

# Nearly Time-Optimal Pure State Tomography with Pauli Measurements

Quantum state tomography is the problem of learning an unknown quantum state from measurement outcomes on independent copies. It is a fundamental problem in quantum information and quantum computing, with a history dating back to the 1950s [2]. In this work, we study tomography of *pure*  $n$ -qubit states. The learner receives copies of an unknown state  $|\psi\rangle \in (\mathbb{C}^2)^{\otimes n}$  and must output an estimate  $|\hat{\psi}\rangle$  with high fidelity, e.g.,  $1 - |\langle\hat{\psi}|\psi\rangle|^2 \leq \varepsilon$  with probability at least  $1 - \delta$ .<sup>1</sup>

If the learner can perform arbitrary measurements, the copy complexity of worst-case pure-state tomography is known to be  $\Theta(2^n/\varepsilon)$ . This rate is achieved by several procedures and is information-theoretically optimal; see, e.g., [1, 5–7, 10–13]. Moreover, the measurements do not need to be entangled *across copies* to attain this scaling [6, 8, 14]. The remaining question is whether entanglement is required *within each  $n$ -qubit copy*. Such highly entangled measurements are infeasible in practice, whereas restricting to simpler measurements would make tomography accessible for larger quantum systems. This motivates the central question of our work:

*Can one achieve the optimal  $\Theta(2^n/\varepsilon)$  copy complexity using only nonadaptive single-qubit (product basis) measurements?*

An even stronger goal is to achieve optimal pure-state tomography using only Pauli basis measurements (i.e. measurements diagonalizing operators  $\{X, Y, Z\}^{\otimes n}$ ), a standard and well-studied class of single-qubit measurements. Previously, the strongest guarantee for pure-state tomography using Pauli measurements (and, more generally, any single-qubit measurements) was due to [5], who achieved copy complexity  $\tilde{O}(3^n/\varepsilon)$ .

Their estimator works as follows. First, each copy of the unknown state is measured in a uniformly random Pauli product basis. Each measurement outcome is then converted into a matrix-valued estimate, and these matrices are averaged across samples. By construction, this average equals the true state in expectation. However, the number of samples required for the estimator to concentrate is  $\tilde{O}(3^n/\varepsilon)$ , and their variance bound is tight even for pure product states. Consequently, any approach following this paradigm – namely, forming an average of possibly reweighted matrices derived from Pauli measurement outcomes and proving concentration via the matrix Bernstein inequality – cannot asymptotically improve upon the  $3^n$  dependence. This leaves open whether a fundamentally different use of Pauli measurement data can avoid this barrier.

Our main result shows that the  $3^n$  scaling is not a fundamental limitation of *all* Pauli measurement schemes. We give a different, but still simple, learning algorithm that achieves essentially optimal copy complexity using only Pauli basis measurements.<sup>2</sup>

**Theorem 0.1.** *There exists an algorithm that, given copies of an unknown  $n$ -qubit pure state  $|\psi\rangle$ , samples  $\tilde{O}(2^n \log(1/\delta)/\varepsilon)$  Pauli product bases, measures one copy of  $|\psi\rangle$  in each sampled basis, and outputs an estimate  $|\hat{\psi}\rangle$  satisfying  $|\langle\hat{\psi}|\psi\rangle|^2 \geq 1 - \varepsilon$  with probability at least  $1 - \delta$ . The algorithm runs in time  $\tilde{O}(2^n/\varepsilon)$ .*

Thus, up to polylogarithmic factors, our copy complexity matches the  $\Omega(2^n/\varepsilon)$  lower bound that holds even when the learner is allowed *arbitrary* measurements [13].

Our algorithm also attains essentially optimal running time. **This is the first algorithm for pure-state tomography with near-optimal running time, addressing a question explicitly raised in [4].** Before our work, the fastest algorithms required  $\tilde{O}(4^n/\varepsilon)$  time [1], while, among algorithms restricted to Pauli-basis measurements, the best known running time was  $\tilde{O}(8^n/\varepsilon)$  [5]. Our result therefore shows that simple, nonadaptive single-qubit measurements suffice to achieve both near-optimal statistical efficiency and near-optimal computational efficiency.

