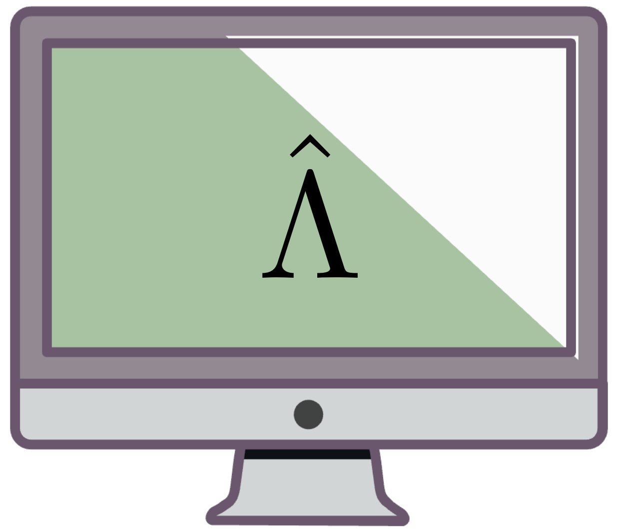
Optimal tomography of quantum channels



$\|\hat{\Lambda} - \Lambda\|_\diamond \leq \varepsilon$

Antonio Anna Mele

PhD student at Freie Universität Berlin,
Google PhD fellow

arXiv > quant-ph > arXiv:2512.10214

Quantum Physics

[Submitted on 11 Dec 2025 (v1), last revised 16 Jan 2026 (this version, v2)]

Optimal learning of quantum channels in diamond distance

Antonio Anna Mele, Lennart Bittel

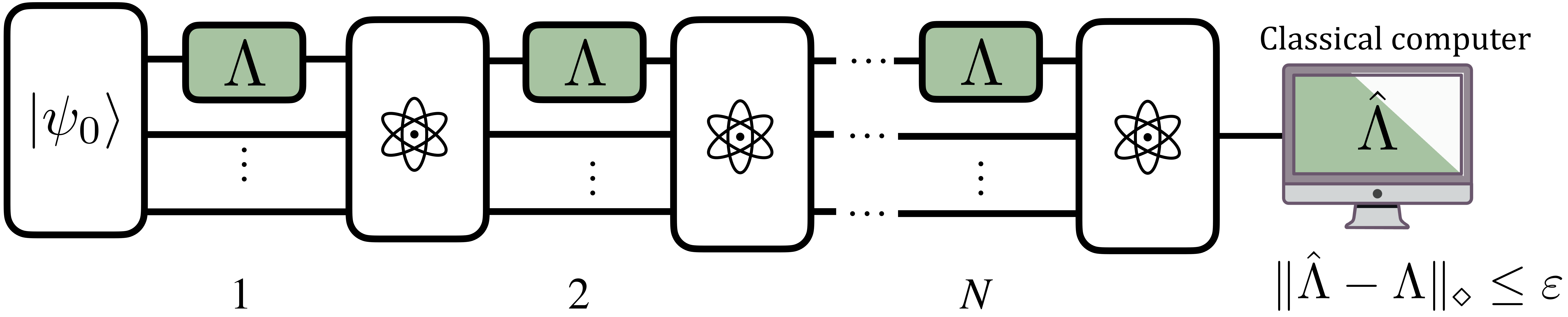
Joint work with

Lennart Bittel

Photo taken during a road trip in South Africa (Nov 2025), where we started thinking about the problem.

Quantum process tomography

Learn/estimate an unknown quantum channel Λ from black-box queries.



The **diamond distance** $\|\hat{\Lambda} - \Lambda\|_\diamond$ is the standard error metric for channels.

