, \tag{304}$$

then

$$F_Q(\rho_\theta(t)) \le 4\|\mathcal{G}_\theta(t)\|^2. \tag{305}$$

G.2 Proof of Lemma 5.1

We first prove the lower bound for one coefficient, which is Lemma 5.1, and we restate it as follows.

Lemma G.3 (Lower bound for single-parameter estimation)**.** *There is a two-sparse one-qubit family*

$$H_\theta = \Lambda(\cos\theta Z + \sin\theta X), \qquad |\theta| \le \frac{\pi}{6}, \tag{306}$$

such that the following holds. For every $0 < \varepsilon \le \Lambda/4$, any control-free protocol with deterministic total evolution time T_{tot} that estimates the X-coefficient to accuracy ε with success probability at least $2/3$ for all $|\theta| \le \pi/6$ must satisfy

$$T_{\mathrm{tot}} = \Omega\left(\frac{\Lambda}{\varepsilon^2}\right). \tag{307}$$

Proof. At $\theta = 0$, we have $H_0 = \Lambda Z$ and $\partial_\theta H_\theta\big|_{\theta=0} = \Lambda X$. For one shot of duration t,

$$\mathcal{G}_0(t) = \int_0^t e^{i\Lambda Z s}(\Lambda X)e^{-i\Lambda Z s}ds = \frac{\sin(2\Lambda t)}{2}X - \frac{1-\cos(2\Lambda t)}{2}Y. \tag{308}$$

Hence, we have $\|\mathcal{G}_0(t)\|^2 = \sin^2(\Lambda t) \le \Lambda t$. The same bound holds for every θ, since the family H_θ is just a rotation of H_0.

Now consider the full transcript of an adaptive protocol. Fisher information adds over shots, after conditioning on the previous transcript. Using Fact G.2 on each shot, $I(\theta) \le 4\Lambda T_{\mathrm{tot}}$ for every θ. Choose $\theta_\varepsilon \in (0, \pi/6]$ by $\Lambda \sin\theta_\varepsilon = 2\varepsilon$. At $+\theta_\varepsilon$, the X-coefficient is $+2\varepsilon$, while at $-\theta_\varepsilon$, it is -2ε. Therefore, any ε-accurate estimator can distinguish these two cases by the sign of its output. Since we require the success probability to be at least $2/3$, we need

$$\mathrm{TV}(P_{+\theta_\varepsilon}, P_{-\theta_\varepsilon}) \ge \frac{1}{3}. \tag{309}$$

On the other hand, Fact G.1 gives

$$\mathrm{TV}(P_{+\theta_\varepsilon}, P_{-\theta_\varepsilon}) \le \frac{1}{2}\int_{-\theta_\varepsilon}^{\theta_\varepsilon} \sqrt{4\Lambda T_{\mathrm{tot}}}d\xi = 2\theta_\varepsilon\sqrt{\Lambda T_{\mathrm{tot}}}. \tag{310}$$

Since $\sin x \ge x/2$ on $[0, \pi/6]$, we have $\theta_\varepsilon \le 2\sin\theta_\varepsilon = \frac{4\varepsilon}{\Lambda}$. Combining the last three displays gives $\frac{1}{3} \le \frac{8\varepsilon}{\Lambda}\sqrt{\Lambda T_{\mathrm{tot}}}$. This rearranges to

$$T_{\mathrm{tot}} = \Omega\left(\frac{\Lambda}{\varepsilon^2}\right). \tag{311}$$

□

G.3 Hard family

Next, we will prove that the log factor is also essential. The idea is to hide one label in each of M blocks. To recover the support, the learner must find all M hidden labels.

The next lemmas make this intuition precise for adaptive protocols. We account for the complete adaptive transcript through a coherent comparison state. Recovering all hidden labels from this state with constant success probability requires size $\Omega(M \log M)$.