<sup>1</sup>For pure states, learning with respect to distances such as trace distance, Frobenius distance, and  $\chi^2$  distance is equivalent, since each is a monotone function of fidelity.

<sup>2</sup>The term *Pauli measurement* can refer to two distinct measurement models. In the first, each qubit is measured independently in one of the  $X$ ,  $Y$ , or  $Z$  bases, yielding an  $n$ -bit outcome; we refer to this as a *Pauli basis measurement*. In the second, the two-outcome projective measurement  $\{(I + P)/2, (I - P)/2\}$  associated with an  $n$ -qubit Pauli operator  $P$  is applied, yielding a single outcome in  $\{\pm 1\}$ ; we refer to this as a *Pauli observable measurement*. Under the latter model, it is known that  $\Omega(d^2/\varepsilon)$  samples are necessary for pure state tomography [3, 9].

## Technical Overview

Our algorithm has two components. First, we show how to perform pure-state tomography assuming access to an estimator that approximates the Frobenius distance between an unknown state and a candidate state  $\sigma$ , using only nonadaptive Pauli measurements that do not depend on  $\sigma$ . The second component constructs such an estimator with the required sample and time complexity.

### Tomography via Frobenius Distance Estimation

Let  $|\psi\rangle$  be a pure  $n$ -qubit state. Using  $\tilde{O}(2^n/\varepsilon)$  copies, we aim to output  $|\hat{\psi}\rangle$  with  $|\langle\psi|\hat{\psi}\rangle|^2 \geq 1 - \varepsilon$ . For any  $x \in \{0, 1\}^\ell$ , let  $p_x$  be the probability of obtaining outcome  $x$  when measuring the first  $\ell$  qubits of  $|\psi\rangle$  in the computational basis, and let  $|\psi_x\rangle$  be the normalized post-measurement state on the remaining  $n - \ell$  qubits conditioned on that outcome.

We reconstruct  $|\psi\rangle$  recursively along the binary tree of prefixes. At depth  $k$ , we maintain estimates  $\{|\hat{\psi}_x\rangle : x \in \{0, 1\}^k\}$ . Given estimates at depth  $k + 1$ , we show how to build estimates at depth  $k$ . The base case is  $k = n - 1$ , where each  $|\psi_x\rangle$  is a 1-qubit state, unique up to global phase. Iterating this process up to  $k = 0$  yields  $|\hat{\psi}_\emptyset\rangle \approx |\psi\rangle$ .

We now describe the recursive estimation procedure. Fix  $x \in \{0, 1\}^k$ . Given  $|\hat{\psi}_{x0}\rangle \approx |\psi_{x0}\rangle$  and  $|\hat{\psi}_{x1}\rangle \approx |\psi_{x1}\rangle$ , we seek coefficients  $\hat{\alpha}_{x0}, \hat{\alpha}_{x1}$  such that

$$|\hat{\psi}_x\rangle := \hat{\alpha}_{x0} |0\rangle |\hat{\psi}_{x0}\rangle + \hat{\alpha}_{x1} |1\rangle |\hat{\psi}_{x1}\rangle \approx |\psi_x\rangle.$$

A stability property of this merge step guarantees the existence of coefficients whose Frobenius error is comparable to the weighted errors of the two children. It therefore suffices to approximately solve this two-parameter optimization problem. Its objective has an explicit closed-form expression as a two-variable trigonometric function, so the main task is to implement an oracle for evaluating it. We do so using samples from  $|\psi_x\rangle$ , obtained by measuring the first  $k$  qubits of  $|\psi\rangle$  in the computational basis.