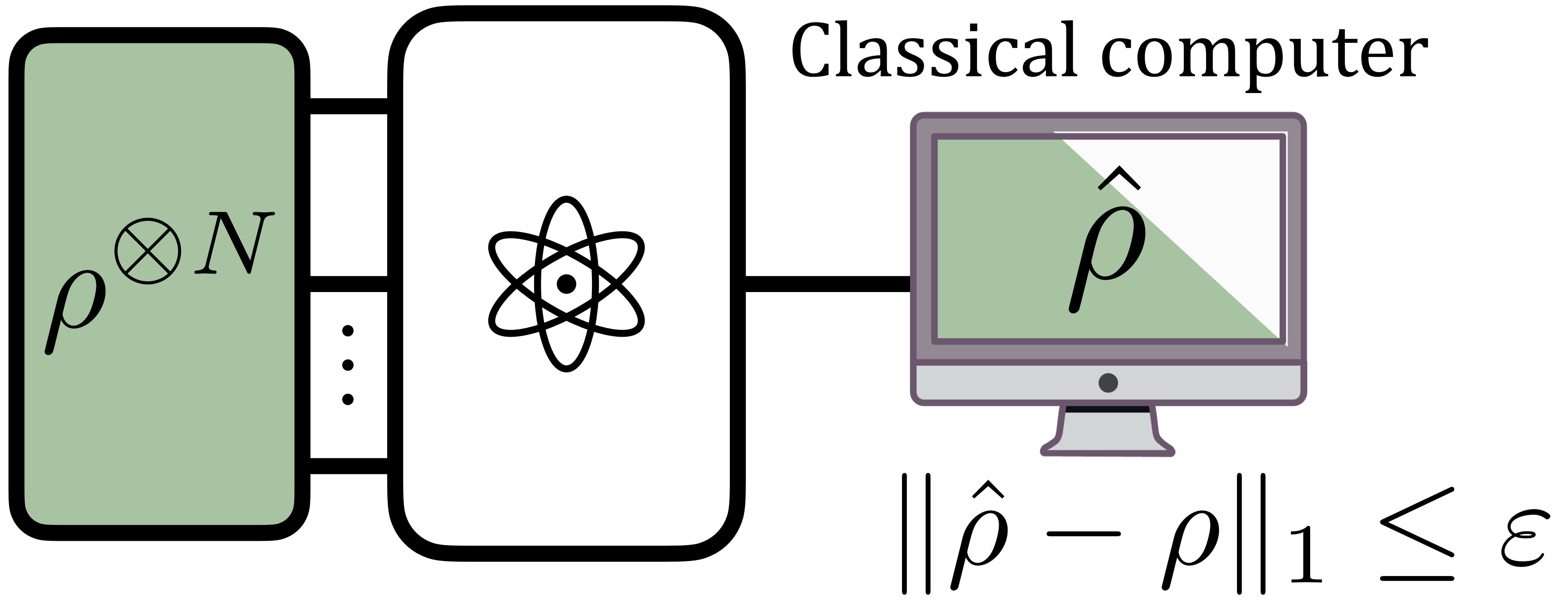
Open problem

Optimal *query complexity* for channel tomography.

How many uses of Λ are sufficient/necessary to estimate it within ε-accuracy in diamond distance?

Optimal number of queries for tomography

- This has been settled only for the special case of **quantum state tomography**.



- $N = \Theta(rd/\varepsilon^2)$ sufficient and necessary to learn an unknown r-rank state ρ.

Pure states $(r = 1)$	**Mixed states** $(r \leq d)$
[Hayashi (1998)] [Guta, Kahn, Kueng, Tropp (2018)]	[O'Donnell, Wright (2015)] [Haah, Harrow, Ji, Wu, Yu (2015)]

- For channels, this was solved only for unitaries (Kraus rank $r = 1$).

Unitaries $(r = 1)$	**General channels** $(r \leq d^2)$
[Haah, Kothari, O'Donnell, Tang (2022)] $N = O(d^2/\varepsilon)$	??? [This work] $N = O(rd^2/\varepsilon^2)$

Quantum channel Λ: CPTP map

$$\Lambda : \mathcal{L}(\mathbb{C}^{d_{\rm in}}) \to \mathcal{L}(\mathbb{C}^{d_{\rm out}})$$

Kraus rank r: minimum number of Kraus operators $\{K_i\}_{i=1}^r \subset \mathbb{C}^{d_{\rm out} \times d_{\rm in}}$.

$$\Lambda(\rho) = \sum_{i=1}^{r} K_i \rho K_i^\dagger, \qquad \sum_{i=1}^{r} K_i^\dagger K_i = I_{\rm d_{in}}. \qquad \lceil d_{\rm in}/d_{\rm out} \rceil \leq r \leq d_{\rm in} d_{\rm out}$$

Diamond distance:

$$\begin{aligned} \|\widehat{\Lambda} - \Lambda\|_\diamond &= \sup_{\rho \in \mathcal{D}(\mathbb{C}^{d_{\rm in}} \otimes \mathbb{C}^{d_{\rm in}})} \left\| \big((\widehat{\Lambda} - \Lambda) \otimes \mathrm{id}\big)(\rho) \right\|_1 \\ &= 4 \cdot \max_{\text{protocol}} \left(\Pr[\text{correctly distinguish } \widehat{\Lambda} \text{ vs. } \Lambda] - \tfrac{1}{2} \right) \end{aligned}$$

- **Previous channel-tomography query-complexity bounds were not tight.**

  (e.g. $d_{\text{in}} = d_{\text{out}} = d$, constant ε):

$$\text{Upper bound:} \quad N = O(d^6),$$

[Surawy-Stepney, Kahn, Kueng, Guta, 2021]
[Oufkir, 2023]

↓ quadratic gap ↓

$$\text{Lower bound:} \quad N = \widetilde{\Omega}(d^4).$$

[Rosenthal, Aaronson, Subramanian, Datta, Gur, 2024]

- A Kraus rank-r channel can be described by $\Theta(r\, d_{\text{in}}\, d_{\text{out}})$ parameters.

- **Natural conjecture** (parameter counting):

$$N \stackrel{?}{=} \Theta(r\, d_{\text{in}}\, d_{\text{out}}) = O(rd^2) = O(d^4)$$

We prove this and so close the gap.

Main Theorem (informal)

Let $\Lambda : \mathcal{L}(\mathbb{C}^{d_{\text{in}}}) \to \mathcal{L}(\mathbb{C}^{d_{\text{out}}})$ be an unknown channel of Kraus rank $\leq r$.

There exists a process-tomography algorithm that uses

$$N = O\left(\frac{d_{\text{in}}\, d_{\text{out}}\, r}{\varepsilon^2}\right)$$

queries to Λ, and outputs $\hat{\Lambda}$ satisfying

$$\|\hat{\Lambda} - \Lambda\|_\diamond \leq \varepsilon$$

- For any fixed ε constant, $N = \Omega(d_{\text{in}}\, d_{\text{out}}\, r)$ queries are necessary.

