Learning shallow quantum circuits with many-qubit gates

Francisca Vasconcelos   FRANCISCA@BERKELEY.EDU
UC Berkeley

Hsin-Yuan Huang   HSINYUAN@CALTECH.EDU
Caltech, Google AI

Editors: Nika Haghtalab and Ankur Moitra

Abstract

The seminal work of [LMN'93] established a cornerstone result for classical complexity, with profound implications for learning theory. By proving low-degree Fourier concentration of AC^0, the work demonstrated that Boolean functions computed by constant-depth circuits can be efficiently PAC-learned via low-degree Fourier sampling. This breakthrough provided the first sample- and time-efficient (quasi-polynomial) algorithm for learning AC^0.

Proposed by [Moore'99] as a natural quantum analog of AC^0, QAC^0 is the class of constant-depth quantum circuits composed of arbitrary single-qubit gates and polynomial CZ gates of unbounded width. In this work, we present the first algorithm for efficient average-case learning of QAC^0 circuits with logarithmic ancilla. Namely, our algorithm achieves quasi-polynomial sample- and time-complexity for learning unknown QAC^0 unitaries to inverse-polynomially small error. We further show that these learned unitaries can be efficiently synthesized via poly-logarithmic depth circuits, making progress towards proper learning of QAC^0. Since in finite-dimensional circuit geometries QAC^0 circuits require polynomial depth to implement, this result significantly expands the family of efficiently learnable quantum circuits.

Keywords: Quantum Computing, Quantum Learning, Circuit Learning, QAC^0

1. Introduction

Motivated by both theory and practice, shallow quantum circuits have become central to quantum complexity. Theoretically, shallow quantum circuits have proven more powerful than their classical counterparts (Bravyi et al., 2018; Watts et al., 2019; Bravyi et al., 2020; Watts and Parham, 2023), with output distributions that are expected to be classically hard to simulate (Terhal and DiVincenzo, 2002; Gao et al., 2017; Bermejo-Vega et al., 2018; Haferkamp et al., 2020; Hangleiter and Eisert, 2023). Experimentally, current quantum hardware is noisy and only capable of simulating short time-scale quantum processes. In order to characterize and continue improving quantum hardware, it is crucial that we develop efficient learning procedures for these shallow quantum circuits.

Classically, efficient learning of shallow circuits was an immediate consequence of a seminal result in classical circuit complexity. Namely, Linial et al. (1993) established low-degree Fourier concentration of Boolean functions implementable in AC^0—the family of constant-depth, polynomial-sized circuits with unbounded fan-in and fan-out. This result had profound implications for areas such as learning theory, hardness of approximation, communication complexity, pseudorandomness, and circuit lower-bounds. For learning, specifically, Linial et al. (1993) connected Fourier concentration to the learnability of circuit classes, resulting in the first quasi-polynomial sample- and time-complexity algorithm for PAC-learning any Boolean function implementable in AC^0.

A natural quantum analog of AC^0 is the QAC^0 circuit class, proposed by Moore (1999). Specifically, QAC^0 is the class of constant-depth (shallow) quantum circuits comprised of arbitrary single-qubit gates and polynomial CZ_k gates of unbounded width k[1] (many-qubit gates). It is a long-standing open question as to whether the quantum fan-out operation can be implemented in QAC^0, i.e. whether $\mathsf{QAC}^0 = \mathsf{QAC}^0_f$ (where the f denotes "with fan-out")[2]. If fan-out is in QAC^0, then powerful subroutines such as sorting, arithmetic operations, phase estimation, and the quantum Fourier transform could be approximately implemented in QAC^0 (Høyer and Špalek, 2005). Furthermore, QAC^0 contains QNC^0 (the quantum analog of NC^0), which is defined similarly to QAC^0, but constrained to constant-width CZ_k gates, i.e. $k = O(1)$. Since constant-width gates are what we can feasibly implement on real quantum hardware and $\mathsf{QNC}^0 \subset \mathsf{QAC}^0 \subset \mathsf{QNC}^1$, learning QAC^0 makes progress towards learning practical logarithmic-depth circuits.

Recent work has made exciting progress towards efficient learning of QAC^0. Notably, Huang et al. (2024) established an efficient worst-case algorithm for learning QNC^0 unitaries. However, this algorithm relies crucially on the fact that QNC^0 gates act only on a constant number of qubits, making it unclear how to generalize the approach to QAC^0 (with many-qubit gates). In another vein, Nadimpalli et al. (2024) proved a quantum analog of Linial et al. (1993)'s celebrated Fourier-concentration result. In particular, they demonstrated that the Pauli spectrum[3] of single-output QAC^0 circuit channels (with limited ancillas) are low-degree concentrated. Interestingly, however, unlike the classical Fourier-concentration result of Linial et al. (1993), this does not straightforwardly imply an efficient learning algorithm for full characterization of QAC^0 circuits.

To see why, note that in the classical setting of Linial et al. (1993), although learned Fourier coefficients are not exact, there exists a simple projection from the learned monomial onto the space of valid Boolean functions. This adds negligible time-complexity to the overall learning algorithm. However, in learning single-output channels of n-qubit QAC^0 circuits (Nadimpalli et al., 2024), the central object of study is an $O(4^n)$-dimensional matrix (corresponding to the channel's Choi representation). Projecting the algorithm's learned matrix onto the space of valid Choi representations requires solving a semi-definite programming problem, which can require exponential time in the worst-case. Thus, the overall runtime of Nadimpalli et al. (2024)'s learning algorithm is far from efficient. Furthermore, even if the algorithm were time-efficient, it would only describe one of the QAC^0 circuit's n outputs and is limited to QAC^0 circuits with at most logarithmic ancilla.

In this paper, we address most limitations of the prior work, by proposing the first sample- and *time-efficient* algorithm for learning the *full unitary* implemented by a QAC^0 circuit. In particular, we show that n-qubit QAC^0 unitaries are learnable to an average-case distance, with high probability, in quasi-polynomial sample- and time-complexity. This matches both the sample- and time-complexity of Linial et al. (1993). Furthermore, we offer an efficient circuit synthesis procedure for our learned unitaries, making progress towards proper learning of QAC^0 circuits. The key remaining limitation, as in the prior work of Nadimpalli et al. (2024), is that our learning algorithm's provable guarantees only extend to QAC^0 circuits with up to logarithmic ancilla. However, we offer a conjecture about the Pauli spectrum of QAC^0 circuits, which would extend our guaran-

---

1. The CZ_k gate applies a -1 phase to a state if all k inputs are in the $|1\rangle$ state and otherwise leaves the state unchanged. This is a natural quantum analog of the classical AND operation.

2. Quantum fan-out is a many-qubit gate, with one control qubit that applies bit-flips to many target qubits. Since quantum fan-out is equivalent to parity via Hadamard conjugation, if fan-out $\in \mathsf{QAC}^0$, this would imply parity $\in \mathsf{QAC}^0$. This is in contrast to AC^0, which trivially contains fan-out, but does not contain parity (Linial et al., 1993).

3. The Pauli spectrum is the quantum analog of the Fourier spectrum of classical Boolean functions. The notion of quantum Boolean functions was originally proposed by Montanaro and Osborne (2010).

Circuit Class	Type	Fan-in	Paper	Learned Object	# Outputs	Complexity		# Ancilla
						Time	Sample	
AC^0	Classical	∞	Linial et al. (1993)	Monomial	1	$O(n^{\log^d n})$	$O(n^{\log^d n})$	–
QNC^0	Quantum	$O(1)$	Huang et al. (2024)	Unitary	n	$O(n)$	$O(n)$	∞
QAC^0	Quantum	∞	Nadimpalli et al. (2024)	Choi Rep	1	$O(2^n)$	$O(n^{\log^d n})$	$O(\log n)$
			This work	Unitary	n	$O(n^{\log^d n})$		

Table 1: Comparison of learning algorithms for shallow (depth-d) classical and quantum circuits. For the quantum algorithms we also report the number of ancilla the guarantees hold for.

tees to QAC^0 circuits with polynomially many ancilla. Table 1 offers a summary of our algorithm's learning guarantees and comparison to prior work.

2. Overview of Results and Contributions

The primary contribution of this work is the first sample- and time-efficient algorithm for learning shallow quantum circuits with many-qubit gates. Namely, we prove the following main result:

Theorem 1 (Efficient learning of n-output QAC^0 unitaries) *Consider an unknown n-qubit, depth-d QAC^0 circuit, implementing unitary C, and failure probability $\delta \in (0,1)$. With high probability, $1 - \delta$, we can learn a $2n$-qubit unitary U, such that the average gate fidelity[4] $\mathcal{D}_{avg}$ is bounded as*

$$\mathcal{D}_{avg}(U, C \otimes C^{\dagger}) \leq 1/\text{poly}(n). \tag{1}$$

Moreover, U is learnable with quasi-polynomial, $O(n^{\log^d n} \log(1/\delta))$, sample- and time-complexity.

Note that our procedure makes use of classical shadow tomography (Huang et al., 2020) and, thus, falls within the standard quantum-PAC learning model. However, we use shadow tomography to learn the Heisenberg-evolved single-qubit Pauli observables of U, meaning we require query access to both U and $U^{\dagger}$, with stabilizer states as input and Pauli-basis output measurements.

We also prove the following unitary synthesis result, which makes progress towards a quasi-polynomial time proper learning algorithm for QAC^0 circuits. Namely, we show that the learned unitary can be synthesized in QAC, or the class of all polylogarithmic-depth quantum circuits consisting of arbitrary single-qubit gates and many-qubit CZ gates.

Proposition 2 (Learning QAC implementations of QAC^0) *Given the learned unitary U which is $1/\text{poly}(n)$-close to QAC^0 circuit C, there exists a quasi-polynomial time algorithm to learn a QAC circuit implementing unitary U^* such that $\mathcal{D}_{avg}(U^*, C \otimes C^{\dagger}) \leq 1/\text{poly}(n)$.*

Our work builds upon and extends the results of Nadimpalli et al. (2024) and Huang et al. (2024). While Nadimpalli et al. (2024) studied the Choi representations of single-output QAC^0 channels, Huang et al. (2024) focused on learning and "sewing" Heisenberg-evolved single-qubit Pauli observables of QNC^0 circuits. To combine these techniques, we establish a connection between the Choi representations of single-output QAC^0 channels and Heisenberg-evolved single-qubit Pauli observables of QAC^0 circuits. We re-prove several key results from Nadimpalli et al. (2024) for QAC^0 Heisenberg-evolved single-qubit Pauli observables, establishing their low-degree concentration. Furthermore, we make the key observation that QAC^0 observables are not just low-degree

---

4. As discussed in Appendix A.2, the average gate fidelity is a measure of how distinct two unitary channel outputs are, for pure input states averaged over the Haar (uniform) measure.

concentrated, but are in fact low-support concentrated—i.e. they are (poly $\log n$)-juntas, with non-zero Pauli terms supported on only poly $\log n$ of the n qubits. This strengthened concentration result is critical in achieving the accuracy guarantees and efficient computational complexity of our learning algorithm. Overall, we believe that these concentration results for Heisenberg-evolved Pauli observables are of independent interest beyond the learning algorithm studied in this work.

Our algorithm improves upon both Huang et al. (2024) and Nadimpalli et al. (2024) in several key ways. Foremost, we successfully generalize the learning procedure of Huang et al. (2024) to circuits with *many-qubit* gates, a setting to which their proof techniques do not directly apply. This generalization is made possible by the new observable concentration results, which show that learning observables with *restricted support* suffices to approximate QAC^0 observables. The restricted support of these learned observables is crucial in obtaining the learning algorithm's efficient sample and time complexity. In contrast to Nadimpalli et al. (2024), which focuses on learning the low-degree approximation of the *Choi representation* of a *single-output* QAC^0 channel (necessitating the tracing out of $n-1$ qubits), our algorithm accomplishes efficient learning of the *unitary* of the *entire n-output* QAC^0 circuit. Most importantly, our algorithm achieves quasi-polynomial sample and time complexity, whereas Nadimpalli et al. (2024) attains quasi-polynomial sample complexity, but suffers an exponential time complexity.

While our algorithm represents a significant advancement in learning QAC^0 circuits, it has two key limitations compared to previous work. A notable drawback of our learning algorithm compared to that of Huang et al. (2024) is that ours only offers an average-case learning guarantee, not a worst-case guarantee[5]. However, we observe that the (Huang et al., 2024, Proposition 3) lower-bound for hardness of learning logarithmic-depth quantum circuits to diamond-norm distance also applies to QAC^0 circuits. Thus, we justify our use of an average-case measure by showing that no algorithm can efficiently worst-case learn arbitrary QAC^0 circuits:

Proposition 3 (Hardness of learning QAC^0) *Learning an unknown n-qubit unitary U, generated by a QAC^0 circuit, to $\frac{1}{3}$ dimond distance with high probability requires $\exp(\Omega(n))$ queries.*

Furthermore, it was not immediately obvious that the worst-case learning guarantees of Huang et al. (2024) would translate to the average-case setting for QAC^0. The proof of these average-case guarantees was a substantial and notable contribution of our work, elaborated in Appendix D.

Another limitation of our learning algorithm, compared to Huang et al. (2024), is that it only works for QAC^0 circuits with a logarithmic ancilla qubits. Note, however, that this ancilla constraint matches that of the learning algorithm in (Nadimpalli et al., 2024, Theorem 39). This constraint arises from the nature of many-qubit QAC^0 gates, which allow light-cones to rapidly encompass a large number of ancilla qubits, even in constant depth. In contrast, the constant-width gate setting of Huang et al. (2024) ensures that at most a constant number of ancillas are in any output qubit's light-cone, implying that most ancilla qubits do not affect the computation. Our algorithm's reliance on light-cone style arguments about the support of Heisenberg-evolved observables necessitates restricting the number of ancilla qubits to prevent error blow-up. Moreover, as noted in Nadimpalli et al. (2024), the pre-specified input value of ancilla qubits poses a challenge for average-case arguments and removal of large-CZ_k gates. To make progress towards the many-ancilla case, we show that if the following conjecture, which is a strengthening of (Nadimpalli et al., 2024, Conjecture 1), is proven true, then our learning procedure for pollynomially-many ancilla qubits:

---

5. Huang et al. (2024)'s worst-case guarantee implies accurate outputs for *any* input. In contrast, average-case learning guarantees accurate outputs for *most* inputs.

Conjecture 4 (Ancilla-independent low-support concentration) *Suppose C is a depth-d, size-s QAC^0 circuit performing clean computation $C(I \otimes |0^a\rangle) = A \otimes |0^a\rangle$ on $n + a$ qubits, for $\mathrm{poly}(n)$ ancilla a. Let $O_{P_i,n} = A(P_i \otimes I^a)A^\dagger$ be a Heisenberg-evolved single-qubit Pauli observable with ancilla restriction. For support $\mathcal{S}$ such that $|\mathcal{S}| = k^d$, the weight outside $\mathcal{S}$ is bounded as*

$$\boldsymbol{W}^{\notin \mathcal{S}}[O_{P_i,n}] \leq \mathrm{poly}(s) \cdot 2^{-\Omega(k^{1/d})}. \tag{2}$$

Corollary 5 (Learning QAC^0 with polynomial ancilla) *Assume Conjecture 4 holds. With quasi-polynomial samples and time, we can learn $2n$-qubit unitary A_{sew}, with high probability, such that*

$$\mathcal{D}_{avg}(A_{sew}, A \otimes A^\dagger) \leq 1/\mathrm{poly}(n). \tag{3}$$

This corollary suggests a potential pathway for handling a broader class of QAC^0 circuits.

3. Prior Work

Sample-efficient learning algorithms have been developed for a wide range of quantum processes. For example, studies of generalization in quantum machine learning models have yielded sample-efficient algorithms for average-case learning of polynomial-size quantum circuit unitaries (Caro et al., 2022, 2023; Zhao et al., 2023). Additionally, Huang et al. (2022) established sample-efficient average-case learning algorithms for polynomial-complexity channels. These results leverage techniques such as shadow tomography (Aaronson, 2018; Bădescu and O'Donnell, 2021; Huang et al., 2020) to achieve sample efficiency. However, while the algorithms are sample-efficient, they often require predicting exponentially many observables and, thus, are not computationally efficient for the full learning task. To understand whether this is a limitation of all algorithms, Zhao et al. (2023) established quantum computational hardness for average-case learning of polynomial-sized quantum circuits, assuming quantum hardness of learning with errors (Regev, 2009).

Computational efficiency necessitates focusing on specific prediction tasks or more structured families of unitaries and channels. For instance, Huang et al. (2023) and Chen et al. (2024b) provided a quasi-polynomial time average-case learning algorithm for accurate prediction of arbitrary n-qubit channels. These algorithms are efficient when predicting a given observable on the output of the unknown quantum channel, for most input states. Restricting to certain families of unitaries and channels is another common strategy to attain computational efficiency. Computationally-efficient learning algorithms have been developed for various families such as unitaries generated by short-time Hamiltonian dynamics (Yu et al., 2023; Huang et al., 2023; Bakshi et al., 2024; Haah et al., 2024; Stilck França et al., 2024), Pauli channels under local or sparsity constraints (Flammia and Wallman, 2020; Flammia and O'Donnell, 2021; Chen et al., 2022, 2023a; Caro, 2024; Chen et al., 2024a), Clifford circuits with limited non-Clifford gates (Lai and Cheng, 2022; Grewal et al., 2023; Leone et al., 2024; Grewal et al., 2024; Du et al., 2024), and quantum juntas (Chen et al., 2023b; Bao and Yao, 2023). Especially relevant to our work are (Nadimpalli et al., 2024)'s algorithm for learning single-output QAC^0 channels and Huang et al. (2024)'s algorithm for learning shallow quantum circuits comprised of constant-width gates (QNC^0), which we now describe in more detail.

Learning shallow quantum circuits. Huang et al. (2024) gave the first polynomial sample- and time-complexity algorithm for learning unitaries of shallow quantum circuits with constant-width gates (QNC^0). The key insight was that, although these circuits can generate highly non-local and

classically-hard output distributions, the associated unitaries can be efficiently reconstructed (to high-accuracy) via the local light-cones of each output qubit. Namely, for an n-qubit shallow circuit U, they developed an efficient algorithm for learning the $3n$ Heisenberg-evolved single-qubit Pauli observables $O_{P_i} = U P_i U^\dagger$. They proved that the learned $\widetilde{O}_{P_i}$ observables are close to the true O_{P_i} observables, with respect to the operator norm, and can be efficiently "sewed" into unitary

$$U_{\text{sew}} = \text{SWAP}^{\otimes n} \prod_{i=1}^{n} \left[\text{Proj}_\infty \left(\frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} \widetilde{O}_{P_i} \otimes P_i \right) \right], \tag{4}$$

where Proj_∞ is the projection onto the unitary minimizing the operator norm distance with respect to the input matrix and $\text{SWAP}^{\otimes n}$ is the SWAP operation between the top and bottom n qubits. Finally, they proved that sewn unitary U_{sew} is close to the true unitary $U \otimes U^\dagger$ under diamond distance. Note that the algorithm's sample and time efficiency rely heavily on the constant-sized support of observables O_{P_i} and $\widetilde{O}_{P_i}$. However, QAC^0 observables can have polynomial-sized support.

On the Pauli Spectrum of QAC^0. Nadimpalli et al. (2024) achieved a quantum analog of Linial et al. (1993)'s Fourier concentration bound for AC^0. For an n-qubit QAC^0 circuit implementing unitary C, they defined single-qubit output channel $\mathcal{E}_C(\rho) = \text{Tr}_{n-1} \left(C \rho C^\dagger \right)$, where ρ is an n-qubit density matrix and Tr_{n-1} denotes the partial trace over $n - 1$ of the qubits, leaving a singular designated output qubit. Via the standard Choi-Jamiołkowski isomorphism, they mapped this channel into a Choi representation, decomposable into the Pauli basis as

$$\Phi_{\mathcal{E}_C} = \left(I^{\otimes n} \otimes \mathcal{E}_C \right) |\text{EPR}_n\rangle\langle\text{EPR}_n| = \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} \widehat{\Phi}_{\mathcal{E}_C}(P) \cdot P, \tag{5}$$

where $|\text{EPR}_n\rangle = \sum_{x \in [n]} |x\rangle \otimes |x\rangle$ denotes the un-normalized Bell state and $\widehat{\Phi}_{\mathcal{E}_C}(P)$ are the Pauli (Fourier) coefficients[6]. They proved that $\Phi_{\mathcal{E}_C}$ is low-degree concentrated by showing that the Pauli weight for all Paulis P of degree $> k$, denoted $\mathbf{W}^{>k}[\Phi_{\mathcal{E}_C}]$, decays exponentially in k. Formally, if size s denotes the total number of CZ_k gates in the circuit, $|P|$ denotes the degree[7], and $d = \mathcal{O}(1)$ denotes the circuit depth,

$$\mathbf{W}^{>k}[\Phi_{\mathcal{E}_C}] = \sum_{|P|>k} \left| \widehat{\Phi}_{\mathcal{E}_C}(P) \right|^2 \leq \mathcal{O}\left(\frac{s^2}{2^{k^{1/d}}} \right). \tag{6}$$

Since $\Phi_{\mathcal{E}_C}$ is low-degree concentrated, Nadimpalli et al. (2024) learned $\Phi_{\mathcal{E}_C}$'s low-degree Pauli coefficients, i.e. $\widetilde{\Phi}_{\mathcal{E}_C}(P)$ for all $P \in \mathcal{F} = \{Q \in \mathcal{P}^n : |Q| < \text{poly} \log(n)\}$, to construct approximate Choi representation $\widetilde{\Phi}_{\mathcal{E}_C} = \sum_{P \in \mathcal{F}} \widetilde{\Phi}_{\mathcal{E}_C}(P) \cdot P$ and proved that learned matrix $\widetilde{\Phi}_{\mathcal{E}_C}$ is close to the true Choi representation $\Phi_{\mathcal{E}_C}$ (according to the normalized Frobenius distance). This resulted in the first quasi-polynomial sample-complexity algorithm for learning single-output QAC^0 channels. However, since projecting the learned matrix onto a valid Choi representation requires solving a exponential-dimensional semi-definite program, the runtime is exponential.

---

6. For more background on Pauli analysis, we refer the reader to Appendix A.

7. The degree of a Pauli string is the number of qubits upon which the Pauli operator P acts non-trivially.

4. Proof Overview

At a high level, our algorithm for learning n-qubit QAC^0 circuits is a nontrivial synthesis of Nadimpalli et al. (2024)'s QAC^0 concentration results and Huang et al. (2024)'s sewing technique. Our main technical contribution is bridging the high-level concepts introduced in the two works and developing an efficient end-to-end learning procedure with provable guarantees. We will now provide a high-level overview of and intuition for the key results and techniques introduced in this work.