We now define the hard family. Let $M, R \geq 2$, and set $r \coloneqq \lceil \log_2 R \rceil$. Use M marker qubits $m_1, \ldots, m_M$ and r shared label qubits. Choose R distinct commuting Hermitian Pauli strings $P_1, \ldots, P_R$ on the label register. For example, one may take distinct Z-type strings. For $a \in [M]$ and $b \in [R]$, define

$$Q_{a,b} \coloneqq \left(\prod_{c<a} Z_{m_c} \right) X_{m_a} P_b. \tag{312}$$

Let $q_{a,b} \in V$ be the Pauli label of $Q_{a,b}$. If $a < a'$, then $Q_{a,b}$ and $Q_{a',b'}$ anticommute on marker qubit m_a. All other tensor factors commute. Therefore, any two strings from different blocks anticommute. And $M \leq 2n + 1$ by Proposition 9 of Ref. [41]. For $x = (x_1, \ldots, x_M) \in [R]^M$ and $\sigma = (\sigma_1, \ldots, \sigma_M) \in \{0,1\}^M$, we define

$$H_{x,\sigma} \coloneqq \frac{\Lambda}{\sqrt{M}} \sum_{a=1}^{M} (-1)^{\sigma_a} Q_{a,x_a}. \tag{313}$$

The support label set is $S_x \coloneqq \{q_{a,x_a} : a \in [M]\}$. The selected Pauli strings pairwise anticommute, so $H_{x,\sigma}^2 = \Lambda^2 I$ and $\|H_{x,\sigma}\| = \Lambda$. The signs σ_a are included to hide the block signs. They do not change the support S_x. For $g \in \{0,1\}^M$, define

$$R_g \coloneqq \prod_{a=1}^{M} Z_{m_a}^{g_a}. \tag{314}$$

Then

$$R_g Q_{a,b} R_g^\dagger = (-1)^{g_a} Q_{a,b}, \qquad R_g H_{x,\sigma} R_g^\dagger = H_{x,\sigma+g}, \tag{315}$$

where addition is over $\mathbb{F}_2$.

The signs σ_a do not change the support S_x. They only choose whether the active Pauli string in block a has a plus sign or a minus sign. The identity above says that changing σ to $\sigma + g$ is the same as conjugating the whole experiment by R_g.

By observed data, we mean the list of all settings chosen so far and all outcomes observed so far. In the proof below, we condition on the complete observed data and simulate each shot pointwise in x, σ.

G.4 Proof of Theorem 5.2

For the rest of the proof, use the $R = 2$ subfamily with $P_1 = I$ and $P_2 = Z$. It uses $M + 1$ qubits; additional qubits may be padded by identities.

The protocol class in this section permits an arbitrary input state and an arbitrary terminal POVM in each shot, with one uninterrupted application of $e^{-iH\tau}$. All subsequent settings and stopping decisions may depend on the complete previous classical transcript. Between shots, the protocol retains only classical information.

Let $\mathcal{K}$ have orthonormal basis $\{|a, b\rangle : a \in [M],\ b \in \{1,2\}\}$, and write

$$|f_{x,\sigma}\rangle = \frac{1}{\sqrt{M}} \sum_{a=1}^{M} (-1)^{\sigma_a} |a, x_a\rangle. \tag{316}$$

For a unit vector $f \in \mathcal{K}$ and a real parameter $B \geq 0$, define the normalized coherent state on the symmetric Fock space over $\mathcal{K}$ by

$$|\sqrt{B} f\rangle_{\mathrm{coh}} = e^{-B/2} \sum_{k=0}^{\infty} \frac{B^{k/2}}{\sqrt{k!}} |f\rangle^{\otimes k}, \tag{317}$$

where the $k = 0$ term is the vacuum $|\mathrm{vac}\rangle$. The parameter B is an additive information resource in this comparison argument. For $0 \le \nu \le B$, a modewise beam-splitter isometry, independent of f, gives the splitting rule

$$|\sqrt{B}f\rangle_{\mathrm{coh}} \longmapsto |\sqrt{\nu}f\rangle_{\mathrm{coh}} \otimes |\sqrt{B-\nu}f\rangle_{\mathrm{coh}}. \tag{318}$$

For completeness, this follows by appending vacuum modes and applying the same two-mode rotation with amplitudes $\sqrt{\nu/B}$ and $\sqrt{1-\nu/B}$ to every mode. The case $B = 0$ is trivial.

