arXiv:2403.09033v4 [quant-ph] 19 Sep 2025

Adaptivity is not helpful for Pauli channel learning

Xuan Du Trinh and Nengkun Yu

Department of Computer Science, Stony Brook University, Stony Brook, NY 11794, USA

We prove that adaptive strategies offer no advantage over non-adaptive ones for learning and testing Pauli channels using entangled inputs. This key observation allows us to characterize the query complexity for several fundamental tasks by translating optimal classical estimation algorithms into the quantum setting. First, we determine the tight query complexity for learning a Pauli channel under the general ℓ_p norm, providing results that improve upon or match the best-known bounds for the ℓ_1, ℓ_2, and ℓ_∞ distances. Second, we resolve the complexity of testing whether a Pauli channel is a white noise source. Finally, we show that the optimal query complexities for estimating the Shannon entropy and support size of the channel's error distribution, and for estimating the diamond distance between two Pauli channels, are all $\Theta\left(\frac{4^n}{n\epsilon^2}\right)$.

1 Introduction

Quantum noise is a fundamental concept in quantum physics, challenging our ability to demonstrate and exploit the advantages of quantum information processing. In quantum cryptography, noise primarily influences quantum communication tasks by generating errors, thereby reducing the secret key rate of quantum key distribution protocols. To build large-scale and fault-tolerant quantum computers, mitigating quantum noise and performing error correction are essential. For these reasons, modeling and evaluating quantum noise always remain a topic of interest.

In the context of quantum error correction, bit flips and phase flips represent the most fundamental types of errors that illustrate the impact of noise on qubit systems. Consequently, the Pauli channel is widely adopted as a standard noise model.

Definition 1 (Pauli Channel) *For an n-qubit system with density operator ρ, the Pauli channel $\mathcal{P}$ is the completely positive and trace-preserving (CPTP) map*

$$\rho \longmapsto \sum_{i\in\{0,1,2,3\}^n} P(i)\,\tau_i \rho \tau_i^\dagger, \tag{1}$$

where $\tau_i = \sigma_{i_1} \otimes \sigma_{i_2} \otimes \cdots \otimes \sigma_{i_n}$ is a tensor product of Pauli operators chosen from

$$\sigma_0 = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_1 = X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$
$$\sigma_2 = Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

and $P(i)$ is the probability associated with the error operator τ_i.

Xuan Du Trinh: xtrinh@cs.stonybrook.edu
Nengkun Yu: nengkun.yu@cs.stonybrook.edu

The probability distribution P characterizes the channel $\mathcal{P}$. When a quantum state ρ passes through a noisy device modeled by the Pauli channel, information about P can be obtained from measurements on the output state $\mathcal{P}(\rho)$. By preparing known input states and selecting appropriate measurement settings, either projective or generalized, one can collect samples from P.

One might consider a specific parameterization for P while analyzing data samples based on physical models. In particular, several well-studied single-parameter Pauli channels have been analyzed in the literature, such as the bit flip channel, the dephasing channel and the depolarizing channel [1]. However, our analysis treats P in full generality. Learning and testing a probability distribution without presupposing its form presents a challenging and contemporary problem in statistics and learning theory. This paper explores the complexity of various related tasks, including learning the Pauli channel and testing some aspects of this model.

Definition 2 (Learning Complexity) *For $\epsilon, \delta \in (0,1)$, the learning complexity is defined as the number of queries to the Pauli channel required to produce an estimate $\hat{P}$ that is ϵ-close to P with respect to a predefined channel distance, with error probability at most δ.*

To simplify the analysis, when the soundness parameter δ is not specified, it is taken to be a constant, namely $\delta = 1/3$. In the literature, learning complexity has been investigated in two senses: either in terms of the number of Pauli channel applications or the number of measurements. In this paper, we find that distinguishing between these two notions is unnecessary.

Having defined learning complexity, it is natural to introduce a metric to compare the true parameters of the quantum channel with the distribution learned from data samples.

Definition 3 (Channel Distance Metrics) *The distance ℓ_p, with $p \geq 1$, between two probability distributions P and $\hat{P}$ is given by*

$$\ell_p(\hat{P}, P) = \|\hat{P} - P\|_p = \left(\sum_{i \in \{0,1,2,3\}^n} |\hat{P}(i) - P(i)|^p \right)^{1/p}. \tag{2}$$

In particular, $\ell_1(\hat{P}, P) = \sum_i |\hat{P}(i) - P(i)|$ equals twice the total variation distance. The metric $\ell_\infty(\hat{P}, P) = \max_i |\hat{P}(i) - P(i)|$ measures the maximum pointwise difference between the two distributions. Finally, $\ell_2(\hat{P}, P) = \sqrt{\sum_i |\hat{P}(i) - P(i)|^2}$ corresponds to the standard Euclidean distance.

Previous research has yielded remarkable results in estimating Pauli channels under various conditions, including scenarios with and without entanglement and employing different metrics for evaluation (see Table 1). In the entanglement-free scheme, Flammia and Wallman introduced an algorithm requiring $O(n2^n/\epsilon^2)$ measurements to estimate Pauli channels in norm ℓ_2 in [2]. Recently, Fawzi et *al.* refined this algorithm and established an upper bound of $O(n2^{3n}/\epsilon^2)$ measurements in the context of norm ℓ_1 [3]. While focusing on the ℓ_∞ norm and without-entanglement condition, Flammia and O'Donnell related learning of Pauli channels to the Population Recovery problem. They provided an algorithm that uses the channel $O(1/\epsilon^2)\log(n/\epsilon)$ times [4]. In another study in the same context, Chen et *al.* demonstrated that $\Theta(2^n/\epsilon^2)$ measurements are required for learning each eigenvalue $P(i)$ [5]. Earlier, Chen et al. showed that entanglement provides an exponential advantage, with an algorithm that applies the channel $O(n/\epsilon^2)$ times [6], leading to an upper bound of $O(4^n/\epsilon^2)$ in the ℓ_1 distance via the Parseval–Plancherel identity [3]. In

another results, Flammia and O'Donnell highlighted that $\Theta(1/\epsilon^2)$ samples of P are both necessary and sufficient for classically learning P in the ℓ_∞ norm. They further stated that Bell measurements can obtain these samples while ancilla, i.e., entanglement, is available [4]. These studies, alongside others [7–9], highlight the complexity and diversity of techniques in quantum channel discrimination. Motivated by this diversity, our paper aims to establish tight bounds for the learning problem across any ℓ_p norm, where $p \geq 1$.

p	**This study**	**With entanglement**	**Without entanglement**
$p = 1$	$\Theta(4^n/\epsilon^2)$	$\tilde{O}(4^n/\epsilon^2)$[3, 6]	$O(n2^{3n}/\epsilon^2)$[2]
$p = 2$	$\Theta(1/\epsilon^2)$	$None$	$O(n2^n/\epsilon^2)$[2]
$p = \infty$	$\Theta(1/\epsilon^2)$	$\Theta(1/\epsilon^2)$[4][1]	$\Omega(1/\epsilon^2)$, $O(1/\epsilon^2)\log(n/\epsilon)$[4]

Table 1: Comparison of learning complexities obtained in Corollary 1 with existing sample complexity bounds in literature for different common values of p. The two last columns list the results previously obtained with and without entanglement resources in terms of the number of measurement rounds or the number of Pauli channel applications.

