Near-optimal incoherent tomography of low-rank quantum channels

Kean Chen* Aadil Oufkir†

Abstract

We study tomography for quantum channels with input dimension d_1, output dimension d_2, and Kraus rank at most r, to within diamond norm error ε, using adaptive experiments that retain no quantum memory between channel queries.

- For quantum channels whose non-zero Choi eigenvalues are bounded below by $\Omega(d_1/r)$, we establish optimal query upper and lower bounds $\Theta(d_1 d_2 r^2/\epsilon^2)$. The upper bound is achieved by a nonadaptive algorithm that uses the estimator from [SSKKG22], together with a new diamond-norm analysis. The lower bound applies to arbitrary adaptive incoherent protocols and follows from a new local family of channels and a uniform one-query Fisher-information bound.
- For general channels, we establish an upper bound $O(d_1 d_2 r^2 \log(2d_1)/\epsilon^2)$, nearly matching the above lower bound $\Omega(d_1 d_2 r^2/\varepsilon^2)$. To achieve this, we generalize the above nonadaptive algorithm by adapting the input state over $O(\log(2d_1))$ rounds with the Matrix Multiplicative Weight Update algorithm.

*University of Pennsylvania, USA. Email: keanchen.gan@gmail.com
†University Mohammed VI Polytechnic, Morocco. Email: aadil.oufkir@gmail.com

1 Introduction

Quantum channel learning is the task of reconstructing an unknown quantum channel from experimental data. It provides an approximate classical description of the input-output behavior of a physical device and is therefore useful for controlling and validating quantum devices, characterizing noise, and comparing experiments with theory [PCZ97, MRL08, EHW+20]. This task has been extensively studied since the early works of [CN97, PCZ97], through compressed-sensing methods [KKEG19, KRT17], shadow-tomography approaches [HCP23, KTCT23], and projected least-squares methods [SSKKG22].

In this work, we study the *query complexity* of quantum channel tomography: given black-box access to a quantum channel $\mathcal{E} : \mathcal{L}(\mathbb{C}^{d_1}) \to \mathcal{L}(\mathbb{C}^{d_2})$, how many queries to $\mathcal{E}$ are necessary and sufficient to produce a classical description of a quantum channel $\widehat{\mathcal{E}}$ satisfying $\left\|\widehat{\mathcal{E}} - \mathcal{E}\right\|_{\diamond} \leq \epsilon$, with probability at least $2/3$? We measure the accuracy of the estimate in diamond norm, which is an operational natural metric since it quantifies the worst-case distinguishability of two channels when arbitrary reference systems, input states and final measurements are allowed.

The query complexity of this problem depends crucially on the access model. In the coherent setting, a learning algorithm has quantum memory, may choose entangled input state, apply several copies of the channel $\mathcal{E}$ coherently and perform entangled final measurement. In the incoherent setting, a learning algorithm has no quantum memory and has to measure the output state after each use of the channel $\mathcal{E}$. Classical information can be used in the design of new input and measurement. For full-rank state tomography and channel tomography, coherent algorithms provably outperform incoherent ones. For tomography of a full-rank d-dimensional quantum state, the coherent query complexity is $\Theta(d^2/\epsilon^2)$ [OW16, HHJ+17] and the incoherent one is $\Theta(d^3/\epsilon^2)$ [HHJ+17, GKKT20, CHL+23]. For full-rank channel tomography, a similar tradeoff occurs for both the diamond norm and the Choi trace norm: the coherent query complexity is $\Theta(d_1^2 d_2^2/\epsilon^2)$ [MB25, CGO+26] and the incoherent one is $\Theta(d_1^3 d_2^3/\epsilon^2)$ [Ouf23, BPGM26].

Usually, physical processes are far from full rank. A channel with small Choi rank r has a Choi operator supported on a low-dimensional subspace and hence can be described using $O(d_1 d_2 r)$ parameters. In the coherent setting, the query complexity of channel learning is mostly known: [CGO+26] shows that the query complexity is $\Theta(d_1 d_2 r/\epsilon^2)$ whenever the dilation rate $\tau \coloneqq r d_2/d_1 \geq 1 + \Omega(1)$ (away-from-boundary regime). At the boundary regime $\tau = 1$, the complexity is $\Theta(d_1 d_2 r/\epsilon)$, exhibiting Heisenberg scaling [CGO+26, HLS+26]. In the incoherent setting, the rank-dependent query complexity was not fully characterized beyond the special cases $r = 1$ [CGO+26] and $r = d_1 d_2$ [Ouf23, BPGM26]. This motivates the main question of this work.

Question 1.1. *What is the query complexity of quantum channel tomography under the diamond norm in the incoherent model?*

In this paper, we answer this question (up to logarithmic factors). We prove that the query complexity in the incoherent setting is $\widetilde{\Theta}(d_1 d_2 r^2/\epsilon^2)$. In the away-from-boundary regime $\tau \geq 1+\Omega(1)$, this represents a factor $\widetilde{O}(r)$ overhead compared to the coherent query complexity. Moreover, the incoherent lower bound has classical $1/\epsilon^2$ scaling throughout the feasible parameter regime, including the boundary regime $\tau = 1$.

1.1 Main results

Let $\mathsf{QChan}^r_{d_1,d_2}$ denote the set of quantum channels $\mathcal{E} : \mathcal{L}(\mathbb{C}^{d_1}) \to \mathcal{L}(\mathbb{C}^{d_2})$ with Kraus rank at most r. An *incoherent* protocol uses the channel once in each round, with an arbitrary input entangled with a fresh reference, and immediately measures or discards all quantum registers. Subsequent inputs and measurements may depend on the classical transcript, consisting of the measurement outcomes and any private random seed, but no quantum information is retained between rounds. We count channel uses and impose no computational restriction on the measurements or classical postprocessing.

Write $D := d_1 d_2$. A channel of Kraus rank exactly r exists if and only if $r \le D$ and $d_1 \le r d_2$. We consider the nontrivial case $d_2 \ge 2$ and therefore assume

$$d_2 \ge 2, \qquad 1 \le r \le D = d_1 d_2, \qquad d_1 \le r d_2. \tag{1}$$

When $d_2 = 1$, the trace map is the only channel and no query is needed.

Our first result concerns quantum channels that have $\Omega(d_1/r)$-gapped Choi spectrum.

Theorem 1.1 (Upper bound for the gapped case, Theorem 3.2 restated)**.** *For every $\mathcal{E} \in \mathsf{QChan}^r_{d_1,d_2}$ such that the non-zero eigenvalues of its Choi operator are at least $\Omega(d_1/r)$, the nonadaptive incoherent protocol in Algorithm 1 outputs a channel $\widehat{\mathcal{E}}$ satisfying* $\Pr\left\{\left\|\widehat{\mathcal{E}} - \mathcal{E}\right\|_\diamond > \epsilon\right\} \le 1/3$ *using at most*

$$O\left(\frac{d_1 d_2 r^2}{\epsilon^2}\right)$$

queries.

Algorithm 1 uses the linear estimator of [SSKKG22] with Haar random measurements on the Choi state. We develop a new analysis of this protocol to handle the diamond distance.

For general channels, the spectrum need not be gapped. We show that adapting the input state leads to a similar complexity up to a logarithmic factor.

Theorem 1.2 (Upper bound, Theorem 3.4 restated)**.** *For every $\mathcal{E} \in \mathsf{QChan}^r_{d_1,d_2}$, the adaptive incoherent protocol in Algorithm 3 outputs a channel $\widehat{\mathcal{E}}$ satisfying* $\Pr\left\{\left\|\widehat{\mathcal{E}} - \mathcal{E}\right\|_\diamond > \epsilon\right\} \le 1/3$ *using at most*

$$O\left(\frac{d_1 d_2 r^2}{\epsilon^2} \log(2d_1)\right)$$

queries.

Algorithm 3 uses $O(\log(2d_1))$ adaptive updates of the input state and an independent Haar-random basis measurement after each channel use.

We complement our upper bounds with the following lower bound.

Theorem 1.3 (Lower bound, Theorem 4.1 restated)**.** *Every adaptive incoherent protocol that learns every channel in $\mathsf{QChan}^r_{d_1,d_2}$ to diamond-norm error ϵ with probability at least $2/3$ uses at least*

$$\Omega\left(\frac{d_1 d_2 r^2}{\epsilon^2}\right)$$

queries. The hard instance family used to prove the lower bound can be chosen so that every non-zero eigenvalue of its Choi operator belongs to $[d_1/(4r), 4d_1/r]$.

Our results are summarized in Table 1.

Table 1: Incoherent query complexity of quantum channel tomography at constant success probability.

Setting	Upper bound	Lower bound
Gapped, nonadaptive	$O\left(\frac{d_1 d_2 r^2}{\epsilon^2}\right)$ Theorem 3.2	$\Omega\left(\frac{d_1 d_2 r^2}{\epsilon^2}\right)$ Theorem 4.1
General, adaptive	$O\left(\frac{d_1 d_2 r^2 \log(2d_1)}{\epsilon^2}\right)$ Theorem 3.4	$\Omega\left(\frac{d_1 d_2 r^2}{\epsilon^2}\right)$ Theorem 4.1

1.2 Overview of techniques

Upper bound, Gapped case. When the smallest positive eigenvalue of the Choi operator $C_{\mathcal{E}}$ is at least $\Omega(d_1/r)$ (Gapped case), the protocol of [SSKKG22] is optimal up to constants (at constant success probability). It consists of measuring the Choi state with Haar random bases, averaging the formed Haar snapshots (Definition 2.8), and projecting to a valid Choi state $\widetilde{X}$. In order to control the diamond distance, we split the error $\Delta = \widetilde{X} - \frac{1}{d_1} C_{\mathcal{E}}$ to four blocks according to the projection $P_{\mathcal{E}}$ onto the support of the Choi operator $C_{\mathcal{E}}$. The support block $P_{\mathcal{E}} \Delta P_{\mathcal{E}}$ contributes $\kappa(\mathcal{E})\|\Delta\|_\infty$ where $\kappa(\mathcal{E}) = d_1 \|\operatorname{tr}_B P_{\mathcal{E}}\|_\infty$. For $\Omega(d_1/r)$-gapped channels $P_{\mathcal{E}} \preceq O(\frac{r}{d_1}) C_{\mathcal{E}}$ so $\kappa(\mathcal{E}) \le O(r)$ and the contribution of the support block is at most $O(r\|\Delta\|_\infty)$. The same contribution (up to confidence-dependent factors) can be proven for the complement block $(I - P_{\mathcal{E}})\Delta(I - P_{\mathcal{E}})$ and this is the technical heart: Since we measure with a global Haar basis, and we project on the complement, the law of Δ is invariant under every unitary supported on $(\operatorname{supp} C_{\mathcal{E}})^\perp$, which can be used to flatten the input marginal (Lemma 3.10). The cross blocks are handled by Cauchy Schwarz inequality. A scalar Bernstein argument combined with a sphere covering net (Section 2.3), using the dimension-free subexponential moments of every directional Haar snapshot, gives $\|\Delta\|_\infty = O(\sqrt{(D + \log(1/\delta))/n})$ in the relevant range (Lemma 2.10). Together with the block analysis, this yields $\|\widehat{\mathcal{E}} - \mathcal{E}\|_\diamond = O(r\sqrt{D/n})$ at constant success probability.

Upper bound, general case. From the sketch above, the difficulty for general channels comes from bounding $\kappa(\mathcal{E})$ which can be much larger than r. In this case, we go beyond fixed maximally entangled input case, allowing for adaptive input states. Now, for an arbitrary input state σ_{in} we prepare the joint output state $X = \mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)$ and the orthogonal projector on its support P_X. Following the same previous strategy, the quantity one would like to control is $\kappa_{\sigma_{\text{in}}}(\mathcal{E}) := \left\|\sigma_{\text{in}}^{-1/2}(\operatorname{tr}_B P_X)^{\mathrm{T}}\sigma_{\text{in}}^{-1/2}\right\|_\infty$. It turns out that, for every channel $\mathcal{E}$, one can show that there is input state σ_{in} for which $\kappa_{\sigma_{\text{in}}}(\mathcal{E}) = O(r)$ and thus can be used to obtain near optimal upper bound on the query complexity. However, finding such input state is nontrivial. The difficulty comes partially from the projector P_X which is a discontinuous function of X. To overcome this difficulty, we soften the projector P_X using the following filter for $s > 0$:

$$f_s(X) = X(X + sI)^{-1} \underset{s\downarrow 0}{\longrightarrow} P_X.$$

The new quantity that takes the role of $\kappa_{\sigma_{\text{in}}}(\mathcal{E})$ is then $\|H_s(\sigma_{\text{in}})\|_\infty$ where $H_s(\sigma_{\text{in}}) = \Phi_{\sigma_{\text{in}}}(f_s(X))^\dagger(I_{\mathrm{B}})$ and $\Phi_{\sigma_{\text{in}}}(M)$ is the linear map whose Choi operator is $\sigma_{\text{in}}^{-\mathrm{T}/2} M \sigma_{\text{in}}^{-\mathrm{T}/2}$. Indeed, we have from Lemma 2.14 that $\lim_{s\downarrow 0}\|H_s(I/d_1)\|_\infty = \kappa(\mathcal{E})$. So the goal now is to find an input state σ_{in} such that $\|H_s(\sigma_{\text{in}})\|_\infty = O(r)$.

Given the filter $f_s(X)$ and defining $X_s = X - sf_s(X) \approx X$ we can decompose the channel $\mathcal{E} = \Phi_{\sigma_{\text{in}}}(X_s) + s\Phi_{\sigma_{\text{in}}}(f_s(X))$. Since $\|s\Phi_{\sigma_{\text{in}}}(f_s(X))\|_\diamond = s\|H_s(\sigma_{\text{in}})\|_\infty$, by setting $s = O(\epsilon/r)$, we can omit the second term in the decomposition once the goal $\|H_s(\sigma_{\text{in}})\|_\infty \le O(r)$ is achieved. The first term of the decomposition $\Phi_{\sigma_{\text{in}}}(X_s)$ motivates the shift of focus to $X_s = X^2(X+sI)^{-1}$. Furthermore, estimating X_s permits to control $\|H_s(\sigma_{\text{in}})\|_\infty$ since $H_s(\sigma_{\text{in}}) = \frac{1}{s}\left(I_{\text{A}} - \Phi_{\sigma_{\text{in}}}(X_s)^\dagger(I_B)\right)$.

The operator X_s can be written as $X_s = XRX$ with $R = (X+sI)^{-1}P_X$ which suggests an estimation using Nyström sandwich: we prepare two *independent* estimators $\widehat{X}_0$ and $\widehat{X}_1$ of X using averages of Haar snapshots. We estimate $\widehat{R} = (S+sI)^{-1}P_S$ where S retains the r largest eigenvalues of $\widehat{X}_0$ and set the remaining to zero. We then estimate X_s with $\widehat{X}_s = \widehat{X}_1\widehat{R}\widehat{X}_1$.

The sandwich form ensures that $\widehat{X}_s \succeq 0$ and thus $\Phi_{\sigma_{\text{in}}}(\widehat{X}_s)$ is completely positive. Moreover, since $\widehat{R}$ is independent of $\widehat{X}_1$, the error can be controlled by writing it as

$$\widehat{X}_s - X_s = X(\widehat{R} - R)X + E\widehat{R}X + X\widehat{R}E + E\widehat{R}E$$

where $E = \widehat{X}_1 - X$. The cross terms can be controlled using a Löwner Cauchy-Schwarz inequality. The same expression can be used to get a multiplicative estimate $\widehat{H}(\sigma_{\text{in}}) = \frac{1}{s}(I_{\text{A}} - \Phi_{\sigma_{\text{in}}}(\widehat{X}_s)^\dagger(I_B))$ of $H_s(\sigma_{\text{in}})$. Concentration inequalities show that the near-optimal upper bound is proven whenever $\|H_s(\sigma_{\text{in}})\|_\infty \le 10r$ (Proposition 3.15).

The last part of this upper bound is the design of input state σ_{in} achieving $\|H_s(\sigma_{\text{in}})\|_\infty \le 10r$. It turns out that $H_s(\sigma_{\text{in}}) = (\mathcal{E}^c)^\dagger((\mathcal{E}^c(\sigma_{\text{in}}) + sI_r)^{-1})$ where $\mathcal{E}^c$ is the complementary channel of $\mathcal{E}$ according to a fixed Kraus list(Lemmas 2.12 and 2.13) so $\operatorname{tr}(\sigma_{\text{in}} H_s(\sigma_{\text{in}})) \le r$ for every σ_{in}. Roughly speaking, we want to turn an average guarantee $\operatorname{tr}(\sigma_{\text{in}} H_s(\sigma_{\text{in}})) \le r$ to a worst case guarantee $\|H_s(\sigma_{\text{in}})\|_\infty \le 10r$. This can be achieved by using the matrix exponentiated-gradient update [TRW05]:

$$\rho_{j+1} = \frac{\exp(\log\rho_j + \log A_j)}{\operatorname{tr}(\exp(\log\rho_j + \log A_j))}, \qquad \sigma_j = \frac{1}{2}\rho_j + \frac{I_{\text{A}}}{2d_1}.$$

Here, A_j is obtained from $\widehat{H}(\sigma_j) + rI_{\text{A}}$ by replacing every eigenvalue below $r/2$ by $r/2$. This update raises the weight of input directions on which H_s is large. It can be shown that $T = \max\{1, \lceil \log_2 d_1 \rceil\}$ iterations suffice using two ingredients. Golden–Thompson inequality gives $\operatorname{tr}(\exp(\log\rho_j + \log A_j)) \le \operatorname{tr}(\rho_j A_j)$. On the concentration event, A_j is a multiplicative approximation to $A(\rho_j) = H_s(\sigma_j) + rI_{\text{A}}$, while the trace budget $\operatorname{tr}(\rho_j A(\rho_j)) \le 3r$ follows from $\rho_j \preceq 2\sigma_j$ and $\operatorname{tr}(\sigma_j H_s(\sigma_j)) \le r$. Consequently, $\operatorname{tr}(\rho_j A_j) = O(r)$. The second ingredient is the operator convexity of $\rho \mapsto \log A(\rho)$ (Corollary 3.20), which translates a bound on the average of $\log A_j$ to a bound on $\log(H_s(\bar{\sigma}) + rI_{\text{A}})$, where $\bar{\sigma} = \frac{1}{T}\sum_{j=1}^T \sigma_j$. After T iterations, $H_s(\bar{\sigma}) \preceq 9rI_{\text{A}}$ (Proposition 3.21), so $\bar{\sigma}$ can be used as an input state, leading to a near-optimal query bound.

Lower bound. We prove the lower bound by the standard strategy of constructing a hard family and using Fano's inequality to show that any learning algorithm must query the channel at least a certain number of times to learn a channel chosen randomly from the hard family (see e.g., [FGLE12, HHJ+17, LN25, FOF25]). In the regime $d_1 \le rd_2/2$, one could use the hard family constructed in [CGO+26] and, instead of bounding the mutual information by the Holevo information as in [OG26], use log-Sobolev inequalities for Haar measure [Mec19, Theorem 5.16]. In order to obtain a general bound for all parameters such that $d_1 \le rd_2$ and $r \le d_1d_2$, we construct a new family of hard channels by first constructing a base channel whose non-zero Choi eigenvalues lie between $\frac{d_1}{2r}$ and $\frac{2d_1}{r}$. Then, the admissible channels of the family are constructed by a small perturbation of the Kraus operators of the base channel and normalization (to ensure the trace-preserving property). This set has dimension $\Omega(d_1d_2r)$ (Lemma 4.3), and the hard family is constructed by taking Gaussian weights on a basis. Let $\mathcal{E}_\Theta$ be a random hard channel in this family, and let Z be the transcript of classical outcomes of a learning algorithm. The lower bound is proved through bounding the mutual information $I(\Theta; Z)$ between Θ and Z. On the one hand, learning $\mathcal{E}$ to diamond error ϵ permits the learner to find Θ in a small ball (Lemma 4.12). Correctness of the learning algorithm and Fano's inequality imply that $I(\Theta; Z) \geq \Omega(d_1 d_2 r)$ (Lemma 4.13). On the other hand, each query reveals only a little information. In fact, using the tester representation of single-query experiments, one can show that the trace of the one-query Fisher information is at most $O(d_1 d_2/r)$ uniformly (Lemma 4.7). The crucial point is that, even for adaptive incoherent algorithms, Fisher information adds along the transcript, and the trace of the n-query Fisher information $I_Z(\Theta)$ is at most $O(nd_1 d_2/r)$ (see Equation (117)). The Gaussian log-Sobolev inequality (Lemma 4.10) translates the bound on the Fisher information into a bound on the mutual information, $I(\Theta; Z) \leq O\left(\frac{\epsilon^2}{d_1 d_2}\right) \operatorname{tr} I_Z(\Theta) \leq O(n\epsilon^2/r)$, leading to the lower bound $n \geq \Omega\left(\frac{d_1 d_2 r^2}{\epsilon^2}\right)$.

1.3 Related work

Incoherent quantum channel tomography. For general Kraus rank r, [KRT17, KKEG19] use compressed-sensing techniques to establish low-rank recovery in Schatten norms. Similar guarantees are established using projected least squares [GKKT20, SSKKG22]. Converting these bounds to diamond distance, using standard inequalities, yields suboptimal bounds in general. The ancilla-assisted projected least squares protocol of [SSKKG22] is the estimator we use in the gapped case.

For full rank $r = d_1 d_2$, [Ouf23] showed the near optimal query complexity $\widetilde{\Theta}(d_1^3 d_2^3/\epsilon^2)$ for nonadaptive incoherent channel tomography under diamond distance. This was generalized to the adaptive incoherent setting by [BPGM26] who established the complexity $\Theta(d_1^3 d_2^3/\epsilon^2)$.

When the rank is minimal $r = 1$, channel tomography becomes isometry tomography. [HKOT23] learn a unitary channel on $\mathbb{C}^d$ to diamond error ϵ using $O(d^2/\epsilon^2)$ incoherent measurements. Their sharper $O(d^2/\epsilon)$ bound uses a sequential bootstrap procedure that repeatedly composes the unknown unitary with a current classical estimate. It therefore falls outside the incoherent model. Then, [CGO+26] generalized this result to learn an isometry channel of dimensions (d_1, d_2) to diamond error ϵ using $O(d_1 d_2/\epsilon^2)$ incoherent measurements. Our results extend both results to higher Kraus rank r.

Coherent quantum channel tomography. The upper bound $O(d_1 d_2 r/\epsilon^2)$ on the coherent query complexity was established by [MB25, CGO+26] and the lower bound $\Omega(d_1 d_2 r/\epsilon^2)$ was proved by [CGO+26] when the dilation rate $\tau = rd_2/d_1 \geq 1 + \Omega(1)$. At the boundary $\tau = 1$, channel tomography exhibits Heisenberg scaling $\Theta(d_1 d_2 r/\epsilon)$ [CGO+26, HLS+26], which is not the case in the incoherent setting where $\Omega(d_1 d_2 r^2/\epsilon^2)$ is needed for all $\tau \geq 1$ (Theorem 1.3).

Quantum state tomography. State tomography can be seen as a special case of channel tomography with trivial input, i.e., $d_1 = 1$. The coherent query complexity is $\Theta(d_2 r/\epsilon^2)$ [OW16, HHJ+17]. The full rank incoherent query complexity $\Theta(d_2^3/\epsilon^2)$ is due to [HHJ+17, GKKT20] for the upper bound and to [CHL+23] for the lower bound. The rank-r explicit dependency was upper bounded by $O(d_2 r^2/\epsilon^2)$ [HHJ+17, GKKT20]. It was lower bounded by $\Omega(d_2 r^2/\epsilon^2)$ by [LN25] for nonadaptive strategies. Theorem 1.3 specialized to $d_1 = 1$ generalized this result to adaptive strategies so the incoherent rank-r state query complexity is $\Theta(d_2 r^2/\epsilon^2)$. We note that this state lower bound was independently proved by [KLMR26, NZ26] as a special case. The general framework for lower bounds was used in [FGLE12, HHJ+17, LN25, FOF25]. Related Fisher-information and log-Sobolev arguments also appear in [KLMR26, NZ26].

1.4 Discussion

Our results answer Question 1.1 by characterizing the cost of quantum memory in low-rank quantum channel tomography. For $\Omega(d_1/r)$-gapped channels, a nonadaptive strategy, with a fixed maximally entangled input state, achieves the optimal incoherent query complexity

$$\Theta\left(\frac{d_1 d_2 r^2}{\epsilon^2}\right)$$

for constant success probability. For general channels, adaptive input selection permits to overcome the gapped spectrum assumption at the cost of an overhead of $\log(2d_1)$ in query complexity. Consequently, compared with the coherent channel tomography results in [CGO+26], we see that the advantages of quantum memory lie in the Choi rank r, and also the error ε (in the boundary and near-boundary regimes), up to a logarithmic factor.

A limitation of our upper bounds is that both algorithms are ancilla-assisted as they prepare a purification of the input state and measure the joint output-reference system. It remains open whether ancilla-free algorithms can achieve comparable query complexities for general Kraus rank r (for minimal rank $r = 1$ and for full rank $r = d_1 d_2$ this is true according to [CGO+26] and [Ouf23, BPGM26] respectively).

Moreover, our algorithms use exact Haar-random bases. It is natural to ask whether finite exact or approximate unitary t-designs can replace Haar randomness while preserving the query complexity of channel tomography.

