

Repeat-Until-Success: Non-deterministic decomposition of single-qubit unitaries

Adam Paetznick

David R. Cheriton School of Computer Science and
Institute for Quantum Computing,
University of Waterloo

Krysta M. Svore

Quantum Architectures and Computation Group,
Microsoft Research

October 21, 2014

Abstract

We present a decomposition technique that uses non-deterministic circuits to approximate an arbitrary single-qubit unitary to within distance ϵ and requires significantly fewer non-Clifford gates than existing techniques. We develop “Repeat-Until-Success” (RUS) circuits and characterize unitaries that can be exactly represented as an RUS circuit. Our RUS circuits operate by conditioning on a given measurement outcome and using only a small number of non-Clifford gates and ancilla qubits. We construct an algorithm based on RUS circuits that approximates an arbitrary single-qubit Z -axis rotation to within distance ϵ , where the number of T gates scales as $1.26 \log_2(1/\epsilon) - 3.53$, an improvement of roughly three-fold over state-of-the-art techniques. We then extend our algorithm and show that a scaling of $2.4 \log_2(1/\epsilon) - 3.28$ can be achieved for arbitrary unitaries and a small range of ϵ , which is roughly twice as good as optimal deterministic decomposition methods.

1 Introduction

As quantum devices continue to mature, there is an emerging need for algorithms that can efficiently and accurately map a high-level quantum algorithm into a low-level fault-tolerant circuit representation. The mapping of a quantum algorithm into its equivalent fault-tolerant circuit representation requires first the choice of a universal basis or gate set, and second a decomposition algorithm that can translate a quantum circuit into a sequence of gates drawn from that basis. The choice of basis is predominantly dictated by the existence of resource-efficient, fault-tolerant quantum error correction protocols for each gate; a common set is CNOT plus the universal single-qubit basis $\{H, T\}$, where $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and $T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$. For many quantum error-correcting codes, a fault-tolerant H requires transversal application of the gate, and a fault-tolerant T requires magic state distillation. The cost of a $\{H, T\}$ circuit is defined to be the number of T gates, given that the resource cost of a fault-tolerant T gate is up to an order of magnitude larger than the resource cost of a fault-tolerant H gate [RHG07, FDJ13].

The decomposition algorithm should minimize the desired cost function, such as the T count of the ϵ -approximate gate sequence. The Solovay-Kitaev theorem [Kit97, KSV02], guarantees that a single-qubit unitary operation can be efficiently approximated to within error ϵ by a sequence of $O(\log^c(1/\epsilon))$ gates from a discrete universal basis, where $c = 1$ is the theoretical lower bound [Kni95]. Fowler gave an exponential-time algorithm that achieves the lower bound, resulting in an approximating sequence containing $2.95 \log_2(1/\epsilon) + 3.75 T$ gates, on average [Fow11]. However, the exponential time complexity limits the achievable accuracy. A database search algorithm based on canonical forms for $\{H, T\}$ circuits was given by Bocharov and Svore [BS12] that also achieves the lower bound and enables search to slightly better accuracy. Recently, efficient algorithms that achieve the lower bound have been developed. Kliuchnikov, Maslov and Mosca (KMM) developed an algorithm which yields $3.21 \log_2(1/\epsilon) - 6.93 T$ gates for the rotation $R_Z(1/10)$ [KMM12b]. Selinger’s algorithm ϵ -approximates a single-qubit Z -axis rotation, $R_Z(\theta) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix}$, using $4 \log_2(1/\epsilon) + 11 T$ gates in the worst case [Sel12]. Subsequent improvement by Ross and Selinger yields a scaling of $3 \log(1/\epsilon) + O(\log \log(1/\epsilon))$ in typical cases [RS14].

For a given single-qubit unitary U and error ϵ , the above algorithms output a fixed sequence of single-qubit gates from the set $\{H, T\}$, without the use of ancillary qubits or measurements. In this paper, we present a circuit framework and algorithm to minimize the T gates required to approximate a given single-qubit unitary. We show that by incorporating ancilla qubits and measurements, the expected number of T gates required to approximate a random Z -axis rotation can be significantly reduced to

$$\text{Exp}_Z[T] = 1.26 \log_2(1/\epsilon) - 3.53 \quad , \quad (1)$$

an improvement of roughly three-fold over [Sel12] and more than two-fold over [Fow11], [KMM12b] and [RS14]. For arbitrary single-qubit unitaries, our results indicate a significantly reduced T -count scaling of

$$\text{Exp}_U[T] = 2.4 \log_2(1/\epsilon) - 3.28 \quad , \quad (2)$$

roughly 50 percent better than using (1) for each Z rotation (three are required in general) and up to four-fold better than traditional ancilla-free decomposition.

Our circuits are distinct from those output by Fowler, KMM and Selinger in that they are *non-deterministic*. Each circuit, when conditioned on a particular measurement outcome, exactly implements a desired unitary, and otherwise implements a unitary that can be reversed at little or no cost; it can then be repeated until the desired unitary is obtained. We call our circuits “Repeat-Until-Success” (RUS) circuits. A significant advantage of RUS circuits is the extremely low resource cost, in non-Clifford gates and ancillary qubits.

Our paper is structured as follows. We begin in Section 2 by discussing existing single-qubit unitary decomposition techniques, and the presence of RUS circuits in previous work. We then characterize unitaries that can be exactly implemented ($\epsilon = 0$) as an RUS circuit in Section 3. In Section 4, we present an optimized direct search algorithm for synthesizing RUS circuits with extremely low T count and in Section 5, we construct a corresponding database of RUS circuits. Leveraging our database, we develop a decomposition algorithm to approximate a given unitary using compositions of RUS circuits in Section 6. We then present a variety of applications of RUS circuits, including a circuit for the V_3 gate that results in state-of-the-art single-qubit decomposition. Finally, we discuss future directions and open problems in Section 7.

2 Existing methods for single-qubit unitary decomposition

In addition to the techniques discussed above [DN05, Fow11, KMM12b, Sel12, RS14], a variety of other methods for single-qubit unitary decomposition have been developed. So-called “phase kickback” involves preparing a special ancilla state based on the quantum Fourier transform and then using phase estimation [KSV02]. Non-deterministic circuits called “programmable ancilla rotations” (PAR) use a cascading set of prepared ancilla states along with gate teleportation [JWM⁺12]. Similar use of non-deterministic circuits to produce a “ladder” of non-stabilizer states, and in turn to approximate an arbitrary single-qubit unitary, has also been proposed [DS12]. The number of T gates required for these ancilla-based methods is larger than for ancilla-free methods, but the total resources are comparable in some architectures [Jon13a]. For this reason, we compare our results to the Fowler, KMM, Selinger, and Ross-Selinger methods.

Non-deterministic circuits have also been proposed for decomposition into alternate gate sets. Bocharov, Gurevich and Svore (BGS) showed that arbitrary single-qubit unitaries can be approximated using the gate set $\{H, S = T^2, V_3\}$, where $V_3 = (I + 2iZ)/\sqrt{5}$, with a typical scaling of $3 \log_5(1/\epsilon)$ in the number of V_3 gates [BGS13]. They suggest a fault-tolerant implementation of the V_3 gate (see Fig. 1a) using an RUS circuit which requires eight T gates, four for each Toffoli (see [Jon13b]). Later, Jones improved this circuit, using only a single Toffoli gate [Jon13a]. Using our optimized direct search algorithm, we find an improved RUS circuit for V_3 that uses only four T gates, as shown in Fig. 1c, and is exact ($\epsilon = 0$). By contrast, an approximation to within $\epsilon = 10^{-6}$ using the KMM algorithm requires 67 T gates. Furthermore, when used to implement V_3 , our circuit results in $\{H, S, V_3\}$ -decomposition achieving substantially lower T count (on average) than $\{H, T\}$ -decomposition methods.

Repeat-until-success circuits have also been used by Wiebe and Kliuchnikov [WK13], who proposed a family of tree-like, hierarchical RUS circuits that yield T counts superior to Selinger and KMM for small-angle Z -axis rotations. In contrast, our results show that RUS circuits can be used for large- and small-angle Z -axis rotations, as well as rotations about an arbitrary axis. We also provide a general characterization of RUS circuits, and a general framework for their construction.

A summary of the T count costs of our method, labeled RUS, and the above algorithms is given in Tables 1 and 2 for non-axial and axial rotations, respectively.

RUS circuits have been considered in other contexts, as well. The term was first used by [LBK04] to describe the implementation of a CZ gate by repeated operations in linear optics. More recently, [SO13] adapted deterministic ancilla-driven methods [AOK⁺10, KOB⁺09] to allow for non-determinism. Our use of repetition is similar to [LBK04] and [SO13], but we generate a family of circuits each of which are intended for use in conjunction with a fault-tolerant gate set, rather than at the physical level.

3 Repeat-Until-Success circuits

To describe RUS circuits, we begin with an example. Consider the circuit shown in Fig. 1a, which performs the single-qubit unitary $V_3 = (I + 2iZ)/\sqrt{5}$. This circuit involves two measurements in the Pauli X -basis. If both measurement outcomes are zero, then the output is equivalent to $V_3 |\psi\rangle$. If any other outcome occurs, then the output is $I |\psi\rangle = |\psi\rangle$. Thus, the circuit may be repeated until obtaining the all zeros outcome, and the number of repetitions will vary according to a geometric probability distribution. (In this case the probability of getting both zeros is $5/8$.) Upon measuring

Method	Description	T count	Comments
Solovay-Kitaev [DN05]	Converging ϵ -net based on group commutators.	$O(\log^{3.97} 1/\epsilon)$	Computationally efficient, but sub-optimal T count.
Ladder states [DS12]	Hierarchical distillation based $ H\rangle$ states.	$O(\log^{1.75} 1/\epsilon)$	Some of the cost can be shifted “offline”.
Direct search [Fow11, BS12]	Optimized exponential-time search.	$2.95 \log_2(1/\epsilon) + 3.75$	Optimal ancilla-free T count.
BGS [BGS13]	Direct search decomposition with V_3 .	$T_V(3 \log_5 1/\epsilon)$	T_V is the T count for choice of fault-tolerant implementation of V_3 .
RUS (non-axial)	Database lookup.	$2.4 \log_2(1/\epsilon) - 3.28$	Limited approximation accuracy.

Table 1: Decomposition methods for arbitrary single-qubit unitaries using the gate set $\{H, S, T\}$.