Lemma G.4 (Pointwise simulation of one shot)**.** *Fix any complete classical history and the resulting next input state, duration τ, and terminal POVM. Put $\nu = \sin^2(\Lambda\tau)$. If $\nu > 0$, one copy of $|\sqrt{\nu}f_{x,\sigma}\rangle_{\mathrm{coh}}$ permits an exact simulation of this shot on a heralded successful branch of probability $p_\tau = e^{-\nu}/5 \ge 1/(5e)$, for every x, σ. The simulator is independent of x, σ. If $\nu = 0$, the shot can be simulated without a comparison resource.*

Proof. Purify the selected input to a known vector $|\psi\rangle$, and tensor all $Q_{a,b}$ with the identity on its purification register. Define

$$W_\psi|a,b\rangle = Q_{a,b}|\psi\rangle. \tag{319}$$

We first show $\|W_\psi\| \le 2$. For each fixed b the operators $Q_{a,b}$ are Hermitian and pairwise anticommute. Consequently, for any $z \in \mathbb{C}^M$,

$$\left\|\sum_a z_a Q_{a,b}\right\| \le \left(\sum_a (\mathrm{Re}\, z_a)^2\right)^{1/2} + \left(\sum_a (\mathrm{Im}\, z_a)^2\right)^{1/2} \le \sqrt{2}\|z\|_2. \tag{320}$$

Applying this to the two b-blocks and then using Cauchy–Schwarz gives $\|W_\psi\| \le 2$.

Write $c = \cos(\Lambda\tau)$ and $s = \sin(\Lambda\tau)$. On the vacuum and one-particle sectors define

$$K_{\psi,\tau}|\mathrm{vac}\rangle = \frac{c}{\sqrt{5}}|\psi\rangle, \tag{321}$$

$$K_{\psi,\tau}|a,b\rangle = -\frac{i\ \mathrm{sgn}(s)}{\sqrt{5}}Q_{a,b}|\psi\rangle, \tag{322}$$

and set $K_{\psi,\tau} = 0$ on all higher-particle sectors. For a vector with vacuum amplitude α and one-particle component z,

$$\|K_{\psi,\tau}(\alpha|\mathrm{vac}\rangle + z)\| \le \frac{|c||\alpha| + 2\|z\|}{\sqrt{5}} \le \sqrt{|\alpha|^2 + \|z\|^2}. \tag{323}$$

Thus $K_{\psi,\tau}$ is a valid successful Kraus operator. By $H_{x,\sigma}^2 = \Lambda^2 I$ and $\nu = s^2$,

$$K_{\psi,\tau}|\sqrt{\nu}f_{x,\sigma}\rangle_{\mathrm{coh}} = \frac{e^{-\nu/2}}{\sqrt{5}}\left(cI - \frac{is}{\sqrt{M}}\sum_a (-1)^{\sigma_a}Q_{a,x_a}\right)|\psi\rangle \tag{324}$$

$$= \frac{e^{-\nu/2}}{\sqrt{5}}e^{-iH_{x,\sigma}\tau}|\psi\rangle. \tag{325}$$

The target state has unit norm, proving the stated success probability. On success, performing the selected POVM, tensored with the identity on the purification register, gives exactly the original shot distribution. On failure, discard the extracted resource and retain only the failure flag. Both flag probabilities are independent of x, σ for this fixed history. The operation acts only on the extracted factor in Eq. (318); conditioned on x, σ, its remaining factor is unchanged after either flag. If $s = 0$, the evolution is a known global phase times the identity, so no comparison resource is needed. □