1.5 Organization

In Section 2, we introduce the notation and tools used throughout the paper. In Section 3, we prove the upper bounds for channel tomography. We begin in Section 3.1 with a nonadaptive protocol based on a fixed maximally entangled input state and establish a query upper bound in terms of the parameter $\kappa(\mathcal{E})$; this yields the optimal query complexity for gapped channels. We then turn in Section 3.2 to general channels, where adaptive selection of the input state removes the spectral-gap assumption at a logarithmic overhead. The technical lemmas used in both upper bounds are collected in Section 3.3.

In Section 4, we prove the lower bound for adaptive incoherent protocols. The hard family consists of gapped channels, so the lower bound applies both to the gapped subclass and to the full class of channels of Kraus rank at most r. The main argument appears in Section 4.1, and the geometric and information-theoretic estimates are proved in Section 4.2.

2 Preliminaries

2.1 Quantum channels and Choi operators

For a finite-dimensional Hilbert space $\mathcal{H}$, write $\mathcal{L}(\mathcal{H})$ for its linear operators and $\mathcal{D}(\mathcal{H})$ for its density operators. The symbols $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ denote the trace, Hilbert–Schmidt, and operator norms, respectively; vector norms are Euclidean. On a real linear space of complex matrices, orthogonality is with respect to the real Hilbert–Schmidt inner product $\langle M, N\rangle_{\mathbb{R}} = \operatorname{Re}\operatorname{tr}(M^\dagger N)$.

For Hermitian M, its support is $\operatorname{supp} M = (\ker M)^\perp$. The Löwner order $M \preceq N$ means that $N - M$ is positive semidefinite. By a *multiplicative estimate* $\widehat{M}$ of a positive definite matrix M, with relative error $0 \le \eta < 1$, we mean $(1-\eta)M \preceq \widehat{M} \preceq (1+\eta)M$. All logarithms are natural unless a base is specified, and $\widetilde{O}$ and $\widetilde{\Theta}$ suppress logarithmic factors.

Let A and B denote the input and output quantum systems, with Hilbert spaces $\mathcal{H}_\mathrm{A} \cong \mathbb{C}^{d_1}$ and $\mathcal{H}_\mathrm{B} \cong \mathbb{C}^{d_2}$. We retain system labels in subscripts: I_A denotes the identity operator on $\mathcal{H}_\mathrm{A}$, id_A the identity map on $\mathcal{L}(\mathcal{H}_\mathrm{A})$, and tr_A the partial trace over $\mathcal{H}_\mathrm{A}$; analogous notation applies to other systems.

We fix the computational basis, and define transposes and complex conjugates with respect to this basis. Complex conjugation is denoted by a superscript $\star$; in particular, $|\psi^\star\rangle$ is the complex conjugate of $|\psi\rangle$. For $K : \mathcal{H}_\mathrm{A} \to \mathcal{H}_\mathrm{B}$, we write

$$|K\rangle\!\rangle = \sum_{a=1}^{d_1} (K|a\rangle) \otimes |a\rangle \in \mathcal{H}_\mathrm{B} \otimes \mathcal{H}_\mathrm{A}, \qquad |I\rangle\!\rangle = \sum_{a=1}^{d_1} |a\rangle \otimes |a\rangle \in \mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{A}. \tag{2}$$

The inverse reshaping map is written mat, so $\mathrm{mat}(|K\rangle\!\rangle) = K$. Therefore, we can easily see

$$\langle\!\langle K|L\rangle\!\rangle = \mathrm{tr}(K^\dagger L), \quad |KM\rangle\!\rangle = (I \otimes M^\mathrm{T})|K\rangle\!\rangle, \quad \mathrm{tr}_\mathrm{B}(|K\rangle\!\rangle\langle\!\langle L|) = (L^\dagger K)^\mathrm{T}. \tag{3}$$

For a linear map $\Phi : \mathcal{L}(\mathcal{H}_\mathrm{A}) \to \mathcal{L}(\mathcal{H}_\mathrm{B})$, define its Choi operator $C_\Phi \in \mathcal{L}(\mathcal{H}_\mathrm{B} \otimes \mathcal{H}_\mathrm{A})$ by

$$C_\Phi = (\Phi \otimes \mathrm{id}_\mathrm{A})(|I\rangle\!\rangle\langle\!\langle I|) = \sum_{a,a'} \Phi(|a\rangle\langle a'|) \otimes |a\rangle\langle a'|. \tag{4}$$

For a channel $\mathcal{E} \in \mathsf{QChan}^r_{d_1,d_2}$, we have $\mathrm{tr}\, C_\mathcal{E} = d_1$. The Choi state refers to the normalized Choi operator $C_\mathcal{E}/d_1$. Complete positivity of Φ is equivalent to $C_\Phi \succeq 0$, and trace preservation is equivalent to $\mathrm{tr}_\mathrm{B}\, C_\Phi = I_\mathrm{A}$. A Kraus representation and its Choi representation are

$$\mathcal{E}(\rho) = \sum_{i=1}^{r} K_i \rho K_i^\dagger, \qquad \sum_i K_i^\dagger K_i = I_\mathrm{A}, \qquad C_\mathcal{E} = \sum_i |K_i\rangle\!\rangle\langle\!\langle K_i|. \tag{5}$$

The minimum number of Kraus operators, called the *Kraus rank* or *Choi rank*, equals $\mathrm{rank}(C_\mathcal{E})$.

We write $\mathsf{QChan}^r_{d_1,d_2}$ for the set of quantum channels $\mathcal{E} : \mathcal{L}(\mathbb{C}^{d_1}) \to \mathcal{L}(\mathbb{C}^{d_2})$ with Kraus rank at most r.

Proposition 2.1 (Diamond norm and Choi operators [Wat18]). *Let $\Phi : \mathcal{L}(\mathcal{H}_\mathrm{A}) \to \mathcal{L}(\mathcal{H}_\mathrm{B})$ be Hermiticity-preserving, with Choi operator C_Φ defined in Equation (4). Then*

$$\|\Phi\|_\diamond = \max_{\rho \in \mathcal{D}(\mathcal{H}_\mathrm{A})} \left\| (I_\mathrm{B} \otimes \sqrt{\rho^\mathrm{T}}) C_\Phi (I_\mathrm{B} \otimes \sqrt{\rho^\mathrm{T}}) \right\|_1. \tag{6}$$

In particular, any two channels $\mathcal{E}, \mathcal{F} : \mathcal{L}(\mathcal{H}_\mathrm{A}) \to \mathcal{L}(\mathcal{H}_\mathrm{B})$ satisfy

$$\frac{1}{d_1}\|C_\mathcal{E} - C_\mathcal{F}\|_1 \le \|\mathcal{E} - \mathcal{F}\|_\diamond \le \|C_\mathcal{E} - C_\mathcal{F}\|_1. \tag{7}$$

We also record two elementary trace-norm consequences of the Löwner order.

Proposition 2.2 (Order and trace-norm estimates). *Let $G \succeq 0$ and let Q be Hermitian, both acting on the same finite-dimensional space.*

(i) *If* $-G \preceq Q \preceq G$ *then* $\|Q\|_1 \le \operatorname{tr} G$.

(ii) *If* $\|Q\|_\infty \le \eta$ *and* Π *is the orthogonal projection onto* $\operatorname{supp} Q$, *then* $\|ZQZ^\dagger\|_1 \le \eta \operatorname{tr}(Z\Pi Z^\dagger)$ *for every matrix* Z.

Proof. (i) Let $\Pi_\pm$ be the spectral projections of Q onto non-negative and negative eigenspaces. Then $\|Q\|_1 = \operatorname{tr}(Q\Pi_+) - \operatorname{tr}(Q\Pi_-) \le \operatorname{tr}(G\Pi_+) + \operatorname{tr}(G\Pi_-) = \operatorname{tr} G$. (ii) Write $Q = |Q|^{1/2}S|Q|^{1/2}$ with $\|S\|_\infty \le 1$. Then $ZQZ^\dagger = (Z|Q|^{1/2}S)(|Q|^{1/2}Z^\dagger)$ and, by $\|MK\|_1 \le \|M\|_2 \|K\|_2$, $\|ZQZ^\dagger\|_1 \le \operatorname{tr}(Z|Q|Z^\dagger) \le \eta \operatorname{tr}(Z\Pi Z^\dagger)$, using $|Q| \preceq \eta\Pi$. □

2.2 Single-query testers

A single-query experiment prepares a state on $\mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{R}$, applies $\mathcal{E} \otimes \mathrm{id}_\mathrm{R}$, and measures the resulting state. Any mixed input can be simulated by purifying it, enlarging the reference system, and extending the measurement trivially to the purifying register. We may therefore take the input to be a pure state $|\psi\rangle \in \mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{R}$.

We allow the measurement to have an arbitrary standard-Borel outcome space $\mathcal{Y}$. Thus the measurement is described by a POVM

$$\mathsf{M} : \mathcal{B}(\mathcal{Y}) \longrightarrow \mathcal{L}(\mathcal{H}_\mathrm{B} \otimes \mathcal{H}_\mathrm{R}), \qquad \mathsf{M}(\mathcal{Y}) = I_\mathrm{B} \otimes I_\mathrm{R}. \tag{8}$$

For every Borel set $B \subseteq \mathcal{Y}$, the outcome law is

$$\Pr_{\mathcal{E}}\{Y \in B\} = \operatorname{tr}[\mathsf{M}(B)(\mathcal{E} \otimes \mathrm{id}_\mathrm{R})(|\psi\rangle\langle\psi|)]\,. \tag{9}$$

Using the copied input space in Equations (2) and (4), write $|\psi\rangle = (I_\mathrm{A} \otimes L)|I\rangle\!\rangle$, where $L : \mathcal{H}_\mathrm{A} \to \mathcal{H}_\mathrm{R}$. Then

$$L^\dagger L = \operatorname{tr}_\mathrm{R}(|\psi\rangle\langle\psi|)^\mathrm{T}. \tag{10}$$

Define the operator-valued tester measure

$$\mathsf{T}(B) := (I_\mathrm{B} \otimes L^\dagger)\mathsf{M}(B)(I_\mathrm{B} \otimes L). \tag{11}$$

Since the outcome state is $(I_\mathrm{B} \otimes L)C_\mathcal{E}(I_\mathrm{B} \otimes L^\dagger)$, cyclicity of the trace gives

$$\Pr_{\mathcal{E}}\{Y \in B\} = \operatorname{tr}(\mathsf{T}(B)C_\mathcal{E}), \qquad \mathsf{T}(\mathcal{Y}) = I_\mathrm{B} \otimes \operatorname{tr}_\mathrm{R}(|\psi\rangle\langle\psi|)^\mathrm{T}. \tag{12}$$

The tester depends only on the chosen input and measurement, not on $\mathcal{E}$.

If $\mathcal{Y}$ is countable, then $T_y := \mathsf{T}(\{y\})$ and

$$\Pr_{\mathcal{E}}\{Y = y\} = \operatorname{tr}(T_y C_\mathcal{E}), \qquad \sum_{y \in \mathcal{Y}} T_y = \mathsf{T}(\mathcal{Y}). \tag{13}$$

Incoherent protocol. An incoherent protocol repeats single-query experiments without retaining quantum memory between rounds.

Specifically, an n-query incoherent protocol chooses, in round t, a state on $\mathcal{H}_\mathrm{A} \otimes \mathcal{H}_{\mathrm{R}_t}$ and a POVM on $\mathcal{H}_\mathrm{B} \otimes \mathcal{H}_{\mathrm{R}_t}$, as functions only of its previous classical outcomes, where R_t is the ancilla system used in round t. It applies $\mathcal{E}_{A\to B} \otimes \mathrm{id}_{\mathrm{R}_t}$ once and performs the POVM. All quantum registers are then discarded. After n rounds it outputs a channel determined by the transcript. The protocol is called nonadaptive when the inputs and POVMs are independent of earlier outcomes, otherwise it is called adaptive.

2.3 Covering and packing nets

We collect the elementary results of covering and packing nets used in this paper; see also [Ver18, Secs. 4.2 and 4.4]. All distances in this subsection are Euclidean. In particular, the unit sphere of $\mathbb{C}^d$ is viewed as a subset of $\mathbb{R}^{2d}$ when estimating cardinalities.

Definition 2.3 (Covering and packing nets)**.** Let S be a subset of a finite-dimensional Euclidean space and let $\eta > 0$. A finite subset $\mathcal{N} \subseteq S$ is an *η-covering net* of S if every $x \in S$ has some $x_0 \in \mathcal{N}$ with $\|x - x_0\|_2 \le \eta$. A finite subset $\mathcal{P} \subseteq S$ is an *η-packing net* if $\|x - y\|_2 > \eta$ for any distinct $x, y \in \mathcal{P}$. A packing net is *maximal* if no further point of S can be added while preserving this separation; maximality is with respect to inclusion, not a requirement of maximum cardinality.

Lemma 2.4 (Covering nets from packing nets)**.** *Let S be a nonempty subset of the Euclidean unit ball in $\mathbb{R}^m$. For every $\eta > 0$, there exists a maximal η-packing net of S. Every such packing net is also an η-covering net and has cardinality at most $(1 + 2/\eta)^m$. Consequently, for $d \ge 1$, the unit sphere of $\mathbb{C}^d$ has an η-covering net $\mathcal{N}$ satisfying*

$$|\mathcal{N}| \le \left(1 + \frac{2}{\eta}\right)^{2d}. \tag{14}$$

The same bounds hold for the unit sphere of any d-dimensional complex subspace.

Proof. For any finite η-packing net $\mathcal{P} \subseteq S$, the Euclidean balls of radius $\eta/2$ centered at its points are pairwise disjoint and lie inside the ball of radius $1 + \eta/2$ centered at the origin. Writing v_m for the volume of the unit ball in $\mathbb{R}^m$, comparison of volumes gives

$$|\mathcal{P}|\, v_m(\eta/2)^m \le v_m(1 + \eta/2)^m, \qquad |\mathcal{P}| \le (1 + 2/\eta)^m.$$

Starting with one point of S and repeatedly adding a point at distance strictly greater than η from all previously chosen points must therefore terminate. The resulting packing net is maximal. Any maximal η-packing net covers S at radius η: otherwise, an uncovered point could be added, contradicting maximality. Apply the real-dimensional bound with $m = 2d$ to obtain Equation (14). □

Lemma 2.5 (Norm approximation by covering nets)**.** *Let $\mathcal{N}_q$ be an η-covering net of the unit sphere of $\mathbb{C}^q$, where $0 < \eta < 1$.*

(i) *For every $A \in \mathbb{C}^{p\times q}$,*

$$\|A\|_\infty \le \frac{1}{1-\eta} \max_{|y\rangle \in \mathcal{N}_q} \|A|y\rangle\|_2. \tag{15}$$

In particular, for every $|z\rangle \in \mathbb{C}^q$,

$$\||z\rangle\|_2 \le \frac{1}{1-\eta} \max_{|y\rangle \in \mathcal{N}_q} |\langle y|z\rangle|. \tag{16}$$

(ii) *If $\eta < 1/2$ and $\mathcal{N}_p$ is an η-covering net of the unit sphere of $\mathbb{C}^p$, then every $A \in \mathbb{C}^{p\times q}$ satisfies*

$$\|A\|_\infty \le \frac{1}{1-2\eta} \max_{\substack{|x\rangle \in \mathcal{N}_p \\ |y\rangle \in \mathcal{N}_q}} |\langle x|A|y\rangle|. \tag{17}$$

(iii) *If $\eta < 1/2$ and $M \in \mathbb{C}^{q\times q}$ is Hermitian, then*

$$\|M\|_\infty \leq \frac{1}{1-2\eta} \max_{|y\rangle\in\mathcal{N}_q} |\langle y|M|y\rangle|. \tag{18}$$

(iv) *For $A \in \mathbb{C}^{p\times q}$, define $s_{\min}(A) = \min_{\||y\rangle\|_2=1} \|A|y\rangle\|_2$. Then*

$$s_{\min}(A) \geq \min_{|y\rangle\in\mathcal{N}_q} \|A|y\rangle\|_2 - \eta \|A\|_\infty . \tag{19}$$

Proof. For each unit vector $|y\rangle$, choose $|y_0\rangle \in \mathcal{N}_q$ with $\||y\rangle - |y_0\rangle\|_2 \leq \eta$. The triangle inequality gives

$$|\|A|y\rangle\|_2 - \|A|y_0\rangle\|_2| \leq \|A(|y\rangle - |y_0\rangle)\|_2 \leq \eta \|A\|_\infty .$$

Taking the supremum over $|y\rangle$ and rearranging proves Equation (15); applying it to the linear functional $A = \langle z|$ proves Equation (16). Taking the infimum instead proves Equation (19).

For (ii), given unit vectors $|x\rangle, |y\rangle$, choose $|x_0\rangle \in \mathcal{N}_p$ and $|y_0\rangle \in \mathcal{N}_q$ within distance η. Since $|x_0\rangle$ is also a unit vector,

$$\begin{aligned}|\langle x|A|y\rangle - \langle x_0|A|y_0\rangle| &\leq |(\langle x| - \langle x_0|)A|y\rangle| + |\langle x_0|A(|y\rangle - |y_0\rangle)| \\ &\leq 2\eta \|A\|_\infty .\end{aligned}$$

Use $\|A\|_\infty = \sup_{\||x\rangle\|_2=\||y\rangle\|_2=1} |\langle x|A|y\rangle|$ and rearrange. For (iii), the same estimate with $|x\rangle = |y\rangle$ and $|x_0\rangle = |y_0\rangle$ gives

$$|\langle y|M|y\rangle - \langle y_0|M|y_0\rangle| \leq 2\eta \|M\|_\infty .$$

Now use the Hermitian identity $\|M\|_\infty = \sup_{\||y\rangle\|_2=1} |\langle y|M|y\rangle|$ and rearrange. □

Corollary 2.6 (Covering-net union bounds)**.** *For a random Hermitian $d\times d$ matrix M and a random complex $p\times q$ matrix A, the following hold for every $t > 0$:*

$$\Pr\{\|M\|_\infty > 2t\} \leq 9^{2d} \sup_{\||x\rangle\|_2=1} \Pr\{|\langle x|M|x\rangle| > t\}, \tag{20}$$

$$\Pr\{\|A\|_\infty > 2t\} \leq 9^{2(p+q)} \sup_{\||x\rangle\|_2=\||y\rangle\|_2=1} \Pr\{|\langle x|A|y\rangle| > t\}. \tag{21}$$

In the second line, $|x\rangle \in \mathbb{C}^p$ and $|y\rangle \in \mathbb{C}^q$.

Proof. Choose deterministic 1/4-covering nets as in Lemma 2.4. For M, apply Equation (18) and a union bound over one covering net. For A, apply Equation (17) and a union bound over the Cartesian product of two covering nets. The respective cardinality bounds are 9^{2d} and $9^{2(p+q)}$. □

2.4 Haar-random basis measurement

Definition 2.7 (Haar-random basis)**.** Let $U \in \mathbb{U}_d$ be a Haar-random unitary. We call $\{U|1\rangle, \ldots, U|d\rangle\}$ a Haar-random basis on the d-dimensional Hilbert space span$\{|1\rangle, \ldots, |d\rangle\}$.

Definition 2.8 (Haar snapshot)**.** Let ρ be a quantum state on $\mathbb{C}^d$. Suppose we apply a Haar-random basis measurement on ρ, obtaining an outcome vector $|\psi\rangle$. We call

$$Y = (d+1)|\psi\rangle\langle\psi| - I_d, \tag{22}$$

a *Haar snapshot* of ρ. For n independent repetitions, obtaining outcome vectors $|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_n\rangle$, we call the average of the resulting snapshots an n-snapshot estimator:

$$\widehat{X}_n = \frac{1}{n}\sum_{t=1}^{n} \left((d+1)|\psi_t\rangle\langle\psi_t| - I_d\right). \tag{23}$$

The snapshot Y has trace one and satisfies $\mathbb{E}\, Y = \rho$, but need not be positive semidefinite. To establish unbiasedness and concentration, we use the following concentration inequality.

Fact 2.9 (Scalar Bernstein for subexponential random variables [Ver18, Proposition 2.7.1 and Theorem 2.8.1])**.** *For a real-valued random variable Z, define*

$$\|Z\|_{\psi_1} := \inf\{s > 0 : \ \mathbb{E}\exp(|Z|/s) \le 2\}.$$

Up to absolute constants, $\|Z\|_{\psi_1} = \sup_{p\ge 1} p^{-1}\left(\mathbb{E}\,|Z|^p\right)^{1/p}$. Let $Z_1, \ldots, Z_n$ be independent centered real-valued random variables, put $K_t = \|Z_t\|_{\psi_1}$, $V = \sum_t K_t^2$, and $K = \max_t K_t$. There is an absolute constant $c > 0$ such that, for every $s \ge 0$,

$$\Pr\left\{\left|\sum_{t=1}^{n} Z_t\right| > s\right\} \le 2\exp\left(-c\min\left\{\frac{s^2}{V}, \frac{s}{K}\right\}\right).$$

In particular, if $K_t \le K_0$ for every t, then there is an absolute constant $c_1 > 0$ such that, for every $u \ge 1$,

$$\Pr\left\{\left|\frac{1}{n}\sum_{t=1}^{n} Z_t\right| > c_1 K_0\left(\sqrt{\frac{u}{n}} + \frac{u}{n}\right)\right\} \le 2e^{-u}.$$

The same conclusions hold for complex-valued variables, up to a change of the absolute constants, by applying the real-valued statement to their real and imaginary parts.

Lemma 2.10 (Haar outcomes and snapshot moments)**.** *Let ρ be a quantum state on $\mathbb{C}^d$ and let $|\psi_1\rangle, \ldots, |\psi_n\rangle$ be the independent outcome vectors obtained by measuring ρ in Haar-random orthonormal bases. Let $\widehat{X}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$, with $Y_j = (d+1)|\psi_j\rangle\langle\psi_j| - I_d$ as in Equation (22), and write $Y = (d+1)|\psi\rangle\langle\psi| - I_d$ for a single Haar snapshot. Then:*

(i) *the observed vector $|\psi\rangle$ has density $p_\rho(|\psi\rangle) = d\langle\psi|\rho|\psi\rangle \le d$ with respect to the uniform (Haar) probability measure on the unit sphere of $\mathbb{C}^d$;*

(ii) $\mathbb{E}\, Y = \rho$, $\mathbb{E}\, Y^2 = (d-1)\rho + dI$, $\mathbb{E}(Y-\rho)^2 \preceq 2dI$ *and* $\|Y - \rho\|_\infty \le d+1$*;*

(iii) *there is an absolute constant c_{H} such that, for every $0 < \zeta < 1$,*

$$\Pr\left\{\left\|\widehat{X}_n - \rho\right\|_\infty > c_{\mathrm{H}}\left(\sqrt{\frac{d + \log(2/\zeta)}{n}} + \frac{d + \log(2/\zeta)}{n}\right)\right\} \le \zeta; \tag{24}$$

(iv) *if U is a unitary with $U\rho U^\dagger = \rho$, then $U\widehat{X}_n U^\dagger$ has the same law as $\widehat{X}_n$.*

Proof. (i) Each vector of a Haar-random orthonormal basis is marginally Haar distributed on the sphere, and outcome i occurs with Born probability $\langle\psi_i|\rho|\psi_i\rangle$. Hence for a test function f, $\mathbb{E}\, f(|\psi\rangle) = \sum_{i=1}^d \mathbb{E}_{\mathrm{Haar}}[\langle\psi_i|\rho|\psi_i\rangle f(|\psi_i\rangle)] = d\, \mathbb{E}_{|\psi\rangle\sim\mathrm{Haar}}[\langle\psi|\rho|\psi\rangle f(|\psi\rangle)] = \mathbb{E}_{|\psi\rangle\sim\mathrm{Haar}}[p_\rho(|\psi\rangle) f(|\psi\rangle)]$.

For (ii) and (iii) we use the standard moment formula

$$\mathbb{E}_{|\varphi\rangle\sim\mathrm{Haar}}[(|\varphi\rangle\langle\varphi|)^{\otimes k}] = \binom{d+k-1}{k}^{-1} \Pi_{\mathrm{sym}}^{(k)}, \qquad \Pi_{\mathrm{sym}}^{(k)} = \frac{1}{k!}\sum_{\pi\in S_k} W_\pi, \tag{25}$$

where $W_\pi = \sum_{i_1,\ldots,i_k} |i_1 i_2 \cdots i_k\rangle\langle i_{\pi^{-1}(1)} i_{\pi^{-1}(2)} \cdots i_{\pi^{-1}(k)}|$ permutes tensor factors according to the permutation $\pi \in S_k$: the left side is a positive operator commuting with every $U^{\otimes k}$ and supported on the symmetric subspace, hence proportional to $\Pi_{\mathrm{sym}}^{(k)}$ by Schur's lemma, and the constant is fixed by the trace [Wat18, Sec. 7.1].

(ii) By (i) and Equation (25) with $k = 2$,

$$\mathbb{E}|\psi\rangle\langle\psi| = d\ \mathrm{tr}_1[(\rho\otimes I)\, \mathbb{E}_{\mathrm{Haar}}(|\varphi\rangle\langle\varphi|)^{\otimes 2}] = \frac{2d}{d(d+1)}\, \mathrm{tr}_1\Big[(\rho\otimes I)\frac{I + W_{(12)}}{2}\Big] = \frac{I+\rho}{d+1},$$

hence $\mathbb{E}\, Y = \rho$. Since $Y^2 = (d+1)(d-1)|\psi\rangle\langle\psi| + I$ we get $\mathbb{E}\, Y^2 = (d-1)\rho + dI$, and $\mathbb{E}(Y-\rho)^2 = \mathbb{E}\, Y^2 - \rho^2 \preceq (d-1)I + dI \preceq 2dI$. The eigenvalues of Y are d and -1, so $\|Y-\rho\|_\infty \le d+1$.

