The Dissertation Committee for Sabee Grewal
certifies that this is the approved version of the following dissertation:

Efficient Learning of Quantum States and Circuits

Committee:

Scott Aaronson, Supervisor

Nick Hunter-Jones

William Kretschmer

John Wright

Efficient Learning of Quantum States and Circuits

by

Sabee Grewal

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2026

To my mom and dad

Acknowledgments

I've had tremendous good fortune during my PhD, for which I feel a great deal of gratitude: generous funding, travelling around the world, working on fun problems, and finding myself in a warm community at the intersection of theoretical computer science and quantum computing. Below I thank a few people who have made a particularly strong impact on my time in graduate school, though this list is by no means exhaustive.

I'm grateful to my advisor, Scott Aaronson. He is contagiously enthusiastic, hyper-curious, well-read, and interacts with his group and the broader community in such an open and genuine manner (if you don't believe me, check out his blog!). Just being around Scott made me a better researcher: my instincts and taste were shaped by seeing which open problems caused him to lift out of his chair (and which ones caused him to retreat to his phone), and by hearing his takes on the latest SciRate dump and the recent goings-on in quantum and AI. Despite being insanely busy, Scott was always available whenever I needed him: I received nearly instant replies to most emails, and he always made time to meet when I asked.

Let me also acknowledge the rest of my committee: Nick Hunter-Jones, William Kretschmer, and John Wright. Thank you for making graduating a stress-free process.

A brief acknowledgement is also due to Codex, which I used in preparing this thesis, for tasks such as making notation consistent, finding typos, and dealing with the bibliography. I'm happy that my dissertation is being completed just barely before AI can solve major open problems in mathematics. The work presented in this thesis is vintage human-generated research.

There are a few people who I think were crucial in helping me get started. Peter Johnson, Dax Enshan Koh, and Patrick Rall were especially generous with their time when I first started out. Shortly after that, William Kretschmer and Daniel Liang—at the time, two slightly more senior students in Scott's group—agreed to work with Vishnu Iyer and me on some problems related to stabilizer states. Getting to work "up close" with William and Daniel early in my PhD was a formative experience; I got to see how they attacked

problems thoughtfully and with discipline. And even before all of this, Scott's writing and Ryan O'Donnell's YouTubing helped spark my early interest in theoretical computer science.

Thank you to my co-authors (including those whose work doesn't appear in this thesis): Scott Aaronson, Srinivasan Arunachalam, Mohammad Bavarian, Toby Cubitt, Matthew DeCross, Justin A. Gerber, Kevin Gilmore, Dan Gresh, Giulio Gueltrini, Meghal Gupta, David Hayes, William He, Nicholas Hunter-Jones, Vishnu Iyer, Dax Enshan Koh, William Kretschmer, Vinayak Kumar, Daniel Liang, Simon Marshall, Karl Mayer, Brian Neyenhuis, Ryan O'Donnell, Marien Raat, Ronak Ramachandran, Dorian Rudolph, Aniruddha Sen, Mihir Singhal, John Wright, and Justin Yirka. I feel very lucky to have worked with you all.

Thank you to everyone at the Quantum Information Center I interacted with over the years: Ryan Anselm, Michelle Ding, Bryce Fuller, Minki Hhan, Shih-Han Hung, Nick Hunter-Jones, Matteo Ippoliti, Vishnu Iyer, Siddhartha Jain, Andreas Karch, Niels Kornerup, William Kretschmer, Brian La Cour, Jiawei Li, Daniel Liang, Jiahui Liu, Shivan Mittal, Sebastian Oberhoff, Corey Ostrove, Jason Pollack, Patrick Rall, Ronak Ramachandran, Andrea Rocchetto, Aniruddha Sen, Shyam Shankar, Justin Yirka, and Yuxuan Zhang.

I'm grateful to Srinivasan Arunachalam, Dar Gilboa, and Peter Johnson for being excellent intern hosts. Thank you to John Wright for hosting me as a graduate student visitor at Simons for a summer, and to Umesh Vazirani (and the CIQC) for supporting my participation in a few programs there.

Enormous thanks to Ryan Sweke for inviting me to give a tutorial at the African Institute for Mathematical Sciences in Cape Town, South Africa. I cherish my interactions with the students there. Thank you also to Joe Fitzsimons at Horizon Quantum Computing for inviting me to their hackathon (shoutout Dissipation Nation, and Matthew Hagan for saving my life).

One hears about how stressful, competitive, or isolating academic life can be; I'm happy to say that my experience in our community has been quite the opposite. I'm grateful for all the friends I get to see every so often at QIP, TQC, or some workshop or another. I

won't try to list everyone here, but if you're reading this, there's a good chance I'm talking about you!

I'd also like to express gratitude for the UT CS Theory Lab, in particular the cohort of students and postdocs around my year with whom I overlapped the most: Jeff Champion, Gautam Chandrasekharan, Joshua Cook, Jesse Goodman, Shivam Gupta, Minki Hhan, Vishnu Iyer, Michael Jaber, Sid Jain, Vinayak Kumar, Lin Lin Lee, Geoffrey Mon, Aditya Parulekar, Rojin Rezvan, Kristin Sheridan, Konstantinos Stavropoulos, and Justin Yirka. It has been an honor going through graduate school with you all. There's plenty I'll miss, but it'll mostly just be the simple stuff: coffee runs down Waller Creek, popping by people's desks to chat, and continuously being in awe of how incredibly loud Vinny can be. I'm glad to have lived through the Golden Age for as long as I did. Thank you also to my labmates down the hall in WNCG: Somi Agarwal, Matt Jordan, Khang Le, Josh Ong, Advait Parulekar, and Harshit Sikchi.

Thank you, Dr. Patt, for supporting me academically for over a decade. I always look forward to our semi-regular trips to Jason's Deli.

To my friends and family outside academia, thank you for reminding me that there's much more to life than research.

I leave my last words for my family. Thank you to Banjo and Satchmo for filling our home with happiness and love. And to Kelsey, for her unlimited patience and for supporting me every step of the way.

Abstract

Efficient Learning of Quantum States and Circuits

Sabee Grewal, PhD
The University of Texas at Austin, 2026

SUPERVISOR: Scott Aaronson

Preparing complex quantum states is of limited value without efficient methods for extracting useful classical information from them. This thesis studies when such information can be extracted efficiently, given copies of an unknown quantum state or query access to an unknown unitary transformation.

We begin with pure-state tomography. We present an algorithm that uses only nonadaptive single-qubit Pauli measurements yet achieves optimal copy and time complexities up to logarithmic factors. This is the first full state tomography algorithm with near-optimal running time. We then show how additional structure can make tomography time-efficient, giving algorithms for two structured classes of quantum states: non-interacting fermions and near-stabilizer states. We also introduce a new learning task, agnostic tomography, where the goal is to find the best approximation to an arbitrary quantum state from a structured class of states, and give efficient algorithms in this model.

Turning to property testing, we give an efficient tester for stabilizer dimension, which implies that $\Omega(n)$ T-gates are necessary for any Clifford+T circuit to prepare n-qubit computationally pseudorandom states. We also prove copy-complexity lower bounds for testing low-degree phase states, showing that testing is nearly as hard as learning these states. More strongly, for every fixed degree d, ensembles of n-qubit degree-d phase states form exponentially accurate approximate state t-designs for $t = \Omega(n^{\lfloor (d-1)/2 \rfloor})$.

Finally, we extend these questions from states to unitary transformations. We develop a general framework for learning unitary transformations by exploiting low-complexity Pauli spectra. This framework unifies and extends prior learning algorithms for Clifford and near-Clifford circuits, shallow circuits, quantum juntas, and fermionic matchgate circuits.

Table of Contents

Chapter 1: Introduction

This thesis develops new algorithms and lower bounds for learning and testing quantum states and circuits. In the basic setting, we are given access to copies of an unknown quantum state and seek to extract useful information about it. We model this access by a black box that prepares a fresh, identical copy of the state on demand, abstracting away the physical mechanism by which the state is produced. The black box might, for example, represent an experimental apparatus that repeatedly prepares a physical system, or a quantum computer executing a state-preparation circuit.

The only way to extract classical information from a quantum state is by measuring it. A quantum measurement produces a classical outcome, but with two important caveats. First, measurements generally disturb the state: after measuring a copy, we cannot simply reuse it as a fresh copy of the original state. Second, measurement outcomes are probabilistic, with their probabilities determined by the Born rule. Thus, a single copy typically reveals only limited information about an unknown state. This is why our access model provides many fresh copies. The algorithm may perform different measurements on different copies—or even a joint measurement on several copies at once—and must somehow piece together the resulting classical outcomes into useful information about the underlying state.

What constitutes "useful information" depends on the learning task. At one extreme, *quantum state tomography* asks us to reconstruct a classical description of the entire state. Other tasks may ask for only partial information about the state, such as a succinct approximation to it or whether the state has a particular property.

Two basic resources will be of particular interest: the number of copies used by the algorithm, which we call its *copy complexity*, and its *time complexity*. The latter accounts for both the quantum computation required to perform the measurements and the classical computation required to process their outcomes. Thus, the central questions are: how many copies are needed, what measurements suffice, and how much computation is required? Quantum states will be our basic setting, though later we will consider analogous questions for other quantum objects, namely unknown quantum circuits.

1.1 Motivations for Quantum Learning

Quantum state learning is already an important tool in experimental physics: it allows phenomena that cannot be observed in a single experimental run to be reconstructed and studied from many repeated trials. As one example, Ma et al. [Ma+12] demonstrated quantum teleportation over a distance of 143 km. To verify that teleportation had succeeded, they performed state tomography on the photon at the receiving end and compared the reconstructed state with the state that had been sent. To collect the data for this reconstruction, the experimenters repeatedly carried out the teleportation procedure, with each successful run providing a fresh teleported photon that could be measured once. The reported reconstruction used 605 successful detection events collected over roughly 6.5 hours.

In a rather different experiment, Deléglise et al. [Del+08] prepared a microwave field in a Schrödinger-cat state—a superposition of two macroscopically distinguishable states of the field—and reconstructed the state at successive points in time. These reconstructions allowed them to observe the gradual decoherence of the superposition, directly probing the transition from quantum to classical behavior. In both examples, learning the quantum state turned many individual measurement outcomes into a classical picture of the underlying quantum behavior.

Quantum state tomography and related certification routines are also important tools in building fault-tolerant quantum computers. They allow encoded states to be characterized and compared with the underlying physical states to determine whether error correction actually helped [Gup+24; DBBHK25].

These examples use quantum learning to *characterize a system relative to some prior expectation*. But quantum learning has a broader role to play in quantum science and computation. As quantum devices improve, we will be able to prepare and manipulate increasingly complicated quantum systems whose behavior is difficult to predict or describe classically. A central goal of quantum learning is to develop the interface that turns these systems into a classical picture that we can understand and use.

To understand the difficulty of this goal, we must confront one of the great boons—and curses—of quantum mechanics: its *exponentiality*. Even in the simplest setting of n qubits—n two-level quantum systems—a general quantum state requires exponentially many parameters to describe. Quantum theory gives us precise and well-understood laws governing how such systems evolve, but these laws do not by themselves give us a useful understanding of a large interacting system. Extracting the consequences we actually care about—such as its collective behavior or the correlations it exhibits—can be extraordinarily difficult.

Exponentiality puts *efficiency* at the heart of quantum learning. Quantum mechanics tells us, in principle, what experiments can be performed on a system and what information their outcomes contain. Quantum learning lets us ask a more computational question: which experiments on a quantum system admit efficient implementations? For a large quantum system, even *writing down* a complete classical description may require exponential time, so full quantum state tomography quickly becomes impractical.

Fortunately, a complete description is often more than we need. We may instead want to predict the outcomes of particular measurements, understand correlations or entanglement between different parts of the system, or find a simpler model that explains its behavior. The goal is therefore to understand which useful information can be extracted experimentally with a feasible number of copies and a feasible amount of computation.

At a mathematical level, quantum learning is a natural extension of classical learning theory and statistics to a noncommutative setting. Classical probability distributions embed naturally as quantum states, and many familiar problems from the classical literature have quantum analogues. For example, tomography generalizes distribution learning, quantum property testing generalizes classical testing, and agnostic tomography mirrors agnostic learning.

Efficient learning algorithms can also provide a useful lens on the structure of quantum systems. Designing such an algorithm requires identifying features of the unknown system that can be exploited computationally. Sometimes this means turning known physical

or algebraic structure into an algorithm; other times, the search for an algorithm reveals structure that was not previously apparent.

1.2 A Quantum Primer

Before stating our results, we give a brief and informal review of the quantum formalism needed to read them.[1] The conventions and notation used throughout the thesis are collected in Chapter 2.

1.2.1 What is a qubit?

The notion of a *state* appears throughout mathematics, science, and engineering. Informally, the state of a system is a snapshot containing all the information needed to describe its current condition. For example, the state of a computer might include the contents of its memory and registers, while the state of a classical mechanical system specifies the positions and velocities of its particles. The state of a quantum system serves the same purpose: it specifies all the information needed to determine how the system evolves under quantum operations and the probability distribution produced by any measurement.

A *qubit* is the simplest quantum system. A pure state of a qubit is a unit vector in the two-dimensional complex Hilbert space $\mathbb{C}^2$. The standard basis of $\mathbb{C}^2$, conventionally called the *computational basis*, consists of

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Thus, any pure state can be written as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \qquad |\alpha|^2 + |\beta|^2 = 1,$$

where the complex numbers α and β are called *amplitudes*. Here, we are using Dirac's *bra–ket notation*: a ket $|\psi\rangle$ is a column vector, the corresponding bra $\langle\psi|$ is its conjugate transpose, and $\langle\phi|\psi\rangle$ denotes the inner product between $|\phi\rangle$ and $|\psi\rangle$.

[1] For more complete introductions, see Watrous [Wat25] and Nielsen and Chuang [NC10]. For video lectures, see Cleve [Cle20] and O'Donnell [ODo24]; for quantum learning theory specifically, see Wright [Wri24].

The simplest measurement is in the computational basis. Applied to $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, a computational-basis measurement returns 0 with probability $|\alpha|^2$ and 1 with probability $|\beta|^2$. More generally, we may measure a qubit in any orthonormal basis $\{|u\rangle, |v\rangle\}$. The probabilities of observing u and v are $|\langle u|\psi\rangle|^2$ and $|\langle v|\psi\rangle|^2$, respectively.

The computational-basis probabilities depend only on the magnitudes of the amplitudes, and therefore do not determine the state. For example, the states

$$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \qquad \text{and} \qquad |-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}$$

have the same computational-basis probabilities, but a measurement in the basis $\{|+\rangle, |-\rangle\}$ distinguishes them perfectly.

To describe the difference, suppose both amplitudes are nonzero and write $\alpha = |\alpha|e^{i\theta_0}$ and $\beta = |\beta|e^{i\theta_1}$. The difference $\theta_1 - \theta_0$, understood modulo 2π, is called their *relative phase*. It is 0 for $|+\rangle$ and π for $|-\rangle$. As the example above shows, relative phase can affect measurement statistics and is therefore physically meaningful. By contrast, multiplying both amplitudes by the same *global phase* $e^{i\theta}$ changes no measurement statistics. Thus, states that differ only by a global phase describe the same physical state.

At this point, it is worth separating the mathematical abstraction from its physical realization. A qubit is not a particular particle or device; it is an *abstract two-level quantum system*. Much like a classical bit can be physically represented in many different ways—by a voltage, a magnetic orientation, and so on—a qubit can be realized using many different physical systems. The states, and the operations and measurements we perform on them, describe the abstraction; how these are implemented physically depends on the particular hardware.

Thus, it is perfectly natural to ask what it means to *physically* measure a qubit in the computational basis, or to perform some particular quantum operation. The answer depends on how the qubit has been realized. In this thesis we mostly work one level above these implementation details: we take quantum states, operations, and measurements as our basic primitives and study what can be learned from them.

1.2.2 How do qubits combine?

The state space of n qubits is the n-fold tensor product

$$\mathcal{H}_n = (\mathbb{C}^2)^{\otimes n} \cong \mathbb{C}^{2^n}.$$

Its computational basis consists of vectors

$$|x\rangle := |x_1\rangle \otimes \cdots \otimes |x_n\rangle, \qquad x = x_1 \cdots x_n \in \{0,1\}^n,$$

and a general pure state is

$$|\psi\rangle = \sum_{x \in \{0,1\}^n} \alpha_x |x\rangle, \qquad \sum_x |\alpha_x|^2 = 1.$$

Thus a general n-qubit state vector has 2^n amplitudes. This is the exponential growth in description size discussed earlier.

When two systems are prepared separately in states $|\psi\rangle$ and $|\varphi\rangle$, their joint state is the tensor product $|\psi\rangle \otimes |\varphi\rangle$, which we typically abbreviate as $|\psi\rangle|\varphi\rangle$. For example,

$$|+\rangle \otimes |0\rangle = \frac{|00\rangle + |10\rangle}{\sqrt{2}}.$$

Not every two-qubit state has this product form. The Bell state

$$|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

cannot be written as a state of the first qubit tensored with a state of the second. Such a state is called *entangled*.

1.2.3 What is a mixed state?

Pure states do not describe every situation. To handle more general situations, it is useful to represent a pure state $|\psi\rangle$ by the rank-one projector $|\psi\rangle\langle\psi|$. This representation lets us describe randomized preparations in the same language: if a procedure produces $|\psi_i\rangle$ with probability p_i, the resulting state has density matrix

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|.$$

Different ensembles can give the same density matrix. For example,

$$\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -| = \frac{I}{2}.$$

Given only the resulting quantum system, no measurement can determine which of these two ensembles was used: the density matrix, rather than a particular ensemble, is the physical description of the state.

The possible density matrices are exactly the positive semidefinite matrices with trace one. *Pure states* are the rank-one cases $\rho = |\psi\rangle\langle\psi|$; all other density matrices are called *mixed states*. Mixed states can arise from classical uncertainty in the preparation, from noise, or from describing only part of a larger entangled system. For m independently prepared copies of ρ, the joint state is $\rho^{\otimes m}$.

1.2.4 What does a measurement reveal?

As mentioned earlier, a *basis measurement* is specified by an orthonormal basis $\{|u_y\rangle\}_y$. On a pure state $|\psi\rangle$, the Born rule gives

$$\Pr[Y = y] = |\langle u_y|\psi\rangle|^2.$$

More generally, a measurement with outcomes y is described by positive semidefinite operators $\{M_y\}_y$ satisfying

$$\sum_y M_y = I.$$

When the state is ρ, the outcome probabilities are

$$\Pr[Y = y] = \mathrm{tr}(M_y \rho).$$

A basis measurement is the special case $M_y = |u_y\rangle\langle u_y|$. These conditions ensure that the probabilities are nonnegative and sum to one.

Our learning algorithms specify measurements to perform on copies of the unknown state, followed by classical postprocessing of their outcomes. An important question is how complicated these measurements need to be.

Within a single multiqubit copy, we may measure each qubit separately, or perform a joint measurement involving several qubits at once. We refer to the former as *single-qubit measurements*. For example, the two-outcome measurement

$$\{|\Phi^+\rangle\langle\Phi^+|,\ I - |\Phi^+\rangle\langle\Phi^+|\}$$

directly tests whether a two-qubit state is the Bell state $|\Phi^+\rangle$; this is a joint measurement of the two qubits and cannot be implemented simply by measuring each qubit separately.

There is a separate distinction across copies. An algorithm may measure each copy individually, or it may keep several copies in quantum memory and perform a *collective measurement* on their joint state $\rho^{\otimes m}$. Collective measurements can be substantially more powerful, but also require more demanding quantum operations.

Finally, measurements may be *adaptive*: the measurement performed later may depend on the classical outcomes of earlier measurements. In a *nonadaptive* protocol, all measurements are fixed in advance. Thus, among the simplest protocols are those that use nonadaptive, single-qubit measurements, while some learning problems require more powerful forms of measurement.

1.2.5 How do we compare quantum states?

A learning algorithm rarely recovers an unknown state exactly, so we need a way to say when its estimate is close to the truth. For pure states, a common measure is the *fidelity*,

$$F(\psi, \phi) := |\langle\psi|\phi\rangle|^2.$$

It lies between zero and one: it equals one exactly when the states differ by only a global phase, and it equals zero when the states are orthogonal. For example, $|0\rangle$ and $|1\rangle$ have fidelity zero, whereas $|0\rangle$ and $|+\rangle$ have fidelity $1/2$. More generally, when comparing a pure state $|\psi\rangle$ with a mixed state ρ, we write

$$F(\psi, \rho) := \langle\psi|\rho|\psi\rangle.$$

Operationally, this is the probability that the two-outcome measurement $\{|\psi\rangle\langle\psi|, I - |\psi\rangle\langle\psi|\}$ returns the outcome corresponding to $|\psi\rangle$.

Another important measure is the *trace distance*,

$$d_{\mathrm{tr}}(\rho, \sigma) \coloneqq \frac{1}{2}\|\rho - \sigma\|_1,$$

where $\|A\|_1$ is the sum of the singular values of A. Trace distance also lies between zero and one, and characterizes how well the two states can be distinguished by measurements. States of trace distance zero are indistinguishable, while states of trace distance one can be distinguished perfectly. For pure states,

$$d_{\mathrm{tr}}(\psi, \phi) = \sqrt{1 - F(\psi, \phi)}.$$

Thus $|0\rangle$ and $|1\rangle$ have trace distance one, while $|0\rangle$ and $|+\rangle$ have trace distance $1/\sqrt{2}$. The general definitions and relationships between these distances are collected in Section 2.3.

1.2.6 What is a quantum circuit?

Quantum states evolve from one state to another through *unitary transformations*: linear maps U satisfying $U^\dagger U = UU^\dagger = I$. If the system begins in state $|\psi\rangle$, applying U produces the state $U|\psi\rangle$. A *quantum circuit* builds such a transformation by composing elementary unitary gates, each acting on a small number of qubits.

The simplest gates act on a single qubit. For example, the Hadamard gate

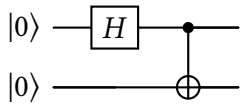
$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

maps $|0\rangle$ to $|+\rangle$. We also use gates that act on two qubits at once. An important example is the *controlled-NOT* (CNOT) gate, which acts on computational-basis states as

$$\mathrm{CNOT}|x, y\rangle = |x, x \oplus y\rangle, \qquad x, y \in \{0, 1\}.$$

In other words, the second qubit is flipped if the first qubit is 1, and is left unchanged if the first qubit is 0. More generally, quantum circuits can be built by composing gates acting on only one or two qubits.

Combining a Hadamard and a CNOT gives a simple circuit for preparing the Bell state $|\Phi^+\rangle$ from $|00\rangle$:



Here, the dot and target symbol together denote the CNOT gate. The state evolves as

$$|00\rangle \xrightarrow{H\otimes I} \frac{|00\rangle + |10\rangle}{\sqrt{2}} \xrightarrow{\text{CNOT}} \frac{|00\rangle + |11\rangle}{\sqrt{2}} = |\Phi^+\rangle.$$

This example views a circuit as a procedure for preparing a state from a fixed input. But a unitary transformation specifies how the circuit acts on *every* possible input, and learning this transformation is a different task from learning a single output state. For example, the identity and the phase-flip gate $Z = \mathrm{diag}(1, -1)$ both fix $|0\rangle$, but they send $|+\rangle$ to $|+\rangle$ and $|-\rangle$, respectively. The final group of results in this thesis concerns learning an unknown unitary transformation in this sense. Its access model and notion of approximation will be introduced when those results are stated.

1.3 Main Results

We now state the main results of the thesis, following the order of the technical chapters.

1.3.1 Tomography with Simple Measurements

We begin with quantum state tomography, the task of recovering a classical description of an unknown quantum state using as few copies and as little computation as possible. Suppose, for simplicity, that the unknown state is a pure state $|\psi\rangle$ on n qubits, and write $d = 2^n$. Given an accuracy parameter ϵ and failure probability δ, the goal is to output a classical description of a state $|\hat{\psi}\rangle$ satisfying

$$F\left(\hat{\psi}, \psi\right) \geq 1 - \epsilon$$

with probability at least $1 - \delta$.

Tomography is one of the most basic tasks in quantum learning, with a history dating back at least to the 1950s [Fan57]. A precursor is the question of whether a quantum state is determined by its measurement statistics in a small collection of bases, which goes back to Pauli in the 1930s [Pau33]. Beyond its foundational role, tomography is

also routinely used to reconstruct states in experiments and to characterize and validate quantum devices [Ma+12; Del+08; Gup+24; DBBHK25].

Even if the learner may perform arbitrary measurements, the optimal worst-case copy complexity of pure-state tomography is $\Theta(d/\epsilon)$ for constant failure probability [Hay98]. Moreover, this complexity can be achieved without performing joint measurements across multiple copies of the state [GKKT20]. This is in contrast to tomography of rank-r mixed states, where achieving the optimal copy complexity requires collective measurements across copies [HHJWY16; OW16; SSW25].

This leaves two natural questions about the complexity of pure-state tomography. First, how simple can the measurements be? Can we attain the optimal copy complexity while measuring each qubit separately, with all measurements chosen nonadaptively in advance?

Second, can tomography be made computationally optimal as well? Previous copy-efficient algorithms require measurements whose implementation has circuit depth exponential in n.

The main result of Chapter 3 answers both questions in the affirmative. We give a pure-state tomography algorithm with optimal copy and time complexity, up to logarithmic factors, using only nonadaptive single-qubit measurements. In particular, each measurement can be implemented by a depth-1 quantum circuit. To the best of our knowledge, this is the first pure-state tomography algorithm with near-optimal running time.

Theorem 1.1 (Pure-state tomography with Pauli-basis measurements). *Let $n \geq 1$ and $\epsilon, \delta \in (0,1)$. There is a nonadaptive algorithm which, given copies of an unknown n-qubit pure state $|\psi\rangle$, measures each copy separately in a Pauli product basis and uses*

$$\widetilde{O}\left(\frac{2^n \log(1/\delta)}{\epsilon}\right)$$

copies in total. With probability at least $1-\delta$, it outputs a state $|\hat{\psi}\rangle$ such that

$$F\left(\hat{\psi}, \psi\right) \geq 1 - \epsilon.$$

The algorithm runs in time $\widetilde{O}(2^n \log(1/\delta)/\epsilon)$.

The algorithm takes a *divide-and-conquer* approach to tomography. To learn an n-qubit state $|\psi\rangle$, write

$$|\psi\rangle = \alpha_0|0\rangle|\psi_0\rangle + \alpha_1|1\rangle|\psi_1\rangle,$$

where $|\psi_0\rangle$ and $|\psi_1\rangle$ are the $(n-1)$-qubit states obtained by conditioning on the first qubit being 0 or 1. We recursively learn these two smaller states and then stitch them back together.

The difficult part is recovering the relative phase between the two branches. Learning $|\psi_0\rangle$ and $|\psi_1\rangle$ separately determines each only up to a global phase, but their relative phase is physically meaningful in the full state $|\psi\rangle$. Our algorithm estimates this missing phase information using only single-qubit Pauli-basis measurements. Recursing on the remaining qubits ultimately reduces the problem to single-qubit tomography.

1.3.2 Learning Structured Quantum States

Next we turn to tomography of n-qubit pure states that are promised to have additional structure. While learning an arbitrary quantum state requires resources exponential in n, many physically and computationally interesting families of states admit much more succinct descriptions. This raises a natural question: when can such structure be exploited to learn the state efficiently? We give two examples where such structure can be exploited to obtain efficient learning algorithms.

Non-interacting fermions. Fermions are particles, such as electrons, that obey the *Pauli exclusion principle*: no two fermions can occupy the same mode. It is useful to think of the modes as bins and the fermions as balls, with at most one ball allowed in each bin. Thus, a configuration of n fermions among m modes can be identified with a subset $S \subseteq [m]$ of size n. Since these configurations can occur in superposition, a general pure state has the form

$$|\psi\rangle = \sum_{\substack{S\subseteq[m] \\ |S|=n}} \alpha_S|S\rangle,$$

and lives in a Hilbert space of dimension $\binom{m}{n}$.

A pure non-interacting fermion state, also called a *Slater determinant*, has much more structure. It is specified by an $m \times n$ column-orthonormal matrix A. Concretely, its amplitudes are determinants:

$$|\psi_A\rangle = \sum_{\substack{S \subseteq [m] \\ |S|=n}} \det(A_S)|S\rangle,$$

where A_S is the $n \times n$ submatrix of A formed by the rows indexed by S. Thus, although the state has $\binom{m}{n}$ amplitudes, it admits a description using only mn parameters. In Chapter 4, we present an efficient algorithm for recovering such a description.

Theorem 1.2 (Tomography of non-interacting fermion states). *Let $|\Psi\rangle$ be a Slater determinant of n fermions in m modes, and let $\epsilon, \delta \in (0,1)$. There is an algorithm using*

$$O\left(\frac{m^3 n^2 \log(m/\delta)}{\epsilon^4}\right)$$

copies of $|\Psi\rangle$ and

$$O\left(\frac{m^4 n^2 \log(m/\delta)}{\epsilon^4}\right)$$

classical time. With probability at least $1-\delta$, it outputs an $m \times n$ column-orthonormal matrix $\widehat{A}$ whose associated Slater determinant $|\widehat{\Psi}\rangle$ satisfies

$$d_{\mathrm{tr}}\left(\widehat{\Psi}, \Psi\right) \leq \epsilon.$$

The algorithm reconstructs A by interfering pairs of modes and measuring their occupation numbers. The main analytic step shows that sufficiently accurate estimates of these measurements determine an accurate estimate of the entire many-body state. Thus, tomography of a state with $\binom{m}{n}$ amplitudes reduces to learning a polynomial-size matrix.

Our work initiated the study of efficient tomography for Slater determinants, and subsequent work improved these guarantees [BMEL25; CZ26]. The best currently known polynomial-time algorithm, due to Christensen and Zhao [CZ26], uses

$$O\left(\frac{mn^2 \log(m/\delta)}{\epsilon^2}\right)$$

copies and runs in time

$$O\left(\frac{m^3 n^3 \log(m/\delta)}{\epsilon^2}\right).$$

Near-stabilizer states. Our second example comes from the *stabilizer formalism*. An n-qubit *stabilizer state* is a state that is an eigenvector of n independent Pauli operators. Equivalently, stabilizer states are precisely the states that can be prepared from $|0^n\rangle$ by Clifford circuits. These states are highly structured, and Clifford circuits can be simulated efficiently on a classical computer [Got97; Got98; AG04]. On the other hand, Clifford circuits together with the non-Clifford T-gate form a universal gate set for quantum computation.

In many settings, the number of non-Clifford gates—and in particular the number of T-gates—is itself a key complexity parameter. For example, the cost of classical simulation algorithms and of fault-tolerant implementations is often analyzed in terms of the T-count of the circuit [BG16; BK05]. This suggests a natural learning question: given that stabilizer states can be learned in polynomial time [Mon17], does efficient learnability persist for states prepared using only a small number of non-Clifford gates?

Our answer is phrased in terms of *stabilizer dimension*, the number of independent Pauli operators for which the state is an eigenvector. A stabilizer state has the maximum possible stabilizer dimension n. More generally, we consider states whose stabilizer dimension is at least $n - t$. Any state prepared by a Clifford circuit together with at most s arbitrary single-qubit non-Clifford gates has stabilizer dimension at least $n - 2s$; see Lemma 4.8. Thus, our results in particular apply to states prepared with few T-gates.

The class is considerably more general than this circuit model. For example, for an arbitrary t-qubit state $|\phi\rangle$, the state

$$|0^{n-t}\rangle \otimes |\phi\rangle$$

has stabilizer dimension at least $n - t$, regardless of how complicated $|\phi\rangle$ is to prepare. In Chapter 4, we show that every pure state with sufficiently large stabilizer dimension can nevertheless be learned efficiently [GIKL25].

Theorem 1.3 (Tomography from high stabilizer dimension). *Let $t \in \{0, \dots, n\}$, and let $|\psi\rangle$ be an n-qubit pure state of stabilizer dimension at least $n - t$. Given $\epsilon, \delta \in (0, 1)$, there are algorithms that, with probability at least $1 - \delta$, output a classical description of a state $|\widehat{\psi}\rangle$*

satisfying

$$d_{\mathrm{tr}}\left(\psi, \widehat{\psi}\right) \leq \epsilon.$$

Their copy complexities and running times are polynomial in n, 2^t, $1/\epsilon$, *and* $\log(1/\delta)$. *One algorithm uses Bell measurements between copies; the other uses only single-copy measurements.*

A key step in both algorithms is a *compression lemma*. Roughly speaking, from copies of $|\psi\rangle$ we learn a Clifford circuit C such that $C|\psi\rangle$ is close to a state of the form

$$|\phi\rangle \otimes |x\rangle,$$

where $|\phi\rangle$ is an arbitrary state on t qubits and $|x\rangle$ is a computational-basis state on the remaining $n - t$ qubits. Thus, all of the unstructured part of the state has been compressed into a t-qubit register. The string x is easy to learn, and we can finish by running general pure-state tomography only on $|\phi\rangle$.

1.3.3 Agnostic Tomography

The structured tomography problems above assume that the unknown state belongs to a particular class. In Chapter 5, we consider a more flexible learning model in which this promise is removed.

Let $\mathcal{C}$ be a class of pure states and let ρ be an arbitrary quantum state. Define

$$F_{\mathcal{C}}(\rho) := \sup_{|\varphi\rangle \in \mathcal{C}} \langle\varphi|\rho|\varphi\rangle.$$

$F_{\mathcal{C}}(\rho)$ is the fidelity achieved by the best approximation to ρ from the class $\mathcal{C}$. In *agnostic tomography*, a model we introduced in [GIKL26a], the goal is to output a succinct classical description of a state $|\hat{\varphi}\rangle \in \mathcal{C}$ satisfying

$$\langle\hat{\varphi}|\rho|\hat{\varphi}\rangle \geq F_{\mathcal{C}}(\rho) - \epsilon.$$

This generalizes structured tomography. If the unknown state is promised to lie in $\mathcal{C}$, then $F_{\mathcal{C}}(\rho) = 1$, and the task reduces to learning the state itself. Agnostic tomography

removes this promise entirely: the input need not belong to $\mathcal{C}$, or arise from any prescribed noise model. Instead, the learner competes with the best state in $\mathcal{C}$. Thus, a learning algorithm for a structured family can remain useful even when the input state does not exactly satisfy the assumed model.

As an example, consider the class Stab_n of n-qubit stabilizer states. Stabilizer states can be learned efficiently when the input is promised to be a stabilizer state. We ask whether one can still efficiently find the best stabilizer approximation to an arbitrary pure state. We give an efficient algorithm in the high-fidelity regime.

Theorem 1.4 (High-fidelity stabilizer approximation). *Let $|\psi\rangle$ be an n-qubit pure state, let $\delta \in (0,1)$, and suppose that*

$$F_{\mathsf{Stab}_n}(|\psi\rangle) \geq \cos^2(\pi/8) + \gamma$$

for some $\gamma > 0$. There is an algorithm that outputs the unique fidelity-maximizing stabilizer state with probability at least $1-\delta$. It uses

$$O\left(n + \frac{\log(n/\delta)}{\gamma^2}\right)$$

copies of $|\psi\rangle$ and runs in time

$$O\left(n^3 + \frac{n^2 \log(n/\delta) + n \log^2(1/\delta)}{\gamma^2}\right).$$

The threshold $\cos^2(\pi/8)$ is not an artifact of the analysis. At the threshold, the closest stabilizer state need not be unique: for example,

$$\cos(\pi/8)|0\rangle + \sin(\pi/8)|1\rangle$$

has equal fidelity with $|0\rangle$ and $|+\rangle$. Above the threshold, the closest stabilizer is unique, and our algorithm recovers it efficiently.

We discuss further examples and subsequent work on agnostic tomography in Chapter 5.

1.3.4 Testing Properties of Quantum States

The learning problems above ask for a classical description of an unknown state. In *quantum property testing*, we ask for much less: given a family $\mathcal{P}$ of pure states, the goal is only to decide whether the unknown state belongs to $\mathcal{P}$ or is far from every state in $\mathcal{P}$. More precisely, an ϵ-tester must distinguish $|\psi\rangle \in \mathcal{P}$ from the case that

$$\max_{|\phi\rangle \in \mathcal{P}} F(\psi, \phi) \leq 1 - \epsilon,$$

under the promise that one of the two cases holds. The hope is that deciding membership may require far fewer copies than learning the state itself.

Testing stabilizer dimension. Recall that an n-qubit state has stabilizer dimension k if it has k independent Pauli symmetries. For any k, we give an efficient test for whether a state has stabilizer dimension at least k. Perhaps surprisingly, the complexity is independent of k.

Theorem 1.5 (Testing stabilizer dimension). *Let $|\psi\rangle$ be an n-qubit pure state, let $k \in \{1, \ldots, n\}$, and let $\epsilon \in (0, 3/8)$ and $\delta \in (0, 1)$. There is an ϵ-tester with perfect completeness for the property of having stabilizer dimension at least k that succeeds with probability at least $1 - \delta$, uses*

$$O\left(\frac{n + \log(1/\delta)}{\epsilon}\right)$$

copies and runs in time

$$O\left(\frac{n^3 + n^2 \log(1/\delta)}{\epsilon}\right).$$

Our tester has a consequence for quantum pseudorandomness. It implies that pseudorandom states cannot be prepared by Clifford circuits with too few non-Clifford gates: in particular, any Clifford+T construction of pseudorandom states requires at least n T-gates [GIKL24].

Testing low-degree phase states. In classical low-degree testing, one is given query access to a function $f : \mathbb{F}_q^m \to \mathbb{F}_q$ and asked to decide whether f is a degree-d polynomial or is far from every such polynomial. Efficient classical testers are known for this problem [AKKLR05; BKSSZ10; JPRZ04]. We ask what happens when the values of the polynomial are instead encoded in the phases of a quantum state.

Let q be prime and let $\omega_q = e^{2\pi \mathrm{i}/q}$. For a polynomial $f : \mathbb{F}_q^m \to \mathbb{F}_q$, define the *phase state*

$$|\psi_f\rangle = \frac{1}{q^{m/2}} \sum_{x \in \mathbb{F}_q^m} \omega_q^{f(x)} |x\rangle.$$

Every amplitude has the same magnitude; all information about f is stored in the phases. In Chapter 6, we prove that testing whether such a state comes from a low-degree polynomial can require many copies.

Theorem 1.6 (Testing low-degree phase states, informal). *For every fixed prime q and degree $d \geq 3$, testing degree-d phase states over $\mathbb{F}_q^m$ requires*

$$\Omega\left(m^{\lfloor (d-1)/2 \rfloor}\right)$$

copies for constant error parameters. Consequently, there is no copy complexity that is polynomial jointly in m, d, and q.

For comparison, learning degree-d phase states requires $\Theta(m^{d-1})$ copies [ABDY23]. Thus, in this setting, testing is nearly as hard as learning: at best, one could hope to save only a factor of about two in the exponent.

Using a separate argument, we show that testing degree-2 phase states requires $\Omega(m)$ copies. Degree-2 phase states are a subset of stabilizer states, even though the larger family of all stabilizer states can be tested using only six copies. Thus, restricting to a seemingly simpler subfamily can make testing substantially harder.

Our lower bounds follow from a stronger pseudorandomness phenomenon: random low-degree phase states already approximate Haar-random ensembles. At the two extremes, degree-1 phase states are completely learnable from a single copy via the Bernstein–Vazirani algorithm, while fully random phase states are known to approximate Haar-random states.

Our results fill in the landscape between these cases. Already for $d = 3$, the uniform ensemble of degree-3 phase states forms an exponentially accurate approximate state t-design for $t = \Omega(m)$. More generally, degree-d yields design order $\Omega(m^{\lfloor (d-1)/2 \rfloor})$.

1.3.5 Learning Structured Quantum Circuits

Finally, Chapter 7 turns from learning quantum states to learning quantum transformations. Given black-box access to an unknown n-qubit unitary U, the goal is to reconstruct its action up to small diamond distance, the natural analogue of trace distance for quantum transformations. Without further assumptions, this problem has exponential cost: learning an arbitrary unitary requires $\Theta(4^n/\epsilon)$ queries [HKOT23]. We therefore ask what kinds of structure can make a unitary efficiently learnable.

Our approach is based on the Pauli expansion of operators. Every n-qubit operator has an expansion

$$A = \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} \widehat{A}(P) P,$$

and we call the coefficients $\{\widehat{A}(P)\}_P$ the *Pauli spectrum* of A, in analogy with the Fourier spectrum of a classical function. Our main technical contribution is efficient learning algorithms for unitary operators whose Pauli spectra have low *sparsity* or low *dimensionality*. Low sparsity means that only a small number of Pauli coefficients are nonzero, while low dimensionality means that the support is contained in a Pauli subgroup generated by only a small number of operators.

To learn a unitary U, we study how it transforms Pauli operators in the Heisenberg picture: $P \longmapsto U^\dagger P U$. Since the single-qubit X's and Z's generate the Pauli group, it suffices to understand this transformation on these $2n$ operators. Our framework learns the operators $U^\dagger P U$ whenever their Pauli spectra have sufficiently low complexity, and then compiles these estimates into an approximation of U. Thus, learning a structured unitary is reduced to understanding the Pauli structure of a small collection of Heisenberg-evolved operators.

This framework recovers learning algorithms for several previously studied circuit classes, including Clifford and near-Clifford circuits, shallow circuits, quantum juntas,

and fermionic matchgate circuits. It also gives algorithms for new combinations of these structures. For example, we efficiently learn compositions of depth-$O(\log \log n)$ circuits with near-Clifford circuits containing $O(\log n)$ non-Clifford single-qubit gates, as well as compositions of matchgate and Clifford circuits.

1.4 Bibliographical Notes

The results presented in this thesis are joint work with collaborators. Chapter 3 is based on work with Meghal Gupta, William He, Aniruddha Sen, and Mihir Singhal [GGHSS26]. Chapter 4 combines work with Scott Aaronson on non-interacting fermions [AG23] and with Vishnu Iyer, William Kretschmer, and Daniel Liang on near-stabilizer states [GIKL25]. The high-fidelity stabilizer result in Chapter 5 is from further work with Iyer, Kretschmer, and Liang [GIKL24]; subsequent work with the same collaborators introduced agnostic tomography and proved the stabilizer-product result stated there [GIKL26a]. Chapter 6 contains two further results with Iyer, Kretschmer, and Liang: the stabilizer-dimension test [GIKL25] and its consequence for the non-Clifford complexity of pseudorandom states [GIKL24]. The low-degree phase-state results are joint work with Omar Alrabiah, Srinivasan Arunachalam, and John Wright [AAGW26]. Finally, Chapter 7 is based on work with Daniel Liang [GL25].

Several additional papers were completed during the PhD but are not included in this thesis [GY24; AGIMR25; GIKL26b; GK24; Aar+24; Kre+25; GK25; GR26].

Chapter 2: Preliminaries

This chapter collects the notation and quantum-information facts that are used across several later chapters. It is intended as a reference rather than as a prerequisite to be read linearly: readers familiar with the basic formalism may consult individual sections as needed. More specialized background—including fermionic linear algebra, coding theory, and distances tailored to unitary channels—is introduced in the chapter where it first becomes relevant.

2.1 Conventions

All Hilbert spaces are finite-dimensional and complex. For a positive integer m, write $[m] \coloneqq \{1, \dots, m\}$. The identity operator is denoted by I, with its dimension supplied as a subscript when it is not clear from context. We write $A^\dagger$ for the conjugate transpose of A, $\mathrm{tr}(A)$ for its trace, and $\mathrm{rank}(A)$ for its rank. For a function or distribution f, its nonzero support is denoted by $\mathrm{supp}(f)$, and $\mathrm{span}(S)$ denotes the linear span of a set of vectors S. For vectors, $a \cdot b$ denotes the standard dot product over the ambient field. Unless explicitly stated otherwise, arithmetic on vectors in $\mathbb{F}_2^m$ is over $\mathbb{F}_2$.

The symbol log denotes the natural logarithm; when the base matters, it is written explicitly, as in $\log_2$ or $\log_q$. The notation $\widetilde{O}(f)$ suppresses factors logarithmic in the displayed parameters. Accuracy and failure parameters are generally written as ϵ and δ, respectively. Individual chapters may reuse letters such as n, d, and t; their meanings are specified locally.

For an n-qubit system, the Hilbert space is

$$\mathcal{H}_n \coloneqq (\mathbb{C}^2)^{\otimes n}, \qquad d \coloneqq \dim(\mathcal{H}_n) = 2^n.$$

Computational-basis vectors are indexed interchangeably by bit strings $x \in \{0,1\}^n$ and binary vectors $x \in \mathbb{F}_2^n$.

2.2 Quantum States and Measurements

Definition 2.1 (Quantum states). A *pure state* in a Hilbert space $\mathcal{H}$ is represented by a unit vector $|\psi\rangle \in \mathcal{H}$, or equivalently by its density operator $|\psi\rangle\langle\psi|$. More generally, a quantum state is a positive semidefinite operator ρ on $\mathcal{H}$ satisfying $\mathrm{tr}(\rho) = 1$. It is pure when $\mathrm{rank}(\rho) = 1$ and mixed otherwise.

The joint state of independently prepared systems is represented by a tensor product. In particular, m independent copies of ρ have state $\rho^{\otimes m}$. If ρ_{AB} is a state on $\mathcal{H}_A \otimes \mathcal{H}_B$, its reduced state on A is $\rho_A = \mathrm{tr}_B(\rho_{AB})$, where tr_B is the partial trace over B.

Definition 2.2 (Measurements). A measurement with finite outcome set $\mathcal{Y}$ is a positive operator-valued measure, or POVM, $\{M_y : y \in \mathcal{Y}\}$, whose elements satisfy

$$M_y \succeq 0, \qquad \sum_{y \in \mathcal{Y}} M_y = I.$$

On input ρ, the probability of outcome y is $\Pr[y] = \mathrm{tr}(M_y \rho)$. A projective measurement is a POVM whose nonzero elements are mutually orthogonal projectors.

Measuring a Hermitian observable $A = \sum_y \lambda_y \Pi_y$ means performing the projective measurement $\{\Pi_y\}_y$ and reporting the corresponding eigenvalue λ_y. For a Pauli observable P, this produces an outcome in $\{-1, +1\}$ with expectation $\mathrm{tr}(P\rho)$.

The state-learning and state-testing algorithms in this dissertation receive independent copies of an unknown state. Their *copy complexity* is the number of copies consumed. Measurements may act jointly on several copies (*collective measurements*) or separately on each copy (*single-copy measurements*), and may be adaptive or nonadaptive depending on whether later measurement choices depend on earlier outcomes. We account separately for copy complexity, quantum circuit cost, and classical running time. The unitary-learning model of Chapter 7 instead counts queries to an unknown transformation and is defined there.

2.3 Norms and Distance Measures

Let $A \in \mathbb{C}^{m\times n}$, set $r = \min\{m, n\}$, and let $\sigma_1(A), \dots, \sigma_r(A)$ be its singular values, including zeros. The Schatten norms used below are

$$\|A\|_1 = \sum_{i=1}^{r} \sigma_i(A), \qquad \|A\|_{\mathrm{F}} = \left(\sum_{i=1}^{r} \sigma_i(A)^2\right)^{1/2}, \qquad \|A\|_{\mathrm{op}} = \max_i \sigma_i(A).$$

They are, respectively, the trace, Frobenius, and operator norms. Each is unitarily invariant: for compatible unitaries U, V, $\|UAV\| = \|A\|$. We will use the standard inequalities

$$\|A\|_{\mathrm{op}} \le \|A\|_{\mathrm{F}} \le \|A\|_1 \le \sqrt{r}\,\|A\|_{\mathrm{F}} \le r\,\|A\|_{\mathrm{op}}, \qquad r = \operatorname{rank}(A). \tag{2.1}$$

For an identity operator I_d, these norms also satisfy

$$\|I_d \otimes A\|_{\mathrm{F}} = \sqrt{d}\,\|A\|_{\mathrm{F}}, \qquad \|I_d \otimes A\|_{\mathrm{op}} = \|A\|_{\mathrm{op}}. \tag{2.2}$$

Definition 2.3 (Fidelity and trace distance). For density operators ρ and σ, their *fidelity* and *trace distance* are

$$F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1^2, \qquad d_{\mathrm{tr}}(\rho, \sigma) := \frac{1}{2}\|\rho - \sigma\|_1.$$

The associated error $1 - F(\rho, \sigma)$ is called infidelity.

If the first state is pure, then

$$F(\psi, \rho) := F(|\psi\rangle\langle\psi|, \rho) = \langle\psi|\rho|\psi\rangle. \tag{2.3}$$

In particular,

$$F(\psi, \phi) = |\langle\psi|\phi\rangle|^2. \tag{2.4}$$

We freely use a ket in an argument of $F(\cdot,\cdot)$ or $d_{\mathrm{tr}}(\cdot,\cdot)$ as shorthand for its rank-one projector.

The fidelity and trace distance satisfy the Fuchs–van de Graaf inequalities

$$1 - \sqrt{F(\rho, \sigma)} \le d_{\mathrm{tr}}(\rho, \sigma) \le \sqrt{1 - F(\rho, \sigma)}. \tag{2.5}$$

For two pure states, the upper bound is an equality:

$$d_{\mathrm{tr}}(\psi, \phi) = \sqrt{1 - |\langle\psi|\phi\rangle|^2}. \tag{2.6}$$

The Frobenius distance between states is simply $\|\rho - \sigma\|_{\mathrm{F}}$. For pure states we use the abbreviation

$$\begin{aligned} d_F(|\psi\rangle, |\phi\rangle) &:= \||\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\|_{\mathrm{F}} \\ &= \sqrt{2(1 - |\langle\psi|\phi\rangle|^2)} = \sqrt{2}\, d_{\mathrm{tr}}(\psi, \phi). \end{aligned} \tag{2.7}$$

Thus, for pure states, infidelity and squared trace distance agree, while squared Frobenius distance differs from them by a factor of two.

We will also use that fidelity cannot decrease when a register is discarded. If $\rho_A = \mathrm{tr}_B(\rho_{AB})$ and $\sigma_A = \mathrm{tr}_B(\sigma_{AB})$, then

$$F(\rho_A, \sigma_A) \geq F(\rho_{AB}, \sigma_{AB}).$$

2.4 Concentration Inequalities

We record the concentration inequalities used repeatedly in the technical chapters. The random variables below need not be identically distributed unless this is stated explicitly.

Lemma 2.4 (Hoeffding's inequality). *Let $X_1, \dots, X_m$ be independent random variables such that $X_i \in [a_i, b_i]$ almost surely. If $S = \sum_{i=1}^m X_i$, then, for every $t \geq 0$,*

$$\Pr[S - \mathbb{E}[S] \geq t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^m (b_i - a_i)^2}\right),$$
$$\Pr[\mathbb{E}[S] - S \geq t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^m (b_i - a_i)^2}\right),$$

and consequently

$$\Pr[|S - \mathbb{E}[S]| \geq t] \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^m (b_i - a_i)^2}\right).$$

Lemma 2.5 (Bernstein's inequality). *Let $Y_1, \dots, Y_m$ be independent mean-zero random variables satisfying $|Y_i| \leq B$ almost surely, and put*

$$V := \sum_{i=1}^m \mathbb{E}[Y_i^2].$$

Then, for every $t \geq 0$,

$$\Pr\left[\left|\sum_{i=1}^m Y_i\right| \geq t\right] \leq 2\exp\left(-\frac{t^2}{2(V + Bt/3)}\right). \tag{2.8}$$

The following consequence is convenient when the variance is controlled by the mean.

Corollary 2.6 (Bernstein's inequality on $[0, B]$). *Let $X_1, \dots, X_m$ be independent random variables with $X_i \in [0, B]$ almost surely, and set*

$$S := \sum_{i=1}^{m} X_i, \qquad \mu := \mathbb{E}[S].$$

Then, for every $\gamma \geq 1$,

$$\Pr\Big[|S - \mu| > \gamma(\sqrt{B\mu} + B)\Big] \leq 2\exp(-\gamma/2).$$

Proof. Apply Lemma 2.5 to $Y_i = X_i - \mathbb{E}[X_i]$. Since

$$\mathbb{E}[Y_i^2] \leq \mathbb{E}[X_i^2] \leq B\,\mathbb{E}[X_i],$$

the variance proxy satisfies $V \leq B\mu$. Taking $t = \gamma(\sqrt{B\mu} + B)$ in (2.8) gives

$$\frac{t^2}{2(B\mu + Bt/3)} \geq \frac{\gamma}{2} \qquad (\gamma \geq 1),$$

which proves the claim. □

Lemma 2.7 (Multiplicative Chernoff bound). *Let $X_1, \dots, X_m$ be independent Bernoulli random variables, not necessarily with the same success probability, and set $S = \sum_{i=1}^{m} X_i$ and $\mu = \mathbb{E}[S]$. For $\eta \in [0, 1]$,*

$$\Pr[S \leq (1 - \eta)\mu] \leq \exp(-\eta^2\mu/2),$$

while, for every $\eta \geq 0$,

$$\Pr[S \geq (1 + \eta)\mu] \leq \exp\left(-\frac{\eta^2\mu}{2 + \eta}\right).$$

Several later algorithms require a prescribed number of successful trials, rather than a fixed multiplicative approximation to their mean. The next form also allows the success probability of a trial to depend on earlier outcomes.

Lemma 2.8 (Chernoff bound for conditional successes). *Let $p \in (0, 1]$, and let $X_1, \dots, X_m \in \{0, 1\}$ be random variables such that, for every i and every history of the preceding variables with positive probability,*

$$\Pr[X_i = 1 \mid X_1, \dots, X_{i-1}] \geq p.$$

Then, for every integer $d \in \{0, \dots, m\}$,

$$\Pr\left[\sum_{i=1}^{m} X_i < d\right] \le \exp(d - mp/2).$$

Consequently, for $\delta \in (0,1)$, the failure probability is at most δ whenever

$$m \ge \frac{2}{p}\left(d + \log\frac{1}{\delta}\right).$$

Proof. For every i, conditional on the preceding variables,

$$\mathbb{E}[2^{-X_i} \mid X_1, \dots, X_{i-1}] \le 1 - p/2.$$

Iterated conditional expectation and Markov's inequality therefore give

$$\begin{aligned}\Pr\left[\sum_{i=1}^{m} X_i < d\right] &\le 2^d\, \mathbb{E}\left[2^{-\sum_{i=1}^{m} X_i}\right] \\ &\le 2^d (1 - p/2)^m \\ &\le \exp(d \log 2 - mp/2) \le \exp(d - mp/2).\end{aligned}$$

Substituting the stated lower bound on m proves the second assertion. □

2.5 Pauli Operators and Symplectic Linear Algebra

Let

$$\mathcal{P}_n := \left\{\alpha P_1 \otimes \cdots \otimes P_n \,:\, P_j \in \{I, X, Y, Z\}, \alpha \in \{\pm 1, \pm \mathrm{i}\}\right\}$$

be the n-qubit Pauli group. It is convenient to choose one Hermitian representative from each Pauli operator modulo phase. For $x = (a, b) \in \mathbb{F}_2^n \times \mathbb{F}_2^n$, define

$$W_x := \mathrm{i}^{a' \cdot b'} \bigotimes_{j=1}^{n} X^{a_j} Z^{b_j}, \tag{2.9}$$

where $a', b' \in \mathbb{Z}^n$ are the natural integer embeddings of a, b. The family $\{W_x \,:\, x \in \mathbb{F}_2^{2n}\}$ is the set of *Weyl operators*. Every coset of $\mathcal{P}_n/\{\pm 1, \pm \mathrm{i}\}$ has exactly one such representative. Each W_x is Hermitian and unitary, and $W_x^2 = I$. Note that the integer embeddings in Eq. (2.9) matter: the exponent records the number of positions at which both a_j and b_j are one, not merely its parity. This convention makes, for example, $W_{(1,1)} = Y$ on one qubit.

Definition 2.9 (Symplectic product). For $x = (a, b)$ and $y = (c, d)$ in $\mathbb{F}_2^n \times \mathbb{F}_2^n$, their *symplectic product* is

$$[x, y] \coloneqq a \cdot d + b \cdot c \in \mathbb{F}_2.$$

This bilinear form encodes Pauli commutation:

$$W_x W_y = (-1)^{[x,y]} W_y W_x. \tag{2.10}$$

Thus W_x and W_y commute exactly when $[x, y] = 0$. We often pass freely between a vector x and its Weyl representative W_x. The two coordinate subspaces

$$\mathcal{X} \coloneqq \mathbb{F}_2^n \times 0^n, \qquad \mathcal{Z} \coloneqq 0^n \times \mathbb{F}_2^n$$

index the strings containing only X's and only Z's, respectively.