Method	Description	T count	Comments
Phase kickback [KSV02]	Uses Fourier states and phase estimation.	$O(\log 1/\epsilon)$ (implementation dependent)	$O(\log 1/\epsilon)$ ancillas. Optimizations make it cost competitive with Selinger and KMM.
PAR [JWM ⁺ 12]	Cascading gate teleportation.	$O(\log 1/\epsilon)$	Constant depth (on average), higher T count than phase kickback.
Selinger [Sel12]	Round-off followed by exact decomposition.	$4 \log(1/\epsilon) + 11$	T count is optimal for worst-case rotations.
Ross-Selinger [RS14]	Round-off followed by exact decomposition.	$3 \log(1/\epsilon) + O(\log \log 1/\epsilon)$	T count is near-optimal for typical rotations.
KMM [KMM12b]	Round-off followed by exact decomposition.	$3.21 \log_2(1/\epsilon) - 6.93$	T count based on scaling for $R_Z(1/10)$.
Floating-point [WK13]	A family of tree-like RUS circuits	$1.14 \log_2(10^\gamma) + 8 \log_2(10^{-\gamma}/\epsilon)$	For small angle $\theta = a \times 10^{-\gamma}$, T count is roughly $1.14 \log_2(1/\theta)$.
RUS (axial)	Database lookup.	$1.26 \log_2(1/\epsilon) - 3.53$	Approximation to within $\epsilon = 10^{-6}$.

Table 2: Decomposition methods for Z -axis rotations using the gate set $\{H, S, T\}$. Approximation of an arbitrary single-qubit unitary is possible by using the relation $U = R_Z(\theta_1)HR_Z(\theta_2)HR_Z(\theta_3)$.

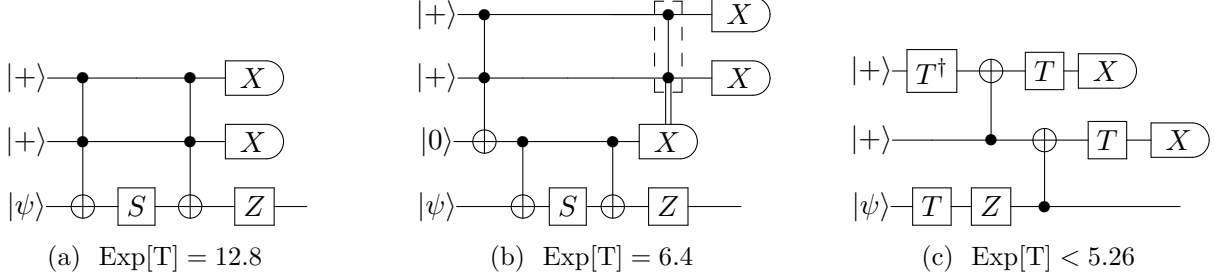


Figure 1: Repeat-Until-Success circuits for $V_3 = (I+2iZ)/\sqrt{5}$. Each of the circuits above implements V_3 conditioned on an X -basis measurement outcome of zero on each of the top two ancilla qubits. If any other measurement outcome occurs, then each circuit implements the identity. The probability of measuring 00 is $5/8$ for each circuit. Repeating the circuit until success yields an expectation value for the number of T gates, as indicated. (a) A slight modification of the circuit presented in [NC00] pp. 198. Each Toffoli gate can be implemented with four T gates (see [Jon13b]). (b) A circuit proposed by Jones that requires just a single Toffoli gate [Jon13a]. (c) An alternative circuit found by direct search. Measurement of the first qubit can be performed before interaction with the data qubit. Thus the top-left part of the circuit can be repeated until measuring zero. The probability of measuring zero on the first qubit is $3/4$. The probability of measuring zero on the second qubit, conditioned on zero outcome of the first qubit, is $5/6$. The T gate applied directly to $|\psi\rangle$ can be freely commuted through the CNOT. In the case that an even number of attempts are required, the T gates can be combined into the Clifford gate $T^2 = S$.

all zeros, the unitary V_3 is implemented *exactly*, even though the overall circuit is non-deterministic.

We define a Repeat-Until-Success (RUS) circuit over a gate set G to be of the following general structure:

1. Prepare m ancilla qubits in the state $|0^m\rangle$.
2. Given an input state $|\psi\rangle$ on n qubits, apply a unitary W to all of the $n + m$ qubits using gates from G .
3. Measure each ancilla qubit in the computational basis. The output is given by $\Phi_i |\psi\rangle$, where Φ_i is a quantum channel on n qubits that depends on the measurement outcome $i \in \{0, 1\}^m$.
4. If the measurement outcome indicates “failure”, apply a recovery operation and repeat.

The measurement outcomes are partitioned into two sets: “success” and “failure”. Success corresponds to some set of desired operations $\{\Phi_i : i \in \text{success}\}$; failure corresponds to some set of undesired operations $\{\Phi_i : i \in \text{failure}\}$. In the case of success, no further action is required. In the case of failure i , a recovery operation Φ_i^{-1} is applied, and the circuit is repeated. For practical purposes, the recovery operations should be implementable for relatively low cost compared to W .

We restrict to the case in which $|\psi\rangle$ is a single qubit and the $\{\Phi_i\}$ are unitary. We also limit to a single “success” output $U |\psi\rangle$, for some unitary U , though U may correspond to multiple measurement outcomes. The operation W is then given by a $2^{m+1} \times 2^{m+1}$ unitary matrix of the

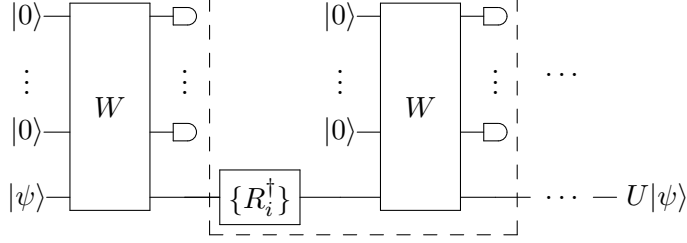


Figure 2: A Repeat-Until-Success circuit that implements the unitary U . Ancilla qubits are prepared in $|0\rangle$, then the unitary W is performed on both the ancillas and $|\psi\rangle$. Upon measuring the ancillas, a unitary operation is effected on $|\psi\rangle$ which is either U or one of $\{R_i\}$, depending on the measurement outcome. If the measurement outcome indicates R_i , then the recovery operation R_i^\dagger is performed, and the process can be repeated.

form

$$W = \frac{1}{\sqrt{\sum_i |\alpha_i|^2}} \begin{pmatrix} \alpha_0 U & \dots \\ \alpha_1 R_1 & \ddots \\ \vdots & \\ \alpha_l R_l & \end{pmatrix}, \quad (3)$$

where U, R_1, \dots, R_l are 2×2 unitary matrices, and $\alpha_0, \dots, \alpha_l \in \mathbb{C}$ are scalars. Since the ancillas are prepared in $|0^m\rangle$, only the first two columns of W are of consequence. Contents of the remaining columns are essentially unrestricted, except that W must be unitary. Each of the $l + 1 = 2^m$ measurement outcomes corresponds to application of a unitary from $U \cup \{R_i\}$ on the input qubit $|\psi\rangle$. Without loss of generality, we select the all zeros outcome to correspond with application of U , since outcomes can be freely permuted. The entire protocol is illustrated in Fig. 2.

To ensure compatibility with existing fault-tolerance schemes, we require that W can be synthesized using the gate set $G = \{\text{Clifford}, T\}$, where Clifford denotes the Clifford group generated by $\{H, S, \text{CNOT}\}$; note that our framework and algorithms can be extended to other gates sets with little difficulty. A unitary matrix is exactly implementable by $\{\text{Clifford}, T\}$ if and only if its entries are contained in the ring extension $\mathbb{Z}[i, \frac{1}{\sqrt{2}}]$ [GS12]. Thus, we require that $\alpha_0 U, \alpha_1 R_1, \dots, \alpha_l R_l \in \mathbb{Z}[i, \frac{1}{\sqrt{2}}]$.

Furthermore, the normalization $1/\sqrt{\sum_i |\alpha_i|^2}$ must also be in the ring. The unitarity condition on W then requires that

$$\sum_i |\alpha_i|^2 = 2^k \quad (4)$$

for some integer k .

If all of the recovery operations R_1, \dots, R_l are exactly implementable by $\{\text{Clifford}, T\}$, then we may assume that $\alpha_1, \dots, \alpha_l \in \mathbb{Z}[i, \frac{1}{\sqrt{2}}]$. If α_0 is an integer, then Lagrange's four-square theorem implies that (4) can be satisfied using at most $m = 2$ ancilla qubits.

3.1 Characterization

Consider a 2×2 unitary matrix U such that

$$U = \begin{pmatrix} u_{00} & u_{01} \\ u_{10} & u_{11} \end{pmatrix} = \frac{1}{\sqrt{2^k} \alpha} \begin{pmatrix} \beta_{00} & \beta_{01} \\ \beta_{10} & \beta_{11} \end{pmatrix}, \quad (5)$$

for $\alpha \in \mathbb{R}$, $\beta_{00}, \dots, \beta_{11} \in \mathbb{Z}[i, \sqrt{2}]$ and integer $k \geq 0$. We are concerned with exactly implementing U only up to a global unit phase $e^{i\phi}$ for some $\phi \in [0, 2\pi)$. Accordingly, we may assume without loss of generality that α is real and non-negative since for any $\beta \in \mathbb{C}$, $\frac{\beta\beta^*}{|\beta|} \geq 0$. The restriction to $\mathbb{Z}[i, \sqrt{2}]$ rather than $\mathbb{Z}[i, \frac{1}{\sqrt{2}}]$ is also without loss of generality, since k can be chosen to eliminate any denominators. Then choosing $\alpha_0 = \sqrt{2^k}\alpha$ we have

$$\alpha_0 = \sqrt{|\beta_{00}|^2 + |\beta_{10}|^2} = \sqrt{x + y\sqrt{2}} \ , \quad (6)$$

where $x = a_{00}^2 + c_{00}^2 + a_{10}^2 + c_{10}^2 + 2(b_{00}^2 + d_{00}^2 + b_{10}^2 + d_{10}^2)$, $y = a_{00}b_{00} + c_{00}d_{00} + a_{10}b_{10} + c_{10}d_{10}$ for integers $a_{00}, b_{00}, c_{00}, d_{00}, a_{10}, b_{10}, c_{10}, d_{10}$.

Any target unitary U must have this form due to (3). In other words, the *only* unitaries that can be obtained by $\{\text{Clifford}, T\}$ circuits of the form shown in Fig. 2 are those that can be expressed by entries in $\mathbb{Z}[i, \sqrt{2}]$ after multiplying by a scalar. Nonetheless, this restricted class can be used to approximate arbitrary unitaries more efficiently than unitaries limited to $\mathbb{Z}[i, \frac{1}{\sqrt{2}}]$, as we show in Section 5 and Section 6.