(iii) We next prove the dimension-log-free estimate in Equation (24). Fix a unit vector $|x\rangle$, put $P_x = |x\rangle\langle x|$, and set $Z = (d+1)|\langle x|\psi\rangle|^2$. For every integer $k \ge 1$, (i) and Equation (25) give

$$\mathbb{E}\, Z^k = \frac{d(d+1)^k}{\binom{d+k}{k+1}}\, \mathrm{tr}\big[(\rho\otimes P_x^{\otimes k})\Pi_{\mathrm{sym}}^{(k+1)}\big] \le (k+1)!.$$

Indeed, after expanding the symmetrizer, every permutation contributes either $\mathrm{tr}\,\rho = 1$ or $\mathrm{tr}(\rho P_x) = \langle x|\rho|x\rangle \le 1$, and $d(d+1)^k/\binom{d+k}{k+1} \le (k+1)!$. Since $\langle x|(Y-\rho)|x\rangle = Z - \mathbb{E}\, Z$, the elementary inequality $|a-b|^k \le 2^{k-1}(|a|^k + |b|^k)$ shows that these centered random variables have subexponential norm $\|Z - \mathbb{E}\, Z\|_{\psi_1}$ bounded by an absolute constant, uniformly in d, ρ, and $|x\rangle$. Scalar Bernstein (Fact 2.9) therefore yields absolute constants $c, c_1 > 0$ such that

$$\Pr\big\{\big|\langle x|(\widehat{X}_n - \rho)|x\rangle\big| > t\big\} \le 2\exp(-cn\min\{t^2, t\}).$$

Apply the 1/4-covering-net union bound in Equation (20) to $M = \widehat{X}_n - \rho$ to obtain

$$\Pr\{\big\|\widehat{X}_n - \rho\big\|_\infty > 2t\} \le 2\exp(2d\log 9 - cn\min\{t^2, t\}).$$

Taking $t = c_1(\sqrt{(d+\log(2/\zeta))/n} + (d+\log(2/\zeta))/n)$ proves Equation (24) after enlarging the absolute constant.

(iv) By (i) the density of $U|\psi\rangle$ is $d\langle\psi|U^\dagger\rho U|\psi\rangle = d\langle\psi|\rho|\psi\rangle$, so $U|\psi\rangle$ and $|\psi\rangle$ have the same law; the snapshots are independent and identically distributed. □

Haar experiments. For the upper bounds in Section 3, we use single-query experiments to obtain Haar snapshots of the joint output state. Let $\mathcal{E} \in \mathsf{QChan}^r_{d_1,d_2}$ have Kraus operators $K_1, \ldots, K_r$, and let $\sigma_{\mathrm{in}} \in \mathcal{D}(\mathcal{H}_{\mathrm{A}})$. Prepare its purification $|\sqrt{\sigma_{\mathrm{in}}}\rangle\rangle \in \mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{A}}$ and apply $\mathcal{E}$ to the first system (in the rest of the paper, when a quantum channel acts on a bipartite system, we assume it acts on the first subsystem, with the identity on the second subsystem). The resulting joint state is

$$\begin{aligned}\mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|) &:= (\mathcal{E}\otimes \mathrm{id})(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)\\ &= (I_{\mathrm{B}}\otimes \sqrt{\sigma_{\text{in}}^{\mathrm{T}}})C_{\mathcal{E}}(I_{\mathrm{B}}\otimes \sqrt{\sigma_{\text{in}}^{\mathrm{T}}})\\ &= \sum_i |K_i\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle K_i\sqrt{\sigma_{\text{in}}}|.\end{aligned}\tag{26}$$

It satisfies $\operatorname{rank}\mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)\le r$, $\operatorname{tr}_{\mathrm{B}}\mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)=\sigma_{\text{in}}^{\mathrm{T}}$, and $\operatorname{tr}\mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)=1$.

Measure this state in an independently chosen Haar-random orthonormal basis of $\mathcal{H}_{\mathrm{B}}\otimes\mathcal{H}_{\mathrm{A}}$. The observed basis vector $|\psi\rangle$ yields the Haar snapshot $Y=(D+1)|\psi\rangle\langle\psi|-I_D$ from Equation (22), with $D=d_1d_2$. Conditional on the chosen basis, the measurement has D discrete outcomes.

2.5 Weighted Choi operators and regularization

Fix a channel $\mathcal{E}$ of Kraus rank r and a positive definite $\sigma_{\text{in}}\in\mathcal{D}(\mathcal{H}_{\mathrm{A}})$. The following definitions relate estimates of the joint output state $\mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)$ in Equation (26) to channel reconstruction and diamond-norm error.

Reconstruction map. For each $M\in\mathcal{L}(\mathcal{H}_{\mathrm{B}}\otimes\mathcal{H}_{\mathrm{A}})$, we use $\Phi_{\sigma_{\text{in}}}(M)$ to denote a linear map from $\mathcal{L}(\mathcal{H}_{\mathrm{A}})$ to $\mathcal{L}(\mathcal{H}_{\mathrm{B}})$ by

$$(\Phi_{\sigma_{\text{in}}}(M)\otimes \mathrm{id}_{\mathrm{A}})(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)=M.$$

Since $\sigma_{\text{in}}\succ 0$, this determines $\Phi_{\sigma_{\text{in}}}(M)$ uniquely. Equivalently, its Choi operator is

$$C_{\Phi_{\sigma_{\text{in}}}(M)}=(I_{\mathrm{B}}\otimes(\sigma_{\text{in}}^{\mathrm{T}})^{-1/2})M(I_{\mathrm{B}}\otimes(\sigma_{\text{in}}^{\mathrm{T}})^{-1/2}).\tag{27}$$

We can see that $\Phi_{\sigma_{\text{in}}}(\cdot)$ is also a linear map and $\Phi_{\sigma_{\text{in}}}(M)$ is not necessarily a quantum channel. In fact, $\Phi_{\sigma_{\text{in}}}(M)$ is a channel exactly when $M\succeq 0$ and $\operatorname{tr}_{\mathrm{B}}M=\sigma_{\text{in}}^{\mathrm{T}}$. In particular,

$$\Phi_{\sigma_{\text{in}}}(\mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|))=\mathcal{E}.$$

So we call $\Phi_{\sigma_{\text{in}}}(\cdot)$ the reconstruction map and $\Phi_{\sigma_{\text{in}}}(M)$ the reconstructed map from M using input state σ_{in}.

Taking the partial trace and transpose in Equation (27) gives

$$(\operatorname{tr}_{\mathrm{B}}C_{\Phi_{\sigma_{\text{in}}}(M)})^{\mathrm{T}}=\sigma_{\text{in}}^{-1/2}(\operatorname{tr}_{\mathrm{B}}M)^{\mathrm{T}}\sigma_{\text{in}}^{-1/2}.\tag{28}$$

For fixed σ_{in}, the map $M\mapsto(\operatorname{tr}_{\mathrm{B}}C_{\Phi_{\sigma_{\text{in}}}(M)})^{\mathrm{T}}$ is linear and positive. For Hermitian M,

$$\Phi_{\sigma_{\text{in}}}(M)^{\dagger}(I_{\mathrm{B}})=(\operatorname{tr}_{\mathrm{B}}C_{\Phi_{\sigma_{\text{in}}}(M)})^{\mathrm{T}},\tag{29}$$

and note that $\operatorname{tr}\Phi_{\sigma_{\text{in}}}(M)(\rho)=\operatorname{tr}(\rho\,\Phi_{\sigma_{\text{in}}}(M)^{\dagger}(I_{\mathrm{B}}))$ for every state ρ.

Lemma 2.11 (Weighted diamond-norm estimates)**.** *For $\rho\in\mathcal{D}(\mathcal{H}_{\mathrm{A}})$, define*

$$L_{\rho,\sigma_{\text{in}}}:=I_{\mathrm{B}}\otimes(\sqrt{\rho^{\mathrm{T}}}\,(\sigma_{\text{in}}^{\mathrm{T}})^{-1/2}),$$

$$L_{\rho,\sigma_{\rm in}}\mathcal{E}(|\sqrt{\sigma_{\rm in}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\rm in}}|)L_{\rho,\sigma_{\rm in}}^\dagger = \mathcal{E}(|\sqrt{\rho}\rangle\!\rangle\langle\!\langle\sqrt{\rho}|).$$

For every $M \in \mathcal{L}(\mathcal{H}_{\rm B}\otimes\mathcal{H}_{\rm A})$,

$$\operatorname{tr}(L_{\rho,\sigma_{\rm in}} M L_{\rho,\sigma_{\rm in}}^\dagger) = \operatorname{tr}(\rho\,(\operatorname{tr}_{\rm B} C_{\Phi_{\sigma_{\rm in}}(M)})^{\rm T}), \tag{30}$$

and for Hermitian M,

$$\|\Phi_{\sigma_{\rm in}}(M)\|_\diamond = \max_{\rho\in\mathcal{D}(\mathcal{H}_{\rm A})} \left\|L_{\rho,\sigma_{\rm in}} M L_{\rho,\sigma_{\rm in}}^\dagger\right\|_1. \tag{31}$$

Consequently:

(a) *if $G\succeq 0$ then $\|\Phi_{\sigma_{\rm in}}(G)\|_\diamond = \|\Phi_{\sigma_{\rm in}}(G)^\dagger(I_{\rm B})\|_\infty = \left\|\operatorname{tr}_{\rm B} C_{\Phi_{\sigma_{\rm in}}(G)}\right\|_\infty$;*

(b) *if $-aG \preceq M \preceq aG$ with $G\succeq 0$ and $a\geq 0$, then $\|\Phi_{\sigma_{\rm in}}(M)\|_\diamond \leq a\left\|\operatorname{tr}_{\rm B} C_{\Phi_{\sigma_{\rm in}}(G)}\right\|_\infty$;*

(c) *for $U,V\in\mathbb{C}^{D\times k}$, where $D=d_1d_2$,*

$$\left\|\Phi_{\sigma_{\rm in}}(UV^\dagger+VU^\dagger)\right\|_\diamond \leq 2\sqrt{\left\|\operatorname{tr}_{\rm B} C_{\Phi_{\sigma_{\rm in}}(UU^\dagger)}\right\|_\infty \left\|\operatorname{tr}_{\rm B} C_{\Phi_{\sigma_{\rm in}}(VV^\dagger)}\right\|_\infty}.$$

Proof. Cyclicity of the trace gives Equation (30); substituting Equation (27) into Equation (6) gives Equation (31). Part (a) follows by positivity, and part (b) by Proposition 2.2 (i). For part (c), write $L = L_{\rho,\sigma_{\rm in}}$ and use

$$\begin{aligned}\left\|L(UV^\dagger+VU^\dagger)L^\dagger\right\|_1 &\leq 2\,\|LU\|_2\,\|LV\|_2\\ &= 2\sqrt{\operatorname{tr}(\rho\,\Phi_{\sigma_{\rm in}}(UU^\dagger)^\dagger(I_{\rm B}))\ \operatorname{tr}(\rho\,\Phi_{\sigma_{\rm in}}(VV^\dagger)^\dagger(I_{\rm B}))}.\end{aligned}$$

Maximizing over ρ proves the bound. ☐

Regularization. For $s>0$, let $f_s(t):=t/(t+s)$ for $t\geq 0$. For $X\succeq 0$, define

$$f_s(X) = X(X+sI)^{-1}, \qquad X_s := Xf_s(X) = X^2(X+sI)^{-1} = X - sf_s(X). \tag{32}$$

Thus $0\preceq X_s\preceq X$, with eigenvalue map $\lambda\mapsto\lambda^2/(\lambda+s)$. For the joint output state $X := \mathcal{E}(|\sqrt{\sigma_{\rm in}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\rm in}}|)$, define the *regularized input-sensitivity operator*

$$H_s(\sigma_{\rm in}) := \Phi_{\sigma_{\rm in}}(f_s(X))^\dagger(I_{\rm B}). \tag{33}$$

Let $\mathcal{E}^c$ be the complementary channel of $\mathcal{E}$ associated with a choice of Kraus operators, as given in Equation (5):

$$\mathcal{E}^c(\rho) = \sum_{i,j}\operatorname{tr}(\rho K_j^\dagger K_i)\cdot|i\rangle\langle j|.$$

Its adjoint $(\mathcal{E}^c)^\dagger$ is thus $(\mathcal{E}^c)^\dagger(Q) = \sum_{i,j}\langle i|Q|j\rangle\cdot K_i^\dagger K_j$.

Lemma 2.12 (Complementary-channel formula)**.** *For $s>0$,*

$$H_s(\sigma_{\rm in}) = (\mathcal{E}^c)^\dagger((\mathcal{E}^c(\sigma_{\rm in})+sI_r)^{-1}). \tag{34}$$

Hence, for every $\rho\in\mathcal{D}(\mathcal{H}_{\rm A})$,

$$\operatorname{tr}(\rho H_s(\sigma_{\rm in})) = \operatorname{tr}(\mathcal{E}^c(\rho)(\mathcal{E}^c(\sigma_{\rm in})+sI_r)^{-1}). \tag{35}$$

Moreover,

$$\operatorname{tr}(\sigma_{\rm in}H_s(\sigma_{\rm in})) = \operatorname{tr} f_s(\mathcal{E}(|\sqrt{\sigma_{\rm in}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\rm in}}|)) \leq r, \qquad 0\preceq H_s(\sigma_{\rm in})\preceq s^{-1}I_{\rm A}.$$

Proof. Let $B_{\sigma_{\text{in}}} : \mathbb{C}^r \to \mathcal{H}_{\text{B}} \otimes \mathcal{H}_{\text{A}}$ have columns $|K_i\sqrt{\sigma_{\text{in}}}\rangle\!\rangle$. Then

$$\mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|) = B_{\sigma_{\text{in}}} B_{\sigma_{\text{in}}}^\dagger, \qquad B_{\sigma_{\text{in}}}^\dagger B_{\sigma_{\text{in}}} = \mathcal{E}^c(\sigma_{\text{in}})^{\mathrm{T}}.$$

For $Q := (\mathcal{E}^c(\sigma_{\text{in}})^{\mathrm{T}} + sI_r)^{-1}$ and $X := \mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)$, the resolvent identity and Equations (3), (28) and (29) give

$$\begin{aligned} f_s(X) &= B_{\sigma_{\text{in}}} Q B_{\sigma_{\text{in}}}^\dagger, \\ H_s(\sigma_{\text{in}}) &= \sum_{i,j} Q_{ij} K_j^\dagger K_i = (\mathcal{E}^c)^\dagger(Q^{\mathrm{T}}). \end{aligned}$$

This proves Equation (34); adjoint duality gives Equation (35). Equation (30) with $\rho = \sigma_{\text{in}}$ gives $\operatorname{tr}(\sigma_{\text{in}} H_s(\sigma_{\text{in}})) = \operatorname{tr} f_s(\mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)) \le \operatorname{rank} \mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|) \le r$. Since $0 \preceq (\mathcal{E}^c(\sigma_{\text{in}}) + sI_r)^{-1} \preceq s^{-1} I_r$, positivity and unitality of $(\mathcal{E}^c)^\dagger$ imply $0 \preceq H_s(\sigma_{\text{in}}) \preceq s^{-1} I_{\text{A}}$.

□

Lemma 2.13 (Regularized channel decomposition). *Let $X := \mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)$, and reconstruct its regularization X_s (see Equation (32)) as the map*

$$\mathcal{E}_s := \Phi_{\sigma_{\text{in}}}(X_s).$$

Then $\mathcal{E}_s$ and $\mathcal{E} - \mathcal{E}_s$ are completely positive and trace-nonincreasing, with

$$\mathcal{E} - \mathcal{E}_s = s\Phi_{\sigma_{\text{in}}}(f_s(X)).$$

The regularization bias and output trace are determined by

$$\begin{aligned} \|\mathcal{E} - \mathcal{E}_s\|_\diamond &= s\,\|H_s(\sigma_{\text{in}})\|_\infty\,, \\ \mathcal{E}_s^\dagger(I_{\text{B}}) &= I_{\text{A}} - sH_s(\sigma_{\text{in}}). \end{aligned}$$

Proof. By Equation (32), $X_s \succeq 0$ and $X - X_s = sf_s(X) \succeq 0$. Reconstruction gives complete positivity and the displayed decomposition. The norm identity follows from Lemma 2.11(a) and Equation (33); the adjoint identity follows from the decomposition, Equation (33), and $\mathcal{E}^\dagger(I_{\text{B}}) = I_{\text{A}}$. Each completely positive summand is trace-nonincreasing because their sum is $\mathcal{E}$. □

The Choi-support projection. Let $P_{\mathcal{E}}$ be the orthogonal projection onto $\operatorname{supp} C_{\mathcal{E}}$ and define

$$\kappa(\mathcal{E}) = d_1 \|\operatorname{tr}_{\text{B}} P_{\mathcal{E}}\|_\infty. \tag{36}$$

Lemma 2.14 (Marginal of the Choi-support projection). *Let $\mathcal{E}$ have Kraus rank r, with $\kappa(\mathcal{E})$ defined in Equation (36), and let $\lambda_{\min}^+(C_{\mathcal{E}})$ be the smallest non-zero Choi eigenvalue of its Choi operator. Then*

$$r \le \kappa(\mathcal{E}) \le \min\left\{\frac{d_1}{\lambda_{\min}^+(C_{\mathcal{E}})}, d_1 \min(r, d_2)\right\}. \tag{37}$$

For every positive definite $\sigma_{\text{in}} \in \mathcal{D}(\mathcal{H}_{\text{A}})$, let $P_{\mathcal{E},\sigma_{\text{in}}}$ be the orthogonal projection onto the support of the joint output state $\mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)$. With H_s defined in Equation (33), we have

$$\lim_{s\downarrow 0} H_s(\sigma_{\text{in}}) = \Phi_{\sigma_{\text{in}}}(P_{\mathcal{E},\sigma_{\text{in}}})^\dagger(I_{\text{B}}). \tag{38}$$

In particular, $\lim_{s\downarrow 0} \|H_s(I/d_1)\|_\infty = \kappa(\mathcal{E})$.

Algorithm 1: Gapped-case incoherent quantum channel tomography

Input: $d_1, d_2, r, \epsilon, \delta, c_1$, such that $\lambda^+_{\min}(C_{\mathcal{E}}) \geq \frac{c_1 d_1}{r}$

Output: A channel estimate $\widehat{\mathcal{E}}$

$D \leftarrow d_1 d_2$; $n \leftarrow \left\lceil c_{\mathrm{ub}} \frac{r^2}{c_1^2 \epsilon^2} (D + \log \frac{4}{\delta}) (1 + \log(8/\delta))^2 \right\rceil$;

for $t \leftarrow 1$ **to** n **do**

  Prepare $\frac{1}{\sqrt{d_1}} \sum_{i=1}^{d_1} |i\rangle_{\mathrm{A}} \otimes |i\rangle_{\mathrm{A}}$ and apply $\mathcal{E} \otimes \mathrm{id}$;

  Measure BA in an independent Haar-random basis, obtaining outcome $|\psi_t\rangle$;

$\widehat{X} \leftarrow \frac{1}{n} \sum_{t=1}^{n} ((D+1)|\psi_t\rangle\langle\psi_t| - I_D)$;

$\widetilde{X} = \arg\min_{Y \succeq 0, \operatorname{tr} Y = 1} \left\| Y - \widehat{X} \right\|_2$;

$\widehat{\mathcal{E}} \in \arg\min_{\mathcal{F} \text{ a channel}} \left\| \mathcal{F} - \Phi_{I/d_1}(\widetilde{X}) \right\|_\diamond$;

return $\widehat{\mathcal{E}}$;

Proof. The positive operator $\operatorname{tr}_{\mathrm{B}} P_{\mathcal{E}}$ has trace r and dimension d_1, so $d_1 \| \operatorname{tr}_{\mathrm{B}} P_{\mathcal{E}} \|_\infty \geq r$. If $\lambda = \lambda^+_{\min}(C_{\mathcal{E}})$, then $\lambda P_{\mathcal{E}} \preceq C_{\mathcal{E}}$ implies $\lambda \operatorname{tr}_{\mathrm{B}} P_{\mathcal{E}} \preceq I_{\mathrm{A}}$, giving $\kappa(\mathcal{E}) \leq d_1/\lambda$. For a unit vector $|v\rangle \in \mathcal{H}_{\mathrm{A}}$,

$$\langle v| \operatorname{tr}_{\mathrm{B}} P_{\mathcal{E}} |v\rangle = \operatorname{tr}(P_{\mathcal{E}}(I_{\mathrm{B}} \otimes |v\rangle\langle v|) P_{\mathcal{E}}) \leq \min(r, d_2),$$

because $0 \preceq P_{\mathcal{E}}(I_{\mathrm{B}} \otimes |v\rangle\langle v|) P_{\mathcal{E}} \preceq I$, and its rank is at most $\min(r, d_2)$. This proves the remaining bound. Finally, $f_s(\mathcal{E}(|\sqrt{\sigma_{\mathrm{in}}}\rangle\rangle\langle\langle\sqrt{\sigma_{\mathrm{in}}}|)) \to P_{\mathcal{E},\sigma_{\mathrm{in}}}$, so Equation (38) follows from Equation (33) and continuity of the reconstruction in Equation (27). At $\sigma_{\mathrm{in}} = I_{\mathrm{A}}/d_1$, $C_{\Phi_{I/d_1}(M)} = d_1 M$ and $P_{\mathcal{E},\sigma_{\mathrm{in}}} = P_{\mathcal{E}}$, so the final claim follows from Equation (29) and invariance of the operator norm under transposition. □

In particular, a channel of Kraus rank at most r satisfying $\lambda^+_{\min}(C_{\mathcal{E}}) \geq c d_1/r$ for some $c > 0$ has $\kappa(\mathcal{E}) \leq r/c$. This spectral condition is used in Theorem 3.2.

3 Upper bounds

In this section, we provide the algorithms that achieve our upper bounds in increasing order of difficulty and generality. In Section 3.1, we first treat the gapped case and show that a fixed maximally entangled input state is sufficient and adaptivity is not needed. The complexity we obtain depends on the parameter $\kappa(\mathcal{E})$ and is optimal for channels with $\Omega(d_1/r)$-gapped Choi spectrum (which implies $\kappa(\mathcal{E}) = O(r)$). Then, in Section 3.2, we treat the general (non-gapped) setting where $\kappa(\mathcal{E})$ can be large. The near-optimal upper bound is achieved by an *adaptive* algorithm that chooses the input states adaptively. Their technical lemmas appear in Section 3.3.

3.1 A fixed maximally entangled input

The protocol of [SSKKG22] prepares the maximally entangled state, sends it through the unknown channel, measures the output in a random basis (see Algorithm 1 for the detailed algorithm). We show that this protocol achieves optimal query complexity when $\kappa(\mathcal{E})$, defined in Equation (36), is $O(r)$, which by (37) happens when the smallest non-zero Choi eigenvalue is at least $\Omega(d_1/r)$.

Fixed-input estimator. Prepare the same maximally entangled input $d_1^{-1/2}|I\rangle\!\rangle$ in every experiment. Its input marginal is I_{A}/d_1, so the measured state is $\mathcal{E}(|I/\sqrt{d_1}\rangle\!\rangle\langle\!\langle I/\sqrt{d_1}|) = C_{\mathcal{E}}/d_1$. Measure the output Choi state by Haar random bases (see Definition 2.7), obtaining outcomes $\{|\psi_1\rangle, \dots, |\psi_n\rangle\}$. Using the Haar snapshots (see Definition 2.8), we define $\widehat{X} = \frac{1}{n}\sum_{t=1}^{n}(D+1)|\psi_t\rangle\langle\psi_t| - I_D$, then project it onto $\mathcal{D}(\mathcal{H}_{\mathrm{B}} \otimes \mathcal{H}_{\mathrm{A}})$ in Hilbert–Schmidt norm to obtain $\widetilde{X}$, and output

$$\widehat{\mathcal{E}} \in \arg \min_{\mathcal{F} \text{ a channel}} \|\mathcal{F} - \Phi_{I/d_1}(\widetilde{X})\|_{\diamond}. \tag{39}$$

Theorem 3.1 (Upper bound for a fixed maximally entangled input)**.** *There is an absolute constant c_{ub} such that, for every channel $\mathcal{E}$ with input dimension d_1 and output dimension d_2, the fixed-input estimator satisfies* $\Pr\{\|\widehat{\mathcal{E}} - \mathcal{E}\|_{\diamond} > \epsilon\} \le \delta$ *whenever*

$$n \ge c_{\mathrm{ub}} \frac{\kappa(\mathcal{E})^2}{\epsilon^2} \left(D + \log \frac{4}{\delta}\right) (1 + \log(8/\delta))^2. \tag{40}$$

Proof. Put

$$\theta = c_{\mathrm{H}} \left(\sqrt{\frac{D + \log(4/\delta)}{n}} + \frac{D + \log(4/\delta)}{n} \right).$$

By Lemma 2.10(iii), and Lemma 3.6, with probability at least $1 - \delta/2$,

$$\left\| \widehat{X} - \mathcal{E}\left(|I/\sqrt{d_1}\rangle\!\rangle\langle\!\langle I/\sqrt{d_1}|\right) \right\|_{\infty} \le \theta, \qquad \left\| \widetilde{X} - \mathcal{E}\left(|I/\sqrt{d_1}\rangle\!\rangle\langle\!\langle I/\sqrt{d_1}|\right) \right\|_{\infty} \le 2\theta.$$

The sample-size assumption ensures $\theta \le 2c_{\mathrm{H}}\sqrt{(D + \log(4/\delta))/n}$ and $2\theta < 1$, after increasing c_{ub}. By Lemma 2.10(iv), the distribution of $\widehat{X}$ is invariant under conjugation by any unitary fixing $\mathcal{E}(|I/\sqrt{d_1}\rangle\!\rangle\langle\!\langle I/\sqrt{d_1}|)$. The projection onto states commutes with unitary conjugation by Lemma 3.6. Consequently, Lemma 3.11, with $\eta = 2\theta$ and $\zeta = \delta/2$, gives

$$\|\Phi_{I/d_1}(\widetilde{X}) - \mathcal{E}\|_{\diamond} \le 2c_{\mathrm{conv}}\kappa(\mathcal{E})\theta(1 + \log(8/\delta))$$

with probability at least $1 - \delta$. The final projection onto channels increases this bound by at most a factor of two: by the definition of $\widehat{\mathcal{E}}$ in Equation (39), $\|\widehat{\mathcal{E}} - \Phi_{I/d_1}(\widetilde{X})\|_{\diamond} \le \|\mathcal{E} - \Phi_{I/d_1}(\widetilde{X})\|_{\diamond}$ so by the triangle inequality $\|\mathcal{E} - \widehat{\mathcal{E}}\|_{\diamond} \le 2\|\mathcal{E} - \Phi_{I/d_1}(\widetilde{X})\|_{\diamond}$. Substituting the preceding bound on θ proves Equation (40). □

Theorem 3.2 (Non-zero Choi eigenvalues bounded below)**.** *There is an absolute constant c_{ub} such that, for every $\mathcal{E} \in \mathsf{QChan}^r_{d_1,d_2}$ with non-zero Choi eigenvalues $\ge cd_1/r$ for some $c > 0$, the fixed-input estimator satisfies* $\Pr\{\|\widehat{\mathcal{E}} - \mathcal{E}\|_{\diamond} > \epsilon\} \le \delta$ *whenever*

$$n \ge \frac{c_{\mathrm{ub}}}{c^2} \cdot \frac{r^2}{\epsilon^2} \left(D + \log \frac{4}{\delta}\right) (1 + \log(8/\delta))^2.$$

Proof. The spectral assumption and Equation (37) give $\kappa(\mathcal{E}) \le r/c$. Then, we apply Theorem 3.1. □

In the next section, we search adaptively for input states that spread the weight of the output evenly.

