

# Quantum State Preparation with Optimal T-Count

David Gosset<sup>1,2,3</sup>, Robin Kothari<sup>1</sup>, and Kewen Wu<sup>4</sup>

<sup>1</sup>Google Quantum AI

<sup>2</sup>Department of Combinatorics and Optimization and Institute for Quantum Computing, University of Waterloo

<sup>3</sup>Perimeter Institute for Theoretical Physics

<sup>4</sup>Computing and Mathematical Sciences Department, Caltech

How many  $T$  gates are needed to approximate an arbitrary  $n$ -qubit quantum state to within error  $\varepsilon$ ? Improving prior work of Low, Kliuchnikov, and Schaeffer, we show that the optimal asymptotic scaling is

$$\Theta\left(\sqrt{2^n \log(1/\varepsilon)} + \log(1/\varepsilon)\right)$$

if we allow ancilla qubits. We also show that this is the optimal  $T$ -count for implementing an arbitrary diagonal  $n$ -qubit unitary to within error  $\varepsilon$ . We describe applications in which a tensor product of many single-qubit unitaries can be synthesized in parallel for the price of one.

## Contents

|          |                                                                                      |           |
|----------|--------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Introduction</b>                                                                  | <b>2</b>  |
| 1.1      | Applications . . . . .                                                               | 5         |
| <b>2</b> | <b>Diagonal unitary synthesis with optimal T-count</b>                               | <b>7</b>  |
| 2.1      | Batched synthesis . . . . .                                                          | 10        |
| 2.2      | Mass production . . . . .                                                            | 11        |
| <b>3</b> | <b>Quantum state preparation with optimal T-count</b>                                | <b>12</b> |
| 3.1      | Approximating a state to constant error: Lemma 3.2 . . . . .                         | 15        |
| 3.2      | Approximating a state to $\varepsilon$ error: Lemma 3.4 . . . . .                    | 16        |
| 3.3      | Flagging and exact amplitude amplification: Lemma 3.5 . . . . .                      | 18        |
| <b>4</b> | <b>Lower bounds</b>                                                                  | <b>20</b> |
| 4.1      | State preparation . . . . .                                                          | 23        |
| 4.2      | Unitary synthesis . . . . .                                                          | 26        |
| <b>A</b> | <b>Canonical form of Clifford circuits with magic states and Pauli postselection</b> | <b>32</b> |
| A.1      | Eliminating ancillas in Clifford circuits . . . . .                                  | 36        |
| <b>B</b> | <b>Improved analysis of Low-Kliuchnikov-Schaeffer state preparation</b>              | <b>37</b> |

# 1 Introduction

The Gottesman-Knill theorem [Got98] reveals a remarkable corner of Hilbert space occupied by Clifford circuits and stabilizer states — classically efficiently simulable but nevertheless equipped with genuinely quantum features such as many-body entanglement and superposition. Only non-Clifford operations can move us out of this corner, a requirement for universal quantum computation. A standard choice is to compile quantum circuits using the Clifford+ $T$  gate set, in which only the single-qubit  $T = \text{diag}(1, e^{i\pi/4})$  gate<sup>1</sup> is non-Clifford.

In recent years there has been an interest in quantifying how much non-Cliffordness, aka *magic*, is required to implement quantum states and operations (see e.g., [VMGE14, BSS16, BBC<sup>+</sup>19, LOH22]). This directly relates to the cost of fault-tolerant quantum computation based on 2D stabilizer error correcting codes such as the surface code, where the implementation of  $T$  gates using magic state distillation is vastly more costly than the transversal implementation of Clifford gates [BK05].<sup>2</sup> A second reason to study quantum magic is its relevance to classical simulation of quantum computers: extensions of the Gottesman-Knill theorem give rise to classical simulation algorithms with a cost that scales polynomially in all parameters except the number of non-Clifford gates [BSS16, BG16, BBC<sup>+</sup>19]. Another line of work has explored the role that magic plays in quantum many-body physics and phases of matter [LW22, EKLH21].

So how much magic is needed to make a quantum state, in the worst case? In particular, let us consider the number of  $T$  gates, or  $T$ -count, required to  $\varepsilon$ -approximate an arbitrary  $n$ -qubit state to within  $\varepsilon$ -error in the  $\ell_2$  norm. Here we consider unitary state preparation circuits composed of Clifford and  $T$  gates.

For  $n = 1$ , a remarkable series of works incorporated techniques from algebraic number theory to give a fairly complete answer to this question [KMM13, Sel15, RS16]. The algorithm of Ross and Selinger [RS16] computes a sequence of<sup>3</sup>  $O(\log(1/\varepsilon))$  single-qubit  $H$  and  $T$  gates that  $\varepsilon$ -approximates a given single-qubit unitary  $U$  (or prepares a 1-qubit state  $U|0\rangle$ ), matching information-theoretic lower bounds [HRC02, BCHK20]. These methods leverage beautiful structural properties of the group generated by single-qubit Clifford and  $T$  gates established in [KMM13] to outperform the well-known approximate synthesis technique based on the Solovay-Kitaev theorem [NC10].

These methods can also be used to prepare an  $n$ -qubit state; one can first express it as  $U|0^n\rangle$  where  $U$  is a sequence of  $O(2^n)$  CNOT and single-qubit gates (see e.g., [BBC<sup>+</sup>95]), and then use the Ross-Selinger algorithm [RS16] to approximate each of the single-qubit gates as a sequence of  $H$  and  $T$  gates. In this way one can upper bound the number of  $T$  gates for  $n$ -qubit state preparation as  $O(2^n(n + \log(1/\varepsilon)))$ . Note that this is also an upper bound on the total number of gates used. Surprisingly, [LKS24, Section 3] showed that the  $T$ -count of  $n$ -qubit state synthesis can be improved by almost a square-root factor to

$$O\left(\sqrt{2^n n \log(n/\varepsilon)} + \log^2(n/\varepsilon)\right), \quad (1)$$

if we allow the use of ancilla qubits that are prepared and returned to the all-zeros state.<sup>4</sup>

<sup>1</sup>We use  $\text{diag}(c_1, \dots, c_m)$  to denote the diagonal matrix with  $c_1, \dots, c_m$  on the diagonal in order.

<sup>2</sup>A recent line of research [Lit19, GSJ24] shows that  $T$  gates might not be as expensive as previously thought for certain architectures and fault tolerance schemes.

<sup>3</sup>In this paper, asymptotic notations  $O(\cdot), \Omega(\cdot)$  only hide absolute constants that do not depend on any parameter, and  $\log(x)$  is the base 2 logarithm of  $x$ .

<sup>4</sup>It is possible to tighten their analysis to give better (but still sub-optimal) results. See Section B for

The synthesis algorithm of [LKS24] has two important ingredients. The first ingredient is a state preparation method proposed by Grover and Rudolph [GR02], which we now describe. For simplicity suppose the target state  $|\psi\rangle = \sum_{x \in \{0,1\}^n} \alpha_x |x\rangle$  has non-negative real amplitudes  $\alpha_x \geq 0$ . Then, starting with  $|0^n\rangle$ , we first apply a single-qubit rotation on the first qubit to ensure it has the correct marginal, i.e.,  $|0\rangle \rightarrow \sqrt{\sum_{x:x_1=0} \alpha_x^2} |0\rangle + \sqrt{\sum_{x:x_1=1} \alpha_x^2} |1\rangle$ . Then, controlled on the first qubit, we rotate the second qubit to have the correct marginal. We repeat this process until the  $n$ -th qubit. This reduces state preparation to the task of synthesizing  $n$  multiply-controlled single-qubit unitaries, or (using Euler angles)  $3n$  *diagonal* multiply-controlled single-qubit unitaries. The second ingredient is a method for approximating diagonal controlled single-qubit unitaries that starts with a  $b = O(\log(n/\varepsilon))$ -bit approximation of the rotation angle and then implements a controlled rotation corresponding to each bit. The technique used to implement the latter controlled rotations can be viewed as a generalization of the fact that an  $n$ -qubit *Boolean* diagonal unitary, with nonzero entries in  $\pm 1$ , can be implemented with  $T$ -count  $\Theta(\sqrt{2^n})$  [Nec62, Sch89, Mas16, LKS24].

If we look beyond the headline square-root dependence  $\sqrt{2^n}$ , we can identify contributions to the  $T$ -count from each of these two ingredients that we might hope to improve: an overhead from the Grover-Rudolph method that scales with the number of qubits  $n$  (under the square root in the first term of (1)) as well as an overhead that scales as  $\log^2(n/\varepsilon)$  that arises from implementing  $b$  controlled single-qubit rotations, each to within  $\varepsilon$ -error.

In this work we show that indeed both of these overheads can be avoided. Our main result is the following improvement which gives an optimal upper bound on the  $T$ -count of an  $n$ -qubit state.

**Theorem 1.1** (Quantum state preparation with optimal  $T$ -count). *Any  $n$ -qubit state can be prepared up to error  $\varepsilon$  by a Clifford+ $T$  circuit starting with the all-zeros state using*

$$O\left(\sqrt{2^n \log(1/\varepsilon)} + \log(1/\varepsilon)\right) \quad (2)$$

*$T$  gates and ancillas.<sup>5</sup> Furthermore, no Clifford+ $T$  circuit (even with measurements and adaptivity) can use asymptotically fewer  $T$  gates.*

More precisely, for any  $n$ -qubit state  $|\psi\rangle$ , we show there is a Clifford+ $T$  circuit  $U$  such that  $U|0^n\rangle|0^a\rangle = |\tilde{\psi}\rangle|0^a\rangle$ , where  $\| |\psi\rangle - |\tilde{\psi}\rangle \| \leq \varepsilon$ . The  $T$ -count of  $U$  and  $a$ , the number of ancillas used, are upper bounded by (2).

The lower bound in Theorem 1.1, which we formally establish in Theorem 4.1, holds in an even stronger model that we call *adaptive* Clifford+ $T$  circuits (defined formally in Section 4). In this model, the algorithm may apply mid-circuit measurements and any future gates may depend on the outcomes of prior measurements. Such a circuit naturally outputs a mixed state, and even if we only require this state to be  $\varepsilon$ -close to the target state in trace distance, we show that the lower bound on  $T$ -count in (2) still holds. The lower bound is established combining the  $\Omega(\log(1/\varepsilon))$  lower bound from [BCHK20], which holds for adaptive Clifford+ $T$  circuits, and a slight generalization of the  $\Omega(\sqrt{2^n \log(1/\varepsilon)})$  lower bound in [LKS24], which only holds for unitary Clifford+ $T$  circuits.<sup>6</sup> Thus Theorem 1.1

details.

<sup>5</sup>This construction uses  $O(2^n \log(1/\varepsilon))$  Clifford gates.

<sup>6</sup>We thank Luke Schaeffer for confirming that their  $\Omega(\sqrt{2^n \log(1/\varepsilon)})$  lower bound at the end of [LKS24, Section 5] should actually be  $\Omega(\sqrt{2^n \log(1/\varepsilon)})$ .

represents the best of both worlds, with the upper bound being established in the weak model, but the lower bound being established in the strong model.

The circuit in [Theorem 1.1](#) is obtained by refining each of the two ingredients in the LKS construction to avoid the overheads described above. We now describe our refinements.

For the first ingredient, we use a variant of recent methods from [\[INN<sup>+</sup>22, Ros24\]](#) which reduce quantum state synthesis to the synthesis of  $O(1)$  diagonal unitaries — far fewer than are needed in the Grover-Rudolph method. We start with a coarse approximation: we use just two *Boolean* diagonal unitaries to construct a state  $|\phi\rangle$  that has a constant overlap with the target state  $|\psi\rangle$  (i.e.,  $\varepsilon$  is a constant). Specifically, we show that there exist two diagonal unitaries  $B_1, B_2$  with diagonal entries in  $\pm 1$ , such that  $|\phi\rangle := B_2 H^{\otimes n} B_1 H^{\otimes n} |0^n\rangle$  has a constant overlap with  $|\psi\rangle$ . Starting from this coarse approximation, we use an idea from [\[Ros24\]](#) to improve the accuracy to the desired value of  $\varepsilon$ . Since  $|\phi\rangle$  already has a constant overlap with  $|\psi\rangle$ , intuitively their difference  $|\psi\rangle - |\phi\rangle$  should have a small length. Then our goal becomes to construct  $|\psi\rangle - |\phi\rangle$ , for which we again have a coarse approximation. Iterating this, we obtain  $|\psi\rangle$  in the limit, where the error decays exponentially:  $|\psi\rangle$  is  $\varepsilon$ -close to

$$|\psi'\rangle := \beta \cdot \sum_{k=0}^{O(\log(1/\varepsilon))} \gamma^k |\phi_k\rangle \quad (3)$$

for some constants  $\beta$  and  $\gamma$ , where each  $|\phi_k\rangle = B_2^{(k)} H^{\otimes n} B_1^{(k)} H^{\otimes n} |0^n\rangle$  is a coarse approximation of the previous step. To prepare this state, we use the LCU (linear combination of unitaries) method along with amplitude amplification, see [Subsection 3.3](#) for details.

For the second ingredient, we give a new method to synthesize diagonal unitaries with optimal  $T$ -count.

**Theorem 1.2** (Diagonal unitary synthesis with optimal  $T$ -count). *Any diagonal unitary on  $n$  qubits can be implemented up to error  $\varepsilon$  by a Clifford+ $T$  circuit using*

$$O\left(\sqrt{2^n \log(1/\varepsilon)} + \log(1/\varepsilon)\right) \quad (4)$$

*$T$  gates and ancillas.<sup>7</sup> Furthermore, no Clifford+ $T$  circuit (even with measurements and adaptivity) can use asymptotically fewer  $T$  gates.*

Here, the distance between two unitaries is the operator norm (aka spectral norm) of their difference, i.e.,  $U$  implements  $V$  to error  $\varepsilon$  if  $\|U - V\| = \varepsilon$ . The lower bound in [Theorem 1.2](#), which is formally proved in [Theorem 4.2](#), holds in the stronger model of adaptive Clifford+ $T$  circuits, as before. The lower bound is proved by a reduction to the previous lower bound; see [Section 4](#) for details.

To prove the upper bound from [Theorem 1.2](#), we first note that an  $n$ -qubit diagonal unitary  $D$  can be regarded as a single-qubit diagonal unitary  $G_y$  controlled on  $n - 1$  qubits  $|y\rangle$ . Now we use the fact that any single-qubit unitary  $G_y$  can be approximated by a sequence of  $O(\log(1/\varepsilon))$  Hadamard and  $T$  gates (using no ancillas) [\[KMM13, Sel15, KMM15, RS16\]](#). To synthesize  $D$ , we start by applying a Boolean oracle  $B: |y\rangle|z\rangle|0\rangle \rightarrow |y\rangle|z\rangle|s_y\rangle$ , where  $y \in \{0, 1\}^{n-1}$ ,  $z \in \{0, 1\}$ , and  $s_y$  is a binary string of length  $O(\log(1/\varepsilon))$  that describes a sequence of  $H$  and  $T$  gates that  $\varepsilon$ -approximates  $G_y$ . This Boolean oracle  $B$  can be implemented with  $T$ -count  $O\left(\sqrt{2^n \log(1/\varepsilon)}\right)$  [\[Mas16, LKS24\]](#). After this step, we apply a sequence of controlled-Hadamard and controlled- $T$  gates on the  $n$ -th qubit, each

---

<sup>7</sup>This construction uses  $O(2^n \log(1/\varepsilon))$  Clifford gates.

controlled on a bit of  $|s_y\rangle$ . The total  $T$ -cost of this last step is  $O(\log(1/\varepsilon))$ , proportional to the length of  $s_y$ . Finally, we uncompute the Boolean oracle  $B$ , returning the third register to the all-zeros state.

We note that it is possible to directly improve on the state synthesis technique from [LKS24] using Theorem 1.2; this gives a better  $T$ -count than (1), but still worse than the optimal bound (2). See Section B for details.

## 1.1 Applications

There are some immediate gains by using our bounds in place of those from [LKS24] in known applications. For instance, we obtain slight improvements to the state-of-the-art lower bound on the stabilizer rank of magic states, due to Mehraban and Tahmasbi [MT24].<sup>8</sup> Our results also improve some related lower bounds concerning decompositions of Boolean functions as linear combination of quadratic functions [FHH<sup>+</sup>14]; see [MT24] for details. We can also obtain improved bounds for “partial unitary synthesis” where the goal is to implement the first  $K$  columns of an  $n$ -qubit unitary. Here state preparation corresponds to the case of  $K = 1$  and unitary synthesis is the case of  $K = 2^n$ . Using Householder reflections [Hou58], this can be reduced to the task of preparing up to  $K$  different  $n$ -qubit states [Kli13, LKS24]. Via this reduction and with our Theorem 1.1, we obtain a  $T$ -count of  $O\left(K\sqrt{2^n \log(K/\varepsilon)} + K \log(K/\varepsilon)\right)$ , which similarly shaves polylogarithmic factors from the bounds in [LKS24].

Below we discuss two further applications of Theorem 1.2 to synthesizing tensor products of single-qubit unitaries.

**Batched Synthesis of Single-Qubit Unitaries.** Results from [Sel15, BCHK20] show that synthesizing a single-qubit unitary has  $T$ -count  $\Theta(\log(1/\varepsilon))$ . What is the cost of implementing  $m$  single-qubit unitaries  $U_1 \otimes \cdots \otimes U_m$  to error  $\varepsilon$ ?<sup>9</sup> The most natural bound would be  $O(m \log(m/\varepsilon))$  by implementing each up to  $\varepsilon/m$  error. Surprisingly, a corollary of our previous result implies that the total cost remains the same  $\log(1/\varepsilon)$  even if  $m$  has slightly superconstant dependence on  $1/\varepsilon$ .

**Theorem 1.3.** *Let  $m = O(\log \log(1/\varepsilon))$  be an integer. A tensor product of  $m$  (potentially different) single-qubit unitaries  $U_1 \otimes \cdots \otimes U_m$  can be implemented up to error  $\varepsilon$  by a Clifford+ $T$  circuit using  $O(\log(1/\varepsilon))$   $T$  gates and ancillary qubits.<sup>10</sup>*

Theorem 1.3 is a simple consequence of Theorem 1.2. To see this, note that we can use the Euler angle decomposition to express each single-qubit unitary  $U_j$  as a sequence of Hadamards and at most 3 diagonal unitaries, i.e.,  $U_j = D_{3,j} H D_{2,j} H D_{1,j}$ , where  $D_{i,j}$  are diagonal single-qubit unitaries for  $i \in \{1, 2, 3\}$  and  $j \in [m]$ . This reduces our synthesis task to the case in which all the single-qubit unitaries are diagonal; clearly it is enough to synthesize  $D_{i,1} \otimes D_{i,2} \otimes \cdots \otimes D_{i,m}$  for  $i = 1, 2, 3$ . But now we can apply Theorem 1.2 which implies that any  $\log \log(1/\varepsilon)$ -qubit diagonal unitary can be implemented with  $T$ -count  $O(\log(1/\varepsilon))$ .

An interesting direction for future work would be to improve on this batched synthesis result (or show that improvements are not possible). In particular, it is an open question

<sup>8</sup>In particular, we can shave off two factors of  $\log m$  to get a slightly improved bound for the approximate stabilizer rank  $\chi_\delta(|T\rangle^{\otimes m}) = \Omega\left(\frac{m^2}{\log^2 m}\right)$  for  $\delta = \Omega(1)$ .

<sup>9</sup>Here  $\varepsilon$  is the operator-norm error in approximating the entire unitary  $U_1 \otimes \cdots \otimes U_m$ .

<sup>10</sup>This construction uses  $O(\log(1/\varepsilon))$  Clifford gates.

whether or not an exponentially strengthened version of the above theorem holds with  $m = \Omega(\log(1/\varepsilon))$ .

**Mass Production of a Single-Qubit Unitary.** The batched synthesis result ([Theorem 1.3](#)) focused on implementing  $m$  *different* single-qubit unitaries. It is natural to ask what happens if we want to prepare  $m$  *identical* single-qubit unitaries. This is called “mass production” [[Uli74](#), [Uhl92](#), [Kre23](#)] in circuit synthesis and lives in a richer family of direct sum problems; see the survey [[Pan12](#)] for details. As an application of [Theorem 1.2](#), we prove the following optimal  $T$ -count for mass production of a single-qubit unitary.

**Theorem 1.4.** *Let  $U$  be an arbitrary single-qubit unitary. Then  $U^{\otimes m}$  can be implemented up to error  $\varepsilon$  by a Clifford+ $T$  circuit using  $O(m + \log(1/\varepsilon))$   $T$  gates and ancillary qubits.<sup>11</sup> Furthermore, no Clifford+ $T$  circuit (even with measurements and adaptivity) can use asymptotically fewer  $T$  gates.*

The bound in [Theorem 1.4](#) is indeed better than the one in [Theorem 1.3](#) if the goal is to implement the same unitary several times: [Theorem 1.4](#) shows that using  $O(\log(1/\varepsilon))$   $T$  gates, we can prepare  $m = O(\log(1/\varepsilon))$  copies of a given single-qubit unitary  $U$ , in contrast to [Theorem 1.3](#), which says we can implement  $m = O(\log \log(1/\varepsilon))$  (potentially different) unitaries.