Choi Representations and Heisenberg-Evolved Observables.   While Nadimpalli et al. (2024) proved low-degree concentration of the *Choi representation* of single-output channels, the Huang et al. (2024) learned low-degree approximations of single-qubit Heisenberg-evolved Pauli *observables*. In Appendix B, we build a translator between these two quantum primitives for QAC^0. To summarize, while the Choi representation encodes information about the QAC^0 channel with respect to any input and measurement observable, the Heisenberg-evolved observable restricts the channel with respect to a single measurement observable. Mathematically, for a QAC^0 circuit C, with single-output channel $\mathcal{E}_C$, Choi representation $\Phi_{\mathcal{E}_C}$, and single-qubit measurement observable O, the two representations are related via the dual-channel as

$$\mathcal{E}_C^*(O) = C^\dagger(I_{n-1} \otimes O)C = \mathrm{Tr}_{\mathrm{out}}\left((O_{\mathrm{out}} \otimes I_{\mathrm{in}})\Phi_{\mathcal{E}_C}\right)^\top. \tag{7}$$

Concentration of QAC^0 Heisenberg-Evolved Observables.   Either the Choi representation or Heisenberg-evolved observable perspective could have been suitable for our work, but we chose observables due to their simpler mathematical nature and because learning requires circuit measurements with respect to select observables. This choice enables straightforward use of Huang et al. (2024)'s Heisenberg-evolved observable sewing procedure, but means some translational work is necessary to leverage Nadimpalli et al. (2024)'s Choi representation concentration result. Thus, the first contribution of our work is a proof similar to (Nadimpalli et al., 2024, Theorem 21), establishing low-degree concentration of QAC^0 Heisenberg-evolved single-qubit Pauli observables.

Proposition 6 (Low-degree concentration)   *For depth-d, size-s QAC^0 circuit C acting on n qubits and Heisenberg-evolved single-qubit Pauli observable $O_{P_i} = C^\dagger P_i C$, for every degree $k \in [n]$:*

$$\boldsymbol{W}^{>k}[O_{P_i}] = \sum_{|Q|>k} |\widehat{O}_{P_i}(Q)|^2 \leq \mathcal{O}\left(s^2 2^{-k^{1/d}}\right). \tag{8}$$

Conceptually, the proof of this result consists of two key steps. First, we show that if the QAC^0 circuit has no CZ gates of width greater than $k^{1/d}$, then the Heisenberg-evolved observable's weight for degree $> k$ is zero. Then, we show that removing these "large" CZ gates does not significantly change the Heisenberg-evolved observable under the normalized Frobenius distance measure, i.e.

Lemma 7 (Large CZ_k removal error)   *For QAC^0 circuit C and QAC^0 circuit $\widetilde{C}$, which is simply C with all m CZ_k's of size $k > \kappa$ removed, let $O_{P_i} = CP_iC^\dagger$ and $O_{P_i}^* = \widetilde{C}P_i\widetilde{C}^\dagger$ be the Heisenberg-evolved single-qubit Pauli observables. The average-case distance between these observables is*

$$\frac{1}{2^n}\left\|O_{P_i} - O_{P_i}^*\right\|_F^2 \leq \epsilon^* = \frac{9m^2}{2^\kappa}. \tag{9}$$

Note that this distance bound is critical in proving the algorithm's learning guarantees. Furthermore, as will be discussed shortly, the size κ (of the smallest CZ_k removed) must be carefully selected

Algorithm 1: Heisenberg-Evolved Observable Learning Procedure

1. Learn all the Pauli coefficients of O_{P_i} of degree $\leq \ell$ to precision η, as in Equation (14).

2. Find the subset of Paulis supported on ℓ qubits, $\mathcal{F}_{\{s\}} \in \mathfrak{F}_\ell$, with max weight amongst the learned coefficients,

$$T_\ell = \arg\max_{\mathcal{F}_{\{s\}} \in \mathfrak{F}_\ell} \sum_{Q \in \mathcal{F}_{\{s\}}} \left| \widetilde{O}_{P_i}(Q) \right|^2. \tag{11}$$

3. Set all coefficients outside of this maximal-weight support to zero,

$$\widetilde{O}_{P_i}^{(\ell)}(Q) = \begin{cases} \widetilde{O}_{P_i}(Q), & \text{if } Q \in T_\ell, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

4. Output the learned observable, which is fully supported on only ℓ qubits,

$$\widetilde{O}_{P_i}^{(\ell)} = \sum_{Q \in T_\ell} \widetilde{O}_{P_i}^{(\ell)}(Q) \cdot Q. \tag{13}$$

in order to learn the observable to high-precision, yet efficiently. Beyond showing that O_{P_i} is low-degree concentrated, we also show that it is low-support, by adapting the proof of Theorem 6.

Lemma 8 (Low-support concentration) *For ϵ^* the same as in Theorem 7 and $\mathcal{S}^* = \text{supp}(O_{P_i}^*)$, the weight of O_{P_i} outside the support of $O_{P_i}^*$ is upper-bounded as*

$$\boldsymbol{W}^{\notin \mathcal{S}^*}[O_{P_i}] = \sum_{Q \in \mathcal{S}^*} |\widehat{O}_{P_i}(Q)|^2 \leq \frac{1}{2^n} \left\| O_{P_i} - O_{P_i}^* \right\|_F^2 = \epsilon^*. \tag{10}$$

For more details about these concentration results and proofs, we refer the reader to Appendix C.

Efficient Learning of QAC^0 Heisenberg-Evolved Observables. Although efficient, Huang et al. (2024)'s algorithm for learning QNC^0 Heisenberg-evolved observables relies crucially on the fact that these observables consist only of constant-width gates and, thus, have constant support. In the QAC^0 setting, however, we only have query access to O_{P_i}, which can have $\mathcal{O}(n)$ support. Thus, we cannot directly use Huang et al. (2024)'s algorithm to learn QAC^0 Heisenberg-evolved observables.

Meanwhile, Nadimpalli et al. (2024) proposed an algorithm with efficient sample complexity for learning approximate Choi representations of single-output QAC^0 channels. The algorithm leverages low-degree concentration of the QAC^0 Choi representation to efficiently approximate it, by only learning low-degree Pauli coefficients. Given our low-degree concentration result for QAC^0 Heisenberg-evolved observables (Theorem 20), one might naturally assume we could simply learn the approximate low-degree observable $\widetilde{O}_{P_i} = \sum_{|Q| \leq \ell} \widetilde{O}_{P_i}(Q) \cdot Q$, where $\widetilde{O}_{P_i}(Q)$ are the learned Pauli coefficients. However, as we will elaborate shortly, it is *crucial* that the learned observables have *low-support*, so as to ensure that the unitary projection step in the eventual observable sewing procedure can be computed efficiently. Therefore, we need a more sophisticated learning procedure that learns a low-support observable.

In light of these difficulties, we propose a new learning algorithm (Algorithm 1) that efficiently learns approximate QAC^0 Heisenberg-evolved Pauli observables $\widetilde{O}_{P_i}^{(\ell)}$ with guaranteed low-support ℓ. Similar to the algorithm of Nadimpalli et al. (2024), this algorithm uses classical shadow
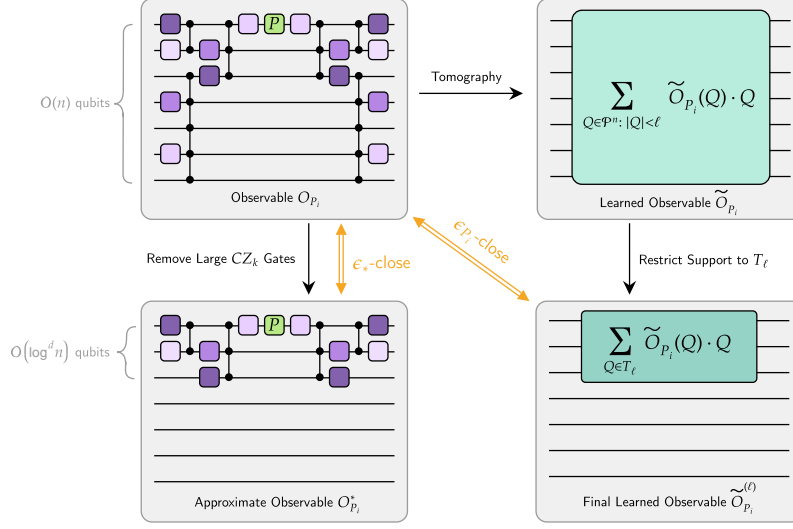
Figure 1: A conceptual illustration of all the observables involved in the QAC^0 Heisenberg-Evolved Learning procedure and their relations.

tomography Huang et al. (2020) to approximately learn all Pauli coefficients $\widetilde{O}_{P_i}(Q)$ for all Paulis $|Q| \leq \ell$. However, after all these coefficients are learned, the Pauli coeffcients are grouped into sets of ℓ-qubit support, $\mathcal{F}_{\{s\}} \in \mathfrak{F}_\ell$. All coefficients which lie outside the set T_ℓ of maximal weight, as expressed in Equation (11), are set to zero. This enforces that the final learned observable $\widetilde{O}_{P_i}^{(\ell)}$, with decomposition given in Equation (13), is supported on only ℓ qubits.

Leveraging our low-support concentration result (Theorem 8), we bound the Frobenius distance between $\widetilde{O}_{P_i}^{(\ell)}$ and O_{P_i} according to the learning accuracy and distance between O_{P_i} and $O_{P_i}^*$.

Lemma 9 (Algorithm 1 error bound) *Let O_{P_i} be a QAC^0 Heisenberg-evolved Pauli observable, which is ϵ^*-close to the observable $O_{P_i}^*$, with all gates of width $\geq \kappa$ removed. Furthermore, suppose that we can learn all the degree-ℓ Pauli coefficients of O_{P_i} to precision η, i.e.*

$$\left| \widehat{O}_{P_i}(Q) - \widetilde{O}_{P_i}(Q) \right| \leq \eta, \quad \forall Q \in \{P \in \mathcal{P}^n : |P| \leq \ell\}. \tag{14}$$

Leveraging these learned coefficients, Algorithm 1 will produce learned observable $\widetilde{O}_{P_i}^{(\ell)}$, such that

$$\frac{1}{2^n} \left\| \widetilde{O}_{P_i}^{(\ell)} - O_{P_i} \right\|_F^2 = \epsilon_{P_i} \leq 2 \cdot 4^\ell \cdot \eta^2 + \epsilon^* \tag{15}$$

Therefore, leveraging Theorem 7 and setting $\kappa = \mathcal{O}(\log n)$ (which implies $\ell = \mathcal{O}(\log^d n)$), we establish that, in quasi-polynomial sample- and time-complexity, we can learn an approximate observable $\widetilde{O}_{P_i}^{(\ell)}$, supported on $\mathcal{O}(\log^d n)$ qubits, which is $1/\text{poly}(n)$-close to O_{P_i}. Figure 1 offers a conceptual illustration of the key observables involved in our algorithm and their relations. For a detailed description of these learning results and proofs, we refer the reader to Appendix D.1.

Efficient Sewing of QAC^0 Heisenberg-Evolved Observables. This procedure is used to efficiently learn all $3n$ of the QAC^0 Heinsenberg-evolved single-qubit Pauli observables. Huang et al. (2024)'s light-cone sewing procedure (with a modified projection operator) is then used to sew all $3n$ learned observables $\widetilde{O}_{P_i}^{(\ell)}$ into approximate unitary C_{sew}, close to $C \otimes C^\dagger$ in average-case distance.

Lemma 10 (Sewing error bound) *Suppose C is an n-qubit QAC^0 circuit, which has a set of Heisenberg-evolved observables $\{O_{P_i}\}_{i,P}$, corresponding to each of the $3n$ single-qubit Paulis P_i. Let $\left\{\widetilde{O}_{P_i}^{(\ell)}\right\}_{i,P}$ denote the set of learned observables, which are at most ϵ_{P_i}-far from the true observables. By "sewing" the learned observables, construct the unitary*

$$C_{sew}\left(\{\widetilde{O}_{P_i}^{(\ell)}\}_{i,P}\right) := \mathrm{SWAP}^{\otimes n} \prod_{i=1}^{n}\left[\mathrm{Proj}_U\left(\frac{1}{2}I\otimes I + \frac{1}{2}\sum_{P\in\{X,Y,Z\}}\widetilde{O}_{P_i}^{(\ell)}\otimes P_i\right)\right], \qquad (16)$$

where Proj_U is the projection onto the unitary minimizing the Frobenius norm distance and $\mathrm{SWAP}^{\otimes n}$ swaps the first and last n qubits. The average-case distance between C_{sew} and $C\otimes C^\dagger$ is at most

$$\mathcal{D}_{avg}(C_{sew}, C\otimes C^\dagger) \le \frac{1}{2}\sum_{i=1}^{n}\sum_{P\in\{X,Y,Z\}}\epsilon_{P_i}. \qquad (17)$$

Since our result leverages average-case distance measures, we had to re-prove most of the learning guarantees of Huang et al. (2024), which were only proven for worst-case distance measures. Furthermore, to ensure that C_{sew} is unitary, the average-case version of the sewing procedure leverages the Proj_U projection onto the unitary minimizing the Frobenius norm distance. Solving for this unitary is the solution to the well-established Orthogonal Procrustes Problem, which requires computing a singular value decomposition (SVD) and setting all singular values to one. In general, computing the SVD of a $2n$-qubit unitary requires exponential time. However, we imposed that the learned observables $\widetilde{O}_{P_i}^{(\ell)}$ are supported on only $\mathcal{O}(\log^d n)$ qubits. This implies that each matrix, $\frac{1}{2}I\otimes I + \frac{1}{2}\sum_{P\in\{X,Y,Z\}}\widetilde{O}_{P_i}^{(\ell)}\otimes P_i$, in the projection also has $\mathcal{O}(\log^d n)$ support. Since we only need to compute the SVD of the non-trivial subsystem within the support, the computational complexity of the sewing procedure is only quasi-polynomial, i.e. $\mathcal{O}(2^{\mathrm{poly}\log n})$. This is the central reason that we impose that the learned observables are not only low-degree, but also low-support.

By plugging our $1/\mathrm{poly}(n)$ error bound for ϵ_{P_i} into Theorem 10, as well as summing the sample and time complexities of the learning and sewing procedures, we can guarantee that our algorithm efficiently learns n-output QAC^0 unitaries, thereby proving Theorem 1. For a more detailed description of these sewing results and proofs, we refer the readers to Appendix D.2.

Learning QAC Circuits with Improved Depth. With our efficient procedure for learning the unitary corresponding to an n-output QAC^0 circuit, one could ask whether we could learn a QAC^0 circuit synthesizing this unitary, also known as "proper learning" of QAC^0. In this work, while we do not achieve a proper learning algorithm for QAC^0, we do make progress towards one. A naive attempt at implementing the learned QAC^0 unitary would require a QAC circuit of quasi-polynomial depth. However, in Theorem 2 we establish that the circuit depth can be reduced to poly-logarithmic.

Begin by noticing that the learned QAC^0 unitary C_{sew}, as given in Equation (16), sews the learned unitaries $W_i = \mathrm{Proj}_U\left(\frac{1}{2}I\otimes I + \frac{1}{2}\sum_{P\in\{X,Y,Z\}}\widetilde{O}_{P_i}^{(\ell)}\otimes P_i\right)$ via SWAP gate operations, which are implementable in QAC^0. To achieve a poly-logarithmic depth QAC implementation of C_{sew}, we prove the following two key results: 1) each unitary W_i can be implemented to high accuracy via a QAC^0 circuit and 2) there exists a sewing order with parallelization such that all n of the W_i unitaries can be implemented and sewed in worst-case poly-logarithmic depth.

To prove the first result, we decompose the QAC^0 circuit structure into two key parts: the CZ_k gates and the arbitrary single-qubit rotation gates. Since W_i is supported on only $\mathcal{O}(\log^d n)$ qubits, a QAC^0 circuit operating on $\mathcal{O}(\log^d n)$ qubits has at most quasi-polynomial configurations of CZ_k gates with $2 \leq k \leq \mathcal{O}(\log n)$. For each such configuration, we show that it is possible to construct a polynomial sized $1/\mathrm{poly}(n)$-net over the polynomial number of arbitrary single-qubit rotation gates. Therefore the $1/\mathrm{poly}(n)$-net over all such QAC^0 circuit architectures has quasi-polynomial size and a simple brute-force search can be leveraged to find the circuit closest to W_i.

To prove the second result, demonstrating that the learned QAC^0 circuits can be sewn into a poly-logarithmic depth QAC circuit, we leverage a graph-coloring argument similar to (Huang et al., 2024, Lemma 13). Note that the poly-logarithmic depth arises due to the poly-logarithmic support of the W_i unitaries, resulting in a poly-logarithmic degree and, thus, worst-case coloring of the graph. For more details about this unitary synthesis procedure, we refer the reader to Appendix D.3.

Concentration and Learning of QAC^0 with Ancilla. In Appendix E, we discuss the applicability of our learning algorithm to QAC^0 circuits with ancilla. Similar to Huang et al. (2024), we only consider circuits where the ancillas are initialized to the $|0^a\rangle$ state and the computation is clean (meaning ancillas are reverted to $|0^a\rangle$ by the end of the computation). This implies that the action of C on the $(n+a)$-qubit system is equivalent to the action of another unitary A on just the n-qubit system without ancillas, i.e. $C(I \otimes |0^a\rangle) = A \otimes |0^a\rangle$. Thus, we define the Heisenberg-evolved Pauli observables of this system "without ancilla restriction" as $O_{P_i,n+a} = C(P_i \otimes I^a)C^\dagger$ and "with ancilla restriction" as $O_{P_i,n} = (I \otimes \langle 0^a|) \cdot O_{P_i,n+a} \cdot (I \otimes |0^a\rangle) = A P_i A^\dagger$. Via Pauli analysis, we prove that the Pauli weight spectrums of $O_{P_i,n+a}$ and $O_{P_i,n}$ are related by a 2^a multiplicative factor or, in other words, an exponential blow-up factor in the number of ancilla.

Lemma 11 (Effect of ancilla) *Let $\mathcal{S} \subseteq \mathcal{P}^n$ be a subset of the set of n-qubit Paulis, then*

$$\mathbf{W}^{\in \mathcal{S}}[O_{P_i,n}] \leq 2^a \cdot \mathbf{W}^{\in \mathcal{S}}[O_{P_i,n+a}]. \tag{18}$$

We also show that the distance between $O_{P_i,n}$ and $O^*_{P_i,n}$ (the circuit with large CZ gates removed) can be bounded by the distance between $O_{P_i,n+a}$ and $O^*_{P_i,n+a}$, but that a 2^a multiplicative blow-up factor also arises relative to the distance in the ancilla-free case.

Lemma 12 (Large CZ_k removal error with ancillas) *The average-case distance between observables of QAC^0 circuit C and QAC^0 circuit $\widetilde{C}$ (C with all m CZ_k's of size $k > \kappa$ removed) satisfies*

$$\frac{1}{2^n} \left\| O_{P_i,n} - O^*_{P_i,n} \right\|_F^2 \leq \frac{1}{2^n} \| O_{P_i,n+a} - O^*_{P_i,n+a} \|_F^2 \leq 2^a \cdot \frac{9m^2}{2^\kappa}. \tag{19}$$

Therefore, by using Algorithm 1 to learn the n-qubit observables with ancilla restriction,

$$\frac{1}{2^n} \left\| \widetilde{O}^{(\ell)}_{P_i,n} - O_{P_i,n} \right\|_F^2 \leq 2 \cdot 4^\ell \cdot \eta^2 + \mathbf{W}^{\notin \mathcal{S}^*}[O_{P_i,n}] \leq 2 \cdot 4^\ell \cdot \eta^2 + 2^a \cdot \frac{9m^2}{2^\kappa} \tag{20}$$

where η is the learning accuracy. To retain our algorithm's efficient sample and time complexity, we maintain that $\kappa \leq \mathcal{O}(\log(n+a)) \leq \mathcal{O}(\log n)$. However, due to the 2^a factor in the numerator of the $\mathbf{W}^{\notin \mathcal{S}^*}[O_{P_i}]$ term, the distance is guaranteed to be $\leq 1/\mathrm{poly}(n)$ only if the number of ancillas is logarithmic, i.e. $a = \mathcal{O}(\log n)$. Thus, we prove that our learning procedure holds for QAC^0 circuits with a logarithmic number of ancilla.

However, if the 2^a term was eliminated from the upper-bound on $\mathbf{W}^{\notin \mathcal{S}^*}[O_{P_i,n}]$ in Equation (20), then the distance would be $\leq 1/\text{poly}(n)$ and $\kappa \leq \mathcal{O}(\log(n+a)) \leq \mathcal{O}(\log n)$ would be satisfied for a number of ancillas a polynomial in n. This observation motivates Theorem 4, a strengthening of (Nadimpalli et al., 2024, Conjecture 1). If proven true, Theorem 5 implies a quasi-polynomial algorithm for learning unitaries of QAC^0 circuits with polynomially many ancillas.

Hardness of Learning QAC^0**.** We conclude the work by proving Theorem 3, as discussed in Section 2 and Appendix F, which demonstrates that QAC^0 cannot be efficiently learned according to the diamond-norm distance measure – motivating our use of an average-case measure.

5. Discussion and Future Work

There are several interesting directions for future work that would strengthen our understanding of QAC^0 and improve the applicability of this learning procedure.

Ancilla. Although our learning procedure handles logarithmic ancilla, many quantum algorithms require far more ancilla to, e.g., perform block-encodings and error correction. We proved that our algorithm works for polynomial ancilla under Theorem 4, but proving this conjecture remains an important and challenging open problem. Concurrent work by Anshu et al. (2024) proved that QAC^0 projectors are well-approximated by low-degree projectors according to the spectral norm, with guarantees for slightly superlinear ancilla, but it is not clear if this result extends to our setting. Furthermore, in this work, we prove algorithmic guarantees for QAC^0 circuits that perform clean computation on their ancillas. However, Nadimpalli et al. (2024) also demonstrated that their low-degree QAC^0 concentration results hold for circuits which do not perform clean computation and even for circuits with dirty ancilla qubits. Therefore, it could be interesting to explore the feasbility of our learning algorithm in these different ancilla settings.