Algorithm 2: Two-batch estimator

Input: $\mathcal{E}, \sigma \succ 0, s > 0, n, d_1, d_2, r$

Output: $\widehat{Z}$ and $\widehat{H}$

$D \leftarrow d_1 d_2$;

for $b \in \{0,1\}$ **do**

  for $t \leftarrow 1$ **to** n **do**

    Prepare $|\sqrt{\sigma}\rangle\rangle = \sum_{i,j=1}^{d_1} (\sqrt{\sigma})_{i,j} |i\rangle_{\mathrm{A}} \otimes |j\rangle_{\mathrm{A}}$ and apply $\mathcal{E} \otimes \mathrm{id}$;

    Measure BA in an independent Haar-random basis, obtaining outcome $|\psi_{b,t}\rangle$;

  $\widehat{X}_b \leftarrow \frac{1}{n} \sum_{t=1}^{n} \left((D+1)|\psi_{b,t}\rangle\langle\psi_{b,t}| - I_D\right)$;

$S \leftarrow$ positive part of $\widehat{X}_0$ restricted to its r largest positive eigenvalues;

$P_S \leftarrow$ projector onto $\operatorname{ran} S$; $R \leftarrow (S + sI_D)^{-1} P_S$;

$\widehat{Z} = \widehat{X}_1 R \widehat{X}_1$; $\widehat{H} \leftarrow \frac{1}{s}\left(I_{\mathrm{A}} - \Phi_\sigma(\widehat{Z})^\dagger(I_{\mathrm{B}})\right)$;

return $(\widehat{Z}, \widehat{H})$;

3.2 Adaptive upper bound

We remove the gapped hypothesis of Theorem 3.2 at the cost of a logarithmic overhead in the query complexity. To this end, we rely on a two-batch estimator that returns an estimate of the channel and a multiplicative estimate of $H_s(\sigma_{\mathrm{in}}) + rI_{\mathrm{A}}$ whose operator norm plays the role of $\kappa(\mathcal{E})$ at a general input state (rather than the maximally entangled state). An update of the input marginal then chooses an input at which $\|H_s(\sigma_{\mathrm{in}})\|_\infty = O(r)$. We state the procedures first and use their technical guarantees to prove Theorem 3.4. The two-batch estimator is detailed in Algorithm 2 and the adaptive incoherent channel learning algorithm is detailed in Algorithm 3.

Definition 3.3 (Two-batch estimator)**.** Fix a channel $\mathcal{E} \in \mathsf{QChan}^r_{d_1,d_2}$, a state $\sigma_{\mathrm{in}} \in \mathcal{D}(\mathcal{H}_{\mathrm{A}})$ with $\sigma_{\mathrm{in}} \succeq I_{\mathrm{A}}/(2d_1)$, a regularization parameter $s > 0$, and a batch size n. From two independent batches of n Haar snapshots at input σ_{in}, form $\widehat{X}_0, \widehat{X}_1$ using Equation (23). Let S retain the r largest positive eigenvalues of $\widehat{X}_0$ and set its other eigenvalues to zero. With P_S the projection onto $\operatorname{ran} S$, set

$$R = (S + sI_D)^{-1} P_S,$$

and return

$$\widehat{Z} = \widehat{X}_1 R \widehat{X}_1, \qquad \widehat{H} = \frac{I_{\mathrm{A}} - \Phi_{\sigma_{\mathrm{in}}}(\widehat{Z})^\dagger(I_{\mathrm{B}})}{s}. \tag{41}$$

Here the reconstruction is given by Equation (27). The operator $\widehat{Z}$ is positive semidefinite and the procedure uses $2n$ queries. The first batch chooses R independently of the second batch; this independence is used in Proposition 3.15.

Input update. For $\rho \in \mathcal{D}(\mathcal{H}_{\mathrm{A}})$ define

$$\sigma_{\mathrm{in}}(\rho) = \frac{\rho}{2} + \frac{I_{\mathrm{A}}}{2d_1}, \qquad A(\rho) = H_s(\sigma_{\mathrm{in}}(\rho)) + rI_{\mathrm{A}}. \tag{42}$$

The mixture guarantees that every queried input is bounded below by $I_{\mathrm{A}}/(2d_1)$. Given a positive definite estimate A_j of $A(\rho_j)$, we use the matrix exponentiated-gradient update [TRW05, AK16]

Algorithm 3: Adaptive incoherent quantum channel tomography

Input: $d_1, d_2, r, \epsilon, \delta$

Output: A channel estimate $\widehat{\mathcal{E}}$

$D \leftarrow d_1 d_2$; $\quad s \leftarrow \frac{\epsilon}{64r}$; $\quad T \leftarrow \max\{1, \lceil \log_2 d_1 \rceil\}$;

$\delta' \leftarrow \frac{\delta}{T+1}$; $\quad b \leftarrow D + \left(1 + \frac{d_1}{r}\right) \log \frac{6}{\delta'}$; $\quad n \leftarrow \lceil K \frac{b}{s^2} \rceil$;

$\rho_1 \leftarrow \frac{1}{d_1} I_{\mathrm{A}}$;

for $j \leftarrow 1$ **to** T **do**

  $\sigma_j \leftarrow \frac{1}{2}\rho_j + \frac{1}{2d_1} I_{\mathrm{A}}$;

  $(\widehat{Z}_j, \widehat{H}_j) \leftarrow$ TwoBatch$(\mathcal{E}, \sigma_j, s, n, d_1, d_2, r)$ using fresh copies;

  $A_j \leftarrow \widehat{H}_j + r I_{\mathrm{A}}$ with every eigenvalue below $r/2$ replaced by $r/2$;

  $z_j = \operatorname{tr} \exp(\log \rho_j + \log A_j)$, $\qquad \rho_{j+1} = \frac{1}{z_j} \exp(\log \rho_j + \log A_j)$;

$\bar{\sigma} \leftarrow \frac{1}{T} \sum_{j=1}^{T} \sigma_j$;

$(\widehat{Z}, \widehat{H}) \leftarrow$ TwoBatch$(\mathcal{E}, \bar{\sigma}, s, n, d_1, d_2, r)$ using fresh copies;

$\widehat{\mathcal{E}} \in \arg\min_{\mathcal{F} \text{ a channel}} \|\mathcal{F} - \Phi_{\bar{\sigma}}(\widehat{Z})\|_{\diamond}$.;

return $\widehat{\mathcal{E}}$;

update

$$z_j = \operatorname{tr} \exp(\log \rho_j + \log A_j), \qquad \rho_{j+1} = \frac{\exp(\log \rho_j + \log A_j)}{z_j}.$$

Theorem 3.4. *There is an absolute constant c such that, for every $\mathcal{E} \in \mathsf{QChan}^r_{d_1,d_2}$, the adaptive incoherent protocol in Algorithm 3 outputs a channel $\widehat{\mathcal{E}}$ satisfying*

$$\Pr\{\|\widehat{\mathcal{E}} - \mathcal{E}\|_{\diamond} > \epsilon\} \leq \delta$$

using at most

$$c \frac{r^2}{\epsilon^2} \log(2d_1) \left[D + \left(1 + \frac{d_1}{r}\right) \log\left(\frac{18 \log(2d_1)}{\delta}\right) \right] \tag{43}$$

queries. The protocol uses $O(\log(2d_1))$ adaptive updates of the input marginal and an independent Haar-random basis measurement after each channel use.

Proof. The choice $s = \epsilon/(64r)$ satisfies $s \leq \min\{1, \sqrt{d_1/r}\}$, as required by Proposition 3.15. We condition on the transcript preceding each two-batch estimator. It fixes the input, and both batches are fresh. By Proposition 3.15, with conditional probability at least $1 - \delta'$, the estimate satisfies

$$\frac{3}{4} A(\rho_j) \preceq \widehat{H}_j + r I_{\mathrm{A}} \preceq \frac{5}{4} A(\rho_j).$$

The same proposition applies to the final estimator. A conditional union bound shows that all $T+1$ guarantees hold simultaneously with probability at least $1 - \delta$. From now on, we work on this event.

Input selection. Since $A(\rho_j) \succeq r I_{\mathrm{A}}$ and the lower bound in Equation (74) gives $\widehat{H}_j + r I_{\mathrm{A}} \succeq \frac{3}{4} A(\rho_j) \succeq \frac{3r}{4} I_{\mathrm{A}}$, no eigenvalue is replaced by the clipping step. Hence $A_j = \widehat{H}_j + r I_{\mathrm{A}}$. Therefore, Proposition 3.21 gives

$$H_s(\bar{\sigma}_{\mathrm{in}}) \preceq 9 r I_{\mathrm{A}}. \tag{44}$$

Estimation error. Conditional on the transcript from the input-selection rounds, the final input is fixed. Combining Equation (44) with Proposition 3.15 yields $\|\Phi_{\bar{\sigma}_{\rm in}}(\widehat{Z}) - \mathcal{E}\|_\diamond \le 11rs$. The true channel is feasible in the final minimization, so

$$\|\widehat{\mathcal{E}} - \mathcal{E}\|_\diamond \le 2\|\Phi_{\bar{\sigma}_{\rm in}}(\widehat{Z}) - \mathcal{E}\|_\diamond \le 22rs = \frac{22}{64}\epsilon < \epsilon.$$

Query complexity. The algorithm uses $2n(T+1)$ queries. Since $T+1 = O(\log(2d_1))$ and $\delta' = \delta/(T+1)$, substituting $s = \epsilon/(64r)$ and $T+1 \le 3\log(2d_1)$ gives Equation (43).

□

3.3 Technical lemmas

Building on the identities proved in Section 2, we establish the error bounds for a fixed maximally entangled input and the guarantees for regularized estimation and input selection.

3.3.1 Error estimates for a fixed maximally entangled input

Lemma 3.5 ([GKKT20]). *Let ρ be a state of rank r on a D-dimensional space and σ any state. Then*

$$\|\sigma - \rho\|_1 \le 2r \|\sigma - \rho\|_\infty .$$

Proof. Set $M = \sigma - \rho$. If M had $r+1$ negative eigenvalues, the span V of the corresponding eigenvectors would meet $\ker\rho$, of dimension $D - r$, in a non-zero vector $|v\rangle$; but $\langle v|M|v\rangle < 0$ and $\langle v|M|v\rangle = \langle v|\sigma|v\rangle \ge 0$, a contradiction. So M has at most r negative eigenvalues, and since $\operatorname{tr} M = 0$ its positive and negative parts have equal trace; hence $\|M\|_1 = 2\operatorname{tr} M_- \le 2r\|M\|_\infty$. □

Lemma 3.6 (Hilbert–Schmidt projection onto states [GKKT20]). *Let $\widehat{X}$ be Hermitian with $\operatorname{tr}\widehat{X} = 1$ and let $\widetilde{X} = \arg\min_{\sigma\in\mathcal{D}(\mathbb{C}^D)} \|\sigma - \widehat{X}\|_2$. Then $\widetilde{X}$ is obtained by keeping the eigenvectors of $\widehat{X}$ and replacing its eigenvalues x_i by $(x_i - q)_+$, where $q \in \mathbb{R}$ is the unique value with $\sum_i (x_i - q)_+ = 1$, and $a_+ = \max\{a, 0\}$. Consequently*

$$\|\widetilde{X} - X\|_\infty \le 2\|\widehat{X} - X\|_\infty \quad \textit{for every state } X,$$

and the map $\widehat{X} \mapsto \widetilde{X}$ commutes with unitary conjugation: $\widetilde{U\widehat{X}U^\dagger} = U\widetilde{X}U^\dagger$ for every unitary U.

Proof. The eigenvalue formula for the Hilbert–Schmidt projection onto density operators is given in [SGS12] and [GKKT20, Sec. 4.2]. Since this formula changes only the eigenvalues, the projection commutes with unitary conjugation.

If $\widehat{X} = X$, the claim is immediate. Otherwise, put $\theta = \|\widehat{X} - X\|_\infty$ and note that $q \mapsto \operatorname{tr}(\widehat{X} - qI)_+$ is nonincreasing. Since $\widehat{X} \preceq X + \theta I$ and $Z \mapsto \operatorname{tr} Z_+ = \max_{0\preceq\Pi\preceq I}\operatorname{tr}(Z\Pi)$ is Löwner monotone, $\operatorname{tr}(\widehat{X} - \theta I)_+ \le \operatorname{tr} X_+ = 1$, so $q \le \theta$. If we had $q \le -\theta$ then $\widehat{X} - qI \succeq \widehat{X} + \theta I \succeq X \succeq 0$ and hence $\operatorname{tr}(\widehat{X} - qI)_+ = \operatorname{tr}\widehat{X} - qD = 1 - qD \ge 1 + \theta D > 1$, a contradiction; so $|q| \le \theta$. Finally, for each eigenvalue, $|(x_i - q)_+ - x_i|$ equals $|q|$ if $x_i \ge q$ and $|x_i| \le \max(\theta, |q|)$ otherwise, hence is at most θ in both cases. Thus $\|\widetilde{X} - \widehat{X}\|_\infty \le \theta$ and the triangle inequality finishes the proof. □

Gaussian estimates. A standard complex Gaussian scalar has independent real and imaginary parts, each distributed as $N(0, 1/2)$. The Gaussian matrix estimates in this subsection use this normalization.

Fact 3.7 (Real Gaussian quadratic forms [LM00, Lem. 1])**.** *Let $\xi_1, \xi_2, \ldots$ be i.i.d. real standard Gaussians and $a_i \geq 0$ with $\sum_i a_i < \infty$. Then for $u \geq 0$,*

$$\Pr\Big\{\sum_i a_i \xi_i^2 \geq \sum_i a_i + 2\|a\|_2\sqrt{u} + 2\|a\|_\infty u\Big\} \leq e^{-u}, \quad \Pr\Big\{\sum_i a_i \xi_i^2 \leq \sum_i a_i - 2\|a\|_2\sqrt{u}\Big\} \leq e^{-u}.$$

Corollary 3.8 (Complex Gaussian quadratic forms)**.** *Let $T \succeq 0$ act on $\mathbb{C}^N$ and let $|g_1\rangle, \ldots, |g_m\rangle$ be i.i.d. standard complex Gaussian vectors in $\mathbb{C}^N$: their coordinates have independent real and imaginary parts distributed as $N(0, 1/2)$. Then for every $u \geq 0$,*

$$\Pr\Big\{\sum_{j=1}^m \langle g_j|T|g_j\rangle \geq m \operatorname{tr} T + \sqrt{2m \operatorname{tr}(T^2)\, u} + \|T\|_\infty\, u\Big\} \leq e^{-u}.$$

Proof. Diagonalize $T = \sum_k \lambda_k |u_k\rangle\langle u_k|$. Then $\sum_j \langle g_j|T|g_j\rangle = \sum_k \lambda_k \sum_j |\langle u_k|g_j\rangle|^2$, and the $|\langle u_k|g_j\rangle|^2$ are i.i.d. mean-one exponentials, i.e. $\frac{1}{2}\chi_2^2$. Hence the sum equals $\sum_i a_i \xi_i^2$ with each λ_k appearing $2m$ times with weight $a = \lambda_k/2$. Then $\sum_i a_i = m \operatorname{tr} T$, $\|a\|_2^2 = \frac{m}{2}\operatorname{tr}(T^2)$ and $\|a\|_\infty = \frac{1}{2}\|T\|_\infty$, and Fact 3.7 gives the claim. ☐

Lemma 3.9 (Extreme singular values of complex Gaussian matrices)**.** *Let G be an $N \times m$ matrix whose entries have independent real and imaginary parts distributed as $N(0, 1/2)$. Write $s_{\max}(G)$ and $s_{\min}(G)$ for its largest and smallest singular values. Then for every $t \geq 0$,*

$$\Pr\{s_{\max}(G) \geq 4(\sqrt{N} + \sqrt{m} + t)\} \leq e^{-t^2}, \tag{45}$$

and if $N \geq 125\, m$,

$$\Pr\{s_{\min}(G) \leq \tfrac{1}{5}\sqrt{N}\} \leq 2e^{-N/16}.$$

(Sharper constants are classical; see Davidson and Szarek [DS01, Thm. II.13] and Vershynin [Ver18, Thm. 4.6.1]. The following proof gives the constants used here.)

Proof. For a fixed unit $|x\rangle \in \mathbb{C}^m$ the vector $G|x\rangle$ is standard complex Gaussian in $\mathbb{C}^N$, so $\|G|x\rangle\|_2^2 = \frac{1}{2}\chi_{2N}^2$, i.e. it is $\sum_i a_i \xi_i^2$ with $2N$ weights $a_i = \frac{1}{2}$. Thus $\sum_i a_i = N$, $\|a\|_2 = \sqrt{N/2}$, $\|a\|_\infty = \frac{1}{2}$, and Fact 3.7 gives, for every $u \geq 0$,

$$\Pr\{\|G|x\rangle\|_2^2 \geq N + \sqrt{2Nu} + u\} \leq e^{-u}, \qquad \Pr\{\|G|x\rangle\|_2^2 \leq N - \sqrt{2Nu}\} \leq e^{-u}. \tag{46}$$

Largest singular value. By Lemma 2.4, choose a $\frac{1}{2}$-covering net $\mathcal{S}$ of the unit sphere of $\mathbb{C}^m$ with $|\mathcal{S}| \leq 5^{2m} \leq e^{3.3m}$. Equation (15) gives $s_{\max}(G) \leq 2\max_{|x\rangle\in\mathcal{S}} \|G|x\rangle\|_2$. Taking $u = 3.3m + t^2$ in the first bound of Equation (46) and a union bound, $\max_{|x\rangle\in\mathcal{S}} \|G|x\rangle\|_2^2 \leq N + \sqrt{2N(3.3m + t^2)} + 3.3m + t^2 \leq (\sqrt{N} + 1.9\sqrt{m} + t)^2$ except with probability e^{-t^2}, which gives Equation (45).

Smallest singular value. By Lemma 2.4, choose a $\frac{1}{24}$-covering net $\mathcal{S}'$ of the unit sphere of $\mathbb{C}^m$ with $|\mathcal{S}'| \leq 49^{2m} \leq e^{7.8m}$. Equation (19) gives $s_{\min}(G) \geq \min_{|x\rangle\in\mathcal{S}'} \|G|x\rangle\|_2 - \frac{1}{24}s_{\max}(G)$. Using Equation (46) with $u = N/8$ and a union bound gives $\min_{|x\rangle\in\mathcal{S}'} \|G|x\rangle\|_2^2 \geq N/2$ except with probability $e^{7.8m - N/8} \leq e^{-N/16}$ when $N \geq 125m$. By Equation (45) with $t = \sqrt{N}$ and $m \leq N$, $s_{\max}(G) \leq 12\sqrt{N}$ except with probability e^{-N}. On the intersection, $s_{\min}(G) \geq \sqrt{N/2} - \frac{12}{24}\sqrt{N} \geq \frac{1}{5}\sqrt{N}$. ☐

Partial traces and diamond-norm error.

Lemma 3.10 (Partial trace after a Haar-random conjugation)**.** *There is an absolute constant c with the following property. Let $\eta > 0$ and $S \subseteq \mathcal{H}_{\mathrm{B}} \otimes \mathcal{H}_{\mathrm{A}}$ have $\dim S = r$, and let $T \succeq 0$ be supported on $S^{\perp}$ with*

$$\|T\|_{\infty} \leq \eta, \qquad \operatorname{tr} T \leq 2r\eta, \qquad rd_2 \geq d_1,$$

and let V be Haar distributed on the unitary group of $S^{\perp}$. Then for every $0 < \zeta < 1$, with probability at least $1 - \zeta$,

$$d_1 \left\|\operatorname{tr}_{\mathrm{B}}(VTV^{\dagger})\right\|_{\infty} \leq cr\eta\,\ell, \qquad \ell := 1 + \log(2/\zeta). \tag{47}$$

Proof. Write $Z = VTV^{\dagger}$, which is again positive, supported on $S^{\perp}$, with the same norm and trace as T. We can easily see two deterministic bounds:

$$\text{(D1)} \quad d_1 \left\|\operatorname{tr}_{\mathrm{B}} Z\right\|_{\infty} \leq d_1 d_2 \left\|Z\right\|_{\infty} \leq D\eta, \qquad \text{(D2)} \quad d_1 \left\|\operatorname{tr}_{\mathrm{B}} Z\right\|_{\infty} \leq d_1 \operatorname{tr} Z \leq 2d_1 r\eta,$$

using $\langle\varphi| \operatorname{tr}_{\mathrm{B}} Z|\varphi\rangle = \operatorname{tr}(Z(I_{\mathrm{B}} \otimes |\varphi\rangle\langle\varphi|)) \leq \min\{d_2 \left\|Z\right\|_{\infty}, \operatorname{tr} Z\}$. Let K be a large absolute constant, chosen at the end. If $D \leq Kr\ell$ then (D1) already gives Equation (47) with $c = K$; if $d_1 \leq K$ then (D2) gives it with $c = 2K$. So assume from now on

$$D > Kr\ell \qquad \text{and} \qquad d_1 > K. \tag{48}$$

Set $D' = \dim S^{\perp} = D - r$. From Equation (48), $r < D/K$, so $D' > D/2$ for $K \geq 2$; from $d_1 \leq rd_2$ we get $D \leq rd_2^2$, hence $d_2^2 > K\ell \geq K$ and $d_2 > \sqrt{K}$; and $D' > D/2 = d_1 d_2/2 \geq K d_2/2 \geq 125\, d_2$ for $K \geq 250$.

Step 1 (covering net). By Lemma 2.4, choose a $\frac{1}{4}$-covering net $\mathcal{N}$ of the unit sphere of $\mathcal{H}_{\mathrm{A}}$ with $|\mathcal{N}| \leq 9^{2d_1}$. Applying Equation (18) to the positive operator $\operatorname{tr}_{\mathrm{B}} Z$ gives

$$\left\|\operatorname{tr}_{\mathrm{B}} Z\right\|_{\infty} \leq 2 \max_{|\varphi\rangle \in \mathcal{N}} \langle\varphi| \operatorname{tr}_{\mathrm{B}} Z|\varphi\rangle.$$

Thus it suffices to bound this quadratic form for each fixed $|\varphi\rangle$ and take a union bound. Set

$$u := \log(3|\mathcal{N}|/\zeta) \leq 2d_1 \log 9 + \log(3/\zeta) \leq 4.4\, d_1 + 2\ell. \tag{49}$$

Step 2 (reduction to a projection of rank at most d_2). Fix a unit $|\varphi\rangle \in \mathcal{H}_{\mathrm{A}}$. Because Z is supported on $S^{\perp}$,

$$\langle\varphi| \operatorname{tr}_{\mathrm{B}} Z|\varphi\rangle = \operatorname{tr}(Z\, Q_{\varphi}), \qquad Q_{\varphi} := \Pi_{S^{\perp}}(I_{\mathrm{B}} \otimes |\varphi\rangle\langle\varphi|)\Pi_{S^{\perp}}.$$

Q_{φ} is a positive contraction of rank at most d_2; let R_{φ} be the orthogonal projection onto its range, of rank $m \leq d_2$. Then $Q_{\varphi} \preceq R_{\varphi}$ and hence $\langle\varphi| \operatorname{tr}_{\mathrm{B}} Z|\varphi\rangle \leq \operatorname{tr}(TV^{\dagger}R_{\varphi}V)$.