3.2 Success probability and expected cost

The success probability, i.e., the probability of obtaining the zero outcome for all ancilla measurements, can be computed from (4) and is given by

$$\Pr[\text{success}] = \frac{\alpha_0^2}{2^k} \leq \frac{\alpha_0^2}{2^{\lceil 2 \log_2 \alpha_0 \rceil}} \ , \quad (7)$$

where since $\alpha_0^2 < 2^k$, we may use $k \geq \lceil 2 \log_2 \alpha_0 \rceil$. The circuits in Fig. 1, for example, each yield a value of $\alpha_0 = \sqrt{5}$ and therefore a success probability of $5/8$. If U appears multiple times in (3), then we have

$$\Pr[\text{success}] = \frac{t\alpha_0^2}{2^k} \leq \frac{t\alpha_0^2}{2^{\lceil \log_2 t\alpha_0^2 \rceil}} \ , \quad (8)$$

where t is the number of times that U appears. This upper bound can be made arbitrarily close to one for large enough t .

The expected number of repetitions required in order to achieve success is given by a geometric distribution with expectation value $1/p$, and variance $(1-p)/p^2$, where $p = \Pr[\text{success}]$. If $C(W)$ is the cost of implementing the unitary W , then the expected cost of the RUS circuit is given by $C(W)/p$ with a variance of $C(W)(1-p)/p^2$. The resources required to implement a $\{\text{Clifford}, T\}$ fault-tolerant circuit are often dominated by the cost of implementing the T gate. We therefore define $C(W)$ as the number of T gates in the circuit used to implement W .

The T -gate count is not the only reasonable cost function. Other possibilities include circuit size, width, area or volume, or the total number of measurements. The utility of a particular cost function varies depending on the target quantum computing architecture. For architectures that use the surface code, for example, total volume can be a more complete metric than T count [FDJ13, Jon13a].

Here we choose to use T -gate count as the cost function because it is simple, and is consistent with other $\{\text{Clifford}, T\}$ -decomposition algorithms [KMM12a, AMMR12, Sel12, KMM12b, WK13, GKMR13, RS14]. However, RUS circuits require techniques not present in the circuits produced by previous decomposition methods, such as rapid classical feedback and control, and active

synchronization due to variable time scales per RUS circuit. Thus, while T count allows for direct comparison of RUS circuits with other methods, a more complete metric may be required in the future for resource calculations on a particular hardware architecture.

3.3 Amplifying the success probability

The action of the multi-qubit unitary W may be described by

$$W |0^m\rangle |\psi\rangle = \sqrt{p} |0^m\rangle U |\psi\rangle + \sqrt{1-p} |\Phi^\perp\rangle , \quad (9)$$

where $|\Phi^\perp\rangle$ is a state that depends on $|\psi\rangle$ and satisfies $(|0^m\rangle \langle 0^m| \otimes I) |\Phi^\perp\rangle = 0$. That is, W outputs a state which has amplitude \sqrt{p} on the “success” subspace, and amplitude $\sqrt{1-p}$ on the “failure” subspace. We show that in some cases we may apply amplitude amplification to boost the success probability and reduce the expected T count of an RUS circuit.

Traditional amplitude amplification [BHMT00] proceeds by applying the operator $(RS)^j$ on the initial state $W |0^m\rangle |\psi\rangle$ for some integer $j > 0$ and reflections

$$\begin{aligned} S &= I - 2 |0^m\rangle \langle 0^m| |\psi\rangle \langle \psi| , \\ R &= W S W^\dagger = I - 2 W |0^m\rangle \langle 0^m| |\psi\rangle \langle \psi| W^\dagger . \end{aligned} \quad (10)$$

In the two-dimensional subspace spanned by $\{|0^m\rangle U |\psi\rangle, |\Phi^\perp\rangle\}$, RS acts as a rotation by 2θ where $\sin(\theta) = \sqrt{p}$. Therefore $(RS)^j (W |0^m\rangle |\psi\rangle) = \sin((2j+1)\theta) |0^m\rangle U |\psi\rangle + \cos((2j+1)\theta) |\Phi^\perp\rangle$. The goal then is to choose j appropriately so as to minimize the expected number of T gates.

The problem in this case is that $|\psi\rangle$ is unknown, and therefore we cannot directly implement S . We can, however, implement

$$S' = \overline{\text{CZ}}(m) \otimes I , \quad (11)$$

where $\overline{\text{CZ}}(m) = X^{\otimes m} \text{CZ}(m) X^{\otimes m}$ and $\text{CZ}(m)$ is the generalized controlled- Z gate on m qubits defined by

$$\text{CZ}(m) |x_1, x_2, \dots, x_m\rangle = (-1)^{x_1 x_2 \dots x_m} |x_1, x_2, \dots, x_m\rangle . \quad (12)$$

We could, therefore, apply $(W S' W^\dagger S')^j$ instead of $(RS)^j$.

In the case $m = 1$ (one ancilla qubit) this procedure corresponds to so-called “oblivious” amplitude amplification.

Lemma 3.1 (Oblivious amplitude amplification on $n + 1$ qubits [BCC⁺13]). *Consider a unitary W that satisfies (9) for $m = 1$. Let $S_1 := Z \otimes I$. Then for any $j \in \mathbb{Z}$,*

$$(-W S_1 W S_1)^j W |0\rangle |\psi\rangle = \sin((2j+1)\theta) |0\rangle U |\psi\rangle + \cos((2j+1)\theta) |1\rangle |\phi\rangle , \quad (13)$$

where $\sin(\theta) = \sqrt{p}$.

In fact, oblivious amplitude amplification can be generalized to accommodate any number of ancilla qubits.

Corollary 3.2 (Oblivious amplitude amplification on $n + m$ qubits). *Consider a unitary W that satisfies (9). Oblivious amplitude amplification on $|0^m\rangle U |\psi\rangle$ can be performed using the operator $W S' W^\dagger S'$, where $S' = \overline{\text{CZ}}(m) \otimes I$. More precisely, for any $j \in \mathbb{Z}$*

$$(-W S' W^\dagger S')^j (W |0^m\rangle |\psi\rangle) = \sin((2j+1)\theta) |0^m\rangle U |\psi\rangle + \cos((2j+1)\theta) |\Phi^\perp\rangle , \quad (14)$$

where $\sin(\theta) = \sqrt{p}$.

Proof. The main technical part the proof of Lemma 3.1 in [BCC⁺13] is accomplished by another Lemma called the 2D Subspace Lemma (see Lemma 3.6 of [BCC⁺13]). Like Lemma 3.1, the 2D Subspace Lemma is stated specifically for the $m = 1$ case. However, the proof still holds if $|0\rangle$ is replaced by $|0^m\rangle$. In that case, we find that the state

$$|\Psi^\perp\rangle := W^\dagger \left(\sqrt{1-p} |0^m\rangle U |\psi\rangle - \sqrt{p} |\Phi^\perp\rangle \right) \quad (15)$$

is both orthogonal to $|0^m\rangle |\psi\rangle$ and satisfies $(|0^m\rangle \langle 0^m| \otimes I) |\Psi^\perp\rangle = 0$. This allows us to calculate the behavior of W^\dagger within the two-dimensional subspace spanned by $|0^m\rangle U |\psi\rangle$ and $|\Phi^\perp\rangle$. We have

$$\begin{aligned} W^\dagger(|0^m\rangle U |\psi\rangle) &= \sqrt{p} |0^m\rangle |\psi\rangle + \sqrt{1-p} |\Psi^\perp\rangle \\ W^\dagger |\Phi^\perp\rangle &= \sqrt{1-p} |0^m\rangle |\psi\rangle - \sqrt{p} |\Psi^\perp\rangle \end{aligned} \quad (16)$$

Just as in [BCC⁺13], this permits simple calculations yielding

$$-WS'W^\dagger S'(|0^m\rangle U |\psi\rangle) = \cos(2\theta) |0^m\rangle U |\psi\rangle + \sin(2\theta) |\Phi^\perp\rangle \quad (17)$$

and

$$-WS'W^\dagger S |\Psi^\perp\rangle = \sin(2\theta) |0^m\rangle U |\psi\rangle + \cos(2\theta) |\Phi^\perp\rangle \quad (18)$$

The conclusion is that $-WS'W^\dagger S'$ acts as a rotation by 2θ in the two-dimensional subspace of interest. \square

If $m \leq 2$, then S' can be implemented with only Clifford gates, i.e., X and either Z or CZ . Then, for a fixed value of j , the total number of T gates in the corresponding amplified circuit is given by $(2j+1)T_0$. In order for amplitude amplification to yield an improvement in the expected number of T gates, we therefore require that

$$(2j+1) \sin^2(\theta) < \sin^2((2j+1)\theta) \quad , \quad (19)$$

a condition that holds if and only if $0 \leq p < 1/3$. Thus a sensible course of action is to apply amplitude amplification for all RUS circuits for which $p < 1/3$, and leave higher probability circuits unchanged.

Consider, for example, an RUS circuit that contains 15 T gates and has a success probability of 0.1. In this case, using amplitude amplification with a value of $j = 1$ yields a new circuit with success probability 0.676 and 45 T gates, an improvement in the expected number of T gates by a factor of 2.25. The effects of amplitude amplification on our database of RUS circuits are discussed in Section 5.

Cost analysis of amplitude amplification for circuits with more than two ancilla qubits is more complicated because the reflection operator $S' = \overline{CZ}(m)$ is not a Clifford gate. For three ancilla qubits, for example, S' requires the controlled-controlled- Z gate, which can be implemented with 4 T gates [Jon13b]. Larger versions of $CZ(m)$ could be synthesized directly [Kli13, WGMAG13], or by using a recursive procedure [NC00]. The circuits presented in Section 5 use at most two ancilla qubits, however, so more complicated amplification circuits are not an issue in our analysis.

4 Direct search algorithm

While equations (3) and (6) restrict the kinds of unitaries that can be exactly obtained with RUS circuits, they indicate very little about how to implement the multi-qubit unitary W . Given W explicitly, it is possible to synthesize a corresponding $\{\text{Clifford}, T\}$ circuit with a minimum number of T gates [GKMR13], at least for W with small T count. However, given a unitary U of the form (5), there are potentially many choices of W , and an efficient way to find the W that will result in the minimum number of T gates is unknown (and a direction for future research).

