# Quantum computing, norms and polynomials

Escudero Gutiérrez, F.

# Chapter 7

# Testing and learning quantum Hamiltonians

## 7.1 Introduction

A fundamental and important challenge with building quantum devices is being able to characterize and calibrate its behavior. One approach to do so is *Hamiltonian learning* which seeks to learn the Hamiltonian governing the dynamics of a quantum system given finite classical and quantum resources. Beyond system characterization, it is also carried out during validation of physical systems and designing control strategies for implementing quantum gates [IBF$^+$20]. However, learning an $n$-qubit Hamiltonian is known to be difficult, requiring complexity that scales exponential in the number of qubits unless a coarse metric is used [Car23].

In practice, however, prior knowledge on the structure of Hamiltonians is available e.g., those of engineered quantum devices [SMCG16] where the underlying Hamiltonians primarily involve local interactions with few non-local interactions, and even naturally occurring physical quantum systems such as those with translationally invariant Hamiltonians. To highlight these structural properties, consider an $n$-qubit Hamiltonian $H$ (which is a self-adjoint operator acting on $(\mathbb{C}^2)^{\otimes n}$) expanded in terms of the $n$-qubit Pauli operators:

$$H = \sum_{x \in \{0,1,2,3\}^n} \lambda_x \sigma_x,$$

We call the set of Paulis with non-zero coefficients $\lambda_x$ as the Pauli spectrum of the Hamiltonian denoted by $\mathcal{S} = \{x \in \{0,1,2,3\}^n : \lambda_x \neq 0\}$. Of particular relevance are *k-local* Hamiltonians which involve Pauli operators that act non-trivially on all but at most $k$ qubits and *s-sparse* Hamiltonians whose Pauli expansion contains at most $s$ non-zero Pauli operators i.e., $|\mathcal{S}| \leq s$.

There has thus been a growing suite of Hamiltonian learning results that have shown that when the underlying $n$-qubit Hamiltonian $H$ satisfies these structural properties, learning can be performed with only $\text{poly}(n)$ query complexity, either by making "queries" to the unitary evolution operator $U(t) = \exp(-iHt)$ [dSLCP11, HBCP15, ZYLB21, HKT22, YSHY23, DOS23, HTFS23, LTN$^+$23, SFMD$^+$24, GCC24, Zha24, HMG$^+$25], or by assuming one has access to Gibbs state [AAKS21, HKT22, RSF23, ORSFW23, BLMT23, GCC24]. Notably, [BLMT24] considered the problem of learning Hamiltonians that are both local and sparse, without prior knowledge of the support. Several of the learning algorithms mentioned above however require assumptions on the support of the Hamiltonian beyond locality or sparsity, such as [HTFS23] which considers *geometrically-local* Hamiltonians (a subset of local Hamiltonians) and [YSHY23] which requires assumptions on the support.

Moreover, before learning, it might be desirable to uncover *what is the structure* of an unknown Hamiltonian in order to choose specialized learning algorithms. Even deciding if a Hamiltonian has a particular structure is a fundamental challenge and constitutes the problem of *testing* if an unknown Hamiltonian satisfies a certain structural property. This line of investigation is nascent with only a few works on Hamiltonian *property* testing [SY23, ACQ22, LW22] with Blum et al. [BCO24b] having considered the problem of testing local Hamiltonians and the problem of testing sparse Hamiltonians yet to be tackled. This leads us to the motivating question of this chapter:

*What is the query complexity of learning and testing structured Hamiltonians?*

### Problem statement

Before we state our results answering the question above, we clearly mention our learning and testing problems first. If $H$ is the Hamiltonian describing the dynamics of a certain physical system, then the state of that system evolves according to the *time evolution operator* $U(t) = e^{-iHt}$. This means that if $\rho(0)$ is the state at time 0, at time $t$ the state would have evolved to $\rho(t) = U(t)\rho(0)U^\dagger(t)$. Hence, to test and learn a Hamiltonian one can do the following: prepare a desired state, apply $U(t)$ or tensor products of $U(t)$ with identity to the state, and finally measure in a chosen basis. From here onwards, this is what we mean by *querying* the unitary $U(t)$. It is usual to impose the normalization condition $\|H\|_{\mathrm{op}} \leq 1$ (i.e., that the eigenvalues of $H$ are bounded in absolute value by 1). We will assume this normalization unless otherwise stated, but we will also work out the dependence on $\|H\|_{\mathrm{op}}$ for our learning algorithms. Throughout this paper, we will consider the normalized Frobenius norm as the distance between Hamiltonians, unless otherwise stated. This distance is

$$d(H, H') = \|H - H'\|_2 = \sqrt{\frac{\mathrm{Tr}[(H - H')^2]}{2^n}},$$

and it equals the $\ell_2$-norm of the Pauli spectrum, $d(H, H') = \sqrt{\sum |\lambda_x - \lambda'_x|^2}$. A *property* of a Hamiltonian, denoted $\mathcal{H}$ is a class of Hamiltonians that satisfy the property (here we will be interested in sparse and local properties). We say that $H$ is $\varepsilon$-far from having a property $\mathcal{H}$ if $d(H, H') > \varepsilon$ for every $H' \in \mathcal{H}$, and otherwise is $\varepsilon$-close. Now, we are ready to state the testing and learning problems.

Let $\mathcal{H}$ be a property and let $H$ be an unknown Hamiltonian with $\|H\|_{\mathrm{op}} \leq 1$ and $\mathrm{Tr}[H] = 0$.

**Problem 7.1** (Tolerant testing)**.** Promised $H$ is either $\varepsilon_1$-close or $\varepsilon_2$-far from satisfying property $\mathcal{H}$, decide which is the case by making queries to $U(t)$.

**Problem 7.2** (Hamiltonian learning)**.** Promised $H \in \mathcal{H}$, output a classical description of $\widetilde{H} \in \mathcal{H}$ such that $\|H - \widetilde{H}\|_2 \leq \varepsilon$ by making queries to $U(t)$.

### Summary of results

The main results of this chapter are query-efficient algorithms for testing and learning Hamiltonians that are local and/or sparse. We summarize our results in Table 7.1 (for simplicity we state our results for constant accuracy). Before we discuss our results in more detail, we make a few remarks about our main results.

| | **Testing** | **Learning** |
|---|---|---|
| $s$-sparse | poly($s$) | poly($s$) |
| $k$-local | $O(1)$ | $\exp(k^2)$ |

**Table 7.1:** Query complexity for learning and testing $n$-qubit structured Hamiltonians.

(*i*) As far as we know, by the time of the publication, the results of this chapter are the first: (*a*) with complexities that are *independent* of $n$ [1], and (*b*) that does not assume knowledge of the support.[2]

(*iv*) We give the first learning algorithm for Hamiltonians that are only promised to be sparse, and not necessarily local. Similarly, our local Hamiltonian learning problem doesn't assume geometric locality which was assumed in several prior works.

(*iii*) Our testing algorithms are tolerant, i.e., they can handle the setting where $\varepsilon_1 \neq 0$. As far as we know, there are only a handful of polynomial-time tolerant testers for quantum objects.

(*iv*) Our learning algorithms are based on a subroutine that learns arbitrary $n$-qubit Hamilotmians with $O(1/\varepsilon^4)$ queries, albeit in the coarser metric of the $\ell_\infty$-norm of the Pauli coefficients. As far as we know, this is the only best result for unstructured Hamiltonians. Notably, it is also the first time-efficient proposal for this problem.

We remark that most previous works on Hamiltonian learning (that we highlighted earlier) are done under the distance induced by the supremum norm of the Pauli spectrum and with extra constraints apart from locality [dSLCP11, HBCP15, ZYLB21, HKT22, WKR$^+$22, YSHY23, Car23, DOS23, HTFS23, LTN$^+$23, SFMD$^+$24, GCC24]. When transformed into learning algorithms under the finer distance induced by the $\ell_2$-norm of the Pauli spectrum, these proposals yield complexities that depend polynomially on $n^k$ and only work for a restricted family of $k$-local Hamiltonians. The works that explicitly consider the problem of learning under the $\ell_2$-norm have complexities depending on $n$ and assume a stronger access model [CW23, BLMT24].

### Results

**Testing.** Recently, Bluhm, Caro and Oufkir proposed a non-tolerant testing algorithm, meaning that it only works for the case $\varepsilon_1 = 0$, whose query complexity is $O(n^{2k+2}/(\varepsilon_2 - \varepsilon_1)^4)$ and with total evolution time $O(n^{k+1}/(\varepsilon_2 - \varepsilon_1)^3)$. They posed as

[1] There are a few works that achieve $n$-independent complexities for learning local Hamiltonians in the $\infty$-norm of the Pauli coefficients, but when transformed into 2-norm learners they yield complexities depending on $n^k$.

[2] Soon after [Esc24b], Bakshi et al. [BLMT24] presented a learning algorithm that does not require prior knowledge of the support, achieving Heisenberg scaling using heavy machinery.

open questions whether the dependence on $n$ could be removed and whether an efficient tolerant-tester was possible [BCO24a, Section 1.5]. Our first result gives positive answer to both questions.