Definition 2.10 (Symplectic complement). For a subspace $A \subseteq \mathbb{F}_2^{2n}$, its *symplectic complement* is

$$A^{\perp_s} \coloneqq \{x \in \mathbb{F}_2^{2n} \,:\, [x, a] = 0 \text{ for every } a \in A\}.$$

The subspace A is *isotropic* if $A \subseteq A^{\perp_s}$, and it is *Lagrangian* if $A = A^{\perp_s}$.

Equivalently, $A^{\perp_s}$ indexes all Weyl operators commuting with every element indexed by A. Nondegeneracy of the symplectic form gives

$$\begin{gathered} (A^{\perp_s})^{\perp_s} = A, \qquad \dim(A) + \dim(A^{\perp_s}) = 2n, \\ A \subseteq B \iff B^{\perp_s} \subseteq A^{\perp_s}, \qquad (A + B)^{\perp_s} = A^{\perp_s} \cap B^{\perp_s}. \end{gathered} \tag{2.11}$$

An isotropic subspace has dimension at most n; it is Lagrangian exactly when its dimension is n.

We record two elementary subspace identities used later.

Lemma 2.11 (Symplectic subspace identities). *Let $A, B, C \subseteq \mathbb{F}_2^{2n}$ be subspaces.*

(i) *If $C \subseteq A$, then $A \cap (B + C) = (A \cap B) + C$.*

(ii) *If A is isotropic and B is Lagrangian, then $A^{\perp_s} \cap (A + B)$ is Lagrangian.*

Proof. For the first identity, if $c \in C \subseteq A$, then $b + c \in A$ if and only if $b \in A$. For the second, use Eq. (2.11) and the first identity to obtain

$$\begin{aligned}\big(A^{\perp_s} \cap (A + B)\big)^{\perp_s} &= A + (A + B)^{\perp_s} \\ &= A + \big(A^{\perp_s} \cap B^{\perp_s}\big) \\ &= A^{\perp_s} \cap (A + B^{\perp_s}).\end{aligned}$$

Since $B = B^{\perp_s}$, the final expression is the original subspace. □

The following character-orthogonality identity is the Fourier-analytic form of taking a symplectic complement.

Proposition 2.12 (Symplectic character orthogonality). *For every subspace $A \subseteq \mathbb{F}_2^{2n}$ and every $a \in \mathbb{F}_2^{2n}$,*

$$\mathop{\mathbb{E}}_{x \in A^{\perp_s}} (-1)^{[a,x]} = 1_{a \in A}.$$

Proof. If $a \in A = (A^{\perp_s})^{\perp_s}$, every term is one. Otherwise, there is a $y \in A^{\perp_s}$ with $[a, y] = 1$. Pairing each x with $x + y$ cancels the sum. □

Lemma 2.13 (Computing a symplectic complement). *Given m vectors spanning $A \subseteq \mathbb{F}_2^{2n}$, a basis for $A^{\perp_s}$ can be computed in*

$$O\big(mn \min\{m, n\}\big)$$

time.

Proof. Place the vectors into the rows of an $m \times 2n$ matrix and exchange its left and right $m \times n$ blocks. The nullspace of the resulting matrix is exactly $A^{\perp_s}$. Gaussian elimination gives the stated running time. □

2.5.1 Pauli Expansions

The Weyl operators form an orthonormal basis for matrices under the normalized Hilbert–Schmidt inner product

$$\langle A, B \rangle_{\mathrm{nHS}} := 2^{-n} \operatorname{tr}(A^\dagger B). \tag{2.12}$$

This fixes the normalization of every Pauli coefficient used below.

Definition 2.14 (Pauli expansion). For $A \in \mathbb{C}^{2^n \times 2^n}$, define

$$\widehat{A}(x) \coloneqq 2^{-n} \operatorname{tr}(W_x A).$$

Then A has the unique expansion

$$A = \sum_{x \in \mathbb{F}_2^{2n}} \widehat{A}(x) W_x.$$

The coefficients $\{\widehat{A}(x)\}_x$ form the *Pauli spectrum* of A.

Parseval's identity in this normalization is

$$\|A\|_{\mathrm{F}}^2 = 2^n \sum_{x \in \mathbb{F}_2^{2n}} |\widehat{A}(x)|^2. \tag{2.13}$$

For a density operator ρ, the expansion becomes

$$\rho = 2^{-n} \sum_{x \in \mathbb{F}_2^{2n}} \operatorname{tr}(W_x \rho) W_x, \qquad 2^{-n} \sum_x \operatorname{tr}(W_x \rho)^2 = \operatorname{tr}(\rho^2). \tag{2.14}$$

The squared expectations are real because the Weyl operators are Hermitian.

2.6 Clifford Operations and Stabilizer Dimension

The Clifford group consists of the unitaries that normalize the Pauli group. It is generated by Hadamard, phase, and CNOT gates. If C is a Clifford unitary, conjugation induces a linear map on $\mathbb{F}_2^{2n}$, which we also denote by C, determined by

$$W_{C(x)} = \pm C W_x C^\dagger. \tag{2.15}$$

This map is a symplectic automorphism: it preserves symplectic products, inclusions, dimensions, and symplectic complements. In particular, Clifford changes of basis preserve all of the stabilizer-dimension notions introduced below.

Definition 2.15 (Unsigned stabilizer group). For an n-qubit pure state $|\psi\rangle$, its *unsigned stabilizer group* is

$$\mathrm{Weyl}(|\psi\rangle) \coloneqq \{x \in \mathbb{F}_2^{2n} \,:\, W_x|\psi\rangle = \pm|\psi\rangle\}.$$

Two Pauli operators that stabilize the same pure state must commute. Moreover, multiplication of Pauli operators corresponds, up to sign, to addition of their labels in $\mathbb{F}_2^{2n}$. It follows that $\mathrm{Weyl}(|\psi\rangle)$ is a linear subspace of $\mathbb{F}_2^{2n}$ whose elements have pairwise-zero symplectic product. In other words, $\mathrm{Weyl}(|\psi\rangle)$ is isotropic, and therefore $\dim \mathrm{Weyl}(|\psi\rangle) \leq n$.

Definition 2.16 (Stabilizer dimension). The *stabilizer dimension* of a pure state $|\psi\rangle$ is

$$\dim \mathrm{Weyl}(|\psi\rangle).$$

An n-qubit pure state of stabilizer dimension n is a *stabilizer state*. The quantity $n - \dim \mathrm{Weyl}(|\psi\rangle)$ is its *stabilizer nullity* [BCHK20].

Thus a pure state has stabilizer dimension k exactly when it is stabilized, up to signs, by an abelian group of 2^k Pauli operators. Under Clifford conjugation, the unsigned stabilizer group is simply relabeled by the induced linear map C:

$$\mathrm{Weyl}(C|\psi\rangle) = C\big(\mathrm{Weyl}(|\psi\rangle)\big) = \{C(x) \,:\, x \in \mathrm{Weyl}(|\psi\rangle)\}. \tag{2.16}$$

2.7 The Characteristic and Bell-Difference Distributions

For a pure state, its squared Pauli expectations can be normalized to form a probability distribution that records its Pauli symmetries.

Definition 2.17 (Characteristic distribution). For an n-qubit pure state $|\psi\rangle$, define

$$p_\psi(x) := 2^{-n}\langle\psi|W_x|\psi\rangle^2, \qquad x \in \mathbb{F}_2^{2n}.$$

Indeed, Eq. (2.14) gives

$$\sum_{x\in\mathbb{F}_2^{2n}} p_\psi(x) = 1. \tag{2.17}$$

Moreover, $0 \leq p_\psi(x) \leq 2^{-n}$, and $p_\psi(x) = 2^{-n}$ for every $x \in \mathrm{Weyl}(|\psi\rangle)$. Purity is important here: the same squared coefficients of a mixed state have total mass $\mathrm{tr}(\rho^2)$, which need not equal one.

Fact 2.18 (Large Weyl expectations commute). *Let*

$$M_{1/2}(\psi) := \{x \in \mathbb{F}_2^{2n} \,:\, 2^n p_\psi(x) > 1/2\}.$$

Every pair of elements of $M_{1/2}(\psi)$ has symplectic product zero.

Proof. If $x, y \in M_{1/2}(\psi)$ anticommute, apply the Schrödinger uncertainty relation to the observables W_x, W_y; the form used here is reviewed in [Sch30; AB08]. Since $W_x^2 = W_y^2 = I$, its left-hand side is

$$\big(1 - 2^n p_\psi(x)\big)\big(1 - 2^n p_\psi(y)\big) < \frac{1}{4}.$$

Anticommutation gives $\{W_x, W_y\} = 0$, so the squared covariance term on the right is $\langle\psi|W_x|\psi\rangle^2\langle\psi|W_y|\psi\rangle^2 = 4^n p_\psi(x)p_\psi(y) > 1/4$, a contradiction. □

Bell-difference sampling, introduced by Gross, Nezami, and Walter [GNW21], is a four-copy measurement with outcomes in $\mathbb{F}_2^{2n}$. It may be implemented by performing a Bell-basis measurement on each of two independent pairs of copies and reporting the difference of the two outcome labels.

Definition 2.19 (Bell-difference distribution). For an n-qubit pure state $|\psi\rangle$, define

$$q_\psi(a) := \sum_{x \in \mathbb{F}_2^{2n}} p_\psi(x)p_\psi(a+x), \qquad a \in \mathbb{F}_2^{2n}. \tag{2.18}$$

The output of the Bell-difference measurement has distribution q_ψ.

The symplectic character identity in Proposition 2.12 gives the following common duality.

Proposition 2.20 (Bell-difference duality). *For every pure state $|\psi\rangle$ and every subspace $A \subseteq \mathbb{F}_2^{2n}$,*

$$\sum_{a \in A} p_\psi(a) = \frac{|A|}{2^n} \sum_{x \in A^{\perp_s}} p_\psi(x). \tag{2.19}$$

$$\sum_{a \in A} q_\psi(a) = |A| \sum_{x \in A^{\perp_s}} p_\psi(x)^2. \tag{2.20}$$

These identities are stated in this form in [GIKL24, Theorems 3.1 and 3.2].

The most important consequence is that Bell-difference samples cannot leave the symplectic complement of the exact Pauli symmetries.

Proposition 2.21 (Support of Bell-difference sampling). *For every n-qubit pure state $|\psi\rangle$,*

$$\operatorname{supp}(q_\psi) \subseteq \operatorname{Weyl}(|\psi\rangle)^{\perp_s}.$$

Proof. Let $S = \operatorname{Weyl}(|\psi\rangle)$. For every $x \in S$, we have $p_\psi(x) = 2^{-n}$. Thus Eq. (2.20) and $|S||S^{\perp_s}| = 4^n$ give

$$\sum_{a \in S^{\perp_s}} q_\psi(a) = |S^{\perp_s}| \sum_{x \in S} p_\psi(x)^2 = \frac{|S^{\perp_s}||S|}{4^n} = 1.$$

Thus all of the q_ψ-mass lies in $S^{\perp_s}$. □

For later reductions, it is also useful to compare Bell-difference mass with characteristic mass.

Proposition 2.22 (Comparison of characteristic and Bell-difference mass). *For every pure state $|\psi\rangle$ and subspace $A \subseteq \mathbb{F}_2^{2n}$,*

$$\sum_{a \in A} q_\psi(a) \leq \sum_{a \in A} p_\psi(a).$$

Proof. Use $p_\psi(x) \leq 2^{-n}$ in Eq. (2.20), then apply Eq. (2.19). □

Lemma 2.23 (Span captures almost all sampled mass). *Let $\mathcal{D}$ be a distribution on $\mathbb{F}_2^N$ whose support lies in a subspace of dimension at most d. Let $\eta, \delta \in (0,1)$, and suppose that*

$$m \geq \frac{2\log(1/\delta) + 2d}{\eta}.$$

If A is the span of m independent samples from $\mathcal{D}$, then

$$\sum_{x \in A} \mathcal{D}(x) \geq 1 - \eta$$

with probability at least $1 - \delta$.

Proof. Let A_i be the span of the first i samples. Until A_i contains at least $1-\eta$ of the mass, call the next sample successful when it enlarges the span; once the target mass has been reached, call every remaining sample successful. Every conditional success probability is at least η. Failure after m samples implies that fewer than d successes occurred, so Lemma 2.8 gives

$$\Pr\left[\sum_{x\in A_m} \mathcal{D}(x) < 1-\eta\right] \le \exp\left(-\frac{m\eta}{2} + d\right) \le \delta.$$

□

Finally, Clifford changes of basis only relabel these distributions. From Eq. (2.15), if $|\psi'\rangle = C|\psi\rangle$, then

$$p_{\psi'}(C(x)) = p_\psi(x), \qquad q_{\psi'}(C(x)) = q_\psi(x). \tag{2.21}$$

Because Clifford changes of basis only relabel the distributions, an algorithm can transform a learned isotropic subspace to a standard Z-type subspace without changing the relevant probability masses.

Chapter 3: Nearly Time-Optimal Pure State Tomography

Quantum state tomography is the problem of learning an unknown quantum state from measurement outcomes on independent copies. In this chapter, we focus on 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., $1 - |\langle\hat{\psi}|\psi\rangle|^2 \leq \epsilon$ with probability at least $1 - \delta$.[1] We follow the conventions established in Sections 2.2 and 2.3.

If the learner can perform arbitrary measurements, the copy complexity of worst-case pure-state tomography is known to be $\Theta(2^n/\epsilon)$. This rate is achieved by several procedures and is information-theoretically optimal; see, e.g., [Hay98; HHJWY16; OW16; GKKT20; ACGN23; PSTW25; SSW25; PSW25]. Moreover, it is well known that the measurements do not need to be entangled *across copies* to attain this scaling [Vor13; KRT17; HHJWY16]. What remains unclear, however, is whether entanglement is required *within each n-qubit copy*. Product measurements avoid this within-copy entanglement and are substantially simpler to implement. This motivates the central question of this chapter:

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

An even more ambitious 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 [GKKT20], who achieved copy complexity $\widetilde{O}(3^n/\epsilon)$.

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

[1] For pure states, both trace distance and Frobenius distance are determined by the fidelity; see Equations (2.4) and (2.7).

matrix-valued estimate, and these matrices are averaged across samples.[2] By construction, this average equals the true state in expectation. However, the number of samples required for this estimator to concentrate is $\widetilde{O}(3^n/\epsilon)$. Moreover, their variance bound is shown to be tight even for pure product states. Consequently, any approach that follows this paradigm—namely, forming an average of (possibly reweighted) matrices derived from Pauli measurement outcomes and arguing concentration via the matrix Bernstein inequality—cannot asymptotically improve upon the 3^n dependence.

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 while using only Pauli-basis measurements.[3]

Theorem 3.1. *Let $n \geq 1$ and $\epsilon, \delta \in (0,1)$. There exists an algorithm that, given copies of an unknown n-qubit pure state $|\psi\rangle$, samples $\widetilde{O}(2^n \log(1/\delta)/\epsilon)$ Pauli product bases, measures one copy of $|\psi\rangle$ in each sampled basis, and outputs an estimate $|\hat{\psi}\rangle$ satisfying $|\langle\hat{\psi}|\psi\rangle|^2 \geq 1-\epsilon$ with probability at least $1-\delta$. The algorithm runs in time $\widetilde{O}(2^n \log(1/\delta)/\epsilon)$.*

In particular, the copy complexity matches, up to polylogarithmic factors, the $\Omega(2^n/\epsilon)$ lower bound that holds even when the learner is allowed *arbitrary* measurements [SSW25].

In addition to achieving near-optimal copy complexity, the algorithm also attains essentially optimal running time. To the best of our knowledge, this is the first algorithm for pure-state tomography with near-optimal running time, addressing a question explicitly raised in [GIKL25]. Before this result, the fastest algorithms required $\widetilde{O}(4^n/\epsilon)$ time [ACGN23], while, among algorithms restricted to Pauli-basis measurements, the best known running time was $\widetilde{O}(8^n/\epsilon)$ [GKKT20]. Thus, simple nonadaptive single-qubit

[2] They then apply a projected least-squares step to enforce positivity and unit trace; if one wishes to output a pure state, one may further post-process by taking the eigenvector corresponding to the largest eigenvalue. Neither of these steps has a substantial effect on the accuracy of the estimate.

[3] 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 $\Omega(d^2/\epsilon)$ samples are necessary for pure state tomography [FGLE12; LN22].

measurements suffice to achieve not only near-optimal statistical efficiency but also near-optimal computational efficiency.

3.1 An Aside: Mixed-State Tomography

Our pure-state tomography algorithm can also be combined with the random-purification reduction of Pelecanos, Spilecki, Tang, and Wright [PSTW25]. Their reduction coherently transforms copies of a rank-r, d-dimensional mixed state ρ into copies of a common random purification $|\rho_U\rangle \in \mathbb{C}^d \otimes \mathbb{C}^r$. Conditioned on the random choice of U, the output is therefore an ordinary pure-state tomography instance in dimension dr.

Applying the algorithm of this chapter to these purified copies gives a near-copy-optimal mixed-state tomography algorithm with a striking feature: all coherent processing across copies occurs inside the random-purification channel. Afterward, every copy is measured independently using only nonadaptive single-qubit Pauli-basis measurements.

Corollary 3.2 (Pauli tomography after random purification). *Let ρ be an n-qubit state of rank at most r, where $1 \le r \le d = 2^n$, and suppose that r is known. For $\epsilon, \delta \in (0,1)$, there is an algorithm using*

$$\widetilde{O}\left(\frac{rd\log(1/\delta)}{\epsilon}\right)$$

copies of ρ that outputs a density matrix $\hat{\rho}$ in factored form, namely as $AA^\dagger$ for $A \in \mathbb{C}^{d\times 2^{\lceil \log_2 r\rceil}}$, satisfying

$$F(\rho,\hat{\rho}) \ge 1-\epsilon$$

with probability at least $1-\delta$. The random-purification channel is the only stage that acts coherently across distinct input copies. After purification, every copy is measured independently using nonadaptive single-qubit Pauli-basis measurements, and the post-purification algorithm, including construction of this factored output, runs in time

$$\widetilde{O}\left(\frac{rd\log(1/\delta)}{\epsilon}\right).$$

Proof. Apply the random-purification channel to the input copies and condition on its random purification $|\rho_U\rangle$. After embedding the r-dimensional reference system into $\lceil \log_2 r\rceil$

qubits, the purified state has dimension at most $2dr$. Apply Theorem 3.1 with failure probability $\delta/2$. This gives the stated copy complexity, running time, and measurement structure.

Write the learned purification as $|\widehat{\rho_U}\rangle = \sum_{i,j} A_{ij}|i\rangle|j\rangle$. Tracing out the reference register gives $\hat{\rho} = AA^\dagger$, so returning A gives the stated factored representation without explicitly materializing all d^2 entries of $\hat{\rho}$. Monotonicity of fidelity under partial trace preserves the fidelity guarantee.

The ideal random-purification channel is exact. For an approximate implementation, choose diamond-distance error at most $\delta/2$; this changes the overall failure probability by at most $\delta/2$ and leaves the asymptotic resource bounds unchanged. □

For constant failure probability, the copy complexity is $\widetilde{O}(rd/\epsilon)$, which is optimal up to polylogarithmic factors. The point of this corollary is not an improved copy bound, but rather the measurement structure: after the initial collective purification step, the rest of the algorithm uses only single-qubit Pauli measurements.

3.2 Related Work

Pauli measurement tomography for mixed states. Direct analogues of this result appear in recent work by Acharya, Dharmavarapu, Liu, and Yu [ADLY25a; ADLY25b; Yu20] on mixed-state tomography using Pauli-basis measurements. These works show that, up to polylogarithmic factors, $\Theta(10^n/\epsilon)$ copies are necessary and sufficient for algorithms using nonadaptive Pauli measurements to perform tomography on general mixed states. Interestingly, while nonadaptive Pauli measurements cannot match the copy complexity achieved by general measurement schemes in the mixed-state setting, the result of this chapter shows that they can do so for pure-state tomography. As explained in Section 3.1, the direct-measurement restriction in these lower bounds is essential: coherent purification can first process the original copies, after which product Pauli measurements recover near-optimal scaling.

A separate line of work studies tomography from *Pauli observable measurements* through the lens of compressed sensing and low-rank matrix recovery, initiated by Gross,

Liu, Flammia, Becker, and Eisert [GLFBE10] and developed further in, e.g., [Liu11; FGLE12]. These works show that a rank-r state in dimension d can be uniquely reconstructed from $\widetilde{O}(rd)$ randomly chosen Pauli observables when their expectation values are known exactly. In the tomography setting, however, we only have noisy empirical estimates of these expectation values. Accounting for this statistical noise gives sample complexity bounds on the order of $\widetilde{O}_\epsilon(r^2d^2)$ for tomography. This matches, in its dependence on r and d, the corresponding lower bound of $\Omega(r^2d^2/\epsilon)$ samples for tomography from Pauli observable measurements [FGLE12; LN22].

Direct fidelity estimation. We present an algorithm that estimates the Frobenius distance $\|\rho - \sigma\|_{\mathrm{F}}$ between two (potentially mixed) n-qubit quantum states ρ and σ using nonadaptive Pauli measurements; this serves as a key subroutine in our pure-state tomography algorithm. A closely related problem was studied in the direct fidelity estimation (DFE) procedure of Flammia and Liu [FL11], which estimates fidelity to a known pure target state; in particular, it applies in the pure–pure regime relevant to our algorithm. Their procedure estimates this fidelity to within additive error $\pm\gamma'$, using $\widetilde{O}(2^n/\gamma'^2)$ copies of the unknown state in the worst case. When both states are pure, the fidelity is linearly related to the *squared* Frobenius distance, so estimating $\|\rho-\sigma\|_{\mathrm{F}}$ to additive error γ requires fidelity accuracy $O(\gamma^2)$ and hence $\widetilde{O}(2^n/\gamma^4)$ copies. While this guarantee suffices to obtain a tomography algorithm using $\widetilde{O}(2^n/\operatorname{poly}(\epsilon))$ copies, our stronger Frobenius-distance estimator is required to achieve the optimal dependence on ϵ.

Quantum state certification. Related to tomography is *quantum state certification*, in which one must determine whether an unknown state is close to a given hypothesis state. Recent work on certifying pure states using single-qubit measurements shows that adaptivity is necessary to avoid exponentially large copy complexity; see, for example, [HPS25; GHO25]. This contrasts with the tomography setting studied here, where nonadaptive single-qubit measurements are essentially as powerful as general measurements.

3.3 Technical Overview

The 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 σ, using only nonadaptive Pauli measurements that do not depend on σ. The second component is an implementation of this estimator.

3.3.1 Tomography via Frobenius Distance Estimation

Let $|\psi\rangle$ be a pure n-qubit state. Using $\widetilde{O}(2^n/\epsilon)$ copies, we aim to output $|\hat{\psi}\rangle$ with $|\langle\psi|\hat{\psi}\rangle|^2 \geq 1-\epsilon$. For any $x \in \{0,1\}^\ell$, let p_x be the probability of obtaining outcome x when measuring the first ℓ 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 ℓ, we maintain estimates $\{|\hat{\psi}_x\rangle \,:\, x \in \{0,1\}^\ell\}$. The leaves at depth n are the unique normalized zero-qubit state. Given estimates at depth $\ell+1$, we show how to build estimates at depth ℓ; the first gluing step therefore learns the one-qubit conditional states at depth $n-1$. Iterating up to the root yields $|\hat{\psi}_\emptyset\rangle \approx |\psi\rangle$.

We now describe how to perform this recursive estimation procedure. Fix $x \in \{0,1\}^\ell$. 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

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

By Lemma 3.3, there exist coefficients achieving an error (in Frobenius distance) comparable to the weighted errors of the two children, so it suffices to approximately solve this two-parameter optimization problem. The candidate states form a two-dimensional subspace. Four distance evaluations—at the two child branches and at two fixed superpositions—determine the associated 2×2 overlap matrix up to the error of the distance oracle. We take the top eigenvector of this matrix to obtain the gluing coefficients.

Each of these four evaluations reduces to estimating the distance between the unknown state $|\psi_x\rangle\langle\psi_x|$ and a known candidate σ. This is exactly where the Frobenius-distance estimator from Section 3.3.2 is used: it provides, for any fixed σ, an estimate of $\||\psi_x\rangle\langle\psi_x| - \sigma\|_{\mathrm{F}}$ from appropriate measurements on copies of $|\psi_x\rangle$.

The only subtlety is that the measurements on the last $n - \ell$ 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 $m = \widetilde{O}(2^n p/\epsilon)$ measurement settings (independent of x) that would let the Frobenius estimator compare $|\psi_x\rangle$ to any fixed σ, which is exactly what Section 3.3.2 provides; we then repeat each setting $\Theta((1/p)\,\mathrm{poly}(n))$ times, always measuring the first ℓ 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, so we obtain exactly the measurement data needed to run the distance estimator for that x. Finally, since the probabilities p_x vary with x, we run the above procedure over dyadic values of p down to the smallest relevant scale, which is $\Theta(\tau/2^n) = \widetilde{\Theta}(\epsilon/2^n)$. Prefixes below this scale may be assigned arbitrary estimates within the final error budget, while every remaining p_x lies within a constant factor of one of the chosen dyadic values.

3.3.2 Frobenius Distance Estimation

Suppose that ρ and σ are (potentially mixed) n-qubit states. We are given access to copies of ρ and to a full classical description of σ. Our goal is to estimate the quantity $\|\rho - \sigma\|_{\mathrm{F}}$ up to additive error γ, using a fixed (nonadaptive) list of measurements that does not depend on either ρ or σ, and using $\widetilde{O}(d/\gamma^2)$ samples.

In the description of our algorithm, we will assume access to samples of both ρ and σ, rather than a full description of σ. This is more general, because a classical description of σ allows us to simulate measurements on it. The first step is to note that $\|\rho - \sigma\|_{\mathrm{F}} = 2\sqrt{d}\sqrt{\mathbb{E}_P[v_P^2]}$, where v is a real vector of length 4^n, indexed by n-qubit Pauli matrices, defined by $v_P = \frac{1}{2}\,\mathrm{tr}(P(\rho - \sigma))$ (so that we always have $|v_P| \leq 1$).

By measuring ρ and σ 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 ρ and σ. Therefore, estimating the Frobenius distance using nonadaptive Pauli measurements reduces to the following problem: estimate $\sqrt{r} \coloneqq \sqrt{\mathbb{E}_k[v_k^2]}$ to additive error α 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 need the number of samples T to be $\widetilde{O}(1/\alpha^2)$.

To motivate the algorithm, let us imagine that all entries of v are either *small* (close to 0) or *big* (have absolute value close to 1). Write $r = r_{\text{small}} + r_{\text{big}}$, where r_{small} and r_{big} are the contributions to r from small and big indices, respectively. 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 estimator for $\sqrt{r}$ with additive α error. At a constant factor loss, it suffices to estimate $\hat{r}_{\text{small}}$ and $\hat{r}_{\text{big}}$ separately with guarantees

$$|\hat{r}_{\text{small}} - r_{\text{small}}| \lesssim \alpha\sqrt{r_{\text{small}}} + \alpha^2, \quad |\hat{r}_{\text{big}} - r_{\text{big}}| \lesssim \alpha\sqrt{r_{\text{big}}} + \alpha^2. \tag{3.1}$$

1. To estimate r_{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 get accurate estimates of their values, throwing out the values that are too large. We then output the average of the remaining values $v_{k_i}^2$. If r_{small} is nonnegligible, there must be many small indices, so sampling only T_{small} indices suffices to hit enough of them. On the other hand, many samples per index are needed to estimate their values accurately enough to satisfy (3.1).
2. To estimate r_{big}, we instead sample a large number T_{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 if they contribute significantly to r, but only a few samples per index suffice, since we need only a coarse estimate of v_{k_t} 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 not sum to one, leading to under- or over-counting in expectation. To fix both of these issues at once, before estimating v_{k_t}, we run a fixed process that uses the Rademacher samples to

classify the index as small or big, so that it only contributes to one of the two categories, and then independently compute the estimate of v_{k_t}.

For the general case, rather than splitting indices into only two categories, we partition them into $\log(1/\alpha)$ level sets, where the jth level consists roughly of indices k with $|v_k| \in [2^{-j}, 2^{-j+1}]$. It turns out that to achieve the desired approximation guarantee, we can take $T_j \approx \alpha^{-2}/4^j$ indices in level j, and take $m_j \approx 4^j$ samples of each index. Summing over all levels yields a total sample complexity of $\widetilde{O}(1/\alpha^2)$ as desired.

Time Efficiency

Implementing the above algorithm requires one to output Rademacher random variables with expectation $\mathrm{tr}(\sigma P)$ given a description of σ and P. Naively, this requires time 2^n when $\sigma = |\psi\rangle\langle\psi|$ and we have a classical description of $|\psi\rangle$ as a list of amplitudes. Since this must be repeated for roughly 2^n different Pauli matrices P, the resulting runtime is 4^n, whereas our target is $\widetilde{O}(2^n)$ after suppressing the dependence on ϵ and δ.

To perform this Rademacher sampling in near-linear time, we use the fact that Pauli matrices are phased permutation matrices in the computational basis. Draw $x \in \{0,1\}^n$ with probability $|\langle x|\psi\rangle|^2$. The Pauli string pairs x with a basis vector $x \oplus a$ and assigns a readily computable phase to this transition. From the two amplitudes on this pair, one can compute a number in $[-1, 1]$ whose average is $\langle\psi|P|\psi\rangle$, and then sample a Rademacher variable with that conditional mean. Proposition 3.11 gives the precise construction, including complex amplitudes.

To draw x with probability $|\langle x|\psi\rangle|^2$, we use the alias method [Wal77]. At the beginning of the distance estimator, we construct in time $\widetilde{O}(2^n)$ a data structure that produces such samples in time $O(1)$.

3.4 Recursive Tomography Algorithm

It will be useful to describe $|\psi\rangle$ as a binary tree. Set

$$|\psi_\varnothing\rangle := |\psi\rangle, \qquad p_\varnothing := 1.$$

For any string $x \in \{0,1\}^{<n}$ (identified by a node in the complete binary tree of depth n), we write

$$|\psi_x\rangle = \alpha_{x0}|0\rangle \otimes |\psi_{x0}\rangle + \alpha_{x1}|1\rangle \otimes |\psi_{x1}\rangle, \tag{3.2}$$

where $|\psi_{x0}\rangle$ and $|\psi_{x1}\rangle$ are normalized states on the remaining $n - \mathrm{len}(x) - 1$ qubits and $|\alpha_{x0}|^2 + |\alpha_{x1}|^2 = 1$. Recursively, we define the weight at each node x by

$$p_x \coloneqq \Pr[\text{measuring the first } \mathrm{len}(x) \text{ qubits gives } x].$$

When $p_x > 0$, the coefficients above satisfy $|\alpha_{xb}|^2 = p_{xb}/p_x$. When $p_x = 0$, we choose the normalized conditional state $|\psi_x\rangle$ arbitrarily. This description is not unique even when $p_x > 0$, since phases can be absorbed either into the α_{xb} or into the conditional states, but any consistent choice suffices for our analysis.

The main goal of our algorithm is to reconstruct $|\psi\rangle$ in a "bottom-up" fashion on this tree. The formal leaves at depth n are trivial zero-qubit states; we first learn the one-qubit conditional states $|\psi_x\rangle$ for $\mathrm{len}(x) = n - 1$, then the states at depth $n - 2$, and so on up to the root.

3.4.1 Gluing Branches and Error Accumulation

We first characterize the optimal way to combine two child estimates, and then turn this characterization into an efficient gluing procedure.

Optimal Gluings

The main subroutine we need is the following: given estimates $|\hat{\psi}_{x0}\rangle$ and $|\hat{\psi}_{x1}\rangle$, we wish to glue them together to form an estimate $|\hat{\psi}_x\rangle$. First, we show that if we have good estimates $|\hat{\psi}_{x0}\rangle$ and $|\hat{\psi}_{x1}\rangle$, then there exists a way to glue them together that does not worsen the error (but not necessarily that we can find it algorithmically).

Lemma 3.3. *Let $p_x > 0$. Assume that, for every $b \in \{0,1\}$ with $p_{xb} > 0$, the estimate $|\hat{\psi}_{xb}\rangle$ satisfies, for some $a_b \geq 0$,*

$$d_F(|\hat{\psi}_{xb}\rangle, |\psi_{xb}\rangle) \leq \sqrt{a_b}.$$

Then there exist $\hat{\alpha}_{x0}$ *and* $\hat{\alpha}_{x1}$ *with* $|\hat{\alpha}_{x0}|^2 + |\hat{\alpha}_{x1}|^2 = 1$ *such that*

$$d_F(\hat{\alpha}_{x0}|0\rangle \otimes |\hat{\psi}_{x0}\rangle + \hat{\alpha}_{x1}|1\rangle \otimes |\hat{\psi}_{x1}\rangle, |\psi_x\rangle) \le \sqrt{\sum_{b\,:\,p_{xb}>0} \frac{p_{xb}}{p_x}\, a_b}.$$

Proof. Let $w_b := p_{xb}/p_x$ for $b \in \{0,1\}$ (so $w_0 + w_1 = 1$), and write

$$|\psi_x\rangle = \sqrt{w_0}|0\rangle \otimes |\psi_{x0}\rangle + \sqrt{w_1}|1\rangle \otimes |\psi_{x1}\rangle$$

(absorbing any relative phase into $|\psi_{xb}\rangle$). For b with $w_b > 0$, set $c_b := \langle\psi_{xb}|\hat{\psi}_{xb}\rangle$, and set $c_b = 0$ when $w_b = 0$. For pure states and every b with $w_b > 0$,

$$d_F(|\phi\rangle, |\psi\rangle)^2 = 2\big(1 - |\langle\phi|\psi\rangle|^2\big),$$

and hence

$$1 - |c_b|^2 = \frac{1}{2} d_F(|\hat{\psi}_{xb}\rangle, |\psi_{xb}\rangle)^2 \le \frac{a_b}{2}.$$

Let $s := \sum_{b\,:\,w_b>0} w_b |c_b|^2$. If $s = 0$, every positive-weight branch has squared Frobenius error 2, so the right-hand side of the claimed bound is at least $\sqrt{2}$ and any normalized gluing suffices. Otherwise define

$$\hat{\alpha}_{xb} := \begin{cases} \sqrt{w_b}\, c_b^* / \sqrt{s}, & w_b > 0, \\ 0, & w_b = 0, \end{cases} \qquad (b \in \{0,1\}),$$

so $|\hat{\alpha}_{x0}|^2 + |\hat{\alpha}_{x1}|^2 = 1$. Let

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

Using $\langle 0|1\rangle = 0$,

$$\langle\psi_x|\hat{\psi}_x\rangle = \sqrt{w_0}\hat{\alpha}_{x0} c_0 + \sqrt{w_1}\hat{\alpha}_{x1} c_1 = \sqrt{s},$$

so

$$d_F(|\hat{\psi}_x\rangle, |\psi_x\rangle)^2 = 2(1 - |\langle\psi_x|\hat{\psi}_x\rangle|^2) = 2(1 - s) = 2 \sum_{b\,:\,w_b>0} w_b(1 - |c_b|^2) \le \sum_{b\,:\,w_b>0} w_b a_b.$$

Taking square roots gives the stated bound. □

Algorithmic Gluing

Next, we design an algorithm to approximate the coefficients $\hat{\alpha}_{x0}, \hat{\alpha}_{x1}$ guaranteed by Lemma 3.3. This is the subroutine that calls our Frobenius distance estimator. Fix a working accuracy $\tau > 0$. For $x \in \{0,1\}^\ell$, define

$$\epsilon_x := \frac{\tau}{2^\ell p_x} \quad \text{if } p_x > 0, \tag{3.3}$$

and set $\epsilon_x = +\infty$ if $p_x = 0$.

The following perturbation bound reduces this optimization to four distance evaluations.

Lemma 3.4. *Let $|e_0\rangle$ and $|e_1\rangle$ be orthonormal states, and let $|\phi\rangle$ be a normalized state. For a unit vector $u = (u_0, u_1) \in \mathbb{C}^2$, write*

$$|\chi_u\rangle := u_0|e_0\rangle + u_1|e_1\rangle, \qquad \Delta(u) := d_F(|\chi_u\rangle, |\phi\rangle).$$

Define

$$|e_+\rangle := \frac{|e_0\rangle + |e_1\rangle}{\sqrt{2}}, \qquad |e_{\mathrm{i}}\rangle := \frac{|e_0\rangle - \mathrm{i}|e_1\rangle}{\sqrt{2}}.$$

There is an absolute constant C_{g} with the following property. Suppose that $0 < \eta \le 10^{-2}$ and that we are given additive η-estimates of the four distances from $|\phi\rangle$ to $|e_0\rangle, |e_1\rangle, |e_+\rangle, |e_{\mathrm{i}}\rangle$. From these estimates one can compute a unit vector $\hat{u} \in \mathbb{C}^2$ satisfying

$$\Delta(\hat{u}) \le \min_{\|u\|_2=1} \Delta(u) + C_{\mathrm{g}}\eta.$$

Proof. Let $z_b := \langle e_b|\phi\rangle$ and set $A := zz^\dagger \in \mathbb{C}^{2\times 2}$. For every unit vector u,

$$|\langle \chi_u|\phi\rangle|^2 = u^\dagger A u.$$

Let $D_0, D_1, D_+, D_{\mathrm{i}}$ be the four exact distances, and let $F_j := 1 - D_j^2/2$ be the corresponding fidelities. Then

$$F_0 = A_{00}, \qquad F_1 = A_{11},$$

and

$$\mathrm{Re}(A_{01}) = F_+ - \frac{F_0 + F_1}{2}, \qquad \mathrm{Im}(A_{01}) = F_{\mathrm{i}} - \frac{F_0 + F_1}{2}.$$

Given the estimated distances $\widetilde{D}_j$, define $\widetilde{F}_j \coloneqq 1 - \widetilde{D}_j^2/2$ and construct the Hermitian matrix $\widetilde{A}$ from the same formulas. Since $D_j \le \sqrt{2}$ and $|\widetilde{D}_j - D_j| \le \eta$,

$$|\widetilde{F}_j - F_j| = \frac{1}{2}|\widetilde{D}_j - D_j|\,|\widetilde{D}_j + D_j| \le \frac{\eta}{2}(2\sqrt{2} + \eta) \le 2\eta.$$

Hence the diagonal entries of $\widetilde{A} - A$ have magnitude at most 2η, while the real and imaginary parts of its off-diagonal entry have magnitude at most 4η. Therefore

$$\|\widetilde{A} - A\|_{\mathrm{op}} \le \|\widetilde{A} - A\|_{\mathrm{F}} \le 9\eta.$$

Put $\kappa \coloneqq 9\eta$ and let $\hat{u}$ be a unit eigenvector of $\widetilde{A}$ corresponding to its largest eigenvalue.

Write $s \coloneqq \|z\|_2$. If $s = 0$, then every unit vector is optimal, so assume $s > 0$ and put $u_\star \coloneqq z/s$. Thus $A = s^2 u_\star u_\star^\dagger$, and the minimum distance is

$$\Delta_\star = \sqrt{2(1 - s^2)}.$$

Let $E \coloneqq \widetilde{A} - A$. By the variational definition of $\hat{u}$,

$$\hat{u}^\dagger A \hat{u} \ge s^2 - 2\kappa.$$

Consequently, if

$$r \coloneqq s^2 - \hat{u}^\dagger A \hat{u},$$

then $0 \le r \le 2\kappa$ and

$$\Delta(\hat{u})^2 = \Delta_\star^2 + 2r.$$

If $\Delta_\star \ge 1/2$, then

$$\Delta(\hat{u}) - \Delta_\star = \frac{2r}{\Delta(\hat{u}) + \Delta_\star} \le 8\kappa.$$

Suppose instead that $\Delta_\star < 1/2$. Then $s^2 > 7/8$. Let $v \in \mathbb{C}^2$ be a unit vector orthogonal to $u_\star$, and let $\hat{\lambda}$ be the largest eigenvalue of $\widetilde{A}$. Since $\kappa < 1/8$,

$$\hat{\lambda} \ge u_\star^\dagger \widetilde{A} u_\star \ge s^2 - \kappa \ge \frac{3}{4}.$$

Projecting the eigenvalue equation $\widetilde{A}\hat{u} = \hat{\lambda}\hat{u}$ onto v gives

$$\hat{\lambda}\langle v, \hat{u}\rangle = \langle v, E\hat{u}\rangle,$$

and therefore $|\langle v, \hat{u}\rangle| \le 4\kappa/3$. It follows that $r = s^2|\langle v, \hat{u}\rangle|^2 \le 2\kappa^2$. Therefore

$$\Delta(\hat{u}) - \Delta_\star \le \sqrt{2r} \le 2\kappa.$$

Taking, for example, $C_{\mathrm{g}} = 72$ proves the claim. □

Lemma 3.5. *Let $p_x > 0$. There exists a classical procedure* FIND-COEFFS *which takes the two child estimates, conditional measurement data, an accuracy ϵ^*, and a failure probability δ_0. The data are generated by a nonadaptive Pauli measurement scheme on copies of $|\psi_x\rangle$ that depends only on $n - \mathrm{len}(x)$, ϵ^*, and δ_0, not on the child estimates. The procedure always outputs $\hat{\alpha}_{x0}, \hat{\alpha}_{x1} \in \mathbb{C}$ satisfying $|\hat{\alpha}_{x0}|^2 + |\hat{\alpha}_{x1}|^2 = 1$.*

Suppose that $\epsilon^* \in (0, 1]$, $\epsilon^* \le 0.1\epsilon_x$, and, for each b with $p_{xb} > 0$,

$$d_F(|\hat{\psi}_{xb}\rangle, |\psi_{xb}\rangle) \le (n - \mathrm{len}(x))\sqrt{\epsilon_{xb}}.$$

Then, with probability at least $1 - \delta_0$, the output satisfies

$$d_F\Big(\hat{\alpha}_{x0}|0\rangle \otimes |\hat{\psi}_{x0}\rangle + \hat{\alpha}_{x1}|1\rangle \otimes |\hat{\psi}_{x1}\rangle, |\psi_x\rangle\Big) \le (n - \mathrm{len}(x) + 1)\sqrt{\epsilon_x}.$$

The procedure consumes $\widetilde{O}(2^{n-\mathrm{len}(x)} \log(1/\delta_0)/\epsilon^*)$ copies of $|\psi_x\rangle$ and has matching running time.

Proof. Let $m := n - \mathrm{len}(x)$, and define the orthonormal states

$$|e_0\rangle := |0\rangle \otimes |\hat{\psi}_{x0}\rangle, \qquad |e_1\rangle := |1\rangle \otimes |\hat{\psi}_{x1}\rangle.$$

Let

$$\Delta_\star := \min_{|u_0|^2+|u_1|^2=1} d_F(u_0|e_0\rangle + u_1|e_1\rangle, |\psi_x\rangle).$$

By Lemma 3.3 and the hypotheses of this lemma,

$$\Delta_\star^2 \le m^2 \sum_{b\,:\,p_{xb}>0} \frac{p_{xb}}{p_x}\epsilon_{xb} = m^2 \sum_{b\,:\,p_{xb}>0} \frac{\tau}{2^{\mathrm{len}(x)+1} p_x} \le m^2\epsilon_x. \tag{3.4}$$

Fix

$$\eta := c\sqrt{\epsilon^*},$$

where $c > 0$ is an absolute constant chosen so that $c \le 10^{-2}$ and $C_g c\sqrt{0.1} \le 1$. Apply Theorem 3.10 to estimate the distances from $|\psi_x\rangle$ to the four known states

$$|e_0\rangle, \qquad |e_1\rangle, \qquad \frac{|e_0\rangle + |e_1\rangle}{\sqrt{2}}, \qquad \frac{|e_0\rangle - \mathrm{i}|e_1\rangle}{\sqrt{2}},$$

each to additive error η and with failure probability at most $\delta_0/4$. The physical Pauli queries for these four calls depend only on m, η, and δ_0, so they may be fixed before the child estimates are known. Conditioning on the child estimates, the four candidate states are fixed and the corresponding query blocks remain independent, so the distance-estimation guarantees apply. The procedure constructs the corresponding Hermitian matrix and returns a unit top eigenvector regardless of the estimated values, so its output is always normalized.

On the event that all four calls succeed, which has probability at least $1 - \delta_0$, Lemma 3.4 returns a unit vector $\hat{u} = (\hat{\alpha}_{x0}, \hat{\alpha}_{x1})$ satisfying

$$d_F(\hat{\alpha}_{x0}|e_0\rangle + \hat{\alpha}_{x1}|e_1\rangle, |\psi_x\rangle) \le \Delta_\star + C_g\eta.$$

Since $\epsilon^* \le 0.1\epsilon_x$, our choice of c gives $C_g\eta \le \sqrt{\epsilon_x}$. Thus (3.4) gives

$$d_F(\hat{\alpha}_{x0}|e_0\rangle + \hat{\alpha}_{x1}|e_1\rangle, |\psi_x\rangle) \le (m+1)\sqrt{\epsilon_x}.$$

Each distance estimate uses

$$\widetilde{O}\left(\frac{2^m \log(1/\delta_0)}{\eta^2}\right) = \widetilde{O}\left(\frac{2^m \log(1/\delta_0)}{\epsilon^*}\right)$$

copies of $|\psi_x\rangle$ and matching time. There are four calls, so the stated resource bounds follow. □

3.4.2 The Overall Algorithm

We now make the postselection and scale selection explicit. For a target infidelity ϵ and failure probability δ, set

$$\tau := \frac{2\epsilon}{(n+1)^2}, \qquad \delta_0 := \frac{\delta}{4(2^n - 1)}, \qquad L := C_0\big(n + \log(2/\tau) + \log(4/\delta)\big),$$

where C_0 is a sufficiently large universal constant. At depth ℓ, use the level-dependent grid

$$\mathcal{E}_\ell := \left\{ \frac{2^j \tau}{128\, 2^\ell} \;:\; j \geq 0,\ \frac{2^j \tau}{128\, 2^\ell} \leq 1 \right\} \cup \{1\}.$$

The physical Pauli queries used by the Frobenius estimator depend on the dimension, accuracy, and failure probability, but not on the classically described candidate state. We may therefore sample in advance enough query blocks for the four distance-oracle calls made by Find-Coeffs, independently across depths, scales, and blocks. Let $\mathcal{Q}_{\ell,a}$ denote the resulting list for depth ℓ and accuracy $a \in \mathcal{E}_\ell$. Its elements are *tagged query slots*: two occurrences of the same Pauli operator remain distinct and receive independent data. We have

$$|\mathcal{Q}_{\ell,a}| = \widetilde{O}\left(\frac{2^{n-\ell} \log(1/\delta_0)}{a} \right). \tag{3.5}$$

Algorithm 1 Build-Measurement-Set(n, τ, δ)

Input: Number of qubits n, working accuracy τ, and failure probability δ.
Output: Measurement multiset $\mathcal{M}$.

$\mathcal{M} \leftarrow \emptyset$.
for $\ell = 0, \ldots, n-1$ **do**
  for $a \in \mathcal{E}_\ell$ **do**
    Sample the tagged list $\mathcal{Q}_{\ell,a}$ described above.
    $R_{\ell,a} \leftarrow \lceil C\, 2^\ell (a/\tau) L \rceil$.
    for each tagged slot $q \in \mathcal{Q}_{\ell,a}$, with suffix Pauli P_q **do**
      Add $R_{\ell,a}$ copies of the setting (ℓ, a, q, P_q) to $\mathcal{M}$.
    end for
  end for
end for
return $\mathcal{M}$.

For a setting (ℓ, a, q, P_q), we measure the first ℓ qubits in the computational basis and the remaining qubits in a product basis diagonalizing P_q; on identity factors, we use the computational basis. We retain the observed prefix, the Pauli outcome, and the slot tag. All settings are fixed before any outcome is observed, and each is a Pauli product basis.

Algorithm 2 TOMOGRAPHY-FROM-MEASUREMENTS$(\tau, \delta, \mathcal{M})$

Input: Outcomes from $\mathcal{M} =$ BUILD-MEASUREMENT-SET(n, τ, δ).
Output: Estimate $|\hat{\psi}\rangle$.

for $x \in \{0,1\}^n$ **do**
  Initialize $|\hat{\psi}_x\rangle$ to the unique normalized zero-qubit state.
end for
for $\ell = n-1, \dots, 0$ **do**
  for $x \in \{0,1\}^\ell$ **do**
    built $\leftarrow$ FALSE.
    for $a \in \mathcal{E}_\ell$ in increasing order **do**
      if every $q \in \mathcal{Q}_{\ell,a}$ has an outcome with prefix x **then**
        Use the first such outcome from each tagged block as the conditional data.
        $(\hat{\alpha}_{x0}, \hat{\alpha}_{x1}) \leftarrow$ FIND-COEFFS$(|\hat{\psi}_{x0}\rangle, |\hat{\psi}_{x1}\rangle,$ conditional data$, a, \delta_0)$.
        $|\hat{\psi}_x\rangle \leftarrow \hat{\alpha}_{x0}|0\rangle|\hat{\psi}_{x0}\rangle + \hat{\alpha}_{x1}|1\rangle|\hat{\psi}_{x1}\rangle$; *built* $\leftarrow$ TRUE; **break**.
      end if
    end for
    if *built* = FALSE **then**
      $|\hat{\psi}_x\rangle \leftarrow |0\rangle \otimes |\hat{\psi}_{x0}\rangle$.
    end if
  end for
end for
return $|\hat{\psi}_\varnothing\rangle$.

Theorem 3.6. *Let $n \geq 1$ and $\epsilon, \delta \in (0,1)$. There exists an algorithm* TOMOGRAPHY(ϵ, δ) *that, given copies of an unknown n-qubit pure state $|\psi\rangle$, samples $\widetilde{O}(2^n \log(1/\delta)/\epsilon)$ Pauli product bases, measures one copy of $|\psi\rangle$ in each sampled basis, and outputs an estimate $|\hat{\psi}\rangle$ satisfying $|\langle\hat{\psi}|\psi\rangle|^2 \geq 1-\epsilon$ with probability at least $1-\delta$. The algorithm runs in time $\widetilde{O}(2^n \log(1/\delta)/\epsilon)$.*

Proof. We first establish a simultaneous availability event. Fix a node $x \in \{0,1\}^\ell$ with $p_x > 0$ and $\epsilon_x < 2$, and let a_x be the largest point of $\mathcal{E}_\ell$ satisfying $a_x \leq 0.01\epsilon_x$. The smallest grid point is at most $\epsilon_x/128$, so a_x exists; because the grid is dyadic and $\epsilon_x < 2$, it also satisfies

$$0.005\epsilon_x < a_x \leq 0.01\epsilon_x. \tag{3.6}$$

For a tagged slot $q \in \mathcal{Q}_{\ell,a_x}$, the expected number of copies in its block whose prefix is x is

$$p_x R_{\ell,a_x} \geq C\frac{a_x}{\epsilon_x}L \geq 0.005CL.$$

Hence the probability that this block contains no such copy is at most $\exp(-0.005CL)$. The logarithm of the total number of relevant nodes, scales, and tagged slots is $O(n+\log(2/\tau)+\log(4/\delta))$. A union bound, with C and C_0 sufficiently large, shows that with probability at least $1-\delta/4$, every slot required at every scale a_x is populated.

Conditioned on a slot being populated, the first outcome in that block having prefix x is distributed exactly as the measurement of P_q on $|\psi_x\rangle$. Availability depends only on the prefix, so this conditioning does not bias the retained suffix outcome. Distinct tags use disjoint blocks and therefore give independent conditional outcomes. Moreover, correctness is required from FIND-COEFFS at no more than 2^n-1 positive-probability nodes with $\epsilon_x < 2$. The total failure probability of these calls is at most $(2^n-1)\delta_0 \le \delta/4$. We work on the intersection of these two good events, whose probability is at least $1-\delta/2$, which is stronger than required.

We prove by downward induction that, for every positive-probability node x at depth ℓ,

$$d_F(|\psi_x\rangle, |\hat{\psi}_x\rangle) \le (n-\ell+1)\sqrt{\epsilon_x}. \tag{3.7}$$

At depth n, the true and estimated conditional states are both the unique zero-qubit state. Now take an internal node and put $m = n-\ell$. If $\epsilon_x \ge 2$, even an arbitrary normalized estimate satisfies (3.7), since $d_F \le \sqrt{2} \le (m+1)\sqrt{\epsilon_x}$. If $\epsilon_x < 2$, the scale a_x is complete on the availability event. The reconstruction uses the first complete scale, which is no larger than a_x and hence is at most $0.01\epsilon_x$. For every positive-probability child, the induction hypothesis gives

$$d_F(|\psi_{xb}\rangle, |\hat{\psi}_{xb}\rangle) \le m\sqrt{\epsilon_{xb}}.$$

Thus Lemma 3.5 applies and proves the invariant at x.

At the root, $\epsilon_\emptyset = \tau$, so

$$d_F(|\psi\rangle, |\hat{\psi}\rangle) \le (n+1)\sqrt{\tau}.$$

The choice $\tau = 2\epsilon/(n+1)^2$ therefore gives

$$1 - |\langle\hat{\psi}|\psi\rangle|^2 = \frac{1}{2} d_F(|\psi\rangle, |\hat{\psi}\rangle)^2 \le \epsilon.$$

It remains to verify the resources. By (3.5), a fixed depth–scale pair contributes

$$|\mathcal{Q}_{\ell,a}|R_{\ell,a} = \widetilde{O}\left(\frac{2^{n-\ell}}{a} \cdot \frac{2^{\ell} aL}{\tau}\right) = \widetilde{O}(2^n/\tau)$$

measurements. There are only n levels and logarithmically many scales, so the full measurement count is $\widetilde{O}(2^n/\tau)$.

The classical reconstruction has the same order of growth. Indeed, at a fixed depth and scale, every node that executes FIND-COEFFS must appear as a prefix in the block for any one fixed tagged slot. Thus at most $R_{\ell,a}$ nodes execute the procedure at that pair, and their total work is bounded by $\widetilde{O}(|\mathcal{Q}_{\ell,a}|R_{\ell,a})$; indexing and scanning all nodes adds only $\widetilde{O}(2^n)$ work. Finally, $1/\tau = (n+1)^2/(2\epsilon)$, whose polynomial factor is absorbed by the soft-O notation. This proves both claimed bounds. □

3.5 Frobenius Distance Estimation

We present our algorithm for estimating the Frobenius distance between two (potentially mixed) quantum states using nonadaptive Pauli observable measurements.

Definition 3.7 (Nonadaptive Pauli measurement scheme). A *nonadaptive Pauli measurement scheme* with M measurements is a (possibly randomized) procedure that chooses Pauli observables

$$P_1, \ldots, P_M \in \{I, X, Y, Z\}^{\otimes n}$$

before any measurement is performed. Given an n-qubit quantum state ρ, for each $i \in \{1, \ldots, M\}$, the scheme measures P_i on an independent copy of ρ and records an outcome $X_i \in \{-1, +1\}$ satisfying

$$\mathbb{E}[X_i] = \operatorname{tr}(\rho P_i).$$

Theorem 3.8 (Frobenius distance estimation). *Let ρ and σ be quantum states on n qubits, let $d = 2^n$, and let $\gamma, \delta \in (0,1)$. There exists a nonadaptive Pauli measurement scheme (depending only on d, γ, and δ) using $\widetilde{O}(d \log(1/\delta)/\gamma^2)$ measurements on independent copies of ρ and σ that outputs an estimate $\hat{D} \in \mathbb{R}$ satisfying*

$$\left|\hat{D} - \|\rho - \sigma\|_{\mathrm{F}}\right| \leq \gamma$$

with failure probability at most $(\gamma/d)^{10}\delta$.

The main ingredient is the following classical norm-estimation theorem.

Theorem 3.9 (Rademacher norm estimation). *Let* $v = (v_1, \dots, v_N) \in [-1, 1]^N$ *be an unknown vector. Let* $\alpha \in (0, 1)$. *Suppose that we may (nonadaptively) pick queries* $k_1, \dots, k_M \in \{1, \dots, N\}$ *(possibly with repetitions). For each* $j \in \{1, \dots, M\}$, *we then receive a sample of a Rademacher random variable* $X_j \in \{-1, +1\}$ *with* $\mathbb{E}[X_j] = v_{k_j}$. *Then it is possible to make* $M = \widetilde{O}(1/\alpha^2)$ *nonadaptive queries and output a nonnegative estimate* $\hat{q}$ *such that*

$$\left| \hat{q} - \sqrt{\mathop{\mathbb{E}}_{k \sim [N]} [v_k^2]} \right| \leq \alpha,$$

with probability at least $2/3$.

