Agnostic Tomography of Stabilizer Product States

Sabee Grewal[1], Vishnu Iyer[1], William Kretschmer[1,2], and Daniel Liang[3,4]

[1]University of Texas at Austin, Austin, TX 78712, USA
[2]Simons Institute for the Theory of Computing, University of California, Berkeley, Berkeley, CA 94720, USA
[3]Portland State University, Portland, OR 97201, USA
[4]Rice University, Houston, TX 77005, USA

arXiv:2404.03813v5 [quant-ph] 4 Mar 2026

We define a quantum learning task called *agnostic tomography*, where given copies of an arbitrary state ρ and a class of quantum states $\mathcal{C}$, the goal is to output a succinct description of a state that approximates ρ at least as well as any state in $\mathcal{C}$ (up to some small error ε). This task generalizes ordinary quantum tomography of states in $\mathcal{C}$ and is more challenging because the learning algorithm must be robust to perturbations of ρ.

We give an efficient agnostic tomography algorithm for the class $\mathcal{C}$ of n-qubit stabilizer product states. Assuming ρ has fidelity at least τ with a stabilizer product state, the algorithm runs in time $n^{O(\log(2/\tau))}/\varepsilon^2$, which is $\mathsf{poly}(n/\varepsilon)$ for any constant τ.

1 Introduction

Exponentiality is at the heart of quantum mechanics. A system of merely 20 qubits already requires over a million parameters to describe. This exponentiality is both a boon and a curse. On the one hand, it plays a key role in quantum speedups, such as Shor's algorithm for factoring integers [1]. On the other hand, it seems we need an exponential number of amplitudes to describe systems of interest, preventing them from having tractable classical representations.

A common approach to handling the exponential complexity of physically and computationally relevant systems is to search for an *approximation* by a simple state (sometimes called an *ansatz*). For example, there is an extensive body of work dedicated to finding approximations by product state (i.e., unentangled) ansatzes, especially for groundstates of local Hamiltonians; see [2, 3] for an overview. Of course, product states are not the only possible ansatz: Hilbert space has many interesting classes of quantum states that are tractable—states that we can describe or classically simulate in polynomial time. Examples include matrix product states [4] and stabilizer states [5, 6], which have both seen use as groundstate ansatzes [7, 8].

The problem of approximating an arbitrary state by a simpler one is equally natural in the setting of quantum learning, where the goal is to recover a classical description of an unknown quantum state from examples. If we wish to learn an unknown state ρ from

Sabee Grewal: sabee@cs.utexas.edu, https://sabeegrewal.com/
Vishnu Iyer: vishnu.iyer@utexas.edu, https://vishnuiyer.org/
William Kretschmer: kretsch@cs.utexas.edu, https://wkretschmer.github.io/
Daniel Liang: daniel.liang@ll.mit.edu, https://daniel-you-liang.github.io/

nature, we cannot always assume that ρ has an exact succinct classical description, but we might have an intuition that it is well-approximated by one of the families described above. Alternatively, if the state or learning algorithm is corrupted by even a small amount noise, it might be impossible to learn ρ exactly, forcing us to settle for an approximation.

Nevertheless, the problem of efficiently learning an ansatz state approximation to an unknown ρ has not seen much rigorous study in quantum algorithms, despite the ubiquity of ansatz-based approaches in fields like Hamiltonian complexity. Of course, there are many efficient quantum algorithms for learning highly expressive classes of quantum states, including matrix product states [9], (near)-stabilizer states [10–13], free-fermion states [14], low-degree phase states [15], and outputs of constant-depth circuits [16]. But generally speaking, these algorithms are extremely brittle: correctness requires that ρ belong to the corresponding state class *exactly*, and they may fail even if ρ is affected by 1% noise.

On the other hand, there is a rich study of a similar task in classical learning theory, known as *agnostic learning*. In agnostic learning, the goal is to find a classifier for a dataset that performs close to optimally from some fixed class of models. The motivation for designing agnostic learning algorithms is much the same as for algorithms that approximate quantum states by simpler ones: real-world datasets rarely match abstract models exactly, and some fraction of the data could be corrupted adversarially by noise.

1.1 Agnostic Tomography

Inspired by classical agnostic learning, we define a learning task that captures the analogous problem for quantum states.

Definition 1.1 (Agnostic tomography)**.** *Let $\mathcal{C}$ be some class of pure quantum states.* Agnostic tomography *with respect to $\mathcal{C}$ is the following task: given copies of an arbitrary unknown mixed state ρ and $0 < \varepsilon < 1$, output a succinct classical description*[1] *of a state $|\phi\rangle$ such that*

$$\langle\phi|\rho|\phi\rangle \geq \sup_{|\varphi\rangle\in\mathcal{C}} \langle\varphi|\rho|\varphi\rangle - \varepsilon.$$

Proper agnostic tomography *is the same, but where we additionally require that $|\phi\rangle \in \mathcal{C}$.*

We suggest the name *agnostic tomography* because quantum tomography is the task of recovering a full classical description of an unknown quantum state [17]. Agnostic tomography captures tomography of states in $\mathcal{C}$ (to error ε in fidelity) as a special case: this corresponds to imposing the promise that the unknown state ρ is itself in $\mathcal{C}$.

Note that similar notions have appeared in prior work, albeit with different names. For example, O'Donnell and Wright [18] gave an exponential-time algorithm for learning the optimal rank-k approximation of a mixed state ρ. Grewal, Iyer, Kretschmer, and Liang [19, Theorem 1.3] introduced an exponential-time agnostic tomography algorithm for stabilizer states, under the condition that the input state is pure. They additionally gave a polynomial-time version of the algorithm under the promise that the optimal fidelity is at least $\cos^2(\pi/8)$. Our definition is also close to the definition of quantum Hypothesis Selection given by Bădescu and O'Donnell [20], except that we view the class $\mathcal{C}$ as fixed in advance, rather than a variable input to an algorithm. A survey by Anshu and Arunachalam [21] further pointed out the analogy between agnostic learning and quantum Hypothesis Selection.

[1] "Succinct classical description" is deliberately vague: the appropriate notion may depend on $\mathcal{C}$ or the application in mind. For most purposes, a reasonable choice is a short quantum circuit that prepares $|\phi\rangle$.

Informally, the goal of agnostic tomography is to find a simple ansatz from $\mathcal{C}$ for a much more complicated quantum state. The ansatz should attain a fidelity that is close to the optimal achievable within $\mathcal{C}$. Operationally, we envision experimentalists running agnostic tomography algorithms to learn an ansatz for their complicated physical system, and then working with the ansatz to extract additional meaningful information about the system. Agnostic tomography might also find theoretical applications: for example, [19] suggested that an agnostic tomography algorithm for stabilizer states could be useful as a subroutine for finding low-rank stabilizer decompositions of magic states, which could in turn improve the runtime of certain classical algorithms for simulating quantum circuits [22].

The choice to make $\mathcal{C}$ a class of pure states is not essential; this is only for simplicity. One could straightforwardly generalize our definition to include mixed states by using the mixed state fidelity $F(\rho,\sigma) = \left(\mathrm{tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)^2$ as the measure of correlation.[2] However, we have good reason to demand that algorithms operate on mixed input states ρ instead of only pure states, because one of the key applications of agnostic tomography is learning in the presence of noise, and noise will generally affect the purity of a state.

We emphasize that agnostic tomography is much harder than ordinary tomography, even for extremely simple classes of states $\mathcal{C}$. Intuitively, this is because tomography algorithms get to exploit additional structure from assuming that ρ belongs to $\mathcal{C}$, and this structure may be lost for states ρ that are only slightly outside of $\mathcal{C}$. By contrast, agnostic tomography algorithms must be robust to small perturbations: if ρ has 90% fidelity with some $|\phi\rangle \in \mathcal{C}$, the algorithm should detect this. It is also important to note that agnostic tomography algorithms cannot simply assume that the input state ρ is obtained from a state in $\mathcal{C}$ by some fixed noise channel (e.g., depolarizing noise), and thus agnostic tomography is distinct from a variety of noisy learning tasks considered in prior work [23–26].