In addition to quantum channel discrimination, adaptive quantum learning algorithms play a significant role. These refer to the strategy of dynamically adjusting the choice of quantum measurements and input states based on the outcomes of previous measurements (Figure 1). This approach contrasts with non-adaptive strategies, where the measurement bases are predetermined before experimentation. Notably, the algorithms in [2] employ adaptivity as a crucial component. Subsequently, Fawzi et *al.* in their recent work explored lower bounds for adaptive and non-adaptive strategies in the context of ℓ_1 and ℓ_∞ and specifically in scenarios devoid of entanglement resources [3]. Their results show gaps between the aforementioned lower and upper bounds. This raises an essential question regarding the actual efficacy of adaptivity.

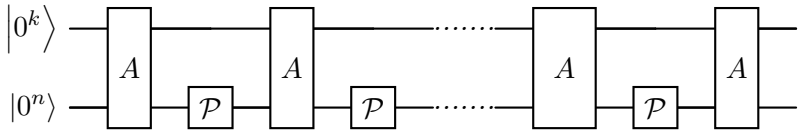

Figure 1: Adaptive strategy for the sampling task: The main n-qubit system and a possible ancilla of k qubits are initialized respectively in the state $|0^n\rangle$ and $|0^k\rangle$. The operation A denotes an adaptive procedure which may involve quantum operations and measurements whose settings can be chosen based on previous outcomes. The gate $\mathcal{P}$ processes the main system in order to extract information from P. The circuit outputs a sample from P.

We present the main theorems of our study, addressing the role of adaptivity in the learning process of the Pauli channel when entanglement resources and ancillary registers are accessible. These theorems enable further analysis of other aspects of Pauli channels.

Theorem 1 *Given a Pauli channel $\mathcal{P}$, the associated random variable P, and the capability to prepare maximally entangled two-qubit states. It is established that a bisimulation*

[1]Flammia and O'Donnell connected the problem of learning $\mathcal{P}$ to the estimation of the discrete distribution P in [4], citing the classical sample complexity of $\Theta(1/\epsilon^2)$ for learning P as reported in [10]. However, they did not provide a proven lower bound for the quantum query complexity.

relationship exists between the Pauli channel $\mathcal{P}$ and the sampling process of its associated probability distribution P. Specifically, the following statements hold

- *The Pauli channel $\mathcal{P}$ can be simulated utilizing the random variable P.*
- *Obtaining one individual sample from P necessitates exactly one application of the channel $\mathcal{P}$.*

This theorem gives an insight that, in terms of information, accessing the channel $\mathcal{P}$ is equivalent to accessing the random variable P. The intuition behind the first point is straightforward since P effectively characterizes $\mathcal{P}$, and the second point will be demonstrated by introducing a sampling algorithm. The equivalence signifies that tasks such as learning or estimating properties of $\mathcal{P}$ can be efficiently executed by sampling from the random variable P and then applying a classical processing algorithm. The impossibility of drawing a larger number of samples N than the number of channel applications M (i.e., $N > M$) is a key element in the proof of Theorem 2. The important implication here is that if the sample complexity for a task is $\mathcal{C}$, then the query complexity on $\mathcal{P}$ directly aligns with $\mathcal{C}$. For instance, achieving an ϵ-close approximation in learning the distribution P within the ℓ_p norm requires $\Theta(poly(\frac{1}{\epsilon}, 2^n))$ samples, then an equal number of queries to the channel $\mathcal{P}$ is required. The upper and lower bounds are all translated from sample complexity to query complexity. A direct and powerful consequence of this one-to-one correspondence between channel applications and samples is that adaptive strategies offer no advantage, a result we formalize in the following theorem.

Theorem 2 *In estimating any properties of channel $\mathcal{P}$ using samples drawn from distribution P, adaptivity does not help to improve channel query complexity compared to non-adaptive strategies.*

Theorem 2 implies that there exists a non-adaptive algorithm that achieves optimal query complexity for the tasks that use samples from P. There will not be an improvement if one uses any adaptive strategy. The proofs of Theorem 1 and Theorem 2 are presented in Section 2. The main technique used is an algorithm for the sampling of Pauli distribution P, that transfers one channel application to yield one sample of P. From these results, subsequent sections will discuss its applications in learning and testing problems, providing Corollaries 1, 2, 3 and 4 while referencing earlier groundbreaking studies [11–13]. Corollaries 3 and 4 will be more formally restated in Corollaries 5 and 6 while the characterization of Pauli channel's noise level and the diamond distance between two Pauli channels will be rigorously defined in the corresponding discussions in Section 3.

Corollary 1 (Learning Complexity of Pauli Channels) *Let $\epsilon \in (0, 1)$ and let $\mathcal{P}$ be an n-qubit Pauli channel. To learn $\mathcal{P}$ within ϵ with respect to the ℓ_p distance and with constant success probability, the number N of required channel applications scales, up to p-dependent constant factors, as*

$$N = \begin{cases} \Theta(4^{\frac{n(2-p)}{p}}\epsilon^{-2}) & \text{for } 1 \le p \le 2 \text{ and } \epsilon \le 4^{-\frac{n(p-1)}{p}} \\ \Theta(4^n) & \text{for } 1 \le p \le 2 \text{ and } 4^{-\frac{n(p-1)}{p}} \le \epsilon \le 2 \cdot 4^{-\frac{n(p-1)}{p}} \\ \Theta(\left(\frac{1}{\epsilon}\right)^{\frac{p}{p-1}}) & \text{for } 1 \le p \le 2 \text{ and } \epsilon \ge 2 \cdot 4^{-\frac{n(p-1)}{p}} \\ \Theta(\epsilon^{-2}) & \text{for } 2 \le p \le \infty \end{cases} \tag{3}$$

Corollary 2 (Complexity of White Noise Testing) *Let $\epsilon \in (0,1)$ and let P denote the discrete distribution characterizing an n-qubit Pauli channel $\mathcal{P}$. Up to constant factors, Table 2 shows the necessary and sufficient number of Pauli channel applications required to distinguish between the following two hypotheses with constant success probability:*

$$H_0 : P = U_{4^n} \quad \text{and} \quad H_1 : \ell_p(P, U_{4^n}) > \epsilon$$

where U_{4^n} is the uniform distribution over 4^n errors.

