Learning Quantum Processes and Hamiltonians via the Pauli Transfer Matrix

MATTHIAS C. CARO, Freie Universität Berlin, Berlin, Germany and California Institute of Technology, Pasadena, United States

---

Learning about physical systems from quantum-enhanced experiments can outperform learning from experiments in which only classical memory and processing are available. Whereas quantum advantages have been established for state learning, quantum process learning is less understood. We establish an exponential quantum advantage for learning an unknown n-qubit quantum process $\mathcal{N}$. We show that a quantum memory allows to efficiently solve the following tasks: (a) learning the Pauli transfer matrix (PTM) of an arbitrary $\mathcal{N}$, (b) predicting expectation values of Pauli-sparse observables measured on the output of an arbitrary $\mathcal{N}$ upon input of a Pauli-sparse state, and (c) predicting expectation values of arbitrary observables measured on the output of an unknown $\mathcal{N}$ with sparse PTM upon input of an arbitrary state. With quantum memory, these tasks can be solved using linearly-in-n many copies of the Choi state of $\mathcal{N}$. In contrast, any learner without quantum memory requires exponentially-in-n many queries, even when using adaptively designed experiments. In proving this separation, we extend existing shadow tomography bounds from states to channels. Moreover, we combine PTM learning with polynomial interpolation to learn arbitrary Hamiltonians from short-time dynamics. Our results highlight the power of quantum-enhanced experiments for learning highly complex quantum dynamics.

CCS Concepts: • **Theory of computation** → **Quantum computation theory**; *Machine learning theory*; • **Mathematics of computing** → *Probability and statistics;*

Additional Key Words and Phrases: Quantum channel learning, Hamiltonian learning, shadow tomography, quantum advantage, sample complexity upper bounds, query complexity lower bounds

ACM Reference Format:
Matthias C. Caro. 2024. Learning Quantum Processes and Hamiltonians via the Pauli Transfer Matrix. *ACM Trans. Quantum Comput.* 5, 2, Article 14 (June 2024), 53 pages. https://doi.org/10.1145/3670418

---

1 Introduction

A complete characterization of an unknown quantum system or quantum process requires a number of copies of the unknown object that scales (at least) linearly in the dimension [42, 65, 66], and thus exponentially in the number of qubits for a multi-qubit system. Hence, there is a need for more resource-efficient alternatives that still suffice to make meaningful (while necessarily incomplete) predictions about the unknown quantum object. In the case of quantum states, the

---

This work was supported by a DAAD PRIME fellowship. The Institute for Quantum Information and Matter is an NSF Physics Frontiers Center.

Author's Contact Information: Matthias C. Caro, Freie Universität Berlin, Berlin, Germany and California Institute of Technology, Pasadena, California, United States; e-mail: matthias.caro@fu-berlin.de.

This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2024 Copyright held by the owner/author(s).

ACM 2643-6817/2024/06-ART14
https://doi.org/10.1145/3670418

frameworks of classical shadows and shadow tomography, motivated by learning-theoretic perspectives on the task of predicting properties of an unknown quantum state, have led to results on sample-efficiently predicting a plethora of physically relevant properties of a quantum system [1, 2, 6, 8, 9, 12, 13, 49, 54–56, 67, 74]. Classical shadows have the desirable feature that they do not rely on complicated quantum processing, and interesting predictions can already be made with non-adaptive single-copy measurements, which are amenable to implementation on noisy intermediate-scale (NISQ) [70] devices. However, there are limitations to what can be learned efficiently with this kind of access to an unknown quantum state or process. As shown in the work of Chen et al. [27], state shadow tomography can in general only be solved sample-efficiently in quantum-enhanced experiments, with access to a quantum memory and to quantum processing. Chen et al. [27] identified several other quantum learning problems with similarly strong query complexity separations, leading to a quantum advantage that becomes realizable on current quantum devices [51].

Whereas state shadow tomography succeeds without any assumptions on the unknown state to be learned and on the observables of interest, we cannot hope to query-efficiently and accurately predict, without prior assumptions, expectation values of arbitrary measurements performed on the output of an arbitrary quantum channel upon input of an arbitrary state. This can be seen by embedding the problem of learning an arbitrary Boolean function into the task of learning an arbitrary quantum channel (as discussed in other works [24, 50]). Nevertheless, results about learning quantum processes are beginning to emerge [18, 21, 29, 32, 35, 36, 52, 53, 60, 61], with different restrictions on the channel to be learned or on the input states and output observables for which expectation values are to be predicted.

Despite the insights of these prior works, the landscape of quantum channel learning is still largely unexplored, especially for general channels with potentially exponential complexity. In particular, while we now know an exponential quantum advantage of quantum-enhanced over conventional experiments in shadow tomography of static quantum systems, analogous results for the case of quantum evolutions are lacking. A central challenge in understanding the potential for a quantum advantage here is that conventional experiments for learning an unknown quantum channel can involve intricate adaptive strategies for choosing both the input states and the measurements to be performed. Thus, tightly bounding the capabilities of such strategies for channel learning has so far been out of reach. This naturally leads us to study the importance of quantum-enhanced experiments for tasks of channel learning, with a focus on their potential to outperform their adaptive classical counterparts. Namely, we investigate the following two questions:

(1) What can we learn about a quantum channel without prior assumptions?
(2) Are quantum-enhanced experiments necessary for quantum channel learning?

1.1 Overview of the Results

We identify learning all entries of the Pauli transfer matrix (PTM) of an unknown quantum channel as a task that can be solved efficiently by a learner with quantum memory, but that requires exponentially many queries to the unknown channel if no quantum memory is used. Here, we view a quantum channel as a linear map between matrices with complex entries, and the PTM is its matrix representation w.r.t. the normalized Pauli basis, an orthonormal basis (ONB) for the space of matrices of importance in quantum computing. Moreover, we show that this quantum advantage carries over to two restricted variants of a channel shadow tomography task. Finally, we demonstrate that PTM learning can be used as a subroutine to learn an arbitrary quantum Hamiltonian from short-time dynamics. Thus, we have found one answer to Question (1): PTM and Hamiltonian learning are meaningful channel learning problems that can be solved efficiently without any prior assumptions on the object to be learned. Additionally, query-efficient PTM learning requires quantum-enhanced experiments, thus emphasizing their relevance to channel learning in answer to Question (2).

Our first main result is an exponential query complexity separation between learners with quantum memory and learners without quantum memory for the task of learning the PTM of an unknown quantum channel.

THEOREM 1.1 (PTM LEARNING WITH AND WITHOUT QUANTUM MEMORY). *There is a learning algorithm with quantum memory that uses* $\mathcal{O}(n/\varepsilon^4)$ *copies of the Choi state of an unknown* n*-qubit quantum channel* $\mathcal{N}$ *to output simultaneously* ε*-accurate estimates for all the* 16^n *entries* $\frac{1}{2^n}\operatorname{tr}[\sigma_A \mathcal{N}(\sigma_B)]$, $A, B \in \{0,1,2,3\}^n$, *of its PTM, with high success probability. In contrast, any learning algorithm without quantum memory, even when allowed to access the unknown channel on subsystems of adaptively chosen input states and perform adaptively chosen output measurements, has to query the unknown channel* $\Omega(4^n/\varepsilon^2)$ *times to achieve the same guarantee.*

Here, the lower bound for learners without quantum memory holds even if the unknown channel is promised in advance to be doubly-stochastic, entanglement-breaking, and to have a sparse PTM. We demonstrate that this query complexity separation carries over to the tasks (a) of predicting arbitrary expectation values of the form $\operatorname{tr}[O\mathcal{N}(\rho)]$, with O a bounded Pauli-sparse observable, ρ a Pauli-sparse quantum state, and $\mathcal{N}$ an arbitrary quantum channel, and (b) of predicting arbitrary expectation values of the form $\operatorname{tr}[O\mathcal{N}(\rho)]$, with O an arbitrary bounded observable, ρ an arbitrary quantum state, and $\mathcal{N}$ a quantum channel with sparse PTM. In fact, for task (a), we give a learning algorithm with quantum memory that query-efficiently builds a classical memory-efficient representation $\hat{\mathcal{N}}$ of $\mathcal{N}$ from which any M such expectation values can be estimated query- and time-efficiently.

COROLLARY 1.2 (PREDICTING PAULI-SPARSE EXPECTATION VALUES FOR ARBITRARY CHANNELS). *There is a learning algorithm with quantum memory that uses* $\mathcal{O}(n/\varepsilon^4)$ *copies of the Choi state of an unknown* n*-qubit quantum channel* $\mathcal{N}$ *to build a classical representation* $\hat{\mathcal{N}}$ *of* $\mathcal{N}$ *from which any* M *expectation values of the form* $\operatorname{tr}[O\mathcal{N}(\rho)]$, *with* O *a bounded Pauli-sparse observable and* ρ *a Pauli-sparse quantum state, can be predicted to accuracy* ε, *with high success probability. The classical representation consists of* $\mathcal{O}(n^2/\varepsilon^4)$ *real numbers stored in classical memory, and the expectation values are predicted using* $\mathcal{O}(\log(M)/\varepsilon^2)$ *additional Choi state copies and using classical computation time* $\mathcal{O}(Mn^2/\varepsilon^4)$.

Notably, these efficiency guarantees do not require any assumptions on the unknown quantum channel $\mathcal{N}$, so this channel may be a quantum process of arbitrary complexity. In Corollary 1.2, we implicitly assumed the observables and states to have $\mathcal{O}(1)$-sparse Pauli expansions. However, as our guarantees scale polynomially with the sparsity parameters, we still get an efficient prediction protocol if we allow for polynomially-in-n Pauli-sparse observables and states. This can then be viewed as a restricted generalization of shadow tomography from states to channels. While we formulate Theorem 1.1 and Corollary 1.2 for the PTM, we also give similar protocols with polynomial-in-n copy complexity for transfer matrices w.r.t. general (appropriately normalized) unitary orthonormal bases and for predicting expectation values with states and observables having a sparse expansion in such a basis, again for an arbitrary unknown channel.

As an application of our protocol for learning the PTM with quantum memory, in our second main result we show how this can be combined with polynomial interpolation for derivative estimation to efficiently learn the Pauli coefficients of an unknown Hamiltonian H. Here, we assume query access to the unitary that implements time evolution along H for (different) short times, all on the order of $\mathcal{O}(1/\|H\|)$.

THEOREM 1.3 (LEARNING ARBITRARY HAMILTONIANS WITH QUANTUM MEMORY). *There is a learning algorithm with quantum memory that uses* $\tilde{O}\left({}^{n\|H\|^4}/_{\varepsilon^4}\right)$ *parallel queries to* $O({}^{1}/_{\|H\|})$*-time evolutions under an unknown n-qubit Hamiltonian H with Pauli expansion* $H = \sum_{A\in\{0,1,2,3\}^n} \alpha(A)\sigma_A$ *with a total evolution time of* $\tilde{O}\left({}^{n\|H\|^3}/_{\varepsilon^4}\right)$ *to estimate all the* $4^n - 1$ *non-trivial Pauli coefficients* $\alpha(A)$, $A \in \{0,1,2,3\}^n \setminus \{0^n\}$, *of H to accuracy* ε, *with high success probability.*

If the goal is to estimate only M Pauli coefficients of H, then the estimation procedure uses classical computation time scaling linearly in M and polynomially in n and ${}^{\|H\|}/_{\varepsilon}$. Moreover, the required classical memory also scales polynomially in n and ${}^{\|H\|}/_{\varepsilon}$. Similarly to how Theorem 1.1 does not require any assumptions on the unknown quantum channel, Theorem 1.3 works query-efficiently for arbitrary Hamiltonians, assuming only a polynomial operator norm bound to be guaranteed in advance. In particular, the unknown Hamiltonian can contain non-local all-to-all interaction terms.

1.2 Techniques and Proof Overview

Upper Bounds for Learning with Quantum Memory. To achieve the complexity upper bounds in Theorem 1.1 and Corollary 1.2, we rephrase the task of PTM learning as a shadow tomography task for the Choi state of the unknown quantum channel. Here, the (normalized) Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ of an n-qubit channel $\mathcal{N}$ is obtained by applying $\mathcal{N}$ to one-half of a maximally entangled state: $\frac{1}{2^n}\Gamma^{\mathcal{N}} = (\mathrm{id}\otimes\mathcal{N})(|\Omega\rangle\langle\Omega|)$, with $|\Omega\rangle$ maximally entangled. Concretely, we observe that PTM entries can be rewritten in terms of expectation values of Pauli observables on the Choi state as $\frac{1}{2^n}\operatorname{tr}[\sigma_A\mathcal{N}(\sigma_B)] = \operatorname{tr}[(\sigma_B^\top\otimes\sigma_A)\frac{1}{2^n}\Gamma^{\mathcal{N}}]$. This now allows us to appeal to an existing procedure for state shadow tomography with Pauli observables from Huang et al. [55], which is based on 2-copy Bell measurements.

We point out that channel learning tasks cannot generically be reduced to state learning tasks in a query-efficient manner. One important reason for this is the dimension factor of 2^n incurred when translating from Choi state expectation values to input-output expectation values of the associated channel. However, with our chosen learning task, the normalization of the PTM entries exactly cancels the dimension factor. This allows us to rely on Choi expectation value estimates without any loss in accuracy so that we inherit the good complexity scaling of state tomography.

We can go from estimates for PTM entries to expectation value estimates under Pauli-sparsity assumptions on either the state and observable or on the channel by expanding the respective objects in the Pauli ONB. Here, the sparsity assumption guarantees that the transfer matrix (TM) learning guarantee, which can be viewed as achieving good ∞-norm error, also leads to a good error for other p-norms. To obtain the analogous results (albeit without bounds on the classical memory and computation time) for more general orthonormal bases and associated transfer matrices as well as for suitably sparse objects, we follow the same line of reasoning but rely on the general state shadow tomography results in other works [2, 9]. Consequently, any improvements to state shadow tomography will also directly improve our results.

Lower Bounds for Learning without Quantum Memory. For our exponential query complexity lower bounds, we extend a technique introduced in the work of Chen et al. [27] to prove query complexity lower bounds for state shadow tomography without quantum memory to a channel learning task of shadow tomography type. Our proof consists of three steps. First, following the framework proposed by Chen et al. [27], we model an adaptive learning procedure that uses only a single copy of an unknown quantum channel at a time in terms of a so-called learning tree. Here, branchings in the tree correspond to different adaptive choices of the learner, depending on observed measurement outcomes. Second, we consider the distributions over the leaves of the learning tree that arise when the learner acts on different (from their perspective unknown) channels, and we upper bound how distinguishable those distributions are. Here, distinguishability is to be understood in terms of uniform one-sided likelihood ratio lower bounds for the leaf probability distributions. Third, via Le Cam's two-point method, such likelihood bounds gives rise to query complexity lower bounds for a corresponding distinguishing task. We then argue that any learner would in particular be able to solve that distinguishing task, thus inheriting the lower bound.

Even with the general learning tree framework in place, analyzing the distinguishability of the induced distributions over learning trees remains a formidable technical challenge, which has so far not been resolved for general quantum channels. For state shadow tomography without quantum memory, Chen et al. [27] were able to overcome this technical hurdle and to establish exponential query complexity lower bounds. Our first observation in trying to instantiate the learning tree framework for lower bounds in PTM learning is that a small modification to the reasoning of Chen et al. [27] already suffices to obtain analogous lower bounds for predicting PTM entries of an unknown quantum channel if the learner only has access to copies of its Choi state. Concretely, we obtain the lower bound via the intermediate task of distinguishing random Choi states of the form $\frac{1}{2^n}(\mathbb{1} \pm 3\varepsilon P)$, where $\pm P$ is a signed $(2n)$-qubit Pauli, from the maximally mixed state. This modification involves determining the second moment for a certain sub-ensemble of all n-qubit Paulis. However, the $\Omega\left({}^{4^n}/_{\varepsilon^2}\right)$ query complexity lower bound obtained in this way does not yet apply to general adaptive quantum channel learning procedures without quantum memory, which can also adaptively choose inputs to the unknown channel.

Our main technical contribution is to prove the one-sided likelihood bounds required for Le Cam's two-point method even in this fully adaptive setting. To achieve this, we identify a quantity that, given a set of PTM entries, characterizes the hardness of simultaneously predicting those entries without using a quantum memory. We analyze this quantity for a suitable set of TM entries and for a suitable ensemble of unknown quantum channels, whose Choi states are either $\frac{1}{2^n}(\mathbb{1} \pm 3\varepsilon P)$, with a random signed Pauli string $\pm P$, or maximally mixed. Our analysis, which combines a second moment calculation for Pauli sub-ensembles with matrix-analytic tools, provides a new way of utilizing the framework of Chen et al. [27] and establishes the $\Omega\left({}^{4^n}/_{\varepsilon^2}\right)$ query complexity lower bound for general channel learning procedures without quantum memory. This analysis may serve as a template for proving query complexity lower bounds for further channel learning tasks without access to quantum-enhanced experiments.

Hamiltonian Learning from PTM Learning. We prove that our PTM learning protocol gives rise to the Hamiltonian learning algorithm of Theorem 1.3 in three steps. First, reinterpreting observations made in other works [40, 43, 75], we note that any single Pauli expansion coefficient $\alpha(A)$ of the unknown Hamiltonian $H = \sum_{A\in\{0,1,2,3\}^n} \alpha(A)\sigma_A$ can be related to the first-order time derivative of a specific time-dependent PTM entry of the associated unitary time evolution, evaluated at time 0. Importantly, this works for Pauli coefficients of arbitrarily high weight. Second, relying on tools from polynomial approximation [48], we can estimate the first-order derivative of a function via that of its Chebyshev interpolating polynomial. This requires us to prove bounds on higher-order derivatives of the function of interest. Thus, the third step in our proof is to control higher-order derivatives of functions of the form $t \mapsto \frac{1}{2}\operatorname{tr}\left[\sigma_B e^{-itH}\rho e^{itH}\right]$ for an arbitrary Pauli σ_B and an arbitrary state ρ. We achieve this via a rewriting in terms of iterated commutators with H. The concrete implementation of this last step deviates from prior work, which used structural assumptions on the Hamiltonian to obtain derivative bounds. In our case, since we do not make any prior assumptions on H, we work with derivative bounds that depend on the spectral norm $\|H\|$. In summary, we use our PTM learning procedure applied to the evolution at different times to produce data points for polynomial interpolation, then estimate the Pauli coefficients of the Hamiltonian via first-order time derivatives from the interpolating polynomial.

1.3 Related Work

Shadow Tomography for Quantum States. Since Aaronson [1] considered a “pretty good” version of state tomography motivated by the framework of probably approximately correct learning, learning quantum states in different models has received significant attention. For example, Aaronson et al. [3] and Chen et al. [30] investigated problems of online learning quantum states. Most closely related to our work is shadow tomography for quantum states, originally proposed by Aaronson [2] and improved upon in the work of Bădescu and O’Donnell [9] and Huang et al. In shadow tomography, the task is to predict many expectation values of an unknown quantum from few copies. Our work can be viewed as partially lifting shadow tomography from states to channels. Here, the “partially” is meant to express that, whereas shadow tomography for states allows to predict expectation values of arbitrary observables, in our channel shadow tomography we have to make some additional assumptions on the input states and output observables.

Classical Shadows for Quantum States and Quantum Processes. While shadow tomography methods are efficient in terms of sample complexity, they rely on measurements on multiple copies of the unknown state and can be inefficient in terms of classical memory and computation time. Classical shadows of quantum states, going back to other works [54, 67], constitute a more practical variant of shadow tomography, since they use single-copy measurements and come with guarantees on classical memory and computation time. However, whereas shadow tomography can predict expectation values for arbitrary observables, classical shadows only work reliably for a suitably restricted set of observables, which depends on the random unitaries used in the shadow protocol. Kunjummen et al. [60] and Levy et al. [61] lifted the classical shadow formalism from states to channels by applying the classical shadow protocol of Huang et al. [54] based on randomized local Paulis to the Choi state of an unknown channel (see also the work of França et al. [75, Appendix H]), which is similar to how we apply shadow tomography to the Choi state to learn TM entries of the channel from Choi access. As the local classical shadow protocol of Huang et al. [54] comes with prediction guarantees only for local observables, the protocol of Kunjummen et al. [60] and Levy et al.[61] can only learn local reduced density matrices of the Choi state. This is not sufficient to learn the whole TM of the unknown channel, which our quantum-enhanced protocol achieves.

Learning Quantum Channels. In the wake of the many insights into learning quantum states, also questions of learning quantum channels have begun to attract attention. For instance, there is a growing literature in variational quantum machine learning exploring how different kinds of complexity bounds for the quantum model, which plays the role of an unknown quantum channel to be learned, lead to sample complexity bounds [4, 14–17, 19–23, 33, 41, 69]. Viewed from the perspective of channel learning, these results make assumptions on the complexity of the unknown channel and use this to bound the information-theoretic complexity of the learning task. These approaches can therefore typically not be applied to arbitrary quantum channels of high complexity.

Also in the broader quantum information theory community, channel learning tasks are gaining traction. Some works [18, 32, 35] prove sample complexity guarantees for learning channels with classical input and quantum output in a probably approximately correct learning setting. When considering channels with quantum input and quantum output, the case of Pauli channels, relevant for modeling quantum noise, is already well studied: some works [36, 37, 44, 45] give efficient procedures for learning the Pauli error rates of an unknown Pauli channel. In work closely related to the present article, Chen et al. [29] proved query complexity separations for learning the Pauli eigenvalues of an unknown Pauli channel, which in our language are its diagonal PTM elements.

They distinguish between learners with quantum memory and learners restricted either by not having access to a quantum memory or only to auxiliary systems of limited size. Our query complexity lower bounds for learning without quantum memory show that the lower bounds of Chen et al. [29, Theorem 2, (i) and (ii)] can be strengthened significantly, achieving an exponential lower bound independently of the size of the allowed auxiliary system, if we go beyond Pauli channels with diagonal PTMs to general quantum channels with general PTMs.

Taking a broader perspective on channel learning, Huang et al. [53] established fundamental results about the conditions under which quantum processes can or cannot be learned from experiments with noise. And in very recent work, Huang et al. [52] extended the classical shadow formalism from states to channels, giving an efficient procedure without quantum memory for learning to predict an arbitrary unknown quantum channel w.r.t. bounded degree k-local observables when averaged over any locally flat distribution of input states. In contrast, our learning algorithm relies on a quantum memory but can predict w.r.t. general Pauli-sparse observables and worst-case w.r.t. Pauli-sparse input states.

Hamiltonian Learning. The task of learning a Hamiltonian underlying the evolution of a quantum physical evolution has seen significant progress in recent years. Some works [10, 38, 62, 72] showed that, at least in principle, generic geometrically local Hamiltonians can be learned from “little information,” such as from a single eigenstate or from few generic pairs of input and output states. Anshu et al. [7] and Rouzé and França [74] gave efficient procedures for learning a geometrically local Hamiltonian from copies of its high-temperature Gibbs state, and this was extended to more general low-intersection Hamiltonians while at the same time improving the sample and time complexity in the work of Haah et al. [43]. Some recent works [40, 75, 79, 82] have highlighted the practical importance of learning an unknown Hamiltonian from access to its dynamics, deriving provable guarantees for geometrically local or low-intersection Hamiltonians and demonstrating the applicability of their Hamiltonian learning procedures for relatively large quantum systems. Recently, Huang et al. [57] proved that even Heisenberg-limited scaling for the total evolution time is achievable in the task of learning low-intersection Hamiltonians from dynamics. Our addition to the Hamiltonian learning literature is, to the best of our knowledge, the first efficient protocol for learning an arbitrary Hamiltonian with an *a priori* unknown structure from parallel access to its short-time dynamics.

Learning from Quantum Experiments. Aharonov et al. [5] and Huang et al. [55] established exponential query complexity separations between certain tasks of learning from quantum experiments with and without quantum memory. Additionally, Huang et al. [55] demonstrated that a class of average-case channel learning tasks does not admit a large quantum query complexity advantage, thereby emphasizing the importance of worst-case considerations. Building on these works, Chen et al. [27] introduced the learning tree framework as a powerful tool for proving query complexity lower bounds for learning without quantum memory, or with limited quantum memory [26]. This mathematical framework has been used by Huang et al. [51] to prove an experimentally demonstrable quantum advantage in learning from quantum physics experiments when using a quantum memory. Moreover, with Chen et al. [28], the learning tree formalism has recently found an application in a complexity-theoretic study of the capabilities of NISQ devices. We use the framework of Chen et al. [27] and develop new technical tools to extend their proof strategy and to derive query complexity lower bounds for PTM learning without quantum memory, thereby extending their lower bounds for state shadow tomography to (Pauli-sparse) channel shadow tomography. Whereas Aharonov et al. [5] and Chen et al. [27] considered “channel distinguishing” tasks, we focus more on “channel learning” tasks. (Note, however, that our query complexity lower bounds are proved via a channel distinguishing task.) And whereas Huang et al. [51, Theorem 3] established a quantum-versus-classical query complexity separation for learning a polynomial-time quantum channel on average over a distribution of states, we consider learning arbitrarily complex quantum channels in a worst-case framework over (certain) input states.

1.4 Summary and Directions for Future Work

In this work, we have given a provably sample-efficient procedure for learning transfer matrices with quantum memory from copies of the Choi state of an arbitrary unknown quantum channel. In the case of the PTM, we obtained efficiency guarantees for classical memory and classical computation time in addition to the copy complexity bounds. This allows to efficiently learn an arbitrary unknown Hamiltonian from dynamics. Moreover, we demonstrated that (Pauli) TM learning can be used as a subroutine for efficiently learning to predict expectation values for (Pauli-) sparse input states and output observables, a task of channel shadow tomography type. In particular, solving these learning problems remains information-theoretically feasible even for arbitrarily complex quantum processes, which is in contrast to the idea of Occam's razor embodied in the established classical statistical learning theory literature based on complexity measures [11, 34, 59, 68, 76, 77].

While the tasks of TM learning and sparse expectation value prediction can be solved information-theoretically efficiently with quantum memory, we proved that no learner without quantum memory can succeed at these tasks with a subexponential number of queries to the unknown quantum channel. Thus, these channel learning tasks show an exponential query complexity separation between learners with and without quantum memory. Our $\Omega(4^n/\varepsilon^2)$ query complexity lower bound for learners without quantum memory is significantly stronger than the $\Omega(2^n/\varepsilon^2)$ obtained by directly applying the existing state shadow tomography lower bound to state preparation channels, showing that the quantum advantage over conventional experiments is even more distinctive in the channel learning case than in the state learning case.

We conclude by outlining some open problems raised by our work. Here, we begin by recalling the two questions underlying this work and aspects thereof that remain open:

- TM learning and Hamiltonian learning are efficiently solvable channel learning tasks without structural assumptions, thus serving as positive examples in answer to Question (1). Can we embed them into a broader class of feasible learning problems with arbitrary unknown quantum processes?
- Our exponential query complexity separation between learners with and without quantum memory answers Question (2) for TM learning. However, it breaks down for certain variants of the task (compare Remarks 5.4, 5.5, 5.9, and 5.10). Can we develop a more general understanding of the conditions conducive to a quantum query complexity advantage in channel learning?

Finally, we mention three questions regarding potential improvements and extensions of our results:

- As our TM learning protocols with quantum memory require only Choi access to the unknown channel, it is natural to ask whether the query complexity can be further improved by allowing for sequential channel access. However, the linear-in-n dependence cannot be improved, even for Pauli channels [29, Theorem 2, (iv)]. Does sequential access allow to improve upon the $\frac{1}{\varepsilon^4}$-scaling?
- Our Hamiltonian learning complexity bounds in Theorem 6.3 exhibit a quartic scaling with $1/\varepsilon$ and a linear or quadratic scaling with n. Can these dependencies be improved to come closer to the guarantees found for restricted Hamiltonians in prior works? Are quantum-enhanced experiments even necessary for learning arbitrary Hamiltonians from dynamics?

— Using the work of França et al. [75, Appendix A], we can employ our PTM learning procedure to learn Lindblad generators with arbitrary Hamiltonian part and arbitrary single-site dissipation terms. Does this extend to more general Lindblad generators with arbitrary dissipative part?

Answering these questions will help understand the role of quantum memory in learning processes in quantum physics from a query complexity perspective motivated by theoretical computer science. Thereby, we can gain a better grasp of quantum learning problems solvable on near-term quantum hardware and of the experimental capabilities that need to be developed for addressing the most challenging quantum learning tasks.

1.5 Structure of the Article

The rest of the article is structured as follows. Section 2 introduces basic notions from quantum information as well as different models of learning from an unknown quantum channel. (See Appendix A for two additional learning models.) In Section 3, we explain how approximations on the level of the TM of a channel translate to expectation value estimates under sparsity assumptions. Section 4 contains our efficient procedures with quantum memory for learning quantum channels via their PTM, with a corresponding query complexity lower bound presented in Appendix E. (More general transfer matrices are considered in Appendix B.) This is then contrasted with exponential query complexity lower bounds for learning without quantum memory in Section 5, the proofs for which can be found in Appendix D and are based on the learning tree formalism [27] shortly reviewed in Appendix C. To showcase an application of TM learning, we demonstrate its applicability to Hamiltonian learning in Section 6, with the main proof again deferred to Appendix D.

2 Preliminaries

2.1 Quantum Channels, the Choi-Jamiolkowski Isomorphism, and the TM Representation

Here, we review basic notions from quantum information and computation that will appear throughout the rest of the article. The presentation here is brief, thus we recommend textbooks (e.g., [46, 64, 78]) or lecture notes (e.g., [71, 81]).

Throughout the article, we work with finite-dimensional Hilbert spaces $\mathbb{C}^d$, with dimension $d \in \mathbb{N}_{\geq 1} = \{1, 2, \ldots\}$. We employ Dirac bra-ket notation to denote elements of $\mathbb{C}^d$ by kets and elements of the dual space $(\mathbb{C}^d)^*$ by bras. We use $\mathcal{B}(\mathbb{C}^d)$ to denote the set of bounded linear operators on $\mathbb{C}^d$. Two norms on $\mathcal{B}(\mathbb{C}^d)$ will be relevant for our purposes. We denote the operator norm (w.r.t. the Euclidean norm on $\mathbb{C}^d$), also called the *spectral norm* or *Schatten* ∞-*norm*, of $X \in \mathcal{B}(\mathbb{C}^d)$ by $\|X\| = \sup_{|\psi\rangle \in \mathbb{C}^d : \langle\psi|\psi\rangle = 1} \langle\psi| X |\psi\rangle$. We denote the norm induced by the Hilbert-Schmidt inner product, also called the *Frobenius norm* or *Schatten* 2-*norm*, of $X \in \mathcal{B}(\mathbb{C}^d)$ by $\|X\|_2 = \sqrt{\mathrm{tr}\left[X^\dagger X\right]}$.