As an illustrative example of the challenges in agnostic tomography, consider $\mathcal{C}$ to be the class of n-qubit product states, meaning the states that decompose into a tensor product $|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle$ of single-qubit states $|\psi_i\rangle$. There is a straightforward tomography algorithm to learn an unknown state $|\psi\rangle$ from $\mathcal{C}$: perform tomography separately on each qubit of $|\psi\rangle$. However, this algorithm fails catastrophically to be a good agnostic learner for $\mathcal{C}$. If we feed in the GHZ state $\rho = \left(\frac{|0^n\rangle+|1^n\rangle}{\sqrt{2}}\right)\left(\frac{\langle 0^n|+\langle 1^n|}{\sqrt{2}}\right)$, the local density matrix of each qubit looks maximally mixed, giving no information. Yet, the global state ρ has fidelity $1/2$ with either $|0^n\rangle, |1^n\rangle \in \mathcal{C}$. This further reveals that any agnostic learner for product states will have to go beyond looking at single-qubit reduced density matrices.[3]

Despite being a perfectly natural learning model, one reason agnostic tomography may not have gained much attention in the literature is that many quantum learning algorithms optimize sample complexity without worrying about computational efficiency [20, 28, 29]. However, agnostic tomography is often *easy* from the perspective of sample complexity alone. Concretely, Aaronson [28] showed that *shadow tomography* on an unknown state ρ lets one estimate $\langle\varphi|\rho|\varphi\rangle$ up to error ε, for every $|\varphi\rangle \in \mathcal{C}$, using a number of samples of ρ that scales polynomially in $\log|\mathcal{C}|$ and ε. So, if $\mathcal{C}$ is any class of n-qubit (pure) states that can be prepared with at most $\mathsf{poly}(n)$ gates (e.g., product states, stabilizer states, or matrix product states of polynomial bond dimension), we have $\log|\mathcal{C}| \leq \mathsf{poly}(n)$, and therefore $\mathcal{C}$ can be agnostically learned with only $\mathsf{poly}(n, 1/\varepsilon)$ samples. For this reason,

[2] Even more generally, one could consider other measures of correlation, such as 1 minus the trace distance. We chose fidelity because it is particularly easy to work with when $\mathcal{C}$ consists of pure states.

[3] A more general example shows that even when the product state fidelity is arbitrarily close to 1, single-qubit local density matrices are insufficient to find the best product state approximation; see [27, Section 2].

the main question of interest surrounding agnostic tomography is whether we can design algorithms that are efficient in both sample complexity *and* runtime.

1.2 Our Result

The main technical result of this work is, to our knowledge, the first efficient agnostic tomography algorithm for a nontrivial class of states. We say "nontrivial" because there are some uninteresting state classes that are very easy to learn agnostically. For example, let $\mathcal{C}$ be any orthonormal basis that can be measured efficiently, such as the computational basis. Then, by repeatedly measuring in the $\mathcal{C}$-basis and outputting the empirical mode, one can efficiently find the best approximation of ρ by a state in $\mathcal{C}$.

We show how to agnostically learn the set of *stabilizer product states*. These states are tensor products of the 6 single-qubit states $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |i\rangle, |-i\rangle\}$ that are ± 1 eigenstates of the single-qubit Pauli operators X, Y, Z. They are so-called because they are precisely the intersection between the set of stabilizer states and the set of product states. We learn this class agnostically in quasipolynomial time, resolving an open question raised by [19]:

Theorem 1.2. *There is a proper agnostic tomography algorithm for the class*

$$\mathcal{C} \coloneqq \{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |i\rangle, |-i\rangle\}^{\otimes n}$$

of n-qubit stabilizer product states. Specifically, given copies of an unknown state ρ and $0 < \varepsilon < 1$, the algorithm outputs a classical description of a stabilizer product state $|\phi\rangle$ such that

$$\langle\phi|\rho|\phi\rangle \geq \max_{|\varphi\rangle \in \mathcal{C}} \langle\varphi|\rho|\varphi\rangle - \varepsilon,$$

with constant probability. Assuming $\max_{|\varphi\rangle \in \mathcal{C}} \langle\varphi|\rho|\varphi\rangle \geq \tau$, the algorithm runs in time

$$\frac{n^{O(\log(2/\tau)}}{\varepsilon^2}.$$

The parameter τ is taken as an auxiliary input to the algorithm. If τ is at least a constant, then the runtime becomes polynomial in n and $1/\varepsilon$. If τ is not known in advance, we can run the algorithm repeatedly and verify the fidelity of its output to perform a binary search over τ.

Despite the apparent simplicity of stabilizer product states (compared to, say, general product states), we remark that the set of stabilizer product states is a useful class of states in its own right that appears in physically relevant scenarios. As an example, a recent work by Bakshi, Liu, Moitra, and Tang [30, Theorem 1.1] showed that sufficiently high-temperature Gibbs states of local Hamiltonians are expressible as probabilistic mixtures of stabilizer product states.

1.3 Main Ideas

To simplify presentation in the main body of this text, we will explain the algorithm and prove its correctness under the assumption that the input ρ is a pure state $|\psi\rangle$. We explain how to prove its correctness on mixed state inputs in Appendix A. This only requires changing a couple of lemmas that appear at the start of the technical sections, and does not materially affect the structure of the proof.

Our algorithm can be seen as a specialization of the algorithm due to Grewal, Iyer, Kretschmer, and Liang [19] that finds a stabilizer state whose fidelity with an input state is

ε-close to maximal. [19]'s algorithm is itself a generalization of Montanaro's algorithm [10] for learning stabilizer states exactly. The basic premise of both algorithms is to identify an n-qubit stabilizer state $|\phi\rangle$ by learning its *stabilizer group*—the abelian group of 2^n Pauli operators of which $|\phi\rangle$ is a $+1$ eigenstate. Both algorithms achieve this via a primitive known as *Bell difference sampling*, which performs a measurement on some copies of the unknown state and induces a distribution over Pauli operators.

In the case of Montanaro's algorithm [10], Bell difference sampling on a stabilizer state $|\psi\rangle$ gives the uniform distribution over Pauli operators in the stabilizer group, modulo a global phase that can be ignored. So, to learn $|\psi\rangle$, it suffices to take $O(n)$ Bell difference samples and compute the stabilizer group generated by them. Overall, this runs in polynomial time.

[19] show that if $|\psi\rangle$ has non-negligible fidelity with some stabilizer state $|\phi\rangle$, then a large fraction of the Pauli operators taken from Bell difference sampling will belong to the stabilizer group S of $|\phi\rangle$ (again, modulo global phase). Moreover, they prove that if $|\phi\rangle$ is the stabilizer state with the largest fidelity, the conditional distribution on S will be "close enough" to uniform over S, in the sense that the conditional distribution cannot be concentrated on any proper subgroup of S. Their algorithm for finding a stabilizer state $|\phi\rangle$ that approximates $|\psi\rangle$ is then the following: perform Bell difference sampling repeatedly on $|\psi\rangle$, iterate through all maximal subsets of the sampled Pauli operators that generate an abelian group S of size 2^n, and finally estimate the fidelity of $|\psi\rangle$ with every state $|\phi\rangle$ stabilized by S (modulo phase). This procedure takes exponential time in general, but in some parameter regimes it is more efficient than a brute-force search over all stabilizer states—for example, if the stabilizer fidelity is at least a constant.

In this paper, we take advantage of the additional structure of stabilizer *product* states. Whereas learning the stabilizer group of an arbitrary n-qubit stabilizer state required identifying n independent generators in both [10] and [19], the stabilizer group of a stabilizer product state can be identified up to phase from just a *single* generator: the unique Pauli stabilizer of weight n. In other words, we only need to assign one of X, Y, or Z to each qubit. Though we are very unlikely to sample the weight-n generator in a single iteration of Bell difference sampling, we show how to identify it using many fewer than n Bell difference samples.

