Hamiltonian Learning via Shadow Tomography of Pseudo-Choi States

Juan Castaneda[1] and Nathan Wiebe[2,3,4]

[1]Department of Physics, University of Toronto, Toronto ON, Canada
[2]Department of Computer Science, University of Toronto, Toronto ON, Canada
[3] Pacific Northwest National Laboratory, Richland WA, USA
[4] Canadian Institute for Advanced Research, Toronto ON, Canada

We introduce a new approach to learn Hamiltonians through a resource that we call the pseudo-Choi state, which encodes the Hamiltonian in a state using a procedure that is analogous to the Choi-Jamiolkowski isomorphism. We provide an efficient method for generating these pseudo-Choi states by querying a time evolution unitary of the form e^{-iHt} and its inverse, and show that for a Hamiltonian with M terms the Hamiltonian coefficients can be estimated via classical shadow tomography within error ϵ in the 2-norm using $\widetilde{O}\left(\frac{M}{t^2\epsilon^2}\right)$ queries to the state preparation protocol, where $t \leq \frac{1}{2\|H\|}$. We further show an alternative approach that eschews classical shadow tomography in favor of quantum mean estimation that reduces this cost (at the price of many more qubits) to $\widetilde{O}\left(\frac{M}{t\epsilon}\right)$. Additionally, we show that in the case where one does not have access to the state preparation protocol, the Hamiltonian can be learned using $\widetilde{O}\left(\frac{\alpha^4 M}{\epsilon^2}\right)$ copies of the pseudo-Choi state. The constant α depends on the norm of the Hamiltonian, and the scaling in terms of α can be improved quadratically if using pseudo-Choi states of the normalized Hamiltonian. Finally, we show that our learning process is robust to errors in the resource states and to errors in the Hamiltonian class. Specifically, we show that if the true Hamiltonian contains more terms than we believe are present in the reconstruction, then our methods give an indication that there are Hamiltonian terms that have not been identified and will still accurately estimate the known terms in the Hamiltonian.

Contents

1 Introduction

Identifying the system Hamiltonian is a central problem in physics, as the Hamiltonian uniquely describes the dynamics of any closed quantum system. Additionally, as quantum computers evolve, the need to characterize errors in their behaviour is becoming increasingly important, and as a result, several classes of quantum certification protocols have been designed to ensure quantum devices behave as intended. There exist a wide variety of such protocols, each providing different amounts and types of information about the processes in question, and operating under different assumptions with varying resource costs [1]. In contrast to efficient protocols like randomized benchmarking, which only provide information about the average error rate of a quantum gate set [2–4], Hamiltonian learning can provide information about the systematic error in a unitary evolution under some fixed but erroneous Hamiltonian, without requiring exponential resources like protocols such as full process tomography which can provide more information about noise processes. Learning the Hamiltonian yields low-level information about the quantum dynamics that can be directly used for quantum control, and thus is an indispensable task for calibration and certification of quantum devices since it can be used to debug quantum Hamiltonian simulation algorithms by learning the simulated Hamiltonian and comparing the result to the ideal target.

Despite the importance of this problem, systematic methods for inferring the Hamiltonian and bounding the sample complexity needed to do so has only recently become a major focus in quantum information science. The most important thrusts include direct inference from short time experiments [5], approximate Bayesian inference [6–9], PAC (Probably Approximate Correct) methods based on learning from thermal states [10, 11], as well as PAC learning by querying the time evolution unitary $U = e^{-iHt}$ [11–17]. These latter techniques provide rigorous sample bounds on the evidence needed to learn the Hamiltonian within fixed error and probability of failure using thermal states or time evolution as the learning resource. However, in addition to potentially costly state preparation methods, approaches for Hamiltonian learning from thermal states are generally suitable for the high-temperature regime, not at very low temperatures. [10, 11].

Much of the recent progress made in the field has instead used the time evolution operator $U = e^{-iHt}$ as the learning resource. This approach naturally leads to applications for certifying quantum computers, where one would like to check whether a desired time evolution is actually performed in practice (this applies to scenarios where one would like to verify their own computations, as well as to scenarios where one party performs the computation, and a second party is interested in determining what went on under the hood). It also adds the possibility of using quantum control to improve the efficiency of learning techniques. Even so, the focus has typically been restricted to physically local Hamiltonians (and the slightly more general low-intersection class) rather than the more generic class of k-local Hamiltonians or beyond. In this work we address this by providing several new algorithms that allow us to efficiently learn Hamiltonians beyond the low-intersection regime, and provide explicit bounds on number of operations needed to produce these state both with and without coherent quantum control. Our method achieves this by introducing a new learning resource - in particular the "pseudo-Choi state" of the Hamiltonian - which we show can be used to efficiently learn the Hamiltonian. The main idea is that these states encode the Hamiltonian in a

way that makes it easy to recover the Hamiltonian coefficients by using few copies of the state to estimate the expectation values for a set of "decoding operators", which are simple operators whose expectation values can be computed on a classical computer in polynomial time. Furthermore, we provide a method of generating these pseudo-Choi states efficiently via controlled queries to the time evolution unitary $U = e^{-iHt}$ and its inverse $U^\dagger$. This access model is similar to that in [11–17], though it is slightly more restrictive because it requires access to the backwards time-evolution $U^\dagger$ as well.

While there have been a few recent approaches focused on learning Hamiltonians beyond the low-intersection class ([12, 15]), our approach improves the dependence of the query complexity on the error, locality of the Hamiltonian terms, and the number of Hamiltonian terms. In addition to this, some of our learning algorithms are robust in the sense that if there are unexpected Hamiltonian terms our algorithm will still correctly estimate the coefficients for terms we expected, and also signal that there were additional terms present. A more in-depth discussion about how our result compares to existing techniques for learning Hamiltonians is left for Section 6. For now, we summarize by saying that there is no clear "best" approach, but rather several good approaches that make sense for different scenarios (i.e. the type of input model considered and the types of Hamiltonians considered). That said, under the admittedly more restrictive input model considered here, the techniques we develop are able to achieve the best scaling in terms of both the number of Hamiltonian terms, as well as the error in estimating the parameters, with the added benefit of robustness. Furthermore, depending on the choice of error metric, our approach is applicable to *any* n-qubit Hamiltonian, including those with exponentially many terms (thus going far beyond even the k-local regime), so long as the norm of the Hamiltonian is bounded by a polynomial in the number of qubits. For these reasons, it remains of interest to study whether pseudo-Choi states can be produced in an efficient manner if one relaxes the input model to only require queries to U.

The paper is laid out as follows. Section 2 summarizes the key results of our work and gives specific statements of the learning problem we consider. In addition, we include a summary of the advantages and drawbacks of our approach when compared to existing Hamiltonian learning methods. Section 3 introduces the pseudo-Choi state of a Hamiltonian and discusses how it can be used as a resource to efficiently solve the Hamiltonian learning problem. A brief overview of the classical shadows tomography procedure introduced in [18] is also included. Section 4 examines the cost of learning the Hamiltonian by querying blackboxes that implement $U = e^{-iHt}$ and $U^\dagger$ using classical shadows in the learning procedure. An alternative method is introduced in Section 5, which uses a quantum algorithm introduced by *Huggins et al.* [19] for estimating expectation values. This approach quadratically improves how the query complexity scales with respect to the error and evolution time. Section 6 discusses the types of Hamiltonians our learning approach is well suited for and includes a comparison to prior work on Hamiltonian learning. Finally, we discuss the robustness of our procedure in Section 7, before concluding in Section 8.

2 Problem Statements and Summary of Results

The problem that we wish to tackle is a natural one. Let us assume that we have a set of possible Hamiltonian terms whose weighted sum comprises a Hamiltonian. Our task is to learn, within a fixed probability of failure, the weights for the Hamiltonian coefficeints within fixed error. This requirement falls within the scope of the quantum PAC learning paradigm, which was first introduced for theoretical purposes [20–22] but has recently become a practical tool used in characterization [18,

19]. The formal statement of this intuitive problem is given below.

Definition 1 (Hamiltonian Learning Problem)**.** *Let $\mathcal{H}_S$ be a Hilbert space of dimension $d = 2^n$, and let $\{H_m \mid m = 0, \dots M-1\}$ be a set of Hermitian and unitary operators acting on the space such that $\{H_m\}$ forms an orthonormal operator basis with respect to the inner product $\langle H_j, H_k \rangle = d^{-1}\mathrm{Tr}(H_j H_k)$ and let the Hamiltonian H be parameterized for $\mathbf{c} \in \mathbb{R}^M$ such that*

$$H = \sum_{m=0}^{M-1} c_m H_m$$

The Hamiltonian learning problem is to compute a vector $\hat{\mathbf{c}}$ such that with probability at least $1-\delta$ we have that $\|\hat{\mathbf{c}} - \boldsymbol{c}\|_2 \leq \epsilon$ where $\|\cdot\|_2$ is the vector L^2-norm.

In the most general case, there are up to d^2 unique Hamiltonian terms in the decomposition. These dense Hamiltonians generically cannot be learned efficiently within the operator basis considered as they require that we learn exponentially many terms in the Hamiltonians. However, focusing on special classes of Hamiltonians can reduce this number. Several recent papers ([11, 13, 14]) have focused on the **low-intersection** class of Hamiltonians, which was introduced in [11], and we describe in Definition 2. The low-intersection class contains all spatially local Hamiltonians, which are the focus of the previous work by Anshu et al. [10], and is therefore relevant for describing many physical systems.

Definition 2 (Low-Intersection Hamiltonian)**.** *A low-intersection Hamiltonian on n qubits is of the form $H = \sum_m c_m H_m$ and satisfies the following properties:*

1. *It is **k-local** : Each H_m is a finite dimensional Hermitian operator that acts non-trivially (i.e. as a non-identity operator) on at most $k \in \mathcal{O}(1)$ qubits.*
2. *For each Hamiltonian term H_m there exist at most a constant number of distinct terms H_l such that H_m and H_l act non-trivially on at least one common qubit.*

The second property directly implies that each qubit is acted on non-trivially by only a constant number of Hamiltonian terms. As a result, the number of Hamiltonian terms in low-intersection Hamiltonians scales as $\mathcal{O}(n)$.

In contrast, the class of k-local Hamiltonians that are not low-intersection is also of interest. This includes Hamiltonians that have pairwise interactions between every pair of qubits, as well as Hamiltonians that can be represented by star graphs (i.e. one qubit interacts with all others, but there are no/few interactions between the other qubits). Importantly, though this class is more general than the low-intersection class, k-local Hamiltonians contain at most $\mathcal{O}(n^k)$ Hamiltonian terms, and so the learning approach we introduce in this paper remains tractable. Physically interesting examples of k-local Hamiltonians include hardcore boson models, Heisenberg models on the complete graph, and spin glass models such as the Sherrington-Kirkpatrick and p-spin models. Furthermore, many quantum algorithms use Hamiltonian simulation as a subroutine, so for certification purposes it is useful to expand the scope of Hamiltonian learning beyond the low-intersection class and even to "less physical" Hamiltonians that may not be common in nature.

We consider Hamiltonian learning using two different, yet related, learning resources. The first is a state which we dub the "pseudo-Choi state" of the Hamiltonian, which roughly speaking is the representation of the Hamiltonian as a quantum state in a larger Hilbert space:

$$|\psi_c\rangle := \frac{(H \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA} |0\rangle_C + |\Phi_d\rangle_{SA} |1\rangle_C}{\alpha} \tag{1}$$

where H is dimensionless, α is a normalization constant and $|\Phi_d\rangle_{SA}$ is a maximally entangled state over two sub-systems denoted S and A. The name prescribed to this state is a consequence of the fact that the left-hand side of the state, $(H \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA}$, is similar to a Choi state, which is defined for a unitary $U_S \in \mathcal{H}_S$ as $|\psi_{Choi}\rangle = (U_S \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA}$. However, it is not generally a Choi state because H need not correspond to a CPTP map. The pseudo-Choi state is formally described in Definition 6. While these states do not result from natural processes like in the case of thermal states, we will see that they allow for very efficient learning of the Hamiltonian coefficients, which makes it of interest to find ways of efficiently generating them.

The second learning resource under consideration is a controllable time-evolution black-box $U = e^{-iHt}$. Along with thermal states, the unitary dynamics of a system is a natural resource to consider for learning the system Hamiltonian. In addition to the cost of thermal state preparation, learning from thermal states is often not feasible, such as at very low temperatures. To see this, consider two Hamiltonians $H_0 \succeq 0$ and $H_1 \succeq 0$

$$H' = |\psi\rangle\langle\psi| + (I - |\psi\rangle\langle\psi|)H_0(I - |\psi\rangle\langle\psi|) \tag{2}$$
$$H'' = |\psi\rangle\langle\psi| + (I - |\psi\rangle\langle\psi|)H_1(I - |\psi\rangle\langle\psi|) \tag{3}$$

In both cases, the zero-temperature groundstate is $|\psi\rangle\langle\psi|$ and thus the success probability of successfully distinguishing between both Hamiltonians is $1/2$, which shows that for sufficiently low temperature thermal states it can become impractical to learn the Hamiltonian in question. In such cases, it may be more convenient to use the unitary evolution of the system as the learning resource. In Section 4 we demonstrate that the pseudo-Choi state can be generated efficiently using the time evolution and its inverse.

As the time evolution unitary is merely one of conceivably many ways of generating the pseudo-Choi state, Section 3 considers Hamiltonian learning in the more general scenario where copies of the pseudo-Choi state are provided to the user with no promises about how they were generated. We refer readers to Section 6.3 for more discussion about pseudo-Choi states, and in particular why they may be more powerful than regular Choi states of the unitary evolution. On the other hand, while the time evolution unitary is merely a specific example of how such states can be generated, it is also a more intuitive resource - especially in the context of certifying quantum simulators. Furthermore, it has been used as the learning resource in previous work on Hamiltonian learning ([11–15]), which allows for a more direct comparison between the sample complexities of each approach. Therefore, we consider it as a separate case in Section 4.

2.1 Main Results

It should be noted that while our results below are stated for k-local Hamiltonians, our methods can be applied to more general Hamiltonians. In particular, as long as there are at most **poly**(n) Hamiltonian terms, and we have information about the structure of the Hamiltonian (i.e. which Hamiltonian terms are present), the query complexity remains polynomial in the number of qubits

for both the classical shadows and QME based approaches, as does the computational complexity. Furthermore, if we consider the infinity-norm in learning the Hamiltonian coefficients, our approach will be able to learn *any* n-qubit Hamiltonian with $\|H\| \in \mathbf{poly}(n)$ using only $\mathbf{poly}(n)$ queries to U and $U^\dagger$. Remarkably, this includes Hamiltonians with an exponential number of terms. For these cases with exponentially many Hamiltonian coefficients, while our methods are query-efficient, the computational complexity will be exponential. However, since all the Hamiltonian coefficients can in theory be computed at the same time on a classical computer, this could be somewhat alleviated with access to a large classical memory.

Our first main result is Theorem 3, which provides an upper bound on the number of pseudo-Choi states needed to learn a Hamiltonian.

Theorem 3 (Solving the pseudo-Choi Hamiltonian Learning Problem via Classical Shadows)**.** *The number of pseudo-Choi states required to solve the Hamiltonian learning problem of Definition 1 within error ϵ and failure probability at most δ for a k-local Hamiltonian is at most*

$$N \in \widetilde{\mathcal{O}}\left(\frac{\alpha^4 n^k}{\epsilon^2}\right)$$

where $\alpha \leq 1 + \|H\|$ is the normalization constant from equation (1).

Theorem 3 is presented in more detail as Theorem 16 and is proven in Section 3.2.1. Note that in Theorem 3 we have replaced M by the upper bound $M \in \mathcal{O}(n^k)$ for the number of terms in a k-local Hamiltonian, and the $\widetilde{\mathcal{O}}$ notation hides logarithmic factors.

Our second main result, Theorem 4, gives an upper bound on the number of queries to the time evolution unitary and its inverse needed to learn a Hamiltonian using classical shadows. This result is related to the previous one, as our approach involves first generating copies of the pseudo-Choi state by querying the unitary dynamics of the Hamiltonian, and then proceeding in similar fashion to the approach used to prove Theorem 3.

Theorem 4 (Solving the Unitary Hamiltonian Learning Problem via Classical Shadows)**.** *The number of queries to the time evolution unitary e^{-iHt} and its inverse, for $t \in \mathcal{O}(1/\|H\|)$, required to solve the Hamiltonian learning problem of Definition 1 within error ϵ and failure probability at most δ for a k-local Hamiltonian is at most*

$$N \in \widetilde{\mathcal{O}}\left(\frac{n^k}{t^2\epsilon^2}\right)$$

Theorem 4 is stated in more detail as Theorem 26 and is proven in Section 4.3.2. Again, we have replaced the number of Hamiltonian terms, M, by its upper bound for a k-local Hamiltonian. Importantly, the method we use for generating our resource states produces pseudo-Choi states of the normalized Hamiltonian, rather than the true Hamiltonian, which effectively results in a quadratic improvement of the query complexity in terms of the normalization constant. This is reflected in Theorem 4 by the quadratic scaling in terms of $1/t$, versus the quartic scaling in terms α from Theorem 3.

We also demonstrate that our learning algorithm is robust to two types of errors and in fact provides a witness to the fact that an error has occurred in the learning problem. The first type of error that we show robustness to is under specification of the Hamiltonian, wherein the Hamiltonian contains

further terms that are not specified in the Hamiltonian Learning Problem. In this case, we show that the learning problem will provide the coefficients for the Hamiltonian terms present in the system and the presence of uncharacterized terms can be inferred from the sum of the coefficients observed relative to the coefficient sum inferred from the pseudo-Choi state. The second form of robustness that we consider is to error in the Choi state, in which case we demonstrate that polynomially small noise can be tolerated in the Hamiltonian learning problem which indicates that the protocol is robust. These results are discussed in Section 7.

Our final main result, Theorem 5, gives an upper bound on the number of queries to the time evolution unitary and its inverse needed to learn a Hamiltonian using the mean-estimation results of Huggins et al. [19]. This process requires more quantum computations than the classical shadows approach (in which all post-measurement calculations are classical), but improves the query complexity quadratically with respect to the evolution time and the error in learning the coefficients.

Theorem 5 (Solving the Unitary Hamiltonian Learning Problem via Quantum Mean Estimation)**.** *The number of queries to the time evolution unitary e^{-iHt} and its inverse, for $t \in \mathcal{O}(1/\|H\|)$, required to solve the Hamiltonian learning problem of Definition 1 within error ϵ and failure probability at most δ for a k-local Hamiltonian is at most*

$$N \in \widetilde{\mathcal{O}}\left(\frac{n^k}{\epsilon t}\right)$$

Theorem 5 is restated for a Hamiltonian with M terms and proven as Theorem 33 in Section 5.2. Although our input model requires the backwards time evolution, this is the best query complexity we are aware of for learning the Hamiltonian coefficients to error ϵ in the 2-norm (see Section 6 for further discussion).

One caveat in that should be mentioned is that our approach requires short evolution times on the order of $1/\|H\|$ in order to prepare pseudo-Choi states from U and $U^\dagger$, preventing us from using long evolution times to reduce the query complexity. Furthermore, a good estimate of $\|H\|$ is required in order to achieve the query complexity in Theorems 4 and 5. Without this knowledge, the norm can be overestimated by the number of Hamiltonian terms (leading to shorter-than-necessary evolution times) to ensure the algorithm doesn't break down. This, for example, would lead to an additional factor of n^k in the query complexity in Theorem 5.

3 Learning the Hamiltonian from Pseudo-Choi States

This section is organized in the following way. We begin by explaining the first of the two learning resources (the pseudo-Choi state of the Hamiltonian, see Definition 6) that will be considered in our model of Hamiltonian learning. Next, Section 3.1 summarizes the parts of the classical shadows procedure (introduced in [18]) which are necessary to understand our approach. Finally, Section 3.2 contains the two main results of this section. In particular: Algorithm 1, which describes how one would learn the Hamiltonian from multiple copies of the pseudo-Choi state through the use of classical shadows tomography, and Theorem 16, which describes the sample complexity of the approach.

Note that the use of classical shadows is not strictly necessary in the learning procedure; any shadow tomography protocol - that is, any protocol that allows one to estimate expectation values

of operators for a given quantum state - can be used instead. The main benefits of classical shadows here are two-fold. First, for the operators we are interested in, the sample complexity has favourable scaling in terms of the number of Hamiltonian terms as a result of the shadow norm of these operators being upper bounded by a constant. Second, once all measurements are complete, the remainder of the algorithm is purely classical. While the main approach we focus on uses classical shadows, Section 5 gives an example of an alternative approach.

Our input state, the "pseudo-Choi state" of the Hamiltonian, consists of three subsystems. It contains the subsystem of interest, an ancillary subsystem of the same size, and an additional single-qubit subsystem that is used to ensure that the norm of the Hamiltonian is not lost in the pseudo-Choi state. The formal definition is given below. Note that while lowercase subscripts are used as indices throughout this paper, uppercase subscripts identify the system that corresponds to an operator or state. For example, the ket $|\Phi_d\rangle_{SA}$ resides in the Hilbert space $\mathcal{H}_S \otimes \mathcal{H}_A$.

Definition 6 (Pseudo-Choi State)**.** *Let $\mathcal{H}_S$ be the $d = 2^n$ dimensional Hilbert space upon which the map containing a dimensionless Hamiltonian H acts. Furthermore, let $\mathcal{H}_A$ be the Hilbert space of an ancilla of the same size, and let $\mathcal{H}_C$ be the 2 dimensional Hilbert space of another ancilla qubit.*

The pseudo-Choi state of the dimensionless Hamiltonian H resides on the Hilbert space $\mathcal{H}_S \otimes \mathcal{H}_A \otimes \mathcal{H}_C$ and is defined as

$$|\psi_c\rangle := \frac{(H \otimes \mathbb{I}_A)\,|\Phi_d\rangle_{SA}\,|0\rangle_C + |\Phi_d\rangle_{SA}\,|1\rangle_C}{\alpha} \tag{4}$$

where α is a normalization constant of the form

$$\alpha := \sqrt{\langle\Phi_d|_{SA}\,(H^2 \otimes \mathbb{I}_A)\,|\Phi_d\rangle_{SA} + 1} \tag{5}$$

and

$$|\Phi_d\rangle_{SA} = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle_S \otimes |i\rangle_A \tag{6}$$

is the maximally entangled state between $\mathcal{H}_S$ and $\mathcal{H}_A$.

Expanding α using the representation of the Hamiltonian from Definition 1, it is easily shown (see Appendix B.1) that

$$\alpha = \sqrt{\|\mathbf{c}\|_2^2 + 1} \tag{7}$$

Note that the right-hand side of the pseudo-Choi state is important, since it acts as a reference and allows us to learn the overall sign and norm of the Hamiltonian (which would otherwise show up as a global phase if the reference term was not there).

3.1 Shadow Tomography of the pseudo-Choi State

With the pseudo-Choi state in hand, we can now begin to learn the coefficients of the Hamiltonian through the use of shadow tomography. The idea of shadow tomography is to construct a model that can predict the expectation values of a particular set of observables with low error and high

probability over the set. Unlike conventional tomography, shadow tomography does not aim to provide a specific density operator, which means that the output expectation values need not be precisely consistent with a density operator, nor does the model need to provide a high fidelity estimate of the quantum state. Instead, we require only that it predicts the outcomes of a set of observables. Fortunately, this is all that we need to reconstruct the Hamiltonian and so this tool is precisely what we need to address our problem. A formal statement of the shadow tomography problem is given below.

Problem 7 (Shadow Tomography Problem)**.** *Given an unknown quantum state, ρ, of dimension $d' = 2^\eta$, and a set of L observables, $\{E_i\}$, estimate $Tr(\rho E_i)$ to within error ϵ_s for all $1 \leq i \leq L$, with success probability $1 - \delta_s$using as few copies of ρ as possible. [23]*

This problem can be solved classically in in settings where the operators in question have efficient classical representations, such as Pauli operators. This approach, referred to as classical shadows tomography [18], involves using a small number of copies of ρ to create a classical representation of ρ. Section 3.2 details how classical shadows can be used in conjunction with the pseudo-Choi state to learn the Hamiltonian coefficients, but to make it clear to the reader where some of the objects in later sections come from, we begin with a brief overview of classical shadows tomography.

The version of classical shadows we consider here involves applying Clifford operators from the group $\mathrm{Cl}(2^\eta)$ to an η qubit state ρ. The procedure consists of N rounds in which a Clifford operator U_i is sampled uniformly at random and applied to ρ, followed by a measurement of each qubit of the resulting state in the computational basis. After each measurement, a "snapshot" of the form $\sigma_i = U_i^\dagger |b_i\rangle \langle b_i| U_i$ is stored in classical memory. Here U_i is the randomly sampled Clifford unitary applied to ρ during the i^{th} round, and $|b_i\rangle$ is the bit string of length η that encodes the measurement outcomes of the i^{th} round. The j^{th} bit of $|b_i\rangle$ is the outcome of measuring the j^{th} qubit in the computational basis (i.e. if a bit in $|b_i\rangle$ is zero, then the outcome of measuring the corresponding qubit was $|0\rangle$). These snapshots are stabilizer states, meaning they can be obtained from $|0\rangle^{\otimes\eta}$ using only Clifford operations, and as a result $O(\eta^2)$ classical bits are needed to store each one in memory [21]. In total, the procedure uses N copies of ρ to generate N classical snapshots.