**Result 7.3.** *There is an algorithm that solves Problem 7.1 for k-local Hamiltonians by making* $\mathrm{poly}(1/(\varepsilon_2-\varepsilon_1))$ *queries to the evolution operator and with* $\mathrm{poly}(1/(\varepsilon_2-\varepsilon_1))$ *total evolution time.*

See Theorem 7.12 for a formal statement of this result. Our algorithm to test for locality is simple. It consists of repeating the following process $1/(\varepsilon_2-\varepsilon_1)^4$ times: prepare $n$ EPR pairs, apply $U(\varepsilon_2-\varepsilon_1)\otimes \mathrm{Id}_{2^n}$ to them and measure in the Bell basis. Each time that we repeat this process, we sample from the Pauli sprectrum of $U(\varepsilon_2-\varepsilon_1)$.[3] As $\varepsilon_2-\varepsilon_1$ is small, Taylor expansion ensures that $U(\varepsilon_2-\varepsilon_1)\approx \mathrm{Id}_{2^n}-i(\varepsilon_2-\varepsilon_1)H$, so sampling from the Pauli spectrum of $U(\varepsilon_2-\varepsilon_1)$ allows us to estimate the weight of the non-local terms of $H$. If that weight is big, we output that $H$ is far from $k$-local, and otherwise we conclude that $H$ is close to $k$-local.

Despite the numerous papers in the classical literature studying the problems of testing and learning sparse Boolean functions [GOS+11, NS12, YZ20, EIS22], there are not many results on learning Hamiltonians that are sparse (and not necessarily local) and the only testing result that we are aware of requires $O(sn)$ queries [BCO24b, Remark B.2]. Here, we present the first sparsity testing algorithm whose complexity does not depend on $n$ and the first learning algorithm for sparse Hamiltonians which does not make any assumptions regarding the support of the Hamiltonian beyond sparsity.

**Result 7.4.** *There is an algorithm that solves Problem 7.1 for s-sparse Hamiltonians by making* $\mathrm{poly}(s/(\varepsilon_2-\varepsilon_1))$ *queries to the evolution operator and with* $\mathrm{poly}(s/(\varepsilon_2-\varepsilon_1))$ *total evolution time.*

See Theorem 7.15 for a formal statement. This testing algorithm consists on performing Pauli sampling of $U(\sqrt{(\varepsilon_2^2-\varepsilon_1^2)/s})$ a total of $O(s^4/(\varepsilon_2^2-\varepsilon_1^2)^4)$ times. From these samples one can estimate the sum of the squares of the top $s$ Pauli coefficients of $U$. If this quantity is big enough, we output that the Hamiltonian is close to $s$-sparse, and otherwise that is far. Although from this high-level description the algorithm seems similar to the locality testing one, the analysis is more involved and requires taking the second order Taylor expansion, which is the reason why the dependence on $(\varepsilon_2-\varepsilon_1)$ is worse in this case.

[3] The Pauli spectrum of a unitary $U=\sum_x \widehat{U}_x\sigma_x$ determines a probability distribution because $\sum_x |\widehat{U}_x|^2=1$.

Additionally, we provide a sparsity tester (Theorem 7.16) that only makes $O(s^2/\varepsilon_2^4)$ queries with $O(s^{1.5}/\varepsilon_2^3)$ total evolution time, but only works in the regime $\varepsilon_1 = O(\varepsilon_2/\sqrt{s})$.

**Learning.** We first propose a protocol to learn unstructured Hamiltonians efficiently in the coarser $\ell_\infty$ norm of the Pauli coefficients. Then, we turn it into a learner in the $\ell_2$ norm for local and sparse Hamiltonians.

**Result 7.5.** *There is an algorithm that outputs estimates $\widetilde{\lambda}_x$ such that $|\lambda_x - \widetilde{\lambda}_x| \leq \varepsilon$ for every $x \in \{0,1,2,3\}^n$ by making $O(1/\varepsilon^4)$ queries to the evolution operator with $O(1/\varepsilon^3)$ total evolution time.*

See Theorem 7.18 for a formal result. The learning algorithm has two stages. In the first stage one samples from the Pauli distribution of $U(\varepsilon)$, as in the testing algorithm, and from that one can detect which are the big Pauli coefficients of $H$. In the second stage we learn the big Pauli coefficients via a novel subroutine based on Clifford Shadows (see Lemma 7.17) and set the small to 0.

For Hamiltonians that are $k$-local, we have the following learning result in the $\ell_2$-norm.

**Result 7.6.** *There is an algorithm that solves Problem 7.2 for $k$-local Hamiltonians by making* $\exp(k^2 + k\log(1/\varepsilon))$ *queries to the evolution operator with* $\exp(k^2 + k\log(1/\varepsilon))$ *total evolution time.*

See Theorem 7.19 for a formal statement of this result. In the case that the Hamiltonian is $k$-local, one can ensure that the coefficients not detected as big in the first stage of the algorithm of Result 7.5 have a neglectable contribution to the $\ell_2$-norm, from which Result 7.6 follows. To argue this formally, we use the non-commutative Bohnenblust-Hille inequality, which has been used recently for various quantum learning algorithms [HCP23b, VZ23].

For Hamiltonians that are $s$-sparse, we have the following learning result in the $\ell_2$-norm.

**Result 7.7.** *There is an algorithm that solves Problem 7.2 for $s$-sparse Hamiltonians by making* $\mathrm{poly}(s/\varepsilon)$ *queries to the evolution operator with* $\mathrm{poly}(s/\varepsilon)$ *total evolution time.*

See Theorem 7.21 for a formal statement. Result 7.7 follows by adding a rounding step to the algorithm of Result 7.5 that ensures that all zero coefficients of the Hamiltonians are also zero for the approximating Hamiltonian.

**Direct comparison to previous work.** Comparing the plethora of Hamiltonian learning algorithms can be challenging due to the different assumptions on the structure of the Hamiltonians (local, sparse, geometrical structures, etc.), the different distances to measure the error ($\ell_\infty$ norm of the coefficients, $\ell_2$ norm, etc.), the different complexity measures (queries, total evolution time, number of experiments, etc.), the different access models (coherent/non-coherent queries, with/without memory, etc.) and the different goals of the algorithm (minimizing the dependence on the dimension parameters like $n, s, k$, achieving the Heisenberg scaling $1/\varepsilon$, etc.). Thus, we only include a direct comparison in Table 7.2 with the works that explicitly consider the same structure and the same error metric as us. As a summary, one can say that for constant $\varepsilon$ our results achieve better dependence on the parameters $n, s, k$ than previous work, while also using the weaker model of incoherent queries, where one can perform only one query before measuring, as opposed to the coherent query model. We also want to remark that our result for learning unstructured Hamiltonian is time efficient, while the, to the best of our knowledge, only previous one is not [Car23].

| Hamiltonians | Reference | $t_{total}$ | Queries | Access model |
|---|---|---|---|---|
| Unstructured, $\ell_\infty$ error | [Car23] | $n/\epsilon^4$ | $n/\epsilon^4$ | Coherent queries |
| | Theorem 7.18 | $1/\varepsilon^3$ | $1/\epsilon^4$ | Incoherent queries |
| $s$-sparse, $\ell_\infty$ error | [Zha24]* | $1/\epsilon^4$ | $1/\epsilon^8$ | Coherent queries |
| | [HMG+25]† | $s^2/\epsilon$ | $s^2/\epsilon$ | Coherent queries |
| | Theorem 7.21 | $1/\varepsilon^3$ | $1/\epsilon^4$ | Incoherent queries |
| $k$-local, $\ell_2$ error | [CW23] | $n^k/\epsilon^2$ | $n^k/\epsilon^2$ | Controlled and inverse queries |
| | [MFPT24]° | $(9n)^k/\epsilon$ | $(27n^3)^k/\epsilon^2$ | Coherent queries |
| | Theorem 7.19 | $\exp(k^2)/\varepsilon^k$ | $\exp(k^2)/\varepsilon^k$ | Incoherent queries |

**Table 7.2:** Comparison of algorithms for learning Hamiltonians with $\|H\|_{\text{op}} \leq 1$.
* It can be improved to $O(1/\varepsilon^{2+o(1)})$ total evolution time and $O(1/\varepsilon^{6+o(1)})$ queries by paying huge constant factors.
† This algorithm works for Hamiltonians with $\sup_x |\lambda_x| \leq 1$, a weaker constraint than $\|H\|_{\text{op}} \leq 1$.
° This algorithm is the only one in the table that uses no quantum memory. We provide an algorithm with no quantum memory for $k$-local learning that performs as the one in the last row, but with an extra factor $\log n$.

**Note added.** After sharing Theorem 7.12 with Bluhm et al., they independently improved the analysis of their testing algorithm and showed that it only requires $O(1/(\varepsilon_2 - \varepsilon_1)^3 \varepsilon_2)$ queries and $O(1/(\varepsilon_2 - \varepsilon_1)^{2.5} \varepsilon_2^{0.5})$ total evolution time, which is very similar to our Theorem 7.12 [BCO24b]. In addition, for a wide range of $k = O(n)$, their algorithm does not require the use of auxiliary qubits.