As a concrete example, suppose that we obtain Bell difference samples $XIZII$, $IYZYI$, and $IIIYZ$ from a 5-qubit state $|\psi\rangle$. Imagine we are lucky enough that all of these Paulis belong to the stabilizer group S (up to phase) of the stabilizer product state $|\phi\rangle$ that maximizes fidelity with $|\psi\rangle$. The only full-weight Pauli operator consistent with these generators is $XYZYZ$, because any two Pauli operators in S must commute *locally*—no pair of corresponding single-qubit Paulis can anticommute (e.g., given that $XIZII$ is a stabilizer, there can be no other stabilizers where the first qubit is Y or Z, or where the third qubit is X or Y). The corresponding stabilizer product states consistent with $XYZYZ$ are precisely

$$\mathcal{B} = \{|+\rangle, |-\rangle\} \otimes \{|i\rangle, |-i\rangle\} \otimes \{|0\rangle, |1\rangle\} \otimes \{|i\rangle, |-i\rangle\} \otimes \{|0\rangle, |1\rangle\}.$$

The 2^n states in $\mathcal{B}$ form an orthonormal basis, so we can find the $|\phi\rangle \in \mathcal{B}$ that maximizes fidelity by repeatedly measuring in this basis and outputting the mode.

So, how many Bell difference samples do we need to learn the stabilizer group S corresponding to a stabilizer product state $|\phi\rangle$? As a first intuition, if the Bell difference samples are uniform over S, then the Pauli sampled at each qubit equals I with $1/2$ probability, or else equals some non-identity Pauli X, Y, or Z with $1/2$ probability. By the union bound, then, roughly $\log n$ samples suffice to see a non-identity Pauli on all n qubits, and hence

to learn S. This contrasts with the case of general stabilizer states, where [10, 19] needed to learn n linearly independent Pauli generators from S.

This intuition turns out to be very close to the truth: we show, by an entropy counting argument, that having $O(\log n)$ Bell difference samples from S suffices (with high probability) to see non-identity Paulis on all but at most $O(\log(1/\tau))$ of the qubits, where $\tau = |\langle\phi|\psi\rangle|^2$. The Pauli operators over the remaining qubits can be brute-forced in time $3^{O(\log(1/\tau))} = \mathsf{poly}(1/\tau)$.

The one remaining challenge to overcome is that not all of the Bell difference samples will belong to S, but only a subset. Because we have no way of knowing *a priori* which of the Bell difference samples belong to S, we simply check all locally commuting subsets of size $O(\log n)$. If p is the probability that a Bell difference sample from $|\psi\rangle$ belongs to S, we need $O(\log n/p)$ Bell difference samples for at least $\Omega(\log n)$ of them to come from S with high probability. By a lemma of [19], we argue that $p \geq \tau^4$, which gives an overall runtime bound of roughly

$$\binom{\log n/\tau^4}{\log n}\mathsf{poly}(1/\tau) \leq n^{O(\log(2/\tau))}.$$

One other noteworthy feature of the algorithm is a runtime dependence on a parameter b that appears in the entropy counting argument. The algorithm is correct for any choice of b strictly between $1/2$ and 1, and in the runtime analysis we can take b to be a constant (say, $2/3$). A closed form expression for the optimal choice of b in terms of the other parameters is unlikely to exist. For this reason, we expect that in a practical implementation of the algorithm, it would be best to optimize b numerically.

1.4 Follow-up Work

Since the initial posting of this work, the agnostic tomography model has motivated the exploration of agnostic learning algorithms for other classically-tractable quantum state families. For example, [31] gave efficient agnostic tomography algorithms for two classes that strictly generalize the set of stabilizer product states, namely: the class of arbitrary stabilizer states, and any class of "discrete" product states. [27] also independently obtained an efficient agnostic learner for discrete product states and designed several agnostic tomography algorithms for general product states. More recently, [32] gave an agnostic tomography algorithm for the class of product *mixed* states by establishing a formal connection to robust statistics. These works rely on techniques that differ substantially from ours. Taken together, these various algorithms demonstrate the richness of the agnostic tomography model and the potential to achieve efficiency for yet more expressive classes of quantum states.

Beyond these algorithms, agnostic tomography has inspired new developments in other contexts. [33] initiated the study of agnostic *process* tomography, which aims to recover a structured quantum channel that closely approximates an unknown one. More recent work by Briët and Castro-Silva [34] used the ideas behind agnostic tomography to give new algorithms for some purely classical problems, including a near-optimal quadratic generalization of the Goldreich–Levin algorithm [35]. As corollaries, they obtain a fast self-corrector for quadratic Reed–Muller codes and an algorithmic version of the Gowers Inverse Theorem. Using related ideas, [36] turned agnostic tomography procedures for stabilizer states into an algorithmic version of the Freiman–Ruzsa theorem in combinatorics. Most recently, [37] obtained faster quantum algorithms for PAC learning depth-3 AC^0 circuits by way of a boosting method for agnostic tomography algorithms.

2 Preliminaries

Throughout, $\log_b$ is the logarithm base b, and log is shorthand for the natural logarithm $\log_e$ where $e \approx 2.718$.

2.1 Probability

For a discrete random variable X with probability mass function p,

$$\mathsf{H}(X) \coloneqq \sum_x -p(x)\log_2(p(x))$$

is the binary entropy of X. For $p \in [0,1]$, we also denote by

$$\mathsf{H}(b) \coloneqq -p\log_2(p) - (1-p)\log_2(1-p)$$

the entropy of a Bernoulli random random variable with success probability p. The min-entropy of X is denoted by $\mathsf{H}_\infty(X) \coloneqq \log_2(1/p_{\max})$, where $p_{\max} = \max_x p(x)$.

We use both the binary and min-entropy in the lemma below. It says that if we take a logarithmic number of samples from a well-spread distribution over $\{0,1\}^n$, we are likely to see 1s in all but a logarithmic number of positions. The reader should note that the bound on $|M|$ is essentially tight: we cannot in general hope for anything bigger than $n - \log_2(C)$, because D might be supported on strings that start with $\log_2(C)$ zeros.

Lemma 2.1. *Let D be a probability distribution over $\{0,1\}^n$ with $D(x) \le \frac{C}{2^n}$ for all x. For some $1/2 < b < 1$, suppose we make $k \ge \log_{1/b}(n/\delta)$ independent draws from D and let $M \subseteq [n]$ be the set of positions in which a 1 is seen. Then with probability at least $1-\delta$, $|M| \ge n - \frac{\log_2(C)}{1-\mathsf{H}(b)}$.*

Proof. The entropy of D satisfies

$$\mathsf{H}(D) \ge \mathsf{H}_\infty(D) \ge -\log_2\left(\frac{C}{2^n}\right) = n - \log_2(C).$$

By the sub-additivity of entropy, we have:

$$\sum_{j=1}^{n} \mathsf{H}(D_j) \ge \mathsf{H}(D) \ge n - \log_2(C),$$

where D_j denotes the jth bit of D.

Let $J \coloneqq \{j \in [n] : \mathsf{H}(D_j) \ge \mathsf{H}(b)\}$. Since $\mathsf{H}(D_j) \le 1$ for all j, we know that $|J| \ge n - \frac{\log_2(C)}{1-\mathsf{H}(b)}$. We will now prove the claim by showing that the positions corresponding to indices in J are unlikely to all be 0. If we denote the k samples by $x_1, \ldots, x_k$, then:

$$\begin{aligned}
\mathbf{Pr}\left[|M| \le n - \frac{\log_2(C)}{1-\mathsf{H}(b)}\right] &\le \mathbf{Pr}[\exists j \in J, \forall i \in [k] : x_{i,j} = 0] \\
&\le |J| \max_{j \in J} \mathbf{Pr}[\forall i \in [k] : x_{i,j} = 0] && \text{(Union bound)} \\
&\le |J| b^k && (\mathbf{Pr}[x_{i,j} = 0] \le b) \\
&\le np^k \\
&\le \delta,
\end{aligned}$$

which completes the proof. □

Let p and q be distributions over $[N]$. The ℓ_∞ *distance* between p and q is defined by

$$\mathrm{dist}_{\ell_\infty}(p,q) \coloneqq \max_{i\in[N]} |p(i) - q(i)|.$$

The next lemma quantifies how many samples from a categorical distribution are needed to estimate its parameters. Note the independence of the number of categories.