Note that although $|b_i\rangle$ is a classical bit string, we represent it using Dirac notation due to the simplicity of the notation, and to be consistent with the original notation used to describe classical shadows in [18]. The expectation value of these snapshots over all choices of Clifford unitaries and measurement outcomes can be viewed as a quantum channel. In particular, since the Clifford group forms a unitary 3-design, the process of randomly sampling Clifford operations corresponds to the following depolarizing channel [18]:

$$\begin{aligned} D_{1/(2^\eta+1)}(\rho) &= \frac{\rho}{2^\eta+1} + \left(1 - \frac{1}{(2^\eta+1)}\right)\frac{\mathrm{Tr}(\rho)\mathbb{I}}{2^\eta} \\ &= \frac{\rho + \mathrm{Tr}(\rho)\mathbb{I}}{2^\eta+1} \end{aligned} \tag{8}$$

This channel can be inverted in the following way:

$$\begin{aligned} D^{-1}_{1/(2^\eta+1)}(\rho) &= D_{2^\eta+1}(\rho) \\ &= (2^\eta+1)\rho - \mathrm{Tr}(\rho)\mathbb{I} \end{aligned} \tag{9}$$

where the first line uses the fact that $D_a(D_b(\rho)) = D_{ab}(\rho)$ (this is easily proven using the definition of a depolarizing channel), and the second line follows directly from the definition of a depolarizing channel.

The classical shadow of ρ is then defined as the collection of N snapshots after applying this inverse channel to each of them, as in Definition 8.

Definition 8 (Classical Shadow (From Clifford Measurements))**.** *Given N copies of an η-qubit quantum state ρ, the classical shadows procedure based on random Clifford measurements returns a classical shadow of ρ which is of the form*

$$\hat{\rho} = \{\hat{\rho}_i | i \in \mathbb{Z}_N\}, \tag{10}$$

where

$$\hat{\rho}_i = (2^\eta + 1)(U_i^\dagger |b_i\rangle \langle b_i| U_i) - \mathbb{I} \tag{11}$$

is the result of applying the inverse of the depolarizing channel (equation (9)) to the i^{th} classical snapshot.

As shown by Huang et al. [18], the classical shadow can be used to efficiently predict expectation values of operators on the state ρ, as required by the Shadow Tomography Problem.

Note that this procedure requires randomly sampling Clifford circuits, which is not as easy as sampling single-qubit operators such as Pauli operators. However, recent work has shown that this process can still be done efficiently. In particular, Bravyi and Maslov [24] and van den Berg [25] have proposed algorithms for sampling Clifford circuits uniformly at random with time complexity $\mathcal{O}(n^2)$ and circuit depth $\widetilde{\mathcal{O}}(n)$. The latter algorithm outputs a circuit directly and allows for parallelism, which effectively reduces the time complexity to $\mathcal{O}(n)$, but requires a quantum computer with fully connected topology to achieve the claimed circuit depth [25]. Meanwhile, the former approach achieves a two-qubit gate depth of $9n$ on the much simpler Linear Nearest Neighbour architecture [24].

Another variation of the classical shadows procedure replaces the random Clifford operators from the group $\mathrm{Cl}(2^\eta)$ with tensor products of single-qubit Clifford operators (i.e. operators from the group $\mathrm{Cl}(2)^{\otimes\eta}$). This process is referred to as the "random Pauli measurement" version of classical shadows, because the single-qubit Clifford group is generated by the Hadamard and phase gates, and these operations enact transformations between the Pauli X, Y, and Z bases. Therefore, allowing only single-qubit Clifford operations and measurement in the computational (Z) basis is equivalent to only allowing measurements in the Pauli X, Y, and Z bases. In the next section, we focus on the use of the Clifford measurement version of classical shadows, and leave discussion about the use of the Pauli version to Appendix A.

3.2 Efficient Estimation of the Hamiltonian Coefficients (Clifford Shadows)

In this section we discuss using the Clifford measurement version of classical shadows to extract the Hamiltonian coefficients from the pseudo-Choi state. We refer to the inference of these coefficients as the Pseudo-Choi Hamiltonian Learning problem (PC HLP). Note that the Pauli measurement flavour of classical shadows can also be used, and this approach is included in Appendix A. While the Pauli version only requires single-qubit operations and can reduce the computational complexity of the classical post-processing, the Clifford version has the better sample complexity for k-local systems. However, although the gap between the two sample complexity upper bounds grows exponentially with k, the benefits of the Pauli approach may outweigh the increased sample complexity for small values of k.

Since classical shadows tomography allows for efficient prediction of linear functions of operators, we can solve the PC HLP by choosing a specific set of operators whose expectation values correspond to the Hamiltonian coefficients. For Clifford-based shadows, these "decoding operators" are given in Definition 9.

Definition 9 (Decoding Operators)**.** *Let the set of decoding operators be defined as*

$$\boldsymbol{O} := \{O_l | l \in \mathbb{Z}_M\} \tag{12}$$

where

$$O_l := (H_l \otimes \mathbb{I}) \left|\Phi_d\right\rangle_{SA} \left\langle\Phi_d\right|_{SA} \otimes \left|0\right\rangle_C \left\langle 1\right|_C \tag{13}$$

and H_l is one of the M Hamiltonian terms.

Furthermore, we define

$$O_\alpha := \left|\Phi_d\right\rangle_{SA} \left\langle\Phi_d\right|_{SA} \otimes \left|1\right\rangle_C \left\langle 1\right|_C \tag{14}$$

Recalling Definition 6, the density matrix representing the pseudo-Choi state is given by

$$\begin{aligned}
\rho_c :=& \left|\psi_c\right\rangle \left\langle\psi_c\right| \\
=& \frac{1}{\alpha^2}\Big((H \otimes \mathbb{I}_A) \left|\Phi_d\right\rangle_{SA} \left\langle\Phi_d\right|_{SA} (H \otimes \mathbb{I}_A) \otimes \left|0\right\rangle_C \left\langle 0\right|_C \\
&+ (H \otimes \mathbb{I}_A) \left|\Phi_d\right\rangle_{SA} \left\langle\Phi_d\right|_{SA} \otimes \left|0\right\rangle_C \left\langle 1\right|_C \\
&+ \left|\Phi_d\right\rangle_{SA} \left\langle\Phi_d\right|_{SA} (H \otimes \mathbb{I}_A) \otimes \left|1\right\rangle_C \left\langle 0\right|_C \\
&+ \left|\Phi_d\right\rangle_{SA} \left\langle\Phi_d\right|_{SA} \otimes \left|1\right\rangle_C \left\langle 1\right|_C \Big)
\end{aligned} \tag{15}$$

We show below in Proposition 10 that the expectation values of the decoding operators acting on the pseudo-Choi state give the Hamiltonian coefficients up to a constant factor. Note that the Hamiltonian terms must be known in order to construct the decoding operators. As this work focuses on k-local Hamiltonians, this is not a problem, since there are at most $\mathcal{O}(n^k)$ possible terms.

Proposition 10 (Computing the Hamiltonian Coefficients)**.** *If ρ_c is the pseudo-Choi state (15), and O_l is a decoding operator as defined in Definition 9 then*

$$\mathrm{Tr}(\rho_c O_l) = \frac{c_l}{\alpha^2}, \tag{16}$$

where c_l is the l^{th} Hamiltonian coefficient, and α is the normalization constant of the pseudo-Choi state (7). Furthermore,

$$\mathrm{Tr}(\rho_c O_\alpha) = \frac{1}{\alpha^2} \tag{17}$$

Proof. From the definition of the decoding operators and the pseudo-Choi state, it is easy to see that

$$\mathrm{Tr}(\rho_c O_l) = \frac{\left\langle\Phi_d\right|_{SA} (H \otimes \mathbb{I}) (H_l \otimes \mathbb{I}) \left|\Phi_d\right\rangle_{SA}}{\alpha^2} \tag{18}$$

Expanding the above using the representation of the Hamiltonian from Definition 1 gives

$$\begin{aligned}\mathrm{Tr}(\rho_c O_l) &= \frac{\langle\Phi_d|_{SA}\left(\sum_m c_m H_m \otimes \mathbb{I}\right)(H_l \otimes \mathbb{I})\,|\Phi_d\rangle_{SA}}{\alpha^2} \\ &= \frac{1}{\alpha^2}\sum_m c_m \langle\Phi_d|_{SA}\,(H_m \otimes \mathbb{I})\,(H_l \otimes \mathbb{I})\,|\Phi_d\rangle_{SA}\end{aligned} \tag{19}$$

We can now use the definition of $|\Phi_d\rangle_{SA}$ to expand the term inside the summation:

$$\begin{aligned}\langle\Phi_d|_{SA}\,(H_m \otimes \mathbb{I})\,(H_l \otimes \mathbb{I})\,|\Phi_d\rangle_{SA} &= \left(\frac{1}{\sqrt{d}}\sum_i \langle i|\,\langle i|\right)(H_m \otimes \mathbb{I})(H_l \otimes \mathbb{I})\left(\frac{1}{\sqrt{d}}\sum_j |j\rangle\,|j\rangle\right) \\ &= \frac{1}{d}\sum_i\sum_j \langle i|\,H_m H_l\,|j\rangle\,\langle i|\,|j\rangle \\ &= \frac{1}{d}\sum_i \langle i|\,H_m H_l\,|i\rangle \\ &= \frac{\mathrm{Tr}(H_m H_l)}{d}\end{aligned} \tag{20}$$

Finally, we substitute (20) back in to (19) to get

$$\begin{aligned}\mathrm{Tr}(\rho_c O_l) &= \frac{1}{\alpha^2}\sum_m c_m \frac{\mathrm{Tr}(H_m H_l)}{d} \\ &= \frac{1}{\alpha^2}\sum_m c_m \delta_{ma} \\ &= \frac{c_a}{\alpha^2}\end{aligned} \tag{21}$$

The second claim ($\mathrm{Tr}(\rho_c O_\alpha) = \frac{1}{\alpha^2}$) follows immediately from the definition of the pseudo-Choi state (15) and the definition of O_α.

□

We seek to approximate the expectation values of Proposition 10 using the classical shadows approach introduced in [18]. However, the classical shadows procedure requires Hermitian operators, which O_l are not. Instead, we can use the following pair of Hermitian operators:

$$O_l^+ := O_l + (O_l)^\dagger \tag{22}$$

$$O_l^- := iO_l - i\,(O_l)^\dagger \tag{23}$$

These two operators are Hermitian, and can be inverted to find that

$$\frac{O_l^+ - iO_l^-}{2} = O_l \tag{24}$$

Therefore, the expectation value of O_l can be predicted by using classical shadows to estimate the expectation values of O_l^+ and O_l^-:

$$\langle O_l\rangle_s = \frac{\left\langle O_l^+\right\rangle_s - \left\langle iO_l^-\right\rangle_s}{2} \tag{25}$$

where the subscript s denotes that these are not the true expectation values, but estimates generated by the classical shadows procedure.

We now consider using the Clifford measurement flavor of classical shadows for the pseudo-Choi state. Since the pseudo-Choi state is on $2n+1$ qubits, its classical shadow is made up of N components of the form

$$\hat{\rho}_i = \left(2^{2n+1}+1\right) U_i^\dagger \left|b_i\right\rangle \left\langle b_i\right| U_i - \mathbb{I} \tag{26}$$

However, it is not necessary to store the shadow in this explicit form. Instead, it is helpful to use the form of the classical shadow specified in Definition 11, and introduce the $\mathbb{I}$ term and the factor of $2^{2n+1}+1$ as needed.

Definition 11. *If the random Clifford measurement version of the classical shadows procedure is performed, the resulting classical shadow of size N can be stored in the following way:*

$$\hat{\rho} = \{|\hat{\rho}_i\rangle \, | i \in \mathbb{Z}_N\}, \tag{27}$$

where

$$|\hat{\rho}_i\rangle = U_i^\dagger \left|b_i\right\rangle \tag{28}$$

The full procedure for estimating the vector of Hamiltonian coefficients, $\mathbf{c}$, is given in Algorithm 1. The algorithm also invokes a further subroutine (Algorithm 2) for determining the expectations using classical shadows. Algorithms 1 and 2 do not use the explicit form of the classical shadow given by Definition 8, but instead use the "compressed" version given by Definition 11.

Previous work by Aaronson and Gottesman provides an algorithm for simulating Clifford circuits classically in polynomial time, as well as an algorithm for calculating the inner product between two stabilizer states in $\mathcal{O}\left(n^3\right)$ time [21]. For future reference, we shall refer to this second algorithm (see the end of Section III in [21]) as the Stabilizer Inner Product (SIP) Algorithm. Algorithm 1 invokes the former of these algorithms, while Algorithm 2 leverages the SIP Algorithm to compute expectation values in polynomial time.

Note that in order for the output of Algorithm 1 to solve the PC HLP, the number of samples of ρ_c must be at large enough for the error and success probability to satisfy the requirements of the Hamiltonian Learning Problem 1. This sufficient value is given later as Theorem 16.

To go along with Algorithm 2, it is also helpful to understand how the expectation values involved are explicitly computed using a classical shadow. Proposition 12 gives expressions for the three expectation values of interest. These expressions contain several inner products which can be efficiently computed using the SIP algorithm of [21].

Proposition 12. *Let $\hat{\rho}_i = \left(2^{2n+1}+1\right) U_i^\dagger \left|b_i\right\rangle \left\langle b_i\right| U_i - \mathbb{I}$ be a classical shadow generated from the random Clifford measurement version of the classical shadows procedure, and consider the operators*

$$O_l^+ = O_l + (O_l)^\dagger$$
$$O_l^- = iO_l - i\,(O_l)^\dagger$$

where O_l is a decoding operator as in Definition 9.

Algorithm 1 `FindCoeffClifford`$(\rho_c^{\otimes N}, \boldsymbol{H}, n, N)$: Determine the Vector of Hamiltonian Coefficients from ρ_c using random Clifford measurement based classical shadows

Input:

$\rho_c^{\otimes N}$: Collection of N pseudo-Choi states on $2n+1$ qubits, each on Hilbert space $\mathcal{H}_S \otimes \mathcal{H}_A \otimes \mathcal{H}_C$

$\boldsymbol{H}$: Set of M k-local Hamiltonian terms H_l whose coefficients c_m are desired

▷ Note: $\rho_c^{\otimes N}$ is a quantum state, whereas $\boldsymbol{H}$ is a set of classical operators

n: Number of qubits in the system on Hilbert space $\mathcal{H}_S$

N: Number of pseudo-Choi states

1. **Initialize the array of Hamiltonian coefficients**

   Let $\hat{\mathbf{c}} \leftarrow \Big[0, 0, ..., 0\Big]_{1\times M}$ and let $\hat{\mathbf{c}}_l$ denote the l^{th} element of $\hat{\mathbf{c}}$

2. **Generate and store the classical shadow**

   Use N copies of ρ_c to generate $\hat{\rho}$, a classical shadow of size N of ρ_c, using the classical shadows procedure based on random Clifford measurements from [18].

   Store $\hat{\rho}$ in classical memory (as per Definition 11) using the tableau algorithm introduced in [21] to compute and store each stabilizer state $|\rho_i\rangle = U_i^\dagger |b_i\rangle$

3. **Generate an estimate of** $\mathrm{Tr}(\rho O_l) = \frac{c_l}{\alpha^2}$ **for each corresponding Hamiltonian term**

   `For` l `from` 0 `to` $M-1$:

   $\hat{\mathbf{c}}_l \leftarrow$ `ComputeExpectation`$(\hat{\rho}, H_l, N, l, n)$ ▷ Estimate of $\frac{c_l}{\alpha^2}$

4. **Generate an estimate of** $\mathrm{Tr}(\rho O_\alpha) = \frac{1}{\alpha^2}$

   $\hat{o}_\alpha \leftarrow$ `ComputeExpectation`$(\hat{\rho}, O_\alpha, N, -1, n)$ ▷ Estimate of $\frac{1}{\alpha^2}$

5. **Remove the factor of** $\frac{1}{\alpha^2}$ **from the Hamiltonian coefficients**

   Let $\hat{\mathbf{c}} \leftarrow \frac{\hat{\mathbf{c}}}{\hat{o}_\alpha}$

6. **Output the vector of Hamiltonian coefficients**

   Return $\hat{\mathbf{c}} = \Big[\hat{c}_0, \hat{c}_1, ..., \hat{c}_{M-1}\Big]$

We then have that

$$\begin{aligned}\mathrm{Tr}\left(\hat{\rho}_i O_l^+\right) =& \left(2^{2n+1}+1\right)\Big(\langle b_i| U_i \left(H_l \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA} |0\rangle_C\Big)\Big(\langle\Phi_d|_{SA} \langle 1|_C U_i^\dagger |b_i\rangle\Big) \\ &+ \left(2^{2n+1}+1\right)\Big(\langle b_i| U_i |\Phi_d\rangle_{SA} |1\rangle_C\Big)\Big(\langle\Phi_d|_{SA} \left(H_l \otimes \mathbb{I}_A\right) \langle 0|_C U_i^\dagger |b_i\rangle\Big),\end{aligned} \tag{29}$$

and similarly,

$$\begin{aligned}\mathrm{Tr}\left(\hat{\rho}_i O_l^-\right) =& i\left(2^{2n+1}+1\right)\Big(\langle b_i| U_i \left(H_l \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA} |0\rangle_C\Big)\Big(\langle\Phi_d|_{SA} \langle 1|_C U_i^\dagger |b_i\rangle\Big) \\ &- i\left(2^{2n+1}+1\right)\Big(\langle b_i| U_i |\Phi_d\rangle_{SA} |1\rangle_C\Big)\Big(\langle\Phi_d|_{SA} \left(H_l \otimes \mathbb{I}_A\right) \langle 0|_C U_i^\dagger |b_i\rangle\Big)\end{aligned} \tag{30}$$

Furthermore,

$$\mathrm{Tr}(\hat{\rho}_i O_\alpha) = \left(2^{2n+1}+1\right) |\langle b_i| U_i |\phi_d\rangle_{SA} |1\rangle_C|^2 - 1 \tag{31}$$

Note that $U_i^\dagger |b_i\rangle$ is on the Hilbert space $\mathcal{H}_S \otimes \mathcal{H}_A \otimes \mathcal{H}_C$, so all the inner products above give scalar values.

The proof of Proposition 12 can be found in Appendix B.2. As previously mentioned, $U_i |b_i\rangle$ is a stabilizer state. Furthermore, $H_l |\Phi_d\rangle_{SA}$ is also a stabilizer state for all $l \in M$, since the Hamiltonian terms are Pauli operators, and therefore also Clifford operators. From this, we see that calculating the expectation values in Proposition 12 comes down to computing a few inner products between stabilizer states, which we can do on a classical computer. Algorithm 2 computes these inner products using the SIP Algorithm of [21], and then performs a median of means (as prescribed by the classical shadows procedure [18]) to reduce the variance of the estimate.

Algorithm 2 `ComputeExpectation`$(\hat{\rho}, O, N, l, n)$: Compute Expectation Values with a Classical Shadow

Input:
$\hat{\rho}$: Classical shadow of size N of the resource state ρ, in the compressed form of Definition 11
H_l: Hermitian operator from the set $\boldsymbol{H}$
N: The size of the classical shadow $\hat{\rho}$
l: Integer indicating which of the M Hamiltonian terms O is. $l = -1$ indicates that the algorithm should use the formula for estimating the normalization constant α
n: Number of qubits in the system

1. **Iterate over the components ($|\hat{\rho}_i\rangle = U_i^\dagger |b_i\rangle$) of $\hat{\rho}$ and compute $\mathrm{Tr}(\hat{\rho}_i O)$ for each one**
   Let $\hat{\mathbf{o}} \leftarrow \left[0, 0, ..., 0\right]_{1\times N}$ and let $\hat{\mathbf{o}}_i$ denote the i^{th} element of $\hat{\mathbf{o}}$
   `if` $l == -1$:
      `for` i `from` 0 to $N-1$:
         Compute $\mathrm{Tr}(\hat{\rho}_i O_\alpha)$ using the SIP Algorithm given in [21] to compute the inner product in equation (31)
   `else` :
      `for` i `from` 0 to $N-1$:
         Compute $\mathrm{Tr}\left(\hat{\rho}_i O_l^+\right)$ using the SIP Algorithm given in [21] to compute the four inner products in equation (29)
         Compute $\mathrm{Tr}\left(\hat{\rho}_i O_l^-\right)$ using the SIP Algorithm given in [21] to compute the four inner products in equation (30)
         $\hat{o}_i \leftarrow \frac{\mathrm{Tr}\left(\hat{\rho}_i O_l^+\right) - i\mathrm{Tr}\left(\hat{\rho}_i O_l^-\right)}{2}$ ▷ As per Equation (25)
2. **Perform median of means protocol and output the estimate of $\mathrm{Tr}(\rho_c O_l) = \frac{c_l}{\alpha^2}$**
   Return `MedianOfMeans`$(\hat{\mathbf{o}})$

Algorithm 3 `MedianOfMeans`($\hat{\mathbf{o}}$):

Input:
$\hat{\mathbf{o}}$: Vector of size N

Perform median of means protocol:
Separate the elements of $\hat{\mathbf{o}}$, $\hat{o}_i$, into K groups, G_k, of equal size
For k from 0 to $K-1$:
    $\hat{o}_k \leftarrow \textbf{mean}\,(G_k)$
$\hat{o}_l \leftarrow \textbf{median}\,(\hat{o}_0, \hat{o}_1, ..., \hat{o}_{K-1})$
Return $\hat{o}_l$

The SIP Algorithm given in [21] requires $\mathcal{O}(n^3)$ time to compute inner products between stabilizer states, and Algorithm 2 calls it $\mathcal{O}(1)$ times for each of the N snapshots (see Theorem 16 for the sufficient value of N). In turn, Algorithm 1 calls Algorithm 2 for each of the $M+1$ operators, giving a total time complexity of $\mathcal{O}(MNn^3)$ for computing all the inner products between stabilizer states. However, all these calculations are independent and can be parallelized over both the operators and the classical snapshots. Thus, with a large classical memory, the computation time required can be greatly reduced through massive parallelism.

3.2.1 Sample Complexity Upper Bound

The main result in this section is Theorem 16, a proof of the sample complexity of solving the Hamiltonian learning problem using pseudo-Choi states as resources. Our result depends on the concept of a shadow norm which we define below for reference.

Definition 13 (Shadow Norm for Clifford-Based Shadows)**.** *Let O be a finite-dimensional operator. The shadow norm of the operator O is defined as*

$$\|O\|_{shadow} = \max_{\sigma} \left(\mathbb{E}_{U\in\mathcal{C}} \sum_{b\in\{0,1\}^n} \langle b| U\sigma U^\dagger |b\rangle \langle b| U\mathcal{M}^{-1}(O)U^\dagger |b\rangle^2 \right)^{1/2}$$

where $\mathcal{C}$ is the Clifford group and the maximization is over all quantum states.

Note that the definition of the shadow norm above assumes the classical shadows process is randomly sampling Clifford unitaries, and not unitaries from some other ensemble.

In order to state our proof of Theorem 16 we need to first state Lemma 14 which bounds the shadow norm, as well as Proposition 15 which gives a value of ϵ_s that ensures the total error of the learning process is at most ϵ, as required by the Hamiltonian Learning Problem (Definition 1).

Lemma 14 (Shadow Norm Upper Bounds)**.** *Let the set $\{O_\alpha\} \cup \{O_l^+, O_l^- \mid l \in \mathbb{Z}_M\}$ be denoted by $\boldsymbol{O} \equiv \{O_i \mid i \in \mathbb{Z}_{2M}\}$. If using the Clifford measurement version of the classical shadows procedure to predict the expectation values of the operators in the set, the shadow norm of each operator is at most*

$$\|O_i\|_{shadow}^2 \leq 3\mathrm{Tr}\left((O_i)^2\right) \leq 6$$

The proof of Lemma 14 can be found in Appendix B.3.

Proposition 15. *To solve the Hamiltonian learning problem with an error of at most ϵ, it serves to let the error in the classical shadows procedure be*

$$\epsilon_s = \frac{\epsilon}{\alpha^2 \sqrt{c_{\max}^2 + 1}\sqrt{M}},$$

where $c_{\max}$ denotes the Hamiltonian coefficient with the largest absolute value.

The proof of Proposition 15 can be found in Appendix B.4.

With these results in hand, we can prove the following result which provides the scaling of the number of resource states that are needed to solve our learning problem.

Theorem 16 (Sample Complexity Bound for PC HLP)**.** *The number of copies of the pseudo-Choi state (4) required by Algorithm 1 to solve the Hamiltonian Learning Problem given by Definition 1 is*

$$N \in \mathcal{O}\left(\frac{\alpha^4 (c_{\max} + 1) M \log\left(M/\delta\right)}{\epsilon^2}\right) \tag{32}$$

Proof of Theorem 16. The sample complexity of the shadow tomography procedure depends on the number of expectation values that are to be predicted. Recall that M is the number of terms in the Hamiltonian. Algorithm 1 requires estimating M expectation values $\langle O_l \rangle$ (and each $\langle O_l \rangle$ requires predicting the expectation values of 2 operators), as well as $\langle O_\alpha \rangle$. Therefore, the total number of operators whose expectation values are to be estimated via shadow tomography is $2M + 1$. Let the set of these operators $(\{O_\alpha\} \cup \{O_l^+, O_l^- \mid l \in \mathbb{Z}_M\}$) be denoted by $\mathbf{O} \equiv \{O_i \mid i \in \mathbb{Z}_{2M}\}$.