[Theorem 1.4](#) is also a corollary of [Theorem 1.2](#). Using the Euler angle decomposition, we can reduce to the case where  $U$  is diagonal, say,  $U = \text{diag}(1, e^{i\theta})$ . Then  $U^{\otimes m}|x\rangle = e^{i\theta \cdot h_x}|x\rangle$ , where  $h_x \in \{0, 1, \dots, m\}$  is the Hamming weight of  $x \in \{0, 1\}^m$ . Therefore, we can first compute the Hamming weight into a second register containing  $L = \lceil \log m \rceil$  qubits  $|x\rangle \rightarrow |x\rangle|h_x\rangle$ , then apply the corresponding phase change  $|h_x\rangle \rightarrow e^{i\theta \cdot h_x}|h_x\rangle$ . The first step can be done with  $O(m)$  complexity in a standard way (see [Fact 2.6](#) for details). For the second step, since  $h_x$  has  $L$  bits, the phase change is a diagonal unitary on  $L = \lceil \log m \rceil$  qubits, which can be implemented with  $T$ -count  $O(\sqrt{m \log(1/\varepsilon)} + \log(1/\varepsilon)) \leq O(m + \log(1/\varepsilon))$  by [Theorem 1.2](#).

Note that one can implement a similar but sub-optimal strategy without the use of [Theorem 1.2](#), see, e.g., [[BCHK20](#), Section 3.2]. Indeed, one can implement the second step as  $R(\theta) \otimes R(2\theta) \otimes \dots \otimes R(2^L\theta)|h_x\rangle = e^{i\theta \cdot h_x}|h_x\rangle$  where  $R(\alpha) \equiv \text{diag}(1, e^{i\alpha})$  and then use the Ross-Selinger algorithm to approximate each single-qubit gate  $R(2^k\theta)$  to within error  $\varepsilon/L$ . This strategy gives an overall  $T$ -count of

$$O(m + L \cdot \log(L/\varepsilon)) = O(m + \log(m) \log(\log(m)/\varepsilon)) \quad (5)$$

$$= O(m + \log(m) \log(1/\varepsilon)), \quad (6)$$

which is worse than the optimal bound given above.

Finally we remark that it is easy to see that the  $T$ -count upper bound  $O(m + \log(1/\varepsilon))$  from [Theorem 1.4](#) is asymptotically tight. On the one hand, the  $\Omega(\log(1/\varepsilon))$   $T$ -count lower bound holds even if  $m = 1$  [[BCHK20](#)]. On the other hand, if  $U$  is the  $T$  gate, we cannot hope to compress  $T^{\otimes m}$  in a generic way, which otherwise contradicts the optimality of [Theorem 1.1](#). See details in [Subsection 2.2](#).

**Discussion.** In this work we have characterized the worst-case  $T$ -count of quantum states and diagonal unitaries. A challenging open question is to understand the optimal  $T$ -count for general  $n$ -qubit unitaries. We can prove an  $\tilde{\Omega}(2^n)$   $T$ -count lower bound (see

<sup>11</sup>This construction uses  $O(m \log(1/\varepsilon))$  Clifford gates.



Theorem 4.3 for details). However there is a glaringly large gap between this lower bound and the best known upper bound  $\tilde{O}(2^{1.5n})$  [LKS24, Ros21]. This upper bound is obtained in [LKS24] by reducing the unitary synthesis task to that of preparing  $2^n$   $n$ -qubit states, each of which uses  $\tilde{O}(\sqrt{2^n})$   $T$  gates; [Ros21] also shows how to synthesize a unitary efficiently with  $O(\sqrt{2^n})$  calls to a  $2n$ -qubit Boolean oracles, each of which needs  $O(2^n)$   $T$  gates (see Lemma 2.1).

While our focus is on  $T$ -count and our construction gives the asymptotic optimal bound, the number of Clifford gates in our state preparation circuit is  $O(2^n \log(1/\varepsilon))$ , which can be verified given the explicit description of our synthesis procedure. This bound is close to the optimal  $\Omega(2^n/n)$  lower bound. In particular, if the goal is to minimize total gate count, we can replace Lemma 2.1 and Remark 2.2 with the construction that achieves the  $O(2^n/n)$  gate count for Boolean functions; then our final circuit will have  $O(2^n/n)$  gates, albeit a similar amount of  $T$  gates. It remains an open problem whether it is possible to achieve  $O(\sqrt{2^n})$   $T$ -count and  $O(2^n/n)$  total gate count at the same time.

We emphasize that our savings on the  $T$ -count is made possible with the use of ancillary qubits, which is also necessary by a simple counting argument. By tuning the construction for Lemma 2.1 with different number of ancillas and  $T$ -count, we believe our construction achieves near optimal tradeoffs between ancillas and  $T$ -count (with some logarithmic slack in  $n$  and  $1/\varepsilon$ ). The best achievable tradeoff between these two metrics is still unknown.

A similarly interesting question is to study the  $T$ -depth in place of  $T$ -count. The most depth-consuming part in our construction is the synthesis of (Boolean) diagonal unitaries, which, in light of Remark 2.2, should be  $\text{poly}(n)$ . The optimal  $T$ -depth is still open and we expect that the lower bound side will be significantly more challenging, if given unbounded ancillas.

**Subsequent Work.** A recent followup by Tan [Tan25] showed how to improve the unitary synthesis  $T$ -count to  $\tilde{O}(2^{4n/3})$ . This is a significant improvement over the aforementioned  $\tilde{O}(2^{1.5n})$  bound, but is still far from the  $\Omega(2^n)$  lower bound.

**Paper Organization.** The remainder of the paper is organized as follows. In Section 2, we establish the optimal  $T$ -count for diagonal unitaries (Theorem 1.2) and in Subsection 2.1 and Subsection 2.2 we describe the applications to batched synthesis (Theorem 1.3) and mass production (Theorem 1.4). In Section 3, we bring in other techniques and prove the optimal  $T$ -count for state preparation (Theorem 1.1). Detailed proofs of lower bounds for Theorem 1.1 (as Theorem 4.1) and Theorem 1.2 (as Theorem 4.2) are presented in Section 4. A canonical form for Clifford+ $T$  circuit with Pauli postselections is proved in Section A, which generalizes [BCHK20]. An improved analysis of [LKS24] is in Section B.

## 2 Diagonal unitary synthesis with optimal T-count

In this section, we prove Theorem 1.2 (restated below) and give details of the applications described in the previous section.

**Theorem 1.2** (Diagonal unitary synthesis with optimal  $T$ -count). *Any diagonal unitary on  $n$  qubits can be implemented up to error  $\varepsilon$  by a Clifford+ $T$  circuit using*

$$O\left(\sqrt{2^n \log(1/\varepsilon)} + \log(1/\varepsilon)\right) \quad (4)$$

*T* gates and ancillas.<sup>12</sup> Furthermore, no Clifford+*T* circuit (even with measurements and adaptivity) can use asymptotically fewer *T* gates.

In [Theorem 1.2](#), the diagonal entries of the diagonal unitary *D* can be arbitrary complex numbers of magnitude 1. An important special case is obtained by restricting to “Boolean phase oracles”: diagonal unitaries with  $\pm 1$  diagonal entries. Such unitaries can be implemented using Clifford+*T* circuits with zero error. Moreover, it is known how to achieve this using very few *T* gates [[Nec62](#), [Sch89](#), [Mas16](#), [LKS24](#)]. We shall use the following statement which is phrased in terms of the standard oracle that computes the Boolean function into an ancilla register.

**Lemma 2.1** ([[LKS24](#), Theorem 2]). *Let  $b \geq 1$  be an integer and  $f: \{0, 1\}^n \rightarrow \{0, 1\}^b$  be an arbitrary Boolean function. Define  $U_f$  as the unitary mapping  $|x\rangle|y\rangle$  to  $|x\rangle|y \oplus f(x)\rangle$  for all  $x \in \{0, 1\}^n, y \in \{0, 1\}^b$ . Then  $U_f$  can be implemented exactly by a Clifford+*T* circuit using  $O(\sqrt{b \cdot 2^n})$  *T* gates and ancillary qubits.<sup>13</sup>*

**Remark 2.2.** We sketch the idea to prove [Lemma 2.1](#). Assume *f* has a classical circuit consisting of XOR gates and AND gates with bounded fan-in. By converting XOR gate into CNOT gate (which is Clifford) and AND gate into Toffoli gate (which has a constant *T*-count), we can synthesize  $U_f$  with *T*-count proportional to the number of AND gates. The latter quantity is termed “multiplicative complexity” in the classical complexity theoretic literature and is thoroughly studied in e.g., [[Nec62](#), [Sch89](#)]; here we also give a brief overview of the upper bound.

Assume  $b = 1$  for simplicity. One way to compute *f* is to gradually fix each input bit:  $f(x) = (x_1 \wedge f_1(x_2, \dots, x_n)) \oplus f_2(x_2, \dots, x_n)$ , where  $f_2$  (resp.,  $f_1 \oplus f_2$ ) is *f* with  $x_1$  fixed to 0 (resp., 1). Let  $d \in [n]$  be an integer to be optimized later. If we iterate the above expansion for  $x_1, \dots, x_d$ , then we need  $2^d$  AND gates and it remains to compute  $2^d$  Boolean functions on  $x_{d+1}, \dots, x_n$ . Now observe that once all conjunctions  $\bigwedge_{j \in S} x_j$ , for  $S \subseteq \{d+1, \dots, n\}$  are computed (which takes  $2^{n-d}$  AND gates<sup>14</sup>), we just need to XOR some of them for each one of the remaining  $2^d$  Boolean functions. Therefore, the total number of AND gates is  $2^d + 2^{n-d}$ , which has the claimed minimum value  $O(\sqrt{2^n})$  by setting  $d = n/2$ . In the case of  $b > 1$ , an analogous calculation shows the total number of AND gates is roughly  $b \cdot 2^d + 2^{n-d}$ , which has minimum value  $O(\sqrt{b \cdot 2^n})$  as claimed.

Note that the number of XOR gate used is  $O(b \cdot 2^n)$ .

We shall also use the known *T*-optimal synthesis of single-qubit unitaries without ancillas [[KMM13](#), [Sel15](#), [KMM15](#), [RS16](#)].

**Lemma 2.3** ([[Sel15](#), [RS16](#)]). *Any single-qubit unitary with determinant 1<sup>15</sup> can be implemented up to error  $\varepsilon$  by a Clifford+*T* circuit using  $O(\log(1/\varepsilon))$  *T* gates and without ancillary qubits.<sup>16</sup>*

<sup>12</sup>This construction uses  $O(2^n \log(1/\varepsilon))$  Clifford gates.

<sup>13</sup>This construction uses  $O(b \cdot 2^n)$  Clifford gates.

<sup>14</sup>To see this, note that each  $|S| = k$  is an AND of some size- $(k-1)$  conjunction and one extra variable. Therefore, building on size- $(< k)$  conjunctions, we only need to introduce  $\binom{n-d}{k}$  AND gates to get size- $(\leq k)$  conjunctions. So the total cost is  $\sum_{k \geq 2} \binom{n-d}{k} < 2^{n-d}$ .

<sup>15</sup>As noted in [[Sel15](#)], this determinant condition is needed since a Clifford+*T* circuit has determinant in  $\{e^{i \cdot t\pi/4} : t = 0, 1, \dots, 7\}$ . This is not a problem in practice because the determinant can be adjusted by multiplying the unitary with a global phase.

<sup>16</sup>In particular, the number of Clifford gates is also  $O(\log(1/\varepsilon))$ .



Below we give the proof of [Theorem 1.2](#). The idea is to view the diagonal unitary  $D$  as a single-qubit unitary acting on the  $n$ th qubit, controlled on the state of the first  $n - 1$  qubits. For each fixed value of the control bits, we have a single-qubit unitary that can be approximated by a sequence<sup>17</sup> of Hadamard and  $T$  gates using [Lemma 2.3](#). To implement  $D$ , we first use [Lemma 2.1](#) to compute a Boolean function that takes as input the  $n - 1$  control bits, and outputs a description of the approximating sequence of Hadamard and  $T$  gates into an ancilla register. Controlled on this ancilla register, we then apply the corresponding sequence of single-qubit gates.

*Proof of Theorem 1.2.* Let  $D = \text{diag}(\alpha_1, \dots, \alpha_{2^n})$  where  $\alpha_1, \dots, \alpha_{2^n}$  are complex numbers on the unit circle. Define a diagonal unitary  $D' = \text{diag}(\alpha_1, \bar{\alpha}_1, \dots, \alpha_{2^n}, \bar{\alpha}_{2^n})$  on  $n + 1$  qubits, where  $\bar{\alpha}_j$  is the complex conjugate of  $\alpha_j$ . Note that  $D' = D \otimes |0\rangle\langle 0| + D^\dagger \otimes |1\rangle\langle 1|$ . In order to implement  $D$ , it suffices to implement  $D'$  on our  $n$ -qubit input register along with an ancilla initialized in the  $|0\rangle$  state.

For each  $j = 1, 2, \dots, 2^n$ , define  $U_j = \text{diag}(\alpha_j, \bar{\alpha}_j)$ , which is a single-qubit unitary with determinant 1. Then  $D' = \text{diag}(U_1, \dots, U_{2^n})$ . Moreover, we know from [Lemma 2.3](#) that each  $U_j$  is  $\varepsilon$ -close to a single-qubit unitary  $\tilde{U}_j$  that can be exactly implemented as a sequence of  $O(\log(1/\varepsilon))$  Hadamard and  $T$  gates. It follows that  $\|D' - \tilde{U}\| \leq \varepsilon$ , where  $\tilde{U} = \text{diag}(\tilde{U}_1, \dots, \tilde{U}_{2^n})$ . To complete the proof we describe a Clifford+ $T$  circuit that exactly implements  $\tilde{U}$  using the number of  $T$  gates and ancillas claimed in the Theorem statement.

From [Lemma 2.3](#) we infer that for some  $K = \Theta(\log(1/\varepsilon))$  and each  $U_j$ , we have

$$\tilde{U}_j = H^{a_j^1} T^{b_j^1} H^{a_j^2} T^{b_j^2} \dots H^{a_j^K} T^{b_j^K}, \quad (7)$$

for some Boolean variables  $a_j^1, \dots, a_j^K, b_j^1, \dots, b_j^K \in \{0, 1\}$ . Define a Boolean function  $f: \{0, 1\}^n \rightarrow \{0, 1\}^{2K}$  to record the gate sequence information  $a, b$  for each  $\tilde{U}_j$ :

$$f(j) = (a_j^1, b_j^1, \dots, a_j^K, b_j^K). \quad (8)$$

Then by [Lemma 2.1](#), the corresponding  $(n + 2K)$ -qubit unitary  $F$  mapping  $|j\rangle|y\rangle$  to  $|j\rangle|y \oplus f(j)\rangle$  can be implemented with  $O(\sqrt{K \cdot 2^n}) = O(\sqrt{2^n \log(1/\varepsilon)})$   $T$  gates and ancillas.

Once the gate sequence is computed into an ancilla register, it remains to apply  $H$  or  $T$  on the  $(n + 1)$ -th qubit controlled on this ancilla register. For this purpose, we note that controlled- $H$  and controlled- $T$  are two-qubit unitaries that can be exactly implemented with  $O(1)$   $T$  gates without ancillas [[GS13](#)].

Putting this together, we arrive at the following implementation of  $\tilde{U}$ .

1. Let  $\mathcal{A} = (\mathcal{A}_0, \mathcal{A}_1)$  be the  $(n + 1)$ -qubit register where we implement  $\tilde{U}$ . Note that  $\mathcal{A}_0$  is an  $n$ -qubit register and  $\mathcal{A}_1$  is a single-qubit register.

Let  $\mathcal{B} = (\mathcal{B}_1, \dots, \mathcal{B}_{2K})$  and  $\mathcal{C}$  be auxiliary registers where each  $\mathcal{B}_j$  is single-qubit and  $\mathcal{C}$  is  $O(\sqrt{2^n \log(1/\varepsilon)})$ -qubit. These registers are initialized in the all-zero state and serve as ancillas.

2. Using  $\mathcal{C}$ , we apply the unitary  $F$  described above on  $\mathcal{A}_0$  and  $\mathcal{B}$ .
3. For each  $\ell = 1, 2, \dots, 2K$ , we either apply  $H$  on  $\mathcal{A}_1$  controlled by  $\mathcal{B}_\ell$  (if  $\ell$  is odd), or we apply  $T$  on  $\mathcal{A}_1$  controlled by  $\mathcal{B}_\ell$  (if  $\ell$  is even).

<sup>17</sup>Since [Lemma 2.3](#) uses no ancilla, the Clifford gates are only Hadamard  $H$  and Phase gate  $S = T^2$ . Hence we simply assume the sequence contains only Hadamard and  $T$  gates.

4. Using  $\mathcal{C}$ , we apply the unitary  $F$  again to restore the ancillas in  $\mathcal{B}$  to the all-zeros state.

To see that this implements  $\tilde{U}$  as desired, we can verify the action of each of the above steps when the initial state of registers  $\mathcal{A}_0$  and  $\mathcal{A}_1$  are basis vectors  $|j\rangle$  and  $|x\rangle$ , respectively, where  $j \in \{1, 2, \dots, 2^n\}$  and  $x \in \{0, 1\}$ :

$$|j\rangle|x\rangle_{\mathcal{A}_1}|0\rangle_{\mathcal{B}}|0\rangle_{\mathcal{C}} \rightarrow |j\rangle|x\rangle_{\mathcal{A}_1}|f(j)\rangle_{\mathcal{B}}|0\rangle_{\mathcal{C}} \quad (9)$$

$$\rightarrow |j\rangle\tilde{U}_j|x\rangle_{\mathcal{A}_1}|f(j)\rangle_{\mathcal{B}}|0\rangle_{\mathcal{C}} \quad (10)$$

$$\rightarrow |j\rangle\tilde{U}_j|x\rangle_{\mathcal{A}_1}|0\rangle_{\mathcal{B}}|0\rangle_{\mathcal{C}}. \quad (11)$$

The number of  $T$  gates used in the above implementation is upper bounded as

$$\underbrace{O\left(\sqrt{2^n \log(1/\varepsilon)}\right)}_{\text{Item 2 and Item 4}} + 2K \cdot \underbrace{O(1)}_{\text{each step in Item 3}} = O\left(\sqrt{2^n \log(1/\varepsilon)} + \log(1/\varepsilon)\right). \quad (12)$$

It is also clear that this upper bounds the number of ancillas used.

The tightness of [Theorem 1.2](#) is deferred as [Theorem 4.2](#) to be proved in [Section 4](#).  $\square$

## 2.1 Batched synthesis

We now discuss the applications of [Theorem 1.2](#) mentioned in the introduction.

First we discuss *batched synthesis* of single-qubit unitaries: we will see that we can implement  $m = O(\log \log(1/\varepsilon))$  single-qubit gates to error  $\varepsilon$  with  $T$ -count  $O(\log(1/\varepsilon))$ .

We are interested in synthesizing a general tensor product  $U_1 \otimes U_2 \otimes \dots \otimes U_m$  of single-qubit unitaries. In order to use [Theorem 1.2](#) we will first reduce to the case where each  $U_j$  is diagonal in the standard basis. To this end we can use the well-known Euler angle decomposition (see e.g., [\[NC10\]](#)) which expresses a single-qubit unitary  $U$  as a product  $U = e^{i\delta} e^{i\gamma Z} e^{i\beta X} e^{i\alpha Z}$  for some real numbers  $\alpha, \beta, \gamma, \delta$ . Here  $X, Y, Z$  are the Pauli matrices. Since  $e^{i\beta X} = H e^{i\beta Z} H$  we can summarize this fact as follows.

**Fact 2.4.** *Let  $U$  be a single-qubit unitary. Then there exist single-qubit diagonal unitaries  $A, B, C$  such that  $U = AHBHC$ , where  $H$  is the single-qubit Hadamard matrix.*

Now we reduce the batched synthesis task to [Theorem 1.2](#).

**Corollary 2.5.** *Let  $m \geq 1$  be an integer and  $U_1, U_2, \dots, U_m$  be arbitrary single-qubit unitaries. Define an  $m$ -qubit unitary  $U = U_1 \otimes \dots \otimes U_m$ . Then  $U$  can be implemented up to error  $\varepsilon$  by a Clifford+ $T$  circuit using  $O\left(\sqrt{2^m \log(1/\varepsilon)} + \log(1/\varepsilon)\right)$   $T$  gates and ancilla qubits.<sup>18</sup>*

*Proof.* By [Fact 2.4](#), each  $U_j$  can be expanded as  $A_j H B_j H C_j$  where  $A_j, B_j, C_j$  are single-qubit diagonal unitaries. As a result, we have  $U = A H^{\otimes m} B H^{\otimes m} C$ , where  $A = A_1 \otimes \dots \otimes A_m, B = B_1 \otimes \dots \otimes B_m, C = C_1 \otimes \dots \otimes C_m$  are  $m$ -qubit diagonal unitaries, which can each be implemented using [Theorem 1.2](#) with the claimed number of  $T$  gates and ancillas.  $\square$

[Theorem 1.3](#) then follows directly from [Corollary 2.5](#).