We first reduce the main theorem to Theorem 3.9.

Reduction of Theorem 3.8 to Theorem 3.9. Write $\Delta = \rho - \sigma$. Expanding in the Pauli basis, $\Delta = d^{-1} \sum_P \Delta_P P$ with $\Delta_P = \mathrm{tr}(\Delta P)$, and orthogonality of the Paulis implies

$$\|\Delta\|_{\mathrm{F}}^2 = d^{-1} \sum_P \Delta_P^2.$$

Define $v_P = \frac{1}{2}\Delta_P$. Since $|\mathrm{tr}(\rho P)| \leq 1$ and likewise for σ, we have $|v_P| \leq 1$. If we let the expectation $\mathbb{E}_P[\cdot]$ be over a uniformly random Pauli label P, then

$$\|\rho - \sigma\|_{\mathrm{F}}^2 = d^{-1} \sum_P \Delta_P^2 = d^{-1} \sum_P (2v_P)^2 = 4d \mathop{\mathbb{E}}_P [v_P^2],$$

so

$$\|\rho - \sigma\|_{\mathrm{F}} = 2\sqrt{d} \sqrt{\mathop{\mathbb{E}}_P [v_P^2]}.$$

Note that, for each P, we may obtain a Rademacher (± 1) random variable with expectation v_P using one sample each of ρ and σ, as follows. Measure P on ρ and σ to obtain X and Y, respectively. Then return the random variable Z, which is X or $-Y$ with probability $1/2$ each. Note that $\mathbb{E}[Z] = \frac{1}{2}(\mathrm{tr}(\rho P) - \mathrm{tr}(\sigma P)) = v_P$, as desired. Thus, we may indeed obtain Rademacher queries as required in Theorem 3.9, using $O(1)$ nonadaptive Pauli measurements per query.

Thus, by Theorem 3.9, we may make $M = \widetilde{O}(1/\alpha^2)$ nonadaptive measurements to obtain $\hat{q}$ such that

$$\left|\hat{q} - \sqrt{\mathop{\mathbb{E}}_{P}[\nu_P^2]}\right| \leq \alpha$$

with probability at least $2/3$. Define $\hat{D} = 2\sqrt{d}\hat{q}$. Then

$$\left|\hat{D} - \|\rho - \sigma\|_{\mathrm{F}}\right| = 2\sqrt{d}\left|\hat{q} - \sqrt{\mathop{\mathbb{E}}_{P}[\nu_P^2]}\right| \leq 2\sqrt{d}\alpha.$$

Choosing $\alpha = \gamma/(2\sqrt{d})$ yields $|\hat{D} - \|\rho - \sigma\|_{\mathrm{F}}| \leq \gamma$ with probability at least $2/3$.

Each classical query uses $O(1)$ Pauli measurements on independent copies of ρ and σ, so the total number of Pauli measurements is

$$M = \widetilde{O}\big(1/\alpha^2\big) = \widetilde{O}\big(d/\gamma^2\big).$$

Standard repetition and taking the median amplifies the success probability to at least $1 - (\gamma/d)^{10}\delta$ at the cost of only polylogarithmic factors in d and $1/\gamma$ and $1/\delta$. □

3.5.1 Time Efficiency

For our tomography algorithm, we only use Theorem 3.8 when $\sigma = |\psi\rangle\langle\psi|$ is a known pure state. In this case, we can improve the time efficiency of our Frobenius distance estimator:

Theorem 3.10. *Set $d = 2^n$, and let $\gamma, \delta \in (0,1)$. Let ρ be a mixed quantum state on n qubits. Let $|\psi\rangle$ be a pure quantum state on n qubits. There exists an algorithm that takes as input a description of $|\psi\rangle$ as a list of 2^n amplitudes in the computational basis and the output of a nonadaptive Pauli measurement scheme (depending only on d, γ, and δ) using $\widetilde{O}(d\log(1/\delta)/\gamma^2)$ measurements on independent copies of ρ and outputs an estimate $\hat{D} \in \mathbb{R}$ satisfying*

$$\left|\hat{D} - \|\rho - |\psi\rangle\langle\psi|\|_{\mathrm{F}}\right| \leq \gamma$$

with failure probability at most $(\gamma/d)^{10}\delta$. The algorithm runs in time $\widetilde{O}(d\log(1/\delta)/\gamma^2)$.

Proof. The sampling stage described below and Algorithm 3 both run in the desired runtime, so it suffices to execute the reduction of Theorem 3.8 to Theorem 3.9 efficiently.

The reduction makes $\widetilde{O}(d\log(1/\delta)/\gamma^2)$ Rademacher queries in total. For every sampled Pauli P, one draw consists of sampling the corresponding variables X and Y. Sampling X is part of the physical measurement of ρ and takes $\mathrm{poly}(n) = \widetilde{O}(1)$ time to process. Proposition 3.11 shows how to sample Y, with $\mathbb{E}[Y] = \langle\psi|P|\psi\rangle$, in $O(n)$ time after one near-linear preprocessing step. □

Proposition 3.11. *Let $|\psi\rangle$ be an n-qubit pure state and let $d = 2^n$. There is a data structure using $O(d)$ space such that, given a Hermitian Pauli observable $P \in \{I, X, Y, Z\}^{\otimes n}$, it outputs in time $O(\log d)$ a Rademacher random variable with expectation $\langle\psi|P|\psi\rangle$. Given the computational-basis amplitudes of $|\psi\rangle$, the data structure can be constructed in time $\widetilde{O}(d)$.*

Proof. Using Theorem 3.12, we build a data structure with the desired space bound and $\widetilde{O}(d)$ setup time. Each query returns in $O(1)$ time an independent sample from the distribution $\mathcal{D}_{|\psi\rangle}$ on $\{0,1\}^n$ given by $\Pr[x] = |\langle x|\psi\rangle|^2$. The data structure is based on the alias method:

Theorem 3.12 ([Wal77]). *Let p be a probability distribution on d elements. There exists a data structure and algorithm that takes $O(d)$ space and time $O(1)$ to answer a query with an independent sample from p. Setting up the data structure takes time $\widetilde{O}(d)$ given access to the list of probabilities of p.*

We now describe how to turn an independent sample from $\mathcal{D}_{|\psi\rangle}$ and a Pauli P into a Rademacher random variable with expectation $\langle\psi|P|\psi\rangle$. Write $P = \bigotimes_{j=1}^{n} P_j$, where $P_j \in \{I, X, Y, Z\}$.

Every Pauli string acts on computational basis vectors as a phased permutation. Concretely, there exists a bit string $a \in \{0,1\}^n$ and a phase function $\eta : \{0,1\}^n \to \{\pm 1, \pm \mathrm{i}\}$ such that for every $x \in \{0,1\}^n$,

$$P|x\rangle = \eta(x)|x \oplus a\rangle.$$

Here $a_j = 1$ exactly when $P_j \in \{X, Y\}$, and $a_j = 0$ exactly when $P_j \in \{I, Z\}$. Since P is Hermitian, $\eta(x \oplus a) = \overline{\eta(x)}$. Thus

$$\langle\psi|P|\psi\rangle = \sum_{x\in\{0,1\}^n} \overline{\psi_{x\oplus a}}\eta(x)\psi_x.$$

We split into two cases:

Case 1: $a = 0$. In this case P is diagonal in the computational basis, so $\eta(x) \in \{\pm 1\}$ for all x, and

$$P|x\rangle = \eta(x)|x\rangle.$$

The algorithm is simply:

1. Draw x from the oracle distribution $\Pr[x] = |\psi_x|^2$.
2. Output $b := \eta(x) \in \{\pm 1\}$.

Then

$$\mathbb{E}[b] = \sum_x |\psi_x|^2 \eta(x) = \sum_x \overline{\psi_x}\eta(x)\psi_x = \langle\psi|P|\psi\rangle.$$

Case 2: $a \neq 0$. Now the map $x \mapsto x \oplus a$ has no fixed points, so $\{0,1\}^n$ is partitioned into disjoint pairs

$$\{x, x \oplus a\}.$$

Choose a canonical representative r for each pair, for example the lexicographically smaller of the two strings. Then

$$\langle\psi|P|\psi\rangle = \sum_r \Big(\overline{\psi_{r\oplus a}}\eta(r)\psi_r + \overline{\psi_r}\eta(r\oplus a)\psi_{r\oplus a}\Big),$$

where the sum runs over all representatives r. For each representative r, define

$$m_r := |\psi_r|^2 + |\psi_{r\oplus a}|^2$$

and

$$t_r := \overline{\psi_{r\oplus a}}\eta(r)\psi_r + \overline{\psi_r}\eta(r\oplus a)\psi_{r\oplus a} = 2\,\mathrm{Re}\Big(\overline{\psi_{r\oplus a}}\eta(r)\psi_r\Big).$$

Since P is Hermitian, the quantity t_r is real. Also,

$$|t_r| \le 2|\psi_r||\psi_{r\oplus a}| \le |\psi_r|^2 + |\psi_{r\oplus a}|^2 = m_r,$$

and hence whenever $m_r > 0$,

$$y_r := \frac{t_r}{m_r} \in [-1, 1].$$

We now describe the algorithm.

1. Draw x from $\mathcal{D}_{|\psi\rangle}$.
2. Compute $x \oplus a$, and let
$$r := \min_{\text{lex}}\{x, x \oplus a\}.$$
3. Look up the two amplitudes ψ_r and $\psi_{r\oplus a}$.
4. Compute $m_r = |\psi_r|^2 + |\psi_{r\oplus a}|^2$ and $t_r = \overline{\psi_{r\oplus a}}\eta(r)\psi_r + \overline{\psi_r}\eta(r \oplus a)\psi_{r\oplus a}$. If $m_r = 0$, then this pair is never sampled, so this case may be assigned arbitrarily. Otherwise set
$$y := \frac{t_r}{m_r} \in [-1, 1].$$
5. Output $b \in \{\pm 1\}$ according to
$$\Pr[b = 1 \mid r] = \frac{1 + y}{2}, \qquad \Pr[b = -1 \mid r] = \frac{1 - y}{2}.$$

Conditioned on the event that the sampled pair is r, we therefore have
$$\mathbb{E}[b \mid r] = y_r = \frac{t_r}{m_r}.$$

Now the probability that the oracle sample lands in the pair indexed by r is exactly
$$m_r = |\psi_r|^2 + |\psi_{r\oplus a}|^2,$$
since this happens iff the sample is either r or $r \oplus a$. Therefore
$$\begin{aligned}\mathbb{E}[b] &= \sum_r \Pr[\text{sampled pair } = r]\ \mathbb{E}[b \mid r] \\ &= \sum_r m_r \cdot \frac{t_r}{m_r} = \sum_r t_r = \langle\psi|P|\psi\rangle.\end{aligned}$$

This proves correctness. Once an oracle sample x has been drawn, the remaining work consists of computing $x \oplus a$, choosing the canonical representative r, evaluating the phase function η, reading two amplitudes from the explicit description of $|\psi\rangle$, and performing $O(1)$ arithmetic operations. This takes $O(\log d)$ time. □

3.5.2 Proof of Theorem 3.9

We may assume that the target accuracy satisfies $\alpha \le 1/2$, since otherwise we can run the construction at accuracy $1/2$. Set

$$\beta := \frac{\alpha}{C_1 \log^2(2/\alpha)}, \tag{3.8}$$

where C_1 is a sufficiently large universal constant. The algorithms below use β as their internal accuracy. In both algorithms, set

$$J := \left\lceil \log_2(1/\beta) \right\rceil, \qquad m_0 := 4\lceil C \log(2/\beta) \rceil, \qquad n_0 := m_0/4,$$

where C is another sufficiently large constant.

For every level $j = 0, \dots, J$, set

$$T_j := \left\lceil \beta^{-2} 4^{-j} \right\rceil, \qquad m_j := 4^j m_0.$$

Independently draw $k_{j,1}, \dots, k_{j,T_j}$ uniformly from $[N]$, and, for every pair (j,t), collect m_j independent Rademacher samples $X_{j,t,1}, \dots, X_{j,t,m_j}$ with mean $v_{k_{j,t}}$. All indices are fixed before any outcome is observed, so this sampling stage is nonadaptive.

Algorithm 3 BUILD-ESTIMATOR(β)

Input: Internal accuracy β and samples from the nonadaptive stage above.

Output: A nonnegative estimate $\hat{q}$ of $\sqrt{\mathbb{E}_k[v_k^2]}$.

$J \leftarrow \lceil \log_2(1/\beta) \rceil$; $m_0 \leftarrow 4\lceil C \log(2/\beta) \rceil$; $n_0 \leftarrow m_0/4$.

function EST-V-SQUARED(j,t)
  $m_j \leftarrow 4^j m_0$.
  $\mu^{(1)} \leftarrow \frac{4}{m_j} \sum_{a=1}^{m_j/4} X_{j,t,a}$.
  $\mu^{(2)} \leftarrow \frac{4}{m_j} \sum_{a=m_j/4+1}^{m_j/2} X_{j,t,a}$.
  return $\mu^{(1)}\mu^{(2)}$.
end function

function LEVEL-CHECK(j,t)
  $m_j \leftarrow 4^j m_0$.
  for $b = 0, \dots, j$ **do**
    $n_b \leftarrow 4^b n_0$; $\overline{X}_b \leftarrow \frac{1}{n_b} \sum_{a=m_j/2+1}^{m_j/2+n_b} X_{j,t,a}$.
    if $|\overline{X}_b| > 2^{-b}$ **then**
      return $\mathbf{1}[b = j]$.
    end if
  end for
  return 0.
end function

for $j = 0, \dots, J$ **do**
  $T_j \leftarrow \lceil \beta^{-2} 4^{-j} \rceil$.
  for $t = 1, \dots, T_j$ **do**
    $U_{j,t} \leftarrow$ EST-V-SQUARED(j,t).
    $\widetilde{U}_{j,t} \leftarrow \max\{0, \min\{U_{j,t}, 16 \cdot 4^{-j}\}\}$.
    $Z_{j,t} \leftarrow$ LEVEL-CHECK(j,t); $\hat{r}_{j,t} \leftarrow \widetilde{U}_{j,t} Z_{j,t}$.
  end for
  $\hat{r}_j \leftarrow T_j^{-1} \sum_{t=1}^{T_j} \hat{r}_{j,t}$.
end for
return $\hat{q} \leftarrow \sqrt{\sum_{j=0}^{J} \hat{r}_j}$.

We next formalize the random level selected by LEVEL-CHECK. Fix $x \in [-1, 1]$, draw one i.i.d. stream $Y_1, \dots, Y_{n_J}$ of Rademacher variables with mean x, and put

$$\overline{Y}_b := \frac{1}{n_b} \sum_{i=1}^{n_b} Y_i, \qquad n_b = 4^b n_0.$$

Let $L(x)$ be the least $b \le J$ for which $|\overline{Y}_b| > 2^{-b}$, and set $L(x) = J+1$ if there is no such level. For a level-j call to Level-Check, couple $Y_i = X_{j,t,m_j/2+i}$ for $i \le n_j$, and then extend the stream independently to length n_J. This coupling gives the exact identity

$$Z_{j,t} = 1_{L(v_{k_{j,t}})=j}. \tag{3.9}$$

The two empirical means used to form $U_{j,t}$ use disjoint samples and are therefore independent of each other and of $Z_{j,t}$, conditioned on $k_{j,t}$.

Lemma 3.13. *For every fixed $x \in [-1,1]$, except with probability $O(\beta^4)$ over the stream defining $L(x)$, the following statements hold:*

1. *if $L(x) = j \le J$, then*
$$0.9 \cdot 2^{-j} \le |x| \le 2.2 \cdot 2^{-j};$$
2. *if $L(x) = J+1$, then*
$$|x| \le 2.2 \cdot 2^{-J} = O(\beta).$$

Proof. For every $b \le J$, Hoeffding's inequality (Lemma 2.4) and the definition of n_b give

$$\Pr\left[|\overline{Y}_b - x| > 0.1 \cdot 2^{-b}\right] \le 2\exp(-cn_0).$$

Taking C sufficiently large and applying a union bound over the $J+1$ nested empirical means shows that, except with probability $O(\beta^4)$,

$$|\overline{Y}_b - x| \le 0.1 \cdot 2^{-b} \quad \text{for every } b \le J. \tag{3.10}$$

If $L(x) = j$, the threshold crossing at level j gives the lower bound. For $j \ge 1$, failure to cross at level $j-1$ gives the upper bound; for $j = 0$, the upper bound follows from $|x| \le 1$. If no threshold is crossed, the estimate at level J gives the second assertion. □

Lemma 3.14. *For every $j \le J$ and t, conditioned on $k_{j,t}$, except with probability $O(\beta^4)$,*

$$|\mu^{(1)} - v_{k_{j,t}}| \le 0.05 \cdot 2^{-j}, \qquad |\mu^{(2)} - v_{k_{j,t}}| \le 0.05 \cdot 2^{-j}.$$

Proof. Each mean averages $m_j/4 = 4^j n_0$ independent Rademacher variables. The claim follows from Hoeffding's inequality (Lemma 2.4), increasing C if necessary. □

Lemma 3.15. *For every $j \le J$ and t, conditioned on $k_{j,t}$,*

$$U_{j,t} Z_{j,t} = \widetilde{U}_{j,t} Z_{j,t}$$

except with probability $O(\beta^4)$.

Proof. Write $x = v_{k_{j,t}}$. If $Z_{j,t} = 0$, both sides vanish. If $Z_{j,t} = 1$, then (3.9) gives $L(x) = j$. On the simultaneous events from Lemmas 3.13 and 3.14,

$$0.9 \cdot 2^{-j} \le |x| \le 2.2 \cdot 2^{-j}$$

and both empirical means differ from x by at most $0.05 \cdot 2^{-j}$. Thus $\mu^{(1)}$ and $\mu^{(2)}$ have the same sign as x, so $U_{j,t} \ge 0$, and

$$U_{j,t} \le (2.25)^2 \, 4^{-j} < 16 \cdot 4^{-j}.$$

The two-sided clipping therefore leaves $U_{j,t}$ unchanged whenever $Z_{j,t} = 1$, apart from the stated exceptional event. □

Define

$$r_j \coloneqq \mathop{\mathbb{E}}_{k,L}\Big[v_k^2 \mathbb{1}_{L(v_k)=j} \Big], \qquad 0 \le j \le J+1,$$

where k is uniform on $[N]$. By Lemma 3.15, the independence of the three sample blocks, and (3.9), we have

$$\mathbb{E}[\hat{r}_{j,t} \mid k_{j,t}] = v_{k_{j,t}}^2 \Pr[L(v_{k_{j,t}}) = j \mid k_{j,t}] + O(\beta^2), \tag{3.11}$$

$$\mathbb{E}[\hat{r}_j] = r_j + O(\beta^2). \tag{3.12}$$

Here we used $|U_{j,t} Z_{j,t}|, |\widetilde{U}_{j,t} Z_{j,t}| \le 1$; the weaker $O(\beta^2)$ error is convenient below.

Lemma 3.16. *With probability at least $1 - O(\beta^4)$, simultaneously for every $0 \le j \le J$,*

$$|\hat{r}_j - r_j| \le C_2 \log(2/\beta) \Big(\beta \sqrt{r_j} + \beta^2 \Big).$$

Proof. Fix j, set $\mu_j = \mathbb{E}[\hat{r}_j]$, and note that the independent summands satisfy

$$0 \le \hat{r}_{j,t} \le B_j := 16 \cdot 4^{-j}.$$

Apply Corollary 2.6 to their sum with $\lambda = C_3 \log(2/\beta)$. Since $T_j \ge \beta^{-2} 4^{-j}$, we have

$$\frac{B_j}{T_j} \le 16\beta^2.$$

After dividing the Bernstein bound by T_j, it follows that

$$|\hat{r}_j - \mu_j| \le C_4 \log(2/\beta)\Big(\beta\sqrt{\mu_j} + \beta^2\Big)$$

except with probability $O(\beta^6)$. Equation (3.12) and $\sqrt{\mu_j} \le \sqrt{r_j} + O(\beta)$ yield the claimed inequality. A union bound over $J + 1 = O(\log(2/\beta))$ levels completes the proof. □

Put $r := \mathbb{E}_k[v_k^2]$. The level events partition the probability space, so

$$r = \sum_{j=0}^{J+1} r_j.$$

On the event in Lemma 3.13, $L(v_k) = J + 1$ implies $v_k^2 = O(\beta^2)$; on its complement we use $v_k^2 \le 1$. Consequently,

$$r_{J+1} = O(\beta^2). \tag{3.13}$$

On the simultaneous event from Lemma 3.16, Cauchy–Schwarz gives

$$|\hat{q}^{\,2} - r| \le C_5\Big(\beta \log^{3/2}(2/\beta)\sqrt{r} + \beta^2 \log^2(2/\beta)\Big). \tag{3.14}$$

Indeed, $\sum_{j=0}^{J} \sqrt{r_j} \le \sqrt{J+1}\sqrt{\sum_j r_j} \le \sqrt{J+1}\sqrt{r}$.

Let

$$a := C_6 \beta \log^{3/2}(2/\beta).$$

After increasing C_6, the right-hand side of (3.14) is at most $a\sqrt{r} + a^2$. Since $\hat{q}^{\,2} \ge 0$, if $r \ge a^2$, then

$$|\hat{q} - \sqrt{r}| \le \frac{a\sqrt{r} + a^2}{\sqrt{r}} \le 2a.$$

If $r < a^2$, then $\hat{q}^{\,2} \le r + a\sqrt{r} + a^2 < 3a^2$, and hence $|\hat{q} - \sqrt{r}| < 3a$. The choice of β in (3.8), with C_1 sufficiently large, makes $3a \le \alpha$. The simultaneous event in Lemma 3.16 has probability at least $2/3$, proving the accuracy guarantee in Theorem 3.9.

Finally, the number of nonadaptive queries is

$$\begin{aligned}\sum_{j=0}^{J} T_j m_j &\le \sum_{j=0}^{J} \bigl(\beta^{-2} 4^{-j} + 1\bigr) 4^j m_0 \\ &= O\Bigl(\beta^{-2} \log^2(2/\beta)\Bigr) = \widetilde{O}(\alpha^{-2}).\end{aligned}$$

The estimator can be evaluated in time linear in this number of outcomes. This completes the proof.

Chapter 4: Learning Structured Quantum States

The preceding chapter studied tomography of arbitrary pure states. Although the algorithm there is nearly optimal, its cost is still exponential in the number of qubits. In this chapter, we ask when additional structure can make tomography efficient.

We give polynomial-time tomography algorithms for two structured classes of states. The first is the class of Slater determinants, which describe non-interacting fermions. The second is the class of states with high stabilizer dimension, which includes states prepared by Clifford circuits with few non-Clifford gates as a special case.

Our tomography algorithm for Slater determinants is stated in Theorem 4.1. The corresponding guarantees for states with high stabilizer dimension are given in Corollaries 4.19 and 4.31.

4.1 Non-Interacting Fermion States

Fermions are particles, such as electrons, that obey the *Pauli exclusion principle*: no two fermions can occupy the same mode. Thus, a configuration of n fermions among m modes is specified by choosing which n modes are occupied, giving $\binom{m}{n}$ possible configurations. Since these configurations can occur in superposition, an n-fermion state in m modes lives in a Hilbert space of dimension $\binom{m}{n}$.

We will study a much more structured class of states: pure non-interacting fermion states, also called Slater determinants. Although a Slater determinant lives in this potentially exponentially large Hilbert space, it has a succinct description. Up to an irrelevant global phase, it is specified by an n-dimensional subspace of $\mathbb{C}^m$, or equivalently by the rank-n projector onto that subspace.

Our algorithm learns this projector, also called the one-body reduced density matrix, by estimating one-mode occupation probabilities in several measurement bases. The main technical step is to show that learning this one-body information suffices to learn the entire many-body state.

The main result of this section is the following.

Theorem 4.1 (Fermionic tomography). *Let $|\Psi\rangle$ be a Slater determinant of n fermions in m modes, and let $\epsilon, \delta \in (0,1)$. There is an algorithm that uses*

$$O\left(\frac{m^3 n^2 \log(m/\delta)}{\epsilon^4}\right)$$

copies of $|\Psi\rangle$, $O(m)$ measurement bases, and

$$O\left(\frac{m^4 n^2 \log(m/\delta)}{\epsilon^4}\right)$$

classical time. With probability at least $1-\delta$, the algorithm outputs an $m \times n$ column-orthonormal matrix $\widehat{A}$ whose associated Slater determinant $|\widehat{\Psi}\rangle$ satisfies

$$d_{\mathrm{tr}}\left(\widehat{\Psi}, \Psi\right) \leq \epsilon.$$

4.1.1 Preliminaries and Problem Setup

We briefly recall the fixed-particle-number formalism; see, for example, [Aar22, Chapters 7–10] for a longer introduction. Let $\Lambda_{m,n} \coloneqq \{S \subseteq [m] \,:\, |S| = n\}$. An element $S \in \Lambda_{m,n}$ records which n of the m modes are occupied. The occupation-number states

$$\{|S\rangle \,:\, S \in \Lambda_{m,n}\}$$

form an orthonormal basis for the Hilbert space of n fermions in m modes.

A unitary $U \in \mathbb{C}^{m\times m}$ describes a transformation of the single-particle modes. Such a unitary can be implemented using $O(m^2)$ two-mode beamsplitters and single-mode phaseshifters [RZBB94]. A beamsplitter acts on a pair of modes by a two-dimensional rotation, while a phaseshifter multiplies one mode by a complex phase. On n fermions, U induces a unitary transformation that we denote by $\varphi(U)$, characterized by

$$\langle S|\varphi(U)|T\rangle = \det(U_{S,T}), \qquad S,T \in \Lambda_{m,n}, \tag{4.1}$$

where $U_{S,T}$ is the submatrix with rows indexed by S and columns indexed by T. The determinant appears because of fermionic antisymmetry: exchanging two particles changes the sign of the corresponding amplitude. In the qubit occupation-number representation,

the transformations $\varphi(U)$ are Hamming-weight-preserving matchgate unitaries [Val02; TD02].

Let $|1_n\rangle := |\{1, \dots, n\}\rangle$ denote the state in which the first n modes are occupied. Every Slater determinant can be written as $|\Psi\rangle = \varphi(U)|1_n\rangle$ for some $U \in \mathbb{C}^{m\times m}$. Let $A \in \mathbb{C}^{m\times n}$ be the matrix formed by the first n columns of U. Then $A^\dagger A = I_n$, and Equation (4.1) gives

$$|\Psi\rangle = \sum_{S\in\Lambda_{m,n}} \det(A_S)|S\rangle, \tag{4.2}$$

where A_S is the $n \times n$ submatrix whose rows are indexed by S.

It will be convenient to work with the *kernel matrix*[1]

$$K := AA^\dagger \in \mathbb{C}^{m\times m}.$$

The matrix K is Hermitian and positive semidefinite, satisfies $K^2 = K$, and has $\operatorname{rank}(K) = \operatorname{tr}(K) = n$.

By the Cauchy–Binet identity,

$$\Pr[\text{observe } S] = |\langle S|\Psi\rangle|^2 = |\det(A_S)|^2 = \det(K_S), \tag{4.3}$$

where K_S is the principal submatrix indexed by S. More generally, if $T \subseteq [m]$ and $|T| \le n$, then

$$\Pr[\text{every mode in } T \text{ is occupied}] = \det(K_T).$$

In particular, k_{ii} is the probability that mode i is occupied.

The matrix A is not a unique description of the physical state. If $R \in \mathbb{C}^{n\times n}$ is unitary, then replacing A by AR changes every amplitude in Equation (4.2) by the same phase $\det(R)$, while

$$(AR)(AR)^\dagger = AA^\dagger.$$

Thus, K records the underlying n-dimensional subspace without choosing a particular basis for it.

[1] We use the term "kernel matrix" following our original paper, by analogy with the kernel of a determinantal point process [Mac75]; the more standard fermionic terminology is the 1-body reduced density matrix (1-RDM).

Finally, applying a single-particle unitary $V \in \mathbb{C}^{m\times m}$ simply conjugates the kernel matrix:

$$K \longmapsto VKV^{\dagger}. \tag{4.4}$$

4.1.2 Learning the Kernel Matrix

We now show how to learn the kernel matrix K. The diagonal entries are immediate: as observed above, k_{ii} is exactly the probability that mode i is occupied, so all diagonal entries can be estimated by measuring in the occupation-number basis.

To learn an off-diagonal entry k_{ij}, we interfere modes i and j. Consider the two beamsplitters

$$B_{\mathrm{Re}} \coloneqq \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}, \qquad B_{\mathrm{Im}} \coloneqq \frac{1}{\sqrt{2}}\begin{pmatrix}1 & \mathrm{i}\\ 1 & -\mathrm{i}\end{pmatrix}.$$

After applying B_{Re} to modes i and j, the probability that mode i is occupied becomes

$$\frac{1}{2}\big(k_{ii} + k_{jj} + 2\,\mathrm{Re}(k_{ij})\big). \tag{4.5}$$

Similarly, applying B_{Im} makes this probability

$$\frac{1}{2}\big(k_{ii} + k_{jj} + 2\,\mathrm{Im}(k_{ij})\big). \tag{4.6}$$

Since k_{ii} and k_{jj} are already known, these two experiments recover the real and imaginary parts of k_{ij}. In other words, interference turns the off-diagonal entries of K into ordinary occupation probabilities that we can estimate directly.

We do not need a separate measurement basis for every pair of modes. The edges of the complete graph on the m modes can be partitioned into $O(m)$ matchings. For each matching, we apply the same beamsplitter simultaneously to every pair of modes in the matching. One measurement basis therefore gives information about $O(m)$ entries of K at once, and $O(m)$ bases suffice to learn the entire matrix.

Once the entries have been estimated, we obtain an empirical matrix $\widetilde{K}$. Unlike the true kernel K, this matrix need not be a rank-n projector. We therefore round it by taking its top n eigenvectors; these form the columns of our estimate $\widehat{A}$.

We formalize this procedure in the following algorithm.

Algorithm 4 LEARN-SLATER($|\Psi\rangle, \gamma, \delta$)

Input: Copies of an n-fermion Slater determinant $|\Psi\rangle$ in m modes; entrywise accuracy $\gamma \in (0,1)$; failure probability $\delta \in (0,1)$.

Output: An $m \times n$ column-orthonormal matrix $\widehat{A}$.

Measure $O(\log(m/\delta)/\gamma^2)$ copies in the occupation-number basis, obtaining estimates $\tilde{k}_{ii}$ for every $i \in [m]$.
Choose $O(m)$ matchings whose union contains every unordered pair $\{i,j\} \subseteq [m]$.
for each matching $\mathcal{M}$ **do**
    Apply B_{Re} in parallel to the pairs in $\mathcal{M}$ and estimate the resulting occupation probabilities of the first mode in each pair.
    Apply B_{Im} in parallel to the pairs in $\mathcal{M}$ and estimate the analogous occupation probabilities.
    for each $\{i,j\} \in \mathcal{M}$ **do**
        Use Equations (4.5) and (4.6) to obtain $\mathrm{Re}(\tilde{k}_{ij})$ and $\mathrm{Im}(\tilde{k}_{ij})$.
        Set $\tilde{k}_{ij} \leftarrow \mathrm{Re}(\tilde{k}_{ij}) + \mathrm{i}\,\mathrm{Im}(\tilde{k}_{ij})$ and $\tilde{k}_{ji} \leftarrow \tilde{k}_{ij}^*$.
    end for
end for
Form $\widetilde{K} = (\tilde{k}_{ij})$ and compute an eigendecomposition $\widetilde{K} = Q\Lambda Q^\dagger$, with eigenvalues in nonincreasing order.
Let $\widehat{A}$ be the first n columns of Q.
return $\widehat{A}$.

The algorithm uses $O(m)$ measurement bases and estimates $O(m)$ occupation probabilities in each basis. Assigning failure probability $O(\delta/m^2)$ to each estimate, Hoeffding's inequality (Lemma 2.4) shows that $O(\log(m/\delta)/\gamma^2)$ samples per basis suffice. There are constant overheads when solving for matrix entries using Equations (4.5) and (4.6), but these are absorbed into the big-O notation. A union bound then shows that all reconstructed entries obey

$$|\tilde{k}_{ij} - k_{ij}| \le \gamma \tag{4.7}$$

with probability at least $1 - \delta$. The measurement stage uses

$$O\left(\frac{m \log(m/\delta)}{\gamma^2}\right) \quad \text{copies and} \quad O\left(\frac{m^2 \log(m/\delta)}{\gamma^2}\right) \quad \text{classical time.}$$

The final eigendecomposition takes $O(m^3)$ time.

The algorithm outputs a valid description of a Slater determinant, since $\widehat{A}$ consists of orthonormal columns of Q. It remains to show that the entrywise guarantee on $\widetilde{K}$ implies that the learned many-body state is close to the original one. We do this in the next subsection.

4.1.3 Error Analysis

The goal of this subsection is to show that an accurate estimate of the kernel matrix leads to an accurate estimate of the underlying Slater determinant. We first show that nearby rank-n kernel matrices give rise to nearby Slater determinants. We will use the following elementary bound.

Lemma 4.2. *If $a_1, \dots, a_n \in [0, 1]$, then*

$$1 - \prod_{i=1}^{n} a_i \leq n \max_{i \in [n]} (1 - a_i).$$

Proof. For $x, y \in [0, 1]$,

$$1 - xy = (1 - x) + x(1 - y) \leq (1 - x) + (1 - y).$$

Iterating this inequality gives

$$1 - \prod_{i=1}^{n} a_i \leq \sum_{i=1}^{n} (1 - a_i) \leq n \max_{i \in [n]} (1 - a_i). \qquad \square$$

Lemma 4.3. *Let $A, \widehat{A} \in \mathbb{C}^{m \times n}$ be column-orthonormal, and define*

$$K \coloneqq AA^\dagger, \qquad \widehat{K} \coloneqq \widehat{A}\widehat{A}^\dagger.$$

If $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_n)$ contains the singular values of $\widehat{A}^\dagger A$, then

$$\|I_n - \Sigma^2\|_{\mathrm{op}} \leq \|\widehat{K} - K\|_{\mathrm{op}}.$$

Proof. Let $Q\Sigma V^\dagger$ be a singular value decomposition of $\widehat{A}^\dagger A$. Then

$$V\Sigma^2 V^\dagger = (\widehat{A}^\dagger A)^\dagger \widehat{A}^\dagger A = A^\dagger \widehat{A}\widehat{A}^\dagger A.$$

Since the spectral norm is unitarily invariant and $A^\dagger A = I_n$,

$$\begin{aligned}
\|I_n - \Sigma^2\|_{\mathrm{op}} &= \|I_n - A^\dagger \widehat{A}\widehat{A}^\dagger A\|_{\mathrm{op}} \\
&= \|A^\dagger (A - \widehat{A}\widehat{A}^\dagger A)\|_{\mathrm{op}} \\
&\leq \|A - \widehat{A}\widehat{A}^\dagger A\|_{\mathrm{op}} \\
&= \|(AA^\dagger - \widehat{A}\widehat{A}^\dagger)A\|_{\mathrm{op}} \\
&\leq \|K - \widehat{K}\|_{\mathrm{op}}.
\end{aligned}$$

$\square$

Theorem 4.4 (The kernel matrix suffices). *Let $|\Psi\rangle$ and $|\widehat{\Psi}\rangle$ be Slater determinants of n fermions in m modes, described by column-orthonormal matrices $A, \widehat{A} \in \mathbb{C}^{m\times n}$. If $K = AA^\dagger$ and $\widehat{K} = \widehat{A}\widehat{A}^\dagger$, then*

$$d_{\mathrm{tr}}\big(\widehat{\Psi}, \Psi\big) \le \sqrt{n\|\widehat{K} - K\|_{\mathrm{op}}}.$$

Proof. By Equation (4.2) and the Cauchy–Binet formula,

$$\begin{aligned} d_{\mathrm{tr}}\big(\widehat{\Psi}, \Psi\big) &= \sqrt{1 - |\langle\widehat{\Psi}|\Psi\rangle|^2} \\ &= \sqrt{1 - |\sum_{S\in\Lambda_{m,n}} \det(\widehat{A}_S)^* \det(A_S)|^2} \\ &= \sqrt{1 - |\det(\widehat{A}^\dagger A)|^2}. \end{aligned}$$

Let $Q\Sigma V^\dagger$ be a singular value decomposition of $\widehat{A}^\dagger A$, with singular values $\sigma_1, \dots, \sigma_n$. Because A and $\widehat{A}$ are column-orthonormal, $\sigma_i \le 1$ for every i [HJ91, Chapter 3]. Consequently,

$$|\det(\widehat{A}^\dagger A)|^2 = \prod_{i=1}^{n} \sigma_i^2.$$

Applying Lemma 4.2 and then Lemma 4.3, we obtain

$$\begin{aligned} d_{\mathrm{tr}}\big(\widehat{\Psi}, \Psi\big) &= \sqrt{1 - \prod_{i=1}^{n} \sigma_i^2} \\ &\le \sqrt{n \max_{i\in[n]}(1 - \sigma_i^2)} \\ &= \sqrt{n\|I_n - \Sigma^2\|_{\mathrm{op}}} \\ &\le \sqrt{n\|\widehat{K} - K\|_{\mathrm{op}}}. \end{aligned}$$

□

The remaining issue is that the empirical matrix $\widetilde{K}$ need not itself be a rank-n projector. The algorithm fixes this by taking the top n eigenvectors of $\widetilde{K}$, thereby producing a valid rank-n kernel matrix. We now show that this rounding step preserves accuracy, using the following standard consequence of Weyl's inequality.

Theorem 4.5 (Weyl's inequality [Wey12]). *Let $M = N + E$ be Hermitian matrices of the same dimension. If $\lambda_1(M) \ge \cdots \ge \lambda_m(M)$ and $\lambda_1(N) \ge \cdots \ge \lambda_m(N)$, then*

$$|\lambda_i(M) - \lambda_i(N)| \le \|E\|_{\mathrm{op}} \qquad (i \in [m]).$$

Theorem 4.6. *Suppose that the empirical matrix $\widetilde{K}$ produced by Algorithm 4 satisfies Equation (4.7). Let $\widehat{A}$ be the algorithm's output, and let $|\widehat{\Psi}\rangle$ be its associated Slater determinant. Then*

$$d_{\mathrm{tr}}\left(\widehat{\Psi}, \Psi\right) \leq \sqrt{2nm\gamma}.$$

Proof. Write

$$E \coloneqq \widetilde{K} - K, \qquad \widehat{K} \coloneqq \widehat{A}\widehat{A}^{\dagger}.$$

Since every entry of E has magnitude at most γ,

$$\|E\|_{\mathrm{op}} \leq \|E\|_{\mathrm{F}} \leq m\gamma. \tag{4.8}$$

Let $\widetilde{K} = Q\Lambda Q^{\dagger}$ be the ordered eigendecomposition computed by the algorithm, and set

$$P_n \coloneqq \mathrm{diag}(\underbrace{1, \ldots, 1}_{n}, \underbrace{0, \ldots, 0}_{m-n}).$$

Since $\widehat{A}$ consists of the first n columns of Q, $\widehat{K} = QP_nQ^{\dagger}$. The eigenvalues of the true kernel K are precisely the diagonal entries of P_n. Weyl's inequality therefore implies

$$\|\widehat{K} - \widetilde{K}\|_{\mathrm{op}} = \|P_n - \Lambda\|_{\mathrm{op}} \leq \|E\|_{\mathrm{op}}.$$

By the triangle inequality and Equation (4.8),

$$\|\widehat{K} - K\|_{\mathrm{op}} \leq \|\widehat{K} - \widetilde{K}\|_{\mathrm{op}} + \|\widetilde{K} - K\|_{\mathrm{op}} \leq 2m\gamma. \tag{4.9}$$

Applying Theorem 4.4 to Equation (4.9) completes the proof. □

Proof of Theorem 4.1. Run Algorithm 4 with

$$\gamma \coloneqq \frac{\epsilon^2}{2nm}.$$

With probability at least $1 - \delta$, the entrywise guarantee Equation (4.7) holds, and hence Theorem 4.6 gives

$$d_{\mathrm{tr}}\left(\widehat{\Psi}, \Psi\right) \leq \sqrt{2nm\gamma} = \epsilon.$$

Substituting this value of γ into the resource bounds above gives

$$O\left(\frac{m^3 n^2 \log(m/\delta)}{\epsilon^4}\right) \quad \text{copies}$$

and

$$O\left(\frac{m^4 n^2 \log(m/\delta)}{\epsilon^4}\right) \quad \text{classical time.}$$

The measurements use one occupation basis together with two bases for each of $O(m)$ matchings, and therefore use $O(m)$ bases in total. □

4.2 States with High Stabilizer Dimension

Our second class consists of states with many Pauli symmetries. At one extreme are stabilizer states: a stabilizer state is the common $+1$-eigenvector of an abelian group of 2^n Pauli operators. Montanaro [Mon17] showed that stabilizer states can be learned efficiently. At the other extreme are arbitrary pure states, which, as we saw in Chapter 3, require exponential copies and time to learn. Here we study what happens in between these two extremes.

Specifically, we consider states whose stabilizer dimension (Definition 2.16) is at least $n - t$. The parameter t interpolates between stabilizer states, for which $t = 0$, and arbitrary pure states, for which t can be as large as n. This class includes, in particular, states prepared by Clifford circuits with few non-Clifford gates.

The basic idea behind our algorithms is simple. We first learn enough of the state's unsigned stabilizer group (Definition 2.15) to find a Clifford transformation C that compresses all of the non-stabilizer degrees of freedom onto at most t qubits. Concretely, after applying C, the state approximately takes the form

$$C|\psi\rangle \approx |\phi\rangle \otimes |x\rangle,$$

where $|\phi\rangle$ is a state on at most t qubits and $|x\rangle$ is a computational-basis state on the remaining qubits. We have therefore reduced the problem to learning a t-qubit state, which we can do using the tomography algorithm from Chapter 3.

We give two ways to learn the unsigned stabilizer group needed for this compression. The first uses Bell-difference sampling: each sample uses four copies, but only two copies are measured jointly at a time. This gives the more efficient algorithm. The second uses only single-copy measurements and requires no coherent quantum memory. The resulting tomography guarantees are stated in Corollaries 4.19 and 4.31.

4.2.1 Preliminaries

We use the Weyl and binary-symplectic notation from Section 2.5. For a pure state $|\psi\rangle$, its unsigned stabilizer group $\mathrm{Weyl}(|\psi\rangle)$ is defined in Definition 2.15, and its dimension is the stabilizer dimension of $|\psi\rangle$. We will also use the characteristic distribution p_ψ and the Bell-difference distribution q_ψ, reviewed in Section 2.7.

A natural subclass of states with high stabilizer dimension is Clifford circuits containing only a few non-Clifford gates.

Definition 4.7 (s-doped Clifford circuit). An s-doped Clifford circuit consists of Clifford gates and at most s arbitrary single-qubit non-Clifford gates.

We use s for the number of non-Clifford gates and reserve t for the stabilizer nullity $n - \dim \mathrm{Weyl}(|\psi\rangle)$. The two parameters are related as follows.

Lemma 4.8 ([GIKL24, Lemma 4.2]). *If $|\psi\rangle$ is the output of an s-doped Clifford circuit, then*

$$\dim \mathrm{Weyl}(|\psi\rangle) \geq n - 2s.$$

Proof sketch. A single-qubit non-Clifford gate commutes with at least one quarter of the current unsigned stabilizer group, and therefore decreases its dimension by at most two. Clifford gates preserve stabilizer dimension. Induction on the number of non-Clifford gates proves the claim. □

The single-copy algorithm below will also use uniformly random Clifford circuits. Such circuits of size $O(n^2)$ can be sampled in $O(n^2)$ classical time [KS14; BM21; Ber21].

4.2.2 Heavy Subspaces and Clifford Compression

To carry out the compression described above, we do not need to recover Weyl($|\psi\rangle$) exactly. It is enough to find a large subspace on which the characteristic distribution p_ψ has nearly as much mass as possible. This motivates the following definition.

Definition 4.9 (Heavy subspace). For $t \in \{0, \dots, n\}$ and $\epsilon \in (0, 1)$, a subspace $S \subseteq \mathbb{F}_2^{2n}$ is (t, ϵ)-*heavy* for $|\psi\rangle$ if

$$\dim(S) \geq n - t \qquad \text{and} \qquad \sum_{x \in S} p_\psi(x) \geq \left(1 - \frac{\epsilon^2}{4}\right) \frac{|S|}{2^n}.$$

Recall that $p_\psi(x) \leq 2^{-n}$ for every x. Thus, a heavy subspace is one on which p_ψ is close, on average, to its largest possible value. We will show that this is already enough to recover essentially the same structure that the unsigned stabilizer subgroup would provide.

We begin with the linear-algebraic part of the compression. Any isotropic subspace can be mapped by a Clifford transformation to a standard group of Z-type Paulis.

Lemma 4.10 (Clifford compression). *Given m vectors spanning a d-dimensional isotropic subspace $A \subseteq \mathbb{F}_2^{2n}$, there is an algorithm that outputs a Clifford circuit C satisfying*

$$C(A) = 0^{2n-d} \times \mathbb{F}_2^d.$$

The algorithm runs in $O(mn \min\{m, n\})$ time, and C contains $O(nd)$ gates.

Proof. Gaussian elimination first produces a row-echelon basis for A, written as a $d \times 2n$ stabilizer tableau M. On this tableau, a Hadamard gate swaps columns j and $n + j$, a phase gate adds column j to column $n + j$, and a CNOT from j to k adds column j to column k and column $n + k$ to column $n + j$; see [AG04]. Row operations may also be performed freely, since they do not change the row span.

Process the rows from top to bottom. At the start of iteration i, maintain the following invariant: the first $i - 1$ rows are the first $i - 1$ standard basis vectors in the X block, all other entries in those rows and columns vanish, and the first $i - 1$ columns of the Z block vanish. The invariant is vacuous for $i = 1$.

For each $j \geq i$, phase and Hadamard gates on qubit j map the pair of entries in columns $j, n+j$ of row i either to $(1,0)$, if it was nonzero, or leave it at $(0,0)$. The row is nonzero and has no support in the previously processed columns, so some $k \geq i$ now has an X-entry equal to one. If $k \neq i$, a CNOT with control k and target i creates a pivot in column i. CNOTs with control i and targets $j > i$ then clear the remaining X-entries in row i, and row additions clear column i below the pivot.

All gates in iteration i act only on qubits $i, \ldots, n$, so the previously processed rows and columns remain fixed. The transformed row i is the ith standard X-basis vector. Because Clifford gates preserve the symplectic form and the row span remains isotropic, its symplectic product with every other row is zero; equivalently, column $n+i$ of the Z block vanishes. This proves the invariant for iteration $i+1$. After d iterations the tableau therefore has block form

$$\left(I_d \quad 0 \,\middle|\, 0\right).$$

A layer of Hadamards moves I_d to the first d columns of the Z block. Finally, a permutation of the qubits moves these columns to the last d positions, yielding $\begin{pmatrix} 0 & 0 & I_d \end{pmatrix}$.

The initial elimination costs $O(mn \min\{m,n\})$. The remaining operations use $O(n)$ Clifford gates and $O(d)$ row operations per pivot, for $O(nd)$ gates. Updating the tableau costs $O(nd^2)$, which is no larger than $O(mn \min\{m,n\})$ because $d \leq \min\{m,n\}$. □

Clifford conjugation merely relabels the characteristic distribution (Equation (2.21)). Thus, once we know that a heavy subspace is isotropic, we may use Lemma 4.10 to put it into this standard Z-type form without changing its p_ψ-mass.

We next show that a heavy subspace has the structure required by Lemma 4.10. Recall from Fact 2.18 that every pair of elements of $M_{1/2}(\psi)$ has symplectic product zero.

Lemma 4.11. *If a subspace $S \subseteq \mathbb{F}_2^{2n}$ satisfies*

$$\sum_{x \in S} p_\psi(x) > \frac{3}{4} \frac{|S|}{2^n},$$

then S is isotropic.

Proof. Let $A = S \cap M_{1/2}(\psi)$. If $|A| \le |S|/2$, then

$$\sum_{x \in S} p_\psi(x) \le \frac{|S|}{2}\frac{1}{2^n} + \frac{|S|}{2}\frac{1}{2^{n+1}} = \frac{3}{4}\frac{|S|}{2^n},$$

contrary to the hypothesis. Hence $|A| > |S|/2$, so A cannot be contained in a proper subspace of S and therefore spans S. The large-expectation commutation fact shows that this span is isotropic. □

We next give an operational interpretation of heavy p_ψ-mass, when the stabilizer group is written in the Z-type form in Lemma 4.10. We show that, for such a subgroup, its p_ψ-mass is proportional to the probability that computational-basis measurements on two copies agree. We first record the elementary identity behind this interpretation. For $z \in \mathbb{F}_2^k$, write $Z^z = Z^{z_1} \otimes \cdots \otimes Z^{z_k}$.

Proposition 4.12. *For every k,*

$$\sum_{z \in \mathbb{F}_2^k} |z\rangle\langle z| \otimes |z\rangle\langle z| = \frac{1}{2^k} \sum_{z \in \mathbb{F}_2^k} Z^z \otimes Z^z.$$

Proof. Expand $|z\rangle\langle z| = 2^{-k}\sum_a (-1)^{a\cdot z} Z^a$, tensor the two expansions, and use character orthogonality. □

Proposition 4.13. *Let $S = 0^{n+t} \times \mathbb{F}_2^{n-t}$. If the last $n-t$ qubits of two copies of $|\psi\rangle$ are measured in the computational basis, the probability of observing the same outcome twice is*

$$2^t \sum_{x \in S} p_\psi(x).$$

Proof. Insert Proposition 4.12 into the projector onto equal outcomes:

$$\sum_{z \in \mathbb{F}_2^{n-t}} \langle\psi|^{\otimes 2} \big(I^{\otimes t} \otimes |z\rangle\langle z|\big)^{\otimes 2} |\psi\rangle^{\otimes 2}$$
$$= \frac{1}{2^{n-t}} \sum_{z \in \mathbb{F}_2^{n-t}} \langle\psi| I^{\otimes t} \otimes Z^z |\psi\rangle^2 = 2^t \sum_{x \in S} p_\psi(x).$$

□

A large collision probability means that the measurement distribution is concentrated on a single outcome. For a pure state, this says exactly that the measured register is nearly fixed to a computational-basis state, leaving all remaining freedom on the first t qubits. We formalize this below.

Lemma 4.14. *Let $S = 0^{n+t} \times \mathbb{F}_2^{n-t}$, and suppose*

$$\sum_{x \in S} p_\psi(x) \geq \frac{1-\eta}{2^t}.$$

Then there are a computational-basis state $|z\rangle$ on $n-t$ qubits and a state $|\varphi\rangle$ on t qubits such that

$$F(\psi, \varphi \otimes z) \geq 1 - \eta.$$

Moreover, $|\varphi\rangle$ is obtained by postselecting the last $n-t$ qubits of $|\psi\rangle$ on outcome z.

Proof. Write $|\psi\rangle = \sum_z \alpha_z |\varphi_z\rangle|z\rangle$. The largest $|\alpha_z|^2$ is at least $\sum_z |\alpha_z|^4$, and the latter is the collision probability. Proposition 4.13 and the assumed mass bound show that it is at least $1-\eta$. Choose an outcome attaining the maximum. □

We can now combine the pieces. A heavy subspace is isotropic, so a Clifford transformation puts it into standard Z-type form. Its large characteristic mass then forces the transformed state to be close to a product of an arbitrary t-qubit state and a computational-basis state on the remaining qubits. Conjugating back gives the compression statement we will use in the tomography algorithms.

Lemma 4.15 (Robust product-state structure). *Let $S \subseteq \mathbb{F}_2^{2n}$ be isotropic of dimension $n-t$, and suppose*

$$\sum_{x \in S} p_\psi(x) \geq \frac{1-\eta}{2^t}.$$

Then there is a state $|\widehat{\psi}\rangle$ with $S \subseteq \mathrm{Weyl}(|\widehat{\psi}\rangle)$ and

$$F\left(\psi, \widehat{\psi}\right) \geq 1 - \eta.$$

More explicitly,

$$|\widehat{\psi}\rangle = C^\dagger |\varphi\rangle|z\rangle,$$

where C is the circuit from Lemma 4.10, $|z\rangle$ is an $(n-t)$-qubit basis state, and

$$|\varphi\rangle = \frac{(I \otimes \langle z|)C|\psi\rangle}{\|(I \otimes \langle z|)C|\psi\rangle\|_2}.$$

Proof. Set $|\phi\rangle = C|\psi\rangle$. By (2.21),

$$\sum_{x \in C(S)} p_\phi(x) = \sum_{x \in S} p_\psi(x) \geq \frac{1-\eta}{2^t}.$$

Since $C(S) = 0^{n+t} \times \mathbb{F}_2^{n-t}$, Lemma 4.14 gives $|\varphi\rangle|z\rangle$ with fidelity at least $1-\eta$ to $C|\psi\rangle$. Fidelity is unitarily invariant, and every element of $C(S)$ stabilizes $|\varphi\rangle|z\rangle$ up to sign. Conjugating back proves both conclusions. □

4.2.3 Reduction to Pure State Tomography

Suppose we have found a heavy subspace S. The preceding subsection shows that, after an efficiently computable Clifford transformation C, the state is close to $|\varphi\rangle|z\rangle$, where $|\varphi\rangle$ lives on at most t qubits and $|z\rangle$ is a computational-basis state on the remaining qubits. This almost reduces the problem to ordinary t-qubit tomography.

Indeed, we can learn z by measuring the last qubits of several copies. Conditioning on this outcome then leaves the first t qubits in the residual state $|\varphi\rangle$. We learn this state using any pure-state tomography algorithm and finally apply $C^\dagger$ to reconstruct the original state.

To state this reduction independently of the particular tomography routine used for the residual state, we introduce the following notation.

Definition 4.16 (Residual tomography complexity). Let $N_{r,\eta,\delta}$ and $M_{r,\eta,\delta}$ denote, respectively, the copy and time complexities of pure-state tomography on r qubits to trace distance η, with failure probability at most δ, using only single-copy measurements.

The tomography algorithm from Theorem 3.6 gives

$$N_{r,\eta,\delta} = \widetilde{O}\left(\frac{2^r \log(1/\delta)}{\eta^2}\right), \qquad M_{r,\eta,\delta} = \widetilde{O}\left(\frac{2^r \log(1/\delta)}{\eta^2}\right). \tag{4.10}$$

For now, we leave N and M abstract so that the reduction itself is separated from this particular choice of algorithm.

The reduction is summarized in the following algorithm.

Algorithm 5 Tomography from a heavy stabilizer subspace

Input: Copies of $|\psi\rangle$, parameters $\epsilon, \delta \in (0,1)$, and a basis of $S \subseteq \mathbb{F}_2^{2n}$
Promise: S is (t,ϵ)-heavy for $|\psi\rangle$
Output: A classical description of $|\widehat{\psi}\rangle$ satisfying $d_{\mathrm{tr}}\big(\psi, \widehat{\psi}\big) \le \epsilon$ with probability at least $1-\delta$

Let $\hat{t} \le t$ satisfy $\dim(S) = n - \hat{t}$.
Construct, using Lemma 4.10, a Clifford circuit C that maps S to $0^{n+\hat{t}} \times \mathbb{F}_2^{n-\hat{t}}$.
Measure the last $n-\hat{t}$ qubits of $24\log(3/\delta)$ copies of $C|\psi\rangle$, and let $\hat{z}$ be the majority outcome.
Set $L \leftarrow 2N_{\hat{t},\epsilon/2,\delta/3} + 8\log(3/\delta)$.
for $i = 1, \dots, L$ **do**
    Measure the last $n-\hat{t}$ qubits of a fresh $C|\psi\rangle$.
    if the outcome is $\hat{z}$ **then**
        Supply the unmeasured first $\hat{t}$ qubits to the residual tomography procedure.
    else
        Discard the remaining qubits.
    end if
end for
if fewer than $N_{\hat{t},\epsilon/2,\delta/3}$ residual states were supplied **then**
    return FAIL.
end if
Run residual tomography with trace-distance accuracy $\epsilon/2$ and failure probability $\delta/3$, obtaining $|\hat{\varphi}\rangle$.
return $|\widehat{\psi}\rangle = C^\dagger|\hat{\varphi}\rangle|\hat{z}\rangle$.