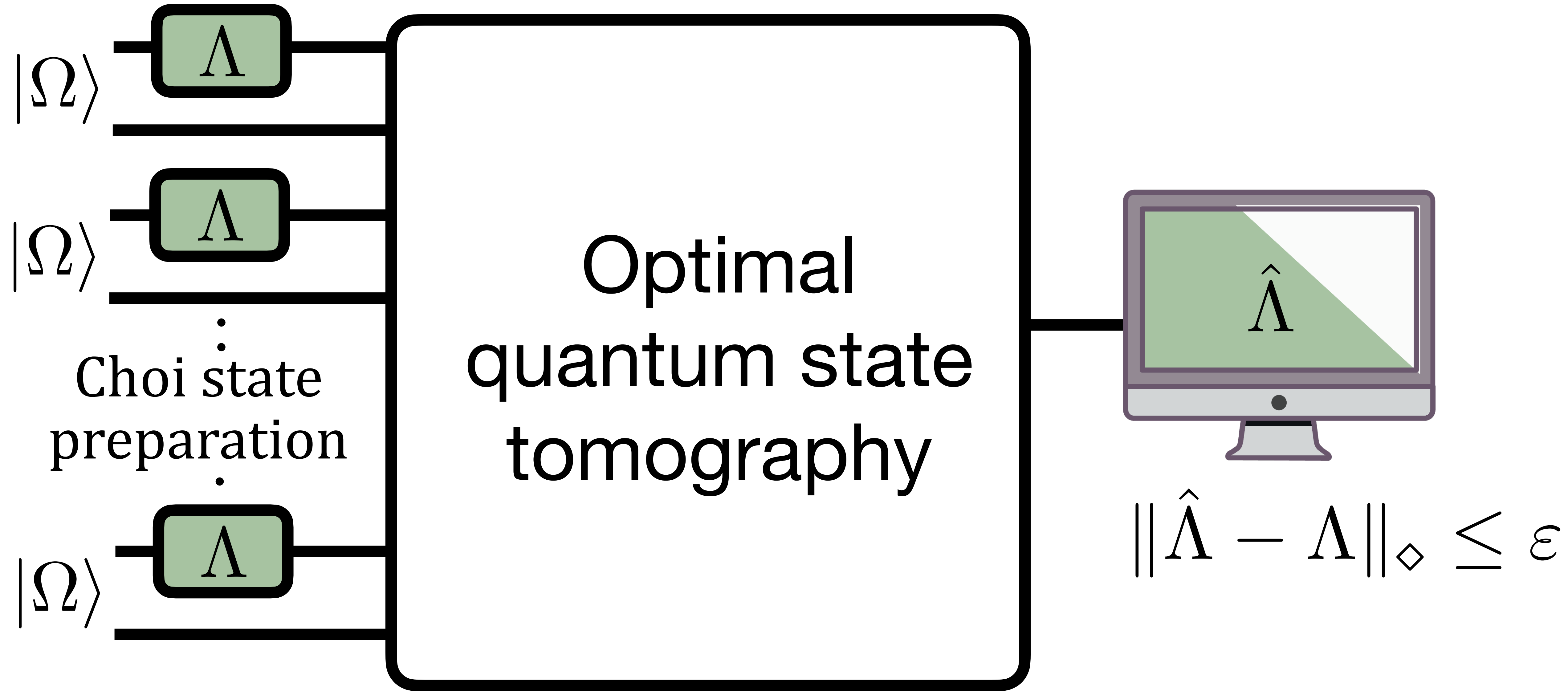
Our idea for the optimal algorithm

What is the first guess you'd make as a student learning basics of quantum channels?

(Assuming the existence of an optimal state tomography algorithm)

(HINT: Choi isomorphism)

Let's run **optimal** state tomography on the Choi state.

