Trading $T$ gates for dirty qubits in state preparation and unitary synthesis

Guang Hao Low$^{1,2}$, Vadym Kliuchnikov$^{1,2}$, and Luke Schaeffer$^{1,3,4}$

1Quantum Architectures and Computation, Microsoft Research, Washington, Redmond, USA
2Azure Quantum, Microsoft, Washington, Redmond, USA
3Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
4Joint Center for Quantum Information and Computer Science, University of Maryland, Maryland, College Park, USA

Efficient synthesis of arbitrary quantum states and unitaries from a universal fault-tolerant gate-set e.g. Clifford+$T$ is a key subroutine in quantum computation. As large quantum algorithms feature many qubits that encode coherent quantum information but remain idle for parts of the computation, these should be used if it minimizes overall gate counts, especially that of the expensive $T$-gates. We present a quantum algorithm for preparing any dimension-$N$ pure quantum state specified by a list of $N$ classical numbers, that realizes a trade-off between space and $T$-gates. Our scheme uses $O(\log (N/\epsilon))$ clean qubits and a tunable number of $\sim (\lambda \log \left( \frac{\log N}{\epsilon} \right))$ dirty qubits, to reduce the $T$-gate cost to $O\left( \frac{N}{\lambda} + \lambda \log N \log \log \frac{N}{\epsilon} \right)$. This trade-off is optimal up to logarithmic factors, proven through an unconditional gate counting lower bound, and is, in the best case, a quadratic improvement in $T$-count over prior ancillary-free approaches. We prove similar statements for unitary synthesis by reduction to state preparation. Underlying our constructions is a $T$-efficient circuit implementation of a quantum oracle for arbitrary classical data.

Contents

1 Introduction 2
2 Data-lookup oracle by a SelectSwap network 3
3 Arbitrary quantum state preparation 5
4 Unitary synthesis by state preparation 6
5 Lower bound 6
6 Conclusion 7
7 Developments after 2018 release of preprint 8
A Purified density matrix preparation 10
B Data-lookup oracle implementation details 11
C Indicator Function Construction 15
D Pure-state preparation implementation details 16

Accepted in Quantum 2024-05-22, click title to verify. Published under CC-BY 4.0.
Above, \( O \) regression [6] queries a classical list of \( N \) quantum speedups for applications such as machine learning [1], simulation of physical systems [2, 3] queried in superposition through a unitary quantum operation. This property is crucial in obtaining 1 Introduction

Many quantum algorithms require coherent access to classical data, that is, data that can be queried in superposition through a unitary quantum operation. This property is crucial in obtaining quantum speedups for applications such as machine learning [1], simulation of physical systems [2, 3] and solving systems of linear equations [4, 5].

The nature of quantum-encoded classical data is itself varied. For example, quantum data regression [6] queries a classical list of \( N \) data-points \( a_x \) through a unitary data-lookup oracle [7]

\[
O|x\rangle|0\rangle = |x\rangle|a_x\rangle. \tag{1}
\]

Other applications, particularly in quantum chemistry [8] instead access Hamiltonian coefficient data through a unitary \( A|0\rangle = |\psi\rangle \) that prepares these numbers as amplitudes in a normalized quantum state, or as probabilities in a purified density matrix. Even more generally, the central challenge is synthesizing some arbitrary unitary \( A \in \mathbb{C}^{N \times N} \) of which \( k \leq N \) columns are either partially or completely specified by a list of complex coefficients that is, say, provided on paper.

Synthesis of these data-access unitaries is typically a dominant factor in the overall algorithmic cost. In any scalable approach to quantum computation, all unitaries decompose into a universal fault-tolerant quantum gate set, such as Clifford gates \{H, S, CNOT\} and T gates [9]. Solovay and Kitaev [9] were the first to recognize that any single-qubit unitary could be \( \epsilon \)-approximated using \( \mathcal{O}(\log^c(1/\epsilon)) \) fault-tolerant gates for \( c = 3.97 \), which was later improved to \( c = 1 \) [10, 11]. By bootstrapping these results, it is well-known that a roughly equal number of \( \mathcal{O}(N \log(N/\epsilon)) \) [12] Clifford and non-Clifford gates suffice for arbitrary dimensions. Notably, the total gate count scaling is optimal in all parameters, following gate-counting arguments [13].

The possibility that T gates could be substantially fewer in number than the Clifford gates, however, is not excluded by known lower bounds. It is believed that fault-tolerant Clifford gates \{H, S, CNOT\} will be cheap in most practical implementations of fault-tolerant quantum computation. In contrast, the equivalent cost of each fault-tolerant non-Clifford T gates, implemented at machine precision, is placed at a space-time volume \( \approx (225 \text{ logical qubits}) \times (10 \text{ Clifford depth}) \) for realistic estimates [14] based on \( |T| \) magic-state distillation at a physical error rate of \( \approx 10^{-3} \).

We present an approach to arbitrary quantum state preparation and unitary synthesis that focuses on minimizing the T count. Unique to our approach is the exploitation of a variable number \( \mathcal{O}(\lambda \log(1/\epsilon)) \) of ancillary qubits, in a manner not considered by prior gate-counting arguments or algorithms. We find a \( \mathcal{O}(\lambda) \) improvement in the T count and circuit depth while keeping the Clifford count unchanged, excluding logarithmic factors. Most surprisingly, its benefit far exceeds the naïve approach of applying these ancillary qubits to producing \(|T| \) magic-states for any \( \lambda = \mathcal{O}(\sqrt{N}) \), as seen in Table 1. In the best-case, the T count of \( \mathcal{O}((\sqrt{N}) \) is a square-root factor smaller than prior art, such as for for preparing arbitrary pure states

\[
|\psi\rangle = \sum_{x=0}^{N-1} \frac{a_x}{\|\vec{a}\|_2} |x\rangle, \quad \|\vec{a}\|_q = \left( \sum_{j=0}^{N-1} |a_x|^q \right)^{1/q}, \tag{2}
\]

| Output \( |\psi\rangle \) | Qubits \( N \) | T count \( \lambda \) | T Depth \( \lambda \) | Valid \( \lambda \) |
|---|---|---|---|---|
| \( |\psi\rangle \) | \( \log N \) | \( N \log \frac{N}{\lambda} \) | \( \frac{N}{\log N} \) \( \log N \) | \( \mathcal{O}(\log N) \cup \Omega(N) \) |
| \( |\psi\rangle \) | \( \log N + \lambda \) | \( N \log \frac{N}{\lambda} \) | \( \frac{N}{\log N} \) \( \log N \) | \( \mathcal{O}(\log N) \cup \Omega(N) \) |

Table 1: Big-\( \mathcal{O}(\cdot) \) cost of preparing to error \( \epsilon \) a dimension \( N \) arbitrary quantum state \( |\psi\rangle \) Eq. (2), quantum state with garbage \( |\psi_{gb}\rangle \) Eq. (4), and an \( N \times N \) unitary \( U \) Eq. (3) where \( K \leq N \) columns are fully specified. Above, \( \epsilon' = \frac{1}{\log N} \), and we have rescaled \( \lambda b \rightarrow \lambda \) where appropriate to facilitate comparison.
Table 2: Upper bounds on cost of possible implementations of the data lookup oracle \( O \) of Eq. (5). See Appendices B.1, B.2 and C for other variations, such as a linear-depth phase-incorrect version of \( \text{Swap} \) using \( \leq 4bN \) \( T \) gates and no additional ancillary qubits. Our results allow for a space-depth trade-off determined by a choice of \( \lambda \in [1,N] \), with a minimized \( T \) gate complexity of \( O(\sqrt{bN}) \) by choosing \( \lambda = O(\sqrt{N/b}) \). Note that \( b\lambda \) qubits of the Fig. 1d implementation may be dirty.

Moreover, we prove this approach realizes an optimal ancillary qubit and \( T \) count trade-off up to log factors.

In particular, our approach is always advantageous as all but a logarithmic number \( O(\log (N/\epsilon)) \) of qubits, independent of \( \lambda \), may be dirty, meaning that they start in, and are returned to the same undetermined initial state. At first glance, the full quadratic speedup is not always desirable as any clean ancillary qubit, initialized in the \( |0\rangle \) state, is a resource that may be better allocated to magic-state distillation. However, dirty qubits may not be used for magic-state distillation, and are a resource typically abundant in many algorithms, such as quantum simulation by a linear combination of unitaries [2]. Even in the most pessimistic scenario where no dirty qubits are available, a reduction in the overall execution time of the algorithm, including the effective cost of magic-state distillation, is possible.