Implementing this low-dimensional optimization reduces to estimating the distance between the unknown state  $|\psi_x\rangle \langle\psi_x|$  and a known candidate  $\sigma$ . The Frobenius-distance estimator described in the next subsection provides, for any fixed  $\sigma$ , an estimate of  $\| |\psi_x\rangle \langle\psi_x| - \sigma \|_F$  from appropriate measurements on copies of  $|\psi_x\rangle$ .

The main subtlety is that the measurements on the last  $n - k$  qubits must be fixed *before* we learn  $x$ : we cannot choose the measurement basis adaptively as a function of the observed prefix. To handle this, fix a probability scale  $p$  and consider all prefixes with  $p_x \approx p$ . It suffices to choose a *global* list of  $m = \tilde{O}(2^n p/\varepsilon)$  measurement settings, independent of  $x$ , that lets the Frobenius estimator compare  $|\psi_x\rangle$  to any fixed  $\sigma$ . We repeat each setting  $\Theta((1/p) \text{poly}(n))$  times, always measuring the first  $k$  qubits in the computational basis and binning outcomes by the observed  $x$ . For any  $x$  with  $p_x \approx p$ , with high probability each setting in the global list is applied at least once within the bin for  $x$ , providing exactly the data required by the distance estimator. Finally, since the probabilities  $p_x$  vary with  $x$ , we run the procedure for  $p \in \{2^{-n}, 2^{-(n-1)}, \dots, 1\}$ . Every  $p_x$  lies within a constant factor of some dyadic  $p$ , so one of these runs handles it.

### Frobenius Distance Estimation

Suppose that  $\rho$  and  $\sigma$  are potentially mixed  $n$ -qubit states. We are given access to copies of  $\rho$  and to a full classical description of  $\sigma$ . Our goal is to estimate  $\|\rho - \sigma\|_F$  up to additive error  $\gamma$ , using a fixed, nonadaptive list of measurements that depends on neither  $\rho$  nor  $\sigma$ , and using  $O(d/\gamma^2)$  samples.

In describing the algorithm, we assume access to samples from both  $\rho$  and  $\sigma$ , rather than a full description of  $\sigma$ . This is more general, since a classical description of  $\sigma$  allows us to simulate measurements on it. The first step is to note that  $\|\rho - \sigma\|_F = 2\sqrt{d} \sqrt{\mathbf{E}_P[v_P^2]}$ , where  $v$  is a real vector of length  $4^n$ , indexed by  $n$ -qubit Pauli matrices, defined by  $v_P = \frac{1}{2} \text{Tr}(P(\rho - \sigma))$ ; in particular,  $|v_P| \leq 1$ .

By measuring  $\rho$  and  $\sigma$  with respect to the observable  $P$ , for any fixed  $P$  we can draw samples from a Rademacher random variable with mean  $v_P$  using just one copy each of  $\rho$  and  $\sigma$ . Therefore, estimating the Frobenius distance using nonadaptive Pauli measurements reduces to the following problem: estimate

$\sqrt{r} := \sqrt{\mathbf{E}_k[v_k^2]}$  to additive error  $\alpha$  for vectors  $v \in [-1, 1]^N$ , given the ability to specify a multiset  $\{k_1, \dots, k_T\}$  and receive Rademacher samples with means  $v_{k_1}, \dots, v_{k_T}$ . We require a total of  $T = \tilde{O}(1/\alpha^2)$  samples.

To motivate the algorithm, imagine first that all entries of  $v$  are either *small*, meaning close to 0, or *big*, meaning that their absolute values are close to 1. Write  $r = r_{\text{small}} + r_{\text{big}}$ , where  $r_{\text{small}}$  and  $r_{\text{big}}$  are the respective contributions to  $r$ . Our goal is to produce an estimator  $\hat{r}$  satisfying  $|\hat{r} - r| \lesssim \alpha\sqrt{r} + \alpha^2$ , since such an estimate can be converted into an additive- $\alpha$  estimate of  $\sqrt{r}$ . At a constant-factor loss, it suffices to estimate the two contributions separately with guarantees

$$|\hat{r}_{\text{small}} - r_{\text{small}}| \lesssim \alpha\sqrt{r_{\text{small}}} + \alpha^2, \quad |\hat{r}_{\text{big}} - r_{\text{big}}| \lesssim \alpha\sqrt{r_{\text{big}}} + \alpha^2. \quad (1)$$