Theorem 4.17 (Heavy-subspace reduction). *Let $|\psi\rangle$ be an n-qubit state. Suppose a subspace $S \subseteq \mathbb{F}_2^{2n}$ is (t,ϵ)-heavy for $|\psi\rangle$. Given a basis of S, Algorithm 5 uses only single-copy measurements and outputs a classical description of $|\widehat{\psi}\rangle$ such that*

$$d_{\mathrm{tr}}\big(\psi, \widehat{\psi}\big) \le \epsilon$$

with probability at least $1-\delta$. It uses

$$2N_{t,\epsilon/2,\delta/3} + 32\log(3/\delta)$$

copies, up to integer ceilings, and

$$M_{t,\epsilon/2,\delta/3} + O\big(n^3 + n^2\big(N_{t,\epsilon/2,\delta/3} + \log(1/\delta)\big)\big)$$

time.

Proof. Set $\hat{t} = n - \dim(S)$, so that $\hat{t} \le t$. Because $\epsilon < 1$, the mass condition in Definition 4.9 and Lemma 4.11 show that S is isotropic. The Clifford compression algorithm therefore constructs the circuit C used above. Applying Lemma 4.15 with $\eta = \epsilon^2/4$ gives a basis string z and a state

$$|\varphi\rangle = \frac{(I \otimes \langle z|)C|\psi\rangle}{\|(I \otimes \langle z|)C|\psi\rangle\|_2}$$

such that

$$d_{\mathrm{tr}}\Big(\psi, C^\dagger|\varphi\rangle|z\rangle\Big) \le \frac{\epsilon}{2}. \tag{4.11}$$

In particular, measuring the final $n - \hat{t}$ qubits of $C|\psi\rangle$ returns z with probability at least $1 - \epsilon^2/4 \ge 3/4$. The multiplicative Chernoff bound in Lemma 2.7 shows that the majority vote fails to identify z with probability at most $\delta/3$.

Condition on identifying z. Each subsequent occurrence of this outcome leaves the first $\hat{t}$ qubits in state $|\varphi\rangle$. Put

$$N := N_{\hat{t},\epsilon/2,\delta/3}, \qquad \ell := \log(3/\delta).$$

For $L = 2N + 8\ell$ postselection trials, the number X of successful postselections has mean $\mu \ge 3L/4 = 3N/2 + 6\ell$. The lower-tail bound in Lemma 2.7 gives

$$\Pr[X < N] \le \exp\left(-\frac{(\mu - N)^2}{2\mu}\right) \le \exp(-\ell) = \frac{\delta}{3},$$

where the exponent is increasing in $\mu \ge N$, and the second inequality follows from

$$\left(\frac{N}{2} + 6\ell\right)^2 \ge 2\ell\left(\frac{3N}{2} + 6\ell\right).$$

Thus the stated number of trials supplies N residual states except with probability at most $\delta/3$. Residual tomography then produces $|\hat{\varphi}\rangle$ with

$$d_{\mathrm{tr}}(\varphi, \hat{\varphi}) \le \frac{\epsilon}{2}$$

except with probability $\delta/3$. Assuming all three events succeed, unitary invariance and the triangle inequality, together with (4.11), give

$$\begin{aligned} d_{\mathrm{tr}}\Big(\psi, C^\dagger|\hat{\varphi}\rangle|\hat{z}\rangle\Big) &\le d_{\mathrm{tr}}\Big(\psi, C^\dagger|\varphi\rangle|z\rangle\Big) + d_{\mathrm{tr}}(|\varphi\rangle|z\rangle, |\hat{\varphi}\rangle|\hat{z}\rangle) \\ &\le \epsilon. \end{aligned}$$

A union bound proves the success probability.

The two sampling stages use the claimed number of copies. Constructing C from a basis of S costs $O(n^3)$ time, and the resulting circuit has $O(n^2)$ gates. Preparing each transformed copy therefore costs $O(n^2)$ time. Residual tomography contributes $M_{t,\epsilon/2,\delta/3}$, yielding the stated running time. □

The important point is that tomography is now exponential only in the stabilizer nullity t, rather than in the total number n of qubits. Indeed, substituting Equation (4.10) into the reduction gives

$$\widetilde{O}\left(\frac{2^t \log(1/\delta)}{\epsilon^2}\right)$$

copies for the residual tomography step. Thus, when $t = O(\log n)$ is small, the tomography step runs in polynomial time.

4.2.4 Tomography Using Bell-Difference Sampling

We first show how Bell-difference sampling can be used to find a heavy subspace. Recall that every Bell-difference sample lies in $\mathrm{Weyl}(|\psi\rangle)^{\perp_s}$ (Proposition 2.21). Thus, if T is the span of several such samples, then

$$T \subseteq \mathrm{Weyl}(|\psi\rangle)^{\perp_s},$$

and hence its symplectic complement $T^{\perp_s}$ has dimension at least the stabilizer dimension of $|\psi\rangle$.

The remaining question is whether $T^{\perp_s}$ is also heavy. If we draw enough samples, their span captures nearly all of the Bell-difference distribution q_ψ. The duality relations from Section 2.7 then convert this large q_ψ-mass into large p_ψ-mass on $T^{\perp_s}$. This gives the following simple algorithm.

Algorithm 6 Heavy-subspace learning from Bell-difference samples

Input: Copies of $|\psi\rangle$ and $\epsilon, \delta \in (0,1)$
Promise: $\dim \mathrm{Weyl}(|\psi\rangle) \geq n - t$
Output: A (t,ϵ)-heavy subspace S for $|\psi\rangle$

1: Draw
$$\frac{8\log(1/\delta) + 16n}{\epsilon^2}$$
independent Bell-difference samples from q_ψ.
2: Let T be the span of the samples.
3: Compute $S = T^{\perp_s}$ using Lemma 2.13.
4: **return** S.

Theorem 4.18 (Heavy subspace from Bell-difference samples). *Let $|\psi\rangle$ have stabilizer dimension at least $n - t$. Algorithm 6 outputs a (t,ϵ)-heavy subspace for $|\psi\rangle$ with probability at least $1 - \delta$. It uses*
$$\frac{32\log(1/\delta) + 64n}{\epsilon^2}$$
copies of $|\psi\rangle$ and
$$O\left(\frac{n^2\log(1/\delta) + n^3}{\epsilon^2}\right)$$
time.

Proof. Let T be the span of the Bell-difference samples and $S = T^{\perp_s}$. By Proposition 2.21, $T \subseteq \mathrm{Weyl}(|\psi\rangle)^{\perp_s}$. Since $\dim \mathrm{Weyl}(|\psi\rangle) \geq n - t$, its symplectic complement has dimension at most $n + t$. Therefore $\dim(S) = 2n - \dim(T) \geq n - t$.

It remains to prove that S is heavy. Applying Lemma 2.23 to q_ψ, with error parameter $\epsilon^2/4$, shows that except with probability δ,
$$\sum_{x \in T} q_\psi(x) \geq 1 - \frac{\epsilon^2}{4}.$$
By Proposition 2.22, the same lower bound holds for the p_ψ-mass of T. Since $T = S^{\perp_s}$, Equation (2.19) then gives
$$\sum_{x \in S} p_\psi(x) = \frac{|S|}{2^n} \sum_{x \in T} p_\psi(x) \geq \left(1 - \frac{\epsilon^2}{4}\right)\frac{|S|}{2^n}.$$
Hence S is (t,ϵ)-heavy.

Each Bell-difference sample consumes four copies, which gives the displayed copy count. Sampling costs $O(n)$ time per outcome, and computing the symplectic complement dominates with

$$O\left(\frac{n^2 \log(1/\delta) + n^3}{\epsilon^2}\right)$$

time. □

Combining this heavy-subspace learner with the reduction from the preceding subsection gives our first tomography algorithm.

Corollary 4.19 (Tomography from Bell-difference sampling). *Let $|\psi\rangle$ have stabilizer dimension at least $n - t$. There is an algorithm that outputs a classical description of $|\widehat{\psi}\rangle$ satisfying*

$$d_{\mathrm{tr}}\left(\psi, \widehat{\psi}\right) \leq \epsilon$$

with probability at least $1 - \delta$. It uses

$$\frac{32 \log(2/\delta) + 64n}{\epsilon^2} + 2N_{t,\epsilon/2,\delta/6} + 32 \log(6/\delta)$$

copies and

$$O\left(\frac{n^2 \log(1/\delta) + n^3}{\epsilon^2} + n^2 N_{t,\epsilon/2,\delta/6}\right) + M_{t,\epsilon/2,\delta/6}$$

time.

Proof. Run Algorithm 6 with failure probability $\delta/2$, then feed its output into Algorithm 5, also with failure probability $\delta/2$. The bounds follow from Theorems 4.17 and 4.18; a union bound gives the claimed success probability. □

4.2.5 Tomography from Single-Copy Measurements

The algorithm based on Bell-difference sampling (Corollary 4.19) is efficient, but it uses entangling measurements across pairs of copies. We now show that this can be avoided altogether. The price is that a single measurement basis reveals less information, so the resulting algorithm will be slower.

Our basic primitive is especially simple: measure two copies separately in the computational basis and take the difference of their outcomes. Recall from Section 2.5 that

$$\mathcal{X} := \mathbb{F}_2^n \times 0^n, \qquad \mathcal{Z} := 0^n \times \mathbb{F}_2^n$$

are the subspaces indexing Pauli strings containing only X's and only Z's, respectively.

Definition 4.20 (Computational difference sampling). Measure one copy of $|\psi\rangle$ in the computational basis to obtain $x \in \mathbb{F}_2^n$, measure a second copy separately to obtain $y \in \mathbb{F}_2^n$, and output $(x + y, 0^n) \in \mathcal{X}$. Let $r_\psi(a)$ denote the probability of output $a \in \mathcal{X}$.

Although a computational difference sample consumes two copies, the copies are never measured jointly. If $|\psi\rangle = \sum_x \alpha_x |x\rangle$, then

$$r_\psi(a, 0^n) = \sum_{x \in \mathbb{F}_2^n} |\alpha_x|^2 |\alpha_{x+a}|^2. \tag{4.12}$$

Thus, r_ψ is simply the self-convolution of the computational-basis measurement distribution.

The usefulness of r_ψ comes from its relation to the characteristic distribution p_ψ. We record a pair of propositions that make this relation explicit.

Proposition 4.21. *For every $a \in \mathbb{F}_2^n$,*

$$\sum_{x \in \mathbb{F}_2^n} |x\rangle\langle x| \otimes |x + a\rangle\langle x + a| = \frac{1}{2^n} \sum_{z \in \mathbb{F}_2^n} (-1)^{a \cdot z} Z^z \otimes Z^z.$$

Proof. Conjugate the second register in Proposition 4.12 by X^a, and use $X^a Z^z X^a = (-1)^{a \cdot z} Z^z$. □

Proposition 4.22. *For every $a \in \mathbb{F}_2^n$,*

$$r_\psi(a, 0^n) = \sum_{z \in \mathbb{F}_2^n} p_\psi(a, z) = \sum_{x \in a \times \mathbb{F}_2^n} p_\psi(x).$$

Proof. Using the preceding proposition,

$$\begin{aligned}\sum_{z\in\mathbb{F}_2^n} p_\psi(a,z) &= \frac{1}{2^n}\sum_z (-1)^{a\cdot z}\langle\psi|X^aZ^z|\psi\rangle^2 \\ &= \sum_x \langle\psi|^{\otimes 2}(X^a\otimes X^a)\big(|x\rangle\langle x|\otimes|x+a\rangle\langle x+a|\big)|\psi\rangle^{\otimes 2} \\ &= \sum_x |\alpha_x|^2|\alpha_{x+a}|^2 = r_\psi(a,0^n),\end{aligned}$$

where the last equality is (4.12). □

Thus, computational difference sampling gives the X-marginal of the characteristic distribution:

$$r_\psi(a,0^n) = \sum_{z\in\mathbb{F}_2^n} p_\psi(a,z).$$

Consequently, concentration of r_ψ on a subspace of $\mathcal{X}$ translates, by symplectic duality, into large p_ψ-mass on a corresponding subspace of the full Pauli space.

Corollary 4.23. *For every subspace $H \subseteq \mathcal{X}$,*

$$\sum_{x\in H} r_\psi(x) = \frac{|H+\mathcal{Z}|}{2^n}\sum_{x\in(H+\mathcal{Z})^{\perp_s}} p_\psi(x).$$

Proof. Proposition 4.22 identifies the left-hand side with the p_ψ-mass of $H+\mathcal{Z}$. Apply Proposition 2.20. □

Computational difference samples cannot point in arbitrary X-directions: their support is determined by the Z-type stabilizers of the state.

Corollary 4.24. *If $x \in \mathcal{X}$ and $r_\psi(x) > 0$, then*

$$x \in \big(\mathrm{Weyl}(|\psi\rangle)\cap\mathcal{Z}\big)^{\perp_s}.$$

Proof. Set

$$H := \big(\mathrm{Weyl}(|\psi\rangle)\cap\mathcal{Z}\big)^{\perp_s}\cap\mathcal{X}.$$

Using Equation (2.11) and Lemma 2.11,

$$(H+\mathcal{Z})^{\perp_s} = \mathrm{Weyl}(|\psi\rangle)\cap\mathcal{Z} \subseteq \mathrm{Weyl}(|\psi\rangle).$$

Every element of the last space has p_ψ-value 2^{-n}. Corollary 4.23 and $|H+\mathcal{Z}||(H+\mathcal{Z})^{\perp_s}| = 4^n$ therefore show that $\sum_{x\in H} r_\psi(x) = 1$. □

This suggests the single-copy algorithm. In any fixed basis, computational difference sampling reveals the stabilizers that are visible as Z-type Pauli operators in that basis. A random Clifford change of basis exposes a different part of the stabilizer group. By repeating this experiment in sufficiently many random Clifford bases and combining the resulting subspaces, we can recover a large heavy subspace.

The number of random bases grows as 2^t: when the stabilizer dimension is $n-t$, a random Z-type Pauli space is less likely to intersect the stabilizer group as t grows.

Algorithm 7 Heavy-subspace learning using single-copy measurements

Input: Copies of $|\psi\rangle$ and $\epsilon, \delta \in (0,1)$
Promise: $\dim \mathrm{Weyl}(|\psi\rangle) \geq n - t$
Output: A (t,ϵ)-heavy subspace S for $|\psi\rangle$

1: Set
$$m_{\mathsf{Clifford}} \leftarrow 2(2^{t+1}+1)\big(n + \log(2/\delta)\big).$$
2: **for** $i = 1, \ldots, m_{\mathsf{Clifford}}$ **do**
3: Sample a uniformly random Clifford circuit C_i, and set $|\psi_i\rangle = C_i|\psi\rangle$.
4: Draw
$$m_{\mathsf{Comp}} := \frac{16n}{\epsilon^2}\big(n + \log(2m_{\mathsf{Clifford}}/\delta)\big)$$
computational difference samples from r_{ψ_i}.
5: Let $\widehat{H}_i \subseteq \mathcal{X}$ be their span, and compute a basis by Gaussian elimination.
6: **end for**
7: Compute
$$S \leftarrow \sum_{i=1}^{m_{\mathsf{Clifford}}} C_i^\dagger\big((\widehat{H}_i + \mathcal{Z})^{\perp_s}\big).$$
8: **return** S.

There are two things to prove about the returned subspace S in order to apply the reduction shown in Theorem 4.17. First, it must have dimension at least $n-t$. Second, it must carry nearly maximal p_ψ-mass. We handle these two properties separately.

The Output Has Large Dimension

We first show that sufficiently many random Clifford bases expose enough Z-type stabilizers to span the entire unsigned stabilizer group. The next lemma quantifies the progress made by a single random basis.

Lemma 4.25. *Let $A \subseteq \mathbb{F}_2^{2n}$ be isotropic of dimension $n-k$, let $T \subsetneq A$, let B be Lagrangian, and let C be a uniformly random Clifford circuit. Then*

$$\Pr\big[C(A \setminus T) \cap B \neq \emptyset\big] \geq \frac{1}{2^{k+1}+1}.$$

Proof. It suffices to consider $\dim(T) = n-k-1$, since a smaller T only increases the event. Write $A = \langle x \rangle + T$. Conditioned on $C(T)$, the vector $C(x)$ is uniform on $C(T^{\perp_s} \setminus T)$. Hence

$$\begin{aligned}\Pr[C(x) \in C(T) + B \mid C(T)] &= \frac{|(C(T)+B) \cap C(T)^{\perp_s}| - |C(T)|}{|C(T)^{\perp_s}| - |C(T)|} \\ &= \frac{2^n - 2^{n-k-1}}{2^{n+k+1} - 2^{n-k-1}} = \frac{1}{2^{k+1}+1}.\end{aligned}$$

The second equality uses Lemma 2.11, since $(C(T)+B) \cap C(T)^{\perp_s}$ is Lagrangian. □

Lemma 4.26. *Suppose $|\psi\rangle$ has stabilizer dimension $n-k$. Let $C_1, \dots, C_m$ be independent uniformly random Clifford circuits and set*

$$S_i := \mathrm{Weyl}(|\psi\rangle) \cap C_i^{\dagger}(\mathcal{Z}).$$

If

$$m \geq 2(2^{k+1}+1)\big(n + \log(1/\delta)\big),$$

then

$$\sum_{i=1}^{m} S_i = \mathrm{Weyl}(|\psi\rangle)$$

with probability at least $1-\delta$.

Proof. Let $T_i = \sum_{j \leq i} S_j$. Until T_i contains $\mathrm{Weyl}(|\psi\rangle)$, the preceding lemma says that the next basis increases $\dim(T_i \cap \mathrm{Weyl}(|\psi\rangle))$ with probability at least

$$p := \frac{1}{2^{k+1}+1}.$$

At most n successful increases are required. Couple the progress indicators from below by independent Bernoulli(p) variables. If $m \geq 2p^{-1}(n + \log(1/\delta))$, Lemma 2.8 gives

$$\Pr[\text{fewer than } n \text{ increases}] \leq \exp\left(n - \frac{mp}{2}\right) \leq \delta.$$

□

The Output Has Large Characteristic Mass

Each individual computational-difference experiment will produce a subspace with nearly maximal characteristic mass. Since the algorithm ultimately takes the sum of many such subspaces, we need to know that this property survives taking sums.

Lemma 4.27. *Suppose $S_1, \dots, S_\ell \subseteq \mathbb{F}_2^{2n}$ satisfy*

$$\sum_{x \in S_i} p_\psi(x) \geq (1 - \eta_i)\frac{|S_i|}{2^n}.$$

Then $S = \sum_i S_i$ satisfies

$$\sum_{x \in S} p_\psi(x) \geq \left(1 - \sum_i \eta_i\right)\frac{|S|}{2^n}.$$

Proof. It is enough to prove the claim for $S = T + S_\ell$ and then induct. Equations (2.11) and (2.19) give

$$\begin{aligned}\sum_{x \in T + S_\ell} p_\psi(x) &= \frac{|T + S_\ell|}{2^n} \sum_{x \in T^{\perp_s} \cap S_\ell^{\perp_s}} p_\psi(x) \\ &\geq \frac{|T + S_\ell|}{2^n}\left(\frac{2^n}{|T|}\sum_{x \in T} p_\psi(x) + \frac{2^n}{|S_\ell|}\sum_{x \in S_\ell} p_\psi(x) - 1\right),\end{aligned}$$

where the inequality is inclusion–exclusion. Substituting the two mass bounds proves the inductive step. □

Lemma 4.28. *Let $S_1, \dots, S_m \subseteq \mathbb{F}_2^{2n}$, and suppose*

$$\sum_{x \in S_i} p_\psi(x) \geq \left(1 - \frac{\eta}{2n}\right)\frac{|S_i|}{2^n}$$

for every i. Then $S = \sum_i S_i$ satisfies

$$\sum_{x \in S} p_\psi(x) \geq (1 - \eta)\frac{|S|}{2^n}.$$

Proof. Some collection of at most $2n$ of the subspaces already spans S, because $S \subseteq \mathbb{F}_2^{2n}$. Apply Lemma 4.27 to that collection. □

Combining this observation with the duality relation for r_ψ gives the mass guarantee needed by the algorithm.

Proposition 4.29. *Let $C_1, \dots, C_m$ be Clifford circuits, set $|\psi_i\rangle = C_i|\psi\rangle$, and let $H_i \subseteq \mathcal{X}$. If*

$$\sum_{x \in H_i} r_{\psi_i}(x) \geq 1 - \frac{\eta}{2n}$$

for every i, then

$$S := \sum_{i=1}^{m} C_i^\dagger\big((H_i + \mathcal{Z})^{\perp_s}\big)$$

satisfies

$$\sum_{x \in S} p_\psi(x) \geq (1-\eta)\frac{|S|}{2^n}.$$

Proof. Let $S_i = C_i^\dagger((H_i + \mathcal{Z})^{\perp_s})$. Clifford invariance, Corollary 4.23, and the complement dimension identity give

$$\begin{aligned}\sum_{x \in S_i} p_\psi(x) &= \sum_{x \in (H_i + \mathcal{Z})^{\perp_s}} p_{\psi_i}(x) \\ &= \frac{2^n}{|H_i + \mathcal{Z}|} \sum_{x \in H_i} r_{\psi_i}(x) \\ &\geq \left(1 - \frac{\eta}{2n}\right)\frac{|S_i|}{2^n}.\end{aligned}$$

Now apply Lemma 4.28. □

Completing the Single-Copy Analysis

We can now combine the dimension and mass arguments.

Theorem 4.30. *Let $|\psi\rangle$ have stabilizer dimension at least $n - t$. Algorithm 7 outputs a (t, ϵ)-heavy subspace for $|\psi\rangle$ with probability at least $1 - \delta$. It uses*

$$O\left(\frac{n\big(n + \log(1/\delta)\big)^2\, 2^t}{\epsilon^2}\right)$$

copies and

$$O\left(\frac{n^3\left(n+\log(1/\delta)\right)^2 2^t}{\epsilon^2}\right)$$

time.

Proof. First consider the dimension condition. Let

$$G_i := \mathrm{Weyl}(|\psi\rangle) \cap C_i^{\dagger}(\mathcal{Z}).$$

Lemma 4.26 and the chosen value of m_{Clifford} show, except with probability $\delta/2$, that $\sum_i G_i = \mathrm{Weyl}(|\psi\rangle)$.

For $|\psi_i\rangle = C_i|\psi\rangle$, define

$$H_i := \left(\mathrm{Weyl}(|\psi_i\rangle) \cap \mathcal{Z}\right)^{\perp_s} \cap \mathcal{X}.$$

Corollary 4.24 guarantees $\widehat{H}_i \subseteq H_i$. Reversing inclusions under symplectic complement and applying Lemma 2.11 gives

$$\begin{aligned}(\widehat{H}_i + \mathcal{Z})^{\perp_s} &\supseteq (H_i + \mathcal{Z})^{\perp_s} \\ &= H_i^{\perp_s} \cap \mathcal{Z} \\ &= \left((\mathrm{Weyl}(|\psi_i\rangle) \cap \mathcal{Z}) + \mathcal{X}\right) \cap \mathcal{Z} \\ &= \mathrm{Weyl}(|\psi_i\rangle) \cap \mathcal{Z} = C_i(G_i).\end{aligned}$$

It follows that the returned subspace obeys

$$S \supseteq \sum_i G_i = \mathrm{Weyl}(|\psi\rangle),$$

and hence $\dim(S) \geq n - t$.

For the mass condition, apply Lemma 2.23 to every r_{ψ_i}, with error $\epsilon^2/(8n)$ and per-round failure probability $\delta/(2m_{\mathsf{Clifford}})$. The chosen m_{Comp} implies, simultaneously for all i,

$$\sum_{x \in \widehat{H}_i} r_{\psi_i}(x) \geq 1 - \frac{\epsilon^2}{8n}$$

except with probability $\delta/2$. Applying Proposition 4.29 with $\eta = \epsilon^2/4$ yields

$$\sum_{x \in S} p_\psi(x) \geq \left(1 - \frac{\epsilon^2}{4}\right)\frac{|S|}{2^n}.$$

A union bound proves correctness.

It remains to count resources. Since $t \leq n$,

$$m_{\mathsf{Comp}} = O\left(\frac{n\big(n + \log(1/\delta)\big)}{\epsilon^2}\right), \qquad m_{\mathsf{Clifford}} = O\big(2^t\big(n + \log(1/\delta)\big)\big).$$

Each computational difference sample consumes two copies, giving the stated copy complexity. Sampling and applying a random Clifford circuit costs $O(n^2)$ per transformed copy. The cost of Gaussian elimination and the final subspace computations is at most $O(m_{\mathsf{Clifford}} m_{\mathsf{Comp}} n^2)$. This is the displayed running time. □

Combining this heavy-subspace learner with the earlier reduction gives a tomography algorithm requiring only single-copy measurements.

Corollary 4.31 (Tomography from single-copy measurements). *Let $|\psi\rangle$ have stabilizer dimension at least $n - t$. Using only single-copy measurements, one can output a classical description of $|\widehat{\psi}\rangle$ satisfying*

$$d_{\mathrm{tr}}\left(\psi, \widehat{\psi}\right) \leq \epsilon$$

with probability at least $1 - \delta$. The algorithm uses

$$O\left(\frac{n\left(n + \log(1/\delta)\right)^2 2^t}{\epsilon^2} + N_{t,\epsilon/2,\delta/6}\right)$$

copies and

$$O\left(\frac{n^3\left(n + \log(1/\delta)\right)^2 2^t}{\epsilon^2} + n^2 N_{t,\epsilon/2,\delta/6}\right) + M_{t,\epsilon/2,\delta/6}$$

time.

Proof. Run Algorithm 7 and then Algorithm 5, assigning failure probability $\delta/2$ to each stage. Combine Theorems 4.17 and 4.30 and use a union bound. □

Finally, if $|\psi\rangle$ is prepared by s non-Clifford gates, then Lemma 4.8 permits $t = 2s$. Both tomography procedures therefore have time and copy complexities polynomial in n, 2^s, $1/\epsilon$, and $\log(1/\delta)$.

Chapter 5: Agnostic Tomography

The algorithms in Chapter 4 rely on exact structural promises: the fermionic input is a Slater determinant, or the unknown state is promised to have nontrivial stabilizer dimension. Such a promise can make tomography efficient, but it is also brittle. If the input lies outside the promised class, the algorithm need not return anything useful, even when the state is very well approximated by a member of that class.

In this chapter, we ask for a more robust guarantee. Given an arbitrary quantum state and a class of candidate states $\mathcal{C}$, our goal is to find a state in $\mathcal{C}$ whose fidelity with the input is nearly as large as possible. The input itself need not belong to $\mathcal{C}$, and we make no assumption about how it was generated or corrupted. We call this task *agnostic tomography*, in analogy with agnostic learning in classical learning theory.

We begin by defining the agnostic tomography model, introduced in joint work with collaborators [GIKL26a]. We then study the problem for stabilizer states. Our main result shows that if a pure state has optimal stabilizer fidelity at least $\cos^2(\pi/8) + \gamma$, then its closest stabilizer state is unique and can be identified exactly in polynomial time. We conclude with some historical context and a discussion of subsequent and ongoing work.

5.1 The Agnostic Tomography Model

Let $\mathcal{C}$ be a class of pure states. For an arbitrary state ρ, define its optimal fidelity with $\mathcal{C}$ by

$$F_{\mathcal{C}}(\rho) := \sup_{|\varphi\rangle \in \mathcal{C}} \langle \varphi | \rho | \varphi \rangle.$$

Thus, $F_{\mathcal{C}}(\rho)$ measures how well ρ can be approximated by a state in $\mathcal{C}$.

Agnostic tomography asks the learner to find a state in $\mathcal{C}$ that approximates ρ nearly as well as the best state in $\mathcal{C}$. We formalize this in the following definition.

Definition 5.1 (Agnostic tomography). Let $\epsilon, \delta \in (0, 1)$. An (ϵ, δ)-*agnostic tomography algorithm* for $\mathcal{C}$ receives copies of an arbitrary unknown state ρ and, with probability at least $1 - \delta$, outputs a succinct classical description of a pure state $|\hat{\varphi}\rangle$ satisfying

$$\langle \hat{\varphi} | \rho | \hat{\varphi} \rangle \geq F_{\mathcal{C}}(\rho) - \epsilon.$$

The algorithm is *proper* if $|\hat{\varphi}\rangle \in \mathcal{C}$.

What counts as a succinct description depends on the class $\mathcal{C}$. For example, it might be a polynomial-size circuit, a stabilizer tableau, or a low-dimensional matrix representation. As in ordinary tomography, our goal is to produce such a classical description rather than to physically prepare the output state.

Structured tomography is a special case. If $\rho = |\psi\rangle\langle\psi|$ for some $|\psi\rangle \in \mathcal{C}$, then $F_{\mathcal{C}}(\rho) = 1$, so an agnostic tomography algorithm must return a state with fidelity at least $1 - \epsilon$ with $|\psi\rangle$. For a general input, the learner instead competes with the best state in $\mathcal{C}$.

Importantly, we make no assumption about why the input lies outside $\mathcal{C}$. In particular, agnostic tomography is more general than learning under a specified noise model, where one assumes that $\rho = \mathcal{N}(|\varphi\rangle\langle\varphi|)$ for some $|\varphi\rangle \in \mathcal{C}$ and a known noise channel $\mathcal{N}$. Here, the input may differ from every state in $\mathcal{C}$ in an arbitrary way.

Even simple classes illustrate why an algorithm for the realizable case need not solve the agnostic problem. Consider the class of n-qubit product states. When the input is promised to be a product state, it can be learned by performing tomography separately on each qubit. For comparison, every one-qubit reduced state of

$$|\mathrm{GHZ}_n\rangle := \frac{|0^n\rangle + |1^n\rangle}{\sqrt{2}}$$

is maximally mixed, even though $|\mathrm{GHZ}_n\rangle$ has fidelity $1/2$ with each of the product states $|0^n\rangle$ and $|1^n\rangle$. The one-qubit marginals therefore miss the global correlations that determine a good product-state approximation. In fact, this problem persists even when the optimal product-state fidelity is arbitrarily close to one [Bak+25, Section 2].

From the standpoint of copy complexity alone, agnostic tomography is often easy. For example, when $\mathcal{C}$ is finite, shadow tomography can estimate the fidelity of every candidate using relatively few copies [Aar20]. The real challenge is to design algorithms that are also computationally efficient.

5.2 High-Fidelity Stabilizer Approximation

We present the simplest (and first) agnostic tomography algorithm. Here, $\mathcal{C} = \mathsf{Stab}_n$ is the class of n-qubit stabilizer states. For a pure state $|\psi\rangle$, write

$$F_{\mathsf{Stab}}(|\psi\rangle) \coloneqq \max_{|\phi\rangle \in \mathsf{Stab}_n} |\langle\phi|\psi\rangle|^2$$

for its stabilizer fidelity.

Our goal is to recover a stabilizer state achieving this maximum when the stabilizer fidelity is sufficiently large. Once $F_{\mathsf{Stab}}(|\psi\rangle)$ exceeds $\cos^2(\pi/8)$ by a constant margin, the optimal stabilizer state is unique, and we can recover it exactly in polynomial time.

Theorem 5.2 (High-fidelity stabilizer approximation). *Let $|\psi\rangle$ be an n-qubit pure state, let $\delta \in (0,1)$, and suppose that*

$$F_{\mathsf{Stab}}(|\psi\rangle) \geq \cos^2(\pi/8) + \gamma$$

for some $\gamma > 0$. There is a proper agnostic tomography algorithm that outputs a classical description of the unique fidelity-maximizing stabilizer state with probability at least $1 - \delta$. It uses

$$O\left(n + \frac{\log(n/\delta)}{\gamma^2}\right)$$

copies of $|\psi\rangle$ and runs in time

$$O\left(n^3 + \frac{n^2 \log(n/\delta) + n \log^2(1/\delta)}{\gamma^2}\right).$$

The algorithm has three steps. First, Bell-difference sampling produces a short list of Weyl operators containing generators of the optimizer's unsigned stabilizer group. Next, squared Pauli expectations let us distinguish those generators from all other sampled Weyl operators. Once the unsigned stabilizer group is known, a final measurement in the corresponding stabilizer basis determines the optimal stabilizer state.

5.2.1 Bell Difference Samples Near an Optimal Stabilizer

Fix a stabilizer state $|\phi^\star\rangle$ maximizing the fidelity with $|\psi\rangle$, and write

$$S^\star := \mathrm{Weyl}(|\phi^\star\rangle), \qquad \tau := |\langle\phi^\star|\psi\rangle|^2 = F_{\mathsf{Stab}}(|\psi\rangle).$$

Our first goal is to show that a small number of Bell-difference samples contains enough elements of $S^\star$ to generate the entire subspace.

We use the characteristic and Bell-difference distributions p_ψ and q_ψ from Section 2.7. The starting point is that high fidelity with $|\phi^\star\rangle$ forces p_ψ to place substantial mass on $S^\star$.

Lemma 5.3. *The characteristic distribution satisfies*

$$\sum_{x\in S^\star} p_\psi(x) \geq \tau^2.$$

Proof. Choose a Clifford circuit C with $C|\phi^\star\rangle = |0^n\rangle$, and write $|\psi'\rangle = C|\psi\rangle$. Clifford conjugation permutes Weyl operators, so

$$\sum_{x\in S^\star} p_\psi(x) = \sum_{z\in\mathbb{F}_2^n} p_{\psi'}(0^n, z).$$

By Cauchy–Schwarz,

$$\begin{aligned}\sum_{z\in\mathbb{F}_2^n} p_{\psi'}(0^n,z) &= \frac{1}{2^n}\sum_{z\in\mathbb{F}_2^n}\langle\psi'|W_{(0^n,z)}|\psi'\rangle^2 \\ &\geq \frac{1}{4^n}\left(\sum_{z\in\mathbb{F}_2^n}\langle\psi'|W_{(0^n,z)}|\psi'\rangle\right)^2 \\ &= |\langle 0^n|\psi'\rangle|^4 = |\langle\phi^\star|\psi\rangle|^4 = \tau^2.\end{aligned}$$

The penultimate equality uses $\sum_z Z^z = 2^n|0^n\rangle\langle 0^n|$. □

It is not enough that q_ψ place noticeable total mass on $S^\star$. To recover all n generators, the samples must continue to reveal new directions even after we have already learned a proper subspace of $S^\star$. The next lemma gives exactly this progress guarantee.

Lemma 5.4 (Progress under Bell-difference sampling). *For every proper subspace $T \subsetneq S^\star$,*

$$\sum_{x\in S^\star\setminus T} q_\psi(x) \geq \frac{2-\sqrt{3}}{2}\,\tau^4.$$

Proof. Enlarge T, if necessary, to a codimension-one subspace of $S^\star$. There is a Clifford circuit taking $|\phi^\star\rangle$ to $|0^n\rangle$, $S^\star$ to $0^n \times \mathbb{F}_2^n$, and T to $0^{n+1} \times \mathbb{F}_2^{n-1}$. Indeed, after mapping $|\phi^\star\rangle$ to $|0^n\rangle$, CNOT gates implement arbitrary invertible linear transformations on its Pauli-Z stabilizers. Both p_ψ and q_ψ are covariant under this Clifford action. We may therefore assume this normal form.

Define

$$c_\psi(x) := 2^{-n/2}\langle\psi|W_x|\psi\rangle, \qquad p_\psi(x) = c_\psi(x)^2,$$

and let

$$\alpha_0 := \langle 0^n|\psi\rangle, \qquad \alpha_1 := \langle 10^{n-1}|\psi\rangle.$$

Choose the global phase so that $\alpha_0 \geq 0$. Since $|0^n\rangle$ maximizes the stabilizer fidelity, $|\alpha_0|^2 = \tau$. Comparing its fidelity with the four stabilizer states

$$|+\rangle|0^{n-1}\rangle, \quad |-\rangle|0^{n-1}\rangle, \quad |\mathrm{i}\rangle|0^{n-1}\rangle, \quad |-\mathrm{i}\rangle|0^{n-1}\rangle$$

shows that the largest of the four fidelities is at least the right-hand side below, since $\max\{|\cos\theta|, |\sin\theta|\} \geq 1/\sqrt{2}$. Hence

$$\alpha_0^2 \geq \frac{1}{2}\Big(\alpha_0^2 + |\alpha_1|^2 + \sqrt{2}\,\alpha_0|\alpha_1|\Big).$$

Solving this quadratic inequality gives

$$|\alpha_1|^2 \leq (2-\sqrt{3})\alpha_0^2.$$

Consequently,

$$\begin{aligned}\sum_{x\in S^\star\setminus T} c_\psi(x) &= 2^{n/2-1}\big(|\alpha_0|^2 - |\alpha_1|^2\big)\\ &\geq 2^{n/2-1}(\sqrt{3}-1)\tau.\end{aligned}$$

The first duality identity in Proposition 2.20, together with Lemma 5.3, gives

$$\sum_{x\in T} p_\psi(x) = \frac{|T|}{2^n}\sum_{x\in T^{\perp_{\mathrm{s}}}} p_\psi(x) \geq \frac{\tau^2}{2},$$

because $|T| = 2^{n-1}$ and $S^\star \subseteq T^{\perp_{\mathrm{s}}}$. A second application of Cauchy–Schwarz gives

$$\begin{aligned}\sum_{x\in S^\star\setminus T} p_\psi(x) &\geq \frac{1}{2^{n-1}}\left(\sum_{x\in S^\star\setminus T}|c_\psi(x)|\right)^2\\ &\geq (2-\sqrt{3})\tau^2.\end{aligned}$$

Finally, adding any $a \in T$ permutes $S^\star \setminus T$, and hence

$$\begin{aligned}\sum_{x \in S^\star \setminus T} q_\psi(x) &\geq \sum_{a \in T} p_\psi(a) \sum_{x \in S^\star \setminus T} p_\psi(x+a) \\ &= \left(\sum_{a \in T} p_\psi(a)\right)\left(\sum_{x \in S^\star \setminus T} p_\psi(x)\right) \\ &\geq \frac{2-\sqrt{3}}{2}\tau^4.\end{aligned}$$

□

The progress bound immediately yields a short generator list.

Lemma 5.5 (Generator-list lemma). *Suppose $F_{\mathsf{Stab}}(|\psi\rangle) \geq \tau$. If*

$$m \geq \frac{8+4\sqrt{3}}{\tau^4}(n + \log(1/\delta)),$$

then, with probability at least $1-\delta$, m independent Bell difference samples contain a complete set of generators for $S^\star$.

Proof. Let T_{i-1} be the span of those among the first $i-1$ samples that belong to $S^\star$. Until $T_{i-1} = S^\star$, call sample i successful if it lies in $S^\star \setminus T_{i-1}$; after the span is complete, call every remaining sample successful. By Lemma 5.4, each conditional success probability is at least

$$c := \frac{2-\sqrt{3}}{2}\tau^4.$$

The number of successes can be coupled from below by Binomial(m, c). Since $S^\star$ has dimension n, failure implies fewer than n successes. Lemma 2.8 therefore gives

$$\Pr[\text{failure}] \leq \exp\left(n - \frac{cm}{2}\right).$$

The stated choice of m, which equals $2(n + \log(1/\delta))/c$, makes this probability at most δ. □

5.2.2 Separating Stabilizers by Pauli Expectations

Bell-difference sampling gives us a short list containing generators of $S^\star$, but it does not tell us which sampled Weyl operators actually belong to $S^\star$. Above the high-fidelity threshold, squared Pauli expectations separate the two cases.

Proposition 5.6. *Suppose that an n-qubit pure state $|\psi\rangle$ has fidelity $\tau \geq 1/2$ with a stabilizer state $|\phi\rangle$. If $x \in \mathrm{Weyl}(|\phi\rangle)$, then*

$$\langle\psi|W_x|\psi\rangle^2 \geq (2\tau - 1)^2.$$

Proof. Write

$$|\psi\rangle = \sqrt{\tau}|\phi\rangle + \sqrt{1-\tau}|\phi^\perp\rangle, \qquad \langle\phi|\phi^\perp\rangle = 0.$$

Choose the sign $P = \pm W_x$ for which $P|\phi\rangle = |\phi\rangle$. The cross terms vanish, so

$$\langle\psi|P|\psi\rangle = \tau + (1-\tau)\langle\phi^\perp|P|\phi^\perp\rangle \geq 2\tau - 1.$$

Because $\tau \geq 1/2$, squaring proves the claim. □

Proposition 5.7. *Under the same assumptions, if $y \notin \mathrm{Weyl}(|\phi\rangle)$, then*

$$\langle\psi|W_y|\psi\rangle^2 \leq 4\tau(1-\tau).$$

Proof. Since $\mathrm{Weyl}(|\phi\rangle)$ is Lagrangian, there exists $x \in \mathrm{Weyl}(|\phi\rangle)$ for which W_x anticommutes with W_y. For any pair of anticommuting Hermitian unitaries A, B and any state,

$$\langle\psi|A|\psi\rangle^2 + \langle\psi|B|\psi\rangle^2 \leq 1.$$

Indeed, every real linear combination $aA+bB$ satisfies $(aA+bB)^2 = (a^2+b^2)I$, and choosing (a,b) parallel to the two expectation values proves the inequality. Apply it to W_x, W_y and use Proposition 5.6:

$$\langle\psi|W_y|\psi\rangle^2 \leq 1 - (2\tau-1)^2 = 4\tau(1-\tau).$$

□

The two ranges are disjoint precisely when $(2\tau - 1)^2 > 4\tau(1-\tau)$, which is equivalent to $\tau > \cos^2(\pi/8)$. This is where the threshold in our theorem comes from.

Corollary 5.8 (Expectation gap). *Suppose*

$$|\langle\phi|\psi\rangle|^2 \geq \cos^2(\pi/8) + \gamma$$

for a stabilizer state $|\phi\rangle$ and some $\gamma > 0$. Then

$$\begin{aligned} x \in \mathrm{Weyl}(|\phi\rangle) \quad &\Longrightarrow \quad \langle\psi|W_x|\psi\rangle^2 \geq \frac{1}{2} + 2\sqrt{2}\gamma + 4\gamma^2, \\ y \notin \mathrm{Weyl}(|\phi\rangle) \quad &\Longrightarrow \quad \langle\psi|W_y|\psi\rangle^2 \leq \frac{1}{2} - 2\sqrt{2}\gamma - 4\gamma^2. \end{aligned}$$

Proof. Substitute $\tau \geq \cos^2(\pi/8) + \gamma$ into Propositions 5.6 and 5.7. □

Corollary 5.9 (Uniqueness above the threshold). *If $F_{\mathsf{Stab}}(|\psi\rangle) > \cos^2(\pi/8)$, then the fidelity-maximizing stabilizer state is unique.*

Proof. Suppose two distinct stabilizer states $|\phi_1\rangle$ and $|\phi_2\rangle$ both have fidelity greater than the threshold. If their unsigned stabilizer groups agree, then they are distinct vectors in the same stabilizer basis and hence orthogonal. Both fidelities exceed $1/2$, which contradicts normalization.

Otherwise, choose $x \in \mathrm{Weyl}(|\phi_1\rangle) \setminus \mathrm{Weyl}(|\phi_2\rangle)$. The strict form of Corollary 5.8 makes $\langle\psi|W_x|\psi\rangle^2$ simultaneously greater than and less than $1/2$, a contradiction. □

The threshold is tight for uniqueness. The one-qubit state

$$\cos(\pi/8)|0\rangle + \sin(\pi/8)|1\rangle$$

has fidelity $\cos^2(\pi/8)$ with both $|0\rangle$ and $|+\rangle$.

5.2.3 Estimating Many Pauli Expectations

The generator list contains only $O(n + \log(1/\delta))$ candidates, and we could estimate their Pauli expectations one at a time. Bell measurements, however, let us estimate all of their squared Pauli expectations simultaneously. The following is a variant of the Pauli estimation routine from [HKP21].

Theorem 5.10 (Simultaneous squared-expectation estimation). *Let $P_1, \dots, P_m$ be n-qubit Weyl operators, and let $|\psi\rangle$ be an unknown pure state. Given $\eta, \delta \in (0,1)$, there is an algorithm that estimates every $\langle\psi|P_i|\psi\rangle^2$ to additive error η, simultaneously with probability at least $1-\delta$, using*

$$2\left\lceil \frac{2\log(2m/\delta)}{\eta^2} \right\rceil$$

copies of $|\psi\rangle$. Its classical processing time is

$$O\left(\frac{nm\log(m/\delta)}{\eta^2}\right).$$

Proof. The observables $W_x \otimes W_x$, for $x \in \mathbb{F}_2^{2n}$, commute: any anticommutation sign between two Weyl operators appears once in each tensor factor and therefore cancels. Their joint eigenbasis is the n-qubit Bell basis. A Bell measurement on two copies consequently produces, for every i, a random sign whose expectation is

$$\langle\psi^{\otimes 2}|P_i \otimes P_i|\psi^{\otimes 2}\rangle = \langle\psi|P_i|\psi\rangle^2.$$

Set

$$N = \left\lceil \frac{2\log(2m/\delta)}{\eta^2} \right\rceil$$

and average these signs over N independent Bell measurements. Hoeffding's inequality (Lemma 2.4) and a union bound over the m observables give the claimed simultaneous accuracy. Each measurement consumes two copies, giving the stated copy bound. Computing the m signs from each outcome takes $O(nm)$ time. □

5.2.4 The Reconstruction Algorithm

We now have all the ingredients. Bell-difference sampling gives a short candidate list, and the expectation gap lets us retain exactly those candidates lying in $S^\star$. Once these elements span $S^\star$, we know the unsigned stabilizer group of the optimum. It remains only to determine the signs, which we do by measuring $|\psi\rangle$ in the corresponding stabilizer basis.

Algorithm 8 High-fidelity stabilizer approximation

Input: Copies of $|\psi\rangle$; parameters $\gamma, \delta > 0$
Promise: $F_{\mathsf{Stab}}(|\psi\rangle) \geq \cos^2(\pi/8) + \gamma$
Output: The unique optimal stabilizer, except with probability δ

1: Set $m \leftarrow \left\lceil \frac{8+4\sqrt{3}}{\cos^8(\pi/8)}(n + \log(3/\delta)) \right\rceil$.
2: Draw m independent Bell-difference samples $x_1, \ldots, x_m \in \mathbb{F}_2^{2n}$.
3: Using Theorem 5.10, estimate every $\langle\psi|W_{x_i}|\psi\rangle^2$ to additive error $2\sqrt{2}\gamma$, with failure probability at most $\delta/3$.
4: Retain the x_i whose estimate is greater than $1/2$, and let S be their linear span.
5: **if** S is not Lagrangian **then**
6: **return** FAIL.
7: **end if**
8: Construct a Clifford circuit C mapping S to $0^n \times \mathbb{F}_2^n$.
9: Measure $C|\psi\rangle$ in the computational basis on $\lceil 4\log(3/\delta) \rceil$ fresh copies, and let $\hat{z}$ be the most frequent outcome.
10: **return** the stabilizer tableau describing $C^\dagger|\hat{z}\rangle$.

Proof of Theorem 5.2. Let $|\phi^\star\rangle$ be a fidelity-maximizing stabilizer state. Set $S^\star = \mathrm{Weyl}(|\phi^\star\rangle)$. This state is unique by Corollary 5.9.

Apply Lemma 5.5 with failure probability $\delta/3$ and $\tau = \cos^2(\pi/8)$. With probability at least $1 - \delta/3$, the Bell-difference samples contain a complete set of generators for $S^\star$.

Condition also on the simultaneous estimates succeeding. For $x_i \in S^\star$, Corollary 5.8 implies that its estimate is at least

$$\frac{1}{2} + 4\gamma^2 > \frac{1}{2}.$$

For $x_i \notin S^\star$, its estimate is at most

$$\frac{1}{2} - 4\gamma^2 < \frac{1}{2}.$$

The threshold step therefore retains exactly the sampled elements of $S^\star$. On the generator-list event they span $S^\star$, so $S = S^\star$ and the algorithm does not return FAIL.

In the stabilizer basis determined by $S^\star$, the outcome corresponding to $|\phi^\star\rangle$ occurs with probability

$$|\langle\phi^\star|\psi\rangle|^2 \geq \cos^2(\pi/8) + \gamma > \frac{1}{2}.$$

For k independent basis measurements, Hoeffding's inequality (Lemma 2.4) gives

$$\Pr[|\phi^\star\rangle \text{ occurs at most } k/2 \text{ times}] \le e^{-k/4}.$$

Thus $k = \lceil 4\log(3/\delta)\rceil$ makes the final failure probability at most $\delta/3$. A union bound over the generator-list, estimation, and basis-measurement events proves correctness.

It remains to account for resources. The m Bell-difference samples use $4m = O(n + \log(1/\delta))$ copies. Applying Theorem 5.10 with $\eta = 2\sqrt{2}\gamma$ uses

$$O\left(\frac{\log(m/\delta)}{\gamma^2}\right)$$

copies, and the final vote uses $O(\log(1/\delta))$. Since $m = O(n + \log(1/\delta))$, the total copy complexity is

$$O\left(n + \frac{\log(n/\delta)}{\gamma^2}\right).$$

Bell-difference sampling takes $O(mn)$ time, and simultaneous estimation takes

$$O\left(\frac{nm\log(m/\delta)}{\gamma^2}\right).$$

Gaussian elimination on the $m \times 2n$ sample matrix, together with construction of the Clifford basis change, takes $O(mn\min\{m,n\})$ time [AG04]. The final basis measurements take $O(n^2\log(1/\delta))$ time. Substituting the value of m and collecting terms gives

$$O\left(n^3 + \frac{n^2\log(n/\delta) + n\log^2(1/\delta)}{\gamma^2}\right),$$

as claimed. □

5.3 Historical Context and Subsequent Work

The high-fidelity stabilizer algorithm of the preceding section was a precursor to the agnostic tomography model: it captured the basic question that eventually led us to formulate the general learning task. At the time, it was unclear whether one could obtain a computationally efficient algorithm without such a high-fidelity assumption.

This question led to subsequent joint work in which we introduced agnostic tomography [GIKL26a]. As our first example, we considered the class of stabilizer product states

$$\mathsf{StabProd}_n := \left\{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |\mathrm{i}\rangle, |-\mathrm{i}\rangle\right\}^{\otimes n}.$$

Unlike the high-fidelity algorithm above, the resulting algorithm works for every value of the optimal fidelity.

Theorem 5.11 (Agnostic tomography of stabilizer product states). *Let $0 < \epsilon < 1$ and $\delta \in (0,1)$. Given copies of an arbitrary n-qubit mixed state ρ, there is a proper agnostic tomography algorithm that outputs $|\hat{\varphi}\rangle \in \mathsf{StabProd}_n$ satisfying*

$$\langle\hat{\varphi}|\rho|\hat{\varphi}\rangle \geq F_{\mathsf{StabProd}_n}(\rho) - \epsilon$$

with probability at least $1-\delta$. Its running time is

$$\frac{n^{O(\log(1/\epsilon))}\log(1/\delta)}{\epsilon^2}.$$

The important point is that the theorem makes no assumption on the optimal fidelity. Thus, unlike the high-fidelity stabilizer algorithm of the preceding section, it applies in all fidelity regimes. For constant ϵ, the running time is polynomial in n; more generally, it is quasipolynomial in n and $1/\epsilon$.

There are two particularly natural ways to generalize stabilizer product states. One can allow entanglement while retaining stabilizer structure, giving the class of arbitrary stabilizer states; or one can retain product structure while allowing arbitrary single-qubit states, giving the class of all product states. Within months of our work, efficient agnostic tomography algorithms were obtained for both classes.

Chen, Gong, Ye, and Zhang [CGYZ25] gave an agnostic tomography algorithm for arbitrary stabilizer states, building on the Bell-difference sampling machinery from the preceding subsection together with one ingenious new idea: when the target stabilizer state cannot yet be recovered, they find a stabilizer projector whose successful postselection boosts the target fidelity by a constant factor, and then iterate. Independently, Bakshi et

al. [Bak+25] gave an agnostic tomography algorithm for arbitrary product states using entirely different techniques.

The model has continued to develop in several directions. It has been extended from states to quantum channels [WLKD25], and to mixed-state ansatz classes via a connection to robust statistics [ADKL25]. A number of subsequent works have further developed the agnostic tomography framework and extended it to other settings [ADGD26; AD26; QXFL26; DJJMP26], and agnostic tomography algorithms have also inspired new classical algorithms [BC26; ACDG26].

Chapter 6: Property Testing Quantum States

The previous chapters asked for a classical description of an unknown quantum state, either exactly or relative to some structured class. In this chapter, we ask for much less. Rather than learning the state, we only want to decide whether it has a given property.

First, we consider testing stabilizer dimension (Definition 2.16), which measures the number of independent Pauli operators that stabilize a pure state. We present a testing algorithm for this property that accepts states with high stabilizer dimension. As a consequence, states prepared by Clifford circuits with too few non-Clifford gates cannot be pseudorandom.

We then turn to low-degree phase states. Classically, testing whether a function is a low-degree polynomial requires only a small number of queries for every fixed degree. The corresponding quantum problem is much harder: testing whether a phase state comes from a low-degree polynomial requires nearly as many copies as learning the state itself.

6.1 The Property Testing Model

Let $\mathcal{P}$ be a set of pure states and let $\Delta(|\psi\rangle, |\phi\rangle)$ be a notion of distance. A state $|\psi\rangle$ is ϵ-far from $\mathcal{P}$ if

$$\inf_{|\phi\rangle \in \mathcal{P}} \Delta(|\psi\rangle, |\phi\rangle) \geq \epsilon.$$

An ϵ-tester using t copies is a two-outcome measurement $\{M, I - M\}$ on t copies of the input such that

$$\begin{aligned} |\psi\rangle \in \mathcal{P} \quad &\implies \quad \operatorname{tr}\left(M |\psi\rangle\langle\psi|^{\otimes t}\right) \geq \frac{2}{3}, \\ |\psi\rangle \text{ is } \epsilon\text{-far from } \mathcal{P} \quad &\implies \quad \operatorname{tr}\left(M |\psi\rangle\langle\psi|^{\otimes t}\right) \leq \frac{1}{3}. \end{aligned}$$

The constants $2/3$ and $1/3$ are conventional and can be amplified by repetition. A tester has *perfect completeness* if it accepts every state in $\mathcal{P}$ with probability one.

The choice of distance is part of the testing problem. Throughout this chapter, distance from a property is measured by infidelity,

$$1 - F(\psi, \phi),$$

using the fidelity convention from Definition 2.3.

6.2 Testing Stabilizer Dimension

We begin by testing stabilizer dimension. Recall that an n-qubit pure state has stabilizer dimension k if it is a common $+1$-eigenvector of an abelian group of 2^k Pauli operators. In Section 4.2, high stabilizer dimension was used as a promise that made tomography easier. Here we ask the question of whether we can design an algorithm to detect if such symmetries exist at all.

For $k \in \{1, \ldots, n\}$, let

$$\mathcal{Q}_k := \{|\phi\rangle \,:\, \dim \mathrm{Weyl}(|\phi\rangle) \geq k\}.$$

Our goal is to distinguish $|\psi\rangle \in \mathcal{Q}_k$ from the case that

$$F_{\mathcal{Q}_k}(|\psi\rangle) := \max_{|\phi\rangle \in \mathcal{Q}_k} F(\psi, \phi) \leq 1 - \epsilon.$$

The tester is based on a simple observation about Bell-difference sampling. If $|\psi\rangle$ has stabilizer dimension at least k, then every Bell-difference sample lies in $\mathrm{Weyl}(|\psi\rangle)^{\perp_s}$, a subspace of dimension at most $2n - k$. Thus, the samples can never span more than $2n - k$ dimensions.

The converse is the interesting direction. If many Bell-difference samples span a subspace T of dimension at most $2n - k$, then its symplectic complement $S = T^{\perp_s}$ has dimension at least k. We will show that, once the samples capture enough of the Bell-difference distribution, S must be an approximate stabilizer subspace. The compression machinery from Section 4.2.2 then implies that $|\psi\rangle$ is close to a state in $\mathcal{Q}_k$. This gives the desired tester.

The only additional structural fact we need is that sufficiently large Bell-difference mass on $S^{\perp_s}$ forces S to be isotropic.

Lemma 6.1 (Heavy Weyl mass implies isotropicity). *Let $S \subseteq \mathbb{F}_2^{2n}$ be a subspace. If*

$$\sum_{x \in S^{\perp_s}} q_\psi(x) > \frac{5}{8},$$

then S is isotropic.

Proof. Set

$$A \coloneqq \{x \in S \,:\, 2^n p_\psi(x) > 1/2\}.$$

By Fact 2.18, all elements of A commute, so $\operatorname{span}(A)$ is isotropic. It therefore suffices to show that A spans S.