<sup>18</sup>This construction uses  $O(2^m \log(1/\varepsilon))$  Clifford gates.

**Theorem 1.3.** *Let  $m = O(\log \log(1/\varepsilon))$  be an integer. A tensor product of  $m$  (potentially different) single-qubit unitaries  $U_1 \otimes \cdots \otimes U_m$  can be implemented up to error  $\varepsilon$  by a Clifford+T circuit using  $O(\log(1/\varepsilon))$  T gates and ancillary qubits.<sup>19</sup>*

*Proof.* We divide the  $m$  single-qubit unitaries into  $K = O(1)$  groups of  $\lceil \log \log(1/\varepsilon) \rceil$  each. It then suffices to implement each group up to error  $\varepsilon/K$ . This is handled by [Corollary 2.5](#).  $\square$

## 2.2 Mass production

Next we consider the *mass production* of a single-qubit unitary. Here the goal is to implement  $U^{\otimes m}$  for some single-qubit unitary  $U$ .

**Theorem 1.4.** *Let  $U$  be an arbitrary single-qubit unitary. Then  $U^{\otimes m}$  can be implemented up to error  $\varepsilon$  by a Clifford+T circuit using  $O(m + \log(1/\varepsilon))$  T gates and ancillary qubits.<sup>20</sup> Furthermore, no Clifford+T circuit (even with measurements and adaptivity) can use asymptotically fewer T gates.*

Before proving this, let us describe our high-level strategy. We can use [Fact 2.4](#) to reduce to the special case of implementing a diagonal single-qubit unitary  $A := \text{diag}(1, e^{i\theta})$ , by choosing the (irrelevant) global phase to be such that all 3 diagonal unitaries are of this form. For any computational basis input  $|x\rangle$  where  $x \in \{0, 1\}^m$ , we have  $A^{\otimes m}|x\rangle = e^{i\theta \cdot |x|}|x\rangle$ , where  $|x|$  is the Hamming weight of  $x$ . A natural approach is then to first compute the Hamming weight into a second register containing  $\lceil \log m \rceil$  qubits, then apply the corresponding phase controlled on this register. The Hamming weight can be computed using  $O(m)$  gates, as we discuss below. In the second step we need to implement a diagonal unitary on  $\lceil \log m \rceil$  qubits. We can do this using [Theorem 1.2](#), incurring a T-count  $O(\sqrt{m \log(1/\varepsilon)} + \log(1/\varepsilon)) = O(m + \log(1/\varepsilon))$ .

We start by showing how to compute the Hamming weight using  $O(m)$  gates.

**Fact 2.6.** *Let  $U_{\text{Ham}}$  be the unitary such that*

$$U_{\text{Ham}}|x\rangle|y\rangle = |x\rangle|y \oplus |x|\rangle \quad (13)$$

*for all  $x \in \{0, 1\}^m$  and  $y \in \{0, 1\}^{\lceil \log m \rceil}$ , where  $|x| \in \{0, 1, \dots, m-1\}$  is the Hamming weight of  $x$ . Then  $U_{\text{Ham}}$  can be exactly prepared by Clifford+T circuit using  $O(m)$  total gates and ancillary qubits.*

We also remark that [\[BP05\]](#) gives the exact multiplicative complexity of computing the Hamming weight function, which upper bounds the T-count of implementing  $U_{\text{Ham}}$  (using the connection in [Remark 2.2](#)). Their construction has slightly lower T-count than [Fact 2.6](#).

*Proof.* We describe a classical Boolean circuit with  $O(m)$  gates to compute the Hamming weight of an  $m$ -bit string. Here the circuit contains AND, OR, NOT gates with bounded fan-in. Then we can convert this to a reversible classical circuit in the standard way using Toffoli gates and  $O(m)$  ancilla qubits; implementing each Toffoli using  $O(1)$  T gates gives [Fact 2.6](#).

<sup>19</sup>This construction uses  $O(\log(1/\varepsilon))$  Clifford gates.

<sup>20</sup>This construction uses  $O(m \log(1/\varepsilon))$  Clifford gates.

Firstly note that given two  $t$ -bit numbers, we can compute their sum as a  $(t + 1)$ -bit number using  $O(t)$  gates, by the textbook algorithm sequentially computing each bit and the corresponding carry bit. The Boolean circuit to compute the Hamming weight will use the above observation iteratively for  $t = 1, 2, \dots, \log m$  as follows. We view the input as  $m$  1-bit numbers. Then we partition it into  $m/2$  pairs and compute the sum of each pair as a 2-bit number. Continuing in this way, at the  $t$ -th stage we will have  $m/2^t$   $t$ -bit numbers. The total gate complexity is

$$\sum_{t=1}^{\lceil \log m \rceil} \frac{m}{2^t} \cdot O(t) = O(m). \quad (14)$$

Given [Fact 2.6](#), we can now prove [Theorem 1.4](#).

*Proof of Theorem 1.4.* By [Fact 2.4](#), we obtain single-qubit diagonal unitaries  $A, B, C$  such that  $U = AHBHC$ . Thus  $U^{\otimes m} = A^{\otimes m} H^{\otimes m} B^{\otimes m} H^{\otimes m} C^{\otimes m}$ . Therefore it suffices to show how to implement  $A^{\otimes m}$ .

By ignoring the (irrelevant) global phase, we assume  $A = \text{diag}(1, e^{i\theta})$ . Observe that  $A^{\otimes m}|x\rangle = e^{i\theta \cdot |x|}|x\rangle$  for all  $x \in \{0, 1\}^m$ . Let  $K = \lceil \log(m) \rceil$ . We implement  $A^{\otimes m}$  as follows:

1. Start with  $|x\rangle|0^K\rangle$ . We apply  $U_{\text{Ham}}$  from [Fact 2.6](#) to obtain  $|x\rangle||x|\rangle$ .

This uses  $O(m)$   $T$  gates and ancillary qubits by [Fact 2.6](#).

2. Define a  $K$ -qubit diagonal unitary  $D$  such that  $D|c\rangle = e^{i\theta \cdot c}|c\rangle$ . We apply  $D$  on the Hamming weight register and obtain  $e^{i\theta \cdot |x|}|x\rangle||x|\rangle$ .

This uses  $O\left(\sqrt{m \log(1/\varepsilon)} + \log(1/\varepsilon)\right) \leq O(m + \log(1/\varepsilon))$   $T$  gates and ancilla qubits by [Theorem 1.2](#).

3. Finally uncompute the Hamming weight by applying  $U_{\text{Ham}}$  from [Fact 2.6](#) again.

This again uses  $O(m)$   $T$  gates and ancilla qubits by [Fact 2.6](#).

Finally we prove the tightness of [Theorem 1.4](#). On the one hand, the  $\Omega(\log(1/\varepsilon))$   $T$ -count lower bound holds even if  $m = 1$  [[BCHK20](#)]. On the other hand, consider the special case where  $U$  is the  $T$  gate. Then  $U^{\otimes m} = T^{\otimes m}$  and it is impossible to implement it using  $o(m)$   $T$  gates. Otherwise, by gate injection  $|T\rangle^{\otimes m}$  can be approximated with  $T$ -count  $o(m)$ . Plugging this in [Theorem 1.1](#), we have that any  $n$ -qubit state can be approximated up to any constant error with  $T$ -count  $o(\sqrt{2^n})$ , contradicting the lower bound for [Theorem 1.1](#).<sup>21</sup>  $\square$

### 3 Quantum state preparation with optimal T-count

In this section we prove our main result, [Theorem 1.1](#) (restated below).

**Theorem 1.1** (Quantum state preparation with optimal  $T$ -count). *Any  $n$ -qubit state can be prepared up to error  $\varepsilon$  by a Clifford+ $T$  circuit starting with the all-zeros state using*

$$O\left(\sqrt{2^n \log(1/\varepsilon)} + \log(1/\varepsilon)\right) \quad (2)$$

*$T$  gates and ancillas.<sup>22</sup> Furthermore, no Clifford+ $T$  circuit (even with measurements and adaptivity) can use asymptotically fewer  $T$  gates.*

<sup>21</sup>The lower bound for [Theorem 1.1](#) holds in the presence of Pauli postselections, as needed here after gate injection. See details in [Section 4](#).

<sup>22</sup>This construction uses  $O(2^n \log(1/\varepsilon))$  Clifford gates.

**Remark 3.1.** Our overall state preparation approach is similar to the one by Rosenthal [Ros24]. Motivated by the Aaronson-Kuperberg problem [AK07, Aar16], [Ros24] focuses on the number of Boolean oracle calls needed for efficient state preparation. While the context is different, the  $T$ -count of the algorithm presented in [Ros24] is near optimal. To obtain our Theorem 1.1, we combine Rosenthal’s algorithm with the crucial missing pieces Theorem 1.2 and Theorem 1.3. Along the way, we also simplify and improve some of Rosenthal’s analysis: our Lemma 3.2 simplifies and improves [Ros24, Lemma 3.2], which was implicit in [INN<sup>+</sup>22]; our Lemma 3.5 improves the query complexity in [Ros24, Theorem 4.2].

We already know from Theorem 1.2 how to implement an arbitrary diagonal unitary using the claimed number of  $T$  gates and ancillas. For the state preparation task, this lets us specialize without loss of generality to the case where the target state has real amplitudes: after preparing a target state with real entries, we can apply a diagonal unitary to apply any desired complex phases.

As a first step, we show how to synthesize a real-valued target state to within a constant approximation error.

**Lemma 3.2.** *Let  $|\psi\rangle$  be an arbitrary  $n$ -qubit state with real amplitudes. There exists an  $n$ -qubit state  $|\phi\rangle$  with real amplitudes such that  $\langle\phi|\psi\rangle \geq \frac{1}{\sqrt{2}}$  and  $|\phi\rangle = B_2 H^{\otimes n} B_1 H^{\otimes n} |0^n\rangle$  for some  $n$ -qubit Boolean phase oracles  $B_1$  and  $B_2$  (i.e., diagonal unitaries in which all diagonal entries are  $\pm 1$ ).*

Results similar to Lemma 3.2 are obtained in [BBC<sup>+</sup>19, INN<sup>+</sup>22, Ros24], but our analysis is arguably simpler and achieves better bounds.<sup>23</sup> The proof of Lemma 3.2 is given in Subsection 3.1.

To implement the Boolean phase oracles in Lemma 3.2, one can apply Lemma 2.1 directly, along with the phase kickback trick [CEMM98]. This is stated as the following Fact 3.3.

**Fact 3.3.** *Let  $B$  be an  $n$ -qubit Boolean phase oracle. Then  $B$  can be prepared exactly by a Clifford+ $T$  circuit using  $O(\sqrt{2^n})$   $T$  gates and ancillary qubits.<sup>24</sup>*

As a result of Fact 3.3, we can prepare the approximation  $|\phi\rangle$  from Lemma 3.2 using only  $O(\sqrt{2^n})$   $T$  gates and ancilla qubits. We see that Lemma 3.2 already establishes the special case of Theorem 1.1 where the error  $\varepsilon$  is a sufficiently large constant.

To handle the small error case, we use the fact that Lemma 3.2 can be used to approximate *any* quantum state  $|\psi\rangle$ . Hence we can reduce the approximation error by approximating the difference state  $|\psi\rangle - |\phi\rangle$ . Continuing iteratively in this way, we are able to reduce the error below any target value, as in [Ros24]. This idea is formalized in Lemma 3.4, which is proved in Subsection 3.2.

**Lemma 3.4.** *Let  $|\psi\rangle$  be an arbitrary  $n$ -qubit state with real amplitudes. For any integer  $T \geq 1$ , there exist  $\zeta \in [\frac{\sqrt{2}-1}{2}, \frac{1}{\sqrt{2}}]$  and  $n$ -qubit states  $|\psi_0\rangle, |\psi_1\rangle, \dots, |\psi_{T-1}\rangle$  such that*

$$\left\| |\psi\rangle - \zeta \cdot \sum_{k=0}^{T-1} 2^{-k/2} \cdot |\psi_k\rangle \right\| \leq \frac{1}{\sqrt{2}-1} \cdot 2^{-T/4}, \quad (15)$$

<sup>23</sup>The constant  $1/\sqrt{2}$  in this lemma is optimal, as evidenced by the cat state  $|\psi\rangle = \frac{1}{\sqrt{2}}(|0^n\rangle + |1^n\rangle)$ .

<sup>24</sup>This construction uses  $O(2^n)$  Clifford gates.

where each  $|\psi_k\rangle = B_2^{(k)} H^{\otimes n} B_1^{(k)} H^{\otimes n} |0^n\rangle$  for some  $n$ -qubit Boolean phase oracles  $B_1^{(k)}$  and  $B_2^{(k)}$ .

Following the strategy from [Ros24], we then use the linear combination of unitaries (LCU) method [WC12, BCK15] to prepare the weighted sum of states in Lemma 3.4, flagged by an additional ancilla. We carefully tune the amplitude of the flagged part of the state to enable the use of exact amplitude amplification [Gro98, BHMT02] to get rid of the junk part and obtain the desired  $\varepsilon$  approximation.

Our implementation of this strategy gives the following Lemma 3.5.

**Lemma 3.5.** *Let  $\varepsilon \in (0, 1/2]$  and  $|\psi\rangle$  be an arbitrary  $n$ -qubit state with real amplitudes. There exist some  $t = \log \log(1/\varepsilon) + O(1)$  and an  $n$ -qubit normalized state  $|\phi\rangle$  such that  $\| |\phi\rangle - |\psi\rangle \| \leq \varepsilon$  and  $|0^t\rangle_{\mathcal{A}} |\phi\rangle_{\mathcal{B}} = R_1 R_2 R_1 R_2 V |0^{n+t}\rangle_{\mathcal{AB}}$  where*

$$R_1 = V \left( 2 |0^{n+t}\rangle \langle 0^{n+t}| - I \right) V^\dagger, \quad (16)$$

$$R_2 = \left( 2 |0^t\rangle \langle 0^t| - I \right)_{\mathcal{A}} \otimes I_{\mathcal{B}}, \quad (17)$$

and  $V$  is a quantum circuit with the following structure in order:

- one layer of single-qubit gates on  $\mathcal{A}$ ;
- Hadamard on  $\mathcal{B}$ ;
- a Boolean phase oracle on  $\mathcal{AB}$ ;
- Hadamard on  $\mathcal{B}$ ;
- another Boolean phase oracle on  $\mathcal{AB}$ ;
- another layer of single-qubit gates on  $\mathcal{A}$ .

The proof of Lemma 3.5 is deferred to Subsection 3.3. We now prove Theorem 1.1.

*Proof of Theorem 1.1.* Using Theorem 1.2 to correct phases, we assume without loss of generality that the target state  $|\psi\rangle$  has real amplitudes. Then it suffices to analyze the  $T$ -count and ancilla count for the circuit in Lemma 3.5. To this end, we analyze the components separately, each of which appears constant number of times in the whole circuit in Lemma 3.5. Assume  $\varepsilon \leq 1/2$ .

- The reflection  $2 |0^{n+t}\rangle \langle 0^{n+t}| - I$  used in  $R_1$  is a Boolean phase oracle and can be constructed by Fact 3.3 with  $O\left(\sqrt{2^{n+t}}\right) = O\left(\sqrt{2^n \log(1/\varepsilon)}\right)$   $T$  gates and ancillas.<sup>25</sup>
- $R_2$  is the same reflection (with a smaller scale) as above, which is also handled by Fact 3.3.
- The Hadamard layer used in  $V$  is Clifford and requires no  $T$  gates or ancilla.
- The Boolean phase oracle on  $\mathcal{AB}$  used in  $V$  can be again prepared using Fact 3.3 with  $O\left(\sqrt{2^{n+t}}\right) = O\left(\sqrt{2^n \log(1/\varepsilon)}\right)$   $T$  gates and ancillas.
- The layer of single-qubit gates on  $\mathcal{A}$  used in  $V$  can be synthesized by Theorem 1.3 since  $\mathcal{A}$  only has  $t = \log \log(1/\varepsilon) + O(1)$  qubits. This uses  $O(\log(1/\varepsilon))$   $T$  gates and ancillas by Theorem 1.3. The approximation error does not blow up since this is only used a constant number of times in total.

Summing over costs of the above components establishes Theorem 1.1. The tightness of Theorem 1.1 is deferred as Theorem 4.1 to be proved in Section 4.  $\square$

<sup>25</sup>This reflection can be implemented with much lower cost. We use Lemma 2.1 as a loose upper bound since it does not affect the asymptotic bound in the end.



### 3.1 Approximating a state to constant error: [Lemma 3.2](#)

In this section we describe how to approximate a state to constant error ([Lemma 3.2](#)) and in the next section we discuss the finer approximation ([Lemma 3.4](#)).

To get the coarse approximation of [Lemma 3.2](#), we take our target state and apply a random diagonal unitary with  $\pm 1$  entries followed by a binary Fourier transform (i.e., the Hadamard matrix). This does a reasonable job of flattening the amplitudes of the state, by which we mean that most of the amplitudes are now of similar magnitude. We then apply a diagonal unitary to make all the phases  $+1$ , which gives a state that has constant overlap with the uniform superposition  $|+\rangle$ . To implement this strategy we use the standard Khintchine inequality.

**Fact 3.6** (Khintchine inequality [[Haa81](#), [Wik24b](#)]). *Let  $N \geq 1$  be an integer and  $\beta_1, \dots, \beta_N$  be arbitrary complex numbers satisfying  $\sum_{j \in [N]} |\beta_j|^2 = 1$ . Then*

$$\frac{1}{\sqrt{2}} \leq \mathbb{E}_{x \sim \{\pm 1\}^N} \left[ \left\| \sum_{j \in [N]} \beta_j \cdot x_j \right\| \right] \leq 1. \quad (18)$$

Using [Fact 3.6](#), we show how to flatten the amplitudes of a quantum state.

**Lemma 3.7.** *Let  $|\psi\rangle$  be an arbitrary  $n$ -qubit state. Let  $F = H^{\otimes n}$  or more generally, let  $F$  be any  $n$ -qubit unitary such that  $|F_{ij}| = 1/\sqrt{2^n}$  for  $i, j \in \{0, 1\}^n$ . Then*

$$\mathbb{E}_{x \sim \{\pm 1\}^{\{0,1\}^n}} [\|F \text{diag}(x)|\psi\rangle\|_1] \geq \frac{\sqrt{2^n}}{\sqrt{2}}. \quad (19)$$

*Proof.* Write  $|\psi\rangle = \sum_{j \in \{0,1\}^n} \beta_j |j\rangle$  with  $\sum_{j \in \{0,1\}^n} |\beta_j|^2 = 1$ . Then

$$\mathbb{E}_x [\|F \text{diag}(x)|\psi\rangle\|_1] = \mathbb{E}_x \left[ \left\| \sum_{i,j} F_{ij} x_j \beta_j |i\rangle \right\|_1 \right] \quad (20)$$

$$= \mathbb{E}_x \left[ \sum_i \left| \sum_j F_{ij} x_j \beta_j \right| \right] = \sum_i \mathbb{E}_x \left[ \left| \sum_j F_{ij} x_j \beta_j \right| \right], \quad (21)$$

where the last equality used the linearity of expectation. By assumption,  $F_{ij} = f_{ij}/\sqrt{2^n}$ , where  $f_{ij}$  is a complex number of unit modulus. Then  $f_{ij}\beta_j$  satisfies the assumption of [Fact 3.6](#) since  $\sum_{j \in \{0,1\}^n} |f_{ij}\beta_j|^2 = 1$ , and hence

$$\mathbb{E}_x [\|F \text{diag}(x)|\psi\rangle\|_1] = \frac{1}{\sqrt{2^n}} \sum_i \mathbb{E}_x \left[ \left| \sum_j f_{ij} x_j \beta_j \right| \right] \geq \frac{1}{\sqrt{2}} \sum_i \frac{1}{\sqrt{2^n}} = \frac{\sqrt{2^n}}{\sqrt{2}} \quad (22)$$

as claimed.  $\square$

After flattening, the state becomes essentially a uniform superposition with different phases. Moreover, in the case of interest where  $|\psi\rangle$  is assumed to have real amplitudes, these phases are simply  $\pm 1$ . Then we can correct the Boolean phases and reverse the flattening procedure to obtain an approximation of the original state  $|\psi\rangle$ .