Lemma 2.2 (Based on [38, Theorem 9])**.** *Let p be an arbitrary distribution over $[N]$, and let $\widehat{p}$ denote the empirical distribution derived from $\frac{2\log(2/\delta)}{\varepsilon^2}$ independent samples from p. Then*

$$\mathbf{Pr}\left[\mathrm{dist}_{\ell_\infty}(\widehat{p},p) > \varepsilon\right] \le \delta.$$

Proof sketch. Letting $m = \frac{2\log(2/\delta)}{\varepsilon^2}$, we have

$$\begin{aligned}\mathbf{Pr}\left[\mathrm{dist}_{\ell_\infty}(\widehat{p},p) > \varepsilon\right] &\le \mathbf{Pr}\left[\mathrm{dist}_K(\widehat{p},p) > \varepsilon/2\right] \\ &\le 2e^{-m\varepsilon^2/2} \\ &\le \delta,\end{aligned}$$

where dist_K is the *Kolmogorov distance*, and the second line is the DKW inequality [39]. □

2.2 Pauli Operators

We denote by $\mathcal{P}_n \coloneqq \{I,X,Y,Z\}^{\otimes n}$ the set of tensor products of the standard single-qubit Pauli operators I,X,Y,Z. The set $\overline{\mathcal{P}_n} \coloneqq \{\pm 1, \pm i\} \times \mathcal{P}_n$, the closure of $\mathcal{P}_n$ under multiplication by ± 1 and $\pm i$ phases, forms a group under multiplication called the *Pauli group*.

$\mathcal{P}_n$ forms an orthonormal basis for the Hermitian matrices in $\mathbb{C}^{2^n\times 2^n}$ under the normalized Hilbert–Schmidt inner product: $\langle A,B\rangle = 2^{-n}\mathrm{tr}(A^\dagger B)$. As such, for any n-qubit state $|\psi\rangle$ we can write

$$|\psi\rangle\langle\psi| = \frac{1}{2^n}\sum_{P\in\mathcal{P}_n} \langle\psi|P|\psi\rangle\, P$$

Orthonormality of $\mathcal{P}_n$ implies that $p_\psi(P) \coloneqq \frac{1}{2^n}\langle\psi|P|\psi\rangle^2$ forms a distribution over $\mathcal{P}_n$. Gross, Nezami, and Walter showed that one can sample from the distribution[4]

$$q_\psi(P) \coloneqq \sum_{Q\in\mathcal{P}_n} p_\psi(Q)p_\psi(PQ) \tag{1}$$

in $O(n)$ time using 4 copies of $|\psi\rangle$, via a procedure known as *Bell difference sampling* [40].[5] q_ψ is a distribution because it equals the convolution $p_\psi * p_\psi$. Bell difference sampling will be a crucial primitive in our algorithm.

[4] The multiplication PQ is understood to be modulo global phase.

[5] [40] uses slightly different notation, identifying $\mathcal{P}_n$ with a $2n$-dimensional symplectic vector space over $\mathbb{F}_2$. We do not use this notation here because the group structure of Pauli operators turns out to be unimportant for us, though we will introduce this notation in Appendix A.

2.3 Stabilizers

Let $|\psi\rangle$ be an n-qubit pure state. We define the *unsigned stabilizer group* of $|\psi\rangle$ to be

$$\mathrm{Stab}_{\pm}(|\psi\rangle) \coloneqq \{P \in \mathcal{P}_n : P|\psi\rangle = \pm|\psi\rangle\}.$$

A *stabilizer state* is a state $|\psi\rangle$ satisfying $|\mathrm{Stab}_{\pm}(|\psi\rangle)| = 2^n$. A *stabilizer product state* is a stabilizer state that is also a tensor product of single-qubit states. The set of n-qubit stabilizer product states is precisely

$$\mathcal{SP} \coloneqq \{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |i\rangle, |-i\rangle\}^{\otimes n}.$$

Observe that the single-qubit stabilizer states are the eigenstates of single-qubit Pauli operators:

$$X|+\rangle = |+\rangle, \quad X|-\rangle = -|-\rangle, \quad Y|i\rangle = |i\rangle, \quad Y|-i\rangle = -|-i\rangle, \quad Z|0\rangle = |0\rangle, \quad Z|1\rangle = -|1\rangle.$$

The *stabilizer product fidelity* of a state $|\psi\rangle$ is the largest fidelity of $|\psi\rangle$ with any stabilizer product state, and is denoted by

$$\mathcal{F}_{\mathcal{SP}}(|\psi\rangle) \coloneqq \max_{|\phi\rangle \in \mathcal{SP}} |\langle\phi|\psi\rangle|^2.$$

We say that an unsigned stabilizer group S is a *stabilizer product group* if S has the form

$$S = \{I, P_1\} \otimes \{I, P_2\} \otimes \cdots \otimes \{I, P_n\}$$

for some single-qubit non-identity Pauli operators $P_1, \ldots, P_n \in \{X, Y, Z\}$. Note that for any stabilizer product state $|\psi\rangle$, $\mathrm{Stab}_{\pm}(|\psi\rangle)$ is a stabilizer product group. Conversely, every stabilizer product group S is the unsigned stabilizer group of exactly 2^n different stabilizer product states $|\psi\rangle$. These states form a basis called the *S-basis*. The S-basis can be measured efficiently given generators for S. For example, if

$$S = \{I, X\} \otimes \{I, Y\} \otimes \{I, Z\},$$

the corresponding basis of stabilizer product states stabilized by S is

$$\{|+\rangle, |-\rangle\} \otimes \{|i\rangle, |-i\rangle\} \otimes \{|0\rangle, |1\rangle\}.$$

In other words, measuring in the S-basis means measuring each qubit according to the non-identity single-qubit Paulis assigned to each qubit by S.

2.4 Additional Notation

Let $P = P_1 \otimes \cdots \otimes P_n$ and $Q = Q_1 \otimes \ldots \otimes Q_n$ be n-qubit Pauli operators. We say P and Q *locally* commute if for all $i \in [n]$, $P_i Q_i = Q_i P_i$. Clearly, if P and Q both belong to a stabilizer product group, then they locally commute.

For $P, Q \in \mathcal{P}_n$, define

$$[[P, Q]] = |\{i \in [n] : P_i Q_i \neq Q_i P_i\}|.$$

Then P and Q locally commute if and only if $[[P, Q]] = 0$.

Suppose a set $S \subset \mathcal{P}_n$ satisfies $[[P, Q]] = 0$ for all $P, Q \in S$. Let T be the set of indices $i \in [n]$ such that for at least one $P \in S$, $P_i \neq I$. Then for every $i \in T$, there exists a

unique $W_i \in \{X, Y, Z\}$ such that for all $x \in S$, $P_i \in \{I, W_i\}$. For $i \notin T$, define $W_i := I$. We define the *local span* of S to be

$$\langle\langle S \rangle\rangle := \{I, W_1\} \otimes \{I, W_2\} \otimes \cdots \otimes \{I, W_n\}.$$

Note that $\langle\langle S \rangle\rangle$ is a group of size $2^{|T|}$, and every $P, Q \in \langle\langle S \rangle\rangle$ satisfies $[[P, Q]] = 0$. Additionally, there are exactly $3^{n-|T|}$ stabilizer product groups $\overline{S}$ that extend $\langle\langle S \rangle\rangle$ (i.e., for which $S \subseteq \overline{S}$): these are obtained by assigning one of $\{X, Y, Z\}$ to each index $i \notin T$. An intuitive definition of the local span is that it is the smallest tensor product of single-qubit commuting groups that contains every Pauli operator in S.

3 Algorithm and Analysis

We are now ready to state the algorithm. In the subsections that follow, we establish its correctness and bound its runtime. As noted in the introduction, the input parameter b is arbitrary, but it affects the runtime; see Section 3.2 for further discussion.