Suppose instead that $|A| \le |S|/2$. The Weyl duality identity in Proposition 2.20 gives

$$\begin{aligned}\sum_{x \in S^{\perp_s}} q_\psi(x) &= |S^{\perp_s}| \sum_{x \in S} p_\psi(x)^2 \\ &\le |S^{\perp_s}| \left(\frac{|S|}{2} \frac{1}{4^n} + \frac{|S|}{2} \frac{1}{4 \cdot 4^n} \right) \\ &= \frac{5}{8},\end{aligned}$$

where the last equality uses $|S||S^{\perp_s}| = 4^n$. This contradicts the hypothesis. Hence $|A| > |S|/2$. No proper subspace of S contains more than half its elements, so A spans S, proving that S is isotropic. □

We can now state the tester.

Algorithm 9 Testing stabilizer dimension

Input: Copies of an n-qubit pure state $|\psi\rangle$; $k \in \{1, \dots, n\}$; $\epsilon \in (0, 3/8)$; $\delta \in (0, 1)$
Promise: Either $|\psi\rangle \in \mathcal{Q}_k$, or $F_{\mathcal{Q}_k}(|\psi\rangle) \le 1 - \epsilon$
Output: Accept in the first case and reject in the second, except with probability at most δ

1: $\eta \leftarrow \epsilon/2$
2: $m \leftarrow \lceil (4n + 2\log(1/\delta))/\eta \rceil$
3: Draw independent Bell-difference samples $x_1, \dots, x_m \sim q_\psi$
4: $T \leftarrow \operatorname{span}\{x_1, \dots, x_m\}$
5: $S \leftarrow T^{\perp_s}$ and $\hat{k} \leftarrow \dim S$
6: **return** ACCEPT if $\hat{k} \ge k$, and REJECT otherwise

Theorem 6.2 (Stabilizer-dimension testing). *Let $|\psi\rangle$ be an n-qubit pure state, let $k \in \{1, \dots, n\}$, and let $\epsilon \in (0, 3/8)$ and $\delta \in (0, 1)$. Under the promise that either*

$$|\psi\rangle \in \mathcal{Q}_k \qquad \text{or} \qquad F_{\mathcal{Q}_k}(|\psi\rangle) \le 1 - \epsilon,$$

Algorithm 9 distinguishes the two cases with probability at least $1 - \delta$. It has perfect completeness, uses

$$O\left(\frac{n + \log(1/\delta)}{\epsilon} \right)$$

copies of $|\psi\rangle$, and runs in

$$O\left(\frac{n^3 + n^2 \log(1/\delta)}{\epsilon}\right)$$

time.

Proof. First suppose that $|\psi\rangle \in \mathcal{Q}_k$. The support property in Proposition 2.21 confines every sample to $\mathrm{Weyl}(|\psi\rangle)^{\perp_s}$. This subspace has dimension at most $2n - k$. Consequently,

$$\dim T \le 2n - k \qquad \text{and} \qquad \hat{k} = \dim T^{\perp_s} \ge k$$

for every possible sequence of samples. The algorithm therefore accepts with probability one.

Now suppose that $F_{\mathcal{Q}_k}(|\psi\rangle) \le 1 - \epsilon$. Apply Lemma 2.23 to q_ψ on the ambient space $\mathbb{F}_2^{2n}$. The choice of m ensures that, with probability at least $1 - \delta$,

$$\sum_{x \in T} q_\psi(x) \ge 1 - \eta.$$

Condition on this event. Since $T = S^{\perp_s}$, Propositions 2.20 and 2.22 yield

$$\sum_{x \in T} p_\psi(x) \ge 1 - \eta,$$
$$\sum_{x \in S} p_\psi(x) = \frac{|S|}{2^n} \sum_{x \in S^{\perp_s}} p_\psi(x) \ge (1 - \eta)\frac{|S|}{2^n}.$$

Moreover, $1 - \eta > 5/8$, so Lemma 6.1 shows that S is isotropic. In particular, $\hat{k} = \dim S \le n$. Writing $t = n - \hat{k}$, the last display becomes

$$\sum_{x \in S} p_\psi(x) \ge \frac{1 - \eta}{2^t}.$$

The robust product-state structure from Lemma 4.15 therefore produces a state $|\widehat{\psi}\rangle$ with stabilizer dimension at least $\hat{k}$ and

$$F\left(\psi, \widehat{\psi}\right) \ge 1 - \eta = 1 - \frac{\epsilon}{2} > 1 - \epsilon.$$

If $\hat{k} \ge k$, then $|\widehat{\psi}\rangle \in \mathcal{Q}_k$, contradicting the no-instance promise. Thus $\hat{k} < k$, and the algorithm rejects on the conditioned event.

Finally, each of the m Bell-difference samples consumes four copies. Computing $T^{\perp_s}$ by Lemma 2.13 takes $O(mn^2)$ time and dominates the remaining work. Substituting the value of m gives the stated bounds. □

The tester is efficient for every threshold k. This contrasts with the tomography algorithms of Section 4.2, whose cost grows exponentially with the stabilizer nullity. For testing, the dimension of the Bell-difference span is already enough: there is no need to reconstruct the residual state.

6.3 Non-Clifford Complexity of Pseudorandom States

The stabilizer-dimension tester has an immediate application to quantum pseudorandomness. A pseudorandom state must be indistinguishable from a Haar-random state by every efficient quantum algorithm, even when the algorithm receives polynomially many copies. States prepared using too few non-Clifford gates, however, have large stabilizer dimension. As a result, their Bell-difference samples are confined to a proper subspace of $\mathbb{F}_2^{2n}$, giving an efficient way to distinguish them from Haar-random states.

This yields linear lower bounds on the non-Clifford resources needed to prepare pseudorandom states. In particular, any preparation using Clifford gates and arbitrary single-qubit non-Clifford gates requires at least $\lceil n/2 \rceil$ non-Clifford gates, while a Clifford+T preparation requires at least n T-gates. We prove these bounds below.

6.3.1 Pseudorandom States and Non-Clifford Gates

We first recall the definition of pseudorandom quantum states and relate the number of non-Clifford gates in a circuit to the stabilizer dimension of its output. Let κ denote the security parameter and $n = n(\kappa)$ the number of output qubits.

Definition 6.3 (Pseudorandom quantum states [JLS18]). A keyed family $\{|\phi_z\rangle : z \in \{0,1\}^\kappa\}$ of n-qubit pure states is *pseudorandom* if it satisfies the following conditions.

(i) There is a polynomial-time quantum algorithm that, on input z, prepares $|\phi_z\rangle$.

(ii) For every quantum algorithm $\mathcal{A}$ that runs in $\mathrm{poly}(\kappa)$ time and every $r = \mathrm{poly}(\kappa)$,

$$\Big| \Pr_{z\sim\{0,1\}^\kappa}\big[\mathcal{A}(|\phi_z\rangle^{\otimes r}) = 1\big] - \Pr_{|\psi\rangle\sim\mu_{\mathrm{Haar},n}}\big[\mathcal{A}(|\psi\rangle^{\otimes r}) = 1\big]\Big| = \mathrm{negl}(\kappa),$$

To violate pseudorandomness, we need an efficiently observable feature that is shared by all states produced with few non-Clifford gates but is absent from a typical Haar-random state. Stabilizer dimension provides exactly such a feature.

Recall from Lemma 4.8 that the output of an s-doped Clifford circuit has stabilizer dimension at least $n - 2s$. For diagonal non-Clifford gates, we can improve the loss.

Lemma 6.4. *Suppose that $|\psi\rangle$ is prepared from $|0^n\rangle$ by Clifford gates and at most s diagonal single-qubit gates. Then*

$$\dim \mathrm{Weyl}(|\psi\rangle) \geq n - s.$$

Proof. Proceed by induction on the number of diagonal gates. The initial stabilizer state has stabilizer dimension n, and Clifford gates preserve the dimension. Consider a diagonal gate D acting on one qubit. It commutes with every Pauli operator whose local factor on that qubit is I or Z. Within any Pauli subgroup, the operators with one of these two local factors form a subgroup of index at most two. Thus at least half of the current unsigned stabilizer group remains after applying D, so its dimension decreases by at most one. Induction proves the claim. □

Bell-difference sampling now turns this stabilizer structure into something observable. Because q_ψ is supported on the symplectic complement of the unsigned stabilizer group, high stabilizer dimension forces all Bell-difference samples to lie in a low-dimensional subspace.

Corollary 6.5 (Low-dimensional Weyl support). *Let $|\psi\rangle$ be the output of an s-doped Clifford state-preparation circuit. Then*

$$\mathrm{supp}(q_\psi) \subseteq \mathrm{Weyl}(|\psi\rangle)^{\perp_s}, \qquad \dim\big(\mathrm{Weyl}(|\psi\rangle)^{\perp_s}\big) \leq n + 2s.$$

If all the non-Clifford gates are diagonal, the second bound improves to

$$\dim\big(\mathrm{Weyl}(|\psi\rangle)^{\perp_s}\big) \le n + s.$$

Proof. The support containment is Proposition 2.21. The two dimension bounds follow from $\dim(S) + \dim(S^{\perp_s}) = 2n$, together with Lemma 4.8 and Lemma 6.4, respectively. □

In particular, if $s < n/2$, then all Bell-difference samples lie in a proper subspace of $\mathbb{F}_2^{2n}$. For diagonal gates the same conclusion holds whenever $s < n$. We next show that Haar-random states behave in the opposite way.

6.3.2 Bell Difference Sampling of Haar-Random States

We have shown that, for states prepared using too few non-Clifford gates, Bell-difference samples are confined to a proper subspace. For a Haar-random state, the opposite happens: with high probability, $O(n)$ Bell-difference samples span all of $\mathbb{F}_2^{2n}$.

To prove this, we first show that no proper subspace can capture too much Bell-difference probability. The needed bound follows from concentration of the Weyl expectation values of a Haar-random state.

A consequence of Lévy's lemma proved in [GIKL23] is that, for every $u > 0$,

$$\Pr_{|\psi\rangle\sim\mu_{\mathrm{Haar},n}}\left[\max_{x\neq 0}|\langle\psi|W_x|\psi\rangle| \ge u\right] \le 2^{2n+1}\exp\left(-\frac{2^n u^2}{36\pi^3}\right). \tag{6.1}$$

For later use, set

$$\beta_n := 2^{2n+1}\exp\left(-\frac{2^n}{36\sqrt{3}\,\pi^3}\right). \tag{6.2}$$

In particular, $\beta_n = \exp(-\Theta(2^n))$.

Lemma 6.6 (Haar anticoncentration on subspaces). *Except for a set of Haar measure at most β_n, an n-qubit pure state $|\psi\rangle$ satisfies*

$$\sum_{x\in T} q_\psi(x) \le \frac{2}{3}$$

simultaneously for every proper subspace $T \subsetneq \mathbb{F}_2^{2n}$.

Proof. It is enough to prove the claim for hyperplanes, since every proper subspace is contained in one. Fix a hyperplane T. Its symplectic complement is spanned by a nonzero vector y. Weyl duality from Proposition 2.20 gives

$$\sum_{x\in T} q_\psi(x) = 2^{2n-1} \sum_{x\in T^{\perp_s}} p_\psi(x)^2$$
$$= \frac{1 + |\langle\psi|W_y|\psi\rangle|^4}{2}.$$

Thus this mass is at most $2/3$ whenever $|\langle\psi|W_y|\psi\rangle| \le 3^{-1/4}$. Applying (6.1) with $u = 3^{-1/4}$ proves the simultaneous statement and gives precisely the exception probability in (6.2). □

Conditioned on this event, whenever the samples currently span a proper subspace, the next Bell-difference sample escapes that subspace with probability at least $1/3$. Thus $O(n + \log(1/\delta))$ samples suffice to span all $2n$ dimensions.

Lemma 6.7 (Spanning under Haar measure). *Let $\delta \in (0,1)$, and set*

$$m := \lceil 12n + 24\log(2/\delta)\rceil.$$

If $|\psi\rangle$ is Haar-random and $x_1,\dots,x_m$ are independent samples from q_ψ, then

$$\Pr\big[\mathrm{span}\{x_1,\dots,x_m\} \neq \mathbb{F}_2^{2n}\big] \le \frac{\delta}{2} + \beta_n,$$

where the probability is over both the Haar-random state and the samples.

Proof. Condition on the event in Lemma 6.6. Let X_i indicate that x_i enlarges the span of the preceding samples, with $X_i = 1$ by convention if the span is already all of $\mathbb{F}_2^{2n}$. For every history,

$$\Pr[X_i = 1 \mid X_1,\dots,X_{i-1}] \ge \frac{1}{3}.$$

Consequently, $\sum_i X_i$ stochastically dominates a binomial random variable Y with parameters m and $1/3$.

Write $L = \log(2/\delta)$ and $\mu = \mathbb{E}[Y] = m/3$. The choice of m gives $\mu \geq 4n + 8L$, so $2n \leq \mu/2$. The multiplicative Chernoff bound in Lemma 2.7 yields

$$\begin{aligned}\Pr[Y < 2n] &\leq \Pr[Y \leq \mu/2] \leq \exp(-\mu/8) \\ &\leq \exp(-n/2 - L) \leq \frac{\delta}{2}.\end{aligned}$$

Fewer than $2n$ span increases occur exactly when the samples fail to span $\mathbb{F}_2^{2n}$. Removing the conditioning adds at most β_n to the failure probability. □

6.3.3 From Distinguishing to Non-Clifford Lower Bounds

We can now put the two sides together. For a state prepared using too few non-Clifford gates, Bell-difference samples are trapped in a proper subspace. For a Haar-random state, the samples span the full Pauli space with high probability. The distinguisher therefore only needs to sample repeatedly and compute the dimension of their span.

Algorithm 10 Few-non-Clifford states versus Haar

Input: Copies of an n-qubit pure state $|\psi\rangle$ and $\delta \in (0, 1)$
Promise: Either $|\psi\rangle$ is Haar-random, or it is the output of an s-doped Clifford state-preparation circuit with $s < n/2$
Output: Return HAAR in the first case and FEW in the second

1: $m \leftarrow \lceil 12n + 24 \log(2/\delta) \rceil$
2: Draw independent Bell-difference samples $x_1, \ldots, x_m \sim q_\psi$
3: $d \leftarrow \dim \operatorname{span}\{x_1, \ldots, x_m\}$
4: **return** HAAR if $d = 2n$, and FEW otherwise

Theorem 6.8 (Few-non-Clifford versus Haar). *Let $\delta \in (0, 1)$ satisfy $\beta_n \leq \delta/2$. Algorithm 10 distinguishes an n-qubit Haar-random state from the output of any s-doped Clifford state-preparation circuit with $s < n/2$, with success probability at least $1 - \delta$. It uses*

$$O\big(n + \log(1/\delta)\big)$$

copies and

$$O\big(n^3 + n^2 \log(1/\delta)\big)$$

time. On the circuit-generated input, its conclusion is deterministic.

Proof. For a circuit-generated input with $s < n/2$, Corollary 6.5 confines all samples to a subspace of dimension at most $n + 2s < 2n$. Their span can therefore never equal $\mathbb{F}_2^{2n}$, and the algorithm always returns FEW.

For a Haar-random input, Lemma 6.7 shows that the samples span the whole space except with probability at most $\delta/2 + \beta_n \leq \delta$. The algorithm then returns HAAR. Each sample consumes four copies of the input state. Computing the rank of an $m \times 2n$ binary matrix by Gaussian elimination takes $O(mn^2)$ time, which gives the stated bounds. □

Any efficiently generated pseudorandom-state family must evade this distinguisher, which immediately forces a linear number of non-Clifford gates.

Corollary 6.9 (Linear non-Clifford count). *Suppose a family of $n = n(\kappa)$-qubit pseudorandom states is generated by $s = s(\kappa)$-doped Clifford state-preparation circuits, where $n(\kappa) \to \infty$. Then, for all sufficiently large security parameters,*

$$s \geq \left\lceil \frac{n}{2} \right\rceil.$$

Proof. If $s < n/2$ for infinitely many security parameters, run Algorithm 10 with $\delta = 1/3$. For all sufficiently large such parameters, $\beta_n \leq 1/6$, so the algorithm labels a Haar-random state correctly with probability at least $2/3$, while it labels every state in the keyed ensemble as FEW with probability one. Its time and copy complexities are polynomial in n, hence polynomial in κ because efficient generation implies $n \leq \mathrm{poly}(\kappa)$. This gives a constant distinguishing advantage for infinitely many security parameters, contradicting Definition 6.3. □

Corollary 6.10 (Linear T-count). *Any family of Clifford+T circuits that prepares an ensemble of $n = n(\kappa)$-qubit pseudorandom states, where $n(\kappa) \to \infty$, must use at least n T-gates for all sufficiently large security parameters.*

Proof. The T-gate is diagonal. If preparing circuits used $s < n$ such gates for infinitely many security parameters, the diagonal refinement in Corollary 6.5 would confine $\mathrm{supp}(q_\psi)$ to a subspace of dimension at most $n + s < 2n$. The same distinguisher would separate every state in the ensemble from a Haar-random state with constant advantage for infinitely many security parameters, contradicting pseudorandomness. □

6.4 Lower Bounds for Testing Low-Degree Phase States

Classical low-degree testing is famously efficient: given query access to a function, one can test whether it is a low-degree polynomial using only a small number of queries. We ask what happens when the values of the polynomial are instead encoded in the phases of a quantum state.

Let q be prime, let $\omega_q = e^{2\pi \mathrm{i}/q}$, and let $f : \mathbb{F}_q^m \to \mathbb{F}_q$ be a polynomial of degree at most d. The corresponding *phase state* is

$$|\psi_f\rangle := \frac{1}{q^{m/2}} \sum_{x \in \mathbb{F}_q^m} \omega_q^{f(x)} |x\rangle.$$

We consider the problem of testing whether an unknown pure state is a degree-d phase state.

We prove that, in contrast with the classical setting, this problem can require many copies. For every fixed prime q and fixed degree d, our main lower bound is

$$\Omega\Big(m^{\lfloor (d-1)/2 \rfloor}\Big)$$

copies. The quadratic case requires a separate argument, which gives the lower bound $m - 2$.

The proof of the general result passes through coding theory. We first associate a family of phase states with an arbitrary linear code and show that these states are hard to distinguish from Haar-random states when the dual code can reliably decode a certain random-error distribution. We then apply this framework to Reed–Muller codes, which correspond precisely to low-degree phase states. The quadratic case behaves differently and is treated separately at the end of the section.

6.4.1 Codeword Phase States

Let $C \leq \mathbb{F}_q^n$ be a linear code. For $c \in C$, define

$$|\psi_c\rangle := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_q^{c_i} |i\rangle, \qquad \mathcal{P}_C := \{|\psi_c\rangle \,:\, c \in C\}.$$

These are the *C-codeword phase states*. The connection with low-degree phase states comes from the q-ary Reed–Muller code $\mathrm{RM}_q[m,d]$, obtained by evaluating every m-variate polynomial of degree at most d at all q^m points of $\mathbb{F}_q^m$. We always identify polynomials that induce the same function, or equivalently work in

$$\mathbb{F}_q[x_1,\dots,x_m]/(x_1^q-x_1,\dots,x_m^q-x_m).$$

With the computational basis indexed by $\mathbb{F}_q^m$, the $\mathrm{RM}_q[m,d]$-codeword phase states are precisely the degree-d phase states. We use

$$D_{d,q}(r) \coloneqq \left|\{\alpha\in\{0,\dots,q-1\}^r \,:\, \alpha_1+\cdots+\alpha_r\le d\}\right|$$

for the number of degree-at-most-d monomials in r variables. In particular, $D_{d,q}(r) = \dim \mathrm{RM}_q[r,d]$. We will also use the standard duality relation

$$\mathrm{RM}_q[m,d]^\perp = \mathrm{RM}_q[m,m(q-1)-d-1], \qquad 0\le d<m(q-1).$$

The random-error distribution relevant to the quantum problem is the following. Sample $a_1,\dots,a_t$ independently and uniformly from $[n]$, and set

$$e_t \coloneqq \delta_{a_1}+\cdots+\delta_{a_t}\in\mathbb{F}_q^n, \tag{6.3}$$

where δ_i is the ith standard basis vector. We denote this distribution by $\mathcal{E}_{n,q}^t$. Repeated locations are allowed, and the sum is taken in $\mathbb{F}_q$.

Definition 6.11 (MAP decoding probability). Let $V\le\mathbb{F}_q^n$ have parity-check matrix P. The quantity $p_{\mathrm{decode}}^t(V)$ is the largest probability with which a decoder, given Pe_t for $e_t\sim\mathcal{E}_{n,q}^t$, can recover e_t. Equivalently,

$$p_{\mathrm{decode}}^t(V)=\sum_s \max_{x\,:\,Px=s}\Pr[e_t=x].$$

The formula is achieved by the maximum a posteriori, or MAP, decoder. Our first goal is to relate this classical decoding probability to quantum indistinguishability. For $t\ge 1$, define the average moment state

$$\rho_C^t \coloneqq \mathop{\mathbb{E}}_{c\sim C}\left[|\psi_c\rangle\langle\psi_c|^{\otimes t}\right], \qquad \rho_{\mathrm{all}}^t\coloneqq\rho_{\mathbb{F}_q^n}^t. \tag{6.4}$$

We next develop a block decomposition of these states. Write $|+_n\rangle = n^{-1/2}\sum_{i=1}^{n}|i\rangle$, and, for $v \in \mathbb{F}_q^n$, let Z^v be the diagonal unitary with $(Z^v)_{i,i} = \omega_q^{v_i}$. For a linear code $V \le \mathbb{F}_q^n$, define the dephasing channel

$$\mathcal{Z}_V^t(\sigma) := \mathop{\mathbb{E}}_{v\sim V}\left[(Z^v)^{\otimes t}\sigma(Z^{-v})^{\otimes t}\right].$$

By construction,

$$\rho_V^t = \mathcal{Z}_V^t\left(|+_n\rangle\langle+_n|^{\otimes t}\right). \tag{6.5}$$

Let $M \in \mathbb{F}_q^{\ell\times n}$ be a generator matrix of V. For $s \in \mathbb{F}_q^{\ell}$, set

$$T_s^V := \left\{(i_1,\dots,i_t) \in [n]^t \;:\; M\sum_{j=1}^{t}\delta_{i_j} = s\right\},$$

and let Π_s^V project onto the span of the computational basis vectors indexed by T_s^V.

Lemma 6.12 (Dephasing as a projective measurement). *For every state σ on $(\mathbb{C}^n)^{\otimes t}$,*

$$\mathcal{Z}_V^t(\sigma) = \sum_{s\in\mathbb{F}_q^{\ell}} \Pi_s^V \sigma \Pi_s^V.$$

Proof. For $a = (a_1,\dots,a_t)$ and $b = (b_1,\dots,b_t)$ in $[n]^t$, character orthogonality gives

$$\begin{aligned}\langle a|\mathcal{Z}_V^t(\sigma)|b\rangle &= \sigma_{a,b}\mathop{\mathbb{E}}_{v\sim V}\left[\omega_q^{\left\langle\sum_j\delta_{a_j}-\sum_j\delta_{b_j},v\right\rangle}\right]\\ &= \sigma_{a,b}\,\mathbf{1}\left\{M\sum_j\delta_{a_j} = M\sum_j\delta_{b_j}\right\}.\end{aligned}$$

These are exactly the matrix entries of the claimed projective channel. □

Lemma 6.13 (Projection probabilities). *For every $s \in \mathbb{F}_q^{\ell}$,*

$$\left\|\Pi_s^V|+_n\rangle^{\otimes t}\right\|_2^2 = \Pr[Me_t = s].$$

Proof. The left-hand side is $n^{-t}|T_s^V|$, which is precisely the probability on the right-hand side under Eq. (6.3). □

For a distribution $\mathcal{D}$ on a finite set, write

$$|\sqrt{\mathcal{D}}\rangle := \sum_x \sqrt{\mathcal{D}(x)}|x\rangle, \qquad \operatorname{diag}(\mathcal{D}) := \sum_x \mathcal{D}(x)|x\rangle\langle x|,$$

and define

$$v(\mathcal{D}) := \frac{1}{2}\left\| |\sqrt{\mathcal{D}}\rangle\langle\sqrt{\mathcal{D}}| - \operatorname{diag}(\mathcal{D}) \right\|_1.$$

Lemma 6.14. *Every finite distribution $\mathcal{D}$ satisfies*

$$v(\mathcal{D}) \le \sqrt{1 - \sum_x \mathcal{D}(x)^2}.$$

Proof. Let $A = |\sqrt{\mathcal{D}}\rangle\langle\sqrt{\mathcal{D}}| - \operatorname{diag}(\mathcal{D})$. As a rank-one positive semidefinite matrix minus a positive semidefinite matrix, A has at most one positive eigenvalue. Since $\operatorname{tr}(A) = 0$, this eigenvalue equals $\frac{1}{2}\|A\|_1 = v(\mathcal{D})$. Consequently,

$$v(\mathcal{D})^2 \le \operatorname{tr}(A^2) = 1 - \sum_x \mathcal{D}(x)^2.$$

□

6.4.2 Decoding and Indistinguishability

We can now state the main connection between coding theory and the quantum problem: if the dual code can decode these random errors with high probability, then few copies of a random codeword phase state reveal almost no information that distinguishes it from a completely random phase state.

Theorem 6.15 (Decodability implies indistinguishability). *For every linear code $C \le \mathbb{F}_q^n$,*

$$d_{\mathrm{tr}}\left(\rho_C^t, \rho_{\mathrm{all}}^t\right) \le \sqrt{1 - p_{\mathrm{decode}}^t(C^\perp)^2}.$$

Proof. Let $G \in \mathbb{F}_q^{k\times n}$ be a generator matrix of C, let $\mathcal{S}$ be the distribution of Ge_t, and, for each s in the support of $\mathcal{S}$, let $\mathcal{E}_s$ be the conditional distribution of e_t given $Ge_t = s$.

For $\mathbb{F}_q^n$, take the identity matrix as the generator matrix. Then the projectors from Lemma 6.12 obey

$$\Pi_s^C = \sum_{x\,:\,Gx=s} \Pi_x^{\mathrm{all}}.$$

For each x with $\Pr[e_t = x] > 0$, normalize $\Pi_x^{\mathrm{all}}|+_n\rangle^{\otimes t}$ to a unit vector $|u_x\rangle$. The vectors $|u_x\rangle$ are mutually orthogonal. Within the s block, the normalized state arising from ρ_C^t is therefore

$$\left(\sum_x \sqrt{\mathcal{E}_s(x)}|u_x\rangle\right)\left(\sum_y \sqrt{\mathcal{E}_s(y)}\langle u_y|\right),$$

whereas the corresponding normalized block of ρ_{all}^t is $\sum_x \mathcal{E}_s(x)|u_x\rangle\langle u_x|$. Both blocks have weight $\mathcal{S}(s)$, by Lemma 6.13. Orthogonality of the blocks now gives the exact identity

$$d_{\mathrm{tr}}\left(\rho_C^t, \rho_{\mathrm{all}}^t\right) = \mathop{\mathbb{E}}_{s\sim\mathcal{S}}[\nu(\mathcal{E}_s)]. \tag{6.6}$$

By Lemma 6.14, Jensen's inequality, and the definition of the MAP decoder,

$$\begin{aligned}
\mathop{\mathbb{E}}_{s\sim\mathcal{S}}[\nu(\mathcal{E}_s)] &\le \sqrt{1 - \sum_x \mathop{\mathbb{E}}_{s\sim\mathcal{S}}[\mathcal{E}_s(x)^2]},\\
\sum_x \mathop{\mathbb{E}}_{s\sim\mathcal{S}}[\mathcal{E}_s(x)^2] &= \sum_s \mathcal{S}(s)\sum_x \mathcal{E}_s(x)^2\\
&\ge \sum_s \mathcal{S}(s)\left(\max_x \mathcal{E}_s(x)\right)^2\\
&\ge \left(\sum_s \mathcal{S}(s)\max_x \mathcal{E}_s(x)\right)^2\\
&= p_{\mathrm{decode}}^t(C^\perp)^2.
\end{aligned}$$

Here G is also a parity-check matrix of $C^\perp$, so the final quantity is exactly the decoding probability in Definition 6.11. Combining the two displays with Eq. (6.6) proves the theorem. □

To convert indistinguishability into a testing lower bound, we compare codeword phase states with Haar-random states. Let

$$\rho_{\mathrm{Haar}}^t := \mathop{\mathbb{E}}_{|\Psi\rangle\sim\mathrm{Haar}(\mathbb{C}^n)}\left[|\Psi\rangle\langle\Psi|^{\otimes t}\right].$$

Theorem 6.16 (Random finite-field phase states). *For every prime q,*

$$d_{\mathrm{tr}}\left(\rho_{\mathrm{all}}^t, \rho_{\mathrm{Haar}}^t\right) \le \frac{t(t-1)}{n}.$$

The proof of Theorem 6.16 is given in Section 6.5.1.

Corollary 6.17. *For every linear code $C \le \mathbb{F}_q^n$,*

$$d_{\mathrm{tr}}(\rho_C^t, \rho_{\mathrm{Haar}}^t) \le \sqrt{1 - p_{\mathrm{decode}}^t(C^\perp)^2} + \frac{t(t-1)}{n}.$$

Proof. Apply the triangle inequality to Theorems 6.15 and 6.16. □

The remaining ingredient is that a Haar-random state is very unlikely to be close to any flat state. Define

$$\mathcal{P}_{\mathrm{flat}} := \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e^{i\theta_i} |i\rangle \;:\; \theta_i \in \mathbb{R} \right\}.$$

Lemma 6.18 (Haar measure of flat-state neighborhoods). *Fix $\epsilon < 1 - \pi/4$. There are constants $c_\epsilon > 0$ and n_ϵ such that, for all $n \ge n_\epsilon$,*

$$\Pr_{|\Psi\rangle \sim \mathrm{Haar}(\mathbb{C}^n)} \left[\max_{|\theta\rangle \in \mathcal{P}_{\mathrm{flat}}} |\langle \Psi | \theta \rangle|^2 \ge 1 - \epsilon \right] \le 2e^{-c_\epsilon n}.$$

Proof. For $|\psi\rangle = \sum_i \alpha_i |i\rangle$, set

$$L(|\psi\rangle) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |\alpha_i|.$$

Choosing each phase to align with α_i gives

$$\max_{|\theta\rangle \in \mathcal{P}_{\mathrm{flat}}} |\langle \psi | \theta \rangle|^2 = L(|\psi\rangle)^2.$$

For a Haar-random state $|\Psi\rangle = \sum_i A_i |i\rangle$, one has $|A_1|^2 \sim \mathrm{Beta}(1, n-1)$. Hence

$$\mathbb{E}\, L(|\Psi\rangle) = \sqrt{n}\; \mathbb{E}|A_1| = \sqrt{n}\, \frac{\Gamma(3/2)\Gamma(n)}{\Gamma(n+1/2)}.$$

The Wendel–Gautschi inequalities [Wen48; Gau59] imply

$$\frac{\sqrt{\pi}}{2} \le \mathbb{E}\, L(|\Psi\rangle) \le \frac{\sqrt{\pi}}{2} \sqrt{1 + \frac{1}{2n}}.$$

Set $a_\epsilon = \sqrt{1-\epsilon}$ and $\Delta_\epsilon = a_\epsilon - \sqrt{\pi}/2 > 0$. For all sufficiently large n,

$$\mathbb{E}\, L(|\Psi\rangle) \le a_\epsilon - \frac{\Delta_\epsilon}{2}.$$

Moreover, L is 1-Lipschitz, because

$$|L(|\psi\rangle) - L(|\phi\rangle)| \le \frac{1}{\sqrt{n}} \sum_i |\alpha_i - \beta_i| \le \| |\psi\rangle - |\phi\rangle \|_2.$$

Lévy's lemma, in the form recorded for example in [Ger13], therefore yields

$$\begin{aligned}\Pr[L(|\Psi\rangle)^2 \geq 1-\epsilon] &\leq \Pr\left[|L(|\Psi\rangle) - \mathbb{E}\, L(|\Psi\rangle)| \geq \frac{\Delta_\epsilon}{2}\right] \\ &\leq 2\exp\left(-\frac{n\Delta_\epsilon^2}{36\pi^3}\right).\end{aligned}$$

Thus the claim holds with $c_\epsilon = \Delta_\epsilon^2/(36\pi^3)$. □

Lemma 6.19. *Let $\mathcal{P}$ be a property of pure states, all of which are flat, and suppose*

$$d_{\mathrm{tr}}\left(\rho_{\mathcal{P}}^t, \rho_{\mathrm{Haar}}^t\right) \leq \gamma,$$

where $\rho_{\mathcal{P}}^t$ is the tth moment of a distribution supported on $\mathcal{P}$. Define

$$\eta_\epsilon(\mathcal{P}) \coloneqq \Pr_{|\Psi\rangle\sim\mathrm{Haar}}\left[\max_{|\phi\rangle\in\mathcal{P}} |\langle\Psi|\phi\rangle|^2 \geq 1-\epsilon\right].$$

If

$$\eta_\epsilon(\mathcal{P}) < \frac{1}{2} - \frac{3\gamma}{2},$$

then every ϵ-tester for $\mathcal{P}$ uses at least $t+1$ copies.

Proof. Suppose that a tester uses $k \leq t$ copies, and let M be its accepting POVM element. Taking a partial trace of the moment states shows that their kth moments are also within trace distance γ. Completeness therefore gives

$$\mathrm{tr}(M\rho_{\mathrm{Haar}}^k) \geq \mathrm{tr}(M\rho_{\mathcal{P}}^k) - \gamma \geq \frac{2}{3} - \gamma.$$

On the other hand, a Haar-random state lies in the ϵ-neighborhood of $\mathcal{P}$ with probability $\eta_\epsilon(\mathcal{P})$. The tester accepts with probability at most 1 on that neighborhood and, by soundness, at most $1/3$ outside it. Thus

$$\mathrm{tr}(M\rho_{\mathrm{Haar}}^k) \leq \eta_\epsilon(\mathcal{P}) + \frac{1}{3}\big(1-\eta_\epsilon(\mathcal{P})\big) = \frac{1}{3} + \frac{2}{3}\eta_\epsilon(\mathcal{P}).$$

Combining the two inequalities forces $\eta_\epsilon(\mathcal{P}) \geq \frac{1}{2} - \frac{3\gamma}{2}$, a contradiction. □

Theorem 6.20 (General codeword-state testing lower bound). *Let $C \le \mathbb{F}_q^n$, fix $0 < \epsilon < 1 - \pi/4$, and define*

$$\gamma := \sqrt{1 - p_{\mathrm{decode}}^t(C^\perp)^2} + \frac{t(t-1)}{n}.$$

For every fixed $\delta > 0$ and all sufficiently large n, if $\gamma \le 1/3 - \delta$, then every ϵ-tester for $\mathcal{P}_C$ requires at least $t+1$ copies.

Proof. By Corollary 6.17, the tth moment of a uniformly random C-codeword state is within trace distance γ of the Haar moment. Since $\mathcal{P}_C \subseteq \mathcal{P}_{\mathrm{flat}}$, Lemma 6.18 gives $\eta_\epsilon(\mathcal{P}_C) \le 2e^{-c_\epsilon n}$. For sufficiently large n, this is strictly smaller than $3\delta/2$, while

$$\frac{1}{2} - \frac{3\gamma}{2} \ge \frac{3\delta}{2}.$$

The claim follows from Lemma 6.19. □

6.4.3 Lower Bounds for Reed–Muller Codeword Phase States

To apply our framework, it remains to show that the relevant Reed–Muller dual codes can decode the random errors above. The required decoding estimate is stated next; its proof is deferred to Section 6.5.2.

Corollary 6.21 (Random-error decoding estimate). *Let q be prime, let $m \ge 1$, and suppose*

$$0 \le d \le \left\lfloor \frac{m(q-1)-2}{2} \right\rfloor, \qquad 0 \le r \le m.$$

Set $C = \mathrm{RM}_q[m, 2d+1]$. If $1 \le t \le D_{d,q}(r)$, then

$$\sqrt{1 - p_{\mathrm{decode}}^t(C^\perp)^2} \le \sqrt{2tq^{r-m}}.$$

Lemma 6.22. *Let q be prime and fix $0 < \epsilon < 1 - \pi/4$. Let d_0, d satisfy*

$$0 \le d_0 \le \left\lfloor \frac{m(q-1)-2}{2} \right\rfloor, \qquad 2d_0 + 1 \le d \le m(q-1).$$

For all sufficiently large m, every ϵ-tester for $\mathcal{P}_{\mathrm{RM}_q[m,d]}$ requires

$$\Omega\Big(\max_{0 \le r \le m} \min\big\{ D_{d_0,q}(r),\, q^{m-r},\, q^{m/2} \big\} \Big)$$

copies.

Proof. Fix $r \in \{0, \dots, m\}$ and abbreviate

$$T_r := \min\left\{D_{d_0,q}(r),\ q^{m-r},\ q^{m/2}\right\}.$$

Choose a sufficiently small absolute constant $c \in (0,1)$. If $T_r < 2/c$, the claimed $\Omega(T_r)$ bound follows from the trivial one-copy lower bound: the property is nonempty, and, for sufficiently large m, its ϵ-far set is nonempty by Lemma 6.18. We may therefore suppose that $T_r \geq 2/c$ and take $t = \lfloor cT_r \rfloor$. Then

$$1 \leq \frac{cT_r}{2} \leq t \leq cT_r.$$

Let

$$C = \mathrm{RM}_q[m, d], \qquad C' = \mathrm{RM}_q[m, 2d_0 + 1].$$

The degree assumption gives $C' \subseteq C$, hence $C^\perp \subseteq C'^\perp$. A smaller code supplies at least as much syndrome information, so

$$p^t_{\mathrm{decode}}(C^\perp) \geq p^t_{\mathrm{decode}}(C'^\perp).$$

Since $t \leq D_{d_0,q}(r)$, Corollary 6.21 yields

$$\sqrt{1 - p^t_{\mathrm{decode}}(C^\perp)^2} \leq \sqrt{2tq^{r-m}}.$$

The Hilbert-space dimension is $n = q^m$, and therefore Corollary 6.17 bounds the distance between the codeword and Haar moment states by

$$\gamma \leq \sqrt{2tq^{r-m}} + \frac{t^2}{q^m}.$$

Our choice of t ensures both $t \leq cq^{m-r}$ and $t \leq cq^{m/2}$, so

$$\gamma \leq \sqrt{2c} + c^2.$$

Choose c small enough that the right-hand side is below $1/3 - \delta$ for some constant $\delta > 0$. Then Theorem 6.20 rules out a tester using t copies. This proves an $\Omega(T_r)$ bound; maximizing over r completes the proof. □

Theorem 6.23 (Low-degree phase-state testing lower bound). *Let q be prime, let $1 \le d \le m(q-1)$, fix $0 < \epsilon < 1 - \pi/4$, and set*

$$d_0 := \left\lfloor \frac{d-1}{2} \right\rfloor.$$

Suppose

$$d_0 \le \left\lfloor \frac{m(q-1)-2}{2} \right\rfloor.$$

For all sufficiently large m, every ϵ-tester for m-variate degree-d phase states over $\mathbb{F}_q$ requires

$$\Omega\left(\sum_{j=0}^{\min\{\lfloor m/2 \rfloor, d_0\}} \binom{\lfloor m/2 \rfloor}{j}\right)$$

copies.

Proof. Apply Lemma 6.22 with $r = \lfloor m/2 \rfloor$. Since $D_{d_0,q}(r) \le q^r$, we have

$$D_{d_0,q}(r) \le q^{m-r}, \qquad D_{d_0,q}(r) \le q^{m/2}.$$

The general lower bound is therefore $\Omega(D_{d_0,q}(r))$. Counting only squarefree monomials gives

$$D_{d_0,q}(r) \ge \sum_{j=0}^{\min\{r, d_0\}} \binom{r}{j},$$

which proves the theorem. □

In particular, the above theorems imply the simpler lower bound

$$\Omega\left(\binom{\lfloor m/2 \rfloor}{\min\{\lfloor m/2 \rfloor, \lfloor (d-1)/2 \rfloor\}}\right).$$

For fixed d, this gives

$$\Omega\left(m^{\lfloor (d-1)/2 \rfloor}\right)$$

copies. Thus, unlike classical low-degree testing, the quantum copy complexity has an exponent that grows with the degree. In particular, there is no bound polynomial jointly in m, d, and q.

6.4.4 Quadratic Phase States

Degree-2 phase states are stabilizer states in the standard prime-dimensional qudit formalism, and stabilizer states in that formalism admit a constant-copy test [GNW21]. One might therefore expect quadratic phase states to be equally easy to recognize. In fact, testing this smaller family requires a linear number of copies, matching the order of the copy complexity for learning the state [Röt09].

We use notation tailored to this case. Let $\mathbf{1}_t \in \mathbb{F}_q^t$ be the all-ones vector. For $x_1, \dots, x_t \in \mathbb{F}_q^m$, let $X \in \mathbb{F}_q^{m\times t}$ be the matrix with these columns and write $|X\rangle = |x_1, \dots, x_t\rangle$. Given $Q \in \mathbb{F}_q^{m\times m}$ and $\ell \in \mathbb{F}_q^m$, define

$$f_{Q,\ell}(x) := x^\top Q x + \langle \ell, x\rangle.$$

As Q and ℓ range uniformly, this parametrizes the uniform distribution over degree-at-most-2 polynomial functions with zero constant term. Constant terms contribute only a global phase. Let

$$\rho_0 := \mathop{\mathbb{E}}_{Q,\ell}\Big[|\psi_{f_{Q,\ell}}\rangle\langle\psi_{f_{Q,\ell}}|^{\otimes t}\Big].$$

Lemma 6.24. *The state ρ_0 has the expansion*

$$\rho_0 = q^{-mt} \sum_{X,Y\in\mathbb{F}_q^{m\times t}} \mathbf{1}\{XX^\top = YY^\top \text{ and } X\mathbf{1}_t = Y\mathbf{1}_t\}|X\rangle\langle Y|.$$

Proof. Expanding the moment state and applying character orthogonality twice gives

$$\begin{aligned}\rho_0 &= q^{-mt} \sum_{X,Y} \mathop{\mathbb{E}}_{Q}\Big[\omega_q^{\langle Q, XX^\top - YY^\top\rangle}\Big] \mathop{\mathbb{E}}_{\ell}\Big[\omega_q^{\langle \ell, X\mathbf{1}_t - Y\mathbf{1}_t\rangle}\Big]|X\rangle\langle Y| \\ &= q^{-mt} \sum_{X,Y} \mathbf{1}\{XX^\top = YY^\top\}\mathbf{1}\{X\mathbf{1}_t = Y\mathbf{1}_t\}|X\rangle\langle Y|.\end{aligned}$$

□

For nonzero $v \in \mathbb{F}_q^m$, let

$$H_v := \{x \in \mathbb{F}_q^m : \langle x, v\rangle = 0\}, \qquad P_v := \sum_{x\in H_v} |x\rangle\langle x|.$$

For a quadratic polynomial f, define the phase state truncated to this hyperplane by

$$|\psi_{v,f}\rangle := q^{-(m-1)/2} \sum_{x \in H_v} \omega_q^{f(x)} |x\rangle = \sqrt{q}\, P_v |\psi_f\rangle.$$

For any quadratic phase state $|\psi_g\rangle$,

$$\left|\langle \psi_g | \psi_{v,f} \rangle\right| \le \frac{|H_v|}{q^{m/2} q^{(m-1)/2}} = \frac{1}{\sqrt{q}}.$$

Every truncated state is consequently at infidelity at least $1 - 1/q$ from every quadratic phase state.

In the averages below, v is uniform on $\mathbb{F}_q^m \setminus \{0\}$, while $f = f_{Q,\ell}$ is drawn independently from the distribution defining ρ_0.

Define the average truncated moment

$$\rho_1 := \mathop{\mathbb{E}}_{\substack{v \neq 0 \\ f}} \left[|\psi_{v,f}\rangle\langle\psi_{v,f}|^{\otimes t} \right].$$

The projection identity above gives

$$\rho_1 = q^t \mathop{\mathbb{E}}_{v \neq 0} \left[P_v^{\otimes t} \rho_0 P_v^{\otimes t} \right]. \tag{6.7}$$

Equivalently,

$$\langle X | \rho_1 | Y \rangle = q^t \mathop{\mathbb{E}}_{v \neq 0} [\mathbf{1}\{X \subseteq H_v\} \mathbf{1}\{Y \subseteq H_v\}] \langle X | \rho_0 | Y \rangle, \tag{6.8}$$

where $X \subseteq H_v$ means that every column of X lies in H_v.

Let

$$\Pi_{=t} := \sum_{\substack{X \in \mathbb{F}_q^{m \times t} \\ \mathrm{rank}(X) = t}} |X\rangle\langle X|, \qquad \Pi_{<t} := I - \Pi_{=t}.$$

We first record a linear-algebra observation. If $\mathrm{rank}(X) = t$, then $X^\top$ is surjective and X is injective, so

$$\mathrm{im}(XX^\top) = \mathrm{col}(X), \qquad \mathrm{rank}(XX^\top) = t.$$

It follows that whenever $\mathrm{rank}(X) = t$ and $XX^\top = YY^\top$, one also has $\mathrm{rank}(Y) = t$ and $\mathrm{col}(Y) = \mathrm{col}(X)$. By Lemma 6.24 and Eq. (6.8), both ρ_0 and ρ_1 are therefore block diagonal with respect to $\Pi_{=t} + \Pi_{<t}$.

Lemma 6.25. *The two blocks satisfy*

$$\|\Pi_{=t}(\rho_0 - \rho_1)\Pi_{=t}\|_1 \le q^{t-m}$$

and

$$\|\Pi_{<t}(\rho_0 - \rho_1)\Pi_{<t}\|_1 \le (q+1)q^{t-m}.$$

Proof. Suppose X and Y have rank t and $\langle X|\rho_0|Y\rangle \neq 0$. Their column spaces coincide, and the condition $X \subseteq H_v$ is therefore equivalent to both $X \subseteq H_v$ and $Y \subseteq H_v$. There are $q^{m-t} - 1$ nonzero vectors orthogonal to this t-dimensional column space. Thus

$$\mathop{\mathbb{E}}_{v \neq 0}[\mathbf{1}\{X \subseteq H_v\}\mathbf{1}\{Y \subseteq H_v\}] = \frac{q^{m-t}-1}{q^m - 1}.$$

Using Eq. (6.8),

$$\Pi_{=t}\rho_1\Pi_{=t} = \frac{q^m - q^t}{q^m - 1}\,\Pi_{=t}\rho_0\Pi_{=t}.$$

Since the latter operator is positive semidefinite,

$$\begin{aligned}\|\Pi_{=t}(\rho_0 - \rho_1)\Pi_{=t}\|_1 &= \left(1 - \frac{q^m - q^t}{q^m - 1}\right)\operatorname{tr}(\Pi_{=t}\rho_0)\\ &\le q^{t-m}.\end{aligned}$$

For the rank-deficient block, positivity and the triangle inequality give

$$\|\Pi_{<t}(\rho_0 - \rho_1)\Pi_{<t}\|_1 \le \operatorname{tr}(\Pi_{<t}\rho_0) + \operatorname{tr}(\Pi_{<t}\rho_1).$$

The probability that t uniformly random vectors of $\mathbb{F}_q^m$ are linearly dependent is at most q^{t-m}. Conditional on a hyperplane, the corresponding probability in its $(m-1)$-dimensional vector space is at most q^{t-m+1}. These are the two traces above, proving the second inequality. □

Proposition 6.26.

$$\|\rho_0 - \rho_1\|_1 \le (q+2)q^{t-m}.$$

Proof. Add the two bounds in Lemma 6.25, using the block-diagonal decomposition. □

Theorem 6.27 (Quadratic phase-state lower bound). *For every $0 < \epsilon \le 1 - 1/q$, any ϵ-tester for m-variate degree-2 phase states over $\mathbb{F}_q$ requires at least $m - 2$ copies.*

Proof. Suppose that a tester uses t copies, and let M be its accepting POVM element. Every state in the ensemble ρ_0 is a quadratic phase state, whereas every state in the ensemble ρ_1 has infidelity at least $1 - 1/q \geq \epsilon$ from all quadratic phase states. Completeness, soundness, and linearity therefore imply

$$\frac{1}{3} \leq \operatorname{tr}\big(M(\rho_0 - \rho_1)\big) \leq \frac{1}{2}\|\rho_0 - \rho_1\|_1 \leq \frac{q+2}{2} q^{t-m},$$

where the last inequality is Proposition 6.26. Consequently,

$$t \geq m + \log_q\left(\frac{2}{3(q+2)}\right).$$

For every prime $q \geq 2$, the logarithmic term is greater than -3. Since t is an integer, $t \geq m - 2$. □

6.5 Deferred Proofs

This final part supplies the two longer ingredients deferred above. The first compares uniformly random $\mathbb{F}_q$ phase states with Haar-random states. The second proves the Reed–Muller decoding estimate used in the general low-degree lower bound.

6.5.1 Random Phase States Are Close to Haar Random

Proof of Theorem 6.16. If $t > n$, the claimed upper bound is at least 1 and is therefore trivial. Assume henceforth that $t \leq n$.

Index the computational basis of $(\mathbb{C}^n)^{\otimes t}$ by strings $a = (a_1, \dots, a_t) \in [n]^t$, and let

$$m_a(j) := |\{\ell \in [t] \,:\, a_\ell = j\}|$$

be the occupation number of j. The matrix entries of the random-phase moment are

$$\begin{aligned} \langle a|\rho_{\mathrm{all}}^t|b\rangle &= \frac{1}{n^t} \mathop{\mathbb{E}}_{v \sim \mathbb{F}_q^n} \left[\omega_q^{\sum_{\ell=1}^t (v_{a_\ell} - v_{b_\ell})} \right] \\ &= \frac{1}{n^t} \mathbf{1}\{m_a(j) \equiv m_b(j) \pmod q \text{ for all } j \in [n]\}. \end{aligned} \tag{6.9}$$

For the Haar moment, the standard symmetric-subspace formula [Har13] gives

$$\rho_{\mathrm{Haar}}^t = \frac{1}{n(n+1)\cdots(n+t-1)} \sum_{\pi \in S_t} U_\pi,$$

and hence

$$\langle a|\rho_{\mathrm{Haar}}^{t}|b\rangle = \frac{|\{\pi \in S_t \,:\, a = b \circ \pi\}|}{n(n+1)\cdots(n+t-1)}. \tag{6.10}$$

Let Π_{distinct} project onto the span of strings with no repeated coordinate. If both a and b are collision-free, every occupation number is 0 or 1. Since $q \geq 2$, congruence modulo q is then the same as equality. Thus Eqs. (6.9) and (6.10) have the same nonzero pattern on this subspace: they are nonzero exactly when a and b are permutations of one another, and there is a unique matching permutation.

Define the unnormalized operator

$$A \coloneqq \sum_{\substack{a,b\in[n]^t \text{ collision-free} \\ \{a_1,\ldots,a_t\}=\{b_1,\ldots,b_t\}}} |a\rangle\langle b|.$$

Then

$$\Pi_{\mathrm{distinct}}\rho_{\mathrm{all}}^{t}\Pi_{\mathrm{distinct}} = \frac{A}{n^t}, \qquad \Pi_{\mathrm{distinct}}\rho_{\mathrm{Haar}}^{t}\Pi_{\mathrm{distinct}} = \frac{A}{n(n+1)\cdots(n+t-1)}.$$

Writing $(n)_t = n(n-1)\cdots(n-t+1)$, we have $\mathrm{tr}(A) = (n)_t$. The weights of the collision-free block are therefore

$$p_{\mathrm{all}} = \frac{(n)_t}{n^t}, \qquad p_{\mathrm{Haar}} = \frac{(n)_t}{n(n+1)\cdots(n+t-1)}.$$

After normalization, both restrictions equal the same state $\tau = A/(n)_t$.

Both moment states are block diagonal with respect to $\Pi_{\mathrm{distinct}} \oplus (I - \Pi_{\mathrm{distinct}})$. This is immediate for the Haar moment because permutations preserve occupation numbers. For the phase moment, suppose a is collision-free and the entry in Eq. (6.9) is nonzero. At each of the t locations occupied by a, the corresponding occupation number of b is congruent to 1 modulo q, and is therefore at least 1. Since the occupation numbers of b sum to t, these are all exactly 1, and every other occupation number is 0. Thus b is also collision-free.

There consequently exist states ρ_{bad} and σ_{bad}, supported on the collision subspace, such that

$$\rho_{\mathrm{all}}^{t} = p_{\mathrm{all}}\tau + (1 - p_{\mathrm{all}})\rho_{\mathrm{bad}}, \qquad \rho_{\mathrm{Haar}}^{t} = p_{\mathrm{Haar}}\tau + (1 - p_{\mathrm{Haar}})\sigma_{\mathrm{bad}}.$$

The triangle inequality gives

$$\begin{aligned} d_{\mathrm{tr}}\big(\rho_{\mathrm{all}}^{t}, \rho_{\mathrm{Haar}}^{t}\big) &\le \frac{1}{2}\big(|p_{\mathrm{all}} - p_{\mathrm{Haar}}| + 1 - p_{\mathrm{all}} + 1 - p_{\mathrm{Haar}}\big) \\ &= 1 - p_{\mathrm{Haar}}, \end{aligned}$$

where the equality uses $p_{\mathrm{Haar}} \le p_{\mathrm{all}} \le 1$. Finally,

$$\begin{aligned} p_{\mathrm{Haar}} &= \prod_{j=0}^{t-1} \frac{n-j}{n+j} = \prod_{j=0}^{t-1}\Big(1 - \frac{2j}{n+j}\Big), \\ 1 - p_{\mathrm{Haar}} &\le \sum_{j=0}^{t-1} \frac{2j}{n+j} \le \frac{t(t-1)}{n}. \end{aligned}$$

This proves the theorem. □

6.5.2 Random-Error Decoding for Reed–Muller Codes

We now prove Corollary 6.21. The main combinatorial step is that a small random collection of columns from a Reed–Muller generator matrix is linearly independent with high probability. This extends the binary analysis of Abbe, Shpilka, and Wigderson [ASW15] to arbitrary prime fields.

For $0 \le d \le m(q-1)$, let

$$\mathcal{P}_{m,d,q} := \mathrm{span}_{\mathbb{F}_q}\{x^{\alpha} \,:\, \alpha \in \{0,\dots,q-1\}^m,\ \alpha_1 + \cdots + \alpha_m \le d\}.$$

This is the vector space of degree-at-most-d polynomial functions on $\mathbb{F}_q^m$. Let $\mathcal{M}_{m,d,q}$ be its displayed monomial basis. For $x \in \mathbb{F}_q^m$, define the evaluation functional

$$E_x : \mathcal{P}_{m,d,q} \to \mathbb{F}_q, \qquad E_x(f) = f(x).$$

In the monomial basis, E_x is the column indexed by x of the standard generator matrix of $\mathrm{RM}_q[m,d]$.

For a linear code $D \le \mathbb{F}_q^N$, define

$$\mathrm{supp}(D) := \{i \in [N] \,:\, y_i \neq 0 \text{ for some } y \in D\}.$$

Its ath generalized Hamming weight is

$$d_a(D) \coloneqq \min\{|\mathrm{supp}(D')| \,:\, D' \leq D,\ \dim D' = a\}.$$

We use the following characterization of the generalized Hamming weights of q-ary Reed–Muller codes [HP98, Theorem 5.10].

Theorem 6.28 (Heijnen–Pellikaan). *Let $Q = \{0, \dots, q-1\}$, and let α be the ath element in lexicographic order of*

$$\left\{\beta \in Q^m \,:\, \sum_i \beta_i > (q-1)m - d - 1\right\}.$$

Then

$$d_a(\mathrm{RM}_q[m,d]) = \sum_{i=1}^{m} \alpha_i q^{m-i} + 1.$$

For $A \subseteq \mathbb{F}_q^m$, let

$$\mathrm{punc}_A : \mathcal{P}_{m,d,q} \to \mathbb{F}_q^A, \qquad \mathrm{punc}_A(f) = f|_A,$$

and define

$$I_d(A) \coloneqq \ker(\mathrm{punc}_A), \qquad h_d(A) \coloneqq \mathrm{rank}(\mathrm{punc}_A).$$

The dual map of punc_A sends the coordinate functional at a to E_a. Consequently,

$$h_d(A) = \dim \mathrm{span}\{E_a \,:\, a \in A\}. \tag{6.11}$$

Define the degree-d closure

$$\mathrm{cl}_d(A) \coloneqq \left\{x \in \mathbb{F}_q^m \,:\, f(x) = 0 \text{ for every } f \in I_d(A)\right\}.$$

Lemma 6.29. *For every $A \subseteq \mathbb{F}_q^m$ and $x \in \mathbb{F}_q^m$,*

$$x \in \mathrm{cl}_d(A) \quad \iff \quad E_x \in \mathrm{span}\{E_a \,:\, a \in A\}.$$

Moreover,

$$I_d(\mathrm{cl}_d(A)) = I_d(A), \qquad h_d(\mathrm{cl}_d(A)) = h_d(A).$$

Proof. We have $I_d(A) = \bigcap_{a \in A} \ker(E_a)$. If $E_x \in \operatorname{span}\{E_a : a \in A\}$, then E_x vanishes on this intersection, so $x \in \mathrm{cl}_d(A)$. Conversely, if $x \notin \mathrm{cl}_d(A)$, there is an $f \in I_d(A)$ such that $E_x(f) \neq 0$; this rules out $E_x \in \operatorname{span}\{E_a : a \in A\}$.