# 7.2 Preliminaries

### Notation

Every $n$-qubit operator $H$ can be written down in its Pauli decomposition as

$$H = \sum_{x \in \{0,1,2,3\}^n} \lambda_x \sigma_x,$$

where the real-valued coefficients $\lambda_x$ are given by $\lambda_x = \frac{1}{2^n} \operatorname{Tr}(H\sigma_x)$. Parseval's identity states that the normalized Frobenius norm of $H$ (denoted as $\|H\|_2$) equals the $\ell_2$-norm of its Pauli spectrum, i.e.,

$$\|H\|_2 = \sqrt{\frac{\operatorname{Tr}[H^\dagger H]}{2^n}} = \sqrt{\sum_{x \in \{0,1,2,3\}^n} |\lambda_x|^2}.$$

We will repeatedly use that $\|H\|_2 \leq \|H\|_{\mathrm{op}}$, which holds because $\|H\|_2^2$ is the average of the squares of the eigenvalues of $H$. We will also consider the $\ell_\infty$ norm of the Pauli coefficients of an operator, which is given by

$$\|H\|_{\ell_\infty} = \sup_x |h_x|.$$

Additionally, we will use $\|H\| := \max\{\|H\|_{\mathrm{op}}, 1\}$.

Given $x \in \{0,1,2,3\}^n$, define $|x|$ as the number of indices $i \in [n]$ where $x_i \neq 0$, define

$$H_{>k} = \sum_{|x|>k} \lambda_x \sigma_x$$

and $H_{\leq k}$ as $\sum_{|x| \leq k} \lambda_x \sigma_x$. From the formulation of the 2-norm in terms of the Pauli coefficients it follows that $\|H_{>k}\|_2 \leq \|H\|_2$. We note that the distance of a Hamiltonian $H$ from the space of $k$-local Hamiltonians is given by $\|H_{>k}\|_2$, as $H_{\leq k}$ is the $k$-local Hamiltonian closest to $H$. The $\ell_2$-distance of $H$ to being $s$-sparse also has a nice expression. Assign labels from $[4^n]$ to $x \in \{0,1,2,3\}^n$ in a way that and $|\lambda_{x_1}| \geq |\lambda_{x_2}| \cdots \geq |\lambda_{x_{4^n}}|$. Then, $\sum_{i \in [s]} \lambda_{x_i} \sigma_{x_i}$ is the closest $s$-sparse Hamiltonian to $H$, so the $\ell_2$-distance of $H$ to the space of $s$-sparse Hamiltonians is $\sqrt{\sum_{i>s} |\lambda_{x_i}|^2}$.

### Necessary subroutines

Suppose $U$ is a unitary and we write out its Pauli decomposition as $U = \sum_x \widehat{U}_x \sigma_x$, then by Parseval's identity $\sum |\widehat{U}_x|^2 = \operatorname{Tr}[U^\dagger U]/2^n = 1$, i.e., $\{|\widehat{U}_x|^2\}_x$ is a *probability* *distribution.* We will be using the fact below extensively.

*Fact* 7.8. Given access to a unitary $U$, one can sample from the distribution $\{|\widehat{U}_x|^2\}_x$.

*Proof.* The proof simply follows by applying $U \otimes \mathrm{Id}_{2^n}$ to $n$ EPR pairs (i.e., preparing the Choi-Jamiolkowski state of $U$) and measuring in the Bell basis, because

$$U \otimes \mathrm{Id}_{2^n} |\mathrm{EPR}_n\rangle = \sum_{x \in \{0,1,2,3\}^n} \widehat{U}_x \bigotimes_{i \in [n]} (\sigma_{x_i} \otimes \mathrm{Id}_2 |\mathrm{EPR}\rangle),$$

and the Bell states can be written as $\sigma_x \otimes \mathrm{Id}_2 |\mathrm{EPR}\rangle$ for $x \in \{0,1,2,3\}$. □

We will also use that given a Hamiltonian $H$, the Taylor expansion of the exponential allows us to approximate the time evolution operator as

$$U(t) = e^{-itH} = \mathrm{Id}_{2^n} - itH + ct^2 R_1(t) \|H\|_{\mathrm{op}}^2 \tag{7.1}$$

for $t \leq 1/2$, where the first order remainder $R_1(t)$ is bounded $\|R_1(t)\|_{\mathrm{op}} \leq 1$ and $c > 1$ is a universal constant.

We will also use the celebrated Classical Shadows by Huang, Chen and Preskill.

**Theorem 7.9** (Clifford shadows [HKP20])**.** *Let $\rho$ be an $n$-qubit state and let $\{O_i\}_{i \in [M]}$ be $n$-qubit traceless observables. Assume that $\sup_i \mathrm{Tr}[O_i^2] = O(1)$. Then, Algorithm 1 obtains estimates $\widetilde{O}_{i,\rho}$ such that, with probability $1-\delta$, satisfy*

$$|\mathrm{Tr}[O_i \rho] - \widetilde{O}_{i,\rho}| \leq \varepsilon$$

*for every $i \in [M]$. The algorithm uses $O\left(\frac{\log(M/\delta)}{\varepsilon^2}\right)$ copies of $\rho$.*

## 7.3 Technical results

In this section, we will first prove our main structural theorems for Hamiltonians and provide subroutines which will be used later for testing and learning these structured Hamiltonians.

### Structural lemma for local Hamiltonians

First, we prove a lemma regarding the discrepancy on the weights of non-local terms of the short-time evolution operator for close-to-local and far-from-local Hamiltonians.

**Algorithm 1** Clifford shadows

**Input:** Copies of a quantum state $\rho$, target set of observables $\{O_i\}_{i\in[M]}$, error parameter $\varepsilon \in (0,1)$, and failure parameter $\delta \in (0,1)$

```
Set T = O(log(M/δ)/ε²) and J = O(log(M/δ))
for j ∈ [J] do
    for k ∈ [T/J] do
        Apply a uniformly random Clifford gate C to a copy of ρ
        Measure in the computational basis. Let |b_{j,k}⟩ be the outcome
        for i ∈ [M] do
            Let Õ_{i,j,k} = (2^n + 1)⟨b_{j,k}|C^{-1} O_i C|b_{j,k}⟩
        end for
    end for
    for i ∈ [M] do
        Let Õ_{i,j} = Mean((Õ_{i,j,k})_k)
    end for
end for
for i ∈ [M] do
    Set Õ_i := Median((O_{i,j})_j)
end for
```

**Output**: $(\widetilde{O}_i)_{i\in[M]}$

**Lemma 7.10.** *Let* $0 \leq \varepsilon_1 < \varepsilon_2$. *Let* $\alpha = (\varepsilon_2 - \varepsilon_1)/(3c)$ *and* $H$ *be an* $n$*-qubit Hamiltonian with* $\|H\|_{\mathrm{op}} \leq 1$, *where* $c$ *is the constant appearing in Taylor expansion (see Eq.* (7.1)*). If* $H$ *is* $\varepsilon_1$*-close* $k$*-local, then*

$$\|U(\alpha)_{>k}\|_2 \leq (\varepsilon_2 - \varepsilon_1)\frac{2\varepsilon_1 + \varepsilon_2}{9c},$$

*and if* $H$ *is* $\varepsilon_2$*-far from being* $k$*-local, then*

$$\|U(\alpha)_{>k}\|_2 \geq (\varepsilon_2 - \varepsilon_1)\frac{\varepsilon_1 + 2\varepsilon_2}{9c}.$$

*Proof.* Recall that $U(\alpha) = \mathrm{Id}_{2^n} - i\alpha H + c\alpha^2 R(\alpha)$ by Eq (7.1) where $\|R\|_{\mathrm{op}} \leq 1$. For simplicity, we set $U = U(\alpha)$ and $R = R_1(\alpha)$. First, assume that $H$ is $\varepsilon_1$-close $k$-local, then by definition we have that $\|H_{>k}\|_2 \leq \varepsilon_1$. Then

$$\|U_{>k}\|_2 \leq \alpha\|H_{>k}\|_2 + c\alpha^2\|R_{>k}\|_2 \leq \frac{\varepsilon_2 - \varepsilon_1}{3c}\varepsilon_1 + c\left(\frac{\varepsilon_2 - \varepsilon_1}{3c}\right)^2 = (\varepsilon_2 - \varepsilon_1)\frac{2\varepsilon_1 + \varepsilon_2}{9c},$$

where in the first inequality we have used the triangle inequality, and in the second that $H$ is $\varepsilon_1$-close to $k$-local and that $\|R_{>k}\|_2 \leq \|R\|_2 \leq \|R\|_{\mathrm{op}} \leq 1$. Now, assume that $H$ is $\varepsilon_2$-far from being $k$-local (i.e., $\|H_{>k}\|_2 \geq \varepsilon_2$). Then

$$\|U_{>k}\|_2 \geq \alpha\|H_{>k}\|_2 - c\alpha^2\|R_{>k}\|_2 \geq \frac{\varepsilon_2 - \varepsilon_1}{3c}\varepsilon_2 - c\left(\frac{\varepsilon_2 - \varepsilon_1}{3c}\right)^2 \geq (\varepsilon_2 - \varepsilon_1)\frac{\varepsilon_1 + 2\varepsilon_2}{9c},$$

where in first inequality we have used triangle inequality on $i\alpha H = ct^2R(\alpha) - U(\alpha)$ to conclude $\alpha\|H_{>k}\|_2 \leq \|U_{>k}\|_2 + c\alpha^2\|R_{>k}\|_2$, and in the second the fact that $H$ is $\varepsilon_2$-far from $k$-local. □