As a step towards synthesizing RUS circuits and understanding their scope, we design an optimized direct search algorithm that synthesizes RUS circuits up to a given T -gate count. Our direct search algorithm is as follows:

1. Select the number of ancilla qubits and the number of gates.
2. Construct a $\{\text{Clifford}, T\}$ circuit and compute the resulting unitary matrix W .
3. Partition the first two columns of W into 2×2 matrices.
4. Identify and remove matrices that are proportional to Clifford gates.
5. If the remaining matrices are all proportional to the same unitary matrix, then keep the corresponding circuit.

We restrict the recovery operations R_i of the circuits in our direct search to the set of single-qubit Cliffords. This choice is motivated by our use of the T count as a cost function; Clifford gates, and therefore the recovery operations are assigned a cost of zero, therefore such recovery operations are inexpensive.

In order to identify relevant search parameters for step 1 and circuit constructions for step 2, we initially performed a random search over a wide range of circuit widths (number of qubits) and sizes (number of gates). Our search produced ample results for small numbers of ancilla qubits, large numbers of T gates, and just one or two entangling gates. We therefore focus our current study on circuits of the form shown in Fig. 3, which contain one ancilla qubit and two CZ gates, interleaved with single-qubit Clifford gates.

Naively, the number of circuits of the form given in Fig. 3 is $O(3^n)$, where n is the maximum number of (non-CZ) gates in the circuit, and the base of three is the size of the set $\{H, S, T\}$. In order to reduce the time complexity of direct search, we constructed each single-qubit gate sequence using the canonical form proposed in [BS12]. A canonical form sequence is the product of three 2×2 unitary matrices $g_2 C g_1$ where g_1, g_2 belong to the single-qubit Clifford group, and C is the product of some number of “syllables” TH and $SHTH$. The canonical form yields a unique representation of all single-qubit circuits over $\{H, T\}$; there are $2^{t-3} + 4$ canonical circuits of T -count at most t . The canonical representation yields more than a quadratic improvement in time complexity compared to naive search, since the number of T gates is roughly one-half the total number of gates.

In general, the canonical form requires conjugation by the full single-qubit Clifford group, which contains 24 elements. Given a product of syllables C , each of the $24^2 = 576$ circuits $g_2 C g_1$ are unique. However, when multiple canonical form circuits are composed in a larger circuit, as in Fig. 3, some combinations of Clifford gates can be eliminated. For example, when $g_2 C g_1$ is applied to the state $|0\rangle$, g_1 need only be an element of $\{I, X, SH, SHX, HSH, HSHX\}$ since diagonal gates act trivially on $|0\rangle$. Similar simplifications for Fig. 3 are shown in Fig. 4. In total, these Clifford optimizations further reduce the search space by a factor of more than 10^5 .

Despite these optimizations, our direct search algorithm still requires time exponential in the number of T gates. To further reduce the time complexity, we partitioned the search into thousands

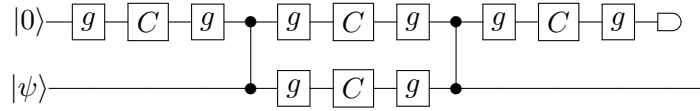


Figure 3: The general form of most RUS circuits in our database. Each of the gates labeled g represents an element of the single-qubit Clifford group. Each of the gates labeled C represents a single-qubit canonical circuit as defined in [BS12].

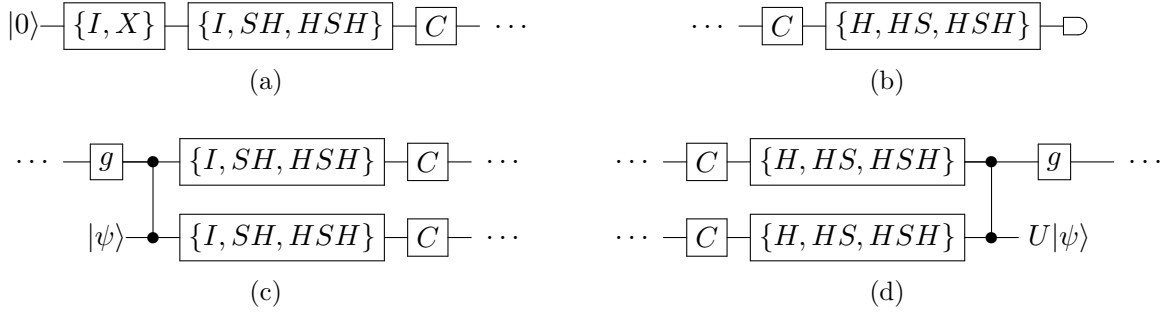


Figure 4: Some gates g in Fig. 3 can be restricted to a subset of the single-qubit Clifford group. (a) Circuits that begin with diagonal gates can be eliminated since they add a trivial phase to $|0\rangle$. (b) Similarly, diagonal gates have no impact on the Z -basis measurement. (c) Pauli gates and S gates can be commuted through the CZ and absorbed into either $|\psi\rangle$ or the preceding g gate. (d) Analogously, Pauli and S gates occurring before the CZ can be absorbed by the trailing g gate or by the output.

of small computations running in parallel on a large cluster and collected the results in a central database. We were able to exhaustively synthesize circuits of the form given in Fig. 3 up to a total (raw) T count of 15 in roughly one week running on hundreds of cores. The results of our direct search algorithm are presented in the next section.

5 Direct search results

Our search yielded many RUS circuits that implement the same unitary U , but with different T -gate counts and success probabilities. To eliminate redundancy we construct a database containing only the circuit with the minimum expected T count for a given unitary U . The resulting database contains 2194 RUS circuits each of which contains at most 15 T gates. Upon success, each circuit exactly implements a unique non-Clifford single-qubit unitary U , and otherwise implements a single-qubit Clifford operation. The database statistics are shown in Fig. 5. For circuits with success probability less than $1/3$, we used amplitude amplification to improve performance (see Section 3.3). Most RUS circuits result in high success probability and low expected T count. Fig. 5b illustrates the impact of amplitude amplification on the expected T count. Amplification improved the performance of circuits with relatively high expected T count, but did not improve circuits with expected T count of 30 or less. In general, RUS circuits exhibit very low expected T counts around 15–20. Note that the database also includes some circuits that were found by preliminary searches not of the form of Fig. 3.

Of the 2194 RUS circuits, 1659 are axial rotations, i.e., unitaries which, modulo conjugation by Cliffords, are rotations about the Z -axis of the Bloch sphere, and 535 are non-axial rotations. The number of axial rotations is noteworthy since, modulo Clifford conjugation, only one non-trivial single-qubit rotation can be exactly synthesized with $\{\text{Clifford}, T\}$ and without measurement, namely T [KMM12a]. Our results show that *many* axial rotations can be implemented exactly (conditioned on success) when measurement is allowed.

Remarkably, the non-axial rotations in our database offer an expected T count that is dramatically better than the T count obtained by approximation algorithms [Sel12, KMM12b, RS14]. For each RUS circuit in the database we computed the number of T gates required to approximate the corresponding unitary to within a distance of 10^{-6} using the algorithm of KMM. Fig. 6 shows the ratio of the T count given by KMM vs. the expected T count for the RUS circuit. (KMM and Ross-Selinger achieve similar T count scaling so we expect similar ratios when comparing to Ross-Selinger.) Our results show a typical improvement of about a factor of three for axial rotations and a typical improvement of about a factor of about 12 for non-axial rotations. The larger improvement for non-axial rotations is expected since the KMM algorithm requires the unitary to be first decomposed into a sequence of three axial rotations.

As an example, the RUS circuit shown in Fig. 7 implements the non-axial single-qubit rotation $U = (2X + \sqrt{2}Y + Z)/\sqrt{7}$ with four T gates and a probability of success of $7/8$. By contrast, approximating U to within $\epsilon = 10^{-6}$ using the KMM algorithm requires a total of 182 T gates. Thus the circuit in Fig. 7 not only implements the intended unitary exactly, but does so at a cost over 40 times less than the best approximation methods.

Our database is too large to offer an analysis of each circuit in detail. However, we highlight some particularly important examples. The smallest circuit in our database contains two T gates and is shown in Fig. 8. Upon measuring zero, which occurs with probability $3/4$, the circuit implements $(I + i\sqrt{2}X)/\sqrt{3}$ and upon measuring one implements I . This circuit was predicted to

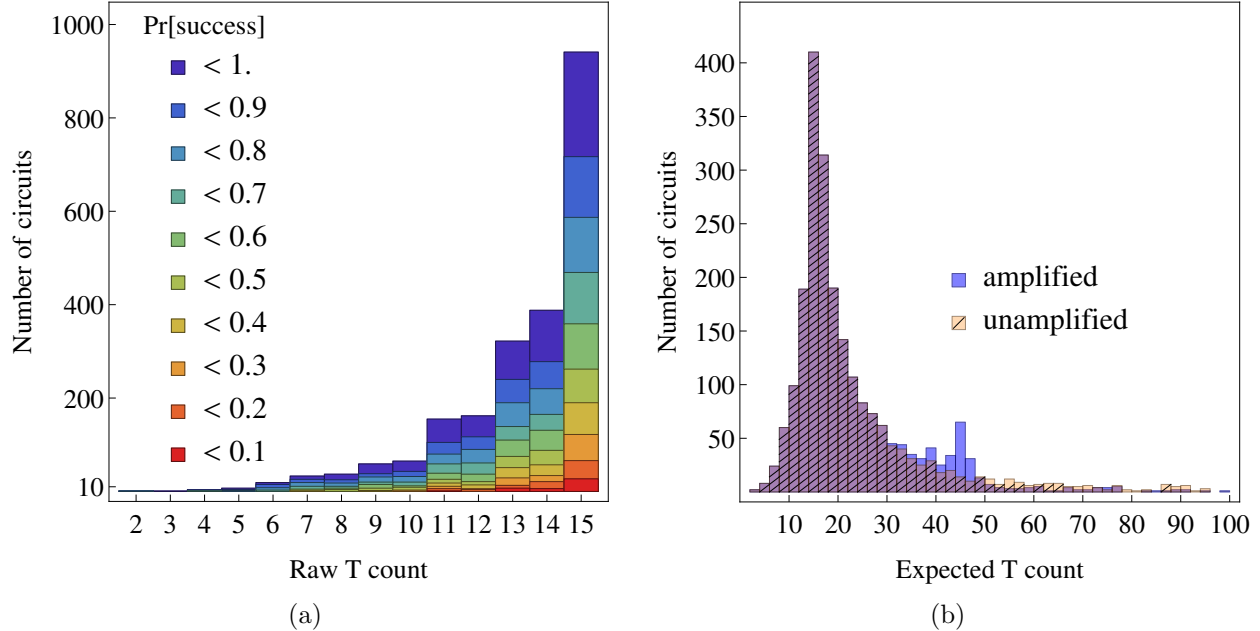


Figure 5: Statistics for the database of repeat-until-success circuits, including all circuits of the form of Fig. 3 up to a T count of 15. (a) The total number of circuits grouped by (raw) T gate count and success probability. (b) The total number of circuits grouped by expected T count, both before amplitude amplification and after amplitude amplification. The two histograms (before amplification and after amplification) are overlaid, where the darker hatched bars indicate circuits that are unaffected by amplification. Only circuits with an expected T count of at most 100 are shown.

exist by Gosset and Nagaj [GN13]. They required a $\{\text{Clifford}, T\}$ circuit that exactly implemented $R = (\sqrt{2}I - iY)/\sqrt{3}$ with a constant probability of success. The unitary implemented by Fig. 8 is equivalent to R up to conjugation by Clifford gates.

