Resource-Optimal Single-Qubit Quantum Circuits

Alex Bocharov and Krysta M. Svore

Quantum Architectures and Computation Group, Microsoft Research, Redmond, Washington 98052, USA

(Received 29 June 2012; revised manuscript received 4 October 2012; published 8 November 2012)

Determining the optimal implementation of a quantum gate is critical for designing a quantum computer. We consider the crucial task of efficiently decomposing a general single-qubit quantum gate into a sequence of fault-tolerant quantum operations. For a given single-qubit circuit, we construct an optimal gate sequence consisting of fault-tolerant Hadamard (H) and $\pi/8$ rotations (T). Our scheme is based on a novel canonical form for single-qubit quantum circuits and the corresponding rules for exactly reducing a general single-qubit circuit to our canonical form. The result is optimal in the number of T gates. We demonstrate that a precomputed epsilon net of canonical circuits in combination with our scheme lowers the depth of approximation circuits by up to 3 orders of magnitude compared to previously reported results.

DOI: 10.1103/PhysRevLett.109.190501 PACS numbers: 03.67.Ac, 03.65.Fd, 03.67.Lx

Introduction.—Quantum algorithms can be described by unitary transformations and projective measurements of a quantum state vector. A unitary transformation can be described by a sequence of unitary matrices, each of which we call a quantum gate. A sequence of one or more quantum gates is called a quantum circuit. A quantum circuit representing a quantum algorithm uses general quantum gates, despite potential challenges with their physical implementations. Therefore, a scalable quantum computer will require the processing of a general quantum gate into a fault-tolerant, implementable sequence of quantum gates. Various techniques for decomposing quantum gates into a sequence of gates drawn from an implementable set, called a discrete gate set, are known [1–5]. However, it is crucial that the resulting gate sequence be an optimal implementation in which resources such as circuit depth, the number of gates, or the number of qubits are minimized. Achieving lower complexity gate sequences is necessary in order to achieve shorter execution time as well as a smaller probability of error.

In this Letter, we address the challenge of optimally decomposing quantum circuits that act on a single qubit. To produce optimal gate sequences, we derive a *canonical form* for single-qubit unitaries and corresponding rules for reducing a single-qubit circuit into our canonical form. We then develop an algorithm for finding an exact, resource-optimal decomposition of a single-qubit gate, if it exists; if it does not exist, our algorithm finds an approximation with precision ϵ that significantly reduces the resource cost of the circuit. We choose to decompose into the gate library of Hadamard (H) and $\pi/8$ (T) rotations, denoted as $\{H, T\}$, since these gates can be implemented fault tolerantly in the Steane code [6] and the surface code [7], both of which have been shown to yield high error thresholds. We choose to minimize the number of T gates, called the T count of the sequence, since the fault-tolerant implementation of T is significantly more expensive than the H gate. Our approach simultaneously reduces circuit depth.

Background.—Decomposition of a single-qubit quantum circuit most often results in a gate sequence that is approximately equal to the original gate, while exact equivalence is achieved in rare cases. When exact decomposition is possible, Amy *et al.* [8] provide an algorithm for decomposing the quantum circuit into a depth-optimal gate sequence in time $O(d|\mathcal{B}|^{d/2})$, where d is the circuit depth and $\mathcal{B}$ is the basis of gates. For single-qubit circuits, our technique improves the runtime and finds an exact decomposition in time $O(d + |\mathcal{B}|^{d/4})$, for $\mathcal{B} = \{H, T\}$.

When exact equivalence is not possible or when the circuit must be resource optimized at the expense of precision, the Solovay-Kitaev theorem [4] guarantees that any single-qubit circuit can be approximated to precision ϵ with a gate sequence of depth $\Theta(\log^c(1/\epsilon))$, where c is a small constant. Dawson and Nielsen [9] developed an algorithm to find an approximation with precision ϵ in time $O(\log^{2.71}(1/\epsilon))$. Their algorithm begins with a base approximation to a single-qubit circuit and proceeds recursively n times, resulting in a circuit that grows in depth as $O(5^n)$, where n depends on the desired level of precision. Optimizing the base approximation is especially important since the resulting circuit heavily depends on the base circuit; if a base circuit has suboptimal cost, then this inefficiency is amplified upon recursion. Fowler gives an exponential-time algorithm (albeit much faster than brute-force search) for improving the base circuit [10,11] that finds its depth-optimal ϵ-approximation.