Since $A \subseteq \mathrm{cl}_d(A)$, one inclusion $I_d(\mathrm{cl}_d(A)) \subseteq I_d(A)$ is immediate. The definition of closure says that every polynomial in $I_d(A)$ also vanishes throughout $\mathrm{cl}_d(A)$, giving the reverse inclusion. Equality of the two ranks follows by rank–nullity. □

Lemma 6.30. *Let $d \geq 1$. If $A \subseteq \mathbb{F}_q^m$ satisfies*

$$h_d(A) \leq D_{d,q}(r),$$

then $|A| \leq q^r$.

Proof. The claim is immediate if $r = m$. If $d = m(q-1)$, then $\mathcal{P}_{m,d,q}$ is the full space of functions on $\mathbb{F}_q^m$, so

$$h_d(A) = |A|, \qquad D_{d,q}(r) = q^r,$$

and the result again follows. We may therefore assume $1 \leq d < m(q-1)$ and $r < m$.

Set

$$N = D_{d,q}(m), \qquad a = N - D_{d,q}(r).$$

We claim that

$$d_a(\mathrm{RM}_q[m,d]) = q^m - q^r. \tag{6.12}$$

To see this, let $Q = \{0, \dots, q-1\}$ and $\mu = (q-1, \dots, q-1)$. The map $\beta \mapsto \mu - \beta$ reverses lexicographic order and is a bijection from

$$B_{\leq d} := \left\{ \beta \in Q^m : \sum_i \beta_i \leq d \right\}$$

onto

$$F_{>(q-1)m-d-1} := \left\{ \alpha \in Q^m : \sum_i \alpha_i > (q-1)m - d - 1 \right\}.$$

The first $D_{d,q}(r)$ elements of $B_{\leq d}$ are exactly those whose first $m-r$ coordinates vanish. The next element is

$$\beta^\star = (0, \dots, 0, 1, 0, \dots, 0),$$

with the 1 in coordinate $m-r$. It follows that $\alpha^\star = \mu - \beta^\star$ is the ath element of the latter set. By Theorem 6.28,

$$d_a(\mathrm{RM}_q[m,d]) = \sum_{i=1}^{m} \alpha_i^\star q^{m-i} + 1 = (q^m - 1) - q^r + 1 = q^m - q^r,$$

proving Eq. (6.12).

Now consider the code $C = \mathrm{RM}_q[m,d]$. By hypothesis,

$$\dim \ker(\mathrm{punc}_A) = N - h_d(A) \geq N - D_{d,q}(r) = a.$$

Choose an a-dimensional subcode D of this kernel. Every codeword in D vanishes on A, so $\mathrm{supp}(D) \subseteq A^c$. Therefore,

$$q^m - |A| \geq |\mathrm{supp}(D)| \geq d_a(C) = q^m - q^r,$$

which is equivalent to $|A| \leq q^r$. □

Theorem 6.31 (Random Reed–Muller columns). *Let q be prime, let $m \geq 1$, let $0 \leq d \leq m(q-1)$, and let $0 \leq r \leq m$. If $v_1, \ldots, v_t$ are independent uniformly random points of $\mathbb{F}_q^m$ and $t \leq D_{d,q}(r)$, then*

$$\Pr[E_{v_1}, \ldots, E_{v_t} \text{ are linearly independent}] \geq 1 - tq^{r-m}.$$

Proof. If $d = 0$, then $D_{0,q}(r) = 1$, so $t \leq 1$ and the conclusion is immediate. Assume $d \geq 1$.

For $0 \leq j \leq t$, let $\mathcal{A}_j$ be the event that $E_{v_1}, \ldots, E_{v_j}$ are linearly independent, and set $A_j = \{v_1, \ldots, v_j\}$. On $\mathcal{A}_j$, Eq. (6.11) gives $h_d(A_j) = j$. By Lemma 6.29,

$$h_d(\mathrm{cl}_d(A_j)) = j.$$

For $0 \leq j \leq t-1$, we have $j < t \leq D_{d,q}(r)$, so Lemma 6.30 implies

$$|\mathrm{cl}_d(A_j)| \leq q^r.$$

Conditional on $\mathcal{A}_j$, the first new dependence occurs at sample $j+1$ precisely when

$$E_{v_{j+1}} \in \mathrm{span}\{E_{v_1}, \ldots, E_{v_j}\},$$

which, by Lemma 6.29, is equivalent to $v_{j+1} \in \mathrm{cl}_d(A_j)$. Since v_{j+1} is fresh and uniform,

$$\Pr[\mathcal{A}_j \cap \mathcal{A}_{j+1}^c] \le q^{r-m}.$$

The events on the left are disjoint for $j = 0, \dots, t-1$, and their union is $\mathcal{A}_t^c$. Summing proves the theorem. □

We next translate column independence into unique decoding. Let $E_q(m,d)$ be the matrix whose rows are indexed by $\mathcal{M}_{m,d,q}$, whose columns are indexed by $x \in \mathbb{F}_q^m$, and whose (f,x) entry is $f(x)$. Thus $E_q(m,d)$ generates $\mathrm{RM}_q[m,d]$ and, by Reed–Muller duality, is a parity-check matrix of $\mathrm{RM}_q[m, m(q-1)-d-1]$.

For error patterns $e, e' \in \mathbb{F}_q^{q^m}$, define

$$e \sim_d e' \quad \Longleftrightarrow \quad \sum_{x \in \mathbb{F}_q^m} e(x) f(x) = \sum_{x \in \mathbb{F}_q^m} e'(x) f(x) \quad \text{for every } f \in \mathcal{P}_{m,d,q}.$$

Lemma 6.32. *For all $e, e' \in \mathbb{F}_q^{q^m}$,*

$$E_q(m,d)e = E_q(m,d)e' \quad \Longleftrightarrow \quad e \sim_d e'.$$

Proof. The row indexed by a monomial $g \in \mathcal{M}_{m,d,q}$ computes $\sum_x e(x)g(x)$. Equality of all these coordinates is equivalent, by linearity and the fact that $\mathcal{M}_{m,d,q}$ is a basis, to equality for every $f \in \mathcal{P}_{m,d,q}$. □

For $U = \{u_1, \dots, u_t\} \subseteq \mathbb{F}_q^m$, let U^d denote the submatrix of $E_q(m,d)$ formed by the columns indexed by U.

Lemma 6.33 (Independent columns give unique errors). *Let*

$$e = \sum_{i=1}^{t} c_i \delta_{u_i},$$

where the u_i are distinct and $c_i \in \mathbb{F}_q^\times$. If the columns of U^d are linearly independent, then no error $e' \neq e$ of weight at most t satisfies $e' \sim_{2d+1} e$.

Proof. Column independence means that the evaluation map

$$\mathcal{P}_{m,d,q} \to \mathbb{F}_q^t, \qquad f \mapsto (f(u_1), \dots, f(u_t))$$

is surjective. Choose $f_1, \dots, f_t \in \mathcal{P}_{m,d,q}$ satisfying $f_i(u_j) = \delta_{i,j}$.

Write an error of weight $t' \le t$ as

$$e' = \sum_{k=1}^{t'} c'_k \delta_{v_k},$$

where the v_k are distinct and $c'_k \neq 0$, and suppose $e' \sim_{2d+1} e$. Define

$$g_i := (f_i(v_1), \dots, f_i(v_{t'}))^\top, \qquad W' := \operatorname{diag}(c'_1, \dots, c'_{t'}).$$

Since $f_i f_j$ has degree at most $2d$, equivalence of the errors gives

$$\begin{aligned} g_i^\top W' g_j &= \sum_{k=1}^{t'} c'_k f_i(v_k) f_j(v_k) \\ &= \sum_{k=1}^{t} c_k f_i(u_k) f_j(u_k) = c_i \delta_{i,j}. \end{aligned}$$

The Gram matrix of $g_1, \dots, g_t$ under the bilinear form induced by W' is therefore the nonsingular diagonal matrix $\operatorname{diag}(c_1, \dots, c_t)$. Hence the g_i are linearly independent, so $t \le t'$. Since $t' \le t$, equality holds and the g_i form a basis of $\mathbb{F}_q^t$.

For $\ell \in [m]$, set

$$D_\ell := \operatorname{diag}((v_1)_\ell, \dots, (v_t)_\ell).$$

The polynomial $x_\ell f_i f_j$ has degree at most $2d + 1$, so the same equivalence gives

$$(D_\ell g_i)^\top W' g_j = (u_i)_\ell c_i \delta_{i,j} = ((u_i)_\ell g_i)^\top W' g_j$$

for every i, j. Since the g_j form a basis and the form induced by W' is nondegenerate,

$$D_\ell g_i = (u_i)_\ell g_i.$$

For each i, the identity $g_i^\top W' g_i = c_i \neq 0$ ensures that g_i has a nonzero coordinate, say k. Reading that coordinate in the preceding eigenvector equation gives $(v_k)_\ell = (u_i)_\ell$ for every

ℓ, and hence $v_k = u_i$. Thus $U \subseteq \{v_1, \dots, v_t\}$; equality of cardinalities gives equality of the two supports.

Relabel so that $v_i = u_i$. Applying $e' \sim_{2d+1} e$ to f_i now gives

$$c_i = \sum_j c_j f_i(u_j) = \sum_j c'_j f_i(v_j) = c'_i.$$

Therefore $e' = e$. □

Corollary 6.34. *Let q be prime, let*

$$0 \le d \le \left\lfloor \frac{m(q-1)-2}{2} \right\rfloor, \qquad 0 \le r \le m,$$

and let $t \le D_{d,q}(r)$. If $e_t \sim \mathcal{E}^t_{q^m,q}$, then with probability at least $1 - tq^{r-m}$, it is the unique error pattern of weight at most t having syndrome $E_q(m, 2d+1)e_t$.

Proof. Identify the q^m coordinates with $\mathbb{F}_q^m$, and write the sampled locations in Eq. (6.3) as $a_1, \dots, a_t$. By Theorem 6.31, the functionals $E_{a_1}, \dots, E_{a_t}$ are linearly independent with probability at least $1 - tq^{r-m}$. On this event the sampled locations are distinct, so e_t has weight t, and Lemma 6.33 makes it the unique weight-at-most-t member of its $\sim_{2d+1}$ class. Lemma 6.32 identifies that class with its syndrome. □

Proof of Corollary 6.21. Let $C = \mathrm{RM}_q[m, 2d+1]$, and let $G = E_q(m, 2d+1)$ be its standard generator matrix. Reed–Muller duality gives

$$C^\perp = \mathrm{RM}_q[m, m(q-1) - (2d+2)],$$

so G is a parity-check matrix of $C^\perp$. Set $\gamma = tq^{r-m}$. By Corollary 6.34, a decoder that searches for the unique weight-at-most-t error with the observed syndrome succeeds with probability at least $1 - \gamma$. The MAP decoder is optimal, and therefore

$$p^t_{\mathrm{decode}}(C^\perp) \ge 1 - \gamma.$$

If $\gamma \ge 1/2$, the desired bound is immediate because its right-hand side $\sqrt{2\gamma}$ is at least 1. If $\gamma < 1/2$, then

$$\sqrt{1 - p^t_{\mathrm{decode}}(C^\perp)^2} \le \sqrt{1 - (1-\gamma)^2} \le \sqrt{2\gamma} = \sqrt{2tq^{r-m}}.$$

This proves the corollary. □

Chapter 7: Learning Structured Quantum Circuits

We now turn from learning quantum states to learning quantum transformations. Given black-box access to an unknown n-qubit unitary U, the general problem is expensive: learning U to diamond distance ϵ requires $\Theta(4^n/\epsilon)$ queries [HKOT23].

As with states, additional structure can make learning efficient. Algorithms are known for several important circuit classes, including Clifford circuits, shallow circuits, fermionic matchgate circuits, and quantum juntas, but these algorithms use rather different techniques. In this chapter, we give a common explanation for many of them.

The key is to study how U acts on Pauli operators under conjugation. If a generating set of Pauli operators remains simple after the transformation

$$P \longmapsto U^\dagger P U,$$

then U itself can be learned efficiently. We measure this simplicity using two properties of the resulting Pauli spectra: sparsity and dimensionality.

7.1 Main Results

7.1.1 The Unifying Framework

Let G be a generating set of Pauli operators. Because conjugation is multiplicative, knowing

$$U^\dagger g U \qquad (g \in G)$$

determines the action of U on every Pauli operator, and hence determines the channel implemented by U. Our framework turns this observation into a learning algorithm when the operators $U^\dagger g U$ have low-complexity Pauli spectra.

We use two notions of complexity. Recall that an operator A has a Pauli expansion

$$A = \sum_x \widehat{A}(x) W_x.$$

Its *Pauli sparsity* is the number of nonzero coefficients $\widehat{A}(x)$, while its *Pauli dimensionality* is the dimension of the span of their support. These are quantum analogues of Fourier sparsity and Fourier dimensionality in Boolean function analysis [GOSSW11].

The following informal statement summarizes the framework; precise versions appear in Theorems 7.47 and 7.49.

Theorem 7.1 (Informal). *Let G be a known generating set of Pauli operators, and suppose that every $U^\dagger g U$, for $g \in G$, has Pauli dimensionality at most k. If these operators can be learned in a phase-fixed, efficiently implementable form, then there is an efficient algorithm which, given access to U and $U^\dagger$, implements the channel of U to diamond distance ϵ with probability at least $1-\delta$, using*

$$\mathrm{poly}(2^k, n)\frac{\log(1/\delta)}{\epsilon}$$

queries.

An analogous statement holds when the conjugated generators are s-Pauli-sparse and can be efficiently rounded to unitaries, with query complexity

$$\mathrm{poly}(s, n)\frac{\log(1/\delta)}{\epsilon^2}.$$

Thus, learning U reduces to learning a collection of individual operators with simple Pauli spectra. Our main technical contribution is to solve this latter problem for both low Pauli dimensionality and low Pauli sparsity.

Theorem 7.2 (Informal). *Let U be an n-qubit k-Pauli-dimensional unitary. There is an efficient algorithm that outputs a unitary V satisfying $\mathsf{d}_\diamond(U, V) \le \epsilon$ with constant success probability using*

$$O(2^k k/\epsilon)$$

queries. Conversely, for $2 \le k \le 2n$, learning all unitaries of Pauli dimension at most k requires $\Omega(2^k/\epsilon)$ queries.

If instead U is s-Pauli-sparse, there is an efficient improper learner that outputs a Pauli-sparse matrix A satisfying $\mathsf{d}_{\mathrm{ph,op}}(U, A) \le \epsilon$ using $O(s^2/\epsilon^2)$ queries. Learning a proper unitary hypothesis requires $\Omega(s/\epsilon)$ queries when $4 \le s \le 4^n$.

The Pauli-dimensional learner outputs a unitary directly. For sparsity, there is an additional difficulty: the learner first produces a structured matrix A close to U, and such

a matrix cannot in general be efficiently rounded to a nearby structured unitary. We identify several settings where this rounding is possible. If the goal is only to implement a channel close to U, the rounding step can instead be avoided using a block-encoded polar decomposition; see Remark 7.45.

At the extreme $k = 2n$, its refined query bound recovers the $\Theta(4^n/\epsilon)$ complexity of unrestricted unitary tomography [HKOT23].

7.1.2 Applications

The framework recovers several known learning results and, more importantly, applies to compositions of circuit classes that were not covered by previous methods. We highlight three examples.

Quantum k-juntas. A quantum k-junta is a unitary that acts nontrivially on only k of its n qubits. Our framework gives a query-optimal learning algorithm.

Corollary 7.3 (Informal). *Let U be an n-qubit quantum k-junta. Given forward-query access to U, there is a time-efficient algorithm that outputs V satisfying $\mathsf{d}_\diamond(U, V) \le \epsilon$ with probability at least $1 - \delta$, using*

$$O\left(\frac{4^k}{\epsilon} \log \frac{1}{\delta}\right)$$

queries. For constant δ, this is optimal up to constant factors.

In particular, the algorithm remains efficient for $k = O(\log n)$.

Compositions of shallow and near-Clifford circuits. Shallow circuits and near-Clifford circuits can each be learned efficiently on their own. Our framework also learns their composition.

Theorem 7.4 (Informal). *Let U be an n-qubit unitary of the form $U = QC$ or $U = CQ$, where Q is a depth-d circuit and C is a Clifford circuit augmented with t single-qubit non-Clifford gates. Then U can be learned efficiently for $d = O(\log \log n)$ and $t = O(\log n)$.*

This is the first result that simultaneously handles these two forms of structure. The dependence on t is exponential and the dependence on d is doubly exponential; our lower bounds show that such behavior is unavoidable for the larger classes considered later.

Compositions of fermionic matchgates and Clifford circuits. Fermionic matchgates are the transformation analogue of the non-interacting fermion states studied in Section 4.1. Previous work gives efficient algorithms for learning matchgate circuits [ODMZ22; CZ26]. Our framework extends these results to compositions with Clifford circuits.

Theorem 7.5 (Informal). *Let U be an n-qubit unitary of the form $U = FC$ or $U = CF$, where F is a fermionic matchgate circuit and C is a Clifford circuit. Then the channel of U can be learned efficiently using $O(n^7/\epsilon^2)$ queries and polynomial time, for constant success probability.*

The framework applies to still more general compositions. For example, it learns constantly many alternating layers of near-Clifford circuits and arbitrary $O(\log n)$-juntas; see Corollary 7.64.

7.2 Preliminaries

We use the Pauli and symplectic notation from Sections 2.5 and 2.5.1, together with the matrix norms from Section 2.3. Here we collect the additional notation and distance measures used specifically for learning unitary transformations.

7.2.1 Pauli Support and Dimensionality

Recall that every n-qubit operator A has a Pauli expansion

$$A = \sum_{x \in \mathbb{F}_2^{2n}} \widehat{A}(x) W_x.$$

We will use two simple measures of the complexity of this expansion: the number of Pauli operators that appear, and the dimension of the subspace that they generate.

Definition 7.6 (Pauli support). The *Pauli support* of $A \in \mathbb{C}^{2^n \times 2^n}$ is

$$\mathrm{supp}(A) := \{x \in \mathbb{F}_2^{2n} \,:\, \widehat{A}(x) \neq 0\}.$$

The size of $\mathrm{supp}(A)$ measures the sparsity of the Pauli expansion. A different notion of complexity asks only how many independent Pauli directions are needed to contain the support.

Definition 7.7 (Pauli dimensionality). The *Pauli dimension* of A is

$$\dim \mathrm{span}(\mathrm{supp}(A)).$$

We say that A is k-Pauli-dimensional if its Pauli dimension is k.

Since every support and span in this chapter is taken with respect to the Pauli basis, we will often simply refer to the support or its span.

The following particular Pauli subspaces will appear repeatedly.

Definition 7.8. For $a, b, n \in \mathbb{N}$ with $a + b \leq n$, define

$$\mathcal{W}_{a,b} := 0^{n-a} \times \mathbb{F}_2^a \times 0^{n-a-b} \times \mathbb{F}_2^{a+b}.$$

As operators, this subspace corresponds to

$$I^{\otimes(n-a-b)} \otimes \{I, Z\}^{\otimes b} \otimes \{I, X, Y, Z\}^{\otimes a}.$$

In words, $\mathcal{W}_{a,b}$ consists of all n-qubit Pauli operators that act trivially on the first $n - a - b$ qubits, act as either I or Z on the next b qubits, and act arbitrarily on the final a qubits.

7.2.2 Distances Between Unitary Channels

Our learning guarantees will be stated in diamond distance, which captures the maximum distinguishability between two quantum channels. For the analysis, however, it will often be more convenient to work with an operator-norm distance after optimizing over global phase.

The diamond distance between two unitaries U and V is the maximum trace distance between the outputs of the corresponding channels, allowing an arbitrary ancilla.

Definition 7.9 (Diamond distance). For unitaries $U, V \in \mathbb{C}^{d\times d}$,

$$\mathsf{d}_\diamond(U,V) \coloneqq \max_\rho \|(I_A \otimes U)\rho(I_A \otimes U^\dagger) - (I_A \otimes V)\rho(I_A \otimes V^\dagger)\|_1$$

where I_A is the identity operator on the ancilla register.

It is often easier to work with the following distance measure, which implies bounds on the diamond distance.

Definition 7.10 (Phase-aligned operator distance). For matrices $A, B \in \mathbb{C}^{d\times d}$,

$$\mathsf{d}_{\mathrm{ph,op}}(A,B) \coloneqq \min_{\theta\in[0,2\pi)} \|e^{\mathrm{i}\theta}A - B\|_{\mathrm{op}}.$$

Fact 7.11 ([HKOT23, Proposition 1.6]). *When $U, V \in \mathbb{C}^{d\times d}$ are unitary,*

$$\mathsf{d}_\diamond(U,V) \le 2\mathsf{d}_{\mathrm{ph,op}}(U,V) \le 2\mathsf{d}_\diamond(U,V).$$

Thus, for unitary channels, controlling phase-aligned operator distance is equivalent up to a constant factor to controlling diamond distance.

Finally, we mention a second important distance measure used in prior work (see [MW16, Section 5.1.1] for details). It can be seen as an average-case distance, since $\mathsf{d}_{\mathrm{phaseF}}(U,V) \le \mathsf{d}_{\mathrm{ph,op}}(U,V) \le \sqrt{d}\cdot\mathsf{d}_{\mathrm{phaseF}}(U,V)$. It is also unitarily invariant.

Definition 7.12 (Phase-aligned normalized Frobenius distance). For unitaries $U, V \in \mathbb{C}^{d\times d}$,

$$\mathsf{d}_{\mathrm{phaseF}}(U,V) = \frac{1}{\sqrt{d}} \min_{\theta\in[0,2\pi)} \|e^{\mathrm{i}\theta}U - V\|_{\mathrm{F}}.$$

7.3 Pauli Projection via Linear Combination of Unitaries

In this section, we introduce Pauli projection, which lets us "filter out" certain Pauli operators from the spectrum of a unitary. Specifically, given a subspace of Pauli operators S and query access to a unitary U, our goal is to apply the operator obtained from U by zeroing out all of its Pauli coefficients outside of S. The procedure relies on block encodings and the linear combination of unitaries technique.

Definition 7.13 (Pauli projection). Let $n \in \mathbb{N}$, let $A \in \mathbb{C}^{2^n \times 2^n}$ be a matrix, and let $S \subseteq \mathbb{F}_2^{2n}$ be a subspace. The Pauli projection of A onto S, denoted $\Pi_S(A)$, is defined as

$$\Pi_S(A) := \frac{1}{2^n} \sum_{x \in S} \operatorname{tr}(AW_x) W_x.$$

In other words, we zero out all Pauli coefficients of A that are outside of S.

The Pauli projection of a unitary channel can be expressed as a convex combination of unitaries. In fact, it can be written as a Pauli twirl.

Lemma 7.14. *Let $U \in \mathbb{C}^{2^n \times 2^n}$ be a unitary matrix, and let $S \subseteq \mathbb{F}_2^{2n}$ be a subspace. The Pauli projection of U onto S can be expressed as a Pauli twirl:*

$$\Pi_S(U) = \mathop{\mathbb{E}}_{q \sim S_s^\perp} [W_q U W_q^\dagger],$$

where the Pauli operator W_q is chosen uniformly at random from $S_s^\perp$.

Proof. Let W_x be an arbitrary Weyl operator. We have

$$\mathop{\mathbb{E}}_{q \sim S_s^\perp} [W_q W_x W_q^\dagger] = \mathop{\mathbb{E}}_{q \sim S_s^\perp} [(-1)^{[q,x]}] W_x. \tag{7.1}$$

If $x \in S$, then $\mathbb{E}_q[(-1)^{[q,x]}] = 1$. However, if $x \notin S$, then $\mathbb{E}_q[(-1)^{[q,x]}] = 0$ by Proposition 2.12. Hence, by linearity, we have that, for any operator A, the Pauli twirl will zero out all $W_x \notin S$, which is precisely the action of $\Pi_S(\cdot)$. □

Next, we recall a standard tool in quantum algorithms: block encodings.

Definition 7.15 (Block encoding; see [TK23, Definition 1.1], [GSLW19, Definition 43], and [Ral20, Definition 1]). Let $A \in \mathbb{C}^{r \times c}$, and let $U \in \mathbb{C}^{2^n \times 2^n}$ be a unitary matrix. We say that U is a *Q-block encoding* of A if U is implementable with $O(Q)$ gates and has the form

$$U = \begin{pmatrix} A & \cdot \\ \cdot & \cdot \end{pmatrix},$$

where $\cdot$ denotes unspecified block matrices.

Lemma 7.16 ([TK23, Lemma 1.5]). *Let $U \in \mathbb{C}^{2^n \times 2^n}$ be a Q-block encoding of $A \in \mathbb{C}^{r \times c}$, and let $|\psi\rangle \in \mathbb{C}^c$ be an input state. Then there is a quantum circuit using 1 query to U (and therefore $O(Q)$ gates) that prepares the state $\frac{A|\psi\rangle}{\|A|\psi\rangle\|_2}$ with probability $\|A|\psi\rangle\|_2^2$.*

We use the linear combination of unitaries (LCU) method to construct block encodings of more general operators. This technique allows one to turn a weighted sum of unitaries into a block encoding of the corresponding operator.

Lemma 7.17 ([CKS17; GSLW19]). *Let $A = \sum_k a_k U_k$ be a linear combination of unitary operators where $a_k \geq 0$. Define the preparation operator* PREP *as*

$$\mathrm{PREP}|0\rangle = \sum_k \sqrt{\frac{a_k}{\sum_k a_k}}|k\rangle,$$

and the select operator SEL *as*

$$\mathrm{SEL}|k\rangle|\psi\rangle = |k\rangle U_k|\psi\rangle.$$

Then the circuit $\mathrm{PREP}^\dagger \cdot \mathrm{SEL} \cdot \mathrm{PREP}$ is a block encoding of $\frac{A}{\sum_k a_k}$.

We are now ready to present the main algorithm of this section. In particular, given a single query to U and a subspace of Pauli operators S, one can perform the mapping

$$|\psi\rangle \longmapsto \frac{\Pi_S(U)|\psi\rangle}{\|\Pi_S(U)|\psi\rangle\|_2}$$

with postselection.

Lemma 7.18. *Let $U \in \mathbb{C}^{2^n \times 2^n}$ be a unitary matrix, let $S \subseteq \mathbb{F}_2^{2n}$ be a subspace, and let $|\psi\rangle$ be an n-qubit input state. There is a quantum algorithm that uses a single query to U and prepares the state $\frac{\Pi_S(U)|\psi\rangle}{\|\Pi_S(U)|\psi\rangle\|_2}$ with probability $\|\Pi_S(U)|\psi\rangle\|_2^2$. The algorithm uses $O(n \cdot \dim(S_{\mathrm{s}}^\perp)) = O(n^2)$ gates and requires $\dim(S_{\mathrm{s}}^\perp) \leq 2n$ ancilla qubits.*

Proof. By Lemma 7.14, we can write $\Pi_S(U) = \mathbb{E}_{q \sim S_{\mathrm{s}}^\perp}[W_q U W_q^\dagger]$, where the expectation is over uniformly random $q \in S_{\mathrm{s}}^\perp \subseteq \mathbb{F}_2^{2n}$. This is a convex combination of unitaries, so the LCU normalization factor in Lemma 7.17 is one. Let $\ell = \dim S_{\mathrm{s}}^\perp$, choose a basis for $S_{\mathrm{s}}^\perp$, and

identify its 2^ℓ elements with bit strings $k \in \{0,1\}^\ell$. Let PREP denote a unitary that maps $|0^\ell\rangle$ to the uniform superposition, i.e., Hadamard gates on ℓ qubits. Let

$$C_W := \sum_k |k\rangle\langle k| \otimes W_k, \qquad \text{SEL} := C_W(I^{\otimes \ell} \otimes U)C_W^\dagger.$$

Thus $\text{SEL}|k\rangle|\psi\rangle = |k\rangle W_k U W_k^\dagger |\psi\rangle$, and it uses two known controlled-Pauli operations and one uncontrolled query to U. Then, by Lemma 7.17, $\text{PREP}^\dagger \cdot \text{SEL} \cdot \text{PREP}$ is a block encoding of $\Pi_S(U)$. Finally Lemma 7.16 implies that we can prepare $\frac{\Pi_S(U)|\psi\rangle}{\|\Pi_S(U)|\psi\rangle\|_2}$ with probability $\|\Pi_S(U)|\psi\rangle\|_2^2$.

Constructing the unitary $|k\rangle\langle k| \otimes W_k$ can be done by first taking generators of $S_s^\perp = \text{span}\{g_1, \ldots, g_\ell\}$. Then control each W_{g_i} on qubit i of the control register. Each controlled W_{g_i} can be constructed as the product of at most $2n$ controlled single-qubit Pauli operations (with the control, this becomes a two-qubit gate), for a total of at most $O(n\ell)$ two-qubit gates. □

7.4 Learning Approximately Block-Diagonal Unitaries

In this section, we develop an algorithm for learning unitary channels that are approximately block-diagonal. The algorithm for learning k-Pauli-dimensional unitary channels, which we present in Section 7.6, will follow from a reduction to the approximately block-diagonal setting. We begin by formally defining "approximately block-diagonal."

Definition 7.19 (Block-diagonal matrix). Let $a, b, n \in \mathbb{N}$ with $a + b \leq n$, and let $A \in \mathbb{C}^{2^n \times 2^n}$ be a matrix. We say that A is (a,b)-block-diagonal if there exist 2^b matrices $A_y \in \mathbb{C}^{2^a \times 2^a}$ such that $A = I^{\otimes n-a-b} \otimes \left(\bigoplus_{y\in\{0,1\}^b} A_y\right)$. That is, A is a block-diagonal matrix whose diagonal blocks are A_1, ..., A_{2^b}, each of size $2^a \times 2^a$, and this $2^{a+b} \times 2^{a+b}$ block structure is repeated 2^{n-a-b} times along the diagonal.

Let $A = I^{\otimes n-a-b} \otimes \left(\bigoplus_{y\in\{0,1\}^b} A_y\right)$ be an (a,b)-block-diagonal matrix. It is helpful to view A as acting on n qubits, grouped into three registers: the first $n-a-b$ qubits are unaffected by A; the second b qubits select which diagonal block we are in; and the last a qubits index a column within that block. For $x \in \{0,1\}^{n-a-b}$, $y \in \{0,1\}^b$, and $z \in \{0,1\}^a$, we have $A(|x\rangle|y\rangle|z\rangle) = |x\rangle|y\rangle(A_y|z\rangle)$.

Definition 7.20 (Block-diagonal unitary). A matrix $U \in \mathbb{C}^{2^n \times 2^n}$ is an (a, b)-block-diagonal unitary if it is both unitary and an (a, b)-block-diagonal matrix.

Definition 7.21 (Approximately block-diagonal unitary). Let $a, b, n \in \mathbb{N}$ with $a + b \leq n$, and let $U \in \mathbb{C}^{2^n \times 2^n}$ be a unitary matrix. We say U is (a, b, ϵ)-approximately block-diagonal (with respect to a given matrix norm $\|\cdot\|$) if there exists an (a, b)-block-diagonal unitary V satisfying $\|U - V\| \leq \epsilon$.

Unless otherwise stated, (a, b, ϵ)-approximately block-diagonal unitaries are always with respect to the operator norm $\|\cdot\|_{\mathrm{op}}$.

With these definitions in place, we can now state the main technical result of this section: a tomography algorithm that efficiently estimates approximately block-diagonal unitaries in phase-aligned operator distance, and hence in diamond distance up to a constant factor.

Theorem 7.22. *Let $a, b, n \in \mathbb{N}$ with $a + b \leq n$, and let $\epsilon, \delta \in (0, 1)$. There is a tomography algorithm that, given query access to an (a, b, ϵ)-approximately block-diagonal unitary channel $U \in \mathbb{C}^{2^n \times 2^n}$, applies U at most $O\Big(2^{a+b} \cdot (2^a + b)\frac{\log(1/\delta)}{\epsilon^2}\Big)$ times and outputs an estimate V satisfying $\mathsf{d}_{\mathrm{ph,op}}(U, V) \leq O(\epsilon)$ with probability at least $1 - \delta$. Moreover:*

- *The output V is an (a, b)-block-diagonal unitary.*
- *The algorithm runs in time* $\mathrm{poly}\big(n,\ 2^{2a+b},\ \epsilon^{-1},\ \log \delta^{-1}\big)$.
- *The algorithm uses only forward access to U (i.e., it does not require $U^\dagger$ or controlled-U).*
- *The algorithm uses $2n - 2a - b$ additional qubits of space.*

For constant accuracy and success probability, parameter counting gives a lower bound of $\Omega(2^{2a+b})$ queries. The upper bound above is

$$O\Big(2^{2a+b} + b2^{a+b}\Big),$$

so it matches the lower bound when $b = O(2^a)$; the possible gap is confined to the highly isotropic regime $b \gg 2^a$. Achieving this scaling requires some care. Ideally, we would simply

treat U as truly block diagonal and learn all of its relevant blocks in parallel. Learning the blocks sequentially would require $O(4^{a+b})$ queries to recover their relative phases. However, because U is only ϵ-close to block-diagonal, its off-diagonal blocks behave like noise, and a naïve parallel approach will fail.

To overcome this, we replace U with a Pauli projection $\Pi_{\mathcal{W}_{a,b}}(U)$ (Definitions 7.8 and 7.13) that has three crucial properties: (i) it is exactly (a,b)-block diagonal, (ii) it is within 2ϵ of U in operator norm, and (iii) it can be efficiently applied using LCUs and block encodings. With this operator in hand, our algorithm essentially applies $\Pi_{\mathcal{W}_{a,b}}(U)$ to states of our choice and then uses state tomography to recover the columns of all blocks of $\Pi_{\mathcal{W}_{a,b}}(U)$ simultaneously, thereby attaining the displayed $O(2^{a+b}(2^a+b))$ dependence.

We now present our learning algorithm.

Algorithm 11 Learning approximately block-diagonal unitaries

Input: Query access to an (a,b,ϵ)-approximately block-diagonal unitary U, where $\epsilon \le 1/K$ for a universal constant $K \ge 1$

Output: A description of an (a,b)-block-diagonal unitary V such that $\mathsf{d}_{\mathrm{ph,op}}(U,V) \le O(\epsilon)$ with probability at least $2/3$

for all $z \in \{0,1\}^a$ **do**
  Apply $\Pi_{\mathcal{W}_{a,b}}(U)$ to $|0^{n-a-b}\rangle|+\rangle^{\otimes b}|z\rangle$ using Lemma 7.18
  Discard the first register, leaving the $(a+b)$-qubit state $|\psi_z\rangle$
  Run the pure-state tomography algorithm of Lemma 7.27 on $|\psi_z\rangle$ with error $\epsilon\sqrt{2^a/(2^a+b)}$ and failure probability $\exp(-5\cdot 2^{a+b})$; denote its output by $|\widehat{\psi_z}\rangle$
end for
for all $y \in \{0,1\}^b$ **do**
  Construct $A_y \in \mathbb{C}^{2^a\times 2^a}$ whose zth column is $\sqrt{2^b}(\langle y|\otimes I^{\otimes a})|\widehat{\psi_z}\rangle$
  Compute an SVD $A_y = L_y\Sigma_y R_y^\dagger$ and set $V_y = L_y R_y^\dagger$
end for
Set $V = I^{\otimes n-a-b}\otimes\left(\bigoplus_{y\in\{0,1\}^b} V_y\right)$
Apply the preceding procedure to $U(I^{\otimes n-a}\otimes H^{\otimes a})$ and denote the result by $V' = I^{\otimes n-a-b}\otimes\left(\bigoplus_{y\in\{0,1\}^b} V'_y\right)$
Let $D \in \mathbb{C}^{2^a\times 2^a}$ be the diagonal unitary obtained by applying Lemma 7.28 to V_0 and V'_0
return $V(I^{\otimes n-a}\otimes D) = I^{\otimes n-a-b}\otimes\left(\bigoplus_{y\in\{0,1\}^b} V_y D\right)$

The first two lemmas show that this projection is exactly (a,b)-block diagonal, remains close to U, and can be rounded blockwise to a unitary.

Fact 7.23. *For a matrix A, the following are all equivalent conditions:*

1. *A is (a,b)-block diagonal.*
2. $\mathrm{supp}(A) \subseteq \mathcal{W}_{a,b}$.
3. $\Pi_{\mathcal{W}_{a,b}}(A) = A$.

Proof. We start by showing that the first two conditions are equivalent. Suppose A is an (a,b)-block diagonal matrix. Then

$$A = I^{\otimes n-a-b} \otimes \left(\bigoplus_{y \in \{0,1\}^b} A_y \right) = I^{\otimes n-a-b} \otimes \sum_{y \in \{0,1\}^b} |y\rangle\langle y| \otimes A_y,$$

where each $A_y \in \mathbb{C}^{2^a \times 2^a}$. It is easy to check that the b-qubit operator $|y\rangle\langle y|$ is only supported on $\{I,Z\}^{\otimes b}$ for all $y \in \{0,1\}^b$. Each A_y is trivially supported on $\{I,X,Y,Z\}^{\otimes a}$, because $\{I,X,Y,Z\}^{\otimes a}$ is a complete basis for operators on a qubits. Hence, $\mathrm{supp}(A) \subseteq \mathcal{W}_{a,b}$. Conversely, suppose that A is supported in $\mathcal{W}_{a,b}$. Each operator in $\mathcal{W}_{a,b}$ is (a,b)-block diagonal, and so any linear combination of them is also (a,b)-block diagonal.

Now we show that conditions 2 and 3 are equivalent. Let A be an operator with $\mathrm{supp}(A) \subseteq \mathcal{W}_{a,b}$. Then it is obvious that $\Pi_{\mathcal{W}_{a,b}}$ will have no effect on A, because it removes only Pauli coefficients outside of $\mathcal{W}_{a,b}$. Conversely, if $\Pi_{\mathcal{W}_{a,b}}(A) = A$, then all of the Pauli coefficients outside of $\mathcal{W}_{a,b}$ must be 0, i.e., $\mathrm{supp}(A) \subseteq \mathcal{W}_{a,b}$. □

Lemma 7.24. *If U is (a,b,ϵ)-approximately block diagonal for any unitarily invariant norm $\|\cdot\|$, then $\|U - \Pi_{\mathcal{W}_{a,b}}(U)\| \leq 2\epsilon$.*

Proof. Because U is (a,b,ϵ)-approximately block diagonal, there exists an (a,b)-block diagonal unitary V such that $\|U - V\| \leq \epsilon$. Then

$$\begin{aligned}
\|U - \Pi_{\mathcal{W}_{a,b}}(U)\| &\leq \|U - V\| + \|V - \Pi_{\mathcal{W}_{a,b}}(U)\| && (7.2)\\
&= \|U - V\| + \|\Pi_{\mathcal{W}_{a,b}}(V - U)\| && (7.3)\\
&= \|U - V\| + \|\mathop{\mathbb{E}}_q \left[W_q (V - U) W_q \right]\| && (7.4)\\
&\leq \|U - V\| + \mathop{\mathbb{E}}_q \left[\|W_q (V - U) W_q\| \right] && (7.5)
\end{aligned}$$

$$= \|U - V\| + \|U - V\| \tag{7.6}$$

$$\leq 2\epsilon. \tag{7.7}$$

The first line uses the triangle inequality. The second line uses the fact that $\Pi_{\mathcal{W}_{a,b}}(V) = V$ by Fact 7.23. The third line follows from Lemma 7.14. The fourth line is the triangle inequality. The fifth line follows from the fact that the norm is unitarily invariant. □

Next, we show that if a unitary U is close to some (a, b)-block diagonal matrix (not necessarily unitary), then U is also close to a genuine (a, b)-block diagonal unitary. Thus the approximate block-diagonal structure can be rounded to an exact one without losing more than a constant factor in error. Moreover, this rounding can be done in polynomial time in the matrix dimension. In the learning algorithm, the nonunitary matrix below is the estimate of $\Pi_{\mathcal{W}_{a,b}}(U)$ before its blocks are rounded by an SVD.

Lemma 7.25. *Let U be a unitary and suppose there exists an (a, b)-block diagonal matrix A (not necessarily unitary) such that $\|U - A\| \leq \epsilon$ for some unitarily invariant norm $\|\cdot\|$. Then U is an $(a, b, 2\epsilon)$-approximately block diagonal unitary.*

Moreover, given a classical description of A, there is an algorithm that outputs an (a, b)-block diagonal unitary V satisfying $\|U - V\| \leq 2\epsilon$, and this algorithm runs in $O(2^{3a+b})$ time.

Proof. We can write A in block form as $A = I^{\otimes n-a-b} \otimes \left(\bigoplus_{y\in\{0,1\}^b} A_y\right)$. For each block $A_y \in \mathbb{C}^{2^a \times 2^a}$, compute its SVD $A_y = L_y \Sigma_y R_y^\dagger$. Define $V := I^{\otimes n-a-b} \otimes \left(\bigoplus_{y\in\{0,1\}^b} L_y R_y^\dagger\right)$. By construction, V is (a, b)-block diagonal, and, in particular, it is the unitary obtained from the polar decomposition of A.

It is a standard fact that, in any unitarily invariant norm, the unitary from the polar decomposition is the closest unitary to a given matrix. Thus,

$$\|V - A\| \leq \|U - A\| \leq \epsilon,$$

since U is also unitary. The triangle inequality gives

$$\|U - V\| \leq \|U - A\| + \|A - V\| \leq 2\epsilon,$$

so U is an $(a, b, 2\epsilon)$-approximately block diagonal unitary.

Finally, the runtime: computing an SVD (or equivalently, the polar decomposition) of a $2^a \times 2^a$ matrix takes $O(2^{3a})$ time. Since there are 2^b blocks, the total cost is $O(2^{3a+b})$. □

Together, Lemmas 7.24 and 7.25 show that the following two conditions are equivalent up to constant factors:

1. U is close to some (a, b)-block-diagonal unitary V, and
2. U is close to its Pauli projection $\Pi_{\mathcal{W}_{a,b}}(U)$.

In what follows, we work with U through its Pauli projection.

The next lemma analyzes the cost of applying the Pauli projection $\Pi_{\mathcal{W}_{a,b}}(U)$. In particular, it bounds the number of queries to U required in order to generate a sufficient number of copies of the projected state for use in the state tomography procedure.

Lemma 7.26. *Let $0 < \epsilon < 1/2$, $\delta \in (0, 1)$, and let U be an (a, b, ϵ)-approximately block-diagonal unitary. For any n-qubit state $|\psi\rangle$ and any constant $c > 0$, set*

$$d := \left\lceil \frac{c\, 2^{a+b}}{\epsilon^2} \right\rceil, \qquad M := \left\lceil \frac{2}{(1-2\epsilon)^2} \left(d + \log \frac{1}{\delta} \right) \right\rceil.$$

There is a procedure that, given access to copies of $|\psi\rangle$ and using M queries to U, prepares d copies of the state

$$\frac{\Pi_{\mathcal{W}_{a,b}}(U)|\psi\rangle}{\|\Pi_{\mathcal{W}_{a,b}}(U)|\psi\rangle\|_2}$$

with probability at least $1 - \delta$.

Proof. By Lemmas 7.16 and 7.17, we successfully prepare a single copy of the desired state using a single query to U with probability $\|\Pi_{\mathcal{W}_{a,b}}(U)|\psi\rangle\|_2^2$. By Lemma 7.24, $\|\Pi_{\mathcal{W}_{a,b}}(U)|\psi\rangle\|_2 \geq 1 - 2\epsilon$. The success probability is therefore at least $(1 - 2\epsilon)^2$.

Repeat the procedure for the stated M trials. Applying Lemma 2.8 with $p = (1 - 2\epsilon)^2$ shows that at least d trials succeed except with probability at most δ. □

We need two further technical ingredients from [HKOT23]. The first is a state tomography algorithm whose error lies in a Haar-random direction. The second is a post-processing routine that aligns the relative phases of the reconstructed unitary's columns.

Lemma 7.27 ([HKOT23, Proposition 2.2] and [GKKT20, Theorem 2, Section 4.3]). *There is a pure state tomography algorithm with the following behavior. Given access to copies of an n-qubit pure state $|\psi\rangle$, it sequentially and nonadaptively makes von Neumann measurements on $O\left(\frac{2^n+\log(1/\delta)}{\epsilon^2}\right)$ copies of $|\psi\rangle$. Then, after classically collating and processing the measurement outcomes in $O(8^n)$ time, it outputs (a classical description of) an estimate pure state*

$$|\widehat{\psi}\rangle = \phi\sqrt{1-\hat{\epsilon}^2}|\psi\rangle + \hat{\epsilon}|w\rangle$$

such that, except with probability at most δ: (1) ϕ is a complex phase; (2) the trace distance $\hat{\epsilon}$ is at most ϵ; (3) the vector $|w\rangle$ is distributed Haar-randomly among all states orthogonal to $|\psi\rangle$.

Lemma 7.28 (Proof of [HKOT23, Proposition 2.3]). *Let $V, V' \in \mathbb{C}^{2^n\times 2^n}$ be classical descriptions of unitary matrices. Suppose there exist diagonal unitaries $\Phi_V, \Phi_{V'}$ and an operator $A \in \mathbb{C}^{2^n\times 2^n}$ such that $\|A\Phi_V - V\|_{\mathrm{op}} \le \epsilon \le \frac{1}{8}$ and $\|AH^{\otimes n}\Phi_{V'} - V'\|_{\mathrm{op}} \le \epsilon \le \frac{1}{8}$. Then there is an algorithm that outputs a description of a diagonal unitary D such that $\mathsf{d}_{\mathrm{ph,op}}(D, \Phi_V^\dagger) \le 24\epsilon$ using $O(8^n)$ classical time.*

The proof of [HKOT23, Proposition 2.3] assumes that A is unitary. However, the argument does not use this, and in fact the proof goes through for arbitrary A.

We now prove Theorem 7.22 by analyzing Algorithm 11.

Proof. We may assume $\epsilon \le 1/C$ for some constant $C > 1$ to be fixed later. If instead $\epsilon > 1/C$, output the identity: it is (a,b)-block diagonal and $\mathsf{d}_{\mathrm{ph,op}}(U, I) \le 2 \le 2C\epsilon$, while the resource bounds are immediate. In addition, we assume $a+b$ is a sufficiently large constant so that, for a universal constant C' to be specified later, $\frac{C'(a+1)}{2^{a+b}} \le \frac{1}{2}$. Note that if $a+b = O(1)$, then our desired query complexity can be achieved by naïvely learning the entire $2^{a+b}\times 2^{a+b}$ block to Frobenius distance (as in [CNY23]). This would still have $O(8^{a+b}/\epsilon^2) = O(1/\epsilon^2)$ query complexity as desired.

By Fact 7.23, the Pauli projection $\Pi_{\mathcal{W}_{a,b}}(U)$ can be expressed in block-diagonal form as

$$\Pi_{\mathcal{W}_{a,b}}(U) = I^{\otimes n-a-b} \otimes \bigoplus_{y\in\{0,1\}^b} B_y. \tag{7.8}$$

Since $\Pi_{\mathcal{W}_{a,b}}(U)$ is obtained via the block-encoding in Lemma 7.18, it is a contraction:

$$\|\Pi_{\mathcal{W}_{a,b}}(U)\|_{\mathrm{op}} \leq 1.$$

Furthermore, Lemma 7.24 gives that $\|\Pi_{\mathcal{W}_{a,b}}(U) - U\|_{\mathrm{op}} \leq 2\epsilon$. Thus for every normalized state $|\psi\rangle$, $\|\Pi_{\mathcal{W}_{a,b}}(U)|\psi\rangle\|_2 \geq 1 - 2\epsilon$.

Define $\alpha_z := \|\Pi_{\mathcal{W}_{a,b}}(U)|0^{n-a-b}\rangle|+\rangle^{\otimes b}|z\rangle\|_2 \geq 1 - 2\epsilon$. In Algorithm 11, we apply the postselection procedure of Lemma 7.18 to prepare copies of the normalized state

$$\frac{\Pi_{\mathcal{W}_{a,b}}(U)|0^{n-a-b}\rangle|+\rangle^{\otimes b}|z\rangle}{\alpha_z} = |0^{n-a-b}\rangle\left(\frac{1}{\sqrt{2^b}}\sum_{y\in\{0,1\}^b}|y\rangle \otimes \frac{B_y}{\alpha_z}|z\rangle\right) =: |0^{n-a-b}\rangle|\psi_z\rangle \tag{7.9}$$

which are then passed to the tomography step (Algorithm 11). By Lemma 7.26,

$$O\left(\frac{2^{a+b}}{\epsilon^2(1-2\epsilon)^2}\frac{2^a+b}{2^a}\right) = O\left(\frac{2^{a+b}}{\epsilon^2}\left(1+\frac{b}{2^a}\right)\right)$$

queries to U suffice for each z, with failure probability at most $\exp(-5\cdot 2^{a+b})$. A union bound gives simultaneous success for all 2^a columns with constant probability. The overall query complexity is therefore $O\left(2^{a+b}(2^a+b)/\epsilon^2\right)$.

For each $z \in \{0,1\}^a$, the tomography call in Algorithm 11 produces a classical approximation $|\widehat{\psi_z}\rangle$ of the state $|\psi_z\rangle$ defined in Eq. (7.9). Conditioned on the tomography succeeding, the algorithm returns a state of the form

$$|\widehat{\psi_z}\rangle = \phi_z\sqrt{1-\epsilon_z^2}\left(\frac{1}{\sqrt{2^b}}\sum_{y\in\{0,1\}^b}|y\rangle \otimes \frac{B_y}{\alpha_z}|z\rangle\right) + \epsilon_z|w_z\rangle = \phi_z\sqrt{1-\epsilon_z^2}|\psi_z\rangle + \epsilon_z|w_z\rangle \tag{7.10}$$

where $|w_z\rangle$ is an $(a+b)$-qubit Haar-random state orthogonal to $|\psi_z\rangle$, $\epsilon_z \leq \epsilon\sqrt{\frac{2^a}{2^{a+b}}}$ is the accuracy of the tomography algorithm for the zth column, and ϕ_z is a random global phase. By Lemma 7.27, the tomography algorithm succeeds with probability at least $1-\exp(-5\cdot 2^{a+b})$. By a union bound, all 2^a tomography algorithms succeed simultaneously except with probability at most $1/100$. In what follows, we condition on this success event.

Recall that our goal is to recover the matrices B_y that appear in Eq. (7.8). The next step of the algorithm (Algorithm 11) is to collate our state estimates (Algorithm 11) into block

matrices A_y. For the analysis, express each A_y in the form

$$A_y = B_y D \Phi E + R_y F,$$

where the matrices D, Φ, E, R_y, F are defined as follows. $D = \mathrm{diag}(\{\alpha_z^{-1}\})$ is the diagonal matrix of scaling factors α_z^{-1}; $\Phi = \mathrm{diag}(\{\phi_z\})$ is the diagonal matrix of phases ϕ_z; $E \coloneqq \mathrm{diag}(\{\sqrt{1-\epsilon_z^2}\}_z)$ and $F \coloneqq \mathrm{diag}(\{\epsilon_z\}_z)$ are the diagonal matrices of error terms. Lastly, for each $y \in \{0,1\}^b$, R_y is the error matrix whose columns are given by

$$\sqrt{2^b} \cdot (\langle y| \otimes I)|w_z\rangle.$$

We next bound the operator norms of the random error matrices R_y. Introduce independent uniform angles $\theta_z \in [0, 2\pi)$ and independent random variables ξ_z, each distributed as $|\langle 0|\mathrm{Haar}\rangle|^2$ for a Haar-random state in dimension 2^{a+b}. Define

$$|h_z\rangle \coloneqq \sqrt{\xi_z} e^{i\theta_z}|\psi_z\rangle + \sqrt{1-\xi_z}|w_z\rangle.$$

By construction, the vectors $|h_z\rangle$ are independent Haar-random states on $a+b$ qubits. Let $M \in \mathbb{C}^{2^{a+b}\times 2^a}$ be the matrix whose columns are the $|h_z\rangle$, and for each $y \in \{0,1\}^b$, let

$$M_y \coloneqq (\langle y| \otimes I^{\otimes a})M.$$

Standard random-matrix estimates [Ver18, Theorems 3.4.6 and 4.6.1] show that, for a fixed y,

$$\sqrt{2^b}\,\|M_y\|_{\mathrm{op}} \le O\left(\sqrt{\frac{2^a+b}{2^a}}\right)$$

except with probability at most $\exp(-5(2^a+b))$. A union bound gives this estimate simultaneously for all y, except with probability at most $\exp(-5\cdot 2^a)$.

From the definition of $|h_z\rangle$, we have

$$\sqrt{2^b} M_y = B_y D \Delta_0 + R_y \Delta_1, \qquad \Delta_0 = \mathrm{diag}\big(\{\sqrt{\xi_z} e^{i\theta_z}\}_z\big), \quad \Delta_1 = \mathrm{diag}\big(\{\sqrt{1-\xi_z}\}_z\big),$$

and hence

$$\|R_y\|_{\mathrm{op}} \le \frac{\sqrt{2^b}\,\|M_y\|_{\mathrm{op}} + \|B_y D\|_{\mathrm{op}} \max_z \sqrt{\xi_z}}{\sqrt{1-\max_z \xi_z}}.$$

The variables ξ_z are subexponential on scale $2^{-(a+b)}$. Standard tail bounds [Ver18, Proposition 2.7.1, Lemma 2.7.6, and Theorem 3.4.6] imply that

$$\max_z \xi_z \le \frac{C'(a+1)}{2^{a+b}}$$

with constant failure probability, for a universal constant $C' > 0$. By our initial assumption this quantity is at most $1/2$. Moreover, $\|B_y D\|_{\mathrm{op}} \le (1-2\epsilon)^{-1}$. Combining these bounds gives

$$\|R_y\|_{\mathrm{op}} \le O\left(\sqrt{\frac{2^a+b}{2^a}}\right) \tag{7.11}$$

for every y, with constant overall success probability. This is the same use of random-matrix concentration as in [HKOT23].

It follows that, for every $y \in \{0,1\}^b$,

$$\begin{aligned}\|B_y\Phi - A_y\|_{\mathrm{op}} &\le \|B_y - B_y D\|_{\mathrm{op}} + \|B_y D\Phi - A_y\|_{\mathrm{op}} \\ &\le \|B_y - B_y D\|_{\mathrm{op}} + \|B_y D\|_{\mathrm{op}}\|I - E\|_{\mathrm{op}} + \|R_y\|_{\mathrm{op}}\|F\|_{\mathrm{op}} \\ &\le O(\epsilon). \end{aligned} \tag{7.12}$$

Here

$$\|I - E\|_{\mathrm{op}} \le \max_z \epsilon_z^2, \qquad \|F\|_{\mathrm{op}} \le \epsilon\sqrt{\frac{2^a}{2^a+b}},$$

while $\|B_y\|_{\mathrm{op}} \le 1$ and submultiplicativity give

$$\|B_y - B_y D\|_{\mathrm{op}} \le \|I - D\|_{\mathrm{op}} \le \frac{2\epsilon}{1-2\epsilon}.$$

It remains to justify rounding A_y, since $B_y\Phi$ need not itself be unitary. Let $Q = I^{\otimes n-a-b} \otimes \bigoplus_y Q_y$ be the block-diagonal unitary promised by the hypothesis. The triangle inequality gives $\|B_y - Q_y\|_{\mathrm{op}} \le 3\epsilon$. Consequently A_y is $O(\epsilon)$-close to the unitary $Q_y\Phi$. Its polar factor V_y is no farther from A_y than $Q_y\Phi$ is, so another triangle inequality gives

$$\|V_y - B_y\Phi\|_{\mathrm{op}} = O(\epsilon). \tag{7.13}$$

We now correct for the relative phases in Φ. By the group structure of $\mathcal{W}_{a,b}$, we have that

$$\Pi_{\mathcal{W}_{a,b}}\big(U(I^{\otimes n-a}\otimes H^{\otimes a})\big) = \Pi_{\mathcal{W}_{a,b}}(U)(I^{\otimes n-a}\otimes H^{\otimes a}) = I^{\otimes n-a-b}\otimes\left(\bigoplus_{y\in\{0,1\}^b} B_y H^{\otimes a}\right)$$

is an (a,b)-block diagonal matrix that is close to $U(I^{\otimes n-a}\otimes H^{\otimes a})$. By the same reasoning as above, Algorithm 11 recovers an (a,b)-block-diagonal unitary whose blocks V'_y are $O(\epsilon)$-close to $B_y H^{\otimes a}\Phi'$, and in particular at most $1/8$-far when the universal constant in the hypothesis is chosen sufficiently large, for another diagonal matrix of phases Φ'. Hence, invoking Lemma 7.28 on V_0 and V'_0 in Algorithm 11 yields a diagonal unitary D such that $\mathsf{d}_{\mathrm{ph,op}}(\Phi^\dagger, D)\le O(\epsilon)$.