Corollary 3 (Complexity of Noise Level Estimation) *Let $\mathcal{P}$ be a Pauli channel characterized by the distribution P. To estimate the noise level of $\mathcal{P}$ in terms of the Shannon entropy of P and the support size of P (the number of Pauli errors occurring with non-zero probability), the necessary and sufficient number of Pauli channel applications is $\Theta\left(\frac{4^n}{n\epsilon^2}\right)$.*

Corollary 4 (Complexity of Diamond Distance Estimation) *To estimate the diamond distance between two Pauli channels $\mathcal{P}_1$ and $\mathcal{P}_2$, the necessary and sufficient number of applications of each channel is $\Theta(\frac{4^n}{n\epsilon^2})$.*

p value	Regime	Necessary	Sufficient
$1 \le p \le 2$	$\epsilon \ge 2^{-\frac{2n(p-1)}{p}}$	$\epsilon^{-\frac{p}{2(p-1)}}$	$\epsilon^{-\frac{p}{2(p-1)}}$
	$\epsilon < 2^{-\frac{2n(p-1)}{p}}$	$2^{\frac{n(4-3p)}{p}}\epsilon^{-2}$	$2^{\frac{n(4-3p)}{p}}\epsilon^{-2}$
$2 < p < \infty$	$\Theta\left(\frac{n}{4^n}\right) \ge \epsilon$	$\frac{n}{4^n\epsilon^2}$	$\frac{1}{2^n\epsilon^2}$
	$\Theta\left(\frac{n}{4^n}\right) < \epsilon \le \frac{1}{2^n}$	$\frac{1}{\epsilon}$	$\frac{1}{2^n\epsilon^2}$
	$\epsilon > \frac{1}{2^n}$	$\frac{1}{\epsilon}$	$\frac{1}{\epsilon}$
$p = \infty$	$\Theta\left(\frac{n}{4^n}\right) \ge \epsilon$	$\frac{n}{4^n\epsilon^2}$	$\frac{n}{4^n\epsilon^2}$
	$\Theta\left(\frac{n}{4^n}\right) < \epsilon$	$\frac{1}{\epsilon}$	$\frac{1}{\epsilon}$

Table 2: White noise testing query complexity bounds. The bounds indicate the necessary and sufficient number of channel queries to distinguish Pauli error distribution P from the uniform distribution U_{4^n}, with respect to the ℓ_p-distance ϵ.

2 Sampling with Bell State Measurement as An Optimal Strategy

2.1 Sampling Algorithm

In the theory of quantum operations, Choi's isomorphism offers a representation of quantum channels. In this representation, the channel $\mathcal{P}$ acting on the n-qubit system A is fully characterized by the quantum state

$$\Lambda = (\mathbb{I} \otimes \mathcal{P})(|\Phi_n\rangle \langle\Phi_n|) \tag{4}$$

where $|\Phi_n\rangle = \frac{1}{\sqrt{2^n}} \sum_{i=0}^{2^n-1} |ii\rangle$ is the maximally entangled state in the extension $A \otimes A'$ with A' isomorphic to A. The fact that Λ encodes all information about $\mathcal{P}$ implies that extracting information from Λ is equivalent to "measuring" the quantum channel. It suggests a method for detecting Pauli errors using entangled states. Nevertheless, entangled

measurement on the total system $A \otimes A'$ is generally resource-intensive. So, we will employ n Bell measurements on n pairs of qubits. This strategy was introduced by [14], where the authors demonstrated its application in extracting information and identifying errors in quantum circuits. Our main contribution lies in the discovery of Theorems 1 and 2 with their implications, where the strategy is specifically applied to Pauli channels. We adapt our notations and conventions as described below.

A simple example will give an intuition of the approach. We start with the preparation of a pair of qubits in a maximally entangled state $\left|\Phi^+\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. One of the two qubits is then subjected to a Pauli channel $\mathcal{P}$. Subsequently, a Bell measurement is performed on the qubits (Figure 2).

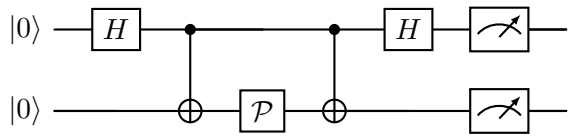

Figure 2: The first Hadamard and CNOT gates create a maximally entangled state. Then, the second qubit is subject to a single-qubit Pauli channel. The second CNOT and Hadamard gates help to project Bell measurement to the measurement in the computational basis.

Effect of Pauli Errors: If a Pauli error occurs on one qubit among an EPR pair, it would transform the initial entangled state $\left|\Phi^+\right\rangle$ into another Bell state. To summarize, the measurement outcomes in a Bell measurement provide information about the types of Pauli errors that may have affected the qubits:

- $|\sigma_0\rangle = \left|\Phi^+\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = (\mathbb{I}_2 \otimes \sigma_0)\left|\Phi^+\right\rangle$ indicates no errors ($\sigma_0 = \mathbb{I}_2$).
- $|\sigma_1\rangle = \left|\Psi^+\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) = (\mathbb{I}_2 \otimes \sigma_1)\left|\Phi^+\right\rangle$ indicates a Pauli-X (σ_1) error.
- $|\sigma_2\rangle = \left|\Psi^-\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) = -i(\mathbb{I}_2 \otimes \sigma_2)\left|\Phi^+\right\rangle$ indicates a Pauli-Y (σ_2) error.
- $|\sigma_3\rangle = \left|\Phi^-\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) = (\mathbb{I}_2 \otimes \sigma_3)\left|\Phi^+\right\rangle$ indicates a Pauli-Z (σ_3) error.

A an interesting feature of this protocol is that the summary remains true regardless of which qubit in each pair is subjected to Pauli errors. Such a strategy can be generalized to measure the n-qubit quantum channel $\mathcal{P}$ using n pairs of maximally entangled qubits for $n \geq 2$ thanks to the fact that

$$|\Phi_n\rangle = \left|\Phi^+\right\rangle^{\otimes n}. \tag{5}$$

From each pair, we pick one qubit to form a set. Then, we have two sets of n qubits. Subsequently, one set passes through the channel $\mathcal{P}$. Bell measurements detect errors on each qubit and its corresponding pairs. For the simplest case $n = 2$, the circuit in Figure 3 shows how the method is employed.

Intuitively, if the tuple of errors $(\sigma_{i_1}, \sigma_{i_2}, ..., \sigma_{i_n})$ is the output of the measurements on n qubit pairs, we get a sample of the distribution P with the value $i = (i_1, i_2, ..., i_n) \in \{0, 1, 2, 3\}^n$. Then, one measurement experiment exploiting one use of $\mathcal{P}$ results in one sample, providing an upper bound for the query complexity of the sampling task. We formalize this direct correspondence in the following lemma.

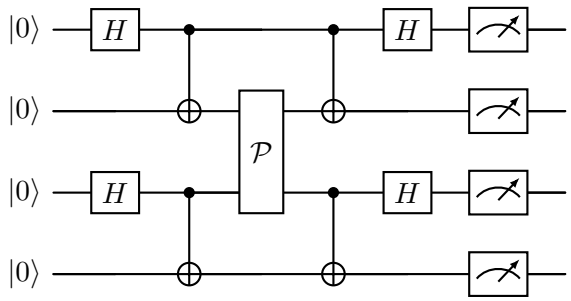

Figure 3: The Pauli channel is applied on the second and third qubits. Then, the errors on these two qubits are detected by Bell measurements on the initial pairs.

Lemma 1 *If Bell measurements are performed on n pairs of qubits using the above strategy, the resulting tuple of errors* $(\sigma_{i_1}, \sigma_{i_2}, \ldots, \sigma_{i_n})$ *forms a sample corresponding to the outcome* $i = (i_1, i_2, \ldots, i_n) \in \{0,1,2,3\}^n$ *drawn from the distribution* P.