Lemma G.5 (Simulation of an adaptive transcript)**.** *Suppose a control-free protocol has deterministic total evolution-time budget T and succeeds with probability at least $2/3$ on every member of the hard family. There is a measurement of the single comparison state $|\sqrt{B}f_{x,\sigma}\rangle_{\mathrm{coh}}$, with $B = 30e\Lambda T$, that recovers x with probability at least $1/2$ whenever success of the original protocol determines x.*

Proof. The case $T = 0$ cannot satisfy the success assumption: its transcript is independent of x, whose uniform-prior recovery probability is at most 2^{-M}. Thus $T > 0$. First consider a simulator with unlimited comparison resource. Condition on the complete raw transcript, including the learner's private randomness. For the next duration τ_t, repeatedly apply Lemma G.4 using fresh resource of size $\nu_t = \sin^2(\Lambda\tau_t)$ until it succeeds. Each retry uses the same selected input, duration, and POVM. The retries are internal to the simulator: their failure flags are not passed to the simulated learner and do not advance its transcript. Its success probability is independent of the hidden parameters, so failures do not change their posterior and the accepted outcome has exactly the required conditional law. This proves exact simulation by induction. The expected resource consumed in that round is

$$\frac{\nu_t}{p_{\tau_t}} = 5e^{\nu_t}\nu_t \le 5e\nu_t. \tag{326}$$

Let $\mathcal{R}$ be the sum of the resource parameters over all attempts. Because $\sin^2 u \le u$ for $u \ge 0$ and the original protocol obeys its budget on every transcript,

$$\mathbb{E}[\mathcal{R}] \le 5e\,\mathbb{E}\sum_t \sin^2(\Lambda\tau_t) \le 5e\Lambda T. \tag{327}$$

These bounds hold pointwise in x, σ.

Now start with $B = 30e\Lambda T$. At each attempt, use Eq. (318) to extract the requested resource, and keep the remaining factor for later rounds. The splitting operation depends only on the known remaining budget and the current classical history. Abort if a request exceeds the remaining budget. Coupling with the unlimited simulation, the probability of such an abort is at most

$$\Pr[\mathcal{R} > B] \le \frac{\mathbb{E}[\mathcal{R}]}{B} \le \frac{1}{6}. \tag{328}$$

The resulting simulation therefore succeeds with probability at least $2/3 - 1/6 = 1/2$. Every operation used is independent of the hidden parameters, so the entire simulation and decoder form a POVM on the initial comparison state. □

Lemma G.6 (Decoding a coherent support string)**.** *Let x be uniform on $\{1,2\}^M$. Even if σ is given to the decoder, the optimal probability of recovering all of x from $|\sqrt{B}f_{x,\sigma}\rangle_{\mathrm{coh}}$ is*

$$p_{\mathrm{all}}(B) = \left(\frac{1+\sqrt{1-e^{-2B/M}}}{2}\right)^M \le \exp\left(-\frac{M}{4}e^{-2B/M}\right). \tag{329}$$

For $M \ge 16$, success probability at least $1/2$ requires $B \ge (M/4)\log M$.

Proof. Given σ, apply its known phase to both modes in each block. The coherent state then factorizes across the M blocks. With $\alpha = \sqrt{B/M}$, the two equiprobable block states are $|\alpha, 0\rangle_{\mathrm{coh}}$ and $|0, \alpha\rangle_{\mathrm{coh}}$, whose overlap is $e^{-B/M}$. Write their density matrices as ρ_1, ρ_2 and let

$$Y = \frac{\rho_1 + \rho_2 + |\rho_1 - \rho_2|}{4}. \tag{330}$$

Then $Y \succeq \rho_b/2$ for $b = 1, 2$ and

$$\mathrm{Tr}\, Y = \frac{1}{2} + \frac{1}{4}\|\rho_1 - \rho_2\|_1 = \frac{1+\sqrt{1-e^{-2B/M}}}{2} =: p. \tag{331}$$