Using classical shadows, this gives a sample complexity of

$$N_s = \mathcal{O}\left(\frac{\log\left(M/\delta_s\right)}{\epsilon_s^2} \max_i \|O_i\|_{shadow}^2\right) \tag{33}$$

to estimate all the expectation values, each within sampling error ϵ_s and with a total probability of failure at most δ. [18]

Since we use the random Clifford measurement version of the classical shadows procedure, Lemma 14 implies that

$$\max_i \|O_i\|_{shadow}^2 \leq 6 \tag{34}$$

This result implies that the sample complexity for predicting the expectation values in Algorithm 1 within an error ϵ_s with probability δ_s is

$$N_s = \mathcal{O}\left(\frac{\log\left(M/\delta_s\right)}{\epsilon_s^2}\right). \tag{35}$$

We would like to determine an upper bound on the number of samples such that with probability $1 - \delta$, the l-2 error in the vector of Hamiltonian coefficients is at most ϵ, as per the Hamiltonian learning problem 1. The total probability of failure of Algorithm 1 is that of the classical shadows procedure, so $\delta = \delta_s$. Additionally, to minimize the number of samples needed while still ensuring that the total error in the Hamiltonian learning process is at most ϵ, it serves to choose the value of ϵ_s given as per Proposition 15. Combining this value of ϵ_s with Equation (35), we arrive at the claimed result. □

It is worth noting that if coherent quantum queries were made to the quantum state preparation routine then we can learn the coefficients with quadratically better scaling with ϵ than the above procedure using results by Huggins et al. [19]. However, since doing so necessitates the introduction of a more powerful oracle that allows coherent queries we ignore such optimizations here, and instead revisit this approach in Section 5.

4 Learning the Hamiltonian from a Unitary Time Evolution Blackbox

While the pseudo-Choi state might seem like a strange choice to use as the learning resource, the time evolution blackbox of the form

$$U = e^{-iHt} \tag{36}$$

is perhaps a more intuitive choice of learning resource, and has been recently studied in the context of Hamiltonian learning [11–14]. The goal of this section is to demonstrate that by querying this time evolution unitary and its inverse for $|t| \leq \frac{\pi}{2\|H\|}$, we can produce states similar to the pseudo-Choi states of the previous section, and as a result learn the Hamiltonian using a procedure very similar to the one in Algorithm 1. This leads to Theorem 26, which is the main result of this section and gives an upper bound on the number of queries to U and $U^\dagger$ needed to solve the Hamiltonian learning problem.

The process of generating the pseudo-Choi state contains three main steps. First, the time evolution black-box is used to generate a unitary matrix that encodes the Hamiltonian (referred to as a block-encoding of the Hamiltonian). Second, this block-encoding is applied in a controlled fashion to a maximally entangled state ($|\phi_d\rangle_{SA}$) and an ancilla qubit. This generates a relative phase between a term that applies the block-encoding to the input state and a term that acts as the identity on it. Finally, the pseudo-Choi state is produced probabilistically by measuring the ancilla qubit in the computational basis. The outcome of this measurement indicates whether or not the state was successfully generated.

An important issue that arises here is that we need to assume that the time evolution can be implemented controllably. This is often the case when we are considering learning a Hamiltonian applied within a simulator, but may not necessarily be straight forward to implement when learning a physical Hamiltonian that cannot be so manipulated. In such cases, we can still perform the control using controlled swap operations if a $+1$ eigenstate of the unitary is known. Further, it is worth noting that in practice we also need the inverse of the evolution using our approach. If these assumptions are met, however, we are able to show that we can efficiently generate the learning resource. Additionally, there exist classes of Hamiltonians ("time-reversible Hamiltonians") for which the backward time evolution can be performed efficiently even if one only has access to a black-box that performs the forward evolution. For example, certain Heisenberg models that can be represented as bipartite - the simplest being a 1-dimensional Ising model - can be easily time-reversed given access to the time evolution black-box. Therefore, if a Hamiltonian can be efficiently broken up into relatively few such Hamiltonians, our approach remains feasible as one can learn the time-reversible Hamiltonians one by one until all the coefficients of the original Hamiltonian are recovered.

4.1 Generating pseudo-Choi States

As mentioned above, the process of producing the pseudo-Choi state begins by using the time evolution unitary to create a block-encoding of the Hamiltonian. Recall that the pseudo-Choi state consists of three subsystems. The map containing the Hamiltonian acts on the subsystem S (denoted by the subscript S), which resides in a d dimensional Hilbert space $\mathcal{H}_S$. Subsystem A is an ancilla of the same size as S. Lastly, subsystem C is a single qubit used to retain information about the overall sign and norm of the Hamiltonian. In addition to these three, we will now require an additional subsystem, B, which is a single qubit used for block-encoding purposes, and will be referred to as the "block-encoding qubit". We formalize the notion of block-encoding below.

Definition 17 (Block-Encoding)**.** *Let U_{block} be a unitary operator acting on the Hilbert space $\mathcal{H}_B \otimes \mathcal{H}_S$. We then say that this unitary is a block-encoding (or more specifically a $(\Delta, \dim(\mathcal{H}_B), 0)$-block-encoding as defined in Section 4 in [26]) of a Hamiltonian H if*

$$\frac{H}{\Delta} = (\langle 0|_B \otimes \mathbb{I}_S) U_{block} (|0\rangle_B \otimes \mathbb{I}_S) \tag{37}$$

where Δ is a normalization factor.

In our context, this unitary block-encoding plays the same role as a unitary that prepares a purified Gibbs state in other approaches to quantum Hamiltonian learning. Rather than a Gibbs state, U_{block} is used to prepare a pseudo-Choi state, from which we can then recover the Hamiltonian coefficients as in Algorithms 1 and 2 in Section 3.2.

Gilyén et al. [26] developed a method for generating a block-encoding of the logarithm of an operator, which is precisely what we seek to do with the time evolution operator. Their results, modified to fit our problem, are summarized in Lemma 18 below.

Lemma 18 (Producing a Block-Encoded Hamiltonian from a Time Evolution Unitary [26])**.** *Let $U = e^{-iHt}$ where the H is Hermitian and $\|Ht\| \leq \frac{1}{2}$. Let $\epsilon_b \in (0, \frac{1}{2}]$ be the block-encoding error. Using*

$$\mathcal{O}\left(\log(1/\epsilon_b)\right)$$

controlled queries to U and U^{-1}, as well as $\mathcal{O}\left(\log(1/\epsilon_b)\right)$ 2-qubit gates and one ancilla qubit, a block-encoding of the Hamiltonian of the form

$$U_{block} = \begin{bmatrix} \frac{2\widetilde{H}t}{\pi} & I \\ J & K \end{bmatrix} \tag{38}$$

can be produced such that

$$\left\| Ht - \widetilde{H}t \right\| \leq \epsilon_b \tag{39}$$

The proof of Lemma 18 can be found in [26] (Lemma 18 is identical to Corollary 71 in [26], only with H replaced by $\widetilde{H}t$), but the general idea is that $\sin(Ht) = \frac{e^{iHt} - e^{-iHt}}{2i}$ can be implemented via Linear Combination of Unitaries (LCU) techniques, resulting in a block-encoding of $\sin(Ht)$. The Quantum Singular Value Transform (QSVT) can then be used to construct a polynomial that approximates

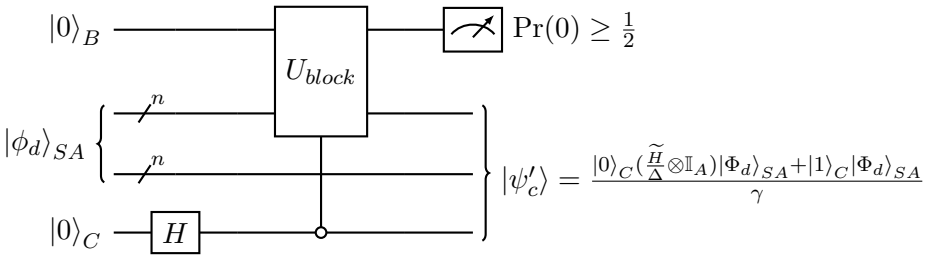

Figure 4.1: A quantum circuit that produces the pseudo-Choi state in Lemma 19. If the qubit in register B is found to be in the state $|0\rangle$ after measuring it in the computational basis, the output of the remaining registers is the pseudo-Choi state $|\psi_c'\rangle$. This outcome occurs with probability equal to $\frac{\|\mathbf{c}\|_2^2}{2\Delta^2} + \frac{1}{2}$, where $\|\mathbf{c}\|_2$ is the 2-norm of the vector of Hamiltonian coefficients, and $\Delta = \frac{\pi}{2t}$.

the arcsin of this quantity, which results in an approximate block-encoding of $\frac{2Ht}{\pi}$. This can be achieved using either quantum singular value transformation methods or linear combinations of unitaries [26–28]

Using the block-encoding of the Hamiltonian from Lemma 18, we can now generate pseudo-Choi states, as per Lemma 19. Because the Hamiltonian in the block-encoding contains a factor of $\frac{2t}{\pi}$, the pseudo-Choi state produced will be slightly different than the one in Section 3 (Equation (4)). That is, rather than a pseudo-Choi state of $\widetilde{H}$, the state produced is a pseudo-Choi state of $\frac{2\widetilde{H}t}{\pi}$:

$$|\psi_c'\rangle = \frac{|0\rangle_C \left(\frac{\widetilde{H}}{\Delta} \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA} + |1\rangle_C |\Phi_d\rangle_{SA}}{\gamma}, \tag{40}$$

where

$$\gamma = \sqrt{\frac{\|\widetilde{\mathbf{c}}\|_2^2}{\Delta^2} + 1}, \tag{41}$$

and

$$\Delta = \frac{\pi}{2t}. \tag{42}$$

The vector $\widetilde{\mathbf{c}}$ above is the vector of Hamiltonian coefficients that correspond to the block-encoded Hamiltonian $\widetilde{H}$ (see Definition 20 in the following section).

Lemma 19 (Generating pseudo-Choi States from the Hamiltonian Block-Encoding)**.** *Let t be a positive number in $(0, 1/2\|H\|]$. There exists an algorithm that yields a pseudo-Choi state of the form $|\psi_c'\rangle = \frac{|0\rangle_C(\frac{\widetilde{H}}{\Delta}\otimes\mathbb{I}_A)|\Phi_d\rangle_{SA}+|1\rangle_C|\Phi_d\rangle_{SA}}{\gamma}$, where $\Delta = \frac{\pi}{2t}$ and $\gamma = \sqrt{\frac{\|\widetilde{\mathbf{c}}\|_2^2}{\Delta^2} + 1}$, using a single controlled application of U_{block} whose success is heralded by a measurement that occurs with probability strictly greater than $\frac{1}{2}$.*

Figure 4.1 gives an explicit circuit for generating the pseudo-Choi state from the Hamiltonian block-encoding U_{block}, and the details of the proof of Lemma 19 can be found in Appendix B.5. Note that the restriction on the evolution time considered above is inherited from Lemma 18.

4.2 Learning via Shadow Tomography

Since the block-encoding U_{block} is used to produce our resource states, the learning procedure will generate an approximation of the Hamiltonian within the block-encoding. This means that there will be additional error introduced due to the fact that the Hamiltonian in the block-encoding will not be exactly equal to the true Hamiltonian.

We now rephrase the Hamiltonian learning problem in terms of the true Hamiltonian, its block-encoded approximation, and the Hamiltonian recovered by the learning procedure, as well as their corresponding coefficient vectors, as follows:

Definition 20. *Let us take the following three definitions for the true Hamiltonian H, the block-encoded Hamiltonian $\tilde{H}$, and the approximation $\hat{H}$ that results from the learning process:*

1. *Let $H := \sum_{m=0}^{M-1} c_m H_m$ be the Hamiltonian we seek to learn, and let $\boldsymbol{c}$ be the vector of its Hamiltonian coefficients (that is, the coefficients of its representation in some orthonormal basis $\{H_m \mid m \in \mathbb{Z}_M\}$).*
2. *Let $\widetilde{H} := \sum_{m=0}^{M-1} \widetilde{c}_m H_m$ be the approximation of H that results from the block-encoding procedure, and let $\widetilde{\boldsymbol{c}}$ be the vector of its Hamiltonian coefficients.*
3. *Let $\hat{H} := \sum_{m=0}^{M-1} \hat{c}_m H_m$ be the approximation of $\widetilde{H}$ that results from the learning procedure, and let $\hat{\boldsymbol{c}}$ be the vector of its Hamiltonian coefficients.*

These above definitions naturally lead us to the following definition for the Hamiltonian learning problem.

Definition 21 (Unitary Hamiltonian Learning Problem)**.** *The Hamiltonian learning problem is to compute, for any $\epsilon > 0$ and $\delta > 0$, a vector $\hat{\boldsymbol{c}}$ such that with probability at least $1 - \delta$,*

$$\|\hat{\boldsymbol{c}} - \boldsymbol{c}\|_2 \leq \epsilon, \tag{43}$$

by querying $U = e^{-iHt}$ and $U^\dagger$.

We also define the task of learning the approximate Hamiltonian $\widetilde{H}$ which appears in the block-encoding as follows:

Definition 22 (Block-Encoded Hamiltonian Learning Problem)**.** *The block-encoded Hamiltonian learning problem is to compute a vector $\hat{\boldsymbol{c}}$ such that with probability at least $1 - \delta$,*

$$\|\hat{\boldsymbol{c}} - \widetilde{\boldsymbol{c}}\|_2 \leq \epsilon_c, \tag{44}$$

by querying U_{block}

Recall that Lemma 18 also gives us ϵ_b, the upper bound on the error of the block-encoded Hamiltonian.

From the pseudo-Choi state (40), we can learn the Hamiltonian coefficients in similar fashion to the PC HLP considered above. Specifically, recall that

$$O_l = |0\rangle_C \langle 1|_C \otimes (H_l \otimes \mathbb{I}) |\Phi_d\rangle_{SA} \langle \Phi_d|_{SA} \tag{45}$$

where H_l is a Hamiltonian term, and let

$$O_\gamma = |1\rangle_C \langle 1|_C \otimes |\Phi_d\rangle_{SA} \langle \Phi_d|_{SA} \tag{46}$$

Define the Pseudo-Choi state for the density operator ρ'_c according to the following

$$\rho'_c := |\psi'_c\rangle \langle \psi'_c| , \tag{47}$$

so that

$$\mathrm{Tr}(\rho'_c O_l) = \frac{c_l}{\Delta\gamma^2}, \tag{48}$$

where c_l is the l^{th} Hamiltonian coefficient and

$$\mathrm{Tr}(\rho'_c O_\gamma) = \frac{1}{\gamma^2} \tag{49}$$

where γ is defined in (41) The entire learning process, outlined in Algorithm 4, therefore consists of first generating a sufficient amount of pseudo-Choi states (40) from the Hamiltonian, running Algorithm 1, and then scaling the result by Δ.

4.3 Query Complexity for Hamiltonian Learning from Time Evolution Unitaries

Rather than sample complexity, it is more meaningful to look at the query complexity for this model - that is, the number of (controlled) queries to $U = e^{-iHt}$. The main result of this section is Theorem 26, which gives an upper bound on the number of times the time-evolution unitary must be queried in order to solve the Unitary HLP 21. We begin by finding a similar upper bound for an easier task, in particular solving the Block-Encoded HLP 22.

Theorem 23 (Query Complexity Upper Bound for Solving the Block-Encoded HLP)**.** *The number of queries to the block-encoding* (37) *required to solve the Block-Encoded Hamiltonian Learning Problem 22 for a Hamiltonian whose M Hamiltonian terms are known, is at most*

$$\widetilde{N} \in \mathcal{O}\left(\frac{M\gamma^2(\widetilde{c}^2_{max} + \frac{1}{t^2})\log(M/\delta)}{\epsilon_c^2}\right) \tag{50}$$

Proof of Theorem 23. The sample complexity of predicting the expectation values of the $2M$ operators in Algorithm 4 using classical shadows is

$$N_s \in \mathcal{O}\left(\frac{\log(M/\delta_s)}{\epsilon_s^2}\right) \tag{51}$$

This is the number of pseudo-Choi states (40) required to predict $\frac{1}{\gamma^2}$ and $\hat{o}_m = \frac{\widetilde{c}_m}{\gamma^2\Delta}$ to within error ϵ_s for all $m \in \mathbb{Z}_M$ with probability at least $1-\delta_s$. Multiplying each $\hat{o}_m$ by $\gamma^2\Delta$ gives the coefficients, $\widetilde{c}_m$, of the block-encoded Hamiltonian within error

$$\epsilon_m \equiv \gamma^2\epsilon_s\sqrt{\widetilde{c}_m^2 + \Delta^2}. \tag{52}$$

Algorithm 4 `FindCoeffUnitary`$(U, U^\dagger, \Delta, \boldsymbol{H}, n, N_s)$: Learn the Hamiltonian Coefficients by Querying $U = e^{-iHt}$ and $U^\dagger$

Input:

U: Time evolution operator e^{-iHt} acting on Hilbert space $\mathcal{H}_S$

Δ : Normalization constant equal to $\frac{\pi}{2t}$ where $t \in \left(0, \frac{1}{2\|H\|}\right]$ is the evolution time used in U, which dictates the probability of success for projection onto the pseudo-Choi state measurement.

$\boldsymbol{H}$: Set of M k-local Hamiltonian terms H_l whose coefficients c_m are desired

▷ Note: $\boldsymbol{H}$ is a set of classical operators

n: Number of qubits in the system on Hilbert space $\mathcal{H}_S$

N_s: Number pseudo-Choi states desired

1. **Generate pseudo-Choi states**

   Run the circuit from Figure 4.1 on $N \in \widetilde{\mathcal{O}}(N_s)$ independent input states to produce N_s pseudo-Choi states of the form $\rho'_c = |\psi'_c\rangle\langle\psi'_c|$, where $|\psi'_c\rangle = \frac{|0\rangle_C(\frac{\widetilde{H}}{\Delta}\otimes\mathbb{I}_A)|\Phi_d\rangle_{SA}+|1\rangle_C|\Phi_d\rangle_{SA}}{\gamma}$ with high probability.

   ▷ An explicit upper bound on N is given in Theorem 26

   ▷ Recall that U_{block} can be produced by querying U and $U^\dagger$ as per Lemma 18, using the methods in [26]

2. **Run Algorithm 1 to estimate** $\frac{\hat{\mathbf{c}}}{\Delta}$

   `Set` $\hat{\mathbf{c}} \leftarrow$ `FindCoeffClifford`$(\rho_c'^{\otimes N_s}, \boldsymbol{H}, n, N_s)$

   ▷ Algorithm 5 can be used as well

3. **Rescale the result by** Δ **as per equation** (48)

   `Return` $\Delta\hat{\mathbf{c}}$

Recall that learning these coefficients is the goal of the Block-Encoded HLP in Definition 22. The l_2 error between the vector of Hamiltonian coefficients returned by Algorithm 4 and that of the block-encoded Hamiltonian is

$$\begin{aligned} \|\widetilde{\mathbf{c}} - \hat{\mathbf{c}}\|_2 &\leq \sqrt{M} \max_m \epsilon_m \\ &= \sqrt{M}\gamma^2\epsilon_s\sqrt{\widetilde{c}_{max}^2 + \Delta^2}, \end{aligned} \tag{53}$$

where $\widetilde{c}_{max}$ is the coefficient with the largest absolute value. To satisfy the Block-Encoded HLP 22, we let the upper bound on this error be equal to ϵ_c. It follows that the following value of ϵ_s is sufficient to solve the Block-Encoded HLP:

$$\epsilon_s = \frac{\epsilon_c}{\sqrt{M}\gamma^2\sqrt{\widetilde{c}_{max}^2 + \Delta^2}}. \tag{54}$$

Substituting this value for ϵ_s into Equation (51) means that the total number of pseudo-Choi states we need to solve the Hamiltonian Learning Problem is

$$N_s \in \mathcal{O}\left(\frac{M\gamma^4(\widetilde{c}_{max}^2 + \Delta^2)\log\left(M/\delta_s\right)}{\epsilon_c^2}\right) \tag{55}$$

Now that we have an upper bound on the number of pseudo-Choi states (40) required to solve the Block-Encoded HLP 22 with probability at least $1-\delta_s$, we would like to know how many queries to U_{block} are needed to solve the Block-Encoded HLP with probability at least $1-\delta$. We must therefore consider the failure probability of the learning algorithm.

To overestimate the failure probability of the entire learning procedure, we assume that it will always fail if less than N_s pseudo-Choi states are produced, and also that the success probability of classical shadows is the same as long as the number of pseudo-Choi states generated is at least N_s. The probability of our learning algorithm failing is then

$$\begin{aligned} \mathbf{P}_{fail} &\leq \delta_s + \delta_{N_s} \\ &\equiv \delta, \end{aligned} \tag{56}$$

where δ_s is the probability of classical shadows failing if we have N_s pseudo-Choi states, δ_{N_s} is the probability of producing less than N_s pseudo-Choi states after $\widetilde{N}$ queries to U_{block}, and the third line comes from the requirement of the Block-Encoded HLP 22 which says that the failure probability must be at most δ. We then let $\delta_s \equiv \delta_{N_s} \equiv \frac{\delta}{2}$. This choice of δ_s gives

$$N_s \in \mathcal{O}\left(\frac{M\gamma^4(\tilde{c}_{max}^2+\Delta^2)\log\left(M/\delta\right)}{\epsilon_c^2}\right), \tag{57}$$

Next, it can be shown via Chernoff bound that

$$\widetilde{N} \in \mathcal{O}\left(\frac{N_s}{\gamma^2}\right) \tag{58}$$

queries to the the block-encoding U_{block} are needed to ensure N_s pseudo-Choi states are produced with probability at least $1-\delta$. Note that the asymptotic notation may be misleading here since γ^2 is larger than 1, but $\widetilde{N}$ is in fact larger than N_s (see Appendix B.6). Combining this with the upper bound on N_s above, and using $\Delta = \frac{\pi}{2t}$, we get

$$\widetilde{N} \in \mathcal{O}\left(\frac{M\gamma^2(\tilde{c}_{max}^2+\frac{1}{t^2})\log\left(M/\delta\right)}{\epsilon_c^2}\right), \tag{59}$$

proving the result in Theorem 23.

□

4.3.1 Choosing Sufficient Values for the Block-Encoding Error and Block-Encoding Learning Error

Theorem 23 gives an upper bound on the number of Hamiltonian block-encodings needed to solve the Block-Encoded HLP 22. Note that solving the Block-Encoded HLP is just a sub-task of solving the Unitary HLP 21, with the missing step being the generation of the Hamiltonian block-encodings from the time evolution unitary e^{-iHt}. This means that to determine the total query complexity, we have to factor in the number of queries required to produce the desired number of block-encodings of the Hamiltonian. In addition, since the block-encoding generated by the procedure described in Lemma 18 has an error of ϵ_b, the learning algorithm will not approximate the coefficients of the Hamiltonian terms of H, but those of the approximation of H in the block-encoding. Thus, the block-encoding error ϵ_b will have a further impact on the total query complexity.

We now seek values for the block-encoding error ϵ_b and the block-encoding learning error ϵ_c, such that the total error in learning the Hamiltonian is at most ϵ, as required by the Unitary HLP 21. We begin with the following proposition:

Proposition 24. *The error in the vector, $\widetilde{\mathbf{c}}$, of Hamiltonian coefficients of the block-encoded Hamiltonian $\widetilde{H}$ is upper bounded by*

$$\|\widetilde{\boldsymbol{c}} - \boldsymbol{c}\|_2 \leq \frac{\sqrt{M}\epsilon_b}{t},$$

where M is the number of Hamiltonian terms, t is the evolution time used to generate the block-encoding, and ϵ is the block-encoding error, as in Lemma 18.

Proof of Proposition 24.

$$\begin{aligned}
|\widetilde{c}_m - c_m| &= \left|\frac{1}{d}\mathrm{Tr}(\widetilde{H}H_m) - \frac{1}{d}\mathrm{Tr}(HH_m)\right| \\
&= \frac{1}{d}\left|\mathrm{Tr}((\widetilde{H} - H)H_m)\right| \\
&\leq \frac{1}{d}\sum_{i=1}^{k}\sigma_i(\widetilde{H} - H)\sigma_i(H_m) \\
&\leq \frac{1}{d}\sigma_{max}(\widetilde{H} - H)\sum_{i=1}^{k}\sigma_i(H_m) \\
&= \frac{1}{d}\sigma_{max}(\widetilde{H} - H)\sum_{i=1}^{k}|\lambda_i(H_m)| \\
&= \frac{1}{d}\left|\sigma_{max}(\widetilde{H} - H)\right|\mathrm{Tr}\left(\sqrt{H_m^\dagger H_m}\right) \\
&= \left|\sigma_{max}(\widetilde{H} - H)\right| \\
&= \left\|\widetilde{H} - H\right\|_2
\end{aligned} \tag{60}$$

In the first inequality we have used the Von Neumann trace inequality, where $\sigma_i(A)$ is the i^{th} of the k singular values of A when sorted in descending order. Likewise, $\sigma_{max}(\widetilde{H} - H)$ is the singular value of $\widetilde{H} - H$ with the largest magnitude.