Proof. We denote the post-measurement state corresponding to the tuple of error $\tau_i = (\sigma_{i_1}, \sigma_{i_2}, ..., \sigma_{i_n})$ by $|\tau_i\rangle = |\sigma_{i_1}\rangle |\sigma_{i_2}\rangle ... |\sigma_{i_n}\rangle$. Considering the image of the Pauli channel under Choi isomorphism with the remark that $|\sigma_{i_k}\rangle = (\mathbb{I}_2 \otimes \sigma_{i_k}) |\Phi^+\rangle$ and $|\Phi_n\rangle \langle\Phi_n| = (|\Phi^+\rangle \langle\Phi^+|)^{\otimes n}$, we have

$$(\mathbb{I}_A \otimes \mathcal{P})(|\Phi_n\rangle \langle\Phi_n|) = \sum_{i\in\{0,1,2,3\}^n} P(i) (\mathbb{I}_A \otimes \tau_i) |\Phi_n\rangle \langle\Phi_n| (\mathbb{I}_A \otimes \tau_i^\dagger) \tag{6}$$

$$= \sum_{i\in\{0,1,2,3\}^n} P(i) \bigotimes_{k=1}^{n} (\mathbb{I}_2 \otimes \sigma_{i_k}) \left|\Phi^+\right\rangle \left\langle\Phi^+\right| (\mathbb{I}_2 \otimes \sigma_{i_k}^\dagger) \tag{7}$$

$$= \sum_{i\in\{0,1,2,3\}^n} P(i) \bigotimes_{k=1}^{n} |\sigma_{i_k}\rangle \langle\sigma_{i_k}| \tag{8}$$

$$= \sum_{i\in\{0,1,2,3\}^n} P(i) |\tau_i\rangle \langle\tau_i| . \tag{9}$$

So with probability $P(i)$, the Bell measurements detect the tuple error τ_i.

Proof of ***Theorem 1****.* We start from the first point: the Pauli channel $\mathcal{P}$ can be simulated utilizing the random variable P. Consider a quantum system consisting of n qubits initially in the state ρ, on which we aim to simulate the action of the Pauli channel $\mathcal{P}$, given access to the random variable P. This process is straightforward. A sample $i = (i_1, i_2, ..., i_n)$ is drawn from the distribution P and then the corresponding Pauli operation $\tau_i = (\sigma_{i_1}, \sigma_{i_2}, ..., \sigma_{i_n})$ is applied to the n-qubit state. The outcome will be the quantum state $\tau_i \rho \tau_i^\dagger$, with probability $P(i)$, i.e. the quantum state is presented by the density operator

$$\mathcal{P}(\rho) = \sum_{i\in\{0,1,2,3\}^n} P(i) \tau_i \rho \tau_i^\dagger . \tag{10}$$

We now consider the second point: obtaining one individual sample from P necessitates exactly one application of the channel $\mathcal{P}$. The lower bound is evident because, given the Pauli channel $\mathcal{P}$, to draw one sample from the distribution P, the quantum circuit must query $\mathcal{P}$ at least once. Lemma 1 confirms that one query is enough. The upper bound and the lower bound match. Therefore, Theorem 1 is proven □.

2.2 Optimality

Let $\mathcal{A}$ denote the collection of all algorithms that sample from P using Pauli channel queries. For $a \in \mathcal{A}$, the algorithm applies the Pauli channel M_a times and produces N_a samples from P. After Theorem 1, an interesting question arises: is there any algorithm $a \in \mathcal{A}$ so that $N_a > M_a$?

The natural conjecture is that no such algorithm exists. Our insight is that the amount of useful information extracted from M_a channel applications cannot exceed the information contained in M_a random variables, which are independent and identically distributed (i.i.d.) according to P. Theorem 1 confirms the equivalence between the information contained in the random variable P and the Pauli channel $\mathcal{P}$. Before presenting a rigorous proof, our intuitive discussion, following simple logic, will suggest that M_a channel applications can provide at most M_a samples of P, emphasizing the insight of this bisimulation relation.

Considering $a \in \mathcal{A}$, Theorem 1 implies that M_a channel applications can be simulated by M_a i.i.d. copies of the random variable P. In other words, the M_a channel applications in algorithm a can be replaced by a random variable $X \sim P^{\otimes M_a}$, which produces an output $Y \sim P^{\otimes N_a}$. This results in a modified version of algorithm a, and we do not distinguish between the simulated version and the original. We denote the simulation efficiency of a as:

$$r_a = \frac{N_a}{M_a}.$$

Observation. Given only N finite samples from a discrete probability distribution P, no method can produce an estimate $\hat{P}$ that approximates P arbitrarily well with high probability. Equivalently, with high probability the distance $\ell_p(\hat{P}, P)$ is lower-bounded by some error $\epsilon(N) > 0$, regardless of the learning algorithm employed [11].

Suppose, for contradiction, that there exists a simulation algorithm with unbounded simulation efficiency. This would enable the generation of infinitely many samples from a finite initial sample set. A learning algorithm, as described in [11], could then exploit these simulated samples to violate the lower bound $\epsilon(N)$. This yields a contradiction. Hence, the simulation efficiency r_a must be bounded for all algorithms $a \in \mathcal{A}$. In fact, we will further show that $r_a \leq 1$ for all $a \in \mathcal{A}$.

Assume that there exists $a \in \mathcal{A}$ such that $r_a > 1$, i.e., $N_a > M_a$. We describe a family of algorithms $\{b_k\}_{k\in\mathbb{N}} \subset \mathcal{A}$. Algorithm b_k takes M_a copies of P and runs a to obtain N_a copies of P. Then, it recycles M_a out of the N_a output copies as input for another execution of algorithm a. Each time algorithm a is executed, we gain $N_a - M_a$ additional copies of P. Using this procedure, algorithm b_k, after k recycles, produces $M_a + k(N_a - M_a)$ copies of P starting from M_a initial copies. The simulation efficiency of b_k is given by:

$$r_{b_k} = 1 + k\frac{N_a - M_a}{M_a}.$$

This simulation efficiency grows arbitrarily large as k increases, which contradicts the fact that simulation efficiency is bounded. Therefore, the assumption that $r_a > 1$ is incorrect. Hence, the answer to the initial question is that $r_a \leq 1$ for all $a \in \mathcal{A}$. We state our finding in the lemma below and technically prove it with no-cloning principle [15].

Lemma 2 (Tight Bounds of The Sampling Task) *Obtaining $N > 0$ samples of P requires exactly $M = N$ channel applications.*

Proof.

- *Upper bound.* The upper bound $M \leq N$ is achieved by the algorithm in Lemma 1.