*Proof of [Lemma 3.2](#).* Let  $|\psi\rangle$  be an  $n$ -qubit state with real amplitudes. By [Lemma 3.7](#), there exists a Boolean phase oracle  $B_2$  such that  $|\tilde{\psi}\rangle := H^{\otimes n} B_2 |\psi\rangle$  has  $\ell_1$  norm

$$\|\tilde{\psi}\|_1 \geq \frac{\sqrt{2^n}}{\sqrt{2}}. \quad (23)$$

Since  $|\psi\rangle$  has real amplitudes, so does  $|\tilde{\psi}\rangle$ . Let  $B_1$  be the Boolean phase oracle such that  $\langle x|B_1|x\rangle = \text{sign}(\langle x|\tilde{\psi}\rangle)$  for all  $x \in \{0, 1\}^n$ .

Letting  $|\phi\rangle = B_2 H^{\otimes n} B_1 H^{\otimes n} |0^n\rangle$ , we have

$$\langle \psi|\phi\rangle = \langle \psi|B_2 H^{\otimes n} B_1 H^{\otimes n}|0\rangle = \langle \tilde{\psi}|B_1 H^{\otimes n}|0\rangle = \frac{1}{\sqrt{2^n}} \cdot \|\tilde{\psi}\|_1 \geq \frac{1}{\sqrt{2}}, \quad (24)$$

as desired.  $\square$

### 3.2 Approximating a state to $\varepsilon$ error: Lemma 3.4

Now we discuss how to reduce the approximation error. We prove the following statement and show that Lemma 3.4 follows from it. The proof follows the reasoning from [Ros24, Lemma 3.1], but completes an edge case (see “Case  $T > j^*$ ” in the proof) missed in the previous analysis.

**Lemma 3.8.** *Let  $\alpha \in (0, 1]$  be a real number. For any  $n$ -qubit state  $|\phi\rangle$ , let  $U_\phi$  be an  $n$ -qubit unitary that satisfies  $\Re\langle\phi|U_\phi|0^n\rangle \geq \alpha$ .<sup>26</sup> Define*

$$\gamma = \min\left\{\alpha, \frac{1}{\sqrt{2}}\right\} \quad \text{and} \quad \beta = \sqrt{1 - \gamma^2}. \quad (25)$$

*Then for every  $n$ -qubit state  $|\psi\rangle$  and integer  $T \geq 1$ , there exist  $\gamma \cdot (1 - \beta) \leq \zeta \leq \gamma$  and  $|\psi_0\rangle, |\psi_1\rangle, \dots, |\psi_{T-1}\rangle$  such that*

$$\left\| |\psi\rangle - \zeta \cdot \sum_{k=0}^{T-1} \beta^k \cdot U_{\psi_k} |0^n\rangle \right\| \leq \frac{\gamma}{1 - \beta} \cdot \beta^{T/2}. \quad (26)$$

*Proof.* Since  $\Re\langle\phi|U_\phi|0^n\rangle \geq \alpha \geq \gamma$ , we have

$$\| |\phi\rangle - \gamma \cdot U_\phi |0^n\rangle \| = \sqrt{1 + \gamma^2 - 2\gamma \cdot \Re\langle\phi|U_\phi|0^n\rangle} \leq \sqrt{1 - \gamma^2} = \beta. \quad (27)$$

Since  $\gamma > 0$ , we know  $\beta < 1$ . Thus our task of approximating  $|\psi\rangle$  is now reduced to that of approximating  $|\psi\rangle - \gamma \cdot U_\psi |0^n\rangle$  which has a smaller length. We will show that (26) follows by iteratively applying (27).

Formally, define  $|\psi_0\rangle = |\tilde{\psi}_0\rangle = |\psi\rangle$  and recursively for  $j \geq 1$

$$|\tilde{\psi}_j\rangle = |\psi\rangle - \gamma \cdot \sum_{k=0}^{j-1} \beta^k \cdot U_{\psi_k} |0^n\rangle \quad (28)$$

and

$$|\psi_j\rangle = \frac{|\tilde{\psi}_j\rangle}{\| |\tilde{\psi}_j\rangle \|} \quad (29)$$

until some  $j^* \geq 1$  for which  $|\tilde{\psi}_{j^*}\rangle$  becomes a zero vector. Set  $j^* = +\infty$  if  $|\tilde{\psi}_j\rangle$  is never zero.

---

<sup>26</sup>For a complex number  $v$ ,  $\Re v$  equals its real part.

**Case  $T \leq j^*$ .** We first handle the case where  $T \leq j^*$ . Below we give an inductive proof that

$$\left\| |\tilde{\psi}_j\rangle \right\| \leq \beta^j \quad (30)$$

for each  $1 \leq j \leq j^*$ . Then we simply set  $\zeta = \gamma$  and (26) follows from (30) as  $\gamma, \beta \in [0, 1]$ .

To establish (30), first note that the base case  $j = 1$  follows directly from (27). For  $j \geq 2$ , we have

$$\left\| |\tilde{\psi}_j\rangle \right\|^2 = \left\| |\tilde{\psi}_{j-1}\rangle - \gamma\beta^{j-1} \cdot U_{\psi_{j-1}}|0^n\rangle \right\|^2 \quad (31)$$

$$= \left\| \left\| |\tilde{\psi}_{j-1}\rangle \right\| \cdot |\psi_{j-1}\rangle - \gamma\beta^{j-1} \cdot U_{\psi_{j-1}}|0^n\rangle \right\|^2 \quad (32)$$

$$= \left\| |\tilde{\psi}_{j-1}\rangle \right\|^2 + \gamma^2\beta^{2j-2} - 2\gamma\beta^{j-1} \left\| |\tilde{\psi}_{j-1}\rangle \right\| \cdot \Re\langle \psi_{j-1} | U_{\psi_{j-1}} | 0^n \rangle \quad (33)$$

$$\leq \left\| |\tilde{\psi}_{j-1}\rangle \right\|^2 + \gamma^2\beta^{2j-2} - 2\gamma^2\beta^{j-1} \left\| |\tilde{\psi}_{j-1}\rangle \right\|. \quad (\text{since } \Re\langle \psi_{j-1} | U_{\psi_{j-1}} | 0^n \rangle \geq \gamma)$$

Since  $\gamma \leq \frac{1}{\sqrt{2}}$  and  $\beta = \sqrt{1 - \gamma^2}$ , we know  $\gamma \leq \beta$ . Then by the induction hypothesis we have

$$\left\| |\tilde{\psi}_j\rangle \right\|^2 \leq \max_{0 \leq x \leq \beta^{j-1}} x^2 + \gamma^2\beta^{2j-2} - 2\gamma^2\beta^{j-1}x \quad (34)$$

$$= \max \left\{ \gamma^2\beta^{2j-2}, \beta^{2j-2} - \gamma^2\beta^{2j-2} \right\} \quad (35)$$

$$= \max \left\{ \gamma^2\beta^{2j-2}, \beta^{2j} \right\} = \beta^{2j} \quad (\text{since } \gamma^2 + \beta^2 = 1 \text{ and } \gamma \leq \beta)$$

as desired.

**Case  $T > j^*$ .** Now we turn to the case where  $T > j^*$ . For each  $j \geq j^* \geq 1$ , define  $|\psi_j\rangle = |\psi_{j \bmod j^*}\rangle$ . Let  $t = \left\lfloor \frac{T}{j^*} \right\rfloor$  and  $m = T \bmod j^*$ . Then

$$\sum_{k=0}^{T-1} \beta^k \cdot U_{\psi_k} | 0^n \rangle = \sum_{\ell=0}^{t-1} \beta^{\ell \cdot j^*} \sum_{k=0}^{j^*-1} \beta^k \cdot U_{\psi_k} | 0^n \rangle + \sum_{k=0}^m \beta^{t \cdot j^* + k} \cdot U_{\psi_k} | 0^n \rangle \quad (36)$$

$$= \sum_{\ell=0}^{t-1} \frac{\beta^{\ell \cdot j^*}}{\gamma} \cdot |\psi\rangle + \sum_{k=0}^m \beta^{t \cdot j^* + k} \cdot U_{\psi_k} | 0^n \rangle \quad (\text{by the definition of } j^*)$$

$$= \frac{1 - \beta^{t \cdot j^*}}{\gamma \cdot (1 - \beta^{j^*})} \cdot |\psi\rangle + \sum_{k=0}^m \beta^{t \cdot j^* + k} \cdot U_{\psi_k} | 0^n \rangle. \quad (37)$$

Define  $\zeta = \frac{\gamma \cdot (1 - \beta^{j^*})}{1 - \beta^{t \cdot j^*}}$ . Then we have

$$\gamma \geq \zeta \geq \gamma \cdot (1 - \beta^{j^*}) \geq \gamma \cdot (1 - \beta) \quad (38)$$

as claimed. In addition,

$$\text{LHS of (26)} = \left\| \zeta \cdot \sum_{k=0}^m \beta^{t \cdot j^* + k} \cdot U_{\psi_k} | 0^n \rangle \right\| \leq \zeta \cdot \sum_{k=0}^m \beta^{t \cdot j^* + k} \quad (39)$$

$$\leq \zeta \cdot \sum_{k=0}^{j^*-1} \beta^{t \cdot j^* + k} = \zeta \cdot \frac{1 - \beta^{j^*}}{1 - \beta} \cdot \beta^{t \cdot j^*} \quad (\text{since } m = T \bmod j^* \leq j^* - 1)$$

$$\begin{aligned}
&\leq \gamma \cdot \frac{1 - \beta^{j^*}}{1 - \beta} \cdot \beta^{t \cdot j^*} \leq \frac{\gamma}{1 - \beta} \cdot \beta^{t \cdot j^*} && (\text{since } \zeta \leq \gamma) \\
&= \frac{\gamma}{1 - \beta} \cdot \beta^{\lfloor \frac{T}{j^*} \rfloor \cdot j^*} && (\text{since } t = \lfloor \frac{T}{j^*} \rfloor) \\
&\leq \frac{\gamma}{1 - \beta} \cdot \beta^{T/2}, && (\text{since } T \geq j^* \text{ and } \lfloor x \rfloor \geq x/2 \text{ for } x \geq 1)
\end{aligned}$$

which verifies (26).  $\square$

Given Lemma 3.8, we immediately obtain Lemma 3.4.

*Proof of Lemma 3.4.* In light of Lemma 3.8, we set  $\alpha = \frac{1}{\sqrt{2}}$  and design  $U_\phi$  as the  $BHBBH$  circuit from Lemma 3.2. Note that all states used in Lemma 3.8 have real amplitudes since we start with  $|\psi\rangle$  that has real amplitudes. Therefore circuits  $U_\phi$  are all well-defined.  $\square$

### 3.3 Flagging and exact amplitude amplification: Lemma 3.5

In this Section we describe the circuit that prepares an  $n$ -qubit state with optimal  $T$ -count, establishing Lemma 3.5.

We first use the linear combination of unitaries technique to deduce a circuit that prepares the approximating state from Lemma 3.4.

**Lemma 3.9.** *Let  $\varepsilon \in (0, 1/2]$  and  $|\psi\rangle$  be an arbitrary  $n$ -qubit state with real amplitudes. There exist some  $t = \log \log(1/\varepsilon) + O(1)$ , an  $n$ -qubit normalized state  $|\phi\rangle$ , and an  $(n+t)$ -qubit unnormalized state  $|\tau\rangle$  such that*

1.  $\| |\phi\rangle - |\psi\rangle \| \leq \varepsilon$ .
2.  $(|0^t\rangle \langle 0^t| \otimes I_n) |\tau\rangle = 0$ , where  $I_n$  is the  $n$ -qubit identity operator.
3. For some  $\xi \geq \frac{4\sqrt{2}-4}{5} \geq 0.33$ , the normalized state  $\xi|0^t\rangle_{\mathcal{A}}|\phi\rangle_{\mathcal{B}} + |\tau\rangle_{\mathcal{AB}}$  can be prepared as follows, starting from the initial state  $|0^t\rangle_{\mathcal{A}}|0^n\rangle_{\mathcal{B}}$ :
  - (a) Apply one layer of single-qubit gates on  $\mathcal{A}$ .
  - (b) Apply Hadamard on  $\mathcal{B}$ .
  - (c) Apply a Boolean phase oracle on  $\mathcal{AB}$ .
  - (d) Apply Hadamard on  $\mathcal{B}$ .
  - (e) Apply another Boolean phase oracle on  $\mathcal{AB}$ .
  - (f) Apply another layer of single-qubit gates on  $\mathcal{A}$ .

*Proof.* Let  $T = 2^t$  and  $\beta = \frac{1}{\sqrt{2}}$ . By Lemma 3.4, there exist states  $|\psi_0\rangle, |\psi_1\rangle, \dots, |\psi_{T-1}\rangle$  and Boolean phase oracles  $B_1^{(k)}, B_2^{(k)}$  for  $0 \leq k \leq T-1$ , such that  $|\psi_k\rangle = B_2^{(k)} H^{\otimes n} B_1^{(k)} H^{\otimes n} |0^n\rangle$  and

$$\left\| |\psi\rangle - \zeta \cdot \sum_{k=0}^{T-1} \beta^k |\psi_k\rangle \right\| \leq \frac{1}{\sqrt{2}-1} \cdot 2^{-T/4} \leq \frac{\varepsilon}{10}, \quad (40)$$

where  $\zeta \in \left[ \frac{\sqrt{2}-1}{2}, \frac{1}{\sqrt{2}} \right]$  and for the last inequality we used the fact that  $t = \log \log(1/\varepsilon) + O(1)$ .

Define  $\ell = \left\| \zeta \cdot \sum_{k=0}^{T-1} \beta^k |\psi_k\rangle \right\|$  and

$$|\phi\rangle = \frac{\zeta}{\ell} \cdot \sum_{k=0}^{T-1} \beta^k |\psi_k\rangle \quad \text{and} \quad \xi = \frac{\ell \cdot (1 - \beta)}{\zeta \cdot (1 - \beta^T)}. \quad (41)$$

To get the circuit for Item 3 we use the linear combination of unitaries technique:

- In [Item 3a](#) we apply a layer of single-qubit gates on  $\mathcal{A}$  that transforms  $|0^t\rangle_{\mathcal{A}}$  to

$$\sqrt{\frac{1-\beta}{1-\beta^T}} \prod_{\ell=0}^{t-1} (|0\rangle + \beta^{2^\ell/2} |1\rangle)_{\mathcal{A}} = \sqrt{\frac{1-\beta}{1-\beta^T}} \sum_{k=0}^{T-1} \beta^{k/2} |k\rangle_{\mathcal{A}}. \quad (42)$$

- [Items 3b](#) to [3e](#) correspond to, controlled on index  $k$  at register  $\mathcal{A}$ ,  $B_2^{(k)} H^{\otimes n} B_1^{(k)} H^{\otimes n}$  is applied on  $|0^n\rangle_{\mathcal{B}}$  to obtain

$$\sqrt{\frac{1-\beta}{1-\beta^T}} \sum_{k=0}^{T-1} \beta^{k/2} |k\rangle_{\mathcal{A}} \otimes |\psi_k\rangle_{\mathcal{B}}. \quad (43)$$

Here we use the fact that, since each  $B_1^{(k)}, B_2^{(k)}$  are Boolean phase oracles, their controlled versions are Boolean phase oracles on registers  $\mathcal{AB}$ .

- In [Item 3f](#) we apply the inverse of the layer of single-qubit gates from [Item 3a](#), which gives

$$|0^t\rangle_{\mathcal{A}} \cdot \frac{1-\beta}{1-\beta^T} \sum_{k=0}^{T-1} \beta^k |\psi_k\rangle_{\mathcal{B}} + |\tau\rangle_{\mathcal{AB}} = \xi |0^t\rangle_{\mathcal{A}} |\phi\rangle_{\mathcal{B}} + |\tau\rangle_{\mathcal{AB}} \quad (44)$$

by (41). From the above we see that [Item 2](#) is satisfied.

It remains to prove [Item 1](#) and verify the range of  $\xi$  in [Item 3](#). By (40), we have  $\| |\psi\rangle - \ell \cdot |\phi\rangle \| \leq \frac{\beta^{T/2}}{\sqrt{2}-1}$  and hence

$$1 - \frac{\beta^{T/2}}{\sqrt{2}-1} \leq \ell \leq 1 + \frac{\beta^{T/2}}{\sqrt{2}-1}. \quad (45)$$

Therefore  $\| |\phi\rangle - |\psi\rangle \| \leq \| |\psi\rangle - \ell \cdot |\phi\rangle \| + |1 - \ell| \leq \frac{2\beta^{T/2}}{\sqrt{2}-1} \leq \varepsilon$ , which verifies [Item 1](#). Plugging (40) and (45) into (41), we have

$$\begin{aligned} \xi &\geq \frac{(1 - \beta^{T/2}) \cdot (1 - \beta)}{\zeta \cdot (1 - \beta^T)} = \frac{1 - \beta}{\zeta \cdot (1 + \beta^{T/2})} \geq \frac{1 - \beta}{\zeta \cdot (1 + \frac{\varepsilon}{2})} && (\text{since } \beta^{T/2} \leq \frac{\varepsilon}{2}) \\ &\geq \frac{1 - \beta}{\frac{1}{\sqrt{2}} \cdot \frac{5}{4}} && (\text{since } \varepsilon \leq 1/2 \text{ and } \zeta \leq \frac{1}{\sqrt{2}}) \\ &= \frac{4 \cdot (\sqrt{2} - 1)}{5} \geq 0.33 && (\text{since } \beta = \frac{1}{\sqrt{2}}) \end{aligned}$$

as claimed.  $\square$

[Item 3](#) in [Lemma 3.9](#) is a coarse flagging scheme where the approximation  $|\phi\rangle$  of the target  $|\psi\rangle$  is flagged with some undetermined amplitude  $\xi \geq 0.33$ . In order to enable the use of exact amplitude amplification, it will be convenient to fix this amplitude to a known value.

**Corollary 3.10.** *[Lemma 3.9](#) holds with the replacement  $\xi = \sin(\frac{\pi}{10})$ .*

*Proof.* By [Lemma 3.9](#),  $\xi \geq 0.33 \geq 0.31 \geq \sin(\frac{\pi}{10})$ . Therefore there exists a single-qubit unitary  $G$  such that

$$G|0\rangle = \frac{\sin(\pi/10)}{\xi} \cdot |0\rangle + \sqrt{1 - \frac{\sin^2(\pi/10)}{\xi^2}} \cdot |1\rangle. \quad (46)$$

Recall that [Lemma 3.9](#) constructs the state  $\xi|0^t\rangle_{\mathcal{A}}|\phi\rangle_{\mathcal{B}} + |\tau\rangle_{\mathcal{AB}}$ . Now add an extra ancilla in  $\mathcal{A}$  and apply  $G$  to it. This gives the state  $\sin(\frac{\pi}{10}) \cdot |0^{t+1}\rangle_{\mathcal{A}}|\phi\rangle_{\mathcal{B}} + |\tau'\rangle_{\mathcal{AB}}$ , where  $(|0^{t+1}\rangle\langle 0^{t+1}| \otimes I_n)|\tau'\rangle = 0$ . Finally, note that the additional gate  $G$  can be absorbed into the layer of single-qubit gates applied in [Item 3f](#) of the circuit described in [Lemma 3.9](#). We have shown that we can safely set  $\xi = \sin(\frac{\pi}{10})$  in [Lemma 3.9](#).  $\square$

The constant  $\sin(\frac{\pi}{10})$  is chosen because it allows us to prepare  $|\phi\rangle$  exactly after two rounds of amplitude amplification [[Gro98](#), [BHMT02](#)]. We now complete the proof of [Lemma 3.5](#).

*Proof of Lemma 3.5.* Let  $V$  be the circuit in [Corollary 3.10](#) to exactly prepare

$$|\rho\rangle := \sin\left(\frac{\pi}{10}\right) \cdot |0^t\rangle_{\mathcal{A}}|\phi\rangle_{\mathcal{B}} + |\tau\rangle_{\mathcal{AB}}. \quad (47)$$

By [Lemma 3.9](#) and [Corollary 3.10](#),  $V$  has the structure claimed in [Lemma 3.5](#).

Observe that  $R_1 = V(2|0^{n+t}\rangle\langle 0^{n+t}| - I)V^\dagger$  is the reflection along the state  $|\rho\rangle$ . Also recall that  $R_2 = (2|0^t\rangle\langle 0^t| - I)_{\mathcal{CA}} \otimes I_{\mathcal{B}}$  is the reflection on register  $\mathcal{CA}$  along  $|0^t\rangle$ . Using the standard analysis of amplitude amplification [[Gro98](#), [BHMT02](#)] and the fact that  $\sin(5 \cdot \frac{\pi}{10}) = 1$ , we arrive at

$$R_1 R_2 R_1 R_2 V |0^{n+t}\rangle = (-R_1 R_2)^2 |\rho\rangle = |0^t\rangle |\phi\rangle, \quad (48)$$

as desired.  $\square$

## 4 Lower bounds

In this section we prove  $T$ -count lower bounds matching the upper bounds in [Theorem 1.1](#) and [Theorem 1.2](#). Our lower bounds apply in a very general model of adaptive Clifford+ $T$  circuits that we now describe.