Algorithm 1: Stabilizer Product State Approximation

Input: Copies of $|\psi\rangle$, $0 < \varepsilon \leq \tau < 1$, $1/2 < b < 1$
Promise: $|\psi\rangle$ has fidelity at least τ with some stabilizer product state
Output: A stabilizer product state $|\phi\rangle$ such that $|\langle\phi|\psi\rangle|^2 \geq \mathcal{F}_{\mathcal{SP}}(|\psi\rangle) - \varepsilon$ with probability at least $1/8$

$k := \log_{1/b}(2n)$
$m_{\mathsf{clique}} := \frac{1}{\tau^4}(k+1)$
$t := \frac{4\log_2(1/\tau)}{1-\mathsf{H}(b)}$
$m_{\mathsf{est}} := \frac{8\log\left(16\binom{m_{\mathsf{clique}}}{k}3^t\right)}{\varepsilon^2}$
Initialize an empty graph G
repeat m_{clique} **times**
    Using 4 copies of $|\psi\rangle$, perform Bell difference sampling to obtain $W \in \mathcal{P}_n$
    Add a vertex for W in G and an edge to all vertices W' in G such that $[[W, W']] = 0$
foreach clique $(V_1, \ldots, V_k) \in G$ of size k **do**
    Compute $S := \langle\langle V_1, \ldots, V_k \rangle\rangle$
    if $\log_2|S| \geq n - t$ **then**
        foreach stabilizer product group $\overline{S}$ extending S **do**
            Measure m_{est} copies of $|\psi\rangle$ in the $\overline{S}$-basis
            foreach $|\phi\rangle$ in the $\overline{S}$-basis **do**
                Compute the empirical probability $\widehat{o}_\phi$ with which $|\phi\rangle$ was measured
return whichever $|\phi\rangle$ maximizes $\widehat{o}_\phi$

Note that the success probability in Algorithm 1 can be amplified from $1/8$ to $1 - \delta$ in a black-box fashion, by repeating the algorithm $O(\log(1/\delta))$ times and checking which outputs $|\phi\rangle$ have largest fidelity with $|\psi\rangle$; we omit the details.

3.1 Correctness

We begin with two basic lemmas about the Bell difference sampling distribution. These will turn out to be the *only* two lemmas that must be generalized in Appendix A in order to

prove Algorithm 1's correctness on mixed state inputs. The first shows that Bell difference sampling cannot place too much probability on any one Pauli operator.

Lemma 3.1. *For all n-qubit quantum states $|\psi\rangle$ and $P \in \mathcal{P}_n$,*

$$q_\psi(P) \leq \frac{1}{2^n}.$$

Proof. This is a simple consequence of q_ψ being a convolution of p_ψ with itself:

$$\begin{aligned} q_\psi(P) &= \sum_{Q \in \mathcal{P}_n} p_\psi(Q) p_\psi(PQ) \\ &\leq \frac{1}{2^n} \sum_{Q \in \mathcal{P}_n} p_\psi(Q) && (\forall P \in \mathcal{P}_n,\, p_\psi(P) \leq \frac{1}{2^n}) \\ &= \frac{1}{2^n}. \end{aligned}$$

□

This next lemma shows that if a state $|\psi\rangle$ has large fidelity with a stabilizer state $|\phi\rangle$, then Bell difference samples have a good probability of belonging to the unsigned stabilizer group of $|\phi\rangle$. A conceptually similar statement was already proved in [19, Lemma 5.6].

Lemma 3.2. *Let $|\psi\rangle$ have fidelity at least τ with stabilizer state $|\phi\rangle$ and let $S = \mathrm{Stab}_\pm(|\phi\rangle)$. Then*

$$\sum_{P \in S} q_\psi(P) \geq \tau^4.$$

Proof.

$$\begin{aligned} \sum_{P \in S} q_\psi(P) &= \sum_{P \in S} \sum_{Q \in \mathcal{P}_n} p_\psi(Q) p_\psi(PQ) \\ &\geq \sum_{P \in S} \sum_{Q \in S} p_\psi(Q) p_\psi(PQ) \\ &= \left(\sum_{P \in S} p_\psi(P) \right)^2 \\ &\geq \tau^4, && ([19, \text{Lemma 5.2}]) \end{aligned}$$

where we have used the fact that S is a subgroup of $\mathcal{P}_n$, modulo phase. □

We now prove the lemma that is most central to our algorithm's correctness. Below, think of b as some constant (say, 2/3). The lemma shows that if we take $O(\log n)$ Bell difference samples, then conditioned on all belonging to the unsigned stabilizer group S of a stabilizer product state of fidelity τ, the local span of the samples is likely to capture all but $O(\log(1/\tau))$ generators of S.

Lemma 3.3. *Let $|\psi\rangle$ have fidelity at least τ with stabilizer product state $|\phi\rangle$. Let $S = \mathrm{Stab}_\pm(|\phi\rangle)$. Let $V_1, \ldots, V_k \sim q_\psi \mid S$, meaning that each V_i is sampled from q_ψ conditioned on being in S. Fix $1/2 < b < 1$. If $k \geq \log_{1/b}(n/\delta)$, then with probability at least $1 - \delta$, $\log_2 |\langle\langle V_1, \ldots, V_k \rangle\rangle| \geq n - \frac{4 \log_2(1/\tau)}{1 - \mathsf{H}(b)}$.*

Proof. Let $\mathcal{Y}$ denote a draw from $q_\psi \mid S$. Observe that for any $P \in S$, we have

$$\begin{aligned}\Pr_{V\sim\mathcal{Y}}[V = P] &= \frac{q_\psi(P)}{\sum_{P'\in S} q_\psi(P')} \\ &\leq \frac{1}{2^n \sum_{P'\in S} q_\psi(P')} && \text{(Lemma 3.1)} \\ &\leq \frac{1}{\tau^4 2^n} && \text{(Lemma 3.2).}\end{aligned}$$

Now, let D be the distribution over $\{0,1\}^n$ obtained by sampling $V \sim \mathcal{Y}$ and putting 1s at the positions where V has a non-identity Pauli. Then for all $x \in \{0,1\}^n$, $D(x) \leq \frac{1}{\tau^4 2^n}$ because $\mathcal{Y}$ is supported over the stabilizer product group S, and the mapping $\mathcal{Y}$ to D is a bijection between S and $\{0,1\}^n$. By Lemma 2.1, with probability at least $1-\delta$, k independent draws from D will see 1s in at least $n - \frac{4\log_2(1/\tau)}{1-\mathsf{H}(b)}$ locations. Therefore, the local span $\langle\langle V_1, \ldots, V_k\rangle\rangle$ of the corresponding samples from $\mathcal{Y}$ will generate a group of order at least $2^{n-\frac{4\log_2(1/\tau)}{1-\mathsf{H}(b)}}$. □

Lemma 3.2 guarantees that at least τ^4 of the q_ψ mass is on the stabilizer group S^* of the fidelity-maximizing stabilizer product state. So, it suffices to take $m_{\mathsf{clique}} = O(\log n/\tau^4)$ Bell difference samples to have a good chance that some subset of size $\Omega(\log n)$ all belong to S. Hence, we can appeal to Lemma 3.3 to argue that the algorithm has a decent chance of finding S^* in one of its iterations.

Lemma 3.4. *Let $|\varphi\rangle$ be the stabilizer product state that has the largest fidelity with $|\psi\rangle$, and let $S^* = \mathrm{Stab}_\pm(|\varphi\rangle)$. Then with probability at least $\frac{1}{4}$, $|\varphi\rangle$ appears as one of the states $|\phi\rangle$ in Line 14.*

Proof. Let $W_1, \ldots, W_{m_{\mathsf{clique}}}$ be the Bell difference samples obtained in Line 6. Applying Lemma 3.2, each W_i has probability at least τ^4 of belonging to S^*. Hence, with probability at least $1/2$, there exists a subset $\{V_1, \ldots, V_k\} \subset \{W_1, \ldots, W_{m_{\mathsf{clique}}}\}$ such that $V_i \in S^*$ for every $i \in [k]$, because the median of a binomial distribution with $N = m_{\mathsf{clique}}$ trials and rate $P \geq \tau^4$ is at least $\lfloor NP \rfloor \geq k$.

Lemma 3.3 additionally tells us that conditioned on $V_i \in S^*$ for every i, we have $\log_2|\langle\langle V_1, \ldots, V_k\rangle\rangle| \geq n - t$ with probability at least $1/2$. So, with overall probability at least $1/4$, there exists a subset $\{V_1, \ldots, V_k\} \subset \{W_1, \ldots, W_{m_{\mathsf{clique}}}\}$ such that $V_i \in S^*$ for every i *and* $\log_2|\langle\langle V_1, \ldots, V_k\rangle\rangle| \geq n - t$.