- *Lower bound.* In the discussion of the simulation efficiency of sampling algorithms, we primarily used reasoning based on classical information processing. We will now reaffirm our finding using the no-cloning principle. Assume that there exists $a \in \mathcal{A}$ such that $r_a > 1$. Then, in the corresponding family $\{b_k\}_{k\in\mathbb{N}} \subset \mathcal{A}$ described above, there exists b_κ such that $r_{b_\kappa} \geq 2$. This means that from M_a i.i.d. samples of P, we can produce $2M_a$ i.i.d. samples of P. We note that any classical distribution can be written in a density matrix form and in our quantum computing framework, any operation we may perform on a quantum state must be modeled by a CPTP map. Let us consider the density matrix ρ that represents the classical state of M_a i.i.d. samples of P. Then, algorithm b_κ induces a CPTP map $\mathcal{E}_{b_\kappa}$ such that for all classical state ρ that may represent P,

$$\mathcal{E}_{b_\kappa}(\rho \otimes |int\rangle \langle int|) = \rho \otimes \rho, \tag{11}$$

  where $|int\rangle \langle int|$ is the initial state of an ancillary system, which is used to contain information copied from ρ. One can easily find that it is a contradiction. Indeed, considering two classical states ρ_1 and ρ_2, we have that $\rho_{12} = \frac{1}{2}(\rho_1 + \rho_2)$ is also a classical state. Algorithm b_κ is supposed to ensure that

$$\begin{cases} \mathcal{E}_{b_\kappa}(\rho_1 \otimes |int\rangle \langle int|) = \rho_1 \otimes \rho_1 \\ \mathcal{E}_{b_\kappa}(\rho_2 \otimes |int\rangle \langle int|) = \rho_2 \otimes \rho_2 \\ \mathcal{E}_{b_\kappa}(\rho_{12} \otimes |int\rangle \langle int|) = \rho_{12} \otimes \rho_{12}. \end{cases} \tag{12}$$

  However in general, if $\rho_1 \neq \rho_2$, these three equalities violate the linearity of CPTP map $\mathcal{E}_{b_\kappa}$. This contradiction proves that no such algorithm b_κ can exist, which invalidates our initial assumption that $r_a > 1$. Therefore, for any algorithm $a \in \mathcal{A}$, the simulation efficiency $r_a = N_a/M_a \leq 1$. This implies $M_a \geq N_a$, establishing the lower bound.

Proof of ***Theorem 2****.* Lemma 2 directly leads to the proof of Theorem 2. Specifically, given the Pauli channel $\mathcal{P}$, in order to prove that adaptivity offers no advantage in sampling from the distribution P, we established two points:

1. Obtaining N samples of P requires $M = \Theta(N)$ channel applications.
2. There exists a non-adaptive algorithm that achieves this query complexity. □

To elaborate further on this inference: Consider the general strategy shown in Figure 1 in the context of Theorem 1, the redundancy of adaptivity is evident. Theorem 1 indicates that just one application of the gate $\mathcal{P}$ suffices to produce a sample from the distribution P. After this initial use, the entire system may undergo some possible quantum operations. Subsequently, it is subjected to quantum measurement. Any subsequent steps to interpret the measurement outcome are purely classical and do not involve further quantum operations on the system. The possible quantum operations before the measurement cannot be adaptive because no measurement outcome was available.

3 Learning and Testing Pauli Channels

Revisiting the exploration of learning and testing discrete distributions based on ℓ_p norms reveals a longstanding research interest. Waggoner has solidified the definitive bounds for the complexities of learning and uniformity testing [11]. Leveraging his main results allows

us to infer the learning complexity, and the white noise testing for Pauli channels determining if P represents a completely random variable of size 2^n. The logic is very simple. As discussed above, Theorem 1 and Lemma 2 identify the number of samples required to estimate any properties of P with the necessary number of queries to $\mathcal{P}$. Corollaries 1 and 2 are the consequences of the results in [11] and the translation between the two kinds of complexities.

3.1 Learning Algorithm

We now consider the classical processing part of learning channel $\mathcal{P}$. The algorithm proposed in the reference [11] is employed to learn the distribution P. The procedure for drawing samples has been described in the previous section.

Algorithm 1 Learning the discrete distribution P

1: Input: channel $\mathcal{P}$; number of qubits n ; distance parameter $p \in [1, \infty]$ and $\epsilon, \delta \in (0, 1)$
2: Choose N sufficient for p, n, ε, δ according to proven upper bounds in Lemma 3.
3: Draw N samples.
4: Let X_i be the number of samples drawn of each coordinate $i \in \{0, 1, 2, 3\}^n$.
5: Compute each $\hat{P}_i = \frac{X_i}{N}$.
6: Output $\hat{P}$.

We have the exact sufficient numbers of samples to learn the discrete distribution P using Algorithm 1 according to Theorem 5.2 in [11], applied in the context of distribution size of 4^n, as written in Lemma 3. In addition, Theorem 5.4 in [11] shows the lower bounds up to constant factors of learning discrete distributions in ℓ_p, presented in Lemma 4.

Lemma 3 (Exact Upper Bounds on Learning Sample Complexity) *To learn P within error $\epsilon \in (0, 1)$ with respect to the ℓ_p distance and with success probability at least $1 - \delta$, it suffices to run Algorithm 1 with the following number of samples:*

$$N_O = \frac{1}{\delta} \begin{cases} 4^{\frac{n(2-p)}{p}} \epsilon^{-2} & \text{if } 1 \leq p \leq 2 \text{ and } \epsilon \leq 2 \cdot 4^{-\frac{n(p-1)}{p}} \\ \frac{1}{4} \left(\frac{2}{\epsilon}\right)^{\frac{p}{p-1}} & \text{if } 1 \leq p \leq 2 \text{ and } \epsilon \geq 2 \cdot 4^{-\frac{n(p-1)}{p}} \\ \epsilon^{-2} & \text{if } 2 \leq p \leq \infty. \end{cases} \tag{13}$$

Lemma 4 (Lower Bounds on Learning Sample Complexity) *To learn P within error $\epsilon \in (0, 1)$ with respect to the ℓ_p distance and with constant success probability, the required number of samples is at least*

$$N_\Omega = \begin{cases} \Omega\left(4^{\frac{n(2-p)}{p}} \epsilon^{-2}\right) & \text{if } 1 \leq p \leq 2 \text{ and } \epsilon \leq 4^{-\frac{n(p-1)}{p}} \\ \Omega\left(\left(\frac{1}{\epsilon}\right)^{\frac{p}{p-1}}\right) & \text{if } 1 \leq p \leq 2 \text{ and } \epsilon \geq 4^{-\frac{n(p-1)}{p}} \\ \Omega\left(\epsilon^{-2}\right) & \text{if } 2 \leq p \leq \infty \end{cases} \tag{14}$$

The lower bounds and upper bounds in Lemma 3 and 4 match up to constant factors. We conclude the tight bounds for learning Pauli channel $\mathcal{P}$ in Corollary 1.

*Proof of **Corollary 1**.* The translation from distribution sample complexities to channel query complexities is automatic thanks to Theorem 1 and 2. Except the tight bound

$\Theta(4^n)$ in the regime where $1 \leq p \leq 2$ and $4^{-\frac{n(p-1)}{p}} \leq \epsilon \leq 2 \cdot 4^{-\frac{n(p-1)}{p}}$, the derivation of the other values of N directly follows from the matching, up to constant factors, between the bounds in Lemma 3 and 4. Considering this distinct regime, given the fact that $\epsilon = \Theta(4^{-\frac{n(p-1)}{p}})$, the lower bound and the upper bound converge to the same order, i.e. $(\frac{1}{\epsilon})^{\frac{p}{p-1}} = \Theta(4^n)$ and $4^{\frac{n(2-p)}{p}}\epsilon^{-2} = \Theta(4^n)$ □.