**Adaptive Clifford+ $T$  Circuits.** Consider a quantum computation  $\mathcal{A}$  with an  $n$ -qubit input register as well as an  $a$ -qubit ancilla register that is initialized in the all-zero state. The computation proceeds via a sequence of Clifford gates,  $T$  gates, and single-qubit measurements in the computational basis. These operations may act on any of the  $n + a$  qubits. Moreover, each Clifford or  $T$  gate may be classically controlled on the outcomes of measurements that have occurred previously.

Let  $r \in \{0, 1\}^*$  be a binary string that describes the measurement outcomes in order obtained during the course of the computation.<sup>27</sup> For each  $r$ , there is a Kraus operator  $K_r$  which describes the corresponding sequence of  $T$  gates, Clifford gates, and single-qubit projectors  $|r_i\rangle\langle r_i|$  for  $i = 1, 2, \dots$ . The adaptive Clifford+ $T$  circuit implements a quantum channel which maps the  $n$ -qubit input state  $\rho$  to an  $(n + a)$ -qubit output state:

$$\mathcal{A}(\rho) = \sum_r K_r (\rho \otimes |0^a\rangle\langle 0^a|) K_r^\dagger. \quad (49)$$

For example, in the simplest case where there are no measurements and the circuit simply applies an  $(n + a)$ -qubit Clifford+ $T$  unitary  $U$ ,  $\mathcal{A}(\rho) = U(\rho \otimes |0^a\rangle\langle 0^a|)U^\dagger$ . As another example, if the circuit applies a unitary  $U$  and then measures the first qubit, then we have  $\mathcal{A}(\rho) = \sum_{r \in \{0, 1\}} (|r\rangle\langle r| \otimes I)U(\rho \otimes |0^a\rangle\langle 0^a|)U^\dagger(|r\rangle\langle r| \otimes I)$ .

---

<sup>27</sup>For maximum generality, we do not assume that the number of single-qubit measurements is bounded. If we knew there were exactly  $m$  measurements, then  $r$  would be an  $m$ -bit string.



We can view the output state  $\mathcal{A}(\rho)$  as a probabilistic mixture of the states

$$\sigma_r(\rho) \equiv \frac{K_r(\rho \otimes |0^a\rangle\langle 0^a|) K_r^\dagger}{\text{Tr}(K_r^\dagger K_r(\rho \otimes |0^a\rangle\langle 0^a|))}, \quad (50)$$

where  $\sigma_r(\rho)$  occurs with probability  $p_r(\rho) \equiv \text{Tr}(K_r^\dagger K_r(\rho \otimes |0^a\rangle\langle 0^a|))$ .

**T-Count.** Let  $\mathcal{T}_r$  be the number of  $T$  gates used by the circuit, conditioned on measurement outcome  $r$ . The maximal number of  $T$  gates used by  $\mathcal{A}$  is

$$\mathcal{T}_{\max}(\mathcal{A}) = \sup_{r \in \{0,1\}^*} \mathcal{T}_r. \quad (51)$$

We also consider the expected  $T$ -count of  $\mathcal{A}$  starting with the all-zeros input state:

$$\mathcal{T}^0(\mathcal{A}) = \sum_r p_r(|0^n\rangle\langle 0^n|) \mathcal{T}_r. \quad (52)$$

Finally, we define the expected  $T$ -count of  $\mathcal{A}$  for the worst-case input state as follows

$$\mathcal{T}(\mathcal{A}) = \sup_{\rho} \sum_r p_r(\rho) \mathcal{T}_r, \quad (53)$$

where  $\rho$  is any  $n$ -qubit mixed state and  $\sum_r p_r(\rho) \mathcal{T}_r$  is the expected number of  $T$  gates used in  $\mathcal{A}(\rho)$ . For example, if the circuit is unitary and has no measurements, then all three of these measures equal the  $T$ -count of the unitary.

Note that clearly we have

$$\mathcal{T}^0(I \otimes \mathcal{A}) = \mathcal{T}^0(\mathcal{A}) \leq \mathcal{T}(I \otimes \mathcal{A}) = \mathcal{T}(\mathcal{A}) \leq \mathcal{T}_{\max}(\mathcal{A}) = \mathcal{T}_{\max}(I \otimes \mathcal{A}). \quad (54)$$

Here we used the fact that, if  $\mathcal{A}$  is an adaptive Clifford+ $T$  circuit with an  $n$ -qubit input register, then so is  $\mathcal{A}' = I \otimes \mathcal{A}$  where the first register can have any number of qubits  $n' \geq 0$  and the second register is  $n$  qubits (in particular,  $\mathcal{A}'$  is just  $\mathcal{A}$  applied to the second register).

In the following we will prove lower bounds on the expected  $T$ -count  $\mathcal{T}^0(\mathcal{A})$  (for state preparation) and  $\mathcal{T}(\mathcal{A})$  (for unitary synthesis). These in turn imply lower bounds on the maximal  $T$ -count due to (54).

**State Preparation.** We say that  $\mathcal{A}$  prepares state  $|\psi\rangle$  to within error  $\varepsilon$  in trace distance if

$$\frac{1}{2} \cdot \|\mathcal{A}(|0^n\rangle\langle 0^n|) - |\psi\rangle\langle\psi| \otimes |0^a\rangle\langle 0^a|\|_1 \leq \varepsilon, \quad (55)$$

where  $\|\cdot\|_1$  here is the trace norm.

Note that one could instead consider a variant of the approximate state preparation task where we do not require that the ancillas are returned to the all-zeros state. In this case we would require the following closeness condition, which might seem to be less stringent than (55):

$$\frac{1}{2} \cdot \|\text{Tr}_{\text{anc}}(\mathcal{A}(|0^n\rangle\langle 0^n|)) - |\psi\rangle\langle\psi|\|_1 \leq \varepsilon, \quad (56)$$

where  $\text{Tr}_{\text{anc}}$  denotes the partial trace over the  $a$ -qubit ancilla register. However it is not hard to see that we can work with the closeness condition (55) without loss of generality. Indeed, if an adaptive Clifford+ $T$  circuit  $\mathcal{A}$  satisfies (56), then we can construct another

adaptive Clifford+ $T$  circuit  $\mathcal{B}$  that satisfies (55) and with the same expected  $T$ -count  $\mathcal{T}^0(\mathcal{A}) = \mathcal{T}^0(\mathcal{B})$ . In particular  $\mathcal{B}$  is obtained by first applying  $\mathcal{A}$ , then measuring the ancilla register in the computational basis, then applying a Pauli gate on the ancilla register which is controlled on the measurement outcome  $z \in \{0, 1\}^a$ , that maps  $|z\rangle \rightarrow |0^a\rangle$ . Then  $\mathcal{B}(|0^n\rangle\langle 0^n|) = \text{Tr}_{\text{anc}}(\mathcal{A}(|0^n\rangle\langle 0^n|)) \otimes |0^a\rangle\langle 0^a|$  and the closeness condition (55) holds for  $\mathcal{B}$ .

The following [Theorem 4.1](#) is a strengthening of the tightness of [Theorem 1.1](#), which we will prove in this section.

**Theorem 4.1.** *There is an  $n$ -qubit state  $|\psi\rangle$  such that any adaptive Clifford+ $T$  circuit  $\mathcal{A}$  that prepares  $|\psi\rangle$  up to error  $\varepsilon$  in trace distance satisfies*

$$\mathcal{T}^0(\mathcal{A}) = \Omega\left(\sqrt{2^n \log(1/\varepsilon)} + \log(1/\varepsilon)\right).$$

This theorem lower bounds the expected  $T$ -count  $\mathcal{T}^0(\mathcal{A})$  of an adaptive Clifford+ $T$  circuit—this is the most general setting we are able to handle. Note that if  $\mathcal{A}$  is in fact a unitary  $U$  composed of Clifford gates and  $T$  gates (with no measurement), then its output state is a pure state  $\mathcal{A}(\rho) = U(\rho \otimes |0^a\rangle\langle 0^a|)U^\dagger$  where  $a$  is the number of ancillas used. In particular, there is only one Kraus operator  $K = U$  in (49). In this case the number of  $T$  gates in the circuit for  $U$  is equal to  $\mathcal{T}_{\max}(\mathcal{A})$  which is also equal to  $\mathcal{T}(\mathcal{A})$ .

**Unitary Synthesis.** For an  $n$ -qubit unitary  $U$ , consider the channel  $\mathcal{U}$  defined by

$$\mathcal{U}(\rho) \equiv U\rho U^\dagger \otimes |0^a\rangle\langle 0^a|. \quad (57)$$

We say that  $\mathcal{A}$  implements  $U$  to within error  $\varepsilon$  in the diamond distance if

$$\|\mathcal{A} - \mathcal{U}\|_\diamond \leq \varepsilon, \quad (58)$$

where  $\|\cdot\|_\diamond$  here is the diamond norm. As a strengthening of the lower bound in [Theorem 1.2](#), we will prove the following [Theorem 4.2](#).

**Theorem 4.2.** *There is an  $n$ -qubit diagonal unitary  $D$  such that any adaptive Clifford+ $T$  circuit  $\mathcal{A}$  that implements  $D$  up to error  $\varepsilon$  in diamond distance satisfies*

$$\mathcal{T}(\mathcal{A}) = \Omega\left(\sqrt{2^n \log(1/\varepsilon)} + \log(1/\varepsilon)\right).$$

Note that in the case where  $\mathcal{A}$  is a unitary Clifford+ $T$  circuit with no measurements, the diamond distance is equivalent to the usual distance with respect to the operator norm, up to a global phase.

Finally, we establish the following  $T$ -count lower bound for synthesis of general unitaries.

**Theorem 4.3.** *There is an  $n$ -qubit unitary  $U$  such that any adaptive Clifford+ $T$  circuit  $\mathcal{A}$  that implements  $U$  up to error  $\varepsilon$  in diamond distance satisfies*

$$\mathcal{T}(\mathcal{A}) = \Omega\left(2^n \sqrt{\log(1/\varepsilon)} + \log(1/\varepsilon)\right).$$

The rest of the section is devoted to the proofs of [Theorem 4.1](#), [Theorem 4.2](#), and [Theorem 4.3](#).

## 4.1 State preparation

In this section we prove [Theorem 4.1](#).

As an intermediate step in the proof, it will be convenient to work with a model of Clifford circuits with Pauli postselection, defined as follows.

An operator  $P$  is an  $n$ -qubit Pauli iff  $P = i^b P'$  for some  $b \in \{0, 1, 2, 3\}$  and  $P' \in \{I, X, Y, Z\}^{\otimes n}$ . In addition, if  $b \in \{0, 2\}$ , then  $P$  is a Hermitian Pauli. An  $n$ -qubit Pauli postselection is specified by an  $n$ -qubit Hermitian Pauli operator  $P$ . On input state  $|\phi\rangle$ , the postselection produces a normalized state  $|\psi\rangle$  of  $|\phi\rangle$  projected to the  $+1$  eigenspace of  $P$ , i.e.,  $|\psi\rangle$  is proportional to  $(I + P)|\phi\rangle$ , denoted by  $|\psi\rangle \propto (I + P)|\phi\rangle$ . More precisely,  $|\psi\rangle = \frac{(I+P)|\phi\rangle}{\|(I+P)|\phi\rangle\|}$  and it is defined arbitrarily if  $(I + P)|\phi\rangle$  becomes zero. Note that Pauli postselection is not a linear map.

**Definition 4.4** (Clifford circuit with Pauli postselection). A Clifford circuit with Pauli postselection is a sequence of Clifford gates and Pauli postselections applied to the input state.

A Clifford circuit with Pauli postselection and copies of the single-qubit magic state  $|T\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i\pi/4}|1\rangle)$  can perform  $T$  gates by gate injection.<sup>28</sup> In the following we shall consider state preparation circuits of the following form:  $\mathcal{C}$  is a Clifford circuit with  $m$  Pauli postselections,  $n$  input qubits, and  $a$  ancillas such that

$$|\phi_{\text{out}}\rangle|0^{t+a}\rangle = \mathcal{C}(|\phi_{\text{in}}\rangle|T\rangle^{\otimes t}|0^a\rangle), \quad (59)$$

where  $t$  is the number of magic states. Note that we can explicitly write out (59) as proportional to

$$C_{m+1}M_m C_m \cdots M_1 C_1 (|\phi_{\text{in}}\rangle|T\rangle^{\otimes t}|0^a\rangle), \quad (60)$$

where each  $C_j$  is an  $(n + t + a)$ -qubit Clifford and  $M_j = I + P_j$  for some  $(n + t + a)$ -qubit Hermitian Pauli; and  $C_j, M_j$  depend only on  $\mathcal{C}$ , not on  $|\phi_{\text{in}}\rangle, |\phi_{\text{out}}\rangle$ .

The following Claim shows that, if there is an adaptive Clifford+ $T$  circuit  $\mathcal{A}$  that approximately prepares a given state with expected  $T$ -count  $\mathcal{T}^0(\mathcal{A})$ , then there is a Clifford circuit with Pauli postselections and  $2\mathcal{T}^0(\mathcal{A})$  magic states that also approximately prepares the state. This allows us to shift our focus to the model of Clifford circuits with Pauli postselection. The proof uses an averaging argument similar to the one used in [BCHK20, Lemma 5.4].

**Claim 4.5.** Assume  $\mathcal{A}$  is an adaptive Clifford+ $T$  circuit that prepares an  $n$ -qubit state  $|\psi\rangle$  up to error  $\varepsilon$  in trace distance. Let  $t = 2 \cdot \mathcal{T}^0(\mathcal{A})$ . Then for some integers  $m, a \geq 0$ , there exists a Clifford circuit  $\mathcal{C}$  with  $m$  Pauli postselections and  $a$  ancillas such that  $\mathcal{C}(|0^n\rangle|T\rangle^{\otimes t}|0^a\rangle) = |\phi\rangle|0^{t+a}\rangle$  and  $\frac{1}{2} \||\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\|_1 \leq \sqrt{6\varepsilon}$ .

*Proof.* Assume  $\mathcal{A}$  uses  $a$  ancillas. Define  $\rho = |0^n\rangle\langle 0^n|$ . Recall from (50) that

$$\mathcal{A}(\rho) = \sum_r p_r(\rho) \cdot \sigma_r(\rho), \quad (61)$$

where  $p_r(\rho)$  is a distribution. In addition,  $K_r$  uses  $\mathcal{T}_r$   $T$  gates and

$$\sigma_r(\rho) = |\pi_r\rangle\langle\pi_r|, \quad (62)$$

---

<sup>28</sup>In particular, we can implement a  $T$  gate using the fact that  $(I \otimes \langle 0|)\text{CNOT}|\psi\rangle|T\rangle = \frac{1}{\sqrt{2}}T|\psi\rangle$  for any single qubit  $|\psi\rangle$ .

where  $|\pi_r\rangle \propto K_r(|0^n\rangle|0^a\rangle)$  is a pure state. We will find some  $r$  such that  $|\pi_r\rangle$  is close to  $|\psi\rangle|0^a\rangle$  and  $K_r$  uses a small number of  $T$  gates.

By assumption,  $\mathcal{A}(\rho)$  is  $\varepsilon$ -close to  $|\psi\rangle\langle\psi| \otimes |0^a\rangle\langle 0^a|$  in trace distance. Hence the fidelity  $F$  between  $\mathcal{A}(\rho)$  and  $|\psi\rangle\langle\psi| \otimes |0^a\rangle\langle 0^a|$  is

$$F = \langle 0^a | \langle \psi | \mathcal{A}(\rho) | \psi \rangle | 0^a \rangle \geq (1 - \varepsilon)^2 \geq 1 - 2\varepsilon, \quad (63)$$

where we use the inequality  $1 - \sqrt{F} \leq \varepsilon$  (see e.g., [Wil11, Theorem 9.3.1]). Plugging this into (61) and (62), we have

$$\mathbb{E}_{r \sim p_r(\rho)} \left[ \underbrace{|\langle \pi_r | \psi \rangle \langle 0^a | 0^a \rangle|^2}_{F_r} \right] \geq 1 - 2\varepsilon. \quad (64)$$

Hence by Markov's inequality, for at least  $2/3$  mass of  $r$  (under distribution  $p_r(\rho)$ ), the fidelity between  $|\pi_r\rangle$  and  $|\psi\rangle|0^a\rangle$  is  $F_r \geq 1 - 6\varepsilon$ . On the other hand, since  $\mathcal{T}^0(\mathcal{A}) = \mathbb{E}_{r \sim p_r(\rho)} [\mathcal{T}_r]$ , for at least  $1/2$  mass of  $r$  we have  $\mathcal{T}_r \leq 2 \cdot \mathcal{T}^0(\mathcal{A})$ . Hence there exists some  $r$  such that  $\mathcal{T}_r \leq 2 \cdot \mathcal{T}^0(\mathcal{A})$  and  $F_r \geq 1 - 6\varepsilon$ .

Fix any such  $r$  and write  $|\pi_r\rangle = \alpha|\phi\rangle|0^a\rangle + \sqrt{1-\alpha}|\perp\rangle$ , where  $|\phi\rangle$  is a normalized state and  $(I \otimes |0^a\rangle\langle 0^a|)|\perp\rangle = 0$ . Since

$$1 - 6\varepsilon \leq F_r = |\langle \pi_r | \psi \rangle \langle 0^a | 0^a \rangle|^2 = \alpha^2 |\langle \phi | \psi \rangle|^2 \leq |\langle \phi | \psi \rangle|^2, \quad (65)$$

we know that the fidelity between  $|\phi\rangle$  and  $|\psi\rangle$  is at least  $1 - 6\varepsilon$ . This implies that the trace distance between  $|\phi\rangle$  and  $|\psi\rangle$  is at most  $\sqrt{6\varepsilon}$  (see e.g., [Wil11, Theorem 9.3.1]).

Finally we observe that  $|\phi\rangle|0^a\rangle$  is obtained from  $|\pi_r\rangle$  followed by postselecting the last  $a$  qubits being all-zero. This means we can construct  $|\phi\rangle|0^a\rangle$  by applying  $K_r$  and the above postselection on  $|0^n\rangle|0^a\rangle$ . This process uses  $t = 2 \cdot \mathcal{T}^0(\mathcal{A})$   $T$  gates,<sup>29</sup> each of which can be implemented using a magic state  $|T\rangle$  and gate injection. This gives a Clifford circuit  $\mathcal{C}$  with Pauli postselection and input state  $|0^n\rangle|T\rangle^{\otimes t}|0^a\rangle$  which prepares  $|\phi\rangle|0^a\rangle$ .  $\square$

Given Claim 4.5, we now focus on Clifford+ $T$  circuits with Pauli postselections and ancillas. To complete the proof of Theorem 4.1, it is sufficient to prove the following proposition.

**Proposition 4.6.** *There is an  $n$ -qubit state  $|\psi\rangle$  such that the following holds. Assume  $\mathcal{C}$  is a Clifford circuit with  $m$  Pauli postselections and  $a$  ancillas. Assume  $\mathcal{C}(|0^n\rangle|T\rangle^{\otimes t}|0^a\rangle) = |\phi\rangle|0^{t+a}\rangle$  and  $\frac{1}{2} \|\phi\rangle\langle\phi| - |\psi\rangle\langle\psi|\|_1 \leq \varepsilon$ . Then  $t = \Omega\left(\sqrt{2^n \log(1/\varepsilon)} + \log(1/\varepsilon)\right)$ .*

The first lower bound of  $\Omega(\log(1/\varepsilon))$  holds even for single-qubit state preparation. This is proved by Beverland, Campbell, Howard, and Kliuchnikov [BCHK20, Lemma 5.9]. The second  $T$ -count lower bound is a counting argument as in [LKS24]: the number of distinct circuits should be at least the number of distinguishable states in question. The latter quantity follows directly from standard sphere packing bounds for  $\ell_2$  balls (see e.g., [Wik24a]).

**Fact 4.7.** *There are  $N = (1/\varepsilon)^{\Theta(2^n)}$   $n$ -qubit states  $|\tau_1\rangle, \dots, |\tau_N\rangle$  with pairwise trace distance at least  $\varepsilon$ .*

<sup>29</sup>Though  $\mathcal{T}_r \leq 2 \cdot \mathcal{T}(\mathcal{A})$ , we can always ensure that there are exactly  $t$   $T$  gates (say by applying dummy  $T$  gates acting on one of the ancillas in the  $|0\rangle$  state at the beginning of the circuit; note  $T|0\rangle = |0\rangle$ ).