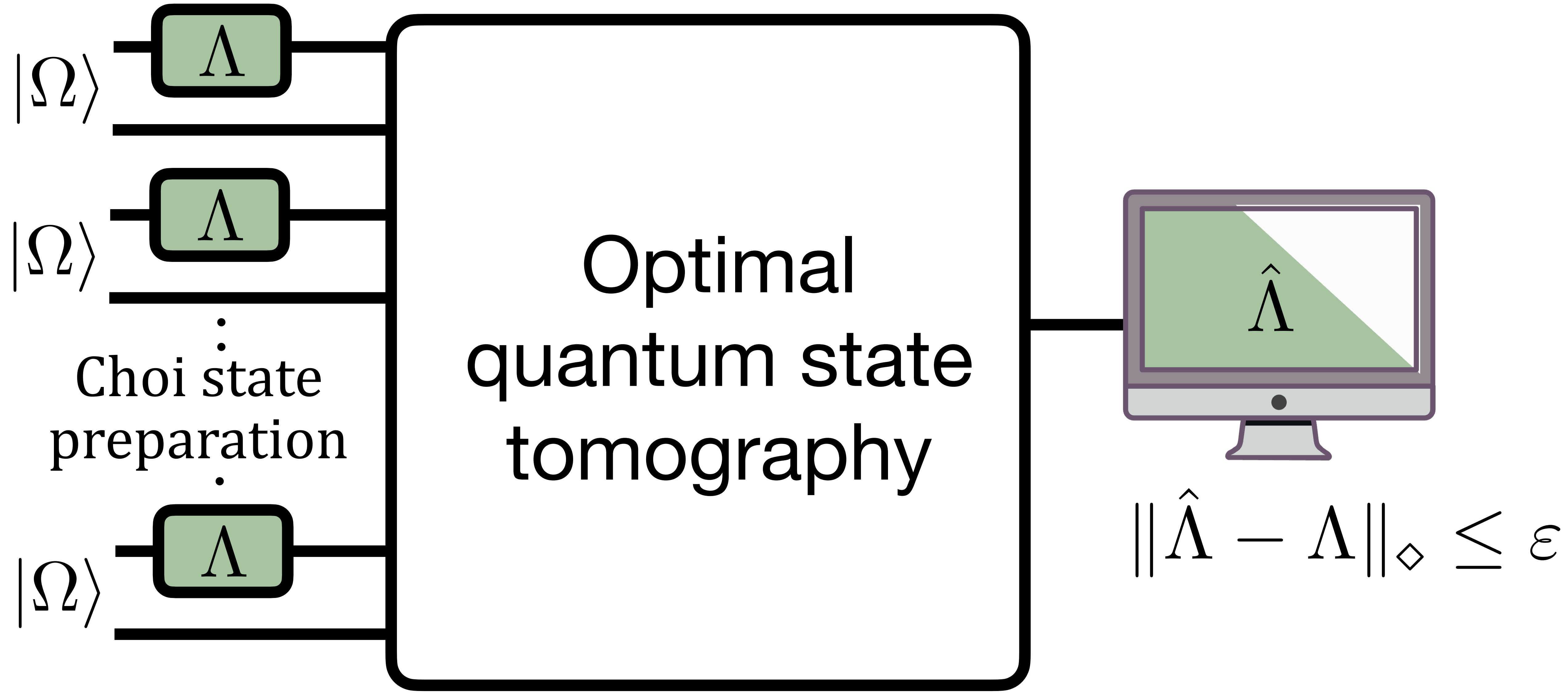


Spoiler, it suffices:

Optimal channel tomography = Optimal state tomography on Choi-state

Take-home message

Choi-state tomography suffices.



Common objection: doesn't this incur a dimensional overhead?

Answer: No (if you analyze the 'error state' directly in diamond norm)

The optimal state-tomography algorithm we use

Mixed state tomography reduces to pure state tomography

Angelos Pelecanos* Jack Spilecki* Ewin Tang* John Wright*

arXiv:2511.15806

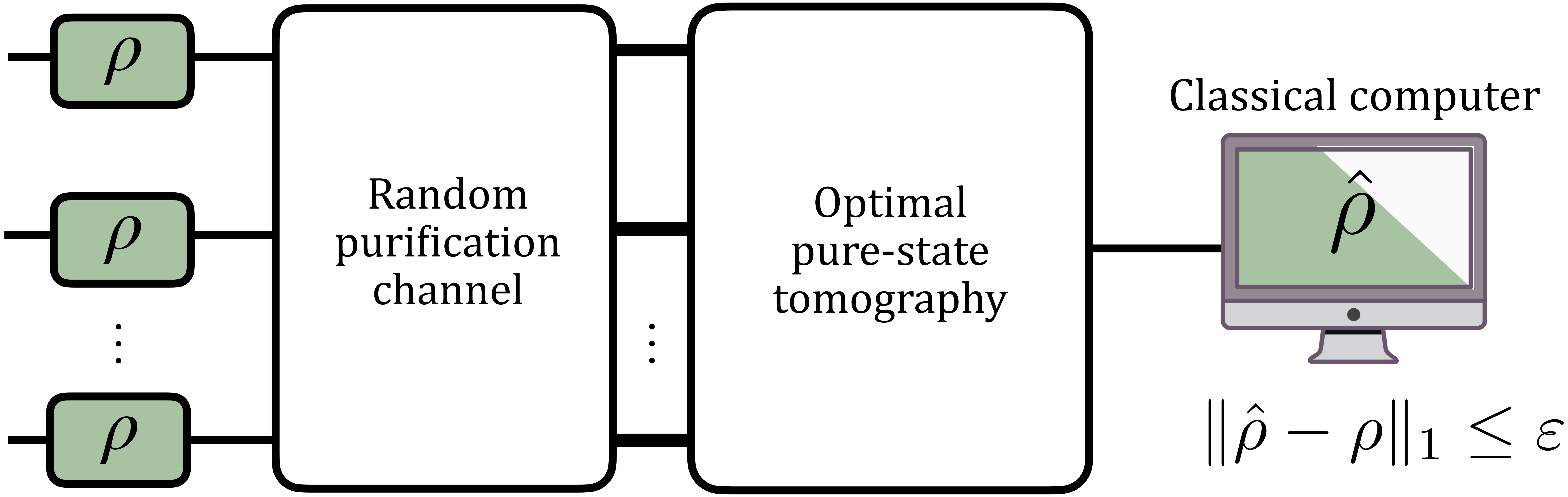
$$\|\hat{\rho} - \rho\|_1 \leq \varepsilon$$