Assume henceforth that this has occurred. Because $V_i \in S^*$ for every i, and because S^* is a stabilizer product group, $[[V_i, V_j]] = 0$ for each i, j. Therefore, $(V_i, \ldots, V_k)$ will be one of the cliques in the graph found in Line 9, and the subspace $S = \langle\langle V_1, \ldots, V_k\rangle\rangle$ will pass the check at Line 11. For some choice of $\overline{S}$ in Line 12 we will have $\overline{S} = S^*$, so one of the states $|\phi\rangle$ in the $\overline{S}$-basis will equal $|\varphi\rangle$. □

We can now complete the proof of Algorithm 1's correctness.

Theorem 3.5. *Algorithm 1 finds a stabilizer product state $|\phi\rangle$ such that $|\langle\phi|\psi\rangle|^2 \geq \mathcal{F}_{\mathcal{SP}}(|\psi\rangle) - \varepsilon$ with probability at least $1/8$.*

Proof. Lemma 2.2 implies that in any fixed iteration of Line 12, all of the estimates $\widehat{o}_\phi$ satisfy

$$|\widehat{o}_\phi - |\langle\phi|\psi\rangle|^2| \leq \varepsilon/2,$$

except with failure probability at most

$$\frac{1}{8\binom{m_{\mathsf{clique}}}{k}3^t}.$$

The loop in Line 12 is executed at most $\binom{m_{\mathsf{clique}}}{k}3^t$ times, because:

- Line 9 runs at most $\binom{m_{\mathsf{clique}}}{k}$ iterations;
- Line 12 runs at most 3^t iterations per iteration of Line 9, because there are at most 3^t stabilizer product groups $\overline{S}$ that extend S with size 2^{n-t} (this corresponds to assigning one of Pauli X, Y, or Z to each unassigned qubit).

Hence, by the union bound, *all* of the estimates $\widehat{o}_\phi$ satisfy

$$|\widehat{o}_\phi - |\langle\phi|\psi\rangle|^2| \leq \varepsilon/2, \tag{2}$$

except with failure probability at most $1/8$.

Let $|\varphi\rangle$ be the stabilizer product state that has the largest fidelity with $|\psi\rangle$, meaning $|\langle\varphi|\psi\rangle|^2 = \mathcal{F}_{\mathcal{SP}}(|\psi\rangle)$. By Lemma 3.4, $|\varphi\rangle$ appears as at least one of the states $|\phi\rangle$ in Line 14, except with failure probability at most $3/4$.

Assume that both of the above steps succeed, which occurs with probability at least $1/8$ by the union bound. Letting $|\phi\rangle$ be the state that maximizes $\widehat{o}_\phi$, we have

$$\begin{aligned}
|\langle\phi|\psi\rangle|^2 &\geq \widehat{o}_\phi - \varepsilon/2 && \text{(Equation (2))}\\
&\geq \widehat{o}_\varphi - \varepsilon/2 && (|\phi\rangle \text{ maximizes } \widehat{o}_\phi)\\
&\geq |\langle\varphi|\psi\rangle|^2 - \varepsilon && \text{(Equation (2))}\\
&= \mathcal{F}_{\mathcal{SP}}(|\psi\rangle) - \varepsilon,
\end{aligned}$$

which shows that the algorithm is correct. □

3.2 Time Complexity

The runtime of Algorithm 1 is easily seen to be dominated by computing the estimates $\widehat{o}_\phi$ in Line 14. Of course, Line 14 should not literally be implemented by iterating through all 2^n states $|\phi\rangle$ in the $\overline{S}$-basis, because in general $m_{\mathsf{est}} \ll 2^n$—it is more efficient to maintain a dictionary of the states $|\phi\rangle$ that were measured, along with their measurement counts. If implemented in this fashion, then the overall runtime of the algorithm is bounded by

$$\binom{m_{\mathsf{clique}}}{k}3^t m_{\mathsf{est}} \cdot O(n)$$

because Line 9 runs at most $\binom{m_{\mathsf{clique}}}{k}$ iterations, Line 12 runs at most 3^t iterations (as argued in the proof of Theorem 3.5 above), Line 14 runs m_{est} iterations, and each iteration of Line 14 takes $O(n)$ time to perform the measurement in a stabilizer product basis. Plugging in m_{est}, the runtime becomes

$$\binom{m_{\mathsf{clique}}}{k}3^t \log\left(\binom{m_{\mathsf{clique}}}{k}3^t\right) \cdot O(n/\varepsilon^2). \tag{3}$$

If we take b to be some constant such as $2/3$, then $3^t \leq \mathsf{poly}(1/\tau)$ and

$$\binom{m_{\mathsf{clique}}}{k} \leq \binom{O(\log n/\tau^4)}{O(\log n)} \leq \mathsf{poly}(1/\tau)^{O(\log n)} \leq n^{O(\log(2/\tau))}.$$

Therefore, the total runtime of the algorithm is at most

$$\frac{n^{O(\log(2/\tau))}}{\varepsilon^2}.$$

Note that the factor 2 in the $O(\log(2/\tau))$ is an arbitrary constant greater than 1, and is only necessary when τ is sufficiently close to 1. If we assume τ at most some constant bounded away from 1, it becomes correct to write $O(\log(1/\tau)$ in the exponent.

A bound on the explicit constant in the exponent requires more careful accounting. Consider

$$f(n,\tau,b) \coloneqq \binom{m_{\mathsf{clique}}}{k} 3^t,$$

so that the runtime becomes quasilinear in f, i.e.,

$$f(n,\tau,b)\log(f(n,\tau,b)) \cdot O(n/\varepsilon^2).$$

First we bound the term corresponding to Line 9:

$$\begin{aligned}
\binom{m_{\mathsf{clique}}}{k} &= \binom{\frac{1}{\tau^4}(\log_{1/b}(2n)+1)}{\log_{1/b}(2n)} \\
&\le \left(\left(1+\frac{1}{\log_{1/b}(2n)}\right)^{\log_{1/b}(2n)} \binom{\frac{1}{\tau^4}\log_{1/b}(2n)}{\log_{1/b}(2n)}\right) && \binom{a+b}{k} \le \left(\frac{a+b}{a}\right)^k \binom{a}{k} \\
&\le e\binom{\frac{1}{\tau^4}\log_{1/b}(2n)}{\log_{1/b}(2n)} && (1+1/x)^x \le e \\
&\le e(e\tau^{-4})^{\log_{1/b}(2n)} && \binom{a}{k} \le \left(\frac{ae}{k}\right)^k \\
&= e(2n)^{\log_{1/b}(e/\tau^4)} \\
&\le e(2n)^{4\log_{1/b}(e/\tau)}.
\end{aligned}$$

Then

$$f(n,\tau,b) \le e(2n)^{4\log_{1/b}(e/\tau)} 3^{\frac{4\log_2(1/\tau)}{1-\mathrm{H}(b)}}.$$

Optimizing the choice of b for general values of n and τ seems difficult, but the bound does not suggest any clear benefit to choosing b particularly close to $1/2$ or 1 instead of some fixed constant in between.

4 Discussion and Open Problems

Can the runtime of Algorithm 1 be improved to polynomial in n and $1/\tau$, rather than quasipolynomial? At minimum, it might be possible to further optimize the leading constant in the exponent of the runtime. We also suspect that the algorithm could be made more efficient in practice. For example, rather than fixing m_{clique} in advance, one could take Bell difference samples in an online fashion until the graph G contains a sufficiently large clique. This probably occurs faster in practice than our worst-case bounds predict.

Are there other classes of quantum states with simple classical descriptions that admit efficient agnostic tomography algorithms? Other examples to consider might include outputs of low-depth circuits and free-fermionic states.

Acknowledgments

SG is supported (via Scott Aaronson) by a Vannevar Bush Fellowship from the US Department of Defense, the Berkeley NSF-QLCI CIQC Center, a Simons Investigator Award, and the Simons "It from Qubit" collaboration. VI is supported by an NSF Graduate Research Fellowship. WK acknowledges support from the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator. DL is supported by NSF award FET-2243659.