*Proof.* We start with a set of  $n$ -qubit states  $|\zeta_1\rangle, \dots, |\zeta_M\rangle$  that are pairwise  $(10\varepsilon)$ -far in  $\ell_2$  distance. By the standard  $\ell_2$  sphere packing, we can choose  $M = (1/\varepsilon)^{\Theta(2^n)}$ . Note that the trace distance between two pure states is asymptotically lower bounded by the  $\ell_2$  distance with a global phase, i.e.,

$$\frac{1}{2} \|\langle\phi|\langle\phi| - |\psi\rangle\langle\psi|\|_1 = \sqrt{1 - |\langle\phi|\psi\rangle|^2} \geq \sqrt{1 - |\langle\phi|\psi\rangle|} = \frac{1}{\sqrt{2}} \sqrt{2 - 2|\langle\phi|\psi\rangle|} \quad (66)$$

$$= \frac{1}{\sqrt{2}} \sqrt{2 - 2 \max_{\theta \in [0, 2\pi)} \Re(e^{i\theta} \langle\phi|\psi\rangle)} \quad (67)$$

$$= \frac{1}{\sqrt{2}} \cdot \min_{\theta \in [0, 2\pi)} \|\langle\phi| - e^{i\theta} \langle\psi|\| \quad (68)$$

By discretizing the rotation angle  $\theta$ , each  $|\zeta_j\rangle$  can be  $\varepsilon$ -close to  $O(1/\varepsilon)$  other  $|\zeta_k\rangle$ 's in trace distance. Hence we can find  $N = \Theta(\varepsilon M) = (1/\varepsilon)^{\Theta(2^n)}$  states  $|\tau_1\rangle, \dots, |\tau_N\rangle$  from  $|\zeta_1\rangle, \dots, |\zeta_M\rangle$  such that their pairwise trace distance is at least  $\varepsilon$ .  $\square$

To upper bound the number of distinct circuits, we express it in a canonical form as described in [Lemma 4.8](#). This was established in [\[BCHK20, Section A.7\]](#). Crucially, this canonical form has size independent of the number of Pauli postselections and ancillas.

**Lemma 4.8** ([\[BCHK20\]](#)). *Let  $\mathcal{C}$  be a Clifford circuit with  $m$  Pauli postselections,  $n$  input qubits, and  $a$  ancillas. Let  $|\phi\rangle$  be an  $n$ -qubit state. Assume*

$$|\phi\rangle|0^{t+a}\rangle = \mathcal{C}(|0^n\rangle|T\rangle^{\otimes t}|0^a\rangle). \quad (69)$$

*Then there exists an  $(n+t)$ -qubit Clifford unitary  $C$  and matrices  $M_1, M_2, \dots, M_t$  such that*

$$|\phi\rangle|0^t\rangle \propto CM_t \cdots M_2 M_1 (|0^n\rangle|T\rangle^{\otimes t}), \quad (70)$$

*where each  $M_j$  is  $I + P_j$  for some  $(n+t)$ -qubit Hermitian Pauli  $P_j$ .*

We remark that [Lemma 4.8](#) provides a canonical form for state preparation circuits. With a slight tweak, the original proof works for a more general setting including unitary synthesis circuits. For completeness and future reference, we present the statement as [Theorem A.2](#) and prove it in [Section A](#).

We are now ready to prove [Proposition 4.6](#). This, combined with [Claim 4.5](#), completes the proof of [Theorem 4.1](#).

*Proof of Proposition 4.6.* Assume  $\varepsilon < 1/8$ . This ensures that the trace distance between  $|\phi\rangle$  and  $|\psi\rangle$  is at most  $1/8$  (see e.g., [\[Wil11, Section 9\]](#)). Then the  $\Omega(\log(1/\varepsilon))$  lower bound follows from [\[BCHK20, Lemma 5.9\]](#).

To get the  $\Omega(\sqrt{2^n \log(1/\varepsilon)})$  lower bound, assume for every  $n$ -qubit state  $|\psi\rangle$  there is a Clifford circuit  $\mathcal{C}$  with Pauli postselection such that  $\mathcal{C}(|0^n\rangle|T\rangle^{\otimes t}|0^a\rangle) = |\phi\rangle|0^{t+a}\rangle$  and the trace distance of  $|\phi\rangle, |\psi\rangle$  is at most  $\varepsilon$ .

By setting  $|\psi\rangle = |\tau_j\rangle$  for each  $j \in [N]$  from [Fact 4.7](#), we apply [Lemma 4.8](#) to get a state preparation circuit in canonical form. Now let us count the number of possible state preparation circuits of the form (70). To this end note that there are  $2^{O((n+t)^2)}$  Clifford unitaries  $C$  on  $n+t$  qubits, and there are  $2 \cdot 4^{n+t}$  choices for each Hermitian Pauli  $P_j$ . So we have  $2^{O(n^2+t^2)}$  possible state preparation circuits of the form (70). By the construction of  $|\tau_j\rangle$ 's in [Fact 4.7](#), each  $|\tau_j\rangle$  gives a different circuit. Thus  $2^{O(n^2+t^2)} \geq (1/\varepsilon)^{\Theta(2^n)}$ , which gives  $t = \Omega(\sqrt{2^n \log(1/\varepsilon)})$  since  $\varepsilon < 1/8$ .  $\square$

## 4.2 Unitary synthesis

Now we turn to the unitary synthesis  $T$ -count lower bounds ([Theorem 4.2](#) and [Theorem 4.3](#)). We will reduce them to state preparation tasks, by considering the action of an adaptive Clifford+ $T$  circuit on the maximally entangled state and comparing this with the Choi state corresponding to the unitary  $U$  of interest.

**Definition 4.9** (Maximally entangled state and Choi state). We use

$$|\iota\rangle = 2^{-n/2} \sum_{x \in \{0,1\}^n} |x\rangle|x\rangle$$

to denote the  $2n$ -qubit maximally entangled state. For an  $n$ -qubit unitary  $U$ , we use  $|\iota_U\rangle$  to denote the Choi state, defined by

$$|\iota_U\rangle = 2^{-n/2} \sum_{x \in \{0,1\}^n} |x\rangle \otimes (U|x\rangle). \quad (71)$$

The following fact follows directly from (58) and the definition of the diamond norm.

**Fact 4.10.** Assume an adaptive Clifford+ $T$  circuit  $\mathcal{A}$ , using  $a$  ancillas, implements an  $n$ -qubit unitary  $U$  up to error  $\varepsilon$  in diamond distance. Then  $(I \otimes \mathcal{A})(|\iota\rangle\langle\iota|)$  is  $\varepsilon$ -close in trace distance to  $|\iota_U\rangle\langle\iota_U| \otimes |0^a\rangle\langle 0^a|$ .

By an almost identical argument as [Claim 4.5](#), we have the following claim.

**Claim 4.11.** Assume an adaptive Clifford+ $T$  circuit  $\mathcal{A}$  implements an  $n$ -qubit unitary  $U$  up to error  $\varepsilon$  in the diamond distance. Let  $t = 2 \cdot \mathcal{T}(\mathcal{A})$ . Then for some integers  $m, a \geq 0$ , there exists a Clifford circuit  $\mathcal{C}$  with  $m$  Pauli postselections and  $a$  ancillas such that  $\mathcal{C}(|0^{2n}\rangle|T\rangle^{\otimes t}|0^a\rangle) = |\phi\rangle|0^{t+a}\rangle$  and  $\frac{1}{2} \| |\iota_U\rangle\langle\iota_U| - |\phi\rangle\langle\phi| \|_1 \leq \sqrt{6\varepsilon}$ .

*Proof.* By [Fact 4.10](#), we know that  $(I \otimes \mathcal{A})(|\iota\rangle\langle\iota|)$  is  $\varepsilon$ -close in trace distance to  $|\iota_U\rangle\langle\iota_U| \otimes |0^a\rangle\langle 0^a|$  for some integer  $a \geq 0$ . Then following the proof of [Claim 4.5](#) (but replacing the all-zeros input state with  $\rho = |\iota\rangle\langle\iota|$  and the expected  $T$ -count  $\mathcal{T}^0(\mathcal{A})$  by the worst-case expected  $T$ -count  $\mathcal{T}(\mathcal{A})$ ), we obtain a Clifford circuit  $\mathcal{C}'$  with Pauli postselection such that

$$\mathcal{C}'(|\iota\rangle|T\rangle^{\otimes t}|0^a\rangle) = |\phi\rangle|0^{t+a}\rangle, \quad (72)$$

where the trace distance between  $|\iota_U\rangle$  and  $|\phi\rangle$  is at most  $\sqrt{6\varepsilon}$ . Moreover,  $\mathcal{C}'$  uses  $t = 2 \cdot \mathcal{T}(I \otimes \mathcal{A}) = 2 \cdot \mathcal{T}(\mathcal{A})$  magic states, where the second equality follows from (54).

Now notice that the maximally entangled state

$$|\iota\rangle = 2^{-n/2} \sum_{x \in \{0,1\}^n} |x\rangle|x\rangle = \left( \frac{|00\rangle + |11\rangle}{\sqrt{2}} \right)^{\otimes n} \quad (73)$$

can be constructed by  $n$  single-qubit Hadamard gates and  $n$  CNOTs on  $|0^{2n}\rangle$ . Let  $\mathcal{C}$  be the circuit applying these Clifford operations followed by  $\mathcal{C}'$ . Then we have  $\mathcal{C}(|0^{2n}\rangle|T\rangle^{\otimes t}|0^a\rangle) = |\phi\rangle|0^{t+a}\rangle$  as claimed.  $\square$

By [Claim 4.11](#), it suffices to prove the following [Proposition 4.12](#) and [Proposition 4.13](#) to obtain [Theorem 4.2](#) and [Theorem 4.3](#) respectively.

**Proposition 4.12.** *There is an  $n$ -qubit diagonal unitary  $D$  such that the following holds. Assume  $\mathcal{C}$  is a Clifford circuit with  $m$  Pauli postselections and  $a$  ancillas. Assume*

$$\mathcal{C}(|0^{2n}\rangle|T\rangle^{\otimes t}|0^a\rangle) = |\phi\rangle|0^{t+a}\rangle$$

and  $\frac{1}{2} \|\langle\phi|\langle\phi| - |\iota_D\rangle\langle\iota_D|\|_1 \leq \varepsilon$ . Then  $t = \Omega\left(\sqrt{2^n \log(1/\varepsilon)} + \log(1/\varepsilon)\right)$ .

**Proposition 4.13.** *There is an  $n$ -qubit unitary  $U$  such that the following holds. Assume  $\mathcal{C}$  is a Clifford circuit with  $m$  Pauli postselections and  $a$  ancillas. Assume*

$$\mathcal{C}(|0^{2n}\rangle|T\rangle^{\otimes t}|0^a\rangle) = |\phi\rangle|0^{t+a}\rangle$$

and  $\frac{1}{2} \|\langle\phi|\langle\phi| - |\iota_U\rangle\langle\iota_U|\|_1 \leq \varepsilon$ . Then  $t = \Omega\left(2^n \sqrt{\log(1/\varepsilon)} + \log(1/\varepsilon)\right)$ .

Following the proof of [Theorem 4.1](#), we now need to construct a large set of distinguishable Choi states to apply the counting argument. For this purpose we relate the trace distance between Choi states to normalized Hilbert–Schmidt (aka Frobenius) distance of the corresponding unitaries.

In particular, if  $A$  is a  $d \times d$  matrix we write  $\|A\|_{\text{HS}} = \sqrt{\text{Tr}(A^\dagger A)/d}$ .

**Fact 4.14.** *Let  $U$  and  $V$  be two  $n$ -qubit unitaries. The trace distance between their Choi states is*

$$\frac{1}{2} \|\langle\iota_U|\langle\iota_U| - |\iota_V\rangle\langle\iota_V|\|_1 \geq \frac{1}{\sqrt{2}} \cdot \min_{\theta \in [0, 2\pi)} \|U - e^{i\theta} \cdot V\|_{\text{HS}}, \quad (74)$$

where  $\|\cdot\|_{\text{HS}}$  is the normalized Hilbert-Schmidt norm.

*Proof.* By (68), it suffices to observe the following

$$\|\langle\iota_U| - e^{i\theta} \langle\iota_V|\|^2 = \left\| 2^{-n/2} \sum_{x \in \{0,1\}^n} |x\rangle \otimes \left( (U - e^{i\theta} V) |x\rangle \right) \right\|^2 \quad (75)$$

$$= 2^{-n} \sum_{x \in \{0,1\}^n} \left\| (U - e^{i\theta} V) |x\rangle \right\|^2 \quad (76)$$

$$= 2^{-n} \cdot \text{Tr} \left( (U - e^{i\theta} V)^\dagger (U - e^{i\theta} V) \right) \quad (77)$$

$$= \|U - e^{i\theta} V\|_{\text{HS}}^2. \quad \square$$

Note that we have similar packing bounds.

**Fact 4.15.** *There are  $N = (1/\varepsilon)^{\Theta(2^n)}$   $n$ -qubit diagonal unitaries  $D_1, \dots, D_N$  such that the pairwise trace distance between their Choi states is at least  $\varepsilon$ .*

*Proof.* By [Fact 4.14](#), we work with the normalized Hilbert-Schmidt norm. The phase issue will be handled by discretization, same as the proof of [Fact 4.7](#).

Observe that the normalized Hilbert-Schmidt distance between two diagonal unitaries is the normalized  $\ell_2$  distance between their diagonals (viewed as vectors of length  $2^n$ ). Then by standard sphere packing results, we can get  $M = (1/\varepsilon)^{\Theta(2^n)}$  diagonal unitaries  $\tilde{D}_1, \dots, \tilde{D}_M$  that are pairwise  $(10\varepsilon)$ -far in normalized Hilbert-Schmidt distance. By discretizing the rotation angle of the global phase, each  $\tilde{D}_j$  can be  $(\sqrt{2}\varepsilon)$ -close to  $O(1/\varepsilon)$  other  $\tilde{D}_k$ 's in normalized Hilbert-Schmidt distance even with the presence of a global phase. This, combined with [Fact 4.14](#), implies that we can find  $N = \Theta(\varepsilon M) = (1/\varepsilon)^{\Theta(2^n)}$  diagonal unitaries  $D_1, \dots, D_N$  from  $\tilde{D}_1, \dots, \tilde{D}_M$  such that the pairwise trace distance between their Choi states is at least  $\varepsilon$ .  $\square$



**Fact 4.16.** *There are  $N = (1/\varepsilon)^{\Theta(4^n)}$   $n$ -qubit unitaries  $U_1, \dots, U_N$  such that the pairwise trace distance between their Choi states is at least  $\varepsilon$ .*

*Proof.* We only show how to obtain  $M = (1/\varepsilon)^{\Theta(4^n)}$  unitaries that are  $(\sqrt{2}\varepsilon)$ -far from each other in normalized Hilbert-Schmidt distance. The remaining proof is the same as Fact 4.15.

Let  $A$  be a skew Hermitian matrix (i.e.,  $A + A^\dagger = 0$ ). Then  $e^A$  is a unitary. By relating the normalized Hilbert-Schmidt distance before and after matrix exponentiation, we construct the desired unitaries via “well-separated skew Hermitian matrices” (see e.g., [BL18, Appendix D]). To this end, let  $B$  be another skew Hermitian matrix and assume  $\|A\|_{\text{HS}}, \|B\|_{\text{HS}} \leq 1/2$ . Then notice that

$$\|e^A - e^B\|_{\text{HS}} = \left\| \sum_{k=1}^{+\infty} \frac{A^k - B^k}{k!} \right\|_{\text{HS}} \quad (\text{Taylor expansion})$$

$$\geq \|A - B\|_{\text{HS}} - \sum_{k \geq 2} \frac{\|A^k - B^k\|_{\text{HS}}}{k!} \quad (78)$$

$$= \|A - B\|_{\text{HS}} - \sum_{k \geq 2} \frac{\left\| \sum_{\ell=1}^k A^{k-\ell}(A - B)B^{\ell-1} \right\|_{\text{HS}}}{k!} \quad (79)$$

$$\geq \|A - B\|_{\text{HS}} - \sum_{k \geq 2} \frac{k \cdot 2^{-(k-1)} \cdot \|A - B\|_{\text{HS}}}{k!} \quad (\text{since } \|A\|_{\text{HS}}, \|B\|_{\text{HS}} \leq 1/2)$$

$$= (2 - \sqrt{e}) \|A - B\|_{\text{HS}} \geq \|A - B\|_{\text{HS}} / 3. \quad (80)$$

Hence we just need to obtain skew Hermitian matrices  $A_1, \dots, A_M$  of norm  $\|A_j\|_{\text{HS}} \leq 1/2$  and pairwise distance  $\|A_j - A_k\|_{\text{HS}} \geq 3\sqrt{2}\varepsilon$ . The desired bound on  $M$  follows from standard sphere packing in  $\ell_2$  norm, by viewing the upper triangular part of a  $2^n \times 2^n$  skew Hermitian matrices as a vector of length  $2^n(2^n + 1)/2 = \Theta(4^n)$ .  $\square$

At this point, we use the canonical circuit from Lemma 4.8 and prove Proposition 4.12 and Proposition 4.13, which, combined with Claim 4.11, complete the proof of Theorem 4.2 and Theorem 4.3.

*Proof of Proposition 4.12.* Assume  $\varepsilon$  is a small constant. Then the  $\Omega(\log(1/\varepsilon))$  lower bound also follows from [BCHK20, Lemma 5.9]. To get the  $\Omega(\sqrt{2^n \log(1/\varepsilon)})$  lower bound, we follow the same calculation as in the proof of Proposition 4.6. Suppose that for every diagonal unitary  $D$  there is a Clifford circuit  $\mathcal{C}$  with Pauli postselection such that  $\mathcal{C}(|0^{2n}\rangle|T\rangle^{\otimes t}|0^a\rangle) = |\phi\rangle|0^{t+a}\rangle$  and the trace distance between  $|\phi\rangle$  and  $|\iota_D\rangle$  is at most  $\varepsilon$ .

In light of Fact 4.14, we consider the diagonal unitaries  $D = D_j$  for each  $j \in [N]$  from Fact 4.15. Then by Lemma 4.8, we obtain a canonical form for the  $(2n + t)$ -qubit Clifford circuit with Pauli postselection that approximately prepares  $|\iota_{D_j}\rangle$  for each  $j$ . But there are at most  $2^{O(n^2+t^2)}$  different canonical forms on  $2n + t$  qubits. By Fact 4.14 and Fact 4.15, each  $|\iota_{D_j}\rangle$  gives a different circuit. Thus  $2^{O(n^2+t^2)} \geq (1/\varepsilon)^{\Theta(2^n)}$ , which gives  $t = \Omega(\sqrt{2^n \log(1/\varepsilon)})$  since  $\varepsilon < 1/8$ .  $\square$

*Proof of Proposition 4.13.* As above, the  $\Omega(\log(1/\varepsilon))$  lower bound follows from [BCHK20]. The  $\Omega(2^n \sqrt{\log(1/\varepsilon)})$  follows by applying the same counting argument used in the proof of Proposition 4.12, but now with the unitaries  $U = U_j$  for each  $j \in [N]$  from Fact 4.16 and  $N = (1/\varepsilon)^{\Theta(4^n)}$ .  $\square$

## Acknowledgments

We thank Ryan Babbush, Dominic Berry, Michael Beverland, Oscar Higgott, Tanuj Khattar, Guang Hao Low, Dmitri Maslov, Luke Shaeffer, and Rolando Somma for helpful discussions. We thank Tanuj Khattar for pointing out that the analysis of the algorithm in [LKS24] can be improved. We thank anonymous reviewers for helpful comments. KW wants to thank Daniel Grier, Jiaqing Jiang, Saeed Mehraban, and Anurudh Peduri for relevant references.

DG is a CIFAR fellow in the quantum information science program. This work was done while KW was at Google.

## Author contribution

The authors jointly discussed the technical contents of the paper and wrote the paper, including all the revisions. Large language model tools were only used to address typesetting and package collision issues when switching to this template in the final stage, after which the authors also proofread.

## References

- [Aar16] Scott Aaronson. The complexity of quantum states and transformations: from quantum money to black holes. *arXiv preprint arXiv:1607.05256*, 2016. [arXiv:1607.05256](#), [doi:10.48550/arXiv.1607.05256](#).
- [AK07] Scott Aaronson and Greg Kuperberg. Quantum versus classical proofs and advice. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 115–128. IEEE, 2007. [doi:10.1109/CCC.2007.27](#).
- [BBC<sup>+</sup>95] Adriano Barenco, Charles H. Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter. Elementary gates for quantum computation. *Phys. Rev. A*, 52:3457–3467, Nov 1995. [doi:10.1103/PhysRevA.52.3457](#).
- [BBC<sup>+</sup>19] Sergey Bravyi, Dan Browne, Pádraic Calpin, Earl Campbell, David Gosset, and Mark Howard. Simulation of quantum circuits by low-rank stabilizer decompositions. *Quantum*, 3:181, 2019. [doi:10.22331/q-2019-09-02-181](#).
- [BCHK20] Michael Beverland, Earl Campbell, Mark Howard, and Vadym Kliuchnikov. Lower bounds on the non-Clifford resources for quantum computations. *Quantum Science and Technology*, 5(3):035009, 2020. [doi:10.1088/2058-9565/ab8963](#).
- [BCK15] Dominic W. Berry, Andrew M. Childs, and Robin Kothari. Hamiltonian simulation with nearly optimal dependence on all parameters. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 792–809. IEEE, 2015. [doi:10.1109/FOCS.2015.54](#).
- [BG16] Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by Clifford gates. *Physical review letters*, 116(25):250501, 2016. [doi:10.1103/PhysRevLett.116.250501](#).
- [BHMT02] Gilles Brassard, Peter Hoyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002. [doi:10.1090/conm/305/05215](#).