Let $V = I^{\otimes n-a-b}\otimes\bigoplus_{y\in\{0,1\}^b} V_y$ be the output of Algorithm 11. The phase in the definition of $\mathsf{d}_{\mathrm{ph,op}}(\Phi^\dagger, D)$ is common to every block. Choosing that same phase in the following distance and using the direct-sum operator norm gives

$$\begin{aligned}
&\mathsf{d}_{\mathrm{ph,op}}\big(\Pi_{\mathcal{W}_{a,b}}(U), V(I^{\otimes n-a}\otimes D)\big)\\
&\quad\le \max_{y\in\{0,1\}^b}\|B_y - V_y\Phi^\dagger\|_{\mathrm{op}} + \mathsf{d}_{\mathrm{ph,op}}(\Phi^\dagger, D)\\
&\le O(\epsilon). \qquad (7.14)
\end{aligned}$$

The last line follows from Eq. (7.13) and the fact that $\mathsf{d}_{\mathrm{ph,op}}(\Phi^\dagger, D)\le O(\epsilon)$. We note that it is essential to invoke Lemma 7.28 on a $2^a\times 2^a$ block (rather than the entire unitary) to avoid an exponential dependence on n in the final query and time complexity.

Finally, the triangle inequality bounds the distance from the algorithm's output to U:

$$\begin{aligned}
\mathsf{d}_{\mathrm{ph,op}}(U, V(I^{\otimes n-a}\otimes D)) &\le \mathsf{d}_{\mathrm{ph,op}}\big(U, \Pi_{\mathcal{W}_{a,b}}(U)\big) + \mathsf{d}_{\mathrm{ph,op}}\big(\Pi_{\mathcal{W}_{a,b}}(U), V(I^{\otimes n-a}\otimes D)\big)\\
&\le \|U - \Pi_{\mathcal{W}_{a,b}}(U)\|_{\mathrm{op}} + O(\epsilon)\\
&\le O(\epsilon).
\end{aligned}$$

The second line follows from Eq. (7.14), and the last line follows from Lemma 7.24.

The constant-success guarantee of Algorithm 11 can be amplified to success probability $1-\delta$ by the standard procedure of [HKOT23, Proposition 2.4], multiplying the query

complexity by $O(\log(1/\delta))$. For completeness, each of the two passes makes 2^a calls to the state-tomography routine of Lemma 7.27. A call takes $O(8^{a+b})$ time, while collecting the required copies and assembling and rounding the 2^b blocks incur only polynomial overhead in n, 2^{a+b}, and $1/\epsilon$. Standard confidence amplification contributes a polynomial factor in $\log(1/\delta)$. Thus the total runtime is

$$\mathrm{poly}\big(n,\ 2^{2a+b},\ \epsilon^{-1},\ \log \delta^{-1}\big),$$

as claimed. □

7.5 Reducing Pauli Dimensionality to Block Diagonality

In this section, we describe how to reduce the problem of learning a k-Pauli-dimensional unitary channel to the problem of learning an (a, b, ϵ)-approximately block-diagonal unitary channel, for suitable integers a, b with $2a + b = k$. Our reduction has two phases. In the first phase, we learn a subspace that captures almost all of the Pauli weight of the unknown unitary U. We give two algorithms for this task: one that requires only forward access to U, and another that additionally uses inverse access; the latter achieves a quadratic improvement in query complexity. In the second phase, we construct a Clifford circuit that maps the subspace found in the first phase into the canonical form $\mathcal{W}_{a',b'}$.

Together, these two phases give the desired reduction. Conjugating U by the Clifford from the second phase produces an approximately block-diagonal unitary, which can be learned using Section 7.4. We now describe both phases in detail.

7.5.1 Learning the Support Subspace

The first step of our reduction is to identify a subspace that captures almost all of the Pauli weight of the unknown unitary U. Recall that any n-qubit unitary can be expressed in the Pauli basis as

$$U = \sum_{x \in \mathbb{F}_2^{2n}} \alpha_x W_x,$$

where the coefficients form a probability distribution when squared in magnitude, since $\sum_x |\alpha_x|^2 = 1$.

A convenient way to access this distribution is via the Choi state of U, defined as

$$|\Phi_U\rangle := \big(I^{\otimes n} \otimes U\big)\frac{1}{\sqrt{2^n}}\sum_{x\in\mathbb{F}_2^n}|x\rangle\otimes|x\rangle.$$

The states $\{|\Phi_{W_x}\rangle : x \in \mathbb{F}_2^{2n}\}$ form the *Bell basis*. It is a standard fact (see, e.g., [MO10]) that measuring $|\Phi_U\rangle$ in the Bell basis yields outcome x with probability $|\alpha_x|^2$. Thus, repeated Bell-basis measurements of the Choi state provide us with independent samples from the distribution supported on $\mathrm{supp}(U)$.

Learning Without the Inverse

Our strategy without inverse access is simply to collect enough Bell samples and infer a low-dimensional subspace that captures most of the probability mass. The sampling lemma in Lemma 2.23 guarantees that $O((d + \log(1/\delta))/\eta)$ samples suffice when the distribution is supported on a d-dimensional subspace; see also [GIKL25, Lemma 2.3]. Gaussian elimination returns a succinct set of generators for the sampled span. Applying this result to Bell samples from the Choi state $|\Phi_U\rangle$ gives an efficient procedure for learning a subspace that captures nearly all of the Pauli weight of U.

Learning With the Inverse

If we also have access to $U^\dagger$, then we can use amplitude amplification to obtain a more efficient support-learning algorithm. The key ingredient is fixed-point amplitude amplification, which allows us to boost the probability of detecting computational-basis strings with non-negligible support.

Lemma 7.29 (Fixed-point amplitude amplification [YLC14]). *Let $|\psi\rangle$ be an n-qubit quantum state and Π a diagonal projector in the computational basis such that*

$$\langle\psi|\Pi|\psi\rangle \geq \eta$$

for some known $\eta > 0$. Suppose we have unitaries $U, U^\dagger$ with $U|0^n\rangle = |\psi\rangle$. Then there is an algorithm that outputs a computational-basis state $|x\rangle$ satisfying $\langle x|\Pi|x\rangle = 1$ and $\langle x|\psi\rangle \neq 0$ with probability at least $1-\delta$. The algorithm uses

$$O\left(\frac{\log(1/\delta)}{\sqrt{\eta}}\right)$$

queries to $U, U^\dagger$, in addition to the gates required to implement the two known reflections used by the fixed-point search.

This tool allows us to find basis states in the support of $|\psi\rangle$ more efficiently than by simple sampling. Combining it with the iterative spanning procedure from Lemma 2.23 yields the following corollary, which quadratically improves the dependence of the query complexity on η.

Corollary 7.30. *Let $|\psi\rangle = \sum_{x\in\mathbb{F}_2^n} \beta_x |x\rangle$ be an n-qubit state, let $\eta, \delta \in (0,1)$, and suppose $R|0^n\rangle = |\psi\rangle$ with query access to R and $R^\dagger$. Suppose there exists a subspace $D \subseteq \mathbb{F}_2^n$ of dimension d such that*

$$\sum_{x\in D} |\beta_x|^2 = 1.$$

Then there is an algorithm that outputs a subspace $A \subseteq \mathbb{F}_2^n$ such that

$$\sum_{x\in A} |\beta_x|^2 \geq 1 - \eta$$

with probability at least $1-\delta$. Its resource costs are

$$\begin{aligned}
&O\left(\frac{d}{\sqrt{\eta}} \log(d/\delta)\right) && \text{queries to } R \text{ and } R^\dagger,\\
&O\left(\frac{nd^2}{\sqrt{\eta}} \log(d/\delta)\right) && \text{elementary gates,}\\
&O(d^3 n) && \text{classical processing time.}
\end{aligned}$$

Proof. We build up the target subspace iteratively. Let $A_0 = \{0\}$, and for $i = 1, \ldots, d$ repeat:

1. Define Π_i to be the projector onto basis states outside of A_{i-1}.
2. Run the fixed-point search sequence from Lemma 7.29 with parameter η and failure probability δ/d, and measure its output in the computational basis.
3. If the outcome x lies outside A_{i-1}, set $A_i = \mathrm{span}(A_{i-1} \cup \{x\})$; otherwise set $A_i = A_{i-1}$.

The fixed-point sequence uses only the reflection about $|\psi\rangle$ and a diagonal reflection, so its state remains supported on the computational-basis support of $|\psi\rangle$. Consequently

every vector added to A_i lies in D. Whenever

$$\sum_{x \notin A_{i-1}} |\beta_x|^2 \geq \eta,$$

the promise of Lemma 7.29 holds, and that round finds a new vector outside A_{i-1}, except with probability δ/d. A union bound therefore shows that, with probability at least $1-\delta$, either the residual weight drops below η during the procedure or all d rounds add linearly independent vectors from the d-dimensional space D. In the first case the current span, and hence the final span, captures weight at least $1-\eta$; in the second case $A_d = D$.

Finally, Lemma 7.29 uses $O\left(\frac{1}{\sqrt{\eta}} \log(d/\delta)\right)$ queries and $O\left(\frac{nd}{\sqrt{\eta}} \log(d/\delta)\right)$ gates per iteration: reversible membership in the current span can be implemented from a row-reduced basis using $O(nd)$ gates. Over d iterations, this yields the claimed query and gate complexities. Each update of A_i requires Gaussian elimination in $O(nd^2)$ time, so across all d iterations the total classical cost is $O(d^3 n)$. □

7.5.2 Mapping to a Block-Diagonal Unitary via Clifford Circuits

Having identified a subspace that captures most of the Pauli weight of U, we now show how to map it into the canonical form $\mathcal{W}_{a',b'}$. Along the way, we determine how accurate the learned subspace must be for the reduction to go through.

A basic fact we rely on is that Clifford circuits normalize the Pauli group: for any Clifford C and any $x \in \mathbb{F}_2^{2n}$,

$$CW_xC^\dagger = \pm W_y$$

for some $y \in \mathbb{F}_2^{2n}$. Thus, conjugating U by a suitable Clifford circuit that transforms the span of its Pauli support into $\mathcal{W}_{a,b}$ yields a unitary $CUC^\dagger$ that is block-diagonal. For a subspace $S \subseteq \mathbb{F}_2^{2n}$, we formally define

$$CSC^\dagger := \{y \in \mathbb{F}_2^{2n} \,:\, \exists x \in S \text{ such that } CW_xC^\dagger = \pm W_y\}.$$

Every Pauli subspace $S \subseteq \mathbb{F}_2^{2n}$ admits a canonical basis with respect to the symplectic inner product. Specifically, S can always be generated by vectors of the form

$$\{x_1, z_1, \ldots, x_a, z_a, z_{a+1}, \ldots, z_{a+b}\},$$

where the symplectic products satisfy $[x_i, x_j] = [z_i, z_j] = 0$ and $[x_i, z_j] = \delta_{ij}$ for all i, j. In this representation, S naturally decomposes into a *symplectic part* of dimension $2a$, spanned by the pairs (x_i, z_i), and an *isotropic part* of dimension b, spanned by $\{z_{a+1}, \dots, z_{a+b}\}$. This (a, b)-decomposition directly determines the block structure: if U has Pauli support contained in such an S, then after a suitable Clifford conjugation U becomes (a, b)-block diagonal.

We can now state the main technical lemma of this section. It shows that if a learned subspace captures sufficient Pauli weight, then we can efficiently construct a Clifford $\widetilde{C}$ for which $\widetilde{C}U\widetilde{C}^\dagger$ is approximately block diagonal, with parameters determined by the learned subspace.

Lemma 7.31. *Let U be a k-Pauli-dimensional unitary and let $S := \mathrm{span}(\mathrm{supp}(U))$. Suppose S has an (a, b)-decomposition, so $2a + b = k$. If one can learn a subspace $T \subseteq S$ satisfying*

$$\frac{1}{4^n} \sum_{x \in T} |\mathrm{tr}(UW_x)|^2 \geq 1 - \frac{\epsilon^2}{2^{a+b+2}},$$

then there exists a Clifford circuit $\widetilde{C}$ such that $\widetilde{C}U\widetilde{C}^\dagger$ is (a', b', ϵ)-approximately block diagonal for some $a', b' \geq 0$ with $a' + b' \leq a + b$. Moreover, $\widetilde{C}$ can be constructed in time $O(nk^2)$.

To prove this lemma, we require two results about subspaces of symplectic vector spaces. The first is a semi-folklore decomposition procedure. Specifically, given subspaces $T \subseteq S \subseteq \mathbb{F}_2^{2n}$, one can efficiently find a basis for T split into symplectic and isotropic parts, and extend this to a basis for S such that the basis of T is contained in that of S. This result generalizes the Symplectic Gram-Schmidt procedure (see e.g., [FCYBC04, Lemma 2], [Wil09], and [GIKL24, Lemma 5.1]) to work simultaneously on two subspaces (one a subspace of the other). An explicit algorithm for obtaining this decomposition is somewhat subtle and, to the best of our knowledge, does not appear in the literature. We therefore give its proof in Section 7.11, after the main learning results.

Lemma 7.32. *Let $T \subseteq S \subseteq \mathbb{F}_2^{2n}$, where T has symplectic type (a', b') and S has type (a, b). There is an integer ℓ satisfying*

$$0 \leq \ell \leq \min\{b', a - a'\}, \qquad b - b' + \ell \geq 0,$$

and a simultaneous symplectic basis such that

$$\begin{aligned} T &= \operatorname{span}\{z_1, x_1, \dots, z_{a'}, x_{a'}, z_{a'+1}, \dots, z_{a'+b'}\}, \\ S &= \operatorname{span}\Big(T \cup \{x_{a'+1}, \dots, x_{a'+\ell}\} \\ &\qquad \cup \{z_{a'+b'+1}, x_{a'+b'+1}, \dots, z_{a+b'-\ell}, x_{a+b'-\ell}\} \\ &\qquad \cup \{z_{a+b'-\ell+1}, \dots, z_{a+b}\}\Big), \end{aligned}$$

where $[x_i, x_j] = [z_i, z_j] = 0$ *and* $[x_i, z_j] = \delta_{ij}$ *for all* i, j*. Moreover, such a basis can be found in time* $O(n(a+b)^2)$ *given any generating sets for* T *and* S*.*

The second result is another semi-folklore result that can be attributed to the techniques of [AG04]. This algorithm is a minor generalization of [Ber21, Section 4.2] and [GIKL25, Lemma 3.2].

Lemma 7.33. *Given generators*

$$\{z_1, x_1, \dots, z_{a'}, x_{a'}, z_{a'+1}, \dots, z_{a'+b'}\}$$

for a subspace $T \subseteq \mathbb{F}_2^{2n}$*, and additional generators*

$$\{x_{a'+1}, \dots, x_{a'+\ell}, z_{a'+b'+1}, x_{a'+b'+1}, \dots, z_{a+b'-\ell}, x_{a+b'-\ell}, z_{a+b'-\ell+1}, \dots, z_{a+b}\}$$

that extend this to a generating set for a larger subspace $S \supseteq T$*, where for all* i, j*,* $[x_i, x_j] = [z_i, z_j] = 0$ *and* $[x_i, z_j] = \delta_{ij}$*, there is a Clifford circuit* C *such that*

$$CTC^\dagger = \mathcal{W}_{a',b'}, \qquad CSC^\dagger \subseteq \mathcal{W}_{a+b'-\ell,\, b-b'+\ell}.$$

Moreover, this Clifford circuit can be found in time $O(n(a+b))$*.*

Before proving the main lemma of this section, we record one final technical fact. It shows that if most of the Pauli weight of an (a, b)-block diagonal unitary is concentrated on a smaller subspace $\mathcal{W}_{a',b'}$, then the unitary is close (in operator norm) to its Pauli projection onto $\mathcal{W}_{a',b'}$.

Lemma 7.34. *Let U be an (a,b)-block diagonal unitary, and suppose there exists $a', b' \in \mathbb{N}$ with $a' + b' \leq a + b$ such that*

$$\frac{1}{4^n} \sum_{x \in \mathcal{W}_{a',b'}} |\mathrm{tr}(W_x U)|^2 \geq 1 - \frac{\epsilon^2}{2^{a+b}}$$

Then $\|\Pi_{\mathcal{W}_{a',b'}}(U) - U\|_{\mathrm{op}} \leq \epsilon$.

Proof. Because U is (a,b)-block diagonal, we can write $U = I^{\otimes n-a-b} \otimes U'$ for some $(a+b)$-qubit unitary U' (recall that an (a,b)-block diagonal matrix is also $(a+b,0)$-block diagonal). Observe that

$$\frac{1}{2^n} \mathrm{tr}((I^{\otimes n-a-b} \otimes W_x) \cdot U) = \frac{1}{2^n} \mathrm{tr}(I^{\otimes n-a-b} \otimes (W_x \cdot U')) = \frac{1}{2^{a+b}} \mathrm{tr}(W_x U').$$

Now let

$$\Pi_{\mathcal{W}_{a',b'}}(U) = I^{\otimes n-a-b} \otimes \Pi_{\mathcal{W}_{a',b'}}(U').^{1}$$

By the Pauli Parseval identity Eq. (2.13) and the fact that $\frac{1}{4^{a+b}} \sum_{x \in \mathrm{supp}(U')} |\mathrm{tr}(W_x U')|^2 = 1$,

$$\begin{aligned}
\|\Pi_{\mathcal{W}_{a',b'}}(U') - U'\|_{\mathrm{F}} &= \sqrt{2^{a+b}} \cdot \sqrt{\frac{1}{4^{a+b}} \sum_{x \in \mathrm{supp}(U') \setminus \mathcal{W}_{a',b'}} |\mathrm{tr}(W_x U')|^2} \\
&= \sqrt{2^{a+b}} \cdot \sqrt{1 - \frac{1}{4^{a+b}} \sum_{x \in \mathcal{W}_{a',b'}} |\mathrm{tr}(W_x U')|^2} \\
&= \sqrt{2^{a+b}} \cdot \sqrt{1 - \frac{1}{4^n} \sum_{x \in \mathcal{W}_{a',b'}} |\mathrm{tr}((I^{\otimes n-a-b} \otimes W_x) \cdot U)|^2} \\
&\leq \epsilon.
\end{aligned}$$

Finally, note that

$$\begin{aligned}
\|\Pi_{\mathcal{W}_{a',b'}}(U) - U\|_{\mathrm{op}} &= \|I^{\otimes n-a-b} \otimes \left(\Pi_{\mathcal{W}_{a',b'}}(U') - U'\right)\|_{\mathrm{op}} \\
&= \|\Pi_{\mathcal{W}_{a',b'}}(U') - U'\|_{\mathrm{op}} && \text{(Eq. (2.2))} \\
&\leq \|\Pi_{\mathcal{W}_{a',b'}}(U') - U'\|_{\mathrm{F}} && \text{(Eq. (2.1))} \\
&\leq \epsilon. && \square
\end{aligned}$$

[1] Technically, these $\mathcal{W}_{a',b'}$ are different, as the number of preceding identity matrices changes from $n-a-b$ to $(a+b)-(a'+b')$.

By combining Lemmas 7.32 to 7.34, we can prove Lemma 7.31.

Proof. Let us assume for a moment that the full support span S is known. Lemma 7.32 guarantees that T and S satisfy the hypotheses of Lemma 7.33. Hence there is a Clifford circuit C for which $CTC^\dagger = \mathcal{W}_{a',b'}$ and $CSC^\dagger \subseteq \mathcal{W}_{\bar{a},\bar{b}}$, where $\bar{a} = a + b' - \ell$, $\bar{b} = b - b' + \ell$, and $\bar{a} + \bar{b} = a + b$. Let $U_{\mathrm{BD}} := CUC^\dagger$. Using Lemmas 7.25 and 7.34 and the strengthened constant in the hypothesis, we see that U_{BD} is (a', b', ϵ)-approximately block-diagonal.

The algorithm does not know S; the Clifford C above serves only as a witness. From the generators of T alone, construct a Clifford $\widetilde{C}$ satisfying $\widetilde{C}T\widetilde{C}^\dagger = \mathcal{W}_{a',b'}$; this takes $O(nk^2)$ time by Lemma 7.33. Now define the Clifford $C_2 = C\widetilde{C}^\dagger$, which normalizes

$$C_2\big(\mathcal{W}_{a',b'}\big)C_2^\dagger = C_2\Big(\widetilde{C}T\widetilde{C}^\dagger\Big)C_2^\dagger = CTC^\dagger = \mathcal{W}_{a',b'}.$$

Observe that this also implies

$$C_2^\dagger\big(\mathcal{W}_{a',b'}\big)C_2 = \mathcal{W}_{a',b'}.$$

As U_{BD} is within ϵ of some (a', b')-block-diagonal unitary V in operator distance, unitary invariance gives

$$\|U_{\mathrm{BD}} - V\|_{\mathrm{op}} = \|C_2^\dagger U_{\mathrm{BD}} C_2 - C_2^\dagger V C_2\|_{\mathrm{op}} \le \epsilon.$$

Finally, by Fact 7.23 note that $\mathrm{supp}(V) \subseteq \mathcal{W}_{a',b'}$. Therefore $\mathrm{supp}\Big(C_2^\dagger V C_2\Big) \subseteq \mathcal{W}_{a',b'}$, so $C_2^\dagger V C_2$ is *also* (a', b')-block diagonal, by Fact 7.23 once again. As we established that $C_2^\dagger U_{\mathrm{BD}} C_2$ is close to this (a', b')-block diagonal unitary $C_2^\dagger V C_2$, it follows that

$$C_2^\dagger U_{\mathrm{BD}} C_2 = \widetilde{C}C^\dagger\Big(CUC^\dagger\Big)C\widetilde{C}^\dagger = \widetilde{C}U\widetilde{C}^\dagger$$

is (a', b', ϵ)-approximately block-diagonal, completing the proof. □

7.6 Learning k-Pauli Dimensionality

In this section, we present our algorithm for learning k-Pauli-dimensional unitary channels. We build on the preceding two sections, which reduced the problem to learning approximately block-diagonal unitaries and gave an algorithm for that task. We first describe a

base algorithm whose query complexity scales quadratically with the desired precision. We then apply the bootstrapping technique of [HKOT23] to achieve Heisenberg scaling. The resulting fine-grained bound gives a query-optimal algorithm for quantum juntas in Section 7.9.1.

7.6.1 Base Algorithm for k-Pauli Dimensionality

Theorem 7.35. *Let $a, b, n \in \mathbb{N}$ with $a+b \le n$. Let $U \in \mathbb{C}^{2^n \times 2^n}$ be a k-Pauli-dimensional unitary whose support span has a $2a$-dimensional symplectic part and a b-dimensional isotropic part; equivalently, for some Clifford C,*

$$C \operatorname{span}(\operatorname{supp}(U)) C^\dagger = \mathcal{W}_{a,b}.$$

There is a tomography algorithm that, given query access to $U \in \mathbb{C}^{2^n \times 2^n}$ and parameters $\delta, \epsilon > 0$, outputs an estimate V satisfying $\mathsf{d}_{\mathrm{ph,op}}(U, V) \le \epsilon$ with probability at least $1 - \delta$. Moreover, the algorithm satisfies the following properties:

- $\operatorname{supp}(V) \subseteq \operatorname{span}(\operatorname{supp}(U))$.
- *The algorithm makes at most $O\left(2^{a+b}(2^a + b)\frac{\log(1/\delta)}{\epsilon^2}\right)$ queries to U and only requires forward access; that is, it does not require $U^\dagger$ or controlled-U.*
- *The algorithm runs in time $\operatorname{poly}\left(n,\ 2^{2a+b},\ \epsilon^{-1},\ \log \delta^{-1}\right)$.*
- *The algorithm uses at most $2n$ additional qubits of space.*

Proof. Sample from the Choi state of U, using n ancilla qubits, and let A be the span of the observed Pauli labels. Using Lemma 7.31 we need

$$\frac{1}{4^n} \sum_{x \in A} |\operatorname{tr}(W_x U)|^2 \ge 1 - \frac{\epsilon^2}{K \cdot 2^{a+b}}$$

to invoke Lemma 7.31. For sufficiently large universal constants, that lemma constructs a Clifford $\widetilde{C}$ for which $\widetilde{C} U \widetilde{C}^\dagger$ is $(a', b', \epsilon/K')$-approximately block diagonal. Here $a' \le a$, $a' + b' \le a + b$, and $2a' + b' \le 2a + b$. By Lemma 2.23, the forward-only support-learning query cost is

$$O\left(\frac{2^{a+b}}{\epsilon^2}(2a + b) \log \frac{1}{\delta}\right).$$

Finally, apply Theorem 7.22 to the $(a', b', \epsilon/K')$-approximately block-diagonal unitary. Its query cost is

$$O\left(2^{a'+b'}(2^{a'}+b')\frac{\log(1/\delta)}{\epsilon^2}\right) = O\left(2^{a+b}(2^a+b)\frac{\log(1/\delta)}{\epsilon^2}\right).$$

The equality uses $b' \le 2a+b$ and $2a = O(2^a)$. The forward-only support-learning cost is bounded by the same expression, and each query to $\widetilde{C}U\widetilde{C}^\dagger$ makes one query to U. If V_{BD} is the block learner's output, return

$$V = \widetilde{C}^\dagger V_{\mathrm{BD}} \widetilde{C}.$$

Unitary invariance gives the claimed error. Moreover, $\mathrm{supp}(V_{\mathrm{BD}}) \subseteq \mathcal{W}_{a',b'} = \widetilde{C}A\widetilde{C}^\dagger$, so $\mathrm{supp}(V) \subseteq A \subseteq \mathrm{span}(\mathrm{supp}(U))$.

The Choi-state and block-learning workspaces can be reused and require at most $2n$ additional qubits. Constructing the Clifford takes $O(n(a+b)^2)$ time, and the support-learning and block-learning stages both have runtime

$$\mathrm{poly}(n, 2^{2a+b}, \epsilon^{-1}, \log \delta^{-1}).$$

□

7.6.2 Bootstrapping to Heisenberg Scaling

To achieve the optimal $1/\epsilon$ Heisenberg scaling, we augment Theorem 7.35 using the bootstrapping technique of [HKOT23]. The high-level idea is as follows. Suppose we first learn an approximation V_1 with $\mathsf{d}_{\mathrm{ph,op}}(U, V_1) \le \epsilon$. Then $UV_1^\dagger$ is itself ϵ-close to the identity. Applying the base algorithm to $(UV_1^\dagger)^2$ yields a unitary V_2 such that $\mathsf{d}_{\mathrm{ph,op}}(V_2, (UV_1^\dagger)^2) \le \epsilon$. Taking a square root, $\sqrt{V_2}$ provides an approximation to $UV_1^\dagger$ that is accurate up to $\epsilon/2$, so that $\mathsf{d}_{\mathrm{ph,op}}(U, \sqrt{V_2}V_1) \le \epsilon/2$. Iterating this process with higher powers progressively reduces the error, ultimately yielding Heisenberg scaling. The procedure is captured in the following lemma.

Lemma 7.36 ([HKOT23, Theorem 3.3, Remark 3.4]). *Suppose we are given oracle access to an unknown unitary $U \in \mathbb{C}^{d\times d}$ belonging to a subgroup G of unitaries that is closed under*

fractional powers (i.e., $U^{1/p} \in G$ for all $U \in G$ and $p \in \mathbb{N}$). Assume further that there exists an algorithm $\mathcal{A}$ that, given oracle access to $U \in G$, outputs a unitary $V \in G$ with $\mathsf{d}_{\mathrm{ph,op}}(U, V) \leq \frac{1}{600}$ with probability at least 0.51. Then, for any error parameters $\epsilon, \delta \in (0, 1)$, there is an algorithm that outputs a unitary $V' \in G$ with the following guarantees:

- $\mathsf{d}_{\mathrm{ph,op}}(U, V') \leq \epsilon$ *with probability at least* $1 - \delta$*;*
- $V' \in G$.

Moreover, if $\mathcal{A}$ uses Q queries, then the new algorithm uses only $O\left(\frac{Q}{\epsilon} \log(1/\delta)\right)$ queries.

Let T denote the runtime of $\mathcal{A}$ (to achieve constant error and constant success probability), D the time to compute distances between elements in G, P the time to compute a fractional power of a unitary in G, M the time to multiply elements of G, and S the time to synthesize a circuit for elements of G given their classical descriptions. Then the overall runtime of the new algorithm is

$$O\left(\left(\frac{S \cdot Q}{\epsilon} + (T + P + M + D \log(1/\delta)) \log(1/\epsilon)\right) \log(1/\delta)\right).$$

An immediate corollary of Theorem 7.35 and Lemma 7.36 is a nearly query-optimal algorithm for learning k-Pauli-dimensional unitary channels; it is optimal in the parameter regimes identified below. For the bootstrapping to work, it is crucial that the output V of Theorem 7.35 lies in the same unitary algebra as U, namely the algebra spanned by $\{W_x : x \in \mathrm{span}(\mathrm{supp}(U))\}$.

Corollary 7.37. *Let $a, b, n \in \mathbb{N}$ with $a + b \leq n$. Let $U \in \mathbb{C}^{2^n \times 2^n}$ be a k-Pauli-dimensional unitary whose support span has a $2a$-dimensional symplectic part and a b-dimensional isotropic part, so that $C\,\mathrm{span}(\mathrm{supp}(U))C^\dagger = \mathcal{W}_{a,b}$ for some Clifford circuit C. There is a tomography algorithm that, given query access to $U \in \mathbb{C}^{2^n \times 2^n}$ and parameters $\delta, \epsilon > 0$, outputs an estimate V satisfying $\mathsf{d}_\diamond(U, V) \leq \epsilon$ with probability at least $1 - \delta$. Moreover, the algorithm satisfies the following properties:*

- $\mathrm{supp}(V) \subseteq \mathrm{span}(\mathrm{supp}(U))$.

- *The algorithm makes at most* $O\left(2^{a+b}(2^a+b)\frac{\log(1/\delta)}{\epsilon}\right)$ *queries to* U *and requires only forward access; it does not require* $U^\dagger$ *or controlled-*U.
- *The algorithm runs in time* $\mathrm{poly}\left(n, 2^{2a+b}, \epsilon^{-1}, \log \delta^{-1}\right)$.
- *The algorithm uses at most* $2n$ *additional qubits of space.*

Proof. Let $A = \mathrm{span}(\mathrm{supp}(U))$ and let $\mathfrak{A}$ be the linear span of $\{W_x : x \in A\}$. Because A is a binary subspace, $\mathfrak{A}$ is a finite-dimensional unital $*$-algebra. Its unitary elements form a group G, and functional calculus shows that every fractional power of an element of G remains in G. Since Theorem 7.35 always outputs an element of G, Lemma 7.36 gives Heisenberg scaling. Finally, use Fact 7.11 to convert the distance to diamond distance.

After the Clifford reduction, $\mathfrak{A}$ is represented as a direct sum of 2^b blocks of size 2^a. Distances, products, functional calculus, and circuit synthesis can therefore all be performed in time polynomial in n, 2^{2a+b}, the requested precision, and the confidence parameter. Together with the base learner, this gives the stated runtime. □

Finally, let us state the performance of the algorithm solely in terms of k, since a and b are generally unknown. This is immediate from Corollary 7.37 and the identity $2a + b = k$.

Corollary 7.38. *Let* $U \in \mathbb{C}^{2^n \times 2^n}$ *be a* k*-Pauli-dimensional unitary. There is a tomography algorithm that, given query access to* U *and parameters* $\delta, \epsilon > 0$*, outputs an estimate* V *satisfying* $\mathsf{d}_\diamond(U, V) \leq \epsilon$ *with probability at least* $1 - \delta$*. Moreover, the algorithm satisfies the following properties:*

- $\mathrm{supp}(V) \subseteq \mathrm{span}(\mathrm{supp}(U))$.
- *The algorithm makes at most* $O\left(2^k k \frac{\log(1/\delta)}{\epsilon}\right)$ *queries to* U *and requires only forward access; it does not require* $U^\dagger$ *or controlled-*U.
- *The algorithm runs in* $\mathrm{poly}(n, 2^k, \epsilon^{-1}, \log \delta^{-1})$ *time.*
- *The algorithm uses at most* $2n$ *additional qubits of space.*

7.7 Learning s-Pauli Sparsity

In this section, we present our algorithm for learning unitaries with sparse Pauli support. At a high level, we reduce the problem to quantum state learning. In particular, we show that obtaining a poly(s)-time improper learning algorithm reduces to the following state-tomography task: given copies of a state $|\psi\rangle$ supported on s computational-basis states, output an approximate classical description of $|\psi\rangle$ in poly(s) time. We then give an optimal algorithm for this task.

The resulting unitary-learning algorithm outputs an estimate A that is not necessarily unitary. To obtain a proper learner, one must round A to a nearby unitary while preserving its structure. It is unclear how to do this efficiently in general, and we leave it as an open problem. We do, however, identify two natural settings in which efficient rounding is possible, and we explain how the classical rounding step can be avoided if one only needs to apply a quantum channel close to the unknown unitary.

Definition 7.39 (Pauli sparsity). A matrix A is s-Pauli-sparse if $|\mathrm{supp}(A)| \leq s$.

The following lemma establishes that we can efficiently learn the support of the unknown sparse unitary.

Lemma 7.40. *Let $\epsilon \in (0,1]$ and $\delta \in (0,1)$. Given an unknown s-Pauli-sparse n-qubit unitary $U = \sum_{x\in\mathrm{supp}(U)} \alpha_x W_x$, using*

$$\left\lceil 2\frac{s+\log(1/\delta)}{\epsilon^2} \right\rceil$$

queries, $O\Big(\frac{n}{\epsilon^2}(s+\log(1/\delta))\Big)$ time, and n ancilla qubits, one can, with probability at least $1-\delta$, learn a set $S \subseteq \mathrm{supp}(U)$ such that

$$\sum_{x\notin S} |\alpha_x|^2 \leq \epsilon^2.$$

Proof. Bell-sample U, and let S_i be the set of distinct outcomes observed after i samples. Until the residual mass $r_i = \sum_{x\notin S_i} |\alpha_x|^2$ falls below ϵ^2, the conditional probability that the next sample reveals a new support element is $r_i > \epsilon^2$. After the target condition has been

met, declare every subsequent trial a success. The resulting adaptive success indicators satisfy the hypotheses of Lemma 2.8 with $p = \epsilon^2$.

Once s successes have occurred, either the residual-mass condition was reached or all at most s support elements were found. Thus, for $m = \lceil 2(s + \log(1/\delta))/\epsilon^2 \rceil$, the set S_m has the desired property with probability at least $1 - \delta$. □

Next, we give an optimal state tomography algorithm for states supported on a small number of computational basis states.

Lemma 7.41 (Copy-optimal tomography of sparse quantum states). *Let $1 \le s \le 2^n$, and let $|\psi\rangle$ be an n-qubit quantum state supported on at most s computational-basis states. For $\epsilon \in (0, 1]$ and $\delta \in (0, 1)$, there is an algorithm that, given copies of $|\psi\rangle$, outputs a classical description of a state $|\hat{\psi}\rangle$ such that, with probability at least $1 - \delta$,*

- *$|\hat{\psi}\rangle$ is ϵ-close to $|\psi\rangle$ in trace distance;*
- *$|\hat{\psi}\rangle$ is supported on a subset of the support of $|\psi\rangle$;*
- *the algorithm uses at most $O\left(\frac{s+\log(1/\delta)}{\epsilon^2}\right)$ copies of $|\psi\rangle$;*
- *the algorithm runs in time $O\left(ns\,\frac{s+\log(1/\delta)}{\epsilon^2} + s^3\right)$.*

Moreover, for $2 \le s \le 2^n$ and $0 < \epsilon \le 1/2$, $\Omega(s/\epsilon^2)$ copies are necessary at constant success probability.

Proof. Let $B \subseteq \mathbb{F}_2^n$ be the computational-basis support and write $|\psi\rangle = \sum_{x \in B} \beta_x |x\rangle$. Computational-basis sampling and the argument of Lemma 7.40, with accuracy $\epsilon/2$ and failure probability $\delta/3$, produce a set $S \subseteq B$ such that

$$p_S := \sum_{x \in S} |\beta_x|^2 \ge 1 - \epsilon^2/4.$$

This step uses at most

$$m_0 := \left\lceil \frac{8(s + \log(3/\delta))}{\epsilon^2} \right\rceil$$

copies.

To prepare the coherent truncation, reversibly compute the predicate $x \in S$ into an ancilla and postselect that ancilla on 1. The data register is then in the state

$$|\phi\rangle := \frac{1}{\sqrt{p_S}} \sum_{x \in S} \beta_x |x\rangle.$$

A direct membership test takes $O(sn)$ gates per attempt. The resulting state lies in an s-dimensional space. Run the tomography algorithm of Lemma 7.27 with accuracy $\epsilon/2$ and failure probability $\delta/3$. For some absolute constant C, it requires at most

$$d := \left\lceil C \frac{s + \log(3/\delta)}{\epsilon^2} \right\rceil$$

copies of $|\phi\rangle$. Conditional on the support-learning step succeeding, each postselection attempt succeeds with probability at least $1 - \epsilon^2/4$. Hence Lemma 2.8 shows that

$$M := \left\lceil \frac{2}{1 - \epsilon^2/4} \left(d + \log \frac{3}{\delta} \right) \right\rceil = O\left(\frac{s + \log(1/\delta)}{\epsilon^2} \right)$$

attempts, and therefore as many additional copies of $|\psi\rangle$, provide the required d successful postselections except with probability at most $\delta/3$.

The trace distance between the two pure states is exactly

$$\sqrt{1 - |\langle \phi | \psi \rangle|^2} = \sqrt{1 - p_S} \leq \epsilon/2.$$

By the triangle inequality, an $\epsilon/2$ estimate of $|\phi\rangle$ is therefore an ϵ-accurate estimate of $|\psi\rangle$ in trace distance. A union bound over support learning, postselection, and tomography gives total failure probability at most δ. The copy and time bounds follow from m_0, M, and the $O(sn)$-time membership test used in each postselection attempt.

For the lower bound, restrict to states in a fixed s-dimensional coordinate subspace. Learning an arbitrary pure state in this subspace to trace distance ϵ, with constant success probability, requires $\Omega(s/\epsilon^2)$ copies [SSW25]. □

We can now present our algorithm for learning unitaries with sparse Pauli support. The algorithm proceeds by reducing to the task solved in Lemma 7.41.

Theorem 7.42 (Improper tomography of Pauli-sparse unitaries). *Let $s \in \mathbb{N}$ with $1 \le s \le 4^n$. Let $U \in \mathbb{C}^{2^n \times 2^n}$ be a s-Pauli-sparse unitary. There is a tomography algorithm that, given query access to $U \in \mathbb{C}^{2^n \times 2^n}$ and parameters $\delta, \epsilon > 0$, outputs a classical description of a matrix V satisfying $\mathsf{d}_{\mathrm{ph,op}}(U, V) \le \epsilon$ with probability at least $1 - \delta$. Moreover, the algorithm satisfies the following properties:*

- $\mathrm{supp}(V) \subseteq \mathrm{supp}(U)$.
- *The algorithm makes at most $O\Big(\frac{s}{\epsilon^2}(s + \log(1/\delta))\Big)$ queries to U and requires only forward access; it does not require $U^\dagger$ or controlled-U.*
- *The output V is not required to be unitary.*
- *The algorithm runs in time $O\Big(ns^2 \frac{s+\log(1/\delta)}{\epsilon^2} + s^3\Big)$.*
- *The algorithm uses n additional qubits of space.*

Proof. Observe that, in the Bell basis, the Choi state of U is

$$|\Phi_U\rangle = \sum_{x \in \mathrm{supp}(U)} \alpha_x |\Phi_{W_x}\rangle,$$

so it is an s-sparse $2n$-qubit state. Applying the standard Bell-basis change reduces its tomography to sparse tomography in the computational basis. Run Lemma 7.41 with trace-distance accuracy $\epsilon/\sqrt{2s}$. If the output is $\sum_x \hat{\alpha}_x |\Phi_{W_x}\rangle$, then, after choosing a global phase, the Euclidean distance between the two amplitude vectors is at most $\sqrt{2}$ times their trace distance. Hence

$$\left(\sum_x |\alpha_x - \hat{\alpha}_x|^2\right)^{1/2} \le \frac{\epsilon}{\sqrt{s}}.$$

Define $V := \sum_x \hat{\alpha}_x W_x$. By Cauchy–Schwarz,

$$\mathsf{d}_{\mathrm{ph,op}}(U, V) \le \sum_x |\alpha_x - \hat{\alpha}_x| \le \sqrt{s}\left(\sum_x |\alpha_x - \hat{\alpha}_x|^2\right)^{1/2} \le \epsilon.$$

Using Lemma 7.41, we obtain this estimate with $O\Big(\frac{s}{\epsilon^2}(s + \log(1/\delta))\Big)$ Choi states, each prepared with one query to U. □

With inverse access, one may instead use [ACGN23, Theorem 26], obtaining $\widetilde{O}(s^{3/2}/\epsilon)$ queries.

We next discuss when the improper estimate in Theorem 7.42 can be efficiently rounded to a unitary.

Fact 7.43. *Any matrix A that can be expressed as a sum over $s < n$ Weyl operators can be rounded to its closest unitary operator (in any unitarily invariant distance) in time* $\mathrm{poly}(n, s) + \exp(O(s))$*, while increasing the Pauli sparsity to at most 2^s.*

Proof. The Pauli sparsity being s also implies that the Pauli dimension is at most s. Putting the support span into canonical form costs polynomial time in n and s, after which the polar decomposition is computed in a space of dimension at most 2^s and takes $\exp(O(s))$ time. □

Fact 7.44. *Any nonzero Hermitian matrix A that is a linear combination of s mutually anticommuting Weyl operators can be rounded by normalizing its coefficient vector to the closest Hermitian unitary in their span in Frobenius distance. This takes $O(ns)$ classical time, and the rounded unitary has the same Pauli support as A.*

Proof. Let S be the anticommuting set and write $A = \sum_{x \in S} \alpha_x W_x$, where the α_x are real. Then

$$\begin{aligned} A^2 &= \sum_{x \in S} \alpha_x^2 I + \sum_{x<y} \alpha_x \alpha_y (W_x W_y + W_y W_x) \\ &= \left(\sum_{x \in S} \alpha_x^2 \right) I. \end{aligned}$$

It follows that $\dfrac{A}{\sqrt{\sum_{x \in S} \alpha_x^2}}$ must be a Hermitian unitary operator.

Every Hermitian unitary in the span has the form $B = \sum_{x \in S} \beta_x W_x$ with $\sum_x \beta_x^2 = 1$. Orthogonality of the Weyl basis gives

$$\|A - B\|_{\mathrm{F}}^2 = 2^n \sum_{x \in S} (\alpha_x - \beta_x)^2.$$

Cauchy–Schwarz shows that the minimizing vector is $\beta = \alpha/\|\alpha\|_2$, so the polar factor $A/\|\alpha\|_2$ is the desired unitary.

When the anticommuting family is known to be Clifford-equivalent to a subset of the Jordan–Wigner Majoranas, the same symplectic reduction constructs that Clifford, and the normalized combination can be synthesized as a matchgate. This is precisely the setting used in Theorem 7.67. □

Remark 7.45. One can avoid classically rounding to the nearest unitary if it suffices to implement a CPTP *quantum channel* that approximates the unknown unitary. Let $A = \sum_{x\in\mathrm{supp}(A)} \alpha_x W_x$ be the output of our algorithm, which is not necessarily unitary. Using Lemma 7.17, one can efficiently build a block-encoding U_A of $\frac{A}{\sum_{x\in\mathrm{supp}(A)}|\alpha_x|}$ and then apply [QR22, Corollary 1, Eq. (8)] to approximately apply the polar decomposition of A. (Note that scaling A by a positive constant does not affect the polar decomposition.) Let $L = \sum_x |\alpha_x|$, and let κ denote the inverse of the smallest singular value of A/L. If A is ϵ-close in operator norm to a unitary, then

$$\kappa \le \frac{L}{1-\epsilon} \le \frac{\sqrt{s}\,\|\alpha\|_2}{1-\epsilon}.$$

Moreover, $\|\alpha\|_2 = \|A\|_{\mathrm{F}}/\sqrt{2^n} \le 1+\epsilon$. Thus, for $\epsilon \le 1/2$, the polar implementation of [QR22, Corollary 1, Eq. (8)] uses $O(\sqrt{s}\log(1/\epsilon))$ calls to U_A and $U_A^\dagger$.

The same block-encoded polar construction applies in the framework of Section 7.8 whenever the desired output is an implementation of a channel close to $U^\dagger \otimes U$.

7.8 A Framework for Learning Structured Quantum Circuits

7.8.1 The Framework

We present our framework for learning structured unitaries. The basic idea is to learn the Heisenberg evolution $U^\dagger g U$ of a generating set of Pauli operators. We first show that these operators determine the unknown unitary information-theoretically, and then show how to combine efficient implementations of their estimates into an implementation of the unknown unitary channel. To state the condition precisely, we begin by defining a Pauli generating set.

Definition 7.46 (Pauli generating set). A subset $G \subset \mathcal{P}_n$ of the n-qubit Pauli group is called a *generating set* if every Pauli operator $P \in \{I, X, Y, Z\}^{\otimes n}$ can be expressed as a product of elements from G, up to a global phase. Formally, for each such P, there exist a phase $c_P \in \{\pm 1, \pm \mathrm{i}\}$ and a sequence of $\ell_P \leq L$ generators $g_1, \ldots, g_{\ell_P} \in G$ such that $P = c_P \prod_{j=1}^{\ell_P} g_j$. The minimum such upper bound L is called the *length* of G.

Theorem 7.47 (Learning generating sets suffices). *Let G be a Pauli generating set with length L. Let U and V be n-qubit unitary channels. If V matches the Heisenberg evolution of U on every generator $P \in G$ to operator norm accuracy ϵ':*

$$\|V^\dagger P V - U^\dagger P U\|_{\mathrm{op}} \leq \epsilon',$$

then the diamond distance between the unitary channels is at most

$$\mathsf{d}_\diamond(U, V) \leq 4L\epsilon'.$$

Proof. Let $W = VU^\dagger$. By assumption, for all $P \in G$,

$$\|W^\dagger P W - P\|_{\mathrm{op}} = \|UV^\dagger P V U^\dagger - UU^\dagger P U U^\dagger\|_{\mathrm{op}} = \|V^\dagger P V - U^\dagger P U\|_{\mathrm{op}} \leq \epsilon'.$$

By the unitary invariance of the operator norm, multiplying by W on the left bounds the commutator: $\|WP - PW\|_{\mathrm{op}} \leq \epsilon'$.

By Definition 7.46, any n-qubit Pauli operator $P \in \{I, X, Y, Z\}^{\otimes n}$ can be written as a product of at most $\ell \leq L$ generators, $P = c_P \prod_{j=1}^{\ell} g_j$. We use a telescoping sum to bound the commutator of W with any Pauli P. Since $\|g_j\|_{\mathrm{op}} = 1$, we have

$$\|WP - PW\|_{\mathrm{op}} \leq \sum_{j=1}^{\ell} \|W g_j - g_j W\|_{\mathrm{op}} \leq \ell\epsilon' \leq L\epsilon'.$$

To bound the global distance between the channels, we apply a standard Pauli twirling identity:

$$\frac{1}{4^n} \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} PWP = \frac{\mathrm{tr}(W)}{2^n} I =: \alpha I.$$

Specifically, we rewrite the difference between W and its identity component in terms of the commutators.

$$\alpha I - W = \frac{1}{4^n} \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} (PWP - W) = \frac{1}{4^n} \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} (PW - WP)P.$$

Taking the operator norm and applying the triangle inequality, we have

$$\|W - \alpha I\|_{\mathrm{op}} \leq \frac{1}{4^n} \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} \|(PW - WP)P\|_{\mathrm{op}} = \frac{1}{4^n} \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} \|WP - PW\|_{\mathrm{op}} \leq L\epsilon'.$$

Because W is unitary, $\|W\|_{\mathrm{op}} = 1$. By the reverse triangle inequality, $1 = \|W\|_{\mathrm{op}} \leq \|\alpha I\|_{\mathrm{op}} + \|W - \alpha I\|_{\mathrm{op}} = |\alpha| + \|W - \alpha I\|_{\mathrm{op}}$, forcing $|\alpha| \geq 1 - L\epsilon'$. Writing $\alpha = |\alpha| e^{\mathrm{i}\theta}$, we factor out the global phase:

$$\|e^{-\mathrm{i}\theta} W - |\alpha| I\|_{\mathrm{op}} = \|W - \alpha I\|_{\mathrm{op}} \leq L\epsilon'.$$

Thus, we can bound the distance from W to the identity matrix:

$$\|e^{-\mathrm{i}\theta} W - I\|_{\mathrm{op}} \leq \|e^{-\mathrm{i}\theta} W - |\alpha| I\|_{\mathrm{op}} + \big||\alpha| - 1\big| \leq L\epsilon' + L\epsilon' = 2L\epsilon'.$$

The diamond distance between U and V is equal to the diamond distance between W and the identity. Thus, we conclude the proof by applying Fact 7.11. □

It is easy to see that $L \leq 2n$ for any generating set G. This is because Pauli operators (up to phase) can be identified with elements of $\mathbb{F}_2^{2n}$, so any such operator can always be expressed with $2n$ generating elements.

Theorem 7.47 establishes that learning the Heisenberg-evolved operators of a generating set information-theoretically suffices to determine an unknown unitary operator. However, to obtain an algorithm, we must also specify how to construct a description of the unknown unitary operator from the learned Heisenberg-evolved operators. We present such a construction now, which generalizes the approach of Huang, Liu, Broughton, Kim, Anshu, Landau, and McClean [Hua+24]. The runtime of the resulting algorithm will depend on the number of generators needed to express the weight-1 Pauli operators, which we call the *local length* of a generating set.

Definition 7.48 (Local length of a generating set). Let $G \subseteq \mathcal{P}_n$ be a generating set. Define the *local length* of G as

$$\ell_G := \max_{i\in[n],\, Q\in\{X_i,Y_i,Z_i\}} \min\Big\{t \,:\, \exists\, P_1,\dots,P_t \in G,\ \omega\in\{\pm 1,\pm \mathrm{i}\} \text{ such that } Q = \omega\prod_{j=1}^{t} P_j\Big\}.$$

Theorem 7.49 (Efficient unitary learning via local-length generating sets). *Let G be a known generating set of local length $L = \ell_G$, and put $m = |G|$. Suppose that, for every $g \in G$, there is an algorithm which, given query access to $U_g := U^\dagger g U$, outputs a circuit for a unitary $\widehat{V}_g$ satisfying*

$$\|\widehat{V}_g - U^\dagger g U\|_{\mathrm{op}} \le \eta$$

with probability at least $1-\delta'$, using $q(\eta,\delta')$ queries to U_g and $t(\eta,\delta')$ time. Suppose also that products of these circuits and the polar factors of the constant-term linear combinations used below can be implemented efficiently from their classical descriptions. Then, given queries to U and $U^\dagger$, there is an algorithm that outputs an implementation V on $2n$ qubits satisfying

$$\mathsf{d}_\diamond(U^\dagger \otimes U, V) \le \epsilon$$

with probability at least $1-\delta$. It uses

$$O\Big(m\, q\Big(\frac{\epsilon}{6nL}, \frac{\delta}{m}\Big)\Big)$$

queries to each of U and $U^\dagger$. The calls to the learning algorithms take total time

$$m\, t\Big(\frac{\epsilon}{6nL}, \frac{\delta}{m}\Big),$$

in addition to the stated circuit-composition and polar-factor costs.

Proof. Learn every generator image to ordinary operator-norm accuracy $\eta = \epsilon/(6nL)$ and failure probability δ/m. A union bound gives simultaneous success. Each query to $U_g = U^\dagger g U$ is implemented using one query each to U and $U^\dagger$, together with the known Pauli g. If $P = c_P \prod_{j=1}^{\ell_P} g_j$ is a weight-one Pauli and $\ell_P \le L$, define

$$\widehat{V}_P = c_P \prod_{j=1}^{\ell_P} \widehat{V}_{g_j}.$$

A telescoping sum gives

$$\|\widehat{V}_P - U^\dagger P U\|_{\mathrm{op}} \le L\eta = \frac{\epsilon}{6n}.$$

Let $\mathrm{SWAP} = \prod_{i=1}^{n} \mathrm{SWAP}_i$, where SWAP_i exchanges qubits i and $i+n$. The identity used in [Hua+24] is

$$U^\dagger \otimes U = \left(\prod_{i=1}^{n} O_i\right)\mathrm{SWAP}, \qquad O_i = (U^\dagger \otimes I)\,\mathrm{SWAP}_i(U \otimes I). \tag{7.15}$$

Since

$$O_i = \frac{1}{2}\Big(I \otimes I + (U^\dagger X_i U) \otimes X_i + (U^\dagger Y_i U) \otimes Y_i + (U^\dagger Z_i U) \otimes Z_i\Big),$$

substituting the three estimates gives a matrix $\widehat{O}_i$ with

$$\|\widehat{O}_i - O_i\|_{\mathrm{op}} \le \frac{\epsilon}{4n}.$$

The matrix $\widehat{O}_i$ is a short linear combination of known implementable unitaries. Its LCU block encoding and the polar construction of Remark 7.45 therefore produce an implementable unitary $\widehat{W}_i$. Since O_i is unitary, the rounding bound in Lemma 7.25 gives

$$\|\widehat{W}_i - O_i\|_{\mathrm{op}} \le \frac{\epsilon}{2n}.$$

Finally set $V = (\prod_i \widehat{W}_i)\,\mathrm{SWAP}$. A telescoping sum yields $\|V - U^\dagger \otimes U\|_{\mathrm{op}} \le \epsilon/2$, and Fact 7.11 converts this to diamond distance at most ϵ. □

The output in Theorem 7.49 is the doubled unitary $U^\dagger \otimes U$, rather than a phase-fixed circuit for U itself. This still implements the channel of U: prepare any fixed state on the first register, place the desired input on the second, apply the doubled unitary, and discard the first output register. The distinction matters only if the learned unitary is to be used as an input to another coherent learning procedure.

Corollary 7.50. *Let G be a known generating set. Let U be an n-qubit unitary such that every $U^\dagger g U$, $g \in G$, has Pauli dimension at most k (respectively, s-Pauli-sparse and efficiently roundable). Suppose that the phase ambiguity in each structural estimate can be removed efficiently, yielding an implementable unitary estimate in ordinary operator norm, and that*

the polar factors required in Theorem 7.49 can be implemented efficiently. Then, given queries to U and $U^\dagger$, there is an algorithm that outputs a $2n$-qubit unitary channel V satisfying $\mathsf{d}_\diamond(U^\dagger \otimes U, V) \le \epsilon$ with probability at least $1-\delta$. Its query complexity is polynomial in $n, 2^k, m, L, \log(1/\delta)$ and $1/\epsilon$ in the dimensional case, and polynomial in $n, s, m, L, \log(1/\delta)$ and $1/\epsilon^2$ in the sparse case, where $L = \ell_G$ and $m = |G|$.

Proof. Combine Corollary 7.38 (respectively, Theorem 7.42) with Theorem 7.49. The additional hypothesis is needed because a phase-aligned matrix estimate does not by itself fix the sign of the reflection $U^\dagger g U$. □

Lemma 7.51 (Fixing the sign of a learned reflection). *Let $R = U^\dagger g U$, where g is a nonidentity Pauli. Suppose an efficiently implementable Hermitian unitary $\widehat{R}$ satisfies*

$$\min_{\sigma\in\{\pm1\}} \|\widehat{R} - \sigma R\|_{\mathrm{op}} \le \eta \qquad (\eta < 1/3),$$

and a $+1$ eigenstate of $\widehat{R}$ can be prepared efficiently. Then the sign σ can be identified with failure probability δ using $O(\log(1/\delta))$ forward queries to U.