The optimal state-tomography algorithm we use

Mixed state tomography reduces to pure state tomography

Angelos Pelecanos* Jack Spilecki* Ewin Tang* John Wright*

arXiv:2511.15806



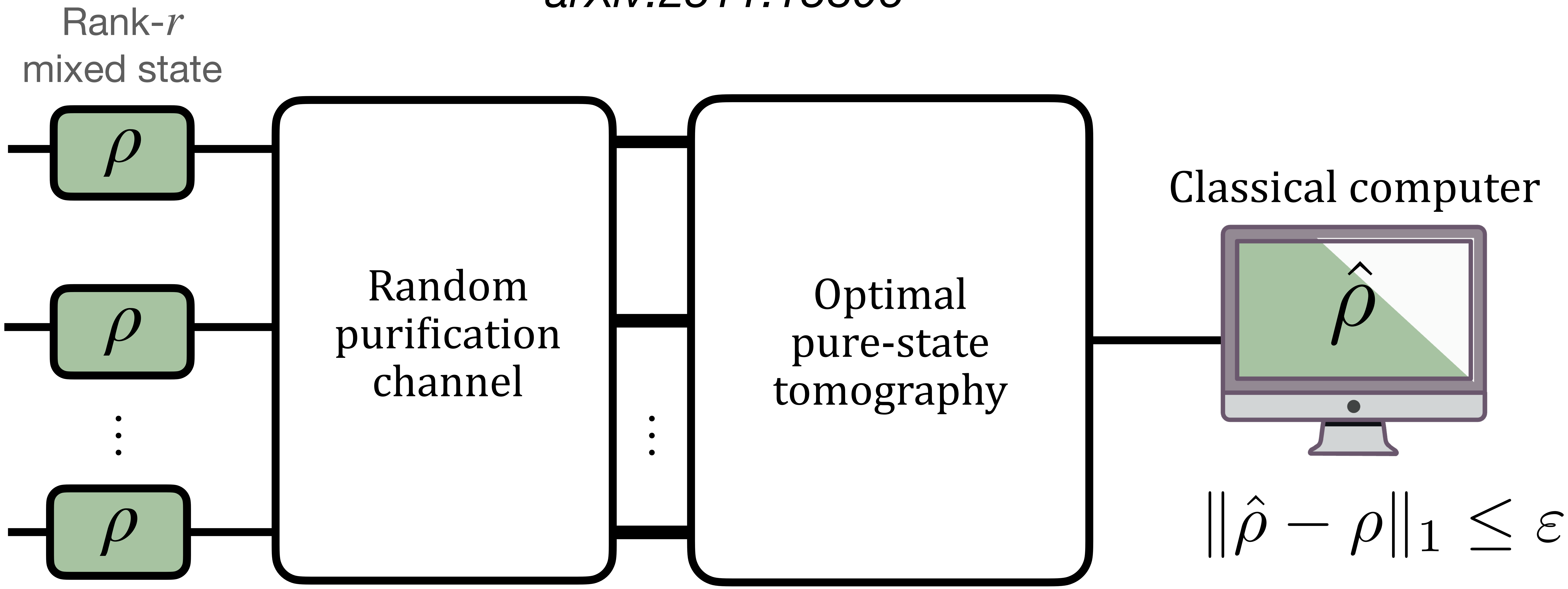
$$\|\hat{\rho} - \rho\|_1 \leq \varepsilon$$

The optimal state-tomography algorithm we use

Mixed state tomography reduces to pure state tomography

Angelos Pelecanos* Jack Spilecki* Ewin Tang* John Wright*

arXiv:2511.15806



$$\|\hat{\rho} - \rho\|_1 \leq \varepsilon$$

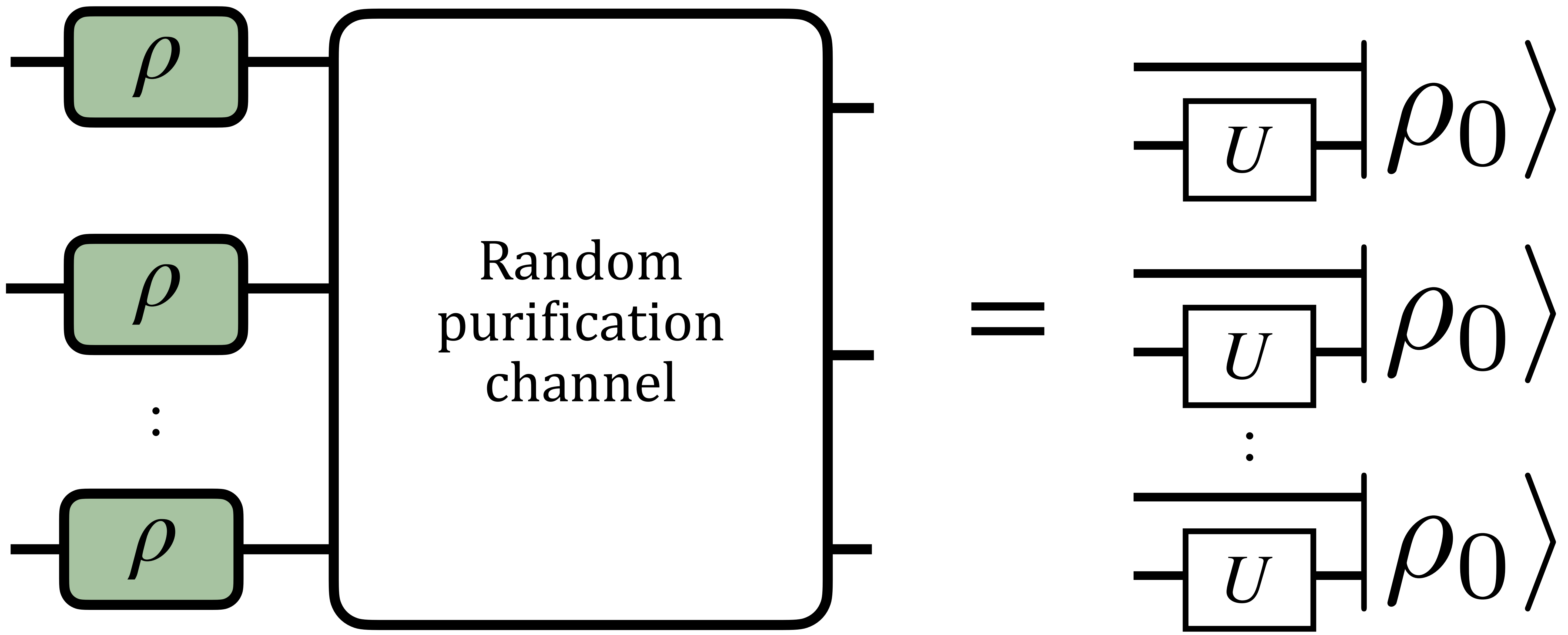
$$\rho^{\otimes N} \longrightarrow |\rho\rangle^{\otimes N} \longrightarrow |\widehat{\rho}\rangle \longrightarrow \widehat{\rho} = \mathrm{Tr}_{\mathbb{C}^r}(|\widehat{\rho}\rangle\langle\widehat{\rho}|)$$