The set of d-dimensional density matrices, the mathematical description for quantum states, is denoted by $\mathcal{S}(\mathbb{C}^d) = \left\{\rho \in \mathcal{B}(\mathbb{C}^d) \mid \rho \geq 0 \ \wedge \ \mathrm{tr}[\rho] = 1\right\}$. The rank-1 projections in $\mathcal{S}(\mathbb{C}^d)$ are called *pure states*. We typically identify a pure state $|\psi\rangle\langle\psi| \in \mathcal{S}(\mathbb{C}^d)$ with the corresponding normalized vector $|\psi\rangle \in \mathbb{C}^d$. If $\rho \in \mathcal{S}(\mathbb{C}^d \otimes \mathbb{C}^{d'})$ is a state on a composite system, we say that ρ is *separable* if it can be written as a convex combination of tensor products of states on the two tensor factors, otherwise we call ρ *entangled*.

We describe measurements in terms of positive operator-valued measures (POVMs). Here, a d-dimensional M-outcome POVM is a set $\{E_i\}_{i=1}^M$ of bounded linear operators $E_i \in \mathcal{B}(\mathbb{C}^d)$ satisfying

$0 \leq E_i \leq \mathbb{1}_d$ for all $1 \leq i \leq M$ as well as $\sum_{i=1}^{M} E_i = \mathbb{1}_d$. According to Born's rule, the probability of observing outcome i when measuring the POVM $\{E_i\}_{i=1}^{M}$ on the state ρ is given by $\operatorname{tr}[E_i \rho]$.

We now turn our attention to evolutions of quantum systems. These are described mathematically by quantum channels.

Definition 2.1 (Quantum Channels). A linear superoperator $\mathcal{N} : \mathcal{B}(\mathbb{C}^{d_{\text{in}}}) \longrightarrow \mathcal{B}(\mathbb{C}^{d_{\text{out}}})$ is a quantum channel (in the Schrödinger picture) if $\mathcal{N}$ is completely positive (i.e., $\text{id}_{\text{aux}} \otimes \mathcal{N}$ is positivity-preserving for any auxiliary system) and trace-preserving. If $\mathcal{N}$ is also unital (i.e., satisfies $\mathcal{N}(\mathbb{1}_{d_{\text{in}}}) = \mathbb{1}_{d_{\text{out}}}$), then we call $\mathcal{N}$ a *doubly-stochastic quantum channel* (see [63]).

We call a quantum channel $\mathcal{N}$ *entanglement-breaking* if $(\text{id}_{\text{aux}} \otimes \mathcal{N})(\rho)$ is separable for any state $\rho \in \mathcal{S}(\mathbb{C}^{d_{\text{aux}}} \otimes \mathbb{C}^{d_{\text{in}}})$ (see [47]). Our main focus will be on n-qubit channels, for which $d_{\text{in}} = d_{\text{out}} = 2^n$.

Linear superoperators are isomorphically related to bounded linear operators on a tensor product space. A well-known isomorphism particularly useful for quantum information purposes is given in the following.

Proposition 2.2 (Choi-Jamiolkowski Isomorphism [31, 58]). *Given an ONB $\mathcal{B} = \{|i\rangle\}_{i=1}^{d_{\text{in}}} \subseteq \mathbb{C}^{d_{\text{in}}}$, the linear map that takes as input a linear superoperator $\mathcal{N} : \mathcal{B}(\mathbb{C}^{d_{\text{in}}}) \longrightarrow \mathcal{B}(\mathbb{C}^{d_{\text{out}}})$ and maps it to the bounded operator*

$$\frac{1}{d_{\text{in}}} \Gamma^{\mathcal{N}} := (\text{id}_{\text{in}} \otimes \mathcal{N})(\Omega) \in \mathcal{B}(\mathbb{C}^{d_{\text{in}}} \otimes \mathbb{C}^{d_{\text{out}}}), \tag{2.1}$$

where $\Omega = |\Omega\rangle\langle\Omega|$ with $|\Omega\rangle = \frac{1}{\sqrt{d_{\text{in}}}} \sum_{i=1}^{d_{\text{in}}} |ii\rangle$ is the canonical maximally entangled state w.r.t. $\mathcal{B}$, is a linear isomorphism and is called the Choi-Jamiolkowski isomorphism (w.r.t. the basis $\mathcal{B}$).

Given the Choi state $\frac{1}{d_{\text{in}}} \Gamma^{\mathcal{N}}$ of a quantum channel, we can describe the action of $\mathcal{N}$ on an input operator $X \in \mathcal{B}(\mathbb{C}^{d_{\text{in}}})$ via $\mathcal{N}(X) = d_{\text{in}} \operatorname{tr}_{\text{in}}[(X_{\text{in}}^{\top} \otimes \mathbb{1}_{\text{out}}) \frac{1}{d_{\text{in}}} \Gamma^{\mathcal{N}}]$. More generally, the action of $\text{id}_{\text{aux}} \otimes \mathcal{N}$ on a bipartite input operator $X \in \mathcal{B}(\mathbb{C}^{d_{\text{aux}}} \otimes \mathbb{C}^{d_{\text{in}}})$ is given by $(\text{id}_{\text{aux}} \otimes \mathcal{N})(X) = d_{\text{in}} \operatorname{tr}_{\text{in}}[(X^{\top_{\text{in}}} \otimes \mathbb{1}_{\text{out}})(\mathbb{1}_{\text{aux}} \otimes \frac{1}{d_{\text{in}}} \Gamma^{\mathcal{N}})]$.

From now on, we always imagine that a basis $\mathcal{B}$ of $\mathbb{C}^{d_{\text{in}}}$ has been fixed, so we will not mention the choice of basis for the canonical maximally entangled state, for the Choi-Jamiolkowski isomorphism, or for matrix transpositions explicitly anymore. We collect useful properties of the Choi-Jamiolkowski isomorphism in the next proposition.

Proposition 2.3 (Properties of the Choi-Jamiolkowski Isomorphism (see, e.g., [81, Proposition 2.1])). *Let $\mathcal{N} : \mathcal{B}(\mathbb{C}^{d_{\text{in}}}) \longrightarrow \mathcal{B}(\mathbb{C}^{d_{\text{out}}})$ be a linear superoperator and let $\frac{1}{2^n} \Gamma^{\mathcal{N}}$ be its normalized Choi-Jamiolkowski operator as in Equation (2.1). Then:*

(1) $\mathcal{N}$ is Hermiticity-preserving if and only if $\frac{1}{2^n} \Gamma^{\mathcal{N}}$ is Hermitian.
(2) $\mathcal{N}$ is completely positive if and only if $\frac{1}{2^n} \Gamma^{\mathcal{N}}$ is positive semidefinite.
(3) $\mathcal{N}$ is trace-preserving if and only if $\operatorname{tr}_{\text{out}}[\frac{1}{2^n} \Gamma^{\mathcal{N}}] = \frac{1}{2^{d_{\text{in}}}} \mathbb{1}_{\text{in}}$.
(4) $\mathcal{N}$ is unital if and only if $\operatorname{tr}_{\text{in}}[\frac{1}{2^n} \Gamma^{\mathcal{N}}] = \frac{1}{2^{d_{\text{out}}}} \mathbb{1}_{\text{out}}$.

In particular, we see that $\mathcal{N}$ is a quantum channel if and only if $\frac{1}{2^n} \Gamma^{\mathcal{N}}$ is a quantum state with a maximally mixed first marginal. Moreover, recalling that $\mathcal{N}$ is entanglement-breaking if and only if $\frac{1}{2^n} \Gamma^{\mathcal{N}}$ is separable [47], we see that $\mathcal{N}$ is a doubly-stochastic and entanglement-breaking quantum channel if and only if $\frac{1}{2^n} \Gamma^{\mathcal{N}}$ is a separable quantum state whose first and second marginal both are maximally mixed.

The Choi-Jamiolkowski isomorphism already provides us with a useful representation for a quantum channel. The next definition introduces an alternative representation obtained by choosing an ONB for the space of bounded operators and then considering the matrix representation of a quantum channel w.r.t. that basis. For convenience, we focus on the case of multi-qubit channels.

Definition 2.4 (General Transfer Matrices). Let $\mathcal{Q} = \{Q_i\}_{i=1}^{4^n}$ be an ONB for $\mathcal{B}((\mathbb{C}^2)^{\otimes n})$ w.r.t. the Hilbert-Schmidt inner product, consisting of Hermitian bounded linear operators $Q_i = Q_i^\dagger \in \mathcal{B}((\mathbb{C}^2)^{\otimes n})$. Let $\mathcal{N} : \mathcal{B}(\mathbb{C}^{2^n}) \to \mathcal{B}(\mathbb{C}^{2^n})$ be an n-qubit quantum channel. The *TM* of $\mathcal{N}$ w.r.t. $\mathcal{Q}$ is the matrix $R_{\mathcal{N}}^{\mathcal{Q}} \in \mathbb{C}^{4^n \times 4^n}$ with entries

$$\left(R_{\mathcal{N}}^{\mathcal{Q}}\right)_{i,j} := \operatorname{tr}\left[Q_i \mathcal{N}(Q_j)\right], \quad \forall 1 \leq i, j \leq 4^n. \tag{2.2}$$

As we are interested in channels acting on systems of multiple qubits, we can consider a relevant special case of Definition 2.5 by choosing the overall ONB to consist of tensor products of elements of a single-qubit ONB.

Definition 2.5 (Local Transfer Matrices). Let $\mathcal{Q} = \{Q_i\}_{i=1}^{4}$ be a local ONB for $\mathcal{B}(\mathbb{C}^2)$ w.r.t. the Hilbert-Schmidt inner product, consisting of Hermitian bounded linear operators $Q_i = Q_i^\dagger \in \mathcal{B}(\mathbb{C}^2)$, and assume $Q_0 = \frac{1}{\sqrt{2}}\mathbb{1}_2$. Let $\mathcal{N} : \mathcal{B}(\mathbb{C}^{2^n}) \to \mathcal{B}(\mathbb{C}^{2^n})$ be an n-qubit quantum channel. The *local TM* of $\mathcal{N}$ w.r.t. $\mathcal{Q}$ is the matrix $R_{\mathcal{N}}^{\mathcal{Q}} \in \mathbb{C}^{\{0,1,2,3\}^n \times \{0,1,2,3\}^n}$ with entries

$$\left(R_{\mathcal{N}}^{\mathcal{Q}}\right)_{A,B} := \operatorname{tr}[Q_A \mathcal{N}(Q_B)], \quad \forall A, B \in \{0,1,2,3\}^n, \tag{2.3}$$

where we used the notation

$$Q_A = \bigotimes_{k=1}^{n} Q_{A_k}, \quad \forall A = (A_1, \ldots, A_n) \in \{0,1,2,3\}^n. \tag{2.4}$$

Note that we use $A, B, C, \ldots$ to denote elements of $\{0,1,2,3\}^n$ (or sometimes of $\{0,1,2,3\}^{2n}$), which leads to operators indexed by $A, B, C, \ldots$ to express tensor products. This is not to be confused with the indexing of operators by the subsystem they act on, for which sometimes a similar notation is used. When an indexing by subsystems becomes necessary, we will do so by giving the subsystems names such as in (for "input"), out (for "output"), or aux (for "auxiliary").

We obtain the prime example for a (local) TM when considering the ONB of normalized Pauli strings.

Example 2.6 (Pauli Transfer Matrix). Consider the local ONB of normalized single-qubit Paulis, $\mathcal{P} = \{\frac{1}{\sqrt{2}}\sigma_a\}_{a \in \{0,1,2,3\}}$. For an n-qubit quantum channel $\mathcal{N} : \mathcal{B}(\mathbb{C}^{2^n}) \to \mathcal{B}(\mathbb{C}^{2^n})$, this gives rise to the PTM $R_{\mathcal{N}}^{\mathcal{P}} \in \mathbb{C}^{\{0,1,2,3\}^n \times \{0,1,2,3\}^n}$, a local TM, with entries

$$\left(R_{\mathcal{N}}^{\mathcal{P}}\right)_{A,B} := \frac{1}{2^n} \operatorname{tr}[\sigma_A \mathcal{N}(\sigma_B)], \quad \forall A, B \in \{0,1,2,3\}^n. \tag{2.5}$$

Note that for a quantum channel $\mathcal{N}$, all PTM entries are real and lie in the interval $[-1, 1]$ (compare, e.g., [39, Section 2.1.3]).

2.2 Learning Quantum Channels with and without Quantum Memory

In this section, we describe different models of learning quantum channels. These models differ in the type of access to the unknown quantum channel—for example, depending on whether the learner has a quantum memory at their disposal or whether the learner can actively choose the input states on which the unknown quantum channel is queried. We first recall two definitions from Chen et al. [25]. The first considers general adaptive procedures for learning an unknown quantum channel if the learner does not have access to a quantum memory.

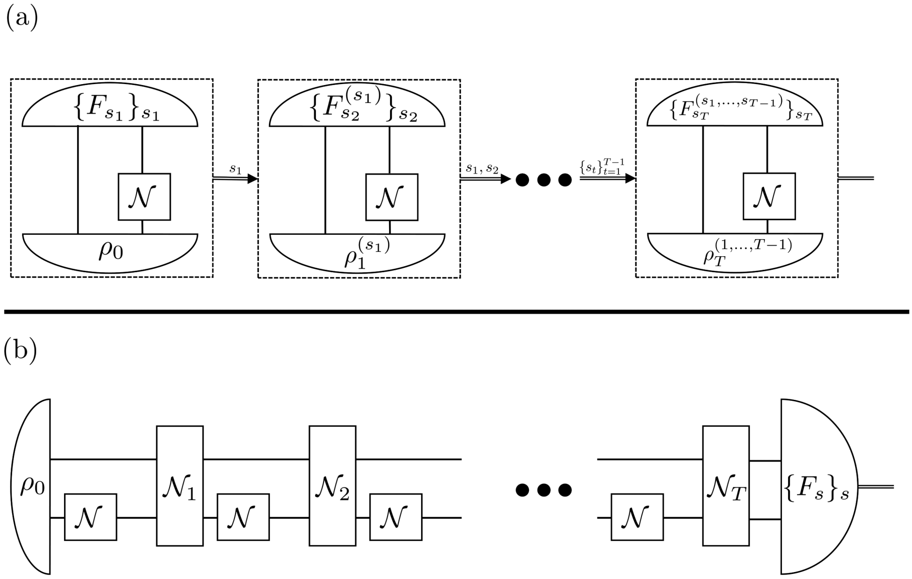

Fig. 1. Illustration of learning quantum channels from general channel access, each panel to be read from left to right. Panel (a) depicts a learner without quantum memory, and panel (b) depicts a learner with quantum memory.

Definition 2.7 (Learning Quantum Channels without Quantum Memory [25, Definition 4.17]). An algorithm for learning an unknown n-qubit quantum channel $\mathcal{N}$ without quantum memory can obtain classical data from the channel oracle by preparing an arbitrary input state $\rho \in \mathcal{B}((\mathbb{C}^2)^{\otimes n_{\text{aux}}} \otimes (\mathbb{C}^2)^{\otimes n})$, evolving under $\mathcal{N}$ to yield the output state $(\text{id}_{\text{aux}} \otimes \mathcal{N})(\rho)$, and performing an arbitrary POVM measurement $\{F_s\}_s \subseteq \mathcal{B}((\mathbb{C}^2)^{\otimes n_{\text{aux}}} \otimes (\mathbb{C}^2)^{\otimes n})$ on that output state to obtain outcome s with probability $\text{Tr}[F_s(\text{id}_{\text{aux}} \otimes \mathcal{N})(\rho)]$. Here, the selection of the input state and the POVM can be adaptive—that is, depend on previously chosen input states, previously chosen POVMs, and previously observed measurement outcomes. After T oracle accesses, the algorithm uses the entirety of the obtained measurement outcomes to predict properties of $\mathcal{N}$.

General learning procedures in which the learner does have a quantum memory at their disposal are encapsulated in the following definition.

Definition 2.8 (Learning Quantum Channels with Quantum Memory [25, Definition 4.18]). An algorithm for learning an unknown n-qubit quantum channel $\mathcal{N}$ with quantum memory can access the channel oracle as a quantum channel during a mixed-state quantum computation. The resulting state of the quantum memory after T oracle accesses is

$$\rho_T^{\mathcal{N}} = \mathcal{N}_T(\text{id}_{\text{aux},T-1} \otimes \mathcal{N}) \dots \mathcal{N}_2(\text{id}_{\text{aux},1} \otimes \mathcal{N})\mathcal{N}_1(\text{id}_{\text{aux},0} \otimes \mathcal{N})(\rho_0), \tag{2.6}$$

with some input state $\rho_0 \in \mathcal{S}((\mathbb{C}^2)^{\otimes n_{\text{aux},0}} \otimes (\mathbb{C}^2)^{\otimes n})$ and some n_{t-1}-to-n_t qubit quantum channels $\mathcal{N}_t : \mathcal{B}((\mathbb{C}^2)^{\otimes n_{t-1}}) \to \mathcal{B}((\mathbb{C}^2)^{\otimes n_t})$, for $t = 1, \dots, T$, where $n_0 = n$. After T oracle accesses, the algorithm performs a POVM $\{F_s\}_s \subseteq \mathcal{B}((\mathbb{C}^2)^{\otimes n_T})$ on the quantum memory state $\rho_T^{\mathcal{N}}$ to predict properties of $\mathcal{N}$.

Definitions 2.7 and 2.8 are illustrated in Figure 1. We will also consider variants of the previous two definitions. To obtain these variants, we suppose that the learner has access to the unknown

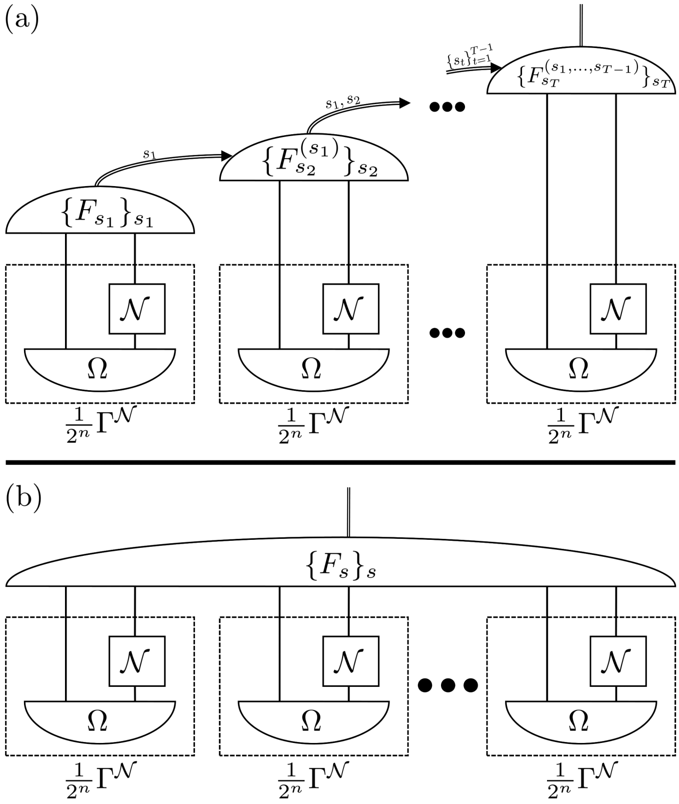
(a)
$\{s_t\}_{t=1}^{T-1}$
$\{F_{s_T}^{(s_1,\ldots,s_{T-1})}\}_{s_T}$
s_1, s_2
s_1
$\{F_{s_2}^{(s_1)}\}_{s_2}$
$\{F_{s_1}\}_{s_1}$
$\mathcal{N}$
Ω
$\frac{1}{2^n}\Gamma^{\mathcal{N}}$
(b)
$\{F_s\}_s$
$\mathcal{N}$
Ω
$\frac{1}{2^n}\Gamma^{\mathcal{N}}$

Fig. 2. Illustration of learning quantum channels from Choi access, each panel to be read from bottom to top. Panel (a) depicts a learner without quantum memory, and panel (b) depicts a learner with quantum memory.

quantum channel $\mathcal{N}$ only via its Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$. In analogy to Chen et al. [25, Definitions 4.15 and 4.16], this leads us to the following two definitions.

Definition 2.9 (Learning Quantum Channels without Quantum Memory from Choi Access). An algorithm for learning an unknown n-qubit quantum channel $\mathcal{N}$ without quantum memory from Choi access can obtain classical data from the Choi state oracle by performing arbitrary adaptively chosen POVM measurements on single copies of the Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$. In other words, for each oracle access, the algorithm can select a POVM $\{F_s\}_s \subseteq \mathcal{B}((\mathbb{C}^2)^{\otimes n} \otimes (\mathbb{C}^2)^{\otimes n})$ and obtain the classical outcome s with probability $\operatorname{tr}\left[F_s \frac{1}{2^n}\Gamma^{\mathcal{N}}\right]$. Here, the selection of the POVM can be adaptive (i.e., depend on previously chosen POVMs and previously observed measurement outcomes). After T oracle accesses, the algorithm uses the entirety of the obtained measurement outcomes to predict properties of $\mathcal{N}$.

Definition 2.10 (Learning Quantum Channels with Quantum Memory from Choi Access). An algorithm for learning an unknown n-qubit quantum channel $\mathcal{N}$ with quantum memory from Choi access can obtain copies of the Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ from the Choi state oracle and store those copies in the quantum memory. After T oracle accesses, the algorithm performs a joint POVM $\{F_s\}_s \subseteq \mathcal{B}(((\mathbb{C}^2)^{\otimes n} \otimes (\mathbb{C}^2)^{\otimes n})^{\otimes T})$ on $(\frac{1}{2^n}\Gamma^{\mathcal{N}})^{\otimes T}$ to predict properties of $\mathcal{N}$.

We illustrate learning channels from Choi access in Figure 2. Any procedure for learning a quantum channel from Choi access in particular is an instance of a general procedure for learning a quantum channel. To see this, simply use independent copies of a maximally entangled state as input in Definitions 2.7 and 2.8. However, as the implementation of an n-qubit quantum channel via teleportation of its Choi state in general has a success probability 4^{-n}, not every general procedure for learning a quantum channel may be realizable via learning from Choi access. However, under additional assumptions on shared local unitary invariances of the Choi states of the unknown channels, the success probability of implementation via teleportation can be increased (compare, e.g., [81, Section 2.1]), so reducing general channel learning to learning from Choi access may become feasible. Finally, in Appendix A, we describe a model of channel learning from parallel access, which is intermediate between Choi access and general channel access.

3 Predicting Expectation Values From TM Entries Under Sparsity Assumptions

Before establishing our results on learning transfer matrices, we first discuss how approximations to entries of the TM of a channel $\mathcal{N}$ can be used to obtain estimates for expectation values of the form $\operatorname{tr}[O\mathcal{N}(\rho)]$. Namely, we first observe the following. The ability to simultaneously approximate entries $(R^{\mathcal{Q}}_{\mathcal{N}})_{i,j}$ of the TM of a channel $\mathcal{N}$ up to accuracy ε allows to approximate the expectation value $\operatorname{tr}[O\mathcal{N}(\rho)]$ up to accuracy $\tilde{\varepsilon}$, where $\tilde{\varepsilon}/\varepsilon$ can be controlled in terms of the 1-norms of the $\mathcal{Q}$-basis coefficient vectors of O and ρ. This is particularly useful if both O and ρ have sparse $\mathcal{Q}$-basis expansions. Our second observation is similar. If the TM of $\mathcal{N}$ is sparse, then ε-accurate estimates of those entries translate to $\tilde{\varepsilon}$-accurate estimates of $\operatorname{tr}[O\mathcal{N}(\rho)]$ for any O and ρ, where $\tilde{\varepsilon}/\varepsilon$ depends on the sparsity of the TM. In summary, under sparsity assumptions, accurate estimates of TM entries give rise to accurate estimates of expectation values.

For this section, we always take $\mathcal{Q} = \{Q_i\}_{i=1}^{4^n}$ to be a Hermitian ONB of $\mathcal{B}((\mathbb{C}^2)^{\otimes n})$. Accordingly, the transfer matrices in this subsection will be as in Definition 2.4.

3.1 Sparse Input States and Sparse Output Observables

We begin by showing how estimates for the TM translate to expectation value estimates under 1-norm assumptions on the coefficient vectors of state and observable.

Lemma 3.1. *Let $\rho \in \mathcal{S}(\mathbb{C}^{2^n})$ be an n-qubit quantum state whose coefficient vector in the $\mathcal{Q}$-basis expansion has 1-norm $B_\rho > 0$. Let $O \in \mathcal{B}(\mathbb{C}^{2^n})$ be a self-adjoint n-qubit observable whose coefficient vector in the $\mathcal{Q}$-basis expansion has 1-norm $B_O > 0$. Assume that we have ε-accurate estimates $\hat{r}^{\mathcal{Q}}_{i,j}$ for the entries of the $\mathcal{Q}$-TM of an n-qubit quantum channel $\mathcal{N} : \mathcal{B}(\mathbb{C}^{2^n}) \to \mathcal{B}(\mathbb{C}^{2^n})$. Then, from these estimates, we can obtain an $(\varepsilon B_\rho B_O)$-accurate estimate $\hat{\mu}$ of* $\operatorname{tr}[O\mathcal{N}(\rho)]$.

Proof. Consider the $\mathcal{Q}$-basis expansions

$$\rho = \sum_{i=1}^{4^n} \alpha_i Q_i,\ \alpha_i = \operatorname{Tr}[\rho Q_i], \tag{3.1}$$

$$O = \sum_{i=1}^{4^n} \beta_i Q_i,\ \beta_i = \operatorname{Tr}[O Q_i], \tag{3.2}$$

of ρ and of O. By expanding in the $\mathcal{Q}$-basis, we can rewrite our quantity of interest, $\operatorname{Tr}[O\mathcal{N}(\rho)]$, as

$$\operatorname{Tr}[O\mathcal{N}(\rho)] = \sum_{i,j=1}^{4^n} \alpha_i \beta_j (R^{\mathcal{Q}}_{\mathcal{N}})_{j,i}\,. \tag{3.3}$$

Accordingly, we define the estimate $\hat{\mu}$ as

$$\hat{\mu} := \sum_{i,j=1}^{4^n} \alpha_i \beta_j \hat{r}^{\mathcal{Q}}_{j,i}\,, \tag{3.4}$$

with the the estimates $\hat{r}^{\mathcal{Q}}_{i,j}$ for the entries of the $\mathcal{Q}$-TM satisfying $|\hat{r}^{\mathcal{Q}}_{i,j} - (R^{\mathcal{Q}}_{\mathcal{N}})_{j,i}| \leq \varepsilon$ for all $1 \leq i, j \leq 4^n$. Then, we can argue for the accuracy of $\hat{\mu}$ as follows:

$$\begin{aligned} |\hat{\mu} - \mathrm{Tr}[O\mathcal{N}(\rho)]| &= \left| \sum_{i,j=1}^{4^n} \alpha_i \beta_j (\hat{r}^{\mathcal{Q}}_{j,i} - (R^{\mathcal{Q}}_{\mathcal{N}})_{j,i}) \right| && (3.5)\\ &\leq \varepsilon \cdot \sum_{i,j=1}^{4^n} |\alpha_i| \cdot |\beta_j| && (3.6)\\ &= \varepsilon \|\vec{\alpha}\|_1 \cdot \|\vec{\beta}\|_1 && (3.7)\\ &= \varepsilon B_\rho B_O\,. && (3.8) \end{aligned}$$

Here, the inequality is an application of Hölder's inequality (with $p = 1$ and $q = \infty$). □

As sparsity assumptions lead to 1-norm bounds, this has the following consequence.

Corollary 3.2. *Let $\rho \in \mathcal{S}(\mathbb{C}^{2^n})$ be an n-qubit quantum state whose coefficient vector in the $\mathcal{Q}$-basis expansion is s_ρ-sparse. Let $O \in \mathcal{B}(\mathbb{C}^{2^n})$ be a self-adjoint n-qubit observable whose coefficient vector in the $\mathcal{Q}$-basis expansion is s_O-sparse. Assume that we have ε-accurate estimates $\hat{r}^{\mathcal{Q}}_{i,j}$ for the entries of the $\mathcal{Q}$-TM of an n-qubit quantum channel $\mathcal{N} : \mathcal{B}(\mathbb{C}^{2^n}) \to \mathcal{B}(\mathbb{C}^{2^n})$. Then, from these estimates, we can obtain an $\left(\varepsilon s_\rho \sqrt{s_O} \|O\|_2 \max_{1\leq i\leq 4^n} \|Q_i\|_\infty\right)$-accurate estimate of* $\mathrm{tr}[O\mathcal{N}(\rho)]$.