Proper Learning. While our unitary synthesis procedure guarantees a poly-logarithmic depth circuit implementation, we believe that proper learning of the QAC^0 circuit should be possible. Our current circuit synthesis approach is via brute-force search and an ε-net, which we believe leaves room for improvement:

Open Problem 1 (Quasi-polynomial proper learning QAC^0**)** *Given the learned unitary* C_{sew}*, which is* $1/\text{poly}(n)$*-close to* QAC^0 *circuit* C*, does there exist a quasi-polynomial time algorithm to learn a* QAC^0 *circuit implementing unitary* C_{sew}^* *such that* $\mathcal{D}_{avg}(C_{sew}^*, C \otimes C^\dagger) \leq 1/\text{poly}(n)$*?*

Although it is widely believed that efficient AC^0 proper learners do not exist, AC^0 and QAC^0 with sublinear ancilla are fundamentally incomparable[8]. Thus, efficient proper learning of QAC^0 with sublinear ancilla would not contradict the classical belief.

Time Complexity. Classically, Kharitonov (1993) showed that Linial et al. (1993)'s quasi-polynomial complexity for learning AC^0 is optimal, under the standard classical cryptographic hardness-of-factoring assumption. Since factoring is not quantumly hard, Arunachalam et al. (2021) established a quasi-polynomial time lower-bound for *quantum* learning of AC^0 via a reduction to Ring Learning with Errors (RLWE). Although there are fundamental differences between AC^0 and QAC^0, we

---

8. To see why: Anshu et al. (2024) showed that Parity$\notin \text{QAC}^0$[sublinear]. Since Parity=Fanout quantumly, so Fanout $\notin \text{QAC}^0$[sublinear]. However, Fanout$\in \text{AC}^0$ by definition.

believe that it should be possible (potentially through similar crypotgraphic reductions) to prove a quasi-polynomial lower-bound for learning QAC^0. This would imply optimality of our algorithm.

Sample Complexity. Our learning algorithm's core subroutine can be interpreted as learning a junta approximation of QAC^0 Heisenberg-evolved Pauli observables. To this end, there is a growing literature of sophisticated algorithms for quantum junta learning (Chen et al., 2023b; Bao and Yao, 2023; Gutierrez, 2024). While these results do not directly apply to learning the entire n-qubit unitary of a QAC^0 circuit, they open up intriguing possibilities for improving our algorithm's sample complexity with respect to parameters such as circuit size.

Beyond QAC^0. Finally, our work contributes to the burgeoning field of research seeking to identify provably trainable and learnable quantum circuit families. Our results extend Huang et al. (2024)'s efficient learnability of QNC^0 to the broader class of QAC^0. This advance raises intriguing questions for future exploration. For instance, since QAC^0 requires linear depth to implement in 1D geometry, can we efficiently learn other polynomial-depth circuit families in the 1D geometry? Furthermore, since $QAC^0 \subseteq QNC^1 \subseteq QNC$, a natural progression would be to investigate efficient learnability of polylog-depth QNC circuits or other superclasses of QAC^0. By uncovering rigorous algorithms for training and learning quantum circuits, we may enable automated design of quantum circuits, protocols, and algorithms, in addition to characterization of experimental quantum devices.

Acknowledgments

References

Scott Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*, pages 325–338, 2018.

Anurag Anshu, Yangjing Dong, Fengning Ou, and Penghui Yao. On the Computational Power of QAC0 with Barely Superlinear Ancillae. *arXiv preprint arXiv:2410.06499*, 2024.

Srinivasan Arunachalam, Alex Bredariol Grilo, and Aarthi Sundaram. Quantum hardness of learning shallow classical circuits. *SIAM Journal on Computing*, 50(3):972–1013, 2021.

Costin Bădescu and Ryan O'Donnell. Improved quantum data analysis. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1398–1411, 2021.

Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. Structure learning of Hamiltonians from real-time evolution. *arXiv preprint arXiv:2405.00082*, 2024.

Zongbo Bao and Penghui Yao. Nearly optimal algorithms for testing and learning quantum junta channels. *arXiv preprint arXiv:2305.12097*, 2023.

Charles H Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and weaknesses of quantum computing. *SIAM journal on Computing*, 26(5):1510–1523, 1997.

Juan Bermejo-Vega, Dominik Hangleiter, Martin Schwarz, Robert Raussendorf, and Jens Eisert. Architectures for quantum simulation showing a quantum speedup. *Physical Review X*, 8(2): 021010, 2018.

Sergey Bravyi, David Gosset, and Robert König. Quantum advantage with shallow circuits. *Science*, 362(6412):308–311, 2018.

Sergey Bravyi, David Gosset, Robert König, and Marco Tomamichel. Quantum advantage with noisy shallow circuits. *Nature Physics*, 16(10):1040–1045, 2020.

Jacob C Bridgeman and Christopher T Chubb. Hand-waving and interpretive dance: an introductory course on tensor networks. *Journal of physics A: Mathematical and theoretical*, 50(22):223001, 2017.

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

Matthias C Caro, Hsin-Yuan Huang, Marco Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J Coles. Generalization in quantum machine learning from few training data. *Nature communications*, 13(1):4919, 2022.

Matthias C Caro, Hsin-Yuan Huang, Nicholas Ezzell, Joe Gibbs, Andrew T Sornborger, Lukasz Cincio, Patrick J Coles, and Zoë Holmes. Out-of-distribution generalization for learning quantum dynamics. *Nature Communications*, 14(1):3751, 2023.

Senrui Chen, Sisi Zhou, Alireza Seif, and Liang Jiang. Quantum advantages for Pauli channel estimation. *Physical Review A*, 105(3):032435, 2022.

Senrui Chen, Yunchao Liu, Matthew Otten, Alireza Seif, Bill Fefferman, and Liang Jiang. The learnability of Pauli noise. *Nature Communications*, 14(1):52, 2023a.

Senrui Chen, Changhun Oh, Sisi Zhou, Hsin-Yuan Huang, and Liang Jiang. Tight bounds on pauli channel learning without entanglement. *Physical Review Letters*, 132(18):180805, 2024a.

Sitan Chen, Jaume de Dios Pont, Jun-Ting Hsieh, Hsin-Yuan Huang, Jane Lange, and Jerry Li. Predicting quantum channels over general product distributions. *arXiv preprint arXiv:2409.03684*, 2024b.

Thomas Chen, Shivam Nadimpalli, and Henry Yuen. Testing and learning quantum juntas nearly optimally. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1163–1185. SIAM, 2023b.

Yuxuan Du, Min-Hsiu Hsieh, and Dacheng Tao. Efficient Learning for Linear Properties of Bounded-Gate Quantum Circuits. *arXiv preprint arXiv:2408.12199*, 2024.

Steven T Flammia and Ryan O'Donnell. Pauli error estimation via population recovery. *Quantum*, 5:549, 2021.

Steven T Flammia and Joel J Wallman. Efficient estimation of Pauli channels. *ACM Transactions on Quantum Computing*, 1(1):1–32, 2020.

Xun Gao, Sheng-Tao Wang, and L-M Duan. Quantum supremacy for simulating a translation-invariant ising spin model. *Physical review letters*, 118(4):040502, 2017.

Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. Efficient learning of quantum states prepared with few non-clifford gates. *arXiv preprint arXiv:2305.13409*, 2023.

Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. Improved stabilizer estimation via bell difference sampling. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1352–1363, 2024.

Francisco Escudero Gutierrez. Learning junta distributions and quantum junta states, and QAC0 circuits. *arXiv preprint arXiv:2410.15822*, 2024.

Jeongwan Haah, Robin Kothari, and Ewin Tang. Learning quantum Hamiltonians from high-temperature Gibbs states and real-time evolutions. *Nature Physics*, pages 1–5, 2024.

Jonas Haferkamp, Dominik Hangleiter, Adam Bouland, Bill Fefferman, Jens Eisert, and Juani Bermejo-Vega. Closing gaps of a quantum advantage with short-time hamiltonian dynamics. *Physical Review Letters*, 125(25):250501, 2020.

Dominik Hangleiter and Jens Eisert. Computational advantage of quantum random sampling. *Reviews of Modern Physics*, 95(3):035001, 2023.

Peter Høyer and Robert Špalek. Quantum fan-out is powerful. *Theory of computing*, 1(1):81–103, 2005.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, et al. Quantum advantage in learning from experiments. *Science*, 376(6598):1182–1186, 2022.

Hsin-Yuan Huang, Sitan Chen, and John Preskill. Learning to predict arbitrary quantum processes. *PRX Quantum*, 4(4):040337, 2023.

Hsin-Yuan Huang, Yunchao Liu, Michael Broughton, Isaac Kim, Anurag Anshu, Zeph Landau, and Jarrod R. McClean. Learning shallow quantum circuits, 2024.

Michael Kharitonov. Cryptographic hardness of distribution-specific learning. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, pages 372–381, 1993.

Ching-Yi Lai and Hao-Chung Cheng. Learning quantum circuits of some T gates. *IEEE Transactions on Information Theory*, 68(6):3951–3964, 2022.

Lorenzo Leone, Salvatore FE Oliviero, and Alioscia Hamma. Learning t-doped stabilizer states. *Quantum*, 8:1361, 2024.

Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *Journal of the ACM (JACM)*, 40(3):607–620, 1993.

Ashley Montanaro and Tobias J Osborne. Quantum boolean functions. *Chicago Journal of Theoretical Computer Science*, 1:1–45, 2010.

Cristopher Moore. Quantum circuits: Fanout, parity, and counting. *arXiv preprint quant-ph/9903046*, 1999.

Shivam Nadimpalli, Natalie Parham, Francisca Vasconcelos, and Henry Yuen. On the Pauli Spectrum of QAC0. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, page 1498–1506, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703836. doi: 10.1145/3618260.3649662. URL https://doi.org/10.1145/3618260.3649662.

Michael A Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Physics Letters A*, 303(4):249–252, October 2002. ISSN 0375-9601. doi: 10.1016/s0375-9601(02)01272-0. URL http://dx.doi.org/10.1016/S0375-9601(02)01272-0.

Michael A Nielsen and Isaac L Chuang. *Quantum Computation and Quantum Information*. Cambridge university press, 2010.

Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM (JACM)*, 56(6):1–40, 2009.

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

Barbara M Terhal and David P DiVincenzo. Adaptive quantum computation, constant depth quantum circuits and arthur-merlin games. *arXiv preprint quant-ph/0205133*, 2002.

Adam Bene Watts and Natalie Parham. Unconditional quantum advantage for sampling with shallow circuits. *arXiv preprint arXiv:2301.00995*, 2023.

Adam Bene Watts, Robin Kothari, Luke Schaeffer, and Avishay Tal. Exponential separation between shallow quantum circuits and unbounded fan-in shallow classical circuits. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 515–526, 2019.

Mark M Wilde. *Quantum Information Theory*. Cambridge university press, 2013.

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

Haimeng Zhao, Laura Lewis, Ishaan Kannan, Yihui Quek, Hsin-Yuan Huang, and Matthias C Caro. Learning quantum states and unitaries of bounded gate complexity. *arXiv preprint arXiv:2310.19882*, 2023.

Appendix A. Preliminaries

In this section, we will briefly establish notation and important concepts to be used throughout the paper. This manuscript will assume familiarity with the basics of quantum computational and information theory. For more background, we refer the interested reader to Nielsen and Chuang (2010); Wilde (2013). In particular, the preliminary sections of Nadimpalli et al. (2024); Huang et al. (2024) are especially relevant to this work. For more information on tensor networks, as used in Figure 2, we refer the reader to Bridgeman and Chubb (2017).

In terms of notation, we will generally use n to refer to the number of qubits in the computational register of a quantum system and a to refer to the number of ancilla qubits. Let $[n] = \{1, 2, \ldots, n\}$ denote the set of qubit indices. We will use the notation I_k to refer to an identity matrix of dimension $2^k \times 2^k$ and, when the dimension is implied by the context, may drop the k subscript altogether.

Given an n-qubit unitary U corresponding to a quantum circuit, we will define the circuit's corresponding unitary channel as

$$\mathcal{U}(\rho) = U\rho U^\dagger. \tag{21}$$

Implicitly, we will assume that unitary channels are perfectly implemented, without any noise. Furthermore, in this work, ancilla qubits are assumed to be restricted to input $|0\rangle\langle 0|$. For the same quantum circuit, implementing unitary U, and a given Hermitian measurement observable M, the circuit's Heisenberg-evolved observable (corresponding to the dual channel) is defined as

$$O_M = U^\dagger M U. \tag{22}$$

Note that if the measurement observable M is unitary, then O_M is also unitary.

A.1. Circuit Classes

This work will refer to several classical and quantum circuit classes. We briefly review them here for ease of reference.

Classically, NC^k is the class of circuits of depth $\mathcal{O}(\log^k n)$ and size $\mathrm{poly}(n)$, comprised of AND, OR, and NOT gates of fan-in ≤ 2. The circuit class $\mathsf{NC} = \bigcup_{k \geq 1} \mathsf{NC}^k$ refers to the union of all these circuit classes. The circuit classes AC^k and AC are defined analogously, but allow for AND and OR gates of unbounded fan-in.

Quantumly, QNC^k is the class of quantum circuits of depth $\mathcal{O}(\log^k n)$, comprised of unlimited arbitrary single-qubit gates and polynomially many CZ gates (that act on 2-qubits). The circuit class $\mathsf{QNC} = \bigcup_{k \geq 1} \mathsf{QNC}^k$ refers to the union of all these circuit classes. The circuit classes QAC^k and QAC are defined analogously, but allow for CZ gates that simultaneously operate on an unbounded number number of qubits.

A.2. The Average-Case Distance Measure

This work will leverage average-case distance measures, so we will now review their important properties. The presentation and intuition are largely based on prior works Huang et al. (2024); Nielsen (2002). To begin, a standard notion of distance for quantum states is the fidelity measure.

Definition 13 (Fidelity) *Given two quantum states ρ, σ the fidelity of the states is defined as*

$$\mathcal{F}(\rho, \sigma) = \mathrm{Tr}\left(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)^2. \tag{23}$$

If one of the states is a pure state, e.g. $\sigma = |\psi\rangle\langle\psi|$, then the fidelity expression reduces to

$$\mathcal{F}(\rho, \sigma) = \langle\psi| \rho |\psi\rangle. \tag{24}$$

Note that for any input (ρ, σ), the fidelity is bounded as $\mathcal{F}(\rho, \sigma) \in [0, 1]$. Furthermore, the maximal value of 1 is obtained if and only if the states are identical, i.e. $\mathcal{F}(\rho, \rho) = 1$.

For quantum channels, we consider the standard Haar distance measure, which is the same notion of average-case distance for comparing quantum channels as utilized in (Huang et al., 2024, Definition 3).

Definition 14 (Average-Case Distance) *The average-case distance between two n-qubit CPTP maps $\mathcal{E}_1$ and $\mathcal{E}_2$ is defined as*

$$\mathcal{D}_{avg}(\mathcal{E}_1, \mathcal{E}_2) = \mathop{\mathbb{E}}_{|\psi\rangle\sim\mathrm{Haar}} \left[1 - \mathcal{F}\left(\mathcal{E}_1(|\psi\rangle\langle\psi|), \mathcal{E}_2(|\psi\rangle\langle\psi|)\right)\right], \tag{25}$$

where $|\psi\rangle$ is sampled from the Haar (uniform) measure and $\mathcal{F}$ is the fidelity.

Intuitively, this notion of average-case distance measures how distinct the two channel outputs are, for pure input states averaged over the Haar (uniform) measure. In the case of unitary channels the Haar distance measure simplifies to the average gate fidelity measure. As such, we will often abuse notation for unitary channels and write $\mathcal{D}_{\mathrm{avg}}(U_1, U_2)$ to mean $\mathcal{D}_{\mathrm{avg}}(\mathcal{U}_1, \mathcal{U}_2)$.

Fact 1 (Average Gate Fidelity - Nielsen (2002)) *For unitaries U_1 and U_2, with corresponding unitary channels $\mathcal{U}_1$ and $\mathcal{U}_2$, the average-case distance satisfies*

$$\mathcal{D}_{avg}(\mathcal{U}_1, \mathcal{U}_2) = \frac{2^n}{2^n + 1} \left(1 - \frac{1}{4^n} \left|\mathrm{Tr}\left(U_1^\dagger U_2\right)\right|^2\right). \tag{26}$$

In this work we will also leverage properties of the well-established normalized Frobenius and global phase-invariant distance measures for unitaries.

Definition 15 (Normalized Frobenius Distance) *The normalized Frobenius distance between two n-qubit unitaries U and V is defined as*

$$\mathcal{D}_F(U, V) = \frac{1}{2^n}\|U - V\|_F^2 \tag{27}$$

Definition 16 (Global Phase-Invariant Distance) *The global phase-invariant distance between two n-qubit unitaries U and V is defined as*

$$\mathcal{D}_P(U, V) = \min_{\phi\in\mathbb{R}} \frac{1}{2^n}\|e^{i\phi}U - V\|_F^2 \tag{28}$$

As distance measures, both satisfy the triangle inequality. Furthermore, from these definitions, it trivially follows that the normalized Frobenius distance is lower-bounded by the global phase-invariant distance.

Fact 2 *For unitaries U_1 and U_2, the global phase-invariant distance is upper-bounded by the normalized Frobenius distance,*

$$\mathcal{D}_P(U_1, U_2) \leq \mathcal{D}_F(U_1, U_2). \tag{29}$$

Huang et al. (2024) also showed that this average-case distance can be upper-bounded by the Frobenius norm distance between the unitaries.

Fact 3 *For unitaries U_1 and U_2, with corresponding unitary channels $\mathcal{U}_1$ and $\mathcal{U}_2$, the average-case distance is upper-bounded by the global phase-invariant distance,*

$$\mathcal{D}_{avg}(\mathcal{U}_1, \mathcal{U}_2) \leq \mathcal{D}_P(U_1, U_2). \tag{30}$$

A.3. Pauli Analysis

In this work, we will make frequent use of concepts from so-called "Pauli analysis" or "quantum Boolean functions", as originally defined by Montanaro and Osborne (2010), which is a quantum analog of classical analysis of Boolean functions O'Donnell (2014).

Central to Pauli analysis are the four standard Pauli operators:

$$I := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{31}$$

Note that the identity matrix is often thought of as the "trivial" Pauli operator, as it has no effect on a quantum system. We will denote the set of single-qubit Pauli operators as $\mathcal{P} = \{I, X, Y, Z\}$. Let $\mathcal{P}^n = \{I, X, Y, Z\}^{\otimes n}$ denote the set of all n-qubit Pauli strings.

In general, given an n-qubit Pauli string $Q \in \mathcal{P}^n$, we use Q_i to refer to the ith Pauli of Pauli string Q. Similarly, given a set of indices $\mathcal{S} \subseteq [n]$, $Q_{\mathcal{S}}$ refers to the Pauli sub-string

$$Q_{\mathcal{S}} = \bigotimes_{i \in [n]} Q_i^{\delta\{i \in \mathcal{S}\}}, \tag{32}$$

with the convention that $Q_i^0 = I$. Note, however, that we will abuse notation and, in the case of single-qubit Pauli $P \in \mathcal{P}$ use P_i to refer to the "single-qubit Pauli"

$$P_i := I^{\otimes(i-1)} \otimes P \otimes I^{\otimes(n-i)}, \tag{33}$$

which applies P to the ith qubit in an n-qubit system. We define the support of a Pauli string $Q \in \mathcal{P}^n$ to be the set of qubit indices upon which the Pauli acts non-trivially, that is

$$\text{supp}(Q) = \{i \in [n] : Q_i \neq I\}. \tag{34}$$

The degree of a Pauli string Q is the size of its support,

$$|Q| = |\text{supp}(Q)|. \tag{35}$$

Central to Pauli analysis is the observation that the set of Pauli strings $\mathcal{P}^n$ forms an orthonormal basis for any n-qubit quantum unitary $U \in U(2^n)$,

$$U = \sum_{P \in \mathcal{P}^n} \widehat{U}(P) \cdot P, \quad \text{where } \widehat{U}(P) = \frac{1}{2^n} \text{Tr}\left(U^\dagger P\right). \tag{36}$$

Note that $\widehat{U}(P)$ are typically referred to as the Pauli coefficients of U.

For a subset of Pauli strings, $\mathcal{S} \subseteq \mathcal{P}^n$, we define the unitary's Pauli weight on that subset as

$$\mathbf{W}^{\in \mathcal{S}}[U] = \sum_{Q \in \mathcal{S}} |\widehat{U}(Q)|^2. \tag{37}$$

For notational convenience, we will use the notation $\mathbf{W}^{\notin \mathcal{S}}$ to refer to the weight in subset $\bar{\mathcal{S}} = \mathcal{P}^n \backslash \mathcal{S}$ and the notation $\mathbf{W}^{>k}$ to refer to the weight of Paulis with degree greater than k, i.e. $\mathcal{S} = \{P \in \mathcal{P}^n : |P| > k\}$. Finally, note that Parseval's formula holds for the Pauli decomposition.

Fact 4 (Parseval's Formula) *For unitary $U \in U(2^n)$,*

$$\frac{1}{2^n} \|U\|_F^2 = \sum_{P \in \mathcal{P}^n} |\widehat{U}(P)|^2 = \mathbf{W}^{\in \mathcal{P}^n}[U] = 1 \tag{38}$$

Finally, important to this work, we will distinguish between the support and degree of a unitary. The support of a unitary is the set of all qubits upon which the unitary acts non-trivially. Formally, for the n-qubit unitary U defined in Equation (36), its support is the union of the supports of all the Pauli strings P for which $\widehat{U}(P) \neq 0$, i.e.

$$\mathrm{supp}(U) = \bigcup_{P \in \mathcal{P}^n : \hat{U}(P) \neq 0} \{i \in [n] : P_i \neq I\}. \tag{39}$$

If n-qubit unitary has support k, where k is small relative to n, we refer to U as low-support. In analogy to the classical setting, this is similar to a k-junta. Meanwhile, the degree of a unitary is equal to the maximum degree of any Pauli string P for which $\widehat{U}(P) \neq 0$. Formally, for the n-qubit unitary U defined in Equation (36), its degree is

$$|U| = \max_{P \in \mathcal{P}^n : \hat{U}(P) \neq 0} |P|. \tag{40}$$

If n-qubit unitary has degree k, where k is small relative to n, we refer to U as low-degree. It is clear that any unitary with support k has degree k. However, a unitary with degree k could have, for example, non-zero Pauli coefficients on all Pauli strings of degree k and, thus, support n. Therefore, a unitary of degree k does not necessarily have support k. This implies that any low-support unitary is neccesarily a low-degree unitary, but any low-degree unitary is not neccesarily a low-support unitary. Note that this is analogous to how, in the classical setting, all k-juntas necessarily have degree-k, but a degree-k Boolean function is not necessarily a k-junta.

A.4. Classical Shadow Tomography

Central to our learning results will be the classical shadow tomography procedure of Huang et al. (2020), which enables efficient learning of $\mathrm{Tr}(O_i \rho)$ for an arbitrary quantum state ρ and a set of observables $\{O_i\}_i$. In particular, we will leverage the improved sample-complexity achieved by the shadow-norm result of Huang et al. (2023) for Pauli observables.