For every string x, tensor-product order gives $Y^{\otimes M} \succeq 2^{-M}\bigotimes_a \rho_{x_a}$. Hence every joint POVM $\{D_x\}$ satisfies

$$2^{-M}\sum_x \mathrm{Tr}\left(D_x \bigotimes_a \rho_{x_a}\right) \le \sum_x \mathrm{Tr}(D_x Y^{\otimes M}) = p^M. \tag{332}$$

The product of optimal binary measurements attains this bound. The inequality in Eq. (329) follows from $\sqrt{1-z} \le 1 - z/2$ and $1 - z \le e^{-z}$. Thus $p_{\mathrm{all}}(B) \ge 1/2$ implies

$$B \ge \frac{M}{2}\log\frac{M}{4\log 2} \ge \frac{M}{4}\log M, \qquad M \ge 16. \tag{333}$$

□

We now put the above pieces together. We choose M so that each active coefficient has size about ε. Then support recovery is the same as recovering the hidden vector x. Now, we prove Theorem 5.2, which is restated as follows.

Theorem G.7 (Lower bound for ε-detectable support)**.** *There is an absolute constant $c > 0$ such that the following holds. Let $0 < \varepsilon \le \Lambda/16$, and assume the number of qubits is large enough to embed the hard family in Eq.* (313)*, for example $n \ge C_0 \frac{\Lambda^2}{\varepsilon^2}$ for a sufficiently large absolute constant C_0. Assume also that the sparsity bound in Problem 2.1 is at least $m \coloneqq \lfloor(\Lambda/(4\varepsilon))^2\rfloor$. Then any control-free protocol with deterministic total evolution-time budget T_{tot} that solves Problem 2.1 with success probability at least $2/3$ must have total evolution time*

$$T_{\mathrm{tot}} \ge c\frac{\Lambda}{\varepsilon^2}\log\frac{\Lambda}{\varepsilon} \tag{334}$$

where one may take $c = 2^{-14}$.

Proof. We set $M \coloneqq \lfloor(\frac{\Lambda}{4\varepsilon})^2\rfloor$ and $R \coloneqq 2$. Since $0 < \varepsilon \le \Lambda/16$, we have $M \ge 16$ and $\frac{\Lambda^2}{32\varepsilon^2} \le M \le \frac{\Lambda^2}{16\varepsilon^2}$. The family uses $M + 1 \le 2M$ qubits.

For every Hamiltonian in the hard family, each active coefficient has a magnitude $\frac{\Lambda}{\sqrt{M}} \ge 4\varepsilon$. All inactive coefficients are zero. On the event of coefficientwise error at most ε, the set

$$\widetilde{S} = \{u \ne 0 : |\widehat{\lambda}_u| > 2\varepsilon\} \tag{335}$$

equals S_x: active estimates have magnitude at least 3ε and inactive estimates have magnitude at most ε. Since $x \mapsto S_x$ is one-to-one, support recovery is the same as recovering x.

Put the uniform prior on x and on the signs σ. A worst-case success probability of at least $2/3$ implies an average success probability of at least $2/3$. Lemma G.5 gives a decoder of the comparison state with $B = 30e\Lambda T_{\mathrm{tot}}$ and success probability at least $1/2$. Giving σ to this decoder can only help. Applying Lemma G.6 yields

$$T_{\mathrm{tot}} \ge \frac{M\log M}{120e\Lambda}. \tag{336}$$

Finally, note that $M \ge \frac{\Lambda^2}{32\varepsilon^2}$ and $\log M \ge \frac{3}{4}\log\frac{\Lambda}{\varepsilon}$, where the second inequality uses $\varepsilon \le \Lambda/16$. Therefore

$$T_{\mathrm{tot}} \ge \frac{1}{5120e}\frac{\Lambda}{\varepsilon^2}\log\frac{\Lambda}{\varepsilon} \ge 2^{-14}\frac{\Lambda}{\varepsilon^2}\log\frac{\Lambda}{\varepsilon}. \tag{337}$$