As discussed in Section 1, our database contains a circuit that implements V_3 . In addition to the circuit shown in Fig. 1c, our search also found a circuit that implements V_3 with the same number of T gates, but with just a single ancilla qubit, as shown in Fig. 9. The expected T count of the single-ancilla circuit is slightly worse than that of Fig. 1c, though, since all four of the T gates on the ancilla must be performed “online”.

The V_3 gate is one of a family of V -basis gates for which the normalization factor is $1/\sqrt{5}$. In addition to single-qubit unitary decomposition based on V_3 , [BGS13] also offers the possibility of decomposing single-qubit unitaries using V -basis gates with normalization factors $1/\sqrt{p}$ where p is a prime. These “higher-order” V gates cover $SU(2)$ more rapidly than V_3 and therefore offer potentially more efficient decomposition algorithms. A number of such V -basis gates can be found in our database, including axial versions for $p \in \{13, 17, 29\}$, as shown in Fig. 10, offering the first fault-tolerant implementations of these gates. The prospect of decomposition algorithms with these circuits is discussed in Section 6.1.

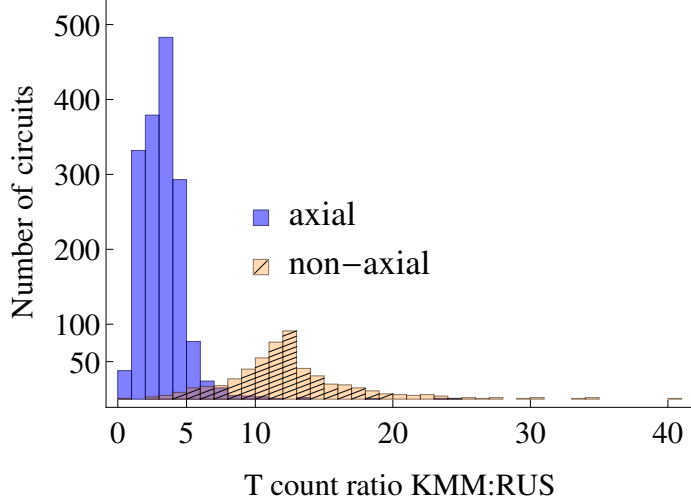


Figure 6: RUS circuits database split into axial and non-axial single-qubit rotations. For each circuit, the number of T gates required to approximate the corresponding “success” unitary U to within 10^{-6} was calculated using the algorithm of [KMM12b]. The x -axis represents the ratio of the KMM T count vs. the expected number of T gates for the RUS circuit.

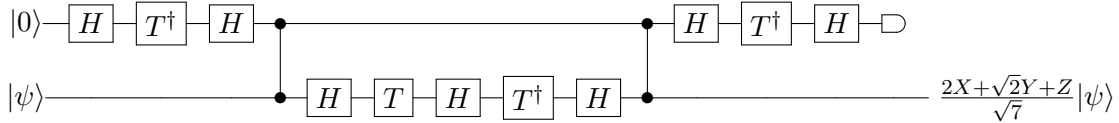


Figure 7: An RUS circuit to implement the unitary $U = (2X + \sqrt{2}Y + Z)/\sqrt{7}$ with probability $7/8$, and Z otherwise. Approximation of U without ancillas requires 182 T gates (roughly 40 times more) for $\epsilon = 10^{-6}$.

6 Applications

One application of RUS circuits is in the construction of universal sets of gates. Our RUS circuits offer exact, fault-tolerant implementations of a large set of single-qubit unitary gates. The Clifford group plus any one non-Clifford gate is universal for quantum computation (see, e.g., [CAB12] Appendix D). Thus any of our RUS circuits can be used to construct a new universal gate set. The question, though, is whether or not RUS circuits can be used to decrease resource costs of unitary approximation methods.

In this section, we show that RUS circuits can be used to significantly improve upon approximate decomposition of single-qubit unitaries. First we discuss the use of our improved V_3 circuit for

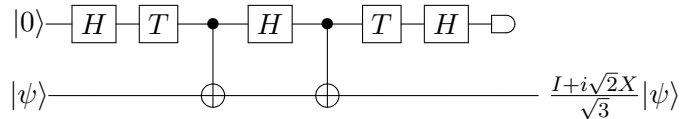


Figure 8: The smallest circuit in our database. Upon measuring zero, with probability $3/4$, it implements $(I + i\sqrt{2}X)/\sqrt{3}$ on the input state $|\psi\rangle$. Upon measuring one, it implements the identity.

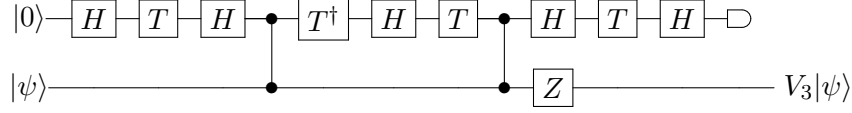


Figure 9: A circuit, like the circuits in Fig. 1, to implement V_3 with probability $5/8$ and identity with probability $3/8$, using only one ancilla qubit and one measurement.

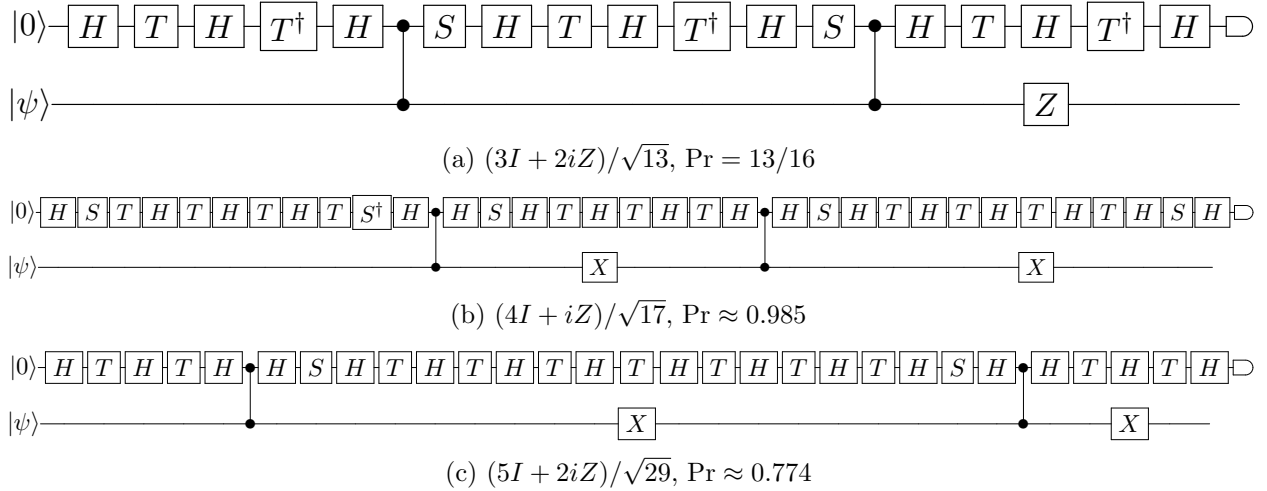


Figure 10: RUS circuits for V -basis gates with prime normalization factors (a) $p = 13$ (b) $p = 17$ and (c) $p = 29$. The values under each circuit indicate the unitary effected upon success and the success probability, respectively. Each circuit implements the identity upon failure.

decomposition into $\{\text{Clifford}, V_3\}$. Then we show how to compose RUS circuits in series in order to expand the size and density of the database. The expanded database can be used to approximate single-qubit unitaries up to an accuracy that is sufficient for a number of important quantum algorithms. In particular, in Section 6.3, we show how to use circuits in our database for applications using the quantum phase estimation algorithm.

6.1 Decomposition with V_3

The RUS circuit for V_3 , shown in Fig. 1c, can be used directly in the decomposition algorithm of [BGS13]. The BGS algorithm produces an ϵ -approximation of a given single-qubit unitary with $3 \log_5(1/\epsilon)$ V_3 gates in most cases. Multiplying by an expected T -cost of 5.26, using the circuit in Fig. 1c, yields an algorithm with an expected T count of

$$15.78 \log_5(1/\epsilon) . \quad (20)$$

This is an improvement over the estimated T count of $3(3.21 \log_2(3/\epsilon) - 6.93)$ [KMM12b] for all $\epsilon < 0.25$.

The database also contains V -basis gates with prime normalization factors larger than 5. In [BGS13], the authors conjecture that the decomposition algorithm for $p = 5$ extends to other primes with a T -count scaling of $4 \log_p(1/\epsilon)$. However, whereas $p = 5$ requires only the single V_3 gate, higher prime values require implementation of multiple V gates. For simplicity, assume that each of the required V gates can be implemented with T -count T_p . Then the decomposition achieved for prime p will be better than that obtained with V_3 if

$$1 < \frac{5.26}{T_p} \log_5(p) . \quad (21)$$

Unfortunately, our database contains only a single V -basis gate for each of $p = \{13, 17, 29\}$. For the sake of argument, we calculate (21) under the optimistic assumption that for each p , the remaining V gates can someday be implemented at the same cost T_p . Using the circuits in Fig. 10 we obtain

$$5.26/7.38 \log_5 13 \approx 1.13, \quad (22a)$$

$$5.26/11.17 \log_5 17 \approx 0.83, \quad (22b)$$

$$5.26/14.22 \log_5 17 \approx 0.77 . \quad (22c)$$

Based on these calculations we conclude that, while improved decomposition may be possible using $p = 13$, higher values of p are unlikely to yield cost benefits on their own.