Our optimization scheme is based on a canonical form for single-qubit circuits and can be used for the base approximation in the Dawson-Nielsen algorithm. Our canonical form is similar in spirit to the normal form for single-qubit circuits over the $\{H, T\}$ gate library given by Matsumoto and Amano [12]. However, their normal form is expressed in $SU(2)$, which contains a nontrivial two-element center that makes the algebra sensitive to the sign of the global phase. In contrast, our canonical form uses group identities in the projective special unitary group

$PSU(2)$, allowing further optimization of quantum circuits. More formally, it is a unique representative of a double coset of circuits with respect to the Clifford group. By expressing a circuit in its canonical form, we can compress its circuit depth and minimize its T count.

A canonical form and canonical reduction of circuits.— We introduce our canonical form and rules for reducing a single-qubit quantum circuit to its canonical form. Throughout, we use $\cdot$ to represent gate composition. We use $\{\cdot\}$ to indicate the basis elements of a group and $\langle\cdot\rangle$ to indicate the group generated by those elements, where here elements are single-qubit quantum gates.

We begin with $PSU(2)$ representations of the Hadamard gate H and the $\pi/8$-gate T,

$$H = \begin{bmatrix} i/\sqrt{2} & i/\sqrt{2} \\ i/\sqrt{2} & -i/\sqrt{2} \end{bmatrix}, \qquad T = \begin{bmatrix} e^{-i\pi/8} & 0 \\ 0 & e^{+i\pi/8} \end{bmatrix}.$$

The phase gate $S = T^2$ and the Hadamard gate H together generate a 24-element subgroup in $PSU(2)$. We denote this group, the Clifford group, as $\mathcal{C}$.

We introduce the following two circuits, which we call syllables, each of which are composed of two quantum gates: $TH = T.H$ and $SH = S.H$. In $PSU(2)$, syllable TH is a group element of infinite order (see Sec. 4.5.3 in Ref. [13]), whereas syllable SH is a group element of order 3, $SH.SH.SH = (SH)^3 = I$.

Consider the set of all circuits generated by various compositions of TH and SH. We note that the basis of gates $\{TH, SH\}$ is equivalent to the basis of gates $\{H, T\}$ due to the following identities:

$$H = TH(SH)^2TH; \qquad T = (TH)^2(SH)^2TH.$$

Since $SH^3 = I$, any circuit in $\langle TH, SH\rangle$ with subsequences $(SH)^k$, where $k > 2$, can be immediately reduced to a circuit limited to $k = 1$ or 2. Furthermore, since $TH(SH)^2TH = H$, it is intuitively clear that $(SH)^2$ should not occur in a well-formed circuit. We further find that even single occurrences of SH can be algebraically removed from the initial segment of a circuit.

Definition. A nonempty circuit in $\langle TH, SH\rangle$ is said to be *normalized* if it ends with TH and does not explicitly contain $(SH)^2$. A normalized circuit is either the identity I or a nonempty normalized circuit.

In other words, a normalized circuit is either the identity I or follows one of two patterns, $n.TH$ or $n.SHTH$, where n is a shorter normalized circuit.

Definition. A normalized circuit is said to be *canonical* if it does not contain SH earlier than the fifth syllable.

The shortest canonical circuit that contains SH is $(TH)^4.SH.TH$.

Proposition 1. Each $\langle H, T\rangle$ circuit U can be represented as either $U = n.g$ or $U = H.n.g$, where n is a normalized circuit and $g \in \mathcal{C}$. This representation is built at a cost linear in the length of U.

The proof of this proposition is based on the H, S representation of the $\mathcal{C}$ group. See Supplemental Material at Ref. [14] for the proof, the H, S representation of $\mathcal{C}$, and a set of $\mathcal{C}/T$ commutation relations.

It is important to note that, as per the $SH.T = H.SH.TH.S.SH$ relation, given a normalized circuit n starting with TH, $SH.n$ can be rewritten as $H.SH.n'$ where n' is a normalized circuit.