### Structural lemma for sparse Hamiltonians

Similar to local Hamiltonians, we show a discrepancy in the sum of the top Pauli coefficients of the short-time evolution operator for close-to-sparse and far-from-sparse Hamiltonians. To formally state this result we need to introduce the concept of *top energy*. Let $U(t)$ the time evolution operator at time $t$ and let $\{\widehat{U}(t)\}_x$ be its Pauli coefficients. We assign labels from $\{x_0, \dots, x_{4^n-1}\}$ to $x \in \{0,1,2,3\}^n$ in a way that $\widehat{U}_{x_0} = \widehat{U}_{0^n}$ and $|\widehat{U}_{x_1}| \geq |\widehat{U}_{x_2}| \geq \dots \geq |\widehat{U}_{x_{4^n-1}}|$. Now, we define the top energy at time $t$ as

$$\mathsf{TopEnergy}(t; s) := |\widehat{U}_{x_0}(t)|^2 + \sum_{i \in [s]} |\widehat{U}_{x_i}(t)|^2,$$

**Lemma 7.11.** *Let $H$ be a $n$-qubit Hamiltonian with $\|H\|_{\text{op}} \leq 1$ and $\text{Tr}[H] = 0$. Let $t \in (0,1)$. On the one hand, if $H$ is $\varepsilon_1$-close to $s$-sparse, then*

$$\mathsf{TopEnergy}(t; s) \geq 1 - \varepsilon_1^2 t^2 - O(t^3 s).$$

*On the other hand, if $H$ is $\varepsilon_2$-far from $s$-sparse, then*

$$\mathsf{TopEnergy}(t; s) \leq 1 - \varepsilon_2^2 t^2 + O(t^3 s).$$

*Proof.* For this proof we need to consider the 2nd order Taylor expansion of $U(t)$,

$$U(t) = \text{Id} - itH - t^2H^2/2 + O(t^3)R_2,$$

where $R_2$ is the remainder of the series of order 2 that satisfies $\|R_2\|_{\text{op}} \leq 1$, because $\|H\|_{\text{op}} \leq 1$. Since $\text{Tr}[H] = 0$ (so $\lambda_{0^n} = 0$), we have

$$\widehat{U}_0(t) = 1 - \frac{t^2}{2} \cdot \sum_{x \in \{0,1,2,3\}^n} \lambda_x^2 + O(t^3),$$

so, using that $|a^2 - b^2| = |a-b||a+b|$, we have that

$$\Big||\widehat{U}_0(t)|^2 - \big(1 - t^2 \sum_{x\in\{0,1,2,3\}^n} \lambda_x^2\big)\Big| = O(t^3). \tag{7.2}$$

To control $|U_x(t)|$ for $x \neq 0^{2n}$, we use the first order Taylor expansion of $U(t) = \mathrm{Id}_{2^n} - itH + ct^2 R_1(t)$ and get

$$\big||\widehat{U}_x(t)| - |t\lambda_x|\big| \leq |\widehat{U}_x(t) - (-it\lambda_x)| \leq \|U(t) - (-itH)\|_2 \leq O(t^2)\|R_1\|_2 \leq O(t^2), \tag{7.3}$$

where we again used that $\|R_1\|_2 \leq 1$. From this it follows that

$$\begin{aligned} \big||\widehat{U}_x(t)|^2 - t^2\lambda_x^2\big| = \Big|\big(|\widehat{U}_x(t)| - |t\lambda_x|\big)\cdot\big(|\widehat{U}_x(t)| + |t\lambda_x|\big)\Big| &= O(t^2)(|U_x| + |t\lambda_x|) \\ &= O(t^2)(2|t\lambda_x| + O(t^2)) = O(t^3), \end{aligned} \tag{7.4}$$

where the second and third equality both used Eq. (7.3); and in the last line used $|\lambda_x| \leq \|H\|_{\mathrm{op}} \leq 1$. In particular, the above implies that

$$|\widehat{U}_x(t)|^2 \geq t^2|\lambda_x|^2 - O(t^3) \tag{7.5}$$

Now we will define a quantity similar to the top energy, but now we will define the top coefficients as the top coefficients of $H$. To be precise, we assign labels to $\{y_0, \ldots, y_{4^n-1}\}$ to the elements of $\{0,1,2,3\}^n$ in a way such that $y_0 = 0^{2n}$ and $|\lambda_{y_1}| \geq \cdots \geq |\lambda_{y_{4^n-1}}|$. We now define

$$\mathsf{TopEnergy}_H(t;s) := \Big(1 - t^2 \sum_{x\in\{0,1,2,3\}^n} \lambda_x^2\Big) + \sum_{i\in[s]} (t\lambda_{y_i})^2.$$

If the top $s$ Pauli coefficients of $H$ coincided with the ones of $U(t)$ and there was no error in the Taylor expansion, then $\mathsf{TopEnergy}_H(t;s)(t) = \mathsf{TopEnergy}(t;s)$. However, this may not be true in general. Nevertheless, we show that both quantities are close to each other. To this end,

$$\begin{aligned}\mathsf{TopEnergy}(t;s) &= |\widehat{U}_{x_0}(t)|^2 + \sum_{i\in[s]} |\widehat{U}_{x_i}(t)|^2 \\ &\geq |\widehat{U}_{y_0}(t)|^2 + \sum_{i\in[s]} |\widehat{U}_{y_i}(t)|^2 \\ &\geq \big(1 - t^2 \sum_{x\in\{0,1,2,3\}^n} \lambda_x^2\big) + \sum_{i\in[s]} (t\lambda_{y_i})^2 - (s+1)O(t^3) \\ &= \mathsf{TopEnergy}_H(t;s) - (s+1)O(t^3),\end{aligned}$$

where in the first inequality we used that $x_1,\ldots,x_s$ correspond to the $s$ largest coefficients of $U(t)$, so $\sum_{i\in[s]} |\widehat{U}_{x_i}(t)|^2$ is larger than the sum of the squares of any other $s$ coefficients of $U$; in the second inequality we used Eqs. (7.2) and (7.5). Similarly, one can check that $\mathsf{TopEnergy}_H(t;s) \geq \mathsf{TopEnergy}(t;s) - (s+1)O(t^3)$, so

$$|\mathsf{TopEnergy}_H(t;s) - \mathsf{TopEnergy}(t;s)| \leq O(st^3).$$

Now, the claimed result follows by noticing that

$$\mathsf{TopEnergy}_H(t;s) = 1 - t^2 \sum_{i>s} |\lambda_{y_i}|^2,$$

and that $\sum_{i>s} |\lambda_{y_i}|^2$ is the square of the $\ell_2$-distance of $H$ to the space of $s$-sparse Hamiltonians, because $\sum_{i\in[s]} \lambda_{y_i}\sigma_{y_i}$ is the $s$-sparse Hamiltonian closest to $H$. □

## 7.4 Testing Hamiltonians

In this section, we give our testing algorithms for local Hamiltonians.

### 7.4.1 Testing local Hamiltonians

We now state our locality testing algorithm and prove its guarantees.