At the critical value $p = 2$, all the tight bounds in the corollary converge to $\Theta(1/\epsilon^2)$. Interestingly, for all $p \geq 2$, and specifically for the ℓ_∞ norm, the required number of channel applications remains the same at $\Theta(1/\epsilon^2)$. When $p = 1$, we consistently find ourselves within the regime $\epsilon \leq 2^{-\frac{2n(p-1)}{p}} = 1$, leading to $N = \Theta(4^n/\epsilon^2)$. Remarkably, the complexities presented in the corollary are optimal up to only constant factors, effectively covering all possible scenarios for general ℓ_p norms.

The algorithm employing maximally entangled qubit pairs and Bell measurements demonstrates significant quantum advantages as shown by the comparisons in Table 1. For the ℓ_∞ distance, it refines strategies from previous entanglement-free settings, reducing their complexities from $\Theta(2^n/\epsilon^2)$ in [5] and $O(1/\epsilon^2)\log(n/\epsilon)$ in [4] to $\Theta(1/\epsilon^2)$. The bounds for the case of learning in the distance ℓ_1 also benefit from quantum advantage, improved from $O(n2^{3n}/\epsilon^2)$ to $\Theta(4^n/\epsilon^2)$, and tighter than the previous one using entanglement $\tilde{O}(4^n/\epsilon^2)$. Moreover, for the ℓ_2 distance, the algorithm reduces the upper bound from $O(n2^n/\epsilon^2)$ to merely $\Theta(1/\epsilon^2)$, eliminating the dependency on the number of qubits. Thus, the principal contribution of this part is the establishment of tight query complexity bounds for learning Pauli channels across any ℓ_p distance. Furthermore, the effectiveness of our non-adaptive algorithm further confirms that adaptivity does not provide any additional benefit in learning the Pauli channel when entanglement resources are deployed.

3.2 White Noise Testing

We now consider the problem of testing whether the Pauli channel is a white noise source, that is, whether P is the uniform distribution U_{4^n} over the 4^n elements of $\{0,1,2,3\}^n$. More formally, the task is to distinguish between the two hypotheses

$$H_0 : P = U_{4^n} \qquad \text{and} \qquad H_1 : \ell_p(P, U_{4^n}) > \epsilon$$

with high probability, where deviation from uniformity is measured in the ℓ_p distance exceeding ϵ.

The query complexity for this task follows by translating the classical sample complexity stated in Lemma 5, originally from [11]. Substituting the domain size $k = 4^n$ into the classical bounds immediately yields the results presented in Corollary 2.

Lemma 5 (Classical Uniformity Testing Bounds) *Let P be a distribution over the domain $[k] = \{1, 2, \ldots, k\}$, and let U_k denote the uniform distribution over $[k]$. For $\epsilon \in (0, 1)$, the necessary and sufficient sample complexities for distinguishing between the hypotheses*

$$H_0 : P = U_k \qquad \textit{and} \qquad H_1 : \ell_p(P, U_k) > \epsilon$$

with constant success probability are summarized in Table 3.

3.3 Estimating Noise Level

We begin our discussion on estimating the noise level of Pauli channels by introducing two key characteristic quantities.

p value	Regime	Necessary	Sufficient
$1 \le p \le 2$	$\epsilon \ge k^{-\frac{(p-1)}{p}}$	$\epsilon^{-\frac{p}{2(p-1)}}$	$\epsilon^{-\frac{p}{2(p-1)}}$
	$\epsilon < k^{-\frac{(p-1)}{p}}$	$k^{\frac{(4-3p)}{2p}}\epsilon^{-2}$	$k^{\frac{(4-3p)}{2p}}\epsilon^{-2}$
$2 < p < \infty$	$\Theta\left(\frac{\log k}{k}\right) \ge \epsilon$	$\frac{\log k}{k\epsilon^2}$	$\frac{1}{\sqrt{k}\epsilon^2}$
	$\Theta\left(\frac{\log k}{k}\right) < \epsilon \le \sqrt{k}$	$\frac{1}{\epsilon}$	$\frac{1}{\sqrt{k}\epsilon^2}$
	$\epsilon > \sqrt{k}$	$\frac{1}{\epsilon}$	$\frac{1}{\epsilon}$
$p = \infty$	$\Theta\left(\frac{\log k}{k}\right) \ge \epsilon$	$\frac{\log k}{k\epsilon^2}$	$\frac{\log k}{k\epsilon^2}$
	$\Theta\left(\frac{\log k}{k}\right) < \epsilon$	$\frac{1}{\epsilon}$	$\frac{1}{\epsilon}$

Table 3: Classical sample complexity for uniformity testing. The bounds indicate the necessary and sufficient number of samples to distinguish a distribution P over k elements from the uniform distribution U_k, with respect to the ℓ_p-distance ϵ.

Definition 4 (Shannon Entropy) *The Shannon entropy of a distribution P over the domain $[k] = \{1, 2, \ldots, k\}$ is defined as*

$$H(P) = -\sum_{i=1}^{k} P(i) \log_2 P(i),$$

where, by convention, $0 \log_2 0 = 0$.

Definition 5 (Support Size) *The support size of a distribution P over the domain $[k] = \{1, 2, \ldots, k\}$, denoted by $S(P)$, is defined as the number of elements $i \in [k]$ such that $P(i) \neq 0$.*

Entropy is a fundamental measure of uncertainty in a probability distribution. A higher entropy value indicates greater unpredictability. In this work, we focus on the Shannon entropy, which quantifies the expected amount of information contained in a random variable.

In the context of the Pauli channel, the Shannon entropy of the distribution P over the domain $\{0, 1, 2, 3\}^n$ ranges from 0 to $2n$ bits. This provides insight into the diversity and randomness of errors affecting the n-qubit system, enabling a characterization of the channel's noise level. A natural question arises: even if we can estimate the Pauli error probabilities with high precision, does the resulting estimate $\hat{P}$ yield an accurate approximation of the entropy?

Fortunately, the error in entropy estimation can be bounded using the error in estimating the underlying distribution. Specifically, if the total variation distance between $\hat{P}$ and P satisfies $TV(\hat{P}, P) = \epsilon < \frac{1}{2}$, the Fannes–Audenaert inequality [16] ensures that:

$$|H(\hat{P}) - H(P)| \le \epsilon \log_2(4^n - 1) + h_{\text{bin}}(\epsilon), \tag{15}$$

where $h_{\text{bin}}(\epsilon) = -\epsilon \log_2 \epsilon - (1 - \epsilon) \log_2(1 - \epsilon)$ is the binary entropy function.

Example. Consider two Pauli distributions P and $\hat{P}$ over $n = 10$ qubits:

$$P = \left\{0.99, \frac{0.01}{4^{10} - 1}, \ldots, \frac{0.01}{4^{10} - 1}\right\}, \quad \hat{P} = \left\{0.9, \frac{0.1}{4^{10} - 1}, \ldots, \frac{0.1}{4^{10} - 1}\right\}.$$

In this case:

$$TV(P, \hat{P}) = \epsilon = 0.09, \quad h_{\text{bin}}(\epsilon) \approx 0.436,$$
$$H(P) \approx 0.281, \quad H(\hat{P}) \approx 2.469.$$

The difference between the two entropy values is approximately 2.188, and the theoretical upper bound from Inequality 15 is evaluated to about 2.236. While the bound is fairly tight for small systems, it seems to become increasingly loose as n grows. This motivates the need for improved estimators that remain accurate in high-dimensional regimes.