Lemma 17 (Classical Shadows for Low-Degree Pauli Observables - Huang et al. (2020, 2023))
Assume we are given error parameter $\epsilon > 0$, failure probability $\delta \in [0, 1]$, degree $\ell \geq 1$, and

$$N = \mathcal{O}\left(\frac{3^\ell}{\epsilon^2} \log\left(\frac{n^\ell}{\delta}\right)\right) \tag{41}$$

copies of unknown n-qubit state ρ that we can make random Pauli measurements on. Let $\mathcal{M}$ be a set of n-qubit Pauli matrices of degree $\leq \ell$, i.e.

$$\mathcal{M} \subseteq \{P \in \mathcal{P}^n : |P| \leq \ell\}. \tag{42}$$

With probability at least $1 - \delta$, we can output an estimate $\widetilde{s}(P)$ for each $P \in \mathcal{M}$ such that

$$|\widetilde{s}(P) - \mathrm{Tr}(P\rho)| \leq \epsilon. \tag{43}$$

The computational complexity of this procedure scales as $\mathcal{O}(|\mathcal{M}| \cdot N)$.

A.5. Unitary Projection

Our sewing procedure will require an efficient procedure for projecting non-unitary matrices onto unitary matrices. Specifically, we require a projection which minimizes the Frobenius norm distance. As such, for an arbitrary matrix A, we will define $\mathrm{Proj}_U(A)$ to be the projection of A onto the unitary matrix that minimizes Frobenius norm distance,

$$\mathrm{Proj}_U(A) = \min_{B \in U(2^n)} \|A - B\|_F^2. \tag{44}$$

Note that this minimization task is the well-established "Orthogonal Procrustes Problem" and has an efficiently computable, simple solution.

Fact 5 (Orthogonal Procrustes Problem) *For $m \times m$ matrix A with singular value decomposition $A = U\Sigma V^\dagger$,*

$$\mathrm{Proj}_U(A) = UV^\dagger, \tag{45}$$

which can be computed in $\mathcal{O}(m^3)$ time.

A.6. The CZ_k Gate

In this work, we will use CZ_k to refer to a control-Z gate acting on k-qubits. Note that, unlike other quantum "controlled" operations, the CZ_k does not need to distinguish between target and control qubits, since it only applies a phase when activated by the all ones state. Mathematically, the CZ_k gate can be described by its action on the computational basis states $|i\rangle \in \{|0\rangle, |1\rangle\}^{\otimes k}$, as

$$CZ_k |i\rangle = \begin{cases} -|i\rangle, & \text{if } |i\rangle = |1_k\rangle \\ |i\rangle, & \text{otherwise} \end{cases}, \tag{46}$$

where we use $|1_k\rangle = |111...1\rangle$ to denote the all ones state. Thus, the CZ_k unitary can be decomposed in terms of projectors as

$$CZ_k = -|1_k\rangle\langle 1_k| + \sum_{\substack{j \in \{0,1\}^k \\ j \neq 1_k}} |j\rangle\langle j|. \tag{47}$$

Alternatively, the CZ_k unitary can be expressed in terms of its Pauli decomposition.

Lemma 18 (CZ_k **Pauli Decomposition**) *The Pauli decomposition of the CZ_k gate acting on k qubits is given by*

$$CZ_k = \sum_{P_Z \in \{I,Z\}^{\otimes k}} \widehat{\alpha}_{CZ_k}(P_Z) \cdot P_Z, \tag{48}$$

with Pauli coefficients

$$\widehat{\alpha}_{CZ_k}(P_Z) = \begin{cases} 1 - 2^{-k+1}, & \text{if } \deg(P_Z) = 0 \\ 2^{-k+1}, & \text{if } \deg(P_Z) = \text{odd}, \\ -2^{-k+1}, & \text{if } \deg(P_Z) \neq 0 \text{ and } \deg(P_Z) = \text{even}. \end{cases} \tag{49}$$

Note that we use notation P_Z to emphasize that Pauli strings consisting of any X or Y Paulis have zero Fourier mass.

Proof [Proof of Theorem 18] Using the CZ_k projector decomposition given in Equation (47), we can calculate the CZ_k Pauli coefficient of for any Pauli string $P \in \{I, Z\}^{\otimes k}$ as

$$\widehat{\alpha}_{CZ_k}(P) = \frac{1}{2^k} \text{Tr}\left(P^\dagger CZ_k\right) \tag{50}$$

$$= -\frac{1}{2^k} \text{Tr}\left(P |1_k\rangle\langle 1_k|\right) + \frac{1}{2^k} \sum_{\substack{j \in \{0,1\}^k \\ j \neq 1_k}} \text{Tr}\left(P |j\rangle\langle j|\right) \tag{51}$$

$$= -\frac{1}{2^k} \langle 1_k| P |1_k\rangle + \frac{1}{2^k} \sum_{\substack{j \in \{0,1\}^k \\ j \neq 1_k}} \langle j| P |j\rangle \tag{52}$$

$$= -\frac{1}{2^k} \langle 1_k| P |1_k\rangle + \frac{1}{2^k} \left(\text{Tr}(P) - \langle 1_k| P |1_k\rangle\right) \tag{53}$$

$$= \frac{1}{2^k} \text{Tr}(P) - \frac{2}{2^k} \langle 1_k| P |1_k\rangle \tag{54}$$

$$= \frac{1}{2^k} \cdot 2^k \cdot \delta(P = I_k) - \frac{1}{2^{k-1}} \cdot (-1)^{\delta(\deg(P)=\text{odd})} \tag{55}$$

$$= \delta(P = I_k) + \frac{1}{2^{k-1}} \cdot (-1)^{\delta(\deg(P)=\text{even})}. \tag{56}$$

∎

Note that this k-qubit decomposition can straightforwardly be extended to a CZ_k gate acting on a k-qubit subset of an n-qubit system, as

$$CZ_k \otimes I_{n-k} = \sum_{P_Z \in \{I,Z\}^{\otimes k}} \widehat{\alpha}_{CZ_k}(P_Z) \cdot P_Z \otimes I_{n-k}, \tag{57}$$

where it is assumed without loss of generality that the CZ_k gate acts on the first k qubits of the system. From hereon, when we write CZ_k, the identity on the remaining $n - k$ qubits, i.e. I_{n-k}, will be assumed.

Appendix B. Choi Representations and Heisenberg-Evolved Observables

The work of Nadimpalli et al. (2024) studied the Choi representation of quantum channels, while the work of Huang et al. (2024) considered the Heisenberg-evolved observables of quantum circuits. Similar to Huang et al. (2024), in this work, we consider QAC^0 Heisenberg-evolved single-qubit Pauli observables. However, we will now establish that they are in fact closely related to the single-output QAC^0 channel Choi representations of Nadimpalli et al. (2024).

In particular, the work of Nadimpalli et al. (2024) studied channels of the form

$$\mathcal{E}_C(\rho) = \mathrm{Tr}_{n-1}(C\rho C^\dagger), \tag{58}$$

where C is the unitary implemented by a QAC^0 circuit and ρ is a density matrix. Their results were presented in terms of the channel's Choi representation, given by

$$\Phi_{\mathcal{E}_C} = \sum_{x,y \in \{0,1\}^n} |x\rangle\langle y| \otimes \mathcal{E}_C(|x\rangle\langle y|). \tag{59}$$

Note that $\Phi_{\mathcal{E}_C}$ operates on $n+1$-qubits, where we denote the n-qubit register corresponding to the channel input as the "in" register and the single qubit corresponding to the channel output as the "out" register. From the definition of the Choi representation, it follows that

$$\mathcal{E}_C(\rho) = \mathrm{Tr}_{\mathrm{in}}\left(\Phi_{\mathcal{E}_C}\left(I_{\mathrm{out}} \otimes \rho_{\mathrm{in}}^\top\right)\right). \tag{60}$$

Measuring the output of this channel with respect to a single-qubit observable O results in the expectation value $\mathrm{Tr}(O\mathcal{E}_C(\rho))$. Via algebraic manipulation of this expectation, we can solve for the dual channel $\mathcal{E}_C^\dagger(O)$,

$$\mathrm{Tr}(O\mathcal{E}_C(\rho)) = \mathrm{Tr}\left(O \cdot \mathrm{Tr}_{n-1}(C\rho C^\dagger)\right) = \mathrm{Tr}\left((I_{n-1} \otimes O)C\rho C^\dagger\right) = \mathrm{Tr}\left(\underbrace{(C^\dagger(I_{n-1} \otimes O)C)}_{\mathcal{E}_C^\dagger(O)}\rho\right). \tag{61}$$

Therefore, the Heisenberg-evolved single-qubit observable of this single-output QAC^0 circuit is

$$\mathcal{E}_C^\dagger(O) = C^\dagger(I_{n-1} \otimes O)C. \tag{62}$$

Alternatively, leveraging Equation (60),

$$\mathrm{Tr}\left(O\mathcal{E}_C(\rho)\right) = \mathrm{Tr}\left(O\,\mathrm{Tr}_{\mathrm{in}}\left(\Phi_{\mathcal{E}_C}(I_{\mathrm{out}} \otimes \rho_{\mathrm{in}}^\top)\right)\right) \tag{63}$$

$$= \mathrm{Tr}\left((O_{\mathrm{out}} \otimes I_{\mathrm{in}})\Phi_{\mathcal{E}_C}(I_{\mathrm{out}} \otimes \rho_{\mathrm{in}}^\top)\right) \tag{64}$$

$$= \mathrm{Tr}\left(\mathrm{Tr}_{\mathrm{out}}\left((O_{\mathrm{out}} \otimes I_{\mathrm{in}})\Phi_{\mathcal{E}_C}\right)\rho_{\mathrm{in}}^\top\right) \tag{65}$$

$$= \mathrm{Tr}\left(\underbrace{\mathrm{Tr}_{\mathrm{out}}\left((O_{\mathrm{out}} \otimes I_{\mathrm{in}})\Phi_{\mathcal{E}_C}\right)^\top}_{\mathcal{E}_C^\dagger(O)}\rho_{\mathrm{in}}\right), \tag{66}$$
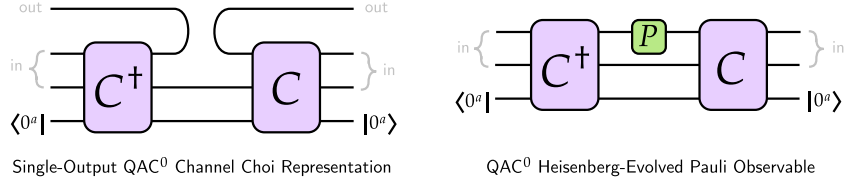
23

Figure 2: Tensor network diagrams illustrating the similar structures of the single-output QAC^0 channel Choi representations studied in the work of Nadimpalli et al. (2024) and the QAC^0 Heisenberg-evolved Pauli observables considered in Huang et al. (2024) and in this work. Here, C is the QAC^0 circuit's unitary and P is a single-qubit observable.

we can also express the Heisenberg-evolved observable in terms of the Choi representation, as

$$\mathcal{E}_C^\dagger(O) = \mathrm{Tr}_{\mathrm{out}}\left((O_{\mathrm{out}} \otimes I_{\mathrm{in}})\Phi_{\mathcal{E}_C}\right)^\top. \tag{67}$$

Thus, Equation (62) and Equation (67) directly establish the link between QAC^0 Heisenberg-evolved single-qubit observables and single-qubit output QAC^0 Choi representations.

Intuitively, the Choi representation contains a full description of the channel and can be used to calculate the channel expectation for any input state and observable pair (ρ, O). Meanwhile, the Heisenberg-evolved observable restricts the channel output to a single measurement observable O, but can be used to compute the expectation for any input state ρ. This relationship is illustrated by the tensor network diagrams in Figure 2.

Finally, we can explicitly relate the Pauli coefficients of each QAC^0 Heisenberg-evolved single-qubit Pauli observable to a Fourier coefficient of the single-output QAC^0 channel Choi representation.

Proposition 19 *For an n-qubit Pauli $Q \in \mathcal{P}^n$, the Pauli coefficient of the Heisenberg-evolved observable $O_{P_{out}}$ is related to that of the single-output Choi representation $\Phi_{\mathcal{E}_C}$ as*

$$\widehat{O}_{P_{out}}(Q) = 2 \cdot \widehat{\Phi}_{\mathcal{E}_C}(P_{out} \otimes Q) \cdot (-1)^{\delta\{Q \text{ has an odd \# of Pauli } Ys\}}. \tag{68}$$

Proof [Proof of Theorem 19] For an n-qubit Pauli $Q \in \mathcal{P}^n$, leveraging Equation (67),

$$\widehat{O}_{P_{\mathrm{out}}}(Q) = \frac{1}{2^n}\,\mathrm{Tr}(O_{P_{\mathrm{out}}} \cdot Q) \tag{69}$$

$$= \frac{1}{2^n}\,\mathrm{Tr}\left(\mathrm{Tr}_{\mathrm{out}}((P_{\mathrm{out}} \otimes I_{\mathrm{in}})\Phi_{\mathcal{E}_C})^\top Q\right) \tag{70}$$

$$= \frac{1}{2^n}\,\mathrm{Tr}\left(\mathrm{Tr}_{\mathrm{out}}\left((P_{\mathrm{out}} \otimes I_{\mathrm{in}})\Phi_{\mathcal{E}_C}(I_{\mathrm{out}} \otimes Q^\top)\right)\right) \tag{71}$$

$$= \frac{1}{2^n}\,\mathrm{Tr}\left(\mathrm{Tr}_{\mathrm{out}}\left((P_{\mathrm{out}} \otimes Q^\top) \cdot \Phi_{\mathcal{E}_C}\right)\right) \tag{72}$$

$$= 2 \cdot \frac{1}{2^{n+1}}\,\mathrm{Tr}\left((P_{\mathrm{out}} \otimes Q^\top) \cdot \Phi_{\mathcal{E}_C}\right) \tag{73}$$

$$= 2 \cdot \widehat{\Phi}_{\mathcal{E}_C}\left(P_{\mathrm{out}} \otimes Q^\top\right) \tag{74}$$

$$= 2 \cdot \widehat{\Phi}_{\mathcal{E}_C} \left(P_{\text{out}} \otimes Q \right) \cdot (-1)^{\delta \{ Q \text{ has an odd \# of Pauli } Ys \}}. \tag{75}$$

This completes the proof. ∎

Appendix C. Concentration of QAC^0 Heisenberg-Evolved Observables

Assume we are given an n-qubit QAC^0 circuit governed by unitary,

$$C = U_1 \cdot CZ_{k_1} \cdot U_2 \cdot CZ_{k_2} \cdots CZ_{k_{m-1}} \cdot U_m \cdot CZ_{k_m} \cdot U_{m+1} = \prod_{i=1}^{m} \left(U_i \cdot CZ_{k_i} \right) \cdot U_{m+1}. \tag{76}$$

where $\{k_1, ..., k_m\} \in [\kappa, n]$ are the sizes of the large CZ_{k_i} gates in the circuit, such that $\kappa \geq 2$ is the size of the smallest. For reasons that will become apparent later, we will assume that $\kappa = \mathcal{O}(\log(n))$. The unitaries U_i correspond to circuits consisting of arbitrary single-qubit gates and CZ_k gates of width $k \leq \kappa$. The single qubit Heisenberg-evolved observables of circuit C are denoted

$$O_{P_i} = C^\dagger P_i C, \tag{77}$$

where $P_i \in \{X, Y, Z\}_i \otimes I_{[n] \setminus i}$ is a non-trivial Pauli on the i-th qubit. Since the QAC^0 circuit can have CZ_k gates of unbounded width, in the worst case it is supported on all qubits, i.e.

$$|\text{supp}(O_{P_i})| \leq \mathcal{O}(n). \tag{78}$$

Finally, we denote the Pauli decomposition of this observable as

$$O_{P_i} = \sum_{Q \in \mathcal{P}^n} \widehat{O}_{P_i}(Q) \cdot Q. \tag{79}$$

In this section, we will show that circuit C is in fact well approximated by the same circuit with all the CZ_{k_i} gates of size $k_i \geq \kappa = \mathcal{O}(\log(n))$ removed. We will denote this approximate circuit as

$$\widetilde{C} = U_1 \cdot U_2 \cdots U_m \cdot U_{m+1} = \prod_{i=1}^{m+1} U_i \tag{80}$$

and its single-qubit Heisenberg-evolved observables as

$$O_{P_i}^* = \widetilde{C}^\dagger P_i \widetilde{C}. \tag{81}$$

Since the circuit only has gates of width $\leq \kappa = \mathcal{O}(\log(n))$, note that it has a much smaller support than C, i.e.

$$\ell = |\text{supp}(O_{P_i}^*)| \leq \mathcal{O}\left(\log^d n\right), \tag{82}$$

where d is the circuit depth (which is constant for QAC^0 circuits). The Pauli decomposition of these approximate observables will be expressed as

$$O_{P_i}^* = \sum_{Q \in \mathcal{P}^n : |Q| < \ell} \widehat{O}_{P_i}^*(Q) \cdot Q. \tag{83}$$

With these definitions, we will now leverage the proof techniques of Nadimpalli et al. (2024) to establish low-degree and low-support concentration of Heisenberg-evolved observables, so as to obtain an upperbound for $\mathbf{W}^{>k}[O_{P_i}]$ by relating it to $O_{P_i}^*$.

C.1. Low-Degree Concentration

We begin by demonstrating that the low-degree spectral concentration result of (Nadimpalli et al., 2024, Theorem 21) for single-output QAC^0 channel Choi representations also holds for QAC^0 Heisenberg-evolved single-qubit Pauli observables. Note, that this proof will only be for QAC^0 circuits without ancilla qubits. For more discussion on the effect of ancillas we refer the reader to Appendix E.

Proposition 20 (Low-Degree Concentration) *Suppose C is a depth-d, size-s QAC^0 circuit acting on n qubits. Let $O_{P_i} = C^\dagger P_i C$ be a Heisenberg-evolved single-qubit Pauli observable. Then for every degree $k \in [n]$,*

$$\mathbf{W}^{>k}[O_{P_i}] \leq \mathcal{O}\left(s^2 2^{-k^{1/d}}\right). \tag{84}$$

At a high-level, this proof will follow that of (Nadimpalli et al., 2024, Theorem 21) and consist of two key steps. First, we will establish that if the QAC^0 circuit has no CZ gates of width greater than $k^{1/d}$, then $\mathbf{W}^{>k}[O_{P_i}] = 0$. Second, we will show that removing these "large" CZ gates does not significantly change the Heisenberg-evolved observable under the average-case measure.

Analogous to (Nadimpalli et al., 2024, Lemma 20), we begin by showing that the weight spectrum of the QAC^0 Heisenberg-evolved observable is zero for any degree greater than the size of the observable's support.

Lemma 21 *Let O_{P_i} be an observable corresponding to a circuit C measured with respect to P_i. If O_{P_i} is supported on ℓ qubits, i.e. $|\mathrm{supp}(O_{P_i})| = \ell$, then the observable's weight is zero for any degree $> \ell$, i.e.*

$$\mathbf{W}^{>\ell}[O_{P_i}] = 0. \tag{85}$$

Proof [Proof of Theorem 21] Given the circuit C and an observable P_i, we will decompose the circuit as

$$C = D_{P_i} L_{P_i}, \tag{86}$$

where L_{P_i} is the unitary corresponding to all gates in the circuit that are in the backwards light-cone of P_i and D_{P_i} is the unitary corresponding to the gates which are not. Thus, by the standard light-cone argument,

$$C^\dagger P_i C = D_{P_i}^\dagger L_{P_i}^\dagger P_i L_{P_i} D_{P_i} = L_{P_i}^\dagger P_i L_{P_i}. \tag{87}$$

Plugging this into our expression for the weight of observable O_{P_i} at degree $> k$,

$$\mathbf{W}^{>k}[O_{P_i}] = \sum_{|Q|>k} \widehat{O}_{P_i}(Q)^2 = \sum_{|Q|>k} \frac{1}{4^n} \mathrm{Tr}(O_{P_i} Q)^2 = \sum_{|Q|>k} \frac{1}{4^n} \mathrm{Tr}\left(C^\dagger P_i C Q\right)^2 \tag{88}$$

$$= \sum_{|Q|>k} \frac{1}{4^n} \mathrm{Tr}\left(L_{P_i}^\dagger P_i L_{P_i} Q\right)^2, \tag{89}$$

we see that the weight only depends on gates in the backwards light-cone of P_i. If Q has degree greater than $|\mathrm{supp}(O_{P_i})| = \ell$, this implies that there must exist at least one non-identity term in

Q which corresponds to an identity term in O_{P_i}, meaning $\mathrm{Tr}\left(L_{P_i}^\dagger P_i L_{P_i} Q\right) = 0$. Therefore, for $|Q| > \ell$, $\widehat{O}_{P_i}(Q)^2 = 0$, which implies that $\mathbf{W}^{>\ell}[O_{P_i}] = 0$. ∎

Analogous to (Nadimpalli et al., 2024, Lemma 23), we can also upper-bound the distance between the true observable, O_{P_i}, and the observable with "large" CZ gates removed, $O_{P_i}^*$.