**Theorem 7.12.** *Algorithm 2 solves the locality testing problem (Problem 7.1 with the property of being k-local) with probability* $\geq 1-\delta$*, by making* $O(1/(\varepsilon_2-\varepsilon_1)^4 \cdot \log(1/\delta))$ *queries to the evolution operator and with* $O(1/(\varepsilon_2-\varepsilon_1)^3 \cdot \log(1/\delta))$ *total evolution time.*

*Proof.* Let $t = (\varepsilon_2 - \varepsilon_1)/(3c)$ and let $U = U(t)$. For notational simplicity, let $\alpha_k :=$

**Algorithm 2** Locality tester

**Input:** Query access to the time evolution of $U(t) = e^{-itH}$, closeness and farness parameters $\varepsilon_1,\ \varepsilon_2 \in (0,1)$, locality parameter $k \in \mathbb{N}$ and failure parameter $\delta \in (0,1)$

Set $T = O(\log(1/\delta)/(\varepsilon_2 - \varepsilon_1)^4)$
Let $t = (\varepsilon_2 - \varepsilon_1)/(3c)$ and $U = U(t)$
Initialize $\alpha'_k = 0$
**for** $i = 1, ..., T$ **do**
    Perform Pauli sampling from $U$. Let $x \in \{0,1,2,3\}^n$ be the outcome.
    **if** $|x| > k$ **then**
        $\alpha'_k \leftarrow \alpha'_k + 1/T$
    **end if**
**end for**Set $\alpha''_k = 0$
**for** $i = 1, ..., T$ **do**
    Perform Pauli sampling from $U$. Let $x \in \{0,1,2,3\}^n$ be the outcome.
    If $|x| > k$, $\alpha''_k \leftarrow \alpha''_k + 1/T$
**end for**

**Output**: If $\alpha'_k \geq (3/4)(\varepsilon_2 - \varepsilon_1)^2$ or $\alpha''_k \geq (\varepsilon_2 - \varepsilon_1)(\varepsilon_1 + 2\varepsilon_2)/(9c) - (\varepsilon_2 - \varepsilon_1)^2/(18c)$ output that $H$ is far from local, and close to local otherwise

$\|U_{>k}\|_2^2$. We will first estimate $\alpha_k$ upto error $(\varepsilon_2 - \varepsilon_1)^2/4$. To do that we sample from $\{|\widehat{U}_x|^2\}_x$ using Fact 7.8 a total of $T = O(1/(\varepsilon_2 - \varepsilon_1)^4 \log(1/\delta))$ times, which can be done with $T$ queries. If $x_1, \ldots, x_T$ are the outcomes of those samples, we define our estimate as

$$\alpha'_k := \frac{1}{T} \sum_{i \in [T]} [|x_i| > k].$$

By the Hoeffding bound, we have that indeed $|\alpha'_k - \alpha_k| \leq (\varepsilon_2 - \varepsilon_1)^2/4$ with probability $\geq 1 - \delta/2$.

If $\alpha'_k \geq (3/4)(\varepsilon_2 - \varepsilon_1)^2$, then $\alpha_k \geq (\varepsilon_2 - \varepsilon_1)^2/2$, so by Lemma 7.10 we conclude that $H$ is far from $k$-local. Otherwise, if $\alpha'_k \leq (3/4)(\varepsilon_2 - \varepsilon_1)^2$, then $\alpha_k \leq (\varepsilon_2 - \varepsilon_1)^2$. Now we take again $T$ samples from $y_1, \ldots, y_T$ from $\{|\widehat{U}_x|^2\}_x$ and define a new estimate

$$\alpha''_k = \frac{1}{T} \sum_{i \in [T]} [|y_i| > k].$$

By definition $\alpha''_k$ equals $\alpha_k$ in expectation. Furthermore, $\alpha_k$ is the empirical average of random variables whose variance is considerably small, because

$$\mathbb{E}[[|y| > k]^2] = \mathbb{E}[[|y| > k]] = \|U_{>k}\|_2^2 \leq (\varepsilon_2 - \varepsilon_1)^2.$$

Then, an application of Bernstein's inequality (Lemma 2.22) shows that $\alpha''_k$ approximates $\|U_{>k}\|_2^2$ up to error $((\varepsilon_2-\varepsilon_1)^2/(18c))^2$ with success probability $1-\delta/2$. At this point, using our structure Lemma 7.10, this is sufficient for testing $k$-locality. □

*Remark* 7.13. We remark that the algorithm for testing locality can be used in more generality for testing if the support of the Hamiltonians is a given $\mathcal{S} \subseteq \{0,1,2,3\}^n$. Also, by a union bound one can test for $M$ supports $\mathcal{S}_1,\ldots,\mathcal{S}_M$ by paying a factor $\log(M)$.

**Theorem 7.14.** *Let $H$ be a $n$-qubit Hamiltonian and let $\mathcal{S}_1,\ldots,\mathcal{S}_M \subseteq \{0,1,2,3\}^n$. Then, with $O(1/(\varepsilon_2-\varepsilon_1)^4\log(M/\delta))$ queries and $O(1/(\varepsilon_2-\varepsilon_1)^3\log(M/\delta))$ total evolution time one can simultaneously, for every $i\in[M]$, test if $H$ is $\varepsilon_1$-close or or $\varepsilon_2$-far from being supported on $\mathcal{S}_i$.*

Theorem 7.12 is one case of Theorem 7.14 where $M=1$ and $\mathcal{S}_1=\{x\in\{0,1,2,3\}^n : |x|\le k\}$.

### 7.4.2 Testing sparse Hamiltonians

Now we state our sparsity testing algorithm and prove its guarantees.

---

**Algorithm 3** Fully tolerant sparsity tester

---

**Input:** Query access to the time evolution of $U(t)=e^{-itH}$, closeness and farness parameters $\varepsilon_1,\ \varepsilon_2\in(0,1)$, sparsity parameter $s\in\mathbb{N}$ and failure parameter $\delta\in(0,1)$

1: Set $T=O(s^6/(\varepsilon_2^2-\varepsilon_1^2)^6\cdot\log(1/\delta))$
2: Let $t=O((\varepsilon_2^2-\varepsilon_1^2)/s)$ and $U=U(t)$
3: Perform Pauli sampling from $U$ a total of $T$ times. Let $(|\alpha_x|^2)_{x\in\{0,1,2,3\}^n}$ the empirical estimate of $(|\widehat{U}_x|^2)_x$ obtained this way.
4: Let $|\alpha_{x_1}|^2,\ldots,|\alpha_{x_s}|^2$ the $s$-biggest elements of $(|\alpha_x|^2)_{x\in\{0,1,2,3\}^n-\{0^n\}}$
5: Set $\Gamma=|\alpha_{0^n}|^2+\sum_{i\in[s]}|\alpha_{x_i}|^2$.

**Output**: If $\Gamma\ge 1-\varepsilon_1^2\frac{(\varepsilon_2^2-\varepsilon_1^2)^2}{s^2}-\frac{1}{2}\frac{(\varepsilon_2^2-\varepsilon_1^2)^3}{s^2}$ output that $H$ is close to sparse, and far from sparse otherwise

---

**Theorem 7.15.** *Algorithm 3 solves the $s$-sparsity testing problem with probability $\ge 1-\delta$, by making $O(s^6/(\varepsilon_2^2-\varepsilon_1^2)^6\cdot\log(1/\delta))$ queries to the evolution operator and with $O(s^5/(\varepsilon_2^2-\varepsilon_1^2)^5\cdot\log(1/\delta))$ total evolution time.*

*Proof.* Let $t=O((\varepsilon_2^2-\varepsilon_1^2)/s)$. By Lemma 7.11 we have that if $H$ is $\varepsilon_1$-close to being sparse, then

$$\mathsf{TopEnergy}(t;s)\ge 1-\varepsilon_1^2\frac{(\varepsilon_2^2-\varepsilon_1^2)^2}{s^2}-\frac{1}{3}\frac{(\varepsilon_2^2-\varepsilon_1^2)^3}{s^2},$$

while if $H$ is $\varepsilon_2$-far from $s$-sparse, then

$$\mathsf{TopEnergy}(t;s) \leq 1 - \varepsilon_2^2 \frac{(\varepsilon_2^2 - \varepsilon_1^2)^2}{s^2} + \frac{1}{3}\frac{(\varepsilon_2^2 - \varepsilon_1^2)^3}{s^2}.$$

From here, it follows that to test it suffices to estimate $\mathsf{TopEnergy}(t;s)$ up to error

$$\begin{aligned}\varepsilon &= \frac{1}{2}\left(1 - \varepsilon_1^2 \frac{(\varepsilon_2^2 - \varepsilon_1^2)^2}{s^2} - \frac{1}{3}\frac{(\varepsilon_2^2 - \varepsilon_1^2)^3}{s^2} - \left\{1 - \varepsilon_2^2 \frac{(\varepsilon_2^2 - \varepsilon_1^2)^2}{s^2} + \frac{1}{3}\frac{(\varepsilon_2^2 - \varepsilon_1^2)^3}{s^2}\right\}\right) \\ &= \frac{(\varepsilon_2^2 - \varepsilon_1^2)^3}{6s^2}.\end{aligned}$$

To do that we will obtain an estimate $(\{|\alpha_x|^2\}_x$ of $\{|\widehat{U}_x|^2\}_x$ and use it to approximate $\mathsf{TopEnergy}(t;s)$. Using Fact 2.6, we obtain an empirical distribution $\{|\alpha_x|^2\}_x$ that is obtained after $T = O(s^2 \log(1/\delta)/\varepsilon^2)$ samples from $\{|\widehat{U}_x|^2\}_x$ (which can be performed with $T$ queries to $U(t)$ thanks to Fact 7.8) satisfies that

$$\left||\alpha_x|^2 - |\widehat{U}_x|^2\right| \leq \frac{\varepsilon}{2s+1} \tag{7.6}$$

for all $x \in \{0,1,2,3\}^n$ with probability $\geq 1-\delta$. We assign new labels $y_0, y_1, \ldots, y_{4^n-1}$ to $\{0,1,2,3\}^n$ in a way such that $|\alpha_{y_0}|^2 = |\alpha_{0^n}|^2$ and $|\alpha_{y_1}|^2 \geq \cdots \geq |\alpha_{y_{4^n-1}}|^2$. Now, we define our estimate for $\mathsf{TopEnergy}(t;s)$ as

$$\mathsf{TopEnergy}'(t;s) = |\alpha_{y_0}(t)|^2 + 2\sum_{i\in[s]} |\alpha_{y_i}(t)|^2.$$