Proposition 2. Each $\langle H, T\rangle$ circuit U can be represented as $U = g_1.c.g_2$, where c is a canonical circuit and $g_1, g_2 \in \mathcal{C}$. This representation is built at a cost linear in the length of U.

Proof. The proof is based on special $PSU(2)$ relations. See Supplemental Material at Ref. [14] for the list of relations. These relations imply that a normalized circuit with T count $0 < t \leq 4$ can be written as $n = g_1.(TH)^{t-1}.T.g_2$, where $g_1, g_2 \in \mathcal{C}$.

Consider $U = [H.][SH.]n.g$, where n is a normalized circuit starting with the TH syllable. The $[\cdot]$ indicates two possible cases: presence or absence of the term. We assume that the normalized circuit n has T count $t > 4$. Consider the shortest prefix of the circuit n spanned by its leftmost four TH syllables and apply one of the transformations from the list of relations to that prefix, thus obtaining the reduction $U = g_1.(TH)^3.T.(g'.H).n'.g$, where $g_1, g \in \mathcal{C}$, n' is a normalized circuit, T count$(n') > 0$, and $g' \in \{I, X, Z, S, S^\dagger, Z.X, S.X, S^\dagger.X\}$.

We consider the (Proposition 1) normalization V of $(g'.H).n'.g$, rewritten to start with either H or TH.

In the case where n' starts with TH, V can start with TH only if $(g'.H)$ commutes with T. In the case where n' starts with $SH.TH$, V can start with TH only if $(g'.H.SH)$ commutes with T. By inspection of the commutation table, we conclude that neither of these can happen and therefore V cannot start with TH.

Thus, V starts with H and $g_1.(TH)^3.T.V$ is the desired canonical form. ■

Definition. The T count of a normalized circuit n is the number of TH syllables in n.

T count is an invariant of the gate represented by a canonical circuit, which follows from Theorem 1 below. The importance of canonical circuits stems from the fact that they enumerate the double $\mathcal{C}$ cosets of $\langle H, T\rangle$, which also follows from Theorem 1:

Theorem 1. If c_1, c_2 are $\mathcal{C}$-equivalent canonical circuits, i.e., $\exists g_1, g_2 \in \mathcal{C}$ such that c_2 and $g_1.c_1.g_2$ evaluate to the same gate in $PSU(2)$, then c_1 and c_2 are equal as $\langle TH, SH\rangle$ circuits.

We outline a proof of this theorem.

Proof. We prove by contradiction that if a normalized circuit n evaluates to a circuit $g \in \mathcal{C}$, then $n = g = I$. Let n be a normalized circuit with T count$(n) > 0$.

Consider the adjoint action of $PSU(2)$ on its Lie algebra $\mathcal{L} = su(2)$, where $ad_u[m] = u.m.u^\dagger$, $u \in PSU(2)$, $m \in \mathcal{L}$. $\mathcal{L}$ is spanned over $\mathbb{R}$ by the Pauli matrices X, Y, Z.

The adjoint action of the $\mathcal{C}$ subgroup on $\mathcal{L}$ is the symmetry group of the octahedron with vertices at $\pm X$, $\pm Y$, $\pm Z$. In particular, for each $g \in \mathcal{C}$, $ad_g[Z]$ must be one of these vertices. To obtain a contradiction, it suffices to show that for a nonidentity normalized circuit n, $ad_n[Z]$ cannot be in $\{\pm X, \pm Y, \pm Z\}$.

Let $A \in \mathcal{L}$ be a matrix over $\mathbb{Z}[\frac{1}{\sqrt{2}}]$ represented as

$$(\sqrt{2})^l A = (x_0 + x_1\sqrt{2})X + (y_0 + y_1\sqrt{2})Y + (z_0 + z_1\sqrt{2})Z,$$

where x_0, x_1, y_0, y_1, z_0, z_1 are integers.

We show that if $A = ad_n[Z]$, then (1) x_0 is odd and (2) y_0, z_0 have the opposite parity. Property (1) implies that the coefficient at X is nonzero, and property (2) implies that at least one other coefficient (at Y or at Z) is nonzero; together they imply that $ad_n[Z]$ cannot be proportional to any one Pauli matrix.