Proof. Prepare a $+1$ eigenstate $|\psi\rangle$ of $\widehat{R}$, apply U, and measure g. The expectation of the outcome is $\langle\psi|R|\psi\rangle$, which lies within η of σ. Independent repetition and Hoeffding's inequality (Lemma 2.4) distinguish the two signs. □

We briefly explain how the extra hypothesis in the corollary is met in the applications below. If a structured estimate A is phase-aligned with the reflection R, then $2^{-n}\operatorname{tr}(A^2)$ approximates the square of the unknown phase. Choosing either square root, removing that phase, Hermitianizing, and taking the Hermitian polar factor produces a reflection $\widehat{R}$ close to either R or $-R$. For a low-Pauli-dimensional estimate, these operations reduce, after Clifford normalization, to diagonalizing 2^b blocks of size 2^a; this block description also supplies a $+1$ eigenstate. For an estimate in an anticommuting Pauli span, Hermitianization preserves the span and polar rounding simply normalizes the real coefficient vector. The symplectic reduction maps the anticommuting family to standard Majoranas, after which a fermionic rotation gives an explicit $+1$ eigenstate. All of these steps are polynomial in the corresponding structured description. Lemma 7.51 then fixes the remaining sign.

Remark 7.52 (Explicit query complexity). Applied to the two cases in Corollary 7.50, the dimension and sparsity instantiations use, respectively,

$$Q_{\text{dim}} = O\left(mnL\, 2^k k \frac{\log(m/\delta)}{\epsilon}\right),$$
$$Q_{\text{sparse}} = O\left(mn^2 L^2 s \frac{s + \log(m/\delta)}{\epsilon^2}\right).$$

7.9 Applications

We now apply the framework to quantum juntas, compositions of shallow and near-Clifford circuits, and compositions of matchgate and Clifford circuits. In each case, the main task is to find a suitable Pauli generating set for which the operators $U^\dagger g U$ have low-complexity Pauli spectra.

7.9.1 Quantum k-Juntas

We begin by recalling the definition of a quantum k-junta.

Definition 7.53 (Quantum k-junta). A quantum k-junta unitary channel acts nontrivially on at most k qubits.

In other words, up to permutation of qubits, it acts as $I^{\otimes n-k} \otimes U$, where U is some arbitrary k-qubit unitary matrix. The following query-optimal tomography algorithm follows from Corollary 7.37.

Corollary 7.54 (Junta learner without inverse). *Let $U \in \mathbb{C}^{2^n \times 2^n}$ be a k-junta unitary. There is a tomography algorithm that, given query access to $U \in \mathbb{C}^{2^n \times 2^n}$ as well as parameters $\delta, \epsilon > 0$, outputs an estimate V satisfying $\mathsf{d}_\diamond(U, V) \leq \epsilon$ with probability at least $1 - \delta$. Moreover, the algorithm satisfies the following properties:*

- *V is a k' junta for $k' \leq k$.*
- *The algorithm makes at most $O\left(4^k \frac{\log(1/\delta)}{\epsilon}\right)$ queries to U (and only requires forward access, i.e., it does not require $U^\dagger$ or controlled-U).*
- *The algorithm runs in $\text{poly}(n, 4^k, \epsilon^{-1}, \log \delta^{-1})$ time.*

- *The algorithm uses at most $2n$ additional qubits of space.*

Proof. After permuting qubits, a k-junta takes the form $I^{\otimes n-k} \otimes U$, so its support span is contained in $\mathcal{W}_{k,0}$. Let (a, b) be the type of its actual support span. Since this span lies in the $2k$-dimensional symplectic space on the active qubits, $a + b \leq k$. Apply Corollary 7.37 without the inverse. Its query complexity is at most

$$O\left(2^k(2^k + k)\frac{\log(1/\delta)}{\epsilon}\right) = O\left(4^k\frac{\log(1/\delta)}{\epsilon}\right).$$

The construction therefore uses only forward queries. □

7.9.2 Composition of Shallow and Near-Clifford Circuits

We now give an algorithm for learning unitary channels that can be expressed as the composition of a shallow circuit with a near-Clifford circuit (in either order). By a shallow circuit we mean a depth-d quantum circuit with arbitrary one- and two-qubit gates, and by a near-Clifford circuit we mean a Clifford circuit augmented with at most t arbitrary single-qubit gates. Our algorithm learns such unitaries to accuracy ϵ in diamond distance in polynomial time whenever $2^d + t = O(\log n)$. In particular, this includes $d \leq \log_2 \log_2 n + O(1)$ and $t = O(\log n)$.

Clifford Nullity

A priori, it is not obvious which natural classes of unitary channels (besides shallow depth circuits) satisfy the conditions of Corollary 7.50. In this subsection, we show that Clifford circuits doped with a small number of single-qubit non-Clifford gates do satisfy these conditions, which in turn yields an alternative learning algorithm for this class. In fact, in Section 7.9.2 we will learn shallow circuits composed with the more general class of near-Clifford unitaries involving small Clifford nullity [JW23], which we define below.

Definition 7.55 (Clifford nullity). The *Clifford nullity* of an n-qubit unitary U is

$$\min_S \operatorname{codim}(S),$$

where the minimum is over subspaces $S \subseteq \mathbb{F}_2^{2n}$ such that, for every $x \in S$, there is a $y \in \mathbb{F}_2^{2n}$ with $U^\dagger W_x U = \pm W_y$.

Intuitively, Clifford nullity measures how much of the Pauli group is normalized by U: nullity 0 corresponds exactly to Clifford circuits. If U has Clifford nullity t and S achieves the minimum in Definition 7.55, then the conjugates $\{U^\dagger W_x U : x \in S\}$ correspond to a Pauli subspace S' of the same dimension, and hence of codimension t. Conjugation preserves commutation, so S and S' also have the same symplectic and isotropic dimensions.

To understand Clifford nullity, we establish the following structural result.

Fact 7.56. *Let $U \in \mathbb{C}^{2^n \times 2^n}$ be an n-qubit unitary channel with Clifford nullity t. Then U can be decomposed into $C_2 U' C_1$ where C_2 and C_1 are Clifford unitary channels and U' is t-Pauli-dimensional. Furthermore, if the subspace S achieving the minimum in Definition 7.55 has an $(n-a-b, b)$ decomposition into a symplectic part of dimension $2(n-a-b)$ and an isotropic part of dimension b (so that $2a+b=t$), then U' has Pauli support in $\mathcal{W}_{a,b}$.*

Proof. Let S' be the Pauli subspace obtained by conjugating S by U, and put $T = \mathcal{W}_{a,b}^{\perp_s}$. Conjugation by U restricts to a symplectic isomorphism from S to S'. Choose a Clifford D_1 that maps T to S. The induced map from T to S' extends to an ambient symplectic transformation. Choose a Clifford implementing this transformation. Its sign discrepancies on a basis of T form a linear character; by nondegeneracy of the ambient symplectic form, composing with a suitable Pauli corrects that character. We may therefore choose D_2 so that

$$D_2 W_x D_2^\dagger = U^\dagger (D_1 W_x D_1^\dagger) U \qquad \text{for every } x \in T.$$

Set $U' = D_1^\dagger U D_2$. Then

$$(U')^\dagger W_x U' = D_2^\dagger U^\dagger (D_1 W_x D_1^\dagger) U D_2 = W_x$$

for every $x \in T$. Hence U' commutes with every Weyl operator in T, which forces

$$\operatorname{supp}(U') \subseteq T_s^\perp = \mathcal{W}_{a,b}.$$

If $\operatorname{span}(\operatorname{supp}(U'))$ had dimension $r < t$, then U' would commute with its $2n-r$-dimensional symplectic complement. Conjugating this subspace by D_1 and D_2 would certify that U

has Clifford nullity at most r, contradicting the minimality of t. Thus U' has Pauli dimension exactly t. Finally, $U = D_1 U' D_2^\dagger$, giving the required Clifford–residual–Clifford decomposition. □

Conversely, a decomposition $U = C_2 U' C_1$ with Clifford C_1, C_2 and $\mathrm{supp}(U') \subseteq \mathcal{W}_{a,b}$, $2a+b = t$, certifies Clifford nullity at most t. Thus the structural form characterizes bounded Clifford nullity. In particular, a k-Pauli-dimensional unitary has Clifford nullity at most k: every Weyl operator in $\mathrm{span}(\mathrm{supp}(U))_{\mathrm{s}}^{\perp}$, a subspace of dimension $2n-k$, commutes with U.

We next record how Pauli dimension behaves under conjugation.

Fact 7.57. *Let A and B be k- and ℓ-Pauli-dimensional matrices, respectively. Then $A^\dagger BA$ has Pauli dimension at most $k+\ell$, with support contained in* $\mathrm{span}(\mathrm{supp}(A) \cup \mathrm{supp}(B))$.

Proof. Write $A = \sum_{x \in G_1} \alpha_x W_x$ and $B = \sum_{y \in G_2} \beta_y W_y$, where G_1 and G_2 have dimensions k and ℓ, respectively. Then

$$\begin{aligned} A^\dagger BA &= \sum_{\substack{x_1, x_2 \in G_1 \\ y \in G_2}} \alpha_{x_1}^* \alpha_{x_2} \beta_y W_{x_1} W_y W_{x_2} \\ &= \sum_{\substack{x_1, x_2 \in G_1 \\ y \in G_2}} \alpha_{x_1}^* \alpha_{x_2} \beta_y (-1)^{[y, x_2]} W_{x_1} W_{x_2} W_y. \end{aligned}$$

Since $x_1, x_2 \in G_1$, the support of $A^\dagger BA$ lies in $G_1 + G_2$, whose dimension is at most $k+\ell$. □

Remark 7.58. If A and B are k- and ℓ-Pauli-dimensional, respectively, then AB has Pauli dimension at most $k+\ell$. The benefit of Fact 7.57 is that $A^\dagger BA$ obeys the same bound, rather than the naive $(2k+\ell)$-dimensional bound.

Using Facts 7.56 and 7.57, we can now show that $U^\dagger AU$ remains low-Pauli-dimensional when A has low Pauli dimension and U has small Clifford nullity.

Lemma 7.59. *Let $U \in \mathbb{C}^{2^n \times 2^n}$ have Clifford nullity t, and let A be a k-Pauli-dimensional matrix. Then $U^\dagger AU$ has Pauli dimension at most $k+t$.*

Proof. From Fact 7.56, $U = C_2 U' C_1$, where C_1, C_2 are Clifford unitaries and U' is t-Pauli-dimensional. Observe that

$$\begin{aligned} U^\dagger A U &= (C_2 U' C_1)^\dagger A (C_2 U' C_1) \\ &= C_1^\dagger \Big((U')^\dagger (C_2^\dagger A C_2) U' \Big) C_1 \end{aligned}$$

and $B \coloneqq C_2^\dagger A C_2$ remains k-Pauli-dimensional. Therefore $D \coloneqq (U')^\dagger B U'$ has Pauli dimension at most $k+t$ by Fact 7.57. Finally, $C_1^\dagger D C_1$ still has Pauli dimension at most $k+t$ because Clifford conjugation permutes Weyl operators. □

To connect this definition with a familiar circuit model, we recall a standard fact from the stabilizer literature: Clifford circuits doped with a small number of single-qubit non-Clifford gates have small Clifford nullity [LOLH24; GIKL24].

Fact 7.60. *A Clifford circuit augmented by t single-qubit non-Clifford gates has Clifford nullity at most $2t$.*[2]

This is actually a special case of the more general fact that alternations of juntas and Clifford circuits have bounded Clifford nullity. It follows that the same machinery also learns such circuits in composition with a shallow-depth circuit.

Fact 7.61. *Let U be a circuit that alternates between unitaries with nullity t_i and quantum juntas on k_i qubits. Then U has Clifford nullity at most $\sum_i (2k_i + t_i)$.*

Tomography Algorithm

We now establish the key technical fact: for U decomposable into a shallow circuit and a unitary of bounded Clifford nullity, the Heisenberg-evolved single-qubit Paulis remain low-dimensional.

[2] If the non-Clifford gates are limited to single-qubit Pauli rotations, such as the T gate, then the nullity is at most t.

Lemma 7.62. *Let $U \in \mathbb{C}^{2^n \times 2^n}$ be expressible as $U = QC$, where Q is a depth-d quantum circuit and C has Clifford nullity t. Put $q = \min\{2^d, n\}$. Then, for every weight-one Pauli operator P_i that acts nontrivially only on qubit i, $U^\dagger P_i U$ has Pauli dimension at most $2q + t$. More precisely, there are a Clifford C_2 and integers $a, b \geq 0$ such that $a + b \leq n$,*

$$2a + b \leq 2q + t, \qquad a + b \leq q + t, \qquad b \leq t,$$

and

$$\operatorname{supp}(C_2 U^\dagger P_i U C_2^\dagger) \subseteq \mathcal{W}_{a,b}.$$

Proof. The backward light cone of P_i under Q contains at most q qubits, so $Q^\dagger P_i Q$ has support in a Pauli subspace of type $(q, 0)$. In the proof of Lemma 7.59, conjugation by C enlarges the Pauli subspace containing this support by at most the t-dimensional support of the Clifford-nullity residual. The resulting subspace has dimension at most $2q + t$ and contains the original $2q$-dimensional nondegenerate subspace. If its canonical type is (a, b), the latter observation also gives $b \leq t$. Therefore $a + b \leq q + t$. A Clifford maps this subspace to $\mathcal{W}_{a,b}$, proving the claim. □

The preceding bounds give the learning theorem for a shallow circuit composed with a unitary of bounded Clifford nullity. Together with Fact 7.60, it also applies when the bounded-nullity circuit is a Clifford circuit augmented with s single-qubit non-Clifford gates, using $t \leq 2s$ below. For single-qubit Pauli rotations, one may take $t \leq s$.

Theorem 7.63. *Let $U \in \mathbb{C}^{2^n \times 2^n}$ be an n-qubit unitary that can be written either as $U = QC$ or $U = CQ$, where Q is a depth-d circuit and C has Clifford nullity t. Then, given query access to U and $U^\dagger$, there exists an algorithm that, with probability at least $1 - \delta$, outputs a $2n$-qubit unitary channel V such that $\mathsf{d}_\diamond(U^\dagger \otimes U, V) \leq \epsilon$ using*

$$O\left(2^{2^d + t}\left(2^{2^d} + t\right)\frac{n^2}{\epsilon}\log(n/\delta)\right)$$

queries to U and $U^\dagger$. Under the synthesis assumptions of Theorem 7.49, the runtime is

$$\operatorname{poly}\left(n, 2^{2^{d+1}+t}, \epsilon^{-1}, \log(1/\delta)\right).$$

Proof. First suppose $U = QC$. By Lemma 7.62, every $U^\dagger P_i U$ has support in $\mathcal{W}_{a,b}$ for parameters satisfying $2a + b \le 2q + t$, $a + b \le q + t$, and $b \le t$, where $q = \min\{2^d, n\}$. Substituting these parameters into Corollary 7.37 gives

$$2^{a+b}(2^a + b) = 2^{2a+b} + b2^{a+b} \le 2^{2q+t} + t2^{q+t} \le 2^{2^d+t}(2^{2^d} + t).$$

The standard single-qubit Pauli generating set has $m = 2n$ and constant local length. Applying Corollary 7.50 and Lemma 7.51 yields the stated query bound. If instead $U = CQ$, then $U^\dagger = Q^\dagger C^\dagger$ has the order just analyzed. Applying the algorithm to $U^\dagger$ gives an approximation to $U \otimes U^\dagger$; swapping the two n-qubit registers gives the required approximation to $U^\dagger \otimes U$. □

By Fact 7.61, an alternation of juntas and bounded-nullity unitaries has bounded Clifford nullity. The preceding theorem therefore gives the following corollary.

Corollary 7.64. *Let $U \in \mathbb{C}^{2^n \times 2^n}$ be an n-qubit unitary that can be written as*

$$U := \prod_i (J_i C_i)$$

where J_i is a junta on k_i qubits and C_i is a unitary with Clifford nullity t_i. Then, given query access to U and $U^\dagger$, there exists an algorithm that, with probability at least $1 - \delta$, outputs a $2n$-qubit unitary channel V such that $\mathsf{d}_\diamond(U^\dagger \otimes U, V) \le \epsilon$ using

$$O\left(2^t(t+1)\frac{n^2}{\epsilon}\log(n/\delta)\right)$$

queries to U and $U^\dagger$, with runtime polynomial in n, 2^t, $1/\epsilon$, and $\log(1/\delta)$, where $t = \sum_i (2k_i + t_i)$.

Proof. By Fact 7.61, the Clifford nullity of U is at most t. We then apply Theorem 7.63. □

7.9.3 Composition of Matchgate and Clifford Circuits

The states of non-interacting fermions studied in Section 4.1 have a parallel transformation model. Matchgate circuits are nonuniversal and classically simulable, but form an expressive circuit class [Val02; BGKT19]. Under the Jordan–Wigner transformation, they are equivalent to fermionic Gaussian unitaries [TD02; Kni01].

Definition 7.65 (Jordan–Wigner Majoranas). The Jordan–Wigner Majorana operators on n qubits are

$$\gamma_{2a-1} \coloneqq Z^{\otimes(a-1)} \otimes X \otimes I^{\otimes(n-a)}, \qquad \gamma_{2a} \coloneqq Z^{\otimes(a-1)} \otimes Y \otimes I^{\otimes(n-a)}$$

for $a \in [n]$.

Observe that the Jordan–Wigner Majoranas are a generating set of minimal size $2n$, but local length at most $2n$. They are also mutually anticommuting, which we will need to apply Fact 7.44.

Definition 7.66 (Matchgate circuit). A unitary U is a matchgate circuit if, for every $a \in [2n]$,

$$U^\dagger \gamma_a U = \sum_{b=1}^{2n} M_{ba} \gamma_b$$

for some real orthogonal matrix $M \in O(2n)$.

Thus the Heisenberg-evolved Jordan–Wigner Majoranas of a matchgate circuit are $2n$-Pauli-sparse.

Theorem 7.67. *Let $U \in \mathbb{C}^{2^n \times 2^n}$ be an n-qubit unitary that can be written either as $U = MC$ or $U = CM$, where M is a matchgate circuit and C is a Clifford circuit. Then, given query access to U and $U^\dagger$, there exists an algorithm that, with probability at least $1 - \delta$, outputs a $2n$-qubit unitary channel V such that $\mathsf{d}_\diamond(U^\dagger \otimes U, V) \leq \epsilon$ using*

$$O\left(n^6 \frac{n + \log(1/\delta)}{\epsilon^2}\right)$$

queries to U and $U^\dagger$, and

$$O\left(n^8 \frac{n + \log(1/\delta)}{\epsilon^2}\right)$$

time.

Proof. It suffices first to consider $U = MC$. Conjugating a Jordan–Wigner Majorana by a matchgate produces a linear combination of $2n$ mutually anticommuting Paulis. A subsequent Clifford conjugation preserves both anticommutation and sparsity. Thus Fact 7.44

gives an efficiently implementable unitary estimate, and Lemma 7.51 fixes its sign. Apply Corollary 7.50 with $|G| = 2n$, $s = 2n$, and local length $L = 2n$. If $U = CM$, apply the proved case to $U^\dagger = M^\dagger C^\dagger$, then swap the two output registers exactly as in the shallow–near-Clifford case. □

We next prove lower bounds for several of the structural parameters in these learning guarantees.

7.10 Lower Bounds

In this section, we prove query lower bounds for several unitary learning problems. We first consider low-Pauli-dimensional and Pauli-sparse unitaries, as well as quantum juntas. We then discuss lower bounds for learning unitaries that can be expressed as compositions of shallow and near-Clifford circuits.

7.10.1 Lower Bounds for Pauli Dimensionality, Sparsity, and Quantum Juntas

Our lower bounds leverage [HKOT23, Theorem 1.2], which showed that $\Omega(d^2/\epsilon)$ queries are necessary to learn $d \times d$ unitary matrices, along with padding arguments. These lower bounds establish the query optimality of Corollary 7.54 and show that Corollary 7.38 is optimal in its exponential and precision dependence, up to its polynomial factor in k. For sparsity, we prove an $\Omega(s/\epsilon)$ lower bound, so a gap remains between that and our $O(s^2/\epsilon^2)$ upper bound in Theorem 7.42.

Lemma 7.68 ([HKOT23, Theorem 1.2]). *Let $d \geq 2$, and let $\mathcal{A}$ be an algorithm that, for an unknown unitary $U \in \mathbb{C}^{d\times d}$ accessible through black-box oracles that implement $U, U^\dagger$, $cU = |0\rangle\langle 0| \otimes I_d + |1\rangle\langle 1| \otimes U$, and $cU^\dagger = |0\rangle\langle 0| \otimes I_d + |1\rangle\langle 1| \otimes U^\dagger$, can output a classical description of a unitary V such that $\mathsf{d}_\diamond(U, V) < \epsilon < \frac{1}{8}$ with probability $\geq \frac{2}{3}$. Then $\mathcal{A}$ must use $\Omega(d^2/\epsilon)$ oracle queries.*

Theorem 7.69. *Let $1 \leq k \leq n$, and let $\mathcal{A}$ be an algorithm that, for an unknown k-junta $U \in \mathbb{C}^{2^n\times 2^n}$ accessible through black-box oracles that implement $U, U^\dagger$, $cU = |0\rangle\langle 0| \otimes I_{2^n} + |1\rangle\langle 1| \otimes U$,*

and $cU^\dagger = |0\rangle\langle 0| \otimes I_{2^n} + |1\rangle\langle 1| \otimes U^\dagger$*, can output a classical description of a unitary* V *such that* $\mathsf{d}_\diamond(U, V) < \epsilon < \frac{1}{8}$ *with probability* $\geq \frac{2}{3}$*. Then* $\mathcal{A}$ *must use* $\Omega(4^k/\epsilon)$ *oracle queries.*

Proof. By Lemma 7.68, learning arbitrary unitaries on k qubits requires $\Omega(4^k/\epsilon)$ queries. Padding each such unitary with $I^{\otimes n-k}$ gives an n-qubit k-junta. A learner for this padded family using $o(4^k/\epsilon)$ queries would therefore contradict Lemma 7.68; the unused $n-k$ qubits merely act as ancillas. □

Corollary 7.70. *Let* $2 \leq k \leq 2n$*, and let* $\mathcal{A}$ *be an algorithm that, for an unknown unitary of Pauli dimension at most* k*,* $U \in \mathbb{C}^{2^n \times 2^n}$ *accessible through black-box oracles that implement* U*,* $U^\dagger$*,* $cU = |0\rangle\langle 0| \otimes I_{2^n} + |1\rangle\langle 1| \otimes U$*, and* $cU^\dagger = |0\rangle\langle 0| \otimes I_{2^n} + |1\rangle\langle 1| \otimes U^\dagger$*, can output a classical description of a unitary* V *such that* $\mathsf{d}_\diamond(U, V) < \epsilon < \frac{1}{8}$ *with probability* $\geq \frac{2}{3}$*. Then* $\mathcal{A}$ *must use* $\Omega(2^k/\epsilon)$ *oracle queries.*

Proof. Put $r = \lfloor k/2 \rfloor$. Every r-junta has Pauli dimension at most $2r \leq k$, so a learner for the class of unitaries of Pauli dimension at most k would learn all r-juntas. By Theorem 7.69, it therefore needs

$$\Omega(4^r/\epsilon) = \Omega(2^k/\epsilon)$$

queries. □

We can also show an $\Omega(s/\epsilon)$ lower bound for Pauli sparsity.

Corollary 7.71. *Let* $4 \leq s \leq 4^n$*, and let* $\mathcal{A}$ *be an algorithm that, for an unknown* s*-Pauli-sparse unitary* $U \in \mathbb{C}^{2^n \times 2^n}$ *accessible through black-box oracles that implement* U*,* $U^\dagger$*,* $cU = |0\rangle\langle 0| \otimes I_{2^n} + |1\rangle\langle 1| \otimes U$*, and* $cU^\dagger = |0\rangle\langle 0| \otimes I_{2^n} + |1\rangle\langle 1| \otimes U^\dagger$*, can output a classical description of a unitary* V *such that* $\mathsf{d}_\diamond(U, V) < \epsilon < \frac{1}{8}$ *with probability* $\geq \frac{2}{3}$*. Then* $\mathcal{A}$ *must use* $\Omega(s/\epsilon)$ *oracle queries.*

Proof. Let $s' = 2^{\lfloor \log_2 s \rfloor}$, so $s/2 \leq s' \leq s$. The hard family of Pauli dimension $\log_2 s'$ is s'-Pauli sparse and hence also s-Pauli-sparse. By Corollary 7.70, learning it requires $\Omega(s'/\epsilon) = \Omega(s/\epsilon)$ queries. □

7.10.2 Lower Bounds for the Composition of Shallow and Near-Clifford Circuits

We now prove lower bounds for learning compositions of near-Clifford circuits with shallow circuits. An $\Omega(2^t)$ dependence in the query complexity of Theorem 7.63 is unavoidable in general, because the class of unitaries with Clifford nullity at most $2t$ contains all quantum t-juntas. The lower bound of Theorem 7.69 therefore applies. However, it remains open whether the same lower bound holds in the more restricted setting of Clifford circuits augmented with t single-qubit non-Clifford gates.

We also give a short proof that $\exp(\exp(\Omega(d)))$ query complexity is necessary even when inverse queries are allowed. This follows from the lower bound of [BBBV97] for Grover search via a padding argument. The corresponding non-padded argument appears in [Hua+24, Proposition 3], which shows a weaker $\Omega(\exp(n))$ lower bound for $d = O(\log n)$.

Lemma 7.72. *Circuits of depth d require $\exp(\exp(\Omega(d)))$ queries to learn to diamond distance $\epsilon < 1$ with constant success probability.*

Proof. A multi-controlled phase gate on k qubits can be implemented in depth $d = O(\log k)$ using clean ancillas.[3] This multi-controlled phase gate can be used to build the following family of unitary channels $\{U_y\}_{y\in\{0,1\}^k}$ on k qubits using two extra layers of Pauli X gates:

$$U_y|x\rangle = \begin{cases} (-1)|x\rangle & x = y \\ |x\rangle & x \neq y \end{cases}$$

meaning that this gate can also be implemented in depth $d = O(\log k)$. Distinguishing the identity from an unknown U_y is the unstructured search problem on 2^k possible marked strings. This provably requires $\Omega(2^{k/2}) = 2^{\exp(\Omega(d))}$ queries to achieve constant success probability [BBBV97].

Each U_y is maximally far from the identity in diamond distance. Indeed, for any nonzero bit string $e \in \{0,1\}^k$,

$$|\psi_y\rangle = \frac{|y\rangle + |y \oplus e\rangle}{\sqrt{2}}$$

[3] This is the relevant form of the inclusion $\mathsf{QAC}^0 \subseteq \mathsf{QNC}^1$.

is mapped by U_y to $(-|y\rangle + |y \oplus e\rangle)/\sqrt{2}$, which is orthogonal to $|\psi_y\rangle$. Therefore, learning to diamond distance strictly less than 1 allows one to distinguish the identity channel from the U_y. It follows that such an algorithm must use $\exp(\exp(\Omega(d)))$ queries even with inverse access, as $U_y = U_y^\dagger$ and $I = I^\dagger$. □

7.11 A Simultaneous Symplectic Basis Extension

The reduction in Section 7.5.2 used a simultaneous symplectic basis for a learned subspace T and the full support span S. We postponed the constructive linear-algebra proof to keep the main learning argument together. We now supply it, thereby completing the proof of Lemma 7.32.

Lemma 7.32. *Let $T \subseteq S \subseteq \mathbb{F}_2^{2n}$, where T has symplectic type (a', b') and S has type (a, b). There is an integer ℓ satisfying*

$$0 \le \ell \le \min\{b', a - a'\}, \qquad b - b' + \ell \ge 0,$$

and a simultaneous symplectic basis such that

$$\begin{aligned} T = \ & \operatorname{span}\{z_1, x_1, \dots, z_{a'}, x_{a'}, z_{a'+1}, \dots, z_{a'+b'}\}, \\ S = \ & \operatorname{span}\big(T \cup \{x_{a'+1}, \dots, x_{a'+\ell}\} \\ & \qquad \cup \{z_{a'+b'+1}, x_{a'+b'+1}, \dots, z_{a+b'-\ell}, x_{a+b'-\ell}\} \\ & \qquad \cup \{z_{a+b'-\ell+1}, \dots, z_{a+b}\}\big), \end{aligned}$$

where $[x_i, x_j] = [z_i, z_j] = 0$ and $[x_i, z_j] = \delta_{ij}$ for all i, j. Moreover, such a basis can be found in time $O(n(a+b)^2)$ given any generating sets for T and S.

Proof. Choose a basis G_T for T, and extend it by a list G_R so that $G_T \cup G_R$ is a basis for S. We use the elementary symplectic elimination operation associated with a pair z, x satisfying $[x, z] = 1$:

$$v \longmapsto v + [v, x]z + [v, z]x.$$

It replaces v by a vector orthogonal to both x and z. Applying the same elementary column operations to a generating list preserves its span.

First run ordinary symplectic Gram–Schmidt on G_T. Whenever two current generators anticommute, record them as a pair z_i, x_i and orthogonalize every remaining vector of both G_T and G_R against that pair. When no such pair remains, the recorded vectors form a' symplectic pairs, and the b' vectors left in G_T form a basis R_T for the radical of the restricted form on T. In particular, the vectors in R_T commute with all of T.

The second phase determines which radical vectors of T acquire symplectic partners in the larger space S. If $z \in R_T$ and a remaining vector $x \in G_R$ satisfy $[x, z] = 1$, record z, x as a pair and orthogonalize all remaining vectors against it. For a remaining vector $z' \in R_T$, this update only adds a multiple of z, so the updated radical list still spans a subspace of T. Continue until no such pair exists, and let ℓ be the number of pairs obtained. Relabel these as

$$z_{a'+1}, x_{a'+1}, \dots, z_{a'+\ell}, x_{a'+\ell}.$$

The $b' - \ell$ unpaired elements of R_T commute with every remaining vector of S; label them $z_{a'+\ell+1}, \dots, z_{a'+b'}$.

Finally, run ordinary symplectic Gram–Schmidt on the vectors still in G_R. This produces symplectic pairs and an isotropic remainder. Because the type of S is (a, b), the number of new pairs is $a - a' - \ell$, and the number of new isotropic vectors is $b - b' + \ell$.

Label the new pairs by the indices

$$a' + b' + 1, \dots, a + b' - \ell$$

and the new isotropic vectors by

$$a + b' - \ell + 1, \dots, a + b.$$

This gives exactly the two generating lists in the statement. It also proves

$$0 \le \ell \le \min\{b', a - a'\}, \qquad b - b' + \ell \ge 0.$$

Each elimination step makes all later vectors orthogonal to the pair just chosen, so the resulting vectors satisfy $[x_i, x_j] = [z_i, z_j] = 0$ and $[x_i, z_j] = \delta_{ij}$.

At most $O((a+b)^2)$ symplectic products and vector additions are performed. Each costs $O(n)$ over $\mathbb{F}_2$, giving total time $O(n(a+b)^2)$. □

Bibliography

[AAGW26] Omar Alrabiah, Srinivasan Arunachalam, Sabee Grewal, and John Wright. *No Low-Degree Tests for Quantum States*. 2026. DOI: 10.48550/arXiv.2608.00265. arXiv: 2608.00265 [quant-ph].

[Aar+24] Scott Aaronson, Mohammad Bavarian, Toby Cubitt, Sabee Grewal, Giulio Gueltrini, Ryan O'Donnell, and Marien Raat. *Computability Theory of Closed Timelike Curves*. 2024. DOI: 10.48550/arXiv.1609.05507. arXiv: 1609.05507 [quant-ph].

[Aar20] Scott Aaronson. "Shadow Tomography of Quantum States". In: *SIAM Journal on Computing* (2020). DOI: 10.1137/18M120275X.

[Aar22] Scott Aaronson. *Introduction to Quantum Information Science II: Lecture Notes*. 2022. URL: https://scottaaronson.com/qisii.pdf.

[AB08] A. Angelow and M.-C. Batoni. *About Heisenberg Uncertainty Relation (by E. Schrödinger)*. 2008. arXiv: quant-ph/9903100 [quant-ph].

[ABDY23] Srinivasan Arunachalam, Sergey Bravyi, Arkopal Dutt, and Theodore J. Yoder. "Optimal Algorithms for Learning Quantum Phase States". In: *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*. 2023. DOI: 10.4230/LIPIcs.TQC.2023.3.

[ACDG26] Srinivasan Arunachalam, Davi Castro-Silva, Arkopal Dutt, and Tom Gur. "Classical and Quantum Polynomial Freiman–Ruzsa Algorithms". In: *17th Innovations in Theoretical Computer Science Conference (ITCS 2026)*. 2026. DOI: 10.4230/LIPIcs.ITCS.2026.11.

[ACGN23] 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.

[AD26] Srinivasan Arunachalam and Arkopal Dutt. *Tomography of Quantum States with Bounded Extent*. 2026. DOI: 10.48550/arXiv.2606.07425. arXiv: 2606.07425 [quant-ph].

[ADGD26] Srinivasan Arunachalam, Arkopal Dutt, Alexandru Gheorghiu, and Michael De Oliveira. "Learning Depth-3 Circuits via Quantum Agnostic Boosting". In: *Proceedings of the 39th Conference on Learning Theory*. 2026. URL: https://proceedings.mlr.press/v336/arunachalam26a.html.

[ADKL25] Alvan Arulandu, Ilias Diakonikolas, Daniel Kane, and Jerry Li. *Agnostic Product Mixed State Tomography via Robust Statistics*. 2025. DOI: 10.48550/arXiv.2510.08472. arXiv: 2510.08472 [quant-ph].

[ADLY25a] Jayadev Acharya, Abhilash Dharmavarapu, Yuhan Liu, and Nengkun Yu. *Pauli Measurements Are Near-Optimal for Single-Qubit Tomography*. 2025. arXiv: 2507.22001.

[ADLY25b] Jayadev Acharya, Abhilash Dharmavarapu, Yuhan Liu, and Nengkun Yu. "Pauli measurements are not optimal for single-copy tomography". In: *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*. 2025.

[AG04] Scott Aaronson and Daniel Gottesman. "Improved Simulation of Stabilizer Circuits". In: *Physical Review A* (2004). DOI: 10.1103/PhysRevA.70.052328.

[AG23] Scott Aaronson and Sabee Grewal. "Efficient Tomography of Non-Interacting-Fermion States". In: *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*. 2023. DOI: 10.4230/LIPIcs.TQC.2023.12.

[AGIMR25] Scott Aaronson, Sabee Grewal, Vishnu Iyer, Simon C. Marshall, and Ronak Ramachandran. "PDQMA = DQMA = NEXP: QMA with Hidden Variables and Non-Collapsing Measurements". In: *45th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2025)*. 2025. DOI: 10.4230/LIPIcs.FSTTCS.2025.3.

[AKKLR05] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. "Testing Reed–Muller Codes". In: *IEEE Transactions on Information Theory* (2005). DOI: 10.1109/TIT.2005.856958.

[ASW15] Emmanuel Abbe, Amir Shpilka, and Avi Wigderson. "Reed–Muller Codes for Random Erasures and Errors". In: *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*. 2015. DOI: 10.1145/2746539.2746575.

[Bak+25] Ainesh Bakshi, John Bostanci, William Kretschmer, Zeph Landau, Jerry Li, Allen Liu, Ryan O'Donnell, and Ewin Tang. "Learning the Closest Product State". In: *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*. 2025. DOI: 10.1145/3717823.3718207.

[BBBV97] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. "Strengths and Weaknesses of Quantum Computing". In: *SIAM J. Comput.* (1997). DOI: 10.1137/S0097539796300933.

[BC26] Jop Briët and Davi Castro-Silva. "A Near-Optimal Quadratic Goldreich–Levin Algorithm". In: *Proceedings of the 2026 Annual ACM–SIAM Symposium on Discrete Algorithms*. 2026. DOI: 10.1137/1.9781611978971.224.

[BCHK20] Michael Beverland, Earl Campbell, Mark Howard, and Vadym Kliuchnikov. "Lower Bounds on the Non-Clifford Resources for Quantum Computations". In: *Quantum Science and Technology* (2020). DOI: 10.1088/2058-9565/ab8963.

[Ber21] Ewout van den Berg. "A Simple Method for Sampling Random Clifford Operators". In: *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*. 2021. DOI: 10.1109/QCE52317.2021.00021.

[BG16] Sergey Bravyi and David Gosset. "Improved Classical Simulation of Quantum Circuits Dominated by Clifford Gates". In: *Physical Review Letters* (2016). DOI: 10.1103/PhysRevLett.116.250501.

[BGKT19] Sergey Bravyi, David Gosset, Robert König, and Kristan Temme. "Approximation algorithms for quantum many-body problems". In: *Journal of Mathematical Physics* (2019). DOI: 10.1063/1.5085428.

[BK05] Sergey Bravyi and Alexei Kitaev. "Universal Quantum Computation with Ideal Clifford Gates and Noisy Ancillas". In: *Physical Review A* (2005). DOI: 10.1103/PhysRevA.71.022316.

[BKSSZ10] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. "Optimal Testing of Reed–Muller Codes". In: *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. 2010. DOI: 10.1109/FOCS.2010.54.

[BM21] Sergey Bravyi and Dmitri Maslov. "Hadamard-Free Circuits Expose the Structure of the Clifford Group". In: *IEEE Transactions on Information Theory* (2021). DOI: 10.1109/TIT.2021.3081415.

[BMEL25] Lennart Bittel, Antonio Anna Mele, Jens Eisert, and Lorenzo Leone. "Optimal Trace-Distance Bounds for Free-Fermionic States: Testing and Improved Tomography". In: *PRX Quantum* (2025). DOI: 10.1103/pzx6-nkfb.

[CGYZ25] Sitan Chen, Weiyuan Gong, Qi Ye, and Zhihan Zhang. "Stabilizer Bootstrapping: A Recipe for Efficient Agnostic Tomography and Magic Estimation". In: *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*. 2025. DOI: 10.1145/3717823.3718191.

[CKS17] Andrew M. Childs, Robin Kothari, and Rolando D. Somma. "Quantum Algorithm for Systems of Linear Equations with Exponentially Improved Dependence on Precision". In: *SIAM Journal on Computing* (2017). DOI: 10.1137/16M1087072.

[Cle20] Richard Cleve. *Quantum Information Processing*. Video lecture playlist, University of Waterloo. YouTube. 2020. URL: https://www.youtube.com/playlist?list=PLidiQIHRzpXJY_EpEXqslQ7FjcpC_dIyy (visited on 08/10/2026).

[CNY23] Thomas Chen, Shivam Nadimpalli, and Henry Yuen. "Testing and Learning Quantum Juntas Nearly Optimally". In: *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2023. DOI: 10.1137/1.9781611977554.ch43.

[CZ26] Aria Christensen and Andrew Zhao. *Learning Fermionic Linear Optics with Heisenberg Scaling and Physical Operations*. 2026. DOI: 10.48550/arXiv.2602.05058. arXiv: 2602.05058 [quant-ph].

[DBBHK25] Lucas Daguerre, Robin Blume-Kohout, Natalie C. Brown, David Hayes, and Isaac H. Kim. "Experimental Demonstration of High-Fidelity Logical Magic States from Code Switching". In: *Physical Review X* (2025). DOI: 10.1103/dck4-x9c2.

[Del+08] Samuel Deléglise, Igor Dotsenko, Clément Sayrin, Julien Bernu, Michel Brune, Jean-Michel Raimond, and Serge Haroche. "Reconstruction of Non-Classical Cavity Field States with Snapshots of Their Decoherence". In: *Nature* (2008). DOI: 10.1038/nature07288.

[DJJMP26] Arkopal Dutt, Dale Jacobs, John Jeang, Saeed Mehraban, and Vladimir Podolskii. *Learning Clifford-Structured Quantum Unitaries and Hamiltonians*. 2026. DOI: 10.48550/arXiv.2608.09912. arXiv: 2608.09912 [quant-ph].

[Fan57] 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.

[FCYBC04] David Fattal, Toby S. Cubitt, Yoshihisa Yamamoto, Sergey Bravyi, and Isaac L. Chuang. *Entanglement in the stabilizer formalism*. 2004. arXiv: quant-ph/0406168 [quant-ph].

[FGLE12] 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. arXiv: 1205.2300 [quant-ph].

[FL11] Steven T. Flammia and Yi-Kai Liu. "Direct fidelity estimation from few Pauli measurements". In: *Physical review letters* (2011). DOI: 10.1103/PhysRevLett.106.230501.

[Gau59] Walter Gautschi. "Some Elementary Inequalities Relating to the Gamma and Incomplete Gamma Function". In: *Journal of Mathematics and Physics* (1959). DOI: 10.1002/sapm195938177.

[Ger13] Manuel Gerken. *Measure Concentration: Lévy's Lemma*. 2013.

[GGHSS26] Sabee Grewal, Meghal Gupta, William He, Aniruddha Sen, and Mihir Singhal. *Nearly Time-Optimal Pure State Tomography with Pauli Measurements*. To appear in the 67th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2026). 2026. DOI: 10.48550/arXiv.2601.04444. arXiv: 2601.04444 [quant-ph].

[GHO25] Meghal Gupta, William He, and Ryan O'Donnell. *Few Single-Qubit Measurements Suffice to Certify Any Quantum State*. 2025. arXiv: 2506.11355.

[GIKL23] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. "Low-Stabilizer-Complexity Quantum States Are Not Pseudorandom". In: *14th Innovations in Theoretical Computer Science Conference*. 2023. DOI: 10.4230/LIPIcs.ITCS.2023.64.

[GIKL24] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. "Improved Stabilizer Estimation via Bell Difference Sampling". In: *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*. 2024. DOI: 10.1145/3618260.3649738.

[GIKL25] 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.

[GIKL26a] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. “Agnostic Tomography of Stabilizer Product States”. In: *Quantum* (2026). DOI: 10.22331/q-2026-03-13-2027.

[GIKL26b] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. “Lower Bounds on the Non-Clifford Cost of Pseudoentanglement”. In: *Physical Review A* (2026). DOI: 10.1103/v493-8s1q. arXiv: 2404.00126 [quant-ph].

[GK24] Sabee Grewal and Vinayak M. Kumar. *Improved Circuit Lower Bounds and Quantum-Classical Separations*. Electronic Colloquium on Computational Complexity, Report TR24-130. 2024. DOI: 10.48550/arXiv.2408.16406. arXiv: 2408.16406 [quant-ph].

[GK25] Sabee Grewal and William Kretschmer. *Unentanglement and Post-Measurement Branching in Quantum Interactive Proofs*. Electronic Colloquium on Computational Complexity, Report TR25-145; to appear in TQC 2026. 2025. DOI: 10.48550/arXiv.2509.15319. arXiv: 2509.15319 [quant-ph].

[GKKT20] Madalin Guţă, Jonas Kahn, Richard Kueng, and Joel A Tropp. “Fast state tomography with optimal error bounds”. In: *Journal of Physics A: Mathematical and Theoretical* (2020).

[GL25] Sabee Grewal and Daniel Liang. *Efficient Learning of Structured Quantum Circuits via Pauli Dimensionality and Sparsity*. 2025. DOI: 10.48550/arXiv.2510.00168. arXiv: 2510.00168.

[GLFBE10] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. “Quantum state tomography via compressed sensing”. In: *Phys. Rev. Lett.* (2010). DOI: 10.1103/PhysRevLett.105.150401. arXiv: 0909.3304 [quant-ph].

[GNW21] David Gross, Sepehr Nezami, and Michael Walter. “Schur–Weyl Duality for the Clifford Group with Applications: Property Testing, a Robust Hudson Theorem, and de Finetti Representations”. In: *Communications in Mathematical Physics* (2021). DOI: 10.1007/s00220-021-04118-7.

[GOSSW11] Parikshit Gopalan, Ryan O’Donnell, Rocco A. Servedio, Amir Shpilka, and Karl Wimmer. “Testing Fourier Dimensionality and Sparsity”. In: *SIAM Journal on Computing* (2011). DOI: 10.1137/100785429.

[Got97] Daniel Gottesman. “Stabilizer Codes and Quantum Error Correction”. PhD thesis. California Institute of Technology, 1997. DOI: 10.7907/rzr7-dt72.

[Got98] Daniel Gottesman. *The Heisenberg Representation of Quantum Computers*. 1998. arXiv: quant-ph/9807006 [quant-ph]. URL: https://arxiv.org/abs/quant-ph/9807006.

[GR26] Sabee Grewal and Dorian Rudolph. “On the Pure Quantum Polynomial Hierarchy and Quantified Hamiltonian Complexity”. In: *53rd International Colloquium on Automata, Languages, and Programming (ICALP 2026)*. 2026. DOI: 10.4230/LIPIcs.ICALP.2026.103.

[GSLW19] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. “Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics”. In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. 2019. DOI: 10.1145/3313276.3316366.

[Gup+24] Riddhi S. Gupta, Neereja Sundaresan, Thomas Alexander, Christopher J. Wood, Seth T. Merkel, Michael B. Healy, Marius Hillenbrand, Tomas Jochym-O’Connor, James R. Wootton, Theodore J. Yoder, Andrew W. Cross, Maika Takita, and Benjamin J. Brown. “Encoding a Magic State with Beyond Break-Even Fidelity”. In: *Nature* (2024). DOI: 10.1038/s41586-023-06846-3.

[GY24] Sabee Grewal and Justin Yirka. “The Entangled Quantum Polynomial Hierarchy Collapses”. In: *39th Computational Complexity Conference (CCC 2024)*. 2024. DOI: 10.4230/LIPIcs.CCC.2024.6.

[Har13] Aram W. Harrow. *The Church of the Symmetric Subspace*. 2013. DOI: 10.48550/arXiv.1308.6595. arXiv: 1308.6595 [quant-ph].

[Hay98] Masahito Hayashi. “Asymptotic estimation theory for a finite-dimensional pure state model”. In: *Journal of Physics A: Mathematical and General* (1998).

[HHJWY16] 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.

[HJ91] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. 1991. DOI: 10.1017/CBO9780511840371.

[HKOT23] Jeongwan Haah, Robin Kothari, Ryan O’Donnell, and Ewin Tang. “Query-optimal estimation of unitary channels in diamond distance”. In: *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*. 2023. DOI: 10.1109/FOCS57990.2023.00028.

[HKP21] Hsin-Yuan Huang, Richard Kueng, and John Preskill. “Information-Theoretic Bounds on Quantum Advantage in Machine Learning”. In: *Physical Review Letters* (2021). DOI: 10.1103/PhysRevLett.126.190505.

[HP98] Peter Heijnen and Ruud Pellikaan. “Generalized Hamming Weights of q-ary Reed–Muller Codes”. In: *IEEE Transactions on Information Theory* (1998). DOI: 10.1109/18.651015.

[HPS25] Hsin-Yuan Huang, John Preskill, and Mehdi Soleimanifar. “Certifying almost all quantum states with few single-qubit measurements”. In: *Nature Physics* (2025).

[Hua+24] Hsin-Yuan Huang, Yunchao Liu, Michael Broughton, Isaac Kim, Anurag Anshu, Zeph Landau, and Jarrod R. McClean. “Learning Shallow Quantum Circuits”.

In: *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*. 2024. DOI: 10.1145/3618260.3649722.

[JLS18] Zhengfeng Ji, Yi-Kai Liu, and Fang Song. "Pseudorandom Quantum States". In: *Advances in Cryptology—CRYPTO 2018*. 2018. DOI: 10.1007/978-3-319-96878-0_5.

[JPRZ04] Charanjit S. Jutla, Anindya C. Patthak, Atri Rudra, and David Zuckerman. "Testing Low-Degree Polynomials over Prime Fields". In: *45th Annual IEEE Symposium on Foundations of Computer Science*. 2004. DOI: 10.1109/FOCS.2004.64.

[JW23] Jiaqing Jiang and Xin Wang. "Lower Bound for the T Count Via Unitary Stabilizer Nullity". In: *Phys. Rev. Appl.* (3 2023). DOI: 10.1103/PhysRevApplied.19.034052.

[Kni01] E. Knill. *Fermionic Linear Optics and Matchgates*. 2001. arXiv: quant-ph/0108033 [quant-ph]. URL: https://arxiv.org/abs/quant-ph/0108033.

[Kre+25] William Kretschmer, Sabee Grewal, Matthew DeCross, Justin A. Gerber, Kevin Gilmore, Dan Gresh, Nicholas Hunter-Jones, Karl Mayer, Brian Neyenhuis, David Hayes, and Scott Aaronson. *Demonstrating an Unconditional Separation Between Quantum and Classical Information Resources*. 2025. DOI: 10.48550/arXiv.2509.07255. arXiv: 2509.07255 [quant-ph].

[KRT17] 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.

[KS14] Robert Koenig and John A. Smolin. "How to Efficiently Select an Arbitrary Clifford Group Element". In: *Journal of Mathematical Physics* (2014). DOI: 10.1063/1.4903507.

[Liu11] Yi-Kai Liu. *Universal low-rank matrix recovery from Pauli measurements*. 2011. arXiv: 1103.2816 [quant-ph].

[LN22] Angus Lowe and Ashwin Nayak. *Lower bounds for learning quantum states with single-copy measurements*. 2022. arXiv: 2207.14438.

[LOLH24] Lorenzo Leone, Salvatore F. E. Oliviero, Seth Lloyd, and Alioscia Hamma. "Learning Efficient Decoders for Quasichaotic Quantum Scramblers". In: *Physical Review A* (2024). DOI: 10.1103/PhysRevA.109.022429.

[Ma+12] Xiao-Song Ma, Thomas Herbst, Thomas Scheidl, Daqing Wang, Sebastian Kropatschek, William Naylor, Bernhard Wittmann, Alexandra Mech, Johannes Kofler, Elena Anisimova, Vadim Makarov, Thomas Jennewein, Rupert Ursin, and Anton Zeilinger. "Quantum Teleportation over 143 Kilometres Using Active Feed-Forward". In: *Nature* (2012). DOI: 10.1038/nature11472.

[Mac75] Odile Macchi. "The Coincidence Approach to Stochastic Point Processes". In: *Advances in Applied Probability* (1975). DOI: 10.2307/1425855.

[MO10] Ashley Montanaro and Tobias J. Osborne. *Quantum boolean functions*. 2010. arXiv: 0810.2435 [quant-ph].

[Mon17] Ashley Montanaro. *Learning Stabilizer States by Bell Sampling*. 2017. DOI: 10.48550/arXiv.1707.04012. arXiv: 1707.04012 [quant-ph].

[MW16] Ashley Montanaro and Ronald de Wolf. *A Survey of Quantum Property Testing*. 2016. DOI: 10.4086/toc.gs.2016.007. URL: https://doi.org/10.4086/toc.gs.2016.007.

[NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. 10th anniversary. 2010. DOI: 10.1017/CBO9780511976667.

[ODMZ22] Michał Oszmaniec, Ninnat Dangniam, Mauro E. S. Morales, and Zoltán Zimborás. "Fermion Sampling: A Robust Quantum Computational Advantage Scheme Using Fermionic Linear Optics and Magic Input States". In: *PRX Quantum* (2022). DOI: 10.1103/PRXQuantum.3.020328.

[ODo24] Ryan O'Donnell. *Quantum Computer Programming in 100 Easy Lessons*. Video lecture playlist. YouTube. 2024. URL: https://www.youtube.com/playlist?list=PLm3J0oaFux3bF48kurxGR6jrmPaQf6lkN (visited on 08/10/2026).

[OW16] Ryan O'Donnell and John Wright. "Efficient quantum tomography". In: *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. 2016.

[Pau33] Wolfgang Pauli. "Die allgemeinen Prinzipien der Wellenmechanik". In: *Handbuch der Physik*. 2nd ed. .1. 1933.

[PSTW25] Angelos Pelecanos, Jack Spilecki, Ewin Tang, and John Wright. *Mixed state tomography reduces to pure state tomography*. 2025. arXiv: 2511.15806.

[PSW25] Angelos Pelecanos, Jack Spilecki, and John Wright. *The debiased Keyl's algorithm: a new unbiased estimator for full state tomography*. 2025. arXiv: 2510.07788.

[QR22] Yihui Quek and Patrick Rebentrost. "Fast algorithm for quantum polar decomposition and applications". In: *Phys. Rev. Res.* (1 2022). DOI: 10.1103/PhysRevResearch.4.013144.

[QXFL26] Wentao Qi, Boyan Xu, Shiguang Feng, and Lvzhou Li. *Agnostic Learning of Qudit Stabilizer States*. 2026. DOI: 10.48550/arXiv.2607.15559. arXiv: 2607.15559 [quant-ph].

[Ral20] Patrick Rall. "Quantum algorithms for estimating physical quantities using block encodings". In: *Phys. Rev. A* (2 2020). DOI: 10.1103/PhysRevA.102.022408.

[Röt09] Martin Rötteler. "Quantum Algorithms to Solve the Hidden Shift Problem for Quadratics and for Functions of Large Gowers Norm". In: *Proceedings of the 34th International Symposium on Mathematical Foundations of Computer Science*. 2009. DOI: 10.1007/978-3-642-03816-7_56.

[RZBB94] Michael Reck, Anton Zeilinger, Herbert J. Bernstein, and Philip Bertani. "Experimental Realization of Any Discrete Unitary Operator". In: *Physical Review Letters* (1994). DOI: 10.1103/PhysRevLett.73.58.

[Sch30] Erwin Schrödinger. "The Uncertainty Relations in Quantum Mechanics: Zum Heisenbergschen Unschärfeprinzip". In: *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse* (1930).

[SSW25] Thilo Scharnhorst, Jack Spilecki, and John Wright. *Optimal lower bounds for quantum state tomography*. 2025. arXiv: 2510.07699.

[TD02] Barbara M. Terhal and David P. DiVincenzo. "Classical Simulation of Noninteracting-Fermion Quantum Circuits". In: *Physical Review A* (2002). DOI: 10.1103/PhysRevA.65.032325.

[TK23] Ewin Tang and Christopher Kang. *Quantum and quantum-inspired linear algebra*. 2023. URL: https://ewintang.com/assets/tang-pcmi-lectures.pdf.

[Val02] Leslie G. Valiant. "Quantum Circuits That Can Be Simulated Classically in Polynomial Time". In: *SIAM Journal on Computing* (2002). DOI: 10.1137/S0097539700377025.

[Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. 1st. 2018. DOI: 10.1017/9781108231596.

[Vor13] Vladislav Voroninski. *Quantum Tomography From Few Full-Rank Observables*. 2013. arXiv: 1309.7669 [math-ph]. URL: https://arxiv.org/abs/1309.7669.

[Wal77] Alastair J Walker. "An efficient method for generating discrete random variables with general distributions". In: *ACM Transactions on Mathematical Software (TOMS)* (1977).

[Wat25] John Watrous. *Understanding Quantum Information and Computation*. 2025. DOI: 10.48550/arXiv.2507.11536. arXiv: 2507.11536 [quant-ph]. URL: https://arxiv.org/abs/2507.11536.

[Wen48] James G. Wendel. "Note on the Gamma Function". In: *The American Mathematical Monthly* (1948). DOI: 10.2307/2304460.

[Wey12] Hermann Weyl. "Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)". In: *Mathematische Annalen* (1912). DOI: 10.1007/BF01456804.

[Wil09] Mark M. Wilde. "Logical operators of quantum codes". In: *Phys. Rev. A* (6 2009). DOI: 10.1103/PhysRevA.79.062322.

[WLKD25] Chirag Wadhwa, Laura Lewis, Elham Kashefi, and Mina Doosti. "Agnostic Process Tomography". In: *PRX Quantum* (2025). DOI: 10.1103/q2nb-zg9m.

[Wri24] John Wright. *Quantum Learning Theory*. Course website and scribe notes for UC Berkeley CS 294. 2024. URL: https://people.eecs.berkeley.edu/~jswright/quantumlearningtheory24/.

[YLC14] Theodore J. Yoder, Guang Hao Low, and Isaac L. Chuang. "Fixed-Point Quantum Search with an Optimal Number of Queries". In: *Phys. Rev. Lett.* (21 2014). DOI: 10.1103/PhysRevLett.113.210501.

[Yu20] Nengkun Yu. *Sample efficient tomography via Pauli Measurements*. 2020. arXiv: 2009.04610.