

Quantum memory advantage for quantum process tomography

Carlos Bravo-Prieto,¹ Weiyan Gong,² and Antonio Anna Mele^{1,3}

¹*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*

²*School of Engineering and Applied Sciences, Harvard University, Allston, MA 02134, USA*

³*Google Quantum AI, Venice, California 90291, USA*

Learning an unknown quantum process is a common task in quantum technologies. It underlies device characterization, validation, benchmarking, and the certification of noisy quantum dynamics. For an unknown channel $\Phi : \mathcal{L}(\mathbb{C}^{d_{\text{in}}}) \rightarrow \mathcal{L}(\mathbb{C}^{d_{\text{out}}})$, a natural goal is to output an estimator $\tilde{\Phi}$ that is close to Φ in diamond norm. A central question is which physical resources are responsible for a more efficient process tomography. In particular, coherent protocols may preserve quantum information across different uses of the channel and process the resulting systems jointly. Incoherent protocols, by contrast, must measure after each channel use and keep only classical information between rounds. The remaining issue is whether classical adaptive protocols can compensate for the absence of quantum memory.

Results: We prove that they cannot. Any adaptive incoherent protocol that estimates every channel $\Phi : \mathcal{L}(\mathbb{C}^{d_{\text{in}}}) \rightarrow \mathcal{L}(\mathbb{C}^{d_{\text{out}}})$ to diamond-norm error ε with constant success probability requires

$$\Omega\left(\frac{d_{\text{in}}^3 d_{\text{out}}^3}{\varepsilon^2}\right)$$

channel queries, even when each individual channel use is allowed to be ancilla-assisted and the choice of input state and measurement at each round may depend arbitrarily on all previous classical outcomes. This matches the best known incoherent upper bounds up to logarithmic factors [1, 2], and therefore establishes the optimal incoherent sample complexity $\tilde{\Theta}(d_{\text{in}}^3 d_{\text{out}}^3 / \varepsilon^2)$. Since coherent process tomography with quantum memory achieves $\Theta(d_{\text{in}}^2 d_{\text{out}}^2 / \varepsilon^2)$ for the same worst-case diamond-norm task [3, 4], our result gives a strict learning separation between tomography with and without quantum memory.

Main ideas: The proof combines the process-tester formalism with a Bayesian anti-concentration argument that is robust to full classical adaptivity. We construct a local family of hard channels around the completely depolarizing channel, represent every one-shot incoherent experiment as a tester acting linearly on the Choi operator, and then condition on a full transcript of the adaptive protocol. Once the transcript is fixed, the adaptive choices become a deterministic sequence of tester elements. The key technical step is a transcript-wise likelihood tilt bound over local Schatten neighborhoods. This shows that, below the scale $d_{\text{in}}^3 d_{\text{out}}^3 / \varepsilon^2$, the posterior distribution cannot concentrate in the Choi trace ball that would be forced by any uniformly diamond-norm accurate estimator.

Potential impact: The result identifies quantum memory as the resource responsible for the known improvement in full process tomography. It also shows that the optimal incoherent protocol need not be adaptive, up to logarithmic factors: classical adaptivity alone cannot close the gap to coherent tomography. More broadly, the transcript-wise posterior anti-concentration method may be useful for other adaptive quantum learning tasks in which each round is quantum but the learner stores only classical information.

I. Problem setting and resource model. Let $A \simeq \mathbb{C}^{d_{\text{in}}}$ and $B \simeq \mathbb{C}^{d_{\text{out}}}$, and write $D := d_{\text{in}} d_{\text{out}}$. We use the normalized Choi representation

$$J_{\Phi} = (\text{id}_{A'} \otimes \Phi)(|\Omega\rangle\langle\Omega|), \quad |\Omega\rangle = \frac{1}{\sqrt{d_{\text{in}}}} \sum_{i=1}^{d_{\text{in}}} |i\rangle_{A'} |i\rangle_A,$$

so that Φ is a channel if and only if $J_{\Phi} \geq 0$ and $\text{Tr}_B J_{\Phi} = I_A / d_{\text{in}}$ [5, 6]. We will use the elementary implication $\|J_{\Phi} - J_{\Psi}\|_1 \leq \|\Phi - \Psi\|_{\diamond}$, which allows us to convert diamond-norm accuracy into Choi trace-norm concentration.