Random purification channel

[Tang, Wright, Zhandry (2025)]

$$\Phi^{d,r}_{\mathrm{Purify}}\left(\rho^{\otimes n}\right) = \mathbb{E}_{|\rho\rangle}\left[\,|\rho\rangle\langle\rho|^{\otimes n}\right]$$

$|\rho\rangle \in \mathbb{C}^d \otimes \mathbb{C}^r$



$, U \sim \mathrm{Haar}(r)$

Proof of correctness

(mixed state tomography)

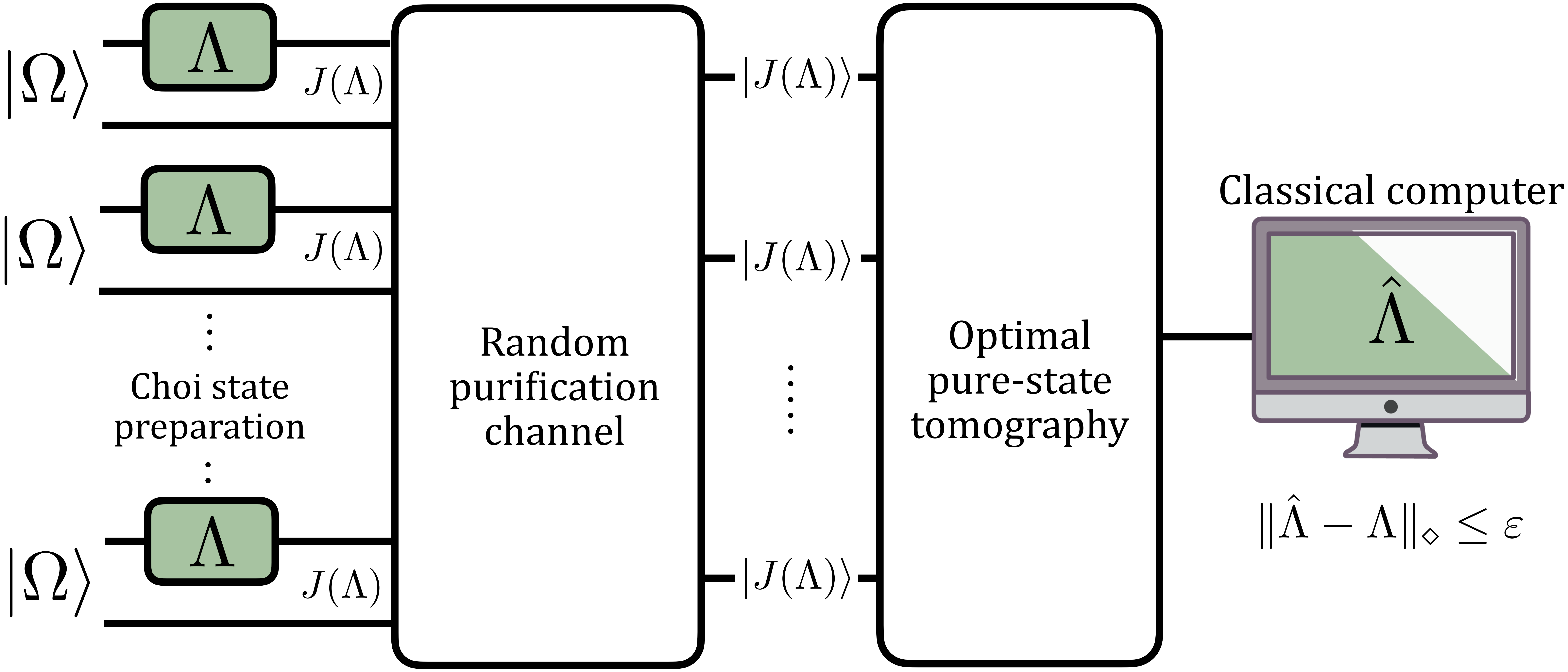
ρ rank-r mixed state $\longrightarrow$ $|\rho\rangle \in \mathbb{C}^d \otimes \mathbb{C}^r$ $\longrightarrow$ $N = O\left(\frac{rd}{\varepsilon^2}\right)$

(Optimal pure-state tomography guarantees) $\downarrow$

$$\|\hat{\rho} - \rho\|_1 = \left\| \mathrm{Tr}_{\mathbb{C}^r}\big(|\hat{\rho}\rangle\langle\hat{\rho}| - |\rho\rangle\langle\rho|\big) \right\|_1 \leq \left\| |\hat{\rho}\rangle\langle\hat{\rho}| - |\rho\rangle\langle\rho| \right\|_1 \leq \varepsilon.$$

$$N = O\left(\frac{rd}{\varepsilon^2}\right) \implies \|\hat{\rho} - \rho\|_1 \leq \varepsilon.$$

Our channel-tomography algorithm



Channel tomography reduces to pure state tomography!