1. To estimate  $r_{\text{small}}$ , we sample a small number of indices  $k_1, \dots, k_{T_{\text{small}}}$  and take many samples of each  $v_{k_t}$  to estimate its value accurately, discarding values that are too large. We then output the average of the remaining values  $v_{k_t}^2$ . If  $r_{\text{small}}$  is nonnegligible, there must be many small indices, so sampling only  $T_{\text{small}}$  indices suffices to encounter enough of them. On the other hand, many samples per index are needed to estimate their values accurately enough to satisfy (1).
2. To estimate  $r_{\text{big}}$ , we instead sample a large number  $T_{\text{big}}$  of indices  $k_1, \dots, k_{T_{\text{big}}}$ , take a small number of samples of each  $v_{k_t}$ , discard indices whose empirical averages are too small, and output the average of the corresponding values  $v_{k_t}^2$ . Here, many indices are required to ensure that large entries are encountered whenever they contribute significantly to  $r$ , but only a few samples per index are needed, since a coarse estimate of  $v_{k_t}$  suffices when  $v_{k_t}^2$  is large.

Thresholding based on the empirical values of  $v_{k_t}$  introduces two issues. First, the resulting estimate of  $v_k$  is no longer unbiased when conditioned on an index being classified as small or big. Second, for indices whose true values lie near the threshold, the thresholding rule may assign the index to neither or both categories with nonzero probability. Consequently, the probabilities with which an index contributes to the small and big parts may fail to sum to one, leading to under- or over-counting in expectation. To resolve both issues, before estimating  $v_{k_t}$  we run a fixed process, using independent Rademacher samples, that assigns the index to exactly one of the two categories. We then independently compute the estimate of  $v_{k_t}$ .

For a general vector, rather than splitting the indices into only two categories, we partition them into  $\log(1/\alpha)$  level sets, where the  $j$ th level consists roughly of indices  $k$  with  $|v_k| \in [2^{-j}, 2^{-j+1}]$ . It turns out that the desired approximation guarantee can be obtained by taking  $T_j \approx \alpha^{-2}/4^j$  indices at level  $j$  and  $m_j \approx 4^j$  samples of each index. Thus, as the entries become smaller, the algorithm examines fewer coordinates but estimates each one more accurately. Summing over all levels yields total sample complexity  $\tilde{O}(1/\alpha^2)$ , as desired.

**Time Efficiency.** Implementing the above algorithm requires producing Rademacher random variables with expectation  $\text{Tr}(\sigma P)$  given descriptions of  $\sigma$  and  $P$ . Naively, when  $\sigma = |\psi\rangle\langle\psi|$  and  $|\psi\rangle$  is given as a list of amplitudes, this requires time  $2^n$  for each  $P$ . Repeating the computation for roughly  $2^n$  Pauli matrices gives a runtime of  $4^n$ , whereas we require runtime  $\tilde{O}(2^n)$ , ignoring  $\varepsilon$  and  $\delta$ .

To perform this sampling in near-linear time, we use the fact that Pauli matrices are phased permutation matrices in the computational basis. If  $|\psi\rangle$  has real computational-basis amplitudes, then  $\langle\psi|P|\psi\rangle$  is a weighted sum of values in  $\{\pm 1, \pm i\}$ . This gives the following procedure: draw  $x \in \{0, 1\}^n$  with probability  $|\langle x|\psi\rangle|^2$  and compute the entry of  $P$  that is nonzero in the  $x$ th column, which takes time  $O(n)$  when  $P$  is given as a Pauli string. Straightforward reweighting and signing produces a value  $v$  such that, conditioned on  $x$ , one can output a Rademacher random variable with expectation  $v$ . The case where  $|\psi\rangle$  has phases is analogous.