Proof. We use the notation from the previous proof for $\mathcal{Q}$-basis coefficients. First observe that, by orthonormality of the Q_i, Parseval's equality yields

$$\|\rho\|_2^2 = \sum_{i=1}^{4^n} |\alpha_i|^2 = \|\vec{\alpha}\|_2^2\,, \quad \|O\|_2^2 = \sum_{i=1}^{4^n} |\beta_i|^2 = \|\vec{\beta}\|_2^2\,. \tag{3.9}$$

Now, if $\vec{\alpha}$ is s_ρ-sparse and $\vec{\beta}$ is s_O-sparse, Hölder's inequality tells us that

$$\|\vec{\alpha}\|_1 \leq s_\rho \|\vec{\alpha}\|_\infty = s_\rho \max_{1\leq i\leq 4^n} |\mathrm{Tr}[\rho Q_i]| \leq s_\rho \max_{1\leq i\leq 4^n} \|\rho\|_1 \cdot \|Q_i\|_\infty = s_\rho \max_{1\leq i\leq 4^n} \|Q_i\|_\infty\,, \tag{3.10}$$

as well as

$$\|\vec{\beta}\|_1 \leq \sqrt{s_O} \|\vec{\beta}\|_2 = \sqrt{s_O} \|O\|_2\,. \tag{3.11}$$

Plugging these bounds into the statement of the previous lemma finishes the proof. □

Remark 3.3. In the preceding proofs, we have chosen to expand both ρ and O w.r.t. $\mathcal{Q}$. This leads to the norm bounds $\|\vec{\alpha}\|_\infty \leq \max_{1\leq i\leq 4^n} \|Q_i\|_\infty$ and $\|\vec{\beta}\|_\infty \leq \|O\|_2$. From a quantum information perspective, one may instead consider the alternative expansions $\rho = \sum_{i=1}^{4^n} \tilde{\alpha}_i \max_{1\leq j\leq 4^n} \|Q_j\|_\infty Q_j$ with $\tilde{\alpha}_i = \frac{1}{\max_{1\leq j\leq 4^n} \|Q_i\|_\infty} \mathrm{Tr}[\rho Q_i]$, as well as $O = \sum_{i=1}^{4^n} \tilde{\beta}_i \frac{1}{\max_{1\leq j\leq 4^n} \|Q_j\|_\infty} Q_i$ with $\tilde{\beta}_i = \max_{1\leq j\leq 4^n} \|Q_j\|_\infty \mathrm{Tr}[OQ_i]$. These lead to the norm bounds $\|\vec{\tilde{\alpha}}\|_\infty \leq 1$ and $\|\vec{\tilde{\beta}}\|_\infty \leq \max_{1\leq j\leq 4^n} \|Q_i\|_\infty \|O\|_2$. In particular, if $\mathcal{Q} = \mathcal{P}$ is the Pauli ONB, we have the operator norm bound $\max_{1\leq j\leq 4^n} \|Q_i\|_\infty = \frac{1}{\sqrt{2^n}}$, and if O has bounded operator norm $\|O\| \leq B$, we get $\|\vec{\tilde{\alpha}}\|_\infty \leq 1$ and $\|\vec{\tilde{\beta}}\|_\infty \leq B$. Of course, this change in normalization does not change the sparsity of the respective expansions and also leaves the product of the 1-norms of the coefficient vectors invariant. Thus, the results of Lemma 3.1 and Corollary 3.2 remain untouched by such a simultaneous change in normalization.

In the rest of this article, we will sometimes for brevity call O (or ρ) s_O-sparse (or s_ρ-sparse) w.r.t. $\mathcal{Q}$ if the respective coefficient vector $(\mathrm{tr}[OQ_i])_{i=1}^{4^n}$ (or $(\mathrm{tr}[\rho Q_i])_{i=1}^{4^n}$) in the $\mathcal{Q}$-basis is s_O-sparse (or s_ρ-sparse). If $\mathcal{Q} = \mathcal{P}$ is the Pauli ONB, we will speak of Pauli-sparse observables and states. To illustrate potential applications, we next discuss some examples of observables and states with a sparse ONB expansion. We do so for the relevant special case of the Pauli ONB.

Example 3.4 (Pauli-sparse Observables). The following constitute examples for observables with a sparse Pauli-ONB expansion:

- All n-qubit Pauli observables of arbitrarily high weight have Pauli-sparsity $s_O = 1$, independently of n.
- All geometrically k-local n-qubit observables have Pauli-sparsity $s_O = \mathcal{O}(4^k n^D)$, assuming the underlying lattice is a subset of $\mathbb{Z}^D$.
- All k-local n-qubit observables have Pauli-sparsity $s_O \leq \binom{n}{k} \cdot 4^k \leq \mathcal{O}((4n)^k)$.
- All k-local degree-d n-qubit observables—that is, k-local n-qubit observables made up of summands such that each qubit is acted on by at most d of them (see the work of Huang et al. [52] for a formal definition)—have Pauli-sparsity $s_O \leq \mathcal{O}(nd4^k)$.

Remark 3.5. For these last two classes of observables, one can alternatively use the work of Huang et al. [52, Corollaries 3 and 4] to get better bounds on the 1-norm of the vector of Pauli coefficients of O. Note, however, that in another work, Huang et al. [52] consider the unnormalized Pauli expansion of an observable. If we account for this (compare also the discussion in Remark 3.3), then in the notation of Lemma 3.1, the work of Huang et al. [52, Corollary 3] combined with Hölder's inequality implies that for k-local O, we have

$$\|\vec{\beta}\|_1 \leq \|\vec{\beta}\|_{2k/(k+1)} s_O^{(k-1)/2k} \leq \frac{3\sqrt{2^n}}{C(k)} \|O\| \cdot \mathcal{O}\left((4n)^{\frac{k-1}{2}} \right), \tag{3.12}$$

where $C(k) = (\exp(\Theta(k \log k)))^{-1}$. Even more directly, the work of Huang et al. [52, Corollary 4] implies that for a k-local degree-d observable O,

$$\|\vec{\beta}\|_1 \leq \frac{3\sqrt{2^n}}{C(k,d)} \|O\|, \tag{3.13}$$

with $C(k,d) = (\sqrt{d} \exp(\Theta(k \log k)))^{-1}$. Both of these 1-norm bounds improve in terms of n-dependence upon the more naive $\|\vec{\beta}\|_1 \leq \sqrt{s_O} \|O\|_2 \leq \sqrt{s_O 2^n} \|O\|$ that we would obtain by combining the respective sparsity observed in Example 3.4 with the reasoning that led from Lemma 3.1 to Corollary 3.2. This demonstrates that the bound of Corollary 3.2, based on exact sparsity, can be looser than the 1-norm-based bound of Lemma 3.1.

Example 3.6 (Pauli-sparse States). The following constitute examples for states with a sparse Pauli-ONB expansion:

- For any non-identity n-qubit Pauli σ_A, indexed by a string $A \in \{0,1,2,3\}^n$ with $A \neq 0^n$, the mixed state $\rho_\pm = \frac{1}{2^n}(\mathbb{1}_{2^n} \pm \sigma_A)$ proportional to the projector on the ±1 eigenspace of σ_A has Pauli-sparsity $s_{\rho_\pm} = 2$, independently of n.
- Let $\sigma_{A_1}, \dots, \sigma_{A_k}$ be pairwise distinct n-qubit Paulis. Then, the mixed state $\rho_\pm = \frac{1}{2^n} \prod_{\ell=1}^k (\mathbb{1}_{2^n} \pm \sigma_{A_\ell}) = \frac{1}{2^n} \sum_{b \in \{0,1\}^k} (-1)^{|b|} \sigma_{A_1}^{b_1} \dots \sigma_{A_k}^{b_k}$ proportional to the projector on the joint ±1 eigenspace of $\sigma_{A_1}, \dots, \sigma_{A_k}$ has Pauli-sparsity $s_{\rho_\pm} = |\langle \sigma_{A_1}, \dots, \sigma_{A_k} \rangle| = 2^k$.

Having given these examples of observables and states that we can cover, we remark on an important limitation on the class of sparse states.

Remark 3.7. The class of quantum states with an s_ρ-sparse $\mathcal{Q}$-basis expansion only contains states with a limited purity. With the notation used in the proof of Lemma 3.8, this can be seen as follows. If ρ is s_ρ-sparse w.r.t. $\mathcal{Q}$, then its purity can be bounded as

$$\operatorname{tr}\left[\rho^2\right] = \|\rho\|_2^2 = \|\vec{\alpha}\|_2^2 \leq s_\rho \|\vec{\alpha}\|_\infty^2 \leq s_\rho \max_{1\leq i\leq 4^n} \|Q_i\|_\infty^2. \tag{3.14}$$

As $1 = \operatorname{tr}[\rho] \leq \operatorname{rank}(\rho)\operatorname{tr}\left[\rho^2\right]$, this also implies that $\operatorname{rank}(\rho) \geq (s_\rho \max_{1\leq i\leq 4^n}\|Q_i\|_\infty^2)^{-1}$. For our results in later sections, we will assume $\max_{1\leq i\leq 4^n}\|Q_i\|_\infty = \frac{1}{\sqrt{2^n}}$ and $s_\rho \leq \mathcal{O}(\mathrm{poly}(n))$, with n the number of qubits. Under these assumptions, we see that any s_ρ-sparse state ρ has purity at most $\mathcal{O}(\frac{\mathrm{poly}(n)}{2^n})$ and rank at least $\Omega(\frac{2^n}{\mathrm{poly}(n)})$. Thus, such states are highly mixed and high-rank. Note, however, that sparse states can be far from maximally mixed in trace distance. For instance, if $A \in \{0,1,2,3\}^n$ with $A \neq 0^n$, then $\rho_\pm = \frac{\mathbb{1}_{2^n} \pm \sigma_A}{2^n}$ has Pauli-sparsity 2 and $\left\|\rho_\pm - \frac{1}{2^n}\mathbb{1}_{2^n}\right\|_1 = 1$.

We also note that the results of this subsection as well as the results of later sections all straightforwardly extend to observables O that are close to sparse in operator norm and to states that are close to sparse in 1-norm, since the true expectation value is then well approximated by the one of the sparse approximations. This, however, does not add conceptual insight beyond the case of exact sparsity, so we do not explicitly mention the approximate sparsity variant of the result each time.

3.2 Sparse Quantum Channel

In the previous subsection, we made assumptions on the states and observables but no assumptions on the unknown quantum channel. Here, we consider the converse, by making a sparsity assumption on the channel but allowing for arbitrary states and observables.

Lemma 3.8. *Let $\rho \in \mathcal{S}(\mathbb{C}^{2^n})$ be an n-qubit quantum state, and let $O \in \mathcal{B}(\mathbb{C}^{2^n})$ be a self-adjoint n-qubit observable. Let $\mathcal{N} : \mathcal{B}(\mathbb{C}^{2^n}) \to \mathcal{B}(\mathbb{C}^{2^n})$ be an n-qubit quantum channel with an $s_\mathcal{N}$-sparse TM w.r.t. $\mathcal{Q}$. Assume that we have ε-accurate estimates $\hat{r}_{i,j}^{\mathcal{Q}}$ for the non-zero entries of the $\mathcal{Q}$-TM of $\mathcal{N}$. Then, from these estimates, we can obtain an $(\varepsilon s_\mathcal{N}\|O\|_2 \max_{1\leq i\leq 4^n}\|Q_i\|_\infty)$-accurate estimate of* $\operatorname{tr}[O\mathcal{N}(\rho)]$.

Proof. Let us denote the pairs of indices corresponding to non-zero entries of the $\mathcal{Q}$-TM of $\mathcal{N}$ by $(i_1, j_1), \ldots, (i_{s_\mathcal{N}}, j_{s_\mathcal{N}}) \in \{1,\ldots,4^n\}^2$. Expanding in the $\mathcal{Q}$-basis, we again start from Equations (3.3) and (3.4). We can make use of our sparsity assumption to reduce the number of summands and then again apply Hölder to obtain the following:

$$|\hat{\mu} - \operatorname{Tr}[O\mathcal{N}(\rho)]| = \left|\sum_{i,j=1}^{4^n} \alpha_i\beta_j(\hat{r}_{j,i}^{\mathcal{Q}} - (R_\mathcal{N}^{\mathcal{Q}})_{j,i})\right| \tag{3.15}$$

$$= \left|\sum_{s=1}^{s_\mathcal{N}} \alpha_{j_s}\beta_{i_s}\left(\hat{r}_{i_s,j_s}^{\mathcal{Q}} - (R_\mathcal{N}^{\mathcal{Q}})_{i_s,j_s}\right)\right| \tag{3.16}$$

$$\leq \varepsilon \cdot \sum_{s=1}^{s_\mathcal{N}} |\alpha_{j_s}| \cdot |\beta_{i_s}|, \tag{3.17}$$

where we used our assumption that $|\hat{r}^{Q}_{i_s,j_s} - (R^{Q}_{\mathcal{N}})_{i_s,j_s}| \leq \varepsilon$ for all $1 \leq s \leq s_{\mathcal{N}}$. Yet another application of Hölder gives us

$$\begin{aligned} \sum_{s=1}^{s_{\mathcal{N}}} |\alpha_{j_s}| \cdot |\beta_{i_s}| &\leq \left(\sum_{s=1}^{s_{\mathcal{N}}} |\alpha_{j_s}|\right) \cdot \max_{1\leq s\leq s_{\mathcal{N}}} |\beta_{i_s}| && (3.18)\\ &\leq s_{\mathcal{N}} \|\vec{\alpha}\|_\infty \cdot \|\vec{\beta}\|_\infty && (3.19)\\ &\leq s_{\mathcal{N}} \|\vec{\alpha}\|_\infty \cdot \|\vec{\beta}\|_2 && (3.20)\\ &\leq s_{\mathcal{N}} \cdot \left(\max_{1\leq i\leq 4^n} \|Q_i\|_\infty\right) \cdot \|O\|_2 , && (3.21) \end{aligned}$$

where the final step uses the inequality $\|\vec{\alpha}\|_\infty \leq \max_{1\leq i\leq 4^n} \|Q_i\|_\infty$ and the equality $\|\vec{\beta}\|_2 = \|O\|_2$, both of which were established in the proof of Corollary 3.2. Combining the two chains of inequalities gives the claimed approximation guarantee. □

We emphasize that applying Lemma 3.8 requires estimates for the non-zero TM entries as well as the knowledge that all other TM entries are equal to zero. In addition, similarly to how we could allow for approximate sparsity in the states and observables in the previous subsection, it is straightforward to extend Lemma 3.8 to channels that are close to a channel with sparse TM in 1-to-1 norm. We conclude this subsection with some examples of quantum channels with a sparse (Pauli) TM.

Example 3.9 (Pauli-sparse Channels). The following constitute examples for channels with a sparse PTM:

- Let $\mathcal{N}(\cdot) = \sum_{C\in\{0,1,2,3\}^n} \gamma_C \sigma_C(\cdot)\sigma_C$ be an n-qubit Pauli channel with Pauli error rates satisfying $\gamma_C \geq 0$ and $\sum_{C\in\{0,1,2,3\}^n} \gamma_C = 1$. Then, for any $A, B \in \{0,1,2,3\}^n$, we have

$$\begin{aligned} \frac{1}{2^n}\operatorname{tr}[\sigma_A \mathcal{N}(\sigma_B)] &= \frac{1}{2^n} \sum_{C\in\{0,1,2,3\}^n} \gamma_C \operatorname{tr}[\sigma_A\sigma_C\sigma_B\sigma_C] && (3.22)\\ &= \left(\sum_{\substack{C\in\{0,1,2,3\}^n\\ [\sigma_B,\sigma_C]=0}} \gamma_C - \sum_{\substack{D\in\{0,1,2,3\}^n\\ \{\sigma_B,\sigma_D\}=0}} \gamma_D \right) \delta_{A,B}. && (3.23) \end{aligned}$$

  In other words, for any $B \in \{0,1,2,3\}^n$, the normalized Pauli $\frac{\sigma_B}{\sqrt{2^n}}$ is an eigenoperator of $\mathcal{N}$ with associated Pauli eigenvalue given by

$$\lambda_B = \sum_{\substack{C\in\{0,1,2,3\}^n\\ [\sigma_B,\sigma_C]=0}} \gamma_C - \sum_{\substack{D\in\{0,1,2,3\}^n\\ \{\sigma_B,\sigma_D\}=0}} \gamma_D . \qquad (3.24)$$

  In other words, the PTM of a Pauli channel is a diagonal matrix with its eigenvalues on the diagonal. Therefore, for Pauli channels, PTM learning is equivalent to the task of Pauli eigenvalue learning considered in the work of Chen et al. [29]. Accordingly, a Pauli channel has an s-sparse PTM if and only if its spectrum contains (at most) s non-zero eigenvalues.
- An n-qubit channel $\mathcal{N}$ has an s-sparse PTM if and only if its Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ has an s-sparse Pauli ONB expansion. (This can be seen as a direct consequence of Equation (4.1).) Thus, any example of a Pauli-sparse $2n$-qubit state ρ with $\operatorname{tr}_{n+1,\ldots,2n}[\rho] = \frac{1}{2^n}\mathbb{1}_{2^n}$ gives rise to a quantum channel with s-sparse PTM. Concretely, if the Pauli expansion of the Choi state is $\frac{1}{2^n}\Gamma^{\mathcal{N}} =$

$\sum_{C\in\{0,1,2,3\}^{2n}} \gamma(C)\frac{\sigma_C}{\sqrt{2^{2n}}}$, with $\gamma(B0^n) = \frac{1}{\sqrt{2^n}}\delta_{B,0^n}$ to ensure that $\mathcal{N}$ is trace-preserving, then the corresponding channel $\mathcal{N}$ acts on an input state $\rho = \sum_{A\in\{0,1,2,3\}^n} \alpha(A)\frac{\sigma_A}{\sqrt{2^n}}$ via

$$\mathcal{N}(\rho) = 2^n \sum_{A,B\in\{0,1,2,3\}^n} (-1)^{|\{1\leq i\leq n \mid A_i=2\}|}\alpha(A)\gamma(AB)\frac{\sigma_B}{\sqrt{2^n}}\,. \tag{3.25}$$

— To make the abstract construction of the previous bullet point more concrete, consider the Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}} = \frac{\mathbb{1}_{2^{2n}}+\sigma_A}{2^{2n}}$, for some $A\in\{0,1,2,3\}^{2n}$ with $A_{n+1}\ldots A_{2n}\neq 0^n$. The action of the associated n-qubit quantum channel is given by

$$\mathcal{N}(\cdot) = 2^n \operatorname{tr}_{\text{in}}\left[((\cdot)_{\text{in}}^{\top}\otimes \mathbb{1}_{\text{out}})\frac{1}{2^n}\Gamma^{\mathcal{N}}\right] = \frac{\operatorname{tr}[(\cdot)]\mathbb{1}_{2^n} + s_A \operatorname{tr}\left[(\cdot)\sigma_{A_1\ldots A_n}\right]\sigma_{A_{n+1}\ldots A_{2n}}}{2^n}, \tag{3.26}$$

where the sign s_A is given by $s_A = (-1)^{|\{1\leq i\leq n \mid A_i=2\}|}$. As $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ has a 2-sparse Pauli expansion, this channel $\mathcal{N}$ has a 2-sparse PTM.

We conclude this section with a remark about limitations of the class of sparse channels.

Remark 3.10. If $\mathcal{N}$ has an $s_{\mathcal{N}}$-sparse $\mathcal{Q}$-TM, then that directly implies that this TM has rank at most $s_{\mathcal{N}}$. As the $\mathcal{Q}$-TM of $\mathcal{N}$ is simply its matrix representation w.r.t. the ONB $\mathcal{Q}$, that tells us this: an n-qubit quantum channel with $s_{\mathcal{N}}$-sparse TM has rank at most $s_{\mathcal{N}}$ when viewed as a linear map from $\mathcal{B}((\mathbb{C}^2)^{\otimes n})$ to itself. Thus, our notion of sparse channels can only capture relatively low-rank channels (i.e., channels with a relatively low-dimensional image space).

We can also consider the Choi/Kraus rank of $\mathcal{N}$ instead of its rank as a linear map. For simplicity, let us focus on the case of the Pauli ONB. As mentioned previously, by Equation (4.1), $\mathcal{N}$ having an s-sparse PTM is equivalent to its Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ having an s-sparse Pauli expansion. Following the reasoning in Remark 3.7, we thus see that an s-sparse channel $\mathcal{N}$ has a Choi/Kraus rank of at least $\Omega(\frac{4^n}{s})$. For general ONBs and TMs, we can obtain a Choi/Kraus rank lower bound with a similar reasoning, replacing Equation (4.1) by Equation (B.3).

4 Efficient PTM Learning With Quantum Memory

In this section, we show that a learner with access to a quantum memory can efficiently estimate all the PTM entries of an unknown channels from Choi access. Combining this with the results of Section 3, this gives a procedure for estimating Pauli-sparse expectation values of an unknown channel. Alternatively, we can use it to estimate arbitrary expectation values for an unknown channel promised to have a sparse PTM.

First, we show that the Pauli-specific shadow tomography result of Huang et al. [55] allows us to learn the PTM of an arbitrary unknown channel from a small number of Choi state copies, assuming access to a quantum memory.

Theorem 4.1. *There is a learning procedure with quantum memory from Choi access that, using* $m = O(\frac{n+\log(1/\delta)}{\varepsilon^4})$ *copies of the Choi state* $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ *of an arbitrary unknown n-qubit channel* $\mathcal{N}$*, outputs, with success probability* $\geq 1-\delta$*, numbers* $\hat{r}^{\mathcal{P}}_{A,B}$ *for* $A,B\in\{0,1,2,3\}^n$ *such that* $|\hat{r}^{\mathcal{P}}_{A,B} - (R^{\mathcal{P}}_{\mathcal{N}})_{A,B}|\leq\varepsilon$ *holds simultaneously for all* $A,B\in\{0,1,2,3\}^n$.

Proof. For every $A,B\in\{0,1,2,3\}^n$, we can write the corresponding PTM entry as

$$\left(R^{\mathcal{P}}_{\mathcal{N}}\right)_{A,B} = \frac{1}{2^n}\operatorname{tr}[\sigma_A\mathcal{N}(\sigma_B)] = \operatorname{tr}\left[(\sigma_B^{\top}\otimes\sigma_A)\frac{1}{2^n}\Gamma^{\mathcal{N}}\right] = (-1)^{|\{1\leq i\leq n|B_i=2\}|}\operatorname{tr}\left[(\sigma_B\otimes\sigma_A)\frac{1}{2^n}\Gamma^{\mathcal{N}}\right]. \tag{4.1}$$

Here, the last equality used that $\sigma_0^\top = \sigma_0$, $\sigma_1^\top = \sigma_1$, $\sigma_2^\top = -\sigma_2$, and $\sigma_3^\top = \sigma_3$. Thus, an ε-accurate estimate of the Pauli expectation value $\mathrm{tr}[(\sigma_B \otimes \sigma_A)\frac{1}{2^n}\Gamma^{\mathcal{N}}]$ immediately translates to a ε-accurate estimate of the PTM entry $(R_{\mathcal{N}}^{\mathcal{P}})_{A,B}$.

To learn all these $M = 16^n$ Pauli expectation values simultaneously, we employ the Pauli shadow tomography procedure of Huang et al. [55, Supplementary Material V.B]. According to Huang et al. [55, Theorem 2], this procedure uses at most

$$\mathcal{O}\left(\frac{\log\left(\frac{16^n}{\delta}\right)}{\varepsilon^4}\right) = \mathcal{O}\left(\frac{n + \log\left(\frac{1}{\delta}\right)}{\varepsilon^4}\right) \tag{4.2}$$

copies of the Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ to produce, with success probability $\geq 1-\delta$, simultaneously ε-accurate estimates for all Pauli expectation values. □

In Appendix E, we give a simple proof for a corresponding lower bound of $\Omega(\frac{n}{\varepsilon^2})$, thus establishing the optimality of the linear-in-n scaling for learning the PTM with quantum memory from Choi access. In fact, Chen et al. [29, Theorem 2, (iv)] use teleportation stretching for Pauli channels to show the following: even given general channel access and even when the unknown channel is promised to be a Pauli channel and thus to have a diagonal PTM, the linear-in-n scaling cannot be improved.

When focusing on the task of predicting sparse expectation values, we not only have efficient query complexity, but we can also translate guarantees on the efficiency of the Pauli shadow tomography result of Huang et al. [55] in terms of both classical memory and classical post-processing time to our setting.

Corollary 4.2. *There is a learning procedure with quantum memory from Choi access that, using*

$$m_1 = \mathcal{O}\left(\frac{n+\log(1/\delta)}{\varepsilon^4} \cdot B^4 s_\rho^4 s_O^2\right) \tag{4.3}$$

copies of the Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ *of an unknown n-qubit channel* $\mathcal{N}$*, produces a classical description* $\hat{\mathcal{N}}$ *of* $\mathcal{N}$ *from which any M expectation values of the form* $\mathrm{tr}[O_i\mathcal{N}(\rho_i)]$, $1 \leq i \leq M$, *with* O_i *some n-qubit observables with* s_O*-sparse Pauli basis expansions satisfying* $\|O_i\|_2 \leq B\sqrt{2^n}$ *and* ρ_i *some n-qubit states with* s_ρ*-sparse Pauli basis expansions, can be predicted to accuracy* ε *with* $m_2 = \mathcal{O}(\frac{\log(\min\{Ms_\rho s_O, 16^n\})+\log(1/\delta)}{\varepsilon^2} \cdot B^2 s_\rho^2 s_O)$ *additional copies of* $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ *and classical computation time* $\mathcal{O}(nm_1M) = \mathcal{O}(\frac{n^2+n\log(1/\delta)}{\varepsilon^4} \cdot B^4 s_\rho^4 s_O^2 M)$*, with success probability* $\geq 1-\delta$*. The classical representation* $\hat{\mathcal{N}}$ *of* $\mathcal{N}$ *consists of* $\mathcal{O}(nm_1) = \mathcal{O}(\frac{n^2+n\log(1/\delta)}{\varepsilon^4} \cdot B^4 s_\rho^4 s_O^2)$ *real numbers stored in classical memory.*

Proof. This follows from the work of Huang et al. [55, Theorem 2] and the bounds on classical memory and classical computation time given in the proof thereof (see [55, Supplementary Material V.B]) when combined with Corollary 3.2. To see this, note that, by Corollary 3.2, it suffices to use the protocol of Huang et al. [55, Supplementary Material V.B] to predict the at most $\min\{Ms_\rho s_O, 16^n\}$ relevant non-zero PTM entries, which we can relate to expectation values of Pauli observables on the Choi state by Equation (4.1), each to accuracy $\tilde{\varepsilon} = \frac{\varepsilon}{s_\rho\sqrt{s_O}B}$. □

When talking about computational efficiency in Corollary 4.2, we implicitly assume that the Pauli-sparse observables O_i and the Pauli-sparse states ρ_i are given to the learner in the natural efficient classical representations, $\hat{O}_i$ and $\hat{\rho}_i$, namely as lists of the respective Pauli coefficients. For example, given the freedom in normalization discussed in Remark 3.3, we can work with the classical representations $\hat{\rho}_i = \{(A_s, \operatorname{tr}[\rho_i \sigma_{A_s}])\}_{s=1}^{s_\rho}$ and $\hat{O}_i = \{(B_s, \frac{1}{2^n} \operatorname{tr}[O_i \sigma_{B_s}])\}_{s=1}^{s_O}$. Given these Pauli representations of O_i and ρ_i, our learning procedure can then efficiently predict $\operatorname{tr}[O_i \mathcal{N}(\rho_i)]$. Corollary 4.2 thus tells us that we can predict expectation values of the form $\operatorname{tr}[O\mathcal{N}(\rho)]$ to an inverse-polynomial accuracy from an efficient number of copies of $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ using efficient classical computation time and efficient classical memory, as long as we assume that both O and ρ have a $\mathcal{O}(\text{poly}(n))$-sparse Pauli basis expansions and that $\|O\| \leq \mathcal{O}(\text{poly}(n))$.

While we formulate Corollary 4.2 for sparse observables and states, we note that a similar copy complexity bound holds when replacing sparsity by 1-norm assumptions. This can be proven in the same way, merely replacing Corollary 3.2 by Lemma 3.1. As we saw in Remark 3.5, the latter can lead to tighter bounds than the former. However, when starting from 1-norm assumptions, it is not immediately clear how to translate the statements about computational efficiency, for example, because there is no longer a natural efficient representation for the observables and states.

Remark 4.3. We note some additional features of Corollary 4.2. First, if the classical representations are only approximate, in the sense that we obtain $\hat{\rho}_i = \{(A_s, \hat{\alpha}_s^{(i)})\}_{s=1}^{s_\rho}$ with $|\hat{\alpha}_s^{(i)} - \operatorname{tr}[\rho_i \sigma_{A_s}]| \leq \varepsilon_\rho$ and $\hat{O}_i = \{(B_s, \hat{\beta}_s^{(i)})\}_{s=1}^{s_O}$ with $|\hat{\beta}_s^{(i)} - \frac{1}{2^n}\operatorname{tr}[O_i \sigma_{B_s}]| \leq \varepsilon_O$, then our procedure still gives $(\varepsilon + s_\rho s_O \varepsilon_O + s_\rho \varepsilon_\rho(\sqrt{s_O}B + s_O \varepsilon_O))$-accurate estimates for $\operatorname{tr}[O_i \mathcal{N}(\rho_i)]$. (This can be seen by augmenting the proof of Lemma 3.8 with additional triangle inequalities.)

Second, we inherit two more interesting features of Huang et al. [55, Theorem 2]. On the one hand, in Corollary 4.2, the O_i and ρ_i need not be known in advance, and we can build our classical representation and get told only afterward which expectation values we should predict. On the other hand, the part of the procedure in Corollary 4.2 that produces the classical representation $\hat{\mathcal{N}}$ only ever performs measurements on at most two copies of the Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$, so a quantum memory of $2n$ qubits suffices to build $\hat{\mathcal{N}}$. However, using $\hat{\mathcal{N}}$ to predict M expectation values uses a quantum memory of $2n(m_2 - 1)$ qubits, with m_2 as in Corollary 4.2.

If we are promised in advance that the unknown quantum channel has a sparse PTM but the location of the non-zero entries is unknown, then a combination of Theorem 4.1 with a cut-off argument leads to the following variant of our result.

LEMMA 4.4. *There is a learning procedure with quantum memory from Choi access that, using* $m = \mathcal{O}(\frac{n+\log(1/\delta)}{\varepsilon^4})$ *copies of the Choi state* $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ *of an unknown n-qubit channel* $\mathcal{N}$ *with* $s_{\mathcal{N}}$*-sparse PTM, outputs, with success probability* $\geq 1-\delta$*, a list* $\{((i_s, j_s), \hat{r}^{\mathcal{P}}_{i_s,j_s})\}_{s=1}^{S}$ *of length S such that*

(1) $(i,j) \notin \{(i_s,j_s)\}_{s=1}^{S} \Rightarrow |(R^{\mathcal{P}}_{\mathcal{N}})_{i,j}| \leq \varepsilon$,
(2) $(i,j) \in \{(i_s,j_s)\}_{s=1}^{S} \Rightarrow |(R^{\mathcal{P}}_{\mathcal{N}})_{i,j}| > \varepsilon/6$,
(3) $S \leq s_{\mathcal{N}}$*, and*
(4) $|\hat{r}^{\mathcal{P}}_{i_s,j_s} - (R^{\mathcal{P}}_{\mathcal{N}})_{i_s,j_s}| \leq \varepsilon$ *holds for all* $1 \leq s \leq S$.