We also consider applications of our approach. For instance, a similar speedup to unitary synthesis

\[
U = \left( \sum_{k=0}^{K} |u_k\rangle \langle k| \right) + \cdots,
\]

where \( K \leq N \) columns are specified, follows by a well-known reduction based on Householder reflections. Improvements to state preparation with garbage

\[
|\psi_{gb}\rangle = \sum_{x=0}^{N-1} \sqrt{\frac{a_x}{\|\vec{a}\|_1}} |x\rangle |\text{garbage}_x\rangle.
\]

relevant to the most advanced quantum simulation techniques [16, 17, 8] are also presented in Appendix A.

Underlying our \( T \) gate scaling results is an improved implementation of a data lookup oracle

\[
O|x\rangle|0\rangle|0\rangle = |x\rangle|a_x\rangle|\text{garbage}_x\rangle.
\]

Note the attached garbage state may always be uncomputed by applying \( O \) in reverse. We begin by describing our implementation of Eq. (5), which we call a ‘\( \text{SelectSwap} \)’ network, with costs outlined in Table 2. Subsequently, we apply \( \text{SelectSwap} \) to the state preparation problem using the fact that there exists classical data such that preparing any \( |\psi\rangle \) requires only \( O(\text{polylog}(N)) \) queries and additional primitive quantum gates [18]. The reduction of unitary synthesis to state preparation is then described. Finally, we prove optimality of our approach through matching lower bounds, and discuss the results.

2 Data-lookup oracle by a \( \text{SelectSwap} \) network

The unitary data-lookup oracle of Eq. (5) accepts an input number state \( |x\rangle \in \mathbb{C}^N \) where \( x \in \{0,1,\ldots, N-1\} \equiv [N] \), and returns an arbitrary \( b \)-bit number \( a_x \in \{0,1\}^b \). Our approach combines a multiplexer implementation of \( O \) [20], called \( \text{Select} \) and a unitary swap network \( \text{Swap} \), with costs summarized in Table 2.
The Select operator applies some arbitrary unitary \( U_x \) controlled by the index state \( |x⟩ \), that is

\[
\text{SELECT} = \sum_{x=0}^{N-1} |x⟩⟨x| ⊗ U_x.
\] (6)

Thus \( O \) is realized by choosing \( U_x = X^{a_x} \equiv \bigotimes_{j=0}^{\lambda-1} X^{a_{x,j}} \) [8] to either be identity or the Pauli-\( X \) gate depending on the bit string \( a_x \). As described in Fig. 1a, the costs, excluding \( \{U_x\} \), is \( O(N) \) Clifford+T gates. As controlled-\( X \) is Clifford too, an additional \( O(bN) \) Clifford gates are applied. These CNOTs may be applied in logarithmic depth using an ancillary qubit free quantum fanout discussed in Appendix B.1.

The unitary Swap network moves a \( b \)-qubit quantum register indexed by \( x \) to the \( x = 0 \) register, controlled by the state \( |x⟩ \). For any quantum states \( \bigotimes_{x=0}^{N-1} |φ_x⟩_x \),

\[
\text{SWAP} \left[ \bigotimes_{x=0}^{N-1} |φ_x⟩_x \right] = \bigotimes_{x=0}^{N-1} |φ_x⟩_0 ⊗ \cdots,
\] (7)

where the remaining quantum states (\( \cdots \)) in registers \( x > 0 \) are unimportant.

As illustrated in Fig. 1b, this decomposes into a network of controlled-swap operators. As each controlled-swap operator decomposes into two CNOTs and one TOFFOLI, this network uses \( O(bN) \) Clifford+T gates. An ancillary qubit free logarithmic-depth version of SWAP is discussed in Appendix B.2.

Our SELECT-SWAP network illustrated in Fig. 1c is a simple hybrid of the above two schemes. Similar to the SWAP approach, we duplicate the \( b \)-bit register \( λ \) times, where \( λ \in \{1, \cdots, N\} \) is an integer. For \( λ \) that is not a power of 2, we compute \( |x⟩ \rightarrow |q⟩|r⟩ \), which is the quotient \( q = x/λ \) and remainder \( r = x \mod λ \). This contributes an additive cost of \( O(\log N \log λ) \) gates. SELECT is controlled by \( |q⟩ \) to write multiple values of \( a_x \) simultaneously into these duplicated registers by choosing \( U_x = \bigotimes_{j=0}^{\lambda-1} X^{a_{x/Nλ+j}}, \) where \( x \in \lfloor N/λ \rfloor \). SWAP is then controlled \( |r⟩ \) to move the desired
data entry $|a_x\rangle$ to the output register. As the T gate complexity of $O(\lambda b + \frac{N}{c})$ is determined only by the dimension of the SELECT and SWAP control registers, this is minimized with value $O(\sqrt{Nb})$ at $\lambda = O(\sqrt{N}b)$.

Importantly, all but $b + [\log_2 N]$ of the qubits may be made dirty using a simple modification shown in Fig. 1d. Then for any computational basis state $|\phi\rangle = \otimes_{r=1}^{\lambda} |\phi_r\rangle_r$, and any input state $|x\rangle = |q\rangle |r\rangle$, let us evaluate $|0\rangle |\phi\rangle$ at each dotted line:

$$
\begin{align*}
|0\rangle |\phi\rangle \rightarrow |0\rangle |\phi_r \oplus a_x \rangle_0 \cdots \rightarrow |\phi_r \oplus a_x \rangle |\phi_r \oplus a_x \rangle_0 \cdots \\
\rightarrow |\phi_r \oplus a_x \rangle |\phi\rangle \rightarrow |\phi_r \oplus a_x \rangle |\phi_r \rangle_0 \cdots \\
\rightarrow |a_x \rangle |\phi_r \rangle_0 \cdots \rightarrow |a_x \rangle |\phi\rangle \\
\end{align*}
$$

By linearity, this is true for all quantum states $|\phi\rangle$.

As the T gate complexity begins to increase with sufficiently large $\lambda$, one may simply elect to not use excess available dirty qubits. However, continued reduction of the T depth down to $O(\log N)$ might be a useful property. In Appendix C we discuss an alternate construction that achieves logarithmic T depth and preserves the quadratic T count improvement for larger $\lambda = \Omega(\sqrt{N})$.

3 Arbitrary quantum state preparation

Preparation of an arbitrary dimension $N = 2^n$ quantum state $|\psi\rangle = \frac{1}{\sqrt{|a|}} \sum_{x \in \{0,1\}^n} a_x |x\rangle$ using the data-lookup oracle $O$ of Eq. (5) is well-known in prior art. The basic idea was introduced by [21], and an ancillary-free implementation was presented in [12]. We outline the inductive argument of [18], and evaluate its cost using our SELECT-SWAP implementation of $O$.

For any bit-string $y \in \{0,1\}^w$ of length $w \leq n$, let the probability that the first $w$ qubits of $|\psi\rangle$ are in state $|y\rangle$ be $p_y = \frac{1}{|y|_2^2} \sum_{x: \text{prefix}_w(x) = y} |a_x|^2$. Thus a single-qubit rotation $e^{-i\theta |0\rangle \langle 0|}$ by angle $\theta = \cos^{-1} \sqrt{p_0}$ prepares the state $|\psi_1\rangle = \sqrt{p_0} |0\rangle + \sqrt{p_1} |1\rangle$, where $p_0$ is the probability that the first qubit of $|\psi\rangle$ is in state $|0\rangle$. We then recursively apply single-qubit rotations on the $(w+1)^{th}$ qubit conditioned on the first $w$ qubits being in state $|y\rangle$. The rotation angles $\theta_y = \cos^{-1} \sqrt{p_{0y}/p_y}$ are chosen so that the state produced $|\psi_{w+1}\rangle$ reproduces the correct probabilities on the first $w + 1$ qubits.

These conditional rotations are implemented using a sequence of data-lookup oracles $O_1, \cdots, O_{n-1}$, where $O_w$ stores a $b$-bit approximation of all $\theta_y$ where $y \in \{0,1\}^w$. At the $w^{th}$ iteration,

$$
|\psi_w\rangle = \sum_{y \in \{0,1\}^w} \sqrt{p_y} |y\rangle \rightarrow_{O_w} \sum_{y \in \{0,1\}^w} \sqrt{p_y} |y\rangle |\theta_y\rangle \\
\rightarrow \sum_{y \in \{0,1\}^w} \sqrt{p_y} |y\rangle \left( \sqrt{\frac{p_{0y}}{p_y}} |0\rangle + \sqrt{\frac{p_{1y}}{p_y}} |1\rangle \right) |\theta_y\rangle \\
\rightarrow \sum_{y \in \{0,1\}^{w+1}} \sqrt{p_y} |y\rangle = |\psi_{w+1}\rangle.
$$

Note that we omit any garbage registers as they are always uncomputed. Also, the second line is implemented using $b$ single-qubit rotations each controlled by a bit of $\theta_y$. The complex phases of the target state $|\psi\rangle$ are applied to $|\psi_n\rangle$ by a final step with a data-lookup oracle storing $\phi_x = \arg [a_x / \sqrt{p_x}]$. Thus $O(b \log N)$ single-qubit rotations are applied in total.