Lemma 22 *Let O_{P_i} be an observable corresponding to a QAC^0 circuit C measured with respect to P_i. Let $O_{P_i}^*$ be an observable corresponding to the QAC^0 circuit $\widetilde{C}$ (circuit C with all m CZ_k's of size $k > \kappa$ removed), measured with respect to P_i. The average-case distance between these two is observables is upper-bounded by*

$$\mathcal{D}_F(O_{P_i}, O_{P_i}^*) \le \epsilon^* = \frac{9m^2}{2^\kappa}. \tag{90}$$

Proof [Proof of Theorem 22] Let us define the sub-circuits C_j and $\widetilde{C}_j$ of C and $\widetilde{C}$, respectively, as

$$C_j = U_j \prod_{i=j}^m \left(CZ_{k_i} U_{i+1}\right), \tag{91}$$

$$\widetilde{C}_j = \prod_{i=j}^m U_i. \tag{92}$$

Via unitary-invariance of the Frobenius norm and a hybrid argument,

$$\left\|O_{P_i} - O_{P_i}^*\right\|_F = \left\|C_1^\dagger P_i C_1 - \widetilde{C}_1^\dagger P_i \widetilde{C}_1\right\|_F \tag{93}$$

$$= \left\|CZ_{k_1}^\dagger C_2^\dagger P_i C_2 CZ_{k_1} - \widetilde{C}_2^\dagger P_i \widetilde{C}_2\right\|_F \tag{94}$$

$$= \left\|CZ_{k_1}^\dagger C_2^\dagger P_i C_2 CZ_{k_1} - C_2^\dagger P_i C_2 + C_2^\dagger P_i C_2 - \widetilde{C}_2^\dagger P_i \widetilde{C}_2\right\|_F^2 \tag{95}$$

$$\le \left\|CZ_{k_1}^\dagger C_2^\dagger P_i C_2 CZ_{k_1} - C_2^\dagger P_i C_2\right\|_F^2 + \left\|C_2^\dagger P_i C_2 - \widetilde{C}_2^\dagger P_i \widetilde{C}_2\right\|_F^2 \tag{96}$$

$$\vdots \tag{97}$$

$$\le \sum_{j=1}^m \left\|CZ_{k_j}^\dagger C_{j+1}^\dagger P_i C_{j+1} CZ_{k_j} - C_{j+1}^\dagger P_i C_{j+1}\right\|_F. \tag{98}$$

Denoting $V_j = C_{j+1}^\dagger P_i C_{j+1}$ and using the fact that

$$CZ_{k_j} = I^{\otimes n} - 2 \cdot I^{\otimes n - k_i} \otimes \left|1^{k_j}\right\rangle\!\left\langle 1^{k_j}\right| = I - 2\left|1^{k_j}\right\rangle\!\left\langle 1^{k_j}\right|, \tag{99}$$

the terms in the summation of Equation (98) are upper-bounded as

$$\left\|CZ_{k_j}^\dagger V_j CZ_{k_j} - V_j\right\|_F = \sqrt{2\,\mathrm{Tr}(I) - 2\,\mathrm{Tr}\left(V_j^\dagger CZ_{k_j} V_j CZ_{k_j}^\dagger\right)} \tag{100}$$

$$= \sqrt{2\,\mathrm{Tr}(I) - 2\,\mathrm{Tr}\left(V_j^\dagger (I - 2\left|1^{k_j}\right\rangle\!\left\langle 1^{k_j}\right|) V_j (I - 2\left|1^{k_j}\right\rangle\!\left\langle 1^{k_j}\right|)\right)} \tag{101}$$

$$= \sqrt{2\operatorname{Tr}(I) - 2\operatorname{Tr}\Big((V_j^\dagger - 2V_j^\dagger \left|1^{k_j}\right\rangle\!\!\left\langle1^{k_j}\right|)(V_j - 2V_j \left|1^{k_j}\right\rangle\!\!\left\langle1^{k_j}\right|)\Big)} \tag{102}$$

$$= \sqrt{2\operatorname{Tr}(I) - 2\operatorname{Tr}\Big(I - 4\left|1^{k_j}\right\rangle\!\!\left\langle1^{k_j}\right| + 4V_j^\dagger \left|1^{k_j}\right\rangle\!\!\left\langle1^{k_j}\right| V_j \left|1^{k_j}\right\rangle\!\!\left\langle1^{k_j}\right|\Big)} \tag{103}$$

$$= \sqrt{8\operatorname{Tr}\big(\left|1^{k_j}\right\rangle\!\!\left\langle1^{k_j}\right|\big) - 8\operatorname{Tr}\Big(V_j^\dagger \left|1^{k_j}\right\rangle\!\!\left\langle1^{k_j}\right| V_j \left|1^{k_j}\right\rangle\!\!\left\langle1^{k_j}\right|\Big)} \tag{104}$$

$$= \sqrt{8 \cdot 2^{n-k_j} - 8\operatorname{Tr}\Big(V_j^\dagger \left|1^{k_j}\right\rangle\!\!\left\langle1^{k_j}\right| V_j \left|1^{k_j}\right\rangle\!\!\left\langle1^{k_j}\right|\Big)} \tag{105}$$

$$\leq 3 \cdot 2^{(n-k_j)/2}. \tag{106}$$

Plugging this upper-bound back into Equation (98) and denoting $\kappa = \min_j k_j$, we obtain the desired upper-bound for the Frobenius distance between the two observables,

$$\mathcal{D}_F(O_{P_i}, O_{P_i}^*) \leq \frac{1}{2^n}\left(\sum_{j=1}^m 3 \cdot 2^{(n-k_j)/2}\right)^2 \leq \frac{1}{2^n}\left(3m \cdot \max_j 2^{(n-k_j)/2}\right)^2 = \frac{9m^2}{2^\kappa}. \tag{107}$$

∎

Analogous to (Nadimpalli et al., 2024, Claim 24), we will now show that the weight of O_{P_i} can be upper-bounded by the weight of $O_{P_i}^*$ and its distance to $O_{P_i}^*$.

Lemma 23 *For any degree k, the weight of O_{P_i} is upperbounded as*

$$\mathbf{W}^{>k}[O_{P_i}] \leq \left(\mathbf{W}^{>k}\left[O_{P_i}^*\right]^{1/2} + \frac{1}{\sqrt{2^n}}\left\|O_{P_i} - O_{P_i}^*\right\|_F\right)^2. \tag{108}$$

Proof [Proof of Theorem 23] Note that the weight expression satisfies the triangle inequality. Thus, the weight of O_{P_i} can be decomposed with respect to $O_{P_i}^*$ as follows.

$$\mathbf{W}^{>k}[O_{P_i}] = \mathbf{W}^{>k}\left[O_{P_i}^* + (O_{P_i} - O_{P_i}^*)\right] \tag{109}$$

$$\leq \left(\mathbf{W}^{>k}\left[O_{P_i}^*\right]^{1/2} + \mathbf{W}^{>k}\left[O_{P_i} - O_{P_i}^*\right]^{1/2}\right)^2 \tag{110}$$

$$\leq \left(\mathbf{W}^{>k}\left[O_{P_i}^*\right]^{1/2} + \frac{1}{\sqrt{2^n}}\left\|O_{P_i} - O_{P_i}^*\right\|_F\right)^2. \tag{111}$$

This concludes the proof. ∎

Leveraging Theorem 21, Theorem 22, and Theorem 23, we can now straightforwardly prove Theorem 20.

Proof [Proof of Theorem 20] Let $O_{P_i}^*$ be defined such that all CZ gates of size $\geq \kappa = k^{1/d}$ are removed. Since the QAC^0 circuit is depth-d, the support of $O_{P_i}^*$ is bounded as $|\mathrm{supp}(O_{P_i}^*)| < \left(k^{1/d}\right)^d = k$. Therefore, by Theorem 21,

$$\mathbf{W}^{>k}[O_{P_i}^*] = 0. \tag{112}$$

Furthermore, plugging $\kappa = k^{1/d}$ into Theorem 22, the distance between O_{P_i} and $O_{P_i}^*$ is bounded as

$$\mathcal{D}_F(O_{P_i}, O_{P_i}^*) \leq \frac{9m^2}{2^{k^{1/d}}} \leq \frac{9s^2}{2^{k^{1/d}}}, \tag{113}$$

where we leveraged the fact that the total number of gates removed m, must be less than the size s of the circuit. Plugging Equation (112) and Equation (113) into the degree-k weight upper bound of Theorem 23, we obtain the desired result. ∎

C.2. Low-Support Concentration

For the computational efficiency of the learning algorithm to be presented in this work, it is crucial to establish that, beyond low-degree concentrated, the O_{P_i} observables' weight spectrum is concentrated on a set of Paulis with small support, i.e. low-support concentrated.

To begin, we modify Theorem 21 to show that the weight of a Heisenberg-evolved observable is zero for all Paulis that lie outside of its support.

Lemma 24 *Let O_{P_i} be an observable corresponding to a circuit C measured with respect to P_i. Let $\mathcal{S} = \{P \in \mathcal{P}^n : P_i = I, \forall i \notin \mathrm{supp}(O_{P_i})\}$ denote the set of Pauli strings in O_{P_i}'s support. Then O_{P_i}'s weight is zero for all Paulis acting non-trivially outside of its support, i.e.*

$$\mathbf{W}^{\notin \mathcal{S}}[O_{P_i}] = 0. \tag{114}$$

Proof [Proof of Theorem 24] Given the circuit C and an observable P_i, we will decompose the circuit as

$$C = D_{P_i} L_{P_i}, \tag{115}$$

where L_{P_i} is the unitary corresponding to all gates in the circuit that are in the backwards light-cone of P_i and D_{P_i} is the unitary corresponding to the gates which are not. Thus, by the standard light-cone argument,

$$C^\dagger P_i C = D_{P_i}^\dagger L_{P_i}^\dagger P_i L_{P_i} D_{P_i} = L_{P_i}^\dagger P_i L_{P_i}. \tag{116}$$

Plugging this into our expression for the weight of observable O_{P_i} for Paulis outside the support,

$$\mathbf{W}^{\notin \mathcal{S}}[O_{P_i}] = \sum_{Q \notin \mathcal{S}} \widehat{O}_{P_i}(Q)^2 = \sum_{Q \notin \mathcal{S}} \frac{1}{4^n} \mathrm{Tr}(O_{P_i} Q)^2 = \sum_{Q \notin \mathcal{S}} \frac{1}{4^n} \mathrm{Tr}\left(C^\dagger P_i C Q\right)^2 \tag{117}$$

$$= \sum_{Q \notin \mathcal{S}} \frac{1}{4^n} \mathrm{Tr}\left(L_{P_i}^\dagger P_i L_{P_i} Q\right)^2, \tag{118}$$

we see that the weight only depends on gates in the backwards light-cone of P_i, and thus in the support of O_{P_i}. If Q acts non-trivially outside $\mathcal{S} = \mathrm{supp}(O_{P_i})$, this implies that there must exist at least one non-identity term in Q which corresponds to an identity term in O_{P_i}, meaning $\mathrm{Tr}\left(L_{P_i}^\dagger P_i L_{P_i} Q\right) = 0$. Therefore, for all $Q \notin \mathcal{S}$, $\widehat{O}_{P_i}(Q)^2 = 0$, which implies that $\mathbf{W}^{\notin \mathcal{S}}[O_{P_i}] = 0$. ∎

Leveraging this result and a proof similar to that of Theorem 23, we achieve a low-support concentration result. In particular, we show that the weight of O_{P_i} outside the set of Paulis in the support of $O_{P_i}^*$ is at most the distance between O_{P_i} and $O_{P_i}^*$, which we proved in Theorem 22 to decay with respect to the size of the support of $O_{P_i}^*$.

Lemma 25 *For $\mathcal{S}^* = \mathrm{supp}(O_{P_i}^*)$, the weight of O_{P_i} outside the support of $O_{P_i}^*$ is upper-bounded as*

$$W^{\notin \mathcal{S}^*}[O_{P_i}] \leq \mathcal{D}_F(O_{P_i}, O_{P_i}^*) \leq \epsilon^*. \tag{119}$$

Proof [Proof of Theorem 23] Note that the weight expression satisfies the triangle inequality. Thus, the weight of O_{P_i} can be decomposed with respect to $O_{P_i}^*$ as follows.

$$\mathbf{W}^{\notin \mathcal{S}^*}[O_{P_i}] = \mathbf{W}^{\notin \mathcal{S}^*}\left[O_{P_i}^* + (O_{P_i} - O_{P_i}^*)\right] \tag{120}$$

$$\leq \left(\mathbf{W}^{\notin \mathcal{S}^*}\left[O_{P_i}^*\right]^{1/2} + \mathbf{W}^{\notin \mathcal{S}^*}\left[O_{P_i} - O_{P_i}^*\right]^{1/2}\right)^2 \tag{121}$$

By Theorem 24, we have that $\mathbf{W}^{\notin \mathcal{S}^*}\left[O_{P_i}^*\right] = 0$. Therefore,

$$\mathbf{W}^{\notin \mathcal{S}^*}[O_{P_i}] \leq \mathbf{W}^{\notin \mathcal{S}^*}\left[O_{P_i} - O_{P_i}^*\right] \leq \mathcal{D}_F(O_{P_i}, O_{P_i}^*). \tag{122}$$

This concludes the proof of this lemma. ∎

Appendix D. Efficient Learning of QAC^0 Circuit Unitaries

Leveraging the low-support concentration of Heisenberg-evolved QAC^0 observables, we will prove the main result of this work – a sample and time efficient algorithm for learning n-output QAC^0 unitaries.

Theorem 26 (Learning shallow circuits with many-qubit gates) *Consider an unknown n-qubit, depth-d QAC^0 circuit governed by unitary C. For error parameter $\epsilon = 1/\mathrm{poly}(n)$ and failure probability $\delta \in (0, 1)$, we can learn a $2n$-qubit unitary C_{sew} such that*

$$\mathcal{D}_{avg}(C_{sew}, C \otimes C^\dagger) \leq \varepsilon, \tag{123}$$

with high probability $1 - \delta$. C_{sew} can be learned with quasi-polynomial sample and time complexity.

At a high-level, the proof of this theorem consists of two main parts:

1. In Appendix D.1, we demonstrate that the QAC^0 circuit's Heisenberg-evolved single-qubit Pauli observables can be efficiently learned to high accuracy. This section is where most of the algorithm's novelty lies. Via classical shadow tomography, we show that an observable with $\mathcal{O}(\log^d n)$-support can be learned in quasi-polynomial sample and time complexity. Then, leveraging the previously-established concentration results, we prove that this low-support learned observable is $1/\mathrm{poly}(n)$-close to the true Heisenberg-evolved observable.

2. In Appendix D.2, we leverage (Huang et al., 2024, Section 5.2.2)'s procedure for "sewing" these learned Heisenberg-evolved Pauli observables into a unitary description of the circuit. Note that novel work is done to guarantee that the sewed unitary is close to the true unitary under the *average-case* distance measure. (In Huang et al. (2024), guarantees were given according to the worst-case measure).

In Appendix D.3, we also offer an efficient procedure to synthesize an explicit poly-logarithmic depth QAC circuit that implements a unitary $\frac{1}{\text{poly}(n)}$-close to $C \otimes C^\dagger$.

D.1. Approximate Learning of QAC^0 Heisenberg-Evolved Observables

In this section, we propose an efficient algorithm, Algorithm 1, for learning an observable $\widetilde{O}_{P_i}^{(\ell)}$, supported on $\ell = \mathcal{O}(\log^d n)$ qubits. The majority of the section will focus on proving the following learning guarantee, which establishes that $\widetilde{O}_{P_i}^{(\ell)}$ is $1/\text{poly}(n)$-close to the true QAC^0 Heisenberg-evolved Pauli observable O_{P_i}.

Lemma 27 *Let $b \geq 2$ and $c \geq 3d$ be constants. Let $\delta \in (0,1)$ be a failure probability. For observable O_{P_i}, with high probability, $1 - \delta$, we can learn an approximate observable $\widetilde{O}_{P_i}^{(\ell)}$ such that*

$$\mathcal{D}_F\left(O_{P_i}, \widetilde{O}_{P_i}^{(\ell)}\right) \leq \epsilon_{P_i} \leq \frac{2}{n^b} + \frac{9d^2}{c^2 \cdot n^{c-2}} = \frac{1}{\text{poly}(n)}. \tag{124}$$

The sample and time complexity of this procedure are quasi-polynomial, $\mathcal{O}\left(n^{\text{poly}\log n} \cdot \log\left(1/\delta\right)\right)$.

Importantly, our proposed algorithm fundamentally differs from those of Huang et al. (2024) and Nadimpalli et al. (2024). In particular, we cannot simply apply (Huang et al., 2024, Lemma 10) because that result assumes that the true observable to be learned is low-degree. However, in our setting, the true QAC^0 observable is not low-degree (just *close to* a low-degree observable in average-case distance), meaning that the procedure's guarantees no longer hold. Meanwhile, the learning algorithm of Nadimpalli et al. (2024) simply leverages the low-degree concentration result to efficiently learn a degree-ℓ Choi representation, which is close to the true single-output QAC^0 channel Choi representation. However, as will become apparent in the next section, to achieve an efficient computational complexity for the observable sewing procedure, it is critical that we learn an observable which has an ℓ-qubit *support*. That is, we need to learn an observable which is not only low-degree, but also low-support. We cannot simply sample from the low-degree support, because apriori we do not actually know which qubits the observable is supported on.

D.1.1. THE LOW-SUPPORT OBSERVABLE LEARNING ALGORITHM

We will now describe Algorithm 1 – our proposed procedure for learning the Heisenberg-evolved single-qubit Pauli observables of a QAC^0 circuit. With this algorithm, we aim to learn, for each observable O_{P_i}, an approximation,

$$\widetilde{O}_{P_i} = \sum_{Q \in \mathcal{P}^n} \widetilde{O}_{P_i}(Q) \cdot Q, \tag{125}$$

such that the normalized Frobenius distance is guaranteed to be small,

$$\mathcal{D}_F\left(O_{P_i}, \widetilde{O}_{P_i}\right) = \sum_{Q \in \mathcal{P}^n} \left|\widehat{O}_{P_i}(Q) - \widetilde{O}_{P_i}(Q)\right|^2 \leq \epsilon_{P_i}. \tag{126}$$

Note that O_{P_i} could be supported on all n qubits, meaning $\widehat{O}_{P_i}(Q)$ could be non-zero for all 4^n possible Paulis Q. Thus, if we were simply to try and learn all the Pauli coefficients of O_{P_i}, we would require a learning algorithm with exponential complexity.

However, in Appendix C, we saw that O_{P_i} is close in distance to $O_{P_i}^*$, where $|\text{supp}(O_{P_i}^*)| \leq \ell = O(\log^d n)$. This implies that O_{P_i} is concentrated on the set of Paulis $S_\ell^* = \{P \in \mathcal{P}^n : P_i = I, \forall i \notin \text{supp}(O_{P_i}^*)\}$. Therefore, we should be able to learn a decent approximation of O_{P_i} simply by learning the S_ℓ^*-truncated approximation,

$$\widetilde{O}_{P_i}^* = \sum_{Q \in S_\ell^*} \widetilde{O}_{P_i}(Q) \cdot Q. \tag{127}$$

Note, however, that we will be learning the coefficients directly from O_{P_i} and do not apriori know which qubits are contained in $\text{supp}(O_{P_i}^*)$. Therefore, we will first need to learn approximations of all the degree-ℓ Pauli coefficients of O_{P_i}, i.e.

$$\widetilde{O}_{P_i}(Q), \quad \forall\, Q \in \mathcal{F}_\ell = \{P \in \mathcal{P}^n : |P| \leq \ell\}. \tag{128}$$

Then, we will select the learned ℓ-qubit support with maximal weight.

Formally, to describe this we will need to introduce a bit of notation. Let $S = \{L \subset [n] : |L| = \ell\}$ denote the set of all possible ℓ-qubit subsets of the n total qubits. For an ℓ-qubit subset $s \in S$, define the set of all Paulis supported on that set as

$$\mathcal{F}_{\{s\}} = \left\{P_s \otimes I_{\bar{s}} \mid \forall P \in \mathcal{P}^\ell\right\}. \tag{129}$$

Finally, denote the set of all possible ℓ-qubit supports as

$$\mathfrak{F}_\ell = \left\{\mathcal{F}_{\{s\}} \mid \forall s \in S\right\}. \tag{130}$$

Thus, the set of Paulis in an ℓ-qubit support with maximal weight is defined as

$$T_\ell = \arg\max_{\mathcal{F}_{\{s\}} \in \mathfrak{F}_\ell} \sum_{Q \in S} \left|\widetilde{O}_{P_i}(Q)\right|^2. \tag{131}$$

We will set any learned coefficients outside this set to zero, i.e.

$$\widetilde{O}_{P_i}^{(\ell)}(Q) = \begin{cases} \widetilde{O}_{P_i}(Q), & \text{if } Q \in T_\ell, \\ 0, & \text{otherwise} \end{cases}. \tag{132}$$

Thus, Algorithm 1 will learn the observable

$$\widetilde{O}_{P_i}^{(\ell)} = \sum_{Q \in T_\ell} \widetilde{O}_{P_i}^{(\ell)}(Q) \cdot Q. \tag{133}$$

32

D.1.2. OBSERVABLE LEARNING GUARANTEES

We will now prove guarantees for Algorithm 1 – namely, that the learned observable $\widetilde{O}_{P_i}^{(\ell)}$ is $1/\text{poly}(n)$-close to the true observable O_{P_i}.