PROOF: First apply the learning procedure of Theorem 4.1 with accuracy parameter $\tilde{\varepsilon} = \varepsilon/3$. With success probability $\geq 1-\delta$, this produces estimates $\tilde{r}^{\mathcal{P}}_{i,j}$ satisfying $|\tilde{r}^{\mathcal{P}}_{i,j} - (R^{\mathcal{P}}_{\mathcal{N}})_{i,j}| \leq \varepsilon/3$ simultaneously for all $1 \leq i,j \leq 4^n$. Now define

$$\hat{r}^{\mathcal{P}}_{i,j} = \begin{cases} 0 & \text{if } |\tilde{r}^{\mathcal{P}}_{i,j}| \leq \varepsilon/2 \\ \tilde{r}^{\mathcal{P}}_{i,j} & \text{else} \end{cases}. \tag{4.4}$$

Let $\{((i_s,j_s), \hat{r}^{\mathcal{P}}_{i_s,j_s})\}_{s=1}^{S}$ be the list of remaining non-zero numbers $\hat{r}^{\mathcal{P}}_{i,j}$ and the associated locations (i,j) in the TM. It is now easy to see that, conditioned on the success event, this list has the desired properties:

(1) Assume $(i,j) \notin \{(i_s, j_s)\}_{s=1}^S$. By construction, this is the case if and only if $\hat{r}^{\mathcal{P}}_{i,j} = 0$, which in turn is equivalent to $|\tilde{r}^{\mathcal{P}}_{i,j}| \leq \varepsilon/2$. As $|\tilde{r}^{\mathcal{P}}_{i,j} - (R^{\mathcal{P}}_{\mathcal{N}})_{i,j}| \leq \varepsilon/3$ holds in the case of success, this implies $|(R^{\mathcal{P}}_{\mathcal{N}})_{i,j}| \leq \varepsilon/2 + \varepsilon/3 = 5\varepsilon/6 \leq \varepsilon$ by the triangle inequality.

(2) Assume $(i,j) \in \{(i_s, j_s)\}_{s=1}^S$. By construction, this is the case if and only if $\hat{r}^{\mathcal{P}}_{i,j} \neq 0$ and $\hat{r}^{\mathcal{P}}_{i,j} = \tilde{r}^{\mathcal{P}}_{i,j}$, which in turn is equivalent to $|\tilde{r}^{\mathcal{P}}_{i,j}| > \varepsilon/2$. As $|\tilde{r}^{\mathcal{P}}_{i,j} - (R^{\mathcal{P}}_{\mathcal{N}})_{i,j}| \leq \varepsilon/3$ holds in the case of success, this implies $|(R^{\mathcal{P}}_{\mathcal{N}})_{i,j}| > \varepsilon/2 - \varepsilon/3 = \varepsilon/6$.

(3) As a direct consequence of 2., we see that $S \leq \{(i,j) \mid (R^{\mathcal{P}}_{\mathcal{N}})_{i,j} \neq 0\} = s_{\mathcal{N}}$.

(4) For any $1 \leq s \leq S$, we have $\hat{r}^{\mathcal{P}}_{i_s,j_s} = \tilde{r}^{\mathcal{P}}_{i_s,j_s}$ by construction. As $|\tilde{r}^{\mathcal{P}}_{i_s,j_s} - (R^{\mathcal{P}}_{\mathcal{N}})_{i_s,j_s}| \leq \varepsilon/3$ holds in the case of success, we have $|\hat{r}^{\mathcal{P}}_{i_s,j_s} - (R^{\mathcal{P}}_{\mathcal{N}})_{i_s,j_s}| \leq \varepsilon/3 \leq \varepsilon$. □

Combining Lemma 4.4 with Lemma 3.8 gives a variant of Corollary 4.2 in which we replace the sparsity assumption on states and observables by a sparsity assumption on the unknown channel.

Corollary 4.5. *There is a learning procedure with quantum memory from Choi access that, using*

$$m = O\left(\frac{n + \log(1/\delta)}{\varepsilon^4} \cdot B^4 s_{\mathcal{N}}^4\right) \tag{4.5}$$

copies of the Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ *of an unknown n-qubit channel* $\mathcal{N}$ *with* $s_{\mathcal{N}}$*-sparse PTM, produces, with success probability* $\geq 1 - \delta$*, a classical description* $\hat{\mathcal{N}}$ *of* $\mathcal{N}$ *from which all expectation values of the form* $\operatorname{tr}[O\mathcal{N}(\rho)]$*, where O is an arbitrary n-qubit observable with* $\|O\|_2 \leq B\sqrt{2^n}$ *and* ρ *is an arbitrary n-qubit state, can be predicted to accuracy* ε*.*

Proof. By Lemma 3.8, we can get the desired ε-accurate expectation value estimates from $\tilde{\varepsilon}$-accurate PTM entry estimates if we set $\tilde{\varepsilon} = \frac{\varepsilon\sqrt{2^n}}{s_{\mathcal{N}}\|O\|_2}$. If we plug this accuracy $\tilde{\varepsilon}$ into the copy complexity bound of Lemma 4.4 and insert the assumed inequality $\|O\|_2 \leq B\sqrt{2^n}$, we obtain the stated copy complexity bound. □

Corollary 4.5 describes an information-theoretically efficient way of predicting expectation values for arbitrary input states and arbitrary bounded output observables if the unknown quantum channel is promised to have a polynomially-sparse PTM. Importantly, the learning procedure in Corollary 4.5 does not require knowledge about the sparsity structure of the PTM. It suffices to know only that the TM is sparse, the sparsity structure is then found as part of the learning procedure. In fact, if additionally the sparsity structure of the (P)TM is known in advance, then there is a straightforward information-theoretically efficient learning procedure without quantum memory (compare Remark 5.4).

Remark 4.6. Analogously to Corollary 4.2, one can show that the classical representation $\mathcal{N}$ in Corollary 4.5 can be chosen to consist of $O(\frac{n^2 + n\log(1/\delta)}{\varepsilon^4} \cdot B^4 s_{\mathcal{N}}^4)$ real numbers. However, it is not clear how to translate the computation time bound from Corollary 4.2 to the case of sparse channels and arbitrary states and observables. In particular, there is no longer a natural efficient representation of the states and observables for which the expectation value is to be predicted.

While we focus on the PTM in this section, we can extend most of our results to TMs w.r.t. more general unitary orthonormal bases. We explain this in Appendix B.

5 Exponential Query Complexity Lower Bounds For PTM Learning Without Quantum Memory

In the language introduced in Section 2.2, we can view Section 4 (and Appendix B) as giving us algorithms with quantum memory for learning the PTM (and TM) and for predicting sparse expectation values of an unknown quantum channel from Choi access with a polynomial number of copies. In other words, if we have access to a quantum memory, then we can information-theoretically efficiently solve these two learning tasks even in the weakest of the three quantum access models defined in Section 2.2. In this section, we explore how essential the quantum memory is to achieve the efficient information-theoretic complexity scaling.

5.1 Query Complexity Lower Bounds for PTM Learning without Quantum Memory from Choi Access

We first prove exponential query complexity lower bounds for the weakest of our learning models, namely that of learning without quantum memory from Choi access. While the result of this subsection is implied by that of the next subsection on learning without quantum memory from general channel access, we have chosen to present it separately because the proof in the case of Choi access is technically less involved and can serve as preparation for the proof in the general channel access case. More precisely, the proof for Choi access is a variant of the proof strategy used in the work of Chen et al. [27] to establish exponential query complexity lower bounds for learners without quantum memory for the task of Pauli shadow tomography of quantum states. For learning without quantum memory from Choi access, we have the following query complexity lower bound.

Lemma 5.1. *Any algorithm for learning without quantum memory from Choi access requires* $\Omega(\frac{4^n}{\varepsilon^2})$ *copies of the Choi state of an unknown n-qubit channel* $\mathcal{N}$ *to estimate all entries of* $R^{\mathcal{P}}_{\mathcal{N}}$ *up to accuracy* ε *with success probability* $\geq \frac{2}{3}$.

Proof Sketch. We give a detailed proof in Appendix D. The proof is a modification to that of Chen et al. [25, Corollary 5.9] and proceeds as follows. As learning PTM entries is, by Equation (4.1), equivalent to predicting Pauli expectation values for the Choi state, PTM learning becomes a Pauli shadow tomography problem on the level of the Choi state. For general states, Chen et al. [25, Corollary 5.9] established a Pauli shadow tomography query complexity lower bound for learners without quantum memory by reducing to a many-versus-one state distinguishing task and constructing an ensemble of states for which the latter task is hard. To adapt their proof to our setting, we effectively restrict their state ensemble to the sub-ensemble of valid Choi states and analyze the hardness of the Pauli shadow tomography task for that sub-ensemble. □

We can recover the same asymptotic copy complexity lower bound even if we restrict the unknown channel to be doubly-stochastic and entanglement-breaking, and to have a sparse PTM with unknown sparsity structure.

Lemma 5.2. *Any algorithm for learning without quantum memory from Choi access requires* $\Omega(\frac{4^n}{\varepsilon^2})$ *copies of the Choi state of an unknown doubly-stochastic and entanglement-breaking n-qubit channel* $\mathcal{N}$ *with* $\mathcal{O}(1)$*-sparse PTM to estimate all entries of* $R^{\mathcal{P}}_{\mathcal{N}}$ *up to accuracy* ε *with success probability* $\geq \frac{2}{3}$.

Proof. We give a detailed proof in Appendix D. The proof requires only a small modification compared to that of Lemma 5.1. Namely, we have to further restrict the state ensemble to those that are valid Choi states of doubly-stochastic quantum channels. All states in the ensemble are already separable and $\mathcal{O}(1)$-sparse, so we automatically focus on entanglement-breaking channels with $\mathcal{O}(1)$-sparse PTM. Again, we have to perform a hardness analysis for the Pauli shadow tomography task of this further restricted ensemble. □

If $\mathcal{N}$ is doubly-stochastic, then we have for $A, B \in \{0,1,2,3\}^n$,

$$\operatorname{tr}\left[\sigma_A \mathcal{N}\left(\frac{\mathbb{1}_2^{\otimes n} + \sigma_B}{2^n}\right)\right] = \begin{cases} 1 + \left(R_{\mathcal{N}}^{\mathcal{P}}\right)_{A,B} & \text{if } A = 0^n \\ \left(R_{\mathcal{N}}^{\mathcal{P}}\right)_{A,B} & \text{else} \end{cases}. \tag{5.1}$$

Thus, the ability to estimate expectation values of the form $\operatorname{tr}[\sigma_A \mathcal{N}(\frac{\mathbb{1}_2^{\otimes n} + \sigma_B}{2^n})]$ up to a desired accuracy ε immediately allows one to estimate the PTM entries $(R_{\mathcal{N}}^{\mathcal{P}})_{A,B}$ to the same accuracy, assuming $\mathcal{N}$ is doubly-stochastic. Trivially, the observables σ_A have a 1-sparse Pauli basis expansion, and the states $\frac{\mathbb{1}_2^{\otimes n} + \sigma_B}{2^n}$ have a 2-sparse Pauli basis expansion. In addition, $\|\sigma_A\|_2 = \sqrt{2^n}$. Therefore, Lemma 5.2 has the following consequence.

Corollary 5.3. *Any learning algorithm without quantum memory requires $\Omega(\frac{4^n}{\varepsilon^2})$ copies of the Choi state of an unknown doubly-stochastic and entanglement-breaking n-qubit channel $\mathcal{N}$ with $O(1)$-sparse PTM to produce a classical description $\hat{\mathcal{N}}$ of $\mathcal{N}$ from which any expectation value of the form $\operatorname{tr}[O\mathcal{N}(\rho)]$, with O an n-qubit observable with $O(1)$-sparse Pauli basis expansion satisfying $\|O\|_2 \leq \sqrt{2^n}$ and ρ an n-qubit state with $O(1)$-sparse Pauli basis expansion, can be predicted to accuracy ε with success probability $\geq \frac{2}{3}$.*

This is to be contrasted with Corollaries 4.2 and 4.5 (as well as Corollaries B.4 and B.6), which showed that using a quantum memory allows to solve even more general versions of the same expectation value prediction task using polynomially-in-n or even linearly-in-n many copies of the Choi state of an unknown n-qubit quantum channel. Thus, we have established an exponential separation in query complexity between algorithms with and without quantum memory for the task of predicting Pauli-sparse expectation values of an unknown channel with sparse PTM when given access to copies of its Choi state.

Remark 5.4. The exponential query complexity lower bound for learners without quantum memory from Choi access may no longer apply if stronger assumptions on the unknown quantum channel are made. For example, if $\mathcal{N}$ is promised to be $s_{\mathcal{N}}$-sparse and the sparsity structure of $\mathcal{N}$ (i.e., the positions of non-zero PTM entries) is known in advance, then, by Lemma 3.8, it suffices to estimate the corresponding $s_{\mathcal{N}}$ many Pauli expectation values for the Choi state. Thus, if $s_{\mathcal{N}} \leq O(\operatorname{poly}(n))$, then this can easily be achieved without quantum memory using $O(\frac{\operatorname{poly}(n)\cdot\log(1/\delta)}{\varepsilon^2})$ Choi state copies.

Remark 5.5. Complementary to the previous remark, we note that the exponential query complexity lower bound for learners without quantum memory from Choi access may also fail to hold if stronger assumptions on the input states and output observables are made. For instance, if we only care about "low-weight" observables and states, whose Pauli expansions consist of Paulis acting non-trivially on at most a constant number of qubits, then applying classical shadows to the Choi state suffices, so no quantum memory is needed [60, 61].

5.2 Query Complexity Lower Bounds for PTM Learning without Quantum Memory from Sequential Channel Access

Next, we extend the exponential query complexity lower bounds for the task of PTM learning without quantum memory also to the stronger model of sequential channel access (Definition 2.7).

Theorem 5.6. *Any learning algorithm without quantum memory requires $\Omega(\frac{4^n}{\varepsilon^2})$ copies of an unknown n-qubit channel $\mathcal{N}$ to estimate all entries of $R_{\mathcal{N}}^{\mathcal{P}}$ up to accuracy ε with success probability $\geq \frac{2}{3}$.*

Proof. We give a detailed proof in Appendix D. The proof is based on the learning tree representation of Chen et al. [27] and an extension of their analysis from state shadow tomography to our channel shadow tomography setting. While the underlying framework is the same as in the proof of the Choi state access case, the technical details become more involved due to the possibility of adaptively chosen input states. ☐

Similarly to the previous subsection, we also have a stronger doubly-stochastic, entanglement-breaking, and sparse version.

Theorem 5.7. *Any learning algorithm without quantum memory requires $\Omega(\frac{4^n}{\varepsilon^2})$ copies of an unknown doubly-stochastic and entanglement-breaking n-qubit channel $\mathcal{N}$ with $\mathcal{O}(1)$-sparse PTM to estimate all entries of $R^{\mathcal{P}}_{\mathcal{N}}$ up to accuracy ε with success probability $\geq \frac{2}{3}$.*

Proof. We give a detailed proof in Appendix D. This proof requires only a small modification to that of Theorem 5.6, which again consists of further restricting the ensemble of unknown channels to a sub-ensemble consisting only of doubly-stochastic entanglement-breaking channels with $\mathcal{O}(1)$-sparse PTM. This further restriction then requires a corresponding adaptation to the hardness analysis. ☐

In exactly the same way that Lemma 5.2 gave rise to Corollary 5.3, Theorem 5.7 has the following immediate consequence.

Corollary 5.8. *Any learning algorithm without quantum memory requires $\Omega(\frac{4^n}{\varepsilon^2})$ copies of an unknown doubly-stochastic and entanglement-breaking n-qubit channel $\mathcal{N}$ with $\mathcal{O}(1)$-sparse PTM to produce a classical description $\hat{\mathcal{N}}$ of $\mathcal{N}$ from which any expectation value of the form* $\operatorname{tr}[O\mathcal{N}(\rho)]$, *with O an n-qubit observable with $\mathcal{O}(1)$-sparse Pauli basis expansion satisfying $\|O\|_2 \leq \sqrt{2^n}$ and ρ an n-qubit state with $\mathcal{O}(1)$-sparse Pauli basis expansion, can be predicted to accuracy ε with success probability $\geq \frac{2}{3}$.*

This shows that the exponential query complexity separation for the task of predicting Pauli-sparse expectation values between learning with quantum memory from Choi state copies and learning without quantum memory from Choi state copies persists even if we allow learners without quantum memory to access copies of the unknown sparse channel. Again, a learner with quantum memory can even solve more general tasks than that in Corollary 5.8 (see Corollaries 4.2 and 4.5). Thereby, we have established an exponential separation between the weakest of our three learning models with quantum memory (Definition 2.10) and the strongest of our three learning models without quantum memory (Definition 2.7).

Remark 5.9. Additionally, the exponential query complexity lower bounds in this subsection can fail under *a priori* promises on the unknown channel. Concretely, Chen et al. [29, Corollary 2] give a learning procedure without quantum memory that estimates all (non-trivial) PTM entries of an unknown n-qubit Pauli channel $\mathcal{N}$ using $\mathcal{O}(\frac{n\log(1/\delta)}{\varepsilon^2})$ queries to $\mathcal{N}$, where n-dimensional auxiliary quantum systems are used. Thus, if we are promised that the unknown channel is a Pauli channel, the exponential query complexity separation between PTM learning with and without quantum memory breaks down. In fact, the protocol of Chen et al. [29] achieves the same n-dependence and a better ε-dependence than that of Theorem 4.1, despite the latter using a quantum memory.

Remark 5.10. As a further example of a change in framework can break the exponential query complexity lower bound for learners without quantum memory, we highlight the worst case nature of our task. Namely, we require the learner to make accurate predictions for all Pauli-sparse states and observables. In contrast, [52] recently established that arbitrary quantum channels can be learned efficiently without quantum memory from parallel access in a worst-case sense over certain (k-local degree-d) observables and on-average over certain (locally flat) distributions of quantum states.

Remark 5.11. To conclude our discussion of query complexity lower bounds, we explain the implications of Chen et al. [29, Theorem 2, (iii)] for our learning task. Formulated in our language, Chen et al. [29, Theorem 2, (iii)] say this: without the possibility of querying the unknown channel on a subsystem of an entangled quantum state (as is, for example, required to prepare Choi state copies), the task of PTM learning requires $\Omega(2^{n/3})$ many queries, even if the unknown channel is promised to be a Pauli channel (or, equivalently, to have a diagonal PTM). This teaches us about the importance of allowing for an auxiliary quantum system our definitions of learning with quantum memory. Combined with the query complexity lower bounds for learners without quantum memory established in this section, we see that the efficient query complexity of PTM learning with quantum memory established in Theorem 4.1 requires both the ability to query the unknown channel sequentially without intermediate measurements and the ability to let the unknown channel act on a subsystem of a larger composite quantum system.

6 Hamiltonian Learning With Quantum Memory

As an application of our positive results for PTM learning with quantum memory, we show how these can be combined with polynomial interpolation techniques to give rise to a Hamiltonian learning protocol. The ideas presented in this section follow the approaches based on polynomial interpolation and derivative estimation put forward in other works [40, 43, 75], but our PTM learner allows us to deal with arbitrary Hamiltonians, whereas these prior works require assumptions on the structure of the unknown Hamiltonian.

Our goal is to learn an unknown n-qubit Hamiltonian $H = H^\dagger \in \mathcal{B}((\mathbb{C}^2)^{\otimes n})$, without making prior assumptions on the structure of H. Namely, if H has the Pauli expansion $H = \sum_{A\in\{0,1,2,3\}^n} \alpha(A)\sigma_A$, with $\alpha(A) = \frac{1}{2^n}\operatorname{tr}[H\sigma_A] \in \mathbb{R}$ and w.l.o.g. $\alpha(0^n) = 0$, we aim to approximate the coefficient vector $\vec{\alpha} = (\alpha(A))_{A\in\{0,1,2,3\}^n}$ in ℓ_∞-norm over $\mathbb{R}^{4^n}$. (Note: Here we consider an unnormalized Pauli expansion.)

We will consider learning an unknown Hamiltonian from access to the associated unitary dynamics. For any time $t \geq 0$, we will denote the unitary quantum channel describing the (Schrödinger picture) evolution under the Hamiltonian H for time t by

$$\mathcal{U}_t : \mathcal{B}((\mathbb{C}^2)^{\otimes n}) \to \mathcal{B}((\mathbb{C}^2)^{\otimes n}),\ \mathcal{U}_t(\rho) = e^{-itH}\rho e^{itH}\,. \tag{6.1}$$

By expanding the operator exponentials in their power series, we obtain

$$\mathcal{U}_t(\rho) = \rho - it[H,\rho] + \sum_{k=2}^{\infty}\frac{(-it)^k}{k!}C_k^H(\rho)\,, \tag{6.2}$$

where $C_k^H(\rho)$ denotes the k-fold iterated commutator defined via

$$C_k^X(Y) := \underbrace{[X,[X,[\ldots,[X}_{k\text{ times}},Y]\ldots]]]\,. \tag{6.3}$$

Repeating an observation previously made in other works [40, 43, 75, 82], we can isolate any single coefficient $\alpha(A)$ in the first-order term when acting on a suitable Pauli-sparse input state and measuring a suitable Pauli observable on the obtained output state.

Lemma 6.1. *Let $H = \sum_{A\in\{0,1,2,3\}^n} \alpha(A)\sigma_A$, with $\alpha(A) = \frac{1}{2^n}\operatorname{tr}[H\sigma_A] \in \mathbb{R}$. Let $A \in \{0,1,2,3\}^n$, $A \neq 0^n$. Let $1 \leq j \leq n$ be such that $A_j \neq 0$. Let $B_j \in \{1,2,3\}\backslash\{A_i\}$ and set $B_k = 0$ for all $k \neq i$. Define* *the quantum state* $\rho = \frac{\mathbb{1}_{2^n} + i[\sigma_A, \sigma_B]/2}{2^n}$. *Then*

$$\operatorname{tr}[\sigma_B \cdot (-i[H, \rho])] = 2\alpha(A). \tag{6.4}$$

Proof. This can be seen by a direct computation:

$$\begin{aligned}
\operatorname{tr}[\sigma_B \cdot (-i[H, \rho])] &= \operatorname{tr}[(i[H, \sigma_B]) \cdot \rho] && (6.5)\\
&= -\frac{1}{2^{n+1}} \operatorname{tr}[[H, \sigma_B] \cdot [\sigma_A, \sigma_B]] && (6.6)\\
&= -\frac{1}{2^{n+1}} \sum_{C \in \{0,1,2,3\}^n} \alpha(C) \operatorname{tr}[[\sigma_C, \sigma_B] \cdot [\sigma_A, \sigma_B]] && (6.7)\\
&= \frac{1}{2^{n+1}} \sum_{\substack{C \in \{0,1,2,3\}^n \\ [\sigma_B, \sigma_C] \neq 0}} \alpha(C) \operatorname{tr}[[\sigma_C, \sigma_B] \cdot [\sigma_A, \sigma_B]] && (6.8)\\
&= \frac{1}{2^{n+1}} \sum_{\substack{C \in \{0,1,2,3\}^n \\ [\sigma_B, \sigma_C] \neq 0}} \alpha(C) \operatorname{tr}[2\sigma_B \sigma_C \cdot 2\sigma_A \sigma_B] && (6.9)\\
&= \frac{2}{2^{n}} \sum_{\substack{C \in \{0,1,2,3\}^n \\ [\sigma_B, \sigma_C] \neq 0}} \alpha(C) \operatorname{tr}[\sigma_C \cdot \sigma_A] && (6.10)\\
&= 2\alpha(A), && (6.11)
\end{aligned}$$

as claimed. □

Recalling Equation (6.2), we conclude the following. For any $A \neq 0^n$, if we manage to extract the first-order time derivative at time zero of $\operatorname{tr}[\sigma_B \mathcal{U}_t(\rho)]$ with B and ρ as in Lemma 6.1, then that tells us the coefficient $\alpha(A)$. We can achieve this by performing polynomial interpolation for the function $t \mapsto \operatorname{tr}[\sigma_B \mathcal{U}_t(\rho)]$. To ensure that polynomial interpolation gives an accurate estimate for the first derivative, we need bounds on higher-order derivatives. The next lemma establishes such bounds for a general Hamiltonian.

Lemma 6.2. *Let* $B \in \{0,1,2,3\}^n$ *be an arbitrary string, and let* $\rho \in \mathcal{S}((\mathbb{C}^2)^{\otimes n})$ *be an arbitrary* n*-qubit state. Then, for any* $k \in \mathbb{N}$ *and for any* $\tau \geq 0$,

$$\left| \frac{\mathrm{d}^k}{\mathrm{d}t^k} \operatorname{Tr}[\sigma_B \mathcal{U}_t(\rho)] \Big|_{t=\tau} \right| \leq (2\|H\|)^k . \tag{6.12}$$

Proof. Using the power series expansion of $\mathcal{U}_t(\rho)$, one can evaluate the time derivative of a time-dependent expectation value as $\frac{\mathrm{d}^k}{\mathrm{d}t^k} \operatorname{Tr}[\sigma_B \mathcal{U}_t(\rho)]\big|_{t=\tau} = (i)^k \operatorname{Tr}[C_k^H(\sigma_B) \mathcal{U}_\tau(\rho)]$. As the operator norm of an iterated commutator satisfies $\left\|C_k^X(Y)\right\| \leq (2\|X\|)^k \|Y\|$, an application of Hölder's inequality gives

$$\left| \frac{\mathrm{d}^k}{\mathrm{d}t^k} \operatorname{Tr}[\sigma_B \mathcal{U}_t(\rho)] \Big|_{t=\tau} \right| \leq \left\|C_k^H(\sigma_B)\right\| \cdot \|\mathcal{U}_\tau(\rho)\|_1 \leq (2\|H\|)^k \|\sigma_B\| = (2\|H\|)^k, \tag{6.13}$$

since $\mathcal{U}_\tau$ is completely positive and trace-preserving for every $\tau \geq 0$ and ρ is a quantum state. □

Together with polynomial interpolation guarantees for derivative estimation, our PTM learning protocol from Corollary 4.2 now has the following consequence.

Theorem 6.3. *There is a learning procedure with quantum memory from parallel access that, using*

$$m_1 = \tilde{O}\left(\tfrac{n + \log(1/\delta)}{\varepsilon^4} \cdot \|H\|^4\right) \tag{6.14}$$

parallel queries to $\mathcal{O}(\frac{1}{\|H\|})$*-time evolutions along* H *with a total evolution time of* $T_1 = \tilde{\mathcal{O}}(\frac{n+\log(1/\delta)}{\varepsilon^4} \cdot \|H\|^3)$*, produces a classical description* $\hat{H}$ *of* H *from which any* M *Pauli coefficients* $\alpha(A^{(i)})$*,* $A^{(i)} \in \{0,1,2,3\}^n$*,* $1 \leq i \leq M$*, can be estimated to accuracy* ε *with* $m_2 = \tilde{\mathcal{O}}(\frac{\log(M)+\log(1/\delta)}{\varepsilon^2} \cdot \|H\|^2)$ *additional parallel queries to* $\mathcal{O}(\frac{1}{\|H\|})$*-time evolutions along* H *with a total evolution time of* $T_2 = \tilde{\mathcal{O}}(\frac{n+\log(1/\delta)}{\varepsilon^2} \cdot \|H\|)$ *and classical computation time* $\tilde{\mathcal{O}}(\frac{n^2+n\log(1/\delta)}{\varepsilon^4} \cdot \|H\|^4 M)$*, with success probability* $\geq 1-\delta$*. The classical representation* $\hat{H}$ *of* H *consists of* $\tilde{\mathcal{O}}(\frac{n^2+n\log(1/\delta)}{\varepsilon^4} \cdot \|H\|^4)$ *real numbers stored in classical memory.*

Proof Sketch. We give a detailed proof in Appendix D. The learning procedure works as follows. First, we use Corollary 4.2 to build classical representations $\hat{\mathcal{U}}_t$ of $\mathcal{U}_t$ for different times t. The totality of those classical representations makes up our classical representation $\hat{H}$ of H. When asked to estimate a Pauli coefficients $\alpha(A^{(i)})$, $1 \leq i \leq M$, from that classical representation, we use each $\hat{\mathcal{U}}_t$ to obtain an estimate of $\frac{1}{2}\operatorname{tr}[\sigma_{B^{(i)}}\mathcal{U}_t(\rho^{(i)})]$ with $B^{(i)}$ and $\rho^{(i)}$ as in Lemma 6.1 for different times, and then use those estimates to approximately evaluate the first-order time derivative at time 0 via Chebyshev polynomial interpolation for each i. Motivated by Equation (6.2) and Lemma 6.1, these approximate derivatives are our estimates for the Pauli coefficients $\alpha(A^{(i)})$. □

Assuming an upper bound of $\|H\| \leq \mathcal{O}(\text{poly}(n))$, Theorem 6.3 gives an efficient procedure for learning the Pauli coefficients of H in ℓ_∞-norm. Crucially, and in contrast to results from prior work, we do not need any structural assumptions about H, as it can be an arbitrarily non-local Hamiltonian. However, Theorem 6.3 is not directly comparable to prior works on learning Hamiltonians from real time dynamics. Methods making use of known structure in the Hamiltonian achieve exponentially better n dependence and polynomially better ε dependence than Theorem 6.3, and they do not scale directly with $\|H\|$ but rather with the the ℓ_∞-norm of the vector of Pauli coefficients of H and with, for example, the assumed locality parameter [40, 43, 57, 75].

APPENDICES

A Learning Quantum Channels with Parallel Access

We present one more access model for quantum channel learning, this one lying in between the general access model and the Choi access model. Here, we suppose that the learner has to perform all oracle accesses to the unknown quantum channel in parallel. We formalize this as follows.

Definition A.1 (Learning Quantum Channels without Quantum Memory from Parallel Access). A T-query algorithm for learning an unknown n-qubit quantum channel $\mathcal{N}$ without quantum memory from parallel access is a T-query algorithm for learning an unknown n-qubit quantum channel $\mathcal{N}$ without quantum memory under the additional restriction that the input states must not be chosen adaptively.

Definition A.2 (Learning Quantum Channels with Quantum Memory from Parallel Access). A T-query algorithm for learning an unknown n-qubit quantum channel $\mathcal{N}$ with quantum memory from parallel access is a 1-query algorithm for learning the unknown (nT)-qubit quantum channel $\mathcal{N}^{\otimes T}$ with quantum memory. In other words, such an algorithm can prepare an input state $\rho_0 \in \mathcal{S}((\mathbb{C}^2)^{\otimes n_{\text{aux}}}((\mathbb{C}^2)^{\otimes n})^{\otimes T})$, access the oracle for $\mathcal{N}$ in parallel T times to have the state $\rho_T^{\mathcal{N}} = (\text{id}_{\text{aux}} \otimes \mathcal{N}^{\otimes T})(\rho_0)$ in the quantum memory, and then perform a joint POVM $\{F_s\}_s \subseteq \mathcal{B}((\mathbb{C}^2)^{\otimes n_{\text{aux}}} \otimes ((\mathbb{C}^2)^{\otimes n})^{\otimes T})$ on $\rho_T^{\mathcal{N}}$ to predict properties of $\mathcal{N}$.