We implement these oracles with the SELECT-SWAP network of Fig. 1, using a fixed value of $\lambda$ for all $O_k$. A straightforward sum over the T count of Fig. 1 is $O(b \lambda \log (N) + \frac{N}{c})$, which is then added to the total T count of $O(b \log (\frac{N}{c}))$ for synthesizing all single-qubit rotations each to error $\delta$ using the phase gradient technique [22], outlined in Appendix D. The error of the resulting state $|\psi\rangle$ produced is determined by the number of bits $b$ used to represent the rotation angles, in addition to rotation synthesis errors $\delta$. Adding these errors leads to

$$
\| |\psi\rangle - |\psi\rangle \| = O(\delta) + O(2^{-b} \log N) \leq \epsilon,
$$

Accepted in Quantum 2024-05-22, click title to verify. Published under CC-BY 4.0.
which is bounded by $\epsilon$ with the choice $b = \Theta(\log(\frac{\log N}{\epsilon}))$ and $\delta = \Theta(\epsilon)$. As a function of $\epsilon$, the total $T$ gate complexity is then $O\left(\frac{N}{\epsilon} + b(\lambda n + b)\right)$, where

$$b(\lambda n + b) = O\left(\lambda \log\left(\frac{N}{\epsilon}\right) \log\left(\frac{\log N}{\epsilon}\right)\right)$$

$$= O\left(\lambda \log^2\left(\frac{N}{\epsilon}\right)\right),$$

and is plotted in Fig. 2.

### 4 Unitary synthesis by state preparation

The ability to prepare arbitrary quantum states enables synthesis of arbitrary unitaries $U \in \mathbb{C}^{N \times N}$. Given the matrix elements $\{[u_k] \mid k \in [K]\}$ for the first $K$ columns of $U$, the isometry synthesis problem is to find a quantum circuit that implements a unitary $V$ that approximates $U$ in the first $K$ columns to error $\|U - V\|_{2, K} = \max_{x \in K} \|U(x) - V(x)\| \leq \epsilon$.

We use the Householder reflections decomposition [24] to find a $V$ that is a product of $K$ reflections $I - 2|v_k\rangle\langle v_k|$ and a diagonal gate $\text{diag}(e^{i\phi_1}, \ldots, e^{i\phi_K}, 1, \ldots, 1)$, for some set of quantum states $|v_k\rangle$. Note that this representation is not unique. The diagonal gate can be eliminated by using one ancillary qubit as discussed in [25]. There, it suffices to implement the unitary $\tilde{W} = |0\rangle\langle 1| \otimes U + |1\rangle\langle 0| \otimes U^\dagger$, which is equal to the product of reflections $I - 2|w_k\rangle\langle w_k|$ where $|w_k\rangle = (|1\rangle \otimes |k\rangle - |0\rangle \otimes |u_k\rangle)/\sqrt{2}$ for $k \in [K]$.

Given a state preparation unitary $A_k|0\rangle = |u_k\rangle$, one can prepare state $|w_k\rangle$ starting from $|0\rangle \otimes |e_1\rangle$. Apply Hadamard to the first qubit, then a sequence of CNOT gates to prepare $|1\rangle \otimes |k\rangle + |0\rangle \otimes |0\rangle$, and finally apply controlled-$A_k$ negatively controlled on the first qubit. Note that when this method is applied to the synthesis of sparse isometries, the states being synthesized are again sparse. Moreover, the cost of converting a state into a reflection doubles the number of non-Clifford gates. Thus the number of $T$ gates used to synthesize an isometry is twice that for all the Controlled-$A_k$ operations, and scales like $O\left(K \left(\frac{N}{\epsilon} + \lambda \log\left(\frac{N}{\epsilon}\right) \log\left(\frac{\log N}{\epsilon}\right)\right)\right)$.

### 5 Lower bound

We prove the optimality of our construction through a circuit counting argument. The most general circuit on $q$ qubits that uses $\Gamma$ $T$-gates has the canonical form $[26] C \cdot \prod_{j=1}^q e^{-i\pi P_j/8}$, where each $P_j$ is one of $4^q$ possible Pauli operators, and $C$ is one of $2^{O(q^2)}$ possible Clifford operators. Thus

![Figure 2: T gate count dependence on number of dirty qubits exploited for approximating an arbitrary quantum state of dimension $N = 10^{12,4,6,8}$ to error $10^{-3}$ using our algorithm (dots) in comparison with the standard ancillary-free approach (dashed) [12]. Note that at this error, $b \geq 17$ qubits are required to represent each coefficient in binary, but this may be halved by randomization techniques [23]. Moreover, one may always use fewer than the maximum number of available dirty qubits.](image)
the number of unique quantum circuits is at most

$$\text{Unique quantum circuits} = \mathcal{O}(4^\Gamma + \mathcal{O}(\sqrt{N})).$$

(13)

A lower bound on the qubit and T-gate complexity of the data-lookup oracle of Eq. (5) is obtained by counting the number of unique Boolean functions $f : [N] \rightarrow \{0,1\}^k$. As there $2^{KN}$ such functions, we compare with Eq. (13). This leads to a lower bound on the space-T gate product

$$q\Gamma = \Omega(bN - q^2).$$

(14)

As the SelectSwap complexity in Fig. 1 is $q\Gamma = \mathcal{O}(\lambda^2 b^2 + bN + \log(1/\epsilon)(1/\lambda + \lambda b))$, this is optimal up to logarithmic factors so long as the number of T-gates dominates the qubit count like $\lambda = o(\sqrt{N}/b)$, which is the case in most quantum circuits of interest.

A similar lower bound on state preparation is obtained by counting the number of dimension-$N$ quantum states that a distinguishable with error $\epsilon$. Without loss of generality, we only count quantum states $|\psi\rangle \in \mathbb{R}^N$ with real coefficients. These states live on the surface a unit-ball $\mathbb{B}_N$ of dimension $N$, with area $\text{Area}[\mathbb{B}_N] = \frac{2^{N/2}}{\sqrt{\pi N}}$. Let us now fix a state $|\psi\rangle$. Then the states $|\chi\rangle$ that satisfy $|||\psi\rangle - |\chi\rangle|| \leq \epsilon$ live inside a $\epsilon$-ball $\mathbb{B}_{N-1}$ with volume $\mathcal{O}(\text{Vol}[\mathbb{B}_{N-1}]) = \mathcal{O}(\frac{N^{N/2}}{2^{N-1}})$ [27]. Thus there are at least $\Omega(\frac{\text{Area}[\mathbb{B}_N]}{\text{Vol}[\mathbb{B}_{N-1}]}) = \Omega(\sqrt{N/\epsilon})$ quantum states. Once again by comparing with Eq. (13), we obtain a $T$-gate lower bound of

$$q\Gamma = \Omega(N \log(1/\epsilon) - q^2).$$

(15)

This also matches the cost of our approach in Eq. (13) up to logarithmic factors, so long as $\lambda = o(\sqrt{N}/\log(1/\epsilon))$. The total number of isometries within at least distance $\epsilon$ from each other can also be estimated using Lemma 4.3 on Page 14 in [28], and is roughly $\Omega((1/\epsilon)^{KN})$. An analogous argument can be made for state preparation with garbage by considering the unit simplex instead of the unit ball.