Proof of correctness

(Not as easy as for mixed state tomography)

Λ Kraus rank-r $\longrightarrow$ Choi state $J(\Lambda)$ rank-r $\longrightarrow$ $|J(\Lambda)\rangle \in \mathbb{C}^{d_{\text{in}} d_{\text{out}}} \otimes \mathbb{C}^r$

$\Lambda : \mathcal{L}(\mathbb{C}^{d_{\text{in}}}) \to \mathcal{L}(\mathbb{C}^{d_{\text{out}}})$

$J(\Lambda) \in \mathbb{C}^{d_{\text{in}} d_{\text{out}} \times d_{\text{in}} d_{\text{out}}}$

$$\longrightarrow \; N = O\left(\frac{r\, d_{\text{in}} d_{\text{out}}}{\varepsilon^2}\right).$$

(Optimal pure-state tomography guarantees) $\downarrow$

Naive approach fails for channels:

$$\|\widehat{J} - J(\Lambda)\|_1 = \left\|\mathrm{Tr}_r\left(|\widehat{J}\rangle\langle\widehat{J}| - |J(\Lambda)\rangle\langle J(\Lambda)|\right)\right\|_1 \leq \left\||\widehat{J}\rangle\langle\widehat{J}| - |J(\Lambda)\rangle\langle J(\Lambda)|\right\|_1 \leq \varepsilon$$

$$\|\widehat{\Lambda} - \Lambda\|_\diamond \leq d_{\text{in}} \, \|\widehat{J} - J(\Lambda)\|_1 \leq d_{\text{in}}\varepsilon$$

The naive reduction introduces a d_{in} overhead.

Proof of correctness

(**Done well:** via SDP equivalent definition of the diamond norm)

For Hermitian Choi matrices:

$$\|\Phi\|_\diamond = d_{\text{in}} \max_{\substack{Y=Y^\dagger \\ \sigma\in\mathcal{D}(\mathcal{H}_{\text{in}})}} \operatorname{tr}\big(J(\Phi)\,Y\big)$$

$$\textit{subject to} \quad Y \in \mathcal{L}(\mathcal{H}_{\text{out}} \otimes \mathcal{H}_{\text{in}}),$$

$$-\,\mathbb{I}_{\text{out}} \otimes \sigma \;\le\; Y \;\le\; \mathbb{I}_{\text{out}} \otimes \sigma.$$

For positive Choi matrices:

$$\|\Phi\|_\diamond = d_{\text{in}} \max_{\sigma\in\mathcal{D}(\mathcal{H}_{\text{in}})} \operatorname{tr}\big(J(\Phi)\,(\mathbb{I}_{\text{out}} \otimes \sigma)\big)$$

$$= d_{\text{in}} \left\|\operatorname{tr}_{\text{out}} J(\Phi)\right\|_\infty.$$

Proof of correctness (Why no dimensional overhead)

Key decomposition for pure-state tomography:

[Hayashi (1998)]
[Guta, Kahn, Kueng, Tropp, (2018)]

$$|\hat{J}\rangle = \sqrt{1-\varepsilon^2}\,|J(\Lambda)\rangle + \varepsilon\,|\psi_{\mathrm{err}}\rangle, \qquad |\psi_{\mathrm{err}}\rangle \sim \mathrm{Haar}.$$

Convert to Choi operators:

$$\widehat{J} = \mathrm{Tr}_{\mathbb{C}^r}\left(|\widehat{J}\rangle\langle\widehat{J}|\right), \qquad J(\Lambda) = \mathrm{Tr}_{\mathbb{C}^r}\left(|J(\Lambda)\rangle\langle J(\Lambda)|\right).$$

Choi difference:

$$\widehat{J} - J(\Lambda) = \varepsilon\sqrt{1-\varepsilon^2}(K_{12}+K_{21}) + \varepsilon^2(K_{22}-K_{11}),$$

$$K_{ij} := \mathrm{Tr}_{\mathbb{C}^r}\left(|\Phi_i\rangle\langle\Phi_j|\right), \qquad \Phi_1 := J(\Lambda),\ \Phi_2 := \psi_{\mathrm{err}}.$$

Control the diamond norm of the Choi difference:

$$\|\widehat{\Lambda}-\Lambda\|_\diamond \equiv \|\widehat{J}-J(\Lambda)\|_\diamond \le O(\varepsilon)\Big(\|K_{11}\|_\diamond + \|K_{12}+K_{21}\|_\diamond + \|K_{22}\|_\diamond\Big) \le O(\varepsilon)$$

$$\|K_{11}\|_\diamond = 1, \qquad \|K_{12}+K_{21}\|_\diamond \le 2\sqrt{\|K_{22}\|_\diamond} = O(1), \qquad \|K_{22}\|_\diamond = O(1)$$