It only remains to show that $\mathsf{TopEnergy}'(t;s)$ $\varepsilon$-approximates $\mathsf{TopEnergy}(t;s)$. We will see that in two steps. First,

$$\begin{aligned}\mathsf{TopEnergy}'(t;s) &= |\alpha_{y_0}(t)|^2 + 2\sum_{i\in[s]} |\alpha_{y_i}(t)|^2 \\ &\geq |\alpha_{x_0}(t)|^2 + 2\sum_{i\in[s]} |\alpha_{x_i}(t)|^2 \\ &\geq |u_{x_0}(t)|^2 + 2\sum_{i\in[s]} |u_{x_i}(t)|^2 - \varepsilon \\ &= \mathsf{TopEnergy}(t;s) - \varepsilon,\end{aligned}$$

where the second line is true by definition of $y_0, \ldots, y_{4^n-1}$ and the third line is true because Eq. (7.6). Switching the roles of $\mathsf{TopEnergy}'(t;s)$ and $\mathsf{TopEnergy}(t;s)$, one can prove that $\mathsf{TopEnergy}(t;s) \geq \mathsf{TopEnergy}'(t;s) - \varepsilon$.

**Complexity analysis.** We have queried $U(t)$ a total of $T = O(s^2 \log(1/\delta)/\varepsilon^2)$ times with $\varepsilon = (\varepsilon_2^2 - \varepsilon_1^2)^3/6s^2$ and $t = O((\varepsilon_2^2 - \varepsilon_1^2)/s)$, so the number of queries is

$$O\left(\frac{s^6}{(\varepsilon_2^2 - \varepsilon_1^2)^6} \log(1/\delta)\right)$$

and the total evolution time

$$O\left(\frac{s^5}{(\varepsilon_2^2 - \varepsilon_1^2)^5} \log(1/\delta)\right).$$

□

Furthermore, for the regime where $\varepsilon_1 = O(\varepsilon_2/s^{0.5})$ we propose a more efficient testing algorithm.

**Algorithm 4** Not that tolerant sparsity tester

**Input:** Query access to the time evolution of $U(t) = e^{-itH}$, sparsity parameter $s \in \mathbb{N}$, closeness and farness parameters $\varepsilon_1,\ \varepsilon_2 \in (0,1)$ satisfying $\varepsilon_1 = O(\varepsilon_2/\sqrt{s})$ and failure parameter $\delta \in (0,1)$

1: Set $T = O(s^2/\varepsilon_2^4 \cdot \log(1/\delta))$
2: Let $t = \Omega(\varepsilon_2/\sqrt{s})$ and $U = U(t)$
3: Perform Pauli sampling from $U$ a total of $T$ times. Let $\mathcal{X}$ the set of sampled Paulis.

**Output**: If $|\mathcal{X} - \{0^{2n}\}| \leq s$ output that $H$ is close to sparse, and far from sparse otherwise

**Theorem 7.16.** *Let $H$ be a traceless Hamiltonian with $\|H\|_{\mathrm{op}} \leq 1$. Provided that $\varepsilon_1 = O(\varepsilon_2/s^{0.5})$, Algorithm 4 solves the $s$-sparsity testing problem with probability $\geq 1-\delta$. The algorithm makes $O(s^2/\varepsilon_2^4 \cdot \log(1/\delta))$ queries to the evolution operator and uses $O(s^{1.5}/\varepsilon_2^3 \cdot \log(1/\delta))$ total evolution time.*

*Proof.* Let $C > 1$ be a constant that appears in the first-order Taylor expansion,

$$U(t) = \mathrm{Id} - itH + Ct^2 R_1(t)$$

with $\|R_1\|_{\mathrm{op}} \leq 1$ for $t \in (0,1)$. We will assume that $\delta = 1/3$, as the case $\delta \in (0,1/3)$ follows by a standard majority voting argument. Algorithm 4 is simple. One just performs Pauli sampling of $U = U(t)$ a number of $T$ times, for some $t$ and $T$ to be determined later. Let $\mathcal{X}$ be the labels of the Pauli strings sampled in this process. If

$|\mathcal{X} - \{0^{2n}\}| \leq s$ we output that $H$ is sparse, and otherwise we output that is far from sparse. It remains to analyze the correctness.

**Correctness.** In the case that $H$ is $\varepsilon_1$-close $s$-sparse, there exists $\mathcal{S} \subset \{0,1,2,3\}^n$ of size $s$ where $H$ is $\varepsilon_1$-concentrated. Then, by Taylor expansion,

$$\sqrt{\sum_{x \notin (\mathcal{S} \cup \{0^{2n}\})} |\widehat{U}_x|^2} \leq t \sqrt{\sum_{x \notin (\mathcal{S} \cup \{0^{2n}\})} |\lambda_x|^2} + Ct^2 \leq t\varepsilon_1 + Ct^2 \leq 2Ct^2,$$

where in the last inequality we have assumed that

$$\varepsilon_1 \leq Ct. \tag{7.7}$$

Hence, the probability of sampling an element outside $\mathcal{S} \cup \{0^{2n}\}$ in one sample is at most $4C^2t^4$. Thus, the probability of not sampling an element outside $\mathcal{S} \cup \{0^{2n}\}$ in $T$ samples is at least

$$(1 - 4C^2t^2)^T \geq 1 - 4C^2t^4T.$$

In particular, if

$$T \leq \frac{1}{3}\frac{1}{4C^2t^4} \tag{7.8}$$

it will be satisfied that $|\mathcal{X} - \{0^{2^n}| \leq s$ with probability $\geq 2/3$, as desired.

In the case that $H$ is $\varepsilon_2$-far from $s$-sparse, we will perform an analysis similar to the coupon collector problem. By Taylor expansion we have that for every set $\mathcal{S}$ of size $s$,

$$\sqrt{\sum_{x \notin (\mathcal{S} - \{0^{2n}\})} |\widehat{U}_x|^2} \geq \varepsilon_2 t - Ct^2 \geq \frac{\varepsilon_2 t}{2}, \tag{7.9}$$

where we have assumed that

$$Ct \leq \varepsilon_2/2. \tag{7.10}$$

Let $X_i$ the random variable that accounts for the number of samples between the $(i-1)$-th sampled non-$0^{2n}$-Pauli and the $i$-th sampled non-$0^{2n}$-Pauli. Applying Eq. (7.9) to every $\mathcal{X}_i$, it follows that $\mathbb{E}[X_i] \leq 4/\varepsilon_2^2 t^2$ for every $i \in [s+1]$, so

$$\mathbb{E}[X_1 + \cdots + X_{s+1}] \leq \frac{4(s+1)}{\varepsilon_2^2 t^2}.$$

Hence, by Markov's inequality, if

$$T \geq \frac{\sqrt{3}4(s+1)}{\varepsilon_2^2 t^2} \tag{7.11}$$

it will be satisfied that $|\mathcal{X} - \{0^{2n}\}| \geq s+1$ with probability $\geq 2/3$, as desired.

Finally, we note that we have assumed conditions Eqs. (7.7), (7.8), (7.10) and (7.11) to ensure the correctness of the algorithm. All these equations are satisfied provided that

$$t = \frac{\varepsilon_2}{\sqrt{50C^2(s+1)}} = \Omega\left(\frac{\varepsilon_2}{\sqrt{s}}\right),$$
$$T = \frac{1}{12C^2t^4} = O\left(\frac{s^2}{\varepsilon_2^4}\right),$$
$$\varepsilon_1 \leq \frac{\varepsilon_2}{\sqrt{50(s+1)}} = O\left(\frac{\varepsilon_2}{\sqrt{s}}\right).$$

□

## 7.5 Learning Hamiltonians

### 7.5.1 Learning unstructured Hamiltonians

We start by showing how to efficiently learn an arbitrary $n$-qubit Hamiltonian in $\ell_\infty$ error. To do that, we propose a protocol to estimate a given set of Pauli coefficients $\mathcal{X}$ of a Hamiltonian via Shadow tomography. To describe the protocol, we introduce the following $2n$-qubit observables. Given $x \in \{0,1,2,3\}^n$, we define

$$\mathcal{R}_x := \frac{1}{2}(|\mathrm{Bell}_{0^{2n}}\rangle\langle\mathrm{Bell}_x| + |\mathrm{Bell}_x\rangle\langle\mathrm{Bell}_{0^{2n}}|),$$
$$\mathcal{I}_x := \frac{1}{2}(-i|\mathrm{Bell}_{0^{2n}}\rangle\langle\mathrm{Bell}_x| + i|\mathrm{Bell}_x\rangle\langle\mathrm{Bell}_{0^{2n}}|).$$