Lemma 28 (Observable Learning Guarantees) *Let O_{P_i} be a QAC^0 Heisenberg-evolved Pauli observable, which is ϵ^*-close to the observable $O_{P_i}^*$, with all gates of width $\geq \kappa$ removed. Furthermore, suppose that we can learn all the degree-ℓ Pauli coefficients of O_{P_i} to precision η, i.e.*

$$\left| \widehat{O}_{P_i}(Q) - \widetilde{O}_{P_i}(Q) \right| \leq \eta, \quad \forall Q \in \{P \in \mathcal{P}^n : |P| \leq \ell\}. \tag{134}$$

Leveraging these learned coefficients, Algorithm 1 will produce a learned observable $\widetilde{O}_{P_i}^{(\ell)}$, such that

$$\mathcal{D}_F\left(\widetilde{O}_{P_i}^{(\ell)}, O_{P_i} \right) \leq 2 \cdot 4^\ell \cdot \eta^2 + \epsilon^*. \tag{135}$$

Proof [Proof of Theorem 28] The error can be decomposed as,

$$\mathcal{D}_F\left(\widetilde{O}_{P_i}^{(\ell)}, O_{P_i} \right) = \sum_{Q \in \mathcal{P}^n} \left| \widehat{O}_{P_i}(Q) - \widetilde{O}_{P_i}^{(\ell)}(Q) \right|^2 \tag{136}$$

$$= \sum_{Q \in T_\ell} \left| \widehat{O}_{P_i}(Q) - \widetilde{O}_{P_i}^{(\ell)}(Q) \right|^2 + \sum_{Q \notin T_\ell} \left| \widehat{O}_{P_i}(Q) - \widetilde{O}_{P_i}^{(\ell)}(Q) \right|^2 \tag{137}$$

$$\leq |T_\ell| \cdot \eta^2 + \sum_{Q \notin T_\ell} \left| \widehat{O}_{P_i}(Q) \right|^2. \tag{138}$$

In order to establish the desired error bound, we will upper-bound $\sum_{Q \notin T_\ell} \left| \widehat{O}_{P_i}(Q) \right|^2$. Leveraging Equation (131) and the triangle inequality,

$$\sum_{Q \in S_\ell^*} \left| \widetilde{O}_{P_i}^{(\ell)}(Q) \right|^2 \leq \sum_{Q \in T_\ell} \left| \widetilde{O}_{P_i}^{(\ell)}(Q) \right|^2 \tag{139}$$

$$\leq \sum_{Q \in T_\ell} \left(\left| \widehat{O}_{P_i}(Q) \right|^2 + \left| \widetilde{O}_{P_i}^{(\ell)}(Q) - \widehat{O}_{P_i}(Q) \right|^2 \right) \tag{140}$$

$$\leq |T_\ell| \cdot \eta^2 + \sum_{Q \in T_\ell} \left| \widehat{O}_{P_i}(Q) \right|^2, \tag{141}$$

which implies that

$$\sum_{Q \in T_\ell} \left| \widehat{O}_{P_i}(Q) \right|^2 \geq \left(\sum_{Q \in S_\ell^*} \left| \widetilde{O}_{P_i}^{(\ell)}(Q) \right|^2 \right) - |T_\ell| \cdot \eta^2. \tag{142}$$

Leveraging Parseval's (Fact 4) and the fact that $\widetilde{O}_{P_i}^{(\ell)}$ only has non-zero coefficients for Paulis in T_ℓ,

$$\sum_{Q \notin T_\ell} \left| \widehat{O}_{P_i}(Q) \right|^2 \leq |T_\ell| \cdot \eta^2 + \sum_{Q \notin S_\ell^*} \left| \widetilde{O}_{P_i}^{(\ell)}(Q) \right|^2 = |T_\ell| \cdot \eta^2 + \sum_{Q \notin S_\ell^*, Q \in T_\ell} \left| \widetilde{O}_{P_i}^{(\ell)}(Q) \right|^2 \tag{143}$$

Using the triangle inequality and bound from Theorem 25, we obtain an upper-bound in terms of η and ϵ^*,

$$\sum_{Q \notin T_\ell} \left| \widehat{O}_{P_i}(Q) \right|^2 \leq |T_\ell| \cdot \eta^2 + \sum_{Q \notin S_\ell^*, Q \in T_\ell} \left| \widehat{O}_{P_i}(Q) - \widetilde{O}_{P_i}^{(\ell)}(Q) \right|^2 + \sum_{Q \notin S_\ell^*, Q \in T_\ell} \left| \widehat{O}_{P_i}(Q) \right|^2 \quad (144)$$

$$\leq |T_\ell| \cdot \eta^2 + \sum_{Q \in T_\ell} \left| \widehat{O}_{P_i}(Q) - \widetilde{O}_{P_i}^{(\ell)}(Q) \right|^2 + \sum_{Q \notin S_\ell^*} \left| \widehat{O}_{P_i}(Q) \right|^2 \quad (145)$$

$$\leq |T_\ell| \cdot \eta^2 + |T_\ell| \cdot \eta^2 + \epsilon^* \quad (146)$$

$$= 2 \cdot |T_\ell| \cdot \eta^2 + \epsilon^* \quad (147)$$

Noting that T_ℓ is the set of Pauli coefficients in the support of ℓ-qubits, $|T_\ell| = 4^\ell$, we obtain the desired result. ∎

Leveraging Theorem 28, classical shadow tomography Huang et al. (2020) (as described in Theorem 17), and some of our prior concentration results, we can now prove Theorem 27. In particular, we will now prove that the learning procedure requires quasi-polynomial sample and time complexity.

Proof [Proof of Theorem 27] Leveraging our bound on the distance between O_{P_i} and $\widetilde{O}_{P_i}^{(\ell)}$ from Theorem 28, our goal is to show that

$$\mathcal{D}_F \left(O_{P_i}, \widetilde{O}_{P_i}^{(\ell)} \right) \leq \epsilon_{P_i} = 2 \cdot 4^\ell \cdot \eta^2 + \epsilon^* \leq \frac{1}{\text{poly}(n)}. \quad (148)$$

Setting $\ell = C \cdot \log^d n$ (where $C = c^d$ and $c \geq 3d$), this implies that $m = d \cdot \lfloor \frac{n}{\kappa} \rfloor$ gates of width at least $\kappa = c \cdot \log n$ are removed from O_{P_i} to obtain $O_{P_i}^*$. Plugging this into Theorem 22 implies that the distance between O_{P_i} and $O_{P_i}^*$ is bounded as

$$\epsilon^* \leq \frac{9m^2}{2^\kappa} \leq \frac{9 \cdot d^2 \cdot \lfloor \frac{n}{\kappa} \rfloor^2}{2^\kappa} \leq \frac{9 \cdot d^2 \cdot n^2}{\kappa^2 \cdot 2^\kappa} = \frac{9 \cdot d^2 \cdot n^2}{c^2 \log^2 n \cdot 2^{\log(n^c)}} \leq \frac{9d^2}{c^2 n^{c-2}} \quad (149)$$

Therefore, setting the observable learning accuracy to

$$\eta^2 = \frac{1}{n^b \cdot 4^\ell}, \quad (150)$$

for some constant $b \geq 2$, achieves the desired error bound:

$$\epsilon_{P_i} = 2 \cdot 4^\ell \cdot \eta^2 + \epsilon^* \leq \frac{2}{n^b} + \frac{9d^2}{c^2 \cdot n^{c-2}} \leq \frac{2}{n^b} + \frac{1}{n^{3d-2}} = \frac{1}{\text{poly}(n)}. \quad (151)$$

Now all that remains is to prove the sample and computational complexity. Denote the set of n-qubit Paulis of degree $\leq \ell$ as

$$\mathcal{F}_\ell = \{ P \in \mathcal{P}^n : |P| \leq \ell \}. \quad (152)$$

Via classical shadow tomography, with the state $\rho = O_{P_i}$ and the set of bounded-degree Pauli observables $\mathcal{F}_\ell$, we can produce an η-estimate $\widetilde{O}_{P_i}(Q)$ of

$$\widehat{O}_{P_i}(Q) = \text{Tr}(Q \cdot O_{P_i}) \quad (153)$$

for each $Q \in \mathcal{F}_\ell$. Since the Pauli observables are all of degree at most ℓ, by Theorem 17, the sample complexity of this classical shadow tomography procedure is

$$\mathcal{O}\left(\frac{3^\ell}{\eta^2} \log\left(\frac{n^\ell}{\delta}\right)\right) = \mathcal{O}\left(3^\ell \cdot n^b \cdot 4^\ell \cdot \log\left(\frac{n^\ell}{\delta}\right)\right) = \mathcal{O}\left(n^b \cdot 4^{2\ell} \cdot \log\left(\frac{n^\ell}{\delta}\right)\right). \tag{154}$$

Furthermore, since the size of $\mathcal{F}_\ell$, or the total number of Pauli of degree $\leq \ell$, is upper-bounded as

$$|\mathcal{F}_\ell| = \sum_{k=1}^{\ell} 3^k \cdot \binom{n}{k} \leq \mathcal{O}(n^\ell), \tag{155}$$

the computational complexity is

$$\mathcal{O}\left(n^{\ell+b} \cdot 4^{2\ell} \cdot \log\left(\frac{n^\ell}{\delta}\right)\right). \tag{156}$$

As described in Algorithm 1, to produce the approximation $\widetilde{O}_{P_i}^{(\ell)}$, we also need to find the set of Paulis supported on ℓ qubits with maximal weight and set all other Pauli coefficients to zero. Since there are $\binom{n}{\ell}$ possible supports of size ℓ and calculating the weight of each support involves summing over 3^ℓ different Pauli coefficients, the computational complexity of finding the set T_ℓ is quasi-polynomial, i.e.

$$\binom{n}{\ell} \cdot 3^\ell \leq \mathcal{O}(n^\ell). \tag{157}$$

To achieve the final, explicit sample and computational complexity, plug in the value $\ell = C \cdot \log^d n$, where $C = c^d$ and $c \geq 3d$. Thus, the overall sample complexity is quasi-polynomial

$$\mathcal{O}\left(n^b \cdot 4^{2\ell} \cdot \log\left(\frac{n^\ell}{\delta}\right)\right) = \mathcal{O}\left(n^b \cdot 2^{4C \log^d n} \cdot \log\left(\frac{n^{C \log^d n}}{\delta}\right)\right) \tag{158}$$

$$= \mathcal{O}\left(n^b \cdot 2^{4C \log^d n} \cdot \left(C \log^d n \cdot \log n - \log \delta\right)\right) \tag{159}$$

$$= \mathcal{O}\left(n^b \cdot 2^{4c^d \log^d n} \cdot \left(\log^{d+1} n + \log(1/\delta)\right)\right) \tag{160}$$

$$= \mathcal{O}\left(2^{\text{poly} \log n} \cdot \log(1/\delta)\right), \tag{161}$$

and the computational complexity is also quasi-polynomial

$$\mathcal{O}\left(n^{\ell+b} \cdot 4^{2\ell} \cdot \log\left(\frac{n^\ell}{\delta}\right) + n^\ell\right) = \mathcal{O}\left(n^{\ell+b} \cdot 4^{2\ell} \cdot \log\left(\frac{n^\ell}{\delta}\right)\right) \tag{162}$$

$$= \mathcal{O}\left(n^{c^d \log^d n + b} \cdot 2^{4c^d \log^d n} \cdot \left(\log^{d+1} n + \log(1/\delta)\right)\right) \tag{163}$$

$$= \mathcal{O}\left(n^{\text{poly} \log n} \cdot \log(1/\delta)\right). \tag{164}$$

∎

D.2. Sewing QAC^0 Heisenberg-Evolved Observables

We will now use a procedure analogous to that of (Huang et al., 2024, Section 5.2.2) to sew these Heisenberg-evolved Pauli observables and project them onto a unitary which is $1/\mathrm{poly}(n)$-close to the true QAC^0 circuit unitary, with respect to the average-case distance.

Analogous to (Huang et al., 2024, Lemma 9), we begin by showing that the error in the sewing procedure can be upper-bounded by the sum of the learned observables' learning error (from the last section). Importantly, note that our proof differs from that of Huang et al. (2024), since we are leveraging an average-case instead of a worst-case distance measure.

Lemma 29 (Observable Sewing Guarantees) *Suppose C is an n-qubit QAC^0 circuit, which has a set of Heisenberg-evolved observables $\{O_{P_i}\}_{i,P}$, corresponding to each of the $3n$ possible single-qubit Paulis P_i. Let $\left\{\widetilde{O}_{P_i}^{(\ell)}\right\}_{i,P}$ denote the set of learned observables, which are each at most ϵ_{P_i}-far from the respective true observable. Construct the unitary*

$$C_{sew} := \mathrm{SWAP}^{\otimes n} \prod_{i=1}^{n} \left[\mathrm{Proj}_U \left(\frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} \widetilde{O}_{P_i}^{(\ell)} \otimes P_i \right) \right], \tag{165}$$

by "sewing" the learned observables, where Proj_U is the projection onto the unitary minimizing Frobenius norm, as defined in Equation (44), and $\mathrm{SWAP}^{\otimes n}$ swaps the first and last n qubits. The average-case distance between C_{sew} and $C \otimes C^\dagger$ is at most

$$\mathcal{D}_{avg}(C_{sew}, C \otimes C^\dagger) \le \frac{1}{2} \sum_{i=1}^{n} \sum_{P \in \{X,Y,Z\}} \epsilon_{P_i}. \tag{166}$$

Proof [Proof of Theorem 29] To begin, note that $C \otimes C^\dagger$, can be decomposed in terms of the true observables as

$$C \otimes C^\dagger = \mathrm{SWAP}^{\otimes n} \prod_{i=1}^{n} \left[\frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} O_{P_i} \otimes P_i \right]. \tag{167}$$

Also, define the following three matrices:

$$V_i = \frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} O_{P_i} \otimes P_i \tag{168}$$

$$\widetilde{W}_i = \frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} \widetilde{O}_{P_i}^{(\ell)} \otimes P_i \tag{169}$$

$$W_i = \mathrm{Proj}_U \left(\widetilde{W}_i \right). \tag{170}$$

Noting that C_{sew} and $C \otimes C^\dagger$ implement unitary channels, we can leverage Fact 3 and Fact 2 to upper-bound the average gate fidelity of the channels by the normalized Frobenius distance of the corresponding unitaries,

$$\mathcal{D}_{\mathrm{avg}}(C_{\mathrm{sew}}, C \otimes C^\dagger) \le \mathcal{D}_P(C_{\mathrm{sew}}, C \otimes C^\dagger) \le \mathcal{D}_F(C_{\mathrm{sew}}, C \otimes C^\dagger). \tag{171}$$

Leveraging the facts that the Frobenius norm is unitary invariant and satisfies the triangle inequality, we perform the following hybrid argument:

$$\mathcal{D}_{\mathrm{avg}}(\mathcal{C}_{\mathrm{sew}}, \mathcal{C} \otimes \mathcal{C}^{\dagger}) \leq \mathcal{D}_F(\mathcal{C}_{\mathrm{sew}}, \mathcal{C} \otimes \mathcal{C}^{\dagger}) \tag{172}$$

$$= \mathcal{D}_F\left(S\prod_{i=1}^{n} W_i, S\prod_{i=1}^{n} V_i\right) \tag{173}$$

$$= \mathcal{D}_F\left(\prod_{i=1}^{n} W_i, \prod_{i=1}^{n} V_i\right) \tag{174}$$

$$\leq \mathcal{D}_F\left(\prod_{i=1}^{n} W_i, \prod_{i=1}^{n-1} W_i V_n\right) + \mathcal{D}_F\left(\prod_{i=1}^{n-1} W_i V_n, \prod_{i=1}^{n} V_i\right) \tag{175}$$

$$\leq \mathcal{D}_F\left(W_n, V_n\right) + \mathcal{D}_F\left(\prod_{i=1}^{n-1} W_i, \prod_{i=1}^{n-1} V_i\right) \tag{176}$$

$$\vdots \tag{177}$$

$$\leq \sum_{i=1}^{n} \mathcal{D}_F\left(W_i, V_i\right). \tag{178}$$

Upper-bounding each $\mathcal{D}_F\left(W_i, V_i\right)$ term as

$$\mathcal{D}_F\left(W_i, V_i\right) \leq \mathcal{D}_F\left(W_i, \widetilde{W}_i\right) + \mathcal{D}_F\left(\widetilde{W}_i, V_i\right), \tag{179}$$

we are now interested in the values of $\mathcal{D}_F\left(W_i, \widetilde{W}_i\right)$ and $\mathcal{D}_F\left(\widetilde{W}_i, V_i\right)$. However, leveraging the definition of Proj_U, $\mathcal{D}_F\left(W_i, \widetilde{W}_i\right)$ can be expressed in terms of $\mathcal{D}_F\left(\widetilde{W}_i, V_i\right)$, as

$$\mathcal{D}_F\left(\widetilde{W}_i, W_i\right) = \frac{1}{2^{2n}}\left\|\widetilde{W}_i - W_i\right\|_F^2 = \frac{1}{2^{2n}}\min_{U \in \mathcal{U}(2^{2n})}\left\|\widetilde{W}_i - U\right\|_F^2 \leq \frac{1}{2^{2n}}\left\|\widetilde{W}_i - V_i\right\|_F^2 = \mathcal{D}_F\left(\widetilde{W}_i, V_i\right). \tag{180}$$

Therefore, Equation (179) simplifies to

$$\mathcal{D}_F\left(W_i, V_i\right) \leq 2 \cdot \mathcal{D}_F\left(\widetilde{W}_i, V_i\right). \tag{181}$$

Furthermore, we can upper-bound $\mathcal{D}_F\left(\widetilde{W}_i, V_i\right)$, as

$$\mathcal{D}_F\left(\widetilde{W}_i, V_i\right) = \frac{1}{2^{2n}}\left\|\widetilde{W}_i - V_i\right\|_F^2 \tag{182}$$

$$\leq \frac{1}{2^{2n}}\left\|\left(\frac{1}{2}I \otimes I + \frac{1}{2}\sum_{P \in \{X,Y,Z\}} \widetilde{O}_{P_i}^{(\ell)} \otimes P_i\right) - \left(\frac{1}{2}I \otimes I + \frac{1}{2}\sum_{P \in \{X,Y,Z\}} O_{P_i} \otimes P_i\right)\right\|_F^2 \tag{183}$$

$$\leq \frac{1}{2^{2n}}\left\|\frac{1}{2}\sum_{P \in \{X,Y,Z\}}\left(\widetilde{O}_{P_i}^{(\ell)} - O_{P_i}\right) \otimes P_i\right\|_F^2 \tag{184}$$

$$\leq \frac{1}{2^{2n}} \cdot \frac{1}{4} \operatorname{Tr} \left(\sum_{P \in \{X,Y,Z\}} \left(\widetilde{O}_{P_i}^{(\ell)} - O_{P_i} \right)^{\dagger} \otimes P_i \cdot \sum_{Q \in \{X,Y,Z\}} \left(\widetilde{O}_{P_i}^{(\ell)} - O_{Q_i} \right) \otimes Q_i \right)$$
$$(185)$$

$$= \frac{1}{2^{2n}} \cdot \frac{1}{4} \sum_{P,Q \in \{X,Y,Z\}} \operatorname{Tr} \left(\left(\widetilde{O}_{P_i}^{(\ell)} - O_{P_i} \right)^{\dagger} \left(\widetilde{O}_{P_i}^{(\ell)} - O_{Q_i} \right) \otimes P_i Q_i \right) \tag{186}$$

$$= \frac{1}{2^{2n}} \cdot \frac{1}{4} \sum_{P,Q \in \{X,Y,Z\}} \operatorname{Tr} \left(\left(\widetilde{O}_{P_i}^{(\ell)} - O_{P_i} \right)^{\dagger} \left(\widetilde{O}_{P_i}^{(\ell)} - O_{Q_i} \right) \right) \operatorname{Tr} \left(P_i Q_i \right) \tag{187}$$

$$= \frac{1}{2^{2n}} \cdot \frac{1}{4} \sum_{P,Q \in \{X,Y,Z\}} \operatorname{Tr} \left(\left(\widetilde{O}_{P_i}^{(\ell)} - O_{P_i} \right)^{\dagger} \left(\widetilde{O}_{P_i}^{(\ell)} - O_{Q_i} \right) \right) \cdot 2^n \delta(P_i = Q_i) \tag{188}$$

$$= \frac{1}{2^n} \cdot \frac{1}{4} \sum_{P \in \{X,Y,Z\}} \operatorname{Tr} \left(\left(\widetilde{O}_{P_i}^{(\ell)} - O_{P_i} \right)^{\dagger} \left(\widetilde{O}_{P_i}^{(\ell)} - O_{P_i} \right) \right) \tag{189}$$

$$= \frac{1}{4} \sum_{P \in \{X,Y,Z\}} \frac{1}{2^n} \left\| \widetilde{O}_{P_i}^{(\ell)} - O_{P_i} \right\|_F^2 \tag{190}$$

$$= \frac{1}{4} \sum_{P \in \{X,Y,Z\}} \epsilon_{P_i}. \tag{191}$$

Therefore, combining our bounds from Equation (178), Equation (181), and Equation (191), we obtain the desired upper-bound:

$$\mathcal{D}_{\text{avg}}(C_{\text{sew}}, C \otimes C^{\dagger}) \leq \sum_{i=1}^{n} \mathcal{D}_F \left(W_i, V_i \right) \leq \sum_{i=1}^{n} 2 \cdot \mathcal{D}_F \left(\widetilde{W}_i, V_i \right) \leq \frac{1}{2} \sum_{i=1}^{n} \sum_{P \in \{X,Y,Z\}} \epsilon_{P_i}. \tag{192}$$

$\blacksquare$

With these results establishing efficient learning and sewing of QAC^0 Heisenberg-evolved single-qubit Pauli observables, we can now prove the main result of the section, Theorem 26.

Proof [Proof of Theorem 26] Leveraging Algorithm 1 and Theorem 27, we can learn the set of observables $\{\widetilde{O}_{P_i}^{(\ell)}\}_{i,P}$ such that

$$\mathcal{D}_F \left(O_{P_i}, \widetilde{O}_{P_i}^{(\ell)} \right) \leq \epsilon_{P_i} \leq \frac{2}{n^b} + \frac{9d^2}{c^2 \cdot n^{c-2}} \tag{193}$$

for all $3n$ single-qubit Paulis P_i, with quasi-polynomial sample and computational complexity. Leveraging the result of Theorem 29, we can then "sew" these learned observables into the unitary

$$C_{\text{sew}} := \text{SWAP}^{\otimes n} \prod_{i=1}^{n} \left[\text{Proj}_U \left(\frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} \widetilde{O}_{P_i}^{(\ell)} \otimes P_i \right) \right], \tag{194}$$

which, for constants $b \geq 2$ and $c \geq 3d$, satisfies

$$\mathcal{D}_{\text{avg}}(\mathcal{C}_{\text{sew}}, \mathcal{C} \otimes \mathcal{C}^{\dagger}) \leq \frac{1}{2} \sum_{i=1}^{n} \sum_{P \in \{X,Y,Z\}} \epsilon_{P_i} \leq \frac{3n}{2} \cdot \arg\max_{P_i} \epsilon_{P_i} \leq \frac{3}{n^{b-1}} + \frac{9d^2}{c^2 \cdot n^{c-3}} = \frac{1}{\text{poly}(n)}, \tag{195}$$

meaning that C_{sew} is $1/\text{poly}(n)$-close to $C \otimes C^\dagger$ in average-case distance.

All that remains is to verify that the computational complexity of the sewing procedure of Equation (194) is in fact quasi-polynomial. Note that the construction of C_{sew} requires computing

$$\widetilde{U}_i = \text{Proj}_U \left(\frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} \widetilde{O}_{P_i}^{(\ell)} \otimes P_i \right) \tag{196}$$

n times. As described in Fact 5, computing Proj_U reduces to computing the singular value decomposition of the matrix. In general, computing the SVD of a $2n$-qubit matrix has exponential complexity $\mathcal{O}(2^{6n})$. However, in Algorithm 1 we specifically imposed that the learned observable $\widetilde{O}_{P_i}^{(\ell)}$ only have support on $\ell = \mathcal{O}(\log^d n)$ qubits. Since P_i only has support on 1 qubit, the total support of the matrix $\frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} \widetilde{O}_{P_i}^{(\ell)} \otimes P_i$ is $\ell + 1$ qubits. Therefore, we only need to compute the SVD of the sub-matrix corresponding to the non-trivial support, which has quasi-polynomial computational complexity $\mathcal{O}(2^{\text{poly} \log n})$. Therefore, the total sewing procedure involves performing n of these SVDs, which is still quasi-polynomial complexity. ∎

D.3. Learning QAC^0 with Optimized Depth

Now that we have learned a unitary C_{sew} which is close in average-case distance to $C \otimes C^\dagger$, we will explore circuit-synthesis procedures that implement unitaries close to C_{sew}. In other words, we will look into *proper learning* of the QAC^0 circuit. We will show that, while a naive compilation procedure would produce a worst-case quasi-polynomial-depth circuit, we can generate in quasi-polynomial time an explicit (poly-logarithmic depth) QAC circuit C_{sew}^* that implements a unitary $\frac{1}{\text{poly}(n)}$-close to C_{sew}. In particular, we will prove the following theorem.