Let us now establish a lower bound on state preparation that holds when measurements and arbitrary number of ancillae are used. For the purpose of the lower bound we also allow the use of post-selected measurements of multiple-qubit Pauli observables. Every preparation of an $n$-qubit state by Clifford+T circuit with ancillae and post-selected Pauli measurements can be rewritten [29] as the following sequence of operations: 1) initialization of $\Gamma$ qubits into T state $|0\rangle + e^{i\pi/4}\langle 1|$; 2) post-selected measurement of $\Gamma - n$ commuting Pauli observables; 3) application of a Clifford unitary on $\Gamma$ qubits. After the three steps first $\Gamma - n$ qubits are in a zero state and last $n$ qubits are in the state being prepared. Let us count the number of distinct states that can be prepared by the steps described above. For the step two, there are at most $2 \cdot 4^\Gamma$ ways of choosing first $\Gamma - n$ qubits observable and at most $4^\Gamma$ ways of choosing each of the remaining $\Gamma - n - 1$ observables because each of them needs to commute with the first observable. Therefore, on step two we have at most $2 \cdot 4^\Gamma(\Gamma - n)$ choices of Pauli observables. For the step three, two distinct Clifford unitaries can lead to preparing the same state and counting total number of Clifford unitaries on $\Gamma$ qubits leads to an overestimate. The prepared $|\psi\rangle$ state is completely described by $4^n$ numbers $\alpha_P = \text{Tr}(P|\psi\rangle\langle\psi|)$ where $P$ goes over all $n$-qubit Pauli matrices $\{I, X, Y, Z\}^\otimes n$. Let us count how many distinct $4^n$ dimensional vectors of $\alpha_P$ we can get on the step three by applying different Clifford unitaries $C$. Let $\rho'$ be a density matrix describing a state of all qubits after step two, then $\alpha_P = \text{Tr}(C\rho'C^\dagger(I_{\Gamma-n} \otimes P))$ which is equal to $\text{Tr}(\rho'C^\dagger(I_{\Gamma-n} \otimes P)C)$. We see that the vector of $\alpha_P$ is uniquely defined by action of $C$ on $2n$ Pauli operators $Z_{\Gamma-n+k}, X_{\Gamma-n+k}$ for $k$ from 1 to $n$. There are at most $2 \cdot 4^{2n}$ ways of choosing the action of $C$ on the listed Pauli operators. Therefore we can prepare at most $2 \cdot 4^{\Gamma(\Gamma+n)}$ distinct states. This leads to the lower bound on the number of required T gates $\Omega\left(\sqrt{N} \log N \log(1/\epsilon)\right)$.

6 Conclusion

We have shown that arbitrary quantum states with $N$ coefficients, or unitaries with $KN$ values specified by classical data may be synthesized with a T gate complexity that is an optimal $\mathcal{O}(\sqrt{N})$.
reduction over prior art. As these subroutines are ubiquitous in many quantum algorithms we expect this result to be widely applicable.

We also expect our approach to be practical due to its almost exclusive usage of dirty qubits, which are typically abundant in larger quantum algorithms. Though our results are asymptotically optimal, constant factor and logarithmic improvements in costs could still be possible through careful optimization [30, 31]. For instance, our approach can be modified to use only $O(1)$ additional clean qubits, but this increases the $T$ count by a logarithmic factor. As more limited trade-offs between $T$ gates and ancillary qubits are observed in other quantum circuits, such as for addition [22] or AND [19], a major open question highly relevant to implementation in nearer-term quantum computers, is whether such a property could be generic for many other quantum circuits and algorithms.

6.1 Acknowledgements
We thank Nicolas Delfosse, Jeongwan Haah, Robin Kothari, Jessica Lemieux, Martin Roetteler, and Matthias Troyer for insightful discussions.

7 Developments after 2018 release of preprint

There have been numerous improvements in trade-offs [15] between qubits, depth, and overall spacetime volume [32] of quantum circuits for state preparation and unitary synthesis [15], especially in connection to the block-encoding framework [16, 33] for matrices represented by classical data [34]. A notable new direction is error-resilient table-lookup [35], which achieves logarithmic error scaling by using a large number of $O(N)$ qubits. However, minimizing use of expensive non-Clifford gates is rarely a priority in these results, and almost all exhibit the same scaling for number of Clifford and non-Clifford gates. We have updated Table 1 with a comparison to recent work [15] that claims an optimal qubit–depth trade–off. To facilitate comparison, we rescale $\lambda$ and we use the tighter bound Eq. (11) on the $T$ gate cost that was originally Eq. (12).

Our SelectSwap architecture of table lookup remains state-of-art. To date, all methods with an asymptotic advantage in non-Clifford gate cost for state preparation or unitary synthesis [34] do so by reduction to SelectSwap. The circuit implementation of SelectSwap has also seen some optimizations. By allowing intermediate circuit measurements, the multi-target CNOT gates in Appendix B.1 can be implemented in constant, rather than logarithmic depth, though at the cost of using an additional $O(n)$ clean qubits. Intermediate measurements also enable uncomputing the garbage register Eq. (5) of SelectSwap using $4\lceil \frac{n}{\lambda} \rceil + 4\lambda$ T gates and $\lambda + \lceil \log (N/\lambda) \rceil$ clean qubits, which is independent of $b$ [36].

The $T$ count reductions enabled by SelectSwap have been key to numerous state-of-art resource estimates for quantum computing applications, such as in chemistry [36, 37, 38, 39], and other classical-data-intense routines [40, 34].

References

[1] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. “Quantum principal component analysis”. *Nature Physics* 10, 631–633 (2014).
[2] Dominic W. Berry, Andrew M. Childs, Richard Cleve, Robin Kothari, and Rolando D. Somma. “Simulating Hamiltonian dynamics with a truncated Taylor series”. *Physical Review Letters* 114, 090502 (2015).
[3] Guang Hao Low and Isaac L. Chuang. “Optimal Hamiltonian simulation by quantum signal processing”. *Physical Review Letters* 118, 010501 (2017).
[4] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. “Quantum algorithm for linear systems of equations”. *Physical Review Letters* 103, 150502 (2009).
[5] Chunhao Wang and Leonard Wossnig. “A quantum algorithm for simulating non-sparse Hamiltonians”. *Quantum Information & Computation* 20, 597–615 (2020).
[6] Nathan Wiebe, Daniel Braun, and Seth Lloyd. “Quantum algorithm for data fitting”. *Physical Review Letters* 109, 050505 (2012).
[7] Dorit Aharonov and Amnon Ta-Shma. “Adiabatic quantum state generation and statistical zero knowledge”. Proceedings of the thirty-fifth ACM symposium on Theory of computing, Page 20 (2003).

[8] Ryan Babbush, Craig Gidney, Dominic W. Berry, Nathan Wiebe, Jarrod McClean, Alexandru Paler, Austin Fowler, and Hartmut Neven. “Encoding electronic spectra in quantum circuits with linear T complexity”. Physical Review X 8, 041015 (2018).

[9] Michael A. Nielsen and Isaac L. Chuang. “Quantum computation and quantum information”. Cambridge University Press, (2010). 10th anniversary edition.

[10] Vadym Klüchnikov, Dmitri Maslov, and Michele Mosca. “Fast and efficient exact synthesis of single-qubit unitaries generated by Clifford and T gates”. Quantum Information & Computation 13, 607–630 (2013).

[11] Neil J Ross and Peter Selinger. “Optimal ancilla-free Clifford+T approximation of Z-rotations”. Quantum Information & Computation 16, 901–953 (2016).

[12] V. V. Shende, S. S. Bullock, and I. L. Markov. “Synthesis of quantum-logic circuits”. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 25, 1000–1010 (2006).

[13] Aram W. Harrow, Benjamin Recht, and Isaac L. Chuang. “Efficient discrete approximations of quantum gates”. Journal of Mathematical Physics 43, 4445–4451 (2002).

[14] Daniel Litinski. “A game of surface codes: large-scale quantum computing with lattice surgery”. Quantum 3, 128 (2019).

[15] Xiaoming Sun, Guojing Tian, Shuai Yang, Pei Yuan, and Shengyu Zhang. “Asymptotically optimal circuit depth for quantum state preparation and general unitary synthesis”. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 42, 3301–3314 (2023).

[16] Guang Hao Low and Isaac L. Chuang. “Hamiltonian simulation by qubitization”. Quantum 3, 163 (2019).

[17] Guang Hao Low and Isaac L. Chuang. “Hamiltonian simulation by uniform spectral amplification” (2017). arXiv:1707.05391.

[18] Scott Aaronson. “The complexity of quantum states and transformations: From quantum money to black holes” (2016). arXiv:1607.05256.

[19] 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”. Physical Review A 52, 3457–3467 (1995).

[20] Andrew M. Childs, Dmitri Maslov, Yunseong Nam, Neil J. Ross, and Yuan Su. “Toward the first quantum simulation with quantum speedup”. Proceedings of the National Academy of Sciences 115, 9456–9461 (2018).

[21] Lov Grover and Terry Rudolph. “Creating superpositions that correspond to efficiently integrable probability distributions” (2002). arXiv:quant-ph/0208112.

[22] Craig Gidney. “Halving the cost of quantum addition”. Quantum 2, 74 (2018).

[23] Guang Hao Low. “Halving the cost of quantum multiplexed rotations” (2021). arXiv:2110.13439.

[24] Alston S. Householder. “Unitary triangularization of a nonsymmetric matrix”. Journal of the ACM 5, 339–342 (1958).

[25] Vadym Klüchnikov. “Synthesis of unitaries with Clifford+T circuits” (2013). arXiv:1306.3200.