**Lemma 7.17.** *Let $H$ be an $n$-qubit traceless Hamiltonian and $\mathcal{X} \subseteq \{0,1,2,3\}^n$. Then, Algorithm 5 allows one to estimate the Pauli coefficients corresponding to $\mathcal{X}$ with success probability $\geq 1-\delta$. It uses $O((\log|\mathcal{X}|/\delta)\|H\|^4/\varepsilon^4)$ queries and $O(\log(|\mathcal{X}|/\delta)\|H\|^2/\varepsilon^3)$ total evolution time. The minimum evolution time is $\varepsilon/\|H\|^2$, the number of ancillas is $n$, and the time complexity is $O(\mathrm{poly}(n)|\mathcal{X}|\|H\|^4/\varepsilon^4 \cdot \log(|\mathcal{X}|/\delta))$.*

*Proof.* **Correctness of the algorithm:** Let $t_0 = \Theta(\varepsilon/\|H\|^2)$ and $U = U(t_0)$. As

**Algorithm 5** Estimating a given set of Pauli coefficients of a Hamiltonian

**Input:** Query access to the time evolution of $U(t) = e^{-itH}$, target set of Pauli coefficients $\mathcal{X} \subseteq \{0,1,2,3\}^n - \{0^n\}$, error parameter $\varepsilon \in (0,1)$, and failure parameter $\delta \in (0,1)$

```
Set T = O(‖H‖^4/ε^4 · log(|X|/δ)) and t_0 = Θ(ε/‖H‖^2)
Set U = U(t_0)
for j ∈ [T] do
    Prepare |J(U)⟩ = (U ⊗ Id_{2^n})|Bell_n⟩
    Apply a uniformly random Clifford gate C
    Measure in the computational basis. Let |b_j⟩ be the outcome
    for x ∈ X do
        Let R_{x,j} = (2^n + 1)⟨b_j|C^{-1} R_x C|b_j⟩ and I_{x,j} = (2^n + 1)⟨b_j|C^{-1} I_x C|b_j⟩
    end for
end for
for x ∈ X do
    Set R̃_x := MedianOfMeans(R_{x,j})_j and Ĩ_x := MedianOfMeans(I_{x,j})_j
end for
```

**Output**: $((\widetilde{R}_x + i\widetilde{I}_x)/(-it))_{x\in\mathcal{X}}$

$\mathrm{Tr}[\mathcal{R}_x^2] = \mathrm{Tr}[\mathcal{I}_x^2] = 2$, by Theorem 7.9, the numbers $\widetilde{R}_x$ and $\widetilde{I}_x$ that Algorithm 5 outputs satisfy

$$|\,\mathrm{Tr}[\mathcal{R}_x|J(U)\rangle\langle J(U)|] - \widetilde{R}_x| \leq \frac{\varepsilon^2}{\|H\|^2}, \quad |\,\mathrm{Tr}[\mathcal{I}_x|J(U)\rangle\langle J(U)|] - \widetilde{I}_x| \leq \frac{\varepsilon^2}{\|H\|^2}, \tag{7.12}$$

for every $x \in \mathcal{X}$ with probability $\geq 1-\delta$. By Taylor expansion, as $\lambda_{0^{2n}} = 0$, we have that $|\widehat{U}_{0^{2n}} - 1| \leq O(t_0^2\|H\|^2)$. Thus,

$$\mathrm{Tr}[\mathcal{R}_x|J(U)\rangle\langle J(U)|] = \frac{1}{2}(\widehat{U}_x\widehat{U}^*_{0^{2n}} + \widehat{U}_{0^{2n}}\widehat{U}^*_x) = \mathrm{Re}(\widehat{U}_x\widehat{U}^*_0) = \mathrm{Re}(\widehat{U}_x) \pm O(t_0^2\|H\|^2), \tag{7.13}$$

and similarly $\mathrm{Tr}[\mathcal{I}_x|J(U)\rangle\langle J(U)|] = \mathrm{Im}(\widehat{U}_x) \pm O(t_0^2\|H\|^2)$. Hence, combining Eqs. (7.12) and (7.13) we have that

$$|\widehat{U}_x - (\widetilde{R}_x + i\widetilde{I}_x)| \leq \frac{\varepsilon^2}{\|H\|^2} + O(t_0^2\|H\|^2) \leq O\left(\frac{\varepsilon^2}{\|H\|^2}\right),$$

for every $x \in \mathcal{X}$. Finally, by Taylor expansion we have that $|\widehat{U}_x/(-it_0) - \lambda_x| \leq O(t_0\|H\|^2)$, so

$$\left|\lambda_x - \frac{\widetilde{R}_x + i\widetilde{I}_x}{-it_0}\right| \leq O\left(\frac{\varepsilon^2}{t_0\|H\|^2}\right) + O(t_0\|H\|^2) = O(\varepsilon),$$

for every $x \in \mathcal{X}$, as claimed.

**Time complexity:** The time complexity is dominated by the first loop in Algorithm 5, whose time complexity is $O(|\mathcal{X}| \cdot T \cdot (t_{est} + \mathrm{poly}(n))$, where the poly($n$) comes from applying a random Clifford gate and $t_{est}$ is the time taken to compute $\langle b|C^{-1}\mathcal{R}_x C|b\rangle$ for an $n$-qubit Clifford gate $C$ and a computational basis state $|b\rangle$. Now, expanding $R_x$ one can write $\langle b|C^{-1}\mathcal{R}_x C|b\rangle$ as an algebraic expression of a finite number of terms of the kind $\langle y|D|z\rangle$, where $|y\rangle$ and $|z\rangle$ are computational basis states and $D$ a Clifford gate. Hence, via Gottesman-Knill theorem [Got98, AG04] follows that $t_{est} = O(n^2)$, so the total time complexity is $O(\mathrm{poly}(n)|\mathcal{X}|\|H\|^4/\varepsilon^4 \cdot \log(|\mathcal{X}|/\delta))$. □

Now, we are ready to present our learning algorithm for arbitrary Hamiltonians with no promise about its structure.

**Algorithm 6** Learning unstructured Hamiltonians

**Input:** Query access to the time evolution of $U(t) = e^{-itH}$, error parameter $\varepsilon \in (0,1)$, and failure parameter $\delta \in (0,1)$

1: Set $T = O(\|H\|^4/\varepsilon^4 \cdot \log(\|H\|^2/\varepsilon^2\delta))$ and $t_0 = \Theta(\varepsilon/\|H\|^2)$
2: Set $U = U(t_0)$
3: Set $\mathcal{X} = \emptyset$
4: **for** $j \in [T]$ **do**
5: Prepare $|J(U)\rangle = (U \otimes \mathrm{Id}_{2^n})|\mathrm{Bell}_n\rangle$
6: Measure in the Bell basis and add the outcome $x \in \{0,1,2,3\}^n$ to $\mathcal{X}$ if $x \neq 0^{2^n}$
7: **end for**
8: Run Algorithm 5 run with $U(t)$, $\mathcal{X}$, $\varepsilon$ and $\delta$ as inputs. Let $(\widetilde{\lambda}_x)_{x\in\mathcal{X}}$ the output.

**Output**: $\widetilde{H} = \sum_{x\in\mathcal{X}} \widetilde{\lambda}_x \sigma_x$

**Theorem 7.18** (Learning unstructured Hamiltonians)**.** *Let $H$ be an $n$-qubit and traceless Hamiltonian. Then, Algorithm 6 $\varepsilon$-learns all Pauli coefficients of $H$ with success probability $\geq 1-\delta$. It uses $\widetilde{O}((\|H\|/\varepsilon)^4)$ queries to the evolution operator and $\widetilde{O}(\|H\|^2/\varepsilon^3)$ total evolution time. The minimum evolution time is $\Theta(\varepsilon/\|H\|^2)$, the algorithm uses $n$ ancilla qubits and only one round of adaptivity, and the time complexity is* $\mathrm{poly}(n, 1/\varepsilon, \|H\|)$.

*Proof.* Let $t_0 = \Theta(\varepsilon/\|H\|^2)$ and $U = U(t_0)$ and let $T = O(\|H\|^4/\varepsilon^4 \cdot \log(\|H\|^2/\varepsilon^2\delta))$, as in Algorithm 6.

**Correctness**: We claim that with probability $\geq 1-\delta$ the set $\mathcal{X}$ generated in Algorithm 6 contains all $x$ such that

$$|\lambda_x| \geq \varepsilon, \tag{7.14}$$

and that

$$|\mathcal{X}| \leq \widetilde{O}\left(\frac{\|H\|^4}{\varepsilon^4}\right). \tag{7.15}$$

To show Eq. (7.14) we note that by Taylor expansion, if $|\lambda_x| \geq \varepsilon$, then $|\widehat{U}_x| = \Omega((\varepsilon^2/\|H\|^2))$, so $|\widehat{U}_x|^2 = \Omega((\varepsilon^4/\|H\|^4))$. Hence, the probability that such an $x$ does not belong to $\mathcal{X}$, which stores the non-$0^{2n}$ outcomes of sampling from $(|\widehat{U}_x|^2)_x$, is at most

$$\left(1 - |\widehat{U}_x|^2\right)^T \leq e^{-T|\widehat{U}_x|^2} \leq \frac{\varepsilon^2\delta}{\|H\|^2}.$$

Hence, as there is at most $\|H\|^2/\varepsilon^2$ coefficients with $|\lambda_x| \geq \varepsilon$, because $\sum_x |\lambda_x|^2 \leq \|H\|^2$, Eq. (7.14) follows from a union bound. Eq. (7.15) holds because $|\mathcal{X}| \leq T$.