Diamond SDP definition + Haar concentration:

$$\left(\|K_{22}\|_\diamond = d_{\mathrm{in}}\,\|\mathrm{Tr}_{\mathrm{out}}(K_{22})\|_\infty = O(1),\right)$$

$$\left\|\mathrm{Tr}_{\mathrm{out},r}\left(|\psi_{\mathrm{err}}\rangle\langle\psi_{\mathrm{err}}|\right)\right\|_\infty = O(1/d_{\mathrm{in}}).$$

$$\boxed{\|\widehat{\Lambda}-\Lambda\|_\diamond \le O(\varepsilon)}$$

Theorem (upper bound)

Let $\Lambda : \mathcal{L}(\mathbb{C}^{d_{\text{in}}}) \to \mathcal{L}(\mathbb{C}^{d_{\text{out}}})$ be an unknown channel of Kraus rank $\leq r$.

There exists a channel-tomography protocol that uses

$$N = O\left(\frac{d_{\text{out}}\, d_{\text{in}}\, r}{\varepsilon^2} + \frac{d_{\text{in}}}{\varepsilon^2} \log\frac{1}{\delta}\right)$$

queries to Λ, and outputs $\hat{\Lambda}$ satisfying $\|\hat{\Lambda} - \Lambda\|_\diamond \leq \varepsilon$ with probability at least $1 - \delta$.

Any channel tomography algorithm must use:

(with Kraus rank r larger than minimal)

$$N = \Omega\left(\frac{d_{\text{out}}\, d_{\text{in}}\, r}{\varepsilon^\beta} + \frac{d_{\text{in}}}{\varepsilon^2} \log\frac{1}{\delta}\right)$$

Reduction to channel identification + packing-net argument

Reduction to learning the bias of many classical coins

The unification of optimal tomography protocols

$$N = O\left(\frac{d_{\mathrm{in}}\, d_{\mathrm{out}}\, r}{\varepsilon^2}\right) \implies \|\hat{\Lambda} - \Lambda\|_\diamond \leq \varepsilon$$

Optimal state tomography $(d_{\mathrm{in}} = 1, \quad \|\cdot\|_\diamond = \|\cdot\|_1)$

Optimal tomography of isometries (previously not known) and unitaries $(r = 1)$

Choi-state already pure (purification step not needed).

Thus, our protocol = single-copy Choi-state tomography. [Guta, Kahn, Kueng, Tropp (2018)]

Heisenberg scaling ε^{-1} achievable for unitaries as well. [Haah, Kothari, O'Donnell, Tang (2022)]

Optimal binary POVMs tomography (previously not known) $(d_{\mathrm{out}} = 2, d_{\mathrm{in}} \equiv d, r = d)$

(i.e., channels diagonal in the computational basis)

Concurrent work:

[Submitted on 15 Dec 2025 (v1), last revised 2 Feb 2026 (this version, v2)]

Quantum channel tomography and estimation by local test

Kean Chen, Nengkun Yu, Zhicheng Zhang

Follow-up:

[Submitted on 15 Jan 2026]

Optimal lower bound for quantum channel tomography in away-from-boundary regime

Kean Chen, Zhicheng Zhang, Nengkun Yu

Minimal Kraus rank: $r_{\min} = \lceil d_{\text{in}}/d_{\text{out}} \rceil$

$$r \geq 2r_{\min} \implies N = \Omega\left(\frac{d_{\text{out}} d_{\text{in}} r}{\varepsilon^2}\right) \overset{\text{(together with our result)}}{\implies} N = \Theta\left(\frac{d_{\text{out}} d_{\text{in}} r}{\varepsilon^2}\right)$$

- In the relevant case $d_{\text{in}} = d_{\text{out}} = d$, this implies:

$r \geq 2:$	$r = 1:$
$N = \Theta\left(\frac{d^2 r}{\varepsilon^2}\right)$	$N = \Theta\left(\frac{d^2}{\varepsilon}\right)$ [Haah, Kothari, O'Donnell, Tang (2022)]
(Classical scaling)	(Heisenberg scaling)

Conclusions

- **Channel tomography** with query-**optimal** dimensional-scaling.
- **Choi-state learning suffices.**
- **Adaptivity does not help** for optimal channel tomography.
  (We use the channel only to prepare Choi state copies)
- **Entangled/Coherent measurements here help** (coherence is needed only to purify)**.**
  (Single-copy non-adaptive protocols are suboptimal [Oufkir, 2023])
  (Entangled measurements provably necessary for non-adaptive protocols)

Open problems

- **Adaptivity helps for single-copy channel tomography?** (For states: [Chen, Huang, Li, Liu, Sellke, 2023])
- **Close the gap in the ε-dependence** left open near the minimal Kraus-regime.

THANK YOU!

$$r_{\min} \leq r \leq 2r_{\min} \quad \Longrightarrow \quad N = \Theta\left(\frac{d_{\text{out}} d_{\text{in}} r}{\varepsilon^{\beta}}\right), \quad \beta = ?$$

- **Is adaptivity necessary for optimal unitary tomography** with Heisenberg scaling ε^{-1}?
- **Limited entangled measurements** trade-off. (For states: [Chen, Li, Liu, 2025])
- Generalise to arbitrary **process tensors** (combs): **"A complete and unified theory of tomography"**.