Noting that this result does not depend on m, we use it to finish the proof as follows:

$$\begin{aligned}
\|\widetilde{\mathbf{c}} - \mathbf{c}\|_2 &\leq \sqrt{M}\,\|\widetilde{\mathbf{c}} - \mathbf{c}\|_\infty \\
&= \sqrt{M}\max_m |\widetilde{c}_m - c_m| \\
&\leq \sqrt{M}\left\|\widetilde{H} - H\right\|_2 \\
&\leq \frac{\sqrt{M}\epsilon_b}{t}
\end{aligned} \tag{61}$$

where we use the result of Lemma 18 in the last line. □

The next step in choosing sufficient values for ϵ_b and ϵ_c is to upper bound the total learning error from solving the Unitary HLP in terms of ϵ_b and ϵ_c, as in Lemma 25.

Lemma 25 (Upper Bound on Error for the Unitary HLP)**.** *The total error between the vector of the true Hamiltonian coefficients and the vector recovered by using Algorithm 4 to solve the Unitary HLP is upper bounded by*

$$\|\hat{\boldsymbol{c}} - \boldsymbol{c}\|_2 \leq \epsilon_c + \frac{\sqrt{M}\epsilon_b}{t}$$

where ϵ_c is the error in learning the block-encoded Hamiltonian (see Definition 22) and ϵ_b is the error in the block-encoding itself (see Lemma 18).

Proof of Lemma 25.

$$\begin{aligned}\|\hat{\mathbf{c}} - \mathbf{c}\|_2 &= \|\hat{\mathbf{c}} - \widetilde{\mathbf{c}} + \widetilde{\mathbf{c}} - \mathbf{c}\|_2 \\ &\leq \|\hat{\mathbf{c}} - \widetilde{\mathbf{c}}\|_2 + \|\widetilde{\mathbf{c}} - \mathbf{c}\|_2 \\ &\leq \epsilon_c + \|\widetilde{\mathbf{c}} - \mathbf{c}\|_2 \end{aligned} \tag{62}$$

Note that the second inequality is guaranteed by querying the Hamiltonian block-encoding U_{block} the sufficient amount of times to solve the Block-Encoded HLP (22). Theorem 23 gives an upper bound on this number of queries.

Combining this with the result of Proposition 24 gives the result in Lemma 25. □

Finally, to achieve the desired upper bound of $\|\hat{\mathbf{c}} - \mathbf{c}\|_2 \leq \epsilon$ for the total error of the Hamiltonian learning problem (21), we simply choose ϵ_b and ϵ_c such that the right side of the inequality in Lemma 25 is less than ϵ. Since the learning process is much more expensive than the block-encoding process, it makes sense to allocate most of the error budget to ϵ_c, as will be discussed in the following section.

4.3.2 Query Complexity Upper Bound

With the previous results in place, we can now state our bounds on the query complexity for the Unitary HLP 21. Unlike in the Block-Encoded HLP, and the PC HLP, we now consider unitary preparation of the pseudo-Choi states via the process described in Section 4.1. Note that in the following we do not make use of amplitude amplification as the success probability for this learning problem is at least 1/2, which implies that amplitude amplification will not lead to asymptotic advantages.

Theorem 26 (Query Complexity Upper Bound for Solving the Unitary HLP)**.** *The number of queries to the time evolution unitary e^{-iHt} and its inverse required to solve the Unitary Hamiltonian Learning Problem 21 for a Hamiltonian with M Hamiltonian terms, is at most*

$$N \in \mathcal{O}\left(\frac{M\log\left(M/\delta\right)}{t^2\epsilon^2}\log\left(\frac{M}{t\epsilon}\right)\right) \tag{63}$$

where $t \in \mathcal{O}(1/\|H\|)$

Proof of Theorem 26. From Theorem 23 we have that the number of block-encodings needed to solve the Block-Encoded HLP 22 is

$$\widetilde{N} \in \mathcal{O}\left(\frac{M\gamma^2(\widetilde{c}_{max}^2 + \frac{1}{t^2})\log\left(M/\delta\right)}{\epsilon_c^2}\right) \tag{64}$$

Lemma 18 states that $\mathcal{O}\left(\log\left(\frac{1}{\epsilon_b}\right)\right)$ queries to the time evolution blackbox $U = e^{-iHt}$ and its inverse are required to generate each block-encoding. Since we apply Lemma 18 to produce the block-encoding, it is important to note the preconditions on ϵ and $\epsilon_b \leq 1/2$. The latter of which is guaranteed by our assumption on ϵ if $\epsilon_b \leq \epsilon$. Next, recall that $\frac{\gamma^2}{2}$ is the probability of projecting onto a pseudo-Choi state upon measuring the block-encoding qubit. Lemma 19 then implies that $\gamma^2 \in \mathcal{O}(1)$ under the assumption that $t \in \mathcal{O}(1/\|H\|)$. Furthermore, we make the assumption that each Hamiltonian coefficient satisfies $|c_m| \leq 1$. As the norm of the Hamiltonian coefficients can be rescaled by choosing the evolution time to be larger or smaller, we can make such a choice for our state learning without restricting the class of Hamiltonians that is learnable within our approach.

Therefore, the upper bound on the total query complexity for the Hamiltonian learning problem (21) is

$$N \in \mathcal{O}\left(\frac{M\log\left(M/\delta\right)}{t^2\epsilon_c^2}\log\left(\frac{1}{\epsilon_b}\right)\right) \tag{65}$$

To satisfy Problem 21, the total error must be $\|\hat{\mathbf{c}} - \mathbf{c}\|_2 \leq \epsilon$. From Lemma 25 we see that this can be accomplished by choosing ϵ_c and ϵ_b such that $\epsilon_c + \frac{M\epsilon_b}{t} \leq \epsilon$. If we assign an error budget of $\frac{\epsilon}{2}$ to each of these terms, that is choose

$$\epsilon_c := \epsilon/2, \tag{66}$$

and

$$\epsilon_b := \frac{\epsilon t}{2M}, \tag{67}$$

we arrive at the query complexity given in Theorem 26. □

Note that the choice of error budget is somewhat arbitrary for the purposes of determining the asymptotic scaling, as it only affects the query complexity by a constant factor. However, since the query complexity is quadratic in ϵ_c^{-1} but only logarithmic in ϵ_b^{-1}, it is logical to assign most of the error budget to ϵ_c. For example, choosing

$$\epsilon_c \leq \left(1 - \frac{1}{M}\right)\epsilon \tag{68}$$

and

$$\epsilon_b \leq \frac{t\epsilon}{M^2} \tag{69}$$

keeps the total error to at most ϵ and limits the increase in query complexity to a constant factor of approximately 2 for large values of M. Our assumption that $\epsilon_b \leq 1/2$ then immediately follows from $\epsilon \leq M^2/(2t)$.

5 Hamiltonian Learning via Quantum Mean-Value Estimation

The approach considered previously has great strengths. Notably, the quantum operations required are simple Pauli or Clifford operations (depending on the flavour of classical shadows used), and

after measurements are complete, all expectation values can be calculated efficiently on a classical computer. It should be noted that it also allows us to learn the Hamiltonian without having an explicit dependence on the number of terms in the Hamiltonian if we look at the l_∞ error in the Hamiltonian coefficients (though we have focused on the l_2 error metric throughout this paper, as it better highlights the differences between the query complexities of our approach and previous ones). However, the most significant drawback of this approach is that the method scales poorly with $1/\epsilon$. This can be ameliorated by dropping the use of shadows and using a quantum computer to directly estimate all of the terms directly from the PC state by computing the expectation values of the Hamiltonian using mean-value estimation algorithms.

In [19] an algorithm is proposed for estimating the expectation value of a vector of observables. The central idea behind this approach is to take the problem of mean estimation and reduce it to a gradient estimation procedure on the exponential of the sum of Hamiltonian terms. The gradient estimation procedure can then be executed optimally using the technique of [29], which then allows the vector of mean values to be estimated using a number of queries that also scales optimally with the parameters involved. The key result is that for some state ψ, the mean values of a set of M operators with norm at most ν_{max} can be computed within l_∞ error ϵ' and probability of failure at most δ using $\widetilde{\mathcal{O}}(\nu_{max}\sqrt{M}/\epsilon')$ queries to the unitary that prepares ψ as well as to oracles that perform the exponentials of each of the operators. In Lemma 29 we show that the number of queries of the same form required to learn a Hamiltonian using this technique is $\widetilde{\mathcal{O}}(\alpha^2 M/\epsilon)$, where ϵ is the l_2 error in the Hamiltonian coefficients and α^2 is the normalization factor in the pseudo-Choi state.

The main result of this section is given later in Theorem 33, which gives an upper bound on the number of queries to the time-evolution unitary that are needed to solve the Hamiltonian learning problem. However, we begin in Section 5.1 by proving Lemma 29, which solves a simpler problem. In particular, it assumes we have a unitary oracle that prepares our resource state, and gives an upper bound on the number of queries to the oracle needed to solve the Hamiltonian learning problem. Next, in Section 5.2 we show how to generate the resource state in unitary fashion by querying the time-evolution operator and its inverse and then prove Theorem 33.

5.1 Learning the Hamiltonian Coefficients from a Unitary Preparation of the Resource State

As in the case of Pauli-based classical shadows, we shall start with the pseudo-Choi state and perform a partial trace over the ancillary system A, resulting in the following resource state:

$$\rho = \frac{1}{d\alpha^2}\Big(H^2 \otimes |0\rangle_C \langle 0|_C + H \otimes |0\rangle_C \langle 1|_C + H \otimes |1\rangle_C \langle 0|_C + \mathbb{I} \otimes |1\rangle_C \langle 1|_C \Big), \tag{70}$$

where

$$\alpha = \sqrt{\|\mathbf{c}\|_2^2 + 1} \tag{71}$$

and $\mathbf{c}$ is the vector of Hamiltonian coefficients.

Next, we define the decoding operators we will use to extract the Hamiltonian coefficients from our resource state.

Definition 27 (Decoding Operators)**.** *Let the set of decoding operators be defined as*

$$\boldsymbol{O} \equiv \{O_l | l \in \mathbb{Z}_M\} \tag{72}$$

where

$$O_l = \frac{H_l \otimes X_C}{2} \tag{73}$$

and H_l is one of the M Hamiltonian terms.

Furthermore, we define

$$O_\alpha \equiv \mathbb{I} \otimes |1\rangle_C \langle 1|_C \tag{74}$$

Proposition 28 shows that for our resource state ρ, the expectation values of the decoding operators can be used to recover the Hamiltonian coefficients, as well as the normalization factor of the resource state.

Proposition 28 (Computing the Hamiltonian Coefficients)**.** *If ρ is the resource state (70), then $\forall$ $l \in \mathbb{Z}_M$ we have*

$$\mathrm{Tr}(\rho O_l) = \frac{c_l}{\alpha^2}, \tag{75}$$

where c_l is the l^{th} Hamiltonian coefficient, and α is the normalization constant of the pseudo-Choi state (7). Furthermore,

$$\mathrm{Tr}(\rho O_\alpha) = \frac{1}{\alpha^2} \tag{76}$$

Proof of Proposition 28. From the definitions of O_l and the resource state ρ, it is easy to see that

$$\mathrm{Tr}(\rho O_l) = \frac{\mathrm{Tr}(H H_l)}{d\alpha^2}. \tag{77}$$

Expanding the above using the representation of the Hamiltonian from Definition 1 gives

$$\begin{aligned} \mathrm{Tr}(\rho O_l) &= \frac{\sum_m c_m \mathrm{Tr}(H_m H_l)}{d\alpha^2} \\ &= \frac{1}{d\alpha^2} \sum_m c_m \delta_{ml} d \\ &= \frac{c_l}{\alpha^2} \end{aligned} \tag{78}$$

The second claim ($\mathrm{Tr}(\rho O_\alpha) = \frac{1}{\alpha^2}$) follows immediately from Equation (135) and the definition of O_α. □

Next, we show that the Hamiltonian learning problem can be efficiently solved by querying a unitary oracle that prepares the resource state.

Lemma 29. *Let U_ρ be a unitary oracle that prepares the state ρ from equation (70). The Hamiltonian learning problem in Definition 1 can be solved using*

$$\widetilde{\mathcal{O}}\left(\frac{\alpha^2 M}{\epsilon}\right)$$

queries to U_ρ and $U_\rho^\dagger$, as well as to an oracle that implements controlled-$e^{-i\theta O_l}$ for each decoding operator O_l.

Proof. By Proposition 28, we have

$$\mathrm{Tr}(\rho O_l) = \frac{c_l}{\alpha^2}, \tag{79}$$

From this we can see that the Hamiltonian can be estimated by estimating a mean-value of a set of Hamiltonian operators. Specifically, consider the vector $\mathbf{d}$

$$\mathbf{d} := [\mathrm{Tr}(\rho O_0), \ldots, \mathrm{Tr}(\rho O_{M-1})] \tag{80}$$

from which the vector of Hamiltonian coefficients can be computed via

$$c_l = \alpha^2 d_l \tag{81}$$

Since the Hamiltonian learning problem considers Hamiltonian terms to have a norm that is at most a constant (so $\nu_{max} \in \mathcal{O}(1)$), the mean value estimation algorithm of [19] can be used to estimate $\mathbf{d}$ within infinity norm ϵ' using

$$N \in \widetilde{\mathcal{O}}(\sqrt{M}/\epsilon') \tag{82}$$

queries to U_ρ and $U_\rho^\dagger$, as well as to controlled-$e^{-i\theta O_l}$ $\forall l \in [0, M-1]$. All that is left now is to determine how small ϵ' should be to guarantee that the l_2 error in the estimate of the vector of Hamiltonian coefficients is at most ϵ.

If we define $\hat{c}_l$ to be our estimate of c_l, then

$$\begin{aligned} |\hat{c}_l - c_l| &\leq \alpha^2 \epsilon' \sqrt{c_l^2 + 1} \\ &\in O\left(\alpha^2 \epsilon'\right), \end{aligned} \tag{83}$$

since we need to scale the result of the mean estimation by α^2 (as in equation (81)). Therefore, if we want the l_∞ error in estimating c_l to be at most ϵ_∞, we require

$$\epsilon' \in \mathcal{O}(\epsilon_\infty/\alpha^2) \tag{84}$$

Finally, for the 2-norm of the error in the estimate of the vector of Hamiltonian coefficients to be at most ϵ, we choose $\epsilon_\infty \in \Theta(\epsilon/\sqrt{M})$ (because $\|\cdot\|_2 \leq \sqrt{M}\|\cdot\|_\infty$ from Cauchy Schwarz). This leads to a query complexity of

$$N \in \widetilde{\mathcal{O}}(\alpha^2 M/\epsilon). \tag{85}$$

□

We now wish to use the result of Lemma 29 to determine how many queries to the time evolution blackbox will be necessary to solve the Unitary HLP.

5.2 Constructing a Unitary Preparation of the Resource State

In Lemma 29 we showed that the Hamiltonian Learning Problem could be solved using $\widetilde{\mathcal{O}}\left(\frac{\alpha^2 M}{\epsilon}\right)$ queries to a unitary operation that prepares the state ρ from equation (70). We would now like

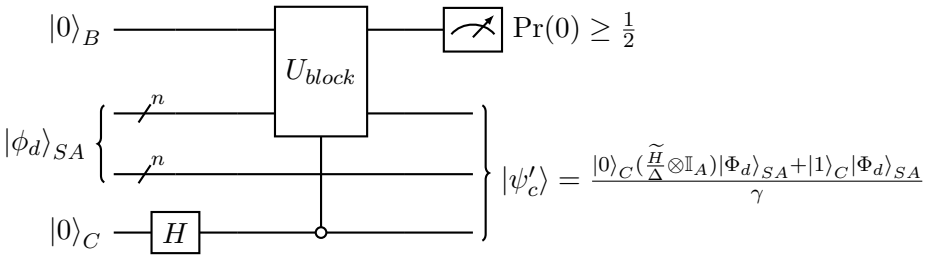

Figure 5.1: A quantum circuit that produces the pseudo-Choi state of $\widetilde{H}/\Delta$. If the qubit in register B is found to be in the state $|0\rangle$ after measuring it in the computational basis, the output of the remaining registers is the pseudo-Choi state. This outcome occurs with probability equal to $\frac{\|\mathbf{c}\|_2^2}{2\Delta^2} + \frac{1}{2}$, where $\|\mathbf{c}\|_2$ is the 2-norm of the vector of Hamiltonian coefficients, and $\Delta = \frac{\pi}{2t}$.

to show that this state can be prepared in unitary fashion by querying the time-evolution operator $U = e^{-iHt}$ and its inverse. The idea is that if we know the number of queries to U and $U^\dagger$ it takes to prepare this state, we can use Lemma 29 to find the number of queries to U and $U^\dagger$ needed to solve the Unitary HLP. We begin by considering the circuit from Section 4.1 (reproduced here in Figure 5.1) that was used to generate the pseudo-Choi state. Recall that U_{block} contains a block-encoding, $\widetilde{H}$, of the Hamiltonian H, and Lemma 18 shows that U_{block} can be efficiently generated by querying the time-evolution unitary $U = e^{-iHt}$ and its inverse. A full analysis of the circuit can be found in Appendix B.5.

The first step in producing the state in Equation 70 is to produce the pseudo-Choi state, in similar fashion to Figure 5.1. However, since we would like to use the result of Lemma 29, we must make note of two things. First, the pseudo-Choi state we get from querying the time-evolution operator is not of the Hamiltonian H, but of $\widetilde{H}/\Delta$ (where $\widetilde{H}$ is the block-encoding of the Hamiltonian and $\Delta = \frac{\pi}{2t}$), meaning that we will end up with a state that is slightly different from the one in equation 70. We will address this later when looking at the query complexity. The second thing to take note of is that Lemma 29 requires a unitary preparation of the resource state, but the circuit in Figure 5.1 involves measuring the block-encoding qubit. We can get around this measurement by using fixed-point amplitude amplification to approximate the pseudo-Choi state as follows.

The fixed-point amplitude amplification (FPAA) process of Yoder et al. [30] considers a starting state $|s\rangle$, a circuit A that prepares the starting state, a target state $|T\rangle$, and an oracle U_{AA} that flips an ancilla qubit when given the target state. It then applies a circuit S_L (which makes $\mathcal{O}(L)$ queries U, A, and $A^\dagger$) to the initial state such that $\|\langle T| S_L |s\rangle\|^2 \geq 1 - \delta_{AA}$ for a given $\delta_{AA} \in [0, 1]$. In our case, the preparation circuit, A, is simply the circuit from Figure 5.1, excluding the measurement of register B. Recall that this circuit queries the time evolution unitary, which is our learning resource, and prepares the state

$$|s\rangle \equiv \frac{1}{\sqrt{2}}\left(|0\rangle_C |0\rangle_B \left(\frac{\widetilde{H}}{\Delta} \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA} + |0\rangle_C |1\rangle_B (J_S \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA} + |1\rangle_C |0\rangle_B |\Phi_d\rangle_{SA} \right), \quad (86)$$

which serves as the initial state for the FPAA circuit. Our target state is the pseudo-Choi state

$$|T\rangle \equiv \frac{|0\rangle_C |0\rangle_B (\frac{\widetilde{H}}{\Delta} \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA} + |1\rangle_C |0\rangle_B |\Phi_d\rangle_{SA}}{\gamma}, \quad (87)$$

where $\gamma = \sqrt{\frac{\|\widetilde{\mathbf{c}}\|_2^2}{\Delta^2} + 1}$.

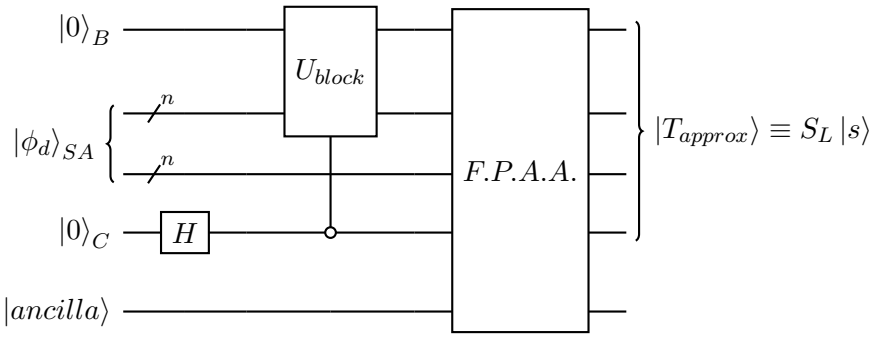

Figure 5.2: A quantum circuit that produces an approximation of the pseudo-Choi state of $\widetilde{H}/\Delta$, where $F.P.A.A.$ refers to the circuit performing fixed-point amplitude amplification due to Yoder et al. [30] . Performing a partial trace over subsystem A results in an approximation of the state in equation 70. Since this process is unitary, Lemma 29 gives an upper bound on the number of queries to this circuit needed to solve the Unitary HLP.

To perform FPAA, we need a unitary that flips an ancilla qubit when it acts on the target state. While we do not know the pseudo-Choi state (seeing as it contains all the unknown Hamiltonian coefficients), we know that if the block-encoding qubit is in the state $|0\rangle_B$, the remaining registers contain the pseudo-Choi state. Therefore, the unitary we desire is simply a zero-controlled $CNOT$ operation using register B as the control. With this, we can apply FPAA to $|s\rangle$ to get the state

$$|T_{approx}\rangle \equiv S_L\,|s\rangle\,, \tag{88}$$

which is very close to the pseudo-Choi state, $|T\rangle$, of the block-encoded Hamiltonian $\widetilde{H}/\Delta$. Following this, we simply ignore all qubits in register A, resulting in

$$\chi \equiv \mathrm{Tr}_A(|T_{approx}\rangle\langle T_{approx}|), \tag{89}$$

which is an approximation of the state

$$\begin{aligned}\widetilde{\rho} &\equiv \mathrm{Tr}_A(|T\rangle\langle T|)\\ &= \frac{1}{d\gamma^2}\Big(\frac{\widetilde{H}^\dagger\widetilde{H}}{\Delta^2}\otimes|0\rangle_C\langle 0|_C + \frac{\widetilde{H}}{\Delta}\otimes|0\rangle_C\langle 1|_C + \frac{\widetilde{H}^\dagger}{\Delta}\otimes|1\rangle_C\langle 0|_C + \mathbb{I}\otimes|1\rangle_C\langle 1|_C\Big)\otimes|0\rangle\langle 0|_B.\end{aligned} \tag{90}$$

In turn, $\widetilde{\rho}$ is essentially an approximation of the state ρ from equation (70), with the block-encoding procedure being the source of error. We must also highlight the distinction that the Hamiltonian is normalized by Δ in this case, whereas the Hamiltonian in ρ was not.

Now that we have a unitary preparation of this state, we need to determine how many iterations of FPAA are needed in order to ensure that after using the results of Lemma 29, the total error in predicting the Hamiltonian coefficients is small enough to satisfy the Unitary Hamiltonian Learning Problem from Definition 21. This will tell us the total number of queries to the time evolution operator U and it's inverse, which is the total query complexity of the algorithm. An upper bound on this value is given in Theorem 33, which we prove at the end of this section. We begin by establishing a few lemmas that capture the query complexity of specific parts of the total algorithm.

Yoder et al. [30] describe the query complexity of their fixed-point amplitude amplification algorithm in terms of the success probability of obtaining the target state after the FPAA process, which we summarize in Lemma 30.

Lemma 30 (Fixed Point Amplitude Amplification)**.** *Let $|s\rangle$ be some initial state, $|T\rangle$ be the desired target state, with their overlap given by $\langle T|s\rangle = \sqrt{\lambda}e^{i\xi}$. Let A be a circuit that prepares $|s\rangle$ from the zero state, and U_{AA} be a unitary oracle that flips an ancilla qubit when the input is $|T\rangle$. Additionally, let S_L be the circuit that performs the fixed-point amplitude amplification process using $\mathcal{O}(L)$ queries to A, $A^\dagger$, and U_{AA}. The success probability of obtaining the target state is given by $P_L = |\langle T| S_L |s\rangle|^2$. For a given δ_{AA}, the value of L needed to extract the target state with probability $P_L \geq 1 - \delta_{AA}$ is*

$$L \in \mathcal{O}\left(\frac{\log(1/\delta_{AA})}{\sqrt{\lambda}}\right)$$

Proof. See Yoder et al. [30] □

Since our preparation of the pseudo-Choi state should be unitary, and thus does not involve measurement, we restate this query complexity in Lemma 31 in terms of the trace distance between the target state $|T\rangle$ and the output state $|T_{approx}\rangle = S_L |s\rangle$, rather than in terms of a success probability. Note that since our ultimate goal is to determine the number of queries to the time evolution unitary, we only care about the queries to the preparation circuit A, which queries the time evolution unitary, and not the number of queries to U_{AA}, which as we discussed previously is just a simple 2-qubit operation.