Step 3 (Gaussian model for the random subspace). If $m = 0$, the quadratic form is zero and the desired bound is immediate. Assume $m \geq 1$. Then $V^{\dagger}R_{\varphi}V$ is a Haar-random rank-m orthogonal projection of $S^{\perp} \simeq \mathbb{C}^{D'}$, so it is distributed as $G(G^{\dagger}G)^{-1}G^{\dagger}$ with G an $D' \times m$ standard complex Gaussian matrix. Since $D' \geq 125\, d_2 \geq 125\, m$, Lemma 3.9 gives $\Pr\{G^{\dagger}G \succeq \frac{D'}{25} I_m\} \geq 1 - 2e^{-D'/16}$, and on that event

$$\operatorname{tr}(TG(G^{\dagger}G)^{-1}G^{\dagger}) = \operatorname{tr}((G^{\dagger}TG)(G^{\dagger}G)^{-1}) \leq \frac{25}{D'} \operatorname{tr}(G^{\dagger}TG).$$

Corollary 3.8 bounds $\operatorname{tr}(G^\dagger T G) \le m \operatorname{tr} T + \sqrt{2m \operatorname{tr}(T^2) u} + \eta u$ except with probability e^{-u}.

Step 4 (bounding the exceptional probabilities). By Equation (48) and $d_2 > \sqrt{K}$ we have $D' > D/2 \ge d_1\sqrt{K}/2$, and also $D' > Kr\ell/2 \ge K\ell/2$. Hence $D'/16 \ge \frac{\sqrt{K}}{64} d_1 + \frac{K}{64}\ell$, which for K a large enough absolute constant exceeds $4.4d_1 + 2\ell + \log 2 \ge u + \log 2$ by Equation (49). Therefore $2e^{-D'/16} \le e^{-u}$ and the two exceptional events together have probability at most $2e^{-u} \le \zeta/|\mathcal{N}|$.

Step 5 (union bound and final estimate). On the intersection over $|\varphi\rangle \in \mathcal{N}$, an event of probability at least $1-\zeta$, using $m \le d_2$, $\operatorname{tr} T \le 2r\eta$, $\operatorname{tr}(T^2) \le \|T\|_\infty \operatorname{tr} T \le 2r\eta^2$ and $D' > D/2$,

$$d_1 \left\|\operatorname{tr}_{\mathrm{B}} Z\right\|_\infty \le \frac{50\, d_1}{D'}\Big(2r\eta d_2 + 2\eta\sqrt{d_2 r u} + \eta u\Big) \le \frac{100}{d_2}\Big(2r\eta d_2 + 2\eta\sqrt{d_2 r u} + \eta u\Big).$$

By Equation (49) and $d_1 \le r d_2$ we have $u/d_2 \le 4.4r + 2\ell \le 7r\ell$ and $\sqrt{ru/d_2} \le \sqrt{4.4r^2 + 2r\ell} \le 2.1r + 1.5r\ell \le 4r\ell$. Therefore $d_1 \left\|\operatorname{tr}_{\mathrm{B}} Z\right\|_\infty \le (200 + 800 + 700)\, r\eta\ell$, which is Equation (47). $\square$

Lemma 3.11 (Diamond-norm error under unitary invariance)**.** *There are absolute constants c, c_{conv} with the following property. Let $\mathcal{E} \in \mathsf{QChan}^r_{d_1,d_2}$ have Kraus rank exactly r. Let P be the orthogonal projector onto $\operatorname{supp}(\mathcal{E}(|I/\sqrt{d_1}\rangle\!\rangle\langle\!\langle I/\sqrt{d_1}|))$, set $P^\perp = I_D - P$, and put $\kappa = \kappa(\mathcal{E}) = d_1\|\operatorname{tr}_{\mathrm{B}} P\|_\infty$ as in Equation (36). Recall the reconstruction map $\Phi_{I/d_1}(\cdot)$ defined in Equation (27), where $C_{\Phi_{I/d_1}(M)} = d_1 M$. Let $\widetilde{X}$ be a random density operator on $\mathcal{H}_{\mathrm{B}} \otimes \mathcal{H}_{\mathrm{A}}$ whose distribution is invariant under conjugation by every unitary of the form $U = I_P \oplus V$, where I_P is the identity on $\operatorname{supp}(P)$ and V is unitary on $\operatorname{supp}(P^\perp)$. Set $\Delta = \widetilde{X} - \mathcal{E}(|I/\sqrt{d_1}\rangle\!\rangle\langle\!\langle I/\sqrt{d_1}|)$. Then for every $\eta, \zeta \in (0,1)$, except on an event of probability at most ζ, $\|\Delta\|_\infty \le \eta$ implies*

$$\left\|\Phi_{I/d_1}(\widetilde{X}) - \mathcal{E}\right\|_\diamond \le c_{\mathrm{conv}}\, \kappa\eta\,\ell, \qquad \ell := 1 + \log(4/\zeta).$$

More precisely, we have

$$\left\|\Phi_{I/d_1}(\widetilde{X}) - \mathcal{E}\right\|_\diamond \le \underbrace{cr\eta\ell}_{P^\perp\Delta P^\perp} + \underbrace{2\eta\sqrt{cr\kappa\ell}}_{\textit{off-diagonal terms}} + \underbrace{\kappa\eta}_{P\Delta P}. \tag{50}$$

Proof. Write $L_\rho = L_{\rho, I/d_1} = I_{\mathrm{B}} \otimes (\sqrt{d_1 \rho^{\mathrm{T}}})$, so that by Lemma 2.11

$$\left\|\Phi_{I/d_1}(\Delta)\right\|_\diamond = \max_{\rho \in \mathcal{D}(\mathcal{H}_{\mathrm{A}})} \left\|L_\rho \Delta L_\rho\right\|_1. \tag{51}$$

Also, $\operatorname{tr}(L_\rho G L_\rho) = d_1 \operatorname{tr}(G(I_{\mathrm{B}} \otimes \rho^{\mathrm{T}})) \le d_1 \left\|\operatorname{tr}_{\mathrm{B}} G\right\|_\infty$ for $G \succeq 0$. Throughout, we will work on the event $\mathsf{E} = \{\|\Delta\|_\infty \le \eta\}$. Note that the distribution of Δ is invariant under the group $\mathcal{G} = \{I_P \oplus V\}$. On E, Lemma 3.5 gives $\|\Delta\|_1 \le 2r\eta$. Put

$$T_1 = P^\perp \Delta P^\perp, \qquad T_2 = P^\perp |\Delta| P^\perp.$$

Both are positive semidefinite (for T_1, we know $P^\perp \mathcal{E}(|I/\sqrt{d_1}\rangle\!\rangle\langle\!\langle I/\sqrt{d_1}|)P^\perp = 0$ and $\widetilde{X} \succeq 0$), and on E both satisfy $\|T_i\|_\infty \le \eta$ and $\operatorname{tr} T_i \le \|\Delta\|_1 \le 2r\eta$. We also know $rd_2 \ge d_1$ since $\mathcal{E}$ is a quantum channel.

Randomization. Let V be Haar-random on $\operatorname{supp}(P^\perp)$, and $U = I_P \oplus V$. We know that $U\Delta U^\dagger$ has the same distribution as Δ, and $T_i(U\Delta U^\dagger) = UT_i(\Delta)U^\dagger$ because $P^\perp$ commutes with U and $|U\Delta U^\dagger| = U|\Delta|U^\dagger$. Consequently, for the event

$$\mathcal{F} := \mathsf{E} \cap \{d_1 \left\|\operatorname{tr}_{\mathrm{B}} T_i\right\|_\infty > cr\eta\ell \text{ for some } i \in \{1,2\}\},$$

we have $\Pr(\mathcal{F}) = \Pr(\mathcal{F}')$ where $\mathcal{F}'$ is the same event computed for $U\Delta U^\dagger$. Conditioning on Δ and applying Lemma 3.10 with $\zeta/2$ to each of the two fixed admissible operators T_1, T_2 (legitimate because its conclusion holds for every fixed admissible T) gives $\Pr(\mathcal{F}) \le \zeta$. From here on we work on $\mathsf{E} \setminus \mathcal{F}$, i.e. we may use

$$d_1 \left\|\mathrm{tr}_{\mathrm{B}}\, T_i\right\|_\infty \le cr\eta\ell, \qquad i = 1, 2. \tag{52}$$

The four blocks. Fix $\rho \in \mathcal{D}(\mathcal{H}_{\mathrm{A}})$ and split $\Delta = P\Delta P + P\Delta P^\perp + P^\perp \Delta P + T_1$.

The $P^\perp \Delta P^\perp$ term. $L_\rho T_1 L_\rho \succeq 0$, so its trace norm is its trace, and Equation (51) with Equation (52) gives $\|L_\rho T_1 L_\rho\|_1 \le cr\eta\ell$.

The $P\Delta P$ term. $P\Delta P$ is supported in $\operatorname{ran} P$ and has operator norm at most η, so Proposition 2.2(ii) and Equation (51) give

$$\|L_\rho P\Delta P L_\rho\|_1 \le \eta\ \mathrm{tr}(L_\rho P L_\rho) \le \eta\, d_1 \left\|\mathrm{tr}_{\mathrm{B}}\, P\right\|_\infty = \kappa\eta.$$

The coefficient is $\kappa(\mathcal{E})$ from Equation (36).

The off-diagonal terms. By $\|MK\|_1 \le \|M\|_2 \|K\|_2$ with $M = L_\rho P^\perp \Delta$ and $K = PL_\rho$,

$$\left\|L_\rho P^\perp \Delta P L_\rho\right\|_1^2 \le \mathrm{tr}(L_\rho P^\perp \Delta^2 P^\perp L_\rho)\ \mathrm{tr}(L_\rho P L_\rho).$$

On E we have $\Delta^2 \preceq \eta|\Delta|$, hence $P^\perp \Delta^2 P^\perp \preceq \eta T_2$; using Equation (51) twice, once with T_2 and Equation (52) and once with P,

$$\left\|L_\rho P^\perp \Delta P L_\rho\right\|_1^2 \le (\eta \cdot cr\eta\ell) \cdot \kappa, \qquad \text{i.e.} \qquad \left\|L_\rho P^\perp \Delta P L_\rho\right\|_1 \le \eta\sqrt{cr\kappa\ell},$$

and the adjoint block obeys the same bound.

Adding the four contributions and maximizing over ρ gives Equation (50). Finally $r \le \kappa$ by Lemma 2.14 and $\ell \ge 1$, so $\sqrt{r\kappa\ell} \le \kappa\ell$ and each of the three terms is at most a constant multiple of $\kappa\eta\ell$. □

3.3.2 Regularized estimation

Lemma 3.12 (Positive semidefinite rank truncation)**.** *Let $X \succeq 0$ be a $D \times D$ matrix of rank at most r, and let $\widehat{X}$ be Hermitian with $\left\|\widehat{X} - X\right\|_\infty \le \theta$. Let S be obtained from $\widehat{X}$ by keeping its r largest eigenvalues if they are positive and setting all other eigenvalues to zero. Then*

$$S \succeq 0, \qquad \operatorname{rank} S \le r, \qquad \|S - X\|_\infty \le 2\theta. \tag{53}$$

Proof. By Weyl's inequality every eigenvalue of $\widehat{X}$ that is negative has modulus at most θ, and if $r < D$ its $(r+1)$-st largest eigenvalue is at most $\lambda_{r+1}(X) + \theta = \theta$. Every eigenvalue discarded in forming S therefore has modulus at most θ, so $\left\|S - \widehat{X}\right\|_\infty \le \theta$, and the triangle inequality gives Equation (53). □

Lemma 3.13 (Weighted quadratic estimate)**.** *There is an absolute constant c_{B} with the following property. Let $\sigma_{\mathrm{in}} \in \mathcal{D}(\mathcal{H}_{\mathrm{A}})$ satisfy $\sigma_{\mathrm{in}} \succeq I_{\mathrm{A}}/(2d_1)$, assume $d_1 \le rd_2$ and $1 \le r \le D$, and use the reconstruction in Equation (27). Let X be a state on $\mathcal{H}_{\mathrm{B}} \otimes \mathcal{H}_{\mathrm{A}}$, let $\widehat{X}$ be the linear estimator in Equation (23) built from n independent Haar snapshots of X, and put $E = \widehat{X} - X$. Let W be a* *fixed $D \times k$ matrix of rank at most r, independent of those snapshots. Then, for every $0 < \delta < 1$, with probability at least $1 - \delta$,*

$$\left\|\Phi_{\sigma_{\text{in}}}(EWW^{\dagger}E)^{\dagger}(I_{\text{B}})\right\|_{\infty}^{1/2} \;\le\; c_{\text{B}}\,\|W\|_{\infty}\sqrt{d_1}\left(\sqrt{\frac{rd_2+\log(4/\delta)}{n}}+\frac{rd_2+\log(4/\delta)}{n}\right). \tag{54}$$

The statement holds conditionally on any earlier classical transcript that determines σ_{in} and W, provided the n snapshots come from fresh measurements.

Proof. A compact singular value decomposition replaces W by a factor with at most r columns, so we may assume $k \le r$ without changing $WW^{\dagger}$ or $\|W\|_{\infty}$; the case $W = 0$ is trivial. Write $|w_1\rangle, \dots, |w_k\rangle$ for the columns of W and define, for a $D \times D$ matrix M, the $(kd_2) \times d_1$ matrix

$$\mathcal{A}_W(M) := \begin{pmatrix} \mathrm{mat}(M|w_1\rangle)\,\sigma_{\text{in}}^{-1/2} \\ \vdots \\ \mathrm{mat}(M|w_k\rangle)\,\sigma_{\text{in}}^{-1/2} \end{pmatrix}.$$

By Equation (3), $\mathrm{mat}(|v\rangle)^{\dagger}\mathrm{mat}(|v\rangle) = (\mathrm{tr}_{\text{B}}|v\rangle\langle v|)^{\text{T}}$, so summing over the blocks gives the identity

$$\mathcal{A}_W(E)^{\dagger}\mathcal{A}_W(E) = \Phi_{\sigma_{\text{in}}}(EWW^{\dagger}E)^{\dagger}(I_{\text{B}}), \qquad \|\mathcal{A}_W(E)\|_{\infty} = \left\|\Phi_{\sigma_{\text{in}}}(EWW^{\dagger}E)^{\dagger}(I_{\text{B}})\right\|_{\infty}^{1/2}. \tag{55}$$

It therefore suffices to control $\|\mathcal{A}_W(E)\|_{\infty}$.

Step 1: direct Haar moment bounds. Let $|\varphi\rangle$ be Haar distributed on the unit sphere of $\mathbb{C}^D$. We claim that, for every complex $D \times D$ matrix M and every real $p \ge 2$,

$$\left(\mathbb{E}\,|D\langle\varphi|M|\varphi\rangle - \mathrm{tr}\,M|^{p}\right)^{1/p} \le c_1 p\,\|M\|_2\,. \tag{56}$$

We first take $M = M^{\dagger}$ and $\mathrm{tr}\,M = 0$. For every positive integer m, the Haar moment identity Equation (25) gives

$$\mathbb{E}\,(D\langle\varphi|M|\varphi\rangle)^m = \frac{D^m}{D(D+1)\cdots(D+m-1)}\sum_{\pi\in S_m}\prod_{c\in\mathcal{C}(\pi)}\mathrm{tr}(M^{|c|}), \tag{57}$$

where $\mathcal{C}(\pi)$ is the set of cycles of π and $|c|$ is the length of the cycle c.

If π has a one-cycle, its contribution to Equation (57) vanishes because $\mathrm{tr}\,M = 0$. For every cycle length $\ell \ge 2$,

$$|\,\mathrm{tr}(M^{\ell})| \le \sum_i |\lambda_i(M)|^{\ell} \le \left(\sum_i |\lambda_i(M)|^2\right)^{\ell/2} = \|M\|_2^{\ell}\,.$$

Therefore, for every even integer $m \ge 2$,

$$\mathbb{E}\,|D\langle\varphi|M|\varphi\rangle|^m \le \frac{D^m}{D(D+1)\cdots(D+m-1)}\sum_{\pi\in S_m}\|M\|_2^m \le m!\,\|M\|_2^m\,.$$

Hence

$$\left(\mathbb{E}\,|D\langle\varphi|M|\varphi\rangle|^m\right)^{1/m} \le (m!)^{1/m}\,\|M\|_2 \le m\,\|M\|_2\,. \tag{58}$$

For a general Hermitian M, set

$$M_0 = M - \frac{\operatorname{tr} M}{D} I_D.$$

Then

$$\|M_0\|_2^2 = \|M\|_2^2 - \frac{|\operatorname{tr} M|^2}{D} \le \|M\|_2^2 .$$

Given $p \ge 2$, choose an even integer m such that $p \le m \le p+2$. Monotonicity of L^p norms and Equation (58) then give

$$(\mathbb{E}\,|D\langle\varphi|M|\varphi\rangle - \operatorname{tr} M|^p)^{1/p} \le m\,\|M\|_2 \le 2p\,\|M\|_2 .$$

Finally, for a general complex matrix M, write

$$M = M_1 + iM_2, \qquad M_1 = \frac{M+M^\dagger}{2}, \qquad M_2 = \frac{M-M^\dagger}{2i}.$$

Both M_1 and M_2 are Hermitian and

$$\|M_1\|_2^2 + \|M_2\|_2^2 = \|M\|_2^2 .$$

Applying the Hermitian estimate to the real and imaginary parts and using the triangle inequality proves Equation (56).

Step 2: transfer to the observed Haar outcome. Let $|\psi\rangle$ be the outcome obtained by measuring the state X in a Haar-random basis. By Lemma 2.10(i), $|\psi\rangle$ has density

$$q_X(|\varphi\rangle) = D\langle\varphi|X|\varphi\rangle$$

with respect to spherical Haar measure. The second Haar moment gives

$$\begin{aligned}\mathbb{E}\,q_X^2 &= D^2\mathbb{E}_{\mathrm{Haar}}(\langle\varphi|X|\varphi\rangle)^2 \\ &= \frac{D((\operatorname{tr} X)^2 + \operatorname{tr}(X^2))}{D+1} = \frac{D(1+\operatorname{tr}(X^2))}{D+1} \le 2, \end{aligned} \tag{59}$$

where we used $\operatorname{tr} X = 1$ and $\operatorname{tr}(X^2) \le 1$.

For the snapshot

$$Y = (D+1)|\psi\rangle\langle\psi| - I_D,$$

and a fixed complex matrix M, define

$$\xi_M := \operatorname{tr}((Y-X)M).$$

As a function of a Haar vector $|\varphi\rangle$,

$$\begin{aligned}\xi_M(|\varphi\rangle) &= (D+1)\langle\varphi|M|\varphi\rangle - \operatorname{tr} M - \operatorname{tr}(XM) \\ &= \frac{D+1}{D}\,(D\langle\varphi|M|\varphi\rangle - \operatorname{tr} M) + \frac{\operatorname{tr} M}{D} - \operatorname{tr}(XM). \end{aligned} \tag{60}$$

The deterministic term is bounded by

$$\Big|\frac{\operatorname{tr} M}{D} - \operatorname{tr}(XM)\Big| \le \frac{|\operatorname{tr} M|}{D} + |\operatorname{tr}(XM)|$$
$$\le \frac{\|M\|_2}{\sqrt{D}} + \|X\|_1 \|M\|_\infty \le 2\|M\|_2 \,. \tag{61}$$

Combining Equations (56), (60) and (61) shows that, for every $p \ge 2$,

$$\left(\mathbb{E}\,|\xi_M|^{2p}\right)^{1/(2p)} \le c_1 p \|M\|_2 \,.$$

Cauchy–Schwarz and Equation (59) now give

$$\begin{aligned}(\mathbb{E}_{q_X} |\xi_M|^p)^{1/p} &= (\mathbb{E}_{\mathrm{Haar}}\,[q_X(|\varphi\rangle)|\xi_M(|\varphi\rangle)|^p])^{1/p} \\ &\le \left(\mathbb{E}\,|q_X|^2\right)^{1/(2p)} \left(\mathbb{E}\,|\xi_M|^{2p}\right)^{1/(2p)} \le c_1 p \|M\|_2 \,.\end{aligned}$$

For $1 \le p < 2$, the same estimate follows by monotonicity from the case $p = 2$. Thus, by the equivalent moment characterization of the ψ_1 norm stated in Fact 2.9, there is an absolute constant c_1 such that

$$\|\xi_M\|_{\psi_1} \le c_1 \|M\|_2 \,. \tag{62}$$

Moreover,

$$\mathbb{E}_{q_X} \xi_M = \operatorname{tr}((\mathbb{E}Y - X)M) = 0$$

by Lemma 2.10(ii).

Let $Y_1, \ldots, Y_n$ be the independent snapshots forming $\widehat{X}$ and write

$$\xi_{M,t} := \operatorname{tr}((Y_t - X)M).$$

The scalar Bernstein inequality for independent centered subexponential variables (Fact 2.9), applied to Equation (62), gives, for every $u \ge 1$,

$$\Pr\left\{\left|\frac{1}{n}\sum_{t=1}^{n} \xi_{M,t}\right| > c_1 \|M\|_2 \left(\sqrt{\frac{u}{n}} + \frac{u}{n}\right)\right\} \le 2e^{-u}. \tag{63}$$

For a complex M, this follows by applying the real-valued inequality to the Hermitian and skew-Hermitian parts and taking a union bound; the factor from this union bound is absorbed by replacing u with $u + \log 2$ and enlarging the absolute constant. Since

$$E = \widehat{X} - X = \frac{1}{n}\sum_{t=1}^{n}(Y_t - X),$$

Equation (63) can equivalently be written as

$$\Pr\left\{|\operatorname{tr}(EM)| > c_1 \|M\|_2 \left(\sqrt{\frac{u}{n}} + \frac{u}{n}\right)\right\} \le 2e^{-u}. \tag{64}$$

Step 3: fixed bilinear forms of $\mathcal{A}_W(E)$. Fix unit vectors

$$|a\rangle \in \mathbb{C}^{kd_2}, \qquad |v\rangle \in \mathbb{C}^{d_1},$$

and decompose $|a\rangle = (|a_1\rangle, \ldots, |a_k\rangle)$ with $|a_j\rangle \in \mathbb{C}^{d_2}$. Define

$$|h\rangle := \sigma_{\text{in}}^{-1/2}|v\rangle, \qquad |q_j\rangle := |a_j\rangle \otimes |h^\star\rangle, \qquad M_{a,v} := \sum_{j=1}^{k} |w_j\rangle\langle q_j|.$$

For every $D \times D$ matrix M, we have

$$\langle a_j|\text{mat}(M|w_j\rangle)|h\rangle = (\langle a_j| \otimes \langle h^\star|)M|w_j\rangle.$$

Therefore,

$$\begin{aligned}\langle a|\mathcal{A}_W(M)|v\rangle &= \sum_{j=1}^{k} \langle a_j|\text{mat}(M|w_j\rangle)\sigma_{\text{in}}^{-1/2}|v\rangle \\ &= \sum_{j=1}^{k} \langle q_j|M|w_j\rangle = \text{tr}(MM_{a,v}).\end{aligned} \tag{65}$$

Let Q be the $D \times k$ matrix with columns $|q_1\rangle, \ldots, |q_k\rangle$. Then $M_{a,v} = WQ^\dagger$. Moreover,

$$\begin{aligned}\|Q\|_2^2 &= \sum_{j=1}^{k} \||a_j\rangle \otimes |h^\star\rangle\|_2^2 = \sum_{j=1}^{k} \||a_j\rangle\|_2^2 \||h^\star\rangle\|_2^2 \\ &= \||a\rangle\|_2^2 \langle v|\sigma_{\text{in}}^{-1}|v\rangle \le 2d_1,\end{aligned}$$

because $\sigma_{\text{in}} \succeq I_{\text{A}}/(2d_1)$. It follows that

$$\|M_{a,v}\|_2 = \left\|WQ^\dagger\right\|_2 \le \|W\|_\infty \|Q\|_2 \le \sqrt{2d_1}\, \|W\|_\infty. \tag{66}$$

Applying Equation (64) with $M = M_{a,v}$ and using Equations (65) and (66), we obtain, for every fixed pair of unit vectors $(|a\rangle, |v\rangle)$ and every $u \ge 1$,

$$\Pr\left\{|\langle a|\mathcal{A}_W(E)|v\rangle| > c_1 \|W\|_\infty \sqrt{d_1}\left(\sqrt{\frac{u}{n}} + \frac{u}{n}\right)\right\} \le 2e^{-u}. \tag{67}$$

Step 4: passage from fixed vectors to the operator norm. By Lemma 2.4, choose deterministic 1/4-covering nets $\mathcal{N}_1$ and $\mathcal{N}_2$ of the unit spheres of $\mathbb{C}^{kd_2}$ and $\mathbb{C}^{d_1}$, respectively, with

$$|\mathcal{N}_1| \le 9^{2kd_2}, \qquad |\mathcal{N}_2| \le 9^{2d_1}.$$

Set

$$u = c_0(kd_2 + d_1) + \log\frac{4}{\delta},$$

where c_0 is a sufficiently large absolute constant. Taking a union bound in Equation (67) over $\mathcal{N}_1 \times \mathcal{N}_2$, we find that, with probability at least $1 - \delta$,

$$\max_{\substack{|a\rangle \in \mathcal{N}_1 \\ |v\rangle \in \mathcal{N}_2}} |\langle a|\mathcal{A}_W(E)|v\rangle| \le c_1 \|W\|_\infty \sqrt{d_1}\left(\sqrt{\frac{kd_2 + d_1 + \log(4/\delta)}{n}} + \frac{kd_2 + d_1 + \log(4/\delta)}{n}\right). \tag{68}$$

The passage from this maximum to the operator norm is Equation (17) with $\eta = 1/4$, applied to $B = \mathcal{A}_W(E)$.