These two definitions are illustrated in Figure 3. Any procedure for learning an unknown quantum channel from Choi access can also be realized with parallel access to the unknown channel.

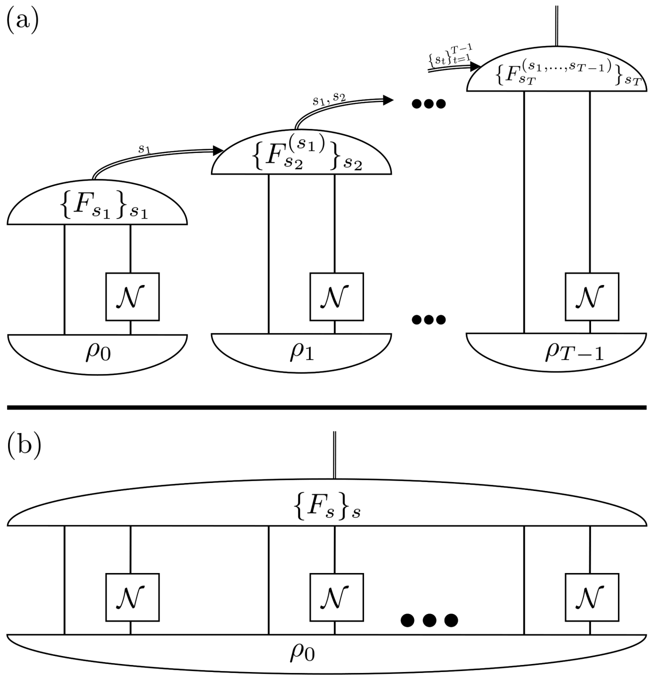

Fig. 3. Illustration of learning quantum channels from parallel access, each panel to be read from bottom to top. Panel (a) depicts a learner without quantum memory, and panel (b) depicts a learner with quantum memory.

Additionally, any procedure with parallel access is clearly an instance of a general channel learning procedure. Thus, the parallel access model sits in between Choi access and general channel access.

B TM Learning With Quantum Memory From Polynomially Many Queries

While the focus in Section 4 was on the PTM, we can extend our reasoning to general TMs when replacing the Pauli shadow tomography procedure of Huang et al. [55] by the more general shadow tomography procedure of Aaronson [2] and Bădescu and O'Donnell [2, 9]. In this appendix, we describe how that leads to a query-efficient procedure for learning general TMs, assuming a quantum memory and Choi state access. According to the results of Section 3, this allows us to estimate sparse expectation values for an unknown quantum channel. Moreover, as discussed in Section 2.2 and Appendix A, this then particularly implies information-theoretically efficient learners with quantum memory from parallel access and from more general sequential access in terms of its power.

First, we demonstrate that applying existing state shadow tomography methods to the Choi state of the unknown quantum channel is sufficient to obtain accurate estimates for all TM entries simultaneously, while using only polynomially many Choi state copies.

Theorem B.1. *There is a learning procedure with quantum memory from Choi access that, using*

$$m = \tilde{O}\left(\frac{n^3+n\log(1/\delta)}{\varepsilon^4}\cdot\left(2^n\cdot\max_{1\le i\le 4^n}\|Q_i\|_\infty^2\right)^4\right) \tag{B.1}$$

copies of the Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ *of an unknown n-qubit channel* $\mathcal{N}$*, outputs, with success probability* $\geq 1-\delta$*, numbers* $\hat{r}^{Q}_{i,j}$ *for* $1 \leq i,j \leq 4^n$ *such that* $|\hat{r}^{Q}_{i,j} - (R^{Q}_{\mathcal{N}})_{i,j}| \leq \varepsilon$ *holds simultaneously for all* $1 \leq i,j \leq 4^n$*.*

Notice that the copy complexity upper bound in Theorem B.1 becomes polynomial in n if we assume $\|Q_i\|_\infty = \frac{1}{\sqrt{2^n}}$ for all $1 \leq i \leq 4^n$. This is, for example, satisfied in the case of the Pauli ONB. In fact, it is easy to see from $\|Q_i\|_2 \leq \sqrt{2^n}\|Q_i\|$ that $\max_{1\leq i\leq 4^n}\|Q_i\| = \frac{1}{\sqrt{2^n}}$ holds if and only if Q is an ONB of normalized unitaries.

Proof. We first define a set of effect operators, to which we then later apply the improvement by Bădescu and O'Donnell [9] to the shadow tomography procedure of Aaronson [2]. Namely, for each $1 \leq i,j \leq 4^n$, we define

$$E_{i,j} = \frac{1}{2}\left(\mathbb{1}\otimes\mathbb{1} + \frac{1}{\|Q_j\|_\infty}Q_j^\top \otimes \frac{1}{\|Q_i\|_\infty}Q_i\right). \tag{B.2}$$

Notice that since every Q_i is Hermitian, so is every $E_{i,j}$. Moreover, since each $\frac{1}{\|Q_j\|_\infty}Q_j^\top \otimes \frac{1}{\|Q_i\|_\infty}Q_i$ is normalized in operator norm, an application of the triangle inequality shows that $0 \leq E_{i,j} \leq \mathbb{1}\otimes\mathbb{1}$ holds for all $1 \leq i,j \leq 4^n$. Thus, every $E_{i,j}$ is a valid effect operator. In addition, note that by definition, we have

$$\begin{aligned}\operatorname{Tr}\left[Q_i\mathcal{N}(Q_j)\right] &= 2^n \operatorname{Tr}\left[(Q_j^\top \otimes Q_i)\frac{1}{2^n}\Gamma^{\mathcal{N}}\right] &\text{(B.3)}\\ &= 2^n \cdot \|Q_i\|_\infty \cdot \|Q_j\|_\infty \cdot \operatorname{Tr}\left[\left(\frac{1}{\|Q_i\|_\infty}Q_i^\top \otimes \frac{1}{\|Q_j\|_\infty}Q_j\right)\frac{1}{2^n}\Gamma^{\mathcal{N}}\right] &\text{(B.4)}\\ &= 2^n \cdot \|Q_i\|_\infty \cdot \|Q_j\|_\infty \cdot 2 \cdot \left(\operatorname{Tr}\left[E_{i,j}\frac{1}{2^n}\Gamma^{\mathcal{N}}\right] - \frac{1}{2}\right). &\text{(B.5)}\end{aligned}$$

Hence, an $\tilde{\varepsilon}$-accurate estimate for the expectation value $\operatorname{Tr}[E_{i,j}\frac{1}{2^n}\Gamma^{\mathcal{N}}]$ gives rise to a $(\tilde{\varepsilon}\cdot 2^{n+1}\cdot \max_{1\leq i\leq 4^n}\|Q_i\|_\infty^2)$-accurate estimate of the TM entry $(R^{Q}_{\mathcal{N}})_{i,j} = \operatorname{Tr}[Q_i\mathcal{N}(Q_j)]$.

Letting $\varepsilon > 0$ and $\delta \in (0,1)$ be arbitrary, define $\tilde{\varepsilon} = \frac{\varepsilon}{2^{n+1}\cdot \max_{1\leq i\leq 4^n}\|Q_i\|_\infty^2} > 0$. According to Bădescu and O'Donnell [9, Theorem 1.4], if we apply their shadow tomography procedure for the $M = 16^n$ effect operators $E_{i,j}$, $1 \leq i,j \leq 4^n$, and the $(2n)$-qubit quantum state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$, we see that

$$\frac{\left(\log^2(M) + \log\left(\frac{2n}{\delta\varepsilon}\right)\right)\cdot 2n}{\tilde{\varepsilon}^4}\cdot \mathcal{O}\left(\log\left(\frac{2n}{\delta\varepsilon}\right)\right) = \tilde{\mathcal{O}}\left(\frac{n^3 + n\log(1/\delta)}{\varepsilon^4}\cdot\left(2^n\cdot\max_{1\leq i\leq 4^n}\|Q_i\|_\infty^2\right)^4\right) \tag{B.6}$$

copies of $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ suffice to produce estimates $\tilde{r}_{i,j}$ satisfying $|\tilde{r}_{i,j} - \operatorname{Tr}[E_{i,j}\frac{1}{2^n}\Gamma^{\mathcal{N}}]| \leq \tilde{\varepsilon}$ for all $1 \leq i,j \leq 4^n$, with success probability $\geq 1-\delta$. By our preceding reasoning, these simultaneously $\tilde{\varepsilon}$-accurate expectation value estimates now give us simultaneously ε-accurate estimates

$$\hat{r}^{Q}_{i,j} = 2^n\cdot\|Q_i\|_\infty\cdot\|Q_j\|_\infty\cdot 2\cdot\left(\tilde{r}_{i,j} - \frac{1}{2}\right) \tag{B.7}$$

of the TM entries, with success probability $\geq 1-\delta$. □

Remark B.2. There are two immediate extensions of Theorem B.1 that we want to highlight. First, we have phrased Theorem B.1 in a way suggesting that the learner knows the ONB Q in advance. This is not needed, as the learner only needs to know an upper bound on $\max_{1\leq i\leq 4^n}\|Q_i\|_\infty^2$ in advance to determine the suitable copy complexity. To see this, recall that the shadow tomography procedure in the work of Bădescu and O'Donnell [9] allows for adaptively chosen measurements, so also in our case the same copy complexity bound holds if the learner does not know Q in advance and the ONB elements are only revealed adaptively. This is in contrast to the Pauli ONB case: when formulating a variant of Theorem 4.1 for estimating M PTM entries, these entries must be known to the learner in advance. This limitation is inherited from the Pauli shadow tomography procedure of Huang et al. [55], which does not allow for adaptively chosen Pauli observables. Similarly, the observables O_i and states ρ_i in Corollary 4.2 must be known to the learner in advance for our sample complexity bound to apply.

Second, while there is only a single ONB $\mathcal{Q}$ in Theorem B.1, we can also allow for N different ONBs $\mathcal{Q}^{(1)}, \dots, \mathcal{Q}^{(N)}$. This will increase the number of effect operators in the proof from 16^n to $N \cdot 16^n$, but since that number only enters the sample complexity bound for shadow tomography logarithmically, the final copy complexity will still be bounded as $\tilde{O}(\frac{n^3+n\log(N)+n\log(1/\delta)}{\varepsilon^4} \cdot (2^n \cdot \max_{1\leq i\leq 4^n, 1\leq L\leq N} \|Q_i^{(L)}\|_\infty^2)^4)$.

In summary, we can sample-efficiently simultaneously learn TMs w.r.t. exponentially many different adaptively chosen ONBs, if we know in advance how many different ONBs there are and that the maximum operator norm of any ONB element is $\frac{1}{\sqrt{2^n}}$. These two extensions also carry over to the next result, but we do not state the corresponding variants explicitly.

Remark B.3. Comparing Theorem B.1 with Theorem 4.1, we see that our bound in the Pauli-specific case provides an asymptotic copy complexity improvement by a factor of n^2 compared to simply applying the bound obtained when applying Theorem B.1 to the Pauli ONB (up to logarithmic factors). Thus, both Theorems 4.1 and B.1 give polynomial-in-n copy complexity bounds for learning the PTM, but the former gives an effectively cubic n-dependence, which the latter improves to linear. Moreover, as the general state shadow tomography results of Aaronson [2] and Bădescu and O'Donnell [9] do not come with bounds on classical memory or computational complexity, we also do not have such bounds in the case of a general TM, in contrast to the PTM case (Corollary 4.2).

Combining Theorem B.1 with Corollary 3.2, we obtain the following.

Corollary B.4. *There is a learning procedure with quantum memory from Choi access that, using*

$$m = \tilde{O}\left(\frac{n^3 + n\log(1/\delta)}{\varepsilon^4} \cdot B^4 s_\rho^4 s_O^2 \cdot \left(\sqrt{2^n} \cdot \max_{1\leq i\leq 4^n} \|Q_i\|_\infty\right)^{12}\right) \tag{B.8}$$

copies of the Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ *of an* n*-qubit channel* $\mathcal{N}$*, produces, with success probability* $\geq 1-\delta$*, a classical description* $\hat{\mathcal{N}}$ *of* $\mathcal{N}$ *from which all expectation values of the form* $\operatorname{tr}[O\mathcal{N}(\rho)]$*, where* O *is an* n*-qubit observable with* s_O*-sparse* $\mathcal{Q}$*-basis expansion such that* $\|O\|_2 \leq B\sqrt{2^n}$ *and* ρ *is an* n*-qubit state with* s_ρ*-sparse* $\mathcal{Q}$*-basis expansion, can be predicted to accuracy* ε*.*

Proof. By Corollary 3.2, we can get the desired ε-accurate expectation value estimates from $\tilde{\varepsilon}$-accurate TM entry estimates if we set $\tilde{\varepsilon} = \frac{\varepsilon}{s_\rho\sqrt{s_O}\|O\|_2 \max_{1\leq i\leq 4^n}\|Q_i\|_\infty}$. Plugging this accuracy $\tilde{\varepsilon}$ into the copy complexity bound of Theorem B.1 and inserting the assumed inequality $\|O\|_2 \leq B\sqrt{2^n}$ now gives the stated copy complexity bound. □

Note that, in particular, any n-qubit observable O with $\|O\|_\infty \leq B$ satisfies the Hilbert-Schmidt norm bound $\|O\|_2 \leq B\sqrt{2^n}$. Thus, the corollary can be applied to bounded $\mathcal{Q}$-sparse observables.

Exactly analogous to the Pauli case, if we are promised in advance that the unknown quantum channel has a sparse TM, then we can combine Theorem B.1 with a cut-off argument to obtain the following variant of our result.

Lemma B.5. *There is a learning procedure with quantum memory from Choi access that, using*

$$m = \tilde{O}\left(\frac{n^3+n\log(1/\delta)}{\varepsilon^4}\cdot\left(2^n\cdot\max_{1\leq i\leq 4^n}\|Q_i\|_\infty^2\right)^4\right) \tag{B.9}$$

copies of the Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ *of an unknown n-qubit channel* $\mathcal{N}$ *with* $s_{\mathcal{N}}$*-sparse TM w.r.t.* $\mathcal{Q}$*, outputs, with success probability* $\geq 1-\delta$*, a list* $\{((i_s,j_s),\hat{r}^{\mathcal{Q}}_{i_s,j_s})\}_{s=1}^S$ *of length S such that*

1. $(i,j)\notin\{(i_s,j_s)\}_{s=1}^S \Rightarrow |(R^{\mathcal{Q}}_{\mathcal{N}})_{i,j}|\leq\varepsilon$,
2. $(i,j)\in\{(i_s,j_s)\}_{s=1}^S \Rightarrow |(R^{\mathcal{Q}}_{\mathcal{N}})_{i,j}|>\varepsilon/6$,
3. $S\leq s_{\mathcal{N}}$, *and*
4. $|\hat{r}^{\mathcal{Q}}_{i_s,j_s}-(R^{\mathcal{Q}}_{\mathcal{N}})_{i_s,j_s}|\leq\varepsilon$ *holds for all* $1\leq s\leq S$.

Proof. The proof is essentially identical to that of Corollary 4.5, and we only have to replace Theorem 4.1 by Theorem B.1. □

Combining Lemma B.5 with Lemma 3.8, we get a variant of Corollary B.4 in which we replace the sparsity assumption on states and observables by a sparsity assumption on the unknown channel.

Corollary B.6. *There is a learning procedure with quantum memory from Choi access that, using*

$$m = \tilde{O}\left(\frac{n^3+n\log(1/\delta)}{\varepsilon^4}\cdot B^4 s_{\mathcal{N}}^4\cdot\left(\sqrt{2^n}\cdot\max_{1\leq i\leq 4^n}\|Q_i\|_\infty\right)^{12}\right) \tag{B.10}$$

copies of the Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ *of an n-qubit channel* $\mathcal{N}$ *with* $s_{\mathcal{N}}$*-sparse TM w.r.t.* $\mathcal{Q}$*, produces, with success probability* $\geq 1-\delta$*, a classical description* $\hat{\mathcal{N}}$ *of* $\mathcal{N}$ *from which all expectation values of the form* $\mathrm{tr}[O\mathcal{N}(\rho)]$*, where O is an n-qubit observable with* $\|O\|_2\leq B\sqrt{2^n}$ *and ρ is an n-qubit state, can be predicted to accuracy ε.*

Proof. Recalling Lemma 3.8, we see that $\tilde{\varepsilon}$-accurate TM entry estimates are sufficient to get the desired ε-accurate expectation value estimates if we take $\tilde{\varepsilon}=\frac{\varepsilon}{s_{\mathcal{N}}\|O\|_2\max_{1\leq i\leq 4^n}\|Q_i\|_\infty}$. Plugging this accuracy $\tilde{\varepsilon}$ into the copy complexity bound of Lemma B.5 and inserting the assumed inequality $\|O\|_2\leq B\sqrt{2^n}$ now gives the stated copy complexity bound. □

If $\mathcal{Q}$ is an ONB of normalized unitaries, then Corollary B.6 describes a procedure for sample-efficiently predicting expectation values for arbitrary input states and arbitrary bounded output observables if the unknown quantum channel is promised to have a polynomially-sparse TM. Here, no *a priori* knowledge about the structure of the (sparse) TM is needed.

C Learning Tree Formalism for Learning Without Quantum Memory

Here, we recall (part of) the learning tree formalism introduced in the work of Chen et al. [27] to establish query complexity lower bounds for learners without quantum memory. More precisely, we need the following definition of the tree representation for learning a quantum channel without quantum memory.

Definition C.1 (Tree Representation for Learning Channels [25, Definition 7.1]). Let $\mathcal{N}$ be an n-qubit quantum channel. An algorithm for learning from T queries to $\mathcal{N}$ without quantum memory can be represented as a rooted tree $\mathcal{T}$ of depth T in which each node encodes all measurement outcomes the algorithm has received thus far. The tree has the following properties:

- Each node u has an associated probability $p^{\mathcal{N}}(u)$.
- The probability associated to the root r of the tree equals 1, $p^{\mathcal{N}}(r)=1$.

- At each non-leaf node u, we prepare a state $|\phi_u\rangle \in (\mathbb{C}^2)^{\otimes n_{\mathrm{aux}}} \otimes (\mathbb{C}^2)^{\otimes n}$, apply the channel $\mathrm{id}_{\mathrm{aux}} \otimes \mathcal{N}$ to $|\phi_u\rangle\!\langle\phi_u|$, and measure a ($u$-dependent) rank-1 POVM $\{\sqrt{w_s^u 2^n 2^{n_{\mathrm{aux}}}}\,|\psi_s^u\rangle\!\langle\psi_s^u|\}_s \subseteq \mathcal{B}((\mathbb{C}^2)^{\otimes n_{\mathrm{aux}}} \otimes (\mathbb{C}^2)^{\otimes n})$ to obtain a classical outcome s.
- If v is a child node of u, then the probability of the child node is related to the probability of its parent via
$$p^{\mathcal{N}}(v) = p^{\mathcal{N}}(u) \cdot w_s^u 2^n 2^{n_{\mathrm{aux}}} \cdot \langle\psi_v|\,(\mathrm{id}_{\mathrm{aux}} \otimes \mathcal{N})(|\phi_u\rangle\!\langle\phi_u|)\,|\psi_v\rangle\,. \tag{C.1}$$
- Each root-to-leaf path in $\mathcal{T}$ is of length T. If ℓ is a leaf node, then $p^{C}(\ell)$ is the probability of the classical memory being in state ℓ after the learning procedure.

While Definition C.1 implicitly assumes all input states to be pure and all measurements to be rank-1 POVMs, these assumptions can be made without loss of generality via standard purification arguments and via simulating general POVMs with rank-1 POVMs. (Compare the work of Chen et al. [25, Lemma 4.8 and Remark 4.19] for the latter.) In addition, we can without loss of generality assume the auxiliary system size n_{aux} to be the same for each query to the channel oracle.

D Additional Proofs

D.1 Proofs of Lemmas 5.1 and 5.2

To prove Lemmas 5.1 and 5.2, we make use of the following two variations of Chen et al. [25, Theorem 5.5].

Lemma D.1 (Shadow Tomography Lower Bound for Choi States). *Let $O_1, \ldots, O_M \in \mathcal{B}((\mathbb{C}^2)^{\otimes 2n})$ be $M \in 2\mathbb{N}$ traceless and self-adjoint $(2n)$-qubit observables satisfying the following properties:*

(i) For every $1 \leq i \leq M$, $\sigma(O_i) \subseteq \{-1, 1\}$. In particular, $\|O_i\| = 1$ holds for all $1 \leq i \leq M$.
(ii) For every $1 \leq i \leq M$, $\mathrm{tr}_{n+1,\ldots,2n}[O_i] = 0$.
(iii) For every $1 \leq i \leq M/2$, $O_i = -O_{i+M/2}$.

Any algorithm without quantum memory requires
$$\Omega\left(\frac{1}{\varepsilon^2 \delta(O_1, \ldots, O_M)}\right) \tag{D.1}$$
copies of the Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ of an unknown n-qubit channel $\mathcal{N}$ to simultaneously predict all expectation values $\mathrm{tr}[O_i \frac{1}{2^n}\Gamma]$, $1 \leq i \leq M$, up to accuracy ε with a success probability $\geq \frac{2}{3}$. Here, we defined
$$\delta(O_1, \ldots, O_M) := \sup_{|\phi\rangle \in (\mathbb{C}^2)^{\otimes 2n} : \|\phi\| = 1} \frac{2}{M} \sum_{i=1}^{M/2} \langle\phi|\, O_i\, |\phi\rangle^2\,. \tag{D.2}$$

Proof. We only have to notice that the maximally mixed state on $2n$ qubits is a valid Choi state, and that the assumptions on the O_i ensure that
$$\frac{1}{2^n}\Gamma_i := \frac{\mathbb{1}_2^{\otimes 2n} + 3\varepsilon O_i}{2^{2n}} \tag{D.3}$$
is a valid Choi state for every $1 \leq i \leq M$, as long as $\varepsilon < 1/3$. Now, the remainder of the proof is exactly the same as the proof of Chen et al. [25, Theorem 5.5]. □

Lemma D.2 (Shadow Tomography Lower Bound for Choi States of Doubly-stochastic Sparse Channels). *Let $O_1, \ldots, O_M \in \mathcal{B}((\mathbb{C}^2)^{\otimes 2n})$ be $M \in 2\mathbb{N}$ traceless and self-adjoint $(2n)$-qubit observables satisfying the following properties:*

(i) For every $1 \leq i \leq M$, $\sigma(O_i) \subseteq \{-1, 1\}$. In particular, $\|O_i\| = 1$ holds for all $1 \leq i \leq M$.
(ii) For every $1 \leq i \leq M$, $\mathrm{tr}_{1,\ldots,n}[O_i] = 0$ and $\mathrm{tr}_{n+1,\ldots,2n}[O_i] = 0$.
(iii) For every $1 \leq i \leq M/2$, $O_i = -O_{i+M/2}$.
(iv) For every $1 \leq i \leq M$, O_i has a $\mathcal{O}(1)$-sparse Pauli ONB expansion.
(v) For every $1 \leq i \leq M$, O_i is a tensor product of $2n$ single-qubit observables.

Any algorithm without quantum memory requires

$$\Omega\left(\frac{1}{\varepsilon^2 \delta(O_1, \ldots, O_M)}\right) \tag{D.4}$$

copies of the Choi state $\frac{1}{2^n}\Gamma^{\mathcal{N}}$ of an unknown doubly-stochastic and entanglement-breaking n-qubit channel $\mathcal{N}$ with $\mathcal{O}(1)$-sparse PTM to simultaneously predict all expectation values $\mathrm{tr}[O_i \frac{1}{2^n}\Gamma]$, $1 \leq i \leq M$, up to accuracy ε with a success probability $\geq \frac{2}{3}$. Here, we defined

$$\delta(O_1, \ldots, O_M) := \sup_{|\phi\rangle \in (\mathbb{C}^2)^{\otimes 2n} : \|\phi\|=1} \frac{2}{M} \sum_{i=1}^{M/2} \langle\phi| O_i |\phi\rangle^2 . \tag{D.5}$$

Proof. We only have to notice that the maximally mixed state on $2n$ qubits is a valid Choi state of a doubly-stochastic and entanglement-breaking quantum channel with $\mathcal{O}(1)$-sparse PTM, and that the assumptions (i) and (ii) on the O_i ensure that

$$\frac{1}{2^n}\Gamma_i := \frac{\mathbb{1}_2^{\otimes 2n} + 3\varepsilon O_i}{2^{2n}} \tag{D.6}$$

is a valid Choi state of a doubly-stochastic quantum channel $\mathcal{N}_i$ for every $1 \leq i \leq M$, as long as $\varepsilon < 1/3$. Moreover, the channel $\mathcal{N}_i$ has a $\mathcal{O}(1)$-sparse PTM because of Equation (4.1) and assumption (iv), and $\mathcal{N}_i$ is entanglement-breaking because its Choi state $\frac{1}{2^n}\Gamma_i$ is separable by assumptions (i) and (v), which can be seen by expanding O_i in its eigenbasis consisting of tensor products of eigenvectors. Now, the remainder of the proof is exactly the same as the proof of Chen et al. [25, Theorem 5.5]. □

To apply these two lemmata, we need to establish upper bounds on $\delta(O_1, \ldots, O_M)$ for sets of Pauli observables satisfying the respective requirements. We do so in the following two lemmata, with a reasoning quite similar to that of Chen et al. [25, Lemma 5.8].

Lemma D.3. *Consider the following $2 \cdot 4^n(4^n - 1)$ many $(2n)$-qubit Pauli observables*

$$P_1, \ldots, P_{4^n(4^n-1)} \in \{+\sigma_A\}_{A \in \{0,1,2,3\}^{2n}} \setminus \{+\sigma_B \otimes \mathbb{1}_2^{\otimes n}\}_{B \in \{0,1,2,3\}^n} \tag{D.7}$$

$$P_{4^n(4^n-1)+1}, \ldots, P_{2 \cdot 4^n(4^n-1)} \in \{-\sigma_A\}_{A \in \{0,1,2,3\}^{2n}} \setminus \{-\sigma_B \otimes \mathbb{1}_2^{\otimes n}\}_{B \in \{0,1,2,3\}^n} . \tag{D.8}$$

The set of observables $\{P_i\}_{i=1}^{2 \cdot 4^n(4^n-1)}$, when suitably ordered, satisfies conditions (i) through (iii) from Lemma D.1. Moreover,

$$\delta(P_1, \ldots, P_{2 \cdot 4^n(4^n-1)}) = \frac{1}{4^n}. \tag{D.9}$$

Proof. Each P_i is a Pauli observable and thus clearly satisfies condition (i). In addition, each P_i satisfies condition (ii) because, for every $A \in \{0,1,2,3\}^{2n}$, $\mathrm{tr}_{n+1,\ldots,2n}[\sigma_A] \neq 0$ holds if and only if $A_j = 0$ for all $n+1 \leq j \leq 2n$. Condition (iii) is satisfied by construction (when ordering the observables suitably). It remains to compute $\delta(P_1, \ldots, P_{2 \cdot 4^n(4^n-1)})$. To this end, let $|\phi\rangle \in (\mathbb{C}^2)^{\otimes 2n}$ be an arbitrary normalized pure state, $\|\phi\| = 1$. Then,

$$\begin{aligned}
&\frac{1}{4^n(4^n-1)} \sum_{i=1}^{4^n(4^n-1)} \langle\phi| P_i |\phi\rangle^2 && \text{(D.10)}\\
&= \frac{1}{4^n(4^n-1)} \operatorname{tr}\left[\left(\sum_{i=1}^{4^n(4^n-1)} P_i \otimes P_i\right)(|\phi\rangle\langle\phi|)^{\otimes 2}\right] && \text{(D.11)}\\
&= \frac{1}{4^n(4^n-1)} \operatorname{tr}\left[\left(\left(\sum_{A\in\{0,1,2,3\}^{2n}} \sigma_A \otimes \sigma_A\right)\right.\right. && \text{(D.12)}\\
&\qquad\qquad \left.\left. - \left(\sum_{B\in\{0,1,2,3\}^{n}} (\sigma_B \otimes \mathbb{1}_2^{\otimes n}) \otimes (\sigma_B \otimes \mathbb{1}_2^{\otimes n})\right)\right)(|\phi\rangle\langle\phi|)^{\otimes 2}\right] && \text{(D.13)}\\
&= \frac{1}{4^n(4^n-1)} \left(\operatorname{tr}\left[2^{2n}\,\mathrm{SWAP}_{1,\ldots,2n}(|\phi\rangle\langle\phi|)^{\otimes 2}\right] - \operatorname{tr}\left[2^{n}\,\mathrm{SWAP}_{1,\ldots,n}(\operatorname{tr}_{n+1,\ldots,2n}[|\phi\rangle\langle\phi|])^{\otimes 2}\right]\right) && \text{(D.14)}\\
&= \frac{1}{4^n(4^n-1)} \left(2^{2n} - 2^n \operatorname{tr}\left[\operatorname{tr}_{n+1,\ldots,2n}[|\phi\rangle\langle\phi|]^2\right]\right). && \text{(D.15)}
\end{aligned}$$

Here, we used the following well-known identity:

$$\sum_{B\in\{0,1,2,3\}^{n}} \sigma_B \otimes \sigma_B = 2^n\,\mathrm{SWAP}_{1,\ldots,n}\,. \tag{D.16}$$

This is a special case of the general representation of the swap operator in any ONB of operators (see, e.g., [81, Example 1.2]).