Lemma 31 (Query Complexity of Fixed Point Amplitude Amplification)**.** *Let $|s\rangle$ be the initial state as defined in equation (86), $|T\rangle$ be the target state as defined in equation (87), and A be a circuit that prepares $|s\rangle$ from the zero state. Additionally, let S_L be the circuit that performs the fixed-point amplitude amplification process using $\mathcal{O}(L)$ queries to A and $A^\dagger$. In order for the trace distance between the target state $|T\rangle$ and the output state $|T_{approx}\rangle = S_L |s\rangle$ to be at most D, the number of queries to A and $A^\dagger$ needed is*

$$L \in \mathcal{O}\left(\log(1/D)\right)$$

Proof. Lemma 30 guarantees that $\mathcal{O}\left(\frac{\log(1/\delta_{AA})}{\sqrt{\lambda}}\right)$ queries to A and $A^\dagger$ are enough to ensure that $|\langle T| S_L |s\rangle|^2 \geq 1 - \delta_{AA}^2$, which implies that

$$\frac{1}{\delta_{AA}} \leq \frac{1}{\sqrt{1 - |\langle T| S_L |s\rangle|^2}}. \tag{91}$$

Therefore, we can express the number of queries as

$$L \in \mathcal{O}\left(\frac{1}{\sqrt{\lambda}} \log\left(\frac{1}{\sqrt{1 - |\langle T| S_L |s\rangle|^2}}\right)\right) \tag{92}$$

Using the definitions of $|s\rangle$ and $|T\rangle$ from equations (86) and (87), it can be shown that $\langle T|s\rangle \in [1/2, 1]$ (the calculation is identical to that of calculating the success probability of measuring the block-encoding qubit in the desired state, which can be found in Appendix B.5). Therefore, since $\langle T|s\rangle = \lambda e^{i\xi}$, we need not consider λ in the query complexity, as in the worst case it only affects the number of queries by a constant factor.

Since $S_L |s\rangle$ and $|T\rangle$ are pure states, the trace distance between them is given by $D = \frac{1}{2}\sqrt{1 - |\langle T| S_L |s\rangle|^2}$. With this, we can express equation (92) as

$$L \in \mathcal{O}\left(\log\left(\frac{1}{D}\right)\right) \tag{93}$$

□

Next, we prove Lemma 32, which is simply a reformulation of Lemma 29 using the state $\widetilde{\rho}$ from equation (90) rather than the state ρ from equation (70).

Lemma 32 (Query Complexity of Quantum Mean Estimation)**.** *Let $U_{\widetilde{\rho}}$ be a unitary oracle that prepares the state $\widetilde{\rho}$ from equation (90). The Hamiltonian learning problem in Definition 1 can be solved using*

$$\widetilde{\mathcal{O}}\left(\frac{\Delta M}{\epsilon}\right)$$

queries to $U_{\widetilde{\rho}}$ and $U_{\widetilde{\rho}}^\dagger$, as well as to an oracle that implements controlled-$e^{-i\theta O_l}$ for each decoding operator of the form $O_l = \frac{H_l \otimes X_C}{2}$.

Proof. Using the definition the block-encoded Hamiltonian $\widetilde{H}$ from Definition 20, a short calculation shows that

$$\mathrm{Tr}(\widetilde{\rho} O_l) = \frac{\mathrm{Re}[\widetilde{c}_l]}{\Delta\gamma^2} \tag{94}$$

Note that while the coefficients c_m of the Hamiltonian H are real, the coefficients $\widetilde{c}_m$ of the approximation in the block-encoding $\widetilde{H}$ may be complex, though we only learn their real component here. Fortunately, this is of no consequence; in Section 4.3 we showed that $|\widetilde{c}_l - c_l| \leq \frac{\epsilon_b}{t}$, where t is the evolution time and ϵ_b is the block-encoding error, and writing $\widetilde{c}_l$ as the sum of its real and imaginary parts it is easy to see that $|\mathrm{Re}[\widetilde{c}_l] - c_l| \leq |\widetilde{c}_l - c_l|$, which means that the upper bound still holds. We can now continue in similar fashion to Lemma 29, using the quantum mean estimation protocol introduced by Huggins et al. [19] to estimate the vector

$$\mathbf{d} := [\mathrm{Tr}(\widetilde{\rho} O_0), ..., \mathrm{Tr}(\widetilde{\rho} O_{M-1})], \tag{95}$$

from which we approximate the Hamiltonian coefficients by

$$\widetilde{c}_l := \Delta\gamma^2 d_l, \tag{96}$$

where, again, we ignore any possible imaginary component. By the same argument as in Lemma 29, we require

$$\widetilde{\mathcal{O}}(\Delta\gamma^2 M/\epsilon) \tag{97}$$

queries to $U_{\widetilde{\rho}}$ and $U_{\widetilde{\rho}}^\dagger$ to predict the vector of coefficients $\widetilde{\mathbf{c}} = [\widetilde{c}_0, ..., \widetilde{c}_{M-1}]$ to within l_2 error ϵ.

To complete the proof, we refer to Appendix B.5, where we showed that $\gamma \in [1, \sqrt{2}] \in \mathcal{O}(1)$, and therefore we state the query complexity as

$$\widetilde{\mathcal{O}}(\Delta M/\epsilon) \tag{98}$$

□

We now have the sufficient resources to prove Theorem 33.

Theorem 33 (Solving the Hamiltonian Learning Problem via Quantum Mean Estimation)**.** *Let $U = e^{-iHt}$ be the unitary operator that performs time evolution under the Hamiltonian H for time t. The number of queries to U and $U^\dagger$ needed to solve the Unitary Hamiltonian Learning Problem for the Hamiltonian H is at most*

$$N \in \widetilde{\mathcal{O}}\left(\frac{M}{\epsilon t}\right)$$

Proof. The total number of queries to the time evolution operator and its inverse can be calculated by looking at the procedure as 4 nested queries.

By Lemma 32, the quantum mean estimation process makes

$$\widetilde{\mathcal{O}}(\Delta M/\epsilon') \tag{99}$$

queries to the unitary $U_{\widetilde{\rho}}$ that prepares the state $\widetilde{\rho}$ (equation (90)), where $\Delta = \frac{\pi}{2t}$. $U_{\widetilde{\rho}}$ makes use of the fixed-point amplitude amplification circuit, which makes

$$\mathcal{O}\left(\log(1/D)\right) \tag{100}$$

queries the circuit A that prepares state $|s\rangle$ (equation (86)), as well as to $A^\dagger$, where D is the trace distance between the target state and the approximation of the target state that results from the FPAA process. Finally, the circuit A makes a single query to the block-encoding of the Hamiltonian, which is prepared using

$$\mathcal{O}\left(\log(1/\epsilon_b)\right) \tag{101}$$

queries to the time evolution operator $U = e^{-iHt}$ and its inverse, where ϵ_b is the block-encoding error. Altogether, this results in a query complexity of

$$\mathcal{O}\left(\frac{M\log(1/D)\log(1/\epsilon_b)}{\epsilon' t}\right). \tag{102}$$

Since each step of the process involves approximation, we must choose sufficiently small values for the learning error ϵ', the block-encoding error ϵ_b, and the FPAA error $\epsilon_{AA} := D$ such that the total l_2 error in the estimate of the Hamiltonian coefficients is at most ϵ, as required by the Hamiltonian learning problem.

Recall from Definition 20 that $\mathbf{c}$ is the vector of Hamiltonian coefficients we wish to learn, $\widetilde{\mathbf{c}}$ is the vector of coefficients for the block-encoded Hamiltonian, and $\hat{\mathbf{c}}$ is the vector of coefficients that is estimated by the learning procedure. The total error in learning the coefficients is then

$$\begin{aligned}\|\hat{\mathbf{c}} - \mathbf{c}\|_2 &= \|\hat{\mathbf{c}} - \widetilde{\mathbf{c}} + \widetilde{\mathbf{c}} - \mathbf{c}\|_2 \\ &\leq \|\hat{\mathbf{c}} - \widetilde{\mathbf{c}}\|_2 + \|\widetilde{\mathbf{c}} - \mathbf{c}\|_2 \\ &\leq \|\hat{\mathbf{c}} - \widetilde{\mathbf{c}}\|_2 + \frac{M\epsilon_b}{t},\end{aligned} \tag{103}$$

where we have used the result of Proposition 24 in the last line. Conceptually, what we have done is separate the total error into the sum of the error in learning the block-encoded coefficients and the error in the block-encoded coefficients themselves. Next, we separate $\|\hat{\mathbf{c}} - \widetilde{\mathbf{c}}\|_2$ into the error that comes from the quantum mean estimation, and the error resulting from using the state χ to approximate the QME resource state $\widetilde{\rho}$:

$$\|\hat{\mathbf{c}} - \widetilde{\mathbf{c}}\|_2 \leq \|\hat{\mathbf{c}} - \mathbf{c}'\|_2 + \|\mathbf{c}' - \widetilde{\mathbf{c}}\|_2 , \tag{104}$$

where $\mathbf{c}'$ is the vector of Hamiltonian coefficients that would be recovered if the the state $\widetilde{\rho}$ was used in the QME process. We upper bound $\|\hat{\mathbf{c}} - \mathbf{c}'\|_2$ by considering the error in estimating a single coefficient:

$$\begin{aligned}
|\hat{c}_l - c'_l| &= \Delta\gamma^2 \left|\mathrm{Tr}(\chi O_l) - \mathrm{Tr}(\widetilde{\rho} O_l)\right| \\
&= \Delta\gamma^2 \left|\mathrm{Tr}((\chi - \widetilde{\rho}) O_l)\right| \\
&\leq \Delta\gamma^2 \sum_{i=1}^{k} \sigma_i(\chi - \widetilde{\rho}) \sigma_i(O_l) \\
&\leq \Delta\gamma^2 \sigma_{max}(O_l) \sum_{i=1}^{k} \sigma_i(\chi - \widetilde{\rho}) \\
&= \Delta\gamma^2 \|O_l\|_2 \sum_{i=1}^{k} |\lambda_i(\chi - \widetilde{\rho})| \\
&= \frac{\Delta\gamma^2}{2} \mathrm{Tr}\left(\sqrt{(\chi - \widetilde{\rho})^\dagger (\chi - \widetilde{\rho})}\right) \\
&= \Delta\gamma^2 T(\chi, \widetilde{\rho})
\end{aligned} \tag{105}$$

where $T(\chi, \widetilde{\rho})$ is the trace distance between χ and $\widetilde{\rho}$. We have made use of Von Neumann's trace inequality in the third line, where $\sigma_i(\cdot)$ represents the i^{th} singular value of the corresponding operator. In the fifth line we used the fact that $\|O_l\|_2 = 1/2$ (O_l is simply a Pauli operator multiplied by a factor of $1/2$), as well as expressed the singular values of $\chi - \widetilde{\rho}$ as the absolute values of its eigenvalues (this is valid since $\chi - \widetilde{\rho}$ is Hermitian, and therefore normal).

By the Cauchy-Schwarz inequality, we then have

$$\begin{aligned}
\|\hat{\mathbf{c}} - \mathbf{c}'\|_2 &\leq \sqrt{M} \max_l |\hat{c}_l - c'_l| \\
&\leq \sqrt{M} \Delta\gamma^2 T(\chi, \widetilde{\rho})
\end{aligned} \tag{106}$$

Recall that χ is the state we get after tracing out system A for the state $|T_{approx}\rangle$ that results from the FPAA process, and similarly $\widetilde{\rho}$ is the state resulting from tracing out system A for the target state $|T\rangle$ of the FPAA. Therefore, the trace distance between χ and $\widetilde{\rho}$ is less than or equal to the trace distance between $|T_{approx}\rangle$ and $|T\rangle$ [31].

As per Lemma 31, performing $\mathcal{O}(\log(1/D))$ iterations of FPAA guarantees that the trace distance between $|T_{approx}\rangle$ and $|T\rangle$ is at most D. Therefore, we arrive at the following expression for the total error in predicting the Hamiltonian coefficients:

$$\begin{aligned}
\|\hat{\mathbf{c}} - \mathbf{c}\|_2 &\leq \sqrt{M} \Delta\gamma^2 D + \epsilon' + \frac{M\epsilon_b}{t} \\
&= \frac{\pi\gamma^2\sqrt{M}D}{2t} + \epsilon' + \frac{M\epsilon_b}{t}.
\end{aligned} \tag{107}$$

Assigning an error budget of $\epsilon/3$ to each of the three terms (i.e. choosing $D := \frac{2\epsilon t}{3\pi\gamma^2\sqrt{M}}$, $\epsilon' := \epsilon/3$, and $\epsilon_b := \frac{\epsilon t}{3M}$) ensures the total error is at most ϵ, and gives a final query complexity of

$$N \in \mathcal{O}\left(\frac{M \log^2\left(\frac{M}{\epsilon t}\right)}{\epsilon t}\right). \tag{108}$$

Using $\widetilde{\mathcal{O}}$ which suppresses subdominant logarithmic terms, we get

$$N \in \widetilde{\mathcal{O}}\left(\frac{M}{\epsilon t}\right), \tag{109}$$

proving the result in Theorem 33.

Note that the failure probability does not lead to an increase in the number of queries (in $\widetilde{O}$ notation) because unlike in the classical shadows approach, we replace the measurement in the preparation circuit with fixed-point amplitude amplification so that instead of getting our desired resource state with some probability, we are guaranteed to have an approximation of the resource state. This means the only failure probability comes from the quantum mean estimation process, which guarantees the failure probability is smaller than 1/3 [19]. By Chernoff arguments this probability can be made less than ϵ using a logarithmic number of extra queries, thus the $\widetilde{O}$ query complexity is unaffected.

□

6 Applications

Since in general a Hamiltonian can have exponentially many terms, one of the focuses of research on Hamiltonian learning has been to find suitable classes of Hamiltonians for which learning is more tractable. One way to approach this problem is to limit the scope to Hamiltonians with some known structure. Along this vein, several recent approaches to Hamiltonian learning have considered low-intersection Hamiltonians whose Hamiltonian terms have limited overlap in their support [11, 13, 14]. This class of Hamiltonians contains all spatially local Hamiltonians, and as such contains many physically relevant Hamiltonians.

While the low-intersection class of Hamiltonians is undoubtedly rich, it does not capture all physically relevant Hamiltonians. Our approach is well suited to more general Hamiltonians, such as those that are k-local but not low-intersection (for example, if you have pairwise interactions between all pairs of qubits). This is the first regime where our methods outperform previous approaches, under the assumption that we can query $U = e^{-iHt}$ as well as $U^\dagger$. However, the learning algorithms we develop are suitable even for Hamiltonians beyond the k-local class. Furthermore, in addition to physically notable Hamiltonians like the examples given in this section, several quantum algorithms use Hamiltonian simulation as a subroutine. Notable examples include algorithms for solving the quantum linear systems problem [32, 33], phase estimation [34], semi-definite programming [35], linear PDEs [36], and recent work on simulating systems of coupled oscillators [37]. Therefore, it is of interest to develop learning techniques for different and more general classes of Hamiltonians, and not just the low-intersection class.

In this section we present a few examples of Hamiltonians that are k-local, but not low-intersection. We also include a discussion about the settings in which previous Hamiltonian techniques work

well, and in which settings our approach outperforms them - in several cases reducing the query complexity polynomially, or even exponentially.

6.1 Fermionic and Hard-core Boson Models

A particularly relevant application of Hamiltonian learning is the learning problem for systems of particles. These models are naturally represented in terms of creation and annihilation operators. The nature of these operators vary, however the common thread between them is that the number operator (which measures the number of particles located in a particular mode) can be expressed as

$$n = a^\dagger a \tag{110}$$

where a is an annihilation operator and $a^\dagger$ is a creation operator. The form of these operators depends on the symmetries of the underlying particles; however, popular examples include bosonic case where $(a^\dagger)^p |0\rangle = \sqrt{p!} |p\rangle$ where $|p\rangle$ is a state that has p particles in it and the fermionic or hard-core bosonic cases where $(a^\dagger)^p |0\rangle = \delta_{p\leq 1} |p\rangle$. The bosonic cases have commuting operators; whereas the fermionic case uses anti-commuting creation and annihilation operators.

As a particular example, let us consider the following model which describes the hopping of particles on a graph $G = (V, E)$ with interactions between nearest neighbors in the graph. For simplicity, let us assume that the creation and annihilation operators are either fermionic or hard-core bosonic. Under these assumptions, the Hamiltonian can be expressed as

$$H = \sum_{j\in V} c_j a_j^\dagger a_j + \sum_{(i,j)\in E} d_{i,j}(a_i^\dagger a_j + a_j^\dagger a_i) + e_{i,j} a_i^\dagger a_i a_j^\dagger a_j \tag{111}$$

These Hamiltonian terms can be simplified by realizing that $\hat{n} = a^\dagger a = (I + Z)/2$

$$H = \sum_{j\in V} c'_j Z_j + \sum_{(i,j)\in E} d_{i,j}(a_i^\dagger a_j + a_j^\dagger a_i) + e_{i,j} Z_i Z_j / 4 \tag{112}$$

In this case the Hamiltonian satisfies the above assumptions and further a specific qubit representation of the creation and annihilation operators are not needed for this case provided $\mathrm{Tr}(a_i^\dagger a_j a_k^\dagger a_\ell) = 0$ for distinct i, j and k, ℓ unless $i = \ell$ and $j = k$. To see this, note that a_i is an off-diagonal matrix as $a_i |1\rangle = |0\rangle$. Thus the only way we can build a diagonal operator is by pairing a_i with $a_i^\dagger$. This can only occur, under the assumption of distinctness if $i = \ell$ and similarly $j = k$. Thus this conforms the the Hamiltonian Learning Problem from Definition 1 while not needing a specific Pauli decomposition of the creation and annihilation operators. In contrast, many existing approaches to Hamiltonian learning require a low-intersection Hamiltonian and such a Hamiltonian need not be low intersection. Specifically, if one considers the case of a fermionic Hamiltonian then the Jordan-Wigner representation does not yield a low-intersection Hamiltonian except for the case of one-dimensional graph G. Thus our method differs from many approaches because such Hamiltonians can be learned directly rather than going first through a representation of for their creation and annihilation operators.

6.2 Spin Glass Models

Other models that lend themselves well to our approach to Hamiltonian learning are spin glasses. One particular example on the complete graph is the Sherrington-Kirkpatrick (SK) model [38], an

Ising-like model in which any pair of spins on a lattice can be coupled, regardless of the distance between them. The SK model was introduced to study the magnetic properties of spin glasses, and is also of interest in combinatorial optimization and the study of neural networks [39].

In the presence of a transverse field, the Sherrington-Kirkpatrick Hamiltonian takes the form

$$H_{SK} = \frac{1}{\sqrt{n}} \sum_{1 \leq i < j \leq n} J_{ij} Z_i Z_j + g \sum_{i=1}^{n} X_i, \tag{113}$$

where the J_{ij} are i.i.d. random variables sampled from the standard Gaussian distribution, and $g > 0$ is the strength of the transverse field.

Since couplings are between pairs of spins, the Hamiltonian is 2-local and thus contains $M \in \mathcal{O}(n^2)$ terms. However, the Hamiltonian is very much not low-intersection, as each term has overlapping support with $\mathcal{O}(M)$ other terms. Note that while Heisenberg models on the complete graph contain only pairwise interactions, our method applies to more general spin glass models which allow for interactions between more than two spins. These models, often referred to as p-spin models, are p-local Hamiltonians with infinite range p-body interactions and thus the query complexity of the learning process increases, as there are $M \in \mathcal{O}(n^p)$ Hamiltonian coefficients to learn. Additionally, while these spin glass models only contain Pauli Z interactions, our methods also apply to models with more general interactions such as Heisenberg models on dense graphs.

6.3 Prior Work

In this section we consider several recent works on Hamiltonian learning. Note that the varying cost and error metrics used by different authors makes comparing approaches a difficult endeavour, so our goal here is mainly to highlight the scenarios where each approach may be particularly useful and mention how our approach fares in each case. Regarding error metrics, we would like to note that the infinity-norm, where one would like to simultaneously estimate every individual Hamiltonian coefficient to within some error ($\max_i |\hat{c}_i - c_i| \leq \epsilon$), is popular as a metric. However, we feel that the 2-norm ($\|\hat{\mathbf{c}} - \mathbf{c}\|_2 \leq \epsilon$) better highlights the cost of learning, as the infinity-error obscures some of the dependence on the number of Hamiltonian terms. Thus, unless otherwise stated, references to query complexity will be for learning the Hamiltonian coefficients to error ϵ in the 2-norm. While all the approaches we will discuss use the same basic learning resource (black-box access to $U = e^{iHt}$), we separate them into two classes: those that require quantum control, and those that do not. The distinguishing factor between these two cases is that experiments without quantum control use only a single application of U.

First, we consider approaches that have the best scaling with respect to the error ϵ in learning the Hamiltonian coefficients. Namely, the results by *Huang et al.* [14] and *Dutkiewicz et al.* [17], which are able to achieve Heisenberg-limited scaling ($\mathcal{O}(\epsilon^{-1})$) in the evolution time. As shown in [17], this is possible only through the use of quantum control, which makes their approach well suited for learning low-intersection Hamiltonians when accuracy is important. Our QME-based approach also makes use of quantum control to achieve Heisenberg-limited scaling, and has the added advantage that we are able to learn more general Hamiltonians, such as k-local Hamiltonians. Our shadows-based approach does not achieve this error scaling, but is robust to unexpected Hamiltonian terms, as we will discuss in the following section. A direct comparison between our approach and the one of [14] is difficult because of differences in error metrics, complexity metrics, and input models.

Importantly, the input models differ in that we require backwards time evolution, but only require one controlled application of U and $U^\dagger$ for a short time. Meanwhile, in [14] they require the ability to rapidly interleave many short-time evolutions. One case where our approach excels is when the Hamiltonian contains high-weight Pauli terms (i.e. terms that act non-trivially on k qubits, for large k). Here, the techniques of [14] require exponential (in k) queries and total evolution time. This is particularly bad if $k \in \mathcal{O}(n)$. On the other hand, our approach is able to handle this case with $\mathbf{poly}(n)$ queries to U and $U^\dagger$.

There are also several approaches that do not beat the standard quantum limit([11–13, 15, 16]) but have other notable advantages - perhaps the most significant being that aside from [12] they do not require quantum control. Several of these approaches also aim to look beyond low-intersection Hamiltonians. The results of *Franca et al.* [16] scale particularly well with then number of Hamiltonian terms, requiring $\widetilde{\mathcal{O}}\left(\frac{M}{\epsilon^2}\right)$ queries to U to learn a Hamiltonian with M terms. However, their constant factor is exponentially large in k, and so their approach is unfavourable if the Hamiltonian contains high-weight Pauli terms. Conversely, the approach of *Gu et al.* [13] does not scale as favourably for Hamiltonians with many terms (requiring $\mathcal{O}(M^5)$ queries), but is able to handle high-weight Pauli terms without incurring an exponential (in k) overhead. These approaches are very well-suited for leaning low-intersection Hamiltonians, and even some (but not all) more general types of Hamiltonians, but are not ideal when there are high-weight terms or the number of terms is large.

The techniques of *Yu et al.* [12] and *Caro* [15] seek to remedy this and consider even broader classes of Hamiltonians. The approach of [15], allows for any Hamiltonian to be learned (in the 2-norm) using $\mathbf{poly}(n)$ queries to U, so long as the Hamiltonian contains at most $\mathbf{poly}(n)$ non-zero terms, has some known structure (for example, k-locality), and $\|H\| \in \mathbf{poly}(n)$. In [12], the authors dropped the need for an underlying structure to be known a priori, and instead only require that the Hamiltonians are sparse. The main drawback is that the resource complexity of these learning procedures does not scale well with the error in learning the coefficients, as both require $\mathcal{O}(1/\epsilon^4)$ queries to U. In addition, the effects of the evolution time on the query complexity of [12] are not clear.

The result due to *Caro* [15] is particularly relevant to our work. Firstly, the procedure makes use of Choi states of the time evolution unitary, whereas we use pseudo-Choi states of the Hamiltonian itself. Our improved error scaling appears to indicate that pseudo-Choi states are more powerful in some ways than regular Choi states, and we believe this comes from the use of quantum control in preparing pseudo-Choi states. As mentioned above, [17] demonstrated the power of quantum control in achieving better error scaling, so it is perhaps not unexpected that our error scaling is not matched by an approach that uses only time-evolved states. Interestingly, both approaches also share similar bounds on evolution times. In addition to this, our results and those of [15] have a key advantage that to the best of our knowledge no other approach shares, which is that our shadows-based approach, along with that of [15], can be used to learn *any* Hamiltonian query efficiently if we consider the error in the infinity norm. As mentioned earlier, this error metric perhaps blurs the dependence on the number of Hamiltonian terms, but in situations where one really only cares about this metric, these two approaches can be used to learn Hamiltonians with even an exponential number of terms using a polynomial number of queries as long as the norm of the Hamiltonian is polynomially bounded in the number of qubits (though our algorithm remains more efficient). Of course, the computational complexity would be exponential as we would need to output estimates for an exponential number of coefficients, but the coefficients could be computed

in parallel on classical computers, which could drastically reduce the wall time needed.