- [BK05] Sergey Bravyi and Alexei Kitaev. Universal quantum computation with ideal Clifford gates and noisy ancillas. *Physical Review A—Atomic, Molecular, and Optical Physics*, 71(2):022316, 2005. doi:[10.1103/PhysRevA.71.022316](https://doi.org/10.1103/PhysRevA.71.022316).
- [BL18] Thomas Barthel and Jianfeng Lu. Fundamental limitations for measurements in quantum many-body systems. *Physical Review Letters*, 121(8):080406, 2018. doi:[10.1103/PhysRevLett.121.080406](https://doi.org/10.1103/PhysRevLett.121.080406).
- [BP05] Joan Boyar and René Peralta. The exact multiplicative complexity of the Hamming weight function. In *Electronic Colloquium on Computational Complexity (ECCC’05)*, (049), 2005.
- [BSS16] Sergey Bravyi, Graeme Smith, and John A Smolin. Trading classical and quantum computational resources. *Physical Review X*, 6(2):021043, 2016. doi:[10.1103/PhysRevX.6.021043](https://doi.org/10.1103/PhysRevX.6.021043).
- [CEMM98] Richard Cleve, Artur Ekert, Chiara Macchiavello, and Michele Mosca. Quantum algorithms revisited. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1969):339–354, 1998. doi:[10.1098/rspa.1998.0164](https://doi.org/10.1098/rspa.1998.0164).
- [CG97] Richard Cleve and Daniel Gottesman. Efficient computations of encodings for quantum error correction. *Physical Review A*, 56(1):76, 1997. doi:[10.1103/PhysRevA.56.76](https://doi.org/10.1103/PhysRevA.56.76).
- [EKLH21] Tyler D. Ellison, Kohtaro Kato, Zi-Wen Liu, and Timothy H. Hsieh. Symmetry-protected sign problem and magic in quantum phases of matter. *Quantum*, 5:612, 2021. doi:[10.22331/q-2021-12-28-612](https://doi.org/10.22331/q-2021-12-28-612).
- [FHH<sup>+</sup>14] Yuval Filmus, Hamed Hatami, Steven Heilman, Elchanan Mossel, Ryan O’Donnell, Sushant Sachdeva, Andrew Wan, and Karl Wimmer. Real analysis in computer science: A collection of open problems, 2014. URL: <https://simons.berkeley.edu/sites/default/files/openprobsmerged.pdf>.
- [Got98] Daniel Gottesman. The heisenberg representation of quantum computers. *arXiv preprint quant-ph/9807006*, 1998. arXiv:[quant-ph/9807006](https://arxiv.org/abs/quant-ph/9807006), doi:[10.48550/arXiv.quant-ph/9807006](https://doi.org/10.48550/arXiv.quant-ph/9807006).
- [GR02] Lov Grover and Terry Rudolph. Creating superpositions that correspond to efficiently integrable probability distributions. *arXiv preprint quant-ph/0208112*, 2002. arXiv:[quant-ph/0208112](https://arxiv.org/abs/quant-ph/0208112), doi:[10.48550/arXiv.quant-ph/0208112](https://doi.org/10.48550/arXiv.quant-ph/0208112).
- [Gro98] Lov K Grover. Quantum computers can search rapidly by using almost any transformation. *Physical Review Letters*, 80(19):4329, 1998. doi:[10.1103/PhysRevLett.80.4329](https://doi.org/10.1103/PhysRevLett.80.4329).
- [GS13] Brett Giles and Peter Selinger. Exact synthesis of multiqubit Clifford+T circuits. *Physical Review A—Atomic, Molecular, and Optical Physics*, 87(3):032332, 2013. doi:[10.1103/PhysRevA.87.032332](https://doi.org/10.1103/PhysRevA.87.032332).
- [GSJ24] Craig Gidney, Noah Shuitty, and Cody Jones. Magic state cultivation: growing t states as cheap as cnot gates. *arXiv preprint arXiv:2409.17595*, 2024. doi:[10.48550/arXiv.2409.17595](https://doi.org/10.48550/arXiv.2409.17595).
- [Haa81] Uffe Haagerup. The best constants in the Khintchine inequality. *Studia Mathematica*, 70(3):231–283, 1981. doi:[10.4064/sm-70-3-231-283](https://doi.org/10.4064/sm-70-3-231-283).
- [Hou58] Alston S Householder. Unitary triangularization of a nonsymmetric matrix. *Journal of the ACM (JACM)*, 5(4):339–342, 1958. doi:[10.1145/320941.320947](https://doi.org/10.1145/320941.320947).
- [HRC02] Aram W. Harrow, Benjamin Recht, and Isaac L. Chuang. Efficient discrete ap-

- proximations of quantum gates. *Journal of Mathematical Physics*, 43(9):4445–4451, 2002. doi:[10.1063/1.1495899](https://doi.org/10.1063/1.1495899).
- [INN<sup>+</sup>22] Sandy Irani, Anand Natarajan, Chinmay Nirkhe, Sujit Rao, and Henry Yuen. Quantum search-to-decision reductions and the state synthesis problem. In *37th Computational Complexity Conference*, 2022. doi:[10.4230/LIPIcs.CCC.2022.5](https://doi.org/10.4230/LIPIcs.CCC.2022.5).
- [Kli13] Vadym Kliuchnikov. Synthesis of unitaries with Clifford+T circuits. *arXiv preprint arXiv:1306.3200*, 2013. doi:[10.48550/arXiv.1306.3200](https://doi.org/10.48550/arXiv.1306.3200).
- [KMM13] Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. Fast and efficient exact synthesis of single-qubit unitaries generated by Clifford and T gates. *Quantum Information & Computation*, 13(7-8):607–630, 2013. doi:[10.26421/QIC13.7-8-4](https://doi.org/10.26421/QIC13.7-8-4).
- [KMM15] Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. Practical approximation of single-qubit unitaries by single-qubit quantum Clifford and T circuits. *IEEE Transactions on Computers*, 65(1):161–172, 2015. doi:[10.1109/TC.2015.2409842](https://doi.org/10.1109/TC.2015.2409842).
- [Kre23] William Kretschmer. Quantum mass production theorems. In *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2023. doi:[10.4230/LIPIcs.TQC.2023.10](https://doi.org/10.4230/LIPIcs.TQC.2023.10).
- [Lit19] Daniel Litinski. Magic state distillation: Not as costly as you think. *Quantum*, 3:205, 2019. doi:[10.22331/q-2019-12-02-205](https://doi.org/10.22331/q-2019-12-02-205).
- [LKS24] Guang Hao Low, Vadym Kliuchnikov, and Luke Schaeffer. Trading T gates for dirty qubits in state preparation and unitary synthesis. *Quantum*, 8:1375, 2024. doi:[10.22331/q-2024-06-17-1375](https://doi.org/10.22331/q-2024-06-17-1375).
- [LOH22] Lorenzo Leone, Salvatore FE Oliviero, and Alioscia Hamma. Stabilizer Rényi entropy. *Physical Review Letters*, 128(5):050402, 2022. doi:[10.1103/PhysRevLett.128.050402](https://doi.org/10.1103/PhysRevLett.128.050402).
- [LW22] Zi-Wen Liu and Andreas Winter. Many-body quantum magic. *PRX Quantum*, 3(2):020333, 2022. doi:[10.1103/PRXQuantum.3.020333](https://doi.org/10.1103/PRXQuantum.3.020333).
- [Mas16] Dmitri Maslov. Optimal and asymptotically optimal nct reversible circuits by the gate types. *Quantum Information & Computation*, 16(13-14):1096–1112, 2016. doi:[10.26421/QIC16.13-14-2](https://doi.org/10.26421/QIC16.13-14-2).
- [MT24] Saeed Mehraban and Mehrdad Tahmasbi. Quadratic lower bounds on the approximate stabilizer rank: A probabilistic approach. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 608–619, 2024. doi:[10.1145/3618260.3649733](https://doi.org/10.1145/3618260.3649733).
- [NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010. doi:[10.1017/CB09780511976667](https://doi.org/10.1017/CB09780511976667).
- [Nec62] Eduard I. Nechiporuk. On the complexity of schemes in some bases containing nontrivial elements with zero weights. *Problemy kibernetiki*, 8:123–160, 1962.
- [Pan12] Denis Pankratov. Direct sum questions in classical communication complexity. *Master’s thesis, University of Chicago*, 2012.
- [Ros21] Gregory Rosenthal. Query and depth upper bounds for quantum unitaries via grover search. *arXiv preprint arXiv:2111.07992*, 2021. arXiv:[2111.07992](https://arxiv.org/abs/2111.07992), doi:[10.48550/arXiv.2111.07992](https://doi.org/10.48550/arXiv.2111.07992).
- [Ros24] Gregory Rosenthal. Efficient quantum state synthesis with one query. In *Pro-*

- ceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 2508–2534. SIAM, 2024. doi:10.1137/1.9781611977912.89.
- [RS16] Neil J. Ross and Peter Selinger. Optimal ancilla-free Clifford+T approximation of Z-rotations. *Quantum Inf. Comput.*, 16(11&12):901–953, 2016. doi:10.26421/QIC16.11-12-1.
- [Sch89] Claus-Peter Schnorr. The multiplicative complexity of Boolean functions. In *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes: 6th International Conference, AAECC-6 Rome, Italy, July 4–8, 1988 Proceedings 6*, pages 45–58. Springer, 1989. doi:10.1007/3-540-51083-4\_47.
- [Sel15] Peter Selinger. Efficient Clifford+T approximation of single-qubit operators. *Quantum Information & Computation*, 15(1-2):159–180, 2015. doi:10.26421/QIC15.1-2-10.
- [Tan25] Xinyu Tan. Unitary synthesis with fewer t gates. *arXiv preprint arXiv:2509.25702*, 2025. doi:10.48550/arXiv.2509.25702.
- [Uhl92] Dietmar Uhlir. Networks computing Boolean functions for multiple input values. In *Proceedings of the London Mathematical Society Symposium on Boolean Function Complexity*, pages 165–173, 1992. doi:10.1017/CB09780511526633.013.
- [Uli74] D Ulig. On the synthesis of self-correcting schemes from functional elements with a small number of reliable elements. *Mathematical Notes of the Academy of Sciences of the USSR*, 15:558–562, 1974. doi:10.1007/BF01152835.
- [VMGE14] Victor Veitch, S. A. Hamed Mousavian, Daniel Gottesman, and Joseph Emerson. The resource theory of stabilizer quantum computation. *New Journal of Physics*, 16(1):013009, 2014. doi:10.1088/1367-2630/16/1/013009.
- [WC12] Nathan Wiebe and Andrew Childs. Hamiltonian simulation using linear combinations of unitary operations. In *APS March Meeting Abstracts*, volume 2012, pages T30–003, 2012. doi:10.26421/QIC12.11-12-1.
- [Wik24a] Wikipedia. Covering number — Wikipedia, the free encyclopedia. <http://en.wikipedia.org/w/index.php?title=Covering%20number&oldid=1190804299>, 2024. [Online; accessed 20-September-2024].
- [Wik24b] Wikipedia. Khintchine inequality — Wikipedia, the free encyclopedia. <http://en.wikipedia.org/w/index.php?title=Khintchine%20inequality&oldid=1200765288>, 2024. [Online; accessed 11-June-2024].
- [Wil11] Mark M Wilde. From classical to quantum shannon theory. *arXiv preprint arXiv:1106.1445*, 2011. doi:10.48550/arXiv.1106.1445.

## A Canonical form of Clifford circuits with magic states and Pauli postselection

To describe our generalization of Lemma 4.8, we need to recall the definition of stabilizer group.

**Definition A.1** (Stabilizer group). For a state  $|\tau\rangle$ , its stabilizer group  $\text{Stab}(|\tau\rangle)$  is the set of Pauli operators  $P$  satisfying  $P|\tau\rangle = |\tau\rangle$ . For a set  $\Lambda$  of states, we define its stabilizer group as  $\text{Stab}(\Lambda) = \bigcap_{|\tau\rangle \in \Lambda} \text{Stab}(|\tau\rangle)$ . Note that the stabilizer group is a set of commutative Hermitian Pauli operators and does not contain  $-I$ . We also remark that the size of a stabilizer group is an integer power of two.

Now we present the formal statement for the canonical form as Theorem A.2.



**Theorem A.2.** Let  $\mathcal{C}$  be a Clifford circuit with  $m$  Pauli postselections,  $n$  input qubits, and  $a$  ancillas. Let  $\{|\phi_\lambda\rangle\}_{\lambda \in \mathcal{S}}$  and  $\{|\psi_\lambda\rangle\}_{\lambda \in \mathcal{S}}$  be two sets of  $n$ -qubit states indexed by  $\mathcal{S}$ . Assume

$$|\psi_\lambda\rangle|0^{t+a}\rangle = \mathcal{C}\left(|\phi_\lambda\rangle|T\rangle^{\otimes t}|0^a\rangle\right) \quad (81)$$

holds for all  $\lambda \in \mathcal{S}$ , where we recall the circuit expression in [Definition 4.4](#). Then there exists an  $(n+t)$ -qubit Clifford unitary  $C$  and matrices  $M_1, M_2, \dots, M_c$  such that

$$|\psi_\lambda\rangle|0^t\rangle \propto CM_c \cdots M_2 M_1 \left(|\phi_\lambda\rangle|T\rangle^{\otimes t}\right) \quad (82)$$

holds for all  $\lambda \in \mathcal{S}$ , where each  $M_j$  is  $I + P_j$  for some  $(n+t)$ -qubit Hermitian Pauli  $P_j$  and

$$c = t - \log(|\text{Stab}(\{|\phi_\lambda\rangle\}_{\lambda \in \mathcal{S}})|) + \log(|\text{Stab}(\{|\psi_\lambda\rangle\}_{\lambda \in \mathcal{S}})|). \quad (83)$$

Intuitively, [Theorem A.2](#) says that the number of ancillary qubits and the number of Pauli postselections are upper bounded by the number of  $T$  gates.

- If  $\mathcal{S}$  is a singleton set and  $|\phi\rangle = |0^n\rangle$ , then  $\text{Stab}(|\phi\rangle) = 2^n$  and  $\text{Stab}(|\psi\rangle) \leq 2^n$ . Thus  $c \leq t$  and [Theorem A.2](#) specializes as [Lemma 4.8](#), which justifies the above intuition for state preparation task, as also proved in [\[BCHK20\]](#).
- If  $\{|\phi_\lambda\rangle\}_{\lambda \in \mathcal{S}}$  ranges over all  $n$ -qubit states with  $|\psi_\lambda\rangle = V|\phi_\lambda\rangle$  for an  $n$ -qubit unitary  $V$ , then  $|\text{Stab}(\{|\phi_\lambda\rangle\}_{\lambda \in \mathcal{S}})| = |\text{Stab}(\{|\psi_\lambda\rangle\}_{\lambda \in \mathcal{S}})| = 1$ . Thus  $c = t$  and [Theorem A.2](#) justifies the above intuition for unitary synthesis of  $V$ .

For intermediate tasks like implementing the first  $K$  columns of a unitary (aka partial unitary synthesis), [Theorem A.2](#) can also be used, by setting  $\{|\phi_\lambda\rangle\}_{\lambda \in \mathcal{S}}$  to be the first  $K$  computational basis states and analyzing the stabilizer group of the corresponding outputs  $\{|\psi_\lambda\rangle\}_{\lambda \in \mathcal{S}}$ .

The proof of [Theorem A.2](#) follows the analysis in [\[BCHK20, Section A.7\]](#). We start with the following convenient fact.

**Fact A.3.** Let  $P, Q$  be two anticommuting Hermitian Pauli operators. Then  $-iPQ$  is Hermitian Pauli and  $\frac{I+PQ}{\sqrt{2}}$  is Clifford.

*Proof.* Consider the Clifford that conjugates  $P$  to  $X \otimes I$  and  $Q$  to  $Z \otimes I$ . Then the result follows.  $\square$

In light of [\(60\)](#) and [\(82\)](#), we prove the following lemma.

**Lemma A.4.** Let  $C_1, \dots, C_{r+1}$  be  $K$ -qubit Clifford unitaries and  $P_1, \dots, P_r$  be  $K$ -qubit Hermitian Pauli operators. Define  $M_j = I + P_j$  for each  $j \in [r]$ . Let  $\{|\rho_\lambda\rangle\}_{\lambda \in \mathcal{E}}$  and  $\{|\tau_\lambda\rangle\}_{\lambda \in \mathcal{E}}$  be two sets of  $K$ -qubit states indexed by  $\mathcal{E}$ . Assume

$$|\tau_\lambda\rangle \propto C_{r+1} M_r C_r \cdots M_1 C_1 |\rho_\lambda\rangle \quad (84)$$

holds for all  $\lambda \in \mathcal{E}$ . Then there exist an integer  $0 \leq r' \leq r$ , a  $K$ -qubit Clifford  $C$ , and  $K$ -qubit Hermitian Pauli operators  $Q_1, \dots, Q_{r'}$  such that the following holds.

1. Let  $A_j = I + Q_j$ . Then  $|\tau_\lambda\rangle \propto C A_{r'} \cdots A_2 A_1 |\rho_\lambda\rangle$  holds for all  $\lambda \in \mathcal{E}$ .
2. Let  $\mathcal{G} = \text{Stab}(\{|\rho_\lambda\rangle\}_{\lambda \in \mathcal{E}})$ . For each  $j \in [r']$ , we have  $Q_j \notin \langle Q_1, \dots, Q_{j-1}, \mathcal{G} \rangle$  and  $Q_j$  commutes with any  $P \in \langle Q_1, \dots, Q_{j-1}, \mathcal{G} \rangle$ .
3.  $|\text{Stab}(\{|\rho_\lambda\rangle\}_{\lambda \in \mathcal{E}})| \cdot 2^{r'} \leq |\text{Stab}(\{|\tau_\lambda\rangle\}_{\lambda \in \mathcal{E}})|$ .

*Proof.* Observe that

$$C_{r+1}M_rC_r \cdots M_1C_1 = C_{r+1}M_rC_r \cdots M_2C_2C_1 \left( C_1^\dagger M_1C_1 \right) \quad (85)$$

$$= C_{r+1}M_rC_r \cdots M_2C_2C_1A_1 \quad (\text{define } A_1 = C_1^\dagger M_1C_1)$$

$$= \cdots = \quad (86)$$

$$= C_{r+1} \cdots C_2C_1 \cdot A_r \cdots A_2A_1 \quad (\text{define } A_j = (C_j \cdots C_1)^\dagger M_j(C_j \cdots C_1))$$

$$= CA_r \cdots A_2A_1. \quad (\text{define } C = C_{t+1} \cdots C_2C_1)$$

Since each  $C_j$  is Clifford, the final  $C$  is also Clifford. Since each  $M_j = I + P_j$  for some Hermitian Pauli  $P_j$ , we have  $A_j = I + (C_j \cdots C_1)^\dagger P_j(C_j \cdots C_1) =: I + Q_j$  where  $Q_j$  is also a Hermitian Pauli. Combined with (84), this already guarantees [Item 1](#).

To ensure [Item 2](#), we perform a clean-up procedure inductively for  $j = 1, 2, \dots$ . To this end, we divide into the following cases.

- If  $Q_j$  anticommutes with some  $R \in \mathcal{G}$ , then we show it can be discarded. Observe that

$$\begin{aligned} (Q_j R) A_{j-1} \cdots A_1 |\rho_\lambda\rangle &= Q_j A_{j-1} \cdots A_1 R |\rho_\lambda\rangle && (\text{by induction and } \text{Item 2}) \\ &= Q_j A_{j-1} \cdots A_1 |\rho_\lambda\rangle. && (\text{since } R \in \mathcal{G} \subseteq \text{Stab}(|\rho_\lambda\rangle)) \end{aligned}$$

Hence we can safely replace  $A_j = I + Q_j$  as  $I + Q_j R \propto \frac{I+Q_j R}{\sqrt{2}}$ , which is Clifford due to [Fact A.3](#). Then we can apply the same trick in (85) to merge with the Clifford  $C$  outside.