Since $k \le r$ and $d_1 \le rd_2$,

$$kd_2 + d_1 \le 2rd_2.$$

Combining Equations (17) and (68) and absorbing numerical constants gives

$$\|\mathcal{A}_W(E)\|_\infty \le c_{\mathrm{B}} \|W\|_\infty \sqrt{d_1} \left(\sqrt{\frac{rd_2 + \log(4/\delta)}{n}} + \frac{rd_2 + \log(4/\delta)}{n} \right).$$

Finally, Equation (55) yields

$$\left\|\Phi_{\sigma_{\mathrm{in}}}(EWW^\dagger E)^\dagger(I_{\mathrm{B}})\right\|_\infty^{1/2} \le c_{\mathrm{B}} \|W\|_\infty \sqrt{d_1} \left(\sqrt{\frac{rd_2 + \log(4/\delta)}{n}} + \frac{rd_2 + \log(4/\delta)}{n} \right),$$

which is Equation (54). $\square$

Lemma 3.14 (Stability of the regularized quadratic form)**.** *Let $s > 0$ and let $X, S \succeq 0$ be $D \times D$ matrices with $\|S - X\|_\infty \le \eta \le s/4$. Let $F = f_s(X)$ and $Z = X_s$ be the matrices defined in Equation (32), let P_S be the projection onto $\operatorname{ran} S$, and set*

$$R := (S + sI_D)^{-1} P_S.$$

Then $0 \preceq R \preceq s^{-1} I_D$, $\operatorname{rank} R \le \operatorname{rank} S$, and

$$-3\eta F \;\preceq\; XRX - Z \;\preceq\; 3\eta F. \tag{69}$$

Moreover, writing F^+ for the Moore–Penrose inverse and $F^{+1/2} = (F^+)^{1/2}$,

$$W := RXF^{+1/2} \quad \textit{satisfies} \quad WF^{1/2} = RX, \qquad \operatorname{rank} W \le \operatorname{rank} S, \qquad \|W\|_\infty \le 2. \tag{70}$$

Proof. Write $R_0 = (S + sI)^{-1}$ and $R_X = (X + sI)^{-1}$, so $F = XR_X = R_X X$ and $Z = XR_X X$. Iterating the resolvent identity $R_0 = R_X - R_X(S - X)R_0$ once gives the exact second-order form

$$R_0 = R_X - R_X(S - X)R_X + R_X(S - X)R_0(S - X)R_X,$$

and multiplying by X on both sides,

$$XR_0X - Z = -F(S - X)F + F(S - X)R_0(S - X)F.$$

Since $-\eta I \preceq S - X \preceq \eta I$ and $0 \preceq R_0 \preceq s^{-1} I$, conjugation gives

$$-\eta F^2 \;\preceq\; XR_0X - Z \;\preceq\; (\eta + \eta^2/s)F^2. \tag{71}$$

Next, $SQ = 0$ for $Q := I - P_S$ implies $R_0 Q = s^{-1} Q$, hence $R = R_0 - s^{-1} Q$ and $XRX = XR_0X - s^{-1} XQX$. Also $\|XQ\|_\infty = \|(X - S)Q\|_\infty \le \eta$ and $0 \preceq QXQ \preceq \eta Q$. On $\operatorname{supp} X = \operatorname{supp} F$ one has $F^{-1} = (X + sI)X^{-1}$, so $XF^+X = X^2 + sX$ (both sides also vanish on $\ker X$) and

$$\left\|F^{+1/2} XQ\right\|_\infty^2 = \|QXF^+XQ\|_\infty = \left\|Q(X^2 + sX)Q\right\|_\infty \le \eta^2 + s\eta,$$

using $\|QX^2Q\|_\infty = \|XQ\|_\infty^2$. Since $F^{+1/2}F^{1/2} = P_X$ and $XP_X = X$, we may write $QX = (QXF^{+1/2})F^{1/2}$, and therefore

$$XQX = (QX)^\dagger(QX) = F^{1/2}(QXF^{+1/2})^\dagger(QXF^{+1/2})F^{1/2} \preceq (\eta^2 + s\eta)F,$$

that is,

$$s^{-1}XQX \preceq (\eta + \eta^2/s)F. \tag{72}$$

Subtracting Equation (72) from Equation (71) and using $F^2 \preceq F$ and $\eta/s \le 1/4$ gives Equation (69), since both coefficients are then at most $2\eta + \eta^2/s \le \frac{9}{4}\eta \le 3\eta$.

For Equation (70), note $XF^{+1/2} = (X + sI)F^{1/2}$ (both sides vanish on $\ker X$ and agree on $\operatorname{supp} X$). Hence

$$W = R(X + sI)F^{1/2} = [R(S + sI) + R(X - S)]F^{1/2} = [P_S + R(X - S)]F^{1/2},$$

so $\|W\|_\infty \le (1 + \eta/s)\left\|F^{1/2}\right\|_\infty \le \frac{5}{4} \le 2$ and $\operatorname{rank} W \le \operatorname{rank} R \le \operatorname{rank} S$. Finally $F^{+1/2}F^{1/2} = P_X$ and $XP_X = X$, so $WF^{1/2} = RX$. □

Proposition 3.15 (Error bounds for the two-batch estimator)**.** *There is an absolute constant K with the following property. Let $0 < s \le \min\{1, \sqrt{d_1/r}\}$, $0 < \delta < 1$, and $\mathcal{E} \in \mathsf{QChan}^r_{d_1,d_2}$. Fix $\sigma_{\text{in}} \in \mathcal{D}(\mathcal{H}_\text{A})$ with $\sigma_{\text{in}} \succeq I_\text{A}/(2d_1)$, and let $H = H_s(\sigma_{\text{in}})$, where H_s is defined in Equation (33). Set*

$$b := D + \left(1 + \frac{d_1}{r}\right)\log\frac{6}{\delta}, \qquad n = \left\lceil K\,\frac{b}{s^2}\right\rceil. \tag{73}$$

Let $\widehat{Z}, \widehat{H}$ be the outputs of the two-batch estimator in Definition 3.3 with this batch size, and use the reconstruction rule $M \mapsto \Phi_{\sigma_{\text{in}}}(M)$ from Equation (27). With probability at least $1 - \delta$,

$$\tfrac{3}{4}(H + rI_\text{A}) \preceq \widehat{H} + rI_\text{A} \preceq \tfrac{5}{4}(H + rI_\text{A}), \tag{74}$$

$$\left\|\Phi_{\sigma_{\text{in}}}(\widehat{Z}) - \mathcal{E}\right\|_\diamond \le \left(s + \tfrac{3s}{64}\right)\|H\|_\infty + 4s\sqrt{q\,r\,\|H\|_\infty} + q\,rs, \qquad q := 2^{-16}. \tag{75}$$

In particular, if $\|H\|_\infty \le 10r$ then $\left\|\Phi_{\sigma_{\text{in}}}(\widehat{Z}) - \mathcal{E}\right\|_\diamond \le 11\,rs$.

Proof. Write $E = \widehat{X}_1 - \mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)$, $\eta = s/64$, and let $F = f_s(\mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|))$ and $Z = \mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)F$ be as in Equation (32).

Step 1: the three events. By Lemma 2.10(iii) and Lemma 3.12,

$$\|S - \mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)\|_\infty \le \eta \tag{76}$$

with probability at least $1 - \delta/3$, provided $n \ge c(D + \log(6/\delta))/s^2$, which follows from Equation (73) for K large. Condition on the first batch and on Equation (76). Then R and $W = R\mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)F^{+1/2}$ are fixed and independent of E. For the fixed unknown state $\mathcal{E}(|\sqrt{\sigma_{\text{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\text{in}}}|)$, the matrix W is determined by the first batch; it is used only in the proof and need not be computed by the protocol. By Lemma 3.14, $\operatorname{rank} W \le r$, $\|W\|_\infty \le 2$, $\operatorname{rank} R^{1/2} \le r$ and $\left\|R^{1/2}\right\|_\infty \le s^{-1/2}$. Apply Lemma 3.13 twice — once to W and once to $R^{1/2}$ — each with failure probability $\delta/3$. Put $a = rd_2 + \log(12/\delta)$. After an absolute adjustment of K, Equation (73) gives $n \ge Kd_1a/(rs^2)$, and hence

$$\sqrt{d_1}\left(\sqrt{\frac{a}{n}} + \frac{a}{n}\right) \le s\sqrt{\frac{r}{K}} + \frac{rs^2}{K\sqrt{d_1}} \le s\sqrt{r}\left(\frac{1}{\sqrt{K}} + \frac{1}{K}\right),$$

where the last inequality uses $s \le \sqrt{d_1/r}$.

Taking K sufficiently large in terms of c_{B} and q makes the last expression at most $\frac{s\sqrt{qr}}{c_{\mathrm{B}}}$. Since $\|W\|_\infty \le 2$ and $\|R^{1/2}\|_\infty \le s^{-1/2}$, Equation (54) yields

$$\left\|\Phi_{\sigma_{\mathrm{in}}}(EWW^\dagger E)^\dagger(I_{\mathrm{B}})\right\|_\infty \le 4qrs^2, \qquad \left\|\Phi_{\sigma_{\mathrm{in}}}(ERE)^\dagger(I_{\mathrm{B}})\right\|_\infty \le qrs. \tag{77}$$

A union bound gives total failure probability at most δ.

Step 2: the exact expansion. By Lemma 3.14, $R\mathcal{E}(|\sqrt{\sigma_{\mathrm{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\mathrm{in}}}|) = WF^{1/2}$ and $\mathcal{E}(|\sqrt{\sigma_{\mathrm{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\mathrm{in}}}|)R = F^{1/2}W^\dagger$, so

$$\widehat{Z} - Z = \underbrace{\left(\mathcal{E}(|\sqrt{\sigma_{\mathrm{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\mathrm{in}}}|)R\mathcal{E}(|\sqrt{\sigma_{\mathrm{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\mathrm{in}}}|) - Z\right)}_{\text{approximation error}} + \underbrace{\left(EWF^{1/2} + F^{1/2}W^\dagger E\right)}_{\text{linear}} + \underbrace{ERE}_{\succeq 0}. \tag{78}$$

For every $a > 0$, positivity of $(a^{1/2}F^{1/2} \mp a^{-1/2}W^\dagger E)^\dagger(a^{1/2}F^{1/2} \mp a^{-1/2}W^\dagger E)$ gives

$$\pm(EWF^{1/2} + F^{1/2}W^\dagger E) \preceq aF + a^{-1}EWW^\dagger E. \tag{79}$$

Step 3: the multiplicative estimate. Combining Equations (69), (78) and (79) and $ERE \succeq 0$,

$$-[(3\eta + a)F + a^{-1}EWW^\dagger E] \preceq \widehat{Z} - Z \preceq (3\eta + a)F + a^{-1}EWW^\dagger E + ERE.$$

Apply the positive linear map in Equation (28). By Equations (29) and (33), $\Phi_{\sigma_{\mathrm{in}}}(F)^\dagger(I_{\mathrm{B}}) = H$, and Equation (41) and Lemma 2.13 give $\widehat{H} - H = -\Phi_{\sigma_{\mathrm{in}}}(\widehat{Z} - Z)^\dagger(I_{\mathrm{B}})/s$. With $a = s/16$ and $\eta = s/64$, so that $(3\eta + a)/s = 7/64$, the two estimates in Equation (77) give

$$-\left[\tfrac{7}{64}H + 65qrI_{\mathrm{A}}\right] \;\preceq\; \widehat{H} - H \;\preceq\; \left[\tfrac{7}{64}H + 64qrI_{\mathrm{A}}\right].$$

Since $7/64 \le 1/4$ and $65q \le 1/4$, the right-hand sides are at most $\frac{1}{4}(H + rI_{\mathrm{A}})$ in the Löwner order, so $-\frac{1}{4}(H + rI_{\mathrm{A}}) \preceq \widehat{H} - H \preceq \frac{1}{4}(H + rI_{\mathrm{A}})$. Hence, by adding $H + rI_{\mathrm{A}}$, we obtain $\frac{3}{4}(H + rI_{\mathrm{A}}) \preceq \widehat{H} + rI_{\mathrm{A}} \preceq \frac{5}{4}(H + rI_{\mathrm{A}})$, which is Equation (74).

Step 4: the channel estimate. Write $\widehat{Z} - \mathcal{E}(|\sqrt{\sigma_{\mathrm{in}}}\rangle\!\rangle\langle\!\langle\sqrt{\sigma_{\mathrm{in}}}|) = (\widehat{Z} - Z) - sF$ and apply $M \mapsto \Phi_{\sigma_{\mathrm{in}}}(M)$. By Lemma 2.11, the approximation term of Equation (78) contributes at most $3\eta\,\|H\|_\infty$ (Part (b) with $G = F$); the linear term of Equation (78) contributes at most $2\sqrt{\|\Phi_{\sigma_{\mathrm{in}}}(EWW^\dagger E)^\dagger(I_{\mathrm{B}})\|_\infty\,\|H\|_\infty} \le 4s\sqrt{qr\,\|H\|_\infty}$ (Part (c) with $U = EW$, $V = F^{1/2}$); the term $ERE \succeq 0$ of Equation (78) contributes exactly $\|\Phi_{\sigma_{\mathrm{in}}}(ERE)^\dagger(I_{\mathrm{B}})\|_\infty \le qrs$ (Part (a)); and $-sF$ contributes exactly $s\,\|H\|_\infty$ (Lemma 2.13). Adding gives Equation (75). If $\|H\|_\infty \le 10r$ then the right-hand side is at most $10rs(1 + \frac{3}{64}) + 4rs\sqrt{10q} + qrs \le 11rs$.

□

3.3.3 Input selection

We first record the matrix inequalities used to choose the input marginal.

Lemma 3.16 (Operator monotonicity and concavity; cf. [Bha97, Ch. V])**.** *On positive definite matrices, $x \mapsto x^{-1}$ is operator convex and operator decreasing, and $x \mapsto \log x$ and, for each $s > 0$, $x \mapsto x/(x + s)$ are operator monotone and operator concave.*

Proof. The inverse. For $M \succ 0$ and $|x\rangle \in \mathbb{C}^D$, $\langle x|M^{-1}|x\rangle = \sup_{|z\rangle}(2\,\mathrm{Re}\langle x|z\rangle - \langle z|M|z\rangle)$, attained at $|z\rangle = M^{-1}|x\rangle$. Each quadratic form of M^{-1} is thus a supremum of functions affine in M, hence convex, and visibly nonincreasing in M; so $M \mapsto M^{-1}$ is operator convex and operator decreasing.

The others. For $x > 0$ and $s > 0$,

$$\frac{x}{x+s} = 1 - \frac{s}{x+s}, \qquad \log x = \int_0^\infty \Big(\frac{1}{1+t} - \frac{1}{x+t}\Big)dt,$$

the second being the standard integral representation (both sides vanish at $x = 1$ and have derivative $1/x$, since $\int_0^\infty (x+t)^{-2}dt = 1/x$). By the previous paragraph applied to $M + tI$, each integrand $\frac{1}{1+t}I - (M+tI)^{-1}$ is operator concave and operator monotone in M, and these properties are preserved by taking nonnegative combinations and pointwise limits; the same applies to $I - s(M+sI)^{-1}$. □

Fact 3.17 (Golden–Thompson [Bha97, Thm. IX.3.7])**.** *For Hermitian M, N one has* $\operatorname{tr} e^{M+N} \leq \operatorname{tr}(e^M e^N)$.

Lemma 3.18 (Trace bound for $A(\rho)$)**.** *Fix $s > 0$ and, as in Equation (42), set $\sigma_{\mathrm{in}}(\rho) = \rho/2 + I_{\mathrm{A}}/(2d_1)$ and $A(\rho) = H_s(\sigma_{\mathrm{in}}(\rho)) + rI_{\mathrm{A}}$, where H_s is defined in Equation (33). Then every $\rho \in \mathcal{D}(\mathcal{H}_{\mathrm{A}})$ satisfies $\operatorname{tr}(\rho A(\rho)) \leq 3r$.*

Proof. Put $\nu = \sigma_{\mathrm{in}}(\rho) \succeq \rho/2$. By Equation (35) and positivity of $\mathcal{E}^c$, $\mathcal{E}^c(\rho) \preceq 2\mathcal{E}^c(\nu)$, so

$$\operatorname{tr}(\rho H_s(\nu)) = \operatorname{tr}(\mathcal{E}^c(\rho)(\mathcal{E}^c(\nu) + sI_r)^{-1}) \leq 2\operatorname{tr}(\mathcal{E}^c(\nu)(\mathcal{E}^c(\nu) + sI_r)^{-1}) \leq 2r.$$

Adding $\operatorname{tr}(\rho \cdot rI_{\mathrm{A}}) = r$ finishes the proof. □

Lemma 3.19 (Operator convexity of $Q \mapsto \log\Phi(Q^{-1})$)**.** *Let $\Phi : \mathcal{L}(\mathbb{C}^m) \to \mathcal{L}(\mathcal{H}_{\mathrm{A}})$ be completely positive with $\Phi(I_m) \succ 0$. Then the map $Q \mapsto \log\Phi(Q^{-1})$ is operator convex on positive definite Q.*

Proof. Take a Stinespring representation $\Phi(Y) = V^\dagger(I_k \otimes Y)V$ with $V : \mathcal{H}_{\mathrm{A}} \to \mathbb{C}^k \otimes \mathbb{C}^m$. The hypothesis $\Phi(I_m) = V^\dagger V \succ 0$ makes V injective, so $V^\dagger$ is onto. Set $\Psi(Q) := [\Phi(Q^{-1})]^{-1}$. For every $|x\rangle \in \mathcal{H}_{\mathrm{A}}$,

$$\langle x|\Psi(Q)|x\rangle = \min\{\langle y|(I_k \otimes Q)|y\rangle : \ V^\dagger|y\rangle = |x\rangle\}. \tag{80}$$

Indeed, for positive definite M the constrained minimum of $\langle y|M|y\rangle$ subject to $V^\dagger|y\rangle = |x\rangle$ equals $\langle x|(V^\dagger M^{-1}V)^{-1}|x\rangle$, attained at $|y\rangle = M^{-1}V(V^\dagger M^{-1}V)^{-1}|x\rangle$; take $M = I_k \otimes Q$ and note $V^\dagger(I_k \otimes Q^{-1})V = \Phi(Q^{-1})$.

By Equation (80) each quadratic form of Ψ is an infimum of functions affine in Q, hence concave; so Ψ is operator concave. Since log is operator monotone and operator concave (Lemma 3.16), $\log\circ\Psi$ is operator concave: for $t \in [0,1]$,

$$\log\Psi(tQ_0 + (1-t)Q_1) \succeq \log(t\Psi(Q_0) + (1-t)\Psi(Q_1)) \succeq t\log\Psi(Q_0) + (1-t)\log\Psi(Q_1).$$

Finally $\log\Phi(Q^{-1}) = -\log\Psi(Q)$, which is therefore operator convex. □

Corollary 3.20 (Operator convexity of $\rho \mapsto \log A(\rho)$)**.** *Fix $s > 0$ and let $A(\rho) = H_s(\rho/2 + I_{\mathrm{A}}/(2d_1)) + rI_{\mathrm{A}}$ be the matrix-valued map defined in Equation (42), with H_s defined in Equation (33). Then $\rho \mapsto \log A(\rho)$ is operator convex on $\mathcal{D}(\mathcal{H}_{\mathrm{A}})$.*

Proof. On $\mathbb{C}^r \oplus \mathcal{H}_\mathrm{A}$ define the affine, positive definite matrix $Q(\rho) = \mathrm{diag}(\mathcal{E}^c(\sigma_\mathrm{in}(\rho)) + sI_r,\ I_\mathrm{A})$ and the completely positive map $\Phi(\mathrm{diag}(Y_1, Y_2)) = (\mathcal{E}^c)^\dagger(Y_1) + rY_2$, extended to all of $\mathcal{L}(\mathbb{C}^r \oplus \mathcal{H}_\mathrm{A})$ by compressing to the two diagonal blocks first. Then $\Phi(I) = (1+r)I_\mathrm{A} \succ 0$ and, by Equation (34), $\Phi(Q(\rho)^{-1}) = H_s(\sigma_\mathrm{in}(\rho)) + rI_\mathrm{A} = A(\rho)$. Apply Lemma 3.19 and compose with the affine map $\rho \mapsto Q(\rho)$. □

Proposition 3.21 (Input selection from multiplicative estimates)**.** *Fix $s > 0$, and let $\sigma_\mathrm{in}(\rho)$ and $A(\rho)$ be the maps defined in Equation (42), formed from H_s in Equation (33). Put $T := \max\{1, \lceil \log_2 d_1 \rceil\}$ and $\rho_1 := I_\mathrm{A}/d_1$. Suppose that for each $j = 1, \dots, T$ a positive definite A_j is given with*

$$\tfrac{3}{4}\, A(\rho_j) \ \preceq\ A_j \ \preceq\ \tfrac{5}{4}\, A(\rho_j), \tag{81}$$

and define

$$z_j = \operatorname{tr}\exp(\log \rho_j + \log A_j), \qquad \rho_{j+1} = \frac{\exp(\log \rho_j + \log A_j)}{z_j}. \tag{82}$$

Then $\bar\rho := \frac{1}{T}\sum_{j=1}^{T} \rho_j$ satisfies

$$A(\bar\rho) \preceq 10rI_\mathrm{A}, \qquad \textit{equivalently} \qquad H_s(\sigma_\mathrm{in}(\bar\rho)) \preceq 9rI_\mathrm{A}. \tag{83}$$

Proof. Each ρ_j is a density operator, positive definite by induction. By Golden–Thompson (Fact 3.17), Equation (81) and Lemma 3.18,

$$z_j \le \operatorname{tr}(\rho_j A_j) \le \tfrac{5}{4}\operatorname{tr}(\rho_j A(\rho_j)) \le \tfrac{15r}{4}.$$

Taking logarithms in Equation (82) is exact and gives $\log \rho_{j+1} = \log \rho_j + \log A_j - (\log z_j) I_\mathrm{A}$; summing over $j = 1, \dots, T$,

$$\sum_{j=1}^{T} \log A_j = \log \rho_{T+1} - \log \rho_1 + \Big(\sum_{j=1}^{T} \log z_j\Big) I_\mathrm{A} \preceq (\log d_1) I_\mathrm{A} + T \log \tfrac{15r}{4}\, I_\mathrm{A},$$

using $\rho_{T+1} \preceq I_\mathrm{A}$, hence $\log \rho_{T+1} \preceq 0$, and $-\log \rho_1 = (\log d_1) I_\mathrm{A}$. By operator monotonicity of the logarithm (Lemma 3.16) and the lower bound in Equation (81), $\log A(\rho_j) \preceq \log A_j + \log \frac{4}{3}\, I_\mathrm{A}$. Therefore

$$\frac{1}{T}\sum_{j=1}^{T} \log A(\rho_j) \preceq \Big(\frac{\log d_1}{T} + \log(5r)\Big) I_\mathrm{A},$$

because $\frac{15}{4} \cdot \frac{4}{3} = 5$. Corollary 3.20 bounds $\log A(\bar\rho)$ by the left-hand side, so the largest eigenvalue of $\log A(\bar\rho)$ is at most $\log(5r) + (\log d_1)/T$; exponentiating this *scalar* inequality gives $A(\bar\rho) \preceq 5r\, d_1^{1/T} I_\mathrm{A} \preceq 10rI_\mathrm{A}$, since $d_1^{1/T} \le 2$ by the choice of T. Subtracting rI_A gives Equation (83). □

4 Lower bound

In this section we prove the lower bound of Theorem 1.3 which we restate here as Theorem 4.1.

Theorem 4.1. *There are absolute constants $c, \epsilon_0 > 0$ such that the following holds. Assume $d_2 \geq 2$, $1 \leq r \leq D$, $d_1 \leq rd_2$, $Dr \geq 12$, and $0 < \epsilon \leq \epsilon_0$. Every adaptive incoherent protocol that learns every channel in $\mathsf{QChan}^r_{d_1,d_2}$ to diamond-norm error ϵ with probability at least $2/3$ uses at least*

$$c\frac{Dr^2}{\epsilon^2}$$

queries. The hard instance family used to prove the lower bound can be chosen so that every non-zero eigenvalue of its Choi operator belongs to $[d_1/(4r), 4d_1/r]$. The conclusion therefore also holds for learning channels with $\Omega(d_1/r)$-gapped Choi spectrum.

For the proof, we construct a local family of exactly trace-preserving channels with $\Theta(Dr)$ real parameters. We then show that the Fisher information matrix of a single incoherent query has trace at most $16D/r$. A Gaussian prior with variance $\Theta(t^2/D)$ in each coordinate, conditioned on the parameter domain, yields the mutual-information bound $O(nt^2/r)$ between the channel parameter and the algorithm transcript. Finally, learning to error $\Theta(t)$ requires $\Omega(Dr)$ nats of information, which gives the desired lower bound.