Accordingly, we get

$$\begin{aligned}
\delta(P_1,\ldots,P_{2\cdot 4^n(4^n-1)}) &= \sup_{|\phi\rangle\in(\mathbb{C}^2)^{\otimes 2n}:\|\phi\|=1} \frac{1}{4^n(4^n-1)} \sum_{i=1}^{4^n(4^n-1)} \langle\phi| P_i |\phi\rangle^2 && \text{(D.17)}\\
&= \frac{2^{2n} - 2^n\cdot\frac{1}{2^n}}{4^n(4^n-1)} && \text{(D.18)}\\
&= \frac{1}{4^n}, && \text{(D.19)}
\end{aligned}$$

and the supremum is attained at $|\phi\rangle$ if and only if $\operatorname{tr}_{n+1,\ldots,2n}[|\phi\rangle\langle\phi|]$ is maximally mixed, which is equivalent to $|\phi\rangle$ being maximally entangled across the cut $\{1,\ldots,2n\} = \{1,\ldots,n\} \cup \{n+1,\ldots,2n\}$. □

Lemma D.4. *Consider the following* $2(4^{2n}-2\cdot 4^n+1) = 2(4^n-1)^2$ *many* $(2n)$*-qubit Pauli observables*

$$\begin{aligned}
P_1,\ldots,P_{(4^n-1)^2} &\in \{+\sigma_A\}_{A\in\{0,1,2,3\}^{2n}:\exists 1\le i\le n,\, n+1\le j\le 2n: A_i\neq 0\neq A_j} && \text{(D.20)}\\
P_{4^{2n}-2\cdot 4^n+2},\ldots,P_{2(4^n-1)^2} &\in \{-\sigma_A\}_{A\in\{0,1,2,3\}^{2n}:\exists 1\le i\le n,\, n+1\le j\le 2n: A_i\neq 0\neq A_j}\,. && \text{(D.21)}
\end{aligned}$$

The set of observables $\{P_i\}_{i=1}^{2(4^n-1)^2}$*, when suitably ordered, satisfies conditions (i) through (v) from Lemma D.2. Moreover,*

$$\delta(P_1,\ldots,P_{2((4^n-1)^2)}) = \frac{1}{4^n-1}. \tag{D.22}$$

Proof. Each P_i is a Pauli tensor product observable and thus clearly satisfies conditions (i), (iv), and (v). Additionally, each P_i satisfies condition (ii) because, for every $A \in \{0,1,2,3\}^{2n}$, $\operatorname{tr}_{n+1,\ldots,2n}[\sigma_A] \neq 0$ holds if and only if $A_j = 0$ for all $n+1 \le j \le 2n$, and $\operatorname{tr}_{1,\ldots,n}[\sigma_A] \neq 0$ holds if and only if $A_i = 0$ for all $1 \le i \le n$. Condition (iii) is satisfied by construction (when ordering the observables suitably). It remains to compute $\delta(P_1, \dots, P_{2(4^n-1)^2})$. To this end, let $|\phi\rangle \in (\mathbb{C}^2)^{\otimes 2n}$ be an arbitrary normalized pure state, $\|\phi\| = 1$. Then,

$$
\begin{aligned}
&\frac{1}{(4^n-1)^2} \sum_{i=1}^{(4^n-1)^2} \langle\phi| P_i |\phi\rangle^2 && \text{(D.23)}\\
&= \frac{1}{(4^n-1)^2} \operatorname{tr}\left[\left(\sum_{i=1}^{(4^n-1)^2} P_i \otimes P_i\right)(|\phi\rangle\langle\phi|)^{\otimes 2}\right] && \text{(D.24)}\\
&= \frac{1}{(4^n-1)^2} \operatorname{tr}\left[\left(\left(\sum_{A\in\{0,1,2,3\}^{2n}} \sigma_A \otimes \sigma_A\right) - \left(\sum_{B\in\{0,1,2,3\}^{n}} (\sigma_B \otimes \mathbb{1}_2^{\otimes n}) \otimes (\sigma_B \otimes \mathbb{1}_2^{\otimes n})\right)\right.\right. && \text{(D.25)}\\
&\qquad \left.\left. - \left(\sum_{B\in\{0,1,2,3\}^{n}} (\mathbb{1}_2^{\otimes n} \otimes \sigma_B) \otimes (\mathbb{1}_2^{\otimes n} \otimes \sigma_B)\right) + \mathbb{1}_2^{\otimes 2n}\right)(|\phi\rangle\langle\phi|)^{\otimes 2}\right] && \text{(D.26)}\\
&= \frac{1}{(4^n-1)^2}\Big(\operatorname{tr}\left[2^{2n}\, \mathrm{SWAP}_{1,\dots,2n}(|\phi\rangle\langle\phi|)^{\otimes 2}\right] - \operatorname{tr}\left[2^{n}\, \mathrm{SWAP}_{1,\dots,n}(\operatorname{tr}_{n+1,\dots,2n}[|\phi\rangle\langle\phi|])^{\otimes 2}\right] && \text{(D.27)}\\
&\qquad - \operatorname{tr}\left[2^{n}\, \mathrm{SWAP}_{n+1,\dots,2n}(\operatorname{tr}_{1,\dots,n}[|\phi\rangle\langle\phi|])^{\otimes 2}\right] + 1\Big) && \text{(D.28)}\\
&= \frac{1}{(4^n-1)^2}\left(2^{2n} - 2^n\left(\operatorname{tr}\left[\operatorname{tr}_{n+1,\dots,2n}[|\phi\rangle\langle\phi|]^2\right] + \operatorname{tr}\left[\operatorname{tr}_{1,\dots,n}[|\phi\rangle\langle\phi|]^2\right]\right) + 1\right), && \text{(D.29)}
\end{aligned}
$$

where we again used Equation (D.16) multiple times in the computation.

Accordingly, we get

$$
\begin{aligned}
\delta(P_1, \dots, P_{2(4^n-1)^2}) &= \sup_{|\phi\rangle\in(\mathbb{C}^2)^{\otimes 2n}: \|\phi\|=1} \frac{1}{(4^n-1)^2} \sum_{i=1}^{(4^n-1)^2} \langle\phi| P_i |\phi\rangle^2 && \text{(D.30)}\\
&= \frac{2^{2n} - 2^n \cdot \frac{2}{2^n} + 1}{(4^n-1)^2} && \text{(D.31)}\\
&= \frac{1}{4^n-1}, && \text{(D.32)}
\end{aligned}
$$

and the supremum is attained at $|\phi\rangle$ if and only if both $\operatorname{tr}_{1,\dots,n}[|\phi\rangle\langle\phi|]$ and $\operatorname{tr}_{n+1,\dots,2n}[|\phi\rangle\langle\phi|]$ are maximally mixed, which is equivalent to $|\phi\rangle$ being maximally entangled across the cut $\{1, \dots, 2n\} = \{1, \dots, n\} \cup \{n+1, \dots, 2n\}$. □

We are now in a position to complete the proof of Lemmas 5.1 and 5.2.

Proof of Lemmas 5.1 and 5.2. Recalling from the proof of Theorem 4.1 that estimating all entries of $R_{\mathcal{N}}^{\mathcal{P}}$ up to accuracy ε is equivalent to estimating all expectation values $\operatorname{tr}\left[\sigma_A \frac{1}{2^n}\Gamma^{\mathcal{N}}\right]$, $A \in \{0,1,2,3\}^{2n}$, up to accuracy ε, we can now combine Lemmas D.1 and D.3 to obtain Lemma 5.1, and combine Lemmas D.2 and D.4 to obtain Lemma 5.2. □

D.2 Proof of Theorems 5.6 and 5.7

Next, we prove Theorems 5.6 and 5.7. To this end, we first establish analogues of Lemmas D.1 and D.2 for the case of sequential channel access. The proof uses the notation introduced in Definition C.1.

Lemma D.5. *Let $O_1, \dots, O_M \in \mathcal{B}((\mathbb{C}^2)^{\otimes 2n})$ be $M \in 2\mathbb{N}$ traceless and self-adjoint $(2n)$-qubit observables satisfying the following properties:*

(i) *For every $1 \leq i \leq M$, $\sigma(O_i) \subseteq \{-1, 1\}$. In particular, $\|O_i\| = 1$ holds for all $1 \leq i \leq M$.*

(ii) For every $1 \leq i \leq M$, $\mathrm{tr}_{n+1,\ldots,2n}[O_i] = 0$.
(iii) For every $1 \leq i \leq M/2$, $O_i = -O_{i+M/2}$.

Any algorithm for learning without quantum memory requires

$$\Omega\left(\frac{1}{\varepsilon^2 \Delta(O_1,\ldots,O_M)}\right) \tag{D.33}$$

queries to an n-qubit channel $\mathcal{N}$ *to simultaneously predict all expectation values* $\mathrm{tr}\left[O_i \frac{1}{2^n}\Gamma\right]$, $1 \leq i \leq M$, *up to accuracy* ε *with a success probability* $\geq \frac{2}{3}$. *Here, we defined*

$$\Delta(O_1,\ldots,O_M) := \sup_{\substack{n_{\mathrm{aux}} \in \mathbb{N} \\ |\phi\rangle \in (\mathbb{C}^2)^{\otimes n_{\mathrm{aux}}} \otimes (\mathbb{C}^2)^{\otimes n} : \|\phi\| = 1 \\ \rho \in \mathcal{S}((\mathbb{C}^2)^{\otimes n_{\mathrm{aux}}} \otimes (\mathbb{C}^2)^{\otimes n})}} \frac{2}{M} \sum_{i-1}^{M/2} \left(\frac{\langle\phi|\, \mathrm{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes O_i)]\, |\phi\rangle}{\langle\phi|\, (\rho_{\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}})\, |\phi\rangle}\right)^2 . \tag{D.34}$$

Analogously to how $\delta(O_1,\ldots,O_M)$ was interpreted by Chen et al. [27] as characterizing the hardness of state shadow tomography without quantum memory for the observables $O_1,\ldots,O_M$, we can view $\Delta(O_1,\ldots,O_M)$ as a quantity characterizing how challenging it is to perform Choi state shadow tomography without quantum memory for these observables, even given query access to the corresponding channel. In particular, if $O_1,\ldots,O_M$ are chosen to be Pauli observables $\sigma_{B_1} \otimes \sigma_{A_1},\ldots,\sigma_{B_M} \otimes \sigma_{A_M}$, then, via Equation (4.1), the quantity $\Delta(O_1,\ldots,O_M)$ expresses the hardness of simultaneously predicting the PTM entries $(\mathcal{R}^{\mathcal{P}}_{\mathcal{N}})_{A_i,B_i}$ when given general access to $\mathcal{N}$, but without a quantum memory being available.

Proof. We first show that any algorithm that can solve the learning task can also solve a certain many-versus-one distinguishing task. Then we show the query complexity lower bound for the latter distinguishing task. As in the proof of Lemma D.1, consider the states

$$\frac{1}{2^n}\Gamma_i := \frac{\mathbb{1}_2^{\otimes 2n} + 3\varepsilon O_i}{2^{2n}}, \quad 1 \leq i \leq M, \tag{D.35}$$

which are valid Choi states by conditions (i) and (ii). Notice that also the maximally mixed state $\frac{1}{2^{2n}}\mathbb{1}_2^{\otimes 2n}$ is a valid Choi state. Moreover, our assumptions on the observables O_i imply both that $\mathrm{tr}[O_i \frac{1}{2^n}\Gamma_i] = 3\varepsilon$ and that $\mathrm{tr}[O_i \frac{1}{2^{2n}}\mathbb{1}_2^{\otimes 2n}] = 0$ for all $1 \leq i \leq M$. Therefore, any algorithm that can predict all the expectation values $\mathrm{tr}[O_i \frac{1}{2^n}\Gamma]$, $1 \leq i \leq M$, up to accuracy ε with success probability $\geq \frac{2}{3}$ immediately gives rise to an algorithm that, with success probability $\geq \frac{2}{3}$, solves the many-versus-one distinguishing task between the unknown channel having maximally mixed Choi state $\frac{1}{2^{2n}}\mathbb{1}_2^{\otimes 2n}$ and the unknown channel being drawn uniformly at random from the M channels $\mathcal{N}_i$ with Choi states $\frac{1}{2^n}\Gamma_i$. Hence, any query complexity lower bound for achieving the latter task without a quantum memory directly implies the same query complexity lower bound for the former task under the same restrictions on the algorithm. Thus, we now establish a query complexity lower bound for performing the many-versus-one distinguishing task without quantum memory.

Consider the tree representation $\mathcal{T}$ of the learning algorithm. The probability of arriving at a leaf ℓ when having access to $\mathcal{N}$ is

$$p^{\mathcal{N}}(\ell) = \prod_{t=1}^{T} w_{v_t} 2^n 2^{n_{\mathrm{aux}}} \left\langle \psi_{v_t} \right| (\mathrm{id}_{\mathrm{aux}} \otimes \mathcal{N})(\left|\phi_{v_{t-1}}\right\rangle\!\left\langle\phi_{v_{t-1}}\right|) \left|\psi_{v_t}\right\rangle . \tag{D.36}$$

Using that, for every bipartite state $\rho \in \mathcal{S}((\mathbb{C}^2)^{n_{\mathrm{aux}}} \otimes (\mathbb{C}^2)^{n_{\mathrm{in}}})$, we have

$$(\mathrm{id}_{\mathrm{aux}} \otimes \mathcal{N})(\rho) = 2^n\, \mathrm{tr}_{\mathrm{in}}\left[(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})\left(\mathbb{1}_{\mathrm{aux}} \otimes \frac{1}{2^n}\Gamma^{\mathcal{N}}_{\mathrm{in,out}}\right)\right], \tag{D.37}$$

we can compute

$$(\mathrm{id}_{\mathrm{aux}} \otimes \mathcal{N}_i)(\rho) = \frac{\rho_{\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}} + 3\varepsilon\, \mathrm{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes O_i)]}{2^n} \tag{D.38}$$

as well as

$$(\mathrm{id}_{\mathrm{aux}} \otimes \mathcal{N}_{\mathrm{max-mixed}})(\rho) = \frac{\rho_{\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}}}{2^n}. \tag{D.39}$$

Using these expressions as well as the notational shorthands $\rho_{v_{t-1}} = |\phi_{v_{t-1}}\rangle\langle\phi_{v_{t-1}}|$ and $p_t = \langle\psi_{v_t}|\, \rho_{v_{t-1},\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}}\, |\psi_{v_t}\rangle$, we can now perform the following calculation:

$$\begin{aligned}
&\frac{\mathbb{E}_{i\sim\mathrm{Uniform}(\{1,\ldots,M\})}[p^{\mathcal{N}_i}(\ell)]}{p^{\mathcal{N}_{\mathrm{max-mixed}}}(\ell)} && \text{(D.40)}\\
&= \mathbb{E}_{i\sim\mathrm{Uniform}(\{1,\ldots,M\})}\left[\prod_{t=1}^{T} \frac{w_{v_t} 2^n 2^{n_{\mathrm{aux}}}\left(p_t + 3\varepsilon\,\langle\psi_{v_t}|\,\mathrm{tr}_{\mathrm{in}}[(\rho_{v_{t-1}}^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes O_i)]\,|\psi_{v_t}\rangle\right)}{w_{v_t} 2^n 2^{n_{\mathrm{aux}}} p_t}\right] && \text{(D.41)}\\
&= \mathbb{E}_{i\sim\mathrm{Uniform}(\{1,\ldots,M\})}\left[\exp\left(\sum_{t=1}^{T} \log\left(1 + 3\varepsilon \frac{\langle\psi_{v_t}|\,\mathrm{tr}_{\mathrm{in}}[(\rho_{v_{t-1}}^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes O_i)]\,|\psi_{v_t}\rangle}{p_t}\right)\right)\right] && \text{(D.42)}\\
&\geq \exp\left(\sum_{t=1}^{T} \mathbb{E}_{i\sim\mathrm{Uniform}(\{1,\ldots,M\})}\left[\log\left(1 + 3\varepsilon \frac{\langle\psi_{v_t}|\,\mathrm{tr}_{\mathrm{in}}[(\rho_{v_{t-1}}^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes O_i)]\,|\psi_{v_t}\rangle}{p_t}\right)\right]\right) && \text{(D.43)}\\
&= \exp\left(\sum_{t=1}^{T} \frac{1}{M} \sum_{i-1}^{M/2} \log\left(1 - 9\varepsilon^2 \left(\frac{\langle\psi_{v_t}|\,\mathrm{tr}_{\mathrm{in}}[(\rho_{v_{t-1}}^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes O_i)]\,|\psi_{v_t}\rangle}{p_t}\right)^2\right)\right) && \text{(D.44)}\\
&\geq \exp\left(-\sum_{t=1}^{T} \frac{18}{M} \sum_{i-1}^{M/2} \varepsilon^2 \left(\frac{\langle\psi_{v_t}|\,\mathrm{tr}_{\mathrm{in}}[(\rho_{v_{t-1}}^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes O_i)]\,|\psi_{v_t}\rangle}{p_t}\right)^2\right) && \text{(D.45)}\\
&\geq \exp\left(-9T\varepsilon^2 \Delta(O_1,\ldots,O_M)\right). && \text{(D.46)}
\end{aligned}$$

Here, the first two steps are simple rewritings. The third step is by Jensen's inequality. The fourth step uses condition (iii). The fifth step uses the numerical inequality $\log(1-x) \geq -2x$, which is valid for $x \in [0, 0.79]$, and which we can apply for $\varepsilon > 0$ small enough. The final step holds by definition of $\Delta(O_1,\ldots,O_M)$. With this, we have shown that

$$\frac{\mathbb{E}_{i\sim\mathrm{Uniform}(\{1,\ldots,M\})}[p^{\mathcal{N}_i}(\ell)]}{p^{\mathcal{N}_{\mathrm{max-mixed}}}(\ell)} \geq \exp\left(-9T\varepsilon^2\Delta(O_1,\ldots,O_M)\right) \geq 1 - 9T\varepsilon^2\Delta(O_1,\ldots,O_M). \tag{D.47}$$

Using the one-sided version of Le Cam's two-point method [25, Lemma 5.4], this tells us that the learning algorithm without quantum memory solves the many-versus-one distinguishing task correctly with probability at most $9T\varepsilon^2\Delta(O_1,\ldots,O_M)$. Thus, to achieve a success probability $\geq 2/3$, the number of queries has to satisfy

$$T \geq \Omega\left(\frac{1}{\varepsilon^2\Delta(O_1,\ldots,O_M)}\right). \tag{D.48}$$

By the reduction between learning and distinguishing discussed at the beginning of the proof, this establishes the claimed query complexity lower bound. □

By slightly varying the assumptions on the observables, we get the following version for doubly-stochastic and entanglement-breaking channels with sparse PTM.

Lemma D.6. *Let* $O_1, \dots, O_M \in \mathcal{B}((\mathbb{C}^2)^{\otimes 2n})$ *be* $M \in 2\mathbb{N}$ *traceless and self-adjoint* $(2n)$*-qubit observables satisfying the following properties:*

(i) For every $1 \leq i \leq M$, $\sigma(O_i) \subseteq \{-1, 1\}$. *In particular,* $\|O_i\| = 1$ *holds for all* $1 \leq i \leq M$.
(ii) For every $1 \leq i \leq M$, $\mathrm{tr}_{1,\dots,n}[O_i] = 0$ *and* $\mathrm{tr}_{n+1,\dots,2n}[O_i] = 0$.
(iii) For every $1 \leq i \leq M/2$, $O_i = -O_{i+M/2}$.
(iv) For every $1 \leq i \leq M$, O_i *has a* $\mathcal{O}(1)$*-sparse Pauli ONB expansion.*
(v) For every $1 \leq i \leq M$, O_i *is a tensor product of* $2n$ *single-qubit observables.*

Any algorithm for learning without quantum memory requires

$$\Omega\left(\frac{1}{\varepsilon^2 \Delta(O_1, \dots, O_M)}\right) \tag{D.49}$$

queries to a doubly-stochastic and entanglement-breaking n*-qubit channel* $\mathcal{N}$ *with* $\mathcal{O}(1)$*-sparse PTM to simultaneously predict all expectation values* $\mathrm{tr}[O_i \frac{1}{2^n}\Gamma]$, $1 \leq i \leq M$, *up to accuracy* ε *with a success probability* $\geq \frac{2}{3}$. *Here, we again defined*

$$\Delta(O_1, \dots, O_M) := \sup_{\substack{n_{\mathrm{aux}} \in \mathbb{N} \\ |\phi\rangle \in (\mathbb{C}^2)^{\otimes n_{\mathrm{aux}}} \otimes (\mathbb{C}^2)^{\otimes n} : \|\phi\| = 1 \\ \rho \in \mathcal{S}((\mathbb{C}^2)^{\otimes n_{\mathrm{aux}}} \otimes (\mathbb{C}^2)^{\otimes n})}} \frac{2}{M} \sum_{i-1}^{M/2} \left(\frac{\langle\phi| \mathrm{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes O_i)] |\phi\rangle}{\langle\phi| (\rho_{\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}}) |\phi\rangle}\right)^2. \tag{D.50}$$

Proof. The proof follows the same steps as the proof of Lemma D.5, we only have to notice that, with the changed condition (ii) and the added conditions (iv) and (v), all the $\frac{1}{2^n}\Gamma_i$ as well as the maximally mixed state are valid Choi states of doubly-stochastic and entanglement-breaking quantum channels with $\mathcal{O}(1)$-sparse PTM. □

As before, to make use of these lower bounds, we need to evaluate (or at least bound) the quantity $\Delta(O_1, \dots, O_M)$ for sets of observables of interest. We will do so for the same sets of observables that we have already considered in Lemmas D.3 and D.4.

Lemma D.7. *Consider the following* $2 \cdot 4^n(4^n - 1)$ *many* $(2n)$*-qubit Pauli observables*

$$P_1, \dots, P_{4^n(4^n-1)} \in \{+\sigma_A\}_{A \in \{0,1,2,3\}^{2n}} \setminus \{+\sigma_B \otimes \mathbb{1}_2^{\otimes n}\}_{B \in \{0,1,2,3\}^n} \tag{D.51}$$

$$P_{4^n(4^n-1)+1}, \dots, P_{2\cdot 4^n(4^n-1)} \in \{-\sigma_A\}_{A \in \{0,1,2,3\}^{2n}} \setminus \{-\sigma_B \otimes \mathbb{1}_2^{\otimes n}\}_{B \in \{0,1,2,3\}^n}. \tag{D.52}$$

The set of observables $\{P_i\}_{i=1}^{2\cdot 4^n(4^n-1)}$, *when suitably ordered, satisfies conditions (i) through (iii) from Lemma D.5. Moreover,*

$$\Delta(P_1, \dots, P_{2\cdot 4^n(4^n-1)}) = \frac{1}{4^n - 1}. \tag{D.53}$$

Proof. We have already established conditions (i) through (iii) when proving Lemma D.3. We upper bound $\Delta(P_1, \dots, P_{2\cdot 4^n(4^n-1)})$ as follows. Let $n_{\mathrm{aux}} \in \mathbb{N}$, $|\phi\rangle \in (\mathbb{C}^2)^{\otimes n_{\mathrm{aux}}} \otimes (\mathbb{C}^2)^{\otimes n}$ with $\|\phi\| = 1$, and let $\rho \in \mathcal{S}((\mathbb{C}^2)^{\otimes n_{\mathrm{aux}}} \otimes (\mathbb{C}^2)^{\otimes n})$ be arbitrary. Then, using Equation (D.16) similarly to before, we obtain the following:

$$\frac{1}{4^n(4^n - 1)} \sum_{i=1}^{4^n(4^n-1)} \langle\phi| \mathrm{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes P_i)] |\phi\rangle^2 \tag{D.54}$$

$$= \frac{1}{4^n(4^n-1)} \sum_{i=1}^{4^n(4^n-1)} \operatorname{tr}\left[\operatorname{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes P_i)]\,|\phi\rangle\langle\phi|\right]^2 \tag{D.55}$$

$$= \frac{1}{4^n(4^n-1)} \sum_{i=1}^{4^n(4^n-1)} \operatorname{tr}\left[\left((\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes P_i)\right)(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\right]^2 \tag{D.56}$$

$$= \frac{1}{4^n(4^n-1)} \operatorname{tr}\left[\left(\sum_{i=1}^{4^n(4^n-1)} (\mathbb{1}_{\mathrm{aux}} \otimes P_i) \otimes (\mathbb{1}_{\mathrm{aux}} \otimes P_i)\right)\left((\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})\right)^{\otimes 2}\right] \tag{D.57}$$

$$= \frac{1}{4^n(4^n-1)} \Big(\operatorname{tr}\left[2^{2n}\,\mathrm{SWAP}_{\mathrm{in,out}}(\operatorname{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})])^{\otimes 2}\right] \tag{D.58}$$

$$- \operatorname{tr}\left[2^{n}\,\mathrm{SWAP}_{\mathrm{in}}(\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})])^{\otimes 2}\right]\Big) \tag{D.59}$$

$$= \frac{1}{4^n(4^n-1)} \Big(2^{2n} \operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})])^2\right] \tag{D.60}$$

$$- 2^n \operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})])^2\right]\Big)\,. \tag{D.61}$$

Therefore, we have

$$\frac{1}{4^n(4^n-1)} \sum_{i=1}^{4^n(4^n-1)} \left(\frac{\langle\phi|\operatorname{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes P_i)]\,|\phi\rangle}{\langle\phi|(\rho_{\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}})|\phi\rangle}\right)^2 \tag{D.62}$$

$$= \frac{1}{4^n(4^n-1)} \Bigg(2^{2n} \frac{\operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})])^2\right]}{\langle\phi|(\rho_{\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}})|\phi\rangle^2} \tag{D.63}$$

$$- 2^n \frac{\operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})])^2\right]}{\langle\phi|(\rho_{\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}})|\phi\rangle^2}\Bigg)\,. \tag{D.64}$$

Notice that

$$\operatorname{tr}\left[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})\right] = \langle\phi|(\rho_{\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}})|\phi\rangle = \operatorname{tr}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)(\rho \otimes \mathbb{1}_{\mathrm{out}})], \tag{D.65}$$

since $\operatorname{tr}_{\mathrm{in}}[\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}}] = \rho_{\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}}$. As the partial trace of an operator has the same trace as the original operator, we can thus rewrite the preceding as follows:

$$\frac{1}{4^n(4^n-1)} \sum_{i=1}^{4^n(4^n-1)} \left(\frac{\langle\phi|\operatorname{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes P_i)]\,|\phi\rangle}{\langle\phi|(\rho_{\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}})|\phi\rangle}\right)^2 \tag{D.66}$$

$$= \frac{1}{4^n(4^n-1)} \left(2^{2n} \frac{\operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})])^2\right]}{\operatorname{tr}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)(\rho \otimes \mathbb{1}_{\mathrm{out}})]^2} - 2^n\right). \tag{D.67}$$

It remains to bound the first summand. We first rewrite it using the identity $\operatorname{tr}_A[XY] = \operatorname{tr}_A[X^{\top_A} Y^{\top_A}]\ \forall X, Y \in \mathcal{B}(\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B})$ together with the identity $(\theta_{\mathrm{in}} \otimes \mathrm{id}_{\mathrm{out}}) \circ \operatorname{tr}_{\mathrm{aux}} = \operatorname{tr}_{\mathrm{aux}} \circ (\theta_{\mathrm{in}} \otimes \mathrm{id}_{\mathrm{aux,out}})$ for the transpose map θ_{in} viewed as a linear superoperator on $\mathcal{B}((\mathbb{C}^2)^{\otimes n})$, which allows us to replace $\rho^{\top_{\mathrm{in}}}$ by ρ in the enumerator. (We explain this in full detail in the proof of Lemma D.8.) Then, we do a direct computation. Suppose $|\phi\rangle_{\mathrm{aux,out}}$ has Schmidt decomposition $|\phi\rangle_{\mathrm{aux,out}} = \sum_{i=1}^{\min\{2^{n_{\mathrm{aux}}}, 2^n\}} \sqrt{\lambda_i}\,|e_i\rangle_{\mathrm{aux}} \otimes |f_i\rangle_{\mathrm{out}}$, where the $\lambda_i \geq 0$ satisfy $\sum_{i=1}^{\min\{2^{n_{\mathrm{aux}}}, 2^n\}} \lambda_i = 1$, and where $\{|e_i\rangle\}_{i=1}^{2^{n_{\mathrm{aux}}}}$ and $\{|f_i\rangle\}_{i=1}^{2^n}$ are ONBs of $(\mathbb{C}^2)^{\otimes n_{\mathrm{aux}}}$ and $(\mathbb{C}^2)^{\otimes n}$, respectively. As $n_{\mathrm{in}} = n = n_{\mathrm{out}}$, we can expand $\rho_{\mathrm{in,aux}}$ w.r.t. the tensor product basis $\{|f_i\rangle\langle f_j| \otimes |e_k\rangle\langle e_\ell|\}_{i,j,k,\ell}$ of

$\mathcal{B}((\mathbb{C}^2)^{\otimes n} \otimes (\mathbb{C}^2)^{\otimes n_{\mathrm{aux}}})$ as

$$\rho_{\mathrm{in,aux}} = \sum_{i,j,k,\ell} \rho_{ijk\ell} \left|f_i\middle\rangle\middle\langle f_j\right|_{\mathrm{in}} \otimes |e_k\rangle\langle e_\ell|_{\mathrm{aux}} . \tag{D.68}$$

Using these explicit expansions for $|\phi\rangle$ and ρ, we can compute

$$(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\rho \otimes \mathbb{1}_{\mathrm{out}})]) = \sum_{i,j,k,\ell,n} \sqrt{\lambda_i\lambda_j}\rho_{k\ell jn} \,|f_k\rangle\langle f_\ell|_{\mathrm{in}} \otimes |e_i\rangle\langle e_n|_{\mathrm{aux}} \otimes \left|f_i\middle\rangle\middle\langle f_j\right|_{\mathrm{out}} , \tag{D.69}$$

leading to

$$\mathrm{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\rho \otimes \mathbb{1}_{\mathrm{out}})] = \sum_{i,j,k,\ell} \sqrt{\lambda_i\lambda_j}\rho_{k\ell ji} \,|f_k\rangle\langle f_\ell|_{\mathrm{in}} \otimes \left|f_i\middle\rangle\middle\langle f_j\right|_{\mathrm{out}} . \tag{D.70}$$

From this, we obtain

$$\begin{aligned}
\mathrm{tr}\left[\mathrm{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})]\right]^2 &= \mathrm{tr}\left[\sum_{i,j,k,\ell} \sqrt{\lambda_i\lambda_j}\rho_{k\ell ij} \,|f_k\rangle\langle f_\ell|_{\mathrm{in}} \otimes \left|f_i\middle\rangle\middle\langle f_j\right|_{\mathrm{out}}\right]^2 &\text{(D.71)}\\
&= \left(\sum_{i,\ell} \lambda_i \rho_{\ell\ell ii}\right)^2 &\text{(D.72)}\\
&= \sum_{i,j,k,\ell} \lambda_i\lambda_j \rho_{\ell\ell ii}\rho_{kkjj} &\text{(D.73)}
\end{aligned}$$

as well as

$$\begin{aligned}
&\mathrm{tr}\left[\left(\mathrm{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})]\right)^2\right] &\text{(D.74)}\\
&= \mathrm{tr}\left[\sum_{i,j,k,\ell,m,n,s,t} \sqrt{\lambda_i\lambda_j\lambda_s\lambda_t}\rho_{k\ell ji}\rho_{mnts} \,|k\rangle\langle\ell|_{\mathrm{in}} \cdot |m\rangle\langle n|_{\mathrm{in}} \otimes |i\rangle\langle j|_{\mathrm{out}} \cdot |s\rangle\langle t|_{\mathrm{out}}\right] &\text{(D.75)}\\
&= \sum_{i,j,k,\ell} \lambda_i\lambda_j \rho_{\ell kij}\rho_{k\ell ji}. &\text{(D.76)}
\end{aligned}$$