We prove the desired properties (1) and (2) by induction on the T count of n. By direct computation,

$$\begin{aligned} ad_{TH}[X] &= Z, \\ ad_{TH}[Y] &= (X - Y)/\sqrt{2}, \\ ad_{TH}[Z] &= (X + Y)/\sqrt{2}, \\ ad_{SHTH}[X] &= Y, \\ ad_{SHTH}[Y] &= (-X + Z)/\sqrt{2}, \\ ad_{SHTH}[Z] &= (X + Z)/\sqrt{2}, \end{aligned}$$

and, in particular, properties (1) and (2) hold for $ad_{TH}[Z] = (X + Y)/\sqrt{2}$ ($x_0 = 1$, $y_0 = 1$, $z_0 = 0$).

Given matrix $A \in \mathcal{L}$ presented as above, we have

$$\begin{aligned} (\sqrt{2})^{l+1} ad_{TH}[A] = {} & [(y_0 + z_0) + (y_1 + z_1)\sqrt{2}]X \\ & + [(z_0 - y_0) + (z_1 - y_1)\sqrt{2}]Y \\ & + (2x_1 + x_0\sqrt{2})Z, \\ (\sqrt{2})^{l+1} ad_{SHTH}[A] = {} & [(z_0 - y_0) + (z_1 - y_1)\sqrt{2}]X \\ & + (2x_1 + x_0\sqrt{2})Y + [(y_0 + z_0) \\ & + (y_1 + z_1)\sqrt{2}]Z. \end{aligned}$$

By induction, y_0, z_0 have opposite parity; therefore, the new x_0 that is equal to either $y_0 + z_0$ or $z_0 - y_0$ is odd in both cases. In the expression for $ad_{TH}[A]$, the new $y_0' = z_0 - y_0$ is odd, but the new $z_0' = 2x_1$ is even. In the expression for $ad_{SHTH}[A]$, the new $y_0' = 2x_1$ is even, but the new $z_0' = y_0 + z_0$ is odd.

Since each nontrivial normalized circuit is either $n_1.TH$ or $n_1.SHTH$, where n_1 is a shorter normalized circuit, this concludes the induction. ■

The remainder of the proof is based on the following two lemmas:

Lemma 1. Let n be a normalized circuit that is either I or starts with TH. Let u be any $\langle TH, SH\rangle$ circuit that evaluates to the same gate as n. Then n is equal to the normalization of u.

Lemma 2 [15]. Let n be a normalized circuit, and let g_1, $g_2 \in \mathcal{C}$. If both n and normalization of $g_1.n.g_2$ start with $(TH)^4$, then $g_1 = I$.

We note that Theorem 1 is trivially true for canonical circuits of T count ≤ 4. Thus. we assume that both c_1 and c_2 start with $(TH)^4$.

By Lemma 1, c_2 is equal to the normalization of $g_1.c_1.g_2$, and hence g_1 is the identity per Lemma 2. Since $c_1.g_2$ is its own normalization, $c_2 = c_1.g_2$. Finally, g_2 is a suffix of c_2 and thus a normalized circuit of T count zero, i.e., $g_2 = I$.

In our proposed canonical form, we have minimized the T count of the circuit. The T count and circuit depth are closely tied; e.g., in an $\{H, T\}$ canonical form there are at least T count -1 and at most T count $+1$ Clifford gates in the sequence, and all but at most two of these gates are either H or $HSH = \sqrt{X}$. Thus, when minimizing T count we in turn also optimize for circuit depth.

Approximation of single-qubit circuits.—We can use our canonical form and reduction rules to produce an optimized gate sequence that approximates a single-qubit circuit to precision ϵ. First, we build a database of canonical circuits by iterating over T count; in practice, the database can be used to perform exact or approximate circuit decomposition. Throughout, we compute precision ϵ using the trace distance between single-qubit circuits U and V, although different distance measures can be used:

$$\mathrm{dist}\,(U, V) = \sqrt{\frac{2 - |\mathrm{tr}(U.V^\dagger)|}{2}}.$$

The following remarkable observation leads to an efficient algorithm for finding an ϵ approximation:

Corollary 1. Given a single-qubit gate $U \in PSU(2)$, U can be ϵ approximated with an $\langle H, T\rangle$ circuit with T count $<t$ if and only if one of the gates in the double coset $\mathcal{C}.U.\mathcal{C} = \{g_1.U.g_2 | g_1, g_2 \in \mathcal{C}\}$ can be ϵ approximated by a canonical circuit with T count $<t$.