Another important characteristic of the Pauli channel is its potential sparsity: not all error $\tau_i \in \{0, 1, 2, 3\}^n$ necessarily occur with non-zero probability. Moreover, errors may be concentrated in specific regions of the quantum system or limited to certain error types (e.g., bit flips or phase flips). Given the exponential growth of the distribution size, 4^n, obtaining a large enough sample to observe this sparsity directly would require significant resources.

However, if we can accurately estimate the actual support size $S(P)$ using only a small number of samples, this would complement the entropy estimate and provide a fuller picture of the channel's noise structure. Together, entropy and support size serve as two foundational indicators of the Pauli channel's behavior.

Breakthrough work by Valiant and Valiant [12, 13] proposes a method to estimate such distributional properties—including both entropy and support size, using sublinear sample complexity. We adopt their estimator, referred to as *"Estimate Unseen"*, and formalize its performance in the lemma below.

Lemma 6 (Estimate Unseen) *Let P be a distribution over $[k] = \{1, 2, \ldots, k\}$, and suppose $k \in \mathbb{N}$ is sufficiently large. There exist positive constants α and γ such that for any $\epsilon \in (0, 1)$, the "Estimate Unseen" algorithm, using $\frac{\gamma}{\epsilon^2} \cdot \frac{k}{\log k}$ samples from P, returns with probability at least $1 - e^{-k^\alpha}$:*

- *An estimate $\hat{H}(P)$ satisfying $|\hat{H}(P) - H(P)| < \epsilon$,*
- *An estimate $\hat{S}(P)$ satisfying $\frac{|\hat{S}(P) - S(P)|}{k} < \epsilon$, provided that $P(i) \notin (0, \frac{1}{k})$ for all i.*

Furthermore, the sample complexity $\Theta\left(\frac{1}{\epsilon^2} \cdot \frac{k}{\log k}\right)$ is optimal up to constant factors for this task.

This algorithm functions as a classical information-theoretic tool for analyzing the noise level in the Pauli channel $\mathcal{P}$. We apply Lemma 6 to the probability distribution P that characterizes the Pauli channel $\mathcal{P}$, using the Bell measurement strategy once again to draw samples from P to obtain the formal restatement for Corollary 3 in the following corollary. This result builds on the complexity translation framework developed in Theorem 1 and Theorem 2.

Corollary 5 (Complexity of Noise Level Estimation) *There exists an algorithm satisfying the following property: for $n \in \mathbb{N}$ sufficiently large, there exist positive constants α and γ such that for any $\epsilon \in (0, 1)$, using $\frac{\gamma}{\epsilon^2} \cdot \frac{4^n}{n}$ queries to the Pauli channel $\mathcal{P}$, the algorithm returns with probability at least $1 - e^{-4^{n\alpha}}$:*

- *An estimate $\hat{H}(P)$ satisfying $|\hat{H}(P) - H(P)| < \epsilon$,*
- *An estimate $\hat{S}(P)$ satisfying $\frac{|\hat{S}(P) - S(P)|}{4^n} < \epsilon$, provided that $P(i) \notin (0, \frac{1}{4^n})$ for all i.*

Moreover, the query complexity $\Theta\left(\frac{1}{\epsilon^2} \cdot \frac{4^n}{n}\right)$ is optimal up to constant factors.

3.4 Estimating Diamond Distance between Two Pauli Channels

In this part, we investigate the task of distinguishing between two Pauli channels using the diamond distance. Consider a quantum system composed of n qubits, represented by the Hilbert space A, along with an ancillary system described by a finite-dimensional Hilbert space B. The Pauli channels of interest, denoted by $\mathcal{P}_1$ and $\mathcal{P}_2$, are modeled as CPTP maps acting on $\mathcal{L}(A)$, the space of linear operators on A. Meanwhile, $\mathcal{I}_B$ denotes the identity super-operator on $\mathcal{L}(B)$.

Definition 6 (Diamond Distance) *Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two CPTP maps over $\mathcal{L}(A)$. The diamond distance between $\mathcal{P}_1$ and $\mathcal{P}_2$ is defined as*

$$\|\mathcal{P}_1 - \mathcal{P}_2\|_\diamond = \max_{\chi \geq 0,\ \|\chi\|_1 = 1} \|(\mathcal{I}_B \otimes (\mathcal{P}_1 - \mathcal{P}_2))\,\chi\|_1\,, \tag{16}$$

where χ is a density operator on $B \otimes A$, and $\|\cdot\|_1$ denotes the trace norm.

This measure quantifies the maximum possible deviation between the actions of the two quantum channels, even in the presence of entanglement between the system A and an arbitrary ancillary system B. Specifically, it captures the difference between two quantum channels even when their action are only on part of an arbitrarily large system. This property makes the diamond distance especially relevant for evaluating the fidelity of quantum operations, where seemingly small discrepancies between an ideal channel and its physical implementation can lead to significant changes in computational outcomes. Consequently, the diamond norm serves as a fundamental metric for verifying the precision and robustness of quantum information processing tasks.

In the context of Pauli channels, where two CPTP maps $\mathcal{P}_1$ and $\mathcal{P}_2$ are defined by probability distributions P_1 and P_2, respectively, an insightful result shown in [17] establishes that the diamond distance between these two maps is precisely twice the total variation distance (equivalently, the ℓ_1-distance) between P_1 and P_2. Formally, this relationship is given by

$$\|\mathcal{P}_1 - \mathcal{P}_2\|_\diamond = \sum_{i \in \{0,1,2,3\}^n} |P_1(i) - P_2(i)| = \ell_1(P_1, P_2). \tag{17}$$

According to Theorem 1, achieving an ϵ-accurate approximation in ℓ_1-distance for the distributions P_1 and P_2, via empirical estimates $\hat{P}_1$ and $\hat{P}_2$, requires a sample complexity of $\Theta(4^n/\epsilon^2)$. This implies that approximating the diamond distance between two Pauli channels up to an additive error of 2ϵ can be accomplished using $O(4^n/\epsilon^2)$ queries to the channels. Specifically, the triangle inequality yields

$$|\ell_1(P_1, P_2) - \ell_1(\hat{P}_1, \hat{P}_2)| \leq \ell_1(P_1, \hat{P}_1) + \ell_1(P_2, \hat{P}_2) < 2\epsilon. \tag{18}$$

However, the question regarding the optimality of this approach arises. By exploiting the work of Valiant et al. in [12, 13] and our Bell measurement strategy, we find that the actual number of applications of the Pauli channels needed to estimate $\|\mathcal{P}_1 - \mathcal{P}_2\|_\diamond$ effectively reduces to $\Theta(\frac{4^n}{n\epsilon^2})$, indicating a more efficient estimation process. To achieve this improvement, we utilize the *distance estimator* for the ℓ_1-distance introduced in [13], formalized in the following lemma.