As ρ is Hermitian and positive semidefinite, we have, for any i,j,k,ℓ, $\rho_{\ell kij} = \overline{\rho_{k\ell ji}}$ and $0 \leq \det\begin{pmatrix} \rho_{kkii} & \rho_{k\ell ij} \\ \overline{\rho_{k\ell ji}} & \rho_{\ell\ell ii}\end{pmatrix} = \rho_{\ell\ell ii}\rho_{kkjj} - |\rho_{\ell kij}|^2 = \rho_{\ell\ell ii}\rho_{kkjj} - \rho_{\ell kij}\rho_{k\ell ji}$. The latter holds because the matrix $\begin{pmatrix} \rho_{kkii} & \rho_{k\ell ij} \\ \overline{\rho_{k\ell ji}} & \rho_{\ell\ell ii}\end{pmatrix}$ is a principal submatrix of the positive semidefinite matrix ρ and thus itself positive semidefinite. Altogether, we now have the following:

$$\begin{aligned}
\mathrm{tr}\left[\left(\mathrm{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})]\right)^2\right] &= \sum_{i,j,k,\ell} \underbrace{\lambda_i\lambda_j}_{\geq 0}\; \underbrace{\rho_{\ell kij}\rho_{k\ell ji}}_{\leq \rho_{\ell\ell ii}\rho_{kkjj}} &\text{(D.77)}\\
&\leq \sum_{i,j,k,\ell} \lambda_i\lambda_j \rho_{\ell\ell ii}\rho_{kkjj} &\text{(D.78)}\\
&= \mathrm{tr}\left[\mathrm{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})]\right]^2 . &\text{(D.79)}
\end{aligned}$$

Plugging this inequality back into our expression of interest, we have shown

$$\frac{1}{4^n(4^n-1)}\sum_{i=1}^{4^n(4^n-1)}\left(\frac{\langle\phi|\,\mathrm{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}}\otimes P_i)]\,|\phi\rangle}{\langle\phi|\,(\rho_{\mathrm{aux}}\otimes\mathbb{1}_{\mathrm{out}})\,|\phi\rangle}\right)^2 \leq \frac{2^{2n}-2^n}{4^n(4^n-1)} \tag{D.80}$$

$$= \frac{2^n-1}{2^n(4^n-1)} \tag{D.81}$$

$$\leq \frac{1}{4^n-1}. \tag{D.82}$$

To finish the proof, it only remains to notice that equality can be attained, for example, if $|\phi\rangle_{\mathrm{aux,out}}$ and $\rho_{\mathrm{in,aux}}$ factorize as $|\phi\rangle_{\mathrm{aux,out}} = |\psi\rangle_{\mathrm{aux}}\otimes|\varphi\rangle_{\mathrm{out}}$ and $\rho_{\mathrm{in,aux}} = |\varphi\rangle\langle\varphi|_{\mathrm{in}}\otimes|\psi\rangle\langle\psi|_{\mathrm{aux}}$. ☐

Lemma D.8. *Consider the following* $2(4^{2n}-2\cdot 4^n+1) = 2(4^n-1)^2$ *many* $(2n)$*-qubit Pauli observables*

$$P_1,\dots,P_{(4^n-1)^2} \in \{+\sigma_A\}_{A\in\{0,1,2,3\}^{2n}:\exists 1\leq i\leq n,\, n+1\leq j\leq 2n: A_i\neq 0\neq A_j} \tag{D.83}$$

$$P_{4^{2n}-2\cdot 4^n+2},\dots,P_{2(4^n-1)^2} \in \{-\sigma_A\}_{A\in\{0,1,2,3\}^{2n}:\exists 1\leq i\leq n,\, n+1\leq j\leq 2n: A_i\neq 0\neq A_j}. \tag{D.84}$$

The set of observables $\{P_i\}_{i=1}^{2(4^n-1)^2}$*, when suitably ordered, satisfies conditions (i) through (v) from Lemma D.6. Moreover,*

$$\frac{1}{(2^n+1)^2} \leq \Delta(P_1,\dots,P_{2((4^n-1)^2)}) \leq \frac{1}{(2^n-1)^2}. \tag{D.85}$$

Proof. We have already established conditions (i) through (v) when proving Lemma D.4. We bound $\Delta(P_1,\dots,P_{2(4^n-1)^2})$ as follows. Let $n_{\mathrm{aux}}\in\mathbb{N}$, $|\phi\rangle\in(\mathbb{C}^2)^{\otimes n_{\mathrm{aux}}}\otimes(\mathbb{C}^2)^{\otimes n}$ with $\|\phi\| = 1$, and $\rho\in\mathcal{S}((\mathbb{C}^2)^{\otimes n_{\mathrm{aux}}}\otimes(\mathbb{C}^2)^{\otimes n})$ be arbitrary. Then, using Equation (D.16) similarly to before, we obtain the following:

$$\frac{1}{(4^n-1)^2}\sum_{i=1}^{(4^n-1)^2}\langle\phi|\,\mathrm{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}}\otimes P_i)]\,|\phi\rangle^2 \tag{D.86}$$

$$= \frac{1}{(4^n-1)^2}\sum_{i=1}^{(4^n-1)^2}\mathrm{tr}\left[\mathrm{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}}\otimes P_i)]\,|\phi\rangle\langle\phi|\right]^2 \tag{D.87}$$

$$= \frac{1}{(4^n-1)^2}\sum_{i=1}^{(4^n-1)^2}\mathrm{tr}\left[\left((\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}}\otimes P_i)\right)(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\right]^2 \tag{D.88}$$

$$= \frac{1}{(4^n-1)^2}\,\mathrm{tr}\left[\left(\sum_{i=1}^{(4^n-1)^2}(\mathbb{1}_{\mathrm{aux}}\otimes P_i)\otimes(\mathbb{1}_{\mathrm{aux}}\otimes P_i)\right)\left((\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})\right)^{\otimes 2}\right] \tag{D.89}$$

$$= \frac{1}{(4^n-1)^2}\Bigg(\mathrm{tr}\left[2^{2n}\,\mathrm{SWAP}_{\mathrm{in,out}}(\mathrm{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})])^{\otimes 2}\right] \tag{D.90}$$

$$-\,\mathrm{tr}\left[2^n\,\mathrm{SWAP}_{\mathrm{in}}(\mathrm{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})])^{\otimes 2}\right] \tag{D.91}$$

$$-\,\mathrm{tr}\left[2^n\,\mathrm{SWAP}_{\mathrm{out}}(\mathrm{tr}_{\mathrm{aux,in}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})])^{\otimes 2}\right] \tag{D.92}$$

$$+\,\mathrm{tr}\left[(\mathbb{1}_{\mathrm{aux}}\otimes\mathbb{1}_{\mathrm{in,out}})(\mathrm{tr}_{\mathrm{aux,in}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})])^{\otimes 2}\right]\Bigg) \tag{D.93}$$

$$= \frac{1}{(4^n-1)^2}\Bigg(2^{2n}\,\operatorname{tr}\left[\left(\operatorname{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})]\right)^2\right] \tag{D.94}$$

$$- 2^n\,\operatorname{tr}\left[\left(\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})]\right)^2\right] \tag{D.95}$$

$$- 2^n\,\operatorname{tr}\left[\left(\operatorname{tr}_{\mathrm{aux,in}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})]\right)^2\right] \tag{D.96}$$

$$+ \operatorname{tr}\left[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})\right]^2\Bigg). \tag{D.97}$$

Therefore, we have

$$\frac{1}{(4^n-1)^2}\sum_{i=1}^{(4^n-1)^2}\left(\frac{\langle\phi|\operatorname{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}}\otimes P_i)]\,|\phi\rangle}{\langle\phi|\,(\rho_{\mathrm{aux}}\otimes\mathbb{1}_{\mathrm{out}})\,|\phi\rangle}\right)^2 \tag{D.98}$$

$$= \frac{1}{(4^n-1)^2}\Bigg(2^{2n}\,\frac{\operatorname{tr}\left[\left(\operatorname{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})]\right)^2\right]}{\langle\phi|\,(\rho_{\mathrm{aux}}\otimes\mathbb{1}_{\mathrm{out}})\,|\phi\rangle^2} \tag{D.99}$$

$$- 2^n\,\frac{\operatorname{tr}\left[\left(\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})]\right)^2\right]}{\langle\phi|\,(\rho_{\mathrm{aux}}\otimes\mathbb{1}_{\mathrm{out}})\,|\phi\rangle^2} \tag{D.100}$$

$$- 2^n\,\frac{\operatorname{tr}\left[\left(\operatorname{tr}_{\mathrm{aux,in}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})]\right)^2\right]}{\langle\phi|\,(\rho_{\mathrm{aux}}\otimes\mathbb{1}_{\mathrm{out}})\,|\phi\rangle^2} \tag{D.101}$$

$$+ \frac{\operatorname{tr}\left[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})\right]^2}{\langle\phi|\,(\rho_{\mathrm{aux}}\otimes\mathbb{1}_{\mathrm{out}})\,|\phi\rangle^2}\Bigg). \tag{D.102}$$

Notice that

$$\operatorname{tr}\left[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})\right] = \langle\phi|\,(\rho_{\mathrm{aux}}\otimes\mathbb{1}_{\mathrm{out}})\,|\phi\rangle = \operatorname{tr}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})], \tag{D.103}$$

since $\operatorname{tr}_{\mathrm{in}}[\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}}] = \rho_{\mathrm{aux}}\otimes\mathbb{1}_{\mathrm{out}} = \operatorname{tr}_{\mathrm{in}}[\rho\otimes\mathbb{1}_{\mathrm{out}}]$. As the partial trace of an operator has the same trace as the original operator, we can thus rewrite the preceding as follows:

$$\frac{1}{(4^n-1)^2}\sum_{i=1}^{(4^n-1)^2}\left(\frac{\langle\phi|\operatorname{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}}\otimes P_i)]\,|\phi\rangle}{\langle\phi|\,(\rho_{\mathrm{aux}}\otimes\mathbb{1}_{\mathrm{out}})\,|\phi\rangle}\right)^2 \tag{D.104}$$

$$= \frac{1}{(4^n-1)^2}\Bigg(2^{2n}\,\frac{\operatorname{tr}\left[\left(\operatorname{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})]\right)^2\right]}{\operatorname{tr}\left[\operatorname{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})]\right]^2} \tag{D.105}$$

$$- 2^n\,\frac{\operatorname{tr}\left[\left(\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})]\right)^2\right]}{\operatorname{tr}\left[\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})]\right]^2} \tag{D.106}$$

$$- 2^n\,\frac{\operatorname{tr}\left[\left(\operatorname{tr}_{\mathrm{aux,in}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})]\right)^2\right]}{\operatorname{tr}\left[\operatorname{tr}_{\mathrm{aux,in}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\!\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})]\right]^2} + 1\Bigg). \tag{D.107}$$

Next we use the identity

$$\operatorname{tr}[XY] = \operatorname{tr}\left[X^{\top_A}Y^{\top_A}\right], \quad \forall X, Y \in \mathcal{B}(\mathbb{C}^{d_A}\otimes\mathbb{C}^{d_B}), \tag{D.108}$$

which can, for example, be checked by expanding in a tensor product basis. Let θ denote the transpose viewed as linear superoperator. We clearly have $(\theta_{\mathrm{in}}\otimes\mathrm{id}_{\mathrm{out}})\circ\operatorname{tr}_{\mathrm{aux}} = \operatorname{tr}_{\mathrm{aux}}\circ(\theta_{\mathrm{in}}\otimes\mathrm{id}_{\mathrm{aux,out}})$.

Combining this with Equation (D.108), we can rewrite the enumerator in the first summand above as

$$
\begin{aligned}
&\operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})])^2\right] &\text{(D.109)}\\
&=\operatorname{tr}\left[\left((\theta_{\mathrm{in}}\otimes\mathrm{id}_{\mathrm{out}})\left(\operatorname{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})]\right)\right)^2\right] &\text{(D.110)}\\
&=\operatorname{tr}\left[\left(\operatorname{tr}_{\mathrm{aux}}\left[(\theta_{\mathrm{in}}\otimes\mathrm{id}_{\mathrm{aux,out}})\left((\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})\right)\right]\right)^2\right] &\text{(D.111)}\\
&=\operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux}}\left[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})\right])^2\right]. &\text{(D.112)}
\end{aligned}
$$

Similarly, we can combine the identity $\operatorname{tr}[XY]=\operatorname{tr}[X^\top Y^\top]\ \forall X,Y\in\mathcal{B}(\mathcal{H})$ with the commutation $\theta_{\mathrm{in}}\circ\operatorname{tr}_{\mathrm{aux,out}}=\operatorname{tr}_{\mathrm{aux,out}}\circ(\theta_{\mathrm{in}}\otimes\mathrm{id}_{\mathrm{aux,out}})$ to rewrite the enumerator in the second summand above as

$$
\begin{aligned}
&\operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})])^2\right] &\text{(D.113)}\\
&=\operatorname{tr}\left[\left(\theta_{\mathrm{in}}(\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})])\right)^2\right] &\text{(D.114)}\\
&=\operatorname{tr}\left[\left(\operatorname{tr}_{\mathrm{aux,out}}\left[(\theta_{\mathrm{in}}\otimes\mathrm{id}_{\mathrm{aux,out}})\left((\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})\right)\right]\right)^2\right] &\text{(D.115)}\\
&=\operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})])^2\right]. &\text{(D.116)}
\end{aligned}
$$

Finally, using the identity $\operatorname{tr}_A[XY]=\operatorname{tr}_A[X^{\top_A}Y^{\top_A}]\ \forall X,Y\in\mathcal{B}(\mathbb{C}^{d_A}\otimes\mathbb{C}^{d_B})$, we can rewrite the enumerator in the third summand above as

$$
\begin{aligned}
&\operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux,in}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})])^2\right] &\text{(D.117)}\\
&=\operatorname{tr}\left[\left(\operatorname{tr}_{\mathrm{aux,in}}\left[(\theta_{\mathrm{in}}\otimes\mathrm{id}_{\mathrm{aux,out}})\left((\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})\right)\right]\right)^2\right] &\text{(D.118)}\\
&=\operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux,in}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})])^2\right]. &\text{(D.119)}
\end{aligned}
$$

In summary, we now have the following expression:

$$
\begin{aligned}
&\frac{1}{(4^n-1)^2}\sum_{i=1}^{(4^n-1)^2}\left(\frac{\langle\phi|\operatorname{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}}\otimes\mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}}\otimes P_i)]\,|\phi\rangle}{\langle\phi|\,(\rho_{\mathrm{aux}}\otimes\mathbb{1}_{\mathrm{out}})\,|\phi\rangle}\right)^2 &\text{(D.120)}\\
&=\frac{1}{(4^n-1)^2}\left(2^{2n}\frac{\operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})])^2\right]}{\operatorname{tr}\left[\operatorname{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})]\right]^2}\right. &\text{(D.121)}\\
&\qquad -2^n\frac{\operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})])^2\right]}{\operatorname{tr}\left[\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})]\right]^2} &\text{(D.122)}\\
&\qquad \left.-2^n\frac{\operatorname{tr}\left[(\operatorname{tr}_{\mathrm{aux,in}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})])^2\right]}{\operatorname{tr}\left[\operatorname{tr}_{\mathrm{aux,in}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})]\right]^2}+1\right). &\text{(D.123)}
\end{aligned}
$$

To simplify the second and third summand further, we use that the partial trace is cyclic w.r.t. operators that act non-trivially only on the system(s) being traced over. This allows us to do the following rewritings. As $|\phi\rangle\langle\phi|=(|\phi\rangle\langle\phi|)^2$, we get

$$
\begin{aligned}
\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})]&=\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)^2\,(\rho\otimes\mathbb{1}_{\mathrm{out}})] &\text{(D.124)}\\
&=\operatorname{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)\,(\rho\otimes\mathbb{1}_{\mathrm{out}})\,(\mathbb{1}_{\mathrm{in}}\otimes|\phi\rangle\langle\phi|)]. &\text{(D.125)}
\end{aligned}
$$

And as $\rho = \sqrt{\rho}^2$, we get

$$\begin{aligned}\mathrm{tr}_{\mathrm{aux,in}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\rho \otimes \mathbb{1}_{\mathrm{out}})] &= \mathrm{tr}_{\mathrm{aux,in}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\sqrt{\rho} \otimes \mathbb{1}_{\mathrm{out}})^2] &\text{(D.126)}\\ &= \mathrm{tr}_{\mathrm{aux,in}}[(\sqrt{\rho} \otimes \mathbb{1}_{\mathrm{out}})\,(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\sqrt{\rho} \otimes \mathbb{1}_{\mathrm{out}})]. &\text{(D.127)}\end{aligned}$$

Therefore, both $\mathrm{tr}_{\mathrm{aux,out}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\rho \otimes \mathbb{1}_{\mathrm{out}})]$ and $\mathrm{tr}_{\mathrm{aux,in}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\rho \otimes \mathbb{1}_{\mathrm{out}})]$ are partial traces of positive semidefinite operators and thus themselves positive semidefinite. As the inequality $|\mathrm{tr}[X^2]| = \mathrm{tr}[X^2] \leq (\mathrm{tr}[X])^2$ holds for any positive semidefinite X, we now arrive at the following upper bound:

$$\begin{aligned}&\frac{1}{(4^n-1)^2}\sum_{i=1}^{(4^n-1)^2}\left(\frac{\langle\phi|\,\mathrm{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes P_i)]\,|\phi\rangle}{\langle\phi|\,(\rho_{\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}})\,|\phi\rangle}\right)^2 &\text{(D.128)}\\ &\leq \frac{1}{(4^n-1)^2}\left(2^{2n}\frac{\mathrm{tr}\left[(\mathrm{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\rho \otimes \mathbb{1}_{\mathrm{out}})])^2\right]}{\mathrm{tr}\left[\mathrm{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\rho \otimes \mathbb{1}_{\mathrm{out}})]\right]^2} + 2\cdot 2^n + 1\right). &\text{(D.129)}\end{aligned}$$

It remains to recall from the proof of Lemma D.7 that

$$\mathrm{tr}\left[(\mathrm{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\rho \otimes \mathbb{1}_{\mathrm{out}})])^2\right] \leq \mathrm{tr}\left[\mathrm{tr}_{\mathrm{aux}}[(\mathbb{1}_{\mathrm{in}} \otimes |\phi\rangle\langle\phi|)\,(\rho \otimes \mathbb{1}_{\mathrm{out}})]\right]^2. \tag{D.130}$$

Therefore, we have shown the upper bound:

$$\begin{aligned}\frac{1}{(4^n-1)^2}\sum_{i=1}^{(4^n-1)^2}\left(\frac{\langle\phi|\,\mathrm{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes P_i)]\,|\phi\rangle}{\langle\phi|\,(\rho_{\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}})\,|\phi\rangle}\right)^2 &\leq \frac{2^{2n} + 2\cdot 2^n + 1}{(4^n-1)^2} &\text{(D.131)}\\ &= \frac{(2^n+1)^2}{(4^n-1)^2} &\text{(D.132)}\\ &= \frac{1}{(2^n-1)^2}. &\text{(D.133)}\end{aligned}$$

This finishes the proof for the upper bound. To establish the lower bound, we can, for example, consider the case in which $|\phi\rangle_{\mathrm{aux,out}}$ and $\rho_{\mathrm{in,aux}}$ factorize as $|\phi\rangle_{\mathrm{aux,out}} = |\psi\rangle_{\mathrm{aux}} \otimes |\varphi\rangle_{\mathrm{out}}$ and $\rho_{\mathrm{in,aux}} = |\varphi\rangle\langle\varphi|_{\mathrm{in}} \otimes |\psi\rangle\langle\psi|_{\mathrm{aux}}$. Plugging these choices into our exact expression above, we get

$$\begin{aligned}\frac{1}{(4^n-1)^2}\sum_{i=1}^{(4^n-1)^2}\left(\frac{\langle\phi|\,\mathrm{tr}_{\mathrm{in}}[(\rho^{\top_{\mathrm{in}}} \otimes \mathbb{1}_{\mathrm{out}})(\mathbb{1}_{\mathrm{aux}} \otimes P_i)]\,|\phi\rangle}{\langle\phi|\,(\rho_{\mathrm{aux}} \otimes \mathbb{1}_{\mathrm{out}})\,|\phi\rangle}\right)^2 &= \frac{2^{2n} - 2\cdot 2^n + 1}{(4^n-1)^2} &\text{(D.134)}\\ &= \frac{(2^n-1)^2}{(4^n-1)^2} &\text{(D.135)}\\ &= \frac{1}{(2^n+1)^2}. &\text{(D.136)}\end{aligned}$$

This concludes the proof. □

Having established these lemmata, we can complete the proof of Theorems 5.6 and 5.7.

PROOF OF THEOREMS 5.6 AND 5.7. Recalling from the proof of Theorem 4.1 that estimating all entries of $R_{\mathcal{N}}^{\mathcal{P}}$ up to accuracy ε is equivalent to estimating all expectation values $\mathrm{tr}[\sigma_A \frac{1}{2^n}\Gamma^{\mathcal{N}}]$, $A \in \{0,1,2,3\}^{2n}$, up to accuracy ε, we can now combine Lemmas D.5 and D.7 to obtain Theorem 5.6, and combine Lemmas D.2 and D.8 to obtain Theorem 5.7. □

Remark D.9. A natural alternative proof strategy for establishing Theorems 5.6 and 5.7 would be to reduce them to their respective Choi access analogues. It is, however, not immediately obvious how to obtain such a reduction information-theoretically efficiently because protocols like implementation of a channel via Choi state teleportation have a success probability that is exponentially small in the number of qubits. While this success probability can be boosted if local unitary invariances of the Choi state are known in advance (compare, e.g., [81, Section 2.1]), we would need the unknown quantum channels to share the same invariances for this insight to be useful in a reduction for our lower bounds. We were unable to identify a suitable shared invariance for the Choi states used in our proofs and therefore resorted to a proof strategy other than a reduction to learning from Choi access. It would be interesting to see whether the proof of Chen et al. [29, Theorem 2, (iv)], which is based on teleportation stretching for Pauli channels, can be modified to obtain a similar reduction from general channel access to Choi access in our setting, as this may simplify our proofs.

D.3 Proof of Theorem 6.3

Here, we give the complete proof for our Hamiltonian learning guarantee in Theorem 6.3.

Proof of Theorem 6.3. We follow the strategy sketched in the main text. We choose the parameters for the polynomial interpolation as follows. We set the maximal evolution time to be $T = 1/\|H\|$ and the Chebyshev degree to be $L = \lceil 2\log(\frac{8\|H\|}{\sqrt{2\pi}\ln(2)\varepsilon})\rceil$. Additionally, we set our interpolation times to be $t_\ell = \frac{T}{2}(1 + z_\ell)$, where $z_\ell = -\cos(\frac{2\ell-1}{2L}\pi)$, for $1 \leq \ell \leq L$. Moreover, we set $\tilde{\varepsilon} = \frac{3T\varepsilon}{4(L-1)L(2L-1)}$.

As first part of the procedure, for each $1 \leq \ell \leq L$, we apply Corollary 4.2 to obtain a classical description $\hat{\mathcal{U}}_{t_\ell}$ of $\mathcal{U}_{t_\ell}$ that allows to predict Pauli-sparse expectation values, with success probability $\geq 1 - \delta$. This uses $\mathcal{O}(\frac{n+\log(1/\delta)}{\tilde{\varepsilon}^4})$ parallel queries to $\mathcal{U}_{t_\ell}$ for each $1 \leq \ell \leq L$. In the second part of the procedure, to estimate M Pauli coefficients $\alpha(A^{(i)})$, $A^{(i)} \in \{0,1,2,3\}^n \setminus \{0^n\}$, $1 \leq i \leq M$, we first use the classical representations $\hat{\mathcal{U}}_\ell$ to obtain $\tilde{\varepsilon}$-accurate estimates $2\hat{\alpha}_{t_\ell}^{(i)}$ of the expectation values $\operatorname{tr}\left[\sigma_{B^{(i)}}\mathcal{U}_{t_\ell}(\rho^{(i)})\right]$, with $B^{(i)}$ and $\rho^{(i)}$ as in Lemma 6.1. This uses $m_2 = \mathcal{O}(\frac{\log(M/\delta)}{\tilde{\varepsilon}^2})$ additional parallel queries to $\mathcal{U}_{t_\ell}$ as well as classical computation time $\mathcal{O}(\frac{n+\log(1/\delta)}{\tilde{\varepsilon}^4})$, for each $1 \leq \ell \leq L$. Next, we perform Chebyshev interpolation with these estimated values, which gives Chebyshev coefficients $\hat{b}_0^{(i)} = \frac{1}{L}\sum_{\ell=1}^{L}\hat{\alpha}_{t_\ell}^{(i)}$ and $\hat{b}_m^{(i)} = \frac{2}{L}\sum_{\ell=1}^{L}\hat{\alpha}_{t_\ell}^{(i)}T_m(z_\ell)$, $1 \leq m \leq L-1$, where T_m is the mth Chebyshev polynomial (of the first kind). These Chebyshev coefficients in turn give rise to the first-order Taylor coefficients $\hat{c}_1^{(i)} = -\frac{2}{T}\sum_{m=0}^{L-1}(-1)^m\hat{b}_m^{(i)}m^2$, $1 \leq i \leq M$, as shown, for example, in the work of Gu et al. [40, Lemma 6].

It remains to show that indeed $|\hat{c}_1^{(i)} - \alpha(A^{(i)})| \leq \varepsilon$ holds simultaneously for all $1 \leq i \leq M$, with success probability $\geq 1-\delta$. To see this, first recall that by Corollary 4.2, we know that, with success probability $\geq 1 - \delta$,

$$\left|2\hat{\alpha}_{t_\ell}^{(i)} - \operatorname{tr}\left[\sigma_{B^{(i)}}\mathcal{U}_{t_\ell}(\rho^{(i)})\right]\right| \leq \tilde{\varepsilon} \quad \forall 1 \leq i \leq M. \tag{D.137}$$

For the remainder of the proof, we condition on this success event. Consider "ideal" datasets

$$\left(t_1, \tfrac{1}{2}\operatorname{tr}\left[\sigma_{B^{(i)}}\mathcal{U}_{t_1}(\rho^{(i)})\right]\right), \ldots, \left(t_L, \tfrac{1}{2}\operatorname{tr}\left[\sigma_{B^{(i)}}\mathcal{U}_{t_L}(\rho^{(i)})\right]\right) \tag{D.138}$$

to which the algorithm does not have access, but which is a useful tool for our analysis. Let $b_m^{(i)}$ be the coefficients of the Chebyshev interpolation polynomial for that ideal dataset. Let $c_1^{(i)} = -\frac{2}{T}\sum_{m=0}^{L-1}(-1)^m b_m^{(i)} m^2$ be the corresponding estimates for the first-order Taylor coefficients. By the triangle inequality, $|\hat{c}_1^{(i)} - \alpha(A^{(i)})| \leq |\hat{c}_1^{(i)} - c_1^{(i)}| + |c_1^{(i)} - \alpha(A^{(i)})|$. The first of these two summands is easy to bound: as we are in the success event, we have $|\hat{b}_m^{(i)} - b_m^{(i)}| \leq 2\tilde{\varepsilon}$ for every $0 \leq m \leq L-1$ and for every $1 \leq i \leq M$. Therefore,

$$|\hat{c}_1^{(i)} - c_1^{(i)}| \leq \frac{2}{T}\sum_{m=0}^{L-1}|\hat{b}_m - b_m|m^2 \leq \frac{4\tilde{\varepsilon}}{T}\sum_{m=0}^{L-1} m^2 = \frac{2\tilde{\varepsilon}(L-1)L(2L-1)}{3T} = \frac{\varepsilon}{2}, \tag{D.139}$$

where the last inequality is by our choice of $\tilde{\varepsilon}$. For the second summand, we know—from Lemma 6.1 and from the definition of $c_1^{(i)}$—that $\alpha(A^{(i)})$ and $c_1^{(i)}$ are exactly the first-order Taylor coefficients of the smooth $\mathbb{R}$-valued function $t \mapsto \frac{1}{2}\operatorname{tr}[\sigma_{B^{(i)}}\mathcal{U}_t(\rho^{(i)})] \equiv f^{(i)}(t)$ and of the corresponding degree-$(L-1)$ Chebyshev interpolation, respectively. We can therefore use the work of Howell [48, Theorem 3], which tells us

$$|c_1^{(i)} - \alpha(A^{(i)})| \leq \left\|\omega^{(1)}\right\|_\infty \cdot \frac{\left\|\frac{\mathrm{d}^L}{\mathrm{d}t^L} f^{(i)}\right\|_\infty}{1! \cdot (L-1)!}, \tag{D.140}$$

where $\omega(t) := (t-t_0)\cdot(t-t_1)\cdot \ldots \cdot (t-t_N)$. To bound the right-hand side, we combine Lemma 6.2 with the observation that

$$\begin{aligned}
\left\|\omega^{(1)}\right\|_\infty &= \max_{\tau\in[0,T]} \frac{\mathrm{d}}{\mathrm{d}t}\left((t-t_1)\cdot(t-t_2)\cdot\ldots\cdot(t-t_L)\right)\Big|_{t=\tau} && \text{(D.141)}\\
&= \max_{\tau\in[0,T]} \sum_{i=1}^{L}\prod_{k\neq i}(t-t_k)\Big|_{t=\tau} && \text{(D.142)}\\
&\leq \max_{\zeta\in[-1,1]} \sum_{i=1}^{L}\prod_{k\neq i}\frac{T}{2}(z-z_k)\Big|_{z=\zeta} && \text{(D.143)}\\
&\leq \sum_{i=1}^{L} T^{L-1} && \text{(D.144)}\\
&\leq LT^{L-1}. && \text{(D.145)}
\end{aligned}$$

Combining these bounds, we obtain

$$|c_1^{(i)} - \alpha(A^{(i)})| \leq \frac{LT^{L-1}}{(L-1)!}\cdot(2\|H\|)^L = 2L\cdot\|H\|\cdot\frac{(2T\|H\|)^{L-1}}{(L-1)!}. \tag{D.146}$$

We can further upper bound this using the Stirling approximation bounds $\sqrt{2\pi(L-1)}(\frac{L-1}{e})^{L-1}e^{\frac{1}{12(L-1)+1}} \leq (L-1)! \leq \sqrt{2\pi(L-1)}(\frac{L-1}{e})^{L-1}e^{\frac{1}{12(L-1)}}$ from the work of Robbins [73], which give us

$$\begin{aligned}
2L\cdot\|H\|\cdot\frac{(2T\|H\|)^{L-1}}{(L-1)!} &\leq 2L\cdot\|H\|\cdot\frac{(2eT\|H\|)^{L-1}}{\sqrt{2\pi(L-1)}(L-1)^{L-1}}e^{-\frac{1}{12(L-1)+1}} && \text{(D.147)}\\
&\leq \frac{2L}{\sqrt{2\pi(L-1)}}\cdot\|H\|\cdot\left(\frac{1}{2}\right)^{L-1} && \text{(D.148)}\\
&\leq \frac{4\|H\|}{\sqrt{2\pi}\ln(2)}\cdot\ln(2)Le^{-\ln(2)L} && \text{(D.149)}
\end{aligned}$$

$$\leq \frac{4\|H\|}{\sqrt{2\pi}\ln(2)} \cdot \left(\frac{1}{2}\right)^{\frac{L}{2}} \tag{D.150}$$

$$\leq \frac{\varepsilon}{2}, \tag{D.151}$$

where the last inequality is by our choice of L. In that computation, we also used the fact that $xe^{-x} \leq e^{-\frac{x}{2}}$ holds for any $x \geq 0$. Altogether, we have shown that, conditioned on our success event, $|\hat{c}_1^{(i)} - \alpha(A^{(i)})| \leq \varepsilon$ holds for all $1 \leq i \leq M$.