On the other hand, given implementations of multiple V gates, there is no reason to limit to a single value of p . One could imagine an algorithm that combined multiple classes of V gates, using largely V_3 and using more expensive high-order V gates selectively. We do not consider such an algorithm directly. In the next section, however, we study the effect of optimally combining all of the RUS circuits in our database, not just V gates.

6.2 Decomposition by composition of RUS circuits

It is possible to approximate a given single-qubit unitary U to within any ϵ by composing Clifford gates and circuits from our database. But finding the optimal composition sequence among all

possible compositions of circuits is a challenging task. Ideally, we could construct an efficient decomposition algorithm based on algebraic characterization of the set of RUS circuits, similar to algorithms for other gate sets [Sel12, KMM12b, BGS13, RS14]. But the current theoretical characterization of RUS circuits remains open is a direction for future work. Here, we develop decomposition algorithm based on exhaustive composition of RUS circuits, which is similar in nature to the methods of [Fow11] and [BS12].

Starting with the set of RUS circuits found by our direct search algorithm, we compute all products of pairs of circuits, keeping those that produce a unitary which is not yet in the database. Composite circuits of arbitrary size can be constructed in this manner: triples of circuits can be constructed from singles and pairs, and so on. Call a circuit a class- k circuit if it is composed of a k -tuple of RUS circuits from the original database. Then the number N_k of class- k circuits is bounded by

$$N_k \leq N_1 \cdot N_{k-1} \leq N_1^k , \quad (23)$$

where N_1 is the number of circuits in the original database.

To manage the database expansion, we keep only those circuits that yield an expected T count of at most some fixed value T_0 . This has the simultaneous effect of discarding poorly performing circuits and reducing the value of N_k so that construction of class- $(k + 1)$ circuits is less computationally expensive. Furthermore, circuits can be partitioned into equivalence classes by Clifford conjugation. The unitaries of the initial set of circuits are of the form $g_0 U g_1$, where U is the unitary obtained from the RUS circuit, and g_0, g_1 are single-qubit Cliffords. Thus, the product of k such circuits has the form

$$g_0 U_1 g_1 U_2 g_2 \dots U_k g_k . \quad (24)$$

The set of class- $(k + 1)$ circuits can then be constructed by using

$$g_0 U_1 g_1 U_2 g_2 \dots U_k g_k (g_k U_{k+1} g_{k+1}) = g_0 U_1 g_1 U_2 g_2 \dots U_k g_k'' U_{k+1} g_{k+1} , \quad (25)$$

so that the Clifford g_k is unnecessary. Furthermore, g_0 can always be prepended later, and so we instead express each class- k unitary as

$$U_1 g_1 U_2 g_2 \dots U_k . \quad (26)$$

To find an equivalence class representative of U , we first remove the global phase by multiplying by $u^*/\sqrt{|u|^2}$, where u is the first non-zero entry in the first row of U . Next, we conjugate U by all possible pairs of single-qubit Cliffords. The first element of a lexicographical sort then yields the representative $g_1 U g_2$ for some Cliffords g_1, g_2 .

Once the expanded database has been constructed up to a desired size, the decomposition algorithm is straightforward. Given a single-qubit unitary U and $\epsilon \in [0, 1]$, select all database entries V such that $D(U, V) \leq \epsilon$, where

$$D(U, V) = \sqrt{\frac{2 - |\text{Tr}(U^\dagger V)|}{2}} \quad (27)$$

is the distance metric defined by [Fow11] and also used by [Sel12, KMM12b, BGS13, WK13, RS14]. Then, among the selected entries, find and output the circuit with the lowest expected T count.

6.2.1 Results: decomposition with axial rotations

An arbitrary single-qubit unitary can be decomposed into a sequence of three Z -axis rotations and two Hadamard gates [NC00]. Therefore, approximate decomposition of Z -axis rotations suffices to approximate any single-qubit unitary. If we limit to Z -axis, i.e, diagonal, rotations only, then a few additional simplifications are possible. In particular, each unitary can be represented by a single real number corresponding to the rotation angle in radians. The result of a sequence of such rotations is then given by the sum of the angles. Furthermore, up to conjugation by $\{X, S\}$, all Z -axis rotations can be represented by an angle in the range $[0, \pi/4]$. This allows for construction of a database of Z -axis rotations which is much larger than a database of arbitrary (non-axial) unitaries.

Using the database expansion procedure described above, we construct a database containing all combinations of RUS circuits with expected T count at most 30. The maximum distance (according to (27)) between any two neighboring rotations is less than 2.8×10^{-6} , and can be improved to 2×10^{-6} by selectively filling the largest gaps. So the resulting database permits approximation of any Z -axis rotation to within $\epsilon = 10^{-6}$.

To approximate a Z -axis rotation by an angle θ , we select all entries that are within the prescribed distance ϵ , and then choose the one with the smallest expected T count. This procedure is efficient since the database can be sorted according to rotation angle. Then the subset of entries that are within ϵ can be identified by binary search.

In order to assess the performance of this method, we approximate, for various values of ϵ , a sample of 10^5 randomly generated angles in the range $[0, \pi/4]$. Results are shown in Fig. 11 and Table 3. A fit of the mean expected T count for each ϵ yields a scaling given by (1), with a slope roughly 2.4 times smaller than that reported by [KMM12b] for the rotation $R_Z(1/10)$.

By way of comparison, Wiebe and Kliuchnikov report a scaling of $1.14 \log_2(1/\theta)$ for small angles θ . However, their RUS circuits are specially designed for small angles. For arbitrary angles they report an expected T count of about

$$1.14 \log_2(10^\gamma) + 8 \log_2(10^{-\gamma}/\epsilon) \ , \quad (28)$$

where $\theta = a \times 10^{-\gamma}$ for some $a \in (0, 1)$ and integer $\gamma > 0$. Using (28) to calculate costs for the same 10^5 random angles as above, we obtain a fit function of

$$6 \log_2(1/\epsilon) - 2.2 \ . \quad (29)$$

Equation (29) indicates that the efficiency of the circuits in [WK13] does not extend to coarse angles. Nevertheless, in Section 6.2.2 we show how to combine the circuits of Wiebe and Kliuchnikov with our RUS circuits to achieve good cost scaling for relatively high accuracies.

Equation (1) also implies that RUS Z -axis rotations can be used to approximate *arbitrary* single-qubit unitaries with a scaling approaching that of optimal ancilla-free decomposition. Since an arbitrary unitary can be expressed as a product of three axial rotations, the expected T count for approximating an arbitrary single-qubit unitary is given by $3.9 \log_2(3/\epsilon) - 8.37$. On the other hand, Fowler calculates an optimal T -count of $2.95 \log_2(1/\epsilon) + 3.75$ (on average) without using ancillas [Fow11].

Since our circuits are non-deterministic, we are also concerned with the probability distribution of the number of T gates. For each composite circuit in the database, we calculate the variance σ^2 of the T count based on the variance of each individual circuit. We then obtain a confidence

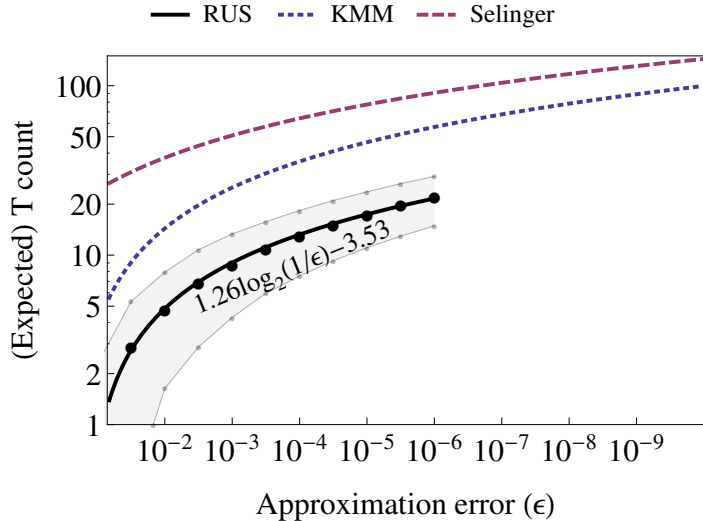


Figure 11: The expected number of T gates required to approximate a single-qubit Z -axis rotation to within a distance ϵ over 10^5 real numbers selected in the range $[0, \pi/4]$ uniformly at random. For each value θ , the RUS circuit with the smallest expected T count within ϵ of the unitary $R_Z(\theta)$ was selected. The mean for each value of ϵ is plotted, yielding a fit-curve of $1.26 \log_2(1/\epsilon) - 3.53$. The gray region is an estimate of the interval containing the actual number of T gates with probability 95%. The other curves are included for reference: KMM = $3.21 \log_2(1/\epsilon) - 6.93$ [KMM12b], Selinger = $4 \log_2(1/\epsilon) + 11$ [Sel12].

$\log_{10}(1/\epsilon)$	Exp T (σ^2)	$\pm 95\%$ (σ^2)
1	1.1 (1.1)	1.2 (3.6)
1.5	2.9 (2.2)	2.5 (2.9)
2	4.8 (3.4)	3.1 (2.9)
2.5	6.8 (3.9)	4.0 (3.8)
3	8.8 (4.3)	4.5 (4.7)
3.5	10.9 (4.6)	4.9 (5.2)
4	12.9 (4.8)	5.4 (5.5)
4.5	15.1 (5.3)	5.9 (5.7)
5	17.4 (5.7)	6.3 (5.8)
5.5	19.6 (6.0)	6.7 (6.1)
6	22.0 (6.4)	7.1 (6.5)

Table 3: Expected T count required to approximate a random single-qubit Z -axis rotation with an RUS circuit. The middle column indicates the expected T count based on a sample of 10^5 random angles. The right-hand column indicates the expected 95 percent confidence interval of the T count for the best RUS circuit, given a random angle θ . The variance of each expected value is indicated in parenthesis.

interval using Chebyshev’s inequality

$$\Pr(|\text{Actual}[T] - \text{Exp}[T]| \geq k\sigma) \leq \frac{1}{k^2} . \quad (30)$$

Table 3 shows the mean expected T count for each ϵ . By also calculating the mean variance σ^2 , we obtain an estimate of the corresponding 95% confidence interval, shown by the gray region in Fig. 11. That is, for a randomly chosen angle θ , the actual number of T gates required to implement $R_Z(\theta)$ is within the given interval around $1.26 \log_2(1/\epsilon) - 3.53$, with probability 0.95.