Lemma 7 *Let P_1 and P_2 be probability distributions over the domain $[k] = \{1, 2, \ldots, k\}$, where $k \in \mathbb{N}$ is sufficiently large. Then there exist positive constants α and γ such that for*

any $\epsilon \in (0,1)$*, the* distance estimator*, using* $\frac{\gamma}{\epsilon^2} \cdot \frac{k}{\log k}$ *independent samples from each of* P_1 *and* P_2*, returns, with probability at least* $1 - e^{-k^\alpha}$*, an estimate* $\hat{d}$ *of* $\ell_1(P_1, P_2)$ *satisfying*

$$|\hat{d} - \ell_1(P_1, P_2)| < \epsilon. \tag{19}$$

Furthermore, the sample complexity $\Theta\left(\frac{1}{\epsilon^2} \cdot \frac{k}{\log k}\right)$ *is optimal up to constant factors for this task.*

By applying the same complexity translation technique as in Theorem 1 and Theorem 2, this result directly leads to the following corollary tailored to Pauli channels.

Corollary 6 (Complexity of Diamond Distance Estimation) *There exists an algorithm with the following property: for sufficiently large* n*, there exist positive constants* α *and* γ *such that for any* $\epsilon \in (0,1)$*, the algorithm uses* $\frac{\gamma}{\epsilon^2} \cdot \frac{4^n}{n}$ *queries to each of the Pauli channels* $\mathcal{P}_1$ *and* $\mathcal{P}_2$*, and with probability at least* $1 - e^{-4^{n\alpha}}$*, returns an estimate* $\hat{d}$ *of* $\ell_1(P_1, P_2)$ *satisfying*

$$|\hat{d} - \ell_1(P_1, P_2)| < \epsilon. \tag{20}$$

Moreover, the query complexity $\Theta\left(\frac{1}{\epsilon^2} \cdot \frac{4^n}{n}\right)$ *is optimal up to constant factors for this task.*

4 Conclusion

Our work has addressed the complexities involved in learning and testing Pauli channels, establishing lower and upper bounds across all ℓ_p norms, most of which are tight. Building on foundational work in discrete probability distributions and employing novel algorithms such as "Estimate-Unseen", we have provided deeper insights into quantum noise characterization in Pauli channels. Our exploration highlights the cost of accurately estimating critical properties of the error distribution, such as its Shannon entropy, support size, and the diamond distance to between channels. Furthermore, our findings show that adaptivity provides no advantage when maximally entangled qubit pairs and Bell measurements are available. This work contributes to the theoretical framework for quantum information processing and offers practical insights for developing more efficient quantum computing systems. Future research may extend these methodologies to other quantum channel models, opening new directions for quantum error correction and noise mitigation.

References

[1] Konstantinos Georgopoulos, Clive Emary, and Paolo Zuliani. Modeling and simulating the noisy behavior of near-term quantum computers. *Phys. Rev. A*, 104:062432, Dec 2021. DOI: 10.1103/PhysRevA.104.062432. URL https://doi.org/10.1103/PhysRevA.104.062432.

[2] Steven T. Flammia and Joel J. Wallman. Efficient estimation of pauli channels. *ACM Transactions on Quantum Computing*, 1(1), December 2020. DOI: 10.1145/3408039. URL https://doi.org/10.1145/3408039.

[3] Omar Fawzi, Aadil Oufkir, and Daniel Stilck França. Lower bounds on learning pauli channels with individual measurements. *IEEE Transactions on Information Theory*, 71(4):2642–2661, 2025. DOI: 10.1109/TIT.2025.3527902. URL https://doi.org/10.1109/TIT.2025.3527902.

[4] Steven T. Flammia and Ryan O'Donnell. Pauli error estimation via population recovery. *Quantum*, 5:549, September 2021. DOI: 10.22331/q-2021-09-23-549. URL https://doi.org/10.22331/q-2021-09-23-549.

[5] Senrui Chen, Changhun Oh, Sisi Zhou, Hsin-Yuan Huang, and Liang Jiang. Tight bounds on pauli channel learning without entanglement. *Phys. Rev. Lett.*, 132:180805, May 2024. DOI: 10.1103/PhysRevLett.132.180805. URL https://doi.org/10.1103/PhysRevLett.132.180805.

[6] Senrui Chen, Sisi Zhou, Alireza Seif, and Liang Jiang. Quantum advantages for pauli channel estimation. *Phys. Rev. A*, 105:032435, Mar 2022. DOI: 10.1103/PhysRevA.105.032435. URL https://doi.org/10.1103/PhysRevA.105.032435.

[7] Joel J. Wallman and Joseph Emerson. Noise tailoring for scalable quantum computation via randomized compiling. *Phys. Rev. A*, 94:052325, Nov 2016. DOI: 10.1103/PhysRevA.94.052325. URL https://doi.org/10.1103/PhysRevA.94.052325.

[8] S. Chen, Y. Liu, M. Otten, et al. The learnability of pauli noise. *Nature Communications*, 14:52, 2023. DOI: 10.1038/s41467-022-35759-4. URL https://doi.org/10.1038/s41467-022-35759-4. Received: 4 July 2022; Accepted: 23 December 2022; Published: 4 January 2023.

[9] Cambyse Rouzé and Daniel Stilck França. Efficient learning of the structure and parameters of local pauli noise channels. 2023. DOI: 10.48550/arXiv.2307.02959. URL https://doi.org/10.48550/arXiv.2307.02959.

[10] Clément L. Canonne. A short note on learning discrete distributions. 2020. DOI: 10.48550/arXiv.2002.11457. URL https://doi.org/10.48550/arXiv.2002.11457.

[11] Bo Waggoner. Lp testing and learning of discrete distributions. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, ITCS '15, page 347–356, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450333337. DOI: 10.1145/2688073.2688095. URL https://doi.org/10.1145/2688073.2688095.

[12] Gregory Valiant and Paul Valiant. Estimating the unseen: an n/log(n)-sample estimator for entropy and support size, shown optimal via new clts. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, STOC '11, page 685–694, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450306911. DOI: 10.1145/1993636.1993727. URL https://doi.org/10.1145/1993636.1993727.

[13] Gregory Valiant and Paul Valiant. Estimating the unseen: Improved estimators for entropy and other properties. *J. ACM*, 64(6), October 2017. ISSN 0004-5411. DOI: 10.1145/3125643. URL https://doi.org/10.1145/3125643.

[14] Dominik Hangleiter and Michael J. Gullans. Bell sampling from quantum circuits. *Physical Review Letters*, 133(2):020601, 2024. DOI: 10.1103/PhysRevLett.133.020601. URL https://doi.org/10.1103/PhysRevLett.133.020601.

[15] William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802–803, 1982. DOI: 10.1038/299802a0. URL https://doi.org/10.1038/299802a0.

[16] Gerardo Adesso and Fabrizio Illuminati. Entanglement in continuous variable systems: recent advances and current perspectives. *Journal of Physics A: Mathematical and Theoretical*, 40(28):7821–7880, 2007. DOI: 10.1088/1751-8113/40/28/S18. URL https://doi.org/10.1088/1751-8113/40/28/S18.

[17] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A*, 85:042311, Apr 2012. DOI: 10.1103/PhysRevA.85.042311. URL https://doi.org/10.1103/PhysRevA.85.042311.