[26] David Gosset, Vadym Klüchnikov, Michele Mosca, and Vincent Russo. “An algorithm for the T-count”. Quantum Information & Computation 14, 1261–1276 (2014).

[27] Caterina E. Mora and Hans J. Briegel. “Algorithmic complexity and entanglement of quantum states”. Physical Review Letters 95, 200503 (2005).

[28] E. Knill. “Approximation by quantum circuits” (1995). arXiv:quant-ph/9508006.

[29] Michael Beverland, Earl Campbell, Mark Howard, and Vadym Klüchnikov. “Lower bounds on the non-clifford resources for quantum computations”. Quantum Science and Technology 5, 035009 (2020).

[30] Olivia Di Matteo, Vlad Gheorghiu, and Michele Mosca. “Fault-tolerant resource estimation of quantum random-access memories”. IEEE Transactions on Quantum Engineering 1, 1–13 (2020).
A Purified density matrix preparation

In some applications, particular in quantum simulation based on a linear combination of unitaries or qubitization \[2, 16\], it suffices to prepare the density matrix \( \rho = \sum_{x=0}^{N-1} \frac{a_x}{\| \vec{a} \|_1} |x\rangle\langle x| \) through a quantum state \( |\psi\rangle = \sum_{x=0}^{N-1} \sqrt{\frac{a_x}{\| \vec{a} \|_1}} |x\rangle |\text{garbage}_x\rangle \) of Eq. (4) where the number state \( |x\rangle \) is entangled with some garbage that depends only on \( x \). By allowing garbage, it was shown by \[8\] that strictly linear T gate complexity in \( N \) is achievable, using a SELECT data-lookup oracle corresponding to the \( \lambda = 1 \) case of Table 2. We outline the original idea, then generalize the procedure using the SELECT\textsc{Swap} network, which enables sublinear T gate complexity and better error scaling than the garbage-free approach. As density matrices have positive diagonals, we only consider the case of positive \( a_x \geq 0 \).

The original approach is based on a simple observation. By comparing a \( b \)-bit number state \( |a\rangle \) together with a uniform superposition state \( |u_{2^b}\rangle = \frac{1}{\sqrt{2^b}} \sum_{j=0}^{2^b-1} |j\rangle \) over \( 2^b \) elements, \( |a\rangle \) may be mapped to

\[
|a\rangle \rightarrow |a\rangle \left( \sqrt{\frac{a}{2^b}} |0\rangle_{|a\rangle} + \sqrt{\frac{2^b-a}{2^b}} |1\rangle_{|a\rangle} \right),
\]

(16)
where we denote a uniform superposition after the first $a$ elements by $|a_{\geq a}\rangle = \sum_{j=0}^{a-1} |j\rangle$. This may be implemented using quantum addition \cite{41}, which costs $O(b)$ Clifford+T gates with depth $O(b)$.

This observation is converted to state-preparation in four steps. First, the normalized coefficients $a_{j}N_{2^{b}} \approx a_{j}$ are rounded to nearest integer values such that $\|a'\|_1 = N2^{b}$. Second, the data-lookup oracle that writes two numbers $a''_{x} \in \{2^{b}\}$ and $f(x) \in \{N\}$ such that $a'_{x} = a''_{x} + \sum_{y \in \{f^{-1}(x)\}} (2^{b} - a''_{y})$. Thus

$$O|x\rangle|0\rangle|0\rangle = |x\rangle|a''_{x}\rangle|f(x)\rangle,$$

where we have omitted the irrelevant garbage state. Third, the oracle $O$ is applied to a uniform superposition over $|x\rangle$, and the comparator trick of Eq. (16) is applied. This produces the state

$$\sum_{x=0}^{N-1} \frac{|x\rangle}{\sqrt{N}}|a''_{x}\rangle|f(x)\rangle \left( \sqrt{\frac{a''_{x}}{2^{b}}}|0\rangle |u_{a''_{x}}\rangle + \cdots |1\rangle |u_{a''_{N}}\rangle \right).$$

Finally, $|f(x)\rangle$ is swapped with $|x\rangle$, controlled on the $|1\rangle$ state. This leads to a state $|\psi\rangle = \sum_{x=0}^{N-1} \sqrt{\frac{a''_{x}}{2^{b}}} |x\rangle |x\rangle |\text{garbage}_{x}\rangle$. After tracing out the garbage register, the resulting density matrix $\rho'$ approximates the desired state $\rho$ with trace distance

$$\|\rho' - \rho\|_1 = O(2^{-b}) \leq \epsilon.$$  

The T gate complexity is then the cost of the data-lookup oracle of Eq. (17) plus $O(b)$ for the comparator of Eq. (16), plus $O(\log N)$ for the controlled swap with $|f(x)\rangle$. By implementing this data-lookup oracle with the SELECTS\text{W}AP network, one immediately obtains the stated T gate complexity of $O(\lambda (b + \log N) + \frac{\lambda}{2}) = O(\lambda \log (N/\epsilon) + \frac{N}{\epsilon})$, where we choose $b = O(\log (1/\epsilon))$.

### B Data-lookup oracle implementation details

In this section, we present additional details on the implementation of the data-lookup oracle. In particular, we discuss a multi-target CNOT implementation in logarithmic depth without ancillary qubits in Appendix B.1, and a swap network SWAP with similar properties in Appendix B.2. Also evaluated is the T count and Clifford depth of these implementations up to constant factors. We define the Clifford depth, to be the number of layers of two-qubit Clifford gates that cannot be executed in parallel, assuming all-to-all qubit connectivity. We also assume that each T magic-state injection circuit has a Clifford depth of 1.

#### B.1 Quantum fanout in logarithmic depth without ancillary qubits

In this section, we construct a controlled-\text{NOT} gate that targets $n$ qubits, that is,

$$\text{CNOT}_{n} = |0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes X^{\otimes n}.$$  

The most straightforward approach applies $n$ \text{NOT} gates in sequence, each controlled by the same qubit. A slight modification results in logarithmic depth as shown in Table 3.

Given any number of $n$ qubits in states $|x_{j}\rangle$ for $j = 0, 1, \ldots, n - 1$, one may use a ladder of $n - 1$ controlled-\text{NOT} gates to realize the transformation

\[
\begin{align*}
|x_{0}\rangle|x_{1}\rangle|x_{2}\rangle\cdots|x_{n-1}\rangle \\
\rightarrow|x_{0}\rangle|x_{0} \oplus x_{1}\rangle|x_{2}\rangle\cdots|x_{n-1}\rangle \\
\rightarrow|x_{0}\rangle|x_{0} \oplus x_{1}\rangle|x_{0} \oplus x_{1} \oplus x_{2}\rangle\cdots|x_{n-1}\rangle \\
\rightarrow|x_{0}\rangle|x_{0} \oplus x_{1}\rangle|x_{0} \oplus x_{1} \oplus x_{2}\rangle\cdots\oplus_{j=0}^{n-1} x_{j}\rangle.
\end{align*}
\]

Let us call this unitary operation LADDER$_{n}$.

We now introduce a control qubit $|z\rangle$. One implementation of CNOT$_{n}$ is then obtained by applying LADDER$_{n}$$^\dagger$, followed by a \text{NOT} on $|x_{0}\rangle$ controlled by $|z\rangle$, and finally followed by LADDER$_{n}$. This has a Clifford depth of $2n - 1$ as depicted below for the example $n = 4$. 

---

Accepted in Quantum 2024-05-22, click title to verify. Published under CC-BY 4.0.
Table 3: Different implementations of a controlled-\textsc{not} gate Eq. (20) that targets $n$ qubits.

| Approach       | Clifford Depth ($d$) | Clifford Count | Volume       |
|----------------|----------------------|----------------|--------------|
| Linear         | $n$                  | $n$            | $(n+1)d$     |
| Logarithmic Eq. (22) | $2[\log_2(n)] + 1$  | $2n - 1$       | $(n+1)d$     |

\begin{equation}
2[\log_2(n/2 + 1)] \leq 2n - 1 = (n+1)d
\end{equation}

By distributing the controls and targets above in a tree-structure as depicted below for the example $n = 4$, the Clifford depth of $\text{Cnot}_n$ may be reduced to $2[\log_2(n)] + 1$.

\begin{equation}
\text{Cnot}_4\ \text{Ladder}^\dagger\ \text{Ladder}_4
\end{equation}