An adaptive incoherent protocol uses the unknown channel once per round. At round t , after observing a classical history $h_{t-1} = (y_1, \dots, y_{t-1})$, the learner chooses an ancilla R_t , an input state $\rho_{h_{t-1}} \in \mathcal{D}(R_t \otimes A)$, and a POVM on $R_t \otimes B$. The channel is applied once, the output is measured immediately, and the outcome is appended to the transcript. The learner may perform unlimited classical computation and may choose all later experiments as arbitrary functions of the transcript, but no quantum information is retained between different channel uses.

II. Main theorem. Our main theorem rules out any improvement from classical adaptivity in full incoherent process tomography.

Theorem 1 (Adaptive incoherent process tomography lower bound). *There exist universal constants $c > 0$, $\varepsilon_{\star} > 0$, and $D_0 \in \mathbb{N}$ such that the following holds. Let $d_{\text{out}} \geq 2$ and $D = d_{\text{in}} d_{\text{out}} \geq D_0$. Suppose an adaptive incoherent*

protocol, allowing arbitrary standard Borel outcome spaces, uses T queries to an unknown channel $\Phi : \mathcal{L}(A) \rightarrow \mathcal{L}(B)$ and outputs a linear-map-valued estimator $\hat{\Phi}$. If, for every channel Φ ,

$$\mathbb{P}_\Phi \left[\|\hat{\Phi} - \Phi\|_\diamond \leq \varepsilon \right] \geq \frac{2}{3}$$

for some $0 < \varepsilon \leq \varepsilon_\star$, then

$$T \geq c \frac{D^3}{\varepsilon^2} = c \frac{d_{\text{in}}^3 d_{\text{out}}^3}{\varepsilon^2}.$$

Together with the known incoherent upper bounds [1, 2], Theorem 1 provides the optimal complexity of incoherent quantum process tomography up to logarithmic factors. It also identifies the precise resource behind the coherent improvement. Because the theorem already allows arbitrary classical adaptivity, arbitrary ancilla-assisted inputs within each round, and arbitrary outcome-dependent measurements, the remaining separation from coherent tomography cannot be attributed to adaptivity. It is a quantum memory advantage: retaining and jointly processing quantum information across channel uses improves the query complexity by a factor $\tilde{\Theta}(D)$.

III. Proof strategy. The proof proceeds in four steps, which are summarized in the Figure below.

Step 1: Reduce every incoherent round to a process tester. A single incoherent use of the channel can be written in tester form, the one-slot case of the quantum comb formalism [7, 8]. For any input state on $R \otimes A$ and POVM on $R \otimes B$, there exist positive operators $\{T_y\}_y$ on $A \otimes B$ and a state $\tau \in \mathcal{D}(A)$ such that

$$\sum_y T_y = \tau^\top \otimes I_B, \quad \mathbb{P}_\Phi[Y = y] = d_{\text{in}} \text{Tr}(T_y J_\Phi).$$

Thus the statistical dependence on Φ is linear in the Choi operator.

This representation is especially useful for adaptive protocols. Once a complete transcript $z = (y_1, \dots, y_T)$ is fixed, the adaptive choices made along that transcript are fixed as well. Writing $T_t(z)$ for the tester element used at round t along transcript z , the likelihood ratio between two channels factors as

$$\frac{P_\Phi(z)}{P_{\Phi_0}(z)} = \prod_{t=1}^T \frac{\text{Tr}(T_t(z) J_\Phi)}{\text{Tr}(T_t(z) J_{\Phi_0})}.$$

This identity is the mechanism by which we handle full adaptivity: after conditioning on the transcript, the protocol becomes a deterministic sequence of one-round tests.

Step 2: Construct a hard local family of channels. The hard family is centered at the completely depolarizing channel, namely $\Phi_{\text{dep}}(\rho) = \text{Tr}(\rho) I_B / d_{\text{out}}$, and $J_{\text{dep}} = I_{AB} / D$. We decompose the output space as $B = B_0 \oplus B_1 \oplus B_{\text{rem}}$, where $\dim B_0 = \dim B_1 = \lfloor \frac{d_{\text{out}}}{2} \rfloor$, and set $K_0 = A \otimes B_0$, $K_1 = A \otimes B_1$, with $r = \dim K_0 = \dim K_1 = \Theta(D)$. For $X \in \mathbb{C}^{r \times r}$, define the off-diagonal Hermitian embedding

$$E(X) = \begin{pmatrix} 0 & X \\ X^\dagger & 0 \end{pmatrix}$$

on $K_0 \oplus K_1$, extended by zero on the remaining subspace. The key algebraic property is $\text{Tr}_B E(X) = 0$. Hence perturbations by $E(X)$ preserve the trace-preserving constraint. For a small universal constant $\sigma > 0$, define $J_X = I_{AB} / D + \sigma E(X) / D$, with X restricted to an operator-norm ball. These are valid full-rank channel Choi operators. Moreover, for nearby parameters X_0 and $X_0 + W$, then $\|J_{X_0+W} - J_{X_0}\|_1 = 2\sigma \|W\|_1 / D$. Thus Choi trace balls become Schatten-1 balls in the local matrix parameter W .