It follows that the optimal ϵ approximation of U under a certain T count t is immediately derived from the optimal ϵ approximation of *some* gate $G \in \mathcal{C}.U.\mathcal{C}$ under T count t. There, at most $|\mathcal{C}|^2 = 576$ gates in $\mathcal{C}.U.\mathcal{C}$, and for the majority of targets U the double coset consists of small groups of elements at significant distance from each other. Thus, distinct values in $\{|\mathrm{tr}(U.g)| | g \in \mathcal{C}\}$ typically differ by more than 8ϵ, which suggests that the search for an ϵ approximation can be run on parallel threads, each searching over $O(\epsilon 2^{t/2})$ candidates.

This is a foundation for an efficient implementation of an ϵ net of $\langle H, T\rangle$ circuits (described in detail in Ref. [16]). We have built such an ϵ net for $\epsilon = 0.002$ with memory footprint below 45 MB, which includes circuits with T count <26. The use of such an ϵ net within an algorithm for Solovay-Kitaev decomposition dramatically improves the precision and resource cost of the output circuit.

Recall that the Dawson-Nielsen (D-N) algorithm to perform Solovay-Kitaev decomposition [9] is recursive, and finer approximations require greater recursion depth. At depth 0, D-N returns an extrinsic "basic" approximation of a single-qubit gate U. At depth n, it composes an approximation from the depth $n-1$ approximation U_{n-1} and the depth $n-1$ approximations of two auxiliary matrices V_{n-1} and W_{n-1} such that the resulting n approximation is given by

$$U_n = V_{n-1}.W_{n-1}.V_{n-1}^{\dagger}.W_{n-1}^{\dagger}.U_{n-1}. \tag{1}$$

According to the D-N estimate (Sec. 3, Eq (1) in Ref. [9]), tightening the basic approximation precision by a factor of 10 should lead to precision improvement by a factor close to 10^{-6} at recursion depth 4. Experimental evaluation of our implementation based on the above ϵ net with T count <26 agrees with this estimate and delivers approximation precisions in the 10^{-8} range at recursion depth 3. Evaluating approximations at higher recursion depth requires extended arithmetic beyond machine-defined double type.

We use our database and canonical form scheme to obtain an optimized base approximation and compare the T count and precision of D-N using our base approximation and the original base approximation [9]. We created three canonical circuit databases of varying size from which to determine the optimized base approximation; they store circuits up to T count 24, 25, and 26. Figure 1 plots the T count versus the mean precision ϵ, where the mean is calculated over the approximation of 10 000 random unitary gates, and the points indicate recursion levels $n = 0, 1, 2, 3$, for our method and the baseline. Both axes in the graph are plotted on the logarithmic scale.

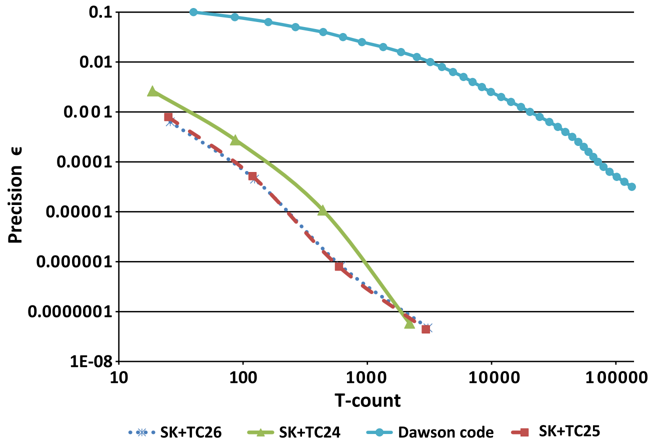

FIG. 1 (color online). T count versus mean precision ϵ (trace distance) of the approximation of 10 000 random unitaries, for recursion levels $n = 0, 1, 2, 3$. The markers indicate recursion level n.