Theorem 30 *Given a QAC^0 circuit C, there exists a quasi-polynomial time algorithm to learn a QAC circuit implementing unitary C_{sew}^* such that*

$$\mathcal{D}_F \left(C_{sew}^*, C \otimes C^\dagger \right) \leq \frac{1}{\text{poly}(n)}. \tag{197}$$

Throughout this section, we will use the following notation:

$$V_i = \frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} O_{P_i} \otimes P_i, \tag{198}$$

$$V_i^* = \frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} O_{P_i}^* \otimes P_i, \tag{199}$$

$$\widetilde{W}_i = \frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} \widetilde{O}_{P_i}^{(\ell)} \otimes P_i, \tag{200}$$

$$W_i = \text{Proj}_U \left(\widetilde{W}_i \right). \tag{201}$$

Furthermore, by the sewing procedure of Huang et al. (2024), if O_{P_i} corresponds to the QAC^0 circuit C and L_i is the light-cone of C with respect to measurement qubit i, then

$$V_i = \frac{1}{2} I \otimes I + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} O_{P_i} \otimes P_i \tag{202}$$

$$= \frac{1}{2} \sum_{P \in \{I,X,Y,Z\}} C^\dagger P_i C \otimes P_i \tag{203}$$

$$= \frac{1}{2} \sum_{P \in \{I,X,Y,Z\}} L_i^\dagger P_i L_i \otimes P_i \tag{204}$$

$$= (L_i^\dagger \otimes I) \left(\frac{1}{2} \sum_{P \in \{I,X,Y,Z\}} P_i \otimes P_i \right) (L_i \otimes I) \tag{205}$$

$$= L_i^\dagger S_i L_i, \tag{206}$$

where S_i denotes the SWAP operation between the i^{i} and $(n+i)^{\text{th}}$ qubits. Similarly, if $O_{P_i}^*$ corresponds to the QAC^0 circuit $\widetilde{C}$ (with large CZ gates removed) and $\widetilde{L}_i$ is the light-cone of $\widetilde{C}$ with respect to measurement qubit i, then

$$V_i^* = \widetilde{L}_i^\dagger S_i \widetilde{L}_i. \tag{207}$$

D.3.1. NAIVE IMPLEMENTATION

We will begin by evaluating the worst-case circuit depth of a naive compilation procedure for our learned unitary. Our naive implementation will leverage the following standard fact about the complexity of unitary synthesis.

Fact 6 (Huang et al. (2024) Fact 4) *Given a unitary U, which acts on k qubits, there is an algorithm that outputs a circuit (acting on k qubits) that consists of at most 4^k two-qubit gates, which exactly implements the unitary U, in time $2^{\mathcal{O}(k)}$.*

Because any two-qubit gate can be generated by a constant number of single-qubit gates and CZ gates, we can immediately obtain the following fact from the above.

Fact 7 (Adapted from Huang et al. (2024) Fact 4) *Given a unitary U, which acts on k qubits, there is an algorithm that outputs a circuit (acting on k qubits) that consists of at most 4^k single-qubit gates or many-qubit CZ gates, which exactly implements the unitary U, in time $2^{\mathcal{O}(k)}$.*

For each qubit measurement index i, we will consider a naive circuit-synthesis procedure that simply compiles the individual projected unitaries W_i into a circuit with gates acting on at most 2 qubits and then sews these compiled circuits into C_{sew}, as specified in Equation (194), via some arbitrary ordering. Since the support of each unitary W_i is $\ell = \text{poly} \log(n)$, by Fact 6, it would be compiled into a circuit acting on $\ell = \text{poly} \log(n)$ qubits consisting of up to $4^{\text{poly} \log(n)}$ gates.

Stitching together all $3n$ of these circuits in an arbitrary order results in a circuit with quasi-polynomial depth. This is substantially deeper than the constant depth of the true QAC^0 circuit that we aimed to learn. Thus, we will now show how improved compilation and ordering of the unitaries in the sewing procedure can reduce this depth down to poly-logarithmic, while still only requiring quasi-polynomial computational complexity.

D.3.2. IMPROVED COMPILATION

We will now describe an improved compilation procedure, that reduces the circuit-synthesis depth for unitaries W_i from $\mathcal{O}(4^{\mathrm{poly}\log n})$ to constant-depth d.

Theorem 31 *The learned unitary W_i can be compiled into a QAC^0 circuit, governed by unitary C_i^*, that is supported on $\mathcal{O}(\log^d n)$ qubits and such that*

$$\mathcal{D}_F\left(W_i, C_i^*\right) \leq \frac{1}{\mathrm{poly}(n)}. \tag{208}$$

The computational complexity of this compilation procedure is quasi-polynomial.

In order to prove this result, we will first show that the learned unitaries W_i are close to the $\mathcal{O}(\log^d n)$-support QAC^0 circuits governed by unitary $V_i^* = \widetilde{L}_i^\dagger S_i \widetilde{L}_i$.

Lemma 32 *Let O_{P_i} be a Heisenberg-evolved observable of the QAC^0 circuit C and $\widetilde{O}_{P_i}$ be a Heisenberg-evolved observable of some other circuit $\widetilde{C}$ such that*

$$\mathcal{D}_F\left(O_{P_i}, \widetilde{O}_{P_i}\right) \leq \epsilon, \quad \forall i \in [n], \ P \in \{X, Y, Z\}. \tag{209}$$

Then,

$$\mathcal{D}_F\left(\frac{1}{2}\sum_{P \in \{X,Y,Z\}} O_{P_i} \otimes P_i, \ \frac{1}{2}\sum_{P \in \{X,Y,Z\}} \widetilde{O}_{P_i} \otimes P_i\right) \leq \epsilon \tag{210}$$

Proof Leveraging Cauchy-Schwarz and the fact that $\|A \otimes B\|_F^2 = \|A\|_F^2 \|B\|_F^2$,

$$\mathcal{D}_F\left(\frac{1}{2}\sum_{P \in \{I,X,Y,Z\}} O_{P_i} \otimes P_i, \ \frac{1}{2}\sum_{P \in \{I,X,Y,Z\}} \widetilde{O}_{P_i} \otimes P_i\right) \tag{211}$$

$$\leq \frac{1}{2^n}\left\|\frac{1}{2}\sum_{P \in \{I,X,Y,Z\}} (O_{P_i} - \widetilde{O}_{P_i}) \otimes P_i\right\|_F^2 \tag{212}$$

$$\leq \frac{1}{4}\sum_{P \in \{I,X,Y,Z\}} \frac{1}{2^n}\left\|(O_{P_i} - \widetilde{O}_{P_i}) \otimes P_i\right\|_F^2 \tag{213}$$

$$\leq \frac{1}{4}\sum_{P \in \{I,X,Y,Z\}} \frac{1}{2^n}\left\|(O_{P_i} - \widetilde{O}_{P_i})\right\|_F^2 \|P_i\|_F^2 \tag{214}$$

$$= \frac{1}{4}\sum_{P \in \{X,Y,Z\}} \mathcal{D}_F\left(O_{P_i}, \widetilde{O}_{P_i}\right) \tag{215}$$

$$\leq \frac{3}{4} \cdot \epsilon. \tag{216}$$

This concludes the proof. ∎

Corollary 33 *The learned unitaries W_i are $1/\mathrm{poly}(n)$-close to the unitaries $V_i^* = \widetilde{L}_i^\dagger S_i \widetilde{L}_i$,*

$$\mathcal{D}_F\left(W_i, V_i^*\right) \leq \frac{1}{\mathrm{poly}(n)}. \tag{217}$$

Proof By Theorem 27,

$$\mathcal{D}_F\left(\widetilde{O}_{P_i}^{(\ell)}, O_{P_i}\right) \leq \epsilon_{P_i} = \frac{1}{\mathrm{poly}(n)}, \tag{218}$$

and, by Theorem 22,

$$\mathcal{D}_F\left(\widetilde{O}_{P_i}^*, O_{P_i}\right) \leq \epsilon^* = \frac{1}{\mathrm{poly}(n)}. \tag{219}$$

By triangle inequality and Theorem 32, we have that

$$\mathcal{D}_F\left(\widetilde{W}_i, V_i^*\right) \tag{220}$$

$$\leq \mathcal{D}_F\left(\widetilde{W}_i, V_i\right) + \mathcal{D}_F\left(V_i, V_i^*\right) \tag{221}$$

$$\leq \frac{1}{2^n} \left\| \frac{1}{2} \sum_{P \in \{X,Y,Z\}} \left(\widetilde{O}_{P_i}^{(\ell)} \otimes P_i - O_{P_i} \otimes P_i\right) \right\|_F^2 + \frac{1}{2^n} \left\| \frac{1}{2} \sum_{P \in \{X,Y,Z\}} \left(O_{P_i} \otimes P_i - O_{P_i}^* \otimes P_i\right) \right\|_F^2 \tag{222}$$

$$\leq \epsilon_{P_i} + \epsilon^* \leq \frac{1}{\mathrm{poly}(n)}. \tag{223}$$

Furthemore, since Proj_U is the projection onto the unitary minimizing the Frobenius norm, for $W_i = \mathrm{Proj}_U(\widetilde{W}_i)$, it must be true that

$$\mathcal{D}_F\left(\widetilde{W}_i, W_i\right) \leq \mathcal{D}_F\left(\widetilde{W}_i, V_i^*\right) \leq \frac{1}{\mathrm{poly}(n)}. \tag{224}$$

Therefore, by triangle inequality we obtain the desired result,

$$\mathcal{D}_F\left(W_i, V_i^*\right) \leq \mathcal{D}_F\left(\widetilde{W}_i, W_i\right) + \mathcal{D}_F\left(\widetilde{W}_i, V_i^*\right) \leq \frac{1}{\mathrm{poly}(n)}. \tag{225}$$

■

This implies that there must exist a circuit of the form $V_i^* = \widetilde{L}_i^\dagger S_i \widetilde{L}_i$ that is $1/\mathrm{poly}(n)$-close to each of the learned unitaries W_i. Note that since the swap gate S_i can be implemented in QAC^0, the circuit V_i^* is contained QAC^0. This implies that, to find a circuit $1/\mathrm{poly}(n)$-close to W_i, rather than searching over all possible QAC^0 architectures, we can restrict our search to QAC^0 circuits of the form of V_i^*, or more precisely QAC^0 circuits in the lightcone $\widetilde{L}_i$. Thus, we will now show how to efficiently construct an ϵ-net over QAC^0 circuits of depth-d, with $\mathcal{O}(\log^d n)$ support, that is guaranteed to contain a QAC^0 circuit $1/\mathrm{poly}(n)$-close to the learned unitary W_i.

Lemma 34 *Let $\mathcal{C}^*$ be the class of all depth-d QAC^0 circuits with CZ gates acting on at most $\kappa = \mathcal{O}(\log n)$ qubits and supported on $\mathcal{O}(\log^d n)$ qubits. $\mathcal{C}^*$ has a $1/\mathrm{poly}(n)$-net, denoted $\mathcal{N}_{1/\mathrm{poly}(n)}(\mathcal{C}^*)$, of quasi-polynomial size that can be constructed in quasi-polynomial time.*

Proof Recall that the general structure of a QAC^0 circuit is alternating layers of CZ gates and layers of arbitrary single-qubit gates. Therefore, we can construct an ϵ-net for $\mathcal{C}^*$ by first enumerating all possible architectures (i.e. placements of CZ gates) and then, for each architecture, creating an ϵ'-net for each possible SU(2) gate.

We will begin by enumerating all possible QAC^0 architectures, i.e. configurations of CZ gates of width at most $\kappa = \mathcal{O}(\log n)$ acting on $\ell = \mathcal{O}(\log^d n)$ qubits across d layers. We begin thinking of a given layer of the architecture as a graph with ℓ vertices, corresponding to each of the qubits in the support. Within this graph framework, if a set of vertices are contained in a k-clique this means the corresponding qubits are acted upon by a CZ_k gate. Therefore, to enumerate the total number of distinct CZ configurations in the layer, we simply need to enumerate the number of distinct graphs comprised of cliques of size at most κ. This is trivially upper-bounded by the number of distinct subgraphs of the complete graph on ℓ nodes, which is quasi-polynomial, i.e.

$$2^{\binom{\ell}{2}} \leq 2^{\ell^2} = 2^{\text{poly}\log(n)}. \tag{226}$$

Since the circuits are depth-d, where d is constant, the total number of CZ configurations across the whole circuit is $d \cdot 2^{\text{poly}\log(n)}$, which is also quasi-polynomial.

As previously mentioned, between the layers of CZ gates are layers of arbitrary single-qubit gates. In total, there are at most $d \cdot \ell \leq \mathcal{O}(\log^d n)$ single-qubit gates. By a standard hybrid argument, it can be shown that the error propagation of the SU(2) ϵ'-net is additive both within and across layers of the single qubit gates. Therefore, to achieve an overall $1/\text{poly}(n)$-net,

$$d \cdot \ell \cdot \epsilon' \leq \text{poly}\log(n) \cdot \epsilon' \leq \frac{1}{\text{poly}(n)}, \tag{227}$$

which implies that $\epsilon' \leq 1/\text{poly}(n)$. An ϵ'-net for $SU(2)$ can be constructed with $\left(\frac{c_0}{\epsilon'}\right)^{c_1}$ elements, which for $\epsilon' = 1/\text{poly}(n)$ is polynomial size.

Therefore, the total size of the net is the size of the $SU(2)$ ϵ'-net times the total number of single-qubit gates times the total number of architectures, which is bounded as

$$\text{poly}(n) \cdot d \cdot \text{poly}\log(n) \cdot 2^{\text{poly}\log(n)} \leq \mathcal{O}(2^{\text{poly}\log(n)}), \tag{228}$$

and therefore quasi-polynomial. ■

We will now show that we can efficiently find an element of the ϵ-net that is $1/\text{poly}(n)$-close to W_i. Combining this with the prior results of the section, we achieve a simple proof of Theorem 31.

Proof [Proof of Theorem 31] By using a brute-force search procedure, we can iterate through the quasi-polynomial elements of the ϵ-net described in Theorem 34 to find the element

$$L_i^* = \underset{L \in \mathcal{N}_{1/\text{poly}(n)}(\mathcal{C}^*)}{\arg\min} \mathcal{D}_F\left(W_i, L^\dagger S_i L\right). \tag{229}$$

Since the swap gate S_i can be implemented in constant-depth in QAC^0, we thus have a constant-depth circuit implementation of the unitary $C_i^* = (L_i^*)^\dagger S_i L_i^*$. By Theorem 33,

$$\mathcal{D}_F\left(W_i, C_i^*\right) \leq \mathcal{D}_F\left(W_i, \widetilde{L}_i^\dagger S_i \widetilde{L}_i\right) = \mathcal{D}_F\left(W_i, V_i^*\right) \leq \frac{1}{\text{poly}(n)}, \tag{230}$$

meaning that C_i^* is $1/\text{poly}(n)$-close to W_i, as desired. ■

D.3.3. IMPROVED ORDERING

Leveraging the improved compilation result, we will now propose an improved ordering for the sewing procedure. This will enable us to construct a QAC circuit which is $1/\text{poly}(n)$-close to C_{sew}, thereby proving Theorem 30.

To begin, we show how the constant-depth learned circuits for C_i^* can be sewed into a worst-case poly-logarithmic depth circuit. Note that our proof approach is similar to that of (Huang et al., 2024, Lemma 13). However, the circuit for each C_i^* has poly-logarithmic support (whereas those of Huang et al. (2023) have constant support), meaning we only achieve poly-logarithmic depth (instead of constant depth).

Lemma 35 *(Sewing into a poly-logarithmic depth circuit) Given $3n$ learned observables $\widetilde{O}_{P_i}^{(\ell)}$, such that for any qubit i, $\left|\bigcup_P \text{supp}\left(\widetilde{O}_{P_i}^{(\ell)}\right)\right| = \mathcal{O}(\text{poly}\log(n))$ and there are only $\text{poly}\log(n)$ qubits j such that*

$$\text{supp}\left(\widetilde{O}_{P_i}^{(\ell)}\right) \cap \text{supp}\left(\widetilde{O}_{P_j}^{(\ell)}\right) \neq \varnothing. \tag{231}$$

There exists a sewing ordering for C_{sew}, as defined in Equation (194), such that it can be implemented by a $\text{poly}\log(n)$-depth quantum circuit. The computational complexity for finding this sewing order is polynomial, i.e. $\mathcal{O}(n\log^d n)$.

Proof [Proof of Theorem 35] Defining $A(i) = \bigcup_P \text{supp}\left(\widetilde{O}_{P_i}^{(\ell)}\right)$, then $\text{supp}\left(C_i^*\right) \subseteq A(i) \cup \{n+i\}$, where C_i^* is the learned constant-depth circuit from Theorem 31.

Now, consider an n-node graph (where each node represents one of the n qubits), such that each pair (i, j) of nodes is connected by an edge if

$$A(i) \cap A(j) \neq \varnothing. \tag{232}$$

The graph only has $\mathcal{O}(n\log^d n)$ many edges and can be constructed as an adjacency list in time $\mathcal{O}(n\log^d n)$. Since the size of the support $A(i)$ is poly-logarithmic, the graph has poly-logarithmic degree. Thus, we can use a $\mathcal{O}(n\log^d n)$-time greedy graph coloring algorithm to color the graph using $\chi = \mathcal{O}(\log^d n)$ colors. For each node i, let $c(i)$ denote the color labeled from 1 to χ.

We can modify the arbitrary sewing order of the $3n$ observables $\widetilde{O}_{P_i}^{(\ell)}$ in Equation (194) to the ordering given by this greedy graph coloring, where we order from the smallest to the largest color. By the definition of graph coloring, for any pair (i, j) of qubits with the same color,

$$A(i) \cap A(j) = \varnothing. \tag{233}$$

Therefore, for each color c', we can implement the $2n$-qubit unitary

$$\prod_{i:c(i)=c'} C_i^* \tag{234}$$

via the constant-depth quantum circuits C_i^*. Since there are at most a poly-logarithmic number of colors, with the color-based ordering, C_{sew} will be poly-logarithmic depth in the worst-case. ∎

Combining this improved ordering result with the improved compilation result of the last section, we can now prove Theorem 30.

Proof [Proof of Theorem 30] For each qubit i, by Theorem 31 we can find a constant-depth QAC^0 circuit C_i^* such that

$$\mathcal{D}_F\left(W_i, C_i^*\right) \leq \frac{1}{\mathrm{poly}(n)}. \tag{235}$$

Performing this for all n measurement qubits requires quasi-polynomial computational complexity. By Theorem 35, we can find a sewing order for all the C_i^* circuits, in polynomial time, which sews them into the poly-logarithmic depth circuit C_{sew}^*.

To conclude, we will prove that C_{sew}^* is $1/\mathrm{poly}(n)$-close to $C \otimes C^\dagger$. By triangle inequality and Theorem 26

$$\mathcal{D}_F(C_{\mathrm{sew}}^*, C \otimes C^\dagger) \leq \mathcal{D}_F(C_{\mathrm{sew}}^*, C_{\mathrm{sew}}) + \mathcal{D}_F(C_{\mathrm{sew}}, C \otimes C^\dagger) \leq \mathcal{D}_F(C_{\mathrm{sew}}^*, C_{\mathrm{sew}}) + \frac{1}{\mathrm{poly}(n)} \tag{236}$$

Therefore, all that remains is to show that C_{sew}^* is $1/\mathrm{poly}(n)$-close to C_{sew}. This can be achieved by leveraging the bound of Theorem 31 and a simple hybrid argument,

$$\mathcal{D}_F(C_{\mathrm{sew}}^*, C_{\mathrm{sew}}) = \mathcal{D}_F\left(\prod_i C_i^*, \prod_{i=1}^n W_i\right) \tag{237}$$

$$\leq \mathcal{D}_F\left(\prod_{i=1}^n C_i^*, W_1 \prod_{i=2}^n C_i^*\right) + \mathcal{D}_F\left(W_1 \prod_{i=2}^n C_i^*, \prod_{i=1}^n W_i\right) \tag{238}$$

$$\leq \mathcal{D}_F\left(C_1^*, W_1\right) + \mathcal{D}_F\left(\prod_{i=2}^n C_i^*, \prod_{i=2}^n W_i\right) \tag{239}$$

$$\leq \frac{1}{\mathrm{poly}(n)} + \mathcal{D}_F\left(\prod_{i=2}^n C_i^*, \prod_{i=2}^n W_i\right) \tag{240}$$

$$\vdots \tag{241}$$

$$\leq n \cdot \frac{1}{\mathrm{poly}(n)} \leq \frac{1}{\mathrm{poly}(n)}. \tag{242}$$

This concludes the proof. ∎

Appendix E. Concentration and Learning of QAC^0 with Ancillas

Now we will show how things change for QAC^0 circuits with ancillas. In this section, let C be the unitary corresponding to an $(n + a)$-qubit QAC^0 circuit, of the same form as Equation (76), operating on n standard qubits and a ancilla qubits. Similar to Huang et al. (2024), we will only consider circuits where the ancillas are initialized to the $|0^a\rangle$ state and the computation is clean (meaning ancillas are reverted to the $|0^a\rangle$ state at the end of the computation). Note that since the computation is clean, the action of C on the $(n + a)$-qubit system is equivalent to the action of another unitary A on just the n-qubit system without ancillas, i.e.

$$C(I \otimes |0^a\rangle) = A \otimes |0^a\rangle. \tag{243}$$

We will define the Heisenberg-evolved Pauli observables of this system "without ancilla restriction" as

$$O_{P_i,n+a} = C(P_i \otimes I^a)C^\dagger \tag{244}$$

and "with ancilla restriction" as

$$O_{P_i,n} = (I \otimes \langle 0^a|) \cdot O_{P_i,n+a} \cdot (I \otimes |0^a\rangle) = (I \otimes \langle 0^a|)C(P_i \otimes I^a)C^\dagger(I \otimes |0^a\rangle) = AP_iA^\dagger. \tag{245}$$

E.1. Concentration of QAC^0 Heisenberg-Evolved Observables with Ancillas

We will now re-prove the concentration results of Appendix C for QAC^0 circuits with ancillas. To begin, generalizing Theorem 22 to QAC^0 circuits with ancillas, we bound the distance between the true observable $O_{P_i,n}$ and the observable $O^*_{P_i,n}$ (with large CZ gates removed).