This work was done in part while SG, VI, and DL were visiting the Simons Institute for the Theory of Computing, supported by NSF QLCI Grant No. 2016245.

References

[1] Peter W. Shor. "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer". SIAM Journal on Computing **26**, 1484–1509 (1997).

[2] Fernando G. S. L. Brandão and Aram W. Harrow. "Product-state approximations to quantum states". Communications in Mathematical Physics **342**, 47–80 (2016).

[3] John Kallaugher, Ojas Parekh, Kevin Thompson, Yipu Wang, and Justin Yirka. "Complexity Classification of Product State Problems for Local Hamiltonians". In Raghu Meka, editor, 16th Innovations in Theoretical Computer Science Conference (ITCS 2025). Volume 325 of Leibniz International Proceedings in Informatics (LIPIcs), pages 63:1–63:32. Dagstuhl, Germany (2025). Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[4] David Perez-Garcia, Frank Verstraete, Michael M. Wolf, and J. Ignacio Cirac. "Matrix product state representations". Quantum Info. Comput. **7**, 401–430 (2007).

[5] Daniel Gottesman. "The Heisenberg Representation of Quantum Computers" (1998). arXiv:quant-ph/9807006.

[6] Scott Aaronson and Daniel Gottesman. "Improved Simulation of Stabilizer Circuits". Physical Review A**70** (2004).

[7] Frank Verstraete, Valentin Murg, and J. Ignacio Cirac. "Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems". Advances in Physics **57**, 143–224 (2008).

[8] Jiace Sun, Lixue Cheng, and Shi-Xin Zhang. "Stabilizer ground states for simulating quantum many-body physics: theory, algorithms, and applications". Quantum **9**, 1782 (2025).

[9] Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. "Efficient quantum state tomography". Nature Communications **1**, 1–7 (2010).

[10] Ashley Montanaro. "Learning stabilizer states by Bell sampling" (2017). arXiv:1707.04012.

[11] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. "Efficient Learning of Quantum States Prepared With Few Non-Clifford Gates". Quantum **9**, 1907 (2025).

[12] Lorenzo Leone, Salvatore F. E. Oliviero, and Alioscia Hamma. "Learning t-doped stabilizer states". Quantum **8**, 1361 (2024).

[13] Dominik Hangleiter and Michael J. Gullans. "Bell sampling from quantum circuits". Phys. Rev. Lett. **133**, 020601 (2024).

[14] 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). Volume 266 of Leibniz International Proceedings in Informatics (LIPIcs), pages 12:1–12:18. (2023).

[15] 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). Volume 266 of Leibniz International Proceedings in Informatics (LIPIcs), pages 3:1–3:24. (2023).

[16] 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. Page 1343–1351. STOC 2024New York, NY, USA (2024). Association for Computing Machinery.

[17] Konrad Banaszek, Marcus Cramer, and David Gross. "Focus on quantum tomography". New Journal of Physics **15**, 125020 (2013).

[18] Ryan O'Donnell and John Wright. "Efficient quantum tomography II". In Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing. Page 962–974. STOC 2017New York, NY, USA (2017). Association for Computing Machinery.

[19] 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. Page 1352–1363. STOC 2024New York, NY, USA (2024). Association for Computing Machinery.

[20] Costin Bădescu and Ryan O'Donnell. "Improved Quantum Data Analysis". In Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing. Page 1398–1411. STOC 2021. Association for Computing Machinery (2021).

[21] Anurag Anshu and Srinivasan Arunachalam. "A survey on the complexity of learning quantum states". Nature Reviews Physics **6**, 59–69 (2024).

[22] Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. "Simulation of quantum circuits by low-rank stabilizer decompositions". Quantum **3**, 181 (2019).

[23] Yingkai Ouyang and Marco Tomamichel. "Learning quantum graph states with product measurements". In 2022 IEEE International Symposium on Information Theory (ISIT). Pages 2963–2968. (2022).

[24] Aravind Gollakota and Daniel Liang. "On the Hardness of PAC-learning Stabilizer States with Noise". Quantum **6**, 640 (2022).

[25] Alexander Poremba, Yihui Quek, and Peter Shor. "The Learning Stabilizers with Noise Problem". In Shubhangi Saraf, editor, 17th Innovations in Theoretical Computer Science Conference (ITCS 2026). Volume 362 of Leibniz International Proceedings in Informatics (LIPIcs), pages 108:1–108:19. Dagstuhl, Germany (2026). Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[26] Andrey Boris Khesin, Jonathan Z. Lu, Alexander Poremba, Akshar Ramkumar, and Vinod Vaikuntanathan. "Average-Case Complexity of Quantum Stabilizer Decoding" (2025). arXiv:2509.20697.

[27] 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. Page 1212–1221. STOC '25New York, NY, USA (2025). Association for Computing Machinery.

[28] Scott Aaronson. "Shadow Tomography of Quantum States". In Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing. Page 325–338. STOC 2018. Association for Computing Machinery (2018).

[29] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements". Nature Physics **16**, 1050–1057 (2020).
[30] Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. "High-temperature Gibbs states are unentangled and efficiently preparable". In 2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS). Pages 1027–1036. (2024).
[31] 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. Page 429–438. STOC '25New York, NY, USA (2025). Association for Computing Machinery.
[32] Alvan Arulandu, Ilias Diakonikolas, Daniel Kane, and Jerry Li. "Agnostic Product Mixed State Tomography via Robust Statistics" (2025). arXiv:2510.08472.
[33] Chirag Wadhwa, Laura Lewis, Elham Kashefi, and Mina Doosti. "Agnostic process tomography". PRX Quantum **6**, 040371 (2025).
[34] Jop Briët and Davi Castro-Silva. "A near-optimal quadratic Goldreich-Levin algorithm" (2025). arXiv:2505.13134.
[35] Oded Goldreich and Leonid A. Levin. "A hard-core predicate for all one-way functions". In Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing. Page 25–32. STOC '89New York, NY, USA (1989). Association for Computing Machinery.
[36] Srinivasan Arunachalam, Davi Castro-Silva, Arkopal Dutt, and Tom Gur. "Algorithmic polynomial Freiman-Ruzsa theorems" (2025). arXiv:2509.02338.
[37] Srinivasan Arunachalam, Arkopal Dutt, Alexandru Gheorghiu, and Michael de Oliveira. "Learning depth-3 circuits via quantum agnostic boosting" (2025). arXiv:2509.14461.
[38] Clément L. Canonne. "A short note on learning discrete distributions" (2020). arXiv:2002.11457.
[39] Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator". The Annals of Mathematical Statistics **27**, 642–669 (1956).
[40] 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". Communications in Mathematical Physics **385**, 1325–1393 (2021).
[41] Koenraad M R Audenaert and Martin B Plenio. "Entanglement on mixed stabilizer states: normal forms and reduction procedures". New Journal of Physics **7**, 170 (2005).

A Agnostic Tomography of Mixed States

In this appendix, we show that Algorithm 1 remains correct when run on mixed state inputs. To do so, we generalize some properties of Bell difference sampling (specifically, Lemmas 3.1 and 3.2) to mixed states. This will require different notation for working with Pauli operators, which we now introduce.

A.1 Weyl Operators

We begin by identifying the unsigned Pauli operators $\mathcal{P}_n$ with a vector space $\mathbb{F}_2^{2n}$ via the formalism of Weyl operators, following Gross, Nezami, and Walter [40]. For $x = (a, b) \in \mathbb{F}_2^{2n}$, the *Weyl operator* W_x is

$$W_x \coloneqq i^{a' \cdot b'} (X^{a_1} Z^{b_1}) \otimes \cdots \otimes (X^{a_n} Z^{b_n}),$$

where $a', b' \in \mathbb{Z}^n$ are the embeddings of a, b into $\mathbb{Z}^n$. Each W_x is an element of $\mathcal{P}_n$ and vice-versa, so this yields a one-to-one correspondence between $\mathbb{F}_2^{2n}$ and $\mathcal{P}_n$.