4.1 Main proof

The local family. For a Kraus list $\{K_1, \ldots, K_r\}$, we write its stack and scaled Choi factor as

$$\mathbf{K} = \begin{pmatrix} K_1 \\ \vdots \\ K_r \end{pmatrix}, \qquad U = \sqrt{\frac{r}{d_1}}[\,|K_1\rangle\!\rangle \;\cdots\; |K_r\rangle\!\rangle\,], \qquad C_{\mathcal{E}} = \frac{d_1}{r}UU^\dagger. \tag{84}$$

The reshaping map

$$\mathsf{S}(M) = \begin{pmatrix} \mathrm{mat}(|m_1\rangle) \\ \vdots \\ \mathrm{mat}(|m_r\rangle) \end{pmatrix}, \qquad M = [|m_1\rangle \;\cdots\; |m_r\rangle], \tag{85}$$

is a Hilbert–Schmidt isometry and satisfies $\mathsf{S}(U) = \sqrt{r/d_1}\,\mathbf{K}$. By Lemma 4.2, choose reference Kraus operators with stack $\mathbf{K}_0$ and scaled Choi factor U_0 satisfying

$$\mathbf{K}_0^\dagger \mathbf{K}_0 = I_{\mathrm{A}}, \qquad \tfrac{1}{2}I_r \preceq U_0^\dagger U_0 \preceq 2I_r. \tag{86}$$

For $\Theta \in \mathbb{C}^{D\times r}$, put

$$T_\Theta = \sqrt{d_1/r}\,\mathsf{S}(\Theta), \qquad \|T_\Theta\|_2 = \sqrt{d_1/r}\,\|\Theta\|_2, \tag{87}$$

and consider the real linear space

$$\mathcal{V} = \{\Theta : U_0^\dagger\Theta = \Theta^\dagger U_0, \quad \mathbf{K}_0^\dagger T_\Theta + T_\Theta^\dagger \mathbf{K}_0 = 0\}. \tag{88}$$

Unitary mixing of the Kraus operators changes U_0 to U_0V, for a unitary V, without changing $U_0U_0^\dagger$. The first constraint makes Θ orthogonal, in the real Hilbert–Schmidt inner product, to all directions U_0W with $W^\dagger = -W$. The second constraint is the linearized trace-preservation condition. Define

$$Q_\Theta = (I_{\mathrm{A}} + T_\Theta^\dagger T_\Theta)^{-1/2}, \qquad \mathbf{K}_\Theta = (\mathbf{K}_0 + T_\Theta)Q_\Theta. \tag{89}$$

For every $\Theta \in \mathcal{V}$,

$$(\mathbf{K}_0 + T_\Theta)^\dagger (\mathbf{K}_0 + T_\Theta) = I_\mathrm{A} + T_\Theta^\dagger T_\Theta, \qquad \mathbf{K}_\Theta^\dagger \mathbf{K}_\Theta = I_\mathrm{A}.$$

Thus $\mathbf{K}_\Theta$ defines a channel $\mathcal{E}_\Theta$. Its scaled factor and Choi operator are

$$U_\Theta = (I_\mathrm{B} \otimes Q_\Theta^\mathrm{T})(U_0 + \Theta), \qquad C_\Theta := C_{\mathcal{E}_\Theta} = \frac{d_1}{r} U_\Theta U_\Theta^\dagger. \tag{90}$$

For $0 < t \le t_0 := 2^{-5}$ restrict to the compact convex set

$$\mathcal{K}_t = \{\Theta \in \mathcal{V} : \|\Theta\|_\infty \le t,\ \|T_\Theta\|_\infty \le t\}. \tag{91}$$

Choose an orthonormal basis of the real Hilbert space $\mathcal{V}$, with inner product $\langle M, N\rangle_\mathbb{R} = \operatorname{Re}\operatorname{tr}(M^\dagger N)$, and denote the coordinates of Θ by $\theta \in \mathbb{R}^k$, where $k = \dim_\mathbb{R} \mathcal{V}$. We identify $\mathcal{K}_t$ with its coordinate image in $\mathbb{R}^k$. Set

$$\sigma := \frac{t}{32\sqrt{D}}, \qquad \gamma_\sigma := \mathcal{N}(0, \sigma^2 I_k). \tag{92}$$

For every Borel set $A \subseteq \mathbb{R}^k$, define

$$\mu_t(A) := \Pr_{\theta\sim\gamma_\sigma} \{\theta \in A \,|\, \theta \in \mathcal{K}_t\}. \tag{93}$$

The dimension, spectral bounds, and prior estimates needed below are proved in Lemmas 4.3, 4.6 and 4.10.

Proof of Theorem 4.1. Take $M = 2^{21}$, $\epsilon_0 = 2^{-26}$, and $t = M\epsilon \le t_0$. Run the protocol on $\mathcal{E}_\Theta$ with $\theta \sim \mu_t$ and let Z be its classical transcript. By Lemma 4.6, every channel in the family has rank exactly r and non-zero Choi eigenvalues in $[d_1/(4r), 4d_1/r]$.

Information available from the transcript. Lemma 4.7 bounds the trace of the one-query Fisher information matrix by $16D/r$, uniformly over the input, measurement, and parameter. Conditional scores have mean zero, so this bound adds over adaptive rounds. The log-Sobolev inequality for μ_t then gives

$$I(\theta; Z) \le \frac{nt^2}{128r} \tag{94}$$

by Proposition 4.11.

Information required for reconstruction. On the success event, Equation (7) implies

$$\frac{1}{d_1}\|C_\Theta - C_{\widehat{\mathcal{E}}(Z)}\|_1 \le \epsilon.$$

By Lemma 4.12, every such ball has prior mass at most $2e^{-Dr/2}$. Fano's inequality for reconstruction sets (Lemma 4.13) and success probability $2/3$ yield

$$I(\theta; Z) \ge \frac{2}{3}\left(\frac{Dr}{2} - \log 2\right) - \log 2 \ge \frac{Dr}{6},$$

where the last inequality holds for $Dr \ge 12$. Comparing with Equation (94) gives

$$n \ge \frac{128}{6}\frac{Dr^2}{M^2\epsilon^2} \ge \frac{21}{M^2}\frac{Dr^2}{\epsilon^2}.$$

This proves the theorem with $c = 21/M^2$. Since the family has rank exactly r, it is also contained in the rank-at-most-r class. $\square$

4.2 Technical lemmas

We establish the geometric estimates for the parameterized family, the mutual-information bound for adaptive transcripts, and the bound on the prior mass of trace-norm balls used in Section 4.1.

4.2.1 Geometry of the lower-bound family

Lemma 4.2 (Reference Kraus operators)**.** *Assume $d_2 \geq 2, 1 \leq r \leq D = d_1 d_2$ and $d_1 \leq r d_2$. There are Kraus operators $K_1, \ldots, K_r \in \mathbb{C}^{d_2 \times d_1}$, with stack $\mathbf{K}_0$ and Choi factor U_0 as in Equation (84), such that*

$$\mathbf{K}_0^\dagger \mathbf{K}_0 = I_{\mathrm{A}}, \qquad \tfrac{1}{2} I_r \ \preceq\ U_0^\dagger U_0 \ \preceq\ 2\, I_r.$$

The K_i may be taken Hilbert–Schmidt orthogonal, so that $U_0^\dagger U_0$ is diagonal.

Proof. Note $(U_0^\dagger U_0)_{ij} = (r/d_1)\,\mathrm{tr}(K_i^\dagger K_j)$.

Case $r \leq d_1$. Partition the input basis into r nonempty blocks of sizes $a_i \in \{\lfloor d_1/r \rfloor, \lceil d_1/r \rceil\}$ summing to d_1. Since $d_1 \leq r d_2$ we have $d_1/r \leq d_2$, and as d_2 is an integer, $\lceil d_1/r \rceil \leq d_2$; so each block admits an isometry into $\mathcal{H}_{\mathrm{B}}$. Let K_i be such an isometry on the i-th block and zero elsewhere. Then $\sum_i K_i^\dagger K_i = I_{\mathrm{A}}$, the K_i have orthogonal supports, and $\mathrm{tr}(K_i^\dagger K_j) = a_i \delta_{ij}$. Hence $U_0^\dagger U_0$ is diagonal with entries $r a_i / d_1$. For $x = d_1/r \geq 1$ one has $\lfloor x \rfloor \geq x/2$ and $\lceil x \rceil \leq 2x$, which gives Equation (86).

Case $r \geq d_1$. Choose integers $k_x \in \{\lfloor r/d_1 \rfloor, \lceil r/d_1 \rceil\}$, $1 \leq x \leq d_1$, with $\sum_x k_x = r$. Since $r \leq d_1 d_2$ we have $\lceil r/d_1 \rceil \leq d_2$, so for each x we may pick k_x orthonormal vectors $|v_{x,1}\rangle, \ldots, |v_{x,k_x}\rangle$ in $\mathcal{H}_{\mathrm{B}}$. Put $K_{x,j} = k_x^{-1/2} |v_{x,j}\rangle\langle x|$. There are exactly r of them, $\sum_{x,j} K_{x,j}^\dagger K_{x,j} = I_{\mathrm{A}}$, and they are Hilbert–Schmidt orthogonal: different x have orthogonal input supports, equal x have orthogonal output vectors. Their squared norms are $1/k_x$, so $U_0^\dagger U_0$ is diagonal with entries $r/(d_1 k_x)$, and the same floor–ceiling estimates applied to $y = r/d_1 \geq 1$ give Equation (86). The two cases agree when $r = d_1$. □

Lemma 4.3 (Dimension of the parameter space)**.** *Assume Equation (1). Let $\mathcal{V}$ be the real linear space defined in Equation (88) from reference factors satisfying Equation (86), and set $k = \dim_{\mathbb{R}} \mathcal{V}$. Then*

$$\tfrac{3}{4}\, Dr \ \leq\ \Big(1 - \tfrac{1}{d_2^2}\Big) Dr \ \leq\ 2Dr - r^2 - d_1^2 \ \leq\ k \ \leq\ 2Dr. \tag{95}$$

Proof. The ambient real dimension is $2Dr$ so $k \leq 2Dr$. The Hermiticity constraint says that the skew-Hermitian part of $U_0^\dagger \Theta$ vanishes, which is r^2 real linear conditions; the trace-preservation constraint takes values in the Hermitian $d_1 \times d_1$ matrices, which is d_1^2 real linear conditions. By rank–nullity $k \geq 2Dr - r^2 - d_1^2$; independence of the constraints is not needed.

By Equation (1) we have $d_1/d_2 \leq r \leq d_1 d_2$, so $(r - d_1/d_2)(r - d_1 d_2) \leq 0$. Expanding, $r^2 + d_1^2 \leq r d_1 d_2 + r d_1/d_2 = Dr(1 + d_2^{-2})$. Since $d_2 \geq 2$ this is at most $\frac{5}{4} Dr$, and Equation (95) follows. □

Lemma 4.4 (Inverse square root)**.** *Let $X, Y \succeq I$ be positive definite. Then, in the operator norm and in the Hilbert–Schmidt norm, we have*

$$\|X^{-1/2} - Y^{-1/2}\| \ \leq\ \tfrac{1}{2} \|X - Y\|.$$

Consequently the Fréchet derivative of $X \mapsto X^{-1/2}$ at any $X \succeq I$ has norm at most $\frac{1}{2}$ in both norms.

Proof. Use $X^{-1/2} = \frac{1}{\pi}\int_0^\infty s^{-1/2}(X+sI)^{-1}\,ds$ and the resolvent identity $(X+sI)^{-1} - (Y+sI)^{-1} = (X+sI)^{-1}(Y-X)(Y+sI)^{-1}$. For $X, Y \succeq I$ both resolvents have operator norm at most $(1+s)^{-1}$, so the integrand is bounded in either norm by $s^{-1/2}(1+s)^{-2}\|X-Y\|$; here we use that a two-sided multiplication is bounded on the Hilbert–Schmidt norm by the product of the operator norms of its factors. Since $\int_0^\infty s^{-1/2}(1+s)^{-2}ds = \pi/2$, the claim follows. The derivative statement follows by taking limit. $\square$

Lemma 4.5 (Remainder estimates)**.** *Let $0 < t \le t_0 = 2^{-5}$ and $\Theta, \Xi \in \mathcal{K}_t$, where $\mathcal{K}_t \subseteq \mathcal{V}$ is the parameter domain defined in Equations (88) and (91). Let Q_Θ and U_Θ be given by Equations (89) and (90), and define*

$$U_\Theta = U_0 + \mathbf{S}_\Theta, \qquad \mathbf{S}_\Theta := \Theta + R_\Theta, \qquad R_\Theta := (I_\mathrm{B} \otimes (Q_\Theta - I_\mathrm{A})^\mathrm{T})(U_0 + \Theta). \tag{96}$$

Put $\Delta = \Theta - \Xi$. For every direction $H \in \mathcal{V}$, with $DU_\Theta[H]$ denoting the Fréchet derivative of $\Theta \mapsto U_\Theta$ in direction H, the following estimates hold:

$$\|Q_\Theta - I_\mathrm{A}\|_\infty \le \tfrac{1}{2}t^2, \qquad \|R_\Theta\|_\infty \le t^2, \qquad \|\mathbf{S}_\Theta\|_\infty \le \tfrac{17}{16}t, \tag{97}$$

$$\|R_\Theta - R_\Xi\|_2 \le \tfrac{5}{4}t\,\|\Delta\|_2\,, \qquad \|\mathbf{S}_\Theta - \mathbf{S}_\Xi\|_2 \le \tfrac{9}{8}\|\Delta\|_2\,, \qquad \|DU_\Theta[H]\|_2 \le 2\,\|H\|_2\,. \tag{98}$$

Proof. Since $\|T_\Theta\|_\infty \le t$ we have $I_\mathrm{A} \preceq I_\mathrm{A} + T_\Theta^\dagger T_\Theta \preceq (1+t^2)I_\mathrm{A}$, so Lemma 4.4 with $Y = I$ gives $\|Q_\Theta - I\|_\infty \le \frac{1}{2}\left\|T_\Theta^\dagger T_\Theta\right\|_\infty \le \frac{1}{2}t^2$. As $\|U_0\|_\infty \le \sqrt{2}$ by Equation (86) and $\|\Theta\|_\infty \le t$,

$$\|R_\Theta\|_\infty \le \tfrac{1}{2}t^2(\sqrt{2}+t) \le t^2, \qquad \|\mathbf{S}_\Theta\|_\infty \le t + t^2 \le \tfrac{17}{16}t,$$

using $t \le 2^{-5}$. This is Equation (97).

For the Hilbert–Schmidt estimates, use the reshaping map S from Equation (85). Under this map, $M \mapsto (I_\mathrm{B} \otimes N^\mathrm{T})M$ is $\mathsf{S}(M) \mapsto \mathsf{S}(M)N$, so for any $N \in \mathbb{C}^{d_1 \times d_1}$,

$$\left\|(I_\mathrm{B} \otimes N^\mathrm{T})(U_0 + \Theta)\right\|_2 = \|\mathsf{S}(U_0+\Theta)N\|_2 \le \|\mathsf{S}(U_0+\Theta)\|_\infty \|N\|_2 \le \sqrt{\tfrac{r}{d_1}}\,(1+t)\,\|N\|_2\,, \tag{99}$$

because $\mathsf{S}(U_0 + \Theta) = \sqrt{r/d_1}(\mathbf{K}_0 + T_\Theta)$ with $\|\mathbf{K}_0\|_\infty = 1$ and $\|T_\Theta\|_\infty \le t$. Combined with $\|T_\Delta\|_2 = \sqrt{d_1/r}\,\|\Delta\|_2$ from Equation (87), the factor $\sqrt{d_1/r}$ cancels, giving a bound independent of the dimensions.

Now decompose

$$R_\Theta - R_\Xi = (I_\mathrm{B} \otimes (Q_\Theta - Q_\Xi)^\mathrm{T})(U_0 + \Theta) + (I_\mathrm{B} \otimes (Q_\Xi - I)^\mathrm{T})\Delta.$$

By Lemma 4.4, $\|Q_\Theta - Q_\Xi\|_2 \le \frac{1}{2}\left\|T_\Theta^\dagger T_\Theta - T_\Xi^\dagger T_\Xi\right\|_2 \le \frac{1}{2}(\|T_\Theta\|_\infty + \|T_\Xi\|_\infty)\,\|T_\Delta\|_2 \le t\sqrt{d_1/r}\,\|\Delta\|_2$. By Equation (99) the first term has Hilbert–Schmidt norm at most $t(1+t)\,\|\Delta\|_2$, and the second at most $\frac{1}{2}t^2\,\|\Delta\|_2$. Hence $\|R_\Theta - R_\Xi\|_2 \le (t + t^2 + \frac{1}{2}t^2)\,\|\Delta\|_2 \le \frac{5}{4}t\,\|\Delta\|_2$ and $\|\mathbf{S}_\Theta - \mathbf{S}_\Xi\|_2 \le (1 + \frac{5}{4}t)\,\|\Delta\|_2 \le \frac{9}{8}\,\|\Delta\|_2$.

For the derivative, differentiate Equation (96) in a direction $H \in \mathcal{V}$:

$$DU_\Theta[H] = H + (I_\mathrm{B} \otimes (DQ_\Theta[H])^\mathrm{T})(U_0 + \Theta) + (I_\mathrm{B} \otimes (Q_\Theta - I)^\mathrm{T})H.$$

Lemma 4.4 gives $\|DQ_\Theta[H]\|_2 \le \frac{1}{2}\left\|T_H^\dagger T_\Theta + T_\Theta^\dagger T_H\right\|_2 \le \|T_\Theta\|_\infty \|T_H\|_2 \le t\sqrt{d_1/r}\,\|H\|_2$, so by Equation (99) the middle term is at most $t(1+t)\,\|H\|_2$ and the last is at most $\frac{1}{2}t^2\,\|H\|_2$. Therefore $\|DU_\Theta[H]\|_2 \le (1 + \frac{5}{4}t)\,\|H\|_2 \le 2\,\|H\|_2$. $\square$

Lemma 4.6 (Spectral and metric control)**.** *Let $0 < t \le t_0 = 2^{-5}$ and $\Theta, \Xi \in \mathcal{K}_t$, with $\mathcal{V}$ and $\mathcal{K}_t$ as in Equations (88) and (91). Use the scaled Choi factor U_Θ and Choi operator $C_\Theta = C_{\mathcal{E}_\Theta}$ defined in Equation (90); derivatives are taken along the real space $\mathcal{V}$. Then*

(i) *C_Θ has rank r and its non-zero eigenvalues lie in $[d_1/(4r), 4d_1/r]$.*

(ii) *$\|DU_\Theta[H]\|_2 \le 2\|H\|_2$ for all $H \in \mathcal{V}$.*

(iii) *$d_1^{-1}\|C_\Theta - C_\Xi\|_2 \ge (2r)^{-1}\|\Theta - \Xi\|_2$.*

(iv) *$d_1^{-1}\|C_\Theta - C_\Xi\|_\infty \le 16t/r$.*

(v) *$\|\Theta - \Xi\|_2^2 \le 4rt^2$.*

Proof. (i) By Equation (86), $\sigma_{\min}(U_0) \ge 2^{-1/2}$ and $\|U_0\|_\infty \le \sqrt{2}$. The matrix $I_{\mathrm{B}} \otimes Q_\Theta^{\mathrm{T}}$ is positive with the spectrum of Q_Θ, hence with eigenvalues in $[(1+t^2)^{-1/2}, 1]$. Therefore, by Equation (90),

$$\sigma_{\min}(U_\Theta) \;\ge\; \frac{2^{-1/2} - t}{\sqrt{1+t^2}} \;>\; \tfrac{1}{2}, \qquad \|U_\Theta\|_\infty \;\le\; \sqrt{2} + t \;<\; 2,$$

for $t \le 2^{-5}$. The non-zero eigenvalues of $(C_\Theta/d_1) = U_\Theta U_\Theta^\dagger/r$ are those of $U_\Theta^\dagger U_\Theta/r$, so they lie in $[1/(4r), 4/r]$, and there are exactly r of them.

(ii) This is the derivative estimate in Equation (98).

(iii) Put $\Delta = \Theta - \Xi \in \mathcal{V}$ and $L(\Delta) = U_0\Delta^\dagger + \Delta U_0^\dagger$. Expanding and using cyclicity,

$$\|L(\Delta)\|_2^2 = 2\operatorname{tr}(\Delta^\dagger \Delta\, U_0^\dagger U_0) + 2\operatorname{tr}((U_0^\dagger \Delta)^2).$$

The Hermiticity constraint in Equation (88) says that $U_0^\dagger\Delta$ is Hermitian, so the second trace is nonnegative; and $U_0^\dagger U_0 \succeq \frac{1}{2}I$ by Equation (86), so the first is at least $\|\Delta\|_2^2$. Hence

$$\|L(\Delta)\|_2 \;\ge\; \|\Delta\|_2\,. \tag{100}$$

The Hermiticity constraint excludes the r^2-dimensional space of perturbations $\Delta = U_0 W$ with $W^\dagger = -W$. Without the constraint, these directions, arising from unitary mixing of the Kraus operators, would satisfy $L(\Delta) = 0$.

Writing $U_\Theta = U_0 + \mathbf{S}_\Theta$ and $\mathbf{S}_\Theta - \mathbf{S}_\Xi = \Delta + (R_\Theta - R_\Xi)$,

$$r((C_\Theta - C_\Xi)/d_1) - L(\Delta) = U_0(R_\Theta - R_\Xi)^\dagger + (R_\Theta - R_\Xi)U_0^\dagger + \mathbf{S}_\Theta\mathbf{S}_\Theta^\dagger - \mathbf{S}_\Xi\mathbf{S}_\Xi^\dagger.$$

By Lemma 4.5 its Hilbert–Schmidt norm is at most

$$2\sqrt{2}\cdot\tfrac{5}{4}t\,\|\Delta\|_2 + (\|\mathbf{S}_\Theta\|_\infty + \|\mathbf{S}_\Xi\|_\infty)\,\|\mathbf{S}_\Theta - \mathbf{S}_\Xi\|_2 \le \Big(\tfrac{5}{2}\sqrt{2} + \tfrac{17}{8}\cdot\tfrac{9}{8}\Big)t\,\|\Delta\|_2 \le 6t\,\|\Delta\|_2\,.$$

With Equation (100) and $6t \le \frac{3}{16}$ this gives $r\,\|(C_\Theta - C_\Xi)/d_1\|_2 \ge (1-6t)\,\|\Delta\|_2 \ge \frac{1}{2}\,\|\Delta\|_2$.

(iv) $\|U_\Theta - U_0\|_\infty = \|\mathbf{S}_\Theta\|_\infty \le \frac{17}{16}t$ and $\|U_\Theta\|_\infty + \|U_0\|_\infty \le 2 + \sqrt{2}$, so $\|(C_\Theta - C_0)/d_1\|_\infty \le r^{-1}\,\|U_\Theta - U_0\|_\infty\,(\|U_\Theta\|_\infty + \|U_0\|_\infty) \le 8t/r$. The triangle inequality through (C_0/d_1) gives (iv).

(v) Θ has rank at most r, so $\|\Theta\|_2 \le \sqrt{r}\,\|\Theta\|_\infty \le \sqrt{r}t$, and likewise for Ξ. □

4.2.2 Information bounds and prior mass estimates

For a fixed single-query experiment, let $\{T_y\}_{y\in\mathcal{Y}}$ be its tester, as in Equation (12). Since

$$\operatorname{tr}\mathsf{T}(\mathcal{Y}) = \operatorname{tr}\big(I_{\mathrm{B}} \otimes \operatorname{tr}_{\mathrm{R}}(|\psi\rangle\langle\psi|)^{\mathrm{T}}\big) = d_2, \tag{101}$$

the formula

$$\nu(B) := \frac{1}{d_2}\operatorname{tr}\mathsf{T}(B), \qquad B \in \mathcal{B}(\mathcal{Y}), \tag{102}$$

defines a probability measure on $\mathcal{Y}$.

The finite-dimensional Radon–Nikodym theorem gives a measurable operator density $T_y \succeq 0$ such that

$$\mathsf{T}(B) = \int_B T_y \, d\nu(y). \tag{103}$$

Moreover,

$$\operatorname{tr}(T_y) = d_2 \qquad \text{for } \nu\text{-almost every } y. \tag{104}$$

Recall that $\theta \in \mathbb{R}^k$ are the coordinates of Θ in an orthonormal basis of the real Hilbert space $\mathcal{V}$. Then, the outcome law under the channel $\mathcal{E}_\Theta$ has density

$$q_\theta(y) = \operatorname{tr}(T_y C_\Theta) = \frac{d_1}{r}\operatorname{tr}(T_y U_\Theta U_\Theta^\dagger) \tag{105}$$

with respect to ν.

The score is $\nabla_\theta \log q_\theta(y)$ when $q_\theta(y) > 0$ and is defined to be zero otherwise. Its Fisher information matrix is

$$I_1(\theta) = \int_{\{y:q_\theta(y)>0\}} \frac{\nabla_\theta q_\theta(y)\nabla_\theta q_\theta(y)^{\mathrm{T}}}{q_\theta(y)} \, d\nu(y). \tag{106}$$

Lemma 4.7 (One-query Fisher information)**.** *Let $0 < t \le t_0 = 2^{-5}$ and $\Theta \in \mathcal{K}_t$. The Fisher information matrix defined in Equation (106) satisfies*

$$\operatorname{tr} I_1(\theta) \le \frac{16D}{r}. \tag{107}$$

Moreover, the score has mean zero.