As the control qubit $|z\rangle$ is only used once in the above circuit, a further reduction in depth is possible by repeatedly using it to apply additional multi-target $\text{Cnot}$ gates in each time-slice. Let us denote by $g(d) = 2^{(d-1)/2}$ the maximum number of qubits targeted by the above circuit in a depth of $d$. Then the total number of qubits $n(d)$ targeted with this additional reduction satisfies the recurrence

\begin{equation}
n(1) = g(1),
\end{equation}

\begin{equation}
n(2) = n(1) + g(1),
\end{equation}

\begin{equation}
n(d) = n(d-1) + g(2[d/2] - 1) = \begin{cases} \frac{3 \cdot 2^{(d-1)/2} - 2}{2^{d/2 - 1}}, & d \text{ odd}, \\ 2^{d/2 - 1}, & d \text{ even}. \end{cases}
\end{equation}

Let us denote by $D[\text{Cnot}_n]$ the depth of this implementation of $\text{Cnot}_n$, which satisfies

\begin{equation}
D(\text{Cnot}_n) \leq \left\lceil 2\log_2 \left( \frac{n + 2}{2} \right) \right\rceil.
\end{equation}

B.2 Implementations of a \textsc{swap} network

In this section, we detail various implementations of the unitary swap network $\text{Swap}$ that moves an $b$-qubit quantum register indexed by $x \in \{0, \ldots, N-1\}$ to the position of the $x = 0$ register, controlled by an index state $|x\rangle$. More precisely, for any set of quantum states $\bigotimes_{x=0}^{N-1} |\phi_x\rangle_x$ in the $b$-qubit register indexed by $x$,

\begin{equation}
\text{Swap} \left[ |x\rangle \bigotimes_{x=0}^{N-1} |\phi_x\rangle_x \right] = |x\rangle |\phi_0\rangle_0 \cdots .
\end{equation}
| Approach          | Clifford Depth ($d_c$) | T count $T$ | T depth ($d_t$) | Volume = $(2n+1)(d_c + d_t)$ |
|-------------------|------------------------|------------|----------------|-----------------------------|
| Linear            | $4n + 4$               | $7n$       | $4$            | $8n^2 + \mathcal{O}(n)$     |
| Logarithmic       | $4D[CNOT]_{[n/2]} + 2$ | $14n$      | $8$            | $\leq 8n\left(\left\lceil \log_2 \frac{n+2}{2} \right\rceil + 2.5\right) + \mathcal{O}(\log n)$ |
| Phase incorrect   | $D[CNOT_n] + 4$        | $4n$       | $4$            | $\leq 2n\left(\left\lceil \log_2 \frac{n+2}{2} \right\rceil + 8\right) + \mathcal{O}(\log n)$ |

Table 4: Different implementations of a controlled-swap between two $n$-qubit registers. The depth $D[CNOT_n] \leq \left\lceil \log_2 \frac{N}{2} \right\rceil$ is from Eq. (25).

where the final quantum states in registers indexed by $x > 0$ are unimportant. Let us express the index $x \equiv x_0x_1 \cdots$ in binary, where $x_0$ is the smallest bit. Then it suffices to perform swaps between all pairs of registers indexed by $\{(i, i+2^j) | i \in 2^{j+1}N_0\}$, controlled by the $j$th qubit $|x_j\rangle$ of the index state $|x\rangle$, in the order of $j = 0, 1, \cdots, \left\lceil \log_2 N \right\rceil - 1$.

Each controlled pair-wise swap may be understood as a circuit $CSwap_n$ that swaps two $n$-qubit quantum registers in any state $|\psi\rangle, |\phi\rangle \in \mathbb{C}^{2^n}$, controlled by a single qubit $|z\rangle \in \mathbb{C}^2$. That is,

$$CSwap_n|z\rangle|\psi\rangle|\phi\rangle = \begin{cases} |z\rangle|\psi\rangle|\phi\rangle, & z = 0, \\ |z\rangle|\phi\rangle|\psi\rangle, & z = 1. \end{cases}$$

(27)

The overall cost of $Swap$ is then the sum of costs of $CSwap_n$, over all $j$, where $n = \left\lceil \frac{N}{2^{j+1}} + \frac{1}{2} \right\rceil$.

We now consider different implementations considered realize trade-offs between Toffoli-gate count, circuit depth, and ancillary qubit usage, as summarized in Table 4.

B.2.1 $CSwap_n$ in linear depth without ancillary qubits

It is simple to construct $CSwap_n$ with depth $\mathcal{O}(n)$ without any ancillary qubits. As the circuit that swaps two qubits is constructed from three CNOT gates as follows,

$$
\begin{array}{c}
|\psi\rangle \\
|\phi\rangle
\end{array} =
\begin{array}{c}
|\phi\rangle \\
|\psi\rangle
\end{array}
$$

a controlled-swap below is obtained by replacing the middle CNOT with a Toffoli gate.

$$
\begin{array}{c}
|z\rangle \\
\vdots
\vdots
\vdots
\end{array} =
\begin{array}{c}
|z\rangle \\
\vdots
\vdots
\vdots
\end{array}
$$

A circuit that implements Eq. (27), a swap between $n$ pairs of qubits, is then the above repeated $n$ times in sequence, each controlled by the same qubit as follows.

$$
\begin{array}{c}
|z\rangle \\
|\psi\rangle_0 \\
|\phi\rangle_0 \\
\vdots \\
|\psi\rangle_{n-1} \\
|\phi\rangle_{n-1}
\end{array}
$$

B.2.2 $CSwap_n$ in logarithmic depth without ancillary qubits

Constructing $CSwap_n$ with depth $\mathcal{O}(\log n)$ without any ancillary qubits requires a little more thought. Let us consider a more general problem. Suppose we have an arbitrary unitary operator $V$ that is self-inverse, meaning $V^2 = I$ – one may verify that the two-qubit swap satisfies this property. Our goal is to implement a multi-target controlled-$V$ gate on $n$ registers

$$
|0\rangle|0\rangle \otimes I + |1\rangle|1\rangle \otimes V^{\otimes n}.
$$

(28)

To begin, consider the following circuit identity, which is motivated by the ‘toggling’ trick in [42].
Observe the the bottom qubit may be dirty – its state does not affect the computation, and remains unchanged at the end of it.

Thus a multi-target controlled-V on \( n \) registers may be constructed by applying \( n \) singly-controlled V gates in parallel before and after a single multiply-controlled \( \text{NOT} \) gate, using a total of \( n \) extra dirty qubits as follows.

As the additional qubits may be dirty, this is easily modified to use no ancillary qubits at all. Let us apply the multi-target V on \( \lceil n/2 \rceil \) registers, using any \( \lfloor n/2 \rfloor \) qubits from the other registers as dirty qubits. When \( n \) is odd, the topmost V may be controlled directly by the \(|z\rangle\) qubit. We then apply the same circuit on the remaining registers by using qubits in the initial targets as control qubits. In total, this uses at most \( 2n \) controlled-V gates and two multiply-controlled \( \text{NOT} \) gates on \( \lceil n/2 \rceil \) qubits, each with cost given by Table 3.

### B.2.3 T-gate decomposition

Each controlled-swap may be decomposed into Clifford+T gates using standard techniques. For instance, the standard synthesis of each Toffoli uses 7 T-gates [12], as seen below.

Thus one might expect that \( \text{CSwap}_n \) in logarithmic depth requires \( 14N \) T-gates. However, simple cancellations using the above decomposition reduces this to \( 10N \) T-gates.

A further reduction to just \( 4N \) T-gates is possible if we allow the output state to be correct up to a phase factor. The decomposition by [19] below, using the gate \( G = S^\dagger \cdot H \cdot T \cdot H \cdot S \), approximates the Toffoli gate up to a minus sign on one of the matrix elements.

Thus a controlled-swap is obtained by a simple modification as follows.
Using this approximate controlled-swap, a version of \text{CSwap}_n that is correct up to a \pm 1 phase may be obtained by replacing the middle \text{CNOT} with \text{CNOT}^\dagger, which may be implemented with logarithmic depth. As this incorrect phase may be absorbed into the garbage state of the data-lookup oracle, we may apply this phase-incorrect swap operation without loss of generality.

\section{Indicator Function Construction}

There is an alternative construction based on implementing an \textit{indicator function}. That is, the function

\[ e^{(n)} : \{0, 1\}^n \to \{0, 1\}^N, \tag{29} \]

where \( N = 2^n \), which maps a string \( x \in \{0, 1\}^n \) to an exponentially long encoding which is 1 at position \( x \) and zeros everywhere else.

We observe that \( e^{(n)} \) can be implemented with the following parameters.

\textbf{Theorem 1.} For \( n \geq 1 \), let \( U_n \) be the unitary for the \( n \)th indicator function, \( e^{(n)} \). That is, the circuit mapping

\[ |x\rangle|y\rangle \mapsto |x\rangle e^{(n)}(x) \oplus y \] \quad \text{(30)}

for all \( x \in \{0, 1\}^n \) and \( y \in \{0, 1\}^N \). There is a circuit computing \( U_n \) in depth \( O(\log^2 N) \) with \( O(N) \) \( T \) gates without any ancillary qubits.