7 Robustness

A last point that we wish to raise about our protocol involves the question of robustness of the learning protocol. Specifically, when learning the Hamiltonian there is always a possibility that the Hamiltonian that we wish to learn contains terms that are absent in the model we wish to describe it with. For example, we could assume that the Hamiltonian is of the form $H = x_1P_1 + x_2P_2$ but the actual Hamiltonian used to generate the pseudo-Choi state is of the form $H_{\text{true}} = x_1P_1 + x_2P_2 + x_3P_3$ for orthogonal Pauli operators P_1, P_2, P_3. In this case, it is clear that our protocols will never be able to learn the true Hamiltonian model for the pseudo-Choi state without including the P_3 term in our model. Nonetheless, we could ask ourselves if the protocol is robust to small errors in the Hamiltonian and if such errors are evident in the reconstruction. Interestingly, we will see that our protocols are both robust and also error evident meaning that we can tell when terms are present in the true Hamiltonian that are not reflected in the model.

Definition 34 (Under Specified Hamiltonian Learning Problem)**.** *Let $H = \sum_{j=1}^{M} c_j H_j + \chi E$ where E is an unknown Hermitian operator such that $\max_j |\text{Tr}(EH_j)| = 0$ and $\frac{1}{d}\text{Tr}(E^2) = 1$ with χ an unknown multiplicative constant. Our aim is to learn $\hat{\mathbf{c}}$ such that $\|\mathbf{c} - \hat{\mathbf{c}}\|_2 \leq \epsilon$ and an estimate $\hat{\chi}$ such that $|\chi - \hat{\chi}| \leq \epsilon_\chi$ with probability greater than $1 - \delta$*

Corollary 35 considers the error in learning the coefficient χ of the extra terms E. In particular, it shows that if we allow the error in learning χ to be larger than some lower bound, the Under Specified HLP can be solved using a number of queries on the same order as for the Unitary HLP. Thus, it is possible to check if any extra terms are present in the Hamiltonian by estimating the value of χ. Note that although the value of ϵ remains the same as for the Unitary HLP, the evolution time t must be small enough to normalize the Hamiltonian with the extra terms in order for the protocol to function.

Corollary 35. *There exists an algorithm that can solve the under specified Hamiltonian learning problem using $\mathcal{O}\left(\frac{M\log(M/\delta)}{t^2\epsilon^2}\log\left(\frac{M}{t\epsilon}\right)\right)$ controlled applications of e^{-iHt} and its inverse under the assumptions of Theorem 26 if $\epsilon_\chi = \Omega(\frac{\epsilon\|\widetilde{\mathbf{c}}\|_1}{|\chi|})$ and $\epsilon_s\gamma^2 \leq 1$.*

Proof. The underspecified Hamiltonian learning problem can be solved in exactly the same manner as the fully specified learning problem. Specifically, let us assume that we apply the algorithm $\texttt{FindCoeffUnitary}(U, U^\dagger, \Delta, \boldsymbol{H}, n, N_s)$ where U here is e^{-iHt} and Δ is chosen to be $\pi/2t$ for $t \in \left(0, \frac{1}{2\|H\|}\right)$. Note that here we need to assume that this constant is chosen to be sufficiently large so that the erroneous Hamiltonian is also normalized given that $\chi \neq 0$.

Let us assume for the moment that we knew the identity of E and performed the full learning protocol on the state. The only differences between the existing protocol and this one are that $M+1$ terms are present, and t must be small enough to normalize the true Hamiltonian. It follows that the total number of queries to the unitary dynamics needed to learn the Hamiltonian scales from Theorem 26 as

$$N \in \mathcal{O}\left(\frac{M\log(M/\delta)}{t^2\epsilon^2}\log\left(\frac{M}{t\epsilon}\right)\right) \tag{114}$$

where the constants here are taken to be the true constants of the Hamiltonian with E present.

Next assume that we construct the same Clifford shadow of the state as we would for solving the Unitary HLP. We then wish to reconstruct only $\mathbf{c}$ from the data. We can achieve this by following the procedure `FindCoeffClifford` after noting that the values of the coefficients of $\mathbf{c}$ are computed independently of the coefficient of E. Thus we can reconstruct $\mathbf{c}$ without reconstructing E simply by skipping the E reconstruction in the state (which wouldn't be possible anyways since the identity of E is not known). Then from Theorem 26, we are guaranteed that the error in the reconstructed $\hat{\mathbf{c}}$ is at most ϵ in this case and we learn $1/\gamma^2$ within precision ϵ_s. Thus we can learn the vector $\mathbf{c}$ within the required error from the exact same procedure discussed above.

The coefficient of E can the be inferred from the result given that

$$\gamma^2 = \frac{\|\tilde{c}\|_2^2 + \chi^2}{\Delta^2} + 1. \tag{115}$$

Thus the value of χ^2 can be inferred from the residual value of the normalization after the vector $\hat{\mathbf{c}}$ is subtracted from it

$$\hat{\chi}^2 = (\hat{\gamma}^2 - 1)\Delta^2 - \|\tilde{c}\|_2^2 \tag{116}$$

given that we compute the value of γ^{-2} within error ϵ_s this implies that

$$|(\gamma^{-2} \pm \epsilon_s)^{-1} - \gamma^2| \leq \frac{\gamma^2}{1 - \epsilon_s\gamma^2} - \gamma^2 = \epsilon_s\gamma^4 + O(\epsilon_s^2\gamma^6). \tag{117}$$

Thus if $\epsilon_s\gamma^2 \leq 1$ then the remainder is $O(\epsilon_s\gamma^4)$. Recall that $\gamma^2 \in \mathcal{O}(1)$, so this assumption is not restrictive. Under the assumption that the magnitudes of the Hamiltonian coefficients are bounded above by 1, we find by substitution of (54) that the error in our estimate of χ^2 is at most

$$\begin{aligned} \epsilon_{\chi^2} &= \sqrt{\left(\frac{\Delta^4\gamma^8\epsilon^2}{M(1+\Delta^2)}\right) + 2\,\|\widetilde{\mathbf{c}}\|_2^2\,\epsilon^2} \\ &\in \mathcal{O}\left(\epsilon\sqrt{\left(\frac{\Delta^2}{M}\right) + \|\widetilde{\mathbf{c}}\|_2^2}\right) \end{aligned} \tag{118}$$

where we use the fact that $\gamma \in \mathcal{O}(1)$ and assume $\Delta \gg 1$ (which is reasonable, as Δ needs to normalize the Hamiltonian).

In order to understand how the error in our estimate of χ scales, we need to consider how an estimate of χ is found from χ^2. The simplest way of performing this is to simply take the square root of our estimate of χ^2, which under the assumption that $\left|\epsilon_{\chi^2}\right| \leq \chi^2/2$ leads us to the conclusion that

$$\epsilon_\chi \in O\left(\frac{\epsilon_{\chi^2}}{|\chi|}\right). \tag{119}$$

This results in

$$\epsilon_\chi \in \mathcal{O}\left(\frac{\epsilon\sqrt{\left(\frac{\Delta^2}{M}\right) + \|\widetilde{\mathbf{c}}\|_2^2}}{|\chi|}\right) \tag{120}$$

Recall that $\Delta \in \mathcal{O}(\|H\|_2)$, and so $\Delta \in \mathcal{O}(\|\widetilde{\mathbf{c}}\|_1)$ if we assume the extra terms are smaller in magnitude than the terms that were expected in our model (that is, $\|\widetilde{\mathbf{e}}\|_1 \leq \|\widetilde{\mathbf{c}}\|_1$). Therefore, we have

$$\epsilon_\chi \in \mathcal{O}\left(\frac{\epsilon \|\widetilde{\mathbf{c}}\|_1}{|\chi|}\right). \tag{121}$$

Note that in the calculations above, we have assumed there was no error in the block-encoding of the additional Hamiltonian terms. In reality, the steps above correspond to the error in learning $\widetilde{\chi}$, the magnitude of the extra Hamiltonian terms in the block-encoding. The difference between χ and $\widetilde{\chi}$ can be upper bounded using Proposition 24, which leads to

$$\begin{aligned} |\chi - \hat{\chi}| &\leq |\chi - \widetilde{\chi}| + |\widetilde{\chi} - \hat{\chi}| \\ &\leq \frac{\epsilon_b}{t} + \epsilon_\chi \end{aligned} \tag{122}$$

Recalling that in Theorem 26 the value of ϵ_b was specifically chosen such that $\frac{\sqrt{M}\epsilon_b}{t}$ was at most $\epsilon/2$, we see that

$$|\chi - \hat{\chi}| \in \mathcal{O}\left(\frac{\epsilon}{2\sqrt{M}} + \frac{\epsilon \|\widetilde{\mathbf{c}}\|_1}{|\chi|}\right), \tag{123}$$

and assuming the magnitude of the extra terms is smaller, or at least not much larger, than the terms that were expected (enough so that $|\chi|$ is smaller than something on the order of $\sqrt{M}\|\widetilde{\mathbf{c}}\|_1$), we have

$$|\chi - \hat{\chi}| \in \mathcal{O}\left(\frac{\epsilon \|\widetilde{\mathbf{c}}\|_1}{|\chi|}\right). \tag{124}$$

Thus if we insist that the error tolerance in the estimation of χ is in $\Omega\left(\frac{\epsilon\|\widetilde{\mathbf{c}}\|_1}{|\chi|}\right)$ then we see that our accuracy criteria is automatically satisfied, and we can solve the under specified Hamiltonian learning problem. □

This shows that the protocol is robust to errors in the underlying Hamiltonian. Specifically, it shows that if there are unknown terms in the Hamiltonian then shadow tomography will learn the correct coefficients of the known terms but also the coefficient of the (normalized) sum of the unknown terms will also be visible from the results. This allows us to identify the presence of such terms, but our protocol provides no direct guidance other than the fact that the terms in question must be orthogonal to the known Hamiltonian terms.

In theory, since all additional terms can be represented as a sum of Pauli operators, one could systematically check the coefficients for the (k+1)-local terms, the (k+2)-local terms, etc. until all the non-zero ones are found (that is, when the magnitude of the additional coefficients is consistent with the magnitude of χ). Furthermore, this could all be done without performing any additional measurements, as this is all performed on a classical computer using the classical shadow. The downside is that if one has no knowledge about the locality (in the Pauli basis) of the extra terms, then in the worst case this would require $\mathcal{O}(4^n)$ time to check all possible Hamiltonian terms (though with a large enough classical memory, theoretically all these checks could be done in parallel), and

with a polynomial number of queries, the resulting exponentially large vector of coefficients would only be epsilon accurate in the infinity norm. However, if one has some intuition about the potential identities of the extra terms, it is straightforward to check if they exist and learn their coefficients. In these cases, Corollary 35 gives the following upper bound on the error in learning all the coefficients, including the ones that were not initially considered:

$$\|\mathbf{c} + \mathbf{e} - (\hat{\mathbf{c}} + \hat{\mathbf{e}})\|_2 \leq \epsilon + \epsilon_\chi, \tag{125}$$

where ϵ is the l_2 error we wanted to achieve in to satisfy the original Hamiltonian learning problem, $\mathbf{e}$ is the vector of coefficients of the extra Hamiltonian terms, and $\hat{\mathbf{e}}$ is our estimate of it. Again, note that we still require t to be small enough to account for the extra terms, so it may be useful to choose a value of t that is slightly smaller than needed for a k-local Hamiltonian, in case there are additional Hamiltonian terms.

The remaining question that we wish to discuss is how accurate the preparation of our pseudo-Choi state needs to be in order to ensure that our estimate of the classical shadow is correct. Let us assume for this discussion that for parameter ω and state operator $\rho^\perp$

$$\tilde{\rho}_c := (1-\omega)|\psi_c'\rangle\langle\psi_c'| + \omega\rho^\perp. \tag{126}$$

This setting can cover not only cases where an inexact pseudo-Choi state is prepared but also the case where an unknown number of qubits are acted on by the Hamiltonian. Although at first glance an explicit analysis similar to the previous corollary would seem appropriate for this case, conceptual hurdles arise in the way in which we define the pseudo-Choi state in this case if we do not know the underlying qubits. The more natural setting is one in which the Hamiltonian terms are coupled to an unknown reservoir for the system which naturally leads to open system dynamics and in turn to a density matrix similar to that described above after a partial trace over the environmental degrees of freedom. For this reason, we focus on this model for the system.

The shadow tomography reconstruction will then use such states to prepare the distributions $\tilde{\tilde{\rho}}_i$ based on the measurement statistics that we glean from measuring the state in the computational basis. Specifically we have from (48) and (49) that the error in reconstructing each of the Hamiltonian coefficients, under the assumption that the spectral norm of each operator obeys $\|O_l\| \leq 1$, is

$$|\mathrm{Tr}(\tilde{\rho}_c O_l) - \mathrm{Tr}(\rho_c{}' O_l)| \leq \|\tilde{\rho}_c - \rho_c'\|_{\mathrm{Tr}} \leq 2\omega \tag{127}$$

Similarly, our estimate of $1/\gamma^2$ will err by at most 2ω. Note that in addition to this error, which is caused by our inaccurate preparation of the pseudo-Choi state, we also have an additional error, ϵ_s, due to the use of classical shadows in estimating the expectation values above.

Let c_l be the true coefficient of the l^{th} Hamiltonian term, and let c_l' be the estimate we would get if the pseudo-Choi state was prepared accurately (that is, $\omega = 0$). Finally, let $\widetilde{c}_l$ be the estimate we get from an inaccurate preparation of the pseudo-Choi state. The error in estimating this coefficient is then

$$\begin{aligned} |c_l - \widetilde{c}_l| &\leq |c_l - c_l'| + |c_l' - \widetilde{c}_l| \\ &\leq 2\gamma^2\Delta(\epsilon_s + 2\omega), \end{aligned} \tag{128}$$

where the factor of $2\gamma^2\Delta$ on each term results from having to multiply the expectation values by $\gamma^2\Delta$, and we have used the fact that the magnitudes of the Hamiltonian coefficients are upper bounded by 1, as well as the assumption that $\Delta \geq 2$ (which is reasonable since $\Delta \in \mathcal{O}(\|H\|_2)$).

The error due to the inaccurate preparation of the pseudo-Choi state will not dominate the error from shadow tomography if this error is subdominant to ϵ_s, which from (54) will occur if

$$\omega \in O\left(\frac{\epsilon_c}{\sqrt{M}\gamma^2\sqrt{\tilde{c}_{\max}^2 + \Delta^2}}\right). \tag{129}$$

From this we see that the result is robust in the sense that the value of ω required shrinks at most polynomially with the number of terms in the Hamiltonian. However, as the number of terms increases polynomially with the number of qubits for many systems of interest, this means that in the worst case scenario a very accurate preparation of the pseudo-Choi state will be needed in order to achieve the requisite accuracy for the entire Hamiltonian.

8 Conclusion and Outlook

In this work we have provided a new approach to Hamiltonian learning that draws inspiration from the state to channel isomorphism to yield a state from which the Hamiltonian can be directly learned and can be prepared using a simple quantum circuit. We are able to show that such approaches can easily learn k-local Hamiltonians, which is a broader class of Hamiltonians than the low-intersection case considered in most other Hamiltonian learning works, and in some cases can learn even much more general Hamiltonians without requiring exponentially many queries. We further consider the learning problem in a scenario where amplitude estimation can be used to further accelerate it using mean-value estimation algorithms and achieve the optimal $1/\epsilon$ scaling with the error tolerance.

There are a number of interesting problems revealed by this work. While our approach extends the scope of recent work on Hamiltonian learning ([11, 13, 14]) to include more general Hamiltonians, a caveat is that we require backwards time evolution to generate the pseudo-Choi state. One area of interest for future work would be to find methods of generating pseudo-Choi-like states efficiently without the need for backwards time evolution, making our techniques very appealing for Hamiltonian learning under a much wider range of scenarios. Additionally, recent work by *Huang et al.* [14] introduced an interesting method for reshaping Hamiltonians into non-interacting patches, allowing different parts of the Hamiltonian to be learned in parallel. If these methods could be applied to our approach, it may be possible to create pseudo-Choi states for separate parts of the Hamiltonian instead of the entire Hamiltonian, which could allow for longer evolution times, potentially reducing the total number of queries.

More broadly, however, this form of a block-encoding of the Hamiltonian within a quantum state allows much more flexible learning than existing approaches permit. While preparing pseudo-Choi states requires a trusted quantum computer, making them less appealing for near-term experimental settings, they open up the potential of tackling many different learning problems by similarly encoding information about the system within states. Such problems include Lindbladian learning for open systems, time-dependent Hamiltonian learning, and even applications of Hamiltonian learning methods to learn the structure of quantum walks. It is our belief that these results represent a step towards a broader understanding of learning dynamical models for quantum systems and in turn a characterization of the physical models that can or cannot be efficiently learned from experiment.

Acknowledgements

JC acknowledges funding from NSERC and CIFAR. NW would like to acknowledge funding for this work from Google Inc. This material is based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Co-design Center for Quantum Advantage (C2QA) under contract number DE-SC0012704 (PNNL FWP 76274).

9 References

References

[1] Jens Eisert et al. "Quantum certification and benchmarking". In: *Nature Reviews Physics* 2.7 (June 2020), pp. 382–390. DOI: 10.1038/s42254-020-0186-4. URL: https://doi.org/10.1038%2Fs42254-020-0186-4.

[2] E. Knill et al. "Randomized benchmarking of quantum gates". In: *Phys. Rev. A* 77 (1 Jan. 2008), p. 012307. DOI: 10.1103/PhysRevA.77.012307. URL: https://doi.org/10.1103/PhysRevA.77.012307.

[3] Timothy Proctor et al. "What Randomized Benchmarking Actually Measures". In: *Phys. Rev. Lett.* 119 (13 Sept. 2017), p. 130502. DOI: 10.1103/PhysRevLett.119.130502. URL: https://doi.org/10.1103/PhysRevLett.119.130502.

[4] J. Helsen et al. "General Framework for Randomized Benchmarking". In: *PRX Quantum* 3 (2 June 2022), p. 020357. DOI: 10.1103/PRXQuantum.3.020357. URL: https://doi.org/10.1103/PRXQuantum.3.020357.

[5] Marcus P. da Silva, Olivier Landon-Cardinal, and David Poulin. "Practical Characterization of Quantum Devices without Tomography". In: *Physical Review Letters* 107.21 (Nov. 2011). ISSN: 1079-7114. DOI: 10.1103/physrevlett.107.210404. URL: https://doi.org/10.1103/PhysRevLett.107.210404.

[6] Christopher E Granade et al. "Robust online Hamiltonian learning". In: *New Journal of Physics* 14.10 (Oct. 2012), p. 103013. DOI: 10.1088/1367-2630/14/10/103013. URL: https://doi.org/10.1088/1367-2630/14/10/103013.

[7] Nathan Wiebe et al. "Hamiltonian Learning and Certification Using Quantum Resources". In: *Phys. Rev. Lett.* 112 (19 May 2014), p. 190501. DOI: 10.1103/PhysRevLett.112.190501. URL: https://doi.org/10.1103/PhysRevLett.112.190501.

[8] Jianwei Wang et al. "Experimental quantum Hamiltonian learning". In: *Nature Physics* 13.6 (June 2017), pp. 551–555. ISSN: 1745-2481. DOI: 10.1038/nphys4074. URL: https://doi.org/10.1038/nphys4074.

[9] Tim J. Evans, Robin Harper, and Steven T. Flammia. "Scalable Bayesian Hamiltonian learning". In: (2019). arXiv: 1912.07636 [quant-ph]. URL: https://doi.org/10.48550/arXiv.1912.07636.

[10] Anurag Anshu et al. "Sample-efficient learning of interacting quantum systems". In: *Nature Physics* 17.8 (May 2021), pp. 931–935. ISSN: 1745-2481. DOI: 10.1038/s41567-021-01232-0. URL: https://doi.org/10.1038/s41567-021-01232-0.

[11] Jeongwan Haah, Robin Kothari, and Ewin Tang. "Learning quantum Hamiltonians from high-temperature Gibbs states and real-time evolutions". In: *Nature Physics* 20.6 (Mar. 2024), pp. 1027–1031. ISSN: 1745-2481. DOI: 10.1038/s41567-023-02376-x. URL: https://doi.org/10.1038/s41567-023-02376-x.

[12] Wenjun Yu et al. "Robust and Efficient Hamiltonian Learning". In: *Quantum* 7 (June 2023), p. 1045. ISSN: 2521-327X. DOI: 10.22331/q-2023-06-29-1045. URL: https://doi.org/10.22331/q-2023-06-29-1045.

[13] Andi Gu, Lukasz Cincio, and Patrick J. Coles. "Practical Hamiltonian learning with unitary dynamics and Gibbs states". In: *Nature Communications* 15.1 (Jan. 2024). ISSN: 2041-1723. DOI: 10.1038/s41467-023-44008-1. URL: https://doi.org/10.1038/s41467-023-44008-1.

[14] Hsin-Yuan Huang et al. "Learning Many-Body Hamiltonians with Heisenberg-Limited Scaling". In: *Phys. Rev. Lett.* 130 (20 May 2023), p. 200403. DOI: 10.1103/PhysRevLett.130.200403. URL: https://doi.org/10.1103/PhysRevLett.130.200403.

[15] Matthias C. Caro. "Learning Quantum Processes and Hamiltonians via the Pauli Transfer Matrix". In: *ACM Transactions on Quantum Computing* 5.2 (June 2024), pp. 1–53. ISSN: 2643-6817. DOI: 10.1145/3670418. URL: https://doi.org/10.1145/3670418.

[16] Daniel Stilck França et al. *Efficient and robust estimation of many-qubit Hamiltonians*. 2022. DOI: 10.1038/s41467-023-44012-5. arXiv: 2205.09567 [quant-ph]. URL: https://doi.org/10.1038/s41467-023-44012-5.

[17] Alicja Dutkiewicz, Thomas E. O'Brien, and Thomas Schuster. "The advantage of quantum control in many-body Hamiltonian learning". In: *Quantum* 8 (Nov. 2024), p. 1537. ISSN: 2521-327X. DOI: 10.22331/q-2024-11-26-1537. URL: https://doi.org/10.22331/q-2024-11-26-1537.

[18] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements". In: *Nature Physics* 16.10 (June 2020), pp. 1050–1057. ISSN: 1745-2481. DOI: 10.1038/s41567-020-0932-7. URL: https://doi.org/10.1038/s41567-020-0932-7.

[19] William J. Huggins et al. "Nearly Optimal Quantum Algorithm for Estimating Multiple Expectation Values". In: *Physical Review Letters* 129.24 (Dec. 2022). ISSN: 1079-7114. DOI: 10.1103/physrevlett.129.240501. URL: https://doi.org/10.1103/PhysRevLett.129.240501.

[20] Rocco A. Servedio and Steven J. Gortler. "Equivalences and Separations Between Quantum and Classical Learnability". In: *SIAM Journal on Computing* 33.5 (2004), pp. 1067–1092. DOI: 10.1137/S0097539704412910. eprint: https://doi.org/10.1137/S0097539704412910. URL: https://doi.org/10.1137/S0097539704412910.

[21] Scott Aaronson and Daniel Gottesman. "Improved simulation of stabilizer circuits". In: *Physical Review A* 70.5 (Nov. 2004). DOI: 10.1103/physreva.70.052328. URL: https://doi.org/10.1103/PhysRevA.70.052328.

[22] Srinivasan Arunachalam and Ronald de Wolf. "Guest Column: A Survey of Quantum Learning Theory". In: *SIGACT News* 48.2 (June 2017), pp. 41–67. ISSN: 0163-5700. DOI: 10.1145/3106700.3106710. URL: https://doi.org/10.1145/3106700.3106710.

[23] Scott Aaronson. “Shadow tomography of quantum states”. In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2018. Los Angeles, CA, USA: Association for Computing Machinery, 2018, pp. 325–338. ISBN: 9781450355599. DOI: 10.1145/3188745.3188802. URL: https://doi.org/10.1145/3188745.3188802.

[24] Sergey Bravyi and Dmitri Maslov. “Hadamard-Free Circuits Expose the Structure of the Clifford Group”. In: *IEEE Transactions on Information Theory* 67.7 (July 2021), pp. 4546–4563. DOI: 10.1109/tit.2021.3081415. URL: https://doi.org/10.1109/TIT.2021.3081415.

[25] Ewout van den Berg. “A simple method for sampling random Clifford operators”. In: (2021). DOI: 10.48550/ARXIV.2008.06011. URL: https://doi.org/10.48550/arxiv.2008.06011.

[26] András Gilyén et al. “Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics”. In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (June 2019). DOI: 10.1145/3313276.3316366. URL: https://doi.org/10.1145/3313276.3316366.

[27] Andrew M Childs and Nathan Wiebe. “Hamiltonian simulation using linear combinations of unitary operations”. In: *Quantum Information & Computation* 12.11-12 (2012), pp. 901–924. URL: https://doi.org/10.26421/QIC12.11-12-1.

[28] Joran Van Apeldoorn et al. “Quantum SDP-solvers: Better upper and lower bounds”. In: *Quantum* 4 (2020), p. 230. URL: https://doi.org/10.22331/q-2020-02-14-230.

[29] András Gilyén, Srinivasan Arunachalam, and Nathan Wiebe. “Optimizing quantum optimization algorithms via faster quantum gradient computation”. In: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA ’19. San Diego, California: Society for Industrial and Applied Mathematics, 2019, pp. 1425–1444. DOI: 10.1137/1.9781611975482.87. URL: https://doi.org/10.1137/1.9781611975482.87.