To draw  $x$  with probability  $|\langle x|\psi\rangle|^2$ , we use the alias method [15]. At the beginning of the distance-estimation procedure, we construct a data structure in time  $\tilde{O}(2^n)$  that subsequently produces each such sample in time  $O(1)$ .

## References

- [1] Joran van Apeldoorn, Arjan Cornelissen, András Gilyén, and Giacomo Nannicini. “Quantum tomography using state-preparation unitaries”. In: *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2023. DOI: [10.1137/1.9781611977554.ch47](https://doi.org/10.1137/1.9781611977554.ch47) (page 1).
- [2] Ugo Fano. “Description of States in Quantum Mechanics by Density Matrix and Operator Techniques”. In: *Reviews of Modern Physics* (1957). DOI: [10.1103/RevModPhys.29.74](https://doi.org/10.1103/RevModPhys.29.74). URL: <https://doi.org/10.1103/RevModPhys.29.74> (page 1).
- [3] Steven T. Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. “Quantum Tomography via Compressed Sensing: Error Bounds, Sample Complexity, and Efficient Estimators”. In: *New J. Phys.* (2012). DOI: [10.1088/1367-2630/14/9/095022](https://doi.org/10.1088/1367-2630/14/9/095022). arXiv: [1205.2300](https://arxiv.org/abs/1205.2300) [quant-ph] (page 1).
- [4] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. “Efficient Learning of Quantum States Prepared With Few Non-Clifford Gates”. In: *Quantum* (2025). DOI: [10.22331/q-2025-11-06-1907](https://doi.org/10.22331/q-2025-11-06-1907) (page 1).
- [5] Madalin Guță, Jonas Kahn, Richard Kueng, and Joel A Tropp. “Fast state tomography with optimal error bounds”. In: *Journal of Physics A: Mathematical and Theoretical* (2020) (page 1).
- [6] Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. “Sample-optimal tomography of quantum states”. In: *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. 2016 (page 1).
- [7] Masahito Hayashi. “Asymptotic estimation theory for a finite-dimensional pure state model”. In: *Journal of Physics A: Mathematical and General* (1998) (page 1).
- [8] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. “Low rank matrix recovery from rank one measurements”. In: *Applied and Computational Harmonic Analysis* (2017). DOI: <https://doi.org/10.1016/j.acha.2015.07.007> (page 1).
- [9] Angus Lowe and Ashwin Nayak. “Lower bounds for learning quantum states with single-copy measurements”. In: *arXiv preprint arXiv:2207.14438* (2022) (page 1).
- [10] Ryan O’Donnell and John Wright. “Efficient quantum tomography”. In: *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. 2016 (page 1).
- [11] Angelos Pelecanos, Jack Spilecki, Ewin Tang, and John Wright. “Mixed state tomography reduces to pure state tomography”. In: *arXiv preprint arXiv:2511.15806* (2025) (page 1).
- [12] Angelos Pelecanos, Jack Spilecki, and John Wright. “The debiased Keyl’s algorithm: a new unbiased estimator for full state tomography”. In: *arXiv preprint arXiv:2510.07788* (2025) (page 1).
- [13] Thilo Scharnhorst, Jack Spilecki, and John Wright. “Optimal lower bounds for quantum state tomography”. In: *arXiv preprint arXiv:2510.07699* (2025) (page 1).
- [14] Vladislav Voroninski. *Quantum Tomography From Few Full-Rank Observables*. 2013. arXiv: [1309.7669](https://arxiv.org/abs/1309.7669) [math-ph]. URL: <https://arxiv.org/abs/1309.7669> (page 1).
- [15] Alastair J Walker. “An efficient method for generating discrete random variables with general distributions”. In: *ACM Transactions on Mathematical Software (TOMS)* (1977) (page 3).