\textbf{Proof.} When \( n = 1, 2 \) there are trivial circuits without ancillary qubits. For \( n > 2 \), let us suppose we have arbitrarily many clean ancillary qubits. We divide the input bits \( x \) into two halves, \( x_{hi} \) and \( x_{lo} \), then recursively compute \( e^{(n/2)}(x_{hi}) \) and \( e^{(n/2)}(x_{lo}) \) in \( O(2^{n/2}) = O(\sqrt{N}) \) clean ancillary qubits each.

Each output of \( e^{(n)}(x) \) is the AND of a bit in \( e^{(n/2)}(x_{lo}) \) and a bit in \( e^{(n/2)}(x_{hi}) \). Each bit is used \( O(\sqrt{N}) \) times, so we can finish XORing \( e^{(n)}(x) \) into \( y \) with an array of Toffoli gates of \( O(\sqrt{N}) \). Alternatively, we can make \( O(\sqrt{N}) \) copies of each vector and do all the Toffoli gates simultaneously.

Now suppose the ancillary qubits are dirty. Dirty qubits cannot store a qubit in the same way as clean qubits; since the initial state of the qubit is arbitrary, the information is encoded in the \textit{change} in state rather than the actual state. To implement a \text{CNOT} controlled on a dirty qubit, for instance, we apply the \text{CNOT} first, flip or not flip the qubit, then apply another \text{CNOT}. If there was no change, the two gates cancel, otherwise exactly one fires.

Recall that the Toffoli gate has an implementation up to a faulty sign (which we can tolerate) as three \text{CNOT} interleaved with \text{G} or \text{G}†. As described above, we can implement each \text{CNOT} with a pair of \text{CNOT} and, more importantly, a recursive call to the subroutine which populates that qubit, i.e., an indicator function. Actually, we use the previously described circuit for \text{CNOT} instead of many individual \text{CNOT} gates for each Toffoli involving that bit.

The depth cost for \( \text{CNOT}_O(\sqrt{N}) \) is \( O(\log(\sqrt{N})) = O(n) \). Additionally, to compute \( U_n \) we need 4 recursive calls to \( U_{n/2} \). The depth satisfies the recurrence \( D(n) = 4D(n/2) + O(n) \), so \( D(n) = O(n^2) = O(\log^2 N) \).

Finally, suppose we apply the array of Toffoli gates in two layers instead of just one simultaneous layer. Since we are only applying Toffolis to half of the \(|y\rangle\) register at any time, the other half can be used as dirty ancillary qubits. Asymptotically, this is much more than the \( O(\sqrt{N}) \) dirty qubits we need to compute \( U_{n/2} \). Hence, we can implement \( U_n \) without any additional qubits. \qed

\begin{itemize}
  \item It is wasteful to compute all of \( e^{(n)}(x) \) in ancillary qubits, but it can be used to compute an arbitrary function. Better is to divide the input into two pieces, similar to how we compute the indicator function.
  \item \textbf{Theorem 2.} Suppose \( f : \{0, 1\}^n \to \{0, 1\}^b \) is an arbitrary function, and let \( U \) be the unitary mapping \[ |x\rangle|y\rangle \mapsto |x\rangle y \oplus f(x), \quad \text{(31)} \]
\end{itemize}
for all $x \in \{0,1\}^n$ and $y \in \{0,1\}^b$.

There is a circuit for $U$ of depth $O(\log^2 N + \log b)$, using $O(bN)$ dirty ancillary qubits and $O(\sqrt{N})$ T gates. Alternatively, there is a circuit for $U$ using $O(\sqrt{bN})$ T gates and only $O(\lambda \sqrt{bN})$ dirty ancillary qubits, but with depth $O(\sqrt{bN}/\lambda + \log^2 N)$, for any $1 \leq \lambda \leq \sqrt{bN}$.

Proof. Divide the input into $n - k$ and $k$ bit pieces, where $k$ is to be determined later. Let $\hat{f}: \{0,1\}^{n-k} \to \{0,1\}^{k \times 2^k}$ be

$$
\hat{f}(x_{hi}) = \left( f(x_{hi},x_{lo}) \right)_{x_{lo} \in \{0,1\}^k},
$$

the function which outputs the function for all possible values of the low order bits, given the high order bits.

Suppose, for the moment, that we have as many clean ancillary qubits as we want. We can naively compute $\hat{f}(x_{hi})$ by constructing $e^{(n-k)}(x_{hi})$, making $b \cdot 2^k$ copies, and computing the parity of some subset of bits of $e^{(n-k)}(x_{hi})$ for each output bit of $\hat{f}(x_{hi})$. Constructing $e^{(n-k)}(x_{hi})$ requires $O(2^{n-k})$ ancillary qubits, $O(2^{n-k})$ T gates and $O((n-k)^2)$ depth. The rest is done with CNOT$_{b \cdot 2^k}$ gates to make copies and CNOT$_{O(2^{n-k})}$ gates conjugated by Hadamards to compute parities. This uses many ancillary qubits--$O(2^{n-k})$ for the original vector times $O(b2^k)$ copies is $O(bN)$ qubits.

Since the layers of CNOT gates compute (in the output bits) a linear function of the inputs, it is not difficult to adapt for dirty ancillary qubits. Just apply the linear function, flip the dirty ancillary qubits that are set to 1, then apply the linear function again, appealing to the identity $T(x \oplus y) \oplus T(y) = T(x)$ for a linear function. Thus, whatever state $|y\rangle$ was in the dirty qubits, we still manage to compute $T(x)$.

We have shown how to compute $\hat{f}(x_{hi})$ and XOR it into dirty ancillary qubits. We also know how to compute $e^k(x_{lo})$ and XOR $b$ copies of it into dirty ancillary qubits. Think of $\hat{f}(x_{hi})$ as a $b \times 2^k$ matrix, then all that remains is to return the correct column of $\hat{f}(x_{hi})$. If we also think of $e^k(x_{lo})$ as a length $2^k$ column vector, then we are computing a matrix/vector product. The simplest way to do this is to make $b$ copies of $e^k(x_{lo})$ and execute $b$ vector/vector products in parallel, at a cost of $O(2^k)$ Toffoli gates for each one.

To compute the Toffoli gates on dirty ancillary qubits, we decompose them into CNOT gates and single qubit gates. The layers of CNOT gates are linear, so it is possible to compute each such layer with dirty ancillary qubits.

Computing $\hat{f}(x_{hi})$ uses $O(bN)$ ancillary qubits, $O(2^{n-k})$ T gates, and has depth $O((n-k)^2 + k + \log b)$. The rest of the circuit has $O(b2^k)$ ancillas to store $\hat{f}(x_{hi})$ and/or copies of $e^k(x_{lo})$, and uses $O(b2^k)$ T gates in depth $O(k^2 + \log b)$. Since the T gate count is $O(2^{n-k} + b2^k)$, we set $k$ such that $2^k \approx \sqrt{N/b}$, and it becomes $O(\sqrt{bN})$.

It is possible to trade off depth and number of ancillary qubits. We only need $O(2^{n-k})$ qubits to store $e^{(n-k)}(x_{hi})$ in the computation of $\hat{f}(x_{hi})$, if we are willing to compute parities for each of the $b2^k$ output bits of $\hat{f}(x_{hi})$ one at a time, in depth $O(b2^k)$. More generally, we can use $O(\lambda 2^{n-k})$ ancillary qubits for any integer $\lambda \in [1,b2^k]$ and use depth $O(b2^k/\lambda)$. For the optimal T count we use the same setting of $k$, giving $O(\sqrt{bN}/\lambda + \log^2 N)$ depth with $O(\lambda \sqrt{bN})$ ancillary qubits and $O(\sqrt{bN})$ T gates.

\end{proof}

D Pure-state preparation implementation details

The approach by Shende, Bullock, and Markov [12] synthesizes a unitary $A$ that prepares a pure state $A |0\rangle = \sum_{x=0}^{2^N} \frac{a_x}{\sqrt{2^N}} |x\rangle = |\psi\rangle$ with arbitrary coefficients in $N = 2^n$ dimensions. The underlying circuit, illustrated below for the example of $N = 8$ for positive coefficients,
is built from $j \in [n]$ multiply-controlled arbitrary single qubit rotations, $U_j$ where

$$U_j = \sum_{x=0}^{2^n-1} |x\rangle\langle x| \otimes e^{i2\pi\theta_{j,x}Z},$$  \hspace{1cm} (33)$$

for some set of rotation angles $\theta_{j,x}$. Note that it suffices to consider $Z$-phase rotations as rotations about the $X,Y$ Pauli operators are equivalent up to a single-qubit Clifford similarity transformation. Each multiplexor is applied twice — once to create a pure state with the right probabilities $|a_x\rangle^2$, and once to apply the correct phase $e^{i\arg[a_x]}$. Below we describe how $U_j$ may be implemented using the data-lookup oracle of Eq. (5)

$$O|x\rangle|0\rangle = |x\rangle|a_x\rangle|\text{garbage}_x\rangle,$$  \hspace{1cm} (34)$$

and evaluate the overall error and cost of state preparation.