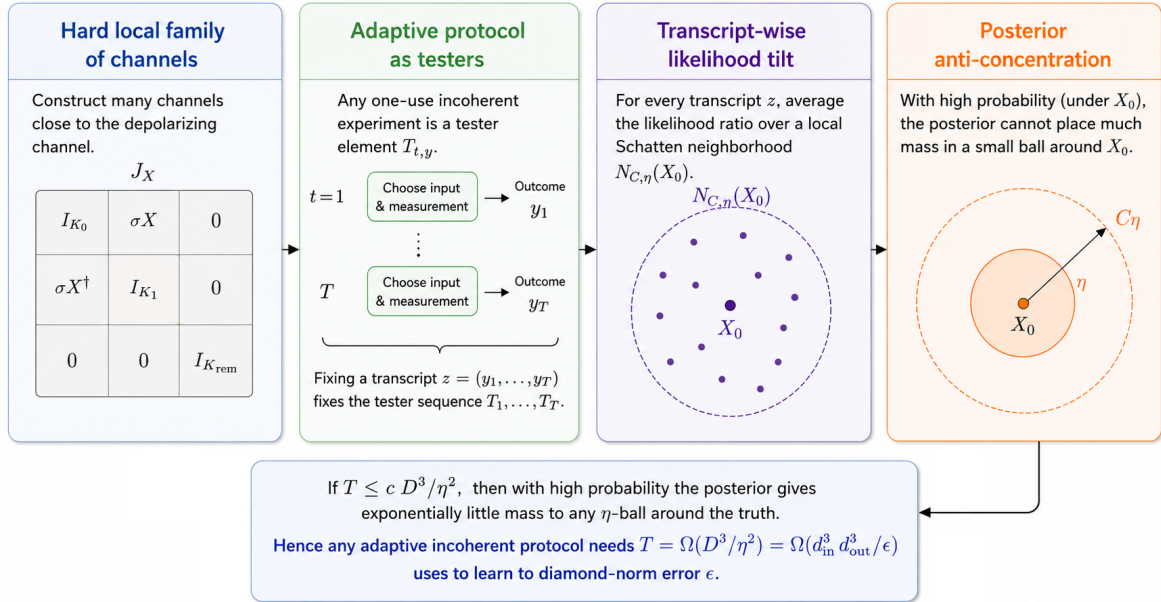
Step 3: Prove a transcript-wise likelihood tilt bound. We place a truncated complex Ginibre prior on X . The proof compares posterior mass in a small Choi trace ball to posterior mass in a larger, regularized neighborhood

$$N_{C,\eta} = \left\{ W : \|W\|_1 \leq CL_\eta, \|W\|_\infty \leq \frac{CL_\eta}{4r} \right\}, \quad L_\eta = \frac{\eta D}{2\sigma}.$$

The Schatten-1 constraint controls trace distance, while the operator-norm constraint keeps one-step likelihood perturbations uniformly small.

For every fixed transcript z and every regular center X_0 , define

$$\Lambda_W(z) = \prod_{t=1}^T \frac{\text{Tr}(T_t(z) J_{X_0+W})}{\text{Tr}(T_t(z) J_{X_0})}.$$



The main technical lemma states that, for W uniformly distributed on $N_{C,\eta}$, $\mathbb{E}_W[\Lambda_W(z)] \geq \exp(-KC^2\eta^2T/D)$ for a universal constant K . The factor $1/D$ per channel use is essential. It comes from the isotropic covariance of the local perturbations, the off-diagonal embedding $E(W)$, and the tester normalization.

This likelihood loss is balanced against a geometric volume gain. The regularized neighborhood has volume larger than the small Schatten-1 ball by a factor at least $C^{2r^2} \exp(-c_{\text{vol}}r^2)$, using standard volume-radius estimates for complex Schatten balls [9]. Since $r = \Theta(D)$, the volume gain is exponential in D^2 , while the likelihood loss is exponential in η^2T/D .