[30] Theodore J. Yoder, Guang Hao Low, and Isaac L. Chuang. “Fixed-Point Quantum Search with an Optimal Number of Queries”. In: *Physical Review Letters* 113.21 (Nov. 2014). ISSN: 1079-7114. DOI: 10.1103/physrevlett.113.210501. URL: https://doi.org/10.1103/PhysRevLett.113.210501.

[31] Alexey E. Rastegin. “Relations for Certain Symmetric Norms and Anti-norms Before and After Partial Trace”. In: *Journal of Statistical Physics* 148.6 (Aug. 2012), pp. 1040–1053. DOI: 10.1007/s10955-012-0569-8. URL: https://doi.org/10.1007%2Fs10955-012-0569-8.

[32] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. “Quantum Algorithm for Linear Systems of Equations”. In: *Phys. Rev. Lett.* 103 (15 Oct. 2009), p. 150502. DOI: 10.1103/PhysRevLett.103.150502. URL: https://doi.org/10.1103/PhysRevLett.103.150502.

[33] Andrew M. Childs, Robin Kothari, and Rolando D. Somma. “Quantum Algorithm for Systems of Linear Equations with Exponentially Improved Dependence on Precision”. In: *SIAM Journal on Computing* 46.6 (2017), pp. 1920–1950. DOI: 10.1137/16M1087072. eprint: https://doi.org/10.1137/16M1087072. URL: https://doi.org/10.1137/16M1087072.

[34] A. Y. Kitaev, A. H. Shen, and Vyalyi M. N. *Classical and Quantum Computation*. Vol. 47. American Mathematical Society, 2002. DOI: 10.1090/gsm/047. URL: https://doi.org/10.1090/gsm/047.

[35] Fernando G. S. L. Brandão et al. "Quantum SDP Solvers: Large Speed-Ups, Optimality, and Applications to Quantum Learning". In: *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*. Ed. by Christel Baier et al. Vol. 132. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019, 27:1–27:14. ISBN: 978-3-95977-109-2. DOI: 10.4230/LIPIcs.ICALP.2019.27. URL: https://doi.org/10.4230/LIPIcs.ICALP.2019.27.

[36] Andrew M. Childs, Jin-Peng Liu, and Aaron Ostrander. "High-precision quantum algorithms for partial differential equations". In: *Quantum* 5 (Nov. 2021), p. 574. ISSN: 2521-327X. DOI: 10.22331/q-2021-11-10-574. URL: https://doi.org/10.22331/q-2021-11-10-574.

[37] Ryan Babbush et al. "Exponential Quantum Speedup in Simulating Coupled Classical Oscillators". In: *Phys. Rev. X* 13 (4 Dec. 2023), p. 041041. DOI: 10.1103/PhysRevX.13.041041. URL: https://doi.org/10.1103/PhysRevX.13.041041.

[38] David Sherrington and Scott Kirkpatrick. "Solvable Model of a Spin-Glass". In: *Phys. Rev. Lett.* 35 (26 Dec. 1975), pp. 1792–1796. DOI: 10.1103/PhysRevLett.35.1792. URL: https://doi.org/10.1103/PhysRevLett.35.1792.

[39] David Sherrington. *Physics and Complexity: a brief spin glass perspective*. 2012. arXiv: 1201.1852 [cond-mat.dis-nn]. URL: https://doi.org/10.48550/arXiv.1201.1852.

[40] N Alon and J.H. Spencer. *The Probabilistic Method (Second edition.)* London: Interscience Publishers, 2000.

A Estimating the Hamiltonian Coefficients (Pauli Shadows)

Previously we considered the use of Clifford-based shadows procedure for estimating the Hamiltonian coefficients, but it is also possible to use the Pauli version of classical shadows. Since the operators involved in the analysis of the Pauli-based shadows are all single-qubit operators, we introduce some notation to indicate the specific 2 dimensional subspace an operator acts on, as well as the qubit that a measurement result corresponds to. First, since the subscript i already denotes the round of measurement each bit string $|b_i\rangle$ came from, we use the notation $|b_i[j]\rangle$ to refer to specific bits of $|b_i\rangle$. More specifically,

$$|b_i\rangle = \bigotimes_{j=0}^{\eta-1} |b_i[j]\rangle \tag{130}$$

Once again, recall that $|b_i\rangle$ is a classical bit string, and Dirac notation is only used for presentation purposes. Second, the random unitaries, U_i, applied during each round of measurement are tensor products of single-qubit Clifford operators. Therefore, we write them as

$$U_i = \bigotimes_{j=0}^{\eta-1} U_{(i,j)}, \tag{131}$$

where $U_{(i,j)}$ is a single-qubit operator acting on the j^{th} qubit. Similarly, we can write the Hamiltonian terms H_l as

$$H_l = \bigotimes_{j=0}^{\eta-1} H_{(l,j)} \tag{132}$$

We again consider the process (viewed as a channel) of taking the expectation value over all measurement outcomes and unitaries in the group $\mathrm{Cl}(2)^{\otimes\eta}$. Rather than a single depolarizing channel (as in the case where Clifford operations from $\mathrm{Cl}(2^\eta)$ were randomly sampled), this can be viewed as η depolarizing channels, each acting on a single qubit [18]. Inverting these channels as before gives the classical shadow specified in Definition 36.

Definition 36 (Classical Shadow (From Pauli Measurements))**.** *Given N copies of an η-qubit quantum state ρ, the classical shadows procedure based on random Pauli measurements returns a classical shadow of ρ which is of the form*

$$\hat{\rho} = \{\hat{\rho}_i | i \in \mathbb{Z}_N\}, \tag{133}$$

where

$$\hat{\rho}_i = \bigotimes_{j=0}^{\eta-1} \left(3U_{(i,j)}^\dagger \left|b_i[j]\right\rangle \left\langle b_i[j]\right| U_{(i,j)} - \mathbb{I}\right) \tag{134}$$

Recall that i denotes the round of measurement and j denotes the qubit; the unitary applied to ρ during the i^{th} round of the classical shadows is $U_i = \bigotimes_{j=0}^{\eta-1} U_{(i,j)}$, and $|b_i[j]\rangle$ refers to the outcome of measuring the j^{th} qubit in the computational basis during the i^{th} round of the classical shadows procedure.

Just as in the case where the Clifford flavour of classical shadows was used, the goal of this section is to extract the Hamiltonian coefficients from the pseudo-Choi state - that is, to solve the Pseudo-Choi Hamiltonian Learning problem (PC HLP). However, instead of generating a classical shadow of the pseudo-Choi state, we begin by performing a partial trace over the ancillary system A to simplify the learning process. This results in the following state which will serve as our learning resource:

$$\rho = \frac{1}{d\alpha^2}\Big(H^2 \otimes |0\rangle_C \langle 0|_C + H \otimes |0\rangle_C \langle 1|_C + H \otimes |1\rangle_C \langle 0|_C + \mathbb{I} \otimes |1\rangle_C \langle 1|_C \Big), \tag{135}$$

where $d = 2^n$ is the dimension of the Hilbert space the Hamiltonian acts on, and $\alpha = \sqrt{\|\mathbf{c}\|_2^2 + 1}$. The set of decoding operators whose expectation values correspond to the Hamiltonian coefficients is also modified as per Definition 37:

Definition 37 (Decoding Operators)**.** *Let the set of decoding operators be defined as*

$$\boldsymbol{O} \equiv \{O_l | l \in \mathbb{Z}_M\} \tag{136}$$

where

$$O_l = \frac{H_l \otimes X_C}{2} \tag{137}$$

and H_l is one of the M Hamiltonian terms.

Furthermore, we define

$$O_\alpha \equiv \mathbb{I} \otimes |1\rangle_C \langle 1|_C \tag{138}$$

By Proposition 28, we have that

$$\mathrm{Tr}(\rho O_l) = \frac{c_l}{\alpha^2}, \tag{139}$$

where c_l is the l^{th} Hamiltonian coefficient, and α is the normalization constant of the pseudo-Choi state (7). Furthermore,

$$\mathrm{Tr}(\rho O_\alpha) = \frac{1}{\alpha^2} \tag{140}$$

Now that we have a set of operators whose expectation values correspond to the Hamiltonian coefficients when acting on the resource state ρ (135), we seek to approximate these expectation values using a classical shadow of ρ. Recall that to obtain ρ we traced out the ancillary system A from the pseudo-Choi state (15). To generate a classical shadow of ρ, starting from the pseudo-Choi state, we must therefore ignore the ancilla system A during the classical shadows procedure, meaning that all random (single-qubit) Clifford operations and measurements are only on $n+1$ qubits (n qubits for the system on Hilbert space $\mathcal{H}_S$ and one qubit for ancilla on Hilbert space $\mathcal{H}_C$). In terms of the definition of the classical shadow, we let $\eta = n+1$ so that a classical shadow is made up of N components of the form

$$\hat{\rho}_i = \bigotimes_{j=0}^{n} \left(3U^\dagger_{(i,j)} \left| b_i[j] \right\rangle \left\langle b_i[j] \right| U_{(i,j)} - \mathbb{I} \right) \tag{141}$$

With this, and the definition of the decoding operators given in Definition 37, we can write the expectation values of interest in the explicit forms given in Claim 38.

Claim 38. *Let X_C denote the single-qubit Pauli X operation acting on $\mathcal{H}_C$, and let $H_{l,j}$ be the single qubit Pauli operator from H_l that acts on the j^{th} qubit (recall that each H_l is an n-qubit Pauli operator on $\mathcal{H}_S$).*

$$\mathrm{Tr}(\hat{\rho}_i O_l) = \frac{3}{2} \left\langle b_i[n] \right| U_{(i,n)} X_C U^\dagger_{(i,n)} \left| b_i[n] \right\rangle \prod_{j=0}^{n-1} \left(3 \left\langle b_i[j] \right| U_{(i,j)} H_{(l,j)} U^\dagger_{(i,j)} \left| b_i[j] \right\rangle - \mathrm{Tr}(H_{(l,j)}) \right), \tag{142}$$

Similarly,

$$\mathrm{Tr}(\hat{\rho}_i O_\alpha) = 3 \left| \left\langle b_i[n] \right| U_{(i,n)} \left| 1 \right\rangle_C \right|^2 - 1 \tag{143}$$

Proof of Claim 38. Recall that $\hat{\rho}_i$ is on $n+1$ qubits, H_l is on n qubits, and X_C is the single qubit

Pauli X operator. Then,

$$\begin{aligned}
\mathrm{Tr}[\hat{\rho}_i O_l] =& \mathrm{Tr}\left[\left(\bigotimes_{j=0}^{n}\left(3U_{(i,j)}^{\dagger}\left|b_i[j]\right\rangle\left\langle b_i[j]\right|U_{(i,j)}-\mathbb{I}\right)\right)\left(\frac{H_l\otimes X_C}{2}\right)\right] \\
=& \frac{1}{2}\mathrm{Tr}\left[\left(\bigotimes_{j=0}^{n}\left(3U_{(i,j)}^{\dagger}\left|b_i[j]\right\rangle\left\langle b_i[j]\right|U_{(i,j)}-\mathbb{I}\right)\right)\left(\left(\bigotimes_{j=0}^{n-1}H_{(l,j)}\right)\otimes X_C\right)\right] \\
=& \frac{1}{2}\mathrm{Tr}\Bigg[\left(\bigotimes_{j=0}^{n-1}\left(3U_{(i,j)}^{\dagger}\left|b_i[j]\right\rangle\left\langle b_i[j]\right|U_{(i,j)}H_{(l,j)}-H_{(l,j)}\right)\right) \\
& \otimes\left(3U_{(i,n)}^{\dagger}\left|b_i[n]\right\rangle\left\langle b_i[n]\right|U_{(i,n)}X_C-X_C\right)\Bigg] \\
=& \frac{3}{2}\mathrm{Tr}\left[U_{(i,n)}^{\dagger}\left|b_i[n]\right\rangle\left\langle b_i[n]\right|U_{(i,n)}X_C\right]\prod_{j=0}^{n-1}\mathrm{Tr}\left[3U_{(i,j)}^{\dagger}\left|b_i[j]\right\rangle\left\langle b_i[j]\right|U_{(i,j)}H_{(l,j)}-H_{(l,j)}\right] \\
=& \frac{3}{2}\left\langle b_i[n]\right|U_{(i,n)}X_CU_{(i,n)}^{\dagger}\left|b_i[n]\right\rangle\prod_{j=0}^{n-1}\left(3\left\langle b_i[j]\right|U_{(i,j)}H_{(l,j)}U_{(i,j)}^{\dagger}\left|b_i[j]\right\rangle-\mathrm{Tr}\left[H_{(l,j)}\right]\right),
\end{aligned} \tag{144}$$

which gives the result of equation (142).

Likewise,

$$\begin{aligned}
\mathrm{Tr}[\hat{\rho}_i O_\alpha] =& \mathrm{Tr}\left[\left(\bigotimes_{j=0}^{n}\left(3U_{(i,j)}^{\dagger}\left|b_i[j]\right\rangle\left\langle b_i[j]\right|U_{(i,j)}-\mathbb{I}\right)\right)\left(\mathbb{I}_S\otimes\left|1\right\rangle_C\left\langle 1\right|_C\right)\right] \\
=& \mathrm{Tr}\Bigg[\left(\bigotimes_{j=0}^{n-1}\left(3U_{(i,j)}^{\dagger}\left|b_i[j]\right\rangle\left\langle b_i[j]\right|U_{(i,j)}-\mathbb{I}\right)\right) \\
& \otimes\left(\left(3U_{(i,n)}^{\dagger}\left|b_i[n]\right\rangle\left\langle b_i[n]\right|U_{(i,n)}-\mathbb{I}\right)\left|1\right\rangle_C\left\langle 1\right|_C\right)\Bigg] \\
=& \prod_{j=0}^{n-1}\left(\mathrm{Tr}\left[3U_{(i,j)}^{\dagger}\left|b_i[j]\right\rangle\left\langle b_i[j]\right|U_{(i,j)}\right]-\mathrm{Tr}\left[\mathbb{I}\right]\right) \\
& \left(3\left\langle b_i[n]\right|U_{(i,n)}\left|1\right\rangle_C\left\langle 1\right|_CU_{(i,n)}^{\dagger}\left|b_i[n]\right\rangle-1\right) \\
=& 3\left\langle b_i[n]\right|U_{(i,n)}\left|1\right\rangle_C\left\langle 1\right|_CU_{(i,n)}^{\dagger}\left|b_i[n]\right\rangle-1,
\end{aligned} \tag{145}$$

which gives the result of equation (143), completing the proof. □

Since only single qubit operations are required, computing each trace takes $\mathcal{O}(n)$ time, and requires $\mathcal{O}(n)$ memory to store all intermediate values. In fact, if we focus on k-local Hamiltonians, each Hamiltonian term H_l only contains at most k non-identity Paulis, so only $\mathcal{O}(k)$ operations are required to compute each value.

As was the case with the random Clifford measurements, it is unnecessary to store the entire classical shadow:

Definition 39. *If the random Pauli measurement version of the classical shadows procedure is performed, the resulting classical shadow of size N can be stored in the following way:*

$$|\hat{\rho}\rangle = \{|\hat{\rho}_i\rangle \, | i \in \mathbb{Z}_N\}, \tag{146}$$

where

$$|\hat{\rho}_i\rangle = \bigotimes_{j=1}^{n+1} U^{\dagger}_{(i,j)} \, |b[j]\rangle_i \tag{147}$$

Here we use Dirac notation to describe the resultant bit-strings, however note that as $|b[j]\rangle_i$ is a classical bit, so $|\hat{\rho}_i\rangle$ is a computational basis vector which corresponds to a one-hot unary representation of a bit string.

Algorithm 5 describes the full procedure for estimating the Hamiltonian coefficients. Note that it uses the "compressed" Pauli shadows given by Definition 39. Also, there is no need to use the Tableau or SIP Algorithms of [21] to store intermediary values or compute the expectation values since all the operations involved are on single qubits.

Since all the operations involved in this process are single-qubit operations, computing the required inner products is more efficient than in the case of Clifford-based shadows. In particular, the number of computations required to do so is at most $\mathcal{O}(MN)$ where a sufficient value of N is given in equation (150). Note that this value of N includes a factor of 4^k, where k is the locality of the Hamiltonian. Meanwhile, the Clifford-based procedure requires $\mathcal{O}(MNn^3)$ computations to calculate the necessary inner products, where the sufficient value of N is given in Theorem 16 and does not contain the factor of 4^k. Therefore, unless the value of k is large, the Pauli-based procedure will have the better post-processing performance, at the expense of having a slightly higher sample complexity. Furthermore, as in the Clifford case, these operations can be done in parallel, greatly reducing the computation time.

A.1 Sample Complexity Upper Bound

The sample complexity of this method is very similar to that of the case where Clifford based classical shadows was used to learn the pseudo-Choi state. The key difference is that when the random Pauli measurement version of classical shadows is used, the upper bound on the shadow norm in Lemma 14 no longer applies. Instead, we have the following result:

Lemma 40 (Shadow Norm Upper Bound (Pauli Shadows))**.** *Let the set $\{O_\alpha\} \cup \{O_l \mid l \in \mathbb{Z}_M\}$ be denoted by $\boldsymbol{O} \equiv \{O_i \mid i \in \mathbb{Z}_{2M}\}$. Note that these operators act non-trivially on at most $k+1$ qubits. If using the random Pauli measurement version of the classical shadows procedure to predict the expectation values of the operators in the set, the shadow norm of each operator is at most*

$$\|O_i\|^2_{shadow} \leq 3^{k+1} \tag{148}$$

Proof of Lemma 40. This results directly from [18], where Huang et al. showed that for classical shadows based on random Pauli measurements

$$\|O\|^2_{shadow} \leq 3^k \tag{149}$$

where O is a single k-local Pauli operator.

□

Algorithm 5 `FindCoeffPauli`$(\rho_c^{\otimes N}, \boldsymbol{H}, n, N)$: Determine the Vector of Hamiltonian Coefficients from ρ_c

Input:
$\rho_c^{\otimes N}$: Collection of N pseudo-Choi states on n qubits, each on Hilbert space $\mathcal{H}_S \otimes \mathcal{H}_A \otimes \mathcal{H}_C$
$\boldsymbol{H}$: Set of M k-local Hamiltonian terms H_l whose coefficients c_m are desired
▷ $\rho_c^{\otimes N}$ is a quantum state, whereas $\boldsymbol{H}$ is a classical set of operators

1. **Initialize the array of Hamiltonian coefficients**

   Let $\hat{\mathbf{c}} \leftarrow \Big[0, 0, ..., 0\Big]_{1\times M}$ and let $\hat{\mathbf{c}}_l$ denote the l^{th} element of $\hat{\mathbf{c}}$

2. **Generate and store the classical shadow**

   Use copies of ρ_c to generate $\hat{\rho}$, a classical shadow of size N of $\mathrm{Tr}_A(\rho_c)$, by limiting the random Pauli measurement based classical shadows procedure to the space $\mathcal{H}_S \otimes \mathcal{H}_C$ to effectively perform the partial trace over $\mathcal{H}_A$.

   Store $\hat{\rho}$ in classical memory (i.e. store each vector $|\hat{\rho}_i\rangle$)

   ▷ See Definition 39 of a classical shadow

3. **Generate an estimate of** $\mathrm{Tr}(\rho O_l) = \frac{c_l}{\alpha^2}$ **for each corresponding Hamiltonian term**

   Let $\hat{\mathbf{o}} \leftarrow \Big[0, 0, ..., 0\Big]_{1\times N}$ and let $\hat{\mathbf{o}}_i$ denote the i^{th} element of $\hat{\mathbf{o}}$

   `For` l `from` 0 `to` $M-1$:

   `For` i `from` 0 to $N-1$:

   $\hat{o}_i \leftarrow \frac{3}{2} \langle b_i[n]| U_{(i,n)} X_C U_{(i,n)}^\dagger |b_i[n]\rangle \prod_{j=0}^{n-1} \Big(3 \langle b_i[j]| U_{(i,j)} H_{(l,j)} U_{(i,j)}^\dagger |b_i[j]\rangle - \mathrm{Tr}(H_{(l,j)})\Big)$

   ▷ Equal to $\mathrm{Tr}\left(\hat{\rho}_i O_l\right)$ as per Equation (142)

   $\hat{\mathbf{c}}_l \leftarrow$ `MedianOfMeans`$(\hat{\mathbf{o}})$

4. **Generate an estimate of** $\mathrm{Tr}(\rho O_\alpha) = \frac{1}{\alpha^2}$

   `For` i `from` 0 to $N-1$:

   $\hat{o}_i \leftarrow 3 \Big|\langle b_i[n]| U_{(i,n)} |1\rangle_C\Big|^2 - 1$ ▷ Equal to $\mathrm{Tr}(\hat{\rho}_i O_\alpha)$ as per Equation (143)

   $\hat{o}_\alpha \leftarrow$ `MedianOfMeans`$(\hat{\mathbf{o}})$

5. **Remove the factor of** $\frac{1}{\alpha^2}$ **from the Hamiltonian coefficients**

   Let $\hat{\mathbf{c}} \leftarrow \frac{\hat{\mathbf{c}}}{\hat{o}_\alpha}$

6. **Output the vector of Hamiltonian coefficients**

   Return $\hat{\mathbf{c}} = \Big[\hat{c}_0, \hat{c}_1, ..., \hat{c}_{M-1}\Big]$

By the same argument as in the case of Clifford based classical shadows, this upper bound on the shadow norm leads to a sample complexity of

$$N \in \mathcal{O}\left(\frac{3^k \alpha^4 (c_{\max}+1) M \log\left(M/\delta\right)}{\epsilon^2}\right) \tag{150}$$

B Proofs and technical results

B.1 Proof of Equation (7)

Proof.

$$\begin{aligned}
\langle\Phi_d|_{SA}\left(H^2\otimes\mathbb{I}_A\right)|\Phi_d\rangle_{SA} &= \frac{\sum_{i=0}^{d-1}\langle i|_S\langle i|_A\left(H^2\otimes\mathbb{I}_A\right)\sum_{j=0}^{d-1}|j\rangle_S|j\rangle_A}{d}\\
&= \frac{\sum_{i=0}^{d-1}\sum_{j=0}^{d-1}\langle i|_S H^2|j\rangle_S\langle i|_A|j\rangle_A}{d}\\
&= \frac{\sum_{i=0}^{d-1}\langle i|_S H^2|i\rangle_S}{d}\\
&= \frac{\mathrm{Tr}\left(H^2\right)}{d}\\
&= \frac{1}{d}\mathrm{Tr}\left(\sum_{l=0}^{M-1}c_l H_l\sum_{m=0}^{M-1}c_m H_m\right)\\
&= \frac{1}{d}\sum_{l=0}^{M-1}\sum_{m=0}^{M-1}c_l c_m\mathrm{Tr}(H_l H_m)\\
&= \sum_{m=0}^{M-1}c_m^2\\
&= \|\mathbf{c}\|_2^2
\end{aligned} \tag{151}$$

Therefore

$$\begin{aligned}
\alpha &:= \sqrt{\langle\Phi_d|_{SA}\left(H^2\otimes\mathbb{I}_A\right)|\Phi_d\rangle_{SA}+1}\\
&= \sqrt{\|\mathbf{c}\|_2^2+1}
\end{aligned} \tag{152}$$

□

B.2 Proof of Proposition 12

Proposition 12 contains equations showing how to explicitly compute three expectation values of interest using a classical shadow. This section outlines the derivation of each of these three expressions.

Recall that classical shadows generated by the random Clifford measurement version of the classical shadows procedure are of the form $\hat{\rho}_i = \left(2^{2n+1}+1\right)U_i^\dagger|b_i\rangle\langle b_i|U_i - \mathbb{I}$, and recall the definitions of the following operators:

$$\begin{aligned}
O_l^+ &= O_l + (O_l)^\dagger\\
O_l^- &= iO_l - i\,(O_l)^\dagger
\end{aligned}$$

where $O_l = (H_l\otimes\mathbb{I})|\Phi_d\rangle_{SA}\langle\Phi_d|_{SA}\otimes|0\rangle_C\langle 1|_C$ is a decoding operator as in Definition 9.

Also, note that $U_i^\dagger|b_i\rangle$ is on the Hilbert space $\mathcal{H}_S\otimes\mathcal{H}_A\otimes\mathcal{H}_C$, and H_l acts on the Hilbert space $\mathcal{H}_S$.