D.1 Multiply-controlled phase gate from data lookup oracles

Consider a multiply-$n$-controlled arbitrary single qubit rotation

$$U = \sum_{x=0}^{2^n-1} |x\rangle\langle x| \otimes e^{i2\pi\theta_xZ},$$  \hspace{1cm} (35)$$

where each rotation angle $\theta_x \in [0,1)$. Given a number state $|x\rangle$ and an arbitrary single-qubit state $|z\rangle$, this unitary performs a controlled-rotation

$$U|x\rangle|z\rangle = |x\rangle e^{i2\pi\theta_xZ}|z\rangle.$$  \hspace{1cm} (36)$$

Each rotation angle has a binary expansion

$$\theta_x = \sum_{k=0}^{\infty} a_{x,k} 2^{-k}.$$  \hspace{1cm} (37)$$

By truncating to the above to $b$-bits of precision, we obtain an integer approximation $a_x$ of $\theta_x$ where

$$a_x = 2^b \sum_{k=0}^{b-1} a_{x,k} 2^{-k}, \quad \left| \frac{a_x}{2^b} - \theta_x \right| < \frac{1}{2^b}. \hspace{1cm} (38)$$$$

Let us encode these values of $a_x$ into the data-lookup oracle, and express its T cost as the function $f(n, b)$. Its output is then

$$O|x\rangle|0\rangle \otimes |0\rangle = |x\rangle \left( \bigotimes_{k=0}^{b-1} |a_{x,k}\rangle \right) |\text{garbage}_x\rangle,$$  \hspace{1cm} (39)$$

where we explicitly represent the number state $|a_x\rangle$ in terms of its component qubits.

D.1.1 Approach using arbitrary single-qubit synthesis

One possible approximation, call it $U'$, of $U$ in Eq. (35) applies the single-qubit rotation $e^{i\pi Z/2^k}$ to the target state $|z\rangle$, controlled by the state $|a_{x,k}\rangle$. The garbage register is then uncomputed by
running $O$ in reverse. Explicitly, this circuit realizes the transformation

$$
|x\rangle\langle 0|\otimes |0\rangle|z\rangle \\
\rightarrow |x\rangle \left( \bigotimes_{k=0}^{b-1} |a_{x,k}\rangle \right) |\text{garbage}_x\rangle|z\rangle \\
\rightarrow |x\rangle \left( \bigotimes_{k=0}^{b-1} |a_{x,k}\rangle \right) e^{i\pi \sum_{k=0}^{b-1} a_{x,k} Z/2^k} |z\rangle \\
= |a_x\rangle|\text{garbage}_x\rangle e^{i2\pi a_x/2^b} |z\rangle \\
\rightarrow |x\rangle\langle 0|\otimes |0\rangle e^{i2\pi a_x/2^b} |z\rangle.
$$

(40)

Now, each controlled-arbitrary phase rotation decomposes into 2 arbitrary single-qubit rotations, and CNOT gates – Note that a decomposition into 1 arbitrary single-qubit rotation is possible if we modify the above for the range $\theta_x \in (-1/2, 1/2)$, but the explanation is slightly more complicated. As any arbitrary single-qubit rotation is approximated to error $\epsilon$ using $O(\log (1/\epsilon))$ T gates [43], we may use a triangle inequality to bound the error to

$$
\|U - U'\| \leq \left\| \sum_{x=0}^{2^n-1} |x\rangle\langle x| \otimes (e^{i2\pi \theta_x Z} - e^{i2\pi a_x Z/2^b}) \right\| + 2b\epsilon \\
\leq \max_{|y|<1/2^b} \left| (e^{i2\pi y} - 1) \right| + 2b\epsilon \\
< \frac{2\pi}{2^b} + 2b\epsilon.
$$

(41)

Thus for any target error $\|U - U'\| = \delta$, we may solve for the $b, \epsilon$ parameters and bound the number of T gates required to

$$
f(n, b) + O(\log^2 (1/\delta)), \quad b = O(\log (1/\delta)).
$$

(42)

D.1.2 Approach using phase gradients

A more efficient approach [22] uses a Fourier state as a resource

$$
\mathcal{F} = \frac{1}{\sqrt{2^b}} \sum_{k=0}^{2^b-1} e^{-2\pi ik/2^b} |k\rangle.
$$

(43)

combined with a reversible adder

$$
\text{ADD} |x\rangle|y\rangle = |x\rangle|y+x \mod 2^b\rangle.
$$

(44)

Observe that

$$
\text{ADD} |x\rangle|\mathcal{F}\rangle = e^{2\pi i x/2^b} |x\rangle|\mathcal{F}\rangle.
$$

(45)

Thus the controlled adder, known to cost $O(b)$ T gates,

$$
\text{CADD} = \text{ADD} \otimes |0\rangle\langle 0| + \text{ADD}^\dagger \otimes |1\rangle\langle 1|
$$

(46)

realizes the controlled-phase rotation

$$
\text{CADD} |x\rangle|\mathcal{F}\rangle|z\rangle = |x\rangle|\mathcal{F}\rangle e^{2\pi i x/2^b} |z\rangle.
$$

(47)

Thus another possible approximation, call it $U''$, uses this adder, controlled by the target state $|z\rangle$, to the registers containing the desired phase rotation $|a_x\rangle$, and the Fourier state. This realizes the same transformation in Eq. (40), using the circuit depicted below.
Assuming that the Fourier state is prepared perfectly, this approximates $U$ with error
\[
\|U - U''\| \leq \left\| \sum_{x=0}^{2^n-1} |x\rangle\langle x| \otimes (e^{i2\pi x z Z} - e^{i2\pi a_x Z/2^b Z}) \right\| \leq \max_{|y| < 1/2^b} \left| (e^{i2\pi y} - 1) \right| < \frac{2\pi}{2^b}.
\] (48)

The state preparation unitary $A''$ applies $n$ such approximations $U''_j$ to the multiply-controlled rotations $U_j$, leading to an overall error bounded by
\[
\left\| \prod_{j=0}^{n-1} U_j - \prod_{j=0}^{n-1} U''_j \right\| \leq \sum_{j=0}^{n-1} \|U_j - U''_j\| < \frac{2n\pi}{2^b}.
\] (49)

Note that the cost of approximating the Fourier state to error $\epsilon$ has a $T$ cost of $O(b \log (1/\epsilon))$. This imperfect Fourier state $|F'\rangle$ contributes a one-time error of $\epsilon$, following the inequality
\[
\|W(|F'\rangle \otimes I) - (W|F\rangle \otimes I)\| \leq \| |F'\rangle - |F\rangle\| \leq \epsilon,
\] (50)
for any arbitrary unitary operator $W$. Thus $A''$ prepares the state $|\psi''\rangle = A''|0\rangle|F'\rangle$, which approximates $|\psi\rangle$ with error
\[
\delta = \| |\psi''\rangle - |\psi\rangle|F\rangle\| \leq \frac{2\pi n}{2^b} + \epsilon.
\] (51)

The total error $\delta$ may be controlled by choosing $b = \Theta(\log (\frac{N}{\delta}))$ and $\epsilon = \Theta(\delta)$.

The total $T$ cost of arbitrary state preparation is then the sum of costs of the data-lookup oracles $f(j, b) = O(\frac{2^n}{\lambda} + b\lambda)$, adders, and the Fourier state:
\[
T \text{ cost } = \left( \sum_{j=0}^{n-1} f(j, b) + O(b) \right) + O(b \log (1/\epsilon))
\]
\[
= O(2^n/\lambda + nb\lambda + nb + b \log (1/\epsilon))
\]
\[
= O \left( \frac{N}{\lambda} + b(\lambda n + \log (1/\delta)) \right)
\]
\[
= O \left( \frac{N}{\lambda} + \log \left( \frac{\log N}{\delta} \right) (\lambda \log N + \log (1/\delta)) \right)
\]
\[
= O \left( \frac{N}{\lambda} + \lambda \log \left( \frac{\log N}{\delta} \right) \log \left( \frac{N}{\delta} \right) \right)
\]
\[
= O \left( \frac{N}{\lambda} + \lambda \log^2 \left( \frac{N}{\delta} \right) \right).
\] (52)