The approximation accuracy permitted by our database is limited by computation time and memory. To maximize efficiency, we used floating-point (accurate to 14 digits) rather than symbolic arithmetic. Construction of all RUS circuit combinations up to expected T count of 30 took roughly 20 hours and 41 GB of memory using Mathematica. Table 4 shows the number of circuit combinations and corresponding rotation angle densities for increasing values of the expected T count. The size and density of the database increases by roughly one order of magnitude for every five T gates. We expect that with a more efficient implementation—in C/C++ for example—the worst-case approximation accuracy could be improved.

Max. exp.			
T count	Size	Mean D	Max D
5	7	0.04	0.08
10	134	0.0021	0.0066
15	2079	0.00013	0.0014
20	27420	0.00001	0.00017
25	320736	0.0000009	0.000016
30	3446708	0.00000008	0.0000028

Table 4: Size and density of the Z -axis rotation database according to the maximum expected number of T gates. The mean and the maximum distances between nearest neighbors is given in columns three and four, respectively.

6.2.2 More accurate axial rotations using gearbox circuits

The approximation accuracy of Z -axis rotations can be improved indirectly by combining our database of axial rotations with the floating-point approach of Wiebe and Kliuchnikov [WK13]. In their approach, a Z -axis rotation by angle $\phi = a \times 10^{-\gamma}$ is approximated with a “gearbox” circuit that multiplies the mantissa $a \in (0, 1)$ by the value $10^{-\gamma}$. The T count of the gearbox circuit scales as

$$\text{Exp}_Z^{\text{WK}}[T] = 2T(a, 10^\gamma \epsilon) + 1.14 \log_2(10^\gamma) + 12.2 \quad , \quad (31)$$

where $T(a, \epsilon)$ is the number of T gates required to approximate $R_Z(a)$ to within a distance ϵ . In [WK13], Selinger’s algorithm is used to approximate the mantissa a . However, any approximation method may be used.

The gearbox circuits are most useful when the angle ϕ is very small, and the number of significant digits $m = \log_{10}(10^{-\gamma}/\epsilon)$ is also small. In that case, (31) is largely determined by the $1.14 \log_2(10^\gamma)$ term, which scales better than any other known methods. The scaling is maintained even for very high accuracy, so long as the required relative precision is low.

If our decomposition method based on RUS circuits is used to approximate $R_Z(a)$ (instead of Selinger’s method), then we obtain

$$\text{Exp}_Z^{\text{WK}}[T] = 2.52 \log_2(10^{-\gamma}/\epsilon) + 1.14 \log_2(10^\gamma) + 5.14 \quad , \quad (32)$$

which is an improvement over the direct methods due to Selinger and KMM, even for large angles. The density of the database presented in Section 6.2.1 permits a maximum of $m = 6$ significant digits; a larger database would permit higher precision.

If full precision is required (i.e., $\gamma = 0$), then a slightly different method can be used. Given an angle θ and error $10^{-6} > \epsilon \geq 10^{-11}$, an approximation of $R_Z(\theta)$ can be obtained by first using the RUS axial rotation database to get $R_Z(\tilde{\theta})$ such that $|\tilde{\theta} - \theta| = \phi \leq 10^{-6}$. Then, a gearbox circuit can be used to approximate $\phi = a \times 10^{-\gamma}$ to within the prescribed distance ϵ , where $R_Z(a)$ is obtained by again using the RUS database. The expected T count is estimated by

$$\text{Exp}_Z^{\text{hybrid}}[T] = 1.26 \log_2(1/\delta) + 2 \cdot 1.26 \log_2(10^{-\gamma}/\epsilon) + 1.14 \log_2(10^\gamma) + 1.61 \quad , \quad (33)$$

where δ is the selected accuracy of the approximation $\tilde{\theta}$. Assuming $\phi \approx \delta$ and therefore $10^\gamma \approx 10/\delta$, we obtain

$$\text{Exp}_Z^{\text{hybrid}}[T] \approx 2.52 \log_2(1/\epsilon) - 0.12 \log_2(1/\delta) - 2.97 \quad . \quad (34)$$

Thus, an effective strategy is to approximate θ to the maximum accuracy permitted by the axial RUS database ($\delta = 10^{-6}$) and then approximate the remaining angle ϕ with a gearbox circuit.

The coarse approximation $\tilde{\theta}$ will often be better than 10^{-6} so the actual scaling may vary from (34). To check, we calculated for $\epsilon \geq 10^{-11}$, the cost of the hybrid approach for the same 100k angles used in Section 6.2.1. The results yield an empirical fit of $2.62 \log_2(1/\epsilon) - 3.1$, which is slightly higher than (34), but still lower than that reported by KMM.

Even higher accuracy can be obtained by recursively applying the hybrid procedure. If the mantissa a of ϕ requires more accuracy than the RUS database can provide, then $R_Z(a)$ can be coarsely approximated using the database and the remainder can be obtained using another gearbox. Asymptotically, such an approach has scaling $\Theta((1/\epsilon)^{1/\log_2(1/\delta)})$, making it practical only for a limited range of $\epsilon > 10\delta^2$.

6.2.3 Results: Decomposition with non-axial rotations

While it suffices to use three Z -axis rotations and two Hadamard gates to decompose an arbitrary single-qubit non-axial rotation, this process, used by [KMM12b], [Sel12] and [RS14], incurs a factor of three increase in cost, since each axial rotation must in turn be decomposed. This effect is illustrated in Fig. 6 by the larger ratios for non-axial unitaries. Using just our axial database for non-axial unitary decomposition results in a similar increase in cost. Although Fowler’s method [Fow11] does not incur the additional cost for arbitrary unitaries, maintaining a scaling of $2.95 \log_2(1/\epsilon) + 3.75$, the method is exponential and does not achieve exact implementation for many unitaries. RUS circuits, on the other hand, offer a large domain of exactly implementable unitaries. As Fig. 6 suggests, composing both axial and non-axial RUS circuits could yield better approximations than using Z -axis rotations alone.

Construction of the database in the non-axial case is significantly more challenging than in the axial case. First, unitaries must be represented by three rotation angles instead of one. Second, composition of circuits requires multiplication in the non-axial case, which is less efficient than for the Z -axis case which only requires addition. Third, organization of the database to enable efficient lookup is more complicated; Z -axis rotations can be sorted by rotation angle, while arbitrary unitaries require a more complicated data structure such as a k -d tree [DN05, Amy13].

However, we can express each unitary by its Clifford equivalence class representative (26), and also avoid conjugating by all 576 pairs of Clifford gates. Since any single-qubit Clifford can be written as a product $g_1 g_2$ where $g_1 \in G_1$, $g_2 \in G_2$ and

$$\begin{aligned} G_1 &= \{I, Z, S, S^\dagger\} \\ G_2 &= \{I, H, X, XH, HS, XHS, HSH, XHSH\} \end{aligned} \quad (35)$$

then we need only conjugate by G_2 . Each resulting unitary can then be decomposed into three rotations

$$g_2 U g_2' = R_Z(\theta_1) R_X(\theta_2) R_Z(\theta_3) \quad (36)$$

The Clifford gates in G_1 are diagonal and only modify θ_1 and θ_3 . Up to conjugation by elements of G_1 , we have

$$R_Z(\theta_1) R_X(\theta_2) R_Z(\theta_3) \equiv R_Z(\theta_1 \bmod \pi/2) R_X(\theta_2) R_Z(\theta_3 \bmod \pi/2) \quad (37)$$

Choosing $0 \leq \theta_1, \theta_2 < \pi/2$, we can find an equivalence class representative without actually conjugating by G_1 , saving a factor of $576/64 = 9$.

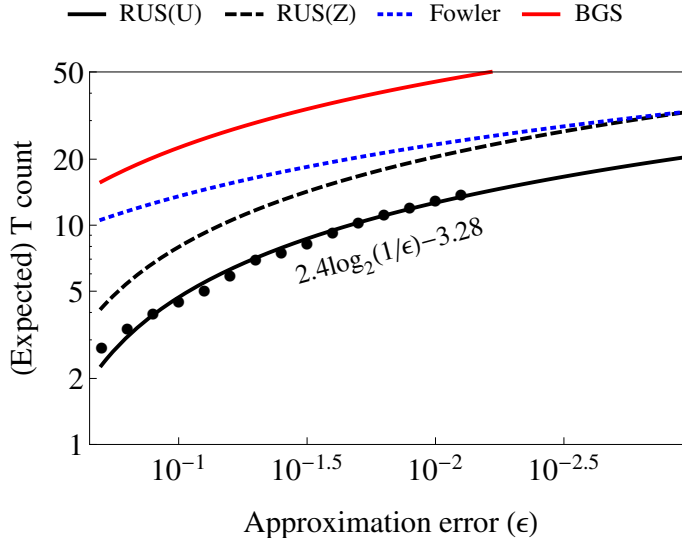


Figure 12: The expected number of T gates required to approximate an arbitrary single-qubit unitary to within distance ϵ . Each point indicates the mean of 100 random unitaries approximated to the corresponding accuracy with our full database of RUS circuits. With 95 percent confidence, the solid black line has slope in the range $[2.29, 2.51]$. The dashed black line indicates the estimated cost of first expressing the unitary as a product of axial rotations, and then decomposing each rotation using the Z -axis RUS database from Section 6.2.1. The solid red line indicates the scaling obtained by using the circuit in Fig. 1c for V_3 decomposition [BGS13]. The scaling is worse than the others, but is valid for $\epsilon \geq 10^{-10}$. The estimated scaling using exponential direct search (Fowler [Fow11]) is shown for reference.

Using these optimizations, we construct a database of size 45526 containing all RUS circuits with expected T count at most 18. We calculated the best circuit for 100 random single-qubit unitaries for a range of $\epsilon \geq 8 \times 10^{-3}$. A fit-curve of the data yields a scaling of $\text{Exp}_U[T] = 2.4 \log_2(1/\epsilon) - 3.28$. Based on the slope, the savings is roughly 18 percent over Fowler; in absolute terms, the savings is roughly a factor of two for modest approximation accuracy. See Fig. 12. Given the relatively large ratios for non-axial unitaries in Fig. 6 and the fact that our database contains only a limited subset of possible RUS circuits, by incorporating a larger set of circuits, we expect the scaling to further improve.

6.3 Quantum algorithms using coarse angles

The accuracy of our decomposition method is limited by the size of the database. Our Z -axis rotation database is capable of approximating arbitrary rotations up to an accuracy of 10^{-6} . To achieve higher accuracy, either the database must be expanded, or an algorithmic decomposition such as that of Section 6.1 must be used. However, a variety of important quantum algorithms require only limited rotation accuracies. Fowler, for example, used numerical analysis to argue that Shor’s algorithm requires rotation angles no smaller than $\theta = \pi/64 \approx 0.05$ with an approximation error of $\epsilon = \pi/256 \approx 0.012$ [FH04].