Lemma 36 *Let C be an $(n+a)$-qubit QAC^0 circuit performing clean computation with respect to a ancillas. Let $O_{P_i,n+a} = C(P_i \otimes I^a)C^\dagger$ be the Heisenberg-evolved P_i observable without ancilla restriction and $O_{P_i,n} = (I \otimes \langle 0^a|) \cdot O_{P_i,n+a} \cdot (I \otimes |0^a\rangle)$ be the same observable with ancilla restriction. Define $O^*_{P_i,n+a} = \widetilde{C}(P_i \otimes I^a)\widetilde{C}^\dagger$ to be the Heisenberg-evolved P_i observable corresponding to the QAC^0 circuit $\widetilde{C}$, which is simply C with all m CZ_k's of size $k > \kappa$ removed. Let $O^*_{P_i,n} = (I \otimes \langle 0^a|) \cdot O^*_{P_i,n+a} \cdot (I \otimes |0^a\rangle)$ be the ancilla-restricted version of this observable. The average-case distance between observables corresponding to circuits C and $\widetilde{C}$ is upper-bounded by*

$$\mathcal{D}_F\left(O_{P_i,n}, O^*_{P_i,n}\right) \leq \mathcal{D}_F(O_{P_i,n+a}, O^*_{P_i,n+a}) \leq 2^a \cdot \frac{9m^2}{2^\kappa}. \tag{246}$$

Proof We derive the following bound, using the facts that $\|AB\|_F \leq \|A\| \cdot \|B\|_F$ and $\|I \otimes |0^a\rangle\| = 1$:

$$\|O_{P_i,n} - O^*_{P_i,n}\|_F = \|(I \otimes \langle 0^a|)\left(C(P_i \otimes I^a)C^\dagger - \widetilde{C}(P_i \otimes I^a)\widetilde{C}^\dagger\right)(I \otimes |0^a\rangle)\|_F \tag{247}$$

$$\leq \|I \otimes \langle 0^a|\| \cdot \|\left(C(P_i \otimes I^a)C^\dagger - \widetilde{C}(P_i \otimes I^a)\widetilde{C}^\dagger\right)(I \otimes |0^a\rangle)\|_F \tag{248}$$

$$\leq \|I \otimes \langle 0^a|\| \cdot \|C(P_i \otimes I^a)C^\dagger - \widetilde{C}(P_i \otimes I^a)\widetilde{C}^\dagger\|_F \cdot \|I \otimes |0^a\rangle\| \tag{249}$$

$$= 1 \cdot \|C(P_i \otimes I^a)C^\dagger - \widetilde{C}(P_i \otimes I^a)\widetilde{C}^\dagger\|_F \cdot 1 \tag{250}$$

$$= \|C(P_i \otimes I^a)C^\dagger - \widetilde{C}(P_i \otimes I^a)\widetilde{C}^\dagger\|_F \tag{251}$$

$$= \|O_{P_i,n+a} - O^*_{P_i,n+a}\|_F. \tag{252}$$

Therefore, if C has m CZ gates of width $\geq \kappa$ to be removed in $\widetilde{C}$, then following from Equation (107),

$$\mathcal{D}_F(O_{P_i,n}, O^*_{P_i,n}) \leq \mathcal{D}_F(O_{P_i,n+a}, O^*_{P_i,n+a}) \tag{253}$$

$$\leq \frac{1}{2^n}\left(\sum_{j=1}^m 3 \cdot 2^{(n+a-k_j)/2}\right)^2 \tag{254}$$

$$\leq \frac{1}{2^n}\left(3m \cdot \max_j 2^{(n+a-k_j)/2}\right)^2 \tag{255}$$

$$= \frac{9m^2}{2^{\kappa-a}}. \tag{256}$$

∎

We will now demonstrate that the Pauli weight of the observable with ancilla restriction $O_{P_i,n}$ can be upper-bounded by that of the observable without ancilla restriction $O_{P_i,n+a}$.

Lemma 37 *Let $\mathcal{S} \subseteq \mathcal{P}^n$ be a subset of the set of n-qubit Paulis, then*

$$\mathbf{W}^{\in\mathcal{S}}[O_{P_i,n}] \leq 2^a \cdot \mathbf{W}^{\in\mathcal{S}}[O_{P_i,n+a}]. \tag{257}$$

Proof [Proof of Theorem 37] We will begin by assuming that $O_{P_i,n+a}$ has the following Pauli decomposition,

$$O_{P_i,n+a} = \sum_{Q\in\mathcal{P}^n, R\in\mathcal{P}^a} \widehat{O_{P_i,n+a}}(Q\otimes R) \cdot Q\otimes R \tag{258}$$

We can relate the Pauli coefficients of $O_{P_i,n+a}$ to those of $O_{P_i,n}$, as

$$\widehat{O_{P_i,n}}(S) = \frac{1}{2^n} \cdot \mathrm{Tr}\left((I\otimes\langle 0^a|)\cdot O_{P_i,n+a}\cdot(I\otimes|0^a\rangle)\cdot S\right) \tag{259}$$

$$= \frac{1}{2^n}\cdot\mathrm{Tr}\left((I\otimes\langle 0^a|)\left(\sum_{Q\in\mathcal{P}^n,R\in\mathcal{P}^a}\widehat{O_{P_i,n+a}}(Q\otimes R)\cdot Q\otimes R\right)(I\otimes|0^a\rangle)\cdot(S\otimes I^a)\right) \tag{260}$$

$$= \frac{1}{2^n}\cdot\sum_{Q\in\mathcal{P}^n,R\in\mathcal{P}^a}\widehat{O_{P_i,n+a}}(Q\otimes R)\cdot\mathrm{Tr}\left(QS\otimes\langle 0^a|R|0^a\rangle\right) \tag{261}$$

$$= \frac{1}{2^n}\cdot\sum_{Q\in\mathcal{P}^n,R\in\mathcal{P}^a}\widehat{O_{P_i,n+a}}(Q\otimes R)\cdot\mathrm{Tr}(QS)\cdot\langle 0^a|R|0^a\rangle \tag{262}$$

$$= \frac{1}{2^n}\cdot\sum_{Q\in\mathcal{P}^n,R\in\mathcal{P}^a}\widehat{O_{P_i,n+a}}(Q\otimes R)\cdot 2^n\cdot\delta\{Q=S\}\cdot\delta\{R\in\{I,Z\}^{\otimes a}\} \tag{263}$$

$$= \sum_{R\in\{I,Z\}^{\otimes a}}\widehat{O_{P_i,n+a}}(S\otimes R). \tag{264}$$

Leveraging this result and the Cauchy-Schwarz inequality, we can prove the desired result as

$$\mathbf{W}^{\in\mathcal{S}}[O_{P_i,n}] = \sum_{Q:Q\in\mathcal{S}}\left|\widehat{O_{P_i,n}}(Q)\right|^2 \tag{265}$$

$$= \sum_{Q:Q\in\mathcal{S}}\left|\sum_{R\in\{I,Z\}^{\otimes a}}\widehat{O_{P_i,n+a}}(Q\otimes R)\right|^2 \tag{266}$$

$$\leq 2^a\cdot\sum_{Q:Q\in\mathcal{S}}\sum_{R\in\{I,Z\}^{\otimes a}}\left|\widehat{O_{P_i,n+a}}(Q\otimes R)\right|^2 \tag{267}$$

47

$$\leq 2^a \cdot \mathbf{W}^{\in \mathcal{S}}[O_{P_i,n+a}]. \tag{268}$$

$\blacksquare$

Note that this result is similar to that of (Nadimpalli et al., 2024, Proposition 27), but we prove the bound for the weight over arbitrary subsets of Paulis $\mathcal{S} \in \mathcal{P}^n$, as opposed to just the weight above a certain degree. This enables us to use the result to achieve both the low-degree and low-support concentration bounds as simple corollaries.

Corollary 38 (Low-Degree Concentration with Ancillas) *Suppose C is a depth-d, size-s QAC^0 circuit performing clean computation on $n + a$ qubits, where a is the number of ancilla. Let $O_{P_i,n} = (I \otimes \langle 0^a|)C(P_i \otimes I^a)C^\dagger(I \otimes |0^a\rangle)$ be a Heisenberg-evolved single-qubit Pauli observable with ancilla restriction. Then for every degree $k \in [n]$,*

$$\mathbf{W}^{>k}[O_{P_i,n}] \leq \mathcal{O}\left(s^2 2^{-k^{1/d}}\right) \cdot 2^a. \tag{269}$$

Proof [Proof of Theorem 38] Let $\mathcal{S}$ be the set of Paulis of degree $> k$, i.e. $\mathcal{S} = \{P \in \mathcal{P}^n : |P| > k\}$. By Theorem 37,

$$\mathbf{W}^{>k}[O_{P_i,n}] \leq 2^a \cdot \mathbf{W}^{>k}[O_{P_i,n+a}]. \tag{270}$$

Since $O_{P_i,n+a}$ is an observable defined over $n+a$ qubits without ancilla restriction, by Theorem 20,

$$\mathbf{W}^{>k}[O_{P_i,n+a}] \leq \mathcal{O}\left(s^2 2^{-k^{1/d}}\right). \tag{271}$$

Plugging this into Equation (270), we obtain the desired result. $\blacksquare$

Corollary 39 (Low-Support Concentration with Ancillas) *For $\mathcal{S}^* = \mathrm{supp}(O^*_{P_i,n})$, the weight of $O_{P_i,n}$ outside the support of $O^*_{P_i,n}$ is upper-bounded as*

$$\mathbf{W}^{\notin \mathcal{S}^*}[O_{P_i,n}] \leq \mathcal{D}_F\left(O_{P_i,n}, O^*_{P_i,n}\right) \leq \mathcal{D}_F\left(O_{P_i,n+a}, O^*_{P_i,n+a}\right) \leq 2^a \cdot \frac{9m^2}{2^\kappa} \tag{272}$$

Proof [Proof of Theorem 39] Note that the weight expression satisfies the triangle inequality. Thus, the weight of $O_{P_i,n}$ can be decomposed with respect to $O^*_{P_i,n}$ as follows.

$$\mathbf{W}^{\notin \mathcal{S}^*}[O_{P_i,n}] = \mathbf{W}^{\notin \mathcal{S}^*}\left[O^*_{P_i,n} + (O_{P_i,n} - O^*_{P_i,n})\right] \tag{273}$$

$$\leq \left(\mathbf{W}^{\notin \mathcal{S}^*}\left[O^*_{P_i,n}\right]^{1/2} + \mathbf{W}^{\notin \mathcal{S}^*}\left[O_{P_i,n} - O^*_{P_i,n}\right]^{1/2}\right)^2 \tag{274}$$

Since $\mathcal{S}^* = \mathrm{supp}(O^*_{P_i,n})$, by Theorem 24,

$$\mathbf{W}^{\notin \mathcal{S}^*}\left[O^*_{P_i,n}\right] = 0. \tag{275}$$

Furthermore, by Theorem 36

$$\mathbf{W}^{\notin \mathcal{S}^*}\left[O_{P_i,n} - O^*_{P_i,n}\right] \leq \mathcal{D}_F\left(O_{P_i,n}, O^*_{P_i,n}\right) \leq \mathcal{D}_F\left(O_{P_i,n+a}, O^*_{P_i,n+a}\right). \tag{276}$$

Plugging these results into Equation (274) and leveraging our upper-bound on the distance from Theorem 36 obtains the desired expression. $\blacksquare$

E.2. Learning QAC^0 Heisenberg-Evolved Observables with Ancillas

In order to learn the unitary of an $(n + a)$-qubit QAC^0 circuit with ancillas, it turns out that we can use the procedure of Appendix D with some slight modifications.

Proposition 40 (Learning QAC^0 with Logarithmic Ancillas) *Suppose we are given an $(n + a)$-qubit depth-d QAC^0 circuit governed by unitary C, performing clean computation*

$$C(I \otimes |0^a\rangle) = A \otimes |0^a\rangle \tag{277}$$

on a logarithmic number of ancilla qubits, i.e. $a = \mathcal{O}(\log n)$. For error parameter $\varepsilon = 1/\mathrm{poly}(n)$ and failure probability $\delta \in (0, 1)$, we can learn a $2n$-qubit unitary A_{sew} which is ε-close to the unitary $A \otimes A^\dagger$, i.e.

$$\mathcal{D}_{avg}(A_{sew}, A \otimes A^\dagger) \leq \varepsilon, \tag{278}$$

with high probability $1 - \delta$, as well as quasi-polynomial sample and computational complexity.

Proof [Proof of Theorem 40] As expressed in Equation (277), since the circuit C performs clean computation, the ancillas return to their starting state by end of the computation. Therefore, we can measure out the ancillas at the end of the computation, to obtain the unitary A which performs the same computation as C, but only acts on the n computation qubits. Thus, in the learning procedure, we will be interested in learning and sewing the n-qubit Heisenberg-evolved single-qubit Paulis *with* ancilla restriction, i.e. $O_{P_i,n}$.

Since these observables with ancilla restriction only act on n qubits, we can straightforwardly apply Algorithm 1 for learning the Heisenberg-evolved single-qubit Pauli observables of the circuit. However, the learning guarantees in this case differ from those of the ancilla-free case. In particular, to show that the learned observable $\widetilde{O}_{P_i,n}^{(\ell)}$ is close to the desired observable with ancilla restriction $O_{P_i,n}$, we need to bound the distance between these operators. Plugging $\widetilde{O}_{P_i,n}^{(\ell)}$ and $O_{P_i,n}$ into Theorem 28, we get that their distance is bounded as

$$\mathcal{D}_F\left(\widetilde{O}_{P_i,n}^{(\ell)}, O_{P_i,n}\right) \leq 2 \cdot 4^\ell \cdot \eta^2 + \mathbf{W}^{\notin \mathcal{S}^*}[O_{P_i,n}] \leq 2 \cdot 4^\ell \cdot \eta^2 + 2^a \cdot \frac{9m^2}{2^\kappa}, \tag{279}$$

where η is the learning accuracy and the second inequality leveraged our low-support concentration result from Theorem 39. Notice that the key difference to the ancilla-free case is a 2^a factor amplifying the error in removing large CZ gates.

As in the ancilla-free case, we need to perform a balancing act to ensure that this learning error is small while ensuring that the algorithm has efficient sample and computational complexity. Since the sample and time complexity are directly related to the supports of the learned observables, to achieve quasi-polynomial complexity, we will want to learn observables with poly $\log n$ support. However, if we were to remove all CZ gates of width $\geq \kappa = c \cdot \log(n + a)$ in this circuit of size $s = \mathrm{poly}(n)$, then

$$\mathbf{W}^{\notin \mathcal{S}^*}[O_{P_i,n}] \leq \frac{9m^2}{2^{\kappa-a}} \leq \frac{9 \cdot s^2 \cdot 2^a}{2^{\log((n+a)^c)}} = \frac{9 \cdot s^2 \cdot 2^a}{(n+a)^c}. \tag{280}$$

Due to the 2^a factor in the numerator of Equation (280), $\mathbf{W}^{\notin \mathcal{S}^*}[O_{P_i}]$ is guaranteed to be $\leq 1/\mathrm{poly}(n)$ only if the number of ancillas is logarithmic, i.e. $a = \mathcal{O}(\log n)$. ∎

Note that the reason our algorithm is restricted to QAC^0 circuits with a logarithmic number of ancillas is the ancilla-dependence in the low-support concentration result of Theorem 39. However, we do not believe this bound to be tight. If (Nadimpalli et al., 2024, Conjecture 1) were proven true, it would imply ancilla-independent low-degree concentration of the Heisenberg-evolved observables.

Conjecture 41 (Nadimpalli et al. (2024) Conjecture 1) *For a size-s, depth-d QAC^0 circuit acting on n-qubits and* $\mathrm{poly}(n)$ *ancilla qubits, then for all $k \in [n+1]$,*

$$W^{>k}[\Phi_{\mathcal{E}_C}] \leq \mathrm{poly}(s) \cdot 2^{-\Omega(k^{1/d})}. \tag{281}$$

Corollary 42 *For a size-s, depth-d QAC^0 circuit acting on n-qubits and* $\mathrm{poly}(n)$ *ancilla qubits, then for all degrees $k \in [n]$,*

$$W^{>k}[O_{P_i,n}] \leq \mathrm{poly}(s) \cdot 2^{-\Omega(k^{1/d})}. \tag{282}$$

Proof By Theorem 19 and Theorem 41, for $k \in [n]$

$$\mathbf{W}^{>k}[O_{P_i,n}] = \sum_{Q \in \mathcal{P}^n : |Q| > k} \left| \widehat{O_{P_i,n}}(Q) \right|^2 \tag{283}$$

$$= 4 \sum_{Q \in \mathcal{P}^n : |Q| > k} \left| \widehat{\Phi_{\mathcal{E}_C}}(Q \otimes P) \right|^2 \tag{284}$$

$$\leq 4 \sum_{R \in \mathcal{P}^{n+1} : |R| > k+1} \left| \widehat{\Phi_{\mathcal{E}_C}}(R) \right|^2 \tag{285}$$

$$\leq 4 \cdot \mathrm{poly}(s) \cdot 2^{-\Omega((k+1)^{1/d})} \tag{286}$$

$$\leq \mathrm{poly}(s) \cdot 2^{-\Omega(k^{1/d})} \tag{287}$$

∎

However, for the purposes of our learning algorithm, it does not suffice to have low-degree concentration. Instead, we need low-support concentration. Thus, we conjecture that the ancilla-dependence of the low-support concentration result of Theorem 39 can be eliminated.

Conjecture 43 (Ancilla-Independent Low-Support Concentration) *For a size-s, depth-d QAC^0 circuit acting on n-qubits and* $\mathrm{poly}(n)$ *ancilla qubits and support $\mathcal{S}$ such that $|\mathcal{S}| = k^d$,*

$$W^{\notin\mathcal{S}}[O_{P_i,n}] \leq \mathrm{poly}(s) \cdot 2^{-\Omega(k^{1/d})}. \tag{288}$$

If this conjecture were proven true, it would imply that our learning algorithm works for QAC^0 circuits with polynomially many ancilla qubits.

Corollary 44 (Efficient learning of QAC^0 with Polynomial Ancillas) *Suppose we are given an $(n+a)$-qubit depth-d QAC^0 circuit governed by unitary C, performing clean computation*

$$C(I \otimes |0^a\rangle) = A \otimes |0^a\rangle \tag{289}$$

on polynomially many ancilla qubits, i.e. $a = \text{poly}(n)$. *For failure probability* $\delta \in (0,1)$, *we can learn a 2n-qubit unitary* A_{sew} *such that*

$$\mathcal{D}_{avg}(A_{sew}, A \otimes A^\dagger) \leq \frac{1}{\text{poly}(n)}, \tag{290}$$

with high probability $1 - \delta$. *The sample and computational complexity of this procedure are quasi-polynomial.*

Proof Following from the proof of Theorem 40, in this case, the distance between $\widetilde{O}_{P_i,n}^{(\ell)}$ and $O_{P_i,n}$ is bounded as

$$\mathcal{D}_F\left(\widetilde{O}_{P_i,n}^{(\ell)}, O_{P_i,n}\right) \leq 2 \cdot 4^\ell \cdot \eta^2 + \mathbf{W}^{\notin \mathcal{S}^*}[O_{P_i,n}] \leq 2 \cdot 4^\ell \cdot \eta^2 + \text{poly}(s) \cdot 2^{-\Omega(k^{1/d})}, \tag{291}$$

where η is the learning accuracy. Since the sample and time complexity are directly related to the supports of the learned observables, to achieve quasi-polynomial complexity, we will want to learn observables with $\text{poly}\log n$ support. In this case, if we remove all CZ gates of width $\geq \kappa = c \cdot \log(n + a)$ where $a = n^b$ in this circuit of size $s = \text{poly}(n)$, then

$$\mathbf{W}^{\notin \mathcal{S}^*}[O_{P_i,n}] \leq \text{poly}(s) \cdot 2^{-\Omega(k^{1/d})} \leq \frac{\text{poly}(s)}{2^{c \cdot \log(n + n^b)}} \leq \frac{\text{poly}(s)}{n^{bc}}. \tag{292}$$

Thus, if c is chosen to be a constant sufficiently large such that $n^{bc} > \text{poly}(s)$, then $\mathbf{W}^{\notin \mathcal{S}^*}[O_{P_i,n}] \leq 1/\text{poly}(n)$ and we obtain the desired learning guarantee. ∎

Appendix F. Hardness of Learning QAC^0

We will conclude with a result on the hardness of learning QAC^0. This result follows straightforwardly from (Huang et al., 2024, Proposition 3), which showed that it is exponentially-hard to learn QAC^0 according to the diamond-norm distance. In particular, the proof leverages a specific worst-case logarithmic-depth circuit U_x, which via a Grover lower-bound is shown to require exponential queries to learn according to the worst-case measure. We simply observe that U_x is in and of itself a QAC^0 circuit to extend the hardness result to QAC^0 circuits. For the sake of completeness, we re-write the theorem statement of (Huang et al., 2024, Proposition 3) in the context of QAC^0 circuits, as well as the proof.

Proposition 45 (Hardness of learning QAC^0) *Consider an unknown n-qubit unitary U generated by a QAC^0 circuit. Then,*

1. *Learning U to $\frac{1}{3}$ dimond distance with high probability requires $\exp(\Omega(n))$ queries.*

2. *Distinguishing whether U equals the identity matrix I or is $\frac{1}{3}$-far from the identity matrix in diamond distance with high probability requires $\exp(\Omega(n))$ queries.*

Proof For $x, y \in \{0,1\}^n$, let U_x be the unitary,

$$U_x \ket{y} = \begin{cases} 1, & x = y, \\ -1 & x \neq y, \end{cases} \tag{293}$$

51

which can be constructed as

$$U_x = \left(\prod_{\substack{i \in [n]: \\ x_i = 0}} X_i \right) CZ_{[n]} \left(\prod_{\substack{i \in [n]: \\ x_i = 0}} X_i \right). \tag{294}$$

Huang et al. (2024) used U_x to prove a learning lower-bound for logarithmic depth circuits comprised of constant-width gates. To begin, they showed that, in the class of circuits comprised solely of constant-width gates, U_x could be synthesized in $\mathcal{O}(\log n)$-depth. Core to the argument is that, if one can learn a unitary U up to $\frac{1}{3}$ error in diamond distance with high probability or distinguish whether U equals the identity I or is $\frac{1}{3}$-far from I in diamond distance with high probability, then one can successfully distinguish I from U_x. However, distinguishing I from one of $U_x, \forall x \in \{0,1\}$ is the Grover search problem. Therefore, by the Grover lower bound (Bennett et al., 1997), the number of queries must be at least $\Omega(2^{n/2}) = \exp(\Omega(n))$.

Note that the unitary U_x is not only contained in the class of logarithmic-depth constant-width circuits, but also is contained in QAC^0, since it is comprised solely of single-qubit gates and an n-qubit CZ gate. Therefore, the argument by and lower-bound of Huang et al. (2024) also applies to QAC^0, concluding the proof. ∎