Proof of equation (29).

$$
\begin{aligned}
\operatorname{Tr}\left(\hat{\rho}_i O_l^+\right) =& \operatorname{Tr}\Big(\Big[\Big(2^{2n+1}+1\Big) U_i^\dagger |b_i\rangle \langle b_i| U_i - \mathbb{I}\Big] \Big[(H_l \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |0\rangle_C \langle 1|_C \\
&+ |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} (H_l \otimes \mathbb{I}_A) \otimes |1\rangle_C \langle 0|_C \Big]\Big) \\
=& \Big(2^{2n+1}+1\Big) \operatorname{Tr}\left(\Big[U_i^\dagger |b_i\rangle \langle b_i| U_i\Big] \Big[(H_l \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |0\rangle_C \langle 1|_C \Big]\right) \\
&- \operatorname{Tr}\Big((H_l \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |0\rangle_C \langle 1|_C \Big) \\
&+ \Big(2^{2n+1}+1\Big) \operatorname{Tr}\left(\Big[U_i^\dagger |b_i\rangle \langle b_i| U_i\Big] \Big[|\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} (H_l \otimes \mathbb{I}_A) \otimes |1\rangle_C \langle 0|_C \Big]\right) \\
&- \operatorname{Tr}\Big(|\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} (H_l \otimes \mathbb{I}_A) \otimes |1\rangle_C \langle 0|_C \Big) \\
=& \Big(2^{2n+1}+1\Big) \operatorname{Tr}\left(\Big[U_i^\dagger |b_i\rangle \langle b_i| U_i\Big] \Big[(H_l \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |0\rangle_C \langle 1|_C \Big]\right) \\
&- \operatorname{Tr}\Big((H_l \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \Big) \operatorname{Tr}\left(|0\rangle_C \langle 1|_C\right) \\
&+ \Big(2^{2n+1}+1\Big) \operatorname{Tr}\left(\Big[U_i^\dagger |b_i\rangle \langle b_i| U_i\Big] \Big[|\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} (H_l \otimes \mathbb{I}_A) \otimes |1\rangle_C \langle 0|_C \Big]\right) \\
&- \operatorname{Tr}\Big(|\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} (H_l \otimes \mathbb{I}_A) \Big) \operatorname{Tr}\Big(|1\rangle_C \langle 0|_C \Big) \\
=& \Big(2^{2n+1}+1\Big) \operatorname{Tr}\left(\Big[U_i^\dagger |b_i\rangle \langle b_i| U_i\Big] \Big[(H_l \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |0\rangle_C \langle 1|_C \Big]\right) \\
&+ \Big(2^{2n+1}+1\Big) \operatorname{Tr}\left(\Big[U_i^\dagger |b_i\rangle \langle b_i| U_i\Big] \Big[|\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} (H_l \otimes \mathbb{I}_A) \otimes |1\rangle_C \langle 0|_C \Big]\right) \\
=& \Big(2^{2n+1}+1\Big) \langle b_i| U_i \Big((H_l \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |0\rangle_C \langle 1|_C \Big) U_i^\dagger |b_i\rangle \\
&+ \Big(2^{2n+1}+1\Big) \langle b_i| U_i \Big(|\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} (H_l \otimes \mathbb{I}_A) \otimes |1\rangle_C \langle 0|_C \Big) U_i^\dagger |b_i\rangle \\
=& \Big(2^{2n+1}+1\Big) \Big(\langle b_i| U_i (H_l \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA} \otimes |0\rangle_C \Big) \Big(\langle\Phi_d|_{SA} \otimes \langle 1|_C U_i^\dagger |b_i\rangle \Big) \\
&+ \Big(2^{2n+1}+1\Big) \Big(\langle b_i| U_i |\Phi_d\rangle_{SA} \otimes |1\rangle_C \Big) \Big(\langle\Phi_d|_{SA} (H_l \otimes \mathbb{I}_A) \otimes \langle 0|_C U_i^\dagger |b_i\rangle \Big)
\end{aligned}
\tag{153}
$$

□

Proof of equation (30).

$$\begin{aligned}
\mathrm{Tr}\left(\hat{\rho}_i O_l^-\right) =& \mathrm{Tr}\Big(\Big[\Big(2^{2n+1}+1\Big) U_i^\dagger |b_i\rangle \langle b_i| U_i - \mathbb{I}\Big] \Big[i \left(H_l \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |0\rangle_C \langle 1|_C \\
& - i |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \left(H_l \otimes \mathbb{I}_A\right) \otimes |1\rangle_C \langle 0|_C \Big] \Big) \\
=& i \Big(2^{2n+1}+1\Big) \mathrm{Tr}\left(\Big[U_i^\dagger |b_i\rangle \langle b_i| U_i\Big] \Big[\left(H_l \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |0\rangle_C \langle 1|_C \Big] \right) \\
& - i \mathrm{Tr}\left(\left(H_l \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |0\rangle_C \langle 1|_C \right) \\
& - i \Big(2^{2n+1}+1\Big) \mathrm{Tr}\left(\Big[U_i^\dagger |b_i\rangle \langle b_i| U_i\Big] \Big[|\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \left(H_l \otimes \mathbb{I}_A\right) \otimes |1\rangle_C \langle 0|_C \Big] \right) \\
& + i \mathrm{Tr}\left(|\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \left(H_l \otimes \mathbb{I}_A\right) \otimes |1\rangle_C \langle 0|_C \right) \\
=& i \Big(2^{2n+1}+1\Big) \langle b_i| U_i \Big(\left(H_l \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |0\rangle_C \langle 1|_C \Big) U_i^\dagger |b_i\rangle \\
& - i \Big(2^{2n+1}+1\Big) \langle b_i| U_i \Big(|\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \left(H_l \otimes \mathbb{I}_A\right) \otimes |1\rangle_C \langle 0|_C \Big) U_i^\dagger |b_i\rangle \\
=& i \Big(2^{2n+1}+1\Big) \Big(\langle b_i| U_i \left(H_l \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA} \otimes |0\rangle_C \Big) \Big(\langle\Phi_d|_{SA} \otimes \langle 1|_C U_i^\dagger |b_i\rangle \Big) \\
& - i \Big(2^{2n+1}+1\Big) \Big(\langle b_i| U_i |\Phi_d\rangle_{SA} \otimes |1\rangle_C \Big) \Big(\langle\Phi_d|_{SA} \left(H_l \otimes \mathbb{I}_A\right) \otimes \langle 0|_C U_i^\dagger |b_i\rangle \Big)
\end{aligned} \tag{154}$$

□

Proof of equation (31). Recall from Definition 9 that $O_\alpha = |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |1\rangle_C \langle 1|_C$ is the decoding operator corresponding to the normalization constant α. From this it follows that

$$\begin{aligned}
\mathrm{Tr}(\hat{\rho}_i O_\alpha) =& \mathrm{Tr}\Big(\Big[\Big(2^{2n+1}+1\Big) U_i^\dagger |b_i\rangle \langle b_i| U_i - \mathbb{I}\Big] \Big[|\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |1\rangle_C \langle 1|_C \Big] \Big) \\
=& \Big(2^{2n+1}+1\Big) \mathrm{Tr}\Big(\Big[U_i^\dagger |b_i\rangle \langle b_i| U_i\Big] \Big[|\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |1\rangle_C \langle 1|_C \Big] \Big) \\
& - \mathrm{Tr}\Big(|\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |1\rangle_C \langle 1|_C \Big) \\
=& \Big(2^{2n+1}+1\Big) \langle b_i| U_i \Big(|\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |1\rangle_C \langle 1|_C \Big) U_i^\dagger |b_i\rangle - 1 \\
=& \Big(2^{2n+1}+1\Big) |\langle b_i| U_i |\phi_d\rangle_{SA} |1\rangle_C|^2 - 1
\end{aligned} \tag{155}$$

□

B.3 Proof of Lemma 14

Proof of Lemma 14. Recall the definitions of the Hermitian operators in Lemma 14:

$$O_\alpha = |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |1\rangle \langle 1| , \tag{156}$$

and

$$O_l^+ = O_l + (O_l)^\dagger \tag{157}$$

$$O_l^- = i O_l - i (O_l)^\dagger \tag{158}$$

where

$$O_l = (H_l \otimes \mathbb{I}) |\Phi_d\rangle_{SA} \langle\Phi_d|_{SA} \otimes |0\rangle \langle 1| \tag{159}$$

is defined for all $0 \leq l \leq M-1$.

First consider the square of the Hilbert-Schmidt norm of O_l^+:

$$\begin{aligned}\mathrm{Tr}\left(\left(O_l^+\right)\left(O_l^+\right)^\dagger\right) &= \mathrm{Tr}\left(\left(O_l^+\right)^2\right) \\ &= \mathrm{Tr}\left(O_l^2\right) + \mathrm{Tr}\left(O_l O_l^\dagger\right) + \mathrm{Tr}\left(O_l^\dagger O_l\right) + \mathrm{Tr}\left(\left(O_l^\dagger\right)^2\right) \\ &= \mathrm{Tr}\left(O_l^2\right) + 2\mathrm{Tr}\left(O_l O_l^\dagger\right) + \mathrm{Tr}\left(\left(O_l^\dagger\right)^2\right) \end{aligned} \tag{160}$$

Examining each term above individually, we see that:

$$\begin{aligned}\mathrm{Tr}\left(O_l^2\right) &= \mathrm{Tr}\left(\left(\left(H_l \otimes \mathbb{I}\right)|\Phi_d\rangle_{SA}\langle\Phi_d|_{SA} \otimes |0\rangle\langle 1|\right)^2\right) \\ &= \mathrm{Tr}\Big(\left(\left(H_l \otimes \mathbb{I}\right)|\Phi_d\rangle_{SA}\langle\Phi_d|_{SA} \otimes |0\rangle\langle 1|\right)\left(\left(H_l \otimes \mathbb{I}\right)|\Phi_d\rangle_{SA}\langle\Phi_d|_{SA} \otimes |0\rangle\langle 1|\right)\Big) \\ &= \mathrm{Tr}\Big(\langle\Phi_d|_{SA}\left(H_l \otimes \mathbb{I}\right)|\Phi_d\rangle_{SA}\langle\Phi_d|_{SA}\left(H_l \otimes \mathbb{I}\right)|\Phi_d\rangle_{SA}\Big)\mathrm{Tr}\Big(|0\rangle\langle 1||0\rangle\langle 1|\Big) \\ &= 0 \end{aligned} \tag{161}$$

Similarly,

$$\mathrm{Tr}\left(\left(O_l^\dagger\right)^2\right) = 0 \tag{162}$$

$$\begin{aligned}\mathrm{Tr}\left(O_l O_l^\dagger\right) &= \mathrm{Tr}\left(\Big(\left(H_l \otimes \mathbb{I}\right)|\Phi_d\rangle_{SA}\langle\Phi_d|_{SA} \otimes |0\rangle\langle 1|\Big)\Big(\left(H_l \otimes \mathbb{I}\right)|\Phi_d\rangle_{SA}\langle\Phi_d|_{SA} \otimes |0\rangle\langle 1|\Big)^\dagger\right) \\ &= \mathrm{Tr}\left(\Big(\left(H_l \otimes \mathbb{I}\right)|\Phi_d\rangle_{SA}\langle\Phi_d|_{SA} \otimes |0\rangle\langle 1|\Big)\Big(|\Phi_d\rangle_{SA}\langle\Phi_d|_{SA}\left(H_l^\dagger \otimes \mathbb{I}\right) \otimes |1\rangle\langle 0|\Big)\right) \\ &= \mathrm{Tr}\left(\left(H_l \otimes \mathbb{I}\right)|\Phi_d\rangle_{SA}\langle\Phi_d|_{SA}\left(H_l^\dagger \otimes \mathbb{I}\right)\right)\mathrm{Tr}\left(|0\rangle\langle 1||1\rangle\langle 0|\right) \\ &= \langle\Phi_d|_{SA}\left(H_l^\dagger \otimes \mathbb{I}\right)\left(H_l \otimes \mathbb{I}\right)|\Phi_d\rangle_{SA} \\ &= \langle\Phi_d|_{SA}|\Phi_d\rangle_{SA} \\ &= 1 \end{aligned} \tag{163}$$

where we have used the fact that H_l is part of an orthonormal basis, so $H_l^\dagger H_l = \mathbb{I}$.

From this we arrive at the following upper bound:

$$\mathrm{Tr}\left(\left(O_l^+\right)^2\right) = 2 \tag{164}$$

Likewise:

$$\mathrm{Tr}\left(\left(O_l^-\right)^2\right) = 2 \tag{165}$$

Finally, it is obvious from the definition of O_α that

$$\mathrm{Tr}\left(O_\alpha^2\right) = 1 \tag{166}$$

Section 5B in [18] proves that $\|O_i\|^2_{shadow} \leq 3\text{Tr}\left((O_i)^2\right)$ for a Hermitian operator O_i.

Since Lemma 14 defines $\mathbf{O} \equiv \{O_i \mid i \in \mathbb{Z}_{2M}\}$ as the set $\{O_\alpha\} \cup \{O_l^+, O_l^- \mid l \in \mathbb{Z}_M\}$, we see that

$$\text{Tr}\left((O_i)^2\right) \leq 2, \tag{167}$$

so

$$\|O_i\|^2_{shadow} \leq 6, \tag{168}$$

completing the proof of Lemma 14. □

B.4 Choosing a Sufficiently Small Classical Shadows Error

To keep the total error of the Hamiltonian learning problem to at most ϵ, the error of the classical shadows procedure, ϵ_s, cannot be too large. However, decreasing ϵ_s requires increasing the number of measurements N. Therefore, we seek the largest value of ϵ_s such that the total error is at most ϵ. This value is given by Proposition 15, which is proven below.

Proof of Proposition 15. Using classical shadows in Algorithm 1, we compute $\hat{o}_i$ and $\hat{o}_\alpha$, which, with a probability of at least δ_s, are estimates of $\frac{1}{\alpha^2}$ and $\frac{c_i}{\alpha^2}$ to within an error of ϵ_s . This means that upon success, the uncertainty in α^2, which we get by taking the reciprocal of our estimate of $\frac{1}{\alpha^2}$, is at most

$$\epsilon_\alpha = \alpha^4 \epsilon_s \tag{169}$$

The next step in Algorithm 1 is to divide $\hat{o}_i$ by $\hat{o}_\alpha$ to get $\hat{c}_i$, our estimate of c_i. The uncertainty in c_i is therefore at most

$$\begin{aligned} \epsilon_i &= c_i \sqrt{\left(\frac{\epsilon_\alpha}{\alpha^2}\right)^2 + \left(\frac{\epsilon_s}{\frac{c_i}{\alpha^2}}\right)^2} \\ &= \epsilon_s \alpha^2 \sqrt{c_i^2 + 1} \end{aligned} \tag{170}$$

The total l_2 error in the vector of Hamiltonian coefficients is then

$$\begin{aligned} \|\hat{\mathbf{c}} - \mathbf{c}\|_2 &\leq \sqrt{M} \max_i \epsilon_i \\ &= \sqrt{M} \epsilon_s \alpha^2 \sqrt{\max_i c_i^2 + 1} \end{aligned} \tag{171}$$

For this total error to be at most ϵ, as required by the Hamiltonian learning problem, it serves to choose

$$\epsilon_s = \frac{\epsilon}{\alpha^2 \sqrt{c_{\max}^2 + 1} \sqrt{M}} \tag{172}$$

□

B.5 Proof of Lemma 19

Here we provide a formal proof of Lemma 19 which provides our key result surrounding the cost of preparing the pseudo-Choi state that represents the Hamiltonian.

Proof of Lemma 19. Begin by preparing the state $|\psi_0\rangle = |0\rangle_C |0\rangle_B |\Phi_d\rangle_{SA}$, and use it as the input state to the circuit in Figure 4.1. Applying the Hadamard gate to register C gives

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}\Big(|0\rangle_C |0\rangle_B |\Phi_d\rangle_{SA} + |1\rangle_C |0\rangle_B |\Phi_d\rangle_{SA} \Big) \tag{173}$$

After the controlled application of U_{block}, which acts on registers B and S, we have the state

$$|\psi_2\rangle = \frac{1}{\sqrt{2}}\Big(|0\rangle_C (U_{block} \otimes \mathbb{I}_A) |0\rangle_B |\Phi_d\rangle_{SA} + |1\rangle_C |0\rangle_B |\Phi_d\rangle_{SA} \Big) \tag{174}$$

Recall that the block-encoding of the Hamiltonian is of the form

$$U_{block} = |0\rangle_B \langle 0|_B \otimes \frac{2\widetilde{H}t}{\pi} + |0\rangle_B \langle 1|_B \otimes I_S + |1\rangle_B \langle 0|_B \otimes J_S + |1\rangle_B \langle 1|_B \otimes K_S \tag{175}$$

where I_S, J_S, and K_S are "junk" operators we do not care about. Thus we can write the previous state as

$$|\psi_2\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle_C |0\rangle_B \left(\frac{\widetilde{H}}{\Delta} \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA} + |0\rangle_C |1\rangle_B (J_S \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA} + |1\rangle_C |0\rangle_B |\Phi_d\rangle_{SA} \right) \tag{176}$$

The probability of obtaining the outcome $|0\rangle_B$ upon measuring the block-encoding qubit is given by:

$$\begin{aligned}
\mathbf{Pr}(0) &= \langle\psi_2| (\mathbb{I}_C \otimes |0\rangle_B \langle 0|_B \otimes \mathbb{I}_{SA}) |\psi_2\rangle \\
&= \langle\psi_2| \left(\frac{|0\rangle_C |0\rangle_B \left(\frac{\widetilde{H}}{\Delta} \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA} + |1\rangle_C |0\rangle_B |\Phi_d\rangle_{SA}}{\sqrt{2}} \right) \\
&= \frac{1}{2}\left(\langle\Phi_d|_{SA} \left(\frac{\widetilde{H}^\dagger \widetilde{H}}{\Delta^2} \otimes \mathbb{I}_A \right) |\Phi_d\rangle_{SA} + 1 \right) \\
&= \frac{\|\widetilde{\mathbf{c}}\|_2^2}{2\Delta^2} + \frac{1}{2},
\end{aligned} \tag{177}$$

where in the last line we inserted the representation of the Hamiltonian $\widetilde{H}$ from Definition 20 (the missing steps are nearly identical to the proof of Equation (7) in Appendix B.1).

If the outcome of the measurement of register B is $|0\rangle_B$, the resulting state is

$$\begin{aligned}
|\psi_3\rangle &= \frac{(\mathbb{I}_C \otimes |0\rangle_B \langle 0|_B \otimes \mathbb{I}_{SA}) |\psi_2\rangle}{\sqrt{\mathbf{Pr}(0)}} \\
&= \frac{|0\rangle_C |0\rangle_B (\frac{\widetilde{H}}{\Delta} \otimes \mathbb{I}_A) |\Phi_d\rangle_{SA} + |1\rangle_C |0\rangle_B |\Phi_d\rangle_{SA}}{\sqrt{2}\sqrt{\frac{\|\widetilde{\mathbf{c}}\|_2^2}{2\Delta^2} + \frac{1}{2}}}
\end{aligned} \tag{178}$$

where $\Delta = \frac{\pi}{2t}$.

Ignoring the extra qubit in register B, this is a pseudo-Choi state of $\frac{\widetilde{H}}{\Delta}$:

$$|\psi_c'\rangle = \left(\frac{|0\rangle_C \left(\frac{\widetilde{H}}{\Delta} \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA} + |1\rangle_C |\Phi_d\rangle_{SA}}{\gamma} \right), \tag{179}$$

where

$$\gamma = \sqrt{\frac{\|\widetilde{\mathbf{c}}\|_2^2}{\Delta^2} + 1}. \tag{180}$$

□

To check that the success probability of producing the pseudo-Choi state (i.e. measuring the block-encoding qubit and getting the outcome $|0\rangle_B$) is a valid probability, we would like to be sure that the term $\langle\Phi_d|_{SA} \left(\frac{\widetilde{H}^\dagger \widetilde{H}}{\Delta^2} \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA}$ is in $[1/2, 1]$. We start by relating $\|\widetilde{\mathbf{c}}\|_2^2$ to the spectral norm of the Hamiltonian:

$$\begin{aligned}
\|\widetilde{\mathbf{c}}\|_2^2 &= \langle\Phi_d|_{SA} \left(\widetilde{H}^\dagger \widetilde{H} \otimes \mathbb{I}_A\right) |\Phi_d\rangle_{SA} \\
&= \frac{1}{d} \sum_{i=1}^{d} \sum_{j=1}^{d} \langle i|_S \langle i|_A \left(\widetilde{H}^\dagger \widetilde{H} \otimes \mathbb{I}_A\right) |j\rangle_S |j\rangle_A \\
&= \frac{1}{d} \sum_{i=1}^{d} \sum_{j=1}^{d} \langle i|_S \widetilde{H}^\dagger \widetilde{H} |j\rangle_S \langle i|_A |j\rangle_A \\
&= \frac{1}{d} \sum_{i=1}^{d} \langle i|_S \widetilde{H}^\dagger \widetilde{H} |i\rangle_S \\
&= \frac{1}{d} \mathrm{Tr}\left(\widetilde{H}^\dagger \widetilde{H}\right) \\
&= \frac{1}{d} \left\|\widetilde{H}\right\|_F^2 \\
&\geq \frac{\left\|\widetilde{H}\right\|_2^2}{d}.
\end{aligned} \tag{181}$$

Using this, we see that

$$\begin{aligned}
\mathbf{Pr}(0) &\geq \frac{1}{2} \left(\frac{\left\|\widetilde{H}\right\|_2^2}{d\Delta^2} + 1 \right) \\
&= \frac{2\left\|\widetilde{H}t\right\|_2^2}{d\pi^2} + \frac{1}{2} \geq \frac{1}{2}.
\end{aligned} \tag{182}$$

where in the second line we used $\Delta = \frac{\pi}{2t}$. Similarly, using the fact that $\left\|\tilde{H}\right\|_F \leq \sqrt{r}\left\|\tilde{H}\right\|_2$ we have that

$$\begin{aligned}\mathbf{Pr}(0) &\leq \frac{1}{2}\left(\frac{r\left\|\widetilde{H}\right\|_2^2}{d\Delta^2}+1\right)\\ &= \frac{2r\left\|\widetilde{H}t\right\|_2^2}{d\pi^2}+\frac{1}{2}.\end{aligned} \tag{183}$$

From Lemma 18 we inherit the following upper bound on the evolution time:

$$t \leq \frac{1}{2\left\|H\right\|}, \tag{184}$$

which means that if the block-encoded Hamiltonian had no error (i.e. $\widetilde{H} = H$) the success probability would certainly be less than 1. However, we must take into account the error from the block-encoding procedure. Thankfully, the result of Lemma 18 guarantees that

$$\left\|Ht - \widetilde{H}t\right\| \leq \epsilon_b, \tag{185}$$

for $\epsilon_b \in (0, \frac{1}{2})$. Therefore, the success probability remains valid as it can never be more than 1 or less than $\frac{1}{2}$.

B.6 Proof of Equation (58)

Since we have an upper bound on N_s, the number of pseudo-Choi states required to solve the Block-Encoded Hamiltonian Learning Problem 22, we would now like to know how many queries to U_{block} will suffice to generate N_s pseudo-Choi states with high probability. An upper bound on this value is given by equation (58), which we now prove.

Proof of equation (58). Consider the following Chernoff bound, from [40]:

Lemma 41. *Let $X_1, ..., X_n \in [0,1]$ be independent random variables with $\mathbb{E}[X_i] = p_i$, let $X = \sum_{i=1}^n X_i$, and let $\mu = \mathbb{E}[X] = \sum_{i=1}^N X_i p_i$. For $\beta \in (0,1)$ we have*

$$\mathbf{Pr}(X \leq (1-\beta)\mu) \leq e^{-\frac{\beta^2\mu}{2}} \tag{186}$$

Now, let X be the number of pseudo-Choi states produced after $\widetilde{N}$ queries to U_{block}. We want X to be greater than or equal to N_s, so we seek an upper bound on $\mathbf{Pr}(X \leq N_s)$. Since the success probability of obtaining each pseudo-Choi state from one query to the block-encoding is $\frac{\gamma^2}{2}$ (see Appendix B.5), we have $\mu = \frac{\widetilde{N}\gamma^2}{2}$.

To use the Chernoff bound 41, we therefore choose

$$\beta = 1 - \frac{2N_s}{\gamma^2\widetilde{N}} \tag{187}$$

Note that since β must be greater than zero to apply the Chernoff bound, we require $\widetilde{N} \geq \frac{2N_s}{\gamma^2}$. By the Chernoff bound we then have:

$$\mathbf{Pr}(X \leq N_s) \leq e^{-\frac{\widetilde{N}\gamma^2}{4}\left(1-\frac{4N_s}{\gamma^2\widetilde{N}}+\frac{4N_s^2}{\gamma^4\widetilde{N}^2}\right)} \tag{188}$$

$$= e^{\left(-\frac{\widetilde{N}\gamma^2}{4}+N_s-\frac{N_s^2}{\gamma^2\widetilde{N}}\right)} \tag{189}$$

$$\leq e^{\left(-\frac{\widetilde{N}\gamma^2}{4}+N_s\right)} \tag{190}$$

We want our failure probability to be at most δ_{N_s}, so we let

$$\delta_{N_s} \equiv e^{\left(-\frac{\widetilde{N}\gamma^2}{4}+N_s\right)}, \tag{191}$$

from which we see that

$$\ln\left(\frac{1}{\delta_{N_s}}\right) = \frac{\widetilde{N}\gamma^2}{4} - N_s \tag{192}$$

Finally, solving for $\widetilde{N}$ gives

$$\widetilde{N} = 4\frac{\ln\left(\frac{1}{\delta_{N_s}}\right)}{\gamma^2} + 4\frac{N_s}{\gamma^2} \tag{193}$$

In general, we assume N_s will be much larger than $\ln\left(\frac{1}{\delta_{N_s}}\right)$. Therefore, to generate at least N_s pseudo-Choi states with probability at least $1-\delta_{N_s}$, the number of queries to U_{block} needed is

$$\widetilde{N} \in \mathcal{O}\left(\frac{N_s}{\gamma^2}\right) \tag{194}$$

□