Step 4: Convert likelihood control into posterior anti-concentration. Balancing the two exponents gives the threshold $T = \Omega(D^3/\eta^2)$. Below this threshold, the posterior cannot concentrate near the true channel. More precisely, there are universal constants $\kappa > 0$ and $c_{\text{ac}} > 0$ such that, whenever $T \leq c_{\text{ac}}D^3/\eta^2$, the posterior ν_Z satisfies, for every regular center X_0 , $\mathbb{P}_{X_0}[\nu_Z(B_\eta(X_0)) \leq e^{-\kappa D^2}] \geq 1 - e^{-D^2}$. Here $B_\eta(X_0)$ is the Choi trace ball of radius η around J_{X_0} within the hard family.

Finally, suppose for contradiction that an adaptive incoherent estimator succeeded uniformly in diamond norm using fewer than order D^3/ϵ^2 queries. Applying the posterior anti-concentration statement with $\eta = 2\epsilon$, the Bayes probability of Choi trace-norm success is at most $1/3$. But the assumed worst-case diamond-norm guarantee implies Bayes Choi trace-norm success at least $2/3$, because $\|J_{\hat{\Phi}} - J_\Phi\|_1 \leq \|\hat{\Phi} - \Phi\|_\diamond$. This contradiction proves Theorem 1.

IV. Discussion and outlook. Our result shows that full process tomography exhibits a genuine quantum memory advantage. The conclusion is stronger than a lower bound against a particular non-adaptive measurement scheme. The learner may use arbitrary ancilla-assisted inputs in each round, arbitrary measurements after each channel use, unlimited classical computation, and fully outcome-dependent choices of future experiments. Nevertheless, without the ability to preserve quantum information across channel invocations, the optimal query complexity remains $\tilde{\Theta}(d_{\text{in}}^3 d_{\text{out}}^3/\epsilon^2)$. The improvement to $\Theta(d_{\text{in}}^2 d_{\text{out}}^2/\epsilon^2)$ therefore comes from quantum memory and joint quantum processing.

The techniques may apply beyond full process tomography. Many quantum learning problems have an adaptive incoherent structure: each round is a quantum experiment, but the learner stores only a classical outcome before selecting the next experiment. Examples include adaptive channel property testing, Hamiltonian and noise learning, tomography of structured process families, and certification tasks with adaptive measurement choices. Whenever each branch admits a tester representation and the local parameter space has sufficient isotropic volume, transcript-wise posterior anti-concentration may provide a route to lower bounds robust to full classical adaptivity.

Several directions remain open. One is to characterize intermediate memory regimes: how much quantum memory is necessary to improve over the incoherent rate? Another is to study structured channel classes, such as unitary channels, low-Kraus-rank channels, Pauli or sparse noise models, local noise processes, and Markovian dynamics. In such settings, classical adaptivity may interact with structure in ways that are absent in worst-case full tomography. A final direction is robustness. Coherent protocols assume the ability to store and jointly process quantum information across channel uses, whereas near-term devices have noisy, finite-depth, and imperfect memory. Extending the separation to noisy or short-lived memory models would help quantify how much coherence is required to realize a practical quantum memory advantage.

-
- [1] T. Surawy-Stepney, J. Kahn, R. Kueng, and M. Guta, Projected least-squares quantum process tomography, [Quantum](#) **6**, 844 (2022).
 - [2] A. Oufkir, Sample-optimal quantum process tomography with non-adaptive incoherent measurements, in [2023 IEEE International Symposium on Information Theory \(ISIT\)](#) (IEEE, 2023) pp. 1919–1924.
 - [3] A. A. Mele and L. Bittel, Optimal learning of quantum channels in diamond distance, [arXiv:2512.10214](#) (2025).
 - [4] K. Chen, F. Girardi, A. Oufkir, N. Yu, and Z. Zhang, Quantum channel tomography: optimal bounds and a heisenberg-to-classical phase transition, [arXiv:2604.17369](#) (2026).
 - [5] M.-D. Choi, Completely positive linear maps on complex matrices, [Linear algebra and its applications](#) **10**, 285 (1975).
 - [6] J. Watrous, [The theory of quantum information](#) (Cambridge university press, 2018).
 - [7] G. Gutoski and J. Watrous, Toward a general theory of quantum games, in [Proceedings of the thirty-ninth annual ACM symposium on Theory of computing](#) (2007) pp. 565–574.
 - [8] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Theoretical framework for quantum networks, [Physical Review A](#) **80**, 022339 (2009).
 - [9] Z. Kabluchko, J. Prochno, and C. Thaele, Exact asymptotic volume and volume ratio of Schatten unit balls, [Journal of Approximation Theory](#) **257**, 105457 (2020).