First, we remark that there is no visible distinction between the performance of databases of size T count 25 and 26. Second, for a given precision ϵ, we produce gate sequences with up to 3 orders of magnitude fewer T's than the baseline approach. For example, at precision $\epsilon = 5 \times 10^{-4}$, we achieve a T count on the order of 10^2, whereas the baseline is on the order of 10^5. Third, for a given T count, we achieve gate sequences that are up to 3 orders of magnitude more precise than the baseline approach. At T count 100, our scheme produces a gate sequence with precision around 10^{-4}, whereas the baseline has precision around 10^{-1}. These improvements are crucial when considering that optimization of the depth and T count significantly affects the execution time and number of errors in a physical implementation of the circuit. With our optimization scheme, previously infeasible quantum circuits may be one step closer to being implemented in a physical device.

Another application of our database and canonical form is to consider "lossy" circuit optimization, where the task is to determine if for a slight decrease in precision ϵ there exists an ϵ-approximate circuit that requires even fewer resources. We can use a database of canonical circuits to answer this question, and the complexity of searching this database is significantly reduced due to the following theorem:

Theorem 2. Canonical circuits with distinct T counts evaluate to unitary matrices with distinct matrix traces.

See Supplemental Material at Ref. [14] for the proof sketch. The theorem implies that if a trace level $L_t = \{U \in PSU(2) || \mathrm{tr}(U)| = t\}$ contains several canonical circuits, all of these circuits have the same T count. It reduces the question to searching over different trace levels for a more optimized circuit. The implication of such a search for optimized circuits is an open research question.

Conclusion.—In summary, we have defined a resource-optimal canonical form and corresponding rules to reduce a single-qubit quantum circuit to our canonical form. Given a single-qubit circuit, our scheme can be used to produce a gate sequence that is exactly equivalent and uses a minimal number of T gates, or it can be used to determine a resource-optimal base approximation. Our database of canonical circuits is significantly smaller in memory than previous methods [10,11]. When using our technique within the D-N algorithm, we achieve up to 3 orders of magnitude improvement in both precision and T count over the baseline. One future direction, as previously mentioned, is to consider lossy circuit optimization. Another future direction is to generalize the definition of a canonical form to other libraries of gates and, furthermore, to extend the form to n-qubit circuits.

We thank Dave Wecker, Burton Smith, Michael Freedman, Zhenghan Wang, John Platt, Austin Fowler, and Peter Selinger for useful discussions. We also thank Rodney Van Meter and Nathan Cody Jones for sharing the baseline D-N data.

[1] P. O. Boykin, T. Mor, M. Pulver, V. Roychowdhury, and F. Vatan, arXiv:quant-ph/9906054.

[2] A. Barenco, C. Bennett, R. Cleve, D. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. Smolin, and H. Weinfurter, Phys. Rev. A **52**, 3457 (1995).

[3] A. Kitaev, Russ. Math. Surv. **52**, 1191 (1997).

[4] A. Kitaev, A. Shen, and M. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, Providence, 2002).

[5] N. C. Jones, J. D. Whitfield, P. L. McMahon, M. Yung, R. van Meter, A. Aspuru-Guzik, and Y. Yamamoto, arXiv:1204.0567v1.

[6] P. Aliferis, D. Gottesman, and J. Preskill, Quantum Inf. Comput. **6**, 97 (2006).

[7] A. G. Fowler, A. M. Stephens, and P. Groszkowski, Phys. Rev. A **80**, 052312 (2009).

[8] M. Amy, D. Maslov, M. Mosca, and M. Roetteler, arXiv:1206.0758.

[9] C. Dawson and M. Nielsen, Quantum Inf. Comput. **6**, 81 (2006).

[10] A. G. Fowler, arXiv:quant-ph/0411206.

[11] A. Fowler, Ph.D. thesis, University of Melbourne, 2005.

[12] K. Matsumoto and K. Amano, arXiv:0806.3834.

[13] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).

[14] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.109.190501 for the reference multiplication tables referred in this Letter, as well as detailed proofs of Proposition 1 and Theorem 2.

[15] P. Selinger (private communication).

[16] A. Bocharov and K. M. Svore (to be published).