- If  $Q_j$  anticommutes with  $Q_{j^*}$  for some  $j^* < j$ , then it can also be discarded. By induction and [Item 1](#), we assume without loss of generality  $j^* = j-1$ . Since  $Q_{j-1}$  is Hermitian Pauli, we have  $Q_{j-1}A_{j-1} = Q_{j-1}(I+Q_{j-1}) = Q_{j-1}+I = A_{j-1}$ . Therefore we can safely replace  $A_j = I + Q_j$  as  $\frac{I+Q_j Q_{j-1}}{\sqrt{2}}$ , which again is Clifford by [Fact A.3](#).
- Now we are in the situation where  $Q_j$  commutes with  $Q_1, \dots, Q_{j-1}$  and elements in  $\mathcal{G}$ . To guarantee [Item 2](#), it suffices to show that if  $Q_j \in \langle Q_1, \dots, Q_{j-1}, \mathcal{G} \rangle$ , then it can be discarded. Assume  $Q_j \in \langle Q_1, \dots, Q_{j-1}, \mathcal{G} \rangle$ . Since  $Q_1, \dots, Q_{j-1}$  and  $\mathcal{G}$  are commuting Hermitian Pauli operators,  $Q_j$  can be expressed as  $R \cdot \prod_{\ell \in L} Q_\ell$  for some  $L \subseteq [j-1]$  and  $R \in \mathcal{G}$ . This means  $Q_j$  lies in the stabilizer group of  $A_{j-1} \cdots A_1 |\rho_\lambda\rangle$  and hence

$$Q_j A_{j-1} \cdots A_1 |\rho_\lambda\rangle = A_{j-1} \cdots A_1 |\rho_\lambda\rangle. \quad (87)$$

This means  $A_j A_{j-1} \cdots A_2 A_1 |\rho_\lambda\rangle \propto A_{j-1} \cdots A_2 A_1 |\rho_\lambda\rangle$  and thus it can be removed directly.

Finally we prove [Item 3](#). By [Item 1](#), we have

$$|\text{Stab}(\{|\tau_\lambda\rangle\}_{\lambda \in \mathcal{E}})| = |\text{Stab}(\{CA_{r'} \cdots A_1 |\rho_\lambda\rangle\}_{\lambda \in \mathcal{E}})| \quad (88)$$

$$= |\text{Stab}(\{A_{r'} \cdots A_1 |\rho_\lambda\rangle\}_{\lambda \in \mathcal{E}})| \quad (89)$$

$$=: |\mathcal{H}|. \quad (\text{since } C \text{ is Clifford})$$

Since  $Q_j A_j = A_j$ , we have  $Q_j \in \mathcal{H}$  for each  $j \in [r']$ . We also have  $P \in \mathcal{H}$  for each  $P \in \mathcal{G}$ . Therefore,  $\langle Q_1, \dots, Q_{r'}, \mathcal{G} \rangle \subseteq \mathcal{H}$ , where, by [Item 2](#), the former set has size  $2^{r'} \cdot |\mathcal{G}|$ . Hence  $|\mathcal{H}| \geq 2^{r'} \cdot |\mathcal{G}|$ . Putting back the definition of  $\mathcal{G}$  and  $\mathcal{H}$  implies [Item 3](#).  $\square$



In addition to [Lemma A.4](#), we need the following technical lemma to remove unnecessary ancillas. [Lemma A.5](#) generalizes [[BCHK20](#), Lemma A.10] and shares a similar proof.

**Lemma A.5.** *Let  $\{|\phi_\lambda\rangle\}_{\lambda \in \mathcal{S}}$  and  $\{|\psi_\lambda\rangle\}_{\lambda \in \mathcal{S}}$  be two sets of  $n$ -qubit states indexed by  $\mathcal{S}$ . Let  $C'$  be an  $(n+m)$ -qubit Clifford unitary where  $m \geq 0$  is an integer. Assume*

$$|\psi_\lambda\rangle|0^m\rangle = C'(|\phi_\lambda\rangle|0^m\rangle) \quad \text{holds for all } \lambda \in \mathcal{S}. \quad (90)$$

*Then there exists an  $n$ -qubit Clifford unitary  $C$  such that  $|\psi_\lambda\rangle = C|\phi_\lambda\rangle$  for all  $\lambda \in \mathcal{S}$ .*

Given [Lemma A.5](#) and [Lemma A.4](#), we first complete the proof of [Theorem A.2](#). The proof of [Lemma A.5](#) is presented in [Subsection A.1](#).

*Proof of Theorem A.2.* By [Definition 4.4](#) and (81),  $\mathcal{C}$  can be expressed by  $(n+t+a)$ -qubit Clifford  $C_1, \dots, C_{m+1}$  and  $(n+t+a)$ -qubit Hermitian Pauli  $R_1, \dots, R_m$  such that

$$|\psi_\lambda\rangle|0^t\rangle|0^a\rangle \propto C_{m+1}W_mC_m \cdots W_1C_1 \left( |\phi_\lambda\rangle|T\rangle^{\otimes t}|0^a\rangle \right) \quad (91)$$

holds for all  $\lambda \in \mathcal{S}$ , where  $W_j = I + R_j$ .

By [Lemma A.4](#), we can further simplify above by an integer  $0 \leq c \leq m$ , an  $(n+t+a)$ -qubit Clifford  $C'$ , and  $(n+t+a)$ -qubit Hermitian Pauli  $Q_1, \dots, Q_c$  such that

$$|\psi_\lambda\rangle|0^t\rangle|0^a\rangle \propto C'A_c \cdots A_2A_1 \left( |\phi_\lambda\rangle|T\rangle^{\otimes t}|0^a\rangle \right) \quad (92)$$

holds for all  $\lambda \in \mathcal{S}$ , where  $A_j = I + Q_j$ . In addition,

$$\begin{aligned} 2^{t+a-c} \cdot |\text{Stab}(\{|\psi_\lambda\rangle\}_{\lambda \in \mathcal{S}})| &= 2^{-c} \cdot \left| \text{Stab} \left( \{|\psi_\lambda\rangle|0^t\rangle|0^a\rangle\}_{\lambda \in \mathcal{S}} \right) \right| \quad (\text{since } \text{Stab}(|0\rangle) = \{I, Z\}) \\ &\geq \left| \text{Stab} \left( \{|\phi_\lambda\rangle|T\rangle^{\otimes t}|0^a\rangle\}_{\lambda \in \mathcal{S}} \right) \right| \quad (\text{by Item 3 of Lemma A.4}) \\ &\geq |\text{Stab}(\{|\phi_\lambda\rangle\}_{\lambda \in \mathcal{S}})| \cdot |\text{Stab}(|0^a\rangle)| \quad (93) \\ &= 2^a \cdot |\text{Stab}(\{|\phi_\lambda\rangle\}_{\lambda \in \mathcal{S}})|, \quad (94) \end{aligned}$$

which implies

$$c \leq t - \log(|\text{Stab}(\{|\phi_\lambda\rangle\}_{\lambda \in \mathcal{S}})|) + \log(|\text{Stab}(\{|\psi_\lambda\rangle\}_{\lambda \in \mathcal{S}})|). \quad (95)$$

By [Item 2](#) of [Lemma A.4](#), we know that each  $Q_j$  commutes with  $\text{Stab}(\{|\phi_\lambda\rangle|T\rangle^{\otimes t}|0^a\rangle\}_{\lambda \in \mathcal{S}})$ , which contains all  $Z$  rotations on the last  $a$ -qubits. Since  $Q_j$  is a Hermitian Pauli, we have

$$Q_j = P_j \otimes Z(v_j) \quad (96)$$

for some  $(n+t)$ -qubit Hermitian Pauli  $P_j$  and some  $v_j \in \{0,1\}^a$ . Here,  $Z(v_j)$  applies the single-qubit  $Z$  gate on the coordinates selected by  $v_j$ . Define  $M_j = I + P_j$ . Then for any  $(n+t)$ -qubit state  $|\rho\rangle$ , we have

$$A_j(|\rho\rangle|0^a\rangle) = (M_j|\rho\rangle) \otimes |0^a\rangle. \quad (97)$$

Therefore,

$$\begin{aligned} |\psi_\lambda\rangle|0^t\rangle|0^a\rangle &\propto C'A_c \cdots A_2A_1 \left( |\phi_\lambda\rangle|T\rangle^{\otimes t}|0^a\rangle \right) \quad (\text{by (92)}) \\ &= C' \left( (M_c \cdots M_2M_1|\phi_\lambda\rangle|T\rangle^{\otimes t}) \otimes |0^a\rangle \right). \quad (\text{by (97)}) \end{aligned}$$

Now define  $|\rho_\lambda\rangle = M_c \cdots M_2 M_1 |\phi_\lambda\rangle |T\rangle^{\otimes t}$ . Then for all  $\lambda \in \mathcal{S}$ ,

$$|\psi_\lambda\rangle |0^t\rangle \otimes |0^a\rangle \propto C'(|\rho_\lambda\rangle \otimes |0^a\rangle).$$

where  $C'$  is Clifford. Hence by [Lemma A.5](#), there exists an  $(n+t)$ -qubit Clifford  $C$  such that

$$|\psi_\lambda\rangle |0^t\rangle \propto C|\rho_\lambda\rangle = CM_c \cdots M_2 M_1 (|\phi_\lambda\rangle |T\rangle^{\otimes t}). \quad (98)$$

To complete the proof of [Theorem A.2](#), we finally remark that it is always possible to pad dummy  $P = I$  to ensure the equality in (95) holds.  $\square$

### A.1 Eliminating ancillas in Clifford circuits

Finally we prove [Lemma A.5](#), which generalizes [[BCHK20](#), Lemma A.10].

**Lemma A.5.** *Let  $\{|\phi_\lambda\rangle\}_{\lambda \in \mathcal{S}}$  and  $\{|\psi_\lambda\rangle\}_{\lambda \in \mathcal{S}}$  be two sets of  $n$ -qubit states indexed by  $\mathcal{S}$ . Let  $C'$  be an  $(n+m)$ -qubit Clifford unitary where  $m \geq 0$  is an integer. Assume*

$$|\psi_\lambda\rangle |0^m\rangle = C'(|\phi_\lambda\rangle |0^m\rangle) \quad \text{holds for all } \lambda \in \mathcal{S}. \quad (90)$$

*Then there exists an  $n$ -qubit Clifford unitary  $C$  such that  $|\psi_\lambda\rangle = C|\phi_\lambda\rangle$  for all  $\lambda \in \mathcal{S}$ .*

*Proof.* Let  $r = \log |\text{Stab}(\{|\phi_\lambda\rangle\})|$ . Since  $\text{Stab}(\{|\phi_\lambda\rangle\})$  is a commutative subgroup of the Pauli group and does not contain  $-I$ , there exists a Clifford  $C_\phi$  conjugating  $\text{Stab}(\{|\phi_\lambda\rangle\})$  to the group of  $Z$  operators on the last  $r$  qubits [[CG97](#)], which means

$$|\phi_\lambda\rangle = C_\phi(|\phi'_\lambda\rangle |0^r\rangle) \quad (99)$$

holds for every  $\lambda \in \mathcal{S}$  and some  $(n-r)$ -qubit state  $|\phi'_\lambda\rangle$ . By (90), we also have

$$\log |\text{Stab}(\{|\psi_\lambda\rangle\})| = r$$

and analogously, there exists a Clifford  $C_\psi$  such that

$$|\psi_\lambda\rangle = C_\psi(|\psi'_\lambda\rangle |0^r\rangle) \quad (100)$$

holds for every  $\lambda \in \mathcal{S}$  and some  $(n-r)$ -qubit state  $|\psi'_\lambda\rangle$ . Now define Clifford  $C'' = (C_\psi \otimes I_m)^\dagger C' (C_\phi \otimes I_m)$ . Then (90) becomes

$$|\psi'_\lambda\rangle |0^{r+m}\rangle = C''(|\phi'_\lambda\rangle |0^{r+m}\rangle) \quad (101)$$

holds for all  $\lambda \in \mathcal{S}$ . In addition,  $\text{Stab}(\{|\phi'_\lambda\rangle\})$  and  $\text{Stab}(\{|\psi'_\lambda\rangle\})$  contain only the trivial stabilizer  $I$ .

We will make another simplification to make sure that  $C''$  commutes with the single-qubit  $Z$  gate on the last qubit, which will help factorize  $C''$  and apply induction. To this end, we observe that  $\text{Stab}(\{|\phi'_\lambda\rangle |0^{r+m}\rangle\}) = \text{Stab}(\{|\psi'_\lambda\rangle |0^{r+m}\rangle\})$  equals the Pauli  $Z$  group on the last  $r+m$  qubits. Hence by (101), for the single-qubit  $Z$  gate on the last qubit we know that

$$C''(I_{n+m-1} \otimes Z)C''^\dagger = I_{n-r} \otimes Z(s) \quad (102)$$

for some  $s \in \{0, 1\}^{r+m}$ . Let  $D$  be an  $(r+m)$ -qubit CNOT circuit reversing this mapping, i.e.,  $DZ(s)D^\dagger$  equals the single-qubit  $Z$  gate on the last qubit. Define Clifford  $C''' = (I_{n-r} \otimes D)C''$ . Then we have the commutativity condition:

$$C'''(I_{n+m-1} \otimes Z)C'''^\dagger = (I_{n-r} \otimes D)(I_{n-r} \otimes Z(s))(I_{n-r} \otimes D)^\dagger \quad (103)$$

$$= I_{n+m-1} \otimes Z. \quad (104)$$

Moreover, (101) remains unchanged:

$$C'''(|\phi'_\lambda\rangle|0^{r+m}\rangle) = (I_{n-r} \otimes D)|\psi'_\lambda\rangle|0^{r+m}\rangle = |\psi'_\lambda\rangle|0^{r+m}\rangle \quad (105)$$

holds for all  $\lambda \in \mathcal{S}$ , where the second equality uses the fact that  $D$  is a CNOT circuit.

At this point, we can factorize  $C'''$  to reduce the number of ancillary qubits. This is achieved by the following [Fact A.6](#) in [\[BCHK20\]](#).

**Fact A.6** ([\[BCHK20, Proposition A.11 and Lemma A.13\]](#)). *Assume  $C$  is a  $K$ -qubit Clifford unitary commuting with the single-qubit  $Z$  gate on the last qubit. Then  $C = C_0 \otimes |0\rangle\langle 0| + C_1 \otimes |1\rangle\langle 1|$ , where  $C_0, C_1$  are  $(K-1)$ -qubit Clifford unitaries.*

Combining [Fact A.6](#) and (103), we know that

$$C''' = C_0 \otimes |0\rangle\langle 0| + C_1 \otimes |1\rangle\langle 1|, \quad (106)$$

where  $C_0$  is an  $(n+m-1)$ -qubit Clifford. Furthermore, by (105), we have

$$C_0(|\phi'_\lambda\rangle|0^{r+m-1}\rangle) = |\psi'_\lambda\rangle|0^{r+m-1}\rangle \quad (107)$$

holds for all  $\lambda \in \mathcal{S}$ .

Iterating the above process to ensure  $C_0$  commutes with the single-qubit  $Z$  gate on the last qubit, we can repeatedly trim off the trailing ancillas until we obtain

$$\tilde{C}(|\phi'_\lambda\rangle|0^r\rangle) = |\psi'_\lambda\rangle|0^r\rangle \quad \text{holds for all } \lambda \in \mathcal{S}, \quad (108)$$

where  $\tilde{C}$  is some  $n$ -qubit Clifford. Recall from (99) and (100) that  $|\phi'_\lambda\rangle|0^r\rangle = C_\phi^\dagger|\phi_\lambda\rangle$  and  $|\psi'_\lambda\rangle|0^r\rangle = C_\psi^\dagger|\psi_\lambda\rangle$ . Hence setting  $C = C_\psi\tilde{C}C_\phi^\dagger$  guarantees  $C|\phi_\lambda\rangle = |\psi_\lambda\rangle$ . This completes the proof of [Lemma A.5](#).  $\square$

## B Improved analysis of Low-Kliuchnikov-Schaeffer state preparation

We observe that the proof of [Theorem 1.2](#) works equally well for diagonals of single-qubit unitaries. With [Fact 2.4](#), we can also decompose arbitrary single-qubit unitary into diagonal unitaries for which [Theorem 1.2](#) directly applies. These are presented as the following [Lemma B.1](#).

**Lemma B.1.** *Let  $n \geq 0$  be an integer and  $U_1, U_2, \dots, U_{2^n}$  be arbitrary single-qubit unitaries. Define  $(n+1)$ -qubit unitary  $U = \text{diag}(U_1, \dots, U_{2^n})$ . Then  $U$  can be implemented up to error  $\varepsilon$  by a Clifford+ $T$  circuit using  $O\left(\sqrt{2^n \log(1/\varepsilon)} + \log(1/\varepsilon)\right)$   $T$  gates and ancillary qubits.*

*Proof.* The proof is identical to the proof [Theorem 1.2](#) once we make sure that the determinant of each  $U_j$  is 1. To achieve this, let  $D = \text{diag}(\alpha_1, \beta_1, \dots, \alpha_{2^n}, \beta_{2^n})$  be an  $(n+1)$ -qubit diagonal unitary such that the determinant of  $U'_j$  is 1 for all  $j$ , where  $U'_j = \text{diag}(\alpha_j, \beta_j)U_j$ . Note that  $D^\dagger$  can be constructed using [Theorem 1.2](#), and then  $DU = \text{diag}(U'_1, \dots, U'_{2^n})$  can be handled in the same way as in [Theorem 1.2](#). Combining both gives  $D^\dagger(DU) = U$ .

An alternative proof is to use [Fact 2.4](#), which represents each  $U_j$  as  $A_j H B_j H C_j$  for single-qubit diagonal unitaries  $A_j, B_j, C_j$ . Then we have

$$U = \text{diag}(A_1, \dots, A_{2^n}) \text{diag}(H, \dots, H) \quad (109)$$

$$\cdot \text{diag}(B_1, \dots, B_{2^n}) \quad (110)$$

$$\cdot \text{diag}(H, \dots, H) \text{diag}(C_1, \dots, C_{2^n}), \quad (111)$$

where  $\text{diag}(H, \dots, H)$  is simply a Hadamard gate on the  $(n+1)$ -th qubit and, moreover,  $\text{diag}(A_1, \dots, A_{2^n})$  (resp.,  $\text{diag}(B_1, \dots, B_{2^n}), \text{diag}(C_1, \dots, C_{2^n})$ ) can be implemented by [Theorem 1.2](#).  $\square$

As a direct application of [Lemma B.1](#), we improve the state synthesis bound in [\[LKS24\]](#). Note that [Corollary B.2](#) is only off by the extra  $n$  in front of  $\log(1/\varepsilon)$ , which does not dominate the overall complexity unless  $\varepsilon$  is extremely small. Given this, this construction may be of practical interest.

**Corollary B.2.** *Any  $n$ -qubit state can be prepared up to error  $\varepsilon$  by a Clifford+ $T$  circuit starting with the all-zeros state using*

$$O\left(\sqrt{2^n \log(1/\varepsilon)} + n \log(1/\varepsilon)\right) T \text{ gates} \quad (112)$$

and

$$O\left(\sqrt{2^n \log(1/\varepsilon)} + \log(1/\varepsilon)\right) \text{ ancillary qubits.} \quad (113)$$

*Proof.* The state synthesis algorithm in [\[LKS24\]](#) uses Grover-Rudolph algorithm [\[GR02\]](#): The first qubit is rotated to have the correct marginal. Then conditioned on the first qubit, the second qubit is rotated to have the correct marginal. This process is iterated until the  $n$ th qubit is rotated conditioned on the first  $n-1$  qubits.

Let  $k = 0, 1, \dots, n-1$ . The  $k$ -th operation above is a rotation on the  $(k+1)$ -th qubit, controlled by the first  $k$  qubits. Therefore, it is a diagonal of single-qubit unitaries, for which we apply [Lemma B.1](#). Let  $\varepsilon_k = \varepsilon/2^{n-k}$ . Then by [Lemma B.1](#), the  $k$ -th operation above can be  $\varepsilon_k$ -approximately implemented with

$$O\left(\sqrt{2^k \log(1/\varepsilon_k)} + \log(1/\varepsilon_k)\right) = O\left(\sqrt{2^k (n-k + \log(1/\varepsilon))} + n-k + \log(1/\varepsilon)\right) \quad (114)$$

$T$  gates and ancillas.

Through all  $k = 0, 1, \dots, n-1$ , the maximal number of ancillas is

$$O\left(\max_k \sqrt{2^k (n-k + \log(1/\varepsilon))} + n-k + \log(1/\varepsilon)\right) = O\left(\sqrt{2^k \log(1/\varepsilon)} + \log(1/\varepsilon)\right). \quad (115)$$

The total error is

$$\sum_{k=0}^{n-1} \varepsilon_k = \varepsilon \cdot \sum_{k=0}^{n-1} 2^{k-n} \leq \varepsilon, \quad (116)$$

and the total number of  $T$  gates used is

$$\begin{aligned} & \sum_{k=0}^{n-1} O\left(\sqrt{2^k (n-k + \log(1/\varepsilon))} + n-k + \log(1/\varepsilon)\right) \\ & \leq \sum_{k=0}^{n-1} O\left(\sqrt{2^k (n-k)} + \sqrt{2^k \log(1/\varepsilon)} + \log(1/\varepsilon)\right) \quad (\text{since } \sqrt{a+b} \leq \sqrt{a} + \sqrt{b}) \\ & = O\left(\sqrt{2^n \log(1/\varepsilon)} + n \log(1/\varepsilon)\right). \end{aligned} \quad (117) \quad \square$$