Another application of coarse angles is in quantum chemistry. Consider a Hamiltonian for a

molecule expressed in second quantized form, where the objective is to determine the ground state energy of the molecule. Wecker et al. [WBCT13] have developed a technique to obtain an estimate of the energy using only angles at most 10^{-6} accuracy in the phase estimation algorithm. Similarly, Jones et al. show how to optimize quantum chemistry simulations by ignoring terms with small norm [JWM⁺12]. They use Z -axis rotations with approximation accuracies in the range $\epsilon = 10^{-5}$. For such algorithms, our method produces rotations at the desired accuracy using extremely few T gates.

7 Conclusions and future work

We have presented a general framework of non-deterministic circuits called “Repeat-Until-Success” (RUS) circuits, and characterize unitaries which can be exactly represented by a RUS circuit. Traditional methods decompose single-qubit unitaries into deterministic sequences of gates. Wiebe and Kliuchnikov showed that by adding measurements and allowing non-deterministic circuits, decompositions with fewer T gates are possible (in expectation) for very small Z -axis rotations [WK13]. Our results extend that conclusion to arbitrary single-qubit unitaries. By synthesizing RUS circuits and then composing them, we can approximate arbitrary single-qubit unitaries to within a distance of 10^{-6} , which is sufficient for many quantum algorithms. Approximation accuracy can be improved by combining our circuits with those of [WK13]. For a random Z -axis rotation, our technique yields an approximation which requires as little as one-third as many T gates as [Sel12], [KMM12b], [RS14], and [Fow11]. Composing axial and non-axial RUS circuits yields even larger improvements in T count costs, where the approximation accuracy is limited by the size of the database.

Our results suggest a number of possible areas for further research. First, circuits of the form shown in Fig. 3 make up only a subset of possible RUS circuits. Expanding the search to include additional types of circuits could improve database density. Second, a formal number-theoretic characterization of RUS circuits needs to be made. A theoretical understanding could lead to efficient decomposition algorithms based on RUS circuits and allow for approximation to much smaller values of ϵ .

Extensions of the RUS circuit framework to multi-qubit unitaries or non-unitary channels should also be considered. In addition, we have restricted the setting to recovery operations that are Clifford operators. That restriction could be modified to allow for larger or alternative classes of operations. On the other hand, fault-tolerance schemes based on stabilizer codes often permit the application of Pauli operators [Kni05] at no cost. Thus, it might be sensible to limit recovery operations to only tensor products of Paulis.

Finally, the non-deterministic nature of RUS circuits imposes some additional constraints on the overall architecture of the quantum computer. Many fault-tolerance schemes already use non-deterministic methods to implement certain gates. But most of the non-determinism occurs “offline”, without impacting the computational data qubits. Since RUS circuits are “online”, the time required to implement a given unitary cannot be determined in advance. Such asynchronicity will require extensive placement and routing techniques and classical control logic. Architecture-specific analysis will be required in order to concretely assess the benefits of using RUS circuits.

Acknowledgements

The authors extend thanks to Vadym Kliuchnikov, Alex Bocharov, Nathan Wiebe, Yuri Gurevich, Andreas Blas, David Gosset and Cody Jones for helpful discussions, and to Dave Wecker for assistance with the implementation of the direct search. Thanks also to Robin Kothari for suggesting the amplitude amplification technique. AEP would like to thank Microsoft Research and the entire QuArC group for their hospitality.

References

- [AMMR12] Matthew Amy, Dmitri Maslov, Michele Mosca, and Martin Roetteler. A meet-in-the-middle algorithm for fast synthesis of depth-optimal quantum circuits. 2012, arXiv:1206.0758.
- [Amy13] Matthew Amy. *Algorithms for the Optimization of Quantum Circuits*. Master’s thesis, University of Waterloo, 2013.
- [AOK⁺10] Janet Anders, Daniel Kuan Li Oi, Elham Kashefi, Dan E. Browne, and Erika Andersson. Ancilla-Driven Universal Quantum Computation. *Physical Review A*, 82:020301, 2010, arXiv:0911.3783.
- [BCC⁺13] Dominic W. Berry, Andrew M. Childs, Richard Cleve, Robin Kothari, and Rolando D. Somma. Exponential improvement in precision for simulating sparse Hamiltonians. 2013, arXiv:1312.1414.
- [BGS13] Alex Bocharov, Yuri Gurevich, and Krysta M. Svore. Efficient Decomposition of Single-Qubit Gates into V Basis Circuits. *Physical Review A*, 88:012313, 2013, arXiv:1303.1411.
- [BHMT00] Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. Quantum Amplitude Amplification and Estimation. 2000, arXiv:0005055.
- [BS12] Alex Bocharov and Krysta M. Svore. A Depth-Optimal Canonical Form for Single-qubit Quantum Circuits. *Physical Review Letters*, 109:19050, 2012, arXiv:1206.3223.
- [CAB12] Earl T. Campbell, Hussain Anwar, and Dan E. Browne. Magic state distillation in all prime dimensions using quantum Reed-Muller codes. *Physical Review X*, 2:041021, 2012, arXiv:1205.3104.
- [DN05] Christopher M. Dawson and Michael A. Nielsen. The Solovay-Kitaev algorithm. *Quantum Information and Computation*, 6(1):81–95, 2005, arXiv:0505030.
- [DS12] Guillaume Duclos-Cianci and Krysta M. Svore. A State Distillation Protocol to Implement Arbitrary Single-qubit Rotations. *Physical Review A*, 88:042325, 2012, arXiv:1210.1980.
- [FDJ13] Austin G. Fowler, Simon J. Devitt, and Cody Jones. Surface code implementation of block code state distillation. *Scientific reports*, 3(1939), 2013, arXiv:1301.7107.

- [FH04] Austin G. Fowler and Lloyd C. L. Hollenberg. Scalability of Shor’s algorithm with a limited set of rotation gates. *Physical Review A*, 70:32329, 2004, arXiv:0306018.
- [Fow11] Austin G. Fowler. Constructing arbitrary Steane code single logical qubit fault-tolerant gates. *Quantum Information and Computation*, 11:867–873, 2011, arXiv:0411206.
- [GKMR13] David Gosset, Vadym Kliuchnikov, Michele Mosca, and Vincent Russo. An algorithm for the T-count. 2013, arXiv:1308.4134.
- [GN13] David Gosset and Daniel Nagaj. Quantum 3-SAT is QMA1-complete. 2013, arXiv:1302.0290.
- [GS12] Brett Giles and Peter Selinger. Exact synthesis of multi-qubit Clifford+T circuits. *Physical Review A*, 87, 032332, 2012, arXiv:1212.0506.
- [Jon13a] Cody Jones. *Logic synthesis for fault-tolerant quantum computers*. PhD thesis, Stanford University, 2013, arXiv:1310.7290.
- [Jon13b] Cody Jones. Low-overhead constructions for the fault-tolerant Toffoli gate. *Physical Review A*, 87, 022328, 2013, arXiv:1212.5069.
- [JWM⁺12] Cody Jones, James D. Whitfield, Peter L. McMahon, Man-Hong Yung, Rodney Van Meter, Alán Aspuru-Guzik, and Yoshihisa Yamamoto. Simulating chemistry efficiently on fault-tolerant quantum computers. *New Journal of Physics*, 14, 115023, 2012, arXiv:1204.0567.
- [Kit97] Alexei Y. Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.
- [Kli13] Vadym Kliuchnikov. Synthesis of unitaries with Clifford+T circuits. 2013, arXiv:1306.3200.
- [KMM12a] Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. Fast and efficient exact synthesis of single qubit unitaries generated by Clifford and T gates. *Quantum Information and Computation*, 13(7&8):607–630, 2012, arXiv:1206.5236.
- [KMM12b] Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. Practical approximation of single-qubit unitaries by single-qubit quantum Clifford and T circuits. 2012, arXiv:1212.6964.
- [Kni95] Emanuel Knill. Approximation by Quantum Circuits. Technical Report LAUR-95-2225, Los Alamos National Laboratory, 1995, arXiv:9508006.
- [Kni05] Emanuel Knill. Quantum Computing with Very Noisy Devices. *Nature*, 434(7029):39–44, 2005, arXiv:0410199.
- [KOB⁺09] Elham Kashefi, Daniel Kuan Li Oi, Daniel E. Browne, Janet Anders, and Erika Andersson. Twisted graph states for ancilla-driven quantum computation. *Proc. 25th Conference on the Mathematical Foundations of Programming Semantics (MFPS 25), ENTCS*, 249:307–331, 2009, arXiv:0905.3354.

- [KSV02] Alexei Y. Kitaev, Alexander H. Shen, and Mikhail N. Vyalyi. *Classical and Quantum Computation*. American Mathematical Society, Providence, RI, 2002.
- [LBK04] Yuan Liang Lim, Almut Beige, and Leong Chuan Kwek. Repeat-Until-Success Quantum Computing. *Physical Review Letters*, 95, 030505, 2004, arXiv:0408043.
- [NC00] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- [RHG07] Robert Raussendorf, Jim Harrington, and Kovid Goyal. Topological fault-tolerance in cluster state quantum computation. *New Journal of Physics*, 9(6):199–199, 2007, arXiv:0703143.
- [RS14] Neil J. Ross and Peter Selinger. Optimal ancilla-free Clifford+T approximation of z-rotations. 2014, arXiv:1403.2975.
- [Sel12] Peter Selinger. Efficient Clifford+T approximation of single-qubit operators. 2012, arXiv:1212.6253.
- [SO13] Kerem Halil Shah and Daniel Kuan Li Oi. Ancilla Driven Quantum Computation with arbitrary entangling strength. In *Proc. 8th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2013)*, 2013, arXiv:1303.2066.
- [WBCT13] Dave Wecker, Bela Bauer, Bryan Clark, and Matthias Troyer. In preparation. 2013.
- [WGMAG13] Jonathan Welch, Daniel Greenbaum, Sarah Mostame, and Alán Aspuru-Guzik. Efficient Quantum Circuits for Diagonal Unitaries Without Ancillas. 2013, arXiv:1306.3991.
- [WK13] Nathan Wiebe and Vadym Kliuchnikov. Floating point representations in quantum circuit synthesis. *New Journal of Physics*, 15:093041, 2013, arXiv:1305.5528.