To finish the proof, let us summarize the different complexity bounds. To build the classical representations in the first phase, we use $O(\frac{n+\log(1/\delta)}{\tilde{\varepsilon}^4})$ queries to $\mathcal{U}_{t_\ell}$ for each $1 \leq \ell \leq L$. Overall, the first phase thus uses $O(L \cdot \frac{n+\log(1/\delta)}{\tilde{\varepsilon}^4}) = \tilde{O}(\frac{n+\log(1/\delta)}{\varepsilon^4} \cdot \|H\|^4)$ parallel queries to time evolutions along H, each with a time of order $O(T) = O(1/\|H\|)$, leading to a total evolution time for the first phase of

$$T_{1,\text{total}} = O\left(\frac{n+\log(1/\delta)}{\tilde{\varepsilon}^4}\right) \cdot \sum_{\ell=1}^{L} t_\ell \tag{D.152}$$

$$= \tilde{O}\left(\frac{n+\log(1/\delta)}{\varepsilon^4} \cdot \|H\|^4\right) \cdot \sum_{\ell=1}^{L} \frac{T}{2}(1+z_\ell) \tag{D.153}$$

$$= \tilde{O}\left(\frac{n+\log(1/\delta)}{\varepsilon^4} \cdot \|H\|^4\right) \cdot \frac{1}{2\|H\|} \sum_{\ell=1}^{L} \left(1 - \cos\left(\frac{2\ell-1}{L}\pi\right)\right) \tag{D.154}$$

$$\leq \tilde{O}\left(\frac{n+\log(1/\delta)}{\varepsilon^4} \cdot \|H\|^4\right) \cdot \frac{L}{\|H\|} \tag{D.155}$$

$$\leq \tilde{O}\left(\frac{n+\log(1/\delta)}{\varepsilon^4} \cdot \|H\|^3\right). \tag{D.156}$$

The second phase produces the $\tilde{\varepsilon}$-accurate estimates $\hat{\alpha}_{t_\ell}^{(i)}$, for $1 \leq \ell \leq L$, $1 \leq i \leq M$. This uses $O(\frac{\log(M/\delta)}{\tilde{\varepsilon}^2})$ additional parallel queries to $\mathcal{U}_{t_\ell}$ for each $1 \leq \ell \leq L$. Accordingly, this uses $O(\frac{L\cdot\log(M/\delta)}{\tilde{\varepsilon}^2}) = \tilde{O}(\frac{\log(M)+\log(1/\delta)}{\varepsilon^2} \cdot \|H\|^2)$ additional parallel queries to $O(1/\|H\|)$-time evolutions along H, and the total evolution time for the second phase is $T_{2,\text{total}} \leq \tilde{O}(\frac{n+\log(1/\delta)}{\varepsilon^2} \cdot \|H\|)$. Producing the estimates $\hat{\alpha}_{t_\ell}^{(i)}$ uses classical computation time $O(LM \cdot \frac{n^2+n\log(1/\delta)}{\tilde{\varepsilon}^4}) = \tilde{O}(\frac{n^2+n\log(1/\delta)}{\varepsilon^4} \cdot \|H\|^4 M)$. This dominates the classical computation time since the Chebyshev interpolation and consequent derivative estimation takes time $O(L)$.

Finally, the classical representation $\hat{H}$ of H consists of the classical representations $\hat{\mathcal{U}}_{t_\ell}$ of $\mathcal{U}_{t_\ell}$ for $1 \leq \ell \leq L$. As each $\hat{\mathcal{U}}_{t_\ell}$ consists of $O(\frac{n^2+n\log(1/\delta)}{\tilde{\varepsilon}^4})$ real numbers, we see that $\hat{H}$ consists of $O(L \cdot \frac{n^2+n\log(1/\delta)}{\tilde{\varepsilon}^4}) = \tilde{O}(\frac{n^2+n\log(1/\delta)}{\varepsilon^4} \cdot \|H\|^4)$ real numbers stored in classical memory. □

E Query Complexity Lower Bound for PTM Learning With Quantum Memory from Choi Access

Complementing the discussion of learning (Pauli) transfer matrices and predicting sparse expectation values from copies of the Choi state from Section 4 and Appendix B, we give an information-theoretic argument to show that the linear-in-n sample complexity scaling of Theorem 4.1 and Corollary 4.2 is optimal for learning with quantum memory from Choi access.

Lemma E.1. *Any algorithm for learning with quantum memory from Choi access requires* $\Omega(\frac{n}{\varepsilon^2})$ *copies of the Choi state of an unknown n-qubit channel* $\mathcal{N}$ *to estimate all entries of* $R_{\mathcal{N}}^{\mathcal{P}}$ *up to accuracy* ε *with success probability* $\geq \frac{2}{3}$.

Proof. Consider the set of observables $\{P_i\}_{i=1}^{2\cdot 4^n(4^n-1)}$ defined in Lemma D.3 and the corresponding Choi states $\frac{1}{2^n}\Gamma_i = \frac{\mathbb{1}_2^{\otimes 2n}+3\varepsilon P_i}{2^{2n}}$. By construction, we have

$$\operatorname{tr}\left[P_i \frac{1}{2^n}\Gamma_j\right] = 3\varepsilon(\delta_{i,j} - \delta_{i+4^n(4^n-1),j}) \quad \forall 1 \leq i,j \leq 2\cdot 4^n(4^n-1), \tag{E.1}$$

where the addition in the index is to be understood modulo $2\cdot 4^n(4^n-1)$. Therefore, the ability to predict $\operatorname{tr}\left[P_i \frac{1}{2^n}\Gamma_j\right]$ up to accuracy ε for every j—which is no harder than estimating all PTM entries up to accuracy ε by Equation (4.1)—from T copies of the Choi state implies the ability to identify i from T copies of the Choi state. If we consider i to be chosen uniformly at random, then this implies that the mutual information between the classical and the quantum subsystem in the classical-quantum state

$$\rho_{\mathrm{cq}} = \mathbb{E}_{I\sim\mathcal{U}(\{1,\ldots,2\cdot 4^n(4^n-1)\})}\left[|I\rangle\langle I| \otimes \left(\frac{1}{2^n}\Gamma_I\right)^{\otimes T}\right] \tag{E.2}$$

is at least $\log_2(2\cdot 4^n(4^n-1)) = \Omega(n)$. In other words, we have

$$\begin{aligned}\Omega(n) &\leq I(\mathrm{c};\mathrm{q})_\rho &\text{(E.3)}\\ &= S\left(\mathbb{E}_{I\sim\mathcal{U}(\{1,\ldots,2\cdot 4^n(4^n-1)\})}\left[\left(\frac{1}{2^n}\Gamma_I\right)^{\otimes T}\right]\right) - \mathbb{E}_{I\sim\mathcal{U}(\{1,\ldots,2\cdot 4^n(4^n-1)\})}\left[S\left(\left(\frac{1}{2^n}\Gamma_I\right)^{\otimes T}\right)\right] &\text{(E.4)}\\ &\leq T\left(2n - \mathbb{E}_{I\sim\mathcal{U}(\{1,\ldots,2\cdot 4^n(4^n-1)\})}\left[S\left(\frac{1}{2^n}\Gamma_I\right)\right]\right). &\text{(E.5)}\end{aligned}$$

Here, the second step used that the quantum mutual information of a classical-quantum state equals its Holevo information (compare the work of Wilde [80, Exercise 11.6.9]) and the third step uses that any $(2nT)$-qubit state has a von Neumann entropy of at most $2nT$, and that the von Neumann entropy of a T-fold tensor power equals T times the von Neumann entropy of a single tensor factor.

Next, we compute $S\left(\frac{1}{2^n}\Gamma_i\right)$ for arbitrary i. To this end, simply observe that half of the eigenvalues of P_i are ± 1, so half of the eigenvalues of $\frac{1}{2^n}\Gamma_i$ are $\frac{1\pm 3\varepsilon}{2^{2n}}$. Therefore, we get

$$\begin{aligned}S\left(\frac{1}{2^n}\Gamma_i\right) &= \frac{2^{2n}}{2}\cdot\frac{1+3\varepsilon}{2^{2n}}\log\left(\frac{2^{2n}}{1+3\varepsilon}\right) + \frac{2^{2n}}{2}\cdot\frac{1-3\varepsilon}{2^{2n}}\log\left(\frac{2^{2n}}{1-3\varepsilon}\right) &\text{(E.6)}\\ &= 2n - \left(1 - \frac{1+3\varepsilon}{2}\log\left(\frac{2}{1+3\varepsilon}\right) - \frac{1-3\varepsilon}{2}\log\left(\frac{2}{1-3\varepsilon}\right)\right) &\text{(E.7)}\\ &= 2n - \mathcal{O}(\varepsilon^2), &\text{(E.8)}\end{aligned}$$

where the last step is via Taylor expansion. Therefore, we have shown

$$\Omega(n) \leq I(\mathrm{c};\mathrm{q})_\rho \leq \mathcal{O}(T\varepsilon^2), \tag{E.9}$$

which we can rearrange to the claimed $T = \Omega(\frac{n}{\varepsilon^2})$. □

The proof strategy used here is not new (see, e.g., the work of Aaronson [2, Section 6] for a similar reasoning). In fact, we can easily apply the same reasoning with the set of observables $\{P_i\}_{i=1}^{2\cdot 4^n(4^n-1)}$ from Lemma D.4 to show that the $\Omega(\frac{n}{\varepsilon^2})$ lower bound of Lemma E.1 even holds if the unknown channel is promised to be doubly-stochastic and to have a $\mathcal{O}(1)$-sparse PTM. Following the same reasoning that we have already used to go from Theorem B.1 to Corollary B.4, this lower bound immediately carries over to the task of predicting expectation values for states and observables with $\mathcal{O}(1)$-sparse Pauli ONB expansion for an unknown doubly-stochastic quantum channel with $\mathcal{O}(1)$-sparse PTM.

Acknowledgments

MCC thanks Hsin-Yuan Huang, Sumeet Khatri, John Preskill, and Asad Raza for insightful discussions, as well as Antonio Anna Mele, Andreas Bluhm, Alexander Nietner, Daniel Stilck França, Srinivasan Arunachalam, and the anonymous reviewers at FOCS 2023 for helpful feedback on an earlier draft of this article.

References

[1] Scott Aaronson. 2007. The learnability of quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 463, 2088 (2007), 3089–3114. https://doi.org/10.1098/rspa.2007.0113

[2] Scott Aaronson. 2018. Shadow tomography of quantum states. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC '18)*. ACM, New York, NY, USA, 325–338. https://doi.org/10.1145/3188745.3188802

[3] Scott Aaronson, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak. 2019. Online learning of quantum states. *Journal of Statistical Mechanics: Theory and Experiment* 2019, 12 (Dec. 2019), 124019. https://doi.org/10.1088/1742-5468/ab3988 arXiv:1802.09025

[4] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. 2021. The power of quantum neural networks. *Nature Computational Science* 1, 6 (2021), 403–409. https://doi.org/10.1038/s43588-021-00084-1

[5] Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi. 2022. Quantum algorithmic measurement. *Nature Communications* 13, 1 (2022), 1–9. https://doi.org/10.1038/s41467-021-27922-0

[6] Ahmed A. Akhtar, Hong-Ye Hu, and Yi-Zhuang You. 2023. Scalable and flexible classical shadow tomography with tensor networks. *Quantum* 7 (June 2023), 1026. https://doi.org/10.22331/q-2023-06-01-1026

[7] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. 2021. Sample-efficient learning of interacting quantum systems. *Nature Physics* 17, 8 (2021), 931–935. https://doi.org/10.1038/s41567-021-01232-0

[8] Mirko Arienzo, Markus Heinrich, Ingo Roth, and Martin Kliesch. 2023. Closed-form analytic expressions for shadow estimation with brickwork circuits. *Quantum Information and Computation* 23, 11-12 (Sept. 2023), 961–993. https://doi.org/10.26421/qic23.11-12-5

[9] Costin Bădescu and Ryan O'Donnell. 2021. Improved quantum data analysis. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 1398–1411. https://doi.org/10.1145/3406325.3451109

[10] Eyal Bairey, Itai Arad, and Netanel H. Lindner. 2019. Learning a local Hamiltonian from local measurements. *Physical Review Letters* 122, 2 (2019), 020504. https://doi.org/10.1103/physrevlett.122.020504

[11] Peter L. Bartlett and Shahar Mendelson. 2002. Rademacher and Gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research* 3 (Nov. 2002), 463–482. https://jmlr.org/papers/v3/bartlett02a.html

[12] Simon Becker, Nilanjana Datta, Ludovico Lami, and Cambyse Rouzé. 2024. Classical shadow tomography for continuous variables quantum systems. *IEEE Transactions on Information Theory* 70, 5 (2024), 3427–3452. https://doi.org/10.1109/TIT.2024.3357972

[13] Christian Bertoni, Jonas Haferkamp, Marcel Hinsche, Marios Ioannou, Jens Eisert, and Hakop Pashayan. 2022. Shallow shadows: Expectation estimation using low-depth random Clifford circuits. *arXiv preprint arXiv:2209.12924* (2022). https://arxiv.org/abs/2209.12924

[14] Kaifeng Bu, Dax Enshan Koh, Lu Li, Qingxian Luo, and Yaobo Zhang. 2021. Rademacher complexity of noisy quantum circuits. *arXiv preprint arXiv:2103.03139* (2021). https://arxiv.org/abs/2103.03139

[15] Kaifeng Bu, Dax Enshan Koh, Lu Li, Qingxian Luo, and Yaobo Zhang. 2022. Statistical complexity of quantum circuits. *Physical Review A* 105, 6 (2022), 062431. https://doi.org/10.1103/PhysRevA.105.062431

[16] Kaifeng Bu, Dax Enshan Koh, Lu Li, Qingxian Luo, and Yaobo Zhang. 2023. Effects of quantum resources and noise on the statistical complexity of quantum circuits. *Quantum Science and Technology* 8, 2 (Feb. 2023), 025013. https://doi.org/10.1088/2058-9565/acb56a

[17] Haoyuan Cai, Qi Ye, and Dong-Ling Deng. 2022. Sample complexity of learning parametric quantum circuits. *Quantum Science and Technology* 7, 2 (2022), 025014. https://doi.org/10.1088/2058-9565/ac4f30

[18] Matthias C. Caro. 2021. Binary classification with classical instances and quantum labels. *Quantum Machine Intelligence* 3 (2021), Article 18, 24 pages. https://doi.org/10.1007/s42484-021-00043-z

[19] Matthias C. Caro and Ishaun Datta. 2020. Pseudo-dimension of quantum circuits. *Quantum Machine Intelligence* 2 (2020), 14. https://doi.org/10.1007/s42484-020-00027-5

[20] Matthias C. Caro, Elies Gil-Fuster, Johannes Jakob Meyer, Jens Eisert, and Ryan Sweke. 2021. Encoding-dependent generalization bounds for parametrized quantum circuits. *Quantum* 5 (2021), 582. https://doi.org/10.22331/q-2021-11-17-582

[21] Matthias C. Caro, Hsin-Yuan Huang, Marco Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J. Coles. 2022. Generalization in quantum machine learning from few training data. *Nature Communications* 13 (2022), Article 4919. https://doi.org/10.1038/s41467-022-32550-3

[22] Matthias C. Caro, Hsin-Yuan Huang, Nicholas Ezzell, Joe Gibbs, Andrew T. Sornborger, Lukasz Cincio, Patrick J. Coles, and Zoë Holmes. 2023. Out-of-distribution generalization for learning quantum dynamics. *Nature Communications* 14 (2023), Article 3751. https://doi.org/10.1038/s41467-023-39381-w

[23] Chih-Chieh Chen, Masaya Watabe, Kodai Shiba, Masaru Sogabe, Katsuyoshi Sakamoto, and Tomah Sogabe. 2021. On the expressibility and overfitting of quantum circuit learning. *ACM Transactions on Quantum Computing* 2, 2 (2021), 1–24. https://doi.org/10.1145/3466797

[24] Sitan Chen. 2022. Learning to Predict Arbitrary Quantum Processes. Retrieved June 4, 2024 from https://memento.epfl.ch/event/learning-to-predict-arbitrary-quantum-processes/

[25] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. 2021. Exponential separations between learning with and without quantum memory. *arXiv preprint arXiv:2111.05881* (2021). https://arxiv.org/abs/2111.05881

[26] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. 2021. A hierarchy for replica quantum advantage. *arXiv preprint arXiv:2111.05874* (2021). https://arxiv.org/abs/2111.05874

[27] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. 2022. Exponential separations between learning with and without quantum memory. In *Proceedings of the 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS '22)*. IEEE, 574–585. https://doi.org/10.1109/FOCS52979.2021.00063

[28] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. 2023. The complexity of NISQ. *Nature Communications* 14, 1 (2023), 6001. https://doi.org/10.1038/s41467-023-41217-6

[29] Senrui Chen, Sisi Zhou, Alireza Seif, and Liang Jiang. 2022. Quantum advantages for Pauli channel estimation. *Physical Review A* 105, 3 (2022), 032435. https://doi.org/10.1103/PhysRevA.105.032435

[30] Xinyi Chen, Elad Hazan, Tongyang Li, Zhou Lu, Xinzhao Wang, and Rui Yang. 2022. Adaptive online learning of quantum states. *arXiv preprint arXiv:2206.00220* (2022). https://arxiv.org/abs/2206.00220

[31] Man-Duen Choi. 1975. Completely positive linear maps on complex matrices. *Linear Algebra and Its Applications* 10, 3 (1975), 285–290. https://doi.org/10.1016/0024-3795(75)90075-0

[32] Kai-Min Chung and Han-Hsuan Lin. 2021. Sample efficient algorithms for learning quantum channels in PAC model and the approximate state discrimination problem. In *16th Conference on the Theory of Quantum Computation, Communication and Cryptology (TQC)*, Min-Hsiu Hsieh (Ed.). Leibniz International Proceedings in Informatics (LIPIcs), Vol. 197. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, Article 3, 22 pages. https://doi.org/10.4230/LIPIcs.TQC.2021.3

[33] Yuxuan Du, Zhuozhuo Tu, Xiao Yuan, and Dacheng Tao. 2022. Efficient measure for the expressivity of variational quantum algorithms. *Physical Review Letters* 128, 8 (2022), 080506. https://doi.org/10.1103/PhysRevLett.128.080506

[34] Richard M. Dudley. 1999. *Uniform Central Limit Theorems*. Cambridge University Press.

[35] Marco Fanizza, Yihui Quek, and Matteo Rosati. 2022. Learning quantum processes without input control. *arXiv preprint arXiv:2211.05005* (2022). https://arxiv.org/abs/2211.05005

[36] Steven T. Flammia and Ryan O'Donnell. 2021. Pauli error estimation via Population Recovery. *Quantum* 5 (2021), 549. https://doi.org/10.22331/q-2021-09-23-549

[37] Steven T. Flammia and Joel J. Wallman. 2020. Efficient estimation of Pauli channels. *ACM Transactions on Quantum Computing* 1, 1 (2020), 1–32. https://doi.org/10.1145/3408039

[38] James R. Garrison and Tarun Grover. 2018. Does a single eigenstate encode the full Hamiltonian? *Physical Review X* 8, 2 (2018), 021026. https://doi.org/10.1103/PhysRevX.8.021026

[39] Daniel Greenbaum. 2015. Introduction to quantum gate set tomography. *arXiv preprint arXiv:1509.02921* (2015). https://arxiv.org/abs/1509.02921

[40] Andi Gu, Lukasz Cincio, and Patrick J. Coles. 2024. Practical Hamiltonian learning with unitary dynamics and Gibbs states. *Nature Communications* 15, 1 (Jan. 2024), Article 312. https://doi.org/10.1038/s41467-023-44008-1

[41] Casper Gyurik, Dyon Vreumingen, van, and Vedran Dunjko. 2023. Structural risk minimization for quantum linear classifiers. *Quantum* 7 (Jan. 2023), 893. https://doi.org/10.22331/q-2023-01-13-893

[42] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. 2016. Sample-optimal tomography of quantum states. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*. 913–925. https://doi.org/10.1109/TIT.2017.2719044

[43] Jeongwan Haah, Robin Kothari, and Ewin Tang. 2021. Optimal learning of quantum Hamiltonians from high-temperature Gibbs states. *arXiv preprint arXiv:2108.04842* (2021). https://arxiv.org/abs/2108.04842

[44] Robin Harper, Steven T. Flammia, and Joel J. Wallman. 2020. Efficient learning of quantum noise. *Nature Physics* 16, 12 (2020), 1184–1188. https://doi.org/10.1038/s41567-020-0992-

[45] Robin Harper, Wenjun Yu, and Steven T. Flammia. 2021. Fast estimation of sparse quantum noise. *PRX Quantum* 2, 1 (2021), 010322. https://doi.org/10.1103/PRXQuantum.2.010322

[46] Teiko Heinosaari and Mário Ziman. 2011. *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement*. Cambridge University Press.

[47] Michael Horodecki, Peter W. Shor, and Mary Beth Ruskai. 2003. Entanglement breaking channels. *Reviews in Mathematical Physics* 15, 6 (2003), 629–641. https://doi.org/10.1142/S0129055X03001709

[48] Gary W. Howell. 1991. Derivative error bounds for Lagrange interpolation: An extension of Cauchy's bound for the error of Lagrange interpolation. *Journal of Approximation Theory* 67, 2 (1991), 164–173. https://doi.org/10.1016/0021-9045(91)90015-3

[49] Hong-Ye Hu, Soonwon Choi, and Yi-Zhuang You. 2021. Classical shadow tomography with locally scrambled quantum dynamics. *arXiv preprint arXiv:2107.04817* (2021). https://arxiv.org/abs/2107.04817

[50] Hsin-Yuan Huang. 2022. Learning to Predict Arbitrary Quantum Processes. Retrieved June 4, 2024 from https://simons.berkeley.edu/events/quantum-colloquium-learning-predict-arbitrary-quantum-processes

[51] Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, and Jarrod R. McClean. 2022. Quantum advantage in learning from experiments. *Science* 376, 6598 (2022), 1182–1186. https://doi.org/10.1126/science.abn7293

[52] Hsin-Yuan Huang, Sitan Chen, and John Preskill. 2023. Learning to predict arbitrary quantum processes. *PRX Quantum* 4, 4 (2023), 040337. https://doi.org/10.1103/PRXQuantum.4.040337

[53] Hsin-Yuan Huang, Steven T. Flammia, and John Preskill. 2022. Foundations for learning from noisy quantum experiments. *arXiv preprint arXiv:2204.13691* (2022). https://arxiv.org/abs/2204.13691

[54] Hsin-Yuan Huang, Richard Kueng, and John Preskill. 2020. Predicting many properties of a quantum system from very few measurements. *Nature Physics* 16, 10 (2020), 1050–1057. https://doi.org/10.1038/s41567-020-0932-7

[55] Hsin-Yuan Huang, Richard Kueng, and John Preskill. 2021. Information-theoretic bounds on quantum advantage in machine learning. *Physical Review Letters* 126, 19 (May 2021), 190505. https://doi.org/10.1103/PhysRevLett.126.190505

[56] Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V. Albert, and John Preskill. 2022. Provably efficient machine learning for quantum many-body problems. *Science* 377, 6613 (2022), eabk3333. https://doi.org/10.1126/science.abk3333

[57] Hsin-Yuan Huang, Yu Tong, Di Fang, and Yuan Su. 2023. Learning many-body Hamiltonians with Heisenberg-limited scaling. *Physical Review Letters* 130, 20 (2023), 200403. https://doi.org/10.1103/PhysRevLett.130.200403

[58] Andrzej Jamiołkowski. 1972. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics* 3, 4 (1972), 275–278. https://doi.org/10.1016/0034-4877(72)90011-0

[59] Michael J. Kearns and Robert E. Schapire. 1994. Efficient distribution-free learning of probabilistic concepts. *Journal of Computer and System Sciences* 48, 3 (1994), 464–497. https://doi.org/10.1016/S0022-0000(05)80062-5

[60] Jonathan Kunjummen, Minh C. Tran, Daniel Carney, and Jacob M. Taylor. 2023. Shadow process tomography of quantum channels. *Physical Review A* 107, 4 (2023), 042403. https://doi.org/10.1103/PhysRevA.107.042403

[61] Ryan Levy, Di Luo, and Bryan K. Clark. 2024. Classical shadows for quantum process tomography on near-term quantum computers. *Physical Review Research* 6, 1 (Jan. 2024), 013029. https://doi.org/10.1103/physrevresearch.6.013029

[62] Zhi Li, Liujun Zou, and Timothy H. Hsieh. 2020. Hamiltonian tomography via quantum quench. *Physical Review Letters* 124, 16 (2020), 160502. https://doi.org/10.1103/PhysRevLett.124.160502

[63] Christian B. Mendl and Michael M. Wolf. 2009. Unital quantum channels—Convex structure and revivals of Birkhoff's theorem. *Communications in Mathematical Physics* 289, 3 (2009), 1057–1086. https://doi.org/10.1007/s00220-009-0824-2

[64] Michael A. Nielsen and Isaac L. Chuang. 2000. *Quantum Computation and Quantum Information*. Cambridge University Press.

[65] Ryan O'Donnell and John Wright. 2016. Efficient quantum tomography. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*. 899–912. https://doi.org/10.1145/2897518.2897544

[66] Ryan O'Donnell and John Wright. 2017. Efficient quantum tomography II. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. 962–974. https://doi.org/10.1145/3055399.3055454

[67] Marco Paini and Amir Kalev. 2019. An approximate description of quantum states. *arXiv preprint arXiv:1910.10543* (2019). https://arxiv.org/abs/1910.10543

[68] David Pollard. 1984. *Convergence of Stochastic Processes*. Springer.

[69] Claudiu Marius Popescu. 2021. Learning bounds for quantum circuits in the agnostic setting. *Quantum Information Processing* 20, 9 (2021), 1–24. https://doi.org/10.1007/s11128-021-03225-7

[70] John Preskill. 2018. Quantum computing in the NISQ era and beyond. *Quantum* 2 (2018), 79. https://doi.org/10.22331/q-2018-08-06-79

[71] John Preskill. 2020. *Quantum Computation (Lecture Notes)*. Caltech. http://theory.caltech.edu/~preskill/ph219/index.html#lecture

[72] Xiao-Liang Qi and Daniel Ranard. 2019. Determining a local Hamiltonian from a single eigenstate. *Quantum* 3 (2019), 159. https://doi.org/10.22331/q-2019-07-08-159

[73] Herbert Robbins. 1955. A remark on Stirling's formula. *American Mathematical Monthly* 62, 1 (1955), 26–29. https://doi.org/10.2307/2308012

[74] Cambyse Rouzé and Daniel Stilck França. 2024. Learning quantum many-body systems from a few copies. *Quantum* 8 (April 2024), 1319. https://doi.org/10.22331/q-2024-04-30-1319

[75] Daniel Stilck França, Liubov A. Markovich, V. V. Dobrovitski, Albert H. Werner, and Johannes Borregaard. 2024. Efficient and robust estimation of many-qubit Hamiltonians. *Nature Communications* 15, 1 (2024), 311. https://doi.org/10.1038/s41467-023-44012-5

[76] Leslie G. Valiant. 1984. A theory of the learnable. *Communications of the ACM* 27, 11 (1984), 1134–1142. https://doi.org/10.1145/1968.1972

[77] Vladimir N. Vapnik and Alexei Ya. Chervonenkis. 1971. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications* 16, 2 (1971), 264–280. https://doi.org/10.1137/1116025

[78] John Watrous. 2018. *The Theory of Quantum Information*. Cambridge University Press.

[79] Frederik Wilde, Augustine Kshetrimayum, Ingo Roth, Dominik Hangleiter, Ryan Sweke, and Jens Eisert. 2022. Scalably learning quantum many-body Hamiltonians from dynamical data. *arXiv preprint arXiv:2209.14328* (2022). https://arxiv.org/abs/2209.14328

[80] Mark M. Wilde. 2011. From classical to quantum Shannon theory. *arXiv preprint arXiv:1106.1445* (2011). https://arxiv.org/abs/1106.1445

[81] Michael M. Wolf. 2012. Quantum Channels and Operations: Guided Tour. Retrieved June 4, 2024 from https://mediatum.ub.tum.de/doc/1701036/1701036.pdf

[82] Wenjun Yu, Jinzhao Sun, Zeyao Han, and Xiao Yuan. 2023. Robust and efficient hamiltonian learning. *Quantum* 7 (June 2023), 1045. https://doi.org/10.22331/q-2023-06-29-1045

Received 11 October 2023; revised 11 October 2023; accepted 24 May 2024