□

H Numerical simulations

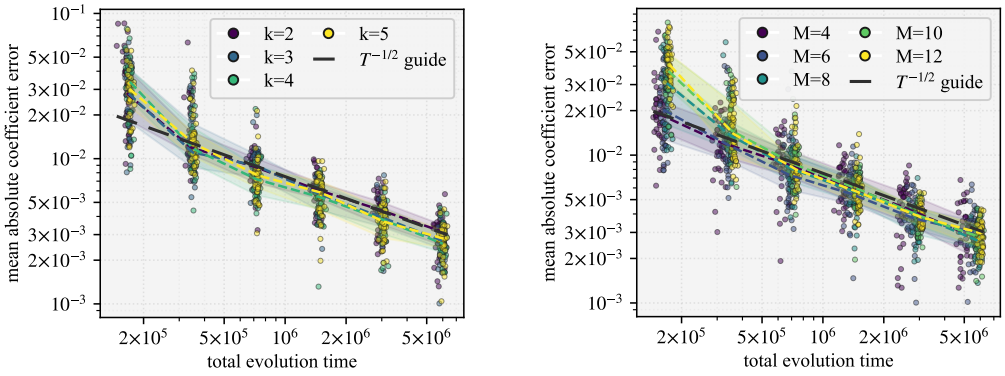

Figure 2: Mean absolute coefficient error versus total evolution time for Algorithm 2 in learning random k-local Hamiltonians (left) and random ansatz-free sparse Hamiltonians (right). The black dashed line is a $T^{-1/2}$ guide.

Here, we validate the complete protocol in Algorithm 2 by numerical simulations. Each data point runs the same two-stage reconstruction pipeline as Algorithm 2: the same-Pauli displacement sieve first constructs the candidate set $\widehat{\mathcal{U}}_{\mathrm{proj}}$, and the cross-Pauli parity estimator then estimates all coefficients in that candidate set. The plotted error metric is the mean absolute coefficient error over the true support set S defined as

$$\mathrm{MAE}_S = \frac{1}{|S|} \sum_{u \in S} |\widehat{\lambda}_u - \lambda_u|, \tag{338}$$

where missed true support elements are counted with $\widehat{\lambda}_u = 0$. It penalizes both missed support and coefficient-estimation error.

For each Hamiltonian instance, we normalize $\|H\| \leq \Lambda$ with $\Lambda = 1$, run Algorithm 2 at a fixed detection scale ε, and vary the sampling budget through the constants multiplying the projection and parity-estimation block sizes. For the unrestricted sparse panel, we use $n = 5$, $M \in \{4, 6, 8, 10, 12\}$, six budget values, and 40 independent trials per (M, budget) pair, for a total of $5 \times 6 \times 40 = 1200$ complete reconstructions. For the exact-locality panel, we use $n = 6$, $k \in \{2, 3, 4, 5\}$, fixed sparsity $M_{\mathrm{loc}} = 8$, the same six budget values, and 40 independent trials per (k, budget) pair, for a total of $4 \times 6 \times 40 = 960$ complete reconstructions. The observed error follows the expected $T^{-1/2}$ trend for both unrestricted sparse and k-local data-generating Hamiltonians.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims in the abstract is the same as the claims in the introduction. We provide pointers with the main claims in the introduction to the result we rigorously proved in the paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations in Section 6.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: We provide proof guidelines in the manuscript and the rigorous proof for every claim we made in the appendix.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Appendix H reports the simulations.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: The simulation code and raw numerical data are not provided.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Appendix H gives the system sizes, normalization and trial counts.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: Only forty independent trials per parameter–budget pair are reported.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)

- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: Hardware specifications are not reported.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This paper focuses on the theory of Hamiltonian learning.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.