For $x, y \in \mathbb{F}_2^{2n}$, we call $[x, y]$ the *symplectic product* of x and y over $\mathbb{F}_2^{2n}$, which is defined by

$$[x, y] = x_1 \cdot y_{n+1} + x_2 \cdot y_{n+2} + ... + x_n \cdot y_{2n} + x_{n+1} \cdot y_1 + x_{n+2} \cdot y_2 + ... + x_{2n} \cdot y_n,$$

Operationally, the symplectic product captures commutation relations: W_x and W_y commute if and only if $[x, y] = 0$.

A.2 Bell Difference Sampling Mixed States

Performing Bell difference sampling on a mixed state ρ induces a distribution q_ρ over $\mathcal{P}_n$ (or equivalently, $\mathbb{F}_2^{2n}$), but it is a bit harder to define than for pure states. [40, Equation 3.7] shows that Bell difference sampling can be written as a projective measurement $\{\Pi_x\}_{x \in \mathbb{F}_2^{2n}}$, where

$$\Pi_x \coloneqq \frac{1}{4^n} \sum_{a \in \mathbb{F}_2^{2n}} (-1)^{[x,a]} W_a^{\otimes 4}.$$

Thus, we can write $q_\rho(x) \coloneqq \mathrm{tr}(\Pi_x \rho^{\otimes 4})$ as the probability that Bell difference sampling obtains $a \in \mathbb{F}_2^{2n}$. We can also define $p_\rho(x) \coloneqq \mathrm{tr}(W_x \rho)^2 / 2^n$ in analogy with p_ψ. We note that for mixed states, unlike for pure states, p_ρ is no longer a distribution (see Lemma A.1), and q_ρ can no longer be expressed as the convolution of p_ρ with itself, as we did in Equation (1). Nonetheless, we shall see that p_ρ and q_ρ are related in useful ways.

A.3 Properties of Bell Difference Sampling

We first show a connection between the purity of a state ρ and p_ρ

Lemma A.1. *Let ρ be a quantum mixed state. Then*

$$\mathrm{tr}(\rho^2) = \sum_{x \in \mathbb{F}_2^{2n}} p_\rho(x).$$

Proof. Because the Weyl operators form an orthogonal basis for the space of Hermitian n-qubit operators, we can express ρ in this basis as:

$$\rho = \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^{2n}} \mathrm{tr}(W_x \rho) W_x.$$

Consequently,

$$\begin{aligned}
\mathrm{tr}(\rho^2) &= \frac{1}{4^n} \mathrm{tr}\left(\sum_{x,y \in \mathbb{F}_2^{2n}} \mathrm{tr}(W_x \rho) \mathrm{tr}(W_y \rho) W_x W_y \right) \\
&= \frac{1}{4^n} \sum_{x,y \in \mathbb{F}_2^{2n}} \mathrm{tr}(W_x \rho) \mathrm{tr}(W_y \rho) \mathrm{tr}\left(W_x W_y\right) \\
&= \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^{2n}} \mathrm{tr}(W_x \rho)^2 && (\mathrm{tr}(W_x W_y) = 2^n \delta_{xy}) \\
&= \sum_{x \in \mathbb{F}_2^{2n}} p_\rho(x). && \square
\end{aligned}$$

With this in hand, we can establish an analogue of Lemma 3.1 for mixed states.

Lemma A.2. *For all n-qubit mixed states ρ and $x \in \mathbb{F}_2^{2n}$,*

$$q_\rho(x) \leq \frac{\mathrm{tr}(\rho^2)}{2^n} \leq \frac{1}{2^n}.$$

Proof.

$$\begin{aligned} q_\rho(x) &= \sum_{a \in \mathbb{F}_2^{2n}} (-1)^{[x,a]} p_\rho(a)^2 \\ &\leq \sum_{a \in \mathbb{F}_2^{2n}} p_\rho(a)^2 \\ &\leq \frac{1}{2^n} \sum_{a \in \mathbb{F}_2^{2n}} p_\rho(a) \\ &\leq \frac{\mathrm{tr}(\rho^2)}{2^n}. \end{aligned}$$

Above, the first line follows from the definitions of q_ρ and p_ρ and the linearity of the trace. The third line follows from the observation that $p_\rho(x) \leq \frac{1}{2^n}$ always. The last line is an application of Lemma A.1. □

Finally, we prove the analogue of Lemma 3.2, showing that q_ρ places large mass on the unsigned stabilizer groups of large-fidelity stabilizer states.

Lemma A.3. *Let ρ be an n-qubit mixed state and let $|\phi\rangle$ be a stabilizer state that has fidelity τ with ρ, meaning $\langle\phi|\rho|\phi\rangle = \tau$. Let $S = \{x \in \mathbb{F}_2^{2n} : W_x \in \mathrm{Stab}_\pm(|\phi\rangle)\}$. Then*

$$\sum_{x \in S} q_\rho(x) \geq \tau^4.$$

Proof. By definition of q_ρ, we have

$$\begin{aligned} \sum_{x \in S} q_\rho(x) &= \frac{1}{4^n} \sum_{x \in S} \sum_{a \in \mathbb{F}_2^{2n}} (-1)^{[x,a]} \mathrm{tr}\left(W_a^{\otimes 4} \rho^{\otimes 4}\right) \\ &= \sum_{x \in S} \sum_{a \in \mathbb{F}_2^{2n}} (-1)^{[x,a]} p_\rho(a)^2 \\ &= \sum_{a \in \mathbb{F}_2^{2n}} p_\rho(a)^2 \sum_{x \in S} (-1)^{[x,a]} \end{aligned}$$

Now, if $a \in S$ then $[x, a] = 0$ for all $x \in S$, because all elements of a stabilizer group must commute. Otherwise, $[x, a] = 0$ for half of the $x \in S$ and $[x, a] = 1$ for the other half. This means that

$$\sum_{x \in S} (-1)^{[x,a]} = \begin{cases} 2^n & a \in S, \\ 0 & a \notin S. \end{cases}$$

As such, we can simplify:

$$\begin{aligned} \sum_{x \in S} q_\rho(x) &= 2^n \sum_{x \in S} p_\rho(x)^2 \\ &= \frac{1}{2^n} \sum_{x \in S} \mathrm{tr}\left(W_x \rho\right)^4 \\ &\geq \frac{1}{16^n} \left(\sum_{x \in S} |\mathrm{tr}\left(W_x \rho\right)| \right)^4 \qquad \text{(Jensen's Inequality)} \end{aligned}$$

To complete the proof, let $\overline{S} \coloneqq \{P \in \overline{\mathcal{P}_n} : P|\phi\rangle = |\phi\rangle\}$ be the (signed) stabilizer group of $|\phi\rangle$. By well-known calculations (e.g., [41, Equation (2)]), the density matrix of a stabilizer state can be written as an average over the stabilizer group:

$$|\phi\rangle\langle\phi| = \frac{1}{2^n}\sum_{P\in\overline{S}} P. \tag{4}$$

Then:

$$\begin{aligned}
\sum_{x\in S} q_\rho(x) &\geq \frac{1}{16^n}\left(\sum_{x\in S}|\mathrm{tr}\left(W_x\rho\right)|\right)^4 \\
&\geq \frac{1}{16^n}\left(\sum_{P\in\overline{S}}\mathrm{tr}\left(P\rho\right)\right)^4 && \text{(Triangle inequality)} \\
&= \frac{1}{16^n}\left(2^n\mathrm{tr}\left(|\phi\rangle\langle\phi|\,\rho\right)\right)^4 && \text{(Linearity of Trace and Equation (4))} \\
&= \langle\phi|\rho|\phi\rangle^4 \\
&= \tau^4.
\end{aligned}$$

□

It now follows that Algorithm 1 remains correct when run on mixed state inputs.

Corollary A.4. *On input a mixed state ρ in place of $|\psi\rangle$, Algorithm 1 outputs a stabilizer product state $|\phi\rangle$ such that $\langle\phi|\rho|\phi\rangle \geq \max_{|\varphi\rangle\in\mathcal{SP}}\langle\varphi|\rho|\varphi\rangle - \varepsilon$ with probability at least $1/8$.*

Proof sketch. Follow the proof steps in Section 3.1, replacing Lemma 3.1 with Lemma A.2, replacing Lemma 3.2 with Lemma A.3, and identifying each $x \in \mathbb{F}_2^{2n}$ with the corresponding Weyl operator $W_x \in \mathcal{P}_n$. □