Now, if Eqs. (7.14) and (7.15) are satisfied, Algorithm 5 provides estimates of the coefficients of $\mathcal{X}$, which contains all labels $x$ of coefficients $|\lambda_x| \geq \varepsilon$.

**Complexities:** The query complexity is $2T = \tilde{O}(\|H\|^4/\varepsilon^4)$, the minimum evolution time $t_0 = \Theta(\varepsilon/\|H\|^2)$ and the total time evolution $2Tt_0 = \widetilde{O}(\|H\|^2/\varepsilon^3)$. Additionally, the time complexity of Algorithm 6 is dominated by the call to Algorithm 5, which runs in time $O(\text{poly}(n)|\mathcal{X}|\|H\|^2/\varepsilon^2)$, which thanks to Eq. (7.15) is $\text{poly}(n, 1/\varepsilon, \|H\|)$. □

### 7.5.2 Learning local Hamiltonians

We now introduce our local Hamiltonian learner and prove its guarantees.

**Algorithm 7** Local Hamiltonian learner

**Input:** Query access to the time evolution of $U(t) = e^{-itH}$, error parameter $\varepsilon \in (0,1)$, locality parameter $k \in \mathbb{N}$ and failure parameter $\delta \in (0,1)$

```
Set T = exp(O(k^2 + k log(1/ε)) log(1/δ)
Let t = ε^{k+1} exp(−k(k+1)/2) and U = U(t)
Set γ = (ε/‖H‖^2)^{k+1} exp(−k(k+1)/2) and β = γε/‖H‖
Learn β-estimates λ'_x of λ_x via Algorithm 6
for |x| ≤ k do
    if |λ'_x| ≤ γ then
        λ̃_x = 0
    else
        λ̃_x = λ'_x
    end if
end for
```

**Output**: $\sum_{x \leq k} \widetilde{\lambda}_x \sigma_x$

**Theorem 7.19.** *Given a $n$-qubit $k$-local Hamiltonian $H$, Algorithm 7 outputs $\widetilde{H}$ such that with probability $\geq 1-\delta$ satisfies $\|H-\widetilde{H}\|_{\ell_2} \leq \varepsilon$. The algorithm makes* $\exp(O(k^2 + k\log(\|H\|^2/\varepsilon))\log(1/\delta)$ *queries to the evolution operator with* $\exp(O(k^2 + k\log(\|H\|^2/\varepsilon))\log(1/\delta)$ *total evolution time.*

To prove this theorem, we use the non-commutative Bohnenblust-Hille inequality by Volberg and Zhang [VZ23].

**Theorem 7.20** (Non-Commutative Bohnenblust-Hille inequality)**.** *Let $H = \sum_x \lambda_x \sigma_x$ be a $k$-local Hamiltonian. Then, there is a universal constant $C$ such that*

$$\widetilde{H} = \sum_{x\in\{0,1,2,3\}^n} |\lambda_x|^{\frac{2k}{k+1}} \leq C^k \|H\|.$$

*Proof of Theorem 7.19.* We only analyze the correctness of Algorithm 7, as the complexity quickly follows from Theorem 7.18. In this proof we also use the notation of Algorithm 7. The $\ell_2$-error of approximating $H$ with $\widetilde{H}$ is

$$\|\widetilde{H} - H\|_{\ell_2}^2 = \sum_{|\lambda'_x|\leq\gamma} |\lambda_x|^2 + \sum_{|\lambda'_x|\geq\gamma, |x|\leq k} |\lambda_x - \lambda'_x|^2. \tag{7.16}$$

We show separately that the two terms are at most $O(\varepsilon^2)$. To bound the contribution of the small Pauli coefficients, we first note that by Theorem 7.18 we have that

$$|\lambda'_x| \leq \gamma \quad \Longrightarrow \quad |\lambda_x| \leq \gamma + \beta = O(\gamma). \tag{7.17}$$

Hence,

$$\sum_{|\lambda'_x|\leq\gamma} |\lambda_x|^2 \leq \sum_{|\lambda_x|\leq O(\gamma)} |\lambda_x|^2 \leq O(\gamma^{\frac{2}{k+1}}) \sum_{x\in\{0,1,2,3\}^n} |\lambda_x|^{\frac{2k}{k+1}} \leq \gamma^{\frac{2}{k+1}} (C^k\|H\|^2)^{\frac{2k}{k+1}} = O(\varepsilon), \tag{7.18}$$

where in the first inequality we have used Eq. (7.17), in the third inequality we have used Theorem 7.20 and in the last inequality that $\gamma = (\varepsilon/\|H\|^2)^{k+1}\exp(-k(k+1)/2)$. To bound the contribution of the coefficients $|\lambda_x| \geq \gamma$ we notice that there is at most $\|H\|^2/\gamma^2$ of them, because $\sum_x |\lambda_x|^2 \leq \|H\|^2$. Thus,

$$\sum_{|\lambda'_x|\geq\gamma, |x|\leq k} |\lambda_x - \lambda'_x|^2 \leq \frac{\|H\|^2}{\gamma^2} \sup_x |\lambda_x - \lambda'_x|^2 \leq \frac{\|H\|^2\beta^2}{\gamma^2} = \varepsilon^2,$$

where in the second inequality we use the $\lambda'_x$ are $\beta$-estimates of $\lambda_x$ and in the last equality we use that $\beta = \gamma\varepsilon/\|H\|$. □

### 7.5.3 Learning sparse Hamiltonians

In this section we introduce our sparse Hamiltonian learner and prove its guarantees.

**Algorithm 8** Sparse Hamiltonian learner

**Input:** Query access to the time evolution of $U(t) = e^{-itH}$, error parameter $\varepsilon \in (0,1)$, sparsity parameter $s \in \mathbb{N}$ and failure parameter $\delta \in (0,1)$

```
1: Learn (ε/2)-estimates λ'_x of λ_x via Algorithm 6
2: for x ∈ {x : λ_x ≠ 0} do
3:     if λ'_x ≤ ε/2 then
4:         λ̃_x = 0
5:     elseλ_x > ε/2
6:         λ̃_x = λ'_x
7:     end if
8: end for
```

**Output**: $\widetilde{H} = \sum_x \widetilde{\lambda}_x \sigma_x$

**Theorem 7.21** (Sparse Hamiltonian learning)**.** *Given an $n$-qubit, $s$-sparse Hamiltonian $H$, Algorithm 8 outputs another Hamiltonian $\widetilde{H} = \sum \widetilde{\lambda}_x \sigma_x$ such that with probability $\geq 1-\delta$ satisfies $\|H - \widetilde{H}\|_{\ell_\infty} \leq \varepsilon$, The algorithms uses $\widetilde{O}(\|H\|^4/\varepsilon^4)$ queries and $\widetilde{O}(\|H\|^2/\varepsilon^3)$ total evolution time.*

*Furthermore, if $\lambda_x = 0$, then $\widetilde{\lambda}_x = 0$. This implies that running Algorithm 8 with $\varepsilon = \varepsilon'/\sqrt{s}$ outputs $\widetilde{H}$ such that $\|H - \widetilde{H}\|_{\ell_2} \leq \varepsilon'$. In this case, the algorithm uses $\widetilde{O}(\|H\|^4 s^2/\varepsilon'^4)$ queries and $\widetilde{O}(\|H\|^2 s^{1.5}/\varepsilon'^3)$ total evolution time.*

*Proof.* The first part, concerning learning in the $\ell_\infty$ error follows from Theorem 7.18. The fact that $\lambda_x = 0$, then $\widetilde{\lambda}_x = 0$ follows from Line 3 of Algorithm 8. Finally, we note that having $\lambda_x = 0 \implies \widetilde{\lambda}_x = 0$ and $|\lambda_x - \lambda_x| \leq \varepsilon'/\sqrt{s}$, implies $\|H - \widetilde{H}\|_{\ell_2} \leq \varepsilon'$. Indeed,

$$\|H - \widetilde{H}\|_{\ell_2} = \sum_{\lambda_x \neq 0} |\lambda_x - \widetilde{\lambda}_x|^2 \leq s \sup_x |\lambda_x - \widetilde{\lambda}_x|^2 = \varepsilon'^2,$$

where in the first step we have used that $\lambda_x = 0 \implies \widetilde{\lambda}_x = 0$, in the second that $|\lambda_x - \lambda_x| \leq \varepsilon'/\sqrt{s}$ and in the third that $H$ is $s$-sparse. □