Proof. For a fixed $G \succeq 0$, define

$$p(U) := \frac{d_1}{r}\operatorname{tr}(GUU^\dagger). \tag{108}$$

On the real Hilbert space of complex matrices with inner product $\operatorname{Re}\operatorname{tr}(X^\dagger Y)$, one has

$$\nabla_U p = 2\frac{d_1}{r} GU \tag{109}$$

and hence

$$\begin{aligned} \|\nabla_U p\|_2^2 &= 4\left(\frac{d_1}{r}\right)^2 \operatorname{tr}(U^\dagger G^2 U) \\ &\le 4\frac{d_1}{r}\operatorname{tr}(G)\, p(U). \end{aligned} \tag{110}$$

The derivative of $\theta \mapsto U_\Theta$ has operator norm at most 2 by Lemma 4.6. Taking $G = T_y$ therefore gives

$$\|\nabla_\theta q_\theta(y)\|_2^2 \le 16\frac{d_1}{r}\operatorname{tr}(T_y) q_\theta(y). \tag{111}$$

Dividing by $q_\theta(y)$ on its positive set and integrating with respect to ν yields

$$\begin{aligned} \operatorname{tr} I_1(\theta) &\le 16\frac{d_1}{r}\int_{\mathcal{Y}} \operatorname{tr}(T_y)\, d\nu(y) \\ &= 16\frac{d_1 d_2}{r} = \frac{16D}{r}, \end{aligned} \tag{112}$$

where we used Equation (103). If $q_\theta(y) = 0$, then Equation (111) implies

$$\nabla_\theta q_\theta(y) = 0. \tag{113}$$

It remains to prove that the score is centered. For every coordinate θ_a,

$$|\partial_a q_\theta(y)| = |\operatorname{tr}(T_y \partial_a C_\Theta)| \le \operatorname{tr}(T_y)\|\partial_a C_\Theta\|_\infty. \tag{114}$$

The derivative of C_Θ is locally bounded, while $\operatorname{tr}(T_y) = d_2$ for ν-almost every y. Differentiation under the integral is therefore justified. Since

$$\int_{\mathcal{Y}} q_\theta(y)\, d\nu(y) = 1, \tag{115}$$

we obtain

$$\begin{aligned} \mathbb{E}_\theta[\nabla_\theta \log q_\theta(Y)] &= \int_{\mathcal{Y}} \nabla_\theta q_\theta(y)\, d\nu(y) \\ &= \nabla_\theta \int_{\mathcal{Y}} q_\theta(y)\, d\nu(y) = 0. \end{aligned} \tag{116}$$

□

Adaptive transcripts. Let $Z = (W, Y_1, \dots, Y_n)$ be the transcript, where W is the private random seed, sampled independently of θ. For each fixed seed $W = w$ and classical history, apply Equation (102) to the tester selected at that history. Since all outcome spaces are standard Borel, the corresponding Radon–Nikodym densities may be chosen jointly measurable in the history and current outcome. The resulting parameter-independent control kernels induce a parameter-independent reference measure on the transcript space. Relative to this measure, the transcript likelihood is the product of the conditional likelihoods.

The transcript score is therefore the sum of the conditional scores. Since each conditional score has conditional mean zero, the cross terms between distinct rounds vanish. Hence

$$I_Z(\theta) = \sum_{j=1}^{n} \mathbb{E}_\theta\left[I_{Y_j|W,Y_{<j}}(\theta)\right], \qquad \operatorname{tr} I_Z(\theta) \le \frac{16nD}{r}. \tag{117}$$

The expectation in Equation (117) is over W and the previous outcomes. The bound is uniform over all seeds and histories.

Lemma 4.8 (Operator norm of a Gaussian matrix)**.** *Let $G \in \mathbb{C}^{p\times q}$ be a centered Gaussian random matrix obtained as a real-linear image of a real Gaussian vector, and suppose that for every pair of complex unit vectors $|x\rangle, |y\rangle$ both* $\mathrm{Re}(\langle x|G|y\rangle)$ *and* $\mathrm{Im}(\langle x|G|y\rangle)$ *have variance at most v^2. Then for $u > 0$,*

$$\Pr\{\|G\|_\infty > u\} \;\le\; 4\exp\Big(2(p+q)\log 9 - \frac{u^2}{16v^2}\Big). \tag{118}$$

Proof. For a fixed pair of unit vectors $|x\rangle \in \mathbb{C}^p$, $|y\rangle \in \mathbb{C}^q$, the event $|\langle x|G|y\rangle| > s$ forces the real or the imaginary part to exceed $s/\sqrt{2}$ in absolute value, an event of probability at most $4e^{-s^2/(4v^2)}$. The 1/4-covering-net union bound in Equation (21), with $s = u/2$, therefore gives

$$\Pr\{\|G\|_\infty > u\} \le 9^{2(p+q)}\, 4e^{-u^2/(16v^2)},$$

which is Equation (118). □

Fact 4.9 (Bakry–Émery log-Sobolev criterion on a convex domain [KM16, Theorem 2.1])**.** *Let $\Omega \subseteq \mathbb{R}^k$ be open and convex, and let $\Phi \in C^2(\Omega)$ satisfy*

$$0 < Z_\Phi := \int_\Omega e^{-\Phi(\theta)}\, d\theta < \infty \tag{119}$$

and

$$\nabla^2\Phi(\theta) \succeq \varkappa I_k \qquad \textit{for every } \theta \in \Omega, \tag{120}$$

where $\varkappa > 0$. Let π be the probability measure on Ω with density $Z_\Phi^{-1}e^{-\Phi}$. Then, for every $f \in C^1(\Omega)$,

$$\mathrm{Ent}_\pi(f^2) \le \frac{2}{\varkappa}\int_\Omega \|\nabla f\|_2^2\, d\pi, \tag{121}$$

where

$$\mathrm{Ent}_\pi(g) := \int_\Omega g\log g\, d\pi - \left(\int_\Omega g\, d\pi\right)\log\left(\int_\Omega g\, d\pi\right). \tag{122}$$

Lemma 4.10 (Mass and log-Sobolev inequality for the prior)**.** *In the orthonormal coordinates fixed above, let γ_σ and μ_t be the measures defined in Equations (92) and (93), and set*

$$\Omega_t := \mathrm{int}_{\mathbb{R}^k}(\mathcal{K}_t). \tag{123}$$

Let $0 < t \le t_0 = 2^{-5}$. Use the space $\mathcal{V}$, reshaped perturbation T_Θ, and domain $\mathcal{K}_t$ from Equations (87), (88) and (91). Then

$$\gamma_\sigma\{\|\Theta\|_\infty \le t/2,\ \|T_\Theta\|_\infty \le t/2\} \;\ge\; \tfrac{1}{2}, \qquad \textit{so in particular} \qquad \gamma_\sigma(\mathcal{K}_t) \;\ge\; \tfrac{1}{2}. \tag{124}$$

Then μ_t is supported on $\mathcal{K}_t$, satisfies $\mu_t \le 2\gamma_\sigma$ as measures, and obeys the logarithmic Sobolev inequality

$$\mathrm{Ent}_{\mu_t}(f^2) \le 2\sigma^2\int_{\Omega_t} \|\nabla f\|_2^2\, d\mu_t \tag{125}$$

for every $f \in C^1(\Omega_t)$.

Proof. For unit vectors $|x\rangle \in \mathbb{C}^D$, $|y\rangle \in \mathbb{C}^r$ we have $\mathrm{Re}(\langle x|\Theta|y\rangle) = \langle |x\rangle\langle y|, \Theta\rangle_{\mathbb{R}}$, whose coefficient in the coordinates of $\mathcal{V}$ is the orthogonal projection of $|x\rangle\langle y|$ onto $\mathcal{V}$, of Hilbert–Schmidt norm at most $\||x\rangle\langle y|\|_2 = 1$; the same holds for the imaginary part with $i|x\rangle\langle y|$. So Θ satisfies the hypothesis of Lemma 4.8 with $v = \sigma$, $p = D$, $q = r$. Since $r \le D$, $p + q \le 2D$, and $u = t/2$ with Equation (92) gives $u^2/(16v^2) = t^2/(64\sigma^2) = 16D$, hence

$$\Pr\{\|\Theta\|_\infty > t/2\} \le 4\exp(4D\log 9 - 16D) \le 4e^{-7D},$$

as $4\log 9 < 9$. For T_Θ: by Equation (87) the corresponding coefficient has norm at most $\sqrt{d_1/r}$, so the hypothesis holds with $v^2 = (d_1/r)\sigma^2$, $p = rd_2$, $q = d_1$. *Here the second feasibility condition enters:* $d_1 \le rd_2$ gives $p + q \le 2rd_2$, and

$$\frac{u^2}{16v^2} = \frac{t^2 r}{64 d_1 \sigma^2} = \frac{r}{d_1} \cdot 16D = 16rd_2,$$

so $\Pr\{\|T_\Theta\|_\infty > t/2\} \le 4\exp(4rd_2\log 9 - 16rd_2) \le 4e^{-7rd_2}$. Both $D \ge 2$ and $rd_2 \ge 2$, so the two exceptional probabilities sum to at most $8e^{-14} < \frac{1}{2}$, proving Equation (124). The event in Equation (124) is contained in $\mathcal{K}_t$, and $\mu_t \le 2\gamma_\sigma$ follows from $\gamma_\sigma(\mathcal{K}_t) \ge \frac{1}{2}$.

The set $\mathcal{K}_t$ is a full-dimensional convex body in $\mathbb{R}^k$. Indeed, under the chosen orthonormal coordinates, $\|\theta\|_2 = \|\Theta\|_2$, and the Euclidean ball centered at the origin with radius $t\min\left\{1, \sqrt{\frac{r}{d_1}}\right\}$ is contained in $\mathcal{K}_t$, because $\|\Theta\|_\infty \le \|\Theta\|_2$, $\|T_\Theta\|_\infty \le \|T_\Theta\|_2 = \sqrt{\frac{d_1}{r}}\|\Theta\|_2$. Consequently, Ω_t is open and convex. Moreover, the boundary of the full-dimensional convex body $\mathcal{K}_t$ has Lebesgue measure zero and hence also γ_σ-measure zero. Therefore, μ_t agrees almost everywhere with the probability measure on Ω_t having density proportional to

$$\exp\left(-\frac{\|\theta\|_2^2}{2\sigma^2}\right). \tag{126}$$

The corresponding potential has Hessian $\sigma^{-2}I_k$. Applying Fact 4.9 on Ω_t with $\varkappa = \sigma^{-2}$ proves Equation (125). $\square$

Proposition 4.11 (Mutual information of an adaptive transcript)**.** *Let* $0 < t \le t_0 = 2^{-5}$*. Let* $\theta \sim \mu_t$*, where* μ_t *is the conditional Gaussian prior in Equation (92) on the parameter domain* $\mathcal{K}_t$ *of Equation (91). The vector* θ *gives orthonormal coordinates for the real Hilbert–Schmidt inner product on* $\mathcal{V}$ *from Equation (88). Let* Z *be the transcript of any* n*-query adaptive incoherent protocol applied to the channel* $\mathcal{E}_\Theta$ *defined in Equation (90). Then its mutual information with the parameter satisfies*

$$I(\theta; Z) \;\le\; \frac{n\,t^2}{128\,r}.$$

Proof. Fix a seed value $W = w$ outside a null set, and let λ_w be the parameter-independent reference probability measure on transcripts obtained from the control kernels. Write

$$q_\theta(z) := \frac{dP_{Z|\theta, W=w}}{d\lambda_w}(z). \tag{127}$$

For λ_w-almost every z, the successive tester densities are fixed. Hence $q_\theta(z)$ is a finite product of functions of the form Equation (105). It is therefore nonnegative and real analytic in θ, and is bounded on the compact set $\mathcal{K}_t$.

For $\eta > 0$, define

$$f_{\eta,z}(\theta) := \sqrt{q_\theta(z) + \eta}. \tag{128}$$

Then $f_{\eta,z} \in C^1(\Omega_t)$, so Equation (125) gives

$$\mathrm{Ent}_{\mu_t}(q_\cdot(z) + \eta) \le \frac{\sigma^2}{2} \int_{\Omega_t} \frac{\|\nabla_\theta q_\theta(z)\|_2^2}{q_\theta(z) + \eta}\, d\mu_t(\theta). \tag{129}$$

Because $q_\theta(z)$ is nonnegative and differentiable on the open set Ω_t, every point at which $q_\theta(z) = 0$ is a local minimum. Therefore,

$$q_\theta(z) = 0 \quad \Longrightarrow \quad \nabla_\theta q_\theta(z) = 0. \tag{130}$$

Because $q_\theta(z)$ is bounded on the compact set $\mathcal{K}_t$, dominated convergence applies to the entropy term on the left-hand side of Equation (129). Moreover,

$$\frac{\|\nabla_\theta q_\theta(z)\|_2^2}{q_\theta(z) + \eta} \uparrow \frac{\|\nabla_\theta q_\theta(z)\|_2^2}{q_\theta(z)} \mathbf{1}_{\{q_\theta(z)>0\}} \tag{131}$$

as $\eta \downarrow 0$, because $q_\theta(z) = 0$ implies $\nabla_\theta q_\theta(z) = 0$. Hence monotone convergence applies to the right-hand side. Hence

$$\mathrm{Ent}_{\mu_t}(q_\cdot(z)) \le \frac{\sigma^2}{2} \int_{\{\theta : q_\theta(z)>0\}} \frac{\|\nabla_\theta q_\theta(z)\|_2^2}{q_\theta(z)}\, d\mu_t(\theta). \tag{132}$$

Integrating with respect to λ_w and applying Tonelli's theorem gives

$$\begin{aligned} I(\theta; Z \mid W = w) &= \int \mathrm{Ent}_{\mu_t}(q_\cdot(z))\, d\lambda_w(z) \\ &\le \frac{\sigma^2}{2} \int_{\Omega_t} \mathrm{tr}\, I_{Z|W=w}(\theta)\, d\mu_t(\theta) \\ &\le \frac{\sigma^2}{2} \frac{16nD}{r}. \end{aligned} \tag{133}$$

Since W is independent of θ and is included in the transcript,

$$I(\theta; Z) = I(\theta; Z \mid W). \tag{134}$$

Averaging over W and using $\sigma^2 = t^2/(1024D)$ therefore gives

$$I(\theta; Z) \le \frac{\sigma^2}{2} \frac{16nD}{r} = \frac{nt^2}{128r}. \tag{135}$$

□

Lemma 4.12 (Prior mass of trace-norm balls)**.** *Assume Equation (1). Let $0 < t \le t_0 = 2^{-5}$ and $0 < \epsilon \le t/M$ with $M := 2^{21}$. Use the Choi operator $C_\Theta = C_{\mathcal{E}_\Theta}$, parameter domain $\mathcal{K}_t$, and conditional Gaussian prior μ_t defined in Equations (90) to (92). Then for every $\varsigma \in \mathbb{C}^{D\times D}$,*

$$\mu_t\{\Theta \in \mathcal{K}_t : \ \frac{1}{d_1} \|C_\Theta - \varsigma\|_1 \le \epsilon\} \ \le \ 2\, e^{-Dr/2}.$$

Proof. Write $\mathcal{S}$ for the set in question and assume it is nonempty, else there is nothing to prove; fix $\Xi \in \mathcal{S}$. For $\Theta \in \mathcal{S}$ the triangle inequality gives $\frac{1}{d_1}\|C_\Theta - C_\Xi\|_1 \le 2\epsilon$. Combining Lemma 4.6 (iii), the singular-value inequality $\|H\|_2^2 \le \|H\|_\infty \|H\|_1$, and Lemma 4.6 (iv),

$$\frac{\|\Theta - \Xi\|_2^2}{4r^2} \;\le\; \frac{1}{d_1^2}\|C_\Theta - C_\Xi\|_2^2 \;\le\; \frac{1}{d_1}\|C_\Theta - C_\Xi\|_\infty \,\frac{1}{d_1}\|C_\Theta - C_\Xi\|_1 \;\le\; \frac{16t}{r}\cdot 2\epsilon,$$

so $\|\Theta - \Xi\|_2^2 \le 128\, rt\epsilon$. Thus $\mathcal{S}$ is contained in the Euclidean ball $\mathcal{B}(\xi,\rho) \subset \mathbb{R}^k$ of radius $\rho = \sqrt{128\, rt\epsilon}$.

The Gaussian density is everywhere at most $(2\pi\sigma^2)^{-k/2}$. Consequently, for any center ξ,

$$\gamma_\sigma(\mathcal{B}(\xi,\rho)) \le \frac{\rho^k}{2^{k/2}\sigma^k\Gamma(k/2+1)} \le \left(\frac{\rho\sqrt{e}}{\sigma\sqrt{k}}\right)^k.$$

Here we used the volume of a Euclidean ball [EG15, Sec. 2.4] and $\Gamma(k/2+1) \ge (k/(2e))^{k/2}$. By Lemma 4.3 and Equation (92), $\sigma\sqrt{k} \ge \frac{t}{32\sqrt{D}}\sqrt{\frac{3}{4}Dr} = \frac{t\sqrt{3r}}{64} \ge t\sqrt{r}/38$, hence

$$\frac{\rho\sqrt{e}}{\sigma\sqrt{k}} \;\le\; \frac{38\sqrt{e}\sqrt{128\, rt\epsilon}}{t\sqrt{r}} \;=\; 38\sqrt{128e}\,\sqrt{\epsilon/t} \;\le\; 709\, M^{-1/2} \;\le\; \tfrac{1}{2},$$

since $M = 2^{21}$ and $709\cdot 2^{-21/2} < \frac{1}{2}$. Using $\mu_t \le 2\gamma_\sigma$ from Lemma 4.10 and $k \ge \frac{3}{4}Dr$ from Lemma 4.3,

$$\mu_t(\mathcal{S}) \;\le\; 2\cdot 2^{-k} \;\le\; 2e^{-\frac{3}{4}(\log 2)Dr} \;\le\; 2e^{-Dr/2}. \qquad \square$$

Lemma 4.13 (Fano's inequality for reconstruction sets)**.** *Let $\theta \sim \mu$ be a random parameter, let Y be data with conditional law $P_{Y|\theta}$, and let $\widehat{\theta} = \widehat{\theta}(Y)$ be an estimator. Fix $\epsilon > 0$ and $0 \le \delta < 1$. For each possible estimate ϑ, let $\mathcal{B}(\vartheta,\epsilon)$ be the measurable set of parameter values reconstructed to accuracy ϵ by that estimate. Assume that the success probability $\Pr\{\theta \in \mathcal{B}(\widehat{\theta}(Y),\epsilon)\} \ge 1-\delta$. If $\mu(\mathcal{B}(\vartheta,\epsilon)) \le \alpha_0$ for every ϑ, where $0 < \alpha_0 \le 1$, then*

$$I(\theta;Y) \;\ge\; (1-\delta)\log\frac{1}{\alpha_0} - \log 2. \tag{136}$$

Proof. Let $\mathcal{T} = \mathbf{1}_{\{\theta\in\mathcal{B}(\widehat{\theta}(Y),\epsilon)\}}$, a binary function of (θ, Y). By the data-processing inequality for relative entropy applied to the joint law $P_{\theta Y}$ and the product $\mu\otimes P_Y$,

$$I(\theta;Y) = D_{\mathrm{KL}}(P_{\theta Y}\,\|\,\mu\otimes P_Y) \;\ge\; h(P_{\theta Y}\{\mathcal{T}=1\},\ (\mu\otimes P_Y)\{\mathcal{T}=1\}),$$

where $h(p,q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$. By hypothesis $P_{\theta Y}\{\mathcal{T}=1\} \ge 1-\delta$, while $(\mu\otimes P_Y)\{\mathcal{T}=1\} = \mathbb{E}_Y\,\mu(\mathcal{B}(\widehat{\theta}(Y),\epsilon)) \le \alpha_0$. For $p = P_{\theta Y}\{\mathcal{T}=1\}$ and $q = (\mu\otimes P_Y)\{\mathcal{T}=1\}$, use $h(p,q) \ge p\log(1/q) - \log 2$ to obtain Equation (136). $\square$

Statement on the use of AI

The upper-bound proof for channels with gapped Choi spectrum was developed by the human authors. After being provided with this proof as context, GPT-6 Astra generated the proof strategy for the general upper bound. GPT-5.6 Sol generated the lower-bound proof after being provided with the earlier construction from [CGO+26] and the suggestion to pursue a van Trees-type argument. The new hard family of channels used in the lower bound was generated by GPT-6 Astra. The authors subsequently checked, revised, and completed all AI-generated arguments and take full responsibility for the correctness and presentation of the results.

References

[AK16] Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programs. *Journal of the ACM*, 63(2), 2016. Article 12.

[Bha97] Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, New York, 1997.

[BPGM26] Carlos Bravo-Prieto, Weiyuan Gong, and Antonio Anna Mele. Quantum memory advantage for quantum process tomography, 2026. arXiv:2607.13476 [quant-ph].

[CGO+26] Kean Chen, Filippo Girardi, Aadil Oufkir, Nengkun Yu, and Zhicheng Zhang. Quantum channel tomography: optimal bounds and a Heisenberg-to-classical phase transition, 2026. arXiv:2604.17369 [quant-ph].

[CHL+23] Sitan Chen, Brice Huang, Jerry Li, Allen Liu, and Mark Sellke. When does adaptivity help for quantum state learning? In *Proceedings of the 64th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 391–404, 2023.

[CN97] Isaac L. Chuang and Michael A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44(11–12):2455–2467, 1997.

[DS01] Kenneth R. Davidson and Stanisław J. Szarek. Local operator theory, random matrices and banach spaces. In *Handbook of the Geometry of Banach Spaces*, volume 1, pages 317–366. North-Holland, Amsterdam, 2001.

[EG15] Lawrence C. Evans and Ronald F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, Boca Raton, FL, revised edition, 2015.

[EHW+20] Jens Eisert, Dominik Hangleiter, Nathan Walk, Ingo Roth, Damian Markham, Rhea Parekh, Ulysse Chabaud, and Elham Kashefi. Quantum certification and benchmarking. *Nature Reviews Physics*, 2(7):382–390, July 2020.

[FGLE12] Steven T. Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New J. Phys.*, 14(9):095022, September 2012.

[FOF25] Omar Fawzi, Aadil Oufkir, and Daniel Stilck França. Lower Bounds on Learning Pauli Channels With Individual Measurements. *IEEE Trans. Inf. Theory*, 71(4):2642–2661, January 2025.

[GKKT20] Mădălin Guţă, Jonas Kahn, Richard Kueng, and Joel A. Tropp. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical*, 53:204001, 2020.

[HCP23] Hsin-Yuan Huang, Sitan Chen, and John Preskill. Learning to predict arbitrary quantum processes. *PRX Quantum*, 4:040337, 2023.

[HHJ+17] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63:5628–5641, 2017.

[HKOT23] Jeongwan Haah, Robin Kothari, Ryan O'Donnell, and Ewin Tang. Query-optimal estimation of unitary channels in diamond distance. In *Proceedings of the 64th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 363–390, 2023.

[HLS+26] Entong He, Zihao Li, Noam Scully, Sisi Zhou, and Yuxiang Yang. Optimal classical shadow estimation of unitary channels at Heisenberg limit, 2026. arXiv:2606.13638 [quant-ph].

[KKEG19] Martin Kliesch, Richard Kueng, Jens Eisert, and David Gross. Guaranteed recovery of quantum processes from few measurements. *Quantum*, 3:171, 2019.

[KLMR26] Ufuk Keskin, Jason Luo, Mahbod Majid, and Matthew Radzihovsky. Tight lower bounds for state tomography with limited entanglement, 2026. arXiv:2609.05718 [quant-ph].

[KM16] Alexander V. Kolesnikov and Emanuel Milman. Riemannian metrics on convex sets with applications to poincaré and log-sobolev inequalities. *Calculus of Variations and Partial Differential Equations*, 55, 2016. Article 77.

[KRT17] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *Applied and Computational Harmonic Analysis*, 42(1):88–116, 2017.

[KTCT23] Jonathan Kunjummen, Minh C. Tran, Daniel Carney, and Jacob M. Taylor. Shadow process tomography of quantum channels. *Physical Review A*, 107:042403, 2023.

[LM00] Béatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000.

[LN25] Angus Lowe and Ashwin Nayak. Lower bounds for learning quantum states with single-copy measurements. *ACM Transactions on Computation Theory*, 17(1):1–42, 2025. Article 7.

[MB25] Antonio Anna Mele and Lennart Bittel. Optimal learning of quantum channels in diamond distance, 2025. arXiv:2512.10214 [quant-ph].

[Mec19] Elizabeth S. Meckes. *The Random Matrix Theory of the Classical Compact Groups*, volume 218 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2019.

[MRL08] M. Mohseni, A. T. Rezakhani, and D. A. Lidar. Quantum-process tomography: Resource analysis of different strategies. *Physical Review A*, 77:032322, 2008.

[NZ26] Ashwin Nayak and Xingyu Zhou. Optimal low-rank quantum state tomography with bounded-sample joint measurements, 2026. arXiv:2609.10514 [quant-ph].

[OG26] Aadil Oufkir and Filippo Girardi. Improved lower bounds for learning quantum channels in diamond distance, 2026. arXiv:2601.04180 [quant-ph].

[Ouf23] Aadil Oufkir. Sample-optimal quantum process tomography with non-adaptive incoherent measurements. In *Proceedings of the 2023 IEEE International Symposium on Information Theory (ISIT)*, pages 1919–1924, 2023.

[OW16] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the 48th ACM Symposium on Theory of Computing (STOC)*, pages 899–912, 2016.

[PCZ97] J. F. Poyatos, J. I. Cirac, and P. Zoller. Complete characterization of a quantum process: The two-bit quantum gate. *Physical Review Letters*, 78:390–393, 1997.

[SGS12] John A. Smolin, Jay M. Gambetta, and Graeme Smith. Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise. *Physical Review Letters*, 108:070502, 2012.

[SSKKG22] Trystan Surawy-Stepney, Jonas Kahn, Richard Kueng, and Mădălin Guţă. Projected least-squares quantum process tomography. *Quantum*, 6:844, 2022.

[TRW05] Koji Tsuda, Gunnar Rätsch, and Manfred K. Warmuth. Matrix exponentiated gradient updates for on-line learning and bregman projection. *Journal of Machine Learning Research*, 6(34):995–1018, 2005.

[Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Number 47 in Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2018.

[Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, Cambridge, 2018.